NON-GENETIC HETEROGENEITY IN DEVELOPMENT AND DISEASE

EDITED BY: Mohit Kumar Jolly, Jean-Pascal Capp and Ankur Sharma
PUBLISHED IN: Frontiers in Genetics, Frontiers in Bioengineering and Biotechnology and Frontiers in Cell and Developmental Biology

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-421-6
DOI 10.3389/978-2-88971-421-6

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

NON-GENETIC HETEROGENEITY IN DEVELOPMENT AND DISEASE

Topic Editors:
Mohit Kumar Jolly, Indian Institute of Science (IISc), India
Jean-Pascal Capp, Institut National des Sciences Appliquées de Toulouse (INSA), France
Ankur Sharma, Harry Perkins Institute of Medical Research, Australia

Citation: Jolly, M. K., Capp, J.-P., Sharma, A., eds. (2021). Non-Genetic Heterogeneity in Development and Disease. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-421-6

Table of Contents

frontiers in Genetics

EDITORIAL
published: 09 August 2021
doi: 10.3389/fgene.2021.731814

Editorial: Non-Genetic Heterogeneity in Development and Disease

Jean-Pascal Capp[1]*, *Mohit Kumar Jolly*[2]* and *Ankur Sharma*[3,4]*

[1] Toulouse Biotechnology Institute, INSA/University of Toulouse, CNRS, INRAE, Toulouse, France, [2] Centre for BioSystems Science and Engineering, Indian Institute of Science, Bangalore, India, [3] QEII Medical Centre and Centre for Medical Research, Harry Perkins Institute of Medical Research, The University of Western Australia, Perth, WA, Australia, [4] Curtin Medical School, Curtin Health Innovation Research Institute, Curtin University, Perth, WA, Australia

Keywords: cancer, inflammation, epigenetics, plasticity, transcription

Editorial on the Research Topic

Non-Genetic Heterogeneity in Development and Disease

Genetically identical populations of mammalian cells can demonstrate inherent heterogeneity in gene expression and consequent functional behavior. This non-genetic heterogeneity can emerge due to stochasticity in gene expression, plasticity emerging due to gene regulatory networks, and the epigenetic state of cells. The non-genetic heterogeneity of cells can determine cell fate decision and differential response and adaptation to varying environmental conditions. Non-genetic heterogeneity of stromal and tissue resident stem cells are also known to play an important role in development, cancer, and infectious diseases. Thus, non-genetic heterogeneity is emerging as a major player in mediating resistance to existing therapies. In this Research Topic, we aimed to assemble a collection of manuscripts that address the following important questions: What is the major source of non-genetic heterogeneity, and what is its role in homeostasis and pathobiology? Why and how does non-genetic resistance occur?

After exploring the literature revealing the stochastic nature of cell differentiation and the role of stochastic gene expression (the so-called "gene expression noise") in this process, Capp and Laforge propose to consider an alternative model of development named ontophylogenesis where the generation of a differentiated state is considered as a constrained random process. The chance-selection principle governing cell differentiation would be based on the randomness of biochemical reactions at lower scales on which the multiscale constraints produced by the dynamical organization of the biological system retroact, thus driving the system toward a stabilized state of equilibrium. Mitchell highlights the experimental and computational systems biology studies that have been instrumental in decoding how B-cells achieve distinct fates and the implications of various mutations. The article focuses on mechanisms leading to cell-to-cell variability in B-cell terminal differentiation, and consequences on population heterogeneity in terms of decision-making timings and population distribution proportions. Next, Sha et al. demonstrate using single-cell transcriptomic data on epithelial-mesenchymal transition (EMT) that intermediate cell-states along the EMT spectrum can play a crucial role in TGFb-induced EMT. Analyzing the trajectory of cell-state transitions induced by various growth factors, they highlight how intermediate states can be instrumental for cell-cell communication, highlighting a role of non-cell-autonomous factors in decision-making. Tonn et al. discuss the "metabolic" phenotypic heterogeneity at single cell resolution. They propose a mixture model for systematic prediction of the impact of biochemical parameters on the metabolite distribution at single cell level. This study opens the avenue for uncharted territory of single cell metabolic heterogeneity.

Gessain et al. discuss the non-genetic heterogeneity in the immune system particularly in macrophages. They specifically discuss the heterogeneity of macrophages in human diseases in

OPEN ACCESS

Edited and reviewed by:
Richard D. Emes,
University of Nottingham,
United Kingdom

***Correspondence:**
Jean-Pascal Capp
capp@insa-toulouse.fr
Mohit Kumar Jolly
mkjolly@iisc.ac.in
Ankur Sharma
ankur.sharma@perkins.org.au

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *28 June 2021*
Accepted: *19 July 2021*
Published: *09 August 2021*

Citation:
Capp J-P, Jolly MK and Sharma A (2021) Editorial: Non-Genetic Heterogeneity in Development and Disease. Front. Genet. 12:731814. doi: 10.3389/fgene.2021.731814

the context of infection, inflammation, metabolism, aging and cancer. Finally they summarize how non-genetic heterogeneity in macrophages may impact the pathophysiology in humans and its implication in therapeutic targeting. Specifically on cancer, Guinn et al. propose in a stimulating article that the role of gene expression noise in metastasis should be investigated by two complementary approaches. On the one hand, the authors discuss the monitoring and cataloging of naturally occurring gene expression variability to establish associations with cancer progression and metastasis, and suggest that three different types of noise-modulated threshold crossing (multistability, hypersensitivity, and irreversibility) should be more particularly studied in the context. On the other hand, they propose to experimentally modulate protein noise independently of the mean through synthetic biological gene circuits to confirm the role of non-genetic heterogeneity in disease development, stress survival, and metastasis.

From a therapeutic viewpoint, Biswas discusses the phenotypic heterogeneity in treatment response, specifically the heterogeneity in cellular response and downstream signaling and its impact on treatment response. Interestingly, he discusses a similar mechanism of action during heterogeneous cellular responses. Farquhar et al. discuss how a combination of computational and experimental approaches helps decoding the design principles of fractional killing and non-genetic heterogeneity implicated in antimicrobial resistance (AMR). They also expound the implications of these ideas in cancer drug resistance and underscore the importance of synthetic biology attempts based on the design principles of regulatory networks, which can help discover effective strategies against AMR.

Deshmukh and Saini adopt a broader perspective by considering the evolutionary implications of phenotypic heterogeneity at all levels of life, from viruses to mammals. On the one side, the authors particularly emphasize the role of non-genetic variability during organismal development (with *Caenorhabditis elegans* as an example) or within a specialized organ system (specifically spermatogenesis). On the other side, its potential initiating and promoting role in the onset of cancer is largely discussed, with detailed examples over apoptosis, signaling, metabolism, as well as drug resistance. Finally, moving beyond investigating phenotypic plasticity and heterogeneity, Clairambault takes the focus on cancer progression from an evolutionary perspective, presenting a breakdown of multicellularity as one of the hallmarks of cancer. The article asks poignant questions about the connection between emergence of multicellularity and that of cancer, and argues that investigating how multicellularity originated can have important insights into understanding how it breaks down during multiple stages of cancer progression. This series of stimulating articles highlight that non-genetic heterogeneity should be considered as a central component in development and disease, and reveal that innovative theories and experiments in modern biology can be elaborated and conducted by considering non-genetic heterogeneity as a driving force in physiological and pathological systems.

AUTHOR CONTRIBUTIONS

J-PC, MJ, and AS conceived, wrote, and edited the final version of this editorial. All authors contributed to the article 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.

frontiers in Cell and Developmental Biology

HYPOTHESIS AND THEORY
published: 23 July 2020
doi: 10.3389/fcell.2020.00659

A Darwinian and Physical Look at Stem Cell Biology Helps Understanding the Role of Stochasticity in Development

Jean-Pascal Capp[1] and Bertrand Laforge[2]*

[1] Toulouse Biotechnology Institute, University of Toulouse, INSA, CNRS, INRAE, Toulouse, France, [2] LPNHE, UMR 7585, Sorbonne Université, CNRS/IN2P3, Université de Paris, Paris, France

Single-cell analysis allows biologists to gain huge insight into cell differentiation and tissue structuration. Randomness of differentiation, both *in vitro* and *in vivo*, of pluripotent (multipotent) stem cells is now demonstrated to be mainly based on stochastic gene expression. Nevertheless, it remains necessary to incorporate this inherent stochasticity of developmental processes within a coherent scheme. We argue here that the theory called ontophylogenesis is more relevant and better fits with experimental data than alternative theories which have been suggested based on the notions of self-organization and attractor states. The ontophylogenesis theory considers the generation of a differentiated state as a constrained random process: randomness is provided by the stochastic dynamics of biochemical reactions while the environmental constraints, including cell inner structures and cell-cell interactions, drive the system toward a stabilized state of equilibrium. In this conception, biological organization during development can be seen as the result of multiscale constraints produced by the dynamical organization of the biological system which retroacts on the stochastic dynamics at lower scales. This scheme makes it possible to really understand how the generation of reproducible structures at higher organization levels can be fully compatible with probabilistic behavior at the lower levels. It is compatible with the second law of thermodynamics but allows the overtaking of the limitations exhibited by models only based on entropy exchanges which cannot cope with the description nor the dynamics of the mesoscopic and macroscopic organization of biological systems.

Keywords: stochastic gene expression, non-genetic heterogeneity, pluripotency, multipotency, biological organization, emergence, constraints, ontophylogenesis

OPEN ACCESS

Edited by:
Andrei V. Chernov, University of California, San Diego, United States

Reviewed by:
Zhixing Cao, East China University of Science and Technology, China
Jonathan Arnold, University of Georgia, United States

****Correspondence:***
Jean-Pascal Capp
capp@insa-toulouse.fr

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Cell and Developmental Biology

Received: *22 January 2020*
Accepted: *01 July 2020*
Published: *23 July 2020*

Citation:
Capp J-P and Laforge B (2020) A Darwinian and Physical Look at Stem Cell Biology Helps Understanding the Role of Stochasticity in Development. Front. Cell Dev. Biol. 8:659. doi: 10.3389/fcell.2020.00659

INTRODUCTION

While pioneering works early suggested its preeminent role in cell differentiation, the notion of stochasticity has been increasingly considered in developmental and stem cell biology only since a decade, thanks to advances in single-cell analyses [see among others (Hayashi et al., 2008; MacArthur et al., 2012; Abranches et al., 2014; Mojtahedi et al., 2016; Moris et al., 2016; Richard et al., 2016; Moussy et al., 2017; Semrau et al., 2017; Stumpf et al., 2017)]. Both the inherent probabilistic behavior of embryonic and adult stem cells related to their highly stochastic gene expression (SGE), and the subsequent randomness of differentiation events, encourage to make this stochasticity a key cellular property on which a modern theory of development should be

based. Some attempts have been made to introduce mechanisms explaining SGE from the chaotic behavior of a presupposed underlying gene network in the cells. Using the ordinary differential equation (ODE) formalism, the notion of attractor is exhibited, which coupled with Waddington's idea of canalization, provides a framework increasingly quoted in the literature. Nevertheless, this approach where SGE results from a chaos-induced noise over a purely deterministic inner cell network has its own inner auto-coherence issues (Kupiec, 2009). On the opposite, both biological and physical considerations allow arguing for a theory that considers the acquisition of differentiated features as the result of decreased cellular stochasticity linked to the appearance of dynamically evolving environmental constraints acting at the cell (through compartmentation and mesoscopic structures such as chromatin, nuclear membrane, etc.), tissue (through cell-cell interactions and communications) and organism (through endocrine, immunity and blood networks) levels. In first approximation, these constraints can be defined as the external conditions that exert on a system. Nevertheless, besides this internal/external dichotomy, in an extended system, local dynamics inside the system can also constitute constraints (see below). Here, these constraints take place concomitantly to the stochastic appearance of cellular phenotypes during developmental processes, and cannot be envisaged independently of cellular behavior. The intertwined and reciprocal influences of both cellular stochasticity and environmental constraints can account for the emergence of developing structures in multicellular organisms and provide a paradigm which is fully compatible with physical and chemical laws at molecular and mesoscopic scales. Importantly, this approach provides an easy explanation of cell reprogramming and cell pluripotency capabilities.

TRANSCRIPTIONAL HETEROGENEITY IN STEM CELLS

Molecular Basis of Transcriptional Heterogeneity in Embryonic Stem Cells

Following the establishment of the first embryonic stem cells (ESC) lines in 1998 (Thomson et al., 1998), a decade of works led to characterization of the ESC state (Young, 2011). Especially the pluripotency network centered on the three pivotal transcription factors (TF) Oct4, Nanog and Sox2 has been defined (Orkin and Hochedlinger, 2011; Li and Izpisua Belmonte, 2018). First, these factors act to positively regulate their own promoters creating an interconnected autoregulatory loop. Second, they occupy and activate the expression of genes necessary to maintain the ESC state, and contribute to repress genes encoding lineage-specific TF. Thus they prevent exit from the pluripotent state (Young, 2011).

Moreover ESC possess an unusual nuclear structure where DNA is arranged in a less compacted chromatin structure compared to differentiated cells. This is associated with the rapid turnover of chromatin proteins which is a hallmark of weak interactions with DNA (Meshorer et al., 2006). Accordingly, ESC are enriched in epigenetic marks associated with elevated gene expression, and possess less marks that compact chromatin enough to prevent any gene expression (Spivakov and Fisher, 2007). This enrichment has later been linked to widespread, generalized, stochastic and pervasive gene expression in ESC (Efroni et al., 2008), with most of the tissue-specific genes analyzed being sporadically expressed at low level while they were not expected in ESC. On the opposite, a large-scale repression of gene expression occurs during differentiation with expression profiles becoming more specific and stable in differentiated cells. Thus stem cells cannot be defined as a cell type characterized by stable phenotypes. They harbor a permanent unstable state and do not stably express specific genes.

Some TF considered to be specific of ESC, such as Nanog (Chambers et al., 2007) and others (Hayashi et al., 2008) involved in pluripotency are expressed with high heterogeneity from cell-to-cell. Interestingly, the propensity to differentiate depends on their expression levels (Kalmar et al., 2009; Radzisheuskaya et al., 2013). Especially, mouse ESC expressing Nanog at a lower level can differentiate more easily and can be considered as being in a "primed" state (Abranches et al., 2014). Cells expressing Nanog at higher levels possess more stable gene expression profiles in various growth conditions, and a minimal expression of differentiation genes (Abranches et al., 2014). Nevertheless, ESC populations cannot be simply divided into two distinct subpopulations with higher and lower levels of pluripotency factors. There is a continuum of states from high to low pluripotency (Klein et al., 2015). Exit from pluripotency occurs only through a continuum of intermediary states and not as an abrupt transition toward a new state (Hough et al., 2014). Thus pluripotency appears to be more a state of dynamic heterogeneity of a population driven by transcriptional noise than a discrete state dependent on the fixed expression of a small set of genes (Kalmar et al., 2009). Moreover the function of the gene regulatory network centered around Nanog might be to control this dynamic heterogeneity (Kalmar et al., 2009).

Transcriptional Heterogeneity in Adult Multipotent Stem Cells

Generalized transcription activity was already observed in hematopoietic stem cells (HSC) in the early 2000s when terminal differentiation genes were shown to be present before any commitment, with a large part of the genome expressed in a less differentiated state (Terskikh et al., 2003). This phenomenon has been later studied by the most powerful tools in molecular and cellular biology which confirmed that differentiation of hematopoietic cells is clearly probabilistic (Hume, 2000; Chang et al., 2008; Pina et al., 2012). The pioneering work led by Sui Huang on progenitors of the hematopoietic system led to two major observations (Chang et al., 2008). On the one hand, these cells heterogeneously express Sca-1 which is associated with HSC, and when a subpopulation with homogenous Sca1 levels is isolated and regrown, the initial heterogeneity is restored upon a few days. On the other hand, the Sca1 level is linked to the propensity to differentiate, reflecting the important bias associated with its expression.

More molecular details about the phenomena beyond the stochastic behavior of HSC were obtained when other authors observed that different regulators of HSC differentiation vary independently and stochastically in these multipotent cells, what has been linked to priming toward different cell types (Pina et al., 2012; Moignard et al., 2015; Moussy et al., 2017; Giladi et al., 2018). For instance, the erythroid lineage is primed while many markers of other lineages are also expressed (Pina et al., 2012). During cell fate specification of haematopoietic multipotential progenitor cells, mixed-lineage intermediates with concurrent expression of haematopoietic stem cell/progenitor and myeloid progenitor TF seem to be mandatory (Olsson et al., 2016). These results showed that there is no coordination in the HSC expression pattern. Bone marrow differentiation is by far less hierarchical than previously assumed (Giladi et al., 2018). The early differentiation steps occur sporadically by the stochastic expression of differentiation genes in individual cells (Muers, 2012). Study of the early differentiation of T cells led to the same conclusion that, contrary to the expression of only one set of lineage-specific TF, there is an ubiquitous high-level co-expression of antagonistic TF in individual cells (Fang et al., 2013). Moreover, cytokines, which are generally correlated to specific lineage, are stochastically expressed in small subpopulations without being associated to specific TF (Fang et al., 2013). More generally, genes are continuously repurposed across lineages and fates, bringing single cells from different lineages closer in transcriptional space and skewing models for differentiation dynamics (Giladi et al., 2018).

The first analysis at the single-cell scale of the development of an entire mammalian organ was performed on the hematopoietic system (Moignard et al., 2015). By studying the early development of blood cells in mice, it was shown that the early differentiation of HSC is asynchronous, with cells at multiple stages expressing similar combinations of developmental regulators. There is a continuous emergence of blood-specified cells underlying an increase in the proportion of cells expressing a given gene between stages. Graded expression changes over time are not consistent with a discrete switch at a specific developmental moment. This work showed that cells destined to become blood and endothelial cells arise at all stages of the analyzed time course rather than in a synchronized fashion at one precise moment.

Most of these works observed and acknowledged the existence of heterogeneity in cell differentiation, but still argued for a deterministic model with a certain degree of noise. Nevertheless, the best illustration of the dynamic nature and variable timing of cell fate commitment of HSC was provided by Moussy et al. (2017) and led to the conclusion that cell fate decision is a dynamic and complex process characterized by fluctuations that are far from the simple deterministic and binary switches between opposite options that are usually envisioned.

Necessity for a "Statistical Mechanics" of Pluripotency

What these works performed in the last decade revealed is at the opposite of the model predominant in the early 2000s where stem cells were supposed to be finely regulated to homogeneously express only few genes allowing pluripotency and self-renewal, and to homogeneously react to differentiation cues. Only few researchers early highlighted the fact, mainly based on the hematopoietic system, that it seems indeed impossible to define a molecular "signature" of stem cells and that generalized and variable gene expression is what defined stemness (Zipori, 2004). From this perspective, one can argue that differentiation is above all the suppression of this widespread SGE (Efroni et al., 2009) and even propose a model where it is highly determined by the transition of the chromatin organization from a dynamic and open state to a more stable and closed state (Ram and Meshorer, 2009).

In the light of these results revealing the importance of randomness in the stem cell behavior, pioneering works by Paldi suggested the central role of the chromatin in the suppression of stochasticity and proposed an explicit relationship between metabolism and the level of cellular variations (Paldi, 2003). MacArthur and Lemischka (2013) proposed in 2013 to apply the concepts of statistical mechanics to pluripotency. Their main idea is that pluripotency is not a property of a cell but the property of a cell population which can consequently be analyzed using the concept of entropy that measures how the different patterns of genetic expression are distributed among the space of all possible states. The equilibrium between states is thought to be ergodic, i.e., recovered very quickly when disturbed, so that every single cell experiences quickly a vast fraction of the possible expression patterns. This allows explaining how quickly an erased phenotypic part of the population can be recovered among cells as observed in different experiments. The permissive chromatin in stem cells imposes weak constraints on gene expression and makes impossible the definition of the pluripotent state at the single-cell level. Functional pluripotency emerges from the dynamic variability of the pluripotent states (MacArthur and Lemischka, 2013).

In the view of the authors, the equilibrium between states is still an intrinsic property of the cell through its genetic regulation network and stochastic microscopic events are seen as a noise over an intrinsically determined cell dynamics regulating the equilibrium between states. Although not always acknowledged by researchers using stochastic networks to describe the differentiation of ESC who would not entertain noise superimposed over a deterministic system, this hypothesis made to save genetic determinism is unfortunately introducing an incoherence in the model since it does not explain how external constraints have to be integrated in the picture. From their point of view, the primed cells in a specific state of the pluripotency pool of states are then determined to be able to respond to specific external signals. In this approach, the integration of those signals by the cell is not included in the model to explain how they can change the equilibrium of states and restrict the number of possible states. This remark also raises the question of the number of accessible states to a cell so that the process remains ergodic. Are the number of states really reduced or is it only their probability of appearance which is modified so that many states become unlikely to actually form? An elegant way to solve this issue is to assume that

cumulative constraints from the inner cell structure (chromatin and compartments) is always strong enough so that the space of genetic patterns is always small enough for the equilibrium between states to remain ergodic. Complementary, the regulation of the equilibrium should not only be intrinsic to the cells to accommodate the so-called external signals. The dynamics of the chromatin organization, which is the main regulator of the accessible states equilibrium, should be allowed to depend on larger scale constraints at tissue or organism level, i.e., not only to cell level constraints.

Consequently, this dynamic variability should be spatiotemporally regulated *in vivo* while these restrictions would be largely released *in vitro* and the intrinsic variability in the population should become apparent depending on culture conditions as seen experimentally (MacArthur and Lemischka, 2013). Thus, one can assume that these stochastic expression variations give rise to a state of dynamic equilibrium in which single cells transit randomly between distinct metastable states, and these highly variable and high-entropy expression patterns are likely to be responsible and necessary for the large number of possible developmental commitments (MacArthur and Lemischka, 2013). Transcriptional heterogeneity in stem cells is thus considered as an advantage for diversifying phenotypes in populations that require diverse potentialities (Torres-Padilla and Chambers, 2014). Then constraints, established when differentiation progresses, stabilize some phenotypes and decrease population entropy because chromatin acquires a less permissive state and expression patterns become more tightly constrained.

STOCHASTIC DIFFERENTIATION PROCESSES

Probabilistic Differentiation

Until recently, cellular differentiation was widely considered as deterministically induced by microenvironmental signals. Nevertheless, stochastic "choices" between several lineages are now demonstrated in many cases, without any role of inducing signals. For instance, in the HSC niche, the less differentiated cells are a mixed population of already primed cells (Yamamoto et al., 2013). This study shows that priming is not the result of niche factors guiding HSC differentiation. Instead, it is largely generated by intrinsic cellular factors (Yamamoto et al., 2013). Also, the fate of individual B cells following activation leading to their differentiation into diverse populations depends on the competition between intracellular processes varying stochastically (Duffy et al., 2012). These variations explain how lymphocytes produce the different terminal populations in reproducible proportions (Duffy et al., 2012). The probabilistic distribution does not result from specific instructions received by individual cells but from collective behavior in response to environmental cues, what could be generalized for differentiation processes from embryos to the hematopoietic system (Duffy et al., 2012).

Other developmental phenomena beginning with a probabilistic choice followed by microenvironmental stabilization are the formation of the retinal mosaic (Wernet et al., 2006), the acquisition of neuronal identity (Lesch et al., 2009) or the intestinal cell differentiation (Raj et al., 2010). Adult stem cells seem to differentiate because of stochastic intracellular phenomena rather than extracellular regulation. Environmental factors may only bias the internal stochastic processes (Wabik and Jones, 2015). This can be interpreted as a selective phenomenon among phenotypes that randomly appeared. Indeed, it is proposed that the varied expression of genes associated with each alternative fate in single cells can act as a substrate for selection by signals (Moris et al., 2016). An example is provided by Guye et al. (2016) who showed that cell fate decisions occur as a stochastic switch integrating cell-autonomous (here Gata6 levels that randomly fluctuate) and non-autonomous (cell types nearby) parameters. Only cells expressing by chance a given level of protein in the adequate environmental context start differentiating. This proposal has been largely developed in a theoretical model of cell differentiation based on Darwinian selection called ontophylogenesis (Kupiec, 1996, 1997, 2009) (see below).

Lineage tracing techniques also revealed that adult stem cells have a far larger differentiation potential when they are out of their initial tissue environment and transplanted in other locations (Goodell et al., 2015). These observations suggest that cell fate is not irreversibly predestined and that the microenvironment where these heterogeneous cells reside restrains lineage choice (Goodell et al., 2015). Differentiating cells are clearly not hierarchically organized and differentiation pathways are not unidirectional (Zipori, 2005; Knaan-Shanzer et al., 2008; Tetteh et al., 2015). An uncertainty exists in terms of origin and fate for each cell, together with variability in the timing and pathway toward differentiation for cells acquiring the same phenotypes. This suggests that stochasticity in differentiation processes finds its origin in the stochastic appearance of differentiation features.

Modulation of Stochastic Gene Expression During Differentiation

Recent works now give evidence that a degree of coordination of SGE is modulated during development by cell-cell communication. The canonical example is provided by the phenomena underlying fate decisions in the inner cell mass (ICM) of the blastocyst stage embryo into the epiblast (EPI) and primitive endoderm (PE) lineages [for review, see Simon et al. (2018)]. Single-cell transcriptomics analyses revealed that cells of the preimplantation mouse embryo uncover considerable transcriptional heterogeneity of ICM cells at the 32-cell stage (Guo et al., 2010; Ohnishi et al., 2014). Following this initial phase of stochastic expression of individual genes, FGF/ERK signaling interfaces with the Nanog and Gata6 TF to ensure cell differentiation in the ICM (Simon et al., 2018). Signaling and SGE are interconnected and influence one-another, and cell communication exploits and buffers noise to enable formation of the tissue with reproducible proportions of EPI and PE cell types (Simon et al., 2018). For a physical perspective, signals can be seen as constraints on the cell, a feedback of the tissue. In that

perspective, organogenesis and cell differentiation are two folds of a single multiscale process.

Other convincing examples of reduction of SGE and progressive coordination of gene expression profiles from cell-to-cell in tissues during development appeared in the last years (Featherstone et al., 2011, 2016; Ji et al., 2013; Pelaez et al., 2015). For instance, when cell differentiation is tracked *in vivo* during development of the pituitary tissue in transgenic rats in which luciferase was expressed from the prolactin (a pituitary hormone) promoter (Featherstone et al., 2011), it appeared that the first endocrine cells showed highly pulsatile expression which is then stabilized as tissue develops and cell number increases. The pulsative and heterogeneous luminescence was restored when cells were dissociated, showing that the stabilized expression pattern depends on the tissue architecture and cellular interactions. More precisely, direct cell contacts involving gap junctions ensure this coordination of prolactin gene expression because trypsin-mediated digestion of extracellular proteins or pharmacological inhibition of intercellular gap junctions reduced transcriptional coordination between cells (Featherstone et al., 2016). Thus, cell communication reduces SGE and phenotypic heterogeneity among differentiating cells.

Finally, only recent multiscale analyses of differentiation showed that it must be understood as a release of previous constraints that maintained SGE at relative lower levels in progenitors cells, followed by peak of variability that occurs before a reduction of SGE at lower levels than initially because new constraints are applied (Richard et al., 2016; Moussy et al., 2017). Indeed, while gene expression is initially widespread and highly variable when growth conditions ensuring the maintenance of the stem state are released, cells progressively transit toward more homogeneous, coordinated and restricted gene expression profiles (Efroni et al., 2008; Richard et al., 2016; Moussy et al., 2017). This cellular entropy appears to decrease from this transient unstable state with the highest SGE to the final stable differentiated state obtained under new environmental constraints (Richard et al., 2016). This trend is observed in single-cell transcriptomic data sets from various cell differentiation systems that have been recently studied in a more systematic way (Gao et al., 2020). Thus cell fate decision appears to be a process of spontaneous variation and selective stabilization in which each cell randomly explores many different possibilities by expressing a large variety of genes before finding a stable combination corresponding to the actual environment (Moussy et al., 2017). SGE and environmental constraints concomitantly act to produce stabilized phenotypes. This is in remarkable agreement with earlier theoretical predictions, especially in the ontophylogenesis theory. Nevertheless, this interplay has been diversely interpreted in different models of cell differentiation.

MODELS OF CELL DIFFERENTIATION BASED ON STOCHASTIC GENE EXPRESSION

SGE is now recognized as a driving force for diversifying phenotypes in undifferentiated cell populations, and as a responsible for lineage priming in stem cells. Alternative theories of development are needed to take into account this initial phenotypic plasticity, and its progressive reduction producing stable differentiated cell types. Especially, several models of cell differentiation have been formulated in which a peak in the gene expression variability is expected to occur.

Attractors and Self-Organization

The epigenetic landscape proposed by Conrad Waddington more than 60 years ago (Slack, 2002) has gained lots of attention in the last decade because it metaphorically describes the cell "canalization" toward differentiation. In this scheme, cells look like balls rolling in a landscape made of hills and valleys, until a final point of minimal altitude. Each valley corresponds to a precise lineage differentiation pathway, and gravity in the driving force leading to terminal differentiated types. But what is the biological reality corresponding to the gravitational force in Waddington's landscape? Why do cells roll toward points of minimal altitudes corresponding to stable cell types?

Several authors have used the dynamical systems theory to reintroduce the concept of attractor states first proposed by Stuart Kauffman decades ago (Kauffman, 1969). Sui Huang especially considers the stem cell state and differentiated states as attractor states created by the global architecture of the gene regulatory network (GRN) which orchestrates gene activities to produce the variety of stable cellular states (Huang, 2011). In this framework, state transitions, such as differentiation of a stem cell, arise when varying environmental conditions together with the action of SGE destabilize the current attractor state enough to generate the transition toward another attracting state in the GRN (Huang et al., 2009; Mojtahedi et al., 2016). Only regulatory signals through activation of alternative signal transduction pathways and/or highly stochastic fluctuations of gene expression levels can change enough the expression of a specific set of genes of the network and produce state transition. In this model, cell communications act to coordinate state transitions and to ensure the "correct" canalization of the various cell types in tissues but attractors are fundamentally self-organizing and self-stabilizing states leading to stable gene expression profiles. Moreover, these states are particularly robust to stochastic expression variations of individual genes. Random fluctuations in gene expression are only rarely able to trigger state transition.

Kunihiko Kaneko also employed the concept of attractor state and Waddington's epigenetic landscape to explain stable differentiation states, but insisted on the necessity to intimately couple both SGE and cell-cell interactions for modeling cell differentiation (Furusawa and Kaneko, 2012). Especially, cell interactions with other differentiating or differentiated cells both stabilize the attractor state and decrease SGE as cells become more differentiated, so that the range of states that the cell could visit decreased with the developmental course (Kaneko, 2011). This synchronization explains the influence of cell interactions in maintaining a low level of protein variations among differentiated cells, but this decrease is only a collateral effect of the attractor state that remains the driving force for differentiation.

In the Kaufmann and Huang's attractor model, gene expression profiles self-organize toward attractor states because

these states are "encoded" in the GRN. Instructive signals produced by extracellular entities are needed to provoke destabilization and state transition. Natural selection during evolution would have conserved only optimized developmental trajectories in the landscape, especially to prevent attracted cells from changing direction toward other attractors without destabilizing signals. But this concept is not a real alternative to the genetic program because one can consider that the attractor states virtually present in the GRN constitute a new form of finalist explanation in development similar to the genetic program. Indeed, one can hardly argue that these attractors are not conceived as final states encoded in the genome, even if they are acquired through a self-organizing and self-stabilizing process.

Moreover, in the Huang/Kauffman model, the valleys (the attractors) guide (canalize) cells, each valley corresponding to a phenotype. However, it is now shown that cells fluctuate between two phenotypes before stabilizing (Moussy et al., 2017). Thus they "jump" from one valley to another before stabilizing, a phenomenon that was previously predicted (Kupiec, 1996) and modeled (Laforge et al., 2005). This observation is in clear contradiction with the attractor model based on Waddington's canalization: valleys cannot be conceived as canalizing cells if cells can jump from one to the other. A true non-finalist model would need a selective extracellular agent acting among phenotypes stochastically generated by SGE. Kaneko introduces this external element but his model still integrates the attractor concept to explain why the ball rolls down along the Waddington's landscape.

Historically, the origin of the self-organization theory is the idea that a system can spontaneously organize which is meant by the prefix "self." Generally, for this reason, the proponents of self-organization have rejected or overtaken Darwinism which on the contrary implies that organization is driven by an external constraint, i.e., natural selection by the environment. Finally, self-organization has never questioned the model of genetic regulation from Jacob and Monod: the GRN relies on such regulations. It assumes that small fluctuations can produce switches between multistable states, but these states are attractor states determined by stable gene networks. In such an approach, the multiscale constraints have no power to change the number nor the properties of the attractor states but only help to reduce the formation of a fraction of those states. Consequently, these models do not predict the emergence of novelty inside the cell just like genetic determinism and do not provide a way of thinking of an intrinsic evolution of species due to stochastic processes inside cells (Kupiec, 2009). Small fluctuations can initiate the switching process but they are then followed by deterministic evolution that can be modeled by deterministic differential equations. Thus the ontology of self-organization is intrinsic stability as the default state. That is why self-organization theorists still consider that SGE is only noise.

Ontophylogenesis

The ontophylogenesis theory has been originally formulated in 1983 by Kupiec following the main idea that stochasticity of biochemical reactions could play a major role in the cellular differentiation process at a time when this process was mainly thought in terms of genetic program (Kupiec, 1983). It considers cell differentiation as a variation/selection process analogous to evolution (Kupiec, 1996, 1997, 2009). Variations are created by stochastic fluctuations of gene expression, and some patterns are selectively stabilized through interactions with the environment and neighboring cells so that they can be a positive force to create patterning in development see for instance Guye et al. (2016). Thus, it is clearly based on another basis compared to the models mentioned above, a chance-selection principle. This theory is suited to take into account initial stochastic and widespread gene expression and its progressive restriction during differentiation, but without any instructive or finalist aspect.

The Darwinian theory previously broke with the postulate of species stability by putting variation as the *primum movens* for the evolution of species. In the context of cell differentiation, ontophylogenesis makes a step further by also considering variability as the default state at the cell level. Stability is then an optical illusion resulting from a state of equilibrium at a given moment. What appears to be a stable state is the state of an intrinsic dynamical system stabilized transiently by the multiscale constraints that act on the cell fate from the inner cellular level to the external environment including of course the structures at all scales in the organism.

Ontophylogenesis has been extensively described with both theoretical and experimental arguments in articles and books by its author (Kupiec, 1983, 1996, 1997, 2009). In a few words, the stochastic nature of molecular interactions and gene expression makes the appearance of phenotypes, especially differentiation features, a probabilistic phenomenon. This phenotypic heterogeneity would constitute a "substrate" for a selection process by the cellular environment which would stabilize the adequate differentiation or metabolic features in a given environment and at a given time. This environment is constituted by many molecules (diffusible or not) and other cells. Differentiating cells able to interact and communicate with these surrounding molecules and cells are stabilized by cell signaling and post-translational modifications of chromatin proteins which have made the expression possible (Kupiec, 1997). This view is fully compatible with the idea that functional pluripotency is an average property of cells (MacArthur and Lemischka, 2013). As each individual stem cell harbors a variable propensity to be primed because of stochastic expression of key regulators, functional pluripotency can hardly be defined at the single-cell level. But this property ensures that a certain proportion of the cell population has the ability to start differentiating at each time. Again, it is the combination of cell-autonomous (variations in gene expression) and non-autonomous (environmental context) that make priming possible.

Development is conceived as a cellular Darwinian process where gene expression is initially unstable, differentiation features appear stochastically, and cells are selected and stabilized if they express by chance the adequate genes at the right place at the right time to interact with surrounding cells or molecules (Laforge et al., 2005). This model does not exclude a certain degree of intracellular regulation that would repress expression of other differentiation genes when some of them are stabilized

for instance to favor canalization (Sunadome et al., 2014) but this mechanism starts by and is not possible without initial stabilization of some genes by cellular interactions. It does not exclude neither that metabolic aspects can be crucial because cells must draw the energy needed to degrade and re-synthesize their components for phenotypic changes (Paldi, 2012). Thus the source of this energy could be a selection factor (a constraint in physicalist terms) which intervenes in the stabilization of certain phenotypes but also in the destabilization of others. Recent works showed that metabolic competition for a resource induces complementary phenotypes and population structure (Varahan et al., 2019).

Finally ontophylogenesis is both the contrary of an instructive mechanism and of a self-organizing phenomenon constrained by GRN. If not stabilized, cells continue to fluctuate or die. Here differentiating cells are also progressively canalized but this canalization is explained by the selective action of the cellular environment, and not by the architecture of the GRN. This dynamic interplay between cellular stochasticity and environmental constraints that act concomitantly and contribute equally to the appearance of stabilized phenotypes paves the way to a better understanding of the causality that explain cell trajectories in multicellular organisms.

Even if the GRN theory has embedded stochastic switches and contains the constraints of signals and cell-to-cell communication, these external constraints only modulate the GRN, they do not alter its structure that remains the essential causal driving force of the system. On the contrary, the main idea behind the ontophylogenesis model is that all organizations at all scales are important for the fate of a given biological system. In that sense, giving a causal role to a structure at a certain scale can be a good effective model but cannot be a fundamental approach of causality. Otherwise, we need to provide an explanation why that specific level is the causal level. More importantly, we need to have a conceptual idea of how that level gets organized. Without this conceptual explanation, we think this is a finalist explanation. The ontophylogenesis model does not have this issue since it takes into account the full diversity of structures that can constrain underlying dynamics. It is the equilibrium between the different levels of organization, which dynamically and reciprocally interact, that accounts for the structuration of the system. Ontophylogenesis is then an extension of Darwinism to ontogenesis.

In this framework, the usual divergence observed between *in vitro* and *in vivo* experimental results finds a natural explanation because all scales have causal importance for the system evolution and structuration. The comparison between experiment and theory becomes more difficult to be done and requires the simultaneous use of *in silico* simulations and *in vivo* measurements of the complex structures of the system under study. When the initial state structures are implemented in a simulation in which the theory is expressed as local relationships between components of the system, one can compare how the real and simulated systems evolve when submitted to the same perturbations.

One should notice that ontophylogenesis is consequently not a self-organization theory [see extensive review in Kupiec (2009)]. Indeed the self-organization theory provides organization from a bottom-up perspective where the concept of spontaneous emergence is the main source of generation of new features inside a system. Ontophylogenesis proposes a different perspective which is neither a bottom-up nor a top-down process but both at the same time. Any effective structure retroacts on the other levels of organization, "effective" meaning that the structure results from the constraints coming from both the developmental process and the environment in which the organism is living. For instance, cell-to-cell interactions modulate gene expression probability and alternative editing of RNA at the protein synthesis level while, at the same time, a living being is changing its environment (Levins and Lewontin, 1985). Ontophylogenesis is then indeed a hetero-organization theory providing a satisfactory integration of the holistic and local points of view which, instead of being contradictory, complete each other. This paradigm is illustrated in **Figure 1** [adapted from Kupiec (2009)] where we have added a bidirectional arrow between cell and (sub)organismal structure and the natural selection levels since organisms are also structuring their local environment.

Importantly, ontophylogenesis also provides a change of paradigm that reconnects biology to physics and chemistry first principles. Indeed ontophylogenesis is a theory of global order from local disorder on the opposite to all the other mentioned approaches that assume that global order at the scale of an organism derives directly from the local information located in the germinal cell. When epigenetic signaling is similar to the introduction of epicycles on the Aristotelian model of the universe pushing a step forward a level of explanation that escapes to the thought fundamental genetic level, ontophylogenesis offers an alternative giving a paradigm

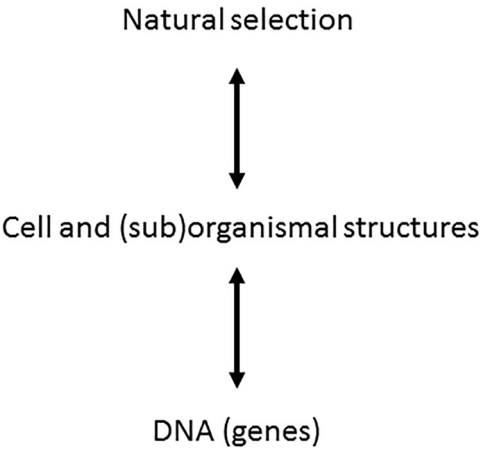

FIGURE 1 | Extension of evolutionary synthesis to ontogenesis. In the frame of the ontophylogenesis model, the relationship between genes and phenotypes is no longer unidirectional but bidirectional. The cell and (sub)organismal structures, subjected to natural selection, sort the molecular interactions, which means that natural selection is a cause of ontogenesis. The genes provide the proteins while the phenotypic structure sorts their interactions from among the possible combinations. The two processes of natural selection and ontogenesis are but one single process of ontophylogenesis [adapted from Kupiec (2009)].

where biological phylogenesis and evolution can be thought coherently. Ontophylogenesis exhibits the importance of the dynamics while biology is usually thought through approaches giving a primordial role to effective structures and functions. These assertions can be enlightened by adopting a physicalist point of view that we propose to now consider.

PHYSICAL CONSIDERATIONS TO CELLULAR PROCESSES AND THEIR CONSEQUENCES ON HOW ORGANIZATION OF BIOLOGICAL SYSTEMS EMERGES

Chaotic Gene Networks Are Not Necessary to Provide Emergence

Understanding from the first principles what is the causality that rules the fate of cells inside an organism is one of the key problems of theoretical biology. The question is of course very difficult since it involves the organization of a complex system with a large variety of components and organized structures that can retroact on the future evolution of a given cell depending on its location in the structured system. So the question also concerns cell collective structure formations and their dynamics.

It is a scientific fact that biological systems at cell scale cannot be modeled using a description in terms of their fundamental constituents at the particle level invoking our current knowledge of the fundamental interactions. Already describing a bound state such as an atom out of its quark and electron content is beyond the scope of the present possibilities. Fortunately, one can build a so-called effective model of the hydrogen atom based on effective components, a proton and an electron to describe its atomic properties such as its emission and absorption spectra. In that specific case, the proton, known to be built from 3 u quarks of charge 1/3 each, needs to have a charge $+1$ to have properties in line with our fundamental understanding of the structure of matter components. Similarly, building effective models of living systems requires the use of fundamental concepts that are coherent with our current knowledge of physics and chemistry laws at low scales, especially if the objective is to provide a broadest and deepest understanding of the biological system organization and dynamics.

Unfortunately, since tens of years now, some of the key concepts used to describe biological systems seem to break this objective. Especially, the concept of information, which is driving theoretical approaches since the beginning of early works of molecular biology, really impacts the way the dynamics of the biological systems is considered. This information concept is broadly understood by biologists as a deterministic path of cell differentiation along the information stored in the DNA. So in that view, the future of the cell inside a body is determined by its gene expression patterns and the dynamics is not so important since the final state is known. This information theory has then a large impact on how biology is overweighting the role of structures in living systems while considering dynamical processes of lower importance since the final state is driven by the information. For instance, looking at embryo development, embryology defines different stages of development characterized by a biological structure while physics would observe the distribution of the times at which the related structure appeared in each embryo by starting a clock at initial time $t = 0$ for each embryo. From the structural point of view, the fact that most embryos develop the related structures provides the impression of a uniform process while the observation of the times at which each structure appears for each embryo is exhibiting the variability of the process producing the structure. One should notice that the information discussed by most biologists is very different from the information theory developed in statistical physics based on the concept of entropy where the level of information in a system is related to the number of its possible microscopic configurations that provides a similar macroscopic state. It is often a subject of misunderstanding between physicists and biologists. In physics, entropy is driven by a very strong principle, the second law of thermodynamics, that states that the entropy of an isolated system can only increase, giving rise to a more and more disordered system. Consequently the increase of order inside an organism requires this system to continuously exchange matter and energy with its environment. Theoretically speaking, it requires to develop a theory of thermodynamics far from the equilibrium. In that perspective, an organism is then a dissipative structure (Prigogine, 1967), i.e., a structure exchanging energy and matter with its external world.

On the opposite side, information in biology is usually related to the deterministic way that will transform a genetic information into a specific structure or function in the organism. The biological paradigm of information is also constraining how the discovery of SGE inside cells is interpreted: the term of "noise" used extensively in the literature generally expresses the idea that this stochasticity occurs on top of a well-defined and information-based behaviors of cells. This denomination forbids thinking stochasticity as a potential driving force of change inside biological systems and as the possible motor of the divergence of living systems as stated by the Darwinian theory of Evolution. In that context, ontophylogenesis suggests that the control of cell stochasticity requires much more than genetic networks (Elowitz et al., 2002), since supracellular structures retroact on the inner cell dynamics.

The observed general difficulty to associate a single gene to a single phenotypic property or biological function and the need of explaining many features with a limited number of genes inside the cells drove genetics into the paradigm of inter-relations between genes to provide a vast number of phenotypic states as a solution to this problem. This idea leads to the definition of gene networks dynamically structured to provide different differentiated states to cells (Emmert-Streib et al., 2014). One should notice that this point of view is just a generalization of the one gene – one molecule – one function paradigm in which the larger complexity is only coming from the combinatorics of several genes. The difficulty of that vision is that a gene network is not well-suited to describe how the cell environment can influence its fate as shown by experimental data. So, one usually envisages the change of a cell state as an intrinsic internal instability (chaotic behavior of a non-linear differential equation

system) or as an external event which is very difficult to plug on the gene network from a theoretical point of view. Indeed, a gene network expresses the relationships between the expression patterns of the various genes inside the cells. It describes how an either positive or negative retroaction of each gene expression can modulate the expression of the other genes.

This conception raises the question of how the external constraints and external conditions such as temperature, pressure, chemical concentrations in the tissue can be included in the models. This can be done to a certain level by making the parameters of the gene network to depend on those conditions so that the full potential of the chaotic behavior inside the cell can allow the gene network to evolve with the changes of external conditions. From a theoretical point of view, this requires to explicitly change the number of parameters needed to implement such a dependence if the number of external parameters changes (for instance a new molecular gradient or new possible inter-cells interactions appears in the vicinity of the cell) and is then an ill-defined procedure. Moreover, keeping the gene network behavior to be the underlying explanation for cell state switches requires its associated set of differential equations to have non-linearities. Indeed, they are the necessary conditions to have a possible chaotic switch between different final states. This is in our opinion putting a too strong constraint on the possible models of cell differentiation (Villani et al., 2011). In particular, this approach ignores the possible role of the external environment changes on the cell fate and forgets about very important scientific facts such as the stochasticity coming from the local molecular behavior that can be moderated by the organization of the cell, especially of its chromatin. While being partially captured by stochastic gene networks, the intrinsic stochasticity of molecular processes allowing the cell to explore different configurations requires to go beyond genetic networks that only provide a finite number of possible states, while biological systems need to escape this limitation to cope with the full diversity of external conditions.

On the other hand, there is a real physical motivation to assume that such an intrinsic stochasticity inside cells can come either from a fundamental reason related to quantum mechanics when atom/molecular scale processes occur (proton/ion pumps, chromatin binding. . .) or from temperature effects which imply a distribution of configuration for the various components in the cell. Including such considerations in the model building allows us to offer a much simpler situation where differential equation systems do not need to be non-linear as soon as this intrinsic stochasticity is able to trigger switches between a large variability of cellular states with a related diversity of patterns of gene expression. From a statistical physics point of view, this stochasticity accounts for the fluctuations allowed by the limited number of components involved at the molecular level in cell mechanisms.

Such an approach can circumvent the need of chaos to explain statistical variation at the gene expression levels which is usually the expected paradigm from usual differential equation formalism (ODE) used to represent the underlying gene network. In chaos theory, the observed apparent stochasticity occurs from the possible switch of cell trajectory in the cell state space from one strange attractor to another (a point in the configuration space around which the system trajectory is attracted). The number of attractors and their location in the configuration space result from the inner properties of the set of differential equations used to describe the system evolution. This is sketched in **Figure 2** where x(t), y(t), and z(t) could, for instance, represent the level of expression of 3 genes giving rise to a variety of observed changes in the cellular state {x,y,z} with time.

The aggregation of the two basic ideas that there is a well physically motivated intrinsic stochasticity inside the cell and that a theoretical approach is needed to include external constraints on the dynamics of biological systems allow to propose a very interesting framework to address biological organization, evolution and emergence of new biological functions in the same theoretical paradigm. Such an approach was first introduced in the theory of ontophylogenesis (Kupiec, 1996, 1997, 2009)

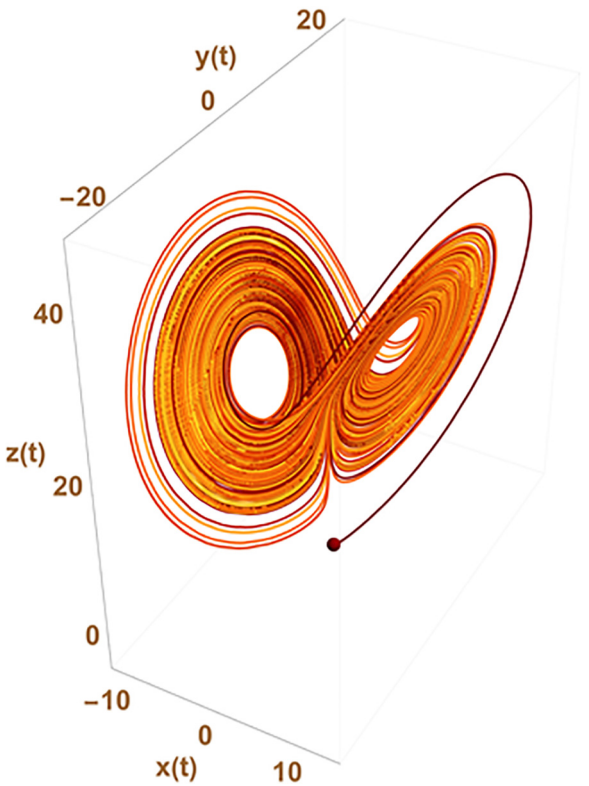

FIGURE 2 | Chaotic trajectories from the historical Lorentz problem. The figure depicts the possibility that states {x,y,z} of the system change apparently randomly from one to the other two well defined regions of the parameter space depicting what has given its name to the "butterfly effect" in chaos theory. The trajectory of the system in the state space switches randomly from trajectories cycling around one of the two strange attractors existing in the ODE system:

$$\begin{cases} \frac{dx}{dt} = -3(x(t) - y(t)), \\ \frac{dy}{dt} = -x(t) \times z(t) + 26.5x(t) - y(t), \\ \frac{dz}{dt} = x(t) \times y(t) - z(t), \end{cases}$$

with initial conditions x(0) = y(0) = 0 and z(0) = 1.

in which dynamics takes a major role while usual genetic information still finds a natural place becoming only an important constraint on the dynamics occurring inside the cell. This idea can be extended to also include other structures inside the cells (compartments, large scale molecular machinery. . .) and in the organism (organs, vascularization. . .) on the same basis (Laforge, 2009). From the physical point of view, this approach allows us to overtake the standard statistical approaches where the key concept used as a proxy for the information is entropy. This latter physical quantity provides in fact a change of the information concept as usually understood and discussed by biologists and of the associated determinism that is usually invoked to speak about the cell fate in terms of gene expression patterns. Indeed, as discussed a bit earlier, entropy, as a statistical physics concept developed to explain the time's arrow in statistical systems, expresses the level of order inside a physical system (Blum, 1968). Nevertheless, one should notice that entropy does not provide an adequate way to describe the evolution of organized dissipative structures. Indeed, from a statistical physics point of view, the system is always a set of microscopic components while its mesoscopic or macroscopic description requires the introduction of a mathematical object able to account for this scale of organization.

Emergence as the Result of a Competition Between Internal Dynamics and External Dynamical Constraints

The second principle of thermodynamics states that a higher order in a structure cannot be obtained without an exchange of entropy with the outside of the system. This means that building models in which a switch in the organization could only come from an internal reorganization is difficult to support. It directly means that using a system of ODE, the only possible hope is to describe a switch between two states with the same level of order at most. This approach is then unsuitable to describe growing structures and organization inside a cell unless something more is added to take into account these entropy exchanges. Importantly, this argument states that some constraints on the organization process of a biological system are arising from physical limitations to energy and matter fluxes exchanged between the cell and its local environment. This argument holds at any scale and also gives inputs to understand how inner cell structures evolve such as chromatin, cell compartments or even proteins during the macromolecular processes involved in gene expression. In these various cases, one envisages different structures with different outsides which show that there is a continuity of scales that need to be considered to understand how biological systems get structured macroscopically. Indeed speaking for instance about chromatin requires to consider cell compartmentation and nuclear membrane in eukaryotic cells as structures that act on the possible fate of chromatin.

The second principle of thermodynamics is coherent with the ontophylogenesis framework explanation of stochasticity control inside a cell as a result of changes in its local environment, including cellular exchanges with its neighbors. It also puts an important light on how emergence of novelty can occur in biology giving a very strong input to the Darwinian theory (Laforge, 2009). Ontophylogenesis, by pushing Darwinian logic inside the cell, i.e., by introducing a random turn-over of different primed states resulting from intrinsic stochasticity of physico-chemical reactions at low scale, provides a key reason to explain the divergence of species seen as a successive collections of individuals intrinsically subject to variations possibly aggregated in seminal cells affecting their whole descendance (Kupiec, 2009). Interestingly, this framework allows external causes to imply a change of an individual's organization as a special case of constraints on the organism dynamics during its development.

An idea of how emergence occurs in biological systems (and in complex systems in general) can be depicted using the physical concept of internal energy of a system that describes the kinetic energies of its components and the potential energy arising from their geometric organization. In the case of the chromatin system, this corresponds mostly to all the binding energies of molecular interactions that organize its geometrical configuration and to the kinetic energies of the related molecules that have an average speed determined by the temperature inside the cell. **Figure 3A** illustrates a very simple case in which the internal energy of the system is represented as a function of two conformation variables (x and y).

When the system is isolated, i.e., is not experiencing any external constraints, general physical variational principles state that the system evolves toward the configuration giving a minimum of its internal energy. The configuration (x = 0, y = 0) occupied by the blue point is then depicting the equilibrium position of the system in the energetic landscape. If the system is now put in an environment that is acting on it through external forces, the situation requires to add an external potential energy like the one shown on **Figure 3B**. In that case, the physics drives the systems so that it minimizes the sum of the two potential energies. Then the final state is not given by the minimal internal energy configuration of the system only but is obtained as a mixed result between the internal and external potentials configuration. Translated to a biological case, if one looks at the differentiation of cells in different conditions of pressure, consequences from those conditions are expected on the cell differentiation as it has been demonstrated *in vivo* (Farge, 2003). Analogously, an impact of external local forces acting on cell membrane is expected on the cell division axis as it has been shown *in vitro* (Thery and Bornens, 2006). **Figure 3** illustrates that different final states can be obtained in the evolution of the system if its initial state starts from different positions in the configuration space since there are different minima in the energetic landscape build as the sum of the two potentials. Then the final state depends from the starting point of the system in the energetic landscape. If we now consider that there is a dynamical process ongoing with specific times of evolution for both internal and external potentials, the end point of the evolution depends on the competition between the dynamics of internal and external constraints which is the precise case in which **Figures 4A,B** have been obtained (the blue point is the position of the system in its state space (x,y) after the same amount of time in the two different simulations) from different

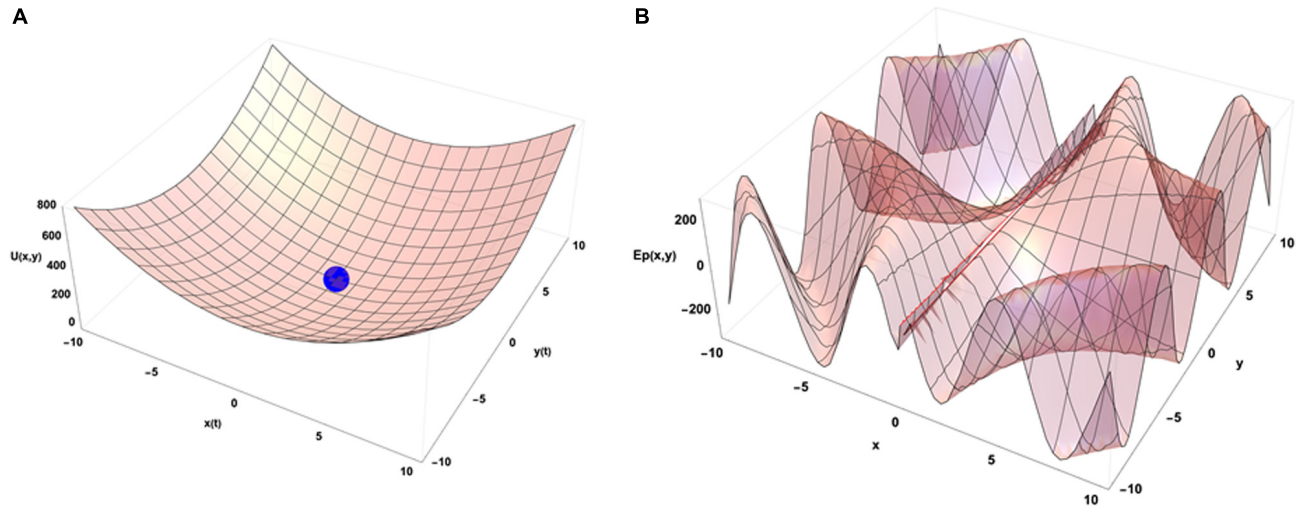

FIGURE 3 | Energetic point of view of the dynamical organization of a system with 2 degrees of freedom in interaction with its environment. **(A)** Internal energy landscape of a system as a function of two variables measuring its geometrical organization. The final configuration of the system is (0,0) without external constraints. **(B)** A potential energy landscape describing the interactions of the system with its external environment as a function of its location in the (x,y) configuration space (t = 0). The functional form used to draw that illustrative example is U(x,y) = 4 (x2+y2) for **(A)** and Ep(x,y) = 300 × Sin(0.1 × y + 3 t) × Cos[1/(x + 0.1) + 3 t] for **(B)**. In the latter case, at dependence of the external potential is used to illustrate the impact of external potential energy dynamics on the final organization of the system (see text).

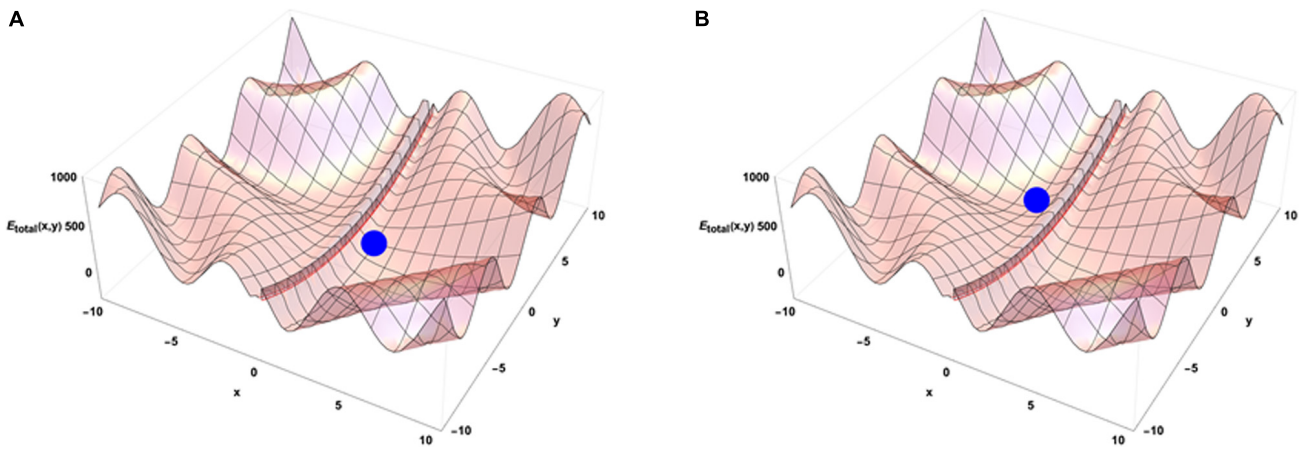

FIGURE 4 | Emergence from the competition of dynamics of external and internal constraints. The subfigures **(A,B)** show that different minima of energy can be reached by the system in its time evolution. The point reached at the same time step of the simulation depends on the position of the system in its initial state and on the evolution dynamics of the external potential.

initial states under the same dynamics of the external potential. The figures show the sum of the two potential energies at the final time of the simulation. At each step in time, the simulation moves the state of the system toward the closest local minimum around the current position of the system with a speed that is representative to the characteristic time of evolution of the inner dynamics of the system. To illustrate the situation, one can argue that the typical time of evolution of the internal energy is related to the inertia of the components of the system and to the average duration of bounded states inside the system. The external dynamics is characterized by a typical time of external potential energy variations. The meaning of the results of this simulation is that, when one looks at an organized system (one only accesses to the final state of the evolution!), there is no hope to understand its structure from the properties of its local components since we have lost the history of the changes that occurred in its vicinity while it was being structured. The real situation is even a bit more complex if one considers that the matter exchanged between the system and its external world changes deeply the inner potential. This can be seen as an extra internal dynamic of the inner potential allowing us to think about this issue in the heuristic approach proposed here.

This simple example can be extended to address the case for an open system where energy and matter is being exchanged with its external environment where the general situation explained above remains. Rephrased in the language of statistical physics, it seems that one cannot just address the organization of a complex system from its inner components properties only. Indeed, dissipative structures put emergent constraints of the dynamics of the microscopic components. One should also notice that from a theoretical point of view, one needs to use a mathematical structure to describe and follow these mesoscopic or macroscopic organizations in the system. Indeed, it is not enough to just define the evolution of the system of interest by accounting for the amount of entropy it exchanged with the external world. One needs to indicate how the new possible ordering power given by the decrease of the entropy of the system due to this exchange with its external world is used inside the system. This means that the traditional statistical physics approach is ill-defined to address the question of complex system organization with spatial and temporal heterogeneity. Such systems require much more than the phase transition paradigm to be understood. It also means that numerical simulations are, in the context of our current knowledge, the only tool to face these kinds of problems but they also require the definition of what one calls a mesoscopic scale organized structure. Indeed, from the point of view of the microscopic components of the system, any organized structure is just a simple configuration of its parts. The fact that such a structure holds in time is not a property of the local components but comes from the interactions at different scales between mesoscopic structures (Marsland and England, 2018).

One important remark along that approach is that it gives a natural explanation to the apparition of levels in biological systems. Indeed, those levels are effective structures (by opposition to fundamental or ontologic) formed from very long lifetime and precise characteristic size dynamic structures. Ontophylogenesis hence provides a reasonable scientific framework to think about emergent properties in biological systems.

The ontophylogenesis paradigm has been tested using numerical simulations (Laforge et al., 2005) to demonstrate its ability to provide good properties of organization during a developmental process. In that study, it was demonstrated using a very simple model with 2 cell types and a mechanism of cell-to-cell signaling based on physics principles that the ontophylogenesis framework can provide reproducible emergent properties relevant to a developmental process. In particular, *in silico* simulations exhibited the formation of a bilayer of cells whose structure completely results from an equilibrium between the numerical values of the parameters of the simulation and from the choice of the simulation rules conditioning how the local environment of each cell stabilizes its chance of changing its phenotype. Importantly, the simulations showed that the emergent bilayer can reach a stabilized form corresponding to an end of the developmental process without an "end of growth" gene. In that simulation, the emergence of the bilayer is associated with a reduction of the variance of expression patterns of every single cell inside the bilayer. Even more interesting, this work presented further results suggesting that cancer could be a problem of tissue disorganization rather than an initial mutation inside a cancer cell. This is supported by theoretical considerations and experimental evidence over cancer initiation at the tissue level (Capp, 2005, 2017). In a subsequent work (Glisse, 2009), it was demonstrated that a sizable amount of stochasticity in gene expression can provide a quicker and more effective way of building the bilayer structure than a very precisely driven expression. Furthermore, this work also demonstrated that the introduction of apoptosis mechanism can provide a more robust formation of the bilayer structure giving a Darwinian explanation to apoptosis inside multicellular organisms.

The present discussion allows a revisit of the Waddington landscape paradigm often invoked in cell state evolution in the context of noise as discussed previously. In the ontophylogenesis framework, that landscape is dynamic because of the evolution of the external constraints applying to the dynamical system under study. In more biological words, this landscape is dynamic because of the dynamics of the selection conditions that occur in the vicinity of the biological system. On the other hand, the system owns an intrinsic capability to explore the landscape which is a consequence of the stochastic nature of its dynamics. This allows the system to change its route in the landscape with low probability of switching from one valley to the other if the valleys are deep. Cell reprogramming to adapt to new conditions is then a natural process inside the ontophylogenesis framework. The organization of biological systems is then a hetero-organization where the intrinsic stochastic dynamics of the fundamental ingredients of the system and the retroaction of organized structures (seen as dynamical states of equilibrium with a certain time of life) on the local dynamics of the parts have a shared causality in the determination of the fate of the system. In some sense, this framework reconciles both Darwinism (intrinsic divergence) and a form of neo-Lamarckism (action of the environment to guide the evolution of a single individual development but without finalism) as two faces of the same underlying coherent framework to think biology.

CONCLUSION

Both biological and physical considerations highlight the need to reconsider developmental processes in the light of the inherent stochasticity of cellular behavior. Considering its origin at the gene expression level, whose stochasticity appears to be modulated by environmental constraints during development, leads to put it forward when wanting to build a coherent framework that would no more try to accommodate classical deterministic or semi-deterministic views with this biological randomness. Being supported by its coherence with long-accepted conceptions in physics, a theory such as ontophylogenesis has the necessary requirements to include both the demonstrated probabilistic behavior at the cellular level, and the stabilizing environmental influence on the ongoing processes by coupling their concomitant actions in the emergence of structured tissues. Adopting such an alternative viewpoint would render developmental biology closer to modern physics, and make biologists enter a new era full of innovative research

avenues. It would in particular give a very strong case to further develop simulation approaches of biological systems based on multiagent formalism where interactions at low scale can be input in the model while heterogeneous mesoscopic or macroscopic structures retroaction on the evolution of the simulated biological system are dynamically provided by the simulated geometrical organization of the microscopic components of the system. As initial structures result from the previous evolution of the system under a dynamics of external constraints that cannot be known from the observation of the structures themselves, initial state of the simulation should be set from precisely measured structures observed in the real systems that the simulation is to be compared with.

Importantly, ontophylogenesis allows cell differentiation, development and evolution to be thought in the same framework as the cumulative evolution of germinal cells inside successive individuals. Its proposal to see deterministic processes as stochastic processes with a probability of one, as the results of the constraints on the system dynamics, accommodates the known mainly deterministic properties of biological systems. Complementary, its ability to predict the restriction of stochasticity in the process of cell differentiation (Laforge et al., 2005) which has been experimentally observed afterward (Moussy et al., 2017) validates its nature of scientific theory.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

J-PC and BL formulated the hypotheses and wrote the manuscript.

ACKNOWLEDGMENTS

We are very grateful to Thomas Heams, Jean-Jacques Kupiec, and Andràs Paldi for critical reading of the manuscript and helpful comments.

REFERENCES

Abranches, E., Guedes, A. M., Moravec, M., Maamar, H., Svoboda, P., Raj, A., et al. (2014). Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. *Development* 141, 2770–2779. doi: 10.1242/dev.108910

Blum, H. F. (1968). *Time's Arrow and Evolution*. Princeton, NJ: Princeton University Press.

Capp, J. P. (2005). Stochastic gene expression, disruption of tissue averaging effects and cancer as a disease of development. *Bioessays* 27, 1277–1285. doi: 10.1002/bies.20326

Capp, J. P. (2017). Tissue disruption increases stochastic gene expression thus producing tumors: cancer initiation without driver mutation. *Int. J. Cancer* 140, 2408–2413. doi: 10.1002/ijc.30596

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., et al. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature* 450, 1230–1234. doi: 10.1038/nature06403

Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* 453, 544–547. doi: 10.1038/nature06965

Duffy, K. R., Wellard, C. J., Markham, J. F., Zhou, J. H., Holmberg, R., Hawkins, E. D., et al. (2012). Activation-induced B cell fates are selected by intracellular stochastic competition. *Science* 335, 338–341. doi: 10.1126/science.1213230

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D. J., Dash, C., et al. (2008). Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* 2, 437–447.

Efroni, S., Melcer, S., Nissim-Rafinia, M., and Meshorer, E. (2009). Stem cells do play with dice: a statistical physics view of transcription. *Cell Cycle* 8, 43–48. doi: 10.4161/cc.8.1.7216

Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science* 297, 1183–1186. doi: 10.1126/science.1070919

Emmert-Streib, F., Dehmer, M., and Haibe-Kains, B. (2014). Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Front. Cell Dev. Biol.* 2:38. doi: 10.3389/fcell.2014.00038

Fang, M., Xie, H., Dougan, S. K., Ploegh, H., and Van Oudenaarden, A. (2013). Stochastic cytokine expression induces mixed T helper cell States. *PLoS Biol.* 11:e1001618. doi: 10.1371/journal.pbio.1001618

Farge, E. (2003). Mechanical induction of Twist in the Drosophila foregut/stomodeal primordium. *Curr Biol* 13, 1365–1377. doi: 10.1016/s0960-9822(03)00576-1

Featherstone, K., Harper, C. V., Mcnamara, A., Semprini, S., Spiller, D. G., Mcneilly, J., et al. (2011). Pulsatile patterns of pituitary hormone gene expression change during development. *J. Cell Sci.* 124, 3484–3491. doi: 10.1242/jcs.088500

Featherstone, K., Hey, K., Momiji, H., Mcnamara, A. V., Patist, A. L., Woodburn, J., et al. (2016). Spatially coordinated dynamic gene transcription in living pituitary tissue. *eLife* 5:e08494.

Furusawa, C., and Kaneko, K. (2012). A dynamical-systems view of stem cell biology. *Science* 338, 215–217. doi: 10.1126/science.1224311

Gao, N. P., Gandrillon, O., Páldi, A., Herbach, U., and Gunawan, R. (2020). Universality of cell differentiation trajectories revealed by a reconstruction of transcriptional uncertainty landscapes from single-cell transcriptomic data. *bioRxiv* 2020.2004.2023.056069.

Giladi, A., Paul, F., Herzog, Y., Lubling, Y., Weiner, A., Yofe, I., et al. (2018). Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. *Nat. Cell Biol.* 20, 836–846. doi: 10.1038/s41556-018-0121-4

Glisse, J. (2009). *Simulation d'un Modèle Darwinien de Différenciation Cellulaire*. PhD. Paris: University Pierre et Marie Curie.

Goodell, M. A., Nguyen, H., and Shroyer, N. (2015). Somatic stem cell heterogeneity: diversity in the blood, skin and intestinal stem cell compartments. *Nat. Rev. Mol. Cell Biol.* 16, 299–309. doi: 10.1038/nrm3980

Guo, G., Huss, M., Tong, G. Q., Wang, C., Li Sun, L., Clarke, N. D., et al. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* 18, 675–685. doi: 10.1016/j.devcel.2010.02.012

Guye, P., Ebrahimkhani, M. R., Kipniss, N., Velazquez, J. J., Schoenfeld, E., Kiani, S., et al. (2016). Genetically engineering self-organization of human pluripotent stem cells into a liver bud-like tissue using Gata6. *Nat. Commun.* 7:10243.

Hayashi, K., De Sousa Lopes, S. M. C., Tang, F., Lao, K., and Surani, M. A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* 3, 391–401. doi: 10.1016/j.stem.2008.07.027

Hough, S. R., Thornton, M., Mason, E., Mar, J. C., Wells, C. A., and Pera, M. F. (2014). Single-cell gene expression profiles define self-renewing, pluripotent, and lineage primed states of human pluripotent stem cells. *Stem Cell Rep.* 2, 881–895. doi: 10.1016/j.stemcr.2014.04.014

Huang, S. (2011). Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 2247–2259. doi: 10.1098/rstb.2011.0008

Huang, S., Ernberg, I., and Kauffman, S. (2009). Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin. Cell Dev. Biol.* 20, 869–876. doi: 10.1016/j.semcdb.2009.07.003

Hume, D. A. (2000). Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. *Blood* 96, 2323–2328. doi: 10.1182/blood.v96.7.2323.h8002323_2323_2328

Ji, N., Middelkoop, T. C., Mentink, R. A., Betist, M. C., Tonegawa, S., Mooijman, D., et al. (2013). Feedback control of gene expression variability in the *Caenorhabditis elegans* Wnt pathway. *Cell* 155, 869–880. doi: 10.1016/j.cell.2013.09.060

Kalmar, T., Lim, C., Hayward, P., Munoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., et al. (2009). Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* 7:e1000149. doi: 10.1371/journal.pbio.1000149

Kaneko, K. (2011). Characterization of stem cells and cancer cells on the basis of gene expression profile stability, plasticity, and robustness: dynamical systems theory of gene expressions under cell-cell interaction explains mutational robustness of differentiated cells and suggests how cancer cells emerge. *Bioessays* 33, 403–413. doi: 10.1002/bies.201000153

Kauffman, S. (1969). Homeostasis and differentiation in random genetic control networks. *Nature* 224, 177–178. doi: 10.1038/224177a0

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., et al. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201. doi: 10.1016/j.cell.2015.04.044

Knaan-Shanzer, S., Van Der Velde-Van Dijke, I., Van De Watering, M. J., De Leeuw, P. J., Valerio, D., Van Bekkum, D. W., et al. (2008). Phenotypic and functional reversal within the early human hematopoietic compartment. *Stem Cells* 26, 3210–3217. doi: 10.1634/stemcells.2007-0117

Kupiec, J. J. (1983). A probabilist theory for cell differentiation, embryonic mortality and DNA c-value paradox. *Specul. Sci. Technol.* 6, 471–478.

Kupiec, J. J. (1996). A chance-selection model for cell differentiation. *Cell Death Differ.* 3, 385–390.

Kupiec, J. J. (1997). A Darwinian theory for the origin of cellular differentiation. *Mol. Gen. Genet.* 255, 201–208. doi: 10.1007/s004380050490

Kupiec, J. J. (2009). *The Origin of Individuals*. Singapore: World Scientific Publishing Company.

Laforge, B. (2009). Emergent properties in biological systems as a result of competition between internal and external dynamics. *J. Biol. Phys. Chem.* 9, 5–9. doi: 10.4024/13la09a.jbpc.09.01

Laforge, B., Guez, D., Martinez, M., and Kupiec, J. J. (2005). Modeling embryogenesis and cancer: an approach based on an equilibrium between the autostabilization of stochastic gene expression and the interdependence of cells for proliferation. *Prog. Biophys. Mol. Biol.* 89, 93–120. doi: 10.1016/j.pbiomolbio.2004.11.004

Lesch, B. J., Gehrke, A. R., Bulyk, M. L., and Bargmann, C. I. (2009). Transcriptional regulation and stabilization of left-right neuronal identity in *C. elegans*. *Genes Dev.* 23, 345–358. doi: 10.1101/gad.1763509

Levins, R., and Lewontin, R. (1985). *The Dialectic Biologist*. Boston, MA: Harvard University Press.

Li, M., and Izpisua Belmonte, J. C. (2018). Deconstructing the pluripotency gene regulatory network. *Nat. Cell Biol.* 20, 382–392. doi: 10.1038/s41556-018-0067-6

MacArthur, B. D., and Lemischka, I. R. (2013). Statistical mechanics of pluripotency. *Cell* 154, 484–489. doi: 10.1016/j.cell.2013.07.024

MacArthur, B. D., Sevilla, A., Lenz, M., Muller, F. J., Schuldt, B. M., Schuppert, A. A., et al. (2012). Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity. *Nat. Cell Biol.* 14, 1139–1147. doi: 10.1038/ncb2603

Marsland, R., and England, J. (2018). Limits of predictions in thermodynamic systems: a review. *Rep. Prog. Phys.* 81:016601. doi: 10.1088/1361-6633/aa9101

Meshorer, E., Yellajoshula, D., George, E., Scambler, P. J., Brown, D. T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* 10, 105–116. doi: 10.1016/j.devcel.2005.10.017

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A. J., Tanaka, Y., Wilkinson, A. C., et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* 33, 269–276. doi: 10.1038/nbt.3154

Mojtahedi, M., Skupin, A., Zhou, J., Castano, I. G., Leong-Quong, R. Y., Chang, H., et al. (2016). Cell fate decision as high-dimensional critical state transition. *PLoS Biol.* 14:e2000640. doi: 10.1371/journal.pbio.2000640

Moris, N., Pina, C., and Arias, A. M. (2016). Transition states and cell fate decisions in epigenetic landscapes. *Nat. Rev. Genet.* 17, 693–703. doi: 10.1038/nrg.2016.98

Moussy, A., Cosette, J., Parmentier, R., Da Silva, C., Corre, G., Richard, A., et al. (2017). Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment. *PLoS Biol.* 15:e2001867. doi: 10.1371/journal.pbio.2001867

Muers, M. (2012). Development: many roads lead to commitment. *Nat. Rev. Genet.* 13:224. doi: 10.1038/nrg3202

Ohnishi, Y., Huber, W., Tsumura, A., Kang, M., Xenopoulos, P., Kurimoto, K., et al. (2014). Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat. Cell Biol.* 16, 27–37. doi: 10.1038/ncb2881

Olsson, A., Venkatasubramanian, M., Chaudhri, V. K., Aronow, B. J., Salomonis, N., Singh, H., et al. (2016). Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* 537, 698–702. doi: 10.1038/nature19348

Orkin, S. H., and Hochedlinger, K. (2011). Chromatin connections to pluripotency and cellular reprogramming. *Cell* 145, 835–850. doi: 10.1016/j.cell.2011.05.019

Paldi, A. (2003). Stochastic gene expression during cell differentiation: order from disorder? *Cell Mol. Life Sci.* 60, 1775–1778. doi: 10.1007/s00018-003-23147-z

Paldi, A. (2012). What makes the cell differentiate? *Prog. Biophys. Mol. Biol.* 110, 41–43. doi: 10.1016/j.pbiomolbio.2012.04.003

Pelaez, N., Gavalda-Miralles, A., Wang, B., Navarro, H. T., Gudjonson, H., Rebay, I., et al. (2015). Dynamics and heterogeneity of a fate determinant during transition towards cell differentiation. *eLife* 4:e08924.

Pina, C., Fugazza, C., Tipping, A. J., Brown, J., Soneji, S., Teles, J., et al. (2012). Inferring rules of lineage commitment in haematopoiesis. *Nat. Cell Biol.* 14, 287–294. doi: 10.1038/ncb2442

Prigogine, I. (1967). "Structure, dissipation and life," in *Proceedings of the 1st International Conference on Physique Théorique et Biologie*, ed. Marois Amsterdam.

Radzisheuskaya, A., Chia Gle, B., Dos Santos, R. L., Theunissen, T. W., Castro, L. F., Nichols, J., et al. (2013). A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nat. Cell Biol.* 15, 579–590. doi: 10.1038/ncb2742

Raj, A., Rifkin, S. A., Andersen, E., and Van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature* 463, 913–918. doi: 10.1038/nature08781

Ram, E. V., and Meshorer, E. (2009). Transcriptional competence in pluripotency. *Genes Dev.* 23, 2793–2798. doi: 10.1101/gad.1881609

Richard, A., Boullu, L., Herbach, U., Bonnafoux, A., Morin, V., Vallin, E., et al. (2016). Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. *PLoS Biol.* 14:e1002585. doi: 10.1371/journal.pbio.1002585

Semrau, S., Goldmann, J. E., Soumillon, M., Mikkelsen, T. S., Jaenisch, R., and Van Oudenaarden, A. (2017). Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. *Nat. Commun.* 8:1096.

Simon, C. S., Hadjantonakis, A. K., and Schroter, C. (2018). Making lineage decisions with biological noise: lessons from the early mouse embryo. *Wiley Interdiscip. Rev. Dev. Biol.* 7, e319. doi: 10.1002/wdev.319

Slack, J. M. (2002). Conrad Hal Waddington: the last renaissance biologist? *Nat. Rev. Genet.* 3, 889–895. doi: 10.1038/nrg933

Spivakov, M., and Fisher, A. G. (2007). Epigenetic signatures of stem-cell identity. *Nat. Rev. Genet.* 8, 263–271. doi: 10.1038/nrg2046

Stumpf, P. S., Smith, R. C. G., Lenz, M., Schuppert, A., Muller, F. J., Babtie, A., et al. (2017). Stem cell differentiation as a non-markov stochastic process. *Cell Syst.* 5, 268–282.e267.

Sunadome, K., Suzuki, T., Usui, M., Ashida, Y., and Nishida, E. (2014). Antagonism between the master regulators of differentiation ensures the discreteness and robustness of cell fates. *Mol. Cell* 54, 526–535. doi: 10.1016/j.molcel.2014.03.005

Terskikh, A. V., Miyamoto, T., Chang, C., Diatchenko, L., and Weissman, I. L. (2003). Gene expression analysis of purified hematopoietic stem cells and committed progenitors. *Blood* 102, 94–101. doi: 10.1182/blood-2002-08-2509

Tetteh, P. W., Farin, H. F., and Clevers, H. (2015). Plasticity within stem cell hierarchies in mammalian epithelia. *Trends Cell Biol.* 25, 100–108. doi: 10.1016/j.tcb.2014.09.003

Thery, M., and Bornens, M. (2006). Cell shape and cell division. *Curr. Opin. Cell Biol.* 18, 648–657. doi: 10.1016/j.ceb.2006.10.001

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., et al. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* 282, 1145–1147. doi: 10.1126/science.282.5391.1145

Torres-Padilla, M. E., and Chambers, I. (2014). Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. *Development* 141, 2173–2181. doi: 10.1242/dev.102624

Varahan, S., Walvekar, A., Sinha, V., Krishna, S., and Laxman, S. (2019). Metabolic constraints drive self-organization of specialized cell groups. *elife* 8:e46735.

Villani, M., Barbieri, A., and Serra, R. (2011). A dynamical model of genetic networks for cell differentiation. *PLoS One* 6:e17703. doi: 10.1371/journal.pone.0017703

Wabik, A., and Jones, P. H. (2015). Switching roles: the functional plasticity of adult tissue stem cells. *Embo J.* 34, 1164–1179. doi: 10.15252/embj.201490386

Wernet, M. F., Mazzoni, E. O., Celik, A., Duncan, D. M., Duncan, I., and Desplan, C. (2006). Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* 440, 174–180. doi: 10.1038/nature04615

Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K. L., et al. (2013). Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell* 154, 1112–1126. doi: 10.1016/j.cell.2013.08.007

Young, R. A. (2011). Control of the embryonic stem cell state. *Cell* 144, 940–954. doi: 10.1016/j.cell.2011.01.032

Zipori, D. (2004). The nature of stem cells: state rather than entity. *Nat. Rev. Genet.* 5, 873–878. doi: 10.1038/nrg1475

Zipori, D. (2005). The stem state: plasticity is essential, whereas self-renewal and hierarchy are optional. *Stem Cells* 23, 719–726. doi: 10.1634/stemcells.2005-0030

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.

frontiers in Bioengineering and Biotechnology

PERSPECTIVE
published: 18 September 2020
doi: 10.3389/fbioe.2020.583415

Advancing Antimicrobial Resistance Research Through Quantitative Modeling and Synthetic Biology

*Kevin S. Farquhar[1], Harold Flohr[2] and Daniel A. Charlebois[2]**

[1] Precision for Medicine, Houston, TX, United States, [2] Department of Physics, University of Alberta, Edmonton, AB, Canada

Antimicrobial resistance (AMR) is an emerging global health crisis that is undermining advances in modern medicine and, if unmitigated, threatens to kill 10 million people per year worldwide by 2050. Research over the last decade has demonstrated that the differences between genetically identical cells in the same environment can lead to drug resistance. Fluctuations in gene expression, modulated by gene regulatory networks, can lead to non-genetic heterogeneity that results in the fractional killing of microbial populations causing drug therapies to fail; this non-genetic drug resistance can enhance the probability of acquiring genetic drug resistance mutations. Mathematical models of gene networks can elucidate general principles underlying drug resistance, predict the evolution of resistance, and guide drug resistance experiments in the laboratory. Cells genetically engineered to carry synthetic gene networks regulating drug resistance genes allow for controlled, quantitative experiments on the role of non-genetic heterogeneity in the development of drug resistance. In this perspective article, we emphasize the contributions that mathematical, computational, and synthetic gene network models play in advancing our understanding of AMR to discover effective therapies against drug-resistant infections.

Keywords: antimicrobial resistance, gene regulatory networks, mathematical modeling and simulation, non-genetic heterogeneity, stochastic gene expression, synthetic biology

OPEN ACCESS

Edited by:
Mohit Kumar Jolly,
Indian Institute of Science (IISc), India

Reviewed by:
Simon Mitchell,
Brighton and Sussex Medical School, United Kingdom
Mario Andrea Marchisio,
Tianjin University, China
Sangram Bagh,
Saha Institute of Nuclear Physics (SINP), India

****Correspondence:***
Daniel A. Charlebois
dcharleb@ualberta.ca

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *14 July 2020*
Accepted: *02 September 2020*
Published: *18 September 2020*

Citation:
Farquhar KS, Flohr H and Charlebois DA (2020) Advancing Antimicrobial Resistance Research Through Quantitative Modeling and Synthetic Biology. Front. Bioeng. Biotechnol. 8:583415. doi: 10.3389/fbioe.2020.583415

INTRODUCTION

Antimicrobial resistance (AMR) is an emerging health crisis that is undermining modern medicine (World Health Organization, 2014). AMR arises when bacteria, fungi, viruses or other microbes no longer respond to the antimicrobial drugs used to treat them. As of 2016, 700,000 deaths per year are attributed to AMR (O'Neill and The Review on Antimicrobial Resistance, 2016). If unmitigated, it is estimated that by 2050, AMR will kill 10 million people per year globally and result in a cumulative lost global production cost of 100 trillion USD. Though it has been argued that these figures may be over-estimates (de Kraker et al., 2016), there is undoubtedly a large and increasing clinical and public health burden associated with AMR. Drug resistance during chemotherapy also continues to be the major limiting factor for successfully treating patients with cancer (Vasan et al., 2019). In order to mitigate drug resistance, we need to establish new quantitative tools to study the drug resistance process, to discover new drugs, and to develop novel treatment strategies that extend the "lifespan" of antimicrobial and chemotherapy drugs.

It is well established that drug resistance can develop through genetic mutation (**Figure 1A**) that causes a permanent change in a micro-organism's DNA or through the acquisition of a drug

resistance gene (e.g., horizontal gene transfer that occurs in bacteria) (Ochman et al., 2000). More recently, research has uncovered a new form of non-genetic stress resistance that can arise from fluctuations in gene expression in clonal cell populations (**Figure 1B**; Fraser and Kærn, 2009; Geiler-Samerotte et al., 2013; van Boxtel et al., 2017); this, for example, includes the non-genetic drug resistance associated with the increased expression of genes that encode efflux proteins that pump antimicrobial drugs out of pathogenic yeasts such as *Candida glabrata* (Ben-Ami et al., 2016) and *Cryptococcus neoformans* (Mondon et al., 1999). Targeting this phenomenon will be important for mitigating AMR, as some non-genetically drug-resistant pathogens are not easily detected by standard laboratory tests (Sears and Schwartz, 2017) and non-genetic drug resistance may be associated with the failure of antimicrobial therapies (Ben-Ami et al., 2016; Wuyts et al., 2018) and chemotherapies (Brock et al., 2009). Non-genetic heterogeneity resulting in drug resistance has been shown to be modulated by gene regulatory network structure (e.g., in the PDR network discussed below) (Charlebois et al., 2014; Inde and Dixon, 2018; Camellato et al., 2019) and may accumulate through multiple slightly asymmetric cell divisions (Mitchell et al., 2018; Tripathi et al., 2020). The emerging paradigm is that drug resistance is a multi-stage process and that acute, non-genetic drug resistance can facilitate the evolution of permanent, genetic drug resistance (**Figure 1C**). Non-genetic mechanisms are now thought to facilitate genetic resistance by increasing the population size under drug treatment and hence the chance of acquiring genetic mutations (Brock et al., 2009; Charlebois et al., 2011; Farquhar et al., 2019), and through synergism between adaptive mutations and non-genetic heterogeneity (Bódi et al., 2017; Salgia and Kulkarni, 2018). Furthermore, it is known that mutations in PDR1, a gene that regulates PDR5 in the pleiotropic drug resistance (PDR) network in yeast (**Figure 2A**), can cause full resistance to the antifungal drug fluconazole (Ferrari et al., 2009). Though, more research is needed to elucidate the interplay between non-genetic and genetic forms of drug resistance.

Mathematical models of drug resistance have been used for over three decades (Lavi et al., 2012); many older mathematical studies were based on ABC (ATP-binding cassette) transporters, such as the PDR5 gene that is regulated by PDR1 in the PDR network, as the main mechanism of resistance. These models are now beginning to include more contemporary knowledge of AMR mechanisms and incorporate how drug resistance gene networks function and evolve during treatment (Charlebois et al., 2014; Farquhar et al., 2019). Mathematical models have the potential to predict the effectiveness of various treatment strategies, such as using combination drug therapies to overcome AMR (Baym et al., 2016), and can guide laboratory experiments by identifying experimental targets and by narrowing down the immense number of ways that antimicrobial drugs can be applied. Additionally, these models can elucidate mechanisms underlying the development of AMR (e.g., Farquhar et al., 2019) and predict AMR from experimental data (Arepyeva et al., 2017).

Synthetic biology is rapidly becoming part of the solution to many of our needs in medicine, agriculture, and energy production (El Karoui et al., 2019). A particularly promising application is to genetically engineer micro-organisms to carry synthetic gene networks to study AMR in a more quantitative, controlled, and efficient manner than has been possible using traditional ("natural" or non-genetically modified) model micro-organisms (González et al., 2015). At present, it is extremely challenging to develop and experimentally validate mathematical models using pathogens, where drug resistance genes have evolved to be highly connected to the host genome; for instance, the expression of MDR1/p-glycoprotein (responsible for multiple drug resistance (MDR) of tumors to chemotherapy; Gottesman et al., 2002) is regulated by a multitude of factors, making it difficult to quantitatively study how non-genetic mechanisms may contribute to AMR and drug resistance in cancer (Camellato et al., 2019). Furthermore, unlike synthetic drug resistance networks, many native resistance networks are still not known completely. Nevertheless, the design of synthetic gene networks is a model-guided process (Sakurai and Hori, 2018) and these networks are constructed to mimic known natural drug resistance networks using techniques from genetic engineering (Cameron et al., 2014; Bartley et al., 2017).

MATHEMATICAL MODELING OF NON-GENETIC ANTIMICROBIAL RESISTANCE

Modeling Non-genetic Gene Expression Heterogeneity in Drug Resistance

Early work on non-genetic drug resistance focused on the amplitude of fluctuations or "noise" in the expression of drug resistance genes. Models predicted that low gene expression noise would be beneficial under low drug concentrations and that high gene expression noise would be beneficial under high drug concentrations (**Figure 1B**, inset; Blake et al., 2006; Zhuravel et al., 2010); these predictions were confirmed experimentally in the same studies. Subsequently, a more general theoretical framework was developed that incorporated the frequency of gene expression noise, as well as the amplitude of the expression noise (Charlebois et al., 2011). Importantly, using this quantitative framework it was hypothesized that drug resistance can develop independently of mutation, provided that the fluctuation timescales are sufficiently long. Cell population models (Arino and Kimmel, 1993; Henson, 2003; Charlebois and Balázsi, 2019) have also been used to incorporate the multi-scale nature of AMR. For instance, a stochastic model of gene expression was combined with a population simulation algorithm to computationally investigate the evolution of gene expression noise (Charlebois, 2015).

Modeling Drug Resistance Networks in Microbes and Mammalian Cells

Mathematical models have been used to investigate the effect gene network structures or motifs have on AMR. For instance, it was shown computationally that gene network motifs can enhance drug resistance by modulating gene expression noise within a cell population (Charlebois et al., 2014). Charlebois et al. showed

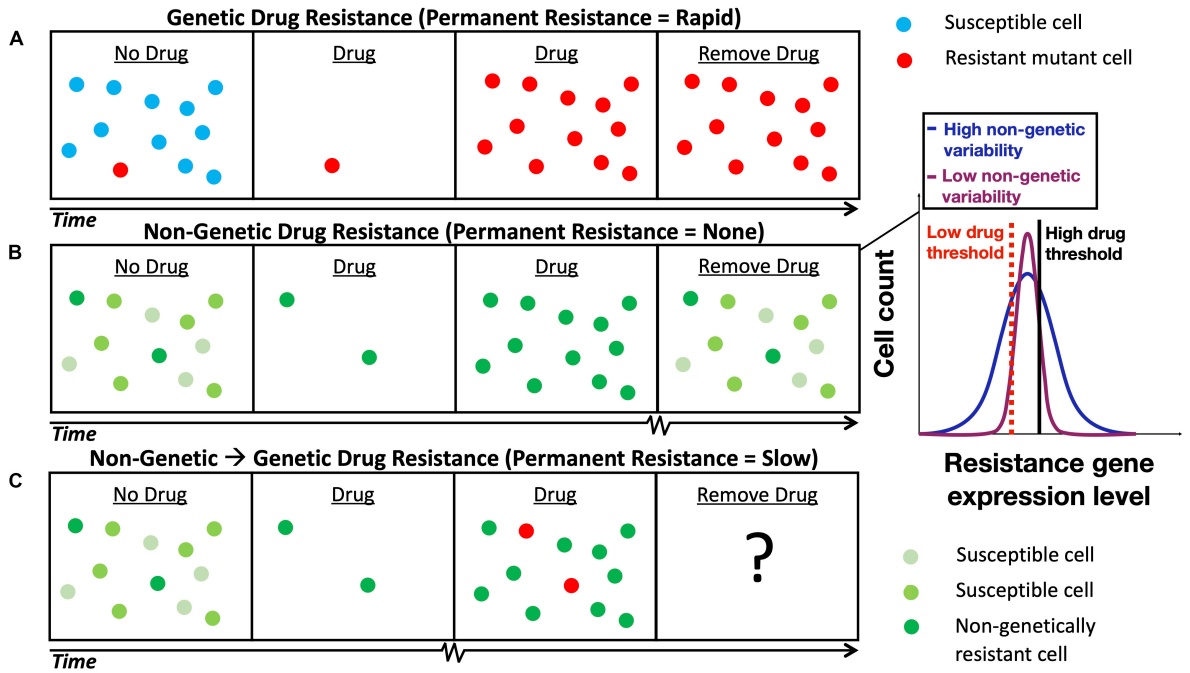

FIGURE 1 | Schematic depicting the development of non-genetic and genetic drug resistance. **(A)** The development of genetic drug resistance via evolution by natural selection of a pre-existing drug resistance mutation. **(B)** The development of non-genetic drug resistance in a clonal cell population via the selection cells with sufficiently long gene expression fluctuation timescales. The shade of green denotes the level of gene expression of a drug resistance gene inside the cell; lighter and darker shades of green represent lower levels and higher levels of gene expression, respectively. Inset illustrates gene expression histograms typically obtained from clonal cell populations with low and high levels of non-genetic, cell-to-cell variability. A high level of non-genetic variability is advantageous at high drug concentrations and a low level of non-genetic variability is advantageous at low drug concentrations (cells with resistance gene expression levels below a given drug threshold perish). **(C)** The evolution of longer-term, genetic drug resistance is facilitated by shorter-term, non-genetic drug resistance; the ultimate outcome will be determined based on the condition-dependent relative fitness of each subpopulation. Squiggly lines on the time axes in **(B,C)** represent longer timescales. Notably, the relaxation timescales of non-genetic fluctuations in the expression of drug resistance genes regulated by positive feedback gene circuits have been estimated to be 58 h in mammalian cells (Farquhar et al., 2019) and 283 h in yeast (Nevozhay et al., 2012).

that feedforward and positive feedback loops, the same network motifs that have been found to be imbedded in some gene networks regulating AMR in pathogenic yeast (Kolaczkowski et al., 1998) and human cancer cells (Misra et al., 2005), enhance drug resistance *in silico*. This new understanding of how gene network structure regulates AMR opens up new lines of research and identifies new potential therapeutic targets (e.g., targeting regulator genes in the network, rather than the drug resistance genes they control) against drug-resistant pathogens and cancers to be investigated experimentally.

Mathematical modeling and computer simulations have been used to predict how drug efflux pump proteins affect gene network function and fitness in prokaryotic and eukaryotic organisms. In Langevin and Dunlop (2018) it was found experimentally that the cellular fitness benefit of AcrAB-TolC, a well-known multi-drug resistance pump in *E. coli*, depended on the rate of stress induction; fits to data allowed the fitness benefit that the pumps conferred under different stress induction rates to be accurately predicted by mathematical models. In another study, it was predicted that incorporating negative feedback via drug efflux pumps into synthetic gene networks can increase the response of the gene network at low antibiotic inducer concentrations (Diao et al., 2016). This prediction was confirmed experimentally in the same study using synthetic gene networks in *S. cerevisiae* and was found to be the result of reduced regulator gene expression.

In Farquhar et al. (2019) the authors developed a stochastic population dynamics model to infer mechanisms for drug resistance in mammalian cells. The stochastic population model predicted that gene network motifs facilitate the development of acute drug resistance and that non- or slow-growing subpopulations of "persister-like" cells that do not succumb are critical reservoirs for the development of fast growing, heritably drug-resistant mutants enabling longer-term drug resistance (see Brauner et al., 2016; Rosenberg et al., 2018; Berman and Krysan, 2020 for the distinction between "tolerance," "heteroresistance" or "persistence," and "resistance"). This study compliments previous work in bacteria that demonstrated that antibiotic tolerant non- or slow-growing mutant cells precede the developed genetic drug resistance during intermittent antibiotic exposure (Levin-Reisman et al., 2017). The persistence phenotype (e.g., Kussell et al., 2005) and stochastic phenotype switching (e.g., Acar et al., 2008) have also been investigated in mathematical models and experiments on genetically engineered micro-organisms and found to affect fitness in fluctuating environments.

Ultimately, mathematical and computation models of AMR must be validated by performing quantitative drug resistance experiments; genetically engineered cells that harbor synthetic gene networks controlling the expression of drug resistance genes is proving to be an effective experimental model system.

SYNTHETIC DRUG RESISTANCE GENE NETWORKS AND ANTIMICROBIAL RESISTANCE EXPERIMENTS

Genetic engineering techniques are used to synthetize and combine DNA to build synthetic gene networks or "circuits" (Cameron et al., 2014) that control drug resistance genes. Common methods used to engineer synthetic gene networks include recombinant molecular cloning, Gibson assembly (Gibson et al., 2009; Santos-Moreno and Schaerli, 2019), and CRISPR-Cas9 gene editing (Nissim et al., 2014; Jusiak et al., 2016). Cell-to-cell heterogeneity may cause unexpected deviations from intended synthetic gene circuit behavior (Beach et al., 2017). However, statistical tools can enhance the design process and reliability of synthetic gene networks (Sakurai and Hori, 2018). With proper design, synthetic gene networks can be precisely tuned to control gene expression mean and noise levels using chemical inducers that do not adversely affect the micro-organisms harboring these networks.

Synthetic Antimicrobial Resistance Gene Networks

Synthetic gene networks have been engineered to regulate drug resistance and have been shown to serve as well-characterized models of natural stress response modules in evolution experiments (González et al., 2015; Bódi et al., 2017; Farquhar et al., 2019; Gouda et al., 2019).

Nevozhay et al. (2012) constructed a two-gene positive feedback network that enables bi-stable gene expression to control a Zeocin antibiotic resistance gene in *S. cerevisiae*.

In this work, a computational approach based on stochastic cellular movement in "gene expression space" was used to predict cell population fitness of low- and high-expressing subpopulations. The authors found an optimum on the fitness landscape that balances the costs and benefits of expressing a drug resistance gene in various experimental antibiotic inducer and drug conditions. In a subsequent microbial evolution study using the same positive feedback yeast strain, it was found that the synthetic gene network was fine-tuned by evolution to modulate the network's noisy response and optimize fitness via specific "intra-circuit" and "extra-circuit" DNA mutations (González et al., 2015), which can lead to loss of gene circuit function that can be regained in certain conditions under drug selection (Gouda et al., 2019). The study by Gouda et al. (2019) also suggests that slow growth due to antibiotics may allow cells to access hidden drug-resistant states prior to genetic changes. Computational models based on fitness and gene expression properties have been developed to predict specific aspects of evolutionary dynamics (including the speed at which the ancestral genome disappears from the population and the types and number of mutant alleles that establish in each experimental condition) in different inducer and drug conditions (González et al., 2015). These computational models were validated in the same studies by microbial evolution experiments on the genetically engineered positive feedback yeast strain (Nevozhay et al., 2012; González et al., 2015; Gouda et al., 2019).

Genetically engineered networks have also been designed to control the expression of genes that encode efflux proteins that lead to AMR. Diao et al. (2016) used synthetic negative feedback gene networks, inducible by the antibiotic doxycycline, to regulate the expression of PDR5. This study found that the addition of a second layer of negative feedback (resulting from pumping doxycycline out of the cell by the PDR5 protein) altered the dose-responses of the original gene circuits in a manner that was predictable by mathematical modeling. In another study, Camellato et al. (2019) engineered a synthetic gene network in yeast to mimic the PDR5 and MDR1 networks that underly multi-drug resistance in yeast and human breast cancer cells (**Figure 2B**). In agreement with computational predictions made years earlier (Charlebois et al., 2014), the authors found that feedforward and positive feedback network motifs enabled rapid, self-sustained activation of gene expression leading to enhanced cell survival in the presence of a cytotoxic drug. It has been proposed that activating the expression of genes that encode multi-drug resistance efflux pump proteins in the absence of antibiotic pressure may allow susceptible bacteria to outcompete resistant bacteria, which normally down-regulate the expression of resistance genes in conditions without antibiotics to eliminate the associated fitness cost (Wang et al., 2019).

Synthetic Drug Resistance Gene Circuits in Mammalian Cells

To experimentally investigate the role of non-genetic cell-to-cell variability in cancer drug resistance, it is imperative to precisely control this non-genetic heterogeneity that can drive adaptation to chemotherapeutic agents. Synthetic gene circuits integrated in mammalian cells can be designed to precisely control gene expression noise for drug resistance genes, while keeping the mean expression levels the same (**Figure 2F**; Aranda-Díaz et al., 2017; Farquhar et al., 2019). This approach allows synthetic gene circuits to separate key biological variables contributing to resistance from other confounding variables like mean expression and genetic background.

In Chinese Hamster Ovary (CHO) cells with a recombinase-mediated integration site known as a "Flp-In" landing pad (Wirth and Hauser, 2004), Farquhar et al. (2019) designed, constructed, and integrated into the landing pad a mammalian negative feedback (mNF) synthetic gene circuit (**Figure 2C**) based on a humanized tetracycline repressor (hTetR) gene (Nevozhay et al., 2013); the mNF circuit demonstrated doxycycline-inducible expression of a purmoycin drug resistance gene (PuroR) with low gene expression noise (**Figure 2E**). Highlighting the advantages of mathematical modeling in synthetic gene circuit design, the mNF circuit was based on another gene circuit transferred that applied modeling to predict the effects of multiple design

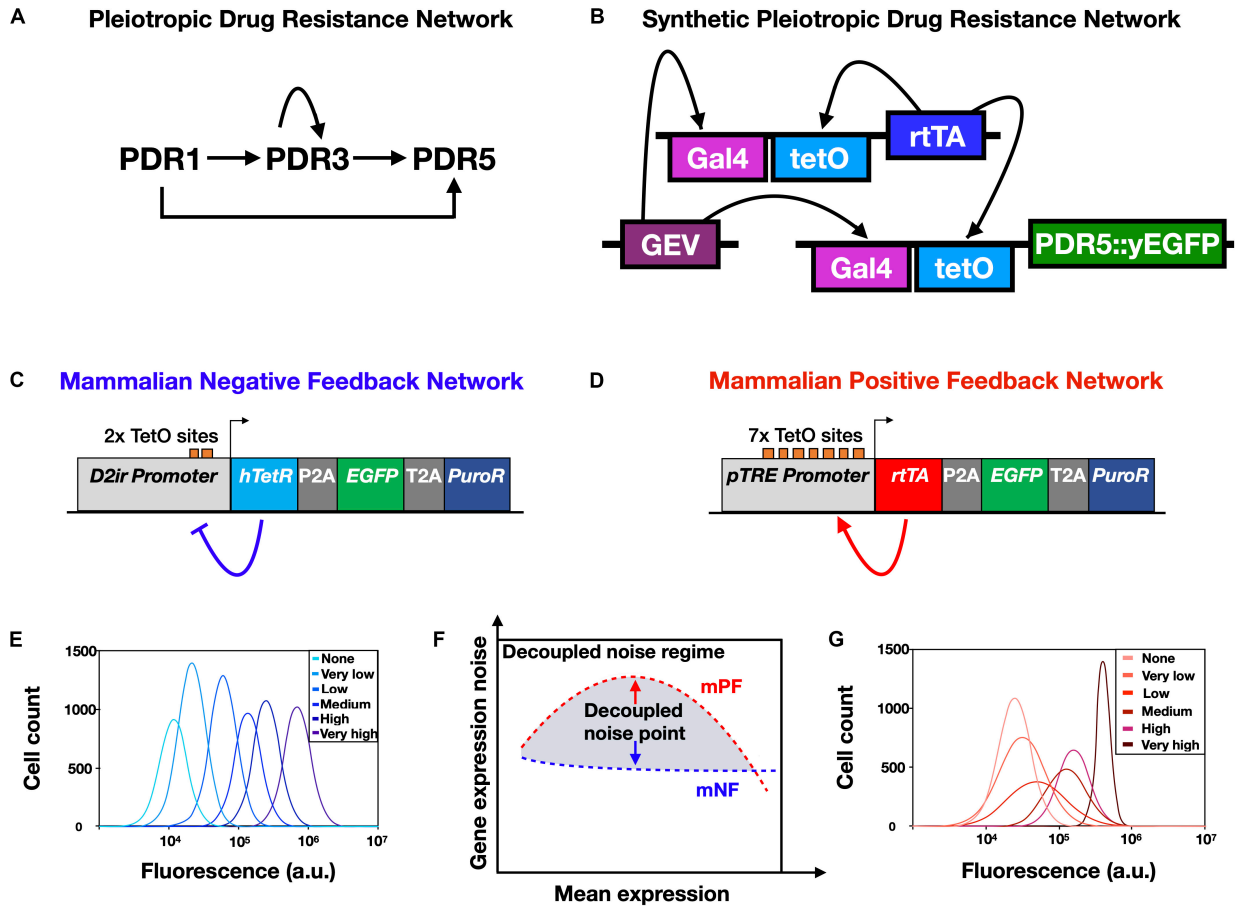

FIGURE 2 | Synthetic gene networks engineered to mimic natural drug resistance gene networks and study the effects of non-genetic heterogeneity on AMR. **(A)** Yeast pleiotropic drug resistance (PDR) gene network. Imbedded in this gene network structure are positive feedback regulation (self-activation of PDR3) and feedforward regulation (PDR1 indirectly activates PDR5 through PDR3) and direct activation (PDR1 activates PDR5). **(B)** A synthetic gene network engineered to have the same network motif as the PDR network shown in **(A)**. Note that fluorescence reporter genes, such as yEGFP, are fused to drug resistance genes, such as PDR5, to enable experimental measurement. **(C)** Schematic of the mammalian negative feedback (mNF) gene network, which expresses the humanized tetracycline repressor (hTetR) gene, the puromycin resistance gene (PuroR), and the fluorescence reporter EGFP separated by self-cleaving 2A elements. **(D)** Schematic of the mammalian positive feedback (mPF) gene network, which expresses the reverse tetracycline transactivator (rtTA), PuroR, and EGFP separated by self-cleaving 2A elements. **(E)** Schematic of representative gene expression histograms obtained from a dose response of the mNF strain. **(F)** The decoupled noise regime (gray shading) is composed of decoupled noise points (red and blue arrows), which occur where the mean gene expression noise for the high-noise mPF and the low-noise mNF gene networks are decoupled from mean gene expression. **(G)** Schematic of representative gene expression histograms obtained from a dose response of the mPF strain. The legends in **(E,G)** indicate the inducer (doxycycline) level for each distribution. Panel **(A)** was reproduced from Charlebois et al. (2014) with permission, panel **(B)** was adapted with permission from Camellato (2018), and panels **(C–G)** were adapted from Farquhar et al. (2019), CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

iterations, leading to increased fold change and minimal gene expression noise (Nevozhay et al., 2013). Complementing the low noise mNF gene circuit, Farquhar et al. (2019) also constructed a mammalian positive feedback (mPF) gene circuit (**Figure 2D**) regulated by a reverse tetracycline trans-activator (rtTA), integrated into the same CHO genomic integration site as the low-noise mNF circuit, leading to doxycycline-inducible expression of PuroR with high levels of gene expression noise (**Figure 2G**). Notably, no bimodal gene expression regime was observed for the mPF gene circuit at intermediate doxycycline concentrations, though bimodality is observed for the PF gene circuit in yeast (e.g., Nevozhay et al., 2012). Though the exact mechanism for the lack of observed bimodality in the mPF gene circuit remains unknown, this highlights that genetically engineered circuits do not always function the same way in different organisms. When inducing the two circuits in mammalian CHO cells to express the same PuroR mean expression level (**Figure 2F**) and treating the CHO cells with various concentrations of puromycin, the authors found that adaptation to low concentrations of drug was more rapid for the mNF circuit with low gene expression noise. On the other hand, high gene expression noise from the mPF circuit facilitated adaptation to high levels of puromycin, while cells with the mNF circuit treated at a high puromycin concentration did not adapt at all. This validated the approach to investigating drug resistance and noise in mammalian cells using synthetic gene networks,

which allowed gene expression noise to be decoupled from mean drug resistance gene expression in isogenic cells; this approach could also help to further elucidate the role of rare-cell expression and drug-induced reprogramming in mammalian drug resistance (Shaffer et al., 2017).

DNA sequencing of the gene circuits after adaptation to puromycin and monitoring expression and survival after temporary removal of drug revealed adaptation mechanisms (Farquhar et al., 2019). The self-repression from the tetracycline repressor in the mNF circuit tended to break down through intra-circuit mutations, leading to higher PuroR expression and irreversible resistance to puromycin even without circuit induction. In the mPF circuit, no intra-circuit mutations were found despite PuroR expression levels remaining elevated above pre-treatment mean expression levels, which was reversible and led to re-sensitization to puromycin after inducer removal. Epigenetic factors and chromatin modifications may have driven the elevated expression at the genomic locus which was evolutionarily selected for during adaptation (Berger, 2007). By using synthetic gene networks containing a drug resistance gene in isogenic mammalian cells, Farquhar et al. addressed a long-standing question regarding how non-genetic heterogeneity could lead to initial cell survival during chemotherapy which then facilitates the development of genetic drug resistance in cancer (Brock et al., 2009).

DISCUSSION

A new interdisciplinary field of research is emerging that combines multi-scale quantitative models with synthetic biology to rationally design gene networks using engineering principles for AMR research. One important goal is to use these models to predict the effects of non-genetic drug resistance on the evolution of genetic drug resistance. Another important goal is to advance pharmaceutical and clinical AMR research by investigating new "resistance proof" antimicrobial compounds and novel therapeutic treatment strategies.

Moving forward, a challenge that must be addressed is how to adapt the mathematical models and translate the experimental discoveries made using synthetic systems to pathogens with complex and highly interconnected gene regulatory networks. More research on pathogenic micro-organisms and mammalian cells is needed to elucidate the underpinnings of non-genetic resistance at the molecular and single-cell levels. Specifically, capturing the complexity of native resistance mechanisms, which are not completely understood, with synthetic gene networks presents both an obstacle and an opportunity. Quantifying known drug resistance effects in genetically engineered organisms may elucidate native resistance mechanisms in pathogens. One possible approach involves adding additional regulatory interactions iteratively to a well-understood synthetic gene network controlling a drug resistance gene and making predictions for their impact on drug resistance. Eventually, the networks become complex enough to mimic the phenotypes caused by native resistance mechanisms. In the cases of cancer subtypes, introducing synthetic gene circuits controlling a gene with specific mutations associated with chemotherapy resistance will be challenging, with genomic instability possibly corrupting the gene circuit. Targeted genomic integration in various cancer cell lines and primary cell strains will also differ in efficiency, making comparisons between cell types difficult. To address these challenges, a better understanding of the gene regulatory networks, mutations, and signal transduction pathways associated with chemoresistance in specific cancers is needed.

Research incorporating quantitative modeling and genetically engineered networks will be critical to fully understand how non-genetic and genetic mechanisms interact in the development of drug resistance, and to discover effective strategies that target acute non-genetic drug resistance to alleviate the development of permanent genetic drug resistance in infectious diseases and cancers. Several promising approaches include synergistically using noise-enhancing compounds to reactivate latent HIV to increase sensitivity to existing antiviral drugs (Dar et al., 2014), using combined drug treatment regimens to target non-proliferating *M. tuberculosis* persisters to reduce treatment times (Zhang et al., 2012), eliminating bacterial persisters using engineering approaches that target bacterial metabolism (Allison et al., 2011), and the use of epigenetic modifiers in combination with targeted therapies to reduce the ability of a cancerous cell to switch phenotypes to acquire a drug-resistant state (Salgia and Kulkarni, 2018).

Overall, combining mathematical models and synthetic gene networks is leading to new quantitative model systems for drug resistance research, which are desperately needed to advance our fundamental understanding of the multi-stage drug resistance process. Ultimately it remains to be seen how discoveries made using these quantitative model systems will translate to pathogens and cancer. However, the potential of this new area of research to help mitigate the socio-economic costs of drug resistance warrants its relentless pursuit.

AUTHOR CONTRIBUTIONS

DC conceptualized and supervised the study. KF, HF, and DC contributed to the literature review. HF and DC developed the figures. KF and DC wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

Precision provided indirect funding support for this manuscript by way of employment of KF. No additional funding was provided for this manuscript by Precision or any other commercial entity. DC was supported in part by funding from the Government of Canada's New Frontiers in Research Fund – Exploration grant program (NFRFE-2019-01208).

ACKNOWLEDGMENTS

We are grateful to Prof. Gábor Balázsi for helpful comments on the manuscript. We thank Mr. Mark Igmen for assisting with the literature review. A preprint of this manuscript has been posted on the arXiv (Farquhar et al., 2020).

REFERENCES

Acar, M., Mettetal, J. T., and van Oudenaarden, A. (2008). Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* 40, 471–475. doi: 10.1038/ng.110

Allison, K. R., Brynildsen, M. P., and Collins, J. J. (2011). Metabolite-enabled eradication of bacterial persisters by aminoglycosides. *Nature* 473, 216–221. doi: 10.1038/nature1006

Aranda-Díaz, A., Mace, K., Zuleta, I., Harrigan, P., and El-Samad, H. (2017). Robust Synthetic Circuits for Two-Dimensional Control of Gene Expression in Yeast. *ACS Synth. Biol.* 6, 545–554. doi: 10.1021/acssynbio.6b00251

Arepyeva, M. A., Kolbin, A. S., Sidorenko, S. V., Lawson, R., Kurylev, A. A., Balykina, Y. E., et al. (2017). A mathematical model for predicting the development of bacterial resistance based on the relationship between the level of antimicrobial resistance and the volume of antibiotic consumption. *J. Glob. Antimicrob Resist.* 8, 148–156. doi: 10.1016/j.jgar.2016.11.010

Arino, O., and Kimmel, M. (1993). Comparison of approaches to modeling of cell population dynamics. *SIAM J. Appl. Math.* 53, 1480–1504. doi: 10.1137/0153069

Bartley, B. A., Kim, K., Medley, J. K., and Sauro, H. M. (2017). Synthetic biology: engineering living systems from biophysical principles. *Biophys. J.* 112, 1050–1058. doi: 10.1016/j.bpj.2017.02.013

Baym, M., Stone, L. K., and Kishony, R. (2016). Multidrug evolutionary strategies to reverse antibiotic resistance. *Science* 351:aad3292. doi: 10.1126/science.aad3292

Beach, R. R., Ricci-Tam, C., Brennan, C. M., Moomau, C. A., Hsu, P.-H., Hua, B., et al. (2017). Aneuploidy causes non-genetic individuality. *Cell* 169, 229–242. doi: 10.1016/j.cell.2017.03.021

Ben-Ami, R., Zimmerman, O., Finn, T., Amit, S., Novikov, A., Wertheimer, N., et al. (2016). Heteroresistance to Fluconazole is a continuously distributed phenotype among Candida glabrata clinical strains associated with in vivo persistence. *mBio* 7, e616–e655. doi: 10.1128/mBio.00655-16

Berger, S. (2007). The complex language of chromatin regulation during transcription. *Nature* 447, 407–412. doi: 10.1038/nature05915

Berman, J., and Krysan, D. J. (2020). Drug resistance and tolerance in fungi. *Nat. Rev. Microbiol.* 18, 319–331. doi: 10.1038/s41579-019-0322-2

Blake, W. J., Balázsi, G., Kohanski, M. A., Isaacs, F. J., Murphy, K. F., Kuang, Y., et al. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell.* 24, 853–856. doi: 10.1016/j.molcel.2006.11.003

Bódi, Z., Farkas, Z., Nevozhay, D., Kalapis, D., Lázár, V., Csörgö, B., et al. (2017). Phenotypic heterogeneity promotes adaptive evolution. *PLoS Biol.* 15:e2000644. doi: 10.1371/journal.pbio.1002607

Brauner, A., Fridman, O., Gefen, O., and Balaban, N. Q. (2016). Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nature* 14, 320–330. doi: 10.1038/nrmicro.2016.34

Brock, A., Chang, H., and Huang, S. (2009). Non-genetic heterogeneity-a mutation-independent driving force for the somatic evolution of tumors. *Nat. Rev. Genet.* 10, 336–342. doi: 10.1038/nrg2556

Camellato, B. (2018). *Gene regulatory networks are a mechanism for drug resistance.* Master's thesis, (Canada: University of Ottawa).

Camellato, B., Roney, I. J., Azizi, A., Charlebois, D., and Kærn, M. (2019). Engineered gene networks enable non-genetic drug resistance and enhances cellular robustness. *Eng. Biol.* 3, 72–79. doi: 10.1049/enb.2019.0009

Cameron, D. E., Bashor, C. J., and Collins, J. J. (2014). A brief history of synthetic biology. *Nat. Rev. Microbiol.* 12, 381–390. doi: 10.1038/nrmicro3239

Charlebois, D. A. (2015). Effect and evolution of gene expression noise on the fitness landscape. *Phys. Rev. E.* 92:022713. doi: 10.1103/PhysRevE.92.022713

Charlebois, D. A., Abdennnur, N., and Kærn, M. (2011). Gene expression noise facilitates adaptation and drug resistance independently of mutation. *Phys. Rev. Lett.* 107:218101. doi: 10.1103/PhysRevLett.107.218101

Charlebois, D. A., and Balázsi, G. (2019). Modeling cell population dynamics. *In Silico Biol.* 13, 21–39. doi: 10.3233/ISB-180470

Charlebois, D. A., Balázsi, G., and Kærn, M. (2014). Coherent feedforward transcriptional regulatory motifs enhance drug resistance. *Phys. Rev. E.* 89:052708. doi: 10.1103/PhysRevE.89.052708

Dar, R. D., Hosmane, N. N., Arkin, M. R., Siliciano, R. F., and Weinberger, L. S. (2014). Screening for noise in gene expression identifies drug synergies. *Science* 344, 1392–1396. doi: 10.1126/science.1250220

de Kraker, M. E. A., Stewardson, A. J., and Harbarth, S. (2016). Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050? *PLoS Med.* 13:e1002184. doi: 10.1371/journal.pmed.1002184

Diao, J., Charlebois, D. A., Nevozhay, D., Bodi, Z., Csaba, P., and Balázsi, G. (2016). Efflux pump control alters synthetic gene circuit function. *ACS Synth. Biol.* 5, 619–631. doi: 10.1021/acssynbio.5b00154

El Karoui, M., Hoyos-Flight, M., and Fletcher, L. (2019). Future trends in synthetic biology-a report. *Front. Bioeng. Biotechnol.* 7:175. doi: 10.3389/fbioe.2019.00175

Farquhar, K. S., Charlebois, D. A., Szenk, M., Cohen, J., Nevozhay, D., and Balázsi, G. (2019). Role of network-mediated stochasticity in mammalian drug resistance. *Nat. Commun.* 10:2766. doi: 10.1038/s41467-019-10330-w

Farquhar, K. S., Flohr, H., and Charlebois, D. A. (2020). Advancing Drug Resistance Research Through Quantitative Modeling and Synthetic Biology. *arXiv*

Ferrari, S., Ischer, F., Calabrese, D., Posteraro, B., Sanguinetti, M., Fadda, G., et al. (2009). Gain of function mutations in CgPHD1 of Candida glabrata not only mediate antifungal resistance but also enhance virulence. *PLoS Pathog.* 5:e1000268. doi: 10.1371/journal.ppat.1000268

Fraser, D., and Kærn, M. (2009). A chance at survival: gene expression noise and phenotypic diversification strategies. *Mol. Microbiol.* 71, 1333–1340. doi: 10.1111/j.1365-2958.2009.06605.x

Geiler-Samerotte, K. A., Bauer, C. R., Li, S., Ziv, N., Gresham, D., and Siegal, M. L. (2013). The details in the distributions: why and how to study phenotypic variability. *Curr. Opin. Biotechnol.* 24, 752–759. doi: 10.1016/j.copbio.2013.03.010

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. III., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/NMETH.1318

González, C., Ray, J. C. J., Manhart, M., Adams, R. M., Nevozhay, D., Morozov, A. V., et al. (2015). Stress-response balance drives the evolution of a network module and its host genome. *Mol. Syst. Biol.* 11:827. doi: 10.15252/msb.20156185

Gottesman, M. M., Fojo, T., and Bates, S. E. (2002). Multidrug resistance in cancer: role of ATP-dependent transporters. *Nat. Rev. Cancer* 2, 48–58. doi: 10.1038/nrc706

Gouda, M. K., Manhart, M., and Balázsi, G. (2019). Evolutionary regain of lost gene circuit function. *Proc. Natl. Acad. Sci. USA* 116, 25162–25171. doi: 10.1073/pnas.1912257116

Henson, M. A. (2003). Dynamic modeling of microbial cell populations. *Curr. Opin. Biotech.* 14, 460–467. doi: 10.1016/S0958-1669(03)00104-6

Inde, Z., and Dixon, S. J. (2018). The impact of non-genetic heterogeneity on cancer cell death. *Crit. Rev. Biochem. Mol. Biol.* 53, 99–114. doi: 10.1080/10409238.2017.1412395

Jusiak, B., Cleto, S., Perez-Piñera, P., and Lu, T. K. (2016). Engineering synthetic gene circuits in living cells with CRISPR technology. *Trends Biotechnol.* 34, 353–547. doi: 10.1016/j.tibtech.2015.12.014

Kolaczkowski, M., Kolaczoska, A., Luczynski, J., Witek, S., and Goffeau, A. (1998). In Vivo characterization of the drug resistance profile of the major ABC transporters and other components of the yeast pleiotropic drug resistance network. *Microb. Drug. Resist.* 4, 143–158. doi: 10.1089/mdr.1998.4.143

Kussell, E., Kishony, R., Balaban, N. Q., and Leibler, S. (2005). Bacterial Persistence: A Model of Survival in Changing Environments. *Genetics* 169, 1807–1814. doi: 10.1534/genetics.104.035352

Langevin, A. M., and Dunlop, M. J. (2018). Stress introduction rate alters the benefit of AcrAB-TolC efflux pumps. *J. Bacteriol.* 200, e517–e525. doi: 10.1128/JB.00525-17

Lavi, O., Gottesman, M. M., and Levy, D. (2012). The dynamics of drug resistance: A mathematical perspective. *Drug Resist. Updat.* 15, 90–97. doi: 10.1016/j.drup.2012.01.003

Levin-Reisman, I., Ronin, I., Gefen, O., Braniss, I., Shoresh, N., and Balaban, N. Q. (2017). Antibiotic tolerance facilitates the evolution of resistance. *Science* 355, 826–830. doi: 10.1126/science.aaj2191

Misra, S., Ghatak, S., and Toole, B. P. (2005). Regulation of MDR1 Expression and Drug Resistance by a Positive Feedback Loop Involving Hyaluronan, Phosphoinositide 3-Kinase, and ErbB2. *J. Biol. Chem.* 280, 20310–20315. doi: 10.1074/jbc.M500737200

Mitchell, S., Roy, K., Zangle, T. A., and Hoffmann, A. (2018). Nongenetic origins of cell-to-cell variability in B lymphocyte proliferation. *Proc. Natl. Acad. Sci. USA* 115, E2888–E2897. doi: 10.1073/pnas.1715639115

Mondon, P., Petter, R., Amalfitano, G., Luzzati, R., Concia, E., Polacheck, I., et al. (1999). Heteroresistance to fluconazole and voriconazole in Cryptococcus neoformans. *Antimicrob. Agents Chemother.* 43, 1856–1861. doi: 10.1128/AAC.43.8.1856

Nevozhay, D., Adams, R. M., Itallie, E. V., Bennett, M. R., and Balázsi, G. (2012). Mapping the environmental fitness landscape of a synthetic gene circuit. *PLoS Comput. Biol.* 8:1–17. doi: 10.1371/journal.pcbi.1002480

Nevozhay, D., Zal, T., and Balázsi, G. (2013). Transferring a synthetic gene circuit from yeast to mammalian cells. *Nat. Commun.* 4, 1–11. doi: 10.1038/ncomms2471

Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P., and Lu, T. K. (2014). Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol. Cell.* 54, 698–710. doi: 10.1016/j.molcel.2014.04.022

Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304. doi: 10.1038/35012500

O'Neill, J., and The Review on Antimicrobial Resistance. (2016). *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations*. London: Wellcome Trust.

Rosenberg, A., Ene, I. V., Bibi, M., Zakin, S., Segal, E. S., Ziv, N., et al. (2018). Antifungal tolerance is a subpopulation effect distinct from resistance and is associated with persistent candidemia. *Nat. Commun.* 9:2470. doi: 10.1038/s41467-018-04926-x

Sakurai, Y., and Hori, Y. (2018). Optimization-based synthesis of stochastic biocircuits with statistical specifications. *J. R. Soc. Interface* 15:20170709. doi: 10.1098/rsif.2017.0709

Salgia, R., and Kulkarni, P. (2018). The genetic/non-genetic duality of drug 'resistance' in cancer. *Trends Cancer* 4, 110–118. doi: 10.1016/j.trecan.2018.01.001

Santos-Moreno, J., and Schaerli, Y. (2019). A framework for the modular and combinatorial assembly of synthetic gene circuits. *ACS Synth. Biol.* 8, 1691–1697. doi: 10.1021/acssynbio.9b00174

Sears, D., and Schwartz, B. S. (2017). Candida Auris: an emerging multidrug-resistant pathogen. *Int. J. Infect. Dis.* 63, 95–98. doi: 10.1016/j.ijid.2017.08.017

Shaffer, S. M., Dunagin, M. C., Torborg, S. R., Torre, E. A., Emert, B., Krepler, C., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* 546, 431–435. doi: 10.1038/nature22794

Tripathi, S., Chakraborty, P., Levine, H., and Jolly, M. K. (2020). A mechanism for epithelial-mesenchymal heterogeneity in a population of cancer cells. *PLoS Comput. Biol.* 16:e1007619. doi: 10.1371/journal.pcbi.1007619

van Boxtel, C., van Heerden, J. H., Nordholt, N., Schmidt, P., and Bruggeman, F. J. (2017). Taking chances and making mistakes: non-genetic phenotypic heterogeneity and its consequences for surviving in dynamic environments. *J. R. Soc. Interface* 14:20170141. doi: 10.1098/rsif.2017.0141

Vasan, N., Baselga, J., and Hyman, D. M. (2019). A view on drug resistance in cancer. *Nature* 575, 299–309. doi: 10.1038/s41586-019-1730-1

Wang, T., Kunze, C., and Dunlop, M. J. (2019). Salicylate increases fitness cost associated with MarA-mediated antibiotic resistance. *Biophys. J.* 117, 563–571. doi: 10.1016/j.bpj.2019.07.005

Wirth, D., and Hauser, H. (2004). Flp-mediated integration of expression cassettes into FRT-tagged chromosomal loci in mammalian cells. *Methods Mol. Biol.* 267, 467–476. doi: 10.1385/1-59259-774-2:467

World Health Organization. (2014). *Antimicrobial Resistance: Global report of surveillance*. Switzerland: WHO.

Wuyts, J., Van Dijck, P., and Holtappels, M. (2018). Fungal persister cells: the basis for recalcitrant infections? *PLoS Pathog.* 14:e1007301. doi: 10.1371/journal.ppat.1007301

Zhang, Y., Yew, W. W., and Barer, M. R. (2012). Targeting persisters for tuberculosis control. *Antimicrob. Agents Chemother.* 56, 2223–2230. doi: 10.1128/AAC.062886211

Zhuravel, D., Fraser, D., St-Pierre, S., Tepliakova, L., Pang, W. L., Hasty, J., et al. (2010). Phenotypic impact of regulatory noise in cellular stress-response pathways. *Syst. Synth. Biol.* 4, 105–116. doi: 10.1007/s11693-010-9055-2

Conflict of Interest: KF was employed by the company Precision for Medicine.

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.

frontiers in Genetics

OPINION
published: 30 October 2020
doi: 10.3389/fgene.2020.586726

Observation and Control of Gene Expression Noise: Barrier Crossing Analogies Between Drug Resistance and Metastasis

Michael Tyler Guinn[1,2,3]*, *Yiming Wan*[1,2], *Sarah Levovitz*[1,2], *Dongbo Yang*[4], *Marsha R. Rosner*[4] *and Gábor Balázsi*[1,2]*

[1] Biomedical Engineering Department, Stony Brook University, Stony Brook, NY, United States, [2] Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY, United States, [3] Stony Brook Medical Scientist Training Program, Stony Brook, NY, United States, [4] Ben May Department for Cancer Research, The University of Chicago, Chicago, IL, United States

Keywords: synthetic biology, non-genetic heterogeneity, noise, control, metastasis, cancer, drug resistance

OPEN ACCESS

Edited by:
Mohit Kumar Jolly,
Indian Institute of Science (IISc), India

Reviewed by:
Shensi Shen,
Institut Gustave Roussy, France
Jean Clairambault,
Institut National de Recherche en Informatique et en Automatique (INRIA), France

***Correspondence:**
Michael Tyler Guinn
mtguinn@gmail.com
orcid.org/0000-0002-7891-8765
Gábor Balázsi
gabor.balazsi@stonybrook.edu
orcid.org/0000-0002-6865-5818

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *23 July 2020*
Accepted: *29 September 2020*
Published: *30 October 2020*

Citation:
Guinn MT, Wan Y, Levovitz S, Yang D, Rosner MR and Balázsi G (2020) Observation and Control of Gene Expression Noise: Barrier Crossing Analogies Between Drug Resistance and Metastasis. Front. Genet. 11:586726. doi: 10.3389/fgene.2020.586726

INTRODUCTION

Cancer metastasis is still the main cause of death for most cancer types (Dillekås et al., 2019). The molecular causes of metastasis are diverse, complex, and poorly understood, including genetic and other molecular changes that transcend genetic sequence. Despite their complexity and diversity, a new emerging theme posits these changes generate cellular heterogeneity that can promote cancer metastasis (Fidler, 1978; Lee et al., 2014; Nguyen et al., 2016; Caswell and Swanton, 2017). Cellular heterogeneity can be genetic or non-genetic. Besides genetic mutations, non-genetic heterogeneity allows otherwise identical cells to develop drastically different phenotypes due to variations in molecular players that accumulate and compound effects over time.

How could cellular heterogeneity affect metastasis? A successfully metastasizing cell must cross multiple physical and molecular barriers: it must detach from the primary site, intravasate, survive the bloodstream or lymphatic vessels, extravasate, overcome immune attack, and start growing. Therefore, understanding how cellular heterogeneity affects barrier crossing is quintessential to understand its role in metastasis. Moreover, other barrier-crossing phenomena, such as drug resistance, may be unexpectedly informative, or even analogous to various steps in metastasis.

Non-genetic heterogeneity and biological noise are broad terms we consider synonymous here. They include variation in essentially any cellular property that is not genetic in origin, such as cell size, protein levels, cell function, and lifespan. A subtype of non-genetic heterogeneity relevant to this article is gene expression noise, which manifests as varying messenger RNA (mRNA) or protein levels in cells with identical genomes. While there are different uses of the term "noise" across the fields of biology, we define noise here as in physics and engineering disciplines besides biology: as a general stochastic process that does not exclude heritability. Indeed, the existence of noise with various spectra (1/f noise, colored noise, etc.) in nature implies that there can be various short- and long-term components of randomness.

Understanding the role of gene expression noise in metastasis can be based on two complementary investigative approaches, which are conceptual generalizations of forward and reverse genetics (Gurumurthy et al., 2016), respectively. Namely, *forward (observational) investigation* (Schuh et al., 2020; Shaffer et al., 2020) monitors and catalogs naturally occurring gene expression variability at various molecular levels in different cell types, seeking associations with cancer progression from an observational perspective. By contrast, *reverse (perturbational) investigation* (Kang et al., 2013; Nguyen et al., 2016) studies cancerous phenotypes arising upon

artificially imposed gene expression noise, using noise-controlling genetic devices, methods or chemicals (Desai et al., 2020). Here, we discuss both perspectives regarding gene expression noise and metastasis and outline how the reverse approach may be necessary due to a natural coupling between the noise and mean of gene expression, and how it may be accomplished through synthetic biology.

NON-GENETIC HETEROGENEITY AND THRESHOLD CROSSING IN CELLULAR AND MOLECULAR PROCESSES

Genetic heterogeneity among cells, tissues, and organisms has long been known to play roles in generating the phenotypic diversity that life exhibits (Nichol et al., 2019). Genetic heterogeneity can cause two main types of clonal variation in a cell population (Agozzino et al., 2020). First, coding sequence mutations generate clonal populations with protein molecules of missing or altered function (e.g., a mutant enzyme loses or improves its affinity for a substrate). Second, changes in non-coding sequence or gene copy number create clones with altered mRNA or protein levels, without any changes in protein function (e.g., the amount of unaltered enzyme molecules increases or decreases). This second type of genetic heterogeneity resembles non-genetic heterogeneity or gene expression noise, especially if the latter is heritable (i.e., cellular memory is long) (Acar et al., 2005; Nevozhay et al., 2012; Shaffer et al., 2020), since they both manifest as lasting and propagating cell–cell differences in the number of protein or RNA molecules. Indeed, non-genetic variability, like genetic mutations, can play important roles in physiological processes, disease development, and evolution (Brock et al., 2009; Balázsi et al., 2011; Chattwood and Thompson, 2011; Frank and Rosner, 2012; Bai et al., 2013).

Two key characteristics of gene expression noise are its amplitude and its memory. First, the amplitude of gene expression noise [measured by the standard deviation or coefficient of variation (CV)] characterizes how far molecule numbers can deviate from the mean. Noise amplitudes can range from slight (CV $<$ 20%) to dramatic (CV $>$ 200%), giving rise to vastly different cellular phenotypes (Kærn et al., 2005; Raj and van Oudenaarden, 2008). Second, the cellular memory of gene expression noise characterizes the heritable aspects of random cellular differences. Cellular memory is, at least conceptually, independent from the noise amplitude (Acar et al., 2005; Nevozhay et al., 2012) and can range from $<$1 cell cycle time to hundreds of cell generations (Nevozhay et al., 2012), enabling a non-genetic version of clonal expansion that affects genetic evolutionary dynamics (González et al., 2015; Bódi et al., 2017; Kheir Gouda et al., 2019).

Gene expression noise can have many sources, including biochemical reactions of transcription, translation, post-translational modifications, mRNA/protein degradation, and other cellular processes (Kærn et al., 2005; Balázsi et al., 2011). These processes affect biological noise, which can be segmented as "intrinsic" or "extrinsic" (Thattai and van Oudenaarden, 2001; Elowitz et al., 2002; Swain et al., 2002). Intrinsic noise comprises variation intrinsic to gene expression due to stochastic effects from biochemical reactions involving low copy numbers of molecular species along the central dogma of molecular biology (Ozbudak et al., 2002; Swain et al., 2002; Quarton et al., 2020). This contrasts with extrinsic noise that describes variations in more global factors affecting gene expression, such as global regulators, ribosomes, polymerases, cofactor concentrations, microenvironmental variation, and activity of other cellular players extrinsic to the gene (Swain et al., 2002; Stamatakis et al., 2011). In addition, non-genetic heterogeneity (noise) in protein levels can be preexisting or induced. Unlike tightly regulated deterministic stress programs where most cells turn on the same stress response, induced heterogeneity implies random cell–cell differences emerging upon exposure to stress (Gasch et al., 2017; Farquhar et al., 2019), where some cells turn on various forms of stress response while some other cells do not.

Both intrinsic and extrinsic noise can affect developmental and evolutionary processes. Underlying such effects are threshold-crossing cellular processes that noise can promote or suppress. Noise of drug resistance protein levels can promote microbial and cancer cell populations to cross survival thresholds and thus resist high levels of drug treatment (Blake et al., 2006; Brock et al., 2009; Fraser et al., 2009; Shaffer et al., 2017). On the other hand, noise can also hinder short-term cell survival at low levels of drug treatment (Blake et al., 2006). In addition, recent evidence indicates that preexisting or stress-induced noise can play similar dichotomous roles during long-term evolution of cell populations (Fraser et al., 2009; Marusyk et al., 2012; Farquhar et al., 2019). While less established, there are suggestions that protein noise promotes oncogenesis (Brock et al., 2009), epithelial-to-mesenchymal transition (EMT), mesenchymal–epithelial transition (MET), and the initiation of metastasis (Lee et al., 2014; Nguyen et al., 2016; George et al., 2017; Jolly and Celià-Terrassa, 2019). These findings provoke the question: Could there be common principles of non-genetic heterogeneity underlying all these biologically different, but mechanistically similar processes? For example, higher, heritable noise can cause phenotypic changes in more members of a cell population by pushing and keeping cells above thresholds (Charlebois et al., 2011) that arise from multiple sources (**Figure 1A**), such as multistability, hypersensitivity, and irreversibility, which we discuss next.

Multistability is the property of a system to permit multiple potential steady states (Gardner et al., 2000; Macía et al., 2009) (two at the minimum), such as restraint vs. commitment to sugar utilization (Novick and Weiner, 1957), oocyte maturation (Xiong and Ferrell, 2003), and stem cell differentiation (Macarthur et al., 2009), which could be imagined as valleys in a landscape (Yuan et al., 2017; Kang et al., 2019; Agozzino et al., 2020) (**Figure 1A**). Stability of any steady state implies that effects of small, temporary external perturbations decay over time, so cells will return to their valley bottoms after weak noise or transient environmental fluctuations push them slightly away. However, sufficiently large temporary perturbations can alter protein means or noise to a degree that moves cells uniformly or individually beyond the crest (separatrix) separating two valleys, causing them to fall into the neighboring valley. One common

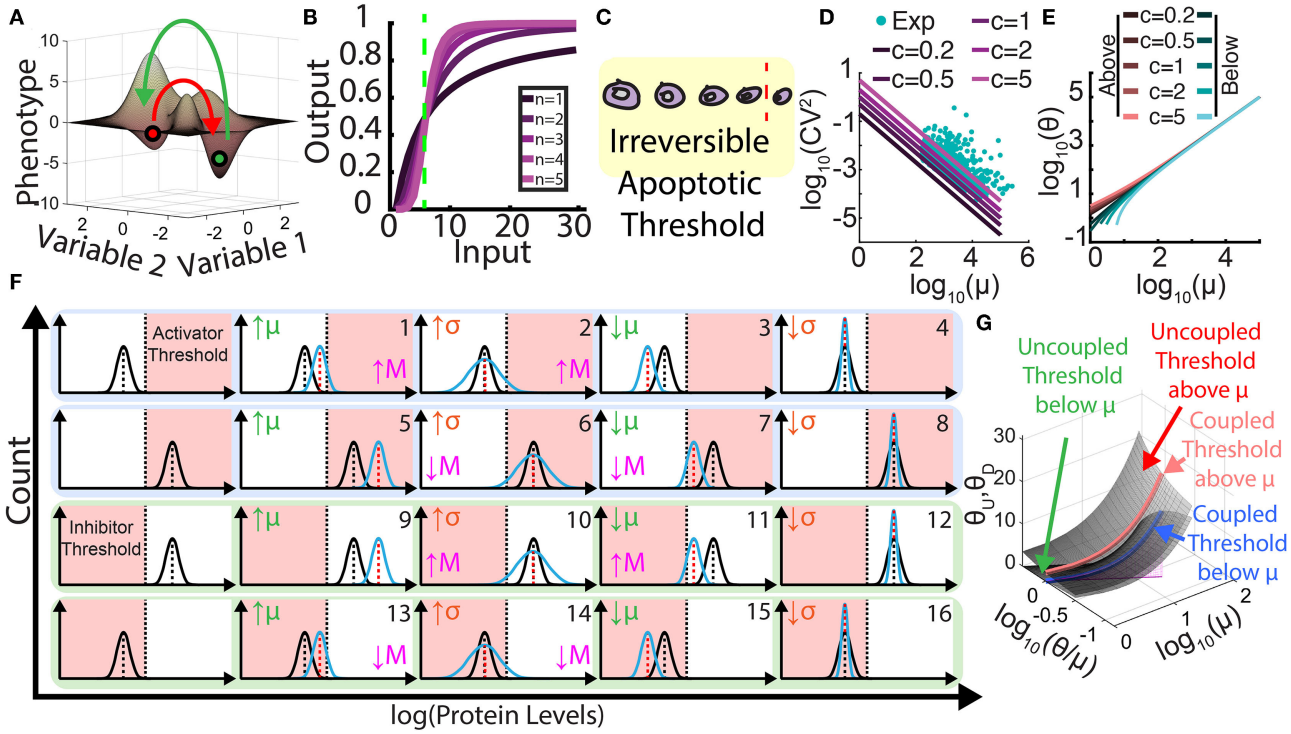

FIGURE 1 | Role of heterogeneity and its control in threshold crossing and metastasis. **(A)** Multistable landscape illustrating two steady states (phenotypes). Red arrow represents going from the first phenotype to the second across the crest (threshold), and the green arrow represents the reverse transition. **(B)** Ultrasensitivity is the abrupt change in output for small input changes, captured by Hill functions with high Hill coefficients, $n \geq 2$. Green line represents threshold where the system is most sensitive to input. **(C)** Irreversibility stems from a physical or chemical threshold that acts as a point of no return when cells surpass it. **(D)** Graph illustrating the natural tendency of gene expression noise [coefficient of variation squared (CV^2)] to decrease hyperbolically as mean levels increase: $CV^2 = c/\text{mean}$. The values of $c \neq 1$ correspond to non-Poisson behaviors, such as bursting ($c > 1$) or noise suppression ($c < 1$, e.g., by negative feedback). Teal scatter dots represent rescaled yeast experimental data (Newman et al., 2006) illustrating the natural tendency of coupling between the CV and the mean for hundreds of genes. **(E)** Graph of threshold levels θ that can be crossed as a function of mean protein levels μ, with natural mean-noise coupling for bursting parameter c. Orange colors represent downward threshold crossing (heuristically, $\theta < \mu - \sigma$) when the mean is above the threshold, whereas the blue colors represent upward threshold crossing ($\theta > \mu + \sigma$) when the mean is below the threshold. **(F)** Threshold crossing effects of independently changing the mean μ or standard deviation σ for four starting positions of cell populations relative to phenotypic thresholds. The graphs with blue background (1–8) relate to metastatic activator levels, while the graphs with green background (9–16) relate to metastatic inhibitor levels. Pink areas represent metastasis (M), while white areas represent no metastasis, with pink arrows indicating increased or decreased likelihood. **(G)** Landscapes and curves illustrating phenotypic threshold levels that can be overcome (crossed) when the noise (CV) and mean (μ) vary in a coupled or uncoupled manner, respectively. The higher surface corresponds to upward threshold crossing [top and bottom rows of **(F)**, 1–4 and 13–16], while the lower surface corresponds to downward threshold crossing [middle rows of **(F)**, 5–8, 9–12]. Thresholds crossed by naturally coupled mean and noise values are overlaid on each surface as blue and orange lines.

theme underlying natural and engineered multistability is positive feedback embedded in biomolecular networks (Gardner et al., 2000; Angeli et al., 2004; Nevozhay et al., 2012), which enable cell survival and resistance to various environmental stresses (Charlebois et al., 2011; Farquhar et al., 2019) and cancer cell transitions (Lee et al., 2014). Two examples of barrier crossing while switching steady states may be EMT and MET, which are vital processes in embryonic development, tissue repair, and cancer metastasis (Zhang et al., 2014; Nieto et al., 2016; Jolly et al., 2017, 2018, 2019; Li and Balazsi, 2018; Gómez Tejeda Zañudo et al., 2019). However, it is increasingly accepted that the phenotypic spectrum between epithelial (E) and mesenchymal (M) cell states includes one or more intermediate states, so noise may induce such intermediate cross-state transitions. Moreover, noise-induced barrier crossing may lead to coexistent E, M, and hybrid E/M phenotypes, as well as emergence of stem-like circulating tumor cells (CTCs), thereby causing collective dissemination of primary tumors (Kudo-Saito et al., 2009; Jolly et al., 2015; Grigore et al., 2016; Bocci et al., 2019a,b), variation in tumor-seeding abilities (Neelakantan et al., 2017; Grosse-Wilde et al., 2018), and differences in drug sensitivity (Creighton et al., 2009; He et al., 2019; Tièche et al., 2019). Therefore, drug resistance, full and partial EMT (Aiello et al., 2018), and metastasis may all have underlying threshold-crossing mechanisms through multistability (Lee et al., 2014; Li and Balazsi, 2018).

Ultrasensitivity is the second threshold-generating property related to sharp input–output transfer functions with switch-like characteristics (Ferrell and Ha, 2014) in monostable systems. Monostable cells are those for which mathematical models predict a single steady state. From an experimental perspective, monostable cells return to their original state (protein and

mRNA levels) upon a temporary perturbation, even if they are highly sensitive and thus deviate far. Ultrasensitivity leads to abrupt, large cellular responses to small, persistent input differences within a narrow input range (**Figure 1B**). For example, monostable cells can be ultrasensitive when their response to an internal or external factor is sigmoidal. For a system exhibiting ultrasensitivity, a threshold can be defined as the stimulus level (i.e., metabolite, protein, or cofactor concentration) at which the system is maximally sensitive (Louis and Becskei, 2002; Zhang et al., 2013).

The last mechanism of threshold generation is irreversibility, whereby external factors restrict or block the reversion of cellular processes, as in embryonic development (Caplan and Ordahl, 1978). Both monostable and multistable cells may approach physical or biochemical barriers that, once crossed, prevent reversion causing permanent outcomes. Such irreversibility can stabilize new cellular states. Typical examples are commencement of DNA synthesis, cell lysis, and apoptosis, which, after a certain progression, cannot revert (**Figure 1C**). Likewise, when a tumor cell enters the bloodstream and travels away, it is very unlikely to return to its original site (Scott et al., 2013), and even if it does, it will be already altered due to its time in a different environment.

Each of these three threshold-generating mechanisms can produce biological consequences in populations of cells ranging from cell division and neuron depolarization to apoptosis when a certain threshold is exceeded (Mateo et al., 2011; Xie et al., 2011; Sato et al., 2013). A common conceptual way to connect cell population phenotypes, such as drug resistance, oncogenesis, and metastasis to single-cell behaviors is through noise-modulated threshold crossing. However, altering the mean can also move cell populations closer to or farther from thresholds without any change in the noise. Therefore, investigating the role of noise in such processes requires fixing the mean. Yet, in natural systems, the mean and noise of protein levels have a tendency to be coupled (Newman et al., 2006; Dar et al., 2016), where higher means often associate with lower noise, along a hyperbolic ($CV^2 = c/\text{mean}$) interdependence. Typically, this relationship holds for various bursting regimes ($c > 1$ in **Figure 1D**). From an observational standpoint, naturally occurring mutations or other changes make it rarely possible (Dar et al., 2014; You et al., 2019) to parse out specifically how the noise of a single protein affects threshold crossing and phenotypes independently of the mean. For example, **Figure 1E** shows the threshold levels that a given protein can naturally overcome based on the experimentally demonstrated inverse relationship (Newman et al., 2006; Dar et al., 2016) between its expression mean and noise (**Figure 1D**). Overall, due to their natural coupling tendency, both the mean and noise will change and affect threshold crossing. To decipher their individual impact on biological phenotypes, we need engineering approaches to independently control the protein noise and the mean because observing natural decoupling scenarios is far from trivial. In the next section, we explore ways of decoupling means and noise, with implications on drug resistance and metastasis.

CONTROLLING NON-GENETIC HETEROGENEITY OF METASTASIS THROUGH SYNTHETIC BIOLOGICAL GENE CIRCUITS

To confirm observational suggestions on the role of non-genetic heterogeneity in disease development, stress survival, and metastasis, one must control protein noise independently of the mean in living organisms. Engineering approaches from the field of synthetic biology (Elowitz and Leibler, 2000; Gardner et al., 2000) enabled the inception of such control (Blake et al., 2006; Fraser et al., 2009), followed by identification of many methods that reduce or amplify gene expression noise (Maamar et al., 2007; Cagatay et al., 2009; Nevozhay et al., 2013; Shimoga et al., 2013; Farquhar et al., 2019; Guinn and Balázsi, 2019). In natural systems, a change in a protein's mean will often change the noise since the two parameters tend to be coupled (**Figure 1D**). Therefore, biological threshold crossing typically does not utilize the mean and noise as two fully independent degrees of freedom (**Figure 1E**). As opposed to natural mechanisms, synthetic biological systems can allow independent changes in the mean and noise such that they are no longer coupled. Synthetic gene circuits thus provide an increasing number of ways to allow noise-mean decoupling (Aranda-Díaz et al., 2017; Farquhar et al., 2019) in studies of noise-modulated phenotypic transitions.

Two main approaches have accomplished decoupling the gene expression mean and noise from one another in synthetic gene circuits: (i) different gene circuits to express the same gene with different noise-vs.-mean dependencies (Blake et al., 2006; Süel et al., 2006; Kim and Sauro, 2010; Farquhar et al., 2019) and (ii) combinatorial induction of cascaded modules within the same gene circuit (Aranda-Díaz et al., 2017). Decoupling the noise and mean from one another adds a new degree of freedom to tune threshold crossing and reveal individual contributions of mean and noise on cellular processes, such as metastasis.

As an illustrative classification inspired by studies examining how mean and noise affect drug resistance (Blake et al., 2006; Farquhar et al., 2019), we consider a threshold for a metastasis activator (or a threshold for a metastasis inhibitor). Bistability, ultrasensitivity, and irreversibility underlying such thresholds are frequent themes in metastasis and EMT (Lee et al., 2014; Zhang et al., 2014; Jolly et al., 2015). Based on the position of the threshold (above or below) relative to the mean, and the ability to tune the mean and variance independently up and down, there are 16 possible phenotypic scenarios (**Figure 1F**). To start, assume metastasis activator levels are below a phenotypic threshold when the cells are not metastatic (**Figure 1F**, 1–4). In such a scenario, both the mean and variance of activator levels can be tuned up or down. Two of these changes (tuning mean or variance up) should promote metastasis by enhanced threshold crossing, while the two opposite changes (tuning mean or variance down) should hinder metastasis. Alternatively, for a metastatic cell population with hyperthreshold activator distribution (**Figure 1F**, 5–8), tuning the mean down or variance up should hinder metastasis,

while the opposite changes should promote metastasis. Overall, for underthreshold populations, elevated activator mean and variance consistently aid overcoming the threshold. In contrast, for hyperthreshold populations, elevated activator variance still aids, but elevated mean hinders threshold crossing. Analogously, non-metastatic and metastatic cell populations could also have metastasis inhibitor levels above or below a phenotypic threshold, respectively (**Figure 1F**, 9–16). Tuning the variance and mean of a metastasis suppressor up or down will affect threshold crossing according to the above principles but with opposite phenotypic effects. Similar reasoning can predict the effect of simultaneous changes in the mean and the noise, although the scenarios can be numerous and complicated.

Protein mean- and noise-dependent crossings of various thresholds in synthetic and natural scenarios can also be explored visually as landscapes (**Figure 1G**). Synthetic biological control can explore surfaces with two degrees of freedom, whereas natural coupling restricts movement to a single degree of freedom along paths on such surfaces. These landscapes and curves illustrate what phenotypic threshold levels can be crossed when the noise and mean are coupled or uncoupled, respectively. In the future, it will be important to examine similar but multimolecular threshold crossings in higher dimensions, or the joint effects of the mean, the noise, as well as of the higher moments (i.e., skewness, kurtosis, etc.). Moreover, specific fitness landscapes (Nevozhay et al., 2012; González et al., 2015) will need to replace threshold approximations in many realistic scenarios. Synthetic biological tools will be indispensable for addressing these future questions.

Experimental evidence for how gene expression mean and noise independently affect metastasis threshold crossing is relatively lacking, but synthetic biology is already shedding light on drug resistance (Farquhar et al., 2019), which may give insights to the role of noise in metastasis as an evolutionary process. According to the threshold-crossing principles, high noise aids drug resistance evolution when stress is high but hampers survival when stress is low, mimicking the effects of noise on short-term survival (Blake et al., 2006; Fraser et al., 2009; Farquhar et al., 2019). Analogously, high noise should facilitate metastasis initiation for pre-metastatic cells before dissemination. In contrast, high noise may hinder the rate of metastasis for cells that have already acquired invasive characteristics. These findings relate to where cells sit (below or above) relative to thresholds at which specific phenotypes emerge. We anticipate that utilizing noise amplifying or reducing gene circuits (Becskei et al., 2001; Hooshangi et al., 2005; Weinberger et al., 2005; Tan et al., 2009; Diao et al., 2016; Farquhar et al., 2019) will give similar insights into processes underlying metastasis. Using the growing repertoire of noise-controlling synthetic biology tools and chemicals will certainly uncover unknown roles of gene expression in processes of full or partial EMT (Aiello et al., 2018), metastasis, and oncogenesis.

AUTHOR CONTRIBUTIONS

MG, YW, and GB conceived and wrote the manuscript. SL surveyed the literature and added the references. DY and MR contributed some critical concepts and feedback on the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Institutes of Health, NIGMS MIRA Program R35 GM122561 (to GB), R01 GM121735-01 (to MR), the Rustandy fund for Innovative Cancer Research (to MR), a National Defense Science and Engineering Graduate (NDSEG) Fellowship (to MG), and by the Laufer Center for Physical and Quantitative Biology (to GB).

REFERENCES

Acar, M., Becskei, A., and van Oudenaarden, A. (2005). Enhancement of cellular memory by reducing stochastic transitions. *Nature* 435, 228–232. doi: 10.1038/nature03524

Agozzino, L., Balázsi, G., Wang, J., and Dill, K. A. (2020). How do cells adapt? Stories told in landscapes. *Ann. Rev. Chem. Biomol. Eng.* 11, 155–182. doi: 10.1146/annurev-chembioeng-011720-103410

Aiello, N. M., Maddipati, R., Norgard, R. J., Balli, D., Li, J., Yuan, S., et al. (2018). EMT subtype influences epithelial plasticity and mode of cell migration. *Dev. Cell* 45, 681–695.e684. doi: 10.1016/j.devcel.2018.05.027

Angeli, D., Ferrell, J. E., and Sontag, E. D.Angeli, D., Ferrell, J. E., and Sontag, E. D. Jr. (2004). Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl. Acad. Sci. U.S.A.* 101, 1822–1827. doi: 10.1073/pnas.0308265100

Aranda-Díaz, A., Mace, K., Zuleta, I., Harrigan, P., and El-Samad, H. (2017). Robust synthetic circuits for two-dimensional control of gene expression in yeast. *ACS Synth. Biol.* 6, 545–554. doi: 10.1021/acssynbio.6b00251

Bai, F., Che, Y.-S., Kami-ike, N., Ma, Q., Minamino, T., Sowa, Y., et al. (2013). Populational heterogeneity vs. temporal fluctuation in Escherichia coli flagellar motor switching. *Biophys J.* 105, 2123–2129. doi: 10.1016/j.bpj.2013.09.043

Balázsi, G., van Oudenaarden, A., and Collins, J. J. (2011). Cellular decision making and biological noise: from microbes to mammals. *Cell* 144, 910–925. doi: 10.1016/j.cell.2011.01.030

Becskei, A., Séraphin, B., and Serrano, L. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* 20, 2528–2535. doi: 10.1093/emboj/20.10.2528

Blake, W. J., Balázsi, G., Kohanski, M. A., Isaacs, F. J., Murphy, K. F., Kuang, Y., et al.Blake, W. J., et al. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell.* 24, 853–865. doi: 10.1016/j.molcel.2006.11.003

Bocci, F., Gearhart-Serna, L., Boareto, M., Ribeiro, M., Ben-Jacob, E., Devi, G. R., et al. (2019a). Toward understanding cancer stem cell heterogeneity in the tumor microenvironment. *Proc. Natl. Acad. Sci. U.S.A.* 116, 148–157. doi: 10.1073/pnas.1815345116

Bocci, F., Levine, H., Onuchic, J., and Jolly, M. K. (2019b). Deciphering the dynamics of epithelial-mesenchymal transition and cancer stem cells in tumor progression. *Curr. Stem. Cell Rep.* 5, 1–11. doi: 10.1007/s40778-019-0150-3

Bódi, Z., Farkas, Z., Nevozhay, D., Kalapis, D., Lázár, V., Csörgo, B., et al. (2017). Phenotypic heterogeneity promotes adaptive evolution. *PLoS Biol.* 15:e2000644. doi: 10.1371/journal.pbio.2000644

Brock, A., Chang, H., and Huang, S. (2009). Non-genetic heterogeneity–a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.* 10, 336–342. doi: 10.1038/nrg2556

Cagatay, T., Turcotte, M., Elowitz, M. B., Garcia-Ojalvo, J., and Süel, G. M. (2009). Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell* 139, 512–522. doi: 10.1016/j.cell.2009.07.046

Caplan, A. I., and Ordahl, C. P. (1978). Irreversible gene repression model for control of development. *Science* 201:120. doi: 10.1126/science.351805

Caswell, D. R., and Swanton, C. (2017). The role of tumour heterogeneity and clonal cooperativity in metastasis, immune evasion and clinical outcome. *BMC Med.* 15:133. doi: 10.1186/s12916-017-0900-y

Charlebois, D. A., Abdennur, N., and Kaern, M. (2011). Gene expression noise facilitates adaptation and drug resistance independently of mutation. *Phys. Rev. Lett.* 107:218101. doi: 10.1103/PhysRevLett.107.218101

Chattwood, A., and Thompson, C. R. (2011). Non-genetic heterogeneity and cell fate choice in *Dictyostelium discoideum*. *Dev. Growth Differ.* 53, 558–566. doi: 10.1111/j.1440-169X.2011.01270.x

Creighton, C. J., Li, X., Landis, M., Dixon, J. M., Neumeister, V. M., Sjolund, A., et al. (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13820–13825. doi: 10.1073/pnas.0905718106

Dar, R. D., Hosmane, N. N., Arkin, M. R., Siliciano, R. F., and Weinberger, L. S. (2014). Screening for noise in gene expression identifies drug synergies. *Science* 344, 1392–1396. doi: 10.1126/science.1250220

Dar, R. D., Shaffer, S. M., Singh, A., Razooky, B. S., Simpson, M. L., Raj, A., et al. (2016). Transcriptional bursting explains the noise-versus-mean relationship in mRNA and protein levels. *PLoS ONE.* 11:e0158298. doi: 10.1371/journal.pone.0158298

Desai, R. V., Hansen, M. M. K., Martin, B., Yu, C., Ding, S., Thomson, M., et al. (2020). Discovery of a cellular mechanism regulating transcriptional noise. *bioRxiv* 2020.2006.2029.128439v128431.

Diao, J., Charlebois, D. A., Nevozhay, D., Bódi, Z., Pál, C., and Balázsi, G. (2016). Efflux pump control alters synthetic gene circuit function. *ACS Synth. Biol.* 5, 619–631. doi: 10.1021/acssynbio.5b00154

Dillekås, H., Rogers, M. S., and Straume, O. (2019). Are 90% of deaths from cancer caused by metastases? *Cancer Med.* 8, 5574–5576 doi: 10.1002/cam4.2474

Elowitz, M. B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338. doi: 10.1038/35002125

Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science* 297:1183. doi: 10.1126/science.1070919

Farquhar, K. S., Charlebois, D. A., Szenk, M., Cohen, J., Nevozhay, D., and Balázsi, G. (2019). Role of network-mediated stochasticity in mammalian drug resistance. *Nat. Commun.* 10:2766. doi: 10.1038/s41467-019-10330-w

Ferrell, J. E., and Ha, S. H. (2014). Ultrasensitivity part I: michaelian responses and zero-order ultrasensitivity. *Trends Biochem. Sci.* 39, 496–503. doi: 10.1016/j.tibs.2014.08.003

Fidler, I. J. (1978). Tumor heterogeneity and the biology of cancer invasion and metastasis. *Cancer Res.* 38, 2651–2660.

Frank, S. A., and Rosner, M. R. (2012). Nonheritable cellular variability accelerates the evolutionary processes of cancer. *PLoS Biol.* 10:e1001296. doi: 10.1371/journal.pbio.1001296

Fraser, D., and Kaern, M.Fraser, D., and Kærn, M. (2009). A chance at survival: gene expression noise and phenotypic diversification strategies. *Mol. Microbiol.* 71, 1333–1340. doi: 10.1111/j.1365-2958.2009.06605.x

Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342. doi: 10.1038/35002131

Gasch, A. P., Yu, F. B., Hose, J., Escalante, L. E., Place, M., Bacher, R., et al. (2017). Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. *PLoS Biol.* 15:e2004050. doi: 10.1371/journal.pbio.2004050

George, J. T., Jolly, M. K., Xu, S., Somarelli, J. A., and Levine, H. (2017). Survival outcomes in cancer patients predicted by a partial emt gene expression scoring metric. *Cancer Res.* 77, 6415–6428. doi: 10.1158/0008-5472.CAN-16-3521

Gómez Tejeda Zañudo, J., Guinn, M. T., Farquhar, K., Szenk, M., Steinway, S. N., Balázsi, G., et al. (2019). Towards control of cellular decision-making networks in the epithelial-to-mesenchymal transition. *Phys. Biol.* 16, 031002–031002. doi: 10.1088/1478-3975/aaffa1

González, C., Ray, J. C., Manhart, M., Adams, R. M., Nevozhay, D., Morozov, A. V., et al. (2015). Stress-response balance drives the evolution of a network module and its host genome. *Mol. Syst. Biol.* 11:827. doi: 10.15252/msb.20156185

Grigore, A. D., Jolly, M. K., Jia, D., Farach-Carson, M. C., and Levine, H. (2016). Tumor budding: the name is EMT. Partial EMT. *J. Clin. Med.* 5:51. doi: 10.3390/jcm5050051

Grosse-Wilde, A., Kuestner, R. E., Skelton, S. M., MacIntosh, E., d'Hérouël, A. F., Ertaylan, G., et al. (2018). Loss of inter-cellular cooperation by complete epithelial-mesenchymal transition supports favorable outcomes in basal breast cancer patients. *Oncotarget* 9, 20018–20033. doi: 10.18632/oncotarget.25034

Guinn, M. T., and Balázsi, G. (2019). Noise-reducing optogenetic negative-feedback gene circuits in human cells. *Nucleic Acids Res.* 47, 7703–7714. doi: 10.1093/nar/gkz556

Gurumurthy, C. B., Grati, M., Ohtsuka, M., Schilit, S. L., Quadros, R. M., and Liu, X. Z. (2016). CRISPR: a versatile tool for both forward and reverse genetics research. *Hum. Genet.* 135, 971–976. doi: 10.1007/s00439-016-1704-4

He, J., Lee, H. J., Saha, S., Ruan, D., Guo, H., and Chan, C. H.He, J., et al. (2019). Inhibition of USP2 eliminates cancer stem cells and enhances TNBC responsiveness to chemotherapy. *Cell Death Dis.* 10:285. doi: 10.1038/s41419-019-1512-6

Hooshangi, S., Thiberge, S., and Weiss, R. (2005). Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl. Acad. Sci. U.S.A.* 102:3581. doi: 10.1073/pnas.0408507102

Jolly, M. K., Boareto, M., Huang, B., Jia, D., Lu, M., Ben-Jacob, E., et al.Jolly, M. K., et al. (2015). Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. *Front. Oncol.* 5:155. doi: 10.3389/fonc.2015.00155

Jolly, M. K., and Celià-Terrassa, T. (2019). Dynamics of phenotypic heterogeneity associated with EMT and stemness during cancer progression. *J. Clin. Med.* 8:1542. doi: 10.3390/jcm8101542

Jolly, M. K., Mani, S. A., and Levine, H. (2018). Hybrid epithelial/mesenchymal phenotype(s): the 'fittest' for metastasis? *Biochim. Biophys. Acta Rev. Cancer* 1870, 151–157. doi: 10.1016/j.bbcan.2018.07.001

Jolly, M. K., Somarelli, J. A., Sheth, M., Biddle, A., Tripathi, S. C., Armstrong, A. J., et al. (2019). Hybrid epithelial/mesenchymal phenotypes promote metastasis and therapy resistance across carcinomas. *Pharmacol. Ther.* 194, 161–184. doi: 10.1016/j.pharmthera.2018.09.007

Jolly, M. K., Ware, K. E., Gilja, S., Somarelli, J. A., and Levine, H. (2017). EMT and MET: necessary or permissive for metastasis? *Mol. Oncol.* 11, 755–769. doi: 10.1002/1878-0261.12083

Kærn, M., Elston, T. C., Blake, W. J., and Collins, J. J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* 6, 451–464. doi: 10.1038/nrg1615

Kang, T., White, J. T., Xie, Z., Benenson, Y., Sontag, E., and Bleris, L. (2013). Reverse engineering validation using a benchmark synthetic gene circuit in human cells. *ACS Synth. Biol.* 2, 255–262. doi: 10.1021/sb300093y

Kang, X., Wang, J., and Li, C. (2019). Exposing the underlying relationship of cancer metastasis to metabolism and epithelial-mesenchymal transitions. *iScience* 21, 754–772. doi: 10.1016/j.isci.2019.10.060

Kheir Gouda, M., Manhart, M., and Balázsi, G. (2019). Evolutionary regain of lost gene circuit function. *Proc. Natl. Acad. Sci. U.S.A.* 116, 25162–25171. doi: 10.1073/pnas.1912257116

Kim, K. H., and Sauro, H. M. (2010). Fan-out in gene regulatory networks. *J. Biol. Eng.* 4:16. doi: 10.1186/1754-1611-4-16

Kudo-Saito, C., Shirako, H., Takeuchi, T., and Kawakami, Y. (2009). Cancer metastasis is accelerated through immunosuppression during Snail-induced EMT of cancer cells. *Cancer Cell.* 15, 195–206. doi: 10.1016/j.ccr.2009.01.023

Lee, J., Lee, J., Farquhar, K. S., Yun, J., Frankenberger, C. A., Bevilacqua, E., et al. (2014). Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. *Proc. Natl. Acad. Sci. U.S.A.* 111, E364–E373. doi: 10.1073/pnas.1304840111

Li, C., and Balazsi, G. (2018). A landscape view on the interplay between EMT and cancer metastasis. *NPJ Syst. Biol Appl.* 4:34. doi: 10.1038/s41540-018-0068-x

Louis, M., and Becskei, A. (2002). Binary and graded responses in gene networks. *Sci. STKE* 2002:pe33. doi: 10.1126/stke.2002.143.pe33

Maamar, H., Raj, A., and Dubnau, D. (2007). Noise in gene expression determines cell fate in bacillus subtilis. *Science* 317:526. doi: 10.1126/science.1140818

Macarthur, B. D., Ma'ayan, A., and Lemischka, I. R. (2009). Systems biology of stem cell fate and cellular reprogramming. *Nat. Rev. Mol. Cell Biol.* 10, 672–681. doi: 10.1038/nrm2766

Macía, J., Widder, S., and Solé, R. (2009). Why are cellular switches Boolean? General conditions for multistable genetic circuits. *J. Theor. Biol.* 261, 126–135. doi: 10.1016/j.jtbi.2009.07.019

Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* 12, 323–334. doi: 10.1038/nrc3261

Mateo, C., Avermann, M., Gentet, L. J., Zhang, F., Deisseroth, K., and Petersen, C. C. (2011). *In vivo* optogenetic stimulation of neocortical excitatory neurons drives brain-state-dependent inhibition. *Curr. Biol.* 21, 1593–1602. doi: 10.1016/j.cub.2011.08.028

Neelakantan, D., Zhou, H., Oliphant, M. U. J., Zhang, X, Simon, L. M., Henke, D. M., (2017). EMT cells increase breast cancer metastasis via paracrine GLI activation in neighbouring tumour cells. *Nat. Commun.* 8:15773. doi: 10.1038/ncomms15773

Nevozhay, D., Adams, R. M., Van Itallie, E., Bennett, M. R., and Balázsi, G. (2012). Mapping the environmental fitness landscape of a synthetic gene circuit. *PLoS Comput. Biol.* 8:e1002480. doi: 10.1371/journal.pcbi.1002480

Nevozhay, D., Zal, T., and Balázsi, G. (2013). Transferring a synthetic gene circuit from yeast to mammalian cells. *Nat. Commun.* 4:1451. doi: 10.1038/ncomms2471

Newman, J. R., Ghaemmaghami, S., Ihmels, S., Breslow, D. K., Noble, M., et al. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441, 840–846. doi: 10.1038/nature04785

Nguyen, A., Yoshida, M., Goodarzi, H., and Tavazoie, S. F. (2016). Highly variable cancer subpopulations that exhibit enhanced transcriptome variability and metastatic fitness. *Nat. Commun.* 7:11246. doi: 10.1038/ncomms11246

Nichol, D., Robertson-Tessi, M., Anderson, A. R. A., and Jeavons, P. (2019). Model genotype-phenotype mappings and the algorithmic structure of evolution. *J. R. Soc. Interf.* 16:20190332. doi: 10.1098/rsif.2019.0332

Nieto, M. A., Huang, R. Y., Jackson, R. A., and Thiery, J. P. (2016). EMT: 2016. *Cell.* 166, 21–45 doi: 10.1016/j.cell.2016.06.028

Novick, A., and Weiner, M. (1957). Enzyme induction as an all-or-none phenomenon. *Proc. Natl. Acad. Sci. U.S.A.* 43, 553–566. doi: 10.1073/pnas.43.7.553

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., and van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nat. Genet.* 31, 69–73. doi: 10.1038/ng869

Quarton, T., Kang, T., Papakis, V., Nguyen, K., Nowak, C., Li, Y., et al. (2020). Uncoupling gene expression noise along the central dogma using genome engineered human cell lines. *Nucleic Acids Res.* 48, 9406–9413. doi: 10.1093/nar/gkaa668

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216–226. doi: 10.1016/j.cell.2008.09.050

Sato, Y., Yasuhara, K., Kikuchi, J., and Sato, T. N. (2013). Synthetic cell division system: controlling equal vs. unequal divisions by design. *Sci Rep.* 3:3475. doi: 10.1038/srep03475

Schuh, L., Saint-Antoine, M., Sanford, E. M., Emert, B. L., Singh, A., Marr, C., et al. (2020). Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer. *Cell Syst.* 10, 363–378 e312. doi: 10.1016/j.cels.2020.03.004

Scott, J. G., Basanta, D., Anderson, A. R., and Gerlee, P. (2013). A mathematical model of tumour self-seeding reveals secondary metastatic deposits as drivers of primary tumour growth. *J. R. Soc. Interface.* 10:20130011. doi: 10.1098/rsif.2013.0011

Shaffer, S. M., Dunagin, M. C., Torborg, S. R., Torre, E. A., Emert, B., Krepler, C., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* 546, 431–435. doi: 10.1038/nature22794

Shaffer, S. M., Emert, B. L., Reyes Hueros, R. A., Cote, C., Harmange, G., Schaff, D. L., et al. (2020). Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors. *Cell* 182, 947–959 e917. doi: 10.1016/j.cell.2020.07.003

Shimoga, V., White, J. T., Li, Y., Sontag, E., and Bleris, L. (2013). Synthetic mammalian transgene negative autoregulation. *Mol. Syst. Biol.* 9, 670–670. doi: 10.1038/msb.2013.27

Stamatakis, M., Adams, R. M., and Balázsi, G. (2011). A common repressor pool results in indeterminacy of extrinsic noise. *Chaos* 21:047523. doi: 10.1063/1.3658618

Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M., and Elowitz, M. B. (2006). An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440, 545–550. doi: 10.1038/nature04588

Swain, P. S., Elowitz, M. B., and Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 99:12795. doi: 10.1073/pnas.162041399

Tan, C., Marguet, P., and You, L. (2009). Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol.* 5, 842–848. doi: 10.1038/nchembio.218

Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* 98:8614. doi: 10.1073/pnas.151588598

Tièche, C. C., Gao, Y., Bührer, E. D., Hobi, N., Berezowska, S. A., Wyler, K., et al. (2019). Tumor initiation capacity and therapy resistance are differential features of EMT-related subpopulations in the NSCLC cell line A549. *Neoplasia.* 21, 185–196. doi: 10.1016/j.neo.2018.09.008

Weinberger, L. S., Burnett, J. C., Toettcher, J. E., Arkin, A. P., and Schaffer, D. V. (2005). Stochastic gene expression in a lentiviral positive-feedback loop: hiv-1 tat fluctuations drive phenotypic diversity. *Cell* 122, 169–182. doi: 10.1016/j.cell.2005.06.006

Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R., and Benenson, Y. (2011). Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* 333, 1307–1311. doi: 10.1126/science.1205527

Xiong, W., and Ferrell, J. E. (2003). A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* 426, 460–465. doi: 10.1038/nature02089

You, S. T., Jhou, Y. T., Kao, C. F., and Leu, J. Y. (2019). Experimental evolution reveals a general role for the methyltransferase Hmt1 in noise buffering. *PLoS Biol.* 17:e3000433. doi: 10.1371/journal.pbio.3000433

Yuan, R., Zhu, X., Wang, G., Li, S., Ao, P., Yuan, R., et al. (2017). Cancer as robust intrinsic state shaped by evolution: a key issues review. *Rep. Prog. Phys.* 80:042701. doi: 10.1088/1361-6633/aa538e

Zhang, J., Tian, X. J., Zhang, H., Teng, Y., Li, R., Bai, F., et al. (2014). TGF-beta-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Sci Signal.* 7:ra91. doi: 10.1126/scisignal.2005304

Zhang, Q., Bhattacharya, S., and Andersen, M. E. (2013). ltrasensitive response motifs: basic amplifiers in molecular signalling networks. *Open Biol.* 3:130031. doi: 10.1098/rsob.130031

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.

frontiers in Genetics

OPINION
published: 19 November 2020
doi: 10.3389/fgene.2020.579738

Stepping From Modeling Cancer Plasticity to the Philosophy of Cancer

Jean Clairambault [1,2]*

[1] Laboratoire Jacques-Louis Lions, BC 187, Sorbonne Université, Paris, France, [2] Inria, Paris, France

Keywords: philosophy, therapeutics, evolution, multicellularity, cancer

1. INTRODUCTION

Coherent multicellular organisms are not only cohesive from a spatial and anatomical point of view but also coherent from the phenotypic and cell-functional point of view of compatibility, cooperativity and division of tasks between cells and tissues. This is mandatory to make possible the achievement of a stable, functional, and reproductive whole.

Leaving aside the possibility of spontaneous "emergence of order from chaos," this article proposes a simpler hypothesis of a system of communication ways between trees of differentiation, relying on the control of transcription factors that determine differentiation: "the cohesion watch." These are considered a part of the immune system, whose armed force is the immune response, innate as well as adaptive, humoral and cellular, but is not the whole of it. Indeed, this paper poses that the immune system is the coordinator of the unity of the organism. Within the immune system in this extended vision that is thus more general than the immune response, the cohesion watch is in charge of the control of compatibilities and cooperation between the anatomical and the phenotypic/cell-functional systems as well as within each of these systems. It is a mandatory component of multicellularity that ultimately leads to an anatomically cohesive and functionally coherent organism.

The immune system in this extended sense should thus comprise the following: (a) the equivalent in all Metazoans of the major histocompatibility complex (MHC) of jawed vertebrates, in charge of characterizing all cells of a given individual within its species (this article postulates the existence in all Metazoans of a coding system analogous to the MHC of jawed vertebrates that is present in all its forerunners in animal evolution); (b) the immune response; and (c) the cohesion watch. The latter is here assumed to be a complementary histocompatibility complex in charge of the following:

- driving indifferently (i.e., in an equal way) in each organism-to-be in a given species the body plan (or Bauplan, described with regulatory mechanisms in Davidson et al., 1995 for bilaterians and for older animals in evolution in Müller et al., 2004) by launching morphogens during embryonic development;
- simultaneous with the indifference, guiding from pluripotent stem cells attached to the body plan the development of the trees of cell differentiation; and
- establishing between the trees and twigs of differentiation that stem from the body plan acellular compatibility regulatory mechanism, the main compatibility task of the cohesion watch.

These trees, in the Waddington view (inverse, in its three-dimensional presentation, of the tree expansion metaphor), are none else than the epigenetic landscapes, and they are controlled by transcription factors and epigenetic enzymes. All differentiations lead terminally to mature cell types, between 200, and say, 400 (according to various evaluations) (Jacob, 1977) in the human species but only 20 in the sponges Porifera (Müller, 2001). They are in any event of a fixed number for every organism in a given species. Inscribed, much like the body plan and the functionality differentiation trees, in the genome of each cell, this cohesion watch should manifest itself materially

OPEN ACCESS

Edited by:
Mohit Kumar Jolly,
Indian Institute of Science (IISc), India

Reviewed by:
Antoni Celià-Terrassa,
Mar Institute of Medical Research (IMIM), Spain
David Jordan Wooten,
Pennsylvania State University (PSU), United States
Charles Lineweaver,
Australian National University, Australia

***Correspondence:**
Jean Clairambault
jean.clairambault@inria.fr

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *03 July 2020*
Accepted: *03 October 2020*
Published: *19 November 2020*

Citation:
Clairambault J (2020) Stepping From Modeling Cancer Plasticity to the Philosophy of Cancer. Front. Genet. 11:579738. doi: 10.3389/fgene.2020.579738

during development as a net of communication between and within differentiation trees. At the chromatin level, it should control the non-expression (in closed-state chromatin), expression or repression (in open-state chromatin) of genes at nodes in the cell differentiation trees, and it should also control the stability of the body plan.

2. MODELING PLASTICITY IN CANCER CELL POPULATIONS

In a series of papers starting in 2013, a team of mathematicians (the author included) at Laboratoire Jacques-Louis Lions, Sorbonne University, Paris, and some followers elsewhere were initially stimulated by an article published in 2010 (Sharma et al., 2010) that reported reversible drug resistance in a cancer cell culture. The culture, exposed to massive doses of drugs, developed in sparsely distributed resistant subpopulations (named persisters), and such resistance, shown to be of a non-genetic nature, was completely reversed when the drug was withdrawn from the culture. Driven by this biological observation of reversible resistance in cancer cells, we tackled the question of understanding and predicting the dynamics of these cancer cell populations by using mathematical models. The behavior of these highly *plastic* cell populations was relevantly described by phenotype-structured partial differential equations. In these equations, the structuring variable, i.e., the parameter-like one that codes for the biological variability of interest, is assumed to store the *heterogeneity* of the cell population with respect to the expression of drug resistance. It was chosen to be a positive real variable representing the expression of a resistance phenotype *continuously* from 0 (totally sensitive) to 1 (totally resistant) (Perthame, 2007, 2015; Lavi et al., 2013; Lorz et al., 2013, 2015; Chisholm et al., 2015, 2016a,b; Lorenzi et al., 2016; Almeida et al., 2018, 2019; Cho and Levy, 2018a,b; Clairambault, 2019; Clairambault and Pouchol, 2019; Nguyen et al., 2019).

These models, intended to represent the effects of a cancer treatment on cell populations, and ultimately on patients, with the aim to overcome their capacities of resistance induced by the treatment itself, naturally give rise to the proposal of theoretically optimized therapeutic strategies. Such strategies, which have recently been the object of active research (reviewed in Jarrett et al., 2020) aim to contain or eradicate cancer growth, avoiding the two major pitfalls of treatments in clinical oncology, namely unwanted toxic side effects in healthy cell populations and the emergence of resistance in cancer populations (Pouchol, 2018; Pouchol et al., 2018). These strategies, which are still theoretical, may be too recent to be widely accepted by oncologists as plausibly efficacious and challenged by preliminary experiments, in Petri dishes or in laboratory rodents. In the meantime, many questions arise about the nature of plasticity in cancer cells and cancer cell populations that underly them (reviewed in Shen and Clairambault, 2020). Cancer is a disease of multicellular organisms that are normally functionally constituted of terminally and irreversibly differentiated cells. We have made progress in understanding the causes and the mechanisms of the reversion of differentiations that make cancer cells so plastic. Are we are able to quickly adapt their phenotypes to a changing environment, such as deadly drug pressure, while healthy cells cannot?

3. QUESTIONS ABOUT MULTICELLULARITY AND CANCER

Some motivations for the interest of stepping away from such therapeutically oriented models of drug resistance in cancer cell populations, and thus of plasticity in cancer, to more general considerations can be seen as arising from observed facts. Such questions are most of the time dodged, likely being perceived as too complex to be solved by specialists of one domain only, in the field of cancer biology:

- Cancer can be found throughout the animal kingdom (Aktipis et al., 2015) and beyond, but plants, however, are not lethally affected by it (Doonan and Sablowski, 2010); investigating the earliest stages of multicellularity in animals (Müller, 2001; Müller and Müller, 2003; Müller et al., 2004), i.e., searching for its failures, may therefore be a natural way to understand how some somatic cells become cheaters to their established multicellular community.
- The genes that are altered in cancers are the same that serve a multicellularity design (Domazet-Lošo and Tautz, 2008, 2010; Davies and Lineweaver, 2011; Vincent, 2011a,b; Lineweaver et al., 2014); can we methodically collect these genes?
- What defines identical organisms? A "self" conserved during sequences of differentiation that in humans developed from the first embryonic cell to the about "200 terminally differentiated cell types"? Interesting answers are suggested in different works dealing with the philosophy of biology or the "philosophy of cancer" (books by Pradeu, 2012, 2019; Bertolaso, 2016; Laplane, 2016; Plutynski, 2018, and others).
- Can we envision the immune system not as limited to the immune response to pathogens and abnormal host cells but rather as a law of cohesion for the whole organism construction?
- Would not the immune response be in this extended vision of the immune system only its "sword arm," a police patrol and pale reflection of the law itself, whereas a hidden part of the immune system would be the "spirit of laws" (analogous, *mutatis mutandis*, to Rousseau's unwritten social contract in human societies)?
- What holds together, normally without conflict, the cell types? Is it not something that governs development from the beginning—something more than what the immune system uses when it recognizes as non-self (foe rather than friend) a cancer cell?
- Is there a complementary relationship between the maintenance of such coherence and the major histocompatibility complex (MHC), or rather its likely forerunners in non-vertebrates, yielding early adaptive immunity?
- What is the primary function of the immune system if not to ensure organism cohesion (of tissues), and how does such coherence (of signals) operate? If it is so, what is the impact

of this (extended, i.e., going beyond the classical cellular and humoral immune response and earlier than it in the construction of multicellularity) version of the immune system on cell differentiation?
- Is the immune system the "glue" (Pradeu, 2012, 2019) that holds together the cells and functions of the multicellular organisms we all are constituted of together until such cohesion/coherence is altered in cancer?

