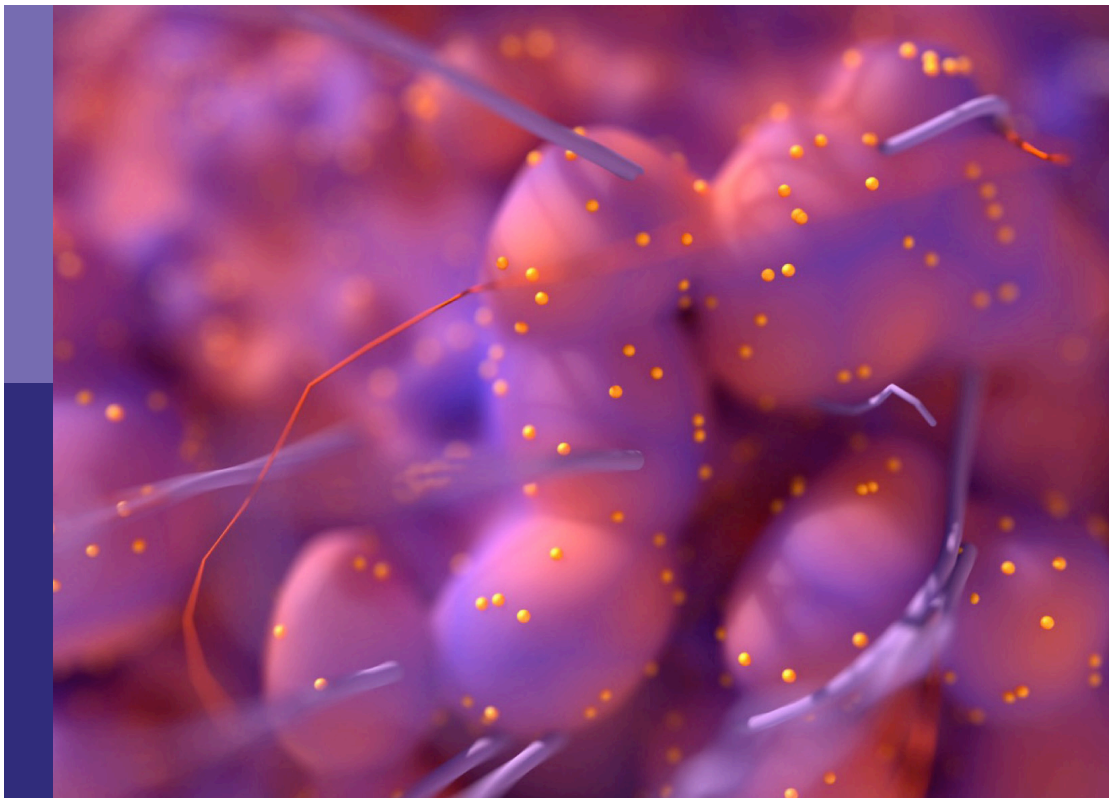
frontiers | Research Topics

Evolutionary biology of tumor initiation and progression

Edited by
Yusuke Suenaga, Erich Bornberg-Bauer and Mamoru Kato

Published in
Frontiers in Oncology



ISSN 1664-8714
ISBN 978-2-8325-5099-1
DOI 10.3389/978-2-8325-5099-1

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

Evolutionary biology of tumor initiation and progression

Topic editors

Yusuke Suenaga — Chiba Cancer Center, Japan

Erich Bornberg-Bauer — University of Münster, Germany

Mamoru Kato — National Cancer Centre, Japan

Citation

Suenaga, Y., Bornberg-Bauer, E., Kato, M., eds. (2024). *Evolutionary biology of tumor initiation and progression*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-5099-1

Table of contents

frontiers | Frontiers in Oncology

BRIEF RESEARCH REPORT
published: 08 July 2022
doi: 10.3389/fonc.2022.888114

Clonal Evolution at First Sight: A Combined Visualization of Diverse Diagnostic Methods Improves Understanding of Leukemic Progression

Sarah Sandmann[1†], Yvonne Lisa Behrens[2†], Claudia Davenport[2], Felicitas Thol[3], Michael Heuser[3], Daniela Dörfel[4], Friederike Löhr[5], Agnes Castrup[6], Doris Steinemann[2], Julian Varghese[1], Brigitte Schlegelberger[2], Martin Dugas[1,7] and Gudrun Göhring[2]*

[1] Institute of Medical Informatics, University of Münster, Münster, Germany, [2] Department of Human Genetics, Hannover Medical School, Hannover, Germany, [3] Department of Hematology, Hemostasis, Oncology and Stem Cell Transplantation, Hannover Medical School, Hannover, Germany, [4] Department of Hematology, Oncology and Immunology, Klinikum Region Hannover (KRH) Klinikum Siloah, Hannover, Germany, [5] Department of Hematology and Oncology, Klinikum Braunschweig, Braunschweig, Germany, [6] Hämato-Onkologische Praxis, Hämato-Onkologische Praxis im Medicum, Bremen, Germany, [7] Institute of Medical Informatics, Heidelberg University Hospital, Heidelberg, Germany

OPEN ACCESS

Edited by:
Robert Ohgami,
University of California, San Francisco, United States

Reviewed by:
Babal Jha,
Cleveland Clinic, United States
Michael Diamantidis,
University Hospital of Larissa, Greece

***Correspondence:**
Yvonne Lisa Behrens
Behrens.Yvonne@mh-hannover.de

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Hematologic Malignancies, a section of the journal Frontiers in Oncology

Received: 02 March 2022
Accepted: 10 June 2022
Published: 08 July 2022

Citation:
Sandmann S, Behrens YL, Davenport C, Thol F, Heuser M, Dörfel D, Löhr F, Castrup A, Steinemann D, Varghese J, Schlegelberger B, Dugas M and Göhring G (2022) Clonal Evolution at First Sight: A Combined Visualization of Diverse Diagnostic Methods Improves Understanding of Leukemic Progression. Front. Oncol. 12:888114. doi: 10.3389/fonc.2022.888114

Patients with myeloid neoplasia are classified by the WHO classification systems. Besides clinical and hematological criteria, cytogenetic and molecular genetic alterations highly impact treatment stratification. In routine diagnostics, a combination of methods is used to decipher different types of genetic variants. Eight patients were comprehensively analyzed using karyotyping, fluorescence *in situ* hybridization, array-CGH and a custom NGS panel. Clonal evolution was reconstructed manually, integrating all mutational information on single nucleotide variants (SNVs), insertions and deletions (indels), structural variants and copy number variants (CNVs). To allow a correct integration, we differentiate between three scenarios: 1) CNV occurring prior to the SNV/indel, but in the same cells. 2) SNV/indel occurring prior to the CNV, but in the same cells. 3) SNV/indel and CNV existing in parallel, independent of each other. Applying this bioinformatics approach, we reconstructed clonal evolution for all patients. This generalizable approach offers the possibility to integrate various data to analyze identification of driver and passenger mutations as well as possible targets for personalized medicine approaches. Furthermore, this model can be used to identify markers to assess the minimal residual disease.

Keywords: clonal evolution, bioinformatics, single nucleotide variants, copy number variants, cancer cell fraction, leukemic progression

INTRODUCTION

Myeloid neoplasia, including acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS) are heterogeneous hematopoietic stem cell disorders, which are marked by the acquisition of somatic alterations and clonal evolution (1–4). Besides clinical and hematological criteria, cytogenetic and molecular genetic alterations highly impact treatment stratification (2). In recent

years, high-throughput sequencing technologies have led to the identification of many driver and passenger alterations that enable precision medicine (e.g. *IDH1* inhibitor ivosidenib in AML patients) (2, 5–9). The different subgroups of myeloid neoplasia (e.g. MDS and AML) show a highly heterogeneous cytogenetic and molecular genetic profile. Multiple subclones may exist in one patient. Distribution of recurrently mutated genes and clonal architecture are different, e.g. in MDS/MPN subtypes (4, 10, 11). Consequently, leukemic progression may be influenced by several leukemic clones, and the clonal composition identified at diagnosis can differ from the composition identified at relapse or progression (12). Patients with myeloid neoplasia, especially patients with MDS, often progress to AML (sAML) through a process of clonal evolution (2, 11–13). Clonal evolution can be associated with poor prognosis, relapse and therapeutic resistance (14, 15). To this day different models of tumor evolution have been reported (e.g. linear, branching or neutral evolution) (16).

Recent publications have shown that reconstruction of clonal evolution and characterization of clonal architecture are important for the understanding of tumor development and treatment failure (14). Furthermore, they can help in finding new treatment strategies (e.g. combination of drugs, or new drugs for patients with relapse) (11, 12, 14). Different bioinformatic approaches already exist for the calculation of clonal evolution of bulk sequencing data. In addition, the single cell sequencing method has expanded the field of reconstructing clonal evolution (14, 17–19). A study of Morita et al. has shown both genetic and phenotypic evolution in AML by single cell sequencing and cell surface protein analyses (14). However, in routine diagnostics, a combination of methods (e.g. karyotyping and NGS panel) is used to identify alterations that are important for treatment stratification of patients with myeloid neoplasia, and which may not be detected using only a single method (20). These methods decipher different types of genetic variants, including single nucleotide variants (SNVs), insertions/deletions (indels), structural variants (SVs) and copy number variations (CNVs). The combination of a wide spectrum of methods could lead to the identification of many alterations, which, in turn, can make interpretation difficult. Unfortunately, an approach that combines all identified molecular genetic and cytogenetic aberrations and that reconstructs and visualizes the genetic architecture and clonal evolution, is still lacking.

In the present study, we performed karyotyping, fluorescence *in situ* hybridization, array-CGH and used a custom NGS panel in a cohort of eight patients with myeloid neoplasia. Here, we propose a bioinformatic approach as well as a visualization of occurrences of genetic alterations to improve our understanding of leukemic progression (driver and passenger alterations) and assist in finding the best treatment stratification paving the way for personalized medicine approaches. Clonal evolution and clonal architecture not only play a role in therapeutic resistance, relapse and poor outcomes in myeloid neoplasia, but also in other non-solid tumors (14). Our proposed bioinformatic model is a general approach, which may be used on all kinds on non-solid tumors showing clonal evolution.

MATERIALS AND METHODS

Study Population

A total of eight patients with myeloid neoplasia and clonal evolution were included in this study. The patients were analyzed between 2011 and 2021 at our department. The ethical review boards of Hannover Medical School approved this study and all patients gave their written consent. Our cohort was analyzed at one or more time points. All results of performed analyses as well as clinical data are shown in **Table 1** and in more detail in **Table S1**.

Karyotyping and Fluorescence *In-Situ* Hybridization (FISH) Analysis

Cytogenetic and molecular cytogenetic analyses were performed on bone marrow aspirates or peripheral blood cultures. Chromosome preparation and fluorescence R-banding were performed as described previously (21, 22). Altogether, whenever possible, 15 to 25 metaphases were examined per patient. The karyotype was described according to guidelines of the International System for Human Cytogenetic Nomenclature (23). Depending on the cytogenetic aberrations in each patient, FISH analyses on interphase nuclei were performed using: (1) break apart probes for the loci 3q26 (EVI1 (MECOM) Break Apart; Cytocell, Cambridge, UK), 11q23 and 12p13 (Vysis MLL and ETV6 FISH Probe Kit; Abbott, IL, USA); (2) dual color probes for the loci 5p15.2, 5q31, CEP7, 7q31, 9q34.1, 17p13.1, CEP17, 22q11.2 (Vysis EGR1/D5S23, D5S721, D7S486/CEP7, BCR/ABL and TP53/CEP 17 FISH Probe Kit; Abbott); and (3) single color probe for the locus 20q12 (Vysis D20S108 FISH Probe Kit; Abbott). At least 200 interphase nuclei were analyzed for each probe.

Array-CGH (Array-Based Comparative Genomic)

According to the manufacturer's protocol (e-Array design 84704, Agilent Technologies, Waldbronn, Germany), three patients (patients c1, 6 and 7) were screened for CNVs by high-resolution array-CGH. Microarray slides were scanned immediately using an Agilent microarray scanner at a resolution of 2 µm. Fluorescence ratios were calculated using Feature Extraction Software and copy number states analyzed using the CGH data analysis software Genomic Workbench (Agilent Technologies).

NGS With IDT Custom Panel

DNA sequencing was performed for all patients using an IDT custom panel (Integrated DNA Technologies Inc, Iowa, USA) including 148 leukemia-associated genes. Sequencing was performed on an Illumina NextSeq sequencer. All samples were aligned against the reference genome GRCh37 using BWA mem (24). The pipeline uses eight open-source tools for independent variant calling. Subsequently, raw calls are automatically combined and characterized with respect to data quality. Default settings were used (minimum number of reads 50, minimum number of reads with the alternate allele 20, minimum variant allele frequency VAF 0.01). Excluding low-quality calls, the remaining variants are

TABLE 1 | Overview of analyzed patients (n=8) (All results of performed analyses as well as clinical data are shown in **Table S1**).

patient (#)	WHO diagnosis	status	cytogenetic aberrations	DNA analysis NGS (somatic)			treatment
				gene	variant	variant allele frequeny (VAF)	
#1	AML*1	initial	del(5)(q14q33), -17, add(18)(q22),+mar1,+mar2,+mar3	*ASXL1*	c.2317G>T p.(Glu773*)	10.83%	not known
				DNMT3A	c.2645G>A p.(Arg882His)	18.69%	
				IDH1	c.394C>T p.(Arg132Cys)	13.85%	
				TP53	c.427G>A p.(Val143Met)	20.49%	
#2	MDS*2	initial	+8,+10	*BCOR*	c.4639+1G>A	66.76%	Azacitidine plus Venetoclax
				DNMT3A	c.2645G>A p.(Arg882His)	39.43%	
				KRAS	c.34G>C p.(Gly12Arg)	4.03%	
				SF3B1	c.2098A>G p.(Lys700Glu)	1.31%	
				STAG2	c.3362_3365dup p.(Ser1123Hisfs*14)	52.12%	
				U2AF1	c.101C>T p.(Ser34Phe)	31.50%	
#3	t-MN*6	initial	t(9;20)(q11;q11),t(12;22)(p13;q11)	*NF1*	c.2033del p.(Pro678Arg*10)	49.52%	CPY-351, allogenic stem cell planned
#4	t-MN	initial	del(5)(q14q34),+8,i(8)(q10)x2	*IDH2*	c.419G>A p.(Arg140Gln)	46.79%	Azacitidine plus Venetoclax
				RUNX1	c.420T>G p.(Ser140Arg)	43.00%	
				SRSF2	c.284C>T p.(Pro95Leu)	39.78%	
				TET2	c.1455del p.(Asn486Thr*11)	47.20%	
				TET2	c.3473del p.(Ala1158Glu*68)	41.43%	
				TP53	c.844C>T p.(Arg282Trp)	90.92%	
		progression	t(2;3)(p23;q27),del(5)(q14q34),del(7)(q21),+8,i(8)(q10)x2	*IDH2*	c.419G>A p.(Arg140Gln)	33.28%	
				KRAS	c.35G>C p.(Gly12Ala)	1.96%	
				NRAS	c.35G>C p.(Gly12Ala)	1.70%	
				PTPN11	c.1508G>C p.(Gly503Ala)	2.80%	
				RUNX1	c.420T>G p.(Ser140Arg)	38.29%	
				SRSF2	c.284C>T p.(Pro95Leu)	40.88%	
				TET2	c.1398_1402dup p.(His468Leu*20)	5.11%	
				TET2	c.1455del p.(Asn486Thr*11)	44.10%	
				TET2	c.3473del p.(Ala1158Glu*68)	36.71%	
				TP53	c.844C>T p.(Arg282Trp)	74.41%	
#5	AML	initial	t(4;14;11)(q22;q32;q23),add(10)(p14)	*ASXL1*	c.2077C>T p.(Arg693*)	0.68%	7+3
				KRAS	c.38G>A p.(Gly13Asp)	0.35%	
				NRAS		0.50%	

(Continued)

TABLE 1 | Continued

patient (#)	WHO diagnosis	status	cytogenetic aberrations	DNA analysis NGS (somatic)			treatment
				gene	variant	variant allele frequeny (VAF)	
					c.182A>G p.(Gln61Arg)		
		progression	t(4;14;11)(q22;q32;q23),+der(4)t(4;14;11),+8,+9,add(10)(p14),+19,+21	*ASXL1*	c.2077C>T p.(Arg693*)	7.84%	
				KRAS	c.38G>A p.(Gly13Asp)	7.59%	
				NRAS	c.182A>G p.(Gln61Arg)	5.34%	
#6	initial: MPN*3; 6-years later: suspicion on sAML*4	initial	–	*JAK2*	c.1849G>T p.(Val617Phe)	73.76%	Hydroxyurea & phlebotomy/blood letting; ~6 years later: Jakavi
		Progression	add(5)(q12),-7,-13,del(14)(q12q31),der(17)t(13;17)(q21;p12)	*JAK2*	c.1849G>T p.(Val617Phe)	97.53%	
				TP53	c.814G>A p.(Val272Met)	19.56%	
		progression	add(2)(q37),add(5)(q12),-7,-13,del(14)(q12q31),der(17)t(13;17)(q21;p12),del(20)(q12q13)	*JAK2*	c.1849G>T p.(Val617Phe)	98.12%	11 cycles Azacytidin + 2 cycles LD Ara-C
				TP53	c.814G>A p.(Val272Met)	23.36%	
#7	sAML from atypical CML*5	initial	ins(9;12)(q34;p12p13),+12	no variants detected			7+3 plus Dasatinib, allogenic Tx
		progression	+X,ins(9;12)(q34; p12p13), +11,+12,del(12)(p13),+19	no variants detected			
		remission	-	no variants detected			
		relapse	der(7)t(7;9)(q35;q21),ins(9;12)(q34;p12p13),+12	no variants detected			
#8	AML without matu-ration	initial	del(9)(q21q31)	*NRAS*	c.37G>C p.(Gly13Arg)	12.69%	7+3, allogenic transplantation, after rezidive: FLA-V-IDA (=FLAG-IDA with Venetoclax) plus donor lymphocytes
				WT1	c.1136_1142dup p.(Ala382Thr*5)	23.26%	
		remission	–	no variants detected			
		relapse	t(1;16)(p12;q21),del(9)(q21q31)	*NRAS*	c.37G>C p.(Gly13Arg)	12.78%	
				NRAS	c.37G>T p.(Gly13Cys)	5.03%	
				WT1	c.1141_1144dup p.(Ala382Val*4)	4.51%	
				WT1	c.1136_1142dup p.(Ala382Thr*5)	11.35%	
				WT1	c.1128dup p.(Thr377Asp*8)	1.48%	
				WT1	c.1110dup p.(Val371Cys*14)	3.69%	
		progression	t(1;16)(p12;q21),?add(9)(p12),del(9)(q21q31),add(10)(p12),add(17)(q22)	*NRAS*	c.37G>C p.(Gly13Arg)	17.38%	
				NRAS	c.37G>T p.(Gly13Cys)	4.93%	
				WT1	c.1141_1144dup p.(Ala382Val*4)	4.03%	
				WT1	c.1136_1142dup p.(Ala382Thr*5)	17.18%	
				WT1	c.1128dup p.(Thr377Asp*8)	4.76%	
				WT1	c.1110dup p.(Val371Cys*14)	4.96%	
		remission	–	no variants detected			

**1 AML, acute myeloid leukemia; *2 MDS, myelodysplastic syndrome; *3 MPN, myeloproliferative neoplasm; *4 sAML, secondary AML; *5 CML, chronic myeloid leukemia*

further characterized (e.g. presence in common databases and in silico effect prediction). Combining all information, an artifact- and a polymorphism score is calculated for every SNV and small indel. The scores allow for automatic classification of the calls as true variants, polymorphisms and artifacts. Additionally, all detected variants were manually checked using the Integrative Genomics Viewer (IGV) to confirm automatic classification (25, 26). All detected variants were classified according to the standards and guidelines of the American College of Medical Genetics and Genomics (ACMG) (27). Here, we report only variants of unknown significance (VUS), likely pathogenic and pathogenic variants.

Bioinformatic Approach

SNVs and short indels were detected using targeted NGS and variant calling pipeline appreci8 (28). SVs and CNVs were detected performing karyotyping, FISH and array-CGH analysis. Clonal evolution was reconstructed manually, integrating all mutational information on SNVs, indels, SVs and CNVs as described in Reutter et al. (29). The results were visualized by fishplots (30). In short, percentages of cancer cell fractions (CCFs) for SNVs and indels were estimated based on VAFs, assuming heterozygous variants (2*VAF=CCF). Percentages of CCFs for SVs and CNVs were estimated based on cell counts reported for karyotyping and FISH analyses. For SVs and CNVs that were only detected in array-CGH percentages of CCF were estimated based on $\log_2$Ratio of fluorescence intensities for reference versus tumor probes. In case of CNVs overlapping the position of an SNV or indel, the calculation of CCF is less straightforward. Altogether, we differentiate between three possible scenarios:

Scenario 1: The CNV occurred prior to the SNV/indel, but in the same cells.

Scenario 2: The SNV/indel occurred prior to the CNV, but in the same cells.

Scenario 3: SNV/indel and CNV exist in parallel, independent of each other.

We defined w as the ratio of cells featuring a CNV and an SNV/indel. Analogously, x, y and z were defined in the same manner. The sum over all cell ratio was always $w+x+y+z=1.0$. Furthermore, the known CCF for the CNV was defined as $CCF_{CNV}=w+x$. Additionally, two formulas were derived from the model (see **Figure 1A**):

$$VAF_{SNV/indel} = (cnv_{value} \cdot w + y)/ \quad (1)$$
$$(1 \cdot w + 1 \cdot x + 2 \cdot y + 2 \cdot z) \quad if \ \ CNV \ \ is \ \ deletion$$

$$VAF_{SNV/indel} = (cnv_{value} \cdot w + y)/ \quad (2)$$
$$(3 \cdot w + 3 \cdot x + 2 \cdot y + 2 \cdot z) \ \ if \ \ CNV \ \ is \ \ duplication$$

We defined cnv_{value} as a parameter to model the number of alleles with the SNV/indel depending on the CNV. For scenarios 1 and 3 the $cnv_{value}=1$ and for scenario 2 the cnv_{value} can be 0, 1 or 2. Dependent on the three scenarios we considered, the two equations may be further simplified: For scenario 1, the CNV occurring prior to the SNV/indel, $y=0$ (see **Figure 1B**). Thus, $CCF_{CNV}=w+x$ and $CCF_{SNV/indel}=w$. As $VAF_{SNV/indel}$ is also known, w and $CCF_{SNV/indel}$ can be easily determined. For scenario 2, a distinction of cases has to be made. Both, the allele with or without SNV/indel may be affected by CNV. Regarding deletions, $cnv_{value}=0$ if the allele with SNV/indel is deleted, and $cnv_{value}=1$ if the allele without SNV/indel is deleted. Regarding duplications, $cnv_{value}=2$ if the allele with SNV/indel is duplicated, and $cnv_{value}=1$ if the allele without SNV/indel is duplicated. Additionally, $x=0$, so $CCF_{CNV}=w$. The unknown

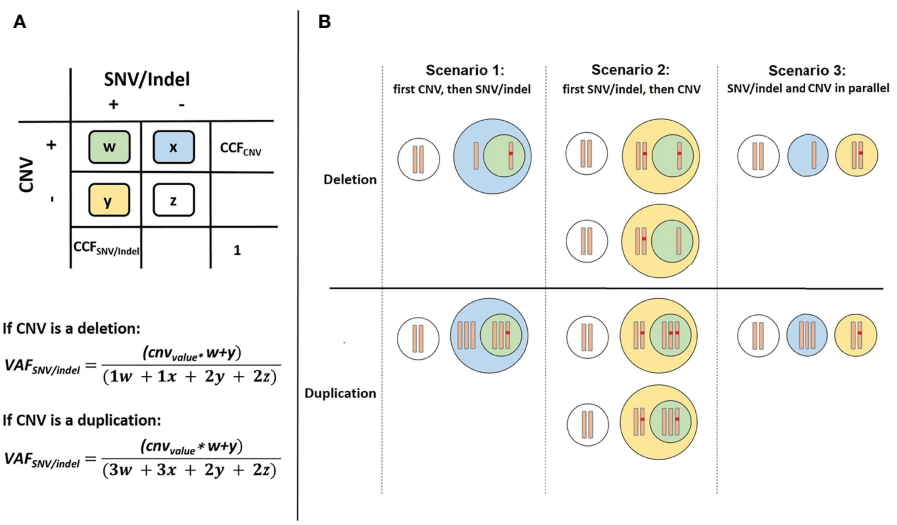


FIGURE 1 | Overview of cell types in case of a CNV overlapping the position of an SNV/indel. **(A)** Color scheme used for 4 possible cases. **(B)** Differentiation between the three possible scenarios.

value of y and thus $CCF_{SNV/indel}=w+y$ can be determined using $VAF_{SNV/indel}$, following the distinction of cases. For the third scenario of SNV/indel and CNV occurring in parallel, independent of each other, $w=0$. Thus, $CCF_{CNV}=x$ and $CCF_{SNV/indel}=y$. As $VAF_{SNV/indel}$ is known, the unknown value of $CCF_{SNV/indel}$ can be determined. Although $CCF_{SNV/indel}$ can always be determined using the above-mentioned formulas, the result may not always be distinct. Information on whether the CNV occurred prior or after the SNV/indel is in general not available. Therefore, for application of our approach, all scenarios have to be considered and four possible values for $CCF_{SNV/indel}$ have to be calculated. Scenarios may be excluded if the basic assumption $w+x+y+z=1.0$ is violated.

RESULTS

Eight patients with myeloid neoplasia were comprehensively analyzed using karyotyping, fluorescence *in situ* hybridization (FISH), and a custom NGS panel. In addition, array-CGH analysis was performed whenever possible. All patients showed clonal evolution with many genomic variants, either at the initial time point and/or during the course of the disease.

Identification of Linear Clonal Evolution in Patients Analyzed at One Time Point

In order to show that clonal evolution can be reconstructed and visualized even when only one time point is available, we examined three patients (patient 1, 2 and 3) with the aforementioned bioinformatic approach. The total number of aberrations in a patient ranged from 3 (patient 3) to 11 (patient 1). Due to this approach it was possible to determine clones by clustering mutations. These clones were sorted according to CCF, in order of their appearance. Through reconstruction and visualization of clonal evolution it was possible to identify the initial clone (stemline) as well as the passenger aberrations for all patients. For example, patient 1 (AML) showed a deletion in 5q (del(5)(q14q33)) in the initial clone, patient 2 (MDS) showed a trisomy 8 as well as a pathogenic variant in *DNMT3A* in the initial clone, and patient 3 (therapy related myeloid neoplasm) showed a translocation t(12;22) in combination with a pathogenic variant in *NF1* in the initial clone (see **Figure 2** and **Table 1**/**S1**). The total number of clones in a patient ranged from 2 (patient 3) to 7 (patient 1). In summary, our approach was successful in identifying linear clonal evolution in all patients.

Identification of Different Clonal Evolution Models (Linear, Branching or Neutral) in Patients Analyzed at More Than One Time Point

Next, we analyzed five patients where data was available for more than one time point (2 to 5 time points of analysis) to show that the bioinformatic approach can reconstruct different models of clonal evolution. At each time point, all patients were analyzed by a combination of methods, including karyotyping, FISH and a custom NGS-Panel. The estimated therapy effect was included in the reconstruction of clonal evolution. The approach reconstructed a distinct model of clonal evolution for patients 5, 7 and 8 (see **Figures 2C–E**). Patient 5 was analyzed at two time points – initial and progression. The bioinfomatic approach reconstructed a linear clonal evolution pattern. The initial clone showed a translocation t(4;14;11) and a structural aberration in chromosome arm 10p (add (10)(p14)). At this time point, pathogenic variants in *KRAS*, *ASXL1* and *NRAS* were identified with a very low percentage of CCF (close to 1%). However, the pathogenic variants were only identified by deep sequencing. At the second time point (progression), additional chromosome aberrations (+der(4)(4;14;11), +8, +9, +19 and +21) were identified and this subclone became the mainline (see **Figure 2D**). Patient 7 was analyzed at four time points – initial, progression, remission and relapse. Stem cell transplantation was performed between the third and fourth time point of analysis. The initial clone showed an insertion of a segment of chromosome 12 into chromosome 9 (ins(9;12)) and a trisomy 12. At time point two, the percentage of cells only harboring the stemline's mutations were greatly reduced (~5%) and a subclone developed. This clone showed a deletion in 12p (del(12)(p13)) and a trisomy 11, 19 and X. For the subclone, the approach calculated linear clonal evolution. At time point three, the subclone was no longer detectable, but a new subclone was identified at time point four. This subclone presented a derivative chromosome 7 with a translocation t(7;9) (der(7)t(7;9)(q35;q21)). At this time point, the clonal evolution pattern of patient 7 changed from linear to branched (see **Figure 2E**). Patient 8 was analyzed at five time points – initial, remission, progression, relapse and remission. Stem cell transplantation took place between the first and second time point. The initial clone showed a deletion in 9q (del(9)(q21q31)) and a subclone with a likely pathogenic variant in *WT1* as well as a pathogenic variant in *NRAS*. At this time point, a linear evolution pattern was reconstructed. At time point two, after stem cell transplantation, no variants were detected – neither by karyotyping nor deep sequencing. However, the aforementioned variants were present at time point three (progression). At this time point, a linear clonal evolution pattern was reconstructed in one subclone (light red). Furthermore, four additional subclones were observed. All subclones developed independently of the initial clone with a deletion in 9q (del(9)(q21q31)). At the fourth time point of analysis, three additional subclones were detected. The clonal evolution pattern of patient 8 changed from linear to neutral. At the last time point of analysis (remission), no variants were detected (see **Figure 2F**). In summary, a linear evolution pattern was reconstructed for patient 5, a branched evolution pattern was reconstructed for patient 7 and a neutral evolution pattern was reconstructed for patient 8.

Reconstruction of clonal evolution in patients 4 and 6 showed more than one possible clonal evolution pattern. All possible versions of reconstruction are shown in **Figure 3**. Patient 4 was analyzed at two time points (initial and progression). The initial clone showed a deletion in 5q (del(5)(q14q34)). Additionally, four subclones developed. At this time point, patient 4 showed a linear clonal evolution pattern. At time point two, four additional subclones were detected and the evolution pattern changed from a linear to a branched evolution pattern. Although the model of clonal evolution might be clear, it was not possible to reconstruct

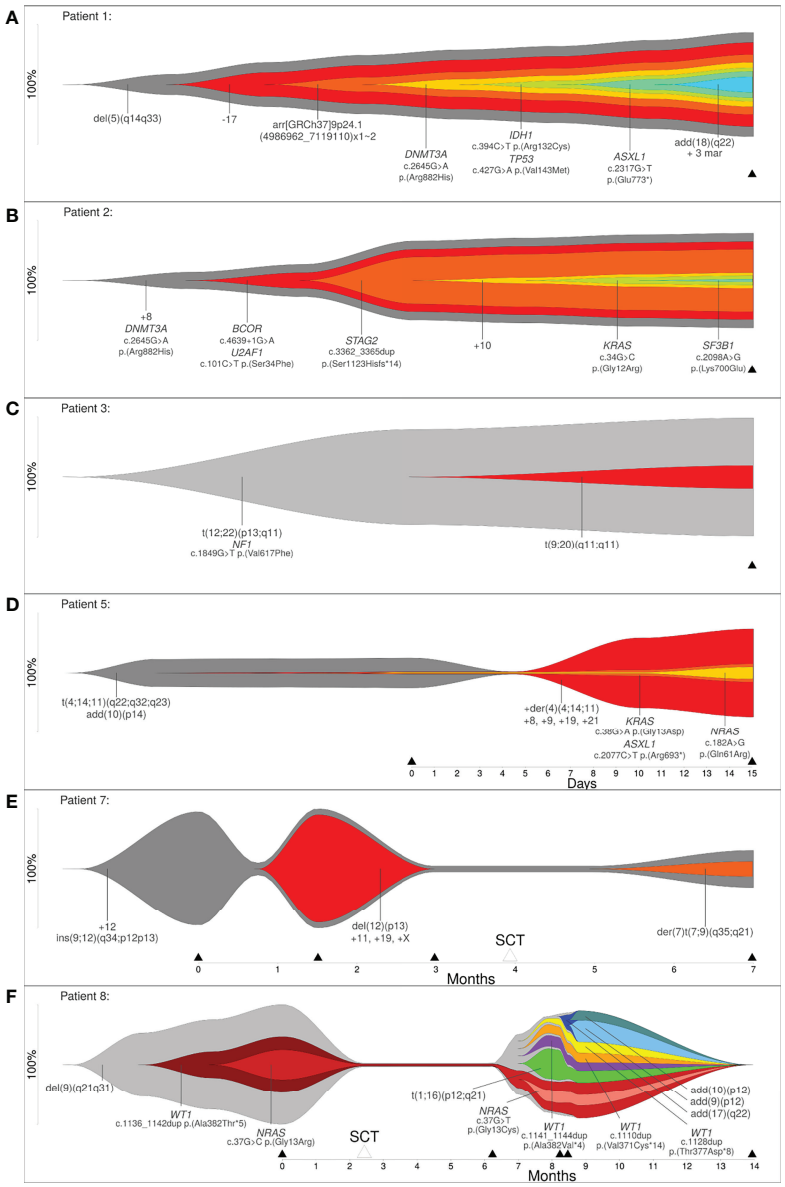


FIGURE 2 | **(A–C)** Reconstruction of clonal evolution of three patients analyzed at one time point. **(D–F)** Clonal evolution patterns of patient 5 (linear evolution), patient 7 (branched evolution) and patient 8 (neutral evolution). Y-axis represents the percentage of clone size, black triangles indicate the analyzed time point and white triangles indicate time point of stem cell transplantation.

the precise subclonal relationship. The smallest subclone with (likely) pathogenic variants in *KRAS*, *NRAS* and *PTPN11* showed a low percentage of CCF (5%). Based on the available data, the parent clone of the smallest subclone cannot be definitely determined and, therefore, patient 4 showed three possible versions of branched evolution patterns (see **Figure 3A**). Patient 6 was analyzed at three time points – initial and two progression time points. The initial clone showed a pathogenic variant in *JAK2*. At the second time point of analysis, we identified two events: (1) a likely pathogenic variant in *TP53* and (2) a derivative chromosome 17 resulting in loss of *TP53* (der (17)t(13;17)(q21;p12)) with further aberrations. Based on the available data, it was not possible to reconstruct which of the two events took place first. Furthermore, it was not possible to calculate if both events affected the same allele or different alleles. As described in Materials and Methods, four different cases have to be considered (see **Figure 1A**). Two out of four cases were in line with the basic assumption w+x+y+z=1.0. Both cases were confirmed by the observed percentages of CCF and VAFs at the time points one and two of analysis. According to the first case, the CNV event took place first, followed by the likely pathogenic variant in *TP53*. Both aberrations affected different alleles. For

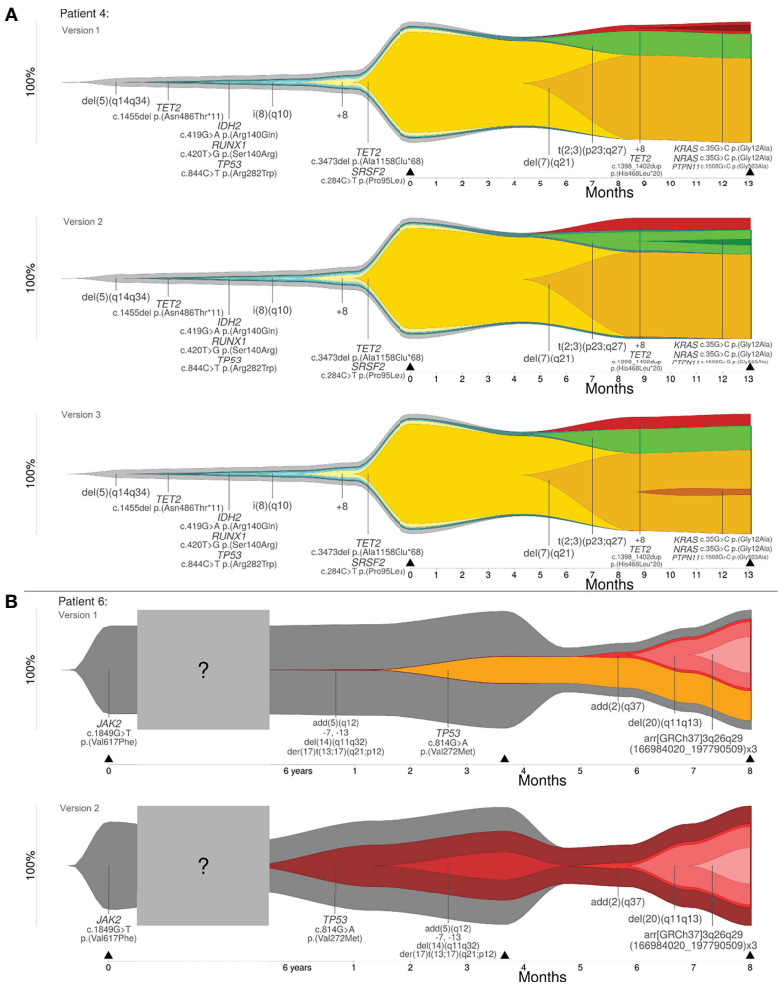


FIGURE 3 | **(A)** Clonal evolution patterns of patient 4: three versions of branched evolution. **(B)** Clonal evolution patterns of patient 6: linear and branched evolution. Y-axis represents the percentage of clone size and black triangles indicate the analyzed time point.

this case, a branched clonal evolution pattern was reconstructed (see **Figure 3B**). Mathematically, it is also possible that the point mutation in *TP53* took place first and subsequently the mutated allele was affected by the deletion. For this case, a linear clonal evolution pattern was reconstructed. However, a biallelic inactivation of *TP53* frequently occurs during disease progression of hematologic neoplasms (e.g. MDS) (31). In summary, three possible branched evolution patterns were reconstructed for patient 4, and two models of clonal evolution (linear and branching) were reconstructed for patient 6.

DISCUSSION

In this study we propose a bioinformatic approach to analyze clonal evolution and the genetic architecture comprising SNVs, indels and SVs, especially CNVs in patients with myeloid neoplasia. In order to show the possibility of reconstruction and visualization of clonal evolution at one time point of analysis, we analyzed three patients with myeloid neoplasia. Here, all patients showed a linear clonal evolution pattern. Furthermore, with this bioinformatic approach, we addressed all alterations that may play a role in the pathogenesis of the disease (driver) and alterations, which may occur during disease development (passenger). For example, patient 1 and 2 showed a *DNMT3A*^{R882} mutation at the initial time point of analysis. Pathogenic variants in *DNMT3A* occur in ~30% of AML patients and studies of clonal architecture have shown that pathogenic variants in *DNMT3A* occur as an early event in leukemogenesis. For that reason, *DNMT3A* could be a possible therapeutic target in the future (32).

In recent years, several targeted therapies have been approved for the treatment of patients with myeloid neoplasia. For example, the development of genome sequencing has identified numerous somatic alterations in patients with AML. Some of these alterations ("actionable mutations") can be targeted by specific drugs to improve the outcome of the patients (33, 34).

For example, patient 1 had an *IDH1*^{R132} mutation, which can be found in 5-10% of AML patients. Patients with an *IDH1*^{R132} mutation can be treated with ivosidenib. A recent study has shown that in these patients single-agent ivosidenib treatment leads to a complete remission (CR) or CR with incomplete haematological recovery rates of 42.4% (5). A phase trial study has analyzed a cohort of AML patients with a pathogenic variant in *IDH1*, which were treated with ivosidenib-and-azacitidine or with placebo-and-azacitidine. Patients treated with ivosidenib-and-azacitidine showed a significantly longer event-free survival and overall survival (24.0 month vs. 7.9 month) in comparison to patients treated with placebo-and-azacitidine (35). In addition, Bolton et al. have analyzed the clonal hematopoiesis in patients with therapy-related myeloid neoplasms. The study has shown that pathogenic variants show a different clonal evolution based on exposures (e.g. radiation, topoisomerase II inhibitors), for example, clones with pathogenic variants in DNA damage response genes outcompetes other clones (36). Thus, reconstruction of clonal evolution as well as identification of driver and passenger alterations allow a better understanding of the mechanisms of leukemogenesis and improvement of (future) treatment strategies (personalized therapy).

We analyzed five patients at more than one time point to show that the approach can reconstruct different models of clonal evolution (i.e. linear, branching and neutral evolution). For all of these patients the estimated therapy effect was included. The analyzed data of three out of these five patients allowed a distinct reconstruction of clonal evolution (patients 5, 7 and 8). Linear evolution is characterized by a dominant clone which overgrows the ancestral clone after acquisition of additional mutations in a stepwise manner (3, 4) - a linear mutation pattern was reconstructed for patient 5. Branching evolution is characterized by the occurrence of different subclones from on ancestral clone (3, 4) – as detected in patient 7. Neutral evolution is an extreme case of branching evolution, which is characterized by the accumulation of random mutations over time (16) – as was observed in patient 8. A possible advantage of reconstructing clonal evolution could be to show if both pathogenic variants co-occur in the same clone or in different clones. This could influence the treatment strategy. For example, patient 8 showed a pathogenic variant in *NRAS* and in *WT1*. Pathogenic variants in *NRAS* occur in ~12% of all AML cases and co-occur with pathogenic variants in epigenetic modifiers (*TET2/IDH/WT1*). The co-occurrence of pathogenic variants in *NRAS* and *WT1* showed a sensitivity to MAPK kinase inhibition, which has been shown in patient samples and mouse models providing a possible treatment strategy in the future (32). Especially patients with AML show a high number of heterogeneous pathogenic variants at diagnosis and relapse. These pathogenic variants are organized into a hierarchy of clones and they are able to adapt and evolve in response to therapeutic pressure (5). In the future, the clonal evolution pattern and genetic architecture of many patients must be analyzed in order to learn more about therapy resistance and new possibilities for therapy strategies to improve patient outcome.

For the remaining two patients (patient 4 and 6) the analyzed data showed more than one possible clonal evolution pattern. The proposed bioinformatic approach can be used to analyze the order and occurrence of genetic alterations, which could be relevant for treatment decisions and for the understanding of leukemic progression. For example, patient 6 showed a *JAK2/TP53* clone at the transformation to AML, which replaced the dominant *JAK2* clone in the myeloproliferative neoplasia (initial diagnosis). This observation is in concordance with a recent publication, where a significant alteration in clonal architecture or a "clonal sweep" with emergence of new dominant clone(s) was observed with single cell sequencing (17). These observations can be used for therapy stratification during the disease course.

Altogether, we present a generalizable approach that enables the combination of a wide spectrum of methods and to analyze retrospective samples to visualize leukemic progression and genetic architecture. Recent publications have reconstructed clonal evolution with single cell sequencing where the available data has allowed definitive identification of clonal architecture (17, 19, 37). Single cell sequencing has many advantages: exact distinction of different clones, measurement of accurate clonal complexity or resolution of mutational order (17, 38). But there are also many disadvantages, including limited depth of sequencing, no possibility of analyzing retrospective samples (no suitable material), or missing information of chromosomal aberrations (important for treatment stratification). Even though the used methods in this project do not offer the possibility to analyze clonal evolution at a single cell level, the inclusion of cytogenetic and molecular genetic alterations allow the identification of driver and passenger alterations, which is currently impossible using only single cell sequencing. For future projects, a combination of methods (bulk and single cell), including karyotyping and single cell sequencing, could be helpful for the exact reconstruction of clonal evolution.

In conclusion, this bioinformatic approach offers the possibility of analyzing clonal evolution and the order and occurrence of many cytogenetic and molecular genetic alterations (genetic architecture) at one or more time points of analysis. Different models of clonal evolution (i.e. linear, branching and neutral) can be reconstructed with this approach. As the approach describes integration of data for reconstruction of clonal evolution in general, its application is not limited to myeloid neoplasms. Instead, it may be applied on all kinds of non-solid tumors showing clonal evolution. The visualization of the results in fishplots contributes to a better understanding of genetic architecture and leukemic progression. This approach helps to identify possible targets for the disease (personalized therapy) and can be used to identify markers in order to assess minimal residual disease.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: PRJNA847150 link to https://www.ncbi.nlm.nih.gov/sra/PRJNA847150.

ETHICS STATEMENT

The study was conducted according to the guidelines of the Declaration of Helsinki. The ethical review boards of Hannover Medical School approved this study (Nr. 8657_BO_K_2019) and all patients gave their written consent. Informed consent was obtained from all subjects involved in the study. Reported results can be found in **Supplementary Materials**.

AUTHOR CONTRIBUTIONS

The study was conceived by SS, YB, and GG. Data and material of the study was generated by YB, GG, SS, DS, FT, MH, AC, DD, FL, JV, MD, and BS. Data collection was performed by SS, YB, GG. Interpretation and analysis were conducted by YB, GG, SS, MD, and BS. SS and YB wrote the manuscript. The paper was edited by all authors. All authors contributed to the article and approved the submitted version.

FUNDING

GG and BS have been supported by the BMBF MyPred (01GM1911A).

ACKNOWLEDGMENTS

We would like to thank Dr. Marius Wöste for his input on analyzing and interpreting array-CGH data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.888114/full#supplementary-material

REFERENCES

1. Doulatov S, Papapetrou EP. Studying Clonal Evolution of Myeloid Malignancies Using Induced Pluripotent Stem Cells. *Curr Opin Hematol* (2021) 28(1):50–6. doi: 10.1097/MOH.0000000000000620
2. E Swerdlow SH Campo, NL Harris, ES Jaffe, SA Pileri, H Stein and J Thiele eds. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Revised 4th edition*. Lyon, France: International Agency for Research on Cancer (2017).
3. da Silva-Coelho P, Kroeze LI, Yoshida K, Koorenhof-Scheele TN, Knops R, van de Locht LT, et al. Clonal Evolution in Myelodysplastic Syndromes. *Nat Commun* (2017) 8:15099. doi: 10.1038/ncomms15099
4. Vosberg S, Greif PA. Clonal Evolution of Acute Myeloid Leukemia From Diagnosis to Relapse. *Genes Chromosomes Canc* (2019) 58(12):839–49. doi: 10.1002/gcc.22806
5. Dohner H, Wei AH, Lowenberg B. Towards Precision Medicine for AML. *Nat Rev Clin Oncol* (2021) 18(9):577–90. doi: 10.1038/s41571-021-00509-w
6. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* (2016) 374(23):2209–21. doi: 10.1056/NEJMoa1516192
7. Metzeler KH, Herold T, Rothenberg-Thurley M, Amler S, Sauerland MC, Gorlich D, et al. Spectrum and Prognostic Relevance of Driver Gene Mutations in Acute Myeloid Leukemia. *Blood* (2016) 128(5):686–98. doi: 10.1182/blood-2016-01-693879
8. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of Genetic Lesions in 944 Patients With Myelodysplastic Syndromes. *Leukemia* (2014) 28(2):241–7. doi: 10.1038/leu.2013.336
9. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and Biological Implications of Driver Mutations in Myelodysplastic Syndromes. *Blood* (2013) 122(22):3616–27; quiz 99. doi: 10.1182/blood-2013-08-518886
10. Palomo L, Meggendorfer M, Hutter S, Twardziok S, Adema V, Fuhrmann I, et al. Molecular Landscape and Clonal Architecture of Adult Myelodysplastic/Myeloproliferative Neoplasms. *Blood* (2020) 136(16):1851–62. doi: 10.1182/blood.2019004229
11. Makishima H, Yoshizato T, Yoshida K, Sekeres MA, Radivoyevitch T, Suzuki H, et al. Dynamics of Clonal Evolution in Myelodysplastic Syndromes. *Nat Genet* (2017) 49(2):204–12. doi: 10.1038/ng.3742
12. Onecha E, Rapado I, Luz Morales M, Carreno-Tarragona G, Martinez-Sanchez P, Gutierrez X, et al. Monitoring of Clonal Evolution of Acute Myeloid Leukemia Identifies the Leukemia Subtype, Clinical Outcome and Potential New Drug Targets for Post-Remission Strategies or Relapse. *Haematologica* (2021) 106(9):2325–33. doi: 10.3324/haematol.2020.254623
13. Menssen AJ, Walter MJ. Genetics of Progression From MDS to Secondary Leukemia. *Blood* (2020) 136(1):50–60. doi: 10.1182/blood.2019000942
14. Morita K, Wang F, Jahn K, Hu T, Tanaka T, Sasaki Y, et al. Clonal Evolution of Acute Myeloid Leukemia Revealed by High-Throughput Single-Cell Genomics. *Nat Commun* (2020) 11(1):5327. doi: 10.1038/s41467-020-19119-8
15. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* (2017) 168(4):613–28. doi: 10.1016/j.cell.2017.01.018
16. Davis A, Gao R, Navin N. Tumor Evolution: Linear, Branching, Neutral or Punctuated? *Biochim Biophys Acta Rev Cancer* (2017) 1867(2):151–61. doi: 10.1016/j.bbcan.2017.01.003
17. Miles LA, Bowman RL, Merlinsky TR, Csete IS, Ooi AT, Durruthy-Durruthy R, et al. Single-Cell Mutation Analysis of Clonal Evolution in Myeloid Malignancies. *Nature* (2020) 587(7834):477–82. doi: 10.1038/s41586-020-2864-x
18. Weber LL, El-Kebir M. Distinguishing Linear and Branched Evolution Given Single-Cell DNA Sequencing Data of Tumors. *Algorith Mol Biol* (2021) 16 (1):14. doi: 10.1186/s13015-021-00194-5
19. Potter N, Miraki-Moud F, Ermini L, Titley I, Vijayaraghavan G, Papaemmanuil E, et al. Single Cell Analysis of Clonal Architecture in Acute Myeloid Leukaemia. *Leukemia* (2019) 33(5):1113–23. doi: 10.1038/s41375-018-0319-2
20. Akkari YMN, Baughn LB, Dubuc AM, Smith AC, Mallo M, Dal Cin P, et al. Guiding the Global Evolution of Cytogenetic Testing for Hematologic Malignancies. *Blood* (2022) 139(15):2273–84. doi: 10.1182/blood.2021014309
21. Schlegelberger B, Metzke S, Harder S, Zühlke-Jenisch R, Zhang Y, Siebert R. Classical and Molecular Cytogenetics of Tumor Cells. Berlin, Heidelberg: Springer. In: Springer Lab Manual (1999). p. 151–85.
22. Schlegelberger B, Zwingers T, Harder L, Nowotny H, Siebert R, Vesely M, et al. Clinicopathogenetic Significance of Chromosomal Abnormalities in Patients With Blastic Peripheral B-Cell Lymphoma. Kiel-Wien-Lymphoma Study Group. *Blood* (1999) 94(9):3114–20. doi: 10.1182/blood.V94.9.3114
23. McGowan-Jordan J, Hastings R, Moore S. *An International System for Human Cytogenomic Nomenclature*. Basel: Karger. J McGowan-Jordan, R Hastings and S Moore, editors (2020).
24. Li H. *Aligning Sequence Reads, Clone Sequences and Assembly Contigs With BWA-MEM*. Karger. Basel (2013).
25. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration. *Brief Bioinform* (2013) 14(2):178–92. doi: 10.1093/bib/bbs017
26. Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A, Mesirov JP. Variant Review With the Integrative Genomics Viewer. *Cancer Res* (2017) 77(21):e31–e4. doi: 10.1158/0008-5472.CAN-17-0337

27. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* (2015) 17(5):405–24. doi: 10.1038/gim.2015.30
28. Sandmann S, Karimi M, de Graaf AO, Rohde C, Gollner S, Varghese J, et al. Appreci8: A Pipeline for Precise Variant Calling Integrating 8 Tools. *Bioinformatics* (2018) 34(24):4205–12. doi: 10.1093/bioinformatics/bty518
29. Reutter K, Sandmann S, Rohde J, Muller S, Woste M, Khanam T, et al. Reconstructing Clonal Evolution in Relapsed and non-Relapsed Burkitt Lymphoma. *Leukemia* (2021) 35(2):639–43. doi: 10.1038/s41375-020-0862-5
30. Miller CA, McMichael J, Dang HX, Maher CA, Ding L, Ley TJ, et al. Visualizing Tumor Evolution With the Fishplot Package for R. *BMC Genomics* (2016) 17(1):880. doi: 10.1186/s12864-016-3195-z
31. Bernard E, Nannya Y, Hasserjian RP, Devlin SM, Tuechler H, Medina-Martinez JS, et al. Implications of TP53 Allelic State for Genome Stability, Clinical Presentation and Outcomes in Myelodysplastic Syndromes. *Nat Med* (2020) 26(10):1549–56. doi: 10.1038/s41591-020-1008-z
32. Kishtagari A, Levine RL, Viny AD. Driver Mutations in Acute Myeloid Leukemia. *Curr Opin Hematol* (2020) 27(2):49–57. doi: 10.1097/MOH.0000000000000567
33. Collignon A, Hospital MA, Montersino C, Courtier F, Charbonnier A, Saillard C, et al. A Chemogenomic Approach to Identify Personalized Therapy for Patients With Relapse or Refractory Acute Myeloid Leukemia: Results of a Prospective Feasibility Study. *Blood Cancer J* (2020) 10(6):64. doi: 10.1038/s41408-020-0330-5
34. Carr TH, McEwen R, Dougherty B, Johnson JH, Dry JR, Lai Z, et al. Defining Actionable Mutations for Oncology Therapeutic Development. *Nat Rev Canc* (2016) 16(5):319–29. doi: 10.1038/nrc.2016.35
35. Montesinos P, Recher C, Vives S, Zarzycka E, Wang J, Bertani G, et al. Ivosidenib and Azacitidine in IDH1-Mutated Acute Myeloid Leukemia. *N Engl J Med* (2022) 386(16):1519–31. doi: 10.1056/NEJMoa2117344
36. Bolton KL, Ptashkin RN, Gao T, Braunstein L, Devlin SM, Kelly D, et al. Cancer Therapy Shapes the Fitness Landscape of Clonal Hematopoiesis. *Nat Genet* (2020) 52(11):1219–26. doi: 10.1038/s41588-020-00710-0
37. Pellegrino M, Sciambi A, Treusch S, Durruthy-Durruthy R, Gokhale K, Jacob J, et al. High-Throughput Single-Cell DNA Sequencing of Acute Myeloid Leukemia Tumors With Droplet Microfluidics. *Genome Res* (2018) 28 (9):1345–52. doi: 10.1101/gr.232272.117
38. Brierley CK, Mead AJ. Single-Cell Sequencing in Hematology. *Curr Opin Oncol* (2020) 32(2):139–45. doi: 10.1097/CCO.0000000000000613

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

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

frontiers | Frontiers in Oncology

BRIEF RESEARCH REPORT
published: 19 July 2022
doi: 10.3389/fonc.2022.919278

The Role of Clonal Evolution on Progression, Blood Parameters, and Response to Therapy in Multiple Myeloma

Sarah Sandmann[1†], Katharina Karsch[2†], Peter Bartel[2], Rita Exeler[3], Tobias J. Brix[1], Elias K. Mai[4], Julian Varghese[1], Georg Lenz[2] and Cyrus Khandanpour[2,5*]*

[1] Institute of Medical Informatics, University of Münster, Münster, Germany, [2] Department of Medicine A, Hematology, Oncology and Pneumology, University Hospital Münster, Münster, Germany, [3] Institute of Human Genetics, University Hospital Münster, Münster, Germany, [4] Department of Internal Medicine V, Heidelberg University Hospital, Heidelberg, Germany, [5] University Medical Center Schleswig-Holstein Campus Lübeck, University of Lübeck, Lübeck, Germany

OPEN ACCESS

Edited by:
Claudio Cerchione, Scientific Institute of Romagna for the Study and Treatment of Tumors (IRCCS), Italy

Reviewed by:
Syed Jafar Mehdi, University of Arkansas for Medical Sciences, United States
Anju Kumari, National Cancer Institute (NIH), United States

***Correspondence:**
Sarah Sandmann
sarah.sandmann@uni-muenster.de
Cyrus Khandanpour
cyrus.khandanpour@uni-luebeck.de

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Hematologic Malignancies, a section of the journal Frontiers in Oncology

Received: 13 April 2022
Accepted: 22 June 2022
Published: 19 July 2022

Citation:
Sandmann S, Karsch K, Bartel P, Exeler R, Brix TJ, Mai EK, Varghese J, Lenz G and Khandanpour C (2022) The Role of Clonal Evolution on Progression, Blood Parameters, and Response to Therapy in Multiple Myeloma. Front. Oncol. 12:919278. doi: 10.3389/fonc.2022.919278

Introduction: A variety of biomarkers are considered for diagnosis (e.g., β2-microgobulin, albumin, or LDH) and prognosis [e.g., cytogenetic aberrations detected by fluorescence *in situ* hybridization (FISH)] of multiple myeloma (MM). More recently, clonal evolution has been established as key. Little is known on the clinical implications of clonal evolution.

Methods: We performed in-depth analyses of 25 patients with newly diagnosed MM with respect to detailed clinical information analyzing blood samples collected at several time points during follow-up (median follow-up: 3.26 years since first diagnosis). We split our cohort into two subgroups: with and without new FISH clones developing in the course of disease.

Results: Each subgroup showed a characteristic chromosomal profile. Forty-three percent of patients had evidence of appearing new clones. The patients with new clones showed an increased number of translocations affecting chromosomes 14 (78% vs. 33%; p = 0.0805) and 11, and alterations in chromosome 4 (amplifications and translocations). New clones, on the contrary, were characterized by alterations affecting chromosome 17. Subsequent to the development of the new clone, 6 out of 9 patients experienced disease progression compared to 3 out of 12 for patients without new clones. Duration of the therapy applied for the longest time was significantly shorter within the group of patients developing new clones (median: 273 vs. 406.5 days; p = 0.0465).

Discussion: We demonstrated that the development of new clones, carrying large-scale alterations, was associated with inferior disease course and shorter response to therapy, possibly affecting progression-free survival and overall survival as well. Further studies evaluating larger cohorts are necessary for the validation of our results.

Keywords: multiple myeloma, clonal evolution, survival, prognosis, chromosomal alteration

1 INTRODUCTION

Multiple myeloma (MM) is a hematologic malignancy characterized by monoclonal expansion of plasma cells in the bone marrow (1). Clinical presentation is very heterogeneous, ranging from a rather indolent course to clinically aggressive plasma cell leukemia.

The diagnostic criteria of MM have expanded in recent years and now include biomarkers for disease malignancy such as bone marrow plasma cell count, free light chain ratio, and number of focal lesions on MRI, in addition to the well-established CRAB criteria (2).

The prognostic significance of cytogenetic aberrations is well described; t(4;14), t(14;16), and del(17p) are included in Revised International Staging System (R-ISS) risk stratification (3, 4). A change in the mutational profile or clonal composition between two or more time points is referred to as clonal evolution (5, 6). Models describing clonal evolution can be categorized as linear, branched, punctuated, or neutral (7).

The molecular basics of clonal evolution in MM have been studied and reviewed in depth (5, 8–10). It has been observed that the complexity of the MM tumor genome increased over time (11) and that cytogenetic heterogeneity is of prognostic significance in newly diagnosed MM patients treated with bortezomib (12). However, more information on clinical implications of clonal evolution is needed.

We hypothesize that patients with MM, characterized by new clones emerging after first diagnosis, show poor prognosis compared to patients without new clones. We define a new clone as a gain of aberrations and, thus, increased heterogeneity of a tumor. The gain can affect both healthy cells, leading to branched clonal evolution with independent clones, or clones already present at first diagnosis, leading to a derivative of these clones. To explore our hypothesis, we analyzed a set of 25 patients with MM, split into two subgroups: with or without new clones. For these patients, we correlated clinical data on progression, chromosomal profiles, blood parameters, and therapies, based on blood samples collected at several time points during follow-up.

2 METHODS

2.1 Study Population

A cohort of 25 patients with MM and comparable therapies applied, treated at the University Hospital Münster, was analyzed. All data were collected and analyzed in accordance with relevant ethical guidelines and principles of the Declaration of Helsinki. The ethical review boards of the University of Münster and the Ärztekammer Westfalen-Lippe approved this study (2018-452-f-S). All patients gave their written informed consent. Detailed information on every patient is provided in **Supplementary Table 1**; a summary of the study population's main characteristics is provided in **Supplementary Table 2**.

Data were collected between December 2014 and March 2021. All patients were monitored starting with first diagnosis of MM (exceptions: for UPN09, information on laboratory parameters and therapies applied was available at diagnosis of a plasmocytoma, 2 years prior to diagnosis of MM; for UPN16, information on laboratory parameters was available 4 months prior to first diagnosis of MM). Median time of follow-up after first diagnosis of MM was 3.62 years (IQR = 2.01–5.00 years). Forty percent of patients were female. At first diagnosis, patients were in median 58 years old (IQR = 53–64 years). Nine out of 25 patients received autologous stem cell transplantation (SCT). Five overall survival (OS) events occurred during follow-up (see *Section 2.4 Statistical Methods* for the definition of OS). Since this was the primary endpoint of the study, data on cytogenetic aberrations were available for all patients at several time points. Analyses on blood samples of the patients were performed between 2014 and 2021 at the Institute of Human Genetics Münster. A median of 3 samples per patient was analyzed during follow-up (range: 1–11 analyses during follow-up). Additionally, information on therapy and laboratory parameters (κ and λ light chains measured in serum), lactate dehydrogenase (LDH), and gradient of monoclonal protein (M-gradient) was available for every patient at several time points during follow-up. A visual summary of available data is provided in **Supplementary Figure 1**, detailed information on laboratory parameters is provided in **Supplementary Figures 2–4**.

Based on clonal evolution, we split up our cohort into two subgroups: For subgroup 1, no new clones were detected in the course of disease (patients UPN05 to 16) based on cytogenetic findings using the fluorescence *in situ* hybridization (FISH) approach. All clones were already present at first analysis with FISH. For subgroup 2, at least one completely new clone was developed in the course of the disease, which was not detected at any previous time point (patients UPN17 to 25); i.e., a gain of aberrations occurred. Four patients (UPN01 to 04) were excluded from subgroup analysis, because they had no follow-up FISH samples available.

2.2 Fluorescence *In Situ* Hybridization Analysis

FISH was performed as described previously (13, 14). Briefly, CD138-purified plasma cells were analyzed. The following commercial available probes from Cytocell (Cytocell, Cambridge, UK), MetaSystems (MetaSystems Probe GmbH, Altlußheim, Germany), CytoTest (CytoTest, Rockville, USA), and Abbott Molecular (Abbott, Green Oaks, USA) were used: Cytocell CKS1B/CDKN2C (P18) Amplification/Deletion Probe (1p32.3, 1q21); Cytocell D13S319 Plus Deletion Probe (13q14.2, 13qter); MetaSystems XL TP53/NF1 Deletion Probe (17p13.1, 17q11.2); MetaSystems XL Iso(17q) Deletion Probe (17p13, 17q22); MetaSystems XL MYC BA Break Apart Probe (8q24); MetaSystems XL E2A Break Apart Probe (19q13); CytoTest NSD1/TERT FISH Probe Kit (5q35, 5p15); Abbott Molecular LSI ATM/CEP11 11 FISH Probe (11cen, 11q22); Cytocell probes for centromeric regions (D3Z1, D7Z1, D9Z3, D15Z4); Cytocell IGH/FGFR3 Plus Translocation, Dual Fusion (4p16.3, 14q32.33); Cytocell IGH/MAF Translocation, Dual Fusion (14q32, 16q23); Cytocell IGH/MAFB Translocation, Dual Fusion (14q32, 20q12); Cytocell IGH/MYEOV Translocation, Dual Fusion (11q13, 14q32); and Cytocell IGH Plus Breakapart

Probe (14q32.33). At least 100 interphase nuclei were analyzed for each probe (exception: UPN15 at time point 1: 81 nuclei, time point 2: 56 nuclei). Cutoff levels for all probes were ~5%. Read out was done by human genetics experts in Münster.

2.3 Clonal Evolution

For every patient, data on present and absent chromosomal aberrations were available for 1 to 11 time points. As alterations were determined using FISH, the number of cells affected by alteration divided by the total number of evaluated cells corresponds to the cancer cell fraction (CCF). Clonal evolution was manually reconstructed based on CCF (15). To reconstruct clonal evolution also in the presence of a few time points, we applied an approach estimating the clonal development between measured time points (16). Detailed information on detected alterations, CCFs, and assigned clusters is available in **Supplementary Table 3**.

2.4 Statistical Methods

Plots visualizing clonal evolution were generated using R 4.1.2 (17) and R package "fishplot" (18). Complex plots were developed, combining fishplots with diagrams visualizing therapy information and laboratory parameters over time. In case clonal evolution could not be reconstructed uniquely based on available data, all possible versions were reported. Plots for every patient are available as **Supplementary Figures 5–29**.

Analysis of progression-free survival (PFS) and OS was performed using R 4.1.2 (17). For PFS, we analyzed the time between first diagnosis and first progression. Additionally, we considered the time between last FISH (subgroup 1)/FISH detecting the new clone (subgroup 2) and next progression. For OS, we analyzed the time between first diagnosis and death of a patient. Kaplan–Meier curves were calculated using R 4.1.2 (17) and R packages "survival" (19) and "survminer" (20).

3 RESULTS

In this study, we analyzed clonal evolution in patients with MM, aiming at identifying differences with respect to disease progression, distinct chromosomal profiles, laboratory parameters, and response to therapy.

3.1 Disease Progression

Figure 1A provides an overview of follow-up, time points of performed FISH, and progression data.

For 8 out of 12 patients (67%) in subgroup 1 (without new clone), disease progression could be observed. Similarly, 6 out of 9 patients (67%) in subgroup 2 (with new clone) experienced progression in the course of follow-up. Investigating whether disease progression takes place earlier in subgroup 2, we analyzed PFS (time between first diagnosis and first progression). However, no significant results could be obtained (**Figure 1B**; $p = \mathbf{0.6}$).

Two out of 12 patients (17%) in subgroup 1 died within our observation period. In contrast, 3 out of 9 patients (33%) deceased in subgroup 2. Analysis of OS (time between first diagnosis and death of a patient) did not indicate significant differences between both subgroups (**Figure 1C**; $p = \mathbf{0.51}$).

Additionally, we consider PFS in relation to analysis of chromosomal alterations. The time span between last FISH (subgroup 1) or detection of the new clone (subgroup 2), and the next subsequent progression (**Figure 1D**) is evaluated. Four out of 12 patients in subgroup 1 (patients UPN05, 06, 10, and 12) and 3 out of 9 in subgroup 2 (patients UPN18, 20, and 22) never experienced any progression within time of follow-up. Additionally, 5 out of the remaining 8 patients in subgroup 1 did not experience any further disease progression subsequent to last FISH. Analysis of PFS did not reveal any significant differences between both subgroups ($p = \mathbf{0.45}$). Subsequent to the development of the new clone, 6 out of 9 patients experienced disease progression compared to 3 out of 12 for patients without new clones.

3.2 Chromosomal Profiles

Patients with MM can be characterized by specific chromosomal profiles: Translocations involving chromosome 14 [especially t (4;14)], gains and amplifications affecting chromosomal region 1q21, as well as deletion of 17p are indicators for adverse prognosis (3, 4, 9, 10, 12).

Figure 2A visualizes the chromosomal profiles of all patients considered in this study for the whole time of follow-up (for detailed information on cytogenetic aberrations detected by FISH, see **Supplementary Table 3**; for detailed information on clonal evolution of each patient, see **Supplementary Figures 5–29**). As expected, a majority of patients featured translocations affecting chromosome 14 (12 out of 25; 48%), amplifications affecting chromosome 1 (14 out of 25; 56%), and/or deletions affecting chromosome 17 (9 out of 25; 36%). Additionally, 14 out of 25 patients (56%) showed deletions in chromosome 13. For 10 out of 25 patients (40%), alterations affecting chromosome 11 (translocations and amplifications) were detected.

Comparing subgroups 1 and 2, considerable differences could be observed. While only 33% of all patients in subgroup 1 (4 out of 12) were characterized by translocation of chromosome 14, it was 78% in subgroup 2 (7 out of 9; $p = 0.0805$). With respect to all alterations affecting chromosome 14, a relation of 67% vs. 100% could be observed (8 out of 12 vs. 9 out of 9; $p = 0.1038$). Further major differences affected translocation of chromosome 11 (17% vs. 44%; 2 out of 12 vs. 4 out of 9) and changes in chromosome 4 (17% vs. 44%; 2 out of 12 vs. 4 out of 9).

Within subgroup 2, further differences could be observed comparing alterations detected by the first FISH and alterations characterizing the newly developed clones. **Figure 2B** visualizes all cytogenetic aberrations characterizing the new clones. In **Figure 2C**, a summary on chromosome level is provided.

It appeared striking that only 1 out of 9 patients in subgroup 2 had an initial alteration in chromosome 17. The ratio increased to 5 out of 9 when considering newly developed clones. Similarly, 4 out of 9 patients acquired alterations in chromosome 1. By contrast, only 1 out of 9 patients (UPN21) acquired an additional alteration in chromosome 14 in the course of clonal evolution. However, this patient already had a rearrangement in

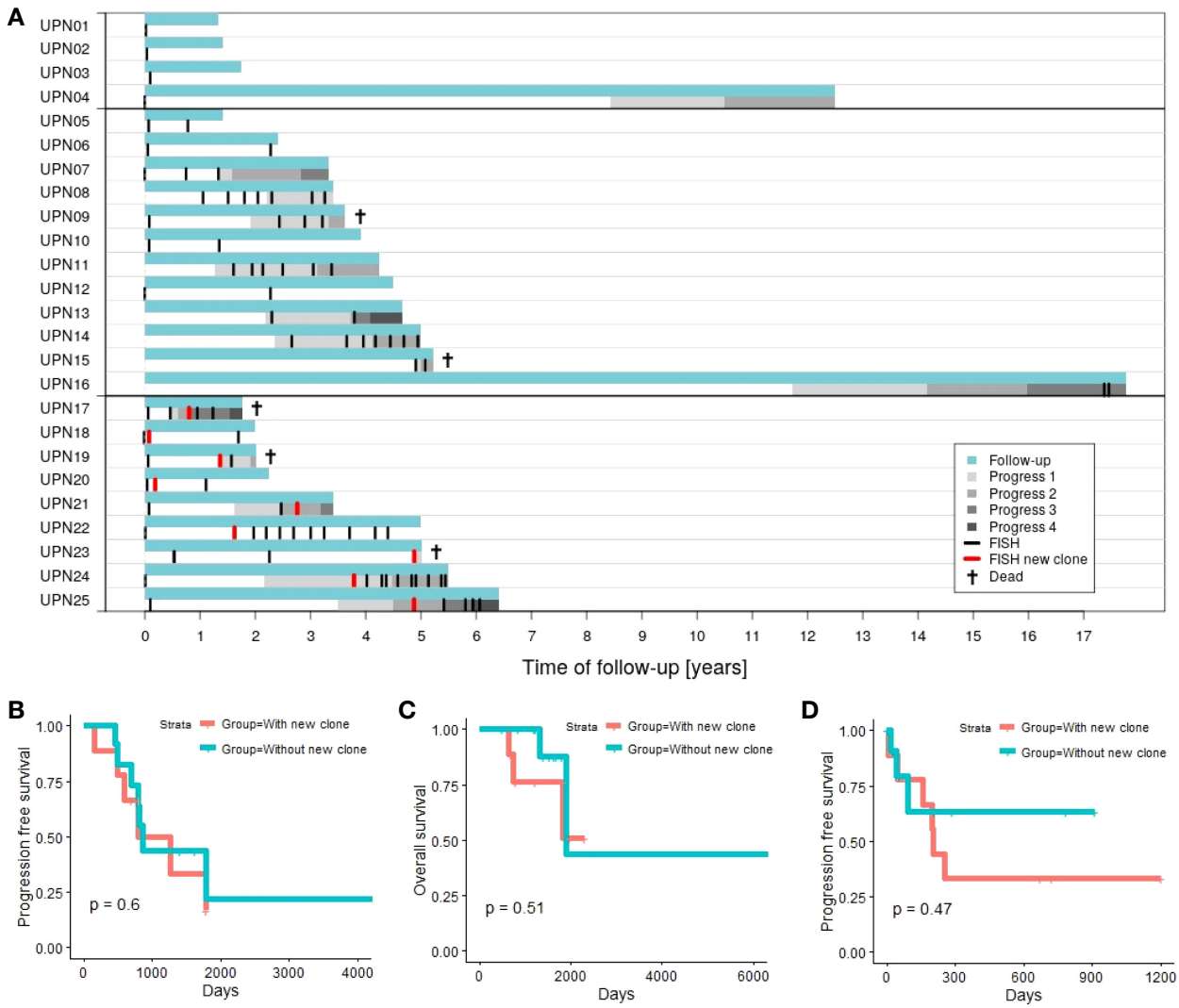


FIGURE 1 | **(A)** Time of follow-up for patients with 1 time point of aberration analysis (patients UPN01 to 04), patients with >1 time point of aberration analysis and no new clone emerging in the course of disease (patients UPN05 to 16; subgroup 1), and patients with >1 time point of aberration analysis and a new clone emerging in the course of disease (patients UPN17 to 25; subgroup 2). **(B)** Progression-free survival (first progression after first diagnosis) comparing subgroup 1 vs. subgroup 2. **(C)** Overall survival comparing subgroup 1 vs. subgroup 2. **(D)** Progression-free survival in relation to chromosomal alterations comparing subgroup 1 (next progression subsequent to last FISH) vs. subgroup 2 (next progression subsequent to detection of the new clone).

chromosome 14 at first diagnosis. Furthermore, no new alterations are acquired in chromosomes 4 (alterations detected at first diagnosis in 44%; 4 out of 9), 3, 5, 8, and 18 (11% each; 1 out of 9).

3.3 Laboratory Parameters

For all patients, information on the presence of κ and λ light chains in serum, increased LDH activity, and gradient of monoclonal protein (M-gradient) was available, measured at several time points in the course of disease. **Figure 3** visualizes the development of laboratory parameters in the two subgroups. Color indicates the time point at which measurements were conducted. For patients without new clones, we compared values measured prior to and after last FISH. For patients with new clones, we compared values measured prior to and after the development of the new clone.

For κ and λ light chains, no correlation between the values measured and the time point at which they were measured could be observed. This observation is true for both subgroups (for details, see **Supplementary Figure 2**).

LDH serum levels measured in subgroup 1 were heterogeneously distributed. The highest as well as the lowest values were scattered over the time of follow-up, independent of therapy. In contrast, correlation with respect to time point could be observed for subgroup 2. In 8 out of 9 cases, the highest values of LDH activity were measured subsequent to the development of the new clones. However, detailed evaluation of the results revealed that values did not show a continuous increase. Instead,

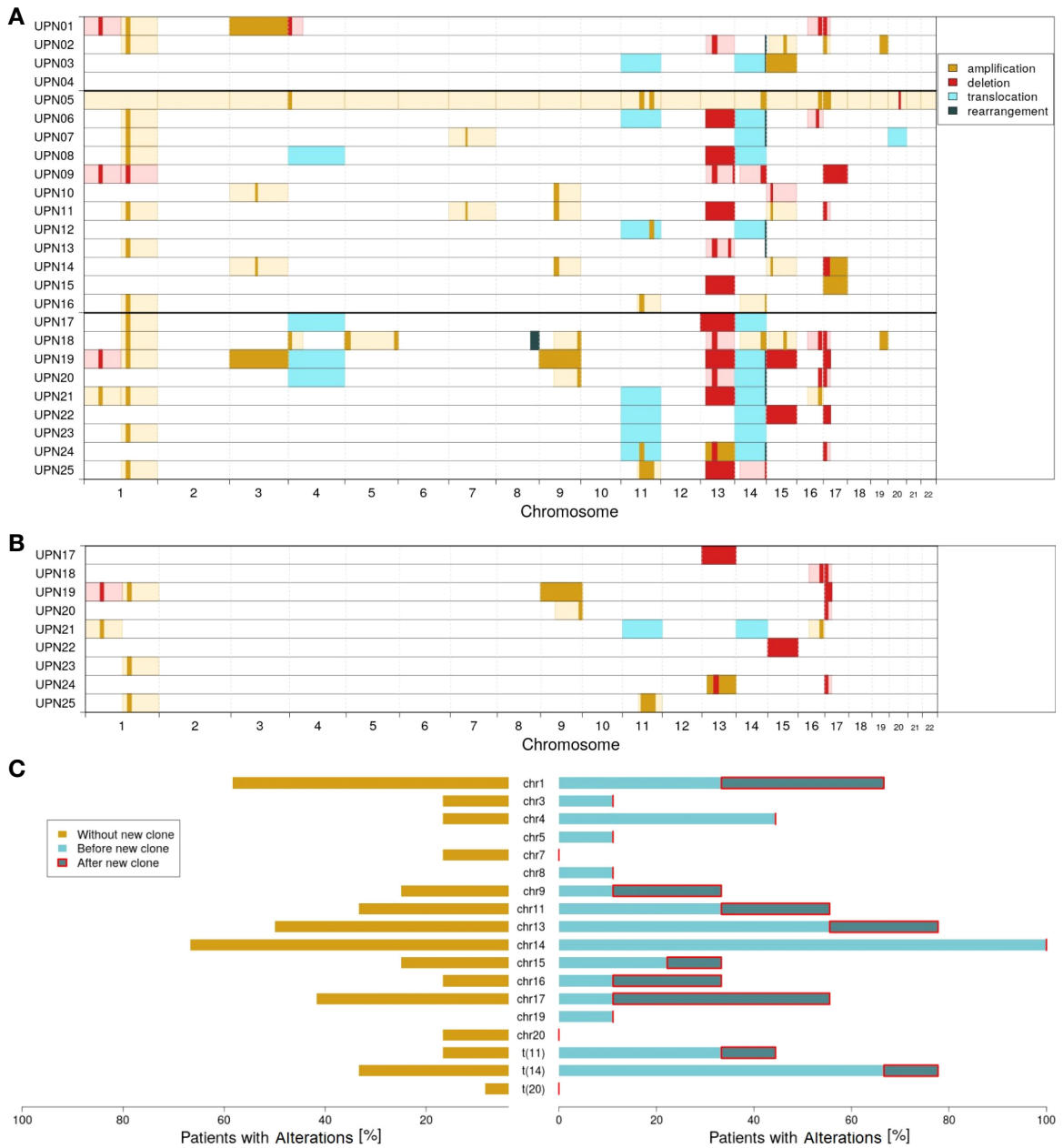


FIGURE 2 | Molecular characterization of the cohort; amplifications (dark yellow), deletions (red), translocations (light blue), and rearrangements (dark green) according to FISH probes. Light colors (yellow and red) indicate likely interpretation of the observed FISH results. **(A)** Overview of all alterations detected. **(B)** Alterations characterizing the new clones observed in patients UPN17 to 25. **(C)** Proportion of patients with altered chromosomes: without the new clone (dark yellow), with the new clone before (light blue), and after (dark blue) detection of the new clone.

peaks were usually observed within 1 year after development of the new clone. Subsequently, LDH activity decreased again until the end of follow-up (**Supplementary Figure 3**). Significant correlation to therapies applied could not be observed.

For M-gradient, no certain pattern was associated prior to the development of myeloma or upon the development of a new clone (for details, see **Supplementary Figure 4**). Patient UPN09 (subgroup 1) is an exception from this observation. A significant increase in M-gradient can be observed ($p_{\text{UPN09}} = 0.0065$). The patient died within less than 4 years after the first diagnosis.

3.4 Response to Therapy

For treatment of MM, a variety of therapies are available. In the course of disease progression, time to therapy failure is known to decrease (21).

FIGURE 3 | Differences in measured laboratory parameters. For patients in subgroup 1, laboratory parameters measured before (light blue) and after (blue/green) last FISH. For patients in subgroup 2, laboratory parameters measured before (light blue) and after (yellow) detection of the new clone. **(A)** κ and λ light chains measured in serum. **(B)** LDH activity. **(C)** Gradient of monoclonal protein (M-gradient).

Figure 4 provides an overview of therapies applied in subgroups 1 and 2 (for detailed information on therapy for every patient, see **Supplementary Figures 5–29**). In general, patients in both subgroups received comparable therapies. However, while patients in subgroup 1 were characterized by many therapies being applied for a long time (**Figure 4A**), switches in therapy after a short time could be observed for subgroup 2 (**Figure 4B**). While all patients received at least one therapy for ≥100 days, 83% (10 out of 12) received at least one therapy for ≥200 days in subgroup 1 vs. 55% (5 out of 9) in subgroup 2 (75% vs. 33% for therapies received for ≥300 days).

We compared both subgroups with respect to the top 5 therapies applied for the longest time (**Figure 4C**). Only marginal differences could be observed for therapies 2 to 5.

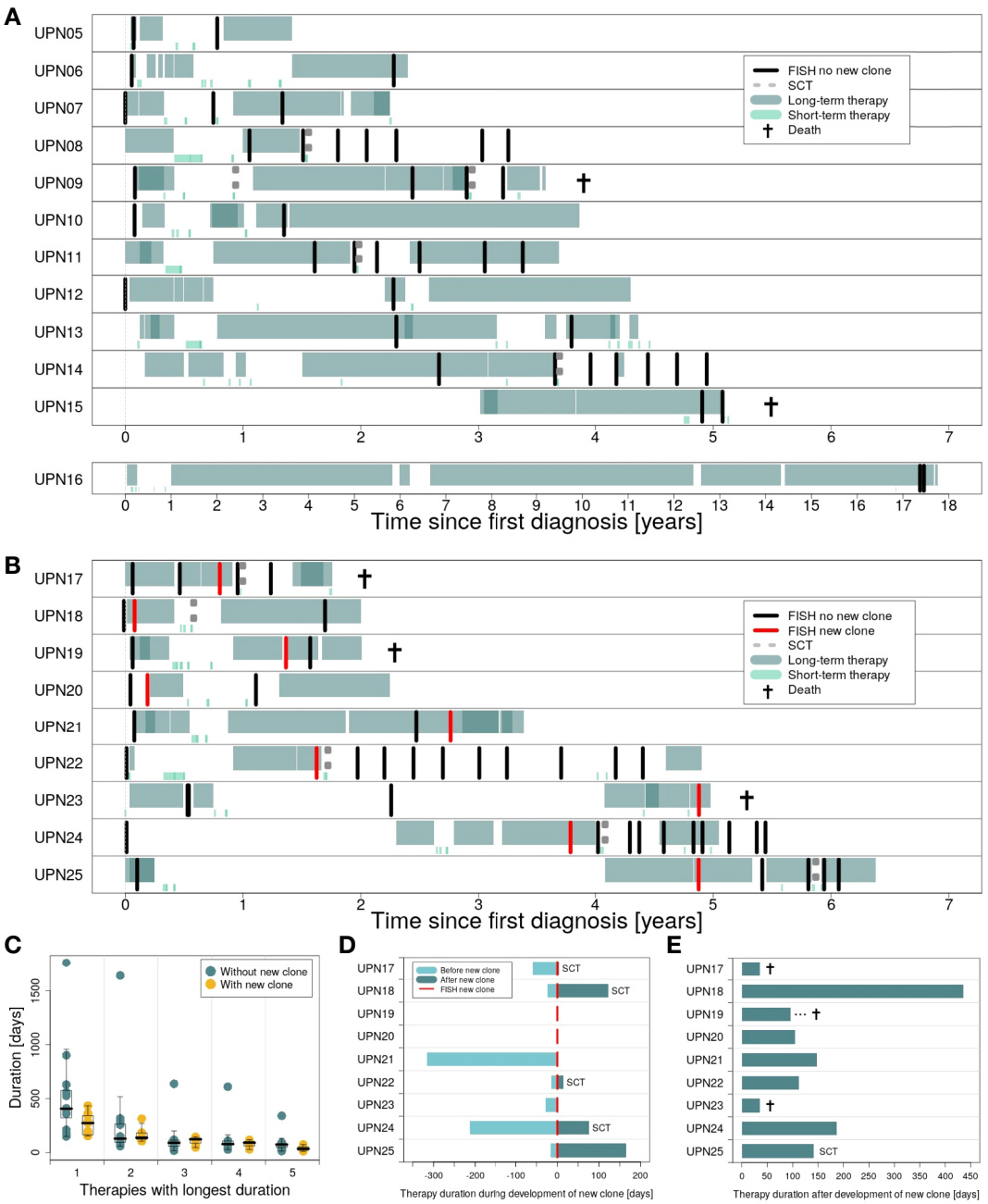


FIGURE 4 | Duration of therapies applied in the course of disease for patients in subgroups 1 and 2. **(A)** Long-term vs. short-term therapies for patients in subgroup 1 (patients UPN05 to 16). **(B)** Long-term vs. short-term therapies for patients in subgroup 2 (patients UPN17 to 25). **(C)** Therapies with the longest duration comparing subgroup 1 vs. subgroup 2. **(D)** Duration of therapies for subgroup 2, applied during development of the new clone. **(E)** Duration of the first therapy applied after development of the new clone.

However, for the therapy applied for the longest time, subgroups showed major differences. Analysis of our data by Mann–Whitney *U* test revealed a significant difference between both subgroups: For patients with new clones, duration of the longest therapeutic regimes was significantly shorter compared to patients without new clones (p = 0.0465; median subgroup 1: 406.5 days, range: 150–1,764 days; median subgroup 2: 273 days, range: 152–434 days).

Detailed information on the duration of therapies applied during and after development of the new clones for patients in subgroup 2 is provided in **Figures 4D, E**. For two patients, UPN19 and 20, the new clone was developed in between two

therapeutic regimes. Five patients in subgroup 2 received SCT, with four out of 5 right after the development of the new clone. Immediate change in therapy after the development of the new clone was observed for 3 patients (UPN17, 21, and 23), of whom two died within less than 50 days.

4 DISCUSSION

Clonal evolution and—as a result—increasing clonal heterogeneity have been studied widely in MM (22–27). It can be observed not only during the disease course as an adaptation to different treatments, but also at the same time point of the disease stage with different clones being present at different locations of the body (e.g., in focal lesions and the iliac crest; 28). Even within a compartment, such as the bone marrow, the expression pattern of plasma cells can differ depending on their position (29). Adding further complexity is the fact that the patient's treatment course can be especially heterogeneous regarding the sequence of therapeutic regimes the patient received. Finally, clonal evolution can be observed on a rather large scale such as chromosomal alterations or smaller scale with mutations or small range deletions or amplifications.

In all cases, clonal evolution contributes to the fact that the disease course of MM is characterized by a shorter PFS from one line of therapy to the other (21). It remains to be defined how different clonal branches and different sites of the body at different time points might be best detected, classified, and accordingly treated. Yet, in clinical routine, performing whole-genome sequencing or biopsies from different sites is not feasible right now. Additionally, identifying these sites, e.g., by PET-CT or MRI, is equally difficult.

In this study, we performed an in-depth analysis of patients with MM. Our analyses focused on clinical differences in patients with and without new clones developing in the course of disease. Clones were characterized by larger-scale alterations, which could be detected by FISH—a method very widely established and rather affordable in an outpatient setting. Integration of data from a variety of sources—disease progression, chromosomal profiles, laboratory parameters, and applied therapies—provided us with the unique option to study the interplay of diverse features characterizing patients with MM. The limiting factors are that we focused on large-scale alterations, clones might have been missed due to limited sensitivity of FISH (30), and the disease course and treatment sequences are quite heterogeneous. Additionally, our patient cohort is rather small. Dividing our cohort into two subgroups—with and without new clones—further limits the statistical power of our results.

With these limitations in mind, we could make the following observations: In our cohort, new large-scale chromosomal aberrations developed quite frequently (9/21 patients—43%). Clones emerging with chromosomal aberrations frequently featured alteration of chromosomes 1 and 17 (44% each; 77% in total). Both chromosomes have been described to play a role in myeloma progression, with del(17p) affecting the function of *TP53* with its known negative factor influencing the course of hematological malignancies (31). In addition, chromosome 1 alterations have been shown to be associated with an inferior disease course leading to the amplification of genes such as *BCL9*, *MCL1*, *CKS1B*, and *ANP32E*. These genes are either implicated in the inhibition of apoptosis, enhancement of cell cycle, or epigenetic modification (9, 10, 12, 25, 32–34).

With all the limitations due to the small size of our cohort and thus limited statistical power, emergence of a large-scale genomic alteration negatively influenced length of response of therapy and potentially PFS and OS, although both did not reach significance. It also led to an overall shorter PFS to each line subsequent to the emergence of a new clone. This observation is in line with previously reported results on clonal evolution negatively influencing prognosis (35). We evaluated whether emergence of a new clone could be correlated with laboratory findings. LDH activity reached higher peaks after clonal evolution, although the specific courses were quite heterogeneous, reflecting that the disease might initially respond to therapy, but then again relapse.

Altogether, we can show that large-scale genomic clonal evolution is associated with an inferior disease course with shorter response duration to subsequent new lines, possibly shorter PFS and OS. It affects, at a high degree, chromosomes 1 and 17, which confer a negative prognosis already at initial diagnosis. These results warrant additional studies with larger groups treated more homogeneously for confirmation. For these groups, collection of detailed information on progression, blood parameters, and response to therapy during follow-up is required in addition to analysis of genomic alterations. The number of patients analyzed is a limitation of our study. Our goal was to use an approach to follow clonal evolution, which is widely used in a clinical setting. We screened, in our database, more than 600 patients and only found these patients, for whom we have consecutive FISH data combined with a variety of clinical data. The strength of our study is that we can connect clonal status with standard laboratory parameters and clinical outcome. Most clinical trials focus on the specific setting with follow-up to the next line of treatment, but long-term data over the course of different courses are not easily available. What we show is "real-world data" regarding frequency of large-scale genomic alterations and that clonal evolution is, as not surprising, really heterogeneous with regard to its association with different clinical parameters. Yet, the search for clonal large-scale genomic alterations can actually be achieved also in an outpatient setting. It would then require discussion with the patient, in which additional therapeutic steps might be required to counteract the effect of such variation.

DATA AVAILABILITY STATEMENT

Original data presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The study was reviewed and approved by ethical review boards of the University of Münster and the Ärztekammer Westfalen-Lippe (2018-452-f-S). The patients provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KK, RE, TJB, and CK collected patient samples. SS and KK performed data analyses. All authors interpreted and discussed the data. SS, KK, PB, CK, and EKM wrote the manuscript. CK, JV, and GL coordinated and supervised the project. All authors read, revised, and approved the final version of the manuscript.

FUNDING

The work was supported by the Jose Carreras Leukämie Foundation (DJCLS 17R/2018), partially by the Deutsche Krebshilfe (70112392), Deutsche Forschungsgemeinschaft (KH331/2-3), the intramural funding of the faculty of Medicine at University Hospital of Muenster (Kha2/002/20), and a grant from Bundesministerium für Bildung und Forschung (HiGHmed 01ZZ1802V).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.919278/full#supplementary-material

REFERENCES

1. Palumbo A, Anderson K. Multiple Myeloma. *N Engl J Med* (2011) 364:1046–60. doi: 10.1056/204NEJMra1011442205
2. Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G, Mateos M-V, et al. International Myeloma Working Group Updated Criteria for the Diagnosis of Multiple Myeloma. *Lancet Oncol* (2014) 15:e538–48. doi: 10.1016/S1470-2045(14)70442-5215
3. Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Richardson LR, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. *J Clin Oncol* (2015) 33:2863–9. doi: 10.1200/JCO.2015.61.2267208
4. Rajkumar SV. Multiple Myeloma: 2020 Update on Diagnosis, Risk-Stratification and Management. *Am J Hematol* (2020) 95:548–67. doi: 10.1002/ajh.25791212
5. Bolli N, Avet-Loiseau H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, et al. Heterogeneity of Genomic Evolution and Mutational Profiles in Multiple Myeloma. *Nat Commun* (2014) 5:496–511. doi: 10.1038/ncomms3997
6. Furukawa Y, Kikuchi J. Molecular Basis of Clonal Evolution in Multiple Myeloma. *Int J Hematol* (2020) 111:496–511. doi: 10.1007/s12185-020-02829-6199
7. Davis A, Gao R, Navin N. Tumor Evolution: Linear, Branching, Neutral or Punctuated? *Biochim Biophys Acta Rev Cancer* (2017) 1867::496–511. doi: 10.1016/j.bbcan.2017.01.003197
8. Castaneda O, Baz R. Multiple Myeloma Genomics - A Concise Review. *Acta Med Acad* (2019) 48:57–67. doi: 10.5644/ama2006-124.242192
9. Merz M, Jauch A, Hielscher T, Mai EK, Seckinger A, Hose D, et al. Longitudinal Fluorescence *in Situ* Hybridization Reveals Cytogenetic Evolution in Myeloma Relapsing After Autologous Transplantation. *Haematologica* (2017) 102:1432–8. doi: 10.3324/haematol.2017.168005
10. Merz M, Hielscher T, Schult D, Mai EK, Raab MS, Hillengass J, et al. Cytogenetic Subclone Formation and Evolution in Progressive Smoldering Multiple Myeloma. *Leukemia* (2019) 34:1192–6. doi: 10.1038/s41375-019-0634-2
11. Keats JJ, Chesi M, Egan JB, Garbitt VM, Palmer SE, Braggio E, et al. Clonal Competition With Alternating Dominance in Multiple Myeloma. *Blood* (2012) 120:1067–76. doi: 10.1182/blood-2012-01-405985
12. Merz M, Jauch A, Hielscher T, Bochtler T, Schönland SO, Seckinger A, et al. Prognostic Significance of Cytogenetic Heterogeneity in Patients With Newly Diagnosed Multiple Myeloma. *Blood Adv* (2017) 2:1–9. doi: 10.1182/bloodadvances.2017013334
13. Bayani J, Squire JA. Fluorescence *in Situ* Hybridization (FISH). *Curr Protoc Cell Biol* (2004) 22:22.4.1–52. doi: 10.1002/0471143030.cb2204s23187
14. Brune T, Riepe FG, Beier K, Exeler R, Louwen F, Garritsen H. Differentiation Of188single Populations in a Bidirectional Mixed Lymphocyte Culture Using X and Y Chromosome-Specific FiSH Markers. *J Immunol Methods* (2002) 266:105–10. doi: 10.1016/s0022-1759(02)00107-2190
15. da Silva-Coelho P, Kroeze LI, Yoshida K, Koorenhof-Scheele TN, Knops R, van de Locht LT, et al. Clonal Evolution in Myelodysplastic Syndromes. *Nat Commun* (2017) 8:639–43. doi: 10.1038/194ncomms15099. 193.
16. Reutter K, Sandmann S, Rohde J, Müller S, Wöste M, Khanam T, et al. Reconstructing Clonal Evolution in Relapsed and non-Relapsed Burkitt Lymphoma. *Leukemia* (2021) 35:639–43. doi: 10.1038/217s41375-020-0862-5218
17. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing (2020).
18. Miller CA, McMichael J, Dang HX, Maher CA, Ding L, Ley TJ, et al. Visualizing Tumor Evolution With the Fishplot Package for R. *BMC Genomics* (2016) 17. doi: 10.1186/s12864-016-3195-z203
19. Therneau TM. *A Package for Survival Analysis in R*. (2020), R package version 3.2-7. Available at: (https://CRAN.R-project.org/package=survival)
20. Kassambara A, Kosinski M, Biecek P. *Survminer: Drawing Survival Curves Using 'Ggplot2'*. (2021), R package version 0.4.9201. Avaialble at: https://cran.r-project.org/web/packages/survminer/
21. Gandhi UH, Cornell RF, Lakshman A, Gahvari ZJ, McGehee E, Jagosky MH, et al. Outcomes of Patients With Multiple Myeloma Refractory to CD38-Targeted Monoclonal Antibody Therapy. *Leukemia* (2019) 33:2266–75. doi: 10.1038/s41375-019-0435-7
22. Brioli A, Melchor L, Titley I, Vijayaraghavan G, Stephens C, Zeisig A, et al. The Impact of Long-Term Lenalidomide Exposure on the Cellular Composition of Bone Marrow. *Leuk Lymphoma* (2014) 55:2665–68. doi: 10.3109/10428194.2014.900765
23. Dutta AK, Alberge J-B, Sklavenitis-Pistofidis R, Lightbody ED, Getz G, Ghobrial IM. Single-Cell Profiling of Tumour Evolution in Multiple Myeloma - Opportunities for Precision Medicine. *Nat Rev Clin Oncol* (2022) 19:223–36. doi: 10.1038/s41571-021-00593-y
24. Melchor L, Brioli A, Wardell CP, Murison A, Potter NE, Kaiser MF, et al. Single-Cell Genetic Analysis Reveals the Composition of Initiating Clones and Phylogenetic Patterns of Branching and Parallel Evolution in Myeloma. *Leukemia* (2014) 28:1705–15. doi: 10.1038/leu.2014.13
25. Shah V, Johnson DC, Sherborne AL, Ellis S, Aldridge FM, Howard-Reeves J, et al. Subclonal TP53 Copy Number is Associated With Prognosis in Multiple Myeloma. *Blood* (2018) 132:2465–69. doi: 10.1182/blood-2018-06-857250
26. Shen YJ, Mishima Y, Shi J, Sklavenitis-Pistofidis R, Redd RA, Moschetta M, et al. Progression Signature Underlies Clonal Evolution and Dissemination of Multiple Myeloma. *Blood* (2021) 137:2360–72. doi: 10.1182/blood.2020005885
27. Walker BA, Wardell CP, Melchor L, Brioli A, Johnson DC, Kaiser MF, et al. Intraclonal Heterogeneity is a Critical Early Event in the Development of Myeloma and Precedes the Development of Clinical Symptoms. *Leukemia* (2014) 28:384–90. doi: 10.1038/leu.2013.199

28. Rasche L, Chavan SS, Stephens OW, Patel PH, Tytarenko R, Ashby C, et al. Spatial Genomic Heterogeneity in Multiple Myeloma Revealed by Multi-Region Sequencing. *Nat Commun* (2017) 8. doi: 10.1038/s41467-017-00296-y
29. Merz M, Merz AMA, Wang J, Wei L, Hu Q, Hutson N, et al. Deciphering Spatial Genomic Heterogeneity at a Single Cell Resolution in Multiple Myeloma. *Nat Commun* (2022) 13. doi: 10.1038/s41467-022-28266-z
30. Avet-Loiseau H, Lannes R, Perrot A, Mazzotti C, Divoux M, Cazaubiel T, et al. In Multiple Myeloma, High-Risk Secondary Genetic Events Observed at Relapse Are Present From the Diagnosis in Tiny Undetectable Subclones. *Blood* (2021) 138(Supplement 1):77. doi: 10.1182/blood-2021-146869
31. Stengel A, Kern W, Haferlach T, Meggendorfer M, Fasan A, Haferlach C. The Impact of TP53 Mutations and TP53 Deletions on Survival Varies Between AML, ALL, MDS and CLL: An Analysis of 3307 Cases. *Leukemia* (2017) 31:705–11. doi: 10.1038/leu.2016.263
32. Carrasco DR, Tonon G, Huang Y, Zhang Y, Sinha R, Feng B, et al. High-Resolution Genomic Profiles Define Distinct Clinico-Pathogenetic Subgroups of Multiple Myeloma Patients. *Cancer Cell* (2006) 9:313–25. doi: 10.1016/j.ccr.2006.03.019
33. Hanamura I, Stewart JP, Huang Y, Zhan F, Santra M, Sawyer JR, et al. Frequent Gain of Chromosome Band 1q21 in Plasma-Cell Dyscrasias Detected by Fluorescence *in Situ* Hybridization: Incidence Increases From MGUS to Relapsed Myeloma and is Related to Prognosis and Disease Progression Following Tandem Stem-Cell Transplantation. *Blood* (2006) 108:1724–32. doi: 10.1182/blood-2006-03-009910
34. Marchesini M, Ogoti Y, Fiorini E, Samur AA, Nezi L, D'Anca M, et al. ILF2 Is a Regulator of RNA Splicing and DNA Damage Response in 1q21-Amplified Multiple Myeloma. *Cancer Cell* (2017) 32:88–100. doi: 10.1016/j.ccell.2017.05.011
35. Landau DA, Carter SL, Getz G, Wu CJ. Clonal Evolution in Hematological Malignancies and Therapeutic Implications. *Leukemia* (2014) 28:34–43. doi: 10.1038/leu.2013.248

Conflict of Interest: EKM reports consulting or Advisory Role, Honoraria, Research Funding, and Travel Accommodations and Expenses from Bristol Myers Squibb/Celgene, GlaxoSmithKline, Janssen Cilag, Sanofi and Takeda.

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

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

frontiers | Frontiers in Oncology

TYPE Review
PUBLISHED 08 August 2022
DOI 10.3389/fonc.2022.945376

OPEN ACCESS

EDITED BY
Nan-Shan Chang,
National Cheng Kung University, Taiwan

REVIEWED BY
Thomas Seufferlein,
University of Ulm, Germany
Bong Hwan Sung,
Vanderbilt University, United States

*CORRESPONDENCE
Gábor Valcz
valcz.gabor@med.semmelweis-univ.hu

SPECIALTY SECTION
This article was submitted to Molecular and Cellular Oncology, a section of the journal Frontiers in Oncology

RECEIVED 16 May 2022
ACCEPTED 06 July 2022
PUBLISHED 08 August 2022

CITATION
Valcz G, Újvári B, Buzás EI, Krenács T, Spisák S, Kittel Á, Tulassay Z, Igaz P, Takács I and Molnár B (2022) Small extracellular vesicle DNA-mediated horizontal gene transfer as a driving force for tumor evolution: Facts and riddles.
Front. Oncol. 12:945376.
doi: 10.3389/fonc.2022.945376

Small extracellular vesicle DNA-mediated horizontal gene transfer as a driving force for tumor evolution: Facts and riddles

Gábor Valcz[1]*, Beáta Újvári[2], Edit I. Buzás[3,4,5], Tibor Krenács[6], Sándor Spisák[7], Ágnes Kittel[8], Zsolt Tulassay[1], Péter Igaz[1,9,10], István Takács[9] and Béla Molnár[1,9]

[1]MTA-SE Molecular Medicine Research Group, Eötvös Loránd Research Network, Budapest, Hungary, [2]School of Life and Environmental Sciences, Centre for Integrative Ecology, Deakin University, Waurn Ponds, VIC, Australia, [3]Department of Genetics, Cell- and Immunobiology, Semmelweis University, Budapest, Hungary, [4]ELKH-SE Immune-Proteogenomics Extracellular Vesicle Research Group, Semmelweis University, Budapest, Hungary, [5]HCEMM-SU Extracellular Vesicle Research Group, Semmelweis University, Budapest, Hungary, [6]1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary, [7]Institute of Enzymology, Research Centre for Natural Sciences, Budapest, Hungary, [8]Institute of Experimental Medicine, Eötvös Loránd Research Network, Budapest, Hungary, [9]Department of Internal Medicine and Oncology, Semmelweis University, Budapest, Hungary, [10]Department of Endocrinology, Semmelweis University, Budapest, Hungary

The basis of the conventional gene-centric view on tumor evolution is that vertically inherited mutations largely define the properties of tumor cells. In recent years, however, accumulating evidence shows that both the tumor cells and their microenvironment may acquire external, non-vertically inherited genetic properties *via* horizontal gene transfer (HGT), particularly through small extracellular vesicles (sEVs). Many phases of sEV-mediated HGT have been described, such as DNA packaging into small vesicles, their release, uptake by recipient cells, and incorporation of sEV-DNA into the recipient genome to modify the phenotype and properties of cells. Recent techniques in sEV separation, genome sequencing and editing, as well as the identification of new secretion mechanisms, shed light on a number of additional details of this phenomenon. Here, we discuss the key features of this form of gene transfer and make an attempt to draw relevant conclusions on the contribution of HGT to tumor evolution.

KEYWORDS

extracellular vesicles, exosomes, horizontal gene transfer, tumor evolution, cell-cell communication

Introduction

Generally, it is considered that cancer cells divide by mitosis and do not exchange DNA with each other or with the cells of the tumor microenvironment (1, 2). The vertically inherited mutations across successive generations largely (but not exclusively) determine the adaptive reaction of the cell's progeny against the existing selective pressure and result in the expansion or contraction of their subclones (1, 2). Recent studies suggested that a non-vertical transmission of DNA may also occur among the community of cancer cells (oncobiota) and between cancer- and microenvironmental genomes (3, 4). This process, namely horizontal gene transfer (HGT), may provide a selective advantage for the recipient cell if the overall effects of the transferred gene are beneficial (5). This phenomenon is considered as a non-cell-autonomous mechanism, in which higher genotype's adaptive value of the recipient cell is partially linked to the donor cells. HGT may accelerate genomic evolution by allowing a faster adaptation in the group of recipient cells than it would happen by vertical transfer of the same gene (5). This exchange of genetic material occurs among both nearby cells (as paracrine signals) and relatively distant ones (in an endocrine manner) (3, 6). Biologically active cell-free DNA (cfDNA), as a mediator of HGT, can be transported in several forms in the intercellular space (Box 1). Here, we particularly focus on the cfDNA-carrying lipid bilayer membrane-enclosed extracellular vesicles (EVs) which can be intercellular mediators of biological and cellular functions. EVs are secreted by most (if not all) cells both under physiological and pathological conditions (14). The cfDNA encapsulation by EVs confers enhanced stability to the transported genomic material (15). Thus, the increased EV secretion of cancer cells (compared to normal ones) and the expanded appearance of specific, clinically relevant mutations found in conveyed DNA allow for the detection and monitoring of tumors using liquid biopsy applications (16–18). However, the evolutionary effect of the released mutant genes inserted into the genome of recipient cells is less known. As we present in Box 1, EVs are divided into different subpopulations based on their biogenesis. Given that in most instances there is no direct evidence for the biogenetics route of a given vesicle, an operational classification based on EV sizes can be used. From all EV subcategories, apoptotic cell-derived large EVs (i.e., apoptotic bodies) were first described as mediators of oncogenic HGT (6). Active secretion of the EV-conveyed DNA associated with small EVs (sEVs, Box 1) may be particularly important from a clinical perspective, because sEVs from surviving cells may deliver "tried and tested" genes that stimulate fast protective functions against the intense stress factors. Here, we will discuss the main steps of the sEV-DNA-mediated HGT among donor and recipient cells in human cancers, and where it is possible, we compare it to the processes of normal mammalian cells. We particularly focus on the potential biological relevance of sEV-DNA and the controversial issues association with HGT in tumor evolution.

Release of DNA through sEV secretion

The majority of cancers are derived from a single ancestral cell by the generation of a diverse successor population with subclonal architecture (1). Cancer also shapes its own microenvironment into a supportive one (19). The developing genetic-, epigenetic-functional- and phenotypic heterogeneity of individual cells provides a remarkable capacity for a population to adapt to challenging environmental conditions during cancer progression and therapy (20, 21). HGT can significantly influence the evolutionary trajectory of a given tumor by spreading genes encoding for molecules which provide advantages for the cells with suboptimal survival, expansion or metastatic capacity. EV-mediated HGT is not a common event among healthy mammalian cells (22). However, the fundamental differences between cancer and normal cells may change the frequency of the EV-mediated HGT. The first alteration in cancer can be the abnormal transport of the genomic DNA (gDNA) from the nucleus to the cytoplasm. Aberrant expression of nuclear membrane components, abnormalities in chromosome segregation, and mechanical forces from the actin cytoskeleton resulting in the rupture of the nuclear envelope are significantly more often observed in cancer than in normal cells, where the disintegration of the nuclear membrane is

BOX 1 Carriers of cfDNA, EV subcategories, sEVs, exosomes and non-conventionally released vesicles.

CfDNA as a mediator of HGT, can be present in different forms in the extracellular space, namely as DNA fragments, virtosomes [a complex of DNAs, RNAs, proteins, and lipids (7)], nucleosomes (DNA wrapped around an octamer of histone proteins), or packaged into extracellular vesicles (EVs) (8). The EV nomenclature refers to the EV biogenesis including: i) exosomes (sEVs with endosome/multivesicular body (MVB) origin; ~50-100 nm), ii) microvesicles (originated by direct budding/blebbing from the cell surface; 100-1000 nm) and iii) large EVs including apoptotic bodies (products of apoptotic cell disassembly; 1-5 μm) (9). In most instances, only the diameter of EVs can be determined with certainty. In this review, we follow the MISEV 2018 guidelines which suggest the term sEVs for EVs smaller than 200 nm in diameter, regardless of their origin (10). It follows from the above that the sEV term denotes a group of EVs with heterogeneous origins. Exosomes represent a subset of sEVs which are formed intracellularly by the inward budding of the limiting membrane of endosomes/MVBs (with the intrusion of the cytosolic components). Later, the MVBs fuse with the plasma membrane, so their intraluminal vesicles are released into extracellular space as exosomes. After the uptake, exosomes are known to mediate a wide spectrum of effects on the recipient cell (11).

Recently, unconventional sEV release mechanisms have been hypothesized. In this case, migrasomes and *en bloc* released MVB-like EV clusters could possibly serve as sources of sEVs upon rupture of their limiting membrane (12, 13).

transient and limited to the mitosis (23). Disruption of the membrane barrier around the chromosomes allows gDNA to be exposed to cytoplasmic locations outside of the nucleus (Figure 1A) (23). Another form of delivery of gDNA to the cytoplasm may occur through micronucleus formation (Figure 1B), when a chromosome (or part of it) segregates improperly during mitosis and recruits its own nuclear envelope outside of the primary nuclear membrane (24, 25). Other, phenomena which are not studied in association with oncogenic sEV release, such as nucleophagy might also take a part in this process (26). Cytoplasmic or micronucleus-enclosed gDNA can translocate into the intraluminal vesicles of multivesicular bodies (MVBs) (27) (see exosomes in Box 1 and Figure 1C). Although the molecular background of DNA packaging to sEVs remains largely obscure, some interesting details have been described recently. Such a process is the interaction of CD63 exosome-associated tetraspanin with a DNA binding protein (i.e., the presence of CD63-Histone H2B-gDNA complex) which potentially plays a key role in the loading of micronuclear gDNA into sEVs (24). The sEV release is involved in the ablation of potentially harmful or damaged genetic material from the cell (28), suggesting that sEV-mediated gDNA release may be at a baseline level both in normal cells as well as cancer cells to maintain cell homeostasis. However, it is particularly important that this system can adapt quickly to stress (e.g., genotoxic oncotherapies) by increasing micronucleus formation and the concomitant packaging of gDNA into sEVs (sEV-gDNA) (Figure 1D) (24). The dynamic adaptation of sEV-mediated DNA release is further supported by the changing quality and/or quantity of transported genetic content (e.g., the proportion of genomic and mitochondrial (mt) DNA, as well as the average size of DNA fragments) upon diverse environmental effects (29, 30).

Considering that the sEVs-conveyed gDNA fragments represent the entire host genome and contain full-length oncogene sequences (31) (Figure 1E), they are promising tools for the diagnosis and treatment monitoring of cancer patients by detecting mutations characteristic of the given tumor type (32, 33). However, several recently published studies report conflicting results on the actual DNA transport capacity of sEVs (details in Box 2, the described secretion mechanism is indicated in Figure 1F). As described in Box 2, some studies only consider luminal (protected, intravesicular) DNA as a genetic material delivered by sEVs, although DNA can also be associated with the surface of sEVs. Therefore, the DNA transport capacity of sEVs may be underestimated or misinterpreted.

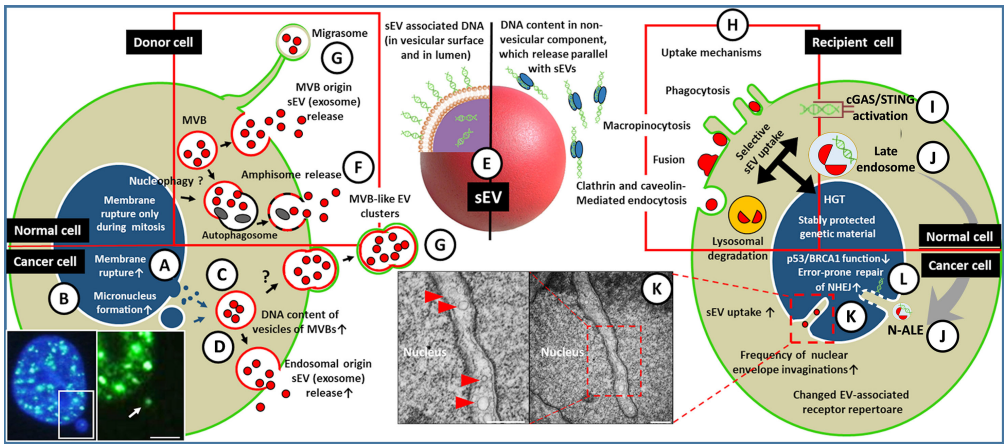


FIGURE 1
A simplified representation of sEV-mediated HGT among tumor/microenvironmental cells. **(A)** The nuclear gDNA is discharged to the cytoplasmic region of the EV releasing cell (left) through the rupture of the nuclear envelope or by the formation of micronucleus (the fluorescent microscopic image shows a double-strand break (DSB) site in the micronucleus of HT-29 cell [white arrow, γH2AX staining, scale bar: 2 μm)] **(B)**. **(C)** From the cytoplasm or from micronuclei the gDNA translocates into the intraluminal vesicles (future exosomes) of multivesicular bodies. **(D)** Both the gDNA content of the exosomes and their release are increased in tumor cells, especially upon the effect of therapeutic stress. **(E)** The gDNA may be transferred either in the lumen of sEVs and/or on the exofacial EV surface, or independently as a non-vesicular component. The gDNA content of sEVs might depend on their origin, like exocytosis of MVBs **(D)**, amphisomes **(F)**, or sEV discharge from *en bloc* released MVB-like EV clusters or migrasomes **(G)**. The red frame indicates that all listed processes (i.e., migrasome formation, exosome secretion, amphisome exocytosis) may occur in both normal and tumor cells. **(H)** Uptake of the released sEVs by a recipient cell (right) may include receptor-mediated processes (characteristic of both normal and tumor cells, indicated by a red frame). **(I)** In the cytoplasm, the sEVs (or their components) may activate DNA recognition pathways e.g., cGAS/STING. **(J)** The EV-containing late endosomes may reach the invaginations of the nucleus (as nuclear envelope invagination-associated late endosome/N-ALE) where its parts (probably including EV-DNA) may enter the nucleus through nuclear pores. **(K)** Small vesicles (<200 nm) are also detectable in association with the nuclear membrane invaginations (red arrowheads in electron microphotographs of HT29 colorectal cancer cells). The origin of these vesicular structures has not been examined (scale bars: 500 nm). **(L)** The integration of gDNA into the recipient genome may require malfunction of the host DNA repair or onco-suppressor mechanisms (e.g., p53, and BRCA1). Many of the processes presented here have only been described in relation to cancers. Further studies are needed to demonstrate whether these occur in healthy cells.

BOX 2 Questions about the ability of sEVs to transport DNA.

It is generally accepted that sEVs carry higher amounts of double-stranded DNA as compared to the single-stranded one determined by using DNases that differentially recognize and digest the two types of DNA (such as Shrimp dsDNase and S1 nuclease) (16). Besides gDNA, the full mitochondrial genome is also identified in sEVs (34, 35). Furthermore, enhanced amounts of sEV-associated mtDNA and tumor-specific gDNA have been described in the blood of cancer patients (35, 36). This latter property allows sEV-based identification of informative mutations from liquid biopsy applications, supporting the early detection and diagnosis of cancer as well as monitoring the treatment response (15). This established view is questioned by the suggestion that DNA released by cells is localized in DNaseI-sensitive, non-vesicular structures (nucleosomes) released by the exocytosis of amphisomes (37). Although the release of other non-membranous, small (~ 30 nm) nanoparticles (i.e., exomers) has also been described, the authors have detected DNA in sEVs with cell-type-specific relative abundance (38). Furthermore, many earlier studies used DNase treatment for the examination of luminal (protected) DNA [e.g., (22, 31, 35, 39)]. Importantly *Thakur et al.* described that the large (> 2.5 kb) double-stranded DNA fragments bind to the exofacial surface of exosomes, while the size of the luminal DNA fragments ranges from 100 bp to 2.5 kbp (16).

Although DNase treatment has confirmed the presence of luminal DNA content in sEVs in these studies, the question arises whether the DNA removed from the exofacial surface can indeed be classified as an artifact or the surface-associated DNA cargo is a native property of sEVs. This latter possibility was confirmed by an artificial selection pressure induced release of both DNA and DNA-binding proteins on the exofacial surface of sEVs *in vitro* (29). Based on this observation, it should be considered, that although DNase digestion removes potential foreign contaminants of sEVs, it may also eliminate the external, physiological, or pathophysiological DNA cargo from their exofacial vesicular surface.

Conceivably, we propose, that DNA from the microenvironment may also be transferred onto the sEV surface in the extracellular space (as part of a biomolecular corona). Thus, the genetic information carried by an sEV is not necessarily limited to a single EV donor cell.

The DNA cargo has been reported to be present only in a small proportion of sEVs (~10% *in vitro,* ~1% *in vivo*) (16, 24) suggesting that these vesicles are heterogeneous in this respect. This heterogeneity has also been confirmed using high-resolution iodixanol density gradients, discriminating high-density sEV fraction with enriched DNA content and a low-density sEV fraction which carried minor quantities of DNA (40). Because the sEV fraction separated by ultracentrifugation can be derived from multiple cell sources, we need to emphasize the heterogeneity both within sEVs of the same biogenetic origin (e.g., exosome subpopulations) and sEVs released from different subcellular structures (i.e., MVBs, migrasomes, MVB-like EV clusters, Figure 1G, Box 1).

Probably, the packaging of the genomic nuclear content to sEVs cannot be simplified to a yes/no question. As detailed above, the gDNA content of the cytoplasmic region may be highly dependent on the existence of certain pathophysiological processes in the cell (e.g., nuclear membrane rupture and micronuclei formation) (23–25). It may lead to an increased gDNA content in MVB-originated sEVs (i.e., exosomes) (24, 27). The rapid change in gDNA content of exosomes by existing stress factors (e.g., effects of artificial selection or the pro-inflammatory microenvironment) have been observed (24, 29, 30), however, it is not known how similar stress conditions may regulate the non-MVB-originated sEV release. Presumably, these sEV release pathways may change dynamically upon exposure to various microenvironmental or therapy-induced stress factors.

Uptake and functions of sEV-gDNA in recipient cells

The tumor can be considered as an ecosystem in which both the cells and the subclonal populations cooperate with each other for acquiring space and resources from the host organism e.g., by sharing molecules with beneficial local and/or systemic effects (41–43). After the sEV-protected gDNA survives the release and is present in the blood or in other body fluids, the recipient cells may internalize it from the extracellular space [the known sEV uptake mechanisms are shown in Figure 1H, and were reviewed by McKelvey et al. (44)]. In fact, this process is not trivial, so the selective sEV uptake may be the first barrier to the spread of vesicle-carried genes with oncogenic potential. This selectivity shows the dependency on the type, oncogene status, and receptor repertoire of the recipient cells. Substantial differences in sEV uptake have been observed between normal epithelial cells compared to the tumorigenic variant of them (45), resulting in functional changes in the latter. This study also described robust sEV uptake by fibroblasts (compared to normal epithelial cells), although these fibroblasts lacked proper tumor suppressor mechanisms (45). In the absence of control fibroblasts, it is questionable whether increased sEV uptake was caused by cell-type dependency or other changes, associated with the malfunction of the tumor suppressor mechanisms (see below). The image is further refined by the observation that sEV adhesion to fibroblasts shows integrin receptor dependency resulting in a non-random (i.e., organotropic) metastasis formation (46). On the other hand, the sEV tetraspanin web also plays a primary role in selective target cell binding (i.e., formation of tetraspanin–integrin complexes) (47), which drives attention to the significance of proper donor and recipient cell selection when designing experiments. We have very limited knowledge about the uptake of unconventionally secreted sEVs [such as migrasomes and *en bloc* released MVB-like EV clusters (12, 13)]. Although uptake of migrasomes by recipient cells has been described *in vitro* (12), the details of this process are unknown.

The sEV-DNA uptake may be a robust process as it has been detected in ~16% of sEV-treated fibroblasts (48). The DNA, which enters the cytoplasm does not necessarily reach the nucleus but can activate cytoplasmic DNA recognition

receptors (Figure 1I). For instance, radiation or chemotherapy-induced, tumor-derived sEV-gDNA triggers antitumor immune response in dendritic cells by activating the cGAS/STING pathway (49, 50). In other cases, sEV-containing late endosomes have been shown to migrate to the invaginations of the nuclear envelope (51), where hypothetically, they can exchange genomic content between the donor sEVs and the recipient cell (Figure 1J). However, evidence for such an exchange is still lacking. The EV components may be released from the envelope invagination-associated late endosomes to the narrow space between the endosomal and nuclear membranes and might translocate into the nucleoplasm *via* nuclear pores (51). Ultrastructural studies of our research group detected single, small (<200 nm) vesicles within the nuclear invaginations of cancer cells (Figure 1K). However, the origin of these vesicles remains unknown, and obviously, their potential role in HGT should be investigated.

The last step of the oncogenic HGT is the process in which the transferred gDNA integrates into the recipient genome. Some factors have been described in prokaryotes and eukaryotes, which influence the rates of acceptance of the horizontally transferred genes, such as physical/biological properties of the acquired DNA (e.g., length, GC content, codon usage, epigenetic marks, and the complexity of interactions with other genes) as well as the location of genetic integration in the recipient genome (52, 53). However, the last phase of HGT between mammalian cells, especially in sEV-mediated processes has been less examined. Interestingly, genome engineering provides an opportunity for a more in-depth study of this process. The integration of donor DNA sequences at off-target double-strand breaks (DSBs) has been described during CRISPR-Cas9-assisted genome editing, which is known to be caused by an error-prone repair of a non-homologous end joining (NHEJ) pathway (54). The presence of bovine gDNA sequences were detectable in the genome of NIH-3T3 fibroblasts using a medium containing 10% fetal bovine serum. This effect was significantly reduced with the use of exosome-depleted but cfDNA-containing medium, suggesting a primary role of sEVs in HGT (54). Tumor suppressor mechanisms such as the one mediated by p53 may affect the success of genome editing (55, 56). For example, the DSBs can be toxic for human pluripotent stem cells in a p53-dependent manner (55). Similarly, to genome editing, in cancer evolution, the erroneous rejoining of DNA has been shown to also generate genomic changes at DSBs sites (57). This raises the possibility that the capture of exogenous oncogenic sequences at DSB sites might be an evolutionary driving force of tumors. Considering that p53 plays a fundamental role in the fidelity control of NHEJ (58), loss of p53 function can improve tumor cell survival, and in parallel, may create the opportunity for possible HGT. This hypothesis is supported by a pioneering work that showed that p53 dysfunction is required for the incorporation of oncogenes into the recipient genome during apoptotic EV-mediated HGT (6). The fidelity of DNA end-joining was impaired also in the case of *breast cancer gene 1* (*BRCA1*) mutations (Figure 1L) (59, 60). The involvement of this tumor suppressor in sEV-mediated HGT is evidenced by the fact that a successful sEV-mediated DNA transfer has been described in *BRCA1*-KO fibroblasts in contrast to wild-type control cells (48).

A successful sEV-mediated HGT can be confirmed *in vitro* by genomic profiling of the transformed cell, detecting integrated gDNA, or its transcription products, phenotypical transformation as well as the new functions associated with the transferred genes. It should not be forgotten that the sEVs transport complex sets of information (61). Thus, the appearance of some new, cancer-associated functions may not only result from the transfer of sEV-gDNA, but also from several other sEV-conveyed regulator molecules. According to the principle of Darwinian selection, the transferred genes can spread in the population, if it carries genetic components that act as drivers associated with the host cell phenotype (5, 6, 45). Consequently, during the study of the long-term effects of sEV-mediated HGT, the tumor evolutionary aspects should also be considered (see below and Box 3).

The potential impact of sEV-mediated HGT in tumor evolution

The tumor-associated sEV signaling, including the delivery of aberrantly released molecule packets (e.g., proteins, lipids, metabolites, coding- and non-coding nucleic acids) influences the evolutionary events of tumors by various, often parallel cellular processes (45, 80). Being the most cancer-specific component of this complex system, here we particularly focused on sEV-delivered gDNA in tumor evolution. Although oncogenic HGT is less known among tumor evolutionary biologists, the *in vitro* and *in vivo* results discussed here clearly indicate that this process is more than a theoretical phenomenon. The presence of clonal heterogeneity in cancer (81) suggests that the HGT-based cooperation among the admixed- or the spatially non-uniformly distributed subclones may be a rare event. Regarding its frequency, it must be emphasized that some phenomena associated with oncogenic HGT can be highly context-dependent (23, 24, 29, 30). Thus, the successful incorporation rate of sEV-mediated HGT may differ greatly depending on the imposed selection pressure. Cellular experiments modeling sEV-gDNA transfer under diverse selective pressures would be clinically relevant for mapping the transfer of resistance mutations between cells of

BOX 3 Horizontal gene transfers to the rescue - Overcoming genomic decay, Muller's ratchet and metabolic exhaustion in cancer cells

A cancer cell's fitness is governed by its own proliferation; thus, the underlying Darwinian dynamics will select for proliferative self-renewal, territorial expansion, migration and invasion properties that procure higher fitness (62, 63). However, the propagation of clonal cancer cells by asexual reproduction exposes them to the emergence and accumulation of recessive mutations (termed "Muller's ratchet"). While cancer progression has largely been attributed to selection driving the accumulation of a certain number of somatic mutations, moderately deleterious mutations with no role in cancer (passengers) can accumulate as they largely evade natural selection, and thus negatively alter the cancer evolutionary landscape (64, 65). In the absence of meiotic recombination that would purge deleterious mutations in sexually reproducing organisms, and thus prokaryotes largely rely on horizontal gene transfer to restore and augment genetic diversity (66, 67). While direct evidence is so far lacking for nuclear cancer genomes to rely on HGT to mediate Muller's ratchet, evidence of capturing host mtDNA to prevent deleterious homoplasmy and loss of mitochondrial function emerges from both human and animal cancer studies (68–70). For example, EVs have been found to harbor and transfer full mitochondrial genomes to cells with impaired metabolism, and thus restore the metabolic activity of breast cancer (34). Mitochondria exchange between leukemic cells and mesenchymal stem cells has also been found to enhance the survival and therapy resistance of leukemia cells (71). In addition, studies show that tumor cells receive mtDNA from other cells of the body in order to maintain optimal cellular respiratory conditions to achieve metastasis (71, 72).

Conquering Muller's ratchet and maintaining metabolic potential is particularly important for the survival of transmissible cancer cell lines that are able to spread across hosts and hence are being passaged infinite number of times (69). One such transmissible cancer cell lines is the Canine Venereal Tumor (CTVT), a sexually transmitted malignant cell line that affects dogs (73). CTVT is the oldest known living cancer with an estimated age of between 4,000 and 11,000 years (74–77). Since its emergence in Asia (77), CTVT has spread across the globe, infected millions of dogs, and most likely experienced the accumulation and homoplasmy of deleterious mtDNA mutations. To avoid genomic melt-downs and metabolic catastrophes, CTVT has been found to capture and incorporate host mtDNA multiple times, as well as to occasionally employ mtDNA recombination and re-assortment during its evolutionary history (78). Replacement of part of the cancer mtDNA genome with sequence from the host mtDNA has also been observed in another transmissible cancer cell line, the bivalve transmissible neoplasia (BTN) from Chile (79). Whether exosomes have been facilitating HGT in these unique cancer cell lines, remain to be answered and an intriguing research area to follow up.

oncobiota. In addition, several experiments listed here focusing on cancer-fibroblast interactions suggested the role of sEV-mediated HGT in adaptive strategies to construct specialized niches. The development of a permissive and subsequently supportive stroma from the tumor suppressor microenvironment is a complex eco-evolutionary process (82), in which the recipient cells may incorporate genetic material from other cells (45, 48). Seemingly, the main requirement for this is that the recipient genome is not stably protected and/or repaired when damaged. This condition is met in carcinoma-associated fibroblasts upon genetic or epigenetic downregulation of *p53* and *BRCA1* [summarized in ref (83)], as well as apoptotic EV-conveyed *HPV16/18 E6* DNA have been shown to impair DNA repair mechanisms. These latter EVs were isolated from cervical cancer and contributed to the disruption of the p53/p21 pathway in primary fibroblasts (84).

By sEV-mediated HGT the recipient cells acquire adaptive benefits which can be manifested in increased proliferation, metastatic capacity, and foci-forming ability, reduced apoptosis, and the potential emergence of HLA-associated immune escape (31, 45, 48). However, these pioneering papers did not investigate HGT under intense therapeutic stress. The importance of therapy in an evolutionary context is highlighted by an observation about apoptotic EV-mediated HGT (6). [Here we note that although the apoptotic EVs arise as typical products of chemo- and radiotherapy, the transfer of full-length (3308 bp) sequences of *H-ras* (one of the examined genes in ref (6)) by sEVs was also described among living cells (31)]. When the incorporated DNA contains an advantageous mutation in the context of the treatment in question, it may become fixed, and it may spread among the recipient offspring cells. Thus, it may contribute to tumor evolution through several generations (6). Accordingly, EV-mediated HGT may greatly affect the sensitivity profile of the cells in residual disease [often undetectable, small population of malignant cells which persist after therapy (85)]. Consequently, its inheritance to the recipient genome may determine the properties of the recurring tumor.

Furthermore, sEVs from the primary tumor may influence critical events of metastasis, such as the preparation of pre-metastatic microenvironment (46) and may induce the formation of potentially metastatic tumor cells at least partly *via* HGT (48). The metastatic spread by genetic material has been known for a long time (86), and it is consistent with Darwin's pangenetic explanation (87). However, the role of sEVs in this process has only recently been studied and described. During the metastatic cascade, the phenotype of cancer cells shows dynamic changes, including epithelial-to-mesenchymal-, and mesenchymal-to-epithelial transitions (88, 89). The transition of *BRCA1*-KO fibroblasts to carcinoma-like cells (i.e., mesenchymal-to-epithelial transition) could be induced by sEV-gDNA (along with sEV-associated regulators) without preceding epithelial-to-mesenchymal transition (48). This suggests that the metastasic colonization is not exclusively due to the migration of primary tumor cells to metastatic sites [see the "seed to soil model" by Stephen Paget (90)], but it may also involve sEV-induced reprogramming of fibroblasts (48). In connection with this completely new phenomenon, further studies are needed to clarify as to whether carcinoma–like cells behave as tumor cells or as supporting microenvironment cells.

It is important to note that the above-mentioned studies examined the transition of one, or a few selected genes and their short functional effect with potential evolutionary benefits. However, parallel with beneficial genes, neutral or deleterious mutations are also conveyed into the recipient cells by HGT (5). Hypothetically, therefore, the role of oncogenic HGT in cancer

may be twofold: transferring deleterious genes may accelerate the irreversible accumulation of mutations which ultimately cause a mutational meltdown. Secondly, it may also increase the genetic diversity required for rapid adaptation by transferring beneficial genes (see Box 3 for further details).

Conclusion

Presumably, under physiological conditions, mammalian sEV-mediated HGT may not be extensive, and it has a restricted evolutionary impact. However, we want to emphasize that cancer-associated alterations in DNA repair, sEV secretion and uptake, and functional integration of the transmitted genes might modify the typical range and effect of HGT. Considering that cancer is a special evolutionary system, with its fast-growing, closely spaced large populations of cells that have similar genomes, the effect of rare events is likely to be increased as compared to the physiologic non-tumorous conditions. Mapping of the complex HGT phenomenon and integrating the knowledge reviewed here into our thinking of cancer development and progression may help to better interpret genomic data and allow the development of more precise tumor evolution models.

Author contributions

GV, BÚ, EIB, TK, SS and ÁK wrote the manuscript. PI, ZT, IT and BM provided the critical revisions. All authors approved the final version of the manuscript for submission and approved it for publication.

Funding

This work was funded by the NVKP_16-1-2016-0004 grant of the Hungarian National Research, Development and Innovation Office (NKFIH), as well as the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary, within the framework of the molecular biology thematic program of Semmelweis University.

Acknowledgments

We would like to thank István Csabai, Gergely Szöllősi, Alexandra Kalmár, Zoltán Szállási, Barbara Barták, Sára Zsigrai and Norbert Solymosi for their ideas and supports.

Conflict of interest

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

Publisher's note

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

References

1. Sidow A, Spies N. Concepts in solid tumor evolution. *Trends Genet* (2015) 31 (4):208–14. doi: 10.1016/j.tig.2015.02.001

2. Merlo LM, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. *Nat Rev Cancer* (2006) 6(12):924–35. doi: 10.1038/nrc2013

3. Trejo-Becerril C, Pérez-Cárdenas E, Taja-Chayeb L, Anker P, Herrera-Goepfert R, Medina-Velázquez LA, et al. Cancer progression mediated by horizontal gene transfer in an *in vivo* model. *PloS One* (2012) 7(12):e52754. doi: 10.1371/journal.pone.0052754

4. Kawamura Y, Yamamoto Y, Sato TA, Ochiya T. Extracellular vesicles as trans-genomic agents: Emerging roles in disease and evolution. *Cancer Sci* (2017) 108(5):824–30. doi: 10.1111/cas.13222

5. Vogan AA, Higgs PG. The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. *Biol Direct* (2011) 6:1. doi: 10.1186/1745-6150-6-1

6. Bergsmedh A, Szeles A, Henriksson M, Bratt A, Folkman MJ, Spetz AL, et al. Horizontal transfer of oncogenes by uptake of apoptotic bodies. *Proc Natl Acad Sci U S A* (2001) 98(11):6407–11. doi: 10.1073/pnas.101129998

7. Gahan PB, Stroun M. The virtosome-a novel cytosolic informative entity and intercellular messenger. *Cell Biochem Funct* (2010) 28(7):529–38. doi: 10.1002/cbf.1690

8. Aarthy R, Mani S, Velusami S, Sundarsingh S, Rajkumar T. Role of circulating cell-free dna in cancers. *Mol Diagn Ther* (2015) 19(6):339–50. doi: 10.1007/s40291-015-0167-y

9. György B, Szabó TG, Pásztói M, Pál Z, Misják P, Aradi B, et al. Membrane vesicles, current state-of-the-art: emerging role of extracellular vesicles. *Cell Mol Life Sci* (2011) 68(16):2667–88. doi: 10.1007/s00018-011-0689-3

10. Théry C, Witwer KW, Aikawa E, Alcaraz MJ, Anderson JD, Andriantsitohaina R, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the international society for extracellular vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles* (2018) 7(1):1535750. doi: 10.1080/20013078.2018.1535750

11. Whiteside TL. Tumor-derived exosomes and their role in cancer progression. *Adv Clin Chem* (2016) 74:103–41. doi: 10.1016/bs.acc.2015.12.005

12. Ma L, Li Y, Peng J, Wu D, Zhao X, Cui Y, et al. Discovery of the migrasome, an organelle mediating release of cytoplasmic contents during cell migration. *Cell Res* (2015) 25(1):24–38. doi: 10.1038/cr.2014.135

13. Valcz G, Buzás EI, Kittel Á, Krenács T, Visnovitz T, Spisák S, et al. En bloc release of MVB-like small extracellular vesicle clusters by colorectal carcinoma cells. *J Extracell Vesicles* (2019) 8(1):1596668. doi: 10.1080/20013078.2019.1596668

14. Yuana Y, Sturk A, Nieuwland R. Extracellular vesicles in physiological and pathological conditions. *Blood Rev* (2013) 27(1):31–9. doi: 10.1016/j.blre.2012.12.002

15. Kalluri R, LeBleu VS. Discovery of double-stranded genomic dna in circulating exosomes. *Cold Spring Harb Symp Quant Biol* (2016) 81:275–80. doi: 10.1101/sqb.2016.81.030932

16. Thakur BK, Zhang H, Becker A, Matei I, Huang Y, Costa-Silva B, et al. Double-stranded DNA in exosomes: a novel biomarker in cancer detection. *Cell Res* (2014) 24(6):766–9. doi: 10.1038/cr.2014.44

17. Valcz G, Galamb O, Krenács T, Spisák S, Kalmár A, Patai ÁV, et al. Exosomes in colorectal carcinoma formation: ALIX under the magnifying glass. *Mod Pathol* (2016) 29(8):928–38. doi: 10.1038/modpathol.2016.72

18. Bebelman MP, Janssen E, Pegtel DM, Crudden C. The forces driving cancer extracellular vesicle secretion. *Neoplasia* (2021) 23(1):149–57. doi: 10.1016/j.neo.2020.11.011

19. Polyak K, Haviv I, Campbell IG. Co-Evolution of tumor cells and their microenvironment. *Trends Genet* (2009) 25(1):30–8. doi: 10.1016/j.tig.2008.10.012

20. Vitale I, Shema E, Loi S, Galluzzi L. Intratumoral heterogeneity in cancer progression and response to immunotherapy. *Nat Med* (2021) 27(2):212–24. doi: 10.1038/s41591-021-01233-9

21. Gerlinger M, McGranahan N, Dewhurst SM, Burrell RA, Tomlinson I, Swanton C. Cancer: evolution within a lifetime. *Annu Rev Genet* (2014) 48:215–36. doi: 10.1146/annurev-genet-120213-092314

22. Fischer S, Cornils K, Speiseder T, Badbaran A, Reimer R, Indenbirken D, et al. Indication of horizontal dna gene transfer by extracellular vesicles. *PloS One* (2016) 11(9):e0163665. doi: 10.1371/journal.pone.0163665

23. Lim S, Quinton RJ, Ganem NJ. Nuclear envelope rupture drives genome instability in cancer. *Mol Biol Cell* (2016) 27(21):3210–13. doi: 10.1091/mbc.E16-02-0098

24. Yokoi A, Villar-Prados A, Oliphint PA, Zhang J, Song X, De Hoff P, et al. Mechanisms of nuclear content loading to exosomes. *Sci Adv* (2019) 5(11):eaax8849. doi: 10.1126/sciadv.aax8849

25. Hatch EM, Fischer AH, Deerinck TJ, Hetzer MW. Catastrophic nuclear envelope collapse in cancer cell micronuclei. *Cell* (2013) 154(1):47–60. doi: 10.1016/j.cell.2013.06.007

26. Papandreou ME, Tavernarakis N. Nucleophagy: from homeostasis to disease. *Cell Death Differ* (2019) 26(4):630–9. doi: 10.1038/s41418-018-0266-5

27. Elzanowska J, Semira C, Costa-Silva B. DNA In extracellular vesicles: biological and clinical aspects. *Mol Oncol* (2021) 15(6):1701–14. doi: 10.1002/1878-0261.12777

28. Takahashi A, Okada R, Nagao K, Kawamata Y, Hanyu A, Yoshimoto S, et al. Exosomes maintain cellular homeostasis by excreting harmful DNA from cells. *Nat Commun* (2017) 8:15287. doi: 10.1038/ncomms15287

29. Németh A, Orgovan N, Sódar BW, Osteikoetxea X, Pálóczi K, Szabó-Taylor KÉ, et al. Antibiotic-induced release of small extracellular vesicles (exosomes) with surface-associated DNA. *Sci Rep* (2017) 7(1):8202. doi: 10.1038/s41598-017-08392-1

30. Domenis R, Cifù A, Mio C, Fabris M, Curcio F. Pro-inflammatory microenvironment modulates the transfer of mutated tp53 mediated by tumor exosomes. *Int J Mol Sci* (2021) 22(12):6258. doi: 10.3390/ijms22126258

31. Lee TH, Chennakrishnaiah S, Audemard E, Montermini L, Meehan B, Rak J. Oncogenic ras-driven cancer cell vesiculation leads to emission of double-stranded DNA capable of interacting with target cells. *Biochem Biophys Res Commun* (2014) 451(2):295–301. doi: 10.1016/j.bbrc.2014.07.109

32. Sharma A, Johnson A. Exosome DNA: Critical regulator of tumor immunity and a diagnostic biomarker. *J Cell Physiol* (2020) 235(3):1921–32. doi: 10.1002/jcp.29153

33. Bernard V, Kim DU, San Lucas FA, Castillo J, Allenson K, Mulu FC, et al. Circulating nucleic acids are associated with outcomes of patients with pancreatic cancer. *Gastroenterology* (2019) 156(1):108–118.e4. doi: 10.1053/j.gastro.2018.09.022

34. Sansone P, Savini C, Kurelac I, Chang Q, Amato LB, Strillacci A, et al. Packaging and transfer of mitochondrial DNA *via* exosomes regulate escape from dormancy in hormonal therapy-resistant breast cancer. *Proc Natl Acad Sci U S A* (2017) 114(43):E9066–75. doi: 10.1073/pnas.1704862114

35. Kahlert C, Melo SA, Protopopov A, Tang J, Seth S, Koch M, et al. Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. *J Biol Chem* (2014) 289(7):3869–75. doi: 10.1074/jbc.C113.532267

36. Keserű JS, Soltész B, Lukács J, Márton É, Szilágyi-Bónizs M, Penyige A, et al. Detection of cell-free, exosomal and whole blood mitochondrial DNA copy number in plasma or whole blood of patients with serous epithelial ovarian cancer. *J Biotechnol* (2019) 298:76–81. doi: 10.1016/j.jbiotec.2019.04.015

37. Jeppesen DK, Fenix AM, Franklin JL, Higginbotham JN, Zhang Q, Zimmerman LJ, et al. Reassessment of exosome composition. *Cell* (2019) 177(2):428–445.e18. doi: 10.1016/j.cell.2019.02.029

38. Zhang H, Freitas D, Kim HS, Fabijanic K, Li Z, Chen H, et al. Identification of distinct nanoparticles and subsets of extracellular vesicles by asymmetric flow field-flow fractionation. *Nat Cell Biol* (2018) 20:332–43. doi: 10.1038/s41556-018-0040-4

39. Cai J, Han Y, Ren H, Chen C, He D, Zhou L, et al. Extracellular vesicle-mediated transfer of donor genomic DNA to recipient cells is a novel mechanism for genetic influence between cells. *J Mol Cell Biol* (2013) 5:227–38. doi: 10.1093/jmcb/mjt011

40. Lázaro-Ibáñez E, Lässer C, Shelke GV, Crescitelli R, Jang SC, Cvjetkovic A, et al. DNA Analysis of low- and high-density fractions defines heterogeneous subpopulations of small extracellular vesicles based on their DNA cargo and topology. *J Extracell Vesicles* (2019) 8(1):1656993. doi: 10.1080/20013078.2019.1656993

41. Archetti M, Pienta KJ. Cooperation among cancer cells: applying game theory to cancer. *Nat Rev Cancer* (2019) 19(2):110–17. doi: 10.1038/s41568-018-0083-7

42. Tabassum DP, Polyak K. Tumorigenesis: it takes a village. *Nat Rev Cancer* (2015) 15(8):473–83. doi: 10.1038/nrc3971

43. Burrell RA, Swanton C. Re-evaluating clonal dominance in cancer evolution. *Trends Cancer* (2016) 2(5):263–76. doi: 10.1016/j.trecan.2016.04.002

44. McKelvey KJ, Powell K, Ashton AW, Morris JM, McCracken SA. Exosomes: Mechanisms of uptake. *J Circ biomark* (2015) 4:7. doi: 10.5772/61186

45. Lee TH, Chennakrishnaiah S, Meehan B, Montermini L, Garnier D, D'Asti E, et al. Barriers to horizontal cell transformation by extracellular vesicles containing oncogenic h-ras. *Oncotarget* (2016) 7(32):51991–2002. doi: 10.18632/oncotarget.10627

46. Hoshino A, Costa-Silva B, Shen TL, Rodrigues G, Hashimoto A, Tesic Mark M, et al. Tumour exosome integrins determine organotropic metastasis. *Nature* (2015) 527(7578):329–35. doi: 10.1038/nature15756

47. Rana S, Yue S, Stadel D, Zöller M. Toward tailored exosomes: the exosomal tetraspanin web contributes to target cell selection. *Int J Biochem Cell Biol* (2012) 44(9):1574–84. doi: 10.1016/j.biocel.2012.06.018

48. Abdouh M, Floris M, Gao ZH, Arena V, Arena M, Arena GO. Colorectal cancer-derived extracellular vesicles induce transformation of fibroblasts into colon carcinoma cells. *J Exp Clin Cancer Res* (2019) 38(1):257. doi: 10.1186/s13046-019-1248-2

49. Kitai Y, Kawasaki T, Sueyoshi T, Kobiyama K, Ishii KJ, Zou J, et al. DNA-Containing exosomes derived from cancer cells treated with topotecan activate a sting-dependent pathway and reinforce antitumor immunity. *J Immunol* (2017) 198(4):1649–59. doi: 10.4049/jimmunol.1601694

50. Diamond JM, Vanpouille-Box C, Spada S, Rudqvist NP, Chapman JR, Ueberheide BM, et al. Exosomes shuttle trex1-sensitive ifn-stimulatory dsdna from irradiated cancer cells to dcs. *Cancer Immunol Res* (2018) 6(8):910–20. doi: 10.1158/2326-6066.CIR-17-0581

51. Rappa G, Santos MF, Green TM, Karbanová J, Hassler J, Bai Y, et al. Nuclear transport of cancer extracellular vesicle-derived biomaterials through nuclear envelope invagination-associated late endosomes. *Oncotarget* (2017) 8(9):14443–61. doi: 10.18632/oncotarget.14804

52. Husnik F, McCutcheon JP. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Microbiol* (2018) 16(2):67–79. doi: 10.1038/nrmicro.2017.137

53. Schönknecht G, Weber AP, Lercher MJ. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *Bioessays* (2014) 36(1):9–20. doi: 10.1002/bies.201300095

54. Ono R, Yasuhiko Y, Aisaki KI, Kitajima S, Kanno J, Hirabayashi Y. Exosome-mediated horizontal gene transfer occurs in double-strand break repair during genome editing. *Commun Biol* (2019) 2:57. doi: 10.1038/s42003-019-0300-2

55. Ihry RJ, Worringer KA, Salick MR, Frias E, Ho D, Theriault K. Et al: p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat Med* (2018) 24(7):939–46. doi: 10.1038/s41591-018-0050-6

56. Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med* (2018) 24(7):927–30. doi: 10.1038/s41591-018-0049-z

57. Rodgers K, McVey M. Error-prone repair of dna double-strand breaks. *J Cell Physiol* (2016) 231(1):15–24. doi: 10.1002/jcp.25053

58. Gatz SA, Wiesmüller L. p53 in recombination and repair. *Cell Death Differ* (2006) 13(6):1003–16. doi: 10.1038/sj.cdd.4401903

59. Coupier I, Baldeyron C, Rousseau A, Mosseri V, Pages-Berhouet S, Caux-Moncoutier V, et al. Fidelity of DNA double-strand break repair in heterozygous cell lines harbouring BRCA1 missense mutations. *Oncogene* (2004) 23(4):914–9. doi: 10.1038/sj.onc.1207191

60. Bau DT, Mau YC, Shen CY. The role of BRCA1 in non-homologous end-joining. *Cancer Lett* (2006) 240(1):1–8. doi: 10.1016/j.canlet.2005.08.003

61. Kalluri R. The biology and function of exosomes in cancer. *J Clin Invest* (2016) 126(4):1208–15. doi: 10.1172/JCI81135

62. Gatenby RA, Gillies RJ. A microenvironmental model of carcinogenesis. *Nat Rev Cancer* (2008) 8(1):56–61. doi: 10.1038/nrc2255

63. Greaves M, Maley CC. Clonal evolution in cancer. *Nature* (2019) 481 (7381):306–13. doi: 10.1038/nature10762

64. McFarland CD, Korolev KS, Kryukov GV, Sunyaev SR, Mirny LA. Impact of deleterious passenger mutations on cancer progression. *Proc Natl Acad Sci U S A* (2013) 110:2910–5. doi: 10.1073/pnas.1213968110

65. Kumar S, Warrell J, Li S, McGillivray PD, Meyerson W, Salichos L, et al. Passenger mutations in more than 2,500 cancer genomes: overall molecular functional impact and consequences. *Cell* (2020) 180(5):915–927.e16. doi: 10.1016/j.cell.2020.01.032

66. Rodriguez-Valera F, Martin-Cuadrado AB, López-Pérez M. Flexible genomic islands as drivers of genome evolution. *Curr Opin Microbiol* (2016) 31:154–60. doi: 10.1016/j.mib.2016.03.014

67. Chu HY, Sprouffske K, Wagner A. Assessing the benefits of horizontal gene transfer by laboratory evolution and genome sequencing. *BMC Evol Biol* (2018) 18:54. doi: 10.1186/s12862-018-1164-7

68. Rebbeck CA, Thomas R, Breen M, Leroi AM, Burt A. Origins and evolution of a transmissible cancer. *Evolution* (2009) 63(9):2340–9. doi: 10.1111/j.1558-5646.2009.00724.x

69. Strakova A, Murchison EP. The cancer which survived: insights from the genome of an 11000 year-old cancer. *Curr Opin Genet Dev* (2015) 30:49–55. doi: 10.1016/j.gde.2015.03.005

70. Ujvari B, Gatenby RA, Thomas F. The evolutionary ecology of transmissible cancers. *Infect Genet Evol* (2016) 39:293–303. doi: 10.1016/j.meegid.2016.02.005

71. Griessinger E, Moschoi R, Biondani G, Peyron JF. Mitochondrial transfer in the leukemia microenvironment. *Trends Cancer* (2017) 3(12):828–39. doi: 10.1016/j.trecan.2017.10.003

72. Berridge MV, Dong L, Neuzil J. Mitochondrial dna in tumor initiation, progression, and metastasis: role of horizontal mtdna transfer. *Cancer Res* (2015) 75(16):3203–8. doi: 10.1158/0008-5472.CAN-15-0859

73. Strakova A, Murchison EP. The changing global distribution and prevalence of canine transmissible venereal tumour. *BMC Vet Res* (2014) 10:168. doi: 10.1186/s12917-014-0168-9

74. Murchison EP, Wedge DC, Alexandrov LB, Fu B, Martincorena I, Ning Z, et al. Transmissible dog cancer genome reveals the origin and history of an ancient cell lineage. *Science* (2014) 343(6169):437–40. doi: 10.1126/science.1247167

75. Ostrander EA, Davis BW, Ostrander GK. Transmissible tumors: Breaking the cancer paradigm. *Trends Genet* (2016) 32(1):1–15. doi: 10.1016/j.tig.2015.10.001

76. Decker B, Davis BW, Rimbault M, Long AH, Karlins E, Jagannathan V, et al. Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor. *Genome Res* (2015) 25(11):1646–55. doi: 10.1101/gr.190314.115

77. Baez-Ortega A, Gori K, Strakova A, Allen JL, Allum KM, Bansse-Issa L, et al. Somatic evolution and global expansion of an ancient transmissible cancer lineage. *Science* (2019) 365(6452):eaau9923. doi: 10.1126/science.aau9923

78. Strakova A, Ní Leathlobhair M, Wang GD, Yin TT, Airikkala-Otter I, Allen JL, et al. Mitochondrial genetic diversity, selection and recombination in a canine transmissible cancer. *Elife* (2016) 5:e14552. doi: 10.7554/eLife.14552

79. Yonemitsu MA, Giersch RM, Polo-Prieto M, Hammel M, Simon A, Cremonte F, et al. A single clonal lineage of transmissible cancer identified in two marine mussel species in south America and Europe. *Elife* (2019) 8:e47788. doi: 10.7554/eLife.47788

80. Zhang L, Yu D. Exosomes in cancer development, metastasis, and immunity. *Biochim Biophys Acta Rev Cancer* (2019) 1871(2):455–68. doi: 10.1016/j.bbcan.2019.04.004

81. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* (2017) 168(4):613–28. doi: 10.1016/j.cell.2017.01.018

82. Frankenstein Z, Basanta D, Franco OE, Gao Y, Javier RA, Strand DW, et al. Stromal reactivity differentially drives tumour cell evolution and prostate cancer progression. *Nat Ecol Evol* (2020) 4(6):870–84. doi: 10.1038/s41559-020-1157-y

83. Du H, Che G. Genetic alterations and epigenetic alterations of cancer-associated fibroblasts. *Oncol Lett* (2017) 13(1):3–12. doi: 10.3892/ol.2016.5451

84. Gaiffe E, Prétet JL, Launay S, Jacquin E, Saunier M, Hetzel G, et al. Apoptotic HPV positive cancer cells exhibit transforming properties. *PloS One* (2012) 7(5): e36766. doi: 10.1371/journal.pone.0036766

85. Valcz G, Buzás EI, Sebestyén A, Krenács T, Szállási Z, Igaz P, et al. Extracellular vesicle-based communication may contribute to the co-evolution of cancer stem cells and cancer-associated fibroblasts in anti-cancer therapy. *Cancers* (2020) 12(8):2324. doi: 10.3390/cancers12082324

86. Bendich A, Wilczok T, Borenfreund E. Circulating DNA as a possible factor in oncogenesis. *Science* (1965) 148(3668):374–6. doi: 10.1126/science.148.3668.374

87. Liu Y. Darwin's pangenesis and medical genetics. *Adv Genet* (2018) 102:67–92. doi: 10.1016/bs.adgen.2018.05.008

88. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer* (2009) 9 (4):265–73. doi: 10.1038/nrc2620

89. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* (2011) 144(5):646–74. doi: 10.1016/j.cell.2011.02.013

90. Paget S. The distribution of secondary growths in cancer of the breast. *Cancer Metastasis Rev* (1989) 8(2):98–101. doi: 10.1016/S0140-6736(00)49915-0

frontiers | Frontiers in Oncology

TYPE Case Report
PUBLISHED 11 January 2023
DOI 10.3389/fonc.2022.1050452

OPEN ACCESS

EDITED BY
Robert Wesolowski,
Comprehensive Cancer Center, The Ohio State University, United States

REVIEWED BY
Clotilde Costa,
Sanitary Research Institute Foundation of Santiago de Compostela, Spain
Daniel Stover,
The Ohio State University, United States

*CORRESPONDENCE
Patrizio Giacomini
patrizio.giacomini@ifo.it

SPECIALTY SECTION
This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology

RECEIVED 21 September 2022
ACCEPTED 19 December 2022
PUBLISHED 11 January 2023

CITATION
Allegretti M, Barberi V, Ercolani C, Vidiri A, Giordani E, Ciliberto G, Giacomini P and Fabi A (2023) Unusual phylogenetic tree and circulating actionable ESR1 mutations in an aggressive luminal/HER2-low breast cancer: Case report.
Front. Oncol. 12:1050452.
doi: 10.3389/fonc.2022.1050452

Unusual phylogenetic tree and circulating actionable ESR1 mutations in an aggressive luminal/HER2-low breast cancer: Case report

Matteo Allegretti[1], Vittoria Barberi[2], Cristiana Ercolani[3], Antonello Vidiri[4], Elena Giordani[1], Gennaro Ciliberto[5], Patrizio Giacomini[6*] and Alessandra Fabi[7]

[1]Translational Oncology Research, IRCSS Regina Elena National Cancer Institute, Rome, Italy, [2]Medical Oncology 1, IRCSS Regina Elena National Cancer Institute, Rome, Italy, [3]Pathology, IRCSS Regina Elena National Cancer Institute, Rome, Italy, [4]Radiology and Diagnostic Imaging, IRCSS Regina Elena National Cancer Institute, Rome, Italy, [5]Scientific Directorate, IRCSS Regina Elena National Cancer Institute, Rome, Italy, [6]Clinical Trial Center, IRCSS Regina Elena National Cancer Institute, Rome, Italy, [7]Precision Medicine in Senology, Scientific Directorate - Department of Women and Child Health, Fondazione Policlinico Universitario "A. Gemelli" IRCCS, Rome, Italy

Under therapeutic pressure aggressive tumors evolve rapidly. Herein, a luminal B/HER2-low breast cancer was tracked for >3 years during a total of 6 largely unsuccessful therapy lines, from adjuvant to advanced settings. Targeted next generation sequencing (NGS) of the primary lesion, two metastases and 14 blood drawings suggested a striking, unprecedented coexistence of three evolution modes: punctuated, branched and convergent. Punctuated evolution of the trunk was supported by *en bloc* inheritance of a large set (19 distinct genes) of copy number alterations. Branched evolution was supported by the distribution of site-specific SNVs. Convergent evolution was characterized by a unique asynchronous expansion of three actionable (OncoKB level 3A) mutations at two consecutive ESR1 codons. Low or undetectable in all the sampled tumor tissues, ESR1 mutations expanded rapidly in blood during HER2/hormone double-blockade, and predicted life-threatening local progression at lung and liver metastatic foci. Dramatic clinical response to Fulvestrant (assigned off-label exclusively based on liquid biopsy) was associated with clearance of all 3 subclones and was in stark contrast to the poor therapeutic efficacy reported in large liquid biopsy-informed interventional trials. Altogether, deconvolution of the tumor phylogenetic tree, as shown herein, may help to customize treatment in breast cancers that rapidly develop refractoriness to multiple drugs.

KEYWORDS

breast cancer, cancer evolution, liquid biopsy, ESR1, HER2, molecular tumor board

1 Introduction

Hormone-receptor (HR) positive breast cancer includes the luminal A and luminal B subtypes and is the most common type of tumor in women diagnosed with early-stage breast cancer. Although its overall prognosis is better than HER2 + and triple-negative breast cancers, individual risk of relapse may differ widely (1, 2).

The mainstay of targeted treatment for luminal breast cancer is endocrine therapy (ET), which includes several agents that either directly target estrogen receptor (ER) or suppress estrogen production. Standard adjuvant treatment in low-risk breast cancer involves Tamoxifen in pre-menopausal patients, and aromatase inhibitors (Ai) in post-menopausal patients. Association of luteinizing hormone-releasing hormone inhibitors is recommended in pre-menopausal patients. Cyclin-dependent kinase inhibitors (CDK4/6i), combined with ET, are recognized as the standard of care for adjuvant treatment in high-risk patients, and in the metastatic setting (1, 2). Trastuzumab Deruxtecan (T-DXd), recently approved for $HER2^+$ advanced tumors, is expected to have a strong clinical impact on HR^+/HER2-low luminal breast cancers (3).

Despite expanding therapeutic options, some invasive luminal breast cancers may rapidly develop resistance to pharmacological treatment. In many tumors, this has been associated with biological aggressiveness and complex sub-clonal dynamics (4, 5). It is becoming increasingly clear that real-time tracking by longitudinal tumor tissue sampling and liquid biopsy (LB) would be useful to instruct clinical decisions (5) but, with notable exceptions (6), phylogenetic trees are mostly reconstructed by genome-wide *post-mortem* analysis of tumor DNA from multiple metastatic foci.

Herein, we describe an aggressive breast carcinoma characterized by the co-occurrence of an extremely fast clinical progression and an unusual molecular evolution. Whereas its luminal B/HER2-low subtype features were conventional and poorly informative, targeted next generation sequencing (NGS) of tissue and liquid biopsies suggested an unprecedented combination of punctuated, branched, and convergent evolution patterns (4). Clinical NGS deciphered tumor phylogenesis and convergent adaptive selection, enabling successful assignment of off-label treatment.

2 Case description

The clinical timeline, treatments and main diagnostic assessments in this patient are summarized in Figure 1A. In July 2017, a 55-year-old woman underwent quadrantectomy and axillary lymphadenectomy for a multifocal (G3) ductal infiltrating breast carcinoma of the luminal B subtype: estrogen receptor (ER) 90%, progesterone receptor (PgR) 5%, Ki67 50%, HER2-low (immunohistochemistry 2+; non-amplified by *in situ* hybridization). In December 2017, one month only after the end of adjuvant therapy, early metastatic dissemination to the lung, bones and parasternal soft tissue called for first-line therapy with CDK4/6i plus Letrozole. Unfortunately, further rapid progression occurred (July 2018) involving the lung and spine (L2/L3), which prompted to surgical removal of a thoracic subcutaneous metastasis for diagnostic re-assessment and targeted NGS (see below). Routine immunohistochemistry of the metastatic tissue confirmed the $ER^+/PgR^+/HER2\ 2^+$ status detected in the primary, and in addition revealed a very slight HER2 copy number gain (HER2/CEP17 SISH ratio: 2.4). Since Trastuzumab deruxtecan (3) was not available in year 2018, a combined regimen was attempted of Trastuzumab plus Pertuzumab and weekly Taxol for double HER2/hormone blockade (2^{nd} line). However, novel pleural and left costal lesions developed after 5 cycles. This prompted us to initiate 3^{rd} line therapy with Trastuzumab emtansine (T-DM1) plus Letrozole. This combination was administered in the context of a LB study, now published (7). LB detected progressive accumulation of 3 actionable ESR1 circulating alterations (see below) during treatment, despite apparently stable disease. Clinical progression (massive dissemination to the lung, pleura and liver) occurred in June 2019 (total body CT scan, Figure 1B). The patient was referred to the Regina Elena intramural Molecular Tumor Board (MTB) in very poor general conditions and dyspnoic. Off-label Fulvestrant was recommended (see below) exclusively based on LB data, in association with Capecitabine (4^{th} line), but the latter had to be discontinued after a single cycle due to severe gastrointestinal intolerance. In contrast, Fulvestrant was well tolerated and, despite administration as single-agent, it resulted in a quick and dramatic improvement with disappearance of most lung lesions and associated pleural exudate, as well as drastic regression of the extensive liver invasion (Figure 1C). Response lasted 8 months until early February 2020, when the patient, who was otherwise in good general conditions and essentially asymptomatic, suddenly experienced dyspnea and chest pain due to a pleural exudate requiring thoracentesis. Salvage treatment with Trastuzumab plus Vinorelbine (5^{th} line) resulted in partial response of the pleural effusion and apparently stable thoracic disease for about 7 additional months, until massive pleural/lung recurrence and death in November 2020.

3 Diagnostic assessment and molecular profiling

Tumor DNAs (tDNAs) were obtained from the three sampled tumor lesions: the primary tumor, the subcutaneous metastasis, and neoplastic cells isolated from the pleural effusion following overnight adherence to plastic dishes. Circulating tumor DNAs (ctDNAs) were obtained from 14 blood

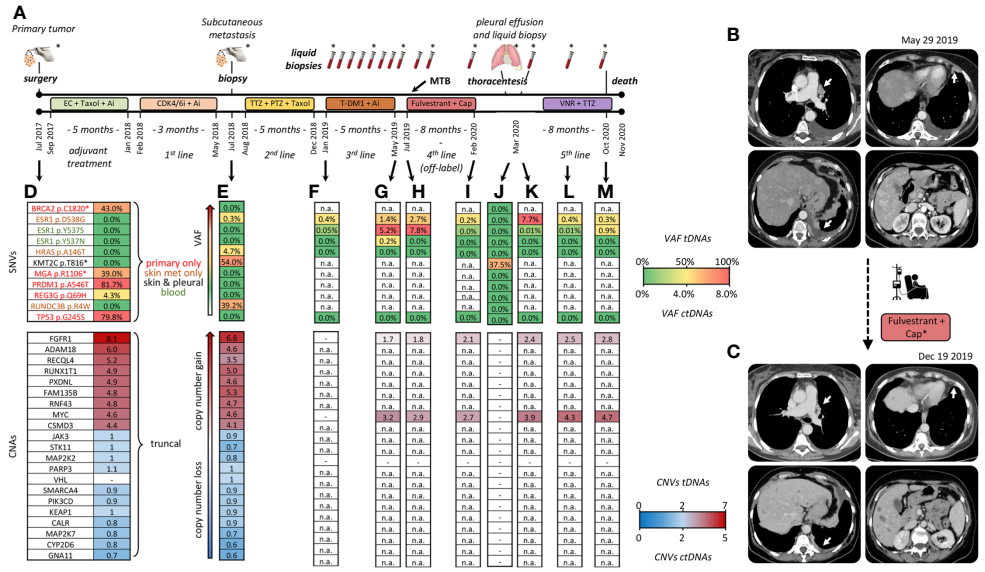


FIGURE 1
Clinical timeline and tumor profiling. **(A)** Treatment history: adjuvant chemotherapy with Epirubicine plus Cyclofosphamide (EC: 90/600 mg/mq, q14), followed by Taxol (175 mg/mq, q14) plus Pegfilgrastim, and Aromatase Inhibitor (Ai: Letrozole 2.5 mg/day). First line therapy with Palbociclib (CDK/6i:125 mg/die for 21 days, every 28 days) plus Aromatase Inhibitor (Ai: Letrozole 2.5 mg/day). Surgical removal of a thoracic skin metastasis. Second line with Trastuzumab (TTZ: 8 mg dose loading → 6 mg) plus Pertuzumab (PTZ: 840 mg dose loading → 420 mg) and weekly Taxol. Third line with T-DM1 (3.6 mg/kg, every 21 days) and Letrozole (2.5 mg/day), beginning of LB monitoring. Fourth line, off-label Fulvestrant (500 mg day 1, 15, 28, then every 28 days) in association with Capecitabine (Cap: 2000 mg/m^2 daily discontinued after a single cycle). Fifth line with Vinorelbine (VNR: 25 mg/mq at day 1 and 8) plus Trastuzumab (8 mg/kg -> 6 mg/kg) every 21 days. Asterisks: tissue and blood samples tested by both NGS and dPCR. All the other samples: testing by dPCR only. **(B, C)** CT scans before and following treatment with Fulvestrant; dates of clinical imaging are noted. Major sites of response to Fulvestrant are indicated by white arrows. Cap*, Capecitabine **(D–M)** Key time points at which tumor tissue/cells and blood drawings were obtained. Single nucleotide variants (SNV) are color-coded by tissue of origin in **(D)**. Palettes rank alterations by variant allele frequency (different %VAFs) and copy numbers (CNA). Copy numbers are expressed per diploid genome: copy number-neutral n=2. tDNA, tumor DNA; ctDNA, circulating tumor DNA. (-) undetectable. n.a., not assessed because SNV/CNA was not included in the LB-grade NGS panel.

drawings. All 3 tDNAs and 8/14 ctDNAs (Figure 1A, asterisks) were tested by NGS (Life Technologies, see Supplementary for technical details) with two targeted panels of different complexity. The tDNA-grade Oncomine™ Comprehensive Assay Plus detected 9 single nucleotide variants (SNVs) and 19 copy number alterations (CNA) in a total of 505 tested genes. These 27 alterations are listed along with variant allele frequencies (VAFs) and estimated copy numbers in Figures 1D, E, J. Due to its smaller size (it detects SNVs, CNAs and fusions in 52, 12 and 92 genes, respectively), the ctDNA-grade Oncomine™ PanCancer Cell-Free Assay could only assess 3 of 9 SNVs and 2 of 19 CNAs. Results are shown for 7 representative blood drawings (Figures 1F–M; n.a.: not assessed).

The distribution of the 9 SNVs among the three tDNAs readily revealed marked phylogenetic divergence during tumor evolution: 5 SNVs (BRCA2 p.C1820, MGA p.R1106*, PRDM1 p.A546T, REG3G p.Q69H and TP53 p.G245S) were exclusive of the primary, 3 (HRAS p.A146T, RUNDC3B p.R4W and ESR1 p.D538G) were exclusive of the subcutaneous metastasis, and only one (KMT2C p.T816) was shared between any two lesions, namely between the subcutaneous metastasis and the neoplastic pleural effusion (Figures 1D, E, J, top section; SNVs color-coded by site of detection in panel d).

Particularly in light of rapid clinical progression, BRCA p.C1820 and TP53 p.G245S losses appeared counterintuitive. The former is a potentially inheritable, likely pathogenic variant (https://varsome.com), and the latter is a loss-of-function cancer driver strongly associated with poor outcome (OncoKB level Px1; https://www.oncokb.org). These considerations prompted to orthogonal testing by custom-designed digital PCR (dPCR) assays. However, NGS calling was confirmed, and dPCR-assessed VAFs differed marginally (within +/- 3%) from NGS-assessed VAFs (see representative TP53 testing in Table S1). Likewise, dPCR confirmed that BRCA2 p.C1820 was undetectable at both tested metastatic sites and in leukocyte DNA (not shown), conclusively ruling out a germ line origin.

Next, we focused on the similarly puzzling observation of a simultaneous presence of three gain-of-function ESR1 SNVs (p.D538G, p.Y537S and p.Y537N), all known to be actionable and associated with resistance to hormone blockade (8). Interestingly, these three SNVs displayed very different origins and accumulation kinetics, as may be seen by comparing tDNA from the skin metastasis, and ctDNAs obtained before the T-

DM1/Letrozole and Fulvestrant treatments, at progression, and at later time points (Figures 1E–M). Specifically, ESR1 p.D538G was the only ESR1 alteration to be detected in tissues. It displayed a very low VAF at onset both in tDNA and ctDNA, and then marked blood increases mostly coincident with clinical progression. In contrast, ESR1 p.Y537S and p.Y537N, initially undetectable, appeared *de novo* in blood, and their association with clinical outcome appeared to be less stringent. None of the three SNVs could be detected in the pleural effusion (Figure 1J). Also these changes were confirmed by orthogonal dPCR validation (representative results shown in Table S1). In summary, BRCA2 p.C1820 and TP53 p.G245S were negatively selected, whereas ESR1 alterations were positively selected by successive events taking place for the most part at undetermined tumor sites.

Strikingly in contrast to the heterogeneous SNV distribution, the primary site and the skin lesion did share a large set of 19 collinear CNAs (Figure 1D, E, bottom section). Most CNAs appeared to be inherited *en bloc*, since copy number gains and gene-to-gene ratios were roughly conserved between the two lesions. Only FAM135B displayed a discordant trend (increase instead of decrease/stability in copy numbers). Unfortunately, not all CNAs were testable by the ctDNA-grade NGS panel. Despite this limitation, we were able to confirm that both FGFR1 and MYC (the only two CNAs testable in blood) were amplified and, similar to tissues, copy number gains persisted in serial blood drawings, undergoing limited absolute and relative variations (Figures 1F–L, bottom). Of special interest, FGFR1 and MYC amplifications were detected in peripheral blood collected at the time of thoracentesis despite tDNA from the pleural neoplastic cells had lost FGFR1, MYC, and all the other CNAs (Figure 1J vs K). The easiest interpretation of these results is that a clonal expansion containing at least FGFR1 and MYC CNAs was present throughout disease course, e.g. in the primary tumor, skin metastatic site and other unidentified tumor sites, as indirectly revealed by LB, but it was lost in the last lesion that had been sampled, e.g. tumor cells from the pleural cavity.

To display multiple SNV and CNA trajectories, longitudinal series of %VAF values and DNA copy numbers were elaborated by the fishplot package for RStudio (9). The 6 ctDNAs not tested by NGS were included by incorporating dPCR data in the model. To portray tumor evolution in its entirety we chose to represent SNVs and CNAs in the same graph, and generated separate fishplots for blood and tumor lesions (Figures 2A, B). For simplified CNA rendering, we focussed on FGFR1 and MYC copy numbers, that are available for both blood and tissues. These were averaged and merged into a single fishplot. However, since %VAF and gene copy numbers are incommensurable, the latter were displayed in background and not to scale. It is acknowledged that no inference can be made about relative SNV/CNA subclonal representation, whereas quantitative relationships among multiple SNVs are faithfully graphed.

Specifically, the fishplot in Figure 2A clearly shows relative abundance and kinetics of ESR1 p.D538G p.Y537S and p.Y537N. Quite strikingly, the two most abundant variants displayed a roughly reciprocal relationship whereby expansion of either is mirrored by contraction of the other. Unsurprisingly, circulating ESR1 alterations were prodromic to life-threatening clinical progression (see case description), which led to seek MTB multidisciplinary advice.

In elaborating therapeutic recommendations, the MTB considered the following points: (a) the 3 actionable (OncoKB

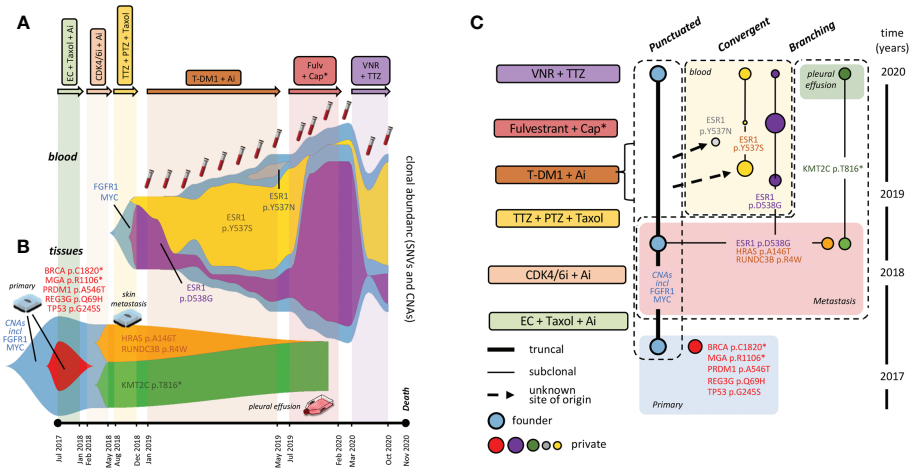


FIGURE 2
Fishplots and tumor phylogenesis. Fishplots displaying alterations in blood **(A)** and tumor lesions **(B)**. Areas are proportional to relative SNV abundances, whereas CNAs are depicted in background, not to scale. **(C)** Tumor phylogenetic tree: boxes surround evolution modes. Connecting lines: parental relationship proven (continuous) and inferred (dotted). ESR1 mutations: dimension of dots (blood variants) is roughly proportional to clonal representation. Tumor clones are consistently color-coded between fishplot areas and the dots of the phylogenetic tree. Abbreviations as in Figure 1.

level 3A) ESR1 SNVs had disabled two consecutive codons in the catalytic ESR1 site; (b) they had occurred in rapid succession within 7 months from the first T-DM1/Letrozole administration; (c) their distinct kinetics and tissue/blood distribution were suggestive of convergent adaptive events in distinct founders; (d) their timing of appearance and subsequent clinical progression at lung and liver foci was compatible with an elective topographical distribution of at least some ESR1 variants at these sites.

Preliminary evidence was also considered from the plasmaMATCH trial, now published (10), that subclonal ESR1 alterations when detected in blood do not confer sensitivity to targeted treatment. Nevertheless, in light of arguments (a-d), and in the absence of alternative therapeutic options, off-label Fulvestrant was recommended exclusively based on LB data, and in association with Capecitabine.

As outlined in case description, only Fulvestrant was tolerated, and dramatic clinical improvement (CT scans in Figure 1C, performed in November) was mirrored, in December 2019, by nearly complete blood clearance of the newly arisen ESR1 p.Y537S and p.Y537N variants, although the ESR1 p.D538G clone (chronologically the first to appear) re-expanded, and FGFR1/MYC CNAs kept expanding steadily in blood (Figure 2A). These results formally prove direct variant (particularly ESR1 p.Y537S) targeting by Fulvestrant, and provide evidence for a dissociated response, both molecular and clinical.

At the time of further pleural relapse in early February 2020, the ESR1 p.D538G variant and the FGFR1 and MYC copy number gains were all abundant in blood (Figure 2A). However, none was detectable in tumor cells obtained on the same day from the neoplastic pleural effusion, that only retained a single 'undruggable' KMT2C SNV (Figure 2B) first seen in the 2018 cutaneous metastasis. While further supporting clearance of ESR1 variants from the pleural space, these results prevented target FGFR1 treatment, and prompted instead to Trastuzumab plus Vinorelbine as a possible salvage therapy. Partial response of the pleural effusion was accompanied by decreases in FGFR1, MYC and ESR1 p.D538G variants, and (re)-expansion of ESR1 p.Y537S (Figure 2A). Subsequent increases of all variants in October 2020 (Figure 2A, last blood drawing) heralded massive pleural/lung recurrence and death in November 2020.

For synoptic view, the CNA trunk and the 11 SNVs detected in tDNAs and ctDNAs were assigned to specific tumor lesions and blood drawings by applying the pigeonhole principle and logical deduction. The inferred tumor phylogenetic tree is shown in Figure 2C.

4 Discussion

Herein, we describe an aggressive HER2-low breast cancer that, based on straightforward immunohistochemical and molecular subtyping, apparently meets the 'standard' definition of a luminal B tumor subtype. However, extensive NGS profiling clearly shows that its hidden hallmark is the unprecedented (to our knowledge) co-existence of punctuated, branched, and convergent evolution (boxed in Figure 2C).

Punctuated evolution is supported by persistence of a large set of collinear CNAs (particularly the FGFR1/MYC amplifications) across the primary, the skin metastasis, and possibly other tumor sites (as indirectly supported by LB). *En bloc* inheritance of multiple alterations in the tumor trunk is a hallmark of punctuated evolution (5).

Convergent evolution of the three ESR1 subclones is supported by their distinct origin and timing of appearance (e.g. one from the skin metastasis and two from unknown metastatic sites), and their striking convergence toward the same phenotypic effect (resistance to hormone blockade).

Finally, branched evolution is best exemplified by the localization of the 'undruggable' KMT2C variant (likely originated from the skin lesion) to the second neoplastic pleural effusion in the absence of all other alterations, including actionable FGFR1 amplification and ESR1 mutations. This is compatible with a dramatic clonal sweep. This event had implications on therapeutic escape, as it left no residual targeting options.

Single-cell omics of tumor tissues and circulating tumor cells have shown that each breast cancer is a collection of multiple molecular subtypes, and may be viewed as a dynamically evolving, adaptively selected cell ecotype (11, 12). It is tempting to speculate that multi-modal evolution, detected herein, is a powerful shortcut to ecotype plasticity and tumor aggressiveness.

The closest precedent of the present study is possibly a case report in which treatment-induced regression of actionable genomic alterations at specific metastatic sites correlated with changes in the corresponding ctDNAs (13). However, in this previous report the tumor architecture was reconstructed *post-mortem*, whereas in the present study the entire deduction process was in real-time and based on *ex vivo* testing. Consequently, NGS/dPCR results could be exploited for therapeutic purposes.

Response of ESR1-mutated variants (OncoKB level 3A when detected in tissues) to Fulvestrant deserves a comment in light of poor success (0/27 objective responses when subclonal ESR1 alterations were detected in blood) of the PlasmaMATCH trial (10). At least four arguments support the idea that the single patient described herein is an exceptional responder: (a) an unusual oncogenic dependence from ESR1, supported by convergent evolution of three variants; (b) selective metastatic seeding at an anatomical site (lung) electively responsible for life-threatening progression in this specific patient; (c) reciprocal expansion/contraction of ESR1 subclones during therapy, e.g. clonal competition for the same ecotype niche, a condition suggested, but not proven, to result in self-limited tumor

spreading (4); and (d) Fulvestrant-mediated simplification of the tumor phylogenetic tree, that may have helped to also achieve partial response to subsequent salvage therapy of the undruggable KT2MC-associated pleural variant.

One may object that the benefit of single-agent treatment with Fulvestrant, although impressive, was rather short-lived. Based on the I-PREDICT drug combination trial (14) we hypothesize that a double FGFR1/ESR1 blockade acting on both the trunk and the first lung/liver tumor branch(es) might have been better to counteract multiple evolution modes, resulting in a more durable response. At any rate, the patient survived 18 months after progression from the last conventional therapeutic option, which is remarkable in light of the short overall survival time (about 40 months) from diagnosis to death.

Tumors with no special or notable features except inexplicable clinical aggressiveness may not be unusual in the clinical practice. Our study suggests that deconvoluting the tumor phylogenetic tree by clinical-grade NGS may suffice to customize treatment in breast cancers that rapidly develop refractoriness to multiple drugs.

5 Patient perspective

The patient was a business manager with key professional responsibilities in her company. During the entire clinical course she was fully aware of her clinical conditions, and specifically asked the medical oncologist to receive all the necessary treatments (including off-label Fulvestrant), as long as she could stay sufficiently fit to attend her duties.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

The studies involving human participants were reviewed and approved by Fondazione Ugo Bietti, *via* Elio Chianesi 53, Rome, Italy. The patients/participants provided their written informed consent to participate in this study.

Author contributions

VB recruited and monitored the patient, MA performed and analysed NGS and dPCR. CE and EG processed tissue and plasma specimens. AV reviewed CT scans. GC supervised and coordinated MTB discussions. MA, PG, and AF planned study activities and wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

Supported by Associazione Italiana per la Ricerca sul Cancro (AIRC, Nuvenia Fellowship id. 19503 to MA), Lazio Innova ERBB2-2D (id. A0375-2020-36630), and IRE Scientific Direction intramural funding.

Acknowledgments

Our Research Nurse Dr. Emanuela Taraborelli is gratefully acknowledged for excellent assistance.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.1050452/full#supplementary-material

References

1. Waks AG, Winer EP. Breast cancer treatment: A review. *JAMA* (2019) 321 (3):288–300. doi: 10.1001/jama.2018.19323

2. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. *Lancet* (2021) 397(10286):1750–69. doi: 10.1016/S0140-6736(20)32381-3

3. Modi S, Jacot W, Yamashita T, Sohn J, Vidal M, Tokunaga E, et al. Trastuzumab deruxtecan in previously treated Her2-low advanced breast cancer. *N Engl J Med* (2022) 387:9–20. doi: 10.1056/NEJMoa2203690

4. Venkatesan S, Swanton C. Tumor evolutionary principles: How intratumor heterogeneity influences cancer treatment and outcome. *Am Soc Clin Oncol Educ book Am Soc Clin Oncol Annu Meeting* (2016) 35:e141–9. doi: 10.1200/edbk_158930

5. Turajlic S, Sottoriva A, Graham T, Swanton C. Resolving genetic heterogeneity in cancer. *Nat Rev Genet* (2019) 20(7):404–16. doi: 10.1038/s41576-019-0114-6

6. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* (2015) 21(7):751–9. doi: 10.1038/nm.3886

7. Allegretti M, Fabi A, Giordani E, Ercolani C, Romania P, Nistico C, et al. Liquid biopsy identifies actionable dynamic predictors of resistance to trastuzumab emtansine (T-Dm1) in advanced Her2-positive breast cancer. *Mol Cancer* (2021) 20(1):151. doi: 10.1186/s12943-021-01438-z

8. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. Oncokb: A precision oncology knowledge base. *JCO Precis Oncol* (2017) 1:1–16. doi: 10.1200/PO.17.00011

9. Miller CA, McMichael J, Dang HX, Maher CA, Ding L, Ley TJ, et al. Visualizing tumor evolution with the fishplot package for r. *BMC Genomics* (2016) 17(1):880. doi: 10.1186/s12864-016-3195-z

10. Turner NC, Kingston B, Kilburn LS, Kernaghan S, Wardley AM, Macpherson IR, et al. Circulating tumour DNA analysis to direct therapy in advanced breast cancer (Plasmamatch): A multicentre, multicohort, phase 2a, platform trial. *Lancet Oncol* (2020) 21(10):1296–308. doi: 10.1016/s1470-2045(20)30444-7

11. Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* (2021) 53(9):1334–47. doi: 10.1038/s41588-021-00911-1

12. Jordan NV, Bardia A, Wittner BS, Benes C, Ligorio M, Zheng Y, et al. Her2 expression identifies dynamic functional states within circulating breast cancer cells. *Nature* (2016) 537(7618):102–6. doi: 10.1038/nature19328

13. Murtaza M, Dawson SJ, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun* (2015) 6:8760. doi: 10.1038/ncomms9760

14. Sicklick JK, Kato S, Okamura R, Schwaederle M, Hahn ME, Williams CB, et al. Molecular profiling of cancer patients enables personalized combination therapy: The I-predict study. *Nat Med* (2019) 25(5):744–50. doi: 10.1038/s41591-019-0407-5

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 27 January 2023
DOI 10.3389/fonc.2023.1100602

OPEN ACCESS

EDITED BY
Yusuke Suenaga,
Chiba Cancer Center, Japan

REVIEWED BY
Barani Kumar Rajendran,
Yale University, United States
Wen Zhou,
Shanghai Veterinary Research Institute (CAAS), China

*CORRESPONDENCE
Masaki Michishita
michishita@nvlu.ac.jp

†PRESENT ADDRESS
Rei Nakahira,
Department of Radiobiology, Institute for Environmental Sciences, Aomori, Japan

SPECIALTY SECTION
This article was submitted to Molecular and Cellular Oncology, a section of the journal Frontiers in Oncology

RECEIVED 17 November 2022
ACCEPTED 16 January 2023
PUBLISHED 27 January 2023

CITATION
Michishita M, Ochiai K, Nakahira R, Azakami D, Machida Y, Nagashima T, Nakagawa T and Ishiwata T (2023) mTOR pathway as a potential therapeutic target for cancer stem cells in canine mammary carcinoma. *Front. Oncol.* 13:1100602. doi: 10.3389/fonc.2023.1100602

mTOR pathway as a potential therapeutic target for cancer stem cells in canine mammary carcinoma

Masaki Michishita[1,2*], Kazuhiko Ochiai[2,3], Rei Nakahira[1†], Daigo Azakami[4], Yukino Machida[1], Tomokazu Nagashima[1], Takayuki Nakagawa[5] and Toshiyuki Ishiwata[6]

[1]Department of Veterinary Pathology, Faculty of Veterinary Science, Nippon Veterinary and Life Science University, Tokyo, Japan, [2]Research Center for Animal Life Science, Nippon Veterinary and Life Science University, Tokyo, Japan, [3]Department of Veterinary Hygiene, Faculty of Veterinary Science, Nippon Veterinary and Life Science University, Tokyo, Japan, [4]Laboratory of Veterinary Clinical Oncology, Faculty of Agriculture, Tokyo University of Agriculture and Technology, Tokyo, Japan, [5]Laboratory of Veterinary Surgery, Graduate School of Agricultural and Life Science, The University of Tokyo, Tokyo, Japan, [6]Division of Aging and Carcinogenesis, Research Team for Geriatric Pathology, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan

Mammary adenocarcinoma, the most common cancer in female dogs, often exhibits the lymph node and lung metastases and has a higher mortality rate. However, mammary adenocarcinoma has no established treatment, except early surgical excision. Canine mammary carcinoma has many common features with human mammary carcinoma, including clinical characteristics, heterogeneity, and genetic aberrations, making it an excellent spontaneous tumor model for human breast cancer. Diverse cancers comprised heterogeneous cell populations originating from cancer stem cells (CSCs) with self-renewal ability. Therefore, in addition to conventional therapy, therapeutic strategies targeting CSCs are essential for cancer eradication. The present study aimed to extract inhibitors of canine mammary CSCs that suppress their self-renewal ability. Sphere-formation assay, which evaluates self-renewal ability, was performed for the canine mammary cancer cell lines CTBp and CNMp. The spheres formed in this assay were used in inhibitor library screening, which identified various signaling pathways such as proteosome, stress inducer, and mammalian target of rapamycin (mTOR). The present study focused on the mTOR signaling pathway. Western blotting showed higher levels of phosphorylated mTOR in sphere-forming CTBp and CNMp cells than in adherent cells. Drug sensitivity examination using the mTOR inhibitors everolimus and temsirolimus revealed dose-dependent reductions in viability among both sphere-forming cells and adherent cells. Expression of phosphorylated mTOR in adherent and sphere-forming cells decreased by everolimus and temsirolimus treatment. In mice transplanted with CTBp-derived spheres, everolimus treatment significantly decreased tumor volume compared to control. These results reveal that the mTOR signaling pathway may be a potential to be a therapeutic target in both cancer cells and CSCs. Novel therapeutic strategies for canine mammary carcinoma are expected to benefit to human breast carcinoma as well.

KEYWORDS

cancer stem cells (CSC), dog, mammary adenocarcinoma, mTOR, sphere-formation assay

Introduction

Canine mammary tumors are the most common tumors in middle-aged and elderly female dogs (1). Canine mammary carcinoma accounts for approximately 50% of all mammary tumors and is associated with poor clinical behavior, including lymph node and lung metastases, and high mortality (1). Canine mammary cancer shares many common features with human breast cancer, including the age of onset, hormonal etiology, stage, tumor diversity, lymph node metastasis, and genetic abnormalities including breast cancer susceptibility gene 2 (BRCA), phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA), and TP53 mutations, and protein expression, including human epidermal growth factor receptor 2 (HER2), estrogen receptor, progesterone receptor, and p63 (2–6). Therefore, canine mammary cancer is considered a spontaneous model of human breast cancer. Sex hormones are closely related to mammary tumor development in dogs. Ovariectomy before the first estrous reduces the mammary tumor incidence by 99%, and ovariectomy before the second and third estrous reduces them by 92% and 74%, respectively (7). Surgical resection is the first choice for mammary tumor treatment in dogs. Adjuvant therapy is performed for inflammatory breast cancer for which quality of life improvement cannot be expected after surgical resection. Adjuvant therapy for canine mammary tumors includes chemotherapy, such as doxorubicin, cyclophosphamide, and docetaxel, and molecular-targeted therapy, such as toceranib, piroxicam, and firocoxib (8–12). Tamoxifen, which is used as estrogen therapy for human breast cancer, is not recommended in dogs due to adverse reactions, such as vulva edema, purulent discharge, pyometra, and conjunctivitis (13). Additionally, aglepristone, an anti-progestin drug, is used for labor induction and pyometra treatment. Aglepristone has not been practiced due to insufficient therapeutic outcome data, although it is expected to treat dogs with mammary cancers (5). Therefore, developing a novel treatment strategy in addition to conventional therapy is necessary because a treatment protocol, including adjuvant therapy for canine mammary cancer, has not yet been established. Human patients with breast cancer are treated with molecular-targeted drugs using monoclonal antibodies, tyrosine kinase inhibitors, cyclin-dependent kinase 4/6 inhibitors, antiangiogenic agents, and poly (ADP-ribose) polymerase inhibitors in addition to conventional chemotherapy (14). Molecular-targeted drugs were developed to directly act on molecular cancer cell abnormalities and selectively target various signaling pathways related to cancer cell proliferation, aggression, and apoptosis, and have yielded more successful results in cancer therapy (14). Canine mammary cancer has many similarities with human breast cancer, and molecular-targeted therapy for human breast cancer is expected to be beneficial in canine mammary cancer treatment.

Cancer stem cells (CSCs) or tumor-initiating cells are a subpopulation of cancer cells and play an important role in cancer development, recurrence, and metastasis. CSCs have self-renewal and differentiation capacities, higher tumorigenicity in nude mice, and radiotherapy and chemotherapy resistance (15, 16). Therefore, CSCs-targeted therapies are essential for cancer eradication. CSCs can be enriched by various techniques, such as surface antigen analysis, side population analysis, aldeflour assay, and sphere-formation assay, in humans and dogs (17, 18). Sphere-formation assay is an excellent method to efficiently enrich cell populations with self-renewal ability. Canine CSCs have been identified in various cancers, such as mammary adenocarcinoma, hepatocellular carcinoma, pulmonary adenocarcinoma, rhabdomyosarcoma, and melanoma since the existence of CSCs was first reported in osteosarcoma in 2007 (19–27). In veterinary medicine, CSC studies for mammary cancer are most advanced in dogs. Sphere-forming cells derived from mammary adenocarcinoma lines have higher stem cell-related gene expression and higher tumorigenicity in immunodeficient mice compared to adherent cells (20). Additionally, sphere-formation assay is used for *in vitro* sensitivity assay of anticancer drugs, such as doxorubicin and carboplatin, as well as small-molecule inhibitors targeting cyclooxygenase-2, and CSCs exhibit lower sensitivity than non-cancer stem cells (20, 28, 29). Therefore, sphere-formation assay is not only useful for elucidating the pathogenesis of mammary cancer, which is as diverse as in humans, but also for searching for inhibitors and molecular-targeted inhibitors that suppress self-renewal ability. High-throughput screening in humans, using sphere-formation assay, is conducted for various cancers, such as breast and lung cancers, for inhibitors that suppress the self-renewal ability of CSCs (30–32). High-throughput screening of canine mammary CSCs has not been reported although a few agents targeting canine mammary CSCs have been identified so far. Therefore, the present study used a molecular-targeted inhibitor library to search for drugs that suppress the self-renewal ability of CSCs derived from canine mammary cancer lines and focused on the candidate inhibitors targeting the mammalian target of rapamycin (mTOR) signaling pathway extracted by the *in vitro* screening.

Materials and methods

Cell lines and culture

The present study used two canine mammary carcinoma cell lines (CTBp and CNMp) (33). The cell lines were maintained in Dulbecco's modified Eagle medium and nutrient F-12 (DMEM/F12, Invitrogen, Carlsbad, CA, USA), supplemented with 10% fetal bovine serum (Hyclone, Logan, UT, USA) and antibiotics (Nakarai Tesque, Kyoto, Japan) at 37°C in an atmosphere containing 5% CO_2.

Sphere-formation assay

The sphere-forming assay was performed as previously described (18). In brief, singly suspended cells derived from CTBp and CNMp were plated at a density of 1×10^3 or 1×10^5 viable cells per ultralow attachment 96-well plate (Coring, NY, USA) for drug sensitivity or 100-mm dish (Coring) for xenograft, respectively. The cells were grown in serum-free DMEM/F12 supplemented with 10 ng/mL of basic fibroblast growth factor (bFGF, Invitrogen, Carlsbad, CA, USA), 10 ng/mL of epidermal growth factor (EGF, Invitrogen), 4 mg/mL of heparin (Sigma-Aldrich, St. Louis, CA, USA), and NeuroBrew-21 (Miltenyi Biotech, Tokyo, Japan).

In vitro drug screening and sensitivity assay

Sphere assay was performed using SCADS inhibitor Kit IV, including 39 molecule-targeted inhibitors, obtained from the Screening Committee of Anticancer Drugs, Japan (Table 1). Singly suspended cells were cultured in the presence of inhibitors (final concentration: 1 μM or 10 μM) from the beginning for 5 days. Everolimus (LC Laboratories, Boston, MA, USA) and temsirolimus (LC Laboratories) were used for inhibitor sensitivity assay of adherent cells and spheres derived from CTBp and CNMp. The adherent cells, seeded at 5×10^3 cells/well on the 96-well plates, were cultured for 24 h and stimulated with a fresh culture medium containing seven different doses (final concentration: 0.0001, 0.001, 0.01, 0.1, 1, 10, or 100 μM) of everolimus or temsirolimus for 48 h. The sphere assay was performed under the same inhibitor concentration and using the *in vitro* screening protocol described above. Each living cell was evaluated using Cell Counting Kit-8 (Dojindo Laboratories, Kumamoto, Japan).

For analysis of mTOR signaling activity after inhibitor treatment, CTBp and CNMp cell lines were seeded at 5×10^4 and 1×10^4 cells in 35 mm dish for cell culture and 6-well plate for sphere-forming assay, respectively. Adherent cultures were replaced with medium containing everolimus or temsirolimus at a final concentration of 10 μM after 3 days of culture. In the sphere-forming assay, after culturing for 5 days, similar inhibitors were added at a final concentration of 10 μM. Cells were harvested 1 and 4 hours after the addition of the inhibitor, and western blotting described below was performed.

Western blotting

The adherent and sphere-forming cells derived from CTBp and CNMp cells were collected by centrifugation and washed with phosphate-buffer saline. The cells were lysed in lysis buffer (Promega, Tokyo, Japan) with a protein inhibitor cocktail for 15 min. Approximately 10 μg of the extracted protein was analyzed with the following specific monoclonal antibodies against mTOR (clone 7C10, Cell signaling Technology, Tokyo, Japan), phospho-mTOR (Ser2448) (clone 49F9, Cell Signaling Technology), 4E-BP (clone 53H11, Cell Signaling Technology) and phospho-4E-BP (Thr37/46) (clone 236B4, Cell Signaling Technology), and polyclonal antibody against β-actin (Santa Cruz Biotechnology). The membranes were incubated with horseradish peroxidase-conjugated immunoglobulin G (IgG) (GE Healthcare, Tokyo, Japan). The immunoreactivity was detected using an ATTO EzWestLumi plus reagent (ATTO, Tokyo, Japan) and ImageQuant LAS4000 mini (GE Healthcare).

Xenograft transplantation

Female BLAB/c nude mice, aged 8 weeks, were purchased from CLEA Inc. (Tokyo, Japan). A suspension of 1×10^6 sphere-forming cells derived from CTBp was subcutaneously injected into the ventrolateral area under anesthesia. We administered saline (control, n = 6/group) or everolimus (Novartis Pharma, Basel, Switzerland, 4 mg/kg; oral n = 4/group) intraorally twice a week for 21 days after tumor formation was macroscopically confirmed. The tumor volume (V) was estimated using the following equation: $V = [(\text{length}) \times (\text{width})^2]/2$. Experiments were approved by the Animal Experiments Committee of Nippon Veterinary and Life Science University and were performed following the Guidelines for Animal Experiments by the Nippon Veterinary and Life Science University.

Histopathology

The tumors formed in nude mice were fixed with 10% neutral-buffered formalin and routinely embedded in paraffin wax for histological examination. Sections were stained with hematoxylin and eosin. Serial sections were immunostained using the streptavidin-biotin-peroxidase method with primary monoclonal antibodies specific for Ki67 (1:100, Dako, Denmark A/S, Glostrup, Denmark) and alpha-smooth muscle actin (SMA, 1:400, Dako), vascular endothelial growth factor (VEGF, 1:100, Santa Cruz Biotechnology, California, USA). Briefly, sections were treated in 0.03% H_2O_2 in 33% methanol at room temperature for 30 min for endogenous peroxidase blocking, following a pretreatment at 121°C for 20 min in citrate buffer (pH 6.0) for Ki67 and SMA, and at 121°C for 15min in citrate buffer (pH 9.0) for VEGF. The validation of antibodies was confirmed by a positive reaction with biopsy samples diagnosed with canine mammary adenocarcinoma or by a negative normal mouse IgG. The intratumor SMA-positive vessel and Ki67 index of tumor cell densities were evaluated as previously described (34). To evaluate the immunostaining intensity of VEGF, 5 high-power field (x400) of tumor tissue were selected and measured using Image J software.

Statistical analysis

The results are presented as means ± standard deviation. Student *t*-test and Welch's *t*-test were used for statistical analyses with R version 4.2.2. *P*-values of <0.05 was considered significant.

Results

In vitro library screening using molecular-targeted inhibitors in canine mammary adenocarcinoma cell lines

A sphere-formation assay was performed using a molecular-targeted inhibitor kit consisting of 39 types to extract inhibitors that suppress the self-renewal ability. A decreased value of ≥50% was found in 23 inhibitors in CTBp compared to control under conditions of final concentrations of 1 μM and 10 μM. Conversely, 4 and 12 inhibitors were extracted at final concentrations of 1 μM and 10 μM, respectively, in CNMp. Eleven inhibitors, such as EGF receptor (lapatinib, gefitinib), proteosome (bortezomib), stress inducer (anisomycin), and mTOR (temsirolimus, everolimus, and torkinib), were common between both lines (Table 1). This study focused on the mTOR signaling pathway evaluated by western blotting and the *in vitro* and *in vivo* antitumor effects of adherent

10.3389/fonc.2023.1100602

TABLE 1 The viability of the sphere-forming cells derived from canine mammary carcinoma cell lines cultured with various concentration of the inhibitors supplied in SCADS inhibitor kit IV.

SCADS kit4 inhibitors		Viability (% of control)			
		CTBp		CNMp	
Targets	Inhibitors	1μM	10μM	1μM	10μM
blank	none (DMSO)	100.00	100.00	100.00	100.00
Bcr-Abl	nilotinib	60.51	64.82	64.25	97.16
Multi-kinases	sorafenib	**12.85**	**11.21**	90.95	65.01
mTOR	temsirolimus	**27.64**	**28.30**	74.66	**29.27**
EGFR/Her2	lapatinib	**13.43**	**12.21**	63.35	**27.28**
Bcr-Abl/Kit	imatinib mesylate	136.30	139.07	68.33	105.99
Multi-kinases	sunitinib malate	**20.27**	**16.76**	139.37	108.65
EGFR	gefitinib	**12.47**	**10.77**	54.75	**20.55**
HDAC	vorinostat	**11.60**	**9.99**	60.63	**13.21**
EGFR	erlotinib	**13.22**	**12.10**	93.67	**22.01**
Proteasome	bortezomib	**13.26**	**11.88**	50.23	**12.01**
Bcr-Abl/Src	dasatinib	**13.22**	**11.88**	90.50	**32.82**
mTOR	everolimus	**31.16**	**30.41**	106.33	**46.29**
Multi-kinases	pazopanib	**13.76**	**12.76**	150.23	63.47
Rho/SRF	CCG-1423	67.68	57.60	106.79	121.89
PIM	PIM1/2 Kinase Inhibitor V	69.62	61.38	143.44	151.68
PIM	PIM1 Inhibitor II	101.99	85.68	113.57	100.97
Hedgehog	AY 9944	**12.76**	**11.43**	99.10	76.12
Hedgehog	cyclopamine	**44.63**	**36.85**	92.76	68.23
Hedgehog	Jervine	**47.20**	**39.29**	94.57	58.35
STAT3	WP1066	**13.39**	**11.99**	134.84	86.98
STAT3	5,15-DPP	67.55	59.49	143.89	128.26
Wnt	IWP-2	91.34	69.92	143.44	117.96
Wnt	IWR-1-endo	68.50	64.04	74.66	158.72
Wnt	FH535	72.81	68.92	75.11	109.54
Notch	DAPT	**47.16**	**38.07**	186.88	86.75
tankyrase-selective PARP	XAV939	112.14	89.57	145.70	178.41
pan-PARP	PJ-34	**41.07**	**30.74**	185.07	137.95
PARP-1/2-selective	Olaparib	**17.24**	**16.32**	115.38	92.66
antipsychotic drug	chlorpromazine hydrochloride	**25.90**	**43.51**	163.35	118.86
depression treatment	desipramine hydrochloride	103.44	80.36	208.14	128.52
golgi inhibitor	brefeldin A	**11.07**	**9.99**	176.02	109.02
stress inducer	anisomycin	**14.94**	**11.07**	**4.58**	**4.61**
thalidomide family	thalidomide	126.68	95.98	92.74	92.43
thalidomide family	lenalidomide	126.15	102.65	86.30	79.14
retinoids	tretinoin	94.34	78.28	66.90	**42.83**
retinoids	tamibarotene	106.01	109.54	**40.30**	51.73

(Continued)

TABLE 1 Continued

SCADS kit4 inhibitors		Viability (% of control)			
		CTBp		CNMp	
Targets	Inhibitors	1μM	10μM	1μM	10μM
DNA alkylation	temozolomide	132.52	103.36	66.54	86.24
EML4-ALK	crizotinib	**42.50**	**11.73**	**43.41**	**4.64**
mTOR	Torkinib	**36.08**	**12.11**	**17.45**	**6.15**

The viability of less than 50% are indicated by boldface. Data represet the mean of triplecate culture.

cells and sphere-forming cells derived from canine mammary adenocarcinoma lines.

Activated mTOR signal pathway was detected in both adherent and sphere-forming cells

Western blotting was performed to confirm the expression of mTOR signal-related proteins, such as mTOR and 4E-BP1 in adherent and sphere-forming cells from canine mammary adenocarcinoma lines. Expression levels of mTOR and phosphorylated mTOR proteins were similar between adherent and sphere-forming cells of both lines (Figure 1). Conversely, 4E-BP1 was expressed in adherent and sphere-forming cells and phosphorylated 4E-BP1 was more highly expressed in sphere-forming than adherent cells of both lines (Figure 1). These results revealed that mTOR signaling was activated in both CTBp and CNMp.

mTOR inhibitors, including everolimus and temsirolimus, inhibit adherent proliferation and sphere formation *in vitro*

In vitro sensitivity assay was performed using the mTOR inhibitors, including everolimus and temsirolimus, to examine inhibitory effects on sphere formation and adherent cell proliferation. Both CTBp and CNMp decreased the number of

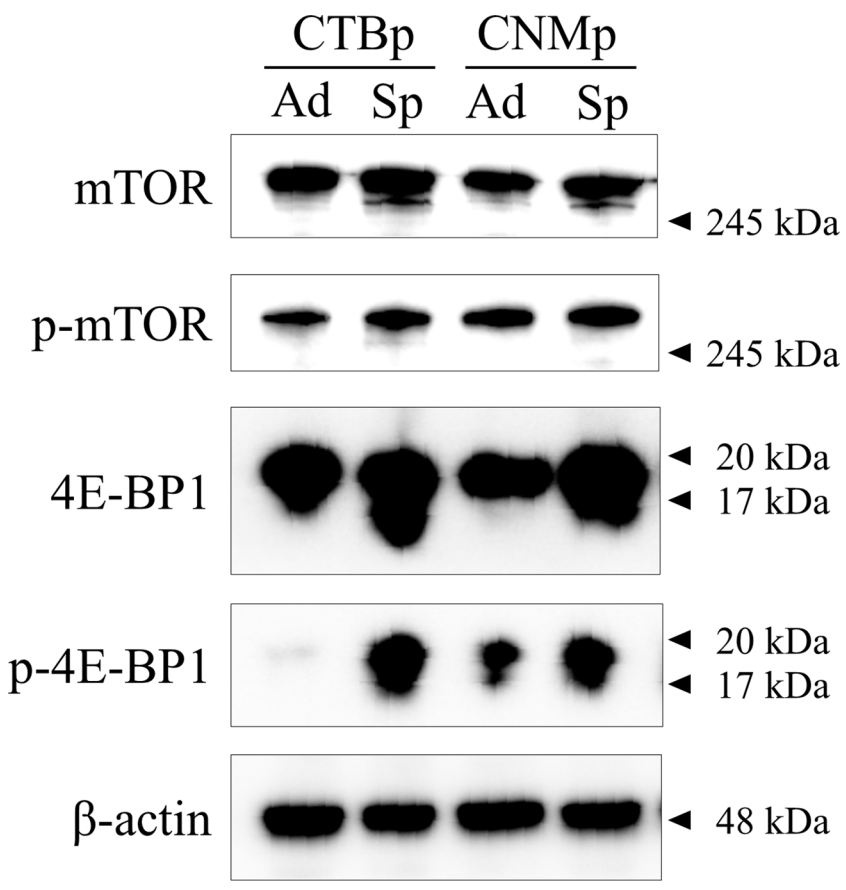


FIGURE 1
Western blot analysis of mTOR signaling in canine mammary adenocarcinoma lines, CTBp and CNMp. Adherent cells, ad; sphere-forming cells, Sp.

sphere-forming and adherent cells in a dose-dependent manner with everolimus and temsirolimus (Figure 2). The IC_{50} for everolimus and temsirolimus was 158.8 nM and 123 nM in CTBp-derived (Figure 2A) and 1.16 μM and 3.13 nM in CNMp-derived sphere-forming cells, respectively (Figure 2B). Whereas, that in CTBp- and CNMp-derived adherent cells were 17.0 nM and 39.5 nM (Figure 2C) and 53.6 μM and 52.9 nM (Figure 2D), respectively. In adherent cultures, the cell numbers of CTBp and CNMp lines were measured 24 and 48 hrs after treatment of inhibitors, and their numbers tended to be time-dependent (Supplemental Figure 1). Furthermore, in the sphere-forming assay, the number of CTBp-derived sphere-forming cells was measured 2 and 4 days after addition and was time-dependent similar to adherent cells (Supplemental Figure 1).

mTOR inhibitors decrease the phosphorylation of mTOR signal in adherent and sphere-forming cells

Western blotting was performed to examine the expression of mTOR signal with 10 μM everolimus and temsirolimus treatment. In both cell lines, adherent cells decreased phosphorylated mTOR and 4E-BP1 expression 1 and 4 hours after inhibitor treatment (Figure 3A). On the other hand, sphere-forming cells treated with everolimus and temsirolimus also decreased phosphorylated 4E-BP1and mTOR expression (Figure 3B).

Everolimus exhibits the antitumor effect in xenograft mice injected with CTBp-derived sphere-forming cells

The *in vivo* antitumor effect of everolimus was investigated using mice transplanted with CTBp-derived sphere-forming cells. A significant tumor volume reduction was observed in the everolimus-administered group compared to the control group 14 and 21 days after administration (Figure 4). Histologically, the tumors formed in mice were similar in both groups and consisted of tubular or solid tumor cell proliferation (Figures 5A, B). Tumor necrosis and inflammatory cells, such as lymphocytes and mast cells, were not observed in both group. The Ki67 index of tumor cells was 12.72 ± 9.17 and 16.43 ± 19.69 in the control and everolimus-administered

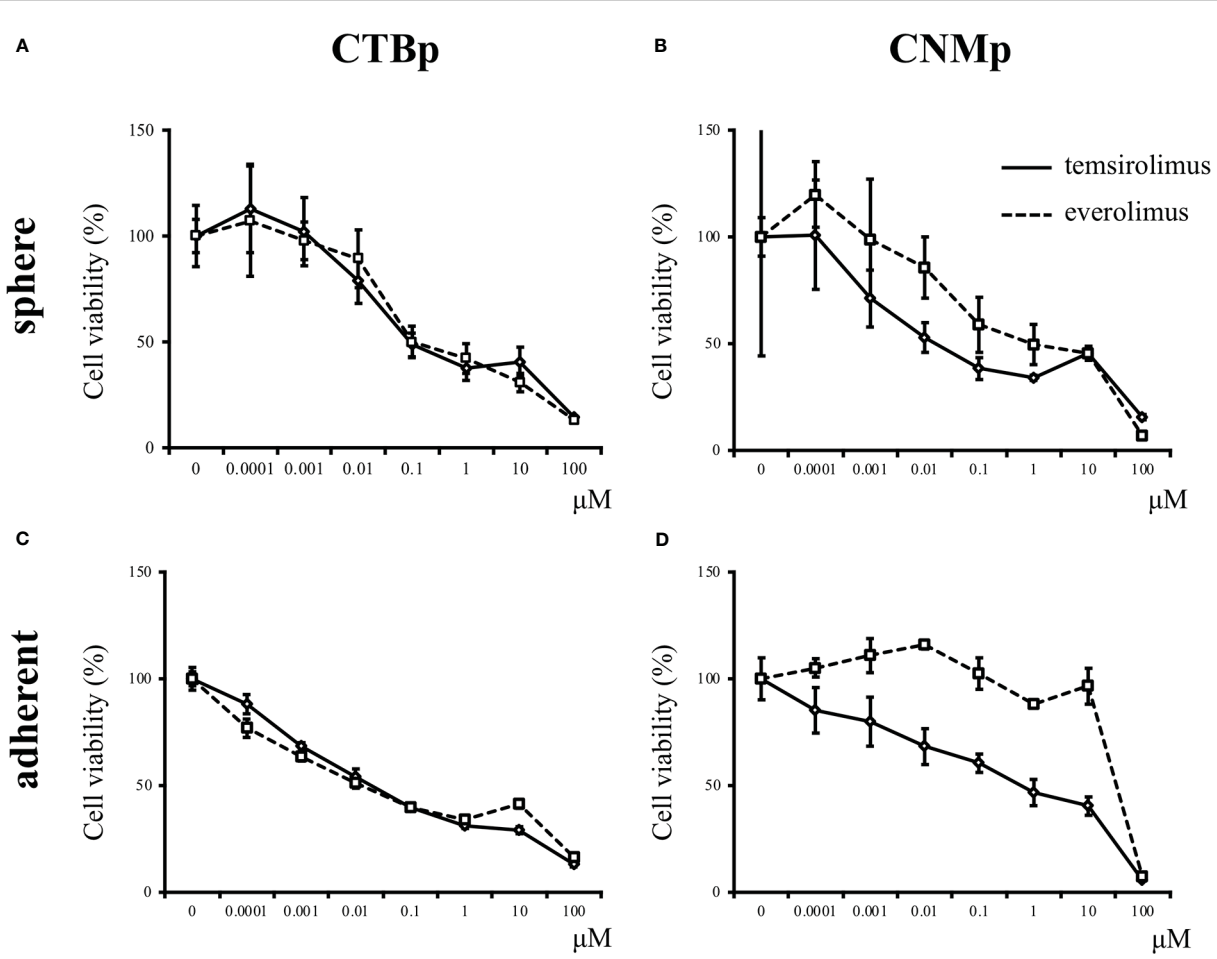


FIGURE 2
In vitro sensitivity assay of mTOR inhibitors, including everolimus and temsirolimus, in canine mammary adenocarcinoma lines, CTBp and CNMp. **(A, C)** CTBp and **(B, D)** CNMp. Upper shows for sphere-forming cells and lower shows the sensitivity assay results for adherent cells. The results shown are representative of at least three independent experiments.

10.3389/fonc.2023.1100602

groups, respectively (Figure 5C). The number of intratumoral SMA-positive vessels was 14.48 ± 4.08 and 11.50 ± 4.51 in the control group and the everolimus-administered group, respectively (Figure 5D). Almost all tumor cells were positive for VEGF. VEGF immunostaining intensity of tumor cells was 214.5 ± 12.3 and 216.3 ± 6.49 in the control and evelorimus-administrated groups, respectively (Figure 5E). A significant difference was found in tumor volume, but with no significant difference between the two groups in both the Ki67 index, VEGF expression of tumor cells and the number of SMA-positive vessels.

Discussion

This study conducted an *in vitro* library screening to suppress the self-renewal ability of spheres derived from canine mammary adenocarcinoma CTBp and CNMp lines and extracted molecular-targeted inhibitors, such as mTOR, hedgehog, and proteasome. This study focused on the mTOR signal. The activation of mTOR was observed in both sphere-forming cells and adherent cells of canine mammary cancer, and 4E-BP, which is an mTOR downstream signal, was activated in sphere-forming cells. Furthermore, *in vitro*

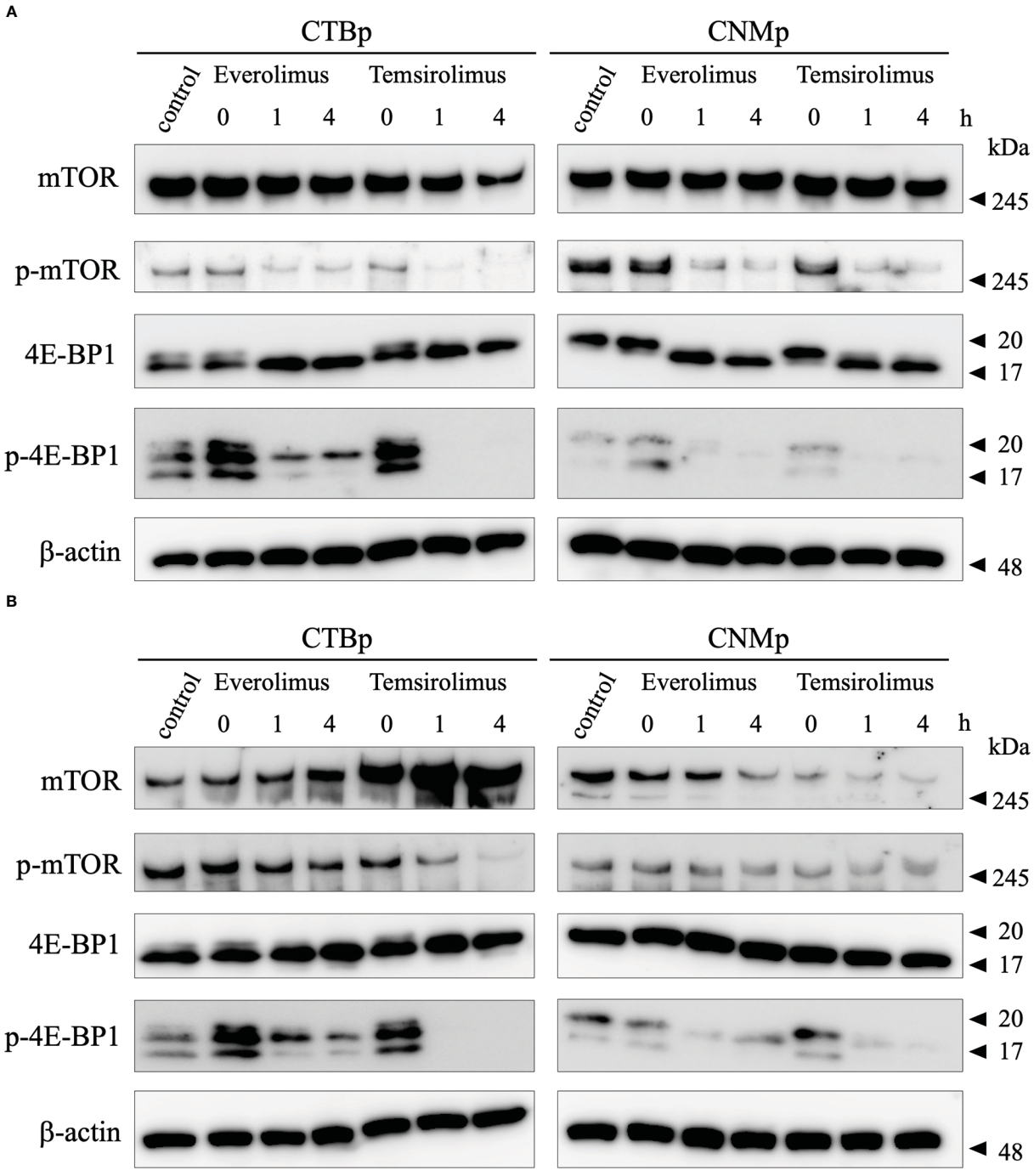


FIGURE 3
Western blot analysis of mTOR signal after evelorimus and temsirolimus treatment. **(A)** adherent cells, **(B)** sphere-forming cells.

sensitivity assay of everolimus and temsirolimus showed their growth suppression in both adherent cells and spheres, and everolimus revealed an antitumor effect in mice transplanted with sphere-forming cells. These results suggest that *in vitro* screening by sphere-formation assay using an inhibitor library is extremely useful for extracting inhibitors that suppress the self-renewal ability of CSCs in canine mammary carcinoma.

mTOR is a serine-threonine kinase that functions as a key downstream target of the phosphatidyl-inositol-3 kinase (PI3K)/AKT signaling pathway and has various regulatory functions, such as cell proliferation, metabolism, angiogenesis, and autophagy (35–37). mTOR forms a complex of mTORC1 and mTORC2, the 4E-BP1 and S6K exist downstream of mTORC1. Additionally, mTOR activation is associated with tumor development and plays an important role in maintaining the stemness of CSCs (36, 38). In human breast cancer, mTOR activation in CSCs is important for colony-forming and tumorigenicity (39). Activation of mTOR in CSCs has been reported in various cancers, such as colon cancer, prostate cancer, salivary gland cancer, and glioblastoma (40–43). mTOR signaling suppression reduces aldehyde dehydrogenase activity, which is abundant in immature cells, such as stem cells, in colon cancer (44). Therefore, mTOR signaling has attracted attention as a therapeutic target for various cancers (45). Everolimus has an inhibitory effect on breast CSCs (46, 47). Additionally, metformin exhibits antitumor effects on breast CSCs *via* AMP-activated protein kinase (AMPK) activation and mTOR suppression (48, 49). In dogs, mTOR phosphorylation has been detected by immunohistochemistry in various tumors, such as mammary tumors, squamous cell carcinoma, trichoblastoma, myxosarcoma, hemangiosarcoma, and prostate cancer (50–54). Phosphorylated mTOR expression in canine mammary tumors is associated with tumorigenesis and negative clinical behavior (51). Additionally, mTOR phosphorylation has been demonstrated in many cancer lines, such as osteosarcoma, melanoma, hemangiosarcoma, mast cell tumor, breast adenocarcinoma, glioma, and lymphoma, by western blotting, and rapamycin, which is an mTOR inhibitor, is associated with decreased mTOR phosphorylation and cell viability (50, 55–58). Cancer therapy targeting the PI3K/mTOR signaling pathway is expected to have an antitumor effect against canine mammary cancer and melanoma (59–61). The dual PI3K/mTOR inhibitor, VDC597, dose-dependently reduces cell proliferation, invasion, and vascular endothelial growth factor production in canine hemangiosarcoma (58). However, the role of mTOR signaling in canine mammary CSCs remains unclear. Therefore, this study

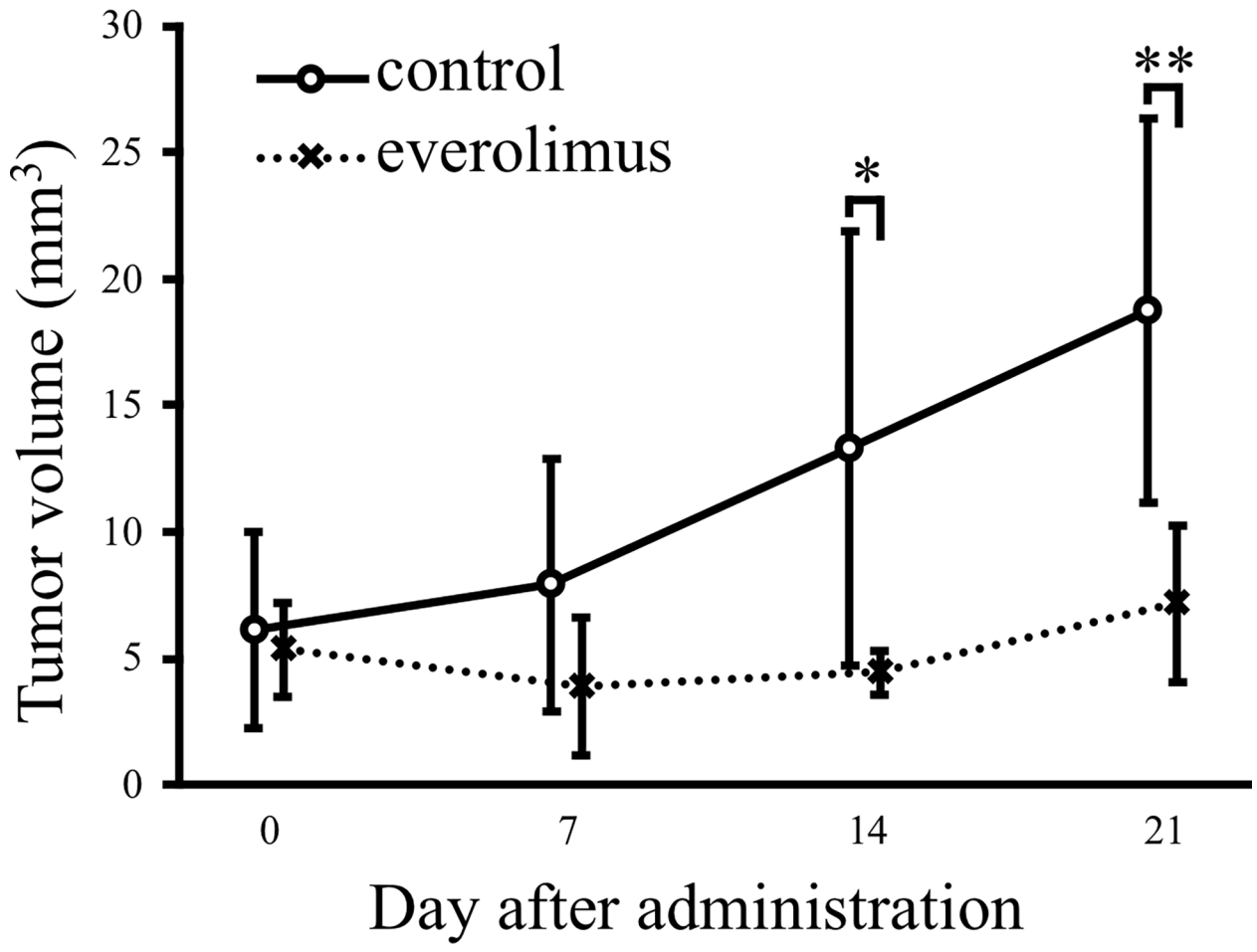


FIGURE 4
Antitumor effects of everolimus in a xenograft model transplanted canine mammary adenocarcinoma CTBp. Everolimus (n = 4, squares) or saline (n = 6, circles) was administrated twice per week for 21 days. The differences were tested by Scheffe's F test. *$P < 0.05$, **$P < 0.01$.

proposes the use of mTOR inhibitors as targeted therapies for CSCs and cancer cells in canine mammary cancers, similar to human breast cancers. Unfortunately, the mechanism of the antitumor effect of everolimus in canine mammary cancer model mice could not be clarified because no difference was found in mitotic figures and angiogenesis between control and mTOR-administered groups. VEGF, which plays an important role in tumor angiogenesis, expression in tumor cells between control and everolimus-administrated groups supports the result that there is no difference in intratumoral angiogenesis between both group. Further studies will reveal the mechanisms underlying the antitumor effect of mTOR inhibitors.

Hedgehog (HH) signaling plays an important role not only in promoting embryonic development and cell differentiation but also in tumor initiation and progression (62). Additionally, HH signaling is essential not only for normal stem cells but also for maintaining CSC stemness (63). HH signaling pathway dysregulation in human breast cancer has been implicated in triple-negative and HER2-positive breast cancers and is persistently activated in CSCs, thereby promoting CSC's self-renewal ability (63–66). Therefore, HH signaling is one of the cancer therapeutic targets. HH signaling in dogs is expressed in cancer cell lines, including osteosarcoma and transitional cell carcinoma, and HH inhibitors, such as cyclopamine, GANT61, and vismodegib, suppress tumor

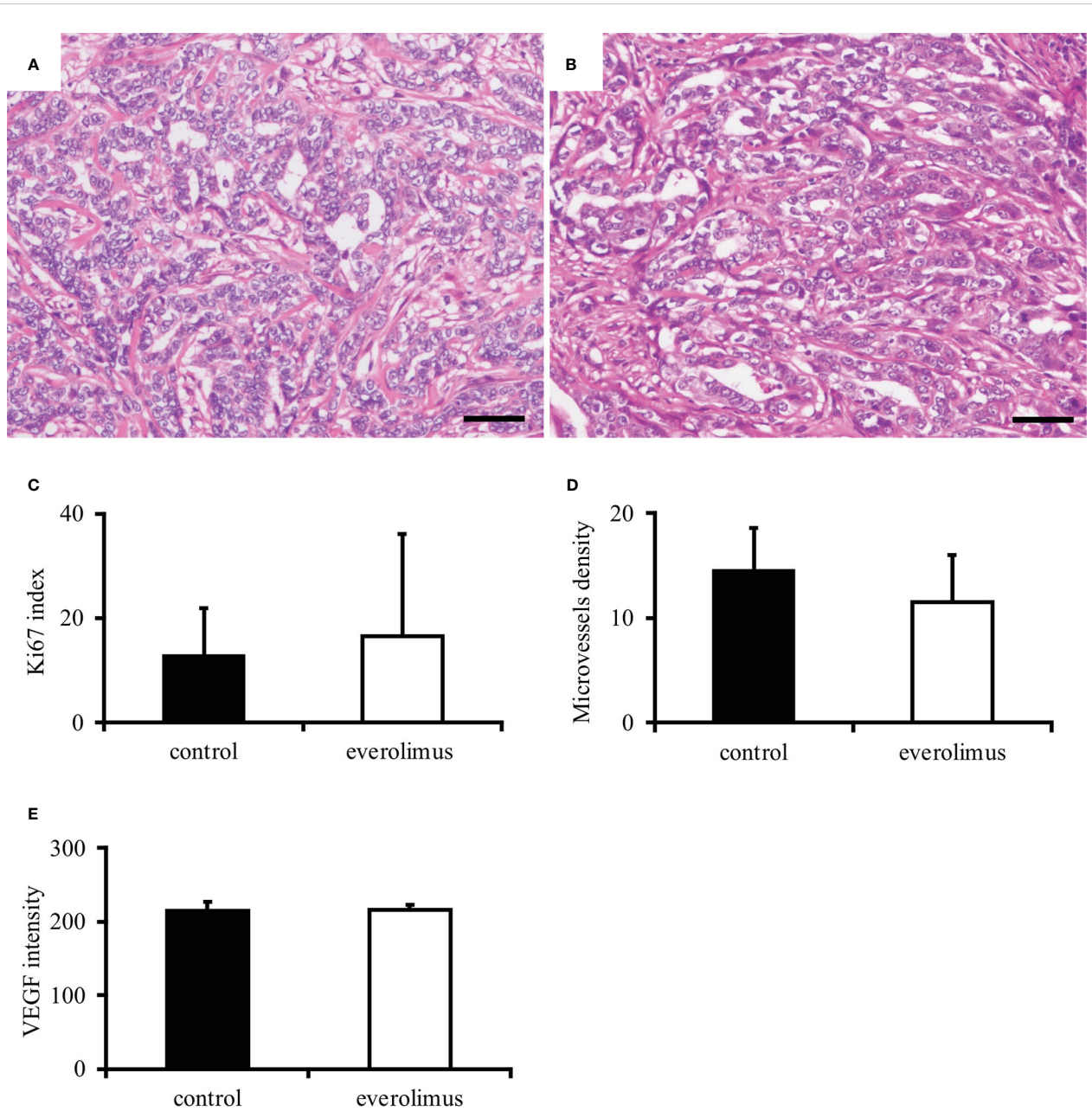


FIGURE 5
Histopathological evaluation of canine mammary carcinoma model mice. Tumors formed in xenografts show similar histology in both control **(A)** and everolimus-administered groups **(B)**. Hematoxylin and eosin. Scale bar = 50 µm. No significant difference was found in the Ki67 index **(C)**, smooth muscle actin-positive microvessels density **(D)**, and VEGF intensity **(E)** between the control and the everolimus-administered groups. The differences were determined by the Student *t*-test and Weltch's *t*-test.

proliferation in these cancers (67–69). However, the antitumor effects in canine mammary cancer-containing CSCs remained unknown. In the present study, HH signaling, such as AY9944, cyclopamine, and jervine, is identified as a candidate to suppress the self-renewal ability of CSCs from the CTBp line, although detailed analysis has not been performed. Therefore, HH signaling may be a potential therapeutic target in canine mammary carcinoma, similar to human breast cancers.

The proteasome inhibitor, bortezomib, showed high sensitivity to canine mammary adenocarcinoma lines in this high-throughput screening. Bortezomib induces cell death *via* proteotoxic stress and alters the pro/anti-apoptotic protein balance by inhibiting ubiquitinated protein degradation by the 20S proteasome (70). Bortezomib is a Food and Drug Administration-approved therapeutic drug for multiple myeloma and mantle cell lymphoma (71, 72). Conversely, bortezomib monotherapy has had poor outcomes in patients with metastatic breast cancer, whereas a study reported 11 months of progression-free survival without adverse events in patients with metastatic triple-negative breast cancer with TP53 mutations (73, 74). CSCs are more resistant to bortezomib than differentiated cancer cells, but bortezomib-encapsulated nanoparticles can affect the stemness of CSCs compared to free bortezomib (75, 76). Veterinary medicine has shown higher sensitivity in canine cancer lines, including prostate cancer, lymphoma, melanoma, and osteosarcoma, as well as antitumor efficacy in melanoma-transplanted mice, but bortezomib sensitivity in CSCs has never been evaluated (77–80). Therefore, molecular-targeted therapy using bortezomib is expected to be beneficial as a cancer treatment for dogs. Further studies will reveal that bortezomib has antitumor effects in canine mammary cancer, although this study performed no detailed bortezomib analysis in canine mammary CSCs.

Sphere-forming cells are less sensitive to drugs, such as doxorubicin, carboplatin, and cyclooxygenase-2, than adherent cancer cells (20, 28, 29). *In vitro* library screening revealed the presence of inhibitors, including Wnt, PIM, and thalidomide family, that showed low sensitivity to the two concentrations used in this study, suggesting that all inhibitors are insentisitive to CSCs. Furthermore, further research is essential to determine whether inhibitors that are sensitive to sphere-forming cells can acquire resistance to them.

Sphere is a cancer stem cell population with self-renewal and differentiation ability (20). Characterization of sphere-forming cells in canine mammary carcinoma will lead not only to the elucidation of the pathogenesis of mammary carcinoma, but also to the development of therapies targeting CSCs (18). The tumor microenvironment plays a critical role in the stemness of CSCs, and also contributes to tumor progression and resistance to therapeutic agents (81, 82). The tumor microenvironment comprises a diverse population of cells, including fibroblasts, cancer-associated fibroblasts (CAFs), mesenchymal stem cells, endothelial cells, immune cells, such as T lymphocytes, macrophages, and dendritic cells (83). However, sphere-forming assay can not construct a microenvironment. Therefore, co-culture of spheres and CAFs can construct a microenvironment that is useful for further characterization of CSCs (84). Furthermore, organoids are formed in 3-dimentional cultures, but, unlike spheres, they form mimics of cancer tissues composed of CSCs, cancer cells, and microenvironment that are construct *in vivo* (85). In further research, in addition to spheres, *in vitro* drug screening targeting cancer organoids will be essential for the development of new therapeutic strategies in veterinary medicine.

Canine mammary cancer is an excellent spontaneous intermediate animal model for human breast cancer study, and new therapeutic studies for canine mammary cancer are a promising area in comparative oncology. However, the results of this study are limited to cell culture and mammary cancer model mice, and the therapeutic effect in dogs with mammary cancer remain unclear. In the future, clinical trials in dogs with breast cancer are essential. Therefore, establishing new therapeutic strategies and developing novel therapeutic protocols for canine mammary cancer is expected to bring beneficial benefits not only to veterinary medicine but also to human breast cancer treatment. Human and canine oncology need to collaborate in breast cancer research following the one health concept.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

The animal study was reviewed and approved by Animal Experimental Committee, Nippon Veterinary and Life Science University.

Author contributions

MM: conception, design, and writing the draft. All authors contributed to data acquisition and analysis, and approved the submitted version.

Funding

This work was supported by a Grant in Aid for Scientific Research of JSPS KAKENHI Grant Numbers 26292161, 19K06391, and 22K05996.

Acknowledgments

The authors thank Ms. Shoko Yashima, and Yuki Katori for histopathological support.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1100602/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Time-susceptibility to everolimus and temsirolimus in mammary adenocarcinoma cells. CTBp **(A, B)** and CNMp **(C, D)** adherent cells treated with evelorimus and temsirolimus. CTBp-derived sphere-forming cells treated with evelorimus **(E)** and temsirolimus **(F)**.

References

1. Misdorp W, Else RW, Hellmén E, Lipscomb TP. *Histological classification of mammary tumors of the dog and the cat. 2nd ed*. Armed Forces Institute of Pathology in Cooperation with the American Registry of Pathology and the World Health Organization Collaborating Centre for Worldwide Reference on Comparative. (Washington, DC, USA: Oncology) (1999).

2. Rajendran BK, Deng CX. Characterization of potential driver mutations involved in human breast cancer by computational approaches. *Oncotarget* (2017) 8:50252–72. doi: 10.18632/oncotarget.17225

3. Abdelmegeed SM, Mohammed S. Canine mammary tumors as a model for human disease. *Oncol Lett* (2018) 15:8195–205. doi: 10.3892/ol.2018.8411

4. Lee KH, Hwang HJ, Noh HJ, Shin TJ, Cho JY. Somatic mutation of *PIK3CA* (H1047R) is a common driver mutation hotspot in canine mammary tumors as well as human breast cancers. *Cancers (Basel).* (2019) 11:2006. doi: 10.3390/cancers11122006

5. Valdivia G, Alonso-Diez Á, Pérez-Alenza D, Peña L. From conventional to precision therapy in canine mammary cancer: A comprehensive review. *Front Vet Sci* (2021) 8:623800. doi: 10.3389/fvets.2021.623800

6. Miao K, Lei JH, Valecha MV, Zhang A, Xu J, Wang L, et al. NOTCH1 activation compensates BRCA1 deficiency and promotes triple-negative breast cancer formation. *Nat Commun* (2020) 11:3256. doi: 10.1038/s41467-020-16936-9

7. Schneider R, Dorn CR, Taylor DO. Factors influencing canine mammary cancer development and postsurgical survival. *J Natl Cancer Inst* (1969) 43:1249–61.

8. Poirier VJ, Hershey AE, Burgess KE, Phillips B, Turek MM, Forrest LJ, et al. Efficacy and toxicity of paclitaxel (Taxol) for the treatment of canine malignant tumors. *J Vet Intern Med* (2004) 18:219–22. doi: 10.1892/0891-6640(2004)18<219:eatopt>2.0.co;2

9. Simon D, Schoenrock D, Baumgärtner W, Nolte IJ. Postoperative adjuvant treatment of invasive malignant mammary gland tumors in dogs with doxorubicin and docetaxel. *Vet Intern Med* (2006) 20:1184–90. doi: 10.1892/0891-6640(2006)20[1184:patimm]2.0.co;2

10. Lavalle GE, De Campos CB, Bertagnolli AC, Cassali GD. Canine malignant mammary gland neoplasms with advanced clinical staging treated with carboplatin and cyclooxygenase inhibitors. *In Vivo.* (2012) 26:375–9.

11. Kuruoglu FE, Ozyigit MO, Nak D, Avci Kupeli Z, Ekici ZM, Koca D, et al. Efficacy and toxicity of doxorubicin and cyclophosphamide for the neoadjuvant treatment of locally advanced stage canine mammary tumors. *Kafkas Univ Vet Fak Derg.* (2020) 26:729–34. doi: 10.9775/kvfd.2020.24112

12. Alonso-Miguel D, Valdivia G, García-San José P, Alonso-Diez Á, Clares I, Portero M, et al. Clinical outcome of dogs diagnosed with canine inflammatory mammary cancer treated with metronomic cyclophosphamide, a cyclooxygenase-2 inhibitor and toceranib phosphate. *Vet Comp Oncol* (2022) 20:179–88. doi: 10.1111/vco.12760

13. Tavares WL, Lavalle GE, Figueiredo MS, Souza AG, Bertagnolli AC, Viana FA, et al. Evaluation of adverse effects in tamoxifen exposed healthy female dogs. *Acta Vet Scand* (2010) 52:67. doi: 10.1186/1751-0147-52-67

14. Demir Cetinkaya B, Biray Avci C. Molecular perspective on targeted therapy in breast cancer: a review of current status. *Med Oncol* (2022) 39:149. doi: 10.1007/s12032-022-01749-1

15. Clarke MF, Dick JE, Dirks PB, Eaves CJ, Jamieson CH, Jones DL, et al. Cancer stem cells–perspectives on current status and future directions: AACR workshop on cancer stem cells. *Cancer Res* (2006) 66:9339–44. doi: 10.1158/0008-5472.CAN-06-3126

16. Visvader JE, Lindeman GJ. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nat Rev Cancer.* (2008) 8:755–68. doi: 10.1038/nrc2499

17. Charafe-Jauffret E, Monville F, Ginestier C, Dontu G, Birnbaum D, Wicha MS. Cancer stem cells in breast: current opinion and future challenges. *Pathobiology* (2008) 75:75–84. doi: 10.1159/000123845

18. Michishita M. Understanding of tumourigenesis in canine mammary tumours based on cancer stem cell research. *Vet J* (2020) 265:105560. doi: 10.1016/j.tvjl.2020.105560

19. Wilson H, Huelsmeyer M, Chun R, Young KM, Friedrichs K, Argyle DJ. Isolation and characterisation of cancer stem cells from canine osteosarcoma. *Vet J* (2008) 175:69–75. doi: 10.1016/j.tvjl.2007.07.025

20. Michishita M, Akiyoshi R, Yoshimura H, Katsumoto T, Ichikawa H, Ohkusu-Tsukada K, et al. Characterization of spheres derived from canine mammary gland adenocarcinoma cell lines. *Res Vet Sci* (2011) 91:254–60. doi: 10.1016/j.rvsc.2010.11.016

21. Nemoto Y, Maruo T, Sato T, Deguchi T, Ito T, Sugiyama H, et al. Identification of cancer stem cells derived from a canine lung adenocarcinoma cell line. *Vet Pathol* (2011) 48:1029–34. doi: 10.1177/0300985810396106

22. Pang LY, Cervantes-Arias A, Else RW, Argyle DJ. Canine mammary cancer stem cells are radio- and chemo- resistant and exhibit an epithelial-mesenchymal transition phenotype. *Cancers (Basel).* (2011) 3:1744–62. doi: 10.3390/cancers3021744

23. Michishita M, Akiyoshi R, Suemizu H, Nakagawa T, Sasaki N, Takemitsu H, et al. Aldehyde dehydrogenase activity in cancer stem cells from canine mammary carcinoma cell lines. *Vet J* (2012) 193:508–13. doi: 10.1016/j.tvjl.2012.01.006

24. Guth AM, Deogracias M, Dow SW. Comparison of cancer stem cell antigen expression by tumor cell lines and by tumor biopsies from dogs with melanoma and osteosarcoma. *Vet Immunol Immunopathol* (2014) 161:132–40. doi: 10.1016/j.vetimm.2014.07.006

25. Michishita M, Ezaki S, Ogihara K, Naya Y, Azakami D, Nakagawa T, et al. Identification of tumor-initiating cells in a canine hepatocellular carcinoma cell line. *Res Vet Sci* (2014) 96:315–22. doi: 10.1016/j.rvsc.2014.01.004

26. Kishimoto TE, Yashima S, Nakahira R, Onozawa E, Azakami D, Ujike M, et al. Identification of tumor-initiating cells derived from two canine rhabdomyosarcoma cell lines. *J Vet Med Sci* (2017) 79:1155–62. doi: 10.1292/jvms.16-0412

27. Tanabe A, Deguchi T, Sato T, Nemoto Y, Maruo T, Madarame H, et al. Radioresistance of cancer stem-like cell derived from canine tumours. *Vet Comp Oncol* (2016) 14:e93–e101. doi: 10.1111/vco.12110

28. Pang LY, Argyle SA, Kamida A, Morrison KO, Argyle DJ. The long-acting COX-2 inhibitor mavacoxib (Trocoxil™) has anti-proliferative and pro-apoptotic effects on canine cancer cell lines and cancer stem cells in vitro. *BMC Vet Res* (2014) 10:184. doi: 10.1186/s12917-014-0184-9

29. Barbieri F, Wurth R, Ratto A, Campanella C, Vito G, Thellung S, et al. Isolation of stem-like cells from spontaneous feline mammary carcinomas: phenotypic characterization and tumorigenic potential. *Exp Cell Res* (2012) 318:847–60. doi: 10.1016/j.yexcr.2012.02.008

30. Gupta PB, Onder TT, Jiang G, Tao K, Kuperwasser C, Weinberg RA, et al. Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* (2009) 138:645–59. doi: 10.1016/j.cell.2009.06.034

31. Choi HS, Kim DA, Chung H, Park IH, Kim BH, Oh ES, et al. Screening of breast cancer stem cell inhibitors using a protein kinase inhibitor library. *Cancer Cell Int* (2017) 17:25. doi: 10.1186/s12935-017-0392-z

32. She X, Gao Y, Zhao Y, Yin Y, Dong Z. A high-throughput screen identifies inhibitors of lung cancer stem cells. *BioMed Pharmacother.* (2021) 140:111748. doi: 10.1016/j.biopha.2021.111748

33. Uyama R, Nakagawa T, Hong SH, Mochizuki M, Nishimura R, Sasaki N. Establishment of four pairs of canine mammary tumour cell lines derived from primary and metastatic origin and their e-cadherin expression. *Vet Comp Oncol* (2006) 4:104–13. doi: 10.1111/j.1476-5810.2006.00098.x

34. Michishita M, Uto T, Nakazawa R, Yoshimura H, Ogihara K, Naya Y, et al. Antitumor effect of bevacizumab in a xenograft model of canine hemangiopericytoma. *J Pharmacol Sci* (2013) 121:339–42. doi: 10.1254/jphs.12268sc

35. Guertin DA, Sabatini DM. Defining the role of mTOR in cancer. *Cancer Cell* (2007) 12:9–22. doi: 10.1016/j.ccr.2007.05.008

36. Chia S, Gandhi S, Joy AA, Edwards S, Gorr M, Hopkins S, et al. Novel agents and associated toxicities of inhibitors of the pi3k/Akt/mtor pathway for the treatment of breast cancer. *Curr Oncol* (2015) 22:33–48. doi: 10.3747/co.22.2393

37. Hua H, Kong Q, Zhang H, Wang J, Luo T, Jiang Y. Targeting mTOR for cancer therapy. *J Hematol Oncol* (2019) 12:71. doi: 10.1186/s13045-019-0754-1

38. Xia P, Xu XY. PI3K/Akt/mTOR signaling pathway in cancer stem cells: from basic research to clinical application. *Am J Cancer Res* (2015) 5:1602–9.

39. Zhou J, Wulfkuhle J, Zhang H, Gu P, Yang Y, Deng J, et al. Activation of the PTEN/mTOR/STAT3 pathway in breast cancer stem-like cells is required for viability and maintenance. *Proc Natl Acad Sci U.S.A.* (2007) 104:16158–63. doi: 10.1073/pnas.0702596104

40. Dubrovska A, Kim S, Salamone RJ, Walker JR, Maira SM, García-Echeverría C, et al. The role of PTEN/Akt/PI3K signaling in the maintenance and viability of prostate cancer stem-like cell populations. *Proc Natl Acad Sci U S A.* (2009) 106:268–73. doi: 10.1073/pnas.0810956106

41. Chen J, Shao R, Li F, Monteiro M, Liu JP, Xu ZP, et al. PI3K/Akt/mTOR pathway dual inhibitor BEZ235 suppresses the stemness of colon cancer stem cells. *Clin Exp Pharmacol Physiol* (2015) 42:1317–26. doi: 10.1111/1440-1681.12493

42. Bahmad HF, Mouhieddine TH, Chalhoub RM, Assi S, Araji T, Chamaa F, et al. The Akt/mTOR pathway in cancer stem/progenitor cells is a potential therapeutic target for glioblastoma and neuroblastoma. *Oncotarget* (2018) 9:33549–61. doi: 10.18632/oncotarget.26088

43. Andrade NP, Warner KA, Zhang Z, Pearson AT, Mantesso A, Guimaraēs DM, et al. Survival of salivary gland cancer stem cells requires mTOR signaling. *Cell Death Dis* (2021) 12:108. doi: 10.1038/s41419-021-03391-7

44. Huang EH, Hynes MJ, Zhang T, Ginestier C, Dontu G, Appelman H, et al. Aldehyde dehydrogenase 1 is a marker for normal and malignant human colonic stem cells (SC) and tracks SC overpopulation during colon tumorigenesis. *Cancer Res* (2009) 69:3382–9. doi: 10.1158/0008-5472.CAN-08-4418

45. Porta C, Paglino C, Mosca A. Targeting PI3K/Akt/mTOR signaling in cancer. *Front Oncol* (2014) 4:64. doi: 10.3389/fonc.2014.00064

46. Zhu Y, Zhang X, Liu Y, Zhang S, Liu J, Ma Y, et al. Antitumor effect of the mTOR inhibitor everolimus in combination with trastuzumab on human breast cancer stem cells *in vitro* and in vivo. *Tumour Biol* (2012) 33:1349–62. doi: 10.1007/s13277-012-0383-6

47. Liu Y, Zhang X, Liu J, Hou G, Zhang S, Zhang J. Everolimus in combination with letrozole inhibit human breast cancer MCF-7/Aro stem cells *via* PI3K/mTOR pathway: an experimental study. *Tumour Biol* (2014) 35:1275–86. doi: 10.1007/s13277-013-1170-8

48. Song CW, Lee H, Dings RP, Williams B, Powers J, Santos TD, et al. Metformin kills and radiosensitizes cancer cells and preferentially kills cancer stem cells. *Sci Rep* (2012) 2:362. doi: 10.1038/srep00362

49. Lee H, Park HJ, Park CS, Oh ET, Choi BH, Williams B, et al. Response of breast cancer cells and cancer stem cells to metformin and hyperthermia alone or combined. *PloS One* (2014) 9:e87979. doi: 10.1371/journal.pone.0087979

50. Murai A, Abou Asa S, Kodama A, Sakai H, Hirata A, Yanai T. Immunohistochemical analysis of the Akt/mTOR/4E-BP1 signalling pathway in canine haemangiomas and haemangiosarcomas. *J Comp Pathol* (2012) 147:430–40. doi: 10.1016/j.jcpa.2012.05.002

51. Delgado L, Gärtner F, Dias Pereira P. Activation of mammalian target of rapamycin in canine mammary carcinomas: an immunohistochemical study. *J Comp Pathol* (2015) 152:138–44. doi: 10.1016/j.jcpa.2014.12.004

52. Rivera-Calderón LG, Fonseca-Alves CE, Kobayashi PE, Carvalho M, Vasconcelos RO, Laufer-Amorim R. P-mTOR, p-4EBP-1 and eIF4E expression in canine prostatic carcinoma. *Res Vet Sci* (2019) 122:86–92. doi: 10.1016/j.rvsc.2018.11.006

53. Sanz Ressel BL, Massone AR, Barbeito CG. Immunohistochemical expression of selected phosphoproteins of the mTOR signalling pathway in canine cutaneous squamous cell carcinoma. *Vet J* (2019) 245:41–8. doi: 10.1016/j.tvjl.2018.12.024

54. Mathew R, Sajitha IS, Balakrishnan-Nair DK, Nair SS, Kariyil BJ, Abraham MJ, et al. Histological and immunohistochemical evaluation of phosphorylated mechanistic target of rapamycin in canine skin tumours. *J Comp Pathol* (2021) 184:60–4. doi: 10.1016/j.jcpa.2021.02.001

55. Gordon IK, Ye F, Kent MS. Evaluation of the mammalian target of rapamycin pathway and the effect of rapamycin on target expression and cellular proliferation in osteosarcoma cells from dogs. *Am J Vet Res* (2008) 69:1079–84. doi: 10.2460/ajvr.69.8.1079

56. Kent MS, Collins CJ, Ye F. Activation of the AKT and mammalian target of rapamycin pathways and the inhibitory effects of rapamycin on those pathways in canine malignant melanoma cell lines. *Am J Vet Res* (2009) 70:263–9. doi: 10.2460/ajvr.70.2.263

57. Chen YT, Tan KA, Pang LY, Argyle DJ. The class I PI3K/Akt pathway is critical for cancer cell survival in dogs and offers an opportunity for therapeutic intervention. *BMC Vet Res* (2012) 8:73. doi: 10.1186/1746-6148-8-73

58. Pyuen AA, Meuten T, Rose BJ, Thamm DH. *In vitro* effects of PI3K/mTOR inhibition in canine hemangiosarcoma. *PloS One* (2018) 13:e0200634. doi: 10.1371/journal.pone.0200634

59. Barbieri F, Thellung S, Ratto A, Carra E, Marini V, Fucile C, et al. *In vitro* and *in vivo* antiproliferative activity of metformin on stem-like cells isolated from spontaneous canine mammary carcinomas: translational implications for human tumors. *BMC Cancer.* (2015) 15:228. doi: 10.1186/s12885-015-1235-8

60. Saeki K, Watanabe M, Tsuboi M, Sugano S, Yoshitake R, Tanaka Y, et al. Anti-tumour effect of metformin in canine mammary gland tumour cells. *Vet J* (2015) 205:297–304. doi: 10.1016/j.tvjl.2015.04.026

61. Wei BR, Michael HT, Halsey CH, Peer CJ, Adhikari A, Dwyer JE, et al. Synergistic targeted inhibition of MEK and dual PI3K/mTOR diminishes viability and inhibits tumor growth of canine melanoma underscoring its utility as a preclinical model for human mucosal melanoma. *Pigment Cell Melanoma Res* (2016) 29:643–55. doi: 10.1111/pcmr.12512

62. Merchant AA, Matsui W. Targeting hedgehog–a cancer stem cell pathway. *Clin Cancer Res* (2010) 16:3130–40. doi: 10.1158/1078-0432.CCR-09-2846

63. Cochrane CR, Szczepny A, Watkins DN, Cain JE. Hedgehog signaling in the maintenance of cancer stem cells. *Cancers (Basel).* (2015) 7:1554–85. doi: 10.3390/cancers7030851

64. Liu S, Dontu G, Mantle ID, Patel S, Ahn NS, Jackson KW, et al. Hedgehog signaling and bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. *Cancer Res* (2006) 66:6063–71. doi: 10.1158/0008-5472.CAN-06-0054

65. Tanaka H, Nakamura M, Kameda C, Kubo M, Sato N, Kuroki S, et al. The hedgehog signaling pathway plays an essential role in maintaining the CD44+CD24-/low subpopulation and the side population of breast cancer cells. *Anticancer Res* (2009) 29:2147–57.

66. Bhateja P, Cherian M, Majumder S, Ramaswamy B. The hedgehog signaling pathway: A viable target in breast cancer? *Cancers (Basel)* (2019) 11(8):1126. doi: 10.3390/cancers11081126

67. Gustafson TL, Kitchell BE, Biller B. Hedgehog signaling is activated in canine transitional cell carcinoma and contributes to cell proliferation and survival. *Vet Comp Oncol* (2017) 15:174–83. doi: 10.1111/vco.12149

68. Baldanza VE, Rogic A, Yan W, Levine CB, Levine RA, Miller AD, et al. Evaluation of canonical hedgehog signaling pathway inhibition in canine osteosarcoma. *PloS One* (2020) 15:e0231762. doi: 10.1371/journal.pone.0231762

69. Nam A, Song WJ, An JH, Rebhun RB, Youn HY, Seo KW. Expression of the hedgehog signalling pathway and the effect of inhibition at the level of smoothened in canine osteosarcoma cell lines. *Vet Comp Oncol* (2022) 20:778–87. doi: 10.1111/vco.12828

70. Manasanch EE, Orlowski RZ. Proteasome inhibitors in cancer therapy. *Nat Rev Clin Oncol* (2017) 14:417–33. doi: 10.1038/nrclinonc.2016.206

71. Richardson PG, Sonneveld P, Schuster MW, Irwin D, Stadtmauer EA, Facon T, et al. Assessment of proteasome inhibition for extending remissions (APEX) investigators. bortezomib or high-dose dexamethasone for relapsed multiple myeloma. *N Engl J Med* (2005) 352:2487–98. doi: 10.1056/NEJMoa043445

72. Raedler L. Velcade (Bortezomib) receives 2 new FDA indications: For retreatment of patients with multiple myeloma and for first-line treatment of patients with mantle-cell lymphoma. *Am Health Drug Benefits.* (2015) 8:135–40.

73. Yang CH, Gonzalez-Angulo AM, Reuben JM, Booser DJ, Pusztai L, Krishnamurthy S, et al. Bortezomib (VELCADE) in metastatic breast cancer: pharmacodynamics, biological effects, and prediction of clinical benefits. *Ann Oncol* (2006) 17:813–7. doi: 10.1093/annonc/mdj131

74. Meißner T, Mark A, Williams C, Berdel WE, Wiebe S, Kerkhoff A, et al. Metastatic triple-negative breast cancer patient with TP53 tumor mutation experienced 11 months progression-free survival on bortezomib monotherapy without adverse events after ending standard treatments with grade 3 adverse events. *Cold Spring Harb Mol Case Stud* (2017) 3:a001677. doi: 10.1101/mcs.a001677

75. Shen S, Du XJ, Liu J, Sun R, Zhu YH, Wang J. Delivery of bortezomib with nanoparticles for basal-like triple-negative breast cancer therapy. *J Control Release.* (2015) 208:14–24. doi: 10.1016/j.jconrel.2014.12.043

76. Giuffrida R, Adamo L, Iannolo G, Vicari L, Giuffrida D, Eramo A, et al. Resistance of papillary thyroid cancer stem cells to chemotherapy. *Oncol Lett* (2016) 12:687–91. doi: 10.3892/ol.2016.4666

77. Thudi NK, Shu ST, Martin CK, Lanigan LG, Nadella MV, Van Bokhoven A, et al. Development of a brain metastatic canine prostate cancer cell line. *Prostate* (2011) 71:1251–63. doi: 10.1002/pros.21341

78. Kojima K, Fujino Y, Goto-Koshino Y, Ohno K, Tsujimoto H. Analyses on activation of NF-κB and effect of bortezomib in canine neoplastic lymphoid cell lines. *J Vet Med Sci* (2013) 75:727–31. doi: 10.1292/jvms.12-0168

79. Ito K, Kobayashi M, Kuroki S, Sasaki Y, Iwata T, Mori K, et al. The proteasome inhibitor bortezomib inhibits the growth of canine malignant melanoma cells *in vitro* and *in vivo*. *Vet J* (2013) 198:577–82. doi: 10.1016/j.tvjl.2013.08.003

80. Patatsos K, Shekhar TM, Hawkins CJ. Pre-clinical evaluation of proteasome inhibitors for canine and human osteosarcoma. *Vet Comp Oncol* (2018) 16:544–53. doi: 10.1111/vco.12413

81. Plaks V, Kong N, Werb Z. The cancer stem cell niche: how essential is the niche in regulating stemness of tumor cells? *Cell Stem Cell* (2015) 16:225–38. doi: 10.1016/j.stem.2015.02.015

82. Yang F, Xu J, Tang L, Guan X. Breast cancer stem cell: the roles and therapeutic implications. *Cell Mol Life Sci* (2017) 74:951–66. doi: 10.1007/s00018-016-2334-2337

83. Melzer C, von der Ohe J, Lehnert H, Ungefroren H, Hass R. Cancer stem cell niche models and contribution by mesenchymal stroma/stem cells. *Mol Cancer.* (2017) 16:28. doi: 10.1186/s12943-017-0595-x

84. Sachs N, de Ligt J, Kopper O, Gogola E, Bounova G, Weeber F, et al. A living biobank of breast cancer organoids captures disease heterogeneity. *Cell* (2018) 172:373–386.e10. doi: 10.1016/j.cell.2017.11.010

85. Lee PJ, Ho CC, Ho H, Chen WJ, Lin CH, Lai YH, et al. Tumor microenvironment-based screening repurposes drugs targeting cancer stem cells and cancer-associated fibroblasts. *Theranostics* (2021) 11:9667–86. doi: 10.7150/thno.62676

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 28 September 2023
DOI 10.3389/fonc.2023.1137966

OPEN ACCESS

EDITED BY
Yusuke Suenaga,
Chiba Cancer Center, Japan

REVIEWED BY
Soumya Basu,
Dr. D. Y. Patil Biotechnology & Bioinformatics Institute, India
Kazuki Takahashi,
The University of Tokyo, Japan

*CORRESPONDENCE
Hiroshi Haeno
haeno@rs.tus.ac.jp

RECEIVED 05 January 2023
ACCEPTED 07 September 2023
PUBLISHED 28 September 2023

CITATION
Yu Q, Kobayashi SS and Haeno H (2023) Mathematical analysis identifies the optimal treatment strategy for epidermal growth factor receptor-mutated non-small cell lung cancer. *Front. Oncol.* 13:1137966. doi: 10.3389/fonc.2023.1137966

Mathematical analysis identifies the optimal treatment strategy for epidermal growth factor receptor-mutated non-small cell lung cancer

Qian Yu[1], Susumu S. Kobayashi[2] and Hiroshi Haeno[3]*

[1]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Science, The University of Tokyo, Kashiwa, Japan, [2]Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States, [3]Research Institute for Biomedical Sciences, Tokyo University of Science, Noda, Japan

Introduction: In Asians, more than half of non-small cell lung cancers (NSCLC) are induced by epidermal growth factor receptor (EGFR) mutations. Although patients carrying EGFR driver mutations display a good initial response to EGFR-Tyrosine Kinase Inhibitors (EGFR-TKIs), additional mutations provoke drug resistance. Hence, predicting tumor dynamics before treatment initiation and formulating a reasonable treatment schedule is an urgent challenge.

Methods: To overcome this problem, we constructed a mathematical model based on clinical observations and investigated the optimal schedules for EGFR-TKI therapy.

Results: Based on published data on cell growth rates under different drugs, we found that using osimertinib that are efficient for secondary resistant cells as the first-line drug is beneficial in monotherapy, which is consistent with published clinical statistical data. Moreover, we identified the existence of a suitable drug-switching time; that is, changing drugs too early or too late was not helpful. Furthermore, we demonstrate that osimertinib combined with erlotinib or gefitinib as first-line treatment, has the potential for clinical application. Finally, we examined the relationship between the initial ratio of resistant cells and final cell number under different treatment conditions, and summarized it into a therapy suggestion map. By performing parameter sensitivity analysis, we identified the condition where osimertinib-first therapy was recommended as the optimal treatment option.

Discussion: This study for the first time theoretically showed the optimal treatment strategies based on the known information in NSCLC. Our framework can be applied to other types of cancer in the future.

KEYWORDS

computational modeling, drug resistance, cancer evolution, lung cancer, optimal treatment strategy

Introduction

Among all the cancer types, lung cancer causes the highest number of cancer-related deaths. 26% of cancer-related deaths in males and 25% in females are induced by lung cancer (1). In Asians, 53% of non-small cell lung cancer (NSCLC) progression is induced by epidermal growth factor receptor (EGFR) mutations, such as the L858R mutation, exon 19 deletion, and exon 20 insertion (2). Besides, EGFR has been recognized as an oncogenic driver of NSCLC, making it increasingly important in the era of precision medicine for lung cancer (3).

EGFR belongs to the receptor tyrosine kinase (RTK) family. EGFR is activated by various ligands in the extracellular environment and transmits cellular responses to mediate many cellular activities, including cell proliferation, survival, growth, and development. It is expressed in many organs, with its abnormal expression associated with a variety of cancers. EGFR has an extracellular ligand-binding domain, hydrophobic transmembrane domain, and cytoplasmic tyrosine kinase domain. The driver mutations in EGFR associated with cancers are concentrated in the tyrosine kinase domain, forming exons 18–21 (4–7). More than 200 types of EGFR mutations have been identified, but the most common types are exon-19 deletion and the L858R mutation in exon 21 (8, 9). Approximately 44% of EGFR-mutated patients harbor exon-19 deletion, and 31% have the L858R mutation (10).

Although EGFR was first identified in 1977, EGFR-targeted antitumor drugs were first reported in 1994 (11). After the first report of EGFR-targeted therapy, first-generation EGFR-Tyrosine Kinase Inhibitors (EGFR-TKIs) were not approved until 2004 (12). Subsequently, the second-generation EGFR-TKI, afatinib, was approved in 2014. First- and second-generation EGFR-TKIs are effective in most cases of lung cancer harboring EGFR driver mutations (13–16). However, acquiring mutations, such as the T790M mutation, causes drug resistance and induces treatment failure (17, 18). In 2015, the third-generation EGFR-TKI (osimertinib), which inhibits both driver mutations and the T790M mutation, was approved as a second-line drug for patients with EGFR mutations (19–22). Although osimertinib is clinically effective, the emergence of additional mutations, such as the C797S mutation, induces resistance to osimertinib (23–25). Clinical observations suggest that optimized treatment schedules can help patients achieve better therapeutic effects (26–28). Thus, predicting resistance evolution and making reasonable treatment schedules in advance are necessary to delay the appearance of drug resistance and prolong the time of recurrence. However, even with knowledge of medical and genetic information in the early stage, such as tumor size and the proportion of different genotypes, it is still difficult to simulate the future development of tumors using traditional biological techniques alone.

Mathematical modeling is an approach for simulating realistic problems using mathematics and computational algorithms. This can offer a better understanding of how tumors evolve during treatment, which can be visualized *in vivo*. Thus, it can help us predict tumor dynamics under certain treatment schedules, compare different treatments, and even suggest optimal treatment strategies. Many studies have demonstrated the capability of mathematical modeling in cancer-related research (29–33). For example, Diaz Jr. et al. developed a mathematical model of cell kinetics during chemotherapy to predict the emergence of resistant genotypes in colorectal cancer (30). Castagnino et al. established a mathematical model of a genetic network to identify novel molecular targets for the treatment of colorectal cancer (34). In this way, we decided to employ mathematical modeling to predict tumor evolution and direct reasonable treatment schedules for lung cancer patients harboring EGFR mutations.

In this study, we establish a novel mathematical model of lung cancer evolution under EGFR-targeted therapy based on clinical observations. Parameter values in the model are estimated from published literature, and the results are confirmed using clinical observations. Moreover, we examine the relationship between the timing of switching drugs and the final number of cells in the tumor. Furthermore, we compare the combinatorial use of EGFR-TKIs to their sequential use to test their potential for clinical application. Finally, we investigate how intratumoral heterogeneity at the initial time of therapy affects treatment outcomes. The simulation results are comprehensively tested by parameter sensitivity analysis in order to identify the condition where each treatment strategy becomes the best option. Our framework is expected to be capable of suggesting reasonable individualized medicine for EGFR-mutated NSCLC.

Materials and methods

Mathematical model

Based on clinical observations (35, 36), we established a mathematical model that describes the dynamics of the four types of EGFR-mutated cells under two types of EGFR-TKIs (Figure 1). There are two different types of EGFR-TKIs in the model: one is "DrugA," representing the first- or second-generation EGFR-TKIs named gefitinib, erlotinib or afatinib; and the other is "DrugB," representing osimertinib. Four cancer cell types are denoted by "Type-W," "Type-X," "Type-Y," and "Type-Z". Type-W is sensitive to both DrugA and DrugB, indicating cancer cells with driver EGFR mutations, such as L858R mutations or exon-19 deletion. Type-X cells are resistant to DrugA but sensitive to DrugB, indicating cells with T790M mutations. Type-Y is sensitive to DrugA but resistant to DrugB, indicating cells with C797S mutations. Type-Z is resistant to both DrugA and DrugB. Summarizing the relationship between drugs and cells, under DrugA treatment, Type-W and Type-Y will decrease, but Type-X and Type-Z will increase, whereas under DrugB treatment, Type-W and Type-X will decrease, but Type-Y and Type-Z will increase. According to published clinical studies (37–39), when DrugA was administered as first-line treatment, Type-X eventually became dominant in the tumor. After switching from DrugA to DrugB, the frequency of Type-X decreased, and only Type-Z continued to grow and dominate the tumor. However, when using DrugB as the first-line treatment, Type-Y will replace Type-W as the major population. After

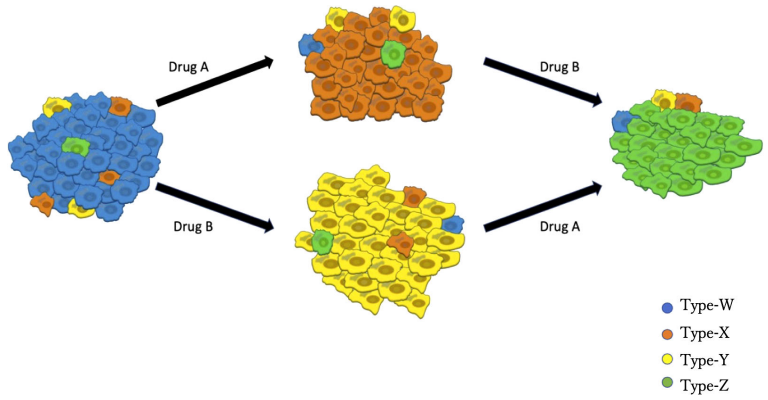


FIGURE 1
Illustration of the Model. Blue cells represent EGFR-TKI sensitive genotypes (for example, $EGFR^{L858R}$ and $EGFR^{del\text{-}19}$), orange cells represent osimertinib sensitive genotypes (such as $EGFR^{L858R\text{-}T790M}$ or $EGFR^{del\text{-}19\text{-}T790M}$), yellow cells represent osimertinib-only resistant genotypes (for instance, $EGFR^{L858R\text{-}C797S}$ or $EGFR^{del\text{-}19\text{-}C797S}$), green cells represent all EGFR-TKI resistant genotypes (like $EGFR^{L858R\text{-}T790M\text{-}C797S}$ and $EGFR^{del\text{-}19\text{-}T790M\text{-}C797S}$). DrugA involved the first- and second-generation EGFR-TKIs (erlotinib, gefitinib, and afatinib), and DrugB is osimertinib.

switching from DrugB to DrugA, only Type-Z became the donor population in the tumor.

In this mathematical model, we assumed that each cell type itself increases in number by cell division and mutates into a resistant type at a low mutation rate. We did not consider back mutations that were resistant to the sensitive cells. Moreover, according to the purpose of this study, we only focused on mutations related to drug resistance and assumed that other mutations are neutral and do not affect the growth kinetics. Then the dynamics of Type-W, Type-X, Type-Y, and Type-Z are given by Eqs. (1)

$$\frac{dw}{dt} = aw \tag{1.1}$$

$$\frac{dx}{dt} = gw + bx \tag{1.2}$$

$$\frac{dy}{dt} = hw + cy \tag{1.3}$$

$$\frac{dz}{dt} = kw + px + qy + fz \tag{1.4}$$

Here, the variables w, x, y, and z represent the cell numbers of Type-W, -X, -Y, and -Z, respectively. Parameters a, b, c, and f are the growth rates of Type-W, Type-X, Type-Y, and Type-Z, respectively, and g, h, k, p, and q are the mutation rates from type-W to Type-X, Type-W to Type-Y, Type-W to Type-Z, Type-X to Type-Z, and Type-Y to Type-Z, respectively. Because no other cell type can mutate into Type-W, the number of Type-W cells is affected by its kinetics only. However, Type-W will mutate into Type-X, Type-Y, and Type-Z.

Solution of equations

The Eqs. (1) can be solved and given by Eqs. (2)

$$w(t) = W_0 e^{at} \tag{2.1}$$

$$x(t) = X_0 e^{bt} + \frac{g}{a-b} W_0 e^{(at-bt)e^{bt}} \tag{2.2}$$

$$y(t) = Y_0 e^{ct} + \frac{h}{a-c} W_0 e^{(at-ct)e^{ct}} \tag{2.3}$$

$$\begin{aligned} z(t) = {} & Z_0 e^{ft} - \frac{e^{ft}}{ab+ac-bc-a^2}\left[\frac{W_0}{a-f} e^{(at-ft)}(a^2k - abk - ack + bck + agp + \right. \\ & ahq - cgp - bhq) \\ & + \frac{X_0}{b-f} e^{(bt-ft)}(a^2p - abp - acp + bcp) \\ & \left. + \frac{Y_0}{c-f} e^{(ct-ft)}(a^2q - abq - acq + bcq)\right] \end{aligned} \tag{2.4}$$

The equations describe the cell number of each type over time (t) during therapy. W_0, X_0, Y_0, and Z_0 represent the initial cell numbers of Type-W, -X, -Y, and Type-Z in the tumor. Please refer to Table 1 for the meaning of each letter in the model.

Parameter evaluation

The parameter values were obtained from the published data (Table 2) (40–42). Since we obtained growth parameters under erlotinib and osimertinib treatments, we regarded these drugs as representative of DrugA and DrugB, respectively. Since Starrett et al. (41) reported that first-line therapy with erlotinib and osimertinib delayed the emergence of secondary mutations in untreated EGFR-mutated NSCLC, thus, for combination therapy, we defined a combinatorial regimen of erlotinib plus osimertinib as DrugC. Based on genome-editing cell line experiments (40), we adopted the growth rate of EGFR-L858R mutated cells for Type-W as -0.17 [/day] under DrugA (a_A) and -0.32 [/day] under DrugB (a_B). Note that the subscript of each growth rate represents the condition of drugs, *i.e.*, a_A represents the growth rate of Type-W under DrugA. From Starrett et al. (41), we adopted the growth rate of EGFR-L858R/C797S mutated cells for Type-Y as -0.13 [/day] under DrugA (c_A) and 0.024 [/day] under DrugB (c_B). The growth

TABLE 1 Parameter notation in the mathematical model.

Cell Type	Cell Number	Initial Cell Number	Growth Rate	Mutation Rate
Type-W	w	W_0	a	/
Type-X	x	X_0	b	g
Type-Y	y	Y_0	c	h
Type-Z	z	Z_0	f	k (from W), p (from X), q (from Y)

rate of EGFR-L858R/T790M/C797S mutated cells for Type-Z is 0.022 [/day] under DrugB (f_B), and the growth rate of EGFR-L858R/C797S mutated cells for Type-Y is -0.0335 [/day] under DrugC (c_C). According to Gunnarsson et al. (42), we set all the mutation rate as 10^{-7} [/day] (g, h, k, p and q).

Based on the adopted parameters, we assumed all the other parameter values. Because the growth rate of Type-Y under DrugC is approximately 26% of that under DrugA (-0.0335/-0.13), we calculated the growth rate of Type-W in DrugC as 26% of that under DrugA, which is -0.064 [/day] (a_C). We assume the growth rate of Type-X under DrugC is same as Type-Y, which is -0.0335 [/day] (b_C). Moreover, we assumed the growth rate of Type-X under DrugB as -0.15 [/day] (b_B), which is smaller than that of Type-Y under DrugA (c_A) based on clinical observation (20, 21) where the first-line treatment by DrugB showed better prognosis than that by DrugA. Based on the same reason, we assumed the growth rate of Type-X cell under DrugA as 0.045 [/day] (b_A). Finally, since Type-Z is resistant to both DrugA and DrugB, we assume its growth rates under DrugA and DrugC are same as that under DrugB effect, which is -0.022 [/day] (f_A and f_C).

As for the initial condition of simulations, the initial total cell number of the tumor is set to be 10^9, and the standard initial cell number of Type-X (X_0), Type-Y (Y_0), Type-Z (Z_0), and Type-W (W_0), is 10^4, 10^4, 10, and the rest component, respectively. The initial total cell number is set to be 10^9 because the diameter of a tumor at this point is about 1cm and a detectable size clinically. About the drug switching time in monotherapy, we set day-307 (sta=307) under DrugA-first therapy and day-567 (stb=567) under DrugB-first therapy based on clinical statistic data of the median Progression-Free Survival (mPFS) (20). The whole treatment time is assumed to be 1000 days (T=1000) in our simulation because 1000 days is long enough to compare the treatment options.

Computational simulation

We used Python (version 3.8.8) to simulate our model. We did time course simulation of different therapies for checking whether our model can express the progression of tumor as clinical observation. Then, we simulated the relationship between drug switching time and final cell number to theoretically figured out the possible affects that could influence therapy effects. Moreover, we simulated how parameters affected final cell number in different therapies. All the codes for simulations can be found in our GitHub open repository: https://github.com/yuqianxigua/EGFR-TKIs-therapy.

Results

Tumor evolution under different treatments

We simulated Eqs. (2) to predict tumor progression under different treatments, including monotherapy and combination therapy. When DrugA was used as first-line treatment (Figures 2A, D), Type-W and Type-Y decreased, whereas Type-X and Type-Z increased. Once Type-X became the dominant population, the tumor started to grow again and would no longer be sensitive to the first treatment. We then changed this drug to DrugB. In this study, we set the drug-switching time at day 307 (t= 307) based on clinical observations of the median Progression-Free Survival (mPFS) of erlotinib treatment (21). Under the second-line medication of DrugB, the growth of Type-W and Type-X was suppressed, but that of Type-Y and Type-Z increased. Finally, Type-Z became the major population. The simulation results

TABLE 2 Parameter values with different therapies.

Parameters / Drug	Growth Rate (/day)				Mutation Rate (/day)				
	a (L858R)	b (L858R/T790M)	c (L858R/C797S)	f (L858R/T790M/C797S)	g (W→X)	h (W→Y)	k (W→Z)	p (X→Z)	q (Y→Z)
A(erlotinib)	-0.17	0.045	-0.13	0.022	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}
B(osimertinib)	-0.32	-0.15	0.024	0.022	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}
C(e+0)	-0.064	-0.0335	-0.0335	0.022	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}

demonstrated a trend of tumor evolution under erlotinib-first treatment. At the end of our simulation period, set as day 1000 (t=1000), the cell number of the tumor was 2.12× 10^{12}. Next, we examined the DrugB-first treatment. The main population changed from Type-W to Type-Y and Type-Z (Figures 2B, E). During the treatment period, Type-W and Type-X decreased, whereas Type-Y and Type-Z increased. Once Type-Y became the main population, the tumor started to grow again and was no longer sensitive to DrugB. Herein, we switched drugs at day 567 (t=567) because the mPFS was approximately 567 days under the osimertinib treatment (21). When DrugA was used as the second-line treatment, Type-Y was suppressed, and Type-Z continued to grow and dominated the tumor. Compared with the presumed evolution (Figure 2E), our model profitably reflected the tumor progression of osimertinib-first treatment. In this treatment schedule, the tumor recurred at day 490, which was longer than that of erlotinib-first therapy. Additionally, at the end of our simulation period (t=1000), the total cell number was 1.09× 10^{12}, which was less than that of the DrugA-first treatment.

Furthermore, we investigated the outcomes of combination therapy (DrugC) by using DrugA and DrugB at the same time as first-line treatment (Figures 2C, F). When DrugC was applied as the

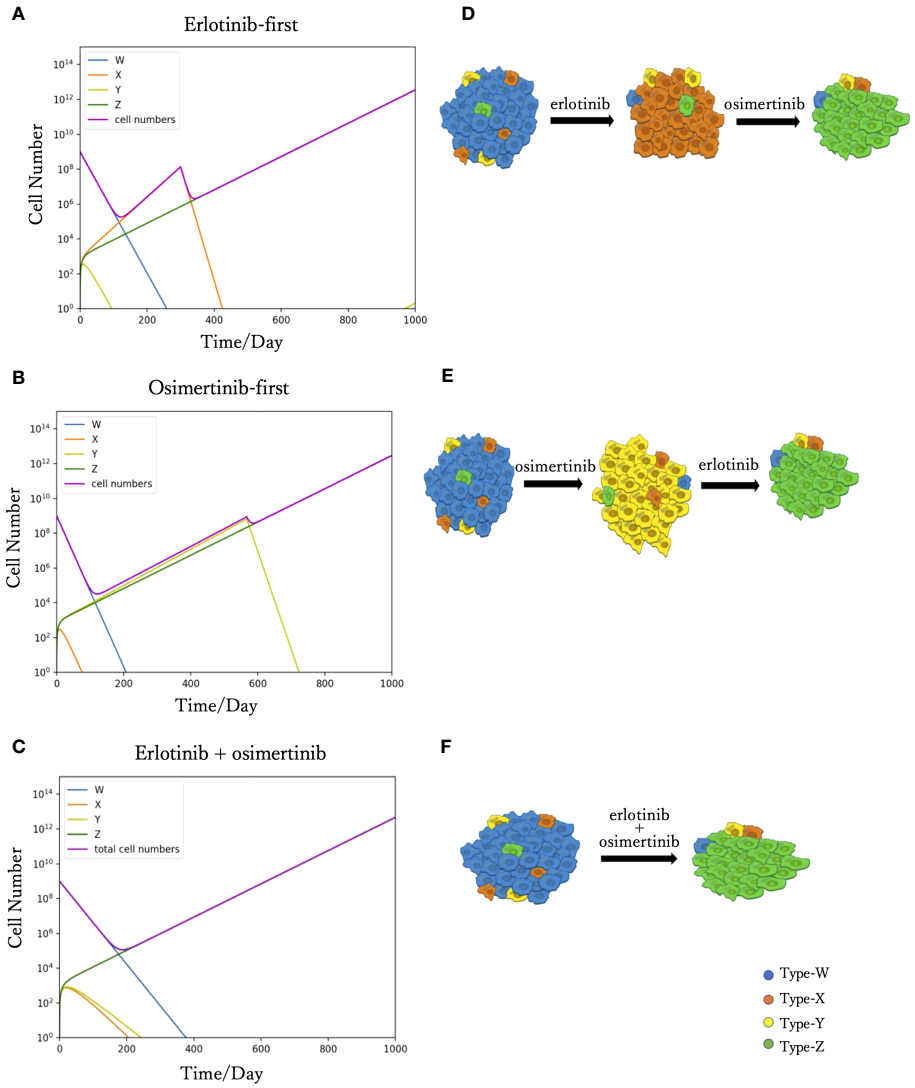


FIGURE 2
Time course simulation results of monotherapy and combination therapy. The results of the simulations are depicted in **(A–C)**. The x-axis is time, and the y-axis is the cell number. The blue, orange, yellow and green curves represent the dynamics of Type-W, -X, -Y, and -Z, respectively. The purple curve represents the total cell number. The expected tumor progression tendencies are depicted in **(D, E, F)**. The blue, orange, yellow, and green cells are Type-W, -X, -Y, and -Z, respectively. In the simulation of erlotinib-first **(A)**, the main population changed from Type-W to Type-X for a while. After changing erlotinib to osimertinib at day 307, Type-X decreased, and Type-Z became the dominant population in the end. This simulation result represents the tumor evolution tendency shown in **(D)**. The simulation product of osimertinib-first is shown in **(B)**. The tumor response to osimertinib increased in the beginning, but as Type-Y became the main population, osimertinib-resistance appeared. After the change to erlotinib at day 567, Type-Y decreased, and the tumor response to treatment increased again. However, Type-Z became the central population causing drug resistance. This simulation result represents the tumor evolution tendency shown in **(E)**. In the combination therapy **(C)**, the main population changed from Type-W to Type-Z. This result represents the tumor evolution tendency shown in **(F)**.

first-line treatment, Type-W, Type-X, and Type-Y decreased, and only Type-Z continued to increase because it was resistant to both DrugA and DrugB. Type-W is the main population in the early phase of treatment, and eventually, Type-Z replaced Type-W to become the main population in the tumor. During medication, neither X nor Type-Y dominated the population. At day 1000, the total cell number was 4.2×10^{12}.

Drug-switching time and final cell number

To examine the relationship between drug-switching time and the development of the total cell number of the tumor, we simulated the tumor dynamics and measured the total cell number at day 1000 with various drug-switching times (Figure 3). In the case of DrugA-first treatment (Figure 3A), the lowest final total number of cells was 2.0×10^{12}, while it was about 1.0×10^{12} in the case of DrugB-first treatment (Figure 3B). This implied that first-line treatment with DrugB displayed better treatment outcomes than DrugA-first treatment. Moreover, the total number of cells at day 1000 remained essentially the same in an appropriate range of drug-switching times under both DrugA- and DrugB-first treatments. This suggested the existence of an optimal drug-switch period, and it was not advisable to switch drugs too early or too late. Furthermore, comparing the suitable drug-switching time period for these two treatments, DrugB-first therapy had a broader range than DrugA-first. In the DrugB-first treatment, switching drugs from days 200 to 900 was acceptable (Figure 3B). However, in DrugA-first therapy, the suitable drug-switching time ranged from day 100 to day 450 (Figure 3A).

Cell initial proportion dependence

To investigate the effect of the initial proportion of different mutant cells on the final cell number, we simulated how the final cell number changes with the increase of mutant cell proportion in different treatment strategies (Figure 4). We explored the effect of only one resistant cell type at one time, keeping other conditions constant as the standard condition. For Type-X and Type-Y, we tested the change in initial proportion from 10^{-8} to 10^{-1}, and for Type-Z from 10^{-9} to 10^{-5}. With the increase of Type-X cell (Figures 4A–C), the final cell number did not change under DrugB-first therapy (Figure 4B) and combination therapy (DrugC) (Figure 4C) but increased in DrugA-first therapy (Figure 4A) once the initial proportion of Type-X exceeded 10^{-4}. Similarly, in Type-Y dependence simulations (Figures 4D–F), the final cell number increased only in DrugB-first therapy (Figure 4E) when the initial proportion of Type-Y became larger than 10^{-3}. In addition, in Type-Z dependence simulations (Figures 4G–I), the final cell number increased once the initial proportion of Type-Z cell became larger than 10^{-7} under all treatments. Within the range of initial cell proportion that did not cause an increase in the final cell number, DrugB-first therapy always showed the smallest number of total cells at day 1000.

Therapy selection map

In order to identify which treatment strategy is optimal in a given case, we compared the final total cell number in different treatments with the change of the initial Type-X and Type-Y cell proportion and summarized the results in a therapy selection map (Figure 5). In this simulation, we kept the initial number of Type-Z constant as 10. By comparing the final total cell number under these three treatment strategies in the different initial proportions of Type-X and Type-Y cells, we determined the best strategy by realizing the smallest cell number at day 1000. The simulations were performed in the same method as used in Figure 2. From this map, we noticed that DrugB-first therapy was the optimal choice when tumors harbored a low initial proportion of Type-Y cells. However, DrugA-first therapy could still be advisable if the initial proportion of Type-Y cells was more significant in the tumor cluster. Furthermore, this map indicated that when both Type-X and Y cells had a high initial proportion in the tumor cluster, combination therapy (DrugC) was the optimal choice.

Parameter sensitivity analysis

To investigate the parameter sensitivity, we analyzed how the total cell number at day 1000 changed with parameters under those

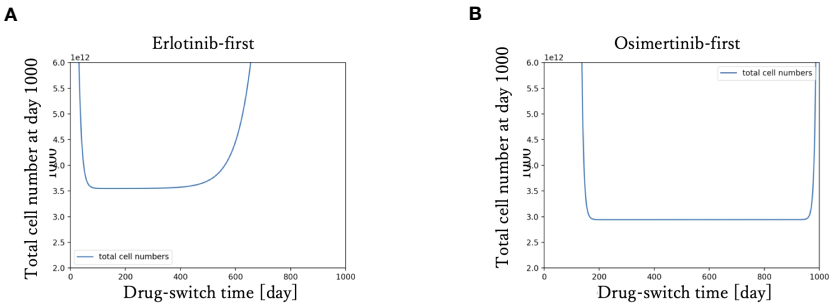


FIGURE 3
Drug switch time and final total cell number. The x-axis is drug-switch time, and the y-axis is the total cell number at day 1000. In panel **(A)**, the simulation result using erlotinib as a first-line treatment is shown. The lowest total cell number at day 1000 is approximately 2×10^{12}. In panel **(B)**, the case of first-line Osimertinib treatment is shown. The lowest total cell number at day 1000 is approximately 1×10^{12}.

10.3389/fonc.2023.1137966

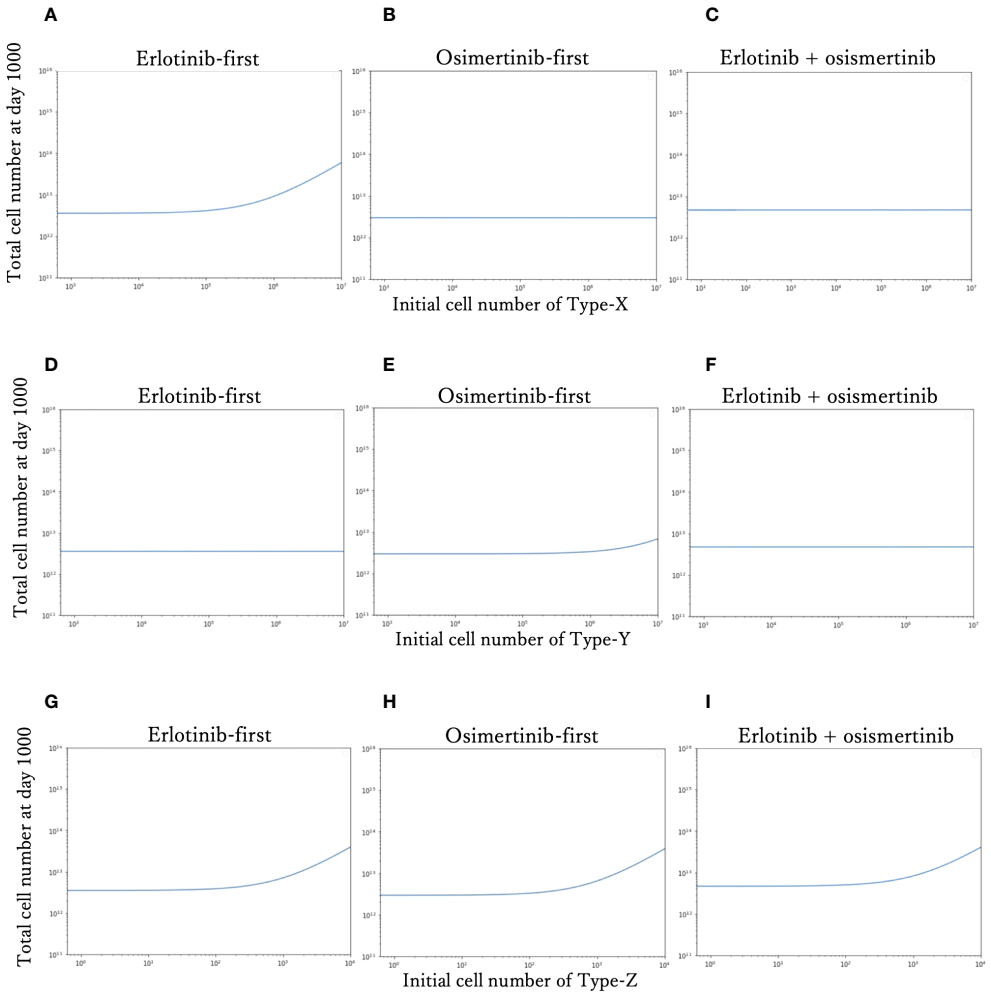


FIGURE 4
Relationship between the initial proportion and final cell number. The total cell number at day 1000 of each therapy with the different initial cell numbers of Type-X cells are shown in **(A–C)**. The final total cell number of each therapy with the different initial cell numbers of Type-Y cells are shown in **(D–F)**. The total cell number at day-1000 of each therapy with the different initial cell numbers of Type-Z cells are shown in **(G–I)**. The x-axis is initial proportion of mutation cells, the y-axis is the final total cell number.

three treatment strategies (Figure 6). In the analysis of the growth rate of Type-W cell (*a*), the final total cell number increased with the increase of a_A under DrugA-first therapy (Figure 6A1); the increase of a_B under DrugB-first therapy (Figure 6A2); and the increase of a_C under the combination therapy (Figure 6A3). As for the growth rate of Type-X cell (*b*) and Type-Y cell (*c*), they did not affect the final total cell number significantly in our simulated value range (Figures 6B, C). Moreover, about the growth rate of Type-Z cell (*f*), the final total cell number increased with f_C under the combination therapy (DrugC) (Figure 6D). Concerning the effect of mutation rates (Figure 6E–I), their influence was different based on therapy strategies. In DrugA-first therapy, the increase of g_A, h_A, k_A, p_A and q_A increased the final total cell number. Meanwhile, the increase of g_B, h_B, k_B, p_B and q_B increased it in DrugB-first therapy. In the combination therapy (DrugC), only k_C increased it (Figure 6).

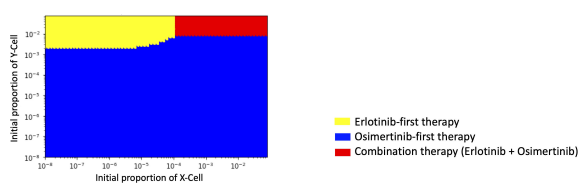


FIGURE 5
Initial Proportions of Type-X and -Y cells and treatment strategy selection. The x-axis is the initial ratio of Type-X, and the y-axis is the initial ratio of Type-Y. The yellow region means that osimertinib-first therapy is the optimal therapy, the blue region means erlotinib-first therapy is the optimal therapy, and the green region means combination therapy (erlotinib+osimertinib) is the optimal choice.

Parameter dependence on the therapy selection map

Since several parameter values were set by our own assumptions, we investigated how these values affected the optimal choice of treatment in detail (Figure 7). In this analysis, we changed one focused parameter value, made a therapy-selection

map as described in Figure 5, calculated the area of each strategy on the map and showed the area composition of each strategy at each parameter value. Especially, we investigated the dependence of growth rates of Type-W (a) and Type-Z (f) cell, and mutation rate from Type-W to Type-Z cell (k) under the three treatment strategies. As a result, the area where DrugB-first therapy exhibited superiority was large in the a_A, a_B, and a_C,-dependent analyses (Figures 7A–C). When the DrugB effect was weak against Type-W cell, the DrugA-first therapy became superior (Figure 7B). Moreover, when we changed growth rates of Type-Z under the three strategies, the DrugB-first therapy was the best option again except the cases where the growth rate of Type-Z under DrugA and DrugB was large (Figures 7D, E), and the growth rate of Type-Z under DrugB and DrugC was small (Figures 7E, F). Finally, changing the mutation rate under the three treatment strategies, DrugB-first therapy was the best option in most cases (Figures 7G–I). When the mutation rate (k) was small under DrugA and DrugC, and large under DrugB, DrugA-first or DrugC therapy became the best option (Figures 7G–I).

Discussion

In this study, we proposed a new mathematical model of EGFR-mutated NSCLC. First, our model successfully reproduced the process of tumor evolution under different treatment schedules, including monotherapy and combination therapy (Figure 2). In the erlotinib-first treatment (i.e., DrugA-first treatment), the drug was switched at day 307, while at day 576 it was switched in the osimertinib-first treatment (i.e., DrugB-first treatment). Next, we compared the effects of the two therapies. Our simulation results indicated that first-line osimertinib therapy was better than erlotinib. Within the same time period, for example, 1000 days in our study, osimertinib-first therapy resulted in a lower total cell number. Furthermore, the tumor recurred at nearly 500 days in osimertinib-first therapy compared to approximately 300 days in erlotinib-first therapy. This implied that first-line osimertinib therapy could suppress the growth of tumors more effectively than first-line erlotinib therapy, and could prolong the time of tumor recurrence. In the FLAURA project, clinical statistical data also revealed that EGFR-mutated NSCLC patients treated with osimertinib-first therapy had longer median mPFS (20, 21). This statistical study indicated the validity of our proposed model.Additionally, we noticed that in monotherapy, the total cell count was relatively low over a range of drug-switching times (Figure 3). This finding describes the existence of a suitable drug-switching phase, which suggests that it is not advisable to change the drug at a very early or late stage. In the suitable range of drug switching times, our simulation results showed that osimertinib-first therapy had a relatively lower total cell number than erlotinib-first therapy at day 1000. This result also indicates the potential of osimertinib as a first-line therapy in clinical applications. In

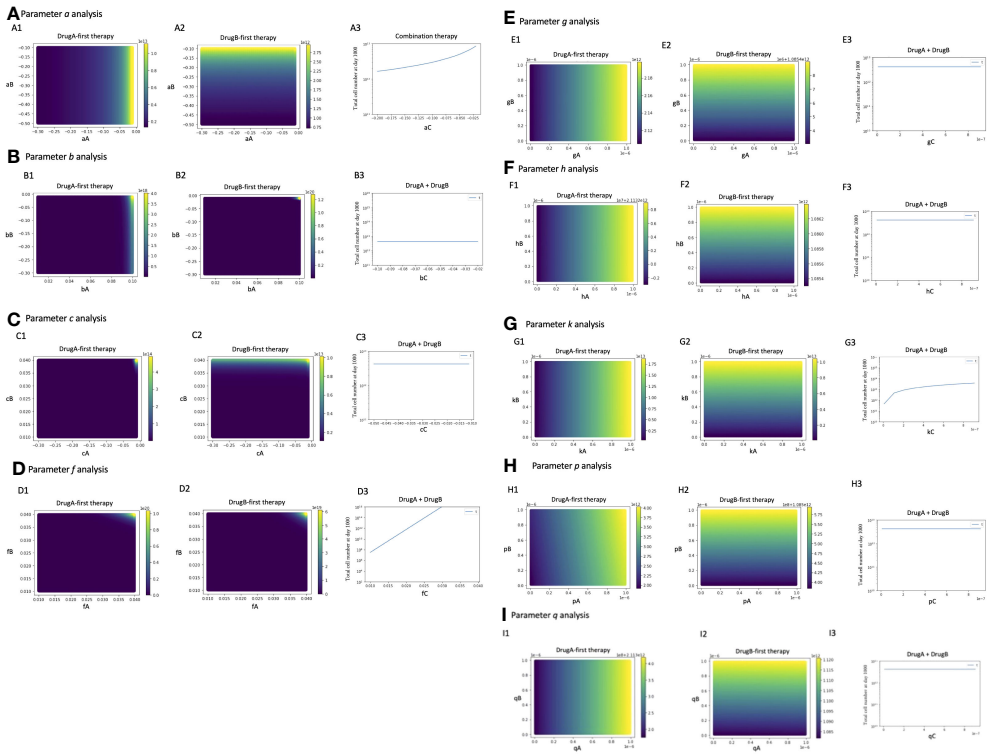


FIGURE 6
Parameter sensitivity analysis. Parameter dependence on the total total cell number at day 1000 under the three therapy strategies was analyzed. In monotherapy related analysis, the x- and y-axis are the parameters in the effect of DrugA and DrugB, and the color bar presented the final total cell number. In combination therapy, the x-axis is the parameter, while the y-axis is the final total cell number. The analysis of growth rates, a, b, c, f is showed in **(A–D)**, the mutation rates analysis is showed in **(E–I)**.

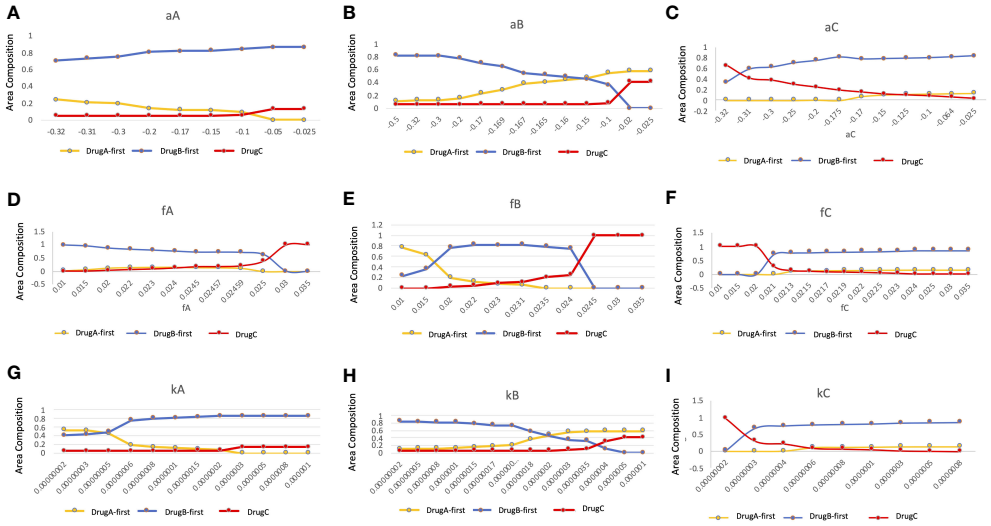


FIGURE 7
Parameter dependence on the area compositions of the three strategies in the optimal strategy map. We analyzed the composition of optimal therapies among DrugA-first therapy, DrugB-first therapy, and DrugC therapy in therapy selection map with the change of parameters. The x-axis is the parameter to be focused, the y-axis is the percentage of the area of the optimal therapies in the therapy selection map. The dependence of growth rates of Type-W **(A–C)**, and Type-Z cell **(D–F)**, and mutation rate from Type-W to Type-Z cell **(G–I)** under the three treatment strategies were tested. The yellow, blue, and red lines are DrugA-first therapy, DrugB-first therapy, and DrugC therapy, respectively.

addition, the appropriate drug-switching time range in osimertinib-first treatment was broader than that in erlotinib-first.

Furthermore, we explored the influence of tumor heterogeneity on therapeutic effects (Figure 4). By analyzing the relationship between the initial cell number and total cell number at the end of the tested time, we learned that the therapeutical effects depended on the initial ratio of resistant types in untreated tumors, and sensitive type did not affect it. According to the simulation results, when the initial ratio of Type-X exceeded the threshold, only the total cell number in the erlotinib-first therapy became large (Figure 4A). In the case of Type-Y, only osimertinib-first therapy resulted in large number of cells (Figure 4E). As for Type-Z, when its number became sufficiently large, the final total cell number developed rapidly in all the tested treatment schedules (Figures 4G–I). These results indicated that a high proportion of drug-resistant cells is associated with poor treatment efficacy. This conclusion suggests that if the tumor harbors a high ratio of Type-X, osimertinib-first is better than erlotinib-first. However, with a high initial ratio of Type-Y, erlotinib-first was better. Importantly, by combining this information, we for the first time theoretically revealed the relationship between the choice of treatment strategy and the initial proportion of Type-X and -Y cell (Figure 5). These findings indicated the advantage of first-line osimertinib treatment and revealed the influencing factors when determining treatment plans. Parameter sensitivity analysis about the total cell number and the best treatment choice confirmed the region where osimertinib-first therapy was superior to other options (Figures 6 and 7). Especially, we noticed that among the parameters, growth rate of Type-W and Type-Z cell and mutation rate from Type-W to Type-Z made a significantly change in the therapy selection map (Figure 7). Thess findings indicated the importance of suppressing all-drug-sensitive (Type-W) and all-drug-resistant (Type-Z) cells. This implied that during the treatment, not only the emergence of secondary resistant cells, but also the response of all-drug-resistant and -sensitive cells to drugs should be considered.

Furthermore, the simulation results showed that the combination of two types of EGFR-TKIs (erlotinib + osimertinib) as first-line therapy has the potential for clinical applications. Recently, several clinical studies have combined two types of EGFR-TKIs (first- and third-generation TKIs) as first-line treatment. In 2017, Wang et al. first reported the combination of erlotinib and osimertinib in patients with EGFR-mutated NSCLC patients (37). They illustrated the expediency of this type of treatment strategy. In addition, Rotow et al. applied gefitinib plus osimertinib as the first-line treatment for untreated patients with EGFR-mutated NSCLC (43). Their results showed the feasibility of conducting EGFR-TKI combination therapy, and survival analysis is in progress. In this study, we explored the combination of erlotinib and osimertinib as first-line therapy and explained the advantages of this method from a theoretical level. Especially, the drug response time with combination therapy was longer than that with monotherapy. Based on our simulation results, the recurrence time under combination therapy was longer than 500 days, whereas it was approximately 450 days in osimertinib-first treatment and 300 days in erlotinib-first therapy. This finding implied that the ability of combination therapy to prevent the emergence of acquired mutations and prolong the drug response time was even better than osimertinib-first treatment, which suggested its potential in clinical applications.

Based on the above, the versatility exhibited by the simulation results suggests that our model has the potential to be applied to simulate other similar cases in different cancer types. For further study, some clinical information about patients, such as age, sex, and the degree of malignancy of the tumor, may be considered in the parameter estimation. Thus, this model can be used to develop individual treatment schedules in the future.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

SSK and HH conceived the idea of the study. QY and HH developed simulation codes. QY conducted computational simulations. QY and HH contributed to the interpretation of the results. SSK and HH supervised the conduct of this study. All authors reviewed the manuscript draft and revised it. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by JSPS KAKENHI Grant Number JP22H04925 (PAGS) (HH), JP22J10101 (QY), and Boehringer-Ingelheim Japan (HH and SSK).

Conflict of interest

HH and SSK report research fund from Boehringer-Ingelheim Japan.

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

The reviewer KT declared a shared parent affiliation with the author QY to the handling editor at the time of review.

Publisher's note

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA: A Cancer J Clin* (2018) 68(1):7–30. doi: 10.3322/caac.21442

2. Saito M, Shiraishi K, Kunitoh H, Takenoshita S, Yokota J, Kohno T. Gene aberrations for precision medicine against lung adenocarcinoma. *Cancer Sci* (2016) 107 (6):713–20. doi: 10.1111/cas.12941

3. Liu X, Wang P, Zhang C, Ma Z. Epidermal growth factor receptor (EGFR): A rising star in the era of precision medicine of lung cancer. *Oncotarget* (2017) 8 (30):50209–20. doi: 10.18632/oncotarget.16854

4. Rossi S, D'Argento E, Basso M, Strippoli A, Dadduzio V, Cerchiaro E, et al. Different EGFR gene mutations in exon 18, 19 and 21 as prognostic and predictive markers in NSCLC: A single institution analysis. *Mol Diagn Ther* (2016) 20(1):55–63. doi: 10.1007/s40291-015-0176-x

5. Kobayashi Y, Togashi Y, Yatabe Y, Mizuuchi H, Jangchul P, Kondo C, et al. EGFR exon 18 mutations in lung cancer: molecular predictors of augmented sensitivity to afatinib or neratinib as compared with first- or third-generation TKIs. *Clin Cancer Res* (2015) 21(23):5305–13. doi: 10.1158/1078-0432.CCR-15-1046

6. Castellano GM, Aisner J, Burley SK, Vallat B, Yu HA, Pine SR, et al. A novel acquired exon 20 EGFR M766Q mutation in lung adenocarcinoma mediates osimertinib resistance but is sensitive to neratinib and poziotinib. *J Thorac Oncol* (2019) 14(11):1982–8. doi: 10.1016/j.jtho.2019.06.015

7. Yasuda H, Park E, Yun CH, Sng NJ, Lucena-Araujo AR, Yeo WL, et al. Structural, biochemical, and clinical characterization of epidermal growth factor receptor (EGFR) exon 20 insertion mutations in lung cancer. *Sci Transl Med* (2013) 5(216):216ra177. doi: 10.1126/scitranslmed.3007205

8. Castellanos E, Feld E, Horn L. Driven by mutations: the predictive value of mutation subtype in EGFR-mutated non-small cell lung cancer. *J Thorac Oncol* (2017) 12(4):612–23. doi: 10.1016/j.jtho.2016.12.014

9. O'Kane GM, Bradbury PA, Feld R, Leighl NB, Liu G, Pisters KM, et al. Uncommon EGFR mutations in advanced non-small cell lung cancer. *Lung Cancer* (2017) 109:137–44. doi: 10.1016/j.lungcan.2017.04.016

10. Costa DB. Kinase inhibitor-responsive genotypes in EGFR mutated lung adenocarcinomas: moving past common point mutations or indels into uncommon kinase domain duplications and rearrangements. *Transl Lung Cancer Res* (2016) 5 (3):331–7. doi: 10.21037/tlcr.2016.06.04

11. Ward WH, Cook PN, Slater AM, Davies DH, Holdgate GA, Green LR. Epidermal growth factor receptor tyrosine kinase. Investigation of catalytic mechanism, structure-based searching and discovery of a potent inhibitor. *Biochem Pharmacol* (1994) 48(4):659–66. doi: 10.1016/0006-2952(94)90042-6

12. Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, et al. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci U.S.A.* (2004) 101(36):13306–11. doi: 10.1073/pnas.0405220101

13. Rotow J, Bivona TG. Understanding and targeting resistance mechanisms in NSCLC. *Nat Rev Cancer* (2017) 17(11):637–58. doi: 10.1038/nrc.2017.84

14. Park K, Tan EH, O'Byrne K, Zhang L, Boyer M, Mok T, et al. Afatinib versus gefitinib as first-line treatment of patients with EGFR mutation-positive non-small-cell lung cancer (LUX-Lung 7): a phase 2B, open-label, randomised controlled trial. *Lancet Oncol* (2016) 17(5):577–89. doi: 10.1016/S1470-2045(16)30033-X

15. Park K, Wan-Teck Lim D, Okamoto I, Yang JC. First-line afatinib for the treatment of EGFR mutation-positive non-small-cell lung cancer in the 'real-world' clinical setting. *Ther Adv Med Oncol* (2019) 11:1758835919836374. doi: 10.1177/1758835919836374

16. Wu YL, Zhou C, Liam CK, Wu G, Liu X, Zhong Z, et al. First-line erlotinib versus gemcitabine/cisplatin in patients with advanced EGFR mutation-positive non-small-cell lung cancer: analyses from the phase III, randomized, open-label, ENSURE study. *Ann Oncol* (2015) 26(9):1883–9. doi: 10.1093/annonc/mdv270

17. Kobayashi S, Ji H, Yuza Y, Meyerson M, Wong KK, Tenen DG, et al. An alternative inhibitor overcomes resistance caused by a mutation of the epidermal growth factor receptor. *Cancer Res* (2005) 65(16):7096–101. doi: 10.1158/0008-5472.CAN-05-1346

18. Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong KK, et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci U.S.A.* (2008) 105(6):2070–5. doi: 10.1073/pnas.0709662105

19. Soejima K, Yasuda H, Hirano T. Osimertinib for EGFR T790M mutation-positive non-small cell lung cancer. *Expert Rev Clin Pharmacol* (2017) 10(1):31–8. doi: 10.1080/17512433.2017.1265446

20. Soria JC, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, Lee KH, et al. FLAURA investigators. Osimertinib in untreated EGFR-mutated advanced

non-small-cell lung cancer. *N Engl J Med* (2018) 378(2):113–25. doi: 10.1056/NEJMoa1713137

21. Ramalingam SS, Vansteenkiste J, Planchard D, Cho BC, Gray JE, Ohe Y, et al. FLAURA investigators. Overall survival with osimertinib in untreated, EGFR-mutated advanced NSCLC. *N Engl J Med* (2020) 382(1):41–50. doi: 10.1056/NEJMoa1913662

22. Cross DA, Ashton SE, Ghiorghiu S, Eberlein C, Nebhan CA, Spitzler PJ, et al. overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discovery* (2014) 4(9):1046–61. doi: 10.1158/2159-8290.CD-14-0337

23. Leonetti A, Sharma S, Minari R, Perego P, Giovannetti E, Tiseo M. Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. *Br J Cancer* (2019) 121(9):725–37. doi: 10.1038/s41416-019-0573-8

24. Oxnard GR, Hu Y, Mileham KF, Husain H, Costa DB, Tracy P, et al. Assessment of resistance mechanisms and clinical implications in patients with EGFR T790M-positive lung cancer and acquired resistance to osimertinib. *JAMA Oncol* (2018) 4 (11):1527–34. doi: 10.1001/jamaoncol.2018.2969

25. Wang S, Tsui ST, Liu C, Song Y, Liu D. EGFR C797S mutation mediates resistance to third-generation inhibitors in T790M-positive non-small cell lung cancer. *J Hematol Oncol* (2016) 9(1):59. doi: 10.1186/s13045-016-0290-1

26. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* (2009) 361 (10):947–57. doi: 10.1056/NEJMoa0810699

27. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol* (2012) 13(3):239–46. doi: 10.1016/S1470-2045(11)70393-X

28. Sequist LV, Yang JC, Yamamoto N, O'Byrne K, Hirsh V, Mok T, et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol* (2013) 31(27):3327–34. doi: 10.1200/JCO.2012.44.2806

29. Michor F, Hughes TP, Iwasa Y, Branford S, Shah NP, Sawyers CL, et al. Dynamics of chronic myeloid leukaemia. *Nature* (2005) 435(7046):1267–70. doi: 10.1038/nature03669

30. Diaz LA Jr, Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* (2012) 486(7404):537–40. doi: 10.1038/nature11219

31. Foo J, Michor F. Evolution of acquired resistance to anti-cancer therapy. *J Theor Biol* (2014) 355:10–20. doi: 10.1016/j.jtbi.2014.02.025

32. Turajlic S, Sottoriva A, Graham T, Swanton C. Resolving genetic heterogeneity in cancer. *Nat Rev Genet* (2019) 20(7):404–16. doi: 10.1038/s41576-019-0114-6

33. Altrock PM, Liu LL, Michor F. The mathematics of cancer: integrating quantitative models. *Nat Rev Cancer* (2015) 15(12):730–45. doi: 10.1038/nrc4029

34. Castagnino N, Maffei M, Tortolina L, Zoppoli G, Piras D, Nencioni A, et al. Systems medicine in colorectal cancer: from a mathematical model toward a new type of clinical trial. *Wiley Interdiscip Rev Syst Biol Med* (2016) 8(4):314–36. doi: 10.1002/wsbm.1342

35. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the evolution of non-small-cell lung cancer. *N Engl J Med* (2017) 376(22):2109–21. doi: 10.1056/NEJMoa1616288

36. Gridelli C, Rossi A, Carbone DP, Guarize J, Karachaliou N, Mok T, et al. Non-small-cell lung cancer. *Nat Rev Dis Primers* (2015) 1:15009. doi: 10.1038/nrdp.2015.9

37. Wang Z, Yang JJ, Huang J, Ye JY, Zhang XC, Tu HY, et al. Lung adenocarcinoma harboring EGFR T790M and in trans C797S responds to combination therapy of first- and third-generation EGFR TKIs and shifts allelic configuration at resistance. *J Thorac Oncol* (2017) 12(11):1723–7. doi: 10.1016/j.jtho.2017.06.017

38. Oda N, Hotta K, Ninomiya K, Minami D, Ichihara E, Murakami T, et al. A phase II trial of EGFR-TKI readministration with afatinib in advanced non-small-cell lung cancer harboring a sensitive non-T790M EGFR mutation: Okayama Lung Cancer Study Group trial 1403. *Cancer Chemother Pharmacol* (2018) 82(6):1031–8. doi: 10.1007/s00280-018-3694-5

39. Nakagawa K, Garon EB, Seto T, Nishio M, Ponce Aix S, Paz-Ares L, et al. RELAY Study Investigators. Ramucirumab plus erlotinib in patients with untreated, EGFR-mutated, advanced non-small-cell lung cancer (RELAY): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol* (2019) 20(12):1655–69. doi: 10.1016/S1470-2045(19)30634-5

40. Kohsaka S, Nagano M, Ueno T, Suehara Y, Hayashi T, Shimada N, et al. A method of high-throughput functional evaluation of EGFR gene variants of unknown significance in cancer. *Sci Transl Med* (2017) 9(416):eaan6566. doi: 10.1126/scitranslmed.aan6566

41. Starrett JH, Guernet AA, Cuomo ME, Poels KE, van Alderwerelt van Rosenburgh IK, Nagelberg A, et al. Drug sensitivity and allele specificity of first-line osimertinib resistance EGFR mutations. *Cancer Res* (2020) 80(10):2017–30. doi: 10.1158/0008-5472.CAN-19-3819

42. Gunnarsson EB, De S, Leder K, Foo J. Understanding the role of phenotypic switching in cancer drug resistance. *J Theor Biol* (2020) 490:110162. doi: 10.1016/j.jtbi.2020.110162

43. Rotow JK, Botelho Costa D, Paweletz CP, Awad MM, Marcoux P, Rangachari D, et al. Concurrent osimertinib plus gefitinib for first-line treatment of EGFR-mutated non-small cell lung cancer (NSCLC). *J Clin Oncol* (2020) 38(15_suppl):9507–7. doi: 10.1200/JCO.2020.38.15_suppl.9507

frontiers | Frontiers in Oncology

TYPE Review
PUBLISHED 07 November 2023
DOI 10.3389/fonc.2023.1282417

OPEN ACCESS

EDITED BY
Yusuke Suenaga,
Chiba Cancer Center, Japan

REVIEWED BY
Xinchang Zheng,
Baylor College of Medicine, United States
Olga Zimmermannova,
Lund University, Sweden

*CORRESPONDENCE
Yosuke Nagahata
nagahata@kuhp.kyoto-u.ac.jp
Hiroshi Kawamoto
kawamoto@infront.kyoto-u.ac.jp

RECEIVED 24 August 2023
ACCEPTED 24 October 2023
PUBLISHED 07 November 2023

CITATION
Nagahata Y and Kawamoto H (2023) Evolutionary reversion in tumorigenesis. *Front. Oncol.* 13:1282417. doi: 10.3389/fonc.2023.1282417

Evolutionary reversion in tumorigenesis

Yosuke Nagahata* and Hiroshi Kawamoto*

Laboratory of Immunology, Institute for Life and Medical Sciences, Kyoto University, Kyoto, Japan

Cells forming malignant tumors are distinguished from those forming normal tissues based on several features: accelerated/dysregulated cell division, disruption of physiologic apoptosis, maturation/differentiation arrest, loss of polarity, and invasive potential. Among them, accelerated cell division and differentiation arrest make tumor cells similar to stem/progenitor cells, and this is why tumorigenesis is often regarded as developmental reversion. Here, in addition to developmental reversion, we propose another insight into tumorigenesis from a phylogeny viewpoint. Based on the finding that tumor cells also share some features with unicellular organisms, we propose that tumorigenesis can be regarded as "evolutionary reversion". Recent advances in sequencing technologies and the ability to identify gene homologous have made it possible to perform comprehensive cross-species transcriptome comparisons and, in our recent study, we found that leukemic cells resulting from a polycomb dysfunction transcriptionally resemble unicellular organisms. Analyzing tumorigenesis from the viewpoint of phylogeny should reveal new aspects of tumorigenesis in the near future, and contribute to overcoming malignant tumors by developing new therapies.

KEYWORDS

tumorigenesis, evolution, unicellular organism, multicellularization, cross-species comparison, transcriptome

Developmental reversion in tumorigenesis

Cells forming malignant tumors are distinguished from those forming normal tissues based on several features. The first is accelerated/dysregulated cell division with disruption of physiological apoptosis, which makes it easy for malignant cells to proliferate and difficult for them to die (1, 2). The second and third are maturation/differentiation arrest and loss of polarity, which make malignant cells different from normal cells in appearance, and enables us to make pathological diagnoses including dysplasia (3, 4). The fourth is invasive potential, which enables malignant cells to invade through the basement membrane and spread to other organs resulting in metastasis (2, 5, 6). Among these features, accelerated cell division and maturation/differentiation arrest make malignant tumor cells similar to normal undifferentiated cells or stem/progenitor cells, and thus tumorigenesis is often regarded as a reversion of differentiation (7, 8) (Figure 1).

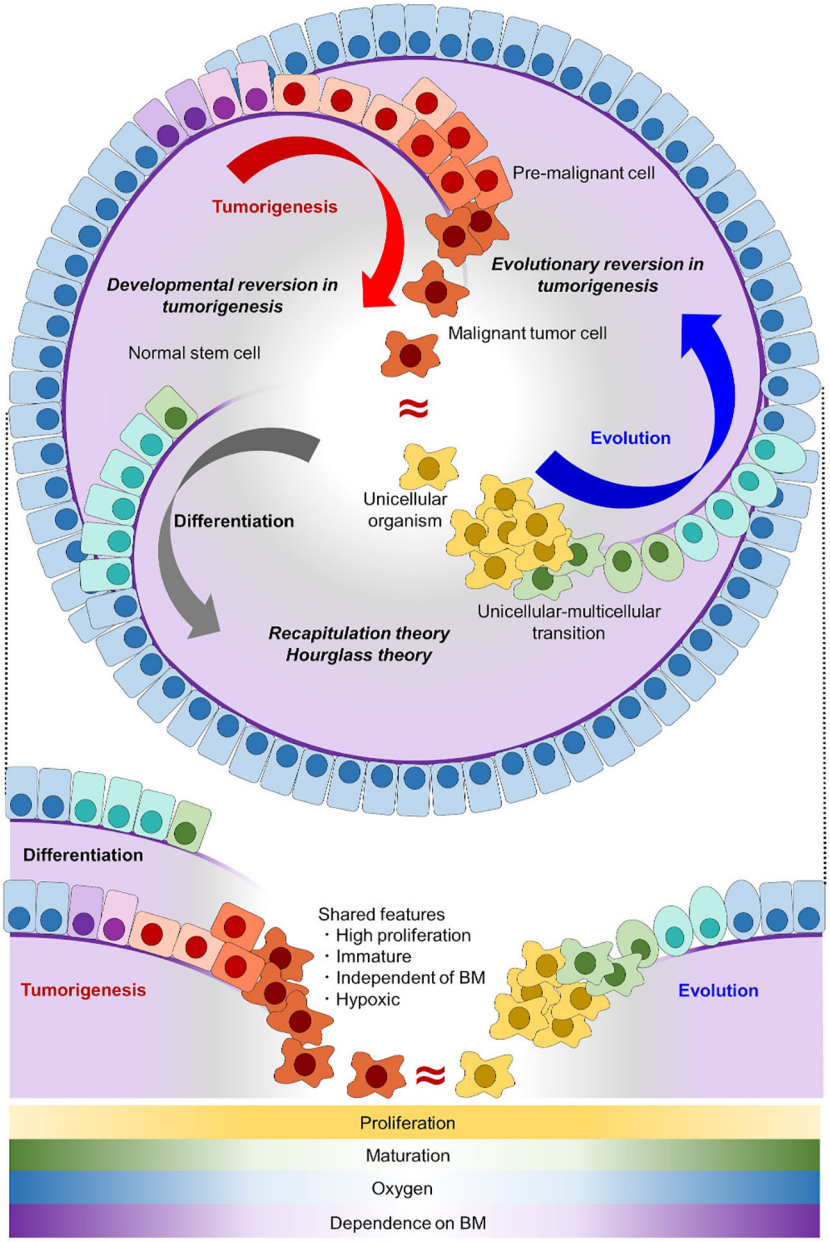


FIGURE 1
Tumorigenesis can be well explained as an evolutionary reversion. In the normal differentiation of epithelial cells (gray arc arrow), tissue stem cells with self-renewal potential generate differentiated progeny cells. In tumorigenesis (red arc arrow), malignant tumor cells acquire self-renewal potential and have an immature phenotype, which makes tumor cells similar to stem cells. In the evolutionary history of early animals (blue arc arrow), ancient unicellular organisms had no basement membrane (purple line) and they proliferated indefinitely. With the onset of multicellularization, early multicellular animals evolved epithelial cells with a basement membrane. Normal stem cells share the proliferative capacity and immature phenotype with malignant tumor cells, but not the invasive potential through a basement membrane. Unicellular organisms share all of these features with malignant tumor cells. Shared features between differentiation/development and evolution are conceptualized as recapitulation theory and hourglass theory (9, 10), and inversely shared features between differentiation/development and tumorigenesis are conceptualized as developmental reversion in tumorigenesis. Here, we proposed that inversely shared features between evolution and tumorigenesis can be conceptualized as evolutionary reversion in tumorigenesis. BM, basement membrane.

The grade of undifferentiated status is clinically important because it is a prognostic factor in some malignant tumors such as thyroid tumors (11) and acute myeloid leukemia (AML) (12–14); undifferentiated tumors are typically more malignant and have an unfavorable prognosis. One limitation of studies until a few decades ago was that comparison was based on the microscopic appearance of tissue sections or cell smear samples, or on expression profiles of genes of probable interest by reverse-transcription PCR or microarray experiments. Although these classifications are valuable and have resulted in improving clinical outcomes, they are neither unbiased nor comprehensive evaluations, and the possibility of unintended bias cannot be completely excluded.

Recent advances in the analysis of cell status with next-generation sequencing, e.g., RNA sequencing, make it possible to

evaluate cell status comprehensively without or at least with minimum bias. We are now able to comprehensively compare grades of undifferentiated status of different cells. Some malignant tumor-specific gene expression profiles have been revealed by comparing transcriptome data of cancer cells and normal cells using RNA sequencing technologies (15). In AML, transcriptomic analysis with RNA sequencing data confirmed that undifferentiated leukemia has a worse prognosis (16–18). These technological advances firmly established that malignant tumor cells have some aspects of reversion of differentiation even when evaluated comprehensively (19, 20). However, while normal stem cells and tumor cells share features such as accelerated proliferation and maturation/differentiation arrest, these two cell types are definitively different in that the latter cells threaten our lives. In other words, reversion in differentiation does not fully explain tumorigenesis and, therefore, it is still important to seek other aspects of tumorigenesis in order to completely understand what malignant tumor cells are and how they arise.

When malignant tumor cells and normal stem cells or benign tumor cells are compared, the invasive potential is an essential feature of the malignant cells. Indeed, this feature makes malignant tumor cells "malignant"; tumor cells invade through a basement membrane and some of them metastasize into other organs (21, 22). In cases with metastatic lesions, patients cannot be cured even if the primary tumor is surgically resected. Loss of polarity is another important feature and pathologist often make a diagnosis based on this (23, 24). Loss of cell polarity is also found in cells in dysplastic tissue, which can be regarded as a premalignant state. Collectively, tumorigenesis cannot be explained simply by the phenomenon of developmental reversion.

Malignant tumor cells share some features with unicellular organism

Here, in addition to developmental reversion, we propose another insight into tumorigenesis from the viewpoint of evolution and phylogeny. When malignant tumor cells and normal cells are compared with unicellular organisms, the tumor cells share more features in common. In an adequate environment, typical unicellular organisms divide indefinitely, lack a basement membrane, and have no/less cell polarity compared with epithelial cells of animals (25) (Figure 1). Among them, lacking basement membrane or invading through it enables unicellular organisms and malignant tumor cells to be distinguished from normal stem/progenitor cell. As invasive potential is an essential for malignancy, evolutionary reversion explains tumorigenesis better than developmental reversion. Together with the fact that multicellular animals including *Homo sapiens* evolved from an ancestral unicellular organism, it is possible to conceptualize that tumorigenesis represents an evolutionary reversion (Figure 1). This evolutionary model in tumorigenesis is different form a traditional tumor micro-evolution model. While the traditional model focuses on tumor evolution in individuals, our model focuses on similarities between tumor cells and unicellular organisms from a view point of phylogeny. Although these two models differ, they can complement each other, and help us understand tumorigenesis more precisely.

Up to now, there are few studies reporting a relationship between tumorigenesis and phylogenetic evolution. Davies and Lineweaver suggested that cancer resembles a prototypic multicellular animal because cancer loses systematic regulation and differentiation (26). In line with this report, Chen et al. reported that the emergence rate of cancer drivers peaked on the deepest branches of multicellular animals, thus, cancer drivers are ancient genes (27). Limitations of this study are that it focused only on tumor-related genes and that only the numbers of genes were examined, while expression levels were not evaluated. Following the study, Trigos et al. and Zhou et al. challenged this limitation. They performed a comprehensive analysis and found that tumor cells more highly expressed these ancient genes compared with normal cells (28–31). Even though a relationship between tumorigenesis and phylogenetics has thus been strongly suggested, another important limitation still remains. These earlier studies compared transcriptional data only among human cells, but not between human cells and unicellular organisms. In other words, it has not been evaluated whether tumor cells truly resemble unicellular organisms in terms of transcriptional profiles, and such a cross-species comparison is needed to make a more precise conclusion.

Another issue has also emerged; some unicellular organisms are not simply unicellular but already have features of multicellularity (32, 33). Whereas it has been suggested in several studies that malignant tumor cells mimic those of early multicellular animals (26, 27, 30), recent advances in the phylogenetics of unicellular and multicellular organisms identified a novel phase for the initiation of multicellularization. Some unicellular organisms display an aggregative state, which should be the prototype of multicellularity (32–42), and cells of some multicellular animals show plasticity like unicellular organisms (43). Thus, it has become possible and important to investigate what it is that malignant tumor cells resemble: early multicellular animals, unicellular organisms, or both.

Comprehensive cross-species comparison of transcriptome data

What makes it difficult to perform cross-species comparisons of transcriptome data is that different species have different genomes. In order to overcome this issue, homologous genes in different species, homologs or orthologs, in other words, should be identified among all genes and throughout all species. Only then can transcriptome data of different species be comprehensively and objectively compared as in a usual analysis of one species (44–46). Recently, by using the OrthoFinder algorithm (47), we performed a cross-species transcriptomic analysis by focusing on four species: mouse, tunicate, sponge, and *Capsaspora owczarzaki*, a unicellular organism (Supplementary Figures 1A, B). Based on such analysis, we succeeded in tracing the evolutionary history of blood cells to the unicellular ancestor of animals (48).

In that study, we have succeeded in comparing transcriptome data from a viewpoint of both phylogenetics and intra-species

lineages. In other words, inter-species lineage analysis was performed by comparing data among i) phagocytic cells, ii) non-phagocytic blood cells, and iii) non-blood cells. We found that macrophages in mouse, tunicate, and sponge are transcriptionally similar to each other and to *C. owczarzaki* (Supplementary Figure 1C). This similarity indicated that macrophage-like phagocytes were the initial blood cells of animals and that their origin can be traced back to unicellular organisms: a common ancestor of animals and *C. owczarzaki*.

When mechanisms of multicellularity are focused on, cell-cell adhesion is a typical feature in multicellularity, and cadherin plays an important role (49–51). Based on our analysis, *C. owczarzaki* does not have cadherin homologs, but proto-cadherin homologs were identified. This proto-cadherin homologous group also contains FAT family genes, which are known to be tumor suppressors and play roles in maintaining cell polarity (52, 53). Interestingly, within *C. owczarzaki*, expression levels of proto-cadherin homolog are higher in the aggregative stage than in the filopodial (amoeboid) stage (Supplementary Figure 1D) (unpublished data obtained by re-analyzing the dataset in our recent report) (48). These results support the idea that the aggregative stage of *C. owczarzaki* represents an intermediate state between unicellular and multicellular organisms, and also support our hypothesis that tumorigenesis has some aspects of loss of multicellularization; i.e. evolutionary reversion.

Evolutionary reversion in tumorigenesis

In the above-mentioned study, we also found that polycomb complexes maintain various blood lineages (T cell, B cell, erythrocyte, and platelet) by repressing phagocyte programs, and that disruption of polycomb complexes led to evolutionary reversion of hematopoiesis; blood in *Ring1a/b* deleted mice was occupied with monocyte/macrophage lineage cells. These findings suggest that various nonphagocytic lineages have evolved from primordial monocytes/macrophages by repressing phagocytic programs with polycomb complexes. Furthermore, these *Ring1a/b*-deleted monocyte/macrophage lineage cells looked like immature monoblasts with CD34 expression, and mice bearing these monoblasts died within a few months, indicating that these were leukemic cells (Figures 2A-D). This is in line with other studies in which disruption of certain polycomb complexes caused AML (54–56).

Surprisingly, we found that *Ring1a/b* KO AML cells were transcriptionally more similar to a eukaryotic unicellular organism, *C. owczarzaki*, than to normal myeloid cells (Figure 2E). We evaluated transcriptional similarities by calculating Pearson's correlation values between mouse normal/leukemic cells and *C. owczarzaki*. AML cells showed higher similarities (correlation values) to all the three stages of *C. owczarzaki*. Thus, malignant tumor cells transcriptionally resemble unicellular organisms; thus,

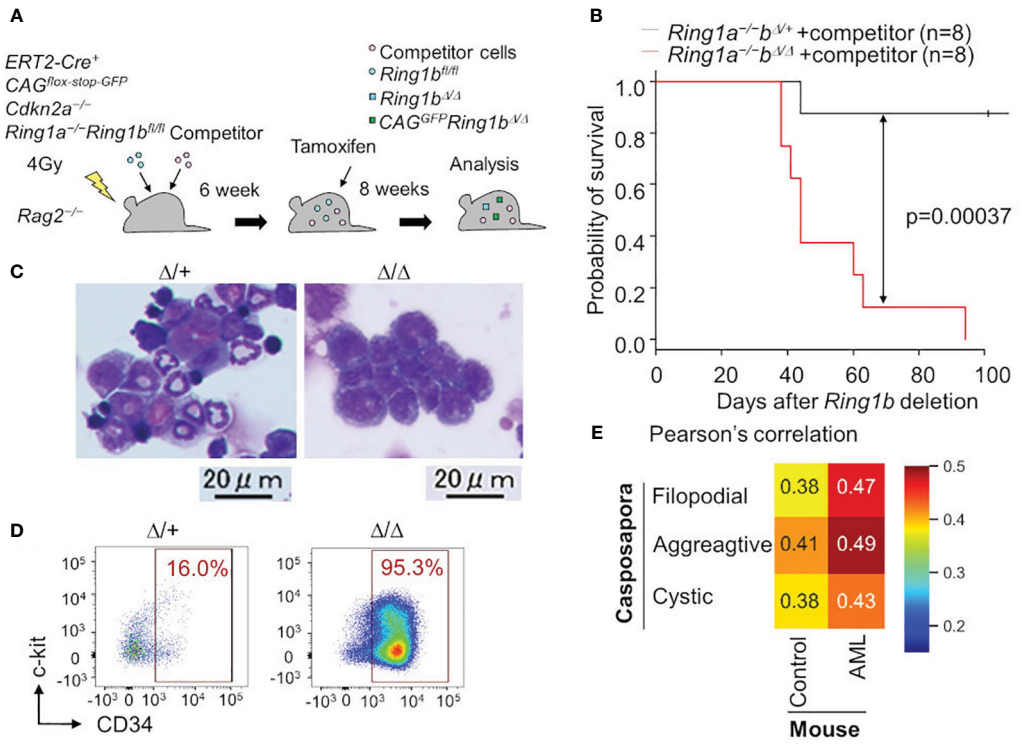


FIGURE 2
Leukemia cells are transcriptionally similar to a unicellular organism. **(A)** *Ring1a/b* were selectively deleted in blood cells using bone marrow chimera mice and the ERT2-Cre system. **(B)** Survival curve with Kaplan-Meier plots of BM transplantation into sublethally irradiated *Rag2*$^{-/-}$ mice with competitor cells. Black and red lines show survival curve of control (*Ring1a–/–Ring1b*$^{\Delta/+}$) and *Ring1a/b*-deleted (*Ring1a*$^{-/-}$*Ring1b*$^{\Delta/\Delta}$) mice on a *Cdkn2a*$^{-/-}$ background, respectively. Statistical significance of differences between the survival rates was calculated with the Log-rank test. **(C, D)** Wright-Giemsa stain of BM smears **(C)** and flow cytometric profiles of whole GFP^{+}CD11b^{+} BM cells **(D)** obtained from control and Ring1a/b-deleted mice with competitor cells. **(E)** Pearson's correlation values between Capsaspora, normal myeloid cells (control) and AML cells.

support for the hypothesis that tumorigenesis has an aspect of evolutionary reversion has become more robust. Although another possibility remains, that disruption of polycomb complexes independently contributes to tumorigenesis as well as evolutionary reversion in hematopoiesis, it is at least reasonable to think that tumorigenesis and phylogenetics have a deep relationship. The fact that histone modification by polycomb complexes is a ubiquitous mechanism also makes this hypothesis more probable (57). Other epigenetic mechanisms such as cohesion are also known to be involved in tumorigenesis (58), thus tumorigenesis, phylogenetics, and epigenetics may have a close relationship. The aspect of phylogenetics is not yet well-accepted in cancer research, therefore, integrating these three research areas is likely to bring about new developments.

In addition to the previous study, we performed cross species analysis adding solid tumor cells and another unicellular organism, choanoflagellate (*Salpingoeca rosetta*), in order to validate our hypothesis of evolutionary reversion in tumorigenesis. In this analysis, AML cells, lung cancer cells, and colon cancer cells were examined and all of them showed higher similarities to both *C. owczarzaki* and *S. rosetta* (Figure 3). We also addressed the issue of traditional cancer evolution from our evolutionary standpoint. In lung cancer, similarities to unicellular organisms were also high in premalignant cells, and it increased along with tumor progression to cancer (Figure 3B). In colon cancer, similarities to unicellular organism got higher along with progression from primary lesion to metastatic lesion (Figure 3C). These data suggested a linear relationship between tumor progression and evolutionary reversion.

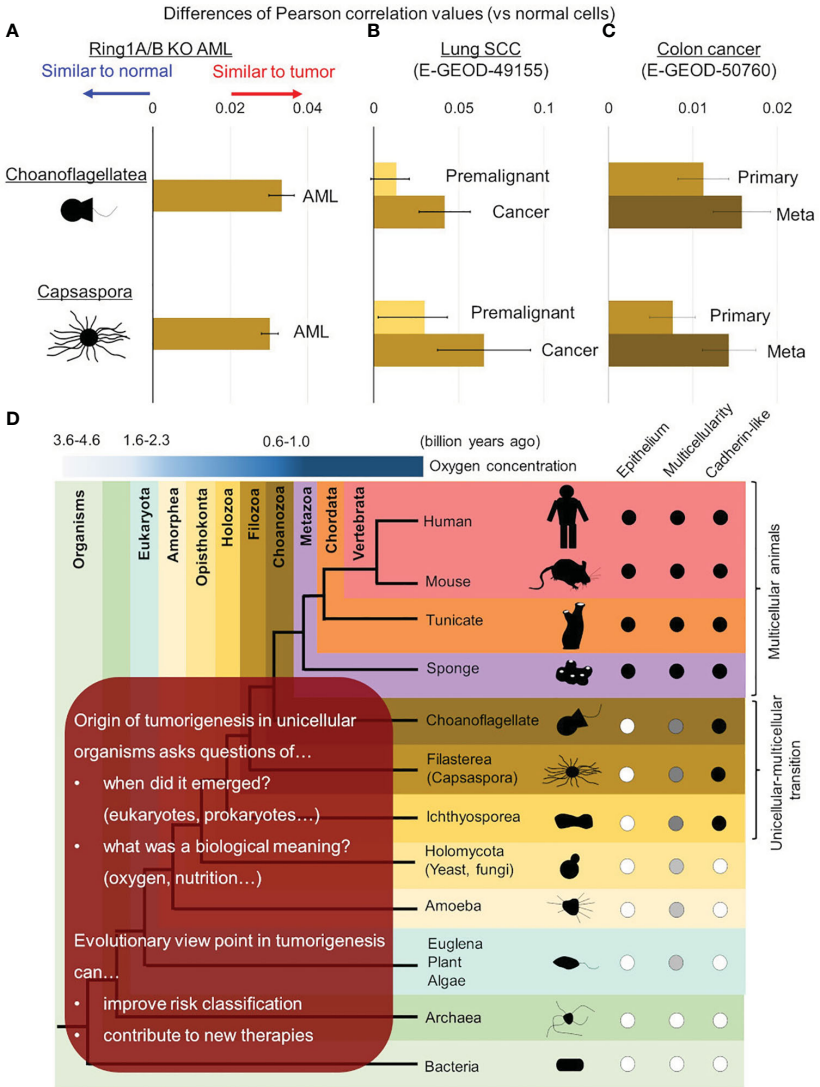


FIGURE 3
Evolutionary reversion in solid tumors and tumor progression. Differences of similarities (Pearson's correlation values) between normal/tumor cells and unicellular organisms were calculated in mouse *Ring1a/b* deleted AML cells **(A)**, lung cancer **(B)**, and colon cancer **(C)**. **(B)** In lung cancer, similarities with unicellular organisms were compared between premalignant cells and cancer cells. **(C)** In colon cancer, similarities with unicellular organisms were compared between primary lesions and metastatic lesions. **(D)** Further investigations are required to reveal origin of tumorigenesis in detail and its biological meaning in ancestral unicellular organisms. Such findings can contribute to overcoming malignant tumors by improving risk classification and developing new therapies.

Further implications of evolutionary aspects in tumorigenesis

It is now strongly suggested that tumorigenesis has an aspect of evolutionary reversion, but many issues are remained to be clarified (Figure 3D). The first issue is when origin of genetic program of tumorigenesis was emerged in ancestral unicellular organisms, because the evolutionary history from the initial organisms to multicellular animals was so long (3 billion years) (33, 59, 60). Although the origin of tumorigenesis may have emerged all at once in an ancestral organism, it is more probable that it emerged step by step during evolution from prokaryotes to unicellular relatives of animals. Further investigations adding various unicellular organism, such as bacteria, archaea, euglena/algae/plants, amoeba, and yeast/fungi, will help us to trace the origin of tumorigenesis more accurately. When the origin of tumorigenesis is traced back to unicellular ancestors, another question emerges; what is a biological meaning or a merit to acquire it in the ancestors? There are many hypotheses worth considering, but oxygen concentration may be essential. As for oxygen concentration, it is well known that tumor microenvironment is hypoxic (24, 61, 62), and it is also suggested that ancient unicellular organisms inhabited hypoxic environment and adjusted to the novel environment with high oxygen (63, 64) (Figure 1). Some ancient genetic program of tumorigenesis may have emerged in such organisms and have brought merits to survive in both of hypoxic and hyperoxic environments. Although such event was beneficial for the ancestral unicellular organisms, risk of tumorigenesis has been inherited to their progenies, and the risk turned out as malignant tumors in animals including human beings. While multicellular animals or their unicellular relatives inherited such dangerous programs, they acquired other genes to control tumorigenesis. These were origins of tumor-suppressor genes, and cell-cell adhesion should have been one of them. Ancestors of animals acquired a cadherin-like protein during evolution from unicellular organisms to multicellular animals (37). This enabled them to create stiff cell-cell adhesion and to form epithelium and multicellularity. Animals have acquired many other cadherin-like proteins including FAT family proteins throughout their evolution, and such proteins have strengthened epithelial cell identity of adhesion. This has worked as safeguard for tumorigenesis preventing premalignant cell from disengaging from epithelium. In other words, losing multicellularization-related genes has been one of the steps of tumorigenesis.

We further argued that evolutionary aspect can contribute to improving clinical outcomes of patients suffering malignant tumors (Figure 3D). In short term, similarities to unicellular organism can make risk-classification more accurate. For example, patients baring tumor cells highly similar to unicellular organism can show worse outcomes, and more intensive therapeutic strategy may overcome the poor prognosis. In long term, evolutionary aspects in tumorigenesis including biological meaning of origin of tumorigenesis can give us a hint to control them and to develop new therapeutic agents or methods. Evaluating effects of agents against unicellular organisms, especially unicellular relatives of animals (e.g., *C. owczarzaki* and *S. rosetta*) will give us informative implications.

In conclusion, cross-species comparisons of transcriptome data provide new insights into tumorigenesis: evolutionary reversion. Further investigations from the novel view point shall help human beings to overcome malignant tumors in the future.

Author contributions

YN: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Writing – original draft, Writing – review & editing. HK: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by funds from Japan Society for the Promotion of Science KAKENHI, Grant-in-Aid for Early-Career Scientists (JP 23K15324), Grant-in-Aid for Scientific Research (B) (JP15H04743), and Grant-in-Aid for Scientific Research on Innovative Areas (JP 19H05747). The Cooperative Research Program (Joint Usage/ Research Center program) of Institute for Life and Medical Sciences, Kyoto University and LiMe Office of Director's Research Grants 2022 (No. 6) and 2023 (No. 7) also supported this work.

Acknowledgments

We thank Peter Burrows (University of Alabama at Birmingham) for critical reading of the manuscript.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1282417/full#supplementary-material

References

1. Hartwell LH, Kastan MB. Cell cycle control and cancer. *Science* (1994) 266:1821–8. doi: 10.1126/science.7997877

2. Compagni A, Christofori G. Recent advances in research on multistage tumorigenesis. *Br J Cancer* (2000) 83:1–5. doi: 10.1054/bjoc.2000.1309

3. Halaoui R, McCaffrey L. Rewiring cell polarity signaling in cancer. *Oncogene* (2015) 34:939–50. doi: 10.1038/onc.2014.59

4. Martin-Belmonte F, Perez-Moreno M. Epithelial cell polarity, stem cells and cancer. *Nat Rev Cancer* (2011) 12:23–38. doi: 10.1038/nrc3169

5. Bogenrieder T, Herlyn M. Axis of evil: molecular mechanisms of cancer metastasis. *Oncogene* (2003) 22:6524–36. doi: 10.1038/sj.onc.1206757

6. Christofori G. New signals from the invasive front. *Nature* (2006) 441:444–50. doi: 10.1038/nature04872

7. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature* (2013) 501:328–37. doi: 10.1038/nature12624

8. Blanpain C, Fuchs E. Stem cell plasticity. Plasticity of epithelial stem cells in tissue regeneration. *Science* (2014) 344:1242281. doi: 10.1126/science.1242281

9. Levit GS, Hossfeld U, Naumann B, Lukas P, Olsson L. The biogenetic law and the Gastraea theory: From Ernst Haeckel's discoveries to contemporary views. *J Exp Zool B Mol Dev Evol* (2022) 338:13–27. doi: 10.1002/jez.b.23039

10. Uesaka M, Kuratani S, Irie N. The developmental hourglass model and recapitulation: An attempt to integrate the two models. *J Exp Zool B Mol Dev Evol* (2022) 338:76–86. doi: 10.1002/jez.b.23027

11. Haddad RI, Nasr C, Bischoff L, Busaidy NL, Byrd D, Callender G, et al. NCCN guidelines insights: thyroid carcinoma, version 2.2018. *J Natl Compr Canc Netw* (2018) 16:1429–40. doi: 10.6004/jnccn.2018.0089

12. Mertelsmann R, Tzvi Thaler H, To L, Gee TS, McKenzie S, Schauer P, et al. Morphological classification, response to therapy, and survival in 263 adult patients with acute nonlymphoblastic leukemia. *Blood* (1980) 56:773–81. doi: 10.1182/blood.V56.5.773.773

13. Cuneo A, Ferrant A, Michaux JL, Boogaerts M, Demuynck H, Van Orshoven A, et al. Cytogenetic profile of minimally differentiated (FAB M0) acute myeloid leukemia: correlation with clinicobiologic findings [see comments]. *Blood* (1995) 85:3688–94. doi: 10.1182/blood.V85.12.3688.bloodjournal85123688

14. Walter RB, Othus M, Burnett AK, Lowenberg B, Kantarjian HM, Ossenkoppele GJ, et al. Significance of FAB subclassification of "acute myeloid leukemia, NOS" in the 2008 WHO classification: analysis of 5848 newly diagnosed patients. *Blood* (2013) 121:2424–31. doi: 10.1182/blood-2012-10-462440

15. Aaltonen LA, Abascal F, Abeshouse A, Aburatani H, Adams DJ, Agranal N, et al Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. *Nature* (2020) 578:82–93. doi: 10.1038/s41586-020-1969-6

16. Zeng AGX, Bansal S, Jin L, Mitchell A, Chen WC, Abbas HA, et al. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. *Nat Med* (2022) 28:1212–23. doi: 10.1038/s41591-022-01819-x

17. Ng SW, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* (2016) 540:433–7. doi: 10.1038/nature20598

18. Huang BJ, Smith JL, Farrar JE, Wang YC, Umeda M, Ries RE, et al. Integrated stem cell signature and cytomolecular risk determination in pediatric acute myeloid leukemia. *Nat Commun* (2022) 13:5487. doi: 10.1038/s41467-022-33244-6

19. Lawson DA, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* (2015) 526:131–5. doi: 10.1038/nature15260

20. Dravis C, Chung CY, Lytle NK, Herrera-Valdez J, Luna G, Trejo CL, et al. Epigenetic and transcriptomic profiling of mammary gland development and tumor models disclose regulators of cell state plasticity. *Cancer Cell* (2018) 34:466–82.e6. doi: 10.1016/j.ccell.2018.08.001

21. Liotta LA, Kohn E. Cancer invasion and metastases. *JAMA* (1990) 263:1123–6. doi: 10.1001/jama.1990.03440080101032

22. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* (2000) 100:57–70. doi: 10.1016/s0092-8674(00)81683-9

23. Fish EM, Molitoris BA. Alterations in epithelial polarity and the pathogenesis of disease states. *N Engl J Med* (1994) 330:1580–8. doi: 10.1056/NEJM199406023302207

24. Weinberg RA. *The biology of cancer*. New York: Garland Science (2014).

25. Nedelcu AM. The evolution of multicellularity and cancer: views and paradigms. *Biochem Soc Trans* (2020) 48:1505–18. doi: 10.1042/BST20190992

26. Davies PC, Lineweaver CH. Cancer tumors as Metazoa 1.0: tapping genes of ancient ancestors. *Phys Biol* (2011) 8:15001. doi: 10.1088/1478-3975/8/1/015001

27. Chen H, Lin F, Xing K, He X. The reverse evolution from multicellularity to unicellularity during carcinogenesis. *Nat Commun* (2015) 6:6367. doi: 10.1038/ncomms7367

28. Trigos AS, Pearson RB, Papenfuss AT, Goode DL. Altered interactions between unicellular and multicellular genes drive hallmarks of transformation in a diverse range of solid tumors. *Proc Natl Acad Sci USA*. (2017) 114:6406–11. doi: 10.1073/pnas.1617743114

29. Trigos AS, Pearson RB, Papenfuss AT, Goode DL. How the evolution of multicellularity set the stage for cancer. *Br J Cancer* (2018) 118:145–52. doi: 10.1038/bjc.2017.398

30. Trigos AS, Pearson RB, Papenfuss AT, Goode DL. Somatic mutations in early metazoan genes disrupt regulatory links between unicellular and multicellular genes in cancer. *Elife* (2019) 8:e40947. doi: 10.7554/eLife.40947

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

32. Sebe-Pedros A, Degnan BM, Ruiz-Trillo I. The origin of Metazoa: a unicellular perspective. *Nat Rev Genet* (2017) 18:498–512. doi: 10.1038/nrg.2017.21

33. Ros-Rocher N, Perez-Posada A, Leger MM, Ruiz-Trillo I. The origin of animals: an ancestral reconstruction of the unicellular-to-multicellular transition. *Open Biol* (2021) 11:200359. doi: 10.1098/rsob.200359

34. Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, et al. Multigene phylogeny of choanozoa and the origin of animals. *PloS One* (2008) 3:e2098. doi: 10.1371/journal.pone.0002098

35. Sebe-Pedros A, de Mendoza A, Lang BF, Degnan BM, Ruiz-Trillo I. Unexpected repertoire of metazoan transcription factors in the unicellular holozoan Capsaspora owczarzaki. *Mol Biol Evol* (2011) 28:1241–54. doi: 10.1093/molbev/msq309

36. Torruella G, Derelle R, Paps J, Lang BF, Roger AJ, Shalchian-Tabrizi K, et al. Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Mol Biol Evol* (2012) 29:531–44. doi: 10.1093/molbev/msr185

37. Suga H, Chen Z, de Mendoza A, Sebe-Pedros A, Brown MW, Kramer E, et al. The Capsaspora genome reveals a complex unicellular prehistory of animals. *Nat Commun* (2013) 4:2325. doi: 10.1038/ncomms3325

38. Sebe-Pedros A, Irimia M, Del Campo J, Parra-Acero H, Russ C, Nusbaum C, et al. Regulated aggregative multicellularity in a close unicellular relative of metazoa. *Elife* (2013) 2:e01287. doi: 10.7554/eLife.01287

39. Sebe-Pedros A, Ballare C, Parra-Acero H, Chiva C, Tena JJ, Sabido E, et al. The dynamic regulatory genome of capsaspora and the origin of animal multicellularity. *Cell* (2016) 165:1224–37. doi: 10.1016/j.cell.2016.03.034

40. Fairclough SR, Chen Z, Kramer E, Zeng Q, Young S, Robertson HM, et al. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta. *Genome Biol* (2013) 14:R15. doi: 10.1186/gb-2013-14-2-r15

41. de Mendoza A, Suga H, Permanyer J, Irimia M, Ruiz-Trillo I. Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. *Elife* (2015) 4:e08904. doi: 10.7554/eLife.08904

42. Torruella G, de Mendoza A, Grau-Bove X, Anto M, Chaplin MA, del Campo J, et al. Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. *Curr Biol* (2015) 25:2404–10. doi: 10.1016/j.cub.2015.07.053

43. Sogabe S, Hatleberg WL, Kocot KM, Say TE, Stoupin D, Roper KE, et al. Pluripotency and the origin of animal multicellularity. *Nature* (2019) 570:519–22. doi: 10.1038/s41586-019-1290-4

44. Shafer MER. Cross-species analysis of single-cell transcriptomic data. *Front Cell Dev Biol* (2019) 7:175. doi: 10.3389/fcell.2019.00175

45. Leong JCK, Li Y, Uesaka M, Uchida Y, Omori A, Hao M, et al. Derivedness index for estimating degree of phenotypic evolution of embryos: A study of comparative transcriptomic analyses of chordates and echinoderms. *Front Cell Dev Biol* (2021) 9:749963. doi: 10.3389/fcell.2021.749963

46. Leong JCK, Uesaka M, Irie N. Distinguishing evolutionary conservation from derivedness. *Life (Basel)* (2022) 12:440. doi: 10.3390/life12030440

47. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* (2019) 20:238. doi: 10.1186/s13059-019-1832-y

48. Nagahata Y, Masuda K, Nishimura Y, Ikawa T, Kawaoka S, Kitawaki T, et al. Tracing the evolutionary history of blood cells to the unicellular ancestor of animals. *Blood* (2022) 140:2611–25. doi: 10.1182/blood.2022016286

49. Yoshida C, Takeichi M. Teratocarcinoma cell adhesion: identification of a cell-surface protein involved in calcium-dependent cell aggregation. *Cell* (1982) 28:217–24. doi: 10.1016/0092-8674(82)90339-7

50. Yoshida-Noro C, Suzuki N, Takeichi M. Molecular nature of the calcium-dependent cell-cell adhesion system in mouse teratocarcinoma and embryonic cells studied with a monoclonal antibody. *Dev Biol* (1984) 101:19–27. doi: 10.1016/0012-1606(84)90112-x

51. Nagafuchi A, Shirayoshi Y, Okazaki K, Yasuda K, Takeichi M. Transformation of cell adhesion properties by exogenously introduced E-cadherin cDNA. *Nature* (1987) 329:341–3. doi: 10.1038/329341a0

52. Mahoney PA, Weber U, Onofrechuk P, Biessmann H, Bryant PJ, Goodman CS. The fat tumor suppressor gene in Drosophila encodes a novel member of the cadherin gene superfamily. *Cell* (1991) 67:853–68. doi: 10.1016/0092-8674(91)90359-7

53. Moeller MJ, Soofi A, Braun GS, Li X, Watzl C, Kriz W, et al. Protocadherin FAT1 binds Ena/VASP proteins and is necessary for actin dynamics and cell polarization. *EMBO J* (2004) 23:3769–79. doi: 10.1038/sj.emboj.7600380

54. Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson AG, et al. Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. *N Engl J Med* (2013) 368:2059–74. doi: 10.1056/NEJMoa1301689

55. Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature* (2018) 562:526–31. doi: 10.1038/s41586-018-0623-z

56. Nagase R, Inoue D, Pastore A, Fujino T, Hou HA, Yamasaki N, et al. Expression of mutant Asxl1 perturbs hematopoiesis and promotes susceptibility to leukemic transformation. *J Exp Med* (2018) 215:1729–47. doi: 10.1084/jem.20171151

57. Piunti A, Shilatifard A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* (2016) 352:aad9780. doi: 10.1126/science.aad9780

58. Ochi Y, Kon A, Sakata T, Nakagawa MM, Nakazawa N, Kakuta M, et al. Combined cohesin-RUNX1 deficiency synergistically perturbs chromatin looping and causes myelodysplastic syndromes. *Cancer Discovery* (2020) 10:836–53. doi: 10.1158/2159-8290.CD-19-0982

59. Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJ, Pisani D. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat Ecol Evol* (2018) 2:1556–62. doi: 10.1038/s41559-018-0644-x

60. Loron CC, Francois C, Rainbird RH, Turner EC, Borensztajn S, Javaux EJ. Early fungi from the Proterozoic era in Arctic Canada. *Nature* (2019) 570:232–5. doi: 10.1038/s41586-019-1217-0

61. Semenza GL. Targeting HIF-1 for cancer therapy. *Nat Rev Cancer* (2003) 3:721–32. doi: 10.1038/nrc1187

62. Terry S, Buart S, Chouaib S. Hypoxic stress-induced tumor and immune plasticity, suppression, and impact on tumor heterogeneity. *Front Immunol* (2017) 8:1625. doi: 10.3389/fimmu.2017.01625

63. Hammarlund EU. Harnessing hypoxia as an evolutionary driver of complex multicellularity. *Interface Focus* (2020) 10:20190101. doi: 10.1098/rsfs.2019.0101

64. Chen SC, Sun GX, Yan Y, Konstantinidis KT, Zhang SY, Deng Y, et al. The Great Oxidation Event expanded the genetic repertoire of arsenic metabolism and cycling. *Proc Natl Acad Sci USA* (2020) 117:10414–21. doi: 10.1073/pnas.2001063117

frontiers | Frontiers in Oncology

TYPE Hypothesis and Theory
PUBLISHED 09 November 2023
DOI 10.3389/fonc.2023.1272981

OPEN ACCESS

EDITED BY
Mamoru Kato,
National Cancer Centre, Japan

REVIEWED BY
Barani Kumar Rajendran,
Yale University, United States
Maik Kschischo,
Koblenz University of Applied Sciences, Germany

*CORRESPONDENCE
Caroline M. Weisman
cweisman@princeton.edu

RECEIVED 04 August 2023
ACCEPTED 20 October 2023
PUBLISHED 09 November 2023

CITATION
Weisman CM (2023) The permissive binding theory of cancer.
Front. Oncol. 13:1272981.
doi: 10.3389/fonc.2023.1272981

The permissive binding theory of cancer

Caroline M. Weisman*

Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, United States

The later stages of cancer, including the invasion and colonization of new tissues, are actively mysterious compared to earlier stages like primary tumor formation. While we lack many details about both, we do have an apparently successful explanatory framework for the earlier stages: one in which genetic mutations hold ultimate causal and explanatory power. By contrast, on both empirical and conceptual grounds, it is not currently clear that mutations alone can explain the later stages of cancer. Can a different type of molecular change do better? Here, I introduce the "permissive binding theory" of cancer, which proposes that novel protein binding interactions are the key causal and explanatory entity in invasion and metastasis. It posits that binding is more abundant at baseline than we observe because it is restricted in normal physiology; that any large perturbation to physiological state revives this baseline abundance, unleashing many new binding interactions; and that a subset of these cause the cellular functions at the heart of oncogenesis, especially invasion and metastasis. Significant physiological perturbations occur in cancer cells in very early stages, and generally become more extreme with progression, providing interactions that continually fuel invasion and metastasis. The theory is compatible with, but not limited to, causal roles for the diverse molecular changes observed in cancer (e.g. gene expression or epigenetic changes), as these generally act causally upstream of proteins, and so may exert their effects by changing the protein binding interactions that occur in the cell. This admits the possibility that molecular changes that appear quite different may actually converge in creating the same few protein complexes, simplifying our picture of invasion and metastasis. If correct, the theory offers a concrete therapeutic strategy: targeting the key novel complexes. The theory is straightforwardly testable by large-scale identification of protein interactions in different cancers.

KEYWORDS

cancer evolution, invasion and metastasis, protein interactions, cancer epigenetic evolution, theory of cancer, evolutionary mechanism

1 Introduction

1.1 The puzzle of invasion and metastasis

The evolution of cancer can be conceptually divided into early and late stages. In the early stage, cells proliferate excessively *in situ*, forming a primary tumor mass. There is now general agreement on what causes this process: genetic mutations. Reproducible "driver mutations" accumulate sequentially, pushing cancer development forward with each hit.

These drivers affect genes in a finite number of pathways whose identities – for example, cell proliferation and apoptosis – are well-suited to explain the phenotypes, like excessive growth, observed in these stages of cancer (1, 2): they require only increases or decreases in their normal functions, which were usually already in operation to some degree in the ancestral cell. This picture constitutes a general explanatory framework for cancer evolution, which I will refer to as the "mutation-centric" framework. It was formally introduced in 1988 in a key paper by Kurt Vogelstein (2) (there referred to as the "genetic model"), whose central figure which I have mildly simplified as Figure 1 below.

The mutation-centric framework is not complete, but is compelling and robust: it is consistent with much of the existing data, and is readily able to explain its major features. When one uses it to consider questions raised by early cancer evolution, one feels that one is approaching the problem in more or less the right way.

In later stages of cancer evolution, cells evolve to invade surrounding tissue, migrate, and colonize new sites in the body, where they form metastases. During this "invasion-metastasis cascade," cells evolve a successive series of abilities that they previously entirely lacked. They break free of epithelia; push past neighboring cells; move into and out of vasculature; survive transport in circulation; and, finally, reach new tissues, only then beginning to adapt to the many new demands of its radically foreign ecology. Colonization is particularly baffling (3), as it seems to require many distinct adaptations, one for each feature of the new environment to which it is not initially suited, that do not seem to have anything to do with one another, or with the adaptations gained in earlier stages (4). Perhaps most remarkable is that cancer passes through each of these life stages in turn, acquiring radical new abilities and forms only briefly before moving on to the new set required by its next phase. How invasion and metastasis happen is a profound question for evolutionary biology.

One is naturally inclined to consider the invasion-metastasis cascade through the mutation-centric framework in trying to understand it. But, even since the introduction of the framework in 1988, as indicated by the striking comparative vagueness of the final arrow in Figure 1 (identical to the original), this has proved largely unsuccessful. Searches for recurrent driver mutations, akin to those found in the early stages, have failed (1, 3, 5–7). Worse, metastases do not even seem to require new mutations relative to their primary tumor counterparts (8, 9), directly contradicting the main tenet of the framework. Beyond the framework's empirical failures, invasion and metastasis present it with conceptual problems. It is not clear how it can accommodate, let alone explain, some basic phenomenology: cancer stem cells, and their relationship to invasion and metastasis (3, 10, 11); polyploid (12) and polyaneuploid (13) giant cancer cells; dramatic genome perturbations like aneuploidy (14) and chromothrypsis (15); and cancer reversion following movement to a normal environment (16–18) or transient reversion of driver mutations (19) It also seems, a priori, a poor match for the evolutionary problem. Many seemingly distinct adaptations are required; if each requires even a few mutations, the total number seems too large to be possible. In brief, the mutation-centric framework, though not strictly disproven, just does not feel like a good fit.

What one considers to be the most explanatorily useful entity in considering a problem is among the most important choices in a scientific enterprise. The correct choice allows discovery of fundamental principles; the wrong one all but precludes it. Is there a type of molecular change that could do better than mutations in explaining observations to date, in guiding future investigation, and in presenting unifying principles through which what is now a pile of disjointed facts may appear a coherent whole? In light of the limitations of the mutation picture, it would be foolish not to search for such alternatives. Indeed, many have already offered proposals (20). Here, I do the same.

1.2 A "binding-centric" framework: the permissive binding theory of oncogenesis

Here, I propose a general molecular mechanism underlying cancer evolution. It was conceived of to address specifically the conceptual gap posed by the later stages of invasion and metastasis, but it may also apply just as well to components of the early ones. In brief, the theory holds that protein binding is substantially permissive, rather than mainly instructive, as we often think. Natural selection has acted (either "intentionally" or as a side effect of other actions) to *limit* binding interactions, pruning many that exist in a baseline state in which they are abundant. Perturbations to cells may generate configurations of proteins that have not been effectively subject to this pruning. These are in the baseline, "permissive state," such that many novel protein binding

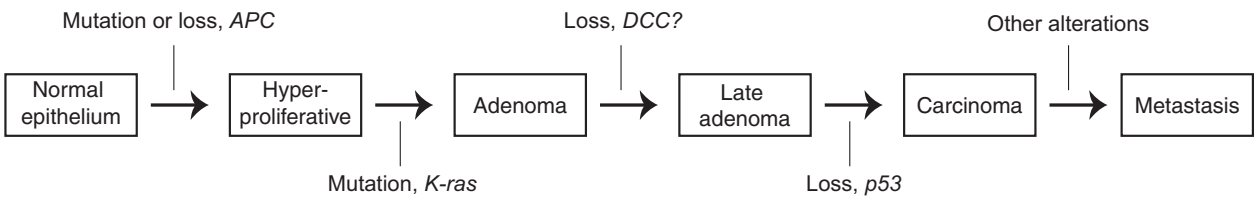


FIGURE 1
A simplified version of the schematization of the mutation-centric theory, introduced in (2), which proposes that the successive accumulation of genetic mutations causes the changes of early oncogenesis. The figure depicts the particular mutations that, under this theory, prompt progression from one stage of cancer to the next in colorectal cancer.

interactions are unleashed. Significant physiological perturbations are ubiquitous in cancer, and especially in its later stages, where they unleash many new binding interactions in this way. A subset of these interactions happen to cause functions that the nascent cancer finds useful; these drive oncogenesis.

Binding interactions are central in the theory, taking the place of mutations as the key entity with both causal and explanatory power. These include interactions between pairs of proteins, but also between proteins and DNA (as transcription factors). They are the direct causal effectors of oncogenesis, as protein interactions are for most biological functions. They are *not* the ultimate cause of oncogenesis: they are determined by a large number of more fundamental physiological factors, like the cell's physical environment, chromatin state, genome, gene expression state, and so on. With perfect knowledge, we would be able to explain all protein interactions in the cell as a function of these underlying factors. But because the relationship between these levels is enormously complex, this ability is unlikely to exist for many years. In its stead, we can gain explanatory power by directly considering the protein interactions themselves.

Below, I will first explain the theory and its rationale, arguing based on three postulates that it is mechanistically principled. Second, I will illustrate how I imagine the theory plays out in the context of cancer to drive its evolution, with attempts to highlight how it is matches the common trajectory of the disease and how it can accommodate diverse observations that have so far lacked explanation. Third, I will discuss the modest amount of direct empirical data that bears on the theory. Fourth, hoping to have inspired interest in testing the theory, I will offer some thoughts on how to do so.

2 The mechanistic basis of the permissive binding theory

2.1 Oncogenesis by activation

The first mechanistic postulate of the theory is that cancer is driven largely by activating functions latent in the cell.

Compared to entire organisms, cancer has an advantage in the struggle to adapt. Many of the challenges that it faces require the gain of functions – migrate, make vessels, secrete cytokines – and adaptation to ecologies – liver, bone, lymph – for which there are ready-made blueprints in its genome. Other cells in the organism, in other contexts, use these as part of their normal function. Evolution is loathe to invent from whole cloth when it can tinker around the edges (21); cancer uses the ample material of the genome, in which is contained every function performed by any cell at any time during the life of the organism.

Consider two functions already widely appreciated as key in the invasion-metastasis cascade. The first is the epithelial-mesenchymal transition (EMT). This is a programmed phenotypic shift in which epithelial cells gain the mesenchymal traits necessary for them to leave the epithelial sheet and become migratory. It comprises multiple processes, including dissolving cell-cell junctions, removing apico-basal polarity, and reorganizing the cytoskeleton. It is essential for embryogenesis, as during the concerted movement of epiblast cells into the embryonic interior during gastrulation and the long migration of the epithelial neural crest to sites around the vertebrate body. It can also be reactivated in select contexts in the adult, as in the migration of cells from the edge to the center of a healing wound. The singular term "EMT" is used across these cases despite their contextual diversity because the genetic program at their core is highly similar. This makes sense: the mechanics required are highly similar, and so evolution has not reinvented the wheel. A carcinoma cell, too, seeking to invade surrounding tissue or metastasize elsewhere, shares the initial state and required mechanics. The EMT would serve its purpose nicely, and, preformed in its genome, is low-hanging fruit. So, indeed, this same EMT is widely recognized to be at the core of carcinoma invasion and metastasis (10, 22).

A similar story is true for angiogenesis. Tumors are heavily reliant on oxygen. But carrying oxygen deep within tissue is not achievable from scratch: the only solution is to activate an existing program, latent in the genome, built for the purpose. In what has been termed the "angiogenic switch (23)," tumors do just this. Some leverage the language of their common genome to manipulate their neighbors, inducing existing vessels to expand by activating familiar angiogenic factors like VEGF (23). Others are more self-reliant, morphing themselves into vasculature with latent differentiation programs that guide them down the endothelial lineage (24). There is more than one road to Rome, but all are well-trodden in normal development.

I use these two examples because they are common and comparatively well-understood. But the principle applies generally. We do not yet know the detailed mechanisms for the many processes driving the diverse adaptations of cancer, but the intuitive stance, supported by these case studies, is that evolution is surely not re-inventing these functions when intact programs encoding them are quite literally at its fingertips.

This postulate is the least controversial, so I will not belabor it. The point for what follows is that the task facing the cancer cell is largely the activation of existing cellular pathways. The programs already exist, and cancer just needs to flip the switches.

2.2 Activation by binding

The second postulate is that functions latent in the cell can be activated through protein binding.

Proteins have two broad modes of function. The first I call "independent functions." A protein acts independently when it is the direct physical executor of a change. Upstream steps involving other proteins may have been necessary to lead here, but now the protein acts under its own steam, without dependencies. These are the key effectors, typified by enzymes, like kinases that act directly to phosphorylate their targets or polymerases that act directly to synthesize nucleic acids. The second mode I call "dependent functions." Here, a protein causes a change entirely by altering the independent activity of another protein. It may activate, repress, or qualitatively modulate this activity, changing its time, place, rate, degree, or targets. (For simplicity I will discuss individual proteins,

but a rigorous application of this language refers to the minimum functional unit, of whatever size, as being independent. This accommodates the case of proteins that have no activity until they are assembled into a complex, in which all members are necessary for function: here, the complex as a whole is the unit that is independent).

An independent function can work in the absence of dependent modulation. A dependent function cannot. Its effect on the cell is expressed entirely through the independent function. Alone, it is impotent, like a transcription factor without a polymerase: utterly unable to make RNA.

The key difference for what follows is the relative difficulty of these modes. By difficulty, I mean something like the necessary level of molecular specificity. Independent functions are, in general, harder. They effect reactions that do not readily occur spontaneously. This necessitates molecular configurations that are a small fraction of the total number possible: they are "highly specific organic catalysts" (25). Such specificity entails substantial work to find this small fraction from natural selection.

Dependent functions, by contrast, can be very easy. At their easiest, they just require some amount of binding – *sticking* – to an independent protein. Sticking to an active or an allosteric site can stop a kinase from phosphorylating. Sticking near and occluding a target lysine can prevent ubiquitination and degradation. Sticking to two linker regions can bring two proteins together, making one the substrate of the other. Compared to independent functions, dependent functions have many more configurations available to them: they need not be nearly as precise.

Dependent functions are a rule, not an exception. They are everywhere, including core cellular pathways. In the cell cycle, securin acts dependently as a crucial checkpoint: it binds the active site of separase, preventing it from cleaving cohesin and driving the cell cycle until phosphorylated by Cdk1. In wnt signaling, disheveled acts dependently to activate beta-catenin, binding and sequestering the "destruction complex" that phosphorylates it and leaving it free to drive transcription. All transcription factors act dependently to increase the affinity of polymerases for their target genes. (I note here that the permissive binding theory is not limited to protein-protein interactions, but includes, for example, protein-DNA interactions, as for transcription factors).

The work of dependent functions is modest at the molecular level, but can have profound effects on the cell. They can, of course, inhibit activities: sticking to a gene's promoter prevents its transcription. They can also activate: sticking to the active site of a kinase that marks a transcription factor for degradation *activates* that transcription factor, driving whatever cellular programs are its target. And they can invent: the binding of an activator which responds to a cue to a novel transcription factor couples transcription of its target genes and the cue, driving a new cellular response.

The point for what follows is that our intuition of what sorts of molecular interactions are required for useful cellular function often sets the bar too high. It is based on independent functions, which are the harder class, and are not representative of much work in the cell. Dependent functions mean that it is easier than we think for proteins to have useful effects. To activate, reduce, or change existing function, all that is needed is *binding*.

2.3 Binding by perturbation

The third postulate is that any sufficiently large perturbation to the cell can produce many new binding interactions.

We imagine protein binding as primarily "instructive." The default state for proteins is that they do not bind to each other. They come to do so only if specifically "instructed" to by the work of natural selection.

I propose that protein binding is actually substantially (of course, not entirely) permissive. Binding is fairly common among proteins for which this has *not* been selected.

Two intuitions underlie the instructive view. The first is that it seems *unlikely* that binding would occur without having been selected for. The space of amino acid sequences is large; sampling randomly within it, approximating the absence of selection, seems very unlikely to result in sequences between which binding is favorable.

There are two errors in this intuition. The first is that proteins are not random, and especially not *with respect to each other*. They are composed substantially of domains: function-conferring subsequences widely shared across proteins. Domains drive much of protein binding (26), implying that proteins sharing domains are predisposed to share binding partners. We should expect some such binding at baseline.

The second error is more direct: even random sequences *do* bind cellular proteins. This seems surprising even to me, but I have been convinced by several recent studies showing that proteins made from random amino acid sequences can bind to cellular proteins and thereby produce essential cellular functions (27–29). The implication here is that protein binding is reasonably common at baseline, even without shared domains.

A second, stronger intuition driving the instructive view is that, because we observe fairly specific patterns of binding, the permissive view has been falsified on empirical grounds. This is incorrect because what we observe is *not* the baseline state: much binding that existed at baseline is not manifest in cells. Rather than binding being actively crafted from a default state in which it is absent, the *absence of binding is actively created from a default state in which it is common*. This is akin to synapse formation in the brain: initially abundant connections are pruned, preserving a much smaller number. This is the inverse of the standard instructive view, and is due to two mechanisms.

First, natural selection actively changes protein sequences to limit binding interactions. This could be either to remove harmful interactions per se or merely as a side effect of strengthening beneficial ones (30–32).

Second, although the "motive" for binding may exist, the "opportunity" often does not, because the proteins are never in the same place at the same time. They are expressed in different cell types, or at different times, or are localized to different places. They may inhabit the same cell, but be hogged in complexes with more abundant proteins, never free when they meet. As above, this arrangement could be due either to selection against harmful

interactions per se or as a side effect of beneficial localization. Whatever the cause of their separation, offered the chance, many pairs would happily bind.

This second point predicts that forcibly colocalizing proteins should frequently produce binding not seen in normal conditions. This is just what is seen, as has been noted for years in comparing results from yeast two-hybrid screens to those from physiological pull-down or cross-linking experiments (33–35). The many "false positive" interactions found in two-hybrid experiments, while not found in normal cells, reflect something essential. Given the opportunity, proteins bind widely.

The permissive view predicts that any perturbation to the cell that changes protein abundance, *relative* abundance, localization, or affinities amounts to producing a "permissive state," and so should produce novel binding interactions. The number of new interactions should scale with the severity of the perturbation, minor changes producing only a few, but dramatic changes producing many.

Some concrete examples may illustrate the intuition. Changes to a protein's amino acid sequences should also tend to reverse selection's drive to circumscribe its binding interactions, creating a permissive state. Changes to posttranslational modifications, which also perturb its physicochemical properties, should have a similar effect. Changes to transcriptional state may change protein abundance, changing binding directly by mass action; *relative* protein abundance, changing binding by competitive effects on mass action, as proteins formerly complexed with one partner become free to bind others following changes to normal stoichiometry; and localization, as proteins overflow into new cellular compartments, changing binding by bringing pairs together for the first time. Changes to protein or RNA abundance or stability, mediated by other proteins that regulate them, like ubiquitinases, RNA binding proteins, and transcription factors, should have the same effects.

I do not mean that *all* changes to these features will produce permissive states. Healthy regulation *is* change. It produces not *novel* binding interactions, but ones that have been tested and approved by evolution, either crafted for their utility or not purged because they are benign. In normal conditions, the cellular milieu is heavily controlled, constrained to this small fraction of states upon which natural selection has acted to ensure good behavior.

By contrast, perturbations to the protein repertoire that do not occur in the normal life of a healthy cell are not pre-screened by natural selection, and so are not guaranteed to be benign. The permissive baseline of protein binding here emerges, unleashing a two-hybrid experiment on a cellular scale. Interactions fall to feral baseline, and widespread binding breaks free.

I will use the term "perturbation" to refer to any change that produces a physiological state sufficiently different than those encountered in normal life that it results in a permissive state and the novel binding that results. This fits our intuitive use of the word, and emphasizes the generality that is a central feature of the theory. The possible causes of perturbations are enormous, and they are not restricted to any particular class of entity. They include internal changes, like mutations, and external changes, like chemicals, external forces, altered substrates, temperature, and much more.

2.4 Summary of the mechanistic basis

In brief, the permissive binding theory holds that many types of environmental and genetic perturbations produce non-physiological combinations of protein abundance and localization. It also holds that binding is – per its name – natively permissive, such that these perturbations generate a permissive state, in which emerge a large number of novel protein binding interactions. A subset of these binding interactions activate existing genetically-encoded functions, by "dependent" modes of action, that alter the activity of other, "independent" actions. A subset of these functions are those that, being useful to the cell, drive oncogenesis, and, as we will see, particularly the later stages.

3 The permissive binding theory at work in cancer

Having laid out the theory, I will first sketch how it might play out in the context of cancer. This is meant to be quite general, providing the spirit and shape of the theory rather than its exact details.

3.1 The overall trajectory of cancer evolution

Cancer initiates according to the mutation-centric framework as described in the introduction. "Driver mutations" occur, increasing proliferation and decreasing apoptosis, leading to the increased cell numbers of benign hyperplasia. These driver genes are generally pleiotropic, as evidenced by their unusually high centrality in cellular networks (36). They have roles in many cellular pathways. The driver mutations therefore have effects on the cell that extend well beyond the pro-proliferative ones for which they are selected. These include higher mutation rates, as in mutations to DNA surveillance genes like p53; and widespread transcriptional and proteomic perturbations, as in mutations to global transcription factors like N-myc, regulators of global transcription factors like CDKN2A, or central signaling hubs like Ras. As just one example, a common driver KRAS mutation (G13D) significantly changes the expression of 6,000 genes (37) and the phosphorylation state of half of all proteins (38). The side effects of driver mutations are profound: they reverberate through the cell, producing not just more growth but a broadly perturbed state.

The hyperplasia present at this stage is benign to the organism, but produces changes to tissue composition, architecture, and environment felt acutely by the hyperplastic cells. Their contact with the basement membrane is altered, changing stiffness, polarization, mechanotransductive signaling (39); their distance to nearby sources of signaling molecules is changed, giving them more or less (40); the disrupted contacts with neighbors changes those neighbors' behaviors, making them secrete more or less or different signals (41). These environmental changes perturb the transcriptional and proteomic state of the cells.

Even at this early and benign stage, this combination of endogenous (produced by the driver mutations) and exogenous (produced by the hyperplasia) changes causes the physiology of the hyperplastic cells to be significantly perturbed. There are many changes to the numbers, locations, and states of the protein repertoire; this cellular state is now one not seen in any normal cell type. Per the permissive binding theory, this is a permissive state, allowing many novel protein binding interactions to emerge.

Per the theory, some of these new binding interactions activate latent cellular functions. These are induced "randomly," depending only on the molecular details of the particular permissive state. There is no bias for functions useful to the cell. If no useful functions are produced, the hyperplastic process continues, bringing continual physical and environmental changes, and inducing new permissive states, new binding, and new functions.

Eventually, a function useful to the tumor emerges, providing the functions necessary for the next stage of its progression. The all-important EMT is induced, conferring the ability to invade and migrate. The relative importance of endogenous and exogenous changes in this emergence is a detail of the theory not yet clear, and will likely vary from case to case. But the exogenous ones are likely significant, as suggested by findings that changes to the stiffness of the extracellular environment can induce the EMT (42, 43) and that EMT activation varies across the spatial dimensions of the tumor, with preferential activation in cells on its edge (44).

The cancer now moves into surrounding tissue. The cells resident there are well-adapted to it and so live happily, but the invasive cells are not. For them, it is a foreign environment, and so further perturbs cellular physiology, creating permissive states. Another wave of novel binding is unleashed. New functions are generated, from which selection can pick those useful to the progression of the cancer at the present time. If the cells are in striking distance of vasculature, for example, functions that allow them to intravasate may be selected. If they are in need of new metabolic strategies, those functions will be selected instead.

Eventually, the cancer moves through circulation and enters foreign tissue. This environment, more unfamiliar than any it has encountered before, deepens physiological perturbation and creates even more permissive states. Sometimes, among the many functions unlocked are a number and kind sufficient to enable successful colonization of this new territory. A metastasis is formed.

We can briefly summarize the overall picture by delineating two phases of oncogenesis. Initiation occurs when driver mutations disable normal cellular guardrails, allowing unregulated proliferation. They also perturb the cell's physiology internally (through disruption of pathways of which the driver genes are members) and externally (by altering the physical environment of the cells), creating the first suite of permissive states. From here, the iterative second phase, which I call innovation, begins. These permissive states create new functions; these enable access to new oncogenic phenotypes; these perturb physiology; this creates new functions – and so on. Genetic perturbations, of generally increasing severity, accumulate throughout, adding the fuel of their own resulting perturbations. A turn of this crank enables each new stage of cancer, which requires its own set of adaptations, in turn. Progression begets progression; pathology begets pathology. This summary is depicted in Figure 2 below.

3.2 Other phenomenological features as explained by the theory

I will now show how the theory might be useful for explaining various other features and observations in cancer. Many other examples are possible; here, I focus on features that the standard mutation-centric picture struggles to explain or points about which readers may be wondering.

3.2.1 Necessity of driver mutations

The theory holds that many different perturbations, including purely environmental ones, can create the permissive states that enable oncogenesis. Why, then do we not observe cancers triggered exclusively by environment, entirely devoid of mutations?

In normal cells, environmental insults do unleash permissive states. But the affected cells are usually then subject to physiological safeguards that, most of the time, successfully prevent excessive proliferation and cancer progression. These safeguards include processes like apoptosis and cell cycle checkpoints, and are carried out by tumor suppressor genes. Successful oncogenesis requires disabling these safeguards, which is generally possible to the extent required only through mutation of these tumor suppressor genes.

The purely environmental induction of cancer is *possible*, as has been shown in experimental conditions (45). That it does not usually occur in nature likely reflects the low likelihood of evading the tumor suppressive mechanisms that are "baked in" to normal physiology without the help of mutations.

Many diseases other than cancer cause widespread physiological perturbations, like the altered hormonal profiles and chronic inflammation of obesity and diabetes. The above argument is consistent with the observation that these diseases increase the risk of cancer (46). Lacking the key initiating mutations, these conditions alone are not sufficient to drive cancer; but, once these mutations occur, the perturbed environment that they produce increases the probability of cancer developing, and/or the speed with which it progresses.

3.2.2 The role of environmental perturbation

Although environmental perturbations alone are not generally sufficient to *induce* cancer, they likely induce many of the permissive states key to its later evolution. This would explain the observed lack of characteristic metastatic mutations and the finding that they do not seem to be necessary (1, 3, 5–7). The invasion-metastasis cascade is partly or largely driven by the environmental perturbations that are rampant in these stages, rather than mutations.

The key role of environment allows the possibility that mutations within a tumor could act indirectly, by perturbing and creating a permissive state in their neighbors. The high genetic heterogeneity of many tumors (47) may reflect such a strategy.

The tumor microenvironment has increasingly been recognized as key in oncogenesis (41, 48). Some of these effects may be by way of generating permissive states in the tumor.

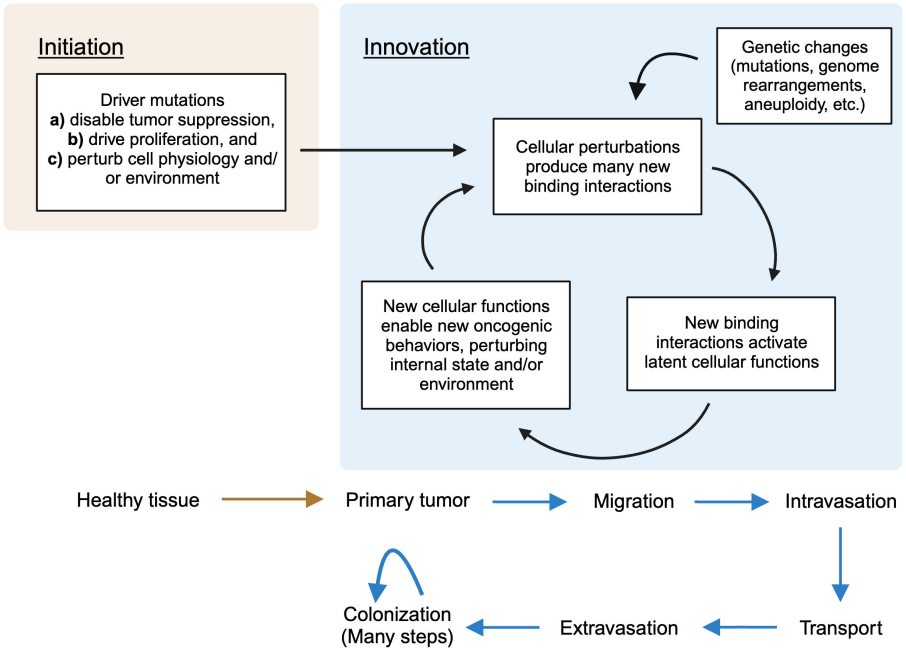


FIGURE 2
A summary of the stages of cancer evolution in terms of the present theory. In the first phase (initiation), driver mutations lead to a primary tumor by disabling tumor suppression, driving proliferation, and perturbing cellular physiology. In the second phase (innovation), an iterative process, new binding interactions are unleashed by perturbed physiology, fueling new oncogenic behaviors, which in turn cause more perturbations as, e.g. the cancer moves into a new environment. Genetic changes accumulate throughout, contributing to perturbation. (Figure created with BioRender.com).

Why do cancers form a primary tumor before they disseminate and metastasize, rather than vice versa? Under the mutation-centric framework, this was assumed to reflect the large number of cell divisions necessary for the mutations enabling spread, an interpretation questioned by the finding that mutations do not seem to be necessary for metastasis, and that metastatic lineages can emerge quite early from primary tumors (49). The essential role of the permissive state produced by the altered physical environment in the primary tumor may be the reason.

3.2.3 Stability and heritability of adaptations

If environmental perturbations cause the abilities that enable oncogenesis, can restoring a normal environment revert the cancer? This has been observed in experimental conditions (16, 17, 50).

Should this always happen? Not necessarily. A strong counterpoint is normal cellular differentiation: a heritable change in cellular state triggered by environmental cues and effected by protein interactions (often, complexes between one or more transcription factors and DNA) that is nonetheless largely irreversible.

A helpful way of conceptualizing this phenomenon is one in which the differentiated cellular state is considered an attractor in a dynamical system, stable against environmental perturbations once reached even though the inciting incident was an environmental change. An analogous model has been proposed for cancer cell states (51). If permissive states unlock new binding events that cause functions which reinforce them, the resulting positive feedback loop may form an attractor, making the cancer cell state stably heritable even when the environmental perturbation is removed. It is counter-intuitive that a non-physiological cellular state like cancer, which has presumably not been produced directly by natural selection, could be stable. But important work in dynamical systems theory has shown that stability in such systems is more common than one might think and does not necessarily require selection (52).

In the case of cancer, it may also be that the duration for which a particular adaptation is advantageous is short, such that it may not need to be as stable as is a normal cell type. For example, it may not be useful for cancers to maintain an activated EMT program after they have already migrated to a new tissue. Especially in this case, but perhaps in general, selection may be able to act to regenerate cells with useful binding interactions despite their not being stably heritable in the very long term.

3.2.4 The role of transcription factors

The theory suggests a key role for transcription factors in oncogenesis. Because cancer cells likely deploy existing genetic programs (e.g. the EMT) in successful invasion and metastasis, transcription factors, as the switches that turn genetic programs on and off, are central. A new binding interaction for a transcription factor, either with a new DNA locus or with a new coactivator or corepressor, has the potential to activate a fully-formed cellular function.

There is also reason to believe that transcription factors should be uniquely vulnerable to forming new binding interactions in perturbed physiology. Many transcription factors bind cooperatively with cofactors, and many bind different targets, or have qualitatively different effects (activation versus repression) on their targets, depending on the particular cofactor to which they are

bound. Small changes in the level of available binding partners can therefore have dramatic changes on target gene expression. That the relative levels (stoichiometry) of transcription factors is a major determinant of gene expression, especially in irreversible cell fate decisions, is well-known (53–56). Physiological perturbations found in cancer, with the capacity to seriously alter protein levels, might therefore be expected to dramatically change the targets of transcription factors.

The large potential of transcription factors to unlock functions useful in cancer combined with the unique vulnerability resulting from their sensitive dependence on stoichiometry suggests that they may play a special role in oncogenesis.

3.2.5 Diversity of environmental insults

A dizzying variety of chemical and environmental exposures increase cancer risk. Structurally diverse chemicals; chronic inflammation; hormonal perturbation; hot liquids; infections; and more all the time (57). We now take for granted that 'almost everything causes cancer,' but such diversity is quite remarkable. One would like a unifying mechanism for the oncogenic activity of such varied agents. It was assumed that most carcinogens are mutagenic, but recent work has shown this not to be the case (58).

The present theory offers an answer. The only requirement to cause oncogenesis, following the acquisition of driver mutations, is any perturbation large enough to generate a permissive state. The theory is agnostic to the specifics of the causal agent. It predicts the striking generality of oncogenic activity that we observe.

3.2.6 Changes in types of mutations over time

Genetic changes accumulate throughout the evolution of cancer. The new environments faced during the invasion-metastasis cascade, particularly colonization, demand a larger number, and more specific types, of adaptations than did the site of the primary tumor. The cancer must activate the specific pathways that can metabolize the new tissue's food sources; adjust to its particular oxygen level; communicate with and coerce its new neighbors; and so on.

Increasingly, it is large, coarse, dramatic genetic changes, like aneuploidy, polyploidy, chromothrypsis, and extracellular DNA, that are useful here (14). Being of very large effect, they dramatically perturb expression levels (59, 60)to create correspondingly permissive states, creating a huge amount of new binding in one fell swoop. This 'everything at the wall' strategy is their best chance of hitting upon the many and specific abilities that they need.

Cells with strong tendencies to produce these dramatic genetic changes in their daughters are especially useful here. They accelerate adaptation by rapidly generating the needed variation. This may be the role of cells that are highly genetically unstable due to very abnormal karyotypes, like polyploid and polyaneuploid giant cancer cells, which have been shown to be capable of seeding metastases with much adaptive potential (12, 61).

These dramatic perturbations are selected for in virtue of the useful interactions that they produce, but also cause many useless, or even harmful, "passenger interactions." Despite not being useful when they occur, they may become useful later, "preadapting" the cancer to new challenges. This may account for at least some of the uncanny ability of cancer cells that have adapted to one environment to adapt to another, despite the two sets of necessary adaptations not obviously having much in common.

3.2.7 Tissue specificity

Because the new binding interactions that drive cancer are due to serendipitous interactions between whatever proteins were present in the perturbed cell, the theory predicts strong dependence of binding events on tissue of origin, based on its characteristic expression profile. As some perturbations are caused by mutations, some mutations should have tissue dependence, as is the case (62).

3.2.8 The EMT-stemness link

It has been observed that cancer cells that have activated the EMT are also strong "cancer stem cells," particularly capable of seeding metastases. The observation is mysterious: the EMT supports the invasion and migration process itself, but is not obviously useful for the different task of colonizing new tissue.

The present theory is consistent with this observation. It holds that any cellular state that is sufficiently dissimilar from those encountered in the normal life of a cell is permissive. The induction of the EMT does not normally occur in the cell types from which tumors have arisen, or in the environments in which they reside. Inducing an EMT program in these foreign contexts pumps its effectors into a cell teeming with proteins utterly unfamiliar to it. A large amount of novel binding is thus unleashed.

This reasoning is consistent with the additional finding that *partial* induction of the EMT, producing an intermediate state neither completely epithelial nor completely mesenchymal, causes stronger stemness than does complete induction. Extended duration of this Frankensteinian intermediate is even more dissimilar to normal physiology, and so is even more permissive.

The theory does not predict that the EMT is special in this regard. Induction of any program that is partial, in a non-physiological cell type, or in a non-physiologic environment should cause oncogenic phenotypes. Just this has been observed in a case of incomplete reprogramming (63).

The EMT is distinctive in that it is strongly selected for because of its utility for migration and invasion. Other programs will not be as often selected, and so will not have the opportunity to drive subsequent adaptation.

3.2.9 New genes and cancer

If protein binding is permissive by default, and such permissive binding is prone to drive oncogenesis, then newly-born genes, for which natural selection has had limited time to prune and limit potentially harmful interactions, should be especially liable to forming such interactions during physiological perturbation, and therefore to promoting cancer.

As I noted previously in a review of so-called "*de novo*" genes, recently been born anew from noncoding DNA, this seems to be the case. Strikingly, the only known functions of human *de novo* genes

are pro-oncogenic ones (64). (It was this curious observation that prompted development of the present theory).

4 Evidence for the theory

My discussion so far has been mostly conceptual, arguing from first principles for the mechanistic plausibility of the theory, and abductive, arguing for its ability to explain and unify much apparently disparate phenomenology in cancer. Is there direct evidence for or against the theory and its specific proposal of the centrality of novel protein interactions?

Direct tests of the theory will ideally begin with identification of binding interactions in cancer cells. This remains technically challenging at medium to large scale, and so relevant data are sparse. But there are some, which I discuss below.

It is important to note that all data discussed below are derived from cell lines, as are the vast majority of experimental data on cancer. Their applicability to the primary aspiration of the theory – the invasion-metastasis cascade – is therefore limited. Although developed primarily to answer the puzzle of invasion and metastasis, there is nothing preventing the theory from operating in earlier cancer, as described above; so I take these results to be promising overall.

4.1 Many novel protein interactions in cancer cell lines

A basic prediction of the theory is that there should be many new binding interactions in cancers. Are there data to this effect?

An important kind of binding interaction is between proteins (as opposed to between, for example, proteins and DNA, as for transcription factors). There is much work aiming to indirectly infer protein interaction networks in cancer lines (from, for example, combinations of transcriptional and genomic data), which has generated results suggestive of many new interactions (65–67). While these results are interesting support for the theory, I will focus here on direct experimental tests, which seem to me more decisive.

One recent pair of studies in a) head and neck (68) and b) breast cancer (69)cell lines systematically identified all binding partners for 30-40 proteins known to drive their respective condition (but not all mutated in the cell lines used). For each cancer type, these protein interactions were identified in two independent cancer lines and, as comparison, in one non-cancerous line from a matched normal tissue.

The number of protein-protein interactions identified in each cell line, and the number found to be shared between them, is shown as Figure 3 below. Two features are striking.

First, all four cancer lines have many more protein-protein interactions, by 2-4 fold, than their non-cancerous counterparts.

Orthogonal support for this conclusion comes from a different study (which, because it took a somewhat narrower and less physiologically relevant approach, I give less primacy). 32 driver mutations were engineered into non-cancerous (HEK293T) cell lines, and interactions between each wild-type and mutated protein and a defined set of ~550 other proteins were identified. The mutated proteins had many more interaction partners (340) than the wild-type.

It is not surprising that cancer cells, being different than normal cells, should have some different interactions. It is less obvious that there should be quite *so many* differences. The addition of new interactions numbering *twice as many as the original total* is profound.

It is also surprising that, in addition to there being many *new interactions*, the *total number* of interactions is so much higher in cancer cells. It might have been that cancers gain and lose a similar number, roughly maintaining the total. This is not what we observe. The large increase in the total number of interactions is specifically consistent with the hypothesis of permissive binding. Interactions are abundant at baseline, pruned in normal conditions, and returned to baseline by perturbation.

Second, within a cancer type, relatively few of the new interactions in each cancer line are shared with the other.

This too is consistent with the hypothesis of permissive binding. Perturbations cause permissive states, but different ones: they allow interactions between whatever proteins are present in the cells, determined by their history, their environment, and the precise nature of the perturbation. Different cancers, differing in all of these features, should have different sets of interactions.

The intersection of these interactions may be the key "drivers" of oncogenic function in these cancers, much as we infer that genes recurrently mutated in many independent tumors are the genetic drivers. The others may, similarly, be "passenger interactions".

4.2 New binding interactions may be the root of differential dependencies

A second kind of evidence comes from experiments identifying genes upon which cancer lines have become dependent during their evolution. This is done by systematically knocking genes out or down, one at a time, and determining which are essential for the survival of the cancer line, but not for normal cells (70).

This work has made two points clear. The first is that there is substantial variation between cancer lines in which genes have become essential, a phenomenon termed "differential dependence." The second is that, while some of the differential dependencies in a given cancer can be explained by the mutations that it bears (for example, gain-of-function mutations to canonical driver genes often result in their becoming essential in "oncogene addiction"), many cannot. One recent analysis of 769 cancer cell lines found 550 genes with differential dependence, of which only 127 – less than a quarter – could be explained by mutations. Adding the cancer type in question to the analysis increased power, allowing 227 of the 550 dependencies to be accounted for (71)– still less than half.

Differentially dependent genes are promising as a clue to the basis of the new abilities gained by cancer cells. Because their role differs dramatically between cancers and their ancestor, they may be the key substrates of change.

In the structure of the permissive binding theory, differentially dependent genes are well-suited to be the proteins directly

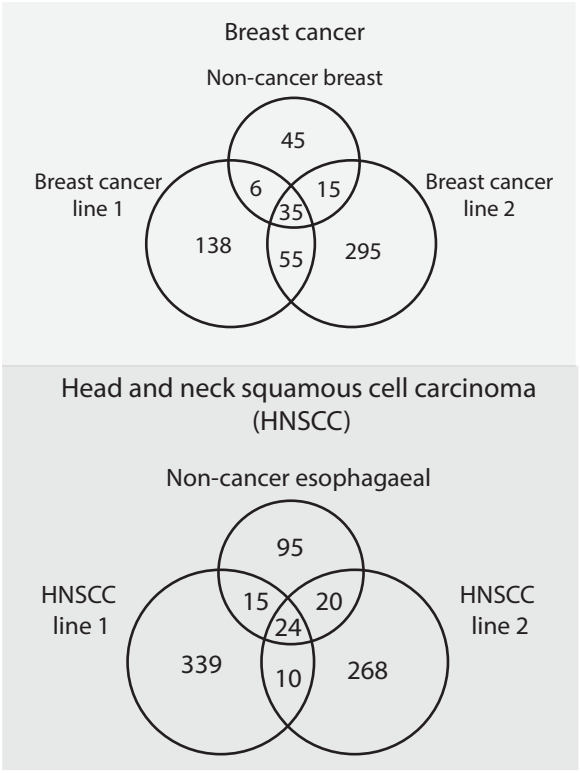


FIGURE 3
The results of two studies (68, 69), that identified protein-protein interactions in cancerous and non-cancerous cell lines. For each of two types of cancer (breast cancer and head and neck squamous cell carcinoma), two cancerous cell lines and one non-cancerous cell line of matched tissue type were profiled. The interactions found in the three lines of each cancer type were then compared to determine how many were shared. The number of unique and shared interactions in each line are depicted in the Venn diagrams.

participating in the novel binding interactions. They are essential specifically in cancer and not in normal cells, due to the functions that they newly drive in cancer. They differ between cancer lines, as new interactions should, having been generated by idiosyncratic endogenous and exogenous perturbations. They are not robustly predictable from mutations, as they result from a wide variety of complex genetic and environmental perturbations. They are much *more* predictable from mutations with the additional context of cell type, which dictates what proteins were present and thus available for binding when the perturbation occurred.

Are differentially dependent genes those involved in the new binding interactions found in cancers? Sadly, there are no experimental data to this effect. But we can ask whether there are known features of these genes suggesting their tendency to be involved in binding interactions in general. A gene ontology analysis of the differentially dependent genes identified in the above analysis reveals that they are remarkably enriched for binding ability. The 5 most statistically significant category enrichments (having removed "molecular function," which, since it is the meta-category selected for use in the analysis, strikes me as trivial) are reproduced (71) as Table 1 below.

The three categories with highest fold enrichment (rather than the lowest p-value; all still statistically significant) are "C-X3-C chemokine binding," "BH3 domain binding," and "alpha-catenin binding." And, of all 109 GO categories found to be significantly enriched, 73 (67%) include the term "binding," compared to only 1878/11236 (17%) of all categories.

4.3 A putative example of the theory in action

There have to my knowledge been no systematic searches for new protein binding interactions that drive the evolved functions essential for cancer, but a beautiful example has been discovered incidentally (72), which I describe below.

Beta-catenin is a transcriptional coactivator that commonly contributes to early oncogenesis. In its canonical role, it is activated in response to wnt signaling, where it acts dependently to displace repressors from transcription factor TCF4, allowing transcription of its target genes. Some colon cancers become dependent on beta-catenin activity following its activation early in their evolution. For example, as shown in Figure 1, mutations reducing the activity of APC, which sequesters beta-catenin, are frequent.

What is the molecular mechanism behind the essential beta-catenin activity in these cancers? One might expect that activation of beta-catenin increases its activity within its normal TCF4 pathway. Surprisingly, a systematic screen for essential genes in these lines failed to find TCF4 pathway genes. Instead, they found

TABLE 1 Results from (71), which first identified "differentially dependent" genes, defined as genes whose essentiality (as determined by causing reduced proliferation when knocked out or down) varies across many cancer cell lines, and then performed a gene ontology (GO) enrichment analysis to determine whether particular functions were enriched among these genes. Shown are the top five most statistically significant GO categories resulting from the analysis.

GO molecular function	Number	Expected number	P-value
Protein binding (GO:000515)	468	361.15	$7.3x10^{-29}$
Binding (GO:0005488)	491	415.98	$7.2x10^{-21}$
Enzyme binding (GO:0019899)	135	57.57	$1.4x10^{-20}$
Protein domain specific binding (GO:0019904)	65	17.83	$3.3x10^{-18}$
Kinase binding (GO:0019900)	67	19.27	$8.8x10^{-18}$

an unexpected dependence on transcriptional coactivator YAP1, canonically involved in Hippo signaling, and transcription factor TBX5. These proteins were shown to assemble into a complex that transcribes antiapoptotic genes including BCL2L1, promoting survival. All three were shown to be necessary for this anti-apoptotic activity and for tumor growth *in vivo*.

Three points are important for our discussion.

First, this is an exquisite case study demonstrating the centrality of a new binding interaction in the survival of a cancer: rather than merely causing more activity in its standard TCF4-related role, beta-catenin forms a new complex containing TBX5 and YAP1. It is relevant to the above data to note that this new binding interaction is the molecular basis of a differential dependence: other colon cancer lines were not dependent on this new complex.

Second, this binding interaction produces function precisely in the manner predicted by the theory. It turns on dependent functions: beta-catenin acts dependently, binding to a coactivator and a transcription factor to bring the transcription factor to DNA, where it activates the existing cellular program of BCL2L1-mediated suppression of apoptosis.

Moreover, the formation of the novel complex was likely dependent on perturbations induced by the driver mutation or subsequent environmental perturbations. Here is a plausible scenario. Following activating mutations, for example to the APC, beta-catenin is more abundant in the cytoplasm. It has a latent propensity to bind YAP1: it does so in the very different conditions of heart development (73), but, even though YAP1 is expressed in most epithelia (72), is normally prevented from doing so in colon cells. The increase in beta-catenin creates the permissive conditions that permit the interaction in this foreign context. Either because this association changes its localization or physicochemical properties, or because of other perturbations to the cell, YAP1 then permissively binds TBX5. Again, it does not normally do so; but, again, it has a latent ability, as shown by experiments in which their binding results from their overexpression (74). The YAP1-TBX5-beta catenin complex is now formed, and binds to the promoters of anti-apoptosis genes. Here again this is not the normal function of the transcription factor TBX5, but again it has latent ability to do so, having been shown to bind these genes when overexpressed (75), and it is likely aided by new binding partner YAP1, which is a coactivator for these targets in other conditions (76). Later in the evolution of such cancers, loss of beta-catenin's normal binding partner TCF4 can further increase the amount available to join this complex, explaining why TCF4 loss promotes tumor progression (77), an observation that, when TCF4 itself was thought to be co-complexed with beta-catenin as part of the key oncogenic complex, was difficult to explain.

Though not discovered with the theory in mind, this looks enormously like a demonstration of it. Here is a case in which new binding interactions, enabled by perturbations to the cell, activate an existing cellular program, driving the differential dependency that reflects this cancer-specific adaptation. A graphical summary showing a simplified version of these findings, and their casting in terms of the mechanistic steps postulated by the theory, is shown below in Figure 4.

Stumbling across a phenomenon at first thought to be unusual has often been the tip of the iceberg, preceding realization that it is actually pervasive. Examples like this may be the rule rather than the exception.

4.4 Summary of the available data

The existing data indicate three things. First, that a new binding interaction underlies at least one ability – the repression of apoptosis – crucial for the evolution of one cancer, through molecular mechanisms consistent with those predicted by the theory. Second, that there are many new binding interactions present in cancers. Third, that genes responsible for a large number of essential cancer-specific abilities are enriched for known functions related to binding. These, to my mind, form a promising initial basis for belief in the utility of the theory, warranting further testing.

5 Testing the theory

An essential feature of the permissive binding theory is that it makes concrete and testable predictions. These are:

1. Cancer cells have more protein interactions than non-cancerous cells.
2. A subset (of unknown size) of these cancer-specific protein interactions cause cancer-specific phenotypes, such as invasion, metastasis, colonization, etc.

Prediction 2 is in many contexts the most important consequence of the theory, but could be true for reasons other than the specific mechanism of permissivity proposed here, which is tested by prediction 1.

Small-scale tests of prediction 1, including the studies discussed above, have already been performed, but larger-scale validation

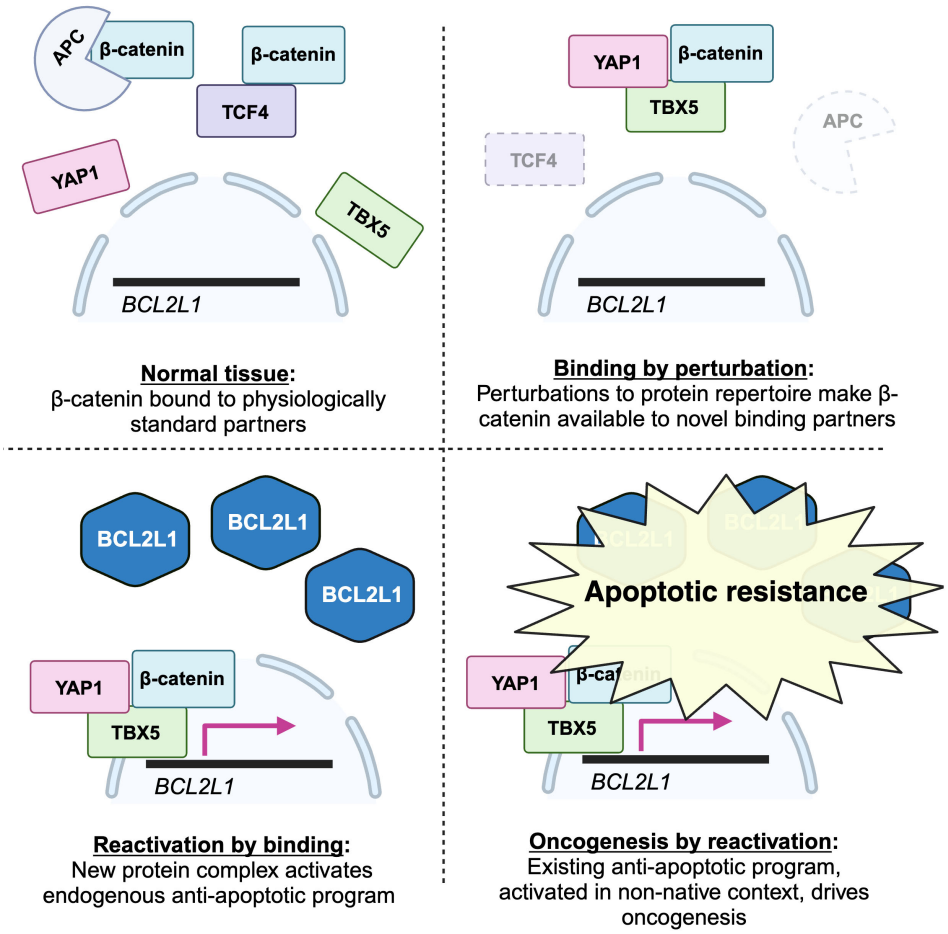


FIGURE 4
A graphical depiction of my interpretation of the results in (72) within the framework of the permissive binding theory. The first event in colorectal cancer progression occurs when deletion or disruption of the APC, a canonical driver mutation, perturbs the cell by decreasing the levels of normal interaction partners of beta-catenin. This leaves beta-catenin free to form new binding interactions with non-standard partners TBX5 and YAP1. Once this novel complex is formed, it can bind to promoters of anti-apoptotic genes like BCL21L, which beta-catenin does not regulate in normal physiological contexts, and drive their expression. This switches on the anti-apoptotic program, which, repressed in normal physiological conditions, confers the new ability of apoptotic resistance on the developing cancer. (Figure created with BioRender.com).

should follow. This requires systematically identifying the binding interactions in many cancers (and non-cancer controls). In an ideal world, this would be done for a variety of cancer types; for primary tumors and especially metastases, rather than cell lines (this is especially important given the importance that the theory places on environment, which will not likely be exactly recapitulated in cell culture); and at proteome-wide scale. Realistically, the limitations of existing technology make many of these desiderata difficult; studies in cell lines focusing on a subset of proteins may be more realistic. To be maximally informative within these constraints, the number of cell lines should be sizable (more on this below), with multiple lines per cancer type and multiple cancer types represented, and the selected 'bait' proteins should be diverse along the axes of function (including receptors/signal transduction machinery, intracellular signaling hubs, transcription factors, etc.) and known oncogenicity (canonical driver proteins and those not as clearly implicated in cancer). Techniques like those used in the analysis of breast cancer and HNSCC lines discussed above (68, 69), as well as many other types of approaches for which there are proof-of-principle demonstrations (66), may be useful here.

Prediction 2 can be tested directly or indirectly. Direct tests can be achieved in two general ways. The first is to *disrupt* a subset of the interactions identified in prediction 1 and assessing whether, and which, oncogenic properties of the cells are affected. Disrupting protein interactions is known to be difficult, and the best strategy will likely depend on the particular interaction in question. For example, there is a growing repertoire of existing small molecule interaction inhibitors, especially for proteins already of therapeutic interest (e.g. Kras (78)); although made to target specific protein pairs, they could also affect interactions between those proteins and other, new partners. New inhibitors could also be sought via standard screening approaches. Genetic approaches, if a particular binding site is known or likely given existing knowledge, may also be possible. The second direct test is the complementary approach: *inducing* a particular protein interaction in a cell line lacking the oncogenic phenotype of interest and determining whether the interaction confers that

phenotype. This strikes me as the more powerful approach, as it minimizes the probability of misleading side effects from the perturbations suggested above. The task of inducing protein interactions is also challenging, but again, the field is progressing fairly rapidly on this front (79, 80). As a true gold standard for causality, as is the case in molecular genetics, one might hope to employ both the perturbation and induction methods in combination: perturbation followed by induction in the same cell line would function analogously to a "knockout" and a "rescue" experiment, which would be strongly compelling. Once an interaction is disrupted or induced, standard *in vitro* and *in vivo* assays for particular oncogenic phenotypes (like wound closure assays for migration, transwell assays for invasiveness, and transplant assays for metastasis (81)) can be used to assess the effect on phenotype.

An indirect test of prediction 2 may be simpler, as can be seen by analogy to the history of the mutation-centric picture. The hypothesis of the causality of mutations in early cancer turned out not only to be correct, but to be realized in a simple form. Rather than each case of cancer evolving by way of its own unique, *idiosyncratic* set of mutations, there is widespread convergence across cancers, both within and between types, resulting in *shared* mutations. When sequencing revealed these shared mutations in cancer after cancer, much more often than could be explained by chance, they were immediately implicated as causal, even in lieu of direct experimental data. Analogously, if the causal interactions posited by prediction 2 include some that are recurrently converged upon by many different cancers, these shared interactions should be found in cancers more often than is expected by chance, and so should be identifiable with statistical testing of datasets generated from prediction 1.

Here is one possibility for such a test. We consider a null model in which the perturbations that characterize cancer probabilistically cause some set of **I** possible new protein interactions, but that none of these is beneficial in cancer progression, such that each cancer acquires some number from this set at random. In this model, the only reason that an interaction would be found in multiple cancer lines is by chance; if we find interactions that occur more frequently than this chance level, we may suspect that they are causal. Suppose that we identify interactions in N total cancers. Under this model, the probability of finding any particular interaction in one of the cancers **C**, which has acquired $\mathbf{C_m}$ interactions from the possible set of **I** interactions, is $P(C) = \mathbf{C_m/I}$. The number of total cancers, **T**, in which a particular interaction appears is distributed as the sum of N independent Bernoulli distributions, one for each of the N cancers, with cancer-specific parameter $\mathbf{p_m} = \mathbf{C_m/I}$. This distribution is also known as the Poisson binomial distribution, and although its analytical form is complex, per the Lyapunov Central Limit Theorem, it can be well-approximated by a normal distribution at even modest values of N, with the same mean and variance as given by the standard sum of its component distributions (82). We can therefore estimate the total number of cancers T in which we expect to find a particular interaction under this null model by a normal distribution with mean $\sum_{m=1}^{N} p_m = \sum_{m=1}^{N} C_m/I$ and variance $\sum_{m=1}^{N} p_m(1-p_m) = \sum_{m=1}^{N} C_m/I(1-C_m/I)$.

What should we use for the number I, the total number of possible interactions from which cancers sample independently? A strict upper bound is the total number of interactions in the human proteome, the square of the number of human proteins, ~$4x10^8$. Realistically, this is vastly too high: many interactions will likely not be detectable even if present due to low abundance or other technical limitations, and many are likely simply not possible to generate even in a perturbed cell. Suppose that in our **N** cancers, we find a total of **U** unique protein interactions. We might use **U** as the value of **I**. This is obviously an underestimate, but an underestimate has the nice property of making for a conservative statistical test, increasing the probability of convergence under the null model.

We can use this model to perform a rough power calculation: how many lines would we need to profile to see significance? This depends on the details of the above parameters, but using numbers on the order of those from Figure 2: if we have 10 cancers, with 200 interactions each, and only mild overlap between them, such that U is 1800 total interactions, we would be able to detect statistical significance at p=0.05, including multiple test correction, for interactions found in 6 or more cancers. If we increase the number of cancers, but hold constant each cancer's number of interactions and the overlap between them, with 20 cancers, we can detect significance at 0.05 for interactions found in 7 or more. We will likely want to include more cancers than this merely for the sake of including a diversity of types; this is merely meant to demonstrate that the prospect of gaining evidence for causality is feasible from the standpoint of statistics.

As realists, we must admit that there is likely to be some dependence on cancer type, as we see for mutations, and on metastatic site, as colonization of different tissues likely requires different abilities. But there may be some complexes that transcend these features, corresponding to functions needed in most or all metastatic sites (e.g. for immune escape). Convergence within these categories seems likelier, but will increase the risk, for which there is not as clear a counterpart in the mutation-centric picture, that shared environmental features rather than causality produce the shared interactions. The ideal scenario is that most interactions are not shared between cancers, suggesting that the above null model is generally a reasonable one, and that a few standouts are widespread. This is more or less what is seen in Figure 2: most interactions are not shared between the cancer lines, suggesting that there is not widespread convergence due to environment alone.

A negative result would not disprove the theory; it could merely be that interactions are idiosyncratic rather than shared. In this case, experimental perturbations, as described above, will be necessary.

6 Conceptual and practical features of the theory

I will now mention two features of the theory that are not related to its probability of being correct, but that are of conceptual and practical interest respectively.

6.1 Compatibility with other theories

An important difference between the present theory and many alternatives is that it includes a causal picture that identifies not only the *ultimate* causes of oncogenesis (here, initiating mutations or environmental perturbations), but its *proximal* causes: novel

binding interactions. Beyond merely including both, the theory is primarily concerned with, and most specific about, this proximal cause. It is essentially agnostic on, and can accommodate wide variation in, the identity and relative importance of ultimate causes.

By contrast, most alternatives are hypotheses about ultimate causes. For example, epigenetic theories may invoke changes in DNA methylation or histone modifications. These leave unspecified the details of *exactly how* these changes produce oncogenic functions.

This difference in level of specification means that many alternative theories can be subsumed without modification under the permissive binding theory. For example, changes to DNA methylation may perturb transcription, leading to imbalances in the levels of proteins present in the cell. In the language of the present theory, DNA methylation may perturb the cell to create a permissive state. The same subsumption can be performed for less mainstream theories, like the genome architecture theory (83), which proposes that oncogenesis is due to dramatic and rapid genome rearrangements, or the tissue organization field theory (84), which proposes that oncogenesis is due to microenvironmental perturbations in the surrounding tissue.

All of these theories may be true to some degree. The permissive binding theory offers a unifying substrate onto which these diverse causes converge in producing their effects. This substrate – protein binding – is well-positioned to both explain and predict the observed effects, being the direct effector of cellular functions.

6.2 Implications for cancer therapy

The theory clearly suggests that targeting the novel protein interactions that cause key oncogenic phenotypes may be of therapeutic benefit. Binding interactions are notoriously difficult to target with small molecules (85), to the extent that they have been referred to as "undruggable" (86), but recent work on the problem, has shown promise (87).

That targeting protein-protein interactions could be a fruitful therapeutic avenue in cancer is not a new idea (88). The permissive binding theory merely emphasizes their importance, arguing that, because they are the most directly causal alteration in cancer cells, they are the single best point of attack.

Importantly, the permissive binding theory also suggests that the causal interactions in cancer are *novel* ones, not present in non-cancerous cells of the same tissue. If true, it is possible that side effects to therapeutics targeting these interactions will be less profound than those resulting from other therapies, which target pathways and mechanisms also found in normal cells.

7 Discussion: hope for simplicity

The permissive binding theory is similar in shape to the mutation-centric picture. Both posit discrete entities (mutations, binding interactions) that change during the evolution of a cancer to directly produce oncogenic functions. But, as we have seen, the consequences of taking protein interactions as the central object in a framework of cancer are quite different than those implied by mutations alone. I think it is better-suited than the mutation picture to explain the invasion-metastasis cascade on conceptual grounds, and am encouraged by initial empirical support. The theory's biggest weakness is that this support is limited to the very small amount of available data that are suited to test it. I hope that this analysis inspires further tests, including but not limited to those suggested above.

In some ways, the permissive binding theory differs from the mutation-centric picture primarily in its esteem for the complexity of biology. It does not limit itself to a picture in which the cell operates akin to clean cartoons in textbooks, in which proteins interact only with their designated playmates, proceeding directly to them and carrying out their jobs in deterministic sequences. It opens itself to one that is more realistic to the physical and evolutionary mechanisms that undergird living systems: movements and interactions of proteins are stochastic, and the cell is only exact as it needs to be, dictated by the normal physiological contexts upon which natural selection primarily acts.

It might also be thought of as differing in accepting the limitations of our knowledge. Proteins are in most cases the final output layer of the cell: they are what makes biology go. The causes of phenotypes must, in the end, be descriptions of *what proteins are doing*. In an ideal world, we would be able to input state variables of any cell – its mutational repertoire, its physical environment, its transcriptional state, and so on – and output what its proteins are doing. But we live far from this ideal world. The permissive binding theory accepts this reality, relinquishing hope of being able to squint through the narrow lens of mutations and accepting the need, more technically difficult though it may be, to look to the proteins themselves to tell us what they are doing.

The permissive binding theory is here discussed in the case of cancer, but has applications beyond it. It may underlie pathological mechanisms in chronic diseases in which the physiological environment is altered as either cause or consequence, like inflammation, diabetes, and cardiovascular disease. Moreover, it is primarily an evolutionary theory, developed to address the evolutionary puzzle of invasion and metastasis, which it does by positing generic perturbations as a source of "unlocking" new functions. This offers a general mechanism that may be at play in other evolutionary contexts.

The permissive binding theory proposes that the waters of our attempts to understand cancer evolution have appeared so muddy because we have focused on the wrong entities. It offers a different level of analysis, at which formerly disjointed observations and mechanisms may collapse into a coherent whole. It is well-positioned to do this, being a claim about the interactions between proteins, which are the final effector of cellular behavior, and so are a natural point of convergence for observations regarding varied underlying mechanisms. Heterogeneous molecular changes, varied in kind and inconsistent in occurrence, that characterize cancers – point mutations, epigenetic mutations, aneuploidies, polyploidies, extrachromosomal DNA, transcriptional changes, translational changes, posttranslational modification changes, metabolite changes, morphological changes – may converge in

their effects on the proteins: by different mechanisms, they may produce the same few key complexes.

I think the theory's biggest virtue is that, if it is true, things could turn out to be simpler than they now appear (89).

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8979493/bin/mmc4.xlsx.

Author contributions

CW: Conceptualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. CW is funded as a Lewis-Sigler Scholar by the Lewis-Sigler Institute of Integrative Genomics at Princeton University.

Acknowledgments

I am grateful to Michael MacArthur, Jacob Boyer, and members of the Ayroles and Akey labs, especially Diogo Melo and Julien Ayroles, for thought and discussion that aided in the development of these ideas, and for feedback on the manuscript.

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.

Publisher's note

References

1. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr., Kinzler KW. Cancer genome landscapes. *Science* (2013) 339:1546–58. doi: 10.1126/science.1235122

2. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* (1990) 61:759–67. doi: 10.1016/0092-8674(90)90186-i

3. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging biological principles of metastasis. *Cell* (2017) 168:670–91. doi: 10.1016/j.cell.2016.11.037

4. Bernards R, Weinberg RA. Metastasis genes: A progression puzzle. *Nature* (2002) 418:823–3. doi: 10.1038/418823a

5. Vanharanta S, Massagué J. Origins of metastatic traits. *Cancer Cell* (2013) 24:410–21. doi: 10.1016/j.ccr.2013.09.007

6. Nguyen B, Fong C, Luthra A, Smith SA, DiNatale RG, Nandakumar S, et al. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* (2022) 185:563–75.e11. doi: 10.1016/j.cell.2022.01.003

7. Martínez-Jiménez F, Movasati A, Brunner S, Nguyen L, Priestley P, Cuppen E, et al. Pan-cancer whole genome comparison of primary and metastatic solid tumors. *bioRxiv* (2022) 2022:6.17.496528. doi: 10.1101/2022.06.17.496528

8. Jacob LS, Vanharanta S, Obenauf AC, Pirun M, Viale A, Socci ND, et al. Metastatic competence can emerge with selection of preexisting oncogenic alleles without a need of new mutations. *Cancer Res* (2015) 75:3713–9. doi: 10.1158/0008-5472.can-15-0562

9. Fidler IJ, Kripke ML. Metastasis results from preexisting variant cells within a Malignant tumor. *Science* (1977) 197:893–5. doi: 10.1126/science.887927

10. Brabletz T, Kalluri R, Nieto MA, Weinberg RA. EMT in cancer. *Nat Rev Cancer* (2018) 18:128–34. doi: 10.1038/nrc.2017.118

11. Mani SA, Guo W, Liao M-J, Eaton E, Ayyanan A, Zhou AY, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* (2008) 133:704–15. doi: 10.1016/j.cell.2008.03.027

12. Niu N, Mercado-Uribe I, Liu J. Dedifferentiation into blastomere-like cancer stem cells *via* formation of polyploid giant cancer cells. *Oncogene* (2017) 36:4887–900. doi: 10.1038/onc.2017.72

13. Amend SR, Torga G, Lin K, Kostecka LG, Marzo A, Austin RH, et al. Polyploid giant cancer cells: Unrecognized actuators of tumorigenesis, metastasis, and resistance. *Prostate* (2019) 79:1489–97. doi: 10.1002/pros.23877

14. Ben-David U, Amon A. Context is everything: aneuploidy in cancer. *Nat Rev Genet* (2020) 21:44–62. doi: 10.1038/s41576-019-0171-x

15. Cortés-Ciriano I, Lee JJ-K, Xi R, Jain D, Jung YL, Yang L, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet* (2020) 52:331–41. doi: 10.1038/s41588-019-0576-7

16. Bussard KM, Smith GH. Human breast cancer cells are redirected to mammary epithelial cells upon interaction with the regenerating mammary gland microenvironment *in-vivo*. *PloS One* (2012) 7:e49221. doi: 10.1371/journal.pone.0049221

17. Mintz B, Illmensee K. Normal genetically mosaic mice produced from Malignant teratocarcinoma cells. *Proc Natl Acad Sci* (1975) 72:3585–9. doi: 10.1073/pnas.72.9.3585

18. Illmensee K, Mintz B. Totipotency and normal differentiation of single teratocarcinoma cells cloned by injection into blastocysts. *Proc Natl Acad Sci* (1976) 73:549–53. doi: 10.1073/pnas.73.2.549

19. Jain M, Arvanitis C, Chu K, Dewey W, Leonhardt E, Trinh M, et al. Sustained loss of a neoplastic phenotype by brief inactivation of MYC. *Science* (2002) 297:102–4. doi: 10.1126/science.1071489

20. Zhu S, Wang J, Zellmer L, Xu N, Liu M, Hu Y, et al. Mutation or not, what directly establishes a neoplastic state, namely cellular immortality and autonomy, still remains unknown and should be prioritized in our research. *J Cancer* (2022) 13:2810–43. doi: 10.7150/jca.72628

21. Jacob F. Evolution and tinkering. *Science* (1977) 196:1161–6. doi: 10.1126/science.860134

22. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. *Science* (2011) 331:1559–64. doi: 10.1126/science.1203543

23. Hanahan D, Folkman J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* (1996) 86:353–64. doi: 10.1016/s0092-8674(00)80108-7

24. Wang R, Chadalavada K, Wilshire J, Kowalik U, Hovinga KE, Geber A, et al. Glioblastoma stem-like cells give rise to tumour endothelium. *Nature* (2010) 468:829–33. doi: 10.1038/nature09624

25. Crick F. *Polypeptides and Proteins: X-Ray Studies*. (1953). Gonville and Caius College, University of Cambridge, Cambridge, UK.

26. Pawson T, Raina M, Nash P. Interaction domains: from simple binding events to complex cellular behavior. *FEBS Lett* (2002) 513:2–10. doi: 10.1016/s0014-5793(01)03292-6

27. Frumkin I, Laub MT. Selection of a de novo gene that can promote survival of E. coli by modulating protein homeostasis pathways. bioRxiv (2023). doi: 10.1101/2023.02.07.527531.

28. Eicholt LA, Aubel M, Berk K, Bornberg-Bauer E, Lange A. Chaperones facilitate heterologous expression of naturally evolved putative *de novo* proteins. *bioRxiv* (2022). doi: 10.1101/2022.03.02.482622

29. Heames B, Buchel F, Aubel M, Tretyachenko V, Loginov D, Novák P, et al. Experimental characterization of *de novo* proteins and their unevolved random-sequence counterparts. *Nat Ecol Evol* (2023) 7:570–80. doi: 10.1038/s41559-023-02010-2

30. Lite T-LV, Grant RA, Nocedal I, Littlehale ML, Guo MS, Laub MT. Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *eLife* (2020) 9:e60924. doi: 10.7554/elife.60924

31. Ghose DA, Przydzial KE, Mahoney EM, Keating AE, Laub MT. Marginal specificity in protein interactions constrains evolution of a paralogous family. *Proc Natl Acad Sci* (2023) 120:e2221163120. doi: 10.1073/pnas.2221163120

32. Nocedal I, Laub MT. Ancestral reconstruction of duplicated signaling proteins reveals the evolution of signaling specificity. *eLife* (2022) 11:e77346. doi: 10.7554/elife.77346

33. Serebriiskii IG, Golemis EA. Two-hybrid systems, methods and protocols. *Methods Mol Biol (Clifton NJ)* (2001) 177:123–34. doi: 10.1385/1-59259-210-4:123

34. Golemis EA, Serebriiskii I, Law SF. The yeast two-hybrid system: criteria for detecting physiologically significant protein-protein interactions. *Curr Issues Mol Biol* (1999) 11(1):34–46. doi: 10.21775/cimb.001.031

35. Tanowitz M, von Zastrow M. Receptor signal transduction protocols. *Methods Mol Biol (Clifton NJ)* (2004) 259:353–69. doi: 10.1385/1-59259-754-8:353

36. Ozturk K, Carter H. Predicting functional consequences of mutations using molecular interaction network features. *Hum Genet* (2022) 141:1195–210. doi: 10.1007/s00439-021-02329-5

37. Charitou T, Srihari S, Lynn MA, Jarboui M-A, Fasterius E, Moldovan M, et al. Transcriptional and metabolic rewiring of colorectal cancer cells expressing the oncogenic KRASG13D mutation. *Br J Cancer* (2019) 121:37–50. doi: 10.1038/s41416-019-0477-7

38. Hammond DE, Mageean CJ, Rusilowicz EV, Wickenden J, Clague MJ, Prior IA. Differential reprogramming of isogenic colorectal cancer cells by distinct activating KRAS mutations. *J Proteome Res* (2015) 14:1535–46. doi: 10.1021/pr501191a

39. Yurchenco PD. Basement membranes: cell scaffoldings and signaling platforms. *Cold Spring Harb Perspect Biol* (2011) 3:a004911. doi: 10.1101/cshperspect.a004911

40. Almagro J, Messal HA, Elosegui-Artola A, Rheenen van J, Behrens A. Tissue architecture in tumor initiation and progression. *Trends Cancer* (2022) 8:494–505. doi: 10.1016/j.trecan.2022.02.007

41. Balkwill FR, Capasso M, Hagemann T. The tumor microenvironment at a glance. *J Cell Sci* (2012) 125:5591–6. doi: 10.1242/jcs.116392

42. Ishihara S, Haga H. Matrix stiffness contributes to cancer progression by regulating transcription factors. *Cancers* (2022) 14:1049. doi: 10.3390/cancers14041049

43. Rice AJ, Cortes E, Lachowski D, Cheung BCH, Karim SA, Morton JP, et al. Matrix stiffness induces epithelial–mesenchymal transition and promotes chemoresistance in pancreatic cancer cells. *Oncogenesis* (2017) 6:e352. doi: 10.1038/oncsis.2017.54

44. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* (2017) 171:1611–1624.e24. doi: 10.1016/j.cell.2017.10.044

45. Maffini MV, Soto AM, Calabro JM, Ucci AA, Sonnenschein C. The stroma as a crucial target in rat mammary gland carcinogenesis. *J Cell Sci* (2004) 117:1495–502. doi: 10.1242/jcs.01000

46. Scully T, Ettela A, LeRoith D, Gallagher EJ. Obesity, type 2 diabetes, and cancer risk. *Front Oncol* (2021) 10:615375. doi: 10.3389/fonc.2020.615375

47. Marusyk A, Polyak K. Tumor heterogeneity: Causes and consequences. *Biochim Biophys Acta (BBA) - Rev Cancer* (2010) 1805:105–17. doi: 10.1016/j.bbcan.2009.11.002

48. Weinberg RA. Coevolution in the tumor microenvironment. *Nat Genet* (2008) 40:494–5. doi: 10.1038/ng0508-494

49. Zhao Z-M, Zhao B, Bai Y, Iamarino A, Gaffney SG, Schlessinger J, et al. Early and multiple origins of metastatic lineages within primary tumors. *Proc Natl Acad Sci* (2016) 113:2140–5. doi: 10.1073/pnas.1525677113

50. Kenny PA, Bissell MJ. Tumor reversion: Correction of Malignant behavior by microenvironmental cues. *Int J Cancer* (2003) 107:688–95. doi: 10.1002/ijc.11491

51. Li Q, Wennborg A, Aurell E, Dekel E, Zou J-Z, Xu Y, et al. Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape. *Proc Natl Acad Sci* (2016) 113:2672–7. doi: 10.1073/pnas.1519210113

52. KAUFFMAN S. Homeostasis and differentiation in random genetic control networks. *Nature* (1969) 224:177–8. doi: 10.1038/224177a0

53. Kumar R, Sharma AK. Transcription factor stoichiometry in cell fate determination. *J Genet* (2021) 100:27. doi: 10.1007/s12041-021-01278-2

54. Clark NM, Fisher AP, Berckmans B, den Broeck LV, Nelson EC, Nguyen TT, et al. Protein complex stoichiometry and expression dynamics of transcription factors modulate stem cell division. *Proc Natl Acad Sci* (2020) 117:15332–42. doi: 10.1073/pnas.2002166117

55. Muraoka N, Ieda M. Stoichiometry of transcription factors is critical for cardiac reprogramming. *Circ Res* (2015) 116:216–8. doi: 10.1161/circresaha.114.305696

56. Graf T. Transcription factor stoichiometry drives cell fate: single-cell proteomics to the rescue. *Cell Stem Cell* (2019) 24:673–4. doi: 10.1016/j.stem.2019.03.002

57. Belpomme D, Irigaray P, Hardell L, Clapp R, Montagnier L, Epstein S, et al. The multitude and diversity of environmental carcinogens. *Environ Res* (2007) 105:414–29. doi: 10.1016/j.envres.2007.07.002

58. Lopez-Bigas N, Gonzalez-Perez A. Are carcinogens direct mutagens? *Nat Genet* (2020) 52:1137–8. doi: 10.1038/s41588-020-00730-w

59. Senger G, Santaguida S, Schaefer MH. Regulation of protein complex partners as a compensatory mechanism in aneuploid tumors. *eLife* (2022) 11:e75526. doi: 10.7554/elife.75526

60. Dürrbaum M, Storchová Z. Effects of aneuploidy on gene expression: implications for cancer. *FEBS J* (2016) 283:791–802. doi: 10.1111/febs.13591

61. Zhang S, Mercado-Uribe I, Xing Z, Sun B, Kuang J, Liu J. Generation of cancer stem-like cells through the formation of polyploid giant cancer cells. *Oncogene* (2014) 33:116–28. doi: 10.1038/onc.2013.96

62. Haigis KM, Cichowski K, Elledge SJ. Tissue-specificity in cancer: The rule, not the exception. *Science* (2019) 363:1150–1. doi: 10.1126/science.aaw3472

63. Ohnishi K, Semi K, Yamamoto T, Shimizu M, Tanaka A, Mitsunaga K, et al. Premature termination of reprogramming *in vivo* leads to cancer development through altered epigenetic regulation. *Cell* (2014) 156:663–77. doi: 10.1016/j.cell.2014.01.005

64. Weisman CM. The origins and functions of *de novo* genes: against all odds? *J Mol Evol* (2022) 90:244–57. doi: 10.1007/s00239-022-10055-3

65. Ivanov AA, Revennaugh B, Rusnak L, Gonzalez-Pecchi V, Mo X, Johns MA, et al. The OncoPPi Portal: an integrative resource to explore and prioritize protein–protein interactions for cancer target discovery. *Bioinformatics* (2017) 34:1183–91. doi: 10.1093/bioinformatics/btx743

66. Tabar MS, Francis H, Yeo D, Bailey CG, Rasko JEJ. Mapping oncogenic protein interactions for precision medicine. *Int J Cancer* (2022) 151:7–19. doi: 10.1002/ijc.33954

67. Li Z, Ivanov AA, Su R, Gonzalez-Pecchi V, Qi Q, Liu S, et al. The OncoPPi network of cancer-focused protein–protein interactions to inform biological insights and therapeutic strategies. *Nat Commun* (2017) 8:14356. doi: 10.1038/ncomms14356

68. Swaney DL, Ramms DJ, Wang Z, Park J, Goto Y, Soucheray M, et al. A protein network map of head and neck cancer reveals PIK3CA mutant drug sensitivity. *Science* (2021) 374:eabf2911. doi: 10.1126/science.abf2911

69. Kim M, Park J, Bouhaddou M, Kim K, Rojc A, Modak M, et al. A protein interaction landscape of breast cancer. *Science* (2021) 374:eabf3066. doi: 10.1126/science.abf3066

70. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. *Cell* (2017) 170:564–576.e16. doi: 10.1016/j.cell.2017.06.010

71. Park TY, Leiserson MDM, Klau GW, Raphael BJ. SuperDendrix algorithm integrates genetic dependencies and genomic alterations across pathways and cancer types. *Cell Genom* (2022) 2:100099. doi: 10.1016/j.xgen.2022.100099

72. Rosenbluh J, Nijhawan D, Cox AG, Li X, Neal JT, Schafer EJ, et al. β-catenin-driven cancers require a YAP1 transcriptional complex for survival and tumorigenesis. *Cell* (2012) 151:1457–73. doi: 10.1016/j.cell.2012.11.026

73. Heallen T, Zhang M, Wang J, Bonilla-Claudio M, Klysik E, Johnson RL, et al. Hippo pathway inhibits wnt signaling to restrain cardiomyocyte proliferation and heart size. *Science* (2011) 332:458–61. doi: 10.1126/science.1199010

74. Murakami M, Nakagawa M, Olson EN, Nakagawa O. A WW domain protein TAZ is a critical coactivator for TBX5, a transcription factor implicated in Holt–Oram syndrome. *Proc Natl Acad Sci* (2005) 102:18034–9. doi: 10.1073/pnas.0509109102

75. He A, Kong SW, Ma Q, Pu WT. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc Natl Acad Sci* (2011) 108:5632–7. doi: 10.1073/pnas.1016959108

76. LeBlanc L, Lee B-K, Yu AC, Kim M, Kambhampati AV, Dupont SM, et al. Yap1 safeguards mouse embryonic stem cells from excessive apoptosis during differentiation. *eLife* (2018) 7:e40167. doi: 10.7554/elife.40167

77. Angus-Hill ML, Elbert KM, Hidalgo J, Capecchi MR. T-cell factor 4 functions as a tumor suppressor whose disruption modulates colon cell proliferation and tumorigenesis. . *Proc Natl Acad Sci* (2011) 108:4914–9. doi: 10.1073/pnas.1102300108

78. Harwood SJ, Smith CR, Lawson JD, Ketcham JM. Selected approaches to disrupting protein–protein interactions within the MAPK/RAS pathway. *Int J Mol Sci* (2023) 24:7373. doi: 10.3390/ijms24087373

79. Marchand A, Hall-Beauvais AKV, Correia BE. Computational design of novel protein–protein interactions – An overview on methodological approaches and applications. *Curr Opin Struct Biol* (2022) 74:102370. doi: 10.1016/j.sbi.2022.102370

80. Weagel EG, Foulks JM, Siddiqui A, Warner SL. Molecular glues: enhanced protein-protein interactions and cell proteome editing. *Med Chem Res* (2022) 31:1068–87. doi: 10.1007/s00044-022-02882-2

81. Bouchalova P, Bouchal P. Current methods for studying metastatic potential of tumor cells. *Cancer Cell Int* (2022) 22:394. doi: 10.1186/s12935-022-02801-w

82. Neammanee K. A refinement of normal approximation to Poisson binomial. *Int J Math Math Sci* (2005) 2005:717–28. doi: 10.1155/ijmms.2005.717

83. Heng J, Heng HH. Genome chaos, information creation, and cancer emergence: searching for new frameworks on the 50th anniversary of the "War on cancer". *Genes* (2021) 13:101. doi: 10.3390/genes13010101

84. Soto AM, Sonnenschein C. The tissue organization field theory of cancer: A testable replacement for the somatic mutation theory. *BioEssays* (2011) 33:332–40. doi: 10.1002/bies.201100025

85. Lu H, Zhou Q, He J, Jiang Z, Peng C, Tong R, et al. Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Signal Transduct Target Ther* (2020) 5:213. doi: 10.1038/s41392-020-00315-3

86. Dang CV, Reddy EP, Shokat KM, Soucek L. Drugging the "undruggable" cancer targets. *Nat Rev Cancer* (2017) 17:502–8. doi: 10.1038/nrc.2017.36

87. Wu D, Li Y, Zheng L, Xiao H, Ouyang L, Wang G, et al. Small molecules targeting protein–protein interactions for cancer therapy. *Acta Pharm Sin B* (2023) 13 (10):4060–88. doi: 10.1016/j.apsb.2023.05.035

88. Ivanov AA, Khuri FR, Fu H. Targeting protein–protein interactions as an anticancer strategy. *Trends Pharmacol Sci* (2013) 34:393–400. doi: 10.1016/j.tips.2013.04.007

89. Weinberg RA. Coming full circle—From endless complexity to simplicity and back again. *Cell* (2014) 157:267–71. doi: 10.1016/j.cell.2014.03.004

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 November 2023
DOI 10.3389/fonc.2023.1213678

OPEN ACCESS

EDITED BY
Tao Liu,
University of New South Wales, Australia

REVIEWED BY
Bruce A. Shapiro,
National Cancer Institute at Frederick (NIH), United States
Takeru Hayashi,
Institute of Microbial Chemistry (IMC), Japan

*CORRESPONDENCE
Yusuke Suenaga
ysuenaga@chiba-cc.jp
Tatsuhito Matsuo
matsuo.tatsuhito@qst.go.jp

RECEIVED 28 April 2023
ACCEPTED 03 November 2023
PUBLISHED 23 November 2023

CITATION
Yamamoto S, Kono F, Nakatani K, Hirose M, Horii K, Hippo Y, Tamada T, Suenaga Y and Matsuo T (2023) Structural characterization of human *de novo* protein NCYM and its complex with a newly identified DNA aptamer using atomic force microscopy and small-angle X-ray scattering. *Front. Oncol.* 13:1213678. doi: 10.3389/fonc.2023.1213678

Structural characterization of human *de novo* protein NCYM and its complex with a newly identified DNA aptamer using atomic force microscopy and small-angle X-ray scattering

Seigi Yamamoto[1], Fumiaki Kono[2], Kazuma Nakatani[1,3,4,5], Miwako Hirose[6], Katsunori Horii[6], Yoshitaka Hippo[1,3,7], Taro Tamada[2,8], Yusuke Suenaga[1*] and Tatsuhito Matsuo[2*]

[1]Laboratory of Evolutionary Oncology, Chiba Cancer Center Research Institute, Chiba, Japan, [2]Institute for Quantum Life Science, National Institutes for Quantum Science and Technology, Chiba, Japan, [3]Graduate School of Medical and Pharmaceutical Sciences, Chiba University, Chiba, Japan, [4]Innovative Medicine CHIBA Doctoral WISE Program, Chiba University, Chiba, Japan, [5]All Directional Innovation Creator Ph.D. Project, Chiba University, Chiba, Japan, [6]Digital Healthcare Business Development Office, NEC Solution Innovators, Ltd., Tokyo, Japan, [7]Laboratory of Precision Tumor Model Systems, Chiba Cancer Center Research Institute, Chiba, Japan, [8]Graduate School of Science, Chiba University, Chiba, Japan

NCYM, a Homininae-specific oncoprotein, is the first *de novo* gene product experimentally shown to have oncogenic functions. NCYM stabilizes MYCN and β-catenin via direct binding and inhibition of GSK3β and promotes cancer progression in various tumors. Thus, the identification of compounds that binds to NCYM and structural characterization of the complex of such compounds with NCYM are required to deepen our understanding of the molecular mechanism of NCYM function and eventually to develop anticancer drugs against NCYM. In this study, the DNA aptamer that specifically binds to NCYM and enhances interaction between NCYM and GSK3β were identified for the first time using systematic evolution of ligands by exponential enrichment (SELEX). The structural properties of the complex of the aptamer and NCYM were investigated using atomic force microscopy (AFM) in combination with truncation and mutation of DNA sequence, pointing to the regions on the aptamer required for NCYM binding. Further analysis was carried out by small-angle X-ray scattering (SAXS). Structural modeling based on SAXS data revealed that when isolated, NCYM shows high flexibility, though not as a random coil, while the DNA aptamer exists as a dimer in solution. In the complex state, models in which NCYM was bound to a region close to an edge of the aptamer reproduced the SAXS data. Therefore, using a combination of SELEX, AFM, and SAXS, the present study revealed the structural properties of NCYM in its functionally active form, thus providing useful information for the possible future design of novel anti-cancer drugs targeting NCYM.

KEYWORDS

***de novo* protein, NCYM, DNA aptamer, solution structure, AFM, SAXS**

1 Introduction

NCYM, a *cis*-antisense gene of *MYCN*, encodes a Hominae-specific oncoprotein (1, 2). In human neuroblastomas, *NCYM* is always co-amplified with *MYCN*, and its expression level is associated with poor prognosis (1). NCYM stabilizes MYCN via inhibition of GSK3β, whereas MYCN stimulates both *MYCN* and *NCYM* transcription (1). This feedback loop contributes to the maintenance of high levels of both MYCN and NCYM expressions in *MYCN*-amplified neuroblastomas (1, 2). NCYM enhances the metastasis of neuroblastomas (1) possibly via inhibition of apoptotic cell death (1, 3, 4) and/or regulation of stemness (5, 6). Furthermore, NCYM has been shown to be associated with progression of adult cancers (2, 7). Therefore, NCYM is a promising target protein for anti-cancer therapy. However, the difficulty in determining its structure hinders drug design (8).

NCYM is a newly evolved coding gene that originated from *MYCN* promoter region during the evolution of the Hominae (1, 2). New genes originating from non-genic regions are known as *de novo* gene birth (9–11), and NCYM is the first *de novo* gene product experimentally shown to have oncogenic functions. Owing to their *de novo* emergence, *de novo* proteins show no homology to known genes and do not have any domains or motifs. The amino acid sequence of *de novo* proteins is similar to a random sequence (12), although a recent report identified the difference between *de novo* proteins and unevolved random-sequence counterparts in that *de novo* proteins exhibit moderately higher solubility in cells (13). Four *de novo* proteins have been structurally characterized to date: Bsc4 (14), NCYM (8), Goddard (15), and AFGP8 (16); however, mainly because of their highly disordered nature, none of the complete structures have been determined. Upon binding to the ice surface, the local structure of the antifreeze glycoprotein AFGP8 make a transition from a disordered to an ordered state (16), indicating the possibility of significant ordering of *de novo* proteins via complex formation with binding partners. Consistent with an earlier prediction that NCYM binds to DNA (17), we have previously found that benzonase treatment significantly improves the solubility of NCYM (8). These observations led us to identify DNA aptamers that bind specifically to NCYM and to consider that analysis of the complex of NCYM and DNA aptamers may contribute to the characterization of the structural dynamics of NCYM.

Here, three types of DNA aptamers were identified by systematic evolution of ligands by exponential enrichment (SELEX) and their interactions with NCYM were characterized by atomic force microscopy (AFM). Moreover, the structure of the NCYM-DNA complex of a representative DNA aptamer (named "No. 1"), which enhances interaction between NCYM and GSK3β, was analyzed using small-angle X-ray scattering (SAXS).

2 Materials and methods

2.1 Aptamer selection procedure

SELEX was performed as previously reported with some modifications (18, 19). Dynabeads™ MyOne™ Carboxylic Acid (CA) magnetic beads (Invitrogen, Waltham, MA) were used for NCYM solidification to segregate NCYM-binding DNA molecules from the non-binding molecules. The target beads were prepared by an amine coupling reaction using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC; Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions and washed with the selection buffer [SB; 40 mM HEPES (pH 7.5), 125 mM NaCl, 5 mM KCl, 1 mM $MgCl_2$, and 0.01% Tween 20]. Briefly, the CA magnetic beads were washed twice with 500 µl of 100 mM 2-Morpholinoethanesulfonic acid mono hydrate (MES) buffer at pH 6.0. After added 50 µl of 100 mM MES buffer and 50 µl of EDC, and the mixture was incubated for 30 minutes at room temperature. The mixture was then mixed with 200 µg of NCYM in 100 mM MES buffer and incubated over night at room temperature to react with the amino group of NCYM and the carboxylic acid of the CA magnetic beads. Next, they were washed twice with 500 µl of PBST (0.1% Tween20 in phosphate buffered saline, PBS) and 500 µl of PBST-BSA (0.1% bovine serum albumin in PBST) was added. Before using the target beads, they were washed twice with 1 ml of SB.

An initial single-stranded DNA (ssDNA) pool, 5′-GGAATGTGGTCCCTCGCAATAAATC-(N30)-GAAATGAGCCCTTTGACCCTGTAC-3′, containing 30 random nucleotides between forward (Fw) and revers (Rv) primer region was purchased from Integrated DNA Technologies (Tokyo, Japan). The selection of aptamers was performed starting from 4.5 nmol of DNAs (~10^{15} molecules) in 100 µl of SB. The pool was mixed for 15 minutes with 250 µg of target beads at 25°C. The beads were then washed with SB, and the bound ssDNA was eluted with 7 M urea. After recovery of the eluted ssDNA using Rv primer beads, polymerase chain reaction (PCR) was performed with KOD Dash DNA polymerase (Toyobo, Osaka, Japan), a forward (Fw) primer (5'-GGAATGTGGTCCCTCGCAATAAATC-3') and a reverse (Rv) primer (5'-GTACAGGGTCAAAGGGCTCATTTC-3') with modification by the 5'-biotin. Next, the amplified double stranded DNA (dsDNA) was bound to MyOne SA C1 magnetic beads, and the Fw chain (ssDNA) was eluted with 0.02M NaOH. The ssDNA was used for the next round.

After eight rounds of selection, the frequency of ssDNA sequences was determined by next-generation sequencing (NGS) from rounds 3 to 8 of SELEX using a MiniSeq System (Illumina, San Diego, CA, USA). Sequencing data were preprocessed by using the program of PRINSEQ++ (20) and adopted above 99.9% of the base calling accuracy (Q score of 30 and above).

2.2 Bio-layer interferometry

All the Bio-Layer Interferometry (BLI) measurements were performed at 25°C using an Octet® RED96e system (Sartorius AG, Goettingen, Germany). All the samples were placed in a 96 well microplate and the sample volume was 200 µl/well. The microplate was shaked at 1,000 rpm during the measurement. As a ligand, each aptamer with 5'-biotin modification was immobilized on an Octet® SA biosensor chip (Sartorius). For kinetics analysis, different concentrations of NCYM (25-400nM) were used. The dissociation constants between the aptamer and NCYM were calculated using a

simple 1:1 biomolecular interaction model according to the manufacturer's instructions.

2.3 Construction of DNA frame structures and introduction of aptamers for AFM imaging

To gain insights into the mechanism of interactions between NCYM and the aptamers obtained by SELEX, NCYM-aptamer binding was studied at the single molecule level using AFM. Scaffolds prepared using the DNA origami method were used for this purpose. The DNA origami method allows the creation of structures of any shape and the introduction of functional molecules anywhere in the structure. Therefore, single-molecule observation using DNA origami structures is suitable for evaluating biomolecules and has been used to observe various molecules (21–29). In this study, we used a DNA frame structure (21). The DNA frame contains a space inside, and dsDNA can be introduced into any sequence.

DNA frames were prepared as previously described (21). A solution containing 10 nM M13mp18 ssDNA (tilibit nanosystems GmbH, Germany), 25 nM staples (2.5 eq), 20 mM Tris-HCl (pH 7.4), 10 mM $MgCl_2$, 1mM EDTA was prepared and annealed at a rate of -1°C per minute from 85°C to 15°C.

Aptamers and DNA oligos for the DNA frame were purchased from Eurofins Genomics K.K. (Tokyo, Japan) and used without further purification. In this study, three types of aptamers, which showed high affinity among the aptamers obtained by SELEX, were employed, and the sequences of which for AFM observation are as follows:

No. 1: 5'-GGAATGTGGTCCCTCGCAATAAATCTATGTA CGTTATTCCCCTTTGACC
AATGCTGAAATGAGCCCTTTGACCCTGTAC **TTTT**TTTCCAGCGGGACTAGCGCGTTGCTC CTCACT-3'

No. 2: 5'-GGAATGTGGTCCCTCGCAATAAATC GGGGAGGGAGGGTGGGGGCGGT
GGGAGGTGGAAATGAGCCCTTTGACCCTGTAC **TTTT**TTTCCAGCGGGACTAGCGCGTTGCTCCT CACT-3'

No. 3: 5'-GGAATGTGGTCCCTCGCAATAAATCGGGCG TTGTGGAGGGGGCGGTG
GGTGGGGGGGAAATGAGCCCTTTGA CCCTGTAC **TTTT** TTTCCAGCGGG ACTAGCGCGTTGCTCCTCACT-3'

All aptamers had a dsDNA complement sequence (underlined) added via a TTTT sequence (bold) at the 3' end for introduction into the DNA frame.

The secondary structures of the aptamers were predicted using RNAfold [http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi]. We used the default settings: minimum free energy (MFE) and partition function, yes; avoid isolated base pairs, yes; incorporate G-Quadruplex formation into the structure prediction algorithm, yes; dangling end options, dangling energies on both sides of a helix in any case; energy parameters, DNA parameters (Mathews model, 2004) at 25 or 38°C with a salt concentration of 1.021M.

The resultant dsDNAs were incorporated into the frame structures described above and observed in the presence of NCYM (Figure S1), which was purified according to a previous method (1).

2.4 Truncation and mutation of the DNA aptamer No. 1

To identify where the aptamers interact with NCYM, we prepared three analogs (short 1, short 2, and short 3) of the No. 1 aptamer based on the secondary structure prediction by RNAfold. The sequences used are as follows:

short 1: 5'-TGGTCCCTCGCAATAAATCTATGTA CGTTATTCCCCTTTGACCAATGC
TGAAATGAGC**TTTT**TTTCCAGCGGGACTAGCG CGTTGCTCCTCACT-3'

short 2: 5'-GGAATGTGGTCCCTCGCAATAAA TCTATGTACGTTATTCCCCTTTGAC
CAATGCCCTTTGACCCTGTAC**TTTT** TTTCCAGCGGGACTAGCGCGTTGCTCCTCACT-3'

short 3: 5'-GGAATGTGGTCCCTCGCATCCCC TTTGACCAATGCTGAAATGAGCCCT
TTGACCCTGTAC**TTTT**TTTCCAGCGGGAC TAGCGCGTTGCTCCTCACT-3'

short 1, short 2, and short 3 lack the 5' end and 3' end side (named "5'-3'-end" herein), the 3' end stem loop, and the central stem loop of the No. 1 aptamer, respectively.

In addition, mutations were introduced into aptamer No.1 without changing its secondary structure. The following sequences are mutants of aptamer No.1, and italic font indicates the introduced mutations.

No.1 mut 1 5'-GGAATGTGGTCCCTCGC *CGCGC*ATCTATGTACG*GCGCG*CCCCTTTGACCAAT GCTGAAATGAGCCCTTTGACCCTGT
ACTTTTTTTCCAGCGGGACTAGCGCGTTGCTCCTCACT-3'

No.1 mut 2 5'-GGAATGTGGTCCCTCGCAATAAATCTATG TACGTTATTCCCCC*CCG*GACCAATGCTGAA ATGAGCCCTTTGACCCTGTACTTTTTTTCCAGCGGGA CTAGCGCGTTGCTCCTCACT-3'

No.1 mut 3 5'-GGAATGTGGTCCCTCGCAATAAATCTATGT ACGTTATTCCCCTTTGACCAATGCTGAAAT GAGCCC*GCG*GACCCTGTACTTTTTTTCCAGCGGGAC TAGCGCGTTGCTCCTCACT-3'

No.1 mut 123 5'-GGAATGTGGTCCCTCGC*CGCGC*ATCTATGTACG*GCGCG*CCCCC*CCG*GACCAATGCTGAAATGAGCCC*GCG*GACCCTGTACTTTTTTTCCAGCGGGACTAGCGCGTTGCTCCTCACT-3'

2.5 AFM imaging

The AFM images were acquired using an AFM system (NanoWizard UltraSpeed, JPK) equipped with a silicon nitride cantilever (Olympus, BL-AC40TS). In all the measurements, NCYM (100 nM) 10eq was added to the adjusted frame structures having aptamer (10 nM) and incubated at 25°C for 2h. Samples were then double-diluted in annealing buffer (20 mM Tris-HCl (pH 7.4), 10 mM $MgCl_2$, 1mM EDTA), adsorbed onto fresh mica plates for 10 minutes at room temperature, and washed three times with the annealing buffer. The observations were performed using the same buffer.

2.6 Immunoprecipitation

20 µl of Dynabeads™ protein G (Thermo Fisher Scientific) slurry was transfered to a clean tube. The tube was placed in a magnetic separation rack for 10-15 seconds, then, the buffer was carefully removed. 2 µg of anti-GSK3β (BD Transduction Laboratories) or mouse IgG (Cell Signaling Technology, Danvers, MA) was dissolved in 200 µl of PBS (0.02% Tween 20) and the solution was added to the beads. The slurry was incubated with rotation at room temperature for 10 minutes. The beads were pelleted using a magnetic separation rack and washed three times with 200 µl of PBS, then resuspended in 100 µl of PBS. 0.17µg (3.7 pmol) of purified GSK3β (Signal Chem, Richmond, Canada), 0.044g (3.7 pmol) of NCYM (1) and DNA aptamers (1eq (3.7 pmol) or 5 eq (18.5 pmol)) were dissolved in 500µl of ice-cold PBS and incubated with rotation at 4°C for 2h. A 100 µl suspension of antibody-conjugated beads was added and incubated for 2h. The beads were pelleted using a magnetic separation rack and washed five times with 1 ml of PBS. The pellet was resuspended in 20 µl of 1X sample buffer and heated at 95°C for 5min.The beads were pelleted using a magnetic separation rack and the supernatant was subjected to the Abby analysis.

2.7 Abby analysis

The NCYM and GSK3β protein levels were measured using a capillary electrophoretic-based immunoassay (the Abby instrument; ProteinSimple, San Jose, CA, USA), according to the manufacturer's protocol. Briefly, the samples were combined with 0.1× sample diluent buffer and 5× fluorescent master mix denaturing buffer to acquire 0.8 µg/µl loading concentration. Subsequently, the samples were denatured for 5 min at 95 °C. The primary antibodies used in this study were anti-NCYM (1) and anti-GSK3β (#9315, Cell Signaling Technology, Danvers, MA, USA). The Abby measurements were performed using a 12–230 kDa separation module with 25-min separation at 375 V, 10-min blocking, 30-min primary antibody incubation, and 30-min secondary antibody incubation (DM-001, ProteinSimple, San Jose, CA, USA). At the end of the run, the chemiluminescent signal was displayed as a virtual blot-like image and an electropherogram based on the molecular weight using Compass (ProteinSimple, San Jose, CA, USA).

2.8 Sample preparation for SAXS measurements

The expression and purification procedures for NCYM were modified from those described previously (8). Recombinant NCYM with glutathione S-transferase (GST) at the N-terminus was expressed in *Escherichia coli* BL21 (DE3) cells using the pGEX-6p-1 vector. The cells were cultured in Luria broth medium containing 0.1 mg/ml ampicillin at 30°C. When OD_{600} reached 0.6, protein expression was induced by 0.1 mM of isopropyl-β-D-thiogalactopyranoside, and culture was continued for 6 h at 30°C. The cells were harvested by centrifugation (3,890×*g*, 15 min, 4°C), and then stored at −30°C until purification.

The frozen cell pellets were thawed and resuspended in phosphate-buffered saline (PBS) supplemented with cOmplete EDTA-free Protease Inhibitor Cocktail (11873580001, Roche, Basel, Switzerland) and lysed by repeated sonication in ice water. The cell lysate was centrifuged at 20,000×*g*, 4°C for 20 min, then the supernatant was loaded onto GSTrap FF column (17513102, Cytiva, Marlborough, MA, USA) equilibrated with PBS, using a peristaltic pump. After washing out the unbound materials with PBS, GST-tagged NCYM (GST-NCYM) was eluted with an elution buffer (50 mM Tris-HCl and 10 mM reduced glutathione, pH 8.0). The flow-through was reloaded onto the re-equilibrated column, and the eluate was collected once more to increase the final protein yield.

The eluted GST-NCYM solution was dialyzed against a buffer containing 50 mM Tris-HCl, 100 mM NaCl, and 1 mM EDTA at pH 8.0. After adding dithiothreitol to the protein solution at a final concentration of 1 mM, the GST-tag was cleaved with 50 U/L-culture of PreScission Protease (27084301, Cytiva) for over 18 h at 4°C with gentle stirring. To exhaustively degrade the remaining nucleic acids, 6,000 U/L-culture of benzonase (71205-3CN/70746-3CN, Merck, Darmstadt, Germany) was added to the cleaved sample along with $MgCl_2$ at a final concentration of 2 mM (Mg^{2+} is required for the activation of benzonase), after which the sample was dialyzed against IEX buffer (20 mM MOPS (pH 7.0) and 1 mM DTT) with 2 mM $MgCl_2$.

The dialyzed sample was loaded onto a HiTrap SP HP column (17115201, Cytiva) equilibrated with IEX buffer, and eluted using a linear gradient of NaCl (0–650 mM). The fractions containing high-purity NCYM, as confirmed by SDS-PAGE, were collected and used for SAXS measurements. The GST-tag did not bind to the column and was detected in the flow-through. The final NCYM yield was 1.25 mg/L-culture, which was estimated with the molecular absorption coefficient of 280 nm ${\epsilon_{280}}^{0.1\%} = 0.558$.

The purified NCYM was dialyzed against a buffer containing 10 mM Tris-HCl (pH8.0), 50 mM NaCl, and 5 mM DTT. Powdered DNA aptamers (No. 1) were dissolved directly in the same buffer. Samples of the NCYM-DNA complex were prepared by mixing these two kinds of solutions at the appropriate molar ratios (see below). These solutions were used for the following SAXS measurements.

2.9 SAXS experiment

SAXS measurements were carried out at BL40B2 in SPring-8 (Hyogo, Japan) on solution samples of NCYM (1.4 and 2.8 mg/ml), the DNA aptamer (1.6 and 5.0 mg/ml), and the NCYM-DNA complex. For the measurements of the complex, two kinds of samples, where the molar ratio of NCYM and DNA was 1:1.2 or 1:1.5 (Table 1), were used to extract the scattering curve of the complex by changing the relative contribution of the unbound DNA aptamers. The wavelength (λ) of the incident X-ray was 1.0 Å and the temperature was 293 K with the sample-to-detector distance of 2.2 m. A pixel detector (PILATUS3S 2 M, Dectris) was used to record the scattering patterns.

Data reduction was conducted using the software SAngler (30): The recorded two-dimensional SAXS patterns were circularly averaged to obtain one-dimensional scattering curves, corrected by the incident flux measured with an ion chamber placed upstream of the samples. The net scattering curves of the scattering particles were obtained by subtracting the scattering curves of the buffer from those of the samples with an appropriate scaling factor based on the scattering particle concentration and the partial specific volumes of proteins (0.73 cm^3/g) or of DNA (0.53 cm^3/g) (31). Finally, the scattering curves were normalized to the absolute scale using H_2O as the standard (32) to estimate the molecular weight of the scattering particles.

Guinier analysis was employed to evaluate the radius of gyration (R_g) of the scattering particle from its scattering curve. A scattering curve I(Q), where Q (=4πsinθ/λ, where 2θ is the scattering angle) denotes the momentum transfer, is represented as follows in a good approximation in the so-called Guinier region ($Q \cdot R_g < 1.3$) (33):

$$I(Q) = I(0)\exp\left(-\frac{1}{3}R_g^2Q^2\right) \quad (1),$$

where I(0) [cm^{-1}] denotes the forward scattering intensity, from which the molecular weight (MW) of the scattering particle is estimated from the following equation in kDa (32):

$$MW = 1500 \times I(0)\frac{1}{c} \quad (2),$$

where c is the weight concentration [g/l].

For rod-like particles such as the DNA aptamer and the NCYM-DNA complex, cross-sectional Guinier analysis was applied to evaluate their cross-sectional radii of gyration (R_c). In this case, I (Q) is approximated as:

$$Q \cdot I(Q) = I_c(0)\exp\left(-\frac{1}{2}R_c^2Q^2\right) \quad (3),$$

where $I_c(0)$ denotes the forward scattering intensity of the cross-section of the scattering particle. Application of Eq. 2 with $I_c(0)$ instead of I(0) yields the scattering mass per unit length.

The scattering curves taken at the lower particle concentrations were merged with those at the higher concentrations at Q = 0.1 $Å^{-1}$ and the resultant curves were used for structural modeling. IGOR Pro software (WaveMetrics, Lake Oswego, OR, USA) was used for (cross-sectional) Guinier analyses and for processing the scattering curves.

TABLE 1 Information on the measured samples of the NCYM-DNA complex.

Molar ratio (NCYM : DNA) in the sample	NCYM concentration in the sample [mg/ml]	DNA concentration in the sample [mg/ml]
1:1.2	1.3	3.2
	0.6	1.5
1:1.5	0.6	1.8

2.10 Structural modeling using the *ab initio* method

Human NCYM comprises109 residues. Because NCYM was expressed with a GST-tag at its N-terminus in this study, eight residues were added to the 109 residues even after cutting the tag. For structural modeling, the program GASBOR (34) was employed, where each residue is represented by a sphere with a constant electron density. The obtained model thus consists of 117 spheres. In this study, the GASBOR runs were repeated 10 times (i.e., 10 best-fit models were obtained). For the modeling of the DNA aptamer, DAMMIF (35) was used, where a molecule is represented by an ensemble of spheres called dummy atoms. The input files required for DAMMIF were generated using AUTORG and DATGNOM (36). The maximum Q value (Q_{max}) used for structural modeling was automatically determined using DATGNOM ($Q_{max}R_g$ < 7–8). The DAMMIF runs were repeated 10 times and the resultant 10 models were averaged, followed by filtration using DAMAVER (37). Structural modeling of the NCYM-DNA complex was carried out using the program MONSA (38), where each of the two phases is represented by a dummy atom model while each phase is assigned a designated electron density value. As an input file of MONSA, the following information is required: The values of the electron density of NCYM and the DNA aptamer were set to be 0.09 $e/Å^3$ and 0.21 $e/Å^3$, respectively, which are typical of these types of molecules (39). The volume fractions of NCYM and the DNA aptamer were obtained from the volumes obtained using GASBOR and DAMMIF, respectively. Using these parameters, the MONSA runs were repeated 10 times. In all of the programs above, the best-fit model, the scattering curve of which reproduces the experimental curve well, is determined by simulated annealing.

2.11 Scattering curve of the NCYM-DNA complex

The scattering curves of the NCYM-DNA complex recorded at 1:1.2 and 1:1.5 molar ratios were found to be superimposable despite the existence of different amounts of unbound DNA aptamers (Figure S2). However, as described in the Results section, NCYM molecules and the DNA aptamers exist as monomers and dimers, respectively, in solution so that the molar ratios of NCYM and the dimeric DNA aptamer in the above samples are 1:0.6 and 1:0.75, respectively, which result in the molar ratios of the complex and the excess amount of NCYM of 0.6:0.4 and 0.75:0.25 assuming that one NCYM monomer binds to a DNA dimer. Because the scattering intensity is proportional to the square of the product of the scattering contrast of a particle and its volume, the scattering contributions of excess NCYM molecules were calculated to be 0.8% and 0.4% for the 0.6:0.4 and 0.75:0.25 samples, respectively. In this estimation, the scattering contrasts of NCYM and DNA were set to be 0.09 e/Å^3 and 0.21 e/Å^3 as described above, and the volumes of a NCYM monomer and a DNA dimer were assumed to be 17000 Å^3 and 68000 Å^3 as obtained by GASBOR and DAMMIF. Thus, it appears that the scattering contribution of free NCYM monomers is negligible, which explains why the scattering curves obtained at NCYM : DNA = 1:1.2 and 1:1.5 are similar to each other within errors. This observation also excluded the possibility that the complex consists of one NCYM monomer and one monomeric DNA aptamer. Based on the above inspection, the curves obtained from the above samples reflect those of the NCYM-DNA complex. The scattering curves taken at NCYM : DNA = 1:1.2 and 1:1.5 were, therefore, averaged to improve the signal-to-noise ratio and the resultant curve was employed as the scattering curve of the NCYM-DNA complex.

3 Results

3.1 Aptamer selection by SELEX

To identify DNA aptamers that specifically bind to NCYM, we performed SELEX and employed NGS technology to monitor the progress of the enrichment sequences that bind to the target in the selection pool. We performed the NGS analysis using an Illumina MiniSeq for the NCYM aptamer selection pools from rounds 3 to 8 to identify the aptamer candidates. The ratio of unique DNA sequences in the selection pool per round showed that DNA sequences were enriched in round 6 (Figure 1A). No. 1 aptamer candidate, which was the most abundant population in round 8, was quickly enriched from round 6 compared to the other DNA sequences. Guanine contents in the selection pools were increased slightly in round 8 (Table S1). BLI measurements and kinetic analyses showed that top three aptamers strongly bind to NCYM protein at K_D value in the range of 53.9 to 299 nM (Table S2); however, the aptamer No. 2 and 3 were predicted to form different secondary structure at 38°C, showing low structural stability at relatively high temperature compared to aptamer No.1 (Figure 1B and Figure S3).

3.2 AFM observation on the NCYM-DNA aptamer complexes

3.2.1 DNA aptamers obtained by SELEX (No. 1, 2, and 3)

To visualize the interaction between NCYM and DNA at the molecular level, we used AFM observation of the aptamer-conjugated DNA frames with or without NCYM (Figure 2). Because the aptamer has a large single-stranded region and is not completely fixed, only dsDNA can be seen. After adding NCYM, white dots were observed on the dsDNA (Figure 2B). The numbers of DNA frames with and without these white dots were counted to assess the affinity of the aptamer for NCYM (Figure 2B). Binding of NCYM to the No. 1 aptamer was observed in 14.5% of the DNA frames (84/580 frames). Binding was also observed for the No. 2 and No. 3 aptamers at 12.0% (80/664 frames) and 14.6% (41/280 frames), respectively. The slightly lower affinity of No. 2 compared to No.1 and No.3 is consistent with the KD value evaluated by BLI measurements.

3.2.2 Truncated and mutated DNA aptamers

Because of the relatively higher stability of the secondary structures of the No.1 aptamer (Figure S2), we focused on aptamer No.1 and further analyzed the regions required for NCYM binding. As shown in Figure 3A, the AFM results revealed a decrease in the number of bonds in short 1 (9.4%, 70/742 frames) and short 2 (5.3%, 75/1425 frames), whereas there was no change in the number of bonds in short 3 (14.2%, 173/1222 frames). Because aptamers No. 2 and No. 3 exhibited GC-rich sequences (Figure 1B), we introduced mutations in the No. 1 aptamer with increasing GC content without affecting the secondary structure, and we found that the mutations at TTT in the central loop showed a significant decrease in the affinity of aptamer No.1 to NCYM (Figure 3B). These results suggest that the 5'-3'-end, the 3' end stem loop, and TTT in the central loop of the No.1 aptamer are required for binding to NCYM.

3.3 The effect of DNA aptamer No. 1 on the interaction between NCYM and GSK3β

To clarify the effect of DNA aptamer No.1 on NCYM function, we examined NCYM binding to GSK3β with or without the aptamer. As reported (1), purified NCYM was co-immunoprecipitated with GSK3β (Figure 4A). Addition of aptamer No.1 enhanced the interaction between NCYM and GSK3β in dose-dependent manner (Figure 4B). This result led us to analyze the NCYM-aptamer No.1 complex because the aptamer may help NCYM to adopt a conformation that facilitates binding to GSK3β.

3.4 SAXS results

3.4.1 Guinier analysis and Kratky plot

To elucidate their structural properties, we employed SAXS analysis to reveal the approximate structures of the complexes in

10.3389/fonc.2023.1213678

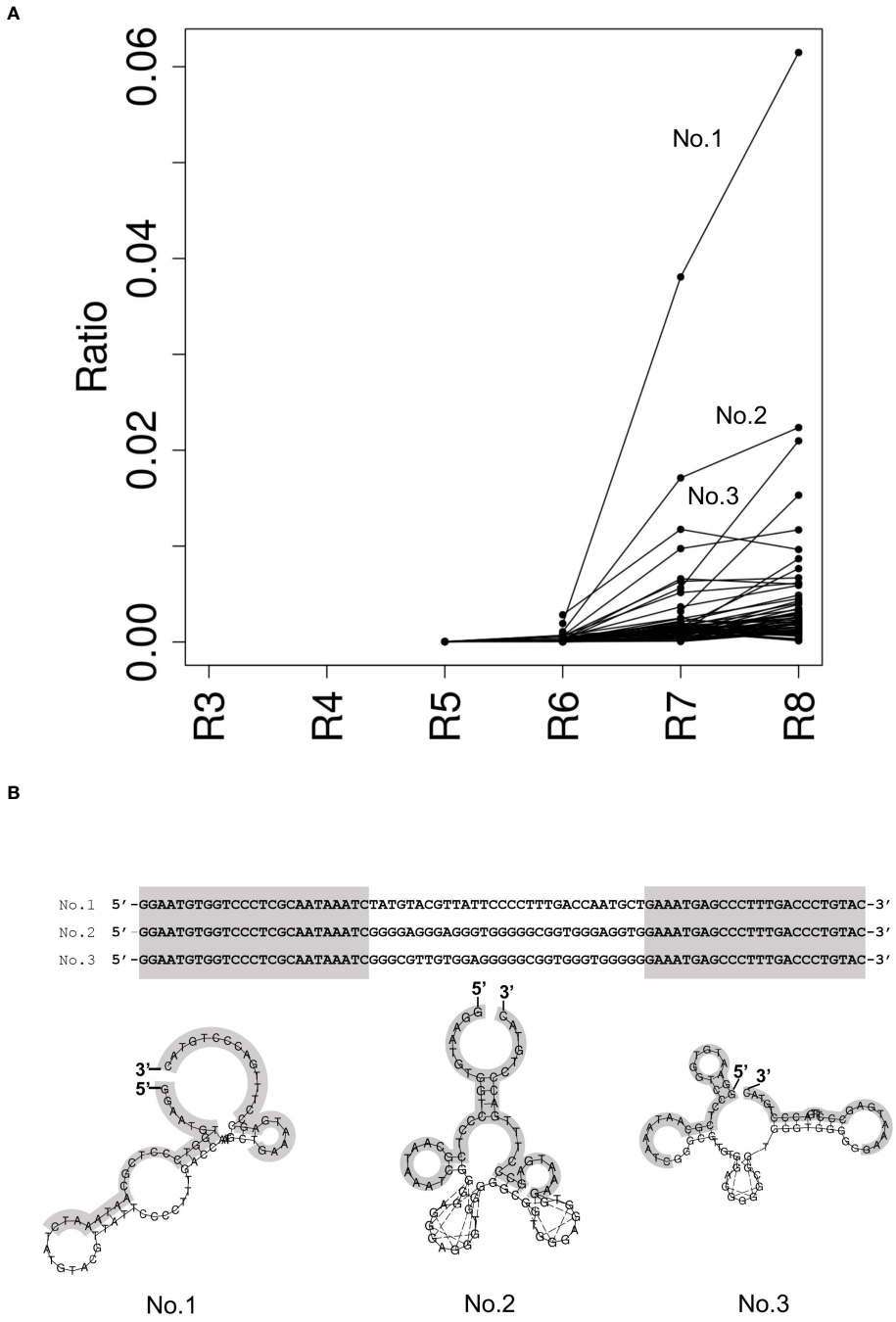
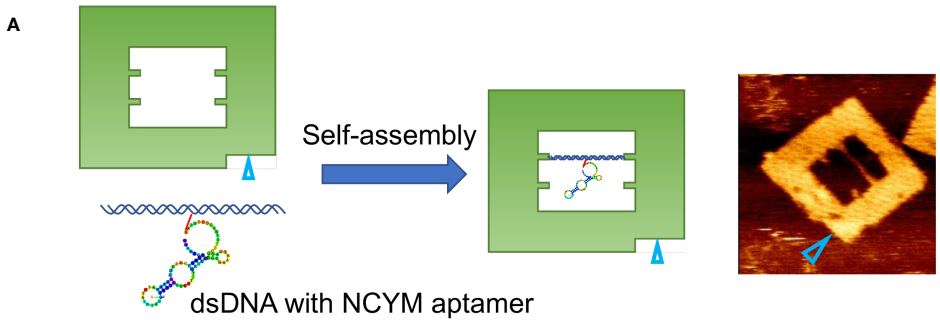


FIGURE 1
Identification of NCYM-bound aptamers. **(A)** Ratio of the enriched ssDNA sequence in the selection pool per round. Each dot represents the ratio of the enriched ssDNA sequences more than 0.001. The lines indicate the same ssDNA sequence between rounds. Top 3 enrichment sequences at round 8 were evaluated for the further structural analysis and named No. 1, No. 2, and No. 3. **(B)** Secondary structure prediction of NCYM-bound aptamer by RNAfold at 25°C. Primer sequences are colored in gray.

solution. Figure 5A shows the results of the Guinier analysis. The radii of gyration (R_g) were estimated to be 25.2 ± 0.7 Å, 46.0 ± 0.3 Å, and 44.8 ± 0.2 Å, for NCYM, the DNA aptamer No. 1, and the NCYM-DNA complex, respectively. The molecular weights of NCYM and the DNA aptamer were estimated to be 7.2, and 54.5 kDa, respectively, indicating that NCYM molecules exist as a monomer in solution whereas DNA aptamers are dimers because the molecular weights of NCYM and the DNA aptamer were 12 kDa, 24.2 kDa, respectively. Although the molecular weight of the complex was estimated to be 46.8 kDa, because the true concentration of the complexes in the sample is not known due to the existence of a small amount of excess NCYM molecules, this value should be interpreted with caution. Scattering curves other than NCYM were found to follow the cross-sectional Guinier approximation, suggesting that the DNA aptamer and the complex adopt a rod-like shape. Figure 5B shows the results of



	No.1	No.2	No.3	Random
occupied	84/580	80/664	41/280	17/618
rate	14.5%	12.0%	14.6%	2.8%

FIGURE 2
Secondary structure prediction model for aptamers and evaluation using DNA origami. **(A)** Design of DNA frame with NCYM aptamer and its AFM image. The open cyan triangles show the orientation marker. **(B)** Examples of AFM images of NCYM-aptamer No.1 complexes on DNA frames (left) and evaluation of the affinity between aptamers obtained by SELEX and NCYM (right). In the "Occupied" row, the numerator and the denominator represent the number of NCYM bound to the aptamers and the total number of aptamers, respectively. The "rate" row denotes the number fraction of NCYM bound to the aptamers calculated from the corresponding value in the "occupied" row.

the cross-sectional Guinier analysis, from which the cross-sectional radii of gyration (R_c) were estimated to be 12.2 ± 0.4 Å and 12.5 ± 0.2 Å for the DNA aptamer and the complex, respectively. This suggests that the overall size of the cross-section is similar between the DNA aptamer and the complex. Both the Guinier analysis and the cross-sectional Guinier analysis can be applied to the current samples because of their relatively short entire length, whereas in many cases, fibrillar proteins are quite long (even on the order of μm) and only the cross-sectional Guinier analysis can be applied (40, 41).

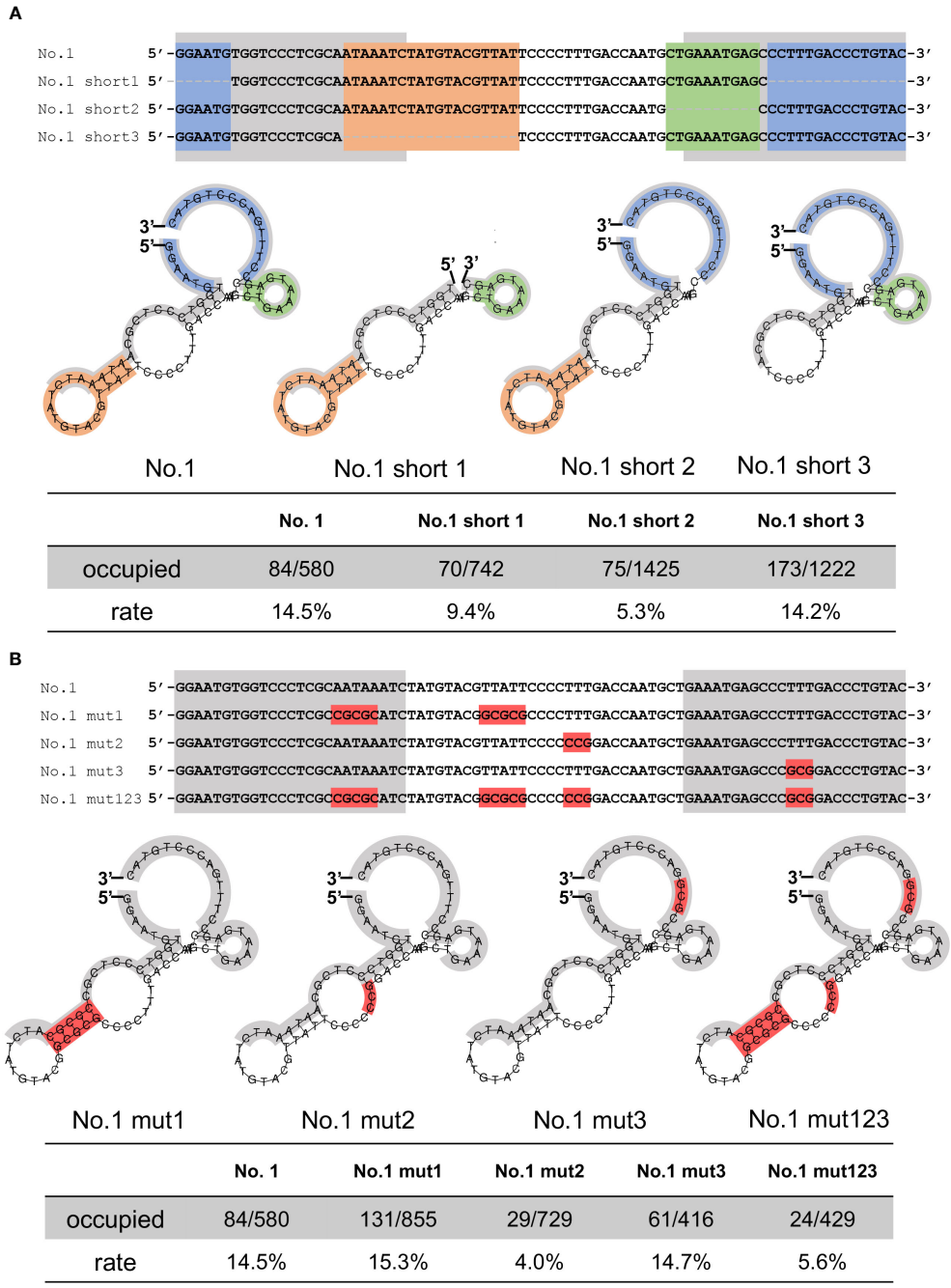
Several plots were used to investigate the structural properties of NCYM, as shown in Figure 5C. The Kratky plot ($Q^2 \cdot I(Q)$ vs Q) (42) shows that the $Q^2 \cdot I(Q)$ value reaches a peak at around Q = 0.1 [$Å^{-1}$] and decreases slightly, followed by the increase, suggesting that NCYM is not a completely unfolded protein, but a partially folded protein. This interpretation is further supported by the Q3 plot ($Q^3 \cdot I(Q)$ vs Q^3), where the $Q^3 \cdot I(Q)$ value reaches a plateau, which is a hallmark of partially folded proteins (43). In contrast, no plateau was observed in the Porod-Debye plot (43), suggesting that NCYM is not a well-folded protein. These results show that NCYM adopts a highly flexible conformation, though it is not a completely unfolded protein, but a partially folded protein. This is in agreement with a previous study showing that NCYM molecules contain defined secondary structures in solution (8).

3.4.2 Solution structure of NCYM, the DNA aptamer, and the NCYM-DNA complex

The results of the structural modeling of NCYM, the DNA aptamer No. 1, and the NCYM-DNA complex are summarized in Figure 6. Figure 6A compares the experimental SAXS curves with those calculated using the best-fit model obtained for each of the three samples. The χ^2 values, which are averaged over 10 models obtained, were 1.17, 1.37, and 1.85, for NCYM, the DNA aptamer, and the complex, respectively. As shown in Figure 6A, the obtained models provide excellent fits to the measured SAXS curves.

As shown in Figure 6B, NCYM was found to have a slightly extended structure in which the bulky and flexible parts are mixed. It is not a completely unfolded protein such as a random coil, as expected from the panels in Figure 5C and our previous study on the secondary structure of NCYM (8). Although the assignment of the secondary structure to the three-dimensional structures of the current NCYM models is not possible at this stage, future structural modeling using SAXS curves with higher Q values would be useful for this purpose (44). The volume of the NCYM models was ~17000 $Å^3$. Regarding the DNA aptamer, a structure with a volume of ~68000 $Å^3$, in which four to five bulky nodes were connected, was found to reproduce the measured SAXS curve, as shown in Figure 6C. Because the aptamers exist as dimers, as evidenced by the molecular weight estimation as described above, a monomeric DNA aptamer corresponds to two or three nodes in this model. The structural features of the DNA model with some nodes here are consistent with those predicted using RNAfold (Figure 1B). Two representative models of the NCYM-DNA complex are shown in Figure 6D. In both cases, the volumes of the NCYM molecule and of the DNA aptamer in the complex were ~19000 $Å^3$ and ~63000 $Å^3$, which are roughly the same as those obtained when these molecules are in isolation. The volume of the complex (~82000 $Å^3$) was essentially the same as the sum of the volumes of NCYM and DNA in isolation (~85000 $Å^3$) within 4% accuracy. It thus follows that the NCYM-DNA complex consists of one NCYM molecule and two DNA aptamers (a dimer). NCYM tends to bind either to or close to a tip of a dimeric DNA aptamer. The slight differences (~10%) in the volumes of each component between in isolation and in the complex may imply that some intramolecular structural

10.3389/fonc.2023.1213678



	No. 1	No.1 short 1	No.1 short 2	No.1 short 3
occupied	84/580	70/742	75/1425	173/1222
rate	14.5%	9.4%	5.3%	14.2%

	No. 1	No.1 mut1	No.1 mut2	No.1 mut3	No.1 mut123
occupied	84/580	131/855	29/729	61/416	24/429
rate	14.5%	15.3%	4.0%	14.7%	5.6%

FIGURE 3
Evaluation of the affinity between truncated or mutated aptamers and NCYM. **(A)** Truncated No.1 aptamers for identification of NCYM binding site. The 5′-3′-end, the central stem loop, the 3′ end stem loop, and the primer region are colored in blue, orange, green, and gray, respectively (upper and middle). Evaluation of the affinity between the truncated aptamers and NCYM (bottom) **(B)** Mutated No.1 aptamers for identification of the NCYM binding site. Mutated sequences are colored in red (upper and middle). Evaluation of the affinity between the mutated aptamers and NCYM (bottom).

changes occur upon binding. A more detailed structure of the complex can be obtained by small-angle neutron scattering (SANS) combined with contrast matching or variation (45), which will be performed in the future. All the 10 models obtained for NCYM, the DNA aptamer, and the NCYM-DNA complex are shown in Figures S4–S6, respectively.

Attempts were made to identify other possible conformations of the complex by changing the volume ratio of each phase in the complex and/or assuming a symmetry in the structure, which are provided as an input file to MONSA. Whereas several models which fit the SAXS curve of the complex quite well in terms of the χ^2 values were obtained, in all cases, deviation of the volumes of NCYM and the DNA aptamer in the complex with regards to those in the unbound state was much larger (15–20%) than the models presented in Figure 6. Considering an independent line of evidence that the volume change of proteins between the folded and the

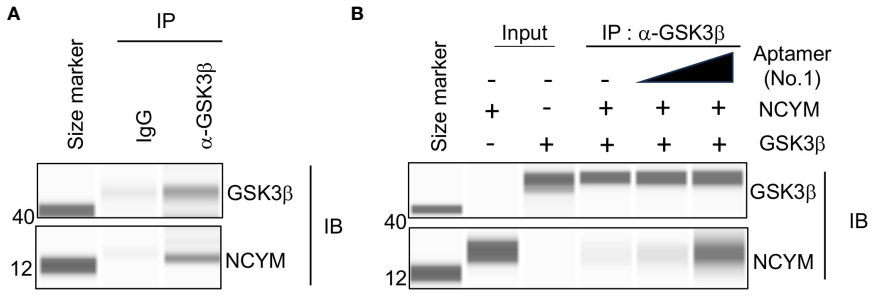


FIGURE 4
Aptamer No.1 increased interaction between NCYM and GSK3β. **(A)** Co-immunoprecipitation of NCYM with GSK3β detected by Abby analysis. **(B)** Aptamer No.1 increased the amount of NCYM co-immunoprecipitated with GSK3β in a dose-dependent manner (0, 1, and 5 eq.).

unfolded states is less than 0.5% (46), the volume change of each component in the complex should be the smallest. Since in the models in Figure 6, the volume of each component in the complex is close to that in the unbound state, and truncation of one edge of a DNA aptamer breaks down the interaction with NCYM as observed by AFM, the models presented here appear to be reasonable.

4 Discussion

In this study, we established a new observation system using aptamers for the single-molecule observation of proteins using DNA origami. To date, protein studies using DNA origami have mainly involved single-molecule observations of proteins that bind

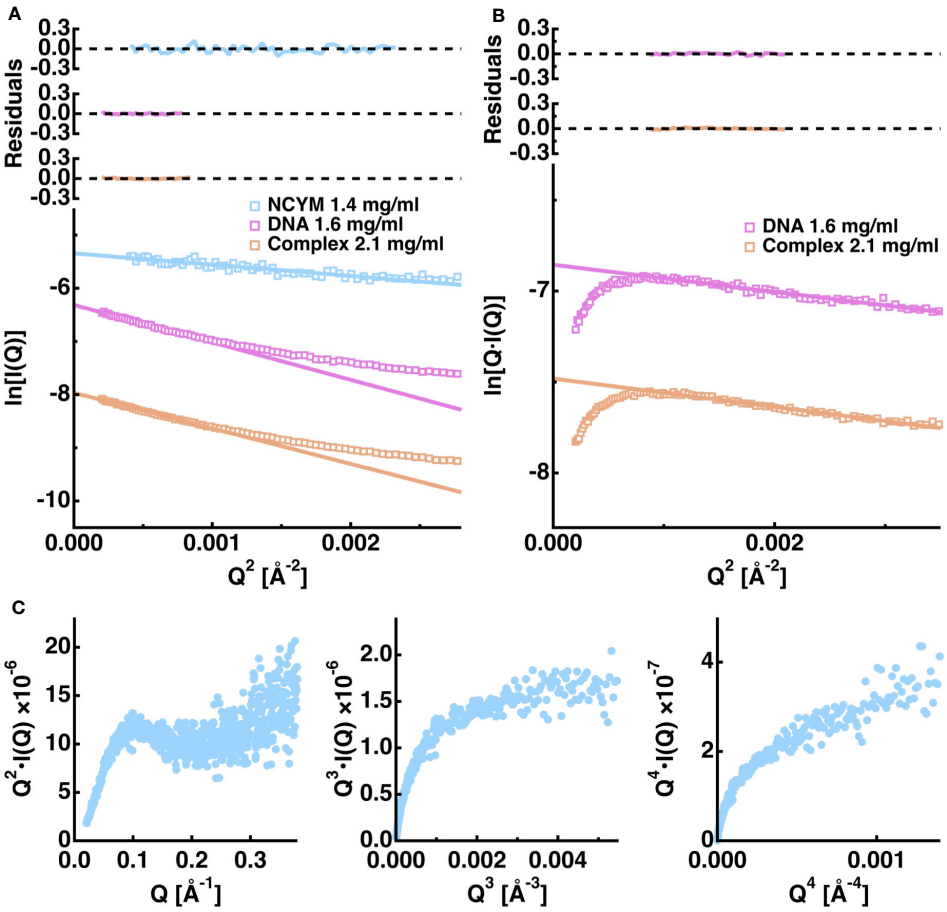


FIGURE 5
Summary of the analysis of the SAXS scattering curves of the NCYM-DNA system. **(A)** Guinier analysis. The logarithm of the scattering intensity is shown as a function of Q^2 for NCYM (cyan), the DNA aptamer No. 1 (magenta), and the NCYM-DNA complex (orange). Upper panels denote the corresponding residuals between the measured and the fitted values. **(B)** Cross-sectional Guinier analysis. Instead of ln[I(Q)] of the Guinier analysis, ln [Q·I(Q)] is plotted as a function of Q^2, from which the cross-sectional radius of gyration is evaluated. **(C)** The left, middle, and right panels show the Kratky plot, Q^3·I(Q) vs Q^3 plot, and the Porod-Debye plot, respectively, of NCYM.

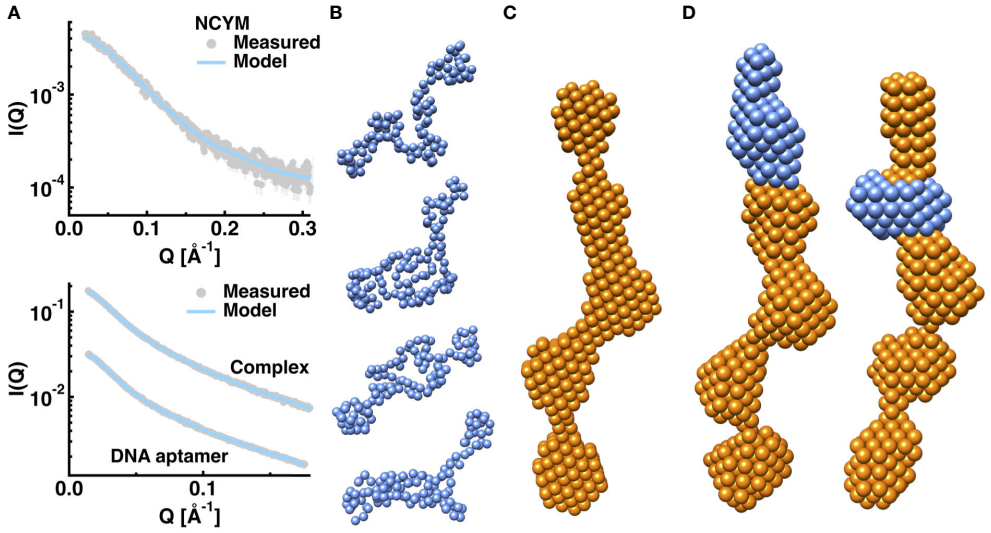


FIGURE 6
Ab initio structural models of NCYM, the DNA aptamer No. 1, and the NCYM-DNA complex. Comparison of the scattering curves between the experiments and the models of NCYM is shown in the upper panel of **(A)** The lower panel of **(A)** shows the comparisons for the DNA aptamer and the NCYM-DNA complex. Experimental values are shown in grey filled circles and the simulated values from the models are shown in cyan solid lines. Error bars are within symbols if not shown. Scattering curves are vertically shifted for clarity. **(B)** Gallery of the NCYM models (arbitrarily chosen 4 models) obtained from GASBOR. Each sphere represents one amino acid residue. **(C)** A representative dummy atom model of the DNA aptamer obtained from DAMMIF and DAMAVER ("damfilt.pdb" is shown). **(D)** Dummy atom models of the NCYM-DNA complex obtained from MONSA. The moieties corresponding to NCYM and the DNA aptamer are shown in marine blue and in orange, respectively.

directly to DNA or using systems based on ligand or avidin-biotin binding (21–29, 47–52). However, DNA origami research using aptamers has mainly focused on functionalizing the DNA origami using already established aptamers (53–56). Therefore, a system that uses DNA origami and aptamers to elucidate the structure of proteins, as in this study, is a new approach that has never been used. In this study, we have also succeeded in roughly identifying the binding site of NCYM to the aptamer obtained using the SELEX method. The method using aptamers can easily fix proteins onto the DNA origami. Therefore, it is expected to be applied to the observation of the interaction between a fixed protein and its target, and to the functional evaluation of proteins for which single molecule observation has not been performed.

In addition to the AFM observation, further structural characterization of NCYM and the DNA aptamer No. 1 that facilitates NCYM binding to GSK3β was conducted using SAXS. Regarding the structure of the DNA aptamer, it was found that it exists as a dimer. This finding is supported by both the molecular weight estimation from the forward scattering intensity and the fitting of the corresponding SAXS curve. Although it is not possible to unambiguously determine the manner in which the two DNA aptamers form dimers, this can be inferred from the current findings. There are three types of arrangements for the two DNA aptamers to form a structure similar to that shown in Figure 6C (Figures 7A–C).

As shown in Figure 7, there are three types of arrangements for two DNA aptamers to form a similar form to that shown in Figure 6C. In the case where the interface between the two aptamers is formed by the central stem loop of one aptamer and the 5'-3'-end of the other aptamer (Figure 7A), NCYM is able to bind to the 5'-3'-end of the upper aptamer. Whereas there is a possibility that other aptamers bind to either edge of the dimer in the same manner, dimers might be more stable than higher-order aggregates as observed for a protein (57) probably due to entropy-enthalpy compensation. In case where two 5'-3'-ends form the interface of the dimer (Figure 7B), NCYM would not be able to bind to the aptamer since these regions are required for NCYM binding as suggested by AFM. If the interface of the dimer is formed by two central stem loops (Figure 7C), at least two NCYM molecules would be able to bind to both edges of the dimer, which is inconsistent with the discussion on the SAXS curve of the complex and our modeling results. The second NCYM binding may be unfavorable in terms of entropy. There are thus two possible models on the mode of dimeric formation of the DNA aptamer (Figures 7A, C). In addition, the observation that NCYM binds to the dimeric DNA aptamer (Figure 6) implies that the affinity between NCYM and an aptamer is lower than that between monomeric aptamers and thus DNA dimers do not dissociate into two monomeric aptamers in the current solution condition.

A concern in the current SAXS analysis is that the deviation of the estimated molecular weight (7.2 kDa) of NCYM from its theoretical value (11.7 kDa) is relatively large. This raises the possibility that the DNA aptamers might exist as trimers or tetramers. The large deviation observed for NCYM is likely to be caused by the small size of NCYM, thus resulting in relatively large experimental errors. However, because the scattering intensity is proportional to the square of the molecular volume, the SAXS curve of the DNA aptamer, which has a larger molecular weight than NCYM, has much lower experimental errors than NCYM (Figure 6A). This results in a more reliable molecular weight

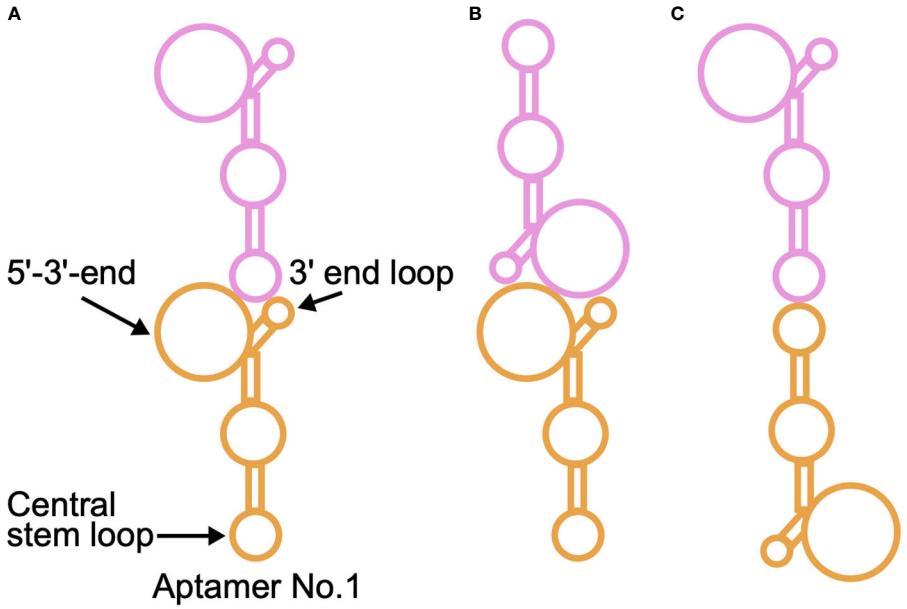


FIGURE 7
Schematic illustration of possible ways by which the DNA aptamer No. 1 forms a dimer. An aptamer is shown in either orange or magenta. The corresponding predicted structure of the aptamer is shown in Figure 1B (left). There are three types of arrangements for two DNA aptamers to take a form similar to that shown in Figure 6C: **(A)** The interface between the two aptamers is formed by the central stem loop of one aptamer and the 5′-3′-end of the other aptamer. **(B)** Two 5′-3′-ends form the interface of the dimer. **(C)** The interface of the dimer is formed by two central stem loops. For a more detailed discussion, please see the main text.

estimation for the DNA aptamer than for NCYM. Furthermore, the structural modeling of an isolated DNA aptamer presented in this work does not require any information on the molecular property including its molecular weight because it is an "ab initio" approach. From the SAXS-derived models with five bumps (Figure 6) and the secondary-structure predicted by RNAfold, which has 2–3 bumps in one aptamer (No. 1 in Figure 1B), it is reasonable to conclude that the DNA aptamer forms a dimer in solution.

In the dummy atom models of the complex, the moiety corresponding to NCYM adopted a compact and well-defined shape whereas NCYM showed flexibility and adoped a slightly extended conformation in the unbound state. This implies that highly flexible NCYM molecules fold upon binding to DNA through the well-known "fly-casting mechanism" (58), in which an unfolded region(s) of a protein binds weakly to the binding site at a relatively large distance, followed by folding as the protein approaches the binding site. As aptamer No.1 facilitated the interaction between NCYM and GSK3β, the compact and well-defined shape of NCYM found in the complex with aptamer No. 1 appears to be the functional structure of NCYM. Folding upon binding has been demonstrated for another *de novo* protein AFGP8 (16). A similar mechanism may also apply to interactions between NCYM and GSK-3β, underlying the mechanism of stabilization of these molecules.

One of the most important advantages of AFM observations with a DNA frame is that the DNA frame is guaranteed to be a monomeric aptamer, which is a feature not observed in other binding assays. This is because dsDNA is bound to an DNA aptamer (Figure 2A) and thus two dsDNAs should be observed by AFM if the DNA aptamers form dimers, which is not the case. Therefore, it is most likely that monomeric aptamers are capable of binding NCYM and dimer formation of DNA aptamers is not a prerequisite for binding of NCYM. Since the SAXS data alone did not show that NCYM could bind to the monomeric form of the aptamer, AFM, in combination with SAXS, provides important insights into the structure of the aptamer-NCYM complex.

Identification of DNA aptamers that can specifically bind to NCYM and facilitate its binding to GSK3β is useful for elucidation of the structure of NCYM in its active form. Although the aptamer-NCYM complex described in this study has not yet been tested by X-ray crystallography, investigation of the structural properties of the complexes may pave the way for the structural characterization of NCYM molecules at the atomic level. Therefore, the present study suggests that the combination of SELEX, AFM, and SAXS is useful for understanding the structural properties of NCYM and the current findings will serve as a foundation for the future design of novel anti-cancer drugs targeting NCYM as well as for elucidating the stabilization mechanism of other cancer-related proteins by NCYM.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

YS and TM designed and supervised the study. SY, FK, KN, MH, KH, TT, YH, YS and TM performed the experiments, acquired, and analyzed the data. SY, FK, KN, TT, YS and TM wrote the manuscript. KN, YH, TT, YS and TM acquired funds. All authors contributed to the article, read, and approved the submitted version.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the AMED eASIA JRP (22jm0210092h0002), and by the JSPS Bilateral Program Number JPJSBP120233212 (SAKURA project).

Acknowledgments

The authors thank SPring-8 for the allocation of the beamtime. The SAXS experiments were carried out under the approval of the SPring-8 Program Review Committee (2022B1242).

Conflict of interest

MH and KH are employees of NEC Solution Innovators. This does not alter the authors' adherence to all Frontiers in Oncology policies.

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1213678/full#supplementary-material

References

1. Suenaga Y, Islam SMR, Alagu J, Kaneko Y, Kato M, Tanaka Y, et al. NCYM, a cis-antisense gene of MYCN, encodes a *de novo* evolved protein that inhibits GSK3β Resulting in the stabilization of MYCN in human neuroblastomas. *PloS Genet* (2014) 10:e1003996. doi: 10.1371/journal.pgen.1003996

2. Suenaga Y, Nakatani K, Nakagawara A. *De novo* evolved gene product NCYM in the pathogenesis and clinical outcome of human neuroblastomas and other cancers. *Jpn J Clin Oncol* (2020) 50:839–46. doi: 10.1093/jjco/hyaa097

3. Shoji W, Suenaga Y, Kaneko Y, Islam SMR, Alagu J, Yokoi S, et al. NCYM promotes calpain-mediated Myc-nick production in human MYCN-amplified neuroblastoma cells. *Biochem Biophys Res Commun* (2015) 461:501–6. doi: 10.1016/j.bbrc.2015.04.050

4. Suenaga Y, Yamamoto M, Sakuma T, Sasada M, Fukai F, Ohira M, et al. TAp63 represses transcription of MYCN/NCYM gene and its high levels of expression are associated with favorable outcome in neuroblastoma. *Biochem Biophys Res Commun* (2019) 518:311–8. doi: 10.1016/j.bbrc.2019.08.052

5. Kaneko Y, Suenaga Y, Islam SMR, Matsumoto D, Nakamura Y, Ohira M, et al. Functional interplay between MYCN, NCYM, and OCT4 promotes aggressiveness of human neuroblastomas. *Cancer Sci* (2015) 106:840–7. doi: 10.1111/cas.12677

6. Islam SMR, Suenaga Y, Takatori A, Ueda Y, Kaneko Y, Kawana H, et al. Sendai virus-mediated expression of reprogramming factors promotes plasticity of human neuroblastoma cells. *Cancer Sci* (2015) 106:1351–61. doi: 10.1111/cas.12746

7. Zhu X, Li Y, Zhao S, Zhao S. LSINCT5 activates Wnt/β-catenin signaling by interacting with NCYM to promote bladder cancer progression. *Biochem Biophys Res Commun* (2018) 502:299–306. doi: 10.1016/j.bbrc.2018.05.076

8. Matsuo T, Nakatani K, Setoguchi T, Matsuo K, Tamada T, Suenaga Y. Secondary structure of human *de novo* evolved gene product NCYM analyzed by vacuum-ultraviolet circular dichroism. *Front Oncol* (2021) 11:688852. doi: 10.3389/fonc.2021.688852

9. Van Oss SB, Carvunis A-R. *De novo* gene birth. *PloS Genet* (2019) 15:e1008160. doi: 10.1371/journal.pgen.1008160

10. Broeils LA, Ruiz-Orera J, Snel B, Hubner N, van Heesch S. Evolution and implications of *de novo* genes in humans. *Nat Ecol Evol* (2023) 7:804–15. doi: 10.1038/s41559-023-02014-y

11. Weisman CM. The origins and functions of *de novo* genes: against all odds? *J Mol Evol* (2022) 90:244–57. doi: 10.1007/s00239-022-10055-3

12. Bornberg-Bauer E, Hlouchova K, Lange A. Structure and function of naturally evolved *de novo* proteins. *Curr Opin Struct Biol* (2021) 68:175–83. doi: 10.1016/j.sbi.2020.11.010

13. Heames B, Buchel F, Aubel M, Tretyachenko V, Loginov D, Novák P, et al. Experimental characterization of *de novo* proteins and their unevolved random-sequence counterparts. *Nat Ecol Evol* (2023) 7:570–80. doi: 10.1038/s41559-023-02010-2

14. Bungard D, Copple JS, Yan J, Chhun JJ, Kumirov VK, Foy SG, et al. Foldability of a natural *de novo* evolved protein. *Structure* (2017) 25:1687–1696.e4. doi: 10.1016/j.str.2017.09.006

15. Lange A, Patel PH, Heames B, Damry AM, Saenger T, Jackson CJ, et al. Structural and functional characterization of a putative *de novo* gene in Drosophila. *Nat Commun* (2021) 12:1667. doi: 10.1038/s41467-021-21667-6

16. Her C, Yeh Y, Krishnan VV. The ensemble of conformations of antifreeze glycoproteins (AFGP8): a study using nuclear magnetic resonance spectroscopy. *Biomolecules* (2019) 9:235. doi: 10.3390/biom9060235

17. Armstrong BC, Krystal GW. Isolation and characterization of complementary DNA for N-cym, a gene encoded by the DNA strand opposite to N-myc. *Cell Growth Differ Mol Biol J Am Assoc Cancer Res* (1992) 3:385–90.

18. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PloS One* (2010) 5:e15004. doi: 10.1371/journal.pone.0015004

19. Minagawa H, Onodera K, Fujita H, Sakamoto T, Akitomi J, Kaneko N, et al. Selection, characterization and application of artificial DNA aptamer containing appended bases with sub-nanomolar affinity for a salivary biomarker. *Sci Rep* (2017) 7:42716. doi: 10.1038/srep42716

20. Cantu VA, Sadural J, Edwards R. PRINSEQ++, a multi-threaded tool for fast and efficient quality control and preprocessing of sequencing datasets. *PeerJ Prepr* (2019) 7:e27553v1. doi: 10.7287/peerj.preprints.27553v1

21. Endo M, Katsuda Y, Hidaka K, Sugiyama H. Regulation of DNA methylation using different tensions of double strands constructed in a defined DNA nanostructure. *J Am Chem Soc* (2010) 132:1592–7. doi: 10.1021/ja907649w

22. Rajendran A, Endo M, Hidaka K, Tran PLT, Mergny J-L, Sugiyama H. Controlling the stoichiometry and strand polarity of a tetramolecular G-quadruplex structure by using a DNA origami frame. *Nucleic Acids Res* (2013) 41:8738–47. doi: 10.1093/nar/gkt592

23. Endo M, Inoue M, Suzuki Y, Masui C, Morinaga H, Hidaka K, et al. Regulation of B-Z conformational transition and complex formation with a Z-form binding protein by introduction of constraint to double-stranded DNA by using a DNA nanoscaffold. *Chem Eur J* (2013) 19:16887–90. doi: 10.1002/chem.201303830

24. Yang Y, Endo M, Suzuki Y, Hidaka K, Sugiyama H. Direct observation of the dual-switching behaviors corresponding to the state transition in a DNA nanoframe. *Chem Commun* (2014) 50:4211–3. doi: 10.1039/C4CC00489B

25. Rajendran A, Endo M, Hidaka K, Sugiyama H. Direct and single-molecule visualization of the solution-state structures of G-hairpin and G-triplex intermediates. *Angew Chem Int Ed Engl* (2014) 53:4107–12. doi: 10.1002/anie.201308903

26. Suzuki Y, Endo M, Cañas C, Ayora S, Alonso JC, Sugiyama H, et al. Direct analysis of Holliday junction resolving enzyme in a DNA origami nanostructure. *Nucleic Acids Res* (2014) 42:7421–8. doi: 10.1093/nar/gku320

27. Yamamoto S, De D, Hidaka K, Kim KK, Endo M, Sugiyama H. Single molecule visualization and characterization of Sox2-Pax6 complex formation on a regulatory DNA element using a DNA origami frame. *Nano Lett* (2014) 14:2286–92. doi: 10.1021/nl4044949

28. Endo M, Xing X, Zhou X, Emura T, Hidaka K, Tuesuwan B, et al. Single-molecule manipulation of the duplex formation and dissociation at the G-quadruplex/i-motif site in the DNA nanostructure. *ACS Nano* (2015) 9:9922–9. doi: 10.1021/acsnano.5b03413

29. Räz MH, Hidaka K, Sturla SJ, Sugiyama H, Endo M. Torsional constraints of DNA substrates impact cas9 cleavage. *J Am Chem Soc* (2016) 138:13842–5. doi: 10.1021/jacs.6b08915

30. Shimizu N, Yatabe K, Nagatani Y, Saijyo S, Kosuge T, Igarashi N. Software development for analysis of small-angle x-ray scattering data. *AIP Conf Proc* (2016) 1741:050017. doi: 10.1063/1.4952937

31. Savelyev A. Assessment of the DNA partial specific volume and hydration layer properties from CHARMM Drude polarizable and additive MD simulations. *Phys Chem Chem Phys* (2021) 23:10524–35. doi: 10.1039/d1cp00688f

32. Orthaber D, Bergmann A, Glatter O. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *J Appl Crystallogr* (2000) 33:218–25. doi: 10.1107/S0021889899015216

33. Svergun D, Koch MH, Timmins P, May RP. *Small angle X-ray and neutron scattering from solutions of biological macromolecules*. United Kingdom: Oxford Universty Press (2013).

34. Svergun DI, Petoukhov MV, Koch MH, Biophys J. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* (2001) 80:2946–53. doi: 10.1016/S0006-3495(01)76260-1

35. Franke D, Svergun DI. DAMMIF, a program for rapidab-initioshape determination in small-angle scattering. *J Appl Crystallogr* (2009) 42:342–6. doi: 10.1107/s0021889809000338

36. Petoukhov MV, Konarev PV, Kikhney AG, Svergun DI. ATSAS 2.1 – towards automated and web-supported small-angle scattering data analysis. *J Appl Crystallogr* (2007) 40:s223–8. doi: 10.1107/s0021889807002853

37. Volkov VV, Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* (2003) 36:860–4. doi: 10.1107/s0021889803000268

38. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* (1999) 76:2879–86. doi: 10.1016/S0006-3495(99)77443-6

39. Gabel F, Engilberge S, Schmitt E, Thureau A, Mechulam Y, Pérez J, et al. Medical contrast agents as promising tools for biomacromolecular SAXS experiments. *Acta Crystallogr Sect D Struct Biol* (2022) 78:1120–30. doi: 10.1107/S2059798322007392

40. Matsuo T, Kono F, Fujiwara S. Effects of the cardiomyopathy-causing E244D mutation of troponin T on the structures of cardiac thin filaments studied by small-angle X-ray scattering. *J Struct Biol* (2019) 205:196–205. doi: 10.1016/j.jsb.2018.12.005

41. Matsuo T, Arata T, Oda T, Fujiwara S. Difference in hydration structures between F-actin and myosin subfragment-1 detected by small-angle X-ray and neutron scattering. *Biophysics (Oxf)* (2013) 9:99–106. doi: 10.2142/biophysics.9.99

42. Porod G. General theory. In: Glatter O, Kratky O, editors. *Small angle X-ray scattering*. London: Plenum Press. (1982) p. 17–52.

43. Rambo RP, Tainer JA. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the porod-debye law. *Biopolymers* (2011) 95:559–71. doi: 10.1002/Bip.21638

44. Matsuo T. Usefulness of medium-angle X-ray scattering for structural characterization of flexible proteins studied by computer simulations. *Biochem Biophys Res Commun* (2020) 525:830–5. doi: 10.1016/j.bbrc.2020.02.150

45. Matsuo T, Arluison V, Wien F, Peters J. Structural information on bacterial amyloid and amyloid-DNA complex obtained by small-angle neutron/X-ray scattering. In: Arluison V, Wien F, Marcoleta A, editors. *Bacterial amyloids: methods and protocols, methods in molecular biology*. New Jersey: Springer, Humana Press. (2022) p. 95–107.

46. Harpaz Y, Gerstein M, Chothia C. Volume changes on protein folding. *Structure* (1994) 2:641–9. doi: 10.1016/S0969-2126(00)00065-4

47. Ko SH, Gallatin GM, Liddle JA. Nanomanufacturing with DNA origami: factors affecting the kinetics and yield of quantum dot binding. *Adv Funct Mater* (2012) 22:1015–23. doi: 10.1002/adfm.201102077

48. Huang J, Suma A, Cui M, Grundmeier G, Carnevale V, Zhang Y, et al. Arranging small molecules with subnanometer precision on DNA origami substrates for the single-molecule investigation of protein–ligand interactions. *Small Struct* (2020) 1:2000038. doi: 10.1002/sstr.202000038

49. Rosier BJHM, Cremers GAO, Engelen W, Merkx M, Brunsveld L, de Greef TFA. Incorporation of native antibodies and Fc-fusion proteins on DNA nanostructures via a modular conjugation strategy. *Chem Commun* (2017) 53:7393–6. doi: 10.1039/C7CC04178K

50. Ngo TA, Dinh H, Nguyen TM, Liew FF, Nakata E, Morii T. Protein adaptors assemble functional proteins on DNA scaffolds. *Chem Commun* (2019) 55:12428–46. doi: 10.1039/C9CC04661E

51. Funke JJ, Dietz H. Placing molecules with Bohr radius resolution using DNA origami. *Nat Nanotechnol* (2016) 11:47–52. doi: 10.1038/nnano.2015.240

52. Voigt NV, Tørring T, Rotaru A, Jacobsen MF, Ravnsbæk JB, Subramani R, et al. Single-molecule chemical reactions on DNA origami. *Nat Nanotechnol* (2010) 5:200–3. doi: 10.1038/nnano.2010.5

53. Zhao S, Tian R, Wu J, Liu S, Wang Y, Wen M, et al. A DNA origami-based aptamer nanoarray for potent and reversible anticoagulation in hemodialysis. *Nat Commun* (2021) 12:358. doi: 10.1038/s41467-020-20638-7

54. Rinker S, Ke Y, Liu Y, Chhabra R, Yan H. Self-assembled DNA nanostructures for distance-dependent multivalent ligand-protein binding. *Nat Nanotechnol* (2008) 3:418–22. doi: 10.1038/nnano.2008.164

55. Chhabra R, Sharma J, Ke Y, Liu Y, Rinker S, Lindsay S, et al. Spatially addressable multiprotein nanoarrays templated by aptamer-tagged DNA nanoarchitectures. *J Am Chem Soc* (2007) 129:10304–5. doi: 10.1021/ja072410u

56. Mei Q, Johnson RH, Wei X, Su F, Liu Y, Kelbauskas L, et al. On-chip isotachophoresis separation of functional DNA origami capture nanoarrays from cell lysate. *Nano Res* (2013) 6:712–9. doi: 10.1007/s12274-013-0347-1

57. Zhou Z, Song X, Berezov A, Zhang G, Li Y, Zhang H, et al. Human glucocorticoid-induced TNF receptor ligand regulates its signaling activity through multiple oligomerization states. *Proc Natl Acad Sci USA* (2008) 105:5465–70. doi: 10.1073/pnas.0711350105

58. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proc Natl Acad Sci USA* (2000) 97:8868–73. doi: 10.1073/pnas.160259697

frontiers | Frontiers in Oncology

TYPE Review
PUBLISHED 07 December 2023
DOI 10.3389/fonc.2023.1250766

OPEN ACCESS

EDITED BY
Yusuke Suenaga,
Chiba Cancer Center, Japan

REVIEWED BY
Barani Kumar Rajendran,
Yale University, United States
Robert J. Canter,
University of California, Davis,
United States

*CORRESPONDENCE
Kazuki Heishima
kheishima@live.jp
Naohiko Aketa
aketa.naohiko@keio.jp
Asuka Kawachi
akawachi@ncc.go.jp

†These authors have contributed equally to this work

RECEIVED 30 June 2023
ACCEPTED 06 November 2023
PUBLISHED 07 December 2023

CITATION
Heishima K, Aketa N, Heishima M and Kawachi A (2023) Hemangiosarcoma in dogs as a potential non-rodent animal model for drug discovery research of angiosarcoma in humans.
Front. Oncol. 13:1250766.
doi: 10.3389/fonc.2023.1250766

Hemangiosarcoma in dogs as a potential non-rodent animal model for drug discovery research of angiosarcoma in humans

Kazuki Heishima[1,2*†], Naohiko Aketa[3*†], Mizuki Heishima[4] and Asuka Kawachi[5,6*†]

[1]Institute for Advanced Study (GUiAS), Gifu University, Gifu, Japan, [2]Center for One Medicine Innovative Translational Research (COMIT), Gifu University, Gifu, Japan, [3]Clinical and Translational Research Center, Keio University Hospital, Tokyo, Japan, [4]IDEXX Laboratories, Tokyo, Japan, [5]Division of Cancer RNA Research, National Cancer Center, Tokyo, Japan, [6]Department of Medical Oncology, National Cancer Center Hospital, Tokyo, Japan

Since the domestication of dogs 10,000 years ago, they have shared their living environment with humans and have co-evolved. The breeding process that dogs have undergone in only a few centuries has led to a significant accumulation of specific genetic alterations that could induce particular diseases in certain breeds. These canine diseases are similar to what is found in humans with several differences; therefore, comparing such diseases occurring in humans and dogs can help discover novel disease mechanisms, pathways, and causal genetic factors. Human angiosarcoma (AS) and canine hemangiosarcoma (HSA), which are sarcomas originating from endothelium, are examples of diseases shared between humans and dogs. They exhibit similar characteristics and clinical behaviors, although with some critical differences resulting from evolution. In this review, we will describe the similarities and differences in terms of clinical and molecular characteristics between human AS and canine HSA, and discuss how these similarities and differences can be applied to advance the treatment of these diseases.

KEYWORDS

angiosarcoma, hemangiosarcoma, cancer heterogeneity, non-conventional animal model, drug discovery

Introduction

Dogs were domesticated more than 10,000 years ago in southern East Asia (1–4). Since then, humans and dogs have co-evolved in a shared living environment, exposed to the same pathological and dietary conditions (5). Dogs have undergone unique evolutionary changes through selective breeding, resulting in a diverse range of breeds with variations in

their morphology, physiology, and behavior. However, these processes have also led to a significant accumulation of genetic alterations and specific diseases in certain breeds (6, 7). Many of the diseases closely resemble disorders that affect humans with several minor differences; therefore, comparing such diseases occurring in humans and dogs can help discover novel disease mechanisms, pathways, and causal genetic factors. Indeed, this approach has successfully contributed to the discovery of novel disease mechanisms, pathways, and causal genetic factors of human diseases, including *NFAT* in SLE-like disease (8), *HAS2* biosynthesis in autoinflammatory disease (9), *LGI2* in remitting focal epilepsy (10), and *SLC4A3* in progressive retinal atrophy (11).

Representative examples of diseases shared between humans and dogs include sarcomas with vascular origins, such as human angiosarcoma (AS) and canine hemangiosarcoma (HSA) (12–14). Both human AS and canine HSA are highly aggressive sarcomas derived from vascular-forming cells, with limited treatment options and high mortality rates. They share many disease characteristics, including molecular profiles and treatment responses. However, they also exhibit critical differences in their incidence rates. Human AS is a rare cancer, accounting for approximately 0.01% of all cancers (12, 14). The rarity of human AS has hindered the development of new therapeutics and biomarkers, despite a significant unmet medical need for new diagnostics and therapies for AS patients. Even basic research tools for AS, such as cell lines and mouse models, are limited. Canine HSA, on the other hand, has an extraordinarily high incidence rate in specific dog breeds (13). The high incidence rate in dogs offers numerous advantages for investigating the clinical responses to therapeutics and the basic biology of the disease, given its clinical and genetic similarities to human AS. Therefore, canine HSA may serve as a unique model for drug discovery research aimed at providing a new treatment option to improve the prognosis of human AS. In this context, we describe the characteristics of canine HSA and discuss how their similarities and differences can be applied to advance the treatment of these diseases.

Classification and general prognosis

Different pathological terminologies have been employed for human AS and canine HSA, which are further differently subclassified based on the disease characteristics of the primary tumor site or etiology.

Human AS encompasses multiple types of endothelial cell-derived sarcomas; i.e., sarcomas derived from endothelial cells of blood vessels (HSA) and lymphatic vessels (lymphangiosarcoma). AS is typically subclassified based on the primary tumor site (cutaneous, soft-tissue, breast, and visceral AS) or etiology (lymphedema-associated and radiation-induced AS) (15–17). Each AS subtype has different prognoses and disease courses. Generally, localized cutaneous AS has a relatively favorable prognosis with 2-year overall survival rates (OS) of 71.6 to 94.1% (18); however, metastatic AS has a poor prognosis with a median OS of 8 to 9.9 months (12, 19).

The canine counterparts of human AS are still referred to separately as HSA or lymphangiosarcoma. HSA is the predominantly reported subtype, while only a small number of cases have been reported as lymphangiosarcoma (20, 21). However, it should be noted that the majority of reports use diagnostic markers such as CD34, CD31, and Factor VIII-related antigen (F8RA), which cannot differentiate between HSA and lymphangiosarcoma. Lymphatic vessel markers like LYVE-1 and PROX-1 are rarely used in veterinary medicine to exclude the possibility of lymphangiosarcoma (22). Although evidence of tumor-associated vessels containing blood cells may help to distinguish HSA from lymphangiosarcoma, it would not completely exclude the possibility of the lesion being lymphangiosarcoma. Therefore, a certain number of cases classified as HSA in veterinary medicine may have been misclassified as lymphangiosarcomas. Nevertheless, canine HSA is typically subclassified based on the primary site: visceral HSA (splenic and hepatic HSA), cardiac HSA, and cutaneous HSA (13). The prognosis is typically grave in splenic, hepatic, and cardiac HSA, with median survival times ranging from 19 to 179 days (23–30). On the other hand, cutaneous HSA has a relatively favorable prognosis, with median survival times ranging from 307 to 1189 days (31, 32).

Epidemiology

Human AS is a very rare cancer, accounting for less than 0.01% of all adult malignancies (14, 33). AS is more likely to occur on the skin of white, elderly individuals, but there are no significant differences in distribution between sexes (33). While AS can arise from any soft tissue or organ with vascular tissues, it most commonly affects the skin of the head, neck, scalp, breast, and extremities. Visceral forms of AS, occurring in the liver, right atrium of the heart, and spleen, are less frequent.

Dogs have a significantly higher incidence of HSA compared to humans, with an estimated 25 to 100 times higher incidence rate (34). HSA accounts for 5% of all non-cutaneous malignant neoplasms in dogs (13, 34) and approximately 50% of all splenic tumors (13, 35). Similar to humans, HSA predominantly affects older animals, although there seems to be a slight male predisposition in dogs (36, 37). Golden Retrievers and German Shepherds are high-risk breeds, and HSA is the leading cause of cancer-associated death in Golden Retrievers (28, 38). Like in humans, HSA can arise from any soft tissue with vasculature, but it most commonly affects the spleen, right atrium of the heart, liver, and skin or subcutaneous tissue (13). The different anatomical distribution observed in human AS and canine HSA is one of the characteristic differences, although the exact underlying cause is unknown.

Etiology

Multiple potential risk factors have been identified in human AS, including chronic lymphedema, radiotherapy, UV radiation, BRCA mutation, familial syndromes, chemical exposure, foreign bodies, and immunosuppression. Radiotherapy and chronic

lymphedema are well-known risk factors for AS (39, 40). Breast cancer patients who receive adjuvant radiotherapy are predisposed to developing chronic lymphedema and subsequent breast AS as unintended side effects of treatment. This condition is known as Stewart-Treves syndrome (39). The highest incidence of AS in breast cancer patients occurs 5-10 years after adjuvant radiotherapy (40). The risk of AS further increases in patients with mutations in the *BRCA1* (185delAG) and *BRCA2* (854delC) genes (41), which are crucial for DNA repair. Milroy's disease and chronic filariasis also cause chronic lymphedema and are associated with the development of AS (15). AS is associated with various familial syndromes such as neurofibromatosis, Maffucci syndrome, Li-Fraumeni syndrome (*TP53* mutations), and Klippel-Trenaunay syndrome (*PIK3CA* mutations) (15), which is consistent with the observed profiles of recurrent mutations in AS. Chemical exposure is another risk factor for AS development, including exposure to vinyl chloride (42), thorium dioxide (43), arsenic, radium, and anabolic steroids (44). Among these, vinyl chloride and thorium dioxide predominantly induce hepatic AS (42, 43). Foreign bodies can also cause AS, such as surgical gauzes (45), vascular prostheses (46), orthopedic prostheses (47), and gouty tophus (48). The causal association between immunosuppression and AS tumorigenesis remains unclear; however, AS has been observed in immunosuppressed patients following renal transplantation, and some epidemiological studies suggest a potential association between AS and AIDS (49).

Not many risk factors have been reported for canine HSA so far; however, the identified risk factors include dog breeds such as German Shepherds or Golden Retrievers (50) and UV radiation (51, 52). Dog breeds such as German Shepherds or Golden Retrievers are considered strong risk factors, possibly due to genetic imbalances resulting from intense inbreeding and selection. A genome-wide association study reported several loci significantly associated with the risk of HSA in Golden Retrievers (50). Like human AS, UV radiation has also been associated with cutaneous HSA (51, 52), which exhibits a high mutation rate and a strong UV mutational signature (53).

Pathology

Both human AS and canine HSA are heterogeneous tumors with significant intra- and intertumoral differences. The hallmark of AS is the proliferation of pleomorphic endothelial cells showing rounded, polygonal, fusiform, or epithelioid morphology without a clear border to normal tissue. Well-differentiated AS may contain abnormal endothelial cells forming vascular sinusoids continuous with normal vascular channels; however, aggressive and poorly differentiated AS tends to lose such architecture and has an epithelioid morphology with a high mitotic rate and areas of hemorrhage and necrosis (15, 54, 55), which grants AS tissue prominent complexity and heterogeneity. AS typically expresses endothelial markers including von Willebrand factor, CD34, CD31, and VEGF. Among these, von Willebrand factor (15) and CD31 are the most commonly used markers for distinguishing AS from other undifferentiated neoplasms (56). However, progressive AS may lose these expressions and gain expression of cytokeratins, which can cause confusion with undifferentiated epithelial malignancies (56).

Canine HSA exhibits similar histological characteristics to human AS. The histologic features include the proliferation of immature, pleomorphic endothelial cells with expression of von Willebrand factor, CD31, claudin-5, CD117, and VEGF (57–61). Similarly, von Willebrand factor and CD31 are the most commonly used markers for distinguishing HSA from other malignancies. An intriguing hypothesis regarding the origin of canine HSA has recently been proposed. Canine HSA has been considered to originate from vascular endothelium based on histopathological findings. However, recent studies have revealed that these malignant cells may originate from pluripotent bone marrow progenitors at the stage of hemangioblasts to angioblasts differentiating into endothelial cells (34, 62, 63). This hypothesis has been reported only in canine HSA; however, given the similarities between human AS and canine HSA, it may also be relevant to the origins of human AS.

Molecular abnormalities

Human AS and canine HSA share many similar genetic abnormalities with minor differences (Figure 1, Table 1). Human AS reportedly has many types of molecular abnormalities; however, common driver pathogenic mutations or copy number aberrations shared in all reported cases have not been identified, likely due to the high heterogeneity in primary locations or etiologic factors of AS. Recurrent mutations in human AS include *KDR* (VEGFR2) (64), *TP53* (67, 73), and *PIK3CA* (64). Other genetic abnormalities of AS include mutations of *KRAS* (69, 74), MAPK (65), *PTPRB*, and *PLCG* (70), as well as amplifications of *KDR* (65, 75), *VEGFA* (72, 73), *MYC* (65, 75), *KIT* (64), and deletions of *CDKN2A* (65, 75). Among these, *PTPRB* and *PLCG* mutations and *MYC* amplification are most frequently observed in secondary AS, such as radiation-induced AS. Other factors are also reported in association with AS pathogenesis: overexpression of *WT1* (Wilms Tumor 1) (76), *LGALS3* (Galectin-3) (77), *ETS1*, metalloproteinases (MMP1, MMP3, and uPA) (78, 79), and *FSCN* (the actin-bundling motility protein) (72). AS has a *KIT* expression (80–82) and amplification (64), whereas no activating mutations in exons 11 (juxtamembrane domain) or 17 (kinase domain) of *KIT* have been observed so far.

Angiogenic pathways have long been recognized for their pivotal roles in the context of Angiosarcoma (AS). Specifically, the central components of these angiogenic pathways, namely *KDR* (VEGFR2) and *VEGFA*, often exhibit mutations and amplifications, as documented in multiple studies (64, 72, 73). KDR serves as the principal receptor tyrosine kinase responsible for mediating VEGFA-induced proangiogenic signaling. As such, it was initially anticipated that the loss of KDR signaling might suppress the proliferation and metastasis of AS, correlating with a more favorable prognosis. Intriguingly, however, the results proved contrary, revealing that the loss of KDR is associated with a less favorable prognosis (83). These findings suggest that KDR may not be a contributing factor to the tumorigenesis of AS but rather a

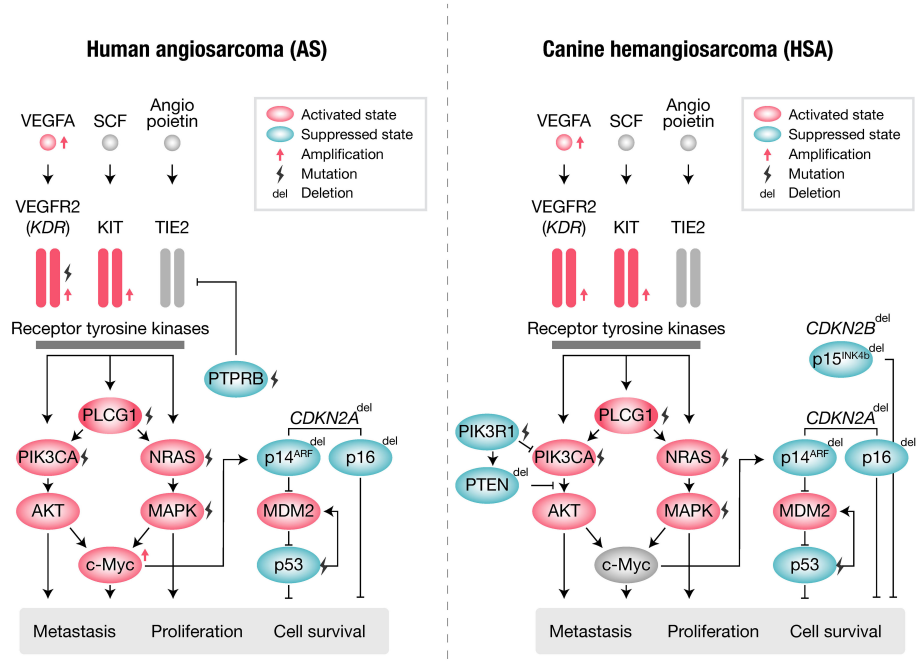


FIGURE 1
Summary of gene abnormalities in human angiosarcoma (AS) and canine hemangiosarcoma (HSA).

potential regulator of endothelial cell differentiation. This may be relevant to the disappointing results of clinical trials for bevacizumab, an antibody targeting VEGF. In a phase II study, only 2 out of 23 patients achieved partial response, and 11 had stable disease (84), and even when combined with Paclitaxel, the response rate was only 28% compared to 45.8% in the combination or monotherapy with Paclitaxel (85).

RAS and its downstream pathways appear to play important roles in AS. The mutation of *NRAS* and *KRAS* has been observed in AS (64, 69, 74). Furthermore, a functional study revealed that the

TABLE 1 Gene abnormalities shared between human angiosarcoma (AS) and canine hemangiosarcoma (HSA).

	Human AS		Canine HSA	
	Mutation	Amplification	Mutation	Amplification
KDR (VEGFR2)	12/47 (26%) Painter et al. (64)	4/34 Murali et al. (65)	–	14/47 (22%) Megquier et al. (66)
TP53	14/47 (30%) Painter et al. (64) 7/13 (53.8%) Naka et al. (67)	–	33/50 (66%) Wang et al. (68) 14/15 (93.33%) Wong et al. (53)	–
PIK3CA	10/47 (21%) Painter et al. (64)	–	23/50 (46%) Wang et al. (68) 14/47 (29.8%) Megquier et al. (66)	–
RAS	NRAS 3/47 (6%) Painter et al. (64) KRAS 8/15 (53.3%) Weihrauch et al. (69)	–	NRAS 12/50 (24%) Wang et al. (68)	–
PLCG	8/47 (17%) Painter et al. (64) 3/34 (8.8%) Behjati et al. (70)	–	2/50 (4%) Wang et al. (68) 1/20 (5%) Wang et al. (71)	–
VEGFA	–	20/20 (100%) Dim et al. (72)	–	9/47 (19%) Megquier et al. (66)
KIT	–	4/47 (9%) Painter et al. (64)	–	8/47 (17%) Megquier et al. (66)
CDKN2A (p14ARF, p16)	Deletion 9/34 (26%) Murali et al. (65)	–	Deletion 10/47 (22%) Megquier et al. (66)	–

introduction of continuously activated *HRAS* into murine endothelial cells produces poorly differentiated AS *in vivo* (86), indicating essential roles of the RAS pathway in the tumorigenesis of AS. Among the downstream pathways of RAS, the activation of the PI3K pathway has been frequently documented in AS patients. Of note, several studies have suggested that PI3K signaling is more important than the MAPK signaling cascade in AS (86, 87). Although activation of PI3K pathways has been frequently noted, the mutation of *PIK3CA* itself had not been reported in AS studies. However, a recent international cooperative project (Angiosarcoma Project) revealed that the *PIK3CA* activating mutation was one of the most frequently mutated genes in human AS (64). These reports suggest a strong contribution of PI3K pathways to the tumorigenesis of AS.

Canine HSA has been reported to bear many shared genetic abnormalities with human AS. The most commonly observed recurrent mutations in canine HSA include *TP53* (66), *PIK3CA* (activating) (66, 71), and *PIK3R1* (66), *PTEN* (inactivating) (71), *NRAS* (68), and *PLCG1* (66, 68). Additionally, canine HSA frequently has other genetic abnormalities, including the deletion of *CDKN2A/B* (66) and *PTEN* (88) and the amplification of *KDR*, *VEGFA*, and *KIT* (66). However, unlike in human AS, the amplification of *MYC* is not frequently observed in canine HSA (89). This is likely due to secondary HSA not being prevalent in dogs.

Many components of the angiogenic pathways are also altered in canine HSA. Consistent with the profiles of human AS, canine HSA frequently exhibits activation of the VEGFA-KDR pathway (66). Similarities have also been observed in the clinical response to inhibitors of angiogenic pathways. Despite the predominant activation of the VEGFA-KDR pathway, a small molecule inhibitor targeting canine KDR, toceranib, failed to induce sufficient clinical responses in canine HSA (90). The results from veterinary clinical trials strongly suggest fundamental similarities in the response to treatment between human AS and canine HSA.

Canine HSA also shares gene abnormalities in the RAS and PI3K pathways, as observed in human AS (68, 71). Intriguingly, activating mutations in *PIK3CA* were first identified in canine HSA by veterinary researchers (71). Initially, these activating mutations in *PIK3CA* were considered to be specific to canines. However, a recent large-scale analysis (Angiosarcoma Project) revealed that human AS also frequently has activating mutations in *PIK3CA* (64). This exemplifies that canine HSA has highly similar molecular characteristics to human AS, indicating the strong potential of canine HSA as a model with similar molecular and clinical characteristics.

Treatment

Currently, there are no effective therapeutic options for both human AS and canine HSA. Human AS has seldom been the primary focus of clinical trials due to its uncommon occurrence; instead, it has frequently been included in prospective clinical trials as part of the sarcoma in general without specialized treatment. To date, there have been no prospective clinical trials for AS that demonstrated a prominent survival benefit of systemic therapy in either the neoadjuvant or adjuvant setting. In line with this, many of the discussions regarding therapeutic options for AS are based on limited evidence reported from retrospective studies, albeit with a few prospective studies.

The primary choice of treatment for human AS with localized lesions remains radical surgery with complete resection. However, due to the invasive nature of AS, achieving complete resection with clear margins is often challenging (17, 91–93). Given the circumstances, neoadjuvant chemotherapy may be performed using gemcitabine, docetaxel, doxorubicin, ifosfamide, and paclitaxel; however, no survival benefit has been reported with the addition of neoadjuvant chemotherapy (94). Concurrent therapy with paclitaxel and radiation has been explored for localized cutaneous AS, and a prospective study has shown improved survival with a 2-year overall survival (OS) of 94.1% compared to 71.6% in the control group (18). In the view of treating metastatic disease, several cytotoxic drugs and regimens are used for this purpose. One of the most commonly used agents is taxane-based regimens involving paclitaxel. A Phase II single arm clinical trial demonstrated that paclitaxel has particular activity in AS and is now often used in the first or second-line setting (18). A prospective study showed that the response rate for this treatment was approximately 19%, with a median progression-free survival (PFS) of 4 months and OS of 8 months (19). Anthracycline-based regimens using doxorubicin are also often utilized, with a response rate of approximately 25%, and a median PFS of 4.9 months and OS of 9.9 months (prospective study) (95).

Canine visceral HSA is usually treated with surgery and adjuvant chemotherapy, and most of the prognostic information in the literature is from retrospective analysis. For splenic HSA, total splenectomy and adjuvant chemotherapy with a single agent or combination protocols involving doxorubicin are recommended (13). However, the therapeutic efficacy is limited, and significant improvement in survival time is rarely achieved. Splenectomy alone has shown median survival times ranging from 19 to 86 days (23–27), with a 2-month survival rate of 31% and a 1-year survival rate of 7% (96). Surgery and adjuvant chemotherapy have resulted in median survival times of 141–179 days, with less than 10% of dogs surviving beyond one year (28–30). Cardiac HSA treated with surgical excision and adjuvant doxorubicin-based chemotherapy showed a similar response to splenic HSA, with median survival times of 183–189 days (97, 98). For cutaneous HSA, surgery with or without adjuvant chemotherapy is generally performed (13). Wider margins are recommended for surgery, although it may be difficult depending on the tumor location, similar to human AS (13). Cutaneous HSA has median survival times of 307 days (with invasion into the surrounding tissue) or greater than 2 years (without evidence of invasion), which is a much better prognosis compared to visceral HSA (31). Cutaneous HSA treated with surgery and adjuvant doxorubicin has shown median survival times as long as 1189 days (32). Only a few studies have reported

the efficacy of radiation therapy for canine HSA, and none have reported a significant improvement in overall survival (99).

Other treatment options and potential treatment under investigation

For the treatment of human AS, several targeted agents have been explored as alternative options. Agents targeting the VEGF-VEGFR angiogenic pathway have been assessed for sarcoma, including bevacizumab and pazopanib. Bevacizumab, a monoclonal antibody targeting VEGFA (84, 100, 101), produced modest results with only 2 out of 23 patients showing a partial response and 11 out of 23 patients with stable disease in a Phase II clinical trial (84). The results were disappointing considering the well-known alteration of the VEGFA-VEGFR pathway in AS. Pazopanib, a tyrosine kinase inhibitor, also yielded modest results with a median progression-free survival (PFS) of 3 months and no significant responses in a retrospective study (102). Immune checkpoint inhibitors (ICIs) have recently emerged as another option for AS. The anti-PD1 (programmed death 1) checkpoint inhibitor pembrolizumab was approved for tumors with a high tumor mutation burden regardless of histology (103). Pembrolizumab showed an exceptional and durable response to 2 out of 3 metastatic AS cases that were refractory to standard therapies (64). Although further studies are required to confirm their efficacy, these reports suggest the promising potential of ICIs for AS treatment.

For canine HSA, multiple agents have been explored as alternative options, including a tyrosine kinase inhibitor targeting KDR (toceranib) (90), a taxane-based agent (Paccal-Vet) (104), immune checkpoint inhibitors (ICIs) (105, 106), other forms of anthracycline (epirubicin) (107), pegylated liposome-encapsulated doxorubicin (108), COX-2 inhibitors (109), and thalidomide (110). Toceranib, a tyrosine kinase inhibitor targeting KDR (90), has been explored for the treatment of canine HSA; however, the results were disappointing: a prospective study showed that the use of toceranib following doxorubicin-based chemotherapy did not improve either disease-free interval or OS in stage I or II canine HSA (a median disease-free interval, 161 days; a median survival time, 172 days) (90). Paccal Vet, a water-soluble, micellar formulation of paclitaxel, has been investigated for treating canine HSA. Paclitaxel has not been used in dogs due to high rates of hypersensitivity reactions when given intravenously (111, 112); however, Paccal Vet is designed not to induce such hypersensitivity, thus it is expected to be useful for the treatment of canine HSA (104). Although ICIs had not been commercially available for dogs, Gilvetmab, the first anti-canine PD-1 antibody, has been approved in October 2023 (Merck Animal Health USA). Currently, Gilvetmab is approved only for 2 tumor types (mast cell tumor and malignant melanoma), and its clinical efficacy for canine HSA is still unclear. Although further studies are required, ICIs have promising potential as a therapeutic modality for canine HSA as in human AS. Several other forms of conventional immunotherapy are documented, including an HSA vaccine (113), liposome-encapsulated muramyl tripeptide phosphatidylethanolamine (LMTP-PE) (114), and polysaccharopeptide (115); however, their efficacy may be limited.

Issues in drug discovery research for human AS

Major issues associated with drug discovery research for human AS are caused by its rarity and can be subdivided into scientific or investment issues. The scientific issues include (1) lack of comprehensive information regarding treatment and biological characteristics and (2) the lack of appropriate research models.

The majority of information regarding therapeutic responses and biological characteristics of AS is based on the results from case series, which may possess potential bias in selection, and the dataset may be incomplete or a mixture of the results from different treatment approaches (12). Randomized trials are lacking, and there are limited prospective studies (12). Nonetheless, recent initiatives for international collaboration, such as The Angiosarcoma Project (64) and the International Rare Diseases Research Consortium (IRDiRC), hold promise for gradually addressing these challenges.

Cell lines and conventional xenograft mouse models useful for AS research are also lacking. The establishment of AS cell lines is extremely difficult. As a result, well-characterized and commonly used cell lines are limited to only a few, including ISO-HAS (116) and ASM (117). Similarly, the availability of cell-derived xenograft models (CDX) and patient-derived xenograft models (PDX) is limited, although a recent study reported a rare successful example (118).

The low incentives for pharmaceutical companies to invest in new drugs for AS are another challenge (119). Drug development specifically for AS carries a high risk of failing to recoup the invested funds and is thus rarely pursued. Additionally, few clinical trials evaluating new compounds for other major cancers include patients with rare cancers such as AS. This is because the inclusion of such patients increases the risk of experiencing unexpected adverse events, which could critically halt the entire development process. Furthermore, the additional costs related to managing compound supply, developing companion diagnostics, and applying for regulatory approvals are not easily justified as investments for pharmaceutical companies. Given these circumstances, current drug discovery research for AS mainly focuses on drug repositioning studies that can be conducted with relatively low-risk investments. However, relying solely on these drug repositioning studies remains challenging to achieve the development of truly effective therapeutic drugs for AS.

The inefficient collection of patients for clinical trials or clinical samples has been a factor that hinders drug discovery research for AS. However, this is getting less problematic due to recently improved cooperation between specialized hospitals or institutions for the registration and enrollment of clinical trials (64).

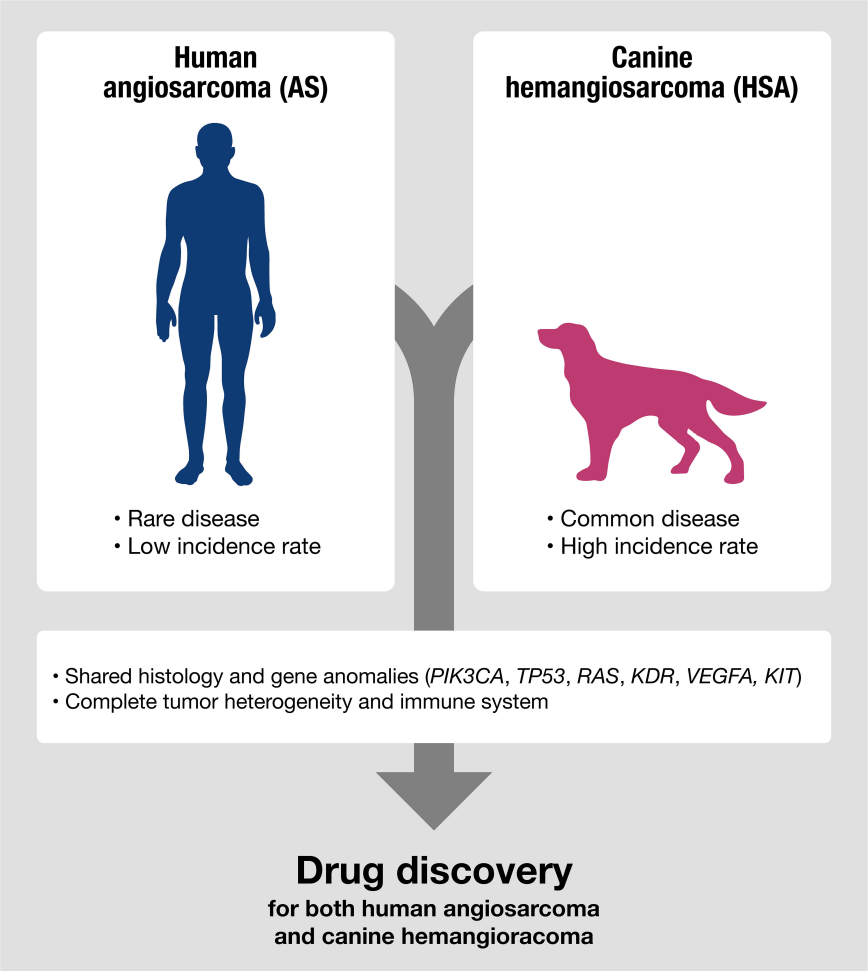


FIGURE 2
Canine HSA shares molecular signatures with human AS, making it a potentially useful model for predicting therapeutic responses in human AS, especially for evaluating the efficacy of drugs targeting shared genetic anomalies between humans and dogs.

Advantages and disadvantages of canine HSA as a drug discovery model

Canine HSA has advantages as a non-conventional model for drug discovery (Figure 2). Despite canine HSA having an exceptionally higher incidence rate than its human analog, it retains significant similarities with human AS, including multistep tumorigenesis that occurs over several years, histological and molecular characteristics, a complete immune system, intra/intertumoral heterogeneity, and living environment. Furthermore, many HSA cell lines have been established and are available for basic research (63, 113, 120–125). The value of the dog model is further highlighted given the advantages in the investigation of immunotherapy. Since the recent broad applications of ICIs, understanding clinical responses to immunotherapy in the heterogeneous tumor microenvironment seen in human cancers is gaining more importance. However, conventional rodent models hardly serve as appropriate models for this purpose due to their non-spontaneous and artificial nature of tumors with dysfunctional immune systems. In contrast, canine HSA is a naturally occurring tumor, which has genetic heterogeneity and an intact immune system that could closely recapitulate the complexity of human cancers. The recent development of a canine version of ICI (Gilvetmab) could further facilitate the investigation of clinical responses to immunotherapy in the setting of genetic and immunologic heterogeneity of tumors. These characteristics and the research environment are useful for analyzing therapeutic responses and the complexities of drug resistance, metastasis, and tumor-host immune interactions in AS patients.

On the other hand, the strategy of using canine HSA as a drug discovery model also has certain disadvantages to consider. Conducting drug discovery research using canine HSA in a clinical setting could be more expensive compared to conventional preclinical studies using laboratory animals such as mice and rats. Dogs have a larger body size than laboratory animals, which requires a higher amount of drugs for testing. Additionally, to acquire high-quality clinical data to evaluate the response to drugs, frequent medical checks during administration may be necessary. These medical checks typically include a set of assessments such as complete blood cell count, blood

biochemistry, urinalysis, ultrasound, X-ray, computed tomography (CT), and magnetic resonance imaging (MRI), which significantly increases the total costs. Moreover, systemic anesthesia is typically required for dogs undergoing CT and MRI, adding extra costs and medical risks to the process of performing these evaluations. Sometimes systemic anesthesia could result in fatal situations, particularly for elderly dogs with cancer, making it challenging to perform these evaluations multiple times from an animal welfare perspective. Furthermore, from a regulatory perspective, the role and purpose of using spontaneous dog tumor models are not clearly defined in the status quo. Therefore, the data collected from canine trials may be considered as supplemental information, which may not have significant impacts on drug approval. Additionally, drug repositioning studies can be directly initiated with humans, reducing the relative necessity of canine models for such purposes. Moreover, although many similarities have been reported between human AS and canine HSA, these diseases are not entirely identical. Therefore, for each study, performing appropriate basic research is necessary to validate that the target mechanism of action is actually conserved between human AS and canine HSA.

Potentials as a model for drug efficacy evaluation

As discussed above, canine HSA is a potentially useful model for evaluating drug efficacy; however, more research is needed to fully utilize its potential. To facilitate the effective use of this model in drug discovery, several key factors should be considered from the current regulatory perspective. The most critical factor to consider is the scientific validity of the model, given that there are currently no established regulatory guidelines for utilizing canine HSA in evaluating drug efficacy (126). Scientific validity is typically determined by whether the model appropriately reflects the molecular mechanisms, physiological conditions, and microenvironment of the target tumors in humans (126). Therefore, conducting basic research to uncover the more detailed characteristics of HSA, as well as establishing guidelines for model evaluation by academic societies, can be an effective strategy to promote the use of canine HSA as a valuable tool in drug discovery. Canine HSA, in this role, holds the potential to help mitigate the risks associated with drug development for human AS. This can be accomplished by early identification of agents with promising activity and safety, effectively distinguishing them from those likely to fail in the drug development journey. Such an approach represents a promising avenue for advancing AS drug development and research progress.

Canine HSA may be useful for basic research aimed at identifying novel mechanisms or therapeutic targets for the development of new drugs, given that it appears to share fundamental similarities with human AS in terms of biological characteristics and therapeutic responses to drugs. Nonetheless, the investigation of new drugs for AS may be challenging due to non-scientific reasons, including low incentives for investing in drug discovery for rare cancers, as mentioned in a previous section. However, a recent approach that targets a limited disease population and recovers the investment through increased drug prices (low population/high margin) may help overcome this situation in drug discovery for rare cancers (119). Several new drugs for rare diseases have been successfully developed using this approach, such as Zolgensma for spinal muscular atrophy (affecting 1 out of 10,000 people) (127) and Tofersen (128) for SOD1-mutated amyotrophic lateral sclerosis (affecting 3 out of 1,000,000 people). Utilizing such an approach may help overcome the current situation filled with difficulties, and the results obtained from basic research using canine HSA still have the potential to pave the way for future new drug development for AS. Additionally, canine HSA may also be a useful research model to explore appropriate doses and intervals in exploratory studies involving the repositioning of drugs that are approved for other cancers. Also, the heterogeneous nature of canine HSA, which includes the complexity of the tumor microenvironment with a complete immune system, may help collect crucial information for understanding factors essential for drug resistance and recurrence.

Potentials as a model for toxicity evaluation

Canine HSA may not be an effective model for toxicity evaluation in drug development as it does not meet the standards of Good Laboratory Practice (GLP) studies. Companion dogs with HSA lack sufficient background or reference data, making it difficult to interpret the results and reducing the reliability of the data compared to toxicity studies conducted under GLP conditions. The nature of this model makes it challenging to determine whether the observed toxicity is caused by disease exacerbation, drug effects, or individual differences within the model. This critical disadvantage hinders the scientific investigation of the underlying causes of toxicity observed in the model. The use of non-conventional animal models may be considered only when general toxicity evaluations are deemed inappropriate. One example is biopharmaceuticals that do not exhibit pharmacological effects in conventional animal models. In such cases, non-conventional animal models can be justified by examining species differences in terms of pharmacological effects, target molecule distribution and expression, and tumor behavior, with reference to ICH guideline S6 (129). However, it is unrealistic to check all these aspects in companion animals; therefore, transgenic animals are commonly used as substitutes for this purpose. The results of efficacy studies described in the previous section may be included as supplemental information for safety evaluations. However, they are typically considered as supplemental reference data with less impact on drug approvals. Therefore, the companion animal model cannot provide the same quality of toxicity data, limiting its use in toxicological studies.

Conclusion

The development of new drug discovery for AS has remained elusive, and the challenges outlined in this review persist. However, it is crucial that these limitations do not deter the commitment of our research community to advancing therapeutic options for patients with AS. As illustrated in this review, canine HSA exhibits significant molecular similarities to human AS, making it a valuable model for predicting therapeutic responses in human AS, especially when evaluating the efficacy of drugs targeting shared genetic anomalies between humans and dogs. This review represents an initial step toward the development of novel drugs for AS and HSA. Nevertheless, further foundational research is imperative to enhance the utilization of canine HSA as a model for drug discovery.

Author contributions

KH, NA, MH, and AK wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

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

Publisher's note

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

References

1. Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* (2013) 495(7441):360–4. doi: 10.1038/nature11837

2. Davis SJ, Valla FR. Evidence for domestication of the dog 12,000 years ago in the Natufian of Israel. *Nature* (1978) 276(5688):608–10. doi: 10.1038/276608a0

3. Pang JF, Kluetsch C, Zou XJ, Zhang AB, Luo LY, Angleby H, et al. Mtdna data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Mol Biol Evol* (2009) 26(12):2849–64. doi: 10.1093/molbev/msp195

4. Skoglund P, Gotherstrom A, Jakobsson M. Estimation of population divergence times from non-overlapping genomic sequences: examples from dogs and wolves. *Mol Biol Evol* (2011) 28(4):1505–17. doi: 10.1093/molbev/msq342

5. Wang GD, Zhai W, Yang HC, Fan RX, Cao X, Zhong L, et al. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat Commun* (2013) 4:1860. doi: 10.1038/ncomms2814

6. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* (2005) 438(7069):803–19. doi: 10.1038/nature04338

7. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, et al. Extensive and breed-specific linkage disequilibrium in Canis familiaris. *Genome Res* (2004) 14(12):2388–96. doi: 10.1101/gr.3147604

8. Wilbe M, Jokinen P, Truve K, Seppala EH, Karlsson EK, Biagi T, et al. Genome-wide association mapping identifies multiple loci for a canine Sle-related disease complex. *Nat Genet* (2010) 42(3):250–4. doi: 10.1038/ng.525

9. Olsson M, Meadows JR, Truve K, Rosengren Pielberg G, Puppo F, Mauceli E, et al. A novel unstable duplication upstream of Has2 predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs. *PloS Genet* (2011) 7(3):e1001332. doi: 10.1371/journal.pgen.1001332

10. Seppala EH, Jokinen TS, Fukata M, Fukata Y, Webster MT, Karlsson EK, et al. Lgi2 truncation causes a remitting focal epilepsy in dogs. *PloS Genet* (2011) 7(7): e1002194. doi: 10.1371/journal.pgen.1002194

11. Downs LM, Wallin-Hakansson B, Boursnell M, Marklund S, Hedhammar A, Truve K, et al. A frameshift mutation in golden retriever dogs with progressive retinal atrophy endorses Slc4a3 as a candidate gene for human retinal degenerations. *PloS One* (2011) 6(6):e21452. doi: 10.1371/journal.pone.0021452

12. Young RJ, Brown NJ, Reed MW, Hughes D, Woll PJ. Angiosarcoma. *Lancet Oncol* (2010) 11(10):983–91. doi: 10.1016/S1470-2045(10)70023-1

13. Vail DM, Thamm DH, Liptak JM. *Withrow and Macewen's Small Animal Clinical Oncology*. Oxford, UK: Elsevier Health Sciences (2019).

14. Lahat G, Dhuka AR, Hallevi H, Xiao L, Zou C, Smith KD, et al. Angiosarcoma: clinical and molecular insights. *Ann Surg* (2010) 251(6):1098–106. doi: 10.1097/SLA.0b013e3181dbb75a

15. Weiss SW, Goldblum JR, Folpe AL. *Enzinger and Weiss's Soft Tissue Tumors*. Oxford, UK: Elsevier Health Sciences (2007).

16. Schlemmer M, Reichardt P, Verweij J, Hartmann JT, Judson I, Thyss A, et al. Paclitaxel in patients with advanced angiosarcomas of soft tissue: a retrospective study of the EORTC soft tissue and bone sarcoma group. *Eur J Cancer* (2008) 44(16):2433–6. doi: 10.1016/j.ejca.2008.07.037

17. Fury MG, Antonescu CR, Van Zee KJ, Brennan MF, Maki RG. A 14-year retrospective review of angiosarcoma: clinical characteristics, prognostic factors, and treatment outcomes with surgery and chemotherapy. *Cancer J* (2005) 11(3):241–7. doi: 10.1097/00130404-200505000-00011

18. Florou V, Wilky BA. Current management of angiosarcoma: recent advances and lessons from the past. *Curr Treat Options Oncol* (2021) 22(7):61. doi: 10.1007/s11864-021-00858-9

19. Penel N, Bui BN, Bay JO, Cupissol D, Ray-Coquard I, Piperno-Neumann S, et al. Phase II trial of weekly paclitaxel for unresectable angiosarcoma: the angiotax study. *J Clin Oncol* (2008) 26(32):5269–74. doi: 10.1200/JCO.2008.17.3146

20. Kelly WR, Wilkinson GT, Allen PW. Canine angiosarcoma (Lymphangiosarcoma). *Vet Pathol* (1981) 18(2):224–7. doi: 10.1177/030098588101800210

21. Curran KM, Halsey CH, Worley DR. Lymphangiosarcoma in 12 dogs: A case series (1998-2013). *Vet Comp Oncol* (2016) 14(2):181–90. doi: 10.1111/vco.12087

22. Halsey CH, Worley DR, Curran K, Charles JB, Ehrhart EJ. The use of novel lymphatic endothelial cell-specific immunohistochemical markers to differentiate Cutaneous angiosarcomas in dogs. *Vet Comp Oncol* (2016) 14(3):236–44. doi: 10.1111/vco.12088

23. Prymak C, McKee LJ, Goldschmidt MH, Glickman LT. Epidemiologic, clinical, pathologic, and prognostic characteristics of splenic hemangiosarcoma and splenic hematoma in dogs: 217 cases (1985). *J Am Vet Med Assoc* (1988) 193(6):706–12.

24. Wood CA, Moore AS, Gliatto JM, Ablin LA, Berg RJ, Rand WM. Prognosis for dogs with stage I or II splenic hemangiosarcoma treated by splenectomy alone: 32 cases (1991-1993). *J Am Anim Hosp Assoc* (1998) 34(5):417–21. doi: 10.5326/15473317-34-5-417

25. Goritz M, Muller K, Krastel D, Staudacher G, Schmidt P, Kuhn M, et al. Canine splenic haemangiosarcoma: influence of metastases, chemotherapy and growth pattern on post-splenectomy survival and expression of angiogenic factors. *J Comp Pathol* (2013) 149(1):30–9. doi: 10.1016/j.jcpa.2012.11.234

26. Batschinski K, Nobre A, Vargas-Mendez E, Tedardi MV, Cirillo J, Cestari G, et al. Canine visceral hemangiosarcoma treated with surgery alone or surgery and doxorubicin: 37 cases (2005-2014). *Can Vet J* (2018) 59(9):967–72.

27. Wendelburg KM, Price LL, Burgess KE, Lyons JA, Lew FH, Berg J. Survival time of dogs with splenic hemangiosarcoma treated by splenectomy with or without

adjuvant chemotherapy: 208 cases (2001–2012). *J Am Vet Med* (2015) 247(4):393–403. doi: 10.2460/javma.247.4.393

28. Brown NO, Patnaik AK, MacEwen EG. Canine hemangiosarcoma: retrospective analysis of 104 cases. *J Am Vet Med Assoc* (1985) 186(1):56–8.

29. Hammer AS, Couto CG, Filppi J, Getzy D, Shank K. Efficacy and toxicity of ac chemotherapy (Vincristine, doxorubicin, and cyclophosphamide) in dogs with hemangiosarcoma. *J Vet Intern Med* (1991) 5(3):160–6. doi: 10.1111/j.1939-1676.1991.tb00943.x

30. Lana S, U'Ren L, Plaza S, Elmslie R, Gustafson D, Morley P, et al. Continuous low-dose oral chemotherapy for adjuvant therapy of splenic hemangiosarcoma in dogs. *J Vet Intern Med* (2007) 21(4):764–9. doi: 10.1892/0891-6640(2007)21[764:clocfa]2.0.co;2

31. Ward H, Fox LE, Calderwood-Mays MB, Hammer AS, Couto CG. Cutaneous hemangiosarcoma in 25 dogs: A retrospective study. *J Vet Intern Med* (1994) 8(5):345–8. doi: 10.1111/j.1939-1676.1994.tb03248.x

32. Bulakowski EJ, Philibert JC, Siegel S, Clifford CA, Risbon R, Zivin K, et al. Evaluation of outcome associated with subcutaneous and intramuscular hemangiosarcoma treated with adjuvant doxorubicin in dogs: 21 cases (2001-2006). *J Am Vet Med Assoc* (2008) 233(1):122–8. doi: 10.2460/javma.233.1.122

33. Rouhani P, Fletcher CD, Devesa SS, Toro JR. Cutaneous soft tissue sarcoma incidence patterns in the U.S. : an analysis of 12,114 cases. *Cancer* (2008) 113(3):616–27. doi: 10.1002/cncr.23571

34. Kim JH, Graef AJ, Dickerson EB, Modiano JF. Pathobiology of hemangiosarcoma in dogs: research advances and future perspectives. *Vet Sci* (2015) 2(4):388–405. doi: 10.3390/vetsci2040388

35. Spangler WL, Culbertson MR. Prevalence, type, and importance of splenic diseases in dogs: 1,480 cases (1985-1989). *J Am Vet Med Assoc* (1992) 200(6):829–34.

36. Gamlem H, Nordstoga K, Arnesen K. Canine vascular neoplasia–a population-based clinicopathologic study of 439 tumours and tumour-like lesions in 420 dogs. *APMIS Suppl* (2008) 116(125):41–54. doi: 10.1111/j.1600-0463.2008.125m4.x

37. Schultheiss PC. A retrospective study of visceral and nonvisceral hemangiosarcoma and hemangiomas in domestic animals. *J Vet Diagn Invest* (2004) 16(6):522–6. doi: 10.1177/104063870401600606

38. Kent MS, Burton JH, Dank G, Bannasch DL, Rebhun RB. Association of cancer-related mortality, age and gonadectomy in golden retriever dogs at a veterinary academic center (1989-2016). *PloS One* (2018) 13(2):e0192578. doi: 10.1371/journal.pone.0192578

39. Stewart FW, Treves N. Lymphangiosarcoma in postmastectomy lymphedema; a report of six cases in Elephantiasis chirurgica. *Cancer* (1948) 1(1):64–81. doi: 10.1002/1097-0142(194805)1:1<64::aid-cncr2820010105>3.0.co;2-w

40. Huang J, Mackillop WJ. Increased risk of soft tissue sarcoma after radiotherapy in women with breast carcinoma. *Cancer* (2001) 92(1):172–80. doi: 10.1002/1097-0142(20010701)92:1<172::aid-cncr1306>3.0.co;2-k

41. West JG, Weitzel JN, Tao ML, Carpenter M, West JE, Fanning C. Brca mutations and the risk of angiosarcoma after breast cancer treatment. *Clin Breast Cancer* (2008) 8(6):533–7. doi: 10.3816/CBC.2008.n.066

42. Bosetti C, La Vecchia C, Lipworth L, McLaughlin JK. Occupational exposure to vinyl chloride and cancer risk: A review of the epidemiologic literature. *Eur J Cancer Prev* (2003) 12(5):427–30. doi: 10.1097/00008469-200310000-00012

43. Ron E. Cancer risks from medical radiation. *Health Phys* (2003) 85(1):47–59. doi: 10.1097/00004032-200307000-00011

44. Locker GY, Doroshow JH, Zwelling LA, Chabner BA. The clinical features of hepatic angiosarcoma: A report of four cases and a review of the English literature. *Med (Baltimore)* (1979) 58(1):48–64. doi: 10.1097/00005792-197901000-00003

45. Ben-Izhak O, Kerner H, Brenner B, Lichtig C. Angiosarcoma of the colon developing in a capsule of a foreign body. Report of a case with associated hemorrhagic diathesis. *Am J Clin Pathol* (1992) 97(3):416–20. doi: 10.1093/ajcp/97.3.416

46. Weiss WM, Riles TS, Gouge TH, Mizrachi HH. Angiosarcoma at the site of a dacron vascular prosthesis: A case report and literature review. *J Vasc Surg* (1991) 14(1):87–91. doi: 10.1016/0741-5214(91)90158-q

47. McDonald DJ, Enneking WF, Sundaram M. Metal-associated angiosarcoma of bone: report of two cases and review of the literature. *Clin Orthop Relat Res* (2002) 396):206–14. doi: 10.1097/00003086-200203000-00031

48. Folpe AL, Johnston CA, Weiss SW. Cutaneous angiosarcoma arising in a gouty tophus: report of a unique case and a review of foreign material-associated angiosarcomas. *Am J Dermatopathol* (2000) 22(5):418–21. doi: 10.1097/00000372-200010000-00006

49. Goedert JJ, Cote TR, Virgo P, Scoppa SM, Kingma DW, Gail MH, et al. Spectrum of aids-associated Malignant disorders. *Lancet* (1998) 351(9119):1833–9. doi: 10.1016/s0140-6736(97)09028-4

50. Tonomura N, Elvers I, Thomas R, Megquier K, Turner-Maier J, Howald C, et al. Genome-wide association study identifies shared risk loci common to two Malignancies in golden retrievers. *PloS Genet* (2015) 11(2):e1004922. doi: 10.1371/journal.pgen.1004922

51. Hargis AM, Ihrke PJ, Spangler WL, Stannard AA. A retrospective clinicopathologic study of 212 dogs with Cutaneous hemangiomas and hemangiosarcomas. *Vet Pathol* (1992) 29(4):316–28. doi: 10.1177/030098589202900406

52. Hargis AM, Lee AC, Thomassen RW. Tumor and tumor-like lesions of perilimbal conjunctiva in laboratory dogs. *J Am Vet Med Assoc* (1978) 173(9):1185–90.

53. Wong K, Ludwig L, Krijgsman O, Adams DJ, Wood GA, van der Weyden L. Comparison of the oncogenomic landscape of canine and feline hemangiosarcoma shows novel parallels with human angiosarcoma. *Dis Model Mech* (2021) 14(7):dmm049044. doi: 10.1242/dmm.049044

54. Fletcher CD. *Diagnostic Histopathology of Tumors*. Oxford, UK: Elsevier Health Sciences (2007).

55. Fletcher C, Bridge JA, Hogendoorn PCW, Mertens F. *Who Classification of Tumours of Soft Tissue and Bone: WHO Classification of Tumours* Vol. vol. 5. Geneva, Switzerland: World Health Organization (2013).

56. Ohsawa M, Naka N, Tomita Y, Kawamori D, Kanno H, Aozasa K. Use of immunohistochemical procedures in diagnosing angiosarcoma. Evaluation of 98 Cases. *Cancer* (1995) 75(12):2867–74. doi: 10.1002/1097-0142(19950615)75:12<2867::aid-cncr2820751212>3.0.co;2-8

57. von Beust BR, Suter MM, Summers BA. Factor Viii-related antigen in canine endothelial neoplasms: an immunohistochemical study. *Vet Pathol* (1988) 25(4):251–5. doi: 10.1177/030098588802500401

58. Ferrer L, Fondevila D, Rabanal RM, Vilafranca M. Immunohistochemical detection of Cd31 antigen in normal and neoplastic canine endothelial cells. *J Comp Pathol* (1995) 112(4):319–26. doi: 10.1016/s0021-9975(05)80013-1

59. Rozolen JM, Teodoro TGW, Sobral RA, Sueiro FAR, Laufer-Amorim R, Elias F, et al. Investigation of prognostic value of claudin-5, Psma, and Ki67 expression in canine splenic hemangiosarcoma. *Anim (Basel)* (2021) 11(8):2406. doi: 10.3390/ani11082406

60. Sabattini S, Bettini G. An immunohistochemical analysis of canine haemangioma and haemangiosarcoma. *J Comp Pathol* (2009) 140(2-3):158–68. doi: 10.1016/j.jcpa.2008.10.006

61. Campos AG, Campos JADB, Sanches DS, Dagli MLZ, Matera JM. Immunohistochemical evaluation of vascular endothelial growth factor (Vegf) in splenic hemangiomas and hemangiosarcomas in dogs. *Open J Vet Med* (2012) 2(4):191–5. doi: 10.4236/ojvm.2012.24030

62. Gorden BH, Kim J-H, Sarver AL, Frantz AM, Breen M, Lindblad-Toh K, et al. Identification of three molecular and functional subtypes in canine hemangiosarcoma through gene expression profiling and progenitor cell characterization. *Am J Pathol* (2014) 184(4):985–95. doi: 10.1016/j.ajpath.2013.12.025

63. Lamerato-Kozicki AR, Helm KM, Jubala CM, Cutter GC, Modiano JF. Canine hemangiosarcoma originates from hematopoietic precursors with potential for endothelial differentiation. *Exp Hematol* (2006) 34(7):870–8. doi: 10.1016/j.exphem.2006.04.013

64. Painter CA, Jain E, Tomson BN, Dunphy M, Stoddard RE, Thomas BS, et al. The angiosarcoma project: enabling genomic and clinical discoveries in a rare cancer through patient-partnered research. *Nat Med* (2020) 26(2):181–7. doi: 10.1038/s41591-019-0749-z

65. Murali R, Chandramohan R, Moller I, Scholz SL, Berger M, Huberman K, et al. Targeted massively parallel sequencing of angiosarcomas reveals frequent activation of the mitogen activated protein kinase pathway. *Oncotarget* (2015) 6(34):36041–52. doi: 10.18632/oncotarget.5936

66. Megquier K, Turner-Maier J, Swofford R, Kim JH, Sarver AL, Wang C, et al. Comparative genomics reveals shared mutational landscape in canine hemangiosarcoma and human angiosarcoma. *Mol Cancer Res* (2019) 17(12):2410–21. doi: 10.1158/1541-7786.MCR-19-0221

67. Naka N, Tomita Y, Nakanishi H, Araki N, Hongyo T, Ochi T, et al. Mutations of P53 tumor-suppressor gene in angiosarcoma. *Int J Cancer* (1997) 71(6):952–5. doi: 10.1002/(sici)1097-0215(19970611)71:6<952::aid-ijc7>3.0.co;2-x

68. Wang G, Wu M, Durham AC, Radaelli E, Mason NJ, Xu X, et al. Molecular subtypes in canine hemangiosarcoma reveal similarities with human angiosarcoma. *PloS One* (2020) 15(3):e0229728. doi: 10.1371/journal.pone.0229728

69. Weihrauch M, Bader M, Lehnert G, Koch B, Wittekind C, Wrbitzky R, et al. Mutation analysis of K-Ras-2 in liver angiosarcoma and adjacent nonneoplastic liver tissue from patients occupationally exposed to vinyl chloride. *Environ Mol Mutagen* (2002) 40(1):36–40. doi: 10.1002/em.10084

70. Behjati S, Tarpey PS, Sheldon H, Martincorena I, Van Loo P, Gundem G, et al. Recurrent Ptprb and Plcg1 mutations in angiosarcoma. *Nat Genet* (2014) 46(4):376–9. doi: 10.1038/ng.2921

71. Wang G, Wu M, Maloneyhuss MA, Wojcik J, Durham AC, Mason NJ, et al. Actionable mutations in canine hemangiosarcoma. *PloS One* (2017) 12(11):e0188667. doi: 10.1371/journal.pone.0188667

72. Dim D, Ravi V, Tan J, Hicks D, Wong M. The actin-bundling motility protein fascin and vascular endothelial growth factor (Vegf) are universally over-expressed in human angiosarcoma. *J Clin Oncol* (2007) 25(18_suppl):10068–. doi: 10.1200/jco.2007.25.18_suppl.10068

73. Zietz C, Rossle M, Haas C, Sendelhofert A, Hirschmann A, Sturzl M, et al. Mdm-2 oncoprotein overexpression, P53 gene mutation, and Vegf up-regulation in angiosarcomas. *Am J Pathol* (1998) 153(5):1425–33. doi: 10.1016/S0002-9440(10)65729-X

74. Garcia JM, Gonzalez R, Silva JM, Dominguez G, Vegazo IS, Gamallo C, et al. Mutational status of K-Ras and Tp53 genes in primary sarcomas of the heart. *Br J Cancer* (2000) 82(6):1183–5. doi: 10.1054/bjoc.1999.1060

75. Wagner MJ, Ravi V, Menter DG, Sood AK. Endothelial cell Malignancies: new insights from the laboratory and clinic. *NPJ Precis Oncol* (2017) 1(1):11. doi: 10.1038/s41698-017-0013-2

76. Ueda T, Oji Y, Naka N, Nakano Y, Takahashi E, Koga S, et al. Overexpression of the Wilms' Tumor gene Wt1 in human bone and soft-tissue sarcomas. *Cancer Sci* (2003) 94(3):271–6. doi: 10.1111/j.1349-7006.2003.tb01432.x

77. Johnson KD, Glinskii OV, Mossine VV, Turk JR, Mawhinney TP, Anthony DC, et al. Galectin-3 as a potential therapeutic target in tumors arising from Malignant endothelia. *Neoplasia* (2007) 9(8):662–70. doi: 10.1593/neo.07433

78. Naito S, Shimizu K, Nakashima M, Nakayama T, Ito T, Ito M, et al. Overexpression of Ets-1 transcription factor in angiosarcoma of the skin. *Pathol Res Pract* (2000) 196(2):103–9. doi: 10.1016/S0344-0338(00)80041-2

79. Dictor M, Bendsoe N, Runke S, Witte M. Major basement membrane components in Kaposi's sarcoma, angiosarcoma and benign vascular neogenesis. *J Cutan Pathol* (1995) 22(5):435–41. doi: 10.1111/j.1600-0560.1995.tb00759.x

80. Miettinen M, Sarlomo-Rikala M, Lasota J. Kit expression in angiosarcomas and fetal endothelial cells: lack of mutations of exon 11 and exon 17 of C-Kit. *Mod Pathol* (2000) 13(5):536–41. doi: 10.1038/modpathol.3880093

81. Hornick JL, Fletcher CD. Immunohistochemical staining for Kit (Cd117) in soft tissue sarcomas is very limited in distribution. *Am J Clin Pathol* (2002) 117(2):188–93. doi: 10.1309/LX9U-F7P0-UWDH-8Y6R

82. Komdeur R, Hoekstra HJ, Molenaar WM, Van Den Berg E, Zwart N, Pras E, et al. Clinicopathologic assessment of postradiation sarcomas: kit as a potential treatment target. *Clin Cancer Res* (2003) 9(8):2926–32.

83. Itakura E, Yamamoto H, Oda Y, Tsuneyoshi M. Detection and characterization of vascular endothelial growth factors and their receptors in a series of angiosarcomas. *J Surg Oncol* (2008) 97(1):74–81. doi: 10.1002/jso.20766

84. Agulnik M, Yarber JL, Okuno SH, von Mehren M, Jovanovic BD, Brockstein BE, et al. An open-label, multicenter, phase II study of bevacizumab for the treatment of angiosarcoma and epithelioid hemangioendotheliomas. *Ann Oncol* (2013) 24(1):257–63. doi: 10.1093/annonc/mds237

85. Ray-Coquard IL, Domont J, Tresch-Bruneel E, Bompas E, Cassier PA, Mir O, et al. Paclitaxel given once per week with or without bevacizumab in patients with advanced angiosarcoma: A randomized phase II trial. *J Clin Oncol* (2015) 33(25):2797–802. doi: 10.1200/JCO.2015.60.8505

86. Arbiser JL, Moses MA, Fernandez CA, Ghiso N, Cao Y, Klauber N, et al. Oncogenic H-Ras stimulates tumor angiogenesis by two distinct pathways. *Proc Natl Acad Sci U.S.A.* (1997) 94(3):861–6. doi: 10.1073/pnas.94.3.861

87. LaMontagne KR Jr., Moses MA, Wiederschain D, Mahajan S, Holden J, Ghazizadeh H, et al. Inhibition of map kinase kinase causes morphological reversion and dissociation between soft agar growth and in vivo tumorigenesis in angiosarcoma cells. *Am J Pathol* (2000) 157(6):1937–45. doi: 10.1016/s0002-9440(10)64832-8

88. Dickerson EB, Thomas R, Fosmire SP, Lamerato-Kozicki AR, Bianco SR, Wojcieszyn JW, et al. Mutations of phosphatase and tensin homolog deleted from chromosome 10 in canine hemangiosarcoma. *Vet Pathol* (2005) 42(5):618–32. doi: 10.1354/vp.42-5-618

89. Thomas R, Borst L, Rotroff D, Motsinger-Reif A, Lindblad-Toh K, Modiano JF, et al. Genomic profiling reveals extensive heterogeneity in somatic DNA copy number aberrations of canine hemangiosarcoma. *Chromosome Res* (2014) 22(3):305–19. doi: 10.1007/s10577-014-9406-z

90. Gardner HL, London CA, Portela RA, Nguyen S, Rosenberg MP, Klein MK, et al. Maintenance therapy with toceranib following doxorubicin-based chemotherapy for canine splenic hemangiosarcoma. *BMC Vet Res* (2015) 11(1):131. doi: 10.1186/s12917-015-0446-1

91. Abraham JA, Hornicek FJ, Kaufman AM, Harmon DC, Springfield DS, Raskin KA, et al. Treatment and outcome of 82 patients with angiosarcoma. *Ann Surg Oncol* (2007) 14(6):1953–67. doi: 10.1245/s10434-006-9335-y

92. Fayette J, Martin E, Piperno-Neumann S, Le Cesne A, Robert C, Bonvalot S, et al. Angiosarcomas, a heterogeneous group of sarcomas with specific behavior depending on primary site: A retrospective study of 161 cases. *Ann Oncol* (2007) 18(12):2030–6. doi: 10.1093/annonc/mdm381

93. Pawlik TM, Paulino AF, McGinn CJ, Baker LH, Cohen DS, Morris JS, et al. Cutaneous angiosarcoma of the scalp: A multidisciplinary approach. *Cancer* (2003) 98 (8):1716–26. doi: 10.1002/cncr.11667

94. Heinhuis KM, IJ NS, van der Graaf WTA, Kerst JM, Schrage Y, Beijnen JH, et al. Neoadjuvant systemic treatment of primary angiosarcoma. *Cancers (Basel)* (2020) 12 (8):2251. doi: 10.3390/cancers12082251

95. Young RJ, Natukunda A, Litiere S, Woll PJ, Wardelmann E, van der Graaf WT. First-line anthracycline-based chemotherapy for angiosarcoma and other soft tissue sarcoma subtypes: pooled analysis of eleven European organisation for research and treatment of cancer soft tissue and bone sarcoma group trials. *Eur J Cancer* (2014) 50 (18):3178–86. doi: 10.1016/j.ejca.2014.10.004

96. Spangler WL, Kass PH. Pathologic factors affecting postsplenectomy survival in dogs. *J Vet Intern Med* (1997) 11(3):166–71. doi: 10.1111/j.1939-1676.1997.tb00085.x

97. Yamamoto S, Hoshi K, Hirakawa A, Chimura S, Kobayashi M, Machida N. Epidemiological, clinical and pathological features of primary cardiac hemangiosarcoma in dogs: A review of 51 cases. *J Vet Med Sci* (2013) 75(11):1433–41. doi: 10.1292/jvms.13-0064

98. Weisse C, Soares N, Beal MW, Steffey MA, Drobatz KJ, Henry CJ. Survival times in dogs with right atrial hemangiosarcoma treated by means of surgical resection with or without adjuvant chemotherapy: 23 cases (1986–2000). *J Am Vet Med* (2005) 226 (4):575–9. doi: 10.2460/javma.2005.226.575

99. Hillers KR, Lana SE, Fuller CR, LaRue SM. Effects of palliative radiation therapy on nonsplenic hemangiosarcoma in dogs. *J Am Anim Hosp Assoc* (2007) 43(4):187–92. doi: 10.5326/0430187

100. Koontz BF, Miles EF, Rubio MA, Madden JF, Fisher SR, Scher RL, et al. Preoperative radiotherapy and bevacizumab for angiosarcoma of the head and neck: two case studies. *Head Neck* (2008) 30(2):262–6. doi: 10.1002/hed.20674

101. Verschraegen CF, Fekrazad HM, Rabinowitz I, Quinn R, Snyder D, Judson P, et al. Phase I/II study of docetaxel (D), gemcitabine (G), and bevacizumab (B) in patients (Pts) with advanced or recurrent soft tissue sarcoma (Sts). *J Clin Oncol* (2007) 25(18_suppl):10056–. doi: 10.1200/jco.2007.25.18_suppl.10056

102. Kollar A, Jones RL, Stacchiotti S, Gelderblom H, Guida M, Grignani G, et al. Pazopanib in advanced vascular sarcomas: an Eortc soft tissue and bone sarcoma group (Stbsg) retrospective analysis. *Acta Oncol* (2017) 56(1):88–92. doi: 10.1080/0284186X.2016.1234068

103. Marabelle A, Fakih M, Lopez J, Shah M, Shapira-Frommer R, Nakagawa K, et al. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 keynote-158 study. *Lancet Oncol* (2020) 21 (10):1353–65. doi: 10.1016/S1470-2045(20)30445-9

104. Khanna C, Rosenberg M, Vail D. A review of paclitaxel and novel formulations including those suitable for use in dogs. *J Vet Intern Med* (2015) 29(4):1006–12. doi: 10.1111/jvim.12596

105. Maekawa N, Konnai S, Takagi S, Kagawa Y, Okagawa T, Nishimori A, et al. A canine chimeric monoclonal antibody targeting Pd-L1 and its clinical efficacy in canine oral Malignant melanoma or undifferentiated sarcoma. *Sci Rep* (2017) 7(1):8951. doi: 10.1038/s41598-017-09444-2

106. Igase M, Nemoto Y, Itamoto K, Tani K, Nakaichi M, Sakurai M, et al. A pilot clinical study of the therapeutic antibody against canine Pd-1 for advanced spontaneous cancers in dogs. *Sci Rep* (2020) 10(1):18311. doi: 10.1038/s41598-020-75533-4

107. Kim SE, Liptak JM, Gall TT, Monteith GJ, Woods JP. Epirubicin in the adjuvant treatment of splenic hemangiosarcoma in dogs: 59 cases (1997-2004). *J Am Vet Med Assoc* (2007) 231(10):1550–7. doi: 10.2460/javma.231.10.1550

108. Sorenmo K, Samluk M, Clifford C, Baez J, Barrett JS, Poppenga R, et al. Clinical and pharmacokinetic characteristics of intracavitary administration of pegylated liposomal encapsulated doxorubicin in dogs with splenic hemangiosarcoma. *J Vet Intern Med* (2007) 21(6):1347–54. doi: 10.1892/06-214.1

109. Kahn SA, Mullin CM, de Lorimier LP, Burgess KE, Risbon RE, Fred RM 3rd, et al. Doxorubicin and deracoxib adjuvant therapy for canine splenic hemangiosarcoma: a pilot study. *Can Vet J* (2013) 54(3):237–42.

110. Bray J, Orbell G, Cave N, Munday J. Does thalidomide prolong survival in dogs with splenic haemangiosarcoma? *J Small Anim Pract* (2018) 59(2):85–91. doi: 10.1111/jsap.12796

111. Rowinsky EK. The development and clinical utility of the taxane class of antimicrotubule chemotherapy agents. *Annu Rev Med* (1997) 48(1):353–74. doi: 10.1146/annurev.med.48.1.353

112. Poirier VJ, Hershey AE, Burgess KE, Phillips B, Turek MM, Forrest LJ, et al. Efficacy and toxicity of paclitaxel (Taxol) for the treatment of canine Malignant tumors. *J Vet Intern Med* (2004) 18(2):219–22. doi: 10.1892/0891-6640(2004)18<219:eatopt>2.0.co

113. U'Ren LW, Biller BJ, Elmslie RE, Thamm DH, Dow SW. Evaluation of a novel tumor vaccine in dogs with hemangiosarcoma. *J Vet Intern Med* (2007) 21(1):113–20. doi: 10.1892/0891-6640(2007)21[113:eoantv]2.0.co;2

114. Vail DM, MacEwen EG, Kurzman ID, Dubielzig RR, Helfand SC, Kisseberth WC, et al. Liposome-encapsulated muramyl tripeptide phosphatidylethanolamine adjuvant immunotherapy for splenic hemangiosarcoma in the dog: a randomized multi-institutional clinical trial. *Clin Cancer Res* (1995) 1(10):1165–70.

115. Brown DC, Reetz J. Single agent polysaccharopeptide delays metastases and improves survival in naturally occurring hemangiosarcoma. *Evid Based Complement Alternat Med* (2012) 2012:384301. doi: 10.1155/2012/384301

116. Masuzawa M, Fujimura T, Hamada Y, Fujita Y, Hara H, Nishiyama S, et al. Establishment of a human hemangiosarcoma cell line (Iso-Has). *Int J Cancer* (1999) 81 (2):305–8. doi: 10.1002/(sici)1097-0215(19990412)81:2<305::aid-ijc22>3.0.co;2-z

117. Krump-Konvalinkova V, Bittinger F, Olert J, Brauninger W, Brunner J, Kirkpatrick CJ. Establishment and characterization of an angiosarcoma-derived cell line, as-M. *Endothelium* (2003) 10(6):319–28. doi: 10.1080/10623320390272316

118. Versleijen-Jonkers YMH, Hillebrandt-Roeffen MHS, Weidema ME, Mooren J, von Rhein DT, de Bitter TJJ, et al. Establishment and characterization of the first patient-derived radiation-induced angiosarcoma xenograft model (Rt-As5). *Sci Rep* (2023) 13(1):2653. doi: 10.1038/s41598-023-29569-x

119. Yates N, Hinkel J. The Economics of moonshots: value in rare disease drug development. *Clin Transl Sci* (2022) 15(4):809–12. doi: 10.1111/cts.13270

120. Thamm DH, Dickerson EB, Akhtar N, Lewis R, Auerbach R, Helfand SC, et al. Biological and molecular characterization of a canine hemangiosarcoma-derived cell line. *Res Vet Sci* (2006) 81(1):76–86. doi: 10.1016/j.rvsc.2005.09.005

121. Fosmire SP, Dickerson EB, Scott AM, Bianco SR, Pettengill MJ, Meylemans H, et al. Canine Malignant hemangiosarcoma as a model of primitive angiogenic endothelium. *Lab Invest* (2004) 84(5):562–72. doi: 10.1038/labinvest.3700080

122. Murai A, Asa SA, Kodama A, Hirata A, Yanai T, Sakai H. Constitutive phosphorylation of the Mtorc2/Akt/4e-Bp1 pathway in newly derived canine hemangiosarcoma cell lines. *BMC Vet Res* (2012) 8:128. doi: 10.1186/1746-6148-8-128

123. Urbasic AS, Hynes S, Somrak A, Contakos S, Rahman MM, Liu J, et al. Oncolysis of canine tumor cells by myxoma virus lacking the Serp2 gene. *Am J Vet Res* (2012) 73(8):1252–61. doi: 10.2460/ajvr.73.8.1252

124. Kim JH, Frantz AM, Anderson KL, Graef AJ, Scott MC, Robinson S, et al. Interleukin-8 promotes canine hemangiosarcoma growth by regulating the tumor microenvironment. *Exp Cell Res* (2014) 323(1):155–64. doi: 10.1016/j.yexcr.2014.02.020

125. Tamburini BA, Trapp S, Phang TL, Schappa JT, Hunter LE, Modiano JF. Gene expression profiles of sporadic canine hemangiosarcoma are uniquely associated with breed. *PloS One* (2009) 4(5):e5549. doi: 10.1371/journal.pone.0005549

126. Cook D, Brown D, Alexander R, March R, Morgan P, Satterthwaite G, et al. Lessons learned from the fate of Astrazeneca's drug pipeline: A five-dimensional framework. *Nat Rev Drug Discovery* (2014) 13(6):419–31. doi: 10.1038/nrd4309

127. Blair HA. Onasemnogene Abeparvovec: A review in spinal muscular atrophy. *CNS Drugs* (2022) 36(9):995–1005. doi: 10.1007/s40263-022-00941-1

128. Miller T, Cudkowicz M, Shaw PJ, Andersen PM, Atassi N, Bucelli RC, et al. Phase 1-2 trial of antisense oligonucleotide tofersen for Sod1 Als. *N Engl J Med* (2020) 383(2):109–19. doi: 10.1056/NEJMoa2003715

129. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. Preclinical safety evaluation of biotechnology-derived pharmaceuticals S6(R1). In: *ICH Harmonised Tripartite Guideline*. Geneva, Switzerland: ICH Expert Working Group (2011).

frontiers | Frontiers in Oncology

TYPE Review
PUBLISHED 19 December 2023
DOI 10.3389/fonc.2023.1292046

OPEN ACCESS

EDITED BY
Yusuke Suenaga,
Chiba Cancer Center, Japan

REVIEWED BY
Barani Kumar Rajendran,
Yale University, United States
Carlos Martinez-Perez,
University of Edinburgh, United Kingdom

*CORRESPONDENCE
Laura Senovilla
laurasenovilla@hotmail.com

RECEIVED 10 September 2023
ACCEPTED 23 November 2023
PUBLISHED 19 December 2023

CITATION
Alvarez-Frutos L, Barriuso D, Duran M, Infante M, Kroemer G, Palacios-Ramirez R and Senovilla L (2023) Multiomics insights on the onset, progression, and metastatic evolution of breast cancer.
Front. Oncol. 13:1292046.
doi: 10.3389/fonc.2023.1292046

Multiomics insights on the onset, progression, and metastatic evolution of breast cancer

Lucia Alvarez-Frutos[1], Daniel Barriuso[1], Mercedes Duran[2], Mar Infante[2], Guido Kroemer[3,4,5], Roberto Palacios-Ramirez[1] and Laura Senovilla[1,3,4]*

[1]Laboratory of Cell Stress and Immunosurveillance, Unidad de Excelencia Instituto de Biomedicina y Genética Molecular (IBGM), Universidad de Valladolid – Centro Superior de Investigaciones Cientificas (CSIC), Valladolid, Spain, [2]Laboratory of Molecular Genetics of Hereditary Cancer, Unidad de Excelencia Instituto de Biomedicina y Genética Molecular (IBGM), Universidad de Valladolid – Centro Superior de Investigaciones Cientificas (CSIC), Valladolid, Spain, [3]Centre de Recherche des Cordeliers, Equipe labellisée par la Ligue contre le cancer, Université Paris Cité, Sorbonne Université, Inserm U1138, Institut Universitaire de France, Paris, France, [4]Metabolomics and Cell Biology Platforms, Institut Gustave Roussy, Villejuif, France, [5]Department of Biology, Institut du Cancer Paris CARPEM, Hôpital Européen Georges Pompidou, Paris, France

Breast cancer is the most common malignant neoplasm in women. Despite progress to date, 700,000 women worldwide died of this disease in 2020. Apparently, the prognostic markers currently used in the clinic are not sufficient to determine the most appropriate treatment. For this reason, great efforts have been made in recent years to identify new molecular biomarkers that will allow more precise and personalized therapeutic decisions in both primary and recurrent breast cancers. These molecular biomarkers include genetic and post-transcriptional alterations, changes in protein expression, as well as metabolic, immunological or microbial changes identified by multiple omics technologies (e.g., genomics, epigenomics, transcriptomics, proteomics, glycomics, metabolomics, lipidomics, immunomics and microbiomics). This review summarizes studies based on omics analysis that have identified new biomarkers for diagnosis, patient stratification, differentiation between stages of tumor development (initiation, progression, and metastasis/recurrence), and their relevance for treatment selection. Furthermore, this review highlights the importance of clinical trials based on multiomics studies and the need to advance in this direction in order to establish personalized therapies and prolong disease-free survival of these patients in the future.

KEYWORDS

breast cancer, early-stage, cancer progression, metastasis, omics, biomarkers

1 Introduction

According to GLOBOCAN 2020, female breast cancer (BC) is the most common type of cancer. In 2020, female BC accounted for nearly 700,000 cancer deaths and 2.3 million new cases worldwide (1). Breast cancer is highly heterogeneous, and its treatment classically depends on its clinical status (early stage, locally advanced or metastatic), histological characteristics and biomarker profile (2). Histologically, breast cancers can be classified as sarcomas or carcinomas depending on whether they originate from connective tissue or epithelial cells. In turn, carcinomas are classified as carcinomas *in situ* if they have not invaded other tissues, or as invasive carcinomas if they have invaded adjacent tissues or other organs of the body. Both *in situ* and invasive carcinomas are found in the lobules and ducts (lobular carcinoma *in situ*, LCIS; ductal carcinoma *in situ*, DCIS; invasive lobular carcinoma, ILC; and invasive ductal carcinoma, IDC) (2). The presence or absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) molecularly determines breast cancer subtypes (3). Estrogen receptor positive (ER+) tumors require estrogen to subsist and grow. PR expression, in turn, is estrogen dependent. Therefore, ER+ and/or PR+ tumors are amenable to endocrine therapies targeting estrogen biosynthesis or estrogen receptors (4, 5). Likewise, HER2 amplification or overexpression determines the use of HER2-targeted therapy (6). Breast cancer can be classified into luminal A, luminal B, HER2-enriched and triple-negative breast cancer (TNBC), also known as basal-like. The luminal A subtype is characterized as ER+ and/or PR+ but HER2−, the luminal B subtype is ER+ and/or PR+/HER2+, the HER2-enriched subtype is characterized by overexpression of HER2 and ER−/PR−, and TNBC is negative for ER, PR and HER2 expression (2, 3, 7, 8). In addition, six TNBC subtypes have been identified: basal-like 1 (BL1), basal-like 2 (BL2), mesenchymal (M), mesenchymal stem-like (MSL), immunomodulatory (IM), and luminal androgen receptor (LAR) (9). In principle, the stratification of patients allows the establishment of treatments adapted to each subtype of breast cancer (2, 10) (Figure 1).

Despite efforts, 700,000 women died from breast cancer in 2020 (1). Apparently, prognostic markers currently used in the clinic, such as ER and HER2 status, histologic subtype, size, lymph node status and Nottingham grade, as well as current predictive tests such as germline *BRCA* status, tumor *PIK3CA* mutation status and programmed death-ligand 1 (PDL-1) status, are not sufficient to determine the most appropriate treatment (11). Relapses are the main obstacle faced by clinicians and are mainly due to the development of resistance to the drugs administered. However, there are still insufficient data to determine whether resistance develops after exposure to the drug or whether resistant cells are present from the onset of the disease, preceding antineoplastic treatments (12). This has prompted the search for biomarkers related to different characteristics of breast cancers, such as genetic alterations, epigenetic reprogramming, tumor-promoting inflammation and immune evasion, dysregulation of cell metabolism, or changes in the microbiota, among others (13, 14). Over the past 50 years, efforts have been made to identify genetic and post-transcriptional alterations, changes in protein expression, and more recently, metabolic changes and even immunological or microbial changes. These changes can be detected at the single molecule or pathway level and can serve as markers for diagnosis and/or discovery of personalized therapies. The detection of these modifications is made possible by technologies known as "omics", such as genomics, transcriptomics, proteomics or metabolomics, or more specific ones such as epigenomics, glycomics or lipidomics, among others. The identification of new biomarkers might allow

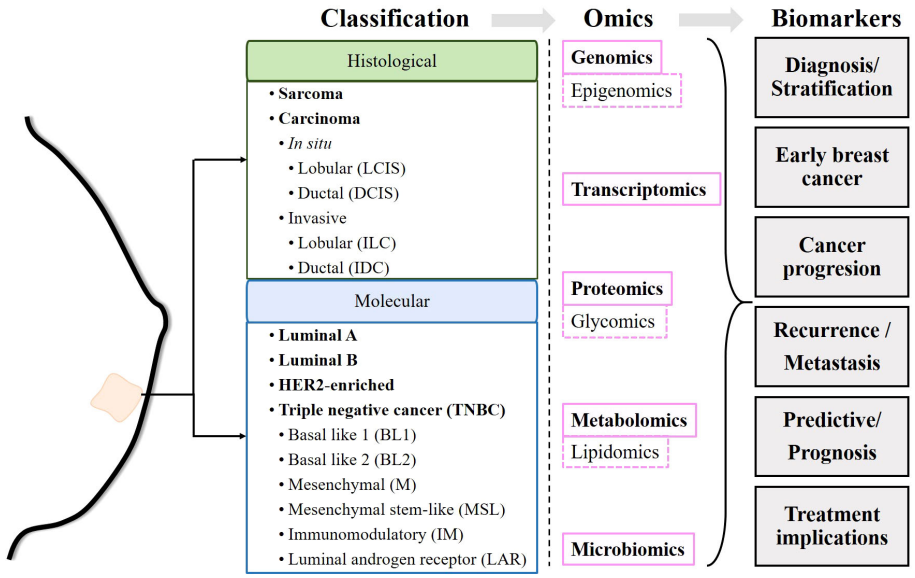


FIGURE 1
Schematic of the workflow performed in the review. Once the commonly accepted classifications of breast cancer have been described, the biomarkers identified by different omics techniques are described.

more precise, and personalized therapeutic decisions in both primary and recurrent cancers (Figure 1).

Genomics collectively characterizes and quantifies all the genes of an organism. Genomic analysis includes techniques such as microarrays, gene expression profiling, serial analysis of gene expression (SAGE), comparative genomic hybridization (CGH), array-based CGH (array-CGH), whole genome amplification (WGA), and high-throughput sequencing or next-generation sequencing (NGS). Different genomic studies have focused on the search for markers with predictive or prognostic value, for patient stratification and for determining the appropriate therapy (15, 16). Transcriptomics studies the set of all RNA transcripts of an organism. The most commonly used techniques to study the transcriptome are microarrays and RNA sequencing (RNA-seq) using high-throughput sequencing or NGS. Transcriptomic analysis allows the measurement of differential gene expression, often presented as hierarchical clustering (17, 18). Proteomics detects and quantifies the presence of proteins produced or modified by an organism. Proteomic analysis is performed by using separation techniques (gas chromatography, liquid chromatography, ultra-high performance liquid chromatography, capillary electrophoresis) combined with mass spectrometry (MS, resulting in GC-MS, LC-MS, UPLC-MS, CE-MS, respectively), nuclear magnetic resonance spectroscopy (NMR), reverse phase protein arrays (RPPA), and sequential windowed acquisition of all theoretical fragment ion mass spectra (SWATH-MS) (19). Glycomics provides insight into the biological significance of N-glycosylation of plasma proteins in cancer (20). Metabolomics approaches based on NMR, LC-MS, GC-MS and desorption electrospray ionization mass spectrometry imaging (DESI-MSI) enable the linking of genotype and phenotype thanks to the knowledge generated by dynamic metabolism (21, 22). Epigenomics (study of gene modifications through the aggregation of chemical compounds, with no changes in the DNA sequence), lipidomics (study of lipids in the biological system), and microbiomics (study of microorganisms present in the human body) are branches that use some of the mentioned above.

Due to the complexity that governs carcinogenesis and tumor progression, recent years have seen efforts to integrate data from different omics into a computational approach that allows for more complex reconstruction of biochemical connections (23). This review aims to summarize the existing knowledge on the advances made thanks to omics studies in human breast cancer. In particular, we focus on 1) the discovery of markers that can be used for diagnosis, 2) molecular and/or signaling pathway alterations in onset, progression, and metastasis, and 3) resistance to therapies and attempts to establish personalized treatments (Figure 1).

2 Implication of omics in the identification of cancer-specific and prognostic biomarkers

A cancer biomarker, found in tissues or body fluids, is used to detect the presence of cancer (differences between tumor and healthy samples or between tumor subtypes) and provides information on prognosis/prediction, cancer progression, and cancer medicine/treatment guidance, among others (24). In this review, we provide an overview of cancer-specific molecules and/or pathways that have been identified as biomarkers in breast cancer using different omics technologies.

2.1 Diagnostic biomarkers

One of the first uses of omics has been to identify biomarkers that differentiate tumor tissue from healthy tissue, as well as to identify biomarkers specific to different subtypes of breast cancer. In addition, the last decade has seen efforts to develop minimally invasive techniques to predict breast cancer subtypes, such as identifying biomarkers in human plasma, saliva, and urine.

The presence, copy number alterations, mutations, or amplifications of various genes are markers of breast cancer in tissue. Table 1 summarizes the most commonly mutated genes (27, 44). Copy number alterations have been observed in *PIK3CA*, *ERBB2*, *TP53*, *MAP2K4*, *MLL3*, *CDKN2A*, *PTEN*, and *RB1* (44). Approximately 35-40% of primary breast cancers harbor mutations in *TP53* and *PIK3CA*, as well as amplification of *ERBB2*, *FGFR1*, and *CCND* (~15%), and mutations in *MUC16*, *AHNAK2*, *SYNE1*, *KMT2C*, *AKT1*, and *GATA3* genes (10%) (25, 45). In 2012, the Cancer Genome Atlas Network (CGAN) identified novel significantly mutated genes (SMGs), including *TBX3*, *RUNX1*, *CBFB*, *AFF2*, *PIK3R1*, *PTPN22*, *PTPRD*, *NF1*, *SF3B1*, and *CCND3* (26).

A proteomic analysis has identified canonical up- and downregulated pathways in breast cancer. Pathways upregulated in breast cancer include glycolysis; metabolic pathways such as pyruvate fatty acid, arginine and proline, and valine, leucine and isoleucine catabolism; protein ubiquitination; RhoA, PI3K/AKT, ILK, 14-3-3-mediated, RAN, aryl hydrocarbon receptor, integrin, clathrin-mediated endocytosis, IGF-1, VEGF, EIF2, actin cytoskeleton, ERK5, GABA receptor, and HER-2 signaling pathways; NRF2-mediated response to oxidative stress; Rho-mediated regulation of actin-based motility; and LPS/IL-1-mediated inhibition of RXR function. Downregulated pathways include the citrate cycle, acute phase response signaling, P53 signaling, primary immunodeficiency signaling, urea cycle and amino group metabolism, Cdc42 signaling, glyoxylate and dicarboxylate metabolism, and autoimmune thyroid disease signaling (46).

A fucosylated triantennary glycan containing three α2-3 sialic acids (also called H6N5F1L3) and a non-fucosylated triantennary glycan containing a combination of α2-3 and α2-6 sialic acids (H6N5L2E1) are found at lower levels in breast cancer patient samples (28), although other studies have reported conflicting data for H6N5F1L3 (47). In addition, a fucosylated tetraantennary glycan containing a combination of α2-3 and α2-6 sialic acids (H7N6F1L1E3) is significantly elevated in breast cancer patients (28). Similarly, elevated levels of other trisialylated triantennary fucosylated glycans (termed H6N5F1S3, consisting of H6N5F1E3, H6N5F1L3, H6N5F1L2E1, H6N5F1L2E1, H6N5F1L1E2, and

TABLE 1 Diagnostic molecular biomarkers of breast cancer detected by different omics.

Omic	Molecular Biomarkers	Sample	REF
Genomic	*PIK3CA, TP53, GATA3, PTEN, AKT1, CDH1, ARID1B, CASP8, BRCA1/2, RB1, MLL3, MAP3K1, MAP3K13, NCOR1, SMARCD1, CDKN1B, TBX3, RUNX1, CBFB, AFF2, PIK3R1, PTPN22, PTPRD, NF1, SF3B1, CCND1, CCND3, MAP2K4, CDKN2A, BARD1, CHECK2, CTLA4, CYP19A1, FGFR1/2, H19, LSP1, MUC16, AHNAK2, SYNE1, CDKN1B, RUNX1, CBFB, AFF2, PTPN22, PTPRD, CCND3, MRE11A, RAD51C, STK11, TERT, TOX3, XRCC2, XRCC3, ERBB2, ARID5B, CTCF, HDAC9, KDM5B, NCOR2, SETD1A, SXL2, ARID1A, CTNND1, NUP107, CHD8, FANC1, CHD9, KEAP1, PCDH18, LAMA2, HDAC9, ARFGEF1, MILT4, FOXO3, GPS2.*	Tissue	(20, 21, 25, 26, 27)
Glycomic	H6N5F1L3, H6N5L2E1, H7N6F1L1E3, H6N5F1S3, Vimentin, keratin 7, enolase 2, pyruvate kinase isozyme M 2 (PKM2), protein disulfide isomerase (PDI) A6, TP and voltage-dependent anion selective channel protein in breast cancer	Tissue	(28, 29, 30, 31)
Metabolomic	N1-Ac-SPD, N8-Ac-SPD	Saliva	(32)
Metabolomic	Acetone, 3-hexanone, 4-heptanone, 2-methyl-5-(methylthio)-furan and acetate	Urine tissue	(33)
Metabolomic Proteomic	taurocholate, taurochenodeoxycholate, glycocholate, allantoin, taurodesoxycholate, glycodesoxycholate, ursodeoxycholate, mannose and fructose, tyramine O-sulfate, N-formylphenylalanine, dopamine 4-sulfate, glycerol, glycerol 3-phosphate, GOT1, LDHB, GSS, GPX3	Plasma	(34)
Metabolomic	Lactate, lysine, alanine, pyruvate and glucose	Serum	(35)
Lipidomic	S1P, ceramides, sphingomyelin phosphodiesterase, palmitate-containing phosphatidylcholines	Tissue	(36, 37, 38)
Lipidomic	Lysophosphatidylcholine, lysoPC a C16:0, PC ae C42:5 and PC aa C34:2	Plasma	(39)
Microbiomic	*Thermus scotoductus, E. coli, Bacillus cereus, Shewanella, Corynbacterium, Bacillus, Staphylococcus, Enterobacteriaceae, Comamondaceae, Bacteroidetes, Mycobacterium fortuitum, Mycobacterium phlei, Fusobacterium, Atopobium, Gluconacetobacter, Hydrogenophaga* and *Lactobacillus*	Tissue	(40, 41, 42, 43)

H6N5F1E3) have been observed (28, 30). Glycomic analysis have identified seven glycosylated proteins with *O*-linked β-D-N-acetylglucosamine (*O*-GlcNAc), an important post-translational modification involving reversible and highly dynamic covalent binding of β-N-GlcNAc to Ser/Thr residues in proteins (48), such as vimentin, keratin 7, enolase 2, pyruvate kinase isozyme M 2 (PKM2), protein disulfide isomerase (PDI) A6, TP, and voltage-dependent anion selective channel protein in breast cancer (31) (Table 1).

Eight pathways have been implicated in breast cancer: protein digestion and absorption, central carbon metabolism in cancer, neuroactive ligand receptor interaction, ABC transporters, mineral uptake, inositol phosphate metabolism, glutathione metabolism, and cysteine and methionine metabolism (49). A combinatorial study of metabolomic and proteomic profiles in human plasma samples has identified metabolic signatures for breast cancer diagnosis. The most abundant metabolites in breast cancer patients include mainly primary and secondary compounds of bile acid metabolism, compounds of fructose, mannose, galactose, tyrosine and glycerolipid metabolism. Critical metabolic pathways in breast cancer include alanine, aspartate and glutamate pathways, glutamine and glutamate metabolic pathways, and arginine biosynthesis pathways. The specific metabolites are listed in Table 1. In addition, thanks to the combined study with the proteome, aspartate aminotransferase (GOT1), L-lactate dehydrogenase B-chain (LDHB), glutathione synthetase (GSS) and glutathione peroxidase 3 (GPX3) have been found to be closely involved in these metabolic pathways (34).

On the other hand, metabolomic analysis of polyamines in saliva before and after surgical treatment has shown a decrease in N1-Ac-SPD and an increase in N8-Ac-SPD in patients after surgical treatment. Thus, the ratio (N8-Ac-SPD)/(N1-Ac-SPD+N8-Ac-SPD) could be an index of health status after surgical treatment (32) (Table 1). A metabolomic study of urine and breast tissue samples has identified dysregulation of lactate, valine, aspartate and glutamine pathways in breast cancer. In addition, five metabolites (acetone, 3-hexanone, 4-heptanone, 2-methyl-5-(methylthio)-furan, and acetate) allow correlation between urine and tissue samples (33). Inositol triphosphate receptor (IP3R) type 2 and 3 expression is increased in breast tumor tissue compared to adjacent healthy tissue. Increased lipoproteins, increased levels of metabolites such as lactate, lysine and alanine, and decreased levels of pyruvate and glucose in the serum of patients with high IP3R expression compared to healthy individuals (35).

The combination of metabolic profiling with tissue protein expression increases the accuracy in characterizing breast cancer patients (35). Lipidomic analyses have shown that the levels of sphingosine-1-phosphate (S1P), ceramides, and other sphingolipids are significantly higher in breast tumors than in normal breast tissue (36). In addition, sphingomyelin phosphodiesterase (SMase), which converts sphingomyelin to ceramide phosphate, is downregulated in 60% of breast tumors (37). An increase in fatty acids such as palmitate-containing phosphatidylcholines (PC) has also been found, especially in ER− and grade 3 tumors compared to healthy breast tissue. Phospholipids may have diagnostic potential as they have been associated with cancer progression and patient survival (38). When plasma samples from breast cancer patients and healthy controls were compared, significantly lower levels of lysophosphatidylcholines (LysoPC) and higher levels of sphingomyelins have been observed in plasma samples from

cancer patients. In addition, three metabolites (LysoPC a C16:0, PC ae C42:5 and PC aa C34:2) differentiate breast cancer patients from healthy controls (39) (Table 1).

Lastly, the bacteria found on the skin have direct access to the breast ducts through the nipple, so the breast tissue contains a wide variety of bacteria, such as *Staphylococcus epidermidis* and *Micrococcus luteus* (50, 51). Different species of bacteria perform different functions. For example, *Lactobacillus* triggers protective mechanisms that include immune activation, competitive inhibition of pathogenic strains, and synthesis of signaling intermediates. In contrast, *E. coli* and Staphylococcus induce DNA damage (40). Changes in the microbiota of breast and intestinal tissues have been associated with the development of breast cancer (50). Several studies even describe correlations between dysbiosis of the tissue microbiome and the development of breast cancer. When characterizing the microbiome of tumor tissue and adjacent non-tumor tissue from different patients, a higher abundance of taxa belonging to the phylum Actinobacteria was observed in the non-tumor samples. In contrast, Firmicutes and Alpha-Proteobacteria were significantly overrepresented in tumor tissue (52). Healthy individuals show a significantly higher abundance of *Lactobacillus, Thermoanaerobacterium thermosaccharolyticum*, *Candidatus Aquiluna* sp., *IMCC13023*, *Anoxybacillus, Leuconostoc*, *Lactococcus, Geobacillus, Methylobacterium, Turicella otitidis* (40), *Prevotella*, *Lactococcus*, *Streptococcus*, *Corynebacterium* and *Micrococcus* (41). In cancer patients, there is an abundance of *Thermus scotoductus*, *E. coli*, *Bacillus cereus*, *Shewanella*, *Corynbacterium* (40), *Bacillus*, *Staphylococcus*, Enterobacteriaceae, Comamondaceae, Bacteroidetes (41), *Mycobacterium fortuitum*, and *Mycobacterium phlei* (42). The breast tissue microbiome of women with enrichment in lower abundance taxa, including the genera *Fusobacterium*, *Atopobium*, *Gluconacetobacter*, *Hydrogenophaga* and *Lactobacillus*, compared to that of women with benign breast disease (43) (Table 1).

2.2 Stratification biomarkers

Genomic alterations have been found in the different molecular subtypes of primary breast cancer. In general, approximately 5-10% of breast cancers are hereditary, mostly due to pathogenic variants or mutations in the *BRCA1* and *BRCA2* genes (53). Mutations in *BRCA1* are associated with ER– and PR– tumors (54, 55), while mutations in *BRCA2* are associated with ER+ and PR+ tumors (56). SMGs are more diverse and recurrent in both luminal subtypes, particularly in luminal A (26), and the heat shock protein (HSP) family has been specifically associated with different cancer types (18). In the luminal subtype A, the most common SMG is *PIK3CA* (45%), followed by *MAP3K1*, *GATA3*, *TP53*, *CDH1*, and *MAP2K4* (26). Copy number gains of *CCND1*, *FGF3*, and *FGFR1* have also been identified (25). The heat shock protein (HSP) genes *DNAJB4*, *DNAJC18*, *HSPA12A*, *HSPA12B*, *HSPB2*, *HSPB6*, *HSPB7*, *CRYAB*, and *SACS* are downregulated, whereas *DNAJC5B*, *DNAJB13*, *DNAJC1*, *DNAJC22*, *HSPB1*, *HSPA6*, and *DNAJC12* are upregulated (18). The luminal subtype B is characterized by *TP53* and *PIK3CA* SMGs (29% each) (26), as well as increased copy number of *CCND1*, *FGF3* and *FGFR1* (25); downregulation of HSP genes such as *DNAJB4*, *DNAJC18*, *HSPA12A*, *HSPA12B*, *HSPB2*, *HSPB6*, *HSPB7*, *CRYAB* and *SACS* and upregulation of *DNAJC5B*, *DNAJB13*, *DNAJC1*, *DNAJC22*, *HSPB1*, *HSPA6*, *CCT5*, *CCT3*, *HSPE1*, *DNAJC9*, *HSPD1*, *DNAJC12*, *DNAJA4*, *HSPH1*, *CCT2*, and *DNAJA3* (18). The HER2-enriched subtype presents with HER2/*ERBB2* amplification (80%), high frequency of mutations in *TP53* (72%) and *PIK3CA* (39%) (26). *PTEN* and *INPP4B* have also been identified as genes of interest. Deletions in *PPP2R2A*, *MTAP*, and *MAP2K4* genes have been reported (45). Among the HSP genes, *CRYAB*, *SACS*, *DNAJB4*, *DNAJC18*, *HSPA12A*, *HSPA12B*, *HSPB2*, *HSPB6*, *HSPB7*, *HSPB8*, *DNAJC5G* and *BBS12* are downregulated, while the upregulated genes are *DNAJC5B*, *DNAJB13*, *DNAJC1*, *DNAJC22*, *HSPB1*, *CCT5*, *CCT3*, *HSPE1*, *DNAJC9*, *HSPD1*, *DNAJA4*, *HSPH1*, *HYOU1*, *DNAJB11*, *CCT6A* and *DNAJB3* (18). Finally, mutations in *TP53* are observed in 80% of TNBC cases, followed by alterations in *PIK3R1* and *NF1* (26). *INPP4B* is another gene of interest in TNBC, and deletions in *PPP2R2A*, *MTAP* and *MAP2K4* genes have also been reported (45). Among the downregulated HSP genes in TNBC, we found *DNAJB4*, *DNAJC18*, *HSPA12A*, *HSPA12B*, *HSPB2*, *HSPB6*, *HSPB7*, *HSPB8*, *DNAJC12* and *DNAJC27*. As for the upregulated genes, *DNAJC5B*, *HSPA6*, *CCT5*, *CCT3*, *HSPE1*, *DNAJC9*, *HSPD1*, *HYOU1*, *DNAJB11*, *CCT6A*, *HSPA5*, *HSPA14*, *CRYAA*, *DNAJC2* and *DNAJC6* were found (18) (Table 2).

MicroRNAs (miRNAs) are small non-coding RNAs that modulate gene expression to regulate various cellular processes, including those involved in breast cancer (66, 67). miR-206 is highly expressed in ER– tumors and also targets the ERα receptor, as do miR-221 and miR-222 (57). In addition, DCIS and LCIS are characterized by the upregulated expression of several miRNAs, which are listed in Table 2 (57).

Through an integrated study combining genomics and epigenomics, pathways unique to TNBC and non-TNBC were identified. The most significant pathways for TNBC are retinal biosynthesis, BAG2, LXR/RXR, EIF2, and P2Y purinergic receptor signaling pathways, whereas in non-TNBC they are UVB-induced MAPK, PCP, endothelial apelin, endoplasmic reticulum stress, and host viral egress mechanisms (68). Based on the genomic and transcriptomic landscape, Xie et al. have established a new classification of immune subtypes of ER+/PR–/HER2– breast cancer, termed clusters 1 to 5. Cluster 1 is characterized by an activated but suppressive immune microenvironment, immune infiltration, increased stromal content, and an elevated TGF-β response signature. Cluster 2 has an inactivated immune phenotype. Cluster 3 has an activated immune phenotype enriched in innate, adaptive, and immunosuppressive cells, as well as interferon (IFN)-γ response, inflammation, macrophage upregulation, and cytolytic signatures. Cluster 4 is characterized by an immunologically inactive phenotype and low infiltration of the microenvironment. Finally, cluster 5 lacks immunologic properties but presents a phenotype associated with hormonal responses (69) (Table 2).

Several studies have investigated the classification of breast cancers using proteomic technologies. By analyzing differential protein expression in tissues, they identified the expression of proteins that characterize luminal subtypes, HER2-enriched

TABLE 2 Stratification biomarkers of breast cancer detected by different omics.

Omic	Biomarkers	Sample	Subtype of BC	REF
Genomic	*BRCA1*, miR-206	Tissue	ER–	(54, 57)
Genomic	*BRCA1*	Tissue	PR-	(47)
Genomic	*BRCA2*	Tissue	ER+, PR+	(56)
Genomic	*BRCA1, INPP4B, PPP2R2A, MTAP, MAP2K4 TP53, PIK3R1, NF1, DNAJC12, DNAJC27, DNAJC2, DNAJC6, DNAJC9, DNAJB11, CCT5, CCT3, CCT6A, HSPE1, HSPD1, HYOU1, HSPB8, HSPA6, HSPA5, HSPA14, CRYAA*		TNBC	(18, 25, 26, 45, 54, 55)
Genomic	*PTEN, INPP4B, PPP2R2A, MTAP, MAP2K4*, HER2, *ERBB2, TP53, PIK3CA, CRYAB, SACS, HSPB1, HSPB8, DNAJC5G, DNAJB13, DNAJC1, DNAJC22, DNAJA4, DNAJC9, DNAJB11, DNAJB3, BBS12, CCT5, CCT3, HSPE1, HSPD1, HSPH1, HYOU1, CCT6A*	Tissue	HER2-enriched	(13, 21, 23)
Genomic	*PIK3CA, MAP3K1, GATA3, TP53, CDH1, MAP2K4, CCND1, FGF3, FGFR1, CRYAB, SACS, DNAJC12, DNAJB13, DNAJC1, DNAJC22, HSPB1, HSPA6*	Tissue	Luminal A	(18, 25, 26)
Genomic	*TP53, PIK3CA, CCND1, FGF3, FGFR1, CRYAB, SACS, HSPB1, HSPA6, CCT5, CCT3, HSPE1, DNAJC9, HSPD1, DNAJC12, DNAJA4, DNAJB13*, DNAJC1, *DNAJA3, DNAJC22, HSPH1, CCT2*	Tissue	Luminal B	(18, 25, 26)
Transcriptomic	miR-21, miR-200c, miR-361-5p, miR-374a, miR-93, miR-182, miR-183, miR-210, miR-221, miR-7b, miR-125b, miR-127-3p, and miR-320	Tissue	DCIS	(57)
Transcriptomic	miR-375, miR-182, miR-183, miR-96, miR-203, miR-425-5p, miR-565	Tissue	LCIS	(57)
Proteomic	PPIaseB, Rho-GDI α, TPM4, Thymosin α1, PGRMC1, Liprin-α1, β-arrestin-1, fascin, DAP5, superoxide dismutase, Ral A binding protein, Galectin-1, uridine phosphorylase 2, cellular retinoic acid-binding protein 1, protein S100-A11, nucleoside diphosphate kinase A, α1-antitrypsin	Tissue	HR+	(58)
Proteomic	HSP90α, laminin, GSTP1, FASN, HSP27, PGK1, GLO, CK19, HNRNPH1, BiP, RKIP CK7, GAPDH, PGK1, FUT8, HEXA, HEXB, MAN2B2, MAN1B1, MAN2A1, GALNT 2,3,6, ADH, ALDH, ACAD, PYCR1,2, PYCRL, PRODH, HNMT, KMO	Tissue	HER2-enriched	(58, 59)
Proteomic	STAT1, PTEN, pMAPK, P38, P27, P21, MASPIN, CD10, FAK, EGFR, Caveolin, CD74, CK14, RCL1, MCM complex proteins, DNA polymerases, DNA damage response proteins, CDK1, CDK2, CDK6, PCNA, PTEN	Tissue	TNBC	(58)
Proteomic	FBP2, FBP1, NDUF, UQCR, SDH, COX subunits, ATP5, ATP6 subunits, CA1, CA2	Tissue	Luminal	(59)
Proteomic	EZH2	Tissue	DCIS	(15)
Genomic Transcriptomic Proteomic	CDH1, TGFBR2, IL11RA, TNFRSF17, CCL15, CCL14, CCR2, CD27, XCL2, IFNAR2 CD40LG, PDCD1 (PD-1), CD274 (PD-L1), CTLA4	Tissue	ILC	(60)
Metabolomic	β-alanine, xanthine, isoleucine, glutamate, taurine	Tissue	ER+	(61)
Metabolomic	Glycochenodeoxycholic acid, alanine, LysoPC (16:1), valine, 2-octenedioic acid	Plasma	ER+	(60)
Metabolomic	Carnitine, LysoPC (20:4), proline, valine, 2-octenedioic acid	Plasma	HER2+	(60)
Metabolomic	L-Tryptophan, LysoPC(14:0), Glycoursodeoxycholic acid, Lysophosphoethanolamines (LysoPE)(18:2)	Plasma	Luminal A	(62)
Metabolomic	LysoPE(18:2), LysoPE(18:1(11Z/9Z)), LysoPC(20: 3), Biliverdin, LysoPE(16:0)	Plasma	Luminal B	(62)
Metabolomic	LysoPE(18:1(11Z)/9Z), LysoPC(0:0/16:0), Biliverdin, L-Tryptophan, LysoPE(18:2)	Plasma	HER2-enriched	(62)
Metabolomic	LysoPE(18:1(11Z)/9Z), LysoPC(0:0/16:0), Biliverdin, L-Tryptophan, LysoPE(18:2) in HER2+; and L-Tryptophan, LysoPC(16:0/0:0), LysoPE(18:1(11Z)/9Z)	Plasma	TNBC	(62)
Lipidomic	taurine (m/z 124.0068), uric acid (m/z 167.0210), ascorbic acid (m/z 175.0241) and glutathione (m/z 306.0765)	Tissue	IBC	(61)
Lipidomic	fatty acids (341.2100 and 382.3736 m/z) and glycerophospholipids (PE (P-16:0/22:6, m/z 746.5099, and PS (38:3), m/z 812.5440)	Tissue	DCIS	(61)
Lipidomic	glycerol-3-phosphate acyltransferase	Tissue	HR-associated	(63)
Lipidomic	GM2	Plasma	ER-	(64)

(Continued)

TABLE 2 Continued

Omic	Biomarkers	Sample	Subtype of BC	REF
Microbiomic	*Bordetella*, *Campylobacter*, *Chlamydia*, *Chlamydophila*, *Legionella*, and *Pasteurella*	Tissue	Luminal B	(65)
Microbiomic	*Arcanobacterium, Bifidobacterium, Cardiobacterium, Citrobacter, and Escherichia*	Tissue	Luminal A	(65)
Microbiomic	*Streptococcus*	Tissue	HER2-enriched	(65)
Microbiomic	*Aerococcus*, *Arcobacter*, *Geobacillus*, *Orientia*, and *Rothia*	Tissue	TNBC	(65)

DCIS, ductal carcinoma *in situ*; ER, estrogen receptor; HR, hormone receptor; IBC, invasive breast cancer; ILC, invasive lobular cancer; LCIS, lobular carcinoma *in situ*; PR, progesterone receptor; TNBC, triple negative breast cancer.

breast cancers, and TNBC. In addition, proteomic profiling of these three breast cancer types revealed functional differences. Cytoskeletal remodeling as well as alterations in the cell adhesion process are found in all BC types. Luminal tumors are characterized by increased "energy metabolism" as indicated by an elevated expression of key proteins in gluconeogenesis, electron transport chain and ATP synthase complex. "Immune response" is altered in both luminal A and luminal B subtypes, while "cell cycle regulation" is important in luminal B tumors. On the other hand, luminal tumors show decreased expression of proteins related to metabolic pathways, including glycolysis, serine synthesis, and glutamine consumption. The HER2-enriched subtype is characterized by decreased "amino acid and energy metabolism", reduced "cellular community" and increased "glycan biosynthesis and metabolism". Finally, the TNBC subtype is characterized by increased "replication and repair", "cell growth and death" and "translation" pathways. Other relevant processes are "immune response" and "blood coagulation". The most relevant proteins involved in these pathways are listed in Table 2 (58, 59, 70). Besides that, the expression of the protein enhancer of zeste homolog 2 (EZH2) is elevated in premalignant atypical ductal hyperplasia (ADH) and even higher in DCIS compared to normal epithelium (15).

Luminal tumors have been classified into 3 proteomic clusters. Luminal cluster-1 is enriched for RNA processing and splicing processes but depleted for immune-related proteins including the ones involved in antigen processing and presentation, and type I and type II IFN signaling. Luminal cluster-2 is enriched for stromal proteins and extracellular matrix (ECM) components. Luminal cluster-3 has high expression of proteins for DNA replication, cell cycle, response to DNA damage, and immune response, while depleted for ECM components, blood coagulation, epithelial cell differentiation, and response to estrogen and steroid hormones compared to luminal clusters-1 and -2. In addition, significantly higher expression of Ki67 is found in luminal cluster-3 compared to luminal cluster-1 and -2 (71). On the other hand, 4 TNBC subgroups have been identified according to 4 proteomic clusters. TNBC cluster-1 has the most favorable survival and is characterized by immune response, antigen processing and presentation, and IFN type I and II signaling processes. TNBC cluster-2 has intermediate survival and is enriched for ECM components, coagulation, and humoral immune response processes. TNBC cluster-3 has intermediate survival and is enriched for lipid metabolism, catabolism, and oxidation-reduction processes. TNBC cluster-4 exhibits the poorest survival and is enriched for DNA replication and cell cycle proteins (71).

A comprehensive genomic, transcriptomic and proteomic analysis of patients with ILC has identified mutations in cadherin-1 (CDH1) and the phosphatidylinositol 3- kinase (PI3K) pathway as the most common molecular alterations in ILC. In addition, two major subtypes of ILC have been identified: an immune-related (IR) subtype and a hormone-related (HR) subtype: The IR subtype is characterized by upregulation of PD-L1 mRNA, PD-1 and CTLA-4, increased sensitivity to DNA damaging agents, and upregulation of lymphoid signaling molecules at the mRNA level (TGFBR2, IL11RA, TNFRSF17, CCL15, CCL14, CCR2, CD27, XCL2, IFNAR2, and CD40LG). In addition, the IR subtype shows upregulated genes in the cytokine-cytokine receptor interaction pathway, suggesting alterations in the composition or functional activity of immune cells within these tumors. Interestingly, the negative regulators of the immune response PDCD1 (PD-1), CD274 (PD-L1) and CTLA4 are expressed at higher mRNA levels in the IR subtype. The related HR subtype is associated with epithelial-to-mesenchymal transition (EMT). Moreover, the HR subtype shows higher levels of estrogen receptors (ESR1) and progesterone receptors (PGR) and upregulation of cell cycle genes and estrogen receptor (ER) target genes (60) (Table 2).

Differential metabolites have been identified when comparing HER2+ with HER2− patients, as well as when comparing ER+ with ER− patients. Plasma samples from HER2+ patients are characterized by increased aerobic glycolysis, gluconeogenesis and fatty acid biosynthesis and decreased Krebs cycle. Specifically, HER2+ is characterized by overexpression of carnitine, LysoPC (20:4), proline, valine, and 2-octenedioic acid. Strong metabolic differences correlate with hormone receptor status. Plasma samples from ER+ patients reflect increased alanine, aspartate, and glutamate metabolism, decreased glycerolipid catabolism and increased purine metabolism. ER+ is characterized by overexpression of glycochenodeoxycholic acid and decreased expression of alanine, LysoPC (16:1), valine, and 2-octenedioic acid. Many glycolytic and glycogenolytic intermediates, components of the glutathione (GSH) pathway, the oncometabolite 2-hydroxyglutarate (2-HG), and the immunomodulatory tryptophan metabolite kynurenine are elevated in ER− compared to ER+ cancers (72, 73). Using deep learning techniques, metabolites, and pathways have been identified

that can discriminate between ER+ and ER− patient samples. Among the identified metabolites, five have been proposed as breast cancer biomarkers: β-alanine, xanthine, isoleucine, glutamate, and taurine (49). Studies performed on plasma samples from breast cancer patients have allowed the identification of specific metabolomic profiles for each cellular subtype: L-tryptophan, LysoPC(14:0), glycoursodeoxycholic acid, lysophosphoethanolamine (LysoPE)(18:2) in luminal A; LysoPE (18:2), LysoPE(18:1(11Z/9Z)), LysoPC(20: 3), biliverdin, LysoPE (16:0) in luminal B; LysoPE(18:1(11Z)/9Z), LysoPC(0:0/16:0), biliverdin, L-tryptophan, LysoPE(18:2) in HER2+; and L-tryptophan, LysoPC(16:0/0:0), LysoPE(18:1(11Z)/9Z) in TNBC (62) Table 2. When plasma samples from patients with mutated *BRCA1* and non-mutated *BRCA1* were compared, the levels of adenine, N6-methyladenosine, and 1-methylguanine were found to be significantly lower in patients with *BRCA1* mutations (74).

The lipidomic profiles of invasive breast cancer (IBC), DCIS and benign surrounding tissue (BAT) have been investigated. IBC is characterized by the presence of polyunsaturated fatty acids, deprotonated glycerophospholipids and sphingolipids. IBC can be distinguished from BAT by the presence of highly saturated lipids and antioxidant molecules. DCIS differs from IBC by lipids associated with cell signaling and apoptosis (61). Lipidomics have allowed the identification of glycerol-3-phosphate acyltransferase (GPAM), an enzyme involved in triacylglycerol and phospholipid biosynthesis, which shows increased cytoplasmic expression in HR-associated breast cancer and improved OS (63) (Table 2).

Finally, Benarjee et al. have identified a local microbial signature associated with each type of breast tumor. *Actinomyces*, *Bartonella*, *Brevundimonas*, *Coxiella*, *Mobiluncus*, *Mycobacterium*, *Rickettsia*, and *Sphingomonas* are common in all types of breast cancer. In the luminal A subtype, *Arcanobacterium*, *Bifidobacterium*, *Cardiobacterium*, *Citrobacter*, and *Escherichia* are detected. *Bordetella*, *Campylobacter*, *Chlamydia*, *Chlamydophila*, *Legionella*, and *Pasteurella* are associated with the luminal B subtype. The HER2-enriched subtype is characterized by the presence of *Streptococcus*, whereas *Aerococcus*, *Arcobacter*, *Geobacillus*, *Orientia*, and *Rothia* are associated to with TNBC (65) (Table 2).

2.3 Prognostic biomarkers

The various omics technologies have allowed the identification of prognostic biomarkers in both solid and liquid samples. Gene expression analysis have identified genes associated with good and poor prognosis in breast cancer (18). In addition, using data from breast cancer databases (TCGA-BRCA and CMI-MBC), a 45-gene optimal prognostic gene signature has been constructed from genes regulated by tumor-associated macrophages (TAMs). All these genes are listed in Table 3 (75).

Not only genomic but also epigenomic differences have been found. By applying a powerful integrative network algorithm to paired DNA methylation and RNA-Seq data from ER+ breast cancer and adjacent healthy tissue, it has been shown that increased levels of DNA methylation and alterations in mRNA expression can predict poor prognosis. In particular, epigenetic silencing of WNT signaling antagonists and bone morphogenetic proteins (BMPs) has been observed in both luminal subtypes, but predominantly in luminal B breast cancer (76). Davalos et al. identified hypermethylation of *RASSF1A* and *PITX2* associated with poor prognosis in early stage of breast cancer (77).

Various proteomic analyses have validated proteins related to apoptosis, cell cycle arrest, cell adhesion, cytokeratins, cell metabolism and lipid binding as prognostic biomarkers of breast cancer OS (78, 79). Elevated expression of serine hydroxymethyltransferase 2 (SHMT2) correlates with poor OS and relapse-free survival (RFS), the amino acid transporter ASCT2 (SLC1A5) correlates with poor RFS (80), endoplasmin (HSP90B1) has been associated with distant metastasis and worse OS, and decorin (DCN) has been associated with lymph node metastasis, increased number of positive lymph nodes and worse OS; and (81). Finally, increased levels of E-cadherin and β-catenin correlate with poor survival in invasive breast cancer but not in lobular carcinoma (82) (Table 3).

An analysis of breast cancer transcriptomic and proteomic data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC)

TABLE 3 Prognostic biomarkers of breast cancer detected by different omics.

Omic	Biomarkers	Prognosis	REF
Genomic	*HSPA2*, *DNAJB5*, *HSCB*, *HSPA12B*	Good	(18)
Genomic	*CCT6A*, *DNAJA2*, *HSPA14*, *CCT7*, *HSPD1*, *CCT2*, *HSPA4*, *DNAJC6*, *CCT5*, *SEC63*, *HSPH1*, *CCT8*, *CCT4*, *HSP90AA1*, *HSPA8*, *DNAJC13*, *HSPA9*, *TCP1*	Poor	(18)
Genomic	*CS*, *SMARCE1*, *IGSF9B*, *SYTL4*, *CEMIP*, *EMC2*, *FHL2*, *RAMP3*, *CISD1*, *PAICS*, *TTI2*, *FIBCD1*, *ZCCHC9*, *VAV3*, *LIMD2*, *TANK*, *PAK6*, *ETFA*, *PRDM16*, *ADAM15*, *NFKBIZ*, *DDAH1*, *CC2D1B*, *SH2B2*, *ACYP2*, *ENDOV*, *KBTBD11*, *AL162595.1*, *PCED1B*, *LYSMD4*, *TRMT2B*, *SLC6A9*, *NOS1AP*, *LINC01291*, *PSMB10*, *RPL12P38*, *ZNF888*, *AL391845.1*, *LINC02585*, *LINC01431*, *AC099520.2*, *CEP95*, *MIR4713HG*, *RBM15B*, *AC061992.2*	-	(75)
Epigenomic	Increased levels of DNA methylation, alterations in mRNA expression, hypermethylation of *RASSF1A* and *PITX2*	Poor	(76, 77)
Proteomic	BCL2, CDH1, CLDN3, CLDN7, NADP, IDH2 CRABP2, SEC14L2	-	(78, 79)
Proteomic	SHMT2, SLC1A5, decorin, endoplasmin, E-cadherin, β-catenin	Poor	(75, 76, 80)
Proteomic Transcriptomic	FLT1, FADD, ALDOA, CXCL, FGFR1, PLCB3, PPP2R2A, RPA1	Poor	(78)

"-" means without information.

resource has identified 2 candidates associated with survival: Fms-related receptor tyrosine kinase 1 (FLT1) in TNBC, Fas-associated death domain (FADD) protein with the luminal type, while 8 candidates: Fructose-bisphosphate aldolase A (ALDOA), C-X-C motif chemokine (CXCL)16, fibroblast growth factor receptor 1 (FGFR1), 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-3 (PLCB3), serine/threonine protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform (PPP2R2A), and replication protein A1 (RPA1) are clearly associated with poor survival in the HER2-enriched type (83). Of note, protein glycosylation correlates with increased tumor burden and poor prognosis in breast cancer (84) (Table 3).

3 Omics data on the onset of breast cancer

Early-stage breast cancer lesions are so small that there may be insufficient material for analysis and it is difficult to obtain accurate data.

The onset of breast cancer is characterized by abnormal paracrine and autocrine signaling, as genes that are highly expressed in healthy breast epithelium are lost in carcinomas, including genes encoding cytokines such as *LIF*, *IL-6*, and *HIN-1*, also known as *SCGB3A1*, and chemokines such as *IL-8*, *GROα*, *GROβ*, and *MIP3α*, also known as *CCL20* (85). In addition, genes silenced by hypermethylation have been identified as responsible for mammary carcinogenesis, including *TWIST, RASSF1A*, *CCND2*, *HIN1*, *BRCA1*, *APC*, *GSTP1*, *BIN1*, *BMP6*, *ESR2*, *CDKN2A*, *CDKN1A*, *TIMP3* and *CST6*, as well as the WNT-negative regulators *WIF1* and *DKK3* (77). In addition, two methylated modifications (H3K9 me2 and me3) of the DNA packaging protein histone H3 decrease during cancer transformation, and the demethylase KDM3A/JMJD1A gradually increases (86). On the other hand, methylated genes such as *ITIH5*, *DKK3*, *RASSF1A*, *SFN*, *CDKN2A*, *MLH1*, *HOXD13* and *PCDHGB7* have been proposed as potential markers for early detection of breast cancer. Hypermethylation of *RASSF1A*, *CCDN2*, *HIN1* and *APC* correlates mainly with HR+ breast cancer, whereas hypermethylation of *CDH1* and *CDH13* is more frequent in TNBC patients (77). Furthermore, the expression of olfactomedin-4 (OLFM4) is higher in non-invasive breast tumors than in invasive breast cancer. Therefore, OLFM4 may also be a biomarker for early breast cancer (87) (Figure 2A).

Alterations in circulating metabolites have been identified in premenopausal women that may predict the development of breast cancer. In fact, ten metabolites listed in Figure 2A have been associated with breast cancer risk (88). Metabolomic signatures of the taurine and hypotaurine pathways and the alanine, aspartate and glutamate pathways obtained from the plasma of breast cancer patients are critical for early diagnosis (89). In addition, six metabolites and eight metabolic pathways have been identified in blood samples that can be used in the early diagnosis of breast cancer. Of the six metabolites, ethyl (R)-3-hydroxyhexanoate, caprylic acid, and hypoxanthine are noteworthy. Of the eight metabolic pathways identified, fatty acid and aminoacyl-tRNA biosynthesis and inositol phosphate metabolism are the pathways most implicated in the early diagnosis of BC (90). Alterations in lipid metabolism favor processes such as growth, proliferation, and motility of cancer cells, favoring tumor progression (91). The detection of lipids in human plasma samples has led to the identification of diagnostic biomarkers that reflect the early stage of TNBC (ES-TNBC). Diacylglycerol (DG) 34:2 is significantly downregulated in the TNBC subtype. Furthermore, a panel of 5 lipids (DG 34:2, PC 40:3, PC 39:8, PC 34:0 and PC 38:9) can differentiate TNBC from non-TNBC and ES-TNBC from ES-non-TNBC. Finally, in TNBC, ceramides are upregulated, whereas DG and LysoPC are downregulated and PC fluctuates (92) (Figure 2A).

4 Omics data on the progression of breast cancer

Genomic evidence suggests that IDC is a consequence of DCIS progression. Genes and/or signaling pathways are altered during tumor progression. The most significant changes occur during the transition from normal tissue to carcinoma *in situ*. In addition to the loss of *LIF*, *IL-6*, *HIN-1*, *IL-8*, *GROα*, *GROβ* and *MIP3α* in carcinomas, glutamine synthase (*GS*) and desmoplakin (*DSP*) are the only two genes specifically upregulated in DCIS, while the metabolic enzymes 3-phosphoglycerate dehydrogenase and glyceraldehyde dehydrogenase and mitochondrial NADH: Ubiquinone dehydrogenase and NADH dehydrogenase 1α, have been observed in invasive carcinomas (85). Videlicet, cancer cells have important metabolic alterations (Figure 2B).

Downregulated genes in DCIS include *TM4SF1*, *NFKB1A*, *PBEF*, *RASD1*, *TNFRSF10B*, *TNFAIP*, *CLU*, *NSEP1*, LITAF/PIG7, *BZW1*, and *CCNL1*, as well as genes encoding several cytokines and chemokines such as *ILF2*, *IL13RA1*, *LIF*, *CLC*, *CCL2*, and *CXCL1*. Some transcripts are frequently overexpressed in the DCIS such as *PKD1-like*, *STARD10*, *EPS8L2*, and *KIAA0545*. Some of these genes are associated with nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) and tumor necrosis factor (TNF) pathways, resulting in impaired apoptosis and sustained proliferation of breast cancer cells (93). In IDC, upregulated genes can be grouped into genes related to cell cycle, extracellular matrix or secreted proteins, cell adhesion and motility, and signal transduction. Several underexpressed genes have also been detected in IDC, such as *TM4SF1*, *TRAF4*, *PPARBP*, *AKR1A1*, *RSRP1*, *MAP1LC3A* and *RBBP6* (93). As for ILCs, they are characterized by *CDH1* alterations, as well as dysregulation in the PI3K/Akt pathway due to mutations in *PIK3CA*, *PTEN* alterations, and mutations in *AKT1* (94) (Figure 2B).

Transcriptomic analysis of tumor cells and their corresponding adjacent microenvironment cells [belonging to the Molecular Taxonomy of Breast Cancer International Consortium "METABRIC" cohort (95)] have shown that *CDH1* mutation may deregulate immune cells in the tumor microenvironment (96). Expression of genes encoding the α and β subunits of the integrins *ITGA4*, *ITGB2*, *ITGAX*, *ITGB7*, *ITGAM*, *ITGAL* and

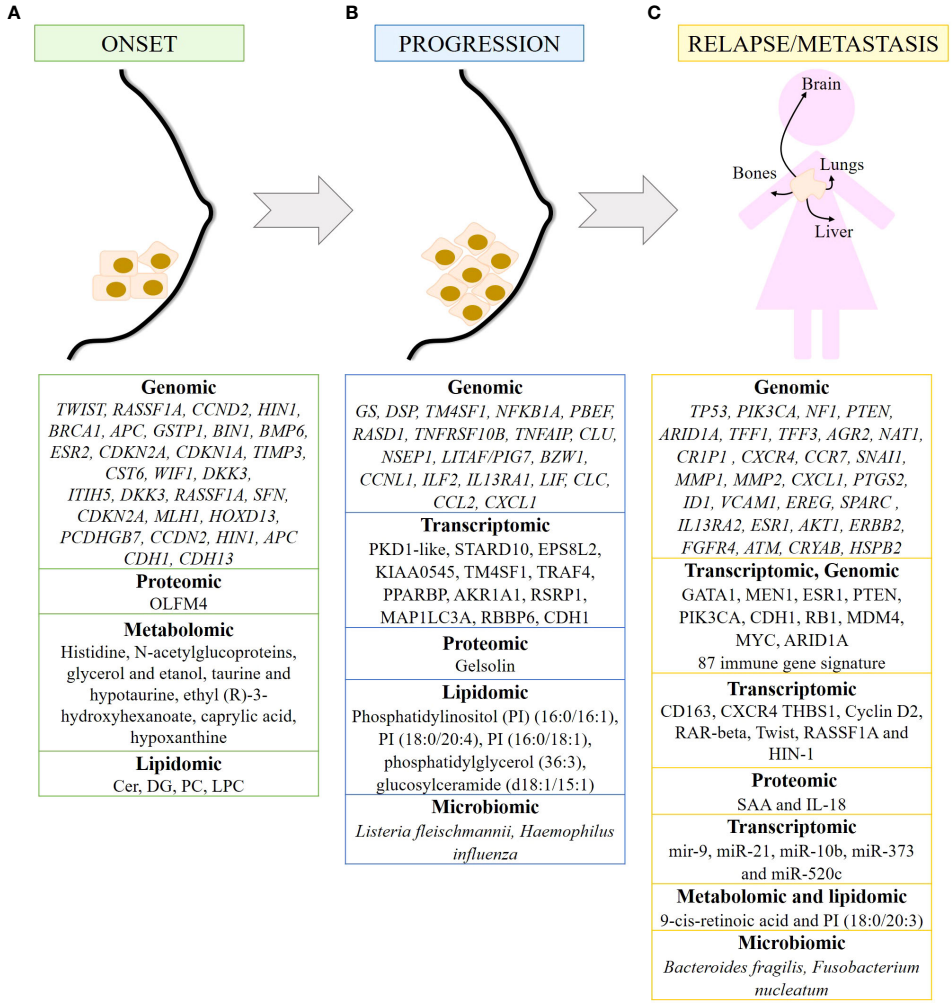


FIGURE 2
Biomarkers of breast cancer at different time points of tumor development detected by omics. Summary of markers observed in breast cancer onset **(A)**, progression **(B)** and metastasis/relapse **(C)** detected by genomic, transcriptomic, proteomic, metabolomics, lipidomic and/or microbiomic. 87 immune gene signature includes *LAG3, RELB, CCL2, IFNG, MSH6, ZC3HAV1, CD68, ORM1, LYZ, USP14, SLA2, HERC5, LAMP3, NONO, BATF, FCER1G, CCR5, REL, DTX3L, HMGB2, C2, CLEC4E, CLEC4D, CLEC7A, IL12A, CXCL10, CXCL11, RASGRP1, HAVCR2, ICOS, ATRIP, TRIM25, RNF166, CCR8, CSF1, NFAM1, TUBB4B, LYAR, CLEC12A, IL27, PIK3CG, XRCC6, PARP9, DNAJC5, MPEG1, TIFA, TLR1, CD47, EXO1, NCF2, SLAMF7, CTSS, GBP5, GBP4, GBP1, CREG1, RNF19B, RC3H2, RAB14, SYK, ACTR2, KCNAB2, OPTN, DDX58, IL2RA, JAK2, CLEC6A, LYST, CCL25, CCL8, HLADRA, RAB27A, PTK2B, PDCD1LG2, IFI30, TLR6, DSN1, HLA-DOB, CXCR6, TNF, IL10, SERPINA1, GSDMD, TRAF3, IL12RB1, CCL5* and *LIG4*.

ITGA8 correlates positively with the presence of immune cell infiltrates in the tumor, with markers of T cell activation and antigen presentation, and with immunosurveillance gene signatures. Expression of these integrins indicates a favorable prognosis in TNBC and HER2-enriched breast cancers. In contrast, expression of *IBSP*, *ITGB3BP*, *ITGB6*, *ITGB*1 and *ITGAV* predict a poor outcome (97) (Figure 2B).

Epigenetic modifications also contribute to breast cancer progression. For example, DNA methylation causes transcriptional silencing of tumor suppressor genes such as *RASSF1A*, *RARB*, *SFN* and *TGM2* (98). Genome-wide methylation analysis revealed that CpG sites were hypermethylated and hypomethylated after *CRY2* silencing. These data suggest that the absence of *CRY2* causes epigenetic dysregulation of genes leading to breast cancer progression (99). The differences in miRNA expression profile are greater in IDC than in ILC compared to their respective carcinomas *in situ*. Thus, IDC is characterized by upregulation of let-7d, miR-210 and miR-221 and downregulation of miR-10b, miR-126, miR-143, miR-218, and miR-335-5p. In contrast, ILC is characterized by upregulation of miR-9, miR-375, miR-182 and miR-183 (57) (Figure 2B).

Proteomic analysis of plasma samples from hereditary BC patients carrying a mutation in the *BRCA1* gene has shown that gelsolin, whose loss negatively correlates with tumor progression, is downregulated in these samples. In addition, its levels are associated with *BRCA1* mutational status (100) (Figure 2B). Furthermore, *O*-GlcNAcylation is increased in primary malignant breast tumors, and this increase is associated with increased expression of *O*-GlcNAc transferase in grades II and III breast tumors (31).

Metabolomic and transcriptomic data integration studies have enabled the identification of genes, pathways, and metabolites as a part of a cancer prediction model and a better understanding of

cancer progression. For example, adenosine monophosphate deaminase 1 (*AMPD1*) and ribonucleotide reductase regulatory subunit M2 (*RRM2*), which are involved in purine metabolism, have been associated with survival in breast cancer patients. Therefore, dysregulation of the purine metabolism pathway may influence breast cancer progression (101). Lipidomic analysis has identified two biomarkers capable of differentiating benign from malignant breast tumors: phosphatidylinositol (PI) (16:0/16:1) and PI (18:0/20:4). In addition, PI (16:0/18:1), phosphatidylglycerol (36:3) and glucosylceramide (d18:1/15:1) have been identified as potential biomarkers for assessing the degree of malignancy of breast tumors (102) (Figure 2B).

Microbiota dysbiosis contributes to breast cancer progression through its effects on skin and breast tissue. The presence of *Listeria fleischmannii* in breast tumor tissue is associated with epithelial-mesenchymal transition (EMT), whereas *Haemophilus influenza* correlates with tumor growth, cell cycle progression, E2F signaling, and mitotic spindle assembly (51). Alterations in the gut microbiome alter the production of bacterial metabolites that may influence tumor progression in breast cancer. Uric acid, glycolic acid, d-mannitol, 2,3-butanediol and trans-ferulic acid exert cytostatic effects, while 3-hydroxyphenylacetic acid, 4-hydroxybenzoic acid and vanillic acid stimulate breast cancer proliferation *in vitro*. In addition, 3-hydroxyphenylacetic acid, 4-hydroxybenzoic acid, 2,3-butanediol and hydrocinnamic acid inhibit EMT, and 2,3-butanediol has both cytostatic and anti-EMT properties (103). Gut microbiota may affect tumor progression by influencing the cancer-immunity dialogue. Gut bacteria elicit a complex and coordinated set of innate and adaptive immune responses to maintain tissue homeostasis. Consequently, when the microbiota-host balance is disrupted and dysbiosis occurs, increased production of inflammatory mediators is observed, which is associated with cancer progression. Low diversity in the gut microbiome is associated with decreased lymphocyte level and increased number of neutrophils, as well as decreased survival in breast cancer patients (51) (Figure 2B).

5 Omics data in relapsed and/or metastatic breast cancer

Between 7% and 11% of early breast cancers recur locally within 5 years after treatment, and 20% to 30% of primary breast cancers develop distant metastases. According to a study based on Surveillance, Epidemiology, and End Results Program (SEER) data for 2010-2013, the most common sites of breast cancer metastasis are bone (30-60%), lung (21-32%), liver (15-32%), and brain (4-10%) (104). Different molecular subtypes of breast cancer are associated with organ-specific metastases. Thus, luminal subtypes A and B metastasize primarily to the bone, the HER2-enriched subtype to the brain and liver, and TNBC to the lung (105). Genomic analyses have shown that metastases retain the same molecular subtype and prognostic signature as their primary tumors. These data suggest that metastatic potential is already determined in the primary tumor. The poor prognostic signature consists of genes that regulate cell cycle, invasion, metastasis, and angiogenesis (106, 107). Common alterations in *TP53* (51%) and *PIK3CA* (49%), as well as mutations/deletions in *NF1* (15%), mutations in *PTEN* (10%), and mutations/deletions in *ARID1A* (15%) have been identified in metastatic breast cancer tumors (108). Bone metastases of breast cancer are characterized by increased expression of the *TFF1*, *TFF3*, *AGR2*, *NAT1* and *CR1P1* genes, as well as the chemokine receptors CXCR4 and C-C chemokine receptor type 7 (CCR7), and upregulation of the zinc finger protein SNAI1 (SNAI1) (45, 109), which is involved in the induction of EMT (110). In breast to lung cancer metastasis, a number of genes such as *MMP1*, *MMP2*, *CXCL1*, *PTGS2*, *ID1*, *VCAM1*, *EREG*, *SPARC*, and *IL13RA2* have been identified in breast to lung cancer metastasis that promote and are clinically correlated; as well as the mitogen-activated protein kinases (MAPK), NFκB and vascular endothelial growth factor (VEGF) signaling pathways (111) (45). Mutations in the *ESR1*, *AKT1*, *ERBB2*, and *FGFR4* genes have been observed in metastatic breast tumors in the liver (45, 112, 113). Transcriptome analysis revealed that the TNF-α pathway is upregulated in lung metastases compared to liver metastases (114). *ATAD2*, *DERL1* and *NEK2A* have been shown to be overexpressed (115–117), whereas *ATM*, *CRYAB* and *HSPB2* genes are often suppressed and/or underexpressed in breast cancer metastases to the brain (45, 118, 119), (Figure 2C). The AURORA study, consisting of genomic and transcriptomic profiling in matched primary tumors and early metastases, has described the key role of somatic mutations *GATA1* and *MEN1* in metastasis. In addition, the enrichment of *ESR1*, *PTEN*, and *PIK3CA* in metastases have been determined, as well as *CDH1* and *RB1* mutations, *MDM4* and *MYC* amplifications, and *ARID1A* deletions (120). Of note, *TP53* mutations and *MYC* amplification are associated with shorter time to relapse (25). Eight common genes have been found to have significant effects on TNBC survival (*ELOB, SLC39A7, TIMM13, BANF1, NDUFS1, NDUFB7, TRAPPC5*, and *MVD*). Finally, a signature of 87 immune genes has been established that is highly predictive of pathologic complete response (pCR), which in turn correlates with improved OS and distant metastasis-free survival (DMFS) (121). These 87 immune genes are shown in Figure 2C.

The immune system is involved in the development of cancer, from tumor initiation to metastasis (122). Patients with stage 3 and 4 breast cancer have a higher percentage of immunosuppressive cells (granulocytic myeloid-derived suppressor cells (MDSCs), $CD14^{+}CD16^{+}$ intermediate monocytes, and $CD127^{-}CD25^{high}FoxP3^{+}$ Treg cells). Inflammation-related genes are differentially expressed in TNBC. In fact, low expression of CD163 and CXCR4 together with high expression of thrombospondin 1 (THBS1) correlates with an increased risk of relapse and poor survival in TNBC (123). The proinflammatory cytokines serum amyloid A (SAA) and IL-18 are elevated in the serum of patients with recurrent breast cancer. Therefore, SAA and IL-18 may be prognostic markers for breast cancer recurrence (124) (Figure 2C).

miRNAs are also involved in cancer migration and metastasis. miR-21, miR-10b, miR-373 and miR-520c promote metastasis, whereas miR-126, miR-335, miR-31, miR-146a, and miR-497 suppress metastasis. miR-9 is associated with local recurrence and ER+ tumors, whereas miR-10 is involved in cell proliferation,

migration, and invasion (57). On the other hand, *Cyclin D2*, *RAR-beta*, *Twist*, *RASSF1A* and *HIN-1* genes show increased methylation in bone, brain and lung metastases compared to primary breast cancer, with *HIN-1* and *RAR-beta* methylation significantly higher in each group (125). Hypermethylation and downregulation of genes involved in breast cancer progression or EMT, such as *LYN, MMP7, KLK10* and *WNT6*, are associated with a significantly lower risk of metastatic relapse (126). Furthermore, H3K4 acetylation has been correlated with breast cancer progression, estrogen responsiveness, and the oncogenic EMT pathway. Therefore, H3K4 is a potential biomarker for tumor progression leading to aggressive metastatic phenotypes (127) (Figure 2C).

Other biomarkers have been identified using other omics techniques. As mentioned above, E-cadherin is considered a good prognostic marker in non-invasive breast cancer, and loss of E-cadherin protein is one of the main features of EMT (82). Furthermore, a metabolomic study has identified 9-cis-retinoic acid as a critical metabolite in breast cancer progression, as it significantly decreased during breast cancer progression to metastasis. This suggests that 9-cis-retinoic acid inhibits tumor progression to metastasis, probably by attenuating cell invasion and migration (128). Combining lipidomic techniques with transcriptomic analysis, PI (18:0/20:3) accumulation has been found to be associated with an increased incidence of lymph node metastasis and activation of the PD-1-related immune checkpoint pathway (129).

Finally, alterations in the microbiota have also been observed to influence breast cancer metastasis. The presence of *Bacteroides fragilis*, a gut-colonizing bacterium, can induce epithelial hyperplasia to promote tumor growth and metastasis via the β-catenin–Notch1 axis (130). In addition, *Fusobacterium nucleatum* promotes tumor progression and metastasis (131) (Figure 2C).

6 Omics and treatment implications

Molecular changes that occur during cancer treatment determine the response to different therapies and guide the optimal choice of treatments to reduce local recurrence and distant metastasis, thereby increasing disease-free survival. The various omics are key to identifying these molecular changes. In fact, several clinical trials are based on one or more omics technologies.

Classically, aromatase inhibitors are recommended for patients with ER+ metastatic breast cancer because they suppress estrogen production. Trastuzumab and lapatinib are administered to patients with HER2-enriched breast tumors, as trastuzumab is a humanized monoclonal antibody against the extracellular domain of HER2, and lapatinib is a tyrosine kinase inhibitor that blocks both HER2 and EGFR activation. However, in light of the omics findings, more refined treatments may be considered according to genomic alterations. For example, both luminal subtypes frequently harbor mutations in *PIK3CA*, so inhibitors of this kinase may be a therapeutic target; *MYC* amplification suggests the use of platinum analogs and taxanes; and, patients with *BRCA1/2* mutations may benefit from poly ADP ribose polymerase (PARP) inhibitors and/or platinum compounds (44). The RAS/RAF/MEK/ERK signaling pathway is altered in tumors that overexpress EGFR and HER2. Since this kinase cascade is critical for survival and apoptosis, alterations in this pathway could affect sensitivity/resistance to anticancer therapies. Mutations in *KRAS*, *HRAS*, *NRAS*, *BRAF* and *NF1* have been observed in all breast cancer subtypes, although their frequency does not exceed 4%. However, copy number alterations (CNAs) of *KRAS* and *BRAF* have been observed in TNBC. Therefore, the use of inhibitors of these kinases in combination with other therapies may delay or induce resistance to treatment (132). *ESR1* methylation is considered a good predictor of survival in tamoxifen-treated patients, whereas *ARHI* methylation predicts survival in non-tamoxifen-treated patients. *BRCA1* hypermethylation, in turn, sensitizes TNBC patients to DNA-damaging chemotherapeutic agents such as cisplatin and PARP inhibitors (77). The prevalence of the PI3K/AKT/mTOR signaling axis has been observed in TNBC. The c-Kit protein, a receptor tyrosine kinase involved in the initiation of this cascade, is overexpressed in 20-25% of TNBC, hence, the tyrosine kinase inhibitors (TKIs) such as imatinib or sunitinib are used in the treatment of these cancers (133) (Table 4).

The endoplasmic reticulum protein KIAA1199, which is involved in tumor growth and invasiveness, is significantly overexpressed in breast tumor samples. Therefore, KIAA1199 is considered a new target for biomarker development and a novel therapeutic target for breast cancer (134, 135). Similarly, cofilin-1 (CFL-1), interleukin-32 (IL-32), proliferating cell nuclear antigen (PCNA), syntenin-1 (SDCBP), and riboforin-2 (RPN-2) have been identified as potential target antigens for HLA-A2+ TNBC immunotherapy (136), or the proto-oncogene RET and kallikrein (KLK)8 as antigens associated with breast cancer in general (137).

Metabolic and lipidomic profiling of TNBC samples combined with transcriptomic and genomic data have identified a number of metabolites as potential therapeutic targets for different transcriptomic subtypes of TNBC. The LAR subtype is characterized by an enrichment of ceramides and fatty acids. Therefore, sphingosine-1-phosphate (S1P), an intermediate of the ceramide pathway, may be a promising drug for the treatment of LAR tumors. In contrast, the basal-like immunosuppressed transcriptomic subtype (BLIS) is characterized by increased metabolites related to oxidation reaction and glycosyl transfer and the lowest level of metabolic dysregulation. In this case, N-acetyl-

TABLE 4 Refined treatments that take into account genomic studies.

Alteration	Treatment	REF
Mutations in PIK3CA	Inhibitors of PI3KA	(44)
Myc amplification	Platinum analogs and taxanes	(44)
Mutation in BRCA1/2	PARP inhibitors and platinum compounds	(44)
BRCA1 hypermethylation	Cisplatin, PARP inhibitors	(77)
c-Kit overexpression	Imatinib, sunitinib	(133)

aspartyl-glutamate has been identified as a key tumor-promoting metabolite and a potential therapeutic target for high-risk BLIS tumors (138). Furthermore, high levels of sphingomyelins are associated with improved DFS in patients with TNBC. Therefore, sphingomyelins and enzymes involved in sphingolipid metabolism could be considered as prognostic markers and potential therapeutic targets, respectively (139). Omics studies can also identify specific pharmacological pathways of resistance and sensitivity of tumor cells to different therapies. For example, a combination of *TP53* deficiency and silencing of *BRCA1*, *BRCA2*, or BRCA1/2-associated genes results in cisplatin sensitivity (140). *PI3KCA* mutations alone or in combination with *PTEN* appear to predict worse outcome after trastuzumab monotherapy or in combination with chemotherapy (141). *BRCA1/2*-deficient cells are sensitive to PARP1 inhibitors. In addition, silencing of kinases such as cyclin-dependent kinase 5 (CDK5), mitogen-activated protein kinase 12 (MAPK12), polo-like kinase 3 (PLK3), polynucleotide phosphatase/bifunctional kinase (PNKP), serine/threonine kinase (STK)22c, and STK36 strongly sensitize to PARP inhibitors (142). Cyclin-dependent kinase (CDK)10 has been identified as a determinant of resistance to endocrine therapies such as tamoxifen, as low CDK10 levels lead to early on tamoxifen treatment (143) (Figure 3A). The classic treatment for HR+ breast cancer is endocrine therapy. However, approximately 30% of patients develop resistance to endocrine therapy. A transcriptomic/proteomic study has identified the set of candidate genes *CEACAM1, KRT19, TMEM81, TMEM119, ESRRA, ERBB3, SRC, AKT1S1, SGEF, SCG5, ALOX12B, CKB, BID, XRCC1, NSL1*, and *CHEK2* that are able to discriminate progression/resistance (PD) from complete response (CR) and correlate significantly with survival (144). In addition, 298 differentially expressed genes were identified between drug-sensitive (DS) and drug-resistant (DR) breast cancer patients prior to neoadjuvant treatment. Among them, the peptidyl-prolyl cis-trans isomerase FKBP4 (FKBP4) and the protein S100-A9 (S100A9) could be putative predictive markers to distinguish the DR group from the DS group of breast cancer patients (145).

Metabolic studies have identified biomarkers that predict response to treatment. For example, glycohyocholic and glucodeoxycholic acids can stratify TNBC patients according to response to neoadjuvant chemotherapy and OS (146). On the other hand, when comparing patients with large primary breast cancer who had received neoadjuvant chemotherapy plus bevacizumab with those who had received chemotherapy alone, higher levels of leucine, acetoacetate and trihydroxybutyrate and lower levels of formate were observed 12 weeks after treatment (147). Furthermore, baseline immunometabolic assessment in combination with ER status could predict the response to neoadjuvant targeted chemotherapy (NATC) based on patient trastuzumab-paclitaxel combination and disease relapse in HER2+ patients. HER2+/ER+ patients have higher levels of T-cell stimulating factors, but also higher levels of cytokines that might be responsible for T-cell suppression. The combination of metabolic data with IL-2 and IL-10 cytokine levels has been shown to be prognostic for relapse (148). In addition, serum metabolites such as leucine, formate, valine, and proline, along with hormone receptor status, have been shown to be discriminators of NATC response. For example, formate, proline, valine, HR+, and HER2− are directly associated with NATC resistance. In contrast, leucine, HR− and HER2+ are directly related to NATC sensitivity. In addition, glyoxylate and dicarboxylate metabolism have been implicated in NATC resistance (149) (Figure 3A).

Pathologic complete response (pCR) and residual disease have been correlated with the genome, transcriptome, and tumor immune microenvironment in patients with early and locally advanced breast cancer undergoing neoadjuvant therapy (150). The pCR is associated with overexpression of driver genes such as *CDKN2A*, *EGFR*, *CCNE1*, and *MYC* and underexpression of *CCND1*, *ZNF703*, and *ESR1*, as well as increased immune activation characterized by enrichment of innate and adaptive

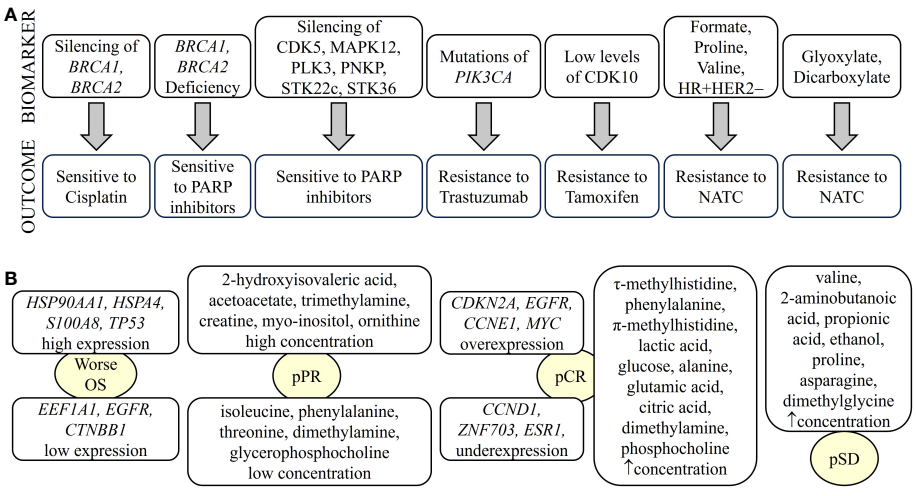


FIGURE 3
Biomarkers of breast cancer defining treatment outcome detected by omics. Biomarkers identified by different omics can determine drug sensitivity/resistance **(A)** or type of response to treatment **(B)**. NATC, neoadjuvant targeted therapy; pCR, pathological complete response; pPR, pathological partial response; pSD, pathological stable disease. "up arrow" means increase.

immune cells. However, tumors with residual disease, particularly HER2− tumors, showed enrichment of EMT and attenuated immune response due to the enrichment of inhibitory $CD56^{dim}$ natural killer cells and regulatory T cells, leading to therapy resistance (150). In transcriptomic profiling, *HSP90AA1*, *EEF1A1*, *APP* and *HSPA4* were associated with recurrence in breast cancer patients with pCR due to neoadjuvant chemotherapy. *TP53*, *EGFR*, *CTNNB1*, *ERBB2* and *HSPB1* may play an important role in the survival of pCR patients. Patients with tumors expressing high levels of *HSP90AA1*, *HSPA4*, *S100A8*, and *TP53* and low levels of *EEF1A1*, *EGFR* and *CTNNB1* showed significantly worse overall survival (OS) (151) (Figure 3B). A set of genes including *AKT1S1*, *NSL1*, *ESRRA*, *TMEM81*, *CKB*, *SGEF*, *KRT19*, *SCG5*, *CEACAM1*, *ALOX12B*, *IDB*, *SRC*, *CHEK2*, *ERBB3* and *XRCC1* have been identified as critical targets of both selective estrogen receptor modulators (SERMs)/selective estrogen receptor downregulators (SERDs) and aromatase inhibitors (AIs) of endocrine resistance (144).

A systemic metabolic study showed that patients with large primary breast cancers undergoing neoadjuvant chemotherapy with poor response have higher citrate levels and lower histidine levels (147). In addition, TNBC patients with pCR had elevated levels of circulating τ-methylhistidine, phenylalanine, π-methylhistidine, lactic acid, glucose, alanine, glutamic acid, citric acid, dimethylamine and phosphocholine, whereas patients with or pathological stable disease (pSD) had elevated levels of valine, 2-aminobutanoic acid, propionic acid, ethanol, proline, asparagine, and N,N-dimethylglycine. Finally, TNBC patients with pathologic partial response (pPR) had high levels of 2-hydroxyisovaleric acid, acetoacetate, trimethylamine, creatine, myo-inositol and ornithine, but low levels of five metabolites, namely isoleucine, phenylalanine, threonine, dimethylamine and glycerophosphocholine (Figure 3B). Therefore, we can infer that alterations in the pathways of glycine, serine, and threonine metabolism; valine, leucine, and isoleucine biosynthesis; and alanine, aspartate, and glutamate metabolism could be used as potential models to predict whether a patient with TNBC is suitable to receive neoadjuvant chemotherapy (152).

The relevance of all these biomarkers lies in their translational potential to identify specific treatments for breast cancer. Several clinical trials are ongoing. A Phase I program conducted at the University of Texas MD Anderson Cancer Center demonstrated that patients who received alteration-matched therapy had a higher objective response rate (ORR), PFS, and OS compared to unmatched therapy (153).

One of the most important clinical trials based on genomic studies is the "Microarray In Node Negative Disease may Avoid ChemoTherapy" (MINDACT). The MINDACT clinical trial aims to demonstrate the clinical relevance of the 70-gene prognostic signature (or MammaPrint™) determined by van't Veer (106) and to compare it with the traditional clinicopathologic prognostic indicators for the assignment of adjuvant chemotherapy in patients with node-negative breast cancer (154). As a result, MammaPrint™ was found to be effective as traditional tools in identifying high-risk patients, but more accurate in identifying low-risk patients who could avoid adjuvant chemotherapy (155, 156). This gene signature outperforms all traditional clinical prognostic factors and clearly discriminates patients with an excellent prognosis from those at high risk of recurrence (157, 158). Phase III results showed that approximately 46% of women with clinically high-risk breast cancer are unlikely to need chemotherapy (159). Women with high clinical risk and low genomic risk younger than 50 years had excellent DMFS when treated with endocrine therapy alone (160). In addition, patients with ER+, HER2− and stage I lymph node-negative tumors ≤2 cm treated with endocrine therapy had significantly fewer breast cancer events (161). On the other hand, no improvement in outcomes was observed with the use of docetaxel-capecitabine compared with anthracycline-based chemotherapy (162). Last year (2022), the Austrian Group Medical Tumor Therapy prospective registry confirmed that the addition of MammaPrint™ to the routine treatment of early luminal breast cancer yields clinically useful results (163).

Concurrent with the MINDACT clinical trial in Europe, the Trial Assigning Individualized Treatment Options Rx (TAILORx) was conducted in North America with the goal of reducing chemotherapy overtreatment by integrating molecular diagnostic testing into the clinical decision-making process. TAILORx is based on a 21-gene based assay (Oncotype DX™) that calculates a recurrence score (RS) and reserves chemotherapy for patients with a low RS (164). Women with a high RS who were treated with adjuvant chemotherapy regimens containing taxanes and/or anthracyclines plus endocrine therapy had an estimated 5-year freedom rate from distant breast cancer recurrence of 93% (165). In patients with ER+, HER2−, lymph node-negative and intermediate RS, adjuvant endocrine and chemoendocrine therapy had similar efficacy, although chemotherapy had some benefit in some women aged 50 years or younger (166). Among women with intermediate RS, Hispanic ethnicity and Asian race were associated with better outcomes. However, Black race was associated with worse clinical outcomes and did not benefit from adjuvant chemotherapy (167).

Other studies are the aforementioned AURORA US Metastatasis project and the TransNEO study. The AURORA US Metastasis project conducted a multiomic study including genomics, epigenomics and transcriptomics in primary tumors and their corresponding metastatic breast cancers (168). In metastatic TNBC, significantly lower expression of MHC class I genes (HLA-A, HLA-B, and HLA-C), DNA methylation of HLA-A, and small focal HLA-A were observed, which were associated with lower immunity and worse OS. Tumors with DNA-methylated HLA-A could be targeted for DNA demethylating drugs in combination with immune checkpoint inhibitors (ICI) (168, 169). The TransNEO molecular profiling study of patients with early and locally advanced breast cancer undergoing neoadjuvant therapy. Genomic, transcriptomic, and tumor immune microenvironment data were combined with clinical and digital pathology data to perform machine learning to create a predictive pCR model. This model is robust and could guide treatment selection in future clinical trials, including in the context of adjuvant therapy (150).

7 Conclusion and perspectives

Currently, treatment selection in breast cancer patients is based on broad clinicopathologic features that fail to accurately classify patients into risk groups, hence resulting into overtreatment of vast segments of patients. Data suggest that ORR, DFS and OS can be improved by the use of tailored therapies. Improved outcomes depend on the identification of new biomarkers that allow for the stratification of patients and eligibility for new therapies. Besides, the identification of biomarkers that predict treatment efficacy will minimize side effects or cumulative toxicity in patients unlikely to benefit from such treatments. Clinical trials, which thus far have mostly been based on genomic, transcriptomic and/or proteomic studies, have been effective in assigning treatment. It is our understanding that multiomics studies, including other omics techniques such as metabolomics, immunomics or microbiomics, are an important step towards precision medicine and hence refine the assignment of the best possible treatment for each patient. Special attention should be paid to the statistical methods used in the analysis of multiomic data to avoid spurious correlations. In fact, correlation coefficients should not be used to explain a process, such as cancer progression, in which multiple variables are involved. In these cases, the use of regression or multivariate analysis techniques will be more appropriate. In addition, validation cohorts will be needed to confirm the reproducibility, robustness, and validity of the results. It is important that validation cohorts have a pre-calculated sample size using statistical power tests and that minimum assay quality criteria have been established. In conclusion, we anticipate that clinical trials based on high-dimensional multiomics data interpreted by artificial intelligence will guide each patient to an optimized and personalized treatment, that will avoid overtreatment, minimize side effects, and improve both DFS and OS.

Author contributions

LA: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. DB: Validation, Writing – original draft, Writing – review & editing. MD: Validation, Writing – original draft, Writing – review & editing. MI: Validation, Writing – original draft, Writing – review & editing. GK: Validation, Writing – original draft, Writing – review & editing. RP: Validation, Writing – original draft, Writing – review & editing. LS: Conceptualization, Funding acquisition, Investigation, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. LS is supported by the Spanish Agencia Estatal de Investigación, AEI (grant number PID2021-126426OB-I00), the Strategic Program from the Institute of Biomedicine and Molecular Genetics (IBGM) of Valladolid (CCVC8485) and the Internationalization Project of the Unit of Excellence IBGM of Valladolid from the Junta de Castilla y León (CL-EI-2021 IBGM), the Institut National du Cancer (INCa, PLBIO2022-093), as well as the "Beatriz Galindo" Program from the Spanish Ministry of Universities. GK is supported by the Ligue contre le Cancer (équipe labellisée); Agence National de la Recherche (ANR) – Projets blancs; AMMICa US23/CNRS UMS3655; Association pour la recherche sur le cancer (ARC); Cancéropôle Ile-de-France; European Research Council Advanced Investigator Grand "ICD-Cancer", Fondation pour la Recherche Médicale (FRM); a donation by Elior; Equipex Onco-Pheno-Screen; European Joint Programme on Rare Diseases (EJPRD); European Research Council (ICD-Cancer), European Union Horizon 2020 Projects Oncobiome and Crimson; Fondation Carrefour; Institut National du Cancer (INCa); Institut Universitaire de France; LabEx Immuno-Oncology (ANR-18-IDEX-0001); a Cancer Research ASPIRE Award from the Mark Foundation; the RHU Immunolife; Seerave Foundation; SIRIC Stratified Oncology Cell DNA Repair and Tumor Immune Elimination (SOCRATE); and SIRIC Cancer Research and Personalized Medicine (CARPEM). This study contributes to the IdEx Université de Paris ANR-18-IDEX-0001. LA-F holds a predoctoral fellowship from the Asociación Española Contra el Cáncer (AECC). DB and RPR are supported by the University of Valladolid.

Conflict of interest

GK has been holding research contracts with Daiichi Sankyo, Eleor, Kaleido, Lytix Pharma, PharmaMar, Osasuna Therapeutics, Samsara Therapeutics, Sanofi, Tollys, and Vascage. GK is on the Board of Directors of the Bristol Myers Squibb Foundation France. GK is a scientific co-founder of everImmune, Osasuna Therapeutics, Samsara Therapeutics and Therafast Bio. GK is in the scientific advisory boards of Hevolution, Institut Servier and Longevity Vision Funds. GK is the inventor of patents covering therapeutic targeting of aging, cancer, cystic fibrosis, and metabolic disorders. GK'wife, Laurence Zitvogel, has held research contracts with Glaxo Smyth Kline, Incyte, Lytix, Kaleido, Innovate Pharma, Daiichi Sankyo, Pilege, Merus, Transgene, 9 m, Tusk and Roche, was on the on the Board of Directors of Transgene, is a cofounder of everImmune, and holds patents covering the treatment of cancer and the therapeutic manipulation of the microbiota. GK's brother, Romano Kroemer, was an employee of Sanofi and now consults for Boehringer-Ingelheim. The funders had no role in the design of the study, in the writing of the manuscript, or in the decision to publish the results.

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660

2. Zubair M, Wang S, Ali N. Advanced approaches to breast cancer classification and diagnosis. *Front Pharmacol* (2020) 11:632079. doi: 10.3389/fphar.2020.632079

3. Parise CA, Bauer KR, Brown MM, Caggiano V. Breast cancer subtypes as defined by the estrogen receptor (Er), progesterone receptor (Pr), and the human epidermal growth factor receptor 2 (Her2) among women with invasive breast cancer in California, 1999-2004. *Breast J* (2009) 15(6):593–602. doi: 10.1111/j.1524-4741.2009.00822.x

4. Patel HK, Bihani T. Selective estrogen receptor modulators (Serms) and selective estrogen receptor degraders (Serds) in cancer treatment. *Pharmacol Ther* (2018) 186:1–24. doi: 10.1016/j.pharmthera.2017.12.012

5. Li Z, Wei H, Li S, Wu P, Mao X. The role of progesterone receptors in breast cancer. *Drug Des Devel Ther* (2022) 16:305–14. doi: 10.2147/DDDT.S336643

6. Schlam I, Swain SM. Her2-positive breast cancer and tyrosine kinase inhibitors: the time is now. *NPJ Breast Cancer* (2021) 7(1):56. doi: 10.1038/s41523-021-00265-1

7. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature* (2000) 406(6797):747–52. doi: 10.1038/35021093

8. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* (2001) 98(19):10869–74. doi: 10.1073/pnas.191367098

9. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* (2011) 121(7):2750–67. doi: 10.1172/JCI45014

10. Lehmann BD, Pietenpol JA. Identification and use of biomarkers in treatment strategies for triple-negative breast cancer subtypes. *J Pathol* (2014) 232(2):142–50. doi: 10.1002/path.4280

11. Allison KH. Prognostic and predictive parameters in breast pathology: A pathologist's primer. *Mod Pathol* (2021) 34(Suppl 1):94–106. doi: 10.1038/s41379-020-00704-7

12. Mitola G, Falvo P, Bertolini F. New insight to overcome tumor resistance: an overview from cellular to clinical therapies. *Life (Basel)* (2021) 11(11). doi: 10.3390/life11111131

13. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* (2011) 144(5):646–74. doi: 10.1016/j.cell.2011.02.013

14. Hanahan D. Hallmarks of cancer: new dimensions. *Cancer Discov* (2022) 12 (1):31–46. doi: 10.1158/2159-8290.CD-21-1059

15. Rennstam K, Hedenfalk I. High-throughput genomic technology in research and clinical management of breast cancer. Molecular signatures of progression from benign epithelium to metastatic breast cancer. *Breast Cancer Res* (2006) 8(4):213. doi: 10.1186/bcr1528

16. Sato F, Saji S, Toi M. Genomic tumor evolution of breast cancer. *Breast Cancer* (2016) 23(1):4–11. doi: 10.1007/s12282-015-0617-8

17. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLoS Comput Biol* (2017) 13(5):e1005457. doi: 10.1371/journal.pcbi.1005457

18. Zoppino FCM, Guerrero-Gimenez ME, Castro GN, Ciocca DR. Comprehensive transcriptomic analysis of heat shock proteins in the molecular subtypes of human breast cancer. *BMC Cancer* (2018) 18(1):700. doi: 10.1186/s12885-018-4621-1

19. Perez de Souza L, Alseekh S, Scossa F, Fernie AR. Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research. *Nat Methods* (2021) 18(7):733–46. doi: 10.1038/s41592-021-01116-4

20. Adamczyk B, Tharmalingam T, Rudd PM. Glycans as cancer biomarkers. *Biochim Biophys Acta* (2012) 1820(9):1347–53. doi: 10.1016/j.bbagen.2011.12.001

21. Claudino WM, Quattrone A, Biganzoli L, Pestrin M, Bertini I, Di Leo A. Metabolomics: available results, current research projects in breast cancer, and future applications. *J Clin Oncol* (2007) 25(19):2840–6. doi: 10.1200/JCO.2006.09.7550

22. Silva C, Perestrelo R, Silva P, Tomas H, Camara JS. Breast cancer metabolomics: from analytical platforms to multivariate data analysis. *A Review Metabolites* (2019) 9 (5). doi: 10.3390/metabo9050102

23. Yugi K, Kubota H, Hatano A, Kuroda S. Trans-omics: how to reconstruct biochemical networks across multiple 'Omic' Layers. *Trends Biotechnol* (2016) 34 (4):276–90. doi: 10.1016/j.tibtech.2015.12.013

24. Sarhadi VK, Armengol G. Molecular biomarkers in cancer. *Biomolecules* (2022) 12(8). doi: 10.3390/biom12081021

25. Fumagalli C, Ranghiero A, Gandini S, Corso F, Taormina S, De Camilli E, et al. Inter-tumor genomic heterogeneity of breast cancers: comprehensive genomic profile of primary early breast cancers and relapses. *Breast Cancer Res* (2020) 22(1):107. doi: 10.1186/s13058-020-01345-z

26. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature* (2012) 490(7418):61–70. doi: 10.1038/nature11412

27. Rajendran BK, Deng CX. Characterization of potential driver mutations involved in human breast cancer by computational approaches. *Oncotarget* (2017) 8 (30):50252–72. doi: 10.18632/oncotarget.17225

28. Vreeker GCM, Vangangelt KMH, Bladergroen MR, Nicolardi S, Mesker WE, Wuhrer M, et al. Serum N-glycan profiles differ for various breast cancer subtypes. *Glycoconj J* (2021) 38(3):387–95. doi: 10.1007/s10719-021-10001-3

29. Abd Hamid UM, Royle L, Saldova R, Radcliffe CM, Harvey DJ, Storr SJ, et al. A strategy to reveal potential glycan markers from serum glycoproteins associated with breast cancer progression. *Glycobiology* (2008) 18(12):1105–18. doi: 10.1093/glycob/cwn095

30. Alley WR Jr., Madera M, Mechref Y, Novotny MV. Chip-based reversed-phase liquid chromatography-mass spectrometry of permethylated N-linked glycans: A potential methodology for cancer-biomarker discovery. *Anal Chem* (2010) 82 (12):5095–106. doi: 10.1021/ac100131e

31. Champattanachai V, Netsirisawan P, Chaiyawat P, Phueaouan T, Charoenwattanasatien R, Chokchaichamnankit D, et al. Proteomic analysis and abrogated expression of O-glcnacylated proteins associated with primary breast cancer. *Proteomics* (2013) 13(14):2088–99. doi: 10.1002/pmic.201200126

32. Takayama T, Tsutsui H, Shimizu I, Toyama T, Yoshimoto N, Endo Y, et al. Diagnostic approach to breast cancer patients based on target metabolomics in saliva by liquid chromatography with tandem mass spectrometry. *Clin Chim Acta* (2016) 452:18–26. doi: 10.1016/j.cca.2015.10.032

33. Silva CL, Perestrelo R, Capelinha F, Tomas H, Camara JS. An integrative approach based on Gc-Qms and Nmr metabolomics data as a comprehensive strategy to search potential breast cancer biomarkers. *Metabolomics* (2021) 17(8):72. doi: 10.1007/s11306-021-01823-1

34. An R, Yu H, Wang Y, Lu J, Gao Y, Xie X, et al. Integrative analysis of plasma metabolomics and proteomics reveals the metabolic landscape of breast cancer. *Cancer Metab* (2022) 10(1):13. doi: 10.1186/s40170-022-00289-6

35. Singh A, Sharma RK, Chagtoo M, Agarwal G, George N, Sinha N, et al. 1h nmr metabolomics reveals association of high expression of inositol 1, 4, 5 trisphosphate receptor and metabolites in breast cancer patients. *PLoS One* (2017) 12(1):e0169330. doi: 10.1371/journal.pone.0169330

36. Nagahashi M, Tsuchida J, Moro K, Hasegawa M, Tatsuda K, Woelfel IA, et al. High levels of sphingolipids in human breast cancer. *J Surg Res* (2016) 204(2):435–44. doi: 10.1016/j.jss.2016.05.022

37. Peetla C, Vijayaraghavalu S, Labhasetwar V. Biophysics of cell membrane lipids in cancer drug resistance: implications for drug transport and drug delivery with nanoparticles. *Adv Drug Deliv Rev* (2013) 65(13-14):1686–98. doi: 10.1016/j.addr.2013.09.004

38. Hilvo M, Denkert C, Lehtinen L, Muller B, Brockmoller S, Seppanen-Laakso T, et al. Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. *Cancer Res* (2011) 71(9):3236–45. doi: 10.1158/0008-5472.CAN-10-3894

39. Qiu Y, Zhou B, Su M, Baxter S, Zheng X, Zhao X, et al. Mass spectrometry-based quantitative metabolomics revealed a distinct lipid profile in breast cancer patients. *Int J Mol Sci* (2013) 14(4):8047–61. doi: 10.3390/ijms14048047

40. Parida S, Sharma D. The power of small changes: comprehensive analyses of microbial dysbiosis in breast cancer. *Biochim Biophys Acta Rev Cancer* (2019) 1871 (2):392–405. doi: 10.1016/j.bbcan.2019.04.001

41. Urbaniak C, Gloor GB, Brackstone M, Scott L, Tangney M, Reid G. The microbiota of breast tissue and its association with breast cancer. *Appl Environ Microbiol* (2016) 82(16):5039–48. doi: 10.1128/AEM.01235-16

42. Thompson KJ, Ingle JN, Tang X, Chia N, Jeraldo PR, Walther-Antonio MR, et al. A comprehensive analysis of breast cancer microbiota and host gene expression. *PLoS One* (2017) 12(11):e0188873. doi: 10.1371/journal.pone.0188873

43. Hieken TJ, Chen J, Hoskin TL, Walther-Antonio M, Johnson S, Ramaker S, et al. The microbiome of aseptically collected human breast tissue in benign and Malignant disease. *Sci Rep* (2016) 6:30751. doi: 10.1038/srep30751

44. Low SK, Zembutsu H, Nakamura Y. Breast cancer: the translation of big genomic data to cancer precision medicine. *Cancer Sci* (2018) 109(3):497–506. doi: 10.1111/cas.13463

45. Bennett C, Carroll C, Wright C, Awad B, Park JM, Farmer M, et al. Breast cancer genomics: primary and most common metastases. *Cancers (Basel)* (2022) 14(13). doi: 10.3390/cancers14133046

46. Fu-Jun L, Shao-Hua J, Xiao-Fang S. Differential proteomic analysis of pathway biomarkers in human breast cancer by integrated bioinformatics. *Oncol Lett* (2012) 4 (5):1097–103. doi: 10.3892/ol.2012.881

47. Saldova R, Haakensen VD, Rodland E, Walsh I, Stockmann H, Engebraaten O, et al. Serum N-glycome alterations in breast cancer during multimodal treatment and follow-up. *Mol Oncol* (2017) 11(10):1361–79. doi: 10.1002/1878-0261.12105

48. Torres CR, Hart GW. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked glcnac. *J Biol Chem* (1984) 259(5):3308–17. doi: 10.1016/S0021-9258(17)43295-9

49. Alakwaa FM, Chaudhary K, Garmire LX. Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data. *J Proteome Res* (2018) 17 (1):337–47. doi: 10.1021/acs.jproteome.7b00595

50. Urbaniak C, Cummins J, Brackstone M, Macklaim JM, Gloor GB, Baban CK, et al. Microbiota of human breast tissue. *Appl Environ Microbiol* (2014) 80(10):3007–14. doi: 10.1128/AEM.00242-14

51. Bernardo G, Le Noci V, Di Modica M, Montanari E, Triulzi T, Pupa SM, et al. The emerging role of the microbiota in breast cancer progression. *Cells* (2023) 12(15). doi: 10.3390/cells12151945

52. Esposito MV, Fosso B, Nunziato M, Casaburi G, D'Argenio V, Calabrese A, et al. Microbiome composition indicate dysbiosis and lower richness in tumor breast tissues compared to healthy adjacent paired tissue, within the same women. *BMC Cancer* (2022) 22(1):30. doi: 10.1186/s12885-021-09074-y

53. Liu L, Hao X, Song Z, Zhi X, Zhang S, Zhang J. Correlation between family history and characteristics of breast cancer. *Sci Rep* (2021) 11(1):6360. doi: 10.1038/s41598-021-85899-8

54. Quante M. [Deafness due to topical neomycin (Author's transl)]. *HNO* (1976) 24 (4):127–8.

55. Lakhani SR, Van De Vijver MJ, Jacquemier J, Anderson TJ, Osin PP, McGuffog L, et al. The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, Her-2, and P53 in patients with mutations in Brca1 and Brca2. *J Clin Oncol* (2002) 20(9):2310–8. doi: 10.1200/JCO.2002.09.023

56. Verhoog LC, Brekelmans CT, Seynaeve C, Dahmen G, van Geel AN, Bartels CC, et al. Survival in hereditary breast cancer associated with germline mutations of Brca2. *J Clin Oncol* (1999) 17(11):3396–402. doi: 10.1200/JCO.1999.17.11.3396

57. Samantarrai D, Dash S, Chhetri B, Mallick B. Genomic and epigenomic cross-talks in the regulatory landscape of Mirnas in breast cancer. *Mol Cancer Res* (2013) 11 (4):315–28. doi: 10.1158/1541-7786.MCR-12-0649

58. Lam SW, Jimenez CR, Boven E. Breast cancer classification by proteomic technologies: current state of knowledge. *Cancer Treat Rev* (2014) 40(1):129–38. doi: 10.1016/j.ctrv.2013.06.006

59. Tyanova S, Albrechtsen R, Kronqvist P, Cox J, Mann M, Geiger T. Proteomic maps of breast cancer subtypes. *Nat Commun* (2016) 7:10259. doi: 10.1038/ncomms10259

60. Michaut M, Chin SF, Majewski I, Severson TM, Bismeijer T, de Koning L, et al. Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. *Sci Rep* (2016) 6:18517. doi: 10.1038/srep18517

61. Santoro AL, Drummond RD, Silva IT, Ferreira SS, Juliano L, Vendramini PH, et al. *In situ* desi-msi lipidomic profiles of breast cancer molecular subtypes and precursor lesions. *Cancer Res* (2020) 80(6):1246–57. doi: 10.1158/0008-5472.CAN-18-3574

62. Diaz-Beltran L, Gonzalez-Olmedo C, Luque-Caro N, Diaz C, Martin-Blazquez A, Fernandez-Navarro M, et al. Human plasma metabolomics for biomarker discovery: targeting the molecular subtypes in breast cancer. *Cancers (Basel)* (2021) 13(1). doi: 10.3390/cancers13010147

63. Brockmoller SF, Bucher E, Muller BM, Budczies J, Hilvo M, Griffin JL, et al. Integration of metabolomics and expression of glycerol-3-phosphate acyltransferase (Gpam) in breast cancer-link to patient survival, hormone receptor status, and metabolic profiling. *J Proteome Res* (2012) 11(2):850–60. doi: 10.1021/pr200685r

64. Shin J, Kim G, Lee JW, Lee JE, Kim YS, Yu JH, et al. Identification of ganglioside Gm2 activator playing a role in cancer cell migration through proteomic analysis of breast cancer secretomes. *Cancer Sci* (2016) 107(6):828–35. doi: 10.1111/cas.12935

65. Banerjee S, Tian T, Wei Z, Shih N, Feldman MD, Peck KN, et al. Distinct microbial signatures associated with different breast cancer types. *Front Microbiol* (2018) 9:951. doi: 10.3389/fmicb.2018.00951

66. Negrini M, Calin GA. Breast cancer metastasis: A microrna story. *Breast Cancer Res* (2008) 10(2):203. doi: 10.1186/bcr1867

67. Ivey KN, Srivastava D. Microrna regulators of differentiation and cell fate decisions. *Cell Stem Cell* (2010) 7(1):36–41. doi: 10.1016/j.stem.2010.06.012

68. Wu J, Mamidi TKK, Zhang L, Hicks C. Unraveling the genomic-epigenomic interaction landscape in triple negative and non-triple negative breast cancer. *Cancers (Basel)* (2020) 12(6). doi: 10.3390/cancers12061559

69. Xie P, An R, Yu S, He J, Zhang H. A novel immune subtype classification of er-positive, Pr-negative and Her2-negative breast cancer based on the genomic and transcriptomic landscape. *J Transl Med* (2021) 19(1):398. doi: 10.1186/s12967-021-03076-x

70. Panis C, Pizzatti L, Herrera AC, Correa S, Binato R, Abdelhay E. Label-free proteomic analysis of breast cancer molecular subtypes. *J Proteome Res* (2014) 13 (11):4752–72. doi: 10.1021/pr500676x

71. Asleh K, Negri GL, Spencer Miko SE, Colborne S, Hughes CS, Wang XQ, et al. Proteomic analysis of archival breast cancer clinical specimens identifies biological subtypes with distinct survival outcomes. *Nat Commun* (2022) 13(1):896. doi: 10.1038/s41467-022-28524-0

72. Fan Y, Zhou X, Xia TS, Chen Z, Li J, Liu Q, et al. Human plasma metabolomics for identifying differential metabolites and predicting molecular subtypes of breast cancer. *Oncotarget* (2016) 7(9):9925–38. doi: 10.18632/oncotarget.7155

73. Tang X, Lin CC, Spasojevic I, Iversen ES, Chi JT, Marks JR. A joint analysis of metabolomics and genetics of breast cancer. *Breast Cancer Res* (2014) 16(4):415. doi: 10.1186/s13058-014-0415-9

74. Roig B, Rodriguez-Balada M, Samino S, Lam EW, Guaita-Esteruelas S, Gomes AR, et al. Metabolomics reveals novel blood plasma biomarkers associated to the brca1-mutated phenotype of human breast cancer. *Sci Rep* (2017) 7(1):17831. doi: 10.1038/s41598-017-17897-8

75. Long M, Wang J, Yang M. Transcriptomic profiling of breast cancer cells induced by tumor-associated macrophages generates a robust prognostic gene signature. *Cancers (Basel)* (2022) 14(21). doi: 10.3390/cancers14215364

76. Gao Y, Jones A, Fasching PA, Ruebner M, Beckmann MW, Widschwendter M, et al. The integrative epigenomic-transcriptomic landscape of Er positive breast cancer. *Clin Epigenet* (2015) 7:126. doi: 10.1186/s13148-015-0159-0

77. Davalos V, Martinez-Cardus A, Esteller M. The epigenomic revolution in breast cancer: from single-gene to genome-wide next-generation approaches. *Am J Pathol* (2017) 187(10):2163–74. doi: 10.1016/j.ajpath.2017.07.002

78. Osz A, Lanczky A, Gyorffy B. Survival analysis in breast cancer using proteomic data from four independent datasets. *Sci Rep* (2021) 11(1):16787. doi: 10.1038/s41598-021-96340-5

79. Geiger T, Madden SF, Gallagher WM, Cox J, Mann M. Proteomic portrait of human breast cancer progression identifies novel prognostic markers. *Cancer Res* (2012) 72(9):2428–39. doi: 10.1158/0008-5472.CAN-11-3711

80. Bernhardt S, Bayerlova M, Vetter M, Wachter A, Mitra D, Hanf V, et al. Proteomic profiling of breast cancer metabolism identifies Shmt2 and Asct2 as prognostic factors. *Breast Cancer Res* (2017) 19(1):112. doi: 10.1186/s13058-017-0905-7

81. Cawthorn TR, Moreno JC, Dharsee M, Tran-Thanh D, Ackloo S, Zhu PH, et al. Proteomic analyses reveal high expression of decorin and endoplasmin (Hsp90b1) are associated with breast cancer metastasis and decreased survival. *PLoS One* (2012) 7(2): e30992. doi: 10.1371/journal.pone.0030992

82. Borcherding N, Cole K, Kluz P, Jorgensen M, Kolb R, Bellizzi A, et al. Re-evaluating E-cadherin and Beta-catenin: A pan-cancer proteomic approach with an emphasis on breast cancer. *Am J Pathol* (2018) 188(8):1910–20. doi: 10.1016/j.ajpath.2018.05.003

83. Hari PS, Balakrishnan L, Kotyada C, Everad John A, Tiwary S, Shah N, et al. Proteogenomic analysis of breast cancer transcriptomic and proteomic data, using *de novo* transcript assembly: genome-wide identification of novel peptides and clinical implications. *Mol Cell Proteomics* (2022) 21(4):100220. doi: 10.1016/j.mcpro.2022.100220

84. Kirmiz C, Li B, An HJ, Clowers BH, Chew HK, Lam KS, et al. A serum glycomics approach to breast cancer biomarkers. *Mol Cell Proteomics* (2007) 6(1):43–55. doi: 10.1074/mcp.M600171-MCP200

85. Porter DA, Krop IE, Nasser S, Sgroi D, Kaelin CM, Marks JR, et al. A sage (Serial analysis of gene expression) view of breast tumor progression. *Cancer Res* (2001) 61 (15):5697–702.

86. Zhao QY, Lei PJ, Zhang X, Zheng JY, Wang HY, Zhao J, et al. Global histone modification profiling reveals the epigenomic dynamics during Malignant transformation in a four-stage breast cancer model. *Clin Epigenet* (2016) 8:34. doi: 10.1186/s13148-016-0201-x

87. Valo I, Raro P, Boissard A, Maarouf A, Jezequel P, Verriele V, et al. Olfm4 expression in ductal carcinoma *in situ* and in invasive breast cancer cohorts by a swath-based proteomic approach. *Proteomics* (2019) 19(21-22):e1800446. doi: 10.1002/pmic.201800446

88. Jobard E, Dossus L, Baglietto L, Fornili M, Lecuyer L, Mancini FR, et al. Investigation of circulating metabolites associated with breast cancer risk by untargeted metabolomics: A case-control study nested within the French E3n cohort. *Br J Cancer* (2021) 124(10):1734–43. doi: 10.1038/s41416-021-01304-1

89. Huang S, Chong N, Lewis NE, Jia W, Xie G, Garmire LX. Novel personalized pathway-based metabolomics models reveal key metabolic pathways for breast cancer diagnosis. *Genome Med* (2016) 8(1):34. doi: 10.1186/s13073-016-0289-9

90. Wei Y, Jasbi P, Shi X, Turner C, Hrovat J, Liu L, et al. Early breast cancer detection using untargeted and targeted metabolomics. *J Proteome Res* (2021) 20 (6):3124–33. doi: 10.1021/acs.jproteome.1c00019

91. Santos CR, Schulze A. Lipid metabolism in cancer. *FEBS J* (2012) 279(15):2610–23. doi: 10.1111/j.1742-4658.2012.08644.x

92. Eghlimi R, Shi X, Hrovat J, Xi B, Gu H. Triple negative breast cancer detection using Lc-Ms/Ms lipidomic profiling. *J Proteome Res* (2020) 19(6):2367–78. doi: 10.1021/acs.jproteome.0c00038

93. Abba MC, Drake JA, Hawkins KA, Hu Y, Sun H, Notcovich C, et al. Transcriptomic changes in human breast cancer progression as determined by serial analysis of gene expression. *Breast Cancer Res* (2004) 6(5):R499–513. doi: 10.1186/bcr899

94. Desmedt C, Zoppoli G, Sotiriou C, Salgado R. Transcriptomic and genomic features of invasive lobular breast cancer. *Semin Cancer Biol* (2017) 44:98–105. doi: 10.1016/j.semcancer.2017.03.007

95. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* (2012) 486(7403):346–52. doi: 10.1038/nature10983

96. Fox NS, Haider S, Harris AL, Boutros PC. Landscape of transcriptomic interactions between breast cancer and its microenvironment. *Nat Commun* (2019) 10(1):3116. doi: 10.1038/s41467-019-10929-z

97. Rojas K, Baliu-Pique M, Manzano A, Saiz-Ladera C, Garcia-Barberan V, Cimas FJ, et al. In silico transcriptomic mapping of integrins and immune activation in basal-like and Her2+ Breast cancer. *Cell Oncol (Dordr)* (2021) 44(3):569–80. doi: 10.1007/s13402-020-00583-9

98. Orlando FA, Brown KD. Unraveling breast cancer heterogeneity through transcriptomic and epigenomic analysis. *Ann Surg Oncol* (2009) 16(8):2270–9. doi: 10.1245/s10434-009-0500-y

99. Mao Y, Fu A, Hoffman AE, Jacobs DI, Jin M, Chen K, et al. The circadian gene Cry2 is associated with breast cancer aggressiveness possibly via epigenomic modifications. *Tumour Biol* (2015) 36(5):3533–9. doi: 10.1007/s13277-014-2989-3

100. Scumaci D, Tamme L, Fiumara CV, Pappaianni G, Concolino A, Leone E, et al. Plasma proteomic profiling in hereditary breast cancer reveals a brca1-specific signature: diagnostic and functional implications. *PloS One* (2015) 10(6):e0129762. doi: 10.1371/journal.pone.0129762

101. Luo X, Yu H, Song Y, Sun T. Integration of metabolomic and transcriptomic data reveals metabolic pathway alteration in breast cancer and impact of related signature on survival. *J Cell Physiol* (2019) 234(8):13021–31. doi: 10.1002/jcp.27973

102. Yang L, Cui X, Zhang N, Li M, Bai Y, Han X, et al. Comprehensive lipid profiling of plasma in patients with benign breast tumor and breast cancer reveals novel biomarkers. *Anal Bioanal Chem* (2015) 407(17):5065–77. doi: 10.1007/s00216-015-8484-x

103. Ujlaki G, Kovacs T, Vida A, Kokai E, Rauch B, Schwarcz S, et al. Identification of bacterial metabolites modulating breast cancer cell proliferation and epithelial-mesenchymal transition. *Molecules* (2023) 28(15). doi: 10.3390/molecules28155898

104. Wu Q, Li J, Zhu S, Wu J, Chen C, Liu Q, et al. Breast cancer subtypes predict the preferential site of distant metastases: A seer based study. *Oncotarget* (2017) 8 (17):27990–6. doi: 10.18632/oncotarget.15856

105. Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, et al. Metastatic behavior of breast cancer subtypes. *J Clin Oncol* (2010) 28(20):3271–7. doi: 10.1200/JCO.2009.25.9820

106. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* (2002) 415 (6871):530–6. doi: 10.1038/415530a

107. Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG, et al. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res* (2005) 65(20):9155–8. doi: 10.1158/0008-5472.CAN-05-2553

108. Akcakanat A, Zheng X, Cruz Pico CX, Kim TB, Chen K, Korkut A, et al. Genomic, transcriptomic, and proteomic profiling of metastatic breast cancer. *Clin Cancer Res* (2021) 27(11):3243–52. doi: 10.1158/1078-0432.CCR-20-4048

109. Smid M, Wang Y, Klijn JG, Sieuwerts AM, Zhang Y, Atkins D, et al. Genes associated with breast cancer metastatic to bone. *J Clin Oncol* (2006) 24(15):2261–7. doi: 10.1200/JCO.2005.03.8802

110. Kaufhold S, Bonavida B. Central role of snail1 in the regulation of Emt and resistance in cancer: A target for therapeutic intervention. *J Exp Clin Cancer Res* (2014) 33(1):62. doi: 10.1186/s13046-014-0062-0

111. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. *Nature* (2005) 436(7050):518–24. doi: 10.1038/nature03799

112. Chen X, Zheng Z, Chen L, Zheng H. Mapk, Nfkappab, and Vegf signaling pathways regulate breast cancer liver metastasis. *Oncotarget* (2017) 8(60):101452–60. doi: 10.18632/oncotarget.20843

113. Tian C, Liu S, Wang Y, Song X. Prognosis and genomic landscape of liver metastasis in patients with breast cancer. *Front Oncol* (2021) 11:588136. doi: 10.3389/fonc.2021.588136

114. Berthelet J, Wimmer VC, Whitfield HJ, Serrano A, Boudier T, Mangiola S, et al. The site of breast cancer metastases dictates their clonal composition and reversible transcriptomic profile. *Sci Adv* (2021) 7(28). doi: 10.1126/sciadv.abf4408

115. Ciro M, Prosperini E, Quarto M, Grazini U, Walfridsson J, McBlane F, et al. Atad2 is a novel cofactor for Myc, overexpressed and amplified in aggressive tumors. *Cancer Res* (2009) 69(21):8491–8. doi: 10.1158/0008-5472.CAN-09-2131

116. Wang J, Hua H, Ran Y, Zhang H, Liu W, Yang Z, et al. Derlin-1 is overexpressed in human breast carcinoma and protects cancer cells from endoplasmic reticulum stress-induced apoptosis. *Breast Cancer Res* (2008) 10(1):R7. doi: 10.1186/bcr1849

117. Wang S, Li W, Liu N, Zhang F, Liu H, Liu F, et al. Nek2a contributes to tumorigenic growth and possibly functions as potential therapeutic target for human breast cancer. *J Cell Biochem* (2012) 113(6):1904–14. doi: 10.1002/jcb.24059

118. Baysal BE, Willett-Brozick JE, Taschner PE, Dauwerse JG, Devilee P, Devlin B. A high-resolution integrated map spanning the sdhd gene at 11q23: A 1.1-Mb Bac Contig, a partial transcript map and 15 new repeat polymorphisms in a tumour-suppressor region. *Eur J Hum Genet* (2001) 9(2):121–9. doi: 10.1038/sj.ejhg.5200585

119. Smith J, Tho LM, Xu N, Gillespie DA. The Atm-Chk2 and Atr-Chk1 pathways in DNA damage signaling and cancer. *Adv Cancer Res* (2010) 108:73–112. doi: 10.1016/B978-0-12-380888-2.00003-0

120. Aftimos P, Oliveira M, Irrthum A, Fumagalli D, Sotiriou C, Gal-Yam EN, et al. Genomic and transcriptomic analyses of breast cancer primaries and matched metastases in Aurora, the breast international group (Big) molecular screening initiative. *Cancer Discov* (2021) 11(11):2796–811. doi: 10.1158/2159-8290.CD-20-1647

121. Vishnubalaji R, Abdel-Razeq H, Gehani S, Albagha OME, Alajez NM. Identification of a gene panel predictive of triple-negative breast cancer response to neoadjuvant chemotherapy employing transcriptomic and functional validation. *Int J Mol Sci* (2022) 23(18). doi: 10.3390/ijms231810901

122. Gonzalez H, Hagerling C, Werb Z. Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes Dev* (2018) 32(19-20):1267–84. doi: 10.1101/gad.314617.118

123. Foulds GA, Vadakekolathu J, Abdel-Fatah TMA, Nagarajan D, Reeder S, Johnson C, et al. Immune-phenotyping and transcriptomic profiling of peripheral blood mononuclear cells from patients with breast cancer: identification of a 3 gene signature which predicts relapse of triple negative breast cancer. *Front Immunol* (2018) 9:2028. doi: 10.3389/fimmu.2018.02028

124. Bera A, Russ E, Manoharan MS, Eidelman O, Eklund M, Hueman M, et al. Proteomic analysis of inflammatory biomarkers associated with breast cancer recurrence. *Mil Med* (2020) 185(Suppl 1):669–75. doi: 10.1093/milmed/usz254

125. Mehrotra J, Vali M, McVeigh M, Kominsky SL, Fackler MJ, Lahti-Domenici J, et al. Very high frequency of hypermethylated genes in breast cancer metastasis to the bone, brain, and lung. *Clin Cancer Res* (2004) 10(9):3104–9. doi: 10.1158/1078-0432.ccr-03-0118

126. Fang F, Turcan S, Rimner A, Kaufman A, Giri D, Morris LG, et al. Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci Transl Med* (2011) 3 (75):75ra25. doi: 10.1126/scitranslmed.3001875

127. Messier TL, Gordon JA, Boyd JR, Tye CE, Browne G, Stein JL, et al. Histone H3 lysine 4 acetylation and methylation dynamics define breast cancer subtypes. *Oncotarget* (2016) 7(5):5094–109. doi: 10.18632/oncotarget.6922

128. Wu J, Yang R, Zhang L, Li Y, Liu B, Kang H, et al. Metabolomics research on potential role for 9-cis-retinoic acid in breast cancer progression. *Cancer Sci* (2018) 109 (7):2315–26. doi: 10.1111/cas.13629

129. Kawashima M, Tokiwa M, Nishimura T, Kawata Y, Sugimoto M, Kataoka TR, et al. High-resolution imaging mass spectrometry combined with transcriptomic analysis identified a link between fatty acid composition of phosphatidylinositols and the immune checkpoint pathway at the primary tumour site of breast cancer. *Br J Cancer* (2020) 122(2):245–57. doi: 10.1038/s41416-019-0662-8

130. Parida S, Wu S, Siddharth S, Wang G, Muniraj N, Nagalingam A, et al. A procarcinogenic colon microbe promotes breast tumorigenesis and metastatic progression and concomitantly activates notch and Beta-catenin axes. *Cancer Discov* (2021) 11(5):1138–57. doi: 10.1158/2159-8290.CD-20-0537

131. Parhi L, Alon-Maimon T, Sol A, Nejman D, Shhadeh A, Fainsod-Levi T, et al. Breast cancer colonization by fusobacterium nucleatum accelerates tumor growth and metastatic progression. *Nat Commun* (2020) 11(1):3259. doi: 10.1038/s41467-020-16967-2

132. Rocca A, Braga L, Volpe MC, Maiocchi S, Generali D. The predictive and prognostic role of Ras-Raf-Mek-Erk pathway alterations in breast cancer: revision of the literature and comparison with the analysis of cancer genomic datasets. *Cancers (Basel)* (2022) 14(21). doi: 10.3390/cancers14215306

133. Gromova I, Espinoza JA, Grauslund M, Santoni-Rugiu E, Moller Talman ML, van Oostrum J, et al. Functional proteomic profiling of triple-negative breast cancer. *Cells* (2021) 10(10). doi: 10.3390/cells10102768

134. Evensen NA, Kuscu C, Nguyen HL, Zarrabi K, Dufour A, Kadam P, et al. Unraveling the role of Kiaa1199, a novel endoplasmic reticulum protein, in cancer cell migration. *J Natl Cancer Inst* (2013) 105(18):1402–16. doi: 10.1093/jnci/djt224

135. Jami MS, Hou J, Liu M, Varney ML, Hassan H, Dong J, et al. Functional proteomic analysis reveals the involvement of Kiaa1199 in breast cancer growth, motility and invasiveness. *BMC Cancer* (2014) 14:194. doi: 10.1186/1471-2407-14-194

136. Ternette N, Olde Nordkamp MJM, Muller J, Anderson AP, Nicastri A, Hill AVS, et al. Immunopeptidomic profiling of Hla-A2-positive triple negative breast cancer identifies potential immunotherapy target antigens. *Proteomics* (2018) 18(12): e1700465. doi: 10.1002/pmic.201700465

137. Yi X, Liao Y, Wen B, Li K, Dou Y, Savage SR, et al. Caatlas: an immunopeptidome atlas of human cancer. *iScience* (2021) 24(10):103107. doi: 10.1016/j.isci.2021.103107

138. Xiao Y, Ma D, Yang YS, Yang F, Ding JH, Gong Y, et al. Comprehensive metabolomics expands precision medicine for triple-negative breast cancer. *Cell Res* (2022) 32(5):477–90. doi: 10.1038/s41422-022-00614-0

139. Purwaha P, Gu F, Piyarathna DWB, Rajendiran T, Ravindran A, Omilian AR, et al. Unbiased lipidomic profiling of triple-negative breast cancer tissues reveals the association of sphingomyelin levels with patient disease-free survival. *Metabolites* (2018) 8(3). doi: 10.3390/metabo8030041

140. Bartz SR, Zhang Z, Burchard J, Imakura M, Martin M, Palmieri A, et al. Small interfering Rna screens reveal enhanced cisplatin cytotoxicity in tumor cells having both Brca network and Tp53 disruptions. *Mol Cell Biol* (2006) 26(24):9377–86. doi: 10.1128/MCB.01229-06

141. Berns K, Horlings HM, Hennessy BT, Madiredjo M, Hijmans EM, Beelen K, et al. A functional genetic approach identifies the pi3k pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer Cell* (2007) 12(4):395–402. doi: 10.1016/j.ccr.2007.08.030

142. Turner NC, Lord CJ, Iorns E, Brough R, Swift S, Elliott R, et al. A synthetic lethal sirna screen identifying genes mediating sensitivity to a parp inhibitor. *EMBO J* (2008) 27(9):1368–77. doi: 10.1038/emboj.2008.61

143. Iorns E, Turner NC, Elliott R, Syed N, Garrone O, Gasco M, et al. Identification of Cdk10 as an important determinant of resistance to endocrine therapy for breast cancer. *Cancer Cell* (2008) 13(2):91–104. doi: 10.1016/j.ccr.2008.01.001

144. Li CL, Moi SH, Lin HS, Hou MF, Chen FM, Shih SL, et al. Comprehensive transcriptomic and proteomic analyses identify a candidate gene set in cross-resistance for endocrine therapy in breast cancer. *Int J Mol Sci* (2022) 23(18). doi: 10.3390/ijms231810539

145. Yang WS, Moon HG, Kim HS, Choi EJ, Yu MH, Noh DY, et al. Proteomic approach reveals Fkbp4 and S100a9 as potential prediction markers of therapeutic response to neoadjuvant chemotherapy in patients with breast cancer. *J Proteome Res* (2012) 11(2):1078–88. doi: 10.1021/pr2008187

146. Diaz C, Gonzalez-Olmedo C, Diaz-Beltran L, Camacho J, Mena Garcia P, Martin-Blazquez A, et al. Predicting dynamic response to neoadjuvant chemotherapy in breast cancer: A novel metabolomics approach. *Mol Oncol* (2022) 16(14):2658–71. doi: 10.1002/1878-0261.13216

147. Debik J, Euceda LR, Lundgren S, Gythfeldt HVL, Garred O, Borgen E, et al. Assessing treatment response and prognosis by serum and tissue metabolomics in breast cancer patients. *J Proteome Res* (2019) 18(10):3649–60. doi: 10.1021/acs.jproteome.9b00316

148. Vignoli A, Muraro E, Miolo G, Tenori L, Turano P, Di Gregorio E, et al. Effect of estrogen receptor status on circulatory immune and metabolomics profiles of Her2-positive breast cancer patients enrolled for neoadjuvant targeted chemotherapy. *Cancers (Basel)* (2020) 12(2). doi: 10.3390/cancers12020314

149. Cardoso MR, Silva AAR, Talarico MCR, Sanches PHG, Sforca ML, Rocco SA, et al. Metabolomics by nmr combined with machine learning to predict neoadjuvant chemotherapy response for breast cancer. *Cancers (Basel)* (2022) 14(20). doi: 10.3390/cancers14205055

150. Sammut SJ, Crispin-Ortuzar M, Chin SF, Provenzano E, Bardwell HA, Ma W, et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature* (2022) 601(7894):623–9. doi: 10.1038/s41586-021-04278-5

151. Takeshita T, Yan L, Peng X, Kimbung S, Hatschek T, Hedenfalk IA, et al. Transcriptomic and functional pathway features were associated with survival after pathological complete response to neoadjuvant chemotherapy in breast cancer. *Am J Cancer Res* (2020) 10(8):2555–69.

152. He X, Gu J, Zou D, Yang H, Zhang Y, Ding Y, et al. Nmr-based metabolomics analysis predicts response to neoadjuvant chemotherapy for triple-negative breast cancer. *Front Mol Biosci* (2021) 8:708052. doi: 10.3389/fmolb.2021.708052

153. Tsimberidou AM, Wen S, Hong DS, Wheler JJ, Falchook GS, Fu S, et al. Personalized medicine for patients with advanced cancer in the phase I program at Md Anderson: validation and landmark analyses. *Clin Cancer Res* (2014) 20(18):4827–36. doi: 10.1158/1078-0432.CCR-14-0603

154. Bogaerts J, Cardoso F, Buyse M, Braga S, Loi S, Harrison JA, et al. Gene signature evaluation as a prognostic tool: challenges in the design of the mindact trial. *Nat Clin Pract Oncol* (2006) 3(10):540–51. doi: 10.1038/ncponc0591

155. Mook S, Van't Veer LJ, Rutgers EJ, Piccart-Gebhart MJ, Cardoso F. Individualization of therapy using mammaprint: from development to the mindact trial. *Cancer Genomics Proteomics* (2007) 4(3):147–55.

156. Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ. Clinical application of the 70-gene profile: the Mindact trial. *J Clin Oncol* (2008) 26(5):729–35. doi: 10.1200/JCO.2007.14.3222

157. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* (2002) 347(25):1999–2009. doi: 10.1056/NEJMoa021967

158. Rutgers E, Piccart-Gebhart MJ, Bogaerts J, Delaloge S, Veer LV, Rubio IT, et al. The Eortc 10041/Big 03-04 Mindact trial is feasible: results of the pilot phase. *Eur J Cancer* (2011) 47(18):2742–9. doi: 10.1016/j.ejca.2011.09.016

159. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med* (2016) 375(8):717–29. doi: 10.1056/NEJMoa1602253

160. Piccart M, van 't Veer LJ, Poncet C, Lopes Cardozo JMN, Delaloge S, Pierga JY, et al. 70-gene signature as an aid for treatment decisions in early breast cancer: updated results of the phase 3 randomised mindact trial with an exploratory analysis by age. *Lancet Oncol* (2021) 22(4):476–88. doi: 10.1016/S1470-2045(21)00007-3

161. Lopes Cardozo JMN, Byng D, Drukker CA, Schmidt MK, Binuya MA, van 't Veer LJ, et al. Outcome without any adjuvant systemic treatment in stage I Er+/Her2- breast cancer patients included in the mindact trial. *Ann Oncol* (2022) 33(3):310–20. doi: 10.1016/j.annonc.2021.11.014

162. Delaloge S, Piccart M, Rutgers E, Litiere S, van 't Veer LJ, van den Berkmortel F, et al. Standard anthracycline based versus docetaxel-capecitabine in early high clinical and/or genomic risk breast cancer in the Eortc 10041/Big 3-04 Mindact phase iii trial. *J Clin Oncol* (2020) 38(11):1186–97. doi: 10.1200/JCO.19.01371

163. Westphal T, Gampenrieder SP, Rinnerthaler G, Balic M, Posch F, Dandachi N, et al. Transferring mindact to daily routine: implementation of the 70-gene signature in luminal early breast cancer - results from a prospective registry of the Austrian group medical tumor therapy (Agmt). *Breast Care (Basel)* (2022) 17(1):1–9. doi: 10.1159/000512467

164. Sparano JA. Tailorx: trial assigning individualized options for treatment (Rx). *Clin Breast Cancer* (2006) 7(4):347–50. doi: 10.3816/CBC.2006.n.051

165. Sparano JA, Gray RJ, Makower DF, Albain KS, Saphner TJ, Badve SS, et al. Clinical outcomes in early breast cancer with a high 21-gene recurrence score of 26 to 100 assigned to adjuvant chemotherapy plus endocrine therapy: A secondary analysis of the tailorx randomized clinical trial. *JAMA Oncol* (2020) 6(3):367–74. doi: 10.1001/jamaoncol.2019.4794

166. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med* (2018) 379(2):111–21. doi: 10.1056/NEJMoa1804710

167. Albain KS, Gray RJ, Makower DF, Faghih A, Hayes DF, Geyer CE, et al. Race, ethnicity, and clinical outcomes in hormone receptor-positive, Her2-negative, node-negative breast cancer in the randomized Tailorx trial. *J Natl Cancer Inst* (2021) 113 (4):390–9. doi: 10.1093/jnci/djaa148

168. Garcia-Recio S, Hinoue T, Wheeler GL, Kelly BJ, Garrido-Castro AC, Pascual T, et al. Multiomics in primary and metastatic breast tumors from the Aurora us network finds microenvironment and epigenetic drivers of metastasis. *Nat Cancer* (2023) 4(1):128–47. doi: 10.1038/s43018-022-00491-x

169. Topper MJ, Vaz M, Marrone KA, Brahmer JR, Baylin SB. The emerging role of epigenetic therapeutics in immuno-oncology. *Nat Rev Clin Oncol* (2020) 17(2):75–90. doi: 10.1038/s41571-019-0266-5

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 08 February 2024
DOI 10.3389/fonc.2024.1237378

OPEN ACCESS

EDITED BY
Luisa Lanfrancone,
European Institute of Oncology (IEO), Italy

REVIEWED BY
Marielle Elizabeth Yohe,
National Institutes of Health (NIH),
United States
Rana A. Youness,
German International University, Egypt

*CORRESPONDENCE
Yusuke Suenaga
ysuenaga@chiba-cc.jp

RECEIVED 09 June 2023
ACCEPTED 15 January 2024
PUBLISHED 08 February 2024

CITATION
Nakatani K, Kogashi H, Miyamoto T, Setoguchi T, Sakuma T, Kugou K, Hasegawa Y, Yamamoto T, Hippo Y and Suenaga Y (2024) Inhibition of OCT4 binding at the *MYCN* locus induces neuroblastoma cell death accompanied by downregulation of transcripts with high-open reading frame dominance.
Front. Oncol. 14:1237378.
doi: 10.3389/fonc.2024.1237378

Inhibition of OCT4 binding at the *MYCN* locus induces neuroblastoma cell death accompanied by downregulation of transcripts with high-open reading frame dominance

Kazuma Nakatani[1,2,3,4], Hiroyuki Kogashi[1,2], Takanori Miyamoto[1], Taiki Setoguchi[5], Tetsushi Sakuma[6], Kazuto Kugou[7], Yoshinori Hasegawa[7], Takashi Yamamoto[6], Yoshitaka Hippo[1,2,8] and Yusuke Suenaga[1]*

[1]Laboratory of Evolutionary Oncology, Chiba Cancer Center Research Institute, Chiba, Japan, [2]Graduate School of Medical and Pharmaceutical Sciences, Chiba University, Chiba, Japan, [3]Innovative Medicine CHIBA Doctoral WISE Program, Chiba University, Chiba, Japan, [4]All Directional Innovation Creator Ph.D. Project, Chiba University, Chiba, Japan, [5]Department of Neurosurgery, Chiba Cancer Center, Chiba, Japan, [6]Graduate School of Integrated Sciences for Life, Hiroshima University, Hiroshima, Japan, [7]Department of Applied Genomics, Kazusa DNA Research Institute, Chiba, Japan, [8]Laboratory of Precision Tumor Model Systems, Chiba Cancer Center Research Institute, Chiba, Japan

Amplification of *MYCN* is observed in high-risk neuroblastomas (NBs) and is associated with a poor prognosis. *MYCN* expression is directly regulated by multiple transcription factors, including OCT4, MYCN, CTCF, and p53 in NB. Our previous study showed that inhibition of p53 binding at the *MYCN* locus induces NB cell death. However, it remains unclear whether inhibition of alternative transcription factor induces NB cell death. In this study, we revealed that the inhibition of OCT4 binding at the *MYCN* locus, a critical site for the human-specific OCT4–MYCN positive feedback loop, induces caspase-2-mediated cell death in *MYCN*-amplified NB. We used the CRISPR/deactivated Cas9 (dCas9) technology to specifically inhibit transcription factors from binding to the *MYCN* locus in the *MYCN*-amplified NB cell lines CHP134 and IMR32. In both cell lines, the inhibition of OCT4 binding at the *MYCN* locus reduced MYCN expression, thereby suppressing MYCN-target genes. After inhibition of OCT4 binding, differentially downregulated transcripts were associated with high-open reading frame (ORF) dominance score, which is associated with the translation efficiency of transcripts. These transcripts were enriched in splicing factors, including MYCN-target genes such as *HNRNPA1* and *PTBP1*. Furthermore, transcripts with a high-ORF dominance score were significantly associated with genes whose high expression is associated with a poor prognosis in NB. Because the ORF dominance score correlates with the translation efficiency of transcripts, our findings suggest that MYCN maintains the expression of transcripts with high translation efficiency, contributing to a poor prognosis in NB. In conclusion, the inhibition of OCT4 binding at the *MYCN* locus resulted in reduced MYCN activity, which in turn led to the downregulation of high-ORF dominance transcripts and subsequently

induced caspase-2-mediated cell death in *MYCN*-amplified NB cells. Therefore, disruption of the OCT4 binding at the *MYCN* locus may serve as an effective therapeutic strategy for *MYCN*-amplified NB.

KEYWORDS

neuroblastoma, MYCN, oct4, open reading frame dominance, CRISPR/dCas9, p53, MDM2, caspase-2

1 Introduction

Neuroblastoma (NB) is the most common extracranial solid tumor in children, accounting for 12%–15% of all cancer-related deaths in children (1–3). At least 40% of all NBs are designated as high-risk tumors and often show *MYCN* amplification (4). Amplification of *MYCN* is observed in 25% of high-risk cases and correlates with poor clinical outcomes in patients with NB (5, 6). *Th-MYCN* mice, which are used as a preclinical *in vivo* model of NB, spontaneously develop NB, highlighting the significance of *MYCN* as a potent oncogene in the pathogenesis of NB (7). Despite current therapeutic advances, therapeutic strategy for targeting MYCN remains a medical challenge (8). Therefore, new MYCN-targeting therapeutic strategies are required to further improve patient outcomes.

MYCN, a basic helix–loop–helix transcription factor, directly regulates the transcription of genes involved in diverse cellular processes, such as cell growth, apoptosis, and differentiation (4). It directly binds to its own intron 1 region and upregulates its own expression and its *cis*-antisense gene *NCYM* by forming a positive autoregulatory loop in NB cells (9–11). In addition to MYCN, other transcription factors bind to the *MYCN* locus to regulate *MYCN* expression in NB. For example, OCT4, a transcription factor that maintains cancer stemness, is highly expressed in NB, regulates multipotency, and contributes to drug-resistant phenotypes of NB (12–17). In our previous study, we found that OCT4 stimulates *MYCN* transcription by binding to the intron 1 of *MYCN* locus, whereas MYCN stimulates *OCT4* transcription by binding to the *OCT4* promoter region (17). The OCT4-binding sequence in intron 1 of *MYCN* is not present in mice but mostly conserved in other mammals (17). In contrast, the E-box in the MYCN-binding region of the *OCT4* promoter is specific to humans and absent even in chimpanzees (17). Thus, OCT4 and MYCN form a human-specific positive feedback loop in NB (17). This human-specific positive feedback loop contributes to the stemness of *MYCN*-amplified NB by maintaining the expression of stem cell-related genes including *LIN28*, *NANOG*, and *SOX2* (17). Additionally, CCCTC-binding factor (CTCF), an insulator protein that is capable of regulating gene expression, stimulates *MYCN* transcription by binding to the *MYCN* promoter region (18). Previous studies investigated the transcriptional regulation of *MYCN* through knockout/knockdown of upstream transcription factors. However, since the expression level of transcription factors themselves are reduced by this method, the expression of downstream genes other than *MYCN* is also altered, and indirect effects of such changes on *MYCN* expression cannot be ruled out. In addition, overexpression of upstream factors used in previous studies is based on expression levels of transcription factors that are not observed under normal physiological conditions. In particular, it is necessary to reevaluate whether binding of CTCF to the *MYCN* region is essential for the *MYCN* expression because CTCF functions as an insulator and affects chromatin status of the entire genome. Therefore, the significance of binding of these transcription factors on *MYCN* locus for regulation of *MYCN* transcription has remained elusive. Recently, a CRISPR/deactivated cas9 (dCas9) system has been developed to specifically inhibit binding of transcription factors without affecting their intrinsic expression levels (19). We previously reported that blocking the p53-binding site on *MYCN* locus using the CRISPR/dCas9 system upregulates *MYCN*, *NCYM*, and p53 expression, inducing apoptotic cell death accompanied by caspase-2 activation (20). Thus, the p53-mediated repression of MYCN/NCYM contributes to the survival of *MYCN*-amplified NB cells (11, 20). However, it remains unclear whether the binding of other transcription factors (OCT4, MYCN, and CTCF) at the *MYCN* locus affects *MYCN* expression and contributes to NB cell survival.

In this study, we evaluated the significance of transcription factors that bind to the *MYCN* locus in NB cells. Our results suggest that the OCT4 binding at the *MYCN* locus plays a crucial role in *MYCN*-amplified NB cell survival.

2 Material and methods

2.1 Cell culture

Human NB cell lines CHP134 and IMR32 were maintained in RPMI-1640 (Nacalai Tesque, Kyoto, Japan) supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, Waltham, MA), 50 U/mL penicillin, and 50 µg/mL streptomycin (Thermo Fisher Scientific, Waltham, MA). Neuroblastoma cell line SK-N-AS was maintained in Dulbecco's Modified Eagle Medium (Sigma-Aldrich, St. Louis, MO) supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, Waltham, MA), 50 U/mL

penicillin, and 50 μg/mL streptomycin (Thermo Fisher Scientific, Waltham, MA).

2.2 Vector construction

To inhibit transcription factor binding at the *MYCN* locus, we designed CRISPR guide RNAs against the MYCN-binding site (9, 10), OCT4-binding site (17), CTCF-binding site A (18), p53-binding site (20), and CTCF-binding site B (data from the UCSC Genome Browser). A CRISPR/dCas9 vector was constructed as follows: pX330A_dCas9-1x2 (Addgene, Watertown, MA; plasmid ID 63596) (21) was treated with BpiI (Thermo Fisher Scientific, Waltham, MA). Thereafter, annealed oligonucleotides (p53-binding site: sense: 5′-CACCGCGCCTGGCTAGCGCTTGCT-3′, antisense: 5′-AAACAGCAAG CGCTAGCCAGGCGC-3′; OCT4-binding site: sense: 5′-CACC AGCAGGGCTTGCAAACCGCC-3′, antisense: 5′-AAACGGCGGTTTGCAAGCCCTGCT-3′; MYCN-binding site: sense: 5′-CACC GGGAGGGGGCATGCAGATGC-3′, antisense: 5′-AAAC GCATCTGCATGCCCCCTCCC-3′; CTCF-A-binding site: sense: 5′-CACC TCTCCGCGA GGTGTCGCCTT-3′, antisense: 5′-AAACAAGGCGACACCT CGCGGAGA-3′; and CTCF-B-binding site: sense: 5′-CACCC CAGCAGGCGGCGATATGCG-3′, antisense: 5′-AAACC GCATATCGCCGCCTGCTGG-3′) were inserted into the digested vector.

2.3 Transfection

Plasmid transfection was performed using the Neon Transfection System (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. We used 2×10^5 cells and 4 μg of the plasmid per transfection. When performing the CUT&RUN assay and RNA isolation for quantitative real-time reverse transcription-polymerase chain reaction (qRT-PCR), plasmid transfections were performed using Lipofectamine 3000 transfection reagent (Invitrogen, Carlsbad, CA), according to the manufacturer's instructions.

2.4 WST assay

Cell proliferation was evaluated using the Cell Counting Kit-8 (CCK-8; Dojindo Laboratories, Kumamoto, Japan), according to the manufacturer's protocol. Briefly, 100 μL of dCas9-transfected cell suspension (5,000 cells/well) was seeded in a 96-well plate. Ninety-six hours after transfection of CRISPR/dCas9, 10 μL of CCK-8 reagent was added into each well of the 96-well plate, and then, the cells were incubated for 2 h at 37°C in a 5% CO_2 incubator. Cell proliferation was monitored at 450 nm using CORONA absorbance microplate reader (MTP-310, CORONA ELECTRIC, Ibaraki, Japan).

2.5 Cytotoxicity assay

To evaluate cell damage, we measured lactate dehydrogenase (LDH) activity released from cells. LDH activity was measured using the LDH Cytotoxicity Assay Kit (Nacalai Tesque, Kyoto, Japan), according to the manufacturer's instructions. Briefly, 100 μL of dCas9-transfected cell suspension (10,000 cells/well) was seeded in a 96-well plate. Ninety-six hours after transfection of CRISPR/dCas9, 100 μL of the substrate solution was added into each well of the 96-well plate. After which, the cells were incubated for 20 min at room temperature under shading condition, and then, 50 μL of the stop solution was added into each well of the 96-well plate. LDH activity was monitored at 490 nm using 2030 ARVO X (PerkinElmer, Kanagawa, Japan).

2.6 CUT&RUN assay

Twenty-four hours after the transfection of CRISPR/dCas9, CUT&RUN (CUT&RUN Assay Kit, #86652, Cell Signaling Technology, Danvers, MA) was performed according to the manufacturer's instructions. For each reaction, 1×10^5 cells were used, and the cells were bound to concanavalin A beads and permeabilized with a digitonin-containing buffer. Antibodies were then added and incubated at 4°C for 2 h. The following antibodies were used in the assay: anti-OCT4 antibody (15 μL/assay, #2750; Cell Signaling Technology, Danvers, MA), anti-RNA Pol II antibody (5 μL/assay, #14958; Cell Signaling Technology, Danvers, MA), anti-RNA PolII C-terminal domain (CTD) phospho Ser2 (Pol II pSer2) antibody (5 μL/assay, # 13499; Cell Signaling Technology, Danvers, MA), anti-RNA PolIICTD phospho Ser5 (Pol II pSer5) antibody (5 μL/assay, # 13523; Cell Signaling Technology, Danvers, MA), and Rabbit (DA1E) mAb IgG XP® Isotype Control (#66362; Cell Signaling Technology, Danvers, MA). As an isotype control for the anti-OCT4 antibody, 15 μL of the anti-IgG antibody was applied per assay. In the case of other antibodies, 5 μL of the anti-IgG antibody was applied per assay. DNA obtained from the CUT&RUN assay was amplified using SYBR Green qRT-PCR with the StepOnePlus™ Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA). The following primer set was used: Primer #1, forward 5′-TCCTGGGAACTGTGTTGGAG-3′ and reverse 5′-CTCG GATGGCTACAGTCTGT -3′; Primer #2, forward 5′-CCCTAATCCTTTTGCAGCCC-3′and reverse 5′-CCGACAGCTCAAACACAGAC-3′; Primer #1 in Figure S4C, forward 5′-TCCTGGGAACTGTGTTGGAG-3′ and reverse 5′-CTCGGATGGCTACAGTCTGT-3′; Primer #2 in Figure S4C, forward 5′-ACTGTAGCCATCCGAGGACA-3′ and reverse 5′-CAAGCCCTGCTCCTTACCTC-3′; Primer #3 in Figure S4C, forward 5′-CTAATATGCCCGGGGGACTG-3′ and reverse 5′-CTCTAGCCAGGATGCCTTCG-3′; Primer #4 in Figure S4C, forward 5′-CCCTAATCCTTTTGCAGCCC-3′ and reverse 5′-CCGACAGCTCAAACACAGAC-3′; Primer #5 in Figure S4C, forward 5′-CGTGCTCGTGAGAGCTAGAA-3′ and reverse 5′-GGCTCCGCAACTTTGGAAAC-3′; Primer #6 in Figure S4C, forward 5′-GTGTCTGTCGGTTGCAGTGT-3′ and reverse 5′-TTAATACCGGGGGTGCTTCC-3′; and Primer #7 in Figure S4C, forward 5′-GGGCATGATCTGCAAGAACC-3′ and reverse 5′-GAAGTCATCTTCGTCCGGGT-3′.The detected DNA levels were normalized by the input signal.

2.7 RNA isolation and qRT-PCR

One day after CRISPR/dCas9 transfection, the total RNA from dCas9-transfected NB cells was isolated using the RNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. cDNA was synthesized using SuperScript II with random primers (Invitrogen, Carlsbad, CA). qRT-PCR was performed using SYBR Green PCR with the StepOnePlus™ Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA). The following primer sets were used: *MYCN*, Primer #3, forward: 5′-TCCATGACAGCGCTAAACGTT-3′ and reverse: 5′-GGAACACACAAGGTGACTTCAACA-3′ and *OCT4*, forward: 5′- GGGTTTTTGGGATTAAGTTCTTC-3′, and reverse: 5′-GCCCCCACCCTTTGTGTT-3′ and *GAPDH*, forward: 5′-GTCTCCTCTGACTTCAACAGCG-3′ and reverse: 5′-ACCACCCTGTTGCTGTAGCCAA-3′. *β-Actin* expression was quantified using the TaqMan real-time PCR assay. The mRNA level of *MYCN* was normalized by *β-Actin* and *GAPDH*.

2.8 Long-read and short-read RNA sequencing

Twenty-four hours after CRISPR/dCas9 transfection, the total RNA from dCas9-transfected NB cells was isolated using the RNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. An Iso-Seq library was prepared as described in the Procedure & Checklist-Iso-Seq Express Template Preparation for Sequel and Sequel II Systems, Version 02, October 2019 (Pacific Biosciences, Menlo Park, CA). Briefly, cDNA was synthesized and amplified using the NEBNext Single Cell/Low Input cDNA Synthesis & Amplification Module (New England Biolabs, Ipswich, MA), Iso-Seq Express Oligo Kit (Pacific Biosciences, Menlo Park, CA), and barcoded primers. The size of the amplified cDNA was selected using ProNex beads (Promega, Madison, WI) under standard conditions. The Iso-Seq library was prepared from the size-selected cDNA using SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA). The Iso-Seq libraries were sequenced on the PacBio Sequel IIe with Sequel ICS v11.0 for 24 h using a single cell of Sequel II SMRT Cell 8M Tray, Sequel II Sequencing Kit 2.0, Sequel II Binding Kit 2.1, and Internal Control 1.0 (Pacific Biosciences, Menlo Park, CA). Circular consensus sequencing (CCS) reads were created using this instrument. An RNA-sequencing (RNA-seq) library was prepared using the NEBNext rRNA Depletion Kit v2 (Human/Mouse/Rat) and the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA). The RNA-seq libraries were sequenced on NextSeq 500 using the NextSeq 500/550 High Output Kit v2.5 (75 cycles) (Illumina, San Diego, CA).

2.9 Bioinformatic analysis

Demultiplexing of CCS reads and removal of cDNA primers were performed using the lima command of SMRT Tools v11.0 (Pacific Biosciences, Menlo Park, CA) with the parameters of Iso-Seq data. Removing of artificial concatemer and reads without polyA tail or with short polyA (less than 20 nt) and trimming of polyA tail were performed using the isoseq3 refine tool with the "–require-polya" parameter. High-quality isoforms were obtained using the isoseq3 cluster with the '–use-qvs' parameter. To collapse the transcripts using the isoseq3 collapse command, the high-quality isoform reads were aligned to the human genome GRCh38 using minimap2 v2.24 (22) with the parameter –preset ISOSEQ. Quality control and filtering of the collapsed transcripts were performed using SQANTI3 (23) with genome annotation (Ensembl GRCh38 release105) to remove 3'-end intrapriming artifact, RT-switching artifact, and low frequency transcript (less than 2 fragments). To identify novel transcripts and remove transcript redundancy in all samples, the filtered transcripts were compared with known transcripts (Ensembl GRCh38 release105) using the TALON v5.0 pipeline (24) with the "–cov 0.95 –identity 0.95 –observed" parameter. Transcript reference sequences, including novel and known transcript sequences, were created using SQANTI3 and used in the following short-read RNA-seq analysis. Salmon v1.9.0 was used to quantify transcript expression levels with the "fldMean 260 –fldSD 73" parameter. Differentially expressed transcripts were analyzed using the high-throughput gene expression data analysis tool DIANE (https://diane.bpmp.inrae.fr/) (25). Differentially expressed transcripts were filtered by setting the $\log_2$ fold change (sgRNA OCT4/no sgRNA) to 0.58 and false discovery rate (FDR) to 0.05 as threshold values.

2.10 Functional annotation analysis

DAVID (https://www.david.ncifcrf.gov) (26) was used to identify the enriched molecular functions and pathways related to the genes of interest. *Q*-values (*P*-values adjusted for FDR) were calculated using the Benjamini–Hochberg method in DAVID.

Enrichr (http://amp.pharm.mssm.edu/Enrichr/) (27–29) was used to analyze the enriched molecular functions and pathways related to the differentially downregulated genes after OCT4-binding inhibition. "ENCODE and ChEA Consensus TFs from ChIP-X," "TF Perturbations Followed by Expression," and "ENCODE TF ChIP-seq 2015" were used as gene-set libraries. *Q*-values (*P*-values adjusted for FDR) were calculated using the Benjamini–Hochberg method in Enrichr.

2.11 Kaplan–Meier analysis-based prognosis classification of transcripts

Differentially downregulated genes detected in the RNA-seq analysis of CHP134 and IMR32 were input into the R2 Genomics Analysis and Visualization Platform (http://r2.amc.nl, Tumor Neuroblastoma - Kocak - 649 -custom - ag44kcwolf, GSE45547) for Kaplan–Meier analysis to extract a set of genes associated with a poor NB prognosis. For the type of survival, overall survival was selected. *Q*-values (*P*-values adjusted for FDR) were calculated using the Benjamini–Hochberg method in R2. We found 734 genes whose high expression was associated with a poor NB

prognosis. The genes were classified into the "high is worse" group. We also identified 622 genes that were not associated with NB prognosis. The genes were classified into the "none" group. Subsequently, ORF dominance of the genes was calculated. Since a gene may potentially have multiple transcript isoforms, each with varying ORF dominance scores, the ORF dominance of a gene was determined by calculating the mean of the ORF dominance scores across all transcript isoforms.

2.12 Western blot analysis

The cells were lysed with RIPA buffer (50 mmol/L Tris-HCl buffer (pH 7.6), 150 mmol/L NaCl, 1(w/v)% Nonidet P40 Substitute, 0.5(w/v)% sodium deoxycholate, protease inhibitor cocktail, and 0.1(w/v)% SDS; # 08714-04, Nacalai Tesque, Kyoto, Japan) and benzonase (Merck Millipore, Billerica, MA) and $MgCl_2$ at final concentrations of 25 U/μL and 2 mM, respectively; incubated at 37°C for 1 h; and centrifuged at 10,000 × *g* for 10 min at 4°C. Thereafter, the supernatant was collected and denatured in SDS sample buffer (125 mM Tris-HCl, pH 6.8, 4% SDS, 10% sucrose, 0.01% BPB, and 10% 2-mercaptoethanol). Cellular proteins were resolved using sodium dodecyl sulfate-polyacrylamide gel electrophoresis before being electroblotted onto polyvinylidene fluoride membranes (#1704156, Bio-Rad Laboratories, Hercules, CA). The membranes were incubated with the following primary antibodies for 60 min at room temperature: anti-Cas9 (1:1000 dilution; #14697S, Cell Signaling Technology, Danvers, MA), anti-MDM2 (1:1000 dilution; OP46, Merck Millipore, Billerica, MA), anti-p53 (1:1000 dilution; #2524, Cell Signaling Technology, Danvers, MA), anti-caspase-2 (1:1000 dilution; sc-5292, Santa Cruz Biotechnology, Dallas, TX), anti-caspase 3 (1:1000 dilution; sc-7148, Santa Cruz Biotechnology, Dallas, TX), and anti-actin (1:1000 dilution; FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan). The membranes were then incubated with horseradish peroxidase-conjugated secondary antibodies (anti-rabbit IgG at 1:5000 dilution or anti-mouse IgG at 1:5000 dilution; both from Cell Signaling Technology, Danvers, MA), and the bound proteins were visualized using a chemiluminescence-based detection kit (ImmunoStar Zeta; ImmunoStar LD, FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan). Chemiluminescence was detected using ImageQuant™ LAS4000 (GE Healthcare, Chicago, IL).

2.13 Abby analysis

The MYCN protein levels were measured using a capillary electrophoretic-based immunoassay (the Abby instrument; ProteinSimple, San Jose, CA), according to the manufacturer's protocol. Briefly, samples were combined with 0.1× sample diluent buffer and 5× fluorescent master mix denaturing buffer to acquire 0.8 μg/μL loading concentration. Subsequently, the samples were denatured for 5 min at 95 °C. The primary antibody used in the present study was MYCN (1:100 dilution; #9405, Cell Signaling Technology, Danvers, MA). The Abby measurement was performed using a 12–230 kDa separation module with 25-min separation at 375 V, 10-min blocking, 30-min primary antibody incubation, and 30-min secondary antibody incubation (DM-001, ProteinSimple, San Jose, CA). RePlex™ Module (RP-001, ProteinSimple, San Jose, CA) was used to detect total proteins. At the end of the run, the chemiluminescent signal was displayed as a virtual blot-like image and electropherogram based on the molecular weight using Compass (ProteinSimple, San Jose, CA).

2.14 Open reading frame dominance score analysis

The transcript sequences detected using long-read RNA-seq analysis were used to calculate ORF dominance, as previously described (30).

2.15 Statistical analysis

Statistical analysis software "R" was used for data analysis. Mann–Whitney *U*-test, Student's *t*-test, and Kruskal–Wallis test were performed as appropriate. A p-value < 0.05 was considered statistically significant.

3 Results

3.1 CRISPR/dCas9 targeting transcription factor-binding sites at the *MYCN* locus reduced the proliferation in *MYCN*-amplified NB cells

Deactivated Cas9 (dCas9) disrupts the binding of transcription factors to specific sites (19). To inhibit transcription factor binding at the *MYCN* locus, we designed CRISPR guide RNAs against the MYCN-binding site (9, 10), OCT4-binding site (17), CTCF-binding site A (18), p53-binding site (20), and CTCF-binding site B (data from the UCSC Genome Browser) (Figure 1A). A previous study has demonstrated that CTCF binds to the upstream region of *MYCN* and promotes its transcription (18). However, using the UCSC Genome Browser, we discovered an additional CTCF-binding site located within the gene body of *MYCN* (Figure S1), whose function has not been investigated in previous study (18). Therefore, we designed CRISPR guide RNAs for both CTCF-binding sites. For convenience, we designated the CTCF-binding site upstream of *MYCN* as CTCF-A and the gene body region as CTCF-B (Figures 1A and S1). We transfected all-in-one CRISPR/dCas9-sgRNA vectors into CHP134 and IMR32 cells, both of which were *MYCN*-amplified NB cells (Figure 1B). MYCN expression in CHP134 and IMR32 cells decreased significantly when targeting the OCT4-binding site with dCas9 (Figure 1C, lane 2 and 8; representative raw data with loading control (total protein level) and MYCN signal are presented in Figure S2). The proliferation of CHP134 and IMR32 cells was significantly reduced by dCas9 targeting the OCT4-binding site, MYCN-binding site, CTCF-A

10.3389/fonc.2024.1237378

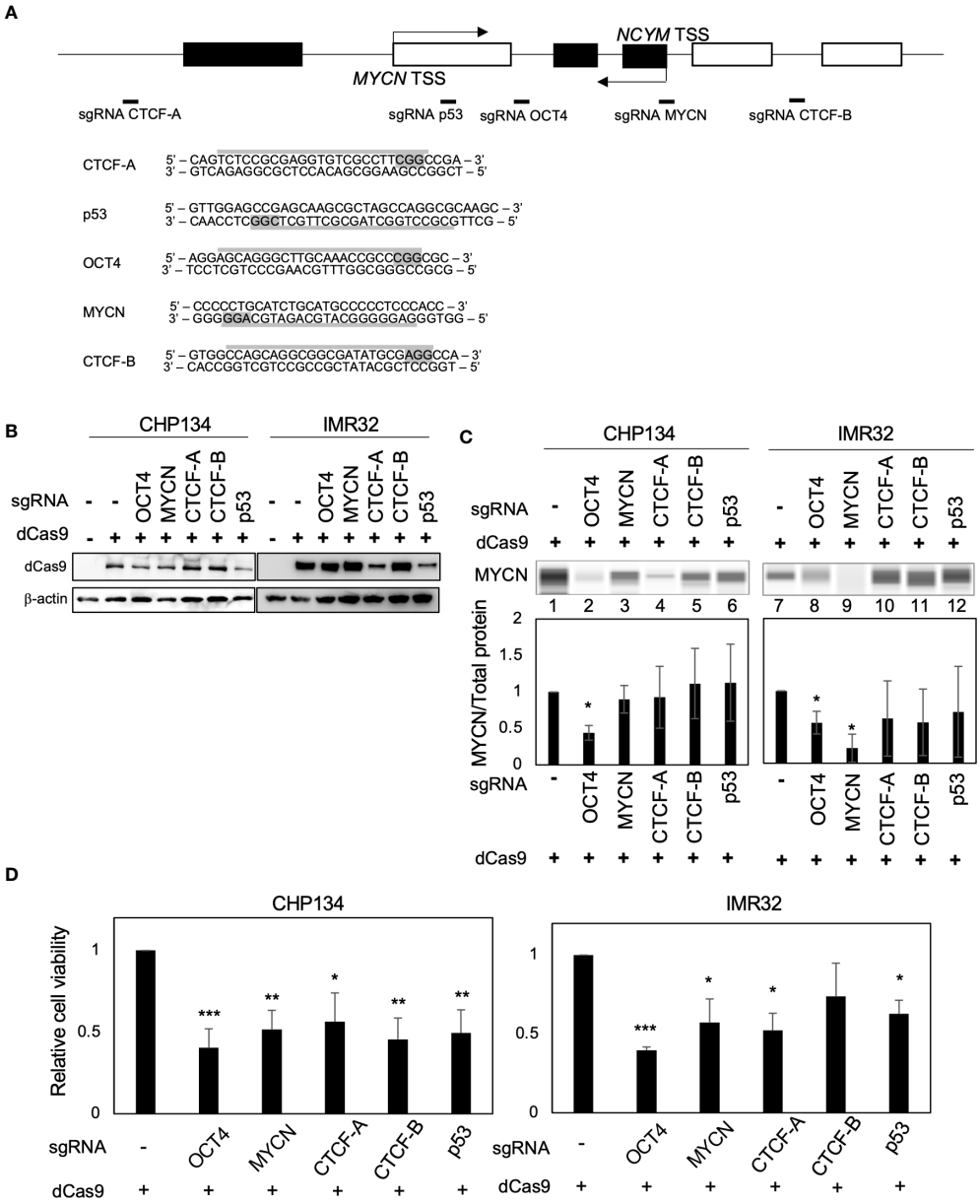


FIGURE 1
CRISPR/dCas9 targeting the *MYCN* locus reduces the proliferation of *MYCN*-amplified neuroblastoma cells. **(A)** A diagram of the MYCN/NCYM locus with the positions of targeting sgRNAs (Upper panel). The white and black boxes indicate the *MYCN* and *NCYM* regions, respectively. The lower panel highlights the targeting and PAM sequences in gray. TSS: transcription start site. **(B)** Western blotting of dCas9 protein in CHP134 and IMR32 cells. Twenty-four hours after CRISPR/dCas9 transfection, these cells were subjected to western blotting. *β-Actin* was used as a loading control. **(C)** Quantitative analysis was performed to measure the MYCN protein level. The MYCN protein level was assessed using the Abby instrument 72 h after transfection. The chemiluminescent signal is displayed as a virtual blot-like image (Upper panel). The quantified MYCN signal, normalized by the total protein level, is represented as a bar graph (Lower panel). Statistical significance (*: $p < 0.05$) was determined using the Student's *t*-test, comparing the results with the those of the no sgRNA control. Error bars represent SEM of three independent experiments. **(D)** Ninety-six hours after CRISPR/dCas9 transfection, the proliferation of CHP134 and IMR32 was measured using the WST assay. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. Data were analyzed using Student's *t*-test (compared with no sgRNA). Error bars represent SEM of six independent experiments.

site, and p53-binding site (Figure 1D). Among these targets, dCas9 targeting the OCT4-binding site was most significantly decreased in the proliferation of both *MYCN*-amplified NBs (CHP134 and IMR32) (Figure 1D). In contrast, the proliferation of the *MYCN*-nonamplified NB cell line SK-N-AS cell line, exhibiting a lower expression of *OCT4 (POU5F1)* mRNA relative to other NB cell lines (Figure S3A), was not affected by dCas9 targeting the OCT4-binding site (Figure S3B).

3.2 Inhibition of OCT4 binding at the *MYCN* locus suppresses *MYCN* mRNA and MYCN activity

As cell proliferation was significantly reduced by dCas9 targeting the OCT4-binding site in both *MYCN*-amplified NB cell lines (Figure 1D), we investigated the effect of inhibition of the OCT4-binding site on MYCN activity. Twenty-four hours after

dCas9 transfection, CRISPR/dCas9 inhibited OCT4 binding at the *MYCN* locus (Figures 2A, B). At this time, the levels of *OCT4* mRNA remained unchanged (Figure S4A), indicating that the decrease in OCT4 binding was not dependent on the expression levels of OCT4. We investigated the potential binding of alternative transcription factors to the OCT4-binding site in *MYCN* locus using the UCSC Genome Browser. However, our findings show that there are no alternative transcription factors found in the same location as the OCT4-binding site in NB (Figure S4B). This indicates that there is no evidence supporting the idea that dCas9 inhibits transcription factors other than OCT4. Additionally, we conducted an investigation into an alternative potential effect of dCas9, specifically its role in interference with transcription elongation (31). We examined the recruitment of RNA polymerase II (RNA pol II) in the proximity of the OCT4-binding site within the *MYCN* locus using the CUT&RUN assay. The results revealed that there was no alteration in both the recruitment of RNA pol II and its phosphorylation status (Figures S4C and S4D), indicating that dCas9 does not interfere with the process of transcription elongation in the proximity of the OCT4-binding site.

The inhibition of OCT4 binding at the *MYCN* locus suppressed the expression of *MYCN* mRNA compared to control (dCas9 without sgRNA) (Figure 2C). A similar trend of *MYCN* mRNA reduction was also observed on transfection of dCas9 targeting MYCN-binding site; however, it did not reach statistical significance (Figure S5). Notably, the reduction in *MYCN* expression resulting from OCT4 inhibition was modest (Figure 2C). The finding seems to contradict the substantial decrease in MYCN protein expression levels observed in Figure 1C. To precisely identify and characterize the transcriptomic changes resulting from OCT4 binding inhibition, we performed short-read RNA-seq combined with long-read RNA-seq of CHP134 and IMR32 cells 24 h after dCas9 transfection. We have listed the detected transcripts and their expression levels in Table S1 (can be accessed on FigShare; https://doi.org/10.6084/m9.figshare.24543067.v1). Through this analysis, we detected 17,601 annotated transcripts (transcript ID starts with ENST~) and 70,753 unannotated transcripts (transcript ID starts with TALONT~) in the combined CHP134 and IMR32 cell samples. Notably, the number of unannotated transcripts was approximately four times higher than the number of annotated transcripts. In the *MYCN* locus, we detected one annotated transcript (Figure 2D (ii)) and four unannotated transcripts (Figure 2D (i), (iii), (iv), and (v)) using long-read RNA-seq analysis. Notably, the expression of ENST00000281043 (Figure 2D (ii)), which encodes the MYCN protein, was suppressed. The result is consistent with the downregulation of MYCN expression in Figure 1C. In contrast, noncoding transcripts of *MYCN* with the highest expression levels showed increased expression (TALONT000261009, Figure 2D (iii)). Because the primer set #3 detected both the coding and noncoding transcripts, the reduction of *MYCN* mRNA appear to be weak in RT-qPCR (Figure 2C). The results suggest that OCT4 regulates promoter usage of *MYCN* gene, and its binding inhibition promotes transcription from the internal promoter (Figure 2D), resulting in reduction of MYCN protein level (Figure 1C). In the *MYCNOS (NCYM)* locus, we detected two annotated transcripts (Figure S6A (viii) and (ix)) and two unannotated transcripts (Figure S6A (vi) and (vii)) using long-read RNA-seq analysis. Among these transcripts, we observed an upregulation in the expression levels of TALONT000260926 (Figure S6B (vii)) and ENST00000419083 (Figure S6B (viii)), both of which have unknown functions. The data provide evidence for the presence of previously unannotated transcripts transcribed from the *MYCN*/*NCYM* locus in NB. Moreover, OCT4 regulates the expression of particular isoforms within the *MYCN*/*NCYM* locus, including the protein-coding isoform of *MYCN*.

In order to examine the potential impact of reducing MYCN expression levels on the downstream pathway of MYCN, we conducted Enrichr analysis (http://amp.pharm.mssm.edu/Enrichr/) (27–29). Enrichr analysis revealed that differentially downregulated genes were enriched in MYCN-target genes (GSE80397: downregulated gene set after *MYCN* knockdown in IMR575) (Figure 2E) and MYC/MAX-target genes (Table S2). On the contrary, in the Enrichr analysis of three independent gene-set libraries (ENCODE and ChEA Consensus TFs from ChIP-X, TF Perturbations Followed by Expression, and ENCODE TF ChIP-seq 2015), enrichment of OCT4-target genes was not observed (Table S2), suggesting no off-target effects of CRISPR/dCas9 on the expression of other OCT4-target genes. These findings indicate that CRISPR/dCas9 specifically inhibited OCT4 binding at the *MYCN* locus and suppressed MYCN activity in *MYCN*-amplified NB.

3.3 Inhibition of OCT4 binding at the *MYCN* locus induces NB cell death accompanied by downregulation of transcripts with high-ORF dominance

We examined how the reduced MYCN activity altered the NB transcriptome. In our previous study, we developed the ORF dominance score, which is defined as the fraction of the longest ORF in the sum of all putative ORF lengths within a transcript sequence (30). This score correlates with translation efficiency of coding transcripts and noncoding RNAs (30). Our previous *in silico*-based analysis suggested that noncoding transcripts with high-ORF dominance are associated with downstream genes of MYCN in humans (30). Therefore, we investigated whether MYCN functions as a regulator of transcripts with high-ORF dominance in NB. We calculated ORF dominance scores of differentially downregulated transcripts using long-read RNA-seq analysis (Table S3). The differentially downregulated transcripts had significantly higher ORF dominance than all transcripts, and this trend was observed for both coding and noncoding RNAs (Figure 3A). Additionally, isoform expression analysis from short-read RNA-seq showed similar results, revealing that the differentially downregulated transcripts had significantly higher ORF dominance in both coding and noncoding transcripts (Figure S7). These findings indicate that MYCN maintains the expression of transcripts with high-ORF dominance in NB. Given that ORF dominance correlates with the translation efficiency of transcripts (30), our results suggest that MYCN maintains the expression of transcripts with high translation efficiency in NB.

MYCN has been reported to globally regulate transcription (32) and splicing (33); however, how this transcription and splicing

10.3389/fonc.2024.1237378

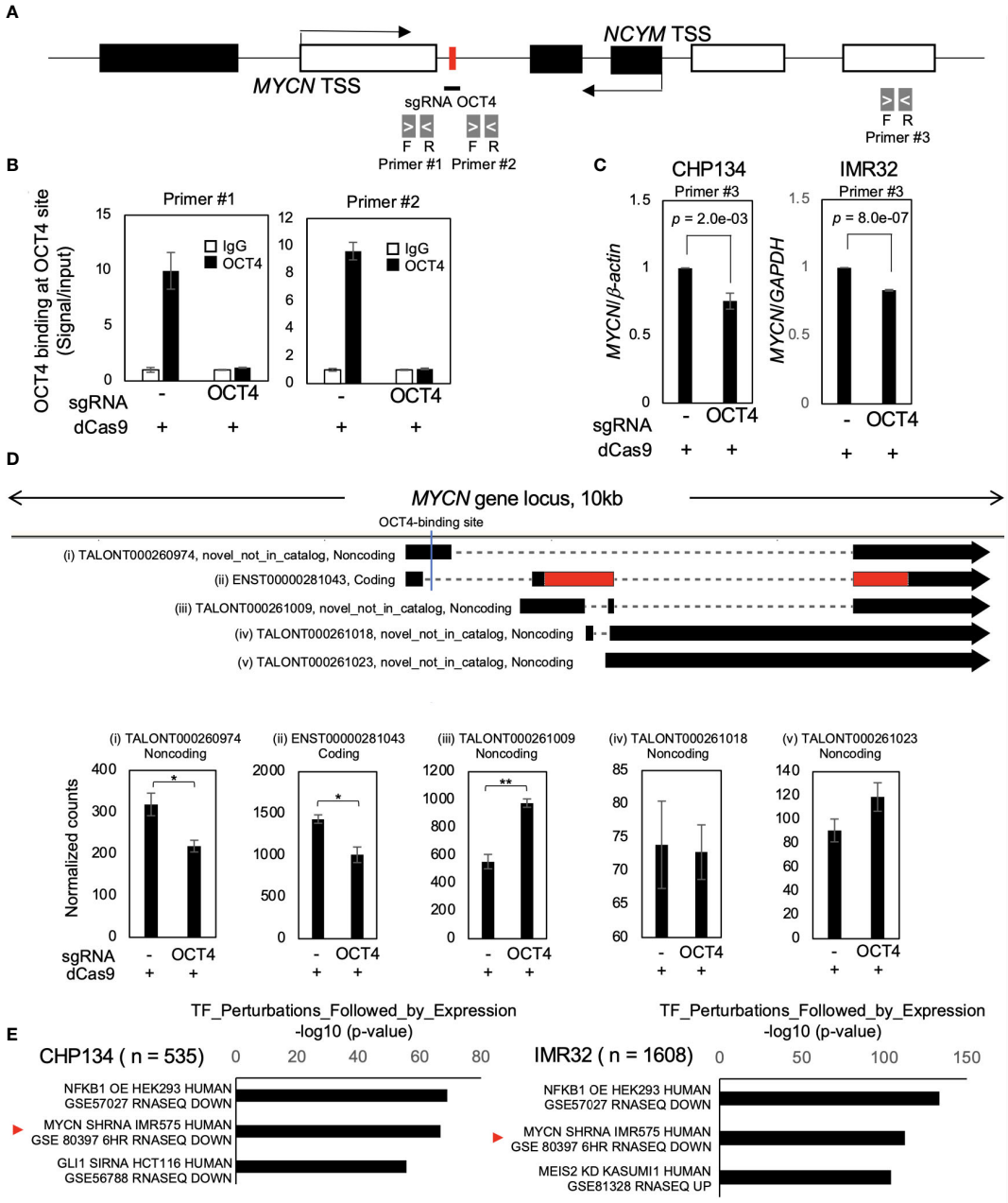


FIGURE 2
Inhibition of OCT4 binding at the *MYCN* locus suppresses *MYCN* mRNA and MYCN activity. **(A)** Schematic depiction of the MYCN/NCYM locus with the location of the primers used in the CUT&RUN assay and quantitative real-time reverse transcription-polymerase chain reaction (qRT-PCR). The OCT4-binding site is indicated with a red box. The white and black boxes indicate the *MYCN* and *NCYM* regions, respectively. The primers employed in the CUT&RUN assay are denoted as primers #1 and #2, whereas that used in qRT-PCR is denoted as primer #3. **(B)** OCT4 binding at the *MYCN* locus was inhibited using CRISPR/dCas9. Twenty-four hours after the transfection of CRISPR/dCas9 targeting the OCT4-binding site, CHP134 cells were subjected to the CUT&RUN assay. Genomic DNA was amplified via qRT-PCR using primers #1and #2 in **(A)**. The signals were normalized by input signals. IgG was used as an isotype control. Error bars represent SEM of three technical replicates. The data presented are representative of the experiment. **(C)** qRT-PCR analyses of *MYCN* in CRISPR/dCas9-transfected CHP134 and IMR32 cells. One day after transfection, *MYCN* mRNA expression levels were measured via qRT-PCR using primer #3 in **(A)** with *β-actin* or *GAPDH* as an internal control. Data were analyzed using Student's *t*-test. Error bars represent SEM of three independent experiments. **(D)** A diagram of transcripts detected at the *MYCN* locus (Upper panel). Black regions indicate *MYCN* transcripts. Red regions indicate coding sequences (CDS). Novel_not_in_catalog means a novel transcript not in the reference produced by a novel splice site. An OCT4-binding site is indicated by a blue line in the diagram. The lower panel demonstrates normalized expression counts (TPM) of *MYCN* transcripts from short-read RNA-seq analysis in CHP134 cells. Error bars represent SEM of three independent experiments. Data were analyzed using the Student's *t*-test. *: $p < 0.05$; **: $p < 0.01$. **(E)** Differentially downregulated genes after inhibition of OCT4 binding at the *MYCN* locus were enriched in MYCN-target genes. Enrichr analysis (http://amp.pharm.mssm.edu/Enrichr/) summary of enriched transcription factor-target genes.

regulations link to the proteome remains elusive. Since ORF dominance correlates with translation efficiency of transcripts (30), our results suggest that these MYCN-induced RNA isoform changes are directional, with increasing potential for efficient translation. Subsequently, we extracted transcripts with high-ORF dominance (ORF dominance > 0.5) from the differentially downregulated transcripts and performed Gene Ontology (GO) analysis using DAVID. Transcripts with high-ORF dominance were associated with the GO terms "mRNA processing," "mRNA splicing via spliceosome," and "RNA splicing" (Figure 3B and Table S4). Notably, the genes encoding the splicing factors *HNRNPA1* and *PTBP1* are the targets of MYCN, and a decrease in MYCN activity induces the downregulation of *HNRNPA1* and *PTBP1* expression and suppresses the proliferation of *MYCN*-amplified NB cells (33). Consistent with this previous report, the expression of *HNRNPA1* and *PTBP1*, the target genes of MYCN, was downregulated after OCT4-binding inhibition in this study (Figure 3C). HNRNPA1 and PTBP1 regulate the alternative splicing of the pyruvate kinase gene (*PKM*) and facilitate the switch from the canonical isoform *PKM1* to the cancer-related isoform *PKM2* (33, 34). The knockdown of *PTBP1*, *HNRNPA1*, and their downstream target *PKM2* represses the proliferation of *MYCN*-amplified NB (33). Similarly, the *PKM2*/*PKM1* ratio was considerably decreased by OCT4-binding inhibition in this study (Figure 3D), suggesting that the splicing switch from *PKM1* to *PKM2* underlies the mechanism of inhibition of NB proliferation after transfection of CRISPR/dCas9 targeting the OCT4-binding site.

We examined the expression of cell death-related proteins using western blotting to gain insights into the mechanism of inhibition of NB proliferation. In our previous study, we found that blocking the p53-binding site at the *MYCN* locus using CRISPR/dCas9 results in the cleavage of caspase-2 and MDM2 and induction of p53 expression (20), which is associated with the p53–MDM2–caspase-2 positive feedback loop (35). Consistent with this report, cleavage of caspase-2 and MDM2, but not caspase-3, and induction of p53 expression were observed in *MYCN*-amplified NB cells at 72 h after transfection of CRISPR/dCas9 (Figure 3E). To evaluate cytotoxicity, the activity of lactate dehydrogenase (LDH) released from cells was measured using a cytotoxicity LDH assay at 96 h after transfection of CRISPR/dCas9. The LDH assay is a method commonly used to evaluate cell damage. LDH is a stable cytoplasmic enzyme that is found in all cells and is released rapidly into the cell culture supernatant when the plasma membrane is damaged (36). The LDH activity was enhanced by CRISPR/dCas9 targeting the OCT4-binding site in CHP134 and IMR32 cells (Figure 3F). This result suggests that inhibition of OCT4 binding at the *MYCN* locus induces apoptosis in NB cells via the activation of the p53–MDM2–caspase-2 positive feedback loop.

3.4 Genes associated with poor NB prognosis encode high-ORF dominance transcripts

Finally, the relationship between NB prognosis and ORF dominance was investigated. For this analysis, the R2 Genomics Analysis and Visualization Platform was used. Differentially downregulated genes were input into the Kaplan–Meier analysis of the R2 database to extract a set of genes associated with poor NB prognosis. We found 734 genes whose high expression is associated with a poor NB prognosis. The genes were classified into the "high is worse" group (Table S5). In addition, 622 genes that were not associated with NB prognosis were identified, which were classified into the "none" group (Table S5). Afterward, ORF dominance of the genes was calculated. Our analysis revealed that coding transcripts in the 'high is worse' group exhibited slightly higher but statistically significant ORF dominance than those in the 'none' group (Figure 4A). Additionally, in the noncoding transcripts, the 'high is worse' group showed higher ORF dominance than those in the 'none' group (Figure 4A). Here, an example of a gene from "high is worse" group is presented. *LSM4*, a MYCN-target gene (37), encodes a member of the LSm family of RNA-binding proteins. LSM4 plays a role in pre-mRNA splicing as a component of the U4/U6-U5 tri-snRNP complex (38). High *LSM4* expression is significantly associated with a poor NB prognosis (Figure 4B). After inhibiting the OCT4 binding at the *MYCN* locus, only the *LSM4* isoform with high-ORF dominance (ENST00000594828, ORF dominance = 0.775) was downregulated, whereas other isoforms remained unchanged (Figure 4B). The results indicate that genes associated with the poor NB prognosis encode high-ORF dominance transcripts, and OCT4 binding inhibition at the *MYCN* locus suppresses these high-ORF dominance transcripts in NB. Because ORF dominance correlates with the translation efficiency of transcripts (30), our results suggest that MYCN maintains the expression of transcripts with high translation efficiency, thereby contributing to a poor prognosis in NB.

4 Discussion

Conventional studies on transcriptional regulation of *MYCN* have largely relied on knockdown/knockout or overexpression of transcription factors, and such methods alter the expression levels of transcription factors, resulting in activation/suppression of downstream target genes other than *MYCN*. Therefore, subsequent alterations in *MYCN* expression by such methods include indirect effects of pathway activation/suppression other than direct effect of transcription factor binding on *MYCN* locus. The CRISPR/dCas9 system was employed to elucidate the significance of transcription factor binding on *MYCN* locus, and found that the inhibition of OCT4 binding was found to be critical for *MYCN* expression in *MYCN*-amplified NB. In our previous study, high *OCT4* mRNA expression was found to be associated with a poor prognosis of *MYCN*-amplified NBs, but not in *MYCN*-non-amplified NBs (17). Consistent with the finding, in the present study, OCT4-binding inhibition in the intron 1 region of *MYCN* decreased the proliferation of *MYCN*-amplified NB cells (CHP134 and IMR32) but not that of *MYCN*-non-amplified NB cells (SK-N-AS), further suggesting that the human-specific OCT4–MYCN network is specifically required for the survival of *MYCN*-amplified NB. Furthermore, long-read sequencing analyses revealed that OCT4 binding on the *MYCN* locus regulates

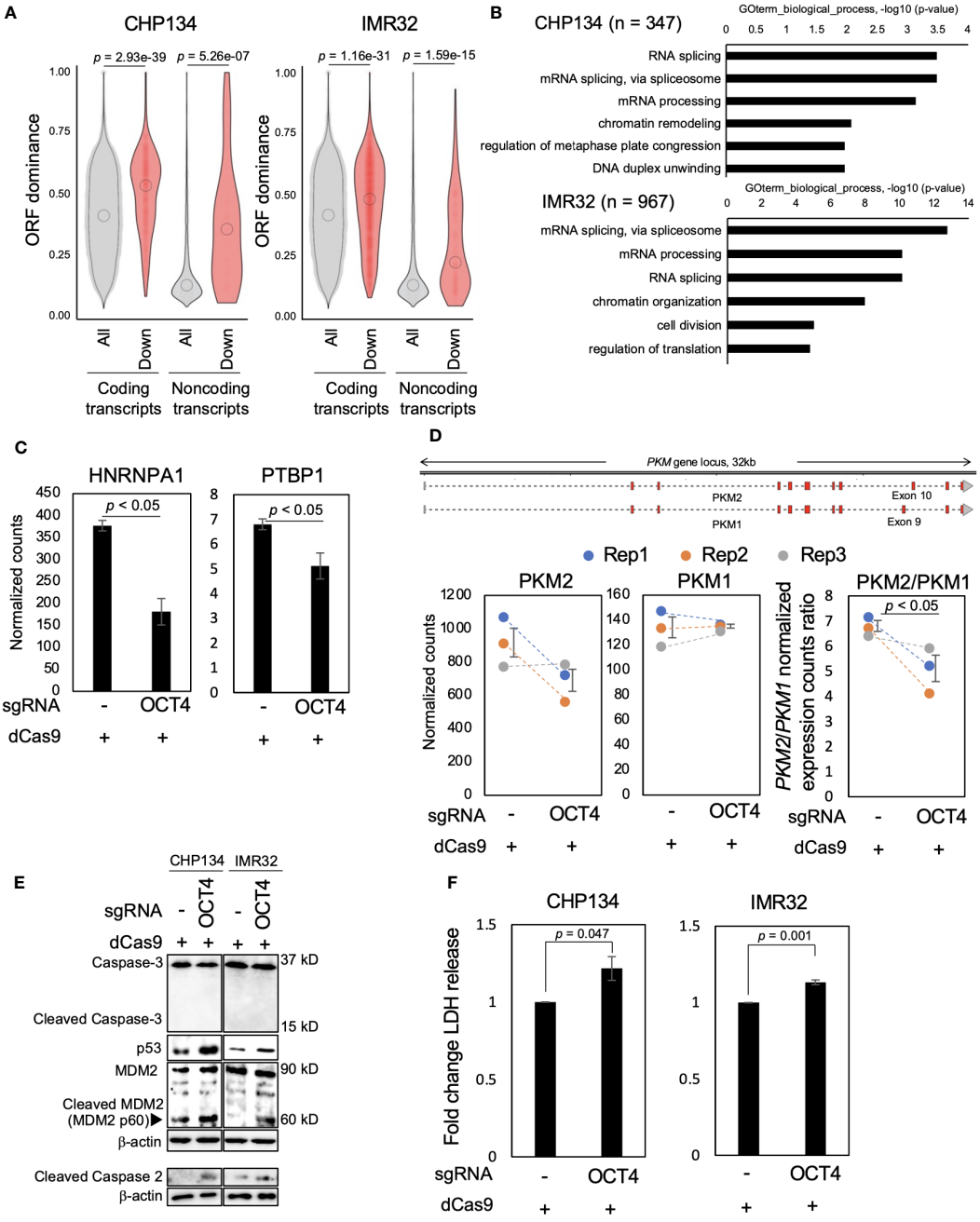


FIGURE 3
Inhibition of OCT4 binding at the *MYCN* locus induces neuroblastoma (NB) cell death accompanied by downregulation of transcripts with high-open reading frame (ORF) dominance. **(A)** Differentially downregulated transcripts were associated with high-ORF dominance in CHP134 (left) and IMR32 cells (right). The number of samples was as follows: coding transcripts (CHP134: all, n = 51,400; down, n = 610, IMR32: all, n = 53,245; down, n = 2,047). Noncoding transcripts (CHP134: all, n = 3,464; down, n = 29, IMR32: all, n = 3,601; down, n = 144). The summary of the data is shown as a violin plot reflecting the data distribution and an open circle indicating the median of the data. *P*-values were calculated using Mann–Whitney *U*-test. **(B)** Gene Ontology (GO) analysis of differentially downregulated transcripts with high-ORF dominance (ORF dominance > 0.5). **(C)** Normalized expression counts (TPM) for *HNRNPA1* and *PTBP1* transcripts from short-read RNA-seq analysis in CHP134 cells. Error bars represent SEM of three independent experiments. Data were analyzed using the Student's *t*-test. **(D)** The CRISPR/dCas9 targeting of the OCT4-binding site led to a reduction in the *PKM2*/*PKM1* ratio. Upper panel: a diagram of transcripts detected using the long-read RNA-seq analysis at the *PKM* locus. Red regions indicate coding sequences (CDS). Lower panel: Normalized expression counts (TPM) for *PKM2* and *PKM1* isoforms and *PKM2/PKM1* normalized expression count ratio from short-read RNA-seq analysis in CHP134 cells. Blue, orange, and gray dots correspond to biological replicates 1, 2, and 3, respectively. Error bars represent SEM of three independent experiments. Data were analyzed using the Student's *t*-test. **(E)** Western blotting of p53, MDM2, caspase-2, and caspase-3 in dCas9-transfected NB cells. Seventy-two hours after transfection, the cells were subjected to western blotting. β-actin was used as the loading control. **(F)** CRISPR/dCas9 targeting the OCT4-binding site induced NB cell death. Ninety-six hours after transfection of CRISPR/dCas9, activity of lactate dehydrogenase (LDH) released from cells was measured using the cytotoxicity LDH assay. Data were analyzed using the Student's *t*-test. Error bars represent SEM of three independent experiments.

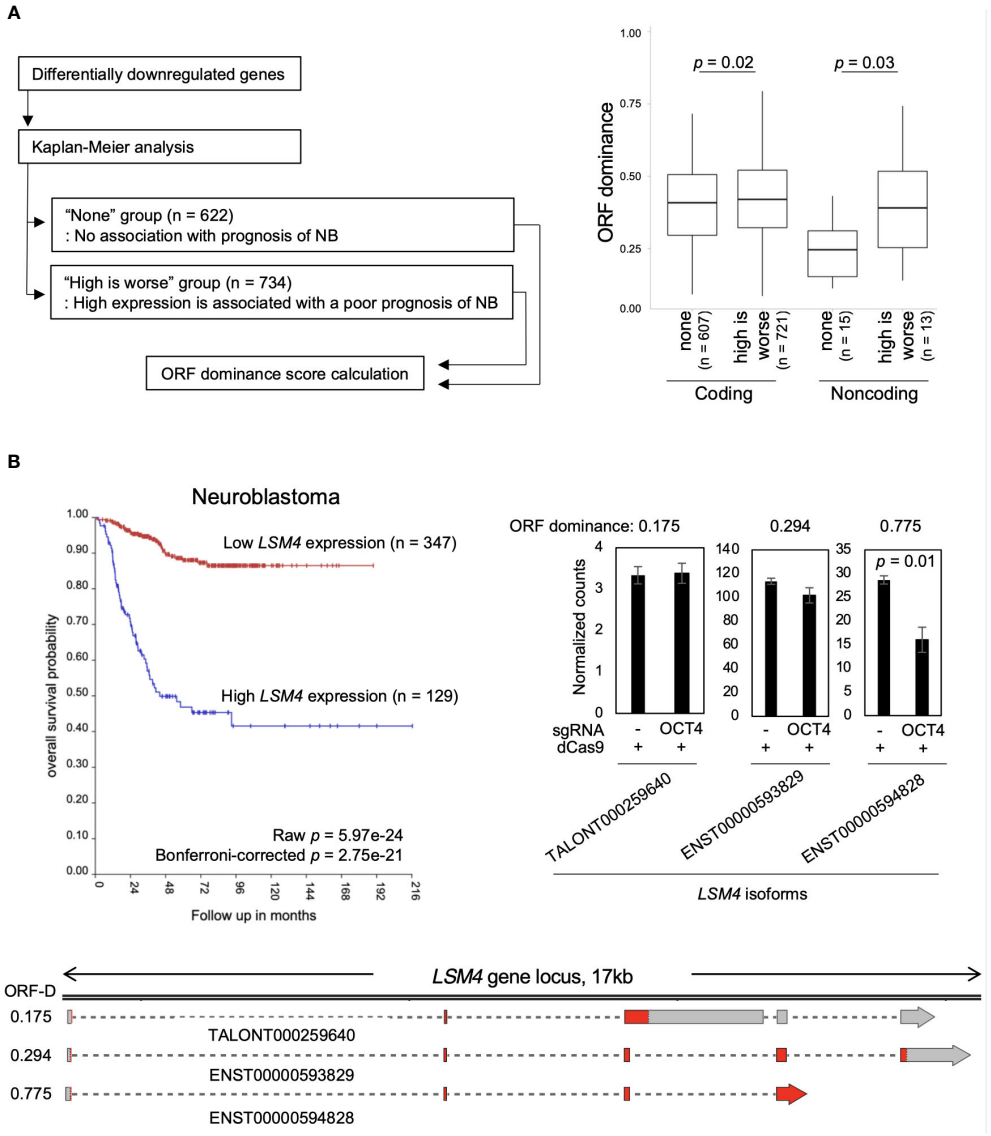


FIGURE 4
Genes associated with the poor neuroblastoma (NB) prognosis encode high-ORF dominance transcripts. **(A)** The "high is worse" transcripts (coding, n = 721; noncoding, n = 13) showed higher ORF dominance than the "none" transcripts (coding, n = 607; noncoding, n = 15). The "high is worse" group contains transcripts whose high expression is associated with a poor NB prognosis. The "none" group transcripts are not associated with NB prognosis. The summary of the data is shown as a boxplot, with the box indicating the IQR, the whiskers showing the range of values that are within 1.5*IQR and a horizontal line indicating the median. *P*-values were calculated using the Mann–Whitney *U*-test. **(B)** High expression of *LSM4* was associated with a poor NB prognosis (Bonferroni-corrected p = 2.75e-21) (Upper left panel). The upper right panel displays normalized expression counts (TPM) of *LSM4* transcripts from short-read RNA-seq analysis in CHP134 cells. Error bars represent SEM of three independent experiments. Data were analyzed using the Student's *t*-test. The lower panel displays a diagram of transcripts detected at the *LSM4* locus in CHP134. Red regions indicate coding sequences (CDS). ORF-D: ORF dominance.

promoter usage in CHP134 and inhibition of the binding resulted in stimulation of transcription of noncoding transcript of *MYCN* from internal promoter. Therefore, combination of the CRISPR/dCas9 system and long-read RNA sequencing clarified the novel mechanism of transcriptional regulation of *MYCN* that could not be revealed by conventional methods.

Inhibition of OCT4 binding at the *MYCN* locus suppressed *MYCN* and its downstream genes, including *HNRNPA1* and *PTBP1*, which are splicing factors. The reduction of *HNRNPA1* and *PTBP1* subsequently decreased splicing activity, leading to a decrease in the *PKM2*/*PKM1* ratio and activation of caspase-2. A previous study by Zhang et al. (33) reported that knockdown of PKM2 suppresses cell proliferation in *MYCN*-amplified NB cells (IMR5), but not in *MYCN*-non-amplified NB cells (SK-N-AS), suggesting a *MYCN*-amplified NB-dependent function for PKM2. Our observation that CRISPR/dCas9 targeting of the OCT4 binding

site suppresses cell proliferation specifically in *MYCN*-amplified NB is therefore consistent with this report. However, the link between PKM2 and caspase-2 remains unclear. One possible explanation is that PKM2 interacts with the CDK1-cyclinB complex to facilitate cell cycle progression in gliomas, and knockdown of PKM2 decreases CDK1 kinase activity (39). Reduced CDK1 activity decreases the inhibitory phosphorylation level of the S340 residue of caspase-2, thereby leading to caspase-2 activation (40). Thus, suppressed *PKM2* expression may activate caspase-2 through the reduction of CDK1-cyclin B kinase activity.

In addition to the OCT4–MYCN network, the inhibition of super enhancer is an interesting approach for targeting MYCN expression in NB. NB is characterized by a core regulatory circuitry (CRC) comprising transcription factors such as PHOX2B, HAND2, and GATA3 that regulate super enhancer (41, 42). The transcription factors form a network with MYCN (43) and are essential for maintaining the cell state in *MYCN*-amplified NB (44). Because blocking of OCT4 or MYCN reduces MYCN expression, simultaneous blocking of OCT4 or MYCN with this CRC, with a focus on targeting transcription factors including HAND2, PHOX2B, and GATA3, may be an effective therapeutic approach for NB.

Previously, we developed the ORF dominance score, defined as the fraction of the longest ORF in the sum of all putative ORF lengths (30). An *in silico*-based analysis suggested that noncoding transcripts with high-ORF dominance are associated with the downstream gene of MYCN in humans (30). However, whether MYCN regulates the expression of transcripts with high-ORF dominance has not yet been experimentally investigated. In this study, we investigated the effect of MYCN activity on the expression of transcripts with high-ORF dominance in *MYCN*-amplified NB cells. Our findings demonstrate that a reduction in MYCN activity led to a decrease in the expression of both coding and noncoding transcripts with high ORF dominance. Because ORF dominance correlates with the translation efficiency of transcripts (30), our results suggest that MYCN maintains the expression of transcripts with high translation efficiency. Importantly, this study identifies MYCN as the first experimentally validated regulator of ORF dominance. However, the mechanism by which MYCN regulates ORF dominance remains unclear. Currently, two potential scenarios are under consideration: (i) MYCN directly transcribes transcripts with high-ORF dominance, or (ii) MYCN maintains expression of transcripts with high-ORF dominance through splicing. In the present study, we observed downregulation of MYCN and its target genes following OCT4 inhibition. As these differentially downregulated transcripts exhibited high ORF dominance, it is plausible that MYCN directly transcribes transcripts with high ORF dominance. However, it has been reported that MYCN transcribes splicing factors (33) and, in the present study, splicing factors *HNRNPA1* and *PTBP1*, which are MYCN-target genes, were downregulated after inhibition of OCT4 binding. As alternative splicing potentially alters the ORF dominance of transcripts, MYCN may maintain the expression of transcripts with high ORF dominance through splicing. Further research is required to elucidate the precise mechanism by which MYCN regulates ORF dominance. Moreover, we found that genes associated with the poor prognosis of NB encode high-ORF dominance transcripts, indicating that ORF dominance may serve as a novel prognostic marker in NB. Coding transcripts in the 'high is worse' group exhibited significantly higher ORF dominance than those in the 'none' group, although this difference was slight. However, the ORF dominance data obtained in this study were based on transcript sequences from cell lines (CHP134 and IMR32), which may not provide the most accurate calculation of ORF dominance. Long-read RNA sequencing data from the large number of clinical samples of NB, which are required for the accurate calculation of ORF dominance, are not currently available. Hence, future studies should investigate whether the ORF dominance score can serve as a prognostic marker in NB using patient-derived transcript data.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

Author contributions

KN, HK, TM, KK, YHa, TSa, TY, and YS performed the experiments and acquired and analyzed the data. KN, KK, YHa, YHi, and YS wrote the manuscript. KN, TSe, YHi, and YS acquired the funds. YS designed and supervised the study. All authors contributed to the article and approved the submitted version.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was partially supported by JST SPRING (Grant Number JPMJSP2109 to KN), e-ASIA Grant from the Japan Agency for Medical Research and Development (NO. JP22jm0210092 to YHi), Grant-in-Aid for Scientific Research (C) JSPS Kakenhi Grant (No. 21K08610 to YS), Grant-in-Aid for Scientific Research (C) JSPS Kakenhi Grant (No. 20K09338 and No. 23K08511 to TSe), and

Takeda Science Foundation (to YS) and the Innovative Medicine CHIBA Doctoral WISE Program (to KN) from Chiba University.

Acknowledgments

We thank Ryo Otaka, Kyoko Takahashi, Miho Kobatake, Taichi Yokoi, and Harumi Saida for their assistance in the study. We would like to thank Editage (www.editage.com) for English language editing.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1237378/full#supplementary-material

References

1. Matthay KK, Maris JM, Schleiermacher G, Nakagawara A, Mackall CL, Diller L, et al. Neuroblastoma. *Nat Rev Dis Primers* (2016) 2:16078. doi: 10.1038/nrdp.2016.78

2. Maris JM. Recent advances in neuroblastoma. *N Engl J Med* (2010) 362:2202–11. doi: 10.1056/NEJMra0804577

3. Brodeur GM. Neuroblastoma: biological insights into a clinical enigma. *Nat Rev Cancer* (2003) 3:203–16. doi: 10.1038/nrc1014

4. Otte J, Dyberg C, Pepich A, Johnsen JI. MYCN function in neuroblastoma development. *Front Oncol* (2020) 10:624079:624079. doi: 10.3389/fonc.2020.624079:624079

5. Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* (1984) 224:1121–4. doi: 10.1126/science.6719137

6. Schwab M, Ellison J, Busch M, Rosenau W, Varmus HE, Bishop JM. Enhanced expression of the human gene N-myc consequent to amplification of DNA may contribute to Malignant progression of neuroblastoma. *Proc Natl Acad Sci U.S.A.* (1984) 81:4940–4. doi: 10.1073/pnas.81.15.4940

7. Weiss WA, Aldape K, Mohapatra G, Feuerstein BG, Bishop JM. Targeted expression of MYCN causes neuroblastoma in transgenic mice. *EMBO J* (1997) 16:2985–95. doi: 10.1093/emboj/16.11.2985

8. Louis CU, Shohet JM. Neuroblastoma: molecular pathogenesis and therapy. *Annu Rev Med* (2015) 66:49–63. doi: 10.1146/annurev-med-011514-023121

9. Suenaga Y, Kaneko Y, Matsumoto D, Hossain MS, Ozaki T, Nakagawara A. Positive auto-regulation of MYCN in human neuroblastoma. *Biochem Biophys Res Commun* (2009) 390:21–6. doi: 10.1016/j.bbrc.2009.09.044

10. Suenaga Y, Islam SMR, Alagu J, Kaneko Y, Kato M, Tanaka Y, et al. NCYM, a cis-antisense gene of MYCN, encodes a *de novo* evolved protein that inhibits GSK3β resulting in the stabilization of MYCN in human neuroblastomas. *PloS Genet* (2014) 10:e1003996. doi: 10.1371/journal.pgen.1003996

11. Suenaga Y, Nakatani K, Nakagawara A. *De novo* evolved gene product NCYM in the pathogenesis and clinical outcome of human neuroblastomas and other cancers. *Jpn J Clin Oncol* (2020) 50:839–46. doi: 10.1093/jjco/hyaa097

12. Pezzolo A, Parodi F, Marimpietri D, Raffaghello L, Cocco C, Pistorio A, et al. Oct-4+/Tenascin C+ neuroblastoma cells serve as progenitors of tumor-derived endothelial cells. *Cell Res* (2011) 21:1470–86. doi: 10.1038/cr.2011.38

13. Hämmerle B, Yañez Y, Palanca S, Cañete A, Burks DJ, Castel V, et al. Targeting neuroblastoma stem cells with retinoic acid and proteasome inhibitor. *PloS One* (2013) 8:e76761. doi: 10.1371/journal.pone.0076761

14. Islam SMR, Suenaga Y, Takatori A, Ueda Y, Kaneko Y, Kawana H, et al. Sendai virus-mediated expression of reprogramming factors promotes plasticity of human neuroblastoma cells. *Cancer Sci* (2015) 106:1351–61. doi: 10.1111/cas.12746

15. Ross RA, Spengler BA, Domènech C, Porubcin M, Rettig WJ, Biedler JL. Human neuroblastoma I-type cells are Malignant neural crest stem cells. *Cell Growth Differ* (1995) 6:449–56.

16. Wei SJ, Nguyen TH, Yang IH, Mook DG, Makena MR, Verlekar D, et al. MYC transcription activation mediated by OCT4 as a mechanism of resistance to 13-cisRA-mediated differentiation in neuroblastoma. *Cell Death Dis* (2020) 11:368. doi: 10.1038/s41419-020-2563-4

17. Kaneko Y, Suenaga Y, Islam SMR, Matsumoto D, Nakamura Y, Ohira M, et al. Functional interplay between MYCN, NCYM, and OCT4 promotes aggressiveness of human neuroblastomas. *Cancer Sci* (2015) 106:840–7. doi: 10.1111/cas.12677

18. Zhao X, Li D, Pu J, Mei H, Yang D, Xiang X, et al. CTCF cooperates with noncoding RNA MYCNOS to promote neuroblastoma progression through facilitating MYCN expression. *Oncogene* (2016) 35:3565–76. doi: 10.1038/onc.2015.422

19. Shariati SA, Dominguez A, Xie S, Wernig M, Qi LS, Skotheim JM. Reversible disruption of specific transcription factor-DNA interactions using CRISPR/Cas9. *Mol Cell* (2019) 74:622–633.e4. doi: 10.1016/j.molcel.2019.04.011

20. Suenaga Y, Yamamoto M, Sakuma T, Sasada M, Fukai F, Ohira M, et al. TAp63 represses transcription of MYCN/NCYM gene and its high levels of expression are associated with favorable outcome in neuroblastoma. *Biochem Biophys Res Commun* (2019) 518:311–8. doi: 10.1016/j.bbrc.2019.08.052

21. Nakagawa Y, Sakuma T, Sakamoto T, Ohmuraya M, Nakagata N, Yamamoto T. Production of knockout mice by DNA microinjection of various CRISPR/Cas9 vectors into freeze-thawed fertilized oocytes. *BMC Biotechnol* (2015) 15:33. doi: 10.1186/s12896-015-0144-x

22. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* (2018) 34:3094–100. doi: 10.1093/bioinformatics/bty191

23. Tardaguila M, de la FuenteL, Marti C, Pereira C, Pardo-Palacios FJ, Risco del H, et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res* (2018) 28:396–411. doi: 10.1101/gr.222976.117

24. Wyman D, Balderrama-Gutierrez G, Reese F, Jiang S, Rahmanian S, Zeng W, et al. A technology-agnostic long-read analysis pipeline for transcriptome discovery and quantification. *bioRxiv* (2019), 672931. doi: 10.1101/672931

25. Cassan O, Lèbre S, Martin A. Inferring and analyzing gene regulatory networks from multi-factorial expression data: a complete and interactive suite. *BMC Genomics* (2021) 22:387. doi: 10.1186/s12864-021-07659-2

26. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using David bioinformatics resources. *Nat Protoc* (2009) 4:44–57. doi: 10.1038/nprot.2008.211

27. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinf* (2013) 14:128. doi: 10.1186/1471-2105-14-128

28. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* (2016) 44:W90–7. doi: 10.1093/nar/gkw377

29. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with Enrichr. *Curr Protoc* (2021) 1:e90. doi: 10.1002/cpz1.90

30. Suenaga Y, Kato M, Nagai M, Nakatani K, Kogashi H, Kobatake M, et al. Open reading frame dominance indicates protein-coding potential of RNAs. *EMBO Rep* (2022) 23:e54321. doi: 10.15252/embr.202154321

31. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* (2013) 152:1173–83. doi: 10.1016/j.cell.2013.02.022

32. Chipumuro E, Marco E, Christensen CL, Kwiatkowski N, Zhang T, Hatheway CM, et al. CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. *Cell* (2014) 159:1126–39. doi: 10.1016/j.cell.2014.10.024

33. Zhang S, Wei JS, Li SQ, Badgett TC, Song YK, Agarwal S, et al. MYCN controls an alternative RNA splicing program in high-risk metastatic neuroblastoma. *Cancer Lett* (2016) 371:214–24. doi: 10.1016/j.canlet.2015.11.045

34. David CJ, Chen M, Assanah M, Canoll P, Manley JL. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* (2010) 463:364–8. doi: 10.1038/nature08697

35. Oliver TG, Meylan E, Chang GP, Xue W, Burke JR, Humpton TJ, et al. Caspase-2-mediated cleavage of Mdm2 creates a p53-induced positive feedback loop. *Mol Cell* (2011) 43:57–71. doi: 10.1016/j.molcel.2011.06.012

36. Kumar P, Nagarajan A, Uchil PD. Analysis of cell viability by the lactate dehydrogenase assay. *Cold Spring Harb Protoc* (2018) 2018. doi: 10.1101/pdb.prot095497

37. Wei JS, Song YK, Durinck S, Chen QR, Cheuk ATC, Tsang P, et al. The MYCN oncogene is a direct target of miR-34a. *Oncogene* (2008) 27:5204–13. doi: 10.1038/onc.2008.154

38. Cooper M, Johnston LH, Beggs JD. Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J* (1995) 14:2066–75. doi: 10.1002/j.1460-2075.1995.tb07198.x

39. Ohba S, Tang Y, Johannessen T-CA, Mukherjee J. PKM2 interacts with the Cdk1-cyclinB complex to facilitate cell cycle progression in gliomas. *Front Oncol* (2022) 12:844861:844861. doi: 10.3389/fonc.2022.844861:844861

40. Andersen JL, Johnson CE, Freel CD, Parrish AB, Day JL, Buchakjian MR, et al. Restraint of apoptosis during mitosis through interdomain phosphorylation of caspase-2. *EMBO J* (2009) 28:3216–27. doi: 10.1038/emboj.2009.253

41. Boeva V, Louis-Brennetot C, Peltier A, Durand S, Pierre-Eugène C, Raynal V, et al. Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat Genet* (2017) 49:1408–13. doi: 10.1038/ng.3921

42. Groningen van T, Koster J, Valentijn LJ, Zwijnenburg DA, Akogul N, Hasselt NE, et al. Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat Genet* (2017) 49:1261–6. doi: 10.1038/ng.3899

43. Xu M, Sun M, Zhang X, Nguyen R, Lei H, Shern JF, et al. HAND2 assists MYCN enhancer invasion to regulate a noradrenergic neuroblastoma phenotype. *Cancer Res* (2023) 83:686–99. doi: 10.1158/0008-5472.CAN-22-2042

44. Durbin AD, Zimmerman MW, Dharia NV, Abraham BJ, Iniguez AB, Weichert-Leahey N, et al. Selective gene dependencies in MYCN-amplified neuroblastoma include the core transcriptional regulatory circuitry. *Nat Genet* (2018) 50:1240–6. doi: 10.1038/s41588-018-0191-z

Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact