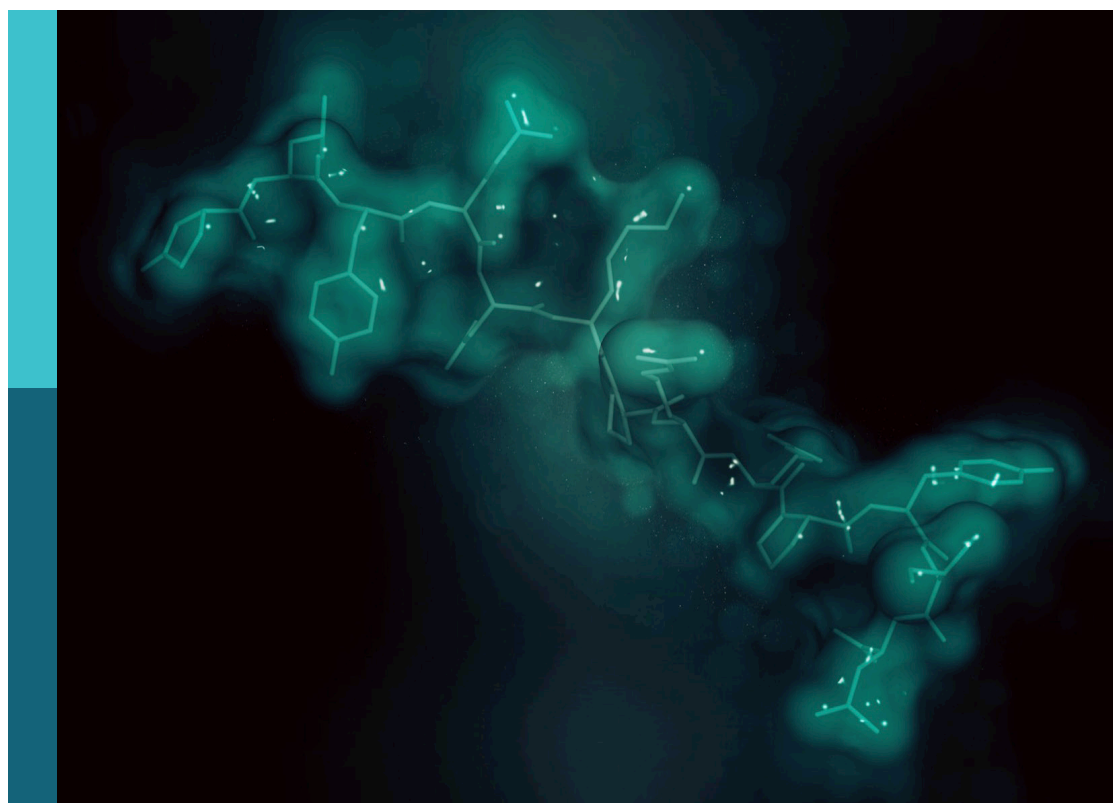# Recent highlights in molecular medicine

Edited by
Ignazio Castagliuolo, Frank Emmert-Streib, Leon J. De Windt,
Vincenzo Cerullo and Giorgio Stassi

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Recent highlights in molecular medicine

Topic editors

Ignazio Castagliuolo — University of Padua, Italy
Frank Emmert-Streib — Tampere University, Finland
Leon J. De Windt — Maastricht University, Netherlands
Vincenzo Cerullo — University of Helsinki, Finland
Giorgio Stassi — University of Palermo, Italy

Table of
contents

Check for updates

*CORRESPONDENCE
Alberto Amaretti,
alberto.amaretti@unimore.it
Stefano Raimondi,
stefano.raimondi@unimore.it

Indole and p-cresol in feces of healthy subjects: Concentration, kinetics, and correlation with microbiome

Francesco Candeliere[1], Marta Simone[1], Alan Leonardi [1], Maddalena Rossi[1,2], Alberto Amaretti[1,2]* and Stefano Raimondi[1,2]*

[1]Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy, [2]Biogest Siteia, University of Modena and Reggio Emilia, Reggio Emilia, Italy

Indole and p-cresol are precursors of the most important uremic toxins, generated from the fermentation of amino acids tryptophan and tyrosine by the proteolytic community of intestinal bacteria. The present study focused on the relationship between the microbiome composition, the fecal levels of indole and p-cresol, and their kinetics of generation/degradation in fecal cultures. The concentration of indole and p-cresol, the volatilome, the dry weight, and the amount of ammonium and carbohydrates were analyzed in the feces of 10 healthy adults. Indole and p-cresol widely differed among samples, laying in the range of 1.0–19.5 μg/g and 1.2–173.4 μg/g, respectively. Higher fecal levels of indole and p-cresol were associated with lower carbohydrates and higher ammonium levels, that are markers of a more pronounced intestinal proteolytic metabolism. Positive relationship was observed also with the dry/wet weight ratio, indicator of prolonged intestinal retention of feces. p-cresol and indole presented a statistically significant negative correlation with OTUs of uncultured Bacteroidetes and Firmicutes, the former belonging to *Bacteroides* and the latter to the families Butyricicoccaceae (genus *Butyricicoccus*), Monoglobaceae (genus *Monoglobus*), Lachnospiraceae (genera *Faecalibacterium*, *Roseburia*, and *Eubacterium ventriosum* group). The kinetics of formation and/or degradation of indole and p-cresol was investigated in fecal slurries, supplemented with the precursor amino acids tryptophan and tyrosine in strict anaerobiosis. The presence of the precursors bursted indole production but had a lower effect on the rate of p-cresol formation. On the other hand, supplementation with indole reduced the net rate of formation. The taxa that positively correlated with fecal levels of uremic toxins presented a positive correlation also with p-cresol generation rate in biotransformation experiments. Moreover other bacterial groups were positively correlated with generation rate of p-cresol and indole, further expanding the range of taxa associated to production of p-cresol (*Bacteroides*, *Alistipes*, *Eubacterium xylanophylum*, and *Barnesiella*) and indole (e.g., *Bacteroides*, *Ruminococcus torques*, *Balutia*, *Dialister*, *Butyricicoccus*). The information herein presented contributes to disclose the relationships between microbiota composition and the production of uremic toxins, that could provide the basis for probiotic intervention on the

gut microbiota, aimed to prevent the onset, hamper the progression, and alleviate the impact of nephropaties.

Introduction

Indole and p-cresol are precursors of toxic metabolites, also referred to as uremic toxins, that are eliminated by healthy kidneys *via* the urine and, in case of renal disease, are related to many complications (Vanholder et al., 2018). Among many uremic toxins, indole and p-cresol present the common feature to be products of bacterial metabolism that occurs mostly in the colon by members of gut microbiota.

Indole derives from metabolism of the essential amino acid L-tryptophan by bacterial tryptophanase. Together with its derivatives indole propionic acid, indole acetic acid, skatole, and tryptamine, are potent bioactive metabolites that affect intestinal barrier integrity and immune cells. Intestinal bacteria can directly influence the type and level of tryptophan-derived metabolites, that participate to intracellular signaling within the microbial community (Lee and Lee, 2010) and can activate pregnane X and aryl hycrocarbon receptors, affecting the mucosal immune response and integrity (De Juan and Segura, 2021). Indoxyl sulphate acts as a protein-bound uremic toxin, whose excretion is reduced in patients with Chronic kidney disease (CKD), with severe consequences in terms of systemic toxicity. p-Cresol is the product of the breakdown of tyrosine by intestinal bacteria (Scott et al., 2013; Saito et al., 2018). It.is another protein-bound uremic toxin accumulating in the body of patients with compromised renal function. For both indole and p-cresol, due the cytotoxic effects, CKD and patients in hemodialysis are at increased risk of vascular damage, morbidity, and mortality.

Humans and the gut microbiota evolved together, establishing a close symbiotic interrelationship, fruitful for both. The human colonic microbiota is a dense and complex community of commensal microbes, mostly bacteria, implicated in a number of biological processes such as resistance to colonization (Ruan et al., 2020), immune system modulation (Saldana-Morales et al., 2021), synthesis of essential vitamins and nutrients (Oliphant and Allen-Vercoe, 2019), and breakdown of undigested polysaccharides and proteins (El Kaoutari et al., 2013; Huang et al., 2017; Raimondi et al., 2021). Gut microbes encode a broad diversity of enzymes capable of processing foreign compounds (e.g., dietary compounds, phytochemicals, environmental pollutants, pharmaceuticals, and other xenobiotics), among which those that transform intermediates of protein breakdown into indole, p-cresol, and derivatives, but also those that can participate to degradation of these uremic toxins (Saito et al., 2018; Popkov et al., 2022).

The microbial composition of intestinal microbiota presents a wide diversity in terms of species composition, and, as a consequence, of metabolic potential. Different microbiota is expected to impact on abundance and diversity of enzymes involved in both biosynthesis and degradation of uremic toxins, with major systemic effects associated to different level of these circulating metabolites and, as a matter of fact, alterations in the gut microbiota in patients with CKD can contribute to CKD progression through different mechanisms, including the increased production of microbiota-derived uremic solutes (Vaziri et al., 2013; Wong et al., 2014; Caggiano et al., 2020; Kim and Song, 2020). Nonetheless, these uremic toxins reach steady levels in healthy subjects, where they have been hypothesized to exert also beneficial effects for body functioning (Vanholder et al., 2022).

Microbe-related levels of indole and p-cresol are expected to vary between individuals, due to specificities in gut microbiota composition and dietary patterns (Marcobal et al., 2013). Albeit the era of genomics, post-genomics, and metagenomics encourages *in silico* studies that extract information on microbiota role and activities from the huge amount of available data, still functional observations can be obtained by biotransformations studies with fecal cultures (Tomas-Barberan et al., 2014; Amaretti et al., 2015; Quartieri et al., 2016a).

Little is known about differences of gut microbial populations in the ability to produce and degrade uremic toxins, and a rigorous investigation focused on these biotransformations has not yet been published. The present study seeks to fill this gap. The efficiency of transformation of the precursors tryptophan and tyrosine into indole and p-cresol, and of the degradation of these uremic toxins was compared in fecal cultures inoculated with different intestinal populations provided by healthy subjects. The study was integrated with bacterial composition of fecal inoculum by16 S rRNA gene profiling, and by characterization of fecal samples in terms of dry and wet weight, pH, levels of indole, p-cresol, ammonium, total carbohydrates, and soluble carbohydrates.

Materials and methods

Fecal samples and slurries preparation

Fecal samples were obtained from ten healthy adults (five men and five women aged 25–50 years) that provided written informed consent, according to the experimental protocol that was approved

with ref. No. 125–15 by the local research ethics committee (Comitato Etico Provinciale, Azienda Policlinico di Modena, Italy). Subjects did not take prebiotics and/or probiotics for the previous 2 weeks or antibiotics for at least 3 months prior to sample collection. Hereinafter, the subjects are referred to as V1–V10. V1 provided 5 fecal samples (referred to as V1a–e), longitudinally collected across a time span of 4 months and spaced at least 7 days from each other. V2 to V10 provided only one sample each.

Fecal samples were collected fresh, sealed within anaerobic plastic bags (AnaeroGen, Oxoid, Basingstoke, United Kingdom), and transferred in a glove box (Concept Plus, Ruskinn Technology, Ltd., Bridgend, United Kingdom) with an atmosphere of 85% N_2, 10% CO_2, and 5% H_2. Within the anerobic cabinet, the feces were homogenized (10%, w/v) in sterile phosphate buffered saline pH 6.5 (PBS, Na_2HPO_4 1 g/L, KCl 0.1 g/L, KH_2PO_4 0.8 g/L, NaCl 8 g/L) or sterile water. PBS suspensions were utilized to for bioconversion experiments. Water suspensions were utilized for the analysis of volatilome, indole and p-cresol, total and soluble carbohydrates, dry weight, pH, and ammonium.

All the chemicals were purchased from Sigma-Aldrich (Merck KGaA, Darmstadt, Germany), unless otherwise stated.

Chemical analysis

The volatilome was determined by solid-phase microextraction (SPME) followed by GC–MS analysis. 2 mL of water suspension of feces were supplemented with 10 μL of 10 M of HCl within a 10-mL vial, the headspace of which was put in contact with a divinylbenzene/carboxen/polydimethylsiloxane fiber (DVB/CAR/ PDMS Supelco; Sigma-Aldrich, St. Louis, MO, United States) for 1 h at 60°C. The VOCs were desorbed and separated in a GC–MS apparatus (7820–5975; Agilent Technologies, Santa Clara, CA, Untied States) equipped with a DB-624 column (30 m $\times$ 250 mm $\times$ 1.4 mm, Agilent Technologies), according to the conditions reported by Raimondi et al. (2019). The identity of the compounds was inferred from the comparison of their mass spectra with NIST 14 spectral library. The compounds whose identity designation presented a quality score > 95% were listed in the volatilome. The peak areas of compounds were taken to be proportional to their abundances. Ammonium was quantified in samples supernatant using the phenol nitroprusside colorimetric assay, based on the formation of the indophenol blue (Amaretti et al., 2019). Total and soluble carbohydrates were analyzed with anthrone colorimetric assay (Amaretti et al., 2019). Indole and p-cresol were quantified by HPLC (1100 System, Agilent Technologies), equipped with a diode array detector and a C18 Kinetex column (Phenomenex, Torrance, CA, United States) as reported by Amaretti et al. (2019). The dry weight of fecal samples, expressed as the g of dry matter per g of wet feces, was measured with a moisture analyser MBG (VWR International s.r.l., Italy).

Bioconversion experiments

PBS fecal suspensions were supplemented with the precursor amino acids tryptophan and tyrosine (TT slurries) or the products indole and p-cresol (IP slurries) at the concentration of 200 mM, and were incubated anaerobically at 37 °C for 24 h. Control slurries (C) without any supplement were incubated under the same conditions. Samples were collected at 0, 4, 8, and 24 h of incubation to analyze indole and p-cresol concentration by HPLC.

16 S rRNA gene profiling

Total DNA was extracted from 2 mL of water suspension utilizing the kit QiAmp PowerFecal DNA (Qiagen, Hilden, Germany), according to the manufacturer's protocol. DNA concentration was adjusted to 5 ng/μL after quantification with a Qubit 3.0 fluorimeter (Thermo Fisher Scientific, Waltham, MA, United States). The regions V1-V3 of 16 S rRNA gene sequences were sequenced by Eurofins Genomics (Ebersberg, Germany) using a MiSeq platform (Illumina Inc., San Diego, CA, United States). Raw sequences were analyzed with QIIME 2.0 version 2021.4 (Bolyen et al., 2019), with CUTADAPT and DADA2 plugins for trimming and denoising into amplicon sequence variants (ASVs) (Martin 2011; Callahan 2016). Taxonomy was assigned with VSEARCH (Rognes et al., 2016), utilizing as reference database SILVA SSU release 138 (https://www.arb-silva.de/download/arb-files/) with similarity threshold set at 0.97.

The appropriate QIIME2 plugins were utilized to compute the alpha- (Chao1, Shannon, and Pielou's evenness) and beta-diversity (Bray-Curtis and Jaccard). Principal Coordinate Analysis (PCoA) was computed with QIIME2, based on the beta-diversity distance matrices.

Correlation between bacterial taxa and VOCs and between taxa and indole/p-cresol transformation values were determined by computing Spearman correlation coefficients ($p < 0.05$) using R package psych. Cytoscape was utilized to visualize correlation networks (Shannon et al., 2003).

PICRUSt2 (Douglas et al., 2020) predicted the abundance of key enzymes in indole e p-cresol biosynthesis, tryptophanase [EC:4.1.99.1] (Yanofsky, 2007) and 2-iminoacetate synthase [ThIH; EC4.1.99.19] and 4-hydroxyphenylacetate decarboxylase [HpdB; EC:4.1.1.83] (Saito et al., 2018; Challand et al., 2010), respectively. Pearson correlation ($p < 0.05$) between the predicted abundances and the and indole/p-cresol transformation values were computed with R package psych.

Aliquots of 5 mL from a fresh culture (16 h) of *B. pseudocatenulatum* WC 0408 were centrifuged and the pellets were washed with PBS buffer pH 6.5. The washed pellets were then resuspended 1:1 in PBS buffer (pH 6.5) supplemented with 1 mM indole. Resting cells were incubated at 37 °C in anaerobic conditions for 48 h and the supernatant was immediately frozen at −20 °C.

TABLE 1 Chemical characterization of fecal samples collected from 10 different volunteers. Feces of Volunteer 1 were collected five times in a period of 4 month (V1a-e). The values are means of 3 replicated measurements ± standard deviation and refer to the wet weight of the feces.

Sample	Dry weight/Wet weight	pH	NH_4^+	Total carbohydrates (mg/g_{wet})	Soluble carbohydrates	p-cresol $(\mu g/g_{wet})$	Indole $(\mu g/g_{wet})$
V1a	0.27 ± 0.02	6.2 ± 0.1	0.42 ± 0.02	6.7 ± 0.7	1.1 ± 0.1	78.6 ± 0.3	17.3 ± 0.2
V1b	0.28 ± 0.02	6.5 ± 0.2	0.29 ± 0.03	12.5 ± 1.5	2.7 ± 0.8	35.6 ± 0.9	27.5 ± 0.6
V1c	0.31 ± 0.00	7.2±0.1	0.43±0.01	9.7 ± 2.8	1.0 ± 0.2	64.7 ± 1.3	12.4 ± 1.2
V1d	0.29 ± 0.01	7.3 ± 0.1	0.46 ± 0.02	22.2 ± 4.0	3.0 ± 1.6	173.4 ± 1.9	12.0 ± 0.3
V1e	0.30 ± 0.01	7.0 ± 0.1	0.30 ± 0.04	12.3 ± 1.1	1.7 ± 0.1	73.5 ± 1.0	8.8 ± 0.1
V2	0.14 ± 0.03	6.4 ± 0.2	0.07 ± 0.00	17.7 ± 0.2	5.8 ± 0.5	45.5 ± 0.9	6.3 ± 0.8
V3	0.18 ± 0.01	6.8 ± 0.2	0.11 ± 0.01	29.5 ± 0.7	2.6 ± 0.2	20.5 ± 0.3	13.6 ± 0.3
V4	0.21 ± 0.00	5.7 ± 0.1	0.48 ± 0.03	15.2 ± 0.2	2.6 ± 0.0	41.6 ± 0.3	7.1 ± 0.5
V5	0.16 ± 0.02	7.1 ± 0.1	0.15 ± 0.02	36.8 ± 0.9	3.2 ± 0.2	52.0 ± 2.1	19.5 ± 1.2
V6	0.10 ± 0.01	6.4 ± 0.1	0.04 ± 0.02	22.9 ± 2.0	7.7 ± 1.9	3.1 ± 0.3	1.0 ± 0.1
V7	0.17 ± 0.02	6.4 ± 0.2	0.04 ± 0.03	26.2 ± 2.2	3.2 ± 0.5	1.2 ± 0.1	4.6 ± 0.1
V8	0.18 ± 0.02	6.7 ± 0.1	0.16 ± 0.02	18.5 ± 1.2	4.5 ± 0.4	36.7 ± 0.2	11.9 ± 0.3
V9	0.23 ± 0.01	6.9 ± 0.1	0.16 ± 0.04	18.2 ± 7.2	2.7 ± 2.8	54.3 ± 0.3	8.8 ± 0.1
V10	0.21 ± 0.01	7.0 ± 0.1	0.16 ± 0.02	15.1 ± 0.6	4.8 ± 2.2	50.9 ± 0.3	4.6 ± 0.1
mean	0.22 ± 0.07	6.7 ± 0.4	0.23 ± 0.16	18.8 ± 8.1	3.3 ± 1.8	52.3 ± 41.8	11.1 ± 6.9

Screening of *Bifidobacteria* and *Lactobacillaceae* for bioconversion of indole and p-cresol

Bifidobacteria and *Lactobacillaceae* (33 and 26 strains, respectively; Supplementary Table S1) were grown on MRS (BD Difco, Sparks, United States) for 16 h, at 37°C in anaerobiosis. Biomass was harvested by centrifugation, washed, and resuspended in 50 mM PBS pH 6.5 at a normalized concentration corresponding to 20 OD_{600}. Indole and p-cresol were separately added at the final concentration of 1.0 mM and resting cells were incubated for 48 h at 37°C in anaerobiosis. At the end of biotransformation, samples were centrifuged, biomass was extracted with methanol, supernatants and extracts were analyzed with HPLC. Experiments were performed in triplicate.

Results

Coprometry, chemical analysis of fecal samples

Fecal samples from 10 healthy adults, *i.e.*, the 5 samples longitudinally collected from the subject V1 and the single ones collected from the other, were analyzed to determine the inter- and intra-individual variability of parameters related to the proteolytic and saccharolytic metabolism, such as pH, content of ammonium, total and soluble carbohydrates, p-cresol and indole, and the dry/wet weight ratio (Table 1, Supplementary Figure S1, S2).

In the cohort of healthy subjects, the dry/wet ratio of feces (w/w) ranged from 0.10 to 0.31, with a mean value of 0.22. This parameter was significantly higher ($p < 0.01$) in the 5 samples from V1 than in the samples from the other subjects (mean = 0.29 and 0.18, respectively), the latter presenting a greater dispersion (σ_{V1a-e} 0.016 vs. σ_{V2-V10} 0.040). Ammonium content spread over a range of more than one magnitude (0.04–0.48 mg/g). The samples V1a–e were characterized by a significantly higher ($p < 0.01$) ammonium level with respect to the samples V2–V10 (mean = 0.38 vs. 0.15 mg/g), but the intra-individual variability within V1 was similar to the inter-individual one between the samples V2–V10. Unlike the dry weight and ammonium, all the other measured parameters did not significantly differ in V1 with respect to the other subjects. Moreover, V1 intra-individual variabilities were similar to the inter-individual ones among V2–V10, with the sole exception of p-cresol (σ_{V1a-e} 52.1 µg/g vs. σ_{V2-V10} 21.6 µg/g).

The pH of all the samples was comprised from 5.7 to 7.3, with a mean value of 6.7. Total and soluble carbohydrates ranged from 6.7 to 36.8 mg/g (mean 18.8 mg/g) and from 1.0 to 7.7 mg/g (mean 3.3 mg/g), respectively. The feces contained very different amounts of indole and p-cresol, with both intra- and inter-individual high variability of concentration of these metabolites. Indole ranged from 1.0 to 27.5 µg/g (mean 11.1 µg/g) and values of p-cresol content were spread among more than two orders of magnitude, ranging from 1.2 to 173.4 µg/g (mean 52.3 µg/g).

The level of dry matter in the feces positively and strongly correlated with the content of ammonium (r = 0.80; Table 2) and was negatively linked to the content of total (r = −0.62) and soluble (r = −0.80) carbohydrates. Accordingly, also ammonium negatively correlated with total and soluble carbohydrates (r = −0.56 and −0.67, respectively). Also the fecal content of the uremic toxins was linked

TABLE 2 Pearson correlation (coefficient r) between the chemical parameters measured for the characterization of the fecal samples. *$p < 0.05$; **$p < 0.01$.

	Dry weight/Wet weight	pH	NH_4^+	Total carbohydrates	Soluble carbohydrates	p-cresol	Indol
Dry weight/ Wet weight		0.34	0.80**	−0.62*	−0.80**	0.70**	0.45
pH	0.34		−0.01	0.21	−0.15	0.26	0.12
NH_4^+	0.80**	−0.01		−0.56*	−0.67**	0.66**	0.45
Total carbohydrates	−0.62*	0.21	−0.56*		0.32	−0.56*	−0.05
Soluble carbohydrates	−0.80**	−0.15	−0.67**	0.32		−0.59*	−0.53*
p-cresol	0.70**	0.26	0.66**	−0.56*	−0.59*		0.32
Indole	0.45	0.12	0.45	−0.05	−0.53*	0.32	

TABLE 3 Accumulation of p-cresol and indole in the fecal slurries. The reported values represent the highest rate observed in the first 8 h of incubation and are expressed as µg of product accumulated per hour per Gram of feces (wet weight). Significantly different mean values compered to the control are indicated; *$p < 0.05$; **$p < 0.01$.

Sample	p-cresol (µg/h/g_{feaces})			Indole (µg/h/g_{feaces})		
	C	+IP	+TT	C	+IP	+TT
V1a	6.0	4.1	15.5	0.9	−1.4	13.5
V1b	6.9	4.6	13.4	1.1	−3.8	10.9
V1c	4.4	5.3	8.9	0.5	0.2	12.3
V1d	7.7	5.4	31.2	1.6	1.8	15.0
V1e	3.2	5.2	12.8	0.8	−1.5	11.4
V2	1.9	3.3	13.5	1.0	−3.0	34.5
V3	3.2	2.6	28.7	0.8	−0.5	37.4
V4	6.8	4.4	14.6	1.0	−1.1	19.3
V5	3.9	−0.7	18.5	1.8	−3.6	37.7
V6	0.0	0.0	0.3	2.1	−0.3	22.6
V7	0.3	1.0	0.1	1.7	1.2	11.1
V8	5.1	2.4	21.6	0.4	−1.5	37.5
V9	6.0	8.1	23.8	1.7	1.1	30.6
V10	6.0	3.2	30.4	0.7	−0.6	36.0
mean	4.4	3.5	16.7**	1.1	−0.9**	23.6**

to the above mentioned parameters. In particular, p-cresol positively correlated with the dry/wet weight ratio (r = 0.70) and the ammonium content (r = 0.66) and was negatively associated with the content of total and soluble carbohydrates (r = −0.56 and −0.59, respectively). Total and soluble carbohydrates were in countertrend also with the content of indole (r = −0.53).

Biotransformations

To determine the kinetics of formation and/or degradation of indole and p-cresol by fecal bacteria, bioconversion experiments

were carried out with resting cells of intestinal microbiota without any addition (controls C), spiked with indole and p-cresol (samples IP), or supplemented with their precursors tryptophan and tyrosine (samples TT; Table 3).

All the controls except V6 accumulated p-cresol with a rate comprised from 0.3 to 6.9 µg/h/g_{feces} (mean 4.4 µg/h/g_{feces}). The intra-individual variability within V1was lower than the inter-individual variability observed among V2-V10 (σ_{V1a-e} 1.8 µg/h/g_{feces} vs. σ_{V2-V10} 2.5 µg/h/g_{feces}). In IP samples, the supplementation of p-cresol did not affect the rate of its accumulation (mean = 3.5 µg/h/g_{feces}), compared to the controls ($p > 0.05$). On the other hand, the presence of the precursor tyrosine in TT samples determined a significant increase in the rate of p-cresol production (mean 16.7 µg/h/g_{feces}, $p < 0.01$). In samples V6 and V7, accumulation of p-cresol was low or negligible and was scarcely affected by the initial spike of p-cresol (IP) or by the presence of tyrosine (TT). The kinetics of p-cresol concentration during the biotransformations was different among the samples (Figure 1A), with the majority of the samples characterized by the highest rates of accumulation in the first 4–8 h, followed by a constant but slower increase of concentration up to 24 h. Conversely, in a few samples (V3, V4, and V10) a plateau was reached after 4–8 h, then the concentration decreased.

The rate of indole accumulation in C samples was lower compared to that of p-cresol (mean 1.1 µg/h/g_{feces}) and was comprised in a narrower range (0.4–2.1 µg/h/g_{feces}). V1 intra-individual variability and V2-V10 inter-individual variability were similar. Indole supplementation determined a significant decrease in kinetic of accumulation, with 10 out of 14 IP samples showing a net reduction of indole concentration during incubation. On the other side, the supplement of tryptophan accelerated the indole accumulation up to 37.7 µg/h/g_{feces}. In both IP and TT samples, the initial supplement increased the variability of the accumulation rate of indole, that lay in the range of −3.8–1.8 and 10.9–37.7 µg/h/g_{feces}, respectively. For indole, the subjects were characterized by different kinetics during the biotransformation, in particular
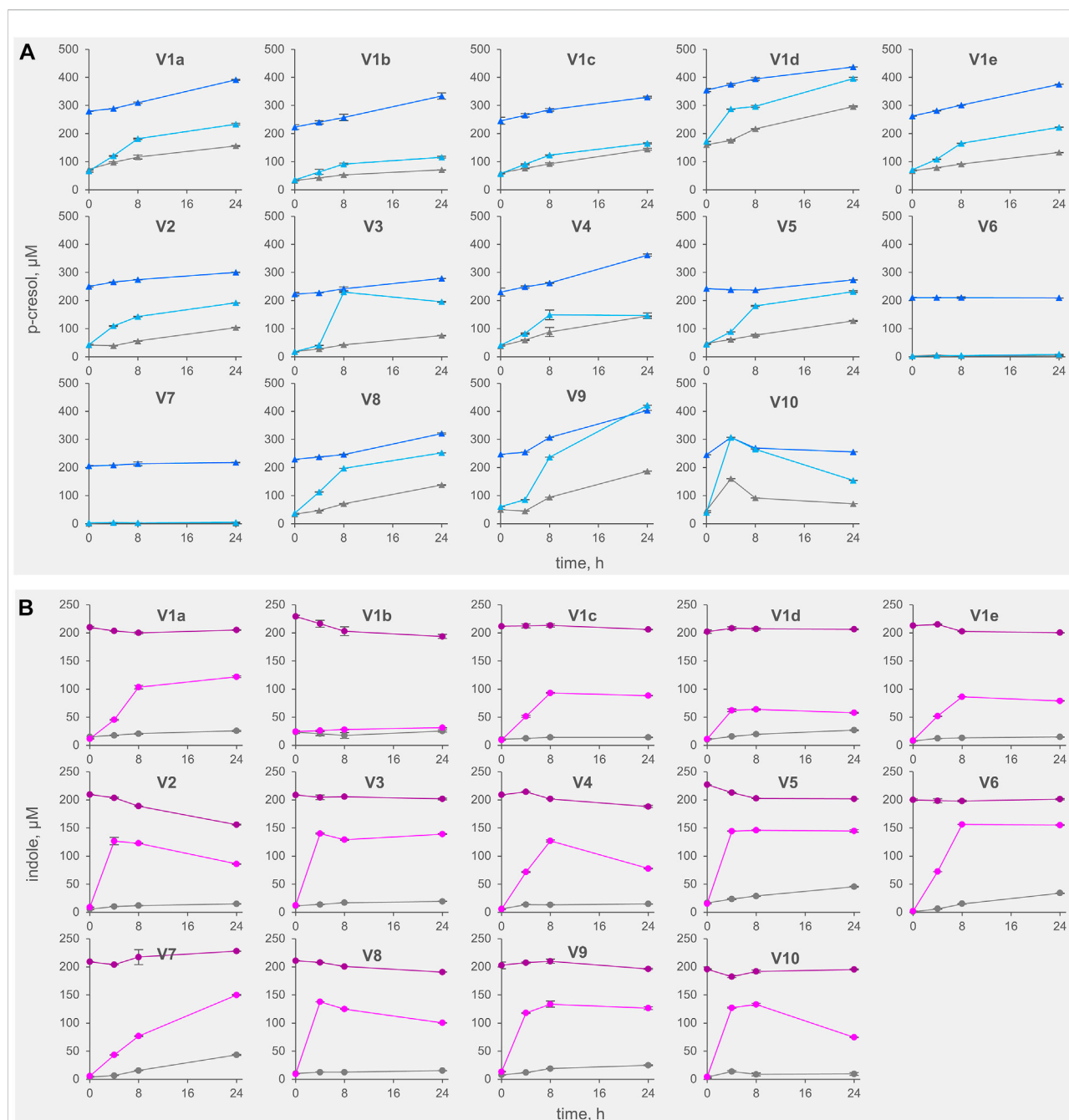
FIGURE 1
Kinetics of accumulation of p-cresol **(A)** and indole **(B)** in the fecal slurries. The values are means of 3 replicated measurements $\pm$ standard deviation. Grey, Controls; dark colors, IP series spiked with indole and p-cresol; light colors, TT series supplemented with tyrosine and tryptophan.

in the TT series (Figure 1B). Most of the TT samples showed the highest rates of accumulation in the first 4–8 h, followed by a slow decrease of the concentration during the last hours of incubation. In some samples (V2, V4, V8, and V10) indole concentration decreased faster, whereas, in two samples (V1a and V7) indole accumulated for 24 h.

VOCs

The headspace of the fecal samples analyzed by SPME-GC-MS yielded a total 57 volatile organic compounds (VOCs) occurring in at least 2 samples, 35 of which contributing at list once for > 1% of the relative abundance (Figure 2,
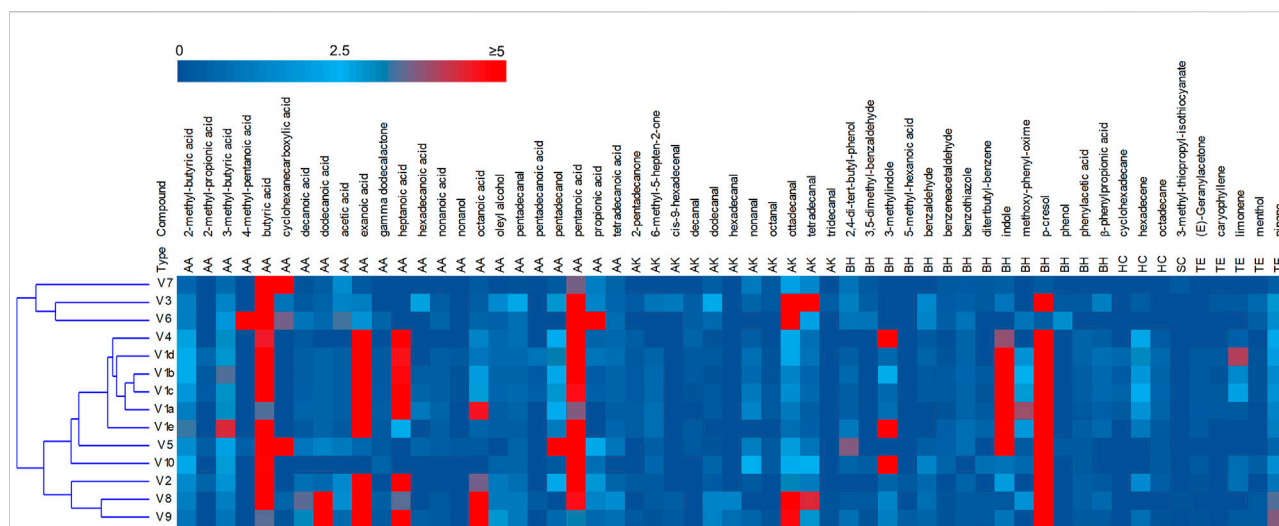
FIGURE 2
Heatmap of the main VOCs identified by HS-SPME-GC-MS analyses of fecal samples. The relative abundance in the gas chromatogram are reported as shades from blue to red. Only the VOCs occurring at least in two samples are reported. AA acids, alcohols and esters; AK aldehydes ketones; BH benzenoids and heterocycles; HC hydrocarbon; SC solphur containing; TE terpene.

Supplementary Figure S3). The main VOCs, occurring most frequently and abundantly, were metabolites derived from the degradation of aromatic amino acids (p-cresol, indole, and 3-methylindole), organic acids (mainly short chain and branched chain fatty acids with 4–8 carbon atoms deriving from the saccharolytic metabolism of sugars and from degradation of aliphatic amino acids), and fatty aldehydes (tetra- and otta-decanal).

PCA of VOCs profiles revealed that the intra-individual variability was less pronounced than the interindividual one (Supplementary Figure S4). The samples V1a–e closely distributed in the lower and right quadrant with V4 and V5, all characterized by high values of aromatic amino acids metabolites. V2, V8, and V9 lay separately and were characterized by short chain fatty acids, whereas the cluster including samples V3, V6, and V7 was featured by high amounts of cyclohexane carboxylic acid, butyric acid and hexadecanal.

16 S rRNA gene profiling

A metataxonomic survey of microbiota used as inoculum was carried out by 16 S rRNA gene profiling. Only the first and the last samples of V1 longitudinal series (V1a and V1e) were analyzed and sample V5 was excluded due to errors in the sequencing procedure. A total of 566,443 sequence reads was obtained, 28,119–182,257 per sample. The reads were dereplicated into 2,064 ASVs hitting a reference sequence in SILVA database and collapsed at the 7th level of taxonomic annotation into 286 features.

As a whole, the fecal microbiota was largely composed by Firmicutes (50.6%–94.6%lower amount of Bacteroidota (4.1%–46.5%, and only minor number of Actinobacteriota (0.2%–3.8%) and Proteobacteria (0.1%–2.3%; Figure 3). Among the bacterial groups recognized as proteolytic (Amaretti et al., 2019; Raimondi et al., 2021), the main taxa were Lachnospiraceae (mean 32.9%), Bacteroidaceae (17.8%), and Ruminococcaceae (13.7%) followed by Peptostreptococcaceae, Clostridiaceae, Streptococcaceae, and Enterobacteriaceae, each of them accounting on average for less than 2.9% of the whole bacterial population.

According to the rarefaction plots of the main alpha diversity indices (Supplementary Figure S5A), the diversity of all samples was entirely captured and sequences were rarefied at the level of the sample with the lowest number of reads (20,251). Non-phylogenetic beta-diversity analysis was performed and reported in Supplementary Figure S5B. Calculating distances among samples with non-quantitative algorithms, the inter-individual distances among the microbiota was higher than the intra-individual one, with V1 samples laying very closely in the PCo1 vs PCo2 plots. On the other hand, with quantitative metrics that calculate distances also considering the abundance of the identified taxa and not only their presence, the samples V1a and V1e lay far from each other, showing that fecal samples from the same subject could be more distant taking into account quantitative composition than samples from different subjects.

PICRUSt2 was utilized to reconstruct the metabolic functions of the microbial communities, with the aim to infer the abundance of the genes encoding the key enzymes involved in metabolism of uremic toxins. The genes encoding tryptophanase,

FIGURE 3
Microbiota composition of the fecal samples analyzed by 16 S rRNA gene profiling. From volunteer 1, only the first and last sample (V1a and V1e), collected 4 months apart, were analyzed. V5 sample sequencing with Illumina platform did not provide useful sequences for analysis. The compositions are reported as relative abundance of the reads collapsed at the level of Phylum **(A)** Family **(B)**, and Genus **(C)**.

2-iminoacetate synthase, and 4-hydroxyphenylacetate decarboxylase were pinpointed in all the microbiomes (Supplementary Figure S6). 2-Iminoacetate synthase, involved in transformation of tyrosine to p-cresol, presented similar abundance in all the samples, laying in the range between 0.07% and 0.09% of the relative abundance of all predicted genes, whereas abundance of 4-hydroxyphenylacetate decarboxylase, involved in the formation of p-cresol, and tryptophanase, responsible of the transformation of tryptophan to indole, exhibited wide differences among samples. 4-Hydroxyphenylacetate decarboxylase ranged from 0.0001% to 0.012%, with the lowest and the highest abundance occurring in V2 and V6, respectively. Such marked difference among samples of genes associated to production of p-cresol was non significantly correlated with differences in p-cresol formation rate, in both control and TT slurries. Tryptophanase ranged from 0.001% to 0.014%, with the lowest values occurring in samples V1a and V6 and the highest ones in V3 and V9. Positive significant correlation between the levels of tryptophanase and indole formation rate in TT slurries (r = 0.69) was identified.

Correlations of microbiome composition with VOCs profiles and with kinetics of p-cresol and indole accumulation

Correlation analysis of VOCs and 16 S rRNA gene profile revealed a statistically significant correlation of p-cresol and indole with some OTUs of Firmicutes (Figure 4, Supplementary Figure S7), mainly uncultured bacteria belonging to the families of Ruminococcaceae (including the genus *Subdoligranulum*), Oscillospiraceae, *Eubacterium coprostanoligenes* group, and Peptostreptococcaceae (genus *Romboutsia*). These bacterial groups also presented a significant positive correlation with 3-methyl indole and other products of proteolytic metabolism, while they were negatively correlated with organic acids (acetic, propionic, and butyric acids), fatty aldehydes, phenol, and benzaldehyde.

P-cresol and indole in both feces and their volatilome presented a statistically significant positive correlation with OTUs of uncultured Bacteroidetes and Firmicutes. The former belonged to the genus *Bacteroides* and the latter to the families Butyricicoccaceae (genus *Butyricicoccus*), Monoglobaceae (genus *Monoglobus*), Lachnospiraceae (genera *Faecalibacterium*, *Roseburia*, and *Eubacterium ventriosum* group), all presenting a general positive correlation (although not always reaching statistical significance) with short chain fatty acids, fatty aldehydes, phenol, and benzaldehyde.

Any correlation between the microbiome composition and the chemical signature of stool samples was searched utilizing Spearman's test. The analysis revealed that the OTU that positively correlated with uremic toxins generally presented a

FIGURE 4
Heatmap of Spearman's correlation between microbial species, VOCs identified by HS-SPME-GC-MS analyses and coprometry parameters.
Microbial taxa and VOCs are filtered to show only features with 1% abundance in at least one subject.
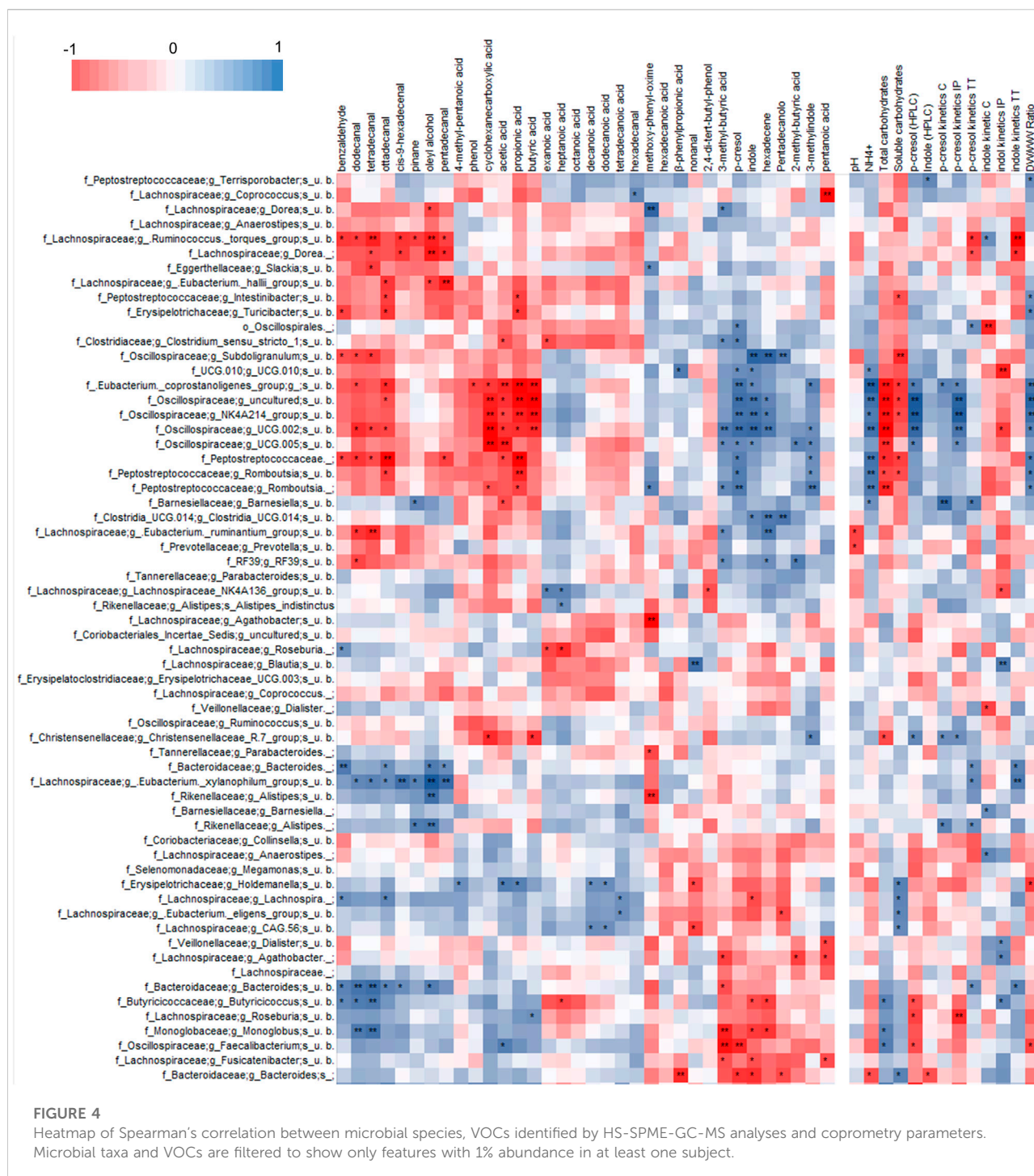
significant positive correlation with the fecal amount of ammonium and with the dry/wet weight ratio, and a significant negative correlation with both soluble and/or total carbohydrates, whereas they did not present any relationship with the pH.

Oscillospiraceae, *Eubacterium coprostanoligenes* group, and Peptostreptococcaceae that positively correlated with the uremic

toxins also presented a general positive correlation with p-cresol generation rate in biotransformation experiments, although the statistical significance was not always reached. Moreover, some other OTUs that did not present any significant correlation with the uremic toxins were positively associated with production rates of p-cresol (e.g., *Bacteroides*, *Alistipes*, *Eubacterium xylanophylum*, and *Barnesiella*) and indole (e.g., *Bacteroides*,

FIGURE 5
Biotranformation of p-cresol and indole with resting cells of
33 Bifidocateria and 26 Lactobacillaceae strains. Box plot reports
the median values, 10th, 25th, 75th, and 90th percentile of the
relative amount of p-cresol and indole removed from the
supernantat referred to the initial values. Significant differences are
indicated. **p < 0.01.
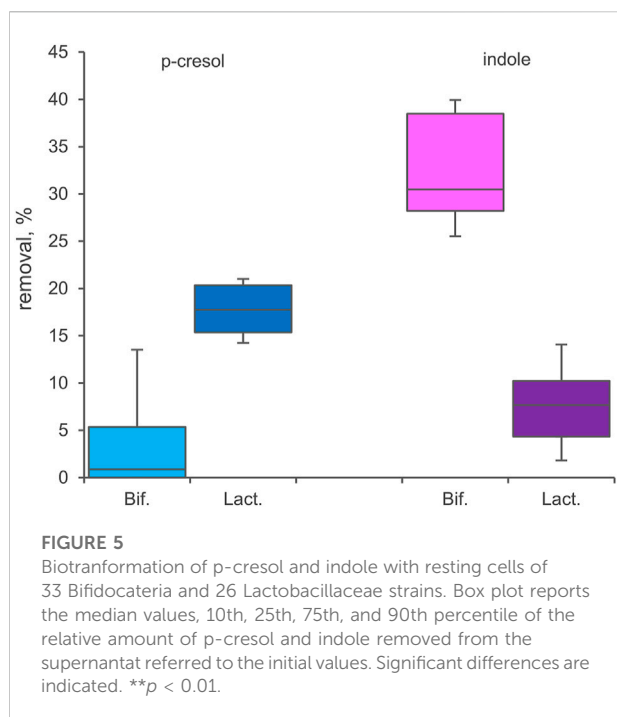
Ruminococcus torques, Balutia, Dialister, Butyricicoccus) by the fecal slurries.

Bioconversion of p-cresol and indole with resting cell of bifidobacteria and lactobacillaceae

33 *Bifidobacterium* strains belonging to 8 different species or subspecies and 26 Lactobacillaceae ascribed to 15 taxa (Supplementary Table S1) were screened for bioconversion of p-cresol and indole in resting cells condition. The percentage of removal of the two molecules from the supernatants and the absorption in the inoculated biomasses were determined after 48 h of incubation. All the tested bifidobacteria strains were able to reduce indole concentration in the supernatant (Figure 5), with values of removal ranging from 24.0% to 41.7% (mean 32.4%). Among the screened species, *B. longum* (subspecies *longum* and *infantis* comprised) was the most effective in indole removal (mean 36.0%, *p* < 0.05). In the tested condition, bifidobacteria were less prone to reduce p-cresol concentration, with a mean value of 4.2% and higher variability among strains (from 0% to 18.2%). In Lactobacillaceae biotranformations, p-cresol was removed from the supernatant more efficiently than indole (mean values of 17.6% and 7.5%, respectively), the latter also characterized by the higher variability among strains (from 0.8% to 16.3%). The tested species showed no significant differences in the amount of removed p-cresol and indole.

Analyzing the biomasses of bifidobacteria and Lactobacillaceae after incubation, always less than 3.0% of absorbed indole or p-cresol was detected.

Discussion

In this study, the volatilome, the microbiota composition, and the physical and chemical features of feces from healthy subjects were investigated to obtain information on the role of intestinal bacteria in the formation of the uremic toxins indole and p-cresol, providing a preliminary view on the intra- and interindividual variability.

Higher fecal levels of indole and p-cresol were associated with lower carbohydrates and higher ammonium levels, that are markers of a more pronounced intestinal proteolytic metabolism. This is in agreement with the positive relationship observed also with the dry/wet weight ratio, since an increased water absorption results from a prolonged retention of feces within the colon that also leads to the onset of protein breakdown as the availability of fermentable carbohydrates becomes lower (Tottey et al., 2017).

The present study also enabled preliminary considerations on intra- and interindividual variability of markers of proteolytic and saccharolytic metabolism affecting the level of uremic toxins. The 5 samples longitudinally provided by the subject V1 over a period of 4 months indicated that the range of concentration of p-cresol and indole can be extremely wide within the same individual, likely as a result of different component of the diets affecting the availability of precursors, and of the impact of the diet on microbiota composition, in its turn affecting the abundance of producing/degrading bacteria. The features of the fecal samples less influenced over the time were dry weight, according to a personal transit time minimally changing in a healthy status, and total carbohydrates, likely reflecting regular eating habits. Albeit subject V1 likely had his own food habits, variations of pH and ammonium concentration likely corresponded to a different intake of proteins and fibers, with resultant differences in terms of proteolytic and saccharolytic metabolic activity in the colon. Fecal specimens from the other 9 volunteers presented a higher content of water and a lower concentration of ammonium, likely reflecting for V1 a longer transit time, which implicates higher water absorption and a shift towards a proteolytic metabolism of colonic bacteria, following depletion of carbohydrates. Accordingly, the mean levels of indole and p-cresol tended to be lower in V2–V10 subjects, when compared to V1, albeit differences were not statistically significant, likely because of the low numerosity of the samples per group and the large variability within the groups.

Previous studies with enrichment cultures of gut microbiota identified main proteolytic bacteria within Lachnospiraceae, Oscillospiraceae, Clostridiaceae, Eubacteriaceae, Peptostrocaccaceae, Sutterellaceae, and Enterobacteriaceae and

revealed that several OTUs of these families correlated with the accumulation of indole and p-cresol within cultures (Amaretti et al., 2019; Raimondi et al., 2021). Some families encompassing potential proteolytic taxa were remarkably abundant in the present study, such as Lachnospiraceae and Ruminococcaceae that represented the 32.9% and 21.7% on average, respectively. On the other hand, other proteolytic families were minor components of the microbiota, with Eubacteriaceae (*Eubacterium corpostanoligenes* group), Peptostreptococcaceae, Clostridiaceae, and Enterobacteriaceae being on average the 4.2%, 2.9%, 1.1%, and 0.1%, respectively.

The headspace of fecal samples contained a mixture of linear and branched SCFA, aromatic compounds, aldehydes, and terpenes. Major markers were organic acids, including butyric and propanoic acids, whereas acetic acid was the less abundant. A high level of indole was a feature of some fecal samples, whereas, with the exceptions of specimens V6 and V7, abundant p-cresol was detected in the headspace of all the specimens. Comparison of VOCs and 16 S rRNA gene profiles pointed out that specific OTUs ascribed to Lachnospiraceae, Oscillospiraceae, and Peptostreptococcaceae presented positive correlation with fecal levels of indole and p-cresol, providing additional evidence of the relationship linking uremic toxins with intestinal protein breakdown and amino acids fermentation. Interestingly, any significant relationship with Enterobacteriaceae, that were major indole producers in fecal cultures, was not found, due to the generally low levels of this family in the samples.

Bioconversions with resting cells obtained from fresh fecal inocula were carried out in order to evaluate the capability of the colonic community to produce or degrade indole and p-cresol and to establish a relationship with bacterial taxa. The presence of the precursors tyrosine and tryptophan significantly increased the accumulation of both the uremic toxins, confirming that colonic bacteria have a strong potential to produce these metabolites. However, the burst of indole production was more marked with respect to p-cresol, suggesting that tryptophan was more limiting in feces compared to tyrosine or that the enzymatic arsenal of gut bacteria was more prone to metabolize this substrate. Interestingly, this attitude was not shared by all the subjects, since V6 and V7 transformations did not efficiently accumulate p-cresol, even in presence of tyrosine. This behavior was consistent with low p-cresol levels in the feces of these subjects, that likely harbored a microbiota with limited metabolic potentials in terms of p-cresol production. The presence of indole reduced its production, suggesting some feedback mechanism depressing indole accumulation when this compound was present in relevant amounts.

Gryp et al. (2020) obtained bacterial isolates belonging to 92 species, spread across several families (mainly within Firmicutes, Bacteroidetes, and Proteobacteria), capable of producing p-cresol under anaerobic conditions from the feces of CKD patients with different severity, indicating that the ability to release uremic toxins is widespread in the microbiome. Coherently, the OTUs that positively correlated with fecal levels of uremic toxins

in the present study were mostly identified as uncultured bacteria belonging to Oscillospiraceae, *Eubacterium coprostanoligenes* group, and Peptostreptococcaceae and presented a positive correlation with also p-cresol generation rate in biotransformation experiments. Likewise, other bacterial groups were positively correlated with generation rate of p-cresol and indole, further expanding the range of taxa associated to production of p-cresol (*Bacteroides*, *Alistipes*, *Eubacterium xylanophylum*, and *Barnesiella*) and indole (e.g., *Bacteroides, Ruminococcus torques, Balutia, Dialister, Butyricicoccus*). Although a causal relationship cannot be established from this data, it seems likely that these bacterial groups are either directly involved in p-cresol or indole generation or are involved in protein breakdown, resulting uremic toxins producers from tyrosine and tryptophan.

Bifidobacteria and Lactobacillaceae have been extensively utilized as probiotics to exert beneficial health effect on the host, despite they do not represent a dominant component of the human gut microbiome (Quartieri et al., 2016b; Rossi et al., 2016; Bottacini et al., 2017; Khalesi et al., 2019). Any potential ability to reduce the levels of indole and p-cresol would be of great importance, as would open new applications of specifically designed probiotics in the amelioration of uremic toxin levels. The screening herein presented revealed that the ability to reduce p-cresol and indole from the environment is widespread among Lactobacillaceae and bifidobacteria, even though with wide strain to strain variability. The mechanism of toxins removal of these bacteria remains to be clarified, particularly for the most promising strains belonging to *Bifodobacterium longum* supsb. *longum*, *Lacticaseibacillus rhamnosus*, and *Lactiplantibacillus plantarum*. Interestingly, lactobacilli and bifidobacteria presented complementary effectiveness in the removal of p-cresol and indole, respectively, suggesting a combined utilization in a probiotic formula.

The present study is the first focusing on the levels of p-cresol and indole, the content of other compound derived by microbial metabolism proteins and carbohydrates, the rate of formation/degradation of these uremic toxins and the correlation with microbiota taxa in the feces of healthy subjects. Despite being a small study, with sample size as a main limitation, the information herein presented contributes to disclose the relationships between microbiota composition and the production of uremic toxins. Therefore, it could represent a hypothesis-generating study and inspire larger confirmatory studies devising nutritional interventions, also combining diet modification with probiotics administration, aimed to prevent the onset, hamper the progression, and alleviate the impact of CKD.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA843561.

Ethics statement

The studies involving human participants were reviewed and approved by Comitato Etico Provinciale, Azienda Policlinico di Modena, Italy. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization: SR, AA, and MR. Investigation: SR, FC, MS, and AL. Data curation: MS, FC, AA, SR, and AL. Formal Analysis: AA. Visualization: SR, FC, and AA. Writing—original draft: all authors. Funding acquisition: MR. Supervision: MR.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
The handling editor [IC] declared a past co-authorship with the author [SR]

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmmed.2022.959189/full#supplementary-material

References

Amaretti, A., Raimondi, S., Leonardi, A., Quartieri, A., and Rossi, M. (2015). Hydrolysis of the rutinose-conjugates flavonoids rutin and hesperidin by the gut microbiota and bifidobacteria. *Nutrients* 7, 2788–2800. doi:10.3390/nu7042788

Amaretti, A., Gozzoli, C., Simone, M., Raimondi, S., Righini, L., Pérez-Brocal, V., et al. (2019). Profiling of protein degraders in cultures of human gut microbiota. *Front. Microbiol.* 10, 2614. doi:10.3389/fmicb.2019.02614

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Author Correction: Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 1091. doi:10.1038/s41587-019-0252-6

Bottacini, F., van Sinderen, D., and Ventura, M. (2017). Omics of bifidobacteria: research and insights into their health-promoting activities. *Biochem. J.* 474, 4137–4152. doi:10.1042/BCJ20160756

Caggiano, G., Cosola, C., Di Leo, V., Gesualdo, M., and Gesualdo, L. (2020). Microbiome modulation to correct uremic toxins and to preserve kidney functions. *Curr. Opin. Nephrol. Hypertens.* 29 (1), 49–56. doi:10.1097/MNH.0000000000000565

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi:10.1038/nmeth.3869

Challand, M. R., Martins, F. T., and Roach, P. L. (2010). Catalytic activity of the anaerobic tyrosine lyase required for thiamine biosynthesis in *Escherichia coli*. *J. Biol. Chem.* 285, 5240–5248. doi:10.1074/jbc.M109.056606

De Juan, A., and Segura, E. (2021). Modulation of immune responses by nutritional ligands of aryl hydrocarbon receptor. *Front. Immunol.* 12, 645168. doi:10.3389/fimmu.2021.645168

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., et al. (2020). PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* 38, 685–688. doi:10.1038/s41587-020-0548-6

El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D., and Henrissat, B. (2013). The abundance and variety of carbohydrate–active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* 11, 497–504. doi:10.1038/nrmicro3050

Gryp, T., Huys, G. R. B., Joossens, M., Van Biesen, W., Glorieux, G., and Vaneechoutte, M. (2020). Isolation and quantification of uremic toxin precursor-generating gut bacteria in chronic kidney disease patients. *Int. J. Mol. Sci.* 21, 1986. doi:10.3390/ijms21061986

Huang, X., Nie, S., and Xie, M. (2017). Interaction between gut immunity and polysaccharides. *Crit. Rev. Food Sci. Nutr.* 57, 2943–2955. doi:10.1080/10408398.2015.1079165

Khalesi, S., Bellissimo, N., Vandelanotte, C., Williams, S., Stanley, D., and Irwin, C. (2019). A review of probiotic supplementation in healthy adults: helpful or hype? *Eur. J. Clin. Nutr.* 73, 24–37. doi:10.1038/s41430-018-0135-9

Kim, S. M., and Song, I. H. (2020). The clinical impact of gut microbiota in chronic kidney disease. *Korean J. Intern. Med.* 35, 1305–1316. doi:10.3904/kjim.2020.411

Lee, J. H., and Lee, J. (2010). Indole as an intercellular signal in microbial communities. *FEMS Microbiol. Rev.* 34, 426–444. doi:10.1111/j.1574-6976.2009.00204.x

Marcobal, A., Kashyap, P. C., Nelson, T. A., Aronov, P. A., Donia, M. S., Spormann, A., et al. (2013). A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice. *ISME J. Oct.* 7 (10), 1933–1943. doi:10.1038/ismej.2013.89

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–17. doi:10.14806/ej.17.1.200

Oliphant, K., and Allen-Vercoe, E. (2019). Macronutrient metabolism by the human gut microbiome: major fermentation by–products and their impact on host health. *Microbiome* 7, 91. doi:10.1186/s40168-019-0704-8

Popkov, V. A., Zharikova, A. A., Demchenko, E. A., Andrianova, N. V., Zorov, D. B., and Plotnikov, E. Y. (2022). Gut microbiota as a source of uremic toxins. *Int. J. Mol. Sci.* 23, 483. doi:10.3390/ijms23010483

Quartieri, A., García-Villalba, R., Amaretti, A., Raimondi, S., Leonardi, A., Rossi, M., et al. (2016a). Detection of novel metabolites of flaxseed lignans *in vitro* and *in vivo*. *Mol. Nutr. Food Res.* 60, 1590–1601. doi:10.1002/mnfr.201500773

Quartieri, A., Simone, M., Gozzoli, C., Popovic, M., D'Auria, G., Amaretti, A., et al. (2016b). Comparison of culture-dependent and independent approaches to characterize fecal bifidobacteria and lactobacilli. *Anaerobe* 38, 130–137. doi:10.1016/j.anaerobe.2015.10.006

Raimondi, S., Luciani, R., Sirangelo, T. M., Amaretti, A., Leonardi, A., Ulrici, A., et al. (2019). Microbiota of sliced cooked ham packaged in modified atmosphere throughout the shelf life: Microbiota of sliced cooked ham in MAP. *Int. J. Food Microbiol.* 289, 200–208. doi:10.1016/j.ijfoodmicro.2018.09.017

Raimondi, S., Calvini, R., Candeliere, F., Leonardi, A., Ulrici, A., Rossi, M., et al. (2021). Multivariate analysis in microbiome description: correlation of human gut protein degraders, metabolites, and predicted metabolic functions. *Front. Microbiol.* 12, 723479. doi:10.3389/fmicb.2021.723479

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584. doi:10.7717/peerj.2584

Rossi, M., Martínez-Martínez, D., Amaretti, A., Ulrici, A., Raimondi, S., and Moya, A. (2016). Mining metagenomic whole genome sequences revealed subdominant but constant *Lactobacillus* population in the human gut microbiota. *Environ. Microbiol. Rep.* 8, 399–406. doi:10.1111/1758-2229.12405

Ruan, W., Engevik, M. A., Spinler, J. K., and Versalovic, J. (2020). Healthy human gastrointestinal microbiome: composition and function after a decade of exploration. *Dig. Dis. Sci.* 65, 695–705. doi:10.1007/s10620-020-06118-4

Saito, Y., Sato, T., Nomoto, K., and Tsuji, H. (2018). Identification of phenol- and p-cresol-producing intestinal bacteria by using media supplemented with tyrosine and its metabolites. *FEMS Microbiol. Ecol.* 94, fiy125. doi:10.1093/femsec/fiy125

Saldana-Morales, F. B., Kim, D. V., Tsai, M. T., and Diehl, G. E. (2021). Healthy intestinal function relies on coordinated enteric nervous system, immune system, and epithelium responses. *Gut Microbes* 13, 1–14. doi:10.1080/19490976.2021.1916376

Scott, K. P., Gratz, S. W., Sheridan, P. O., Flint, H. J., and Duncan, S. H. (2013). The influence of diet on the gut microbiota. *Pharmacol. Res.* 69, 52–60. doi:10.1016/j.phrs.2012.10.020

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303

Tomas-Barberan, F., García-Villalba, R., Quartieri, A., Raimondi, S., Amaretti, A., Leonardi, A., et al. (2014). *In vitro* transformation of chlorogenic acid by human gut microbiota. *Mol. Nutr. Food Res.* 58, 1122–1131. doi:10.1002/mnfr.201300441

Tottey, W., Feria-Gervasio, D., Gaci, N., Laillet, B., Pujos, E., Martin, J. F., et al. (2017). Colonic transit time is a driven force of the gut microbiota composition and metabolism: *In vitro* evidence. *J. Neurogastroenterol. Motil.* 23, 124–134. doi:10.5056/jnm16042

Vanholder, R., Pletinck, A., Schepers, E., and Glorieux, G. (2018). Biochemical and clinical impact of organic uremic retention solutes: A comprehensive update. *Toxins* 10, 33. doi:10.3390/toxins10010033

Vanholder, R., Nigam, S. K., Burtey, S., and Glorieux, G. (2022). What if not all metabolites from the uremic toxin generating pathways are toxic? *Toxins* 14, 221. doi:10.3390/toxins14030221

Vaziri, N. D., Wong, J., Pahl, M., Piceno, Y. M., Yuan, J., DeSantis, T. Z., et al. (2013). Chronic kidney disease alters intestinal microbial flora. *Kidney Int.* 83, 308–315. doi:10.1038/ki.2012.345

Wong, J., Piceno, Y. M., DeSantis, T. Z., Pahl, M., Andersen, G. L., and Vaziri, N. D. (2014). Expansion of urease- and uricase-containing, indole- and p-cresol-forming and contraction of short-chain fatty acid-producing intestinal microbiota in ESRD. *Am. J. Nephrol.* 39, 230–237. doi:10.1159/000360010

Yanofsky, C. (2007). RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. *RNA* 13, 1141–1154. doi:10.1261/rna.620507

Genetics in parkinson's disease: From better disease understanding to machine learning based precision medicine

Mohamed Aborageh[1], Peter Krawitz[2] and Holger Fröhlich[1,3]*

[1]Bonn-Aachen International Center for Information Technology (B-IT), Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany, [2]Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn, Bonn, Germany, [3]Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, Germany

Parkinson's Disease (PD) is a neurodegenerative disorder with highly heterogeneous phenotypes. Accordingly, it has been challenging to robustly identify genetic factors associated with disease risk, prognosis and therapy response via genome-wide association studies (GWAS). In this review we first provide an overview of existing statistical methods to detect associations between genetic variants and the disease phenotypes in existing PD GWAS. Secondly, we discuss the potential of machine learning approaches to better quantify disease phenotypes and to move beyond disease understanding towards a better-personalized treatment of the disease.

KEYWORDS

Parkinson disease, risk, genome-wide association study, machine learning, polygenic risk score

1 Introduction

Parkinson's Disease (PD) is a neurodegenerative disorder (NDD) affecting 7–10 Million patients worldwide. PD patients suffer from motor symptoms like bradykinesia, rigidity, tremor, and postural instability. Speech impairments, characterized by hypokinetic dysarthria, are among the first symptoms (including disruptions in prosody, articulation and, phonation). In addition, non-motor symptoms include cognitive impairment, sleep disorders as well as autonomic and mood dysfunction. The cause of idiopathic PD is unknown, and all currently available treatments (e.g. L-DOPA) are symptomatic. PD has a high subject-to-subject variability of symptoms reflecting disease progression (Poewe et al., 2017).

In recent years, genome-wide association studies (GWAS) have shed light on the polygenic nature of Parkinson's Disease (PD) (Simón-Sánchez et al., 2009; Satake et al., 2009; Kara et al., 2014; Siitonen et al., 2017; Bandres-Ciga et al., 2019; Nalls et al., 2019). First GWAS aimed to identify mutations in coding regions that could be linked to each neurodegenerative trait. Accordingly, variants associated with α-synuclein were detected

FIGURE 1
Advantages and disadvantages of different association tests.
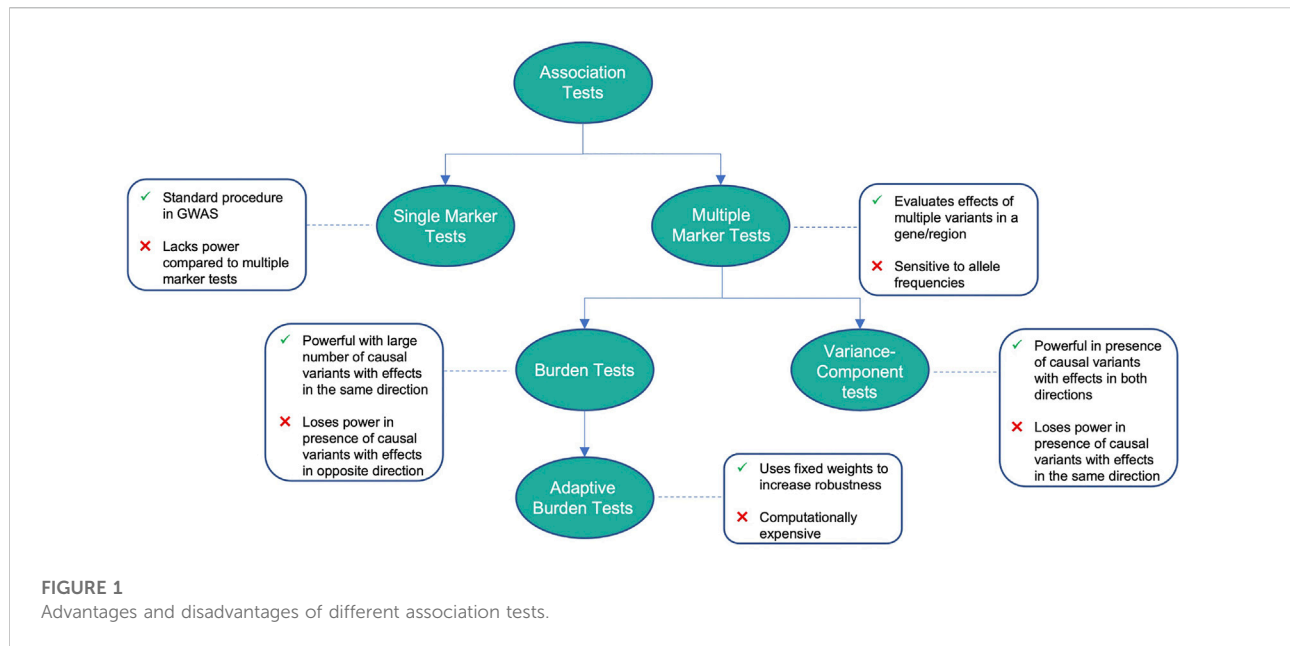
(Mata et al., 2010), one of the hallmark proteins of the disease. However, a meta-analysis of several studies found more variants with smaller effects to be more common in patients than fully penetrant variants (Tran et al., 2020). In addition, larger cohorts now open the possibility to identify less frequent variants and study the interaction with environmental factors. An example is the 23andMe PD cohort, which was able to identify 17 new risk loci for idiopathic PD (Chang et al., 2017; Nalls et al., 2014). Another example is United Kingdom Biobank (UKB), where other authors were able to demonstrate novel gene-environment interactions (Jacobs et al., 2020).

Despite these successes, unraveling the genetic basis of PD, specifically in its sporadic form, remains challenging:

- PD demonstrates a highly heterogeneous phenotype with different long-term outcomes (Aasly, 2020). Accordingly, it is difficult to find genetic associations. So far most research has focused on risk factors for PD diagnosis, but less attention has been paid to identifying genetic variants associated with different long-term outcomes. Notably, a few papers report on genetic risk factors for cognitive impairment in idiopathic PD (Collins and Williams-Gray, 2016; Amer et al., 2018; Planas-Ballvé and Vilas, 2021).
- Sizes of existing cohorts still impose a statistical challenge to identify rare variants.
- Many genetic variants jointly contribute to the phenotype, possibly in a non-linear manner via gene-gene interactions. Finding the true causal subset of variants is still difficult due to the high dimensionality of the GWAS data, the existence of linkage disequilibrium, and statistically low

contributions of rare genetic variants on the population level.
- While in a recent meta-study more than 70 single-nucleotide polymorphisms (SNPs) have been associated with the risk to develop PD, most of them are located in non-coding regions and thus difficult to interpret (Ho et al., 2022).

In this context the goal of this review is two-fold: First, we provide an overview of existing statistical methods that have been employed to detect associations between genetic variants and the disease phenotype as shown in Figure 1 and Table 1. The second goal of this review is to discuss the potential of machine learning approaches, which could allow to better quantify complex phenotypes and to move beyond disease understanding towards a better personalized treatment of PD in the future. While previous reviews focused on the genetic architecture of PD and discuss associated risk factors (Billingsley et al., 2018), gene-specific polymorphisms (Jiménez-Jiménez et al., 2016), gene-gene and gene-environment interactions (Singh et al., 2014; Dunn et al., 2019), our review has thus a distinguishable methodological focus.

2 Variant association tests

In 2011, Sun et al. (2011) considered rare variants as single-nucleotide polymorphisms with minor allele frequencies (MAF) less than 0.01, and have larger effects than common variants. However, when combined, the number of low-frequency variants makes them common. According to the multiple rare variant

(MRV) hypothesis, cases of common inherited diseases are due to the combined effects of highly-penetrant variants (Bodmer and Bonilla, 2008). The genetic composition of PD is often described by two non-mutually exclusive hypotheses: the common disease common variant (CDCV) hypothesis which describes the genetic basis of PD as a result of a large number of common variants with relatively small effects but combined confer significant disease risk (Pritchard and Cox, 2002), and the common disease rare variant (CDRV) hypothesis which speculates that risk components for complex diseases will be rare genetic variants of small or large effects where highly functional or deleterious alleles may exist. This may be noticeable in late-onset diseases like PD where selective pressures are not profound (Billingsley et al., 2018).

Typically, GWA studies focus on variants with MAF greater than 1–5%, and while they were able to identify several variants with evidence of association to disease risk, these common variants only explain 5%–10% of the disease heritability. This led to the conclusion that disease risk is comprised of both common and rare variants (Schork et al., 2009). Variants located near *SNCA*, *MAPT* genes and low frequency coding variants in *GBA* are validated by GWAS to be statistically significant signals associated with PD (Spencer et al., 2011; Lill et al., 2012; Nalls et al., 2014; Chang et al., 2017).

2.1 Single-marker tests

Single-marker testing involves the application of a univariate test for each variant and assessing their significance while using a scaled *p*-value threshold to account for multiple testing (Asimit and Zeggini, 2010). These tests include X^2, Fisher's test, Cochran-Armitage (CA) test for trend and regression analysis, be it logistic regression for testing binary traits or linear regression for quantitative traits. Since each variant is tested independently, corrections for multiple testing should be accounted for to control the family-wise error (FWE) which may result in a loss of power. Instead, controlling the false-discovery rate (FDR) by allowing a small proportion of incorrect null hypotheses may result in a gain of power, especially at a larger number of tests.

If we assume *m* number of variants within an *n* number of samples, a regression model can be fit at each of the *m* variants to test their association with a trait. Assuming that y_i is the phenotype for sample *i* and x_{ij} is the minor allele count of variant *j* for sample *i*, the relationship of variant *i* can be explained by a linear regression model with the following formula:

$$y_i = \alpha_j + \beta_j x_{ij} + \eta_j z_{j_i} + \varepsilon_i$$

Where z_j is a matrix of covariates that may be present, and ε_i is an error term representing independent random variables with

a mean of 0. For that model, a value of $\beta_j = 0$ represents the null hypothesis of no association at variant *j*. For a logistic regression model, y_i is replaced by $\log\left(\frac{p_i}{1-p_i}\right)$ where p_i is the probability of the trait's presence.

The X^2, Fisher's test, and CA tests construct a 2×3 contingency matrix to compare the genotype frequencies between cases and controls, where rows represent disease status and columns represent the three possible genotypes. For X^2 and Fisher's tests, a null hypothesis of equal genotype frequencies for both the cases and controls is considered. Usually, Fisher's test is preferred since it provides exact results of significance, while X^2 test approximates the results with an accuracy that depends on the sample size, which is not ideal in the case of small samples.

If we represent the genotypes as ordered categories *AA*, *Aa*, and *aa*, the CA test is considered a modification of the X^2 test to introduce a suspected ordering of the genotype effects and aims to test a linear effect of the minor allele's copy counts (Slager and Schaid, 2001), which is defined as follows:

$$CA = \frac{n^2 \left(\left(n_{Aa}^0 n^1 - n_{Aa}^1 n^0\right) + 2\left(n_{aa}^0 n^1 - n_{aa}^1 n^0\right)\right)^2}{n^0 n^1 \left(n_{Aa}(n - n_{Aa}) + 4n_{aa}(n - n_{aa}) - 4n_{Aa}n_{aa}\right)}$$

As mentioned earlier, multiple testing needs to be corrected to control the family-wise error (FWE). The Bonferroni correction is used to test an *m* number of variants while assuming the significance level for the *m* independent hypothesis tests is *α*, using *α/m* to calculate the test-specific significance level (Ranstam, 2016). To control the FDR for independent tests, Benjamini & Hochberg (Benjamini and Hochberg, 1995) developed a sequential Bonferroni procedure, where the *m* *p*-values from the individual tests are first ranked: $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$. At FDR level *q*, assume *k* to be the largest *i* such that $P_{(i)} \leq \frac{i}{m} q$, then the null-hypothesis is rejected for *p*-values less than $p_{(k)}$.

A study by Mata et al. (2017) used linear regression to identify genetic variants that may lead to a cognitive decline in PD patients. Eighteen common variants in thirteen genomic regions exceeded the significance threshold for one cognitive test each. However, rare variant analysis did not yield any significance. Another study by Simón-Sánchez et al. (2009) used the Cochran-Armitage test for trend to test associations with PD in European patients. Four SNPs at the *SNCA* locus and three at the *MAPT* locus exceeded Bonferroni-corrected GWAS significance thresholds. An overview about further PD studies and employed statistical tests is provided in Table 2.

2.2 Multiple-marker tests

Multivariate methods can be used as an alternative to testing variants individually by combining information across the variants and testing the multiple variant sites simultaneously.

TABLE 1 Advantages and disadvantages of methods for variant association testing.

Method	Advantages	Disadvantages
Single-marker tests	Standard method to test for association between variants and traits in GWAS, useful for large sample sizes and common variants with large effect sizes	Less powerful for rare variants with similar effect sizes to common variants, leading to the need for stringent significance levels in scenarios with more rare variants, further reducing its power
Multiple-marker tests	Evaluates the effects of multiple variants in a gene or region, instead of testing for each individually. Has higher power than single-marker tests when variants in a group are associated to the same trait or disease	Highly sensitive to allele frequencies
Burden tests	Powerful in scenarios when a large number of variants are causal with effects in the same direction	Lose power with small numbers of causal variants or in the presence of variants with effects in opposite directions
Adaptive burden tests	Uses fixed weights or thresholds to increase robustness	Computationally intensive
Variance-component tests	Powerful in scenarios with a small fraction of causal variants or in presence of variants with effects in opposite directions	Less powerful with large numbers of causal variants or if their effects are in the same direction
Linkage disequilibrium score regression	Robust against confounders and can be used efficiently with large sample sizes	Despite being computationally less intensive than other genetic correlation methods, a practical setback is the need of processing summary statistics from multiple GWAS which can be time consuming
Mendelian randomization	Overcomes limitations of traditional randomized control trials (RCTs) including proneness to confounders, reverse causation and selection bias	Multiple limitations include pleiotropy where a single variant can produce multiple effects, LD where two variants are statistically associated and tend to be inherited together, and bias of precise estimates of causal effects

In that case, a multiple-marker test's power will be higher than that of single-marker tests for multiple moderate SNP effects. Such approaches include Fisher's method, Hotelling's T^2 test, and multiple logistic or linear regression. These tests may be less powerful as they require multiple degrees of freedom.

Fisher's test combines the results of all m single-marker tests, and the test statistic can be represented by $X^2 = -2\sum_{i=1}^{m} \log(p_i)$, assuming p_i are the p-values obtained from the m single-marker tests. However, the test can be anti-conservative when there are dependencies among the m single tests.

Multiple regression can be used to test for the association between the variants and the phenotype in tandem instead of fitting m regression models at each of the rare variants separately. A simple regression model with no covariates for a binary trait can be represented as follows:

$$y_i = \alpha + X\beta_i + \varepsilon_i$$

where X is an $n \times m$ matrix of the minor allele counts for n subjects at m variants, and β is the m vector of regression coefficients. By estimating the associations at each variant collectively, the fit requires m degrees of freedom for the test statistics of each null hypothesis with $\beta_j = 0$ to have $n - m$ degrees of freedom rather than $n - 1$ as in single-marker tests.

Hotelling's two-sample T^2 is a multivariate generalization of the Student's t-test (Xiong et al., 2002) which can be used for case-control studies. Assume we have N_A affected and $N_{\bar{A}}$ unaffected samples. To calculate the test statistic, consider X_{ij} and Y_{ij} as variables defined for the genotype of marker j for individual i from the case and control groups. For N_A we find

$$X_{ij} = \begin{cases} 1, & \text{if} \quad aa \\ 0, & \text{if} \quad Aa \\ -1, & \text{if} \quad AA \end{cases}$$

and Y_{ij} is defined similarly for $N_{\bar{A}}$. Assume $X_i = (X_{i1}, \ldots, X_{im})^T$, $i = 1, \ldots, N_A$ for the cases and $Y_i = (Y_{i1}, \ldots, Y_{ik})^T$, $i = 1, \ldots, N_{\bar{A}}$ for controls, and after establishing the X_i and Y_i's pooled-sample covariance matrix S, Hotelling's two-sample T^2 test statistic can be expressed as

$$T^2 = \frac{N_A N_{\bar{A}}}{N_A + N_{\bar{A}}} (\bar{X} - \bar{Y})^T S (\bar{X} - \bar{Y})$$

and under the null hypothesis, $\frac{N_A + N_{\bar{A}} - m - 1}{m(N_A + N_{\bar{A}} - 2)} T^2$ follows an $F_{m, N_A + N_{\bar{A}} - m - 1}$ distribution.

A drawback to multiple-marker tests is their sensitivity to allele frequencies. A simulation study on rare variants by Li & Leal (Li and Leal, 2008) shows that Hotelling's T^2 test is greatly affected by MAF, and shows a reduction in power in cases of increased numbers of rare causal variants.

Li et al. (2021) used multivariate linear regression to test for variant association to age at onset of PD in the Asian population. Results showed a significant effect of a novel intergenic locus rs9783733 that could delay the age at onset in patients by 2.43 years. Another study by Pankratz et al. (2012) used logistic regression to identify genetic variants associated with pD. Genome-wide significance was reached for variants in SNCA, MAPT, GAK/DGKQ, HLA region and RIT2. Additional tests can be found in Table 2.

TABLE 2 Selected studies on risk variant association utilizing multiple techniques.

Author	Method	Objective	Results
Mata et al. (2017)	Single/Multiple-marker, linear regression/SKAT-O	Identify genetic variants leading to cognitive decline in PD patients	Eighteen common variants in thirteen genomic regions exceeded significance threshold
Simón-Sánchez et al. (2009)	Single-marker, Cochran-Armitage trend test	Studying variant association to PD in European patients	Four SNPs within the *SNCA* locus and three at the *MAPT* locus exceeded Bonferroni corrected GWAS significance threshold
Li et al. (2021)	Multivariate linear regression	Test for variant association to age at onset of PD in the Asian population	Identification of a novel intergenic locus which could delay age at onset of PD by 2.43 years
Tan et al. (2021)	Single-marker, linear regression	Identify genetic variants associated with PD progression	Significant association of *APOE* ε4 tagging variant rs429358 to composite and cognitive progression in PD
Foo et al. (2017)	Multiple logistic regression	Conduct the first Han Chinese GWAS for PD	Presence of some genetic heterogeneity in PD risk between European and East Asian patients
Hernandez et al. (2012)	Multiple-marker, logistic Regression	Identify genetic variants associated with young onset PD in Finnish Patients	Thirteen SNPs that were previously linked to PD showed high significance in the Finnish cohort. However, the study failed to identify any single predominant monogenic causes of the disease in the group
Loesch et al. (2021)	Multiple-marker, logistic regression	Identify PD risk variants in a Latino cohort and describe overlap in genetic structure compared to European ancestry	Genome wide significance shown by *SNCA* locus demonstrating its importance in PD etiology in Latinos
Park et al. (2021)	Multiple-marker, logistic Regression	Identify genetic loci associated with cognitive impairment in patients with sporadic PD	*RYR2* and *CASC17* loci were associated with cognitive impairment based on clinical assessment scores, but none of their SNPs based significance thresholds after Bonferroni correction
Pankratz et al. (2012)	Multiple-marker, logistic Regression	Identification of risk variants associated with PD susceptibility	GWAS significance was reached for previously reported *SNCA*, *MAPT* and *HLA* regions, as well as a novel susceptibility PD locus *RIT2* on chromosome 8
Hill-Burns et al. (2014)	Multiple-marker, logistic regression	Identification of novel PD locus via stratified GWAS study	Identification of a novel locus in chromosome 1p21 in sporadic PD.
Chang et al. (2017)	Multiple-marker, logistic regression	Identification of novel loci associated with PD risk	Identified 17 novel risk loci in a joint analysis of 26,035 cases and 403,190 controls
Hill-Burns et al. (2016)	Multiple-marker, linear regression/Cox regression	Conducting GWAS for age at onset	Two variants, mapped to *LHFPL2* and *TPM1*, were strongly associated to earlier onset PD.
Liu et al. (2011)	Multivariate logistic regression	Identification of risk variants associated to PD in an Ashkenazi Jewish population	The study identified 6 gene regions as candidates for PD using an Ashkenazi Jewish case-control population as discovery set and two other large dataset for replication
Hamza et al. (2010)	Multiple-marker, logistic regression	Conducting a GWAS to identify risk variants in Caucasian population	The study confirmed association with *SNCA* and *MAPT*, replicated *GAK* association and detected novel association with *HLA*, which was replicated in two other datasets
Ryu et al. (2020)	Multiple logistic regression/ Cochran-Armitage trend test	Identify genomic variants associated with motor fluctuations and levodopa-induced dyskinesia (LID)	*FAM129B* SNP rs10760490 was nominally associated with motor fluctuations at 5 years after PD onset, while *GALNT14* SNP rs144125291 was significantly associated to occurrence of LID
Rodrigo and Nyholt, (2021)	Multiple-marker, logistic regression	Reanalyzing an ExomeChip-based NeuroX dataset to identify novel, conditional and joint genetic effects associated with PD	Eleven association signals for PD were identified including five novel signals, three of which are driven by low frequencies and two by rare
Blauwendraat et al. (2019)	Multiple-marker, linear regression	Identification of genetic factors associated with age at onset of PD	Results found two GWAS significant signals at known PD risk loci *SNCA* and a protein-coding variant in *TMEM175*, and Bonferroni corrected signals at other known PD loci including *GBA*, *INPP5F/BAG3*, *FAM47E/SCARB2*, and *MCCC1*
Spencer et al. (2011)	Single/Multiple-marker, logistic regression	Performing a GWAS United Kingdom patients to identify novel risk factors associated to PD	Evidence found for PD independent association in 4q22/*SNCA*, weak but consistent association in previously published associated regions 4p15/*BST1*, 4p16/*GAK* and 1q32/*PARK16* and no significant association for previously reported SNP association in 12q12/*LRRK2*
Saad et al. (2011)	Multiple-marker, logistic regression	Performing a three-stage GWAS to identify common PD risk variants in the European population	Significant association of *SNCA* to PD risk, converging evidence of association with PD on 12q24 and confirming associations on 4p15/*BST1*, previously reported in Japanese data

TABLE 2 (*Continued*) Selected studies on risk variant association utilizing multiple techniques.

Author	Method	Objective	Results
Blauwendraat et al. (2020)	Multiple-marker, logistic/ linear regression	Understand whether genetic variants affect penetrance and age at onset of GBA-associated PD and Lewy body dementia (LBD)	Study shows PD and LBD cases with GBA variants often carry other PD associated risk variants that modify disease risk and age at onset
Spataro et al. (2015)	Combined multivariate and collapsing method (CMC)	Study the contribution of rare variants in the etiology of idiopathic PD	The tests showed significance of dominant genes when analyzing code-altering variants only, while they showed significance of recessive genes when analyzing code-altering, putative code-damaging and putative splice-altering variants
Li et al. (2020)	Weighted sum statistic (WSS)/SKAT-O	Study the association of DnaJ homolog C *DNAJCs* in a large Chinese early-onset PD cohort	Several risk variants showed significance in *DNAJC26*, *DNAJC13*, *DNAJC10* and *DNAJC6*, as well as a novel compound heterozygous mutation in *DNAJC6*
Nalls et al. (2019)	SKAT-O	generate summary statistics of genes passing the inclusion criteria of having at least two coding variants	Out of 113 genes, seven showed high significance including *LRRK2* and *GBA*
Siitonen et al. (2017)	SKAT-O	Identify genetic variants associated to early onset PD in Finnish patients	Novel associations were found in the *CEL* region. However, there is a high chance the finding is a false positive as the *CEL* region has multiple indel mutations
Markopoulou et al. (2021)	SKAT	Understanding the contribution of genetic variants at PD risk genes to individual phenotypic charactersitics of PD	Notable findings show association of *LRRK2* with a prior diagnosis of essential tremors, significant association of *NUCKS1* to Unified PD Risk Scale UPDRS-III motor scores and UPDRS-V (H&Y stage) and association of PD risk SNP rs823118 in the same gene to higher MMSE scores

2.3 Burden tests

Aggregation tests can be used to evaluate the combined effects of multiple variants in a gene or region, rather than testing each of them individually. One class of such tests is called burden tests, which collapse information of multiple variants into a single genetic score and test for its association to a trait (Morgenthaler and Thilly, 2007; Li and Leal, 2008; Zawistowski et al., 2010; Morris and Zeggini, 2010; Asimit et al., 2012). By counting minor alleles across all variants in a set, we can summarize the genotype information, and the statistic is represented by:

$$C_i = \sum_{j=1}^{m} w_j G_{ij}$$

where G_{ij} represents the allele counts of subject i at variant j, and w_j is the weight for variant j.

The summary genetic score C_i can adapt to different assumptions about disease mechanisms. The MZ test (Morris and Zeggini, 2010) utilizes a dominant genetic model instead of an additive one to calculate C_i, which is the number of rare variants for which individual i carries at least a single copy of the minor allele. As for the cohort allelic sums test (CAST) (Morgenthaler and Thilly, 2007), it assumes an increase in disease risk with the presence of any rare variant, and sets the genetic score $C_i = 0$ if there are no minor alleles in the region and $C_i = 1$ otherwise.

We can focus on rare variants by assuming $w_j = 1$ when the MAF of the variant j MAF_j is smaller than a preset threshold or

$w_j = 0$ if otherwise. We can upweight rare variants by using a continuous weight function. Madsen and Browning (Madsen and Browning, 2009) proposed $w_j = 1/[MAF_j(1 - MAF_j)]^{1/2}$ and Wu et al. (2011) proposed the family of Beta densities $w_j = beta$ ($MAF_j, \alpha_1, \alpha_2$) which includes the Madsen and Browning weight as a special case. Information on the functional effects of variants can also be used for weight construction.

Outside of the regression framework, several burden approaches have been presented. The combined multivariate and collapsing method (CMC) (Li and Leal, 2008) collapses rare variants as in CAST, but in different MAF categories and calculates the combined effects of the variants using Hoteling's t test. The Madsen and Browning weighted-sum test (WST) (Madsen and Browning, 2009) uses Wilcoxon's rank-sum test and obtains the p-values by permutation.

All rare variants in a set are assumed to be causal and related to a trait with the same direction and magnitude by burden techniques. Breaking such assumptions can result in a significant loss of power (Neale et al., 2011; Lee et al., 2012a).

Spataro et al. (2015) used different collapsing methods, including the CMC and weighted sum tests, to study the contribution of rare variants in the etiology of idiopathic pD. The tests showed high significance in a Mendelian group of genes that comprise genes of dominant and recessive inheritance. In dominant genes, the tests showed high significance only when analyzing code-altering variants. As for recessive genes, the tests showed significance for code-altering, putative code-damaging, and putative splice-altering variants. Another study by Li et al. (2020) used the weighted sum statistic (WSS) to study the associations of the DNAJC proteins family by genetic analysis

to early onset PD in a large Chinese cohort. The study identified 61 rare variants, two of which showed significance after Bonferroni correction in *DNAJC26*, two in *DNAJC13*, one in *DNAJC10* and one more in *DNAJC6*, as well as a novel compound heterozygous mutation in *DNAJC6*. An overview of further studies using burden tests can be found in Table 2.

2.4 Adaptive burden tests

Adaptive methods were developed to address the limitations posed by the traditional burden tests. These methods are robust in presence of null variants and allow for train-increasing or trait-decreasing variants. Han et al. (Han and Pan, 2010) developed a data-adaptive sum test (aSum) that performs a burden test with estimated directions after first estimating the direction of effect for each variant in a marginal model. It assigns $w_j = -1$ when β_j is likely to be negative and $w_j = 1$ if not. This approach requires permutation for the *p*-values to be calculated. This procedure is improved in the step-up test (Hoffmann et al., 2010), which uses a model-selection framework that assigns $w_j = 0$ when a variant is unlikely to be associated, removing it from consideration.

A more direct approach is utilized by the estimated regression coefficient test (EREC) (Lin and Tang, 2011), which uses estimated regression coefficients for each variant as weights. This is based on the assumption that the true regression coefficient β_j is an optimal weight to maximize power. When minor allele counts (MAC) are small, β_j estimates are unstable, and hence the EREC test stabilizes the estimates by adding a small constant to the estimated β_j, which might reduce the test's optimality. The test uses parametric bootstrap to estimate *p*-values because asymptotic approximation of the test statistic is only accurate for very large samples.

The variable threshold (VT) (Price et al., 2010) is an adaptive modification that chooses the best frequency thresholds for rare variant burden testing and calculates *p*-values analytically or by permutation. Using kernel-based adaptive weighting, the kernel-based adaptive cluster (KBAC) (Liu and Leal, 2010) method combines variant classification of non-risk and risk variants with association tests.

As referenced in the previous section, Li et al. (2020) included the aSUM and KBAC tests with the WSS test to study the associations of the DNAJC proteins family to early onset *p*D. Further information and results of the study can be found in Table2.

2.5 Variance-component tests

This type of association tests uses a variance-component test within a random-effects model and tests for the association of a group of variants by evaluating the distribution of their genetic effects. These tests include the C-alpha test (Neale et al., 2011), the sequence kernel association test (SKAT) (Wu et al., 2010; Wu

et al., 2011), and the sum of squared score test (Pan, 2009). These tests evaluate the distribution of aggregated score test statistics of the individual variants.

SKAT is a non-burden test that uses mixed models and includes the C-alpha test in special cases when covariates are absent, and can also accommodate SNP-SNP interactions. SKAT assumes the regression coefficients β_j are independent and follow a distribution with mean 0 and variance $w_j^2 \tau$, and tests the hypothesis H_0: $\tau = 0$ using a variance-component score test. The SKAT test statistic can be represented as

$$Q_{\text{SKAT}} = \sum_{j=1}^{m} w_j^2 S_j^2$$

which is a weighted sum of squares of the single-variant score statistic S_j. Similar to burden tests, SKAT is robust to groups that include variants with both positive and negative effects, as it collapses S_j^2. When comparing burden and SKAT statistics, it is noted that burden tests collapse the variants first before performing the regression, while SKAT collapses individual variant-test statistics, which explains its robustness to mixed signs of β and large fractions of non-causal variants.

While burden tests are not powerful when the target region has several noncausal variants or causal variants of different associations, they can outperform SKAT in cases where a high proportion of causal variants with effects in a similar direction are present. Lee et al, (2012b) proposed a unified test that is optimal in both scenarios and combines both burden tests and SKAT in a single framework. The test statistic of the unified test is

$$Q_\rho = \rho Q_B + (1 - \rho) Q_S, 0 \le \rho \le 1$$

which is a weighted average of SKAT and burden tests, which reduces to SKAT when $\rho = 0$ or the burden test when $\rho = 1$.

In their meta genome-wide association study, Nalls et al. (2019) used SKAT-O to generate summary statistics of genes with rare coding variants which had an imputation quality larger than 0.8%. 113 genes passed the inclusion criteria of having at least two coding variants. After Bonferroni correction for the 113 genes, seven significant genes were identified including LRRK2 and GBA. Siitonen et al. (2017) also used SKAT-O in their study to analyze variants associated with early onset PD in Finnish patients. The results showed significant associations to PD in the CEL locus which were not previously identified. However, the validity of the result is questioned by the fact that the CEL region has several indel mutations (Taylor et al., 1991; Siitonen et al., 2017).

2.6 Linkage disequilibrium score regression

Linkage Disequilibrium score regression (LDSC) is a method developed by Bulik-Sullivan et al. (2015) that determines if the

distribution of a test statistic in GWAS is inflated due to confounding biases or polygenicity. The idea behind LDSC is that variants in linkage disequilibrium (LD) with a causal variant in an association analysis will show elevated test statistics that are proportional to the LD with the causal variant, while elevations due to confounders like cryptic relatedness or population stratification will not correlate with the LD score. LDSC involves using regression techniques to study the relationship between LD scores and test statistics of SNPs obtained from GWAS studies.

Nalls et al. (2019) used LDSC in their GWAS to examine correlations of PD genetics with that of other traits and diseases using data obtained from GWAS available via LD Hub (Zheng et al., 2017) and biomarker GWAS summary statistics on c-reactive protein and cytokine measures. p-values obtained from the LDSC were adjusted for FDR to account for multiple testing. The authors found four significant correlations, two of which were positive correlations with intracranial volume and putamen volume, and two negative correlations with tobacco use and educational attainment.

Tirozzi et al. (2020) wanted to investigate the genetic overlap between PD and platelet parameters since associations between both have been established but not thoroughly investigated on a genetic level. The authors applied LDSC to summary statistics of a large independent GWAS conducted on Alzheimer's disease (AD), PD, and platelet parameters including mean platelet volume (MPV), platelet count (PLT), and platelet distribution width (PDW) (Jansen et al., 2019). The results showed a significant correlation between PDW and PD risk suggesting the existence of genetic overlap and presenting PDW as a new potential biomarker for PD.

Another study by Andersen et al. (2021) investigates how the immune system contributes to pathogenesis in PD, by studying the enrichment of common variant heritability for PD stratified by immune and brain cell types. The authors performed a stratified LDSC (s-LDSC) analysis using full summary statistics from the meta-analysis of PD GWAS by Nalls et al. (2019) and an earlier meta-analysis by Chang et al. (2017). The results found significant enrichment in open chromatin regions of microglia, with further investigation of expression quantitative locus (eQTL) databases showing the P2RY12 locus to be the most interesting, suggesting it as a microglial gene with PD association signal.

2.7 Mendelian randomization

Mendelian Randomization (MR) is a method that uses measured variation in genes of known function to study the causal effects of a modifiable exposure on disease or health-related outcomes (Lawlor et al., 2008). MR studies use genetic variants as instrumental variables (IV) which can be defined as

FIGURE 2
An illustration of Mendelian Randomization framework.
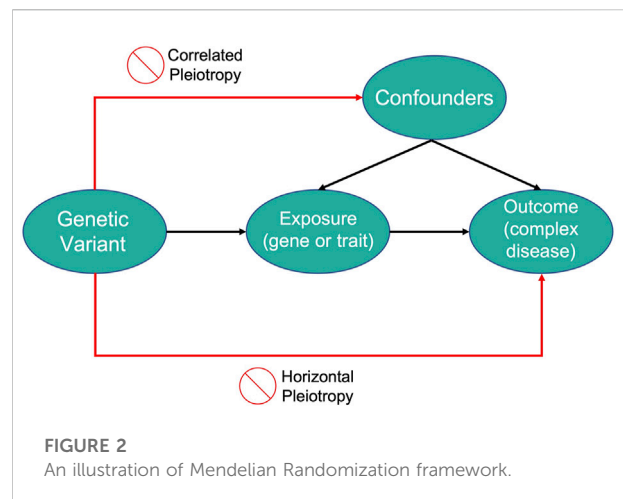
variables that are associated with an outcome only through their robust association with an intermediary variable.

In this context, the aim of MR studies is not to identify genetic variants that are directly associated with the disease but to use the variants as IVs for the modifiable exposure of interest. The genetic variants need to satisfy three assumptions to be considered as IVs in MR studies:

- The variant is associated with the modifiable exposure
- The variant is independent of confounding factors that confound the association of the modifiable exposure to the outcome
- The variant is independent of the outcome given the modifiable exposure and the confounding factors

Therefore, genetic variants that explain variations in an exposure can be used as a proxy to explain how changes in that exposure can influence the outcome of a disease of interest. An illustration of the MR framework is shown in Figure 2.

MR was used by Simon et al. (2014) to investigate whether genetic variants that can predict serum urate levels can predict the rate of progression in patients with early PD, on the basis that higher serum urate levels lower the risk of developing pD. In this study, the authors used SLC2A9 gene as an IV, which explains most of the genetically specified variability in serum urate levels but does not have any known direct associations with the central nervous system. The authors then estimated the association between genetically determined urate levels and PD progression using two-stage regression, where they first fitted a generalized linear regression model with urate levels as the dependent variable, and a SLC2A9 score based on the number of minor alleles at three selected loci, along with potential confounders, as independent variables. Then, a Cox proportional hazards model used the predicted urate levels from the first stage regression as a continuous independent

variable to determine its association with PD progression. The results showed that an increase in the number of *SLC2A9* minor alleles is associated with a decreased serum urate level. Also, the rate of PD progression increased with the number of minor *SLC2A9* alleles associated with lower serum urate levels. Genetic variants other than *SLC2A9* did not show any significant association to lower serum urate levels or rapid PD progression. The results suggest that high serum urate levels are protective of rapid progression in early PD.

Similarly, a study by Domenighetti et al. (2022) uses MR to investigate the association between genetically predicted dairy intake and higher PD risk by using the *LCT* lactase gene's minor allele rs4988235 as an IV, where TT/TC genotypes are associated with lactase persistence and the ability to digest lactose and CC genotype with non-persistence. The authors then used logistic regression to compare the frequency of rs4988235-TC+TT genotypes in patients and controls of European ancestry. Results showed that rs4988235-TC+TT genotypes were more frequent in PD patients than controls, suggesting that higher dairy intake increases PD risk.

Another study by Storm et al. (2021) uses MR to investigate several druggable genes and predict their efficacy as PD drug targets. In this study, the authors considered the expression levels of the druggable genes as the modifiable exposure, while variants associated with expression levels of the genes, called eQTLs, were used as the IVs. The authors sought to use openly available eQTL data for genes under investigation to mimic exposure to corresponding medications (Finan et al., 2017). First, the authors used the cohort collected for the meta-analysis by Nalls et al. (2014). The causal estimates, known as the Wald ratio, were calculated for each SNP, and the ratios were weighted by inverse-variance (IVW) for genes with more than one eQTL available. This identified 31 genes with genetically-determined expression that is highly associated with PD risk. The authors then attempted to replicate the genes with significant association with PD risk in an independent cohort that does not overlap with the original cohort. The authors then used several meta-analysis methods to look for pleiotropy due to confounders including IVW, the MR-Egger intercept test, Cochran's Q test, and the I^2 test. Based on the results, the authors propose the genes *CTSB*, *GPNMB*, *CD38*, *RHD*, *IRAK3* and *LMAN1* as drug targets with the strongest MR evidence.

As previously discussed, Nalls et al. (2019) identified correlations of PD genetics with tobacco consumption, educational attainment, and brain volumes using LDSC. The authors used MR to assess the existence of a causal relationship between PD and the traits. The results showed that cognitive performance and educational attainment had a large causal effect on PD risk, while smoking and brain volumes did not have any significant causal relationship.

2.8 Multiple testing corrections

Multiple testing is one of the major concerns regarding high-dimensional data which results from simultaneous testing of multiple hypotheses, which if not taken into consideration, may lead to rejecting a true null hypothesis by chance, known as a false discovery. This can be accounted for by controlling an appropriate error rate such as the family-wise error rate (FWE) which is the probability of one or more false discoveries. The classical method of controlling FWE is the Bonferroni method (Bland and Altman, 1995), which is an adjustment made to *p*-values when several tests are performed. To perform a Bonferroni correction, assume the critical *p*-value to be α, then divide it by the number of tests made *n*. The new critical *p*-value would then be α/n, and the statistical power of the study is then calculated based on the newly modified *p*-value.

Another method is the Benjamini–Hochberg method which controls the false discovery rate (FDR) (Benjamini and Hochberg, 1995), known as the expected proportion of false rejections out of all rejections. The Benjamini–Hochberg procedure involves ordering all *p*-values from smallest to largest then assigning a ranking to each one, then calculating the critical *p*-value as $(i/m)Q$, where *i* is the rank of the *p*-value, *m* is the total number of tests and *Q* is the chosen FDR. The method then checks the largest *p*-value below the critical rate, and considers any smaller values as significant.

3 Polygenic risk score

The risk of polygenic disorders such as PD cannot be assessed by information conferred from a single variant, but the total set of risk variants that comprise its genetic architecture is required to provide enough information that can help identify individuals at high-risk (Lewis and Vassos, 2020). An individual's risk can be assessed using polygenic risk scores (PRS), calculated as the sum of risk alleles an individual carries, each weighted by their relative effect sizes obtained from the GWAS summary statistics (Ibanez et al., 2019), where the result is a score that represents the individual's genetic load for the disease or trait in question.

In this context linkage disequilibrium (LD) and *p*-values thresholds for individual SNPs have to be considered. Simpler approaches, such as PRSice (Choi and O'Reilly, 2019) and PLINK (Purcell et al., 2007; Gaunt et al., 2007; Chang et al., 2015), only use *p*-value thresholds (clumping + thresholding), whereas more advanced methods, including LDPred (Vilhjálmsson et al., 2015), PRS-CS (Ge et al., 2019), JAMPred (Newcombe et al., 2019), and Lassosum (Mak et al., 2017) additionally take into account based on reference data.

While PRS can provide a simple estimate of the genetic architecture of complex disorders, its additive model generally does not take into account gene-gene interactions (Aschard,

TABLE 3 Polygenic Risk Scores listed in the Polygenic Score Catalog (Lambert et al., 2021).

Author	Reported traits	Ancestry Distribution	Number of variants
Pihlstrøm et al. (2016)	Parkinson's disease, motor decline	European	19
Ibanez et al. (2017)	Parkinson's disease, age at onset	European	16
Paul et al. (2018)	Parkinson's disease, cognitive decline, motor decline	European	23
Nalls et al. (2019)	Parkinson's disease	Multi-ancestry	90, 1805
Bobbili et al. (2020)	Parkinson's disease	European	43
Liu et al. (2021)	Parkinson's disease dementia	Multi-ancestry	3
Sia et al. (2021)	Parkinson's disease	East Asian	6
Chairta et al. (2021)	Parkinson's disease	European	12

2016). Moreover, the typically required pre-filtering of SNPs implies a focus on more common genetic variants.

The largest meta-analysis was performed by Nalls et al. (2014) and was considered the reference for PD-related PRS before including more data from the 23andME (Chang et al., 2017) meta-analysis. The included PRS were associated with PD status, faster motor and cognitive decline (Paul et al., 2018) and age at onset of disease. Another study by Escott-Price et al. (2015) mentions that only PRS built from SNPs with *p*-values below the significant thresholds were associated with PD, suggesting that the genetic architecture of PD includes several common variants with small effects. Another study by Ibanez et al. (2017) shows that PRS from more significant SNPs are also associated with PD risk. Furthermore, PRS were used to show a higher genetic burden in early-onset PD than in late-onset PD (Escott-Price et al., 2015). More studies with established PRS in the PD field can be found in Table 3. A more detailed review of PRS in the PD field can be found in (Dehestani et al., 2021).

4 The perspective of machine learning

4.1 Multi-modal data integration

There is an increasing awareness that PD has to be understood as a complex disease, in which aging, (epi-)genetic variants, environmental pollutants/toxins, lifestyle, and comorbidities jointly contribute to the observed phenotype (Espay et al., 2017; Titova and Chaudhuri, 2017). Whereas variants association tests and PRS have helped to gain a better understanding of the genetic basis of PD, developing algorithms for accurate disease risk assessment, diagnosis, prognosis, and treatment response in the context of precision medicine require combining PRS as well as relevant genetic variants with further data modalities. Hence, predictive machine learning models are needed, which can potentially also overcome one of the typical limitations of PRS, namely lacking variant interactions and thus non-linearities. A recent study shows the combined role of PRS,

rare high-impact variants, and family history in PD risk (Hassanin et al., 2021). Cope et al. demonstrated that a non-linear machine learning algorithm purely trained on genetic variants can result in dramatically improved prediction performances compared to a classical PRS (Cope et al., 2021). Notably, analysis of the model allowed us to identify an interaction between variants in *TMEM175* (coding for a potassium channel in late endosomes) and *GAPDHP25* (glyceraldehyde-3 phosphate dehydrogenase pseudogene 25), which have been linked to PD (Nalls et al., 2014). Another study by Prashanth et al. (2016) used multimodal features to classify early PD subjects from controls using machine learning models. The authors used non-motor features of Rapid Eye Movement (REM) sleep Behaviour Disorder (RBD) and olfactory loss as well as cerebrospinal fluid (CSF) measurements and dopaminergic imaging markers to classify the patients using Naive Bayes, Support Vector Machine (SVM), Boosted Trees and Random Forest classifiers, where SVM gave the highest performance. Based on the results, the authors suggest that the combination of non-motor, CSF, and imaging features can help in the preclinical diagnosis of PD.

A further example is the use of non-linear unsupervised machine learning algorithms by Emon et al. (2020) to identify patient subgroups by exploring the genetic burden by SNPs in genes that have been previously associated with AD and PD, which allowed for a molecular mechanism based stratification of AD and PD patient sub-types. The authors further investigated clinical outcome measures of the patients to confirm whether the patient clusters were disease-associated or reflected general genetic variations in the population and found the clusters to be associated with different clinical symptoms, pathophysiological brain differences, and biological processes that were enriched only in each of the clusters.

Experiences from neurological conditions other than PD suggest that combinations of PRS, (non-linear) combinations of genetic variants, pathway-level burden scores and a detailed description of the clinical phenotype could allow for a rather accurate prediction of disease risk (Khanna et al., 2018; Birkenbihl et al., 2020) and even clinical drug response (de

Jong et al., 2021). Interestingly, in both cases, genetic factors played a comparably small role in the prediction of the clinical outcome. In another study, Makarious et al. (2022) demonstrate the benefits of using multiple data modalities by integrating clinical, genetic, and transcriptomic data in a predictive machine learning framework. Their results showed that integrating multiple data modalities improved PD prediction in mixed populations of cases and controls. They also demonstrated the benefits of using machine learning approaches and the ability to tune the models' parameters and accommodate nonlinearities, as well as identifying important features that contributed the most to the models' predictive performance using model explanation methods such as SHAPley Additive exPlanations (SHAP).

4.2 Deep phenotyping

A few studies have started to focus on genetic risk factors associated to symptoms of idiopathic PD, including cognitive impairment (Collins and Williams-Gray, 2016; Amer et al., 2018; Planas-Ballvé and Vilas, 2021). In this context, it has to be re-emphasized that PD patients suffer from a whole spectrum of motor and non-motor symptoms. Traditionally, these symptoms are assessed via questionnaires, such as the Unified Parkinson's Disease Rating Scale (UPDRS), during a patient's visit to a medical specialist center. The assessment is dependent on the experience of the individual examiner and can thus be subjectively biased. Therefore, during the last years, there has been a strongly growing interest in remote monitoring techniques (RMTs), including wearable sensors and devices (measuring e.g. gait) and smartphone apps (measuring e.g. cognitive abilities). Compared to established questionnaire-based assessments, RMTs offer several potential benefits:

1) They are patient-centric and not biased by a rater's experience.
2) They allow for monitoring disease symptoms within a patient's natural at-home environment, potentially 24/7, hence considering the fact that PD symptoms are variable over the daytime. RMT signals can thus be viewed as real-world data.
3) Digital sensing techniques provide an objective measure of a clinical symptom.

Notably, processing of RMT signals requires advanced data analytical techniques, including machine learning (Fröhlich et al., 2022). The outcome is an abstract set of features representing a patient's phenotype. Following sufficient validation, within clinical studies, these features can result in digital biomarkers, which provide an accurate and quantitative description of PD symptoms. The combination with genetic data thus opens completely new opportunities to identify risk factors for specific PD symptoms, such as cognitive impairment or sleep disturbances. Moreover, machine learning algorithms could potentially be used to combine digital biomarkers with genetic features and other data modalities, including electronic health records, to predict disease risk, prognosis, and response to treatment.

4.3 Parkinson's disease prediction

Multiple studies have used machine learning models to predict PD using different data modalities, analyzing hidden information in data that cannot be interpreted in clinical diagnosis. Wang et al. (2020) investigated the diagnosis of PD based on vowel phonation. Features were obtained from the mPower dataset and improved with additional novel features using a Bayesian correlated t-test. The features were then used as input for an SVM model which performed with moderate accuracy. Bhurane et al. (2022) used SVM with a cubic kernel to classify PD patients and healthy controls. Using features extracted from Electroencephalography (ECG) signals, the proposed approach performed with high accuracy.

Chakraborty et al. (2020) used features extracted from 3T T1-MRI scans to detect neurodegeneration in pD. Using atlas-based segmentation, eight subcortical structures were segmented from the MRI scans, on which feature extraction was performed to extract textural, morphological, and statistical features. The features were then used to train four different machine learning algorithms: an artificial neural network (ANN), XGBoost model, random forest classifier, and an SVM, where the ANN model performed with the highest accuracy. In another study, Ali et al. (2019) used neural networks to detect PD using features obtained from acoustic analysis of voice signals. Linear discriminant analysis (LDA) was used for dimensionality reduction, and a genetic algorithm (GA) to optimize the hyperparameters of the neural network. Initially, the model performed with accuracy which falls after excluding gender-dependent features to eliminate bias.

Peng et al. (2019) used a three-step method for PD gene prediction. The method, called N2A-SVM, uses the Node2vec algorithm to extract vector representations of each gene in the protein-protein interaction (PPI) network. Then it uses an autoencoder to reduce the dimensions of the obtained vector, and an SVM for classification. The performance of N2A-SVM was tested in comparison to the other methods: random walk with restart (RWR) (Li and Patra, 2010), shortest path length (SPL) (Krauthammer et al., 2004) and Euclidean distance (ED) (Díaz-Uriarte and Alvarez de Andrés, 2006), where N2A-SVM showed the highest performance.

Another study by Rastegar et al. (Ahmadi Rastegar et al., 2019) used machine learning models to assess if serum cytokine levels can be used to predict PD progression. The authors used data from the Michael J Fox Foundation *LRRK2* clinical cohort

consortium to assess the variability of inflammatory cytokine levels in patients over a one-year period. Then, the authors used the cytokine measurements with elastic net and random forest models to predict longitudinal clinical outcomes. Using baseline cytokine measurements, random forest models of motor severity showed the best predictive performance, with cytokines $MIP1\alpha$ and $MCP1$ contributing the most to the predictive model.

5 Discussion

The heterogeneous nature of PD imposes specific challenges for finding the underlying genetic causes. We briefly discussed several association tests that were used to identify genetic variants associated with disease risk. Single-marker tests are the simplest approach to studying associations by applying a univariate test to each variant and assessing their significance. However, their statistical power is low for small datasets and requires corrections for multiple testing. These issues were addressed by developing statistical methods that evaluate the associations of multiple variants in specific regions or genes. They are used as a standard method to test for variant association in GWAS, and helped identify several variants associated with PD including *SNCA*, *MAPT*, *GBA* and *HLA* loci as well as others associated with cognitive decline in PD including *APOEε4*, *RYR2* and *CASC17* loci.

Burden tests collapse multiple genetic variants into a single genetic score, which is used to test the association to a trait. Since these tests assume all collapsed rare variants to be causal and associated with the trait under study in a similar direction and magnitude of effect, any changes in said assumptions lead to a loss in their statistical power. Adaptive burden tests address these limitations as they require fewer assumptions about the genetic architecture at each locus, and hence they are suitable in the presence of null variants and trait-increasing or decreasing variants. However, adaptive tests are two-step procedures that may require regression coefficient estimation of individual variants as a first step and can be unstable for rare variants. They also estimate p-values by computationally intensive permutation. The use of burden tests helped us understand the role of different variant types in the etiology of idiopathic PD, and the identified four mendelian mutations of *LRRK2* and *PARK2* loci in idiopathic PD cases (Spataro et al., 2015). Adaptive tests were also used to study the associations of DNAJC proteins family with early onset PD (Li et al., 2020).

Variance-component tests evaluate the distribution of genetic effects for groups of variants to test for their association. Instead of aggregating the variants, they assess the distribution of each of the variants' aggregated score test statistics. Variance-component tests are more powerful than burden tests if the genetic region under study has many non-causal variants or variants with different directions of association,

while burden tests are more powerful when there are more causal variants with the same direction of association. SKAT-O combines both burden tests and SKAT in a single framework but can be less powerful than any of its components if their underlying assumptions are largely true. Nalls et al. (2019) used SKAT-O in their meta GWAS to identify genes with two or more rare coding variants, and 7 significant genes: *LRRK2*, *GBA*, *CATSPER3*, *LAMB2*, *LOC442028*, *NFKB2* and *SCARB2*. SKAT has also been used to study the association of genetic variants to individual phenotypic characteristics of PD, including motor and cognitive functions (Markopoulou et al., 2021).

LD score regression helped researchers distinguish whether inflated GWAS test statistic distributions are due to variants in LD with a causal variant or due to confounding bias or polygenicity. LDSC has been used to examine correlations of PD genetics with different traits, including brain measurements, blood measurements, habitual behaviors, and immune system activity in different cell types (Nalls et al., 2019; Tirozzi et al., 2020; Andersen et al., 2021).

Mendelian Randomization helped understand the causal effects of modifiable exposures on pD. The method uses the genetic variants as instrumental variables in statistical analysis to describe the relationship between the disease and the modifiable exposure of interest. MR was used to investigate the relationship between PD and serum urate levels, suggesting that elevated urate levels are protective of rapid progression in early PD (Simon et al., 2014). MR was used as well to investigate the relationship with lactose tolerance in different PD patient populations, suggesting that high tolerance and increased dairy intake elevate PD risk (Domenighetti et al., 2022). MR also helped propose druggable targets by investigating the expression levels of druggable genes and using them as the modifiable exposure of interest (Storm et al., 2021).

PRS have opened the possibility to assess disease risk on an individual basis rather than purely on the average population level. Limitations of PRS include their additive nature, which neglects gene-gene interactions, and the focus on more common genetic variants. Machine learning models can mitigate this limitation and additionally include further data modalities, such as other molecular and phenotypic data. In this context, electronic health records, as well as digital biomarkers, could help to longitudinally and more objectively characterize disease symptoms. The main challenge with the use of machine learning models is, however, their difficult interpretation, specifically in the case of neural networks. Novel approaches coming from the field of Explainable AI (XAI) could here provide a solution (Linardatos et al., 2020; Arrieta et al., 2020).

In summary, novel methodological developments are necessary to deepen the understanding of the genetic basis of PD and to transfer these insights into better individualized treatment of PD in the future.

Author contributions

MA, HF and PK contributed to arrangement of the literature review. MA wrote the first draft of the manuscript, HF and PK wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmmed.2022.933383/full#supplementary-material

References

Aasly, J. O. (2020). Long-term outcomes of genetic Parkinson's disease. *J. Mov. Disord.* 13 (2), 81–96. doi:10.14802/jmd.19080

Ahmadi Rastegar, D., Ho, N., Halliday, G. M., and Dzamko, N. (2019). Parkinson's progression prediction using machine learning and serum cytokines. *NPJ Park. Dis.* 5, 14. doi:10.1038/s41531-019-0086-4

Ali, L., Zhu, C., Zhang, Z., and Liu, Y. (2019). Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network. *IEEE J. Transl. Eng. Health Med.* 7, 2000410. doi:10.1109/JTEHM.2019.2940900

Amer, H., Shehata, H., Rashed, L. A., Helmy, H., El-Jaafary, S., Sabbah, A., et al. (2018). Genetic influences on cognition in idiopathic Parkinson's disease. *Neurol. Res. Int.* 2018, 5603571. doi:10.1155/2018/5603571

Andersen, M. S., Bandres-Ciga, S., Reynolds, R. H., Hardy, J., Ryten, M., Krohn, L., et al. (2021). Heritability enrichment implicates microglia in Parkinson's disease pathogenesis. *Ann. Neurol.* 89 (5), 942–951. doi:10.1002/ana.26032

Arrieta, A. B., Díaz-Rodríguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* 58, 82–115. doi:10.1016/j.inffus.2019.12.012

Aschard, H. (2016). A perspective on interaction effects in genetic association studies. *Genet. Epidemiol.* 40 (8), 678–688. doi:10.1002/gepi.21989

Asimit, J., and Zeggini, E. (2010). Rare variant association analysis methods for complex traits. *Annu. Rev. Genet.* 44, 293–308. doi:10.1146/annurev-genet-102209-163421

Asimit, J. L., Day-Williams, A. G., Morris, A. P., and Zeggini, E. (2012). ARIEL and AMELIA: Testing for an accumulation of rare variants using next-generation sequencing data. *Hum. Hered.* 73 (2), 84–94. doi:10.1159/000336982

Bandres-Ciga, S., Ahmed, S., Sabir, M. S., Blauwendraat, C., Adarmes-Gómez, A. D., Bernal-Bernal, I., et al. (2019). The genetic architecture of Parkinson disease in Spain: Characterizing population-specific risk, differential haplotype structures, and providing etiologic insight. *Mov. Disord.* 34 (12), 1851–1863. doi:10.1002/mds.27864

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bhurane, A. A., Dhok, S., Sharma, M. D., Rajamanickam, Y., Murugappan, M., and Acharya, U. R. (2022). Diagnosis of Parkinson's disease from electroencephalography signals using linear and self-similarity features. *Expert Syst.* 39. doi:10.1111/exsy.12472

Billingsley, K. J., Bandres-Ciga, S., Saez-Atienzar, S., and Singleton, A. B. (2018). Genetic risk factors in Parkinson's disease. *Cell Tissue Res.* 373 (1), 9–20. doi:10.1007/s00441-018-2817-y

Birkenbihl, C., Emon, M. A., Vrooman, H., Westwood, S., Lovestone, S., Hofmann-Apitius, M., et al. (2020). Differences in cohort study data affect external validation of artificial intelligence models for predictive diagnostics of dementia - lessons for translation into clinical practice. *EPMA J.* 11 (3), 367–376. doi:10.1007/s13167-020-00216-z

Bland, J. M., and Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ* 310 (6973), 170. doi:10.1136/bmj.310.6973.170

Blauwendraat, C., Heilbron, K., Vallerga, C. L., Bandres-Ciga, S., von Coelln, R., Pihlstrøm, L., et al. (2019). Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and α-synuclein mechanisms. *Mov. Disord.* 34 (6), 866–875. doi:10.1002/mds.27659

Blauwendraat, C., Reed, X., Krohn, L., Heilbron, K., Bandres-Ciga, S., Tan, M., et al. (2020). Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia. *Brain* 143 (1), 234–248. doi:10.1093/brain/awz350

Bobbili, D. R., Banda, P., Krüger, R., and May, P. (2020). Excess of singleton loss-of-function variants in Parkinson's disease contributes to genetic risk. *J. Med. Genet.* 57 (9), 617–623. doi:10.1136/jmedgenet-2019-106316

Bodmer, W., and Bonilla, C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* 40 (6), 695–701. doi:10.1038/ng.f.136

Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., et al. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47 (3), 291–295. doi:10.1038/ng.3211

Chairta, P. P., Hadjisavvas, A., Georgiou, A. N., Loizidou, M. A., Yiangou, K., Demetriou, C. A., et al. (2021). Prediction of Parkinson's disease risk based on genetic profile and established risk factors. *Genes* 12 (8), 1278. doi:10.3390/genes12081278

Chakraborty, S., Aich, S., and Kim, H. C. (2020). 3D textural, morphological and statistical analysis of voxel of interests in 3T MRI scans for the detection of Parkinson's disease using artificial neural networks. *Healthc. (Basel)* 8 (1), 34. doi:10.3390/healthcare8010034

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. doi:10.1186/s13742-015-0047-8

Chang, D., Nalls, M. A., Hallgrímsdóttir, I. B., Hunkapiller, J., van der Brug, M., Cai, F., et al. (2017). A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* 49 (10), 1511–1516. doi:10.1038/ng.3955

Choi, S. W., and O'Reilly, P. F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* 8, giz082. giz082. doi:10.1093/gigascience/giz082

Collins, L. M., and Williams-Gray, C. H. (2016). The genetic basis of cognitive impairment and dementia in Parkinson's disease. *Front. Psychiatry* 7, 89. doi:10.3389/fpsyt.2016.00089

Cope, J. L., Baukmann, H. A., Klinger, J. E., Ravarani, C. N. J., Böttinger, E. P., Konigorski, S., et al. (2021). Interaction-based feature selection algorithm outperforms polygenic risk score in predicting Parkinson's disease status. *Front. Genet.* 12, 744557. doi:10.3389/fgene.2021.744557

de Jong, J., Cutcutache, I., Page, M., Elmoufti, S., Dilley, C., Fröhlich, H., et al. (2021). Towards realizing the vision of precision medicine: AI based prediction of clinical drug response. *Brain.* 144 (6), 1738–1750. doi:10.1093/brain/awab108

Dehestani, M., Liu, H., and Gasser, T. (2021). Polygenic risk scores contribute to personalized medicine of Parkinson's disease. *J. Pers. Med.* 11 (10), 1030. doi:10.3390/jpm11101030

Díaz-Uriarte, R., and Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinforma.* 7, 3. doi:10.1186/1471-2105-7-3

Domenighetti, C., Sugier, P. E., Ashok Kumar Sreelatha, A., Schulte, C., Grover, S., Mohamed, O., et al. (2022). Dairy intake and Parkinson's disease: A mendelian randomization study. *Mov. Disord.* 37 (4), 857–864. doi:10.1002/mds.28902

Dunn, A. R., O'Connell, K. M. S., and Kaczorowski, C. C. (2019). Gene-by-environment interactions in Alzheimer's disease and Parkinson's disease. *Neurosci. Biobehav. Rev.* 103, 73–80. doi:10.1016/j.neubiorev.2019.06.018

Emon, M. A., Heinson, A., Wu, P., Domingo-Fernández, D., Sood, M., Vrooman, H., et al. (2020). Clustering of Alzheimer's and Parkinson's disease based on genetic burden of shared molecular mechanisms. *Sci. Rep.* 10 (1), 19097. doi:10.1038/s41598-020-76200-4

Escott-Price, V., Nalls, M. A., Morris, H. R., Lubbe, S., Brice, A., Gasser, T., et al. (2015). Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann. Neurol.* 77 (4), 582–591. doi:10.1002/ana.24335

Espay, A. J., Brundin, P., and Lang, A. E. (2017). Precision medicine for disease modification in Parkinson disease. *Nat. Rev. Neurol.* 13 (2), 119–126. doi:10.1038/nrneurol.2016.196

Finan, C., Gaulton, A., Kruger, F. A., Lumbers, R. T., Shah, T., Engmann, J., et al. (2017). The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* 9 (383), eaag1166. doi:10.1126/scitranslmed.aag1166

Foo, J. N., Tan, L. C., Irwan, I. D., Au, W. L., Low, H. Q., Prakash, K. M., et al. (2017). Genome-wide association study of Parkinson's disease in East Asians. *Hum. Mol. Genet.* 26 (1), 226–232. doi:10.1093/hmg/ddw379

Fröhlich, H., Bontridder, N., Petrovska-Delacréta, D., Glaab, E., Kluge, F., Yacoubi, M. E., et al. (2022). Leveraging the potential of digital technology for better individualized treatment of Parkinson's disease. *Front. Neurol.* 13, 788427. doi:10.3389/fneur.2022.788427

Gaunt, T. R., Rodríguez, S., and Day, I. N. (2007). Cubic exact solutions for the estimation of pairwise haplotype frequencies: Implications for linkage disequilibrium analyses and a web tool 'CubeX. *BMC Bioinforma.* 8, 428. doi:10.1186/1471-2105-8-428

Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A., and Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* 10 (1), 1776. doi:10.1038/s41467-019-09718-5

Hamza, T. H., Zabetian, C. P., Tenesa, A., Laederach, A., Montimurro, J., Yearout, D., et al. (2010). Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat. Genet.* 42 (9), 781–785. doi:10.1038/ng.642

Han, F., and Pan, W. (2010). A data-adaptive sum test for disease association with multiple common or rare variants. *Hum. Hered.* 70 (1), 42–54. doi:10.1159/000288704

Hassanin, E., May, P., Aldisi, R., Krawitz, P., Maj, C., and Bobbili, D. R. (2021)"Assessing the role of polygenic background on the penetrance of monogenic forms in Parkinson's disease. *MedRxiv* [Preprint]. Available at: https://doi.org/10.1101/2021.06.06.21253270 (Accessed September 10, 2022).

Hernandez, D. G., Nalls, M. A., Ylikotila, P., Keller, M., Hardy, J. A., Majamaa, K., et al. (2012). Genome wide assessment of young onset Parkinson's disease from Finland. *PLoS One* 7 (7), e41859. doi:10.1371/journal.pone.0041859

Hill-Burns, E. M., Ross, O. A., Wissemann, W. T., Soto-Ortolaza, A. I., Zareparsi, S., Siuda, J., et al. (2016). Identification of genetic modifiers of age-at-onset for familial Parkinson's disease. *Hum. Mol. Genet.* 25 (17), 3849–3862. doi:10.1093/hmg/ddw206

Hill-Burns, E. M., Wissemann, W. T., Hamza, T. H., Factor, S. A., Zabetian, C. P., and Payami, H. (2014). Identification of a novel Parkinson's disease locus via stratified genome-wide association study. *BMC Genomics* 15, 118. doi:10.1186/1471-2164-15-118

Ho, D., Schierding, W., Farrow, S. L., Cooper, A. A., Kempa-Liehr, A. W., and O'Sullivan, J. M. (2022). Machine learning identifies six genetic variants and alterations in the heart atrial appendage as key contributors to PD risk predictivity. *Front. Genet.* 12, 785436. doi:10.3389/fgene.2021.785436

Hoffmann, T. J., Marini, N. J., and Witte, J. S. (2010). Comprehensive approach to analyzing rare genetic variants. *PLoS One* 5 (11), e13584. doi:10.1371/journal.pone.0013584

Ibanez, L., Dube, U., Saef, B., Budde, J., Black, K., Medvedeva, A., et al. (2017). Parkinson disease polygenic risk score is associated with Parkinson disease status

and age at onset but not with alpha-synuclein cerebrospinal fluid levels. *BMC Neurol.* 17 (1), 198. doi:10.1186/s12883-017-0978-z

Ibanez, L., Farias, F. H. G., Dube, U., Mihindukulasuriya, K. A., and Harari, O. (2019). Polygenic risk scores in neurodegenerative diseases: A review. *Curr. Genet. Med. Rep.* 7, 22–29. doi:10.1007/s40142-019-0158-0

Jacobs, B. M., Belete, D., Bestwick, J., Blauwendraat, C., Bandres-Ciga, S., Heilbron, K., et al. (2020). Parkinson's disease determinants, prediction and gene-environment interactions in the UK Biobank. *J. Neurol. Neurosurg. Psychiatry* 91 (10), 1046–1054. doi:10.1136/jnnp-2020-323646

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* 51 (3), 404–413. doi:10.1038/s41588-018-0311-9

Jiménez-Jiménez, F. J., Alonso-Navarro, H., García-Martín, E., and Agúndez, J. A. (2016). NAT2 polymorphisms and risk for Parkinson's disease: A systematic review and meta-analysis. *Expert Opin. Drug Metab. Toxicol.* 12 (8), 937–946. doi:10.1080/17425255.2016.1192127

Kara, E., Xiromerisiou, G., Spanaki, C., Bozi, M., Koutsis, G., Panas, M., et al. (2014). Assessment of Parkinson's disease risk loci in Greece. *Neurobiol. Aging* 35 (2), e9–442. e16. doi:10.1016/j.neurobiolaging.2013.07.011

Khanna, S., Domingo-Fernández, D., Iyappan, A., Emon, M. A., Hofmann-Apitius, M., and Fröhlich, H. (2018). Using multi-scale genetic, neuroimaging and clinical data for predicting alzheimer's disease and reconstruction of relevant biological mechanisms. *Sci. Rep.* 8 (1), 11173. doi:10.1038/s41598-018-29433-3

Krauthammer, M., Kaufmann, C. A., Gilliam, T. C., and Rzhetsky, A. (2004). Molecular triangulation: Bridging linkage and molecular-network information for identifying candidate genes in alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* 101 (42), 15148–15153. doi:10.1073/pnas.0404315101

Lambert, S. A., Gil, L., Jupp, S., Ritchie, S. C., Xu, Y., Buniello, A., et al. (2021). The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat. Genet.* 53 (4), 420–425. doi:10.1038/s41588-021-00783-5

Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., and Davey Smith, G. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* 27 (8), 1133–1163. doi:10.1002/sim.3034

Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., et al. (2012b). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.* 91 (2), 224–237. doi:10.1016/j.ajhg.2012.06.007

Lee, S., Wu, M. C., and Lin, X. (2012a). Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 13 (4), 762–775. doi:10.1093/biostatistics/kxs014

Lewis, C. M., and Vassos, E. (2020). Polygenic risk scores: From research tools to clinical instruments. *Genome Med.* 12 (1), 44. doi:10.1186/s13073-020-00742-5

Li, B., and Leal, S. M. (2008). Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. *Am. J. Hum. Genet.* 83 (3), 311–321. doi:10.1016/j.ajhg.2008.06.024

Li, C., Ou, R., Chen, Y., Gu, X., Wei, Q., Cao, B., et al. (2020). Mutation analysis of DNAJC family for early-onset Parkinson's disease in a Chinese cohort. *Mov. Disord.* 35 (11), 2068–2076. doi:10.1002/mds.28203

Li, C., Ou, R., Chen, Y., Gu, X., Wei, Q., Cao, B., et al. (2021). Genetic modifiers of age at onset for Parkinson's disease in asians: A genome-wide association study. *Mov. Disord.* 36 (9), 2077–2084. doi:10.1002/mds.28621

Li, Y., and Patra, J. C. (2010). Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network. *Bioinformatics* 26 (9), 1219–1224. doi:10.1093/bioinformatics/btq108

Lill, C. M., Roehr, J. T., McQueen, M. B., Kavvoura, F. K., Bagade, S., Schjeide, B. M., et al. (2012). Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS Genet.* 8 (3), e1002548. doi:10.1371/journal.pgen.1002548

Lin, D. Y., and Tang, Z. Z. (2011). A general framework for detecting disease associations with rare variants in sequencing studies. *Am. J. Hum. Genet.* 89 (3), 354–367. doi:10.1016/j.ajhg.2011.07.015

Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S. (2020). Explainable AI: A review of machine learning interpretability methods. *Entropy (Basel)* 23 (1), 18. doi:10.3390/e23010018

Liu, D. J., and Leal, S. M. (2010). A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. *PLoS Genet.* 6 (10), e1001156. doi:10.1371/journal.pgen.1001156

Liu, G., Peng, J., Liao, Z., Locascio, J. J., Corvol, J. C., Zhu, F., et al. (2021). Genome-wide survival study identifies a novel synaptic locus and polygenic score

for cognitive progression in Parkinson's disease. *Nat. Genet.* 53 (6), 787–793. doi:10.1038/s41588-021-00847-6

Liu, X., Cheng, R., Verbitsky, M., Kisselev, S., Browne, A., Mejia-Sanatana, H., et al. (2011). Genome-wide association study identifies candidate genes for Parkinson's disease in an Ashkenazi Jewish population. *BMC Med. Genet.* 12, 104. doi:10.1186/1471-2350-12-104

Loesch, D. P., Horimoto, A. R. V. R., Heilbron, K., Sarihan, E. I., Inca-Martinez, M., Mason, E., et al. (2021). Characterizing the genetic architecture of Parkinson's disease in latinos. *Ann. Neurol.* 90 (3), 353–365. doi:10.1002/ana.26153

Madsen, B. E., and Browning, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* 5 (2), e1000384. doi:10.1371/journal.pgen.1000384

Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X., and Sham, P. C. (2017). Polygenic scores via penalized regression on summary statistics. *Genet. Epidemiol.* 41 (6), 469–480. doi:10.1002/gepi.22050

Makarious, M. B., Leonard, H. L., Vitale, D., Iwaki, H., Sargent, L., Dadu, A., et al. (2022). Multi-modality machine learning predicting Parkinson's disease. *NPJ Park. Dis.* 8 (1), 35. doi:10.1038/s41531-022-00288-w

Markopoulou, K., Chase, B. A., Premkumar, A. P., Schoneburg, B., Kartha, N., Wei, J., et al. (2021). Variable effects of PD-risk associated SNPs and variants in parkinsonism-associated genes on disease phenotype in a community-based cohort. *Front. Neurol.* 12, 662278. doi:10.3389/fneur.2021.662278

Mata, I. F., Johnson, C. O., Leverenz, J. B., Weintraub, D., Trojanowski, J. Q., Van Deerlin, V. M., et al. (2017). Large-scale exploratory genetic analysis of cognitive impairment in Parkinson's disease. *Neurobiol. Aging* 56, 211e1–211. e7. doi:10.1016/j.neurobiolaging.2017.04.009

Mata, I. F., Shi, M., Agarwal, P., Chung, K. A., Edwards, K. L., Factor, S. A., et al. (2010). SNCA variant associated with Parkinson disease and plasma alpha-synuclein level. *Arch. Neurol.* 67 (11), 1350–1356. doi:10.1001/archneurol.2010.279

Morgenthaler, S., and Thilly, W. G. (2007). A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: A cohort allelic sums test (CAST). *Mutat. Res.* 615 (1-2), 28–56. doi:10.1016/j.mrfmmm.2006.09.003

Morris, A. P., and Zeggini, E. (2010). An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet. Epidemiol.* 34 (2), 188–193. doi:10.1002/gepi.20450

Nalls, M. A., Blauwendraat, C., Vallerga, C. L., Heilbron, K., Bandres-Ciga, S., Chang, D., et al. (2019). Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. *Lancet. Neurol.* 18 (12), 1091–1102. doi:10.1016/S1474-4422(19)30320-5

Nalls, M. A., Pankratz, N., Lill, C. M., Do, C. B., Hernandez, D. G., Saad, M., et al. (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* 46 (9), 989–993. doi:10.1038/ng.3043

Neale, B. M., Rivas, M. A., Voight, B. F., Altshuler, D., Devlin, B., Orho-Melander, M., et al. (2011). Testing for an unusual distribution of rare variants. *PLoS Genet.* 7 (3), e1001322. doi:10.1371/journal.pgen.1001322

Newcombe, P. J., Nelson, C. P., Samani, N. J., and Dudbridge, F. (2019). A flexible and parallelizable approach to genome-wide polygenic risk scores. *Genet. Epidemiol.* 43 (7), 730–741. doi:10.1002/gepi.22245

Pan, W. (2009). Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet. Epidemiol.* 33 (6), 497–507. doi:10.1002/gepi.20402

Pankratz, N., Beecham, G. W., DeStefano, A. L., Dawson, T. M., Doheny, K. F., Factor, S. A., et al. (2012). Meta-analysis of Parkinson's disease: Identification of a novel locus, RIT2. *Ann. Neurol.* 71 (3), 370–384. doi:10.1002/ana.22687

Park, K. W., Jo, S., Kim, M. S., Jeon, S. R., Ryu, H. S., Kim, J., et al. (2021). Genomic association study for cognitive impairment in Parkinson's disease. *Front. Neurol.* 11, 579268. doi:10.3389/fneur.2020.579268

Paul, K. C., Schulz, J., Bronstein, J. M., Lill, C. M., and Ritz, B. R. (2018). Association of polygenic risk score with cognitive decline and motor progression in Parkinson disease. *JAMA Neurol.* 75 (3), 360–366. doi:10.1001/jamaneurol.2017.4206

Peng, J., Guan, J., and Shang, X. (2019). Predicting Parkinson's disease genes based on Node2vec and autoencoder. *Front. Genet.* 10, 226. doi:10.3389/fgene.2019.00226

Pihlstrøm, L., Morset, K. R., Grimstad, E., Vitelli, V., and Toft, M. (2016). A cumulative genetic risk score predicts progression in Parkinson's disease. *Mov. Disord.* 31 (4), 487–490. doi:10.1002/mds.26505

Planas-Ballvé, A., and Vilas, D. (2021). Cognitive impairment in genetic Parkinson's disease. *Park. Dis.* 2021, 8610285. doi:10.1155/2021/8610285

Poewe, W., Seppi, K., Tanner, C. M., Halliday, G. M., Brundin, P., Volkmann, J., et al. (2017). Parkinson disease. *Nat. Rev. Dis. Prim.* 3, 17013. doi:10.1038/nrdp.2017.13

Prashanth, R., Dutta Roy, S., Mandal, P. K., and Ghosh, S. (2016). High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. *Int. J. Med. Inf.* 90, 13–21. doi:10.1016/j.ijmedinf.2016.03.001

Price, A. L., Kryukov, G. V., de Bakker, P. I., Purcell, S. M., Staples, J., Wei, L. J., et al. (2010). Pooled association tests for rare variants in exon-resequencing studies. *Am. J. Hum. Genet.* 86 (6), 832–838. doi:10.1016/j.ajhg.2010.04.005

Pritchard, J. K., and Cox, N. J. (2002). The allelic architecture of human disease genes: Common disease-common variant.or not? *Hum. Mol. Genet.* 11 (20), 2417–2423. doi:10.1093/hmg/11.20.2417

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. doi:10.1086/519795

Ranstam, J. (2016). Multiple P-values and Bonferroni correction. *Osteoarthr. Cartil.* 24 (5), 763–764. doi:10.1016/j.joca.2016.01.008

Rodrigo, L. M., and Nyholt, D. R. (2021). Imputation and reanalysis of ExomeChip data identifies novel, conditional and joint genetic effects on Parkinson's disease risk. *Genes (Basel)* 12 (5), 689. doi:10.3390/genes12050689

Ryu, H. S., Park, K. W., Choi, N., Kim, J., Park, Y. M., Jo, S., et al. (2020). Genomic analysis identifies new loci associated with motor complications in Parkinson's disease. *Front. Neurol.* 11, 570. doi:10.3389/fneur.2020.00570

Saad, M., Lesage, S., Saint-Pierre, A., Corvol, J. C., Zelenika, D., Lambert, J. C., et al. (2011). Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. *Hum. Mol. Genet.* 20 (3), 615–627. doi:10.1093/hmg/ddq497

Satake, W., Nakabayashi, Y., Mizuta, I., Hirota, Y., Ito, C., Kubo, M., et al. (2009). Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat. Genet.* 41 (12), 1303–1307. doi:10.1038/ng.485

Schork, N. J., Murray, S. S., Frazer, K. A., and Topol, E. J. (2009). Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* 19 (3), 212–219. doi:10.1016/j.gde.2009.04.010

Sia, M. W., Foo, J. N., Saffari, S. E., Wong, A. S., Khor, C. C., Yuan, J. M., et al. (2021). Polygenic risk scores in a prospective Parkinson's disease cohort. *Mov. Disord.* 36 (12), 2936–2940. doi:10.1002/mds.28761

Siitonen, A., Nalls, M. A., Hernandez, D. G., Gibbs, J. R., Ding, J., Ylikotila, P., et al. (2017). Genetics of early-onset Parkinson's disease in Finland: Exome sequencing and genome-wide association study. *Neurobiol. Aging* 53, e7. e7-195e10. doi:10.1016/j.neurobiolaging.2017.01.019

Simon, K. C., Eberly, S., Gao, X., Oakes, D., Tanner, C. M., Shoulson, I., et al. Parkinson Study Group (2014). Mendelian randomization of serum urate and Parkinson disease progression. *Ann. Neurol.* 76 (6), 862–868. doi:10.1002/ana.24281

Simón-Sánchez, J., Schulte, C., Bras, J. M., Sharma, M., Gibbs, J. R., Berg, D., et al. (2009). Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.* 41 (12), 1308–1312. doi:10.1038/ng.487

Singh, N. K., Banerjee, B. D., Bala, K., Chhillar, M., and Chhillar, N. (2014). Gene-gene and gene-environment interaction on the risk of Parkinson's disease. *Curr. Aging Sci.* 7 (2), 101–109. doi:10.2174/1874609807666140805123621

Slager, S. L., and Schaid, D. J. (2001). Case-control studies of genetic markers: Power and sample size approximations for armitage's test for trend. *Hum. Hered.* 52 (3), 149–153. doi:10.1159/000053370

Spataro, N., Calafell, F., Cervera-Carles, L., Casals, F., Pagonabarraga, J., Pascual-Sedano, B., et al. (2015). Mendelian genes for Parkinson's disease contribute to the sporadic forms of the disease. *Hum. Mol. Genet.* 24 (7), 2023–2034. doi:10.1093/hmg/ddu616

Spencer, C. C., Plagnol, V., Strange, A., Gardner, M., Paisan-Ruiz, C., Band, G., et al. (2011). Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21. *Hum. Mol. Genet.* 20 (2), 345–353. doi:10.1093/hmg/ddq469

Storm, C. S., Kia, D. A., Almramhi, M. M., Bandres-Ciga, S., Finan, C., Hingorani, A. D., et al. (2021). Finding genetically-supported drug targets for Parkinson's disease using Mendelian randomization of the druggable genome. *Nat. Commun.* 12 (1), 7342. doi:10.1038/s41467-021-26280-1

Sun, X., Namkung, J., Zhu, X., and Elston, R. C. (2011). Capability of common SNPs to tag rare variants. *BMC Proc.* 5 (9), S88. doi:10.1186/1753-6561-5-S9-S88

Tan, M. M. X., Lawton, M. A., Jabbari, E., Reynolds, R. H., Iwaki, H., Blauwendraat, C., et al. (2021). Genome-wide association studies of cognitive and motor progression in Parkinson's disease. *Mov. Disord.* 36 (2), 424–433. doi:10.1002/mds.28342

Taylor, A. K., Zambaux, J. L., Klisak, I., Mohandas, T., Sparkes, R. S., Schotz, M. C., et al. (1991). Carboxyl ester lipase: A highly polymorphic locus on human chromosome 9qter. *Genomics* 10 (2), 425–431. doi:10.1016/0888-7543(91)90328-c

Tirozzi, A., Izzi, B., Noro, F., Marotta, A., Gianfagna, F., Hoylaerts, M. F., et al. (2020). Assessing genetic overlap between platelet parameters and neurodegenerative disorders. *Front. Immunol.* 11, 02127. doi:10.3389/fimmu.2020.02127

Titova, N., and Chaudhuri, K. R. (2017). Personalized medicine in Parkinson's disease: Time to be precise. *Mov. Disord.* 32 (8), 1147–1154. doi:10.1002/mds.27027

Tran, J., Anastacio, H., and Bardy, C. (2020). Genetic predispositions of Parkinson's disease revealed in patient-derived brain cells. *NPJ Park. Dis.* 6, 8. doi:10.1038/s41531-020-0110-8

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., et al. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* 97 (4), 576–592. doi:10.1016/j.ajhg.2015.09.001

Wang, M., Ge, W., Apthorp, D., and Suominen, H. (2020). Robust feature engineering for Parkinson disease diagnosis: New machine learning techniques. *JMIR Biomed. Eng.* 5 (1), e13611. doi:10.2196/13611

Wu, M. C., Kraft, P., Epstein, M. P., Taylor, D. M., Chanock, S. J., Hunter, D. J., et al. (2010). Powerful SNP-set analysis for case-control genome-wide association studies. *Am. J. Hum. Genet.* 86 (6), 929–942. doi:10.1016/j.ajhg.2010.05.002

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* 89 (1), 82–93. doi:10.1016/j.ajhg.2011.05.029

Xiong, M., Zhao, J., and Boerwinkle, E. (2002). Generalized T2 test for genome association studies. *Am. J. Hum. Genet.* 70 (5), 1257–1268. doi:10.1086/340392

Zawistowski, M., Gopalakrishnan, S., Ding, J., Li, Y., Grimm, S., and Zöllner, S. (2010). Extending rare-variant testing strategies: Analysis of noncoding sequence and imputed genotypes. *Am. J. Hum. Genet.* 87 (5), 604–617. doi:10.1016/j.ajhg.2010.10.012

Zheng, J., Erzurumluoglu, A. M., Elsworth, B. L., Kemp, J. P., Howe, L., Haycock, P. C., et al. (2017). LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 33 (2), 272–279. doi:10.1093/bioinformatics/btw613

Check for updates

Insight into genomic organization of pathogenic coronaviruses, SARS-CoV-2: Implication for emergence of new variants, laboratory diagnosis and treatment options

Fikru B. Bedada[1]*, Gezahegn Gorfu[1,2], Shaolei Teng[3] and Marguerite E. Neita[1]

[1]Department of Clinical Laboratory Science, College of Nursing and Allied Health Sciences, Howard University, Washington, DC, United States, [2]Department of Pathology, College of Medicine, Howard University, Washington, DC, United States, [3]Department of Biology, College of Arts and Sciences, Howard University, Washington, DC, United States

SARS-CoV-2 is a novel zoonotic positive-sense RNA virus (ssRNA+) belonging to the genus beta coronaviruses (CoVs) in the Coronaviridae family. It is the causative agent for the outbreak of the disease, COVID-19. It is the third CoV causing pneumonia around the world in the past 2 decades. To date, it has caused significant deaths worldwide. Notably, the emergence of new genetic variants conferring efficient transmission and immune evasion remained a challenge, despite the reduction in the number of death cases, owing to effective vaccination regimen (boosting) and safety protocols. Thus, information harnessed from SARS-CoV-2 genomic organization is indispensable for seeking laboratory diagnosis and treatment options. Here in, we review previously circulating variants of SARS-CoV-2 designated variant of concern (VOC) including the Alpha (United Kingdom), Beta (South Africa), Gamma (Brazil), Delta (India), and recently circulating VOC, Omicron (South Africa) and its divergent subvariants (BA.1, BA.2, BA.3, BA.2.12.1, BA.4 and BA.5) with BA.5 currently becoming dominant and prolonging the COVID pandemic. In addition, we address the role of computational models for mutagenesis analysis which can predict important residues that contribute to transmissibility, virulence, immune evasion, and molecular detections of SARS-CoV-2. Concomitantly, the importance of harnessing the immunobiology of SARS-CoV-2 and host interaction for therapeutic purpose; and use of an in slilico based biocomputational approaches to achieve this purpose *via* predicting novel therapeutic agents targeting PRR such as toll like receptor, design of universal vaccine and chimeric antibodies tailored to the emergent variant have been highlighted.

KEYWORDS

SARS-CoV-2, new genetic variants, genome organization and evolution, molecular diagnosis, ACE2 receptors and COVID-19, vaccines and antiviral drugs, omicron (B.1.1.529), computational model

Introduction

The disease COVID-19 came to prominence in December 2019 after many cases of pneumonia with unknown etiology appeared in Wuhan, China (Chen et al., 2020a; Huang et al., 2020a). Ultimately, the disease quickly spread to other provinces of China and the rest of the world (Chen et al., 2020a; Huang et al., 2020a). Due to its continuing spread, the WHO declared the outbreak as a public health emergency of international concern (Zarocostas, 2020). Subsequently, the virus was renamed as severe acute respiratory syndrome coronaviruse-2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020). To date, it has infected 604, 558, 779 million cases and killed 6,483,819 people worldwide (https://www.worldometers.info/coronavirus/).

Coronaviruses (CoVs) are large, enveloped, single-stranded positive-sense RNA viruses belonging to the coronaviridae family (Chan et al., 2013; Ge et al., 2020). They have four genera namely alpha, beta, delta, and gamma CoVs (Hozhabri et al., 2020). Mammals are mainly infected by alpha and beta CoVs, and the gamma and delta CoVs primarily infect birds (Chen et al., 2020b; Ge et al., 2020; Wu et al., 2020). So far, seven CoVs had been identified causing disease in humans (Wu et al., 2020). The four human CoVs such as HCoV 229E, NL63, OC43, and HKU1 display endemic distribution globally resulting in mild upper respiratory tract infections and collectively are associated with 10%–30% cases of the common cold (Hozhabri et al., 2020; Wu et al., 2020). Conversely, SARS-CoV-2, MERS-CoV and SARS-CoV are the three CoVs that caused the most severe type of illness resulting in lower respiratory tract infections, acute respiratory distress syndrome (ARDS) and death in humans (Chen et al., 2020b; Paules et al., 2020). Accordingly, the case fatality rate of SARS-CoV is 10% whereas that of MERS-CoV is 37% (Huang et al., 2020b; Petersen et al., 2020). For SARS-CoV-2, several factors could affect the case fatality rate, making it a moving target. For instance, age variation, existence of comorbidity, capacity of health care systems, vaccinations, boosting, safety protocols are factors among others (Alimohamadi et al., 2021). Based on current estimates from a systematic review and meta-analysis, the overall estimated pooled case fatality rate of SARS-CoV-2 was 1.0% among the general population and is 19% in patients older than 50 years (Alimohamadi et al., 2021), making SARS-CoV-2 less deadly but more highly transmissible than its close relatives, SARS-CoV and MERS-CoV.

This review is centered on the SARS-CoV-2 genome structural organization, emerging genetic alterations that give rise to VOC, and features that distinguish them in terms of transmissibility, virulence, immune evasion, and the implications of this for diagnostic and therapeutic approaches.

Origin and evolution of SARS-CoV-2, MERS-CoV, and SARS-CoV

Considering their origin and evolution, all CoVs initially existed in diverse species of bats as CoV-related viruses for example, SARS related CoV, MERS related COV and SARS related CoV2 (Ashour et al., 2020; Hozhabri et al., 2020; Lu et al., 2015). Subsequently, sequential mutations and recombination events allowed them to adapt to intermediate hosts for example, civets, camels, and pangolins, ultimately gaining access to human beings (Ashour et al., 2020; Hozhabri et al., 2020). For instance, MERS-CoV originated from bats and later adapted to dromedary camels as an intermediate host (Corman et al., 2014; Fung and Liu, 2019). Similarly, SARS-CoV adapted to civet and raccoon dogs as intermediate hosts. Thus, host jump (spillover) has contributed to the diversity and potential spillover of bat borne CoVs, where such zoonotic pathogens can cross species barriers to ultimately infect humans (Cui et al., 2019; Lu et al., 2015). In the case of SARS-CoV-2, given the challenge of identifying the origin and intermediate host as well as demand for several years of extensive work, utilizing the experience learned from the influenza virus is important (Gao et al., 2013). Moreover, collection of conclusive data in sufficient numbers in terms of genetic similarities and relevant geographical location provides better information (Wang et al., 2021a). In the end, a global search for sarbecoviruses by honing research on *Rhinolophus* bats, pangolins, and minks could be an effective strategy to identify intermediate hosts for SARS-CoV2 (Wang et al., 2021a; Zhao et al., 2020). When one considers that known coronaviruses are zoonotic viruses, these interspecies transmissions in CoVs can lead to evolution of another related novel CoVs (nCoVs) by jumping a natural host *via* a similar mechanism (spillover). Thus, although such undertakings require several years of continuous research, the accumulated data will provide a huge benefit for future origin-tracing endeavors (Gao et al., 2013; Wang et al., 2021a).

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

As noted, the SARS-CoV-2 is a novel zoonotic positive-sense RNA virus belonging to the genus beta coronaviruses in the Coronaviridae family (Chen et al., 2020c; Lu et al., 2020; Paraskevis et al., 2020). The SARS-CoV-2 has led to the outbreak of the disease COVID-19 and is the third CoVs causing pneumonia globally in the past 20 years (Ge et al., 2020). Clearly, the world has witnessed the devastating consequences of SARS-CoV-2 that has resulted in a global pandemic with unprecedented challenges and significant deaths worldwide. Central to the challenge posed by SARS-CoV-2 is its high transmission rate compared to its relatives

SARS-CoV and MERS-CoV (Huang et al., 2020b; Petersen et al., 2020). However, SARS-CoV-2 is less deadly compared with the fatality rate of related species (Huang et al., 2020b; Petersen et al., 2020). Given the broad clinical spectrum and high transmission rate, and mutability of its genome leading to genetic alterations (VOC), eradicating, and managing the spread of SARS-CoV-2 pandemic will remain challenging in the foreseeable future (Petersen et al., 2020). As noted, information gained from genomic organization is indispensable as new variants caused by genetic alterations (mutations) are known to occur in viral infections (Grubaugh et al., 2020; Lauring and Hodcroft, 2021). Certainly, this is evident by the emergence of mutations leading to several new variants of SARS-CoV-2 (VOC) with the ability to spread worldwide. In addition to their complex genetic description reflecting genetic alteration, these emergent new variants are known by less complex WHO naming as Alpha, Beta, Gamma, and Delta which are previously circulating variants of concern (VOCs) (Figure 3), and Omicron variants, which is the current VOC (Figure 4). Thus, understanding the molecular signature of SARS-CoV-2 is important for laboratory identification, genetic surveillance, assessment of virulence, transmissibility, vaccine effectiveness and development of vaccines/drugs (combinatorial therapy).

SARS-CoV-2 genome organization and expression

The genome structure of SARS-CoV-2 comprises single-stranded positive sense RNA (ssRNA+) genome that is around 29.8 kb, providing a trove of information on its genome sequences (Hussain et al., 2020) and Figure 1. This genetic information is central to understanding SARS-CoV-2 evolution, pathogenesis, and the emergence of new variants which occur *via* mutation mediated genetic alterations (Figures 3, 4). The genetic information is particularly important for identification and monitoring of resistant variants and the development of new vaccines tailored to variant epitope, therapeutic monoclonal antibodies (designing chimeric antibodies), and antiviral drugs (combinatorial therapy). In addition, the information gained from the genomic structure is crucial for the molecular diagnostic-based detections. For example, current real time RT-PCR nucleic acid detection is based on the information extracted from specific regions targeting the conserved regions of S, E, N, nsp12, nsp14 and ORF1ab in the genome of novel SARS-CoV-2, which are key for the design of specific primers employed for accurate and specific detection. As shown in Figure 1, the SARS-CoV-2 genome is schematically represented in the pattern of 5′-UTR-ORF1a (yellow), ORF1b (blue), structural proteins (S, E, M, N, orange), accessory genes scattered among structural genes (gray) and 3′-UTR. Typically, the genome of SARS-CoV-2 entails two untranslated regions (UTRs) as 5′-cap structure

and 3′-poly-A tail, which are located at two extreme ends (Hozhabri et al., 2020). These two UTRs flank a single, large overlapping open reading frame (ORF) encoding a nonstructural polyprotein (Chan et al., 2020) followed by genes coding for four structural proteins including Spike (S), Envelope (E), Membrane (M), Nucleocapsid (N) (Figure 2), and accessory genes coding for accessory proteins, some of which are interspersed in the genes of structural proteins and even overlapping with structural genes (Chan et al., 2020; Ge et al., 2020; Hozhabri et al., 2020; Michel et al., 2020; Yoshimoto, 2020).

SARS-CoV-2 ORF1a and ORF1b genes

The ORF1a (yellow) and ORF1b (blue) are two large genes that are translated into polypeptides 1a (PP1a) (yellow) which is short; and PP1ab (yellow/blue) which is long nonstructural proteins (NSP) (Chan et al., 2020; Ge et al., 2020; Hozhabri et al., 2020; Michel et al., 2020; Yoshimoto, 2020) (Figure 1). Of note, the first ORFs of SARS-CoV-2 (ORF1a and ORF1b) comprise large portion (2/3) of the genome length and encode for 16 nonstructural polyproteins (NSPs 1–16), which are directly translated from genomic RNA (Cui et al., 2019). In this regard, it is a -1 ribosomal frameshift located between ORF1a and ORF1b that is responsible for the production of these two large replicase polypeptides (PP1a and PP1ab), key proteins that are important for viral replication (Hozhabri et al., 2020; Michel et al., 2020). Functionally, two virally encoded cysteine proteases, papain-like protease (PLpro) and 3-chymotrypsin-like protease (3CLpro) are key players for further processing of PPs into 16 NSPs (NSPs 1–16) (Baez-Santos et al., 2015; Chen et al., 2020b; Romano et al., 2020). For instance, the NSP14 of CoVs functions as proofreading machinery for prevention of lethal mutagenesis using its N-terminal exoribonuclease (ExoN) domain, and the C-terminal domain as a (guanine-N7) methyl transferase (N7-MTase) for mRNA capping (Ma et al., 2015; Ogando et al., 2020). Given the importance of these proteases and NSP14 (ExoN) for viral replication and transcription which ensures viral propagation, they can be targeted to develop potential broad-spectrum antiviral drugs (pan antiviral drugs) (Ma et al., 2015; Ogando et al., 2020) (Figure 1, green box) and for detection purposes (reviewed in molecular detection section). For example, Remdesivir inhibits RNA-dependent RNA polymerase (RdRp) of CoVs including SARS-CoV-2 by acting as a nucleoside analog; and is FDA-approved for the treatment of COVID-19 patients (Kokic et al., 2021). Further, all CoVs, including SARS-CoV-2, encode two proteases required for PP1A and PP1AB polyproteins processing (Figure 1 yellow and blue) (Reina and Iglesias, 2022). In this context, the key protease chymotrypsin-like (3CL) catalyze the processing of NSP11/16 proteins and is targeted for the development of SARS-CoV-2 antiviral drugs due to the high sequence and structure conservation among all CoVs (Reina and Iglesias, 2022). One such example of an antiviral
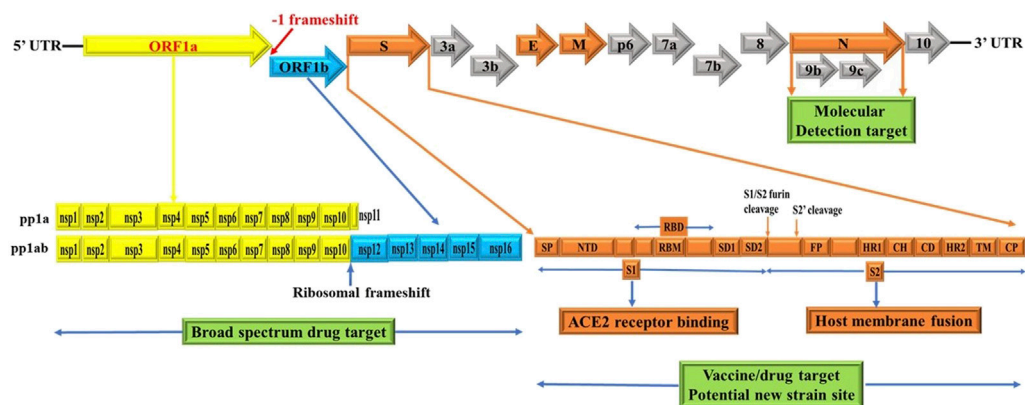
FIGURE 1
Schematics of genome structures of SARS-CoV-2. These schematics are based on information extracted from various studies (Baez-Santos et al., 2015; Berger and Schaffitzel, 2020; Cui et al., 2019; Chen et al., 2020b; Chen et al., 2020c; Chan et al., 2020; Ge et al., 2020; Hozhabri et al., 2020; Johnson et al., 2020; Michel et al., 2020; Peacock et al., 2021; Rabaan et al., 2020; Romano et al., 2020; Wu et al., 2020; Yoshimoto, 2020). Accordingly, the SARS-CoV-2 genome contains two untranslated regions (UTRs), one at 5′-cap structure and the second at the 3′-poly-A tail (Hozhabri et al., 2020). These two UTRs flank a single overlapping open reading frame (ORF) encoding a nonstructural polyprotein (Chan et al., 2020) followed by genes coding for four structural proteins, Spike (S), Envelope (E), Membrane (M), Nucleocapsid (N) and; sub group specific accessory genes coding for several accessory proteins, some of which are interspersed in the genes of structural proteins and even overlapping with structural genes (Chan et al., 2020; Ge et al., 2020; Hozhabri et al., 2020; Michel et al., 2020; Yoshimoto, 2020). SARS-CoV-2 spike glycoprotein domains (S domains) are shown in shades of orange scheme zoomed to magnify the detail. Accordingly, S1 comprises a signal sequence or peptide (SP), NTD, RBD, SD1 and SD2. S2 comprises a second furin cleavage site (S2′) upstream of the fusion peptide (FP), HR1, CH, CD, HR2, TM domain and cytoplasmic C-terminus peptide (CP) (Berger and Schaffitzel, 2020). Green box indicates targets for potential broad spectrum antiviral drugs, molecular detection targets, specific vaccines/drugs targets and genetic alteration hotspots (new variants sites). UTR: untranslated region, ORF: open reading frame, S: spike, E: envelope, M: membrane, N: nucleocapsid, NTD: N-terminal domain, RBD: receptor-binding domain, SD1: subdomain 1, SD2: subdomain 2, S2′: second protease (furin) cleavage site, FP: fusion peptide, HR1: heptad repeat 1, CH: central helix, CD: connector domain, HR2: heptad repeat 2, TM: transmembrane domain and CP: Cytoplasmic C-terminus peptide. Note: Furin cleavage sites are marked with brown arrow and green box show molecular detection target, vaccine/drug target and genetic alteration (new variant) site. The length of genes and proteins are not drawn to the scale. See text for additional information.
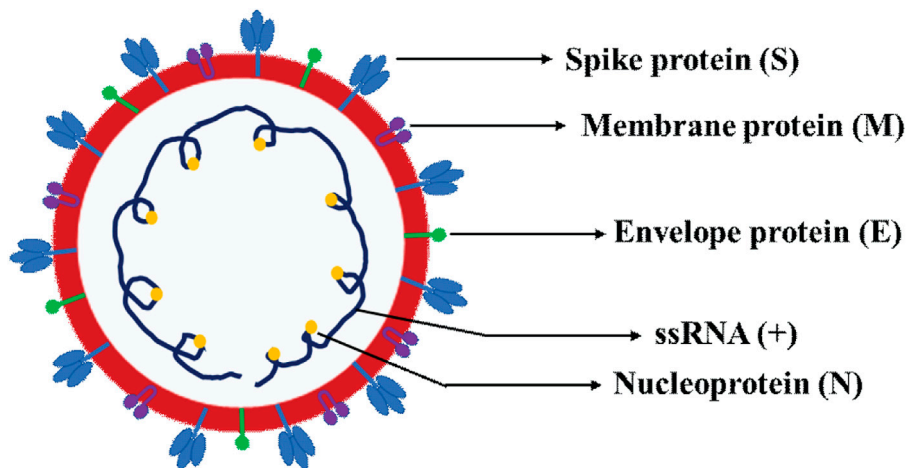
FIGURE 2
Schematic representation of SARS-CoV-2 virion structure. The schematics depicts the various four canonical structural proteins, namely the spike protein (S), envelope protein (E), membrane protein (M) and nucleoprotein (N) and single-stranded positive sense RNA (ssRNA+).

drug targeting a key viral replication machinery (protease) is Paxlovid (nirmatrelvir and ritonavir) from Pfizer (Reina and Iglesias, 2022) which has received emergency use authorization (EUA) by FDA. Similarly, Molnupiravir from Merck increases the frequency of viral RNA mutations by targeting viral RNA-dependent RNA polymerase (RdRp) and impairs SARS-CoV-2 replication (Kabinger et al., 2021) and has received EUA by FDA.

Significance of -1 ribosomal frameshift

As in dictated in Figure 1, the synthesis of PP1ab involves a -1 ribosomal frameshift that occurs in the overlapping region between ORFs1a and ORFs1b. The -1 ribosomal frameshift occurs during the translational elongation step which is required for the expression of protein from CoVs gene1, the largest gene accounting for 2/3 of the CoVs genome (Bekaert and Rousset, 2005; Nakagawa et al., 2016). Thus, PP1ab is encoded by fusion from the two ORFs (ORF1a and ORF1b (Chan et al., 2020; Ge et al., 2020; Hozhabri et al., 2020; Michel et al., 2020; Yoshimoto, 2020). Functionally, most of the mature NSPs produced from these two large PPs are important for synthesizing CoVs RNA (Bekaert and Rousset, 2005; Nakagawa et al., 2016). Thus, -1 ribosomal frameshifting dependent translation of PP1ab is an important step for replication of COVs (Bekaert and Rousset, 2005; Nakagawa et al., 2016). Consistent with this, reduced frameshifting efficiency that affects ratio of PP1a/PP1ab is reported to impact virus infectivity and replication (Ishimaru et al., 2013). As CoVs have evolved to produce optimal levels of -1 ribosomal frameshift for efficient replication (Bekaert and Rousset, 2005; Ishimaru et al., 2013; Nakagawa et al., 2016), it is an evolutionally conserved features of CoVs including SARS-CoV-2, SARS-CoV and MERS-CoV.

SARS-CoV-2 canonical structural genes

Of particular note, the subgenomic RNAs is translated to give rise to all structural and accessory proteins during the transcription/replication of the CoVs genome (Fung and Liu, 2019; Kim et al., 2020). This contrasts with SARS-CoV-2 ORF1a and ORF1b proteins which are directly translated from the RNA genome (Cui et al., 2019). Accordingly, the 3′-end of SARS-CoV-2 genomes contain four canonical structural proteins (orange), namely the S, E, M and N proteins, as common entity to all CoVs (Chen et al., 2020b) and (Figure 2). Of note, the M protein is higher in quantity than any other protein in the virus particle (Masters, 2006; Nal et al., 2005). Functionally, the M protein shapes the virions, promotes membrane curvature, and binds to the nucleocapsid using its three transmembrane domains (Masters, 2006; Nal et al., 2005). Conversely, the N protein binds to NSP3 protein using its two domains which help

tether the genome to replication/transcription complex and allow viral RNA packaging into viral particle during viral assembly (Masters, 2006; Hurst et al., 2009). For molecular diagnostics, it is imperative to have accurate genetic information regarding the target gene in the SARS-CoV-2 genome. Thus, the N protein has significance pertaining to diagnosis. For example, the N gene encoding for N protein has been targeted and is in use to provide much needed information for the design of SARS-CoV-2 specific primers (https://www.raybiotech.com/coronavirus-nucleic-acid-detection-kit/\) and (Figure 1, green box). Further, the N protein is abundantly expressed during infections and has high immunogenic activity. While the E protein is key player for the assembly of the virus and release of the virion from host cells, the S protein has cardinal role in the attachment to host cell receptors, viral entry and determines host tropism (Du et al., 2009; Siu et al., 2008) and Figure 6. Significantly, the S protein is the target for vaccine/drug development, and virtually all emergent new variants have their genetic alterations (mutations) in the spike protein, making it an indispensable part of the SARS-CoV-2 genome for clinical, molecular detection, therapeutic intervention, epidemiological and research purposes.

SARS-CoV-2 spike genes

Apparently, the S glycoprotein is of particular interest because it is a target for vaccine/drug design, development, and production as well as molecular detection. Interestingly, SARS-CoV-2 has a high transmission rate and is prone to genetic alteration at the S gene site coding for S glycoprotein, despite the similarity between SARS-CoV and MERS-CoV (Rabaan et al., 2020; Davies et al., 2021). In this context, the SARS-CoV-2 utilizes its S proteins to bind to angiotensin-converting enzyme 2 (ACE2) and infect human cells (Chen et al., 2020d; Ge et al., 2020; Huang et al., 2020b) and Figure 6. Of note, the S protein of SARS-CoV-2 contains two regions, S1 subunit and S2 subunit, having a size of 180–200 kd (Huang et al., 2020b; Ge et al., 2020) and (Figure 1, orange zoomed to show detail). As shown in Figure 6 depicting host virus interaction, there are two furin (protease) cleavage sites for S protein, one located at the S1/S2 interface and the other at the S2′ position of the S protein, and these are key to viral replication and pathogenesis (Johnson et al., 2020; Peacock et al., 2021). The cleavage of S glycoprotein into S1 and S2 subunits enables fusion of the virus with the host cell membranes (Chen et al., 2020d; Ge et al., 2020; Huang et al., 2020b) and Figure 6. Given that furin cleavage sites facilitate priming that might increase the spread of SARS-CoVs-2, furin inhibitors can be targeted as potential drug therapies for SARS-CoV-2 (Rabaan et al., 2020). Of particular interest, Deletion of H69/V70 at the spike gene enhances infectivity of SARS-CoV-2 *in vitro* and is associated with immune escape in immunocompromised patients (Davies et al., 2021). Further,

the current omicron variant has three mutations at the furin cleavage site (Chen et al., 2021; Chavda and Apostolopoulos, 2022) a site known to increase SARS-CoV-2 infectivity (Chen et al., 2021; Chavda and Apostolopoulos, 2022) and Figure 4. Consistent with these mutations, omicron variant has an increased risk of reinfection compared to predecessor variants (Chen et al., 2021; Chavda and Apostolopoulos, 2022). Generally, the S1 domain is associated with receptor binding and the S2 domain with cell membrane fusion (Chen et al., 2020d; Ge et al., 2020; Huang et al., 2020b) and (Figure 6). Consistent with this, S1 contains N-terminal domain (NTD) and a receptor-binding domain (RBD) which harbors core domain and external subdomain (ESD) (Huang et al., 2020b). S2 contains three functional domains, fusion peptide (FP), and heptad repeat (HR) one and HR2 (Chen et al., 2020d; Ge et al., 2020; Huang et al., 2020b) and (Figure 6). Thus, whether SARS-CoV-2 can combine with host cells or not is determined by the affinity between viral RBD and the ACE2 receptor of human cells (Huang et al., 2020b; Ge et al., 2020). Once RBD binds to the receptor in the initial stage, then S2 changes conformation, facilitating membrane fusion by its three functional domains (FP, HR1 and HR2) (Chen et al., 2020d; Ge et al., 2020; Huang et al., 2020b). Interestingly, SARS-CoV-2 and SARS-CoV have about 75% amino acid identity in their S protein (Chen et al., 2020d; Ge et al., 2020; Huang et al., 2020b). For instance, the S1 functional domain (RBD) of SARS-CoV-2 has 72%–74.9% identity of amino acid sequences in both viruses, making it a close relative to SARS-CoV (Chen et al., 2020d; Ge et al., 2020; Huang et al., 2020b). As for the functional domains of S2, there is no appreciable difference between SARS-CoV-2 and SARS-CoV except for some non-critical amino acid residues in HR1 region (Ge et al., 2020; Wan et al., 2020). These similarities prompt caution for designing unique primers that do not overlap with these similar regions in the spike gene.

SARS-CoV-2 ACE2 receptor

As noted, SARS-CoV-2 infects human cells such as the alveolar endothelium in the lung by binding to ACE2 receptors, membrane receptor (Ge et al., 2020; Yan et al., 2020; Yang et al., 2020) and (Figure 6). Mechanistically, this binding results in endocytosis of the viral complex with consequent local activation of the renin angiotensin aldosterone system (RAAS), resulting in acute lung injury that may progress to acute respiratory distress syndrome (ARDS) in humans (Ge et al., 2020; Yan et al., 2020; Yang et al., 2020). Mounting data show that several critical residues in SARS-CoV-2 RBD have good interactions with human ACE2 receptors (Chen et al., 2020d; Ge et al., 2020; Huang et al., 2020b; Yan et al., 2020). Most residues of the RBD that interact with ACE2 are fully conserved (Chen et al., 2020d;

Ge et al., 2020; Huang et al., 2020b). Other evidence supporting ACE2 as a receptor of cells is that the HR1 and HR2 domain of SARS-CoV-2 can fuse with each other to form a 6-helical bundle (6-HB) akin to SARS-CoV's fusion mechanism (Ge et al., 2020; Xia et al., 2020). Evidence from studies that utilized structural analysis documented that SARS-CoV-2 strongly interacted with ACE2 compared with SARS-CoV (Chen et al., 2020d; Ge et al., 2020). By elucidating the cryo-EM structure of SARS-CoV-2 S protein, it was documented that SARS-CoV-2 bound to ACE2 with 10–20-fold higher affinity than SARS-CoV (Wrapp et al., 2020). Interestingly, the existence of six mutations in the receptor-binding motif (RBM) conferred SARS-CoV-2 S glycoprotein a higher affinity for ACE2 than SARS-CoV (Berger and Schaffitzel, 2020), suggesting mutation induced gain of function (accounting for higher transmission rate of SARS-CoV-2) (Berger and Schaffitzel, 2020; Walls et al., 2020; Wrapp et al., 2020). Mechanistic evidence informs that ACE2 binding triggers conformational changes promoting proteases to further cleave S2, followed by shedding of S1 and activation of S2 refolding into a post-fusion state (Cai et al., 2020). This is promoted by the acquisition of a furin cleavage site between S1 and S2, a feature important for pathogenicity of SARS-CoV-2 (Wrobel et al., 2020) and Figure 6.

SARS-CoV-2 accessory genes

All accessory proteins are translated from subgenomic RNAs (Fung and Liu, 2019; Kim et al., 2020). Functionally, SARS-CoV-2 accessory proteins plays a role for viral release, stability, pathogenicity, and virulence but are not necessary for virus replication (Baruah et al., 2020). Notably, the genes encoding accessory proteins are distinct in different CoVs in terms of number, genomic organization, sequence, and functions (Hozhabri et al., 2020; Song et al., 2019; Wu et al., 2020; Wu and McGoogan, 2020). Accordingly, the 3′-end of the genome of SARS-CoV-2 harbor nine accessory proteins [3a, 3b, p6, 7a, 7b, 8, 9b and 9c (ORF14) early study and ORF10] (Michel et al., 2020; Yoshimoto, 2020) and along with structural proteins (S, E, M and N). In addition, the accessory proteins in CoVs also vary in location and size in different viral subgroups (Ge et al., 2020; Hozhabri et al., 2020; Michel et al., 2020). For example, SARS-CoV-2 and SARS-CoV are significantly different in terms of gene sequence of two accessory proteins (ORF3b and ORF8) (Chen et al., 2020c; Ge et al., 2020; Michel et al., 2020; Romano et al., 2020). In addition, the ORF10 accessory gene is proposed as unique to SARS-CoV-2 and located downstream of N gene and codes for a 38 amino acid long peptide (Michel et al., 2020; Yoshimoto, 2020). This information is critical for designing unique primers for detection purposes.

SARS-CoV-2 genetic variants

The emergence of new genetic variants of SARS-CoV-2 has spurred intense interest among scientific, clinical, and public health experts while creating anxiety for the population at large. Thus, to understand and identify the variants of concern, information inferred from genomic organization is key. It also helps to understand virus evolution and the genomic epidemiology of SARS-CoV-2 (Galloway et al., 2021; Lauring and Hodcroft, 2021). In this context, it is worth noting that viral mutations (genetic alterations) are not uncommon and occur as a natural consequence of viral replication (Grubaugh et al., 2020; Lauring and Hodcroft, 2021; Mohammadi et al., 2021). Typically, when compared with DNA viruses, RNA viruses tend to have higher rates of mutation (Akkiz, 2021; Lauring and Hodcroft, 2021; Mohammadi et al., 2021). Conversely, owing to their ability to encode an enzyme that corrects some of the errors made during replication, CoVs generate fewer mutations than most RNA viruses (Akkiz, 2021; Lauring and Hodcroft, 2021; Mohammadi et al., 2021). Generally, once genetic alterations occur, natural selection dictates the fate of a newly arising mutation (Akkiz, 2021; Lauring and Hodcroft, 2021; Singh and Yi, 2021). For example, while those mutations that reduce viral fitness will diminish the population of incompetent viruses, those mutations that impart a competitive advantage related to viral replication, transmission, or escape from host immunity, will increase within a population. (Lauring and Hodcroft, 2021). Consequently, virus evolution and spread within hosts, in communities, and across countries are shaped at least in part by the natural selection (26,67,68, (Singh and Yi, 2021). Thus, genomic organization of SARS-CoV-2 helps to provide useful genomic information related to evolution of a new SARS-CoV-2 variant and which mutations are being enriched to allow increased circulation in the population (Akkiz, 2021; Lauring and Hodcroft, 2021; Mohammadi et al., 2021). Of particular interest is the identification that these variants have a high number of mutations in the S protein within the amino terminal domain (NTD) and Receptor-binding domain (RBD). These mutations have a direct influence on viral infection rate by enhancing the affinity of RBD for the ACE2 receptor (Akkiz, 2021; Grubaugh et al., 2020; Gomez et al., 2021; Lauring and Hodcroft, 2021; Singh and Yi, 2021) and (Figures 3, 4). Comprehending the adaptive benefit of these mutations on transmissibility, antigenicity, or virulence is of primary importance in any effort geared towards intervention and eventual quelling of community spread (Akkiz, 2021; Grubaugh et al., 2020; Gomez et al., 2021; Lauring and Hodcroft, 2021; Singh and Yi, 2021). In this context, genomic surveillance and laboratory experiments are important players during the emergence of new stains as the former provide information regarding which mutations are emerging, and later can help determine how these mutations change the virus from its predecessor (original virus) in terms of transmissibility, virulence, host evasion, response to interventional therapy and detection (Grubaugh et al., 2020; Lauring and Hodcroft, 2021). In the end, these tools help us identify and comprehend which variants are of greatest concern. Accordingly, several notable variants of concern (VOC) have been identified and are being monitored for implementation of public health strategies that mitigate transmission risks. Next, we will review previously circulating VOC such as Alpha, beta, gamma, and delta (Figure 3), and recently circulating VOC notably omicron and its subvariants (Figure 4).

Previously circulating SARS-CoV-2 variants of concern (VOCs)

Lineage B.1.1.7 also called 501Y.V1 (alpha variant)

The lineage B.1.1.7 also known as 501Y.V1 (Alpha Variant, WHO naming) was a variant of concern (VOC), identified initially in United Kingdom in September 2020. It harbors several mutations in the spike glycoprotein and other sites such as OR1F1ab, Orf8 and N (Grubaugh et al., 2020; Gomez et al., 2021; Lauring and Hodcroft, 2021; Mohammad et al., 2021) and Figure 3. For example, the N501Y mutation occurs in the Receptor-binding domain (RBD) whereas the P681H occurs near the S1/S2 furin cleavage site, a site with high variability in CoVs (Grubaugh et al., 2020; Gomez et al., 2021; Lauring and Hodcroft, 2021) and a potential therapeutic drug target (Rabaan et al., 2020). Data show that acquisition of a P681H mutation at the furin site enhance cleavage of Spike protein (Davies et al., 2021; Shrestha et al., 2022). Further, report from structural modelling of SARS-CoV-2 documented alpha variant having an enhanced furin binding and infectivity (Mohammad et al., 2021). Given the 501Y spike variants are predicted to have a higher affinity for human ACE2 receptors, conceivably, these mutations could potentially influence ACE2 binding and viral replication (Grubaugh et al., 2020; Gomez et al., 2021; Lauring and Hodcroft, 2021; Mohammad et al., 2021), conferring competitive fitness for community spread. Consequently, the expansion of lineage B.1.1.7 was linked to wide spread of SARS-CoV-2 cases, achieving dominance by outcompeting an existing population of circulating variants (Davies et al., 2021; Grubaugh et al., 2020; Gomez et al., 2021; Lauring and Hodcroft, 2021). For instance, population genetic models suggested that lineage B.1.1.7 was spreading 56% more quickly than other lineages (Davies et al., 2021; Lauring and Hodcroft, 2021). Despite the quick spread, evidence showed that vaccines remained effective against this variant lineage (Lauring and Hodcroft, 2021; Wu et al., 2021; Xie et al., 2021).

Lineage B.1.351 also called 501Y.V2 (Beta Variant)

The lineage B.1.351 also known as 501Y.V2 (Beta Variant, WHO naming) was another variant of concern (VOC) that was detected initially in August 2020 in South Africa. It harbors
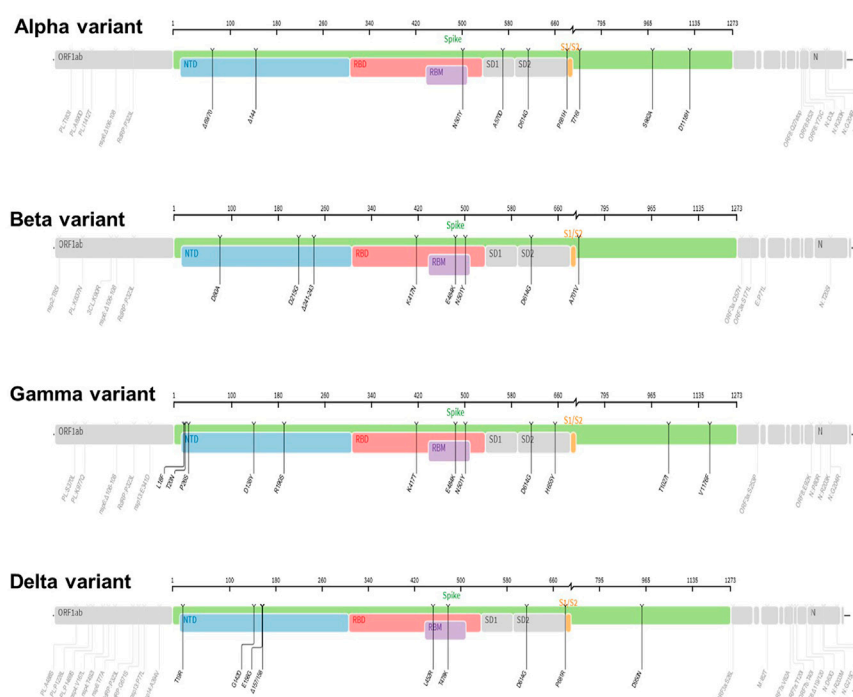
FIGURE 3
The mutation profiles of previously circulating different SARS-CoV-2 variants (VOCs). Accordingly, mutations profiles in the spike regions and other parts of SARS-CoV-2 genome are depicted. The mutation profiles are depicted from top to bottom in the order of Alpha, Beta, Gamma, and Delta VOCs. The pictures from covdb.stanford.edu are licensed under a Creative Commons Attribution-ShareAlike 4.0 International License https://creativecommons.org/licenses/by-sa/4.0/.

multiple mutations in the spike protein notably, changes in the RBD site, K417 N/T, E484K, and N501Y as well as other part of the genome, OR1F1ab, Orf3a, E and N (Galloway et al., 2021; Gomez et al., 2021; Hoffmann et al., 2021) and Figure 3. Based on evidence from reduced neutralization by convalescent and post-vaccination sera, studies indicate that one of the spike protein mutations, E484K may affect neutralization capacity of some polyclonal and monoclonal antibodies (Weisblum et al., 2020; Wang et al., 2021b). Further, the linage B.1.351 evade infection induced antibody responses, certain therapeutic antibodies, and vaccination (Hoffmann et al., 2021). Specifically, acquisition of E484k mutation conferred resistance to NTD raised most mAbs and RBD motif raised multiple individual mAbs (Wang et al., 2021b). In addition, B.1.351 is 9.4-fold more resistant to neutralization by convalescent plasma and 10.3–12.4 fold to sera from vaccinated individual (vaccinee sera) (Wang et al., 2021b).

Lineage P.1 also called 501Y.V3 (gamma variant)

The lineage P.1 also known as 501Y.V3 (Gamma variant, WHO naming) was identified initially in July 2020 in Brazilian travelers to Japan. It acquired mutations in the Receptor-binding domain of spike proteins notably, K417 N/T, E484K, and N501Y as well as another site of the genome including, OR1F1ab, Orf3a,

Orf8 and N (Hoffmann et al., 2021) and Figure 3. This variant was later identified as the dominant variants in Brazil and was subsequently detected in the USA and several other countries (Galloway et al., 2021; Gomez et al., 2021). The emergence of the gamma variant raised concerns owing to increased transmission, propensity for re-infection (Resende et al., 2021) and reduced neutralization by convalescent and post-vaccination sera (Wang et al., 2021c). Report showed that linage P.1 evaded antibody responses to infection, certain therapeutic antibodies, and vaccination (Hoffmann et al., 2021; Wang et al., 2021b).

Lineage B.1.617.2 (delta variant)

The lineage B.1.617.2 (Delta variant, WHO naming) was another variant of concern (VOC) that was initially identified in December 2020 in India, with subsequent detection in the United States, and worldwide distribution (Lopez Bernal et al., 2021; Singh et al., 2021). This variant has acquired distinct mutations in the spike protein including P681R, L452R and D950N as well as mutations on another site of the genome, OR1F1ab, Orf3, M, Orf7a, Orf8 and N (Mlcochova et al., 2021; Zhan et al., 2022), Figure 3. Interestingly, it lacks E484K, and N501Y mutation which are shared by other viral strains (Zhan et al., 2022). Notably, Delta variant has a higher transmission rate and immune evasion capacity (Lopez Bernal et al., 2021; Singh
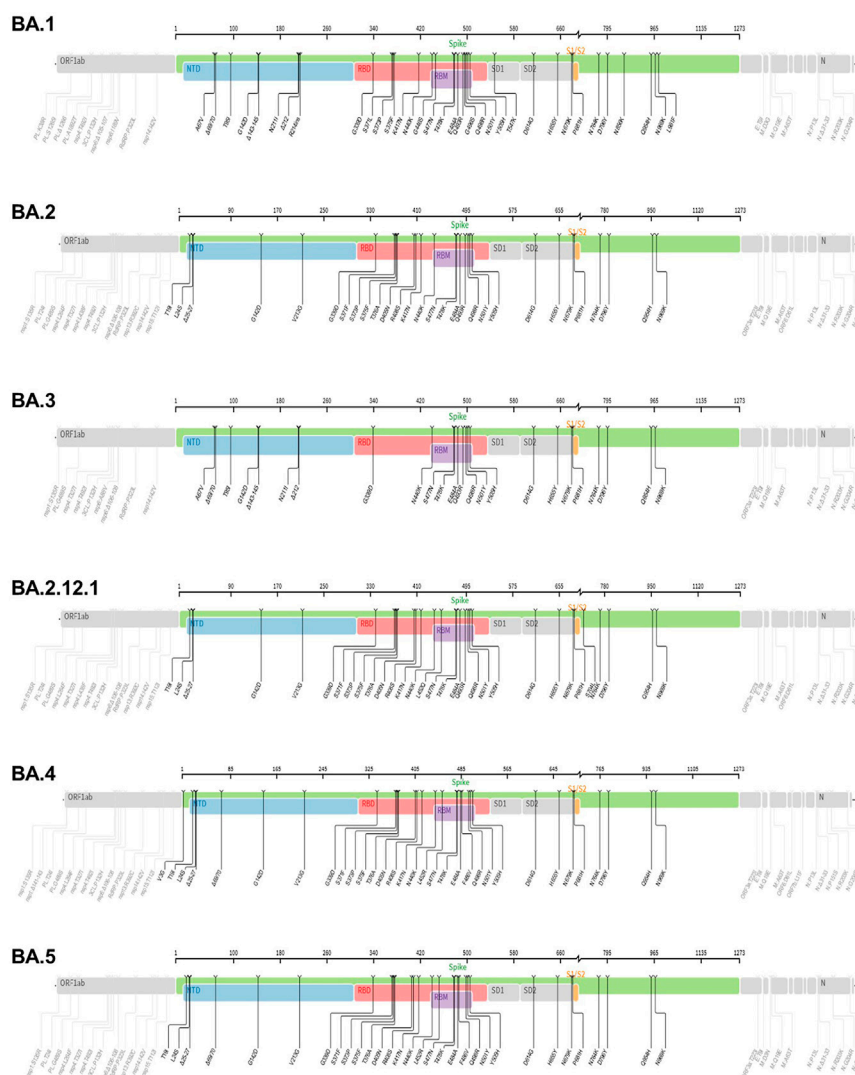
FIGURE 4
The mutation profiles of currently circulating Omicron variant (VOC) subvariants. Accordingly, mutations in the spike regions and other parts of SARS-CoV-2 genome are depicted. The mutation profiles are depicted from top to bottom in the order of BA.1, BA.2, BA.3, BA.2.12.1, BA.4 and BA.5 subvariants. he pictures from covdb.stanford.edu are licensed under a Creative Commons Attribution-ShareAlike 4.0 International License https://creativecommons.org/licenses/by-sa/4.0/.

et al., 2021). As noted, the furin cleavage site is important for determining fusion of the virus with the host cells (Figure 6). Report show that acquisition of a P681R mutation at the furin site enhances cleavage of the Spike protein (Shrestha et al., 2022). Data from a study looking into effectiveness of COVID-19 vaccines against the B.1.617.2 (Delta variant) showed lower efficacy of BNT162b2 or ChAdOx1 nCoV-19 vaccines among persons with the Delta variant than among those with the Alpha variant after receiving single vaccine doses (Lopez Bernal et al., 2021). Further, an *in vitro* study revealed that the delta variant is 6-fold less sensitive to serum neutralizing antibodies obtained from recovered individuals, and 8-fold less sensitive to antibodies elicited from vaccine, in comparison to wild type (Mlcochova

et al., 2021). Next, we will review the VOC currently in circulation.

Lineage B.1.1.529 (omicron variant)

The lineage B.1.1.529 (Omicron variant, WHO naming) is the recent variant of concern (VOC) that was reported initially on 24 November 2021from South Africa (Chavda and Apostolopoulos, 2022). Evaluation of the evolution of SARS-CoV2 using a sequence-based approach revealed that B.1.1.529 variants have acquired many mutations, implicating omicron variant as the most highly mutated strain among all SARS-CoV-2 variants (Nie et al., 2022; Teng et al., 2021; Tuekprakhon et al., 2022). As depicted in Figure 4, the

Omicron variant has significantly more missense mutations accumulated in spike and other genome regions compared to the predecessor VOCs, Alpha, Beta, Gamma, and Delta variants (Figure 4). For example, more than 60 mutations have been identified on several genomic regions of the virus namely, Spike protein, ORF1ab, Envelope, Membrane, Nucleocapsid proteins (Nie et al., 2022). Notably, the spike protein which is the antigenic target for antibody production by infected host and vaccines have been targeted by 26–35 mutations which are unique (nonsynonymous) from the original SARS-CoV-2 variants (Ren et al., 2022). Of note, omicron has high binding affinity to ACE2 receptor due to acquisition of substitutions mutations such as Q493R, N501Y, S371L, S373P, S375F, Q498R, and T478K at the RBD site (Shrestha et al., 2021; Araf et al., 2022) and Figure 4. Additionally, the variant has three mutations at the furin cleavage site, namely N679K, H655Y, and P681H (Chen et al., 2021; Chavda and Apostolopoulos, 2022), a site known to increase SARS-CoV-2 infectivity (Chen et al., 2021; Chavda and Apostolopoulos, 2022). Consistent with these mutations, the Omicron variant has an increased risk of reinfection compared to predecessor variants (Chen et al., 2021; Chavda and Apostolopoulos, 2022). For instance, Omicron spread faster than the Delta variant albeit causing less severe disease (Ren et al., 2022).

Conceivably, the high rate of spread, ability to evade double vaccination, and the host immune system can increase the burden on the health care systems, even though Omicron infections are less fatal than infections with the Delta variant (Cao et al., 2022a; Ren et al., 2022). To counter the burden posed by Omicron, vaccination with a third dose of mRNA vaccine can provide protection against severe disease and hospitalization by supercharging the neutralizing antibody levels and strengthening host immunity (Ren et al., 2022). Further, extending vaccine boosting efforts across different age groups can mitigate the strain on the health care system as evidenced by the BNT162b2 vaccine conferring protection against Omicron variant in children and adolescents (Price et al., 2022).

The Omicron subvariants

The Omicron (B.1.1.529) variant has several subvariants, namely BA.1 (B.1.1.529.1), BA.2 (B.1.1.529.2), BA.3 (B.1.1.529.3), BA.2.12.1, BA.4 and BA.5 (Desingu and Nagarajan, 2022; Majumdar and Sarkar, 2022). They share many of the mutations while also being significantly different (Desingu and Nagarajan, 2022; Majumdar and Sarkar, 2022). Interestingly, these subvariants were detected in South Africa indicating fast evolutionary divergence of the subvariants (Rahimi and Talebi Bezmin Abadi, 2022). Importantly, the existence of many mutations on several regions of the genome (Figure 4) raises questions as to why such genomic instability is occurring in a virus that has an error correcting (proof reading)

exonuclease (the nsp14 protein) (Chen et al., 2021; Gribble et al., 2021). The nsp14 protein mediate viral recombination and is highly conserved among CoVs (Gribble et al., 2021). Despite its role contributing to a low mutation rate and stability of the viral genome, more than six million viral genomes have been documented, suggesting the instability of SARS-CoV-2 genome (Eskier et al., 2020; Thakur et al., 2022). Consistent with this and as shown in Figure 4, the diverse Omicron subvariants (BA.1, BA.2, BA.3, BA.2.12.1, B.4 and BA.5) harbor nsp14 mutations (Figure 4). Further, data reveal mutations in SARS-CoV-2 nsp14 have the strongest association with increased mutation load across the genome compared to nsp7, nsp8 and nsp12 which form the core polymerase complex (Eskier et al., 2020). As discussed above, these subvariants are distinct. For instance, evidence reveals that BA.2 is more transmissible and may cause more severe disease than BA.1 ((Rahimi and Talebi Bezmin Abadi, 2022), (Chen and Wei, 2022)). Further BA.2 is not affected by the therapeutic monoclonal antibodies used to treat people infected with COVID, making BA.2 more resistant to casirivimab, imdevimab and sotrovimab monoclonal antibodies than the original Omicron variant (B.1.1.529) (Iketani et al., 2022) suggesting the fitness of BA.2 subvariant.

The BA.2.12.1 subvariant is diverged from BA.2 by acquiring additional spike mutations S704L and L452Q on top of BA.2 background and causing breakthrough infections in fully vaccinated and boosted individuals (Beheshti Namdar and Keikha, 2022). For example, prior infections with BA.1 appear to confer minimal cross immunity to BA.212.1. As a result, an individual with BA.1 infection can also be infected with BA.2.12.1 (Cao et al., 2022b; Del Rio and Malani, 2022). Early studies document that BA.2.12.1 is about 25% more transmissible than BA.2 but does not appear to cause severe disease (Beheshti Namdar and Keikha, 2022). Consequently, approximately 58% of SARS-CoV-2 isolates sequenced belong to BA.2.12.1 as of May 2022 (Del Rio and Malani, 2022), suggesting fitness of this subvariant.

The BA.4 and BA.5 subvariants also resemble more of BA.2 than BA.1 and emerged in south Africa and Europe displaying competitive advantage in terms of viral replication, transmissibility, and immune evasion (Del Rio and Malani, 2022; Tuekprakhon et al., 2022). Particularly, the acquisition of L452R and F486V substitution mutation confer BA.4 and BA.5 the ability to have increased ACE2-binding affinity, stronger neutralization evasion, as well as higher transmissibility than BA.2 (Del Rio and Malani, 2022; Tuekprakhon et al., 2022). Given the potential of the 69-70del, L452R, and F486V substitution mutations to alter the binding affinity between SARS-CoV-2 spike and human ACE2 (Teng et al., 2021; Tuekprakhon et al., 2022), we have performed an *in silico* analysis to investigate the effects of viral variations on spike-ACE2 interaction. We showed that some mutations may enhance the binding affinity (Teng et al., 2021). For example, L452R is

FIGURE 5
Structural representation of SARS-Cov-2 Spike and human ACE2 proteins. SARS-CoV-2 spike receptor-binding domain (RBD, yellow) and its interaction regions (orange), Human receptor ACE2 (green) and its interaction regions (cyan) were displayed as cartoons. SARS-CoV-2 spike L452 residues (blue) was shown as sticks. The figure was generated using PyMOL (http://www.pymol.org/) based on PDB ID: 6LZG.

located in the interface of spike-ACE2 complex (Figure 5). We observed that L452R can increase the binding affinity ($\Delta\Delta\Delta G$ = −0.395 kcal/mol) between SARS-CoV-2 spike and human ACE2 (Teng et al., 2021). Consistent with our finding, L452R mutation was identified in Omicron BA.4/5 variant and found to play a critical role in the immune escape (Tuekprakhon et al., 2022).

SARS-CoV-2 variants and molecular detection

For detection of SARS-CoV-2 using the real time-polymerase chain reaction (RT-PCR), multiple genes are targeted in a

diagnostic assay (Thakur et al., 2022). This is to ensure specificity and sensitivity of SARS-CoV-2 detection and minimize false positivity. Typical viral genes that are used for such diagnostic purposes include S, N (containing N1 and N2 regions of the N gene), E, nsp12, nsp14, ORF1ab *etc.* (Wang et al., 2020). Particularly, targeting the spike gene ensures specificity as it has unique nucleotide sequences to SARS-CoV-2 and hence minimizes cross-reactivity which would otherwise occur in the presence of other CoVs (Thakur et al., 2022). Given the observed frequent mutation on the S gene, commercial kits and probe sets must be regularly validated to detect new variants and avoid false-negative results (Wang et al., 2020; Thakur et al., 2022). Consistent with this, the S gene target failure has been reported in Alpha and Omicron variants (Wang et al., 2020; Thakur et al., 2022). As a result, a qPCR test that utilize S-gene target failure became useful for rapid detection of the Omicron variant from the Delta variant (Wolter et al., 2022). Conversely, BA.2 lacks the characteristic S-gene target failure, making Delta and BA.2 similar and thus difficult to distinguish the two variants (Rahimi and Talebi Bezmin Abadi, 2022). Owing to the possibility of overlooking BA.2 subvariant (dubbed stealth omicron), S-gene target failure alone may not be sufficient for monitoring the spread of Omicron (Rahimi and Talebi Bezmin Abadi, 2022). Alternatively, BA.2 can be separated from other variants through sequencing or designing primers targeting specific mutations (Rahimi and Talebi Bezmin Abadi, 2022). Of note, although other genes such as N and RdRp (nsp12) are less prone to mutations, any mutation in the primer binding region can reduce assay sensitivity and result in false negative report (Rahman et al., 2021; Wang et al., 2020). For instance, mutations within the N-gene that could affect the sensitivity of RT-PCR tests has been described (Rahman et al., 2021; Wang et al., 2020). Similarly, SNP Q289H which occurred in the N-gene impacted forward primer binding and markedly reduced RT-PCR assay sensitivity (Vanaerschot et al., 2020). Particularly, as omicron variants harbor nine-nucleotide deletions in the N-gene, spanning positions 28,370–28362, single N gene detection-based RT-PCR test assays are expected to produce false negative results (Thakur et al., 2022; Vanaerschot et al., 2020; Wang et al., 2020).

As noted above, the challenge caused by VOCs is clear and there is an urgent need to conduct genetic epidemiological surveys to manage the ongoing pandemic and for better patient care. In this regard, the best approach would be whole genome sequencing (WGS). However, WGS is expensive and time consuming to run on a large scale (Kanjilal and Tang, 2022; Wang et al., 2021d). Alternatively, a rapid and reliable multiplex PCR platform that utilizes a set of four mutation specific PCR based assay has been developed (Wang et al., 2021e). These assays can detect all three RBD spike mutations, including N501Y, E484K, H69-V70del and L450R (Wang et al., 2021e). A similar approach is extended to identify the early circulating SARS-CoV-2 Omicron variant by targeting an Omicron specific Spike (S) insertion-deletion mutation (indel_211–214) observed in B1.1.529/BA.1 lineage and

subvariants (Sibai et al., 2022). Similarly, multiplex platforms that can detect five VOC and three variants of interest (VOI) has been developed by screening for spike protein deletions 69 to 70, and 242 to 244 (S-D242-244) as well as S-N501Y, S-E484K, and S-L452R mutations in clinical samples known to be positive for SARS-CoV-2 during the early 2021 (Wang et al., 2021d). Of interest, these multiplex RT-PCR results have 100% concordance with the strains identified by WGS, corroborating their accuracy (Wang et al., 2021d). Understandably, these multiplex RT-PCR based assays may not have high enough throughput to meet the demand of evolving SARS-CoV-2 variants (Boudet et al., 2021), necessitating the development of high throughput screening tests. CoVarScan is a single PCR assay that use 8-plex fragment analysis, detecting eight mutation regions, three recurrently deleted regions (RDR: RDR1, RDR2, and RDR3-4), three RBD mutations (N501Y, E484K, and L452R), and two ORFs (ORF1A and ORF8) (Clark et al., 2022). Thus, CoVarScan has a merit of identifying almost all VOCs, VOIs and capable of detecting the newly emerging SARS-CoV-2 variants (Aoki et al., 2022; Clark et al., 2022), making it a robust detection strategy to efficiently triage a subset of positive samples for confirmation and negative samples for identification using the time consuming and costly whole genome sequencing (WGS).

Interactions of SARS-CoV-2 with the host and protease required

To initiate successful entry, proteolysis is a fundamental machinery that SARS-CoV-2 uses to deploy fusion peptide for insertion into membrane and gain access into the host cell (Koppisetti et al., 2021). To this end, SARS-CoV-2 binds with the host cell receptor ACE2, which paves the way for the spike protein to undergo proteolysis during infection processes (Naqvi et al., 2020). In this regard, the internal fusion peptide needs to be exposed *via* conformational change (Zhang et al., 2021). These actions ultimately promote plasma membrane fusion (endocytosis), allowing the virus to enter and release its viral RNA into the cytoplasm and for translation to ensue in the host cells (Koppisetti et al., 2021; Naqvi et al., 2020; Zhang et al., 2021) and Figure 6. As shown in Figure 6, three proteases are implicated in the proteolysis of SARS-CoV-2 spike protein, namely furin, a proprotein convertase, TMPRSS2 (transmembrane protease serine 2), a cell surface protease, and cathepsin, a lysosomal protease (Liu et al., 2021). Functionally, furin is a type 1 transmembrane protein that detaches S1 from S2 domain by cleaving the PRRA, a multibasic site at the S1/S2 boundary (Liu et al., 2021). On the other hand, TMPRSS2 is a type 2 transmembrane protein that facilitates membrane fusion by cleaving the S2 site and exposing the internal fusion peptide (Liu et al., 2021; Naqvi et al., 2020). The lysosomal localized cathepsin promote fusion of viral envelope with endosomal membrane by inducing proteolysis after endocytosis of the virion (Liu et al., 2021; Naqvi et al., 2020) and (Figure 6).

Role of SARS-CoV-2 moiety as PAMPs and TLR as PRRs

Notably, the severity and lethality of SARS-CoV-2 infection is associated with cytokines storm, where SARS-CoV-2 induced hyperactivity of the immune system inflicts damage to the implicated tissue or organs (Choudhury and Mukherjee, 2020; Choudhury et al., 2021a; Patra et al., 2021). Mechanistically, the host innate immune system induces inflammatory responses and eliminates pathogens during infection by recognizing pathogen associated molecular patterns (PAMPs) *via* pattern recognition receptors (PRRs) (Choudhury and Mukherjee, 2020; Choudhury et al., 2021a; Patra et al., 2021). A notable example of PRR is the toll like receptor (TLR) among others (Choudhury and Mukherjee, 2020; Choudhury et al., 2021a; Patra et al., 2021). These SARS-CoV-2 sensing, and inflammatory cytokines production are facilitated by the activation of the MyD88-dependent proinflammatory cytokine production (Kurt-Jones et al., 2000; Zheng et al., 2021). Myd88 is a TLR adapter protein that is required to produce inflammatory cytokines such as TNF-α and IL-6 following β-coronavirus infection (Zheng et al., 2021). In this context, several recent works document the role of various moieties of the SARS-CoV-2 as ligand for the TLR, a PRR. For instance, molecular docking studies have demonstrated significant binding of SARS-CoV-2 native spike protein to TLR1, TLR4, and TLR6 with TLR4 displaying the strongest protein-protein interaction with spike protein (Choudhury and Mukherjee, 2020). Interestingly, recent study also documented another SARS-CoV-2 component (envelope protein) as a ligand for TLR2 at the extracellular level (Zheng et al., 2021). In this report, E protein induced inflammation comparable with PAMP3 without the need for viral entry (Zheng et al., 2021). Further, a recent *in silico* study documented the role of intracellular TLRs in SARS-CoV-2 mRNA sensing, where TLR3 binds NSP10, TLR7 binds E protein, and TLR9 binds S2 mRNA, suggesting NSP10, E protein and S2 as possible virus-associated molecular patterns (PAMPs) (Choudhury et al., 2021a). These observations inform the possibility of targeting SARS-CoV-2 moiety (PAMPs) and TLR (PRR) interaction for therapeutic purposes.

Host immunity, immunomodulation, vaccine, chemotherapy or combinatorial approaches against emergent SARS-CoV-2 variants

As noted above, the spike glycoprotein is the hot spot for genetic alterations, making it an interesting region of the SARS-CoV-2 genome for research and monitoring purposes. Consequently, genomic surveillance of SARS-CoV-2 variants has largely focused on mutations in the spike glycoprotein for two main reasons (Lauring and Hodcroft, 2021; Lopez Bernal et al., 2021). Firstly, S protein
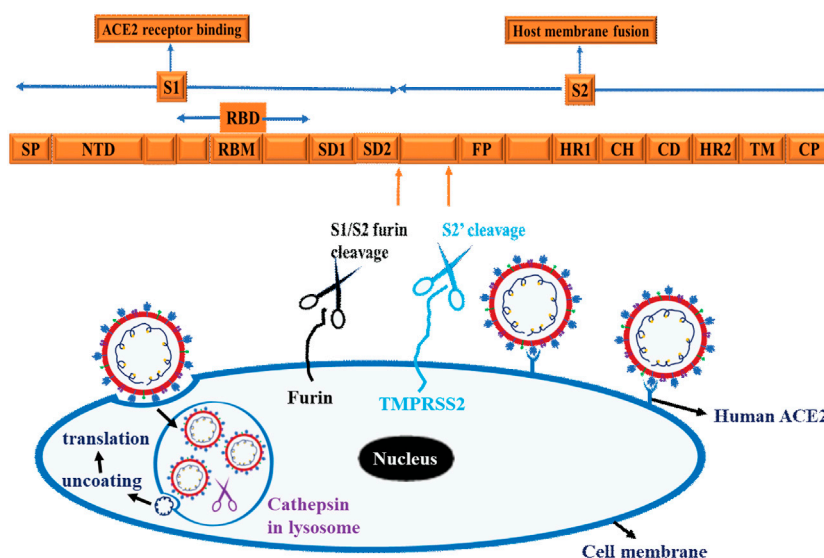
FIGURE 6
Schematic representation of SARS-CoV-2 and Host cell interaction. The Scheme depicts spike protein, cell binding, endocytosis, viral uncoating, and translation. Three proteases are implicated in the proteolysis of SARS-CoV-2 spike protein, namely Furin, a proprotein convertases, TMPRSS2 (transmembrane protease serine 2), a cell surface protease and cathepsin, a lysosomal protease. These schematics are based on information extracted from various studies (Koppisetti et al., 2021; Naqvi et al., 2020; Zhang et al., 2021) and (Liu et al., 2021).

mediates attachment to the host cells ACE2 receptor and secondly, it is a major target of neutralizing antibodies produced by host immune responses either following infection or by vaccination (Lauring and Hodcroft, 2021; Lopez Bernal et al., 2021). Given the potential of these genetic alterations to compromise vaccine effectiveness and escape from host antibodies (immune evasion), the intense interest in mutations occurring in the spike glycoprotein and other sites (Wang et al., 2021c; Resende et al., 2021) is a feasible research strategy. The current Omicron variants and its sub-lineages are good examples (Figure 4). Thus, defining these genetic alterations, and their potential influence on vaccine effectiveness, will require large-scale monitoring of SARS-CoV-2 evolution and host immunity. The rise of more transmissible variants like Delta, currently Omicron and its divergent sub-lineages reinforce the need for further SARS-CoV-2 genomic surveillance and alternative therapeutic approaches. As discussed above, cognizant of spike glycoprotein as main arsenals of SARS-CoV-2 infection, its role as a ligand for human TLR4, warrants further therapeutic avenues targeting TLR4. Consistent with this notion, next we attempt to review adjunct therapeutic options.

Targeting PAMPs/PRRs (SARS-CoV-2 host interaction) with adjunct combinatorial therapy

Cognizant of TLR as the main innate immune receptor recognizing pathogenic ligands and inducing proinflammatory responses, targeting this pathway for therapeutic purpose holds great promise in an effort geared towards mitigation of COVID-19 inflicted immunopathological manifestation (Choudhury and Mukherjee, 2020; Choudhury et al., 2021a; Patra et al., 2021). Thus, in addition to drugs and/or vaccines in current development, understanding the mechanism of inflammatory cytokine production in the host *via* β-coronavirus sensing immunological arsenals such as TLR, opens new approach for adjunct therapeutic strategies to counter the severe illness due to COVID-19 (Choudhury and Mukherjee, 2020; Choudhury et al., 2021a; Patra et al., 2021). In this context, the development of a multi-epitope multi-target chimeric vaccine (named AbhiSCoVac) that stably interact with the TLRs and MHC receptors has been reported using the in slilico based biocomputational approach (Choudhury et al., 2022). Since AbhiSCoVac is hypothesized to work against all pathogenic coronaviruses, it can be a plausible therapeutic approach for managing emerging SARS-CoV-2 variants (Choudhury et al., 2022). Similar approach could be extended for the heart as cardiac injury is another hallmark of covid lethality. Thus, targeting TLR hold great promise for patients inflicted by COVID-19 induced cardiac injury (Choudhury and Mukherjee, 2021). In another context, owing to reduced efficacy of single monoclonal antibodies (MABs) against emerging variants, the need for chimeric antibodies that can target mutant variant and thus display improved efficacy over single MAB has been reported (Das et al., 2021). Importantly, combinatorial therapy that entails chemotherapy and

immunotherapy as curative therapeutic remedies needs further investigation (Choudhury et al., 2021b). Thus, immunotherapy such as vaccine that offers preventive role and restore immune homeostasis and, chemotherapy that could potentially thwart host virus interactions, inhibit viral entry, proliferation, and dampen damaging inflammatory environment in COVID-19 patients (Choudhury et al., 2021b) are viable options to explore.

Concluding remark

Understanding the genome organization of SARS-CoV-2 is indispensable as it familiarizes clinical laboratorians (molecular diagnostics), pathologists (molecular pathology), epidemiologists (molecular epidemiology) and researchers with the products of the viral genome, (proteins enzymes, antigens, and antibodies). As depicted in Figure 1, two-thirds of SARS-CoV-2 genomes encode for viral replicase protein (largest gene coding for ORF1); and the remaining one-third encode for four structural proteins including the S, E, M, and N proteins as well as accessary protein which are embedded within the structural proteins. In this review, we have attempted to provide insight on the genome organization and protein products focusing on the SARS-CoV-2. We have provided useful information regarding the implication of genome structure for understanding emerging variants, such as Alpha, Beta, Delta Gamma (previously circulating VOCs) and Omicron variants (current VOC) and its divergent subvariants (BA.1, BA.2 and BA.3, BA.2.12.1, BA.4 and BA.5) with BA.5 subvariants on the path to become the dominant variant, and is thus being monitored along with others.

Thus, defining these genetic alterations, and their potential influence on detection and therapy/vaccine effectiveness will require large-scale monitoring of SARS-CoV-2 evolution and host immunity. Further, the rise of more transmissible variants like Omicron and its divergent sub-variants reinforce the need for further SARS-CoV-2 genomic surveillance and alternative detection and therapeutic approaches. Thus, future endeavors will focus on the following viable options:

1) The development of rapid and less costly detection tools such as multiplex RT-PCR based assays that has a merit of identifying all VOCs, VOIs and detecting the newly emerging SARS-CoV-2 variants is an important viable option. Strategically, the robustness of such affordable detection tools helps to efficiently triage a subset of positive variants for confirmation and negative samples for

further identification using the expensive and time-consuming whole genome sequencing (WGS).
2) The development of combinatorial therapy that entails immunotherapy such as vaccine that offers preventive role and restore immune homeostasis and, chemotherapy that could potentially thwart host virus interactions, inhibit viral entry, proliferation, and dampen damaging inflammatory environment in COVID-19 patients are viable options to explore.

Taken together, information extracted from the genomic structure of SARS-CoV-2 is of paramount importance in variant identification, genomic surveillance, assessment of mechanism of transmission, virulence, and development of specific vaccines tailored to emergent variants, design of chimeric antibodies, and antiviral drugs or combinational therapies.

Author contributions

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Akkiz, H. (2021). Implications of the novel mutations in the SARS-CoV-2 genome for transmission, disease severity, and the vaccine development. *Front. Med.* 8, 636532. doi:10.3389/fmed.2021.636532

Alimohamadi, Y., Tola, H. H., Abbasi-Ghahramanloo, A., Janani, M., and Sepandi, M. (2021). Case fatality rate of COVID-19: A systematic review and meta-analysis. *J. Prev. Med. Hyg.* 62, E311–E320. doi:10.15167/2421-4248/jpmh2021.62.2.1627

Aoki, A., Mori, Y., Okamoto, Y., and Jinno, H. (2022). PCR-based screening tests for SARS-CoV-2 mutations: What is the best way to identify variants? *Clin. Chem.* 68, 1000–1001. doi:10.1093/clinchem/hvac087

Araf, Y., Akter, F., Tang, Y. D., Fatemi, R., Parvez, M. S. A., Zheng, C., et al. (2022). Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. *J. Med. Virol.* 94, 1825–1832. doi:10.1002/jmv.27588

Ashour, H. M., Elkhatib, W. F., Rahman, M. M., and Elshabrawy, H. A. (2020). Insights into the recent 2019 novel coronavirus (SARS-CoV-2) in light of past human coronavirus outbreaks. *Pathogens* 9, 136. doi:10.3390/pathogens9030186

Baez-Santos, Y. M., St John, S. E., and Mesecar, A. D. (2015). The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. *Antivir. Res.* 115, 21–38. doi:10.1016/j.antiviral.2014.12.015

Baruah, C., Devi, P., and Sharma, D. K. (2020). Sequence analysis and structure prediction of SARS-CoV-2 accessory proteins 9b and ORF14: Evolutionary analysis indicates close relatedness to bat coronavirus. *Biomed. Res. Int.* 2020, 7234961. doi:10.1155/2020/7234961

Beheshti Namdar, A., and Keikha, M. (2022). BA.2.12.1 is a new omicron offshoot that is a highly contagious but not severe disease. *Ann. Med. Surg.* 79, 104034. doi:10.1016/j.amsu.2022.104034

Bekaert, M., and Rousset, J. P. (2005). An extended signal involved in eukaryotic -1 frameshifting operates through modification of the E site tRNA. *Mol. Cell* 17, 61–68. doi:10.1016/j.molcel.2004.12.009

Berger, I., and Schaffitzel, C. (2020). The SARS-CoV-2 spike protein: Balancing stability and infectivity. *Cell Res.* 30, 1059–1060. doi:10.1038/s41422-020-00430-4

Boudet, A., Stephan, R., Bravo, S., Sasso, M., and Lavigne, J. P. (2021). Limitation of screening of different variants of SARS-CoV-2 by RT-PCR. *Diagn. (Basel)* 11, 1241. doi:10.3390/diagnostics11071241

Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S. M., Walsh, R. M., Jr, et al. (2020). Distinct conformational states of SARS-CoV-2 spike protein. *Science* 369, 1586–1592. doi:10.1126/science.abd4251

Cao, Y., Wang, J., Jian, F., Xiao, T., Song, W., Yisimayi, A., et al. (2022). Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* 602, 657–663. doi:10.1038/s41586-021-04385-3

Cao, Y., Yisimayi, A., Jian, F., Song, W., Xiao, T., Wang, L., et al. (2022). BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature* 608, 593–602. doi:10.1038/s41586-022-04980-y

Chan, J. F., Kok, K. H., Zhu, Z., Chu, H., To, K. K. W., Yuan, S., et al. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* 9, 221–236. doi:10.1080/22221751.2020.1719902

Chan, J. F., Lau, S. K., and Woo, P. C. (2013). The emerging novel Middle East respiratory syndrome coronavirus: The "knowns" and "unknowns". *J. Formos. Med. Assoc.* 112, 372–381. doi:10.1016/j.jfma.2013.05.010

Chavda, V. P., and Apostolopoulos, V. (2022). Omicron variant (B.1.1.529) of SARS-CoV-2: Threat for the elderly? *Maturitas* 158, 78–81. doi:10.1016/j.maturitas.2022.01.011

Chen, J., Wang, R., Gilby, N., and Wei, G. (2021). Omicron (B.1.1.529): Infectivity, vaccine breakthrough, and antibody resistance. *Version 1. ArXiv.* arXiv:2112.01318v1.

Chen, J., and Wei, G. W. (2022). Omicron BA.2 (B.1.1.529.2): High potential to becoming the next dominating variant. *J. Phys. Chem. Lett.* 13, 3840–3849. doi:10.21203/rs.3.rs-1362445/v1

Chen, L., Liu, W., Zhang, Q., Xu, K., Ye, G., Wu, W., et al. (2020). RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. *Emerg. Microbes Infect.* 9, 313–319. doi:10.1080/22221751.2020.1725399

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, China: A descriptive study. *Lancet* 395, 507–513. doi:10.1016/S0140-6736(20)30211-7

Chen, Y., Guo, Y., Pan, Y., and Zhao, Z. J. (2020). Structure analysis of the receptor binding of 2019-nCoV. *Biochem. Biophys. Res. Commun.* 525, 135–140. doi:10.1016/j.bbrc.2020.02.071

Chen, Y., Liu, Q., and Guo, D. (2020). Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol.* 92, 2249. doi:10.1002/jmv.26234

Choudhury, A., Das, N. C., Patra, R., and Mukherjee, S. (2021). *In silico* analyses on the comparative sensing of SARS-CoV-2 mRNA by the intracellular TLRs of humans. *J. Med. Virol.* 93, 2476–2486. doi:10.1002/jmv.26776

Choudhury, A., Mukherjee, G., and Mukherjee, S. (2021). Chemotherapy vs. Immunotherapy in combating nCOVID19: An update. *Hum. Immunol.* 82, 649–658. doi:10.1016/j.humimm.2021.05.001

Choudhury, A., and Mukherjee, S. (2020). *In silico* studies on the comparative characterization of the interactions of SARS-CoV-2 spike glycoprotein with ACE-2 receptor homologs and human TLRs. *J. Med. Virol.* 92, 2105–2113. doi:10.1002/jmv.25987

Choudhury, A., and Mukherjee, S. (2021). Taming the storm in the heart: Exploring different therapeutic choices against myocardial inflammation in COVID-19. *Recent Adv. antiinfect. Drug Discov.* 16, 89–93. doi:10.2174/2772434416666210616124505

Choudhury, A., Sen Gupta, P. S., Panda, S. K., Rana, M. K., and Mukherjee, S. (2022). Designing AbhiSCoVac - a single potential vaccine for all 'corona culprits': Immunoinformatics and immune simulation approaches. *J. Mol. Liq.* 351, 118633. doi:10.1016/j.molliq.2022.118633

Clark, A. E., Wang, Z., Ostman, E., Zheng, H., Yao, H., Cantarel, B., et al. (2022). Multiplex fragment analysis for flexible detection of all SARS-CoV-2 variants of concern. *Clin. Chem.* 68, 1042–1052. doi:10.1093/clinchem/hvac081

Corman, V. M., Ithete, N. L., Richards, L. R., Schoeman, M. C., Preiser, W., Drosten, C., et al. (2014). Rooting the phylogenetic tree of Middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat. *J. Virol.* 88, 11297–11303. doi:10.1128/JVI.01498-14

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020). The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5, 536–544. doi:10.1038/s41564-020-0695-z

Cui, J., Li, F., and Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17, 181–192. doi:10.1038/s41579-018-0118-9

Das, N. C., Chakraborty, P., Bayry, J., and Mukherjee, S. (2021). *In silico* analyses on the comparative potential of therapeutic human monoclonal antibodies against newly emerged SARS-CoV-2 variants bearing mutant spike protein. *Front. Immunol.* 12, 782506. doi:10.3389/fimmu.2021.782506

Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., et al. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372, eabg3055. doi:10.1126/science.abg3055

Del Rio, C., and Malani, P. N. (2022). COVID-19 in 2022-the beginning of the end or the end of the beginning? *JAMA* 327, 2389–2390. doi:10.1001/jama.2022.9655

Desingu, P. A., and Nagarajan, K. (2022). Omicron BA.2 lineage spreads in clusters and is concentrated in Denmark. *J. Med. Virol.* 94, 2360. doi:10.1002/jmv.27659

Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., and Jiang, S. (2009). The spike protein of SARS-CoV--a target for vaccine and therapeutic development. *Nat. Rev. Microbiol.* 7, 226–236. doi:10.1038/nrmicro2090

Eskier, D., Suner, A., Oktay, Y., and Karakulah, G. (2020). Mutations of SARS-CoV-2 nsp14 exhibit strong association with increased genome-wide mutation load. *PeerJ* 8, e10181. doi:10.7717/peerj.10181

Fung, T. S., and Liu, D. X. (2019). Human coronavirus: Host-pathogen interaction. *Annu. Rev. Microbiol.* 73, 529–557. doi:10.1146/annurev-micro-020518-115759

Galloway, S. E., Paul, P., MacCannell, D. R., Johansson, M. A., Brooks, J. T., MacNeil, A., et al. (2021). Emergence of SARS-CoV-2 B.1.1.7 lineage - United States, december 29, 2020-january 12, 2021. *MMWR. Morb. Mortal. Wkly. Rep.* 70, 95–99. doi:10.15585/mmwr.mm7003e2

Gao, R., Cao, B., Hu, Y., Feng, Z., Wang, D., Hu, W., et al. (2013). Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* 368, 1888–1897. doi:10.1056/NEJMoa1304459

Ge, H., Wang, X., Yuan, X., Xiao, G., Wang, C., Deng, T., et al. (2020). The epidemiology and clinical information about COVID-19. *Eur. J. Clin. Microbiol. Infect. Dis.* 39, 1011–1019. doi:10.1007/s10096-020-03874-z

Gomez, C. E., Perdiguero, B., and Esteban, M. (2021). Emerging SARS-CoV-2 variants and impact in global vaccination programs against SARS-CoV-2/COVID-19. *Vaccines (Basel)* 9, 243. doi:10.3390/vaccines9030243

Gribble, J., Stevens, L. J., Agostini, M. L., Anderson-Daniels, J., Chappell, J. D., Lu, X., et al. (2021). The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog.* 17, e1009226. doi:10.1371/journal.ppat.1009226

Grubaugh, N. D., Petrone, M. E., and Holmes, E. C. (2020). We shouldn't worry when a virus mutates during disease outbreaks. *Nat. Microbiol.* 5, 529–530. doi:10.1038/s41564-020-0690-4

Hoffmann, M., Arora, P., Gross, R., Seidel, A., Hornich, B. F., Hahn, A. S., et al. (2021). SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. *Cell* 184, 2384–2393 e12. doi:10.1016/j.cell.2021.03.036

Hozhabri, H., Piceci Sparascio, F., Sohrabi, H., Mousavifar, L., Roy, R., Scribano, D., et al. (2020). The global emergency of novel coronavirus (SARS-CoV-2): An update of the current status and forecasting. *Int. J. Environ. Res. Public Health* 17, E5648. doi:10.3390/ijerph17165648

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497–506. doi:10.1016/S0140-6736(20)30183-5

Huang, Y., Yang, C., Xu, X. F., Xu, W., and Liu, S. W. (2020). Structural and functional properties of SARS-CoV-2 spike protein: Potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* 41, 1141–1149. doi:10.1038/s41401-020-0485-4

Hurst, K. R., Koetzner, C. A., and Masters, P. S. (2009). Identification of *in vivo*-interacting domains of the murine coronavirus nucleocapsid protein. *J. Virol.* 83, 7221–7234. doi:10.1128/JVI.00440-09

Hussain, I., Pervaiz, N., Khan, A., Saleem, S., Shireen, H., Wei, D. Q., et al. (2020). Evolutionary and structural analysis of SARS-CoV-2 specific evasion of host immunity. *Genes Immun.* 21, 409–419. doi:10.1038/s41435-020-00120-6

Iketani, S., Liu, L., Guo, Y., Chan, J. F. W., Huang, Y., et al. (2022). Antibody evasion properties of SARS-CoV-2 Omicron sublineages. *Nature* 604, 553–556. doi:10.1038/s41586-022-04594-4

Ishimaru, D., Plant, E. P., Sims, A. C., Yount, B. L., Roth, B. M., Eldho, N. V., et al. (2013). RNA dimerization plays a role in ribosomal frameshifting of the SARS coronavirus. *Nucleic Acids Res.* 41, 2594–2608. doi:10.1093/nar/gks1361

Johnson, B. A., Xie, X., Kalveram, B., Lokugamage, K. G., Muruato, A., Zou, J., et al. (2020). Furin cleavage site is key to SARS-CoV-2 pathogenesis. Version 1. bioRxiv. doi:10.1101/2020.08.26.268854

Kabinger, F., Stiller, C., Schmitzova, J., Dienemann, C., Kokic, G., Hillen, H. S., et al. (2021). Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nat. Struct. Mol. Biol.* 28, 740–746. doi:10.1038/s41594-021-00651-0

Kanjilal, S., and Tang, Y. W. (2022). A tale of trial and triumph: Molecular diagnostics for severe acute respiratory coronavirus 2 over the first two years of the coronavirus disease 2019 pandemic. *Editorial Clin. Lab. Med.* 42 (2). xiii-xv. doi:10.1016/j.cll.2022.04.001

Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell* 181, 914–921. doi:10.1016/j.cell.2020.04.011

Kokic, G., Hillen, H. S., Tegunov, D., Dienemann, C., Seitz, F., Schmitzova, J., et al. (2021). Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat. Commun.* 12, 279. doi:10.1038/s41467-020-20542-0

Koppisetti, R. K., Fulcher, Y. G., and Van Doren, S. R. (2021). Fusion peptide of SARS-CoV-2 spike rearranges into a wedge inserted in bilayered micelles. *J. Am. Chem. Soc.* 143, 13205–13211. doi:10.1021/jacs.1c05435

Kurt-Jones, E. A., Popova, L., Kwinn, L., Haynes, L. M., Jones, L. P., Tripp, R. A., et al. (2000). Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus. *Nat. Immunol.* 1, 398–401. doi:10.1038/80833

Lauring, A. S., and Hodcroft, E. B. (2021). Genetic variants of SARS-CoV-2-what do they mean? *JAMA* 325, 529–531. doi:10.1001/jama.2020.27124

Liu, S., Selvaraj, P., Lien, C. Z., Nunez, I. A., Wu, W. W., Chou, C. K., et al. (2021). The PRRA insert at the S1/S2 site modulates cellular tropism of SARS-CoV-2 and ACE2 usage by the closely related Bat RaTG13. *J. Virol.* 95, JVI.01751–20. doi:10.1128/JVI.01751-20

Lopez Bernal, J., Andrews, N., Gower, C., Gallagher, E., Simmons, R., Thelwall, S., et al. (2021). Effectiveness of covid-19 vaccines against the B.1.617.2 (delta) variant. *N. Engl. J. Med.* 385, 585–594. doi:10.1056/NEJMoa2108891

Lu, G., Wang, Q., and Gao, G. F. (2015). Bat-to-human: Spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol.* 23, 468–478. doi:10.1016/j.tim.2015.06.003

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet* 395, 565–574. doi:10.1016/S0140-6736(20)30251-8

Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., et al. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc. Natl. Acad. Sci. U. S. A.* 112, 9436–9441. doi:10.1073/pnas.1508686112

Majumdar, S., and Sarkar, R. (2022). Mutational and phylogenetic analyses of the two lineages of the Omicron variant. *J. Med. Virol.* 94, 1777–1779. doi:10.1002/jmv.27558

Masters, P. S. (2006). The molecular biology of coronaviruses. *Adv. Virus Res.* 66, 193–292. doi:10.1016/S0065-3527(06)66005-3

Michel, C. J., Mayer, C., Poch, O., and Thompson, J. D. (2020). Characterization of accessory genes in coronavirus genomes. *Virol. J.* 17, 131. doi:10.1186/s12985-020-01402-1

Mlcochova, P., Kemp, S. A., Dhar, M. S., Papa, G., Meng, B., Ferreira, I. A. T. M., et al. (2021). SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* 599, 114–119. doi:10.1038/s41586-021-03944-y

Mohammad, A., Abubaker, J., and Al-Mulla, F. (2021). Structural modelling of SARS-CoV-2 alpha variant (B.1.1.7) suggests enhanced furin binding and infectivity. *Virus Res.* 303, 198522. doi:10.1016/j.virusres.2021.198522

Mohammadi, E., Shafiee, F., Shahzamani, K., Ranjbar, M. M., Alibakhshi, A., Ahangarzadeh, S., et al. (2021). Novel and emerging mutations of SARS-CoV-2: Biomedical implications. *Biomed. Pharmacother.* 139, 111599. doi:10.1016/j.biopha.2021.111599

Nakagawa, K., Lokugamage, K. G., and Makino, S. (2016). Viral and cellular mRNA translation in coronavirus-infected cells. *Adv. Virus Res.* 96, 165–192. doi:10.1016/bs.aivir.2016.08.001

Nal, B., Chan, C., Kien, F., Siu, L., Tse, J., Chu, K., et al. (2005). Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E. *J. Gen. Virol.* 86, 1423–1434. doi:10.1099/vir.0.80671-0

Naqvi, A. A. T., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., et al. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim. Biophys. Acta. Mol. Basis Dis.* 1866, 165878. doi:10.1016/j.bbadis.2020.165878

Nie, C., Sahoo, A. K., Netz, R. R., Herrmann, A., Ballauff, M., and Haag, R. (2022). Charge matters: Mutations in omicron variant favor binding to cells. *Chembiochem.* 23, e202100681. doi:10.1002/cbic.202100681

Ogando, N. S., Zevenhoven-Dobbe, J. C., van der Meer, Y., Bredenbeek, P. J., Posthuma, C. C., and Snijder, E. J. (2020). The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *J. Virol.* 94, 012466–e1320. doi:10.1128/JVI.01246-20

Paraskevis, D., Kostaki, E. G., Magiorkinis, G., Panayiotakopoulos, G., Sourvinos, G., and Tsiodras, S. (2020). Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infect. Genet. Evol.* 79, 104212. doi:10.1016/j.meegid.2020.104212

Patra, R., Chandra Das, N., and Mukherjee, S. (2021). Targeting human TLRs to combat COVID-19: A solution? *J. Med. Virol.* 93, 615–617. doi:10.1002/jmv.26387

Paules, C. I., Marston, H. D., and Fauci, A. S. (2020). Coronavirus infections-more than just the common cold. *JAMA* 323, 707–708. doi:10.1001/jama.2020.0757

Peacock, T. P., Goldhill, D. H., Zhou, J., Baillon, L., Frise, R., Swann, O. C., et al. (2021). The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat. Microbiol.* 6, 899–909. doi:10.1038/s41564-021-00908-w

Petersen, E., Koopmans, M., Go, U., Hamer, D. H., Petrosillo, N., Castelli, F., et al. (2020). Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet. Infect. Dis.* 20, e238–e244. doi:10.1016/S1473-3099(20)30484-9

Price, A. M., Olson, S. M., Newhams, M. M., Halasa, N. B., Boom, J. A., Sahni, L. C., et al. (2022). BNT162b2 protection against the omicron variant in children and adolescents. *N. Engl. J. Med.* 386, 1899–1909. doi:10.1056/NEJMoa2202826

Rabaan, A. A., Al-Ahmed, S. H., Haque, S., Sah, R., Tiwari, R., Malik, Y. S., et al. (2020). SARS-CoV-2, SARS-CoV, and MERS-COV: A comparative overview. *Infez. Med.* 28, 174–184.

Rahimi, F., and Talebi Bezmin Abadi, A. (2022). The Omicron subvariant BA.2: Birth of a new challenge during the COVID-19 pandemic. *Int. J. Surg.* 99, 106261. doi:10.1016/j.ijsu.2022.106261

Rahman, M. S., Hoque, M. N., Islam, M. R., Islam, I., Mishu, I. D., Rahaman, M. M., et al. (2021). Mutational insights into the envelope protein of SARS-CoV-2. *Gene Rep.* 22, 100997. doi:10.1016/j.genrep.2020.100997

Reina, J., and Iglesias, C. (2022). Nirmatrelvir plus ritonavir (Paxlovid) a potent SARS-CoV-2 3CLpro protease inhibitor combination. *Rev. Esp. Quimioter.* 35, 236–240. doi:10.37201/req/002.2022

Ren, S. Y., Wang, W. B., Gao, R. D., and Zhou, A. M. (2022). Omicron variant (B.1.1.529) of SARS-CoV-2: Mutation, infectivity, transmission, and vaccine resistance. *World J. Clin. Cases* 10, 1–11. doi:10.12998/wjcc.v10.i1.1

Resende, P. C., Bezerra, J. F., Teixeira Vasconcelos, R. H., Arantes, I., Appolinario, L., Mendonca, A. C., et al. (2021). Severe acute respiratory syndrome coronavirus 2 P.2 lineage associated with reinfection case, Brazil, june-october 2020. *Emerg. Infect. Dis.* 27, 1789–1794. doi:10.3201/eid2707.210401

Romano, M., Ruggiero, A., Squeglia, F., Maga, G., and Berisio, R. (2020). A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells* 9, E1267. doi:10.3390/cells9051267

Shrestha, L. B., Foster, C., Rawlinson, W., Tedla, N., and Bull, R. A. (2022). Evolution of the SARS-CoV-2 omicron variants BA.1 to BA.5: Implications for immune escape and transmission. *Rev. Med. Virol.* 32, e2381. doi:10.1002/rmv.2381

Shrestha, L. B., Tedla, N., and Bull, R. A. (2021). Broadly-neutralizing antibodies against emerging SARS-CoV-2 variants. *Front. Immunol.* 12, 752003. doi:10.3389/fimmu.2021.752003

Sibai, M., Wang, H., Yeung, P. S., Sahoo, M. K., Solis, D., Mfuh, K. O., et al. (2022). Development and evaluation of an RT-qPCR for the identification of the SARS-CoV-2 Omicron variant. *J. Clin. Virol.* 148, 105101. doi:10.1016/j.jcv.2022.105101

Singh, D., and Yi, S. V. (2021). On the origin and evolution of SARS-CoV-2. *Exp. Mol. Med.* 53, 537–547. doi:10.1038/s12276-021-00604-z

Singh, J., Rahman, S. A., Ehtesham, N. Z., Hira, S., and Hasnain, S. E. (2021). SARS-CoV-2 variants of concern are emerging in India. *Nat. Med.* 27, 1131–1133. doi:10.1038/s41591-021-01397-4

Siu, Y. L., Teoh, K. T., Lo, J., Chan, C. M., KienF.EscriouN., et al. (2008). The M, E, and N structural proteins of the severe acute respiratory syndrome coronavirus are required for efficient assembly, trafficking, and release of virus-like particles. *J. Virol.* 82, 11318–11330. doi:10.1128/JVI.01052-08

Song, Z., Xu, Y., Bao, L., Zhang, L., Yu, P., Qu, Y., et al. (2019). From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses* 11, E59. doi:10.3390/v11010059

Teng, S., Sobitan, A., Rhoades, R., Liu, D., and Tang, Q. (2021). Systemic effects of missense mutations on SARS-CoV-2 spike glycoprotein stability and receptor-binding affinity. *Brief. Bioinform.* 22, 1239–1253. doi:10.1093/bib/bbaa233

Thakur, S., Sasi, S., Pillai, S. G., Nag, A., Shukla, D., Singhal, R., et al. (2022). SARS-CoV-2 mutations and their impact on diagnostics, therapeutics and vaccines. *Front. Med.* 9, 815389. doi:10.3389/fmed.2022.815389

Tuekprakhon, A., Nutalai, R., Dijokaite-Guraliuc, A., Zhou, D., Ginn, H. M., Selvaraj, M., et al. (2022). Antibody escape of SARS-CoV-2 Omicron BA.4 and BA.5 from vaccine and BA.1 serum. *Cell* 185, 2422–2433 e13. doi:10.1016/j.cell.2022.06.005

Vanaerschot, M., Mann, S. A., Webber, J. T., Kamm, J., Bell, S. M., Bell, J., et al. (2020). Identification of a polymorphism in the N gene of SARS-CoV-2 that adversely impacts detection by reverse transcription-PCR. *J. Clin. Microbiol.* 59, 023699–e2420. doi:10.1128/JCM.02369-20

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 183, 281–292. doi:10.1016/j.cell.2020.02.058

Wan, Y., Shang, J., Graham, R., Baric, R. S., and Li, F. (2020). Receptor recognition by the novel coronavirus from wuhan: An analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* 94, 001277–e220. doi:10.1128/JVI.00127-20

Wang, H., Jean, S., Eltringham, R., Madison, J., Snyder, P., Tu, H., et al. (2021). Mutation-specific SARS-CoV-2 PCR screen: Rapid and accurate detection of variants of concern and the identification of a newly emerging variant with spike L452R mutation. *J. Clin. Microbiol.* 59, e0092621. doi:10.1128/JCM.00926-21

Wang, H., Miller, J. A., Verghese, M., Sibai, M., Solis, D., Mfuh, K. O., et al. (2021). Multiplex SARS-CoV-2 genotyping reverse transcriptase PCR for population-level variant screening and epidemiologic surveillance. *J. Clin. Microbiol.* 59, e0085921. doi:10.1128/JCM.00859-21

Wang, P., Casner, R. G., Nair, M. S., Wang, M., Yu, J., Liu, L., et al. (2021). Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. *Cell Host Microbe* 29, 747. bioRxiv. doi:10.1016/j.chom.2021.04.007

Wang, P., Nair, M. S., Liu, L., Iketani, S., Luo, Y., Guo, Y., et al. (2021). Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* 593, 130–135. doi:10.1038/s41586-021-03398-2

Wang, Q., Chen, H., Shi, Y., Hughes, A. C., Liu, W. J., Jiang, J., et al. (2021). Tracing the origins of SARS-CoV-2: Lessons learned from the past. *Cell Res.* 31, 1139–1141. doi:10.1038/s41422-021-00575-w

Wang, R., Hozumi, Y., Yin, C., and Wei, G. W. (2020). Mutations on COVID-19 diagnostic targets. *Genomics* 112, 5204–5213. doi:10.1016/j.ygeno.2020.09.028

Weisblum, Y., Schmidt, F., Zhang, F., DaSilva, J., Poston, D., Lorenzi, J. C., et al. (2020). Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* 9, e61312. doi:10.7554/eLife.61312

Wolter, N., Jassat, W., Walaza, S., Welch, R., Moultrie, H., Groome, M., et al. (2022). Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in south Africa: A data linkage study. *Lancet* 399, 437–446. doi:10.1016/S0140-6736(22)00017-4

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263. doi:10.1126/science.abb2507

Wrobel, A. G., Benton, D. J., Xu, P., Roustan, C., Martin, S. R., Rosenthal, P. B., et al. (2020). SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* 27, 763–767. doi:10.1038/s41594-020-0468-7

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., et al. (2020). Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* 27, 325–328. doi:10.1016/j.chom.2020.02.001

Wu, K., Werner, A. P., Moliva, J. I., Koch, M., Choi, A., Stewart-Jones, G. B. E., et al. (2021). mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. bioRxiv. doi:10.1101/2021.01.25.427948

Wu, Z., and McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72314 cases from the Chinese center for disease control and prevention. *JAMA* 323, 1239–1242. doi:10.1001/jama.2020.2648

Xia, S., Zhu, Y., Liu, M., Lan, Q., Xu, W., Wu, Y., et al. (2020). Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell. Mol. Immunol.* 17, 765–767. doi:10.1038/s41423-020-0374-2

Xie, X., Zou, J., Fontes-Garfias, C. R., Xia, H., Swanson, K. A., Cutler, M., et al. (2021). Neutralization of N501Y mutant SARS-CoV-2 by BNT162b2 vaccine-elicited sera. Version 1. bioRxiv. doi:10.1101/2021.01.07.425740

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367, 1444–1448. doi:10.1126/science.abb2762

Yang, J., Petitjean, S. J. L., Koehler, M., Zhang, Q., Dumitru, A. C., Chen, W., et al. (2020). Molecular interaction and inhibition of SARS-CoV-2 binding to the ACE2 receptor. *Nat. Commun.* 11, 4541. doi:10.1038/s41467-020-18319-6

Yoshimoto, F. K. (2020). The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *Protein J.* 39, 198–216. doi:10.1007/s10930-020-09901-4

Zarocostas, J. (2020). What next for the coronavirus response? *Lancet* 395, 401. doi:10.1016/S0140-6736(20)30292-0

Zhan, Y., Yin, H., and Yin, J. Y. B.1 (2022). B.1.617.2 (delta) variant of SARS-CoV-2: Features, transmission and potential strategies. *Int. J. Biol. Sci.* 18, 1844–1851. doi:10.7150/ijbs.66881

Zhang, Q., Xiang, R., Huo, S., Zhou, Y., Jiang, S., Wang, Q., et al. (2021). Molecular mechanism of interaction between SARS-CoV-2 and host cells and interventional therapy. *Signal Transduct. Target. Ther.* 6, 233. doi:10.1038/s41392-021-00653-w

Zhao, J., Cui, W., and Tian, B. P. (2020). The potential intermediate hosts for SARS-CoV-2. *Front. Microbiol.* 11, 580137. doi:10.3389/fmicb.2020.580137

Zheng, M., Karki, R., Williams, E. P., Yang, D., Fitzpatrick, E., Vogel, P., et al. (2021). TLR2 senses the SARS-CoV-2 envelope protein to produce inflammatory cytokines. *Nat. Immunol.* 22, 829–838. doi:10.1038/s41590-021-00937-x

Strategies to improve safety profile of AAV vectors

Tuisku Suoranta[1], Nihay Laham-Karam[1]* and
Seppo Ylä-Herttuala[1,2,3]*

[1]A. I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland, [2]Heart Center, Kuopio University Hospital, Kuopio, Finland, [3]Gene Therapy Unit, Kuopio University Hospital, Kuopio, Finland

Adeno-associated virus (AAV) vectors are currently used in four approved gene therapies for Leber congenital amaurosis (Luxturna), spinal muscular atrophy (Zolgensma), aromatic L-amino acid decarboxylase deficiency (Upstaza) and Haemophilia A (Roctavian), with several more therapies being investigated in clinical trials. AAV gene therapy has long been considered extremely safe both in the context of immunotoxicity and genotoxicity, but recent tragic deaths in the clinical trials for X-linked myotubular myopathy and Duchenne's muscular dystrophy, together with increasing reports of potential hepatic oncogenicity in animal models have prompted re-evaluation of how much trust we can place on the safety of AAV gene therapy, especially at high doses. In this review we cover genome and capsid engineering strategies that can be used to improve safety of the next generation AAV vectors both in the context of immunogenicity and genotoxicity and discuss the gaps that need filling in our current knowledge about AAV vectors.

Introduction

The last 10 years have been full of great successes for gene therapy and a lot of this success can be attributed to a small virus named Adeno-associated virus (AAV). It was the delivery method of choice for Glybera, the first gene therapy approved by the European Medicines Agency (EMA) in 2012 (Ylä-Herttuala, 2012). Since then, two more *in vivo* AAV gene therapies have been approved by EMA and the United States Food and Drug Administration (FDA): Luxturna and Zolgensma (Keeler and Flotte, 2019), with Upstaza for aromatic L-amino acid decarboxylase deficiency and Roctavian for Haemophilia A having also secured marketing authorisations from EMA (BioMarin Investors, 2022; PTC Therapeutics, 2022).

AAV is a defective parvovirus with a single stranded DNA genome packaged inside a non-enveloped capsid. It is unable to replicate without the presence of a helper virus such as adenovirus or herpes simplex virus (HSV), and instead, in their absence, establishes a latent infection by integrating site specifically into the so-called AAV safe harbour (AAVS1) in 19q13.3 (Atchison et al., 1965; Kotin et al., 1991; Weindler and Heilbronn, 1991). The AAV genome itself contains the bare necessities packed into

FIGURE 1
Schematic representation of wild-type AAV2, its encapsulated genome and the known open reading frames. ITR: inverted terminal repeat. RBE: Rep binding element, trs: terminal resolution site D: D-sequence, P5, P19, P40, and P81: viral promoters, Rep78, Rep68, Rep52, and Rep40: proteins coded by rep, involved in genome replication and packaging. VP1, VP2, and VP3: structural capsid proteins coded by cap, MAAP: membrane associated accessory protein, AAP: assembly activating protein, X: protein X, polyA: viral polyadenylation signal.

4.7 kb (Figure 1). The 145 nt inverted terminal repeats (ITRs) flank both ends of the genome, containing the signals necessary for genome replication and packaging; they are also the only signal required for AAV vectors *in cis*. Between the ITRs reside two genes: the *rep* and the *cap*. For AAV serotype 2 (AAV2) the *rep* codes for four different replicase proteins: Rep78/68, which are responsible for genome replication, and Rep52/40, responsible for genome packaging (King et al., 2001; Stracker et al., 2004). The *cap* codes for three structural proteins: VP1, VP2, and VP3, which form the 22 nm diameter AAV capsid in a 1:1:10 ratio. Three additional proteins have been identified within the AAV2 *cap:* assembly activating protein (AAP), membrane associated accessory protein (MAAP) and protein X (Sonntag et al., 2010; Cao et al., 2014; Ogden et al., 2019; Galibert et al., 2021). The basic structure is the same for all known serotypes, though some differences exist; for example, AAV5 *rep* does not code for Rep68 at all (Fasina and Pintel, 2013).

The natural AAV infection is not associated with any confirmed pathology and is most often established in the liver and bone marrow (Gao et al., 2004). However, the many different serotypes discovered in humans and non-human primates (NHPs) have differing receptor usage, and thus target organs and cell types at varying efficiencies (Wu et al., 2006; Huang et al., 2014). This has been utilised as an advantage in gene therapy since a specific serotype can be selected to optimise transduction of different target tissues. Other attributes also vary–for example AAV9 can cross the blood-brain barrier, while AAV2 cannot (Liu

et al., 2021). Conveniently, AAV is also capable of process called cross-packaging, which allows use of AAV2 ITRs in the vector genome, together with AAV2 *rep* provided *in trans,* to be packaged into the capsid of other serotypes, making switching between capsids relatively effortless (Rabinowitz et al., 2002).

AAV vectors possess many attributes that have made it the delivery vector of choice for *in vivo* gene therapy. In addition to the variety of attributes dictated by the naturally occurring serotypes, the repertoire has been further expanded by engineered capsid variants (Büning et al., 2015). The safety profile also appeared excellent based on the animal studies and early clinical trials. The immunogenicity was found to be relatively low, and the risk of genotoxicity minimal, as the gutted vector showed no active integration (Salmon et al., 2014). Yet, despite the lack of integration, the expression is not transient, as the genomes can persist in an episomal form and continue to produce the transgene even after 10 years (Buchlis et al., 2012).

With more trials now conducted and the demand soaring it has become clear that AAV gene therapy also faces challenges. Firstly, the manufacturing capacity that was sufficient for ultra-rare orphan diseases could not support the pipeline for more common disorders. This has meant moving away from production in monocultures to bioreactors and from ultracentrifugation-based purification to high-throughput affinity chromatography. Though these technologies have quickly evolved, the race to meet the increasing industry demands is ever ongoing (Dobrowsky et al., 2021). The

analytical requirements too are far from simple–the differences to small molecules and simpler biological products means that the regulatory demands have evolved in parallel with the therapies, yet many processes remain unstandardised (Ramsey et al., 2021). Even determining the dose that the patient is receiving can be tricky, as the measured titres can vary up to ten-fold between different laboratories (Lock et al., 2010; Ayuso et al., 2014).

Although some of the technical challenges have been resolved, alarming new concerns around safety have emerged from the exponentially increasing number of studies and trials. Death of four patients in the Astellas-Audentes trial for X-linked myotubular myopathy (XLMTM), one in Pfizer's trial for Duchenne's Muscular Dystrophy (DMD) and the death of one patient administrated with Zolgensma has called into question the high doses used and emphasised the need for AAV vectors with a better efficacy at lower doses. Additional concerns around acute toxicity have arisen from animal studies, with neuropathology and toxicity reported in non-human primates (NHPs) and piglets treated with a high dose of AAV (Hinderer et al., 2018). Furthermore, lesions in dorsal root ganglia were found in the majority of AAV dosed NHPs (Hordeaux et al., 2020). Long term safety profile has also come under scrutiny due to the reports of hepatocellular carcinoma in mice, bile-duct proliferation in rabbits and clonal expansion of transduced hepatocytes in a canine model of Haemophilia A (e.g., Donsante et al., 2001; Bell et al., 2006; Hytönen et al., 2019; Dalwadi et al., 2021b; Li et al., 2021; Nguyen et al., 2021; reviewed in Sabatino et al., 2022). Furthermore, a recent outbreak of hepatitis in Scotland was linked to a wild type AAV2 infection, though the exact connection to the pathophysiology observed remains unclear (Ho et al., 2022).

Safety is of paramount importance for any therapeutic intervention, and the AAV gene therapy field must strive to find–and implement–solutions that address the questions being asked as fully as possible. Here, we have reviewed genome and capsid engineering strategies that can be used to improve the safety profile by promoting immune-evasion, avoiding oncogene activation, and increasing on-target delivery.

Engineering AAV vectors for better safety profile

Many approaches have been taken to engineer AAV vectors. At the vector genome level, this means adding, mutating and deleting sequences; for example, the self-complementary AAV (scAAV) vectors were designed by deleting key signals from the second ITR, causing the genome replication to continue to copy also the second strand instead of termination (McCarty et al., 2003). Consequently, scAAV vectors are no longer dependent on the second strand synthesis, which is a major rate limiting step in the AAV transduction pathway (Ferrari et al., 1996). Other well-

known strategies include codon optimisation of the transgene and the choice and manipulation of elements such as the promoter or polyA sequence.

Likewise, different strategies have been used for capsid engineering. These can be loosely divided into four categories: directed evolution, rational design, computer guided design and combinations thereof. Commonly, error-prone PCR is used to generate capsid mutant libraries or peptides with known or speculated affinities are inserted into the capsid. The intricacies of these methods are summarised in Figure 2 and reviewed elsewhere in more detail (Buchholz et al., 2015; Li and Samulski, 2020; Zolotukhin and Vandenberghe, 2022).

Circumventing innate and adaptive immunity

Though AAVs have traditionally been regarded as having low immunogenicity from the safety point of view, we know that this is not the whole story. Two out of the three severe adverse events reported earlier in the Pfizer DMD trial and the patient who died after Zolgensma administration showed signs of atypical haemolytic uremic syndrome (aHUS)-like complement activation, highlighting that the relatively low immunogenicity can regardless become a serious safety concern (Guillou et al., 2022; Pfizer, 2022). Severe innate immunity related toxicity had also been previously reported after a high-dose AAV-PHP.B i.v administration in one NHP (Hordeaux et al., 2018). In contrast, the challenges posed by neutralising antibodies and cytotoxic T cell responses for efficacy have long been acknowledged (Manno et al., 2006; Vandamme et al., 2017; Korpela et al., 2022). However, while for example the lysis of transduced cells by capsid-specific cytotoxic T cells might not be life-threatening when the transduction levels are low, this is likely to be different when the majority of the target organ is transduced. Additional complications arise if an immune response is mounted against the transgene, which not only will reduce the therapeutic efficacy but may also compromise any protein replacement therapy that the patient has previously relied upon.

Unmethylated CpG motifs in microbial DNA belonging to the Pathogen associated molecular pattern (PAMP), can be sensed by the intracellular innate sensor Toll-like receptor 9 (TLR9). Unmethylated CpG DNA has also been linked to complement activation, which seems to be mediated via both TLR9 dependent and independent mechanisms (Mangsbo et al., 2009). CpG motifs are often unmethylated in AAV vectors, despite the production in mammalian cells, and the activation of the TLR9-MyD88 signalling pathway by unmethylated CpG has been shown to promote adaptive immune responses not only against the capsid but also the transgene (Zhu et al., 2009). Studies in mice have shown that modifying the transgene sequence to eliminate CpG can lead to increased transgene

FIGURE 2
Summary of the most common capsid engineering strategies. PCR; polymerase chain reaction, DARPin: designed ankyrin repeat protein.

FIGURE 3
CpG motifs (in black) and the optimal mouse TLR9 recognition motif (highlighted) found in the AAV2 ITR sequence. ITR: inverted terminal repeat. Flop: hairpin sequence orientation, C, C', B and B': hairpin elements, RBE: Rep binding element, trs: terminal resolution site. Figure generated with SnapGene.

expression and persistence, likely specifically due to the reduced cytotoxic T-cell responses (Faust et al., 2013; Bertolini et al., 2021). Similar results were also obtained by the incorporation of TLR9 oligonucleotide antagonist sequence directly into the AAV vector genome (Chan et al., 2021). In addition to the transgene, CpG motifs are also present on various other vector elements, such as the widely used cytomegalovirus (CMV) enhancer and promoter, so aiming to minimise the amount of CpG motifs in the whole genome could offer further benefit. The ITRs of

AAV2 alone contain 16 CpG motifs each, which may contribute to TLR9-sensing. In their recent study (Pan et al., 2022) successfully modified the ITRs to eliminate all the CpG motifs, though this came at the cost of 3-fold reduced productivity due to reduced genome replication. The effect on immune activation was also not confirmed. Indeed, although TLR9 mediated immunity against the AAV vector genome was hypothesised to have played a part in the loss of expression seen in the BAX 335 Haemophilia B trial (Konkle et al., 2021), the

relevance of TLR9 mediated CpG sensing to AAV immunogenicity in humans is not yet clear. There are known differences between human and mouse TLR9 sensing that should be considered: the expression pattern of TLR9 is much more restricted in humans, and the optimal recognition motif in humans (5′-TCGTW-3′) is markedly different from that of mice (5′-RRCGYY-3′) (Huang and Yang, 2010). The AAV2 ITRs, for example, contain one optimal TLR9 recognition motif for mice, but none for humans (Figure 3). Different formulas for estimating the risk of TLR9 activation by vector genome sequences have been proposed, with retrospective analysis of clinical trial data supporting their usefulness (Wright, 2020). Two other factors also contribute to TLR9 sensing: DNA structure and the dose. It is known that self-complementary AAV vectors are recognised by TLR9 more strongly than their single stranded counterparts, but due to their improved efficacy scAAVs may also allow for the use of a lower dosage, potentially compensating for this difference (Martino et al., 2011).

TLR9 is not the only innate immunity sensor recognising AAV vectors–the dsRNA sensor MDA5 has also been implicated (Shao et al., 2018). This seems to be mediated by antisense transcriptional activity from the sequences within and near the ITRs; one study mapped transcriptional initiation at nucleotides 109 to 145 of the ITR while another identified a binding site for the transcription factor HNF1-α just beyond the D-sequence (Haberman et al., 2000; Logan et al., 2017). As the HNF1-α binding site is outside the actual ITRs it can be safely removed, but it is currently present on some vector plasmids, such as the traditional pSub201. The initiation from within the ITR is trickier as the implicated region contains many important elements, such as the Rep binding element (RBE), terminal resolution site (trs), and the D-sequence, required for packaging (Figure 3). However, adding a polyA element facing away from the 3′-ITR, at the opposite orientation to the transgene, can be used to halt any antisense transcription (Shao et al., 2018), though at the cost of space in the vector.

Many of the strategies tested to avoid immune activation have centred on capsid modifications. For now, activation of a third innate immunity sensor, TLR2, which recognises AAV capsid, is not well enough understood to be circumvented by engineering approaches. Instead, several strategies focus on avoiding the adaptive immune responses. Crucially, adaptive T-cell responses rely on antigen presentation by Major Histocompatibility Complex Class I and II molecules, and several of the major epitopes on AAV1, AAV2, and AAV8 have been mapped (Hui et al., 2015), enabling rational design approaches to modify these sequences. Prediction tools are available for both MHC Class I and II epitopes (Nielsen et al., 2010; Paul et al., 2020)—though the latter are more variable and thus harder to predict–and can be used to screen not only other naturally occurring serotypes, but can also be used in designing engineered novel variants.

MHC presentation can also be avoided *via* a second strategy: by circumventing AAV degradation in the proteasomal pathway, which leads to the generation of peptide-epitopes that can be loaded for antigen presentation. Phosphorylation of certain residues acts as a signal for ubiquitination and proteasomal targeting, so mutation studies have been carried out on the AAV2 capsid surface exposed tyrosine (Y), serine (S) and threonine (T) residues, which can be phosphorylated by different kinases (Zhong et al., 2008; Aslanidi et al., 2012; Aslanidi et al., 2013; Li et al., 2015). From the various mutants screened Y444 + 500 + 730F + T491V showed best transduction (Aslanidi et al., 2013), and though this mutant specifically was not studied in the context of T cell activation, the Y444 + 500 + 730F triple mutant transduction was confirmed to result in less MHC Class I presentation than transduction with the wild type (Martino et al., 2013). Additionally, mutations of the surface lysine (K) residues, which undergo ubiquitination, have been studied, with AAV2 K556E performing the best *in vitro* and *in vivo* hepatic gene transfer (Li et al., 2015). Interestingly this study also found that mutating the same lysine residues in AAV8 did not result in similar improvement *in vivo* as was seen for AAV2. Further studies have identified residues on AAV1, AAV3, AAV5, AAV6 and AAV9 capsids that can be mutated to similar effects, even if direct cross-application of specific mutations between the serotypes is not always possible (Cheng et al., 2012; Martino et al., 2013; Sen et al., 2013; Song et al., 2013).

Due to the adverse effects observed in the clinic we now know that complement activation may pose a serious risk to the patients (Guillou et al., 2022; Pfizer, 2022; Solid Biosciences, 2022). The complement system can be activated *via* three different pathways: classical, lectin mediated and alternative, which all ultimately promote target opsonisation, phagocytosis and increased inflammation (Sarma and Ward, 2011). Blood work from patients with adverse effects in response to AAV9 gene therapy has specifically implicated alternative pathway activation (Guillou et al., 2022; Pfizer, 2022). The alternative pathway is activated by hydrolysis of complement factor C3 in the absence of factor H (or other co-factors) interactions with factor I (Meri, 2016). AAV2 has been shown to activate the complement system, and to interact with both factor H and C3, which together with evidence of catabolism of C3b led to the conclusion that the activation happened primarily *via* the classical pathway (Zaiss et al., 2008). Interestingly, another co-immunoprecipitation study with AAV6 found no specific complement binding (Denard et al., 2013), suggesting that once again the differences between the different serotypes might be significant. As the clinical data so far has specifically implicated AAV9 in the alternative pathway activation, there is need to better understand the differences in complement activation between the serotypes and to also apply this understanding to capsid engineering approaches.

In the classical pathway the target is recognised by the binding of IgM or IgG antibodies, with their Fc regions interacting with the components of the complement system, leading to its activation, and boosting both humoral and cellular adaptive immune responses (Sarma and Ward, 2011). Antibodies also directly interact with different immune cells, promoting inflammation, or block transduction by preventing interactions between the virions and the host cell (Klasse and Sattentau, 2002). Clinical trials have so far been mainly focused on screening for these neutralising antibodies (NAbs) due to the concerns around the loss of efficacy, but while considering the safety it is important to remember that also non-neutralising antibodies interact with the immune system. It is generally difficult to avoid existing anti-capsid antibodies simply by switching a serotype, as the antibody epitopes are fairly well conserved between the different serotypes, resulting in high cross-reactivity (Boutin et al., 2010). Several studies have utilised directed evolution approaches combined with libraries generated by error-prone PCR, DNA recombination (e.g., capsid-shuffling) or random peptide ligation to generate variants that can escape NAbs (e.g., Maheshri et al., 2006; Perabo et al., 2006; Bartel et al., 2012). In these approaches the evolutionary pressure (NAbs) is added either *in vitro* or *in vivo* and the capsids that are able to efficiently escape and mediate transduction are chosen for further screening. As this method requires large amounts of screening and some luck in generating the mutants, the more precise rational approaches have also been employed. Based on 3D modelling Girod *et al.,* 1999 generated AAV2 peptide insertion mutants and screened them against different AAV antibodies: A20, C37-B, D3 and C24-B. Peptide insertions at AAV2 residues 261, 381, 534, 573, and 587 were found to decrease the binding of the different antibodies. In further testing by Huttner *et al.,* 2003 AAV2 vectors with peptide insertions in 534, 573, and 587 were able to transduce different cell lines in the presence of antibody containing human sera, while also demonstrating new tropisms based on the inserted peptide. More precise mapping of the immunogenic epitopes has been achieved by cryo-electron microscopy 3D image reconstruction (Gurda et al., 2012; Gurda et al., 2013; Emmanuel et al., 2022). Key proof of concept study was carried out with AAV8. The footprint of AAV8 NAb ADK8 was mapped to capsid residues 588–592, with mutations in this region allowing evasion of ADK8 neutralisation (Gurda et al., 2012). Furthermore, it was demonstrated that after screening of peptide insertion library at the position 590 the tropism for mouse liver was further enhanced (Raupp et al., 2012).

Minimising the risk of genotoxicity

While the wild-type AAV integrates into the genome in a site-specific manner, this activity is mediated by the large Rep proteins (Stracker et al., 2004), the sequence of which are not present in the gutted vector genomes. Thus, unlike retroviruses, recombinant AAV does not actively integrate, which led many to consider the risk of genotoxicity negligible. Although there is currently no evidence of genotoxicity from any human clinical trials, the observations of hepatocellular carcinoma in mice and clonal hepatocyte expansion in dogs have advocated for a more cautious approach (e.g Donsante et al., 2001; Bell et al., 2006; Dalwadi et al., 2021b; Li et al., 2021; Nguyen et al., 2021). The random integration has been estimated to happen at frequency of 0.001–3%—a fairly rare event by the average estimate (Nowrouzi et al., 2012; Dalwadi et al., 2021a). However, if a 10 kg patient is dosed with 1×10^{14} vg/kg, this would mean 1×10^{10} integration events even by the most conservative value. The integration logically happens mostly in the areas of open chromatin and seems to be mediated by interactions between the naturally occurring random DNA breaks in the host genome and the recruitment of host DNA repair factors to the AAV genome (Miller et al., 2004). Slightly different DNA repair proteins have been shown to interact with ssAAV and scAAV vectors (Cataldi & McCarty, 2010), however we currently do not know whether one might offer a lower genotoxicity risk than the other. Genome integrations by both ssAAV and scAAV vectors have been implicated in different studies (e.g., Donsante et al., 2007; Rosas et al., 2012), even if no direct comparisons of safety have been carried out thus far.

The major genotoxicity risk events can be roughly divided into two categories: silencing of tumour suppressor genes and oncogene activation. These can be countered by either fully avoiding integration or by controlling it, the latter being easier to achieve. In their 2012 study (Wang et al., 2012) added flanking rDNA homology arms to an AAV vector, which then exhibited highly efficient targeted integration and gene correction. Interestingly, the AAV mediated homologous recombination (HR) exhibits significantly higher efficiency than classical HR (Melo et al., 2014). However, the homology arms demand significant space, substantially reducing the already limited packaging capacity. Alternative approaches have been designed to combat this. Barzel et al. (2015) eliminated the promoter requirement by using HR to hitchhike the payload to the end of albumin gene together with a 2A sequence enabling the generation of both albumin and the payload protein from the endogenous albumin promoter. The whole cassette can also be used as a homology template, simply containing the correction(s) required in the parent gene, as was demonstrated in the correction of epidermolysis bullosa causing point mutation in the LAMA3 gene in keratinocytes (Melo et al., 2014). Combinational strategies with CRISPR-Cas9 have also been used for example *in vitro* stem cell editing and also *in vivo* (Dever et al., 2016; Yang et al., 2016). These strategies are somewhat limited in their applicability and reliance on HR, and the utilisation of the natural integration by wtAAV into the AAVS1 seems like an attractive possibility to explore.

FIGURE 4
Genome designs to prevent insertions of partial genomes without transcriptional termination signal. The risk of insertional mutagenesis can be reduced by inverting the genome of self-complementary AAVs with promoter to the mTR **(A)** or deleting the promoter-proximal D-sequence from single-stranded AAVs **(B)**.
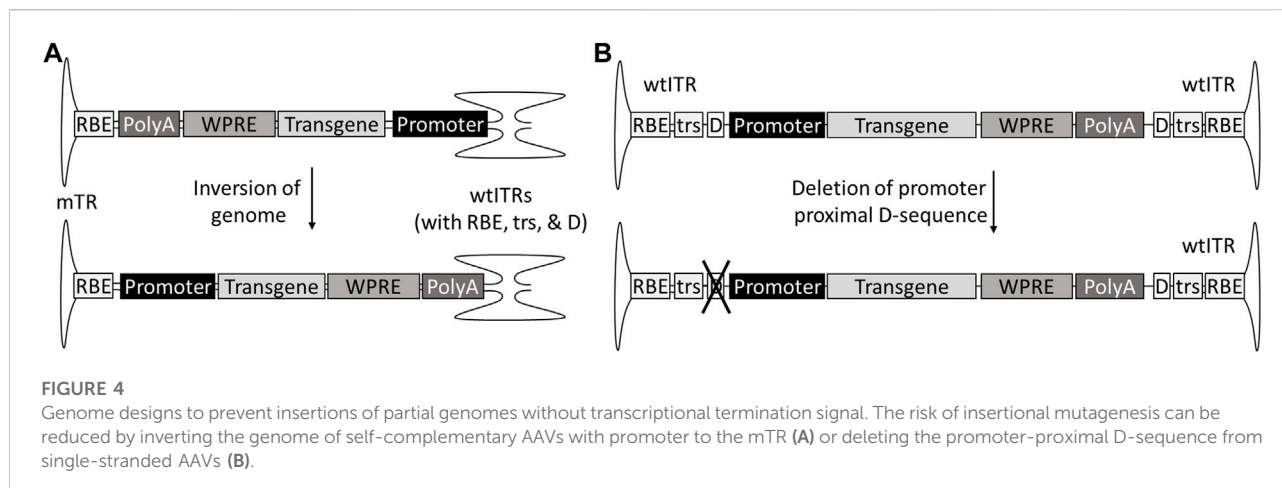
However, this requires the presence of Rep78/68, which is difficult to fit into an AAV vector. To solve this both AAV/ adenovirus and AAV/HSV hybrids have been investigated (Recchia et al., 1999; Heister et al., 2002), but as they only carry parts of the AAV genome packaged into a different viral vector, they can hardly be classified as AAVs.

A part of the genotoxicity risk has been attributed to partial genomes, containing the enhancer-promoter elements without the transcription termination signals (Zhang et al., 2021). These partial genomes can be due to so-called snap back genomes (SBGs) formed during replication when the genome loops back on itself prematurely, forming a self-complementary vector with only part of the sequence. This in turn can be caused for example by hairpin structures such as shRNAs (Xie et al., 2017). Single stranded genomes truncated at the 5′-end are also fairly common, especially in oversized constructs, as the packaging is initiated at the 3′-end ITR (Wu et al.,2010; Tran et al., 2020). For scAAVs, the generation of promoter-only SBGs can be avoided by inverting the genome so that the promoter is proximal to the mutated terminal repeat (mTR), which cannot initiate the packaging (Zhang et al., 2021). In the case of ssAAVs the solution is not as simple; AAV packages genomes of both polarities, meaning both ends of the genome can be truncated. However, this could theoretically be solved by mutating one of the D-sequences required for packaging initiation, and, similar to scAAV, placing the promoter next to the mutated ITR and transcription termination signal, such as polyA, towards the wtITR (Figure 4). As an additional benefit, the deletion of the D(+)-sequence eliminates the binding of NF-κB-repressing factor (NRF), which normally inhibits viral transgene expression (Ling et al., 2015). However, due to half of the genomes present not being available for packaging, this does come at the expense of productivity.

The choice of specific elements in the vector can also either reduce or enhance the risk of genotoxicity. Certain promoters, for example, seem to have a higher risk of genotoxicity than others. In a study of hepatocellular carcinoma in mice, chicken beta actin (CBA) and thyroxine-binding globulin (TBG) enhancer/promoter elements were associated with HCC formation, while human α-1 antitrypsin (hAAT) promoter was not. This was speculated to be due to the increased activation of nearby genes by CBA and TBG promoters (Chandler et al., 2015). It is also known that many enhancers and promoters possess bi-directional activity (Wei et al., 2011). Insulators have been used in retro- and lentiviral vectors to address this and also to minimise read-through transcription, which is another concern for integrated AAV genomes (Browning and Trobridge, 2016). A study by Fitzsimons *et al.*, 2001 also tested insulators in the context of AAV, designing a doxycycline controlled, insulator flanked cassette, which was found functional in rat brain. Indeed, such inducible promoters may also be suitable for improving safety for some therapeutic applications (Chen et al., 2013).

A commonly included element in the AAV vector genome is woodchuck hepatitis virus post-transcriptional regulatory element (WPRE), which promotes transgene expression by supporting transcriptional termination and enhancing mRNA nuclear export (Loeb et al., 1999; Higashimoto et al., 2007). The wild-type WPRE does, however, come with concerns of oncogenesis, as it contains in its beta-element verified promoter activity and the start of the WHV protein X (WHX) ORF. Truncated WHX fragments have been linked to oncogenicity in liver tumours (Wei et al., 1995; Tu et al., 2001). The WHX We1 promoter/enhancer element contained within the wt WPRE sequence has also been validated to be active in the context of vectors (Wei et al., 1995; Kingsman et al., 2005). To avoid this, mutants have been generated with comparable activity, including the so-called mutant 6, with both the promoter and start codon mutated (Zanta-Boussif et al., 2009). A shorter alternative also exists in the so-called mutant 3, which completely

TABLE 1 Genome and capsid engineering solutions for different safety aspects of AAV gene therapy.

Immunogenicity		Genotoxicity
Innate immunity	**Adaptive immunity**	
TLR9 sensing	MHC presentation	Random integration
- Reducing vector CpG content (Faust et al., 2013) - Inhibitor oligonucleotide sequences (Chan et al., 2021)	- Avoidance of proteasomal targeting by mutating key residues (Cheng et al., 2012; Aslanidi et al., 2013; Martino et al., 2013; Sen et al., 2013; Song et al., 2013) - Mutation of MHC presented peptides (Hui et al., 2015)	-Homology directed targeted integration and genome editing (Wang et al., 2012; Melo et al., 2014; Dever et al., 2016)
MDA5 sensing		Promoter/enhancer driven oncogene activation
- Eliminating HNF1-α binding site near the ITR sequence (Logan et al., 2017)	Anti-capsid antibodies	- Keeping the genome size within the wt virus size to minimise partial genome packaging (Wu et al., 2010)
- Inserting a polyA sequence to block ITR originating antisense transcripts (Shao et al., 2018)	- Elimination of antibody recognition by rational design or by directed evolution (Zaiss et al., 2008; Denard et al., 2013)	- Use of tissue specific promoters (Wang et al., 2008; Pacak et al., 2009; Xiong et al., 2019) - scAAV: inverting the genome with promoter proximal to the mTR (Zhang et al., 2021)
Complement activation	Anti-transgene responses	- ssAAV: D-sequence deletion at promoter proximal ITR (Zhang et al., 2021; Ling et al., 2015)
- Alternative pathway: choosing/engineering serotypes which will not trigger alternative pathway (Zaiss et al., 2008; Denard et al., 2013)	- Induction of liver tolerance, e.g. by using a tandem promoter (Colella et al., 2019)	
- Classical pathway: engineering to eliminate anti-capsid antibody epitopes (Maheshri et al., 2006; Perabo et al., 2006; Bartel et al., 2012)	- Minimising overall vector immunogenicity (Colella et al., 2019)	WHV protein X oncogenicity
		- WPRE mutants without the WHVX transcription, (Zanta-Boussif et al., 2009; Choi et al., 2014; Ling et al., 2015; Zhang et al., 2021)

All: Minimising dose by increasing efficacy and maximising on-target delivery by capsid engineering (Buchholz et al., 2015; Büning and Srivastava, 2019)

lacks the beta-element, providing a safer and space saving alternative (Choi et al., 2014). The use of these mutated WPRE sequences seems then advisable as the enhanced expression can support the use of smaller doses, and the more efficient transcriptional termination reduces the risk of transcriptional read-through.

Controlling on-target delivery and off-target expression for better overall safety

Dose affects all aspects of safety; the higher the dose, the higher the chance of immune activation and genome integration. Recent clinical trials utilising very high doses have used systemic delivery to target muscle (Duan, 2018; Shieh et al., 2020). In this case direct delivery to the target tissue is not feasible but comes with a widespread off-target transduction and thus loss of efficacy at the target tissue. Notably, systemic delivery means the vector is circulated through the liver, which most serotypes have some tropism for. In many cases this results in the rise of transaminases, markers of liver toxicity, which is one of the most common adverse effects in AAV clinical trials (Colella et al., 2018). However, some liver transduction may also be beneficial, as transgene expression in the liver seems to promote tolerogenicity (LoDuca et al., 2009; Fuchs et al., 2020). To take advantage of this Colella et al. (2019) multiplexed tissue specific regulatory elements to create liver-muscle and liver-neuron specific

tandem promoters, which successfully prevented anti-transgene immunity in mice. Many engineering strategies have focused on the on-target delivery, but it seems that now the focus needs to shift towards reducing off-target delivery. This means selecting clones that have the best transduction in the target tissue *in vivo* and the least in the control tissues–or similarly screening both the target cell type and any relevant off-targets *in vitro*. It is acknowledged that the translatability of the selection in the *in vitro* cell-based and the *in vivo* animal models to humans can be tricky; hence it is imperative that rational design approaches can also take off-target effects into consideration either at the design stage or later during the screening.

Capsid engineering can come with additional sets of problems in productivity and downstream processing that can hinder translation into the clinics. Alternatively, the expression of the transgene can be limited by using tissue specific promoters or other regulatory elements that induce gene expression in cell-type specific manner (Pacak et al., 2008; Wang et al., 2008; Mushimiyimana et al., 2021). Hence, in systemic delivery of these AAVs despite the high amounts of delivery to non-target tissues, the limited promoter activity can reduce the risk of genotoxicity and toxicity associated with the transgene expression. This has been observed in the context of ocular delivery where the use of non-specific or RPE-specific promoter led to strong RPE toxicity, while the transduction by photoreceptor specific promoter was well tolerated (Xiong et al., 2019).

Discussion

The decades of study into AAV gene therapy have generated a lot of insight and strategies into how to manipulate the AAV genome and capsid for therapeutic benefit. The tropism of the wild type virus can be engineered to suit our needs better by the tools of directed evolution and rational design, and we can make more informed choices on how to design the vector genome as more details become available. Here, in the light of the recent concerns, we have focused on covering the safety aspect, but it is notable that many of the strategies that improve safety can also improve efficacy. For example, CD8[+] T-cell responses cause loss of transgene expression, and off-target delivery results in the "waste" of the vector, requiring higher dosage. We have given an overview of the challenges and the potential solutions suggested in Table 1.

The utilisation of many of these strategies does not, of course, come without challenges. In genome engineering the biggest hurdle is often space; the already limited packaging capacity of AAV is usually utilised to the full, with little room for additional elements or large tissue-specific promoters. For capsid engineering the problem is multifactorial: the yield may not be as good as for the natural serotypes, purification with existing affinity chromatography columns not possible, and the regulatory demands higher. In many cases the ideal elements may be patented, requiring additional licensing. Additionally, the increasing complexity places more demands on the expertise required at multiple different fronts.

Naturally, safety is a complex issue that cannot be purely solved by vector design and may even be influenced by factors that we have not even yet considered. For example, recent studies have compared the axonal transport and anterograde transneuronal spread of the serotypes, finding differences that could have important implications for safety and clinical trial design (Aschauer et al., 2013; Zingg et al., 2017; Yu-Wai-Man et al., 2020). Even the immunomodulatory regimens chosen could play a vital role in determining whether the therapy is successful as wrongly chosen regimen can potentially block the induction of tolerance (Samelson-Jones et al., 2020). At the same time many issues are already well documented, such as the number of empty capsids and their influence on immunogenicity (Verdera et al., 2020). While genome and capsid engineering can be used to alleviate these issues, more solutions are still required at the downstream processing stage (Srivastava et al., 2021). In the animal studies genotoxicity has been strongest in neonatal animals or ones with HCC predisposition (Sabatino et al., 2022), and pre-existing hepatobiliary condition seems to have played a role at least in three of the serious adverse effects observed in the ASPIRO trial (Audentes Therapeutics, 2020). As such we need to gain more understanding on what risk factors to consider and screen for in patients. Sex also seems to matter; liver transduction with AAV2 and AAV5 was significantly higher in male mice than in females, whereas female mice had higher AAV9 transduction in the brain (Davidoff et al., 2003; Maguire et al., 2013). How this translates to human patients remains to be seen.

Bringing more of these insights of AAV into the clinical trials should be a top priority together with continuing the research to gain more in-depth understanding of the underlying biology. Currently, little testing has been done to combine any of the multitude of options covered in this review and seeing the feasibility of these in an "optimally safe AAV" would certainly be of interest to many. For now, the "store-bought" standard fare is then the easiest solution on many accords, but we should consider that in the future it may not be enough, especially as gene therapy aims to expand to target more common, less debilitating diseases. If we want to see more AAV gene therapies succeeding these challenges need to be faced head on, as we aim for safer, more efficacious, and cost-effective therapies.

Author contributions

Conceptualisation, TS, NL-K, and SY-H; Writing–Original Draft, TS; Writing–Review and Editing, TS, NL-K, and SY-H; Visualisation, TS; Supervision, NL-K and SY-H; Funding Acquisition and Resources, SY-H.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aschauer, D. F., Kreuz, S., and Rumpel, S. (2013). Analysis of transduction efficiency, tropism and axonal transport of AAV serotypes 1, 2, 5, 6, 8 and 9 in the mouse brain. *PLoS One* 8, e76310. doi:10.1371/journal.pone.0076310

Aslanidi, G. V., Rivers, A. E., Ortiz, L., Govindasamy, L., Ling, C., Jayandharan, G. R., et al. (2012). High-efficiency transduction of human monocyte-derived dendritic cells by capsid-modified recombinant AAV2 vectors. *Vaccine* 30, 3908–3917. doi:10.1016/j.vaccine.2012.03.079

Aslanidi, G. V., Rivers, A. E., Ortiz, L., Song, L., Ling, C., Govindasamy, L., et al. (2013). Optimization of the capsid of recombinant adeno-associated virus 2 (AAV2) vectors: The final threshold? *PLoS One* 8, 59142. doi:10.1371/journal.pone.0059142

Atchison, R. W., Casto, B. C., and Hammon, W. M. (1965). Adenovirus-associated defective virus particles. *Science* 149, 754–756. doi:10.1126/science.149.3685.754

Audentes Therapeutics (2020). A second patient has tragically died on the AT132 gene therapy trial for XLMTM - letter to XLMTM Patient Community. Available at: https://myotubulartrust.org/audentes-therapeutics-letter-23-june-2020/(Accessed September 7, 2022).

Ayuso, E., Blouin, V., Lock, M., Mcgorray, S., Leon, X., Alvira, M. R., et al. (2014). Manufacturing and characterization of a recombinant adeno-associated virus type 8 reference standard material. *Hum. Gene Ther.* 25, 977–987. doi:10.1089/hum.2014.057

Bartel, M. A., Weinstein, J. R., and Schaffer, D. V. (2012). Directed evolution of novel adeno-associated viruses for therapeutic gene delivery. *Gene Ther.* 19, 694–700. doi:10.1038/gt.2012.20

Barzel, A., Paulk, N. K., Shi, Y., Huang, Y., Chu, K., Zhang, F., et al. (2015). Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. *Nature* 517, 360–364. doi:10.1038/nature13864

Bell, P., Moscioni, A. D., McCarter, R. J., Wu, D., Gao, G., Hoang, A., et al. (2006). Analysis of tumors arising in male B6C3F1 mice with and without AAV vector delivery to liver. *Mol. Ther.* 14, 34–44. doi:10.1016/j.ymthe.2006.03.008

Bertolini, T. B., Shirley, J. L., Zolotukhin, I., Li, X., Kaisho, T., Xiao, W., et al. (2021). Effect of CpG depletion of vector genome on CD8+ T cell responses in AAV gene therapy. *Front. Immunol.* 12, 672449. doi:10.3389/fimmu.2021.672449

BioMarin investors (2022). First gene therapy for adults with severe hemophilia A, BioMarin's ROCTAVIAN (valoctocogene roxaparvovec). Approved by European Commission (EC). Available at: https://investors.biomarin.com/2022-08-24-First-Gene-Therapy-for-Adults-with-Severe-Hemophilia-A,-BioMarins-ROCTAVIAN-TM-valoctocogene-roxaparvovec-,-Approved-by-European-Commission-EC (Accessed September 7, 2022).

Boutin, S., Monteilhet, V., Veron, P., Leborgne, C., Benveniste, O., Montus, M. F., et al. (2010). Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: Implications for gene therapy using AAV vectors. *Hum. Gene Ther.* 21, 704–712. doi:10.1089/hum.2009.182

Browning, D. L., and Trobridge, G. D. (2016). Insulators to improve the safety of retroviral vectors for HIV gene therapy. *Biomedicines* 4, 4. doi:10.3390/biomedicines4010004

Buchholz, C. J., Friedel, T., and Büning, H. (2015). Surface-engineered viral vectors for selective and cell type-specific gene delivery. *Trends Biotechnol.* 33, 777–790. doi:10.1016/j.tibtech.2015.09.008

Buchlis, G., Podsakoff, G. M., Radu, A., Hawk, S. M., Flake, A. W., Mingozzi, F., et al. (2012). Factor IX expression in skeletal muscle of a severe hemophilia B patient 10 years after AAV-mediated gene transfer. *Blood* 119, 3038–3041. doi:10.1182/blood-2011-09-382317

Büning, H., Huber, A., Zhang, L., Meumann, N., and Hacker, U. (2015). Engineering the AAV capsid to optimize vector-host-interactions. *Curr. Opin. Pharmacol.* 24, 94–104. doi:10.1016/j.coph.2015.08.002

Büning, H., and Srivastava, A. (2019). Capsid modifications for targeting and improving the efficacy of AAV vectors. *Mol. Ther. Methods Clin. Dev.* 12, 248–265. doi:10.1016/j.omtm.2019.01.008

Cao, M., You, H., and Hermonat, P. L. (2014). The X gene of adeno-associated virus 2 (AAV2) is involved in viral DNA replication. *PLoS One* 9, 104596. doi:10.1371/journal.pone.0104596

Cataldi, M. P., and McCarty, D. M. (2010). Differential effects of DNA double-strand break repair pathways on single-strand and self-complementary adeno-associated virus vector genomes. *J. Virol.* 84, 8673–8682. doi:10.1128/jvi.00641-10

Chan, Y. K., Wang, S. K., Chu, C. J., Copland, D. A., Letizia, A. J., Verdera, H. C., et al. (2021). Engineering adeno-associated viral vectors to evade innate immune and inflammatory responses. *Sci. Transl. Med.* 13, 3438. doi:10.1126/SCITRANSLMED.ABD3438/SUPPL_FILE/ABD3438_SM

Chandler, R. J., La Fave, M. C., Varshney, G. K., Trivedi, N. S., Carrillo-Carrasco, N., Senac, J. S., et al. (2015). Vector design influences hepatic genotoxicity after adeno-associated virus gene therapy. *J. Clin. Investig.* 125, 870–880. doi:10.1172/JCI79213

Chen, S. J., Johnston, J., Sandhu, A., Bish, L. T., Hovhannisyan, R., Jno-Charles, O., et al. (2013). Enhancing the utility of adeno-associated virus gene transfer through inducible tissue-specific expression. *Hum. Gene Ther. Methods* 24, 270–278. doi:10.1089/hgtb.2012.129

Cheng, B., Ling, C., Dai, Y., Lu, Y., Glushakova, L. G., Gee, S. W. Y., et al. (2012). Development of optimized AAV3 serotype vectors: Mechanism of high-efficiency transduction of human liver cancer cells. *Gene Ther.* 19, 375–384. doi:10.1038/gt.2011.105

Choi, J. H., Yu, N. K., Baek, G. C., Bakes, J., Seo, D., Nam, H. J., et al. (2014). Optimization of AAV expression cassettes to improve packaging capacity and transgene expression in neurons. *Mol. Brain* 7, 1–10. doi:10.1186/1756-6606-7-17

Colella, P., Ronzitti, G., and Mingozzi, F. (2018). Emerging issues in AAV-mediated *in vivo* gene therapy. *Mol. Ther. Methods Clin. Dev.* 8, 87–104. doi:10.1016/j.omtm.2017.11.007

Dalwadi, D. A., Calabria, A., Tiyaboonchai, A., Posey, J., Naugler, W. E., Montini, E., et al. (2021a). AAV integration in human hepatocytes. *Mol. Ther.* 29, 2898–2909. doi:10.1016/j.ymthe.2021.08.031

Dalwadi, D. A., Torrens, L., Abril-Fornaguera, J., Pinyol, R., Willoughby, C., Posey, J., et al. (2021b). Liver injury increases the incidence of HCC following AAV gene therapy in mice. *Mol. Ther.* 29, 680–690. doi:10.1016/j.ymthe.2020.10.018

Davidoff, A. M., Ng, C. Y. C., Zhou, J., Spence, Y., and Nathwani, A. C. (2003). Sex significantly influences transduction of murine liver by recombinant adeno-associated viral vectors through an androgen-dependent pathway. *Blood* 102, 480–488. doi:10.1182/blood-2002-09-2889

Denard, J., Marolleau, B., Jenny, C., Rao, T. N., Fehling, H. J., Voit, T., et al. (2013). C-reactive protein (CRP) is essential for efficient systemic transduction of recombinant adeno-associated virus vector 1 (rAAV-1) and rAAV-6 in mice. *J. Virol.* 87, 10784–10791. doi:10.1128/jvi.01813-13

Dever, D. P., Bak, R. O., Reinisch, A., Camarena, J., Washington, G., Nicolas, C. E., et al. (2016). CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. *Nature* 539, 384–389. doi:10.1038/nature20134

Dobrowsky, T., Gianni, D., Pieracci, J., and Suh, J. (2021). AAV manufacturing for clinical use: Insights on current challenges from the upstream process perspective. *Curr. Opin. Biomed. Eng.* 20, 100353. doi:10.1016/j.cobme.2021.100353

Donsante, A., Miller, D. G., Li, Y., Vogler, C., Brunt, E. M., Russell, D. W., et al. (2007). AAV vector integration sites in mouse hepatocellular carcinoma. *Sci. (80-.)* 317, 477. doi:10.1126/science.1142658

Donsante, A., Vogler, C., Muzyczka, N., Crawford, J. M., Barker, J., Flotte, T., et al. (2001). Observed incidence of tumorigenesis in long-term rodent studies of rAAV vectors. *Gene Ther.* 8, 1343–1346. doi:10.1038/sj.gt.3301541

Duan, D. (2018). Systemic AAV micro-dystrophin gene therapy for Duchenne muscular dystrophy. *Mol. Ther.* 26, 2337–2356. doi:10.1016/j.ymthe.2018.07.011

Emmanuel, S. N., Smith, J. K., Hsi, J., Tseng, Y.-S., Kaplan, M., Mietzsch, M., et al. (2022). Structurally mapping antigenic epitopes of adeno-associated virus 9: Development of antibody escape variants. *J. Virol.* 96, e0125121. doi:10.1128/jvi.01251-21

Fasina, O., and Pintel, D. J. (2013). The adeno-associated virus type 5 small rep proteins expressed via internal translation initiation are functional. *J. Virol.* 87, 296–303. doi:10.1128/jvi.02547-12

Faust, S. M., Bell, P., Cutler, B. J., Ashley, S. N., Zhu, Y., Rabinowitz, J. E., et al. (2013). CpG-depleted adeno-associated virus vectors evade immune detection. *J. Clin. Investig.* 123, 2994–3001. doi:10.1172/JCI68205

Ferrari, F. K., Samulski, T., Shenk, T., and Samulski, R. J. (1996). Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J. Virol.* 70, 3227–3234. doi:10.1128/jvi.70.5.3227-3234.1996

Fitzsimons, H. L., McKenzie, J. M., and During, M. J. (2001). Insulators coupled to a minimal bidirectional tet cassette for tight regulation of rAAV-mediated gene transfer in the mammalian brain. *Gene Ther.* 8, 1675–1681. doi:10.1038/sj.gt.3301582

Fuchs, S. P., Martinez-Navio, J. M., Rakasz, E. G., Gao, G., and Desrosiers, R. C. (2020). Liver-directed but not muscle-directed AAV-antibody gene transfer limits

humoral immune responses in rhesus monkeys. *Mol. Ther. Methods Clin. Dev.* 16, 94–102. doi:10.1016/j.omtm.2019.11.010

Galibert, L., Hyvönen, A., Eriksson, R. A. E., Mattola, S., Aho, V., Salminen, S., et al. (2021). Functional roles of the membrane-associated AAV protein MAAP. *Sci. Rep.* 11, 21698. doi:10.1038/S41598-021-01220-7

Gao, G., Vandenberghe, L. H., Alvira, M. R., Lu, Y., Calcedo, R., Zhou, X., et al. (2004). Clades of adeno-associated viruses are widely disseminated in human tissues. *J. Virol.* 78, 6381–6388. doi:10.1128/JVI.78.12.6381-6388.2004

Girod, A., Ried, M., Wobus, C., Lahm, H., Leike, K., Kleinschmidt, J., et al. (1999). Genetic capsid modifications allow efficient re-targeting of adeno- associated virus type 2. *Nat. Med.* 5, 1052–1056. doi:10.1038/12491

Guillou, J., de Pellegars, A., Porcheret, F., Frémeaux-Bacchi, V., Allain-Launay, E., Debord, C., et al. (2022). Fatal thrombotic microangiopathy case following adeno-associated viral SMN gene therapy. *Blood Adv.* 6, 4266–4270. doi:10.1182/bloodadvances.2021006419

Gurda, B. L., DiMattia, M. A., Miller, E. B., Bennett, A., McKenna, R., Weichert, W. S., et al. (2013). Capsid antibodies to different adeno-associated virus serotypes bind common regions. *J. Virol.* 87, 9111–9124. doi:10.1128/jvi.00622-13

Gurda, B. L., Raupp, C., Popa-Wagner, R., Naumer, M., Olson, N. H., Ng, R., et al. (2012). Mapping a neutralizing epitope onto the capsid of adeno-associated virus serotype 8. *J. Virol.* 86, 7739–7751. doi:10.1128/jvi.00218-12

Haberman, R. P., McCown, T. J., and Samulski, R. J. (2000). Novel transcriptional regulatory signals in the adeno-associated virus terminal repeat A/D junction element. *J. Virol.* 74, 8732–8739. doi:10.1128/jvi.74.18.8732-8739.2000

Heister, T., Heid, I., Ackermann, M., and Fraefel, C. (2002). Herpes simplex virus type 1/adeno-associated virus hybrid vectors mediate site-specific integration at the adeno-associated virus preintegration site, AAVS1, on human chromosome 19. *J. Virol.* 76, 7163–7173. doi:10.1128/jvi.76.14.7163-7173.2002

Higashimoto, T., Urbinati, F., Perumbeti, A., Jiang, G., Zarzuela, A., Chang, L.-J., et al. (2007). The woodchuck hepatitis virus post-transcriptional regulatory element reduces readthrough transcription from retroviral vectors. *Gene Ther.* 14, 1298–1304. doi:10.1038/sj.gt.3302979

Hinderer, C., Katz, N., Buza, E. L., Dyer, C., Goode, T., Bell, P., et al. (2018). Severe toxicity in nonhuman primates and piglets following high-dose intravenous administration of an adeno-associated virus vector expressing human SMN. *Hum. Gene Ther.* 29, 285–298. doi:10.1089/hum.2018.015

Ho, A., Orton, R., Tayler, R., Asamaphan, P., Tong, L., Smollett, K., et al. (2022). *Adeno-associated virus 2 infection in children with non-A-E hepatitis.* medRxiv. doi:10.1101/2022.07.19.22277425

Hordeaux, J., Buza, E. L., Dyer, C., Goode, T., Mitchell, T. W., Richman, L., et al. (2020). Adeno-associated virus-induced dorsal root ganglion pathology. *Hum. Gene Ther.* 31, 808–818. doi:10.1089/hum.2020.167

Hordeaux, J., Wang, Q., Katz, N., Buza, E. L., Bell, P., and Wilson, J. M. (2018). The neurotropic properties of AAV-PHP.B are limited to C57BL/6J mice. *Mol. Ther.* 26, 664–668. doi:10.1016/J.YMTHE.2018.01.018

Huang, L. Y., Halder, S., and Agbandje-Mckenna, M. (2014). Parvovirus glycan interactions. *Curr. Opin. Virol.* 7, 108–118. doi:10.1016/j.coviro.2014.05.007

Huang, X., and Yang, Y. (2010). Targeting the TLR9MyD88 pathway in the regulation of adaptive immune responses. *Expert Opin. Ther. Targets* 14, 787–796. doi:10.1517/14728222.2010.501333

Hui, D. J., Edmonson, S. C., Podsakoff, G. M., Pien, G. C., Ivanciu, L., Camire, R. M., et al. (2015). AAV capsid CD8+ T-cell epitopes are highly conserved across AAV serotypes. *Mol. Ther. Methods Clin. Dev.* 2, 15029. doi:10.1038/mtm.2015.29

Huttner, N. A., Girod, A., Perabo, L., Edbauer, D., Kleinschmidt, J. A., Büning, H., et al. (2003). Genetic modifications of the adeno-associated virus type 2 capsid reduce the affinity and the neutralizing effects of human serum antibodies. *Gene Ther.* 10, 2139–2147. doi:10.1038/sj.gt.3302123

Hytönen, E., Laurema, A., Kankkonen, H., Miyanohara, A., Kärjä, V., Hujo, M., et al. (2019). Bile-duct proliferation as an unexpected side-effect after AAV2-LDLR gene transfer to rabbit liver. *Sci. Rep.* 9, 6934. doi:10.1038/s41598-019-43459-1

Keeler, A. M., and Flotte, T. R. (2019). Recombinant adeno-associated virus gene therapy in light of Luxturna (and Zolgensma and Glybera): Where are ee, and how did we get here? *Annu. Rev. Virol.* 6, 601–621. doi:10.1146/annurev-virology-092818-015530

King, J. A., Dubielzig, R., Grimm, D., and Kleinschmidt, J. A. (2001). DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *EMBO J.* 20, 3282–3291. doi:10.1093/emboj/20.12.3282

Kingsman, S. M., Mitrophanous, K., and Olsen, J. C. (2005). Potential oncogene activity of the woodchuck hepatitis post-transcriptional regulatory element (WPRE). *Gene Ther.* 12, 3–4. doi:10.1038/sj.gt.3302417

Klasse, P. J., and Sattentau, Q. J. (2002). Occupancy and mechanism in antibody-mediated neutralization of animal viruses. *J. Gen. Virol.* 83, 2091–2108. doi:10.1099/0022-1317-83-9-2091

Konkle, B. A., Walsh, C. E., Escobar, M. A., Josephson, N. C., Young, G., von Drygalski, A., et al. (2021). BAX 335 hemophilia B gene therapy clinical trial results: Potential impact of CpG sequences on gene expression. *Blood* 137, 763–774. doi:10.1182/blood.2019004625

Korpela, H., Lampela, J., Airaksinen, J., Järveläinen, N., Siimes, S., Valli, K., et al. (2022). AAV2-VEGF-B gene therapy failed to induce angiogenesis in ischemic porcine myocardium due to inflammatory responses. *Gene Ther.* 1, 322. doi:10.1038/s41434-022-00322-9

Kotin, R. M., Menninger, J. C., Ward, D. C., and Berns, K. I. (1991). Mapping and direct visualization of a region-specific viral DNA integration site on chromosome 19q13-qter. *Genomics* 10, 831–834. doi:10.1016/0888-7543(91)90470-Y

Li, B., Ma, W., Ling, C., Van Vliet, K., Huang, L. Y., Agbandje-McKenna, M., et al. (2015). Site-directed mutagenesis of surface-exposed lysine residues leads to improved transduction by AAV2, but not AAV8, vectors in murine hepatocytes *in vivo. Hum. Gene Ther. Methods* 26, 211–220. doi:10.1089/hgtb.2015.115

Li, C., and Samulski, R. J. (2020). Engineering adeno-associated virus vectors for gene therapy. *Nat. Rev. Genet.* 21, 255–272. doi:10.1038/s41576-019-0205-4

Li, Y., Miller, C. A., Shea, L. K., Jiang, X., Guzman, M. A., Chandler, R. J., et al. (2021). Enhanced efficacy and increased long-term toxicity of CNS-directed, AAV-based combination therapy for Krabbe disease. *Mol. Ther.* 29, 691–701. doi:10.1016/j.ymthe.2020.12.031

Ling, C., Wang, Y., Lu, Y., Wang, L., Jayandharan, G. R., Aslanidi, G. V., et al. (2015). Enhanced transgene expression from recombinant single-stranded D-sequence-substituted adeno-associated virus vectors in human cell lines *in vitro* and in murine hepatocytes *in vivo. J. Virol.* 89, 952–961. doi:10.1128/JVI.02581-14

Liu, D., Zhu, M., Zhang, Y., and Diao, Y. (2021). Crossing the blood-brain barrier with AAV vectors. *Metab. Brain Dis.* 36, 45–52. doi:10.1007/s11011-020-00630-2

Lock, M., McGorray, S., Auricchio, A., Ayuso, E., Beecham, E. J., Blouin-Tavel, V., et al. (2010). Characterization of a recombinant adeno-associated virus type 2 reference standard material. *Hum. Gene Ther.* 21, 1273–1285. doi:10.1089/hum.2009.223

LoDuca, P., Hoffman, B., and Herzog, R. (2009). Hepatic gene transfer as a means of tolerance induction to transgene products. *Curr. Gene Ther.* 9, 104–114. doi:10.2174/156652309787909490

Loeb, J. E., Cordier, W. S., Harris, M. E., Weitzman, M. D., and Hope, T. J. (1999). Enhanced expression of transgenes from adeno-associated virus vectors with the woodchuck hepatitis virus posttranscriptional regulatory element: Implications for gene therapy. *Hum. Gene Ther.* 10, 2295–2305. doi:10.1089/10430349950016942

Logan, G. J., Dane, A. P., Hallwirth, C. V., Smyth, C. M., Wilkie, E. E., Amaya, A. K., et al. (2017). Identification of liver-specific enhancer–promoter activity in the 3′ untranslated region of the wild-type AAV2 genome. *Nat. Genet.* 49, 1267–1273. doi:10.1038/ng.3893

Maguire, C. A., Crommentuijn, M. H. W., Mu, D., Hudry, E., Serrano-Pozo, A., Hyman, B. T., et al. (2013). Mouse gender influences brain transduction by intravascularly administered AAV9. *Mol. Ther.* 21, 1470–1471. doi:10.1038/mt.2013.95

Maheshri, N., Koerber, J. T., Kaspar, B. K., and Schaffer, D. V. (2006). Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. *Nat. Biotechnol.* 24, 198–204. doi:10.1038/nbt1182

Mangsbo, S. M., Sanchez, J., Anger, K., Lambris, J. D., Ekdahl, K. N., Loskog, A. S., et al. (2009). Complement activation by CpG in a human whole blood loop system: Mechanisms and Immunomodulatory Effects. *J. Immunol.* 183, 6724–6732. doi:10.4049/jimmunol.0902374

Manno, C. S., Arruda, V. R., Pierce, G. F., Glader, B., Ragni, M., Rasko, J. J. E., et al. (2006). Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nat. Med.* 12, 342–347. doi:10.1038/nm1358

Martino, A. T., Basner-Tschakarjan, E., Markusic, D. M., Finn, J. D., Hinderer, C., Zhou, S., et al. (2013). Engineered AAV vector minimizes *in vivo* targeting of transduced hepatocytes by capsid-specific CD8+ T cells. *Blood* 121, 2224–2233. doi:10.1182/blood-2012-10-460733

Martino, A. T., Suzuki, M., Markusic, D. M., Zolotukhin, I., Ryals, R. C., Moghimi, B., et al. (2011). The genome of self-complementary adeno-associated viral vectors increases Toll-like receptor 9-dependent innate immune responses in the liver. *Blood* 117, 6459–6468. doi:10.1182/blood-2010-10-314518

McCarty, D. M., Fu, H., Monahan, P. E., Toulson, C. E., Naik, P., and Samulski, R. J. (2003). Adeno-associated virus terminal repeat (TR) mutant generates self-complementary vectors to overcome the rate-limiting step to transduction *in vivo. Gene Ther.* 10, 2112–2118. doi:10.1038/sj.gt.3302134

Melo, S. P., Lisowski, L., Bashkirova, E., Zhen, H. H., Chu, K., Keene, D. R., et al. (2014). Somatic correction of junctional epidermolysis bullosa by a highly recombinogenic AAV variant. *Mol. Ther.* 22, 725–733. doi:10.1038/mt.2013.290

Meri, S. (2016). Self-nonself discrimination by the complement system. *FEBS Lett.* 590, 2418–2434. doi:10.1002/1873-3468.12284

Miller, D. G., Petek, L. M., and Russell, D. W. (2004). Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat. Genet.* 36, 767–773. doi:10.1038/ng1380

Mushimiyimana, I., Niskanen, H., Beter, M., Laakkonen, J. P., Kaikkonen, M. U., Ylä-Herttuala, S., et al. (2021). Characterization of a functional endothelial super-enhancer that regulates ADAMTS18 and angiogenesis. *Nucleic Acids Res.* 49, 8078–8096. doi:10.1093/nar/gkab633

Nguyen, G. N., Everett, J. K., Kafle, S., Roche, A. M., Raymond, H. E., Leiby, J., et al. (2021). A long-term study of AAV gene therapy in dogs with hemophilia A identifies clonal expansions of transduced liver cells. *Nat. Biotechnol.* 39, 47–55. doi:10.1038/s41587-020-0741-7

Nielsen, M., Lund, O., Buus, S., and Lundegaard, C. (2010). MHC class II epitope predictive algorithms. *Immunology* 130, 319–328. doi:10.1111/j.1365-2567.2010.03268.x

Nowrouzi, A., Penaud-Budloo, M., Kaeppel, C., Appelt, U., Le Guiner, C., Moullier, P., et al. (2012). Integration frequency and intermolecular recombination of rAAV vectors in non-human primate skeletal muscle and liver. *Mol. Ther.* 20, 1177–1186. doi:10.1038/mt.2012.47

Ogden, P. J., Kelsic, E. D., Sinai, S., and Church, G. M. (2019). Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. *Sci. (80-.)* 366, 1139–1143. doi:10.1126/science.aaw2900

Pacak, C. A., Sakai, Y., Thattaliyath, B. D., Mah, C. S., and Byrne, B. J. (2009). Erratum: Tissue specific promoters improve specificity of AAV9 mediated transgene expression following intra-vascular gene delivery in neonatal mice (genetic vaccines and therapy). *Genet. Vaccines Ther.* 7, 13–15. doi:10.1186/1479-0556-6-13

Pan, X., Yue, Y., Boftsi, M., Wasala, L. P., Tran, N. T., Zhang, K., et al. (2022). Rational engineering of a functional CpG-free ITR for AAV gene therapy. *Gene Ther.* 29, 333–345. doi:10.1038/s41434-021-00296-0

Paul, S., Croft, N. P., Purcell, A. W., Tscharke, D. C., Sette, A., Nielsen, M., et al. (2020). Benchmarking predictions of MHC class I restricted T cell epitopes in a comprehensively studied model system. *PLoS Comput. Biol.* 16, 1007757. doi:10.1371/journal.pcbi.1007757

Perabo, L., Endell, J., King, S., Lux, K., Goldnau, D., Hallek, M., et al. (2006). Combinatorial engineering of a gene therapy vector: Directed evolution of adeno-associated virus. *J. Gene Med.* 8, 155–162. doi:10.1002/jgm.849

Pfizer, S. (2020). New Phase 1B results of gene therapy in ambulatory boys with Duchenne muscular dystrophy (Dmd) support advancement into pivotal phase 3 study. Available at: https://investors.pfizer.com/investor-news/press-release-details/2020/Pfizers-New-Phase-1b-Results-of-Gene-Therapy-in-Ambulatory-Boys-with-Duchenne-Muscular-Dystrophy-DMD-Support-Advancement-into-Pivotal-Phase-3-Study/default.aspx (Accessed September 7, 2022).

PTC Therapeutics (2022). *Upstaza^TM granted marketing authorization by european commission as first disease-modifying treatment for aadc deficiency.* PTC Therapeutics, Inc. Available at: https://ir.ptcbio.com/news-releases/news-release-details/upstazatm-granted-marketing-authorization-european-commission (Accessed October 21, 2022).

Rabinowitz, J. E., Rolling, F., Li, C., Conrath, H., Xiao, W., Xiao, X., et al. (2002). Cross-packaging of a single adeno-associated virus (AAV) type 2 vector genome into multiple AAV serotypes enables transduction with broad specificity. *J. Virol.* 76, 791–801. doi:10.1128/jvi.76.2.791-801.2002

Ramsey, J. P., Khatwani, S. L., Lin, M., Boregowda, R., Surosky, R., and Andrew Ramelmeier, R. (2021). Overview of analytics needed to support a robust gene therapy manufacturing process. *Curr. Opin. Biomed. Eng.* 20, 100339. doi:10.1016/j.cobme.2021.100339

Raupp, C., Naumer, M., Müller, O. J., Gurda, B. L., Agbandje-McKenna, M., and Kleinschmidt, J. A. (2012). The threefold protrusions of adeno-associated virus type 8 are involved in cell surface targeting as well as post attachment processing. *J. Virol.* 86, 9396–9408. doi:10.1128/jvi.00209-12

Recchia, A., Parks, R. J., Lamartina, S., Toniatti, C., Pieroni, L., Palombo, F., et al. (1999). Site-specific integration mediated by a hybrid adenovirus/adeno-associated virus vector. *Proc. Natl. Acad. Sci. U. S. A.* 96, 2615–2620. doi:10.1073/pnas.96.6.2615

Rosas, L. E., Grieves, J. L., Zaraspe, K., La Perle, K. M. D., Fu, H., and McCarty, D. M. (2012). Patterns of scAAV vector insertion associated with oncogenic events in a mouse model for genotoxicity. *Mol. Ther.* 20, 2098–2110. doi:10.1038/mt.2012.197

Sabatino, D. E., Bushman, F. D., Chandler, R. J., Crystal, R. G., Davidson, B. L., Dolmetsch, R., et al. (2022). Evaluating the state of the science for adeno-associated virus integration: An integrated perspective. *Mol. Ther.* 30, 2646–2663. doi:10.1016/j.ymthe.2022.06.004

Salmon, F., Grosios, K., and Petry, H. (2014). Safety profile of recombinant adeno-associated viral vectors: Focus on alipogene tiparvovec (Glybera®). *Expert Rev. Clin. Pharmacol.* 7, 53–65. doi:10.1586/17512433.2014.852065

Samelson-Jones, B. J., Finn, J. D., Favaro, P., Wright, J. F., and Arruda, V. R. (2020). Timing of intensive immunosuppression impacts risk of transgene antibodies after AAV gene therapy in nonhuman primates. *Mol. Ther. Methods Clin. Dev.* 17, 1129–1138. doi:10.1016/J.OMTM.2020.05.001

Sarma, J. V., and Ward, P. A. (2011). The complement system. *Cell Tissue Res.* 343, 227–235. doi:10.1007/s00441-010-1034-0

Sen, D., Balakrishnan, B., Gabriel, N., Agrawal, P., Roshini, V., Samuel, R., et al. (2013). Improved adeno-associated virus (AAV) serotype 1 and 5 vectors for gene therapy. *Sci. Rep.* 6, 1832. doi:10.1038/srep01832

Shao, W., Earley, L. F., Chai, Z., Chen, X., Sun, J., He, T., et al. (2018). Double-stranded RNA innate immune response activation from long-term adeno-associated virus vector transduction. *JCI Insight* 3, 120474. doi:10.1172/jci.insight.120474

Shieh, P. B., Kuntz, N., Smith, B., Dowling, J., Müller-Felber, W., Bönnemann, C. G., et al. (2020). ASPIRO gene therapy trial In X-linked myotubular myopathy (XLMTM): update on preliminary safety and efficacy findings up to 72 weeks post-treatment, *Neurology* 94.

Solid Biosciences (2018). *Solid Biosciences announces clinical hold on SGT-001 phase I/II clinical trial for Duchenne muscular dystrophy.* Solid Biosci. Press Release. Available at: https://www.mda.org/press-releases/solid-biosciences-announces-clinical-hold-sgt-001-phase-iii-clinical-trial-duchenne (Accessed September 7, 2022).

Song, L., Kauss, M. A., Kopin, E., Chandra, M., Ul-Hasan, T., Miller, E., et al. (2013). Optimizing the transduction efficiency of capsid-modified AAV6 serotype vectors in primary human hematopoietic stem cells *in vitro* and in a xenograft mouse model *in vivo*. *Cytotherapy* 15, 986–998. doi:10.1016/J.JCYT.2013.04.003

Sonntag, F., Schmidt, K., and Kleinschmidt, J. A. (2010). A viral assembly factor promotes AAV2 capsid formation in the nucleolus. *Proc. Natl. Acad. Sci. U. S. A.* 107, 10220–10225. doi:10.1073/pnas.1001673107

Srivastava, A., Mallela, K. M. G., Deorkar, N., and Brophy, G. (2021). Manufacturing challenges and rational formulation development for AAV viral vectors. *J. Pharm. Sci.* 110, 2609–2624. doi:10.1016/j.xphs.2021.03.024

Stracker, T. H., Cassell, G. D., Ward, P., Loo, Y.-M., van Breukelen, B., Carrington-Lawrence, S. D., et al. (2004). The rep protein of adeno-associated virus type 2 interacts with single-stranded DNA-binding proteins that enhance viral replication. *J. Virol.* 78, 441–453. doi:10.1128/jvi.78.1.441-453.2004

Tran, N. T., Heiner, C., Weber, K., Weiand, M., Wilmot, D., Xie, J., et al. (2020). AAV-genome population sequencing of vectors packaging CRISPR components reveals design-influenced heterogeneity. *Mol. Ther. Methods Clin. Dev.* 18, 639–651. doi:10.1016/j.omtm.2020.07.007

Tu, H., Bonura, C., Giannini, C., Mouly, H., Soussan, P., Kew, M., et al. (2001). Biological impact of natural COOH-terminal deletions of Hepatitis B virus X protein in hepatocellular carcinoma tissues. *Cancer Res.* 61, 7803–7810. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11691796 (Accessed January 22, 2020).

Vandamme, C., Adjali, O., and Mingozzi, F. (2017). Unraveling the complex story of immune responses to AAV vectors trial after trial. *Hum. Gene Ther.* 28, 1061–1074. doi:10.1089/hum.2017.150

Verdera, H. C., Kuranda, K., and Mingozzi, F. (2020). AAV vector immunogenicity in humans: A long journey to successful gene transfer. *Mol. Ther.* 28, 723–746. doi:10.1016/j.ymthe.2019.12.010

Wang, B., Li, J., Fu, F. H., Chen, C., Zhu, X., Zhou, L., et al. (2008). Construction and analysis of compact muscle-specific promoters for AAV vectors. *Gene Ther.* 15, 1489–1499. doi:10.1038/gt.2008.104

Wang, Z., Lisowski, L., Finegold, M. J., Nakai, H., Kay, M. A., and Grompe, M. (2012). AAV vectors containing rDNA homology display increased chromosomal integration and transgene persistence. *Mol. Ther.* 20, 1902–1911. doi:10.1038/mt.2012.157

Wei, W., Pelechano, V., Järvelin, A. I., and Steinmetz, L. M. (2011). Functional consequences of bidirectional promoters. *Trends Genet.* 27, 267–276. doi:10.1016/j.tig.2011.04.002

Wei, Y., Etiemble, J., Fourel, G., Buendia, M.-A., and Vitvitski-Trepo, L. (1995). Hepadna virus integration generates virus-cell cotranscripts carrying 3′ truncated X genes in human and woodchuck liver tumors. *J. Med. Virol.* 45, 82–90. doi:10.1002/jmv.1890450116

Weindler, F. W., and Heilbronn, R. (1991). A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication. *J. Virol.* 65, 2476–2483. doi:10.1128/jvi.65.5.2476-2483.1991

Wright, J. F. (2020). Quantification of CpG motifs in rAAV genomes: Avoiding the toll. *Mol. Ther.* 28, 1756–1758. doi:10.1016/j.ymthe.2020.07.006

Wu, Z., Asokan, A., and Samulski, R. J. (2006). Adeno-associated virus serotypes: Vector toolkit for human gene therapy. *Mol. Ther.* 14, 316–327. doi:10.1016/j.ymthe.2006.05.009

Wu, Z., Yang, H., and Colosi, P. (2010). Effect of genome size on AAV vector packaging. *Mol. Ther.* 18, 80–86. doi:10.1038/MT.2009.255

Xie, J., Mao, Q., Tai, P. W. L., He, R., Ai, J., Su, Q., et al. (2017). Short DNA hairpins compromise recombinant adeno-associated virus genome homogeneity. *Mol. Ther.* 25, 1363–1374. doi:10.1016/j.ymthe.2017.03.028

Xiong, W., Wu, D. M., Xue, Y., Wang, S. K., Chung, M. J., Ji, X., et al. (2019). AAV cis-regulatory sequences are correlated with ocular toxicity. *Proc. Natl. Acad. Sci. U. S. A.* 116, 5785–5794. doi:10.1073/pnas.1821000116

Yang, Y., Wang, L., Bell, P., McMenamin, D., He, Z., White, J., et al. (2016). A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat. Biotechnol.* 34, 334–338. doi:10.1038/nbt.3469

Ylä-Herttuala, S. (2012). Endgame: Glybera finally recommended for approval as the first gene therapy drug in the European Union. *Mol. Ther.* 20, 1831–1832. doi:10.1038/mt.2012.194

Yu-Wai-Man, P., Newman, N. J., Carelli, V., Moster, M. L., Biousse, V., Sadun, A. A., et al. (2020). Bilateral visual improvement with unilateral gene therapy injection for Leber hereditary optic neuropathy. *Sci. Transl. Med.* 12, 7423. doi:10.1126/scitranslmed.aaz7423

Zaiss, A. K., Cotter, M. J., White, L. R., Clark, S. A., Wong, N. C. W., Holers, V. M., et al. (2008). Complement is an essential component of the immune response to adeno-associated virus vectors. *J. Virol.* 82, 2727–2740. doi:10.1128/jvi.01990-07

Zanta-Boussif, M. A., Charrier, S., Brice-Ouzet, A., Martin, S., Opolon, P., Thrasher, A. J., et al. (2009). Validation of a mutated PRE sequence allowing high and sustained transgene expression while abrogating WHV-X protein synthesis: Application to the gene therapy of WAS. *Gene Ther.* 16, 605–619. doi:10.1038/gt.2009.3

Zhang, J., Yu, X., Herzog, R. W., Samulski, R. J., and Xiao, W. (2021). Flies in the ointment: AAV vector preparations and tumor risk. *Mol. Ther.* 29, 2637–2639. doi:10.1016/j.ymthe.2021.08.016

Zhong, L., Li, B., Jayandharan, G., Mah, C. S., Govindasamy, L., Agbandje-McKenna, M., et al. (2008). Tyrosine-phosphorylation of AAV2 vectors and its consequences on viral intracellular trafficking and transgene expression. *Virology* 381, 194–202. doi:10.1016/j.virol.2008.08.027

Zhu, J., Huang, X., and Yang, Y. (2009). The TLR9-MyD88 pathway is critical for adaptive immune responses to adeno-associated virus gene therapy vectors in mice. *J. Clin. Investig.* 119, 2388–2398. doi:10.1172/JCI37607

Zingg, B., Chou, X., Zhang, Z., Mesik, L., Liang, F., Tao, H. W., et al. (2017). AAV-Mediated anterograde transsynaptic tagging: Mapping corticocollicular input-defined neural pathways for defense behaviors. *Neuron* 93, 33–47. doi:10.1016/j.neuron.2016.11.045

Zolotukhin, S., and Vandenberghe, L. H. (2022). AAV capsid design: A goldilocks challenge. *Trends Mol. Med.* 28, 183–193. doi:10.1016/j.molmed.2022.01.003

The COVID-19 explorer—An integrated, whole patient knowledge model of COVID-19 disease

Stephan Brock[1], Theodoros G. Soldatos[1,2], David B. Jackson[1], Francesca Diella[1], Klaus Hornischer[1], Anne Schäfer[1], Simon P. Hoerstrup[3,4]* and Maximilian Y. Emmert[3,4,5,6]*

[1]Molecular Health GmbH, Heidelberg, Germany, [2]SRH Hochscule, University of Applied Science, Heidelberg, Germany, [3]Institute for Regenerative Medicine, University of Zurich, Zurich, Switzerland, [4]Wyss Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland, [5]Department of Cardiothoracic and Vascular Surgery, German Heart Institute Berlin, Berlin, Germany, [6]Department of Cardiovascular Surgery, Charité Universitätsmedizin Berlin, Berlin, Germany

Since early 2020 the COVID-19 pandemic has paralyzed the world, resulting in more than half a billion infections and over 6 million deaths within a 28-month period. Knowledge about the disease remains largely disjointed, especially when considering the molecular mechanisms driving the diversity of clinical manifestations and symptoms. Despite the recent availability of vaccines, there remains an urgent need to develop effective treatments for cases of severe disease, especially in the face of novel virus variants. The complexity of the situation is exacerbated by the emergence of COVID-19 as a complex and multifaceted systemic disease affecting independent tissues and organs throughout the body. The development of effective treatment strategies is therefore predicated on an integrated understanding of the underlying disease mechanisms and their potentially causative link to the diversity of observed clinical phenotypes. To address this need, we utilized a computational technology (the Dataome platform) to build an integrated clinico-molecular view on the most important COVID-19 clinical phenotypes. Our results provide the first integrated, whole-patient model of COVID-19 symptomatology that connects the molecular lifecycle of SARS-CoV-2 with microvesicle-mediated intercellular communication and the contact activation and kallikrein-kinin systems. The model not only explains the clinical pleiotropy of COVID-19, but also provides an evidence-driven framework for drug development/repurposing and the identification of critical risk factors. The associated knowledge is provided in the form of the open source COVID-19 Explorer (https://covid19.molecularhealth.com), enabling the global community to explore and analyze the key molecular features of systemic COVID-19 and associated implications for research priorities and therapeutic strategies. Our work suggests that knowledge modeling solutions may offer important utility in expediting the global response to future health emergencies.

1 Introduction

If there is any positive to be gleaned from the devastating COVID-19 pandemic, it could be the scale and impact of global response from the biomedical research community to the study of the SARS-CoV-2 virus. From efforts to characterize molecular disease mechanisms in the search for tractable therapeutic avenues, to drug repurposing and pandemic forecasting, we have witnessed unprecedented levels of multidisciplinary collaboration. Nevertheless, the resultant peer-reviewed insights have come at a rapid velocity and volume that makes it challenging to efficiently capture, integrate and analyze emergent data from both the clinical and molecular domains. The integration of such insights with existing knowledge is pivotal to efficient knowledge transfer and assimilation by the research community, aiding our understanding of disease mechanisms and expediting the generation and testing of associated therapeutic hypotheses. Recognizing the global urgency, we initiated a COVID-19 focused knowledge modeling effort in March 2020 that sought to rapidly address this challenge. Our goal was to develop a whole patient knowledge model of COVID-19 symptomatology and associated molecular knowledge that links the key molecular players in disease pathophysiology to: 1) common symptoms, 2) severe manifestations and 3) outcome and severity-associated risk factors.

To achieve this, we designed a stepwise, expert-driven knowledge modeling strategy that iteratively combined the capacities of both extensive data integration and human insight (see Figure 1). This supervised strategy helped us to 1) manage the rapid pace of new insights, 2) enable the flexible elaboration of more specific disease symptom models and associated hypotheses, and 3) permit the real-time inclusion of important new findings at the whole-patient level. Here, we report on our findings and the functionality of the COVID-19 Explorer web resource. Our results are provided at the level of a whole-patient knowledge model, including the possible causative pathogenic mechanisms underlying COVID-19 phenotypes (Brock et al., 2022). To accommodate usability for a variety of use-case scenarios, results are summarized in different formats: the 'COVID-19 Explorer' provides a detailed, comprehensive and fully interactive view of the relationships and accompanying evidence, while 'The COVID-19 Cockpit' (see Supplementary Table S1) is intended to support clinical researchers in diagnostic and hypothesis generation for new therapeutic strategies (see Figure 1).

By developing a patient-level molecular atlas of COVID-19 pathogenesis as an interactive open-source knowledge model, we provide the biomedical and clinical research communities with an effective tool to decipher COVID-19 hypotheses and to enable more informed development and testing of new diagnostic and therapeutic strategies.

1.1 Materials and methods

We utilized Molecular Health's (MH's) Dataome technology platform in collaboration with disease modeling experts to capture, structure and logically connect diverse clinical and molecular features of COVID-19 pathobiology (see Figure 1). Readouts from these rapid *in silico* analyses were curated and organized into molecular models containing salient information for each symptom (see Figure 2 and Figure 4). The final complete comprehensive model, linking key molecular mechanisms to COVID-19 symptomatology and the related source data, is made publicly available *via* the web-based COVID-19 Explorer (accessible, at http://covid19.molecularhealth.com).

1.1.1 The Dataome technology

Our studies utilized the Dataome technology as the core data integration, knowledgebase and analytical framework. Dataome was designed to enable the constant capture and curation of globally available data sources of clinical and molecular knowledge, with the aim of delivering quality controlled data for clinical decision-making and knowledge discovery in a disease agnostic manner. It consists of three core components (see Supplementary Material S1, Figure 1):

- Dataome Capture—uses an ensemble of public/proprietary algorithms and resources to enable the global harvesting, quality assurance and integration of emergent clinical and molecular data. Structured data is assimilated using automated data integration pipelines that process, normalize and quality assure the data in synchrony with database update cycles. These functions utilize an extensive infrastructure that enables extraction, transformation, and loading (ETL) of source data into a consolidated database framework for information modelling and knowledge extraction. For unstructured data, text and data mining (TDM) technologies provide an ensemble of natural language processing (NLP) functions, such as rule-based linguistic, machine learning and deep learning models, trained to identify critical biomedical terms and relationships from any source of unstructured knowledge (e.g., drug labels, patents and peer-reviewed literature). This machine-reading framework enables targeted extraction of biomedical facts that are fed to a

FIGURE 1
Construction and utility of the COVID-19 molecular symptomology model. Timeline **(left)**: Expert-supervised knowledge modeling strategy enabled the rapid construction of a comprehensive COVID-19 Model using Molecular Health's (MH) Dataome technology. **Timeline (right)**: a constant flow of new observations during the course of the pandemic has provided real-time validations of the resulting model—several findings are today clinically validated and incorporated in multicenter studies. **Initial Data Collection (left; bottom):** a "Seed Model" was constructed based on information derived from the scientific literature available at the early stage of the pandemic, accompanied by supportive data provided by domain experts and relevant knowledge stored in databases. More specifically, the Dataome knowledgebase of clinical and molecular knowledge was queried using a pre-established SARS-CoV-2 protein interaction map of 332 high confidence human protein interactors within the human proteome (Gordon et al., 2020), involving molecular determinants of host cells and the differential impact of the viral infection on the host cell response. **Model Expansion (left; upper):** the 'Seed Model' was normalized, computationally structured and automatically connected to broader biomolecular information to provide the "Base Model"—using Dataome core data integration technology (see also Figure 3), the "Base Model" was iteratively expanded further *via* an expert-guided strategy. This computer augmented evidence prioritization effort revealed a set of converging central molecular pathophysiology mechanisms. **Results (right)**: findings were organized in three different representations providing overlapping and cross-connected functional utilities for the interested clinician and/or translational researcher, namely a network-based view of the COVID-19 Model encompassing a higher-level summary of the eight key pathogenic mechanisms found to be associated with COVID-19 clinical phenotypes (see also Figure 4), a web-based 'COVID-19 Explorer' that links each component of the interconnected COVID-19 Disease Model network to underlying views with additional information and detailed evidence summaries (publicly available at http://covid19.molecularhealth.com), and the "COVID-19 Cockpit"—a guide summarizing the COVID-19 Model in a tabular format with human understandable texts explaining per row reported COVID-19 clinical phenotypes, associated pathogenic mechanisms, and their relation to the COVID-19 Model with respect to each of the eight key SARS-CoV-2 perturbed mechanisms (including links to the respective 'COVID-19 Explorer' description-views). **Clinical significance (right)**: the COVID-19 Model and its representations enable direct clinical or therapeutic support by providing both causal and inferred insights that reflect the symptomology of patients and their individual characteristics (see Supplementary Table S1)—importantly, continuous synchronization with new SARS-CoV-2 data and MH Dataome content is warranted for quick-response, and fast informed data and hypothesis generation towards improved COVID-19 management and better understanding of its multifaceted phenotype manifestations and duration. Treatment modalities indicated by the model include drugs associated with modulation of *ACE2* induction, *B1R* agonists, as well as *Kallikrein* inhibitors, *SerpinG1* enzyme replacement and *SIRT-1* or *PPARγ* activators. Several of these are currently investigated in late phase COVID-19 clinical trial studies (see also (Brock et al., 2022)).

proprietary curation infrastructure for review by biomedical experts. Curated and quality-controlled data is then integrated into the Dataome's Nucleus.

- Dataome Nucleus—a comprehensive data and knowledge resource containing highly interconnected clinical and molecular data, linking clinical phenotypes to underlying molecular knowledge. Nucleus encompasses

data from more than one hundred (100+) public and commercial/private resources, including in-house proprietary databases (see Figure 2 in Supplementary Material S1). Public datasets span a broad range of content, size and formats—from more general, such as literature [e.g., PubMed (Sayers et al., 2021; Kim et al., 2021)], biomedical ontologies [e.g., ICD (Krawczyk et al.,

FIGURE 2
The COVID-19 Explorer Interface. Screenshots from the COVID-19 Explorer highlighting how users can navigate through different levels of detailed visualizations, annotated graphics and reference information. Key pages within the interface include **(A)** Whole patient symptomatology **(B)** eight converging pathogenic disease mechanisms **(C)** a simplified disease model **(D)** an expanded disease model **(E)** symptom mechanisms **(F)** functional mechanisms **(G)** publications and **(H)** terminology references. Each of these views is designed to assist translational researchers in both hypothesis development and testing for COVID-19.

2020), MedDRA (Brown, 2004), ATC (Merabti et al., 2011), MeSH (Lipscomb, 2000), UMLS (Humphreys et al., 2020)] or information about proteins and genes [e.g., from resources like UniProt (UniProt Consortium, 2021), Entrez Gene (Maglott et al., 2011), Ensembl (Howe et al., 2021), UCSC (Gonzalez et al., 2021), or RefSeq (O'Leary et al., 2016)], to more targeted information, such as genomic variant annotations [e.g., ClinVar (Landrum et al., 2020), dbSNP/dbNFSP (Sayers et al., 2021)], information about drugs and their labels, targets or interactors [e.g., DrugBank (Wishart et al., 2018), FDA Orange Book (Ursu et al., 2018) or Drugs@FDA (Ursu et al., 2018)], or biomolecular pathways and interactions [e.g., KEGG (Kanehisa et al., 2021) and Reactome (Jassal et al., 2020)]. Other resources include real-world data (RWD) such as public pharmacovigilance repositories [e.g., VigiBase (Fernandez et al., 2020) or FAERS (Yao et al., 2020)], as well as information regarding clinical trials (e.g., from NCT's clinicaltrials.gov). The system also contains structured information regarding therapeutic guidelines and variant classification [e.g., ACMG (Harrison et al., 2019), NCCN (Koh et al., 2020) or ESMO (Cherny et al., 2017)], as well as further curated datasets pertaining clinical biomarker interpretation,

pathway/interaction relationships, drug and clinical trial information.

- Dataome Analytics—provides a portfolio of analytical solutions designed to derive new insights from the data contained in Nucleus. This data provides the evidence-base to support the development of both commercial decision support technologies [e.g., MH EFFECT (Schotland et al., 2021)].

- and MH GUIDE (Hirotsu et al., 2020) and the efficient development of disease models for any human disease or phenotype, in this case COVID-19. These software tools are complemented by specialized analytical pipelines that integrate bioinformatics, chemoinformatics, systems biology, clinical data science, and AI/machine learning (e.g., with integrated analysis, feature engineering and powerful pre-trained models) based methodologies.

The utility of the integrated Dataome technology has been previously validated across multiple clinically important contexts, including biomarker discovery, drug safety prediction and drug repositioning (Armaiz-Pena et al., 2013; Bohnert et al., 2017; Pradeep et al., 2015; Soldatos et al., 2018; Soldatos and Jackson, 2019; Schotland et al., 2021; Schell et al., 2016). While Dataome provides a flexible approach to the

automated capture and quality assurance of globally available sources of clinical and molecular data and knowledge, a supervised approach was required for the molecular modeling of COVID-19 to account for the rapid emergence of new insights. This permitted the stepwise, expert-guided elaboration of more specific COVID-19 symptomology models and associated hypotheses, flexible enough to enable the near real-time inclusion of important new findings, given the highly dynamic nature of the pandemic.

1.2 COVID-19 knowledge modeling strategy

1.2.1 "Base model" generation

To initiate the model building process, we queried the Dataome Nucleus with the previously reported SARS-CoV-2 protein interaction map of 332 high confidence interactors (Gordon et al., 2020) as a "seed" for knowledge expansion. To optimize the specificity of our analysis, we focused on the molecular determinants of host cells that define them as viral targets and the immediate impact of the viral infection on the host cell response. This was achieved through inclusion of a minimum set of elements defining host cell, host-specific response (e.g., innate immune response) and associated phenotypes (see Figure 1). Domain experts independently inspected resultant data, providing a systematic expansion of the network to include related pathways, protein interactors, and regulatory elements. This so-called "base model" centered on a converging molecular mechanism including the host proteins responsible for virus entry, TMPRSS2 and ACE2, together with significantly differentially down-regulated components of the interferon stimulated genes (ISG) induced by the virus infection (ACE2 and SERPING1) (see Figure 1).

1.2.2 Iterative expansion of the base model

The base model was further expanded through integration of key molecular protagonists associated with COVID-19 pathophysiology and symptomatology (see Figure 1) including:

i) Common disease symptoms (e.g., dry cough, myalgia, anosmia, dysgeusia/ageusia, metallic taste sensation, thick mucus, transient diabetes, silent hypoxia, leukocytopenia, and central nervous system (CNS) manifestations).
ii) Severe manifestations (e.g., Acute Respiratory Distress Syndrome (ARDS), acute lung infection (ALI), lung fibrosis, cardiovascular complications including arrhythmia and acute coronary syndrome (ACS).
iii) Outcome and severity associated risk factors (e.g., age, sex, smoking, air pollution, comorbidities).

Specific symptomatology associated 'pre-models' were then manually defined, driven by the biomedical domain experts who inspected and curated the clinico-molecular data extracted previously *via* the interrogation of the Dataome knowledgebase (see examples in Figure 3). The expert-driven process to elucidate the molecular underpinnings of COVID-19 and the diversity of associated disease phenotypes and risk factors focused on three key goals:

- Unravel the molecular foundations of the systemically observed symptoms
- Assess whether a core innate immune response might explain the multiple post-infection reactions, and
- Identify risk factors and phenotypes towards prioritization of drug candidates.

1.2.3 Pre-model curation and integrated visualization

At each stage of the disease/symptom modeling process, the extracted "pre-models" were inspected and visually modeled using the free open-source software PathVisio, [version 3.2.2 (Kutmon et al., 2015)]. During this process, relevant "pre-model" references were attached to the respective objects in PathVisio and manually complemented with bibliography as necessary. Entities and relationships associated with each mechanism/ symptom were presented by a JavaScript-animated SVG image. Graphical renditions were produced *via* the PathVisio program, with data outputs in GPML and SVG format.

1.3 The web-based knowledge explorer

To enable the effective exploration of the key findings by the community regarding the mechanisms that possibly underlie COVID-19 symptoms and the associated evidence summaries, we developed a dedicated web-based interface providing a comprehensive and fully interactive COVID-19 disease model. Results are also provided in the format of a printable Table (see Supplementary Table S1) An expanded version of this Table is found in the sister article to this one focusing in detail on the molecular hypotheses and their validation status1.

1.3.1 Whole patient pre-model integration

Linkage of the curated pre-models through related molecular protagonists resulted in a whole patient COVID-19 disease model that connects central molecular disease mechanisms (namely, aberrant contact activation system (CAS) and ACE2/ DAK/B1R signaling) (see Figures 4, 5) to eight core pathogenic processes: 1) inflammatory signaling, 2) coagulation, 3) barrier permeability, 4) senescence, 5) autoimmunity, 6) fibrogenic signaling, 7) nociception and 8) exocytosis. The model is completed by functionally intersecting these mechanisms with respective symptoms, associated pathogenic pathways and affected organ-systems (see Figure 4 and Supplementary Table S1). The dynamic nature of the global developments around the
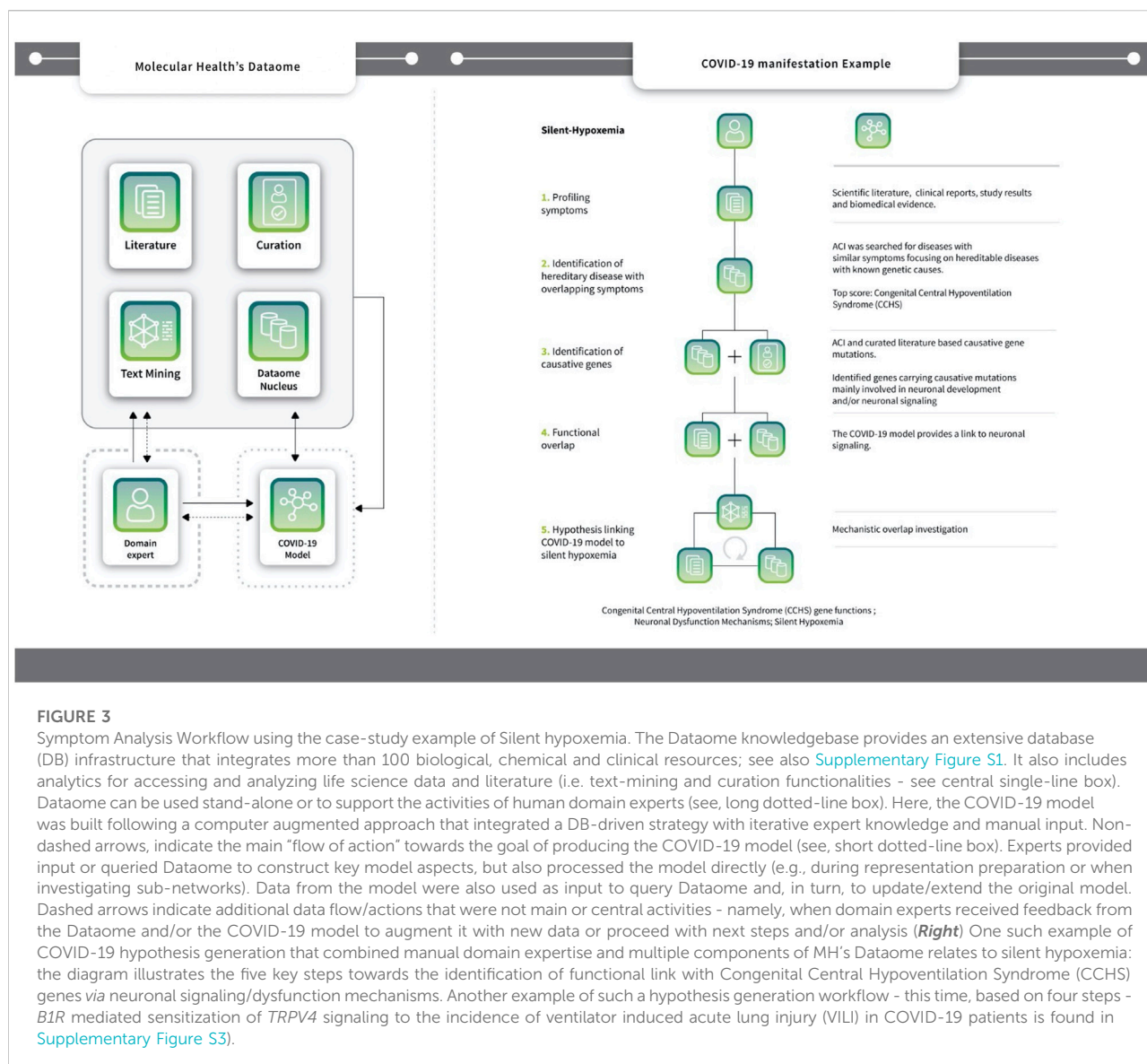
FIGURE 3
Symptom Analysis Workflow using the case-study example of Silent hypoxemia. The Dataome knowledgebase provides an extensive database (DB) infrastructure that integrates more than 100 biological, chemical and clinical resources; see also Supplementary Figure S1. It also includes analytics for accessing and analyzing life science data and literature (i.e. text-mining and curation functionalities - see central single-line box). Dataome can be used stand-alone or to support the activities of human domain experts (see, long dotted-line box). Here, the COVID-19 model was built following a computer augmented approach that integrated a DB-driven strategy with iterative expert knowledge and manual input. Non-dashed arrows, indicate the main "flow of action" towards the goal of producing the COVID-19 model (see, short dotted-line box). Experts provided input or queried Dataome to construct key model aspects, but also processed the model directly (e.g., during representation preparation or when investigating sub-networks). Data from the model were also used as input to query Dataome and, in turn, to update/extend the original model. Dashed arrows indicate additional data flow/actions that were not main or central activities - namely, when domain experts received feedback from the Dataome and/or the COVID-19 model to augment it with new data or proceed with next steps and/or analysis (**Right**) One such example of COVID-19 hypothesis generation that combined manual domain expertise and multiple components of MH's Dataome relates to silent hypoxemia: the diagram illustrates the five key steps towards the identification of functional link with Congenital Central Hypoventilation Syndrome (CCHS) genes *via* neuronal signaling/dysfunction mechanisms. Another example of such a hypothesis generation workflow - this time, based on four steps - *B1R* mediated sensitization of *TRPV4* signaling to the incidence of ventilator induced acute lung injury (VILI) in COVID-19 patients is found in Supplementary Figure S3).

pandemic provided a constant flow of updated observations that were used as real-time validation of the resulting model's core1. Finally, domain expert curation ensured continuous synchronization with new SARS-CoV-2 data and Dataome content.

1.3.2 Webserver modules

A Flask micro web framework serves as the web application displaying the interconnected "sub-models". The application presents an SVG-based graphic for each sub-model, animated by the d3. js JavaScript library. To project the most actual and comprehensive information, the collection of citations supporting the relations of each "sub-model", may range from peer-reviewed articles to very recent conference content (e.g., abstracts) and even ad-hoc communications. This information is provided as auxiliary information *via* respective animated components.

1.4 Results

Constructed during the early stages of the COVID-19 pandemic in 2020, our COVID-19 knowledge model revealed that the multitude and complexity of observed, and seemingly disparate clinical phenotypes may be linked to the pleiotropic activity of eight core molecular mechanisms involved in the host response (see Supplementary Table S1 and 1). In addition, the model revealed functionally connected mechanisms across multiple organ systems allowing for the identification of novel hypotheses for both viral dependent and independent disease
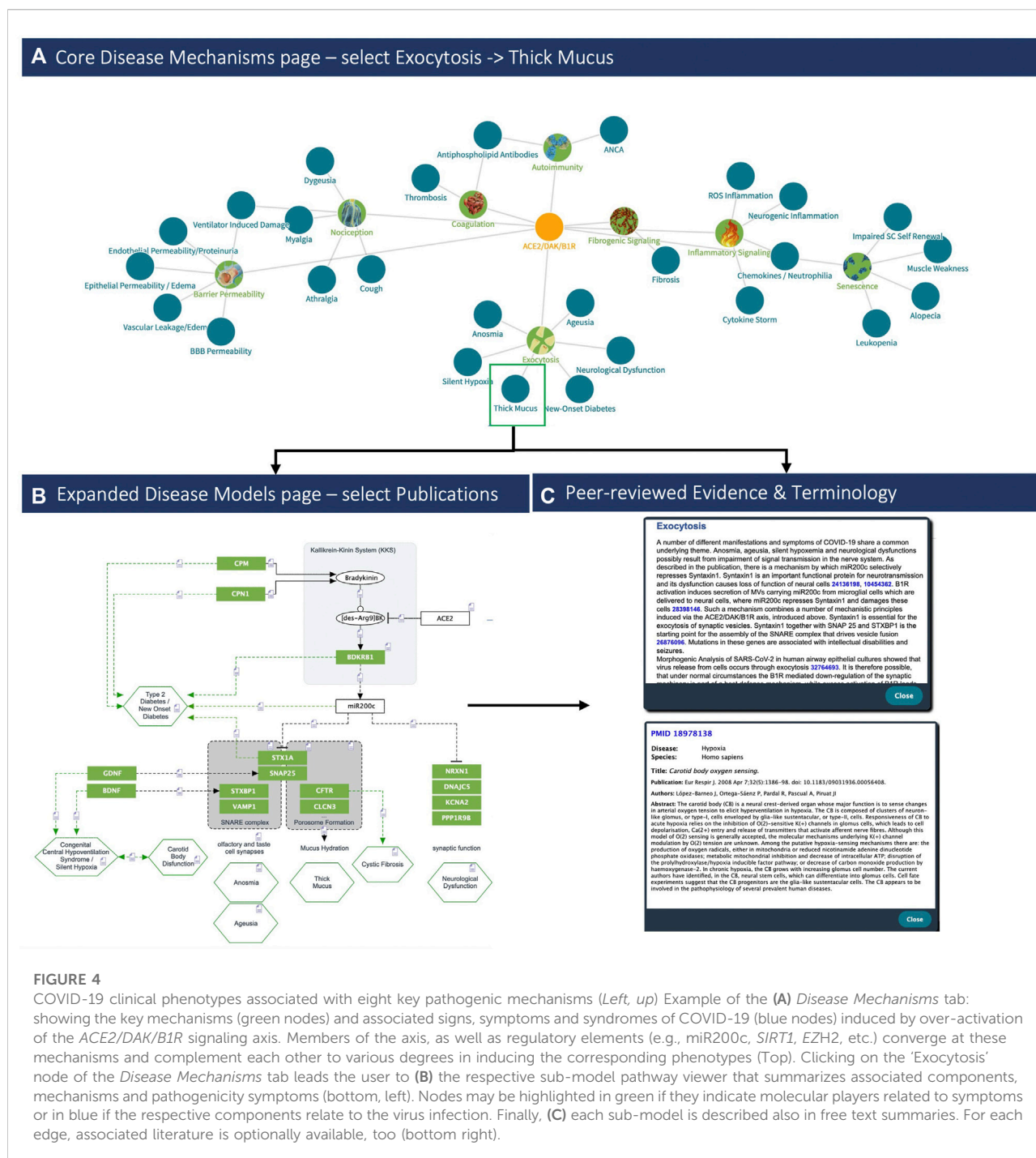
FIGURE 4

COVID-19 clinical phenotypes associated with eight key pathogenic mechanisms (*Left, up*) Example of the **(A)** *Disease Mechanisms* tab: showing the key mechanisms (green nodes) and associated signs, symptoms and syndromes of COVID-19 (blue nodes) induced by over-activation of the *ACE2/DAK/B1R* signaling axis. Members of the axis, as well as regulatory elements (e.g., miR200c, *SIRT1*, *EZ*H2, etc.) converge at these mechanisms and complement each other to various degrees in inducing the corresponding phenotypes (Top). Clicking on the 'Exocytosis' node of the *Disease Mechanisms* tab leads the user to **(B)** the respective sub-model pathway viewer that summarizes associated components, mechanisms and pathogenicity symptoms (bottom, left). Nodes may be highlighted in green if they indicate molecular players related to symptoms or in blue if the respective components relate to the virus infection. Finally, **(C)** each sub-model is described also in free text summaries. For each edge, associated literature is optionally available, too (bottom right).

mechanisms and associated pharmacologic targets that may warrant further investigation for drug repurposing and/or development efforts1.

1.4.1 The COVID-19 explorer

The COVID-19 Explorer represents a comprehensive COVID-19 disease model linking curated molecular

protagonists at the symptom specific level. The COVID-19 Explorer is openly available *via* an interactive web-based interface at: https://covid19.molecularhealth.com.

1.4.1.1 Organization and functionality

The interactive interface provides summarized views between molecular mechanisms and disease processes linked
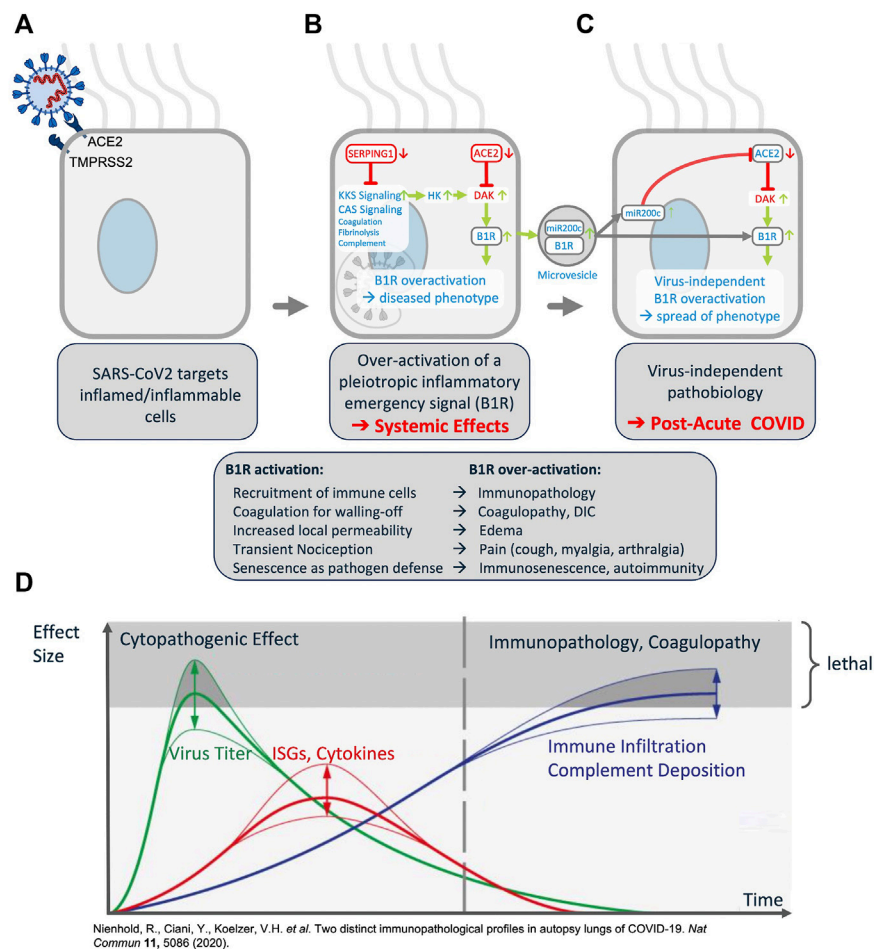
FIGURE 5
Synopsis of the COVID-19 Disease Model. **(A)** The virus targets host cells expressing *ACE2* and *TMPRSS2*, active components of the KKS. **(B)** Within the cell the virus induces downregulation of *SERPING1* and *ACE2*. *SERPING1* downregulation induces activation of the CAS and KKS. *ACE2* downregulation results in accumulation of *DAKs*. Excess *DAKs* activate *B1R*, triggering constitutive activation and auto-amplification. Under normal conditions *B1R* activation triggers a fast, transient inflammatory emergency reaction inducing neutrophil and leukocyte recruitment and infiltration, opening of the epithelial/endothelial barriers for their transmigration, coagulation for local walling-off, fibrogenesis for wound healing, senescence as host defense mechanism and transient nociception. Constitutive activation of this system leads to excess inflammatory signaling, epithelial/endothelial barrier breakdown, induction of thrombosis, fibrosis, pain and other effects. **(C)** *B1R* signaling induces the formation of MVs bearing *B1R* and miR200c. These are transferred to target cells, inducing further expression of miR200c, which leads to *ACE2* downregulation and formation of excess *DAKs* which signal *via B1R*. Thus, MV transfer *via* auto-activation and amplification of the *B1R* system may trigger the virus independent propagation of an inflammatory phenotype. **(D)** Schematic time course of COVID-19 lung disease based on lung autopsy findings (illustration adapted with permission from (Nienhold et al., 2020)). This time course is consistent with the proposed model: early in the disease, an ISGhigh lung profile is observed, with high viral load, high expression of cytokines and ISGs, and sparse immune infiltrates. Late in the disease, an ISGlow lung profile is observed, with low viral load, low local expression of cytokines and ISGs, and strong infiltration of macrophages and lymphocytes. Patients who die early are not able to adequately control SARS-CoV-2, while patients who die late suffer from coagulopathy and immunopathology. This indicates that at a later stage the disease becomes virus independent.

with respective symptoms, associated pathogenic pathways and affected organ-systems. More specifically, the user interface consists of eight components that allow users to explore a number of detailed visualizations and annotated graphics (see Figure 2):

- Organs & Symptoms: consists of a three-dimensional human model that graphically summarizes the major

organs and symptoms associated with COVID-19—the model contains hyperlinks leading to the respective underlying biomedical models.

- Pathogenic Mechanisms: contains an interactive network diagram summarizing connections between COVID-19 symptoms, affected organs, functional mechanisms, and key signaling axes—nodes are hyperlinked to the respective views with the details of each model.

- Disease Model: a schematic synopsis of the central mechanisms identified pertaining to cell damage, even in cells not directly infected by the SARS-CoV-2 virus.
- Expanded Model: visual view of pathways, molecular mechanisms, and biological systems affected by the virus—nodes hyperlink to lists of respective components involved (whether listed under Symptom or Functional Mechanisms, or both).
- Symptom Mechanisms: list of components connected to each molecular model (symptoms, organs, associated genes and proteins, references).
- Functional Mechanisms: list of mechanisms triggered by the SARS-CoV-2 virus, and the related genes and proteins.
- Publications: list of hundreds of citations underlying the COVID-19 Explorer—each reference contains hyperlinks to the respective model page and PubMed record.
- Terminology Reference: organ, disease, symptom and mechanism term groupings as considered for the purpose of this work and interface.

These different views are aimed at enabling multiple use case scenarios, relevant to both the research community and clinical users. The full scope of "COVID-19 Explorer" features and utilities is described in detail within the accompanying User documentation: https://covid19.molecularhealth.com/MHCoronaExplorer_QuickGuide.pdf.

1.4.1.2 Using the explorer: Phenotype associations and predictive potential

Our model was developed during the initial stages of the pandemic. At this time, new and seemingly unrelated clinical phenotypes were concurrently described. Using the Explorer, these could be immediately linked to core mechanisms (e.g., barrier permeability or exocytosis). For instance, first reports of silent hypoxemia emerged in April 2020. The symptoms of silent hypoxemia were mapped against the symptomatology of heritable diseases, identifying Congenital Central Hereditary Hypoventilation Syndrome (CCHS) as a phenotypically related disease (see Figure 3) (Brock et al., 2022). CCHS is caused by dysfunction of the exocytosis machinery in oxygen sensing cells, providing a direct link to our model (Supplementary Material S1, Figure 3) (see also Figure 4). A similar link between clinical phenotypes and our model was established upon the first reports on endotheliitis (Varga et al., 2020), vasculitis and the role of micro-thrombotic events in severe disease (see Supplementary Table S1). Here too, direct links between the molecular etiology of the observed symptoms and our model could be drawn and strikingly, by mid-April, first cases of new onset KWD-like disease were reported in children with COVID-19, thereby also suggesting the predictive potential of the model.

1.4.1.3 Key aspects of the COVID-19 knowledge model

The COVID-19 Explorer and associated knowledge model highlight the key molecular players involved in host responsible for SARS-CoV-2 entry (*ACE2*, *TMPRSS2*), and the host factors of the ISG response that are specifically dysregulated by the SARS-CoV host interaction (*ACE2*, *SERPING1*) (see Figure 5). In turn, these are seen to converge on unifying pleiotropic signaling pathways comprising Renin-Angiotensin System (RAS) and Kallikrein Kinin System (KKS) as part of the Contact Activation System (CAS). Concurrent downregulation of *ACE2* and *SERPING1* may then reciprocally amplify the deregulation of KKS thereby generating a `"perfect storm", which may lead to extreme over-activation of downstream signaling, especially in acute COVID-19. Finally, the model indicates that viral perturbation of eight key mechanisms, alone or in combination, may contribute to the pathogenesis of primary COVID-19 phenotypes.

A full analytical overview of the generated model is summarized in Supplementary Table S1 and made available *via* our COVID-19 Explorer. In addition, extended information about specific mechanisms, pathways, clinical phenotypes and insights into the clinico-molecular hypotheses derived from our model are reported in the associated back-to-back publication (Brock et al., 2022).

1.4.1.4 Host factors mediating SARS-CoV-2 infection

The model identified a converging molecular landscape, delineating host-factor responses to SARS-CoV-2 *via* host proteins responsible for virus entry, as well as significantly differentially down-regulated components of the set of ISG induced by virus infection. More specifically, the model indicates that the cell-entry mechanism (see Figure 5A) and disease-specific ISG signature provides three key active components of SARS-CoV infected cells (*TMPRSS2*, *ACE2* and *SEPRING1*) that may functionally converge in the same pleiotropic signaling systems, namely the CAS and KKS pleiotropic signaling (see Figure 5B).

1.4.1.5 Host-response driven disease mechanisms

The model reveals that SARS-CoV-2 targets cells expressing constituents of a highly inducible inflammatory signaling system causing its excess activation—the pleiotropic nature of this system appears to underpin the diverse clinical manifestations of COVID-19. Importantly, the model indicates also a possible mechanism through which a disease phenotype may be propagated even in the absence of the original viral trigger (see Figure 5C). More specifically:

a) Convergence in pleiotropic KKS may dysregulate the ACE2-DAK-B1R axis, triggering systemic disease (see Figure 5B).
b) Auto-induction of B1R bearing microvesicles (MVs) and interplay with the regulatory miR200c may provide a feed-

forward loop decoupling molecular pathogenesis from virus load (see 5C).

Indeed, analysis of post-mortem COVID-19 lung suggests two distinct stages of disease-progression (see Figure 5D). Early disease has high viral-load and high expression of cytokines and ISGs and sparse immune infiltrates, while in late disease, low viral loads, low local expression of cytokines and ISGs, and strong infiltration of macrophages and lymphocytes prevails. Patients who die early are unable to control SARS-CoV-2, while patients who die later suffer from diffuse tissue-damage and immunopathology (Nienhold et al., 2020) suggesting that late disease stage pathogenesis is apparently decoupled from acute viral-load.

1.4.1.6 Multiple pathologies of COVID-19 phenotypes may converge mechanistically

The model demonstrates that excess activation of inflammatory signaling may turn productive inflammatory response and recruitment of immune cells into a detrimental cytokine storm and immunopathology. Importantly, such mechanisms can be triggered by an imbalance in *ACE2-DAK-B1R* signaling and associated regulatory components (e.g., miR200c or *SIRT1*) (see Figure 5 and Supplementary Table S1).

Altogether, the model suggests that dysregulation/disturbed homeostasis of eight mechanisms, alone or in combination, may contribute to the pathogenesis of major COVID-19 phenotypes. The multiple and seemingly unrelated clinical manifestations of COVID-19, including common disease symptoms (e.g., dry cough, myalgia, anosmia, transient diabetes and silent hypoxia) and severe manifestations (e.g., ARDS, lung fibrosis, acute coronary syndromes and thromboembolic events) may largely be linked to the pleiotropic activity of these core molecular players and mechanisms involved in the host response (see Figures 3, 4). Interestingly, the model reveals that some rarer phenotypes can be matched to other diseases sharing the same symptoms. Silent hypoxemia, for instance, causes the same symptoms as CCHS. The molecular pathologies of both converge on the same molecular mechanism (see Figures 3, 4).

1.4.1.7 Validations, diagnostic and therapeutic implications

The rapid flow of new knowledge and information during the course of the pandemic enabled us to directly examine model-derived hypotheses in real-time, with several emerging as clinically validated or incorporated in specific multicenter studies. For example, COVID-19 organ damage often cannot be entirely explained by the virus' organ tropism and local viral load. In COVID-19 associated kidney disease, for instance, viral load is low and unevenly distributed (Puelles et al., 2020; Su et al., 2020), and cannot explain the extensive kidney damage (Wang et al., 2021). These findings are in line

with the proposed COVID-19 model. Currently, there are also no diagnostic tools that are associated with systemic pathophysiology of COVID-19. A systemic, virus independent disease mechanism requires systemic distribution of a signal that bears the potential to induce a broad spectrum of pathophysiological dysregulation in a variety of organs/tissues. A derailed pleiotropic signaling system such as the KKS/*B1R* signaling axis constitutes a likely candidate.

In this context, circulating MVs enriched in *B1R* and mir200c or circulating mir200c itself could serve as biomarker candidates. Indeed, serum, plasma or PBMC levels of miR200c has been identified as a diagnostic biomarker candidate in different COVID-19 related disease contexts, namely Kawasaki Disease (KWD) (Zhang et al., 2017), pneumonia (Liu et al., 2017), interstitial lung disease (Jiang et al., 2017), COPD (Cao et al., 2014) and fibrosis of multiple tissues (Yang et al., 2012; Ramachandran et al., 2013; Chen et al., 2017). It has also been shown that upregulated circulating miR-200c in plasma may increase the risk of obese individuals to severe COVID-19 (Papannarao et al., 2021).

In addition, the generated COVID-19 disease mechanism model contains target structures with implications for host-directed therapies. According to the model, pharmaceutically tractable target structures include the KKS at multiple levels such as *Kallikrein* inhibitors, *SerpinG1* enzyme replacement or *B1R* inhibitors for example, may represent a preferred therapeutic target, having been under evaluation for the treatment of hyperalgesia and osteoarthritis during the past decade. However, to date, no relevant clinical results have been published and most of the reported trials are inactive or have been stopped or suspended (see (Brock et al., 2022)). Interestingly, the induction of *B1R* also appears to be sensitive to treatment with dexamethasone. Modulation of *ACE2* activity represents another potential candidate for host directed therapy either by direct activation or indirectly by induction (e.g., through *SIRT-1* activators such as Melatonin, Resveratrol and Metformin or activators of *PPARγ* (Dambha-Miller et al., 2020)). Several of these are currently investigated and readouts of such trials utilizing drugs that potentially induce *ACE2* expression (Dambha-Miller et al., 2020) are summarized in the associated back-to-back paper (Brock et al., 2022).

3 Discussion

Our goal in creating the web-based *COVID-19 Explorer* was to provide an easily usable resource summarizing the key symptoms and molecular mechanisms associated with COVID-19 disease at the whole-patient level. To achieve this, we employed the Dataome technology in an iterative, expert-driven approach to:

- Build a comprehensive COVID-19 model
- Examine molecular mechanisms of specific individual symptoms
- Annotate relevant molecular components and pathways with supporting literature and observed evidence
- Visualize these findings, and share them with the community via a webserver that allows review of linked descriptive summaries
- Make hypotheses regarding the molecular etiology of both symptoms and disease and associated therapeutic strategies
- Monitor for the appearance of new findings that (in-)validate our original hypotheses, or modify accordingly

A key value of the COVID-19 Explorer is the way it permits the capture and contextualization of disease specific clinical and molecular information. With a plurality of emergent symptoms reported weekly during March-April 2020, it was critical to connect such clinical phenotypes and risk factors with both existing (e.g., from SARS-CoV-1) and emergent molecular findings. Our work highlights the speed at which bespoke clinico-molecular models can be built, emphasizing the important role that computer augmented disease modeling by domain experts can play in response to global health emergencies. Importantly, the COVID-19 model was built and organized into an updateable data framework driven by data capture, integration and curation activities. The model details molecular factors and systems that may drive COVID-19 and links the pleiotropic symptomology to possible underlying molecular pathology mechanisms. In comparison to other works, the ability to connect such disparate information layers facilitates a unique view in approaching COVID-19 at a whole patient, system-based level (Brock et al., 2022).

The COVID-19 Explorer is one of several valuable COVID-19 knowledge resources to emerge during the course of the pandemic. Prime examples of complementary open-source initiatives include COVID-19 UniprotKB (UniProt Consortium, 2021), Open Targets' COVID-19 Target Prioritization Tool (Carvalho-Silva et al., 2018), and Reactome's SARS-CoV-2 (COVID-19) infection pathway (Acencio, 2020). While these resources add significant value to our armamentarium of COVID-19 focused knowledgebases, the COVID-19 Explorer is the first to contextualize such data at the level of whole patient symptomatology. https://blog.opentargets.org/covid-19-target-prioritisation-tool-released/ https://reactome.org/about/news/161-version-74-released.

Direct comparison of resources complicated by the diversity of starting motivations and utilities. For example, a recent large scale structural analysis has provided unique insights into complex and potentially important mechanisms regarding COVID-19, including viral protein self-assembly, molecular mimicry of human proteins, reversal of post-translational modifications, blockage of host translation, and the disabling of host defenses. In another instance, hospitalized COVID-19 patients were found to be positively correlated with (auto-)

immune responses, not only providing confirmatory observations to our findings, but also highlighting the importance of laboratory-based validation of model-based hypotheses. By providing a whole patient perspective on the molecular etiology of COVID-19 symptomatology, our COVID-19 Explorer resource is complementary to the value.

Despite the broad utility of the whole patient model, several limitations exist. First, our integrated COVID-19 model was generated by a small team of biomedical experts, exploiting the content, technologies and processes of our proprietary Dataome solution. To do so, we partially relied on results gained from the extensive research that has been published on the original SARS coronavirus. For instance, the pathogen specific impact on the differential expression of ISG's has been taken from work on SARS-CoV. We then used structured interaction and pathway data, as well as TDM and subsequent manual curation to identify upstream and downstream processes. All facts are supported by peer-reviewed literature and are transparently reported, though we cannot be certain about the accuracy or reproducibility of these results. We also screened for phenotypes that are associated with molecular perturbations (genetic or pharmacologic) of members within these regulatory networks. However, we limited this work to proposed key factors and the immediate regulatory elements.

Second, we also systematically collected COVID-19 associated phenotypes and risk profiles affecting different organ systems. Since a majority of these phenotypes also occur in other disease contexts, such as symptomatically related hereditary diseases, we identified the molecular mechanisms involved in their respective molecular etiologies. We then screened for convergence/divergence between the disturbed host mechanisms and those underlying the COVID-19 phenotypes. The advantage of such an approach is that key factors and interrelated pleiotropic regulatory concepts are quickly identified. However, our results likely need to be complemented by further systematic extension of the work. It seems logical, that the manifold symptoms and manifestations of COVID-19 result from the dysregulation of a few key elements converging in a pleiotropic mechanism which is connected to completely different phenotypes, depending on the individual tissue and organ context. While in many cases there is also multiple independent evidence linking the disturbed host system to the mechanisms underlying specific symptoms, the causal association with SARS-CoV-2 has yet to be proven. In that respect, we regard our approach as an effort to aggregate and interlink facts and connect existing knowledge that results in a defendable and testable hypothetical model that can inform future targeted research.

In terms of future directions, the current model is based on the molecular phenotype of the cells targeted and infected by SARS-CoV(-2) and provides a basis for explaining the diverse clinical phenotypes, observed risk factors and tractable strategies for therapeutic interventions and prevention.

However, it does not provide a detailed mechanistic model for how direct pathogen host interactions induce and modulate the observed pathogen specific host response. As our initial approach was to reduce complexity by focusing on molecular phenotypes, we now have a consistent and testable base model that permits the systematic integration of additional data such as the full pathogen-host interactome. The model will also be further expanded by the +1-interaction level. The resulting network will then be further enriched by associated disease phenotypes from various sources, disease mechanisms and mode of actions of drugs targeting any of the components in the model. The resulting graph database will lend itself to the application of advance AI analytics to identify hypernodes and, eventually, mechanisms defining causality. This will provide novel angles to detect new strategies for intervention or for the comprehensive evaluation of existing interventional programs.

While our COVID-19 model was initially based on primarily data-mining hypotheses, data from more recent developments have helped update the initial model and also validate multiple predictions. For example, for some clinical phenotypes (e.g., KWD-like syndromes) predicted by our model it appeared that the clinical reality was ultimately superseding our model in real-time given the ongoing, massive global pandemic thereby providing timely validation (Brock et al., 2022). Moreover, our model revealed functionally connected mechanisms across various organ-systems, identified hypotheses for both, viral-dependent and -independent disease mechanisms, and associated pharmacologic targets that may warrant further evaluation (Brock et al., 2022). In this context our model, combined with other laboratory and/or real-world evidence, can be used both as a hypothesis generation and validation point regarding observed experimental or clinical findings.

In summary, our COVID-19 knowledge model links key molecular players in COVID-19 disease pathophysiology to common symptoms, severe manifestations and outcome/ severity-associated risk factors at the whole patient level. We have validated that our COVID-19 Explorer provides a valuable and unique resource to support clinical and translational research audiences in hypothesis generation for new diagnostic and therapeutic strategies. We also anticipate that as the trajectory of scientific discovery continues to correlate the rate of technological advancement, biomedical research will grow increasingly dependent on similar human-focused and systems-based clinico-molecular information systems, capable of summarizing diverse findings in the form of intuitive

whole patient disease models. Our work suggests that computer-augmented modelling of such knowledge by domain experts currently represents the most reliable approach in this regard. Moreover, it also provides a structured format through which future publication of expert reviews may be approached.

Data availability statement

Associated data are downloadable within dedicated sections of the web-based COVID-19 Explorer interface, accessible at http://covid19.molecularhealth.com. Additional data pertaining to this submission may also be made available upon reasonable request. In this context, it is important to emphasize that the development of the COVID-19 Explorer was enabled by MH's Dataome platform, an expansive biomedical data and analytics infrastructure that contains a diverse and integrated array of open-source, proprietary and commercial data sources (over 100) and software (some licensed from third parties).

Author contributions

SB conceived original idea of a host response-centric COVID-19 model, initiated and designed the study, and performed the analysis. KH provided technical support in the development of the COVID-19 Explorer webserver. FD and AS provided domain insight and coordinated expert guidance. ME and SH performed clinical-translational interpretation of the results, provided critical review and input to the manuscript. SB and ME wrote the initial draft. TS performed data management, provided writing review and editing. DJ provided scientific support, writing review and editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmmed.2022.1035215/full#supplementary-material

References

Acencio, M. L. (2020). *SARS-CoV-2 Infection reactome*. WHO. https://reactome.org/content/detail/person/0000-0002-8278-240X

Armaiz-Pena, G., Allen, J. K., Cruz, A., Stone, R. L., Nick, A. M., Lin, Y. G., et al. (2013). Src activation by β-adrenoreceptors is a key switch for tumour metastasis. *Nat. Commun.* 4, 1403. doi:10.1038/ncomms2413

Bohnert, R., Vivas, S., and Jansen, G. (2017). Comprehensive benchmarking of SNV callers for highly admixed tumor data. *PLoS One* 12, e0186175. doi:10.1371/journal.pone.0186175

Brock, S., Jackson, D. B., Soldatos, T. G., Hornischer, K., Schäfer, A., Diella, F., et al. (2022). Whole patient knowledge modeling of COVID-19 symptomatology reveals common molecular mechanisms. *Front. Mol. Med.* 2:1035290. doi:10.3389/fmmed.2022.1035290

Brown, E. G. (2004). Using MedDRA: Implications for risk management. *Drug Saf.* 27, 591–602. doi:10.2165/00002018-200427080-00010

Cao, Z., ZhaNgN., Lou, T., Jin, Y., Wu, Y., Ye, Z., et al. (2014). microRNA-183 down-regulates the expression of BKCaβ1 protein that is related to the severity of chronic obstructive pulmonary disease. *Hippokratia* 18, 328–332.

Carvalho-Silva, D. O., Pierleoni, A., Pignatelli, M., Ong, C., Fumis, L., Karamanis, N., et al. (2018). Open targets platform: New developments and updates two years on. *Nucleic Acids Res.* 47, D1056–D1065. doi:10.1093/nar/gky1133

Chen, J., Cai, J., Du, C., Cao, Q., Li, M., and Liu, B. (2017). Recent advances in miR-200c and fibrosis in organs. *Zhong Nan Da Xue Xue Bao Yi Xue Ban.* 42, 226–232. doi:10.11817/j.issn.1672-7347.2017.02.018

Cherny, N. I., Dafni, U., Bogaerts, J., Latino, N. J., Pentheroudakis, G., Douillard, J. Y., et al. (2017). ESMO-magnitude of clinical benefit scale version 1.1. *Ann. Oncol.* 28, 2340–2366. doi:10.1093/annonc/mdx310

Dambha-Miller, H., Albasri, A., Hodgson, S., Wilcox, C. R., Khan, S., Islam, N., et al. (2020). Currently prescribed drugs in the UK that could upregulate or downregulate ACE2 in COVID-19 disease: A systematic review. *BMJ Open* 10, e040644. doi:10.1136/bmjopen-2020-040644

Fernandez, S., Lenoir, C., Samer, C., and Rollason, V. (2020). Drug interactions with apixaban: A systematic review of the literature and an analysis of VigiBase, the world health organization database of spontaneous safety reports. *Pharmacol. Res. Perspect.* 8, e00647. doi:10.1002/prp2.647

Gonzalez, N., Speir, M. L., Schmelter, D., Rosenbloom, K. R., and Raney, B. J. (2021). The UCSC genome browser database: 2021 update. *Nucleic Acids Res.* 49, D1046–d1057. doi:10.1093/nar/gkaa1070

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468. doi:10.1038/s41586-020-2286-9

Harrison, S. M., Biesecker, L. G., and Rehm, H. L. (2019). Overview of specifications to the ACMG/AMP variant interpretation guidelines. *Curr. Protoc. Hum. Genet.* 103, e93. doi:10.1002/cphg.93

Hirotsu, Y., Schmidt-Edelkraut, U., Nakagomi, H., Sakamoto, I., Hartenfeller, M., Narang, R., et al. (2020). Consolidated BRCA1/2 variant interpretation by MH BRCA correlates with predicted PARP inhibitor efficacy association by MH guide. *Int. J. Mol. Sci.* 21, E3895. doi:10.3390/ijms21113895

Howe, K. L. E., Achuthan, P., Allen, J., Allen, J., Armean, I. M., Azov, A. G., et al. (2021). Ensembl 2021. *Nucleic Acids Res.* 49, D884–d891. doi:10.1093/nar/gkaa942

Humphreys, B. L., Del Fiol, G., and Xu, H. (2020). The UMLS knowledge sources at 30: Indispensable to current research and applications in biomedical informatics. *J. Am. Med. Inf. Assoc.* 27, 1499–1501. doi:10.1093/jamia/ocaa208

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al. (2020). The reactome pathway knowledgebase. *Nucleic Acids Res.* 48, D498–d503. doi:10.1093/nar/gkz1031

Jiang, Z., Tao, J. H., Zuo, T., Li, X. M., Wang, G. S., Fang, X., et al. (2017). The correlation between miR-200c and the severity of interstitial lung disease associated with different connective tissue diseases. *Scand. J. Rheumatol.* 46, 122–129. doi:10.3109/03009742.2016.1167950

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. K. E. G. G. (2021). Kegg: Integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D545–d551. doi:10.1093/nar/gkaa970

Kim, S. P., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* 49, D1388–D1395. doi:10.1093/nar/gkaa971

Koh, W. J., Anderson, B. O., and Carlson, R. W. (2020). NCCN resource-stratified and harmonized guidelines: A paradigm for optimizing global cancer care. *Cancer* 126 (10), 2416–2423. doi:10.1002/cncr.32880

Krawczyk, P., and Święcicki, Ł. (2020). ICD-11 vs. ICD-10 - a review of updates and novelties introduced in the latest version of the WHO International Classification of Diseases. *Psychiatr. Pol.* 54, 7–20. doi:10.12740/PP/103876

Kutmon, M., van Iersel, M. P., Bohler, A., Kelder, T., Nunes, N., Pico, A. R., et al. (2015). PathVisio 3: An extendable pathway analysis toolbox. *PLoS Comput. Biol.* 11, e1004085. doi:10.1371/journal.pcbi.1004085

Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., et al. (2020). ClinVar: Improvements to accessing data. *Nucleic Acids Res.* 48, D835–D844. doi:10.1093/nar/gkz972

Lipscomb, C. E. (2000). Medical subject headings (MeSH)Medical subject headings (MeSH). *Bull. Med. Libr. Assoc.* 88, 265–266.

Liu, Q., Du, J., Yu, X., Xu, J., Huang, F., Li, X., et al. (2017). miRNA-200c-3p is crucial in acute respiratory distress syndrome. *Cell. Discov.* 3, 17021. doi:10.1038/celldisc.2017.21

Maglott, D., Ostell, J., Pruitt, K. D., and Tatusova, T. (2011). Entrez gene: Gene-centered information at NCBI. *Nucleic Acids Res.* 39, D52–D57. doi:10.1093/nar/gkq1237

Merabti, T., Abdoune, H., Letord, C., Sakji, S., Joubert, M., and Darmoni, S. J. (2011). Mapping the ATC classification to the UMLS metathesaurus: Some pragmatic applications. *Stud. Health Technol. Inf.* 166, 206–213.

Nienhold, R., Ciani, Y., Koelzer, V. H., Tzankov, A., Haslbauer, J. D., Menter, T., et al. (2020). Two distinct immunopathological profiles in autopsy lungs of COVID-19. *Nat. Commun.* 11, 5086. doi:10.1038/s41467-020-18854-2

O'Leary, N., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi:10.1093/nar/gkv1189

Papannarao, J. B., Schwenke, D. O., Manning, P., and Katare, R. (2021). Upregulated miR-200c may increase the risk of obese individuals to severe COVID-19. medRxiv.

Pradeep, S., Huang, J., Mora, E. M., Nick, A. M., Cho, M. S., Wu, S. Y., et al. (2015). Erythropoietin stimulates tumor growth via EphB4. *Cancer Cell.* 28, 610–622. doi:10.1016/j.ccell.2015.09.008

Puelles, V. G., Lutgehetmann, M., Lindenmeyer, M. T., Sperhake, J. P., Wong, M. N., Allweiss, L., et al. (2020). Multiorgan and renal tropism of SARS-CoV-2. *N. Engl. J. Med.* 383, 590–592. doi:10.1056/NEJMc2011400

Ramachandran, S., Ilias Basha, H., Sarma, N. J., Lin, Y., Crippin, J. S., Chapman, W. C., et al. (2013). Hepatitis C virus induced miR200c down modulates FAP-1, a negative regulator of Src signaling and promotes hepatic fibrosis. *PLoS One* 8, e70744. doi:10.1371/journal.pone.0070744

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., et al. (2021). Database resources of the national center for biotechnology information in 2023. *Nucleic Acids Res.* 49, gkac1032–d17. doi:10.1093/nar/gkac1032

Schell, M. J., Yang, M., Teer, J. K., Lo, F. Y., Madan, A., Coppola, D., et al. (2016). A multigene mutation classification of 468 colorectal cancers reveals a prognostic role for APC. *Nat. Commun.* 7, 11743. doi:10.1038/ncomms11743

Schotland, P., Racz, R., Jackson, D. B., Soldatos, T. G., Levin, R., Strauss, D. G., et al. (2021). Target adverse event profiles for predictive safety in the postmarket setting. *Clin. Pharmacol. Ther.* 109, 1232–1243. doi:10.1002/cpt.2074

Soldatos, T. G., and Jackson, D. B. (2019). Adverse event circumstances and the case of drug interactions. *Healthc. (Basel)* 7, E45. doi:10.3390/healthcare7010045

Soldatos, T. G., Taglang, G., and Jackson, D. B. (2018). *In silico* profiling of clinical phenotypes for human targets using adverse event data. *High. Throughput.* 7, E37. doi:10.3390/ht7040037

Su, H., Yang, M., Wan, C., Yi, L. X., Tang, F., Zhu, H. Y., et al. (2020). Renal histopathological analysis of 26 postmortem findings of patients with COVID-19 in China. *Kidney Int.* 98, 219–227. doi:10.1016/j.kint.2020.04.003

UniProt Consortium (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–d489. doi:10.1093/nar/gkaa1100

Ursu, O. D., Holmes, J., Bologa, C. G., Yang, J. J., Mathias, S. L., Stathias, V., et al. (2018). DrugCentral 2018: An update. *Nucleic Acids Res.* 47, D963–d970. doi:10.1093/nar/gky963

Varga, Z., Flammer, A. J., Steiger, P., Haberecker, M., Andermatt, R., Zinkernagel, A. S., et al. (2020). Endothelial cell infection and endotheliitis in COVID-19. *Lancet* 395, 1417–1418. doi:10.1016/S0140-6736(20)30937-5

Wang, M., Xiong, H., Chen, H., Li, Q., and Ruan, X. Z. (2021). Renal injury by SARS-CoV-2 infection: A systematic review. *Kidney Dis.* 7, 100–110. doi:10.1159/000512683

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–d1082. doi:10.1093/nar/gkx1037

Yang, S., Banerjee, S., de Freitas, A., Sanders, Y. Y., Ding, Q., Matalon, S., et al. (2012). Participation of miR-200 in pulmonary fibrosis. *Am. J. Pathol.* 180, 484–493. doi:10.1016/j.ajpath.2011.10.005

Yao, X., Tsang, T., Sun, Q., Quinney, S., and Zhang, P. (2020). Mining and visualizing high-order directional drug interaction effects using the FAERS database. *BMC Med. Inf. Decis. Mak.* 20, 50. doi:10.1186/s12911-020-1053-z

Zhang, W., Wang, Y., Zeng, Y., Hu, L., and Zou, G. (2017). Serum miR-200c and miR-371-5p as the useful diagnostic biomarkers and therapeutic targets in Kawasaki disease. *Biomed. Res. Int.* 2017, 8257862. doi:10.1155/2017/8257862

| Frontiers in Molecular Medicine

Whole patient knowledge modeling of COVID-19 symptomatology reveals common molecular mechanisms

Stephan Brock[1], David B. Jackson[1], Theodoros G. Soldatos[1,2], Klaus Hornischer[1], Anne Schäfer[1], Francesca Diella[1], Maximilian Y. Emmert[3,4,5,6]* and Simon P. Hoerstrup[3,4]*

[1]Molecular Health GmbH, Heidelberg, Germany, [2]SRH Hochschule, University of Applied Science, Heidelberg, Germany, [3]Institute for Regenerative Medicine, University of Zurich, Zurich, Switzerland, [4]Wyss Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland, [5]Department of Cardiothoracic and Vascular Surgery, German Heart Institute Berlin, Berlin, Germany, [6]Department of Cardiovascular Surgery, Charité Universitätsmedizin Berlin, Berlin, Germany

Infection with SARS-CoV-2 coronavirus causes systemic, multi-faceted COVID-19 disease. However, knowledge connecting its intricate clinical manifestations with molecular mechanisms remains fragmented. Deciphering the molecular basis of COVID-19 at the whole-patient level is paramount to the development of effective therapeutic approaches. With this goal in mind, we followed an iterative, expert-driven process to compile data published prior to and during the early stages of the pandemic into a comprehensive COVID-19 knowledge model. Recent updates to this model have also validated multiple earlier predictions, suggesting the importance of such knowledge frameworks in hypothesis generation and testing. Overall, our findings suggest that SARS-CoV-2 perturbs several specific mechanisms, unleashing a pathogenesis spectrum, ranging from "a perfect storm" triggered by acute hyper-inflammation, to accelerated aging in protracted "long COVID-19" syndromes. In this work, we shortly report on these findings that we share with the community *via* 1) a synopsis of key evidence associating COVID-19 symptoms and plausible mechanisms, with details presented within 2) the accompanying "COVID-19 Explorer" webserver, developed specifically for this purpose (found at https://covid19.molecularhealth.com). We anticipate that our model will continue to facilitate clinico-molecular insights across organ systems together with hypothesis generation for the testing of potential repurposing drug candidates, new pharmacological targets and clinically relevant biomarkers. Our work suggests that whole patient knowledge models of human disease can potentially expedite the development of new therapeutic strategies and support evidence-driven clinical hypothesis generation and decision making.

KEYWORDS

COVID-19, SARS-CoV-2, molecular mechanisms, evidence-based medicine, hypothesis generation, disease modelling

1 Introduction

Beginning in early 2020, the COVID-19 pandemic has paralyzed the world, with over half a billion infections and >5 million deaths reported by mid-2022 (de Seabra Rodrigues Dias et al., 2022). A multiform systemic disease, COVID-19 presents a multitude of clinical phenotypes afflicting multiple organs and manifesting from mild symptomatology to critical illness (Figure 1A). However, the current knowledge landscape remains both disjointed and diverse, precluding a holistic understanding of COVID-19's complex pathogenesis and associated "long COVID" syndromes. To address this challenge, global holistic clinico-molecular data-mining platforms capable of elaborating whole-patient mechanistic knowledge models are required, to expedite and provide more advanced disease comprehension in response to emergent global health emergencies such as the COVID-19 pandemic.

Recently, we utilized a precision medicine data and technology platform (Dataome), that has previously been validated in clinical decision support and in the elucidation of novel molecular mechanisms associated with drug efficacy and safety (Armaiz-Pena et al., 2013; Pradeep et al., 2015; Schell et al., 2016; Bohnert et al., 2017; Soldatos et al., 2018; Soldatos and Jackson, 2019; Schotland et al., 2021), to deliver a digital whole patient COVID-19 symptomatology model to the community ("*COVID-19 Explorer*"; at https://covid19.molecularhealth.com) (Brock et al., 2022). Here, we demonstrated that it is possible to structure and logically connect diverse clinical and molecular features of COVID-19 pathobiology by using digital health platforms like the Dataome technology platform. Dataome provided a knowledge and data-mining infrastructure that allowed us to rapidly initiate an iterative computer-augmented modeling approach, guided by disease modeling experts that allowed us to build a comprehensive digital model during the early phases of the pandemic (Brock et al., 2022).

In this work, we explore the utility and importance of the holistic patient-level knowledge model within the *COVID-19 Explorer* resource and examine its key features, particularly surrounding the molecular underpinnings of COVID-19 symptomatology. Characteristically, the model suggests that the multitude and complexity of observed, and seemingly disparate COVID-19 clinical phenotypes may be linked to the pleiotropic activity of eight core molecular mechanisms involved in the host response. Moreover, the model revealed functionally connected mechanisms across multiple organ systems and identified novel hypotheses for both viral dependent and independent disease mechanisms. Here, we discuss these molecular perspectives in detail, including the content of the causative pathogenic mechanisms underlying COVID-19 phenotypes, real world confirmatory observations, and examples that demonstrate some of the key analytical utilities (e.g., risk factors). Finally, we focus on the detection of potentially new (or unobvious) clinico-molecular insights across organ-systems and on the identification of potential pharmacologic targets against COVID-19 (Brock et al., 2022).

Our results show that structuring emergent molecular knowledge *via* a pan-symptomatic disease format, helped develop a comprehensive whole patient COVID-19 knowledge model that could expedite evidence-driven hypothesis generation and the discovery of novel clinico-molecular insights. Inspired by the effectiveness of this strategy, we propose that whole-patient knowledge modeling of systemic symptomatology for any disease, as opposed to traditional pathway-specific modeling, may provide important advantages in tackling current unmet medical needs and future health emergencies. Importantly, the whole-patient knowledge model is provided to the community in the form of both an open-source web-sever (found at https://covid19.molecularhealth.com), and a tabular COVID-19 "Cockpit" (Brock et al., 2022). Together they are aimed at enabling a variety of use-case scenarios useful to support translational and clinical researchers in hypothesis generation and the development of new diagnostic and therapeutic strategies.

2 Materials and methods

2.1 Knowledge capture, integration and modelling using the Dataome technology

Our studies were led by a small team of disease modelling experts working with the Dataome Technology Platform to capture, structure and logically connect diverse clinical and molecular features of COVID-19 pathobiology. Dataome consists of three primary modules that enable the constant 1) capture and curation 2) integration and connection and 3) analysis of, globally available data sources of clinical and molecular knowledge (Supplementary Material S1, Figure 1):

The *Dataome Capture* technology uses an ensemble of public/proprietary algorithms and resources to empower the global harvesting, quality assurance and integration of emergent clinical and molecular data. Curated and quality-controlled data is then integrated into the *Data Nucleus*.

The *Dataome Nucleus* encompasses data from >100 public, commercial and proprietary developed resources. Datasets span a broad range of content, size and formats—from more general, such as literature, biomedical ontologies or information on proteins and genes, drugs and their targets, interactors, bio-molecular pathways, and interactions. Other resources include data on millions of patients from diverse real-world data (RWD) databases. Moreover, Dataome contains structured information regarding therapeutic guidelines and variant classification, as well as curated datasets pertaining to clinical biomarker interpretation, pathway/interaction relationships, drug and clinical trial information.

FIGURE 1

COVID-19 disease model—from symptomatology to molecular mechanisms. **(A)** The many faces of COVID-19—a systemic disease: COVID-19 comes with diverse clinical manifestations affecting a broad range of organ systems. Originally identified as respiratory disease (Huang et al., 2020) (pneumonia, acute respiratory distress syndrome (ARDS), Acute Lung Injury (ALI), shortness of breath, cough) with highly viscous mucus (Wu et al., 2020; Zhu et al., 2020), many more manifestations and affected organs were reported over time. These include neurologic manifestations (Mao R. et al., 2020; Ellul et al., 2020)and silent hypoxemia (Couzin-Frankel, 2020), musculoskeletal manifestations (Cipollaro et al., 2020), gastroenterological symptoms (Barrera et al., 2020; Mao Y. et al., 2020), cardiovascular manifestations (Kang et al., 2020; Zheng et al., 2020), endotheliitis (Varga et al., 2020), coagulopathy (Levi et al., 2020), new onset diabetes (Rubino et al., 2020), kidney injury (Selby et al., 2020), leukopenia (Ruan et al., 2020). Visualization was realized using the Biodigital 3D anatomy model (www.biodigital.com). **(B)** Synopsis of the basic COVID-19 Disease Model. **(B)** (sub-panel A) The virus targets host cells expressing ACE2 and TMPRSS2, active components of the KKS. **(B)** (sub-panel B) Within the cell the virus induces downregulation of SERPING1 and ACE2. SERPING1 downregulation induces activation of the CAS and KKS. ACE2 downregulation results in accumulation of DAKs. Excess DAKs activate B1R, triggering a constitutive activation and auto-amplification. Under normal conditions B1R activation triggers a fast, transient inflammatory emergency reaction inducing neutrophil and leukocyte recruitment and infiltration, opening of the epithelial/endothelial barriers for their transmigration, coagulation for local walling-off, fibrogenesis for wound healing, senescence as host defense mechanism and transient nociception. Constitutive activation of this system leads to excess inflammatory signaling, epithelial/endothelial barrier breakdown, induction of thrombosis, fibrosis, pain and other effects. **(B)** (sub-panel C) B1R signaling induces the formation of MVs bearing B1R and miR200c. These are transferred to target cells, inducing further expression of miR200c, which leads to ACE2 downregulation and formation of excess DAKs which signal *via* B1R. Thus, MV transfer *via* auto-activation and amplification of the B1R system may trigger the virus independent propagation of an inflammatory phenotype. **(C)** COVID-19 clinical phenotypes associated to eight key pathogenic mechanisms. Key mechanisms (green) and associated signs, symptoms and syndromes of COVID-19 (blue) induced by over-activation of the ACE2/DAK/B1R signaling axis. Members of the axis, as well as regulatory elements (e.g., miR200c, SIRT1, EZH2, etc.) converge at these mechanisms and complement each other to various degrees in inducing the corresponding phenotypes. **(D)** The key players of the seed model converge at the CAS, Contact Activation System. CAS constitutes a group of plasma proteins that are activated by FXII. FXII activates proinflammatory, procoagulant and host defense pathways. The initial step involves activation of PK, prekallikrein. PK reciprocally activates FXII. Inflammation results from the activation of the KKS, Kallikrein Kinin system resulting in the release of kinins, BK, bradykinin and KD, Kallidin, and the DAK, des-Arg-Kinins DABK and DAKD from high molecular weight kininogen (HK). Coagulation is driven by FXIIa-mediated cleavage of FXI, which eventually leads to thrombin activation and clotting. FXII activation also leads to increased plasmin activation, which breaks down the fibrin clot. Finally, FXII can activate C1r and C1s of the complement pathway. SERPING1 is the key regulator of CAS activation. It is the main inhibitor of the FXIIa, the C1 complex and the conversion of HK to kinins. TMPRSS2 activates tissue kallikrein KLK2, PLAT and fibrinogenic PAR2, which can also be activated by KK. ACE2 metabolizes DAK, thus inhibiting DAK signaling. DAK activates and induces expression of B1R. gC1qR, a pathogen receptor and key activator of the CAS, induces the expression of B1R. B1R activation in turn induces miR200c, which directly targets and suppresses ACE2, constituting a feed-forward loop leading to the generation of excess DAK. miR200c induces Mortalin an inhibitor of the cytotoxic component of the complement cascade. Complement activator MASP2 is a target of the virus N protein. SERPING1 inhibits MASP2 proteolytic activity.

FIGURE 2
Reciprocal regulation of ACE2 and TMPRSS2. ACE2 and TMPRSS2 are reciprocally regulated by miR200c, SIRT1 and EZH2. SIRT1 induces ACE2 expression (Clarke et al., 2014; Yacoub et al., 2014; Moran et al., 2017; Shao et al., 2019) and indirectly represses TMPRSS2 by acting as co-repressor of the AR, androgen receptor (Dai et al., 2007) and by repressing EZH2 (Lu et al., 2011). EZH2 represses ACE2 (Li Y. et al., 2020) and, in an androgen receptor dependent manner, derepresses TMPRSS2 as well as its substrate KLK2 (Kim et al., 2018). miR200c directly represses ACE2 (Liu et al., 2017) and SIRT1 (Carlomosti et al., 2017), through which it may indirectly induce TMPRSS2. AMPK activation increases phosphorylation and stability of ACE2 (Zhang et al., 2018) and via SIRT1 induces expression of ACE2 (Clarke et al., 2014). In turn ACE2 activates AMPK (Murça et al., 2012).

The extensive data input contained within the *Nucleus* provides the evidence-base, required by the analytical technologies and AI-based tools contained within the *Dataome Analytics* part of the technology. These software tools also exist in specialized analytical pipelines that integrate bioinformatics, chemo-informatics, systems biology, clinical data science, and AI/machine learning (integrated analytics, feature engineering and powerful pre-trained models) based methodologies.

2.2 Knowledge modeling workflow

Using these components, the *Dataome Nucleus* was queried with 332 high confidence human protein interactors with the SARS-CoV-2 proteome[1], to provide systematic expansion inclusive of associated pathways, interactors and regulatory elements yielding a massive interactome network mapping the majority of the entire human proteome. To optimize the specificity of analysis, we focused initial modeling on the molecular determinants of host cells that define them as viral

targets and the immediate impact of viral infection on the host cell response. Next, we linked key molecular protagonists to COVID-19 symptomatology, outcome and severity associated risk-factors. These factors were mapped onto pathways, interactors and/or substrates and regulatory networks exploiting interaction databases and data from literature within the *Dataome Nucleus*. Pathways were defined through entities as canonical elements (ACE2, SERPING1) or *via* their substrates (TMPRSS2) and converged into a common pleiotropic signaling cascade, the contact activation system. The *Dataome Nucleus* was also screened for all factors and associated phenotypes linked to naturally occurring variants or functional studies involving pharmacological perturbations or knock-out models, with an initial focus on COVID-19 pulmonary phenotypes.

Readouts from these *in silico* analyses were organized by experts into molecular models containing salient information for each phenotype. The resulting "base model" was then matched to molecular models of diseases that are symptomatically related to specific manifestations of COVID-19 (see Supplementary Material S1, Figure 2). At each stage of the disease/symptom modelling process, the extracted "premodels" were inspected and remodeled in PathVisio (Version: PathVisio 3.2.2). During the remodeling, relevant "premodel" references were attached to the respective objects in PathVisio and manually complemented as necessary. Entities and relationships associated with each mechanism/symptom were presented by a JavaScript-animated SVG image. Graphical renditions were produced *via* the PathVisio program, with data outputs in GPML and SVG format. The web application displaying these "submodels" is served by a Flask micro web framework. The application presents the SVG for a sub-model, animated by the d3.js JavaScript library. To project a most actual and comprehensive information, the collection of citations supporting the relations of each "submodel," range from peer-reviewed articles, very recent conference content (e.g., abstracts) to *ad hoc* communications and is provided as auxiliary information *via* respective animation components.

COVID-19 explorer and data availability: The final comprehensive model, linking key molecular mechanisms to COVID-19 symptomatology and the related source data, is made publicly available *via* a web-based COVID-19 Explorer (http://covid19.molecularhealth.com). For detailed information on data availability please see Supplementary Material.

3 Results

3.1 Insights and hypotheses from the COVID-19 knowledge model

The comprehensive and fully interactive COVID-19 disease model (aka the COVID-19 Explorer, accessible at: https://

covid19.molecularhealth.com), provides a link between possible molecular disease mechanisms [aberrant contact activation system (CAS) and ACE2/DAK/B1R signaling] and eight core pathogenic processes: inflammatory signaling, coagulation, barrier permeability, senescence, autoimmunity, fibrogenic signaling, nociception and exocytosis. These mechanisms are in addition linked with respective symptoms, associated pathogenic pathways and affected organ-systems.

3.2 Host factors mediating SARS-CoV-2 infection

Analysis of the COVID-19 knowledge model revealed a converging molecular network, delineating host-factor responses to SARS-CoV-2 *via* the host proteins responsible for virus entry, together with downregulated components of Interferon Stimulated Genes (ISG's) induced by virus infection. SARS-CoV-2 cell-entry depends on binding of the viral spike (S) protein to the cellular receptor ACE2 and S protein priming by the host cell serine protease TMPRSS2 (Hoffmann et al., 2020) (Figure1B). SARS-CoV-2 targets diverse cell types within the lung, all of which express ACE2 (Hamming et al., 2004; Jia et al., 2005; Xu et al., 2020). TMPRSS2 is highly expressed with broader tissue distribution, suggesting that ACE2 expression may be limiting in cellular susceptibility to infection (Sungnak et al., 2020). The induction of ISG expression is part of the early response to viral infections. Analysis of ISG expression in human airway epithelial cells reveals the ability of SARS-CoV to interfere with this process and avoid anti-viral host-response (Menachery et al., 2014). Induction of ISGs is delayed and occurs after peak titers, a behavior also confirmed for SARS-CoV-2 (Nienhold et al., 2020). While ISG expression is universally increased, ACE2 and SERPING1 are significantly downregulated (Menachery et al., 2014) (Figure 1B). As SERPING1 is one of the proteins with the highest connectivity in the SARS-CoV-1 and SARS-CoV-2 interactomes, it was proposed that SARS-CoV-2 infection directly causes deficiency in C1 esterase inhibitor (Thomson et al., 2020). Recently, clinical samples of COVID-19 patients revealed C1-Inhibitor as one of the most prominently downregulated genes with 80-fold decreased expression (Mast et al., 2021).

The model indicates that the viral cell-entry mechanism and disease-specific ISG signature provides three key active components of SARS-CoV infected cells (TMPRSS2, ACE2 and SERPING1), which may functionally converge in the same pleiotropic signaling systems, namely the Contact Activation System (CAS) and Kallikrein Kinin System (KKS) pleiotropic signaling (Figure 1D):

- TMPRSS2 and ACE2 through common elements of their regulatory network (miR200c, SIRT1, EZH2 and AMPK) are regulated in a reciprocal manner (Figure 2).
- ACE2 is co-expressed with TMPRSS2 in respiratory tract cells and oral mucosa (Hamming et al., 2004; Jia et al.,

2005; Xu et al., 2020; Ziegler et al., 2020) and present in endothelial cells and in the arterial smooth muscle cells of many organs (Hamming et al., 2004). As part of ISG-response, ACE2 plays a tissue-protective role in innate immunity (Menachery et al., 2014; Ziegler et al., 2020). SARS-CoV infection leads to robust down-regulation of ACE2 RNA and protein expression (Kuba et al., 2005; Menachery et al., 2014). Loss of pulmonary ACE2 is a key event in the molecular pathogenesis of acute lung injury (ALI) (Imai et al., 2005; Kuba et al., 2005; Kuba et al., 2006; Imai et al., 2008). The detrimental effect of ACE2 downregulation in SARS-CoV infection is attributed to its role in the Renin-Angiotensin System (RAS) (Tseng et al., 2020). ACE2 converts AngII to Ang (1–7). Its downregulation leads to over-activation of the AngII receptor AT1R (Jia et al., 2005; Banu et al., 2020) (Figures 3, 4). However, on a systemic level, this conversion is ACE2-independent (Serfozo et al., 2020). ACE2 also participates in the KKS. It efficiently inactivates des-Arg-kinins (DAK) (Vickers et al., 2002). Moreover, pharmacological activation and inhibition of ACE2 emphasise it is critical role in inflammation (Figure 5). Thus, our model implicates ACE2 in the context of the KKS, but also in the context of RAS.
- While TMPRSS2 plays a key role in prostate cancer (Tomlins et al., 2005; Kron et al., 2017), it also activates pro-kallikrein-2 (KLK2) (Lucas et al., 2014), resulting in selective cleavage of kininogen to release kallidin (KD), a precursor of the ACE2 substrate des-Arg (Hoffmann et al., 2020)-KD (DAKD). Thus, TMPRSS2 is an activator of the KKS.

SERPING1 is the major regulator of the CAS (Figure 1D). Downregulation of SERPING1 leads to activation of the CAS and excess Bradykinin (Cugno et al., 2009), a precursor of the ACE2 substrate DABK. Aberrant regulation of CAS leads to inflammation and autoimmunity and is involved in diseases like ischemia/reperfusion-syndrome, sepsis, atherosclerosis and diabetes (Davis et al., 2010; Schoenfeld et al., 2016).

3.3 Host-response driven disease mechanisms

Global level analysis of the model reveals that the SARS-CoV-2 virus targets cells expressing constituents of a highly inducible inflammatory signaling system leading to its excess activation. The model further suggests that it is the pleiotropic nature of this system that appears to be responsible for the diverse clinical manifestations of COVID-19. Importantly, it also provides a possible mechanism through which a disease phenotype may be propagated, even in the absence of the
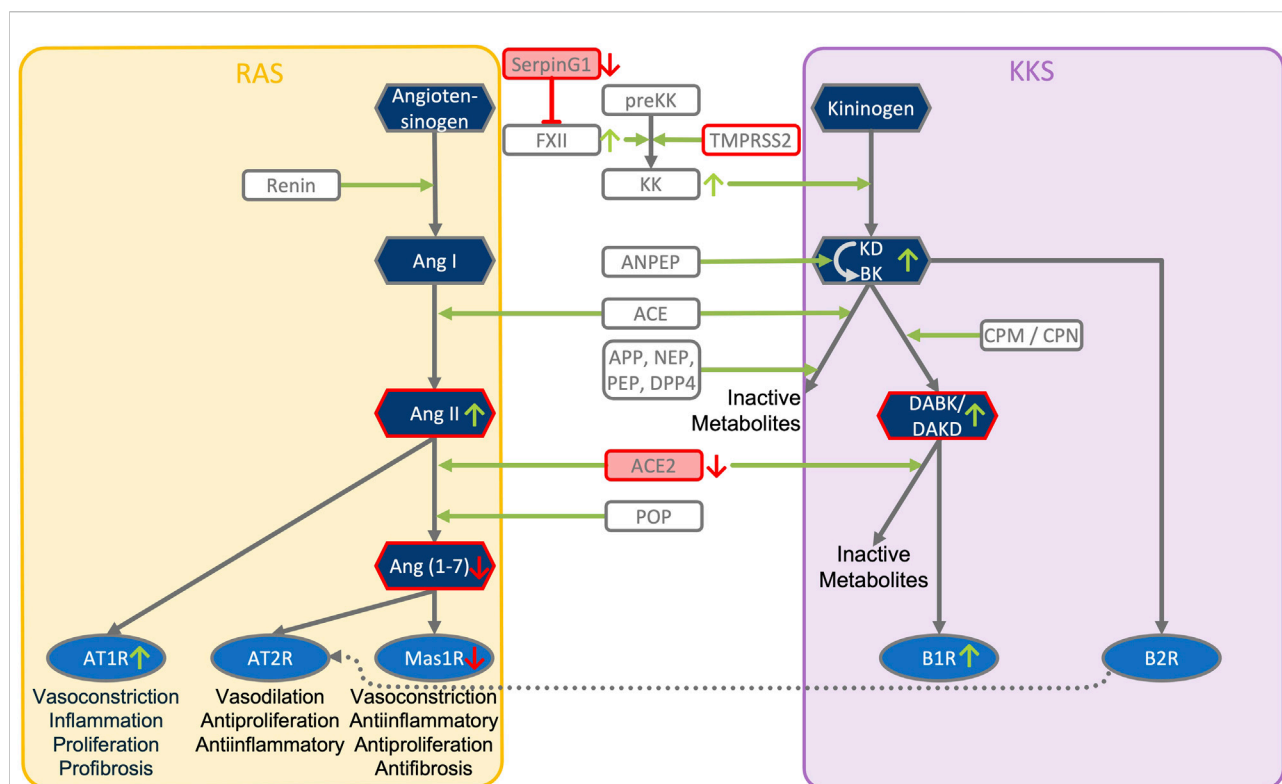
FIGURE 3
ACE2, SERPING1 and ACE2 converge at the crosstalk between KKS and RAS pathways. The KKS comprises the kininogens, the kallikreins, the kinins and the kinin receptors. KK, Kallikrein is generated from preKK, prekallikrein by factor XIIa during activation of CAS. FXIIa can be inhibited by SERPING1 of the complement system. Kinins are generated by cleavage of kininogens by kallikreins. Kinins are BK, bradykinin and Lys-BK (Kallidin, KD) which both activate B2R, and their active des-Arg metabolites, generated by CPM/N, des-Arg (Brock et al., 2022)-BK (DABK) and des-Arg (Hoffmann et al., 2020)-KD (DAKD), which activate B1R. In this figure, the RAS, renin-angiotensin system has been reduced to key components comprising the angiotensins (AngI, II, 1−7), the angiotensin receptors (AT1R, AT2R, Mas receptor) and the ACEs, angiotensin converting enzymes. KKS and RAS are mainly connected by ACE and ACE2. ACE (kininase II) inactivates the kinins and generates angiotensin II (AngII). ACE2 inactivates the B1R agonist DAK, des-Arg kinins and also metabolizes AngII into the Ang-one to seven agonist of AT2R and Mas oncogene. Systemic AngII conversion to Ang (1−7) depends on POP, Prolyloligopeptidase and is ACE2 independent. APP, Aminopeptidase P; NEP, Neprilysin; PEP, Prolylendopeptidase inactivate kinins. AT2R directly interacts with B2R through heterodimerization and AT2R overexpression increases KK activity. Simultaneous downregulation of SERPING1 and ACE2 leads to accumulation of excess DAK (DABK and DAKD).

original viral trigger (Figure1B). The following sections analyze these different perspectives in detail.

3.3.1 Convergence within the pleiotropic KKS dysregulates ACE2-DAK-B1R axis: A "perfect storm," triggering systemic disease

The functions of ACE2, TMPRSS2 and SERPING1 converge within the KKS. Key products of KKS, the kinins bradykinin (BK), Kalladin (KD), and their des-Arg metabolites, des-Arg (Brock et al., 2022)-BK (DABK) and DAKD are major inflammatory mediators. By inactivating DABK and DAKD, ACE2 is a negative regulator of KKS signaling (Jia, 2016). Under normal conditions, DABK and DAKD are readily metabolized (Vickers et al., 2002). Under inflammatory conditions, DABK can accumulate (McLean et al., 2000), with downregulated ACE2 leading to further accumulation of des-Arg-Kinins (DAK) (Sodhi et al., 2018).

Thus, induction of CAS along with downregulation of KKS inhibitor SERPING1 and des-Arg-Kinin inactivator ACE2 may lead to local excess of DAK, resulting in constitutive activation and induction of its receptor B1R (Figure 1B). Indeed, clinical samples from COVID-19 patients revealed that B1R was one of the most prominently upregulated genes with a 260-fold expression increase (Mast et al., 2021).

B1R is part of a rapid response system expressed in inflammation associated cytotypes (Walker et al., 1995; Böckmann and Paegelow, 2000; Wu et al., 2002). It is controlled by inflammatory stimuli, such as IL1β or availability of DABK or DAKD, and rapidly switched-off by internalization in the absence of ligand (Marceau et al., 1983; Dray and Perkins, 1993; Couture et al., 2001; Prado et al., 2002; Calixto et al., 2004; Leeb-Lundberg et al., 2005). Spatial proximity of B1R to enzymes controlling the half-life of its ligands
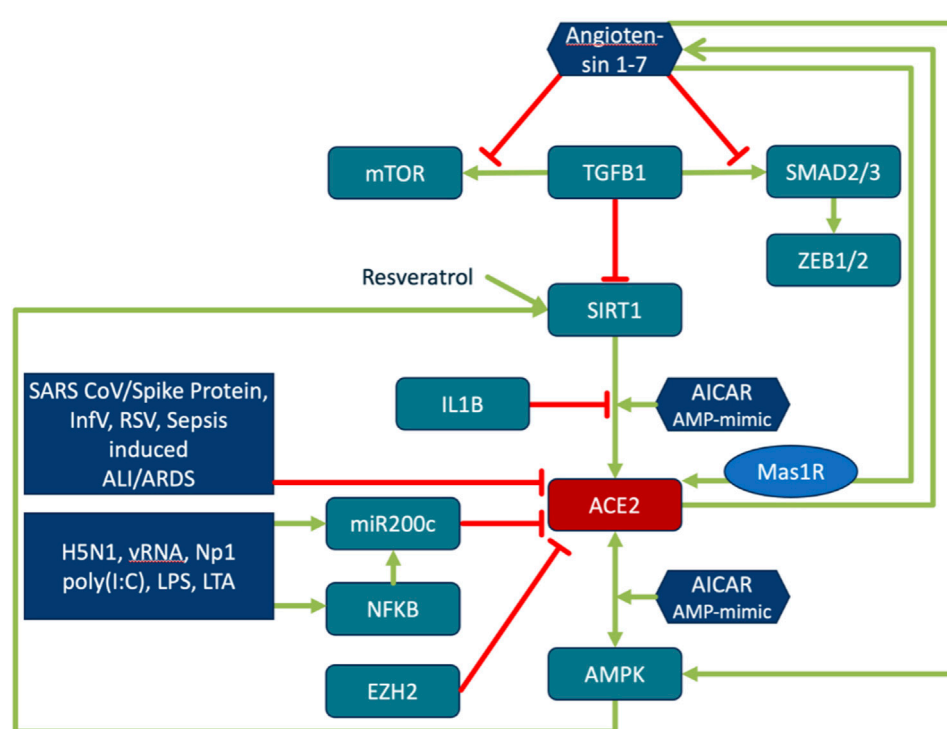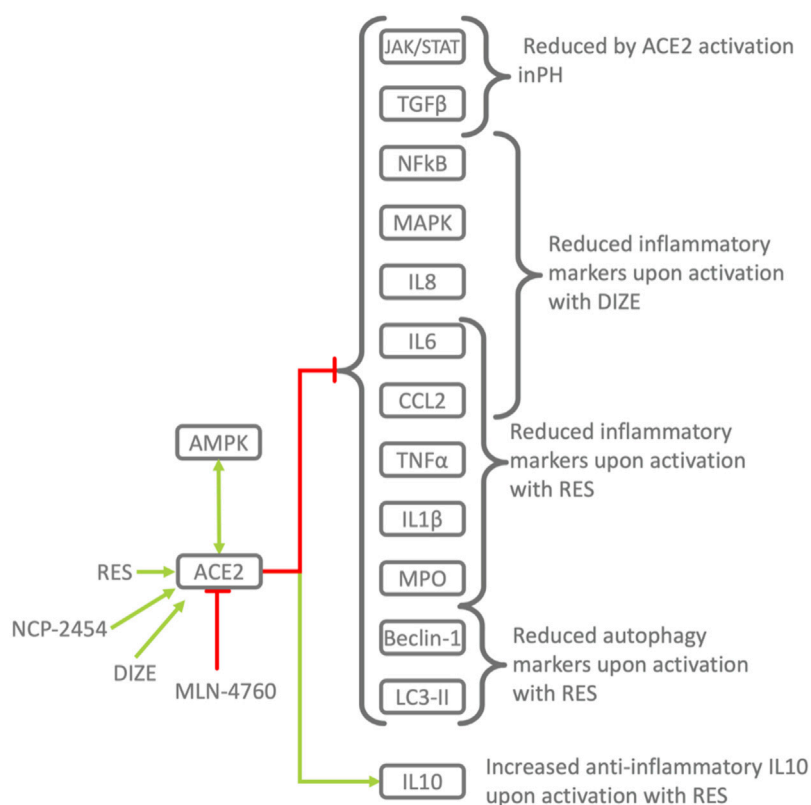
FIGURE 4
Regulation of ACE2 expression and its activity. Regulation of ACE2 is closely interwoven with energy sensing, inflammatory processes and ageing. TGFβ downregulates ACE2 expression in a SIRT1-dependent manner. SIRT1 activator, SRT1720, restores the ACE2 expression. ACE2/Ang (1−7)/MasR are decreased in pulmonary fibrosis and negatively correlate with TGFβ expression. Ang (1−7) directly inhibits TGFβ-induced phosphorylation of SMAD2 and SMAD3 and mTOR, and suppresses the expression of the downstream target genes of TGFβ/SMAD signaling (ZEB1, ZEB2, TWIST, and SNAIL1) (Shao et al., 2019). Activation of the energy status sensing AMPK by AMP mimic AICAR increases phosphorylation of ACE2 and SIRT1 dependent expression of ACE2. Phosphorylation of ACE2 enhances its stability and activity (Zhang et al., 2018). ACE2 in turn activates AMPK. Phosphorylation of AMPK is reduced in ACE2 knock-out, while Ang (1−7) activates AMPK. Resveratrol and Ang (1−7) increase ACE2 expression. SIRT1 antagonist and Mas antagonist block this effect (Liu Y. et al., 2019). SIRT1 binds to the ACE2 promoter. Binding is increased by AICAR treatment and decreased by IL1β (Clarke et al., 2014). ACE2 activation by SIRT1 leads to increased levels of Ang (1−7) (Yacoub et al., 2014). The compound Resveratrol induces SIRT1-dependent upregulation of ACE2 (Moran et al., 2017). The role of miR200c in this context is explained in the sections on miR200c function and the reciprocal regulation of ACE2 and TMPRSS2.

contributes to its regulatory role (Lamb et al., 2002; Lu et al., 2008; Qadri and BaderKinin, 2018). B1R activation promotes cytosolic influx of extracellular Ca^{2+}. Sustained activation increases receptor expression and induces resistance to desensitization and internalization, triggering a positive feed-forward-loop and excess downstream signaling (Faussner et al., 1999; Marceau et al., 2002; Calixto et al., 2004). B1R is involved in signaling cascades controlling numerous mechanisms that when over-activated are directly related to the molecular pathologies underlying COVID-19. Characteristically, it induces 1) an inflammatory response at the site of infection and recruitment of neutrophiles and leukocytes, 2) coagulation to prevent infection, 3) nociception, 4) increased endothelial/epithelial barrier permeability allowing fluid and proteins to move into the interstitium and immune cells to trans-migrate, 5) reduced exocytosis, 6) senescence as mechanisms to fight pathogens and 7) fibrogenesis for wound healing (Figure 1C).

3.3.2 A feed-forward loop decoupling molecular pathogenesis from virus load: Auto-induction of B1R bearing MVs and interplay with the regulatory miR200c

Our patient-level model highlights that in B1R-mediated pathologies, intercellular information exchange *via* MVs may contribute to the dissemination of disease phenotypes (Figure 6). In acute vasculitis, B1R bearing MVs play a role in the bi-directional communication between leukocytes and endothelium. Here, MVs transfer functional receptors to promote kinin-associated inflammation (Kahn et al., 2017; Mossberg et al., 2017). Patients with acute vasculitis show high-levels of circulating B1R-positive endothelial MVs, which induce a neutrophil chemotactic effect. Patient plasma induces release of more B1R-positive MVs from endothelial cells. SERPING1 depleted plasma promotes excessive release of B1R-positive endothelial MVs, while addition of

FIGURE 5
Pharmacological activation/inhibition of ACE2 demonstrates its protective role in inflammation. RES, Resorcinolnaphthalein activates ACE2 in a dose-dependent manner (Hernández Prada et al., 2008) and leads to decrease in pro-inflammatory TNFα, CCL2, IL6 and increase in anti-inflammatory IL10 improving endothelial dysfunction in PH (Li et al., 2012). Activation of ACE2 by RES alleviates the severity of ALI the lung, while ACE2 downregulation and pharmacological inhibition leads to the opposite effect. Similar mechanisms were observed in other diseases. Phosphorylation of ACE2 by AMPK in the endothelium leads to its activation and mitigates PH, pulmonary hypertension (Zhang et al., 2018). In animal models of PH, activation of ACE2 by the RAS, renin-angiotensin system improves endothelia-dependent vasorelaxation, decreases proinflammatory TNFα, CCL2, IL6 and increases anti-inflammatory IL10 (Li et al., 2012) and ACE2 activator reduces monocrotaline-induced PH by suppressing the JAK/STAT and TGFβ (Haga et al., 2015).

SERPING1 or B1R-antagonist abolishes this effect (Mossberg et al., 2017). Thus, B1R signaling induces excretion of B1R bearing MVs in an auto-amplifying manner, while downregulation of SERPING1 via induction of KKS increases ligand supply. Such MV-based intercellular communication also occurs between different airway cells (Gupta et al., 2019). Interestingly, airway epithelial cell secretions are enriched with components of the CAS and KKS, containing key regulators of these systems, indicating that under normal conditions this mechanism constitutes a well-balanced system. This aligns with the observation that pulmonary epithelial cells express active surface peptidases to degrade kinins rapidly, keeping the system balanced and protecting the lung (Ghebrehiwet et al., 2019).

The secretome and exosomes of airway epithelial cells also contain miRNAs. miR200c, a negative regulator of ACE2 expression (Figure 6), is significantly enriched (Gupta et al., 2019) and induced by oxidative stress (Wu et al., 2017) and highly elevated in pneumonia and chronic obstructive pulmonary disease patients (Liu et al., 2017) and positively correlated with disease severity in interstitial lung disease (Jiang et al., 2017). In SARS-CoV, ALI or Acute Respiratory Distress Syndrome (ARDS) is associated with downregulation of ACE2, mediated by increased miR200c. Inhibition of miR200c ameliorates ALI (Liu et al., 2017). By targeting antioxidant proteins, miR200c leads to an increase in reactive oxygen species (ROS) formation (Carlomosti et al., 2017; Wu et al., 2017), contributing to progression of the inflammatory process and endothelial/epithelial barrier dysfunction (Kellner et al., 2017). miR200c has also been shown to directly target FOXO (Carlomosti et al., 2017). FOXO is activated in patients with pulmonary diseases and its deficiency results in suppression of TLR3-dependent epithelial innate immune function and increased pathogen uptake (Totura et al., 2015). In epithelial cells,
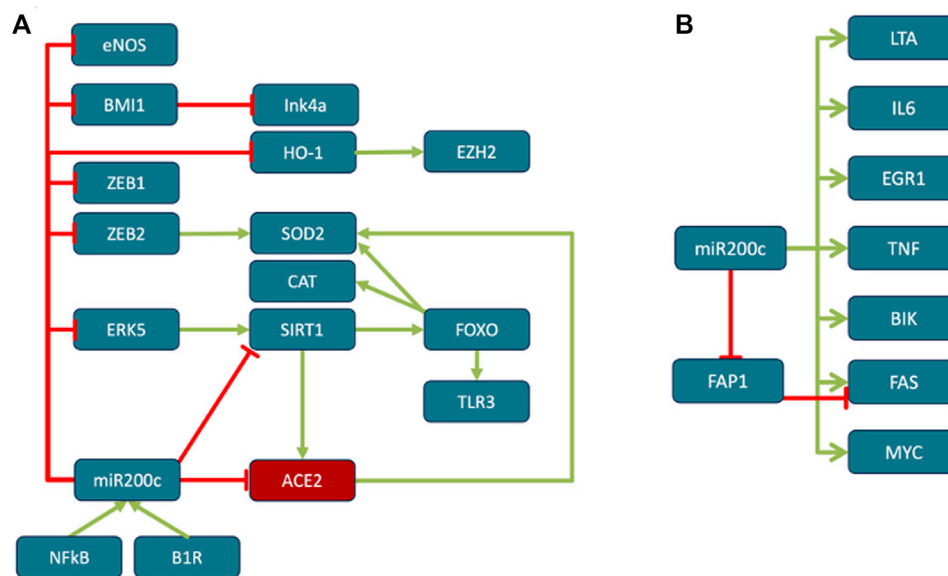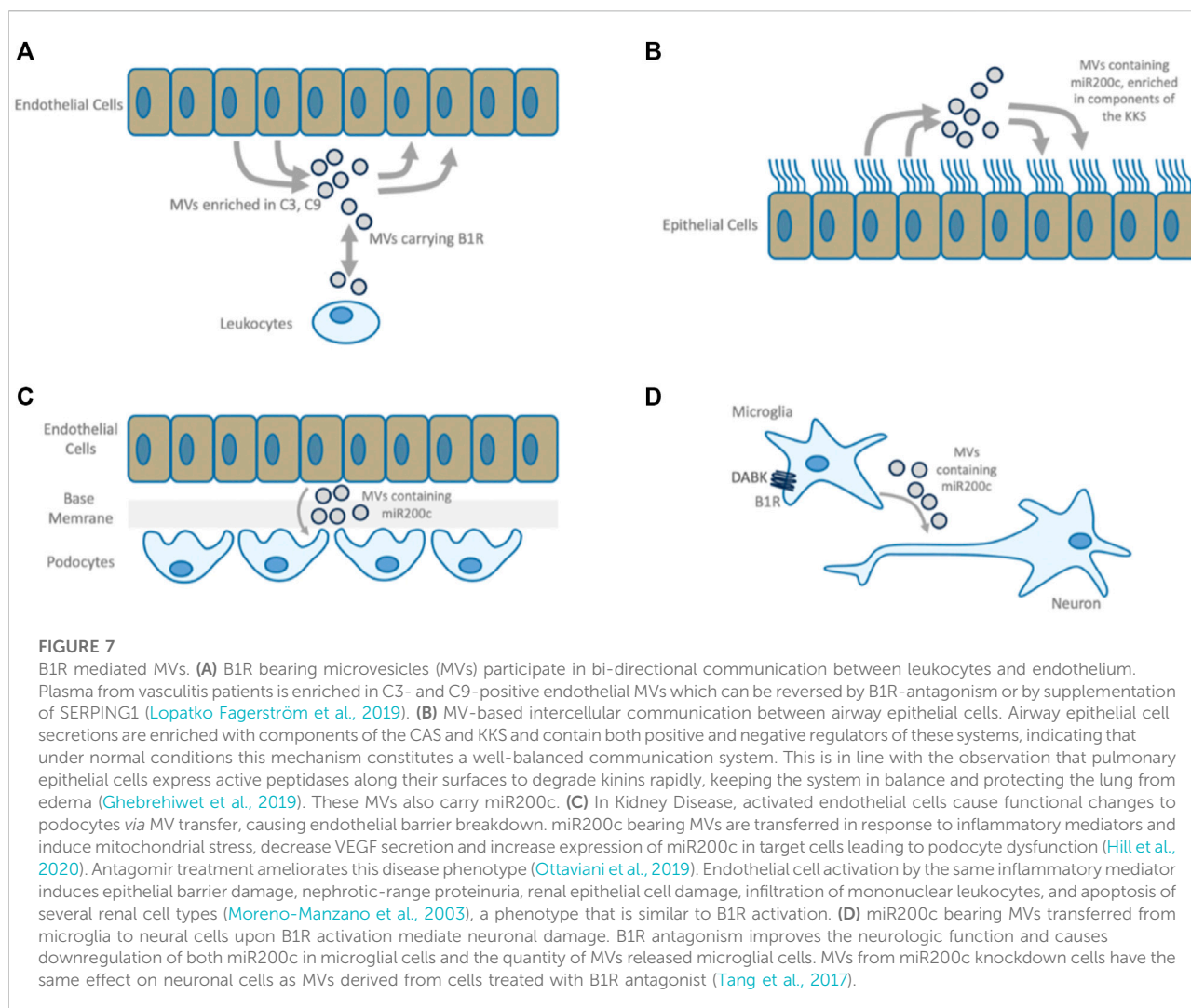
FIGURE 6

Role of miR200c. **(A)** miR200c down-regulates ACE2, reactive oxygen species (ROS) scavengers, eNOS, SIRT1 and derepresses senescence marker p16INK4a. miR200c levels elevated in lung diseases such as fibrosis, chronic obstructive pulmonary disease (COPD) and pneumonia (Cao et al., 2014; Jiang et al., 2017; Liu et al., 2017). B1R activation induces miR200c (Tang et al., 2017), which in turn represses ACE2 in acute respiratory distress syndrome (ARDS) (Liu et al., 2017). miR200c directly represses SIRT1 (Carlomosti et al., 2017), eNOS and indirectly FOXO1 (Carlomosti et al., 2017), which in turn leads to downregulation of ROS scavengers Catalase and SOD2. In addition, oxidative stress induced miR200c downregulates ROS scavenger heme oxidase HO1 (Wu et al., 2017). ROS scavengers alleviate ARDS and septic shock (Janssen and Nozik-Grayck, 2017), while when down-regulated, increased ROS leads to ARDS progression and endothelial/epithelial barrier dysfunction (Kellner et al., 2017; Chen et al., 2018). In addition, HO1 stimulates the expression of EZH2 (He et al., 2019) (see reciprocal regulation of ACE2 and TMPRSS2). Downregulation of SOD2 and SIRT1 by miR200c is mediated through ZEB2 and ERK5 (Wu et al., 2017). FOXO is activated in patients with respiratory tract diseases. TLR3-mediated innate immune responses of bronchial epithelial cells depend on FOXO and its deficiency results in suppression of epithelial innate immune function and increased of pathogen uptake (Totura et al., 2015). miR200c also targets BMI-1 (Cao et al., 2011), absence of which severely affects lymphopoiesis (van der Lugt et al., 1994) (see senescence). **(B)** The effects of miR200c overexpression in epithelial cells. Overexpression induces LTA (19,5x), IL6 (17,5x), EGR1 (8,4x), TNF (7,2x), BIK (4,8x), FAS (4,2x), MYC (2, 8) (Tryndyak et al., 2010). Lymphotoxin (LTA)-signaling participates in airways remodeling during inflammation (Koroleva et al., 2018). Notably, SARS-CoV triggers EGR1 dependent activation of TGFß inducing profibrotic responses (Li et al., 2016), while inhibition of EGR1 ameliorates pulmonary fibrosis (Bhattacharyya et al., 2013). BIK may contribute to lung destruction in COPD (Joyce-Brady and Tuder, 2011) and MYC is a key regulator in sepsis-induced ARDS (Zhang et al., 2019). FAS induces apoptosis. In addition to upregulating FAS, mir200c directly targeting FAS inhibitor FAP1 (Schickel et al., 2010). HCV induced miR200c down modulates FAP-1 resulting in significant increases in expression of collagen and fibroblast growth factor (Ramachandran et al., 2013).

miR200c overexpression strongly induces pro-inflammatory IL6 (Tryndyak et al., 2010). miR200c and B1R carrying MVs play a role in the molecular pathology of many diseases, including kidney disease, vasculitis and Kawasaki Disease (KWD) (Zhang R. et al., 2017) (Figure 7). Next, MVs secreted by stressed cardiomyocytes are highly enriched in miR200c. MV-mediated intercellular communication between microglial and neural cells in cerebral injury (Tang et al., 2017) suggests that B1R signaling is involved in the generation of miR200c carrying MVs. Here, miR200c bearing MVs are transferred from microglia to neural cells mediating neuronal damage. In this context, miR200c selectively represses Syntaxin1, an important functional protein for neurotransmission causing loss of function of neural cells (Davis et al., 1998; Watanabe et al., 2013). Interestingly, these pathological processes can be ameliorated by B1R antagonism or SERPING1 supplementation mimicking the phenotype of Kininogen knockouts.

In summary, the model shows that B1R and miR200c bearing MVs may play an important role in molecular pathologies associated with epithelial and endothelial barrier damage, fibrosis, cardiac and neuronal damage: MVs containing miR200c are involved in communication between airway cells, B1R activation leads to self-propagating dissemination of B1R bearing MVs and can induce the formation of miR200c positive MVs. In targeted cells, B1R is auto-induced and miR200c can suppress ACE2, which in turn leads to accumulation of B1R ligands DAK. This constitutes a mechanism through which a disease phenotype may be propagated even in the absence of the original viral trigger. Indeed, analysis of post-mortem COVID-19 lung suggests two distinct stages of disease-progression. Early disease has high viral-load and high expression of cytokines and ISGs and sparse immune infiltrates, while in late disease, low viral loads, low local expression of cytokines and ISGs, and strong infiltration of macrophages and lymphocytes prevail. Patients

FIGURE 7

B1R mediated MVs. **(A)** B1R bearing microvesicles (MVs) participate in bi-directional communication between leukocytes and endothelium. Plasma from vasculitis patients is enriched in C3- and C9-positive endothelial MVs which can be reversed by B1R-antagonism or by supplementation of SERPING1 (Lopatko Fagerström et al., 2019). **(B)** MV-based intercellular communication between airway epithelial cells. Airway epithelial cell secretions are enriched with components of the CAS and KKS and contain both positive and negative regulators of these systems, indicating that under normal conditions this mechanism constitutes a well-balanced communication system. This is in line with the observation that pulmonary epithelial cells express active peptidases along their surfaces to degrade kinins rapidly, keeping the system in balance and protecting the lung from edema (Ghebrehiwet et al., 2019). These MVs also carry miR200c. **(C)** In Kidney Disease, activated endothelial cells cause functional changes to podocytes *via* MV transfer, causing endothelial barrier breakdown. miR200c bearing MVs are transferred in response to inflammatory mediators and induce mitochondrial stress, decrease VEGF secretion and increase expression of miR200c in target cells leading to podocyte dysfunction (Hill et al., 2020). Antagomir treatment ameliorates this disease phenotype (Ottaviani et al., 2019). Endothelial cell activation by the same inflammatory mediator induces epithelial barrier damage, nephrotic-range proteinuria, renal epithelial cell damage, infiltration of mononuclear leukocytes, and apoptosis of several renal cell types (Moreno-Manzano et al., 2003), a phenotype that is similar to B1R activation. **(D)** miR200c bearing MVs transferred from microglia to neural cells upon B1R activation mediate neuronal damage. B1R antagonism improves the neurologic function and causes downregulation of both miR200c in microglial cells and the quantity of MVs released microglial cells. MVs from miR200c knockdown cells have the same effect on neuronal cells as MVs derived from cells treated with B1R antagonist (Tang et al., 2017).

who die early are unable to control SARS-CoV-2, while patients who die later suffer from diffuse tissue-damage and immunopathology (Nienhold et al., 2020) suggesting that late disease stage pathogenesis is apparently decoupled from acute viral-load.

3.4 Multiple pathologies of COVID-19 phenotypes may converge mechanistically

A hyper-inflammatory response is a major cause of severe disease and mortality (Del Valle et al., 2020). Our model demonstrates how excess activation of inflammatory signaling (Figure 8) turns productive inflammatory response and recruitment of immune cells, into a detrimental cytokine storm and immunopathology. Here, coagulation is no longer restricted to the site of infection/injury but leads to disseminated thrombotic events. Likewise, nociception (Figure 9) becomes chronic pain and increased barrier permeability (Figure 10) causes edema. Under normal inflammatory conditions, exocytosis is downregulated to disturb virus dissemination, however, when inflammation remains unresolved, the same mechanism can lead to impairment of neuronal signal transmission (Figure 11). Inflammation is also known to trigger senescence as a pathogen defense mechanism, which under hyper-inflammatory conditions may impair stem cell function and cause autoimmunity (Figure 12). Finally, inflammation triggers fibrogenesis for healing processes which under hyper-inflammation leads to scar formation. Importantly, these mechanisms can be triggered by an imbalance in ACE2-DAK-B1R signaling and associated regulatory components (e.g., miR200c or SIRT1) (Figure 1C).

Altogether, the model suggests that dysregulated homeostasis of eight mechanisms, alone or in combination, may contribute to the pathogenesis of major COVID-19 phenotypes. The multiple
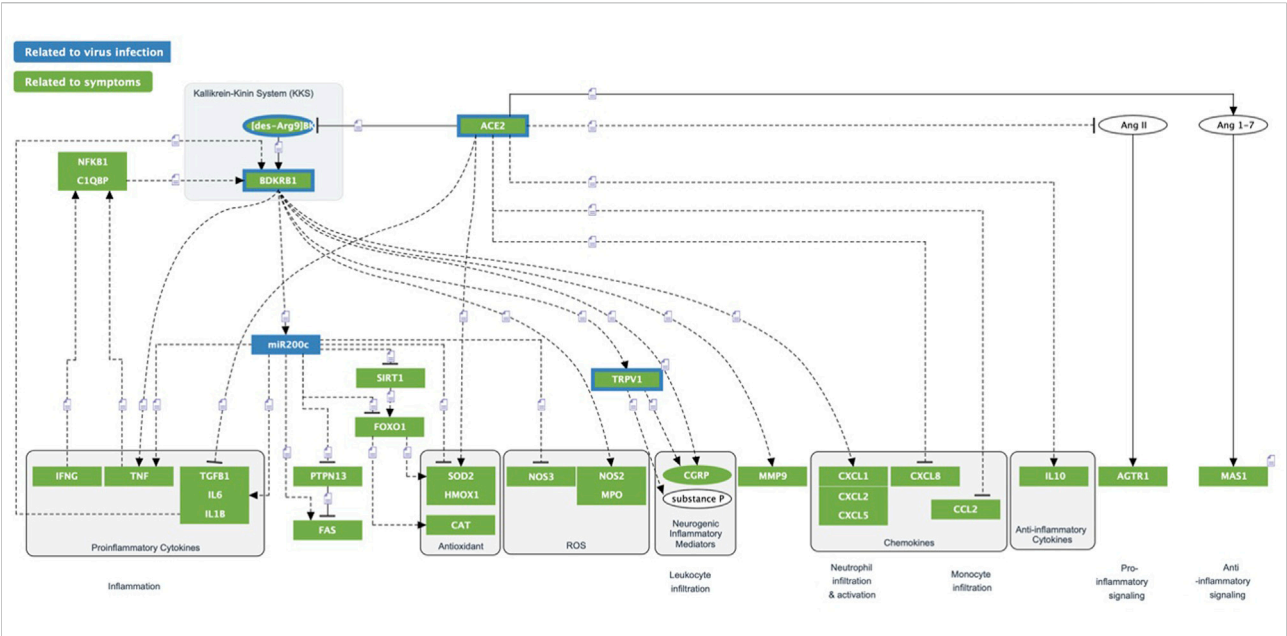
FIGURE 8
ACE2/DAK/B1R triggered inflammatory processes. Multiple inflammatory processes are triggered by the dysregulation of the ACE2/DAK/B1R axis. Excess activation of B1R triggers Cytokines/Chemokines, ROS inflammation and neurogenic inflammation.
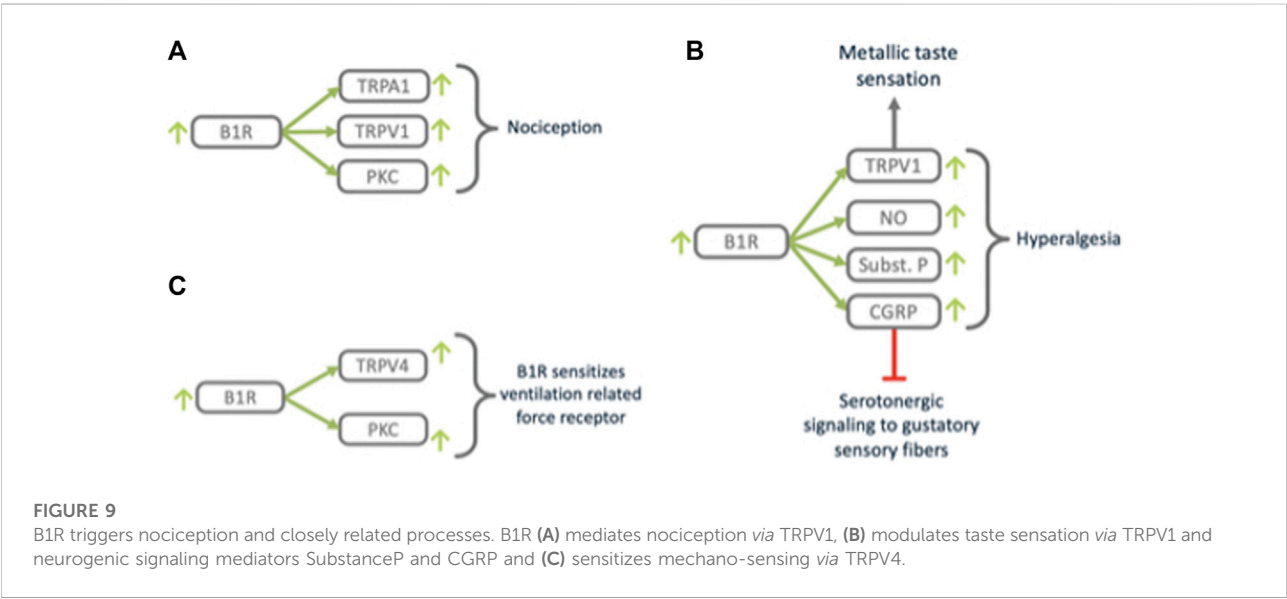
FIGURE 9
B1R triggers nociception and closely related processes. B1R **(A)** mediates nociception *via* TRPV1, **(B)** modulates taste sensation *via* TRPV1 and neurogenic signaling mediators SubstanceP and CGRP and **(C)** sensitizes mechano-sensing *via* TRPV4.

and seemingly unrelated clinical faces of COVID-19, including common disease symptoms (e.g., dry cough, myalgia, anosmia, transient diabetes and silent hypoxia), and severe manifestations (e.g., ARDS, lung fibrosis, acute coronary syndromes and thromboembolic events) may largely be linked to the pleiotropic activity of a few core molecular players and mechanisms involved in the host response (Figure 1). Interestingly, our model reveals that some more rare

phenotypes can be matched to other diseases sharing the same symptoms. Silent hypoxemia, for instance, causes the same symptoms as the Congenital Central Hereditary Hypoventilation Syndrome (CCHS). The molecular pathologies of both converge on the same molecular mechanism (Figure 11). A full analytical overview of the generated model is summarized in Supplementary Table S1 and made available *via* our COVID-19 Explorer.

FIGURE 10

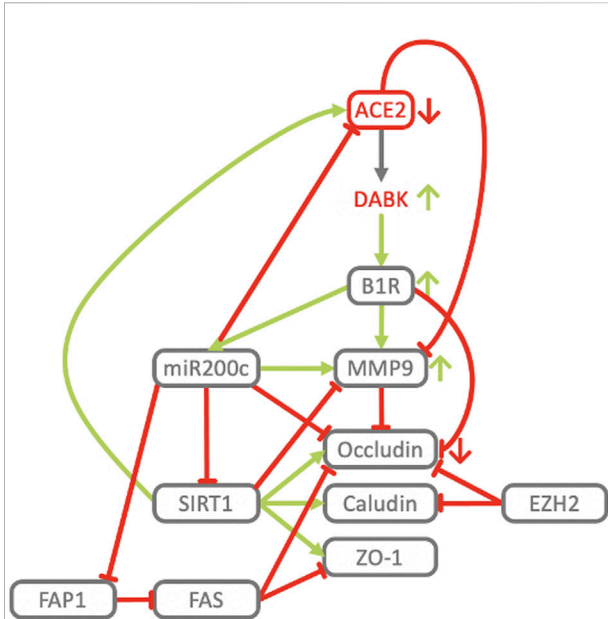Key players of the ACE2/DAK/B1R axis cooperate in the downregulation of tight junction proteins and induction of MMP9. B1R stimulation results in the loss of Occludin expression at tight junctions and an increase of vascular permeability (Mugisho et al., 2019). In neuroinflammatory diseases B1R contributes to inflammation and loss of blood-brain-barrier integrity, while inhibition of B1R protects mice from focal brain injury by reducing blood-brain barrier leakage and inflammation (Raslan et al., 2010). Neutrophils engage the KKS to open up the endothelial barrier in acute inflammation (Kenne et al., 2019). Activation of B1R induces expression and secretion of MMP-9 and MMP-2 (Matus et al., 2016). Active MMP2 and MMP9 degrade components of the alveolar basement membrane (Dunsmore and Rannels, 1996; Vu, 2001), non-matrix components such as integrins (Greenlee et al., 2007; Vaisar et al., 2009), and intercellular targets such as E-cadherin (Symowicz et al., 2007; Li et al., 2010). MMP-9 levels are elevated in ALI/ARDS (Davey et al., 2011) and predictive of the development of ARDS (Hsu et al., 2015). A distinct increase in circulating MMP-9 has been identified in COVID-19 patients with respiratory failure (Ueland et al., 2020). MMP-9 also exacerbates injury pathways in ischemic stroke, impairs and actively degrades components of the BBB, leading to the development of cerebral edema and hemorrhagic transformation (Turner and Sharp, 2016; Brilha et al., 2017). In CKD, MMP-9 activity is associated with resistant albuminuria (Pulido-Olmo et al., 2016). B1R blockade has been shown to dramatically reduce edema formation not only in ARDS but also in models of acute ischemic stroke (Austinat et al., 2009), traumatic brain injury (Raslan et al., 2010) and multiple sclerosis (Göbel et al., 2011). In accordance with B1R downregulating Occludin and inducing MMP-9, ACE2 deficiency has been associated with increased MMP-9 levels in myocardial infarction (Kassiri et al., 2009), while antago-miR200c, potentially *via* derepression of ACE2, inhibits MMP-9, increases Occludin mRNA and protein expression resulting in increased TJ permeability (Al-Sadi et al., 2017). In contrast, ectopic delivery of miR200c transcriptionally and translationally represses Occludin (Elhelw et al., 2017). The effect of other regulatory elements and effectors of our COVID-19 model, like SIRT1, EZH2 and FAS on barrier integrity is consistent with their respective mechanistic roles in regulating the ACE2-DAK-B1R axis. TMPRSS2 activity may also contribute to increased barrier permeability. TMPRSS2 cleaves and thereby activates PAR2 (Wilson et al., 2005). In airways PAR2 activation induces constriction, increases lung vascular and

(*Continued*)

FIGURE 10 (Continued)

epithelial permeability and pulmonary edema (Su et al., 2005). In concordance with its general protective role, activation of SIRT1 by Resveratrol maintains the epithelial barrier by increasing the expression of TJ proteins ZO1, Occludin and Claudin1 (Ma et al., 2014), while it negatively regulates MMP9 in diabetic retinopathy, and reduction of SIRT1 levels through oxidative stress confers an increase in MMP9 (Kowluru et al., 2014). Activation of FAS increases barrier permeability and decreases the expression of Occludin and ZO1 in the alveolar-capillary membrane *in vivo* and in human alveolar epithelium *in vitro* (Herrero et al., 2019). FAS is an effector of miR200c and among the most highly induced genes in response to miR200c overexpression (Tryndyak et al., 2010). At the same time miR200c represses FAP1, a negative regulator of FAS (Schickel et al., 2010). Consistent with this, Hepatitis C Virus induced miR200c down modulates FAP1 and promotes fibrosis (Ramachandran et al., 2013). EZH2-knockdown leads to upregulation of Occludin and Claudins (de Vries et al., 2015). EHZ2 expression increases with age, while aging exacerbates ALI-induced changes of the epithelial barrier, lung function, and inflammation. ALI in old mice showed 6x BALF protein, 2x neutrophils, higher levels of CXCL1, ICAM1, MMP-9 and significantly reduced Occludin (Kling et al., 2017).

3.5 Detailed examples of mechanistic convergence

3.5.1 Exocytosis

A number of different manifestations and symptoms of COVID-19 share a common underlying theme. Anosmia, ageusia, silent hypoxemia and neurological dysfunctions possibly result from impairment of signal transmission in the nervous system. As described above, there is a mechanism by which miR200c selectively represses Syntaxin1. Syntaxin1 is an important functional protein for neurotransmission and its dysfunction causes loss-of-function of neural cells (Davis et al., 1998; Watanabe et al., 2013). B1R activation induces secretion of microvesicles (MVs) carrying miR200c from microglial cells which are delivered to neural cells, where miR200c represses Syntaxin1 and damages these cells (Tang et al., 2017). Such a mechanism combines numerous mechanistic principles induced *via* the ACE2/DAK/B1R axis. Syntaxin1 is essential for the exocytosis of synaptic vesicles. Together with SNAP-25 and STXBP1, Syntaxin1 is the starting point for the assembly of the SNARE complex that drives vesicle fusion (Dawidowski and Cafiso, 2016). Mutations in these genes are associated with intellectual disabilities and seizures. Morphogenic analysis of SARS-CoV-2 in human airway epithelial cultures shows that virus release from cells occurs through exocytosis (Zhu et al., 2020). It is therefore possible that under normal circumstances, the B1R mediated downregulation of the synaptic machinery is part of a host defense mechanism, while excess activation of B1R leads to collateral damage. However, it is not known whether Syntaxin1 or other members of the Syntaxin family are responsible for vesicle fusion in this context.
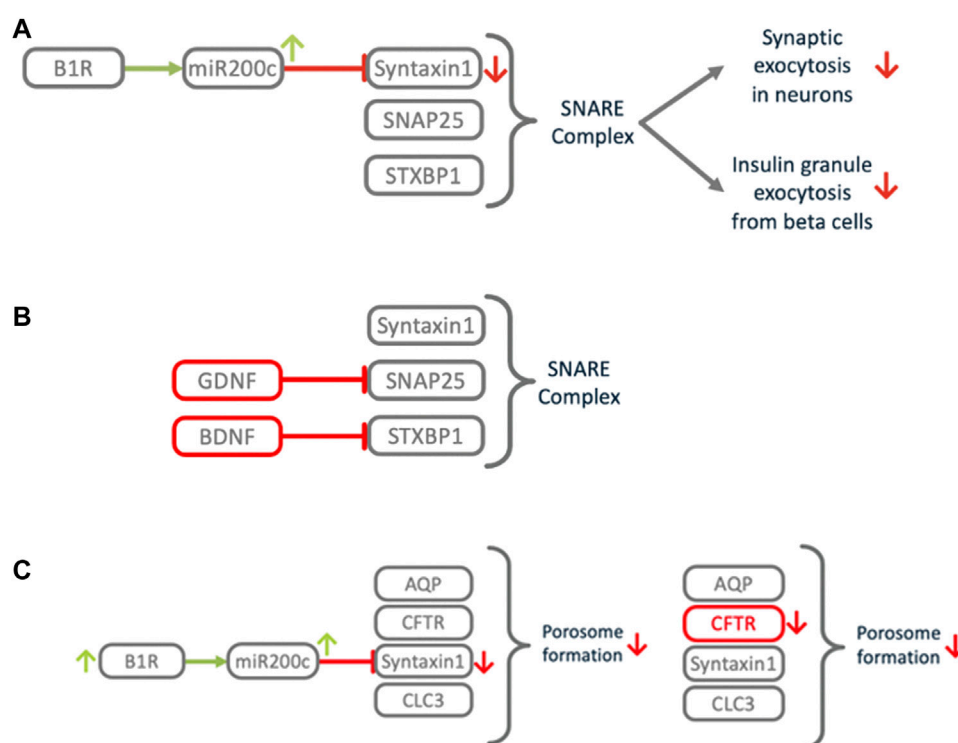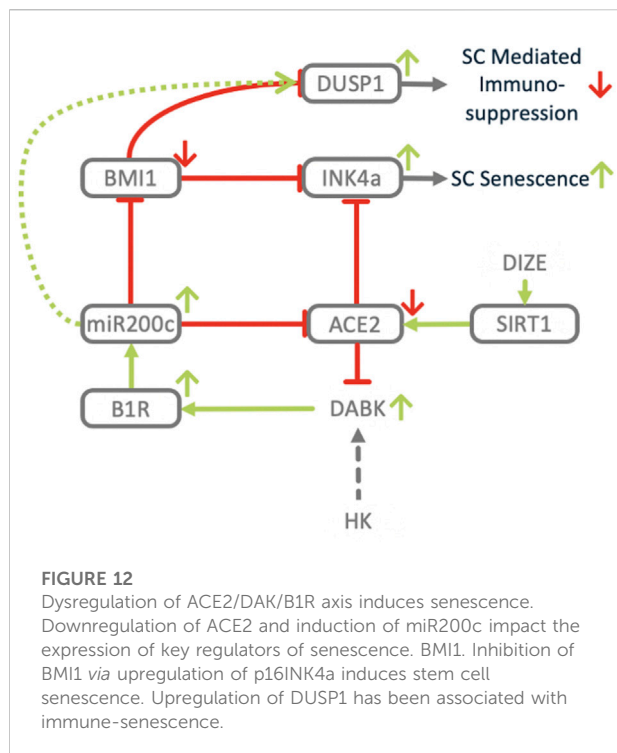
FIGURE 11
Impact of B1R induced miR200c on exocytosis and related processes. **(A)** Downregulation of Syntaxin1 impairs the SNARE complex, a central element of the exocytosis machinery. This impacts signal transduction between neurons, which might cause COVID-19 associated anosmia, ageusia, cognitive impairment or silent hypoxia. **(B)** Congenital central hypoventilation syndrome (CCHS) causes the same symptoms as COVID-associated silent hypoxemia. In CCHS defects in growth factors GDNF or BDNF cause downmodulation of SNAP25 or STXBP1, leading to SNARE dysfunction in the oxygen sensing carotid bodies. **(C)** COVID-19 associated thick mucus resembles the cystic fibrosis (CF) phenotype. In CF, mutations in CFTR cause porosome dysfunction, which impairs the SNARE dependent transient fusion of secretory vesicles at the porosome. Excess activation of B1R may induce local downregulation of Syntaxin1, which is also involved in this process, thus mimicking the CF phenotype.

3.5.2 Neurological dysfunction

COVID-19 patients can develop a range of neurological complications. A population study identified long-term cognitive deficits in patients who have recovered from COVID-19 (Hampshire et al., 2020). The scale of the observed deficits equates to an 8.5-point difference in IQ, which is equivalent to the average 10-year decline. One of the key factors contributing to age-related cognitive decline is the dysregulation of the fronto-temporal SNARE protein interactome (Ramos-Miguel et al., 2018). Global cognitive decline is associated with reduced SNARE complex levels (Syntaxin1, SNAP25, VAMP). Synapse dysfunction is initiated early in this process and occurs independent of neuropathology-driven synapse loss (Honer et al., 2012; Boyle et al., 2013), with the SNARE complex and its interactors representing a fingerprint of synaptic functionality. Reduced expression of Syntaxin1 because of B1R overactivation could provide an explanation for the cognitive decline in COVID-19 patients. The proposed virus independent mechanism could also explain the long-term impairment past acute symptoms.

3.5.3 Anosmia, ageusia

ACE2 is expressed in epithelial cells of oral mucosa and is highly enriched in epithelial cells of tongue (Xu et al., 2020), and in fungiform and circumvallate papillae (Shigemura et al., 2019). Olfactory sensory neurons express neither ACE2 nor TMPRSS2, however, epithelial support cells and stem cells express both genes, as do cells in the nasal respiratory epithelium (Sungnak et al., 2020). Non-neural expression of ACE2 and TMPRSS2 in the olfactory epithelium sustentacular cells occurs at levels comparable to that observed in lung cells (Brann et al., 2020). Thus, infection of these cell types, rather than sensory neurons may be responsible for anosmia in COVID-19 patients. Although sensory neurons are likely not targets of infection, intercellular communication *via* MVs could induce defects in taste and olfactory signal processing through the mechanism by which MV carrying miR200c are transferred from infected cells to sensory cells upon excess DAK and B1R activation. This mechanism could explain the observed complete anosmia and ageusia which is beyond mere taste modulation induced through mechanisms associated with nociception (see above). The

FIGURE 12
Dysregulation of ACE2/DAK/B1R axis induces senescence.
Downregulation of ACE2 and induction of miR200c impact the
expression of key regulators of senescence. BMI1. Inhibition of
BMI1 *via* upregulation of p16INK4a induces stem cell
senescence. Upregulation of DUSP1 has been associated with
immune-senescence.

synaptic release machinery of olfactory sensory neurons is centered on the Syntaxin1-dependent SNARE complex (Marcucci et al., 2009). Also taste cell synapses use the classical Syntaxin1 SNARE machinery for neurotransmitter release in circumvallate taste buds (Yang et al., 2007). Interestingly, Syntaxin1 and the SNARE complex are also potential targets in general anesthesia (van Swinderen and Kottler, 2014) and in the context of Huntington's disease it was demonstrated that miR200c contributes to neuronal dysfunction by targeting genes regulating synaptic function (Jin et al., 2012).

3.5.4 Silent hypoxemia

One of the distinctive features of COVID-19 is severe hypoxemia often associated with near normal respiratory system compliance, a condition that has been termed silent hypoxemia (Gattinoni et al., 2020; Ottestad et al., 2020; Wilkerson et al., 2020; Xie et al., 2020). This distinct symptom shares commonalities with the rare hereditary disease, congenital central CCHS (Figure 11). CCHS is a life-threatening disorder with impaired ventilatory response to hypoxia and hypercapnia which is related to carotid body (CB) dysfunction. CCHS patients show a significant decrease in the number of dopaminergic vesicles in oxygen sensor cells of CBs (López-Barneo et al., 2008). The oxygen sensor cells are strategically located at the bifurcation of the carotid artery, which supplies the brain. Upon arterial hypoxia they transmit signals to the respiratory center, which increases the frequency of

breathing. Dopamine is considered as the predominant transmitter of the CB sensor cells and sensor cells utilize the exocytotic apparatus and its components SNAP25 and Syntaxin1 for signaling (Koerner et al., 2004). BDNF as well as GDNF carry CCHS disease causing mutations (Sasaki et al., 2003). BDNF enhances synaptic functions and restores the severe synaptic dysfunction induced by STXBP1 deficiency. Mutations that disrupt STXBP1 binding to the SNARE severely impair vesicular transmitter release (Lee et al., 2019). GDNF induces expression of SNAP25 (Barrenschee et al., 2015). Defects in GDNF or BDNF lead to SNARE deficiency which impairs exocytosis in CB sensory cells and leads to decreased dopaminergic signaling. In COVID-19 patients B1R/miR200c mediated downregulation of Syntaxin1 in CB sensory cells, which are openly exposed to circulating virus particles, might induce similar impairment of the exocytotic machinery causing silent hypoxemia.

3.5.5 New-onset diabetes

New-onset Diabetes and severe metabolic complications of preexisting Diabetes, including diabetic ketoacidosis and hyperosmolarity, have been observed in patients with COVID-19 (Li et al., 2020a; Chee et al., 2020; Ren et al., 2020). A similar phenotype characterized by virus induced Diabetes was observed with SARS-CoV infections (Yang et al., 2010). In one controlled study (Yang et al., 2010), 39 patients who received no steroid treatment during the disease course and had no concomitant diseases such as pre-existing Diabetes, chronic hepatic, kidney, lung, cardiovascular disorders, cerebrovascular disorders, or blood dyscrasias before infection, were followed-up for 3 years. Fourteen of these 39 patients had Diabetes within 3 days of hospitalization, twenty after 2 weeks. Six patients had Diabetes at discharge and two patients still had Diabetes after 3 years of follow-up. In this study hyperglycemia was a predictor for death.

Pancreatic islet β-cells release Insulin *via* exocytosis of Insulin secretory granules. This process is mediated by distinct membrane fusion machineries (Gaisano, 2012). The fundamental components of this machinery are the three SNARE proteins Syntaxin1, SNAP-SNAP25 and VAMP (Südhof and Rothman, 2009). Syntaxin1 therefore plays a key role in Insulin granule exocytosis and replenishment. In Type-2 Diabetes (T2D), severely reduced islet Syntaxin1 levels contribute to Insulin secretory deficiency (Liang et al., 2017). In COVID-19 patients, ACE2 deficiency and B1R/miR200c mediated downregulation of SyntaxinA1 in β-cells may therefore lead to the induction of new-onset Diabetes. In many instances this phenotype persists even after the virus has been cleared, suggesting a mechanism that can propagate and maintain itself in the absence of virus. ACE2 deficiency has been directly linked to defects in Insulin secretion (Niu et al., 2008) and interestingly, B1R as well as CPN1, the enzyme that converts BK into the B1R ligand DABK have been researched as

preclinical targets in the context of Diabetes (El Akoum et al., 2017; Haddad and Couture, 2017). ACE2 deficiency has been directly linked to defects in Insulin secretion (Niu et al., 2008) and interestingly, B1R as well as CPN1, the enzyme that converts BK into the B1R ligand DABK, have been researched as preclinical targets in the context of Diabetes (El Akoum et al., 2017; Haddad and Couture, 2017). B1R antagonism and pharmacological blockade of CPN1 exerted similar beneficial effects in a rat model of Diabetes (Haddad and Couture, 2017). In mouse models, B1R antagonism has been shown to reverse hyperglycemia (Catanzaro et al., 2010). The anti-diabetic effect of B1R antagonism is consistent with the reduction of glycemia in B1R knockout mice (Seguin et al., 2008). Recently it has been found that B1R-expressing adipose tissue coordinates the metabolic response to diet-induced obesity in a cell-nonautonomous manner and furthers adipose tissue remodeling and the development of metabolic syndrome (Sales et al., 2019). miR200c has also been linked to the molecular pathology of Diabetes. It has been shown that miR200c diminishes Insulin production by inducing pancreatic β-cell damage (Belgardt et al., 2015), while suppression of miR200c improves β-cell function in patients with T2D (Roshanravan et al., 2018). miR200c has also been identified as a mediator of diabetic endothelial dysfunction, Diabetes-associated vascular complications (Zhang et al., 2016) and Diabetes-associated cardiac hypertrophy (Singh et al., 2017). Inhibition of miR200c has been shown to restore endothelial function in diabetic mice (Zhang et al., 2016).

3.5.6 Thick mucus

The mucous secretions found in the airways of COVID-19 patients (Mao L. et al., 2020) are reminiscent of those seen in cystic fibrosis (CF) patients (Martínez-Alemán et al., 2017). CF patients have elevated levels of highly viscous mucus in their lungs resulting from mutations that disrupt the CFTR gene (Cutting, 2015). A key property of mucus is its appropriate viscosity that enables its movement by the underlying cilia. Secretion of more viscous mucus hampers its proper transport, resulting in chronic and fatal airway disease. CFTR is known to interact with Syntaxin1, chloride channel CLC-3, and aquaporins to form the porosome complex. The porosome is essential for mucus hydration and controlling viscosity (Jena, 2014). The process of secretion *via* the porosome is similar to exocytosis: a pore is formed through transient fusion of a secretory vesicle at the porosome base *via* SNARE proteins resulting in the formation of a fusion pore. In COVID-19 patients B1R/miR200c mediated downregulation of Syntaxin1 in lung epithelium, a key component for secretion *via* the porosome, could mimic the molecular pathology of CF.

3.5.7 Barrier permeability

One of the most common causes of hospital admission and death in patients with COVID-19 is ARDS, a clinical syndrome

characterized by acute lung inflammation and increased-permeability pulmonary edema (Ware, 2020). Severe lung failures such as alveolar edema, ARDS and ALI are caused by increased permeability of the alveolar/airway epithelium and exudate formation. This barrier is sealed by TJs between cells, which are composed of Occludin, ZO1 and various claudins. A loss of TJ permselectivity in the airways results in an uncontrolled leakage of high molecular weight proteins and water into the airways, which finally results in the formation of alveolar edema and ARDS (Wittekindt, 2017). Inflammatory lung diseases like asthma, COPD and allergic airway inflammation may predispose patients to severe lung failures. In COPD, for example, the expression of Occludin and ZO1 in alveolar epithelia cells is reduced (Hong, 2012). The same is true for ventilation induced ALI where expression of Occludin is significantly decreased and alveolar permeability is increased. Upregulation of Occludin can reduce ventilation-induced lung injury (Liu et al., 2014). In sepsis-induced ARDS/ALI downregulation of TJ-proteins coincides with upregulation of inflammatory cytokines. Unfractionated heparin attenuates ALI by upregulating Claudin, ZO1 and Occludin (Liu J. et al., 2019). Down-regulation of TJ proteins is also observed in other pathologies involving epithelial barrier defects such as in chronic kidney disease, which is characterized by a marked depletion of TJ proteins (Claudin1, Occludin, and ZO1) in the gastric epithelium (Vaziri et al., 2013) or Kawasaki disease (KWD) where decreased ZO1 levels are associated with intestinal barrier dysfunction (Lai et al., 2020).

Epithelial TJs regulate alveolar air-fluid balance in the lungs, the production of appropriately concentrated urine in the kidney, as well as the absorption of nutrients and containment of bacteria throughout the gastrointestinal tract (Anderson and Van Itallie, 2009). TJs in endothelia maintain intravascular volume and regulate the flux of fluid and solutes between blood vessels and organ parenchyma (Rahimi, 2017). Therefore, endothelial and epithelial barrier dysfunction can result in malabsorption of nutrients, translocation of pathogens, capillary leak, interstitial edema, tissue dysoxia, and ultimately organ failure.

B1R stimulation results in the loss of Occludin expression at TJs and an increase of vascular permeability (Mugisho et al., 2019). In neuroinflammatory diseases B1R contributes to inflammation and loss of Blood: Brain barrier (BBB) integrity. Inhibition of B1R protects mice from focal brain injury by reducing BBB leakage and inflammation (Raslan et al., 2010). Neutrophils engage the KKS to open up the endothelial barrier in acute inflammation (Kenne et al., 2019). MMP9 also plays an important role in BBB break-down through its ability to degrade the base membrane. Levels of MMP9 are elevated in ALI/ARDS (Davey et al., 2011) and MMP9 activity is predictive of the development of ARDS (Hsu et al., 2015). A distinct increase in circulating MMP9 has been identified in COVID-19 patients with respiratory failure (Ueland et al., 2020). MMP9 also exacerbates injury pathways in ischemic stroke, impairs and

actively degrades components of the BBB, leading to the development of cerebral edema and hemorrhagic transformation (Turner and Sharp, 2016; Brilha et al., 2017). In chronic kidney disease, MMP9 activity is associated with resistant albuminuria (Pulido-Olmo et al., 2016). Activation of B1R induces expression and secretion of MMP9 and MMP2 (Matus et al., 2016). Active MMP2 and MMP9 degrade components of the alveolar basement membrane (Dunsmore and Rannels, 1996; Vu, 2001), non-matrix components such as integrins (Greenlee et al., 2007; Vaisar et al., 2009), and intercellular targets such as E-cadherin (Symowicz et al., 2007; Li et al., 2010).

B1R blockade has been shown to dramatically reduce inflammatory processes and edema formation not only in ARDS but also in models of acute ischemic stroke (Austinat et al., 2009), traumatic brain injury (Raslan et al., 2010) and multiple sclerosis (Göbel et al., 2011). Consistent with B1R downregulating Occludin and inducing MMP9, ACE2 deficiency has been associated with increased MMP9 levels in MI (Kassiri et al., 2009), while antago-miR200c, potentially *via* derepression of ACE2, inhibits MMP9, increases Occludin mRNA and protein expression resulting in increased TJ permeability (Al-Sadi et al., 2017). In contrast, ectopic delivery of miR200c transcriptionally and translationally represses Occludin (Elhelw et al., 2017). The effect of other regulatory elements and effectors (Figure 2), like SIRT1, EZH2 and FAS on barrier integrity is consistent with their respective mechanistic roles in regulating the ACE2-DAK-B1R axis. TMPRSS2 activity may also contribute to increased barrier permeability. TMPRSS2 cleaves and thereby activates PAR2 (Wilson et al., 2005).

In airways, PAR2 activation induces constriction, increases lung vascular and epithelial permeability and pulmonary edema, triggers SubstanceP release and increases CXCL2 production (Su et al., 2005). Interestingly, SIRT1 mediates a protective effect on barrier integrity, which is in concordance with its regulatory effect on the ACE2/DAK/B1R signaling axis. Activation of SIRT1 by Resveratrol maintains the epithelial barrier by increasing the expression of TJ proteins ZO1, Occludin and Claudin1 (Ma et al., 2014), while it negatively regulates MMP9 in diabetic retinopathy, and reduction of SIRT1 levels through oxidative stress confers an increase in MMP9 (Kowluru et al., 2014). Activation of FAS increases barrier permeability and decreases the expression of Occludin and ZO1 in the alveolar-capillary membrane *in vivo* and in human alveolar epithelium *in vitro* (Herrero et al., 2019). FAS is an effector of miR200c and among the most highly induced genes in response to miR200c overexpression (Tryndyak et al., 2010). At the same time miR200c represses FAP1, a negative regulator of FAS (Schickel et al., 2010). Consistent with this, Hepatitis C Virus induced miR200c down-modulates FAP1 and promotes fibrosis (Ramachandran et al., 2013). Furthermore, and also consistent with its regulatory role, EZH2-knockdown is accompanied by

upregulation of Occludin and Claudins (de Vries et al., 2015). An important factor influencing barrier integrity is age. Aging exacerbates ALI-induced changes of the epithelial barrier, lung function, and inflammation. ALI in old mice showed 6x BALF protein, 2x neutrophils, and higher CXCL1, ICAM1, MMP9 and significantly reduced Occludin levels (Kling et al., 2017).

3.6 Complex COVID-19 phenotypes

Complex COVID-19 phenotypes may involve a combination of dysregulated mechanisms. While many symptoms are directly related to individual disease mechanisms triggered by B1R overactivation, more complex organ-level syndromes may involve combinations. For instance, in COVID-19 lungs, hyper-inflammatory phenotypes coincide with increased epithelial permeability and resulting edema formation (Carsana et al., 2020), neutrophilia (Fu et al., 2020), micro-thrombotic events (Merrill et al., 2020), endotheliitis (Varga et al., 2020), fibrosis (Grillo et al., 2020) and impaired self-renewal capacity (Shao et al., 2020). Similarly, cardiovascular (Nishiga et al., 2020), kidney (Pei et al., 2020) and neurologic (Bridwell et al., 2020) manifestations arise from a combination of the SARS-CoV triggered pathogenic events. Vasculitis and KWD as well as GBS serve as examples to describe the concerted interplay of the different pathogenic mechanisms triggered in the context of our COVID-19 disease model.

3.6.1 Vasculitis/Kawasaki disease (KWD)

An unusually high incidence of KWD has been reported for children suffering from COVID-19 (Belhadjer et al., 2020; Jones et al., 2020; Moreira, 2020; Nathan et al., 2020; Riphagen et al., 2020; Rivera-Figueroa et al., 2020; Toubiana et al., 2020; Verdoni et al., 2020). The appearance of clinical manifestations resembling KWD has been attributed to a new phenotype of autoimmunity (Rodríguez et al., 2020). Post-mortem examination of COVID-19 patients reveals damage in many organ systems suggesting a general vascular dysfunction (Menter et al., 2020). Cases of COVID-19 associated cutaneous vasculitis have also been reported (Dominguez-Santas et al., 2020; Papa et al., 2020). Loss of TJs and barrier integrity may also play a role in vasculitis and KWD. The TJ protein ZO1 is decreased in KWD and has been associated with intestinal barrier dysfunction of KWD (Lai et al., 2020). Reduced mRNA expression of multiple intestinal TJs was observed in a murine model of KWD (Rivas et al., 2017) and experimental data as well as clinical, genetic, and transcriptome evidence from patients converge to suggest a key role of IL1β in KWD (Noval Rivas et al., 2019), which is a key upstream regulator of B1R.

Transfer of B1R bearing MVs was identified as novel inflammatory mechanism in vasculitis (Kahn et al., 2017), which is also involved in endothelium-neutrophil communication (Tharaux and Dhaun, 2017). Patients with

acute vasculitis show high levels of circulating B1R positive endothelial MVs. B1R positive MVs induce a neutrophil chemotactic effect, which can be blocked by B1 receptor antagonist. Patient plasma induces the release of more B1R positive MVs from endothelial cells, an effect that is dependent on the presence of B1R positive MVs in patient plasma. SERPING1 depleted plasma promoted excessive release of B1R positive endothelial MVs. Addition of SERPING1 or B1R-antagonist inhibited this effect (Mossberg et al., 2017). Serum level of miR200c one of the key regulators in our model has been identified as a suitable diagnostic biomarker and potential target in KWD (Zhang R. et al., 2017).

It has also been demonstrated that plasma from Vasculitis patients has significantly more C3- and C9-positive endothelial MVs than controls. Perfusion of patient acute-phase plasma samples over glomerular endothelial cells induced the release of significantly more complement-positive MVs, in comparison to remission or control plasma. Complement bearing MVs are strongly reduced by B1R antagonism or SERPING1. Likewise, perfusion of glomerular endothelial cells with SERPING1-depleted plasma induced the release of complement-positive MVs, which in turn was significantly reduced by B1R antagonism or SERPING1 (Lopatko Fagerström et al., 2019). Vasculitis often occurs in the context of connective tissue disease. In patients with connective tissue disease, autoantibodies to ACE2 are associated with vasculopathy (Takahashi et al., 2010) and severity of interstitial lung disease is positively correlated with PBMC miR200c, as has been demonstrated for patients with Sjogren's disease (Jiang et al., 2017). Antineutrophilic autoantibodies (ANCAs), which play a role in endogenous systemic vasculitis (Silva de Souza, 2015) have been identified in cases of COVID-19 associated vasculitis and glomerulonephritis (Uppal et al., 2020) and an autoimmune component has been ascribed to the COVID-19-associated KWD phenotype (Rodríguez et al., 2020). Indeed, anti-endothelial cell autoantibodies have been identified in KWD patients (Sakurai, 2019). Thus, the ACE2/DAK/B1R axis associated mechanism driving autoimmunity and stem cell senescence may play a role in the COVID-19 associated pathogenesis of Vasculitis/KWD.

In summary, Vasculitis and KWD are associated with activation of the complement system, KKS signaling, inflammation, loss of TJs and an autoimmune component. All of these are triggered by the same mechanism underlying our disease model. The disease mechanism of KWD converges with the COVID-19 model at the molecular level, as exemplified by miR200c. This micro-RNA is both a disease specific diagnostic biomarker for KWD, and a key element of the disease model contributing to increased vascular permeability and autoimmunity. Intercellular communication mediated by MVs plays an equally central role in both models.

3.6.2 Guillain-Barré (GBS)

An increasing number of case reports suggest that COVID-19 may induce GBS (Camdessanche et al., 2020; Zhao et al., 2020). GBS is an acute inflammatory demyelinating polyradiculoneuropathy resulting from an autoimmune attack on the myelin. Interestingly, B1R activation contributes to demyelination in multiple sclerosis, while B1R inhibition or its genetic deletion decrease the neuroinflammatory response and myelin loss in the spinal cord (Dutra et al., 2011) and the upstream kininogen was established as a key mediator of neurodegeneration (Langhauser et al., 2012). Serum antibodies against glycolipids, mainly gangliosides, are detected in about 60% of patients with GBS and its variants and play a crucial role in the pathogenic mechanisms of GBS (Uchibori and Chiba, 2015). Overactivation of the ACE2/DAK/B1R axis may mechanistically contribute to the development of GBS and drive the autoimmune reaction.

3.7 Risk factors

The model reveals that many of the identified risk-factors comprising obesity, age, sex, smoking and diverse comorbidities such as diabetes or Alzheimer`s disease (Holman et al., 2020; Zhou et al., 2020) are directly related to overlapping pathogenic mechanisms.

3.7.1 Type-2 diabetes (T2D)

It appears that T2D is both a risk factor for severity and in the form of new-onset diabetes, a consequence of COVID-19 (Table 1). Subjects who develop T2D have a complex phenotype with defects in insulin secretion and insulin resistance in target tissues. Key players of our COVID models are associated with both mechanisms. In T2D the ACE2: ACE ratio negatively correlates with HbA1C and loss of ACE2 exacerbates cardiovascular complications of T2D. A potential therapeutic role of ACE2 in the context of T2D has mostly been discussed in the light of its role in the RAS (Batlle et al., 2010). On the other hand, in T2D, SERPING1 is downregulated, while KNG1 is upregulated (Zhang et al., 2013), highlighting a role for the KKS system. This provides additional context to reduced ACE2 activity and its effect on KKS-driven accumulation of DAKs and overactivation of B1R.

According to our model, ACE2 deficiency and B1R/miR200c may mediate the downregulation of Syntaxin1 in β-cells, which could impair exocytosis in pancreatic β-cells and hence hamper insulin secretion. Recently, it has been shown that upregulated circulating miR-200c in plasma may increase the risk of severe COVID-19 for obese individuals (Papannarao et al., 2021). On the other hand, miR200c is a biomarker of insulin resistance in obesity. miR200c diminishes insulin production by inducing pancreatic β-cell damage, while its suppression improves β-cell function in T2D and restores endothelial function.

TABLE 1 Readouts of COVID-19 trials utilizing drugs that have the potential to increase ACE2 expression in the context host response (Chee et al., 2020).

System/Drug group	Drug	Target	Indication (inclusion criteria)	Outcome	References (PMID)
Kinin-Kallikrein	Berinert (C1 inhibitor)	C1R and C1S	Severe COVID-19 pneumonia: SpO2 ≤ 94% in ambient air or Pa02/FiO2 ≤ 300 mmHg	- No change in "time to clinical improvement" - No change to coagulation parameters - Eosinophils increased	3669276 Yamakawa et al. (1987)
	Icatibant (B2R antagonist)	BDKRB2	Severe COVID-19 pneumonia: SpO2 ≤ 94% in ambient air or Pa02/FiO2 ≤ 300 mmHg	- No change in "time to clinical improvement" - No change to coagulation parameters - Eosinophils count increased	3669276 Yamakawa et al. (1987)
	Icatibant (B2R antagonist)	BDKRB2	Severe COVID-19 pneumonia: >3 L/min supplemental oxygen, a CT severity score of ≥7	- 89% (8/9 patients) had reduction of >3 l/min oxygen supplementation 24 h post treatment - No severe adverse events - No clear association with D-dimer levels or fever	32789513 van de Veerdonk et al. (2020)
	Conestat Alfa (C1 inhibitor)	C1R and C1S	Moderate or severe COVID-19 pneumonia by CT scan, C-reactive protein level of at least 30 mg/L oxygen saturation of <93% at rest in ambient air	- Immediate defervescence in 4 out of 5 patients - 2.5x reduction in intubation or death - No change in length of hospitalization	32922409 Urwyler et al. (2020)
ACE Inhibitors and ARBs	Various ACEi and ARBs	ACE and ATR1	COVID-19 patients with ACEi or ARBs prescription before contracting COVID-19. The study compared outcomes between patients randomly assigned to discontinue or continue ACEi or ARBs upon hospitalization	- No difference in outcomes between discontinued or continued use of ACEi or ARBs during COVID-19 hospitalization	33422263 Cohen et al. (2021), 33464336 Lopes et al. (2021)
TRP channels	GSK2798745 (antagonist)	TRPV4	Lung congestion in patients with heart failure: Patients with Heart failure NYHA Class II/III	- Improved lung diffusing capacity for carbon monoxide DLco (only trend not significant) - No serious adverse events	32227554 Stewart et al. (2020), 30637626 Goyal et al. (2019)
Vitamin D	Calcifediol		Patients hospitalized with COVID-19 acute respiratory infection (Radiography confirmed)	- Significant reduction of ICU admission (2% vs. 50%) - Limitation—comparison groups not fully matched - No information on BMI (and thus potential obesity)	32871238 Entrenas Castillo et al. (2020)
Calcium release-activated Calcium (CRAC) channel	Auxora (CM4620) (inhibitor)	ORAI1	Severe COVID-19 pneumonia (chest imaging) with respiratory compromise (e.g. ≥ 30 breaths/min, heart rate ≥125 bpm, $SpO_2 < 93\%$ on room air or $PaO_2/FiO_2 < 300$)	- Reduced median time to recovery (5 vs. 12 days) - Reduced risk of intubation (18% vs. 50%) - Combined death or intubation hazard ratio: 0.23	32795330 Miller et al. (2020)

TABLE 1 (*Continued*) Readouts of COVID-19 trials utilizing drugs that have the potential to increase ACE2 expression in the context host response (Chee et al., 2020).

System/Drug group	Drug	Target	Indication (inclusion criteria)	Outcome	References (PMID)
Non-steroidal anti-inflammatory (NSAIDs)	Acetylsalicylic acid		Patients hospitalized with COVID-19 pneumonia	- Significantly reduced in-hospital death rate	33476420 Meizlish et al. (2021)
	Acetylsalicylic acid + Anti-coagulation (tirofiban, clopidogrel, fondaparinux)		Patients with COIVD19 severe respiratory failure requiring Continuous positive airway pressure (CPAP)	- Reduction in A-a O_2 gradient - Increased PaO2/FiO2 ratio - Earlier weaning from CPAP	32450344 Viecca et al. (2020)
Glucose lowering drugs	Metformin	PRKAB1, GPD1, ETFDH	Study of susceptibility to contract COVID-19 in diabetic patients using Metformin	- No significant association of metformin and risk of COVID-19	33560344 Wang et al. (2021)
	Metformin	PRKAB1, GPD1, ETFDH	Hospitalized COVID-19 patients using Metformin prior to COVID-19 infection	- Reduced mortality/risk of death - Reduced risk of acute ischemic stroke - Some studies point to higher positive effect in females - Some studies show initially more severe course of disease in diabetic patients but overall lower mortality	33745895 Cheng et al. (2021), 33662839 Ghany et al. (2021), 33580540 Bramante et al. (2021), 33519709 Crouse et al. (2021), 33309936 Lalau et al. (2021), 33232684 Lally et al. (2020), 33471718 Li et al. (2020b), Meta-analyses: 32844132 Hariyanto and Kurniawan (2020), 33395778 Lukito et al. (2020)
	Gliptin (DPP4 inhibitor)	DPP4	Meta-analysis of 9 studies with COVID-19 patients using DPP-4 inhibitor	- Reduced risk ration (RR) of death (0.76)	33838614 Rakhmat et al. (2021)

miR200c downregulates IRS1, which is associated with insulin resistance. It also targets PGC1A, which controls the hepatic ratio of IRS1 and IRS2 and low PGC1A is associated with insulin resistance. PGC1A is also negatively regulated by TRPV4, while TRPV4 antagonists reduce high-fat diet-induced obesity, insulin resistance, diabetic nephropathy, retinopathy, and neuropathy. TRPV4 in turn is sensitized by B1R and its activity is mediated *via* a B1R-PKCε axis. B1R sensitizes PKCε and induces its translocation. PKCε activity has been demonstrated to contribute to lipid-induced insulin resistance. Furthermore, B1R in adipose tissue controls the response to diet-induced obesity and its deletion protects from obesity and improves insulin sensitivity.

In summary, combined evidence from the COVID-19 model connects key players both directly and synergistically to defective insulin secretion and insulin resistance in target organs, providing a rationale for both diabetes as risk factor and a consequence of COVID-19.

3.7.2 Co-morbidities and novel onset diseases

Low levels of ACE2 are associated with more severe outcomes and fatality in severe lung diseases. ACE2 plays a role in asthma, COPD, pulmonary fibrosis, PH, ALI and ARDS amongst others. We also examined the role of ACE2 in disease contexts other than lung.

In general, ACE2 plays a protective role and in many of these and it is often downregulated in disease affected tissue. This, for example, is the case in T2D, where a negative correlation between ACE2/ACE vs. HbA1C is observed and loss of ACE2 exacerbates cardiovascular complications in Diabetes (Joshi et al., 2019). ACE2 also plays a role in atherosclerosis, heart failure, cardiac fibrosis, ventricular remodeling, arrhythmia, cerebral ischemia, chronic kidney disease, diabetic nephropathy, liver diseases. Genetic variants of ACE2 are associated with cardiovascular risk, hypertension, hypertensive left ventricular hypertrophy, essential hypertension, AF and cardiomyopathy.

A similar collection of diseases and related phenotypes is affected by pathological B1R signaling (see above). Hence, there is a molecular rationale for the observed COVID-19 co-morbidity risk phenotypes. If the virus directly or indirectly (*via* MVs) affects cells of such diseased organ systems, the downregulation of ACE2 and induction of B1R signaling may aggravate pre-existing conditions or even induce the onset of disease in patients at risk. New-onset Diabetes is one such example of a COVID-19 induced disease phenotype. A bi-directional interplay between the pathogenic mechanisms triggered by COVID-19 and those underlying T2D may both define T2D as severity risk-factor and COVID-19 as a trigger for accelerated development of a Diabetes phenotype.

3.7.3 Smoking and air pollution

Cigarette smoking and air pollution have been associated with higher risk of severe COVID-19 outcomes (Comunian et al., 2020). Since cigarette smoking and nicotine reduce the expression of ACE2, this might contribute to the development of cardiovascular and pulmonary diseases (Jia, 2016; Oakes et al., 2018; Yue et al., 2018). Next, exposure to air particulate matters induces B1R and Kallikrein in lung and heart (Aztatzi-Aguilar et al., 2015) and cigarette smoke induced a significant upregulation in B1R expression level (Al Hariri et al., 2016). In rat lung, cigarette smoke leads to an enhanced expression of B1R (5x) and IL1β (30x), while no increase in levels of B2R or TNFα was observed (Lin et al., 2010). Nicotine also enhances EZH2 expression and EZH2 dependent gene silencing (Vaz et al., 2017; Kumari et al., 2018) which may contribute to the development of COPD (Anzalone et al., 2019). As discussed above, EZH2 is a negative regulator of ACE2 and a positive regulator of TMPRSS2. Thus, air pollution and cigarette smoke may affect pathogenic mechanisms that are involved in COVID-19 pathogenesis at multiple levels, which may increase disease severity in a synergistic manner.

3.7.4 Age and gender

Variation in expression of both TMPRSS2 and ACE2 suggests a credible hypothesis for correlation of vulnerability to SARS-CoV infection to age groups (Wang et al., 2020). However, no significant differences in the infection rate of males and females are found with increasing numbers of COVID-19 patients (Chen J. et al., 2020; Zhang J. J. et al., 2020; Brüssow, 2020; Huang et al., 2020). On the other hand, the link between age and severity of COVID-19 seems well established and the same is true for co-morbidities which are associated with decreased levels of ACE2 which are mostly age-related diseases. Decreased ACE2 levels found in older patients and cardiovascular disease increase the likelihood of severe COVID-19 (AlGhatrif et al., 2010). A negative correlation between ACE2 expression and COVID-19 fatality could be established at both population and molecular levels (Chen N. et al., 2020). Highest ACE2 expression levels were identified Asian females (>30% higher than other ethnic groups) and an age-dependent decrease was observed across all ethnic groups. A strong decrease in ACE2 was found T2D patients and data from human and mice revealed reduced ACE2 expression with inflammatory cytokine treatment and upregulation by estrogen and androgen, both of which decrease with age.

Animal models show that ACE2 expression is dramatically reduced with ageing, with significantly higher ACE2 levels in old females compared to males (Xie et al., 2006). In hematopoietic stem cells, the ACE2/ACE ratio is negatively correlated with age in healthy and in diabetic individuals (Joshi et al., 2019). In airway epithelial cells, DNA methylation near the transcription start site of the ACE2 gene associated with biological age (Corley and Ndhlovu, 2020) and differences in methylation patterns are observed between males and females (Fan et al., 2017). ACE2 is located on the X-Chromosome in a region that escapes X-inactivation (Tukiainen et al., 2017). On the other hand, ACE2 expression is controlled by SRY3, a transcription factor encoded by a gene located on the Y chromosome. SRY3 decreases the activity of ACE2 promoter by 0.5-fold (Milsted et al., 2010), while estrogen upregulates ACE2 expression and causes a protective shift in ACE/ACE2 ratio at the mRNA and protein level (Bukowska et al., 2017). This is consistent with the finding that the rate of degradation of DABK is much higher in women compared with men (Cyr et al., 2001).

The age-dependent expression patterns of EZH2 and SIRT1 are consistent with their respective effects on ACE2 expression. While the activity of ACE2 activator SIRT1 decreases (Hwang et al., 2013), the activity of ACE2 repressor EZH2 increases with age (Dozmorov, 2015; Han and Sun, 2020). Energy sensing via AMPK and SIRT1 plays a role in ACE2 regulation. The negative correlation of ACE2 to T2D is mirrored by the finding that SIRT1 in negatively correlated with hyperglycemia (Yang et al., 2018). Caloric restriction via energy sensing mechanisms results in SIRT1 activation (Waldman et al., 2018). SIRT1 activators have been linked to beneficial effects on COVID-19 severity and mortality. A recent study suggests positive effects of Metformin (Bramante et al., 2020). Also Melatonin has been suggested as potential COVID-19 treatment (Zhang R. et al., 2020) and a clinical trial assessing the efficacy of Melatonin as preventive drug is underway (Ziaei et al., 2020) (NCT04353128). Beneficial effects of SIRT1 activator resveratrol are also being discussed (Filardo et al., 2020). The potential effect of SIRT1 activation is consistent with its role in attenuation of airway inflammation, ARDS, ALI, fibrosis, endothelial barrier dysfunction and edema and maintaining the epithelial barrier by regulating TJs, described above. Metformin, Melatonin and Resveratrol are well tolerated inducers/activators of SIRT1 that could serve as potentially preventive interventions.

3.7.5 Obesity

Numerous components of our COVID-19 disease model are directly associated with obesity. In obesity, factors that contribute to the molecular pathology of COVID-19 are activated, while protective factors are downmodulated. ACE2 reduces cytokine release and inhibits signaling pathways of tissue fibrosis in experimental models of obesity (Rodrigues Prestes et al., 2017) and its deficiency in epicardial adipose tissue worsens inflammation and cardiac dysfunction in response to diet-induced obesity (Patel et al., 2016). The adipocytokine Apelin induces ACE2 expression (Sato et al., 2013) while its expression is decreased in adipose tissue (Wu et al., 2011). The ACE2 activator, SIRT1, shows an inverse correlation with adiposity. SIRT1 levels strongly negatively correlate with the amount of visceral fat (Mariani et al., 2018) B1R in adipocytes regulates glucose tolerance and predisposition to obesity (Mori et al., 2012) and

B1R knockout mice have been shown to be resistant to obesity induced by a high-fat diet (Morais et al., 2015). Furthermore, B1R expression in adipocytes controls adiposity and in turn contributes to whole-body Insulin sensitivity (Sales et al., 2019). It has recently been found that B1R-expressing adipose tissue coordinates the metabolic response to diet-induced obesity and furthers adipose tissue remodeling and the development of metabolic syndrome. Thus, B1R antagonism is considered as therapeutic tool for the treatment of obesity and Diabetes (El Akoum et al., 2017). In conclusion, the same systems that are derailed by the virus in our COVID-19 model are disturbed in obesity and contribute to the development and progression of obesity related diseases (Sales et al., 2019). Thus, B1R antagonism is considered as therapeutic tool for the treatment of obesity and Diabetes (El Akoum et al., 2017). In conclusion, the same systems that are derailed by the virus in our COVID-19 model are disturbed in obesity and contribute to the development and progression of obesity related diseases.

3.7.6 Alzheimer's disease (AD)

APOE4 is the major AD susceptibility gene (Corder et al., 1993). Homozygous APOE4 has also been associated with significant risk for the development of severe COVID-19, as well as death following infection (Kuo et al., 2020). New research has demonstrated that APOE4-associated cognitive decline in AD is associated with breakdown of the BBB, independent of AD pathology (Montagne et al., 2020). APOE4 activates the CypA-MMP9 pathway leading to accelerated BBB breakdown causing neuronal and synaptic dysfunction, while blockade of the CypA-MMP9 pathway restores BBB integrity followed by normalization of neuronal and synaptic functions. This disease mechanism is complementary to the COVID-19 model resulting in barrier dysfunction through downregulation of TJ proteins and induction of MMP9. AD therefore might constitute a risk factor for COVID-19 severity, while COVID-19 from a mechanistic perspective could contribute to AD progression (see above). Interestingly, CypA interacts with the SARS-CoV Nsp1 protein and plays an important role in virus replication (Liu et al., 2020).

3.7.7 GWAS risk factor converging with ACE2, B1R and mir200c signaling

A genome-wide association study (GWAS) involving 1980 patients with COVID-19 and severe disease (defined as respiratory failure) identified a 3p21.31 gene cluster as a genetic susceptibility locus in patients with COVID-19 with respiratory failure and confirmed a potential involvement of the ABO blood-group system. At locus 3p21.31, the association signal spans the genes SLC6A20, LZTFL1, CCR9, FYCO1, CXCR6 and XCR1 (Ellinghaus et al., 2020). Interestingly, the sodium-imino acid transporter 1 (SIT1/SLC6A20) which co-transports with high affinity L-proline, methylaminoisobutyrate, methyl-proline, and hydroxyproline with sodium and chloride,

functionally interacts with ACE2 (Singer and Camargo, 2011). Several members of the SLC6 family transporters, neutral amino acid transporters SLC6A19, SLC6A18 and SLC6A20 interact with ACE2 and its structural homolog TMEM27 (Collectrin). TMEM27 shares 47.8% identity with the non-catalytic extracellular, transmembrane and cytosolic domains of ACE2 and is located next to ACE2 on chromosome Xp22. Expression of ACE2 on the apical membrane of enterocytes as well as TMEM27 in kidney is necessary for the expression of these amino acid transporters in the intestine and kidney, respectively. TMEM27 has been shown to bind proteins involved in intracellular and membrane protein trafficking such Snapin and SNAP25 and plays a role in glucose-stimulated Insulin exocytosis in pancreas. Interestingly, ACE2 deficiency is also linked to defects in Insulin secretion (Niu et al., 2008).

Mutations in SLC6A20 have been associated with Hirschsprung's disease (Xie et al., 2019). The set of genes associated with Hirschsprung include GDNF, which is also linked to the CCHS (Section 3.5.4). Interestingly, CCHS is frequently complicated with neurocristopathies such as Hirschsprung (Sasaki et al., 2003). Interestingly, the molecular pathologies of these diseases (Diabetes, Hirschsprung, CCHS) all converge at the SNARE mediated exocytosis machinery. Mutations in SLC6A19 cause Hartnup disease which is characterized by neutral aminoaciduria. Under certain conditions, patients may also additionally develop symptoms resembling pellagra including photosensitive dermatitis, ataxia, and psychotic behavior, which are secondary to niacin deficiency due to inadequate intestinal absorption of tryptophan, the precursor for niacin synthesis (Hartnup Disease, 1998). Low levels of essential tryptophan may also lead to low levels in Serotonin, Melatonin and Niacin. Similar neuropsychiatric symptoms have been associated with COVID-19 (Ellul et al., 2020). Some COVID-19 patients develop acute Fanconi syndrome including aminoaciduria (Kormann et al., 2020), which resembles the Collectrin KO phenotype, characterized by a Fanconi-like polyuria, which has been linked to its role in SNARE complex formation (Chu and Le, 2014). Significant decreases in Tryptophan levels have been identified with COVID-19 patients (Thomas et al., 2020) and a clinical trial assessing the effect of Niacin on COVID-19 patients is currently underway (Wilkerson et al., 2020).

In summary, the potential COVID-19 severity associated gene, SLC6A20, and other amino acid transporters known to interact with ACE2 functionally converge with ACE2, B1R and mir200c signaling at the SNARE complex. Dysregulation of the SNARE machinery might play a role in different COVID-19 affected organ systems (gastrointestinal system, CNS) and diseases (see new onset Diabetes, silent hypoxemia/hypoventilation).

3.8 Diagnostic and therapeutic implications

Given the importance of the KKS as a key dysregulated pathway in COVID-19, its therapeutic modulation provides a potentially tractable mode-of-action to control the production of active kinins and their des-Arg derivatives. In hereditary angioedema, SERPING1 deficiency leads to accumulation of excess BK, which in turn over-activates B1R. Enzyme replacement therapy and kallikrein inhibition are approved, safe and efficacious treatments for Hereditary Angioedema (HAE). Repurposing these drugs to reestablish cellular control of the CAS might be suitable as host directed therapies for COVID-19 patients. The results of enzyme replacement therapy with C1 Inhibitor are inconclusive, with one study reporting a 2.5 fold reduction in intubation and death, while B2R inhibition had no significant impact (Table 1).

Modulation of ACE2 activity is another potential therapeutic strategy for host directed therapy. Diminazene Aceturate directly enhances ACE2 activity (Qi et al., 2013) and PPARγ activator Rosiglitazone increases ACE2 promoter activity (Scroggin et al., 2012). Activation of SIRT1 increases ACE2 expression, and pharmacological activation of SIRT1 has been shown to attenuate ARDS and lung fibrosis. SIRT1 activators have been linked to beneficial effects on COVID-19 severity and mortality. Melatonin stimulates the SIRT1 signaling pathway (Shah et al., 2017) and has been suggested as a potential COVID-19 treatment (Zhang R. et al., 2020) and a clinical trial assessing its efficacy as preventive drug is underway[218]. Beneficial effects of SIRT1 activator resveratrol are also being discussed (Filardo et al., 2020). Metformin has been demonstrated to exert its protective effect in oxidative injuries *via* enhancing SIRT1 and AMPK expression in human endothelial cells (Hung et al., 2016). Interestingly, Metformin treatment was associated with significantly decreased mortality in COVID-19 patients with Diabetes in a retrospective analysis (Luo et al., 2020). This potential effect of SIRT1 activation is consistent with its role in attenuation of airway inflammation, ARDS, ALI, fibrosis, edema and maintaining endothelial and epithelial barriers. Readouts of trials involving ACE2 modulating drugs are summarized in Table 1. Ongoing COVID-19 trials utilizing ACE2 modulators are listed in Supplementary Table S2 and summarized in Figure 3 of Supplementary Material S1.

According to the COVID-19 model B1R could be a preferred therapeutic target. B1R has been researched as a target for the past 10 years, mainly in the context of hyperalgesia and osteoarthritis. However, no clinical results have been published and most of the reported trials are inactive or have been suspended or discontinued (Supplementary Table S3). In patients hospitalized with severe COVID-19, the use of dexamethasone has been recommended as standard-of-care medication. Dexamethasone results in lower 28-day mortality among patients who received invasive mechanical ventilation or

oxygen (Horby et al., 2020). Interestingly, the induction of B1R is sensitive to treatment with dexamethasone (Phagoo et al., 2005). In the context of our disease model this activity may explain Dexamethasone's striking effects.

COVID-19 associated organ damage cannot be wholly explained by the virus' organ tropism and localized viral load. In COVID-19 associated kidney disease, for example, viral load is low and unevenly distributed (Puelles et al., 2020; Su et al., 2020), making it difficult to explain the extensive kidney damage seen in some patients (Wang et al., 2021). Currently, there are no diagnostic tools associated with such systemic effects. A systemic virus-independent mechanism requires system-wide distribution of a signal that bears the potential to induce a broad spectrum of pathophysiological dysregulation in disparate organs/tissues. A derailed pleiotropic signaling system such as the KKS/B1R signaling axis provides a likely candidate. In this context circulating microvesicles enriched in B1R and mir200c or circulating mir200c itself could serve as biomarker candidates. Indeed, serum, plasma or PBMC levels of miR200c has been identified as a diagnostic biomarker candidate in different COVID-19 related disease contexts, namely Kawasaki Disease (Zhang W. et al., 2017), Pneumonia (Liu et al., 2017), interstitial lung disease (Jiang et al., 2017) COPD (Cao et al., 2014) and fibrosis of multiple tissues (Yang et al., 2012; Ramachandran et al., 2013; Chen et al., 2017). Recently, it has also been shown that upregulated circulating miR-200c in plasma may increase the risk of severe disease in obese individuals (Papannarao et al., 2021).

Our COVID-19 knowledge model contains several drug targets with potential implications for host-directed therapies. These perspectives are summarized in Supplementary Tables S2, S3. Pharmaceutically tractable target structures include the KKS at multiple levels such as Kallikrein inhibitors, SerpinG1 enzyme replacement or B1R inhibitors. Interestingly, the induction of B1R is sensitive to treatment with dexamethasone. Modulation of ACE2 activity presents another potential angle of attack for host directed therapy either by direct activation (e.g., Diminazene Aceturate) or indirectly by induction, e.g., through SIRT-1 activators such as Melatonin, Resveratrol and Metformin or activators of PPARγ (Dambha-Miller et al., 2020). Readouts of trials utilizing drugs that potentially induce ACE2 expression (Dambha-Miller et al., 2020) are summarized in Table 1.

4 Discussion

Working in response to the onset of the COVID-19 pandemic, we developed an iterative, expert-driven strategy to compile published and emergent data into a comprehensive COVID-19 knowledge model. By integrating observed clinical phenotypes with core signaling mechanisms, this interactive resource provides a novel patient-level overview of COVID-19 molecular symptomatology that is intended to aid the rapid and

systematic elucidation of potential COVID-19 pathogenic mechanisms. In this work, we report on the insights and hypotheses gleaned through COVID-19 Explorer (Brock et al., 2022). We demonstrated how patient-level knowledge modeling of disease symptomatology provides an effective strategy to deconvolute the complexity of the COVID-19 systemic disease, covering perspectives ranging from the host factors involved in SARS-CoV-2 infection, to "a perfect storm" triggered by SARS-COV-2 induced acute hyper-inflammation to "accelerated aging" in subacute/chronic-COVID-19 syndrome *via* virus-independent engagement of pro-senescence pathways. Overall, our model suggests that viral perturbation of eight key mechanisms, alone or in combination, contributes to the pathogenesis of the primary COVID-19 phenotypes.

In terms of host factors responsible for SARS-CoV-2 entry, the COVID-19 knowledge model highlights the importance of ACE2, TMPRSS2, and components of the ISG response that are specifically dysregulated by SARS-CoV host interaction (ACE2, SERPING1). According to the model, these factors may converge on unifying pleiotropic signaling pathways comprising RAS and KKS as part of the CAS. The concurrent downregulation of ACE2 and SERPING1 may reciprocally amplify the deregulation of KKS, which may in turn lead to a cascading over-activation of downstream signaling, especially in acute COVID-19.

Beyond, the host factors responsible for infection, we also examined the mechanisms driving systemic disease. At the time the model was developed in early 2020, new and apparently unrelated clinical phenotypes were concurrently described, which were immediately linked to the core mechanisms identified by our model (e.g., barrier permeability or exocytosis). For instance, first reports of silent hypoxemia emerged in April 2020. The symptoms of silent hypoxemia were mapped against the symptomatology of heritable diseases, identifying CCHS as a phenotypically related disease. CCHS is caused by dysfunction of the exocytosis machinery in oxygen sensing cells, providing a direct link to our model. A similar link between clinical phenotypes and our model was established upon the first reports of endotheliitis (Varga et al., 2020), vasculitis and the role of micro-thrombotic events in severe disease (Table 1). Here too, direct links between the molecular etiology of the observed symptoms and our model could be drawn and strikingly, by mid-April, first cases of new onset KWD-like disease were reported in children with COVID-19, thereby also suggesting the predictive potential of the model. Indeed, for some clinical phenotypes predicted by our model, clinical observations in the on-going pandemic provided real-time validation of many hypotheses (see Supplementary Table S1 for summary).

The model clearly identifies eight key mechanisms that alone or in combination can contribute to the pathogenesis of observed COVID-19 phenotypes, revealing several functionally connected mechanisms across various organ-systems and identified numerous hypotheses for both, viral-dependent and -independent disease mechanisms, and associated pharmacologic targets that may warrant further evaluation. In this context, the model suggests that circulating microvesicles enriched in B1R and mir200c or circulating mir200c itself may serve as potential biomarkers. Interestingly, miR200c identified as a key player in our model, serves as diagnostic biomarker in KWD and MV-based intercellular communication, plays a key role in the molecular etiology of endotheliitis. Hence, and not surprisingly, cardiovascular complications have rapidly emerged as key-threats in COVID-19 (Supplementary Table S1). Next, besides pharmaceutically tractable target structures such as the KKS at multiple levels such as Kallikrein inhibitors, SerpinG1 enzyme replacement or B1R inhibitors, our model identified B1R as a potential therapeutic target.

From a clinical and translational perspective, the identification of discrete molecular players, their interactions, and their involvement in various signaling pathways, is central in the systematic understanding of COVID-19 disease course and outcome. Notably, the model also provides a mechanistic framework for new insights on potential risk-factors triggering COVID-19 as well as so-called "long-COVID" syndrome, occurring with multiple transient but also persistent symptoms, despite prior viral clearance. For instance, the accelerated aging phenomena may largely be explained by activated senescence pathways provided in the model. From a clinical viewpoint, the model underlines the increasing medical evidence that COVID-19 is a systemic disease, which makes it imperative to also treat it as such. Thus, this not only calls for more holistic therapeutic strategies, but also requires the development of comprehensive diagnostic algorithms and disease management guidelines to enable clinicians to tailor individual therapy plans for each patient (i.e., a personalized medicine approach), to accelerate its initiation and allow for rapid adjustments where necessary (e.g., management of COVID-19 induced ARDS).

While providing potential opportunities and hypotheses for drug repurposing, new therapeutic strategies and the immediate identification of potential safety concerns, our model may also enable a more rational development and prioritization of drug development, and clinical trial design. Notably, during the ongoing pandemic, numerous repurposed therapeutic regimens were initiated, but then revised repeatedly, with some ultimately discontinued due to safety issues and inefficiency/futility (e.g., Hydroxychloroquine). Hence, our model may also provide an evidence-based platform to assess repurposed drug candidates for their safety and overall clinical utility in COVID-19.

Numerous large-scale data and AI-initiatives have been launched to provide a testbed for pandemic forecasting (Sheridan, 2020). However, it should be emphasized that the majority provide only partial knowledge as opposed to a comprehensive understanding of COVID-19. Therefore, we believe that holistic clinico-molecular data mining platforms

may represent valuable tools to augment 1) the rapid and evidence-driven response to emergent global health emergencies; 2) the in-depth analysis of any complex disease; 3) the development of more personalized disease management guidelines; and d) the identification and development of new drug candidates and biomarkers. From this perspective, such technology platforms may be of critical interest, especially as a framework for global data- and knowledge-sharing. This is particularly relevant in the face of future pandemics and public health emergencies, where multiple stakeholders must rapidly identify and logically structure tomes of real-time information. Patient level disease models may thus prove critical in our future response to such challenges.

Although our COVID-19 model comes with certain limitations (Brock et al., 2022) it does suggest that the overall strategy of aggregating and connecting existing knowledge in a semi-supervised manner, results in a defendable and testable molecular model that may provide an effective means to expedite response to future public health emergencies. This is supported by the fact that several of our original hypotheses have in the meantime been validated through translational research endeavors. Moreover, the work clearly suggests the utility of digital health technologies in evolving a new generation of disease-specific expert-reviews, with broader knowledge coverage and hypothesis generating capacity than current peer-reviewed based approaches.

Beyond the global implications for understanding the molecular basis of COVID-19, our results also suggest that patient-level models may be adapted for evidence-based care. Using knowledge model frameworks such as the MH Corona Explorer, a deeper understanding of any disease can be generated, linking its molecular foundations with its real-world clinical course and outcomes. Applied to COVID-19, the aim is to enable innovative approaches for the diagnosis and therapy of patients and improved patient care. Innovative joint efforts will be crucial to implement systematic studies combining COVID-19 patient data and digital models to identify new drug targets and biomarkers. Findings from such projects may help define predictors for addressing the progression to a severe disease course or for the long-term illness caused by COVID-19. Such collaborations enable knowledge from clinical practice to be combined with the in-depth COVID-19 disease model and be made directly available to doctors for patient care early and record reliable risk assessments of those affected by COVID-19 in order to be able to use them predictively or prognostically. In the future, we hope to be able to preemptively respond to emergent pandemics by networking information on new virus variants and potentially change the course of the disease and its epidemiological consequences. Adaptable treatment strategies that emerge could serve as a key function in care and thus contribute to the management of future pandemics.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

SB conceived original idea of a host response-centric COVID-19 model, initiated and designed the study, and performed the analysis. ME and SH performed clinical-translational interpretation of the results, provided critical review and input to the manuscript. SB, ME and DJ wrote the initial draft. TS performed data management tasks, provided writing review and editing. DJ provided scientific support, writing review and editing.

Acknowledgments

Conflict of interest

Authors SB, DJ, TS, KH, AS, and FD were employed by Molecular Health GmbH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmmed.2022.1035290/full#supplementary-material

References

AlGhatrif, M., Cingolani, O., and Lakatta, E. G. (2020). The dilemma of coronavirus disease 2019, aging, and cardiovascular disease: Insights from cardiovascular aging Science. *JAMA Cardiol.* 5 (7), 747–748. doi:10.1001/jamacardio.2020.1329

Al Hariri, M., Zibara, K., Farhat, W., Hashem, Y., Soudani, N., Al Ibrahim, F., et al. (2016). Cigarette smoking-induced cardiac hypertrophy, vascular inflammation and injury are attenuated by antioxidant supplementation in an animal model. *Front. Pharmacol.* 7, 397. doi:10.3389/fphar.2016.00397

Al-Sadi, R., Rawat, M., and Ma, T. (2017). P-272 MMP-9 modulation of intestinal epithelial tight junction permeability: Role of micro-RNA regulation of occludin. *Inflamm. Bowel Dis.* 23, S88.

Anderson, J. M., and Van Itallie, C. M. (2009). Physiology and function of the tight junction. *Cold Spring Harb. Perspect. Biol.* 1, a002584. doi:10.1101/cshperspect.a002584

Anzalone, G., Arcoleo, G., Bucchieri, F., Montalbano, A. M., Marchese, R., Albano, G. D., et al. (2019). Cigarette smoke affects the onco-suppressor DAB2IP expression in bronchial epithelial cells of COPD patients. *Sci. Rep.* 9, 15682. doi:10.1038/s41598-019-52179-5

Armaiz-Pena, G. N., Allen, J. K., Cruz, A., Stone, R. L., Nick, A. M., Lin, Y. G., et al. (2013). Src activation by β-adrenoreceptors is a key switch for tumour metastasis. *Nat. Commun.* 4, 1403–1412. doi:10.1038/ncomms2413

Austinat, M., Braeuninger, S., Pesquero, J. B., Brede, M., Bader, M., Stoll, G., et al. (2009). Blockade of bradykinin receptor B1 but not bradykinin receptor B2 provides protection from cerebral infarction and brain edema. *Stroke* 40, 285–293. doi:10.1161/STROKEAHA.108.526673

Aztatzi-Aguilar, O. G., Uribe-Ramírez, M., Arias-Montaño, J. A., Barbier, O., and De Vizcaya-Ruiz, A. (2015). Acute and subchronic exposure to air particulate matter induces expression of angiotensin and bradykinin-related genes in the lungs and heart: Angiotensin-II type-I receptor as a molecular target of particulate matter exposure. *Part Fibre Toxicol.* 12, 17. doi:10.1186/s12989-015-0094-4

Banu, N., Panikar, S. S., Leal, L. R., and Leal, A. R. (2020). Protective role of ACE2 and its downregulation in SARS-CoV-2 infection leading to Macrophage Activation Syndrome: Therapeutic implications. *Life Sci.* 256, 117905. doi:10.1016/j.lfs.2020.117905

Barrenschee, M., BottnerM.Harde, J., Lange, C., CossaisF.EbsenM., et al. (2015). SNAP-25 is abundantly expressed in enteric neuronal networks and upregulated by the neurotrophic factor GDNF. *Histochem Cell Biol.* 143, 611–623. doi:10.1007/s00418-015-1310-x

Barrera, F. J., González-González, J. G., and Rodríguez-Gutiérrez, R. (2020). Gastrointestinal and liver involvement in patients with COVID-19. *Lancet Gastroenterol. Hepatol.* 5, 798–799. doi:10.1016/S2468-1253(20)30205-3

Batlle, D., Jose Soler, M., and Ye, M. (2010). ACE2 and diabetes: ACE of ACEs? *Diabetes* 59, 2994–2996. doi:10.2337/db10-1205

Belgardt, B. F., Ahmed, K., Spranger, M., Latreille, M., Denzler, R., Kondratiuk, N., et al. (2015). The microRNA-200 family regulates pancreatic beta cell survival in type 2 diabetes. *Nat. Med.* 21, 619–627. doi:10.1038/nm.3862

Belhadjer, Z., Meot, M., Bajolle, F., Khraiche, D., Legendre, A., Abakka, S., et al. (2020). Acute heart failure in multisystem inflammatory syndrome in children in the context of global SARS-CoV-2 pandemic. *Circulation* 142, 429–436. doi:10.1161/CIRCULATIONAHA.120.048360

Bhattacharyya, S., Fang, F., Tourtellotte, W., and Varga, J. (2013). Egr-1: New conductor for the tissue repair orchestra directs harmony (regeneration) or cacophony (fibrosis). *J. Pathol.* 229, 286–297. doi:10.1002/path.4131

Böckmann, S., and Paegelow, I. (2000). Kinins and kinin receptors: Importance for the activation of leukocytes. *J. Leukoc. Biol.* 68, 587–592. doi:10.1189/jlb.68.5.587

Bohnert, R., Vivas, S., and Jansen, G. (2017). Comprehensive benchmarking of SNV callers for highly admixed tumor data. *PLoS One* 12, e0186175. doi:10.1371/journal.pone.0186175

Boyle, P. A., Wilson, R. S., Yu, L., Barr, A. M., Honer, W. G., Schneider, J. A., et al. (2013). Much of late life cognitive decline is not due to common neurodegenerative pathologies. *Ann. Neurol.* 74, 478–489. doi:10.1002/ana.23964

Bramante, C., Ingraham, N. E., Murray, T. A., Marmor, S., Hovertsen, S., Gronski, J., et al. (2020). Observational study of Metformin and risk of mortality in patients hospitalized with covid-19. *medRxiv*, 20135095. 2020.2006.2019. doi:10.1101/2020.06.19.20135095

Bramante, C. T., Buse, J., Tamaritz, L., Palacio, A., Cohen, K., Vojta, D., et al. (2021). Outpatient metformin use is associated with reduced severity of COVID-19 disease in adults with overweight or obesity. *J. Med. Virol.* 93 (7), 4273–4279. doi:10.1002/jmv.26873

Brann, D. H., Tsukahara, T., Weinreb, C., Lipovsek, M., Van den Berge, K., Gong, B., et al. (2020). Non-neuronal expression of SARS-CoV-2 entry genes in the olfactory system suggests mechanisms underlying COVID-19-associated anosmia. *Sci. Adv.*, 6(31):eabc5801. doi:10.1126/sciadv.abc5801

Bridwell, R., Long, B., and Gottlieb, M. (2020). Neurologic complications of COVID-19. *Am. J. Emerg. Med.* 38, e3–e1549. doi:10.1016/j.ajem.2020.05.024

Brilha, S., Ong, C. W. M., Weksler, B., Romero, N., Couraud, P. O., and Friedland, J. S. (2017). Author Correction: Matrix metalloproteinase-9 activity and a downregulated Hedgehog pathway impair blood-brain barrier function in an *in vitro* model of CNS tuberculosis. *Sci. Rep.* 7, 13956. doi:10.1038/s41598-018-31948-8

Brock, S., Soldatos, T. G., Jackson, D. B., Diella, F., Hornischer, K., Schäefer, A., et al. (2022). The COVID-19 Explorer - an integrated, whole patient knowledge model of COVID-19 disease. *Front. Mol. Med.* 2, 1035215. doi:10.3389/fmmed.2022.1035215

Brüssow, H. (2020). The novel coronavirus - a snapshot of current knowledge. *Microb. Biotechnol.* 13, 607–612. doi:10.1111/1751-7915.13557

Bukowska, A., Spiller, L., Wolke, C., Lendeckel, U., Weinert, S., Hoffmann, J., et al. (2017). Protective regulation of the ACE2/ACE gene expression by estrogen in human atrial tissue from elderly men. *Exp. Biol. Med. (Maywood)* 242, 1412–1423. doi:10.1177/1535370217718808

Calixto, J. B., Medeiros, R., Fernandes, E. S., Ferreira, J., Cabrini, D. A., and Campos, M. M. (2004). Kinin B1 receptors: Key G-protein-coupled receptors and their role in inflammatory and painful processes. *Br. J. Pharmacol.* 143, 803–818. doi:10.1038/sj.bjp.0706012

Camdessanche, J. P., Morel, J., Pozzetto, B., Paul, S., Tholance, Y., and BotElho-NevErs, E. (2020). COVID-19 may induce Guillain-Barré syndrome. *Rev. Neurol. Paris.* 176, 516–518. doi:10.1016/j.neurol.2020.04.003

Cao, Q., Mani, R. S., Ateeq, B., Dhanasekaran, S. M., Asangani, I. A., Prensner, J. R., et al. (2011). Coordinated regulation of polycomb group complexes through microRNAs in cancer. *Cancer Cell* 20, 187–199. doi:10.1016/j.ccr.2011.06.016

Cao, Z., ZhaNg, N., Lou, T., Jin, Y., Wu, Y., Ye, Z., et al. (2014). microRNA-183 down-regulates the expression of BKCaβ1 protein that is related to the severity of chronic obstructive pulmonary disease. *Hippokratia* 18, 328–332.

Carlomosti, F., D'Agostino, M., Beji, S., Torcinaro, A., Rizzi, R., Zaccagnini, G., et al. (2017). Oxidative stress-induced miR-200c disrupts the regulatory loop among SIRT1, FOXO1, and eNOS. *Antioxid. Redox Signal* 27, 328–344. doi:10.1089/ars.2016.6643

Carsana, L., Sonzogni, A., Nasr, A., Rossi, R. S., Pellegrinelli, A., Zerbi, P., et al. (2020). Pulmonary post-mortem findings in a series of COVID-19 cases from northern Italy: A two-centre descriptive study. *Lancet Infect. Dis.* 20, 1135–1140. doi:10.1016/S1473-3099(20)30434-5

Catanzaro, O. L., Dziubecki, D., Obregon, P., Rodriguez, R. R., and Sirois, P. (2010). Antidiabetic efficacy of bradykinin antagonist R-954 on glucose tolerance test in diabetic type 1 mice. *Neuropeptides* 44, 187–189. doi:10.1016/j.npep.2009.12.010

Chee, Y. J., Ng, S. J. H., and Yeoh, E. (2020). Diabetic ketoacidosis precipitated by Covid-19 in a patient with newly diagnosed diabetes mellitus. *Diabetes Res. Clin. Pract.* 164, 108166. doi:10.1016/j.diabres.2020.108166

Chen, J., Cai, J., Du, C., Cao, Q., Li, M., and Liu, B. (2017). Recent advances in miR-200c and fibrosis in organs. *Zhong Nan Da Xue Xue Bao Yi Xue Ban.* 42, 226–232. doi:10.11817/j.issn.1672-7347.2017.02.018

Chen, J., Jiang, Q., Xia, X., Liu, K., Yu, Z., Tao, W., et al. (2020). Individual variation of the SARS-CoV-2 receptor ACE2 gene expression and regulation. *Aging Cell* 19, e13168. doi:10.1111/acel.13168

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, China: A descriptive study. *Lancet* 395, 507–513. doi:10.1016/S0140-6736(20)30211-7

Chen, X., Wang, Y., Xie, X., Chen, H., Zhu, Q., Ge, Z., et al. (2018). Heme oxygenase-1 reduces sepsis-induced endoplasmic reticulum stress and acute lung injury. *Mediat. Inflamm.* 2018, 9413876. doi:10.1155/2018/9413876

Cheng, X., Xin, S., Chen, Y., Li, L., Chen, W., Li, W., et al. (2021). Effects of metformin, insulin on COVID-19 patients with pre-existed type 2 diabetes: A multicentral retrospective study. *Life Sci.* 275, 119371. doi:10.1016/j.lfs.2021.119371

Chu, P. L., and Le, T. H. (2014). Role of collectrin, an ACE2 homologue, in blood pressure homeostasis. *Curr. Hypertens. Rep.* 16, 490. doi:10.1007/s11906-014-0490-4

Cipollaro, L., Giordano, L., Padulo, J., Oliva, F., and Maffulli, N. (2020). Musculoskeletal symptoms in SARS-CoV-2 (COVID-19) patients. *J. Orthop. Surg. Res.* 15, 178. doi:10.1186/s13018-020-01702-w

Clarke, N. E., Belyaev, N. D., Lambert, D. W., and Turner, A. J. (2014). Epigenetic regulation of angiotensin-converting enzyme 2 (ACE2) by SIRT1 under conditions of cell energy stress. *Clin. Sci. (Lond)* 126, 507–516. doi:10.1042/CS20130291

Cohen, J. B., Hanff, T. C., William, P., Sweitzer, N., Rosado-Santander, N. R., Medina, C., et al. (2021). Continuation versus discontinuation of renin-angiotensin system inhibitors in patients admitted to hospital with COVID-19: A prospective, randomised, open-label trial. *Lancet Respir. Med.* 9 (3), 275–284. doi:10.1016/S2213-2600(20)30558-0

Comunian, S., Dongo, D., Milani, C., and Palestini, P. (2020). Air pollution and covid-19: The role of particulate matter in the spread and increase of covid-19's morbidity and mortality. *Int. J. Environ. Res. Public Health* 17, 4487. doi:10.3390/ijerph17124487

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., et al. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Sci. (New York, N.Y.)* 261, 921–923. doi:10.1126/science.8346443

Corley, M. J. N., and Ndhlovu, L. C. (2020). DNA methylation analysis of the COVID-19 host cell receptor, angiotensin I converting enzyme 2 gene (ACE2) in the respiratory system reveal age and gender differences. *Preprints*.

Couture, R., Harrisson, M., Vianna, R. M., and Cloutier, F. (2001). Kinin receptors in pain and inflammation. *Eur. J. Pharmacol.* 429, 161–176. doi:10.1016/s0014-2999(01)01318-8

Couzin-Frankel, J. (2020). The mystery of the pandemic's 'happy hypoxia. *Sci. (New York, N.Y.)* 368, 455–456. doi:10.1126/science.368.6490.455

Crouse, A. B., Grimes, T., Li, P., Might, M., Ovalle, F., and Shalev, A. (2021). Metformin use is associated with reduced mortality in a diverse population with COVID-19 and diabetes. *Front. Endocrinol. (Lausanne)* 11, 600439. doi:10.3389/fendo.2020.600439

Cugno, M., Zanichelli, A., Foieni, F., Caccia, S., and Cicardi, M. (2009). C1-inhibitor deficiency and angioedema: Molecular mechanisms and clinical progress. *Trends Mol. Med.* 15, 69–78. doi:10.1016/j.molmed.2008.12.001

Cutting, G. R. (2015). Cystic fibrosis genetics: From molecular understanding to clinical application. *Nat. Rev. Genet.* 16, 45–56. doi:10.1038/nrg3849

Cyr, M., Lepage, Y., Blais, C., Gervais, N., Cugno, M., Rouleau, J. L., et al. (2001). Bradykinin and des-Arg(9)-bradykinin metabolic pathways and kinetics of activation of human plasma. *Am. J. Physiol. Heart Circ. Physiol.* 281, H275–H283. doi:10.1152/ajpheart.2001.281.1.H275

Dai, Y., Ngo, D., Forman, L. W., Qin, D. C., Jacob, J., and Faller, D. V. (2007). Sirtuin 1 is required for antagonist-induced transcriptional repression of androgen-responsive genes by the androgen receptor. *Mol. Endocrinol. Baltim. Md* 21, 1807–1821. doi:10.1210/me.2006-0467

Dambha-Miller, H., Albasri, A., Hodgson, S., Wilcox, C. R., Khan, S., Islam, N., et al. (2020). Currently prescribed drugs in the UK that could upregulate or downregulate ACE2 in COVID-19 disease: A systematic review. *BMJ Open* 10, e040644. doi:10.1136/bmjopen-2020-040644

Davey, A., McAuley, D. F., and O'Kane, C. M. (2011). Matrix metalloproteinases in acute lung injury: Mediators of injury and drivers of repair. *Eur. Respir. J.* 38, 959–970. doi:10.1183/09031936.00032111

Davis, A. E., 3rd, Lu, F., and Mejia, P. (2010). C1 inhibitor, a multi-functional serine protease inhibitor. *Thromb. Haemost.* 104, 886–893. doi:10.1160/TH10-01-0073

Davis, S., Rodger, J., StephAn, A., Hicks, A., Mallet, J., and Laroche, S. (1998). Increase in syntaxin 1B mRNA in hippocampal and cortical circuits during spatial learning reflects a mechanism of trans-synaptic plasticity involved in establishing a memory trace. *Learn Mem.* 5, 375–390. doi:10.1101/lm.5.4.375

Dawidowski, D., and Cafiso, D. S. (2016). Munc18-1 and the syntaxin-1 N terminus regulate open-closed States in a t-SNARE complex. *Structure* 24, 392–400. doi:10.1016/j.str.2016.01.005

De Seabra Rodrigues Dias, I. R., Cao, Z., and Kwok, H. F. (2022). Adamalysins in COVID-19 -Potential mechanisms behind exacerbating the disease. *Biomed. Pharmacother.* 150, 112970. doi:10.1016/j.biopha.2022.112970

de Vries, N. A., Hulsman, D., Akhtar, W., de Jong, J., Miles, D. C., Blom, M., et al. (2015). Prolonged Ezh2 depletion in glioblastoma causes a robust switch in cell fate resulting in tumor progression. *Cell Rep.* 10, 383–397. doi:10.1016/j.celrep.2014.12.028

Del Valle, D. M., Kim-Schulze, S., Hsin-Hui, H., Beckmann, N. D., Nirenberg, B., Wang, B., et al. (2020). An inflammatory cytokine signature helps predict COVID-19 severity and death. *medRxiv* 2020, 20115758. doi:10.1101/2020.05.28.20115758

Dominguez-Santas, M., Diaz-Guimaraens, B., Garcia Abellas, P., Moreno-GarCia Del Real, C., Burgos-Blasco, P., and SuArez-Valle, A. (2020). Cutaneous small-vessel vasculitis associated with novel 2019 coronavirus SARS-CoV-2 infection (COVID-19). *J. Eur. Acad. Dermatol Venereol.* 34, e536–e537. doi:10.1111/jdv.16663

Dozmorov, M. G. (2015). Polycomb repressive complex 2 epigenomic signature defines age-associated hypermethylation and gene expression changes. *Epigenetics* 10, 484–495. doi:10.1080/15592294.2015.1040619

Dray, A., and Perkins, M. (1993). Bradykinin and inflammatory pain. *Trends Neurosci.* 16, 99–104. doi:10.1016/0166-2236(93)90133-7

Dunsmore, S. E., and Rannels, D. E. (1996). Extracellular matrix biology in the lung. *Am. J. Physiol.* 270, L3–L27. doi:10.1152/ajplung.1996.270.1.L3

Dutra, R. C., Leite, D. F. P., Bento, A. F., Manjavachi, M. N., Patricio, E. S., Figueiredo, C. P., et al. (2011). The role of kinin receptors in preventing neuroinflammation and its clinical severity during experimental autoimmune encephalomyelitis in mice. *PLoS One* 6, e27875. doi:10.1371/journal.pone.0027875

El Akoum, S., Haddad, Y., and Couture, R. (2017). Impact of pioglitazone and bradykinin type 1 receptor antagonist on type 2 diabetes in high-fat diet-fed C57BL/6J mice. *Obes. Sci. Pract.* 3, 352–362. doi:10.1002/osp4.117

Elhelw, D. S., Riad, S. E., Shawer, H., El-Ekiaby, N., Salah, A., Zekri, A., et al. (2017). Ectopic delivery of miR-200c diminishes hepatitis C virus infectivity through transcriptional and translational repression of Occludin. *Arch. Virol.* 162, 3283–3291. doi:10.1007/s00705-017-3449-3

Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., et al. (2020). Genomewide association study of severe covid-19 with respiratory failure. *N. Engl. J. Med.* 383, 1522–1534. doi:10.1056/NEJMoa2020283

Ellul, M. A., Benjamin, L., Singh, B., Lant, S., Michael, B. D., Easton, A., et al. (2020). Neurological associations of COVID-19. *Lancet Neurol.* 19, 767–783. doi:10.1016/S1474-4422(20)30221-0

Entrenas Castillo, M., Entrenas Costa, L. M., Vaquero Barrios, J. M., Alcalá Díaz, J. F., López Miranda, J., Bouillon, R., et al. (2020). Effect of calcifediol treatment and best available therapy versus best available therapy on intensive care unit admission and mortality among patients hospitalized for COVID-19: A pilot randomized clinical study. *J. Steroid Biochem. Mol. Biol.* 203, 105751. doi:10.1016/j.jsbmb.2020.105751

Fan, R., Mao, S. Q., Gu, T. L., Zhong, F. D., Gong, M. L., Hao, L. M., et al. (2017). Preliminary analysis of the association between methylation of the ACE2 promoter and essential hypertension. *Mol. Med. Rep.* 15, 3905–3911. doi:10.3892/mmr.2017.6460

Faussner, A., Bathon, J. M., and Proud, D. (1999). Comparison of the responses of B1 and B2 kinin receptors to agonist stimulation. *Immunopharmacology* 45, 13–20. doi:10.1016/s0162-3109(99)00052-1

Filardo, S., Di Pietro, M., Mastromarino, P., and Sessa, R. (2020). Therapeutic potential of resveratrol against emerging respiratory viral infections. *Pharmacol. Ther.* 214, 107613. doi:10.1016/j.pharmthera.2020.107613

Fu, J., Kong, J., Wang, W., Wu, M., Yao, L., Wang, Z., et al. (2020). The clinical implication of dynamic neutrophil to lymphocyte ratio and D-dimer in COVID-19: A retrospective study in suzhou China. *Thromb. Res.* 192, 3–8. doi:10.1016/j.thromres.2020.05.006

Gaisano, H. Y. (2012). Deploying insulin granule-granule fusion to rescue deficient insulin secretion in diabetes. *Diabetologia* 55, 877–880. doi:10.1007/s00125-012-2483-7

Gattinoni, L., Chiumello, D., Caironi, P., Busana, M., Romitti, F., Brazzi, L., et al. (2020). COVID-19 pneumonia: Different respiratory treatments for different phenotypes? *Intensive Care Med.* 46, 1099–1102. doi:10.1007/s00134-020-06033-2

Ghany, R., Palacio, A., Dawkins, E., Chen, G., McCarter, D., Forbes, E., et al. (2021). Metformin is associated with lower hospitalizations, mortality and severe coronavirus infection among elderly medicare minority patients in 8 states in USA. *Diabetes Metab. Syndr.* 15 (2), 513–518. doi:10.1016/j.dsx.2021.02.022

Ghebrehiwet, B., Geisbrecht, B. V., Xu, X., Savitt, A. G., and Peerschke, E. I. B. (2019). The C1q Receptors: Focus on gC1qR/p33 (C1qBP, p32, HABP-1)(1). *Semin. Immunol.* 45, 101338. doi:10.1016/j.smim.2019.101338

Göbel, K., Pankratz, S., Schneider-Hohendorf, T., Bittner, S., Schuhmann, M. K., Langer, H. F., et al. (2011). Blockade of the kinin receptor B1 protects from autoimmune CNS disease by reducing leukocyte trafficking. *J. Autoimmun.* 36, 106–114. doi:10.1016/j.jaut.2010.11.004

Goyal, N., Skrdla, P., Schroyer, R., Kumar, S., Fernando, D., Oughton, A., et al. (2019). Clinical pharmacokinetics, safety, and tolerability of a novel, first-in-class TRPV4 ion channel inhibitor, GSK2798745, in healthy and heart failure Subjects. *Am. J. Cardiovasc Drugs* 19 (3), 335–342. doi:10.1007/s40256-018-00320-6

Greenlee, K. J., Werb, Z., and Kheradmand, F. (2007). Matrix metalloproteinases in lung: Multiple, multifarious, and multifaceted. *Physiol. Rev.* 87, 69–98. doi:10.1152/physrev.00022.2006

Grillo, F., Barisione, E., Ball, L., Mastracci, L., and Fiocca, R. (2020). Lung fibrosis: An undervalued finding in COVID-19 pathological series. *Lancet Infect. Dis.* 21, e72. doi:10.1016/S1473-3099(20)30582-X

Gupta, R., Radicioni, G., Abdelwahab, S., Dang, H., Carpenter, J., Chua, M., et al. (2019). Intercellular communication between airway epithelial cells is mediated by exosome-like vesicles. *Am. J. Respir. Cell Mol. Biol.* 60, 209–220. doi:10.1165/rcmb.2018-0156OC

Haddad, Y., and Couture, R. (2017). Kininase 1 as a preclinical therapeutic target for kinin B(1) receptor in insulin resistance. *Front. Pharmacol.* 8, 509. doi:10.3389/fphar.2017.00509

Haga, S., Tsuchiya, H., Hirai, T., Hamano, T., Mimori, A., and Ishizaka, Y. (2015). A novel ACE2 activator reduces monocrotaline-induced pulmonary hypertension by suppressing the JAK/STAT and TGF-β cascades with restored caveolin-1 expression. *Exp. Lung Res.* 41, 21–31. doi:10.3109/01902148.2014.959141

Hamming, I., Timens, W., Bulthuis, M. L. C., Lely, A. T., Navis, G. J., and van Goor, H. (2004). Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J. Pathol.* 203, 631–637. doi:10.1002/path.1570

Hampshire, A., Trender, W., Chamberlain, S. R., Jolly, A., Grant, J. E., Patrick, F., et al. (2020). Cognitive deficits in people who have recovered from COVID-19 relative to controls: An N=84,285 online study. *medRxiv*, 20215863. 2020.2010.2020. doi:10.1101/2020.10.20.20215863

Han, X., and Sun, Z. (2020). Epigenetic regulation of KL (klotho) via H3K27me3 (histone 3 lysine [K] 27 trimethylation) in renal tubule cells. *Hypertension* 75, 1233–1241. doi:10.1161/HYPERTENSIONAHA.120.14642

Hariyanto, T. I., and Kurniawan, A. (2020). Metformin use is associated with reduced mortality rate from coronavirus disease 2019 (COVID-19) infection. *Obes. Med.* 19, 100290. doi:10.1016/j.obmed.2020.100290

Hartnup Disease (1998). *Encyclopedia of human nutrition (second edition)*. Available at: https://www.sciencedirect.com/topics/medicine-and-dentistry/hartnup-disease.

He, Z., Zhang, S., Ma, D., Fang, Q., Yang, L., Shen, S., et al. (2019). HO-1 promotes resistance to an EZH2 inhibitor through the pRB-E2F pathway: Correlation with the progression of myelodysplastic syndrome into acute myeloid leukemia. *J. Transl. Med.* 17, 366. doi:10.1186/s12967-019-2115-9

Hernández Prada, J. A., Ferreira, A. J., Katovich, M. J., Shenoy, V., Qi, Y., Santos, R. A. S., et al. (2008). Structure-based identification of small-molecule angiotensin-converting enzyme 2 activators as novel antihypertensive agents. *Hypertension* 51, 1312–1317. doi:10.1161/HYPERTENSIONAHA.107.108944

Herrero, R., Prados, L., Ferruelo, A., Puig, F., Pandolfi, R., Guillamat-Prats, R., et al. (2019). Fas activation alters tight junction proteins in acute lung injury. *Thorax* 74, 69–82. doi:10.1136/thoraxjnl-2018-211535

Hill, N., Michell, D. L., RaMirez-Solano, M., Sheng, Q., Pusey, C., Vickers, K. C., et al. (2020). Glomerular endothelial derived vesicles mediate podocyte dysfunction: A potential role for miRNA. *PLoS One* 15, e0224852. doi:10.1371/journal.pone.0224852

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271e278–280. doi:10.1016/j.cell.2020.02.052

Holman, N., Knighton, P., Kar, P., O'Keefe, J., Curley, M., Weaver, A., et al. (2020). Risk factors for COVID-19-related mortality in people with type 1 and type 2 diabetes in england: A population-based cohort study. *Lancet Diabetes Endocrinol.* 8, 823–833. doi:10.1016/S2213-8587(20)30271-0

Honer, W. G., Barr, A. M., Sawada, K., Thornton, A. E., Morris, M. C., Leurgans, S. E., et al. (2012). Cognitive reserve, presynaptic proteins and dementia in the elderly. *Transl. Psychiatry* 2, e114. doi:10.1038/tp.2012.38

Hong, J. (2012). *Expression of tight junction protein Occludin and ZO-1 in lungs of COPD rats*. Journal of Kunming Medical University. https://www.semanticscholar.org/paper/Expression-of-Tight-Junction-Protein-Occludin-and-Hong/bdb2a6eb336030da096fb7dfdefeca5e32feacf1

Horby, P., Lim, W. S., Emberson, J. R., Mafham, M., Bell, J. L., Linsell, L., et al. (2020). Dexamethasone in hospitalized patients with covid-19 - preliminary report. *N. Engl. J. Med.* doi:10.1056/NEJMoa2021436

Hsu, A. T., Barrett, C. D., DeBusk, G. M., Ellson, C. D., Gautam, S., Talmor, D. S., et al. (2015). Kinetics and role of plasma matrix metalloproteinase-9 expression in acute lung injury and the acute respiratory distress syndrome. *Shock* 44, 128–136. doi:10.1097/SHK.0000000000000386

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497–506. doi:10.1016/S0140-6736(20)30183-5

Hung, C. H., Chan, S. H., Chu, P. M., Lin, H. C., and Tsai, K. L. (2016). Metformin regulates oxLDL-facilitated endothelial dysfunction by modulation of SIRT1 through repressing LOX-1-modulated oxidative signaling. *Oncotarget* 7, 10773–10787. doi:10.18632/oncotarget.7387

Hwang, J. W., Yao, H., Caito, S., Sundar, I. K., and Rahman, I. (2013). Redox regulation of SIRT1 in inflammation and cellular senescence. *Free Radic. Biol. Med.* 61, 95–110. doi:10.1016/j.freeradbiomed.2013.03.015

Imai, Y., Kuba, K., and Penninger, J. M. (2008). The discovery of angiotensin-converting enzyme 2 and its role in acute lung injury in mice. *Exp. Physiol.* 93, 543–548. doi:10.1113/expphysiol.2007.040048

Imai, Y., Kuba, K., Rao, S., Huan, Y., Guo, F., Guan, B., et al. (2005). Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* 436, 112–116. doi:10.1038/nature03712

Janssen, W. J., and Nozik-Grayck, E. (2017). Power of place: Intravascular superoxide dismutase for prevention of acute respiratory distress syndrome. *Am. J. Respir. Cell Mol. Biol.* 56, 147–149. doi:10.1165/rcmb.2016-0407ED

Jena, B. P. (2014). Porosome in cystic fibrosis. *Discov. (Craiova)* 2, e24. doi:10.15190/d.2014.16

Jia, H. P., Look, D. C., Shi, L., Hickey, M., Pewe, L., Netland, J., et al. (2005). ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. *J. Virol.* 79, 14614–14621. doi:10.1128/JVI.79.23.14614-14621.2005

Jia, H. (2016). Pulmonary angiotensin-converting enzyme 2 (ACE2) and inflammatory lung disease. *Shock* 46, 239–248. doi:10.1097/SHK.0000000000000633

Jiang, Z., Tao, J. H., Zuo, T., Li, X. M., Wang, G. S., Fang, X., et al. (2017). The correlation between miR-200c and the severity of interstitial lung disease associated with different connective tissue diseases. *Scand. J. Rheumatol.* 46, 122–129. doi:10.3109/03009742.2016.1167950

Jin, J., Cheng, Y., Zhang, Y., Wood, W., Peng, Q., Hutchison, E., et al. (2012). Interrogation of brain miRNA and mRNA expression profiles reveals a molecular regulatory network that is perturbed by mutant huntingtin. *J. Neurochem.* 123, 477–490. doi:10.1111/j.1471-4159.2012.07925.x

Jones, V. G., Mills, M., Suarez, D., Hogan, C. A., Yeh, D., Segal, J. B., et al. (2020). COVID-19 and Kawasaki disease: Novel virus and novel case. *Hosp. Pediatr.* 10, 537–540. doi:10.1542/hpeds.2020-0123

Joshi, S., Gomez, S., Duran-Mendez, M., Quiroz-Olvera, J., Garcia, C., and Jarajapu, Y. P. (2019). Aging healthy, or with diabetes, is associated with ACE2/ACE imbalance in the hematopoietic stem progenitor cells. *FASEB J.* 33, 514517–517514. doi:10.1096/fasebj.2019.33.1_supplement.514.7

Joyce-Brady, M. F., and Tuder, R. M. (2011). Just in the "bik" of time. *Am. J. Respir. Crit. Care Med.* 183, 1447–1448. doi:10.1164/rccm.201103-0566ED

Kahn, R., Mossberg, M., Stahl, A. L., Johansson, K., Lopatko Lindman, I., Heijl, C., et al. (2017). Microvesicle transfer of kinin B1-receptors is a novel inflammatory mechanism in vasculitis. *Kidney Int.* 91, 96–105. doi:10.1016/j.kint.2016.09.023

Kang, Y., Chen, T., Mui, D., Ferrari, V., Jagasia, D., Scherrer-Crosbie, M., et al. (2020). Cardiovascular manifestations and treatment considerations in COVID-19. *Heart* 106, 1132–1141. doi:10.1136/heartjnl-2020-317056

Kassiri, Z., Zhong, J., Guo, D., Basu, R., Wang, X., Liu, P. P., et al. (2009). Loss of angiotensin-converting enzyme 2 accelerates maladaptive left ventricular remodeling in response to myocardial infarction. *Circ. Heart Fail* 2, 446–455. doi:10.1161/CIRCHEARTFAILURE.108.840124

Kellner, M., Noonepalle, S., Lu, Q., Srivastava, A., Zemskov, E., and Black, S. M. (2017). ROS signaling in the pathogenesis of acute lung injury (ALI) and acute respiratory distress syndrome (ARDS). *Adv. Exp. Med. Biol.* 967, 105–137. doi:10.1007/978-3-319-63245-2_8

Kenne, E., Rasmuson, J., Renne, T., Vieira, M. L., Muller-Esterl, W., Herwald, H., et al. (2019). Neutrophils engage the kallikrein-kinin system to open up the endothelial barrier in acute inflammation. *FASEB J. official Publ. Fed. Am. Soc. Exp. Biol.* 33, 2599–2609. doi:10.1096/fj.201801329R

Kim, J., Lee, Y., Lu, X., Song, B., Fong, K. W., Cao, Q., et al. (2018). Polycomb- and methylation-independent roles of EZH2 as a transcription activator. *Cell Rep.* 25, 28082808–28082820. doi:10.1016/j.celrep.2018.11.035

Kling, K. M., Lopez-Rodriguez, E., Pfarrer, C., Mühlfeld, C., and Brandenberger, C. (2017). Aging exacerbates acute lung injury-induced changes of the air-blood barrier, lung function, and inflammation in the mouse. *Am. J. Physiol. Lung Cell Mol. Physiol.* 312, L12. doi:10.1152/ajplung.00347.2016

Koerner, P., Hesslinger, C., Schaefermeyer, A., Prinz, C., and Gratzl, M. (2004). Evidence for histamine as a transmitter in rat carotid body sensor cells. *J. Neurochem.* 91, 493–500. doi:10.1111/j.1471-4159.2004.02740.x

Kormann, R., Jacquot, A., Alla, A., Corbel, A., Koszutski, M., Voirin, P., et al. (2020). Coronavirus disease 2019: Acute Fanconi syndrome precedes acute kidney injury. *Clin. Kidney J.* 13, 362–370. doi:10.1093/ckj/sfaa109

Koroleva, E. P., Fu, Y. X., and Tumanov, A. V. (2018). Lymphotoxin in physiology of lymphoid tissues - implication for antiviral defense. *Cytokine* 101, 39–47. doi:10.1016/j.cyto.2016.08.018

Kowluru, R. A., Santos, J. M., and Zhong, Q. Sirt1 (2014). Sirt1, a negative regulator of matrix metalloproteinase-9 in diabetic retinopathy. *Invest Ophthalmol. Vis. Sci.* 55, 5653–5660. doi:10.1167/iovs.14-14383

Kron, K. J., Murison, A., Zhou, S., Huang, V., Yamaguchi, T. N., Shiah, Y. J., et al. (2017). TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. *Nat. Genet.* 49, 1336–1345. doi:10.1038/ng.3930

Kuba, K., Imai, Y., and Penninger, J. M. (2006). Angiotensin-converting enzyme 2 in lung diseases. *Curr. Opin. Pharmacol.* 6, 271–276. doi:10.1016/j.coph.2006.03.001

Kuba, K., Imai, Y., Rao, S., Gao, H., Guo, F., Guan, B., et al. (2005). A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat. Med.* 11, 875–879. doi:10.1038/nm1267

Kumari, K., Das, B., Adhya, A., Chaudhary, S., Senapati, S., and Mishra, S. K. (2018). Nicotine associated breast cancer in smokers is mediated through high level of EZH2 expression which can be reversed by methyltransferase inhibitor DZNepA. *Cell death Dis.* 9, 152. doi:10.1038/s41419-017-0224-z

Kuo, C. L., Pilling, L. C., Atkins, J. L., Masoli, J. A. H., Delgado, J., Kuchel, G. A., et al. (2020). APOE e4 Genotype Predicts Severe COVID-19 in the UK Biobank Community Cohort. *J. Gerontol. A Biol. Sci. Med. Sci.* 75, 2231–2232. doi:10.1093/gerona/glaa131

Lai, W.-T., Huang, Y.-H., Lo, M.-H., and Kuo, H.-C. (2020). Tight junction protein ZO-1 in Kawasaki disease. *BMC Pediatr.* 21 (1), 157. doi:10.1186/s12887-021-02622-2

Lalau, J. D., Al-Salameh, A., Hadjadj, S., Goronflot, T., Wiernsperger, N., Pichelin, M., et al. (2021). Metformin use is associated with a reduced risk of mortality in patients with diabetes hospitalised for COVID-19. *Diabetes Metab.* 47 (5), 101216. doi:10.1016/j.diabet.2020.101216

Lally, M. A., Tsoukas, P., Halladay, C. W., O'Neill, E., Gravenstein, S., and Rudolph, J. L. (2021). Metformin is associated with decreased 30-day mortality among nursing home residents infected with SARS-CoV2. *J. Am. Med. Dir. Assoc.* 22 (1), 193–198. doi:10.1016/j.jamda.2020.10.031

Lamb, M. E., Zhang, C., Shea, T., Kyle, D. J., and Leeb-Lundberg, L. M. (2002). Human B1 and B2 bradykinin receptors and their agonists target caveolae-related lipid rafts to different degrees in HEK293 cells. *Biochemistry* 41, 14340–14347. doi:10.1021/bi020231d

Langhauser, F., Gob, E., Kraft, P., Geis, C., Schmitt, J., Brede, M., et al. (2012). Kininogen deficiency protects from ischemic neurodegeneration in mice by reducing thrombosis, blood-brain barrier damage, and inflammation. *Blood* 120, 4082–4092. doi:10.1182/blood-2012-06-440057

Lee, Y. I., Kim, Y. G., Pyeon, H. J., Ahn, J. C., Logan, S., Orock, A., et al. (2019). Dysregulation of the SNARE-binding protein munc18-1 impairs BDNF secretion and synaptic neurotransmission: A novel interventional target to protect the aging brain. *Geroscience* 41, 109–123. doi:10.1007/s11357-019-00067-1

Leeb-Lundberg, L. M., Marceau, F., Müller-Esterl, W., Pettibone, D. J., and Zuraw, B. L. (2005). International union of pharmacology. XLV. Classification of the kinin receptor family: From molecular mechanisms to pathophysiological consequences. *Pharmacol. Rev.* 57, 27–77. doi:10.1124/pr.57.1.2

Levi, M., Thachil, J., Iba, T., and Levy, J. H. (2020). Coagulation abnormalities and thrombosis in patients with COVID-19. *Lancet Haematol.* 7, e438–e440. doi:10.1016/S2352-3026(20)30145-9

Li, C., Lasse, S., Lee, P., Nakasaki, M., Chen, S. W., Yamasaki, K., et al. (2010). Development of atopic dermatitis-like skin disease from the chronic loss of epidermal caspase-8. *Proc. Natl. Acad. Sci. U. S. A.* 107, 22249–22254. doi:10.1073/pnas.1009751108

Li, G., Xu, Y. l., Ling, F., Liu, A. j., Wang, D., Wang, Q., et al. (2012). Angiotensin-converting enzyme 2 activation protects against pulmonary arterial hypertension through improving early endothelial function and mediating cytokines levels. *Chin. Med. J. Engl.* 125, 1381–1388.

Li, J., Wang, X., Chen, J., Zuo, X., Zhang, H., and Deng, A. (2020a). COVID-19 infection may cause ketosis and ketoacidosis. *Diabetes Obes. Metab.* 22, 1935–1941. doi:10.1111/dom.14057

Li, J., Wei, Q., Li, W. X., McCowen, K. C., Xiong, W., Liu, J., et al. (2020b). Metformin use in diabetes prior to hospitalization: Effects on mortality in covid-19. *Endocr. Pract.* 26 (10), 1166–1172. doi:10.4158/EP-2020-0466

Li, Y., Li, H., and Zhou, L. (2020). EZH2-mediated H3K27me3 inhibits ACE2 expression. *Biochem. Biophys. Res. Commun.* 526, 947–952. doi:10.1016/j.bbrc.2020.04.010

Li, S. W., Wang, C. Y., Jou, Y. J., Yang, T. C., Huang, S. H., Wan, L., et al. (2016). SARS coronavirus papain-like protease induces Egr-1-dependent up-regulation of TGF-β1 via ROS/p38 MAPK/STAT3 pathway. *Sci. Rep.* 6, 25754. doi:10.1038/srep25754

Liang, T., Qin, T., Xie, L., Dolai, S., Zhu, D., Prentice, K. J., et al. (2017). New roles of syntaxin-1A in insulin granule exocytosis and replenishment. *J. Biol. Chem.* 292, 2203–2216. doi:10.1074/jbc.M116.769885

Lin, J. C., Talbot, S., Lahjouji, K., Roy, J. P., Senecal, J., Couture, R., et al. (2010). Mechanism of cigarette smoke-induced kinin B(1) receptor expression in rat airways. *Peptides* 31, 1940–1945. doi:10.1016/j.peptides.2010.07.008

Liu, C., von Brunn, A., and Zhu, D. (2020). Cyclophilin A and CD147: Novel therapeutic targets for the treatment of COVID-19. *Med. Drug Discov.* 7, 100056. doi:10.1016/j.medidd.2020.100056

Liu, J., Li, X., Lu, Q., Ren, D., Sun, X., Rousselle, T., et al. (2019). Ampk: A balancer of the renin-angiotensin system. *Biosci. Rep.* 39, BSR20181994. doi:10.1042/BSR20181994

Liu, Y., Mu, S., Li, X., Liang, Y., Wang, L., and Ma, X. (2019). Unfractionated heparin alleviates sepsis-induced acute lung injury by protecting tight junctions. *J. Surg. Res.* 238, 175–185. doi:10.1016/j.jss.2019.01.020

Liu, M., Gu, C., and Wang, Y. (2014). Upregulation of the tight junction protein occludin: Effects on ventilation-induced lung injury and mechanisms of action. *BMC Pulm. Med.* 14, 94. doi:10.1186/1471-2466-14-94

Liu, Q., Du, J., Yu, X., Xu, J., Huang, F., Li, X., et al. (2017). miRNA-200c-3p is crucial in acute respiratory distress syndrome. *Cell Discov.* 3, 17021. doi:10.1038/celldisc.2017.21

Lopatko Fagerström, I., Stahl, A. L., Mossberg, M., Tati, R., Kristoffersson, A. C., Kahn, R., et al. (2019). Blockade of the kallikrein-kinin system reduces endothelial complement activation in vascular inflammation. *EBioMedicine* 47, 319–328. doi:10.1016/j.ebiom.2019.08.020

López-Barneo, J., Ortega-Sáenz, P., Pardal, R., Pascual, A., and Piruat, J. I. (2008). Carotid body oxygen sensing. *Eur. Respir. J.* 32, 1386–1398. doi:10.1183/09031936.00056408

Lopes, R. D., Macedo, A. V. S., de Barros E Silva, P. G. M., Moll-Bernardes, R. J., Dos Santos, T. M., Mazza, L., et al. (2021). Effect of discontinuing vs continuing angiotensin-converting enzyme inhibitors and angiotensin II receptor blockers on days alive and out of the hospital in patients admitted with COVID-19: A randomized clinical trial. *JAMA* 325 (3), 254–264. doi:10.1001/jama.2020.25864

Lu, L., Li, L., Lu, X., Wu, X. s., Liu, D. p., and Liang, C. c. (2011). Inhibition of SIRT1 increases EZH2 protein level and enhances the repression of EZH2 on target gene expression. *Chin. Med. Sci. J.* 26, 77–84. doi:10.1016/s1001-9294(11)60024-2

Lu, Y., Liu, D. X., and Tam, J. P. (2008). Lipid rafts are involved in SARS-CoV entry into Vero E6 cells. *Biochem. Biophys. Res. Commun.* 369, 344–349. doi:10.1016/j.bbrc.2008.02.023

Lucas, J. M., Heinlein, C., Kim, T., Hernandez, S. A., Malik, M. S., True, L. D., et al. (2014). The androgen-regulated protease TMPRSS2 activates a proteolytic cascade involving components of the tumor microenvironment and promotes prostate cancer metastasis. *Cancer Discov.* 4, 1310–1325. doi:10.1158/2159-8290.CD-13-1010

Lukito, A. A., Pranata, R., Henrina, J., Lim, M. A., Lawrensia, S., and Suastika, K. (2020). The effect of metformin consumption on mortality in hospitalized COVID-19 patients: A systematic review and meta-analysis. *Diabetes Metab. Syndr.* 14 (6), 2177–2183. doi:10.1016/j.dsx.2020.11.006

Luo, P., Qiu, L., Liu, Y., Liu, X. L., Zheng, J. L., Xue, H. Y., et al. (2020). Metformin treatment was associated with decreased mortality in COVID-19 patients with diabetes in a retrospective analysis. *Am. J. Trop. Med. Hyg.* 103, 69–72. doi:10.4269/ajtmh.20-0375

Ma, Y., Xu, C., Wang, W., Sun, L., Yang, S., Lu, D., et al. (2014). Role of SIRT1 in the protection of intestinal epithelial barrier under hypoxia and its mechanism. *Zhonghua Wei Chang. Wai Ke Za Zhi* 17, 602–606.

Mao, L., Jin, H., Wang, M., Hu, Y., Chen, S., He, Q., et al. (2020). Neurologic manifestations of hospitalized patients with coronavirus disease 2019 in wuhan, China. *JAMA Neurol.* 77, 683–690. doi:10.1001/jamaneurol.2020.1127

Mao, R., Qiu, Y., He, J. S., Tan, J. Y., Li, X. H., Liang, J., et al. (2020). Manifestations and prognosis of gastrointestinal and liver involvement in patients with COVID-19: A systematic review and meta-analysis. *Lancet Gastroenterol. Hepatol.* 5, 667–678. doi:10.1016/S2468-1253(20)30126-6

Mao, Y., Lin, W., Wen, J., and Chen, G. (2020). Clinical and pathological characteristics of 2019 novel coronavirus disease (COVID-19): A systematic reviews. *medRxiv*, 20025601. 2020.2002.2020. doi:10.1101/2020.02.20.20025601

Marceau, F., Lussier, A., Regoli, D., and Giroud, J. P. (1983). Pharmacology of kinins: Their relevance to tissue injury and inflammation. *Gen. Pharmacol.* 14, 209–229. doi:10.1016/0306-3623(83)90001-0

Marceau, F., Sabourin, T., Houle, S., Fortin, J. P., Petitclerc, E., Molinaro, G., et al. (2002). Kinin receptors: Functional aspects. *Int. Immunopharmacol.* 2, 1729–1739. doi:10.1016/s1567-5769(02)00189-3

Marcucci, F., Zou, D. J., and Firestein, S. (2009). Sequential onset of presynaptic molecules during olfactory sensory neuron maturation. *J. Comp. Neurol.* 516, 187–198. doi:10.1002/cne.22094

Mariani, S., di Giorgio, M. R., Martini, P., Persichetti, A., Barbaro, G., Basciani, S., et al. (2018). Inverse association of circulating SIRT1 and adiposity: A study on underweight, normal weight, and obese patients. *Front. Endocrinol.* 9, 449. doi:10.3389/fendo.2018.00449

Martínez-Alemán, S. R., Campos-Garcia, L., Palma-Nicolas, J. P., Hernandez-Bello, R., Gonzalez, G. M., and Sanchez-Gonzalez, A. (2017). Understanding the entanglement: Neutrophil extracellular traps (NETs) in cystic fibrosis. *Front. Cell Infect. Microbiol.* 7, 104. doi:10.3389/fcimb.2017.00104

Mast, A. E., Wolberg, A. S., Gailani, D., Garvin, M. R., Alvarez, C., Miller, J. I., et al. (2021). Response to comment on 'SARS-CoV-2 suppresses anticoagulant and fibrinolytic gene expression in the lung'. *Elife* 10, e74951. doi:10.7554/eLife.74951

Matus, C. E., Ehrenfeld, P., Pavicic, F., Gonzalez, C. B., Concha, M., Bhoola, K. D., et al. (2016). Activation of the human keratinocyte B1 bradykinin receptor induces expression and secretion of metalloproteases 2 and 9 by transactivation of epidermal growth factor receptor. *Exp. Dermatol* 25, 694–700. doi:10.1111/exd.13038

McLean, P. G., Perretti, M., and Ahluwalia, A. (2000). Kinin B(1) receptors and the cardiovascular system: Regulation of expression and function. *Cardiovasc Res.* 48, 194–210. doi:10.1016/s0008-6363(00)00184-x

Menachery, V. D., Eisfeld, A. J., Schafer, A., Josset, L., Sims, A. C., Proll, S., et al. (2014). Pathogenic influenza viruses and coronaviruses utilize similar and contrasting approaches to control interferon-stimulated gene responses. *mBio* 5, e01174–e01114. doi:10.1128/mBio.01174-14

Menter, T., Haslbauer, J. D., Nienhold, R., Savic, S., Hopfer, H., Deigendesch, N., et al. (2020). Postmortem examination of COVID-19 patients reveals diffuse alveolar damage with severe capillary congestion and variegated findings in lungs and other organs suggesting vascular dysfunction. *Histopathology* 77, 198–209. doi:10.1111/his.14134

Meizlish, M. L., Goshua, G., Liu, Y., Fine, R., Amin, K., Chang, E., et al. (2021). Intermediate-dose anticoagulation, aspirin, and in-hospital mortality in COVID-19: A propensity score-matched analysis. *Am. J. Hematol.* 96 (4), 471–479. doi:10.1002/ajh.26102

Merrill, J. T., Erkan, D., Winakur, J., and James, J. A. (2020). Emerging evidence of a COVID-19 thrombotic syndrome has treatment implications. *Nat. Rev. Rheumatol.* 16, 581–589. doi:10.1038/s41584-020-0474-5

Miller, J., Bruen, C., Schnaus, M., Zhang, J., Ali, S., Lind, A., et al. (2020). Auxora versus standard of care for the treatment of severe or critical COVID-19 pneumonia: Results from a randomized controlled trial. *Crit. Care* 24 (1), 502. doi:10.1186/s13054-020-03220-x

Milsted, A., Underwood, A. C., Dunmire, J., DelPuerto, H. L., Martins, A. S., Ely, D. L., et al. (2010). Regulation of multiple renin-angiotensin system genes by Sry. *J. Hypertens.* 28, 59–64. doi:10.1097/HJH.0b013e328332b88d

Montagne, A., Nation, D. A., Sagare, A. P., Barisano, G., Sweeney, M. D., Chakhoyan, A., et al. (2020). APOE4 leads to blood-brain barrier dysfunction predicting cognitive decline. *Nature* 581, 71–76. doi:10.1038/s41586-020-2247-3

Morais, R. L., Silva, E. D., Sales, V. M., Filippelli-Silva, R., Mori, M. A., Bader, M., et al. (2015). Kinin B1 and B2 receptor deficiency protects against obesity induced by a high-fat diet and improves glucose tolerance in mice. *Diabetes Metab. Syndr. Obes.* 8, 399–407. doi:10.2147/DMSO.S87635

Moran, C. S., Biros, E., Krishna, S. M., Wang, Y., Tikellis, C., Morton, S. K., et al. (2017). Resveratrol inhibits growth of experimental abdominal aortic aneurysm associated with upregulation of angiotensin-converting enzyme 2. *Arterioscler. Thromb. Vasc. Biol.* 37, 2195–2203. doi:10.1161/ATVBAHA.117.310129

Moreira, A. (2020). Kawasaki disease linked to COVID-19 in children. *Nat. Rev. Immunol.* 20, 407. doi:10.1038/s41577-020-0350-1

Moreno-Manzano, V., Mampaso, F., Sepulveda-Munoz, J. C., Alique, M., Chen, S., Ziyadeh, F. N., et al. (2003). Retinoids as a potential treatment for experimental puromycin-induced nephrosis. *Br. J. Pharmacol.* 139, 823–831. doi:10.1038/sj.bjp.0705311

Mori, M. A., Sales, V. M., Motta, F. L., Fonseca, R. G., Alenina, N., Guadagnini, D., et al. (2012). Kinin B1 receptor in adipocytes regulates glucose tolerance and predisposition to obesity. *PLoS One* 7, e44782. doi:10.1371/journal.pone.0044782

Mossberg, M., Stahl, A. L., Kahn, R., Kristoffersson, A. C., Tati, R., Heijl, C., et al. (2017). C1-Inhibitor decreases the release of vasculitis-like chemotactic endothelial microvesicles. *J. Am. Soc. Nephrol.* 28, 2472–2481. doi:10.1681/ASN.2016060637

Mugisho, O. O., Robilliard, L. D., Nicholson, L. F. B., Graham, E. S., and O'Carroll, S. J. (2019). Bradykinin receptor-1 activation induces inflammation and increases the permeability of human brain microvascular endothelial cells. *Cell Biol. Int.* 44, 343–351. doi:10.1002/cbin.11232

Murça, T. M., Moraes, P. L., Capuruco, C. A. B., Santos, S. H. S., Melo, M. B., Santos, R. A. S., et al. (2012). Oral administration of an angiotensin-converting enzyme 2 activator ameliorates diabetes-induced cardiac dysfunction. *Regul. Pept.* 177, 107–115. doi:10.1016/j.regpep.2012.05.093

Nathan, N., Prevost, B., and Corvol, H. (2020). Atypical presentation of COVID-19 in young infants. *Lancet* 395, 1481. doi:10.1016/S0140-6736(20)30980-6

Nienhold, R., Ciani, Y., Koelzer, V. H., Tzankov, A., Haslbauer, J. D., Menter, T., et al. (2020). Two distinct immunopathological profiles in autopsy lungs of COVID-19. *Nat. Commun.* 11, 5086. doi:10.1038/s41467-020-18854-2

Nishiga, M., Wang, D. W., Han, Y., Lewis, D. B., and Wu, J. C. (2020). COVID-19 and cardiovascular disease: From basic mechanisms to clinical perspectives. *Nat. Rev. Cardiol.* 17, 543–558. doi:10.1038/s41569-020-0413-9

Niu, M. J., Yang, J. K., Lin, S. S., Ji, X. J., and Guo, L. M. (2008). Loss of angiotensin-converting enzyme 2 leads to impaired glucose homeostasis in mice. *Endocrine* 34, 56–61. doi:10.1007/s12020-008-9110-x

Noval Rivas, M., Wakita, D., Franklin, M. K., Carvalho, T. T., Abolhesn, A., Gomez, A. C., et al. (2019). Intestinal permeability and IgA provoke immune vasculitis linked to cardiovascular inflammation. *Immunity* 51, 508e506–521. doi:10.1016/j.immuni.2019.05.021

Oakes, J. M., Fuchs, R. M., Gardner, J. D., Lazartigues, E., and Yue, X. (2018). Nicotine and the renin-angiotensin system. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 315, R895–r906. doi:10.1152/ajpregu.00099.2018

Ottaviani, L. M. J., Sansonetti, M., Sampaio-Pinto, V., Halkein, J., el Azzouzi, H., Olieslagers, S., et al. (2019). Abstract 896: Cardiomyocyte-derived mir-200c-3p in exosomes affects endothelial angiogenic capacity and impairs cardiac function. *Circulation Res.* 125. doi:10.1161/res.125.suppl_1.896

Ottestad, W., Seim, M., and Mæhlen, J. O. (2020). COVID-19 with silent hypoxemia. *Tidsskr. Nor. Laegeforen* 140. doi:10.4045/tidsskr.20.0299

Papa, A., Salzano, A. M., Di Dato, M. T., and Varrassi, G. (2020). Images in practice: Painful cutaneous vasculitis in a SARS-cov-2 IgG-positive child. *Pain Ther.* 9, 805–807. doi:10.1007/s40122-020-00174-4

Papannarao, J. B., Schwenke, D. O., Manning, P., and Katare, R. (2021). Upregulated miR-200c may increase the risk of obese individuals to severe COVID-19. *medRxiv*, 21254517. 2021.2003.2029.

Patel, V. B., Mori, J., McLean, B. A., Basu, R., Das, S. K., Ramprasath, T., et al. (2016). ACE2 deficiency worsens epicardial adipose tissue inflammation and cardiac dysfunction in response to diet-induced obesity. *Diabetes* 65, 85–95. doi:10.2337/db15-0399

Pei, G., Zhang, Z., Peng, J., Liu, L., Zhang, C., Yu, C., et al. (2020). Renal involvement and early prognosis in patients with COVID-19 pneumonia. *J. Am. Soc. Nephrol.* 31, 1157–1165. doi:10.1681/ASN.2020030276

Phagoo, S. B., Reddi, K., Silvallana, B. J., Leeb-Lundberg, L. M., and Warburton, D. (2005). Infection-induced kinin B1 receptors in human pulmonary fibroblasts: Role of intact pathogens and p38 mitogen-activated protein kinase-dependent signaling. *J. Pharmacol. Exp. Ther.* 313, 1231–1238. doi:10.1124/jpet.104.083030

Pradeep, S., Huang, J., Mora, E. M., Nick, A. M., Cho, M. S., Wu, S. Y., et al. (2015). Erythropoietin stimulates tumor growth via EphB4. *Cancer cell* 28, 610–622. doi:10.1016/j.ccell.2015.09.008

Prado, G. N., Taylor, L., Zhou, X., Ricupero, D., Mierke, D. F., and Polgar, P. (2002). Mechanisms regulating the expression, self-maintenance, and signaling-function of the bradykinin B2 and B1 receptors. *J. Cell. physiology* 193, 275–286. doi:10.1002/jcp.10175

Puelles, V. G., Lutgehetmann, M., Lindenmeyer, M. T., Sperhake, J. P., Wong, M. N., Allweiss, L., et al. (2020). Multiorgan and renal tropism of SARS-CoV-2. *N. Engl. J. Med.* 383, 590–592. doi:10.1056/NEJMc2011400

Pulido-Olmo, H., Garcia-Prieto, C. F., Alvarez-Llamas, G., Barderas, M. G., Vivanco, F., Aranguez, I., et al. (2016). Role of matrix metalloproteinase-9 in chronic kidney disease: A new biomarker of resistant albuminuria. *Clin. Sci. (Lond)* 130, 525–538. doi:10.1042/CS20150517

Qadri, F., and BaderKinin, M. (2018). Kinin B1 receptors as a therapeutic target for inflammation. *Expert Opin. Ther. Targets* 22, 31–44. doi:10.1080/14728222.2018.1409724

Qi, Y., Zhang, J., Cole-Jeffrey, C. T., Shenoy, V., Espejo, A., Hanna, M., et al. (2013). Diminazene aceturate enhances angiotensin-converting enzyme 2 activity and attenuates ischemia-induced cardiac pathophysiology. *Hypertension* 62, 746–752. doi:10.1161/HYPERTENSIONAHA.113.01337

Rahimi, N. (2017). Defenders and challengers of endothelial barrier function. *Front. Immunol.* 8, 1847. doi:10.3389/fimmu.2017.01847

RakhmatII, Kusmala, Y. Y., Handayani, D. R., Juliastuti, H., Nawangsih, E. N., Wibowo, A., et al. (2021). Dipeptidyl peptidase-4 (DPP-4) inhibitor and mortality in coronavirus disease 2019 (COVID-19) - a systematic review, meta-analysis, and meta-regression. *Diabetes Metab. Syndr.* 15 (3), 777–782. doi:10.1016/j.dsx.2021.03.027

Ramachandran, S., Ilias Basha, H., Sarma, N. J., Lin, Y., Crippin, J. S., Chapman, W. C., et al. (2013). Hepatitis C virus induced miR200c down modulates FAP-1, a negative regulator of Src signaling and promotes hepatic fibrosis. *PLoS One* 8, e70744. doi:10.1371/journal.pone.0070744

Ramos-Miguel, A., Jones, A. A., Sawada, K., Barr, A. M., Bayer, T. A., Falkai, P., et al. (2018). Frontotemporal dysregulation of the SNARE protein interactome is associated with faster cognitive decline in old age. *Neurobiol. Dis.* 114, 31–44. doi:10.1016/j.nbd.2018.02.006

Raslan, F., Schwarz, T., Meuth, S. G., Austinat, M., Bader, M., Renne, T., et al. (2010). Inhibition of bradykinin receptor B1 protects mice from focal brain injury by reducing blood-brain barrier leakage and inflammation. *J. Cereb. Blood Flow. Metab.* 30, 1477–1486. doi:10.1038/jcbfm.2010.28

Ren, H., Yang, Y., Wang, F., Yan, Y., Shi, X., Dong, K., et al. (2020). Association of the insulin resistance marker TyG index with the severity and mortality of COVID-19. *Cardiovasc Diabetol.* 19, 58. doi:10.1186/s12933-020-01035-2

Riphagen, S., Gomez, X., Gonzalez-Martinez, C., Wilkinson, N., and Theocharis, P. (2020). Hyperinflammatory shock in children during COVID-19 pandemic. *Lancet* 395, 1607–1608. doi:10.1016/S0140-6736(20)31094-1

Rivas, M. N. W., Abe, M., Franklin, M. K., Chen, S., Shimada, K., Crother, T. R., et al. (2017). Circulation.

Rivera-Figueroa, E. I., Santos, R., Simpson, S., and Garg, P. (2020). Incomplete Kawasaki disease in a child with covid-19. *Indian Pediatr.* 57, 680–681. doi:10.1007/s13312-020-1900-0

Rodrigues Prestes, T. R., Rocha, N. P., Miranda, A. S., Teixeira, A. L., and Simoes, E. S. A. C. (2017). The anti-inflammatory potential of ACE2/angiotensin-(1-7)/mas receptor Axis: Evidence from basic and clinical research. *Curr. Drug Targets* 18, 1301–1313. doi:10.2174/1389450117666160727142401

Rodríguez, Y., Novelli, L., Rojas, M., De Santis, M., Acosta-Ampudia, Y., Monsalve, D. M., et al. (2020). Autoinflammatory and autoimmune conditions at the crossroad of COVID-19. *J. Autoimmun.* 114, 102506. doi:10.1016/j.jaut.2020.102506

Roshanravan, N., Mahdavi, R., Jafarabadi, M. A., Alizadeh, E., Alipour, S., Ghavami, A., et al. (2018). The suppression of txnip and miR-200c improve beta-cell function in patients with type 2 diabetes: A randomized, double-blind, placebo-controlled trial. *J. Funct. Foods* 48, 481–489. doi:10.1016/j.jff.2018.07.019

Ruan, Q., Yang, K., Wang, W., Jiang, L., and Song, J. (2020). Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med.* 46, 846–848. doi:10.1007/s00134-020-05991-x

Rubino, F., Amiel, S. A., Zimmet, P., Alberti, G., Bornstein, S., Eckel, R. H., et al. (2020). New-onset diabetes in covid-19. *N. Engl. J. Med.* 383, 789–790. doi:10.1056/NEJMc2018688

Sakurai, Y. (2019). Autoimmune aspects of Kawasaki disease. *J. Investig. Allergol. Clin. Immunol.* 29, 251–261. doi:10.18176/jiaci.0300

Sales, V. M., Goncalves-Zillo, T., Castoldi, A., Burgos, M., Branquinho, J., Batista, C., et al. (2019). Kinin B(1) receptor acts in adipose tissue to control fat distribution in a cell-nonautonomous manner. *Diabetes* 68, 1614–1623. doi:10.2337/db18-1150

Sasaki, A., Kanai, M., Kijima, K., Akaba, K., Hashimoto, M., Hasegawa, H., et al. (2003). Molecular analysis of congenital central hypoventilation syndrome. *Hum. Genet.* 114, 22–26. doi:10.1007/s00439-003-1036-z

Sato, T., Suzuki, T., Watanabe, H., Kadowaki, A., Fukamizu, A., Liu, P. P., et al. (2013). Apelin is a positive regulator of ACE2 in failing hearts. *J. Clin. Invest* 123, 5203–5211. doi:10.1172/JCI69608

Schell, M. J., Yang, M., Teer, J. K., Lo, F. Y., Madan, A., Coppola, D., et al. (2016). A multigene mutation classification of 468 colorectal cancers reveals a prognostic role for APC. *Nat. Commun.* 7, 11743. doi:10.1038/ncomms11743

Schickel, R., Park, S. M., Murmann, A. E., and Peter, M. E. (2010). miR-200c regulates induction of apoptosis through CD95 by targeting FAP-1. *Mol. Cell* 38, 908–915. doi:10.1016/j.molcel.2010.05.018

Schoenfeld, A. K., Lahrsen, E., and Alban, S. (2016). Regulation of complement and contact system activation via C1 inhibitor potentiation and factor XIIa activity modulation by sulfated glycans - structure-activity relationships. *PLoS One* 11, e0165493. doi:10.1371/journal.pone.0165493

Schotland, P., Racz, R., Jackson, D. B., Soldatos, T. G., Levin, R., Strauss, D. G., et al. (2021). Target adverse event profiles for predictive safety in the postmarket setting. *Clin. Pharmacol. Ther.* 109, 1232–1243. doi:10.1002/cpt.2074

Scroggin, M. P., Pedersen, K. M., and Lazatigues, E. (2012). The PPAR-γ agonist Rosiglitazone increases angiotensin-converting enzyme 2 (ACE2) promoter activity in neurons. *FASEB J.* 26, 875–813. doi:10.1096/fasebj.26.1_supplement.875.13

Seguin, T., Buleon, M., Destrube, M., Ranera, M. T., CoutuRe, R., Girolami, J. P., et al. (2008). Hemodynamic and renal involvement of B1 and B2 kinin receptors during the acute phase of endotoxin shock in mice. *Int. Immunopharmacol.* 8, 217–221. doi:10.1016/j.intimp.2007.08.008

Selby, N. M., Forni, L. G., Laing, C. M., Horne, K. L., Evans, R. D., Lucas, B. J., et al. (2020). Covid-19 and acute kidney injury in hospital: Summary of NICE guidelines. *Bmj* 369, m1963. doi:10.1136/bmj.m1963

Serfozo, P., Wysocki, J., Gulua, G., Schulze, A., Ye, M., Liu, P., et al. (2020). Ang II (angiotensin II) conversion to angiotensin-(1-7) in the circulation is POP (Prolyloligopeptidase)-Dependent and ACE2 (Angiotensin-Converting enzyme 2)-independent. *Hypertension* 75, 173–182. doi:10.1161/HYPERTENSIONAHA.119.14071

Shah, S. A., Khan, M., Jo, M. H., Jo, M. G., Amin, F. U., and Kim, M. O. (2017). Melatonin stimulates the SIRT1/nrf2 signaling pathway counteracting lipopolysaccharide (LPS)-Induced oxidative stress to rescue postnatal rat brain. *CNS Neurosci. Ther.* 23, 33–44. doi:10.1111/cns.12588

Shao, H., Qin, Z., Geng, B., Wu, J., Zhang, L., Zhang, Q., et al. (2020). Impaired lung regeneration after SARS-CoV-2 infection. *Cell Prolif.* 53, e12927. doi:10.1111/cpr.12927

Shao, M., Wen, Z. B., Yang, H. H., Zhang, C. Y., Xiong, J. B., Guan, X. X., et al. (2019). Exogenous angiotensin (1-7) directly inhibits epithelial-mesenchymal transformation induced by transforming growth factor-β1 in alveolar epithelial cells. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 117, 109193. doi:10.1016/j.biopha.2019.109193

Sheridan, C. (2020). Massive data initiatives and AI provide testbed for pandemic forecasting. *Nat. Biotechnol.* 38, 1010–1013. doi:10.1038/s41587-020-0671-4

Shigemura, N., Takai, S., Hirose, F., Yoshida, R., Sanematsu, K., and Ninomiya, Y. (2019). Expression of renin-angiotensin system components in the taste organ of mice. *Nutrients* 11, 2251. doi:10.3390/nu11092251

Silva de Souza, A. W. (2015). Autoantibodies in systemic vasculitis. *Front. Immunol.* 6, 184. doi:10.3389/fimmu.2015.00184

Singer, D. C., and Camargo, S. M. R. (2011). Collectrin and ACE2 in renal and intestinal amino acid transport. *Channels* 5, 410–423. doi:10.4161/chan.5.5.16470

Singh, G. B., Raut, S. K., Khanna, S., Kumar, A., Sharma, S., Prasad, R., et al. (2017). MicroRNA-200c modulates DUSP-1 expression in diabetes-induced cardiac hypertrophy. *Mol. Cell Biochem.* 424, 1–11. doi:10.1007/s11010-016-2838-3

Sodhi, C. P., Wohlford-Lenane, C., Yamaguchi, Y., Prindle, T., Fulton, W. B., Wang, S., et al. (2018). Attenuation of pulmonary ACE2 activity impairs inactivation of des-Arg(9) bradykinin/BKB1R axis and facilitates LPS-induced neutrophil infiltration. *Am. J. Physiol. Lung Cell Mol. Physiol.* 314, L17–l31. doi:10.1152/ajplung.00498.2016

Soldatos, T. G., and Jackson, D. B. (2019). Adverse event circumstances and the case of drug interactions. *Healthc. (Basel)* 7, 45. doi:10.3390/healthcare7010045

Soldatos, T. G., Taglang, G., and Jackson, D. B. (2018). *In silico* profiling of clinical phenotypes for human targets using adverse event data. *High-throughput* 7, 37. doi:10.3390/ht7040037

Stewart, G. M., Johnson, B. D., Sprecher, D. L., Reddy, Y. N. V., Obokata, M., Goldsmith, S., et al. (2020). Targeting pulmonary capillary permeability to reduce lung congestion in heart failure: A randomized, controlled pilot trial. *Eur. J. Heart Fail* 22 (9), 1641–1645. doi:10.1002/ejhf.1809

Su, H., Yang, M., Wan, C., Yi, L. X., Tang, F., Zhu, H. Y., et al. (2020). Renal histopathological analysis of 26 postmortem findings of patients with COVID-19 in China. *Kidney Int.* 98, 219–227. doi:10.1016/j.kint.2020.04.003

Su, X., Camerer, E., Hamilton, J. R., Coughlin, S. R., and Matthay, M. A. (2005). Protease-activated receptor-2 activation induces acute lung inflammation by neuropeptide-dependent mechanisms. *J. Immunol.* 175, 2598–2605. doi:10.4049/jimmunol.175.4.2598

Südhof, T. C., and Rothman, J. E. (2009). Membrane fusion: Grappling with SNARE and SM proteins. *Sci. (New York, N.Y.)* 323, 474–477. doi:10.1126/science.1161748

Sungnak, W., Huang, N., Bécavin, C., and Berg, M. SARS-CoV-2 entry genes are most highly expressed in nasal goblet and ciliated cells within human airways. ArXiv (2020).

Symowicz, J., Adley, B. P., Gleason, K. J., Johnson, J. J., Ghosh, S., Fishman, D. A., et al. (2007). Engagement of collagen-binding integrins promotes matrix metalloproteinase-9-dependent E-cadherin ectodomain shedding in ovarian carcinoma cells. *Cancer Res.* 67, 2030–2039. doi:10.1158/0008-5472.CAN-06-2808

Takahashi, Y., Haga, S., Ishizaka, Y., and Mimori, A. (2010). Autoantibodies to angiotensin-converting enzyme 2 in patients with connective tissue diseases. *Arthritis Res. Ther.* 12, R85. doi:10.1186/ar3012

Tang, M., Liu, P., Li, X., Wang, J. W., Zhu, X. C., and He, F. P. (2017). Protective action of B1R antagonist against cerebral ischemia-reperfusion injury through suppressing miR-200c expression of Microglia-derived microvesicles. *Neurol. Res.* 39, 612–620. doi:10.1080/01616412.2016.1275096

Tharaux, P. L., and Dhaun, N. (2017). Endothelium-neutrophil communication via B1-kinin receptor-bearing microvesicles in vasculitis. *J. Am. Soc. Nephrol.* 28, 2255–2258. doi:10.1681/ASN.2017030300

Thomas, T., Stefanoni, D., Reisz, J. A., Nemkov, T., Bertolone, L., Francis, R. O., et al. (2020). medRxiv, 20102491. doi:10.1101/2020.05.14.20102491COVID-19 infection results in alterations of the kynurenine pathway and fatty acid metabolism that correlate with IL-6 levels and renal status

Thomson, T. M., Toscano-Guerra, E., Casis, E., and Paciucci, R. (2020). C1 esterase inhibitor and the contact system in COVID-19. *Br. J. Haematol.* 190, 520–524. doi:10.1111/bjh.16938

Tomlins, S. A., Rhodes, D. R., Perner, S., Dhanasekaran, S. M., Mehra, R., Sun, X. W., et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Sci. (New York, N.Y.)* 310, 644–648. doi:10.1126/science.1117679

Totura, A. L., Whitmore, A., Agnihothram, S., Schafer, A., Katze, M. G., Heise, M. T., et al. (2015). Toll-like receptor 3 signaling via TRIF contributes to a protective innate immune response to severe acute respiratory syndrome coronavirus infection. *mBio* 6, e00638–e00615. doi:10.1128/mBio.00638-15

Toubiana, J., Poirault, C., Corsia, A., Bajolle, F., Fourgeaud, J., Angoulvant, F., et al. (2020). Kawasaki-like multisystem inflammatory syndrome in children during the Covid-19 pandemic in paris, France: Prospective observational study. *Bmj* 369, m2094. doi:10.1136/bmj.m2094

Tryndyak, V. P., Beland, F. A., and Pogribny, I. P. (2010). E-cadherin transcriptional down-regulation by epigenetic and microRNA-200 family alterations is related to mesenchymal and drug-resistant phenotypes in human breast cancer cells. *Int. J. Cancer* 126, 2575–2583. doi:10.1002/ijc.24972

Tseng, Y. H., Yang, R. C., and Lu, T. S. (2020). Two hits to the renin-angiotensin system may play a key role in severe COVID-19. *Kaohsiung J. Med. Sci.* 36, 389–392. doi:10.1002/kjm2.12237

Tukiainen, T., Villani, A. C., Yen, A., Rivas, M. A., Marshall, J. L., Satija, R., et al. (2017). Landscape of X chromosome inactivation across human tissues. *Nature* 550, 244–248. doi:10.1038/nature24265

Turner, R. J., and Sharp, F. R. (2016). Implications of MMP9 for blood brain barrier disruption and hemorrhagic transformation following ischemic stroke. *Front. Cell Neurosci.* 10, 56. doi:10.3389/fncel.2016.00056

Uchibori, A., and Chiba, A. (2015). Autoantibodies in guillain-barré syndrome. *Brain Nerve* 67, 1347–1357. doi:10.11477/mf.1416200305

Ueland, T., Holter, J. C., Holten, A. R., Muller, K. E., Lind, A., Bekken, G. K., et al. (2020). Distinct and early increase in circulating MMP-9 in COVID-19 patients with respiratory failure. *J. Infect.* 81, e41–e43. doi:10.1016/j.jinf.2020.06.061

Uppal, N. N., Kello, N., Shah, H. H., Khanin, Y., De Oleo, I. R., Epstein, E., et al. (2020). De novo ANCA-associated vasculitis with glomerulonephritis in COVID-19. *Kidney Int. Rep.* 5, 2079–2083. doi:10.1016/j.ekir.2020.08.012

Urwyler, P., Moser, S., Charitos, P., Heijnen, I. A. F. M., Rudin, M., Sommer, G., et al. (2020). Treatment of COVID-19 with conestat alfa, a regulator of the complement, contact activation and kallikrein-kinin system. *Front. Immunol.* 11, 2072. doi:10.3389/fimmu.2020.02072

Vaisar, T., Kassim, S. Y., Gomez, I. G., Green, P. S., Hargarten, S., Gough, P. J., et al. (2009). MMP-9 sheds the beta2 integrin subunit (CD18) from macrophages. *Mol. Cell. proteomics MCP* 8, 1044–1060. doi:10.1074/mcp.M800449-MCP200

van der Lugt, N. M., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., et al. (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes Dev.* 8, 757–769. doi:10.1101/gad.8.7.757

van de Veerdonk, F. L., Kouijzer, I. J. E., de Nooijer, A. H., van der Hoeven, H. G., Maas, C., Netea, M. G., et al. (2020). Outcomes associated with use of a kinin B2 receptor antagonist among patients with COVID-19. *JAMA Netw. Open* 3 (8), e2017708. doi:10.1001/jamanetworkopen.2020.17708

van Swinderen, B., and Kottler, B. (2014). Explaining general anesthesia: A two-step hypothesis linking sleep circuits and the synaptic release machinery. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 36, 372–381. doi:10.1002/bies.201300154

Varga, Z., Flammer, A. J., Steiger, P., Haberecker, M., Andermatt, R., Zinkernagel, A. S., et al. (2020). Endothelial cell infection and endotheliitis in COVID-19. *Lancet* 395, 1417–1418. doi:10.1016/S0140-6736(20)30937-5

Vaz, M., Hwang, S. Y., Kagiampakis, I., Phallen, J., Patil, A., O'Hagan, H. M., et al. (2017). Chronic cigarette smoke-induced epigenomic changes precede sensitization of bronchial epithelial cells to single-step transformation by KRAS mutations. *Cancer Cell* 32, 360e366–376. doi:10.1016/j.ccell.2017.08.006

Vaziri, N. D., Yuan, J., Nazertehrani, S., Ni, Z., and Liu, S. (2013). Chronic kidney disease causes disruption of gastric and small intestinal epithelial tight junction. *Am. J. Nephrol.* 38, 99–103. doi:10.1159/000353764

Verdoni, L., Mazza, A., Gervasoni, A., Martelli, L., Ruggeri, M., Ciuffreda, M., et al. (2020). An outbreak of severe kawasaki-like disease at the Italian epicentre of the SARS-CoV-2 epidemic: An observational cohort study. *Lancet* 395, 1771–1778. doi:10.1016/S0140-6736(20)31103-X

Vickers, C., Hales, P., Kaushik, V., Dick, L., Gavin, J., Tang, J., et al. (2002). Hydrolysis of biological peptides by human angiotensin-converting enzyme-related carboxypeptidase. *J. Biol. Chem.* 277, 14838–14843. doi:10.1074/jbc.M200581200

Viecca, M., Radovanovic, D., Forleo, G. B., and Santus, P. (2020). Enhanced platelet inhibition treatment improves hypoxemia in patients with severe Covid-19 and hypercoagulability. A case control, proof of concept study. *Pharmacol. Res.* 158, 104950. doi:10.1016/j.phrs.2020.104950

Vu, T. H. (2001). Don't mess with the matrix. *Nat. Genet.* 28, 202–203. doi:10.1038/90023

Waldman, M., Cohen, K., Yadin, D., Nudelman, V., Gorfil, D., Laniado-Schwartzman, M., et al. (2018). Regulation of diabetic cardiomyopathy by caloric restriction is mediated by intracellular signaling pathways involving 'SIRT1 and PGC-1α. *Cardiovasc Diabetol.* 17, 111. doi:10.1186/s12933-018-0754-4

Walker, K., Perkins, M., and Dray, A. (1995). Kinins and kinin receptors in the nervous system. *Neurochem. Int.* 26, 1–16. doi:10.1016/0197-0186(94)00114-a

Wang, J., Cooper, J. M., Gokhale, K., Acosta-Mena, D., Dhalla, S., Byne, N., et al. (2021). Association of Metformin with susceptibility to COVID-19 in people with type 2 diabetes. *J. Clin. Endocrinol. Metab.* 106, 1255–1268. doi:10.1210/clinem/dgab067

Wang, X. D., Povysil, G., Zoghbi, A., Motelow, J., Hostyk, J., Goldstein, D., et al. (2020). *Preprints.*

Ware, L. B. (2020). Physiological and biological heterogeneity in COVID-19-associated acute respiratory distress syndrome. *Lancet Respir. Med.* 8, 1163–1165. doi:10.1016/S2213-2600(20)30369-6

Watanabe, Y., Katayama, N., Takeuchi, K., Togano, T., Itoh, R., Sato, M., et al. (2013). Point mutation in syntaxin-1A causes abnormal vesicle recycling, behaviors, and short term plasticity. *J. Biol. Chem.* 288, 34906–34919. doi:10.1074/jbc.M113.504050

Wilkerson, R. G., Adler, J. D., Shah, N. G., and Brown, R. (2020). Silent hypoxia: A harbinger of clinical deterioration in patients with COVID-19. *Am. J. Emerg. Med.* 38, e2245–e2246. doi:10.1016/j.ajem.2020.05.044

Wilson, S., Greer, B., Hooper, J., Zijlstra, A., Walker, B., Quigley, J., et al. (2005). The membrane-anchored serine protease, TMPRSS2, activates PAR-2 in prostate cancer cells. *Biochem. J.* 388, 967–972. doi:10.1042/BJ20041066

Wittekindt, O. H. (2017). Tight junctions in pulmonary epithelia during lung inflammation. *Pflugers Arch.* 469, 135–147. doi:10.1007/s00424-016-1917-3

Wu, C., Chen, X., Cai, Y., Xia, J., Zhou, X., Xu, S., et al. (2020). Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in wuhan, China. *JAMA Intern Med.* 180, 934–943. doi:10.1001/jamainternmed.2020.0994

Wu, H., Zhang, Z., Du, Q., Yao, H., Li, Z., Wu, L., et al. (2011). Down-regulation of Apelin in obesity-related hypertensive rats induced by high-fat diet: Possible role of angiotensin II–AT1R system. *Int. J. Cardiol.* 152, S50–S51. doi:10.1016/j.ijcard.2011.08.633

Wu, J., Akaike, T., Hayashida, K., Miyamoto, Y., Nakagawa, T., Miyakawa, K., et al. (2002). Identification of bradykinin receptors in clinical cancer specimens and murine tumor tissues. *Int. J. Cancer* 98, 29–35. doi:10.1002/ijc.10142

Wu, Y. H., Lin, H. R., Lee, Y. H., Huang, P. H., Wei, H. C., Stern, A., et al. (2017). A novel fine tuning scheme of miR-200c in modulating lung cell redox homeostasis. *Free Radic. Res.* 51, 591–603. doi:10.1080/10715762.2017.1339871

Xie, J., Tong, Z., Guan, X., Du, B., and Qiu, H. (2020). Clinical characteristics of patients who died of coronavirus disease 2019 in China. *JAMA Netw. Open* 3, e205619. doi:10.1001/jamanetworkopen.2020.5619

Xie, X., Chen, J., Wang, X., Zhang, F., Liu, Y., Xingxiang, W., et al. (2006). Age- and gender-related difference of ACE2 expression in rat lung. *Life Sci.* 78, 2166–2171. doi:10.1016/j.lfs.2005.09.038

Xie, X., He, Q., Huang, L., Li, L., Yao, Y., Xia, H., et al. (2019). Associations of SLC6A20 genetic polymorphisms with Hirschsprung's disease in a Southern Chinese population. *Biosci. Rep.* 39, BSR20182290. doi:10.1042/BSR20182290

Xu, H., Zhong, L., Deng, J., Peng, J., Dan, H., Zeng, X., et al. (2020). High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int. J. Oral Sci.* 12, 8. doi:10.1038/s41368-020-0074-x

Yacoub, R., Lee, K., and He, J. C. (2014). The role of SIRT1 in diabetic kidney disease. *Front. Endocrinol.* 5, 166. doi:10.3389/fendo.2014.00166

Yamakawa, K., Fukuta, S., Yoshinaga, T., Umemoto, S., Itagaki, T., and Kusukawa, R. (1987). Study of immunological mechanism in dilated cardiomyopathy. *Jpn. Circ. J.* 51 (6), 665–675. doi:10.1253/jcj.51.665

Yang, D., Livingston, M. J., Liu, Z., Dong, G., Zhang, M., Chen, J. K., et al. (2018). Autophagy in diabetic kidney disease: Regulation, pathological role and therapeutic potential. *Cell Mol. Life Sci.* 75, 669–688. doi:10.1007/s00018-017-2639-1

Yang, J. K., Lin, S. S., Ji, X. J., and Guo, L. M. (2010). Binding of SARS coronavirus to its receptor damages islets and causes acute diabetes. *Acta Diabetol.* 47, 193–199. doi:10.1007/s00592-009-0109-4

Yang, R., Ma, H., Thomas, S. M., and Kinnamon, J. C. (2007). Immunocytochemical analysis of syntaxin-1 in rat circumvallate taste buds. *J. Comp. Neurol.* 502, 883–893. doi:10.1002/cne.21317

Yang, S., Banerjee, S., de Freitas, A., Sanders, Y. Y., Ding, Q., Matalon, S., et al. (2012). Participation of miR-200 in pulmonary fibrosis. *Am. J. Pathol.* 180, 484–493. doi:10.1016/j.ajpath.2011.10.005

Yue, X., Basting, T. M., Flanagan, T. W., Xu, J., Lobell, T. D., Gilpin, N. W., et al. (2018). Nicotine downregulates the compensatory angiotensin-converting enzyme 2/angiotensin type 2 receptor of the renin–angiotensin system. *Ann. Am. Thorac. Soc.* 15, S126–S127. doi:10.1513/annalsats.201706-464mg

Zhang, H., Liu, J., Qu, D., Wang, L., Luo, J. Y., Lau, C. W., et al. (2016). Inhibition of miR-200c restores endothelial function in diabetic mice through suppression of COX-2. *Diabetes* 65, 1196–1207. doi:10.2337/db15-1067

Zhang, J., Dong, J., Martin, M., He, M., Gongol, B., Marin, T. L., et al. (2018). AMP-Activated protein kinase phosphorylation of angiotensin-converting enzyme 2 in endothelium mitigates pulmonary hypertension. *Am. J. Respir. Crit. Care Med.* 198, 509–520. doi:10.1164/rccm.201712-2570OC

Zhang, J. J., Dong, X., Cao, Y. Y., Yuan, Y. D., Yang, Y. B., Yan, Y. Q., et al. (2020). Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy* 75, 1730–1741. doi:10.1111/all.14238

Zhang, R., Wang, X., Ni, L., Di, X., Ma, B., Niu, S., et al. (2020). COVID-19: Melatonin as a potential adjuvant treatment. *Life Sci.* 250, 117583. doi:10.1016/j.lfs.2020.117583

Zhang, J., Luo, Y., Wang, X., Zhu, J., Li, Q., Feng, J., et al. (2019). Global transcriptional regulation of STAT3- and MYC-mediated sepsis-induced ARDS. *Ther. Adv. Respir. Dis.* 13, 1753466619879840. doi:10.1177/1753466619879840

Zhang, Q., Fillmore, T. L., Schepmoes, A. A., Clauss, T. R. W., Gritsenko, M. A., Mueller, P. W., et al. (2013). Serum proteomics reveals systemic dysregulation of innate immunity in type 1 diabetes. *J. Exp. Med.* 210, 191–203. doi:10.1084/jem.20111843

Zhang, W., Wang, Y., Zeng, Y., Hu, L., and Zou, G. (2017). Serum miR-200c and miR-371-5p as the useful diagnostic biomarkers and therapeutic targets in Kawasaki disease. *Biomed. Res. Int.* 2017, 8257862. doi:10.1155/2017/8257862

Zhang, R., Shi, H., Ren, F., Zhang, M., Ji, P., Wang, W., et al. (2017). The aberrant upstream pathway regulations of CDK1 protein were implicated in the proliferation and apoptosis of ovarian cancer cells. *J. ovarian Res.* 10, 60. doi:10.1186/s13048-017-0356-x

Zhao, H., Shen, D., Zhou, H., Liu, J., and Chen, S. (2020). Guillain-barré syndrome associated with SARS-CoV-2 infection: Causality or coincidence? *Lancet Neurol.* 19, 383–384. doi:10.1016/S1474-4422(20)30109-5

Zheng, Y. Y., Ma, Y. T., Zhang, J. Y., and Xie, X. (2020). COVID-19 and the cardiovascular system. *Nat. Rev. Cardiol.* 17, 259–260. doi:10.1038/s41569-020-0360-5

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in wuhan, China: A retrospective cohort study. *Lancet* 395, 1054–1062. doi:10.1016/S0140-6736(20)30566-3

Zhu, N., Wang, W., Liu, Z., Liang, C., Wang, W., Ye, F., et al. (2020). Morphogenesis and cytopathic effect of SARS-CoV-2 infection in human airway epithelial cells. *Nat. Commun.* 11, 3910. doi:10.1038/s41467-020-17796-z

Ziaei, A., Davoodian, P., Dadvand, H., Safa, O., Hassanipour, S., Omidi, M., et al. (2020). Evaluation of the efficacy and safety of melatonin in moderately ill patients with COVID-19: A structured summary of a study protocol for a randomized controlled trial. *Trials* 21 (1), 882. doi:10.1186/s13063-020-04737-w

Ziegler, C. G. K., Allon, S. J., Nyquist, S. K., Mbano, I. M., Miao, V. N., Tzouanas, C. N., et al. (2020). SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* 181, 10161016–10161035. doi:10.1016/j.cell.2020.04.035

frontiers | Frontiers in Molecular Medicine

Definition of a microbial signature as a predictor of endoscopic post-surgical recurrence in patients with Crohn's disease

Lia Oliver[1], Blau Camps[2], David Julià-Bergkvist[3], Joan Amoedo[1], Sara Ramió-Pujol[1], Marta Malagón[1], Anna Bahí[4], Paola Torres[5], Eugeni Domènech[5], Jordi Guardiola[2], Mariona Serra-Pagès[1], Jesus Garcia-Gil[1,6] and Xavier Aldeguer[1,3,4]*

[1]GoodGut S.L.U, Girona, Spain, [2]Hospital Universitari de Bellvitge, l'Hospitalet de Llobregat, Barcelona, Spain, [3]Hospital Universitari de Girona Doctor Josep Trueta, Girona, Spain, [4]Institut d'Investigació Biomèdica de Girona—IdIBGi Girona, Girona, Spain, [5]Hospital Germans Tries i Pujol, CIBEREHD Badalona, Badalona, Spain, [6]Biology Department, University of Girona, Girona, Spain

Background and aims: Although there are several effective drugs for the treatment of Crohn's disease (CD), almost 70% of patients will require surgical resection during their lifetime. This procedure is not always curative, as endoscopic recurrence occurs in 65%–90% of patients in the first year after surgery. The aetiology of the recurrence is unknown; however, several studies have shown how the resident microbiota is modified after surgery. The aim of this study was to evaluate samples from patients with Crohn's disease before and after an intestinal resection to determine whether there were differences in the abundance of different microbial markers, which may predict endoscopic recurrence at baseline.

Methods: In this observational study, a stool sample was obtained from 25 patients with Crohn's disease before undergoing surgery, recruited at three Catalan hospitals. From each sample, DNA was purified and the relative abundance of nine microbial markers was quantified using qPCR.

Results: An algorithm composed of four microbial markers (*E. coli*, *F. prausnitzii phylogroup I*, *Bacteroidetes*, and *Eubacteria*) showed a sensitivity and specificity of 90.91% and 85.71%, respectively, and a positive and negative predictive value of 83.33% and 92.31%, respectively.

Conclusion: A microbial signature to determine patients who will have post-surgical recurrence was identified. This tool might be very useful in daily clinical practice, allowing the scheduling of personalized therapy and enabling preventive treatment only in patients who really require it.

KEYWORDS

Crohn's disease, gut microbiota, endoscopic recurrence, precision medicine, surgical resection

1 Introduction

Crohn's disease (CD) is a chronic inflammatory bowel disease (IBD) with symptoms evolving in a relapsing and remitting manner that leads to bowel damage and disability (Baumgart and Sandborn, 2012). All segments of the gastrointestinal tract may affected, most commonly the terminal ileum and colon (Torres et al., 2017). Even with the advent of

immunomodulatory therapies, up to 70% of CD patients require intestinal resection during their lifetime (Hamilton et al., 2020), of whom subclinical endoscopic recurrence occurs in the anastomosis in 90% and approximately two-thirds require further surgery in the first year after surgery (De Cruz et al., 2015). Given the significant risk of recurrence after surgery, many clinical factors have been shown to predispose to recurrence, including smoking, younger age of disease onset, length of the resected segment and disease behaviour, although these factors are far from being adequate in predicting recurrence (Gklavas et al., 2017; Zhuang et al., 2021). Currently, the gold standard for diagnosing recurrence is to calculate the Rutgeerts score by performing a colonoscopy within the first year after surgery to adapt treatment (Rutgeerts. et al., 1990). However, early therapeutic intervention to prevent disease recurrence is recommended after surgery (Ding et al., 2016).

Microbiome studies have linked IBD to an alteration in gut microbiota communities in comparison with the predominant composition in healthy controls, and this is known to play a key role in disease initiation and maintenance (Manichanh et al., 2012; Manichanh, 2006; Joossens et al., 2011; Macfarlane et al., 2009; Pascal et al., 2017). Post-operative disease recurrence has also been related to the microbiota, with large differences found between patients with and without recurrence before undergoing surgery (Wright et al., 2017; Sokol et al., 2019; Hamilton et al., 2020). This suggests there may be a microbial signature, detectable before resection, that may be associated with the risk of postoperative recurrence. A surgical prognosis would permit the tailoring the treatment of patients with CD undergoing surgery, allowing early detection and treatment management and, in consequence, improving the quality of life.

Therefore, the aim of this study was to assess whether, at baseline (before surgery), there are differences in the abundance of different microbial markers depending on whether patients undergoing surgery develop endoscopic recurrence, and to evaluate the impact of ileocolonic resection on the abundance of gut microbiota species and whether endoscopic recurrence may be predicted at baseline by a single microbial marker or the definition of a microbial signature capable of differentiating between groups with high diagnostic capacity.

2 Materials and methods

2.1 Study population

The study cohort comprised 25 CD patients scheduled for ileocecal resection surgery recruited between April 2018 and July 2020. This multicentre randomized study was conducted in three centres: The Department of Gastroenterology, Hospital Universitari Dr. Josep Trueta (Girona, Catalonia), Hospital Universitari de Bellvitge (Barcelona, Catalonia), and Hospital Germans Trias i Pujol (Badalona, Catalonia). Ethical approval was obtained from all participating hospitals. Patients underwent habitual clinical procedures. Informed consent was obtained from each patient.

The inclusion criteria were age >18 years, signed informed consent, ileal or ileocolonic CD, and an indication for CD-related intestinal surgery (ileocolonic resection) for the first time. Subjects were excluded if they had received antibiotics in the month before the operation, were pregnant at inclusion or had a history of illness and/or surgery that compromised the digestive tract.

Baseline clinical characteristics of patients are shown in Table 1. A complete resection was systematically made, with all margins free from inflammation.

Study patients were asked to collect a faecal sample no earlier than 1 week before surgery and at 3 and 6 months post-surgery. To ensure sample viability, they had to be stored in one of the following conditions according to the time elapsed between sample deposition and its arrival at the hospital: 1) If the time elapsed was <4 h, the sample could be stored at room temperature, 2) if the time elapsed between sampling and arrival at the hospital was more than 4 but less than 12 h, it had to be kept in a refrigerator at 4 C, and 3) if the sample took more than 12 h before arrival at the hospital, it had to be frozen at −20°C. Samples were kept frozen at −20°C once received at the hospital and stored at −80°C upon analysis.

A stool sample related to colonoscopy results was collected 6 months post-surgery. These samples were asked to be collected prior to bowel preparation, and frozen immediately after deposition at −20°C. The Rutgeerts index was used to assess endoscopic recurrence at approximately 6 months post-surgery, and recurrence was defined as a Rutgeerts score $\geq$ i2.

2.2 DNA extraction from stool samples

Total DNA was isolated from faecal samples using the NucleoSpin Soil DNA kit (Macherey-Nagel GmbH &Co., Düren, Germany), according to the manufacturer's instructions, and eluting DNA in a 100 μL final volume of SE Elution.

2.3 qPCR analysis

The abundance of nine microbial markers was analysed using qPCR: Eubacteria (EUB), *Faecalibacterium prausnitzii* (FAE), and its two phylogroups (PHGI, PHGII), *Akkermansia muciniphila* (AKK), *Escherichia coli* (ECO), Bacteroidetes (BAC), *Ruminococcus* spp. (RUM), and *Methanobrevibacter smithii* (MSM). The targeting primers and probes used are shown in Table 2. These microbial species or taxa were chosen as a representative of the gut microbiota: Eubacteria was analysed as the total bacterial load representation (Nadkarni et al., 2002); *Ruminococcus* spp, *F. prausnitzii* and its phylogroups were analysed as butyrate-producing bacteria and as indicators of the mucosa homeostasis together with *A. muciniphila,* which have been related to IBD (Hanauer, 2019; la Reau and Suen, 2018; Derrien et al., 2017; Miquel et al., 2013; Watterlot et al., 2008; Lopez-Siles et al., 2014a); *E. coli* was analysed as a pro-inflammatory bacteria; *M. smithii* as a methanogen representation; and Bacteroidetes as a representative of mucosal equilibrium.

EUB, AKK, RUM, MSM, and BAC were quantified by preparing single reactions of each biomarker using GoTaq qPCR Bryt Master mix (Promega, Madison, United States). Reactions consisted of 10 μL containing 1X GoTaq® qPCR Master Mix (Promega), between 200 and 300 nM of each primer (specified in Table 2), and between 12 and 20 ng of genomic DNA template. FPRA, PHGI, PHGII, and ECO were quantified by preparing a single reaction for each biomarker using GoTaq qPCR Probe Master Mix (Promega. Madison, United States). Reactions consisted of 10 μL containing 1X GoTaq® qPCR Master Mix (Promega), 300 nM of each primer, between 100 and 250 nM of each probe (specified in Table 3), and between 12 and 20 ng of genomic

TABLE 1 Patient characteristics.

Demographics	Recurrence (Rutgeerts i0-i1)	Remission (Rutgeerts i2-i4)
N, (male)	11 (8)	14 (4)
Mean age at surgery	35	41
Age range	21–47	21–61
Age at diagnosis	N (%)	N (%)
A1 ≤ 16 Year	1 (9.1%)	—
A2 17–40	8 (72.7%)	11 (78,6%)
A3 >40	2 (18.2%)	3 (21.4%)
Disease location at surgery	N (%)	N (%)
L1 Ileum only	6 (54.5%)	9 (64.3%)
L2 Colon only	—	—
L3 Ileum and colon	5 (45.5%)	5 (35.7%)
L4 Upper GI	—	—
Disease phenotype at surgery	N (%)	N (%)
B1 Inflammatory	1 (9.1%)	1 (7.1%)
B2 Stricturing	4 (36.3%)	8 (57.2%)
B3 Penetrating	5 (45.5%)	3 (21.4%)
P Perianal disease	1 (9.1%)	2 (14.3%)

DNA template. The primers used were purchased from Macrogen (Macrogen, Seoul, South Korea).

All quantitative PCRs were run on an AriaMx Real-time PCR System (Agilent Technologies, Santa Clara, United States). Thermal profiles of all qPCRs included an initial denaturation step set at 95°C for 10 min and 40 cycles with 15 s of denaturing at 95°C and 1 min of elongation at 60°C (except for FPRA phylogroups, in which the elongation step was at 64°C). A melting curve step was added at the end of each qPCR with GoTaq qPCR Bryt Master Mix and was used to verify the presence of the expected amplicon size and to control primer dimer formation when applicable. All samples were amplified in duplicate, which were considered valid when the difference between threshold cycles (Ct) was <0.6. A non-template control and a positive control reaction were included in each qPCR run.

Quantification data of each microbial marker was collected and analysed using Aria Software version 1.71 (Agilent Technologies, Santa Clara, United States).

Once the results were obtained, data were normalised by Eubacteria to avoid differences due to laboratory procedures or the abundance of biomarkers.

2.4 Calprotectin quantification

Calprotectin is a calcium-binding protein expressed in neutrophils and monocytes and is used as an inflammatory marker for gastrointestinal diseases (Stříž and Trebichavský, 2004; Ayling and Kok, 2018). Elevated levels of calprotectin in faeces are an indicator of active disease, making it a widely used marker to monitor therapy in patients with IBD (Boschetti et al., 2015). The inflammatory state was quantified by measuring the amount of faecal calprotectin using the PhicalR ELISA Immunodiagnostic Test. The analysis was made at LABCO (Barcelona, Spain). Levels >250 μg/g were associated with remission.

2.5 Statistical analysis

Data normality was assessed using the Kolmogorov-Smirnov test. The non-parametric Kruskal–Wallis test was used to test differences in variables with more than two categories. Pairwise comparisons of subcategories of these variables were analysed using the non-parametric Mann-Whitney test. Paired data were analysed using the McNemar test and correlations using the Pearson test.

The receiver operating characteristic (ROC) curve was used to determine the usefulness of each biomarker in distinguishing between patients with and without post-surgical recurrence. The accuracy of discrimination was measured by the area under the ROC curve (AUC). An AUC approaching 1 indicates that the test is highly sensitive and specific whereas an AUC approaching 0.5 indicates that the test is neither sensitive nor specific. Chi-square tests were performed to assess the influence of categorical variables on the response variable.

All results using microbial markers were performed using their relative abundances (Ct value of each biomarker normalized by EUB Ct value). This normalized the results and the effects derived from the process such as the DNA concentration of each sample were eliminated.

TABLE 2 16S rRNA-targeted primers and probes sequences used in this study. The fluorophore and the quencher of the probes are also shown. NA, not applicable.

Target microbial marker	Primers and probes			Concentration (nM)	
	Code	Sequence 5′→3′	Reference	Forward and reverse primers	Probe
F. prausnitzii	FAE_F	TGTAAACTCCTGTTGTTGAGGAAGATAA	Lopez-Siles et al. (2014b)	250	300
	FAE_R	GCGCTCCCTTTACACCCA			
	FAE_PR	FAM-CAAGGAAGTGACGGCTAACTACGTGCCAG-TAMRA			
F. prausnitzii (phylogroups)	PHG_F	CTCAAAGAGGGGGACAACAGTT	Lopez-Siles et al. (2016)	300	900
	PHG_R	GCCATCTCAAAGCGGATTG			
	PHGI_PR	FAM-TAAGCCCACGACCCGGCATCG-BHQ1			
	PHGII_PR	HEX-TAAGCCCACRGCTCGGCATC-BHQ1			
A. muciniphila	AKK_F	CAGCACGTGAAGGTGGGGAC	Collado et al. (2007)	250	NA
	AKK_R	CCTTGCGGTTGGCTTCAGAT			
Ruminococcus spp.	RUM_F	GGCGGCYTRCTGGGCTTT	Ramirez-Farias et al. (2009)	250	NA
	RUM_R	CCAGGTGGATWACTTATTGTGTTAA			
M. smithii	MSM_F	ACGCAGCTTAAACCACAGTC	Ramió-Pujol et al. (2020)	200	NA
	MSM_R	AAAGACATTGACCCRCGCAT			
E. coli	ECO_F	CATGCCGCGTGTATGAAGAA	Huijsdens et al. (2002)	300	100
	ECO_R	CGGGTAACGTCAATGAGCAAA			
	ECO_PR	TATTAACTTTACTCCCTTCCTCCCCGCTGAA			
Bacteroidetes	BAC_F	CRAACAGGATTAGATACCCT	Bacchetti De Gregoris et al. (2011)	300	NA
	BAC_R	GGTAAGGTTCCTCGCGTAT			
Eubacteria	EUB_F	ACTCCTACGGGAGGCAGCAGT	Nadkarni et al. (2002)	200	NA
	EUB_R	GTATTACCGCGGCTGCTGGCAC			

TABLE 3 Sensitivity, specificity, and area under the curve (AUC) for faecal calprotectin and relative abundance of Bacteroidetes (BAC), *A. muciniphila* (AKK), *Ruminococcus* spp. (RUM), *E. coli* (ECO), *F. prausnitzii* (FAE) and its two pyhlogroups (PHGI and PHGII), and *M. smithii* (MSM).

Marker	Sensitivity (%)	Specificity (%)	AUC
Faecal calprotectin	100	62.5	0.875
BAC	100	36.4	0.737
AKK	66.7	54.5	0.444
RUM	77.8	44.5	0.586
ECO	77.8	36.4	0.576
FAE	88.9	54.5	0.596
PHGI	66.7	36.4	0.576
PHGII	88.9	54.5	0.626
MSM	66.7	36.4	0.384

Sensitivity, specificity, and positive and negative predictive values of the algorithm designed were calculated using Epidat 3.1 (SERGAS, Xunta de Galicia, Spain). All remaining statistical analyses were performed using the SPSS 23.0 statistical package (IBM, NYC, United States). Significance levels were established as $p = \leq 0.05$.

The final algorithm is based on the decision abundance (DA) calculated using the following equation:

$$DA = \frac{\frac{Ct_{ind}-b_{ind}}{m_{ind}}}{\frac{Ct_{EUB}-b_{EUB}}{m_{EUB}}} \tag{1}$$

Where CT is the threshold cycle; b is the intercept point; m is the slope; ind, is the microbial marker; and EUB is eubacteria (total bacterial load).

3 Results

3.1 Microbial marker profile

Sample A (1 week before surgery) was used to analyse whether microbial markers could predict whether there would be post-surgical recurrence. No significant differences were observed between microbial markers when the two groups (recurrence and no recurrence) were compared (*p*-value >0.05) (Figure 1).
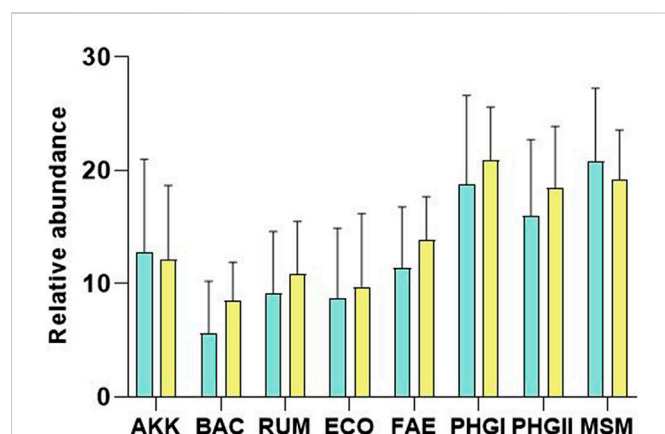
FIGURE 1
Plot of the relative abundances (Microbial marker Ct—Eubacteria Ct) of *A. muciniphila* (AKK), Bacteroidetes (BAC), *Ruminococcus* spp. (RUM), *E. coli* (ECO), *F. prausnitzii* (FAE), *F. prausnitzii* phylogroup I (PHGI), *F. prausnitzii* phylogroup II (PHGII), *M. smithii* (MSM) for remission (green) and recurrence patients (yellow). No significant differences were found (*p*-value>0.05). Note that results are shown in PCR Ct values, so a higher value indicates lower abundance than a lower value.

TABLE 4 Sensitivity, specificity, positive predictive value, and negative predictive value obtained with the algorithm.

Remission vs. Recurrence	Sensitivity (%)
Sensitivity (%)	90.91
Specificity (%)	85.71
Positive predictive value	83.33
Negative predictive value	92.31

Table 3 shows the accuracy of discrimination measured using the area under the ROC curve (AUC) for all microbial markers, with the best sensitivity and specificity results. Calprotectin was the biomarker with the greatest discriminatory capacity and the highest specificity when the remission and the recurrence groups were compared. To prevent DNA quantity bias, relative abundances were calculated by subtracting the value of the total bacterial load (Eubacteria) from the value of each specific marker. The relative abundance of BAC had the highest AUC of microbial markers, with the same sensitivity as calprotectin but less specificity (36.4%). Despite this, the microbial markers were not sufficient to individually discriminate between the two groups of patients.

3.2 Combination of microbial markers to predict post-surgical recurrence

Initially, the specific patient factors were analysed to measure their influence on remission or recurrence. Chi-square tests showed that these factors did not influence the response to our data, with a *p*-value of 0.916 for onset age, 0.821 for disease location, 0.149 for disease phenotype, and 0.087 for patient gender.

Since a single marker could not discriminate between the two target groups with high sensitivity and specificity, we combined

different microbial markers by designing a mathematical algorithm. The algorithm defined had the best predictive capacity to define the surgical prognosis.

Specifically, it combines the relative abundance of three microbial markers (PHGI-EUB, BAC-EUB, and ECO-EUB) where lower counts of PHGI and BAC and higher counts of ECO were associated with recurrence. The algorithm showed a high capacity to discriminate between patients with and without post-surgical recurrence, with a sensitivity and specificity of 90.91% and 85.71%, respectively, and a positive and negative predictive value of 83.33% and 92.31%, respectively (Table 4).

3.3 Evolution of abundance of microbial markers after surgery

The abundance of the microbial markers analysed in stool samples collected 3 and 6 months after surgery were compared with those collected pre-surgery. Patients with remission or relapse post-surgery were analysed separately.

Significant differences were only observed in the postoperative recurrence group, whereas in the remission group the abundance of the markers remained stable over time (Figure 2).

The markers that presented changes in abundance post-surgery were AKK, MSM and ECO; the first two showed a decrease in abundance post-surgery and there was an increase in ECO abundance (all *p*-value = 0.028). As shown in Figure 2, this variation was observed when comparing sample A (pre-surgery) with sample B (3 months later). However, in sample C (6 months post-surgery) the abundance of the markers that showed changes reverted to the levels observed pre-surgery.

Figure 3 shows the effects of surgery on the different markers of patients with post-surgical recurrence in more detail, emphasizing the significant differences observed in the AKK, MSM, and ECO markers from the MANOVA analysis performed.

4 Discussion

Endoscopic recurrence can develop a few days post-surgery and is dependent on the return of the faecal stream and re-establishment of faecal microbiota (Haens et al., 1998). Numerous studies have shown that alterations in the gut microbial profile at the time of surgery or the post-operative follow-up are linked to the post-operative disease course in CD patients (Sokol et al., 2019; Hamilton et al., 2020; Strömbeck et al., 2020; Zhuang et al., 2021). Our results show that, from baseline, there are some differences between the microbial communities of patients with and without relapse. Although these differences were observed in only three of the markers analysed, which may be insufficient to show differences in the microbial ecosystem between groups, we believe that it is sufficient to suggest that patients who relapse have a more unstable and delicate microbiota, since resection affects them more.

The combination of three different species as a predictive tool for post-operative recurrence provides more robust and reliable results. The proposed algorithm combines the relative abundance of three microbial markers (ECO/EUB, PHGI/EUB, and BAC/EUB).

An increase in the abundance *of E. coli* has been associated with the early recurrence of CD in several studies (Neut et al., 2002; Doherty

FIGURE 2
Plot of the relative abundances (Microbial marker Ct—Eubacteria Ct) of *A. muciniphila* (AKK), Bacteroidetes (BAC), *Ruminococcus* spp. (RUM), *E. coli* (ECO), *F. prausnitzii* (FAE), *F. prausnitzii* phylogroup I (PHGI), *F. prausnitzii* phylogroup II (PHGII), *M. smithii* (MSM) for sample A (before surgery), sample B (3 months after surgery) and sample C (6 months after surgery). Note that results are shown in PCR Ct values, so a higher value indicates lower abundance than a lower value.

FIGURE 3
MANOVA plot where the effects of the surgery on the markers are observed. *A. muciniphila* (AKK), Bacteroidetes (BAC), *Ruminococcus* spp. (RUM), *E. coli* (ECO), *F. prausnitzii* (FAE), *F. prausnitzii* phylogroup I (PHGI), *F. prausnitzii* phylogroup II (PHGII), *M. smithii* (MSM) and for sample A (before surgery), sample B (3 months after surgery) and sample C (6 months after surgery).

et al., 2010). As expected, in our study, the counts of *E. coli* in faeces after surgery increased in patients with recurrence. Similarly, whereas we did not observe differences in the abundance of *F. prausnitzii* or its two phylogroups, lower counts of *F. prausnitzii* phylogroup I were related to recurrence in the algorithm designed. *F. prausnitzii* is one of the most abundant commensal butyrate-producing bacteria in the microbiota of healthy subjects, and its prevalence and abundance has been shown to be reduced in IBD patients (Sokol et al., 2008; Miquel

et al., 2013; Lopez-Siles et al., 2014b). Butyrate is a short-chain fatty acid (SCFA), which are volatile fatty acids produced by the gut microbiota in the large bowel as a fermentation product from undigested components. The most abundant SCFA are acetic, propionic, and butyric acid (Ríos-Covián et al., 2016). Our study highlights that, among the benefits of butyrate, a strong effect is produced in IBD, probably as a consequence of mucosal healing and inhibition of inflammation (Russo et al., 2019). This confirms that the clearance abundance of these SCFA-producing bacteria is correlated with a greater predisposition to relapse. Moreover, other studies have found that a low level of SCFA-producing bacteria may be associated with postoperative endoscopic recurrence (Mondot et al., 2016; Wright et al., 2017; Sokol et al., 2019).

In addition, a depleted abundance of Bacteroidetes phylum was paired with endoscopic recurrence in the algorithm designed, which fits with its contribution to gut health. Other studies have suggested that Bacteroidetes species contribute to the homeostasis of the gut microbiota, synthetizing compounds with immunomodulatory properties such as linoleic acid, and that, by improving its diversity and composition, the wellbeing of IBD patients is enhanced (Lee et al., 2013; Delday et al., 2019; Nomura et al., 2021).

The algorithm defined has 90.91% sensitivity and 85.71% specificity, which could lead to the identification of patients prone to recurrence post-surgery, permitting treatment to be personalised, which would increase patients' quality of life and have benefits related to healthcare costs. This would be an improvement in IBD management since there is currently no test that allows the effectiveness of intestinal resection to be predicted.

With respect to the results observed on the effect of surgery on the intestinal microbiota, we found that the counts of the acetate short-chain fatty acid (SCFA)-producing bacteria such as *A. muciniphila* were reduced. Acetate is the most prominent SCFA and substrate for butyrate production and is reported to have effects on lipid metabolism, such as lipogenesis and cholesterogenesis (Morrison et al., 2016). Similarly, the abundance of *Methanobrevibacter smithii* was reduced 3 months post-surgery in patients prone to recurrence, which was reverted at 6 months. *Methanobrevibacter smithii* is an archaeon that converts CO_2 and H_2 into CH_4 and also produces SCFA from carbohydrates; its prevalence in stool samples from IBD patients is significantly lower than in healthy control subjects (Bang et al., 2014). In addition, methane levels are lower in IBD patients than in healthy subjects (Chaudhary et al., 2018).

Our results show intestinal surgery changes the abundance of some of the microbial markers in patients with post-surgical recurrence, but the changes were reverted at 6 months. In contrast, in patients who go into remission, the abundance of no marker was altered, demonstrating a much more stable microbiota from baseline.

The study had some limitations. Despite the promising results obtained, the study was performed in a small cohort and more robust examination in further large-scale studies is needed to validate the algorithm defined prior to its application in clinical practice. Secondly, all the results are based on faecal microbiota and neither data from ileal mucosa-associated microbiota was available nor was a metagenomic analysis performed. However, our markers were initially determined from intestinal biopsies from mucosa. The quantification of these markers was further optimized in stool samples, avoiding one of the drawbacks of clinical microbiology applied to stool markers, which is their vulnerability to dietary changes in different populations (Lopez-Siles et al., 2017). Thirdly,

clinical risk factors such as smoking and the surgical margins inflammation, both important in predicting recurrence were not analysed. Despite this, the objective of our study was to define an algorithm that worked independently of these factors facilitating its use. All these limitations could be addressed with a larger prospective study, which we plan to carry out in the future.

In conclusion, this study shows that surgery has an impact on the microbiota and that postoperative endoscopic recurrence is associated with changes in its composition and abundance. Precision medicine and biomarker-guided surgery prognosis are a necessary step in advancing the clinical effectiveness and improving the quality of life. Hence, the gut microbiota has the potential to become a precise tool for predicting the postoperative evolution and recurrence, which would allow, through microbial therapies, redirection of the microbial profile, and the avoidance of postoperative recurrence.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Comitè d'Ètica de Investigació de l'Hospital Universitari Doctor Josep Trueta de Girona

Author contributions

LO: Study design, analysis, investigation, writing—original draft, and visualization; BC: Patient recruitment, study design, and writing review; DJ-B: Patient recruitment and study design; JA: Study design and analysis; SR-P: Study design, analysis, investigation, writing—review and editing, and visualization; MM: Writing—review and editing; AB: Patient recruitment; PT: Patient recruitment; ED: Patient recruitment; JG: Patient recruitment; MS-P: Conceptualization, study design, analysis, writing—review and editing, and visualization; JG-G: Conceptualization, study design, analysis, writing—review and editing, and visualization; XA: Conceptualization, study design, analysis, writing—review and editing, and visualization.

Acknowledgments

Conflict of interest

GoodGut S.L.U shares the patent PCT/EP 2016/069188 method for the quantification of *Faecalibacterium prausnitzii* phylogroup I and/or phylogroup II and the use thereof as biomarkers, referenced in this study.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ayling, R. M., and Kok, K. (2018). *Fecal Calprotectin*, 87. Amsterdam, Netherlands: Elsevier.

Bacchetti De Gregoris, T., Aldred, N., Clare, A. S., and Burgess, J. G. (2011). Improvement of phylum- and class-specific primers for real-time PCR quantification of bacterial taxa. *J. Microbiol. Methods* 86 (3), 351–356. doi:10.1016/j.mimet.2011.06.010

Bang, C., Weidenbach, K., Gutsmann, T., Heine, H., and Schmitz, R. A. (2014). The intestinal archaea Methanosphaera stadtmanae and Methanobrevibacter smithii activate human dendritic cells. *PLoS One* 9 (6), e99411–e99419. doi:10.1371/journal.pone.0099411

Baumgart, D. C., and Sandborn, W. J. (2012). Crohn's disease. *Lancet* 380 (9853), 1590–1605. doi:10.1016/S0140-6736(12)60026-9

Boschetti, G., Laidet, M., Moussata, D., Stefanescu, C., Roblin, X., Phelip, G., et al. (2015). Levels of fecal calprotectin are associated with the severity of postoperative endoscopic recurrence in asymptomatic patients with Crohn's disease. *Am. J. Gastroenterology* 110 (6), 865–872. doi:10.1038/ajg.2015.30

Chaudhary, P. P., Conway, P. L., and Schlundt, J. (2018). Methanogens in humans: Potentially beneficial or harmful for health. *Appl. Microbiol. Biotechnol.* 102 (7), 3095–3104. doi:10.1007/s00253-018-8871-2

Collado, M. C., Derrien, M., Isolauri, E., De Vos, W. M., and Salminen, S. (2007). Intestinal integrity and Akkermansia muciniphila, a mucin-degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Appl. Environ. Microbiol.* 73 (23), 7767–7770. doi:10.1128/AEM.01477-07

De Cruz, P., Kamm, M. A., Hamilton, A. L., Ritchie, K. J., Krejany, E. O., Gorelik, A., et al. (2015). Crohn's disease management after intestinal resection: A randomised trial. *Lancet* 385 (9976), 1406–1417. doi:10.1016/S0140-6736(14)61908-5

Delday, M., Mulder, I., Logan, E. T., and Grant, G. (2019). Bacteroides thetaiotaomicron ameliorates colon inflammation in preclinical models of Crohn's disease. *Inflamm. Bowel Dis.* 25 (1), 85–96. doi:10.1093/ibd/izy281

Derrien, M., Belzer, C., and de Vos, W. M. (2017). Akkermansia muciniphila and its role in regulating host functions. *Microb. Pathog.* 106, 171–181. doi:10.1016/j.micpath.2016.02.005

Ding, N. S., Hart, A., and De Cruz, P. (2016). Systematic review: Predicting and optimising response to anti-TNF therapy in Crohn's disease - algorithm for practical management. *Aliment. Pharmacol. Ther.* 43 (1), 30–51. doi:10.1111/apt.13445

Doherty, G. A., Bennett, G. C., Cheifetz, A. S., and Moss, A. C. (2010). Meta-analysis: Targeting the intestinal microbiota in prophylaxis for post-operative Crohn's disease. *Aliment. Pharmacol. Ther.* 31 (8), 802–809. doi:10.1111/j.1365-2036.2010.04231.x

Gklavas, A., Dellaportas, D., and Papaconstantinou, I. (2017). Risk factors for postoperative recurrence of Crohn's disease with emphasis on surgical predictors. *Ann. Gastroenterol.* 30 (6), 598–612. doi:10.20524/aog.2017.0195

Haens, G. R. D., Geboes, K., Peeters, M., Baert, F., Penninckx, F., and Rutgeerts, P. (1998). Early lesions of recurrent Crohn's disease caused by infusion of intestinal contents in excluded ileum. *Gastroenterology* 114, 262–267. doi:10.1016/s0016-5085(98)70476-7

Hamilton, A. L., Kamm, M. A., De Cruz, P., Wright, E. K., Feng, H., Wagner, J., et al. (2020). Luminal microbiota related to Crohn's disease recurrence after surgery. *Gut Microbes* 11 (6), 1713–1728. doi:10.1080/19490976.2020.1778262

Hanauer, S. B. (2019). Advanced therapy for inflammatory bowel disease. *J Am Board Fam Med* 15. doi:10.3122/jabfm.2014.03.130224

Huijsdens, X. W., Linskens, R. K., Mak, M., Meuwissen, S. G., Vandenbroucke-Grauls, C. M., and Savelkoul, P. H. (2002). Quantification of bacteria adherent to gastrointestinal mucosa by real-time PCR. *J. Clin. Microbiol.* 40 (12), 4423–4427. doi:10.1128/JCM.40.12.4423-4427.2002

Joossens, M., Huys, G., Cnockaert, M., Preter, V. De., Verbeke, K., Rutgeerts, P., et al. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives *Gut*. 60 (5).631–637. doi:10.1136/gut.2010.223263

la Reau, A. J., and Suen, G. (2018). The ruminococci: Key symbionts of the gut ecosystem. *J. Microbiol.* 56, 199–208. doi:10.1007/s12275-018-8024-4

Lee, S. M., Donaldson, G. P., Mikulski, Z., Boyajian, S., Ley, K., and Mazmanian, S. K. (2013). Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* 501 (7467), 426–429. doi:10.1038/nature12447

Lopez-Siles, M., Garcia-Gil, J., Aldeguer-Manté, X., and Martínez Medina, M. (2017). *PCT/EP20 16/069 188 Method for the quantification of faecalibacterium prausnitzii phylogroup I and/or phylogroup II members and the use thereof as biomarkers.* WO2017025617A1.

Lopez-Siles, M., Martinez-Medina, M., Busquets, D., Sabat-Mir, M., Duncan, S. H., Flint, H. J., et al. (2014). Mucosa-associated Faecalibacterium prausnitzii and *Escherichia coli* co-abundance can distinguish irritable bowel syndrome and inflammatory bowel disease phenotypes. *Int. J. Med. Microbiol.* 304 (3–4), 464–475. doi:10.1016/j.ijmm.2014.02.009

Lopez-Siles, M., Martinez-Medina, M., Busquets, D., Sabat-Mir, M., H Duncan, S., Flint, H., et al. (2014). Mucosa-associated Faecalibacterium prausnitzii and *Escherichia coli* co-abundance can distinguish irritable bowel syndrome and inflammatory bowel disease phenotypes. *Int. J. Med. Microbiol.* 304, 464–475. doi:10.1016/j.ijmm.2014.02.009

Lopez-Siles, M., Martinez-medina, M., Surís-Valls, R., Aldeguer, X., Sabat-Mir, M., Duncan, S. H., et al. (2016). Changes in the abundance of Faecalibacterium prausnitzii phylogroups I and II in the intestinal mucosa of inflammatory bowel disease and patients with colorectal cancer. *Colorectal Cancer* 22 (1), 28–41. doi:10.1097/MIB.0000000000000590

Macfarlane, G. T., Blackett, K. L., Nakayama, T., Steed, H., and Macfarlane, S. (2009). The gut microbiota in inflammatory bowel disease. *Curr. Pharm. Des.* 15, 1528–1536. doi:10.2174/138161209788168146

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., PellEtiEr, E., FrangeuL, L., et al. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 55 (2), 205–211. doi:10.1136/gut.2005.073817

Manichanh, C., Borruel, N., Casellas, F., and Guarner, F. (2012). The gut microbiota in IBD. *Nat. Rev.* 9, 599–608. doi:10.1038/nrgastro.2012.152

Miquel, S., Martín, R., Rossi, O., Bermúdez-Humarán, L. G., Chatel, J. M., Sokol, H., et al. (2013). Faecalibacterium prausnitzii and human intestinal health. *Curr. Opin. Microbiol.* 16, 255–261. doi:10.1016/j.mib.2013.06.003

Mondot, S., Lepage, P., Seksik, P., Allez, M., Tréton, X., Bouhnik, Y., et al. (2016). Structural robustness of the gut mucosal microbiota is associated with Crohn's disease remission after surgery. *Gut* 65 (6), 954–962. doi:10.1136/gutjnl-2015-309184

Morrison, D. J., Preston, T., Morrison, D. J., and Preston, T. (2016). Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* 7 (3), 189–200. doi:10.1080/19490976.2015.1134082

Nadkarni, M. A., Martin, F. E., Jacques, N. A., and Hunter, N. (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology (Reading).* 148(1), 257–266. doi:10.1099/00221287-148-1-257

Neut, C., Bulois, P., Desreumaux, P., Membré, J. M., Lederman, E., Gambiez, L., et al. (2002). Changes in the bacterial flora of the neoterminal ileum after ileocolonic resection for Crohn's disease. *Am. J. Gastroenterology* 97 (4), 939–946. doi:10.1111/j.1572-0241.2002.05613.x

Nomura, K., Ishikawa, D., Okahara, K., Ito, S., Haga, K., and Takahashi, M. (2021). Bacteroidetes species are correlated with disease activity in ulcerative colitis. *J Clin Med* 10 (8), 1749. doi:10.3390/jcm10081749

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., et al. (2017). A microbial signature for Crohn's disease. *Gut* 66 (5), 813–822. doi:10.1136/gutjnl-2016-313235

Ramió-Pujol, S., Amoedo, J., Serra-Pagès, M., Torrealba, L., Bahí, A., Serrano, M., et al. (2020). A novel distinctive form of identification for differential diagnosis of irritable bowel syndrome, inflammatory bowel disease, and healthy controls. *GastroHep* 2 (5), 193–204. doi:10.1002/ygh2.417

Ramirez-Farias, C., Slezak, K., Duncan, A., Holtrop, G., and Louis, P. (2009). Effect of inulin on the human gut: Stimulation of bifidobacterium adolescentis and faecalibacterium prausnitzii. *Br J Nutr.* 101, 541–550. doi:10.1017/S0007114508019880

Ríos-Covián, D., Ruas-Madiedo, P., Margolles, A., Gueimonde, M., De los Reyes-Gavilán, C. G., and Salazar, N. (2016). Intestinal short chain fatty acids and their link with diet and human health. *Front. Microbiol.* 7 (FEB), 185–189. doi:10.3389/fmicb.2016.00185

Russo, E., Giudici, F., Fiorindi, C., Ficari, F., and Scaringi, S. (2019). Immunomodulating activity and therapeutic effects of short chain fatty acids and tryptophan post-biotics in inflammatory bowel disease. *Front Immunol.* 22 (10), 2754. doi:10.3389/fimmu.2019.02754

Rutgeerts, P., KeRRemans, R., and Hiele, M. (1990). Predictability of the postoperative course of Crohn's disease. *Gastroenterology* 99 (4), 956–963. doi:10.1016/0016-5085(90)90613-6

Sokol, H., Brot, L., Stefanescu, C., Auzolle, C., Barnich, N., Buisson, A., et al. (2019). Prominence of ileal mucosa-associated microbiota to predict postoperative endoscopic recurrence in Crohn's disease. *Gut* 69 (34), 462–472. doi:10.1136/gutjnl-2019-318719

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L. G., Gratadoux, J-J., et al. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci.* 105 (43), 16731–16736. doi:10.1073/pnas.0804812105

Stříž, I., and Trebichavský, I. (2004). Calprotectin - a pleiotropic molecule in acute and chronic inflammation. *Physiol. Res.* 53 (3), 245–253. doi:10.33549/physiolres.930448

Strömbeck, A., Lasson, A., Strid, H., Sundin, J., Stotzer, P. O., Simrén, M., et al. (2020). Fecal microbiota composition is linked to the postoperative disease course in patients with Crohn's disease. *BMC Gastroenterol.* 20 (1), 130–210. doi:10.1186/s12876-020-01281-4

Torres, J., Mehandru, S., and Colombel, J. F. (2017). Crohn's disease. *Lancet* 389, 1741–1755. doi:10.1016/S0140-6736(16)31711-1

Watterlot, L., Lakhdari, O., Bermu, L. G., Sokol, H., Bridonneau, C., Furet, J., et al. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Natl. Acad. Sci. U. S. A.* 105 (43), 16731–16736. doi:10.1073/pnas.0804812105

Wright, E. K., Kamm, M. A., Wagner, J., Teo, S. M., Cruz, P. De., Hamilton, A. L., et al. (2017). Microbial factors associated with postoperative Crohn's disease recurrence. *J. Crohns Colitis* 11 (2), 191–203. doi:10.1093/ecco-jcc/jjw136

Zhuang, X., Tian, Z., Li, N., Mao, R., Li, X., Zhao, M., et al. (2021). Gut microbiota profiles and microbial-based therapies in post-operative Crohn's disease: A systematic review. *Front. Med. (Lausanne)* 7, 615858. doi:10.3389/fmed.2020.615858

Cancer immunotherapies: A hope for the uncurable?

Firas Hamdan[1,2,3] and Vincenzo Cerullo[1,2,3,4,5]*

[1]Laboratory of Immunovirotherapy, Drug Research Program, Faculty of Pharmacy, University of Helsinki, Helsinki, Finland, [2]TRIMM, Translational Immunology Research Program, University of Helsinki, Helsinki, Finland, [3]Drug Delivery, Drug Research Program, Division of Pharmaceutical Biosciences, Faculty of Pharmacy, University of Helsinki, Helsinki, Finland, [4]iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki, Helsinki, Finland, [5]Department of Molecular Medicine and Medical Biotechnology and CEINGE, Naples University Federico II, Naples, Italy

The use of cancer immunotherapies is not novel but has been used over the decades in the clinic. Only recently have we found the true potential of stimulating an anti-tumor response after the breakthrough of checkpoint inhibitors. Cancer immunotherapies have become the first line treatment for many malignancies at various stages. Nevertheless, the clinical results in terms of overall survival and progression free survival were not as anticipated. Majority of cancer patients do not respond to immunotherapies and the reasons differ. Hence, further improvements for cancer immunotherapies are crucially needed. In the review, we will discuss various forms of cancer immunotherapies that are being tested or already in the clinic. Moreover, we also highlight future directions to improve such therapies.

KEYWORDS

cancer, immunotherapy, oncolytic virotherapy, antibodies, syngergistic effects

1 Introduction

The pronunciation of the word cancer may alarm the majority, if not all, of the people. This is simply because of its serious health effects which has brought it to be the second leading cause of death, after cardiac conditions (Heron, 2019; Siegel et al., 2020). Scientific surveys from the United Kingdom in 2019 have observed that the public's main support for scientific research concerned cancer research (for Business and Strategy, 2020). Analysts perceive this wide support towards the subject as an eager awaited miracle drug that could cure cancer. Yet, as our understanding of cancer grows the discovery of such miracle drug seems to be lost hope.

Cancers have shown to have a dense complexity regarding its biology. In 2000, Hanahan and Weinberg published one of the most cited reviews in the cancer biology field describing the hallmarks of cancer (Hanahan and Weinberg, 2000). During those times six hallmarks were described to oversimplify the needed characteristics for a successful tumor growth. Based on such hallmarks, traditional therapies were developed ranging from radiation, chemotherapy, and targeted therapies.

To this day, surgery, radiation and chemotherapy represent one of the most used therapies for the treatment of cancer (Harrington and Smith, 2008). The main mode of action of these therapies is the direct tumor cell lysis *via* different mechanisms. This may be perceived as old fashioned since this mode of action has seen not to be that effective in eradicating full tumors. Moreover, these therapies have a growing list of serious side effects detrimental to patients. In spite of this and as our understanding of cancer biology grew, targeted therapies took the stage since they were more effective and safer (Baudino, 2015). Targeted therapies consist of molecules that exploit certain biological difference among

healthy and cancer cells allowing for selectively targeting. Nevertheless, advancements in genomic sequencing have demonstrated that cancer is not a monogenic disease yet a complex and heterogenous disease (Meldrum et al., 2011). This explains why the use of targeted therapies targeting single molecules have not shown an overwhelming success as once expected.

After a decade from the Hanahan and Weinberg review, the authors updated the list of hallmarks by adding two new hallmarks: reprogramming energy metabolism and evading immune responses (Hanahan and Weinberg, 2011a). The latter hallmark consists of a crucial interplay where the immune system has the capabilities to recognize and kill cancer cells. Hence, tumor cells have developed multiple strategies to overcome immune recognition. Also, tumor cells have shown to induce a tumor-promoting inflammation. As a result of this, a novel era of treatments was developed enhancing the immune system to recognize cancer. In 2013, the world-renowned science journal, *Science*, dubbed such treatment as "breakthrough of the year" due to the significant impact in the clinic and in 2018 were the theme for the Nobel prize in physiology and medicine (Couzin-Frankel, 2013). These treatments offer the use of the patient's own immune system to induce anti-tumor response able to sustain a long-lasting anti-tumor killing.

As successful as immunotherapies have been described, looking at the statistical numbers a very small minority (20%–40%) of people do benefit from them (Sharma et al., 2017). Multiple reasons have been formulated over the years concerning why most of the patients do not respond. Nevertheless, a need to improve such treatments is required. This review will provide the current state of immunotherapies and provide novel strategies to further such field.

2 Cancer immunotherapies

Our immune system has a significant role in keeping the integrity of our health. Besides its obvious role in protecting against pathogens, it has a more unobtrusive but highly crucial role in cancer prevention and defence. Already in 1909, Paul Ehrlich postulated that the power of the immune system may be harnessed to control cancer. It was proposed that immune cells are constantly surveilling cells throughout the body, able to recognise and eliminate incipient cancer cells and therefore halt the production of nascent tumours (Hanahan and Weinberg, 2011b). This was validated by striking results where immunocompromised individuals had an increased risk in developing certain cancers (Vajdic and van Leeuwen, 2009). Furthermore, mice models with defective T cells and NK cells were shown to be more susceptible to cancer (Smyth et al., 2000). However, according to such logic, tumours that appear and progress in otherwise healthy individuals should be able to somehow resist or evade elimination by the immune system. Further research indeed indicated that the tumour microenvironment was immunosuppressive and cancer cells are able to develop multiple immune evasion strategies (Vinay et al., 2015). In spite of this, boosting the immune system has been the major target for drug development in the treatment of cancers for the past decade.

Our immune system has a well-known ability to distinguish between self and non-self, especially in the case of infection or malignancy. This process is called immune surveillance and it is crucial in eliminating hundreds of newly formed malignant cells daily. Nevertheless, other than serving as a tumor suppressor the immune system can also shape the tumor immunogenicity in a process called cancer immunoediting. This process is divided into three parts; elimination, equilibrium and escape (Dunn et al., 2004). The first phase consists of a dynamic process in which immune cells recognize tumor cells expressing immunogenic antigens (Kim et al., 2007). This then allows the immune cells to recognize and kill tumor cells. However, not all of tumor cells are immunogenic leading them to not be recognized by immune cells. This adds a bottle neck pressure inducing a positive selection of tumor cells with reduced immunogenicity. These cells then enter the final stage of escape since they are unharmed by the immune system and can proliferate uncontrollably (Dunn et al., 2006; Kim et al., 2007; Swann and Smyth, 2007).

In spite of this, the main objective for cancer immunotherapies is to redirect the immune system towards these cells with reduced immunogenicity (Seliger, 2005). The current cancer immunotherapies in the clinic can be divided into two groups based on mechanism of actions: passive or active immunotherapies. Active immunotherapies involve the direct activation of a tumor-specific immune response. While passive immunotherapies are molecules that are given to patients that cannot either be induced, lowly expressed or non-functioning.

2.1 Passive immunotherapies

In many patients, the ability to induce a proper anti-tumor immune response is hindered by factors of immunosuppression. Thus, passive immunotherapies try to overcome such limitation by fighting cancer directly. These molecules endow intrinsic antitumoral activity and can indirectly or directly target tumor cells. In this section these types of molecules will be further explained.

2.1.1 Cytokines

Cytokines are small molecules expressed by both inflammatory and non-inflammatory cells to coordinate inflammation and other immune responses. In cancer, these molecules have been administered to patients in order to stimulate anti-cancer immune responses in an un-specific way. Two main cytokines that will be discussed are interleukin-2 (IL-2) and granulocyte and macrophage colony stimulating factor (GM-CSF).

IL-2 is a pluripotent cytokine able to stimulate the immune system in many ways. However, one of its crucial roles is in the activation of both natural killer (NK) cells20 and T-cells (Gillis and Smith, 1977; Gillis et al., 1978; Smith, 1988; Lehmann et al., 2001). In specific, high levels of Il-2 can induce T cell expansion and activation for interferon gamma (IFN-gamma) production (Paliard et al., 1988). A recombinant form of Il-2, marketed as Proleukin®, has received FDA approval for the use in metastatic renal cell carcinoma (RCC) (Fyfe et al., 1995) and metastatic melanoma (Amaria et al., 2015). Clinical data has shown that from 270 metastatic melanoma patients, 16% of patients showed objective responses while 6% showed complete response (Davar et al., 2017). Similar results also were seen in metastatic RCC, objective responses were seen in 15% of patients and 8% of patients showed complete responses (Achkar et al., 2017). Moreover, a clear increase in NK and T cell

activation was observed in most of the treated patients. Hence, currently in the clinic Proleukin® is still being tested with other potential synergistic molecules to further improve clinical responses. Nevertheless, systemic administration of IL-2 has been associated with several life-threatening toxicities due to an increased inflammation (Pachella et al., 2015). In spite of this, several strategies are being developed to ensure a targeted release in the tumor microenvironment.

One other widely used cytokine in the clinic is GM-CSF. Compared to IL-2, GM-CSF works with other types of cells in specific APCs. For example, mice defective in GM-CSF had a decrease proliferation and maturation of dendritic cells (DC) and macrophages leading to an increase vulnerability to bacterial infections (Stanley et al., 1994). In 2005, Kurbacher et al. treated 19 cancer patients suffering from breast and female reproductive tract carcinomas with recombinant GM-CSF (Kurbacher et al., 2005). Only one patient had a complete response while six others had a partial response. The main mode of action was shown to be attributed to the activation of DCs and increased antigen presentation. A recombinant protein of GM-CSF (called sagramostim) showed a 100% overall response rate with chronic lymphocytic leukemia patients when combined with chemotherapy. Yet, with chronic myeloid leukemia it was discontinued in all patients due to severe adverse events.

2.1.2 Adoptive cell therapy (ACT)

One other form of passive cancer immunotherapy may come in the form of infusing activated immune cells into patients. This is called adoptive cell therapy (Rosenberg et al., 2008; Perica et al., 2015) and can be divided into two subtypes: adoptive tumor-infiltrating lymphocytes (TILs) and genetically engineered T cells expressing specific T cell receptors (TCRs) or chimeric antigen receptors (CARs). Both these therapies share a common step which is the preconditioning lymphodepletion regimen before treatment. With the use of cyclophosphamide, patients undergo lymphopenia and neutropenia in order to prevent such endogenous cells from attacking injected activated immune cells (Proietti et al., 1998). Moreover, these therapies use the patient's own lymphocytes.

TIL therapy is not a novel form of treatment but can be dated back to 1994 being used in metastatic melanoma (Rosenberg et al., 1994). This therapy consists of isolating tumor-specific T cells within the tumor microenvironment and further expand them ex-vivo. Various regimens for expansion have been described, but the most common is the use of high doses of IL-2. Currently such therapy has been approved by the FDA for metastatic melanoma. Yet, multiple clinical trials are undergoing with different type of expansion regimens or in combination with other treatments.

Other than TILs, scientists have tried to increase the armamentarium of T-cells by genetically engineering them to express specific TCRs or CARs (Rosenberg and Restifo, 2015). In both cases, T cells are first isolated through leukapheresis using peripheral blood. Once T cells have been isolated, using a lentiviral vector a transduction is performed to facilitate expression of a TCR or CAR. Following transductions, these cells are then expanded using high doses of IL-2 and are then ready to be re-infused in patients.

Other than the structure, TCRs and CARs give T cells a different way of killing tumor cells. TCRs are made of αβ heterodimers with each chain consisting of variable and constant region domains.

These receptors associate with CD3 in the surface membrane and recognize major histocompatibility complexes (MHC) loaded with an antigen which induces activation and killing of the tumor cell (Harris and Kranz, 2016; June et al., 2018). As for CARs, these receptors consist of an antigen binding domain, consisting of single-chain variable fragment (scFv) from an antibody, which is connected to an intracellular signaling domain causing cell activation (June et al., 2018). Thus, killing of a tumor cell occurs in a MHC-independent fashion where once the scFv portion of the receptor binds to its specific epitope it triggers T cell activation via its signaling domain (June et al., 2018). This is a clear advantage over the conventional TCR killing since one of the most prominent immune-escape mechanisms a tumor poses the downregulation of MHC from its surface (Blankenstein et al., 2012). Yet, TCRs can recognize intracellular tumor antigens presented by MHC molecules while CAR receptors can only recognized membrane-bound tumor antigens. Moreover, CAR-based T cells are more toxic than TCR based (Restifo et al., 2012). Both forms of treatments can cause neurotoxicity and cytokine release syndrome (CRS), yet with CAR-based T cell therapy it is more severe (Restifo et al., 2012). Also, Currently, six different CAR-T cell therapies have been approved in the clinic. These CAR-T cells consists of a CAR receptors targeting CD19 or BCMA for B-cell based malignancies or multiple myeloma, respectively. As for solid-tumors, CAR-T cells have not managed to obtain clinical approval due to immunosuppressive tumor microenvironment, poor infiltration, poor tumor penetration and other reasons.

2.1.3 Antibody therapy

Antibody therapy in cancer has been one of the most successful types of therapies used in the clinic to treat hematological and solid tumors. The advancement of antibodies in cancer therapy can date back to 1890 when first described as neutralizing substances against diphtheria (Behring, 2013). It was later seen that these substances had a specific property in recognizing specific epitopes and were secreted by our own cells, in specific plasma B cells (Fagraeus, 1947; Van Epps, 2005). It was then hypothesized that each plasma B cell clone was able to produce one specific antibody (Nossal and Lederberg, 1958). This concept led to the isolation of individual plasma B cell clones in order to obtain monoclonal antibodies (Schwaber and Cohen, 1973; Köhler and Milstein, 1975). This technology allowed for the screening of thousands of monoclonal antibodies to identify high-affinity monoclonal antibodies against any desired tumor-associated antigen. The first clinical trial using an antibody began in 1980 for lymphoma patients (Koprowski et al., 1978; Nadler et al., 1980). Sadly, such antibodies provided poor clinical efficacy since such antibodies were murine and induced a human anti-mouse antibody (HAMA) response. Therefore, to optimize antibody therapy techniques to humanize antibodies originating from mice were developed. These techniques included cloning the murine derived variable chains (chimeric) or complementary determining regions (humanized) into human antibody formats. Recent techniques have now allowed for the generation of full human antibodies by using transgenic mice or yeast-phage display (Riechmann et al., 1988; Nelson et al., 2010). As a result of such advancements, antibody therapy has become the most type of drug sold for pharmaceutical purposes.
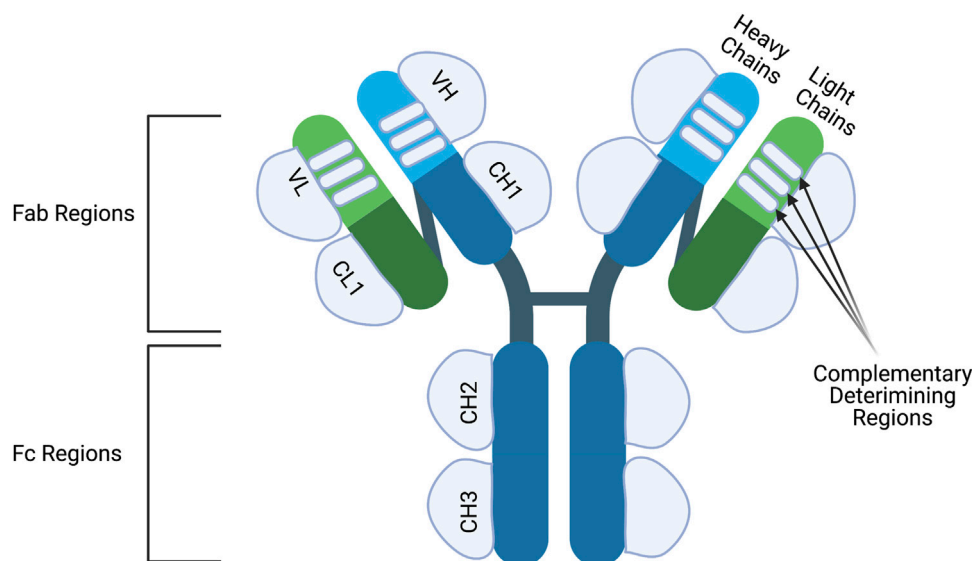
FIGURE 1
The structure of an antibody. Antibodies are made up of two identical heavy chains and light chains. The heavy chains are connected to each other via disulfide bridge and the light chains are connected to the upper part of the heavy chains. Both heavy and light chains consist of variable (V) and constant regions (C). The heavy chain contains three constant domains (CH1-CH3) and one variable domain (VH) while the light chain has one constant domain (CL) and one variable domain (VL). Moreover, the variable chains have three complementary determining regions which dictate the specificity of the antibody. Antibodies can also be classified into two structures; the Fc and Fab region. The Fc regions contains the CH2 and CH3 domains and is important to elicit Fc-effector mechanisms. The Fab regions compromise of CH1, VH, CL1 and VL regions and are important for epitope binding. Figures were created with BioRender.com.

Antibodies are able to directly kill target cells by disrupting or activating receptor signaling. This activity is pertinent to the Fab regions of an antibody which are responsible for binding. However other than direct cell killing, antibodies are also able to orchestrate host-immune response to induced immune-mediated cell death. This dual mechanism of action has made antibody therapy powerful and safe compared to other conventional therapies. This section will describe the structure of an antibody and its use in cancer therapy.

2.1.3.1 Antibody structure

Antibodies are large structures made up of four polypeptides, two heavy and two light chains, joined together *via* disulfide bonds to give a "Y" shaped structure (Figure 1). Both heavy and light chains are made up of two regions: the variable and constant domains. Light chains consist of one variable domain (V_L) and one constant domain (C_L) while heavy chains compromise of one variable domain (V_H) and four constant domains (C_H1, C_H2, C_H3 and C_H4). Furthermore, based on structure, antibodies can be classified into two Fab (Fragment antigen-binding) regions and an Fc (Fragment crystallizable) region. The Fab region compromises of the full light chain and part of the heavy chain (V_H and C_H1) which give the tips of the "Y" shape. The rest of the constant heavy chain domains make up the Fc region which forms the stalk of the "Y" shape.

The variable regions of both the heavy and light chain are then subdivided into four framework regions and three hypervariable regions. The amino-acid composition of the hypervariable regions is the most varied from antibody-to-antibody. Once these regions fold into three β-strands they are then referred to as complementary-

determining regions (CDR) since the shape complements the targeted epitope. The CDRs from the heavy and light chains determine the antibody-binding side but the framework regions also play a minor role. As for the Fc region, it compromises of C_H2-C_H4 and has a vital role for modulating immune cell activity. Immune effector cells can bind to the Fc-region of an antibody through the Fc-receptors subsequently activating effector functions.

2.1.3.2 Heavy and light chains of antibodies

In mammals, two types of light chains of an antibody exist called lambda and kappa. No functional differences have been described for both these chains which are used to build an antibody complex. However, the antibody complex contains two identical light chains, and no mix of kappa and lambda chains usually occurs within one antibody. The proportions of each chain used varies among species and can serve as markers of abnormal proliferation of B cell clones (Hershberg et al., 2015).

As for heavy chains, in mammals there exists five different chains called alpha, gamma, delta, epsilon and micro which give rise to five different antibody classes such as IgA, IgG, IgD, IgE and IgM, respectively. Contrary to light chains, the antibody classes differ in many functional activities, biological properties, and location. This is mostly due to the differential binding of different Fc-receptors since IgA, IgG or IgE bind to Fc-α, Fc-γ, or Fc-ε receptor respectively. These receptors are distinguished based on what type of immune cells express them and signaling properties, explaining the antibody class functions (Figure 2). For example, Fc-ε receptors are found on eosinophils, mast cells and basophils explaining the role of such receptors in allergic responses.
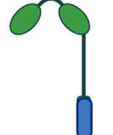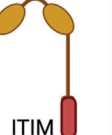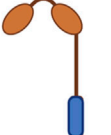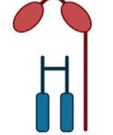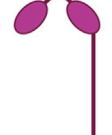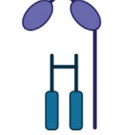
FIGURE 2
The distrubion, structure and function of Fc receptors. The different types of Fc-g and Fc-a receptors found in humans. All activating Fc receptors contain an immunoreceptor tyrosine-basedactivation motif (ITAM) while inhibitory Fc receptors have an immunoreceptor tyrosine-basedinhibitory motif (ITAM) or none. Figures were created with BioRender.com.

2.1.3.3 Direct tumor killing using monoclonal antibodies

As previously mentioned, cancer antibodies have a dual mechanism of action consisting of either direct killing or inducing immune-mediated cell death of opsonized cancer cells. Cancer cells heavily depend on pro-tumor growth and survival signaling provided by different growth factor receptors. Antibodies can perturb such signaling by manipulating the activation or blocking ligand binding subsequently leading to cell death. An example of ligand blocking is the clinically approved Cetuximab, a monoclonal antibody binding to epidermal growth factor receptor (EGFR). EGFR is highly overexpressed on many different types of cancers and when activated can induce proliferation, migration, and invasion of tumor cells (Downward et al., 1984; Gusterson et al., 1984; Ullrich et al., 1984; Cowley et al., 1986). Cetuximab binding to EGFR has been seen to disrupt ligand binding and consequently lead to apoptosis of tumor cells (Li et al., 2005; Patel et al., 2009). Human epidermal growth factor receptor 2 (HER2) is another growth receptor overexpressed by tumor cells to sustain proliferation (Slamon et al., 1989). In specific, HER2 overexpression has been highly seen in ovarian and breast cancer (Slamon et al., 1989). Unlike EGFR, HER2 has no known ligand and is activated by heterodimerization to other growth receptors (Chen et al., 2003). Trastuzumab, a clinically approved monoclonal antibody against HER2, has been shown to disrupt this heterodimerization, consequently leading to tumor cell death (Plosker and Keam, 2006). Trastuzumab has been a clinical success in in treating HER2+ breast cancer patients (Plosker and Keam, 2006).

2.1.3.4 Complement-dependent cytotoxicity

Antibodies can interact with the complement system through the Fc-region to activate the classical component cascade (Figure 2) (Melis et al., 2015; Taylor and Lindorfer, 2016). Once antibodies bind to the target ligand, the available Fc-regions are then able to bind complement protein, C1q. Hexamerization of near-by antibodies allows for efficient C1q binding, which then activates C1r and C1s (Diebolder et al., 2014; Wang et al., 2016). Activation of C1r and C1s leads to the proteolytic cleavage of C4 and C2 to initiate the complement cascade and subsequently complement dependent cytotoxicity (CDC) (Diebolder et al., 2014; Wang et al., 2016). However, only IgG and IgM antibodies are able to elicit CDC since they are the only antibody isotypes that have a C1q binding site. Nevertheless, IgA antibodies have also been observed to elicit CDC via the classical pathway, despite not having a C1q site (Evers et al., 2018; Evers et al., 2020). Yet, this has only been seen in B-cell lymphoma cells and the mechanism has been attributed to other receptors in the B-cell able to bind to C1q.

As an effector function, CDC has been shown to be required for in vivo efficacy. Mice having the genes encoding for C1q knocked out showed no clinical efficacy with anti-CD20 antibody, rituximab (Di Gaetano et al., 2003). Also, follicular lymphoma patients with known polymorphisms in the C1qA gene reducing CDC activity have been correlated with low clinical response to rituximab (Coiffier et al., 2008a). Despite these results, Fc-engineering to increase CDC activity has been extensively done and a successful example of this has been anti-CD20 antibody, ofatumumab. Ofatumumab has been engineered to have an increased ability to hexametrize and bind to C1q subsequently leading to higher CDC activity (Zhang, 2009). This enhancement translated into better clinical outcomes since ofatumumab outperformed rituximab in chronic lymphocytic leukemia patients (Coiffier et al., 2008b).
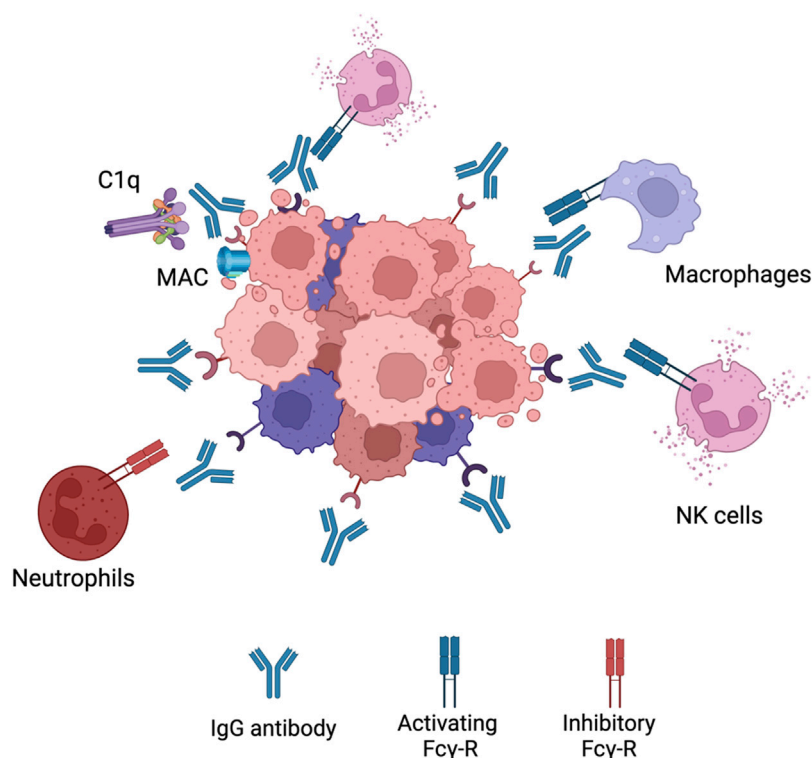
FIGURE 3
IgG1 effector mechanisms. When an IgG antibody opsonizes a cancer cell it can elicit various effector functions. It can interact with C1q complement protein leading to the formation of membrane attack complex (MAC) leading to CDC. The Fc region can also bind to activating Fcg-Rs on NK cells or macrophages to elicit ADCC or ADCP, respectively. Neutrophils express a high level of inhibitory Fcg-Rs leading to very little activation. Figures were created with BioRender.com.

2.1.3.5 Antibody-dependent cell phagocytosis

Macrophages express a variety of Fc receptors and in concrete Fc-γRII (CD32) and Fc-γRI (CD64) allowing for interaction with IgG antibodies against cancer (Figure 2) (Nimmerjahn et al., 2015). This interaction can then lead to cell death through a process called antibody-dependent cell mediated phagocytosis (ADCP) (Hubert et al., 2011). The role of ADCP in clinical efficacy has not been very well studied but there has been some evidence demonstrating a role in antibody efficacy. For example, rats having their macrophages depleted lost significant response towards monoclonal antibody therapy against colon carcinoma (Van Der Bij et al., 2010). Similar results were also shown with SCID-BEIGE mice transplanted with xenografts and treated with monoclonal antibody therapy (Overdijk et al., 2015). These specific mice do not have B or T cells and defective NK cells which then makes macrophages a primary effector immune cell and ADCP the main effector mechanism. These mice showed an *in vivo* clearance of leukemic cells when treated with daratumumab, an anti-CD38 antibody. ADCP efficacy in the clinic was also shown when 11 out of 12 of multiple myeloma patients showed ADCP when cells were cultured and treated with daratumumab *in vitro* (Overdijk et al., 2015).

A reason for why ADCP has not been so clearly correlated with antibody efficacy could be due to the expression of SIRPα and CD47 on macrophages and tumor cells, respectively. The interaction among both receptors leads to a "don't eat me" signal which downregulates ADCP activity (Murata et al., 2018). Blockage of

this axis has been shown to increase antibody therapy by enhancing ADCC activity. Currently, SIRPα and CD47 blockers are being tested in the clinic together with various antibody therapies (Murata et al., 2018).

2.1.3.6 Antibody dependent cell mediated cytotoxicity

In 1965, antibody opsonized cancer cells were shown to be killed *via* a non-phagocytic mechanism, termed antibody-dependent cell mediated cytotoxicity (ADCC) (Möller, 1965). This effector mechanism can be elicited from different types of immune cells such as NK cells, neutrophils, monocytes, and eosinophils (Nimmerjahn and Ravetch, 2008a). However, the way cell death is elicited differs among cells and can range from release of cytotoxic granules, reactive oxygen species release or Fas/FasL signaling (Eischen and Leibson, 1997; Nimmerjahn and Ravetch, 2008b; De Saint Basile et al., 2010). The clinical relevance of ADCC was first described in 2000 where Clynes and colleagues showed that rituximab and trastuzumab relied on ADCC for efficacy (Clynes et al., 2000). Moreover, it was later seen that mice lacking FcγRs or certain mutations limiting ADCC did not respond to monoclonal antibody therapy (De Haij et al., 2010). Within the population, polymorphisms in Fc-γRIIA (CD32a) (Wu et al., 1997) and Fc-γRIIIA (CD16a) (Bibeau et al., 2009) have been found and described to increase IgG affinity and ADCC activity. In several clinical trials with rituximab, it was seen that patients with such polymorphisms

FIGURE 4
IgA effector mechanisms. IgA effector mechanisms differ from IgG antibodies. IgAantibodies do not activate CDC, since they do not have a C1q binding site. The Fc region of IgA bindsto Fca-R on neutrophils cells or macrophages to elicit ADCC or ADCP, respectively. However, since NKcells do not express Fca-R, IgA antibodies do not activate such immune population. Figures were created with BioRender.com.

had a better clinical response (Cartron et al., 2002; Weng and Levy, 2003; Hatjiharissi et al., 2007). Similar results were also shown with cetuximab (Rodríguez et al., 2012) and trastuzumab (Musolino et al., 2008; Boero et al., 2015) treating colorectal cancer and metastatic breast cancer, respectively. Further confirming such results, patients with higher response to trastuzumab also demonstrated higher ADCC activity compared to patients not responding (Arnould et al., 2006).

Since all the cancer antibodies in the clinic are of the IgG isotype, Fc-γRs are the main receptors that mediate ADCC. In humans there exists six different types of Fc-γRs which can be divided into activating (Fc-γRI, Fc-γRIIA, Fc-γRIIC and Fc-γRIIIA) and inhibitory (Fc-γRIIB and Fc-γRIIIB) receptors (Wallace et al., 1994). As the name indicates, the activating receptors elicit ADCC while the inhibitory receptors downregulate effector mechanisms. With IgG therapy, NK cells are the main population that elicit ADCC which is due to the type of Fc-γR expression (Figure 3). NK cells express only one Fc-γR which is the activating Fc-γRIIIA explaining its importance for ADCC mediated by IgG (Wang et al., 2015). Other myeloid and granulocytic cells also express activating Fc-γR but also higher levels of inhibitory Fc-γR. For example, inhibitory Fc-γRIIB expression on neutrophils is seven to five times higher than Fc-γRIIA (Selvaraj et al., 1988). This has been shown to have a negative role on mediating ADCC because of the competition with Fc-γRIIA (van der Kolk et al., 2002; Peipp et al., 2008). This heavy reliance on NK cell for ADCC has been shown to limit efficacy. This is because NK cells have been seen to undergo exhaustion fast and not able to elicit ADCC(97). Only 24 h later NK cells gain the ability to elicit ADCC again (Gill et al., 2012).

2.1.3.7 IgA for cancer therapy

Due to the limitations the IgG isotype poses, preclinical studies have been conducted on the development of cancer therapeutic mAbs with isotypes different from IgG. A potential candidate isotype is IgA (Figure 4). This antibody is the most prominent immunoglobulin isotype found in mucosal sites and the second most frequent antibody isotype in serum, after IgG (98). It consists of two different isotypes; IgA1 and IgA2 with the latter comprising three allotypes; IgA2m (Siegel et al., 2020), IgA2m (Heron, 2019) and IgA2m(n). IgA interacts with immune cells *via* binding to Fc-αR (CD89) (Pakkanen et al., 2010). Such receptor is expressed on cells of the myeloid lineage such as neutrophils, monocytes, distinct macrophage populations and eosinophils (Pakkanen et al., 2010). Initial ADCC experiments with bispecific IgG1-antibodies where one of the F (ab')$_2$ fragments was directed at the FcαR receptor and the other to a target antigen highlighted the potential use of IgA antibodies in the context of malignancies (Valerius et al., 1997). Various reports have also shown that IgA mAbs directed at different tumour antigens showed an increased ability to recruit PMNs as effector cells compared to IgG (Huls et al., 1999; Dechant et al., 2002; Lohse et al., 2011; Boross et al., 2013; Leusen, 2015). This emphasises that IgA antibodies are able to employ a distinct effector population of immune cells against tumour cells compared to IgG. Furthermore, IgA antibodies mediate macrophage dependent tumour cell killing comparable to IgG (Lohse et al., 2012). It has been suggested that IgA mAbs are not able to activate the complement system due to the lack of a C1q-binding site (Bakema and van EgmondImmunoglobulin, 2011). However, certain studies have shown that IgA antibodies (Pascal et al., 2012) or IgG-fab fragments directed against CD20107 have been able to elicit CDC of malignant B-cells through the classical pathway. The mechanism behind it is thought to be due to rearrangements in the IgM or IgG B-cell receptor (BCR) of malignant B-cells (Engelberts et al., 2016) exposing its C1q binding site mediated by the clustering of CD20 after IgA binding. The FcαR represents an advantage over FcγRs since it does not have any inhibitory receptors and no polymorphisms have been reported. This implies that more predictable responses are

Achievable with IgA. Also, the FcαR has not been implicated in shaving leading to CD20 loss. Finally, antibody internalisation occurs less frequently with IgA compared to IgG. These advantages highlight the potential use of the IgA isotype in the development of therapeutic mAbs.

2.1.3.8 IgA and IgG combinational therapy for cancer

Despite the advantage of IgA, this isotype is not able to capitalize on NK cells or complement activation. To maximize on every effector population possible, scientists have tested whether combining IgG and IgA enhanced tumor killing. Bradsma and colleagues showed that using both IgG1 and IgA1 antibodies directed at different TAAs (Tumor Associated Antigens) induced higher killing than the individual antibodies when NK cells and neutrophils were present (Brandsma et al., 2015). However, when the IgG and IgA antibodies were directed towards the same TAA this enhanced effect was not seen. It is hypothesized that it could be due to the competition towards the same TAA leading to one isotype dominating in binding. Further building on this work, TrisomAB was then developed which consisted of an IgG1 antibody directed towards a TAA and Fc-αR (Heemskerk et al., 2021). TrisomAB was shown to increase tumor killing when both NK cells and neutrophils

were present. This data then further supports the use of both antibody isotypes in the treatment of cancer. Similar to TrisomAB, an Fc-fusion peptide against PD-L1 with a chimeric Fc presenting both IgG1 and IgA1 heavy chains was previously described (Hamdan et al., 2021). This Fc-fusion peptide was also shown to activate effector mechanisms of both an IgG1 and IgA1 which resulted in higher tumor killing *in vitro, in vivo* and in patient-derived organoids.

2.1.3.9 Fc-fusion peptides

Antibodies are large complexes which make them very hard to diffuse into large tumors (Strohl and Knight, 2009). Moreover, production is very complex and costly which inflate the price in the clinic (Strohl and Knight, 2009). Regarding such issues, a novel type of antibody-based therapeutics has been developed compromising of peptides fused to an Fc region (Nelson and Reichert, 2009; Strohl, 2009; Beck and Reichert, 2011). These peptides can be of very small size and bind to any desired target. However, when these peptides are administered systemically, they have a very short half-life due to rapid renal filtration. Attaching Fc-regions, in specific of the IgG region, increases the half-life of the peptides due to binding of Fc-neonatal receptors (Suzuki et al., 2010). Moreover, the Fc-regions can also provide Fc-effector mechanisms such as CDC, ADCC or ADCP. Currently, there are 13 different types of Fc-fusion peptides approved in the clinic used for thrombocytopenia, kidney transplants and inflammatory diseases such as arthritis or psoriasis (Strohl, 2009). Currently, no Fc-fusion peptides have been approved for the treatment of cancer. However, many Fc-fusion peptides against cancer have been described and entered clinical testing. For example, a bispecific peptide fused to a Fc of an IgG against HER-1 or HER-2 has showed high anti-tumor efficacy (Sioud et al., 2015). Moreover, the IgG Fc portion was able to elicit Fc-effector mechanisms of a normal IgG antibody.

2.2 Active immunotherapy

In contrast to passive, active immunotherapies are molecules that are used to induce or revitalize anti-tumor responses *in vivo*. This then requires patients to have an active and responsive immune system for successful treatment.

2.2.1 Checkpoint inhibitory therapy

In the thymus, the life of a T cell begins by proliferating and creating a diverse repertoire of TCRs. In order to maintain homeostasis, the immune system needs to distinguish between self and non-self. T-cells go through an initial selection process called central tolerance. In this process T-cells that strongly react to self-peptides, presented by thymocytes, undergo apoptosis. T-cells that weakly respond to self-peptides are released as naive cells to circulate into secondary lymphoid organs. APCs, specifically dendritic cells (DC), are then able to present naive T-cells either with foreign antigens (under infection conditions), overexpressed antigens, neoantigens or mutated self-proteins (under malignancy conditions) resulting in T-cell activation. However, some of the activated T-cells have TCRs which are still able to cross-react with self-antigens. To prevent cross-reactivity to self, multiple checkpoint
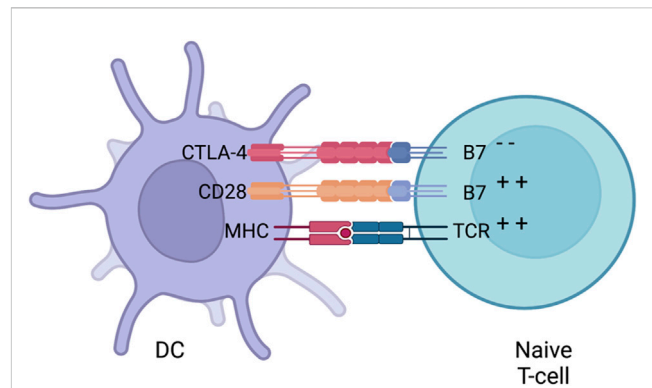
FIGURE 5
CTLA-4 suppression. Naive T-cells migrate to lymph nodes to become activated. Activationis usually provided by both MHC (loaded with an antigen) and co-stimulation from B7 (interacting with CD28) provided by a DC. After early stimulation, CTLA-4 is translocated to the surface of DCswhich then competes with B7 to bind to CD28 and downregulates T-cell activation. Whether a naiveT-cells undergoes activation or anergy is dependent on the balance of between CD28:B7 andCD28:CTLA4 signalling. Figures were created with BioRender.com.

pathways are present during the steps of activation to prevent autoimmunity. Also, checkpoints prevent the immune system to over activate during the course of infection. This process has been termed peripheral tolerance and two main constituents that take the centre of this process are the membrane receptors CTLA-4 and PD-1. Although CTLA-4 and PD-1 have a common function, they are present in different stages of T-cell activation. CTLA-4 is called the "leader" of the immune checkpoints since it regulates the activation of naive T-cells in lymph nodes. Contrary to CTLA-4, PD-1 acts in later stages since it regulates already activated T-cells in the peripheral tissues.

2.2.1.1 The CTLA-4 axis

Activating naive T-cells in the thymus is a complex process that requires more than one signal. In addition to TCR binding to peptide-loaded MHC, several co-stimulatory signals are required for full T-cell activation. An appropriate amount of co-stimulation from either B7-1 (CD80) or B7-2 (CD86), expressed on APCs and binding to CD28 on T cells, is required for activating naive T-cells which then leads to IL-2 production. Stimulatory signals from CD28:B7 also lead to the localisation of CTLA-4 to the surface of T-cells (Masteller et al., 2000). Even though CTLA-4 is a homologue to CD28, it has a higher affinity towards B7 providing competitive binding that results in decreased CD28:B7 interactions (Figure 5). The interaction between CTLA4 and B7 does not produce a stimulatory signal required for naive T-cell activation (Parry et al., 2005). Some data has suggested that CTLA-4 has signalling capabilities able to counteract CD28:B7 stimulatory signals. Other inhibitory mechanisms have also been proposed such as direct inhibition at the TCR immune synapse (Parry et al., 2005) or causing an increased T-cell mobility causing a decreased contact frequency with APCs (Schneider et al., 2006). Thus, CTLA-4 is seen as an inhibitor of the co-stimulation usually supplied by the interaction of CD28 and B7. Whether a naive T-cell undergoes activation or anergy is dependent on the balance between CD28:

B7 and CD28:CTLA4 signalling. What determines this balance still remains a mystery but multiple mechanisms such as ligand competition between B7 and CTLA-4, regulatory cytokines and CTLA-4 signalling has been proposed (Sansom, 2000). When the balance is tilted towards the negative CD28:CTLA4 signalling, IL-2 production is halted preventing cell cycle progression (Krummel and Allison, 1996).

As previously mentioned, CTLA-4 is upregulated on the surface of naive T-cells after CD28:B7 or TCR:MHC binding. Before such stimulation is provided, CTLA-4 is present in the cytoplasm of the cell within vesicles (Linsley et al., 1996). CD28 and TCR stimulation causes the exocytosis of the CTLA-4-containing vesicles, leading to the upregulation of CTLA-4 on the surface. This process is under a positive and graded feedback loop where stronger TCR and CD28 stimulation increases CTLA-4 translocation.

The importance of CTLA-4 in maintaining homeostasis was shown in adult mice, where abrogating CTLA-4 expression caused a systemic inflammation and formation of organ-specific autoantibodies. Moreover, congenital CTLA-4 deficient mice died due to lymphoproliferation (Walunas et al., 1998). Similar observations were also shown in humans where patients with CTLA-4 deficiencies suffer from various autoimmune and autoinflammatory diseases. CTLA-4 is not only expressed on naive T-cells but also on regulatory T cells (Tregs). Unlike in naive T-cells, CTLA-4 is constitutively expressed on Treg cells (Takahashi et al., 2000). This constitutive expression of CTLA-4 makes Treg cells key players in maintaining peripheral tolerance. For example, mice with Treg cells with impaired CTLA-4 had impaired suppressive functions (Walunas et al., 1998).

2.2.1.2 CTLA-4 inhibiton for the treatment of cancer

The rationale behind inhibiting CTLA-4 for treating cancer is not a novel idea but has been reported back in 1996 (Leach et al., 1996). Using preclinical models, it was shown that the blockade of CTLA-4 led to anti-tumour immunity. Mice administered with CTLA-4 antibodies rejected pre-established or injected tumours. Moreover, the rejection resulted in immunity against a second tumour challenge. This was further supported by other studies where administering CTLA-4 antibodies to mice with a pre-established B16-BL6 melanoma resulted in tumour clearance (van Elsas et al., 1999). Based on such preclinical evidence, two CTLA-4 antibodies, ipilimumab and tremelimumab, were developed and entered clinical development. Despite acceptable tolerance and durable responses in patients (Ribas et al., 2005; Hodi et al., 2008; Ribas et al., 2013), tremelimumab did not show statistical significance in overall survival (OS) in a phase III trial with advanced melanoma patients128. However, it is disputed that this may have been due to the crossing over of patients from the chemotherapy-only treatment arm to the chemotherapy and tremelimumab treatment arm. Ipililumab on the other hand has been successful in two recent phase III trials with advanced melanoma patients (Hodi et al., 2010; Robert et al., 2011). While the median survival improved minimally, the success of ipililumab was in the remarkable increase in landmark survival after treatment. After 2 years, 18% patients treated with ipililumab in combination with vaccination against the cancer-specific protein gp100 were alive compared to 5% of patients receiving gp100 vaccination alone. In addition, pooled data from clinical trials testing ipilimumab in advanced melanoma patients showed that 20% of patients had a long-term survival of at least 3 years (Postow et al., 2015a). Not only confined to advanced melanoma, ipililumab has also succeeded with other malignancies. Pancreatic cancer patients receiving ipililumab had an increase in OS compared to patients receiving chemotherapy only (Le et al., 2013). In addition, it also resulted in responses with prostate cancer patients (Slovin et al., 2013).

While the anti-tumour mechanisms of CTLA-4 antibodies are not well understood, the generally believed hypothesis is that blocking CTLA-4 causes an increased activation of proliferation of effector T-cells accompanied with a decrease in activated Treg cells (Fife and Bluestone, 2008). Supporting this hypothesis, good responses in melanoma patients was attributed to a wide and diverse pool of T-cells (Robert et al., 2014). However, other studies observed that a baseline T-cell diversity, before treatment, was associated with higher OS in metastatic melanoma patients (Khunger et al., 2019). Therefore, pre-existing conditions might be prognostic markers for CTLA-4 blockade anti-tumour efficacy rather than post-treatment induced artifacts.

2.2.1.3 PD1/PD-L1 axis

Similar to CTLA-4, PD-1 is part of the CD28/B7 family of co-stimulatory receptors. It is expressed on effector T cells and regulates them by binding to its ligands, PD-L1 and PD-L2, which are expressed by both hematopoeitic and non-hematopoeitic cells (Figure 6). Activation of PD-1 leads to the phosphorylation of both its intracellular immunoreceptor tyrosine-based switch motif (ITSM) and immunoreceptor tyrosine-based inhibitory motif (ITIM). The phosphorylation of these motifs attracts phosphatases, such as SHP-2, that are able to terminate signalling cascades of both CD28 and TCR (Bardhan et al., 2016). This then inhibits the proliferation and survival of T-cells and production of IL-2, IFN-γ and tumour necrosis factor-α (TNF-α). Hence, PD-1 is able to terminate TCR signalling and reduce T-cell activation. PD-1 is not constitutively expressed on T-cells but rather is a marker of "exhaustion". After high levels of stimulations from CD4+ T-cells, effector T-cells start to express PD-1 in order to prevent over-activation (Wherry, 2011). This exhaustion state is commonly observed both in chronic infections and cancer. Therefore, this causes the suboptimal control of infections and cancer progression. For example, mice chronically infected with cytomegalovirus had virus specific CD8+ T-cells present. Yet the T-cells were ineffective since they did not produce cytokines upon antigen challenge. This was also shown in metastatic melanoma patients where exhausted CD8+ T-cells were ineffective in tumour clearance.

Expression and the location between PD-L1 and PD-L2 are different. PD-L1 is expressed on many types of tumours and associated with poor prognosis and high TILs (Hino et al., 2010). PD-L2 resides on DC and monocytes but also on non-immune cells depending on the microenvironment. This contrasting distribution of PD-L1 and PD-L2 causes distinct biological effects when each ligand is bound to PD-1. For example, natural killer T cell (NKT) activation under PD-L1 or PD-L2 signalling were opposing (Akbari et al., 2010). Moreover, PD-L1 and CD80 interaction decreased T-cell response unlike when PD-L2 was blocked, an increased T-helper 2 cell (Th2) activity was noted (Huber et al., 2010). These opposing biological effects provide an explanation on the toxicity levels caused by inhibiting PD-1 and has highlighted the use of PD-L1 inhibitors.
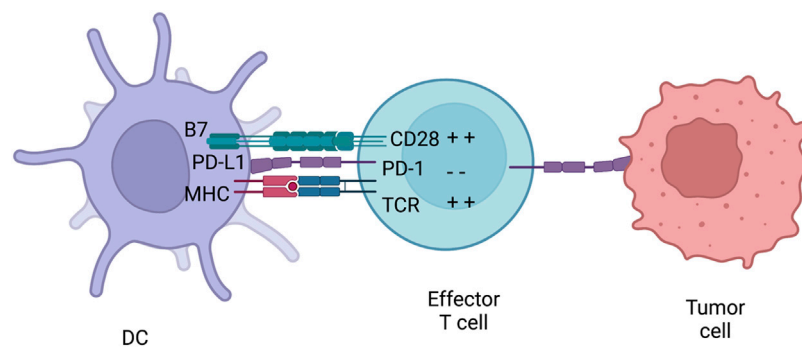
FIGURE 6
PD-1 inhibition. PD-1 is usually expressed on effector T-cells and binds to either PD-L1 or PD-L2. PD-L1 can be expressed both on immune cells and tumours. Therefore, PD-1 can inhibiteffector T-cells at different stages of an immune response. After PD-L1 is activated by the receptor,they can initiate a signalling complex able to counteract MHC and B7 signalling. Figures were created with BioRender.com.

Even though CTLA-4 and PD-1 have similar negative effects on T-cells, there are key differences between the checkpoints. CTLA-4 controls the activation of naive T-cells in lymph nodes whereas PD-1 controls T-cells in the effector phase in the periphery tissues. CTLA-4 expression is confined to T-cells, unlike PD-1 that is expressed on T-cells, B-cells and myeloid cells. Furthermore, the expression and distribution of checkpoint ligands differs. B7 is restricted to professional APCs while PD-L1 and PD-L2 are expressed on leukocytes, non-hematopoeitic cells, and non-lymphoid tissues (Keir et al., 2008). These differences indicate that CTLA-4 down regulates T-cell responses early on in an immune response while PD-1 limits T-cell response later during the effector stage. This then causes different effects *in vivo* when each receptor is inhibited. Blocking CTLA-4 causes an increase in activation and proliferation of effector T-cells regardless of TCR specificity while PD-1 inhibition leads to restoring proper T-cell functions.

2.2.1.4 Targeting the PD-1 axis in cancer therapy

After the success in targeting CTLA-4 during cancer, many antibodies have been designed to disturb the PD-1 axis for a similar purpose. Although the antibodies differ in structure (antibody isotype and chimerised/humanised), they can be categorised in two main groups: antibodies targeting the PD-1 receptor and antibodies targeting the ligand PD-L1. PD-1 antibodies such as nivolumab and pembrolizumab were shown in a phase I clinical trial with advanced melanoma, non-small cell lung cancer, renal cell carcinoma and other tumours to be tolerable and result in high durable responses (Topalian et al., 2012; Topalian et al., 2014). Recently, results from three phase III clinical trials with advanced melanoma have been published. In all three trials the OS was significantly higher in patients receiving nivolumab. In addition to melanoma, renal cell carcinoma patients treated with nivolumab had an OS of 25 months compared to 19.6 months in patients receiving the current standard treatment, everolimus (mTOR inhibitor). Similar results were also shown in a phase III trial with non-small cell lung cancer. The mechanism behind the anti-tumour effects of PD-1 blockade, which occurs during the effector stage of T-cells, involves re-activating peripheral T-cells that have been "exhausted" due to the high exposure of tumour antigens.

Many studies have reported that PD-1 is expressed on TILs (Topalian et al., 2014; Cao et al., 2017). Thus, PD-1 inhibition allows the suppressed TILs to gain back their anti-tumour properties. However, a recent study indicated that ipililumab response was associated with levels of expression of PD-L1 on tumour cells (Taube et al., 2014). When PD-1 is inhibited, the interaction between PD-1:PD-L1 is blocked yet PD-L1 is still able to inhibit T-cells by binding to CD80, a second receptor for this ligand. To overcome such limitation, PD-L1 antibodies have been generated and are able to disrupt both PD-1:PD-L1 and PD-L1: CD80 interactions. These antibodies are also able to keep intact the interaction between PD-1 and PD-L2, required for self-tolerance and thus leading to lower toxicities. Three PD-L1 antibodies (Atezolizumab, Durvalumab and Avelumab) have been clinically approved and have shown durable responses and less toxicity levels in a variety of tumours (Postow et al., 2015b).

2.2.1.5 Fc silencing of checkpoint inhibitors

The Fc-region of antibodies provides the ability to elicit Fc-effector mechanisms which are crucial for clinical efficacy. With respect to checkpoint inhibitors, it has been a subject of debate. For CTLA-4 checkpoint inhibitor, ipilimumab, Fc-effector mechanisms pertinent to the IgG1 isotype has been correlated with clinical efficiency. Advanced melanoma patients with Fc-γIIIA polymorphism V158F increasing IgG1 affinity, have been shown to respond better to ipilimumab (Vargas et al., 2018). Silencing Fc-effector mechanisms by changing the Fc-isotype or adding point mutations have seen to reduce *in vivo* activity of CTLA-4 inhibitors. The mechanism behind this has been argued to be the depletion of immunosuppressive Treg populations. However, opposite results were seen with PD-1 checkpoint inhibitors where Fc-effector mechanisms have lowered the *in vivo* anti-tumour activity (Dahan et al., 2015). PD-1 checkpoint inhibitors with competent Fc regions able to elicit Fc-effector mechanism were shown to deplete crucial CD8+T cell and CD4+ T cell populations (Dahan et al., 2015). These results explain why all of the PD-1 checkpoint inhibitors approved in the clinic are of the IgG4 isotype, an isotype that elicits low levels of ADCC and CDC, and have a S228P mutation decreasing Fc-γR binding. This reduction in Fc-effector activity then

prolongs CD8+T cell binding which subsequently increases anti-tumour activity.

As for PD-L1 checkpoint inhibitors, a safety concern over the addition of Fc-effector mechanisms exists. This is because PD-L1 expression is not solely limited to tumor cells but also can be expressed on healthy cells149. In result, the opsonisation of healthy cells with an antibody able to elicit Fc-effector mechanism can be deleterious. Out of the three approved PD-L1 inhibitors, atezolizumab and durvalumab have point mutations in the IgG1 Fc-region that remove Fc-γR binding. However, *in vivo* data has shown that arming such checkpoint inhibitors can increase anti-tumor efficacy (Dahan et al., 2015; Yi et al., 2021). This was attributed to an increased clearance of tumor cells but also immunosuppressive immune populations. Therefore, a strategy to increase Fc-effector mechanism while maintaining safety concerns is required.

2.2.2 Cancer vaccines

Vaccines have been a major milestone in preventing life threatening infectious diseases. The concept of being able to induce an immune response resulting in a protective immunological memory against cancer is ideal. Not only could this prevent or treat cancer but also help in tumour relapse. Nevertheless, cancer genomics has shown the complexity in achieving this since most of the tumour antigens being highly expressed on tumours are also shared among healthy cells. TAAs such as HER-2, glycoprotein (gp) 100, Telomerase and others are ideal antigen candidates due to their immunogenic properties, yet are expressed on healthy tissue (Finn, 2018; Blass and Ott, 2021). This lack of specificity is concerning due to the "off-target" effects that can be very toxic to a patient. However, in 2010 the first therapeutic cancer vaccine, Sipuleucel-T, was approved by the FDA for asymptomatic or mildly symptomatic metastatic prostate cancer (Plosker, 2011). This vaccine consisted of isolating dendritic cells from patient's PBMCs and expanding/activating them *ex-vivo* using the commonly known TAA called prostatic acid phosphatase (PAP). The approval of such cancer vaccine stimulated other vaccine platforms to be investigated in the clinic. For example, BioNTech have developed a novel RNA lipoplex complex, called FixVac, coding for different TAAs (Sahin et al., 2020). Such platforms can selectively target dendritic cells to induce an appropriate antigen presentation allowing for an effective T-cell immune response.

The perfect type of cancer vaccine would include an antigen selectively expressed on tumour cells. The genome of cancer cells is unstable and undergoes many genetic modifications such as somatic mutations, deletions, duplications and other processes. Due to such instability, neoantigens arise from cancer cells that are not found in healthy cells. Hence, such antigens then represent ideal targets for cancer vaccines. Nonetheless, these antigens are not very immunogenic and fail to induce a sustainable immune response. The first clinical trial evaluating a neoepitope based vaccine was with stage III cutaneous melanoma patients. These patients were injected with A*02:01-specific neoepitopes and a specific CD8[+] T cell response was observed (Lee et al., 2021). Yet, this activation was modest and was not very effective in controlling tumor growth. This field is still a hot topic with multiple type of strategies trying to further strengthen immunogenicity.

2.2.3 Oncolytic viruses

Scientists have stopped hunting for individual tumor suppressor genes or oncogenes and started investigating methods in disrupting whole tumorigenic biological pathways. Oncolytic viruses (OV) are the ideal agents in achieving this. Such viruses are able to thrive in tumor cells where such malignant pathways have been activated or disrupted, and exploit metabolic pathways that characterize tumorigenesis which result in oncolysis (Lawler et al., 2017). Also, oncolytic viruses have extensively been shown to stimulate systemic host immune responses (Lemos de Matos et al., 2020). The tumor microenvironment is immunosuppressive and boosting the immune system has been observed to have significant anti-tumor effects (Zhao et al., 2021). Hence the dual mechanism OV possess makes them interesting therapy agents.

These viruses have been genetically modified to conditionally replicate in cells in which specific cellular pathways are disrupted. This then allows OVs to infect both healthy and tumor cells and only replicate in tumor cells in which cellular pathways are compromised, but be recognized and cleared by healthy cells by the intrinsic immune system. However, studies using immunocompetent and immunocompromised mice have shown that the direct oncolysis of such viruses is not enough to induce tumor clearance (Grote et al., 2001). After tumor lysis, various TAAs are released and made accessible by nearby DCs (Hollingsworth and Jansen, 2019). Such TAAs are then able to be taken up, processed and presented on MHC complexes allowing for adaptive tumor-specific tumor responses to be formed. This mechanism of action has been shown to be key for a successful response. Many different types of DNA and RNA oncolytic viruses have and are currently under clinical development and testing. In this review, a specific focus will be drawn into adenoviruses to be used as OVs.

3 Adenoviruses and their roles as cancer therapies

3.1 Adenoviruses

Adenoviruses were first discovered when scientists were investigating adenoid cells (Rekosh et al., 1977). They observed that these viruses featured a double-stranded DNA genome of about 36 kb packaged into a capsid with an icosahedral shape. Adenoviruses are medium sized, around 100nm, particles with a non-envelope capsid composed of a penton, hexon and fiber knob domain all required for attachment and entry. There has been 57 different serotypes identified to date which can be subdivided into 6 groups (A-F).

Unlike many other viruses, adenoviruses circulate through humans during the whole year and are endemic in children. The mode of transmission of the virus is through water and fomites. Owing to their success of infection, adenoviruses are resilient to harsh environments due to their resistance to chemical and physical agents. For example, the resistance to gastric acid and biliary secretions has allowed such viruses to infect the gastrointestinal tracts (Hierholzer, 1992). Moreover, adenoviruses can withstand being outside the host for up to 3 weeks. To our advantage, despite causing flu-like symptoms these viruses rarely induce serious disease in healthy human but can be generate illness in immune-

compromised patients. No animal reservoirs have been identified making the virus hard to study due to low animal models to mimic disease (Ginsberg and Prince, 1994).

The infection process starts off with the protein-interactions between the adenovirus capsid and host-cellular membranes. The adenovirus capsid is comprised of 240 hexons and 12 pentons. Other minor components such as pIX, pVIII, pVI and IIIa are also present in the capsid. The pentons, which consists of complex of five polypeptide III, provide the base for the trimeric fiber to attach. The fiber contains the knob-fiber domain that is then responsible for attaching to host-cellular membranes. The receptor that the fiber binds to depends on the serotype but the main ones include the coxsackie adenovirus receptor, desmoglein-2 or CD46 (Bergelson et al., 1997; Wu et al., 2004; Marttila et al., 2005; Wang et al., 2011). After initial binding, the pentons in the adenovirus interact with the host-cell integrins ($\alpha_v\beta_3$ or $\alpha_v\beta_5$), leading to activation of certain signaling proteins (GTPases, phosphoinositide-3-OH kinase and MAPK) which induce the uptake of the virus particles via clathrin-coated vesicles (Wickham et al., 1993).

Once inside the vesicles, the acidification of the endosome cues for dismemberment of the viral capsid by proteolytically cleaving protein VI(171). After endosomal escape, the resulting virion is released and transported to the nucleus with the help of dynein and microtubules which interact with capsid proteins (mu, proteins VII and V). Once the adenovirus genome reaches the nucleus, the transcription of genes begins and is divided into two phases; early and late. The early phase consists in the transcription and translation of early gene products (E1A, E1B, E2, E3 and E4) which help in the replication of the adenovirus DNA genome. Moreover, the products of the early genes also then induce the expression of late genes (L1, L2, L3, L4 and L5). The late gene products are required for virion assembly since they represent the structural proteins.

After the genome has been replicated and the structural proteins expressed, virion assembly begins with the hexons and pentons clustering with multiple scaffolding proteins (L4 22-33K) (Ahi and Mittal, 2016a). This induces the insert of the viral DNA inside the virion structure and the final maturation of the virus by the release and cleavage of precursor proteins (L1 52-55K) (Ahi and Mittal, 2016a). The whole replication process of the adenovirus usually takes 24–36 h and can yield about 10,000 virions per cell to be released (Giberson et al., 2012).

3.1.1 Adenovirus genome, replication, and machinery behind it

Despite the small genome of 30–36 kb in length, adenoviruses can encode for multiple genes due to the overlapping open reading frames, alternate splicing, and ability of transcription from both strands of the genome. (Hoeben and Uil, 2013). As described previously, the early gene products are responsible for genome replication and mainly consist of the preterminal protein (pTP), DNA polymerase (Ad Pol) and DNA-binding protein (DBP). The late genes of the adenovirus include proteins involved in virion assembly and encapsulation and are only expressed once the early genes are. The multiple late genes are usually arranged in the adenovirus major late transcription unit (MLTU) which consists of five regions, L1-L5, and are under transcriptional control of the major late promoter (MLP) (Berget et al., 1977; Chow et al., 1977; Nevins and Darnell, 1978; Nevins and Chen-Kiang, 1981; Ramke

et al., 2017). Other than the early and late genes, the adenovirus has also two other gene products, pIX and IVa2, which are often described as intermediate genes since they are not in the MLTU but facilitate the expression of the late genes.

The adenovirus genome is flanked by inverted terminal repeats (ITR) which compromise of around 100 bp each. These ITRs contain a ~50 bp origin of replication which is made up of a core origin and auxiliary origin (Charman et al., 2019). The core region provides the binding site to pTP and and Ad Pol while the auxiliary region provides for cellular transcription factors nuclear factor 1 (NF1) and OCT-1 (De Jong et al., 2003). Moreover, near the ITR regions the adenovirus genome has a packaging sequence (ψ) which is required for encapsulation in virions (Ostapchuk and Hearing, 2003; Ahi and Mittal, 2016b). Finally, to the 5' ends of the genome terminal proteins (TP) is covalently attached which protects the DNA from degradation.

Genome replication starts with the formation of the pre-imitation complex which consists of multiple protein-protein and protein-DNA interactions (Figure 7). Firstly, Ad Pol will covalently attach to pTP via the dCMP nucleotide on its S580 amino acid position (Desiderio and Kelly, 1981; Smart and Stillman, 1982). Following Ad pol binding, DBP then binds to the core origin which then further facilitates the binding of Ad Pol and NF1 to core origin and auxiliary origin, respectively. NF1 and OCT-1 are not necessarily required for genome replication but rather enhance replication. After, Ad Pol dissociates from pTP and the formation of the nascent strand can then begin. This is marked with the dissociation of the pre-initiation complex and allowing DBP to unwind the dsDNA and allowing Ad Pol to form the nascent strand. Interestingly, displaced ssDNA can anneal to itself via the intramolecular/intermolecular interactions of the ITR regions which create dsDNA origins of replications. Hence, both dsDNA and ssDNA can be used as replication intermediates to increase the genome copy number.

3.1.2 Adenovirus-host cell interactions and selective replication

There exist very important interactions between viral and cellular-host proteins that facilitate the adenoviral replication cycle. During the replication cycle of adenovirus, the virus must sequester various cellular proteins to help in genome replication, transcription, and translation. Moreover, while doing so it also has to fight off the intrinsic pathways of the host cell that are designed to shut off cell machinery, induce apoptosis and clear the virus. Therefore, adenoviruses have multiple proteins aiding in facilitating all these processes. Due to the understanding of such mechanisms, scientists have been able to come up with genetic modification allowing adenovirus to conditionally replicate in tumor cells (Heise and Kirn, 2000). This section will describe these crucial interactions and how scientists have taken advantage of them to create conditionally replicating adenoviruses (CrAd).

For the adenovirus to start replicating its genome, the cell must be directed into S-phase (Sha et al., 2010). The adenovirus expresses E1a protein which is responsible in doing so by interacting with retinoblastoma protein (pRb). Under normal conditions, pRb can control the cell-cycle by interacting with DNA-binding transcription factor E2F (Heise et al., 2000). This interaction restricts E2F from binding to DNA and promoting cell replication. E1a can bind to pRB and restrict its interaction with E2F(Hemminki et al., 2015). This then
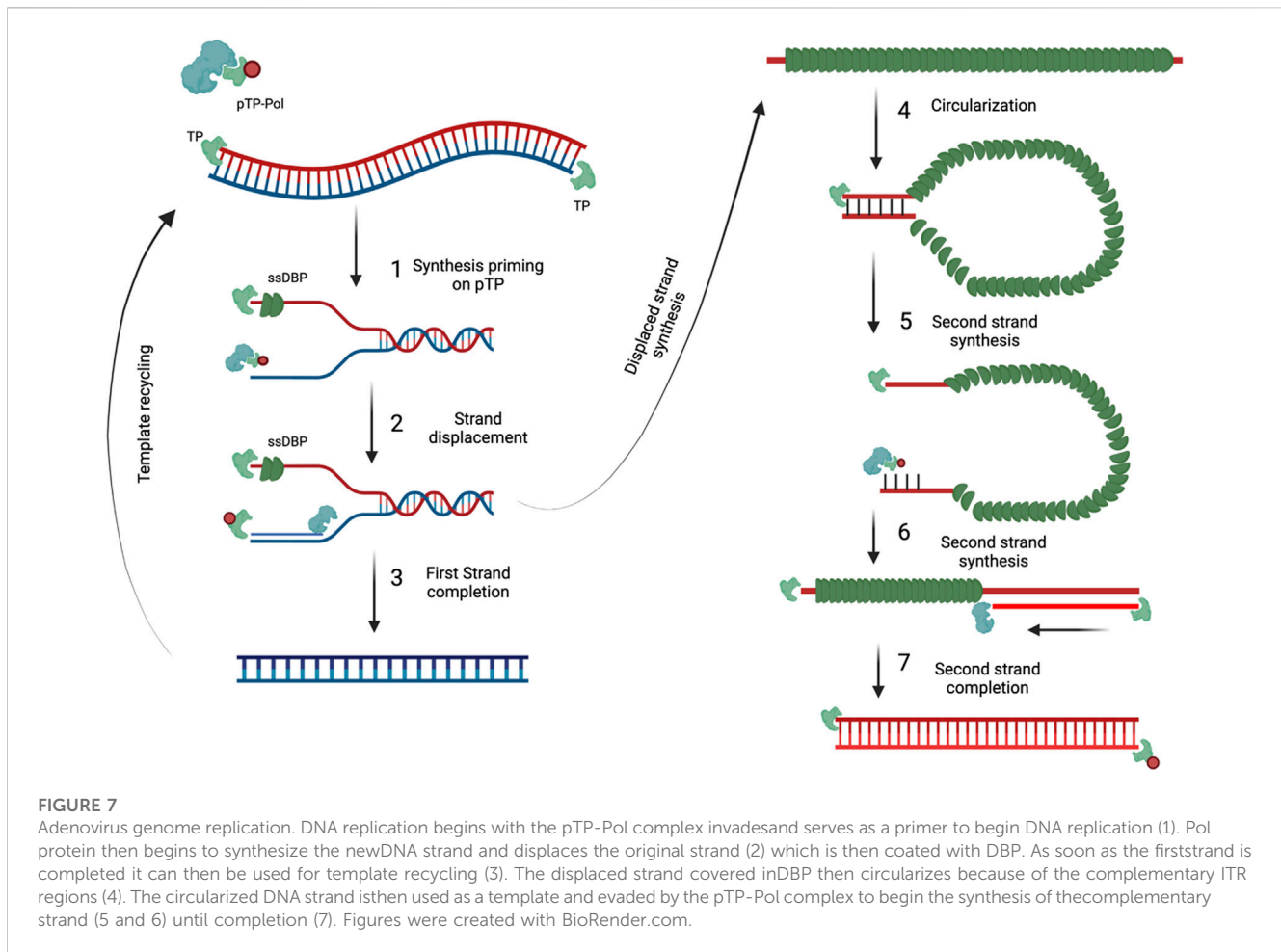
FIGURE 7
Adenovirus genome replication. DNA replication begins with the pTP-Pol complex invadesand serves as a primer to begin DNA replication (1). Pol protein then begins to synthesize the newDNA strand and displaces the original strand (2) which is then coated with DBP. As soon as the firststrand is completed it can then be used for template recycling (3). The displaced strand covered inDBP then circularizes because of the complementary ITR regions (4). The circularized DNA strand isthen used as a template and evaded by the pTP-Pol complex to begin the synthesis of thecomplementary strand (5 and 6) until completion (7). Figures were created with BioRender.com.

leads to E2F to be dissolved and bind freely to DNA and promoting transition into S-phase. Usually in malignant cells, the mechanisms controlling cell replication are defective to sustain cell growth. The majority of cancers have a deficient pRb protein and consequently an E2F roaming freely (Wu and Wu, 2021). In consequence, adenoviruses do not require E1a to replicate in tumor cells and becomes non-essential. Therefore, removing or rendering E1A defective can lead to a selective replication of adenoviruses in tumor cells (deficient in pRb proteins) while unable to replicate in healthy cells due to pRb. A 24 base pair deletion in the E1A protein has been previously described causing the protein unable to bind to pRb (Hemminki et al., 2015). Various clinically tested oncolytic adenoviruses apport this mutation making them selectively replicate in tumor cells and have been shown to be safe.

After adenovirus infection pushes cells into the S-phase, p53 accumulates as a response to induce apoptosis to control cell growth (Lomonosova et al., 2005). To circumvent this, adenoviruses express E1B 55k, E4 orf6 and E1B 19K which are all able to interact with p53 directly or indirectly to avoid apoptosis (Piya et al., 2011). Nevertheless, apoptosis might be disrupted but E1B 19K induces autophagy at the end of the viral replication cycle to release virions. Like pRb, p53 is also mutated in most cancers which prompts for a different strategy to induce selectivity in replication for adenoviruses. Deletion of the E1B gene induces adenoviruses to

replicate in tumor cells while leaving healthy cells free (Cheng et al., 2015). This deletion has also been noted to be clinically safe with different oncolytic viruses, such as ONYX-015 and H101, and in some countries they have been approved as therapy (Heise et al., 1997; Cheng et al., 2015).

3.1.3 Arming oncolytic adenoviruses

As mentioned previously, even though oncolytic adenoviruses can directly infect selectively tumor cells and induce oncolysis this is not enough for clinical efficacy. The release of TAAs from oncolysis leading to a vaccination effect is required for a successful treatment. However, a major limitation from achieving such clinical efficacy is the absence of anti-tumor immune cells and/or preventing their anti-tumor functions. One of the key advantages of using oncolytic adenoviruses is the ability to turn "cold" tumors with poor immune infiltration into "hot" tumors with high immune infiltration (Bramante et al., 2015). Yet, the amount of immune stimulation provided seems not to be enough to sustain clinical efficacy or tumor elimination. Despite this, researchers have armed oncolytic adenoviruses with various molecules ranging from cytokines, antibodies, bi-specific antibodies (BiTEs) and more (Figure 8). Other than expressing adequate levels of immunomodulatory molecules, the oncolytic tropism of the virus may help in

FIGURE 8
Enhancing oncolytic adenoviral therapy. Oncolytic adenoviruses have a specific tumortropism which can be utilized to arm such viruses with immunomodulatory genes. This then leads tothe expression and secretion of the immunomodulatory proteins in the tumor microenvironmentwhich increases immune infiltration. Figures were created with BioRender.com.

circumventing toxicity issues by limiting expression in the tumor microenvironment with minimal leakage to the periphery.

One class of molecules that has been used to arm oncolytic adenoviruses are co-stimulatory molecules. An example is the arming of oncolytic adenoviruses with two immune-activating ligands CD40L and OX40L (Malmström et al., 2010; Pesonen et al., 2012; Loskog et al., 2016; Schiza et al., 2017). CD40L when secreted can interact with CD40 present on APCs and enhance their antigen presentation and co-stimulation capacity (Piechutta and Berghoff, 2019). Moreover, OX40L binds to OX40 found on T cells and induces the survival and homeostasis of memory T-cells (Croft et al., 2009). Another strategy was arming oncolytic adenovirus LOAd703 with CD40L and 4-1BBL (Eriksson et al., 2017). The interaction of 4-1BBL with 4-BBL among T cells and APC lead to the increase of T-cell proliferation and activation. LOAd703 has been tested in clinical trials against many solid tumors and, interestingly, with pancreatic cancer it has been seen to reduce myeloid derived suppressed cells (MDSC) and increase memory T cells in many patients.

The release of cytokines and chemoattractant from oncolytic viruses are a successful strategy to increase immune cell homing to the tumor. The only FDA approved oncolytic virus in the clinic, T-VEC, compromises of a herpes simplex virus expressing GM-CSF (Liu et al., 2003). This cytokine helps in the maturation and antigen presentation

of APC, leading to better induction of T-cell immune responses. A similar version of T-VEC exists, but rather than a herpes simplex virus an adenovirus is used with the 24 base pair deletion in its E1A, previously described, to express GM-CSF (Cerullo et al., 2010). Such virus, called ONCOS-102, is under clinical evaluation and was seen to increase CD8+T cells circulation but more importantly antigen-specific CD8+ T cells in mesothelioma and multiple peritoneal malignancies (Ranki et al., 2016). Many oncolytic adenoviruses have been used to locally express various cytokine such as IL-2 and tumor necrosis factor α (TNF-α), IL-18, IL-24 or IL-12 in order to potentiate anti-tumor immune responses. Other than cytokines, chemokines such as CXCL9 and CXCL10 have also been cloned in oncolytic adenoviruses to recruit T-cells (Ylösmäki and Cerullo, 2020).

BiTEs are small molecules which consist of two scFv directed at different tumor antigens. Conventional BiTEs usually have one of their scFV directed towards CD3 while the other towards a TAA. These BiTE's main mechanism of action is bringing CD3+ T cells into close proximity of tumor cells and induce MHC-independent killing (Scott et al., 2018). These BiTEs have been shown to be excellent therapies for the treatment of lymphomas and leukaemias. For example, Blinatumomab, against CD3 and CD19, is the first BiTE to be approved by the FDA for the use of B-maligancies (Goebeler and Bargou, 2016). However, for solid tumors it has been seen not to be
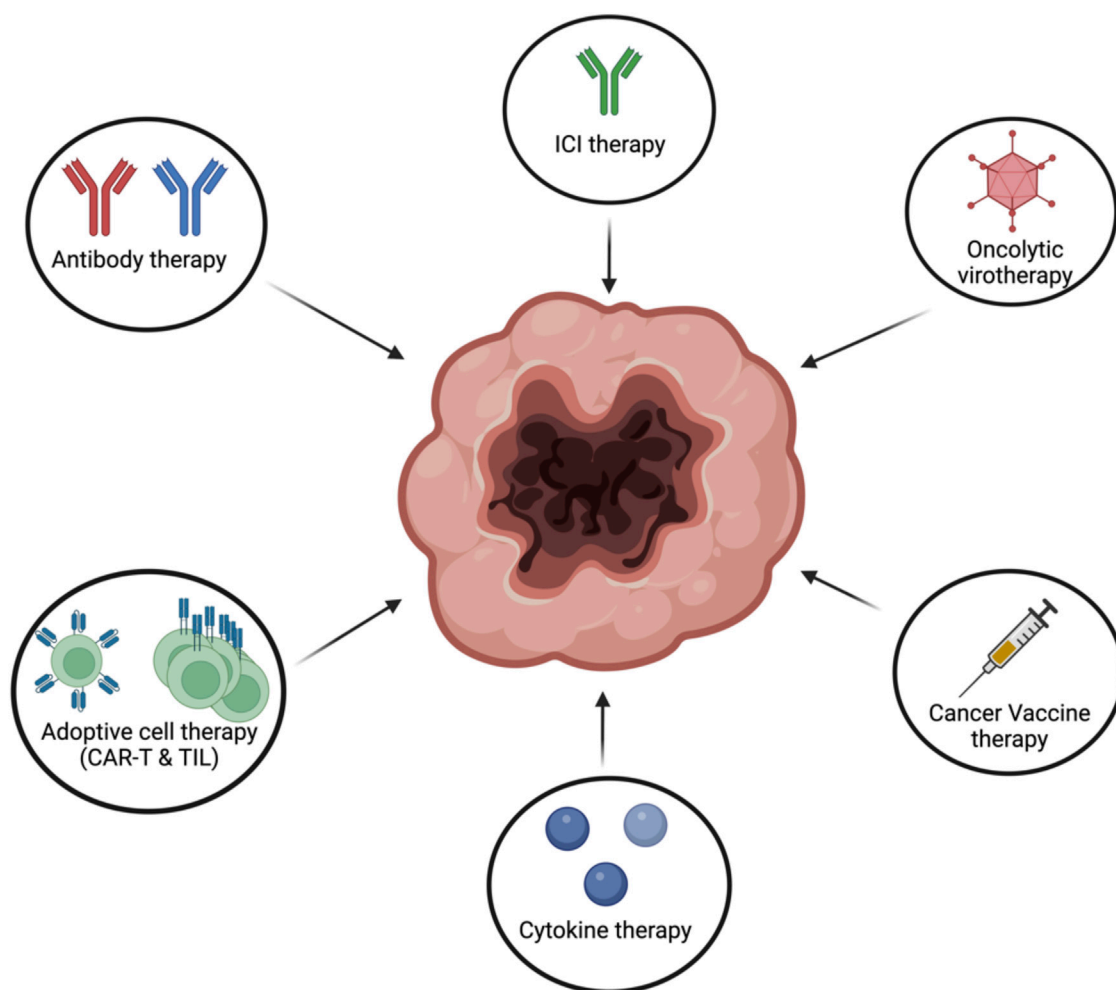
effective since their half-life in blood is short-lived, which consequently requires constant infusion of treatment leading to systemic toxicities. Yet, oncolytic adenoviruses have shown to provide excellent platforms to deliver BiTEs locally and persistently in solid tumors. Enadenotucirev is one of the first oncolytic adenoviruses to express a BiTE, which was directed towards TAA epithelial cell adhesion molecule (EpCAM) and CD3 (Yu et al., 2014). Other than just targeting TAA, a similar oncolytic adenovirus expressing BiTE was also constructed but directed towards fibroblast activation protein (FAP) which is found on cancer-associated fibroblasts (Freedman et al., 2017). The combination of both viruses demonstrated enhanced anti-tumor efficacy and T cell recruitment and function.

The systemic administration of checkpoint inhibitor has been associated with many adverse events. To further improve the safety profile, checkpoint inhibitors have been packaged into the genome of oncolytic adenovirus. Checkpoint inhibitors against CTLA-4 (Dias et al., 2011) and PD-L1 (Wang et al., 2020) have been cloned into oncolytic adenoviruses and have shown to be

effective in controlling tumor growth with a high safety profile. Yet, a limiting factor that needs to be addressed with this strategy is that adenovirus has low capacity for cloning long transgenes in the genome. Hence, cloning whole antibodies consisting of a heavy and light chain can affect the viral fitness.

3.1.4 Construction of adenoviral vectors

The use of adenoviruses for gene therapy, vaccines and cancer immunotherapies has increased throughout the years. This entails the engineering of adenovirus to express any gene of interest (GOI), a process that has been modified several times in order to optimize the procedure. The classical approach that many scientists have used, is the cloning the GOI in a shuttle plasmid containing a 5′-ITR, a packaging signal and sequence of homologous recombination (Stratford-Perricaudet et al., 1992; Mittal et al., 1993). This shuttle plasmid is then transfected into HEK293 cells with an adenovirus vector for homologous recombination to occur and create an adenovirus genome incorporating the GOI. Another used method is the cloning

of the GOI into a similar shuttle plasmid but the homologues recombination sequence is substituted with LoxP site(s) (Hardy et al., 1997). This shuttle is then transfected into HEK293 cells with an adenovirus genome containing LoxP sites. The shuttle vector and adenovirus genome are then joined *via* Cre recombinase-mediated recombination. A separate cloning method is the use of shuttle vector containing a 5′-ITR, a packaging signal and sequence of homologous recombination flanking a kanamycin resistance gene. After the GOI is added to this shuttle, it is linearized and transfected into bacterial cells (BJ5183) along with an ampicillin-resistant adenovirus backbone (He et al., 1998). Colonies are then screened based on kanamycin resistance and the final adenovirus containing the GOI product is linearized and transfected into HEK293 cells.

Each cloning system presented here are well characterized, reproducible and easy to design and carry out. However, these methods are very time consuming, and it can take up to 6 months or more to obtain the final product. This is because homologous recombination has very low efficiency and can take multiple rounds for a positive colony. Moreover, secondary recombination can occur leading to the incorporation of unwanted repeated regions or secondary structures. Hence, the need of novel cloning methods that are faster, easier and reliable are required.

4 Conclusions and future prospects

Cancer immunotherapies have taken the main stage in the treatment of cancer. This is mostly due to the dramatical increase in survival and quality of life for cancer patients. Nevertheless, since cancer is heterogenous not one type of cancer immunotherapy works for all. Depending on multiple factors, certain cancer immunotherapies work better for some cancer patients than others (Figure 9).

To date, only one oncolytic virus has been granted FDA approval for treatment despite years of extensive investigation. One of the reasons is that oncolytic viruses have been generally seen as direct tools for killing cancer due to their tumor-specific tropism. A growing body of evidence has shown that the ability of the virus to activate the immune system is a key attribute with regard to long-term antitumor effects. Therefore, to make more significant advances with such therapies there has been a shift in focus from not solely viewing oncolytic viruses as direct oncolytic tools but also as immunotherapies. This further validates the importance of harnessing the immune system to combat cancer rather than using cytotoxic drugs. Cientists have equipped oncolytic viruses with multiple immune-stimulatory molecules which have enhanced anti-tumor effects. Other than enhancing anti-tumor effects, this has also had a positive effect regarding limiting toxicities since the expression/release of the molecules is limited to the tumor.

Cytokine therapy is an effective strategy to induce an anti-tumor immune response to combat cancer. Yet, safety is one of the main limitations that made clinical approval hard to obtain for cytokine therapy. Usually, these molecules are injected systemically and can induce an overwhelming immune activation leading to various immune-related side effects. One strateg that has been used to circumvent this has been the use of biological carriers, like OV, to limit the immune activation in the tumor microenvironment. Moreover, due to the poor infiltration of CAR-T cells to solid tumors the, the combination of cytokine therapy and CAR-T cells has been tested. Research has shown that cytokine therapy can stimulate higher tumor infiltration of CAR-T cells with solid tumors. These observations further emphasize the testing of different combinations of cancer immunotherapies as possible synergism might exist. Another example of a possible synergy is between cancer vaccines and ICI therapy. Cancer vaccines have been shown to orchestrate an effective T-cell immune response specific towards TAAs. The combination with ICI's, specifically PD-1 or PD-L1 inhibitors, could further enhance the anti-tumor T-cell response by decreasing the inhibitory effects exerted by the tumor cells to escape immune desctruction.

Almost all the cancer immunotherapies in the clinic target PBMCs. Many studies have shown that targeting solely PBMCs does not cause full clearance since the cytotoxic effects mediated are finite. PMNs have been a neglected cell population despite being the largest leukocyte population in blood and highly infiltrated in tumors. This has been mostly due to the use of IgG antibodies which sub-optimally activates neutrophils. This is simply because PMNs highly express CD32b and CD16b which downregulate effector functions or act as a molecular "sink", respectively. This can be dangerous since immune cells have shown to be malleable depending on the microenvironment and stimulus provided. For example, researchers found relatively normal levels of Treg cells in the synovial membrane from rheumatoid arthritis patients compared to healthy individuals. However rather than promoting immune resolution, Tregs cells from the patients were programmed to secrete a powerful pro-inflammatory cytokine IL-17. The mechanism behind was speculated to be most probably due to the influence of the highly inflammatory microenvironment. Despite this, it could explain why the infiltration of neutrophils to the tumor is associated with a lower prognosis since they are not adequately activated (Masucci et al., 2019). This then calls for appropriate molecules able to capitalize such population to be used as an effector population. Moreover, rather than just activating PMNs the main goal in the future would to also involve other effector mechanisms. Multiple studies have shown that involving more than effector population leads to higher tumor killing compared to when each population is used on its own.

Author contributions

FH and VC wrote and corrected the review.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Achkar, T., Arjunan, A., Wang, H., Saul, M., Davar, D., Appleman, L. J., et al. (2017). High-dose interleukin 2 in patients with metastatic renal cell carcinoma with sarcomatoid features. *PLoS One* 12, e0190084. doi:10.1371/JOURNAL.PONE.0190084

Ahi, Y. S., and Mittal, S. K. (2016). Components of adenovirus genome packaging. *Front. Microbiol.* 7, 1503. doi:10.3389/fmicb.2016.01503

Ahi, Y. S., and Mittal, S. K. (2016). Components of adenovirus genome packaging. *Front. Microbiol.* 7, 1503. doi:10.3389/FMICB.2016.01503

Akbari, O., Stock, P., Singh, A. K., Lombardi, V., Lee, W-L., Freeman, G. J., et al. (2010). PD-L1 and PD-L2 modulate airway inflammation and iNKT-cell-dependent airway hyperreactivity in opposing directions. *Mucosal Immunol.* 3, 81–91. doi:10.1038/mi.2009.112

Amaria, R. N., Reuben, A., Cooper, Z. A., and Wargo, J. A. (2015). Update on use of aldesleukin for treatment of high-risk metastatic melanoma. *Immunotargets Ther.* 4, 79–89. doi:10.2147/ITT.S61590

Arnould, L., Gelly, M., Penault-Llorca, F., Benoit, L., Bonnetain, F., Migeon, C., et al. (2006). Trastuzumab-based treatment of HER2-positive breast cancer: An antibody-dependent cellular cytotoxicity mechanism? *Br. J. Cancer* 94, 259–267. doi:10.1038/SJ.BJC.6602930

Bakema, J. E., and van EgmondImmunoglobulin, M. A. (2011). A next generation of therapeutic antibodies? *MAbs* 3, 352–361. doi:10.4161/mabs.3.4.16092

Bardhan, K., Patsoukis, N., Weaver, J., Gordon Freeman, L. L., and Boussiotis, V. A. (2016). PD-1 inhibits the TCR signaling cascade by sequestering SHP-2 phosphatase, preventing its translocation to lipid rafts and facilitating Csk-mediated inhibitory phosphorylation of Lck. *J. Immunol.* 196, 128.15. doi:10.4049/jimmunol.196.supp.128.15

Baudino, T. (2015). Targeted cancer therapy: The next generation of cancer treatment. *Curr. Drug Discov. Technol.* 12, 3–20. doi:10.2174/1570163812666150602144310

Beck, A., and Reichert, J. M. (2011). Therapeutic Fc-fusion proteins and peptides as successful alternatives to antibodies. *MAbs* 3, 415–416. doi:10.4161/MABS.3.5.17334

Behring, E. von (2013). *Ueber das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren.*

Bergelson, J. M., Cunningham, J. A., Droguett, G., Kurt-Jones, E. A., Krithivas, A., Hong, J. S., et al. (1997). Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* 275, 1320–1323. doi:10.1126/SCIENCE.275.5304.1320

Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* 74, 3171–3175. doi:10.1073/PNAS.74.8.3171

Bibeau, F., Lopez-Crapez, E., Fiore, F. Di, Thezenas, S., Ychou, M., Blanchard, F., et al. (2009). Impact of Fc{gamma}RIIa-Fc{gamma}RIIIa polymorphisms and KRAS mutations on the clinical outcome of patients with metastatic colorectal cancer treated with cetuximab plus irinotecan. *J. Clin. Oncol.* 27, 1122–1129. doi:10.1200/JCO.2008.18.0463

Blankenstein, T., Coulie, P. G., Gilboa, E., and Jaffee, E. M. (2012). The determinants of tumour immunogenicity. *Nat. Rev. Cancer* 12, 307–313. doi:10.1038/NRC3246

Blass, E., and Ott, P. A. (2021). Advances in the development of personalized neoantigen-based therapeutic cancer vaccines. *Nat. Rev. Clin. Oncol.* 18 (18), 4215–4229. doi:10.1038/s41571-020-00460-2

Boero, S., Morabito, A., Banelli, B., Cardinali, B., Dozin, B., Lunardi, G., et al. (2015). Analysis of *in vitro* ADCC and clinical response to trastuzumab: Possible relevance of FcγRIIIA/FcγRIIA gene polymorphisms and HER-2 expression levels on breast cancer cell lines. *J. Transl. Med.* 13, 324. doi:10.1186/S12967-015-0680-0

Boross, P., Lohse, S., Nederend, M., Jansen, J. H. M., van Tetering, G., Dechant, M., et al. (2013). IgAEGFR antibodies mediate tumour killing *in vivo*. *EMBO Mol. Med.* 5, 1213–1226. doi:10.1002/emmm.201201929

Bramante, S., Kaufmann, J. K., Veckman, V., Liikanen, I., Nettelbeck, D. M., Hemminki, O., et al. (2015). Treatment of melanoma with a serotype 5/3 chimeric oncolytic adenovirus coding for GM-CSF: Results *in vitro*, in rodents and in humans. *Int. J. Cancer* 137, 1775–1783. doi:10.1002/IJC.29536

Brandsma, A. M., Broeke, T. Ten, Nederend, M., Meulenbroek, L. A. P. M., Van Tetering, G., Meyer, S., et al. (2015). Simultaneous targeting of FcγRs and FcαRI enhances tumor cell killing. *Cancer Immunol. Res.* 3, 1316–1324. doi:10.1158/2326-6066.CIR-15-0099-T

Cao, J., Brouwer, N. J., Richards, K. E., Marinkovic, M., van Duinen, S., Hurkmans, D., et al. (2017). PD-L1/PD-1 expression and tumor-infiltrating lymphocytes in conjunctival melanoma. *Oncotarget* 8, 54722–54734. doi:10.18632/oncotarget.18039

Cartron, G., Dacheux, L., Salles, G., Solal-Celigny, P., Bardos, P., Colombat, P., et al. (2002). Therapeutic activity of humanized anti-CD20 monoclonal antibody and polymorphism in IgG Fc receptor FcgammaRIIIa gene. *Blood* 99, 754–758. doi:10.1182/BLOOD.V99.3.754

Cerullo, V., Pesonen, S., Diaconu, I., Escutenaire, S., Arstila, P. T., Ugolini, M., et al. (2010). Oncolytic adenovirus coding for granulocyte macrophage colony-stimulating factor induces antitumoral immunity in cancer patients. *Cancer Res.* 70, 4297–4309. doi:10.1158/0008-5472.CAN-09-3567

Charman, M., Herrmann, C., and Weitzman, M. D. (2019). Viral and cellular interactions during adenovirus DNA replication. *FEBS Lett.* 593, 3531–3550. doi:10.1002/1873-3468.13695

Chen, J. S., Lan, K., and Hung, M. C. (2003). Strategies to target HER2/neu overexpression for cancer therapy. *Drug Resist Updat* 6, 129–136. doi:10.1016/S1368-7646(03)00040-2

Cheng, P. H., Wechman, S. L., McMasters, K. M., and Zhou, H. S. (2015). Oncolytic replication of e1b-deleted adenoviruses. *Viruses* 7, 5767–5779. doi:10.3390/V7112905

Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell.* 12, 1–8. doi:10.1016/0092-8674(77)90180-5

Clynes, R. A., Towers, T. L., Presta, L. G., and Ravetch, J. V. (2000). Inhibitory Fc receptors modulate *in vivo* cytotoxicity against tumor targets. *Nat. Med.* 6, 443–446. doi:10.1038/74704

Coiffier, B., Lepretre, S., Pedersen, L. M., Gadeberg, O., Fredriksen, H., Van Oers, M. H. J., et al. (2008). Safety and efficacy of ofatumumab, a fully human monoclonal anti-CD20 antibody, in patients with relapsed or refractory B-cell chronic lymphocytic leukemia: A phase 1-2 study. *Blood* 111, 1094–1100. doi:10.1182/BLOOD-2007-09-111781

Coiffier, B., Lepretre, S., Pedersen, L. M., Gadeberg, O., Fredriksen, H., Van Oers, M. H. J., et al. (2008). Safety and efficacy of ofatumumab, a fully human monoclonal anti-CD20 antibody, in patients with relapsed or refractory B-cell chronic lymphocytic leukemia: A phase 1-2 study. *Blood* 111, 1094–1100. doi:10.1182/BLOOD-2007-09-111781

Couzin-Frankel, J. (2013). Breakthrough of the year 2013. Cancer immunotherapy. *Science* 342, 1432–1433. doi:10.1126/SCIENCE.342.6165.1432

Cowley, G. P., Smith, J. A., and Gusterson, B. A. (1986). Increased EGF receptors on human squamous carcinoma cell lines. *Br. J. Cancer* 53, 223–229. doi:10.1038/BJC.1986.39

Croft, M., So, T., Duan, W., and Soroosh, P. (2009). The significance of OX40 and OX40L to T cell biology and immune disease. *Immunol. Rev.* 229, 173–191. doi:10.1111/J.1600-065X.2009.00766.X

Dahan, R., Sega, E., Engelhardt, J., Selby, M., Korman, A. J., and Ravetch, J. V. (2015). *FcγRs modulate the anti-tumor activity of antibodies targeting the PD-1/PD-L1 Axis.* doi:10.1016/j.ccell.2015.08.004

Davar, D., Ding, F., Saul, M., Sander, C., Tarhini, A. A., Kirkwood, J. M., et al. (2017). "High-dose interleukin-2 (hd IL-2) for advanced melanoma: A single center experience from the," in *J. Immunother. Cancer* (Milwaukee, WI: University of Pittsburgh Cancer Institute) 5, 1–10. doi:10.1186/S40425-017-0279-5/FIGURES/3

De Haij, S., Jansen, J. H. M., Boross, P., Beurskens, F. J., Bakema, J. E., Bos, D. L., et al. (2010). *In vivo* cytotoxicity of type I CD20 antibodies critically depends on Fc receptor ITAM signaling. *Cancer Res.* 70, 3209–3217. doi:10.1158/0008-5472.CAN-09-4109

De Jong, R. N., Van Der Vliet, P. C., and Brenkman, A. B. (2003). Adenovirus DNA replication: Protein priming, jumping back and the role of the DNA binding protein DBP. *Curr. Top. Microbiol. Immunol.* 272, 187–211. doi:10.1007/978-3-662-05597-7_7

De Saint Basile, G., Ménasché, G., and Fischer, A. (2010). Molecular mechanisms of biogenesis and exocytosis of cytotoxic granules. *Nat. Rev. Immunol.* 10, 568–579. doi:10.1038/NRI2803

Dechant, M., Vidarsson, G., Stockmeyer, B., Repp, R., Glennie, M. J., Gramatzki, M., et al. (2002). Chimeric IgA antibodies against HLA class II effectively trigger lymphoma cell killing. *Blood* 100, 4574–4580. doi:10.1182/blood-2002-03-0687

Desiderio, S. V., and Kelly, T. J. (1981). Structure of the linkage between adenovirus DNA and the 55,000 molecular weight terminal protein. *J. Mol. Biol.* 145, 319–337. doi:10.1016/0022-2836(81)90208-4

Di Gaetano, N., Cittera, E., Nota, R., Vecchi, A., Grieco, V., Scanziani, E., et al. (2003). Complement activation determines the therapeutic activity of rituximab *in vivo*. *J. Immunol.* 171, 1581–1587. doi:10.4049/JIMMUNOL.171.3.1581

Dias, J. D., Hemminki, O., Diaconu, I., Hirvinen, M., Bonetti, A., Guse, K., et al. (2011). Targeted cancer immunotherapy with oncolytic adenovirus coding for a fully human monoclonal antibody specific for CTLA-4. *Gene Ther.* 1919, 10988–10998. doi:10.1038/gt.2011.176

Diebolder, C. A., Beurskens, F. J., De Jong, R. N., Koning, R. I., Strumane, K., Lindorfer, M. A., et al. (2014). Complement is activated by IgG hexamers assembled at the cell surface. *Science* 343, 1260–1263. doi:10.1126/SCIENCE.1248943/SUPPL_FILE/DIEBOLDER.SM.PDF

Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., et al. (1984). Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences. *Nature* 307, 521–527. doi:10.1038/307521A0

Dunn, G. P., Koebel, C. M., and Schreiber, R. D. (2006). Interferons, immunity and cancer immunoediting. *Nat. Rev. Immunol.* 6 (6), 11836–11848. doi:10.1038/nri1961

Dunn, G. P., Old, L. J., and Schreiber, R. D. (2004). The three Es of cancer immunoediting. *Annu. Rev. Immunol.* 22, 329–360. doi:10.1146/ANNUREV.IMMUNOL.22.012703.104803

Eischen, C. M., and Leibson, P. J. (1997). Role for NK-cell-associated Fas ligand in cell-mediated cytotoxicity and apoptosis. *Res. Immunol.* 148, 164–169. doi:10.1016/S0923-2494(97)84219-8

Engelberts, P. J., Voorhorst, M., Schuurman, J., van Meerten, T., Bakker, J. M., Vink, T., et al. (2016). Type I CD20 antibodies recruit the B cell receptor for complement-dependent lysis of malignant B cells. *J. Immunol.* 197, 4829–4837. doi:10.4049/jimmunol.1600811

Eriksson, E., Milenova, I., Wenthe, J., Hle, M. S., Leja-Jarblad, J., Ullenhag, G., et al. (2017). Shaping the tumor stroma and sparking immune activation by CD40 and 4-1BB signaling induced by an armed oncolytic virus. *Clin. Cancer Res.* 23, 5846–5857. doi:10.1158/1078-0432.CCR-17-0285

Evers, M., Kruse, E., Hamdan, F., Lebbink, R-J., and Leusen, J. H. W. (2018). Comment on "type I CD20 antibodies recruit the B cell receptor for complement-dependent lysis of malignant B cells. *J. Immunol.* 200, 2515–2516. doi:10.4049/JIMMUNOL.1800087

Evers, M., Ten Broeke, T., Jansen, J. H. M., Nederend, M., Hamdan, F., Reiding, K. R., et al. (2020). Novel chimerized IgA CD20 antibodies: Improving neutrophil activation against CD20-positive malignancies. *MAbs* 12, 1795505. doi:10.1080/19420862.2020.1795505

Fagraeus, A. (1947). Plasma cellular reaction and its relation to the formation of antibodies *in vitro. Nature* 159, 499. doi:10.1038/159499A0

Fife, B. T., and Bluestone, J. A. (2008). Control of peripheral T-cell tolerance and autoimmunity via the CTLA-4 and PD-1 pathways. *Immunol. Rev.* 224, 166–182. doi:10.1111/j.1600-065X.2008.00662.x

Finn, O. J. (2018). The dawn of vaccines for cancer prevention. *Nat. Rev. Immunol.* 18, 183–194. doi:10.1038/NRI.2017.140

for Business, D., and Strategy, I. (2020). *Public attitudes to science 2019.*

Freedman, J. D., Hagel, J., Scott, E. M., Psallidas, I., Gupta, A., Spiers, L., et al. (2017). Oncolytic adenovirus expressing bispecific antibody targets T-cell cytotoxicity in cancer biopsies. *EMBO Mol. Med.* 9, 1067–1087. doi:10.15252/EMMM.201707567

Fyfe, G., Fisher, R. I., Rosenberg, S. A., Sznol, M., Parkinson, D. R., and Louie, A. C. (1995). Results of treatment of 255 patients with metastatic renal cell carcinoma who received high-dose recombinant interleukin-2 therapy. *J. Clin. Oncol.* 13, 688–696. doi:10.1200/JCO.1995.13.3.688

Giberson, A. N., Davidson, A. R., and Parks, R. J. (2012). Chromatin structure of adenovirus DNA throughout infection. *Nucleic Acids Res.* 40, 2369–2376. doi:10.1093/NAR/GKR1076

Gill, S., Vasey, A. E., De Souza, A., Baker, J., Smith, A. T., Kohrt, H. E., et al. (2012). Rapid development of exhaustion and down-regulation of eomesodermin limit the antitumor activity of adoptively transferred murine natural killer cells. *Blood* 119, 5758–5768. doi:10.1182/BLOOD-2012-03-415364

Gillis, S., Baker, P. E., Ruscetti, F. W., and Smith, K. A. (1978). Long-term culture of human antigen-specific cytotoxic T-cell lines. *J. Exp. Med.* 148, 1093–1098. doi:10.1084/JEM.148.4.1093

Gillis, S., and Smith, K. A. (1977). Long term culture of tumour-specific cytotoxic T cells. *Nature* 268, 154–156. doi:10.1038/268154A0

Ginsberg, H. S., and Prince, G. A. (1994). The molecular basis of adenovirus pathogenesis. *Infect. Agents Dis.* 3, 1–8.

Goebeler, M. E., and Bargou, R. (2016). Blinatumomab: A CD19/CD3 bispecific T cell engager (BiTE) with unique anti-tumor efficacy. *Leuk. Lymphoma* 57, 1021–1032. doi:10.3109/10428194.2016.1161185

Grote, D., Russell, S. J., Cornu, T. I., Cattaneo, R., Vile, R., Poland, G. A., et al. (2001). Live attenuated measles virus induces regression of human lymphoma xenografts in immunodeficient mice. *Blood* 97, 3746–3754. doi:10.1182/BLOOD.V97.12.3746

Gusterson, B., Cowley, G., Smith, J. A., and Ozanne, B. (1984). Cellular localisation of human epidermal growth factor receptor. *Cell. Biol. Int. Rep.* 8, 649–658. doi:10.1016/0309-1651(84)90045-6

Hamdan, F., Ylosmaki, E., Chiaro, J., Giannoula, Y., Long, M., Fusciello, M., et al. (2021). Novel oncolytic adenovirus expressing enhanced cross-hybrid IgGA Fc PD-L1 inhibitor activates multiple immune effector populations leading to enhanced tumor killing *in vitro, in vivo* and with patient-derived tumor organoids. *J. Immunother. Cancer* 9, e003000. doi:10.1136/JITC-2021-003000

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell.* 144, 646–674. doi:10.1016/J.CELL.2011.02.013

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell.* 144, 646–674. doi:10.1016/j.cell.2011.02.013

Hanahan, D., and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell.* 100, 57–70. doi:10.1016/S0092-8674(00)81683-9

Hardy, S., Kitamura, M., Harris-Stansil, T., Dai, Y., and Phipps, M. L. (1997). Construction of adenovirus vectors through Cre-lox recombination. *J. Virol.* 71, 1842–1849. doi:10.1128/JVI.71.3.1842-1849.1997

Harrington, S. E., and Smith, T. J. (2008). The role of chemotherapy at the end of life: "When is enough, enough?". *JAMA J. Am. Med. Assoc.* 299, 2667–2678. doi:10.1001/JAMA.299.22.2667

Harris, D. T., and Kranz, D. M. (2016). Adoptive T cell therapies: A comparison of T cell receptors and chimeric antigen receptors. *Trends Pharmacol. Sci.* 37, 220–230. doi:10.1016/J.TIPS.2015.11.004

Hatjiharissi, E., Xu, L., Santos, D. D., Hunter, Z. R., Ciccarelli, B. T., Verselis, S., et al. (2007). Increased natural killer cell expression of CD16, augmented binding and ADCC activity to rituximab among individuals expressing the Fc{gamma}RIIIa-158 V/V and V/F polymorphism. *Blood* 110, 2561–2564. doi:10.1182/BLOOD-2007-01-070656

He, T. C., Zhou, S., Da Costa, L. T., Yu, J., Kinzler, K. W., and Vogelstein, B. (1998). A simplified system for generating recombinant adenoviruses. *Proc. Natl. Acad. Sci. U. S. A.* 95, 2509–2514. doi:10.1073/PNAS.95.5.2509

Heemskerk, N., Gruijs, M., Robin Temming, A., Heineke, M. H., Gout, D. Y., Hellingman, T., et al. (2021). Augmented antibody-based anticancer therapeutics boost neutrophil cytotoxicity. *J. Clin. Investig.* 131, e134680. doi:10.1172/JCI134680

Heise, C., Hermiston, T., Johnson, L., Brooks, G., Sampson-Johannes, A., Williams, A., et al. (2000). An adenovirus E1A mutant that demonstrates potent and selective systemic anti-tumoral efficacy. *Nat. Med.* 6 (6), 101134–101139. doi:10.1038/80474

Heise, C., and Kirn, D. H. (2000). Replication-selective adenoviruses as oncolytic agents. *J. Clin. Investigation* 105, 847–851. doi:10.1172/JCI9762

Heise, C., Sampson-Johannes, A., Williams, A., McCormick, F., Von Hoff, D. D., and Kirn, D. H. (1997). ONYX-015, an E1B gene-attenuated adenovirus, causes tumor-specific cytolysis and antitumoral efficacy that can be augmented by standard chemotherapeutic agents. *Nat. Med.* 3 (3), 6639–6645. doi:10.1038/nm0697-639

Hemminki, O., Parviainen, S., Juhila, J., Turkki, R., Linder, N., Lundin, J., et al. (2015). Immunological data from cancer patients treated with Ad5/3-E2F-Δ24-GMCSF suggests utility for tumor immunotherapy. *Oncotarget* 6, 4467–4481. doi:10.18632/ONCOTARGET.2901

Heron, M. (2019). *National vital statistics reports volume 68, number 6, June 24, 2019, deaths: Leading causes for 2017.*

Hershberg, U., Prak, Luning, et al. (2015). The analysis of clonal expansions in normal and autoimmune B cell repertoires. *Philosophical Trans. R. Soc. B Biol. Sci.* 370, 20140239. doi:10.1098/RSTB.2014.0239

Hierholzer, J. C. (1992). Adenoviruses in the immunocompromised host. *Clin. Microbiol. Rev.* 5, 262–274. doi:10.1128/CMR.5.3.262

Hino, R., Kabashima, K., Kato, Y., Yagi, H., Nakamura, M., Honjo, T., et al. (2010). Tumor cell expression of programmed cell death-1 ligand 1 is a prognostic factor for malignant melanoma. *Cancer* 116, 1757–1766. doi:10.1002/cncr.24899

Hodi, F. S., Butler, M., Oble, D. A., Seiden, M. V., Haluska, F. G., Kruse, A., et al. (2008). Immunologic and clinical effects of antibody blockade of cytotoxic T lymphocyte-associated antigen 4 in previously vaccinated cancer patients. *Proc. Natl. Acad. Sci.* 105, 3005–3010. doi:10.1073/pnas.0712237105

Hodi, F. S., O'Day, S. J., McDermott, D. F., Weber, R. W., Sosman, J. A., Haanen, J. B., et al. (2010). Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* 363, 711–723. doi:10.1056/NEJMoa1003466

Hoeben, R. C., and Uil, T. G. (2013). Adenovirus DNA replication. *Cold Spring Harb. Perspect. Biol.* 5, a013003. doi:10.1101/CSHPERSPECT.A013003

Hollingsworth, R. E., and Jansen, K. (20194201). Turning the corner on therapeutic cancer vaccines. *npj Vaccines* 14, 7–10. doi:10.1038/s41541-019-0103-y

Huber, S., Hoffmann, R., Muskens, F., and Voehringer, D. (2010). Alternatively activated macrophages inhibit T-cell proliferation by Stat6-dependent expression of PD-L2. *Blood* 116, 3311–3320. doi:10.1182/blood-2010-02-271981

Hubert, P., Heitzmann, A., Viel, S., Nicolas, A., Sastre-Garau, X., Oppezzo, P., et al. (2011). Antibody-dependent cell cytotoxicity synapses form in mice during tumor-specific antibody immunotherapy. *Cancer Res.* 71, 5134–5143. doi:10.1158/0008-5472.CAN-10-4222

Huls, G., Heijnen, I. A. F. M., Cuomo, E., Van Der Linden, J., Boel, E., Van De Winkel, J. G. J., et al. (1999). Antitumor immune effector mechanisms recruited by phage display-derived fully human IgG1 and IgA1 monoclonal antibodies. *Cancer Res.* 59, 5778–5784.

June, C. H., O'Connor, R. S., Kawalekar, O. U., Ghassemi, S., and Milone, M. C. (2018). CAR T cell immunotherapy for human cancer. *Science* 359, 1361–1365. doi:10.1126/SCIENCE.AAR6711

Keir, M. E., Butte, M. J., Freeman, G. J., and Sharpe, A. H. (2008). PD-1 and its ligands in tolerance and immunity. *Annu. Rev. Immunol.* 26, 677–704. doi:10.1146/annurev.immunol.26.021607.090331

Khunger, A., Rytlewski, J. A., Fields, P., Yusko, E. C., and Tarhini, A. A. (2019). The impact of CTLA-4 blockade and interferon-α on clonality of T-cell repertoire in the tumor microenvironment and peripheral blood of metastatic melanoma patients. *Oncoimmunology* 8, e1652538. doi:10.1080/2162402X.2019.1652538

Kim, R., Emi, M., and Tanabe, K. (2007). Cancer immunoediting from immune surveillance to immune escape. *Immunology* 121, 1–14. doi:10.1111/J.1365-2567.2007.02587.X

Köhler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256, 495–497. doi:10.1038/256495A0

Koprowski, H., Steplewski, Z., Herlyn, D., and Herlyn, M. (1978). Study of antibodies against human melanoma produced by somatic cell hybrids. *Proc. Natl. Acad. Sci. U. S. A.* 75, 3405–3409. doi:10.1073/PNAS.75.7.3405

Krummel, M. F., and Allison, J. P. (1996). CTLA-4 engagement inhibits IL-2 accumulation and cell cycle progression upon activation of resting T cells. *J. Exp. Med.* 183, 2533–2540. doi:10.1084/jem.183.6.2533

Kurbacher, C., Kurbacher, J., Cramer, E., Rhiem, K., Mallman, P., Reichelt, R., et al. (2005). *Continuous low-dose GM-CSF as salvage therapy in refractory recurrent breast or female genital tract carcinoma.* undefined.

Lawler, S. E., Speranza, M. C., Cho, C. F., and Chiocca, E. A. (2017). Oncolytic viruses in cancer treatment: A review. *JAMA Oncol.* 3, 841–849. doi:10.1001/JAMAONCOL.2016.2064

Le, D. T., Lutz, E., Uram, J. N., Sugar, E. A., Onners, B., Solt, S., et al. (2013). Evaluation of ipilimumab in combination with allogeneic pancreatic tumor cells transfected with a GM-CSF gene in previously treated pancreatic cancer. *J. Immunother.* 36, 382–389. doi:10.1097/CJI.0b013e31829fb7a2

Leach, D. R., Krummel, M. F., and Allison, J. P. (1996). Enhancement of antitumor immunity by CTLA-4 blockade. *Science* 271, 1734–1736. doi:10.1126/science.271.5256.1734

Lee, K. L., Schlom, J., and Hamilton, D. H. (2021). Combination therapies utilizing neoepitope-targeted vaccines. *Cancer Immunol. Immunother.* 70, 875–885. doi:10.1007/S00262-020-02729-Y

Lehmann, C., Zeis, M., and Uharek, L. (2001). Activation of natural killer cells with interleukin 2 (IL-2) and IL-12 increases perforin binding and subsequent lysis of tumour cells. *Br. J. Haematol.* 114, 660–665. doi:10.1046/J.1365-2141.2001.02995.X

Lemos de Matos, A., Franco, L. S., and McFadden, G. (2020). Oncolytic viruses and the immune system: The dynamic duo. *Mol. Ther. Methods Clin. Dev.* 17, 349–358. doi:10.1016/J.OMTM.2020.01.001

Leusen, J. H. W. (2015). IgA as therapeutic antibody. *Mol. Immunol.* 68, 35–39. doi:10.1016/j.molimm.2015.09.005

Li, S., Schmitz, K. R., Jeffrey, P. D., Wiltzius, J. J. W., Kussie, P., and Ferguson, K. M. (2005). Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell.* 7, 301–311. doi:10.1016/J.CCR.2005.03.003

Linsley, P. S., Bradshaw, J., Greene, J., Peach, R., Bennett, K. L., and Mittler, R. S. (1996). Intracellular trafficking of CTLA-4 and focal localization towards sites of TCR engagement. *Immunity* 4, 535–543. doi:10.1016/s1074-7613(00)80480-x

Liu, B. L., Robinson, M., Han, Z. Q., Branston, R. H., English, C., Reay, P., et al. (2003). ICP34.5 deleted herpes simplex virus with enhanced oncolytic, immune stimulating, and anti-tumour properties. *Gene Ther.* 10, 292–303. doi:10.1038/SJ.GT.3301885

Lohse, S., Brunke, C., Derer, S., Peipp, M., Boross, P., Kellner, C., et al. (2012). Characterization of a mutated IgA2 antibody of the m(1) allotype against the epidermal growth factor receptor for the recruitment of monocytes and macrophages. *J. Biol. Chem.* 287, 25139–25150. doi:10.1074/jbc.M112.353060

Lohse, S., Derer, S., Beyer, T., Klausz, K., Peipp, M., Leusen, J. H. W., et al. (2011). Recombinant dimeric IgA antibodies against the epidermal growth factor receptor mediate effective tumor cell killing. *J. Immunol.* 186, 3770–3778. doi:10.4049/jimmunol.1003082

Lomonosova, E., Subramanian, T., and Chinnadurai, G. (2005). Mitochondrial localization of p53 during adenovirus infection and regulation of its activity by E1B-19K. *Oncogene* 24 (24), 456796–456808. doi:10.1038/sj.onc.1208836

Loskog, A., Maleka, A., Mangsbo, S., Svensson, E., Lundberg, C., Nilsson, A., et al. (2016). Immunostimulatory AdCD40L gene therapy combined with low-dose cyclophosphamide in metastatic melanoma patients. *Br. J. Cancer* 114, 872–880. doi:10.1038/BJC.2016.42

Malmström, P. U., Loskog, A. S. I., Lindqvist, C. A., Mangsbo, S. M., Fransson, M., Wanders, A., et al. (2010). AdCD40L immunogene therapy for bladder carcinoma - the first phase I/IIa trial. *Clin. Cancer Res.* 16, 3279–3287. doi:10.1158/1078-0432.CCR-10-0385

Marttila, M., Persson, D., Gustafsson, D., Liszewski, M. K., Atkinson, J. P., Wadell, G., et al. (2005). CD46 is a cellular receptor for all species B adenoviruses except types 3 and 7. *J. Virol.* 79, 14429–14436. doi:10.1128/JVI.79.22.14429-14436.2005

Masteller, E. L., Chuang, E., Mullen, A. C., Reiner, S. L., and Thompson, C. B. (2000). Structural analysis of CTLA-4 function *in vivo*. *J. Immunol.* 164, 5319–5327. doi:10.4049/jimmunol.164.10.5319

Masucci, M. T., Minopoli, M., and Carriero, M. V. (2019). Tumor associated neutrophils. Their role in tumorigenesis, metastasis, prognosis and therapy. *Front. Oncol.* 9, 1146. doi:10.3389/FONC.2019.01146

Meldrum, C., Doyle, M. A., and Tothill, R. W. (2011). Next-generation sequencing for cancer diagnostics: A practical perspective. *Clin. Biochem. Rev.* 32, 177–195.

Melis, J. P. M., Strumane, K., Ruuls, S. R., Beurskens, F. J., Schuurman, J., and Parren, P. W. H. I. (2015). Complement in therapy and disease: Regulating the complement system with antibody-based therapeutics. *Mol. Immunol.* 67, 117–130. doi:10.1016/J.MOLIMM.2015.01.028

Mittal, S. K., McDermott, M. R., Johnson, D. C., Prevec, L., and Graham, F. L. (1993). Monitoring foreign gene expression by a human adenovirus-based vector using the firefly luciferase gene as a reporter. *Virus Res.* 28, 67–90. doi:10.1016/0168-1702(93)90090-A

Möller, E. (1965). CONTACT-INDUCED cytotoxicity by lymphoid cells containing foreign isoantigens. *Science* 147, 873–879. doi:10.1126/SCIENCE.147.3660.873

Murata, Y., Tanaka, D., Hazama, D., Yanagita, T., Saito, Y., Kotani, T., et al. (2018). Anti-human SIRPα antibody is a new tool for cancer immunotherapy. *Cancer Sci.* 109, 1300–1308. doi:10.1111/CAS.13548

Musolino, A., Naldi, N., Bortesi, B., Pezzuolo, D., Capelletti, M., Missale, G., et al. (2008). Immunoglobulin G fragment C receptor polymorphisms and clinical efficacy of trastuzumab-based therapy in patients with HER-2/neu-positive metastatic breast cancer. *J. Clin. Oncol.* 26, 1789–1796. doi:10.1200/JCO.2007.14.8957

Nadler, L. M., Stashenko, P., Hardy, R., Kaplan, W. D., Button, L. N., Kufe, D. W., et al. (1980). Serotherapy of a patient with a monoclonal antibody directed against a human lymphoma-associated antigen. *Cancer Res.* 40, 3147–3154.

Nelson, A. L., Dhimolea, E., and Reichert, J. M. (2010). Development trends for human monoclonal antibody therapeutics. *Nat. Rev. Drug Discov.* 9, 767–774. doi:10.1038/NRD3229

Nelson, A. L., and Reichert, J. M. (2009). Development trends for therapeutic antibody fragments. *Nat. Biotechnol.* 27, 331–337. doi:10.1038/NBT0409-331

Nevins, J. R., and Chen-Kiang, S. (1981). Processing of adenovirus nuclear RNA to mRNA. *Adv. Virus Res.* 26, 1–35. doi:10.1016/S0065-3527(08)60419-4

Nevins, J. R., and Darnell, J. E. (1978). Steps in the processing of Ad2 mRNA: poly(A)+ nuclear sequences are conserved and poly(A) addition precedes splicing. *Cell.* 15, 1477–1493. doi:10.1016/0092-8674(78)90071-5

Nimmerjahn, F., Gordan, S., and Lux, A. (2015). FcγR dependent mechanisms of cytotoxic, agonistic, and neutralizing antibody activities. *Trends Immunol.* 36, 325–336. doi:10.1016/J.IT.2015.04.005

Nimmerjahn, F., and Ravetch, J. V. (2008). Analyzing antibody-Fc-receptor interactions. *Methods Mol. Biol.* 415, 151–162. doi:10.1007/978-1-59745-570-1_9

Nimmerjahn, F., and Ravetch, J. V. (2008). Fcgamma receptors as regulators of immune responses. *Nat. Rev. Immunol.* 8, 34–47. doi:10.1038/NRI2206

Nossal, G. J. V., and Lederberg, J. (1958). Antibody production by single cells. *Nature* 181, 1419–1420. doi:10.1038/1811419A0

Ostapchuk, P., and Hearing, P. (2003). Regulation of adenovirus packaging. *Curr. Top. Microbiol. Immunol.* 272, 165–185. doi:10.1007/978-3-662-05597-7_6

Overdijk, M. B., Verploegen, S., Bögels, M., Van Egmond, M., Lammerts Van Bueren, J. J., Mutis, T., et al. (2015). Antibody-mediated phagocytosis contributes to the anti-tumor activity of the therapeutic antibody daratumumab in lymphoma and multiple myeloma. *MAbs* 7, 311–321. doi:10.1080/19420862.2015.1007813

Pachella, R. N., Msn, A. G. P. C. N. P-B. C., Aocnp®, L. A., adsen, A. G. P. C. N. P-B. C., PhD, R. N., Aocns®, L. T., et al. (2015). The toxicity and benefit of various dosing strategies for interleukin-2 in metastatic melanoma and renal cell carcinoma. *J. Adv. Pract. Oncol.* 6, 212. doi:10.6004/jadpro.2015.6.3.3

Pakkanen, S. H., Kantele, J. M., Moldoveanu, Z., Hedges, S., Hakkinen, M., Mestecky, J., et al. (2010). Expression of homing receptors on IgA1 and IgA2 plasmablasts in blood reflects differential distribution of IgA1 and IgA2 in various body fluids. *Clin. Vaccine Immunol.* 17, 393–401. doi:10.1128/CVI.00475-09

Paliard, X., de Waal Malefijt, R., Yssel, H., Blanchard, D., Chrétien, I., Abrams, J., et al. (1988). Simultaneous production of IL-2, IL-4, and IFN-gamma by activated human CD4+ and CD8+ T cell clones. *J. Immunol.* 141, 849–855. doi:10.4049/jimmunol.141.3.849

Parry, R. V., Chemnitz, J. M., Frauwirth, K. A., Lanfranco, A. R., Braunstein, I., Kobayashi, S. V., et al. (2005). CTLA-4 and PD-1 receptors inhibit T-cell activation by distinct mechanisms. *Mol. Cell. Biol.* 25, 9543–9553. doi:10.1128/MCB.25.21.9543-9553.2005

Pascal, V., Laffleur, B., Debin, A., Cuvillier, A., van Egmond, M., Drocourt, D., et al. (2012). Anti-CD20 IgA can protect mice against lymphoma development: Evaluation of the direct impact of IgA and cytotoxic effector recruitment on CD20 target cells. *Haematologica* 97, 1686–1694. doi:10.3324/haematol.2011.061408

Patel, D., Bassi, R., Hooper, A., Prewett, M., Hicklin, D. J., and Kang, X. (2009). Anti-epidermal growth factor receptor monoclonal antibody cetuximab inhibits EGFR/HER-2 heterodimerization and activation. *Int. J. Oncol.* 34, 25–32. doi:10.3892/IJO_00000125/HTML

Peipp, M., Van Bueren, J. J. L., Schneider-Merck, T., Bleeker, W. W. K., Dechant, M., Beyer, T., et al. (2008). Antibody fucosylation differentially impacts cytotoxicity mediated by NK and PMN effector cells. *Blood* 112, 2390–2399. doi:10.1182/BLOOD-2008-03-144600

Perica, K., Varela, J. C., Oelke, M., and Schneck, J. (2015). Adoptive T cell immunotherapy for cancer. *Rambam Maimonides Med. J.* 6, e0004. doi:10.5041/RMMJ.10179

Pesonen, S., Diaconu, I., Kangasniemi, L., Ranki, T., Kanerva, A., Pesonen, S. K., et al. (2012). Oncolytic immunotherapy of advanced solid tumors with a CD40L-expressing replicating adenovirus: Assessment of safety and immunologic responses in patients. *Cancer Res.* 72, 1621–1631. doi:10.1158/0008-5472.CAN-11-3001

Piechutta, M., and Berghoff, A. S. (2019). New emerging targets in cancer immunotherapy: The role of cluster of differentiation 40 (CD40/TNFR5). *ESMO Open* 4, e000510. doi:10.1136/ESMOOPEN-2019-000510

Piya, S., White, E. J., Klein, S. R., Jiang, H., McDonnell, T. J., Gomez-Manzano, C., et al. (2011). The E1B19K oncoprotein complexes with beclin 1 to regulate autophagy in adenovirus-infected cells. *PLoS One* 6, e29467. doi:10.1371/JOURNAL.PONE.0029467

Plosker, G. L., and Keam, S. J. (2006). Trastuzumab: A review of its use in the management of HER2-positive metastatic and early-stage breast cancer. *Drugs* 66, 449–475. doi:10.2165/00003495-200666040-00005

Plosker, G. L. (2011). Sipuleucel-T: In metastatic castration-resistant prostate cancer. *Drugs* 71, 101–108. doi:10.2165/11206840-000000000-00000

Postow, M. A., Callahan, M. K., and Wolchok, J. D. (2015). Immune checkpoint blockade in cancer therapy. *J. Clin. Oncol.* 33, 1974–1982. –82. doi:10.1200/JCO.2014.59.4358

Postow, M. A., Callahan, M. K., and Wolchok, J. D. (2015). Immune checkpoint blockade in cancer therapy. *J. Clin. Oncol.* 33, 1974–1982. –82. doi:10.1200/JCO.2014.59.4358

Proietti, E., Greco, G., Garrone, B., Baccarini, S., Mauri, C., Venditti, M., et al. (1998). Importance of cyclophosphamide-induced bystander effect on T cells for a successful tumor eradication in response to adoptive immunotherapy in mice. *J. Clin. Investig.* 101, 429–441. doi:10.1172/JCI1348

Ramke, M., Lee, J. Y., Dyer, D. W., Seto, D., Rajaiya, J., and Chodosh, J. (2017). The 5′UTR in human adenoviruses: Leader diversity in late gene expression. *Sci. Rep.* 7, 618. doi:10.1038/S41598-017-00747-Y

Ranki, T., Pesonen, S., Hemminki, A., Partanen, K., Kairemo, K., Alanko, T., et al. (2016). Phase I study with ONCOS-102 for the treatment of solid tumors – An evaluation of clinical response and exploratory analyses of immune markers. *J. Immunother. Cancer* 4, 17. doi:10.1186/S40425-016-0121-5

Rekosh, D. M. K., Russell, W. C., Bellet, A. J. D., and Robinson, A. J. (1977). Identification of a protein linked to the ends of adenovirus DNA. *Cell.* 11, 283–295. doi:10.1016/0092-8674(77)90045-9

Restifo, N. P., Dudley, M. E., and Rosenberg, S. A. (2012). Adoptive immunotherapy for cancer: harnessing the T cell response. *Nat. Rev. Immunol.* 12, 269–281. doi:10.1038/NRI3191

Ribas, A., Camacho, L. H., Lopez-Berestein, G., Pavlov, D., Bulanhagui, C. A., Millham, R., et al. (2005). Antitumor activity in melanoma and anti-self responses in a phase I trial with the anti-cytotoxic T lymphocyte-associated antigen 4 monoclonal antibody CP-675,206. *J. Clin. Oncol.* 23, 8968–8977. doi:10.1200/JCO.2005.01.109

Ribas, A., Kefford, R., Marshall, M. A., Punt, C. J. A., Haanen, J. B., Marmol, M., et al. (2013). Phase III randomized clinical trial comparing tremelimumab with standard-of-care chemotherapy in patients with advanced melanoma. *J. Clin. Oncol.* 31, 616–622. doi:10.1200/JCO.2012.44.6112

Riechmann, L., Clark, M., Waldmann, H., and Winter, G. (1988). Reshaping human antibodies for therapy. *Nature* 332, 323–327. doi:10.1038/332323A0

Robert, C., Thomas, L., Bondarenko, I., O'Day, S., Weber, J., Garbe, C., et al. (2011). Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N. Engl. J. Med.* 364, 2517–2526. doi:10.1056/NEJMoa1104621

Robert, L., Tsoi, J., Wang, X., Emerson, R., Homet, B., Chodon, T., et al. (2014). CTLA4 blockade broadens the peripheral T-cell receptor repertoire. *Clin. Cancer Res.* 20, 2424–2432. doi:10.1158/1078-0432.CCR-13-2648

Rodríguez, J., Zarate, R., Bandres, E., Boni, V., Hernández, A., Sola, J. J., et al. (2012). Fc gamma receptor polymorphisms as predictive markers of Cetuximab efficacy in epidermal growth factor receptor downstream-mutated metastatic colorectal cancer. *Eur. J. Cancer* 48, 1774–1780. doi:10.1016/J.EJCA.2012.01.007

Rosenberg, S. A., and Restifo, N. P. (2015). Adoptive cell transfer as personalized immunotherapy for human cancer. *Science* 348, 62–68. doi:10.1126/SCIENCE.AAA4967

Rosenberg, S. A., Restifo, N. P., Yang, J. C., Morgan, R. A., and Dudley, M. E. (2008). Adoptive cell transfer: A clinical path to effective cancer immunotherapy. *Nat. Rev. Cancer* 8, 299–308. doi:10.1038/NRC2355

Rosenberg, S. A., Yannelli, J. R., Yang, J. C., Topalian, S. L., Schwartzentruber, D. J., Weber, J. S., et al. (1994). Treatment of patients with metastatic melanoma with autologous tumor-infiltrating lymphocytes and interleukin 2. *J. Natl. Cancer Inst.* 86, 1159–1166. doi:10.1093/JNCI/86.15.1159

Sahin, U., Oehm, P., Derhovanessian, E., Jabulowsky, R. A., Vormehr, M., Gold, M., et al. (2020). An RNA vaccine drives immunity in checkpoint-inhibitor-treated melanoma. *Nature* 585, 7823585107–7823585112. doi:10.1038/s41586-020-2537-9

Sansom, D. M. (2000). CD28, CTLA-4 and their ligands: Who does what and to whom? *Immunology* 101, 169–177. doi:10.1046/J.1365-2567.2000.00121.X

Schiza, A., Wenthe, J., Mangsbo, S., Eriksson, E., Nilsson, A., Tötterman, T. H., et al. (2017). Adenovirus-mediated CD40L gene transfer increases Teffector/Tregulatory cell ratio and upregulates death receptors in metastatic melanoma patients. *J. Transl. Med.* 15, 79. doi:10.1186/S12967-017-1182-Z

Schneider, H., Downey, J., Smith, A., Zinselmeyer, B. H., Rush, C., Brewer, J. M., et al. (2006). Reversal of the TCR stop signal by CTLA-4. *Science* 313 (1979), 1972–1975. doi:10.1126/science.1131078

Schwaber, J., and Cohen, E. P. (1973). Human x mouse somatic cell hybrid clone secreting immunoglobulins of both parental types. *Nature* 244, 444–447. doi:10.1038/244444A0

Scott, E. M., , et al.Scott, E. M., Duffy, M. R., Freedman, J. D., Fisher, K. D., Seymour, L. W. (2018). Solid tumor immunotherapy with T cell engager-armed oncolytic viruses. *Macromol. Biosci.* 18, 1700187. doi:10.1002/MABI.201700187

Seliger, B. (2005). Strategies of tumor immune evasion. *BioDrugs* 19, 347–354. doi:10.2165/00063030-200519060-00002

Selvaraj, P., Rosse, W. F., Silber, R., and Springer, T. A. (1988). The major Fc receptor in blood has a phosphatidylinositol anchor and is deficient in paroxysmal nocturnal haemoglobinuria. *Nature* 333, 565–567. doi:10.1038/333565A0

Sha, J., Ghosh, M. K., Zhang, K., and Harter, M. L. (2010). E1A interacts with two opposing transcriptional pathways to induce quiescent cells into S phase. *J. Virol.* 84, 4050–4059. doi:10.1128/JVI.02131-09/ASSET/B26DCBCF-2ACD-4DCA-8A8C-A418B02E0AE0/ASSETS/GRAPHIC/ZJV9990930500007.JPEG

Sharma, P., Hu-Lieskovan, S., Wargo, J. A., and Ribas, A. (2017). Primary, adaptive, and acquired resistance to cancer immunotherapy. *Cell.* 168, 707–723. doi:10.1016/J.CELL.2017.01.017

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA Cancer J. Clin.* 70, 7–30. doi:10.3322/CAAC.21590

Sioud, M., Westby, P., Olsen, J. K. E., and Mobergslien, A. (2015). Generation of new peptide-Fc fusion proteins that mediate antibody-dependent cellular cytotoxicity against different types of cancer cells. *Mol. Ther. Methods Clin. Dev.* 2, 15043. doi:10.1038/MTM.2015.43

Slamon, D. J., Godolphin, W., Jones, L. A., Holt, J. A., Wong, S. G., Keith, D. E., et al. (1989). Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 244, 707–712. doi:10.1126/SCIENCE.2470152

Slovin, S. F., Higano, C. S., Hamid, O., Tejwani, S., Harzstark, A., Alumkal, J. J., et al. (2013). Ipilimumab alone or in combination with radiotherapy in metastatic castration-resistant prostate cancer: Results from an open-label, multicenter phase I/II study. *Ann. Oncol.* 24, 1813–1821. doi:10.1093/annonc/mdt107

Smart, J. E., and Stillman, B. W. (1982). Adenovirus terminal protein precursor. Partial amino acid sequence and the site of covalent linkage to virus DNA. *J. Biol. Chem.* 257, 13499–13506. doi:10.1016/S0021-9258(18)33475-6

Smith, K. A. (1988). Interleukin-2: Inception, impact, and implications. *Science* 240, 1169–1176. doi:10.1126/SCIENCE.3131876

Smyth, M. J., Thia, K. Y., Street, S. E., MacGregor, D., Godfrey, D. I., and Trapani, J. A. (2000). Perforin-mediated cytotoxicity is critical for surveillance of spontaneous lymphoma. *J. Exp. Med.* 192, 755–760. doi:10.1084/JEM.192.5.755

Stanley, E., Lieschke, G. J., Grail, D., Metcalf, D., Hodgson, G., Gall, J. A. M., et al. (1994). Granulocyte/macrophage colony-stimulating factor-deficient mice show no major perturbation of hematopoiesis but develop a characteristic pulmonary pathology. *Proc. Natl. Acad. Sci. U. S. A.* 91, 5592–5596. doi:10.1073/PNAS.91.12.5592

Stratford-Perricaudet, L. D., Makeh, I., Perricaudet, M., and Briand, P. (1992). Widespread long-term gene transfer to mouse skeletal muscles and heart. *J. Clin. Investig.* 90, 626–630. doi:10.1172/JCI115902

Strohl, W. R., and Knight, D. M. (2009). Discovery and development of biopharmaceuticals: Current issues. *Curr. Opin. Biotechnol.* 20, 668–672. doi:10.1016/J.COPBIO.2009.10.012

Strohl, W. R. (2009). Optimization of Fc-mediated effector functions of monoclonal antibodies. *Curr. Opin. Biotechnol.* 20, 685–691. doi:10.1016/J.COPBIO.2009.10.011

Suzuki, T., Ishii-Watabe, A., Tada, M., Kobayashi, T., Kanayasu-Toyoda, T., Kawanishi, T., et al. (2010). Importance of neonatal FcR in regulating the serum half-life of therapeutic proteins containing the Fc domain of human IgG1: A comparative study of the affinity of monoclonal antibodies and fc-fusion proteins to human neonatal FcR. *J. Immunol.* 184, 1968–1976. doi:10.4049/JIMMUNOL.0903296

Swann, J. B., and Smyth, M. J. (2007). Immune surveillance of tumors. *J. Clin. Investigation* 117, 1137–1146. doi:10.1172/JCI31405

Takahashi, T., Tagami, T., Yamazaki, S., Uede, T., Shimizu, J., Sakaguchi, N., et al. (2000). Immunologic self-tolerance maintained by CD25(+)CD4(+) regulatory T cells constitutively expressing cytotoxic T lymphocyte-associated antigen 4. *J. Exp. Med.* 192, 303–310. doi:10.1084/jem.192.2.303

Taube, J. M., Klein, A., Brahmer, J. R., Xu, H., Pan, X., Kim, J. H., et al. (2014). Association of PD-1, PD-1 ligands, and other features of the tumor immune microenvironment with response to anti-PD-1 therapy. *Clin. Cancer Res.* 20, 5064–5074. doi:10.1158/1078-0432.CCR-13-3271

Taylor, R. P., and Lindorfer, M. A. (2016). Cytotoxic mechanisms of immunotherapy: Harnessing complement in the action of anti-tumor monoclonal antibodies. *Semin. Immunol.* 28, 309–316. doi:10.1016/J.SMIM.2016.03.003

Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., et al. (2012). Safety, activity, and immune correlates of anti–PD-1 antibody in cancer. *N. Engl. J. Med.* 366, 2443–2454. doi:10.1056/NEJMoa1200690

Topalian, S. L., Sznol, M., McDermott, D. F., Kluger, H. M., Carvajal, R. D., Sharfman, W. H., et al. (2014). Survival, durable tumor remission, and long-term safety in patients with advanced melanoma receiving nivolumab. *J. Clin. Oncol.* 32, 1020–1030. doi:10.1200/JCO.2013.53.0105

Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., et al. (1984). Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* 309, 418–425. doi:10.1038/309418A0

Vajdic, C. M., and van Leeuwen, M. T. (2009). Cancer incidence and risk factors after solid organ transplantation. *Int. J. Cancer* 125, 1747–1754. doi:10.1002/ijc.24439

Valerius, T., Stockmeyer, B., van Spriel, A. B., Graziano, R. F., van den Herik-Oudijk, I. E., Repp, R., et al. (1997). FcαRI (CD89) as a novel trigger molecule for bispecific antibody therapy. *Blood* 90, 4485–4492. doi:10.1182/blood.v90.11.4485.4485_4485_4492

Van Der Bij, G. J., Bögels, M., Otten, M. A., Oosterling, S. J., Kuppen, P. J., Meijer, S., et al. (2010). Experimentally induced liver metastases from colorectal cancer can be prevented by mononuclear phagocyte-mediated monoclonal antibody therapy. *J. Hepatol.* 53, 677–685. doi:10.1016/J.JHEP.2010.04.023

van der Kolk, L. E., de Haas, M., Grillo-López, A. J., Baars, J. W., and van Oers, M. H. J. (2002). Analysis of CD20-dependent cellular cytotoxicity by G-CSF-stimulated neutrophils. *Leukemia* 16, 693–699. doi:10.1038/SJ.LEU.2402424

van Elsas, A., Hurwitz, A. A., and Allison, J. P. (1999). Combination immunotherapy of B16 melanoma using anti-cytotoxic T lymphocyte-associated antigen 4 (CTLA-4) and granulocyte/macrophage colony-stimulating factor (GM-CSF)-producing vaccines induces rejection of subcutaneous and metastatic tumors accompanied by autoimmune depigmentation. *J. Exp. Med.* 190, 355–366. doi:10.1084/jem.190.3.355

Van Epps, H. L. (2005). How heidelberger and avery sweetened immunology. *J. Exp. Med.* 202, 1306. doi:10.1084/JEM20210FTA

Vargas, F. A., Furness, A. J. S., Litchfield, K., Joshi, K., Rosenthal, R., Ghorani, E., et al. (2018). Fc effector function contributes to the activity of human anti-CTLA-4 antibodies. *Cancer Cell.* 33, 649–663.e4. doi:10.1016/J.CCELL.2018.02.010

Vinay, D. S., Ryan, E. P., Pawelec, G., Talib, W. H., Stagg, J., Elkord, E., et al. (2015). Immune evasion in cancer: Mechanistic basis and therapeutic strategies. *Semin. Cancer Biol.* 35, S185–S198. doi:10.1016/j.semcancer.2015.03.004

Wallace, P. K., Howell, A. L., and Fanger, M. W. (1994). Role of Fc gamma receptors in cancer and infectious disease. *J. Leukoc. Biol.* 55, 816–826. doi:10.1002/JLB.55.6.816

Walunas, T. L., Bluestone, J. A., Griffin, M. D., Sharpe, A. H., and Bluestone, J. A. (1998). CTLA-4 regulates tolerance induction and T cell differentiation *in vivo*. *J. Immunol.* 160, 3855–3860. doi:10.4049/jimmunol.160.8.3855

Wang, G., de Jong, R. N., van den Bremer, E. T. J., Beurskens, F. J., Labrijn, A. F., Ugurlar, D., et al. (2016). Molecular basis of assembly and activation of complement component C1 in complex with immunoglobulin G1 and antigen. *Mol. Cell.* 63, 135–145. doi:10.1016/J.MOLCEL.2016.05.016

Wang, G., Kang, X., Chen, K. S., Jehng, T., Jones, L., Chen, J., et al. (2020). An engineered oncolytic virus expressing PD-L1 inhibitors activates tumor neoantigen-specific T cell responses. *Nat. Commun.* 11, 1395111–1395114. doi:10.1038/s41467-020-15229-5

Wang, H., Li, Z. Y., Liu, Y., Persson, J., Beyer, I., Möller, T., et al. (2011). Desmoglein 2 is a receptor for adenovirus serotypes 3, 7, 11, and 14. *Nat. Med.* 17, 96–104. doi:10.1038/NM.2270

Wang, W., Erbe, A. K., Hank, J. A., Morris, Z. S., and Sondel, P. M. (2015). NK cell-mediated antibody-dependent cellular cytotoxicity in cancer immunotherapy. *Front. Immunol.* 6, 368. doi:10.3389/FIMMU.2015.00368

Weng, W. K., and Levy, R. (2003). Two immunoglobulin G fragment C receptor polymorphisms independently predict response to rituximab in patients with follicular lymphoma. *J. Clin. Oncol.* 21, 3940–3947. doi:10.1200/JCO.2003.05.013

Wherry, E. J. (2011). T cell exhaustion. *Nat. Immunol.* 12, 492–499. doi:10.1038/ni.2035

Wickham, T. J., Mathias, P., Cheresh, D. A., and Nemerow, G. R. (1993). Integrins αvβ3 and αvβ5 promote adenovirus internalization but not virus attachment. *Cell.* 73, 309–319. doi:10.1016/0092-8674(93)90231-E

Wu, E., Trauger, S. A., Pache, L., Mullen, T-M., Von Seggern, D. J., Siuzdak, G., et al. (2004). Membrane cofactor protein is a receptor for adenoviruses associated with epidemic keratoconjunctivitis. *J. Virol.* 78, 3897–3905. doi:10.1128/JVI.78.8.3897-3905.2004/ASSET/AF57C77A-6685-4943-BEC9-F129244166DA/ASSETS/GRAPHIC/ZJV0080418530004.JPEG

Wu, J., Edberg, J. C., Redecha, P. B., Bansal, V., Guyre, P. M., Coleman, K., et al. (1997). A novel polymorphism of FcgammaRIIIa (CD16) alters receptor function and predisposes to autoimmune disease. *J. Clin. Investig.* 100, 1059–1070. doi:10.1172/JCI119616

Wu, T., and Wu, L. (2021). The role and clinical implications of the retinoblastoma (RB)-E2F pathway in gastric cancer. *Front. Oncol.* 11, 1954. doi:10.3389/fonc.2021.655630

Yi, M., Niu, M., Xu, L., Luo, S., and Wu, K. (2021). Regulation of PD-L1 expression in the tumor microenvironment. *J. Hematol. Oncol.* 14, 1–13. doi:10.1186/S13045-020-01027-5

Ylösmäki, E., and Cerullo, V. (2020). Design and application of oncolytic viruses for cancer immunotherapy. *Curr. Opin. Biotechnol.* 65, 25–36. doi:10.1016/J.COPBIO.2019.11.016

Yu, F., Wang, X., Guo, Z. S., Bartlett, D. L., Gottschalk, S. M., and Song, X. T. (2014). T-Cell engager-armed oncolytic vaccinia virus significantly enhances antitumor therapy. *Mol. Ther.* 22, 102–111. doi:10.1038/MT.2013.240

Zhang, B. (2009). *Ofatumumab*. MAbs 1, 326–331. doi:10.4161/MABS.1.4.8895

Zhao, Y., Liu, Z., Li, L., Wu, J., Zhang, H., Zhang, H., et al. (2021). Oncolytic adenovirus: Prospects for cancer immunotherapy. *Front. Microbiol.* 12, 1951. doi:10.3389/fmicb.2021.707290

Check for updates

Novel insights into cancer stem cells targeting: CAR-T therapy and epigenetic drugs as new pillars in cancer treatment

Veronica Veschi[1†], Alice Turdo[2†] and Giorgio Stassi[1]*

[1]Department of Surgical, Oncological and Stomatological Sciences (DICHIRONS), University of Palermo, Palermo, Italy, [2]Department of Health Promotion, Mother and Child Care, Internal Medicine and Medical Specialties (PROMISE), University of Palermo, Palermo, Italy

Cancer stem cells (CSCs) represent the most aggressive subpopulation present in the tumor bulk retaining invasive capabilities, metastatic potential and high expression levels of drug efflux pumps responsible for therapy resistance. Cancer is still an incurable disease due to the inefficacy of standard regimens that spare this subpopulation. Selective targeting of CSCs is still an unmet need in cancer research field. Aberrant epigenetic reprogramming promotes the initiation and maintenance of CSCs, which are able to escape the immune system defense. Promising therapeutic approaches able to induce the selective inhibition of this stem-like small subset include immunotherapy alone or in combination with epigenetic compounds. These strategies are based on the specific expression of epitopes and/or epigenetic alterations present only in the CSC and not in the other cancer cells or normal cells. Thus, the combined approach utilizing CAR-T immunotherapy along with epigenetic probes may overcome the barriers of treatment ineffectiveness towards a more precision medicine approach in patients with known specific alterations of CSCs. In this perspective article we will shed new lights on the future applications of epi-immunotherapy in tumors enriched in CSCs, along with its potential side-effects, limitations and the development of therapy resistance.

KEYWORDS

CAR-T cell therapy, CAR-NK, epigenetic inhibitors, cancer stem cells, epigenetic drugs

Introduction

Tumors are a miscellaneous cell composition harboring a subpopulation of stem-like cells endowed with the capability of inducing tumor progression and drug resistance. Several mechanisms have been ascribed to these phenomena including the capability to evade the immune system responses (Turdo et al., 2019). Immunological characteristics specific of cancer stem cells (CSCs) such as the expression of tumor-associated antigens (TAAs), the secretion of cytokines and/or anti-apoptotic molecules along with the upregulation of STAT3 or PI3K/AKT survival signaling pathways, are able to increase resistance to apoptosis and inhibit immune response, facilitating tumor immune escape (Codony-Servat and Rosell, 2015).

Specific aberrant epigenetic modifications in DNA methylation, histone-modifying enzymes, chromatin remodelers and long non-coding RNAs, play a critical role in initiation and maintenance of CSC compartment, thus leading to tumorigenesis (Verona et al., 2022). Treatments aimed at inhibiting the immune characteristics and the epigenetic

alterations specific of CSC subset such as the immuno- and epigenetic-based therapies are thus considered as the new Frontier for the selective targeting of CSCs (D'Accardo et al., 2022; Verona et al., 2022).

Chimeric antigen receptor (CAR)-T cell therapy is among the most promising therapeutic approaches, holding an enormous potential of treating hematological disorders as well as solid tumors. This strategy is based on the *ex vivo* engineering of cancer patient T lymphocytes, in order to express a CAR selectively recognizing TAAs, and subsequent reinfusion in cancer patients. Notwithstanding the binding of engineered CAR-T cells to TAAs always culminates with the activation of cytotoxic signaling, release of granzyme, perforin and cytokines, and consequent elimination of transformed cells, several CAR structures have been developed to ameliorate the efficiency of CAR-T cell killing (Khan and Sarkar, 2022). The CAR is basically composed by an extracellular single-chain variable fragment (scFv) region, followed by a spacer and transmembrane domain and an intracellular region composed by the activation domain. Several adjustments have been made especially in the intracellular domain, which has been improved by the adjunction of i. one or multiple co-stimulatory domains (CD28, 4-1BB, ICOS or OX40), ii. transgenic sequences for pro-inflammatory cytokines release, or iii. IL-2R fragment, allowing JAK/STAT pathway activation (Sadelain et al., 2013; June et al., 2018; Khan and Sarkar, 2022). The basic CAR structure used in engineered T cells is shared by CAR-NK (Lu et al., 2021). In order to ameliorate the safety of CAR, the incorporation of a suicide gene (iC9) into the CAR construct is required to reduce the uncontrolled release of inflammatory cytokines, thus preventing the cytokine release syndrome (CRS) (Guercio et al., 2021).

CAR-T adoptive cell therapy in combination with the epigenetic reprogramming represents a powerful strategy to eradicate tumors due to the capability of overcoming one of the major challenges of cancer treatments, represented by tumor heterogeneity.

CAR-T cell–based therapy to selectively strike cancer stem cells

CSCs downregulate the expression of the major antigens' histocompatibility complex (MHC) class I to elude CD8[+] T lymphocyte recognition. Thus, the CAR-T cell-based therapy by recognizing a specific TAA on cancer cells and overcoming the MHC I restriction, represents the optimal strategy to target CSCs (Sterner and Sterner, 2021). Another consideration is that standard therapies are often unable to distinguish between normal stem cells (NSCs) and CSCs, determining a phenomenon known as "on-target off-tumor toxicity." In this context, the accuracy of CAR-T cells has been improved by engineering CAR-T cells to express the SynNotch receptor, whose binding to the tumor antigen, induces the expression of a second CAR, specific for a different tumor antigen. Double targeting CAR-T cells has indeed proved to be more precise in the killing of cancer cells while sparing normal cells (Roybal et al., 2016; Srivastava et al., 2019). This approach is particularly effective in patient's tumors displaying a downregulation or complete loss of one specific target antigen. Likewise, tumors showing the "antigen escape" resistance patterns

have been successfully targeted by tandem CAR constructs containing two scFv fragments to concomitantly bind multiple TAA (Wilkie et al., 2012).

CARs have indeed been constructed in order to recognize CSCs surface markers, usually used to identify and isolate CSCs, not shared by normal stem cells. A considerable number of anti-cancer therapies based on the use of CSC-targeting CAR-T cells are currently under pre-clinical or clinical evaluation. Pivotal examples are represented by clinical studies testing the efficacy CAR-T cells directed against the CSC surface markers EpCAM (NCT02915445; NCT03563326; NCT03013712; NCT02729493; NCT02725125), CD44v6 (NCT04427449) (Casucci et al., 2013; Porcellini et al., 2020) and c-Met (NCT01837602), involved in cancer cell metastatic potential. Other evidence demonstrated that CAR-T cells in combination with chemotherapy have been applied successfully to treat CD166 (Wang et al., 2019), CD133 (Han et al., 2021) and ROR1 (Srivastava et al., 2021) expressing tumors.

Notably, an unbiased CRISPR-Cas9 screening, on both CAR-T cells and glioblastoma stem cells, identified targetable genes whose knock out could potentiate long-term activation of CAR-T cells and glioblastoma stem cell susceptibility to CAR-T cell killing. Thus, posing CRISPR screens as a novel strategy to further enhance CAR-T cell therapeutic potency (Wang et al., 2021). Similarly, multiple CRISPR-Cas9 gene editing has been performed in CAR-T cells by deleting the *TCRα*, *TCRβ* and *PDCD1* (encoding PD-1), with the scope of decreasing the mispairing of TCR and immune-evasion. The CRISPR-Cas9 engineered CAR-T cells, further modified to express the cancer-specific synthetic transgene NY-ESO-1, displayed a durable post-infusion engraftment in the treated patients (Stadtmauer et al., 2020).

Interestingly, CAR-T cells, targeting the stemness marker GD2 in neuroblastoma, have been equipped with the suicide gene (iC9) as a safe clinical strategy to interrupt the cytotoxic activity of T cells in presence of severe adverse effects such as CRS. Moreover, GD2 CAR-T cells have been modified to express IL-7/IL-15, which significantly reduced the levels of the immune checkpoint PD-L1 on the surface of cancer cells (NCT03721068) (Perkins et al., 2015; Knudson et al., 2020). Indeed, the use of immune checkpoint inhibitors in combination with CAR-T cells therapy revealed a significant improvement in long-term remission of diffuse large B-cell lymphoma patients (NCT04381741). Of note, also CAR-NK can be considered a valid and alternative therapeutic approach (Xie et al., 2020; Christodoulou et al., 2021).

These studies open new scenarios in the field of immunotherapy due to the potential use CARs cell-based therapies, also in combination with chemotherapy, epigenetic-based or targeted therapies, as an effective strategy to disrupt tumor heterogeneity.

Epigenetic drugs targeting cancer stem cells enhance the CAR-T therapy efficacy

Epigenetic-based therapy is an increasingly attracting field in cancer research. Epigenetic compounds such as EZH2, DNMT and HDAC inhibitors are FDA approved, while other histone-methyltransferase (HMT) or histone demethylase (HDM) inhibitors are currently in clinical trials for hematological

FIGURE 1
Potential effects of the combinatorial treatment of epigenetic modulators in both Cancer Stem Cell and CAR-T cell compartments. Epigenetic drugs such as inhibitors of DNMT (DNMTi), BET (BETi), HDAC (HDACi), PRMT (PRMTi) and EZH (EZHi) and SIRT-1 agonists are efficacious against aberrant epigenetic alterations of CSCs leading to impairment of self-renewal and stemness and upregulation of TAA, and on CAR-T cells promoting their memory phenotype and reducing exhaustion.

malignancies and solid tumors. Of note, several epigenetic strategies have been reported to overcome the limits of CAR-T therapy, particularly leading to a more permissive tumor microenvironment (TME). For instance, DNMT inhibitors finely modulate the macrophages immunosuppressive activity and promote the anti-tumor activity of T cells (Stone et al., 2017), while JQ1 a BET inhibitor reduces the expression levels of PD-L1 on tumor cells and immune cells present in TME (Zhu et al., 2016). Among the major obstacles to the effectiveness of CAR-T therapy, the antigen loss and the heterogeneity of antigens may be counteracted by the combination therapy with epigenetic inhibitors (Akbari et al., 2021). Particularly, DNMT and HDAC inhibitors may result to an upregulation of TAAs and of co-stimulatory molecules such as CD40, CD80, CD86 and MHC classes I and II, on tumor cells, in particular in Acute Myeloid Leukemia (AML) cell lines (Maeda et al., 2000; Magner et al., 2000). This epigenetic-based strategy facilitates and potentiates the efficacy of CAR-T cells engineered against AML cells. Of note, EZH2 inhibitors in combination with CAR-T therapy boost the anti-tumor activity of CAR-T in Ewing sarcoma by reducing H3K27me3 mark and inducing the surface expression of GD2, a stemness marker in tumor cells (Kailayangiri et al., 2019). This epigenetic approach is aimed to increase and stabilize the expression of GD2, a tumor-sparse antigen, in Ewing sarcoma cells, in order to enable GD2-CAR T cells to target all tumor cells highly expressing GD2.

Several aberrant druggable epigenetic alterations have been linked to the initiation and maintenance of CSCs across several cancer types (Veschi et al., 2019; Veschi et al., 2020; F. et al., 2022).

Compounds which induce the blockade of these specific epigenetic modifications in CSC subset including DNMT1 and DNMT3 inhibitors, HDAC inhibitors, SIRT-1 agonists and PRMT5 inhibitors are currently available in clinical or preclinical settings (F. et al., 2022). Of note, in pediatric cancers enriched in stem-like cells with high plasticity, alterations of histone modifying enzymes with non-histone targets may be specifically targetable (Veschi et al., 2017; Veschi and Thiele, 2017).

Moreover, epigenetic reprogramming has been used to modulate the differentiation state and to promote the memory phenotypes of CAR-T, to improve CAR-T infiltration and persistence, and finally as an alternative strategy to avoid their exhaustion (Akbari et al., 2021; Alvanou et al., 2023).

Overall, it is becoming increasingly clear that the epigenetic reprogramming induced by the treatment with epi-drugs may exert several effects on both CSC subpopulation and CAR-T cells through i) the impairment of self-renewal and stemness capabilities of CSCs leading to an inhibition of CSC initiation (F. et al., 2022); ii) the upregulation of TAAs specific for CSCs enabling CAR-T cells to specifically target CSCs; iii) the enhancement of the intrinsic properties of T cells by histone, DNA and miRNAs modifications promoting CAR-T cells memory phenotype and reverting their exhaustion (Alvanou et al., 2023).

Thus, combining targeting of epigenetic alterations present on CSCs with CAR-T therapy may boost the epi-immunotherapy for cancers (Akbari et al., 2021; Xu et al., 2022) (Figure 1), although some limitations and side-effects which will be described in the next paragraphs.

Side effects and toxicity of CAR–T cell therapy and their management in clinical practice

As a "living drug," CAR T-cells, armed to recognize tumor antigens once infused in the blood of oncologic patients, can generate multiple side effects with diverse degree of severity. The major described effects include CRS, immune effector cell-associated neurotoxicity syndrome (ICANS) and cytopenia (Schubert et al., 2021).

CRS consists in the massive presence of inflammatory chemokines and cytokines, as IFN-γ, IL-2, TNF-α, GM-CSF, IL-8, IL-10, IL-6 and IL-2, released by activated CAR-T cells, macrophages, monocytes and dendritic cells. CRC manifests in four different grades according to symptoms severity. In order to allow an early management of CRS, temperature, blood pressure and oxygen saturation are constantly monitored. If any imbalance in these parameters is observed, a prompt CRS treatment is strictly advised, by the administration of antipyretics, corticosteroids, tocilizumab (IL-6R inhibitor) and supplemental oxygen administration (Lee et al., 2019; Neelapu, 2019).

The ICANS is observed in more than a half of patients treated with CAR-T cell therapy and can occurs independently of CRS. Patients affected by ICANS experience loss of consciousness, motor weakness and cerebral edema. According to the American Society for Transplantation and Cellular Therapy (ASTCT) grading consensus system, patients are classified in four different grades and are strictly monitored for disease evolution (Lee et al., 2019; Neelapu, 2019). Patients experiencing severe symptoms are treated with corticosteroids, mannitol, hypertonic saline and hyperventilation.

Hematological toxicity causes cytopenia in several patients treated with intravenous infusion of CAR-T cells. If persistent, autologous or allogenic stem cell transplant is recommended to cytopenic patients.

Diverse management strategies have been adopted in order to reduce mild and severe side effects sometimes requiring intensive care unit management. Safety profiles of CAR-T cell therapy have been developed primarily by lowering CAR-T cell doses.

Limitations and challenges of CAR-T cell-based therapy and epigenetic drugs targeting cancer stem cells.

Nevertheless CAR-T cell-based therapy has been successful in hematological malignancies, several obstacles have been experienced in solid tumors, besides the clinical limitations due to side effects.

The type of cancers that can be potentially targeted by both approaches, CAR-T cell-based therapy and epigenetic drugs, are ideally the same in which CAR-T alone has successfully worked: B cells malignancies, AML, chronic lymphocytic leukemia (CLL), acute lymphocytic leukemia (ALL), non-Hodgkin lymphoma (NHL) and Multiple Myeloma (MM). Of note, CAR-T cell therapy encountered several difficulties when translated into the landscape of solid tumors compared to hematologic cancers as follows: heterogeneous antigens that are difficult to be targeted, tumor physical barriers that block the CAR-T infiltration (e.g., CAFs, ECM, increased number of blood vessels), presence of several factors contributing to the immunosuppressive environment (low pH, hypoxia, immune check point blockage molecules) (Qin and Xu, 2022). In particular, the limitations of

these therapies in specifically targeting the CSC compartment are based on peculiarities associated to the stem-like phenotype such as the following: 1) heterogeneous antigens potentially shared with NSCs which lead to "on target off-tumor toxicities"; 2) impaired CAR-T cells persistence and trafficking into the CSC niche in solid tumors; 3) presence of broad variety of immunosuppressive cells and molecules (Masoumi et al., 2021). On target off-tumor toxicities is determined by low levels of expression of some TAAs also in normal cells resulting in CAR-T cells targeting even the healthy cells. Recently, novel strategies to overcome the above described limitations have been proposed such as the use of CARs targeting multiple CSC antigens and the use of dual-specific CAR-T cells, optimization of the *ex vivo* CAR-T cell culture conditions, local delivery of CAR-T cells and the use of CAR-T overexpressing chemokines (Masoumi et al., 2021).

Epigenetic compounds targeting CSCs, such as HDAC inhibitors and particularly the pan-inhibitors, have been reported to induce many side effects such as diarrhea, fatigue, nausea, and anorexia. The main issue in epigenetic drug development is to identify selective compounds with significant *in vitro* cellular activity at nM concentrations without any *in vivo* toxicities. An overview about the limitations and side-effects of the epigenetic inhibitors targeting CSCs, along with the epigenetic mechanisms of resistance to therapy in CSCs, can be found in Turdo et al. (Turdo et al., 2019).

Notably, the evaluation of potential toxicities, adverse events and development of resistance induced by the combinatorial treatment based on CAR-T cells and epigenetic drugs targeting CSCs such as DNMT and HDAC inhibitors is currently ongoing in AML, ALL and NHL, although in a very limited number of clinical trials (NCT03612739, NCT04553393, NCT05797948 and NCT05370547).

Future studies will clarify these aspects and likely evaluate the efficacy of epidrugs targeting CSCs in combination with CAR-T directed against TAAs specific for CSCs.

Conclusion

Epi-immunotherapy is a promising approach, which may overcome the challenge of specifically targeting CSC subpopulation by enhancing the immune response against the tumor cells and CSCs in particular. Notably, immunotherapy approaches based on CARs, such as CAR-T and CAR-NK, have recently revolutionized the treatment of hematological and solid tumors, enriched in CSCs.

CAR-T efficacy in combination with epidrugs is enhanced by several mechanisms which lead to an epigenetic reprogramming of both CSCs and CAR-T cells. Nevertheless, limitations and hurdles of CAR-T therapy especially in solid tumors still remain to be overcome. Several strategies to bypass these limitations, particularly the on target off-tumor toxicity, have been extensively reported. However, further studies are needed for an optimal tailoring of CAR-T cells enabling them to specifically target CSCs especially in solid tumors avoiding the well-known toxicities.

The goal of tumor eradication without recurrence may be achieved by the combination of CAR-T cells targeting TAAs

specific of CSCs and epigenetic drugs specifically targeting CSCs, in association with standard radio and/or chemotherapy.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

VV, AT, and GS conceived, wrote and revised this paper. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Akbari, B., Ghahri-Saremi, N., Soltantoyeh, T., Hadjati, J., Ghassemi, S., and Mirzaei, H. R. (2021). Epigenetic strategies to boost CAR T cell therapy. *Mol. Ther.* 29, 2640–2659. doi:10.1016/j.ymthe.2021.08.003

Alvanou, M., Lysandrou, M., Christophi, P., Psatha, N., Spyridonidis, A., Papadopoulou, A., et al. (2023). Empowering the potential of CAR-T cell immunotherapies by epigenetic reprogramming. *Cancers (Basel).* 15, 1935. doi:10.3390/cancers15071935

Casucci, M., Nicolis Di Robilant, B., Falcone, L., Camisa, B., Norelli, M., Genovese, P., et al. (2013). CD44v6-targeted T cells mediate potent antitumor effects against acute myeloid leukemia and multiple myeloma. *Blood* 122, 3461–3472. doi:10.1182/blood-2013-04-493361

Christodoulou, I., Ho, W. J., Marple, A., Ravich, J. W., Tam, A., Rahnama, R., et al. (2021). Engineering CAR-NK cells to secrete IL-15 sustains their anti-AML functionality but is associated with systemic toxicities. *J. Immunother. Cancer* 9, e003894. doi:10.1136/jitc-2021-003894

Codony-Servat, J., and Rosell, R. (2015). Cancer stem cells and immunoresistance: Clinical implications and solutions. *Transl. Lung Cancer Res.* 4, 689–703. doi:10.3978/j.issn.2218-6751.2015.12.11

D'accardo, C., Porcelli, G., Mangiapane, L. R., Modica, C., Pantina, V. D., Roozafzay, N., et al. (2022). Cancer cell targeting by CAR-T cells: A matter of stemness. *Front. Mol. Med.* 2. doi:10.3389/fmmed.2022.1055028

Guercio, M., Manni, S., Boffa, I., Caruso, S., Di Cecca, S., Sinibaldi, M., et al. (2021). Inclusion of the inducible caspase 9 suicide gene in CAR construct increases safety of CAR.CD19 T cell therapy in B-cell malignancies. *Front. Immunol.* 12, 755639. doi:10.3389/fimmu.2021.755639

Han, Y., Sun, B., Cai, H., and Xuan, Y. (2021). Simultaneously target of normal and stem cells-like gastric cancer cells via cisplatin and anti-CD133 CAR-T combination therapy. *Cancer Immunol. Immunother.* 70, 2795–2803. doi:10.1007/s00262-021-02891-x

June, C. H., O'connor, R. S., Kawalekar, O. U., Ghassemi, S., and Milone, M. C. (2018). CAR T cell immunotherapy for human cancer. *Science* 359, 1361–1365. doi:10.1126/science.aar6711

Kailayangiri, S., Altvater, B., Lesch, S., Balbach, S., Gottlich, C., Kuhnemundt, J., et al. (2019). EZH2 inhibition in ewing sarcoma upregulates G(D2) expression for targeting with gene-modified T cells. *Mol. Ther.* 27, 933–946. doi:10.1016/j.ymthe.2019.02.014

Khan, A., and Sarkar, E. (2022). CRISPR/Cas9 encouraged CAR-T cell immunotherapy reporting efficient and safe clinical results towards cancer. *Cancer Treat. Res. Commun.* 33, 100641. doi:10.1016/j.ctarc.2022.100641

Knudson, K. M., Hicks, K. C., Ozawa, Y., Schlom, J., and Gameiro, S. R. (2020). Functional and mechanistic advantage of the use of a bifunctional anti-PD-L1/IL-15 superagonist. *J. Immunother. Cancer* 8, e000493. doi:10.1136/jitc-2019-000493

Lee, D. W., Santomasso, B. D., Locke, F. L., Ghobadi, A., Turtle, C. J., Brudno, J. N., et al. (2019). ASTCT consensus grading for cytokine release syndrome and neurologic toxicity associated with immune effector cells. *Biol. Blood Marrow Transpl.* 25, 625–638. doi:10.1016/j.bbmt.2018.12.758

Lu, H., Zhao, X., Li, Z., Hu, Y., and Wang, H. (2021). From CAR-T cells to CAR-NK cells: A developing immunotherapy method for hematological malignancies. *Front. Oncol.* 11, 720501. doi:10.3389/fonc.2021.720501

Maeda, T., Towatari, M., Kosugi, H., and Saito, H. (2000). Up-regulation of costimulatory/adhesion molecules by histone deacetylase inhibitors in acute myeloid leukemia cells. *Blood* 96, 3847–3856. doi:10.1182/blood.v96.12.3847

Magner, W. J., Kazim, A. L., Stewart, C., Romano, M. A., Catalano, G., Grande, C., et al. (2000). Activation of MHC class I, II, and CD40 gene expression by histone deacetylase inhibitors. *J. Immunol.* 165, 7017–7024. doi:10.4049/jimmunol.165.12.7017

Masoumi, J., Jafarzadeh, A., Abdolalizadeh, J., Khan, H., Philippe, J., Mirzaei, H., et al. (2021). Cancer stem cell-targeted chimeric antigen receptor (CAR)-T cell therapy: Challenges and prospects. *Acta Pharm. Sin. B* 11, 1721–1739. doi:10.1016/j.apsb.2020.12.015

Neelapu, S. S. (2019). Managing the toxicities of CAR T-cell therapy. *Hematol. Oncol.* 37 (1), 48–52. doi:10.1002/hon.2595

Perkins, M. R., Grande, S., Hamel, A., Horton, H. M., Garrett, T. E., Miller, S. M., et al. (2015). Manufacturing an enhanced CAR T cell product by inhibition of the PI3K/akt pathway during T cell expansion results in improved *in vivo* efficacy of anti-BCMA CAR T cells. *Blood* 126, 1893. doi:10.1182/blood.v126.23.1893.1893

Porcellini, S., Asperti, C., Corna, S., Cicoria, E., Valtolina, V., Stornaiuolo, A., et al. (2020). CAR T cells redirected to CD44v6 control tumor growth in lung and ovary adenocarcinoma bearing mice. *Front. Immunol.* 11, 99. doi:10.3389/fimmu.2020.00099

Qin, Y., and Xu, G. (2022). Enhancing CAR T-cell therapies against solid tumors: Mechanisms and reversion of resistance. *Front. Immunol.* 13, 1053120. doi:10.3389/fimmu.2022.1053120

Roybal, K. T., Rupp, L. J., Morsut, L., Walker, W. J., Mcnally, K. A., Park, J. S., et al. (2016). Precision tumor recognition by T cells with combinatorial antigen-sensing circuits. *Cell* 164, 770–779. doi:10.1016/j.cell.2016.01.011

Sadelain, M., Brentjens, R., and Riviere, I. (2013). The basic principles of chimeric antigen receptor design. *Cancer Discov.* 3, 388–398. doi:10.1158/2159-8290.CD-12-0548

Schubert, M. L., Schmitt, M., Wang, L., Ramos, C. A., Jordan, K., Muller-Tidow, C., et al. (2021). Side-effect management of chimeric antigen receptor (CAR) T-cell therapy. *Ann. Oncol.* 32, 34–48. doi:10.1016/j.annonc.2020.10.478

Srivastava, S., Furlan, S. N., Jaeger-Ruckstuhl, C. A., Sarvothama, M., Berger, C., Smythe, K. S., et al. (2021). Immunogenic chemotherapy enhances recruitment of CAR-T cells to lung tumors and improves antitumor efficacy when combined with checkpoint blockade. *Cancer Cell* 39, 193–208.e10. doi:10.1016/j.ccell.2020.11.005

Srivastava, S., Salter, A. I., Liggitt, D., Yechan-Gunja, S., Sarvothama, M., Cooper, K., et al. (2019). Logic-gated ROR1 chimeric antigen receptor expression rescues T cell-mediated toxicity to normal tissues and enables selective tumor targeting. *Cancer Cell* 35, 489–503. doi:10.1016/j.ccell.2019.02.003

Stadtmauer, E. A., Fraietta, J. A., Davis, M. M., Cohen, A. D., Weber, K. L., Lancaster, E., et al. (2020). CRISPR-engineered T cells in patients with refractory cancer. *Science* 367, eaba7365. doi:10.1126/science.aba7365

Sterner, R. C., and Sterner, R. M. (2021). CAR-T cell therapy: Current limitations and potential strategies. *Blood Cancer J.* 11, 69. doi:10.1038/s41408-021-00459-7

Stone, M. L., Chiappinelli, K. B., Li, H., Murphy, L. M., Travers, M. E., Topper, M. J., et al. (2017). Epigenetic therapy activates type I interferon signaling in murine ovarian cancer to reduce immunosuppression and tumor burden. *Proc. Natl. Acad. Sci. U. S. A.* 114, E10981–E10990. doi:10.1073/pnas.1712514114

Turdo, A., Veschi, V., Gaggianesi, M., Chinnici, A., Bianca, P., Todaro, M., et al. (2019). Meeting the challenge of targeting cancer stem cells. *Front. Cell Dev. Biol.* 7, 16. doi:10.3389/fcell.2019.00016

Verona, F., Pantina, V. D., Modica, C., Lo Iacono, M., D'Accardo, C., Porcelli, G., et al. (2022). Targeting epigenetic alterations in cancer stem cells. *Front. Mol. Med.* 2. doi:10.3389/fmmed.2022.1011882

Veschi, V., Liu, Z., Voss, T. C., Ozbun, L., Gryder, B., Yan, C., et al. (2017). Epigenetic siRNA and chemical screens identify SETD8 inhibition as a therapeutic strategy for p53 activation in high-risk neuroblastoma. *Cancer Cell* 31, 50–63. doi:10.1016/j.ccell.2016.12.002

Veschi, V., and Thiele, C. J. (2017). High-SETD8 inactivates p53 in neuroblastoma. *Oncoscience* 4, 21–22. doi:10.18632/oncoscience.344

Veschi, V., Verona, F., Lo Iacono, M., D'accardo, C., Porcelli, G., Turdo, A., et al. (2020). Cancer stem cells in thyroid tumors: From the origin to metastasis. *Front. Endocrinol. (Lausanne)* 11, 566. doi:10.3389/fendo.2020.00566

Veschi, V., Verona, F., and Thiele, C. J. (2019). Cancer stem cells and neuroblastoma: Characteristics and therapeutic targeting options. *Front. Endocrinol. (Lausanne)* 10, 782. doi:10.3389/fendo.2019.00782

Wang, D., Prager, B. C., Gimple, R. C., Aguilar, B., Alizadeh, D., Tang, H., et al. (2021). CRISPR screening of CAR T cells and cancer stem cells reveals critical dependencies for cell-based therapies. *Cancer Discov.* 11, 1192–1211. doi:10.1158/2159-8290.CD-20-1243

Wang, Y., Yu, W., Zhu, J., Wang, J., Xia, K., Liang, C., et al. (2019). Anti-CD166/4-1BB chimeric antigen receptor T cell therapy for the treatment of osteosarcoma. *J. Exp. Clin. Cancer Res.* 38, 168. doi:10.1186/s13046-019-1147-6

Wilkie, S., Van Schalkwyk, M. C., Hobbs, S., Davies, D. M., Van Der Stegen, S. J., Pereira, A. C., et al. (2012). Dual targeting of ErbB2 and MUC1 in breast cancer using chimeric antigen receptors engineered to provide complementary signaling. *J. Clin. Immunol.* 32, 1059–1070. doi:10.1007/s10875-012-9689-9

Xie, G., Dong, H., Liang, Y., Ham, J. D., Rizwan, R., and Chen, J. (2020). CAR-NK cells: A promising cellular immunotherapy for cancer. *EBioMedicine* 59, 102975. doi:10.1016/j.ebiom.2020.102975

Xu, Y., Li, P., Liu, Y., Xin, D., Lei, W., Liang, A., et al. (2022). Epi-immunotherapy for cancers: Rationales of epi-drugs in combination with immunotherapy and advances in clinical trials. *Cancer Commun. (Lond)* 42, 493–516. doi:10.1002/cac2.12313

Zhu, H., Bengsch, F., Svoronos, N., Rutkowski, M. R., Bitler, B. G., Allegrezza, M. J., et al. (2016). BET bromodomain inhibition promotes anti-tumor immunity by suppressing PD-L1 expression. *Cell Rep.* 16, 2829–2837. doi:10.1016/j.celrep.2016.08.032

Check for updates

A method for temporal-spatial multivariate genomic analysis of acute wound healing via tissue stratification: a porcine negative pressure therapy pilot study

Jacob G. Hodge[1,2], Sumedha Gunewardena[3], Richard A. Korentager[2], David S. Zamierowski[2], Jennifer L. Robinson[4,5,6,7] and Adam J. Mellott[2,8]*

[1]Bioengineering Graduate Program, University of Kansas, Lawrence, KS, United States, [2]Department of Plastic Surgery, University of Kansas Medical Center, Kansas, KS, United States, [3]Department of Cell Biology and Physiology, University of Kansas Medical Center, Kansas, KS, United States, [4]Department of Chemical and Petroleum Engineering, University of Kansas, Lawrence, KS, United States, [5]Department of Orthopaedics and Sports Medicine, University of Washington, Seattle, WA, United States, [6]Department of Mechanical Engineering, University of Washington, Seattle, WA, United States, [7]Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA, United States, [8]Ronawk Inc., Olathe, KS, United States

Introduction: Wound therapies are capable of modulating the complex molecular signaling profile of tissue regeneration. However traditional, bulk tissue analysis results in nonspecific expressional profiles and diluted signaling that lacks temporal-spatial information.

Methods: An acute incisional porcine wound model was developed in the context of negative pressure wound therapy (NPWT). Dressing materials were inserted into wounds with or without NPWT exposure and evaluated over 8-hours. Upon wound explantation, tissue was stratified and dissected into the epidermis, dermis, or subcutaneous layer, or left undissected as a bulk sample and all groups processed for RNAseq. RNAseq of stratified layers provided spatial localization of expressional changes within defined tissue regions, including angiogenesis, inflammation, and matrix remodeling.

Results: Different expressional profiles were observed between individual tissue layers relative to each other within a single wound group and between each individual layer relative to bulk analysis. Tissue stratification identified unique differentially expressed genes within specific layers of tissue that were hidden during bulk analysis, as well as amplification of weak signals and/or inversion of signaling between two layers of the same wound, suggesting that two layers of skin can cancel out signaling within bulk analytical approaches.

Discussion: The unique wound stratification and spatial RNAseq approach in this study provides a new methodology to observe expressional patterns more precisely within tissue that may have otherwise not been detectable. Together these experimental data offer novel insight into early expressional patterns and genomic profiles, within and between tissue layers, in wound healing pathways that could potentially help guide clinical decisions and improve wound outcomes.

1 Introduction

Skin, the largest organ of the human body, serves as an external barrier from the outside elements and provides protection from external insults (Shaw and Martin, 2009). Therefore, a wound can be defined as any injury that results in damage or disruption of the epidermal skin barrier and can often result in exposure of the deeper tissue structures to the outside elements (Shaw and Martin, 2009). Our bodies utilize a natural feedback loop to counteract and repair "open" wounds in order to "close" them and restore the epidermal barrier and protect deeper tissue structures, a process known as acute wound healing (Reinke and Sorg, 2012; Eming et al., 2014). Acute wound healing is an intricate compilation of diverse signaling cascades continuously providing both local and systemic feedback to ensure appropriate progression through the phases of wound healing (Reinke and Sorg, 2012; Eming et al., 2014). Acute wound healing involves multiple cell/tissue types and an intricate balance of dynamic molecular signaling cascades that are continuously changing as the wound tissue evolves. Thus, the ultimate goal of acute wound healing is to reestablish anatomical and functional homeostasis of the injured tissue.

To this day, there is no singular wound care intervention proven most effective for all wounds. However, application of a diverse array of wound dressing materials are often utilized and considered the mainstay of wound care (Dreifke et al., 2015; Hodge et al., 2022). One innovative therapy, that is, widely utilized clinically is negative pressure wound therapy (NPWT), which been shown to modulate the wound healing response for a number of applications, including skin grafting, surgical, traumatic, burn, orthopedic, and diabetic wounds (Argenta and Morykwas, 1997; Morykwas et al., 1997; Agarwal et al., 2019; Norman et al., 2022). NPWT involves the insertion of an open-reticulated biomaterial foam dressing into the wound, such as the polyurethane-derived GranuFoam™, coverage with a semi-permeable adhesive film, and application of subatmospheric pressures (typically 125 mmHg), via a vacuum pump, to mechanically compress the foam and subsequently decrease the wound site volume (Morykwas et al., 2006).

Overall, NPWT is well known for its ability to modulate wound healing, with clinically documented outcomes of increased rates of neovascularization, matrix production, and granulation tissue deposition, paired with augmentation of the immune response and decreases in bacterial burden (Morykwas et al., 2006; Normandin et al., 2021). A number of suggested mechanisms are associated with how NPWT exerts its effects, including the mechanical decrease of interstitial pressure via fluid egress from the wound site, the enhancement of tissue perfusion and oxygenation, and induction of mechanotransductive or foreign body responses to promote granulation tissue production (Morykwas et al., 2001; Morykwas et al., 2006; Agarwal et al., 2019; Hodge et al., 2021; Normandin et al., 2021). Additionally, recent alterations to the NPWT system have been investigated including the use of different foam materials and application of instillation devices, which has further increased the widespread implementation of NPWT as an adjuvant therapy for a variety of

tissue ailments (Kim et al., 2020). Notably, the application of NPWT has also demonstrated the capacity to prepare surgical site tissue with improved grafting outcomes and decreased long-term scarring (Gupta, 2012; Webster et al., 2014). Thus, NPWT can potentially modulate the wound healing response in a number of ways and for a number of different tissue types. Yet, the dynamic mechanism(s) of how NPWT exerts its lasting effects within defined temporal and spatial confinements, remains unresolved. More specifically, it is not fully understood how different NPWT iterations and dressing material properties modulate the initial wound healing response, how these initial responses have resonating/cascading effects on wound healing several days later, what specific cells and signaling process are involved acutely, and where the key cell populations involved in the signaling are localized within the wound tissue.

The interconnected and overlapping nature of wound healing results in a complex array of molecular signaling cascades that varies amongst wound types. Notably, the topographical, architectural, and biological profile of the wound extracellular matrix (ECM) is known to modulate the wound healing process (Reinke and Sorg, 2012; Eming et al., 2014). Similarly, initial interactions between cellular populations and wound dressings properties are critical to the overall progression of the regenerative process, with different material compositions and structural formulations known to modulate cellular activity and gene expression profiles (Thevenot et al., 2008; Boehler et al., 2011; Andorko and Jewell, 2017; Jang et al., 2019). Thus, understanding how interventional wound therapies are capable of modulating wound healing at a molecular level within specific, defined regions of tissue remains a critical area of research for advancing wound therapies. Moreover, investigative new technologies to improve our current understanding of the physiological wound response will ultimately aid in the advancement of effective and targeted wound therapies. To date, wound analyses have exclusively evaluated gene expressional profiles of whole wound isolates/biopsies that compile all/multiple layers of skin tissue into one bulk sample set (Derrick and Lessing, 2014; Ma et al., 2016; Brownhill et al., 2021). Although this methodological processing has yielded large amounts of valuable information, it is limiting in its current form due to the diversity of signaling within different layers of wound tissue.

In this methodological study, a "proof-of-concept" design for a new *in vivo* porcine model is developed for screening acute wound healing interactions with different treatments and dressing materials. Due to the acute nature and interest in understanding initial wound responses, an incisional model was utilized. To provide context within this study NPWT (i.e., Granufoam™ with 125 mmHg vacuum) was utilized as the treatment group and compared to untreated control healing (i.e., no dressing insert or vacuum) within 8 h. Uniquely, this new methodology allows for the capacity to assess both the temporal and spatial components of wound healing within defined regions of the wound tissue due to the unique design, processing, and analyses. Whereas previous techniques utilize "bulk" biopsies and explanted tissue samples to perform gene expression analysis on entire wound samples, this study stratifies the skin into its three component parts, the

epidermis, dermis, and subcutaneous layers. Thus, our new methodology demonstrates the capacity to potentially differentiate within sample sets and perform inter- and intra-skin layer comparisons with multivariate analyses, which offers a new perspective to view wound healing. NPWT has been shown to effect a number of processes; however, there is still much to be understood into when, where, and how it acts on specific cell populations within wounds in order to further advance its therapeutic benefits. We hypothesized that a number of critical wound healing pathways traditionally seen are likely coming from a distinct layer of skin tissue, whereas there may also be never before seen acute signaling pathways revealed due to being previously "washed out" by other layers during "Full/Bulk" tissue analysis.

2 Materials and methods

2.1 Animals

Animal studies were approved by the University of Kansas Medical Center (KUMC) Institutional Animal Care and Use Committee (IACUC) under animal care and use protocol (ACUP) #2019-2535. Two female 4-month-old miniature Yucatan pigs weighing 30–40 kg were procured from Sinclair Bio-resources (Auxvasse, MO), and allowed to acclimate for 14 days in an AAALAC accredited facility at KUMC. Animals were provided with food, water, and social enrichment *ad libitum*.

2.2 Surgery and sample collection

Surgeries were performed on each animal on separate days. Animals were placed under general anesthesia and ophthalmic lubricating ointment was placed to protect the eyes. The animals were prepped with three alternating scrubs of betadine and alcohol. A sterile surgical drape was placed over the animal, except for where the surgical procedure would be carried out. A custom 3D-printed acrylonitrile butadiene styrene stencil was prefabricated and used to appropriately mark the incision sites and dressing borders for the animals in a consistent manner prior to surgery. Additionally, custom designed scalpel guides were prefabricated that allowed for control of incisional depth and a double-scalpel guide was used to control the width and depth of the wound explants. Incisions were cut to a depth of 1.5-cm and a length of 2-cm. A 4 × 4 incisional array was inflicted, with the four columns representing 0-, 2-, 4-, and 8-h timepoints of explantation, and the four rows representing (dorsal to ventral) control (no dressing insert), Owen's Rayon, GranuFoam™ (Kinetic Concepts Inc. [KCI] an Acelity company, San Antonio, TX), and polycaprolactone electrospun mesh dressings (Supplementary Figure S1). Only the 0-h and 8-h GranuFoam™ and control wound samples are discussed in this study to demonstrate "proof-of-concept". A 2-cm × 2-cm × 0.5-cm (thick) piece of each of the three respective dressing materials were inserted in the appropriate wounds for each timepoint. Followed by placement of a Prevena foam over the top of each wound column/timepoint to bridge the dressing materials. The Prevena foams were connected to a vacuum system and set at a constant 125 mmHg for the duration of each listed timepoint. The

right side of each animal received vacuum, whereas the left side of each animal served as a non-vacuum control but did have a Prevena placed over the top of each wound column. With the exception being the time 0-h columns, which were immediately processed.

At each respective timepoints, the double-scalpel guide was used to cut one long strip (entire wound column) and the predetermined depth of 2 cm. Wound explant was immediately processed, and the four treatment groups (rows) were separated, followed by further separation of each individual sample (row). Each 2-cm incisional wound was immediately processed in the operating room by a surgical assistant and cut into four equally-sized sections (5-mm thick) to allow for four different analytical approaches, including RNAlater (Sigma-Aldrich) preservation for future genomic experiments (Figure 1). The newly inflicted wound on the animals, due to wound explants, were packed with surgical gauze and compressed by hand until bleeding ceased. For RNA sample processing, wound samples were bisected in-line with the incision to generate two mirror halves of the wound and refrigerated at 4 °C in RNAlater for 24 h, followed by storage at −80°C. Animals were euthanized while under deep general anesthesia via exsanguination. A photographic overview of the experimental surgical procedure is depicted in Supplementary Figures S1A–K.

2.3 Processing and tissue stratification

RNAlater preserved wound samples were stratified into four groups for each wound, Full/Bulk, epidermis, dermis, and subcutaneous. First, a clean, RNAase-free workspace was created to process the tissue samples. Next, a tub of ice was prepared, an aluminum/metal sheet was placed on top of the ice and washed with 100% isopropanol followed by RNase away. All equipment and supplies were also cleaned with 100% isopropanol and RNase away. For processing, the previously bisected wound samples (two equal mirror halves of same wound) were thawed from −80°C on ice and removed prior to complete thaw so that the tissue was still hard but RNAlater had melted. Next, a Dermatome blade was used to dissect each layer from one-half of each wound sample. The remaining half was processed as a Full/Bulk wound and used to serve as the standard processing technique. Samples were immediately transported to the Genomics Core facility in fresh RNAlater (on ice) for RNA processing and isolation.

2.4 RNA isolation

Epidermis and Subcutaneous Layers. Epidermis and subcutaneous isolates from the pig wound channel were suspended in 1 ml of 4°C Invitrogen Trizol Reagent (Fisher Scientific 15-596-026) in a 5 ml tissue culture tube on ice. The tissue was homogenized using a Power Gen 35 Tissue Homogenizer employing a medium shark's tooth probe—6 mm diameter (Fisher Scientific) for 20 s on ice. The 1 ml lysate is transferred to a Andwin Scientific 5Prime Phase Lock Gel—Heavy 2 ml tube (Fisher Scientific) and the Phase Lock gel purification is performed according to manufacturer's protocol until the separation centrifugation of the aqueous phase and organic phase is complete. The aqueous phase is transferred to a microcentrifuge tube and the volume is determined. An equal volume of 70% ethanol in nuclease free water is added to the aqueous phase. The ethanol adjusted

FIGURE 1

KUts and KUbes Incisional Porcine Model Experimental Design. **(A)** A schematic overview of the incisional porcine wound model depicting the incision depth and insertion of dressing material. **(B)** Overall layout of the 4 × 4 array of incisions on each side of the animals, with one side receiving NPWT. **(C)** Demonstration of the tissue processing and how wound explants would be analyzed. Different materials were inserted into different rows on the animals and at each respective timepoint the column was collected as a strip of tissue, cut into four pieces to separate out the four wounds within each strip and them each of the four wounds were further subdivided into four pieces for analysis (EM = electron microscopy). **(D)** Depiction of the stratification process of each wound explant, separating the "Full/Bulk" wound into three layers, epidermis, dermis, and subcutaneous. **(E)** A depiction of some of the multivariate analyses/comparisons that can be performed with the stratification processing technique.

aqueous phase (with guanidinium salts) is applied directly to a RNeasy Mini kit (Qiagen 74104) purification column and the RNA Clean-up protocol is performed using an on-column DNase treatment with the RNase-Free DNase Set (Qiagen 79254) according to manufacturer's instruction. DNase treated RNA isolates are quality controlled and quantified using a TapeStation 4200 running the RNA ScreenTape assay (Agilent Technologies). RNA Integrity Number (RIN) of ≥8.0 are targeted for proceeding to stranded mRNA library preparation.

Dermis and "Full/Bulk" Layers. Dermis and full tissue isolates (with skin) from the pig wound channel were suspended in 500ul of 4°C Invitrogen Trizol Reagent (Fisher Scientific 15-596-026) in a 5 ml tissue culture tube on ice. The tissue was homogenized using a Power Gen 35 Tissue Homogenizer employing a medium shark's tooth probe—6 mm diameter (Fisher Scientific) for 20 s on ice. The lysate was removed to a new 5 ml tissue culture tube on ice. 500ul of additional chilled Trizol Reagent was added to the remaining un-lysed tissue (mainly skin) and homogenized an additional 20 s on ice leaving ~10% of the original skin undisrupted. The lysate was pooled with the first lysate for 1 ml total. The 1 ml lysate is transferred to an Andwin Scientific 5 Prime Phase Lock Gel—Heavy 2 ml tube (Fisher Scientific) and the Phase Lock gel purification is performed according to manufacturer's protocol until the separation centrifugation of the aqueous phase and organic phase is complete. The aqueous phase is transferred to a microcentrifuge tube and the volume is determined. An equal volume of 70% ethanol in nuclease free water is added to the aqueous phase. The ethanol adjusted aqueous phase (with guanidinium salts) is applied directly to a RNeasy Mini kit (Qiagen 74104) purification column and the RNA Clean-up protocol is performed using an on-column DNase treatment with the RNase-Free DNase Set (Qiagen 79254) according to manufacturer's instruction. DNase treated RNA isolates are quality controlled and quantified using a TapeStation 4200 running the RNA ScreenTape assay (Agilent Technologies).

RNA Integrity Number (RIN) of ≥8.0 are targeted for proceeding to stranded mRNA library preparation.

2.5 RNA library prep

Tecan Universal Plus mRNA-Seq with NuQuant. The stranded mRNA-Seq was performed using the Illumina NovaSeq 6000 Sequencing System at the University of Kansas Medical Center—Genomics Core (Kansas City, KS). Quality control of the total RNA isolates was completed using the Agilent TapeStation 4200 using the RNA ScreenTape Assay kit (Agilent Technologies 5067-5576). Total RNA (500 ng) was used to initiate the library preparation protocol. The 249 RNA isolates representing epidermis, dermis, subcutaneous and full tissue were randomized and divided into 11 sets of samples for library preparation. The total RNA fraction was processed by oligo dT bead capture of mRNA, fragmentation of enriched mRNA, reverse transcription into cDNA, end repair of cDNA, ligation with the appropriate Unique Dual Index (UDI) adaptors and strand selection and library amplification by PCR using the Universal Plus mRNA-seq with NuQuant library preparation kit (Tecan Genomics 0520-A01).

Validation of each sample set of library preparations was performed using the D1000 ScreenTape Assay kit (Agilent Technologies 5067-5582) on the Agilent TapeStation 4200. Concentration of each library was determined with the NuQuant module of the library prep kit using a Qubit 4 Fluorometer (ThermoFisher/Invitrogen), Libraries were normalized to 4 nM concentration and pooled as a sample set. Each multiplexed sample pool set was quantitated, in triplicate, using the Roche Lightcycler96 with FastStart Essential DNA Green Master (Roche 06402712001) and KAPA Library Quant (Illumina) DNA Standards 1-6 (KAPA Biosystems KK4903). Using the qPCR results, each

sample pool set was adjusted to 1.9 nM. All 11 normalized sample pool sets were combined for a final normalized pool to perform multiplexed sequencing.

The normalized and pooled libraries were denatured with 0.2N NaOH (0.04N final concentration) and neutralized with 400 mM Tris-HCl pH 8.0. The pooled libraries were diluted on sequencer to 380 pM prior to onboard clonal clustering of the patterned flow cell using the NovaSeq 6000 S4 Reagent Kit v1.5 (200 cycle) (Illumina 20028313). The 2×101 cycle sequencing profile with dual index reads is completed using the following sequence profile: Read 1–101 cycles x Index Read 1–8 cycles × Index Read 2–8 cycles x Read 2–101 cycles. Following collection, sequence data are converted from. bcl file format to "fastq" file format using bcl2fastq software and de-multiplexed into individual sequences for data distribution using a secure FTP site or Illumina BaseSpace for further downstream analysis.

2.6 Differential gene expression

RNA-Sequencing was performed at a strand specific 100 cycle paired-end resolution, in an Illumina NovaSeq 6000 sequencing machine (Illumina, San Diego, and CA). The analysis was performed on two animals (biological duplicate) with samples obtained from different regions subject to different treatments from each animal as described in the text. The samples were multiplexed in an S1 flow-cell, resulting between 33.6 and 57.9 million reads per sample. The read quality was assessed using the FastQC software. On average, the per sequence quality score measured in the Phred quality scale was above 30 for all the samples. The reads were mapped to the *sus scrofa* genome (Sscrofa11.1; Ensembl release 105) using the STAR software, version 2.6.1c. On average, 95% of the sequenced reads mapped to the genome, resulting between 31.8 and 54.8 million mapped reads per sample, of which on average 92% were uniquely mapped reads. Transcript abundance estimates were calculated using the featureCounts software (Liao et al., 2014). Expression normalization and differential gene expression calculations were performed in DESeq2 software to identify statistically significant differentially expressed genes (Love et al., 2014). The samples were analyzed as matched paired samples. The significance *p*-values were adjusted for multiple hypotheses testing by the Benjamini and Hochberg method establishing a false discovery rate (FDR) for each gene. The differential expression analysis of high-throughput data for this study was performed in R using the DESeq2 package. The DESeq2 package performs an independent filtering of the results by default and therefore requires minimal prefiltering. The input to the software was the raw count matrix with genes with less than one count per million in any of the samples removed. The DESeq2 software performs a median of ratios normalization that accounts for sequencing depth and RNA composition.

2.7 Bioinformatic and statistical analysis

Due to the nature of this study being a 'pilot study' only two animals were utilized. Global gene expression was performed for control wounds at 8-h (no dressing insert and no vacuum) and for NPWT-treated wounds at 8-h (Granufoam™ insert and 125 mmHg vacuum), relative to baseline expression at 0-h control (no dressing insert and no vacuum). Additionally, a second set of analyses was performed for differentially expressed genes (DEGs) for 8-h NPWT-treated wounds relative to 8-h control wounds to assess the effect of NPWT treatment relative to the control response. All analyses were performed for each layer relative to its analogous layer (e.g., epidermis relative to epidermis), in addition to the 0-h control "Full/Bulk" wounds. DEGs were investigated by with Ingenuity Pathways Analysis (IPA) software (Qiagen; Ingenuity Systems Inc.; CA and United States). IPA identified canonical pathways, diseases and functions, upstream regulators, and gene networks that were most significant to the relative gene expression profiles for each set of wounds and categorized the DEGs into specific pathways. Heatmaps, graphs, gene clusters, and network webs were automatically generated within IPA and used to visually depict changes in DEGs. A false discovery rate (FDR) below 0.05 was always utilized to assess for DEGs, and was the only threshold qualification when initially assessing for the heterogeneity in global gene expression. Additional thresholding was performed on subsequent pathway analyses, including a minimum -log (*p*-value) of 1.3, which is equal to $p < 0.05$, and an experimental fold change (FC) of 1.5. The additional thresholding was performed for generation of significant canonical signaling pathways, heatmap generation, and top regulatory network analyses. When performing the intra-wound comparison between the different stratified layers to assess for hidden DEGs within the inflammatory, angiogenic, and matrix remodeling pathways, an FDR of 0.05 and a -log (*p*-value) of 1.3 was used to threshold for significantly altered genes.

3 Results

3.1 KUts and KUbes incisional porcine model experimental design

A schematic overview of the incisional porcine wound model can be found in Figure 1. Incisional wounds were created with a depth that penetrated all three skin layers and into the subcutaneous region, followed by insertion of dressing materials into the wounds to permit tissue interactions with the dressings at each layer (Figure 1A). The incisional model was performed as an array on each side of the animal, with different dressing materials making up the rows and different timepoints (0-, 2-, 4-, and 8-h) making up the columns (Figures 1B, C). However, only the 0- and 8-h wound groups were utilized in this study. Additionally, only the genomic analysis was performed in this "proof-of-concept" study due to its higher sensitivity to detect changes in early wound healing pathways (i.e., within 8 h). A complete photographic depiction of the surgical operations can be found in Supplementary Figures S1A–K. Upon excision of the wounds at each respective timepoint, tissue was processed for downstream analysis accordingly. For RNA analysis, samples were dissected/stratified into their three separate skin layers and subsequently processed (Figures 1D, E). The stratification of the wounds allowed for several different intra- and inter-related analyses that were dressing, temporal, or treatment dependent, as well as untreated control wounds to understand physiological signaling (Figure 1E).
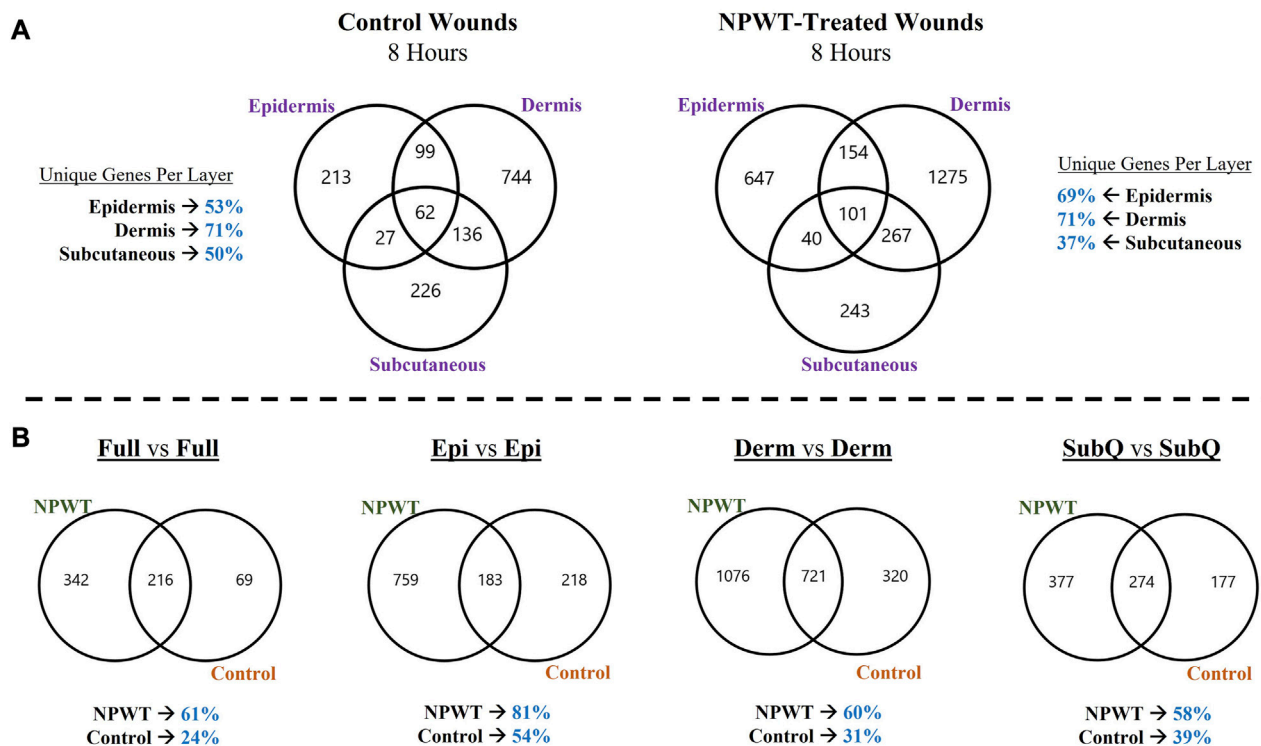
FIGURE 2

Heterogeneity in Global Gene Expression Profiles Between Wound Layers. **(A)** Graphical depiction of differentially expressed genes (DEGs) for Control (*left*) and NPWT-treated (*right*) wounds that were stratified and separated into each specific layer and overlayed. "Layer Specific Genes" denotes the percentage of DEGs that were uniquely expressed in that layer and none of the other layers. **(B)** Direct comparison of unique DEGs for within each layer for either NPWT-treated or control wounds. The percent of uniquely expressed DEGs for each condition is listed below each respective graph. Of note, the graphs do not depict directionality, only overlap in DEGs that have met the predefined threshold of an FDR less than 0.05.

3.2 Heterogeneity in global gene expression profiles between wound layers

To depict the overall heterogeneity of molecular signaling that occurs within the different layers of skin tissue during wound healing, the differentially expressed gene (DEGs) profiles were overlayed and evaluated for uniqueness to specific layers and/or NPWT treatment (NPWT is defined as wounds treated with GranuFoam™ and 125 mmHg vacuum). Of note, the graphs do not depict directionality, only overlap in DEGs that have met the predefined threshold of an FDR less than 0.05 (Figures 2A, B). First, an intra-comparison within a single wound between tissue layers were evaluated for both control healing (no dressing insert and no vacuum at 8-h) and NPWT-treated healing at 8 h, relative to the baseline control wound expression at 0-h, to depict relative uniqueness (Figure 2A). The dermis layer for both control and NPWT wounds exhibited the greatest proportion of DEGs uniquely associated within its layer, at 71%. Whereas the epidermis had the second most and the subcutaneous had the lowest proportion (Figure 2A). Conversely, the relative similarities and overlap between DEGs within the Full vs. Full and Layer vs. Full comparisons, for each layer, is depicted in Supplementary Figure S2 and demonstrates that there was minimal overlap of the layer analyses with the Full analyses, with the average similarity/overlap of the Layer-DEGs with the Full-DEGs coming in at <15% of the total DEGs for each Layer (Supplementary Figure S2). Subsequently, inter-comparison of control versus NPWT-treated wounds within each analogous layer revealed that NPWT treatment modulated the DEG profile, relative to

control wounds, at 8 h 81%, 60%, and 58% of the total DEGs were unique for NPWT treatment for the epidermis, dermis, and subcutaneous regions, respectively (Figure 2B). Notably, the relative proportion of uniquely expressed DEGs in NPWT-treated and 8-h control wounds were similar in the "Full versus Full" comparison group, the total number of DEGs was substantially lower compared to the layer versus layer comparisons. The average total number of DEGs in the layer comparisons was 1130 for NPWT-treated and 631 for 8-h control wounds. On the contrary, in the "Full versus Full" comparison analysis, there was a >50% reduction, with only 558 DEGs for NPWT-treated wounds and 285 in the 8-h control healing wounds.

3.3 Stratified genomic analysis reveals temporal-spatial heterogeneity in canonical signaling pathways

To provide context for the type of data that can be obtained with this methodology, a DEG comparison was performed on the progression of NPWT healing versus untreated control healing at the 8-h timepoint. Gene expression for both NPWT and control 8-h samples were generated relative to the 0-h baseline control wounds (Figure 3A). "Full/Bulk" wound comparisons were performed in parallel to the stratified comparisons, which include the epidermis, dermis, and subcutaneous regions of each wound in the NPWT or 8-h control groups, relative to the analogous region in the 0-h baseline
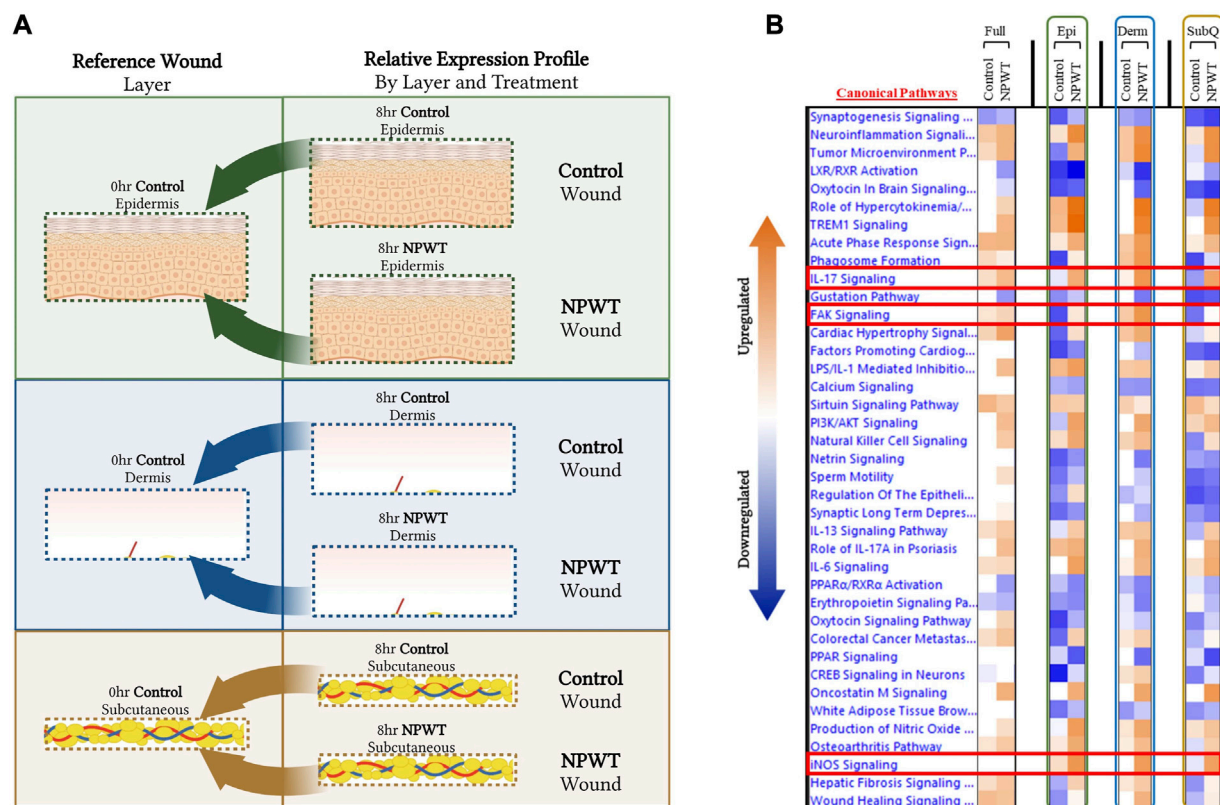
FIGURE 3
Stratified Genomic Analysis Reveals Temporal-Spatial Heterogeneity in Canonical Signaling Pathways. A schematic representation of wounds were stratified and DEGs were generated via relative expression to 0-h control wounds for both 8-h control (untreated) and 8-h NPWT-treated at each layer, including the **(A)** epidermis (*top*), dermis (*middle*), and subcutaneous (*bottom*) regions. **(B)** A heatmap of the top canonical pathways with a FDR of 0.05 and a *p*-value < 0.05. Orange depicts upregulation and blue depicts downregulation of each pathway. White denotes no significant change in pathway activity.

control (Figure 3A). A heatmap of the top canonical pathways was generated for the "Full *versus* Full", "Epidermis *versus* Epidermis", "Dermis *versus* Dermis", and "Subcutaneous *versus* Subcutaneous" comparisons (Figure 3B). When comparing the relative gene expression of the "Full/Bulk" processed wounds, there was minimal changes in expression and canonical signaling, such as that seen in the iNOS signaling pathway, or minimal difference between NPWT and control healing as seen in the IL17 signaling pathway (Figure 3B). However, stratification processing revealed that when focusing on specific regions within the wounds (e.g., epidermis *versus* epidermis), several signaling pathways are revealed to be significantly upregulated or downregulated that were previously not in the "Full/Bulk" comparisons, including inversion of signaling, that is, seen with FAK signaling or the amplification of signaling, that is, seen with iNOS (Figure 3B).

3.4 Tissue stratification reveals canonical signaling and top pathway regulators previously hidden within acute physiological wound healing

To further depict the increased sensitivity and specificity associated with performing gene expression analysis on

stratified wound sections rather than "Full/Bulk", we performed a direct comparative analysis of top canonical pathways and regulators significantly altered during untreated (control) healing at 8 h. Thresholds utilized to determine significance was an FDR 0.05, -log (*p*-value) of 1.3, and a FC of 1.5. When evaluating 8-h control wounds via "Full/Bulk" analysis relative to "Full/Bulk" 0-h baseline control wounds (Figure 4A), only 3 canonical pathways demonstrated a significant modulation, all of which had an FC increase (denoted by orange color), and were nonspecifically associated with inflammation. Expressional changes included the canonical pathways of neuroinflammation, acute phase response, and sirtuin signaling pathways, and the top regulatory network was associated with inhibition of gastrointestinal, lung, and body cavity inflammation (Figures 4B, C). Whereas, when evaluating the same exact 8-h control wounds via stratification of the dermis layer only (Figure 4D), 2 of the 3 same canonical pathways from the "Full/Bulk" analysis were significant, in addition to 17 additional canonical pathways, for a total of 19 significantly altered pathways, most of which exhibited a negative fold change (denoted by blue color). With the top regulatory network being associated with cell movement of leukocytes, homing of cells, and recruitment of granulocytes (Figures 4E, F). Subsequently, the top 6 upstream regulator molecules associated with either activation or inhibition (based

FIGURE 4
Tissue Stratification Reveals Canonical Signaling and Pathway Regulators Previously Hidden in Acute Physiological Wound Healing. **(A, D)** Schematic representations of the **(A)** "Full/Bulk" and **(D)** stratified dermis 8-h and 0-h wounds that were processed for analysis. **(B, E)** Graphical depictions of the top canonical pathways that achieved significance after thresholding, -log (*p*-value) of 1.3 and FC of 1.5. The threshold line is marked at 1.3. Color denotes fold change directionality, with activation/upregulation (*orange*) and inhibition/downregulation (*blue*). **(C, F)** The top regulator effect networks associated with the DEGs for the **(C)** "Full/Bulk" analysis and **(F)** stratified dermis analysis. **(G, H)** The top 6 upstream regulator activators and inhibitors are represented in tabular form for the **(G)** "Full/Bulk" analysis and the **(H)** stratified dermis analysis.

on the z-score) were tabulated for both the "Full/Bulk" and stratified dermis analyses (Figures 4G, H). Of the 12 total markers (6-up/6-down), only 3 were shared between the "Full/Bulk" and the dermis methods, including the top activator TNF, as well as IL1B and IL17A.

3.5 Stratification analysis identifies and localizes key physiological wound responses altered by NPWT not previously observed

Specific wound healing signaling pathways, including inflammation (Figure 5), angiogenesis (Figure 6), and matrix remodeling (Figure 7), were evaluated and DEG analysis was performed with thresholds. All genes included for each pathway passed the thresholds of an FDR of 0.05. Additionally, if a gene reached a significant -log (*p*-value) of 1.3 in at least one of the four comparative groups (Full *vs.* Full, Epi *vs.* Epi, Derm *vs.* Derm, or SubQ *vs.* SubQ), then it was included in the final comparative analysis (Figures 5). For each wound healing signaling pathway, the untreated response was first determined by evaluating the 8-h control relative to the 0-h baseline control wounds (Figures 5A–7A), followed by assessing the direct effect of NPWT via comparing the 8-h NPWT-treated relative to the 8-h control (Figures 5B–7B).

When evaluating inflammatory signaling, there were a total of 23 signaling molecules significantly altered during untreated

healing, whereas when looking specifically at genes that were altered with NPWT relative to the control response (i.e., the effect of NPWT) only 15 were observed, 6 of which were similar between untreated control and NPWT treatment (Figures 5A, B). Notably, 12 of the 23 (52%) genes included in the control response were not significant in the "Full/Bulk" comparison but were within one of the three stratified layer comparisons, including IL1B, IL6, and IL7 (Figure 5A). When evaluating the effect of NPWT, only 3 of the 15 significantly altered genes were observed within the "Full/Bulk" analysis (Figure 5B).

When evaluating angiogenic signaling, there were a total of 28 signaling molecules significantly altered during untreated healing, whereas when looking specifically at genes that were altered with NPWT relative to the control response only 16 were observed, 7 of which were similar between untreated control and NPWT treatment (Figures 6A, B). Notably, 15 of the 28 (54%) genes included in the control response were not significant in the "Full/Bulk" comparison but were within one of the three stratified layer comparisons, including PDGFA, ANGPT2, and INHBB (Figure 6A). Notably, INHBA, SERPINE1, and SMAD2 are all significantly upregulated in all three stratified layers, whereas TIMP4 is significantly downregulated in all three layers. When evaluating the effect of NPWT, 0 of the 16 significantly altered genes were observed within the "Full/Bulk" analysis (Figure 6B). Interestingly, the layer with the most activity in gene expression was the epidermis, with 10 out of 16 of

FIGURE 5
Stratification Analysis Identifies and Localizes Inflammatory Responses Altered by NPWT. An intra-wound analysis that compared DEGs associated with the inflammatory response within each layer of skin and relative to the "Full/Bulk" analyzed wounds that achieved the predetermined significance [significance determined to be a -log (p-value of 1.3 which equals a $p < 0.05$]. Associated DEGs were analyzed in the setting of the **(A)** control wound healing response at 8-h relative to 0-h baseline control and **(B)** NPWT-treated response at 8-h relative to 8-h control. A significance threshold line is denoted at 1.3. Size of bar denotes level of significance. Color of bars denote relative fold change directionality as either upregulated (*orange*) or downregulated (*blue*). DEGs that were common between control and NPWT wounds are "bolded" and denoted with an "*".

the genes significantly modulated, where NPWT resulted in upregulation of TIMP4 and downregulation of SERPINE1. INHBA demonstrated an inversion of signal, with an upregulation in the epidermis and a downregulation within the dermis (Figure 6B). The dermis only had 2 uniquely expressed genes, TIE1 and MMP9, whereas the subcutaneous region had 3 genes, NCOA1, PTGS1, NOTCH4.

When evaluating matrix remodeling signaling, there were a total of 20 signaling molecules significantly altered during untreated healing, whereas when looking specifically at genes that were altered with NPWT relative to the control response only 13 were observed, 6 of which were similar between untreated control and NPWT treatment (Figures 7A, B). Notably, 13 of the 20 (65%) genes included in the control response were not significant in the "Full/Bulk" comparison but were within one of the three stratified layer comparisons, including the notable genes MMP3, MMP8, COL1A1, COL3A1, TIMP1, and TIMP4 (Figure 7A). During untreated acute healing the dermis layer appeared to demonstrate the greatest upregulation of matrix remodeling activity, whereas the subcutaneous exhibited the least (Figure 7A). When evaluating the effect of NPWT, 0 of the 13 significantly altered genes were observed within the "Full/Bulk" analysis (Figure 7B). Interestingly, the subcutaneous region exhibited the greatest alteration of matrix remodeling activity after NPWT, with 8 of the 13 significantly altered genes being uniquely expressed, whereas the epidermis exhibited only 2, and the dermis only 3 (Figure 7B).

4 Discussion

The development of new wound dressings and therapies is an ever-growing field of research, with new targeted and personalized wound dressings for specific applications at the forefront of investigative research (Hodge et al., 2022). However, the skin is a diverse and stratified tissue comprising of diverse populations within the epidermis, papillary dermis, reticular dermis, and subcutaneous adipose (hypodermis) (Shaw and Martin, 2009). Within each stratified layer there are unique compositions of critical structures, including various degrees of lymphatics, blood vessels, sensory nerve endings, glands, and hair follicles. Additionally, each layer is composed of distinct cell populations, as well as supporting cell populations that all ultimately contribute to the processes involved in the maintenance of skin homeostasis and the wound healing response (Shaw and Martin, 2009). Therefore, a targeted approach to analyzing wounds is necessary in order to tease apart and fully understand the immense complexity of the tissue and its regenerative processes. Additionally, generating a fundamental understanding of how acute wound responses relate to long-term trajectory of wound outcomes and how acute interventions and wound dressings modulate the molecular profile of specific cells is critical to advancing targeted therapies.

One such wound therapy, that is, widely utilized is NPWT, and it offers the opportunity for customization and tailoring for specific applications, such as modulating the foam dressing material (e.g.,
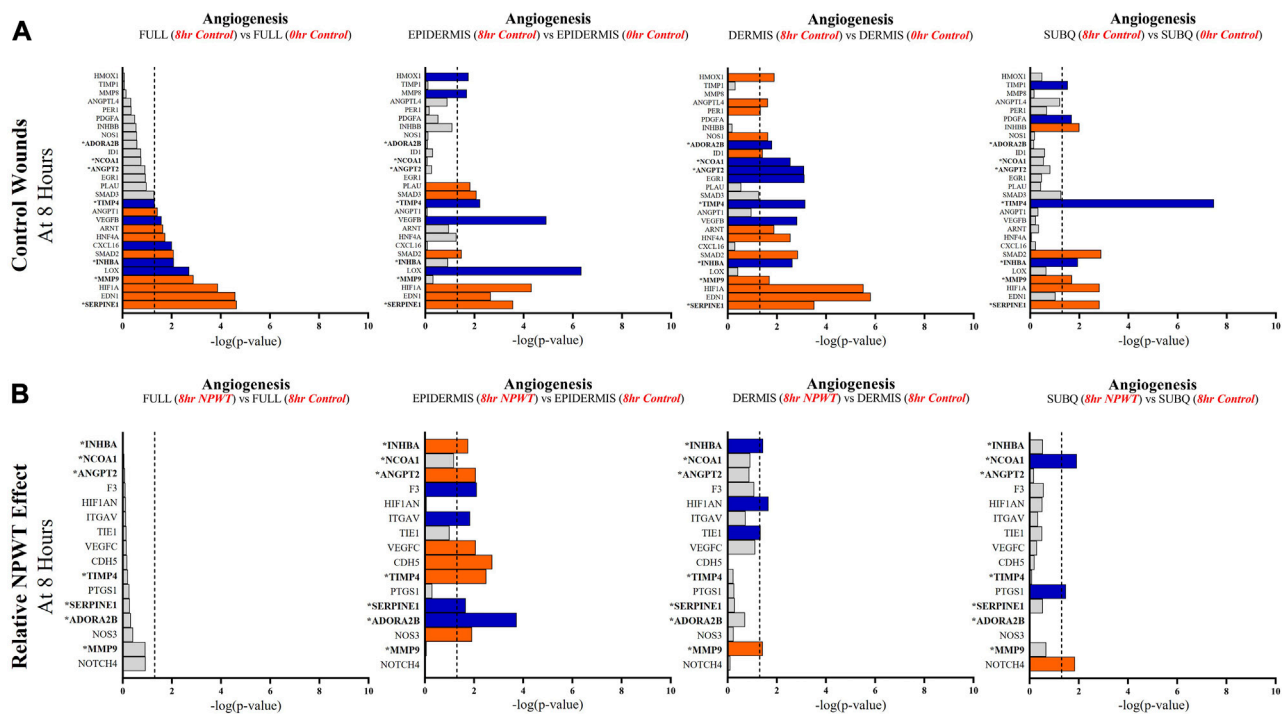
FIGURE 6
Stratification Analysis Identifies and Localizes Angiogenic Responses Altered by NPWT. An intra-wound analysis that compared DEGs associated with the angiogenic response within each layer of skin and relative to the "Full/Bulk" analyzed wounds that achieved the predetermined significance [significance determined to be a -log (p-value of 1.3 which equals a $p < 0.05$]. Associated DEGs were analyzed in the setting of the **(A)** control wound healing response at 8-h relative to 0-h baseline control and **(B)** NPWT-treated response at 8-h relative to 8-h control. A significance threshold line is denoted at 1.3. Size of bar denotes level of significance. Color of bars denote relative fold change directionality as either upregulated (*orange*) or downregulated (*blue*). DEGs that were common between control and NPWT wounds are "bolded" and denoted with an "*".
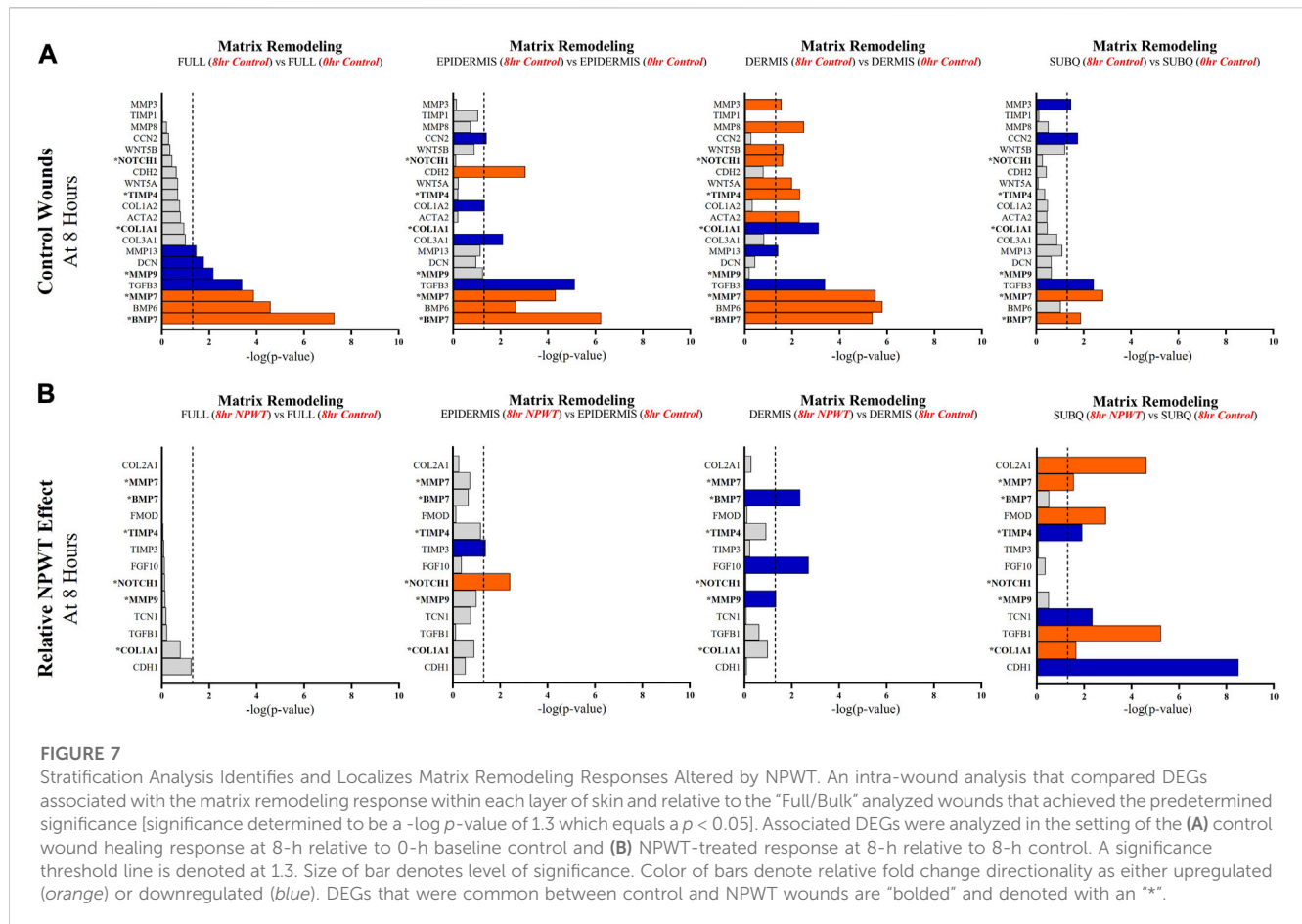
polyurethane *versus* polyvinyl alcohol), structural properties of the dressing (e.g., pore sizes or hydrophobicity), incorporation of instillation and wound irrigation, or may be used in combination with other wound therapies or inclusion of tissue regeneration modalities. However, to date, a basic understanding of the acute temporal-spatial effects of NPWT on wound tissue at a molecular level, relative to physiological healing, is poorly understood. Clinically, we see enhancements of granulation tissue formation, modulation of inflammation, and mitigation of bacterial burden, but how NPWT globally alters specific molecular signaling pathways in both time and space to ultimately achieve these processes is not well documented. Therefore, NPWT was utilized within this methodological study to provide context for a new "proof-of-concept" *in vivo* study design and multiplex molecular analysis approach to provide new insight into NPWT molecular signaling.

In this study a new incisional porcine wound model was created to screen and standardize the processing and analysis of wound interventions in an acute setting, such as NPWT (Supplementary Figure S1). Since understanding the acute physiological responses and interactions of tissue with wound dressing inserts was desired, an incisional model was utilized to maximize the number of timepoints and dressings that could be compared in parallel. Additionally, as we know, early signaling is critical for the proper progression to later downstream signaling pathways, yet there is very little literature on acute wound signaling and genomics. By inflicting a wound with a depth that penetrated through the subcutaneous layer, we were also able to

analyze all three major layers of skin. Each wound was bisected, with one-half serving as the "Full/Bulk" analysis, that is, standard for current wound genomic studies, and the other half was stratified into the 3 layers of the skin to allow for direct comparisons within the same wound (Figures 1D, E). The goal of this study was to depict, in principle, the advantages of this methodology, and thus only a comparison between NPWT-treated (Granufoam™ with 125 mmHg vacuum) and untreated control healing (no dressing insert or vacuum) at 8 h was evaluated.

When assessing for global gene expression changes in control and NPWT-treated wounds, significant diversity was observed in expressional profiles within each layer of skin within the same wounds, with each layer of the control wound exhibiting over 50% of their gene expression unique to that respective layer, with the dermis having the greatest number of unique DEGs at 71%. The NPWT-treated wound exhibited a similar pattern of gene expression diversity (Figure 2A). Thus, the previously utilized "Full/Bulk" approach to wound genomics is an amalgamation of the entire wound healing response within the tissue, however it does not provide any spatial information as to where unique signaling is coming from.

Further extrapolation of the wounds' gene expression profiles was performed via direct comparison of NPWT-treated and control healing at each layer, which depicted extensive heterogeneity in signaling. However, when comparing the global gene expression of the stratified layer analyses *versus* the "Full/Bulk", there was over a 50% decline in total number of DEGs in the "Full/Bulk" group (Figure 2B). This is likely due to competing signals between layers

FIGURE 7
Stratification Analysis Identifies and Localizes Matrix Remodeling Responses Altered by NPWT. An intra-wound analysis that compared DEGs associated with the matrix remodeling response within each layer of skin and relative to the "Full/Bulk" analyzed wounds that achieved the predetermined significance [significance determined to be a -log p-value of 1.3 which equals a $p < 0.05$]. Associated DEGs were analyzed in the setting of the **(A)** control wound healing response at 8-h relative to 0-h baseline control and **(B)** NPWT-treated response at 8-h relative to 8-h control. A significance threshold line is denoted at 1.3. Size of bar denotes level of significance. Color of bars denote relative fold change directionality as either upregulated (*orange*) or downregulated (*blue*). DEGs that were common between control and NPWT wounds are "bolded" and denoted with an "*".

within a wound and dilution of signaling from other layers. This phenomenon is demonstrated in Figure 3, where specific canonical signaling pathways that are minimally altered within the "Full/Bulk" analysis are newly revealed within the stratified analysis or exhibit amplification and/or inversion of signal, which are highlighted with the IL17, FAK, and iNOS signaling pathways (Figure 3B). This suggests that if there is upregulation in one layer of skin and downregulation in another layer, then these two signals could cancel out and result in no signal present when performing a "Full/Bulk" tissue analysis, resulting in a loss in pertinent spatial information regarding the wound healing process. To the authors' knowledge, this is the first depiction of the diverse temporal-spatial modulation of signaling pathways for acute physiological healing in the context of NPWT in such a manner. Thus, the sensitivity and specificity of "Full/Bulk" analytical approaches is drastically hindered due to competing signals within different layers of skin tissue that may result in key hidden signaling pathways.

Next, to further demonstrate the dilution of key signaling pathways in traditional "Full/Bulk" processing and subsequently examine the increased specificity associated with our stratified approach, the authors investigated changes in the top changes in canonical and regulatory pathways within untreated control wounds and compared the traditional "Full/Bulk" approach to changes solely within the dermis to reveal hidden signaling. As expected, there was over a 6-fold increase in significantly altered canonical signaling pathways when performing the stratified analysis with just the dermis, relative to the "Full/Bulk (Figures

4B, E). Moreover, the "Full/Bulk" signaling was less specific and broadly associated with inflammatory activity in a variety of tissues and the top regulatory network was much more complex in nature (Figure 4C). Whereas the dermis stratified analysis revealed 17 hidden canonical pathways, and when assessing the top regulatory network correlated with the dermal gene activity, a more concise and focused network tree was observed (Figures 4E, F). Specifically, the top regulatory network in the dermis was found to be associated with leukocyte and phagocyte migration into the dermis, an expected key wound healing event that occurs in the dermis within the first 24 h (Figure 4F). Unsurprisingly, 3 of the top upstream regulators found in both analyses are IL1B, IL6, and TNF, all of which have broad activity functions during wound healing and the inflammatory response. However, a top upstream regulator found within the dermis that was not noted within the "Full/Bulk" samples is S100A9, a key damage associated molecular pattern (DAMP), that is, found in myeloid cells, such as macrophages, and a known modulator of inflammatory activity, which is recently being investigated for a potential role in re-epithelialization (Figure 4H) (Kerkhoff et al., 2012). Thus, further investigations into physiological regulators may reveal critical prognostic markers for specific wound types (e.g., burns and diabetic ulcers) and potential targets for optimizing current and future wound therapies.

Subsequently, a direct intra-wound and inter-wound comparison between skin layers of DEGs associated with specific wound healing pathways was evaluated, including inflammation, angiogenesis, and matrix remodeling. Genes included in the analysis

for each pathway were derived from a larger gene pool predefined to be associated with each pathway. Thresholding was used to denote DEGs that reached a -log (*p*-value) of 1.3 denoted as significant. However, a DEG had to only reach significance in one of the four groups to be included within the analysis, thus, hidden signaling in a single layer could be identified and its relative expression compared to other layers. This was done for both control wounds at 8-h, relative to the 0-h baseline control, and as a direct comparison of the effect of NPWT at 8-h, relative to 8-h control. It is important to note that the NPWT-treated profile is not the absolute/global profile of the wounds but relative to the 8-h control wounds to assess for the direct effect of NPWT. This approach revealed that over 50% of the significantly expressed DEGs in each of the three pathways being hidden within the "Full/Bulk" analysis for the control group, and over 80% for the NPWT-treated group. Further demonstrating a significant loss in spatial gene expression information in acute wound healing with the "Full/Bulk" analysis.

Further evaluation revealed that several keys signaling markers in all three pathways were being differentially expressed in a specific layer of skin tissue that was not observed in "Full/Bulk" analysis. For example, in the 8-h control groups, the inflammatory markers IL6 and IL1B were only differentially expressed in the epidermis and subcutaneous regions, whereas IL7 and CXCL6 were only expressed in the dermis. Moreover, the epidermis and dermis appear to have similar inflammatory expressional profiles overall, whereas the subcutaneous region had a much less extensive inflammatory role in acute wound healing. Notably, CXCL6 has recently been investigated as a potential biomarker for wound healing in the context of diabetic wounds (Eming et al., 2010; Wang et al., 2019), suggesting a potential origin of this factor may be within the dermis, although this was not a diabetic model and this would need further investigation. Similarly, within the NPWT wounds, the inflammatory marker IL6 was only significantly altered (downregulated) within the epidermis, IL1A was only significantly altered (downregulated) within the dermis, and CSF3 (granulocyte colony-stimulating factor) was only significantly altered (upregulated) within the subcutaneous region (Figures 5A, B). Interestingly, NPWT appeared to exhibit, in the acute setting, an anti-inflammatory stimulus relative to untreated control wounds. Thus, the stratified approach allows for the targeted observation of the inflammatory profile of wounds within defined regions of tissue and the opportunity to assess inflammatory markers for prognostic applications.

Angiogenic and matrix remodeling signaling profiles exhibited similar patterns, with substantial heterogeneity in signaling between layers and 0 differentially expressed genes within the "Full/Bulk" sample groups after NPWT treatment for either pathway. Notably, within the angiogenic pathway, several key markers appear to not be significantly expressed in the acute setting, including VEGFA, NOS1, or FGF2, markers considered to promote angiogenesis in wound healing. Thus, these critical markers may receive input from a different regulatory marker initially before becoming upregulated in wounds, during both untrated control wounds and with NPWT. Interestingly, significant gene expression associated with angiogenesis is occurring within the epidermis, rather than the dermis, including NOS3 (eNOS), CDH5 (VE-Cadherin), and ANGPT2 (Angiopoietin 2). Additionally, there is a direct inversion in signaling for INHBA activity between the epidermis and dermis with NPWT treatment (Figures 6A, B). Conversely, in regard to matrix remodeling, the dermis appears to be the most significant influencer of early signaling of control wounds, with significant modulation of

MMP, TIMP, and BMP signaling. The subcutaneous adipose layer has the least impact during untreated control wound healing. However, this signaling is flipped upon exposure to NPWT, where there is limited additional activity within the dermis but a significant modulation of activity within the subcutaneous layer. This includes an upregulation in TGFB1 activity coming exclusively from the adipose tissue in an acute setting after NPWT treatment, whereas there is not a significant alteration in TGFB1 activity acutely in the control wounds (Figures 7A, B).

An important limitation to point out within this study is the use on only two animals, subsequently resulting in only two biological replicates (n = 2). In order to increase the rigor and external validity of the dataset, a number of steps were taken. First, experts at the Genome Core Facility processed all the tissue and extracted the RNA. A blinded bio-informaticist was utilized to process the raw data counts into DEGs. Then strict parameters were utilized to decrease the likelihood of including false positive and false negative data. Additionally, the DESeq2 package used to obtain DEGs in this study implements advance empirical Bayes methods to estimate gene-specific biological variation under minimal levels of biological replication (as low as two as stated by the authors of the package) by leveraging variance information across genes instead of treating each gene independently (Love et al., 2014; Schurch et al., 2016). Moreover, it is also important to note that the purpose was to introduce a new approach and methodology for processing and analyzing tissue, such as wound tissue, to provide more robust and accurate datasets for downstream analyses.

Although there are a number of studies to date looking into genomic expression profiles of wound healing with or without a wound interventional strategy (e.g., NPWT), there is limited information regarding a temporal and spatial perspective combined, especially in an acute setting (Derrick and Lessing, 2014; Ma et al., 2016; Brownhill et al., 2021). Moreover, there are not any standardized models for assessing wound interventions and dressings, such as that proposed in this study. There are recent advancements in genome sequencing, such as single cell sequencing or 10X spatial transcriptomics, now available that provide a wealth of information and may be more appropriate in certain scenarios. However, there are still limitations to these strategies, including lack of spatial information in single cell sequencing and a need to keep tissue/cells alive and viable, which can be difficult for large scale *in vivo* studies (Williams et al., 2022). Additionally, spatial transcriptomics can be associated with concerns of feasibility in larger studies due to the requirement of imaging every wound/tissue section, in addition to accessibility to sequencing cores with the right equipment and staff, and the economical burdens associated with these newer expensive spatial transcriptomic modalities (Williams et al., 2022). One of the many benefits of this methodology is that it can be paired with any currently available genomics core capable of performing RNA sequencing, or other genomic analysis technique preferred, such as microarrays or RT-PCR. Therefore, the methodology presented in this study provides an alternative to those unable to perform the abovementioned methods, but can also work synergistically with those methods when paired appropriately.

5 Conclusion

Overall, the goal of this study is to highlight the need for more targeted genomic analyses of wound healing modalities and subsequently

evaluate the validity of the approach proposed. This study demonstrated a substantially different genomic perspective when stratification of wound tissue is performed rather than "Full/Bulk" analyses. Thus, there could be specific dynamic molecular mechanisms responsible for many currently utilized wound healing therapies that can be potentially identified with this methodology and tailored for future therapies. Additionally, with the increased sensitivity and specificity of a stratified tissue analysis, biomarkers for chronic and/or complex wounds could be identified and wound dressings can be tailored to release specific biomodulatory compounds within a specific region of the wound. Moreover, pairing the genomic data with additional analytical approaches such as proteomics, histology, electron microscopy, and immunolabeling permits the opportunity to not only better understand how wound interventions augment the healing process, but to also provide new insight into physiological healing at a deeper level. The results from this study and continued investigations may be able to directly translate to the clinical setting in a variety of scenarios via the optimization of therapeutic parameters for current utilized therapies, such as NPWT.

Data availability statement

Raw data uploaded to the NCBI GEO database, series ref. #GSE233863. Please use the following link to access https://www. ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE233863.

Ethics statement

The animal study was reviewed and approved by the University of Kansas Medical Center (KUMC) Institutional Animal Care and Use Committee (IACUC) under animal care and use protocol (ACUP) #2019-2535.

Author contributions

JH, DZ, and AM conceptualized the manuscript and established the experimental design for the manuscript. JH served as the primary surgeon on the day of the experiments. DZ and AM aided in the processing and note taking during the surgical operations. JH prepared and processed all tissue samples. SG processed the raw genomics data and converted the RNA library into an appropriate differential expression data format. SG and JH performed the bioinformatics for the study. JH analyzed the data and generated the figures for the manuscript and wrote the text. SG provided the text for the methodology over the bioinformatic components. JR, SG, DZ, RK, and AM provided feedback and aided in the manuscript revision process. JH, SG, DZ, RK, JR, and AM all provided a final review of the manuscript text to check for accuracy and completion. All authors contributed to the article and approved the submitted version.

Funding

Research reported in this publication was supported by the Zam Research LLC and KUMC Department of Plastic Surgery. The content is solely the responsibility of the authors and does not necessarily represent the official views of the KUMC or Zam Research LLC. The KUMC Genomics Facility is funded via four NIH grants that include Kansas Intellectual and Developmental Disabilities Research Center (NIH U54 HD 090216), the Molecular Regulation of Cell Development and Differentiation COBRE (P30 GM122731-03), the NIH S10 High-End Instrumentation Grant (NIH S10OD021743), and the Frontiers CTSA grant (UL1TR002366). Additional private commercial resources and equipment were provided by Ronawk, Inc.

Acknowledgments

The authors would like to thank the team members from University of Kansas Medical Center (KUMC) Department of Plastic Surgery and Ronawk Inc. for their clinical insight, support, and collaboration on this project. The authors would like to provide specific acknowledgements to Briauna Hawthorne, Braden Stuart, and Kai Simmons for their assistance with the surgical operations and sample collection. The authors would also like to acknowledge the University of Kansas Medical Center Genomics Core facility for their assistance with processing the RNASeq samples in this study. Additionally, the authors would like to thank the veterinary staff at the KUMC Laboratory Animal Research (LAR) facility for their assistance with the animals and surgical operations.

Conflict of interest

All authors declare that they do not have any conflict of interest except DZ and AM. DZ declares that he has sold patents to the VC. to KCI/Acelity and continues to receive royalties on Prevena (KCI/ Acelity, now KCI/3M). DZ is the owner and founder of Zam Research LLC. AM is the co-founder and CEO of Ronawk, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author JR declared that they were an editorial board member of Frontiers at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmmed.2023. 1195822/full#supplementary-material

References

Agarwal, P., Kukrele, R., and Sharma, D. (2019). Vacuum assisted closure (VAC)/ negative pressure wound therapy (NPWT) for difficult wounds: a review. *J. Clin. Orthop. Trauma* 10, 845–848. doi:10.1016/j.jcot.2019.06.015

Andorko, J. I., and Jewell, C. M. (2017). Designing biomaterials with immunomodulatory properties for tissue engineering and regenerative medicine. *Bioeng. Transl. Med.* 2, 139–155. doi:10.1002/btm2.10063

Argenta, L. C., and Morykwas, M. J. (1997). Vacuum-assisted closure: a new method for wound control and treatment: clinical experience. *Ann. Plast. Surg.* 38, 563–577. doi:10.1097/00000637-199706000-00002

Boehler, R. M., Graham, J. G., and Shea, L. D. (2011). Tissue engineering tools for modulation of the immune response. *Biotechniques* 51, 239–240. doi:10.2144/000113754

Brownhill, V. R., Huddleston, E., Bell, A., Hart, J., Webster, I., Hardman, M. J., et al. (2021). Pre-clinical assessment of single-use negative pressure wound therapy during *in vivo* porcine wound healing. *Adv. Wound Care (New Rochelle)* 10, 345–356. doi:10.1089/wound.2020.1218

Derrick, K. L., and Lessing, M. C. (2014). Genomic and proteomic evaluation of tissue quality of porcine wounds treated with negative pressure wound therapy in continuous, noncontinuous, and instillation modes. *Eplasty* 14, e43.

Dreifke, M. B., Jayasuriya, A. A., and Jayasuriya, A. C. (2015). Current wound healing procedures and potential care. *Mater Sci. Eng. C Mater Biol. Appl.* 48, 651–662. doi:10.1016/j.msec.2014.12.068

Eming, S. A., Koch, M., Krieger, A., Brachvogel, B., Kreft, S., Bruckner-Tuderman, L., et al. (2010). Differential proteomic analysis distinguishes tissue repair biomarker signatures in wound exudates obtained from normal healing and chronic wounds. *J. Proteome Res.* 9, 4758–4766. doi:10.1021/pr100456d

Eming, S. A., Martin, P., and Tomic-Canic, M. (2014). Wound repair and regeneration: mechanisms, signaling, and translation. *Sci. Transl. Med.* 6, 265sr6. doi:10.1126/scitranslmed.3009337

Gupta, S. (2012). Optimal use of negative pressure wound therapy for skin grafts. *Int. Wound J.* 9 (1), 40–47. doi:10.1111/j.1742-481X.2012.01019.x

Hodge, J. G., Pistorio, A. L., Neal, C. A., Dai, H., Nelson-Brantley, J. G., Steed, M. E., et al. (2021). Novel insights into negative pressure wound healing from an *in situ* porcine perspective. *Wound Repair Regen.* 30, 64–81. doi:10.1111/wrr.12971

Hodge, J. G., Zamierowski, D. S., Robinson, J. L., and Mellott, A. J. (2022). Evaluating polymeric biomaterials to improve next generation wound dressing design. *Biomater. Res.* 26, 50. doi:10.1186/s40824-022-00291-5

Jang, Y. H., Jin, X., Shankar, P., Lee, J. H., Jo, K., and Lim, K. I. (2019). Molecular-level interactions between engineered materials and cells. *Int. J. Mol. Sci.* 20, 4142. doi:10.3390/ijms20174142

Kerkhoff, C., Voss, A., Scholzen, T. E., Averill, M. M., Zanker, K. S., and Bornfeldt, K. E. (2012). Novel insights into the role of S100A8/A9 in skin biology. *Exp. Dermatol* 21, 822–826. doi:10.1111/j.1600-0625.2012.01571.x

Kim, P. J., Attinger, C. E., Constantine, T., Crist, B. D., Faust, E., Hirche, C. R., et al. (2020). Negative pressure wound therapy with instillation: international consensus guidelines update. *Int. Wound J.* 17, 174–186. doi:10.1111/iwj.13254

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi:10.1093/bioinformatics/btt656

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Ma, Z., Shou, K., Li, Z., Jian, C., Qi, B., and Yu, A. (2016). Negative pressure wound therapy promotes vessel destabilization and maturation at various stages of wound healing and thus influences wound prognosis. *Exp. Ther. Med.* 11, 1307–1317. doi:10.3892/etm.2016.3083

Morykwas, M. J., Argenta, L. C., Shelton-Brown, E. I., and Mcguirt, W. (1997). Vacuum-assisted closure: a new method for wound control and treatment: animal studies and basic foundation. *Ann. Plast. Surg.* 38, 553–562. doi:10.1097/00000637-199706000-00001

Morykwas, M. J., Faler, B. J., Pearce, D. J., and Argenta, L. C. (2001). Effects of varying levels of subatmospheric pressure on the rate of granulation tissue formation in experimental wounds in swine. *Ann. Plast. Surg.* 47, 547–551. doi:10.1097/00000637-200111000-00013

Morykwas, M. J., Simpson, J., Punger, K., Argenta, A., Kremers, L., and Argenta, J. (2006). Vacuum-assisted closure: state of basic research and physiologic foundation. *Plast. Reconstr. Surg.* 117, 121S–126S. doi:10.1097/01.prs.0000225450.12593.12

Norman, G., Shi, C., Goh, E. L., Murphy, E. M., Reid, A., Chiverton, L., et al. (2022). Negative pressure wound therapy for surgical wounds healing by primary closure. *Cochrane Database Syst. Rev.* 4, CD009261. doi:10.1002/14651858.CD009261.pub5

Normandin, S., Safran, T., Winocour, S., Chu, C. K., Vorstenbosch, J., Murphy, A. M., et al. (2021). Negative pressure wound therapy: mechanism of action and clinical applications. *Semin. Plast. Surg.* 35, 164–170. doi:10.1055/s-0041-1731792

Reinke, J. M., and Sorg, H. (2012). Wound repair and regeneration. *Eur. Surg. Res.* 49, 35–43. doi:10.1159/000339613

Schurch, N. J., Schofield, P., Gierlinski, M., Cole, C., Sherstnev, A., Singh, V., et al. (2016). Erratum: how many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22, 1641. doi:10.1261/rna.058339.116

Shaw, T. J., and Martin, P. (2009). Wound repair at a glance. *J. Cell Sci.* 122, 3209–3213. doi:10.1242/jcs.031187

Thevenot, P., Hu, W., and Tang, L. (2008). Surface chemistry influences implant biocompatibility. *Curr. Top. Med. Chem.* 8, 270–280. doi:10.2174/156802608783790901

Wang, X., Li, J., Wang, Z., and Deng, A. (2019). Wound exudate CXCL6: a potential biomarker for wound healing of diabetic foot ulcers. *Biomark. Med.* 13, 167–174. doi:10.2217/bmm-2018-0339

Webster, J., Scuffham, P., Stankiewicz, M., and Chaboyer, W. P. (2014). Negative pressure wound therapy for skin grafts and surgical wounds healing by primary intention. *Cochrane Database Syst. Rev.* 4, CD009261. doi:10.1002/14651858.CD009261.pub2

Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R., and Haque, A. (2022). An introduction to spatial transcriptomics for biomedical research. *Genome Med.* 14, 68. doi:10.1186/s13073-022-01075-1

Frontiers in
Molecular Medicine

Explores mechanisms of human disease using cutting-edge technologies

This transdisciplinary journal explores the molecular mechanisms of human pathologies and the development of diagnostics and therapeutics tools.

Discover the latest Research Topics

See more →

frontiers | Research Topics