4. THE ATAVISTIC THEORY OF CANCER

4.1. The Theory in a Nutshell

According to the atavistic theory, cancer is a local regression of a stable multicellular organism (Metazoa 2.0) to an incoherent state of a cell colony (Metazoa 1.0), non-existent as an evolution entity since it is not stable and is incapable of reproducing itself. This state is supposed to have predated the transition toward established stability that defines a stable and reproductive multicellular organism as a Darwinian selection unit. This point of view has been proposed at least in 1996 (Israel, 1996), and likely earlier, but was popularized in 2011 by Davies, Lineweaver, and Vincent (Davies and Lineweaver, 2011; Vincent, 2011a,b; Lineweaver et al., 2014; Thomas et al., 2017) and then examined from the point of view of the history of genes (Domazet-Lošo and Tautz, 2008, 2010; Wu et al., 2015; Bussey et al., 2017; Cisneros et al., 2017; Trigos et al., 2017, 2018, 2019). The atavistic theory of cancer has also recently been compared (Lineweaver et al., 2020) with the dominating (among cancer biologists) somatic mutation theory (SMT, that is more often compared with tissue organizational theory, TOFT) (Soto and Sonnenschein, 2004; Sonnenschein and Soto, 2016), and popularized in review articles (Goldman et al., 2017a,b). It poses the question of transition to multicellularity, for which we have to elaborate a plausible scenario, not sketched by the abovementioned authors.

4.2. Stage 0, aka Metazoa, The β Version

At this elementary stage of multicellularity, where proliferation limited by apoptosis is the only possible fate for cells (note that the emergence of apoptosis in evolution is studied in depth in Koonin and Aravind, 2002), they stick together in the ocean thanks to a form of collagen glue. Note that the existence of collagen implies enough availability of oxygen in the oceans, which dates this episode back to at least −850 million years. These cells are then able to exchange information, either by paracrine communication or by gap junctions (Trosko, 1987, 2007, 2016), through innexins present, e.g., in Hydra, rather than through connexins (Alexopoulos et al., 2004) or others (Mitchell and Nichols, 2019). Gap junctions allow cells to exchange molecules that can be toxic, such as oxygen, which can be indeed toxic before endosymbiosis of mitochondria in eucaryotes. With regards to the properties of cells at this stage, we assume only proliferation and its dual property, apoptosis, to be both influenced by environmental factors. We also assume a friend-or-foe recognition system to be present in each cell and they can use intercellular communication (paracrine or via gap junctions). Now, what should be the use of such a system if it would not react when a message testing an external intruder registers a foe, i.e., that we are under attack? Assuming no specialization (i.e., no differentiation and no division of work) at this stage, collective fright, fight or flight may be represented by a hedgehog-like attitude, the secretion of toxins in the environment, and collective movement without individual or semi-collective cell motility, respectively. Note that the latter is shown by tumor spheres with inverted polarity, TSIPs, which are moving hedgehogs or urchins (Zajac et al., 2018) encountered in breast and colorectal cancer cell populations. The genome of each of these cells has evolved to grant them such properties, making them able to resist UV radiation, acidity, cytotoxic molecules, hypoxia (after the endosymbiosis of mitochondria in the case of animal cells). A bond between them must exist that defines each of them as a member of a colony—a kind of *self* that controls proper cell division. This *self* and the friend-or-foe recognition system are assumed to be remote ancestors of the major histocompatibility complex (MHC, the common law in jawed vertebrates) and of the humoral immune response of vertebrates, immunoglobulins.

4.3. Stage 1, aka Metazoa 1.0

In the following stage, under the pressure of successive hostile attacks from the environment, begins the *reversible* division of work, i.e., differentiation of subpopulations of cells to allow them to perform specialized tasks, and this does not involve the whole cell population in all the tasks. According to Maynard Keynes and Szathmáry (1995), the first of such specializations could be the constitution of the germen (germinal cells), in charge of propagating the common genome, as opposed to the stroma (stromal cells), in charge of protecting and preserving the germen by all possible means of further specialization, e.g., motility, production of secretions, fast communications, etc. Differentiations producing division of work then appear, and they occur according to molecular determinants inscribed in the DNA, contacts between neighboring cells, and according to physical laws of soft matter that determine them in 3D space (Fleury, 2013). These differentiations are, however, very labile, i.e., reversible; otherwise said, the cells at Stage 1 are endowed with high plasticity with respect to their phenotypes.

Due to such plasticity, which prevents coherent construction of an organized cell colony that could be divided in cooperating subpopulations, no stable structure can emerge at this stage. The sketch of the immune system of Stage 1 has not evolved. On the contrary, something of the emerging self may be lost, as cell divisions may be futile, with junk DNA (the common law is easily trespassed and ignored) and existence of monster or non-viable cells. No working immune system leading to a stable coherent whole can exist in such cell populations. A Stage 1 cell colony is, according to the atavistic hypothesis of cancer (Davies and Lineweaver, 2011) characteristic of cancer cell populations, found in tumors. Many properties available in tumors, such as high individual plasticity, adaptability to external insults, loss of the common self (as all cells are potential defectors—cheaters—with respect to the poor common law of Stage 1) and no regulation of proliferation nor of differentiation, are present. Proliferation (fecundity) and apoptosis are now completed with differentiability and de-differentiability, i.e., extreme cell plasticity. Cooperation between subpopulations may

exist (Tabassum and Polyak, 2015) though not on a perennial nor consistent basis. From a metaphoric Waddington landscape point of view (Waddington, 1957; Huang et al., 2007; Huang, 2011, 2013), the scenery is flat or with unpredictably changing slopes. What can you build with plasticine bricks?

At this stage, the colony of cells is a soft and moving mixed cellular and acellular "soup." To achieve the transition from it to stable multicellularity (Newman, 2016), one can imagine that, if all elements in the genetic roadmap are present at least in some of these cells—in particular if sexed reproduction is also already active, such as in yeast cells—then physical laws of soft matter would drive this soup to a more consistent material. Indeed, mathematical natural gradient dynamics and singularity unfolding (Thom, 1972; Fleury, 2013) can be represented by chemical reaction-diffusion equations (Turing, 1952) at work in morphogen gradient-guided embryology processes. Many attempts to multicellularity may have occurred (and evolutionary biologists tell us that there have been may failed attempts) until a stable cohesion watch (maybe established, e.g., on paracrine or Delta/Notch communication or through gap junctions) can actually emerge and stabilize the structure of the plan. Then any fecundation that launches the division of a fertilized egg can be successful to yield a multicellular organism.

4.4. Stage 2, aka Metazoa 2.0

At Stage 2, an organizational principle emerges from the eddying chaos of Metazoa 1.0 and takes control of differentiations and proliferation. The common law is respected by all cells of the colony, and it can defend itself as a whole entity against attacks and can now inscribe itself in the fate of Darwinian evolution, maintained as a coherent ensemble by a functional immune system and a nerve communication system. The primitive Urmetazoa, as described in Müller (2001), Müller and Müller (2003), Müller et al. (2004), and Srivastastava et al. (2010), may have been a kind of sponge much like Porifera. The multicellularity gene toolkit of Metazoa 2.0 (Davies and Lineweaver, 2011) appeared at this stage, quite early and long before the Cambrian explosion, close to a date around −800 million years (Müller et al., 2004). What is this new collection of genes made of and how has it been hierarchically organized with respect to preexisting genes of unicellularity (e.g., cell cycle control)? What is the common law that defines an individual as any representative of its species (between-species distinction)? What defines a particular individual within its species (within-species distinction)? These questions ought to be documented to better understand what support this point of view may bring to the documentation of the idea of a hierarchical organization of the genome.

The immune system is now not only in charge of friend-or-foe recognition and defense of the colony when it is under attack, but it has, more importantly, emerged as a centralizer principle under the form of a chip present in every cell, ensuring the consistency of the whole construction. This common "law" is inscribed in the genome of each cell. Cheater cells may exist as in every organized society; however, they are sensed by a specialized subpopulation of cells (the police or immune cells) endowed with the mission to contain or destroy them. From a molecular point of view, repeat regions in the genome [in particular LINE-1 (Guler et al., 2017) in connection with the interferon pathway] could be responsible for such sensing. From the metaphoric Waddington viewpoint (Waddington, 1957; Huang et al., 2007; Huang, 2011, 2013), an irreversible differentiation potential (Zhou et al., 2012, 2018) is now present. With regards to the material construction of a stable organism, bricks and enamel are ready to be cooked in an oven, and perennial Assyrian palaces can now be built. What such virtual ovens consist of that will stabilize the multicellular organism during development we do not know; we can only suppose that some genes are silenced throughout this stabilization process.

Yet the fact remains that within the developmental stage of this construction, plasticity (reversibility with respect to a differentiation potential) is necessarily present for a limited time. This is the time of embryological development. After this time, the so-called Yamanaka genes (Takahashi and Yamanaka, 2006) Oct3/4, Sox2, c-Myc, and Klf4, that can reverse differentiation to produce induced pluripotent stem cells (iPSCs) are normally silenced (they can be revived in cancer, disease in which cells have not been properly "cooked" by gene silencing at some differentiation stage). Nevertheless, we know that some Metazoa, like the salamander (or axolotl), are able to locally go back to this developmental stage and regrow a tail or even a limb when it has been severed from the body.

The molecular level at which such control on differentiations is exerted is likely the level of the chromatin, where epigenetic enzymes, themselves coded by epigenetic genes, exert their control on the expression of genes, possibly by controlling transcription factors. The sequence of mutations observed in acute myeloid leukemia (AML), in evolutionary time firstly on epigenetic control genes, then on transcription and differentiation factors, and only finally on genes of proliferation (Hirsch et al., 2016), seems to recapitulate in reverse order the sequence of stages proposed here. One can suspect that a hierarchical relationship, such as that mentioned above about repeat sequences and the immune system, exists among control of gene expression at the chromatin level. Where a repository of an MHC-like common law, i.e., of marks defining not a particular individual but something common to all individuals of a given species and control of differentiations by the immune system, could exist is an open question. Indeed, such epigenetic/immune control of differentiations is not documented but is likely to exist.

To sum up this stage, there is persistent division of work since it appeared at Stage 1 already, but now it is consistently organized as *irreversible* differentiation. This constitutes a new fate (added to proliferation, apoptosis and senescence) in the physiological cellular life in each cell under the control of the cohesion watch during development. Later, added to the cohesion watch, specialized populations of cells, the immune patrol police, have appeared when the organism has been completely built. They oversee surveillance and the containment or destruction of trespassers. The cell colony, now a Metazoan 2.0 endowed with a functional immune system and able to reproduce itself, can successfully go through the tinkering (Jacob, 1977) of Darwinian evolution from sponges to vertebrates. However, in case of malfunction of any of its parts, due to malfunction of the

immune control (insufficient control) on its differentiation fate, this part is likely to revert to Stage 1, aka Metazoa 1.0, according to the atavistic hypothesis of cancer (Davies and Lineweaver, 2011). Conversely, when the police patrols (lymphocytes and macrophages) overreact, wrongly interpreting normal signals as trespasses, this may lead to allergies and auto-immune diseases.

5. WHAT IS A FUNCTIONAL MULTICELLULAR ORGANISM?

5.1. A Borromean System Responsible for the Emergence of Metazoa

The construction of the mind proposed now as common to all individuals in a species thus consists of the following:

(a) a base for the construction—the anatomical system, sets of genes in charge of the spatial embryological development, i.e., the 3D body plan (Müller et al., 2004; Amundson, 2005), and tissue/organ morphogenesis included;
(b) attached on this base to points that are virtual tissue-specific stem cells, domains of differentiation stemming as tree-like structures (inverted Waddington landscapes) of functionalities, i.e., sets of nodes of differentiations specific of a given functionality, e.g., in vertebrates, digestion, circulation and body covering, that in particular will yield the up to 200–400 functional human cell types (Jacob, 1977); and
(c) a hypothesized "cohesion watch" and complementary histocompatibility control system, which is a net made of connections—nervous, hormonal or by cell-to-cell contact—between and within the functionality trees in charge of controlling compatibilities and cooperation within each of the two systems and between the two of them, to achieve a cohesive and coherent multicellular system.

The whole construction should possess the characteristics of a Borromean system (endowed with the Brunnian property: removal of any one component unlinks the entire system) of length 3 (Chichak et al., 2004; Baas et al., 2015). Each subsystem exists independently of the other two, though no common sense can be obtained, in order to achieve the coherent design of a multicellular organism without the simultaneous participation of all three to the design. Furthermore, if any of them dissolves in the environment or fails its task, the other two may continue their separate existences, though this does not lead to a viable organism or else an impaired one. For instance, in the case of failure of control on the human body plan only, and in increasing order of gravity, we could see possible limb agenesis, partial rachischisis (spina bifida), and anencephaly except for the latter case in viable organisms.

The case of cancer, a disease specific of multicellular organisms, and in as much as it may destroy the whole organism, specific of animals [aka Metazoa, characterized by heterotrophicity among multicellular organisms; cancer exists in plants but remains localized and is not lethal (Doonan and Sablowski, 2010)], is the result of primary partial (local) failure of the compatibility control system (the cohesion watch) on the phenotypic coherence of the organism. In cancer, the body plan (in an extended sense, i.e., 3D anatomical shape and functional organ morphogenesis) is usually respected, but failure of control on differentiations (at the level of trees or inverted Waddington landscapes) gradually leads to incoherence in the cooperation tasks (improper division of work) between tissues and organs. Then the natural history of the disease leads to dissolution of the organism as a whole [de-unification of the individual, as Pradeu (2019) writes].

5.2. In More Detail, Why Is It a Borromean Structure?

Should the cohesion watch be firmly attached to the body plan but with missing places there for the trees of functionalities relying on phenotypic differentiation, this could (however unlikely in reality) lead to void shapes that one can figure as development stopped at different embryonic stages, e.g., gastrulation [in triploblastic animals (Seilacher et al., 1998; Martindale et al., 2004)] or neurulation (in vertebrates). If, conversely, it controls all trees responsible for cell-functional phenotypes, when all necessarily cell specializations have been achieved, but the body plan is loose (not cohesive), then division of work is there, and everything is ready for the emergence of a virtual Metazoan except that it cannot be embodied in a stable spatial and functional structure and thus cannot exist. Furthermore, the cohesion watch, an epigenetically controlled non-cellular system of intercellular communication controlling differentiations, must make these differentiations irreversible to yield a stable multicellular organism. Before its appearance in evolution, differentiations were partially or completely reversible, which was in particular useful to making the whole construction able to mobilize enough cells in the colony to face an incoming external aggression. This might be by motility and by specialization into protecting cells, precursors of immune cells, facing them by fight, flight, or fright. In the metaphor of the Waddington landscape, such irreversibility is ensured by the establishment of high epigenetic barriers that prevent de-differentiation or transdifferentiation. Indeed, evolution cannot build anything perennial on moving ground (non-moving meaning here a permanent spatially and functionally organized support within which cell subpopulations can cooperate to establish an individual able to feed on its environment), avoid destruction from it and secure its reproduction.

The cellular immune system cannot appear out of thin air but could emerge from a specialization from a primitive immune-like cell type in the initial cell colony, yielding cells and signaling molecules able to recognize both the MHC, or rather its forerunners in evolution, by tagging an individual in a given multicellular species. It will also be able to recognize common markers, tagging the species, in any cell of the colony. The next stage would be to validate them as faithful elements of the ensemble or else to destroy them or reject them from the cell colony by making use of an armed force, the (cellular and humoral) immune response. These specialized immune cells should then take control of all other cells of both the anatomical development system (the materially established body plan) and of the epigenetic system of differentiations rendered irreversible by

the cohesion watch to then emerge during early embryogenesis a truly stable Metazoan. The article envisions the cohesion watch as a set of intercellular communications assumed to be present in all cells of a Metazoan 2.0 (including the emerging immune cells) under the form of a program that is the basis of the "common law" of the species. Such a dual event, preexisting cohesion watch in all cells—the common law—and enforcement of the cohesion by the materially constituted cohesion law and by the emerging immune cells during embryogenesis—the sword arm or police—is highly evocative of the constitution of an emergent Borromean system. Before its emergence, it can only exist as tumor-like Davies's and Lineweaver's Metazoa 1.0, and after it is constituted, it exists as a cohesive and stable Metazoan 2.0 (a true Metazoan).

5.3. The Basic Anatomic System: The Body Plan in Development

The structure of the body plan (Davidson et al., 1995; Müller et al., 2004) is not easily defined, as it has evolved along with the evolution of species. However, one might define it, independently of the animal species under consideration, as the anatomically based collection of all organism functionalities. Well-known by embryologists for quite a long time, long before the emergence of the study of genetics and the knowledge of the roles of body plan genes, the embryological development of animals has been described from the blastula stage (a 2D sphere made of undifferentiated cells) until the constitution of forms that depend on the species. These forms resort to diploblastism (two layers: the endoderm and ectoderm) in elementary Metazoa, such as placozoa, ctenophora or cnidarians, and later triploblastism (three layers covering a 2D-sphere: the endoderm, ectoderm, and between them, mesoderm) in all others. Triploblastic animals appeared between 1 billion years and 600 million years ago and were later structured by a hard skeleton during the Cambrian explosion, which began 541 million years ago and lasting for about 13–25 million years. In triploblastic animals, particularly in vertebrates, gastrulation and neurulation are dynamic phenomena in which cells follow flows that will constitute their anatomic structures. They have recently been described from a physicist's point of view by Fleury (2013) and from a mathematician's point of view, much earlier, by Thom (1972). No genes are present in the points of view of these authors. However, the explanation of the formation of embryological layers due to the dynamics of morphogen gradients, firstly predicted by Turing (1952), now identified as, e.g., Wnt, and controlled by, e.g., Hox, is presently the norm—all the more so as knock-out embryos (mice and flies) for these genes are currently documented to help us understand their precise roles in anatomical development (Amundson, 2005).

5.4. The Trees of Cell Specialization Controlled by Transcription Factors and Epigenetic Enzymes

Cell functionalities, relying on functional cell phenotypes, were developed in a cell colony with the emergence of transcription factors (de Mendoza and Sebé-Pedrós, 2019). Their combinations forming gene regulatory networks (GRNs) may have occurred very early, as many transcription factors were already present as early as 1.5 billion years ago, in LECA, the last eucaryotic common ancestor (de Mendoza and Sebé-Pedrós, 2019). One may assume that, likely due to the necessity to develop functional capabilities to make individual cells able to adapt to changing and often hostile environments, transcription factors have gradually combined into GRNs, constituting the biological support of the expression of functional phenotypes. Furthermore, differentiations are by nature *epigenetic*, insofar as they occur, leading to very different terminal cell types on the basis of the same genome, which naturally creates a role for epigenetic enzymes at the level of chromatin, partly unraveled in Arney and Fisher (2004) in their relationship with transcription factors, and more recently in Atlasi and Stunnenberg (2017).

Such differentiation phenotypes, achieved by specializations or branching points in the trees that, before the emergence of Metazoa 2.0, were likely all reversible, are modules of elementary adaptation to the external environment already present in unicellular constituents:

- germinal or somatic nature (duality germen/soma in sexed reproduction),
- motility or attachment to a matrix,
- emission/reception of (fast or slow) communication between cells of the colony,
- means of absorption of fueling matter and of elimination of toxic residues,
- activator-inhibitor dynamics, leading to space/time periodic behavior of tissues and of intracellular/intercellular signaling pathways, mandatory to maintain continuity of flows in a limited space, and
- friend-or-foe recognition and elimination of (or fight from) foes.

These cell phenotypes, before the closure of the Borromean node, i.e., before the actual emergence of Metazoa 2.0, are still not fixed by epigenetic constraints, and they are thus widely reversible. In other words, the epigenetic landscape is flat. This will change when some newly established differentiation potential (Zhou et al., 2012, 2018), ensured by the cohesion watch hypothesized to be part of the immune system, will force differentiations to become irreversible.

5.5. The Working Immune System Involves a Cohesion Watch in Charge of Compatibilities

Indeed, the immune control of cell differentiation should consist firstly of checking their coherence (i.e., that cells follow a coherent differentiation path according to simple rules in terms of the complementary histocompatibility complex—the cohesion watch hypothesized earlier in this construction) and secondly of making these differentiations irreversible. The latter implies the constitution of a potential (Zhou et al., 2012, 2018), or of an entropy, at its highest level in stem cells of the tissue (e.g., hematopoietic stem cells for blood) and at its lowest level in the ultimately differentiated cells of the lineage. Among the differentiated blood cells are lymphocytes,

which are in charge of the control of surface antigens of all other tissues.

In more detail, the task of the hypothesized cohesion watch, part of this extended version of the immune system, which must exist already virtually, as inscribed in the self-extracting archive of the genome before fecundation, is thus to ensure compatibilities:

(a) between morphogens of the body plan, able to drive it actually from the zygote in an irreversible way within the 3D space of cells of a given individual (defined by its MHC in vertebrates by some equivalent forerunners in non-vertebrates);
(b) between phenotypic functionalities, ensuring compatibility between differentiation trees that yield lineages within a given subpopulation, and ultimately between cooperating subpopulations (division of work) of terminally differentiated cells; and
(c) between the body plan space distribution and the time distribution of phenotypes in each epigenetic landscape attached to the body plan.

One can think of this cohesion watch as being in charge of irreversibility of differentiations along each tree stemming from the body plan (vertical cohesion), but also of compatibility at each developmental stage between neighboring functionality trees. This involves transversal cohesion, failed for instance in cervical cancer due to histological uncertainty between two different epithelial coverings likely resulting from impaired differentiation of immature renewing cells in one or both lineages, and it is mandatory to form a cohesion net, knitting node after node in all relevant directions.

To mentally illustrate this construction, consider the wickerwork basket. Starting from a circle endowed with lots of connections between its elements, supposed to represent the body plan, functional willow-like twigs stem from each of these elements, representing the great physiological functions of the organism. If no weaving is made between these twigs, the whole set will consist of just flexible differentiation functionalities of a family of cell types, floating freely in the surrounding space unrelated to each other. No cohesion—no division of labor—can result from such unwoven twigs and trees. The task of the cohesion watch is to ensure such weaving during development until tips that are terminally differentiated cells. This naturally includes the solidity of the willow twigs [breaches along the vertical axis resulting in blocked differentiations, as is the case with acute myeloid leukemia (AML)], but the main part of the cohesion watch is to ensure compatibility between (spatially and functionally) neighboring twigs.

Could this hypothesis be tested by evaluation of coherence in the expression of transcription factors responsible for the differentiations of mandatorily compatible tissues at different stages of their differentiations? This could rely on the investigation of intercellular communication means regulating GRNs in different cells, as described in Peter and Davidson (2017) and Erkenbrack et al. (2018).

The emerging capacities of the whole system consisting of the three subsystems, the body plan, trees of phenotypic functionalities giving rise to lineages from virtual pluripotent stem cells and cohesion watch, will now endow the multicellular organism-to-be (after fecundation), in a coherent and stable way. This will consist of making use of division of work and cooperation between the subsystems with functionalities, relying on survival means based on the non-exhaustive list of elementary adaptation phenotypes above and later producing "the great physiological functions" taught to students in medicine and physiology. These capacities will identify a common characteristic of a well-defined species. These might look like the following:

- boundaries with the external environment in both the anatomical spatial and phenotypic (protection) senses,
- strategies to feed on the environment by ingestion of prey,
- friend-or-foe recognition and surveillance against predators,
- abilities to react to hostile environments, whole organism motility (flight) being one such ability,
- integration of all cells by rapid intercellular communication networks,
- reproduction facilities (sexed reproduction by germinal/somatic cell specialization), and
- cognitive processes.

Cognitive processes are indeed among the mandatory functionalities of an evolved multicellular organism (not only vertebrates but including also, for instance, octopuses) under the control of the hypothesized cohesion watch. Conversely, could there exist a support for a possible control of cognitive processes on the immune control of proliferation and differentiation that might explain some inexplicable spontaneous cures of cancer? If so, would the classic immune response (cellular and humoral) be responsible for it, or could it be an effect of the cohesion watch? All physicians are aware of such stories of cures that cannot find any explanation within the corpus of medical knowledge except by a timely intervention of the immune system. An example of a mild one is a plantar wart about to be surgically excised that completely disappeared in one night without any trace on the morning of the intervention; other examples exist for cancer, usually not reported as medical observations, being beyond the scope of contemporary science. This means that, even though the existence of a cohesion watch is primordial for the stability of the organism, it may itself become a part of the organism under the control of superior integrative control, of nervous origin, that unifies a particular individual within a given species with respect to the maintenance of its stability in behavioral life. Michel Jouvet has proposed the interesting hypothesis that the physiological meaning of cortical activity during paradoxical sleep, i.e., dreaming, is a neuronal reprogramming of the individual, a consultation of their genetic program together with their past life personal history, aimed at adapting behavior to be ready to solve issues they will likely meet in their immediate future (Jouvet, 1978, cited by Nathan, 2011).

6. PERSPECTIVES IN CANCER THERAPEUTICS

Within this evolutionary perspective of the design of a multicellular organism, developmental diseases like those mentioned ones above are diseases of the immune system control of the body plan. Assuming a cohesive body plan, which is usually the case, cancer appears as a loss of control of the immune system on the trees of differentiations and on compatibility connections between them. Cancer may thus be the result of flaws in the means of control or the result of incoherence in the control subsystem itself. Auto-immune diseases are clearly due to incoherence in this controlling immune subsystem.

From a cancer therapeutic viewpoint, as stated by Lineweaver et al. (2014), attacking cancer by blocking its proliferation using chemotherapies or radiotherapies is clearly a short-sighted method. It may work completely in some cases, though most often partially and temporarily, but as long as the epigenetic system of control on differentiations fails, the dynamics of cancer will prevail. This may be avoided if the immune response keeps residual cancer cells in check, preventing them from excessive proliferation; this is usually called cancer dormancy and is not clinically distinguishable from a cure if it is indefinitely prolonged.

Despite this limitation, relying on the existing cell-killing therapies, which may be (cytotoxic) chemotherapies, (cytostatic) targeted therapies or immunotherapies, mathematical models have been developed with corresponding theoretically optimized treatment strategies representing monotherapies or more successfully combination therapies in cancer (Pouchol, 2018; Pouchol et al., 2018; Jarrett et al., 2020). The starting point of this article was a solution that—inscribed in the time scale of a human life and not in the billion-year perspective presented above—aims to be immediately useful in the clinic. Taking advantage only of what we know presently of the behavior of cancer cells exposed to cytotoxic and cytostatic drugs in the framework of a cell population and not of the history of their making, as is the goal of the presently proposed billion-year perspective for therapeutics, this has been briefly described in the first section of the present study. Among modern immunotherapies, immune checkpoint inhibitors (ICIs), by boosting the immune response by lymphocytes that attack tumor cells, e.g., in the case of melanoma treated with the combination ipilimumab + nivolumab, may be successful with about 60% of objective response rates in patients, of which 20% of total cases can even reach complete long remissions. Unfortunately, there may also more rarely exist total failures that result in non-responders in 30% of cases and even in so-called hyperprogressors (i.e., experiencing accelerated tumor growth defined by at least a 2-fold tumor growth rate increase compared with pre-immunotherapy rate) in the remaining 10% (Márquez-Rodas et al., 2015; Frelaut et al., 2019; Liu et al., 2019). Such cell-killing strategies may be successful by mending a breach in the control of cell proliferation, but if a fragility remains in the control of differentiations somewhere in the organism, a relapse may occur, possibly with cells that will have been selected for their robustness and will be less sensitive to the treatment.

This should induce us to enhance our understanding of the role of the immune system (and more precisely of the cohesion watch) in the hypothesized Borromean system upon which relies a physiologically well-constituted animal. Rather than fighting uncontrolled proliferation, could we repair altered control on differentiations? Cell-killing strategies, whether they rely on chemotherapies or on modern immune cell-enhancing drugs, miss the basic targets, which are differentiation sites. There are possibly only two known successful non-cell-killing therapies: imatinib in chronic myelogenous leukemia (CML) (Hochhaus et al., 2008), where imatinib [or drugs of the same family of tyrosine kinase inhibitors (TKIs)] blocks the ATP pocket of a chimeric protein, BCR-ABL, which itself is due to a fusion of genes, normalizing proliferation; secondly, there is all-trans retinoic acid (ATRA) in acute promyelocytic leukemia (APL = AML3 in the old French-American-British, FAB, classification of acute myeloid leukemias) (Haferlach, 2008), where ATRA degrades the PML-RARα chimeric protein (that also results from a fusion of genes) that blocks maturation of the myeloid lineage at the promyelocytic stage. Many redifferentiation strategies close to this one have been attempted, but all the others have failed.

Nevertheless, this could be the future of cancer therapeutics: intervention at the differentiation sites on transcription factors or on factors that control them, i.e., enforcing the cohesion watch connection rather than killing cheater cells; in other words, the solution could be in mending a net with a hole in it rather than trying to kill sharks that have escaped containment. Alternatively, this goal can be illustrated with a sociological metaphor. This is indeed relevant as, in the hierarchy of levels of organization that goes from genes to cells and from cells to multicellular organisms, the next level is evolving societies of living multicellular individuals. In light of the above, rather than killing cheater cells through cannonade (i.e., by chemotherapies) or by enforcing the aggressiveness of the police (i.e., by immune checkpoint inhibitors), would it not be better to assess how we enforce the law? The law here is the cohesion watch that exists as a plan in the genome before embodiment in development and later as an acellular communication network between tissues and organs. This could be done by repairing broken local social bonds between functionalities (expressed after embodiment as tissues and organs), as neither the army nor the police are the best means to establish harmonious working links of cooperation between citizens. Citizens in multicellular organisms are the somatic cells in tissues and organs that are normally organized toward a common goal: preservation of the genome toward reproduction, and to that purpose, the preservation of the health of the global society of cells. To be able to do this, a better understanding of the mechanisms of control of differentiation at the level of local transcription factors and at the level of chromatin is needed. The development of epigenetic drugs is promising, widely relying on inhibitors of DNA methyltransferases (iDNMTs) or of histone deacetylases (iHDACs) (Roberti et al., 2019). These could be a starting point, provided that the interactions between epigenetic enzymes and transcription factors can be unraveled (Arney and Fisher, 2004). This could lead to future differentiation-repairing cancer therapies that would be precisely targeted at the best

possible sites of multicellular organisms and would disregard cell-killing therapies, except to accelerate a clearance process, as with ATRA, which is usually delivered together with an anthracyclin, resulting in a complete cure of APL (Haferlach, 2008), in a remote past of cancer medicine. Another route to explore might be to examine, following Davidson's works on intercellular communication means that regulate consistency between intracellular GRNs during development (Peter and Davidson, 2017; Erkenbrack et al., 2018), targets and reestablish such impaired intercellular signaling.

7. CONCLUSION

Moving away from deliberations on the evolution of a cell population at the time scale of a human, which is nevertheless undoubtedly if high interest in therapeutics, an example of which is that what this article advocated along with Robert Gatenby and his colleagues at the Moffitt Cancer Center in Tampa (Gatenby et al., 2009; Gillies et al., 2012; West et al., 2020) in terms of mathematical models designed to optimize strategies based on combined cell-killing therapies (Pouchol, 2018; Pouchol et al., 2018; Jarrett et al., 2020), this article further presents an evolutionary point of view on cancer from a billion-year perspective that, from questions on plasticity in cancer, has guided the development of ideas resorting to what is now named the *philosophy of cancer* (Pradeu, 2012, 2019; Bertolaso, 2016; Laplane, 2016; Plutynski, 2018). The view takes basis in various philosophers of cancer, walking a long and winding path toward a fundamental understanding of multicellularity and of its alterations in cancer. Ultimately, following this path should lead to correct impaired control of differentiation rather than, or at least together with, control of proliferation. Much of what is presented here, as much as it is possible to rely on published observations or opinions, is of speculative nature, in particular with respect to the exploration, discovery and generalization of non-cell-killing therapies, which so far remain elusive in the clinic. Nevertheless, in a time when humanities, mathematics, biology and medicine are uniting their efforts to overcome the struggle against cancer, this approach is hopefully a timely one.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

I am gratefully indebted to my fellows of the informal work group of mathematicians on multicellularity and cancer, Nils Baas, Gregory Ginot, Jean-François Mascari, and Rémy Tuyéras, though the SARS-CoV2 pandemic have restricted our interactions to the strictly virtual when debating constructive ideas. This will hopefully 1 day result in mathematical elements of a geometrical theory of multicellularity and of its alterations in cancer. We hope that such a geometrical theory will provide orientations toward improvement of information extraction from mass cancer data, and in the future of medicine, proposals of therapeutic strategies with respect to precise molecular targets for the correction of altered differentiation mechanisms in cancer.

It is also my pleasure to gratefully acknowledge the friendly help of William S. Levine in editing the final version of the manuscript.

REFERENCES

Aktipis, C. A., Boddy, A. M., Jansen, G., Hibner, U., Hochberg, M. E., Maley, C. C., et al. (2015). Cancer across the tree of life: cooperation and cheating in multicellularity. *Philos. Trans. R. Soc. B* 370:20140219. doi: 10.1098/rstb.2014.0219

Alexopoulos, H., Böttger, A., Fischer, S., Levin, A., Wolf, A., Fujisawa, T., et al. (2004). Evolution of gap junctions: the missing link? *Curr. Biol.* 14, 879–880. doi: 10.1016/j.cub.2004.09.067

Almeida, L., Bagnerini, P., Fabrini, G., Barry, D., Hughes, B. D., and Lorenzi, T. (2019). Evolution of cancer cell populations under cytotoxic therapy and treatment optimisation: insight from a phenotype-structured model. *ESAIM Math. Model. Numer. Anal.* 4, 1157–1190. doi: 10.1051/m2an/2019010

Almeida, L., Chisholm, R. H., Clairambault, J., Lorenzi, T., Lorz, A., Pouchol, C., et al. (2018). "Why is evolution important in cancer and what mathematics should be used to treat cancer? Focus on drug resistance," in *Trends in Biomathematics: Modeling, Optimization and Computational Problems*, ed R. P. Mondaini (Cham: Springer), 107–120. doi: 10.1007/978-3-319-91092-5_8

Amundson, R. (2005). *The Changing Role of the Embryo in Evolutionary Thought*. Cambridge: Cambridge University Press.

Arney, K. L., and Fisher, A. G. (2004). Epigenetic aspects of differentiation. *J. Cell Sci.* 117, 4355–4363. doi: 10.1242/jcs.01390

Atlasi, Y., and Stunnenberg, H. G. (2017). The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* 18, 643–658. doi: 10.1038/nrg.2017.57

Baas, N. A., Seeman, N. C., and Stacey, A. (2015). Synthesising topological links. *J. Math. Chem.* 53, 183–199. doi: 10.1007/s10910-014-0420-3

Bertolaso, M. (2016). *Philosophy of Cancer*. Dordrecht: Springer.

Bussey, K. J., Cisneros, L. H., Lineweaver, C. H., and Davies, P. C. W. (2017). Ancestral gene regulatory networks drive cancer. *Proc. Nat. Acad. Sci. U.S.A.* 114, 6160–6162. doi: 10.1073/pnas.1706990114

Chichak, K. S., Cantrill, S. J., Pease, A. R., Chiu, S. H., Cave, G. W. V., Atwood, J. L., et al. (2004). Molecular Borromean rings. *Science* 304, 1308–1312. doi: 10.1126/science.1096914

Chisholm, R. H., Lorenzi, T., and Clairambault, J. (2016a). Cell population heterogeneity and evolution towards drug resistance in cancer: biological and mathematical assessment, theoretical treatment optimisation. *Biochem. Biophys. Acta* 1860, 2627–2645. doi: 10.1016/j.bbagen.2016.06.009

Chisholm, R. H., Lorenzi, T., and Lorz, A. (2016b). Effects of an advection term in nonlocal Lotka-Volterra equations. *Comm. Math. Sci.* 14, 1181–1188. doi: 10.4310/CMS.2016.v14.n4.a16

Chisholm, R. H., Lorenzi, T., Lorz, A., Larsen, A. K., Almeida, L. N., Escargueil, A., et al. (2015). Emergence of drug tolerance in cancer cell populations: an evolutionary outcome of selection, non-genetic instability and stress-induced adaptation. *Cancer Res.* 75, 930–939. doi: 10.1158/0008-5472.CAN-14-2103

Cho, H., and Levy, D. (2018a). Modeling continuous levels of resistance to multidrug therapy in cancer. *Appl. Math. Model.* 64, 733–751. doi: 10.1016/j.apm.2018.07.025

Cho, H., and Levy, D. (2018b). Modeling the chemotherapy-induced selection of drug-resistant traits during tumor growth. *J. Theor. Biol.* 436, 120–134. doi: 10.1016/j.jtbi.2017.10.005

Cisneros, L. H., Bussey, K. J., Orr, A. J., Miočević, M., Lineweaver, C. H., and Davies, P. (2017). Ancient genes establish stress-induced mutation as a hallmark of cancer. *PLoS ONE* 12:e0176258. doi: 10.1371/journal.pone.0176258

Clairambault, J. (2019). An evolutionary perspective on cancer, with applications to anticancer drug resistance modelling and perspectives in therapeutic control. *J. Math. Study* 52, 470–496. doi: 10.4208/jms.v52n4.19.06

Clairambault, J., and Pouchol, C. (2019). A survey of adaptive cell population dynamics models of emergence of drug resistance in cancer, and open questions about evolution and cancer. *Biomath* 8:23. doi: 10.11145/j.biomath.2019.05.147

Davidson, E. H., Peterson, K. J., and Cameron, R. A. (1995). Origin of bilaterian body plans: evolution of developmental regulatory mechanisms. *Science* 270, 1319–1325. doi: 10.1126/science.270.5240.1319

Davies, P. C. W., and Lineweaver, C. H. (2011). Cancer tumors as Metazoa 1.0: tapping genes of ancient ancestors. *Phys. Biol.* 8:015001. doi: 10.1088/1478-3975/8/1/015001

de Mendoza, A., and Sebé-Pedrós, A. (2019). Origin and evolution of eukaryotic transcription factors. *Curr. Opt. Genet Dev.* 58–59, 25–32. doi: 10.1016/j.gde.2019.07.010

Domazet-Lošo, T., and Tautz, D. (2008). An ancient evolutionary origin of genes associated with human genetic diseases. *Mol. Biol. Evol.* 25, 2699–2707. doi: 10.1093/molbev/msn214

Domazet-Lošo, T., and Tautz, D. (2010). Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoa. BMC *Biol.* 8:66. doi: 10.1186/1741-7007-8-66

Doonan, J. H., and Sablowski, R. (2010). Walls around tumours-why plants do not develop cancer. *Nat. Rev. Cancer* 10, 794–802. doi: 10.1038/nrc2942

Erkenbrack, E. M., Davidson, E. H., and Peter, I. S. (2018). Conserved regulatory state expression controlled by divergent developmental gene regulatory networks in echinoids. *Development* 145:dev167288. doi: 10.1242/dev.167288

Fleury, V. (2013). Development, triploblastism, physics of wetting and the Cambrian explosion. *Acta Biotheor.* 61, 385–396.. doi: 10.1007/s10441-013-9191-7

Frelaut, M., Le Tourneau, C., and Borcoman, E. (2019). Hyperprogression under immunotherapy. *Int. J. Mol. Sci.* 20:2674. doi: 10.3390/ijms20112674

Gatenby, R. A., Silva, A. S., Gillies, R. J., and Frieden, B. R. (2009). Adaptive therapy. *Cancer Res.* 69, 4894–4903. doi: 10.1158/0008-5472.CAN-08-3658

Gillies, R. J., Verduzco, D., and Gatenby, R. A. (2012). Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nat. Rev. Cancer* 12, 487–493. doi: 10.1038/nrc3298

Goldman, A., Kohandel, M., and Clairambault, J. (2017a). Integrating biological and mathematical models to explain and overcome drug resistance in cancer, part 1: biological facts and studies in drug resistance. *Curr. Stem Cell Rep.* 3, 253–259. doi: 10.1007/s40778-017-0097-1

Goldman, A., Kohandel, M., and Clairambault, J. (2017b). Integrating biological and mathematical models to explain and overcome drug resistance in cancer, part 2: from theoretical biology to mathematical models. *Curr. Stem Cell Rep.* 3, 260–268. doi: 10.1007/s40778-017-0098-0

Guler, G. D., Tindell, C. A., Pitti, R., Wilson, C., Nichols, K., and Cheung, T. K. W., et al. (2017). Repression of stress-induced LINE-1 expression protects cancer cell subpopulations from lethal drug exposure. *Cancer Cell* 32, 221–237. doi: 10.1016/j.ccell.2017.07.002

Haferlach, T. (2008). Molecular genetic pathways as therapeutic targets in acute myeloid leukemia. *Hematol. Am. Soc. Hematol. Educ. Program* 2008, 400–411. doi: 10.1182/asheducation-2008.1.400

Hirsch, P., Zhang, Y., Tang, R., Joulin, V., Boutroux, H., Pronier, E., et al. (2016). Genetic hierarchy and temporal variegation in the clonal history of acute myeloid leukemia. *Nat. Comm.* 7:12475. doi: 10.1038/ncomms12475

Hochhaus, A., Druker, B., Sawyers, C., Guilhot, F., Schiffer, C. A., Cortes, J., et al. (2008). Favorable long-term follow-up results over 6 years for response, survival, and safety with imatinib mesylate therapy in chronic-phase chronic myeloid leukemia after failure of interferon-alpha treatment. *Blood* 111, 1039–1043. doi: 10.1182/blood-2007-07-103523

Huang, S. (2011). On the intrinsic inevitability of cancer: from foetal to fatal attraction. *Semin. Cancer. Biol.* 21, 183–199. doi: 10.1016/j.semcancer.2011.05.003

Huang, S. (2013). Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. *Cancer Metast. Rev.* 32, 423–448. doi: 10.1007/s10555-013-9435-7

Huang, S., Guo, Y., May, G., and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* 305, 695–713. doi: 10.1016/j.ydbio.2007.02.036

Israel, L. (1996). Tumour progression: random mutations or an integrated survival response to cellular stress conserved from unicellular organisms? *J. Theor. Biol.* 178, 375–380. doi: 10.1006/jtbi.1996.0033

Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166. doi: 10.1126/science.860134

Jarrett, A. M., Faghihi, D., Hortmuth, D. A. I. I., Lima, E. A. B. F., Virostko, J., Biros, G., et al. (2020). Optimal control theory for personalized therapeutic regimens in oncology: background, history, challenges, and opportunities. *J. Clin. Med.* 9:e1314. doi: 10.3390/jcm9051314

Jouvet, M. (1978). "Does a genetic programming of the brain occur during paradoxical sleep?" in *Cerebral Correlates of Conscious Experience, INSERM Symposium 6*, eds P. Buser and A. Buser-Rougeul (Amsterdam: Elsevier/North-Holland Biomedical Press), 245–261.

Koonin, E. V., and Aravind, L. (2002). Origin and evolution of eukaryotic apoptosis: the bacterial connection. *Cell Death Dis.* 394–404. doi: 10.1038/sj.cdd.4400991

Laplane, L. (2016). *Cancer Stem Cells*. Harvard University Press.

Lavi, O., Greene, J. M., and Levy, D. (2013). The role of cell density and intratumoral heterogeneity in multidrug resistance. *Cancer Res.* 73, 7168–7175. doi: 10.1158/0008-5472.CAN-13-1768

Lineweaver, C. H., and Davies, P. C. W. (2020). Comparison of the atavistic model of cancer to somatic mutation theory: phylostratigraphic analyses support the atavistic model. in *Physics of Cancer*, ed B. S. Gerstman (Singapore: World Scientific). doi: 10.1142/11915

Lineweaver, C. H., Davies, P. C. W., and Vincent, M. D. (2014). Targeting cancer's weaknesses (not its strengths): therapeutic strategies suggested by the atavistic model. *Bioessays* 36, 827–835. doi: 10.1002/bies.201400070

Liu, D., Jenkins, R. W., and Sullivan, R. J. (2019). Mechanisms of resistance to immune checkpoint blockade. *Am. J. Clin. Dermatol.* 20, 41–54. doi: 10.1007/s40257-018-0389-y

Lorenzi, T., Chisholm, R. H., and Clairambault, J. (2016). Tracking the evolution of cancer cell populations through the mathematical lens of phenotype-structured equations. *Biol. Direct* 11:43. doi: 10.1186/s13062-016-0143-4

Lorz, A., Lorenzi, T., Clairambault, J., Escargueil, A., and Perthame, B. (2015). Modeling the effects of space structure and combination therapies on phenotypic heterogeneity and drug resistance in solid tumors. *Bull. Math. Biol.* 77, 1–22. doi: 10.1007/s11538-014-0046-4

Lorz, A., Lorenzi, T., Hochberg, M. E., Clairambault, J., and Perthame, B. (2013). Populational adaptive evolution, chemotherapeutic resistance and multiple anti-cancer therapies. *ESAIM Math. Model. Numer. Anal.* 47, 377–399. doi: 10.1051/m2an/2012031

Márquez-Rodas, I., Cerezuela, P., Soria, A., Berrocal, A., Riso, A., González-Cao, M., et al. (2015). Immune checkpoint inhibitors: therapeutic advances in melanoma. *Ann. Transl. Med.* 3:267. doi: 10.3978/j.issn.2305-5839.2015.10.27

Martindale, M. Q., Pang, K., and Finnerty, J. (2004). Investigating the origins of triploblasty: 'mesodermal' gene expression in a diploblastic animal, the sea anemone Nematostella vectensis (phylum, Cnidaria; class, Anthozoa). *Development* 131, 2463–2474. doi: 10.1242/dev.01119

Maynard Keynes, J., and Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford: Oxford University Press.

Mitchell, J. M., and Nichols, S. A. (2019). Diverse cell junctions with unique molecular composition in tissues of a sponge (Porifera). *EvoDevo* 10:26. doi: 10.1186/s13227-019-0139-0

Müller, W. E. G. (2001). Review: how was metazoan threshold crossed? The hypothetical Urmetazoa. *Compar. Biochem. Physiol. A* 129, 433–460. doi: 10.1016/S1095-6433(00)00360-3

Müller, W. E. G., and Müller, I. M. (2003). Origin of the metazoan immune system: identification of the molecules and their functions in sponges. *Integr. Compar. Biol.* 43, 281–292. doi: 10.1093/icb/43.2.281

Müller, W. E. G., Wiens, M., Adell, T., Gamulin, V., Schröder, H. C., and Müller, I. M. (2004). Bauplan of Urmetazoa: basis for genetic complexity of metazoa. *Int. Rev. Cytol.* 235, 53–92. doi: 10.1016/S0074-7696(04)35002-3

Nathan, T. (2011). *La nouvelle interprétation des r^eves*. Paris: Odile Jacob.

Newman, S. A. (2016). 'Biogeneric' developmental processes: drivers of major transitions in animal evolution. *Philos. Trans. R. Soc. B* 371:20150443. doi: 10.1098/rstb.2015.0443

Nguyen, T. N., Clairambault, J., Jaffredo, T., Perthame, B., and Salort, D. (2019). Adaptive dynamics of hematopoietic stem cells and their supporting stroma: a model and mathematical analysis. *Math. Biosci. Eng.* 16, 4818–4845. doi: 10.3934/mbe.2019243

Perthame, B. (2007). *Transport Equations in Biology*. Birkhäuser: Springer.

Perthame, B. (2015). *Parabolic Equations in Biology*. Cham: Springer.

Peter, I. S., and Davidson, E. H. (2017). Assessing regulatory information in developmental gene regulatory networks. *Proc. Nat. Acad. Sci. U.S.A.* 114, 5862–5869. doi: 10.1073/pnas.1610616114

Plutynski, A. (2018). *Explaining Cancer*. Oxford: Oxford University Press.

Pouchol, C. (2018). *Analysis, control and optimisation of PDEs, application to the biology and therapy of cancer* (Ph.D. thesis), Sorbonne Université, Paris, France. Available online at: https://hal.inria.fr/tel-01889253

Pouchol, C., Clairambault, J., Lorz, A., and Trélat, E. (2018). Asymptotic analysis and optimal control of an integro-differential system modelling healthy and cancer cells exposed to chemotherapy. *J. Math. Pure Appl.* 116, 268–308. doi: 10.1016/j.matpur.2017.10.007

Pradeu, T. (2012). *The Limits of the Self*. Oxford: Oxford University Press.

Pradeu, T. (2019). *Philosophy of Immunology*. Cambridge University Press.

Roberti, A., Valdes, A. F., Torrecillas, R., Fraga, M. F., and Fernandez, A. F. (2019). Epigenetics in cancer therapy and nanomedicine. *Clin. Epigenet.* 11:81. doi: 10.1186/s13148-019-0675-4

Seilacher, A., Bose, P. K., and Pflüger, F. (1998). Triploblastic animals more than 1 billion years ago: trace fossil evidence from India. *Science* 282, 80–84. doi: 10.1126/science.282.5386.80

Sharma, S. V., Lee, D. Y., Li, B., Quinlan, M. P., Takahashi, F., Maheswaran, S., et al. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* 141, 69–80. doi: 10.1016/j.cell.2010.02.027

Shen, S. S., and Clairambault, J. (2020). Plasticity in cancer cell populations. *F1000Res.* 9:F1000 Faculty Rev-635. doi: 10.12688/f1000research.24803.1

Sonnenschein, C., and Soto, A. M. (2016). Carcinogenesis explained within the context of a theory of organisms. *Prog. Biophys. Mol. Biol.* 122, 70–76. doi: 10.1016/j.pbiomolbio.2016.07.004

Soto, A. M., and Sonnenschein, C. (2004). The somatic mutation theory of cancer: growing problems with the paradigm? *Bioessays* 26, 1097–1107. doi: 10.1002/bies.20087

Srivastastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E. A., Mitros, T., et al. (2010). The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* 466, 720–727. doi: 10.1038/nature09201

Tabassum, D. P., and Polyak, K. (2015). Tumorigenesis: it takes a village. *Nat. Rev. Cancer* 15, 473–483. doi: 10.1038/nrc3971

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676. doi: 10.1016/j.cell.2006.07.024

Thom, R. (1972). *Stabilité Structurelle et Morphogén`ese*. Reading: Benjamin Publishing.

Thomas, F., Ujvari, B., Renaud, F., and Vincent, M. (2017). Cancer adaptations: atavism, de novo selection, or something in between? *Bioessays* 39:1700039. doi: 10.1002/bies.201700039

Trigos, A. S., Pearson, R. B., Papenfuss, A. T., and Goode, D. L. (2017). Altered interactions between unicellular and multicellular genes drive hallmarks of transformation in a diverse range of solid tumors. *Proc. Nat. Acad. Sci. U.S.A.* 114, 6406–6411. doi: 10.1073/pnas.1617743114

Trigos, A. S., Pearson, R. B., Papenfuss, A. T., and Goode, D. L. (2018). How the evolution of multicellularity set the stage for cancer. *Br. J. Cancer* 118, 145–152. doi: 10.1038/bjc.2017.398

Trigos, A. S., Pearson, R. B., Papenfuss, A. T., and Goode, D. L. (2019). Somatic mutations in early metazoan genes disrupt regulatory links between unicellular and multicellular genes in cancer. *eLife* 8:e40947. doi: 10.7554/eLife.40947

Trosko, J. E. (1987). Mechanisms of tumor promotion: possible role of inhibited intercellular communication. *Eur. J. Cancer Clin. Oncol.* 23, 599–601. doi: 10.1016/0277-5379(87)90252-5

Trosko, J. E. (2007). Gap junctional intercellular communication as a biological "Rosetta Stone" in understanding, in a systems biological manner, stem cell behavior, mechanisms of epigenetic toxicology, chemoprevention and chemotherapy. *J. Membr. Biol.* 218, 93–100. doi: 10.1007/s00232-007-9072-6

Trosko, J. E. (2016). A conceptual integration of extra-, intra- and gap junctional-intercellular communication in the evolution of multi-cellularity and stem cells: how disrupted cell-cell communication during development can affect diseases later in life. *Int. J. Stem Cell Res. Ther.* 3:21. doi: 10.23937/2469-570X/1410021

Turing, A. (1952). The chemical bases of morhogenesis. *Philos. Trans. R. Soc. B* 237, 37–72. doi: 10.1098/rstb.1952.0012

Vincent, M. D. (2011a). Cancer: a de-repression of a default survival program common to all cells? A life-history perspective on the nature of cancer. *Bioessays* 34, 72–82. doi: 10.1002/bies.201100049

Vincent, M. D. (2011b). Cancer beyond speciation. *Adv. Cancer Res.* 112, 283–350. doi: 10.1016/B978-0-12-387688-1.00010-7

Waddington, C. H. (1957). *The Strategies of the Genes*. London: George Allen & Unwin.

West, J., You, L., Zhang, J., Gatenby, R. A., Brown, J. S., Newton, P. K., et al. (2020). Towards multidrug adaptive therapy. *Cancer Res.* 80, 1578–1589. doi: 10.1158/0008-5472.CAN-19-2669

Wu, A., Zhang, Q., Lambert, G., Khind, Z., Gatenby, R. A., Kim, H. J., et al. (2015). Ancient hot and cold genes and chemotherapy resistance emergence. *Proc. Nat. Acad. Sci. U.S.A.* 112, 10467–10472. doi: 10.1073/pnas.1512396112

Zajac, O., Raingeaud, J., Libanje, F., Lefebvre, C., Sabino, D., Martins, I., et al. (2018). Tumour spheres with inverted polarity drive the formation of peritoneal metastases in patients with hypermethylated colorectal carcinomas. *Nat. Cell Biol.* 20, 296–306. doi: 10.1038/s41556-017-0027-6

Zhou, J. X., Aliyu, M. D. S., Aurell, E., and Huang, S. (2012). Quasi-potential landscape in complex multi-stable systems. *J. R. Soc. Interface* 9, 3539–3553. doi: 10.1098/rsif.2012.0434

Zhou, J. X., Cisneros, L., Knijnenburg, T., Trachana, K., Davies, P., and Huang, S. (2018). Phylostratigraphic analysis of tumor and developmental transcriptomes reveals relationship between oncogenesis, phylogenesis and ontogenesis. *Converg. Sci. Phys. Oncol.* 4:025002. doi: 10.1088/2057-1739/aab1b0

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

REVIEW
published: 30 November 2020
doi: 10.3389/fgene.2020.604528

Phenotypic Heterogeneity in Tumor Progression, and Its Possible Role in the Onset of Cancer

*Saniya Deshmukh and Supreet Saini**

Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai, India

Heterogeneity among isogenic cells/individuals has been known for at least 150 years. Even Mendel, working on pea plants, realized that not all tall plants were identical. However, Mendel was more interested in the discontinuous variation between genetically distinct individuals. The concept of environment dictating distinct phenotypes among isogenic individuals has since been shown to impact the evolution of populations in numerous examples at different scales of life. In this review, we discuss how phenotypic heterogeneity and its evolutionary implications exist at all levels of life, from viruses to mammals. In particular, we discuss how a particular disease condition (cancer) is impacted by heterogeneity among isogenic cells, and propose a potential role that phenotypic heterogeneity might play toward the onset of the disease.

Keywords: metabolism, phenotypic heterogeneity, cancer, signaling, gene regulation

OPEN ACCESS

Edited by:
Mohit Kumar Jolly,
Indian Institute of Science (IISc), India

Reviewed by:
Laasya Samhita,
National Centre for Biological Sciences, India
Subhajyoti De,
Rutgers, The State University of New Jersey, United States
Kevin Farquhar,
Precision for Medicine, United States

***Correspondence:**
Supreet Saini
saini@che.iitb.ac.in

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *09 September 2020*
Accepted: *10 November 2020*
Published: *30 November 2020*

Citation:
Deshmukh S and Saini S (2020) Phenotypic Heterogeneity in Tumor Progression, and Its Possible Role in the Onset of Cancer. Front. Genet. 11:604528. doi: 10.3389/fgene.2020.604528

INTRODUCTION: PHENOTYPIC HETEROGENEITY IN ISOGENIC CELLS

Heterogeneity in a population has been a topic of long-standing interest in populations. Darwin was interested in small variations in a particular trait in a population. He was convinced that it was these small continuous changes in a population, which were responsible for evolutionary change (Darwin, 1859). We now understand that continuous variation can be exhibited because of a trait being a polygenic character, or because of the phenotypic variation among isogenic individuals (Dudley and Lambert, 2004). However, this was not the case in the late 19th century. On the one hand, Mendelians (led by Bateson) were convinced that small, continuous variation of a trait was irrelevant as far as evolutionary change was concerned (Bateson, 1894). It was a discontinuous variant of a trait in a population, which lead to an evolution in populations. On the other hand, Mendelians, lead by Pearson and Weldon, insisted that continuous variation was sufficient to bring about evolutionary change. The debate led to the famous, and often bitter debate between the Biometricians and Mendelians.

Although early efforts at reconciliation began in the early 20th century, it was not until Fisher's publication in 1918, which reconciled the two sides (Fisher, 1918). The strategy adopted by Fisher was partitioning the variance in a phenotype among material causes.

It was recognized since Lamarck's time that the value of a trait is dictated by not just the genetic composition of the individual, but also the environmental conditions surrounding it. Hence, an isogenic population exhibited heterogeneity (**Figure 1**). Here, everything that is not genetic (i.e., DNA sequence of the individual) comprises the environment of an individual.

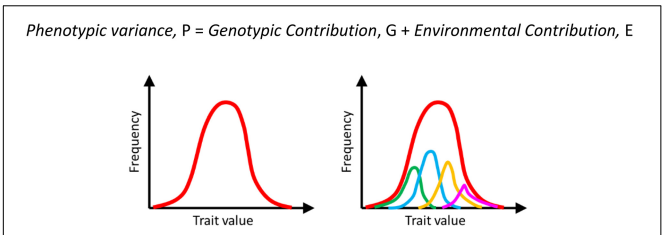

FIGURE 1 | Heterogeneity in a population is a result of both genetic and environmental contributions. The overall manifestation of this heterogeneity in a population is a distribution of the phenotype on the left (red curve). This heterogeneity comprises heterogeneity due to different genotypes (indicated by green, blue, orange, and pink on the right) and the environmental noise (variation within each genotype). For example, the phenotypic distribution represented by the blue curve represents the trait distribution among isogenic individuals. This variation is because of environmental reasons.

A century later, we understand the non-genetic variation observed by Fisher much better. We now can understand the molecular causes and can appreciate the implications of intercellular variation in dictating cellular/organismic fate. The heterogeneity between the phenotype of isogenic cells plays an important role in dictating the evolutionary fate of populations much better. These non-genetic mechanisms result in phenotypic heterogeneity.

In this article, we discuss manifestations of phenotypic heterogeneity at different scales, particularly microbes and in the case of development. Last, we discuss the mechanisms of phenotypic heterogeneity which help us understand the onset and progression of a disease condition (cancer) better.

MECHANISMS OF PHENOTYPIC HETEROGENEITY

At a mechanistic level, why does phenotypic heterogeneity occur? Broadly, it can be classified into two categories: first, isogenic cells/individuals receiving different information from the environment can lead to different manifestations of a phenotypic trait. The second cause is cells exhibiting different phenotypes despite receiving the same environmental information. The former is called extrinsic noise, and the latter, intrinsic noise (Swain et al., 2002) (**Figure 2**).

Phenotypic heterogeneity in biological systems stems from several mechanistic factors.

First, the number of molecules of an average protein in a cell varies from a few dozen to several thousand (Ishihama et al., 2008; Ho et al., 2018). Thus, in the cases where the number of molecules is small in each cell, the trajectory varies from one cell to another (Rao et al., 2002).

Second, fundamental processes in cellular functioning, like transcription, occur in short bursts with large periods of relative inactivity of a promoter, giving rise of heterogeneity in a population. This manifestation means that if we take a snapshot of a population at any instant, significantly different kinetics of production of a particular protein will be observed.

Third, isogenic cells differ because of the noise due to binomial partitioning of cellular resources at the time of division (Huh and Paulsson, 2011a,b). In fact, it is argued that much of the heterogeneity attributed to gene expression stems from the noise of partitioning.

Last, cellular regulatory networks are dictated by feedback. While negative feedback homogenizes behavior, positive feedback increases the cell-cell variation in a population (Mitrophanov and Groisman, 2008; Sauro, 2009). Moreover, positive feedback in networks could lead to manifestations, where a transient commitment to one state can mean that the cell cannot come out of the state (akin to an energy minimization landscape). Such a cell is then "trapped" in that state. The complexity of regulatory networks in cells means that multiple such minima exist. This is thought to be even more important from the perspective of multicellular eukaryotic organisms, where the regulatory networks are much larger and complex as compared to bacteria and there is active communication via signaling between cells (via chemical and physical cues) of a tissue.

PHENOTYPIC HETEROGENEITY IN MICROBIAL SYSTEMS

Starting from seminal work by Delbruck, Benzer (1953), we know that at a single-cell resolution, members of an isogenic population exhibit phenotypic heterogeneity. Benzer's work demonstrated that during infection, the λ-phage exhibits two distinct phases of life-cycle when interacting with an isogenic *E. coli* population. Delbruck quantified the distribution of burst size in a phage. See these reviews for more recent developments on this topic (Smits et al., 2006; Casadesus and Low, 2013; Ackermann, 2015).

Unlike the heterogeneity in **Figure 1**, where the heterogeneity is on a continuous scale, in this form of heterogeneity, two isogenic individuals exhibit two distinct binary responses. This suggests thresholding mechanisms in dictating life-cycle decisions, where a continuous distribution of a protein amount, for instance, can be converted into a phenotypic binary decision. Since Benzer's publications, the phenomenon of phenotypic heterogeneity, in an isogenic microbial population in a well-mixed environment, has been studied in a number of contexts.

In the context of Darwinian fitness, the exhibition of heterogeneity can confer an advantage to the population. A well-studied manifestation is the persister cells in bacterial populations (Balaban et al., 2004; Dhar and McKinney, 2007; Gefen and Balaban, 2009). A small fraction of individuals in an isogenic population, caller persisters, due to their metabolic inactivity, exhibit resistance to antibiotics. Hence, should the population encounter a temporal wave of the antibiotic, these persister cells survive, and resume growth once the wave has passed. Compare this to a microbial population where every member of the population is actively growing, rendering each individual susceptible to the antibiotic. In this context, phenotypic heterogeneity aids the chances of the population surviving an environmental catastrophe. In this bet-hedging strategy, the population places individuals in different phenotypic states, and thereby, ensuring that at least one fraction of the

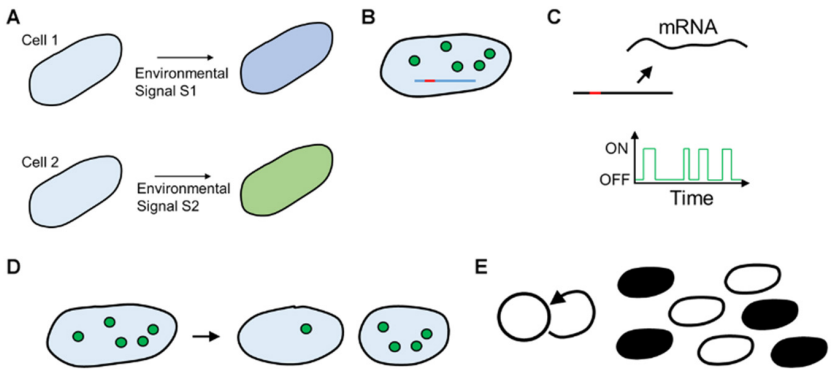

FIGURE 2 | (A) Extrinsic cause of heterogeneity. Two isogenic cells are fed dissimilar signals because of environmental noise, leading to different responses. **(B–E)** Cell–cell heterogeneity because of intracellular noise. **(B)** Small number of Transcription Factor molecules (circles) take widely different time to search for the promoter site (red) on DNA (blue line). **(C)** Cellular processes such as transcription occur with exponentially distributed periods of bursts of activity. **(D)** Cell–cell variation because of unequal partitioning at cell division, **(E)** Feedback in cellular networks leads to all (black) or none (white) states in cellular physiology.

population survives possible stress in the near future. This fitness advantage is context-dependent. In an environment where no antibiotic is encountered, persister cells will not contribute to the growth of the population. Thus, a trait of non-genetic heterogeneity, such as the commitment of a fraction of the population as persisters, is likely an adaptive response under appropriate conditions.

Such a bet-hedging strategy is used by several microbial species to counter environmental stress. Such response to anticipated stress is observed in other contexts too. In *Bacillus subtilis*, the decision to sporulate starts much before the resources run out. When exposed to starvation signals, only a fraction of cells sporulate. The remaining population switches to alternative metabolites for growth. This bet-hedging process is dictated by noise, which thus influences bacterial cell development (Veening et al., 2008). Other manifestations of this bet-hedging strategy have also been reported in other species (Galhardo et al., 2007; Sureka et al., 2008; Tiwari et al., 2010).

Recently reported manifestations of a bet-hedging strategy are more widespread than during anticipation of catastrophic events. One such manifestation has been during the transition from one carbon source to another (Ventela et al., 2003; Solopova et al., 2014). Under identical conditions, the metabolic fate of isogenic cells can have distinctly different fates which, is particularly important from the context of cancer cells (Phan et al., 2014).

In another scenario, phenotypic heterogeneity, via division of labor and cooperation among the participating phenotypic states, aids growth and survival of a microbial population (Varahan et al., 2019). A recent work (Rosenthal et al., 2018) on an isogenic population of *B. subtilis* growing in glucose demonstrated a split into two metabolic states. One converts glucose to acetate, and the other converts the accumulated acetate to acetoin (thus not enabling the accumulation of a toxic intermediate). Such division of labor is, hence, facilitated by a phenotypic heterogeneity in the population, where different parts of the population play distinct roles. The link between phenotypic heterogeneity and adaptive response has been reviewed extensively (Ackermann, 2015).

Phenotypic heterogeneity has also been observed in the context of virulence of pathogenic bacteria. In *Mycobacterium* infections, differences in the microenvironment are known to lead to divergences in the physiological states of the bacteria present in different lesions. The metabolic heterogeneity in the bacterial population, thereafter, has implications in their ability to survive stress such as drugs (Dhar et al., 2017). In *Salmonella,* phenotypic heterogeneity in the intestinal phase of infection helps the population in finding access to the niche in the body (Ellermeier and Slauch, 2007; Saini et al., 2010).

Common threads run in microbial and complex eukaryotes when studying metabolic transitions and heterogeneity. From the context of cancer, Warburg reported that cancer cells undergo aerobic glycolysis and secrete lactate (Kohler, 1973). We now know this to be almost universally true of cancer cells. In addition, the same phenomenon is also seen in yeast (De Deken, 1966) and bacteria (Wolfe, 2005; De Mey et al., 2007). The underlying principles of the logic of metabolism remain conserved across life forms, and when consuming glucose at high rates, flux channels from TCA to lactate/acetate production across bacteria, yeast, or cancer cells.

NON-GENETIC HETEROGENEITY IN METAZOAN SYSTEMS

The development of heterogeneous cell populations in multicellular eukaryotes from an embryo to a developed individual at the time of birth is a classic example of non-genetic heterogeneity. Both intrinsic and extrinsic factors dictate heterogeneity during development. The earliest representation for this diversification was proposed by Waddington, in his landscapes (**Figure 3**) (Waddington, 1956). The initial Waddington landscape was proposed for a developing embryo. However, our current understanding of cellular plasticity considers it to be integral to tissue regeneration in adults. In adult tissues, cells can revert to a progenitor phenotype (de-differentiation) or a mature cell can directly change phenotype

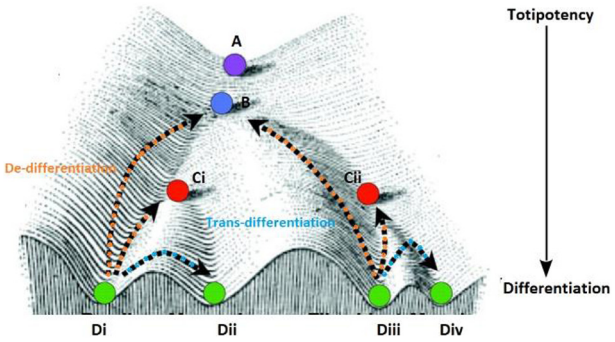

FIGURE 3 | The purple circle at the top **(A)** of the Waddington landscape represents a totipotent cell with the ability to differentiate into all cell lineages **(B–D)** present in a given metazoan system. The higher positions are representative of pluripotent zones and the low-lying zones are lineage-committed zones as denoted by the arrow. As the cell divides and travels down this landscape it retains the ability to for a select number of lineages which is dependent on the path the daughter cell takes downhill (represented by the different sub-types). All the cells represented in this figure originate from the single predecessor cell starting at position **(A)** and are genetically identical however, regulation of cell fate is shaped by the position phenotypic positions specified by points **B,C (Ci–ii)** and **D (i–iv)**, a cell occupies on the landscape. Each position signifies the effect of reprogramming on genetically similar cells for the expression of a fully differentiated or committed phenotype. Some cells can get arrested in a state at higher altitudes (like position **B** or **Cii**) before the can completely differentiate and form the stem cell population, while some cells can revert to a progenitor phenotype (from the terminal phenotypic position **Di** to **Ci**) or change positions laterally (between phenotypic position **Di** to **Dii**) on this landscape.

(*trans*-differentiation) to recuperate after unfavorable conditions (Rajagopal and Stanger, 2016).

As discussed above, noise in biological systems can be introduced due to variation in cell-intrinsic or cell-extrinsic factors (Tsimring, 2014). Cellular noise is ubiquitous and permeates the metazoan cell hierarchy. From a single progenitor, development of all cell types takes place. The scale of this challenge varies (Number of cells in *C. elegans* ∼1000, Drosophila 10^6, humans 10^9). In the face of noise, how does correct cell development and differentiation take place? Wrong developmental decisions (temporally or spatially) could have fitness consequences for individuals.

This section examines the role of non-genetic heterogeneity in the normal development of two diverse multicellular systems. The first example looks at the role of heterogeneity in the holistic development of *Caenorhabditis elegans* from a single-celled zygote. While in the second example, we discuss the human reproductive system as a particular case to study variability at the level of a specialized organ system.

Non-genetic Cellular Variability During Organismal Development

Caenorhabditis elegans is a free-living nematode with a rapid development time (3–5 days) to transition from an embryo to a fully mature adult (Frezal and Felix, 2015). An adult *C. elegans* has about 1000 somatic cells along with 1000–2000 germ cells. As the cell types in the organism are limited, a cell-by-cell reconstruction of the anatomy has been worked out to identify the lineage of each cell. *C. elegans* embryos exhibit two distinct pathways of cell fate determination. First is due to the presence of intracellular determinants in the cell (**Figure 4A**). And, the second pathway as the outcome of the association with the neighboring cells (**Figure 4B**).

We first discuss the intracellular reasons for heterogeneity. Post-fertilization, the sperm entry point polarizes the oocyte cytoplasm, and the site of localization of the male pronucleus forms the posterior end of the organism (Goldstein and Hird, 1996). The assignment of anterior and posterior poles of the embryo leads to the creation of a gradient of cellular proteins (Gotta et al., 2001). Following this, there is an asymmetrical division of cells resulting in unequally sized two cells, namely founder cell (AB) and a stem cell (P1) (Nance and Zallen, 2011). The cell polarity established by the PAR (partitioning defective) proteins mediated signaling pathway plays a vital role in the first division of the zygote occurs (Kemphues et al., 1988). The PAR proteins drive the sperm-derived centrosome to mark the posterior pole (Etemad-Moghadam et al., 1995; Guo and Kemphues, 1995; Boyd et al., 1996; Tabuse et al., 1998) while the gradient of the Gα proteins causes an imbalance in the pulling forces required for translocation of the mitotic spindle (Ajduk and Zernicka-Goetz, 2016). Another factor contributing to cellular variability is the migration of the P-granules, ribonucleoprotein complexes, which are involved in germline specification after fertilization. These granules are membrane-less organelles containing RNA (maternally expressed transcripts) and proteins associated with RNA metabolism

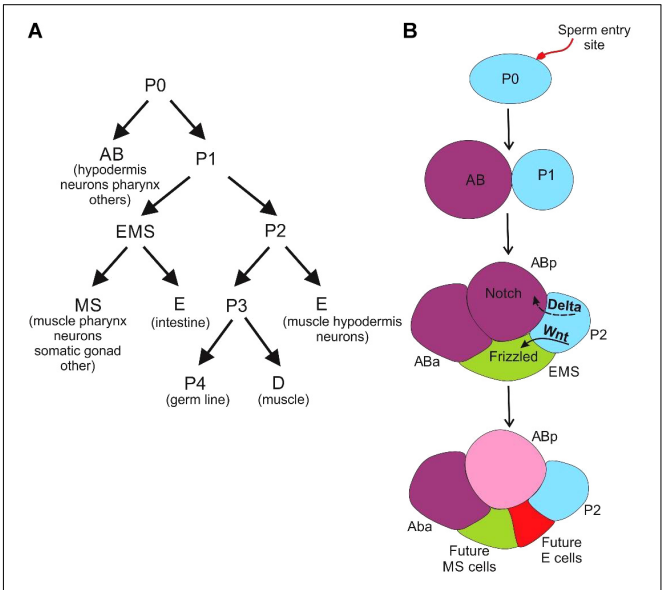

FIGURE 4 | Development of *C. elegans* **(A)** the cells derived from the asymmetric cleavage of the zygote and their future lineages **(B)** the AB divides longitudinally or perpendicular to the anterior-posterior axis to produce ABa and ABp cells whereas the P1 cell divides transversely to produce founder cell EMS and a posterior stem cell (P2). The effect of positional signaling between the cells with respect to P2 cell contact, at the four-celled stage via the Notch and Wnt pathways leads to specification of EMS and ABp cells.

(Marnik and Updike, 2019) and are involved in silencing the non-germline transcripts from accumulating near the nuclear pores in developing gametes. They, hence, play an essential role in the maintenance of germline identity and fertility (Rose and Gonczy, 2014; Seydoux, 2018). During the second division of cleavage, the resultant cells of unequal sizes lead to the inheritance of different cell fate determinants, enabling them to follow distinct developmental paths.

As mentioned above, the other cause of cell-cell heterogeneity is extracellular cues. The positioning of cells during the development of *C. elegans* is crucial for intercellular communication. The cells develop and differentiate according to the signals they receive from their neighbors (**Figure 4B**) (Evans et al., 1994; Mickey et al., 1996; Good et al., 2004; Von Stetina and Mango, 2015).

The specification of cell fates in endodermal lineage depends on interactions between P2 cell and sister cell EMS. The default state of EMS is to develop into mesoderm, which gives rise to the muscles, pharynx and other cells (Rocheleau et al., 1997; Shin et al., 1999). The association with P2 cell causes the EMS cell to polarize and rotate the spindle assembly aligning it with the anteroposterior axis. The end of the EMS cell in contact with P2 cell causes Wnt signaling asymmetry (Thorpe et al., 1997, 2000). If the gradient of Wnt signaling is equalized along with the EMS cell, the resultant daughter cells skip the endodermal fate and develop mesodermal lineages (Herman et al., 1995).

Interestingly, physical contact is not necessary for establishing Wnt asymmetry. A signaling pulse can relay this asymmetry across cell diameters which results in a small shift in the plane of cell division along the AP axis (Bischoff and Schnabel, 2006). This system of cell division and signaling induces cellular variability that aids in the assignment of distinct developmental fates (Maduro, 2010).

During the early development of *C. elegans*, unequal segregation of cellular proteins or intercellular communication gives rise to cellular variation. The process of development relies on the heterogeneity for differentiation of the multipotent predecessors to a stable cellular phenotype.

Non-genetic Cellular Variability Within a Specialized Organ System

As discussed in the case of *C. elegans*, the communication with the extracellular environment provides essential cues to cells for development. As a metazoan embryo develops, there is organogenesis, and specific organ systems are formed. The development of organ systems requires intricate coordination of intercellular signaling within and between tissues. In this section, we consider the development of the reproductive system and subsequent gametogenesis as a model to study non-genetic heterogeneity within an organ system.

The vertebrate gonad has a unique bipotential primordium, and the nature of hormonal signals received dictate the formation of testis or ovaries, governing the phenotypic sex of the organism. The genetic sex is determined in humans by the presence or absence of the Y chromosome (Brennan et al., 2013). The male pathway is dependent on the initiation of male hormones due to gonadal expression of the Y-linked gene, sry. Subsequently, in the absence of these specific testicular hormones, the female pathway is established.

Sex determination in humans is a relatively simple process as compared to sex development. The chromosomes primarily characterize the former while the latter is a multi-parametric process involving genetic, regulatory or hormonal aspects of gonadal development. The outcomes of any abnormality in the development of external or internal genital structures are clinically classified as disorders of sex development (DSD) (Makiyan, 2016). There are multiple non-genetic factors involved in DSD, unlike in case of a chromosomal abnormality. A change in the external environment like exposure to androgens or maternal tumors can act on the bipotential gonad and cause the reversal of phenotypic sex or mosaicism leading to ambiguous development of genitalia where the hormonal factors induce variability in the phenotype of cells with similar chromosomal sex (Witchel, 2018). Ovotesticular disorder, one of the rare cases of DSD, can occur in sry-negative XX males (Ozdemir et al., 2019). The bipotential gonad develops into both, genetically identical ovarian follicles and seminiferous tubules. Potential mechanisms that could be responsible for this heterogeneity in the XX (sry-) individual could be due to the activation of testis specifying genes in the absence of sry and/or inadequate expression of pro-ovary/anti-testis genes (Witchel, 2018).

The above-cited example of a non-genetic variation in the gonadal development is evident only in case of an anomaly. However, there are sources of variation in well-developed gonads. During the process of male and female gametogenesis, cellular heterogeneity is introduced, which is discussed below.

Spermatogenesis is the production of sperm from the primordial germ cells (PGC). The PGCs get incorporated into the sex cords of male embryo and remain dormant. At puberty, the testicular Leydig cells start androgens production under the influence of the Follicle-Stimulating Hormone (FSH) and the Luteinizing Hormone (LH), which are regulated by the hypothalamus (Oduwole et al., 2018). During development, the PGCs divide to form type A1 spermatogonia which establish a pool of self-renewing stem cells (**Figure 5A**). Each A1 spermatogonium divides to produce an A1 spermatogonium and the type A2 spermatogonium. The A2 spermatogonia divide and progress through A3 and A4 spermatogonia stages. This final spermatogonium stage can self-renew, die, or differentiate. The differentiation into the intermediate spermatogonium confirms commitment to becoming spermatozoa, and a subsequent mitotic division forms the type B spermatogonia. They divide to generate the primary spermatocytes which enter meiosis (Dym, 1994).

Spermatogenesis occurs in the lumen of the seminiferous tubules, where the Sertoli and germ cells produce estradiol-17ß (Carreau and Hess, 2010). The spermatocytes give rise to haploid spermatids. All these different stages of developing sperms are in the physical vicinity of the Sertoli cells to draw nutrition. The peculiarity of cellular divisions in spermatogenesis is their incompleteness. As a result, a large number of cells connected by cytoplasmic bridges are formed, which allows the exchange of cytoplasmic constituents from both parental cells thus maintaining clusters of related cells with a varied

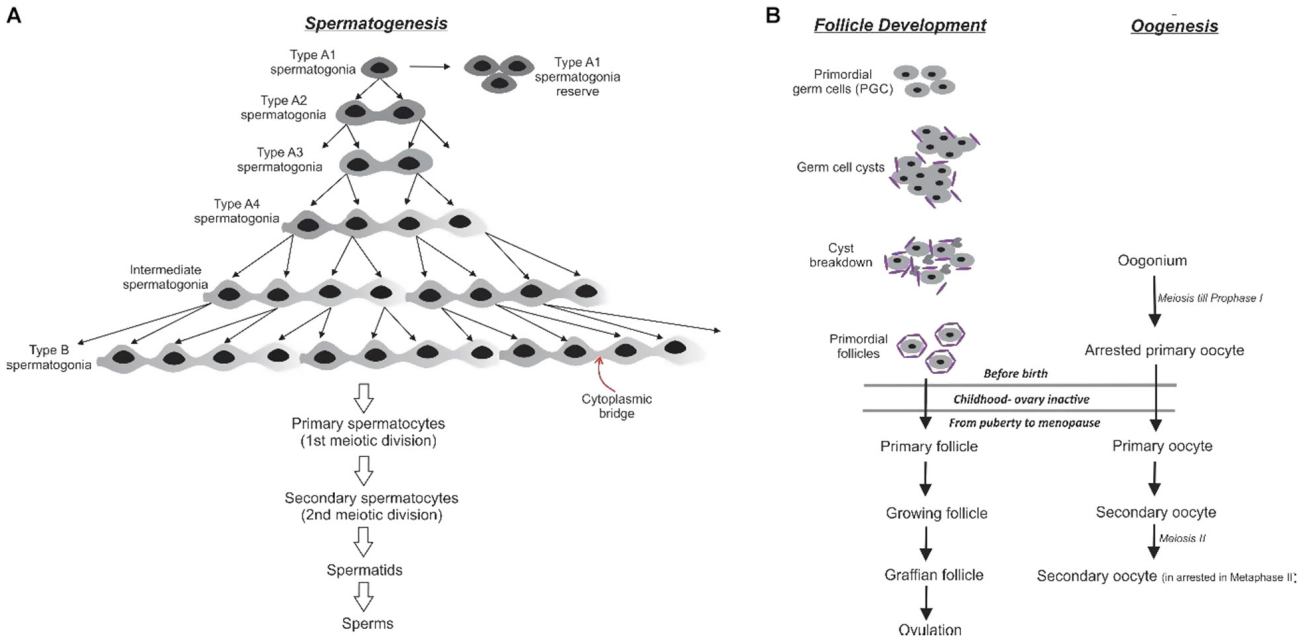

FIGURE 5 | Gametogenesis in the human reproductive system: **(A)** In male spermatogenesis there is the presence of cytoplasmic bridges that connect the pre-meiotic sperm precursor cells allows the sharing of resources via cytoplasmic gradients between diploid cells from the same progenitor germ cells **(B)** in the female follicle development, germ cell cysts are connected oogonia that assist in the development of a primordial follicle.

heterogeneous cytoplasmic composition that are diffusing into each other forming molecular gradients. All cells do not produce all macromolecules in the same concentration as they can procure them from a neighboring connected cell.

The process of oogenesis in females leads to the formation of the ovum (**Figure 5B**). The spermatogenesis in human males occurs from a population of self-renewing stem cells whereas, the oogonia in the ovary are limited as they are devoid of a pool of germline stem cells. At week 6 of gestation, the PGCs arrive and colonize the developing ovary (De Felici, 2013). Upon arrival at the ovary, the PGCs enter synchronous mitotic divisions with incomplete cytokinesis, producing an excess of interconnected oogonia, which forms clusters of related cells, germline cysts (Grive and Freiman, 2015). The functions of these aggregates are not as well characterized in the mammalian systems, and evidence suggests that mitochondria could be exchanged between members of a cyst (Motta et al., 1997; Pepling, 2012). Apoptosis regulates the number of fetal germ cells forming primordial follicles in many organisms (Matova and Cooley, 2001). In human gestation, at around 16 weeks, these cysts breakdown to smaller groups of cells and most of the oocytes undergo apoptosis (Bergeron et al., 1998; Morita et al., 2001; Pepling, 2012). It is suggested that clonally obtained cysts with genetic similarities act in unison to improve oocyte quality. There is a disparity in the future of these cells, only one cell matures to form a mature oocyte while the others act as nurse cells to nourish and act as sinks for damaged cellular components and reservoirs of mitochondria for the dominant oocyte. At the time of meiotic entry, the number of female cysts could act as a determinant of the number of primordial follicles at birth (Lei and Spradling, 2013). This reserve of follicles comprises the ovarian reserve of an adult female, which is cyclically stimulated (Grive and Freiman, 2015).

The phenotypic heterogeneity is an important facet of normal metazoan development. The creation of cellular heterogeneity is a stochastic albeit important event. It creates noise which allows expansion of cellular fate evolution. In the case of early development cited in the *C. elegans*, an unequal division of cellular determinants between daughter cells and their neighboring associations are crucial in the assignment of their fate. The human reproductive system is a specialized system that develops from a group of progenitor cells, this system exploits the cellular disparity to distribute biosynthesis load amongst genetically identical cells. Thus a specialized tissue system utilises noise for resource allocation and energy efficiency of a developing system.

Living systems are shaped by an intricate balance of deterministic laws and randomness (Monod, 1974). Control of noise is critical – as unregulated noise could cause defects too. What happens when the noise goes wrong/out of control? (Raj et al., 2010) The elimination of noise suppression leads to developmental defects and disease.

ONSET OF CANCER

Tumorigenesis and tumor progression has been thought to exemplify a form of somatic representation of Darwinian evolution. Cancer cells are a clonal population with accelerated growth and exhibit intrinsic micro-heterogeneity attributed to non-genetic factors. The switch from normal to the cancerous

state can be fulfilled by any means that is capable of randomly generating heterogeneity, conferring spontaneous individuality to daughter cells (Brock et al., 2009).

To be arrested in a state of cancerous growth, a cell has to "achieve" two goals. One, enter proliferative growth, and second, escape apoptosis, which was presumed to be acquired only through mutations. The resultant changes in the functioning of oncogenes or tumor-suppressing genes via mutations potentially upset the regulatory balance between proliferation and apoptosis, allowing cells to enter the state of cancerous growth.

Assumed to be driven by somatic mutations, which push the cell into a proliferative growth state, and suppress apoptosis. Avoiding proliferative growth is a particular challenge for vertebrates with their long lives, and therefore are likely to acquire a proliferative mutation. Given the mutation rates (for humans, 60 per genome per generation) (Kong et al., 2012; Segurel et al., 2014) and the number of cells in large vertebrates [humans, $O(10^{12})$] – it is almost a surprise that cancer is not more prevalent. Their long lives also impose a regulatory challenge where these organisms have to permit cells to proliferate when needed (e.g., an injury) and stop growing to avoid a cancerous fate. This dilemma suggests that robust mechanisms must exist in the organisms regulatory network to permit and stop proliferation, depending on intracellular and extracellular cues.

Critical among them must be a large number of inhibitory factors, which permit growth only in the strictest of conditions, presence of mitogens being one. Thus, cancer manifestation only takes place when (a) cell(s) enter proliferative growth in an uncontrolled manner, (b) the individual is unable to shed/differentiate this cells undergoing division, (c) regulatory mechanisms intended to control cell growth fail.

One of the most common manifestations of cellular proliferation is the commitment to aerobic glycolysis. Otto Warburg observed that despite the presence of oxygen cancer cells had higher glucose utilization accompanied by lactate accumulation. Warburg (1956) explained the phenomenon through dysfunctional mitochondria, perhaps through mutations. Although this was found not to be the case, research studying the onset of cancer has largely focused on somatic mutations. The idea was first proposed by Boveri, who suggested a role for chromosomal instability in the onset of cancer (Holland and Cleveland, 2009).

Irrespective of the origins, the precise reason for cellular commitment to aerobic glycolysis during cancer is not known. Aerobic glycolysis, although less efficient than mitochondrial oxidation in terms of ATP/glucose yield, permits up to 100 times faster processing of glucose (Shestov et al., 2014). It is particularly essential since tumor environments are crowded (1 cm^3 has 10^9 cells), and hence any opportunity to capture resources must be utilized. Warburg Effect has also been proposed to provide the necessary carbon flux for anabolic demands of rapidly proliferating cells (DeBerardinis et al., 2008; Levine and Puzio-Kuter, 2010; Cairns et al., 2011), or regeneration of NAD from NADPH (Vander Heiden et al., 2009; Lunt and Vander Heiden, 2011). Warburg effect also proposes the alteration of cellular signaling (Wellen and Thompson, 2010, 2012; Locasale and Cantley, 2011). In a different spirit of reasoning, the Warburg effect has been proposed to aid in invasiveness, by altering the tumor-stroma interface via the release of H^+ ions (Estrella et al., 2013).

So, what are the mutations which permit cells to enter the proliferative growth state? The first class of such mutations is one which makes cell division independent of the presence of mitogens (Olmez et al., 2015; Matson and Cook, 2017). Another class of mutations is one where the late G1 cell-cycle checkpoint fails. Escaping apoptosis – this is done in normal tissues too, and the precise signal which helps the cell escape apoptosis is unique in each microenvironment. For example, in epithelia, if cells lose physical contact with neighbors, apoptosis is triggered. Thus, mutations which help the cell escape these apoptotic signals (IGF-1, Atk etc.) (Chen et al., 2017; Wang et al., 2018).

Consistent with these ideas, many theories regarding the origin of cancer involve mutations, which upset the regulatory balance in a cell. In this context, we discuss these ideas before moving to possible mechanisms where non-genetic heterogeneity leads to onset of cancer.

Somatic Mutation Theory

In this context, several theoretical ideas regarding the acquisition of mutations and the onset of cancer have been proposed. The Two-Hit Model in 1971, through a statistical analysis of retinoblastoma of the eye, proposed that two mutations cause this cancer. The two mutations can both occur in somatic cells, or one inherited, and other in somatic cells (KnudsonJr., 1971). Boveri in his 1976 book titled "The Origin of Malignant Tumors" proposed that "the problem of tumors is a cell problem" and that cancer was due to "a certain permanent change in the chromatin complex" which, "without necessitating an external stimulus, forces the cell, as soon as it is mature, to divide again." (Manchester, 1995). Ever since, cancer has become increasingly considered as a problem of cell proliferation due to permanent changes in the "chromatin," a term that in Boveri's time was already known to contain the heritable material.

These theories developed and established the view that while there might be genetic heterogeneity in a tissue, a tumor has clonal origins (Nowell, 1976). This view has been successful in the identification of precise mutations associated with several cancer types (Sawyers, 2004).

Cancer Stem Cell Theory

Through work with stem cells and leukemia in a mice model, the cancer stem cell theory was proposed in 1994. According to this idea, cancer arises from a mutation in a stem cell (a hematopoietic stem cell in this case), which gives rise to a cancer stem cell. This cancer stem cell retains the ability for self-renewal, and also to proliferate (Lapidot et al., 1994; Rosen and Jordan, 2009).

Epigenetic Theory

The establishment of the role of epigenetics in development and gene regulation lead to the proposal of epigenetic reasons for the onset of cancer (Feinberg et al., 2006). Hyper- and Hypomethylation of DNA were demonstrated with silencing the expression of tumor-suppressing genes and activating expression

of oncogenes, respectively (Cho et al., 2000; Jones and Baylin, 2002; Sato et al., 2003).

Tissue Organization Field Theory (TOFT)

Cancer is a tissue-based disease, and that proliferation is the default state of all cells (Soto and Sonnenschein, 2011). The tissue organization field theory (TOFT) states that carcinogenesis takes place at the tissue level of biological organization, as does normal morphogenesis. In this view, a cell is not necessarily a basic unit – for example, without interaction with the ureteric bud, kidney development will fail. Thus, tissue, not a cell, should be viewed as a basic unit of multicellular life. The second premise of TOFT is that the default state of all cells is proliferation.

For instance, when mice are "initiated" by feeding small quantities of a carcinogen, a coal tar derivative, the mice develop tumors long after this exposure (Friedewald and Rous, 1944). This is presumably caused by some change brought into the cells because of the exposure to the carcinogen. However, what was this change which increased the cellular propensity to go into a tumor state? Similar observations exist with experiments with *in vitro* cells (Kennedy et al., 1984). When irradiated with X-rays, it takes these cells many generations to become cancerous. What causes this long duration between exposure and the cells becoming cancerous? The kinetics of this process and the underlying link with the mechanisms that trigger cancer are unclear. This conundrum is laid out in Brash and Cairns (2009a,b) as:

"The prime mystery in carcinogenesis remains the very first step because it is hard to imagine how the numerous genetic changes found in cancer cells could have been produced in any cell as the result of a single exposure to a DNA-damaging agent, or why months or years should have to elapse before the effect of these changes is observed" and "...the picture that emerges from the classical studies of the epidemiology of human cancers and of experimental carcinogenesis in animals is hard to reconcile with what has been learnt about mutagenesis in simple systems such as the bacteria. Initiation seems to be far too efficient to be simply mutagenesis of certain oncogenes and suppressor genes, and the subsequent time-dependent steps are even more obscure."

Hence, an alternate paradigm regarding the onset of cancer is needed.

COULD CANCER ONSET BE TRIGGERED BY NON-GENETIC HETEROGENEITY?

In 1932, the American geneticist Sewall Wright proposed the concept of a fitness landscape (Wright, 1932). Several representations of a landscape exist (Kaplan, 2008), in one, the genetic identity of an individual could be mapped on an N-dimensional space (called sequence space), where each dimension corresponds to a particular locus on the genome. The $N + 1th$ dimension represents the fitness of the individual in a particular environment. Wright proposed that such a structure be called a fitness landscape and that among the topological features of this structure is multiple local optima of fitness.

The analogy can be extended to networks too. The $N + 1$th dimension represents the stability of the network, which can represented as inverse of the potential energy of the system. In such a representation, the N axis represents the amounts of the N regulatory proteins. Regulatory networks are highly interconnected structures, and their potential and stability have been a subject of various studies. Even the simplest regulatory/signaling network where two proteins are mutually repressing has two stable and two unstable steady states. From the perspective of this discussion, this implies that the system has multiple energy minima states available to it. In such a context, the starting point and the consequent noise has a large bearing on the eventual steady state of the system. The manifestations of this idea, in higher dimensionality, could offer many more stable steady-states for the cellular regulatory logic.

Interestingly, at the time of the proposal, the idea of fitness landscapes consisting of multiple peaks and valleys was fiercely contested by the Fisher (1941; Provine, 1986). He proposed that increased dimensionality of the landscapes decreased the probability that a particular genotype corresponded to one of the maxima or minima in all the dimensions of the landscape. Thus, while the concept of valleys and troughs was acceptable in lower dimensions, at an organismal level, the high dimensionality of the structures meant that there was only one global maximum.

In the context of cellular networks, therefore a commitment to alternate steady state, leading to a cancer phenotype remains a distinct possibility. The most common manifestation in cancer is the commitment to cell proliferation and escape from apoptosis. In order to facilitate rapid division, cancer cells commit to aerobic glycolysis. At the same time, the cells escape apoptosis. The molecular pathways dictating cellular commitment to these fates are well understood.

Before we discuss the possibility of cellular commitment to proliferation and escape apoptosis, we discuss two cases where phenotypic heterogeneity has been demonstrated to have adaptive fitness.

Rutherford and Lindquist (1998) demonstrated that a mutant *Hsp90* in Drosophila leads to phenotypic abnormalities in the development of the fly. The observation resulted from a competition for role of *Hsp90* in developmental and a cell stress chaperone. In normal conditions, Hsp90 buffers the variation in a population, which only manifests neutrally. However, when the function of Hsp90 is compromised (mutations or pharmacology), phenotypic variation manifests. Selection acts on this variation, and the selected variants continued to express the variant trait, even after the restoration Hsp90 function. This study provided evidence that genetic backgrounds, which facilitated a greater variation among individuals, were more evolvable. This phenomenon was shown to be a general manifestation of phenotypic heterogeneity across life forms (Queitsch et al., 2002).

Collins and coworkers demonstrated that phenotypic heterogeneity due to transcriptional noise could aid adaptation too (Blake et al., 2006). The authors designed an engineered promoter in yeast, and working with a variety of TATA boxes in the promoter region, demonstrated that promoter designs which exhibited greater variability in the expression of the

downstream gene also conferred a greater ability to withstand acute environmental stress.

These results establish the significance of phenotypic heterogeneity and evolvability of a population. These two studies establish the concept that selection acting on a population, chooses the best available phenotype. The survival of this variant, in case of stress, provides the opportunity to a fraction of the population to pick up a mutation and "solidify" this trait.

A report from Paul Rainey's group followed, demonstrating with *Pseudomonas*, that such phenotypic heterogeneity can be evolved "*de novo*" in a population fairly rapidly. Hence, the link between phenotypic heterogeneity and evolvability was firmly established. Other examples of noise facilitating adaptation exist (Acar et al., 2008; Cagatay et al., 2009).

If cancer can be triggered by phenotypic heterogeneity, the cellular commitment must be so that the cells escape apoptosis, differentiation, and commit to proliferation. For this to manifest, signaling pathways have to be channeled to suppress apoptosis, and metabolism has to be channeled to drive cell division. We next discuss both these facets. We start with a discussion on metabolism.

HIF1 Mediated Feedback and Commitment to Glycolysis

It is well established across different scales of life that faster growth is supported by fermentation and not TCA, despite the lower efficiency of fermentation compared to aerobic respiration. Several ideas have been proposed to explain this, including, surface area availability (Szenk et al., 2017), protein production cost (Kafri et al., 2016), rate of release of energy. This phenomenon is known to be present in microbes (overflow metabolism), yeast (Crabtree effect), and humans (Warburg effect). Interestingly, the Warburg effect is a hallmark of cancer cells. Thus, the first step toward phenotypic heterogeneity "pushing" a cell toward cancer phenotype is a commitment toward aerobic glycolysis. For this purpose, there is active suppression of mitochondrial activity, and the glycolytic pathway is activated in order to channel greater glucose flux through them. The molecular link that facilitates this is the Hypoxia-inducible factor 1 (HIF1).

HIF1 is a dimer of HIFα and HIFβ (Wang et al., 1995; Yang et al., 2005). The presence of oxygen results in the active degradation of HIFα via TCA intermediates (Chan et al., 2005). However, in low oxygen, HIF1 actively represses the expression of pyruvate dehydrogenase kinase and upregulates enzymes in glycolysis (Kim et al., 2006). This double-negative positive feedback is a hallmark of instability in the cellular regulatory network and can lead to altered commitments of individual cells among a population (**Figures 6A,B**).

Grb2 and Plcγ1 Competition for FGFR2 and Cell Proliferation

Recent reports suggest that fibroblast growth factor receptor 2 (FGFR2) expressing cancer cells, which have a low abundance of the protein Gbr2, show a high abundance for metastasis. The Gbr2 and Plcγ1 (phospholipase Cγ1) in a cell compete for access to the FGFR2 protein. Reduced Gbr2 protein levels in the cell, lead to elevated access of Plcγ1, leading to downstream activation of the Atk signaling pathway, eventually leading to cell proliferation (**Figure 6C**) (Timsah et al., 2014, 2016). This demonstration of fluctuations in protein numbers leading to cell fate suggests that it is conceivable that healthy tissue can, via stochastic fluctuations, escape the control of growth and go into a proliferative mode of existence. The competition for an active site between two proteins constitutes the regulatory topology of a cell. In cases like the Gbr2 and Plcγ1 competition, the regulatory topology manifests as the representative of a topology in a cell.

Dual Role of Transcription Factor Myc

Myc is one of the transcription factors which controls the expression of genes necessary for cell proliferation (Henriksson and Luscher, 1996; Roussel et al., 1996; Bouchard et al., 1998). However, the precise regulatory network dictating this activation has a more complex topology (**Figure 6D**). Myc, in a dimer with Max, is an activator of cell proliferation proteins. However, a dimer of Max (or a dimer of Max and one of its many partner proteins), acts as the repressor of the same genes. Thus, the precise control of proliferation or quiescence is controlled by the precise levels of these transcription factors. In contrast with its role in proliferation, Myc is also known to be a regulator of apoptosis in mammalian cells (McMahon, 2014). Myc's role in apoptosis is achieved via the integration of a large number of cell cycle signals (Prendergast, 1999).

The key features of all the regulatory cases discussed above is the presence of bistability in the networks. One of the key characteristics of a cancer cell is proliferation. All the above networks show that control of apoptosis and proliferation is controlled via networks, which could commit to one state or the other, depending on the precise state of the system Numerous check points control cell division, and only when all fail will a cell fall into the proliferative state. Once this rare event of a cell evading cell-cycle control happens, and getting "trapped" in a proliferative state, selection acts to select a mutation which "freezes" this proliferative state.

Tumorigenesis is associated with abnormal cell proliferation, abrogation of apoptotic processes, invasiveness and metastasis. The concept of genetic determinism and non-genetic heterogeneity are intertwined in cancer physiology and progression. The genomic instability leads to genetic heterogeneity in cancer. Whether a novel genotype is the premise for a novel phenotype or vice versa remains as the inherent paradox in cancer evolution (Frank and Rosner, 2012). Thus, the variation observed within a population of clonal cells, within a tumor cannot be explained on the basis of genetic mutations alone.

NON-GENETIC HETEROGENEITY IN CANCER

We highlighted the role of phenotypic heterogeneity in normal developmental processes. Historically, Virchow first observed pleomorphism of cancer cells within tumors establishing

FIGURE 6 | (A) A double negative feedback loop is inherently bistable. The steady state of the system depends on the starting position of the network in the state space. **(B–D)** Regulatory topologies of metabolism **(B)**, signaling **(C)**, and gene regulation **(D)**, which could likely have distinct steady states. The different steady states reflect cellular commitment to proliferation or lack of (or apoptosis).

intratumoral heterogeneity of cellular phenotypes (Almendro et al., 2013). This finding led to a series of studies that have since demonstrated the presence of distinct subpopulations of cancer cells within tumors (Makino, 1956; Fidler, 1978; Heppner and Miller, 1983; Lawson et al., 2018; Keller and Pantel, 2019).

A novel genotype exhibits a new phenotype (Bronstein and Akil, 1990). According to the somatic mutation theory, the evolution of cancer proceeds by the acquisition of genetic changes. In recent years, there has been significant evidence claiming that new non-heritable phenotypic variants can precede genetic variants in cancer evolution (Frank and Rosner, 2012). If the phenotypic variants in a clonal population develop resistance or an advantage over other sub-populations under selective pressures, like changes in the microenvironment or drug treatment, could lead to the selection of a new genetic variant (Yang et al., 2010; Altschuler and Wu, 2010). The phenotypic heterogeneity improves the cellular response to environmental challenges during tumorigenesis and enhances the rate of evolutionary changes (Frank and Rosner, 2012). The complexity of cancer makes it difficult to state if the chronology of phenotypic and genetic variants and their exact contribution to the processes that lead to the progression of cancer.

In this section, we discuss cellular processes which contribute to the phenotypic heterogeneity among cancer cells. These facets of cellular variability are of importance in improving our understanding of cancer progression and design of therapeutic measures.

Signal Transmission and Response

The signaling pathways create a communication web to simultaneously relay information within and between cells, connecting tissue systems to restore homeostasis within the metazoan system (Zhang and Liu, 2002; Guruharsha et al., 2012; Schrier et al., 2016; Schwartz et al., 2016). Cross-talk within the tumor microenvironment (TME) can occur through a diverse range of direct mechanisms like cell-to-cell contact via adhesion molecules, gap junctions, or indirect mechanisms through paracrine signaling by cytokines, extracellular vesicles etc. (Dominiak et al., 2020). Thus, signaling can act as a cause of and be affected due to intratumoral heterogeneity among clonal cells. The non-homogenous response to signals within the TME, can be viewed as a bet-hedging strategy. The diversity in response by malignant cells provides a chance for a fraction of the cells to evade therapy and thereafter lead to a possible relapse (Stumpf and Pybus, 2002; Kussell and Leibler, 2005). There are multiple complications in understanding cell-to-cell communication networks within the TME, as intracellular signaling within individual cells is heterogeneous. Therefore, there is no synchronization of intercellular signals and, this lag in the relay of signals introduces non-genetic heterogeneity in the TME (Thurley et al., 2018).

The aforementioned examples illustrate phenotypic heterogeneity in different cancers and signaling pathways. The mitogen-activated protein kinase (MAPK) pathways are vital, evolutionarily conserved and link extracellular signals to fundamental processes like growth, apoptosis and differentiation (**Figure 7A**). Also, MAPK signaling is often the most misregulated in cancer. There are two arms of MAPK signaling, the ERK pathway and stress-activated MAPKs cascades. The ERK pathway is most well understood of the

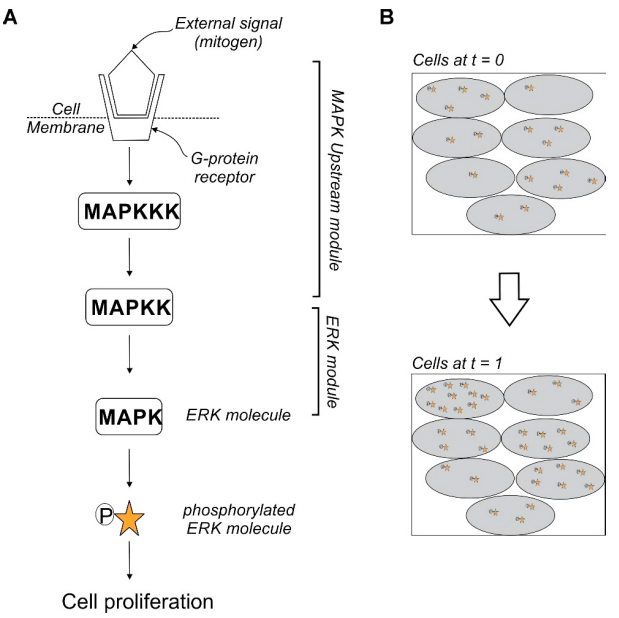

FIGURE 7 | Heterogeneity of MAPK signaling **(A)** depicts the phosphorylation of the primary MAPK molecule ERK (extracellular-signal-regulated kinase) by a MAPKK (mitogen-activated protein kinase kinase). The activation of the core ERK module relays the stimulus to the nucleus for cell proliferation. Factors like ligand availability at the surface and reaction times of other reactions before the phosphorylation of ERK are the upstream module which, effect the core ERK reaction. **(B)** Cartoon of a time-lapse snapshot of cells shows that the concentration of activated ERK at $t = 1$ is dependent on the number of activated ERK molecules at $t = 0$. There is a heterogeneity in the relay of stimulus across cells in the same sub-population. This distributes the response to external stimulus within clonal cells.

mammalian MAPK pathways and is affected in approximately one-third of all human cancers (Dhillon et al., 2007). An extracellular stimulus can activate the ERK (extracellular-signal-regulated kinase) signaling causing it to translocate to the nucleus where the signal is converted to an appropriate output and wired to the next cell. A variation in the levels of ERK activation between clonal cells in culture has been observed (Filippi et al., 2016). The different reaction rates, initial concentration of core signaling molecules and configuration of the upstream signaling cascades feeding into the ERK module could be a potential source of heterogeneity between cells (**Figure 7B**) (Filippi et al., 2016). These intercellular factors reduce the impact of the external signal by managing the distribution of primary MAPK activity on a cell-to-cell basis leading to signal distortion between clonal cells and generating a diverse response within the population to the same stimulus. The variability in response to external stimuli of clonal cells in a tumor distributes their risk of succumbing to immune responses of the body or therapeutic interventions.

The epidermal growth factor receptor (EGFR), from the receptor tyrosine kinases (RTKs) family, has crucial roles in Glioblastoma (GBM) development and progression (Brennan et al., 2013). RTKs pathways are crucial in the regulation of cellular signaling that controls proliferation, metabolism and response to environmental cues (Gschwind et al., 2004; Lemmon et al., 2014). The intertumoral mutational patterns of GBM are stereotypical and less heterogeneous but, striking histological variation s displayed by individual tumors (Lawrence et al., 2013; Sturm et al., 2014). Most GBM samples show the presence of different amplified RTKs, primarily, either EGFR (40–50%) or platelet-derived growth factor receptor alpha polypeptide (PDGFRA) (15%) but a small fraction show both (Furnari et al., 2015). Concurrent amplification of PDGFRA with EGFR is found to occur in 5% of GBM samples (Chakravarty et al., 2017). The RTK cell to cell variation is high. This variation redefines the tumor subpopulations based on the receptor and resultant signaling heterogeneity. The absence of uniformity in RTK introduces a high degree of redundancy in downstream interactions with Phosphoinositide3-Kinase (PI3K) and Mitogen Activated Protein Kinase (MAPK) pathways. The cell-to-cell variability due to heterogeneous RTK expression affects signaling response to RTK-inhibitors, leading to resistance to single target therapeutic approaches.

The Notch signaling pathway is involved in the determination of cellular identity and can elicit tumor suppressive or oncogenic outcomes depending on the simulation (Koch and Radtke, 2007; Ntziachristos et al., 2014). During lung development, the Notch pathway acts as a suppressor of ectopic neuroendocrine differentiation of precursor cells averting small-cell lung cancer (SCLC) (Morimoto et al., 2012; Pietanza et al., 2015). However, endogenous activation of Notch signaling causes neuroendocrine to non-neuroendocrine fate switch in 10–15% tumors. This non-neuroendocrine, Notch-active phenotype interspersed with the tumor of small-cell lung cancer is slow-growing and acts as trophic support for the neuroendocrine phenotype promoting oncogenesis (Lim et al., 2017). This phenotype is relatively chemoresistant, generating a subpopulation of persisters via activation of Notch signaling. These cells can survive chemotherapy and unless it is combined with Notch inhibition there will be inefficient tumor suppression, and relapse in the pre-clinical stages of SCLC.

There is accumulating evidence for the presence of a slow-cycling, dedifferentiated and invasive subpopulations of melanoma cells (Hugo et al., 2016; Tirosh et al., 2016; Fallahi-Sichani et al., 2017). The melanoma cells oscillate between two interchangeable phenotypes using microphthalmia-associated transcription factor (MITF)-rheostat signaling, namely, the proliferative state with high levels of MITF expression (MITFhi) or invasive phenotype with low levels of MITF (MITFlow) (Hoek et al., 2006; Zipser et al., 2011; Kemper et al., 2014). MITF maintains tumor homeostasis by regulation of cell cycle and suppression of apoptosis. The IFNγ signaling plays a crucial role in the regulation of the cytokine-mediated immune signaling. The hypoactivation of the IFNγ signaling inactivates the immunogenicity of the melanoma cells, whereas the hyperactivation creates a dedifferentiated and invasive phenotype which is a stress-induced persister population (Bai et al., 2019). This subpopulation of cells with changes in MITF levels or IFNγ signaling forms a pre-resistant cell phenotype. Together, MITF and IFNγ modulate the oscillation of cell states with constant shifts in cell phenotype of the tumor population to develop immunotherapy resistance.

Non-homogenous Nutrient Supply and Metabolism

The cancerous cells require rapid energy and nutrition for their uncontrolled proliferation. They reform their metabolism, especially glucose, to suit their changing needs and altered microenvironment. Most cancer cells, regardless of oxygen availability, convert glucose to lactate. The glycolytic switch occurs during early carcinogenesis (Vander Heiden et al., 2009). The activation of oncogenic signaling reprograms cell metabolism, to scale up the precursors for macromolecule biosynthesis, for the accumulation of biomass during cell proliferation (Hsu and Sabatini, 2008; Schulze and Harris, 2012). Here we discuss the effect of spatial organization on metabolic reprogramming of individual cells.

The TME is an ecosystem formed by tumor and stromal cells, extracellular matrix (ECM), and secreted factors (Liotta and Kohn, 2001; Shojaei and Ferrara, 2008). The tumor tissue exhibits unique levels of cell differentiation, proliferation, vascularity, immunosuppression, and invasiveness in different pockets and contributes to the phenotypic diversity within subclonal populations (Zuazo-Gaztelu and Casanovas, 2018). The accelerated division of malignant cells causes the tumor microcapillaries to become tortuous and irregularly spaced. They develop pores of different sizes and become hyper permeable, causing the blood to leak plasma and become more viscous. These leaky vessels have reduced nutrient and oxygen-carrying capacity within different sections of the tumor (Chauhan et al., 2012; Martin et al., 2016).

Unlike healthy tissue, compressed blood vessels in tumors leave large tissue volumes without blood flow and oxygen (Baish et al., 2011; Stylianopoulos and Jain, 2013). Thus, as one moves

deeper into the inner mass of the tumor oxygen and nutrient supply decrease due to their distance from vascularization, making the tumor ECM heterogeneous (**Figure 8**) (Polyak and Weinberg, 2009; Polyak et al., 2009; Hanahan and Weinberg, 2011; Quail and Joyce, 2013).

The local irregularities and inefficiencies in the vasculature cause the initiation of unorderly angiogenesis by the tumor cells to draw nutrients. The angiogenesis is switched on by vascular endothelial growth factor (VEGF) signaling, which is upregulated by hypoxia via hypoxia-inducible factor 1α (HIF-1α) and oncogene signaling. Low extracellular pH is another outcome of poor vasculature. Hypoxia also leads to production and build-up of acidic by-products of metabolism such as lactate (Helmlinger et al., 2002; Ward and Thompson, 2012). The acidification of the microenvironment leads to higher proliferation, invasiveness and apoptosis of normal cells.

The cells in a solid tumor have the task to multiply in an environment with heterogeneous zones of hypoxia and pH. They survive by modulating their metabolism and generating diverse phenotypes to secure their resources for rapid proliferation. The level of metabolic reprogramming is fine-tuned to the local conditions like nutrient availability, oxygenation and pH. The over-production or upregulation of VEGF results in better vascularization leading to a subsequent rise in the availability of nutrients and oxygen for the tumor as a whole (Nishida et al., 2006). However, all cells do not need to go into VEGF production overdrive. The mere proximity to the over-producers of the relevant cytokines like VEGF can help non-producer cells conserve energy, and they can hitchhike at the expense of other cancer cells (Kaznatcheev et al., 2017). The free-rider phenotype has an advantage over the producer phenotype and can take over the population by harvesting more nutrients and oxygen from the neovasculature at lower energy expenditure. The different phenotypic strategies used by cells creates a metabolic heterogeneity within the tumor population.

Mobility and Metastasis

Epithelial–mesenchymal plasticity (EMP) is a cellular mechanism, relying on the conversion between epithelium and mesenchyme in developmental milestones, like gastrulation, neural crest formation. Epithelial cells lack mobility with respect to their environment, whereas the mesenchymal cells are mobile (Larue and Bellacosa, 2005). EMP measuredly exploits this disparity in the innate properties of the two cell types during development for migration of cells and the formation of cavities. Similar physio-pathological transitions occur in cancers where there is increased motility and invasiveness during the progression of epithelial tumors.

The non-genetic phenotypic heterogeneity in cancer cells can arise due to reversible processes, epithelial-mesenchymal transition (EMT) and mesenchymal-epithelial transition (MET). The EMP is used by cancer cells for functions like metabolic reprogramming, cell proliferation, metastasis.

Cancer cells within a solid tumor exhibit widespread epithelial-mesenchymal heterogeneity and express epithelial and mesenchymal markers or co-express both and can acquire an epithelial (E), a mesenchymal (M), or one of the hybrid epithelial-mesenchymal (hybrid E/M) phenotypes respectively (Pereira et al., 2015; Hong et al., 2018; Stylianou et al., 2019). The tumor cells can exist in either of the stable phenotypes and can transition spontaneously into another state (Ruscetti et al., 2016). A series of snapshots of any clonal cell population will reveal that genetically identical cells can exist in different EMP phenotypic states over time (Tripathi et al., 2020). The dynamics of phenotypic plasticity between E and M phenotypes are affected by the initial state of the sub-population and random partitioning of parent cell biomolecules (transcription factors, regulatory proteins, miRNA and mRNAs) at cytokinesis.

Similarly, the CSC is a dynamic phenotype and can use the EMP to oscillate between the stem cell- like and differentiated phenotype (Mani et al., 2008; Zomer et al., 2013). The spatiotemporal dynamics of cells with varying EMP can lead to the formation of distinct patterns of phenotypic and functional heterogeneity of the CSCs within the tumor microenvironment (Jolly and Celia-Terrassa, 2019).

The Notch signaling pathway and EMT-inducing signals such as TGF-β together lead to distinct localization of CSCs with varying EMT phenotypes in the tumor. The Notch signaling is activated by binding of Delta or Jagged ligands on the other communicating cell. The ligand displayed by the cell decides the cellular phenotype. At low levels of both Delta and Jagged, cells exist in epithelial phenotype. The subsequent increase in the production of the ligands activates the Notch signaling, and EMT generates more number of cells in the E/M and M phenotypes.. At sites within the tumor, clusters of cells in the hybrid E/M or M phenotype are observed when Notch-Jagged signaling dominates whereas in case of cells segregate when Delta ligand is predominant (Boareto et al., 2016). Bocci et al. (2019) have modeled the diffusion of EMT-inducing signals and Notch signaling controlled non-cell autonomous switch between EMT and CSC fate decision making to reveal a distinct pattern of localization of the mesenchymal CSCs at the invasive edge, while the hybrid E/M CSCs reside in the tumor interior. The Notch-Jagged signaling stabilizes the hybrid, increases the chances of hybrid spatial proximity and expands the CSCs in a tumor (Bocci et al., 2019). The E/M hybrid is associated with higher tumor-initiating ability, a predominant trait of CSCs and drug resistance (Jia et al., 2015; Grosse-Wilde et al., 2018; Tieche et al., 2019).

Drug Resistance

Intratumoral heterogeneity leads to the creation of different tumor subpopulations to sustain growth. A hallmark of different phenotypes of clusters of cancer subpopulations is a wide range of responses to therapeutic agents. The differential response of malignant cells can be attributed to various genetic and non-genetic sources. The CSCs were the first tumor subpopulation to be explored for resistant phenotypes (Lapidot et al., 1994). These quiescent cells help cancer acquire therapy resistance and relapse potential after the initial round of treatment (Al-Hajj et al., 2004; Fabian et al., 2013). We now view the CSCs as a tumor initiating phenotypic state which

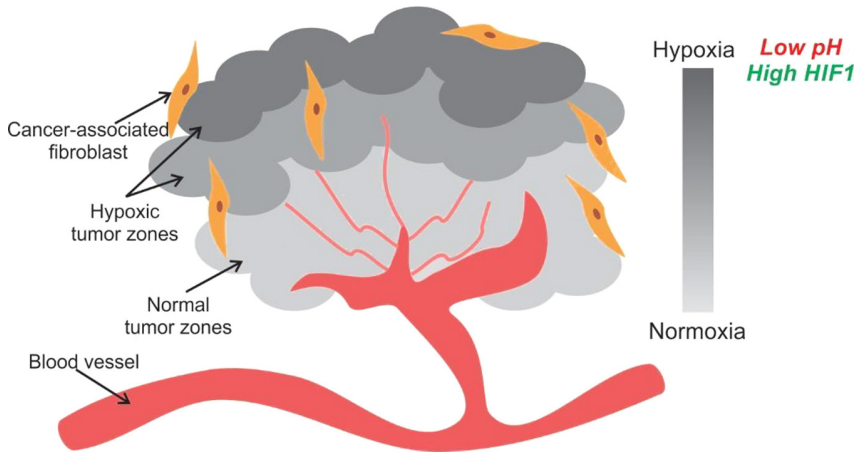

FIGURE 8 | Tumor hypoxia is observed due to spatial arrangement of cells. The tumor cells away from a blood vessel are in a hypoxic and acidic environment. They generate energy for proliferation by aerobic glycolysis. Some cells in the hypoxic zone secrete HIF1 and VEGF to initiate angiogenesis for securing nutrient supply.

has variable markers depending on the type of cancer. The CSC hierarchies are plastic, and interconversion between the CSC and non-CSC is possible due to environmental stimuli (Batlle and Clevers, 2017).

The CSCs can be shielded from blood-borne therapies due to heterogeneous blood flow due to focal hypoxia (Martin et al., 2016). The physical sequestration of a small sub-group of tumor cells can act as seeds for relapse. There are multiple phenotypes used by persister systems to escape therapy and remain quiescent for a relapse. We highlight a few examples where the phenotypic heterogeneity of the cancer cells that helps in the acquisition of drug resistance.

Most chemotherapies target a particular receptor and its downstream effectors. There are cell signaling pathways with a heterozygous expression of surface receptors across cells. The expression of the receptors for a signaling ligand is stochastic, where some cells express either or both receptor types (Patel et al., 2014). Differential expression of ligands and cell surface receptors on a cell within a tumor builds the immunity of the tumor drug and bypasses the treatment. Mosaicism in the expression of cell receptors is widely reported, ranging from glioblastomas to non-small cell lung cancers (Hegde et al., 2013; Iqbal and Iqbal, 2014; Lee et al., 2014). Thus cells exclusively expressing a cell receptor that is not targeted by the therapeutic agent manage to tide over the treatment and cause a relapse. The signaling pathways are common between cancer and normal cells prohibiting the use of multidrug chemotherapy in many cases leaving the door open to relapse initiated by the persister pool of cells. The epithelial-mesenchymal plasticity (EMP) of cells is used by malignancies to disseminate to distant organs and in the metastasis of solid tumors. The malignant cells colonize the secondary sites and reacquire their adhesion properties. The EMP is implicated in contributing to the stemness of the tumoral mass by making it more resistant to cancer therapies (Fischer et al., 2015; Zheng et al., 2015) and evading the immune system (Kudo-Saito et al., 2009). The phenotypes created by EMP differ in their physiological properties like tumor-seeding and sensitivity to drugs (Grosse-Wilde et al., 2018; Tieche et al., 2019). The E/M hybrids form clusters of migratory cell clusters which are more resistant to apoptosis and possess an increased metastatic propensity as compared to cells with a complete mesenchymal phenotype. Thus, the EMP of cells does not only confer mobility on tumor cells but it also contributes to drug resistance.

The cancer cells undergo metabolic reprogramming to switch from OXPHOS to glycolysis which leads to increased glucose uptake to compensate for inefficient breakdown process (Bhattacharya et al., 2016; Potter et al., 2016). The activation of oncogene signaling inevitably spikes the levels of reactive oxygen species (ROS) which can cause apoptosis but are effectively managed by exploiting the inherent cell antioxidant systems activation (Irani et al., 1997; Tanaka et al., 2002). Cells use nicotinamide adenine dinucleotide phosphate (NADPH) as the antioxidant sink for ROS from the glutathione (GSH) and thioredoxin antioxidant systems. The regulation of the NADPH pool is crucial for stimulating the proliferation and survival pathways in malignant cells (Patra et al., 2013; Ciccarese and Ciminale, 2017).

There are some CSC subpopulations in tumors with higher expression of antioxidant genes and low ROS levels which show resistance to radiation therapy (Tanaka et al., 2002). For example, aldehyde dehydrogenase (ALDH) functions as an antioxidant to protect aldehydes from oxidation from byproducts generated by ROS, and the drug-tolerant persister phenotypes are ALDH high (Pribluda et al., 2015).

As illustrated by the examples above, phenotypic heterogeneity leads to the formation of a residual population post a therapeutic intervention. On account of a phenotypic variation from the other tumor cells, this subpopulation is capable of acting as seeds for relapse. There are many different routes by which some cancer cells manage to escape complete elimination. However, these persisters exploit the inherent noise in the system and use it as an asset for survival. Thus, the molecular networks of eukaryotic cells offer a myriad of opportunities for phenotypic

heterogeneity to "lock" cells into phenotype, which can then lead to newer evolutionary pathways, including cancer.

CONCLUSION

Historically viewed as triggered by a mutational event, recent evidence has shaped our understanding regarding non-genetic factors that can trigger cancer. In this view, cell-cell heterogeneity in gene expression leading to altered metabolic states, signaling pathways, resistance states can all "lock" a cell in a state of rapid growth. Thereafter, selection can act on this phenotype, which is then fixed by a mutational event. In this context, we present a survey of possibilities of non-genetic heterogeneity in cancer onset and progression. Experimental manifestation of these possibilities will be an important direction of future work in this area of research.

AUTHOR CONTRIBUTIONS

SD and SS wrote the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

SD is supported by the Institute Post Doctoral Fellowship (IPDF) at IIT Bombay.

ACKNOWLEDGMENTS

The authors thank the reviewers for their comments.

REFERENCES

Acar, M., Mettetal, J. T., and van Oudenaarden, A. (2008). Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* 40, 471–475. doi: 10.1038/ng.110

Ackermann, M. (2015). A functional perspective on phenotypic heterogeneity in microorganisms. *Nat. Rev. Microbiol.* 13, 497–508. doi: 10.1038/nrmicro3491

Ajduk, A., and Zernicka-Goetz, M. (2016). Polarity and cell division orientation in the cleavage embryo: from worm to human. *Mol. Hum. Reprod.* 22, 691–703. doi: 10.1093/molehr/gav068

Al-Hajj, M., Becker, M. W., Wicha, M., Weissman, I., and Clarke, M. F. (2004). Therapeutic implications of cancer stem cells. *Curr. Opin. Genet. Dev.* 14, 43–47.

Almendro, V., Marusyk, A., and Polyak, K. (2013). Cellular heterogeneity and molecular evolution in cancer. *Annu. Rev. Pathol.* 8, 277–302. doi: 10.1146/annurev-pathol-020712-163923

Altschuler, S. J., and Wu, L. F. (2010). Cellular heterogeneity: Do differences make a difference? *Cell* 141, 559–563. doi: 10.1016/j.cell.2010.04.033

Bai, X., Fisher, D. E., and Flaherty, K. T. (2019). Cell-state dynamics and therapeutic resistance in melanoma from the perspective of MITF and IFNgamma pathways. *Nat. Rev. Clin. Oncol.* 16, 549–562. doi: 10.1038/s41571-019-0204-6

Baish, J. W., Stylianopoulos, T., Lanning, R. M., Kamoun, W. S., Fukumura, D., Munn, L. L., et al. (2011). Scaling rules for diffusive drug delivery in tumor and normal tissues. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1799–1803. doi: 10.1073/pnas.1018154108

Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004). Bacterial persistence as a phenotypic switch. *Science* 305, 1622–1625. doi: 10.1126/science.1099390

Bateson, W. (1894). *Material for the Study of Variation*. London: MacMillan and Co.

Batlle, E., and Clevers, H. (2017). Cancer stem cells revisited. *Nat. Med.* 23, 1124–1134.

Benzer, S. (1953). Induced synthesis of enzymes in bacteria analyzed at the cellular level. *Biochim. Biophys. Acta* 11, 383–395. doi: 10.1016/0006-3002(53)90057-2

Bergeron, L., Perez, G. I., Macdonald, G., Shi, L., Sun, Y., Jurisicova, A., et al. (1998). Defects in regulation of apoptosis in caspase-2-deficient mice. *Genes Dev.* 12, 1304–1314. doi: 10.1101/gad.12.9.1304

Bhattacharya, B., Mohd Omar, M. F., and Soong, R. (2016). The Warburg effect and drug resistance. *Br. J. Pharmacol.* 173, 970–979. doi: 10.1111/bph.13422

Bischoff, M., and Schnabel, R. (2006). A posterior centre establishes and maintains polarity of the Caenorhabditis elegans embryo by a Wnt-dependent relay mechanism. *PLoS Biol.* 4:e396. doi: 10.1371/journal.pbio.0040396

Blake, W. J., Balazsi, G., Kohanski, M. A., Isaacs, F. J., Murphy, K. F., and Kuang, Y. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* 24, 853–865. doi: 10.1016/j.molcel.2006.11.003

Boareto, M., Jolly, M. K., Goldman, A., Pietila, M., Mani, S. A., and Sengupta, S. (2016). Notch-Jagged signalling can give rise to clusters of cells exhibiting a hybrid epithelial/mesenchymal phenotype. *J. R. Soc. Interface* 13:20151106. doi: 10.1098/rsif.2015.1106

Bocci, F., Boareto, M., Ribeiro, M., Devi, G. R., Levine, H., and Onuchic, J. N. (2019). Toward understanding cancer stem cell heterogeneity in the tumor microenvironment. *Proc. Natl. Acad. Sci. U.S.A.* 116, 148–157. doi: 10.1073/pnas.1815345116

Bouchard, C., Staller, P., and Eilers, M. (1998). Control of cell proliferation by Myc. *Trends Cell Biol.* 8, 202–206. doi: 10.1016/s0962-8924(98)01251-3

Boyd, L., Guo, S., Levitan, D., Stinchcomb, D. T., and Kemphues, K. J. (1996). PAR-2 is asymmetrically distributed and promotes association of P granules and PAR-1 with the cortex in *C. elegans* embryos. *Development* 122, 3075–3084.

Brash, D., and Cairns, J. (2009a). The mysterious steps in carcinogenesis. *Br. J. Cancer* 101, 379–380. doi: 10.1038/sj.bjc.6605171

Brash, D., and Cairns, J. (2009b). The mysterious steps in carcinogenesis: addendum. *Br. J. Cancer* 101:1490. doi: 10.1038/sj.bjc.6605332

Brennan, C. W., Verhaak, R. G., McKenna, A., Campos, B., Noushmehr, H., Salama, S. R., et al. (2013). The somatic genomic landscape of glioblastoma. *Cell* 155, 462–477.

Brock, A., Chang, H., and Huang, S. (2009). Non-genetic heterogeneity–a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.* 10, 336–342. doi: 10.1038/nrg2556

Bronstein, D., and Akil, H. (1990). Effects of electrical stimulation in the periaqueductal gray on POMC peptides and mRNA in the rat brain. *Prog. Clin. Biol. Res.* 328, 219–222.

Cagatay, T., Turcotte, M., Elowitz, M. B., Garcia-Ojalvo, J., and Suel, G. M. (2009). Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell* 139, 512–522. doi: 10.1016/j.cell.2009.07.046

Cairns, R. A., Harris, I. S., and Mak, T. W. (2011). Regulation of cancer cell metabolism. *Nat. Rev. Cancer* 11, 85–95.

Carreau, S., and Hess, R. A. (2010). Oestrogens and spermatogenesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 1517–1535. doi: 10.1098/rstb.2009.0235

Casadesus, J., and Low, D. A. (2013). Programmed heterogeneity: epigenetic mechanisms in bacteria. *J. Biol. Chem.* 288, 13929–13935. doi: 10.1074/jbc.r113.472274

Chakravarty, D., Pedraza, A. M., Cotari, J., Liu, A. H., Punko, D., Kokroo, A., et al. (2017). EGFR and PDGFRA co-expression and heterodimerization in glioblastoma tumor sphere lines. *Sci. Rep.* 7:9043.

Chan, D. A., Sutphin, P. D., Yen, S. E., and Giaccia, A. J. (2005). Coordinate regulation of the oxygen-dependent degradation domains of hypoxia-inducible factor 1 alpha. *Mol. Cell. Biol.* 25, 6415–6426. doi: 10.1128/mcb.25.15.6415-6426.2005

Chauhan, V. P., Stylianopoulos, T., Martin, J. D., Popovic, Z., Chen, O., and Kamoun, W. S. (2012). Normalization of tumour blood vessels improves the

delivery of nanomedicines in a size-dependent manner. *Nat. Nanotechnol.* 7, 383–388. doi: 10.1038/nnano.2012.45

Chen, N., Fang, W., Lin, Z., Peng, P., Wang, J., Zhan, J., et al. (2017). KRAS mutation-induced upregulation of PD-L1 mediates immune escape in human lung adenocarcinoma. *Cancer Immunol. Immunother.* 66, 1175–1187. doi: 10.1007/s00262-017-2005-z

Cho, M., Grabmaier, K., Kitahori, Y., Hiasa, Y., Nakagawa, Y., Uemura, H., et al. (2000). Activation of the MN/CA9 gene is associated with hypomethylation in human renal cell carcinoma cell lines. *Mol. Carcinog.* 27, 184–189. doi: 10.1002/(sici)1098-2744(200003)27:3<184::aid-mc5>3.0.co;2-p

Ciccarese, F., and Ciminale, V. (2017). Escaping death: mitochondrial redox homeostasis in cancer cells. *Front. Oncol.* 7:117. doi: 10.3389/fonc.2017.00117

Darwin, C. (1859). *On the Origin of Species*. London: John Murray.

De Deken, R. H. (1966). The Crabtree effect: a regulatory system in yeast. *J. Gen. Microbiol.* 44, 149–156. doi: 10.1099/00221287-44-2-149

De Felici, M. (2013). "Origin, migration, and proliferation of human primordial germ cells," in *Oogenesis*, eds G. Coticchio, D. Albertini, and L. De Santis (London: Springer), 19–37. doi: 10.1007/978-0-85729-826-3_2

De Mey, M., De Maeseneire, S., Soetaert, W., and Vandamme, E. (2007). Minimizing acetate formation in *E. coli* fermentations. *J. Ind. Microbiol. Biotechnol.* 34, 689–700. doi: 10.1007/s10295-007-0244-2

DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G., and Thompson, C. B. (2008). The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab.* 7, 11–20. doi: 10.1016/j.cmet.2007.10.002

Dhar, N., McKinney, J., and Manina, G. (2017). "Phenotypic heterogeneity in *Mycobacterium tuberculosis*," in *Tuberculosis and the Tubercle Bacillus*, eds W. R. Jacobs Jr., H. McShane, V. Mizrahi, and I. M. Orme (ASM Press), 671–697. doi: 10.1128/9781555819569.ch32

Dhar, N., and McKinney, J. D. (2007). Microbial phenotypic heterogeneity and antibiotic tolerance. *Curr. Opin. Microbiol.* 10, 30–38. doi: 10.1016/j.mib.2006.12.007

Dhillon, A. S., Hagan, S., Rath, O., and Kolch, W. (2007). MAP kinase signalling pathways in cancer. *Oncogene* 26, 3279–3290. doi: 10.1038/sj.onc.1210421

Dominiak, A., Chelstowska, B., Olejarz, W., and Nowicka, G. (2020). Communication in the cancer microenvironment as a target for therapeutic interventions. *Cancers* 12:1232. doi: 10.3390/cancers12051232

Dudley, J. W., and Lambert, R. J. (2004). 100 generations of selection for oil and protein in corn. *Plant Breed. Rev.* 24, 79–110. doi: 10.1002/9780470650240.ch5

Dym, M. (1994). Spermatogonial stem cells of the testis. *Proc. Natl. Acad. Sci. U.S.A.* 91, 11287–11289.

Ellermeier, J. R., and Slauch, J. M. (2007). Adaptation to the host environment: regulation of the SPI1 type III secretion system in *Salmonella enterica* serovar Typhimurium. *Curr. Opin. Microbiol.* 10, 24–29. doi: 10.1016/j.mib.2006.12.002

Estrella, V., Chen, T., Lloyd, M., Wojtkowiak, J., Cornnell, H. H., and Bailey, K. (2013). Acidity generated by the tumor microenvironment drives local invasion. *Cancer Res.* 73, 1524–1535. doi: 10.1158/0008-5472.can-12-2796

Etemad-Moghadam, B., Guo, S., and Kemphues, K. J. (1995). Asymmetrically distributed PAR-3 protein contributes to cell polarity and spindle alignment in early *C. elegans* embryos. *Cell* 83, 743–752. doi: 10.1016/0092-8674(95)90187-6

Evans, T. C., Crittenden, S. L., Kodoyianni, V., and Kimble, J. (1994). Translational control of maternal glp-1 mRNA establishes an asymmetry in the C. elegans embryo. *Cell* 77, 183–194. doi: 10.1016/0092-8674(94)90311-5

Fabian, A., Vereb, G., and Szollosi, J. (2013). The hitchhikers guide to cancer stem cell theory: markers, pathways and therapy. *Cytometry A* 83, 62–71. doi: 10.1002/cyto.a.22206

Fallahi-Sichani, M., Becker, V., Izar, B., Baker, G. J., Lin, J. R., and Boswell, S. A. (2017). Adaptive resistance of melanoma cells to RAF inhibition via reversible induction of a slowly dividing de-differentiated state. *Mol. Syst. Biol.* 13:905. doi: 10.15252/msb.20166796

Feinberg, A. P., Ohlsson, R., and Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.* 7, 21–33. doi: 10.1038/nrg1748

Fidler, I. J. (1978). Tumor heterogeneity and the biology of cancer invasion and metastasis. *Cancer Res.* 38, 2651–2660.

Filippi, S., Barnes, C. P., Kirk, P. D., Kudo, T., Kunida, K., and McMahon, S. S. (2016). Robustness of MEK-ERK Dynamics and Origins of Cell-to-Cell Variability in MAPK Signaling. *Cell Rep.* 15, 2524–2535. doi: 10.1016/j.celrep.2016.05.024

Fischer, K. R., Durrans, A., Lee, S., Sheng, J., Li, F., and Wong, S. T. (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* 527, 472–476. doi: 10.1038/nature15748

Fisher, R. (1941). Average excess and average effect of a gene substitution. *Ann. Eugen.* 11, 53–63. doi: 10.1111/j.1469-1809.1941.tb02272.x

Fisher, R. A. (1918). The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. R. Soc. Edinb.* 52, 399–433. doi: 10.1017/s0080456800012163

Frank, S. A., and Rosner, M. R. (2012). Nonheritable cellular variability accelerates the evolutionary processes of cancer. *PLoS Biol.* 10:e1001296. doi: 10.1371/journal.pbio.1001296

Frezal, L., and Felix, M. A. C. (2015). *C. elegans* outside the Petri dish. *eLife* 4:e05849.

Friedewald, W. F., and Rous, P. (1944). The initiating and promoting elements in tumor production : an analysis of the effects of tar, benzpyrene, and methylcholanthrene on rabbit skin. *J. Exp. Med.* 80, 101–126. doi: 10.1084/jem.80.2.101

Furnari, F. B., Cloughesy, T. F., Cavenee, W. K., and Mischel, P. S. (2015). Heterogeneity of epidermal growth factor receptor signalling networks in glioblastoma. *Nat. Rev. Cancer* 15, 302–310. doi: 10.1038/nrc3918

Galhardo, R. S., Hastings, P. J., and Rosenberg, S. M. (2007). Mutation as a stress response and the regulation of evolvability. *Crit. Rev. Biochem. Mol. Biol.* 42, 399–435. doi: 10.1080/10409230701648502

Gefen, O., and Balaban, N. Q. (2009). The importance of being persistent: heterogeneity of bacterial populations under antibiotic stress. *FEMS Microbiol. Rev.* 33, 704–717. doi: 10.1111/j.1574-6976.2008.00156.x

Goldstein, B., and Hird, S. N. (1996). Specification of the anteroposterior axis in *Caenorhabditis elegans*. *Development* 122, 1467–1474.

Good, K., Ciosk, R., Nance, J., Neves, A., Hill, R. J., and Priess, J. R. (2004). The T-box transcription factors TBX-37 and TBX-38 link GLP-1/Notch signaling to mesoderm induction in *C. elegans* embryos. *Development* 131, 1967–1978. doi: 10.1242/dev.01088

Gotta, M., Abraham, M. C., and Ahringer, J. (2001). CDC-42 controls early cell polarity and spindle orientation in *C. elegans*. *Curr. Biol.* 11, 482–488. doi: 10.1016/s0960-9822(01)00142-7

Grive, K. J., and Freiman, R. N. (2015). The developmental origins of the mammalian ovarian reserve. *Development* 142, 2554–2563. doi: 10.1242/dev.125211

Grosse-Wilde, A., Kuestner, R. E., Skelton, S. M., MacIntosh, E., d'Herouel, A. F., Ertaylan, G., et al. (2018). Loss of inter-cellular cooperation by complete epithelial-mesenchymal transition supports favorable outcomes in basal breast cancer patients. *Oncotarget* 9, 20018–20033. doi: 10.18632/oncotarget.25034

Gschwind, A., Fischer, O. M., and Ullrich, A. (2004). The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nat. Rev. Cancer* 4, 361–370. doi: 10.1038/nrc1360

Guo, S., and Kemphues, K. J. (1995). par-1, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* 81, 611–620. doi: 10.1016/0092-8674(95)90082-9

Guruharsha, K. G., Kankel, M. W., and Artavanis-Tsakonas, S. (2012). The Notch signalling system: recent insights into the complexity of a conserved pathway. *Nat. Rev. Genet.* 13, 654–666. doi: 10.1038/nrg3272

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

Hegde, G. V., Chiu, C., Alag, N., Schaefer, G., Crocker, L., Ross, S., et al. (2013). Blocking NRG1 and other ligand-mediated Her4 signaling enhances the magnitude and duration of the chemotherapeutic response of non-small cell lung cancer. *Sci. Transl. Med.* 5:171ra18. doi: 10.1126/scitranslmed.3004438

Helmlinger, G., Sckell, A., Dellian, M., Forbes, N. S., and Jain, R. K. (2002). Acid production in glycolysis-impaired tumors provides new insights into tumor metabolism. *Clin. Cancer Res.* 8, 1284–1291.

Henriksson, M., and Luscher, B. (1996). Proteins of the Myc network: essential regulators of cell growth and differentiation. *Adv. Cancer Res.* 68, 109–182. doi: 10.1016/s0065-230x(08)60353-x

Heppner, G. H., and Miller, B. E. (1983). Tumor heterogeneity: biological implications and therapeutic consequences. *Cancer Metastasis Rev.* 2, 5–23. doi: 10.1007/bf00046903

Herman, M. A., Vassilieva, L. L., Horvitz, H. R., Shaw, J. E., and Herman, R. K. (1995). The *C. elegans* gene lin-44, which controls the polarity of certain asymmetric cell divisions, encodes a Wnt protein and acts cell nonautonomously. *Cell* 83, 101–110. doi: 10.1016/0092-8674(95)90238-4

Ho, B., Baryshnikova, A., and Brown, G. W. (2018). Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* proteome. *Cell Syst.* 6, 192–205.e3.

Hoek, K. S., Schlegel, N. C., Brafford, P., Sucker, A., Ugurel, S., and Kumar, R. (2006). Metastatic potential of melanomas defined by specific gene expression profiles with no BRAF signature. *Pigment. Cell Res.* 19, 290–302. doi: 10.1111/j.1600-0749.2006.00322.x

Holland, A. J., and Cleveland, D. W. (2009). Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat. Rev. Mol. Cell Biol.* 10, 478–487. doi: 10.1038/nrm2718

Hong, D., Fritz, A. J., Zaidi, S. K., Nickerson, J. A., Imbalzano, A. N., and Lian, J. B. (2018). Epithelial-to-mesenchymal transition and cancer stem cells contribute to breast cancer heterogeneity. *J. Cell. Physiol.* 233, 9136–9144. doi: 10.1002/jcp.26847

Hsu, P. P., and Sabatini, D. M. (2008). Cancer cell metabolism: warburg and beyond. *Cell* 134, 703–707. doi: 10.1016/j.cell.2008.08.021

Hugo, W., Zaretsky, J. M., Sun, L., Song, C., Moreno, B. H., Pang, J., et al. (2016). Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* 165, 35–44. doi: 10.1016/j.cell.2016.02.065

Huh, D., and Paulsson, J. (2011a). Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.* 43, 95–100. doi: 10.1038/ng.729

Huh, D., and Paulsson, J. (2011b). Random partitioning of molecules at cell division. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15004–15009. doi: 10.1073/pnas.1013171108

Iqbal, N., and Iqbal, N. (2014). Human epidermal growth factor receptor 2 (HER2) in cancers: overexpression and therapeutic implications. *Mol. Biol. Int.* 2014:852748.

Irani, K., Xia, Y., Zweier, J. L., Sollott, S. J., Der, C. J., Fearon, E. R., et al. (1997). Mitogenic signaling mediated by oxidants in Ras-transformed fibroblasts. *Science* 275, 1649–1652. doi: 10.1126/science.275.5306.1649

Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F. U., Kerner, M. J., et al. (2008). Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* 9:102. doi: 10.1186/1471-2164-9-102

Jia, D., Jolly, M. K., Boareto, M., Parsana, P., Mooney, S. M., Pienta, K. J., et al. (2015). OVOL guides the epithelial-hybrid-mesenchymal transition. *Oncotarget* 6, 15436–15448. doi: 10.18632/oncotarget.3623

Jolly, M. K., and Celia-Terrassa, T. (2019). Dynamics of phenotypic heterogeneity associated with EMT and stemness during cancer progression. *J. Clin. Med.* 8:1542. doi: 10.3390/jcm8101542

Jones, P. A., and Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* 3, 415–428. doi: 10.1038/nrg816

Kafri, M., Metzl-Raz, E., Jona, G., and Barkai, N. (2016). The cost of protein production. *Cell Rep.* 14, 22–31. doi: 10.1016/j.celrep.2015.12.015

Kaplan, J. (2008). The end of the adaptive landscape metaphor? *Biol. Philos.* 23, 625–638. doi: 10.1007/s10539-008-9116-z

Kaznatcheev, A., Vander Velde, R., Scott, J. G., and Basanta, D. (2017). Cancer treatment scheduling and dynamic heterogeneity in social dilemmas of tumour acidity and vasculature. *Br. J. Cancer* 116, 785–792. doi: 10.1038/bjc.2017.5

Keller, L., and Pantel, K. (2019). Unravelling tumour heterogeneity by single-cell profiling of circulating tumour cells. *Nat. Rev. Cancer* 19, 553–567. doi: 10.1038/s41568-019-0180-2

Kemper, K., de Goeje, P. L., Peeper, D. S., and van Amerongen, R. (2014). Phenotype switching: tumor cell plasticity as a resistance mechanism and target for therapy. *Cancer Res.* 74, 5937–5941. doi: 10.1158/0008-5472.can-14-1174

Kemphues, K. J., Priess, J. R., Morton, D. G., and Cheng, N. S. (1988). Identification of genes required for cytoplasmic localization in early *C. elegans* embryos. *Cell* 52, 311–320. doi: 10.1016/s0092-8674(88)80024-2

Kennedy, A. R., Cairns, J., and Little, J. B. (1984). Timing of the steps in transformation of C3H 10T 1/2 cells by X-irradiation. *Nature* 307, 85–86. doi: 10.1038/307085a0

Kim, J. W., Tchernyshyov, I., Semenza, G. L., and Dang, C. V. (2006). HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metab.* 3, 177–185. doi: 10.1016/j.cmet.2006.02.002

Knudson, A. G. Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U.S.A.* 68, 820–823. doi: 10.1073/pnas.68.4.820

Koch, U., and Radtke, F. (2007). Notch and cancer: a double-edged sword. *Cell. Mol. Life Sci.* 64, 2746–2762. doi: 10.1007/s00018-007-7164-1

Kohler, R. E. (1973). The background to Otto Warburg's conception of the Atmungsferment. *J. Hist. Biol.* 6, 171–192. doi: 10.1007/bf00127607

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., and Magnusson, G. (2012). Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* 488, 471–475.

Kudo-Saito, C., Shirako, H., Takeuchi, T., and Kawakami, Y. (2009). Cancer metastasis is accelerated through immunosuppression during Snail-induced EMT of cancer cells. *Cancer Cell* 15, 195–206. doi: 10.1016/j.ccr.2009.01.023

Kussell, E., and Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science* 309, 2075–2078. doi: 10.1126/science.1114383

Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Minden, M., et al. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* 367, 645–648. doi: 10.1038/367645a0

Larue, L., and Bellacosa, A. (2005). Epithelial-mesenchymal transition in development and cancer: role of phosphatidylinositol 3' kinase/AKT pathways. *Oncogene* 24, 7443–7454. doi: 10.1038/sj.onc.1209091

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218.

Lawson, D. A., Kessenbrock, K., Davis, R. T., Pervolarakis, N., and Werb, Z. (2018). Tumour heterogeneity and metastasis at single-cell resolution. *Nat. Cell Biol.* 20, 1349–1360. doi: 10.1038/s41556-018-0236-7

Lee, C. C., Shiao, H. Y., Wang, W. C., and Hsieh, H. P. (2014). Small-molecule EGFR tyrosine kinase inhibitors for the treatment of cancer. *Expert Opin. Investig. Drugs* 23, 1333–1348.

Lei, L., and Spradling, A. C. (2013). Mouse primordial germ cells produce cysts that partially fragment prior to meiosis. *Development* 140, 2075–2081. doi: 10.1242/dev.093864

Lemmon, M. A., Schlessinger, J., and Ferguson, K. M. (2014). The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harb. Perspect. Biol.* 6:a020768. doi: 10.1101/cshperspect.a020768

Levine, A. J., and Puzio-Kuter, A. M. (2010). The control of the metabolic switch in cancers by oncogenes and tumor suppressor genes. *Science* 330, 1340–1344. doi: 10.1126/science.1193494

Lim, J. S., Ibaseta, A., Fischer, M. M., Cancilla, B., Cristea, S., and Luca, V. C. (2017). Intratumoural heterogeneity generated by Notch signalling promotes small-cell lung cancer. *Nature* 545, 360–364. doi: 10.1038/nature22323

Liotta, L. A., and Kohn, E. C. (2001). The microenvironment of the tumour-host interface. *Nature* 411, 375–379. doi: 10.1038/35077241

Locasale, J. W., and Cantley, L. C. (2011). Metabolic flux and the regulation of mammalian cell growth. *Cell Metab.* 14, 443–451. doi: 10.1016/j.cmet.2011.07.014

Lunt, S. Y., and Vander Heiden, M. G. (2011). Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. *Annu. Rev. Cell Dev. Biol.* 27, 441–464. doi: 10.1146/annurev-cellbio-092910-154237

Maduro, M. F. (2010). Cell fate specification in the C. *elegans* embryo. *Dev. Dyn.* 239, 1315–1329.

Makino, S. (1956). Further evidence favoring the concept of the stem cell in ascites tumors of rats. *Ann. N. Y. Acad. Sci.* 63, 818–830. doi: 10.1111/j.1749-6632.1956.tb50894.x

Makiyan, Z. (2016). Systematization of ambiguous genitalia. *Organogenesis* 12, 169–182. doi: 10.1080/15476278.2016.1210749

Manchester, K. L. (1995). Theodor Boveri and the origin of malignant tumours. *Trends Cell Biol.* 5, 384–387. doi: 10.1016/s0962-8924(00)89080-7

Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 133, 704–715.

Marnik, E. A., and Updike, D. L. (2019). Membraneless organelles: P granules in *Caenorhabditis elegans*. *Traffic* 20, 373–379. doi: 10.1111/tra.12644

Martin, J. D., Fukumura, D., Duda, D. G., Boucher, Y., and Jain, R. K. (2016). Reengineering the tumor microenvironment to alleviate hypoxia and overcome cancer heterogeneity. *Cold Spring Harb. Perspect. Med.* 6:a027094. doi: 10.1101/cshperspect.a027094

Matova, N., and Cooley, L. (2001). Comparative aspects of animal oogenesis. *Dev. Biol.* 231, 291–320. doi: 10.1006/dbio.2000.0120

Matson, J. P., and Cook, J. G. (2017). Cell cycle proliferation decisions: the impact of single cell analyses. *FEBS J.* 284, 362–375. doi: 10.1111/febs.13898

McMahon, S. B. (2014). MYC and the control of apoptosis. *Cold Spring Harb. Perspect. Med.* 4:a014407. doi: 10.1101/cshperspect.a014407

Mickey, K. M., Mello, C. C., Montgomery, M. K., Fire, A., and Priess, J. R. (1996). An inductive interaction in 4-cell stage *C. elegans* embryos involves APX-1 expression in the signalling cell. *Development* 122, 1791–1798.

Mitrophanov, A. Y., and Groisman, E. A. (2008). Positive feedback in cellular control systems. *Bioessays* 30, 542–555. doi: 10.1002/bies.20769

Monod, J. (1974). *On Chance and Necessity*. London: Palgrave.

Morimoto, M., Nishinakamura, R., Saga, Y., and Kopan, R. (2012). Different assemblies of Notch receptors coordinate the distribution of the major bronchial Clara, ciliated and neuroendocrine cells. *Development* 139, 4365–4373. doi: 10.1242/dev.083840

Morita, Y., Maravei, D. V., Bergeron, L., Wang, S., Perez, G. I., and Tsutsumi, O. (2001). Caspase-2 deficiency prevents programmed germ cell death resulting from cytokine insufficiency but not meiotic defects caused by loss of ataxia telangiectasia-mutated (Atm) gene function. *Cell Death Differ.* 8, 614–620. doi: 10.1038/sj.cdd.4400845

Motta, P. M., Makabe, S., and Nottola, S. A. (1997). The ultrastructure of human reproduction. I. The natural history of the female germ cell: origin, migration and differentiation inside the developing ovary. *Hum. Reprod. Update* 3, 281–295. doi: 10.1093/humupd/3.3.281

Nance, J., and Zallen, J. A. (2011). Elaborating polarity: PAR proteins and the cytoskeleton. *Development* 138, 799–809. doi: 10.1242/dev.053538

Nishida, N., Yano, H., Nishida, T., Kamura, T., and Kojiro, M. (2006). Angiogenesis in cancer. *Vasc. Health Risk Manag.* 2, 213–219.

Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science* 194, 23–28. doi: 10.1126/science.959840

Ntziachristos, P., Lim, J. S., Sage, J., and Aifantis, I. (2014). From fly wings to targeted cancer therapies: a centennial for notch signaling. *Cancer Cell* 25, 318–334. doi: 10.1016/j.ccr.2014.02.018

Oduwole, O. O., Peltoketo, H., and Huhtaniemi, I. T. (2018). Role of follicle-stimulating hormone in spermatogenesis. *Front. Endocrinol.* 9:763. doi: 10.3389/fendo.2018.00763

Olmez, I., Shen, W., McDonald, H., and Ozpolat, B. (2015). Dedifferentiation of patient-derived glioblastoma multiforme cell lines results in a cancer stem cell-like state with mitogen-independent growth. *J. Cell. Mol. Med.* 19, 1262–1272. doi: 10.1111/jcmm.12479

Ozdemir, M., Kavak, R. P., Yalcinkaya, I., and Guresci, K. (2019). Ovotesticular disorder of sex development: an unusual presentation. *J. Clin. Imaging Sci.* 9:34. doi: 10.25259/jcis_45_2019

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., and Wakimoto, H. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401. doi: 10.1126/science.1254257

Patra, K. C., Wang, Q., Bhaskar, P. T., Miller, L., Wang, Z., and Wheaton, W. (2013). Hexokinase 2 is required for tumor initiation and maintenance and its systemic deletion is therapeutic in mouse models of cancer. *Cancer Cell* 24, 213–228. doi: 10.1016/j.ccr.2013.06.014

Pepling, M. E. (2012). Follicular assembly: mechanisms of action. *Reproduction* 143, 139–149. doi: 10.1530/rep-11-0299

Pereira, L., Mariadason, J. M., Hannan, R. D., and Dhillon, A. S. (2015). Implications of epithelial-mesenchymal plasticity for heterogeneity in colorectal cancer. *Front. Oncol.* 5:13. doi: 10.3389/fonc.2015.00013

Phan, L. M., Yeung, S. C., and Lee, M. H. (2014). Cancer metabolic reprogramming: importance, main features, and potentials for precise targeted anti-cancer therapies. *Cancer Biol. Med.* 11, 1–19.

Pietanza, M. C., Byers, L. A., Minna, J. D., and Rudin, C. M. (2015). Small cell lung cancer: will recent progress lead to improved outcomes? *Clin. Cancer Res.* 21, 2244–2255. doi: 10.1158/1078-0432.ccr-14-2958

Polyak, K., Haviv, I., and Campbell, I. G. (2009). Co-evolution of tumor cells and their microenvironment. *Trends Genet.* 25, 30–38. doi: 10.1016/j.tig.2008.10.012

Polyak, K., and Weinberg, R. A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat. Rev. Cancer* 9, 265–273. doi: 10.1038/nrc2620

Potter, M., Newport, E., and Morten, K. J. (2016). The Warburg effect: 80 years on. *Biochem. Soc. Trans.* 44, 1499–1505. doi: 10.1042/bst20160094

Prendergast, G. C. (1999). Mechanisms of apoptosis by c-Myc. *Oncogene* 18, 2967–2987. doi: 10.1038/sj.onc.1202727

Pribluda, A., de la Cruz, C. C., and Jackson, E. L. (2015). Intratumoral heterogeneity: from diversity comes resistance. *Clin. Cancer Res.* 21, 2916–2923. doi: 10.1158/1078-0432.ccr-14-1213

Provine, W. (1986). *Sewall Wright and Evolutionary Biology*. Chicago, IL: University of Chicago Press.

Quail, D. F., and Joyce, J. A. (2013). Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* 19, 1423–1437. doi: 10.1038/nm.3394

Queitsch, C., Sangster, T. A., and Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. *Nature* 417, 618–624. doi: 10.1038/nature749

Raj, A., Rifkin, S. A., Andersen, E., and van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature* 463, 913–918. doi: 10.1038/nature08781

Rajagopal, J., and Stanger, B. Z. (2016). Plasticity in the adult: How should the waddington diagram be applied to regenerating tissues? *Dev. Cell* 36, 133–137. doi: 10.1016/j.devcel.2015.12.021

Rao, C. V., Wolf, D. M., and Arkin, A. P. (2002). Control, exploitation and tolerance of intracellular noise. *Nature* 420, 231–237. doi: 10.1038/nature01258

Rocheleau, C. E., Downs, W. D., Lin, R., Wittmann, C., Bei, Y., Cha, Y. H., et al. (1997). Wnt signaling and an APC-related gene specify endoderm in early C. *elegans* embryos. *Cell* 90, 707–716. doi: 10.1016/s0092-8674(00)80531-0

Rose, L., and Gonczy, P. (2014). Polarity establishment, asymmetric division and segregation of fate determinants in early *C. elegans* embryos. *WormBook* 1–43. doi: 10.1895/wormbook.1.30.2

Rosen, J. M., and Jordan, C. T. (2009). The increasing complexity of the cancer stem cell paradigm. *Science* 324, 1670–1673. doi: 10.1126/science.1171837

Rosenthal, A. Z., Qi, Y., Hormoz, S., Park, J., Li, S. H., and Elowitz, M. B. (2018). Metabolic interactions between dynamic bacterial subpopulations. *eLife* 7:e33099.

Roussel, M. F., Ashmun, R. A., Sherr, C. J., Eisenman, R. N., and Ayer, D. E. (1996). Inhibition of cell proliferation by the Mad1 transcriptional repressor. *Mol. Cell. Biol.* 16, 2796–2801. doi: 10.1128/mcb.16.6.2796

Ruscetti, M., Dadashian, E. L., Guo, W., Quach, B., Mulholland, D. J., and Park, J. W. (2016). HDAC inhibition impedes epithelial-mesenchymal plasticity and suppresses metastatic, castration-resistant prostate cancer. *Oncogene* 35, 3781–3795. doi: 10.1038/onc.2015.444

Rutherford, S. L., and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature* 396, 336–342. doi: 10.1038/24550

Saini, S., Ellermeier, J. R., Slauch, J. M., and Rao, C. V. (2010). The role of coupled positive feedback in the expression of the SPI1 type three secretion system in *Salmonella*. *PLoS Pathog.* 6:e1001025. doi: 10.1371/journal.ppat.1001025

Sato, N., Maitra, A., Fukushima, N., Matsubayashi, H., Rosty, C., and Goggins, M. (2003). Frequent hypomethylation of multiple genes overexpressed in pancreatic ductal adenocarcinoma. *Cancer Res.* 63, 4158–4166.

Sauro, H. M. (2009). Network dynamics. *Methods Mol. Biol.* 541:269.

Sawyers, C. (2004). Targeted cancer therapy. *Nature* 432, 294–297.

Schrier, S. B., Hill, A. S., Plana, D., and Lauffenburger, D. A. (2016). Synergistic communication between CD4+ T cells and monocytes impacts the cytokine environment. *Sci. Rep.* 6:34942.

Schulze, A., and Harris, A. L. (2012). How cancer metabolism is tuned for proliferation and vulnerable to disruption. *Nature* 491, 364–373. doi: 10.1038/nature11706

Schwartz, D. M., Bonelli, M., Gadina, M., and O'Shea, J. J. (2016). Type I/II cytokines, JAKs, and new strategies for treating autoimmune diseases. *Nat. Rev. Rheumatol.* 12, 25–36. doi: 10.1038/nrrheum.2015.167

Segurel, L., Wyman, M. J., and Przeworski, M. (2014). Determinants of mutation rate variation in the human germline. *Annu. Rev. Genomics Hum. Genet.* 15, 47–70. doi: 10.1146/annurev-genom-031714-125740

Seydoux, G. (2018). The P Granules of C. *elegans*: a genetic model for the study of RNA-protein condensates. *J. Mol. Biol.* 430, 4702–4710. doi: 10.1016/j.jmb.2018.08.007

Shestov, A. A., Liu, X., Ser, Z., Cluntun, A. A., Hung, Y. P., and Huang, L. (2014). Quantitative determinants of aerobic glycolysis identify flux through the enzyme GAPDH as a limiting step. *eLife* 3:e03342.

Shin, T. H., Yasuda, J., Rocheleau, C. E., Lin, R., Soto, M., Bei, Y., et al. (1999). MOM-4, a MAP kinase kinase kinase-related protein, activates WRM-1/LIT-1 kinase to transduce anterior/posterior polarity signals in *C. elegans*. *Mol. Cell* 4, 275–280. doi: 10.1016/s1097-2765(00)80375-5

Shojaei, F., and Ferrara, N. (2008). Role of the microenvironment in tumor growth and in refractoriness/resistance to anti-angiogenic therapies. *Drug Resist. Updates* 11, 219–230. doi: 10.1016/j.drup.2008.09.001

Smits, W. K., Kuipers, O. P., and Veening, J. W. (2006). Phenotypic variation in bacteria: the role of feedback regulation. *Nat. Rev. Microbiol.* 4, 259–271. doi: 10.1038/nrmicro1381

Solopova, A., van Gestel, J., Weissing, F. J., Bachmann, H., Teusink, B., Kok, J., et al. (2014). Bet-hedging during bacterial diauxic shift. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7427–7432. doi: 10.1073/pnas.1320063111

Soto, A. M., and Sonnenschein, C. (2011). The tissue organization field theory of cancer: a testable replacement for the somatic mutation theory. *Bioessays* 33, 332–340. doi: 10.1002/bies.201100025

Stumpf, M. P., and Pybus, O. G. (2002). Genetic diversity and models of viral evolution for the hepatitis C virus. *FEMS Microbiol. Lett.* 214, 143–152. doi: 10.1111/j.1574-6968.2002.tb11338.x

Sturm, D., Bender, S., Jones, D. T., Lichter, P., Grill, J., and Becher, O. (2014). Paediatric and adult glioblastoma: multiform (epi)genomic culprits emerge. *Nat. Rev. Cancer* 14, 92–107. doi: 10.1038/nrc3655

Stylianopoulos, T., and Jain, R. K. (2013). Combining two strategies to improve perfusion and drug delivery in solid tumors. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18632–18637. doi: 10.1073/pnas.1318415110

Stylianou, N., Lehman, M. L., Wang, C., Fard, A. T., Rockstroh, A., and Fazli, L. (2019). A molecular portrait of epithelial-mesenchymal plasticity in prostate cancer associated with clinical outcome. *Oncogene* 38, 913–934. doi: 10.1038/s41388-018-0488-5

Sureka, K., Ghosh, B., Dasgupta, A., Basu, J., Kundu, M., and Bose, I. (2008). Positive feedback and noise activate the stringent response regulator rel in mycobacteria. *PLoS One* 3:e1771. doi: 10.1371/journal.pone.0001771

Swain, P. S., Elowitz, M. B., and Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12795–12800. doi: 10.1073/pnas.162041399

Szenk, M., Dill, K. A., and de Graff, A. M. R. (2017). Why do fast-growing bacteria enter overflow metabolism? Testing the membrane real estate hypothesis. *Cell Syst.* 5, 95–104. doi: 10.1016/j.cels.2017.06.005

Tabuse, Y., Izumi, Y., Piano, F., Kemphues, K. J., Miwa, J., and Ohno, S. (1998). Atypical protein kinase C cooperates with PAR-3 to establish embryonic polarity in *Caenorhabditis elegans*. *Development* 125, 3607–3614.

Tanaka, T., Hosoi, F., Nakamura, H., Masutani, H., Ueda, S., and Nishiyama, A. (2002). Thioredoxin-2 (TRX-2) is an essential gene regulating mitochondria-dependent apoptosis. *EMBO J.* 21, 1695–1703. doi: 10.1093/emboj/21.7.1695

Thorpe, C. J., Schlesinger, A., and Bowerman, B. (2000). Wnt signalling in Caenorhabditis elegans: regulating repressors and polarizing the cytoskeleton. *Trends Cell Biol.* 10, 10–17. doi: 10.1016/s0962-8924(99)01672-4

Thorpe, C. J., Schlesinger, A., Carter, J. C., and Bowerman, B. (1997). Wnt signaling polarizes an early *C. elegans* blastomere to distinguish endoderm from mesoderm. *Cell* 90, 695–705. doi: 10.1016/s0092-8674(00)80530-9

Thurley, K., Wu, L. F., and Altschuler, S. J. (2018). Modeling cell-to-cell communication networks using response-time distributions. *Cell Syst.* 6, 355–367.e5.

Tieche, C. C., Gao, Y., Buhrer, E. D., Hobi, N., Berezowska, S. A., and Wyler, K. (2019). Tumor initiation capacity and therapy resistance are differential features of EMT-related subpopulations in the NSCLC Cell Line A549. *Neoplasia* 21, 185–196. doi: 10.1016/j.neo.2018.09.008

Timsah, Z., Ahmed, Z., Ivan, C., Berrout, J., Gagea, M., Zhou, Y., et al. (2016). Grb2 depletion under non-stimulated conditions inhibits PTEN, promotes Akt-induced tumor formation and contributes to poor prognosis in ovarian cancer. *Oncogene* 35, 2186–2196. doi: 10.1038/onc.2015.279

Timsah, Z., Ahmed, Z., Lin, C. C., Melo, F. A., Stagg, L. J., Leonard, P. G., et al. (2014). Competition between Grb2 and Plcgamma1 for FGFR2 regulates basal phospholipase activity and invasion. *Nat. Struct. Mol. Biol.* 21, 180–188. doi: 10.1038/nsmb.2752

Tirosh, I., Izar, B., Prakadan, S. M., Treacy, D., Trombetta, J. J., Rotem, A., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196.

Tiwari, A., Balazsi, G., Gennaro, M. L., and Igoshin, O. A. (2010). The interplay of multiple feedback loops with post-translational kinetics results in bistability of mycobacterial stress response. *Phys. Biol.* 7:036005. doi: 10.1088/1478-3975/7/3/036005

Tripathi, S., Chakraborty, P., Levine, H., and Jolly, M. K. (2020). A mechanism for epithelial-mesenchymal heterogeneity in a population of cancer cells. *PLoS Comput. Biol.* 16:e1007619. doi: 10.1371/journal.pcbi.1007619

Tsimring, L. S. (2014). Noise in biology. *Rep. Prog. Phys.* 77:026601.

Vander Heiden, M. G., Cantley, L. C., and Thompson, C. B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 324, 1029–1033. doi: 10.1126/science.1160809

Varahan, S., Walvekar, A., Sinha, V., Krishna, S., and Laxman, S. (2019). Metabolic constraints drive self-organization of specialized cell groups. *eLife* 8:e46735.

Veening, J. W., Stewart, E. J., Berngruber, T. W., Taddei, F., Kuipers, O. P., and Hamoen, L. W. (2008). Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4393–4398. doi: 10.1073/pnas.0700463105

Ventela, S., Toppari, J., and Parvinen, M. (2003). Intercellular organelle traffic through cytoplasmic bridges in early spermatids of the rat: mechanisms of haploid gene product sharing. *Mol. Biol. Cell* 14, 2768–2780. doi: 10.1091/mbc.e02-10-0647

Von Stetina, S. E., and Mango, S. E. (2015). PAR-6, but not E-cadherin and beta-integrin, is necessary for epithelial polarization in *C. elegans*. *Dev. Biol.* 403, 5–14. doi: 10.1016/j.ydbio.2015.03.002

Waddington, C. H. (1956). *Principles of Embryology*. London: George Allen & Unwin, Ltd.

Wang, G. L., Jiang, B. H., Rue, E. A., and Semenza, G. L. (1995). Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proc. Natl. Acad. Sci. U.S.A.* 92, 5510–5514. doi: 10.1073/pnas.92.12.5510

Wang, M. H., Sun, R., Zhou, X. M., Zhang, M. Y., Lu, J. B., and Yang, Y. (2018). Epithelial cell adhesion molecule overexpression regulates epithelial-mesenchymal transition, stemness and metastasis of nasopharyngeal carcinoma cells via the PTEN/AKT/mTOR pathway. *Cell Death Dis.* 9:2.

Warburg, O. (1956). On the origin of cancer cells. *Science* 123, 309–314.

Ward, P. S., and Thompson, C. B. (2012). Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. *Cancer Cell* 21, 297–308. doi: 10.1016/j.ccr.2012.02.014

Wellen, K. E., and Thompson, C. B. (2010). Cellular metabolic stress: considering how cells respond to nutrient excess. *Mol. Cell* 40, 323–332. doi: 10.1016/j.molcel.2010.10.004

Wellen, K. E., and Thompson, C. B. (2012). A two-way street: reciprocal regulation of metabolism and signalling. *Nat. Rev. Mol. Cell Biol.* 13, 270–276. doi: 10.1038/nrm3305

Witchel, S. F. (2018). Disorders of sex development. *Best Pract. Res. Clin. Obstet. Gynaecol.* 48, 90–102.

Wolfe, A. J. (2005). The acetate switch. *Microbiol. Mol. Biol. Rev.* 69, 12–50. doi: 10.1128/mmbr.69.1.12-50.2005

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. VI Int. Cong. Genetr.* 1, 356–366.

Yang, J., Zhang, L., Erbel, P. J., Gardner, K. H., Ding, K., Garcia, J. A., et al. (2005). Functions of the Per/ARNT/Sim domains of the hypoxia-inducible factor. *J. Biol. Chem.* 280, 36047–36054. doi: 10.1074/jbc.m501755200

Yang, R., Niepel, M., Mitchison, T. K., and Sorger, P. K. (2010). Dissecting variability in responses to cancer chemotherapy through systems pharmacology. *Clin. Pharmacol. Ther.* 88, 34–38. doi: 10.1038/clpt.2010.96

Zhang, W., and Liu, H. T. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res.* 12, 9–18. doi: 10.1038/sj.cr.7290105

Zheng, X., Carstens, J. L., Kim, J., Scheible, M., Kaye, J., and Sugimoto, H. (2015). Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* 527, 525–530. doi: 10.1038/nature16064

Zipser, M. C., Eichhoff, O. M., Widmer, D. S., Schlegel, N. C., Schoenewolf, N. L., Stuart, D., et al. (2011). A proliferative melanoma cell phenotype is responsive to RAF/MEK inhibition independent of BRAF mutation status. *Pigment Cell Melanoma Res.* 24, 326–333. doi: 10.1111/j.1755-148x.2010.00823.x

Zomer, A., Ellenbroek, S. I., Ritsma, L., Beerling, E., Vrisekoop, N., and Van Rheenen, J. (2013). Intravital imaging of cancer stem cell plasticity in mammary tumors. *Stem Cells* 31, 602–606. doi: 10.1002/stem.1296

Zuazo-Gaztelu, I., and Casanovas, O. (2018). Unraveling the role of angiogenesis in cancer ecosystems. *Front. Oncol.* 8:248. doi: 10.3389/fonc.2018.00248

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.

frontiers in Cell and Developmental Biology

REVIEW
published: 14 December 2020
doi: 10.3389/fcell.2020.613116

Non-genetic Heterogeneity of Macrophages in Diseases—A Medical Perspective

Grégoire Gessain[1], Camille Blériot[2]* and Florent Ginhoux[3,4,5]**

[1] Université de Paris, Faculté de Santé, Paris, [2] Inserm U1015, Gustave Roussy, Villejuif, France, [3] Singapore Immunology Network(SIgN), Agency for Science, Technology and Research (A*STAR), Biopolis, Singapore, Singapore, [4] Shanghai Institute of Immunology, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [5] Translational Immunology Institute, SingHealth Duke-NUS Academic Medical Centre, Singapore, Singapore

Macrophages are sessile immune cells with a high functional plasticity. Initially considered as a uniform population of phagocytic scavengers, it is now widely accepted that these cells also assume developmental and metabolic functions specific of their tissue of residence. Hence, the paradigm is shifting while our comprehension of macrophage heterogeneity improves. Accordingly, exploiting this intrinsic versatility appears more and more promising for the establishment of innovative therapeutic strategies. Nevertheless, identifying relevant therapeutic targets remains a considerable challenge. Herein, we discuss various features of macrophage heterogeneity in five main categories of human diseases: infectious, inflammatory, metabolic, age-related, and neoplastic disorders. We summarize the current understanding of how macrophage heterogeneity may impact the pathogenesis of these diseases and propose a comprehensive overview with the aim to help in establishing future macrophage-targeted therapies.

Keywords: macrophage, monocyte, heterogeneity, human diseases, ontogeny, tissue-resident, inflammation

OPEN ACCESS

Edited by:
Ankur Sharma,
Genome Institute of Singapore, Singapore

Reviewed by:
Alexander Mildner,
Helmholtz Association of German Research Centers (HZ), Germany
Connie Jackaman,
Curtin University, Australia

****Correspondence:***
Grégoire Gessain
gregoire.gessain@gmail.com
Camille Blériot
camille.bleriot@gustaveroussy.fr
Florent Ginhoux
Florent_Ginhoux@immunol.a-star.edu.sg

Specialty section:
This article was submitted to Molecular Medicine, a section of the journal Frontiers in Cell and Developmental Biology

Received: *01 October 2020*
Accepted: *10 November 2020*
Published: *14 December 2020*

Citation:
Gessain G, Blériot C and Ginhoux F (2020) Non-genetic Heterogeneity of Macrophages in Diseases—A Medical Perspective. Front. Cell Dev. Biol. 8:613116. doi: 10.3389/fcell.2020.613116

INTRODUCTION

Macrophages are sessile within tissues at steady state and are therefore often named resident tissue macrophages (RTMs). They are considered as the guardians of tissue integrity due to their ability to phagocyte any "non-self" intruders and damaged or dying "self" cells. However, narrowing macrophages to their role of tissue scavengers appears too reductive as more and more immune and non-immune functions are documented (Okabe and Medzhitov, 2016). For example, a very recent study has reported that cardiac macrophages were involved in the elimination of dysfunctional mitochondria ejected from cardiomyocytes (Nicolás-Ávila et al., 2020), a mandatory task to the maintenance of heart homeostasis. It illustrates how macrophage-specific phagocytic abilities have been selected and shaped during evolution. Thus, RTMs should now be more considered as fully integrated and tissue-supportive components of any given tissue rather than only protective innate immune cells. Furthermore, even focusing on their immune functions, phagocytosis of foreign bodies appears only as a single string on their functional bow. Indeed, RTMs (and dendritic cells) have been anticipated as positive initiators of immunity (Janeway, 1989), assuming the original recognition of non-self-antigens. This process was proposed to lead to the generation of second signals strictly required for an efficient adaptive immune response, and these brilliant hypotheses

have been convincingly demonstrated since (Fearon and Locksley, 1996; Hoffmann et al., 1999). So, without RTMs, efficiency of the response as well as immune memory would be altered.

Twenty years ago, the biology of RTMs has been dichotomized into the so-called pro-inflammatory M1 and anti-inflammatory M2 states (Mills et al., 2000). Although outdated since, this was the first conceptual step toward the recognition of the complexity of macrophage biology (Bleriot et al., 2020). Nowadays, it has been clearly demonstrated that RTMs were not only bipolar but could actually harbor a full spectrum of activation states as an integrative response to any signals received (Xue et al., 2014; Ginhoux et al., 2016; Glass and Natoli, 2016). Therefore, RTMs display a plasticity that could be at least comparable or even more pronounced than the one well-recognized for other immune cells such as lymphocytes. It appears fundamental to uncover how this remarkable heterogeneity is generated and how it is modulated in several pathological conditions. Many recent studies described uncharacterized subpopulations of RTMs involved in several pathologies (Chakarov et al., 2019; Jaitin et al., 2019; Ramachandran et al., 2019; Xiong et al., 2019; Zilionis et al., 2019; Katzenelenbogen et al., 2020; Molgora et al., 2020). However, as the other immune and non-immune cells from a same organism, it is essential to remind that macrophages are genetically identical. Thus, to understand this diversity of phenotypes and functions, we must identify factors shaping macrophage identities and responses to stimulation. We have recently proposed to break down these parameters into four interconnected cardinal points, namely, origin, location, time of residence and tissue inflammatory status (Bleriot et al., 2020).

Among these parameters, it is necessary to recall that macrophages are the first immune cells to seed tissues during embryogenesis, as most of them do not derive from adult blood monocytes as it was commonly assumed for decades, but actually derive from embryonic precursors (Ginhoux et al., 2010; Schulz et al., 2012; Hashimoto et al., 2013; Yona et al., 2013; Gomez Perdiguero et al., 2015). Therefore, they are the immune cells forging the earliest ties with their tissue of residence. Indeed, it has been shown that even originating from common embryonic ancestors that could be designed as pre-macrophages, RTMs acquire very early tissue-specific programs depending of their local environment during fetal development (Mass et al., 2016). However, it has been also demonstrated that circulating monocytes seeding adult tissue to give rise to adult RTM also undergo a dramatic reprogramming reflecting the integration of tissue specificities (Bonnardel et al., 2019; Sakai et al., 2019). These diverse environmental cues that drive macrophage differentiation are unique to the niche of residence and involved a tissue-specific cocktail of different cytokines, metabolites, chemokines, and direct cell interactions. Although attempts have been made to describe these environmental programs (T'Jonck et al., 2018), they remain far from being fully described in an extensive manner. This being said, the central point becomes to decipher programs driving macrophage biology and how they evolve across time in healthy tissues or during disease development. In this review, we have split diseases into five main categories: infectious, inflammatory, metabolic, age-related, and neoplastic disorders (**Figure 1**). We discussed thereafter how macrophage biology is profoundly altered when homeostasis is disrupted, and how these changes can support pathogenesis.

MACROPHAGE HETEROGENEITY IN INFECTIOUS DISORDERS

Macrophages and Sepsis

Sepsis is the result of an aberrant host response to infection leading to organ dysfunctions (Singer et al., 2016). Across the globe, more than 30 million cases of sepsis are reported per year and it is the principal cause of death in intensive care units (ICU) (Fleischmann et al., 2016; Reinhart et al., 2017). Of note, it is also a very expensive condition for hospitals, as its annual costs exceed 20 billion dollars in the United States only (Singer et al., 2016). For decades, sepsis mortality was supposed to result solely from an hyperactive inflammatory response leading to harmful effects for the host such as fever, hypotension, tachypnea, tachycardia, coagulation disorders, and multiple-organ failure (Singer et al., 2016). For this reason, preventing excessive inflammation was the main objective in order to find a treatment for sepsis patients. However, a paradigm shift has started to emerge during the last decade because it appeared that the main immune dysfunction associated with high mortality was in reality not an excessive immune activation, but rather a strong immunosuppression, termed "sepsis-induced immuno-paralysis" (Leentjens et al., 2013; Wang et al., 2014). From a clinician perspective, this new paradigm was surprising at first, but rapidly supported by convincing evidences. Indeed, because of sepsis-induced immuno-paralysis, patients are unable to recover from their primary infection and are more likely to develop secondary infections from opportunistic pathogens later on (Otto et al., 2011). As a result, many patients with sepsis do not die from the initial pro-inflammatory hit, but later, from a secondary or opportunistic infection associated with an immunosuppressed state (Muenzer et al., 2010; Boomer et al., 2012; Walton et al., 2014). Therefore, researchers have progressively focused on ways to overcome the immuno-paralysis by developing immuno-stimulatory drugs (Leentjens et al., 2013; Delano and Ward, 2016; Patil et al., 2016).

Resident tissue macrophages play a fundamental role in such immuno-paralysis (Roquilly et al., 2020). For instance, some of the earliest evidence that macrophages may be involved in immuno-paralysis came from experiment of endotoxin-induced tolerance (Foster et al., 2007). In this study, authors showed that long-term endotoxin exposure of macrophages resulted in tolerogenic immune response. Such tolerance was not a result of DNA mutations but was induced by epigenetic modifications. Indeed, recognition of endotoxin by its specific receptor triggers chromatin modifications such as histone acetylation or methylation resulting in silencing of genes coding for pro-inflammatory molecules, but priming of genes coding for antimicrobial peptides (Foster et al., 2007). Of note, the initial dose of the priming agent, i.e., endotoxin in the seminal study discussed here, and pathogen inoculum in primary infection for sepsis patients, appear to

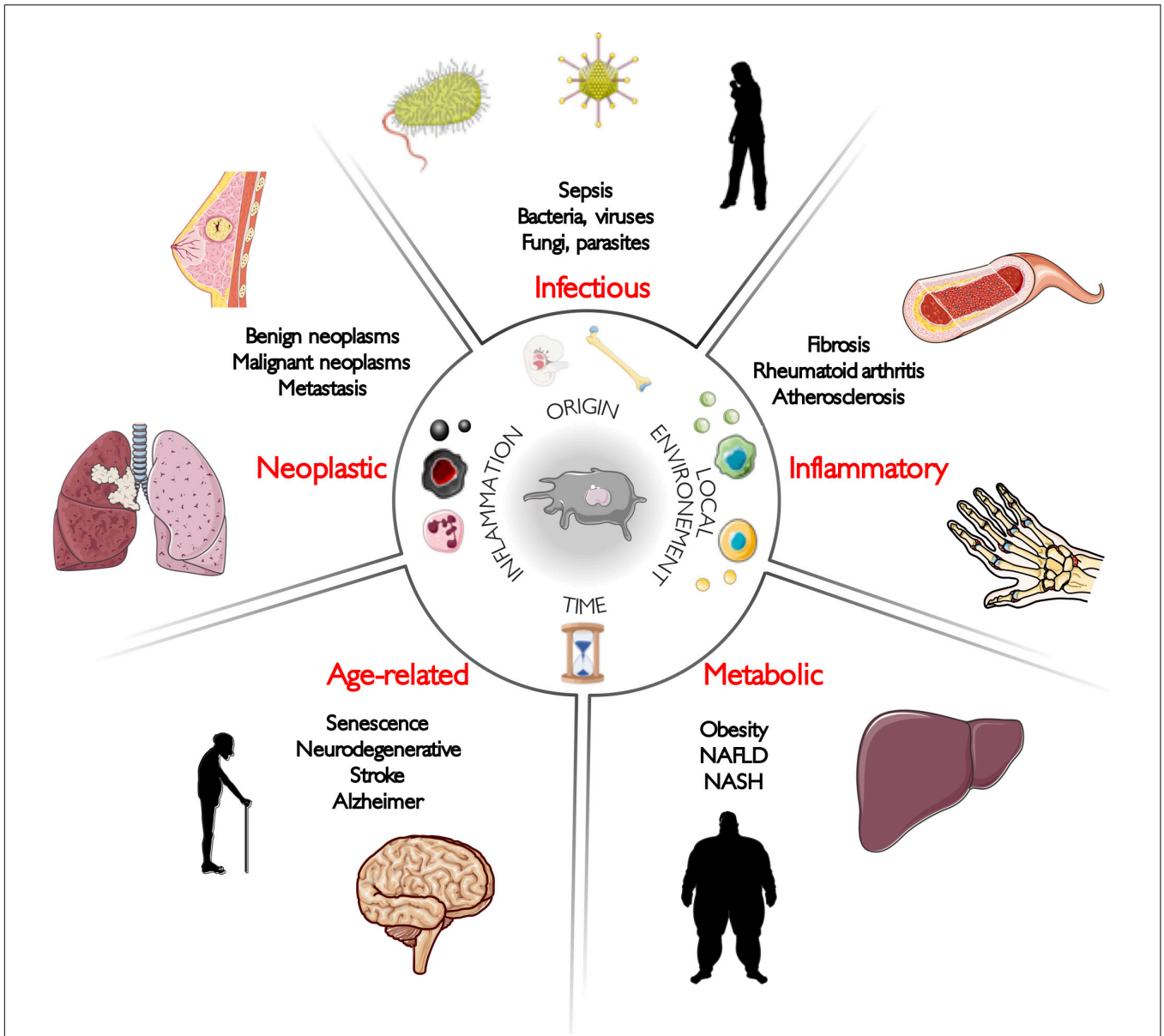

FIGURE 1 | Macrophages are at the crossroads of diseases. Macrophages are implicated in the pathogenesis of almost all human diseases. Here, our aim is to illustrate their non-genetic heterogeneity by describing recent macrophage-related discoveries in five main categories of human diseases: infectious, inflammatory, metabolic, age-related, and neoplastic disorders.

be critical: if a high dose of endotoxin triggered immunoparalysis of macrophages, a continuous exposure to very low doses actually results in enhanced responses to subsequent challenges, a process called trained immunity (Netea et al., 2016; Lajqi et al., 2019).

Similarly, in a double-infection murine model, Roquilly and colleagues showed that after recovering from a primary pneumonia, alveolar macrophages (AMs) displayed poor phagocytic capacity for several weeks (Roquilly et al., 2020). These weakened AMs originated from embryonic resident AMs that experienced a tolerogenic reprogramming of epigenetic nature. Of interest, signal-regulatory protein α (SIRPα), a membrane glycoprotein expressed mainly by myeloid cells, was a major player in the induction of this tolerogenic training. Finally, it was confirmed that AMs from patients with systemic inflammation still harbored reprogramming alterations up to six months after the resolution of inflammation. Very interestingly, *in vitro* inhibition of SIRPα enhanced phagocytosis in monocytes extracted from patients, suggesting that such treatment could modulate epigenetic reprogramming of monocytes and macrophages. Huge efforts are still needed to fully understand the phenomenon of immune tolerance and trained immunity as ones could hypothesize such effects being pathogen-specific. Further studies will integrate this recent concept for a more

efficient targeting of immuno-paralyzed cells and hopefully the design of more efficient clinical treatments.

Macrophages and Bacterial Infection

Tuberculosis (TB) is an airborne infection that principally affects the lungs, caused by the bacterial pathogen *Mycobacterium tuberculosis*. In 2010, it remained a serious global health threat responsible for 1.2 million deaths (Lozano et al., 2012). In the lungs, two macrophage subsets have been described, the alveolar (AMs) and the interstitial macrophages (IMs), AMs being fetal liver-derived whereas IMs being mostly adult monocyte-derived (Chakarov et al., 2019). A recent study has focused on the respective contribution of AMs and IMs in TB pathogenesis (Huang et al., 2018). The authors showed that AMs and IMs mounted divergent responses upon *Mycobacterium tuberculosis* infection: AMs promoted bacterial growth whereas IMs restricted it. Thus, specific targeting of the embryonic-derived AMs could be an interesting avenue for therapeutic research.

Another clinically relevant bacterium that has contributed to elucidate macrophage heterogeneity is *Listeria monocytogenes* (*Lm*). *Lm* is a foodborne pathogen that causes human listeriosis, a systemic infection with one of the highest mortality rates (Charlier et al., 2017). After ingestion of contaminated food, bacteria survive and multiply in the intestinal lumen and actively cross the intestinal barrier. Then, it disseminates within the host (Nikitas et al., 2011) and crosses both the blood–brain barrier and the placental barrier, leading, respectively, to meningitis and encephalitis, as well as abortion and neonatal infection in pregnant women (Lecuit, 2020). *Lm* is a powerful model pathogen that has led to significant discoveries, such as macrophage activation (Mackaness, 1969; Pamer, 2004). Furthermore, in a model of liver infection, *Lm* was able to induce necroptosis of Kupffer cells (KCs), the liver embryonic RTMs (Blériot et al., 2015). Necroptosis of KCs was responsible for (i) the recruitment of bone-marrow pro-inflammatory monocytes to clear the infection and repopulate the empty KC niche and (ii) their conversion after bacterial clearance into liver RTMs that harbor anti-inflammatory phenotype and tissue repair functions. Such a model challenges the M1–M2 paradigm by showing that a single macrophage is actually highly plastic and can be reprogrammed from a M1-like to M2-like phenotype across the course of an infection according to changes in its microenvironment.

Macrophages and Viral Infection

The recently identified severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2) is an enveloped RNA betacoronavirus responsible for the still active coronavirus disease 2019 outbreak (Covid-19). Patients with severe Covid-19 have various clinical symptoms such as pneumonia with fever, cough, and dyspnea (Guan et al., 2020). As in many benign viral infections, the majority of patients presenting with mild disease mounts an efficient immune response (Thevarajan et al., 2020). However, severe viral pneumonia can lead to acute respiratory distress syndrome (ARDS), a complication that occurs in 15.6% of patients with severe Covid-19 (Guan et al., 2020). A disproportionate inflammatory response to the virus contributes to the severity of the symptoms and can result to death (Mehta et al., 2020). Bronchoalveolar fluid (BALF) analysis of patients suffering from severe Covid-19 showed an elevated proportion of mononuclear phagocytes accounting for 80% of total cells, as compared to 40% in healthy controls (Liao et al., 2020). Among them, a near-complete depletion of AMs and an enrichment of inflammatory monocyte-derived macrophages was observed. Such a dramatic decrease of AMs has been confirmed recently by analysis of BALF samples from patients with mild or severe Covid-19 (Silvin et al., 2020). Here again, the contribution to lung damage of embryonic-resident macrophages as opposed to monocyte-derived macrophages remains to be investigated. Interestingly, patients with severe Covid-19 had an accumulation of immuno-suppressive HLA-DR^{low} classical monocytes, a feature commonly found in other severe illnesses (Lukaszewicz et al., 2009), as well as a Covid-19-specific decrease of non-classical $CD14^{low}CD16^{high}$ monocytes. This decrease appears to be a very characteristic biological signature of patients with severe Covid-19, easily measurable with standard diagnostic flow cytometry available in hospitals. Therefore, such monocyte heterogeneity constitutes a valuable predictive biomarker in blood samples that could help to the early detection of severe Covid-19 patients (Schulte-Schrepping et al., 2020; Silvin et al., 2020).

In addition, macrophage diversity could also be involved in the relatively efficient response of children, as compared to adults, to Covid-19 (Castagnoli et al., 2020). Indeed, because the children's immune system is relatively immature, they follow an intense vaccination program that involves several adjuvants. As recently discussed by Mantovani and Netea, adjuvants have been shown to induce trained immunity and thus boost antimicrobial function in myeloid cells (Castagnoli et al., 2020; Mantovani and Netea, 2020). Adjuvants are known to elicit an innate immune response in myeloid cells, which is also efficient against heterologous pathogens. As a consequence, innate immune cells have an increased non-specific response that goes well beyond the antigen in the vaccine (Netea et al., 2020a). Although still theoretical, this vaccination-induced reprogramming of myeloid cells could be one among many reasons why children are relatively resistant to Covid-19. Although others' hypothesis are currently under study (Steinman et al., 2020), the reprogramming of myeloid cells by adjuvants and live vaccines deserved to be analyzed in depth. As a consequence, vaccination of elderly people against *influenzae virus* and *Mycobacterium tuberculosis*, using, respectively, anti-*influenzae virus* vaccines with adjuvants and the live vaccine BCG, may enhance their response to Covid-19 (Mantovani and Netea, 2020; Netea et al., 2020b). Therefore, deciphering macrophage heterogeneity induced by different vaccination protocols might help to solve differences in susceptibility to Covid-19 infection.

Macrophages and Fungal Infections

Cryptococcus neoformans (*Cn*) is an opportunistic fungus that infects immuno-suppressed patients. Worldwide, it is still responsible today for more than 1 million life-threatening infections per year (Brown et al., 2012). A recent study identified the presence of two subsets of AMs at an early stage

of *Cn* pulmonary infection (Xu-Vanpala et al., 2020). One AM population was $CXCL2^{+}$ and had a pro-inflammatory phenotype whereas the second $CXCL2^{-}$ had an anti-inflammatory phenotype, including expression of IL-10. This anti-inflammatory phenotype was regulated at the epigenetic level. Interestingly, these heterogeneous AM responses were not attributable to fungal inoculum, spatial lung localization, nor ontogeny. This heterogeneity was also confirmed in a model of *Aspergillus fumigatus* infection. Important questions remained, as authors did not explain whether intrinsic or extrinsic cues explained the simultaneous appearance of heterogeneous AM subsets with pro- and anti-inflammatory profiles.

MACROPHAGE HETEROGENEITY IN INFLAMMATORY DISORDERS

Macrophages in Tissue Repair and Fibrosis

The T helper 1 (T_H1)–T_H2 paradigm was defined three decades ago (Mosmann et al., 1986), and since then, type 2 immunity was mostly described as a counter-regulatory mechanism dampening type 1 immunity (Wynn and Vannella, 2016). In 2000, Mills and colleagues proposed an elegant parallel between T helper cells and macrophages (Mills et al., 2000). Of note, these two opposite macrophage polarizations were independent of T lymphocytes and were able to influence by themselves opposite immunologic outcomes. Thus, these two distinct populations were termed M1 pro-inflammatory and M2 anti-inflammatory macrophages (Mills et al., 2000). Nowadays, the diverse contributions of type 2 cytokines are more understood. Indeed, in addition to suppressing type 1 response, type 2 immunity and its associated M2 macrophages are well-known contributors of repair and regeneration of injured tissues (Gieseck et al., 2018). Therefore, type 2 response is at the crossroads of two major tasks and its dysregulation can lead to numerous pathological conditions. On the one hand, weak type 2 responses often lead to autoimmune diseases through overstimulation of type 1 responses. On the other hand, chronically activated type 2 responses overactivate wound healing and processes leading to the development of pathological fibrosis (Wynn and Ramalingam, 2012). For many chronic inflammatory diseases, fibrosis is a shared final outcome that can lead to organ failure and death. Fibrotic disorders have been estimated to contribute to 45% of all deaths in the developed world (Wynn, 2004). Numerous fibrotic diseases have been linked to type 2 immunity activation. For instance, chronic helminth infections such as schistosomiasis are associated with fibroproliferative lesions whose mechanism involves type 2 cytokines such as IL-4, IL-5, and IL-13 (Pearce and MacDonald, 2002; Fairfax et al., 2012). As another example, persistence of a chronic injury in the lungs disturbs the wound healing pathways and often leads to fibrosis. Lung fibrosis is of significant medical interest and has been observed in various pulmonary diseases ranging from both acute disorders such as pneumonia, bronchiolitis, ARDS, emphysema, and chronic ones such as idiopathic pulmonary fibrosis (IPF), chronic obstructive pulmonary disease (COPD), asbestosis, asthma, cystic fibrosis, sarcoidosis, and so forth (Gieseck et al., 2018). These diseases affect millions of people globally, and because very few effective treatment options exist, they are one of the leading causes of chronic morbidity and mortality. Type 2 cytokines, as IL-4 and IL-13, are elevated in many of these lung diseases (Grünig et al., 1999; Jakubzick et al., 2003; Keane et al., 2007; Baurakiades et al., 2014; Heitmann et al., 2014; Wills-karp et al., 2016).

Because macrophages and monocytes contribute to the reparation process of injured tissues, from initiation to resolution, several studies suggest a critical role for RTMs in fibrosis pathogenesis. In a model of lung fibrosis, Misharin and colleagues showed that specific depletion of monocyte-derived AMs after their recruitment to the lung ameliorates fibrosis, whereas embryonic AM depletion did not alter fibrosis severity (Misharin et al., 2017). They found that monocyte differentiation to AM occurs progressively during fibrosis and its resolution. Indeed, early monocyte-derived AMs were expressing pro-fibrotic genes, which were then progressively downregulated along their differentiation into mature AMs. Embryonic- and monocyte-derived AMs showed differences in pro-fibrotic gene expression during fibrosis, but ten months later, no more differences were observed. These results revealed remarkable heterogeneity in AM functions according to their origin with important repercussions for the design of innovative myeloid-targeted therapy against fibrosis (Misharin et al., 2017). Our recent work investigated the fibrosis contribution of IMs. First, we demonstrated the existence of two subsets of IMs ($LYVE1^{hi}MHCII^{low}$ and $LYVE1^{low}MHCII^{hi}$) in distinct sub-tissular pulmonary niches and other tissues. Most importantly, we showed that absence of the $LYVE1^{hi}MHCII^{low}$ IMs subpopulation exacerbated immune cell infiltration, tissue inflammation, collagen deposition, and finally fibrotic processes in an experimental model of lung and heart fibrosis. Although these $LYVE1^{hi}MHCII^{low}$ IMs are monocyte-derived, they express high levels of genes linked with wound healing and repair and can restrain experimental fibrosis (Chakarov et al., 2019). These two studies highlighted two levels of RTM heterogeneity within the same organ that are important in fibrosis pathogenesis. First, among AMs, ontogeny seems to have an influence as embryonic- and monocyte-derived AMs display different functions. Second, among IMs, the sub-tissular niches seem to have also a significant impact as two IM subpopulations, located in different regions, display different functions. It would be of interest to take into account such heterogeneity for developing innovative therapies in the future.

Macrophages in Autoimmune Diseases

Rheumatoid arthritis (RA) is a chronic autoimmune disease, in which the small joints are exposed to an inflammatory polyarthritis. RA is a "multicausal" disease that most likely results from a combination of genetic predisposition and various environmental and lifestyle factors. Articular and systemic manifestations in RA can lead to long-term outcomes such as permanent disability. RA is estimated to affect approximately 0.24 to 1 percent of the population (Cross et al., 2014; Hunter et al., 2017). It is defined by a breakdown of tolerance to modified

self-protein and chronic synovitis (Lee and Weinblatt, 2001). Current therapies benefit only to a small proportion of patients as sustained clinical remission is only achieved in 20 to 40% of them (Nagy and van Vollenhoven, 2015). However, a minor fraction of patients has long-term drug-free remission for which underlying mechanisms remain ill defined.

Interestingly, the most abundant synovial immune cells of patients in remission are synovial tissue macrophages (STMs). By comparing STMs obtained from patients with active RA, drug-free remission patients, and healthy donors, Stefano and colleagues have identified two STM populations: MerTK^{neg} STMs were enriched in patients with active RA whereas MerTK^{pos} STMs were predominant in drug-free remission patients and healthy donors. MerTK^{neg} STMs had a pro-inflammatory profile, and MerTK^{pos} STMs were negative regulators of inflammation. Interestingly, they identified two subpopulations among MerTK^{neg} STMs: TREM2^{pos} and FOLR2^{pos}LYVE1^{pos} STMs. These two subsets reside in different locations and have different and complementary immuno-regulatory roles. Here again, the sub-tissular niches seem to have an impact on RTM heterogeneity. Finally, the authors suggest that therapeutic enhancement of the functions of MerTK^{pos} STMs could facilitate restoration of synovial homeostasis (Alivernini et al., 2020). Another recent study highlighted macrophage heterogeneity in the joint of patients suffering from RA (Culemann et al., 2019). At homeostasis, they identified a subset of TREM2^{pos}CX3CR1^{pos} STM that were forming a tight-junction-mediated protective barrier at the synovial lining and physically seclude the joint. These STMs displayed features of epithelial cells and locally proliferated from embryonic-derived interstitial CX3CR1^{neg} macrophages localized within deeper layers of synovial tissue. During RA, this barrier rapidly disintegrated, thus facilitating monocyte infiltration (Culemann et al., 2019). Here again, this study showed very elegantly that ontogeny impacts macrophage function and that RTMs have different roles according to their sub-tissular niche.

Macrophages in Atherosclerosis

Atherosclerosis is a lipid-driven inflammatory disease where the wall of large arteries is slowly filled by atherosclerotic plaques. Even though it remains an asymptomatic disease in the first half of human life, it remains a major contributor to most cardiovascular diseases affecting elderly individuals. Worldwide, stroke and myocardial infarction are a major social and economic burden and one of the leading causes of death (Lozano et al., 2012). Currently, statins, β-blockers and ACE inhibitors are widely used to control hyperlipidemia and hypertension, respectively. However, efficiency of these treatments is limited.

Macrophages are the most abundant immune cells in the plaque (Cochain et al., 2018; Cole et al., 2018) and have been implicated in all stages of the disease (Hansson and Hermansson, 2011). Recent technical advances in immunology such as cytometry by time of flight (CyTOF) and single-cell RNA sequencing (scRNA-Seq) have enabled a comprehensive mapping of the different macrophages in atherosclerotic plaques (Willemsen and de Winther, 2020). Three main populations of macrophages have been defined: (i) resident-like, (ii) pro-inflammatory, and (iii) foamy TREM2^{high} macrophages (Fernandez et al., 2019). The resident-like macrophages are the only subset in healthy mice but are also present in the adventitia of atherosclerotic aorta. They have an embryonic origin, are self-renewing, express highly FOLR2 and LYVE1, and harbor an anti-inflammatory phenotype (Ensan et al., 2016; Cochain et al., 2018; Kim et al., 2018; Winkels et al., 2018). The pro-inflammatory macrophages are monocyte-derived and are exclusively found in the intima of atherosclerotic aorta where they constitute the largest macrophage subset (Cochain et al., 2018; Kim et al., 2018; Winkels et al., 2018), promoting atherosclerosis lesions. Finally, foamy TREM2^{high} macrophages are monocyte-derived lipid-laden foam cells found exclusively in the intima where they take up atherogenic lipoprotein, resulting in the formation of a lipid-rich core that progresses toward necrotic lesions; they are involved in metabolic regulations and seem to have an immunosuppressive phenotype (Cochain et al., 2018). Here again, ontogeny and sub-tissular niches seems to have an impact on the macrophage polarization and thus on the pathogenesis of the disease.

Another noteworthy fact about atherosclerosis is that it has been epidemiologically associated with infections (Thompson et al., 2013). Indeed, according to numerous human epidemiological studies and animal models, the infectious burden might be linked to later atherosclerotic cardiovascular diseases (ASCVD) and acute infections could cause cardiovascular events (Corrales-Medina et al., 2013; Pothineni et al., 2017). One possible hypothesis to explain such observation is that macrophages and monocytes are trained by the successive infectious challenges throughout life. Such trained immunity provides significant protection against reinfection and improves mortality, even in the absence of an effective adaptive immunity (Kleinnijenhuis et al., 2012; Quintin et al., 2012). However, it also contributes to atherosclerosis progression and to acute disruption of existing atherosclerotic plaques (Christ et al., 2016). Trained monocytes and macrophages display a profound proatherogenic phenotype that is mediated by two intracellular mechanisms which are of metabolic and epigenetic nature. This innate immune memory relies both on central and peripheric modifications, resulting in long-term activation of innate immune cells. On the one hand, several studies showed that bone marrow progenitors such as hematopoietic stem cells (HSC) are indeed subjected to an epigenetic reprogramming upon intravenous BCG vaccination (Kaufmann et al., 2018), intraperitoneal administration of β-glucan, a well-known inducer of trained-immunity (Mitroulis et al., 2018), or intraperitoneal administration of endotoxin (de Laval et al., 2020). On the other hand, Yao and colleagues showed that an innate immune memory induced by adenovirus infection was independent of the contribution of monocytes and bone-marrow progenitors, by taking place directly in resident tissue macrophages (Yao et al., 2018). Across all studies, trained myeloid cells were found to be long-lasting as they were still conferring protection up to months after the initial challenge (Machiels et al., 2017; Kaufmann et al., 2018; Yao et al., 2018; de Laval et al., 2020). Of note, in a very elegant study, Réu and colleagues showed that embryonic

RTMs are very long-lasting cells. Indeed, they observed that human microglia were on average 4.2 years old and some of them were found to be more than two decades old (Réu et al., 2017). Such results are of interest because they imply that one single RTM could be challenged several times by different stimuli throughout life, resulting in strong peripheral trained immunity. These results highlight the potential of a better understanding of trained immunity in long-lasting RTM.

To conclude, even if more studies need to be conducted, the epigenetic heterogeneity of macrophages and monocytes might play a significant role in atherosclerosis (Leentjens et al., 2018). Of note, trained immunity can be prevented by pharmacological inhibitors of metabolic pathways, such as glutaminolysis and fatty acid synthesis, and histone methyltransferase blockers (Arts et al., 2016). These could represent innovative strategies to reduce ASCVD risk in patients with acute infections, such as pneumonia. It could also reduce the potential deleterious effects of repeated childhood infections on later ASCVD risk (Leentjens et al., 2018).

MACROPHAGE HETEROGENEITY IN METABOLIC DISORDERS

Macrophages in Obesity-Related Insulin Resistance

Nowadays, in high-income countries, a pandemic of obesity threatens the health population by predisposing them to diabetes, non-alcoholic fatty liver disease (NAFLD) and cardiovascular diseases (Heymsfield and Wadden, 2017). For the first time ever, life expectancy is projected to a potential decline (Olshansky et al., 2005), one of the main reasons being the obesity pandemic and all its related deleterious effects (Ludwig, 2016). Overnutrition induces a positive energy balance that leads to fat accumulation in adipose tissue, which triggers immune responses aimed to restore homeostasis.

Adipose tissue macrophages (ATMs) are the largest immune population in adipose tissue and accumulate even more in obesity where they promote a chronic low-grade inflammation (Weisberg et al., 2003; Xu et al., 2003). The long-term consequences of a persistent inflammation are insulin resistance and loss of metabolic flexibility (Reilly and Saltiel, 2017). In obesity, the M1/M2 paradigm presents several limitations and cannot adequately describe ATM functions. Indeed, obesity converts ATMs into a metabolically activated (MMe) macrophage state that is mechanistically distinct from M1-like or M2-like phenotype (Kratz et al., 2014). Coats and colleagues showed that MMe were associated with both production of inflammatory cytokine (a harmful function) and clearance of dead adipocytes by lysosomal exocytosis in crown-like-structure around dying cells (a beneficial function) (Coats et al., 2017). Traditionally, these opposite functions were attributed to distinct ATM subpopulations: the detrimental one being associated with M1-like ATMs while the beneficial ones being ascribed to M2-like ATMs. However, Coats and colleagues provide evidence that these two functions were the properties of a single MMe macrophage subset that evolves upon diet-induced obesity. These findings were confirmed in a following study using a single-cell sequencing approach. The authors were able to identify a population of $CD9^+$ ATMs that localized to crown-like structure, were enriched in lipids, and upregulated both inflammatory pathways and lysosomal metabolism (Hill et al., 2018). Finally, a recent study identified one population of $CD9^+TREM2^+$ lipid-associated macrophages (LAMs) that arise in obesity conditions from recruited monocytes and formed crown-like structure. These LAMs were able to prevent adipocyte hypertrophy, hypercholesterolemia, inflammation, body-fat accumulation, and glucose intolerance (Jaitin et al., 2019). Together, these results highlight the limitations of the M1/M2 paradigm and showed that one same long-lived macrophage can harbor opposite functions across time, according to the duration of the challenge.

Macrophages in Non-alcoholic Fatty Liver Disease (NALFD) and Its Inflammatory Form, Non-alcoholic Steatohepatitis (NASH)

NAFLD is defined by an excessive fat accumulation in the liver and is associated with obesity and metabolic syndrome. In the United States, 25–30% of the population develops NAFLD, which then may progresses into a more serious form of NAFLD, termed non-alcoholic steatohepatitis (NASH), characterized by chronic liver injury, fibrosis, and inflammation (Cohen et al., 2011; Diehl and Day, 2017; Samuel and Shulman, 2018). NASH can subsequently cause end-stage liver pathologies, as cirrhosis and hepatocellular carcinoma (HCC). Finally, it is a common indication for liver transplantation (Pais et al., 2016; Diehl and Day, 2017).

A recent study discovered specific NASH-associated macrophages (NAMs) that were present both in mice and in humans (Xiong et al., 2019). These NAMs represent about 60% of macrophages from NASH livers; they expressed high amount of TREM2 and CD9 and appear to have a protecting role during NASH pathogenesis. Another study identified in cirrhotic patients a subpopulation of scar-associated macrophages (SAMacs) that differentiates from circulating monocytes, expresses a high amount of TREM2 and CD9, and had a pro-fibrotic phenotype. These SAMacs were also expanded in a cohort of patients suffering from NASH (Ramachandran et al., 2019). Thus, even if macrophages are heterogeneous from one organ to another, it appears that they also share similar properties as illustrated by TREM2 signaling in adipose tissue and the liver. Interestingly, two very recent studies confirmed these findings in mice (Remmerie et al., 2020; Seidman et al., 2020). Indeed, in a model of metabolic-associated fatty liver disease (MAFLD), Remmerie et al. showed that KCs were progressively eliminated along the course of the disease and slowly replaced by monocyte-derived cells. A subset of them was termed "hepatic LAMs" as they had a transcriptome similar to adipose tissue LAMs and fibrotic liver SAMacs (Remmerie et al., 2020). In addition, Seidmann et al. showed that within the diseased liver, different microenvironments are responsible for

distinct differentiations among resident and infiltrating immune cells by remodeling the chromatin status of recruited monocytes but also by modifying the activities of preexisting enhancers of the resident KC population (Seidman et al., 2020).

MACROPHAGE HETEROGENEITY IN AGE-RELATED DISORDERS

Macrophage, Immuno-Senescence, and Inflammaging

Thanks to modern medicine and public health measures, human life expectancy is far better today than it was a century ago. However, a prolonged lifespan goes along with a rise in non-communicable diseases such as cancer and cardiovascular, autoimmune, and neurodegenerative diseases and a higher susceptibility to infections. As a consequence, the aging research community has seen the emergence of geroscience, a research field that aims to extend human longevity (Kennedy et al., 2014). The immune system is impacted by virtually all hallmarks of aging (López-Otín et al., 2013) by undergoing with age a profound remodeling termed immuno-senescence that impacts both arms of our immunity (Grubeck-Loebenstein et al., 2009).

Immuno-senescence is generally associated with a loss of immune functions and defective immune system. All immune cells are affected, ranging from a dysfunction of adaptive T and B cells to functional changes of innate immune cells subsets, such as monocytes and macrophages (Grubeck-Loebenstein et al., 2009). For instance, although the absolute number of monocytes is constant upon aging, the ratio of monocyte subsets is altered: classical monocytes ($CD14^+CD16^-$) are reduced, while intermediate ($CD14^+CD16^+$) and non-classical monocytes ($CD14^{low}CD16^+$) are increased (Seidler et al., 2010; Hearps et al., 2012). Of note, non-classical monocytes have a lower expression of HLA-DR suggesting a decline of antigen presentation function. Likewise, aged macrophages exhibit a lower level of MHC-II expression (Herrero et al., 2002), have a disabled clearance of dead cell capacity (Aprahamian et al., 2008) and a reduced chemotaxis (Solana et al., 2012). However, some innate immunity features seem to be conserved or even increased during immuno-senescence (Franceschi et al., 2000a). Of note, inflammation is not reduced upon aging, and a low-grade, chronic, sterile inflammation, called "inflammaging," seems to be a conserved phenomenon in elderly patients (Franceschi et al., 2000b; Oishi and Manabe, 2016). Indeed, serum levels of IL-6 and CRP increase with age, and their levels are associated with a decline of physical and cognitive performance and predict mortality in the elderly (Puzianowska-Kuźnicka et al., 2016). Human monocytes from the elderly have been shown to express more TNFα (Hearps et al., 2012) and to produce more TLR5-induced IL-8 and IFN-γ-mediated IL-15 (Qian et al., 2012; Lee et al., 2014). In addition, it has been proposed that danger-associated molecular patterns (DAMPs) accumulation, a central driver of inflammation, could be linked with the age-related decline of phagocytic and autophagy activities in macrophages (Oishi and Manabe, 2016; Bulut et al., 2020). Finally, upon aging, senescent cells secrete several inflammatory chemokines and cytokines, a phenomenon termed as the senescence-associated secretory phenotype (SASP). Hall and colleagues have shown that senescent cells can reprogram macrophages, hence termed senescent-associated macrophages (SAMs) (Hall et al., 2016, 2017). SAMs were $p16^+$ and $\beta\text{-gal}^+$, two reliable markers of senescence and displayed both M2-like phenotype and pro-inflammatory profile.

Another contributor of macrophage heterogeneity is clonal hematopoiesis, a process by which genetically distinct subpopulations can be generated from HSC that have underwent DNA point mutations (Beerman et al., 2010). Indeed, it has been widely observed that HSC-derived monocytes gave rise to RTMs upon aging and related accumulated challenges. In addition to addition to leading to the onset of many diseases including blood cancers (Genovese et al., 2014; Jaiswal et al., 2014), this phenomenon could also generate mutated RTM subpopulations involved in specific diseases, cancer notably but also neurodegenerative diseases.

Thus, macrophage heterogeneity will never stop from increasing as we get older. From an innate adaptative point of view, the whole life is a series of successive exposures to various antigens, each of them having an impact on innate immune cells. As a consequence, type, intensity, and temporal sequences of antigen exposure are directly linked to the trained immunity. Recently, the combination of these elements has been called immunological biography or "immuno-biography" (Franceschi et al., 2017). This immuno-biography is considered to be unique for each individual, each one of us having a unique set of heterogeneous trained macrophages. Although still theoretical, medical specialties such as gerontology and geriatrics should pay particular attention in the future, to the immunological anamnesis of each individual to reconstruct their own immuno-biography and predict their subsequent immune responses. However, this will take time before reaching hospital practices as limitless data are impactful and nearly every challenge should be collected: type of delivery (natural *vs.* caesarian), of diet and early nutrition (breast or industrial milk), of infectious diseases and vaccinations, socioeconomic context, ethnicity, psychological status, use of antibiotics, composition of microbiota, and suchlike.

Macrophages in Neurological Diseases

The burden of neurological disorders is increasing as populations are growing and aging. In 2016, disorders of neurological origin were the leading cause of disability-adjusted life years (DALYs) and the second leading cause of deaths. The four largest contributors of neurological DALYs are strokes, migraines, the spectrum of Alzheimer's related dementia and meningitis (Feigin et al., 2019). In the CNS, neuronal and non-neuronal cells are working together. Among non-neuronal cells, "glia" are composed of astrocytes, oligodendrocytes and microglia (Castellani and Schwartz, 2020). As current progresses acknowledge the role of innate immunity and neuroinflammation in driving neurodegenerative disorders, brain-resident macrophages, i.e., microglia, have taken central stage (Heneka et al., 2018; Lenz and Nelson, 2018). Microglia

originate from embryonic-yolk-sac precursors and are self-renewal at steady state. Their apparent heterogeneity has raised several questions regarding their distinct roles in health and diseases (Ginhoux et al., 2010; Ennerfelt and Lukens, 2020).

In a model of neonatal stroke, a recent study reports the contribution of monocytes to microglia during inflammation by using a new mouse fate-mapping model that labels monocyte derivatives (Chen et al., 2020). After neonatal brain ischemia, $CCR2^{+}$ monocytes localized at the ischemic border but were also found in distant peri-infarct sites. At first, these recruited monocytes had an ameboid cell shape and a pro-inflammatory phenotype, but then changed to a more ramified morphology that resemble microglia at day 30, along with the upregulation of microglial gene signatures and M2-like markers. These results suggest a dual function of monocytes after neonatal strokes – i.e., the exacerbation of acute brain damages followed by resolution of inflammation. In addition, they highlighted that infiltrating monocytes undergo *in situ* reprogramming in the brain in order to contribute to the pool of microglia.

Neurodegeneration is defined by an age-related progressive loss of neurons in the central nervous system (CNS), leading to alterations of cognitive performance and dementia (Ramanan and Saykin, 2013). Alzheimer disease (AD) is a neurodegenerative disease with no efficient treatment, characterized by prominent neuroinflammation, extracellular accumulation of amyloid-β (Aβ), and deposition of neurofibrillary tangles in neurons (Castellani and Schwartz, 2020). Here again, microglia have key roles in its pathogenesis. Interestingly, genome-wide association studies (GWAS) in patients with AD have linked mutations in microglial pattern recognition receptors (PRR), including TREM2, with diseased risk (Mhatre et al., 2015). Recently, a new subset of microglia termed "disease-associated microglia" (DAMs) has been identified both in mice and in humans (Keren-Shaul et al., 2017). DAMs expressed genes, which were found associated with AD in human GWAS (Lambert et al., 2013; Keren-Shaul et al., 2017). Genes involved in lipid and metabolic pathways as well as lysosomal and phagocytic capacities are upregulated, including known risk factors of AD, such as APOE and TREM2 (Lambert et al., 2013; Krasemann et al., 2017). DAMs were first detected in the diseased CNS regions, but not in healthy ones. In murine models of AD, DAMs colocalize with Aβ plaques (Keren-Shaul et al., 2017; Mrdjen et al., 2018). Markers of DAM signature were also observed in human AD *postmortem* brains (Friedman et al., 2018). DAMs were shown to be heterogeneous across time as their differentiation appears to be a sequential two-step process: first, microglia shift toward a DAM stage 1 which then moves to a DAM stage 2 (Keren-Shaul et al., 2017; Friedman et al., 2018). This microglial shift, from a homeostatic phenotype to a DAM signature, is believed to rely on the sensing of neurodegeneration-associated molecular patterns (NAMPs). NAMPs, such as Aβ, are danger signals commonly present in various brain conditions, and they are recognized by microglial PRR, such as TREM2 (Deczkowska et al., 2018). TREM2 is a very well-characterized PRR involved in the pathogenesis of AD. Engagement of TREM2 stimulates myeloid cell survival, as well as cytoskeletal reorganization and pro-inflammatory cytokine production (Ulland and Colonna, 2018). Mutations of TREM2 have been reported in patients suffering from late-onset AD (Guerreiro et al., 2013; Jonsson et al., 2013). Several studies have found a beneficial role of TREM2 in Aβ sensing and clearance in various experimental models (Wang et al., 2015, 2016; Ulland et al., 2017; Zhao et al., 2018; Parhizkar et al., 2019). The DAM response is believed to be a protective mechanism aiming at containing neuronal damages, even if several studies are still showing conflicting results, likely due to the heterogeneity of DAMs, as their origin is still yet unclear. However, the discovery of DAMs creates opportunities to develop therapies targeting universal mechanisms of fighting against neuronal death shared by several neurodegenerative conditions (Deczkowska et al., 2018).

MACROPHAGE HETEROGENEITY IN NEOPLASTIC DISORDERS

Among non-communicable diseases, cancer, and all its associated spectrum of diseases, ranks as the leading cause of death (World Health Organization, 2018). Cancer incidence and mortality are still growing worldwide. In 2018, there were an estimated 18.1 million new cases of cancers and 9.6 million deaths from cancers (Bray et al., 2018; Ferlay et al., 2019). Globally, 1 in 6 deaths is due to cancer. In both sexes, lung cancer is the leading cause of cancer death, followed by female breast cancer for incidence and colorectal, stomach, and liver cancers in terms of mortality. In addition, some cancers, such as pancreatic and brain cancers, are less frequent and thus account for a small absolute number of deaths but have a very low five-year survival rate as compared to other cancers like breast cancer.

The emergence of immuno-therapies targeting checkpoint inhibitors during the last decades constituted a major breakthrough in oncology treatment by significantly improving patient prognosis (Couzin-Frankel, 2013). Several inhibitors have reached a market authorization for various cancers, and numerous clinical trials are ongoing worldwide. However, such therapies are expensive, have immune-related adverse events (irAEs) and despite many clinical objective responses, not all patients are responders. Therefore, both clinical and fundamental studies are still urgently needed. Here again, macrophages, called in this context tumor-associated macrophages (TAMs), are the predominant immune cells in the tumor microenvironment (TME) and play a fundamental role in tumor biology.

Macrophages and Lung Cancer

In 2018, 2.1 millions of patients developed a new lung cancer. There were 1.8 million deaths, representing almost 1 in 5 deaths by cancer. The 5-year survival of lung cancer is disastrous, as it is comprised between 15 and 20%. In a mouse model of lung cancer, TAMs have been shown to have a dual origin: they derive both from resident interstitial macrophages (IMs) present before tumorigenesis and from adult monocytes recruited after tumors start to expand (Loyher et al., 2018). However, it has been recently shown that resident IMs also derive from adult monocytes,

whereas AMs are from embryonic origin (Chakarov et al., 2019). As it was shown that AMs were not significant contributors of TAM population in lung cancer (Loyher et al., 2018), we can hypothesize that TAMs actually originate exclusively from adult monocytes, with a varying time of residency within the tissue. In patients suffering from non-small cell lung carcinoma (NSCLC), another recent study highlighted a spatial heterogeneity between TAMs located in the tumor core and TAMs located at invasive margin (Zheng et al., 2020). Indeed, pro-tumoral TAMs were marked especially at the tumor-invasive margin. Moreover, pro-tumoral TAMs were in closer contact to tumor cells as compared to the antitumoral ones. Finally, at the invasive margin, higher proximity of tumor cells to pro-tumoral TAMs and lower proximity to antitumoral TAMs were associated with poor survival (Zheng et al., 2020). Finally, another recent study of human and mouse NSCLC identified a new population of mature dendritic cells enriched in immuno-regulatory molecules (mregDCs) that limit antitumor immunity (Maier et al., 2020). As we can expect to find similar or related programs in TAMs, research in tumoral macrophages needs to be pursued.

Macrophages and Breast Cancer

Worldwide in 2018, 2.1 million newly female breast cancers were diagnosed, accounting for almost 1 in 4 cancer cases among women. A meta-analysis showed that high density of TAMs correlates with poor survival rates and suggested to use TAM density as a prognostic factor (Zhao et al., 2017). Consistent with a pro-tumoral role of TAMs, genetic ablation of *Csf-1* in a murine model of breast cancer (resulting in ablation of macrophages) resulted in delayed tumor development and reduced pulmonary metastasis (Lin et al., 2001). Regarding ontogeny, in a mouse model of mammary tumor (MMTV-PyMT), Franklin et al. showed that TAMs were strictly derived from recruited inflammatory monocytes (Franklin et al., 2014). Interestingly, a study showed in a model of mouse breast cancer, that in response to CCL2 secretion from tumor cells, stromal macrophages were recruited, became intra-epithelial macrophages and induced Wnt-1 production to dismantle E-cadherin junctions, thus promoting early cancer cells dissemination (Linde et al., 2018). In parallel, a recent study (Dawson et al., 2020) identified a new population of ductal macrophages in ductal epithelial structures that were different from the resident stromal macrophages. These ductal macrophages were monocyte-derived and constantly monitored the epithelium throughout breast oncogenesis. Of note, by comparing healthy human breast tissue *vs.* human tissue with lesions of ductal carcinoma *in situ* (DCIS), macrophages were found inside aberrant ductal epithelial structures in between cancer cells that showed reduced E-cadherin levels. Finally, high-grade lesions contained more intra-epithelial macrophages as compared to healthy and low-grade DCIS (Linde et al., 2018).

Besides, another very recent study showed that modulation of TREM2 had a remarkable impact on TAM landscape (Molgora et al., 2020). As in some diseases discussed ahead, TREM2 has already been reported in tumors (Lavin et al., 2017; Song et al., 2019). Indeed, authors showed that *Trem2*$^{-/-}$ mice were more resistant to tumor growth than WT mice in a mammary tumor mouse model. *Trem2* deficiency showed alterations in macrophage populations and an increase of tumor-infiltrating CD8^{+} T cells expressing PD-1. Of note, authors showed that anti-PD-1 therapy was more effective in *Trem2*-deficient mice than in WT mice. Furthermore, anti-TREM2 mAb dampened tumor growth and highly enhanced the efficiency of anti-PD-1 immunotherapy. *Trem2* deficiency and anti-TREM2 mAb were responsible for changes in the tumor-infiltrating macrophages: CX3CR1^{+} and CD206^{+} macrophage subsets declined, while other subsets were induced. Lastly, authors found that TREM2 was a marker of TAM in more than 200 human tumors and that its expression was inversely correlated with greater relapse-free survival and overall survival (OS) and in triple-negative breast cancer. To sum up, TAM remodeling by specifically targeting TREM2 could be a promising avenue for complementing checkpoint immuno-therapy (Molgora et al., 2020). A concomitant study also identified a TREM2^{+} regulatory monocytes in a model of mouse fibrosarcoma (Katzenelenbogen et al., 2020). By coupling scRNA-Seq and intracellular protein activity, authors showed that this population was associated with more dysfunctional CD8^{+} T cells and tumor growth.

Macrophages and Pancreatic Cancer

Although not the most frequent, pancreatic cancer is associated with one of the worst prognoses. In mice, in a model of pancreatic ductal adenocarcinoma, TAM has been shown to originate from both embryonic-RTMs and recruited inflammatory monocytes (Zhu et al., 2017). Embryonic TAMs exhibited a pro-fibrotic profile with increased expression of genes involved in extracellular matrix deposing and remodeling, which is a hallmark of pancreatic ductal adenocarcinoma. In contrast, monocyte-derived TAMs were more efficient antigen-presenting cells (Zhu et al., 2017). Here again, such results argue for a role of macrophage ontogeny in tumor pathogenesis and these findings should be taken into account for future TAM-targeted therapies.

Macrophages and Liver Cancer

Liver dysfunctions from NAFLD to NASH, cirrhosis, and hepatocellular carcinoma (HCC) account for 2 million deaths per year (Williams et al., 2014). HCC was the second leading cause of years of life lost (YLLs) from cancer worldwide between 2005 and 2015 (Yang et al., 2019). According to the staging, various treatments are recommended, ranging from surgical resection to systemic chemotherapy (Marrero et al., 2018). We have recently observed that a subpopulation of FOLR2^{+} TAMs underwent an onco-fetal reprogramming, meaning that these adult cells can acquire a transcriptomic profile similar to fetal macrophages in the specific context of liver cancer. Considering that fetal macrophages are strongly tissue-supportive and take part in the organism development, the reprogramming could explain why TAMs are pro-tumoral. Of note, this phenomenon, mediated by tumoral endothelial cells (ECs) (Sharma et al., 2020), is only partial and does not concerned other TAM subpopulations, such as SPP1^{+} and MT1G^{+} TAMs. Although therapeutic strategies are yet to be designed to modulate this onco-fetal reprogramming, these results unambiguously demonstrate that TAM-oriented immunotherapies need to be very precisely designed in order to be specific and efficient with limited side effects.

Macrophages and Brain Cancer

Brain cancer accounted for only 2.5% of all deaths by cancer in 2018 worldwide. However, as in the pancreas, the five-year mortality rate is disastrous. Studies have identified two TAM subsets in human glioma: one was from embryonic origin and the other one originated from adult bone marrow monocytes (Bowman et al., 2016; Müller et al., 2017). Muller et al. showed that monocyte-derived TAMs, but not embryonic TAMs, were correlated with shorter OS in low-grade glioma. Interestingly, Bowman et al. identified *Itga4* (*Cd49d*) as an effective marker to distinguish embryonic TAMs from monocyte-derived TAMs TAMs both in mice and in humans. They also showed that while embryonic- and monocyte-derived TAMs shared features of tumor education, they exhibited distinct activation states: embryonic TAMs were enriched in pro-inflammatory genes as well as factors involved in extracellular matrix (ECM) remodeling, while monocyte-derived TAMs exhibited an immuno-suppressive signature. Their data suggest that these different faculties resulted from inherent transcriptional networks poised before the onset of tumorigenesis. Therefore, as these chromatin landscapes were established earlier of tumor initiation, it suggests that ontogenically unrelated cells can be engaged in distinct macrophage activation states (Bowman et al., 2016). In addition, Muller et al., showed that embryonic- and monocyte-derived TAMs were enriched in distinct tumor-anatomical structures and that both of them had a gene signature of both M1- and M2-like (Müller et al., 2017).

Macrophages and Metastasis

Macrophages are key players in the formation of pre-metastatic niches (Lopez-Yrigoyen et al., 2020). Indeed, in the primary tumor, they help tumor cells to escape from immune recognition and they prepare distant "pre-metastatic" sites for tumor cells to colonize (Paget, 1889). These "pre-metastatic" niches are shaped by systemic influences of primary tumor through recruitment of monocytes that in turn attract tumor cells by chemokines. Macrophages also remodel the ECM to promote angiogenesis, epithelial-to-mesenchymal transition, and extravasation. Therefore, they enhance both tumor cell tropism and their ability to seed and survives (Psaila and Lyden, 2009). Once tumor cells arrive at these "pre-metastatic" sites, a distinct subset of macrophages termed metastasis-associated-macrophages (MAMs) promotes tumor cell extravasation and growth (Qian et al., 2011). Therefore, MAMs derive exclusively from monocytes, are pro-tumoral (Kitamura et al., 2018), and have been shown to limit the efficacy of classical cancer therapies such as chemotherapy, radiotherapy, and biological therapies (Lopez-Yrigoyen et al., 2020).

Macrophages and the 3 E's Theory

In the early twentieth century, Paul Ehrlich conceived the idea that the immune system could suppress an "*overwhelming frequency*" of carcinoma (Ehrlich, 1909). The revisiting of the Ehrlich proposal had to await the maturation of immunology, and the concept of "*cancer immuno-surveillance*" was proposed in 1957 by Burnett and Thomas. It was defined as follows: *"In large, long-lived animals, like most of the warm-blooded vertebrates, inheritable genetic changes must be common in somatic cells and a proportion of these changes will represent a step toward malignancy. It is an evolutionary necessity that there should be some mechanism for eliminating or inactivating such potentially dangerous mutant cells and it is postulated that this mechanism is of immunological character*." Subsequently, the immuno-surveillance concept was hardly challenged because numerous *in vivo* experiments failed to prove it. Therefore, this theory was rapidly forgotten and relegated to the historical dustbin. A major review published in 2000, which listed the six critical hurdles that a new tumor must circumvent to grow and survive, did not even mention the basal immune response against tumors (Hanahan and Weinberg, 2000). However, because of growing progress in immunology and comprehension of some limitations of mouse model, the new millenium has witnessed the revival of this old debated idea. A second major review published 11 years later by the same authors added 4 supplemental hallmarks. Among them were (i) the avoidance of immune destruction and (ii) the tumor-promoting inflammation by innate immune cells (Hanahan and Weinberg, 2011). Since then, the immuno-surveillance hypothesis has shifted to the "cancer immuno-editing" concept that proceeds through three phases termed "the three Es" for elimination, equilibrium, and escape (Dunn et al., 2002). Although the role of NK and T cells has been very well defined in these processes, TAM implication should also be taken into account.

Indeed, revisiting the 3 E's theory of tumor immuno-editing through the lens of TAM biology is an interesting way to highlight their incredible heterogeneity through time (**Figure 2**). During the earliest stage of tumor onset, i.e., the elimination phase, TAMs are antitumoral and collaborate with adaptive immune cells to recognize and eliminate tumor cells (Dunn et al., 2004). Here, TAM ontogeny remains only partially characterized and should be further investigated to understand its precise contribution to this antitumoral phase. Thereafter, tumor cells able to survive the elimination step can proceed to the equilibrium phase, a state of tumor quiescence where net growth is limited, and cellular immunogenicity edited. During this phase, tumor cells slowly influence the TME to provide support for their growth. There, the majority of TAMs are slowly educated to enhance tumor progression and switch from an antitumoral state to a pro-tumoral one. Of note, ontogeny may have a role in this phenotype shift. Indeed, embryonic TAMs and monocyte-derived TAMs may exhibit different epigenetic profiles that could influence their subsequent education by tumor cells. Thus, understanding the role of ontogenic dimension and the tumoral cues that underlies this antitumoral to pro-tumoral transformation would help blocking or delaying TAM education, which is of course of major interest for preventive medicine. Finally, edited tumors enter the escape phase, where their growth becomes unrestrained and become clinically detected. Patients are generally receiving their first line of treatment during this escape phase. Studies of patients treated with immuno-therapies indicate that the immuno-editing process can reoccur in response to treatment (O'Donnell et al., 2019). Indeed, in case of treatment failure, no objective response is observed, and tumor eventually stays in the escape phase. Alternatively, a more efficient but still inadequate treatment might drive tumor into an on-treatment equilibrium phase, which would be characterized by a partial

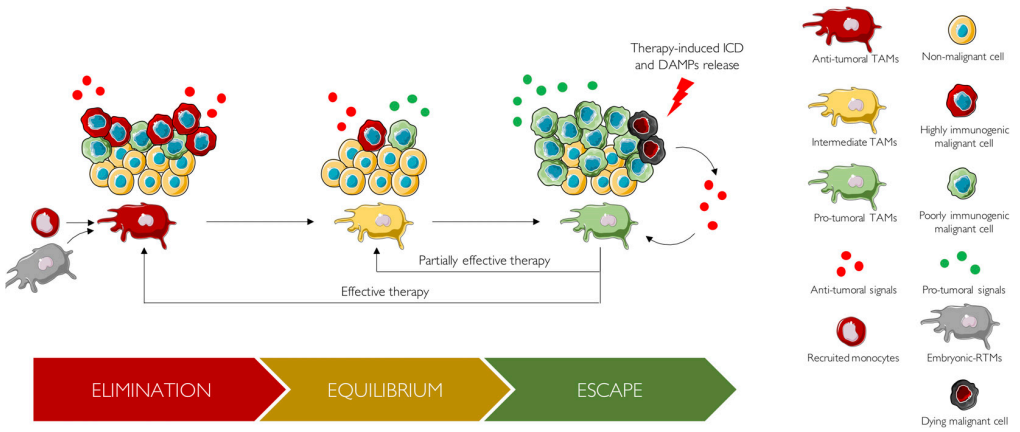

FIGURE 2 | The 3 E's of cancer immuno-editing from a TAM perspective. During the elimination phase, TAMs are antitumoral and collaborate with adaptive immune cells to eliminate highly immunogenic tumor cells. These TAMs arise either from recruited monocytes or from embryonic RTMs. Thereafter, tumor cells able to survive the elimination step, proceed to the equilibrium phase, a state of tumor quiescence where net growth is limited and cellular immunogenicity edited. During this phase, tumor cells slowly influence the TME to provide support for their growth: TAMs are slowly educated to enhance tumor progression and switch from an antitumoral state to a pro-tumoral one. Pro-tumoral cues responsible for TAM reprogramming, as well as the role of ontogeny, should be studied in further details. Finally, edited tumors enter the escape phase. There, tumor growth becomes unrestrained, tumors are clinically detected, and patients are receiving their first line of treatment. In case of treatment failure, no objective response is observed, and tumor eventually stays in the escape phase. Alternatively, a partially effective therapy, might drive tumor into an on-treatment equilibrium phase, which would be characterized by a partial response. Finally, effective therapies can drive tumors back to the elimination phase, which will be reflected by a complete response. Upon treatment, tumor cells undergo immunogenic cell death (ICD) and released DAMPs. These DAMPs boost anticancer immune response by reprogramming antitumoral TAMs into pro-tumoral TAMs.

response. Finally, effective therapies can drive tumors back to the elimination phase, which will be reflected by a complete response.

Interestingly, it is now well known that antineoplastic agents such as anthracyclines, oxaliplatin, and crizotinib stimulate the liberation of DAMPs, from dying cancer cells via the induction of immunogenic cell death (ICD) (Galluzzi et al., 2017). Once released, DAMPs operate as immunological adjuvants and boost anticancer immune responses by converting "cold" into "hot" tumors, characterized by accumulation of pro-inflammatory cytokines and T cell infiltration, resulting in a better response rate to immune-checkpoint blockers (Galluzzi et al., 2017). Numerous DAMPs released by ICD have been already described, such as HMGB1 and ATP (Galluzzi et al., 2020). They have been shown to induce a shift from pro-tumoral to antitumoral TAM (Li et al., 2017). In addition, a very recent study showed that radiation induces the release of microparticles from tumor cells, which induces as well the reprogramming of TAM polarization from a pro-tumoral to an antitumoral phenotype (Wan et al., 2020). To sum up, TAMs are a highly plastic and very heterogeneous population able to change across time as the tumor evolves and upon challenges such as therapies.

CONCLUDING REMARKS

Since their first description by Ilya Ilitch Metchnikov in the late 19th century (Metchnikov, 1883), macrophages have become the focus of many studies. A lot of knowledge has been accumulated over the years shaping our actual understanding of these immune cells in both healthy and diseased tissues. The current paradigm, written in textbooks and taught to students, is that macrophages are phagocytes in charge of the tissue immune surveillance that they exert by ingesting every foreign particle they can catch, making them the prototypal innate immune cells. For this reason, macrophages for which phagocytosis represents the main function are often considered as archaic cells, only assigned as pro-(M1) or anti-(M2) inflammatory and are too often overlooked in the design of innovative therapeutic strategies. Our aim herein was to rehabilitate them as central players of tissue dysbiosis observed in almost all types of human diseases. Indeed, macrophages represent often the most abundant immune cell population in diseased tissues. Furthermore, their inherent plasticity allows them to display a multitude of phenotypes, either supporting or restricting disease development. So, clearly a global strategy neutralizing or depleting macrophage population as a whole is no more conceivable based on the current knowledge. It is now time to target more specifically macrophage populations that support pathogenesis. This could only be done by clarifying the major programs shaping macrophage biology in a time- and spatial-dependent manner. For this, global integration of existing data is needed and is actually in progress in many laboratories and will definitely improve our fundamental knowledge and upcoming therapeutic strategies.

AUTHOR CONTRIBUTIONS

GG, CB, and FG discussed and established the plan of the manuscript and approved the final version for publication. GG and CB wrote the draft. FG provided a substantial intellectual contribution and edited the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

Alivernini, S., MacDonald, L., Elmesmari, A., Finlay, S., Tolusso, B., Gigante, M. R., et al. (2020). Distinct synovial tissue macrophage subsets regulate inflammation and remission in rheumatoid arthritis. *Nat. Med.* 2020:939. doi: 10.1038/s41591-020-0939-8

Aprahamian, T., Takemura, Y., Goukassian, D., and Walsh, K. (2008). Ageing is associated with diminished apoptotic cell clearance in vivo. *Clin. Exp. Immunol.* 152, 448–455. doi: 10.1111/j.1365-2249.2008.03658.x

Arts, R. J. W., Novakovic, B., ter Horst, R., Carvalho, A., Bekkering, S., Lachmandas, E., et al. (2016). Glutaminolysis and Fumarate Accumulation Integrate Immunometabolic and Epigenetic Programs in Trained Immunity. *Cell Metab.* 24, 807–819. doi: 10.1016/j.cmet.2016.10.008

Baurakiades, E., Costa, V. H., Raboni, S. M., De Almeida, V. R. T., Larsen, K. S. K., Kohler, J. N., et al. (2014). The roles of ADAM33, ADAM28, IL-13 and IL-4 in the development of lung injuries in children with lethal non-pandemic acute infectious pneumonia. *J. Clin. Virol.* 61, 585–589. doi: 10.1016/j.jcv.2014.10.004

Beerman, I., Bhattacharya, D., Zandi, S., Sigvardsson, M., Weissman, I. L., Brydere, D., et al. (2010). Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proc. Natl. Acad. Sci. U. S. A.* 107, 5465–5470. doi: 10.1073/pnas.1000834107

Bleriot, C., Chakarov, S., and Ginhoux, F. (2020). Review Determinants of Resident Tissue Macrophage Identity and Function. *Immunity* 52, 957–970. doi: 10.1016/j.immuni.2020.05.014

Blériot, C., Dupuis, T., Jouvion, G., Eberl, G., Disson, O., and Lecuit, M. (2015). Liver-Resident Macrophage Necroptosis Orchestrates Type 1 Microbicidal Inflammation and Type-2-Mediated Tissue Repair during Bacterial Infection. *Immunity* 42, 145–158. doi: 10.1016/j.immuni.2014.12.020

Bonnardel, J., T'Jonck, W., Gaublomme, D., Browaeys, R., Scott, C. L., Martens, L., et al. (2019). Stellate Cells, Hepatocytes, and Endothelial Cells Imprint the Kupffer Cell Identity on Monocytes Colonizing the Liver Macrophage Niche. *Immunity* 51, 638.e–654.e.

Boomer, J. S., To, K., Chang, K. C., Takasu, O., Osborne, D. F., Walton, A. H., et al. (2012). Immunosuppression in Patients Who Die of Sepsis and Multiple Organ Failure. *Surv. Anesthesiol.* 56, 272–273. doi: 10.1097/sa.0b013e3182751ec1

Bowman, R. L., Klemm, F., Akkari, L., Pyonteck, S. M., Sevenich, L., Quail, D. F., et al. (2016). Macrophage Ontogeny Underlies Differences in Tumor-Specific Education in Brain Malignancies. *Cell Rep.* 17, 2445–2459. doi: 10.1016/j.celrep.2016.10.052

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., and White, T. C. (2012). Hidden killers: Human fungal infections. *Sci. Transl. Med.* 4, 1–10.

Bulut, O., Kilic, G., Domínguez-Andrés, J., and Netea, M. G. (2020). Overcoming immune dysfunction in the elderly: trained immunity as a novel approach. *Int. Immunol.* 2020:221075933 doi: 10.1093/intimm/dxaa052

Castagnoli, R., Votto, M., Licari, A., Brambilla, I., Bruno, R., Perlini, S., et al. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection in Children and Adolescents: A Systematic Review. *JAMA Pediatr.* 2, 882–889.

Castellani, G., and Schwartz, M. (2020). Immunological Features of Non-neuronal Brain Cells: Implications for Alzheimer's Disease Immunotherapy. *Trends Immunol.* 41, 794–804. doi: 10.1016/j.it.2020.07.005

Chakarov, S., Lim, H. Y., Tan, L., Lim, S. Y., See, P., Lum, J., et al. (2019). Two distinct interstitial macrophage populations coexist across tissues in specific subtissular niches. *Science* 80:363.

Charlier, C., Perrodeau, É, Leclercq, A., Cazenave, B., Pilmis, B., Henry, B., et al. (2017). Clinical features and prognostic factors of listeriosis: the MONALISA national prospective cohort study. *Lancet Infect. Dis.* 17, 510–519.

Chen, H., Sun, Y., Chen, C., Kuo, Y., Kuan, I. S., Li, Z. T., et al. (2020). Fate mapping via CCR2-CreER mice reveals monocyte-to-microglia transition in development and neonatal stroke. *Sci. Adv.* 6, 1–13.

Christ, A., Bekkering, S., Latz, E., and Riksen, N. P. (2016). Long-term activation of the innate immune system in atherosclerosis. *Semin. Immunol.* 28, 384–393. doi: 10.1016/j.smim.2016.04.004

Coats, B. R., Schoenfelt, K. Q., Barbosa-Lorenzi, V. C., Peris, E., Cui, C., Hoffman, A., et al. (2017). Metabolically Activated Adipose Tissue Macrophages Perform Detrimental and Beneficial Functions during Diet-Induced Obesity. *Cell Rep.* 20, 3149–3161. doi: 10.1016/j.celrep.2017.08.096

Cochain, C., Vafadarnejad, E., Arampatzi, P., Pelisek, J., Winkels, H., Ley, K., et al. (2018). Single-cell RNA-seq reveals the transcriptional landscape and heterogeneity of aortic macrophages in murine atherosclerosis. *Circ. Res.* 122, 1661–1674. doi: 10.1161/circresaha.117.312509

Cohen, J. C., Horton, J. D., and Hobbs, H. H. (2011). Human fatty liver disease: Old questions and new insights. *Science* 332, 1519–1523. doi: 10.1126/science.1204265

Cole, J. E., Park, I., Ahern, D. J., Kassiteridi, C., Abeam, D. D., Goddard, M. E., et al. (2018). Immune cell census in murine atherosclerosis: Cytometry by time of flight illuminates vascular myeloid cell diversity. *Cardiovasc. Res.* 114, 1360–1371. doi: 10.1093/cvr/cvy109

Corrales-Medina, V. F., Musher, D. M., Shachkina, S., and Chirinos, J. A. (2013). Acute pneumonia and the cardiovascular system. *Lancet* 381, 496–505. doi: 10.1016/s0140-6736(12)61266-5

Couzin-Frankel, J. (2013). Breakthrough of the year 2013. *Cancer Immunother. Sci.* 342, 1432–1433.

Cross, M., Smith, E., Hoy, D., Carmona, L., Wolfe, F., Vos, T., et al. (2014). The global burden of rheumatoid arthritis: Estimates from the Global Burden of Disease 2010 study. *Ann. Rheum. Dis.* 73, 1316–1322. doi: 10.1136/annrheumdis-2013-204627

Culemann, S., Grüneboom, A., Nicolás-Ávila, J. Á, Weidner, D., Lämmle, K. F., Rothe, T., et al. (2019). Locally renewing resident synovial macrophages provide a protective barrier for the joint. *Nature* 572, 670–675. doi: 10.1038/s41586-019-1471-1

Dawson, C. A., Pal, B., Vaillant, F., Gandolfo, L. C., Liu, Z., Bleriot, C., et al. (2020). Tissue-resident ductal macrophages survey the mammary epithelium and facilitate tissue remodelling. *Nat. Cell Biol.* 22, 546–558.doi: 10.1038/s41556-020-0505-0

de Laval, B., Maurizio, J., Kandalla, P. K., Brisou, G., Simonnet, L., Huber, C., et al. (2020). C/EBPβ-Dependent Epigenetic Memory Induces Trained Immunity in Hematopoietic Stem Cells. *Cell Stem Cell* 26, 657.e–674.e.

Deczkowska, A., Keren-Shaul, H., Weiner, A., Colonna, M., Schwartz, M., and Amit, I. (2018). Disease-Associated Microglia: A Universal Immune Sensor of Neurodegeneration. *Cell* 173, 1073–1081. doi: 10.1016/j.cell.2018.05.003

Delano, M. J., and Ward, P. A. (2016). Sepsis-induced immune dysfunction: can immune therapies reduce mortality? *J. Clin. Invest.* 126, 23–31.

Diehl, A. M., and Day, C. (2017). Cause, pathogenesis, and treatment of nonalcoholic steatohepatitis. *N. Engl. J. Med.* 377, 2063–2072. doi: 10.1056/nejmra1503519

Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J., and Schreiber, R. D. (2002). Cancer immunoediting: From immunosurveillance to tumor escape. *Nat. Immunol.* 3, 991–998. doi: 10.1038/ni1102-991

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

Ehrlich, P. (1909). Ueber den jetzigen Stand der Karzinomforschung. *Ned. Tijdschr. Geneeskd.* vol. 5, 273–290.

Ennerfelt, H. E., and Lukens, J. R. (2020). The role of innate immunity in Alzheimer's disease. *Immunol. Rev.* 297, 1–22.

Ensan, S., Li, A., Besla, R., Degousee, N., Cosme, J., Roufaiel, M., et al. (2016). Self-renewing resident arterial macrophages arise from embryonic CX3CR1 + precursors and circulating monocytes immediately after birth. *Nat. Immunol.* 17, 159–168. doi: 10.1038/ni.3343

Fairfax, K., Nascimento, M., Huang, S. C. C., Everts, B., and Pearce, E. J. (2012). Th2 responses in schistosomiasis. *Semin. Immunopathol.* 34, 863–871. doi: 10.1007/s00281-012-0354-4

Fearon, D. T., and Locksley, R. M. (1996). The Instructive Role of Innate Immunity in the Acquired Immune Response. *Science* 272, 50–54. doi: 10.1126/science.272.5258.50

Feigin, V. L., Nichols, E., Alam, T., Bannick, M. S., Beghi, E., Blake, N., et al. (2019). Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 18, 459–480.

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., et al. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* 144, 1941–1953. doi: 10.1002/ijc.31937

Fernandez, D. M., Rahman, A. H., Fernandez, N. F., Chudnovskiy, A., Amir, E. D., Amadori, L., et al. (2019). Single-cell immune landscape of human atherosclerotic plaques. *Nat. Med.* 25, 1576–1588.

Fleischmann, C., Scherag, A., Adhikari, N. K. J., Hartog, C. S., Tsaganos, T., Schlattmann, P., et al. (2016). Assessment of global incidence and mortality of hospital-treated sepsis current estimates and limitations. *Am. J. Respir. Crit. Care Med.* 193, 259–272. doi: 10.1164/rccm.201504-0781oc

Foster, S. L., Hargreaves, D. C., and Medzhitov, R. (2007). Gene-specific control of inflammation by TLR-induced chromatin modifications. *Nature* 447, 972–978. doi: 10.1038/nature05836

Franceschi, C., Bonafè, M., and Valensin, S. (2000a). Human immunosenescence: The prevailing of innate immunity, the failing of clonotypic immunity, and the filling of immunological space. *Vaccine* 18, 1717–1720. doi: 10.1016/s0264-410x(99)00513-7

Franceschi, C., Bonafè, M., Valensin, S., Olivieri, F., De Luca, M., Ottaviani, E., et al. (2000b). Inflamm-aging. An evolutionary perspective on immunosenescence. *Ann. N. Y. Acad. Sci.* 908, 244–254. doi: 10.1111/j.1749-6632.2000.tb06651.x

Franceschi, C., Salvioli, S., Garagnani, P., de Eguileor, M., Monti, D., and Capri, M. (2017). Immunobiography and the heterogeneity of immune responses in the elderly: A focus on inflammaging and trained immunity. *Front. Immunol.* 8:1–11. doi: 10.3389/fimmu.2017.00982

Franklin, R. A., Liao, W., Sarkar, A., Kim, M. V., Bivona, M. R., Liu, K., et al. (2014). The cellular and molecular origin of tumor-associated macrophages. *Science* 344, 921–925.

Friedman, B. A., Srinivasan, K., Ayalon, G., Meilandt, W. J., Lin, H., Huntley, M. A., et al. (2018). Diverse Brain Myeloid Expression Profiles Reveal Distinct Microglial Activation States and Aspects of Alzheimer's Disease Not Evident in Mouse Models. *Cell Rep.* 22, 832–847. doi: 10.1016/j.celrep.2017.12.066

Galluzzi, L., Buqué, A., Kepp, O., Zitvogel, L., and Kroemer, G. (2017). Immunogenic cell death in cancer and infectious disease. *Nat. Rev. Immunol.* 17, 97–111. doi: 10.1038/nri.2016.107

Galluzzi, L., Vitale, I., Warren, S., Adjemian, S., Agostinis, P., Martinez, A. B., et al. (2020). Consensus guidelines for the definition, detection and interpretation of immunogenic cell death. *J. Immunother. Cancer* 8, 1–21.

Genovese, G., Kähler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* 371, 2477–2487. doi: 10.1056/nejmoa1409405

Gieseck, R. L., Wilson, M. S., and Wynn, T. A. (2018). Type 2 immunity in tissue repair and fibrosis. *Nat. Rev. Immunol.* 18, 62–76. doi: 10.1038/nri.2017.90

Ginhoux, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., et al. (2010). Fate Mapping Analysis Reveals That Adult Microglia Derive from Primitive Macrophages. *Science* 701, 841–845. doi: 10.1126/science.1194637

Ginhoux, F., Schultze, J. L., Murray, P. J., Ochando, J., and Biswas, S. K. (2016). New insights into the multidimensional concept of macrophage ontogeny, activation and function. *Nat. Immunol.* 17, 34–40. doi: 10.1038/ni.3324

Glass, C. K., and Natoli, G. (2016). Molecular control of activation and priming in macrophages. *Nat. Immunol.* 17, 26–33. doi: 10.1038/ni.3306

Gomez Perdiguero, E., Klapproth, K., Schulz, C., Busch, K., Azzoni, E., Crozet, L., et al. (2015). Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature* 518, 547–551. doi: 10.1038/nature13989

Grubeck-Loebenstein, B., Bella, S., Della Iorio, A. M., Michel, J.-P., Pawelec, G., and Solana, R. (2009). Immunosenescence and vaccine failure in the elderly. *Aging Clin. Exp. Res.* 21, 201–209. doi: 10.1007/bf03324904

Grünig, G., Warnock, M., Wakil, A. E., Venkayya, R., Brombacher, F., Rennick, D. M., et al. (1999). Requirement for IL-13 independently of IL-4 in experimental asthma. *Pneumologie* 53, 6–8.

Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* 382, 1708–1720.

Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., et al. (2013). TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.* 368, 117–127.

Hall, B. M., Balan, V., Gleiberman, A. S., Strom, E., Krasnov, P., Virtuoso, L. P., et al. (2016). Aging of mice is associated with p16(Ink4a)- and β-galactosidasepositive macrophage accumulation that can be induced in young mice by senescent cells. *Aging* 8, 1294–1315. doi: 10.18632/aging.100991

Hall, B. M., Balan, V., Gleiberman, A. S., Strom, E., Krasnov, P., Virtuoso, P., et al. (2017). p16(Ink4a) and senescence−associated β−galactosidase can be induced in macrophages as part of a reversible response to physiological stimuli. *Aging* 9, 1867–1884. doi: 10.18632/aging.101268

Hanahan, D., and Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell* 100, 57–70.

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

Hansson, G. K., and Hermansson, A. (2011). The immune system in atherosclerosis. *Nat. Immunol.* 12, 204–212.

Hashimoto, D., Chow, A., Noizat, C., Teo, P., Beasley, M. B., Leboeuf, M., et al. (2013). Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. *Immunity* 38, 792–804. doi: 10.1016/j.immuni.2013.04.004

Hearps, A. C., Martin, G. E., Angelovich, T. A., Cheng, W. J., Maisa, A., Landay, A. L., et al. (2012). Aging is associated with chronic innate immune activation and dysregulation of monocyte phenotype and function. *Aging Cell* 11, 867–875. doi: 10.1111/j.1474-9726.2012.00851.x

Heitmann, L., Dar, M. A., Schreiber, T., Erdmann, H., Behrends, J., Mckenzie, A. N. J., et al. (2014). The IL-13/IL-4Rα axis is involved in tuberculosis-associated pathology. *J. Pathol.* 234, 338–350. doi: 10.1002/path.4399

Heneka, M. T., McManus, R. M., and Latz, E. (2018). Inflammasome signalling in brain function and neurodegenerative disease. *Nat. Rev. Neurosci.* 19, 610–621. doi: 10.1038/s41583-018-0055-7

Herrero, C., Sebastian, C., Marques, L., Comalada, M., Xaus, J., Valledor, A. F., et al. (2002). Immunosenescence of macrophages: reduced MHC class II gene expression. *Exp. Gerontol.* 37, 389–394.

Heymsfield, S. B., and Wadden, T. A. (2017). Mechanisms, pathophysiology, and management of obesity. *N. Engl. J. Med.* 376, 254–266.

Hill, D. A., Lim, H. W., Kim, Y. H., Ho, W. Y., Foong, Y. H., Nelson, V. L., et al. (2018). Distinct macrophage populations direct inflammatory versus physiological changes in adipose tissue. *Proc. Natl. Acad. Sci. U. S. A.* 115, E5096–E5105.

Hoffmann, J. A., Kafatos, F. C., Janeway, C. A., and Ezekowitz, R. A. B. (1999). Phylogenetic perspectives in innate immunity. *Science* 284, 1313–1318. doi: 10.1126/science.284.5418.1313

Huang, L., Nazarova, E. V., Tan, S., Liu, Y., and Russell, D. G. (2018). Growth of Mycobacterium tuberculosis in vivo segregates with host macrophage metabolism and ontogeny. *J. Exp. Med.* 215, 1135–1152. doi: 10.1084/jem.20172020

Hunter, T. M., Boytsov, N. N., Zhang, X., Schroeder, K., Michaud, K., and Araujo, A. B. (2017). Prevalence of rheumatoid arthritis in the United States adult population in healthcare claims databases, 2004–2014. *Rheumatol. Int.* 37, 1551–1557. doi: 10.1007/s00296-017-3726-1

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B. G., et al. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* 371, 2488–2498.

Jaitin, D. A., Adlung, L., Thaiss, C. A., Weiner, A., Li, B., Descamps, H., et al. (2019). Lipid-Associated Macrophages Control Metabolic Homeostasis in a Trem2-Dependent Manner. *Cell* 178, 686.e–698.e.

Jakubzick, C., Choi, E. S., Joshi, B. H., Keane, M. P., Kunkel, S. L., Puri, R. K., et al. (2003). Therapeutic Attenuation of Pulmonary Fibrosis Via Targeting of IL-4- and IL-13-Responsive Cells. *J. Immunol.* 171, 2684–2693. doi: 10.4049/jimmunol.171.5.2684

Janeway, C. A. (1989). Approaching the asymptote? *Evol.Revol. Immunol. Cold Spring Harb. Symp. Quant. Biol.* 54, 1–13. doi: 10.1101/sqb.1989.054.01.003

Jonsson, T., Stefansson, H., Steinberg, S., Jonsdottir, I., Jonsson, P. V., Snaedal, J., et al. (2013). Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* 368, 107–116.

Katzenelenbogen, Y., Sheban, F., Yalin, A., Yofe, I., Svetlichnyy, D., Jaitin, D. A., et al. (2020). Coupled scRNA-Seq and Intracellular Protein Activity Reveal an Immunosuppressive Role of TREM2 in Cancer. *Cell* 182, 872.e–885.e.

Kaufmann, E., Sanz, J., Dunn, J. L., Khan, N., Mendonça, L. E., Pacis, A., et al. (2018). BCG Educates Hematopoietic Stem Cells to Generate Protective Innate Immunity against Tuberculosis. *Cell* 172, 176.e–190.e.

Keane, M. P., Gomperts, B. N., Weigt, S., Xue, Y. Y., Burdick, M. D., Nakamura, H., et al. (2007). IL-13 Is Pivotal in the Fibro-Obliterative Process of Bronchiolitis Obliterans Syndrome. *J. Immunol.* 178, 511–519. doi: 10.4049/jimmunol.178.1.511

Kennedy, B. K., Berger, S. L., Brunet, A., Campisi, J., Cuervo, A. M., Epel, E. S., et al. (2014). Geroscience: Linking aging to chronic disease. *Cell* 159, 709–713. doi: 10.1016/j.cell.2014.10.039

Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T. K., et al. (2017). A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* 169, 1276.e–1290.e.

Kim, K., Shim, D., Lee, J. S., Zaitsev, K., Williams, J. W., Kim, K. W., et al. (2018). Transcriptome analysis reveals nonfoamy rather than foamy plaque macrophages are proinflammatory in atherosclerotic murine models. *Circ. Res.* 123, 1127–1142. doi: 10.1161/circresaha.118.312804

Kitamura, T., Doughty-Shenton, D., Cassetta, L., Fragkogianni, S., Brownlie, D., Kato, Y., et al. (2018). Monocytes differentiate to immune suppressive precursors of metastasis-associated macrophages in mouse models of metastatic breast cancer. *Front. Immunol.* 8:2004. doi: 10.3389/fimmu.2017.02004

Kleinnijenhuis, J., Quintin, J., Preijers, F., Joosten, L. A. B., Ifrim, D. C., Saeed, S., et al. (2012). Bacille Calmette-Guérin induces NOD2-dependent nonspecific protection from reinfection via epigenetic reprogramming of monocytes. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17537–17542. doi: 10.1073/pnas.1202870109

Krasemann, S., Madore, C., Cialic, R., Baufeld, C., Calcagno, N., El Fatimy, R., et al. (2017). The TREM2-APOE Pathway Drives the Transcriptional Phenotype of Dysfunctional Microglia in Neurodegenerative Diseases. *Immunity* 47, 566.e–581.e.

Kratz, M., Coats, B. R., Hisert, K. B., Hagman, D., Mutskov, V., Peris, E., et al. (2014). Metabolic dysfunction drives a mechanistically distinct proinflammatory phenotype in adipose tissue macrophages. *Cell Metab.* 20, 614–625. doi: 10.1016/j.cmet.2014.08.010

Lajqi, T., Lang, G. P., Haas, F., Williams, D. L., Hudalla, H., Bauer, M., et al. (2019). Memory-Like Inflammatory Responses of Microglia to Rising Doses of LPS: Key Role of PI3Kγ. *Front. Immunol.* 10:1–14. doi: 10.3389/fimmu.2019.02492

Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* 45, 1452–1458.

Lavin, Y., Kobayashi, S., Leader, A., Amir, E., Ad, D., Elefant, N., et al. (2017). Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. *Cell* 169, 750.e–765.e.

Lecuit, M. (2020). Listeria monocytogenes, a model in infection biology. *Cell. Microbiol.* 22, 1–8.

Lee, D. M., and Weinblatt, M. E. (2001). Rheumatoid arthritis. *Lancet* 358, 903–911.

Lee, N., Shin, M. S., Kang, K. S., Yoo, S. A., Mohanty, S., Montgomery, R. R., et al. (2014). Human monocytes have increased IFN-γ-mediated IL-15 production with age alongside altered IFN-γ receptor signaling. *Clin. Immunol.* 152, 101–110. doi: 10.1016/j.clim.2014.03.003

Leentjens, J., Bekkering, S., Joosten, L. A. B., Netea, M. G., Burgner, D. P., and Riksen, N. P. (2018). Trained innate immunity as a novel mechanism linking infection and the development of atherosclerosis. *Circ. Res.* 122, 664–669. doi: 10.1161/circresaha.117.312465

Leentjens, J., Kox, M., Van Der Hoeven, J. G., Netea, M. G., and Pickkers, P. (2013). Immunotherapy for the adjunctive treatment of sepsis: From immunosuppression to immunostimulation time for a paradigm change? *Am. J. Respir. Crit. Care Med.* 187, 1287–1293. doi: 10.1164/rccm.201301-0036cp

Lenz, K. M., and Nelson, L. H. (2018). Microglia and beyond: Innate immune cells as regulators of brain development and behavioral function. *Front. Immunol.* 9:698. doi: 10.3389/fimmu.2018.00698

Li, T. F., Li, K., Wang, C., Liu, X., Wen, Y., Xu, Y. H., et al. (2017). Harnessing the cross-talk between tumor cells and tumor-associated macrophages with a nano-drug for modulation of glioblastoma immune microenvironment. *J. Control. Release* 268, 128–146. doi: 10.1016/j.jconrel.2017.10.024

Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., et al. (2020). Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* 26, 842–844. doi: 10.1038/s41591-020-0901-9

Lin, E. Y., Nguyen, A. V., Russell, R. G., and Pollard, J. W. (2001). Colony-stimulating factor 1 promotes progression of mammary tumors to malignancy. *J. Exp. Med.* 193, 727–739. doi: 10.1084/jem.193.6.727

Linde, N., Casanova-Acebes, M., Sosa, M. S., Mortha, A., Rahman, A., Farias, E., et al. (2018). Macrophages orchestrate breast cancer early dissemination and metastasis. *Nat. Commun.* 9, 1–14.

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. *Cell* 153:1194.

Lopez−Yrigoyen, M., Cassetta, L., and Pollard, J. W. (2020). Macrophage targeting in cancer. *Ann. N. Y. Acad. Sci.* 2020 1–24. doi: 10.1111/nyas.14377

Loyher, P. L., Hamon, P., Laviron, M., Meghraoui-Kheddar, A., Goncalves, E., Deng, Z., et al. (2018). Macrophages of distinct origins contribute to tumor development in the lung. *J. Exp. Med.* 215, 1–18.

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., et al. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380, 2095–2128.

Ludwig, D. S. (2016). Lifespan weighed down by diet. *JAMA - J. Am. Med. Assoc.* 315, 2269–2270. doi: 10.1001/jama.2016.3829

Lukaszewicz, A. C., Grienay, M., Resche-Rigon, M., Pirracchio, R., Faivre, V., Boval, B., et al. (2009). Monocytic HLA-DR expression in intensive care patients: Interest for prognosis and secondary infection prediction. *Crit. Care Med.* 37, 2746–2752. doi: 10.1097/00003246-200910000-00011

Machiels, B., Dourcy, M., Xiao, X., Javaux, J., Mesnil, C., Sabatel, C., et al. (2017). A gammaherpesvirus provides protection against allergic asthma by inducing the replacement of resident alveolar macrophages with regulatory monocytes. *Nat. Immunol.* 18, 1310–1320. doi: 10.1038/ni.3857

Mackaness, G. B. (1969). The influence of immunologically committed lymphoid cells on macrophage activity in vivo. *J. Exp. Med.* 129, 973–992. doi: 10.1084/jem.129.5.973

Maier, B., Leader, A. M., Chen, S. T., Tung, N., Chang, C., LeBerichel, J., et al. (2020). A conserved dendritic-cell regulatory program limits antitumour immunity. *Nature* 580, 257–262. doi: 10.1038/s41586-020-2134-y

Mantovani, A., and Netea, M. G. (2020). Trained Innate Immunity, Epigenetics, and Covid-19. *N. Engl. J. Med.* 383, 1078–1080. doi: 10.1056/NEJMcibr2011679

Marrero, J. A., Kulik, L. M., Sirlin, C. B., Zhu, A. X., Finn, R. S., Abecassis, M. M., et al. (2018). Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. *Hepatology* 68, 723–750. doi: 10.1002/hep.29913

Mass, E., Ballesteros, I., Farlik, M., Halbritter, F., Günther, P., Crozet, L., et al. (2016). Specification of tissue-resident macrophages during organogenesis. *Science* 353:aaf4238. doi: 10.1126/science.aaf4238

Mehta, P., McAuley, D. F., Brown, M., Sanchez, E., Tattersall, R. S., and Manson, J. J. C. O. V. I. D. - (2020). 19: consider cytokine storm syndromes and immunosuppression. *Lancet* 395, 1033–1034. doi: 10.1016/s0140-6736(20)30628-0

Metchnikov, E. (1883). Untersuchungen ueber die mesodermalen Phagocyten einiger Wirbeltiere. *Biol. Cent.* 3, 560–565.

Mhatre, S. D., Tsai, C. A., Rubin, A. J., James, M. L., and Andreasson, K. I. (2015). Microglial Malfunction: The Third Rail in the Development of Alzheimer's Disease. *Trends Neurosci.* 38, 621–636. doi: 10.1016/j.tins.2015.08.006

Mills, C. D., Kincaid, K., Alt, J. M., Heilman, M. J., and Hill, A. M. M. - (2000). 1/M-2 Macrophages and the Th1/Th2 Paradigm. *J. Immunol.* 164, 6166–6173. doi: 10.4049/jimmunol.164.12.6166

Misharin, A. V., Morales-Nebreda, L., Reyfman, P. A., Cuda, C. M., Walter, J. M., McQuattie-Pimentel, A. C., et al. (2017). Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J. Exp. Med.* 214, 2387–2404.

Mitroulis, I., Ruppova, K., Wang, B., Chen, L. S., Grzybek, M., Grinenko, T., et al. (2018). Modulation of Myelopoiesis Progenitors Is an Integral Component of Trained Immunity. *Cell* 172, 147.e–161.e.

Molgora, M., Esaulova, E., Vermi, W., Hou, J., Chen, Y., Luo, J., et al. (2020). TREM2 Modulation Remodels the Tumor Myeloid Landscape Enhancing Anti-PD-1 Immunotherapy. *Cell* 182, 886.e–900.e.

Mosmann, T. R., Cherwinski, H., Bond, M. W., Giedlin, M. A., and Coffman, R. L. (1986). Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J. Immunol.* 136, 2348–2357.

Mrdjen, D., Pavlovic, A., Hartmann, F. J., Schreiner, B., Utz, S. G., Leung, B. P., et al. (2018). High-Dimensional Single-Cell Mapping of Central Nervous System

Immune Cells Reveals Distinct Myeloid Subsets in Health, Aging, and Disease. *Immunity* 48, 380.e–395.e.

Muenzer, J. T., Davis, C. G., Chang, K., Schmidt, R. E., Dunne, W. M., Coopersmith, C. M., et al. (2010). Characterization and modulation of the immunosuppressive phase of sepsis. *Infect. Immun.* 78, 1582–1592. doi: 10.1128/iai.01213-09

Müller, S., Kohanbash, G., Liu, S. J., Alvarado, B., Carrera, D., Bhaduri, A., et al. (2017). Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. *Genome Biol.* 18, 1–14.

Nagy, G., and van Vollenhoven, R. F. (2015). Sustained biologic-free and drug-free remission in rheumatoid arthritis, where are we now? *Arthritis Res. Ther.* 17, 1–7.

Netea, M. G., Domínguez-Andrés, J., Barreiro, L. B., Chavakis, T., Divangahi, M., Fuchs, E., et al. (2020a). Defining trained immunity and its role in health and disease. *Nat. Rev. Immunol.* 20, 375–388. doi: 10.1038/s41577-020-0285-6

Netea, M. G., Giamarellos-Bourboulis, E. J., Domínguez-Andrés, J., Curtis, N., van Crevel, R., van de Veerdonk, F. L., et al. (2020b). Trained Immunity: a Tool for Reducing Susceptibility to and the Severity of SARS-CoV-2 Infection. *Cell* 181, 969–977. doi: 10.1016/j.cell.2020.04.042

Netea, M. G., Joosten, L. A. B., Latz, E., Mills, K. H. G., Natoli, G., Stunnenberg, H. G., et al. (2016). Trained immunity: A program of innate immune memory in health and disease. *Science* 352, 427.

Nicolás-Ávila, J. A., Lechuga-Vieco, A. V., Esteban-Martínez, L., Sánchez-Díaz, M., Díaz-García, E., Santiago, D. J., et al. (2020). A Network of Macrophages Supports Mitochondrial Homeostasis in the Heart. *Cell* 183, 94–109. doi: 10.1016/j.cell.2020.08.031

Nikitas, G., Deschamps, C., Disson, O., Niault, T., Cossart, P., and Lecuit, M. (2011). Transcytosis of Listeria monocytogenes across the intestinal barrier upon specific targeting of goblet cell accessible E-cadherin. *J. Exp. Med.* 208, 2263–2277. doi: 10.1084/jem.20110560

O'Donnell, J. S., Teng, M. W. L., and Smyth, M. J. (2019). Cancer immunoediting and resistance to T cell-based immunotherapy. *Nat. Rev. Clin. Oncol.* 16, 151–167. doi: 10.1038/s41571-018-0142-8

Oishi, Y., and Manabe, I. (2016). Macrophages in age-related chronic inflammatory diseases. *NPJ Aging Mech. Dis.* 2, 1–8.

Okabe, Y., and Medzhitov, R. (2016). Tissue biology perspective on macrophages. *Nat. Immunol.* 17, 9–17. doi: 10.1038/ni.3320

Olshansky, S. J., Passaro, D. J., Hershow, R. C., Layden, J., Carnes, B. A., Brody, J., et al. (2005). potential decline in life expectancy in the United States in the 21st century. *N. Engl. J. Med.* 352, 1138–1145. doi: 10.1056/nejmsr043743

Otto, G. P., Sossdorf, M., Claus, R. A., Rödel, J., Menge, K., Reinhart, K., et al. (2011). The late phase of sepsis is characterized by an increased microbiological burden and death rate. *Crit. Care* 15:R183.

Paget, S. (1889). The distribution of secondary growths in cancer of the breast. *Lancet* 133, 571–573. doi: 10.1016/s0140-6736(00)49915-0

Pais, R., Barritt, A. S., Calmus, Y., Scatton, O., Runge, T., Lebray, P., et al. (2016). and liver transplantation: Current burden and expected challenges. *J. Hepatol.* 65, 1245–1257. doi: 10.1016/j.jhep.2016.07.033

Pamer, E. G. (2004). Immune responses to Listeria monocytogenes. *Nat. Rev. Immunol.* 4, 812–823. doi: 10.1038/nri1461

Parhizkar, S., Arzberger, T., Brendel, M., Kleinberger, G., Deussing, M., Focke, C., et al. (2019). Loss of TREM2 function increases amyloid seeding but reduces plaque-associated ApoE. *Nat. Neurosci.* 22, 191–204. doi: 10.1038/s41593-018-0296-9

Patil, N. K., Bohannon, J. K., and Sherwood, E. R. (2016). Immunotherapy: A promising approach to reverse sepsis-induced immunosuppression. *Pharmacol. Res.* 111, 688–702. doi: 10.1016/j.phrs.2016.07.019

Pearce, E. J., and MacDonald, A. S. (2002). The immunobiology of schistosomiasis. *Nat. Rev. Immunol.* 2, 499–511. doi: 10.1038/nri843

Pothineni, N. V. K., Subramany, S., Kuriakose, K., Shirazi, L. F., Romeo, F., Shah, P. K., et al. (2017). Infections, atherosclerosis, and coronary heart disease. *Eur. Heart J.* 38, 3195–3201.

Psaila, B., and Lyden, D. (2009). The metastatic niche: adapting the foreign soil. *Nat. Rev. Cancer*. 9, 285–293. doi: 10.1038/nrc2621

Puzianowska-Kuźnicka, M., Owczarz, M., Wieczorowska-Tobis, K., Nadrowski, P., Chudek, J., Slusarczyk, P., et al. (2016). Interleukin-6 and C-reactive protein, successful aging, and mortality: The PolSenior study. *Immun. Ageing* 13, 1–12.

Qian, B. Z., Li, J., Zhang, H., Kitamura, T., Zhang, J., Campion, L. R., et al. (2011). CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* 475, 222–225. doi: 10.1038/nature10138

Qian, F., Wang, X., Zhang, L., Chen, S., Piecychna, M., Allore, H., et al. (2012). Age-associated elevation in TLR5 leads to increased inflammatory responses in the elderly. *Aging Cell* 11, 104–110. doi: 10.1111/j.1474-9726.2011.00759.x

Quintin, J., Saeed, S., Martens, J. H. A., Giamarellos-Bourboulis, E. J., Ifrim, D. C., Logie, C., et al. (2012). Candida albicans infection affords protection against reinfection via functional reprogramming of monocytes. *Cell Host Microb.* 12, 223–232. doi: 10.1016/j.chom.2012.06.006

Ramachandran, P., Dobie, R., Wilson-Kanamori, J. R., Dora, E. F., Henderson, B. E. P., Luu, N. T., et al. (2019). Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* 575, 512–518. doi: 10.1038/s41586-019-1631-3

Ramanan, V. K., and Saykin, A. J. (2013). Pathways to neurodegeneration: Mechanistic insights from GWAS in Alzheimer's disease, Parkinson's disease, and related disorders. *Am. J. Neurodegener. Dis.* 2, 145–175.

Reilly, S. M., and Saltiel, A. R. (2017). Adapting to obesity with adipose tissue inflammation. *Nat. Rev. Endocrinol.* 13, 633–643. doi: 10.1038/nrendo.2017.90

Reinhart, K., Daniels, R., Kissoon, N., Machado, F. R., Schachter, R. D., and Finfer, S. (2017). Recognizing Sepsis as a Global Health Priority — A WHO Resolution. *N. Engl. J. Med.* 377, 414–417. doi: 10.1056/nejmp1707170

Remmerie, A., Martens, L., Pearce, E., Saeys, Y., and Scott, C. L. (2020). Osteopontin Expression Identifies a Subset of Recruited Macrophages Distinct from Kupffer Cells in the Fatty Liver. *Immunity* 53, 641.e–657.e. doi: 10.1016/j.immuni.2020.08.004

Réu, P., Khosravi, A., Bernard, S., Mold, J. E., Salehpour, M., Alkass, K., et al. (2017). The Lifespan and Turnover of Microglia in the Human Brain. *Cell Rep.* 20, 779–784. doi: 10.1016/j.celrep.2017.07.004

Roquilly, A., Jacqueline, C., Davieau, M., Mollé, A., Sadek, A., Fourgeux, C., et al. (2020). Alveolar macrophages are epigenetically altered after inflammation, leading to long-term lung immunoparalysis. *Nat. Immunol.* 21, 636–648. doi: 10.1038/s41590-020-0673-x

Sakai, M., Troutman, T. D., Seidman, J. S., Ouyang, Z., Spann, N. J., Abe, Y., et al. (2019). Liver-Derived Signals Sequentially Reprogram Myeloid Enhancers to Initiate and Maintain Kupffer Cell Identity. *Immunity* 51, 655.e–670.e.

Samuel, V. T., and Shulman, G. I. (2018). Nonalcoholic Fatty Liver Disease as a Nexus of Metabolic and Hepatic Diseases. *Cell Metab.* 27, 22–41. doi: 10.1016/j.cmet.2017.08.002

Schulte-Schrepping, J., Reusch, N., Paclik, D., Baßler, K., Schlickeiser, S., Zhang, B., et al. (2020). Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. *Cell* 182, 1419.e–1440.e.

Schulz, C., Perdiguero, E. G., Chorro, L., Szabo-Rogers, H., Cagnard, N., Kierdorf, K., et al. (2012). A lineage of myeloid cells independent of myb and hematopoietic stem cells. *Science* 335, 86–90. doi: 10.1126/science.1219179

Seidler, S., Zimmermann, H. W., Bartneck, M., Trautwein, C., and Tacke, F. (2010). Age-dependent alterations of monocyte subsets and monocyte-related chemokine pathways in healthy adults. *BMC Immunol.* 11:30. doi: 10.1186/1471-2172-11-30

Seidman, J. S., Troutman, T. D., Sakai, M., Gola, A., Spann, N. J., Bennett, H., et al. (2020). Niche-Specific Reprogramming of Epigenetic Landscapes Drives Myeloid Cell Diversity in Nonalcoholic Steatohepatitis. *Immunity* 52, 1057.e–1074.e.

Sharma, A., Seow, J. J. W., Dutertre, C.-A., Pai, R., Blériot, C., Mishra, A., et al. (2020). Onco-fetal Reprogramming of Endothelial Cells Drives Immunosuppressive Macrophages in Hepatocellular Carcinoma. *Cell* 183, 1–18.

Silvin, A., Chapuis, N., Dunsmore, G., Goubet, A.-G., Dubuisson, A., Derosa, L., et al. (2020). Elevated calprotectin and abnormal myeloid cell subsets discriminate severe from mild COVID-19. *Cell* 182, 1401.e–1418.e. doi: 10.1016/j.cell.2020.08.002

Singer, M., Deutschman, C. S., Seymour, C., Shankar-Hari, M., Annane, D., Bauer, M., et al. (2016). The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA - J. Am. Med. Assoc.* 315, 801–810.

Solana, R., Tarazona, R., Gayoso, I., Lesur, O., Dupuis, G., and Fulop, T. (2012). Innate immunosenescence: Effect of aging on cells and receptors of the innate immune system in humans. *Semin. Immunol.* 24, 331–341. doi: 10.1016/j.smim.2012.04.008

Song, Q., Hawkins, G. A., Wudel, L., Chou, P. C., Forbes, E., Pullikuth, A. K., et al. (2019). Dissecting intratumoral myeloid cell plasticity by single cell RNA-seq. *Cancer Med.* 8, 3072–3085. doi: 10.1002/cam4.2113

Steinman, J. B., Lum, F. M., Ho, P. P.-K., Kaminski, N., and Steinman, L. (2020). Reduced development of COVID-19 in children reveals molecular checkpoints gating pathogenesis illuminating potential therapeutics. *Proc. Natl. Acad. Sci. U. S. A.* 0, 1–7.

T'Jonck, W., Guilliams, M., and Bonnardel, J. (2018). Niche signals and transcription factors involved in tissue-resident macrophage development. *Cell. Immunol.* 330, 43–53. doi: 10.1016/j.cellimm.2018.02.005

Thevarajan, I., Nguyen, T. H. O., Koutsakos, M., Druce, J., Caly, L., van de Sandt, C. E., et al. (2020). Breadth of concomitant immune responses prior to patient recovery: a case report of non-severe COVID-19. *Nat. Med.* 26, 453–455. doi: 10.1038/s41591-020-0819-2

Thompson, R. C., Allam, A. H., Lombardi, G. P., Wann, L. S., Sutherland, M. L., Sutherland, J. D., et al. (2013). Atherosclerosis across 4000 years of human history: The Horus study of four ancient populations. *Lancet* 381, 1211–1222. doi: 10.1016/s0140-6736(13)60598-x

Ulland, T. K., and Colonna, M. (2018). TREM2 — a key player in microglial biology and Alzheimer disease. *Nat. Rev. Neurol.* 14, 667–675. doi: 10.1038/s41582-018-0072-1

Ulland, T. K., Song, W. M., Huang, S. C. C., Ulrich, J. D., Sergushichev, A., Beatty, W. L., et al. (2017). TREM2 Maintains Microglial Metabolic Fitness in Alzheimer's Disease. *Cell* 170, 649.e–663.e.

Walton, A. H., Muenzer, J. T., Rasche, D., Boomer, J. S., Sato, B., Brownstein, B. H., et al. (2014). Reactivation of multiple viruses in patients with sepsis. *PLoS One* 9:1–13. doi: 10.1371/journal.pone.0098819

Wan, C., Sun, Y., Tian, Y., Lu, L., Dai, X., Meng, J., et al. (2020). Irradiated tumor cell–derived microparticles mediate tumor eradication via cell killing and immune reprogramming. *Sci. Adv.* 6, 1–14.

Wang, T., Derhovanessian, A., De Cruz, S., Belperio, J. A., Deng, J. C., and Hoo, G. S. (2014). Subsequent infections in survivors of sepsis: Epidemiology and outcomes. *J. Int. Care Med.* 29, 87–95. doi: 10.1177/0885066612467162

Wang, Y., Cella, M., Mallinson, K., Ulrich, J. D., Young, K. L., Robinette, M. L., et al. (2015). TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell* 160, 1061–1071. doi: 10.1016/j.cell.2015.01.049

Wang, Y., Ulland, T. K., Ulrich, J. D., Song, W., Tzaferis, J. A., Hole, J. T., et al. (2016). TREM2-mediated early microglial response limits diffusion and toxicity of amyloid plaques. *J. Exp. Med.* 213, 667–675. doi: 10.1084/jem.20151948

Weisberg, S. P., Leibel, R. L., Anthony, F. Jr., Weisberg, S. P., Mccann, D., Desai, M., et al. (2003). . *J. Clin. Invest.* 112, 1796–1808.

Willemsen, L., and de Winther, M. P. J. (2020). Macrophage subsets in atherosclerosis as defined by single-cell technologies. *J. Pathol.* 250, 705–714. doi: 10.1002/path.5392

Williams, R., Aspinall, R., Bellis, M., Camps-Walsh, G., Cramp, M., Dhawan, A., et al. (2014). Addressing liver disease in the UK: A blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis. *Lancet* 384, 1953–1997. doi: 10.1016/s0140-6736(14)61838-9

Wills-karp, A. M., Luyimbazi, J., Xu, X., Schofield, B., Tamlyn, Y., Karp, C. L., et al. (2016). Interleukin-13: Central Mediator of Allergic Asthma. *Science* 282, 2258–2261.

Winkels, H., Ehinger, E., Vassallo, M., Buscher, K., Dinh, H. Q., Kobiyama, K., et al. (2018). Atlas of the immune cell repertoire in mouse atherosclerosis defined by single-cell RNA-sequencing and mass cytometry. *Circ. Res.* 122, 1675–1688. doi: 10.1161/circresaha.117.312513

World Health Organization. (2018). *World health statistics 2018: monitoring health for the SDGs, sustainable development goals*, Vol. 3. Geneva: World Health Organization.

Wynn, T. A. (2004). Fibrotic disease and the TH1/TH2 paradigm. *Nat. Rev. Immunol.* 4, 583–594. doi: 10.1038/nri1412

Wynn, T. A., and Ramalingam, T. R. (2012). Mechanisms of fibrosis: Therapeutic translation for fibrotic disease. *Nat. Med.* 18, 1028–1040. doi: 10.1038/nm.2807

Wynn, T. A., and Vannella, K. M. (2016). Macrophages in Tissue Repair, Regeneration, and Fibrosis. *Immunity* 44, 450–462. doi: 10.1016/j.immuni.2016.02.015

Xiong, X., Kuang, H., Ansari, S., Liu, T., Gong, J., Wang, S., et al. (2019). Landscape of Intercellular Crosstalk in Healthy and NASH Liver Revealed by Single-Cell Secretome Gene Analysis. *Mol. Cell* 75, 644.e–660.e.

Xu, H., Barnes, G. T., Yang, Q., Tan, G., Yang, D., Chou, C. J., et al. (2003). Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* 112, 1821–1830. doi: 10.1172/jci200319451

Xue, J., Schmidt, S. V., Sander, J., Draffehn, A., Krebs, W., Quester, I., et al. (2014). Transcriptome-Based Network Analysis Reveals a Spectrum Model of Human Macrophage Activation. *Immunity* 40, 274–288. doi: 10.1016/j.immuni.2014.01.006

Xu-Vanpala, S., Deerhake, M. E., Wheaton, J. D., Parker, M. E., Juvvadi, P. R., MacIver, N., et al. (2020). Functional heterogeneity of alveolar macrophage population based on expression of CXCL2. *Sci. Immunol.* 5:eaba7350. doi: 10.1126/sciimmunol.aba7350

Yang, J. D., Hainaut, P., Gores, G. J., Amadou, A., Plymoth, A., and Roberts, L. R. A. (2019). global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat. Rev. Gastroenterol. Hepatol.* 16, 589–604. doi: 10.1038/s41575-019-0186-y

Yao, Y., Jeyanathan, M., Haddadi, S., Barra, N. G., Vaseghi-Shanjani, M., Damjanovic, D., et al. (2018). Induction of Autonomous Memory Alveolar Macrophages Requires T Cell Help and Is Critical to Trained Immunity. *Cell* 175, 1634.e–1650.e.

Yona, S., Kim, K. W., Wolf, Y., Mildner, A., Varol, D., Breker, M., et al. (2013). Fate Mapping Reveals Origins and Dynamics of Monocytes and Tissue Macrophages under Homeostasis. *Immunity* 38, 79–91. doi: 10.1016/j.immuni.2012.12.001

Zhao, X., Qu, J., Sun, Y., Wang, J., Liu, X., Wang, F., et al. (2017). Prognostic significance of tumor-associated macrophages in breast cancer: A meta-analysis of the literature. *Oncotarget* 8, 30576–30586. doi: 10.18632/oncotarget.15736

Zhao, Y., Wu, X., Li, X., Jiang, L. L., Gui, X., Liu, Y., et al. (2018). TREM2 Is a Receptor for β-Amyloid that Mediates Microglial Function. *Neuron* 97, 1023.e–1031.e.

Zheng, X., Weigert, A., Reu, S., Guenther, S., Mansouri, S., Bassaly, B., et al. (2020). Spatial Density and Distribution of Tumor-Associated Macrophages Predict Survival in Non-Small-Cell Lung Carcinoma. *Cancer Res.* canres 2020:0069. doi: 10.1158/0008-5472.can-20-0069

Zhu, Y., Herndon, J. M., Sojka, D. K., Kim, K. W., Knolhoff, B. L., Zuo, C., et al. (2017). Tissue-Resident Macrophages in Pancreatic Ductal Adenocarcinoma Originate from Embryonic Hematopoiesis and Promote Tumor Progression. *Immunity* 47, 323.e–338.e.

Zilionis, R., Engblom, C., Pfirschke, C., Savova, V., Zemmour, D., Saatcioglu, H. D., et al. (2019). Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *Immunity* 50, 1317.e–1334.e.

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.

frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 22 December 2020
doi: 10.3389/fcell.2020.614832

Computation of Single-Cell Metabolite Distributions Using Mixture Models

Mona K. Tonn[1], *Philipp Thomas*[1], *Mauricio Barahona*[1] *and Diego A. Oyarzún*[2,3]*

[1] Department of Mathematics, Imperial College London, London, United Kingdom, [2] School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom, [3] School of Informatics, University of Edinburgh, Edinburgh, United Kingdom

Metabolic heterogeneity is widely recognized as the next challenge in our understanding of non-genetic variation. A growing body of evidence suggests that metabolic heterogeneity may result from the inherent stochasticity of intracellular events. However, metabolism has been traditionally viewed as a purely deterministic process, on the basis that highly abundant metabolites tend to filter out stochastic phenomena. Here we bridge this gap with a general method for prediction of metabolite distributions across single cells. By exploiting the separation of time scales between enzyme expression and enzyme kinetics, our method produces estimates for metabolite distributions without the lengthy stochastic simulations that would be typically required for large metabolic models. The metabolite distributions take the form of Gaussian mixture models that are directly computable from single-cell expression data and standard deterministic models for metabolic pathways. The proposed mixture models provide a systematic method to predict the impact of biochemical parameters on metabolite distributions. Our method lays the groundwork for identifying the molecular processes that shape metabolic heterogeneity and its functional implications in disease.

Keywords: metabolic variability, stochastic gene expression, metabolic modeling, single-cell modeling, mixture model analysis

OPEN ACCESS

Edited by:
Ankur Sharma,
Genome Institute of Singapore, Singapore

Reviewed by:
Yogesh Goyal,
University of Pennsylvania, United States
Alejandro Vignoni,
Max Planck Institute of Molecular Cell Biology and Genetics, Germany

***Correspondence:**
Diego A. Oyarzún
D.Oyarzun@ed.ac.uk

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Cell and Developmental Biology

Received: *07 October 2020*
Accepted: *26 November 2020*
Published: *22 December 2020*

Citation:
Tonn MK, Thomas P, Barahona M and Oyarzún DA (2020) Computation of Single-Cell Metabolite Distributions Using Mixture Models. Front. Cell Dev. Biol. 8:614832. doi: 10.3389/fcell.2020.614832

1. INTRODUCTION

Non-genetic heterogeneity is a hallmark of cell physiology. Isogenic cells can display markedly different phenotypes as a result of the stochasticity of intracellular processes and fluctuations in environmental conditions. Gene expression variability, in particular, has received substantial attention thanks to robust experimental techniques for measuring transcripts and proteins at a single-cell resolution (Golding et al., 2005; Taniguchi et al., 2010). This progress has gone hand-in-hand with a large body of theoretical work on stochastic models to identify the molecular processes that affect expression heterogeneity (Swain et al., 2002; Raj and van Oudenaarden, 2008; Thomas et al., 2014; Dattani and Barahona, 2017; Tonn et al., 2019).

In contrast to gene expression, our understanding of stochastic phenomena in metabolism is still in its infancy. Traditionally, cellular metabolism has been regarded as a deterministic process on the basis that metabolites appear in large numbers that filter out stochastic phenomena (Heinemann and Zenobi, 2011). But this view is changing rapidly thanks to a growing number of single-cell measurements of metabolites and co-factors (Bennett et al., 2009; Imamura et al., 2009; Lemke and Schultz, 2011; Paige et al., 2012; Ibáñez et al., 2013; Yaginuma et al., 2014; Esaki and Masujima, 2015; Xiao et al., 2016; Mannan et al., 2017) that suggest that cell-to-cell metabolite variation is

much more pervasive than previously thought. The functional implications of this heterogeneity are largely unknown but likely to be substantial given the roles of metabolism in many cellular processes, including growth (Weisse et al., 2015), gene regulation (Lempp et al., 2019), epigenetic control (Loftus and Finlay, 2016), and immunity (Reid et al., 2017). For example, metabolic heterogeneity has been linked to bacterial persistence (Radzikowski et al., 2017; Shan et al., 2017), a dormant phenotype characterized by a low metabolic activity, as well as antibiotic resistance (Deris et al., 2013) and other functional effects (Vilhena et al., 2018). In biotechnology applications, metabolic heterogeneity is widely recognized as a limiting factor on metabolite production with genetically engineered microbes (Binder et al., 2017; Schmitz et al., 2017; Liu et al., 2018).

A key challenge for quantifying metabolic variability is the difficulty in measuring cellular metabolites at a single-cell resolution (Amantonico et al., 2010; Takhaveev and Heinemann, 2018; Wehrens et al., 2018). As a result, most studies use other phenotypes as a proxy for metabolic variation, e.g., enzyme expression levels (Kotte et al., 2014; van Heerden et al., 2014), metabolic fluxes (Schreiber et al., 2016), or growth rate (Kiviet et al., 2014; Şimşek and Kim, 2018). From a computational viewpoint, the key challenge is that metabolic processes operate on two timescales: a slow timescale for expression of metabolic enzymes, and a fast timescale for enzyme catalysis. Such multiscale structure results in stiff models that are infeasible to solve with standard algorithms for stochastic simulation (Gillespie, 2007). Other strategies to accelerate stochastic simulations, such as τ-leaping (Rathinam et al., 2003), also fail to produce accurate simulation results due to the disparity in molecule numbers between enzymes and metabolites (Tonn, 2020). These challenges have motivated a number of methods to optimize stochastic simulations of metabolism (Puchałka and Kierzek, 2004; Cao et al., 2005; Labhsetwar et al., 2013; Lugagne et al., 2013; Murabito et al., 2014). Most of these methods exploit the timescale separation to accelerate simulations at the expense of some approximation error. This progress has been accompanied by a number of theoretical results on the links between molecular processes and the shape of metabolite distributions (Levine and Hwa, 2007; Oyarzún et al., 2015; Gupta et al., 2017b; Tonn et al., 2019). Yet to date there are no general methods for computing metabolite distributions that can handle inherent features of metabolic pathways such as feedback regulation, complex stoichiometries, and the high number of molecular species involved.

In this paper we present a widely applicable method for approximating single-cell metabolite distributions. Our method is founded on the timescale separation between enzyme expression and enzyme catalysis, which we employ to approximate the stationary solution of the chemical master equation. The approximate solution takes the form of mixture distributions with: (i) mixture weights that can be computed from models for gene expression or single-cell expression data, and (ii) mixture components that are directly computable from deterministic pathway models. The resulting mixture model can be employed to explore the impact of biochemical parameters on metabolite variability. We illustrate the power of the method in two exemplar systems that are core building blocks of large metabolic networks. Our theory provides a quantitative basis to draw testable hypotheses on the sources of metabolite heterogeneity, which together with the ongoing efforts in single-cell metabolite measurements, will help to re-evaluate the role of metabolism as an active source of phenotypic variation.

2. GENERAL METHOD FOR COMPUTING METABOLITE DISTRIBUTIONS

We consider metabolic pathways composed of enzymatic reactions interconnected by sharing of metabolites as substrates or products. In general, we consider models with M metabolites P_i with $i \in \{1, 2, \ldots, M\}$ and N catalytic enzymes E_j with $j \in \{1, 2, \ldots, N\}$. A typical enzymatic reaction has the form

$$P_i + E_j \underset{k_{\mathrm{b},j}}{\overset{k_{\mathrm{f},j}}{\rightleftharpoons}} C_j \underset{k_{\mathrm{rev},j}}{\overset{k_{\mathrm{cat},j}}{\rightleftharpoons}} P_k + E_j, \tag{1}$$

where P_i and P_k are metabolites, and E_j and C_j are the free and substrate-bound forms of the enzyme. The parameters $(k_{\mathrm{f},j}, k_{\mathrm{b},j})$ and $(k_{\mathrm{cat},j}, k_{\mathrm{rev},j})$ are positive rate constants specific to the enzyme. In contrast to traditional metabolic models, where the number of enzyme molecules is assumed constant, here we explicitly model enzyme expression and enzyme catalysis as stochastic processes. Our models also account for dilution of molecular species by cell growth and consumption of the metabolite products by downstream processes.

Though in principle one can readily write a Chemical Master Equation (CME) for the marginal distribution $\mathbf{P}(P_1, P_2, \ldots P_M)$ given the pathway stoichiometry, analytical solutions of the CME are tractable only in few special cases. To overcome this challenge, we propose a method for approximating metabolite distributions that can be applied in a wide range of metabolic models. We first note that using the Law of Total Probability, the marginal distribution $\mathbf{P}(P_1, P_2, \ldots, P_M)$ can be generally written as:

$$\mathbf{P}(P) = \sum_E \mathbf{P}(E) \times \mathbf{P}(P|E), \tag{2}$$

where $P = (P_1, P_2, \ldots P_M)$ and $E = (E_1, E_2, \ldots, E_N)$ are the vectors of metabolite and enzyme abundances, respectively. The equation in (2) describes the metabolite distribution in terms of fluctuations in gene expression, comprised in the distribution $\mathbf{P}(E)$, and fluctuations in reaction catalysis, described by conditional distribution $\mathbf{P}(P|E)$.

A key observation is that Equation (2) corresponds to a mixture model with weights $\mathbf{P}(E)$ and mixture components $\mathbf{P}(P|E)$. To compute the mixture weights and components, we make use of the timescale separation between gene expression and metabolism. Gene expression operates on a much slower timescale than catalysis (Cao et al., 2005; Levine and Hwa, 2007; Kuntz et al., 2013), with protein half-lives typically comparable to cell doubling times and catalysis operating in the millisecond to second range. Therefore, in the fast timescale of catalysis we

can write a conservation law for the total amount of each enzyme (free and bound):

$$E_{t,j} = E_j + C_j, \tag{3}$$

where $E_{t,j}$ is the total number of enzymes E_j. Note that since our models integrate enzyme kinetics with enzyme expression, the variables $E_{t,j}$ follow their own, independent stochastic dynamics. It is important to note that in our approach, the conservation relation in (3) holds only in the fast timescale of catalysis. This contrasts with classic deterministic models for metabolic reactions, which typically focus on the fast catalytic timescale and assume enzymes as constant model parameters (Cornish-Bowden, 2004).

As a result of the separation of timescales, the weights and components of the mixture in (2) can be computed separately. Specifically, the mixture weights $\mathbf{P}(E)$ can be obtained as solutions of a stochastic model for enzyme expression (Raj and van Oudenaarden, 2008), or taken from absolute single-cell measurements of enzyme expression. Such absolute measurements can be obtained from single-molecule technologies (Okumus et al., 2016), carefully calibrating fluorescence data (Rosenfeld et al., 2006; Bakker and Swain, 2019) or normalization (Taniguchi et al., 2010). The mixture components $\mathbf{P}(P|E)$, on the other hand, can be estimated with suitable approximation techniques. For simplicity, here we choose to employ the Linear Noise Approximation (LNA), which provides a Gaussian estimate of the stationary distribution of a stochastic chemical system (van Kampen, 1992; Elf and Ehrenberg, 2003). The use of the LNA is justified on the basis that metabolites tend to appear in large numbers per cell, a key condition for the LNA to produce accurate results. However, more accurate methods to compute $\mathbf{P}(P|E)$ can be used if required (Andreychenko et al., 2017; Gupta et al., 2017a). In **Figure 1**, we illustrate a schematic of the proposed method.

We thus propose the following procedure for computing single-cell metabolite distributions:

1. Starting from the mixture model in Equation (2), compute the enzyme distribution $\mathbf{P}(E)$ from a stochastic model for gene expression, either analytically (if possible) or numerically with Gillespie's algorithm.
2. To approximate the mixture components $\mathbf{P}(P|E)$ with the LNA, compute the steady state solution $\bar{P}$ of the deterministic rate equation for each enzyme state E:

$$\mathbf{S}v(\bar{P}, E) = 0, \tag{4}$$

   where $\mathbf{S}$ is the stoichiometric matrix and $v(\cdot)$ is the vector of deterministic reaction rates; for ease of notation we have assumed a unit cell volume, and hence the deterministic rates are equal to the propensities of the stochastic model. Note that due to the timescale separation, Equation (4) must be solved assuming constant enzymes E, and its solution depends on the enzyme abundance, i.e., $\bar{P} = \bar{P}(E)$.
3. For each enzyme state E, compute the solution to the Lyapunov equation (Elf and Ehrenberg, 2003):

$$A\Sigma + \Sigma A^T + BB^T = 0, \tag{5}$$

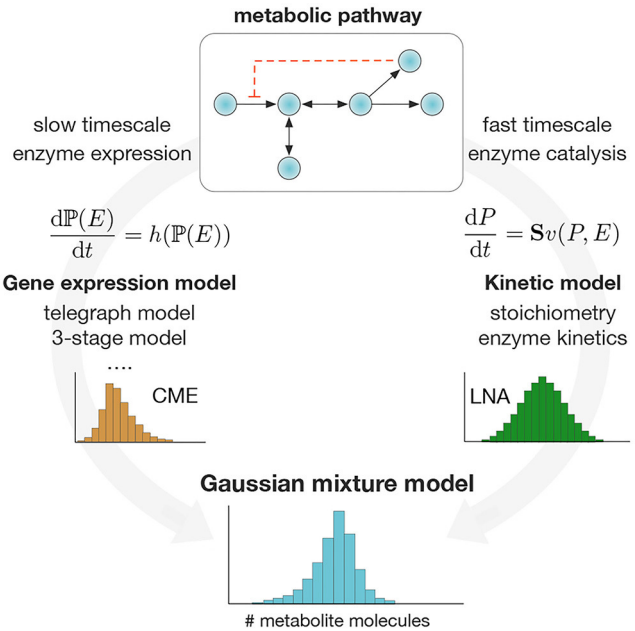

FIGURE 1 | Computation of single-cell metabolite distributions with Gaussian mixture models. We exploit the separation of timescales to compute the weights and components of the mixture model in Equation (2). Mixture weights are computed as stationary solutions to the Chemical Master Equation (CME) for a chosen model for stochastic enzyme expression. The mixture components are computed via the Linear Noise Approximation (Elf and Ehrenberg, 2003) (LNA) applied to the pathway ODE model. The method produces a Gaussian mixture model for metabolite distributions that can be applied in a wide range of metabolic pathways.

   where A is the Jacobian of (4) evaluated at the steady state and $BB^T = \mathbf{S}\mathrm{diag}\{v\}\mathbf{S}^T$. Note that, as in (4), the solution of the Lyapunov equation depends on the enzyme state, i.e., $\Sigma = \Sigma(E)$.
4. Following the LNA, approximate the mixture components $\mathbf{P}(P|E)$ as a multivariate Gaussian distribution with mean $\bar{P}$ and covariance matrix Σ.
5. Combine the weights $\mathbf{P}(E)$ and Gaussian components $\mathbf{P}(P|E)$ through the mixture model in (2).

In the next sections we illustrate the effectiveness of our method in two exemplar systems.

3. REVERSIBLE MICHAELIS-MENTEN REACTION

We first consider a stochastic model that integrates a reversible Michaelis-Menten reaction with a standard model for enzyme expression. As shown in **Figure 2A**, the Michaelis-Menten mechanism includes reversible binding of four species: a metabolic substrate S, a free enzyme E, a substrate-enzyme complex C and a metabolic product P. To model enzyme expression, we use the well-known two-stage scheme for transcription and translation (Thattai and van Oudenaarden, 2001; Shahrezaei and Swain, 2008) (**Figure 2A**). The complete set

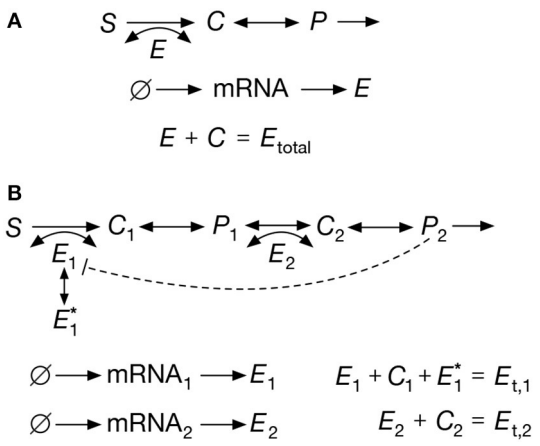

FIGURE 2 | Exemplar metabolic systems. **(A)** Reversible Michaelis-Menten reaction; the full set of reactions are shown in Equations (6)–(9). The model accounts for reversible catalysis of a substrate S into a product P. **(B)** Two-step pathway with noncompetitive end-product inhibition; the reactions are shown in Equations (18)–(25). The product (P_2) sequesters enzyme E_1 into an inactive form E_1^*, thereby reducing the rate of the first reaction. In both examples we assume a constant substrate S and linear dilution of all chemical species. Enzymes are assumed to follow the two-stage model for gene expression (Shahrezaei and Swain, 2008), which includes species for the enzymatic mRNA and protein.

of reactions is:

$$S + E \underset{k_b}{\overset{k_f}{\rightleftharpoons}} C \underset{k_{rev}}{\overset{k_{cat}}{\rightleftharpoons}} P + E, \tag{6}$$

$$\emptyset \xrightarrow{k_{tx}} \text{mRNA} \xrightarrow{k_{tl}} \text{mRNA} + E, \tag{7}$$

$$P \xrightarrow{k_c} \emptyset, \tag{8}$$

$$\text{mRNA} \xrightarrow{k_{deg}} \emptyset, \quad E \xrightarrow{\delta} \emptyset, \quad C \xrightarrow{\delta} \emptyset. \tag{9}$$

The reactions in (6) correspond to a reversible Michaelis-Menten reaction as in (1), while reactions in (7) are the two-stage model for gene expression. We include four additional first-order reactions (8) and (9) to model consumption of the metabolite product with rate constant k_c, mRNA degradation with rate constant k_{deg}, and dilution of all model species with rate constant δ. In what follows we assume that the substrate S remains strictly constant, for example to model cases in which the substrate represents an extracellular carbon source that evolves in much slower timescale than cell doubling times.

Since on the fast timescale of the catalytic reaction, the total number of enzymes can be assumed in quasi-stationary state (Cornish-Bowden, 2004; Tonn et al., 2019), we have that

$$E_{total} = E + C, \tag{10}$$

and therefore the general mixture model in (2) can be written as:

$$\mathbf{P}(P) = \sum_{E_{total}=0}^{\infty} \underbrace{\mathbf{P}(E_{total})}_{\substack{\text{enzyme} \\ \text{distribution}}} \times \underbrace{\mathbf{P}(P|E_{total})}_{\substack{\text{Gaussian} \\ \text{from LNA}}}. \tag{11}$$

The mixture weights $\mathbf{P}(E_{total})$ can be computed from the stochastic model for gene expression in (7). Under the standard assumption that mRNAs are degraded much faster than proteins (Raj and van Oudenaarden, 2008), the stationary solution of the two-stage model can be approximated by a negative binomial distribution (Shahrezaei and Swain, 2008):

$$\mathbf{P}(E_{total}) = \frac{\Gamma(a + E_{total})}{\Gamma(E_{total} + 1)\Gamma(a)} \left(\frac{b}{1+b}\right)^{E_{total}} \frac{1}{(1+b)^a}, \tag{12}$$

where Γ is the Gamma function and the parameters are defined as the burst frequency $a = k_{tx}/\delta$ and burst size $b = k_{tl}/k_{deg}$.

To compute the mixture components $\mathbf{P}(P|E_{total})$ with the LNA, we write the full system of deterministic rate equations [see (35) in section 6] for the three species E, C, and P. Note that in this case, we can further reduce the rate equations by (i) using the conservation law in (10), and (ii) assuming that the binding and unbinding reactions between S and E reach equilibrium faster than the product P, a condition that generally holds in metabolic reactions. After algebraic manipulations, the reduced ODE can be written as:

$$\frac{dP}{dt} = f(P, E_{total}) - g(P, E_{total}) - k_c P \tag{13}$$

where

$$\begin{aligned} f(P, E_{total}) &= E_{total} \frac{k_{cat} S/K_{mS}}{1 + S/K_{mS} + P/K_{mP}}, \\ g(P, E_{total}) &= E_{total} \frac{k_b P/K_{mP}}{1 + S/K_{mS} + P/K_{mP}} \end{aligned} \tag{14}$$

and the parameters are $K_{mS} = (k_b + k_{cat})/k_f$ and $K_{mP} = (k_b + k_{cat})/k_{rev}$.

The mean of each mixture component is simply given by the steady state solution of (13), which we denote as $\bar{P}(E_{total})$. For a given enzyme abundance E_{total}, the variance $\Sigma(E_{total})$ of each Gaussian component is given by the solution to the Lyapunov equation in (5):

$$\Sigma(E_{total}) = \frac{1}{2} \frac{f(\bar{P}(E_{total})) + g(\bar{P}(E_{total})) + k_c \bar{P}(E_{total})}{k_c + g'(\bar{P}(E_{total})) - f'(\bar{P}(E_{total}))}, \tag{15}$$

where f' and g' are first-order derivatives. Combining the negative binomial in (12) with the Gaussian components, we can rewrite Equation (11) to get a Gaussian mixture model for the metabolite:

$$\mathbf{P}(P) = K \sum_{x=0}^{\infty} \frac{1}{\Sigma(x)} \frac{\Gamma(a+x)}{\Gamma(x+1)} \left(\frac{b}{1+b}\right)^x e^{-\frac{1}{2}\left(\frac{P - \bar{P}(x)}{\Sigma(x)}\right)^2}, \tag{16}$$

where both $\bar{P}(x)$ and $\Sigma(x)$ must be computed for each value of $x = E_{total}$ in the summation. The normalization constant in (16) is

$$K = \frac{1}{\sqrt{2\pi}\Gamma(a)(1+b)^a}. \tag{17}$$

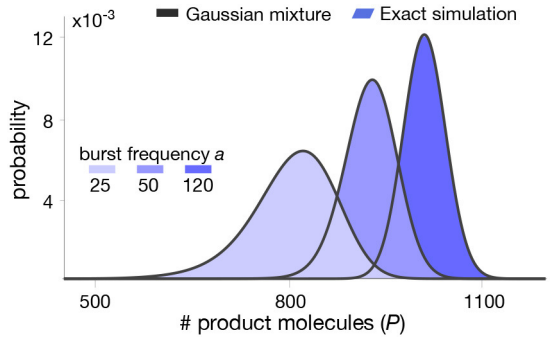

FIGURE 3 | Stationary product distribution of a Michaelis-Menten reaction. The proposed mixture model in (16) provides an excellent approximation for the metabolite distribution obtained with Gillespie's algorithm (Gillespie, 2007). Distributions were computed for varying values of the bursting parameter a. Note that the resulting distributions are almost identical to those predicted in our earlier work using a Poisson mixture (Tonn et al., 2019), since we have deliberately chosen parameters to produce similar distribution in both cases. All parameter values can be found in **Table 1**.

In **Figure 3**, we plot the mixture model (16) for realistic parameter values and compare this approximation with distributions computed from long runs of Gillespie simulations of the whole set of reactions (6)–(9). The results indicate that the mixture model provides an excellent approximation of the metabolite distribution. In the next section we test our methodology in a more complex pathway with feedback regulation.

4. PATHWAY WITH END-PRODUCT INHIBITION

A common regulatory motif in metabolism is end-product inhibition, in which a pathway enzyme can bind to its own substrate as well as the pathway product (see **Figure 2B**). The product thus sequesters enzyme molecules, which reduces the number of free enzymes available for catalysis and slows done the reaction rate. To examine the accuracy of our method in this setting, we study a fully stochastic model for a two-step pathway with noncompetitive end-product inhibition:

$$S + E_1 \underset{k_{b,1}}{\overset{k_{f,1}}{\rightleftharpoons}} C_1 \underset{k_{rev,1}}{\overset{k_{cat,1}}{\rightleftharpoons}} P_1 + E_1 \tag{18}$$

$$P_1 + E_2 \underset{k_{b,2}}{\overset{k_{f,2}}{\rightleftharpoons}} C_2 \underset{k_{rev,2}}{\overset{k_{cat,2}}{\rightleftharpoons}} P_2 + E_2 \tag{19}$$

$$hP_2 + E_1 \underset{k_{rsq}}{\overset{k_{sq}}{\rightleftharpoons}} E^* \tag{20}$$

$$\emptyset \xrightarrow{k_{tx,1}} \text{mRNA}_1 \xrightarrow{k_{tl,1}} \text{mRNA}_1 + E_1 \tag{21}$$

$$\emptyset \xrightarrow{k_{tx,2}} \text{mRNA}_2 \xrightarrow{k_{tl,2}} \text{mRNA}_2 + E_2 \tag{22}$$

$$P_1 \xrightarrow{k_{c,1}} \emptyset, \quad P_2 \xrightarrow{k_{c,2}} \emptyset \tag{23}$$

TABLE 1 | Parameter values for simulations in **Figure 3**.

Figure 3			
δ	0.00025 s^{-1}	k_b	1,000 s^{-1}
a	{25, 50, 120}	k_{cat}	3.6 s^{-1}
b	1	k_{rev}	0.01 s^{-1}
S	3,000 molecules	k_c	0.02 s^{-1}
k_f	$1 \times S$ s^{-1}		

$$\text{mRNA}_1 \xrightarrow{k_{deg,1}} \emptyset, \quad \text{mRNA}_2 \xrightarrow{k_{deg,2}} \emptyset, \tag{24}$$

$$E^* \xrightarrow{\delta} \emptyset, \quad E_1 \xrightarrow{\delta} \emptyset, \quad E_2 \xrightarrow{\delta} \emptyset, \quad C_1 \xrightarrow{\delta} \emptyset, \quad C_2 \xrightarrow{\delta} \emptyset \tag{25}$$

The two reactions in (18) and (19) are reversible Michaelis-Menten kinetics, sharing the intermediate metabolite P_1 as a product and substrate, respectively. The end-product inhibition in (20) consists of reversible binding between h molecules of P_2 and the first enzyme E_1 into a catalytically-inactive complex E^*. The remaining model reactions in (21)–(25) are analogous to the previous example in section 3: reactions in (21) and (22) describe the two-stage model for expression of both enzymes, and with reactions (23)–(25) we model first-order mRNA degradation, product consumption, and dilution by cell growth. For simplicity we also assume that both enzymes are independently expressed, but in general our method can also account for cases in which enzymes are co-expressed or co-regulated (Chubukov et al., 2014). The resulting model has two distinct pools of enzymes, which remain constant over the timescale of catalysis:

$$\begin{aligned} E_{t,1} &= E_1 + E^* + C_1, \\ E_{t,2} &= E_2 + C_2, \end{aligned} \tag{26}$$

and therefore the mixture model in (2) becomes

$$\mathbf{P}(P_1, P_2) = \sum_{E_{t,1}, E_{t,2}} \underbrace{\mathbf{P}(E_{t,1}, E_{t,2})}_{\substack{\text{enzyme} \\ \text{distribution}}} \underbrace{\mathbf{P}(P_1, P_2 | E_{t,1}, E_{t,2})}_{\substack{\text{Gaussian} \\ \text{from LNA}}}, \tag{27}$$

where the summation goes through all $(E_{t,1}, E_{t,2})$ pairs. Since both enzymes are expressed independently, the enzyme distribution is the product of two negative binomials $\mathbf{P}(E_{t,1}, E_{t,2}) = \mathbf{P}(E_{t,1}) \times \mathbf{P}(E_{t,2})$, each one analogous to the distribution in (12).

To compute the mixture components with the LNA, we use the rate equations for the reactions in (18)–(23); the full set of ODEs is listed in Equation (36) in the Methods. As in the first example, by employing the conservation laws in (26) and assuming rapid equilibrium of the complexes C_1 and C_2, the deterministic model can be further simplified to a 2-dimensional ODE:

$$\begin{aligned} \frac{dP_1}{dt} &= f(P_1, P_2) - g(P_1, P_2) - k_{c,1} P_1, \\ \frac{dP_2}{dt} &= g(P_1, P_2) - k_{c,2} P_2, \end{aligned} \tag{28}$$

where for ease of notation we have omitted the dependency on $E_{t,1}$ and $E_{t,2}$. The nonlinear functions in (28) are

$$f(P_1,P_2) = E_{t,1}\frac{\kappa_S S - \kappa_1 P_1}{1+\theta P_2^h + S/K_{m,S} + P_1/K_{m,1}},$$
$$g(P_1,P_2) = E_{t,2}\frac{\kappa_2 P_1 - \kappa_3 P_2}{1+P_1/K_{m,2} + P_2/K_{m,3}}, \quad (29)$$

where $\theta = k_{sq}/k_{rsq}$ is the product-enzyme binding constant and the remaining parameters are defined as $\kappa_S = k_{cat,1}k_{f,1}/(k_{b,1}+k_{cat,1})$, $\kappa_1 = k_{b,1}k_{rev,1}/(k_{b,1}+k_{cat,1})$, $\kappa_2 = k_{cat,2}k_{f,2}/(k_{b,2}+k_{cat,2})$, $\kappa_3 = k_{b,2}k_{rev,2}/(k_{b,2}+k_{cat,2})$, $K_{m,S} = k_{cat,1}/\kappa_S$, $K_{m,1} = k_{b,1}/\kappa_1$, $K_{m,2} = k_{cat,2}/\kappa_2$, and $K_{m,3} = k_{b,2}/\kappa_3$.

As in the previous example, the ODEs in (28) correspond to the full model (36) rewritten in terms of both metabolites assuming that the enzyme-substrate reactions reach equilibrium in a faster timescale than catalysis. This reduced model can be readily employed to obtain approximations for the mixture components with the LNA. If we denote as $\bar{P} = \bar{P}(E_{t,1},E_{t,2})$ the steady state solution of (28), we can write the Lyapunov equation as $A\Sigma + \Sigma A^T + BB^T = 0$ with A and BB^T given by

$$A = \begin{bmatrix} \frac{d}{dP_1}(f-g) - k_{c,1} & \frac{d}{dP_2}(f-g) \\ \frac{dg}{dP_1} & \frac{dg}{dP_2} - k_{c,2} \end{bmatrix}, \quad (30)$$

$$BB^T = \begin{bmatrix} f+g+k_{c,1}P_1 & -g \\ -g & g+k_{c,2}P_2 \end{bmatrix}, \quad (31)$$

where $f(\cdot)$, $g(\cdot)$, and their derivatives are evaluated at the steady state solution $\bar{P}(E_{t,1},E_{t,2})$. The Gaussian components of the mixture model are then

$$\mathbf{P}(P_1,P_2|E_{t,1},E_{t,2}) = \frac{1}{2\pi|\Sigma(E_{t,1},E_{t,2})|}e^{-\frac{1}{2}(P-\bar{P}(E_{t,1},E_{t,2}))^T\Sigma^{-1}(P-\bar{P}(E_{t,1},E_{t,2}))}, \quad (32)$$

where $P = (P_1,P_2)^T$ and $|\cdot|$ is the matrix determinant. After combining the joint distribution of enzymes and the components into Equation (27), we get a Gaussian mixture model for the joint marginal distribution of both metabolites:

$$\mathbf{P}(P_1,P_2) = K\sum_{x,y=0}^{\infty}\frac{\Gamma(a_1+x)\Gamma(a_2+y)}{\Gamma(x+1)\Gamma(y+1)}\left(\frac{b_1}{1+b_1}\right)^x\left(\frac{b_2}{1+b_2}\right)^y \times \frac{1}{|\Sigma(x,y)|}e^{-\frac{1}{2}(P-\bar{P}(x,y))^T\Sigma(x,y)^{-1}(P-\bar{P}(x,y))}, \quad (33)$$

where $\bar{P}(x,y)$ and $\Sigma(x,y)$ need to computed numerically for each pair $(x,y) = (E_{t,1},E_{t,2})$ in the summation. The burst frequencies $a_i = k_{tx,i}/\delta$ and burst sizes $b_i = k_{tl,i}/k_{deg,i}$ are specific to each enzyme, and the normalization constant is given by

$$K = \frac{1}{2\pi\Gamma(a_1)\Gamma(a_2)(1+b_1)^{a_1}(1+b_2)^{a_2}}. \quad (34)$$

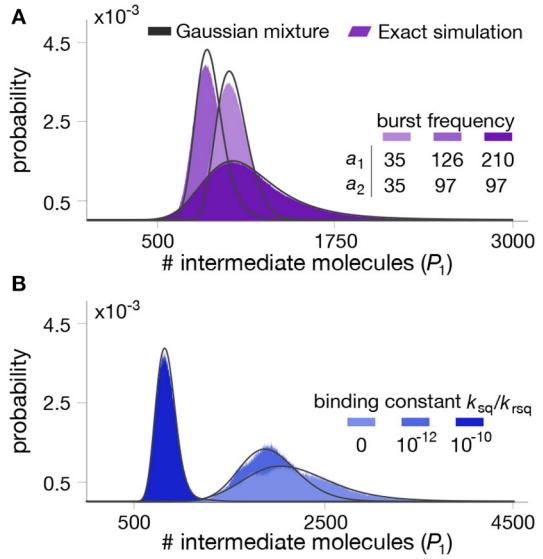

FIGURE 4 | Stationary distributions for the intermediate metabolite in a two-step pathway with end-product inhibition. The panels show the distribution of intermediate metabolite P_1 for different combinations of parameter values. **(A)** Impact of enzyme bursting frequency a_1 and a_2. **(B)** Impact of binding constant between the first enzyme and the end-product. All parameter values can be found in **Table 2**.

To test the quality of the approximation, we numerically computed the mixture model in (33) for various combinations of parameter values, shown in **Figure 4**. We observe that the mixture model offers an excellent approximation as compared to exact Gillespie simulations of the full model (18)–(25). We note that in this case, the full stochastic model has seven species and three different timescales, and therefore the runtime of Gillespie simulations are extremely long, in the order of several hours per run.

To further illustrate the utility of our method, we employed the mixture model to study the impact of parameter perturbations on the metabolite distributions. Without an analytical solution, such a study would require the computation of long Gillespie simulations for each combination of parameter values, which quickly become infeasible due to the long simulation time. In contrast, the mixture model provides a systematic way to rapidly evaluate the influence of model parameters on metabolite distributions. In **Figure 5A** we show summary statistics of the marginal $\mathbf{P}(P_1)$ for various combinations of average enzyme expression levels. The results suggest that expression levels can have a strong impact on the mean and coefficient of variation of the intermediate metabolite. Moreover, in **Figure 5B** we plot the distribution $\mathbf{P}(P_1,P_2)$ for combinations of bursting parameters. The results show that uncorrelated enzyme fluctuations can result in correlated metabolite distributions due to the coupling introduced by the pathway (Levine and Hwa, 2007).

TABLE 2 | Parameter values for simulations in **Figure 4**.

Figure 4					Figure 4A		Figure 4B
δ	0.00025 s^{-1}	$k_{rev,1}$	0.0001 s^{-1}	a_1	{35, 126, 210}	a_1	80
$k_{deg,1}$	0.2 s^{-1}	$k_{c,1}$	0.00025 s^{-1}	a_2	{35, 97, 97}	a_2	80
$k_{deg,2}$	0.2 s^{-1}	$k_{f,2}$	1.5 s^{-1}	b_1	1	b_1	1
S	3,000 molecules	$k_{b,2}$	15,000 s^{-1}	b_2	1	b_2	1
$k_{f,1}$	20 × S s^{-1}	$k_{cat,2}$	150 s^{-1}	k_{sq}	10^{-10} s^{-1}	k_{sq}	{0, 10^{-10}, 10^{-12}} s^{-1}
$k_{b,1}$	15,000 s^{-1}	$k_{rev,2}$	0.001 s^{-1}	k_{rsq}	1 s^{-1}	k_{rsq}	1 s^{-1}
$k_{cat,1}$	22.5 s^{-1}	$k_{c,2}$	0.15 s^{-1}	h	3	h	3

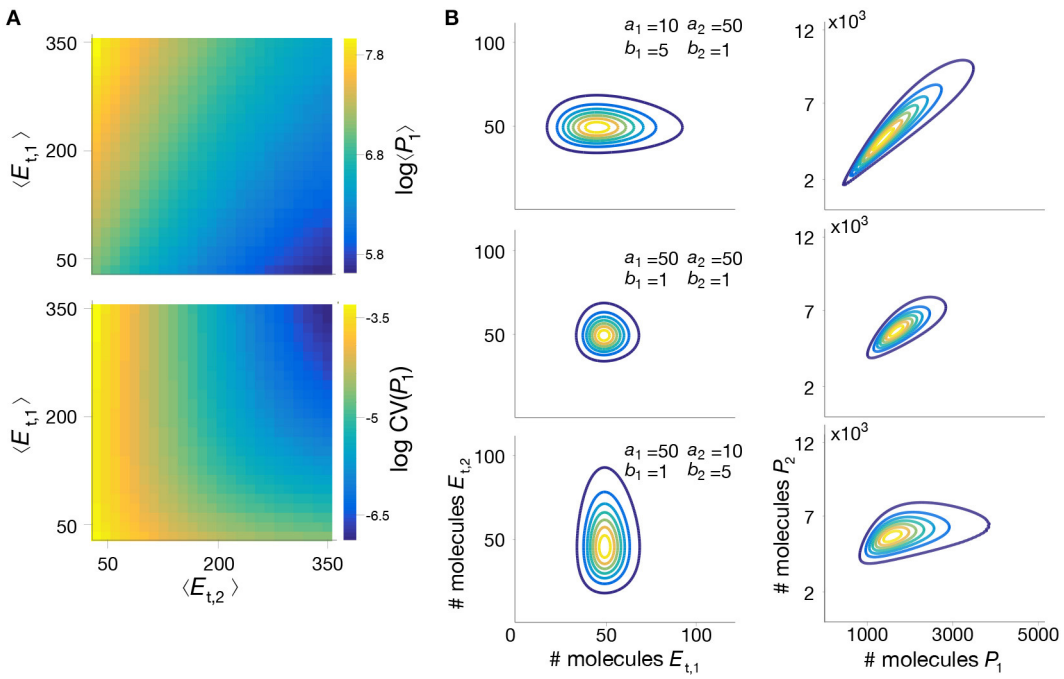

FIGURE 5 | Impact of enzyme expression on metabolite distributions. **(A)** We compute the mean and coefficient of variation (CV) of the intermediate metabolite P_1 in model (18)–(25), for a wide range of mean enzyme expression levels. **(B)** Enzyme expression parameters shape the metabolite distribution; We computed the joint metabolite distribution $\mathbf{P}(P_1, P_2)$ for three combinations of enzyme bursting parameters, chosen to give the same mean expression, and assuming both enzymes are expressed independently. Shown are contour plots of the bivariate distributions of enzymes (left) and metabolites (right). The results suggest that metabolite correlations emerge even when enzymes are uncorrelated, as reported previously in the literature (Levine and Hwa, 2007). All parameter values can be found in **Table 3**.

5. DISCUSSION

Cellular metabolism has traditionally been assumed to follow deterministic dynamics. This paradigm results largely from the observation that cellular metabolites are highly abundant. However, recent data shows that single-cell metabolite distributions can display substantial heterogeneity in their abundance across single cells (Bennett et al., 2009; Imamura et al., 2009; Lemke and Schultz, 2011; Paige et al., 2012; Ibáñez et al., 2013; Yaginuma et al., 2014; Esaki and Masujima, 2015; Xiao et al., 2016; Mannan et al., 2017). It has also been shown that expression of metabolic genes is as variable as any other component of the proteome (Taniguchi et al., 2010), and thus in principle it is plausible that such enzyme fluctuations propagate to metabolites. These observations have begun to challenge the paradigm of metabolism being a deterministic process, suggesting that metabolite fluctuations may play a role in non-genetic heterogeneity.

Here we described a new computational tool to predict the statistics of metabolite fluctuations in conjunction with gene expression. The method is based on a timescale separation argument and leads to a Gaussian mixture model for the stationary distribution of cellular metabolites. Computing distributions from this approximate model is substantially faster than through stochastic simulations, as these can be extremely slow due to the multiple timescales of metabolic pathways. Our technique can therefore be employed to efficiently explore the parameter space and predict the shape of metabolite distributions in different conditions. In earlier work we showed that the product of a single metabolic reaction can be accurately

TABLE 3 | Parameter values for simulations in **Figure 5**.

Figure 5A		Figure 5B	
a_1	[10, 100]	a_1	{10, 50, 50}
a_2	[10, 100]	a_2	{50, 50, 10}
b_1	1	b_1	{5, 1, 1}
b_2	1	b_2	{1, 1, 5}
k_{sq}	10^{-10} s^{-1}	k_{sq}	0 s^{-1}
k_{rsq}	1 s^{-1}	k_{rsq}	1 s^{-1}
h	3	h	3

described by a Poisson mixture model (Tonn et al., 2019). Such approximation allowed the discovery of previously unknown regimes for metabolite distributions, including heavily tailed distributions and various types of bimodality and multimodality. The Poisson approximation, however, is bespoke to single reactions and not valid for more complex systems. In contrast, the Gaussian mixture model discussed here can be applied to multiple kinetic mechanisms, more complex stoichiometries, as well as post-translational regulation.

An advantage of our approach is that the mixture weights can be computed offline from stochastic models for gene expression or single-cell expression data. The model is flexible in that it can readily accommodate gene expression models of various complexity. For the sake of illustration, in our examples we used the simple two-stage model for gene expression, but other models including gene regulation can also be employed (Dattani and Barahona, 2017). Particularly relevant models are those that account for enzyme co-regulation, a widespread feature of bacterial operons (Chubukov et al., 2014), which translates into correlations between expression of different pathway enzymes and the resulting metabolite abundances. A limitation of our method is that in many cases analytic solutions of the CME are not known, particularly for large models with multiple interacting genes. In such cases, the mixture weights $\mathbf{P}(E)$ can be approximated through stochastic simulations (Gillespie, 2007) albeit at the expense of increased computational costs. Most recently, progress in stochastic simulation of genome-scale metabolic networks (Tourigny et al., 2020) can offer an alternative route for studying fluctuations in large metabolic models.

The effectiveness of our method relies on two conditions: the separation of timescales between enzyme expression and enzyme catalysis, and the ability of the LNA to approximate the mixture components accurately. The first condition is satisfied by the vast majority of enzymes because their kinetics operate in regimes that are orders of magnitude faster than gene expression (Chubukov et al., 2014). However, the timescale separation can fail if the metabolic substrate S, typically a carbon source, cannot be assumed to be constant, a suitable assumption in the typical case of abundant nutrient sources with low fluctuations. Our theory would need to be extended in cases when nutrient sources become another source of variability, e.g., under fluctuations dictated by the environment (Dattani and Barahona, 2017). The second condition breaks down when the LNA fails to provide good estimates of the mixture components (Thomas and Grima, 2015; Andreychenko et al., 2017). As explained in section 2, here we have deliberately chosen to employ the LNA because it provides a simple and rapid method to compute the mixture components, $\mathbf{P}(P|E)$, for a broad range of metabolic pathways. Yet in cases where its assumptions do not hold, e.g., low abundance of metabolites, the LNA step in our method can be replaced by more accurate approximations. Such alternative methods include, for example, the conditional system size expansion including terms beyond the LNA, maximum entropy reconstructions using the method of conditional moments, or the finite state projection algorithm (Andreychenko et al., 2017; Gupta et al., 2017a), all of which can be readily incorporated into our mixture model strategy. These methods rely on different assumptions and their approximation quality will vary depending on the specific model parameters; in some cases, estimates for their approximation errors can be obtained with suitable methods, as discussed in a recent review on this topic (Kuntz et al., 2020).

Although our method can account for a large class of metabolic models and post-translational regulation mechanisms, there are a number of promising extensions that would broaden its utility in light of recent experimental advances. First, here we have only considered stationary distributions of metabolites, and a number of experiments have revealed cases in which metabolic heterogeneity emerges during dynamic nutrient shifts (Kotte et al., 2014; van Heerden et al., 2014; Nikolic et al., 2017). Extensions of our method to time-dependent metabolite distributions require the computation of the time-dependent solution of the CME for the enzyme expression model (Shahrezaei and Swain, 2008; Cao and Grima, 2018). As long as the dynamics of gene expression is slow enough to preserve the time scale separation, the computation of the mixture components with the LNA or other methods remains unchanged.

Another promising extension is the inclusion of transcriptional feedback regulation, a topic that has received substantial attention in the literature (Zaslaver et al., 2004; Chubukov et al., 2012; Chaves and Oyarzún, 2019; Lempp et al., 2019). In these systems, some pathway metabolites can bind to transcription factors (TF) that control enzyme expression in the same pathway. Such regulation can be included by using the conditional LNA method (Thomas and Grima, 2015) at the expense of not being able to compute the mixture weights offline anymore. Specifically, this extension would model mixture weights through more elaborate enzyme expression models in which the metabolite-TF interactions are replaced by their conditional averages, leading to an effective feedback model that requires specialized solution methods (Holehouse et al., 2020). A particularly promising application of such extended analysis is in synthetic biology, where there is a growing interest in the interplay between stochastic fluctuations and experimentally tunable parameters of molecular circuits (Briat et al., 2016; Boada et al., 2017). In particular, the use of metabolite-responsive feedback can improve robustness of strains engineered for the production of high-value metabolites (Oyarzún and Stan, 2013; Stevens and Carothers, 2015). Early results in this area (Oyarzún et al., 2015) suggest complex dependencies between metabolite fluctuations and the tunable parameters of the feedback control

system. Such analyses were purely based on lumped models for metabolite-TF binding, and hence a more detailed theory could reveal novel design strategies to mitigate metabolite heterogeneity in production strains.

A number of works have sought to find links between fluctuations across layers of cellular organization, such as gene expression, metabolism and cell growth (Kiviet et al., 2014; Kotte et al., 2014; van Heerden et al., 2014; Nikolic et al., 2017; Thomas et al., 2018). But since measurement of metabolites in single cells remains technically challenging, there is pressing need for computational methods to predict fluctuations in cellular metabolites. Our proposed method provides a systematic approach for such task, paving the way for the generation of hypotheses on the molecular sources of metabolic heterogeneity.

6. METHODS

6.1. Model Simulation

Stochastic simulations were computed with Gillespie's algorithm over long simulation times (several hours) corresponding to thousands of cell cycles. The ODE models and Lyapunov equations were solved in Matlab. In all examples, the negative binomial distribution for gene expression in (12) was computed with its continuum approximation (Gamma distribution).

6.2. Deterministic Rate Equations

6.2.1. Reversible Michaelis Menten

The full set of rate equations for the reversible reaction in (6)–(8) is:

$$\begin{aligned}\frac{\mathrm{d}P}{\mathrm{d}t} &= k_{\mathrm{cat}}C - k_{\mathrm{rev}}EP - k_{\mathrm{c}}P\\ \frac{\mathrm{d}E}{\mathrm{d}t} &= -k_{\mathrm{f}}SE + k_{\mathrm{b}}C + k_{\mathrm{cat}}C - k_{\mathrm{rev}}EP,\\ \frac{\mathrm{d}C}{\mathrm{d}t} &= k_{\mathrm{f}}SE - k_{\mathrm{b}}C - k_{\mathrm{cat}}C + k_{\mathrm{rev}}EP.\end{aligned} \tag{35}$$

To further reduce the above system of ODEs to Equation (13) in the main text, we can substitute the conservation relation in Equation (10), i.e. $C = E_{\mathrm{total}} - E$, and use the fact that the substrate-enzyme complex (C) typically equilibrates much faster than the product P, which means that $\mathrm{d}C/\mathrm{d}t \approx 0$ in the timescale of catalysis.

6.2.2. End-Product Inhibition

The full set of rate equations for the reactions in (18)–(23) is:

$$\begin{aligned}\frac{\mathrm{d}P_1}{\mathrm{d}t} &= k_{\mathrm{cat},1}C_1 - k_{\mathrm{rev},1}E_1P_1 - k_{\mathrm{f},2}E_2P_1 + k_{\mathrm{b},2}C_2 - k_{\mathrm{c},1}P_1\\ \frac{\mathrm{d}P_2}{\mathrm{d}t} &= k_{\mathrm{cat},2}C_2 - k_{\mathrm{rev},2}E_2P_2 - k_{\mathrm{sq}}E_1P_2^h + k_{\mathrm{rsq}}E^* - k_{\mathrm{c},2}P_2.\\ \frac{\mathrm{d}E_1}{\mathrm{d}t} &= -k_{\mathrm{f},1}SE_1 + \left(k_{\mathrm{b},1} + k_{\mathrm{cat},1}\right)C_1 - k_{\mathrm{rev},1}P_1E_1\\ &\quad -k_{\mathrm{sq}}P_2^hE_1 + k_{\mathrm{rsq}}E^*,\\ \frac{\mathrm{d}C_1}{\mathrm{d}t} &= k_{\mathrm{f},1}SE_1 - \left(k_{\mathrm{b},1} + k_{\mathrm{cat},1}\right)C_1 + k_{\mathrm{rev},1}P_1E_1,\\ \frac{\mathrm{d}E^*}{\mathrm{d}t} &= k_{\mathrm{sq}}P_2^hE_1 - k_{\mathrm{rsq}}E^*,\\ \frac{\mathrm{d}E_2}{\mathrm{d}t} &= -k_{\mathrm{f},2}P_1E_2 + \left(k_{\mathrm{b},2} + k_{\mathrm{cat},2}\right)C_1 - k_{\mathrm{rev},2}P_2E_2,\\ \frac{\mathrm{d}C_2}{\mathrm{d}t} &= k_{\mathrm{f},2}P_1E_2 - \left(k_{\mathrm{b},2} + k_{\mathrm{cat},2}\right)C_1 + k_{\mathrm{rev},2}P_2E_2\end{aligned} \tag{36}$$

As in the previous example, we can use the rapid equilibrium assumption and the conservation relations in (26), i.e., $E_{\mathrm{t},1} = E_1 + E^* + C_1$ and $E_{\mathrm{t},2} = E_2 + C_2$, to simplify the 7-dimensional ODE in (28) to the 2-dimensional system in (28) of the main text.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

MT carried out research, model simulation, model analysis, and wrote the paper. PT and MB contributed to model analysis and paper writing. DO designed the research, model analysis, and wrote the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the Human Frontier Science Program through a Young Investigator Grant (RGY0076-2015) awarded to DO, a UKRI Future Leaders Fellowship (MR/T018429/1) awarded to PT, and the EPSRC Centre for Mathematics of Precision Healthcare (EP/N014529/1) awarded to MB.

REFERENCES

Amantonico, A., Urban, P. L., and Zenobi, R. (2010). Analytical techniques for single-cell metabolomics: state of the art and trends. *Anal. Bioanal. Chem.* 398, 2493–2504. doi: 10.1007/s00216-010-3850-1

Andreychenko, A., Bortolussi, L., Grima, R., Thomas, P., and Wolf, V. (2017). *Distribution Approximations for the Chemical Master Equation: Comparison of the Method of Moments and the System Size Expansion*. Springer International Publishing, 39–66.

Bakker, E., and Swain, P. S. (2019). Estimating numbers of intracellular molecules through analysing fluctuations in photobleaching. *Sci. Rep.* 9, 1–13. doi: 10.1038/s41598-019-50921-7

Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* 5, 593–599. doi: 10.1038/nchembio.186

Binder, D., Drepper, T., Jaeger, K. E., Delvigne, F., Wiechert, W., Kohlheyer, D., et al. (2017). Homogenizing bacterial cell factories: analysis and

engineering of phenotypic heterogeneity. *Metab. Eng.* 42, 145–156. doi: 10.1016/j.ymben.2017.06.009

Boada, Y., Vignoni, A., and Picó, J. (2017). Engineered control of genetic variability reveals interplay among quorum sensing, feedback regulation, and biochemical noise. *ACS Synth. Biol.* 6, 1903–1912. doi: 10.1021/acssynbio.7b00087

Briat, C., Gupta, A., and Khammash, M. (2016). Antithetic integral feedback ensures robust perfect adaptation in noisy bimolecular networks. *Cell Syst.* 2, 15–26. doi: 10.1016/j.cels.2016.01.004

Cao, Y., Gillespie, D. T., and Petzold, L. R. (2005). Accelerated stochastic simulation of the stiff enzyme-substrate reaction. *J. Chem. Phys.* 123:144917. doi: 10.1063/1.2052596

Cao, Z., and Grima, R. (2018). Linear mapping approximation of gene regulatory networks with stochastic dynamics. *Nat. Commun.* 9:3305. doi: 10.1038/s41467-018-05822-0

Chaves, M., and Oyarzún, D. A. (2019). Dynamics of complex feedback architectures in metabolic pathways. *Automatica* 99, 323–332. doi: 10.1016/j.automatica.2018.10.046

Chubukov, V., Gerosa, L., Kochanowski, K., and Sauer, U. (2014). Coordination of microbial metabolism. *Nat. Rev. Microbiol.* 12, 327–340. doi: 10.1038/nrmicro3238

Chubukov, V., Zuleta, I. A., and Li, H. (2012) Regulatory architecture determines optimal regulation of gene expression in metabolic pathways. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5127–5132. doi: 10.1073/pnas.1114235109

Cornish-Bowden, A. (2004). *Fundamentals of Enzyme Kinetics, 3rd Edn*. London: Portland Press Ltd.

Dattani, J., and Barahona, M. (2017). Stochastic models of gene transcription with upstream drives: exact solution and sample path characterization. *J. R. Soc. Interf.* 14:20160833. doi: 10.1098/rsif.20016.833

Deris, J. B., Kim, M., Zhang, Z., Okano, H., Hermsen, R., Groisman, A., et al. (2013). The innate growth bistability and fitness landscapes of antibiotic resistant bacteria. *Science* 342:1237435. doi: 10.1126/science.1237435

Elf, J., and Ehrenberg, M. (2003). Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res.* 13, 2475–2484. doi: 10.1101/gr.1196503

Esaki, T., and Masujima, T. (2015). Fluorescence probing live single-cell mass spectrometry for direct analysis of organelle metabolism. *Analyt. Sci.* 31, 1211–1213. doi: 10.2116/analsci.31.1211

Gillespie, D. T. (2007). Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* 1716, 1716–1733. doi: 10.1063/1.1378322

Golding, I., Paulsson, J., Zawilski, S. M., and Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell* 123, 1025–1036. doi: 10.1016/j.cell.2005.09.031

Gupta, A., Mikelson, J., and Khammash, M. (2017a). A finite state projection algorithm for the stationary solution of the chemical master equation. *J. Chem. Phys.* 147:154101. doi: 10.1063/1.5006484

Gupta, A., Milias-Argeitis, A., and Khammash, M. (2017b). Dynamic disorder in simple enzymatic reactions induces stochastic amplification of substrate. *J. R. Soc.* 14, 1–29. doi: 10.1098/rsif.2017.0311

Heinemann, M., and Zenobi, R. (2011). Single cell metabolomics. *Curr. Opin. Biotechnol.* 22, 26–31. doi: 10.1016/j.copbio.2010.09.008

Holehouse, J., Cao, Z., and Grima, R. (2020). Stochastic modeling of auto-regulatory genetic feedback loops: a review and comparative study. *Biophys. J.* 118, 1517–1525. doi: 10.1016/j.bpj.2020.02.016

Ibáñez, A. J., Fagerer, S. R., Schmidt, A. M., Urban, P. L., Jefimovs, K., Geiger, P., et al. (2013). Mass spectrometry-based metabolomics of single yeast cells. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8790–8794. doi: 10.1073/pnas.1209302110

Imamura, H., Nhat, K. P. H., Togawa, H., Saito, K., Iino, R., Kato-Yamada, Y., et al. (2009). Visualization of ATP levels inside single living cells with fluorescence resonance energy transfer-based genetically encoded indicators. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15651–15656. doi: 10.1073/pnas.0904764106

Kiviet, D. J., Nghe, P., Walker, N., Boulineau, S., Sunderlikova, V., and Tans, S. J. (2014). Stochasticity of metabolism and growth at the single-cell level. *Nature* 514, 376–379. doi: 10.1038/nature13582

Kotte, O., Volkmer, B., Radzikowski, J. L., and Heinemann, M. (2014). Phenotypic bistability in Escherichia coli' s central carbon metabolism. *Mol. Syst. Biol.* 10:736. doi: 10.15252/msb.20135022

Kuntz, J., Oyarzún, D. A., and Stan, G. B. V. (2013). "Model reduction of genetic-metabolic networks via time scale separation," in *A Systems Theoretic Approach to Systems and Synthetic Biology*, eds V. Kulkarni, G.-B. Stan, and K. Raman (Springer), 181–210.

Kuntz, J., Thomas, P., Stan, G. B. V., and Barahona, M. (2020). Stationary distributions of continuous-time Markov chains: a review of theory and truncation-based approximations. *SIAM Rev.*

Labhsetwar, P., Cole, J. A., Roberts, E., Price, N. D., and Luthey-Schulten, Z. A. (2013). Heterogeneity in protein expression induces metabolic variability in a modeled *Escherichia coli* population. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14006–14011. doi: 10.1073/pnas.1222569110

Lemke, E. A., and Schultz, C. (2011). Principles for designing fluorescent sensors and reporters. *Nat. Chem. Biol.* 7, 480–483. doi: 10.1038/nchembio.620

Lempp, M., Farke, N., Kuntz, M., Freibert, S. A., Lill, R., and Link, H. (2019). Systematic identification of metabolites controlling gene expression in *E. coli*. *Nat. Commun.* 10:4463. doi: 10.1038/s41467-019-12474-1

Levine, E., and Hwa, T. (2007). Stochastic fluctuations in metabolic pathways. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9224–9229. doi: 10.1073/pnas.0610987104

Liu, D., Mannan, A. A., Han, Y., Oyarzún, D. A., and Zhang, F. (2018). Dynamic metabolic control: towards precision engineering of metabolism. *J. Indus. Microbiol. Biotechnol.* 45, 535–543. doi: 10.1007/s10295-018-2013-9

Loftus, R. M., and Finlay, D. K. (2016). Immunometabolism: cellular metabolism turns immune regulator. *J. Biol. Chem.* 291, 1–10. doi: 10.1074/jbc.R115.693903

Lugagne, J. B., Oyarzún, D. A., and Stan, G. B. (2013). "Stochastic simulation of enzymatic reactions under transcriptional feedback regulation," in *Proceeding of the European Control Conference* (Zurich), 3646–3651.

Mannan, A. A., Liu, D., Zhang, F., and Oyarzún, D. A. (2017). Fundamental design principles for transcription-factor-based metabolite biosensors. *ACS Synth. Biol.* 6, 1851–1859. doi: 10.1021/acssynbio.7b00172

Murabito, E., Verma, M., Bekker, M., Bellomo, D., Westerhoff, H. V., Teusink, B., et al. (2014). Monte-Carlo modeling of the central carbon metabolism of lactococcus lactis: insights into metabolic regulation. *PLoS ONE* 9:e106453. doi: 10.1371/journal.pone.0106453

Nikolic, N., Schreiber, F., Co, A. D., Kiviet, D. J., Bergmiller, T., Littmann, S., et al. (2017). Cell-to-cell variation and specialization in sugar metabolism in clonal bacterial populations. *PLoS Genet.* 13:e1007122. doi: 10.1371/journal.pgen.1007122

Okumus, B., Landgraf, D., Lai, G. C., Bakhsi, S., Arias-Castro, J. C., Yildiz, S., et al. (2016). Mechanical slowing-down of cytoplasmic diffusion allows in vivo counting of proteins in individual cells. *Nat. Commun.* 7, 1–11. doi: 10.1038/ncomms12130

Oyarzún, D. A., Lugagne, J. B., and Stan, G. B. V. (2015). Noise propagation in synthetic gene circuits for metabolic control. *ACS Synth. Biol.* 4, 116–125. doi: 10.1021/sb400126a

Oyarzún, D. A., and Stan, G. B. V. (2013). Synthetic gene circuits for metabolic control: design trade-offs and constraints. *J. R. Soc. Interface* 10:20120671. doi: 10.1098/rsif.2012.0671

Paige, J. S., Nguyen-Duc, T., Song, W., and Jaffrey, S. R. (2012). Fluorescence imaging of cellular metabolites with RNA. *Science* 335:1194. doi: 10.1126/science.1218298

Puchałka, J., and Kierzek, A. M. (2004). Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. *Biophys. J.* 86, 1357–1372. doi: 10.1016/S0006-3495(04)74207-1

Radzikowski, J. L., Schramke, H., and Heinemann, M. (2017). Bacterial persistence from a system-level perspective. *Curr. Opin. Biotechnol.* 46, 98–105. doi: 10.1016/j.copbio.2017.02.012

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its Consequences. *Cell* 135, 216–226. doi: 10.1016/j.cell.2008.09.050

Rathinam, M., Petzold, L. R., Cao, Y., and Gillespie, D. T. (2003). Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method. *J. Chem. Phys.* 119, 12784–12794. doi: 10.1063/1.1627296

Reid, M. A., Dai, Z., and Locasale, J. W. (2017). The impact of cellular metabolism on chromatin dynamics and epigenetics. *Nat. Cell Biol.* 19, 1298–1306. doi: 10.1038/ncb3629

Rosenfeld, N., Perkins, T. J., Alon, U., Elowitz, M. B., and Swain, P. S. (2006). A fluctuation method to quantify *in vivo* fluorescence data. *Biophys. J.* 91, 759–766. doi: 10.1529/biophysj.105.073098

Schmitz, A. C., Hartline, C. J., and Zhang, F. (2017). Engineering microbial metabolite dynamics and heterogeneity. *Biotechnol. J.* 12:1700422. doi: 10.1002/biot.201700422

Schreiber, F., Littmann, S., Lavik, G., Escrig, S., Meibom, A., Kuypers, M. M. M., et al. (2016). Phenotypic heterogeneity driven by nutrient limitation promotes growth in fluctuating environments. *Nat. Microbiol.* 1:16055. doi: 10.1038/nmicrobiol.2016.55

Shahrezaei, V., and Swain, P. S. (2008). Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17256–17261. doi: 10.1073/pnas.0803850105

Shan, Y., Gandt, A. B., Rowe, S. E., Deisinger, J. P., Conlon, B. P., and Lewis, K. (2017). ATP-dependent persister formation in *Escherichia coli*. *mBIO* 8, 1–14. doi: 10.1128/mBio.02267-16

Şimşek, E., and Kim, M. (2018). The emergence of metabolic heterogeneity and diverse growth responses in isogenic bacterial cells. *ISME J.* 12, 1199–1209. doi: 10.1038/s41396-017-0036-2

Stevens, J. T., and Carothers, J. M. (2015). Designing RNA-based genetic control systems for efficient production from engineered metabolic pathways. *ACS Synth. Biol.* 4, 107–115. doi: 10.1021/sb400201u

Swain, P. S., Elowitz, M. B., and Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12795–12800. doi: 10.1073/pnas.162041399

Takhaveev, V., and Heinemann, M. (2018). Metabolic heterogeneity in clonal microbial populations. *Curr. Opin. Microbiol.* 45, 30–38. doi: 10.1016/j.mib.2018.02.004

Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., et al. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329, 533–538. doi: 10.1126/science.1188308

Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8614–8619. doi: 10.1073/pnas.151588598

Thomas, P., and Grima, R. (2015). Approximate probability distributions of the master equation. *Phys. Rev. E* 92:012120. doi: 10.1103/PhysRevE.92.012120

Thomas, P., Popović, N., and Grima, R. (2014). Phenotypic switching in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6994–6999. doi: 10.1073/pnas.1400049111

Thomas, P., Terradot, G., Danos, V., and Weiße, A. Y. (2018). Sources, propagation and consequences of stochasticity in cellular growth. *Nat. Commun.* 9, 1–11. doi: 10.1038/s41467-018-06912-9

Tonn, M. K. (2020). *Stochastic modelling and analysis of metabolic heterogeneity in single cells* (Ph.D. thesis). Imperial College London, London, United Kingdom.

Tonn, M. K., Thomas, P., Barahona, M., and Oyarzún, D. A. (2019). Stochastic modelling reveals mechanisms of metabolic heterogeneity. *Commun. Biol.* 2:108. doi: 10.1038/s42003-019-0347-0

Tourigny, D., Goldberg, A., and Karr, J. (2020). Simulating single-cell metabolism using a stochastic flux-balance analysis algorithm. *bioRxiv*. doi: 10.1101/2020.05.22.110577

van Heerden, J. H., Wortel, M. T., Bruggeman, F. J., Heijnen, J. J., Bollen, Y. J. M., Planqué, R., et al. (2014). Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science* 343:1245114. doi: 10.1126/science.1245114

van Kampen, N. G. (1992). *Stochastic Processes in Physics and Chemistry*. Amsterdam: Elsevier.

Vilhena, C., Kaganovitch, E., Shin, J. Y., Grünberger, A., Behr, S., Kristoficova, I., et al. (2018). A single-cell view of the BtsSR/YpdAB pyruvate sensing network in *Escherichia coli* and its biological relevance. *J. Bacteriol.* 200, 1–13. doi: 10.1128/JB.00536-17

Wehrens, M., Büke, F., Nghe, P., and Tans, S. J. (2018). Stochasticity in cellular metabolism and growth: approaches and consequences. *Curr. Opin. Syst. Biol.* 8, 131–136. doi: 10.1016/j.coisb.2018.02.006

Weiße, A. Y., Oyarzún, D. A., Danos, V., and Swain, P. S. (2015). Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1038–E1047. doi: 10.1073/pnas.1416533112

Xiao, Y., Bowen, C. H., Liu, D., and Zhang, F. (2016). Exploiting non-genetic, cell-to-cell variation for enhanced biosynthesis. *Nat. Chem. Biol.* 12, 339–344. doi: 10.1038/nchembio.2046

Yaginuma, H., Kawai, S., Tabata, K. V., Tomiyama, K., Kakizuka, A., Komatsuzaki, T., et al. (2014). Diversity in ATP concentrations in a single bacterial cell population revealed by. *Sci. Rep.* 4:6522. doi: 10.1038/srep06522

Zaslaver, A., Mayo, A. E., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., et al. (2004). Just-in-time transcription program in metabolic pathways. *Nat. Genet.* 36, 486–491. doi: 10.1038/ng1348

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.

frontiers in Cell and Developmental Biology

BRIEF RESEARCH REPORT
published: 07 January 2021
doi: 10.3389/fcell.2020.594750

High Content Analysis Across Signaling Modulation Treatments for Subcellular Target Identification Reveals Heterogeneity in Cellular Response

*Sayan Biswas**

Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad, India

OPEN ACCESS

Edited by:
Ankur Sharma,
Genome Institute of Singapore, Singapore

Reviewed by:
Aaron Goldman,
Harvard Medical School, United States
Tamal Das,
Tata Institute of Fundamental Research, India
Shamik Sen,
Indian Institute of Technology Bombay, India

***Correspondence:**
Sayan Biswas
sayanbiswasjuee@gmail.com

Specialty section:
This article was submitted to Molecular Medicine, a section of the journal Frontiers in Cell and Developmental Biology

Received: *14 August 2020*
Accepted: *09 November 2020*
Published: *07 January 2021*

Citation:
Biswas S (2021) High Content Analysis Across Signaling Modulation Treatments for Subcellular Target Identification Reveals Heterogeneity in Cellular Response. Front. Cell Dev. Biol. 8:594750. doi: 10.3389/fcell.2020.594750

Cellular phenotypes on bioactive compound treatment are a result of the downstream targets of the respective treatment. Here, a computational approach is taken for downstream subcellular target identification to understand the basis of the cellular response. This response is a readout of cellular phenotypes captured from cell-painting-based light microscopy images. The readouts are morphological profiles measured simultaneously from multiple cellular organelles. Cellular profiles generated from roughly 270 diverse treatments on bone cancer cell line form the high content screen used in this study. Phenotypic diversity across these treatments is demonstrated, depending on the image-based phenotypic profiles. Furthermore, the impact of the treatments on specific organelles and associated organelle sensitivities are determined. This revealed that endoplasmic reticulum has a higher likelihood of being targeted. Employing multivariate regression overall cellular response is predicted based on fewer organelle responses. This prediction model is validated against 1,000 new candidate compounds. Different compounds despite driving specific modulation outcomes elicit a varying effect on cellular integrity. Strikingly, this confirms that phenotypic responses are not conserved that enables quantification of signaling heterogeneity. Agonist-antagonist signaling pairs demonstrate switch of the targets in the cascades hinting toward evidence of signaling plasticity. Quantitative analysis of the screen has enabled the identification of these underlying signatures. Together, these image-based profiling approaches can be employed for target identification in drug and diseased states and understand the hallmark of cellular response.

Keywords: phenotypic similarity, signaling modulation, cellular and organelle behavior, predictive modeling, heterogeneity in responses, mechanism of action, high content imaging screen

1. INTRODUCTION

Measurement of biological activity upon small molecule-based treatment has the potential to illustrate the mechanisms of action by comparing it with profiles of known compounds (Hughes et al., 2000; Lamb et al., 2006; Feng et al., 2009). These measurements from high-throughput target-directed screens have been widely used for their potential application in drug discovery

through unbiased testing of several million compounds per screen (Macarron et al., 2011). Phenotypic screening has also been proposed for efficient assessment of drug candidate testing in biological systems (Lee et al., 2012; Futamura et al., 2013). These approaches are facilitated by quantitative microscopy, widely used in pharmaceutical and academic labs, since it provides a versatile and powerful readout for precise cellular measurements and identifying cellular states (Carpenter, 2007; Futamura et al., 2013). The principle of phenotypic profiling is based on summarizing multiparametric, feature-based analysis of cellular phenotypes of each sample so that sample similarities are reflected on similarities between profiles (Wagner and Clemons, 2009). Transcript expression and proteomics profiling serve as established biological readouts (Hughes et al., 2016; Szalai et al., 2019). In comparison, image-based profiling is cost effective and flexible for scaling between medium and high throughput with relative ease, alongside providing phenotypic details at single-cell resolution (Ljosa et al., 2013). Although image-based screens aim to score samples with respect to one or a few known phenotypes, profiling experiments aim to capture phenotypes not known in advance, using a variety of subtle cellular responses and widely used as predictive models (Ljosa et al., 2013; Kandaswamy et al., 2016; Steigele et al., 2020). A mechanism of action (MoA) usually refers to biochemical interaction through which the drug acts to induce pharmacological effect and phenotypic changes (Kandaswamy et al., 2016), which can be studied based on the phenotype.

This potent research paradigm has been employed over the past few decades by the pharmaceutical and biotechnology sectors (Moffat et al., 2014, 2017). Drug discovery through cell systems biology could significantly reduce the time and cost of new drug development (Butcher, 2007). Automated high-content microscopy imaging and image analysis methods offer an efficient alternative to the traditional target-directed screening approach (Lang et al., 2006; Simm et al., 2018; Nyffeler et al., 2020). This allows researchers to study the cellular phenotype response on molecular perturbations irrespective of putative target activity (Tanaka et al., 2005; Low et al., 2008). Computational application of such methods to study cellular response relies heavily on active measurements that can capture a spectrum of phenotype. Assays with multiple fluorescent markers enable to capture quantitative profiles in high throughput. These methods provide an unbiased approach to study cell states associated with chemical perturbation and disease state to support future probe discovery. Such cellular assays show the value of phenotypic profiling to assist not only in the identification of cellular activity and but also to develop an understanding to elucidate the MoA for drugs whose mode of action or primary targets are unknown (Loo et al., 2007; Young et al., 2008; Caie et al., 2010; Breinig et al., 2015). The ability to identify the targets of candidate molecules in a screen can help overcome one of the bottlenecks to establish it as a drug. Although experimental approaches for target identification in a screen could be labor, resource, and time intensive; computational approaches substantially reduce the work and resource requirement for favorable application (Perlman et al., 2011; Chen et al., 2016; Madhukar et al., 2019). However, a key challenge in the field is the identification of the sub-cellular effects caused by the treatment and also understand the basis of the cellular responses.

In this report, it is aimed to identify the downstream organelle targets by using computation approaches on the "cell-painting" assay screen. A quantitative understanding of the heterogeneous cellular responses in the treatment screen based on the subtle changes in cellular phenotype profiles is demonstrated. This approach fosters the possibility for a quantitative examination of the responses induced by selective pharmacologic agents across cancerous cells. Subsequent analysis demonstrated the role of conserved and differential signatures in the diverse organelle behavior in the multifaceted cellular response. This interconnected dependence is exploited for developing models to predict the overall cellular response based on specific organelle response. Further advancement is achieved through fine quantification, which elucidated the varying cellular response even when the treatment outcome is conserved hinting toward signaling heterogeneity.

2. METHODS

2.1. Dataset

Here the "Cell-Painting" (Bray et al., 2016) assay as documented in BBBC022v1 (Gustafsdottir et al., 2013) has been used. This is publicly available from the Broad Bioimage Benchmark Collection (Ljosa et al., 2012) and is one of the widely used dataset in the field. The raw data have been downloaded as documented in an earlier published report (Gustafsdottir et al., 2013) (from http://www.broadinstitute.org/pubs/gustafsdottir_plosone_2013/). In this dataset, bone carcinoma U2OS cells are imaged on treatment with multiple bioactive compounds. The cells are fluorescently labeled to follow the components: Golgi, endoplasmic reticulum (ER), nuclei (Hoechst), nucleoli (Syto), and mitochondria (Mito). The bio-active compounds are chemical perturbations and are referred to have specific BroadID. To specifically annotate the treatment compounds with the relevance of the induced phenotype or the respective MoA, the "ground truth" of the image data (Corsello et al., 2017) made available as part of the BBBC036 (Bray et al., 2017) from the Broad Bioimage Benchmark Collection (Ljosa et al., 2012) has been used. This allowed ∼270 MoAs to be successfully annotated (**Supplementary Table 1**), which forms the working dataset for this study. There were roughly 1,000 compounds (or BroadIDs) (**Supplementary Table 2**) in the dataset for which MoA could not located based on BBBC036 file. These compounds have been used as test compounds for the prediction model developed as described in **Figure 3C**.

Cell Profiler (Carpenter et al., 2006) pipeline has been engineered (Gustafsdottir et al., 2013) to extract rich quantitative features (**Supplementary Table 3**) at single-cell resolution from the light microscopy images. For locating the origin of features, the occurrences of acronym Golgi, ER, Hoechst, Syto, and Mito in feature labels are used for feature originating from Golgi, endoplasmic reticulum, nucleus, nucleoli, and mitochondria, respectively (**Supplementary Table 3**). These features are used to report the dynamics associated with the specific organelle.

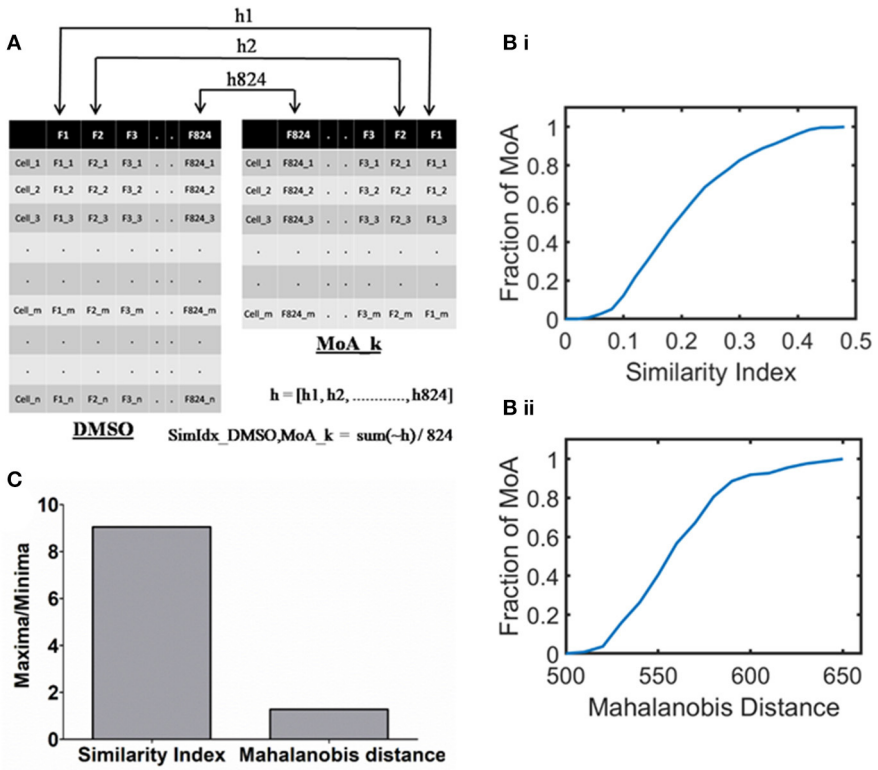

FIGURE 1 | Quantifying the cellular response upon mechanism of action inducing treatment with respect to DMSO. **(A)** A schematic outlining the method of Similarity Index calculation between DMSO (with n cells) and kth MoA (with m cells). Note that 824 single cell features (represented as columns) form the phenotypic profiles. **(B)** Cumulative histogram of the MoAs in the annotated dataset based on (i) SimIdx and (ii) Mahalanobis distance, demonstrating the varied response among the different signaling modulation treatment. **(C)** Quantification of the dynamic range (maxima by minima) calculated for each of the metric.

An example of the dataset with 100 cells is illustrated in **Supplementary Table 5**.

2.2. Analysis

The script developed for the analysis presented in this study is done using MatLab. Required details of the parameters have been enlisted in the respective section 3.

- Similarity index: Two-sample *t*-test has been performed at a 5% significance level. Two tail test has been performed. It is performed to see if a feature has changed significantly in an MoA with respect to the same feature for DMSO treatment (**Figure 1A**). Two sample test is performed to compare between DMSO and the MoA for which hypothesis testing is performed.
- Mahalanobis distance: This computes the distance in multivariate space between a point and distribution. The features of DMSO form the distribution and each cell at MoA forms a point. This measure is often used for outlier filtering in biomedical multivariate data (Laurikkala et al., 2000). Similarly, in this case prior to plotting histogram (**Figure 1B**ii), outlier detection has been performed.
- Multivariate regression: This has been performed using *fitlm* (Holland and Welsch, 1977). All possible combinations of two and three organelles have been used as predictor variables. The output (response) variable is the overall cellular response. The target is to regression model the organelle response (all combinations of two and three variables) to predict overall cellular response.
- Goodness of prediction: The multivariate regression models are used to estimate the goodness of prediction. For the testing, the test set contains data from 1,000 of new compounds. The predictor values are derived from the test dataset and thereby response value is estimated based on the regression model. To check the goodness of prediction, the estimated response value and actual response value are compared. This error is used to determine sum of squares due to error (SSE) and the total sum of squares (SST). R square value is calculated as $1-SSE/SST$. A good prediction would mean low error, which means a higher R square value.

3. RESULTS

3.1. Quantification of the Cellular Response to Signaling Modulation

Signaling modulation through chemical agent treatment causes a spectrum of phenotypic responses (Kitano, 2002; Wawer et al., 2014) in the cells. These responses or the cellular integrity changes, as a result of the treatment, could be

captured from the cellular morphology with the help of quantitative microscopy. The publicly available dataset of "cell painting" extracts these morphological phenotypic profiles from various cellular compartments through microscopy and image quantification at the single-cell resolution. Furthermore, the outcome of the signaling modulation (or the MoA) that the respective compound induces has also been annotated. Hence to study the cellular responses and effect of the treatments, the cell-painting assay dataset has been chosen. Phenotypic profiling summarizes cellular phenotypes upon the treatment, allowing the study of similarities between treatment by studying the profiles (Wagner and Clemons, 2009). The DMSO-treated cells are also profiled to extract the rich quantitative features. Here DMSO serves as the control (Galvao et al., 2014) for the chemical agent treatment. Therefore, to systematically address the cellular response due to the treatment, the similarity in the morphological features (**Supplementary Table 3**) between the treatment and DMSO is assessed.

To quantify this, the significance is tested between respective features of MoA with that of DMSO through p-value as illustrated in **Figure 1A**. Here *h* is a binary array that contains 824 elements, where x^{th} element signifies whether x^{th} feature is similar [0] or not [1] between the MoA treatment and DMSO. The parameter of similarity index (SimIdx) is then quantified for the MoA based on the fraction of similar features (number of zeros in the "h" array) between DMSO and the MoA. Thus, SimIdx is calculated to depict the similarity between the MoA inducing treatment and DMSO in terms of the phenotypic features, which can have a value between 0 and 1 signifying minimum and maximum similarity, respectively.

This process is then iterated across all the MoA treatments in the working dataset. Thus, a cumulative histogram is plotted to show the distribution of SimIdx calculated across all the MoA inducing treatments as shown in **Figure 1Bi**. While a fraction of MoAs has SimIdx close to 0, a significant fraction has it close to 0.5, the highest end of the curve. Therefore, the dynamic band of SimIdx helped to identify the spectrum of responses different MoA poses with respect to DMSO in terms of phenotypic similarity.

To compare this observation, the established method of Mahalanobis distance is also used to determine the phenotypic difference MoA exhibits with that of DMSO. Briefly, this metric helps capture distances in a multivariate feature space. A lesser value of Mahalanobis distance would signify lesser difference between the phenotypes of MoA and DMSO and vice versa. As mentioned earlier, a cumulative histogram is obtained for the Mahalanobis distance metric calculated between the MoAs in the working dataset and DMSO (**Figure 1B**ii). This affirms the varying response the MoAs contained in the dataset exhibit.

Based on the histograms (**Figure 1B**), dynamic range is derived by calculating (dividing maxima by minima) for each of the metrics as indicated in **Figure 1C**. The same working dataset of MoA treatments has been calibrated with both the metric but the SimIdx resolves the innate differential response better than Mahalanobis distance as indicated from the dynamic range. Put together, these two quantitative measures reflect the differential response various MoA exhibits with respect to DMSO.

3.2. Impact on the Organelle Induced by the Treatment

Recognition of sub-cellular compartments affected by a modulation treatment is critical to identify the respective treatment's downstream target. Therefore, the next aim is to monitor the impact on specific organelles. For this, the metric of SimIdx is utilized that enables to compare the phenotypic changes caused by the signaling modulation treatments. The cell painting assay facilitates to address this since it allows simultaneous monitoring of multiple organelles—ER, nucleus (Hoechst), nuclelous (Syto), mitochondria (Mito), and Golgi (Golgi) through targeted fluorescent probes. Thereby the features that originate from the specific organelle targeted fluorescent labels are explicitly identified (**Supplementary Table 3**) among all the features. By specifically comparing the organelle features, the organelle SimIdx has been determined. This calculation paved the understanding of how particular organelle integrity changes upon a signaling modulation treatment. Thus for every MoA treatment, it resulted in five values of organelle SimIdx, one for each of the organelle. These organelle SimIdx values account for changes in that specific organelle integrity due to the signaling modulation treatment. The distribution of this parameter for all the MoA inducing treatments available in the dataset is plotted as a cumulative histogram in **Figure 2A**. The graph shows that this parameter encompasses a diverse range, revealing the variation in the organelle response as well. To investigate this response, the existence of any coupling between the integrity changes in the overall cell and those of specific organelle for respective signaling modulations treatment is examined. Based on the diverse MoA inducing treatments, the plot of specific organelle SimIdx vs. all feature SimIdx is illustrated in **Figure 2B**. The trend shows that changes in the cellular response are reflected as conserved changes in the organelle response. The subtle changes of organelle phenotype are conserved with respect to overall cellular integrity changes, which elucidates an underlying conserved signature in the cellular responses.

As mentioned earlier, each organelle phenotype is profiled based on more than 100 features (**Supplementary Table 3**). It is then assessed; each of these features is affected by how many of the treatments? To extract this information, Impact Index (ImpIdx) is quantified for each of the features. First, a binary array *SimVal* is determined for each feature (in Equation 2 it is represented for 1*st* feature or $f1$), which is an array of 270 elements (number of MoAs in this study). The i^{th} element of this binary array signifies if $f1$ has been affected [1] by the i^{th} MoA inducing treatment or not [0]. ImpIdx value for feature $f1$ is then calculated by adding all elements of $SimVal_{f1}$ as per Equation (2). For every feature, the ImpIdx value would be between zero and the total number of MoA inducing element in the dataset where the extremes would mean that the feature has been impacted significantly for none or all of the MoA inducing treatment. Thus, this parameter is directly proportional to the likelihood estimate of the feature to be impacted upon a treatment.

$$SimVal_{f1} = [0, 1, 1, 0, ..., 1, 0](\sim 270 Elements) \quad (1)$$

$$ImpIdx_{f1} = \sum(SimVal_{f1}) \quad (2)$$

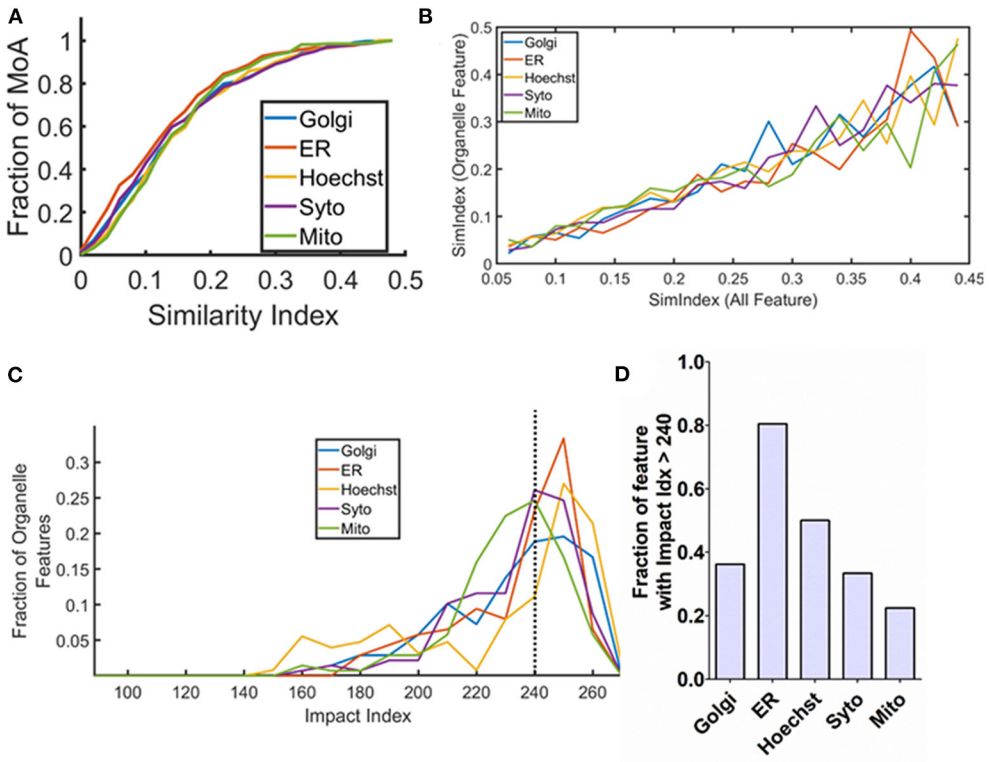

FIGURE 2 | Specific organelle-based responses on the signaling modulation treatment. **(A)** Cumulative histogram of the similarity index calculated based on the organelle features based on the mechanism of actions (MoAs) in the annotated dataset. **(B)** Co-relation curve showing overall SimIdx and organelle-specific SimIdx. The graph is obtained using all the MoAs in the dataset. **(C)** Representation of impact index of the organelle features in the form of a histogram. **(D)** Bar plot showing the fraction of organelle features that has impact index value of more than 240.

In this way, ImpIdx values are obtained for all the 824 features. Using this metric, the aim is to assess which organelles have a higher chance of being impacted downstream of the treatments. To pursue this, ImpIdx from specific organelle features are then collated. The distribution of organelle ImpIdx is shown in **Figure 2C**. Based on this distribution, a finer quantification is performed to identify the characteristic downstream target organelle. This is identified based on the fraction of organelle features that demonstrate ImpIdx of greater than 240 (roughly 90% of its maximum possible value, $0.9 \times 270 = 243$). These organelle fractions are represented in **Figure 2D**, which shows that 80% of ER features express quite high ImpIdx. These revealed that ER is a downstream target for most of the drug treatments performed in this study. In contrast, the mitochondria features express comparatively lesser impact, likely signifying the less pronounced effect by these treatment molecules on mitochondria. Put together, this analysis not only showed coupling between the overall cellular and specific organelle response but also established organelle signatures based on its likelihood of being affected upon treatment.

3.3. Sensitivity Detection and Prediction of Overall Cellular Response

Sensitivity could be one of the hallmarks of biological response and can be useful to extract a direct relationship between the pharmacological agent treatment and resultant downstream response. To address this, first, the correlation curves (**Figure 2B**) are characterized by regression modeling. These regression models are developed separately for each of the organelles and depicted in **Figure 3A** (**Supplementary Table 6**). Next, the first-order derivative is computed on these curves to extract the sensitivity of the organelle response due to the treatment. The resultant sensitivities are shown in **Figure 3B**, which shows there is not any significant bias in terms of organelle sensitivity.

Furthermore, these regression models are also adapted to develop more generalized predictive models. These models shall allow researchers to determine the overall effect of test compounds on the cellular integrity and range its application into orphan compounds. Multivariate regression is performed with the independent variable as the organelle SimIdx and dependent variable as all-feature SimIdx. For multivariate regression models (De'Ath, 2002), the independent variables are more than one. For example, the two organelle regression models contain all the possible combinations of two-organelle (as the independent variable) which shall be correlated with the overall cellular response (dependent variable). These models are iterated with all possible combinations of two and three numbers of independent variables. To evaluate the goodness of the novel drug response identification, the model has been implemented to predict data of a large (>1,000) number of new compounds

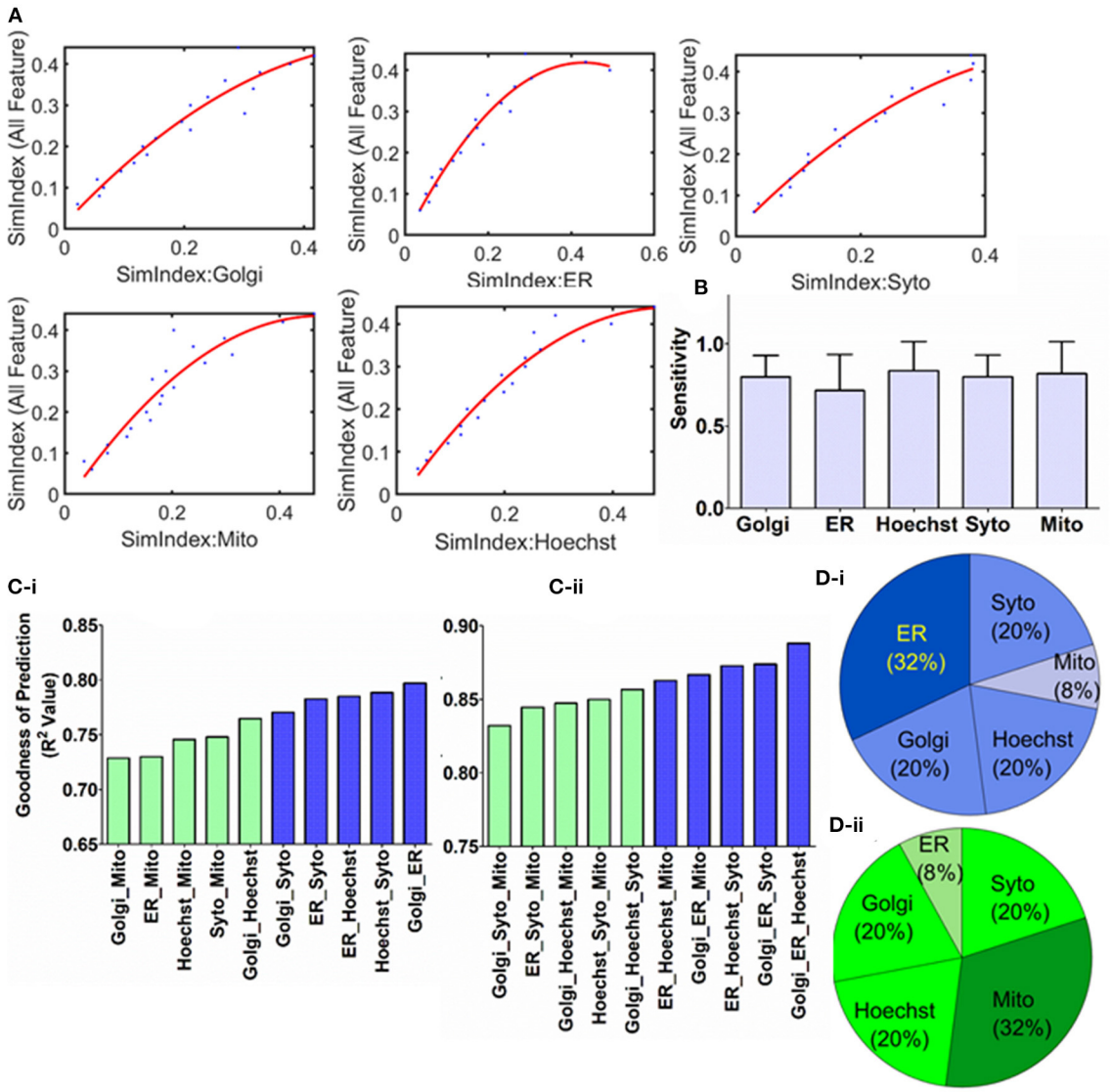

FIGURE 3 | Characterizing the correlation and regression modeling. **(A)** Quadratic polynomial fitted with the organelle response as the independent variable and cellular response as the dependent variable. **(B)** Sensitivity of the cellular response to the organelle response has been quantified through the first-order derivative of the polynomial fitted earlier. **(C)** Multivariate regression analysis is performed with organelle response as independent variable, while the cellular response is the dependent variable. Using this regression model, cellular response is predicted for over 1,000 new compounds. Based on the error between actual and estimated values goodness of prediction is quantified by using R squared values. These calculations are performed for all possible combinations of (i) two and (ii) three organelles as independent variable. Blue and green indicate the combinations that yielded the top 5 and bottom 5 goodness of prediction, respectively, in each case. **(D)** The occurrence of each organelle in (i) top 5 and (ii) bottom 5 models as per goodness of prediction is plotted as pie chart.

(**Supplementary Table 2**). Based on the predictions performed and the actual data, the error is computed by evaluating the R squared values. Overall R squared values calculated from the multivariate regression models have been described in **Figure 3C**. Although all possible combinations in the two and three variable regression models are valuable in making predictions, this accuracy ranking would benefit in understanding the salient organelle that contains signatures to facilitate the predictions. To address this, the organelle combinations that are present in the top 5 accurate models from each of two- and three-variable regression modeling (10 blue colored bars **Figure 3C**) were taken into consideration. Then the repetitions of each organelle were plotted in **Figure 3D**i. Out of the 10 cases (totaling $5 \times 3 + 5 \times 2 = 25$ instances of organelles), ER is featured in 8 (32% of 25 is 8) of those. A similar method is taken for the lower 5 models (10 green colored bars in **Figure 3C**). Then the repetitions of each organelle were plotted in **Figure 3D**ii. Out of these 10 cases (totaling 25 instances of organelles), Mito is featured in 8 (32% of 25 is 8) of those. Evaluated accuracies from the prediction model affirm ER features are pertinent for the prediction while

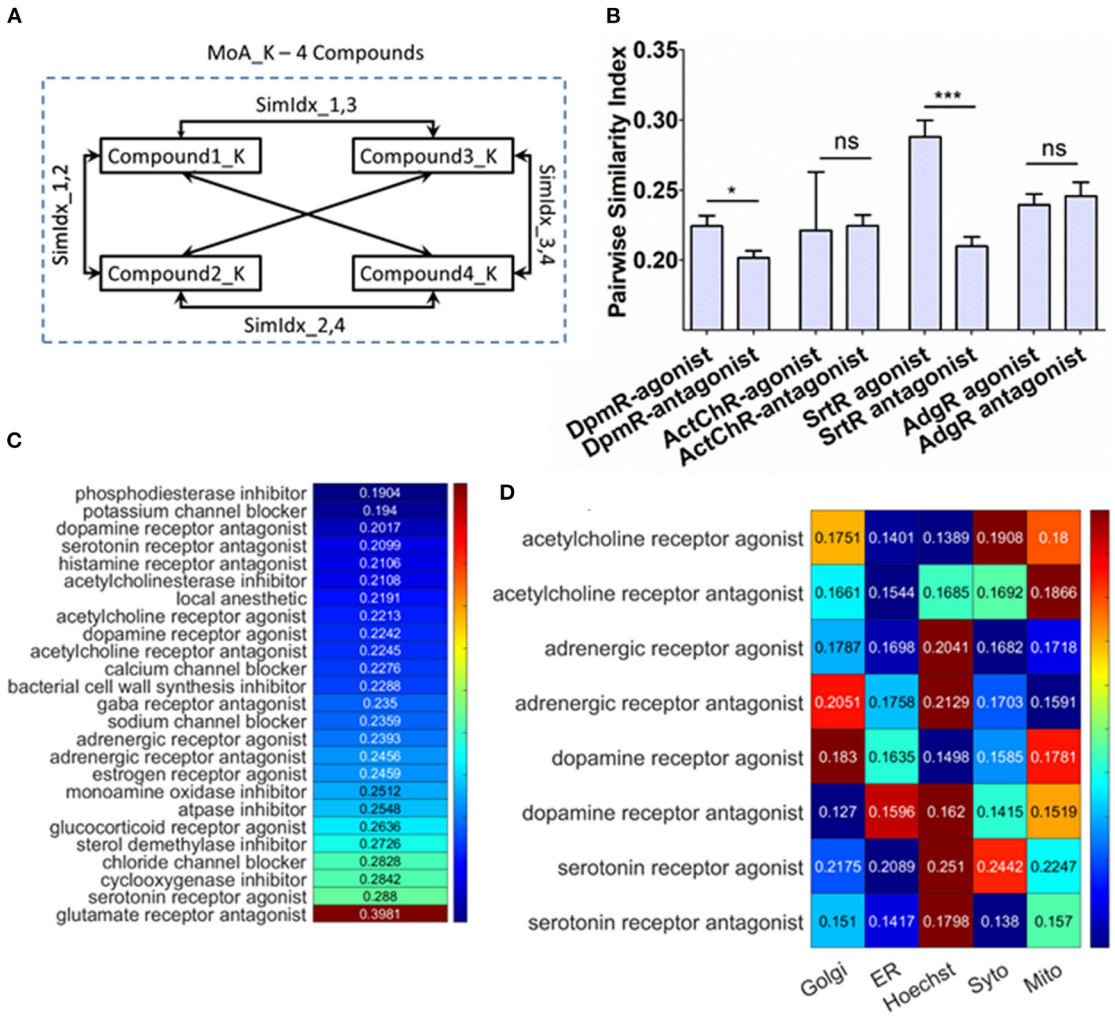

FIGURE 4 | Probing cellular response upon different compound treatment mediating conserved mechanism of action (MoA). **(A)** Schematic of the method to compute the pairwise similarity index between the compounds for the same MoA. **(B)** Pairwise similarity index for antagonist–agonist pair. **(C)** Pairwise similarity index for the MoA enlisted in the dataset, which contains more than three compounds. The colors are column scaled. **(D)** Pairwise similarity index of organelle features specifically for antagonist–agonist pair. The colors are row scaled.

mitochondria features are lower aptness here. Earlier ER is shown to be high ImpIdx or has more likelihood of being affected upon treatments and Mito's lower aptness (**Figure 2D**). Overall, along with studying the sensitivity of organelle-specific response, an efficient cellular response prediction model through multivariate regression is developed.

3.4. Heterogeneous Cellular Response Mediates Conserved MoA

The phenotypic response of cells has now been explored when cells are treated with different signaling modulation treatments. But it would also be interesting to examine the effect on cellular integrity upon treatments with different compounds that enact the same annotated MoA. Since single-cell resolved features can elucidate the heterogeneous response, which can also be used as a biological probe to identify the interactions between cellular machinery. To address this, from the working dataset MoAs were chosen, which were treated with more than three different compounds (**Supplementary Table 4**) and then pairwise similarity index (PSI) among the compounds is determined as shown in **Figure 4A**. To generate PSI, SimIdx is determined by checking fraction of similar features between two compounds and then iterated over all possible combinations of compounds (**Figure 4A**). PSI is similar to SimIdx but is generated by comparing phenotypic profiles between compounds instead of the compound with DMSO. This parameter captured whether the features affected upon these compound treatments are similar (High PSI) or not (Low PSI). If different compounds elicit a similar response that would signify conserved response pathways, which would be captured by higher PSI and vice versa. Interestingly, 24 MoAs are identified for which PSI is <0.3 (**Figure 4C**). In spite of having conserved MoA, these different

compounds exhibit heterogeneity regarding how each of these compounds affects the cellular integrity leading to the low PSI as reported. Thus, it is formulated that different compounds which enacts same outcome (referred here as MoA) might mediate through mechanistically different pathways which enables to evaluate signaling heterogeneity of these MoA cascades.

To probe this furthermore and specifically study how opposing signaling modulations affect the cellular response in terms of the profiled features, the available antagonist and agonist pairs—dopamine receptor (DpmR), acetylcholine receptor (ActChR), serotonine receptor (SrtR), and adregenic receptor (AdgR)—have been chosen. The PSI for these opposing signaling modulations is specifically represented in **Figure 4B**. The DpmR and SrtR agonist has significantly higher PSI, which might mean the agonist pathways are likely to be more conserved (as across compounds similar features are affected resulting in higher PSI) than respective antagonist ones. But, ActChR and AdgR antagonist–agonist pair shows similar PSI. The overall results show that agonist treatments have at least the same or higher PSI in comparison to their antagonist counterparts.

Subsequently, the effect on organelle integrity is determined by computing the PSI particularly on the organelle features. This indicates the similarity in organelle integrity downstream of the compound treatment (**Figure 4A**). If an organelle resembles a high PSI, then the compounds have induced similar changes for that organelle. Based on this calculation of the organelle PSI on the agonist–antagonist pair are shown in **Figure 4D**. This allows inferring that the organelle depicting higher PSI metric are more likely affected (since among the compound treatment this organelle features behaves similarly) upon the respective treatment. In the case of ActChR agonist–antagonist pair, Mitochondria and Syto (nucleolus) features rank as these organelles that get mostly affected through the compounds. Similarly, for AdgR agonist–antagonist pair, Hoechst (Nucleus) and Golgi features are mostly affected through these signaling modulation treatments. Also for the SrtR agonist–antagonist pair, Hoechst (nucleus) is most likely to be affected in both cases followed by Syto (nucleolus) and mitochondria, respectively. In contrast, for DpR the trend reverses. In the case of agonist, Golgi and mitochondria are most likely to be targeted. However, the antagonist treatment targets are different—Hoechst (nucleus) and ER. This observation helps characterize the organelle targets for MoA treatments. For ActChR, AdgR, SrtR agonist–antagonist pair, there is a close resemblance in the most impacted organelle. Since the dopamine receptor affects different targets downstream, this establishes valuable insights regarding signaling plasticity in cancer cells as activation or inactivation of cascades are mediated through different targets. Overall, the quantification helped identify the same MoA inducing treatment could have different downstream targets which hint toward signaling heterogeneity.

4. DISCUSSION

In biomedical applications, it is often important to understand the signatures that chemical perturbation imprints on the cell. Quantitative analysis of fluorescent microscopy enables identification of nascent signatures of perturbation (Rohban et al., 2017) as well as health phenotypes (Way et al., 2020). This work is aimed to identify such underlying cellular response signatures by using a publicly available dataset of high content screen. The departure of phenotypic profiles as compared to DMSO as a reference has provided insights regarding the cellular changes induced. The derived metric of SimIdx (as presented in **Figures 1A,Bi**) from a population of cells is based on the phenotypic impact the signaling modulation treatment causes. SimIdx accredits understanding of phenotypic relationships present in the dataset. This simple yet powerful documentation on diverse data can advance detection of the onset of diseases by labeling the signatures in advance from know datasets.

The subtle changes induced upon treatment are tracked in this study for monitoring phenotypic variations specifically in terms of the organelle. Identification of specific organelle targets could help to target drugs to organelles of maximum relevance. Such a target-directed drug design is critical for maximizing the therapeutic outcome of the drug (Torchilin, 2012). These profiles across various cells aid the identification of novel underlying signatures of organelle-cellular response coupling. Furthermore, the sensitivity analysis of the organelle response (**Figure 3B**) has shown no particular organelle bias, which could be a result of the transfer of impact from one target to another. The overall cellular behavior is dictated by the rich underlying interacting signaling network. However, hyper-activation (Sever and Brugge, 2015) of signaling cascades is also observed in cancer cells. Hence it is likely that the impact of the treatment on some organelle targets might eventually be relayed onto other organelles (Valm et al., 2017; Cohen et al., 2018). This computational study convenes evidence for signaling hyperactivation, which resonates with the literature hence adds to the validation. These approaches on time-lapse microscopy shall resolve these signatures of cellular response in the temporal domain which enables to probe how the underlying connections evolve with time and develop an organelle interactome.

Based on the deterministic response curves, prediction models have been developed to estimate the overall cellular response by using only specific organelle response features. These models were engineered based on multivariate regression, which is extensively used in engineering analytics (Dumouchel and O'Brien, 1989; Prats-Montalbán et al., 2011). The impact of the treatment on the overall cell is then efficiently predicted based on only fewer organelle stains. For validation, the prediction accuracy is measured on a thousand new candidate compounds (**Figure 3C**). This prediction ability open avenues to stain cells with a lesser number of fluorescent labels, yet efficiently determine the overall cellular response (**Figure 3C**) through a simplistic and lesser resource-intensive method. This study also characterizes how ER serves two very critical roles in mediating the cellular response. First, a fraction (80%) of the ER features are affected in at least 240 (out of 270) MoA, making it the most pertinent target organelle (**Figure 2D**) among the ones tested here. Second, ER also acts as a key organelle (**Figure 3Di**) in the cell response prediction models. It is known in the literature that ER is also pivotal for cellular

homeostasis and extracellular response (Xu et al., 2005; Cao and Kaufman, 2014). Additionally, recent studies have also shown that in cancer ER organelle is stressed and associated signaling pathways are often dysregulated (Yadav et al., 2014; Kato and Nishitoh, 2015; Han and Wan, 2018; Lin et al., 2019). This hints that the ER response is likely to be strongly coupled to the cellular response. Hence, the computational findings in this study align with the earlier reported evidence. Overall, such analysis has paved the way to trace rudimentary trends among organelles.

The PSI, another metric characterized in this study, analysis is applied to examine the differential effect on cellular integrity for the same annotated MoA. Here, the response variability itself has directly been used as a biological probe to access information regarding the functional specificity of these molecular mechanisms. If different compounds elicit a similar response that would signify conserved response pathways, which would be captured by higher PSI and vice versa. This calculation has suitably equipped the study to show that different compound treatments cause differential cellular response yet enacts the conserved final MoA (**Figure 4C**). Interestingly, this analysis shows how signaling heterogeneity arises by assessing differential impact on the cell caused by similar treatment. For further comparison of cellular response, selective studying of the agonist–antagonist pair has been performed (**Figure 4B**). This metric has also helped calibrate the trend of organelle (**Figure 4D**) being affected and gain signaling insights. An understanding regarding the organelle targets for the treatments, which can be beneficial for studying drug targets and their effect. The role of dopamine in mediating neuro-synaptic plasticity is already established (Tecuapetla et al., 2007; Ishikawa et al., 2013; Langlois et al., 2018). Dopamine is also useful in cancer treatment as it results in the shrinking of tumor size (Liu et al., 2019) and inhibiting its progression and exerts anticancer effect (Sarkar et al., 2008; Zhang et al., 2017; Kline et al., 2018). Here at single cell level the interaction between dopamine activation and inactivation with cancer is studied. This revealed that the downstream target switches, which could be a result of rewiring in underlying cascades. Hence, this serves as an elementary evidence for signaling plasticity in cancer cells. Further experimental characterization of this plasticity might reveal the machinery involved as well as advance its role in anticancer therapeutics.

Moreover, with the advent of automation in the cell-painting assay, the screen can be substantially increased enabling to integrate these methods to characterize the downstream effect of a larger number of bioactive compounds. The methods developed here enables integration of high content complex data for studying phenotypic responses and cellular signaling. The report shows how quantitative analysis on cellular imaging screens could be used to derive mechanistic evidence regarding cellular signaling and associated activation, heterogeneity, and plasticity. Identification of these characteristics of molecule treatment will not only enhance understanding of cellular function but also can be applied to transitional research to validate drug and therapeutic effects. This shall also benefit drug discovery and personalized medicine by analyzing subtle changes in the effect of diverse molecules.

In summary, taking advantage of the individual-cell measurements in the high content screen, the cellular phenotypic response has been probed. Subsequently, these facilitated the understanding of varying responses in the downstream effect for multiple treatments on cancer cells, specifically the organelle targets, predicting the overall cellular response efficiently for new candidate molecules and finally evaluate the signaling heterogeneity. Since specifics of the treatment would be identified, this will envisage the identification of hallmarks of both molecular as well as disease targets in cells and open promising avenues through interdisciplinary investigation and quantitative models.

DATA AVAILABILITY STATEMENT

The dataset used in this study is available from the Broad Bioimage Benchmark Collection public repository. The original contributions presented in the study are included in the article. For additional resources please refer to https://github.com/sayan08/HCA_Target_Response.

AUTHOR CONTRIBUTIONS

SB conceptualized and designed the project, developed the computational framework, analyzed the data, interpreted the result, and wrote the manuscript.

ACKNOWLEDGMENTS

The author is thankful to Dr. Akash Gulyani, all the lab members and University of Hyderabad, Hyderabad. The author would also like to thank Institute for Stem Cell Science and Regenerative Medicine, Bangalore for past employment, Dr. Nishan Shettigar (InStem), Anirudh Chakravarthy (InStem), Souradeep Sarkar (NCBS), and Shefali (Jadavpur University) for helpful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2020.594750/full#supplementary-material

Supplementary Table 1 | List of all MoA used in the study.

Supplementary Table 2 | List of the of non-annotated compound BroadIDs.

Supplementary Table 3 | List of the profiled features.

Supplementary Table 4 | List of MoA where more than three compounds used.

Supplementary Table 5 | An example of the dataset (100 × 824). Data from 100 randomly chosen cells. These cells are treated with a compound annotated as adrenergic receptor agonist inducing MoA.

Supplementary Table 6 | List of equations used for regression.

REFERENCES

Bray, M.-A., Gustafsdottir, S. M., Rohban, M. H., Singh, S., Ljosa, V., Sokolnicki, K. L., et al. (2017). A dataset of images and morphological profiles of 30,000 small-molecule treatments using the cell painting assay. *Gigascience* 6:giw014. doi: 10.1093/gigascience/giw014

Bray, M.-A., Singh, S., Han, H., Davis, C. T., Borgeson, B., Hartland, C., et al. (2016). Cell painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* 11:1757. doi: 10.1038/nprot.2016.105

Breinig, M., Klein, F. A., Huber, W., and Boutros, M. (2015). A chemical-genetic interaction map of small molecules using high-throughput imaging in cancer cells. *Mol. Syst. Biol.* 11:846. doi: 10.15252/msb.20156400

Butcher, E. (2005). "Can cell systems biology rescue drug discovery?" in *Nature Reviews Drug Discovery 4.6*, 461–467.

Caie, P. D., Walls, R. E., Ingleston-Orme, A., Daya, S., Houslay, T., Eagle, R., et al. (2010). High-content phenotypic profiling of drug response signatures across distinct cancer cells. *Mol. Cancer Therapeut.* 9, 1913–1926. doi: 10.1158/1535-7163.MCT-09-1148

Cao, S. S., and Kaufman, R. J. (2014). Endoplasmic reticulum stress and oxidative stress in cell fate decision and human disease. *Antioxid. Redox Signal.* 21, 396–413. doi: 10.1089/ars.2014.5851

Carpenter, A. E. (2007). Image-based chemical screening. *Nat. Chem. Biol.* 3, 461–465. doi: 10.1038/nchembio.2007.15

Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., et al. (2006). Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* 7:R100. doi: 10.1186/gb-2006-7-10-r100

Chen, X., Yan, C. C., Zhang, X., Zhang, X., Dai, F., Yin, J., et al. (2016). Drug-target interaction prediction: databases, web servers and computational models. *Brief. Bioinform.* 17, 696–712. doi: 10.1093/bib/bbv066

Cohen, S., Valm, A. M., and Lippincott-Schwartz, J. (2018). Interacting organelles. *Curr. Opin. Cell Biol.* 53, 84–91. doi: 10.1016/j.ceb.2018.06.003

Corsello, S. M., Bittker, J. A., Liu, Z., Gould, J., McCarren, P., Hirschman, J. E., et al. (2017). The drug repurposing hub: a next-generation drug library and information resource. *Nat. Med.* 23, 405–408. doi: 10.1038/nm.4306

De'Ath, G. (2002). Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology* 83, 1105–1117. doi: 10.1890/0012-9658(2002)083[1105:MRTANT]2.0.CO;2

Dumouchel, W., and O'Brien, F. (1989). "Integrating a robust option into a multiple regression computing environment," in *Computer Science and Statistics: Proceedings of the 21st Symposium on the Interface* (Alexandria, VA: American Statistical Association), 297–302.

Feng, Y., Mitchison, T. J., Bender, A., Young, D. W., and Tallarico, J. A. (2009). Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. *Nat. Rev. Drug Discov.* 8, 567–578. doi: 10.1038/nrd2876

Futamura, Y., Muroi, M., and Osada, H. (2013). Target identification of small molecules based on chemical biology approaches. *Mol. Biosyst.* 9, 897–914. doi: 10.1039/c2mb25468a

Galvao, J., Davis, B., Tilley, M., Normando, E., Duchen, M. R., and Cordeiro, M. F. (2014). Unexpected low-dose toxicity of the universal solvent DMSO. *FASEB J.* 28, 1317–1330. doi: 10.1096/fj.13-235440

Gustafsdottir, S. M., Ljosa, V., Sokolnicki, K. L., Wilson, J. A., Walpita, D., Kemp, M. M., et al. (2013). Multiplex cytological profiling assay to measure diverse cellular states. *PLoS ONE* 8:e80999. doi: 10.1371/journal.pone.0080999

Han, C. C., and Wan, F. S. (2018). New insights into the role of endoplasmic reticulum stress in breast cancer metastasis. *J. Breast Cancer* 21, 354–362. doi: 10.4048/jbc.2018.21.e51

Holland, P. W., and Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Commun. Stat. Theory Methods* 6, 813–827. doi: 10.1080/03610927708827533

Hughes, C. S., McConechy, M. K., Cochrane, D. R., Nazeran, T., Karnezis, A. N., Huntsman, D. G., et al. (2016). Quantitative profiling of single formalin fixed tumour sections: proteomics for translational research. *Sci. Rep.* 6:34949. doi: 10.1038/srep34949

Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., et al. (2000). Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126. doi: 10.1016/S0092-8674(00)00015-5

Ishikawa, M., Otaka, M., Huang, Y. H., Neumann, P. A., Winters, B. D., Grace, A. A., et al. (2013). Dopamine triggers heterosynaptic plasticity. *J. Neurosci.* 33, 6759–6765. doi: 10.1523/JNEUROSCI.4694-12.2013

Kandaswamy, C., Silva, L. M., Alexandre, L. A., and Santos, J. M. (2016). High-content analysis of breast cancer using single-cell deep transfer learning. *J. Biomol. Screen.* 21, 252–259. doi: 10.1177/1087057115623451

Kato, H., and Nishitoh, H. (2015). Stress responses from the endoplasmic reticulum in cancer. *Front. Oncol.* 5:93. doi: 10.3389/fonc.2015.00093

Kitano, H. (2002). Computational systems biology. *Nature* 420, 206–210. doi: 10.1038/nature01254

Kline, C. L. B., Ralff, M. D., Lulla, A. R., Wagner, J. M., Abbosh, P. H., Dicker, D. T., et al. (2018). Role of dopamine receptors in the anticancer activity of onc201. *Neoplasia* 20, 80–91. doi: 10.1016/j.neo.2017.10.002

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., et al. (2006). The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935. doi: 10.1126/science.1132939

Lang, P., Yeow, K., Nichols, A., and Scheer, A. (2006). Cellular imaging in drug discovery. *Nat. Rev. Drug Discov.* 5, 343–356. doi: 10.1038/nrd2008

Langlois, L. D., Dacher, M., and Nugent, F. S. (2018). Dopamine receptor activation is required for gabaergic spike timing-dependent plasticity in response to complex spike pairing in the ventral tegmental area. *Front. Synaptic Neurosci.* 10:32. doi: 10.3389/fnsyn.2018.00032

Laurikkala, J., Juhola, M., Kentala, E., Lavrac, N., Miksch, S., and Kavsek, B. (2000). "Informal identification of outliers in medical data," in *Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology* (Berlin), 20–24.

Lee, J. A., Uhlik, M. T., Moxham, C. M., Tomandl, D., and Sall, D. J. (2012). Modern phenotypic drug discovery is a viable, neoclassic pharma strategy. *J. Med. Chem.* 55, 4527–4538. doi: 10.1021/jm201649s

Lin, Y., Jiang, M., Chen, W., Zhao, T., and Wei, Y. (2019). Cancer and ER stress: mutual crosstalk between autophagy, oxidative stress and inflammatory response. *Biomed. Pharmacother.* 118:109249. doi: 10.1016/j.biopha.2019.109249

Liu, X., Tang, C., Wen, G., Zhong, C., Yang, J., Zhu, J., et al. (2019). The mechanism and pathways of dopamine and dopamine agonists in prolactinomas. *Front. Endocrinol.* 9:768. doi: 10.3389/fendo.2018.00768

Ljosa, V., Caie, P. D., Ter Horst, R., Sokolnicki, K. L., Jenkins, E. L., Daya, S., et al. (2013). Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. *J. Biomol. Screen.* 18, 1321–1329. doi: 10.1177/1087057113503553

Ljosa, V., Sokolnicki, K. L., and Carpenter, A. E. (2012). Annotated high-throughput microscopy image sets for validation. *Nat. Methods* 9, 637–637. doi: 10.1038/nmeth.2083

Loo, L.-H., Wu, L. F., and Altschuler, S. J. (2007). Image-based multivariate profiling of drug responses from single cells. *Nat. Methods* 4, 445–453. doi: 10.1038/nmeth1032

Low, J., Stancato, L., Lee, J., and Sutherland, J. J. (2008). Prioritizing hits from phenotypic high-content screens. *Curr. Opin. Drug Discov. Dev.* 11, 338–345.

Macarron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T., et al. (2011). Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* 10, 188–195. doi: 10.1038/nrd3368

Madhukar, N. S., Khade, P. K., Huang, L., Gayvert, K., Galletti, G., Stogniew, M., et al. (2019). A bayesian machine learning approach for drug target identification using diverse data types. *Nat. Commun.* 10, 1–14. doi: 10.1038/s41467-019-12928-6

Moffat, J. G., Rudolph, J., and Bailey, D. (2014). Phenotypic screening in cancer drug discovery–past, present and future. *Nat. Rev. Drug Discov.* 13, 588–602. doi: 10.1038/nrd4366

Moffat, J. G., Vincent, F., Lee, J. A., Eder, J., and Prunotto, M. (2017). Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat. Rev. Drug Discov.* 16, 531–543. doi: 10.1038/nrd.2017.111

Nyffeler, J., Willis, C., Lougee, R., Richard, A., Paul-Friedman, K., and Harrill, J. A. (2020). Bioactivity screening of environmental chemicals using imaging-based

high-throughput phenotypic profiling. *Toxicol. Appl. Pharmacol.* 389:114876. doi: 10.1016/j.taap.2019.114876

Perlman, L., Gottlieb, A., Atias, N., Ruppin, E., and Sharan, R. (2011). Combining drug and gene similarity measures for drug-target elucidation. *J. Comput. Biol.* 18, 133–145. doi: 10.1089/cmb.2010.0213

Prats-Montalbán, J. M., De Juan, A., and Ferrer, A. (2011). Multivariate image analysis: a review with applications. *Chemometr. Intell. Lab. Syst.* 107, 1–23. doi: 10.1016/j.chemolab.2011.03.002

Rohban, M. H., Singh, S., Wu, X., Berthet, J. B., Bray, M.-A., Shrestha, Y., et al. (2017). Systematic morphological profiling of human gene and allele function via cell painting. *eLife* 6:e24060. doi: 10.7554/eLife.24060

Sarkar, C., Chakroborty, D., Chowdhury, U. R., Dasgupta, P. S., and Basu, S. (2008). Dopamine increases the efficacy of anticancer drugs in breast and colon cancer preclinical models. *Clin. Cancer Res.* 14, 2502–2510. doi: 10.1158/1078-0432.CCR-07-1778

Sever, R., and Brugge, J. S. (2015). Signal transduction in cancer. *Cold Spring Harbor Perspect. Med.* 5:a006098. doi: 10.1101/cshperspect.a006098

Simm, J., Klambauer, G., Arany, A., Steijaert, M., Wegner, J. K., Gustin, E., et al. (2018). Repurposing high-throughput image assays enables biological activity prediction for drug discovery. *Cell Chem. Biol.* 25, 611–618. doi: 10.1016/j.chembiol.2018.01.015

Steigele, S., Siegismund, D., Fassler, M., Kustec, M., Kappler, B., Hasaka, T., et al. (2020). Deep learning-based HCS image analysis for the enterprise. *SLAS Discov.* 25, 812–821. doi: 10.1177/2472555220918837

Szalai, B., Subramanian, V., Holland, C. H., Alföldi, R., Puskás, L. G., and Saez-Rodriguez, J. (2019). Signatures of cell death and proliferation in perturbation transcriptomics data–from confounding factor to effective prediction. *Nucleic Acids Res.* 47, 10010–10026. doi: 10.1093/nar/gkz805

Tanaka, M., Bateman, R., Rauh, D., Vaisberg, E., Ramachandani, S., Zhang, C., et al. (2005). An unbiased cell morphology-based screen for new, biologically active small molecules. *PLoS Biol.* 3:e128. doi: 10.1371/journal.pbio.0030128

Tecuapetla, F., Carrillo-Reid, L., Bargas, J., and Galarraga, E. (2007). Dopaminergic modulation of short-term synaptic plasticity at striatal inhibitory synapses. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10258–10263. doi: 10.1073/pnas.0703813104

Torchilin, V. P. (2012). Next step in drug delivery: getting to individual organelles. *Drug Deliv. Transl. Res.* 2, 415–417. doi: 10.1007/s13346-012-0102-2

Valm, A. M., Cohen, S., Legant, W. R., Melunis, J., Hershberg, U., Wait, E., et al. (2017). Applying systems-level spectral imaging and analysis to reveal the organelle interactome. *Nature* 546, 162–167. doi: 10.1038/nature22369

Wagner, B. K., and Clemons, P. A. (2009). Connecting synthetic chemistry decisions to cell and genome biology using small-molecule phenotypic profiling. *Curr. Opin. Chem. Biol.* 13, 539–548. doi: 10.1016/j.cbpa.2009.09.018

Wawer, M. J., Li, K., Gustafsdottir, S. M., Ljosa, V., Bodycombe, N. E., Marton, M. A., et al. (2014). Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. *Proc. Natl. Acad. Sci. U.S.A.* 111, 10911–10916. doi: 10.1073/pnas.1410933111

Way, G. P., Kost-Alimova, M., Shibue, T., Harrington, W. F., Gill, S., Piccioni, F., et al. (2020). Predicting cell health phenotypes using image-based morphology profiling. *bioRxiv [preprint]*. doi: 10.1101/2020.07.08.193938

Xu, C., Bailly-Maitre, B., Reed, J. C., et al. (2005). Endoplasmic reticulum stress: cell life and death decisions. *J. Clin. Investig.* 115, 2656–2664. doi: 10.1172/JCI 26373

Yadav, R. K., Chae, S.-W., Kim, H.-R., and Chae, H. J. (2014). Endoplasmic reticulum stress and cancer. *J. Cancer Prevent.* 19:75. doi: 10.15430/JCP.2014.19.2.75

Young, D. W., Bender, A., Hoyt, J., McWhinnie, E., Chirn, G.-W., Tao, C. Y., et al. (2008). Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat. Chem. Biol.* 4, 59–68. doi: 10.1038/nchembio.2007.53

Zhang, X., Liu, Q., Liao, Q., and Zhao, Y. (2017). Potential roles of peripheral dopamine in tumor immunity. *J. Cancer* 8:2966. doi: 10.7150/jca.20850

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 January 2021
doi: 10.3389/fgene.2020.604585

Inference of Intercellular Communications and Multilayer Gene-Regulations of Epithelial–Mesenchymal Transition From Single-Cell Transcriptomic Data

Yutong Sha[1,2], Shuxiong Wang[1], Federico Bocci[1,2], Peijie Zhou[1] and Qing Nie[1,2,3]**

[1] Department of Mathematics, University of California, Irvine, Irvine, CA, United States, [2] The NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, Irvine, CA, United States, [3] Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA, United States

OPEN ACCESS

Edited by:
Ankur Sharma, Genome Institute of Singapore, Singapore

Reviewed by:
Weikang Wang, University of Pittsburgh, United States
David Jordan Wooten, Pennsylvania State University (PSU), United States
Francesc Font Clos, University of Milan, Italy

***Correspondence:**
Peijie Zhou
peijiez1@uci.edu
Qing Nie
qnie@uci.edu

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *09 September 2020*
Accepted: *02 December 2020*
Published: *08 January 2021*

Citation:
Sha Y, Wang S, Bocci F, Zhou P and Nie Q (2021) Inference of Intercellular Communications and Multilayer Gene-Regulations of Epithelial–Mesenchymal Transition From Single-Cell Transcriptomic Data. Front. Genet. 11:604585. doi: 10.3389/fgene.2020.604585

Epithelial-to-mesenchymal transition (EMT) plays an important role in many biological processes during development and cancer. The advent of single-cell transcriptome sequencing techniques allows the dissection of dynamical details underlying EMT with unprecedented resolution. Despite several single-cell data analysis on EMT, how cell communicates and regulates dynamics along the EMT trajectory remains elusive. Using single-cell transcriptomic datasets, here we infer the cell–cell communications and the multilayer gene–gene regulation networks to analyze and visualize the complex cellular crosstalk and the underlying gene regulatory dynamics along EMT. Combining with trajectory analysis, our approach reveals the existence of multiple intermediate cell states (ICSs) with hybrid epithelial and mesenchymal features. Analyses on the time-series datasets from cancer cell lines with different inducing factors show that the induced EMTs are context-specific: the EMT induced by transforming growth factor B1 (TGFB1) is synchronous, whereas the EMTs induced by epidermal growth factor and tumor necrosis factor are asynchronous, and the responses of TGF-β pathway in terms of gene expression regulations are heterogeneous under different treatments or among various cell states. Meanwhile, network topology analysis suggests that the ICSs during EMT serve as the signaling in cellular communication under different conditions. Interestingly, our analysis of a mouse skin squamous cell carcinoma dataset also suggests regardless of the significant discrepancy in concrete genes between *in vitro* and *in vivo* EMT systems, the ICSs play dominant role in the TGF-β signaling crosstalk. Overall, our approach reveals the multiscale mechanisms coupling cell–cell communications and gene–gene regulations responsible for complex cell-state transitions.

Keywords: single-cell RNA sequencing, trajectory inference, gene regulatory network, cell fate decision, cell–cell communication, multi-scale analysis

INTRODUCTION

Epithelial-to-mesenchymal transition (EMT) is a biological process where epithelial cells lose cell–cell adhesion and gain some mesenchymal traits of migration and invasion (Kalluri and Weinberg, 2009; Jolly et al., 2018). EMT not only occurs widely during normal embryonic development, organ fibrosis, and wound healing, but also plays an important role in tumor progression with metastasis (Nieto et al., 2016; Lambert et al., 2017).

Recent studies have underscored that EMT is not a binary process, but instead exists on a spectrum with various hybrid states ranging from epithelial-to-mesenchymal phenotypes (Nieto et al., 2016). Cells undergoing EMT can display mixed epithelial and mesenchymal features and are considered in the intermediate cell states (ICSs; Jolly et al., 2015; Sha et al., 2019; Jia D. et al., 2019). In the context of cancer progression, these ICSs have been proposed as the main drivers of metastasis because of their ability to undergo collective cell migration as highly metastatic multicellular clusters (Jolly et al., 2015). Therefore, understanding the features and role of ICSs during EMT could potentially unlock novel clinical strategies. With the unprecedented opportunities brought by single-cell RNA sequencing (scRNA-seq), the existence of multiple ICSs and their transcriptomic profiles has been observed and analyzed via pseudotemporal ordering or energy landscapes (Qiu et al., 2017; Jin et al., 2018; Li and Balazsi, 2018; Pastushenko et al., 2018; An et al., 2019; Chen et al., 2019). Very recently, specially designed methods have also been proposed to infer EMT trajectories or transition paths from the single-cell transcriptomic (Sha et al., 2020) or imaging data (Wang W. et al., 2020). The integrative analysis combining unsupervised learning of single-cell transcriptomic data and computational modeling of EMT in cancer and embryogenesis successfully uncovered the novel roles of ICSs on adaption, noise attenuation, and transition efficiency (Sha et al., 2020). While these methods have provided insights into the dynamics of EMT from a single-cell perspective, the role of intercellular communication in EMT remains largely unknown.

Indeed, EMT is not necessarily a cell autonomous process. Cells secrete and in turn respond to various growth and differentiation signaling factors secreted by other cells in the extracellular environment, including transforming growth factor β (TGF-β), WNT, and Notch proteins (Moustakas and Heldin, 2007; Xu et al., 2009; Boareto et al., 2016; Bocci et al., 2018). Among them, the well-characterized TGF-β pathway has received much attention as a major inducer of EMT during embryogenesis, cancer progression, and fibrosis (Wendt et al., 2009; Xu et al., 2009). The TGF-β pathway can also crosstalk with other pathways such as WNT and SHH (Zhang et al., 2016), forming the complex response of signaling. In addition, signaling in cell–cell communications has also been found important in the formation and regulation of ICSs (e.g., through Notch pathway; Bocci et al., 2020). This intercellular communication has been shown to play significant roles in regulating gene expression dynamics within individual cells, through analysis of scRNA-seq datasets from several development and cancer systems (Camp et al., 2017; Puram et al., 2017; Zepp et al., 2017; Kumar et al., 2018; Wang S. et al., 2020). Computational methods have been developed to infer cell–cell communication networks based on ligand–receptor interactions (Wang S. et al., 2019; Wang Y. et al., 2019; Cabello-Aguilar et al., 2020; Jin et al., 2020) and elucidate how cell–cell communications propagate to downstream target genes through transcription factors (Browaeys et al., 2020). While methods have been developed to infer EMT gene regulatory network (GRN) from RNA-seq single-cell data (Ramirez et al., 2020), the role of cell–cell communications on gene regulation dynamics along EMT trajectory is poorly understood.

Through both experimental and mathematical modeling studies, the key circuits of EMT involving few epithelial/mesenchymal markers, transcription factors, and signaling molecules have been summarized (Hong et al., 2015; Li et al., 2016; Fazilaty et al., 2019; Kang et al., 2019; Xing and Tian, 2019; Tripathi et al., 2020; Yang et al., 2020). Because of different roles of nodes, the circuits can be modeled as a multilayer network (Kivelä et al., 2014) with hierarchical structures (Browaeys et al., 2020). In the multilayer network, cells communicate with each other and the environment via signal transduction pathways (Layer 1), which directly targets the downstream factors or genes (Layer 2), that subsequently regulate the expression of marker genes of various cell states (Layer 3). In addition, there may be dynamical changes of network structure during EMT, where the temporal (or pseudotemporal) information constitutes another independent dimension of the layer sets. The complex interactions among nodes may exist within the same layers or across different layers, in controlling EMT.

Here we study the time-series scRNA-seq datasets of OVCA420 cancer cell line exposed to various EMT-inducing factors (Cook and Vanderhyden, 2020). We first delineate the underlying transition details at individual cell resolution with a recently developed method, QuanTC. For the cancer cell lines undergoing EMT under three different treatments, we quantify the ICS-regulated trajectories and detect the driver genes in EMT for each case, respectively. While cells undergo TGFb1-driven EMT in a highly synchronized fashion, EMT guided by epidermal growth factor (EGF) and tumor necrosis factor (TNF) is asynchronous. Next, we develop a multilayer network approach to infer and visualize the hierarchical interactions that combine cell–cell communications through the TGF-β pathway, signal transductions, and GRNs from single-cell transcriptomic data. After trajectory inference, we then utilize the multilayer network approach to decipher the role of TGF-β pathway in regulating EMT dynamics with different inducing factors. We also compare the results of *in vitro* cancer cell lines with further analysis of *in vivo* mouse skin squamous cell carcinoma (SCC) dataset (Pastushenko et al., 2018).

RESULTS

Synchronous EMT With Two ICSs Induced by TGFB1

We analyzed the published datasets (Cook and Vanderhyden, 2020) with ovarian OVCA420 cancer cell line capable of undergoing EMT. This cell line, which normally shows an epithelial morphology, was exposed to known EMT-inducing factors: TGFB1, EGF, and TNF, respectively, to promote EMT. We used the samples collected at five distinct time points from day 0 to day 7 after the treatment.

To compare the process of EMT under three treatments, we used QuanTC (Sha et al., 2020) to perform the clustering and transition trajectory reconstruction. QuanTC estimates the optimal number of clusters by analyzing the sorted eigenvalues of symmetric normalized graph Laplacian (**Supplementary Figure 1A**). Four clusters were identified in EMT induced by TGFB1 (**Figure 1A**). A first cluster (C3) was mostly composed by cell subpopulations collected at day 0 and 8 h after induction (**Figure 1B**) and expressed relatively high levels of epithelial markers CDH1 (**Supplementary Figure 1B**). Conversely, a second cluster (C2) consisted of cells collected at days 3 and 7 (**Figures 1A,B**) and expressed relatively high levels of mesenchymal markers FN1 and SNAI2 (**Supplementary Figure 1C**). Furthermore, cells in these clusters had a low Cell Plasticity Index (CPI). CPI employs an entropy-based approach to estimate cell plasticity, so that a higher index implies a higher probability of transition between clusters (see section "*Materials and Methods*"). Based on the CPI values, QuanTC predicted that clusters C2 and C3 have lower percentages of transition cells (TCs; **Figures 1C,D**), thus suggesting that they are the beginning or end of the trajectory. Based on these observations, we identified cluster C3 as the E state and cluster C2 as the M state.

After choosing the E state, C3, as the beginning of the transition, QuanTC computed the most probable transition trajectory, C3–C4–C1–C2, consisting of 67% of the total cell population (**Figure 1E**). The cluster C4 and C1 were thus identified as ICSs I1 and I2, respectively. The marker genes of each state and the transition genes marking the transition between states along the transition trajectory were inferred by QuanTC (**Supplementary Figure 1D**). To characterize the two ICSs, I1 and I2, we performed a Gene Ontology (GO) biological processes analysis (The Gene and Ontology Consortium, 2019) of the top 50 marker genes of each state (**Supplementary Figure 1E**). Both ICSs shared similar biological processes including signaling and localization. Furthermore, I2 also related to adhesion and locomotion. This suggested that the cells in ICSs displayed both epithelial and mesenchymal features and communications with other cells through cell signaling.

Finally, we inspected the population dynamics during TGFB1-driven EMT by considering the pseudotime distribution. Pseudotime quantifies the position of a given cell along the transition trajectory predicted by QuanTC and therefore does not necessarily correlate with the experiment's physical time. In this time series, however, most cells at $t = 0$ days were characterized by a low pseudotime (i.e., they were positioned toward the beginning of the transition trajectory), whereas cells at later time points exhibited progressively higher pseudotime values (**Figure 1F**). In other words, OVCA420 cells started from the E state and progressively transitioned throughout the 7 days of EMT induced by TGFB1 in a nearly synchronous fashion.

Asynchronous EMT Induced by EGF and TNF

Applying QuanTC to the OVCA420 dataset where EMT was induced by EGF, four clusters were also identified based on the biggest eigenvalue gap after the first two eigenvalues because we want to investigate the ICSs during EMT (**Supplementary Figure 2A** and **Figure 2A**). Differently from TGFB1-driven EMT, however, cells collected at different time points colocalized within the same clusters, and no group of cells at any given time point dominated any cluster (**Figure 2B**). Based on the CPI values, the two clusters (C2 and C3) were considered as the E and M states based on the fewer TCs around them (**Figures 2C,D**). Specifically, C2 was then identified as the E state according to the relatively high expression levels of epithelial markers CDH1 (**Supplementary Figure 2B**), and C3 was identified as the M state because of higher expressions of mesenchymal markers FOXC2 and SNAI2 (**Supplementary Figure 2C**).

The most probable transition trajectory was inferred after choosing cluster C2 as the starting state (**Figure 2E**). The two remaining clusters (C1 and C4) between E and M along the transition trajectory had more TCs around them and were identified as I1 and I2, respectively. According to the GO analysis of the top marker genes (**Supplementary Figure 2D**), the I2 state displayed biological processes including adhesion, locomotion, and signaling, showing mixed feature of both epithelial and mesenchymal cells (**Supplementary Figure 2E**).

The average pseudotime values slightly increased along collection time points, hence demonstrating that the EGF stimulus induces an EMT response. Compared to TGFB1-driven EMT, however, pseudotime distribution within each time point had a high variance, thus indicating that the EMT induced by EGF was more asynchronous (**Figure 2F**).

We applied a similar analysis to EMT induced by TNF and also identified four clusters with two ICSs (**Supplementary Figure 3A** and **Figure 3A**). Similar to the case of EGF induction, cells collected at different time points were mixed up in different clusters (**Figure 3B**). After selecting cluster C3 as the E state based on fewer TCs around (**Figures 3C,D**) and expression levels of canonical epithelial and mesenchymal marker genes (**Supplementary Figures 3B,C**), the most probable transition trajectories were revealed (**Figure 3E**). Based on the GO analysis of the top marker genes (**Supplementary Figure 3D**), the two ICSs were different states (**Supplementary Figure 3E**). The I1 state was related to signaling and locomotion indicating the communications with other cells and sharing mesenchymal features.

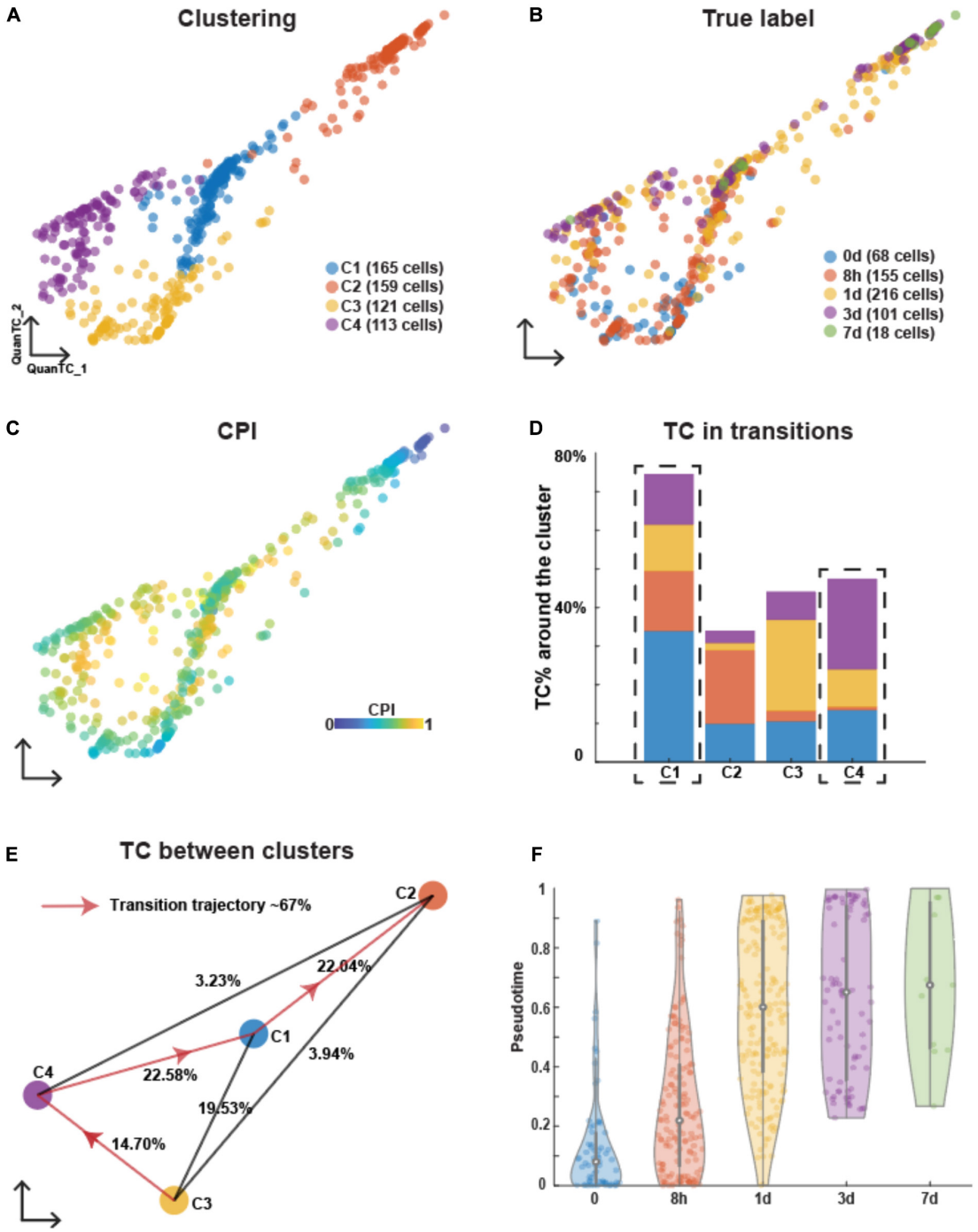

FIGURE 1 | Analyzing OVCA420 cancer cell line undergoing EMT induced by TGFB1 using QuanTC. **(A–C)** Visualization of cells in the two dimensional space by QuanTC. Each circle represents one cell colored by clustering **(A)**, the collection time of the samples after the treatment **(B)**, and CPI values **(C)**. **(D)** Percentage of TC associated with each cluster relative to the total number of TC. The dashed box covers the ICS having more TC around. The parameters to choose TC are given in **Supplementary Table 1**. **(E)** Visualization of cluster centers with color consistent with **(A)**. Each percentage on the line show the percentage of TC between two clusters relative to the total number of cells. Arrowed solid line shows the main transition trajectory. **(F)** Violin plot of pseudotime value of each cell vs the collection time points. Each dot represents a cell colored by collection time points. The circle displays the mean and vertical line shows the interquartile ranges.

Similar to EMT induced by EGF, the average pseudotime values slightly increased across time points with high variance within each time point, thus suggesting the heterogeneity of cells undergoing EMT (**Figure 3F**). Therefore, EMT induced by TNF was also found to be an asynchronous process.

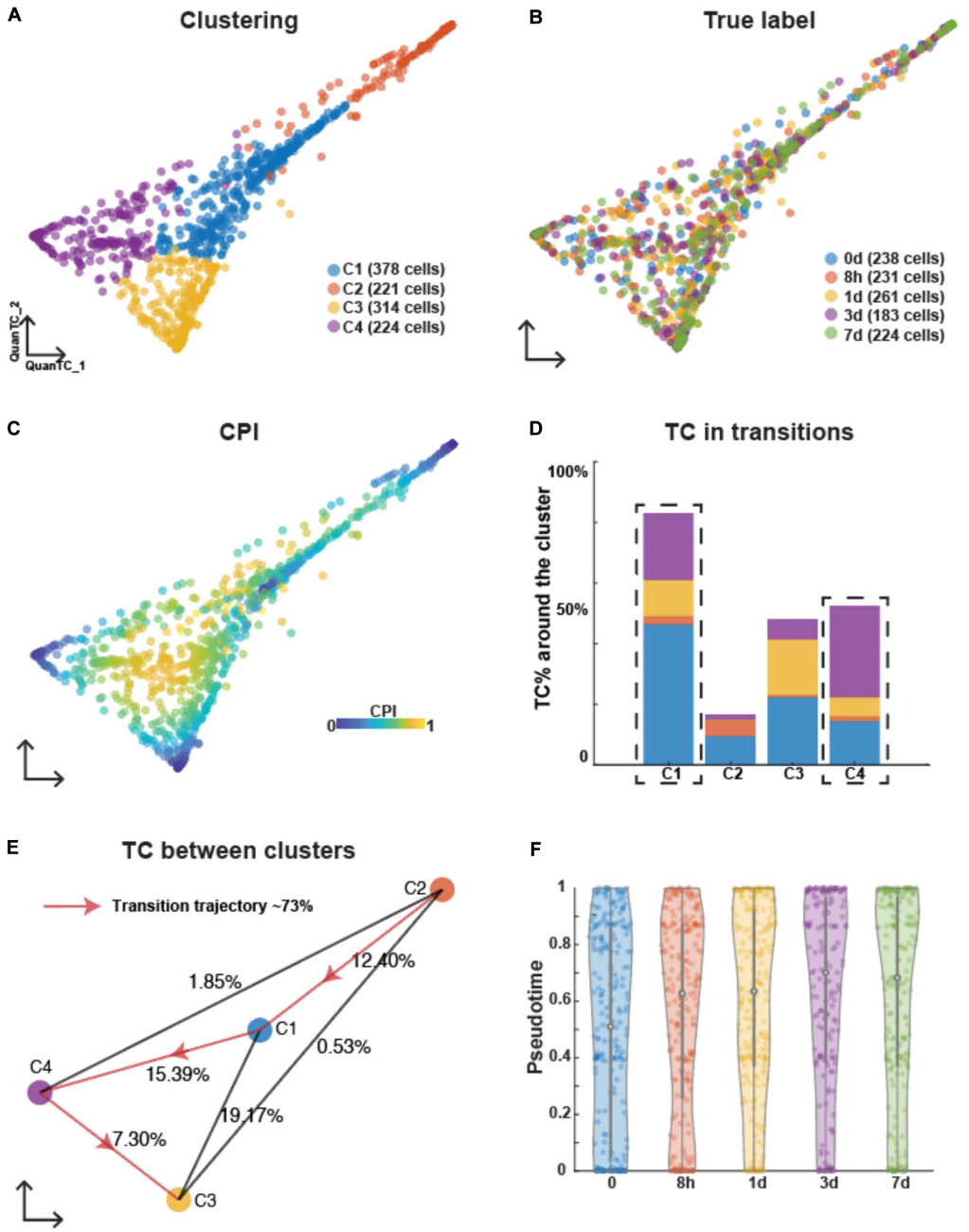

FIGURE 2 | Analyzing OVCA420 cancer cell line undergoing EMT induced by EGF using QuanTC. **(A–C)** Visualization of cells. Each circle represents one cell colored by clustering **(A)**, the collection time of the samples after the treatment **(B)**, and CPI values **(C)**. **(D)** Percentage of TC associated with each cluster relative to the total number of TC. The dashed box covers the ICS having more TC around. The parameters to choose TC are given in **Supplementary Table 1**. **(E)** Visualization of cluster centers with color consistent with **(A)**. Each percentage on the line show the percentage of TC between two clusters relative to the total number of cells. Arrowed solid line shows the main transition trajectory. **(F)** Violin plot of pseudotime value of each cell vs the collection time points. Each dot represents a cell colored by collection time points. The circle displays the mean and vertical line shows the interquartile ranges.

Context-Specific Cellular Communications With Underlying Gene Regulations in TGF-β Signaling

Transforming growth factor-β is a strong promoter of EMT (Hao et al., 2019). TGF-β ligands are not exclusively provided as an external EMT-inducing signal, but can also be secreted by cells, thus raising the possibility of cell–cell communication and EMT driven by intercellular signaling. In order to determine the possible role of TGF-β signaling in EMT, we assembled *in silico* ligand–receptor interaction pairs to explore the crosstalk between ICSs and E/M states. We applied SoptSC (Wang S. et al., 2019)

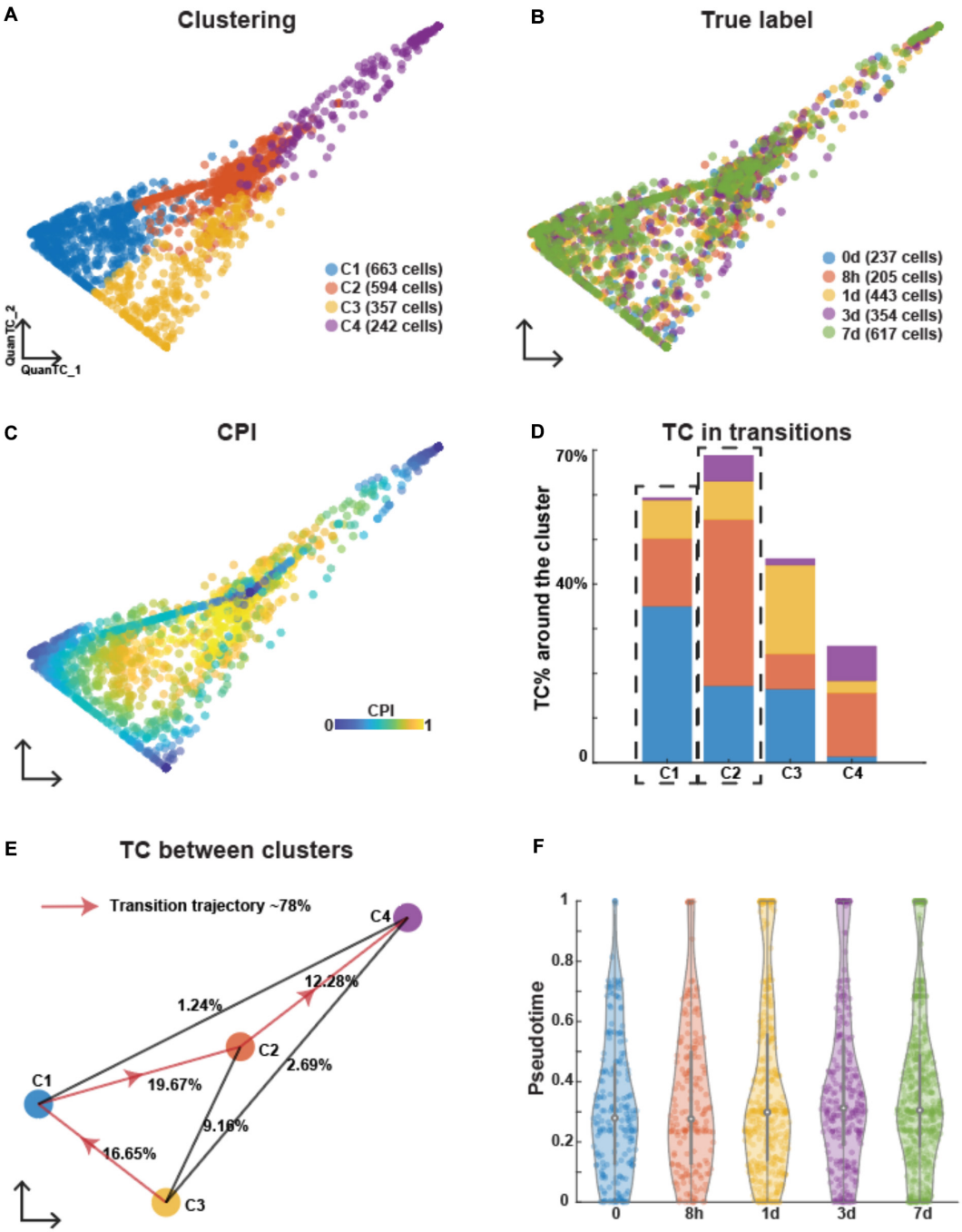

FIGURE 3 | Analyzing OVCA420 cancer cell line undergoing EMT induced by TNF using QuanTC. **(A–C)** Visualization of cells. Each circle represents one cell colored by clustering **(A)**, the collection time of the samples after the treatment **(B)**, and CPI values **(C)**. **(D)** Percentage of TC associated with each cluster relative to the total number of TC. The dashed box covers the ICS having more TC around. The parameters to choose TC are given in **Supplementary Table 1**. **(E)** Visualization of cluster centers with color consistent with **(A)**. Each percentage on the line show the percentage of TC between two clusters relative to the total number of cells. Arrowed solid line shows the main transition trajectory. **(F)** Violin plot of pseudotime value of each cell vs the collection time points. Each dot represents a cell colored by collection time points. The circle displays the mean and vertical line shows the interquartile ranges.

to the expression matrix with inferred states and calculated the signaling probability of each ligand–receptor pair and their downstream targets between pairs of cells. Finally, averaging these pairwise signaling probabilities within each EMT state provides a snapshot of how cells tend to communicate based on their degree of EMT progression (**Figures 4A–C**).

In **Figure 4B**, the directed edges from lower hemisphere to upper hemisphere were inferred between cells where a

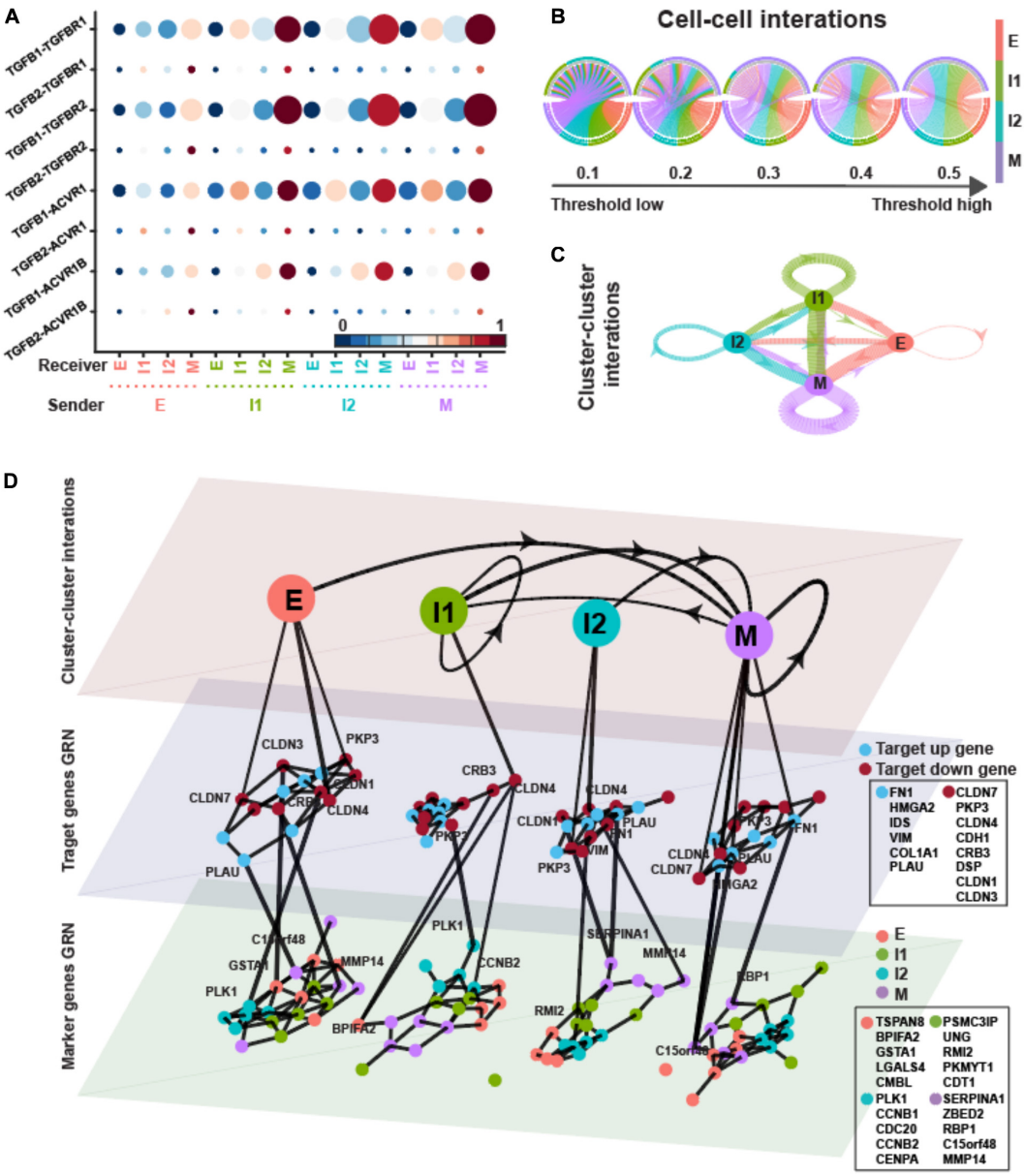

FIGURE 4 | TGFB pathway on OVCA420 cancer cell line undergoing EMT induced by TGFB1. **(A)** Visualization of signaling probability scores of Ligand-Receptor pairs and their downstream signaling components. Dot size represents the number of averaged cells with non-zero probability scores between clusters. Dot color represents the signaling probability scores. **(B)** Circos plot of intercellular network on the top ten ligand-producing and top ten receptor-bearing cells from every cluster. The upper hemisphere of the plot shows receptor-bearing cells. The chords of the plot are colored by the ligand-producing cells in the lower hemisphere. The directed edges from the lower hemisphere to the upper hemisphere represent the probabilities of signaling between cells. The probabilities of signaling between cells above the thresholds are presented. **(C)** Intercluster network. The widths of edges are proportional to the signaling probability scores between clusters. The directed edges are colored by the ligand-producing clusters. **(D)** Multilayer network. The first layer shows the intercluster network as in **(C)** but with higher signaling probabilities greater than 0.5. Second and third layers show gene regulatory networks of target genes and top marker genes of clusters, respectively, using the PIDC algorithm. The target up (down) genes are the up-regulated (down-regulated) target genes of TGF-β signaling. Each dot represents a gene colored by its type. Graph edges indicate the top interactions and the length of the edge is inversely proportional to the interaction strength between genes. The link between first and second layer indicates the target gene are higher expressed within the cluster. The link between second and third layer indicates the strong interaction strength between target and marker genes. The widths of links between layers are proportional to the interaction strength. The ligands, receptors and target genes are given in **Supplementary Table 3**.

high probability of signaling was predicted according to the expressions of ligands in a "sender" (lower hemisphere in the figure) cell and the appropriate expressions of cognate receptors and target genes in a "receiver" cell (upper hemisphere in the figure). The large proportion of M state behaving as "receiver" with high signaling probabilities suggests that

the M state played a dominant role as receiver in TGF-β signaling. All the four states behaved as "sender" in TGF-β signaling.

The cluster–cluster signaling network was then constructed based on the average cell–cell signaling within each cluster (**Figure 4C**). We used strength, closeness, and pagerank as metrics to measure node centrality in the signaling network so that we can quantify the centralities of states in TGF-β signaling. Strength is defined as the sum over weights of the adjacent edges for a given node. Closeness of a node is the inverse of the average length of the shortest path to/from all the other nodes. Pagerank is proportional to the average time spent at a given node during all random walks; therefore, we interpret a high pagerank score as an indication that a node serves as a signaling hub in the network. The pagerank centrality of I1 and that of M were higher, thus showing the signaling hub potential (**Supplementary Table 2**). The I1 and M states had higher in-strength and lower in-closeness indicating that they behaved more like receivers (**Supplementary Table 2**).

To explore the change of the GRNs underlying TGF-β signaling with respect to EMT progress, we applied PIDC (Chan et al., 2017), an algorithm using partial information decomposition to identify GRNs, to the gene expression matrix of target genes and marker genes inferred by QuanTC within each state. In the dataset induced by TGFB1, the first layer of the multilayer network showed the cluster–cluster interactions as in **Figure 4C** but with only higher signaling probabilities greater than 0.5 (**Figure 4D**, top layer). The widths of the directed lines were proportional to the signaling probabilities. The central and bottom layers displayed the GRNs of target genes and marker genes within each state, respectively. The interactions between genes within each state were shown by the edges with lengths inversely proportional to the correlations between genes.

Based on the average correlations between target genes of TGF-β signaling and marker genes (**Supplementary Figure 1F**), both the up-regulated target genes and down-regulated target genes had stronger interactions with marker genes within E and M states. The up-regulated target genes always had largest correlations with marker genes of M stats, whereas the down-regulated target genes had relatively larger correlations with E marker within only E and M states.

In the dataset of EMT induced by EGF, the average TGF-β signaling probabilities suggest that I2 and M states played important roles as receivers, whereas all four states shared similar importance as senders (**Figures 5A–C**). Compared to EMT induced by TGFB1, the pagerank centrality of I2, instead of I1, and M states were higher (**Supplementary Table 2**).

In the multilayer network, the highly varied target genes were quite similar to EMT induced by TGFB1 (**Figures 4**, **5D**). The up-regulated target genes were the same except missing COL1A1, and five out of the eight down-regulated target genes were the same as in **Figure 4D**. However, the top five marker genes of each state varied between the two treatments. Only LGALS4, BPIFA2, and ZBED2 shared marker genes of E and M states. CCNB1 and CCNB2, used to be I2 markers, were I1 markers for EMT induced by EGF.

The average correlations between target genes and marker genes were stronger within the I1 state (**Supplementary Figure 2F**). The up-regulated target genes did not always have largest correlations with marker genes of M state but still with relatively large correlations. The down-regulated target genes had stronger correlations with E markers except in the M state.

In the dataset of EMT induced by TNF, the different EMT states seemed to have similar importance as sender in TGF-β signaling (**Figures 6A–C**). The E and M states behaved as the main receivers. The M state had higher pagerank value showing the potential of signaling hub (**Supplementary Table 2**).

In the multilayer network, the varied up-regulated target genes were the subset of the genes in EMT induced by EGF except having CLDN3, and the down-regulated target genes were the subset of those genes in EMT induced by TGFB1 (**Figures 4–6D**). More than half of the marker genes of E, I1, and M states were the same as in EMT induced by EGF, suggesting the similarity of the EMT under the two treatments.

The target genes and marker genes had higher correlations within the I2 state (**Supplementary Figure 3F**). The up-regulated target genes always had relatively large correlations with marker genes of M state. The down-regulated target genes had stronger correlations with E markers except in the I2 state.

Overall, the M state and part of the ICSs behaved as the signaling hub in the TGF-β signaling of EMT under three different treatments (**Figures 4–6**). The M state was the main receiver in OVCA420 under three treatments with lowest in-closeness (**Supplementary Table 2**), while the underlying GRNs changed between different treatments and along EMT progress. Besides, the top marker genes of different EMT states were quite different among the EMT induced by different treatments, all suggesting the context-specific regulation of GRNs during EMT.

Dominant Role of ICSs *in vivo* During TGF-β Signaling

Finally, we compare the results obtained for OVCA420 cells with *in vivo* data from a skin SCC mouse model to seek whether the defining traits of EMT dynamics are conserved or context-specific. In the original study, a total of six distinct cell populations were identified based on differential expression of cell surface markers (CD106, CD61, and CD51), including four transition states (Pastushenko et al., 2018).

In our previous work (Sha et al., 2020), we identified a total of four EMT states (**Supplementary Figure 4A** and **Figure 7A**) when applying QuanTC unsupervised clustering (Pastushenko et al., 2018). There were two ICSs displaying biological processes including cell–cell adhesion and cell migration indicating hybrid epithelial/mesenchymal features (**Supplementary Figure 4B**).

Compared to the OVCA420 cancer cell line undergoing EMT, the ICSs in SCC had higher probabilities of signaling and played

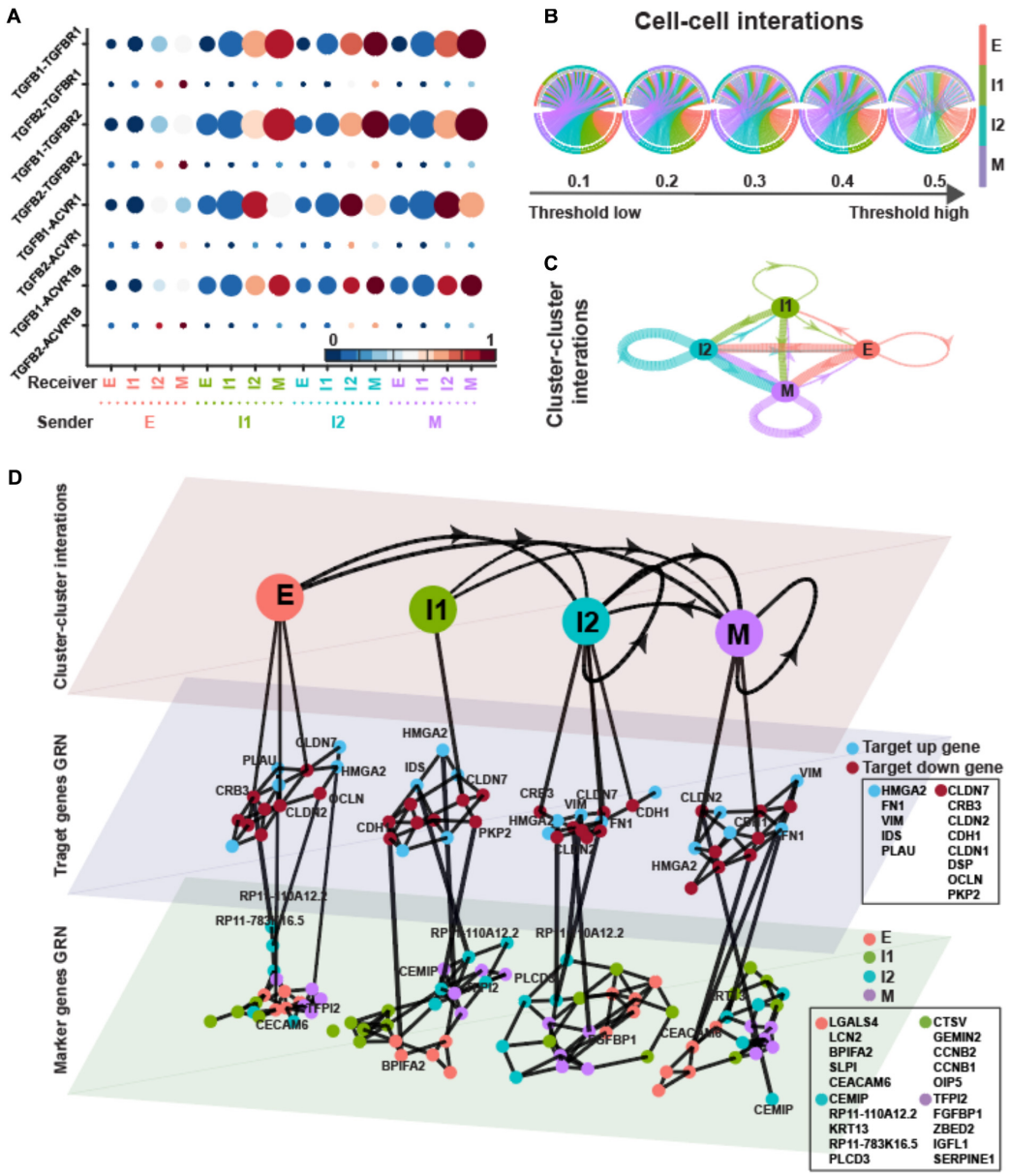

FIGURE 5 | TGFB pathway on OVCA420 cancer cell line undergoing EMT induced by EGF. **(A)** Visualization of signaling probability scores of Ligand-Receptor pairs and their downstream signaling components. Dot size represents the number of averaged cells with non-zero probability scores between clusters. Dot color represents the signaling probability scores. Dot color represents the signaling probability scores. **(B)** Circos plot of intercellular network on the top ten ligand-producing and top ten receptor-bearing cells from every cluster. The upper hemisphere of the plot shows receptor-bearing cells. The chords of the plot are colored by the ligand-producing cells in the lower hemisphere. The directed edges from the lower hemisphere to the upper hemisphere represent the probabilities of signaling between cells. The probabilities of signaling between cells above the thresholds are presented. **(C)** Intercluster network. The widths of edges are proportional to the signaling probability scores between clusters. The directed edges are colored by the ligand-producing clusters. **(D)** Multilayer network. The first layer shows the intercluster network as in **(C)** but with higher signaling probabilities greater than 0.5. Second and third layers show gene regulatory networks of target genes and top marker genes of clusters, respectively, using the PIDC algorithm. The target up (down) genes are the up-regulated (down-regulated) target genes of TGF-β signaling. Each dot represents a gene colored by its type. Graph edges indicate the top interactions and the length of the edge is inversely proportional to the interaction strength between genes. The link between first and second layer indicates the target gene are higher expressed within the cluster. The link between second and third layer indicates the strong interaction strength between target and marker genes. The widths of links between layers are proportional to the interaction strength. The ligands, receptors and target genes are given in **Supplementary Table 3**.

the even more dominant role of cell–cell and cluster–cluster interactions during TGF-β signaling (**Figures 7B–D**). The ICSs, especially the I1 state, had higher Pagerank scores and served as the signaling hub (**Supplementary Table 2**). Both ICSs had lower out-closeness score, indicating that they played the dominant role as the sender in TGF-β signaling. While the M state had by far

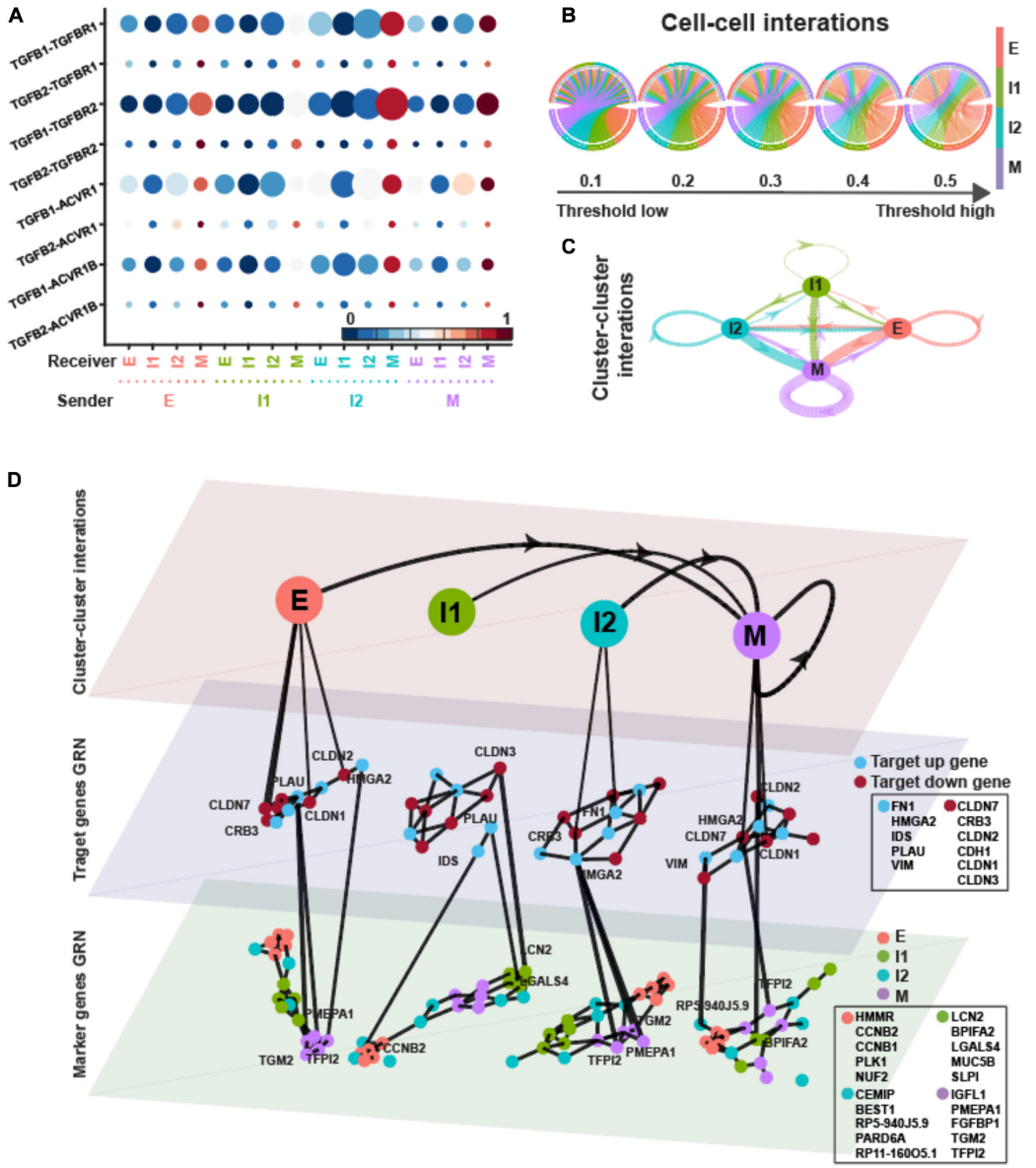

FIGURE 6 | TGFB pathway on OVCA420 cancer cell line undergoing EMT induced by TNF. **(A)** Visualization of signaling probability scores of Ligand-Receptor pairs and their downstream signaling components. Dot size represents the number of averaged cells with non-zero probability scores between clusters. Dot color represents the signaling probability scores. **(B)** Circos plot of intercellular network on the top ten ligand-producing and top ten receptor-bearing cells from every cluster. The upper hemisphere of the plot shows receptor-bearing cells. The chords of the plot are colored by the ligand-producing cells in the lower hemisphere. The directed edges from the lower hemisphere to the upper hemisphere represent the probabilities of signaling between cells. The probabilities of signaling between cells above the thresholds are presented. **(C)** Intercluster network. The widths of edges are proportional to the signaling probability scores between clusters. The directed edges are colored by the ligand-producing clusters. **(D)** Multilayer network. The first layer shows the intercluster network as in **(C)** but with higher signaling probabilities greater than 0.5. Second and third layers show gene regulatory networks of target genes and top marker genes of clusters, respectively, using the PIDC algorithm. The target up (down) genes are the up-regulated (down-regulated) target genes of TGF-β signaling. Each dot represents a gene colored by its type. Graph edges indicate the top interactions and the length of the edge is inversely proportional to the interaction strength between genes. The link between first and second layer indicates the target gene are higher expressed within the cluster. The link between second and third layer indicates the strong interaction strength between target and marker genes. The widths of links between layers are proportional to the interaction strength. The ligands, receptors and target genes are given in **Supplementary Table 3**.

the higher pagerank score in the three OVCA420 datasets, the pagerank score of the M state in SCC was comparable to those of I1 and I2. Consistently, in the original study, the mesenchymal SCC exhibited a "quasi-mesenchymal" phenotype, which was more similar to intermediate state, instead of a fully mesenchymal phenotype (Pastushenko et al., 2018).

The highly varied target genes and marker genes of each state shared no similarity to the OVCA420 cancer line (**Figure 7E**).

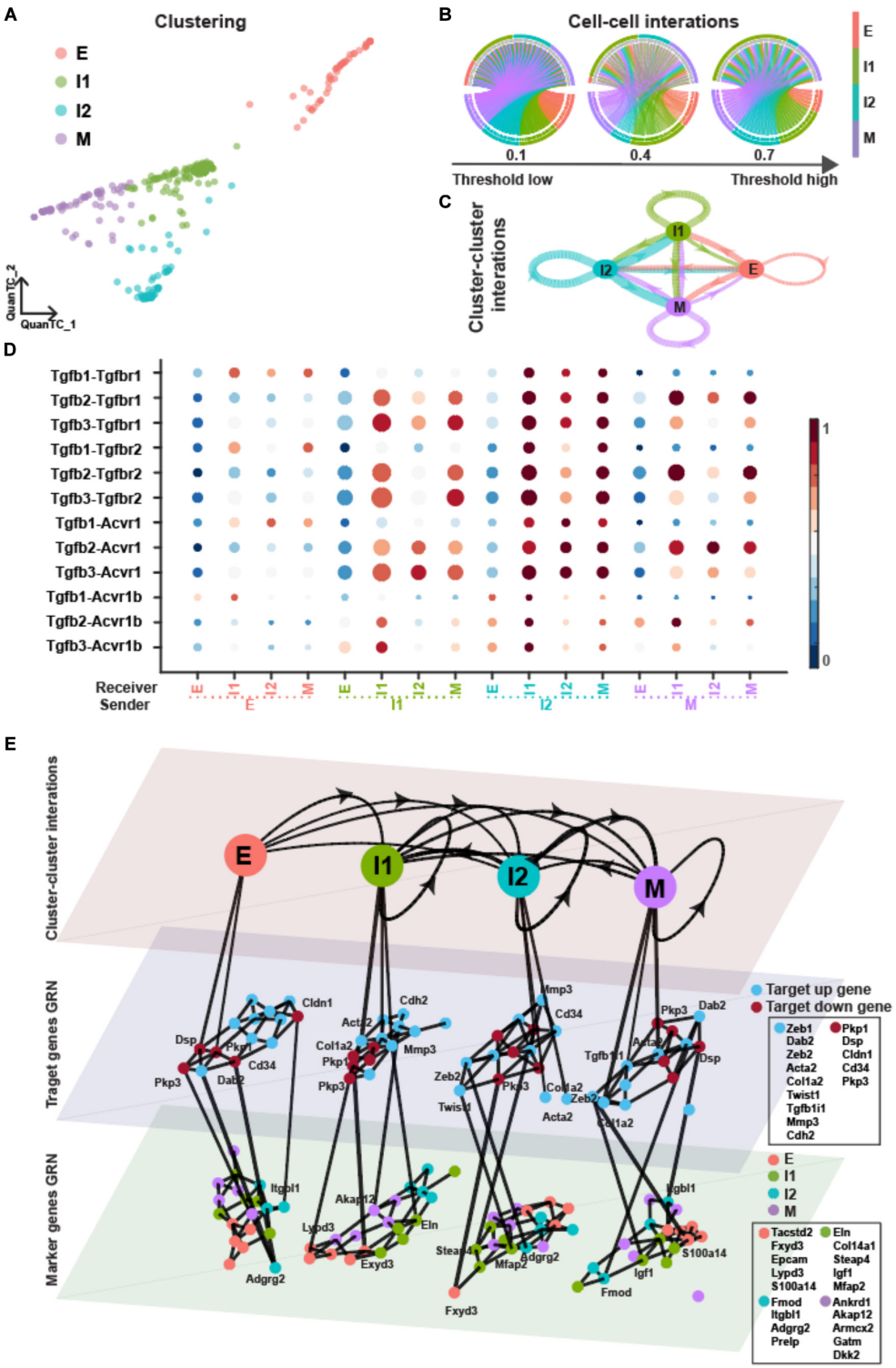

FIGURE 7 | TGF-β pathway on EMT in SCC dataset. **(A)** Visualization of cells using QuanTC. Each circle represents a cell colored by corresponding cell state. **(B)** Circos plot of intercellular network on the top ten ligand-producing and top ten receptor-bearing cells from every cluster. The upper hemisphere of the plot shows receptor-bearing cells. The chords of the plot are colored by the ligand-producing cells in the lower hemisphere. The directed edges from the lower hemisphere to the upper hemisphere represent the probabilities of signaling between cells. The probabilities of signaling between cells above the thresholds are presented. **(C)** Intercluster network. The widths of edges are proportional to the signaling probability scores between clusters. The directed edges are colored by the ligand-producing clusters. **(D)** Visualization of signaling probability scores of Ligand-Receptor pairs and their downstream signaling components. Dot size represents

(Continued)

FIGURE 7 | Continued
the number of averaged cells with non-zero probability scores between clusters. Dot color represents the signaling probability scores. **(E)** Multilayer network. The first layer shows the intercluster network as in **(C)** but with higher signaling probabilities greater than 0.5. Second and third layers show gene regulatory networks of target genes and top marker genes of clusters, respectively, using the PIDC algorithm. The target up (down) genes are the up-regulated (down-regulated) target genes of TGF-β signaling. Each dot represents a gene colored by its type. Graph edges indicate the top interactions and the length of the edge is inversely proportional to the interaction strength between genes. The link between first and second layer indicates the target gene are higher expressed within the cluster. The link between second and third layer indicates the strong interaction strength between target and marker genes. The widths of links between layers are proportional to the interaction strength. The ligands, receptors and target genes are given in **Supplementary Table 3**.

The target genes had strong associations with inferred marker genes within E and I1 states (**Supplementary Figure 4C**). It suggests that EMT varies both between mouse vs human, and *in vitro* vs *in vivo*.

MATERIALS AND METHODS

scRNA-Seq Data Clustering and Transition Trajectory Reconstruction

QuanTC was used to perform clustering and transition trajectory reconstruction. QuanTC can simultaneously detect the ICSs and construct transition trajectories via quantifying the CPI (Sha et al., 2020). The cells with higher CPI values are considered to be transitioning between clusters and are identified as TCs. Via non-negative matrix factorization, QuanTC calculates the probabilities of a given cell belonging to the identified clusters. Cells are projected to a low-dimensional space based on a probabilistic regularized embedding. The transition trajectories are then inferred by summing the cluster-to-cluster transition probabilities that are calculated from cell-to-cluster probabilities and TCs between clusters. The transition genes and marker genes of clusters are obtained through factorizing the gene expression matrix as product of cell-to-cluster probabilities and likelihoods of genes uniquely marking each cluster. In the first step of QuanTC, we applied two additional considerations when choosing the number of identified clusters. First, we know from the original experiment that cells undergo EMT (i.e., there is at least one E state and one M state); furthermore, given that we seek to study ICSs during EMT, we search for at least three total states.

Preprocessing

Single cells with less than 95% expressed genes among all detected genes were considered as low-quality cells and were filtered. Top 3,000 bimodal distributed genes were selected by QuanTC with default parameters to do downstream analysis.

Clustering

A total of 3,000 selected genes and 558 cells of OVCA420 induced by TGFB1, 1,137 cells of OVCA420 induced by EGF, and 1,856 cells of OVCA420 induced by TNF from day 0 to day 7 were retained for clustering. Consensus clustering via SC3 (Kiselev et al., 2017) was performed on the expression matrix to capture the cell–cell similarity. The clusters were defined based on symmetric non-negative factorization as wrapped in QuanTC.

Transition Trajectory

The beginning and end of EMT transition trajectory, E/M states, were identified based on the percentage of TCs around each cluster. The parameters to choose TCs were given in **Supplementary Table 1**. The clusters with fewer TCs around were considered as E/M states, whereas the rest clusters were considered as ICSs along EMT. The E/M states between the two clusters were then identified based on the canonical epithelial and mesenchymal marker genes. The potential transition trajectory was inferred according to the TCs between clusters using "traj" function wrapped in QuanTC. The pseudotime value of each cell was then computed by QuanTC based on the two most probable trajectories.

EMT Marker Genes

The marker genes and transition genes were defined using "markers" function wrapped in QuanTC.

GO Analysis

The analysis of GO biological processes was performed by Metascape (Zhou et al., 2019) on the top 50 markers genes of each ICS selected by QuanTC.

Qualitatively Characterizing Cell–Cell Communications

SoptSC (Wang S. et al., 2019) was used on the datasets without gene filtering to calculate the probability matrix of signals being passed between cells and clusters. Signaling probabilities between cells are defined based on weighted co-expression of signaling pathway activity in sender–receiver cell pairs. With the input of ligand–receptor pairs and target genes (up-regulated or down-regulated in response to pathway activation), SoptSC computes signaling probabilities between sender cells (expressing ligands) and receiver cells (expressing receptors and exhibiting differential target genes activity). Intuitively, given a ligand–receptor pair for a specific signaling pathway, if the ligand is highly expressed in cell i, the cognate receptor is highly expressed in cell j, and the target gene activity in cell j suggests that the signaling pathway may have been activated in this cell, and then there is a chance that communication occurred between these two cells. The signaling passed from cell i to j for a given ligand–receptor pair is quantified by the signaling probability $P_{i,j}$. For a set of ligand–receptor pairs, SoptSC considers the consensus signaling probabilities between cells by taking the average over all signaling probability matrices. The signaling probability passed from cluster u to cluster v is then given

by $P_{u,v} = \frac{\sum_{i \in C_u, j \in C_v} P_{i,j}}{|C_u||C_v|}$, with $|C_u|$ representing the number of cells in cluster u.

The lists of ligands, receptors, and target genes were retrieved from previous studies (Wendt et al., 2009; Xu et al., 2009; Jin et al., 2020) and are given in **Supplementary Table 3**.

Measuring Node Centrality

The centrality of a node (cluster) in cellular communication network is used to quantify its importance in the signaling. We used strength, closeness, and pagerank as metrics to measure node centrality. All these centralities were calculated with the package igraph 1.2.4 (Csardi and Nepusz, 2006).

Strength is one of the basic measures of centrality: it is measured by summing up the edge weights of the adjacent edges for a given node. Our inferred cluster–cluster communication networks are directed, so we calculated in-strength (incoming edges), and out-strength (outgoing edges). Closeness of a given node is defined by the inverse of the average length of the shortest path to/from all the other nodes. In-closeness measures the path to the node, whereas out-closeness measures the paths from the node. We used the normalized values to avoid biases based on the network size. Pagerank is proportional to the average time spent at a given node during all random walks. In the cluster–cluster communication networks, the clusters with high pagerank can be seen as the signaling hub.

Multilayer Regulations of EMT

We utilized the multilayer network framework (Kivelä et al., 2014) to analyze and visualize the changes of complex hierarchical signaling and gene expression regulations in EMT across multiple scales.

Mathematically, the multilayer network can be expressed as the $M = (V_M, E_M, V, \boldsymbol{L})$. Here, V denotes sets of all nodes in the network (as in the regular case), and $\boldsymbol{L} = \{L_a\}_{a=1}^{d}$ denotes d aspects of the network layers, with each aspect $L_a = \{L_a^i\}_{i=1}^{k_a}$ contains k_a elementary layers. Denotes $\times$ as the Cartesian product of sets, and then the node–layer tuple set $V_M \subseteq V \times L_1 \times \cdots \times L_d$ represents all the feasible node–layer combinations in which a node is present in the corresponding layers. The edges set $E_M \subseteq V_M \times V_M$ denotes the weighted links across nodes and layers.

In our context, the nodes set V not only contains cell states $S = \bigcup_{k=1}^{N_c} S_k$ along the EMT trajectories, with N_c denoting the number of cell states, but also contains target genes T of specified signal transduction pathway and marker genes A of each cell state. The layers $\boldsymbol{L} = \{L_H, L_C\}$ has two aspects: The hierarchy aspect $L_H = \{L_H^1, L_H^2, L_H^3\}$ represents the elementary layers of cell–cell communication L_H^1, target genes L_H^2, and marker genes L_H^3, respectively, and the cell states aspect $L_C = \{L_C^k\}_{k=1}^{N_c}$ represents the EMT stages of E state, ICSs, and M state ordered by pseudotime of QuanTC, as we are interested in constructing cell-state–specific regulatory relations. For simplicity, we denote the node–layer tuples in EMT as $V_M = \{(S, L_H^1, \cdot), (T, L_H^2, \cdot), (A, L_H^3, \cdot)\} \subseteq V \times L_H \times L_C$, representing the hierarchical regulation structures at different stages. For instance, (A, L_H^3, L_C^1) denotes the marker genes analyzed in the E state, while (T, L_H^2, L_C^2) represents the target genes considered in the first ICSs. We next specify how the edges E_M are constructed based on the V_M.

The Edges Within Layer $\left(S, L_H^1, \cdot\right)$

The first layer L_H^1 in hierarchy aspect displays the cluster–cluster interactions of intercellular communication, where the aligned nodes show the different EMT states/clusters. Using the notations above, $\left(S, L_H^1, L_C^k\right)$ contains only one node for each k, representing the cell state S_k. The weights for the directed edges to connect $\left(S, L_H^1, L_C^i\right)$ and $\left(S, L_H^1, L_C^j\right)$ are the cluster–cluster interactions between state S_i and state S_j computed by SoptSC above threshold 0.7.

The Edges Within Layer $\left(T, L_H^2, \cdot\right)$

The second layer L_H^2 demonstrates the state-specific interactions among target genes at different stages. The target genes T are the intersection of the list of target genes and the top 3,000 selected informative genes. Given the stage L_C^k, the weighted edges between target gene pair $\left(T_X, L_H^2, L_C^k\right)$ and $\left(T_Y, L_H^2, L_C^k\right)$ were constructed by PIDC algorithm (Chan et al., 2017) using partial information decomposition, only with the cells in cluster S_k. The input to PIDC is an expression matrix with cells from S_k, and the confidence of an edge between a pair of genes is given by $c = F_X\left(U_{X,Y}\right) + F_Y\left(U_{X,Y}\right)$ where $F_X(U)$ is the cumulative distribution function of all the proportional unique contribution scores involving gene X. The top 30% weights were used to embed the inferred network in $\left(T, L_H^2, L_C^k\right)$ using "graph" function in MATLAB based on spectral layout (Koren, 2005). The weights were normalized with max 2 to be comparable with other datasets.

The Edges Within Layer $\left(A, L_H^3, \cdot\right)$

The third layer L_H^3 demonstrates the state-specific interactions among marker genes at different stages. The marker genes selected were identical for $\left(A, L_H^3, L_C^k\right)$ with respect to the choice of k, which represent the union of top five marker genes in each cluster inferred by QuanTC. The edges between marker genes are state-specific for each cell-state layer L_C^k, using the same strategy as for the target genes described above.

The Edges Connecting Layer $\left(S, L_H^1, \cdot\right)$ and $\left(T, L_H^2, \cdot\right)$

These edges quantify the expression of target genes within different states during EMT. The weights for the edges between $\left(S, L_H^1, L_C^k\right)$ and $\left(T, L_H^2, L_C^k\right)$ are the mean expression levels of target genes within cell state S_k, and top 20% weights were shown.

The Edges Connecting Layer $\left(T, L_H^2, \cdot\right)$ and $\left(A, L_H^3, \cdot\right)$

These edges display the regulatory interactions from target genes to marker genes within different states during EMT. The weights for the edges between $\left(T, L_H^2, L_C^k\right)$ and $\left(A, L_H^3, L_C^k\right)$ were inferred by PIDC within cell state S_k, and top 1.5% weights were shown.

DISCUSSION

In this study, we have developed an approach combining unsupervised learning, multivariate information theory, and multilayer network approach to uncover the complex cellular crosstalk and the underlying gene regulatory relationship of EMT from scRNA-seq data.

We started with trajectory reconstruction on the time-series datasets of an OVCA420 cancer cell line undergoing EMT induced by three different external signal (TGFB1, EGF, and TNF) and uncovered the existence of multiple ICSs displaying hybrid epithelial and mesenchymal features. Analysis of scRNA-seq previously demonstrated that EMT induction by TGFB1, EGF, and TNF is carried by context-specific signaling pathways (Cook and Vanderhyden, 2020). Here, we show striking differences in the EMT population dynamics as well. While EMT induced by TGFB1 is synchronous, EGF and TNF induce asynchronous transitions because cells collected at different time points spread all over different clusters. These differences at the cell population level could be explained by the signaling complexity and modularity in response to different EMT inducers. TNF can activate nuclear factor κB (NF-κB) signaling, which in turn crosstalks with several transduction pathways and induces responses to inflammation (Hayden and Ghosh, 2014). TNF–NF-κB signaling has also been proposed as a stability factor for hybrid E/M phenotypes, thus potentially resisting a complete EMT in TNF-induced EMT (Bocci et al., 2019). Similarly, EGF regulation of EMT is not direct, but rather relies on several intermediate signaling steps that could hamper a synchronized transition (Kang et al., 2013). Certainly, future efforts focusing on integrating high-throughput data analysis with *in silico* modeling of the underlying regulatory circuitry will help validate or falsify these hypotheses.

To clarify how cells in different EMT states contribute to cell–cell signaling, we subsequently constructed multilayer networks displaying the TGF-β signaling communication between cells in different EMT states and the underlying GRN that regulates EMT at different EMT stages. We found that ICSs serve as signaling hubs of cell–cell communication, as well as the context-specific response of TGF-β under different treatments. In other words, cells in intermediate EMT states can send and receive inputs from other cells through TGF-β signaling, potentially inducing EMT in their neighbors. Therefore, both cell autonomous TGFB1 induction and intercellular TGFB signaling could contribute to EMT. Future experiments controlling conditional knockouts of TGFB ligands could validate this prediction and quantify the role played by cell–cell communication in EMT. These observations also raise an interesting parallel with Notch signaling, another master regulator of cell–cell communication (Bray, 2016). Signaling through the Notch-Jagged pathway between cancer cells in intermediate EMT states has been proposed as a mechanism that (i) stabilizes intermediate EMT states and (ii) further induces "partial EMT" in other cells (Bocci et al., 2017; Jolly et al., 2017). Our analysis on *in vivo* dataset also suggests that ICS plays the more dominant role in the TGF-β signaling communication.

The core gene circuits for EMT are known to involve multiple molecular components and interactions (Jia et al., 2017; Tian et al., 2019; Yang et al., 2020), providing mechanisms of the EMT transition process (Jolly and Levine, 2017). Recent time-series scRNA-seq data suggest that EMT is indeed highly context-specific (Cook and Vanderhyden, 2020), calling for the need of inferring EMT regulation circuits from a data-driven approach (Tanaka and Ogishima, 2015; Ramirez et al., 2020). Previous works have constructed the GRN of EMT based on the combination of prior knowledge, transcription factor predictions, and model validations from single-cell datasets (Ramirez et al., 2020). Here we have incorporated the intercellular communications in the context of analyzing TCs and ICSs to inspect the dynamical change of regulation interactions along the EMT spectrum.

Our analysis reveals that ICS plays the crucial role in not only interchanging information with both pure epithelial and mesenchymal states, but also communicating with other cells in ICSs during EMT. Previously, the role of ICSs has been studied for tumor metastasis (Jolly et al., 2015) and analyzed through the emergent dynamical properties such as signal adaptation, noise attenuation, and population transition (Ta et al., 2016; Sha et al., 2019; Goetz et al., 2020). Taken together, the EMT cell lineage models with ICS-mediated feedback through cell–cell communications (Lander et al., 2009; Lo et al., 2009) could be further developed to explore the non-linear effects on different cell populations (Jia W. et al., 2019).

The integrative analysis here is a general approach and can be applied to other cell-state transition processes beyond EMT. In particular, the multiplayer gene regulatory and intercellular network provides a multiscale framework to simultaneously explore the cellular communications, the underlying gene regulations, and dynamics of GRNs along transitions. By incorporating additional layers of different transduction elements beyond TGF-β (Jin et al., 2020) and associated transcription factors, one can investigate the more complex regulation processes, such as signal crosstalk and corporation of multiple pathways (Xing and Tian, 2019). In addition, the inclusion of spatial information layer may also facilitate the accuracy of intercellular communication analysis (Cang and Nie, 2020).

Overall, our study provides an initial attempt to investigate the multiscale interactions of intercellular communications and

gene expression regulations during the dynamical process of cell-fate determination.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These datasets can be found here: SCC (GEO: GSE110357) and OVCA420 cancer cell line (GEO: GSE147405) datasets downloaded from the Gene Expression Omnibus.

AUTHOR CONTRIBUTIONS

QN, PZ, and YS conceived the study. YS implemented the algorithm and wrote the codes. YS, SW, and FB performed data analysis. YS, FB, PZ, and QN wrote the manuscript with the help from all the authors. QN and PZ supervised the research. All authors approved the manuscript.

FUNDING

The research was supported by the National Institutes of Health (U01AR073159, R01AR044882, and U54CA217378; in part); National Science Foundation (DMS1763272); and Simons Foundation (594598 to QN).

ACKNOWLEDGMENTS

We thank all the reviewers for their insightful comments and helpful suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.604585/full#supplementary-material

REFERENCES

An, S., Ma, L., and Wan, L. (2019). TSEE: an elastic embedding method to visualize the dynamic gene expression patterns of time series single-cell RNA sequencing data. *BMC Genom.* 20:224. doi: 10.1186/s12864-019-5477-8

Boareto, M., Jolly, M. K., Goldman, A., Pietilä, M., Mani, S. A., Sengupta, S., et al. (2016). Notch-Jagged signalling can give rise to clusters of cells exhibiting a hybrid epithelial/mesenchymal phenotype. *J. R. Soc. Interface* 13:20151106. doi: 10.1098/rsif.2015.1106

Bocci, F., Gearhart-Serna, L., Boareto, M., Ribeiro, M., Ben-Jacob, E., Devi, G. R., et al. (2019). Toward understanding cancer stem cell heterogeneity in the tumor microenvironment. *Proc. Natl. Acad. Sci. U S A.* 116, 148–157. doi: 10.1073/pnas.1815345116

Bocci, F., Jolly, M. K., George, J. T., Levine, H., and Onuchic, J. N. (2018). A mechanism-based computational model to capture the interconnections among epithelial-mesenchymal transition, cancer stem cells and Notch-Jagged signaling. *Oncotarget* 9:29906. doi: 10.18632/oncotarget.25692

Bocci, F., Jolly, M. K., Tripathi, S. C., Aguilar, M., Hanash, S. M., Levine, H., et al. (2017). Numb prevents a complete epithelial-mesenchymal transition by modulating Notch signalling. *J. R. Soc. Interface* 14:20170512. doi: 10.1098/rsif.2017.0512

Bocci, F., Onuchic, J. N., and Jolly, M. K. (2020). Understanding the principles of pattern formation driven by notch signaling by integrating experiments and theoretical models. *Front. Physiol.* 11:929. doi: 10.3389/fphys.2020.00929

Bray, S. J. (2016). Notch signalling in context. *Nat. Rev. Mol. Cell Biol.* 17, 722–735. doi: 10.1038/nrm.2016.94

Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* 17, 159–162. doi: 10.1038/s41592-019-0667-5

Cabello-Aguilar, S., Alame, M., Kon-Sun-Tack, F., Fau, C., Lacroix, M., and Colinge, J. (2020). SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* 48:e55. doi: 10.1093/nar/gkaa183

Camp, J. G., Sekine, K., Gerber, T., Loeffler-Wirth, H., Binder, H., Gac, M., et al. (2017). Multilineage communication regulates human liver bud development from pluripotency. *Nature* 546, 533–538. doi: 10.1038/nature22796

Cang, Z., and Nie, Q. (2020). Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* 11:2084.

Chan, T. E., Stumpf, M. P. H., and Babtie, A. C. (2017). Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Systems* 5, 251–267.e3.

Chen, Z., An, S., Bai, X., Gong, F., Ma, L., and Wan, L. (2019). DensityPath: an algorithm to visualize and reconstruct cell state-transition path on density landscape for single-cell RNA sequencing data. *Bioinformatics* 35, 2593–2601. doi: 10.1093/bioinformatics/bty1009

Cook, D. P., and Vanderhyden, B. C. (2020). Context specificity of the EMT transcriptional response. *Nat. Commun.* 11:2142.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems* 1695, 1–9.

Fazilaty, H., Rago, L., Kass Youssef, K., Ocana, O. H., Garcia-Asencio, F., Arcas, A., et al. (2019). A gene regulatory network to control EMT programs in development and disease. *Nat. Commun.* 10:5115.

Goetz, H., Melendez-Alvarez, J. R., Chen, L., and Tian, X. J. (2020). A plausible accelerating function of intermediate states in cancer metastasis. *PLoS Comput. Biol.* 16:e1007682. doi: 10.1371/journal.pcbi.1007682

Hao, Y., Baker, D., and Ten Dijke, P. (2019). TGF-beta-Mediated epithelial-mesenchymal transition and Cancer metastasis. *Int. J. Mol. Sci.* 20:2767. doi: 10.3390/ijms20112767

Hayden, M. S., and Ghosh, S. (2014). Regulation of NF-kappaB by TNF family cytokines. *Semin. Immunol.* 26, 253–266. doi: 10.1016/j.smim.2014.05.004

Hong, T., Watanabe, K., Ta, C. H., Villarreal-Ponce, A., Nie, Q., and Dai, X. (2015). An Ovol2-Zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS Comput. Biol.* 11:e1004569. doi: 10.1371/journal.pcbi.1004569

Jia, D., George, J. T., Tripathi, S. C., Kundnani, D. L., Lu, M., Hanash, S. M., et al. (2019a). Testing the gene expression classification of the EMT spectrum. *Phys. Biol.* 16:025002. doi: 10.1088/1478-3975/aaf8d4

Jia, W., Deshmukh, A., Mani, S. A., Jolly, M. K., and Levine, H. (2019b). A possible role for epigenetic feedback regulation in the dynamics of the epithelial-mesenchymal transition (EMT). *Phys. Biol.* 16:066004. doi: 10.1088/1478-3975/ab34df

Jia, D., Jolly, M. K., Tripathi, S. C., Den Hollander, P., Huang, B., Lu, M., et al. (2017). Distinguishing mechanisms underlying EMT tristability. *Cancer Converg.* 1:2.

Jin, S., Maclean, A. L., Peng, T., and Nie, Q. (2018). scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics* 34, 2077–2086. doi: 10.1093/bioinformatics/bty058

Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Myung, P., Plikus, M. V., et al. (2020). Inference and analysis of cell-cell communication using cellchat. *bioRxiv* [preprint] doi: 10.1101/2020.07.21.214387

Jolly, M. K., Boareto, M., Debeb, B. G., Aceto, N., Farach-Carson, M. C., Woodward, W. A., et al. (2017). Inflammatory breast cancer: a model for investigating cluster-based dissemination. *NPJ Breast Cancer* 3:21.

Jolly, M. K., Boareto, M., Huang, B., Jia, D., Lu, M., Ben-Jacob, E., et al. (2015). Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. *Front. Oncol.* 5:155. doi: 10.3389/fonc.2015.00155

Jolly, M. K., and Levine, H. (2017). Computational systems biology of epithelial-hybrid-mesenchymal transitions. *Curr. Opin. Systems Biol.* 3, 1–6. doi: 10.1016/j.coisb.2017.02.004

Jolly, M. K., Ward, C., Eapen, M. S., Myers, S., Hallgren, O., Levine, H., et al. (2018). Epithelial-mesenchymal transition, a spectrum of states: role in lung development, homeostasis, and disease. *Dev. Dyn* 247, 346–358. doi: 10.1002/dvdy.24541

Kalluri, R., and Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* 119, 1420–1428. doi: 10.1172/jci39104

Kang, H. W., Crawford, M., Fabbri, M., Nuovo, G., Garofalo, M., Nana-Sinkam, S. P., et al. (2013). A mathematical model for microRNA in lung cancer. *PLoS One* 8:e53663. doi: 10.1371/journal.pone.0053663

Kang, X., Wang, J., and Li, C. (2019). Exposing the underlying relationship of cancer metastasis to metabolism and epithelial-mesenchymal transitions. *iScience* 21, 754–772. doi: 10.1016/j.isci.2019.10.060

Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., et al. (2017). SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* 14, 483–486. doi: 10.1038/nmeth.4236

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *J. Complex Netw.* 2, 203–271.

Koren, Y. (2005). Drawing graphs by eigenvectors: theory and practice. *Comp. Mathemat. Appl.* 49, 1867–1888. doi: 10.1016/j.camwa.2004.08.015

Kumar, M. P., Du, J., Lagoudas, G., Jiao, Y., Sawyer, A., Drummond, D. C., et al. (2018). Analysis of single-Cell RNA-Seq identifies Cell-Cell communication associated with tumor characteristics. *Cell Rep.* 25:e1454.

Lambert, A. W., Pattabiraman, D. R., and Weinberg, R. A. (2017). Emerging biological principles of metastasis. *Cell* 168, 670–691. doi: 10.1016/j.cell.2016.11.037

Lander, A. D., Gokoffski, K. K., Wan, F. Y., Nie, Q., and Calof, A. L. (2009). Cell lineages and the logic of proliferative control. *PLoS Biol.* 7:e15. doi: 10.1371/journal.pbio.1000015

Li, C., and Balazsi, G. (2018). A landscape view on the interplay between EMT and cancer metastasis. *NPJ Systems Biol. Appl.* 4:34.

Li, C., Hong, T., and Nie, Q. (2016). Quantifying the landscape and kinetic paths for epithelial-mesenchymal transition from a core circuit. *Phys. Chem. Chem. Phys.* 18, 17949–17956. doi: 10.1039/c6cp03174a

Lo, W. C., Chou, C. S., Gokoffski, K. K., Wan, F. Y., Lander, A. D., Calof, A. L., et al. (2009). Feedback regulation in multistage cell lineages. *Math. Biosci. Eng.* 6, 59–82. doi: 10.3934/mbe.2009.6.59

Moustakas, A., and Heldin, C. H. (2007). Signaling networks guiding epithelial-mesenchymal transitions during embryogenesis and cancer progression. *Cancer Sci.* 98, 1512–1520. doi: 10.1111/j.1349-7006.2007.00550.x

Nieto, M. A., Huang, R. Y., Jackson, R. A., and Thiery, J. P. (2016). Emt: 2016. *Cell* 166, 21–45.

Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., et al. (2018). Identification of the tumour transition states occurring during EMT. *Nature* 556, 463–468. doi: 10.1038/s41586-018-0040-3

Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., et al. (2017). Single-Cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck Cancer. *Cell* 171:e1624.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., et al. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14, 979–982. doi: 10.1038/nmeth.4402

Ramirez, D., Kohar, V., and Lu, M. (2020). Toward modeling context-specific emt regulatory networks using temporal single cell RNA-Seq data. *Front. Mol. Biosci.* 7:54. doi: 10.3389/fmolb.2020.00054

Sha, Y., Haensel, D., Gutierrez, G., Du, H., Dai, X., and Nie, Q. (2019). Intermediate cell states in epithelial-to-mesenchymal transition. *Phys. Biol.* 16:021001. doi: 10.1088/1478-3975/aaf928

Sha, Y., Wang, S., Zhou, P., and Nie, Q. (2020). Inference and multiscale model of epithelial-to-mesenchymal transition via single-cell transcriptomic data. *Nucleic Acids Res.* 48, 9505–9520. doi: 10.1093/nar/gkaa725

Ta, C. H., Nie, Q., and Hong, T. (2016). Controlling stochasticity in epithelial-mesenchymal transition through multiple intermediate cellular states. *Discrete Continuous Dynamical Systems. Series B* 21:2275. doi: 10.3934/dcdsb.2016047

Tanaka, H., and Ogishima, S. (2015). Network biology approach to epithelial-mesenchymal transition in cancer metastasis: three stage theory. *J. Mol. Cell Biol.* 7, 253–266. doi: 10.1093/jmcb/mjv035

The Gene and Ontology Consortium, (2019). The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47, D330–D338.

Tian, X.-J., Ferro, M. V., and Goetz, H. (2019). "Modeling ncRNA-mediated circuits in cell fate decision," in *Computational Biology of Non-Coding RNA*, (New York, NY: Humana Press), 411–426. doi: 10.1007/978-1-4939-8982-9_16

Tripathi, S., Levine, H., and Jolly, M. K. (2020). The physics of cellular decision making during epithelial-mesenchymal transition. *Annu. Rev. Biophys.* 49, 1–18. doi: 10.1146/annurev-biophys-121219-081557

Wang, S., Drummond, M. L., Guerrero-Juarez, C. F., Tarapore, E., Maclean, A. L., Stabell, A. R., et al. (2020a). Single cell transcriptomics of human epidermis identifies basal stem cell transition states. *Nat. Commun.* 11:4239.

Wang, W., Douglas, D., Zhang, J., Chen, Y.-J., Cheng, Y.-Y., Kumari, S., et al. (2020b). Live cell imaging and analysis reveal cell phenotypic transition dynamics inherently missing in snapshot data. *bioRxiv* [preprint] doi: 10.1101/2019.12.12.874248

Wang, S., Karikomi, M., Maclean, A. L., and Nie, Q. (2019a). Cell lineage and communication network inference via optimization for single-cell transcriptomics. *Nucleic Acids Res.* 47:e66. doi: 10.1093/nar/gkz204

Wang, Y., Wang, R., Zhang, S., Song, S., Jiang, C., Han, G., et al. (2019b). iTALK: an R Package to characterize and illustrate intercellular communication. *bioRxiv* [preprint] doi: 10.1101/507871

Wendt, M. K., Allington, T. M., and Schiemann, W. P. (2009). Mechanisms of the epithelial-mesenchymal transition by TGF-beta. *Future Oncol.* 5, 1145–1168.

Xing, J., and Tian, X. J. (2019). Investigating epithelial-to-mesenchymal transition with integrated computational and experimental approaches. *Phys. Biol.* 16:031001. doi: 10.1088/1478-3975/ab0032

Xu, J., Lamouille, S., and Derynck, R. (2009). TGF-beta-induced epithelial to mesenchymal transition. *Cell Res.* 19, 156–172.

Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., et al. (2020). Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* 21, 341–352.

Zepp, J. A., Zacharias, W. J., Frank, D. B., Cavanaugh, C. A., Zhou, S., Morley, M. P., et al. (2017). Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. *Cell* 170:e1110.

Zhang, J., Tian, X.-J., and Xing, J. (2016). Signal transduction pathways of EMT induced by TGF-β, SHH, and WNT and their crosstalks. *J. Clin. Med.* 5:41. doi: 10.3390/jcm5040041

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10:1523.

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.

frontiers in Cell and Developmental Biology

MINI REVIEW
published: 12 January 2021
doi: 10.3389/fcell.2020.616592

What Will B Will B: Identifying Molecular Determinants of Diverse B-Cell Fate Decisions Through Systems Biology

*Simon Mitchell**

Brighton and Sussex Medical School, University of Sussex, Brighton, United Kingdom

B-cells are the poster child for cellular diversity and heterogeneity. The diverse repertoire of B lymphocytes, each expressing unique antigen receptors, provides broad protection against pathogens. However, B-cell diversity goes beyond unique antigen receptors. Side-stepping B-cell receptor (BCR) diversity through BCR-independent stimuli or engineered organisms with monoclonal BCRs still results in seemingly identical B-cells reaching a wide variety of fates in response to the same challenge. Identifying to what extent the molecular state of a B-cell determines its fate is key to gaining a predictive understanding of B-cells and consequently the ability to control them with targeted therapies. Signals received by B-cells through transmembrane receptors converge on intracellular molecular signaling networks, which control whether each B-cell divides, dies, or differentiates into a number of antibody-secreting distinct B-cell subtypes. The signaling networks that interpret these signals are well known to be susceptible to molecular variability and noise, providing a potential source of diversity in cell fate decisions. Iterative mathematical modeling and experimental studies have provided quantitative insight into how B-cells achieve distinct fates in response to pathogenic stimuli. Here, we review how systems biology modeling of B-cells, and the molecular signaling networks controlling their fates, is revealing the key determinants of cell-to-cell variability in B-cell destiny.

Keywords: systems biology, B-cells, computational modeling, heterogeneity, cell signaling, cell fate, NF-κB, cell-to-cell variability

OPEN ACCESS

Edited by:
Mohit Kumar Jolly,
Indian Institute of Science (IISc), India

Reviewed by:
Leonard Harris,
University of Arkansas, United States
Budhaditya Chatterjee,
Indian Institute of Science (IISc), India

***Correspondence:**
Simon Mitchell
s.a.mitchell@bsms.ac.uk

Specialty section:
This article was submitted to Molecular Medicine, a section of the journal Frontiers in Cell and Developmental Biology

Received: *12 October 2020*
Accepted: *02 December 2020*
Published: *12 January 2021*

Citation:
Mitchell S (2021) What Will B Will B: Identifying Molecular Determinants of Diverse B-Cell Fate Decisions Through Systems Biology. Front. Cell Dev. Biol. 8:616592. doi: 10.3389/fcell.2020.616592

INTRODUCTION

Following antigen exposure, B-cells are activated, often with the help of T-cells, to secrete antibodies essential for resolving infections. In addition to this well-studied humoral immune function, an important role for B-cells in cellular immunity is emerging (Hoffman et al., 2016). B-cell diversity is vital, with loss of diversity correlating with frailty and reductions in overall survival (Gibson et al., 2009). Each B-cell's destiny can range from apoptosis within hours, rapid differentiation into a short-lived plasma blast in the initial days of an infection (Lam and Baumgarth, 2019), to that of a memory B-cell surviving for decades (Seifert and Küppers, 2016), or long-lived plasma cells found

in the bone marrow > 40 years after vaccination (Brynjolfsson et al., 2017). Even B-cells stimulated *ex vivo*, without the complexities of T-cells and the germinal center, will undergo varied fates (Hawkins et al., 2009; Mitchell et al., 2018). Single-cell measurements of B-cells, stimulated with B-cell receptor (BCR)-independent stimuli, show vast cell-to-cell heterogeneity (Shih et al., 2002; Hawkins et al., 2013). Therefore, it seems that non-genetic B-cell diversity is an intrinsic property of B-cells. This has led to substantial efforts to identify the molecular determinants of B-cell destiny with pivotal studies combining insight from experimental models with *in silico* systems biology models. We will first discuss molecular determinants of each fate decision in isolation, followed by the molecular signaling pathways that interpret the cell's environment. Finally, we will put the pieces together to describe how cell-to-cell variability in B-cell fates is understood through systems biology.

CELL CYCLE

In response to antigen challenge, the B-cell population expands due to a portion of the cell population undergoing repeated rounds of cell division. *In vitro*, between 0 and 8 divisions occur, while multiple rounds of proliferation in the germinal center can lead to substantially higher (30+) divisions (Duffy et al., 2012; Tas et al., 2016; Mitchell and Hoffmann, 2018).

Mathematical models have been central to studies of the cell cycle since the 1960s, starting with phenomenological models recapitulating cell-cycle phase transitions (Smith and Martin, 1973). Dowling et al. (2014) observed that time spent in both G1 and S/G2/M phases is highly variable in B-cells. As a result, they proposed an alternative to the highly influential Smith–Martin model, in which all phases of the cell cycle stretch depending on a stochastically determined total division time (Smith and Martin, 1973). The timing of cell-cycle phases was found to be highly correlated in sister cells, suggesting a pre-existing non-genetic source of variability strongly inherited through cell division (Dowling et al., 2014). Interestingly, this stretching of all cell-cycle phases proportional to total cell-cycle length does not seem to be maintained in B lymphoma cell lines (Pham et al., 2018). An inherited molecular source of cell-to-cell variability is consistent with results from lineage-tracking of division times across multiple generations in proliferating B lymphocytes (Duffy et al., 2012; Mitchell et al., 2018). Heinzel et al. (2017) identified c-Myc as this molecule and fit a mathematical model to experimental data based on distributed c-Myc controlling a distributed division destiny. B-cell-specific modeling of cell division has been restricted to phenomenological modeling without explicitly representing molecular processes (Callard and Hodgkin, 2007; Zilman et al., 2010).

Kinetic modeling of the eukaryotic cell cycle became possible as increasing molecular mechanistic detail was revealed in the 1990s (Novak and Tyson, 1993; Csikasz-Nagy, 2009). The foundations for this progress was provided by the seminal work of Novák and Tyson, whose models have a striking ability to generate predictions validated many years later by experiments (Pomerening et al., 2003; Sha et al., 2003; Novák and Tyson, 2004). By adapting metabolic control analysis approaches to this model of the cell cycle, Conradie et al. (2010) found that variation in Cdk2 and its interactions with cyclin-dependent kinase inhibitor ($p27^{Kip1}$) and CyclinE were the most likely sources of cell-to-cell variability in the cell cycle. Later, live-cell Cdk2 tracking identified a bifurcation in Cdk2 trajectories, controlled by p27, as a source of cell-to-cell heterogeneity (Spencer et al., 2013; **Figure 1**). Despite the fact that much of this mechanistic insight has been generated from models of non-lymphatic cell lines, the ability of mechanistic cell-cycle models to generate insights into multiple model organisms from yeast to xenopus suggests that the molecular architecture of such models can also be informative in B-cells (Pomerening et al., 2005; Skotheim et al., 2008). Indeed, a generic model of the mammalian cell cycle was incorporated into a multiscale B-cell model by Shokhirev et al. (2015), which replicated single-cell B-cell proliferation measured by time-lapse microscopy.

CELL DEATH

Mounting an antibody response requires a balance of B-cell proliferation and cell death. Inadequate apoptosis leads to auto-immunity and malignancies while excess apoptosis can cause immunodeficiency (Cossu, 2010; Correia et al., 2015; Li et al., 2016; Yang et al., 2016). Mathematical models have been widely used to illuminate the cell-to-cell variation in the timing of apoptosis (Spencer and Sorger, 2011). The first kinetic models of apoptosis were published two decades ago (Fussenegger et al., 2000), as single-cell analysis revealed switch-like effector caspase kinetics in individual cells (Goldstein et al., 2000; Tyas et al., 2000). This switch-like behavior motivated construction of computational models, with multiple mechanisms being proposed from receptor clustering to cooperativity in apoptosome formation or pore formation (Eissing et al., 2004; Hua et al., 2005; Bagci et al., 2006; Legewie et al., 2006; Ho and Harrington, 2010). Comprehensive mechanistic models constructed by Peter Sorger's group, combined with single-cell fate tracking, have been instrumental in understanding cell-to-cell variability in apoptosis (Albeck et al., 2008; Spencer et al., 2009). These studies attributed variability in cell death timings to differences in BID truncation and a threshold determined by the Bcl-2 family proteins (Spencer et al., 2009; **Figure 1**). Apoptosis timing was found to be correlated in recently divided cells, but correlations between sister cells were lost quickly following cell division ($t_{1/2}$ = 1.5 h) due to intrinsic gene expression noise (Spencer et al., 2009). Similar analysis in B-cells found similar transient correlations in sibling cell apoptosis timing (Hawkins et al., 2009). This may seem inconsistent with results showing that, in proliferating B-cells, the majority of cells that are progeny of a single founder cell will undergo apoptosis in the same generation, even following 100+ hours of proliferation (Hawkins et al., 2009; Mitchell et al., 2018). It seems that the cell-to-cell variability in the decision to undergo apoptosis in a particular generation, and the precise timing of apoptosis, have distinct sources. This is consistent with an analysis of the Albeck et al. (2008) model performed by

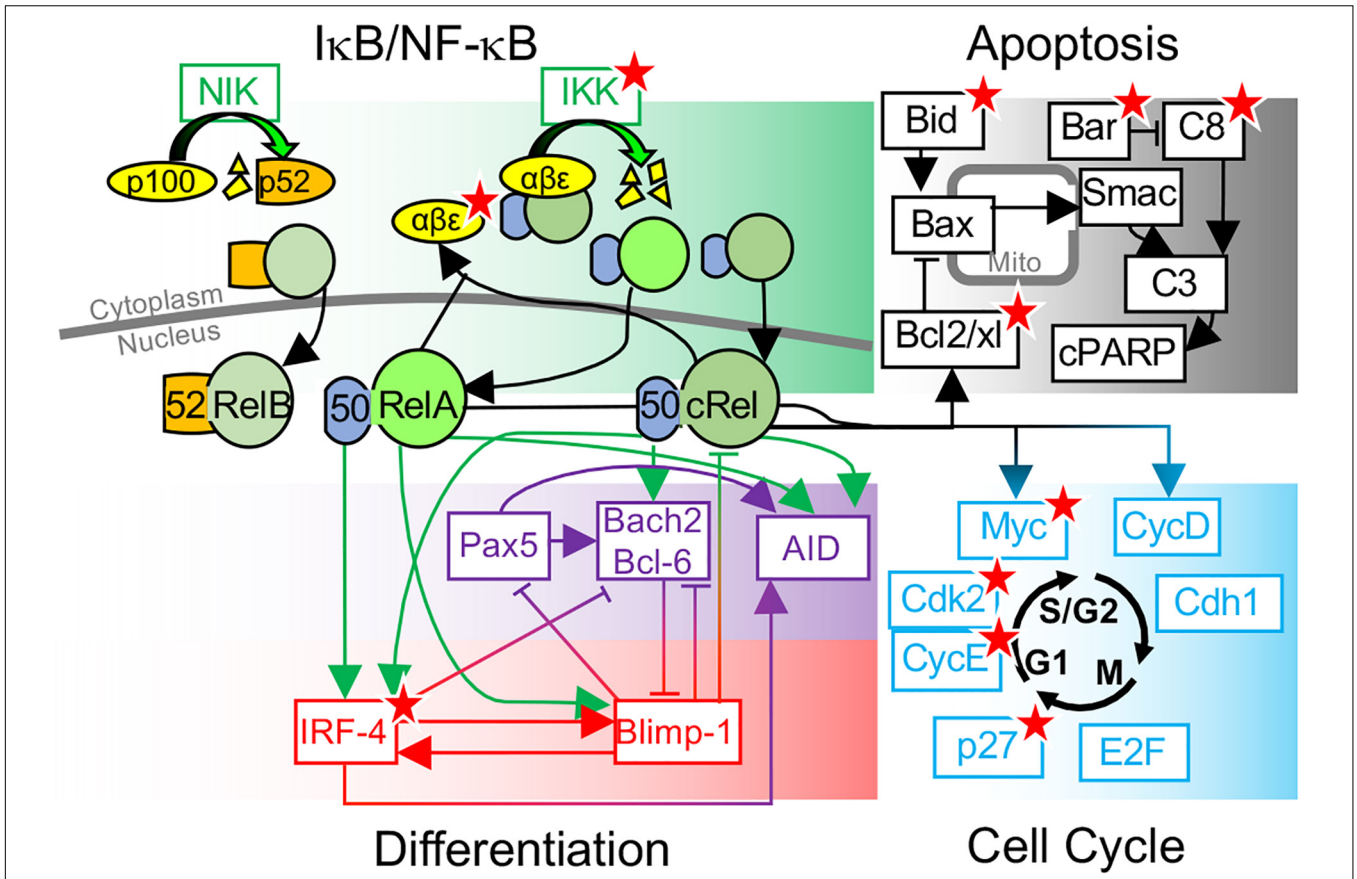

FIGURE 1 | Overview of key molecular determinants of B-cell fate decisions. Schematic of B-cell fate decisions determined by molecular networks controlling NF-κB, Apoptosis, differentiation, and the cell cycle. Key molecular determinants of cell-to-cell variability in B-cell fate decisions, identified through systems biology modeling and experimental studies, are indicated (★).

Loriaux et al. (2013), which found that molecular determinants of the timing of apoptosis are not equivalent to molecular determinants of whether or not a cell undergoes apoptosis. This analysis suggests that Procaspase 8 and its negative regulator Bar are key determinants of cell-to-cell differences in apoptotic decision making (Loriaux et al., 2013). Recent Luria–Delbrück analysis of gene expression revealed that the set of genes whose expression is reliably inherited differs between cell types (Luria and Delbrück, 1943; Shaffer et al., 2020). Therefore, key to predictive mechanistic modeling of B-cell apoptosis will be understanding the molecular determinants of apoptotic decisions, along with how reliably these factors are inherited during proliferation.

B-CELL DIFFERENTIATION

Following proliferation, activated B-cells differentiate into short-lived plasma blasts or long-lived plasma cells, both of which are antibody-secreting cells (ASCs) (Shapiro-Shelef and Calame, 2005). Cell division is required but not sufficient for ASC production, and considerable cell-to-cell differences in the timing of differentiation following activation exist, even *in vitro* (Tangye and Hodgkin, 2004; Zhou et al., 2018).

Recent single-cell RNAseq data indicate a bifurcation during the early stages of B-cell activation, committing a portion of cells to an ASC destiny (Scharer et al., 2020). This requires Interferon Regulatory Factor 4 (IRF4) induction, with higher and sustained activation biasing cells toward ASC fates (Ochiai et al., 2013). This was also seen by Xu et al. (2015) who constructed a minimal mathematical model of mutual inhibition between IRF4 and IRF8 in B-cells, with initial conditions obtained by flow cytometry, and found bifurcating fates recreating experiments showing a fraction of cells undergo rapid differentiation into plasma blasts. Sciammas et al. (2011) modeled the core regulatory network controlling terminal differentiation of activated B-cells including the mutual inhibition between Blimp1 and Bcl6/Bach2, along with the incoherent effects of IRF4 activating both somatic hypermutation (through AID) and differentiation (through Blimp1). This molecular model was incorporated into multiscale stochastic simulations, which revealed that differences in the time spent undergoing class-switch recombination and somatic hypermutation could be explained by the initial rate of IRF4 activation (Sciammas et al., 2011). Subsequent kinetic modeling

found that interactions between Irf4, Bcl6, and Blimp1 were sufficient to capture a broad variety of B-cell differentiation dynamics (Martínez et al., 2012). Taken together, these results show that cell-to-cell differences in terminal differentiation of B-cells result from differences in IRF4 signaling.

NF-κB

NF-κB is a dimeric transcription factor, first discovered in B-cells and later revealed to have near-ubiquitous expression (Sen and Baltimore, 1986; William et al., 1995; Xu et al., 1996; Inlay et al., 2002; Baltimore, 2009). NF-κB's important role in B-cell development, survival, and function has been widely studied (Vallabhapurapu and Karin, 2009; Gerondakis and Siebenlist, 2010; Kaileh and Sen, 2012; Heise et al., 2014; Almaden et al., 2016). In response to increasing BCR activation, B-cells show a digital all-or-nothing NF-κB response, with an increasing number of cells responding, rather than each cell increasing its response, with increasing NF-κB (Shinohara et al., 2014). The all-or-nothing response suggests the presence of a positive feedback loop, enabling cells that cross a cell-specific threshold of activation to invariably achieve maximum activation. Through iterative computational and experimental modeling, a positive feedback was identified between TAK1 (MAP3K7) and inhibitor of NF-κB (IκB) kinase-β (IKKβ) complex, resulting in switch-like single-cell behaviors; disruption of this feedback results in a more graded response (Shinohara et al., 2014). These all-or-nothing responses are consistent with studies applying information theoretic approaches to NF-κB signaling, which reveal that intrinsic noise in NF-κB limits the information the pathway can encode about each cell's environment to only a few states, e.g., absence, low and high stimuli (Cheong et al., 2011; Selimkhanov et al., 2014; Mitchell and Hoffmann, 2018). It seems unlikely that the intricate environmental stimuli received by B-cells through diverse receptors can be accurately encoded through noisy NF-κB signaling in single cells (Rawlings et al., 2012). This may be reconciled by a model-aided analysis that revealed a trade-off between reliable single-cell responses and reliable population-scale responses, with distributed switch-like responses enabling an appropriate fraction of cells within a population to reliably respond (Suderman et al., 2017).

Core to NF-κB signaling is its regulation through sequestration in the cytoplasm by inhibitory proteins (IκBs) (Mitchell et al., 2016). IκBs are themselves induced by nuclear NF-κB, resulting in a negative feedback in which NF-κB inhibits itself with a delay due to gene expression and protein synthesis (**Figure 1**). Such systems can create the oscillatory dynamics seen in NF-κB signaling, and mathematical modeling has been central to understanding NF-κB (Hoffmann et al., 2002; Lipniacki et al., 2004; Nelson et al., 2004; Basak et al., 2012). Each IκB family member has distinct kinetics of induction, degradation, and NF-κB sequestration, resulting in distinct contributions to cell-to-cell variability. IκBα displays rapid and robust stimulus-dependent degradation and subsequent NF-κB-dependent induction, creating a noise-insensitive first peak of NF-κB activity. IκBε has slower kinetics than IκBα (Kearns et al., 2006). Incorporating IκBε with slower negative feedback into mathematical simulations revealed that IκBε enables a more reliable dose-dependent response to sustained signals, minimizing the impact of stochastic gene expression on late-phase NF-κB activity (Longo et al., 2013). Through both kinetic modeling and experimental investigation, IκBε has been found to limit B-cell expansion through limiting NF-κB cRel and RelA (Alves et al., 2014).

Whether cell-to-cell differences in NF-κB signaling result from intrinsically generated noise, such as transcriptional noise, or pre-existing differences between B-cells prior to stimulation has been debated (Williams et al., 2014). Both sources of variation have been simulated through mathematical modeling of NF-κB, with intrinsic noise recreated through stochastic simulation using the Gillespie algorithm (Gillespie, 1977) and pre-existing variability simulated by sampling parameters prior to deterministic simulations (Hayot and Jayaprakash, 2006; Cheng et al., 2015; Hughey et al., 2015). Recent studies combining mathematical modeling with single-cell analysis find that pre-existing cell-to-cell differences best explain distributed single-cell NF-κB dynamics and the similar responses observed in daughter cells (Cheng et al., 2015; Hughey et al., 2015).

PUTTING THE PIECES TOGETHER

In studying the regulatory networks controlling the B-cell fate decisions described above, a pattern emerges. Key molecular determinants of cell-to-cell variability in B-cell fate decisions are NF-κB target genes. Indeed, recent single-cell RNA-sequencing analysis found that the most highly variable genes in lymphoid cells were functionally significant and centered around NF-κB and its target genes, including NFKBIA, MYC, IRF4, and AID (Osorio et al., 2020).

Myc and Bcl2 are NF-κB target genes that have been shown to control B-cell division and apoptosis (Duyao et al., 1990; Chen et al., 1999; Catz and Johnson, 2001; **Figure 1**). This was used by Shokhirev et al. (2015) in order to connect models of NF-κB signaling, the cell cycle, and apoptosis networks discussed above, recapitulating cellular statistics from single-cell time-lapse microscopy and revealing that NF-κB cRel was essential to protect growing B-cells from apoptosis. Mitchell et al. (2018) used this model to determine the source of cell-to-cell variability using single-cell lineage tracking experiments and discovered that B-cell fates were determined by molecular differences in the naïve B-cell population that are reliably inherited during proliferation. Interestingly, predictions of the most significant molecular determinants of cell-to-cell fate variability depend on the magnitude of variability. Perturbing parameters controlling NF-κB signaling resulted in the largest changes in B-cell proliferation; however, this required relatively large parameter changes of twofold or more. Smaller changes in parameters, and logistic regression on simulated cell populations with experimentally determined molecular heterogeneity, did not identify NF-κB-related biochemical processes as the largest determinants of

cell-to-cell variability in B-cell proliferation. Instead, apoptotic signaling regulators such as Bar, Caspase 3, and XIAP were predicted to be the most significant determinants of B-cell proliferative outcome, a result tested through caspase inhibition (Mitchell et al., 2018).

Key determinants of cell-to-cell variability in B-cell terminal differentiation including Blimp1 and IRF4 are also NF-κB target genes (Grumont and Gerondakis, 2000; Morgan et al., 2009; Heise et al., 2014). This led Roy et al. (2019) to add NF-κB regulation to the model of Sciammas et al. (2011) and discover that a previously unidentified regulatory interaction was required to recapitulate experimental results. Roy et al. (2019) discovered that the missing interaction was transcriptional inhibition of NF-κB cRel by Blimp1 and that dynamic downregulation of cRel by Blimp1 was required for plasma cell differentiation (Roy et al., 2019). Once this new regulatory interaction was incorporated into the multiscale model of Shokhirev et al. (2015), the model recapitulated cell-to-cell variability in B-cell proliferation and differentiation dynamics from wild-type and knockout mice (Roy et al., 2019). Given the overlap between NF-κB target genes and key determinants of B-cell fate decisions, well characterized cell-to-cell variability in NF-κB may coordinate diverse B-cell fates. Indeed, if Blimp1 upregulation time is noted in simulations from Roy et al. (2019), this model predicts that B-cells with the highest NF-κB RelA differentiate more quickly (**Figure 2**). As NF-κB integrates BCR and toll-like receptor signaling and induces IRF4, this prediction is consistent with the rapid differentiation by high-affinity BCR-expressing B-cells into plasma blasts (Paus et al., 2006) and the rapidly differentiating subset of cells with high IRF4 activation (Xu et al., 2015). Subsequent cRel downregulation is required to complete differentiation (Roy et al., 2019). The distinct roles of NF-κB cRel and RelA in B-cell survival and differentiation, respectively, seen in these multiscale models are consistent with *in vivo* requirements for germinal center maintenance and plasma cell generation (Heise et al., 2014) and an emerging

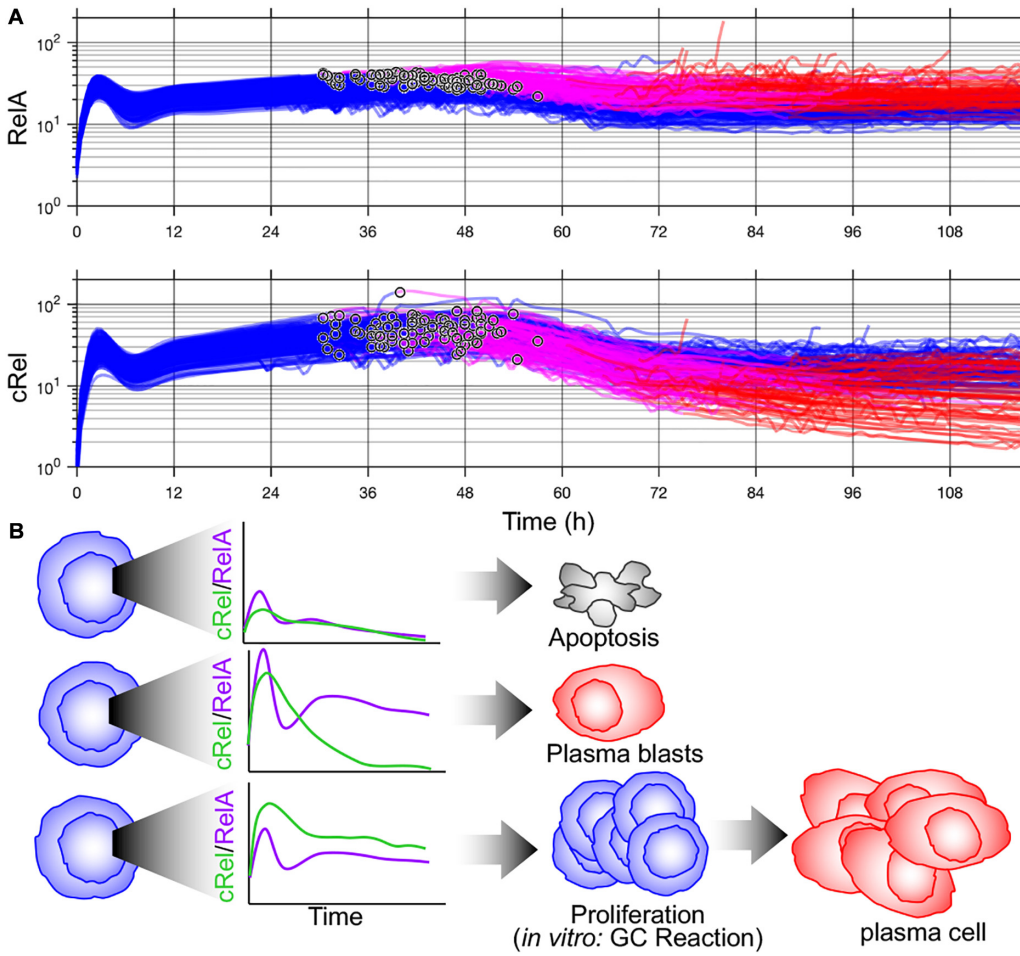

FIGURE 2 | Multiscale modeling of B-cell fates predicts cell-to-cell variability in NF-κB subunits that can orchestrate distinct fates. **(A)** Time course of NF-κB RelA (top) and NF-κB cRel (bottom) from multiscale simulation data from Mitchell et al. (2018). Blimp-1 upregulation time is indicated (○). Activated B-cells (blue), high Blimp1 (pink), and high Blimp1 with low AID (complete differentiation, red). **(B)** Proposed orchestration of cell fates through NF-κB. Inadequate NF-κB induction results in apoptosis (top). High RelA induction followed and subsequent dynamic cRel downregulation results in rapid plasma blast differentiation. High cRel with lower RelA (and therefore lower IRF4) results in a proliferative/germinal center phenotype.

picture of subunit-specific dysregulation of NF-κB in lymphoid malignancies (Kennedy and Klein, 2018).

DISCUSSION

The decision each B-cell faces, between proliferation, apoptosis, and differentiation, is vitally important to thread the needle between autoimmunity and immunodeficiency. An effective immune response requires a portion of B-cells to rapidly express antibodies, in order to buy time for the germinal center reaction to iteratively refine and expand an antigen-specific B-cell population to resolve the infection. Not only does this require careful coordination of multiple cell fates within each B-cell, but it requires cells to reach distinct decisions to the same challenge. Mathematical modeling provides unique opportunities to quantitatively disentangle the cell-intrinsic and extrinsic sources of cell-to-cell variability. Through combined modeling and single-cell experiments, we now know that distinct B-cell fates are achieved through molecular differences in the founder cell of each lineage, which are reliably inherited across many rounds of cell division (Hawkins et al., 2009; Mitchell et al., 2018).

B-cell differentiation takes place in the germinal centers of the spleen and lymph nodes. These structures spatially organize and traffic B-cells, enabling interactions with antigen-presenting cells and T-cells (De Silva and Klein, 2015; Mesin et al., 2016). Recently, these extra- and intercellular processes have been modeled through stochastic approaches (Thomas et al., 2019; Pélissier et al., 2020). Integrating the molecular determinants of B-cell fate decision into models of B-cell fates within the germinal center will be informative for therapeutic targeting of B-cells (Kepler and Perelson, 1993; Figge, 2005; Meyer-Hermann et al., 2012; Robert et al., 2017; Thomas et al., 2019; Pélissier et al., 2020; Verheijen et al., 2020).

The ultimate goal of many of the studies discussed here, and systems biology as a whole, is to use models to enable predictive control over cells in health and disease. While the emerging picture from experiments and models is that measuring one or even 5+ (Mitchell et al., 2018) molecular abundances is unlikely to reliably predict a B-cell's fate, this does not preclude reliable interventions. Modeling has identified molecular targets to control B-cell fates and predicted how mutations will skew proportions and timings of cell fate decisions in experimental systems (Mitchell et al., 2018; Roy et al., 2019). One challenge to predictive modeling is that many models have been parameterized in other cell types, and B-cell specific parameterization is daunting. However, the prevalence of single-cell data, along with promising model-generated experiment-validated results, suggests that a systems biology approach to predictably controlling B-cell responses is a realistic goal.

AUTHOR CONTRIBUTIONS

SM conceived the study and wrote the manuscript.

FUNDING

This work was funded by Leukaemia UK John Goldman Fellowship (2020/JGF/003) and Beat: Cancer Research Grant.

ACKNOWLEDGMENTS

We thank Alexander Hoffman, Koushik Roy, and Haripriya Vaidehi Narayanan for insightful discussion of the topic and many useful comments on the manuscript.

REFERENCES

Albeck, J. G., Burke, J. M., Spencer, S. L., Lauffenburger, D. A., and Sorger, P. K. (2008). Modeling a snap-action, variable-delay switch controlling extrinsic cell death. *PLoS Biol.* 6:e299. doi: 10.1371/journal.pbio.0060299

Almaden, J. V., Liu, Y. C., Yang, E., Otero, D. C., Birnbaum, H., Davis-Turak, J., et al. (2016). B-cell survival and development controlled by the coordination of NF-κB family members RelB and cRel. *Blood J. Am. Soc. Hematol.* 127, 1276–1286. doi: 10.1182/blood-2014-10-606988

Alves, B. N., Tsui, R., Almaden, J., Shokhirev, M. N., Davis-Turak, J., Fujimoto, J., et al. (2014). IκBε is a key regulator of B cell expansion by providing negative feedback on cRel and RelA in a stimulus-specific manner. *J. Immunol.* 192, 3121–3132. doi: 10.4049/jimmunol.1302351

Bagci, E., Vodovotz, Y., Billiar, T., Ermentrout, G., and Bahar, I. (2006). Bistability in apoptosis: roles of bax, bcl-2, and mitochondrial permeability transition pores. *Biophys. J.* 90, 1546–1559. doi: 10.1529/biophysj.105.068122

Baltimore, D. (2009). Discovering NF-κB. *Cold Spring Harbor Perspect. Biol.* 1:a000026. doi: 10.1101/cshperspect.a000026

Basak, S., Behar, M., and Hoffmann, A. (2012). Lessons from mathematically modeling the NF-κB pathway. *Immunol. Rev.* 246, 221–238. doi: 10.1111/j.1600-065x.2011.01092.x

Brynjolfsson, S. F., Mohaddes, M., Kärrholm, J., and Wick, M.-J. (2017). Long-lived plasma cells in human bone marrow can be either CD19+ or CD19–. *Blood Adv.* 1:835. doi: 10.1182/bloodadvances.2017004481

Callard, R., and Hodgkin, P. (2007). Modeling T-and B-cell growth and differentiation. *Immunol. Rev.* 216, 119–129.

Catz, S. D., and Johnson, J. L. (2001). Transcriptional regulation of bcl-2 by nuclear factor κB and its significance in prostate cancer. *Oncogene* 20, 7342–7351. doi: 10.1038/sj.onc.1204926

Chen, F., Demers, L. M., Vallyathan, V., Lu, Y., Castranova, V., and Shi, X. (1999). Involvement of 5′-Flanking κB-like Sites withinbcl-x Gene in Silica-induced Bcl-x Expression. *J. Biol. Chem.* 274, 35591–35595. doi: 10.1074/jbc.274.50.35591

Cheng, Z., Taylor, B., Ourthiague, D. R., and Hoffmann, A. (2015). Distinct single-cell signaling characteristics are conferred by the MyD88 and TRIF pathways during TLR4 activation. *Sci. Signal.* 8:ra69. doi: 10.1126/scisignal.aaa5208

Cheong, R., Rhee, A., Wang, C. J., Nemenman, I., and Levchenko, A. (2011). Information transduction capacity of noisy biochemical signaling networks. *Science* 334, 354–358. doi: 10.1126/science.1204553

Conradie, R., Bruggeman, F. J., Ciliberto, A., Csikász-Nagy, A., Novák, B., Westerhoff, H. V., et al. (2010). Restriction point control of the mammalian cell cycle via the cyclin E/Cdk2: p27 complex. *FEBS J.* 277, 357–367. doi: 10.1111/j.1742-4658.2009.07473.x

Correia, C., Schneider, P. A., Dai, H., Dogan, A., Maurer, M. J., Church, A. K., et al. (2015). BCL2 mutations are associated with increased risk of transformation and shortened survival in follicular lymphoma. *Blood J. Am. Soc. Hematol.* 125, 658–667. doi: 10.1182/blood-2014-04-571786

Cossu, F. (2010). Genetics of SCID. *Italian J. Pediatr.* 36:76.

Csikasz-Nagy, A. (2009). Computational systems biology of the cell cycle. *Briefings Bioinform.* 10, 424–434. doi: 10.1093/bib/bbp005

De Silva, N. S., and Klein, U. (2015). Dynamics of B cells in germinal centres. *Nat. Rev. Immunol.* 15, 137–148. doi: 10.1038/nri3804

Dowling, M. R., Kan, A., Heinzel, S., Zhou, J. H., Marchingo, J. M., Wellard, C. J., et al. (2014). Stretched cell cycle model for proliferating lymphocytes. *Proceed. Natl. Acad. Sci.* 111, 6377–6382. doi: 10.1073/pnas.1322420111

Duffy, K. R., Wellard, C. J., Markham, J. F., Zhou, J. H., Holmberg, R., Hawkins, E. D., et al. (2012). Activation-induced B cell fates are selected by intracellular stochastic competition. *Science* 335, 338–341. doi: 10.1126/science.1213230

Duyao, M. P., Buckler, A. J., and Sonenshein, G. E. (1990). Interaction of an NF-kappa B-like factor with a site upstream of the c-myc promoter. *Proc. Natl. Acad. Sci. U S A.* 87, 4727–4731. doi: 10.1073/pnas.87.12.4727

Eissing, T., Conzelmann, H., Gilles, E. D., Allgöwer, F., Bullinger, E., and Scheurich, P. (2004). Bistability analyses of a caspase activation model for receptor-induced apoptosis. *J. Biol. Chem.* 279, 36892–36897. doi: 10.1074/jbc.m404893200

Figge, M. T. (2005). Stochastic discrete event simulation of germinal center reactions. *Phys. Rev. E* 71:051907.

Fussenegger, M., Bailey, J. E., and Varner, J. (2000). A mathematical model of caspase function in apoptosis. *Nat. Biotechnol.* 18, 768–774. doi: 10.1038/77589

Gerondakis, S., and Siebenlist, U. (2010). Roles of the NF-κB pathway in lymphocyte development and function. *Cold Spring Harbor Perspect. Biol.* 2:a000182. doi: 10.1101/cshperspect.a000182

Gibson, K. L., Wu, Y. C., Barnett, Y., Duggan, O., Vaughan, R., Kondeatis, E., et al. (2009). B-cell diversity decreases in old age and is correlated with poor health status. *Aging Cell* 8, 18–25. doi: 10.1111/j.1474-9726.2008.00443.x

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361. doi: 10.1021/j100540a008

Goldstein, J. C., Waterhouse, N. J., Juin, P., Evan, G. I., and Green, D. R. (2000). The coordinate release of cytochrome c during apoptosis is rapid, complete and kinetically invariant. *Nat. Cell Biol.* 2, 156–162. doi: 10.1038/35004029

Grumont, R. J., and Gerondakis, S. (2000). Rel induces interferon regulatory factor 4 (IRF-4) expression in lymphocytes: modulation of interferon-regulated gene expression by rel/nuclear factor κB. *J. Exp. Med.* 191, 1281–1292. doi: 10.1084/jem.191.8.1281

Hawkins, E., Markham, J., Mcguinness, L., and Hodgkin, P. (2009). A single-cell pedigree analysis of alternative stochastic lymphocyte fates. *Proc. Natl. Acad. Sci.* 106, 13457–13462. doi: 10.1073/pnas.0905629106

Hawkins, E., Turner, M., Wellard, C., Zhou, J., Dowling, M., and Hodgkin, P. (2013). Quantal and graded stimulation of B lymphocytes as alternative strategies for regulating adaptive immune responses. *Nat. Commun.* 4:2406.

Hayot, F., and Jayaprakash, C. (2006). NF-κB oscillations and cell-to-cell variability. *J. Theoret. Biol.* 240, 583–591. doi: 10.1016/j.jtbi.2005.10.018

Heinzel, S., Giang, T. B., Kan, A., Marchingo, J. M., Lye, B. K., Corcoran, L. M., et al. (2017). A Myc-dependent division timer complements a cell-death timer to regulate T cell and B cell responses. *Nat. Immunol.* 18, 96–103. doi: 10.1038/ni.3598

Heise, N., De Silva, N. S., Silva, K., Carette, A., Simonetti, G., Pasparakis, M., et al. (2014). Germinal center B cell maintenance and differentiation are controlled by distinct NF-κB transcription factor subunits. *J. Exp. Med.* 211, 2103–2118. doi: 10.1084/jem.20132613

Ho, K. L., and Harrington, H. A. (2010). Bistability in apoptosis by receptor clustering. *PLoS Comput. Biol.* 6:e1000956. doi: 10.1371/journal.pcbi.1000956

Hoffman, W., Lakkis, F. G., and Chalasani, G. (2016). B cells, antibodies, and more. *Clin. J. Am. Soc. Nephrol.* 11, 137–154.

Hoffmann, A., Levchenko, A., Scott, M. L., and Baltimore, D. (2002). The IκB-NF-κB signaling module: temporal control and selective gene activation. *Science* 298, 1241–1245. doi: 10.1126/science.1071914

Hua, F., Cornejo, M. G., Cardone, M. H., Stokes, C. L., and Lauffenburger, D. A. (2005). Effects of Bcl-2 levels on Fas signaling-induced caspase-3 activation: molecular genetic tests of computational model predictions. *J. Immunol.* 175, 985–995. doi: 10.4049/jimmunol.175.2.985

Hughey, J. J., Gutschow, M. V., Bajar, B. T., and Covert, M. W. (2015). Single-cell variation leads to population invariance in NF-κB signaling dynamics. *Mol. Biol. Cell* 26, 583–590. doi: 10.1091/mbc.e14-08-1267

Inlay, M., Alt, F. W., Baltimore, D., and Xu, Y. (2002). Essential roles of the kappa light chain intronic enhancer and 3′ enhancer in kappa rearrangement and demethylation. *Nat. Immunol.* 3, 463–468. doi: 10.1038/ni790

Kaileh, M., and Sen, R. (2012). NF-κB function in B lymphocytes. *Immunol. Rev.* 246, 254–271. doi: 10.1111/j.1600-065x.2012.01106.x

Kearns, J. D., Basak, S., Werner, S. L., Huang, C. S., and Hoffmann, A. (2006). IκBε provides negative feedback to control NF-κB oscillations, signaling dynamics, and inflammatory gene expression. *J. Cell Biol.* 173, 659–664. doi: 10.1083/jcb.200510155

Kennedy, R., and Klein, U. (2018). Aberrant activation of NF-κB signalling in aggressive lymphoid malignancies. *Cells* 7:189. doi: 10.3390/cells7110189

Kepler, T. B., and Perelson, A. S. (1993). Cyclic re-entry of germinal center B cells and the efficiency of affinity maturation. *Immunol. Today* 14, 412–415. doi: 10.1016/0167-5699(93)90145-b

Lam, J. H., and Baumgarth, N. (2019). The multifaceted B cell response to influenza virus. *J. Immunol.* 202, 351–359. doi: 10.4049/jimmunol.1801208

Legewie, S., Blüthgen, N., and Herzel, H. (2006). Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. *PLoS Comput. Biol.* 2:e120. doi: 10.1371/journal.pcbi.0020120

Li, Y., Takahashi, Y., Fujii, S.-I., Zhou, Y., Hong, R., Suzuki, A., et al. (2016). EAF2 mediates germinal centre B-cell apoptosis to suppress excessive immune responses and prevent autoimmunity. *Nat. Commun.* 7: 10836.

Lipniacki, T., Paszek, P., Brasier, A. R., Luxon, B., and Kimmel, M. (2004). Mathematical model of NF-κB regulatory module. *J. Theoret. Biol.* 228, 195–215. doi: 10.1016/j.jtbi.2004.01.001

Longo, D. M., Selimkhanov, J., Kearns, J. D., Hasty, J., Hoffmann, A., and Tsimring, L. S. (2013). Dual delayed feedback provides sensitivity and robustness to the NF-κ B signaling module. *PLoS Comput. Biol.* 9:e1003112. doi: 10.1371/journal.pcbi.1003112

Loriaux, P. M., Tesler, G., and Hoffmann, A. (2013). Characterizing the relationship between steady state and response using analytical expressions for the steady states of mass action models. *PLoS Comput. Biol.* 9:e1002901. doi: 10.1371/journal.pcbi.1002901

Luria, S. E., and Delbrück, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28:491.

Martínez, M. R., Corradin, A., Klein, U., Álvarez, M. J., Toffolo, G. M., Di Camillo, B., et al. (2012). Quantitative modeling of the terminal differentiation of B cells and mechanisms of lymphomagenesis. *Proc. Natl. Acad. Sci. U S A.* 109, 2672–2677. doi: 10.1073/pnas.1113019109

Mesin, L., Ersching, J., and Victora, G. D. (2016). Germinal center B cell dynamics. *Immunity* 45, 471–482. doi: 10.1016/j.immuni.2016.09.001

Meyer-Hermann, M., Mohr, E., Pelletier, N., Zhang, Y., Victora, G. D., and Toellner, K.-M. (2012). A theory of germinal center B cell selection, division, and exit. *Cell Rep.* 2, 162–174. doi: 10.1016/j.celrep.2012.05.010

Mitchell, S., and Hoffmann, A. (2018). Identifying noise sources governing cell-to-cell variability. *Curr. Opin. Systems Biol.* 8, 39–45. doi: 10.1016/j.coisb.2017.11.013

Mitchell, S., Roy, K., Zangle, T. A., and Hoffmann, A. (2018). Nongenetic origins of cell-to-cell variability in B lymphocyte proliferation. *Proc. Natl. Acad. Sci. U S A.* 115, E2888–E2897.

Mitchell, S., Vargas, J., and Hoffmann, A. (2016). Signaling via the NFκB system. *Wiley Interdiscip. Rev. Systems Biol. Med.* 8, 227–241.

Morgan, M. A., Magnusdottir, E., Kuo, T. C., Tunyaplin, C., Harper, J., Arnold, S. J., et al. (2009). Blimp-1/Prdm1 alternative promoter usage during mouse development and plasma cell differentiation. *Mol. Cell. Biol.* 29, 5813–5827. doi: 10.1128/mcb.00670-09

Nelson, D. E., Ihekwaba, A. E. C., Elliott, M., Johnson, J. R., Gibney, C. A., Foreman, B. E., et al. (2004). Oscillations in NF-κB signaling control the dynamics of gene expression. *Science* 306, 704–708. doi: 10.1126/science.1099996

Novak, B., and Tyson, J. J. (1993). Numerical analysis of a comprehensive model of M-phase control in xenopus oocyte extracts and intact embryos. *J. Cell Sci.* 106, 1153–1168.

Novák, B., and Tyson, J. J. (2004). A model for restriction point control of the mammalian cell cycle. *J. Theoret. Biol.* 230, 563–579. doi: 10.1016/j.jtbi.2004.04.039

Ochiai, K., Maienschein-Cline, M., Simonetti, G., Chen, J., Rosenthal, R., Brink, R., et al. (2013). Transcriptional regulation of germinal center B and plasma cell fates by dynamical control of IRF4. *Immunity* 38, 918–929. doi: 10.1016/j.immuni.2013.04.009

Osorio, D., Yu, X., Zhong, Y., Li, G., Serpedin, E., Huang, J. Z., et al. (2020). Single-Cell expression variability implies cell function. *Cells* 9:14. doi: 10.3390/cells9010014

Paus, D., Phan, T. G., Chan, T. D., Gardam, S., Basten, A., and Brink, R. (2006). Antigen recognition strength regulates the choice between extrafollicular plasma cell and germinal center B cell differentiation. *J. Exp. Med.* 203, 1081–1091. doi: 10.1084/jem.20060087

Pélissier, A., Akrout, Y., Jahn, K., Kuipers, J., Klein, U., Beerenwinkel, N., et al. (2020). computational model reveals a stochastic mechanism behind germinal center clonal bursts. *Cells* 9:1448. doi: 10.3390/cells9061448

Pham, K., Kan, A., Whitehead, L., Hennessy, R., Rogers, K., and Hodgkin, P. (2018). Converse smith-martin cell cycle kinetics by transformed B lymphocytes. *Cell Cycle* 17, 2041–2051. doi: 10.1080/15384101.2018.1511511

Pomerening, J. R., Kim, S. Y., and Ferrell, J. E. Jr. (2005). Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations. *Cell* 122, 565–578. doi: 10.1016/j.cell.2005.06.016

Pomerening, J. R., Sontag, E. D., and Ferrell, J. E. (2003). Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nat. Cell Biol.* 5, 346–351. doi: 10.1038/ncb954

Rawlings, D. J., Schwartz, M. A., Jackson, S. W., and Meyer-Bahlburg, A. (2012). Integration of B cell responses through Toll-like receptors and antigen receptors. *Nat. Rev. Immunol.* 12, 282–294. doi: 10.1038/nri3190

Robert, P. A., Rastogi, A., Binder, S. C., and Meyer-Hermann, M. (2017). "How to simulate a germinal center," in *Germinal Centers. Methods in Molecular Biology*, ed. D. Calado (New York, NY: Humana Press), 303–334. doi: 10.1007/978-1-4939-7095-7_22

Roy, K., Mitchell, S., Liu, Y., Ohta, S., Lin, Y.-S., Metzig, M. O., et al. (2019). A regulatory circuit controlling the dynamics of NFκB cRel transitions B cells from proliferation to plasma cell differentiation. *Immunity* 50, 616.–628.

Scharer, C. D., Patterson, D. G., Mi, T., Price, M. J., Hicks, S. L., and Boss, J. M. (2020). Antibody-secreting cell destiny emerges during the initial stages of B-cell activation. *Nat. Commun.* 11:3989.

Sciammas, R., Li, Y., Warmflash, A., Song, Y., Dinner, A. R., and Singh, H. (2011). An incoherent regulatory network architecture that orchestrates B cell diversification in response to antigen signaling. *Mol. Systems Biol.* 7:495. doi: 10.1038/msb.2011.25

Seifert, M., and Küppers, R. (2016). Human memory B cells. *Leukemia* 30, 2283–2292.

Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A., et al. (2014). Accurate information transmission through dynamic biochemical signaling networks. *Science* 346, 1370–1373. doi: 10.1126/science.1254933

Sen, R., and Baltimore, D. (1986). Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell* 46, 705–716. doi: 10.1016/0092-8674(86)90346-6

Sha, W., Moore, J., Chen, K., Lassaletta, A. D., Yi, C.-S., Tyson, J. J., et al. (2003). Hysteresis drives cell-cycle transitions in *Xenopus laevis* egg extracts. *Proc. Natl. Acad. Sci. U S A.* 100, 975–980. doi: 10.1073/pnas.0235349100

Shaffer, S. M., Emert, B. L., Hueros, R. A. R., Coté, C., Harmange, G., Schaff, D. L., et al. (2020). Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors. *Cell* 182, 947–959. doi: 10.1016/j.cell.2020.07.003

Shapiro-Shelef, M., and Calame, K. (2005). Regulation of plasma-cell development. *Nat. Rev. Immunol.* 5, 230–242. doi: 10.1038/nri1572

Shih, T.-A. Y., Roederer, M., and Nussenzweig, M. C. (2002). Role of antigen receptor affinity in T cell–independent antibody responses in vivo. *Nat. Immunol.* 3, 399–406. doi: 10.1038/ni776

Shinohara, H., Behar, M., Inoue, K., Hiroshima, M., Yasuda, T., Nagashima, T., et al. (2014). Positive feedback within a kinase signaling complex functions as a switch mechanism for NF-κB activation. *Science* 344, 760–764. doi: 10.1126/science.1250020

Shokhirev, M. N., Almaden, J., Davis-Turak, J., Birnbaum, H. A., Russell, T. M., Vargas, J. A., et al. (2015). A multi-scale approach reveals that NF-κB cR el enforces a B-cell decision to divide. *Mol. Systems Biol.* 11:783. doi: 10.15252/msb.20145554

Skotheim, J. M., Di Talia, S., Siggia, E. D., and Cross, F. R. (2008). Positive feedback of G1 cyclins ensures coherent cell cycle entry. *Nature* 454, 291–296. doi: 10.1038/nature07118

Smith, J., and Martin, L. (1973). Do cells cycle? *Proc. Natl. Acad. Sci. U S A.* 70, 1263–1267.

Spencer, S. L., Cappell, S. D., Tsai, F.-C., Overton, K. W., Wang, C. L., and Meyer, T. (2013). The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. *Cell* 155, 369–383. doi: 10.1016/j.cell.2013.08.062

Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M., and Sorger, P. K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459, 428–432. doi: 10.1038/nature08012

Spencer, S. L., and Sorger, P. K. (2011). Measuring and modeling apoptosis in single cells. *Cell* 144, 926–939. doi: 10.1016/j.cell.2011.03.002

Suderman, R., Bachman, J. A., Smith, A., Sorger, P. K., and Deeds, E. J. (2017). Fundamental trade-offs between information flow in single cells and cellular populations. *Proc. Natl. Acad. Sci. U S A.* 114, 5755–5760. doi: 10.1073/pnas.1615660114

Tangye, S. G., and Hodgkin, P. D. (2004). Divide and conquer: the importance of cell division in regulating B-cell responses. *Immunology* 112, 509–520. doi: 10.1111/j.1365-2567.2004.01950.x

Tas, J. M., Mesin, L., Pasqual, G., Targ, S., Jacobsen, J. T., Mano, Y. M., et al. (2016). Visualizing antibody affinity maturation in germinal centers. *Science* 351, 1048–1054. doi: 10.1126/science.aad3439

Thomas, M. J., Klein, U., Lygeros, J., and Rodríguez Martínez, M. (2019). A probabilistic model of the germinal center reaction. *Front. Immunol.* 10:689. doi: 10.3389/fimmu.2019.00689

Tyas, L., Brophy, V. A., Pope, A., Rivett, A. J., and Tavaré, J. M. (2000). Rapid caspase-3 activation during apoptosis revealed using fluorescence-resonance energy transfer. *EMBO Rep.* 1, 266–270. doi: 10.1093/embo-reports/kvd050

Vallabhapurapu, S., and Karin, M. (2009). Regulation and function of NF-κB transcription factors in the immune system. *Ann. Rev. Immunol.* 27, 693–733. doi: 10.1146/annurev.immunol.021908.132641

Verheijen, M., Rane, S., Pearson, C., Yates, A. J., and Seddon, B. (2020). Fate mapping quantifies the dynamics of B cell development and activation throughout life. *Cell Rep.* 33:108376. doi: 10.1016/j.celrep.2020.108376

William, C. S., Liou, H.-C., Tuomanen, E. I., and Baltimore, D. (1995). Targeted disruption of the p50 subunit of NF-κB leads to multifocal defects in immune responses. *Cell* 80, 321–330. doi: 10.1016/0092-8674(95)90415-8

Williams, R. A., Timmis, J., and Qwarnstrom, E. E. (2014). Computational models of the NF-KB signalling pathway. *Computation* 2, 131–158. doi: 10.3390/computation2040131

Xu, H., Chaudhri, V. K., Wu, Z., Biliouris, K., Dienger-Stambaugh, K., Rochman, Y., et al. (2015). Regulation of bifurcating B cell trajectories by mutual antagonism between transcription factors IRF4 and IRF8. *Nat. Immunol.* 16, 1274–1281. doi: 10.1038/ni.3287

Xu, Y., Davidson, L., Alt, F. W., and Baltimore, D. (1996). Deletion of the Ig kappa light chain intronic enhancer/matrix attachment region impairs but does not abolish V kappa J kappa rearrangement. *Immunity* 4, 377–385. doi: 10.1016/s1074-7613(00)80251-4

Yang, J., Zhao, S., Yang, X., Zhang, H., Zheng, P., and Wu, H. (2016). Inhibition of B-cell apoptosis is mediated through increased expression of Bcl-2 in patients with rheumatoid arthritis. *Int. J. Rheumat. Dis.* 19, 134–140. doi: 10.1111/1756-185x.12706

Zhou, J. H., Markham, J. F., Duffy, K. R., and Hodgkin, P. D. (2018). Stochastically timed competition between division and differentiation fates regulates the transition from B lymphoblast to plasma cell. *Front. Immunol.* 9:2053. doi: 10.3389/fimmu.2018.02053

Zilman, A., Ganusov, V. V., and Perelson, A. S. (2010). Stochastic models of lymphocyte proliferation and death. *PLoS One* 5:e12775. doi: 10.1371/journal.pone.0012775

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership