Phylogenetic analysis of pediatric tumors

Andersson, Natalie

2024

Document Version:
Publisher's PDF, also known as Version of record

Phylogenetic analysis of pediatric tumors

NATALIE ANDERSSON
DEPARTMENT OF LABORATORY MEDICINE | FACULTY OF MEDICINE | LUND UNIVERSITY

Phylogenetic analysis of pediatric tumors

Natalie Andersson

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University to be publicly defended on Thursday the 23rd of May 2024 at 13:00 in Fernströmsalen, Faculty of Medicine, Forum Medicum, 22362, Lund.

Faculty opponent
Alexander Anderson, PhD, Moffitt Cancer Center, Tampa, Florida, USA

Organization LUND UNIVERSITY Faculty of Medicine Department of Laboratory Medicine Division of Clinical Genetics Author: Natalie Andersson	**Document name:** Doctoral Dissertation **Date of issue:** 23rd of may 2024 Sponsoring organization

Title and subtitle: Phylogenetic analysis of pediatric tumors

Abstract: The main cause of death in pediatric cancer patients is the development of a metastatic treatment resistant relapse. To cure the children that today die of their disease, novel treatment strategies are needed. Since tumors consist of several subclones, whose prevalences change in time and space and in response to treatment, this should be considered when designing such strategies.

The aim of this thesis was to develop bioinformatical tools to study the relationship between populations of cancer cells in pediatric tumors. Subsequently, our goal was to use these tools to elucidate how cancer cell populations after treatment are related to those before treatment and how metastases relate to both the primary tumor and each other.

In papers I, II and IV, we developed software to investigate the ancestral relationship between subclones of cancer cells. Paper I was focused on evaluating the strength and limitations of our software in systematically delineating intratumoral heterogeneity. In paper II we showed that high-risk neuroblastoma does not exhibit any relapse/resistant specific mutations but is caused by transcriptional changes, although certain chromosomal imbalances (1pq+ and 17q+) was important for neuroblastoma progression independent of treatment. In paper III we showed that neuroblastomas that responded to treatment displayed a strikingly similar evolutionary pattern where subclones dominating before treatment were replaced by subclones from a different most recent common ancestor. Hence, treatment selected for pre-existing, possibly resistant, neuroblastoma cell populations. Extensive phylogenetic branching already at the ancestral states, formed the substrate for the observed pattern. In contrast, tumors that progressed under treatment exhibited linear evolution of subclones.

Paper IV focused on diffuse anaplastic Wilms tumor, where we found that *TP53* mutations initiate an extensive clonal evolution, augmenting intratumoral heterogeneity. These tumors tend to form compartments with clonal populations of cancer cells, stressing the importance to sample different such compartments before initiation of targeted treatment.

In paper V we elucidated how different metastases in the same patient are related to each other and the primary tumor. Our findings suggest that metastases may arise both early, late as well as several times during tumor evolution and that several subclones in the tumor can harbor metastatic capacity. Importantly, intermetastatic spread, where metastases give rise to new metastases, was identified as a common feature across several pediatric tumor types.

This thesis sheds light on the evolutionary dynamics of pediatric tumors and opens up the possibilities to develop novel treatment strategies taking evolutionary dynamics into account.

Key words: Phylogenetics, neuroblastoma, Wilms tumor, Rhabdomyosarcoma, Germ cell tumors, Cancer cell evolution.

Classification system and/or index terms (if any)

Supplementary bibliographical information	**Language:** English	
ISSN and key title: 1652-8220	**ISBN:** 978-91-8021-539-8	
Recipient's notes	**Number of pages:** 190	**Price**
	Security classification	

Signature Date 2024-04-09

Phylogenetic analysis of pediatric tumors

Natalie Andersson

LUND
UNIVERSITY

Front cover art by Natalie Andersson
Back cover photos by Fredrik Bertilsson

Faculty of Medicine
Department of laboratory medicine
Division of clinical genetics

ISBN 978-91-8021-539-8
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2024

Till mamma och pappa

Table of Contents

List of Papers

This thesis is based on the following papers:

Paper I

Andersson N, Chattopadhyay S, Valind A, Karlsson J, Gisselsson D, DEVOLUTION-A method for phylogenetic reconstruction of aneuploid cancers based on multiregional genotyping data. **Communications Biology** 2021 Sep 20;4(1):1103. doi: 10.1038/s42003-021-02637-6.

Paper II

Mañas A, Aaltonen K, Andersson N, Hansson K, Adamska A, Seger A, Yasui H, Van Den Bos H, Radke K, Esfandyari J, Satish Bhave M, Karlsson J, Spierings D, Foijer F, Gisselsson D, Bexell D. *Clinically relevant treatment of PDX models reveals patterns of neuroblastoma chemoresistance*. **Science Advances** 2022 Oct 28;8(43):eabq4617. doi: 10.1126/sciadv.abq4617

Paper III

Karlsson J, Yasui H, Mañas A, Andersson N, Hansson K, Aaltonen K, Jansson C, Durand G, Yang M, Chattopadhyay S, Paulsson K, Spierings D, Foijer F, Valind A, Bexell D, Gisselsson D. *Early evolutionary branching across spatial domains predisposes to clonal replacement under chemotherapy in neuroblastoma*. **In revision.**

Paper IV

Rastegar B, Andersson N, Petersson A, Karlsson J, Chattopadhyay S, Valind A, Jansson C, Durand G, Romerius P, Jirström K, Holmquist Mengelbier L, Gisselsson D. *Resolving the Pathogenesis of Anaplastic Wilms Tumors through Spatial Mapping of Cancer Cell Evolution.* **Clinical Cancer Research** 2023 Jul 14;29(14):2668-2677. doi: 10.1158/1078-0432.CCR-23-0311

Paper V

Andersson N, Ferro M, Jansson C, Chattopadhyay S, Karlsson J, Gisselsson D. *Elucidating the metastatic routes of pediatric tumors*. **Manuscript in preparation.**

Papers not included in this thesis:

Andersson N*, Bakker B*, Karlsson J, Valind A, Holmquist Mengelbier M, CJ Spierings D, Foijer F, Gisselsson D. *Extensive clonal branching shapes the evolutionary history of high-risk pediatric cancers*. *Shared first authors. **Cancer Research** 2020 Apr 1;80(7):1512-1523. doi: 10.1158/0008-5472.CAN-19-3468

Andersson N, H Saba K, Magnusson L, Nilsson J, H Nord K & Gisselsson D, *Inactivation of RB1, CDKN2A and TP53 have distinct effects on genomic stability at side-by-side comparison in karyotypically normal cells*. **Genes, Chromosomes and Cancer** 2023 Feb;62(2):93-100. doi: 10.1002/gcc.23096

Chattopadhyay S, Karlsson J, Valind A, Andersson N, Gisselsson D. *Tracing cancer evolution of aneuploid cancer by multiregional sequencing with CRUST*. **Briefings in Bioinformatics** 2021 Nov 5;22(6):bbab292. doi: 10.1093/bib/bbab292

Petersson A, Andersson N, Olsson Hau S, Eberhard J, Karlsson J, Chattopadhyay S, Valind A, Elebro J, Nodin B, Leandersson K, Gisselsson D and Jirström K. *Branching copy number evolution and parallel immune profiles across the regional tumor space of resected pancreatic cancer*. **Molecular Cancer Research** 2022 May 4; 20(5): 749–761 doi: 10.1158/1541-7786.MCR-21-0986

Dickson E*, Sai Dwijesha A*, Andersson N, Lundh S, Björkqvist M, Petersén Å, Soylu-Kucharz R, *Microarray profiling of hypothalamic gene expression changes in Huntington's disease mouse models*. *Shared first authors. **Frontiers in Neuroscience – Special issue on Huntington disease** 2022 Nov 3;16:1027269. doi: 10.3389/fnins.2022.1027269

Chattopadhyay S, Karlsson J, Ferro M, Mañas A, Kanzaki R, Fredlund E, J. Murphy A, L. Morton C, Andersson N, A. Woolard M, Hansson K, Katarzyna R, M. Davidhoff A, Mohlin S, Pietras K, Bexell D, Gisselsson D. *Evolutionary unpredictability in cancer model systems*. **Submitted**.

Andersson N*, Chattopadhyay S*, Karlsson J, Gisselsson D. *An exhaustive search algorithm identifying all possible evolutionary solutions along with reliability quantification.* *Shared first authors. **Manuscript soon in preparation.**

Abbreviations

aCGH	Array Comparative Genomic Hybridization	MIN	Microsatellite Instability
ADR	Adrenergic	ML	Maximum Likelihood
AHSCT	Autologous Homologous Stem Cell Transplantations	MP	Maximum Parsimony
ARMS	Alveolar rhabdomyosarcoma	MRCA	Most Recent Common Ancestor
BAF	B-Allele Frequency	mRNA	Messenger RNA
BAM	Binary Alignment Map	MSF	Mutated Sample Fraction
BB	Branch-and-Bound method	MTD	Maximum Tolerable Dose
BED	Browser Extensible Data	NB	Neuroblastoma
CCR	Collateral Clonal Replacement	NER	Nucleotide Excision Repair
cfDNA	Cell Free DNA	NGS	Next Generation Sequencing
CIN	Chromosomal Instability	NIN	NER-associated Instability
CNA	Copy Number Alteration	NJ	Neighbor Joining
CNAB	Copy Number Aberration Burden	NNI	Nearest Neighbour Interchange method
CNNI	Copy Number Neutral Imbalance	PCR	Polymerase Chain Reaction
CNV	Copy Number Variation	PDX	Patient Derived Xenograft
COJEC	Chemotherapy protocol including cisplatin, vincristine, carboplatin, etoposide, and cyclophosphamide	PMN	Premetastatic Niche
CSC	Cancer Stem Cell	PSR	Phylogenetic Species Richness
CTC	Circulating Tumor Cell	RMS	Rhabdomyosarcoma
ctDNA	Circulating Tumor DNA	RNAseq	RNA sequencing
DM	Double Minute	SAC	Spindle Assembly Checkpoint
DSB	Double Strand Break	SAM	Sequence Alignment Map
eccDNA	Extrachromosomal circular DNA	ScWGS	Single cell Whole Genome Sequencing
ecDNA	Extrachromosomal DNA	SKY	Spectral Karyotyping
ECM	Extracellular Matrix	SMT	Somatic Mutation Theory
EpSSG	European paediatric Soft Tissue Sarcoma Study Group	SNP	Single Nucleotide Polymorphism

ERMS	Embryonal rhabdomyosarcoma	SNP-array	Single Nucleotide Polymorphisms array
EV	Extracellular Vesicle	SNV	Single Nucleotide Variation
FACS	Fluorescence-Activated Cell Sorting	SPR	Subtree Pruning and Regrafting method
FF	Fresh Frozen	TBR	Tree Bisections and Recombination
FFPE	Formalin-Fixed Paraffin-Embedded	TCF	Tumor Cell Fraction
FISH	Fluorescence In Situ Hybridization	TDS	Targeted Deep Sequencing
FNRMS	Fusion negative rhabdomyosarcoma	TME	Tumor Microenvironment
FPRMS	Fusion positive rhabdomyosarcoma	TOFT	Tissue Organization Field Theory
GCT	Germ Cell Tumor	UPGMA	Unweighted Pair Group Method with Arithmetic mean
gDNA	Genomic DNA	VAF	Variant Allele Frequency
GWAS	Genome Wide Association Study	Vcf	Variant Calling Format
HVA	Homovanillic Acid	VMA	Vanillylmandelic Acid
INRG	International Neuroblastoma Risk Group	vWF	Von Willebrand Factor
INSS	International Neuroblastoma Staging System	WAGR	Wilms tumor-aniridia syndrome
IVA	Chemotherapy protocol including ifosfamide, vincristine & actinomycin D	WES	Whole Exome Sequencing
LOH	Loss Of Heterozygosity	WGD	Whole Genome Duplication
MCF	Mutated Clone Fraction	WGS	Whole Genome Sequencing
MCMC	Markov Chain Monte Carlo	WPGMA	Weighted Pair Group Method with Arithmetic mean
MES	Mesenchymal	WT	Wilms Tumor

Populärvetenskaplig sammanfattning

I princip alla tumörer har förändringar i sin arvsmassa (DNA), vilka kallas för *genetiska förändringar*. Det finns flera olika typer av sådana förändringar såsom mutationer eller ett ökat eller minskat antal kopior av en viss del av DNA-kedjan. Dessa kan påverka cellernas funktion på olika sätt, bland annat genom att få cellerna att bete sig annorlunda än vad de hade utan dessa förändringar.

Redan 1976 föreslogs teorin om cancerevolution av forskaren Peter Nowell. Den baseras på antagandet att en tumör börjar med en enda cell som får ett, mer än normalt, antal genetiska förändringar. Detta får den att omvandlas till en tumörcell som börjar dela sig. När en cell delar sig och blir två, kallar man dessa celler för *dotterceller*, vilka i sin tur kan dela sig och vardera ge upphov till två nya dotterceller och så vidare. De genetiska förändringar som cellen har förs vidare till dess dotterceller, vilka dessutom kan få nya genetiska förändringar. På detta sätt uppstår olika populationer av celler i tumören, som har olika genetiska förändringar. Dessa populationer kallas för *subkloner* (**Figur 1A**). Varje subklon har en unik uppsättning genetiska förändringar, varav vissa kan vara skadliga, medan andra kan vara gynnsamma för cellen. Olika subkloner kan på detta sätt utveckla olika egenskaper. Detta resulterar även i ett naturligt urval av cancerceller där vissa celler, med ogynnsamma egenskaper, dör medan andra, med gynnsamma egenskaper, kan börja dela sig oftare. Subklonerna kan dessutom skilja sig åt i hur känsliga de är mot vissa cancerbehandlingar. De kan även utveckla förmågan att sprida sig till andra delar av kroppen, vilket kallas för *metastasering*, vilket ger upphov till *metastaser* d.v.s. sekundära tumörer. Där kan dessa celler fortsätta dela sig och få nya genetiska förändringar.

Denna evolutionära process gör således att uppsättningen av genetiska förändringar skiljer sig åt mellan olika delar av en och samma tumör i en och samma patient, men även mellan *primärtumören* (den huvudsakliga tumören) och metastaserna samt mellan tumörer hos olika patienter. Detta kallas gemensamt för *tumörheterogenitet*. Om man bara tar en enda biopsi från en tumör kan man därför tro att de genetiska förändringar som man identifierar i detta prov finns representerade i hela tumören samt eventuella metastaser, men det är oftast inte fallet då en annan del av tumören kan uppvisa en delvis annan uppsättning genetiska förändringar. Genom att i stället ta biopsier från olika delar av tumören kan man få en bättre bild av hur den är uppbyggd och har utvecklats.

För att visualisera tumörheterogeniteten kan man skapa släktträd, även kallade *fylogenetiska träd*, baserat på vilka genetiska förändringar cellerna har. Liknande släktträd används ofta inom evolutionsbiologin för att avgöra vilka arter som är besläktade med vilka. Genom att skapa liknande fylogenetiska träd för cancerceller kan man få en uppfattning om i vilken ordning de genetiska

förändringarna uppstått samt vilka olika grupper av celler (subkloner) tumören utgörs av. Här kan subklonerna liknas vid arter hos växter och djur. De fylogenetiska träden kan även användas för att avgöra vilka genetiska förändringar som är tidiga (då de hittas i mer eller mindre alla celler) och vilka som är sena (då de endast återfinns i en liten grupp av celler) i tumörens utveckling. Det går dessutom att med denna metod identifiera genetiska förändringar som är unika för metastaser och återfall. Man kan även se vilka celler som överlevt behandlingen, vilka som metastaserat och vilka som utgör ett eventuellt behandlingsresistent återfall.

Det övergripande målet med studierna som ingår i denna avhandling var att utveckla verktyg för att analysera det genetiska släktskapet mellan cancercellspopulationer i barncancerformerna neuroblastom, Wilms tumör, rhabdomyosarkom och könscellstumörer. Därefter skulle dessa metoder användas för att ta fram ny kunskap om återfall, behandlingsresistens och metastasering för dessa tumörgrupper. *Neuroblastom* är en tumör som utgår från det sympatiska nervsystemet, och oftast uppstår i binjuren. Man har två binjurar som befinner sig precis ovanför njurarna och normalt utsöndrar hormoner som kroppen behöver. *Wilms tumör* är en njurtumör. *Rhabdomyosarkom* är en tumör som uppstår i muskler och *könscellstumörer* uppstår oftast i testiklarna eller äggstockarna.

Avhandlingen består av totalt 5 arbeten. I arbete I tog vi fram ett datorprogram (mjukvara) som skapar släktträd baserat på cancerceller från en och samma patient. Vi utvärderade därefter denna med data från 56 barntumörpatienter samt simulerad data. I arbete II var vi intresserade av att utforska varför ett behandlingsresistent återfall uppstår. Detta gjordes genom att tumörceller från patienter med neuroblastom placerades i möss, där sedan tumörcellerna fick växa. Därefter behandlades mössen med cytostatika. Vissa blev botade, en del tumörer svarade på behandlingen i början (tumören minskade i storlek) men började sedan växa igen, och vissa tumörer svarade inte alls på behandlingen, utan fortsatte växa. Cellprov togs från den första tumören samt från alla möss efter behandling och från återfall. Detta möjliggjorde jämförelse mellan de tumörer som svarat och inte svarat på behandling. Vi använde den mjukvara som publicerats som en del av arbete I samt så utvecklade vi även en ytterligare mjukvara som skapar släktträd från individuella cancerceller. Vi kunde visa att det inte fanns några genetiska förändringar som kunde förklara varför en tumör blev resistent mot behandlingen. Däremot var genetiska förändringar såsom en extra kopia av delar av kromosom 1 och 17 viktiga för tumörens utveckling, oberoende av behandlingen. Vi upptäckte i stället att resistens mot behandling verkar orsakas av så kallade *transkriptionella* förändringar. Detta betyder att cellen antingen har ett överuttryck eller underuttryck av specifika gener, vilket påverkar cellens beteende. Behandlingen som mössen fick gynnade celler som ser ut och beter sig som *mesenkymala celler*, vilka har kopplats till behandlingsresistens i tidigare studier. Som ett nästa steg

kunde vi ta ut celler från de tumörer som återfallit i mössen och odla dem i labbet som så kallade 3D *organoider*, vilket innebar att cellerna odlas i små grupper i stället för individuellt. Dessa cellgrupper (organoider) behöll sina genetiska förändringar, sitt mesenkymala uttryck, och behandlingsresistens som de hade i mössen. Detta kommer möjliggöra för oss att testa nya läkemedel och behandlingsstrategier på dessa celler i laboratoriet i framtida studier.

Även arbete III fokuserade på neuroblastom. Vi analyserade flera områden från tumörer från patienter före och efter behandling. Vissa tumörer svarade på behandlingen medan andra fortsatte växa. Vi kunde identifiera att det evolutionära mönster tumörcellerna uppvisade, berodde på hur effektiv behandlingen var. I tumörer som svarade på behandling byttes landskapet av subkloner ut och ersattes av andra subkloner som inte kunde detekteras vid diagnos. I tumörer som inte svarade signifikant på behandlingen, kunde vi i stället se att de cancercellspopulationer som återfanns efter behandling var vidareutvecklingar av de celler som fanns där innan behandlingen.

Detta har stark klinisk signifikans avseende vilka *målinriktade behandlingar* man kan och bör erbjuda patienterna. Målinriktade behandlingar är läkemedel som inriktar sig specifikt på proteinprodukter av mutationer eller andra defekter som gör cancercellerna annorlunda än normala celler. Dessa läkemedel blir allt vanligare inom cancersjukvården. Om de mål som läkemedlen inriktar sig på, inte längre finns efter behandling, kommer läkemedlet dock vara verkningslöst. Om i stället nya potentiella mål har uppstått efter behandling, öppnar detta upp för ytterligare sätt att behandla patienten. I vårt arbete kunde vi observera ett skifte i vilka subkloner som fanns i tumören före jämfört med efter behandling. Detta belyser vikten av att ta nya biopsier efter avslutad behandling, då subkloner med andra mutationer och behandlingsmål än de som hittats vid diagnos, nu kan utgöra större delen av den kvarvarande tumören efter behandling. Således har detta viktig inverkan på vilka målinriktade behandlingar som är aktuella för just den patienten.

I arbete IV undersökte vi njurtumören Wilms tumör. Det finns flera olika grupper av Wilms tumör såsom blastemala tumörer, stromala tumörer och diffust anaplastisk Wilms tumör (DAWT). DAWT är oftast aggressiv och behandlingsresistent. Patienter med denna tumörtyp har följaktligen oftast sämre prognos än patienter med blastemal eller stromal Wilms tumör. I detta arbete tog vi flera cellprov från olika väl annoterade delar av varje tumör för att ta reda på vilka olika subkloner som fanns i olika områden. Då vi visste var provet var taget kunde vi jämföra hur cellerna såg ut i mikroskop med vilka genetiska förändringar som fanns i det området. Vi kunde se att DAWT uppvisade större intratumoral heterogenitet och mer komplexa fylogenetiska träd med många grenar och genetiska förändringar jämfört med de andra formerna av Wilms tumör. Denna komplexitet visade sig initieras av mutationer i en gen som kallas *TP53*, som oftast

uppstod sent och flera gånger i tumörens utveckling. DAWT bildar även små inkapslade områden med cancerceller som oftast bestod av en enda subklon. Olika sådana områden i samma tumör uppvisade subkloner med olika uppsättningar av genetiska förändringar. Även denna studie belyste därmed vikten av att ta cellprov från olika sådana inkapslade områden innan man initierar målinriktade behandlingar. Som en del av detta arbete utvecklades även en tredje mjukvara som kallas för den modifierade maximum parsimony metoden. Denna skapar också släktträd utifrån cancerceller, men försöker även minimera händelser i trädet som är biologiskt osannolika.

I arbete V var vi intresserade av att identifiera släktskapet mellan primärtumör och metastaser inom en och samma patient. Vi tog flera biopsier från primärtumören och från flera metastaser från samma patient. Metastaserna visade sig ofta har en annan uppsättning genetiska förändringar än primärtumören och olika metastaser inom en och samma patient kunde dessutom uppvisa olika mutationer. Genom att skapa fylogenetiska träd kunde vi se vilka cellgrupper i primärtumören som gav upphov till metastaserna samt hur olika metastaser var besläktade med varandra. Vi fann att metastaser kan uppstå såväl tidigt som sent i tumörens utveckling, att flera olika subkloner i primärtumören kan ha förmågan att sprida sig samt att metastasering kan ske i omgångar vid olika tidpunkter under tumörens utveckling. Slutligen kunde vi visa att metastaser kan ge upphov till nya metastaser, oberoende av primärtumören, vilket kallas *intermetastatisk spridning*. Detta har en viktig klinisk inverkan, då det betyder att nya metastaser kan uppstå även om primärtumören tas bort.

Sammanfattningsvis har denna avhandling resulterat i publikation av tre olika mjukvaror. Dessa har därefter använts för att få ökad förståelse för uppkomst av behandlingsresistens och metastasering i barncancer. Resultaten som presenteras i denna avhandling belyser dessutom vikten av utveckling av nya behandlingsstrategier där man tar hänsyn till tumörheterogenitet och evolutionär dynamik i tumörerna.

Popular science summary

Almost all tumors have alterations in their genetic material (DNA), denoted *genetic alterations*. There are several different types genetic alterations, such as mutations or an increased or decreased number of copies of a certain part of the DNA chain. These can affect the cells' functions in different ways by for example causing the cells to behave differently than they would have without these changes.

As early as 1976, the theory of cancer evolution was proposed by the researcher Peter Nowell. The theory is based on the presumption that a tumor begins with a single cell that obtains a, more than normal, number of genetic changes. This causes it to transform into a tumor cell that begins to divide. When a cell divides and becomes two, these cells are called *daughter cells*, which in turn can divide and each give rise to two new daughter cells and so on. The genetic changes that the cell has are passed on to its daughter cells, which can also acquire new genetic alterations. In this way, different populations of cells arise in the tumor, having different genetic alterations. These populations are called *subclones* (**Figure 1A**). Each subclone has a unique set of genetic alterations, some of which may be harmful, while others may be beneficial to the cell. Different subclones can in this way develop different characteristics. This also results in a natural selection of cancer cells where some cells, with unfavorable characteristics, die while others, with favorable characteristics, can begin to divide more often. The subclones can also differ in their sensitivity to certain cancer treatments. They can also develop the ability to spread to other parts of the body, which is called *metastastatic dissemination* resulting in *metastases* i.e., secondary tumors. There, these cells can continue to divide and acquire new genetic alterations.

This evolutionary process thus causes the set of genetic changes to differ between different parts of the same tumor in the same patient, but also between the primary tumor (the main tumor) and the metastases, and between tumors in different patients. This is collectively referred to as *tumor heterogeneity*. If only a single biopsy is taken from a tumor, one might mistakenly think that the genetic changes that are identified in this sample are represented in the entire tumor as well as all metastases (if present), but this is usually not the case as another part of the tumor can show a partially different set of genetic changes. By instead taking biopsies from different parts of the tumor, it is possible to get a better picture of how the tumor is structured and has developed.

To visualize tumor heterogeneity, family trees, also called *phylogenetic trees*, can be created based on the genetic changes the cells have. Similar family trees are often used in evolutionary biology to determine which species are related to one another. By creating similar phylogenetic trees for cancer cells, one can get an idea of the order in which the genetic changes occurred and which different groups of cells (subclones) the tumor is made up of. Here the subclones can be compared to species in plants and animals. The phylogenetic trees make it possible to determine which genetic changes are early (when they are found in more or less all cells), and which are late (when they are only found in a small group of cells) in the development of the tumor. It is also possible with this method to identify genetic changes that are unique to metastases and relapses and identify which cells have survived the treatment, which have metastasized, and which constitute a possible treatment-resistant relapse.

The overall goal of the studies included in this thesis was to develop tools to analyze the genetic relationship between cancer cell populations in the childhood cancers neuroblastoma, Wilms tumor, rhabdomyosarcoma, and germ cell tumors. These methods were then to be used to gain novel knowledge about recurrence, treatment resistance and metastasis for these tumor groups. *Neuroblastoma* is a tumor that originates from the sympathetic nervous system, and usually arises in the adrenal gland. Each person have two adrenal glands that are located just above the kidneys and normally secrete hormones that the body needs. *Wilms tumor* is a kidney tumor. *Rhabdomyosarcoma* is a tumor that occurs in muscles, and *germ cell tumors* most often occur in the testicles or ovaries.

The thesis consists of a total of 5 scientific papers. In paper I, we developed a computer program (software) that creates phylogenetic trees based on cancer cells from one and the same patient. We then evaluated this with data from 56 childhood tumor patients as well as simulated data. In paper II, we were interested in exploring why a treatment-resistant relapse occurs. This was done by placing tumor cells from patients with neuroblastoma into mice, where the tumor cells were allowed to grow. The mice were then treated with chemotherapeutic drugs. Some mice were cured, some tumors responded to treatment at first (the tumor decreased in size) but then started to grow again, and some tumors did not respond to treatment at all but continued to grow. Cell samples were taken from the first tumor as well as from all mice after treatment and from relapses. This allowed comparison between the tumors that responded and the ones that did not respond to treatment. We used the software published as part of paper I, and also developed an additional software that creates phylogenetic trees based on individual cancer cells. We were able to show that there were no genetic changes that could explain why a tumor became resistant to the treatment. In contrast, genetic changes such as an extra copy of parts of chromosomes 1 and 17 were important for tumor development, regardless of treatment. We discovered instead that resistance to treatment appears to be caused by so-called *transcriptional changes*. This means that the cell either has an over expression or under expression of specific genes, which affects the behavior of the cell. The treatment the mice received favored cells that look and behave like *mesenchymal cells*, which have been linked to treatment resistance in previous studies. As a next step, we were able to take cells from the tumors that had relapsed in the mice and grow them in the lab as so-called 3D *organoids*, which means that the cells are grown in small groups instead of individually. These cell groups (organoids) retained their genetic changes, mesenchymal expression, and treatment resistance that they had in the mice. This will enable us to test new drugs and treatment strategies on these cells in the laboratory in future studies.

Paper III also focused on neuroblastoma. We analyzed several areas from tumors from patients before and after treatment. Some tumors responded to treatment while others continued to grow. We were able to identify that the evolutionary

pattern the tumor cells exhibited depended on how effective the treatment was. In tumors that responded to treatment, the landscape of subclones shifted and was replaced by other subclones that were undetectable at diagnosis. In tumors that did not respond significantly to the treatment, we could instead see that the cancer cell populations that were found after treatment were further developments of the cells that were there before the treatment.

This has strong clinical significance regarding which *targeted treatments* can and should be offered to patients. Targeted therapies are drugs that specifically target the protein products of mutations or other defects that make cancer cells different from normal cells. These drugs are becoming increasingly common in cancer healthcare. If the targets that the drugs are designed to seek are no longer present after treatment, the drug will, however, be ineffective. If, instead, new potential targets have emerged after treatment, this opens up additional ways to treat the patient. In our work, we could observe a shift in which subclones were present in the tumor before compared to after treatment. This highlights the importance of taking new biopsies after completion of treatment, as subclones with other mutations and treatment targets than those found at diagnosis, may now make up the majority of the residual tumor after treatment. Thus, this has important impact on which targeted treatments are relevant for that particular patient.

In paper IV, we investigated the renal tumor Wilms tumor. There are several different groups of Wilms tumor such as blastemal tumors, stromal tumors, and diffuse anaplastic Wilms tumor (DAWT). DAWT is usually aggressive and treatment resistant. Consequently, patients with this tumor type have a worse prognosis than patients with blastemal or stromal Wilms tumor. In this work, we took multiple cell samples from different well-annotated parts of each tumor to find out which different subclones were present in different areas. Since we knew where the sample was taken, we could compare how the cells looked under the microscope with what genetic changes were present in that area. We could see that DAWT exhibited a larger intratumoral heterogeneity and more complex phylogenetic trees with many branches and genetic alterations compared to the other types of Wilms tumor. This complexity was found to be initiated by mutations in a gene called *TP53*, which usually occurred late and multiple times in the tumor's development. DAWT also forms small, encapsulated areas of cancer cells that usually consisted of a single subclone. Different such areas in the same tumor exhibited subclones with different sets of genetic alterations. This study thus also highlighted the importance of taking cell samples from various such encapsulated areas before initiating targeted treatments. As part of this work, a third software called the modified maximum parsimony method was developed. This also creates phylogenetic based on cancer cells, but also tries to minimize events in the tree that are biologically improbable.

In paper V, we were interested in identifying the relationship between the primary tumor and metastases within the same patient. We took several biopsies from the

primary tumor and from several metastases from the same patient. The metastases were often found to have a different set of genetic changes than the primary tumor and different metastases within the same patient could also exhibit different mutations. By creating phylogenetic trees, we could see which cell groups in the primary tumor gave rise to the metastases and how different metastases were related to each other. We found that metastases can occur early in the tumor's development, that several different subclones in the primary tumor can have the ability to spread, and that metastasis can occur several times during the tumor's development. Finally, we were able to show that metastases can give rise to new metastases, independently of the primary tumor, which is denoted *intermetastatic spread*. This has important clinical implications, as it means that new metastases can arise even if the primary tumor is removed.

In summary, this thesis has resulted in the publication of three different software. These have subsequently been used to gain an increased understanding of the emergence of treatment resistance and metastasis in childhood cancer. The results presented in this thesis also highlight the importance of developing new treatment strategies that take into account tumor heterogeneity and evolutionary dynamics in the tumors.

Cancer cell evolution

Neoplasias are a heterogeneous group of diseases driven by somatic genetic or phenotypic alterations, often emerging through a cumulative process. The concept of cancer cell evolution was proposed already in 1976 by Peter Nowell. [1] The theory poses that most tumors originate from a single cell that acquires a certain set of alterations causing it to get a fitness advantage compared to other cells in the tissue. As the cell divides, its alterations are conveyed to its daughter cells. With time, some of these cells might acquire new alterations, that in turn are inherited by their daughter cells. In this way, different subpopulations of cancer cells are established in the tissue, denoted *subclones*. Some subclones might acquire the capability to leave the tumor and colonize other anatomic sites, forming metastases. These metastases and the primary tumor subsequently develop in parallel and might thus diverge from one another in their evolutionary trajectories (**Figure 1A**). The landscape of subclones is further shaped by endogenous factors such as geographical barriers, competition, cooperation between cells in the tumor and interactions with immune cells or other constituents of the tumor microenvironment. In addition, exogenous factors such as antitumoral therapy also affects the subclonal composition of the tumor. Some subclones adapt to these selection pressures and continue to proliferate, while others succumb either through cell death or by entering a senescent state. [2]

Due to this continuous subclonal evolution, it follows that each neoplasm should harbour a unique, and heterogeneous, genetic composition due to the semi-random process of mutations, genetic polymorphisms, and differences in microenvironmental selection pressures. Consequently, the same interventions should yield varied results in different patients due to this heterogeneity. Already in the 1970s-80s, studies indeed confirmed the presence of genetic heterogeneity within patients' tumors. [3, 4] Advancements in next generation sequencing (NGS) allowed for a detailed exploration of this phenomenon and it is now a well-established fact that tumors consist of different populations of cancer cells denoted subclones, defined as at least two tumor cells with identical genomic profiles. Their prevalence can vary both spatially within the tumor (**Figure 1B**) and temporally (**Figure 1C**), akin to Darwinian selection of cancer cells. Heterogeneity within a tumor (intratumoral) as well as between tumors within a patient or between patients (intertumoral) has today been identified across almost all tumor types (**Figure 1D**). [5-11]

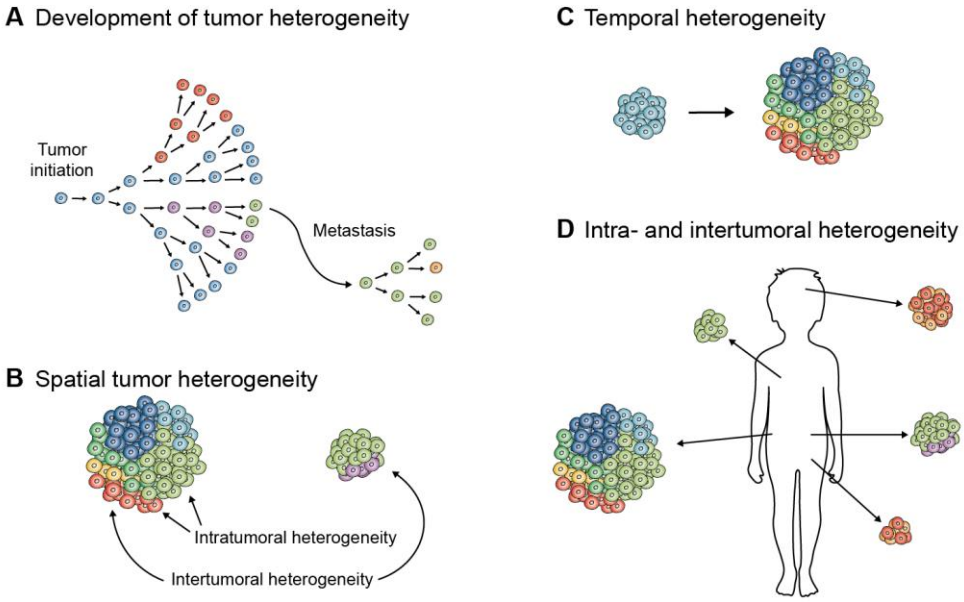


Figure 1. Illustration of the evolutionary processes resulting in intra- and intertumoral heterogeneity. **A)** Initially, a normal cell acquires changes such as genetic alterations causing it to become neoplastic. As the cell divides, these alterations are conveyed to its daughter cells. Over time new alterations are acquired, forming new subclones, indicated here by a change in color. Some cells might acquire the ability to leave the primary tumor site and colonize a distant organ as a metastasis, in which the evolutionary process continues. **B)** This generates genetic heterogeneity both within the tumor (intratumoral heterogeneity) as well as between tumors within the same patient and between patients (intertumoral heterogeneity). **C)** The landscape of subclones also changes over time, resulting in temporal heterogeneity. **D)** Treating a patient harboring a vast intratumoral and intertumoral heterogeneity along with metastatic dissemination, is similar to trying to extinguish a species that is both very heterogeneous as well as geographically dispersed, making the task very difficult.

Cancer can hence be thought of as an evolutionary disease with the tumor initiation resembling a speciation event. [12] Novel treatment strategies should thus preferably take the evolutionary dynamics into account and try to exploit the weaknesses it results in. [13] The most common strategy today is, however, still to use unspecific drugs targeting rapidly dividing cells, denoted chemotherapeutic drugs. This strategy has been in use since the 1950s and has shown significant impact on survival. [14] The drugs are, however, highly unspecific, and often causes a range of side effects by harming healthy cells, that also divide rapidly but are non-malignant. Examples of such are hair follicle cells (causing alopecia), bone marrow cells (causing bone marrow suppression) and intestinal stem cells (causing mucosal wounds, vomiting and diarrhea). [14] Additionally, the treatment strategy is based on the premise that all cells in a tumor are dividing rapidly, which is unlikely. Studies have shown that the proliferative ability of the cells in a tumor is heterogeneous and represents a spectrum with rapidly dividing cells on one end

and cells that do not divide at all on the other. [15] Evidence suggests that some cells even can enter a hibernation state under stress and reawaken when the stress subsides, thus circumventing chemotherapeutic agents and possibly causing a relapse. [16]

Another commonly used method to treat cancer is to surgically remove the tumor, which is a powerful treatment strategy. It is akin to an extinction event where almost all cancer cells are eliminated, only leaving microscopic, small islands of cancer cell survivors scattered across a vast geographical area. Since these islands remain and may regrow, the method is, however, seldom used as the sole treatment for malignant tumors. Unfortunately, the regrowing tumor may also possess a clonal landscape diverging vastly from the one before surgery since the extinction alters the subclonal composition. Surgery itself can also release cancer cells into the bloodstream, causing metastatic spread. The local and systemic inflammatory response after surgery may also contribute to accelerated growth of remaining cancer cells. [17, 18] Radiation therapy can also be used, exploiting the fact that tumor cells often have faulty DNA repair mechanisms. Many tumors are, however, not radiation sensitive and the radiation may also induce additional genetic alterations. [19] None of these treatments thus fully consider the evolutionary dynamics of the disease and they are neither feasible as curative treatment options in a metastatic setting.

Current standard practice often bases the treatment decision on a single biopsy at diagnosis, which may not accurately account for the now well-known heterogeneous subclonal landscape of tumors. This may have a significant impact on the treatment outcome, especially for targeted therapies, which have gained popularity over the past decade due to their ability to be highly cancer specific. With few, but some, important exceptions e.g., Imatinib, trastuzumab, Rituximab and BRAF-inhibitors, targeted therapies have shown surprisingly little effect on long-term survival. [20] When basing the drug choice solely on one biopsy, the suggested target found in this biopsy might be absent in other parts of the tumor. Administering the drug will thus induce a selection pressure for non-target displaying cells. Only the cells having the target will be killed, whereas tumor cells lacking it may continue to grow, now also having access to more space and nutrients, resulting in the tumor rapidly becoming treatment resistant. A similar problem is seen in various other strategies which try to harness the microenvironment via anti-angiogenic drugs and immunotherapies. While these approaches have shown promise in some cases, they suffer from the limitation of only targeting a single molecule and the selection for antigen-negative cells. [21, 22] Sampling several anatomical areas holds the possibility to provide a more comprehensive understanding of the tumour's genetic makeup [6, 23], allowing for improved treatment decisions and patient outcomes, through multidrug targeted therapy.

Despite the significant improvements in survival rates between the 1970s and the 2000s there has been very little progress in the last 20 years for most types of cancer (cancer i siffror 2023, barncancerfonden effektrapport 2018). Despite the increasing interest in cancer cell evolution the last decade, there is still a lack of scientific studies that consider tumor cell evolution to gain knowledge of how resistance and tumor progression develops, especially in pediatric tumors.

In the papers encompassed by this thesis, phylogenetic methods have been used to study tumor cell evolution across several pediatric tumors. The primary aim was to gain increased understanding of the mechanisms underlying the development of treatment resistance, relapse, and metastatic spread. This knowledge might contribute to the development of novel treatment strategies, taking evolutionary dynamics into account, hopefully curing patients that today die of their disease.

The theoretical introduction will commence with an exploration of tumorigenesis, genetic alterations, and their relevance in tracing tumor evolution. This is followed by an introduction to phylogenetics and evolutionary biology. Subsequently, an overview of the pediatric tumor types included in the enclosed studies, will be provided. The introduction concludes with a walkthrough of treatment resistance and the metastatic cascade.

Tumorigenesis

Why does cancer exist?

From an evolutionary perspective, tumor formation is an unavoidable side effect of multicellular organismal life, that has been observed across almost all animal and many plant species. [24] In the latter, tumors primarily arise due to pathogens, such as fungal infections, although instances of spontaneous tumor formation also do occur. Interestingly, since plant cells have a cell wall, they lack motility, and metastases are consequently not seen. [24-26] At the origin of life on earth, all organisms were unicellular, eventually evolving into multicellular colonies and more complex organisms. With multicellular life comes demands and restrictions on the cells making up the organism, necessitating cooperation to enhance the collective nutritional acquisition, like the synergistic dynamics seen in groups of various animal species. Hence, it follows that the cells develop a form of multilevel community dynamics, operating both at the cellular and the organismal level. Cooperation increases the chances of survival and the transmission of genetic material to subsequent generations.

With multicellularity, however, comes drawbacks such as the inherent risk of cancer development. Cancer can be conceptualized as a disruption in the cooperative functioning of the multicellular entity or simply as multicellularity gone wrong [27] and studies have indicated that cancer may be a regression from multicellularity to unicellularity through up- and downregulation of certain cellular pathways (**Figure 2**). [27-29] [30]

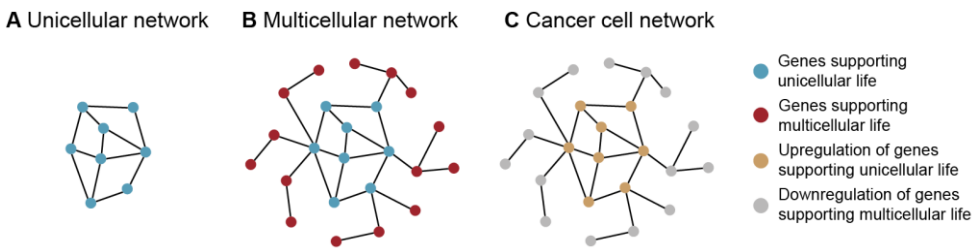


Figure 2. A) A simplified gene network of a unicellular species. **B)** With multicellular life, additional genes are involved, resulting in the development of cooperative behavior. The unicellular network from which it developed is, however, still there. **C)** In cancer, the genes related to multicellular behavior may be downregulated and the ones supporting unicellular life may be upregulated, promoting selfish behavior. Adaptation of Figure 1 in reference. [28]

Despite this proposed regression, cancer paradoxically, in due course, evolves into a distinct multicellular community with its own cooperative networks both between the cancer cells themselves, but also with the immune system and the microenvironment, forming a tumor ecosystem. [31, 32] In this light, cancer could be argued to represent a form of speciation event, representing a natural process of evolution. [12, 33]

Why do we not get more cancer considering the vast number of cells in our bodies?

Considering the millions upon millions of cells in an organism such as the human body, and the continuous cell turnover, it is worth asking why do we not get more cancer than we do? Why is it so rare for a cell to develop into a tumor, such that only one out of millions upon millions of them develop into a tumor? Previously it was thought that the number of cells in an organism should be correlated to the incidence of cancer such that a large animal with many cells should develop cancer more frequently than a small one containing fewer cells. This is, however, not the case. Mice do, for example, get a lot more cancer than elephants, despite their vast differences in size. This lack of correlation between body size, longevity and cancer risk across species is commonly known as *Peto's paradox*. [34] It has been shown that a larger body size and longer life expectancy, is associated with better anticancer mechanisms across mammals. In the elephant it was for example shown that it has 20 copies of the *TP53* tumor suppressor gene. This thus allows for the animal to harness more cells and living longer without the risk of developing cancer. [24, 34]

Cancer can hence be proposed to be rare since we throughout evolution have developed cellular strategies to hinder tumor formation via for example the presence of tumor suppressor genes, which actively inhibit cancer cell behaviour, and through DNA repair systems, apoptosis pathways, redundancy in genetic information, immune surveillance and an evolutionary pressure on the species level favouring traits minimizing cancer. By learning from the anticancer strategies employed by long-living species, we can also gain knowledge of possible treatment strategies for cancer suppression that could potentially be employed in humans. [35]

Tumor initiation

In a healthy individual, each cell has a specific role, and the orchestration of these roles maintains the overall organismal well-being. Cells typically only proliferate in response to external signals and undergo apoptosis if damaged, ensuring a harmonious balance. Tumors form when cells deviate from their designated functions, becoming self-serving entities rather than contributing to the greater good of the organism's cellular community. How tumor initiation occurs has been the subject of debate for centuries, evolving alongside our expanding comprehension of cell biology. Today it is known that a combination of genetic alterations, phenotypic changes and epigenetic changes are likely needed for a cell to develop into a cancer cell and there may be many ways in which tumors can be formed. [36] I will here briefly go through some of the modern theories of tumor initiation (**Table 1**), and will, for reasons that does not have to be explained, exclude older theories such as the balance of the four body fluids which have minimal relevance in contemporary discussions.

The *somatic mutation theory* (SMT) posits that tumor formation is a cell-based disease originating from the accumulation of mutations in somatic cells. These mutations either activate oncogenes or inactivate tumor suppressors, collectively promoting survival, proliferation, and clonal expansion. It assumes that the default state of the cells is quiescence and that the genetic alterations cause the cells to proliferate. The theory has been criticised, mainly by the supporters of the tissue organization field theory (described below) for being too simplistic and the critics additionally argue against the assumption that cells are inherently quiescent. [37] SMT neither explains spontaneous [38] or hormone-driven [39] regression of pediatric and adult cancers, nor the normalization of teratomas injected into blastocysts. [40]

The *tissue organization field theory* (TOFT) challenges the gene-centric perspective by highlighting the importance of tissue organization and communication. It posits that cancer is a tissue-based disease, where carcinogenesis fundamentally is a problem of tissue organization i.e., the maintenance of the normal structure, function, and communication within a tissue. Hence, tumor formation is something that occurs at the tissue level and cannot be reduced to events in individual cells. This may explain the successive transformation from metaplasia, dysplasia to carcinoma and so called pre-cancerous lesions. [41] It assumes that proliferation is the default state of all cells. Since the problem is at the tissue level, the theory also implies that carcinogenesis is a reversible process. [41, 42] TOFT is hence in sharp contrast to the SMT. It does, however, not account for the intratumoral heterogeneity of tumours. [5] Critics also argue that TOFT downplays the role of genetic alterations, which remains crucial factors in understanding the pathogenesis of tumors. [37] Nor does it provide any specific molecular mechanisms explaining the disturbed tissue organization. [43]

In the *classic theory*, also denoted the two-hit hypothesis, a tumor is assumed to arise from genetic alterations in tumor suppressor genes. Knudson proposed the theory in 1971 and claimed that mutations must occur in both alleles of certain genes to initiate tumor formation, as exemplified by the *RB1* tumor suppressor gene in retinoblastoma. This example does, however, constitute a rare exception, and most cancers do not have a particular gene that is always mutated, but the theory does play a role in hereditary cancers. [44-46] The *modified classic theory* is an adaptation of the classical two-hit hypothesis which acknowledges that not all cancers strictly follow a two-hit mechanism. It posits that a combination of several genetic alterations, each of which may have a low selective advantage, in combination with epigenetic changes contributes to tumorigenesis through Darwinian selection of cells harbouring spontaneously occurring mutations. [47-49]

Almost all tumors are aneuploid in some way, [50] laying the foundation of the *aneuploidy theory* positing that abnormal chromosome numbers are the driving force of tumor initiation. It is the misdistribution of chromosomes that is thought to be the cause of cancer rather than mutations. The missegregation of chromosomes could, however, still be a result of mutations and other internal alterations. Several studies have indicated that copy number aberrations both can suppress and promote tumorigenesis. [50-53]

In the last decade the theory of *field cancerization* has gained increased interest. The idea is that tumorigenesis begins long before the lesion is detectable and often also before any morphological changes are seen and is the consequence of the evolution of somatic cells. Somatic cells gain genetic and epigenetic changes that are positively selected for in an otherwise healthy tissue. The interplay between the cells and their microenvironment determines which subclones are selected for. These subclones can grow and produce patches or fields of cells that are predisposed to progress to a neoplasm. [54, 55] The concept was first introduced by Slaughter et al. in the context of oral squamous carcinoma in 1953 [56], but has since then been observed in various cancers including tumors of the oesophagus, skin and lungs. [54]

The *cancer stem cell theory* was originally proposed by Virchow in the 19th century where he claimed that cancer is caused by the activation of dormant embryonic cells present in mature tissues. This theory proposes that a small subset of cells within a tumor, known as cancer stem cells (CSCs), possess the ability to self-renew and initiate tumor growth and are also thought to be more resistant to therapy compared to other cells in the tumor. [57-60] There are several studies where researchers have isolated cancer cells with and without stem cell surface markers, followed by implanting these cells *in vivo*. These two cell groups were shown to possess vastly different abilities to form new tumors, indicating that different cells in the tumor possess different regenerative capabilities. Additionally, it was shown that these stem cells generated tumors having both CSCs but also associated cancer cell populations that were no longer enriched for these markers. [61] Similar

results have been shown for breast cancer [62] and glioblastoma [63] where only a minority of the cancer cells, harbouring stem cell surface markers, were able to generate new tumors *in vivo*. There is some debate regarding the definition of CSCs, which has important implications for how to interpret the results of these studies. Some studies indicate that the state is plastic, and that CSCs can transform into non-CSCs and, interestingly, that both normal cells and non-CSCs tumor cells can spontaneously convert into CSCs *vitro* and *in vivo*. [59, 64] It is, however, not especially surprising that different cell populations have different proliferative potential. Furthermore, it may be that subclones that have a higher proliferative potential also easier form new tumors when implanted *in vivo*. This has not been thoroughly investigated.

The *bad luck theory* founded in 2015 proposes that random mistakes during DNA replication in stem cells (R-mutations) result in the inevitable propagation of mutant clones leading to cancer. It is rather similar to the somatic mutation theory but distinguishes between R-mutations and mutations that are heritable (H-mutations) and those caused by external factors (E-mutations). [65, 66] It suggests that tissues with high cell division rate, regardless of external factors or inherited predispositions, will have a higher risk for tumor formation. The theory has received some criticism since it oversimplifies cancer risk by neither considering the contribution of environmental factors, inherited predispositions nor the impact of the microenvironment. It additionally assumes that cancer risk is dictated entirely by the number of stem cell divisions, which may not be the case. [67] The *ground state theory* on the other hand focuses on the functional state of the cell (its ground state). It suggests that a cell transforms into a cancerous state through alterations in cells that already are in a cancer susceptible state. The susceptibility may vary with age, damage caused by e.g., chemical agents and trauma, and localisation within a tissue. This theory thus acknowledges the importance of epigenetic modifications. [68]

These diverse theories of tumor initiation reflect the complexity of cancer as a biological phenomenon. The ongoing debate regarding how tumor formation is initiated also highlights the need for a comprehensive understanding that integrates genetic, epigenetic, and environmental factors, ultimately guiding the development of more effective and tailored strategies for cancer prevention and treatment.

Table 1. Theories of tumor initiation

Theory	Definition
Somatic mutation theory (SMT) [37]	Cancer is a cell-based disease caused by genetic alterations of the DNA.
Tissue organization field theory (TOFT) [41]	Cancer is a tissue-based disease and cannot be reduced to cellular events in single cells.
Two-hit theory [46, 69]	Both alleles of certain genes must be altered for cancer to develop.
Modified classical theory [47]	A combination of genetic and epigenetic alterations are needed for tumors to develop.
Aneuploidy theory [50]	All tumors possess aneuploidy, and it is this misdistribution of chromosomes that causes tumor initiation.
Field cancerization [55]	Mutant clones in healthy tissues create clonal expansions forming patches of cells, predisposed to cancer development.
Cancer stem cell theory [58]	A small group of cells with stem cell properties initiate and maintain tumor growth. The non-stem cells cannot alone sustain the tumor.
Bad luck theory [68]	Random mistakes during DNA replication in stem cells induce clonal expansions forming tumors.
Ground state theory [68]	The ground state (functional state) of the cell is important. The theory especially acknowledges epigenetic modifications.

Types of genetic alterations

This section will explore the various types of genetic alterations that may occur within a cell, beginning by examining the smallest and progressively move towards the more substantial ones.

Single nucleotide changes

Each DNA molecule consists of two strands, each comprising a long sequence of four different nucleotide bases. These bases - A (adenosine), T (thymine), C (cytosine) or G (guanine) – are accompanied by a deoxyribose sugar and a phosphate group. The two strands bind together through base pairing rules: A pairs with T and C pairs with G, forming a double helix structure. Three consecutive nucleotide bases form a *codon*, that encodes a particular amino acid. With 64 possible combinations of the 4 nucleotide bases in groups of 3, multiple codons can encode the same amino acid. During transcription the sequence is transcribed into messenger RNA (mRNA) that is subsequently translated by the ribosome. The sequence of codons gives instructions to the ribosome regarding which amino acids the corresponding protein is made up of and in which order, allowing for creation of the protein molecule (**Figure 3**).

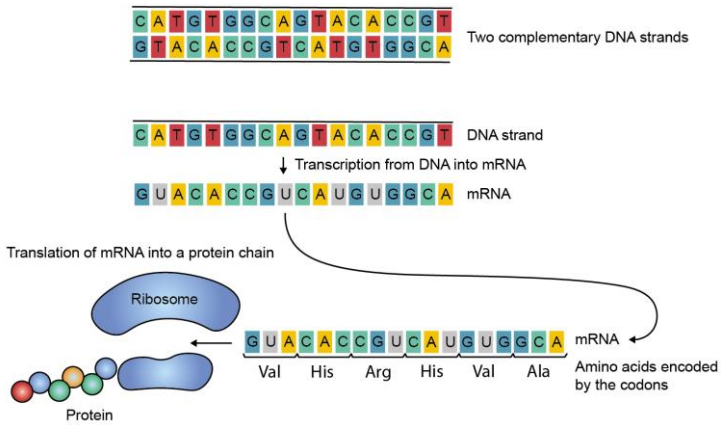


Figure 3. Illustration of the process from DNA to protein. One of the two complementary DNA strands are transcribed into mRNA, which is used by the ribosome to translate the sequence of codons into amino acids, constituting the protein encoded by the corresponding gene.

A *single nucleotide polymorphism* (SNP) is defined as a *germline* substitution of a single nucleotide to another. For it to be called a polymorphism it must be present in a considerable proportion of the population, generally at least 1 %. These alterations might increase the risk of developing certain diseases and result in cancer susceptibility. [70] They are investigated through familial inheritance studies for mendelian traits or genome-wide association studies (GWAS) for polygenic traits [71]

A *single nucleotide variant* (SNV) is a single nucleotide substitution that is not inherited, is less frequent and does not classify as a SNP. Generally, the term is used to refer to point mutations in *somatic* cells, i.e. non-germline cells such as cells in healthy tissues or cancer cells.

SNPs and SNVs (**Figure 5**) can occur in *exons* (protein coding parts of a gene) or *introns* (the sequences that separate a gene's exons). If the substitution occurs in an exon, it can be *synonymous* or *nonsynonymous*. A synonymous mutation does not result in a change in the amino acid sequence. It may, however, still affect the protein's function by slowing down translation, thus allowing the peptide chain to fold into an unusual conformation, making it less functional. It can also affect the splicing. [72] Nonsynonymous mutations on the other hand result in an amino acid change. These can be further divided into either *missense* or *nonsense* mutations. Missense mutations result in an amino acid change in the actual protein chain, which may cause disease. Nonsense mutations result in a premature stop codon and hence a truncated, incomplete, and usually dysfunctional or nonfunctional protein product. Mutations in introns do not change the amino acid sequence of the protein but may result in incorrect intron splicing, which can have deleterious consequences on the protein's structure and function and may play an important role in tumorigenesis. [73]

Several mechanisms contribute to the occurrence of SNPs and SNVs. They might arise through errors during DNA replication where DNA polymerase makes mistakes when reading the sequence, resulting in an incorporation of an incorrect nucleotide. There is, however, a proofreading system and a mismatch repair mechanism, but they are not perfect, and some errors may thus go undetected. Some regions of the genome contain repetitive sequences denoted microsatellites, which are prone to genetic alterations due to replication slippage. Mutations can also arise due to exposure to chemicals, mutagens that modify the structure of the DNA chain, reactive oxygen species or radiation causing direct DNA damage through double strand breaks. This might result in loss of genetic material or that the DNA polymerase misread damaged bases or even skip them. [74]

Recent studies have suggested that spontaneous mutations could stem from a quantum mechanical mechanism. [75] Each base pair in the DNA is held together by two (A-T) or three (C-G) hydrogen bonds. Particles do, however, not have a fixed position but is described by a *wave function*, representing the probability of finding the particle at a certain position if its position would be measured. The nucleotides thus share each proton via hydrogen bonds. In a standard A-T and C-G base pair, the protons are expected to be in their regular form (**Figure 4A**). Through quantum tunnelling there is, however, a small possibility to find the protons closer to the other nucleotide, with which it base pairs, creating a *tautomeric* version of the base pair, as was originally proposed by Watson and Crick in 1953. [76] In their tautomeric form, T can pair with G and A can pair with C, thus violating the classic rules of base pairing (**Figure 4B**). When DNA polymerase reads the DNA sequence it actually performs a quantum measurement. As we know from the double slit experiment, this collapses the wave function, and the proton will now be detected at a specific position. The probability for it to be at a certain position is given by the wave function. With a small, but non zero, probability, this position can be with the base pair in tautomeric form. [75] Consequently, the DNA polymerase might pair the tautomeric nucleotide with the wrong base pair resulting in a SNP or SNV. It also follows that genes that are more frequently replicated and transcribed i.e., quantum measured, should display mutations more frequently, which is also what has been found. [77] Tautomers with alternative proton positions thus seem to be an unavoidable source and driver of mutations, and hence also evolution.

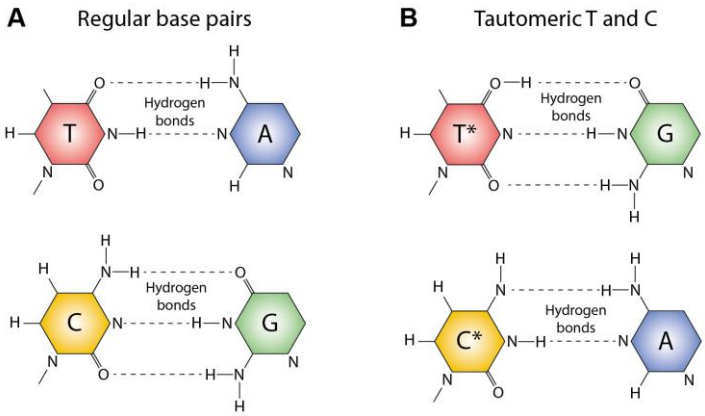


Figure 4. A) Standard, but simplified, T-A and C-G base pairs with protons in their normal positions. **B)** In its tautomeric forms, indicated by a *, the protons are found in the opposite position. Consequently, it is possible for T to base pair with G and C to base pair with A. Also, G and A might be in tautomeric forms, and form base pairs with T and C respectively.

Insertions and deletions

Insertions are incorporations of additional base pairs while *deletions* are removals of base pairs in the DNA sequence (**Figure 5**). They are often denoted together as *indels*. Unless the change is a multiple of 3 it will produce a *frameshift* resulting in a faulty protein chain. Indels have been shown to play an important role in tumorigenesis. [78] If the indel is larger than 50 bases long it is often classified as a structural variant.

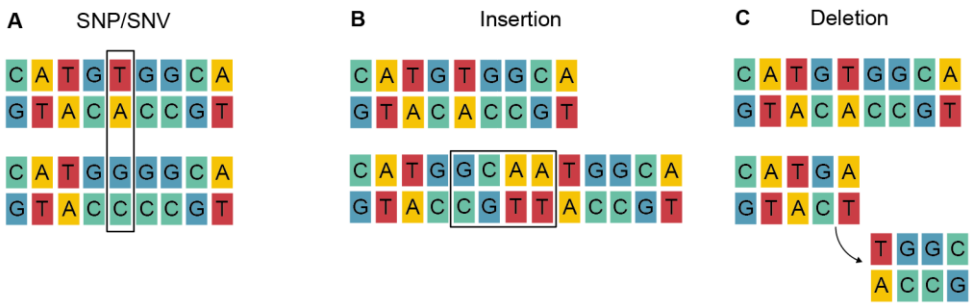


Figure 5. The DNA sequence consists of consecutive pairs of either A and T or C and G. The upper DNA sequence is the reference sequence, and the lower one is the altered sequence. **A)** Ideally the sequence should be identical in all cells and across different individuals. There are sites where the sequence mismatch, constituting a SNP or SNV. **B)** If one or several base pairs are inserted into the DNA-sequence it is denoted an insertion. **C)** If some base pairs are lost from the DNA sequence it is called a deletion.

Chromosomal translocations

A chromosomal translocation refers to the transfer of a segment of a chromosome to another location, either on the same or to another chromosome (**Figure 6**). A *balanced translocation* is an even exchange of material with no genetic information extra or missing. An *unbalanced translocation* on the other hand involves an unequal exchange of chromosomal material, resulting in extra or missing genes. The translocation can also be reciprocal or nonreciprocal. In a *reciprocal translocation* there is an exchange of genetic material between two non-homologous i.e., not identical chromosomes. In a *nonreciprocal translocation* there is a *one-way* transfer of genes from one chromosome to a non-homologous chromosome. Finally, in a *Robertsonian translocation*, two acrocentric chromosomes i.e., chromosomes where one arm is much longer than the other, get attached to each other. Generally, the designation t(A;B) is used to denote a translocation between chromosome A and B.

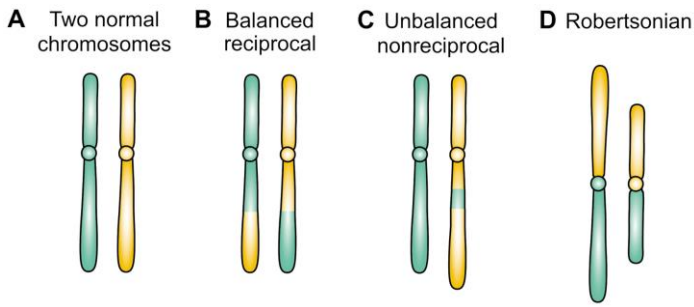


Figure 6. A) Two normal chromosomes. **B)** There has been an even exchange of genetic material between two chromosomes. No material is lost or gained. **C)** There has been an uneven exchange of genetic material between two chromosomes. **D)** A Robertsonian translocation where the q-arms and p-arms, respectively, from two acrocentric chromosomes get attached to each other.

If a translocation joins two genes that are otherwise separated, a *fusion gene* may form. Fusion genes have been shown to have important implications in cancer. The first recurrent rearrangement of this nature was the Philadelphia chromosome in chronic myeloid leukaemia, identified in 1960. [79] It was shown in 1985 that it was due to a translocation between chromosomes 9 and 22, denoted t(9;22), and shown to form the fusion gene *BCR-ABL1*. [80, 81] Since then, several targeted therapies have been created against its protein product. In addition, a long range of other recurrent gene fusion genes and corresponding proteins have been identified across cancers. [82]

Copy number changes

A normal human cell contains 46 chromosomes, consisting of 22 pairs of autosomes and usually one pair of sex chromosomes, encapsulated in the cell

nucleus. Hence, each cell has two copies of each chromosomal segment and is referred to as *diploid*. Each chromosome consists of a single DNA chain wrapped around proteins called *histones*. Several histones together with the DNA wrapped around it forms a small unit denoted a *nucleosome*. The nucleosomes fold up, forming a fiber referred to as *chromatin*. When the cell is not undergoing mitosis, the DNA is kept in this form. Some parts are very tightly packed and is called *heterochromatin*. The genes contained here are not frequently transcribed. On the other hand, there is also lightly packed DNA called *euchromatin*, enriched in genes that are under active transcription. The chromatin can be compressed and folded further, forming condensed *chromatids/chromosomes* during mitosis. Each chromosomal segment is denoted as having the allelic composition 1+1 since there is, normally, one copy of each of the two alleles. Chromosomes are divided into two arms, the p-arm (shorter) and the q-arm (longer) bound together by the *centromere*, to which the mitotic spindles bind during mitosis to separate the two chromosomes. The regions closer to the centromere are referred to as being more proximal than the regions further out towards the telomeres, which are referred to as being distal. To refer to a certain region of a chromosome, the different parts are numbered starting from the centromere and outwards, based on which *chromosome region and band* they belong to. When chromosomes are stained using a dye called Giemsa, heterochromatin stain darker than euchromatin, forming bands. The chromosome area p11 is more proximal than p21 and q11 is more proximal than q21. The first number refers to the chromosome region and the second one the band within the region. In this manner it is possible to identify where a certain alteration is by writing e.g., 17q12 (chromosome 17, on the q-arm, in region 1, band 2).

Copy number variations (CNVs) refer to the normal *germline* variation in the number of copies of a DNA sequence within the population. To be called a variation, it must exist in at least 1 % of the population. An individual may lack e.g., a gene or have additional copies of it. These variations may be associated with certain traits or disorders but can also be neutral. [83] *Copy number alterations* (CNAs), on the other hand, refers to changes in the number of copies of a DNA sequence that occurs de novo within somatic cells. *Aneuploidy* is defined as losses and gains of entire chromosomes. In some definitions losses and gains of chromosome arms are included in this term as well. [50] A *loss* can be *hemizygous* if one copy is lost or *homozygous* if both are lost. If one of the alleles is lost it is also referred to as *loss of heterozygosity* (LOH). If there is a change in the allelic composition, but still resulting in the same number of copies of the segment, it is denoted a *copy number neutral imbalance* (CNNI). A *gain* involves the duplication of a chromosomal segment resulting in additional copies. Losses or gains of parts of a chromosome are denoted *segmental chromosomal aberrations*. The final number of copies of each allele after these losses or gains can be denoted as e.g., 1+0 if one of the alleles is lost or 2+1 if there is an additional copy of one of the alleles (**Figure 7**). CNAs and aneuploid have been observed across almost

all cancer forms. Loosing or gaining a large part of a chromosome may affect a vast number of genes, including oncogenes and tumor suppressors and may have a profound effect on cellular function, causing macro-evolutionary leaps. [48, 50, 53]

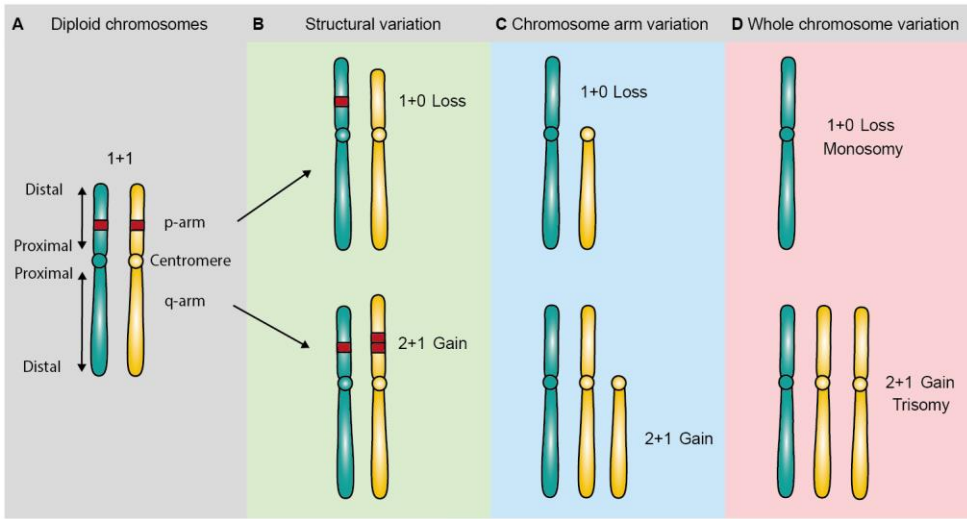


Figure 7. A) Depiction of the different parts of a normal chromosome pair. **B)** Loss or gain of a part of a chromosome arm, denoted a structural variation. **C)** Loss or gain of an entire chromosome arm. **D)** Loss or gain of an entire chromosome, here resulting in one copy of the chromosome (monosomy) or three copies (trisomy), respectively.

There are several proposed mechanisms by which copy number changes can form. During mitosis the telomeres of two chromatids may fuse together. As they are being pulled towards the two poles during anaphase a *chromatin* or *anaphase bridge* can form. Eventually this bridge will break by the pull of the spindle poles, causing a disruption at some random point in the chromosomes resulting in a copy number change. Another mechanism is through *improper mitotic spindle formation*, which may result in the sister chromatids not properly separating from one another during anaphase, referred to as *anaphase lag*. [84] If the cell has more than two centrosomes, it may result in *multipolar mitosis* with multipolar spindles developing. The chromosomes can hence be separated abnormally and result in aneuploidy in both daughter cells. It can cause *monotely* (only one kinetochore, where the spindles usually bind on the chromosome, is attached to a spindle pole), *synthely* (where both sister kinetochores are attached to the same pole) or *merotely* (one sister kinetochore is attached to both poles). [51, 85]

Extrachromosomal DNA

Extrachromosomal DNA (ecDNA) refers to any DNA that exists outside the cells' chromosomal DNA, either inside or outside the nucleus. Eucaryotic cells normally

have ecDNA confined in organelles such as mitochondria. In the context of cancer, however, diverse manifestations of ecDNAs have been identified. [86]

Early observations in cytogenetic metaphase spreads of pediatric neuroblastoma cancer cells in 1965, revealed small chromatin bodies, often appearing in pairs. They were consequently coined *double minutes* (DM). [87] These structures were shown to comprise circular DNA molecules ranging in size from 100 kb to several megabases, allowing them to be visually observed in metaphase spreads. Notably, the DMs found in neuroblastoma cells were found to often harbour a gene, coined *MYCN*. [88, 89] This is now a well-known oncogene, found to be amplified in 20 % of neuroblastoma patients, also conferring a worse prognosis. [90]

There are several proposed mechanisms for how DMs can be generated. One mechanism is through *chromothripsis*, which is a catastrophic event where one or more chromosomes are shattered into multiple DNA-fragments that are subsequently stitched together randomly. Some of the fragments may ligate to themselves or to some of the other fragments, forming double minutes. [91] There are, however, cases where there are DMs but no signs of chromothripsis. Several models have been proposed to explain this. They all have in common that a double strand break initiates the process of DM formation. [92]

Subsequent research has revealed that DMs merely comprise 30 % of all ecDNA. The other group is denoted *extrachromosomal circular DNA* (eccDNA), which are smaller than the DMs, often < 1 kb in size, making them not visible in metaphase spreads. Examples of eccDNA are telomeric circles, small polydispersed DNA elements, and microDNAs. There are several competing theories for how they appear such as after a chromothripsis event or a double strand break, through a breakage-fusion-breakage cycle or via the episome model. [86] Studies have unveiled the prevalence of ecDNA across most cancer types, with chromosomal segments frequently encompassing oncogenes, that are sometimes amplified. [93] The presence of eccDNA is correlated with unfavourable clinical outcomes. [94]

During cell division, the ecDNA is randomly partitioned between the two daughter cells. [95-97] This may result in an unequal segregation of the fragments, generating varying quantities of ecDNA across the cells. This dynamic distribution can quickly increase the ecDNA copy number in a single cell and drive intratumoral heterogeneity. [95, 96, 98, 99]

Gene amplifications

A *gene amplification* is an increase in copy number of a restricted region of a chromosome. The amplified region is referred to as an *amplicon* and ranges in size from kilobases to tens of megabases. The amplified DNA sequence can be localized as ecDNA, repeated sequences at the same locus (forming intrachromosomal homogeneously staining regions; HSRs) or scattered throughout

the genome. Amplifications of genes, alongside an increased expression of the corresponding gene product, are commonly seen across tumor types. Interestingly, a single amplicon may contain DNA from different parts of the genome. [100] Genes that has been shown to be frequently amplified in cancers are *MYCN, CCND1* and *EGFR*. [101] Detection of novel amplified genes in cancer cells is of special interest since they could be candidate oncogenes considering that they are often evolutionarily selected for.

Several theories have been proposed to explain how amplicons may be formed. The most popular is that they are generated through a *breakage-fusion-bridge cycle*. The cycle is initiated by two chromosomes attaching to each other, forming a dicentric derivative of a chromosome. When the cell undergoes mitosis, the mitotic spindles bind to the two centromeres on the chromosome. This results in the chromosome being pulled towards opposite poles, forming a DNA bridge. This bridge will eventually break. Since the chromosome is duplicated during mitosis it might subsequently create a dicentric chromosome again and the cycle repeats. Depending on how the breakage occurs this process can generate many copies of the same chromosome region, resulting in an amplicon. Other models are the translocation-deletion-amplification model and the episome model. [102] [103]

Chromosomal instability and cancer

Our genetic material has evolved over millions of years, resulting in a remarkably stable and well conserved set of chromosomes. In addition, all non-germline cells in the human body exhibit the same number of chromosomes despite having gone through a vast number of proliferations. Despite this, it has been observed that almost all tumors have chromosome numbers deviating from this number, through gains and losses of whole chromosomes or large portions thereof. [48, 101, 104, 105] This feature is especially common in high-grade adult carcinomas [6] and childhood cancers. [23]

Chromosomal instability (CIN) is defined as an elevated rate of chromosomal missegregation. It is important to distinguish CIN from aneuploidy since they are distinct entities. Aneuploidy refers to the *presence* of whole chromosome, or chromosome arm alterations. CIN on the other hand refers to an increased *rate* of chromosomal missegregation. The two do, however, co-occur since an increased CIN also results in aneuploidy. CIN has been shown to be a recurrent feature in aggressive cancers. [106] Genetic instability can also present itself at the nucleotide level, comprising nucleotide excision repair (NER)-associated instability (NIN) and microsatellite instability (MIN), often caused by mutations and epigenetic changes in mismatch repair genes. [105]

There are several mechanisms that may cause CIN [85] such as *telomere dysfunction* triggering breakage-fusion-bridge cycles [107] or defects in *chromosome cohesion* or in the *spindle assembly checkpoint* (SAC). SAC constitutes an important step of the cell cycle through controlling the attachment of the spindles to the chromosomes. Other causes can be faults in the *kinetochore–microtubule attachment* or *cell-cycle regulation*. Finally an increased number of centrosomes, so-called *supernumerary centrosomes*, may result in multipolar spindles increasing the risk of merotely. [108] These changes each increases the risk of missegregation through monotely, synthely, or merotely.

Studies have identified a distinct subset of cancer cells within tumors, characterized by their significantly larger size compared to other cells. These enlarged cells are known as poly-aneuploid cancer cells (PACCs) or polyploid giant cancer cells (PGCCs), and they exhibit unique features. They have been found to be more prevalent in primary tumors of patients with metastases and are even more abundant in the metastases compared to the primary tumor. [109] PACCs demonstrate both polyploidy, resulting from one or more complete whole genome doublings (WGDs), and aneuploidy, involving the loss or gain of chromosomes or chromosome arms. Surviving a WGD grants these cells increased tolerance to chromosome aberrations, promoting CIN. [110, 111] WGDs are also associated with a poorer prognosis across various tumor types, suggesting a potential role in tumor aggressiveness. [110] Moreover, PACCs can emerge in response to stressors like chemotherapy, [112] exhibiting resilience and enhanced evolvability, facilitating their survival in challenging environments. [113] These cells may also express cancer stem cell markers and possess characteristics of stem cells, such as the ability to differentiate into various cell types. [114] Additionally, they can undergo depolyploidization through asymmetric cell division, producing daughter cells with near-diploid DNA content. [115] Recent findings have also highlighted the clinical relevance of targeting PACCs. In ovarian cancer, blocking the formation of PACCs has been shown to enhance the response to PARP inhibitors, suggesting a promising avenue for improving treatment outcomes. [116]

To conclude, CIN and aneuploidy plays a crucial role in tumorigenesis. There are, furthermore, emerging treatments targeting CIN itself by e.g., reactivating cellular pathways that inhibit CIN, by targeting the consequences of aneuploidy, by increasing CIN so that the cell cannot survive. [117]

How genetic alterations can be used to track evolution

Almost all tumors possess genetic alterations of some sort. Due to the evolutionary nature of cancer, different parts of the primary tumor and metastases may display different subclones. Since each subclone has a different set of genetic alterations,

these sets could be seen as a form of unintentional natural barcoding of subclones and could therefore, theoretically, be used to track tumor evolution. Early genetic alterations in the history of the tumor, tend to be present in almost all cells, while later genetic alterations are only present in a subset of cells. By sampling several anatomical regions and analysing them genetically, it is possible to use mathematical methods to determine the temporal order of the alterations and hence also the order in which different subclones developed in the tumor. This is performed using phylogenetic reconstruction, which will be discussed in the next section.

Phylogenetics

Basic tree characteristics and definitions

By analysing the genetic composition of a tumor spatially and temporally, mathematical methods can be employed to reconstruct its evolutionary history. This can further be visualized in the shape of an ancestral tree, denoted a *phylogenetic tree* or *phylogeny*. It is a graph representing the evolutionary relationship between different biological entities and is widely used in evolutionary biology to explore the genetic relationship between species, determine when they diverged from each other temporally, genetically, or phenotypically and to identify similarities and differences between them. Similar algorithms can be employed to produce ancestral trees displaying the relationship between subclones of cancer cells (**Figure 8**). [118]

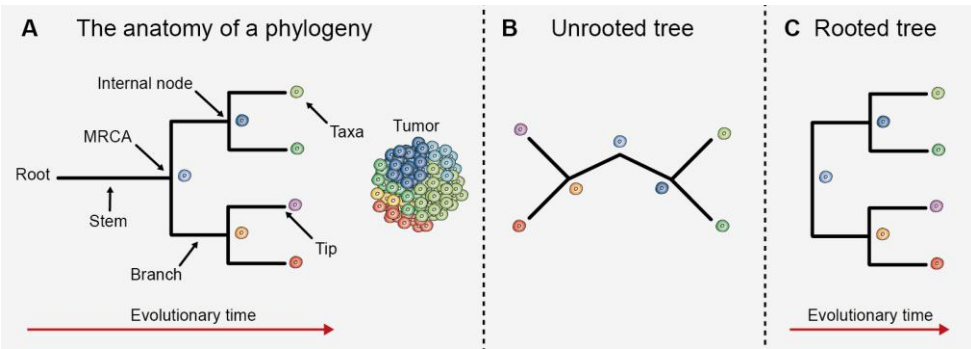


Figure 8. A) A phylogenetic tree and its constitutive parts. The red arrow indicates the direction of time, such that the tree is read from left to right. At the bottom of the tree (leftmost), the root is situated, leading to the most recent common ancestor (MRCA) of all the entities downstream (right). The branches are the horizontal lines connecting a descendant to its ancestor. Their lengths are proportional to the number of alterations separating the entities. The root and the MRCA are connected via a branch denoted the stem, which contains all features shared between all entities downstream (right) of the MCRA. The internal nodes indicate where a speciation event has taken place and is followed by a branch leading to at least one descendant. The tip labels are the biological entities or taxa. Phylogenetic trees can be used to illustrate the evolutionary history of biological entities. In this example it illustrates the evolutionary history of the simplified tumor to the right, where each color represents a distinct subclone. **B)** An unrooted tree is an undirected graph, only showing the evolutionary distance between the entities. **C)** A rooted tree is a directed graph with an inferred MRCA.

Figure 8 depicts a *phylogenetic tree (phylogeny)*. It can be rooted or unrooted. A *rooted* tree is a directed graph where a *most recent common ancestor* (MRCA) is explicitly assigned, i.e., we propose a hypothetical MRCA based on prior knowledge. In the case of tumors, the presumed cell of origin is a normal cell without any genetic alterations. An *unrooted* tree, on the other hand, lacks any assumed ancestral line and does not indicate any beginning or direction of the inferred evolutionary transformation, but merely illustrates the evolutionary distances between the entities. [118]

The *stem* of the phylogeny encompasses the alterations shared by all cells in all sampled regions. These alterations are often referred to as truncal, clonal or *stem alterations*, while alterations present in a subset of cells are called *subclonal alterations*. Private alterations are genetic alterations only present in one taxon. *Internal nodes* represent speciation events, the common ancestor from which at least one lineage descends. *Branches* are horizontal lines that encompass the genetic alterations separating the descendants from the ancestors, and their length is proportional to the number of alterations separating them. A *clade* refers to a group of organisms or cells believed to comprise all the evolutionary descendants to a common ancestor. [118]

In our studies we define a *subclone* as a population of at least two tumor cells with identical genetic profiles which that are not present in all cancer cells in at least one region. Otherwise, it is denoted a *clone*. The *Fitness* is the ability of a tumor cell to survive and proliferate. Subclones with an increased fitness compared to their neighbours will become more prevalent over time. Driver mutations confer a fitness advantage, while passenger mutations do not affect the fitness. A *selective sweep* occurs when a genotype emerges with an extremely high fitness such that it outcompetes all other subclones in the sampled area. [119]

By reconstructing a phylogenetic tree based on multiple samples from the same tumor in a patient, it is possible to investigate its evolutionary dynamics. It is also possible to identify early genetic changes, characterized by their presence in nearly all cells across all samples, constituting possible targets for precision medicine. Furthermore, we it can be used to visualize the heterogeneity of the tumor and to determine which subclones survive therapy, which cause relapses, metastasize and when. It also allows exploration of how different metastases are related to one another and to the primary tumor before and after therapy. Comparing the evolutionary history of tumors across different patients holds the possibility to elucidate similarities and differences, that may account for variations in treatment outcomes.

Methods for phylogenetic reconstruction

There are many ways in which phylogenetic trees can be reconstructed. The oldest and simplest methods are distance-based ones such as the neighbour joining method (NJ), the unweighted pair group method with arithmetic mean (UPGMA) and the weighted pair group method with arithmetic mean (WPGMA). There are, however, other methods, such as the maximum parsimony (MP) method, the maximum likelihood (ML) method and Bayesian methods. [120] The following sections will provide an overview of each of these methods, highlighting their respective strengths and weaknesses. This exploration aims to inform our decision on the most suitable methods for our studies. A comparison between software used in cancer research can be seen in the methods section.

Distance-based methods

Distance based methods, as the name implies, uses the evolutionary distance between sequences to construct a tree that most accurately represents the observed dataset. The branch lengths are intended to mirror the extent of evolutionary change between distinct entities. Each distance can be visualized as the branch length that separates each taxon in the tree, which is the best unrooted tree for that pair of sequences. The overarching objective is to find the tree that best integrates all these individual two-sequence trees into a cohesive, larger structure. [120, 121]

The distance matrix

The distance matrix utilized in distance based phylogenetic methods is a square matrix containing the pairwise distances between the entities to be analysed. For N elements, the matrix will have the dimensions $N\,x\,N$. The matrix is symmetric since the distance between entity i and j is the same as between j and i, $D(i,j) = D(j,i)$. Each matrix element represents the distance between the corresponding entity of the row and column. A larger distance suggests a greater degree of dissimilarity or divergence between a pair of entities, while a smaller suggests a greater similarity. The distances in the matrix can be derived from various data sources such as DNA sequences, RNA sequences, or amino acid sequences in proteins. It can also be binary data, indicating the presence or absence of certain traits or genetic alterations. The choice of method to compute the distance between the entities depends on the available data. Various methods exist for calculating the distance between two sequences. The most used and simplest metric is the Hamming distance [120, 121], but other distance metrics such as the Manhattan distance, and the Euclidean distance are sometimes used (**Figure 9**).

The **Hamming distance** was originally proposed by Richard Hamming in his fundamental paper on Hamming codes in the 1950s [122], and is widely used in error detection systems. It is defined as the number of positions at which two strings of the same length differ from one another, i.e., the minimum number of substitutions needed to transform one string into another.

$$H(A,B) = \sum_{i=1}^{n} \delta(a_i, b_i) \tag{1}$$

Where A and B are two sequences, n is the length of the sequences, a_i and b_i are the characters at position i in sequences A and B, and $\delta(a_i, b_i)$ is the Kronecker delta function, equalling 1 if the two entities are not equal or 0 otherwise. It is important to note that this measure is Boolean, meaning it yields either true or false, and solely focuses on whether the characters differs or not, disregarding their actual values. This is often sufficient when comparing DNA sequences or the presence or absence of a genetic alteration, but it does not capture the complexities of evolutionary processes such as transitions and transversions in nucleotide sequences. It also assumes equal weights for all character changes, which may not be appropriate, but is suitable for cases where only character presence or absence matters.

The *Manhattan distance* is defined by the absolute difference between the actual cartesian coordinates of two entities in a multidimensional space. Only the distance and not which path is taken is important. If the data is binary the Manhattan distance is equal to the Hamming distance.

$$M(A,B) = \sum_{i=1}^{n} |a_i - b_i| \tag{2}$$

Where A and B are two sequences, n is the length of the sequences, a_i and b_i are the characters at position i in the sequences A and B, respectively. It is more robust to outliers compared to the Euclidean distance and it is applicable to both binary and multistate character data. It does, however, assume that there are linear relationships between the characters, and it is also sensitive to the scale used.

The *Euclidean distance* is the length of a straight-line segment between two points.

$$E(A,B) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \tag{3}$$

Where A and B are two sequences, n is the length of the sequences, a_i and b_i are the characters at position i in sequences A and B, respectively. The result is most

often not a whole number and its biological interpretation for genetic alterations is a bit unclear. It is also very sensitive to outliers, which can disproportionately influence the distance measure. The method is well-suited for continuous and multistate data. It does, however, assume linear relationships between characters. In high-dimensional space the curse of dimensionality can impact the interpretation of the distance i.e., the dissimilarity between entities may become less meaningful as the number of dimensions increases. In addition, the distances are not scale invariant, meaning that if the data is multiplied with a common factor, it will change the distances. It is, however, translation invariant such that shifting the entire dataset by a constant value does not affect the distances.

In genetic analyses the data is usually either binary (an event is present or not) or nucleotide sequences (A, T, C and G). Therefore, the most used method for computing evolutionary distances is the Hamming distance, where the absolute number of differences between two sequences are compared.

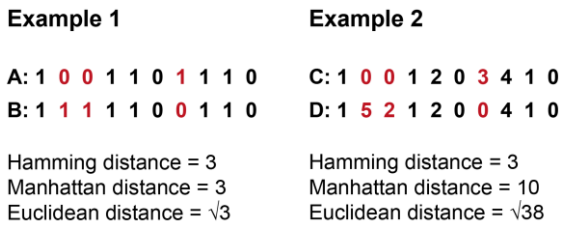


Figure 9. Two examples of how to calculate the Hamming, Manhattan and Euclidean distance between sequence A and B or between C and D. The Hamming distance is the number of instances where the two sequences do not match. The Manhattan distance is the absolute difference between the cartesian coordinates that each position in the sequence represents. The Euclidean distance is the line segment between the two points that the sequences represent.

A limitation of using distance matrices is that they may oversimplify the true underlying evolutionary processes and that information may be lost in the process.

The event matrix

An event matrix is a binary matrix indicating whether a genetic alteration is present or not in a biological entity. For studies of tumor evolution, each column could represent a subclone, and each row a genetic alteration. The matrix elements indicate the presence or absence of each of these, denoted 1 or 0, respectively.

One useful feature of using an event matrix is that all genetic alterations are treated equally during the phylogenetic reconstruction. This is especially advantageous when working with copy number changes, which can span entire chromosomes. Using merely the Hamming distance directly on copy number calling data, could yield millions of differences between the sequences, even

though it might only be due to a very large segmental copy number change or a whole chromosome gain or loss.

To use the event matrix for distance-based phylogenetic reconstruction, the pairwise distances between the entities in the event matrix are computed using e.g., the Hamming distance.

The neighbour joining method

The neighbour joining method was developed by Naruya Saitou and Masatoshi Nei in 1987 [123], and is a clustering algorithm. The algorithm iteratively joins the closest neighbours based on their pairwise distances.

The algorithm works as follows. [120, 123, 124]

1. Start with a star phylogeny where all taxa connect to a central node. The branch lengths are set to half the sum of the pairwise distances for each taxon.
2. Calculate the matrix $\mathrm{Q_{ij}}$

$$\mathrm{Q(i,j)} = (\mathrm{c}-2)\mathrm{d(i,j)} - \sum_{\mathrm{k}=1}^{\mathrm{c}} \mathrm{d(i,k)} - \sum_{\mathrm{k}=1}^{\mathrm{c}} \mathrm{d(j,k)} \qquad (4)$$

   Here *i* and *j* are the two clusters that are being compared, *c* is the current number of clusters, $Q(i,i) = 1 \; \forall \; i$, and *d* is the pairwise distance between the clusters specified in the parentheses. The equation considers how close two clusters are to each other as well as how far away they are from other clusters. Clusters that are close to each other but far away from other clusters are primarily merged. The scaling factor $(c-2)$ is needed for the function to not be taken over by the part of the equation with the sums, which would result in a merging of clusters that are far from each other.
3. The pair of clusters that has the lowest value in the Q-matrix are the ones that should be merged, since they are considered the nearest neighbours.
4. Calculate the branch lengths from the two clusters to their shared node. In the equation, u represents the shared node, f and g are the two clusters to be merged and k represents the other clusters in the dataset. The computation is performed for each of the two clusters individually.

$$\delta(\mathrm{f,u}) = \frac{\mathrm{d(f,g)}}{2} + \frac{1}{2(\mathrm{c}-2)}\left[\sum_{\mathrm{k}=1}^{\mathrm{c}} \mathrm{d(f,k)} - \sum_{\mathrm{k}=1}^{\mathrm{c}} \mathrm{d(g,k)}\right] \qquad (5)$$

$$\delta(g,u) = \frac{d(g,f)}{2} + \frac{1}{2(c-2)}\left[\sum_{k=1}^{c} d(g,k) - \sum_{k=1}^{c} d(f,k)\right] \quad (6)$$

5. Calculate the branch lengths from the other clusters to the new node.

$$d(u,k) = \frac{d(f,k) + d(g,k) - d(f,g)}{2} \quad (7)$$

6. This results in a new distance matrix. Go back to step 2 and redo it. When there are no more taxa the algorithm is ended, resulting in an unrooted tree.

Advantages: It is fast for smaller data sets, and it often produces accurate phylogenies if the distance matrix is correct. Furthermore, it does not assume uniform evolution rates (molecular clock).

Disadvantages: The algorithm relies on an unweighted distance matrix. It may also give negative branch lengths. Additionally, it only yields one phylogeny. IT also has a computational complexity which scales as $O(n^3)$, making it rather slow for larger datasets.

Unweighted pair group method with arithmetic mean (UPGMA)

UPGMA is in principle a hierarchical clustering algorithm, introduced by Sokal and Michener in 1958. [125] The algorithm operates on a distance matrix and scales as $O(n^2 \log(n))$ and is hence slightly computationally faster than the neighbour joining method. Unfortunately, the algorithm employs a strict molecular clock i.e., it assumes a constant rate of genetic change over time across the different entities in the dataset. This means that the amount of genetic change accumulates at a constant rate, resulting in an ultrametric tree where all tips are equidistant from the root.

The algorithm works as follows [120, 125]

1. Begin with a distance matrix, denoted D_1.
2. Identify the pair of taxa a and b separated by the shortest distance in the matrix.
3. Create a new cluster u by combining clusters a and b. The branch lengths (distance, D) from the cluster u to a and b are assumed to be the same.

$$D(a,u) = D(b,u) = \frac{D_1(a,b)}{2} \quad (8)$$

4. Cluster a and b in the distance matrix are replaced by cluster u. The distances in the distance matrix are subsequently updated by calculating the distance from cluster u to all other entities in the distance matrix.

$$D\big((a,b),u\big) = \frac{|a|D(a,u) + |b|D(b,u)}{|a| + |b|} \tag{9}$$

5. Go back to step 2 and repeat the steps using the updated distance matrix. If all entities have been handled, the algorithm is ended.

Advantages: The algorithm is slightly faster than the neighbour joining method and it often produces accurate phylogenetic trees if the distance matrix is correct.

Disadvantages: The UPGMA results in an ultrametric tree, assuming a strict molecular clock. While this assumption may offer a convenient simplification for dating fossils and tracking the divergence of animal species, it is an overly idealized assumption. Several studies indicate that the rate of evolutionary change varies significantly among different biological taxa [126] and in cancer it has been shown that different subclones possess varying mutation frequencies. [127] True phylogenies rarely exhibit ultrametric characteristics and reliance on such an assumption might bias the interpretation of the results. Most phylogenetic methods today therefore employ relaxed clock models, allowing for different rates of molecular evolution among lineages. [128] It is worth noting that certain molecular markers or specific regions of the genome, such as mitochondrial DNA and microsatellites may exhibit relatively constant substitution rates. [129] [130] Conversely, some exhibit such high mutation rates that they comply to the assumption of a strict molecular clock with a linear relationship between mutations and time. Polyguanine tract repeats provide an example of such genomic elements. [131]

Another limitation of the UPGMA method is that the tree may not be additive. This means that the branch lengths may not correspond to the absolute distances between the entities according to the distance matrix.

Weighted pair group method with arithmetic mean (WPGMA)

WPGMA is a distance-based phylogenetic method, also developed by Sokal and Michener in 1958. It is a weighted version of the UPGMA method, aiming to address some of the limitations of UPGMA regarding its assumption of a strict molecular clock. Weights are therefore used to assign varying degrees of influence to characters in the phylogenetic reconstruction. [125] The algorithm operates as follows.

1. Begin with a distance matrix, denoted D_i, where i equals the number of iterations. In the first iteration the matrix is thus denoted D_1.

2. Find the pair of taxa *a* and *b* separated by the shortest distance according to the distance matrix.
3. Create a new cluster *(a,b)* by combining clusters *a* and *b*. The branch lengths from cluster u to *a* and *b* are assumed to be of equal lengths.

$$D_i\big(a,(a,b)\big) = D_i\big(b,(a,b)\big) = \frac{D_i(a,b)}{2} \tag{10}$$

4. Update the event matrix, which now will have one row and one column less. When the distances are updated in the matrix the distances are averaged between each element of the first cluster (*a*,*b*) and each of the remaining elements, *k*. The selection of the weights w_i and w_j, varies depending on method.

$$D_{i+1}\big((a,b),k\big) = \frac{w_i D_i(a,k) + w_j D_i(b,k)}{2} \tag{11}$$

5. Go back to step 2 and repeat until all entities have been handled.

Advantages: The method is simple and computationally efficient. It is also capable to somewhat handling situations where the assumption of a strict molecular clock is not true.

Disadvantages: It still produces ultrametric trees and thus implies a constant rate of evolution across lineages, despite attempts to use weights to certain characters in the phylogenetic reconstruction procedure. It is also sensitive to outliers and, just as for UPGMA, branch lengths may not accurately reflect the actual amount of evolutionary change when the molecular clock hypothesis is violated.

Maximum parsimony

The maximum parsimony (MP) method aims to find the tree that minimizes the total number of evolutionary changes. Hence, the goal is to identify the shortest and easiest tree that explains the observed data, referred to as the most *parsimonious* tree. The foundational work was performed by Edwards and Cavalli-Sforza in 1963 [132] and subsequent advancements were made by James S. Farris in 1970 [133] and Walter M. Fitch 1971. [134] The idea behind the method is that genetic alterations are rare and that the solution encompassing the fewest number of substitutions, while still explaining the data, is preferred over those necessitating additional substitutions.

Finding the most parsimonious tree is challenging due to the large increase in the number of possible phylogenies as the number of taxa increases. Performing an exhaustive search through all possible trees in the *tree space* (a theoretical space

encompassing all possible trees for a given set of taxa) and calculating the length of each, may thus not be possible. Therefore, several algorithms have been developed to perform this tree search to find the most parsimonious tree. [135]

The *branch-and-bound method* (BB) prunes branches that cannot lead to a more parsimonious solution. [136] For a higher number of taxa, however, a *heuristic search* is needed. Heuristic search techniques use various strategies to try to find the optimal tree without looking at all possible solutions. They often use a hill climbing strategy (**Figure 10**). This does, however, not guarantee that the one found is the best, since there is the possibility of ending up in a local optimum. One example of a heuristic search method is the *nearest neighbour interchange method* (NNI), in which subtrees in a tree topology iteratively are exchanged to explore different tree rearrangements. In some versions all nearest neighbours are explored while others halt as soon as a better one is found. If a nearest neighbour tree is found that is more parsimonious than the current tree, this one is used as the current tree for the next iteration of nearest neighbour search. If none of the rearrangements are more parsimonious, the tree is localized at a maximum point in the tree landscape and the search is ended. This might, however, be a local maximum, but since not all trees are investigated, it is not known. [120, 135] Another method is the *subtree pruning and regrafting method* (SPR), which selects and removes a subtree from the main tree and reinserts it in all possible places to create a new node and compute a new parsimony score for each combination. Hence, the SPR method carries out a much wider search than the NNI method and is thus more likely to find a better peak in the tree landscape. [120, 135] The *tree bisections and recombination* (TBR) method is similar, in that it removes a subtree from the main tree, but it iteratively tries all possible connections between the two trees in all possible ways. It thus carries out an even wider search than the SPR method. Note that all these methods assume that we start with an initial tree. This tree can be obtained by random construction, by sequential addition, a star phylogeny or by using e.g., a distance-based method such as the neighbour joining method. [120, 135]

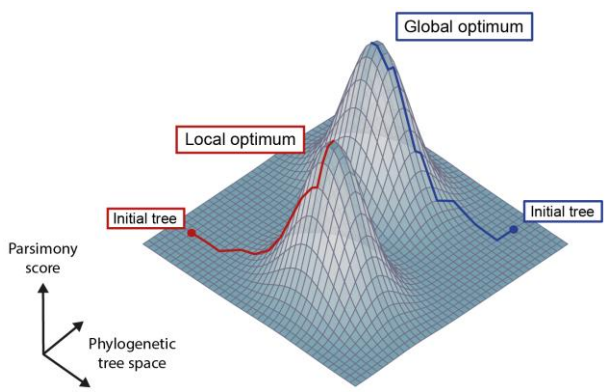


Figure 10. Illustration of a hypothetical tree space. The search is initialized with a starting/initial tree using for example random construction, sequence addition, a star phylogeny or by using a distance-

based method such as the NJ method. This is followed by a heuristic hill-climbing strategy to identify a maximum, here exemplified by trying to find the maximum parsimony score, using e.g., NNI, SPR or TBR. Depending on the start position and the algorithm used, the search might end up in a local optimum (red path) or the global optimum (blue path). By using a method such as the parsimony ratchet this can be circumvented by investigating several paths using different initial trees, choosing the best tree of all searches.

Another method to find the most parsimonious tree is *the parsimony ratchet*. Here, 5-25 % of the characters in the dataset are selected and are given a higher weight compared to the other ones. This is followed by a tree search using e.g. TBR, as described above, until a maximum is reached, resulting in a recommended tree. Similar tree searches are repeated at least 50-200 times, each time with a different set of characters given increased weight. The best tree discovered among all searches is retained. In our studies, we have employed the parsimony ratchet method with 2000 iterations using the TBR method for tree search.

Advantages: MP is conceptually straightforward, and it aligns with Occam's razor. It is also possible to generate several possible solutions.

Limitations: MP assumes that evolutionary change is rare and might thus underestimate the actual evolutionary change. It is also computationally challenging as the tree space might be vast. In addition, the method assumes that all similarity is homologous, and the method can be sensitive to convergent evolution or other processes that result in homoplasy (similarity that is not due to a common ancestor).

Maximum likelihood

The maximum likelihood method (ML) is a widely used method in mathematics. It aims at maximizing the likelihood function L for a given dataset. In phylogenetics this boils down to finding the tree parameters that maximizes the likelihood function, given a certain data set with multiple sequence alignment data. The method was developed by R.A. Fisher in the beginning of the 20th century and has since then been widely used. [137] The likelihood is, by definition, given by

$$\mathrm{L}(\theta) = \mathrm{P}(\mathrm{D}|\theta) \qquad (12)$$

Where θ are the tree parameters and D is the data set representing the taxa. It gives the likelihood, L, given the data set, D, and the tree parameters, θ. The goal in phylogenetic inference is to estimate the tree parameters, θ_E, so that the likelihood function is maximized. The tree parameters θ_E are the *estimate* of the true tree parameters θ. These are the parameters, given the data set (which is not a perfect representation of reality), that maximize the likelihood function. To calculate the

probabilities, a so-called substitution model is used, providing information about how probable a certain state change is in the data set.

Define the data set as $D = (D_1, D_2 \dots D_N)$ where D_i is the data from each cell at position i in the genome. To aid the computational load further, it is often assumed that changes at different sites, as well as the evolution in different lineages, are independent of each other. This aids the mathematical computations greatly. [120] Using the assumptions stated above, the likelihood function can be expressed as

$$\mathrm{L}(\theta) = \mathrm{L}_1 \cdot \mathrm{L}_2 \dots \mathrm{L}_\mathrm{N} = \mathrm{P}(\mathrm{D}|\theta) = \prod_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{P}(\mathrm{D}_\mathrm{i}|\theta) = \prod_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{L}_\mathrm{i}(\theta) \qquad (13)$$

The total likelihood for the whole sequence, $L(\theta)$, is the product of each of these likelihoods. Since the probabilities often are very small numbers it is convenient to use the loglikelihood function instead, which has its maximum at the same parameter values.

$$\ln\big(L(\theta)\big) = \ln(L_1) + \ln(L_2) \dots \ln(L_N) = \sum_{i=1}^{N} P(D_i|\theta) \qquad (14)$$

This function computes the loglikelihood given a certain data set and tree parameters. By maximizing this function, the most likely model parameters, θ^*, and hence the phylogenetic tree, can be obtained. This is done by computing the derivative of the function, setting it equal to zero and by solving for the tree parameters. [120, 138]

To perform the computations, an initial tree is needed, using for example the neighbor joining method. It also requires knowledge of the probability of each state change the phylogeny implies. This information is provided through a *substitution model*, describing the probability of evolutionary change between the states. The state changes could be that between macromolecules, for example the probability of transitions or transversions between A, C, T, and G or transitioning between amino acids, or it can be between binary states. There are many substitution models (**Table 2**) and most assume independence among sites i.e., that an alteration at one site is independent of an alteration in another site. These substitution models also assume that the probability of transforming one state into another only depends on which state it currently is in, and not which state it was in previously. It can hence be seen as a form of Markov chain. [120]

Each model is based on a certain set of parameters. Q is the rate matrix, indicating how frequently a certain state transforms into another. π is the base frequency, indicating how frequent a state is. Many models assume that the base frequency is constant meaning that the prevalence of each state is equal initially. P(t) is the transition probability matrix, indicating the probability of a certain state change. They have the properties such that

$$\pi_i Q_{ij}(t) = \pi_j Q_{ji}(t) \quad (15)$$

$$\pi_i P(t)_{ij} = \pi_j P(t)_{ji} \quad (16)$$

$$\forall i, j, t$$

There is a plethora of different substitution models. Some use a strict molecular clock where the mutation rate is considered constant, regardless of taxa, while others use a relaxed molecular clock which considers a variability in the rate of the molecular clock between evolutionary lineages. Finally, some models are time reversible and do not differentiate between the sequence from the ancestor and the descendant i.e., there are no special taxa in the model (**Table 2**). [120, 139]

Table 2. Overview of *some* of the available substitution models along with a short description of their assumptions. [139] A graphical representation of each model can be seen in **Figure 11**.

Model	Year	Description
The Jukes and Cantor model [140]	1969	Assumes equal transition and transversion rates as well as equal base frequencies.
K2P or K80 [141]	1980	It has two parameters, one for transitions and one for transversions and it assumes equal base frequencies.
K3ST, K3P or K81 [142]	1981	It has three parameters: One for transversions between A and T as well as C and G. Another parameter for transversions between A and C as well as G and T. A separate one for the transitions. It also assumes equal base frequencies.
F81 [138]	1981	Assumes equal transition and transversion rates. Allows the base frequency to change.
HKY85 (K80+F81) [143]	1985	It distinguishes between the rate of transitions and transversions and allows the base frequencies to change.
GTR [144]	1986	It has 11 parameters. It allows all substitution frequencies and base frequencies to vary.
T93 [145]	1993	It has three parameters: One for transversions and then a separate one for transitions between A and G and one for transitions between C and T. The base frequencies are also allowed to change.
Two-state substitution models [120]	1971-78	The Cavender-Farris-Neyman model. It is identical to the JC69 for two states.

In our studies we have used the *Jukes and Cantor model* (JC69). It was proposed by Jukes and Cantor in 1969 and is the oldest and simplest substitution model. It is based on the presumption that time can be discretized into infinitely small subunits, during each of which only one genetic alteration can occur. Additionally, it assumes equal transition rates and equilibrium frequencies (**Figure 11**). [140]

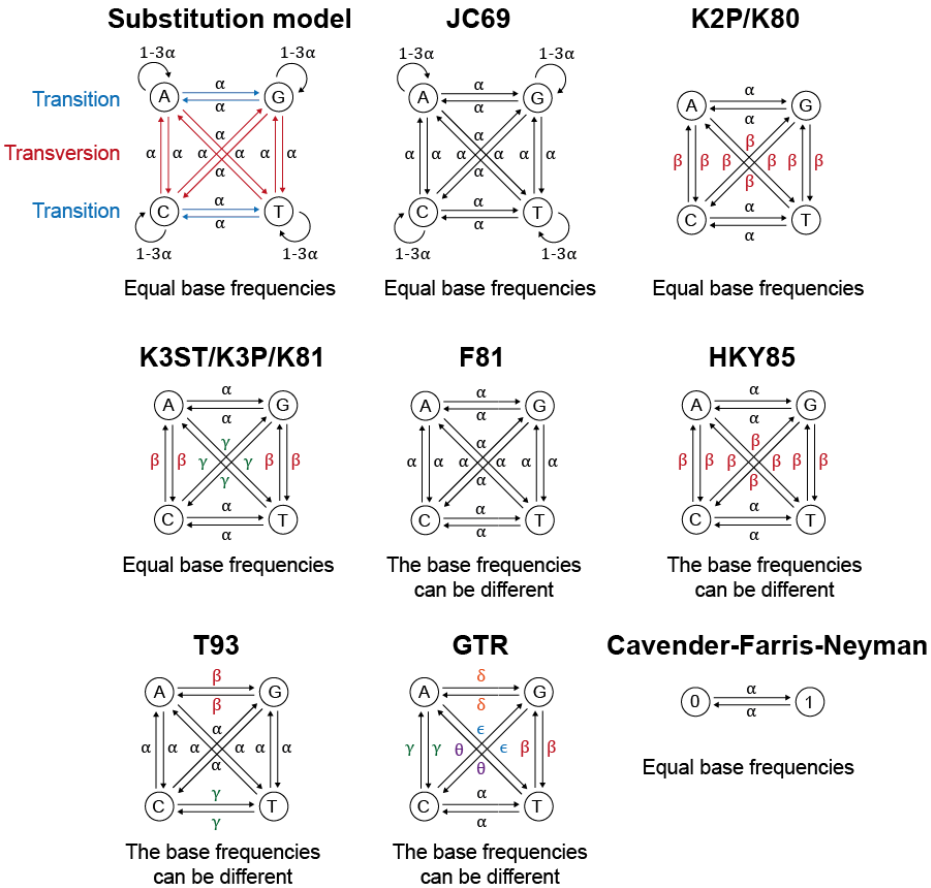


Figure 11. An overview of the substitution models, as listed in **Table 2**. Here, A, C, T and G refers to the nucleotide bases. Transitions are changes between the nucleotide bases A and G, as well as between C and T. Transversions are changes between the nucleotide bases A and T, C and G, A and C as well as between G and T. The annotations α, β, γ, δ, ϵ and θ are substitution rate parameters. For each substitution model there is a visualization of the substitution rate parameters used in each substitution model. For the JC69 there is also a visualization of the probability of maintaining the same state after each time increment. This is the probability 1 minus the sum of the probabilities to change into another base. In this case it is $1 - 3\alpha$. For the other substitution models, I have omitted this part of the figure to make the figures less messy, but all information is there to compute it.

Advantages: The ML method is less affected by sampling error than other methods and is robust to violations of the evolutionary model. Even with short sequences, it outperforms the parsimony and distance-based methods when analysing a large number of characters. [146] It is also statistically well informed, evaluates different tree topologies and allows for direct usage of sequencing data.

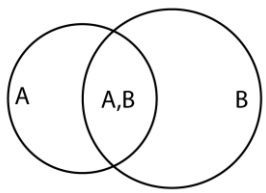
Limitations: It is rather slow, and the result depends on which evolutionary model is chosen. [120]

Bayesian methods

Bayesian methods for phylogenetic reconstruction have been employed since the 60s. They are methods of statistical inference where Bayes' theorem, developed in the 18th century, is used to update the probability for a hypothesis when new evidence or information becomes available. [147] The equation is:

$$P(H|D) = \frac{P(D|H) \cdot P(H)}{P(D)} \tag{17}$$

To explain the equation in an intuitive way, imagine two circles overlapping.



Let $P(A)$ be the probability to get an A and $P(B)$ the probability of getting B. The conditional probability describes the probability of a certain event given the first one. The probability of getting A given B is $P(A|B)$ and can be written as

$$P(A|B) = \frac{P(A,B)}{P(B)} \tag{18}$$

Where $P(A,B)$ is the probability of getting an A and a B. In a similar manner, the probability of getting a B when we have an A is

$$P(B|A) = \frac{P(A,B)}{P(A)} \tag{19}$$

Note that the nominator is the same in (18) and (19). It thus follows that

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \tag{20}$$

Which gives us Bayes' theorem!

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} \tag{21}$$

The probabilities in Bayesian inference are the degree of belief and hence a way to quantify the uncertainty. Bayesian inference in phylogenetics is like the maximum likelihood method, just that it instead uses a prior distribution of the tree being inferred. Now let's look at the theorem again.

$$P(H|D) = \frac{P(D|H) \cdot P(H)}{P(D)} \tag{22}$$

- $P(H|D)$ is the probability of the hypothesis being true given the data. This is denoted the posterior probability.
- $P(D|H)$ is the probability of the data being true given the hypothesis (phylogeny), also denoted the likelihood.
- $P(H)$ is the prior probability of the hypothesis (the suggested phylogeny).
- $P(D)$ is the probability of the dataset.

To use these methods for phylogenetic inference one often uses Markov Chain Monte Carlo (MCMC) methods to draw a random sample from the posterior distribution. MCMC methods are a class of algorithms for sampling from a probability distribution. A Markov chain is a method for generating a sequence of variables where the current value only depends on the previous one. Hence, what happens next depends only on the state right now i.e.

$$P(x_{i+1}|x) \tag{23}$$

A *Markov chain* consists of a set of states. For each pair of states there is a transition matrix that describes the likelihood of moving from one state into another. A *hidden Markov model* is an extension of the Markov chain where the states are hidden, and some variables are observed. If one wants to decipher the most likely Markov chain given the observed data, Bayes theorem is useful. [147]

A *Monte Carlo* method is a technique for randomly sampling a probability distribution and approximating a desired quantity. The most commonly used technique for MCMC is the Metropolis-Hastings method, which can be described as [120, 148]

1. Draw a tree T_i from the posterior distribution.
2. Draw another tree T_j close to T_i.
3. Compute their ratio $R = f(T_j)/f(T_i)$. If the ratio is ≥ 1 it is accepted as the new tree. Else a random number n between 0 and 1 is drawn. If $R < n$ the new tree is rejected. If $R > n$ it is accepted.
4. Return to step 2.

This method is a Markov chain since the next one only depends on the current state. A problem is that this algorithm never terminates so a limit must be set.

Limitations: A serious issue with Bayesian methods is the choice of priors that are chosen subjectively, and greatly influences the results. In addition, Bayesian methods are very computationally demanding. [120, 147]

Advantages: It allows the incorporation of prior information, which can be useful when dealing with complex models or limited data.

Rooting a phylogeny

It is necessary to root a tree i.e., having a predetermined common ancestor of the entities of the phylogeny, to draw conclusions about the order of events. There are two different methods denoted the outgroup criterion and the use of a molecular clock. In the first method, a common ancestor is chosen based on previous knowledge followed. In the second method, it is assumed that the alterations in the tree follow a clocklike behaviour meaning that it aims at finding the point on the tree that has an equal expected amount (not the same as the observed) of change (branch lengths) from there to all tips (an equidistant point), which can be difficult and may not be correct.

Evolutionary biology

Evolution by natural selection

Natural selection describes how organisms with advantageous traits are more likely to survive and reproduce, than those without such traits. This results in a gradual evolution of populations over time, such that the traits that give a survival advantage become more prevalent in the subsequent generations. It is important to note that natural selection acts on the phenotype, rather than the genotype, but in the end it is the genetic composition that is inherited by its offspring and will increase its allele frequency in the population. [149]

In the middle of the 19th century, the naturalist Alfred Russel Wallace independently formulated the theory of natural selection after his extensive travels. Rudimentary versions of the idea had, however, been proposed by various researchers before him. He later composed an essay on the subject, which he sent to Charles Darwin, who had, unbeknownst to Wallace, also been independently developing the idea. A joint paper with Wallace's essay and a writing from Charles Darwin was published in 1858. [150] This was followed by Darwin publishing his book "On the origin of species" in 1859, which he had been working on for many years. The book encompasses a detailed exposition of natural selection and how it can explain the diversity and complexity of life on Earth, which established him as the founder of the theory. [151]

According to Darwin, competition for limited resources such as food, water, partners and space, results in organisms that are better adapted to their environment and have an advantage in producing offspring. This leads to the transfer of advantageous traits to the next generation, resulting in the development of complex adaptations like the ability to fly, swim, or see. Natural selection also explains how organisms can adapt to narrow ecological niches (microevolution), leading to the emergence of new species over time (macroevolution). [149, 151]

For evolution by natural selection to operate, there must be, [149, 151]

1. Heritable variation in the population.
2. Fitness differences in the population.
3. Competition between species.

Natural selection in the context of cancer

Since tumors consist of a genetically and phenotypically heterogeneous population of cancer cells, it has been proposed that cancer evolution may operate through natural selection, but of cancer cells. [1, 29, 31, 152] Let us delve into each of the three criteria, presented above, and discuss what they mean and whether they are fulfilled or not in a tumor.

Heritable variation

Heritable variation is a key component of natural selection. *Variation* arises from mutations, which are generated through a more or less random process that can give rise to variants that are beneficial, neutral, or deleterious. While the origin of a new genetic variant occurs at random, the probability of it being passed on to the next generation *is not* if it impacts the survival and reproductive probability of the organism. It is this non-randomness that forms the basis of natural selection, resulting in a change in the *proportion* of the variants *already present* in the population. Natural selection does not by itself generate variation but acts upon it. In fact, natural selection can even lead to a reduction in genetic variation within populations. Hence, natural selection is by itself incapable of producing new traits. [149]

The generated variants must also be *heritable*. Note that the theory of natural selection was developed before the concepts of modern genetics was developed. Darwin himself believed in something called *pangenesis*, which is an outdated theory of heredity which suggests that every part of the body produces small particles called *gemmules*, which travel to the reproductive organs and are passed on to offspring during reproduction. These gemmules contain information about the traits of the parent organism. When they combine during reproduction, they determine the traits of the offspring. In this way the characteristics acquired during the lifetime of an individual can be passed on to the next generation. It was later pointed out that if this theory would be true it would result in so called *blending inheritance*. This means that if two gemmules fuse together during fertilization the traits of the offspring would be an average of the parents' ones. In time this would result in all traits blending out, resulting in a uniform average phenotype that would make natural selection impossible. [153]

Lamarckian inheritance is another discredited theory of inheritance which was proposed by the French naturalist Jean-Baptiste Lamarck in the early 19th century. It dictates that the characteristics acquired by an organism during its lifetime could be passed on to its offspring. This theory is, however, not supported by modern genetics and is incorrect. [154] [155] In 1892 August Weismann proposed the theory of the *germ plasm*, arguing that the body harnesses a germ plasm which formed the body during the development, but that the body could not influence the germ plasm i.e., that there was a one-way interaction. To prove his theory, Weismann

cut of the tails of mice and showed that the offspring had normal tails. This was used as proof that both pangenesis and Lamarckian inheritance were impossible. [155]

In the middle of the 19th century the monk Gregor Mendel, often referred to as the father of modern genetics, developed Mendel's laws of inheritance. The unification of Darwin's theory of evolution and Gregor Mendel's ideas on heredity into a mathematical framework called *the modern synthesis*, was first coined in 1942 by Julian Huxley in his book "Evolution: The modern synthesis". [155] After this, scientists came to accept natural selection. Furthermore, in 1918 it was shown by R. A. Fisher how continuous variation could come from several discrete genetic loci. Thus, a process generating new variation in combination with the possibility of inheriting this variation is essential for evolution to operate. [156] Since then our understanding of cell biology and genetics has developed markedly, allowing us to explain how genetic alterations can be inherited by offspring.

In the context of tumors, it is a well-established fact that cancer cells harbour both genetic and phenotypic alterations. [48] Furthermore, different parts of the tumor contain different genetic alterations and many cells display the same genetic alterations. These observations implies that the variants generated are passed on to their daughter cells, and that these descendant cells retain the ability to undergo further divisions. [5] Consequently, there exists a heritable variation in the cancer cell population and the first criteria is fulfilled.

Fitness

Fitness refers to how successful an organism with a certain genotype is at reproducing in a particular environment compared to the other organisms in the same environment. Fitness can be defined on the population level as the average fitness of all individuals with a certain genotype, or at an individual level as an individual's ability to reproduce compared to other individuals in the same population. Mutations are completely random with respect to fitness. Natural selection will, in a non-random manner, result in an increase in the proportion of variants with higher fitness and decrease in the proportion of variants with lower fitness. [149]

The same trait can thus have different fitness, depending on the environment. It is also worth noting that traits with high fitness, may have been present long before the current environment arose. Natural selection can only operate here and now and thus changes the proportion of organisms with a certain trait based on the current situation. Natural selection cannot make a trait more prevalent just because it might be advantageous in the future and cannot create new traits. [149]

Herbert Spencer introduced the phrase "survival of the fittest" in 1864 after reading "On the origin of the species", a phrase which Darwin subsequently included in the fifth edition of the book. [157] This term is rather unfortunate since it

often creates confusion. It puts emphasis on survival, while fitness generally refers to the trait's reproductive advantages, and not survival per se. A species can have a longer survival time, but still a lower fitness, than a species with shorter life span but better reproductive ability. The term leads to the confusion that fitness has to do with survival per se, which it does not. [149] The term also implies that the fittest organisms survive. At the same time the organism that survives is also the one that is fittest. This is somewhat a circular argument. Evolution by natural selection may also result in *fixation*, meaning that "survival of the fittest" not necessarily leads to a fitness increase of the fittest individual, but merely that the lesser fit variants are removed from the population while the fittest variant persist in status quo, referred to as fixation. [149]

To aid the visualization of fitness, Sewall Wright introduced the concept of a *fitness landscape* where each position represents a genotype, and the height the fitness. An adaptive landscape on the other hand changes shape with changes in population densities and survival strategies used by the species. [158] A fitness landscape can be useful for visualizing evolutionary trajectories, identifying optimal genotypes and for explaining or predicting evolutionary outcomes. It may also be used to develop strategies to guide organisms to specific positions in the landscape. [159, 160]

Studies have shown that tumor cells can gain various genetic alterations, which can be broadly categorized into three types: (1) Advantageous alterations, conferring an increased fitness; (2) Neutral alterations, having little or no effect on fitness; and (3) Deleterious alterations, resulting in a fitness decrease due to a loss of functionality. Advantageous and neutral alterations are often referred to as driver and passenger mutations, respectively. [161, 162] Since a tumor possess intratumoral genetic heterogeneity and different genetic alterations confer varying levels of fitness, it can be concluded that there are fitness differences among different cellular lineages in a tumor. Consequently, the second criterion is also fulfilled.

Competition

Competition is defined as an interaction between organisms in which the fitness of one is lowered by the presence of another. In theory, all species populations should have the capacity to increase in number exponentially. However, this is rarely observed in nature due to limitations in available resources such as food and physical space, often referred to as "a struggle for existence". [151]

Cancer cells compete for nutrients, oxygen, glucose, space, growth signals, and interact with other tumor cells as well as the tumor microenvironment. [163] Additionally, antitumoral therapies confer a selection pressure where sensitive cells are killed, leaving the resistant ones. [13] Therefore, we can conclude that there is competition between the subclones in a tumor as well as between the subclones

and their environment. Consequently, the third criteria for natural selection to act is fulfilled.

A tumor thus harbours all the necessary conditions for natural selection to operate.

Fundamental types of evolution

Several evolutionary concepts are essential for the understanding of evolutionary biology (**Figure 12**). *Divergent evolution* occurs when two or more populations diverge from a common ancestor, resulting in dissimilarity among their descendants. This process leads to the development of distinct species over time, exemplified in nature by the wolf, fox and dog diverging from a common ancestor. In cancer, divergent evolution is common, manifesting as intratumoral and intertumoral heterogeneity, within a patient, leading to genotypic divergence between subclones over time. [5]

Speciation is the evolutionary process through which populations become distinct species. It often requires reproductive isolation in some way, reducing gene flow. *Allopatric* speciation occurs when populations are separated by geographical barriers, while *peripatric* speciation involves isolation of a small peripheral population. *Parapatric* speciation is caused by species evolving reproductive isolation within a population and *sympatric* speciation occurs within the same geographical areas without any physical barriers to gene flow, which can be seen in tumors through polyploidization such as doubling of the genome. Determining what constitutes a new species can be challenging. Some researchers have proposed that tumor formation itself can be viewed as a form of speciation event that happens convergently across humans and other animals and plants. [12]

Adaptive radiation is a process whereby organisms diversify rapidly from a common ancestor into a plethora of new forms. A classic example of this is the finches in the Galapagos, described by Darwin in "On the origin of species", which display a surprising variety. In cancer the theory of a big bang model of development of colorectal cancer could be viewed as a type of adaptive radiation. [127]

If distantly related organisms, independently evolve similar traits to adapt to similar necessities, it is denoted *convergent evolution*. An excellent example of this from nature is the development of wings in bats, butterflies, and birds, which developed independently to achieve a similar goal. Such structures are also denoted *analogous* structures or *homoplasy* (similar function). In cancer the process of cancer development itself could be seen as convergent evolution, as it develops in a similar fashion across animals and plants independently. [164] Whole

genome alterations and certain mutations that are recurrent in tumors from different patients, can also be considered examples of convergent evolution.

Parallel evolution, on the other hand, occurs when independent species acquire similar characteristics while evolving together in the same environment or eco-space at the same time, often resulting in *homologous* structures. In contrast to the case of convergent evolution, the species are more closely related, and their descendants are thus even more like one another. An example is placental mammals and marsupials as well as the development of arms and legs. In cancer, identical whole chromosome alterations can occur independently, in parallel, in different parts of a tumor. [164]

Coevolution is the process by which two species' evolution affect each other's evolution through natural selection. Examples include flowers adapted for pollination by insects or birds, coevolving alongside their respective pollinators. These flowers exhibit distinct shapes tailored to attract specific insects or birds, facilitating successful pollination and reproduction. In return, the insects and birds derive essential nutrients from their interactions with these specialized flowers. [165] In tumors the interplay between the microenvironment and the cancer cells, can be considered a coevolutionary phenomenon. [166]

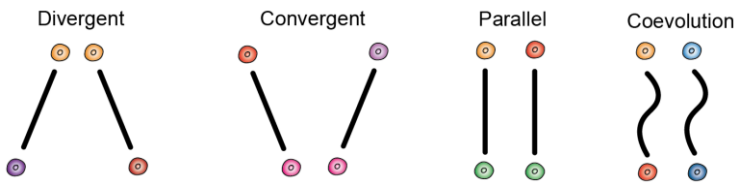


Figure 12. Illustration of the evolutionary concepts divergent, convergent, parallel and coevolution.

Modes of tumor evolution

How tumors grow has been extensively studied, and there are several proposed modes of evolution (**Figure 13**). [167] The oldest and simplest is *linear evolution*, suggesting that mutations are acquired in a stepwise manner, leading to new subclones that outcompete others through consecutive selective sweeps. According to this model, all cancer cells within the tumor should be identical until an additional genetic alteration occurs, causing the new variant to dominate the tumor. [168] This model has, however, faced criticism for presuming a homogenous cancer cell population living in a deterministic environment. [169] Limited experimental evidence supports linear evolution as the sole mode of tumor evolution. If the model was accurate, one would expect to observe a dominant clone with rare persistent clones from previous selective sweeps when analysing a tumor. This contradicts the widespread intratumoral heterogeneity observed in

nearly all tumors. [5] As a result, this theory has been largely set aside in the context of advanced cancer but may still play a role in tumor initiation scenarios with a small number of cancer cells and in benign tumors. [167]

In the *branching evolution* model, subclones are considered to diverge from a common ancestor and evolve in parallel, both diverging subclones having an increased fitness. Clonal sweeps are not as common in this model, leading to the detection of multiple lineages at the same time point. This model is well supported from studies using bulk and single cell DNA sequencing, across various cancer types. [6, 170]

Punctuated tumor evolution suggests that numerous genomic aberrations occur rapidly in the initial stages of tumor evolution, followed by a few dominant clones expanding to form the tumor mass. This results in a long root followed by a branching pattern. This model is also referred to as "the big bang model of tumor evolution". [127] This big bang event could be a chromothripsis or chromoplexy event, frequently observed in tumors. [171 172]

Neutral evolution posits that random passenger alterations, which confer no fitness advantage, accumulate over time. This leads to genetic drift and extensive intratumoral heterogeneity. Consequently, there is no active selection for fitness during most of the tumor's lifespan, and the observed intratumoral heterogeneity is viewed as a by-product without functional significance in tumor growth. [173]

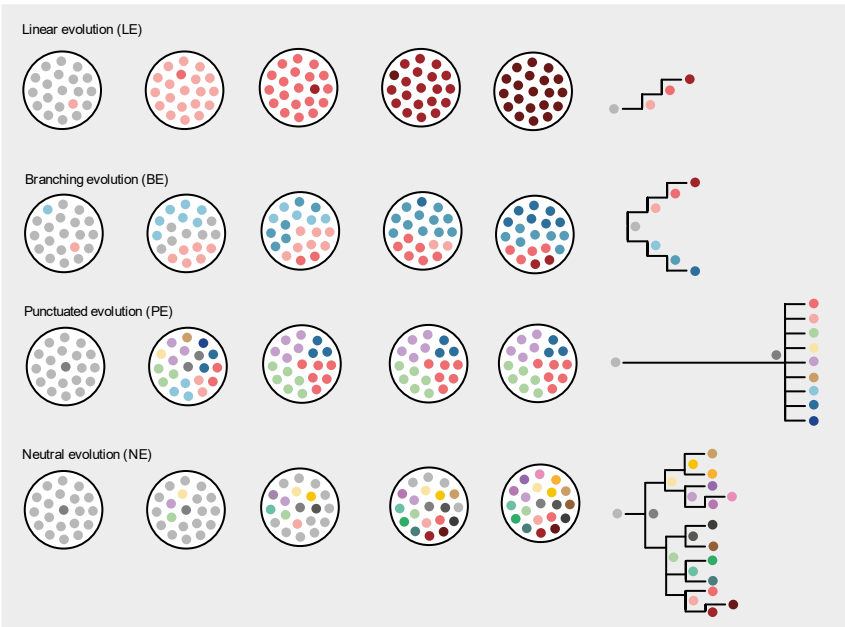


Figure 13. Proposed modes of tumor growth. Linear evolution suggests a stepwise acquisition of mutations with new subclones sweeping and outcompeting the previous ones. Branching evolution

involves the development of several subclones, each with increased fitness compared to the ancestor, and which co-exist in the tumor. Punctuated evolution entails rapid genomic changes followed by a few of these persisting in the dominant clones. Finally, neutral evolution involves the accumulation of random alterations conferring no fitness advantage or disadvantage, resulting in extensive intratumoral heterogeneity, but which is considered merely a by-product.

Antitumoral therapy from an evolutionary perspective

Despite the success of current antitumoral treatments for many cancer patients, a notable subset of patients are not cured. The efficacy of most current therapies is limited by their tendency to be either overly generalized or excessively specific. As Gatenby et al. formulated, "*it is chess, not whack-a-mole*". [20] It is, however, more akin to playing chess, without knowing the rules, and the rules that exist change with time and depending on who you play against. Evolution plays a crucial role in determining the success or failure of current therapies. By understanding why they fail, this knowledge might be used in order to develop new treatment strategies. Let us initiate this exploration by examining current therapeutic modalities through the lens of cancer evolution (**Figure 14**).

Current antitumoral therapeutic modalities

The introduction of *chemotherapeutic agents* has greatly improved survival across cancers. [14] These drugs kill fast dividing cells, such as the cells in a fast-growing tumor, but unfortunately also rapidly dividing normal cells. They are thus highly non-specific and causes extensive side effects by harming non-malignant cells such as mucosal cells, which results in vomiting, extensive wounds in the guts and mouth and diarrhea. The drugs also harm hair follicles, causing hair loss, and affect inflammatory cells, leaving the patient immunocompromised. [14] Additionally, it has been shown that the proliferative potential of different subclones in the tumor is heterogeneous, that is, different subclones proliferate at different rates, resulting in fitness differences between them. [162] Furthermore, some cancer cells can enter a hibernating or a quiescent state under stress, where they do not proliferate, just to reawake as the stressor is relieved. These dormant cells can confer treatment resistance and late relapses. [16, 174]

Introducing chemotherapeutic agents into the tumor environment will thus kill the cells that divide fast, while it may leave those that divide slowly, imposing an intense selective pressure for slowly or non-dividing cells. There may also be cells that are resistant to the agent via the development of resistance strategies such as upregulation of efflux pumps, increased drug metabolism, DNA repair, increased cell size or inactivation of apoptotic pathways. [14] Following the death of the sensitive population, the surviving cells will have access to additional growth

space and more resources. This may promote the surviving population to grow in a manner it could not before, when the dominant population was present. This is called *competitive release*, which is an evolutionary mechanism by which the removal of a dominant species allows for a less dominant species to flourish. [175] This has e.g., been shown to happen during chemotherapy in prostate cancer. [176]

Efforts have been made to investigate the mechanisms by which cancer cells develop resistance to chemotherapy, but fundamentally the treatment will always select for cells that divide slowly, whether they are resistant or not. Some tumors may also be growing so fast that the mitotic frequency surpasses the rate by which cells die of chemotherapy. So, it is important to remember that not all cells surviving therapy have to be resistant. Rapidly dividing cells that are resistant are, however, a major problem. A possible way to remedy this problem would be to develop drugs that hinder the cells from using their resistance mechanisms, hinders them from hibernating or by combining chemotherapy with drugs specifically targeting cells that is starting to develop a hibernating phenotype.

Another common treatment strategy reserved for solid tumors is *surgery*, which in theory is an extinction event or severe mechanical bottleneck. Tumor cells that have escaped the tumor and metastasized will survive the localized extinction. Moreover, the surgery may alter the microenvironment in the area in a way that may be favorable or unfavorable for the surviving cells. [17, 18] Studies have also indicated that removal of the primary tumor can increase the division rate of metastases, possibly through the removal of suppressive signals from the primary tumor. [17]

Radiation therapy also represents a form of extinction therapy and is often used in combination with other therapeutic modalities. Extensive advances has been made in the field of radiation oncology in the last 50 years. [177] The treatment mainly results in DNA double strand breaks. If the damage is not adequately repaired, the cell dies. Studies have indicated that cancer cells exhibit slower DNA repair compared to normal cells. It is important to note that the radiation therapy itself induces DNA damage, potentially leading to additional genomic aberrations if the cell manages to survive. [19]

During the past decade, *targeted therapies* have gained immense popularity. These therapies are highly specific for certain cellular targets, and it was believed that they could act as a golden bullet, killing all cancer cells. Despite the initial interest, they have shown surprisingly little impact on overall long term survival, only prolonging survival with a couple of months. [20] The reason is that in almost all cases, there will be cells that does not display the target, or cells that display a different version of it, which the drug does not react to. This results in the selection of non-target displaying cells, and eventually treatment refractory disease. [20] There are, however, some exceptions such as NRTK-blockers, such as Gleevec, and kinase blockers in hematological cancers. [178]

In recent years, immunotherapies have gained increasing interest, where the goal is to harness the immune cells in the microenvironment to kill the tumor cells. These therapies have shown a surprisingly good effect on initial tumor shrinkage, particularly for malignant melanoma and hematological cancers. [179] A limitation of this method is that current therapies mainly target a single marker on the surface of the tumor cells. This results in a selection pressure for cells not having the marker for destruction. There is also an extensive heterogeneity regarding how patients respond to the treatment. Some patients even progress during treatment and some get vast side effects, such as activating autoimmune diseases, resulting in termination of the treatment. It has been shown that the concept of immune cold (tumors having few infiltrating immune cells) and warm (tumors having many infiltrating immune cells) tumors plays a role in whether the tumor responds to treatment or not, where immune warm tumors show better responses. [22, 179]

Novel therapeutic strategies that exploit evolutionary dynamics

There are several novel antitumoral therapeutic strategies that have been developed in the past 20 years (**Figure 14**). *Adaptive therapy* aims at treating cancer as a chronic disease, focusing on tumor size control. The approach operates on the premise that without therapy, resistant cells have a fitness disadvantage, whereas non-resistant cells have a fitness advantage. Upon the administration of treatment, the fitness of resistant cells rises, while the fitness of the sensitive cells decreases. The tumor is treated until it has shrunk to a certain degree, determined radiologically or through measurement of a biological biomarker in the blood. Upon halting therapy, the fitness of the sensitive cells rises relative to the resistant cells, allowing the sensitive cells to recover while the resistant population is kept in check. This dynamic interplay between the two cell populations is leveraged to extend the patient's lifespan by maintaining tumor control. The strategy also allows administering the minimum effective dose of chemotherapeutics, contributing to an improved quality of life. [180] Early clinical studies in prostate cancer have shown promising results with a significantly increased survival, along with the need for a lesser amount of chemotherapy, compared to standard of care, which minimizes side effects and increases the quality of life. [181]

In *sucker's gambit* antitumoral therapy is administered to shape the evolution of cancer resistance so that the surviving cancer cells become easier to target. This is similar to the *evolutionary double* bind where the tumor is treated with one drug which makes it more susceptible to treatment with another drug. The order in which the drugs are given will consequently have a major impact on the efficacy of the overall treatment. [182]

In *extinction therapy*, also denoted *first-strike second-strike therapy*, antitumoral agents are administered in order to mimic the dynamics of extinctions as seen in nature. The idea is to treat the tumor in a two-step process. In *the first strike*

treatment, the drug is given such that it diminishes the population size dramatically. The surviving population will be small, fragmented and have lower genetic diversity and spatial dispersion compared to the original population. At this stage, the population is less capable of withstanding additional stressors or environmental perturbations which the original population might not have been affected by. This is exploited during *the second strike*, where another treatment is given compared to the first strike, pushing the population over the extinction threshold or minimum viable population, resulting in an extinction vortex from which there is no return. Note that a good second-strike agent does not necessarily represent a good first strike agent and vice versa. Curing cancer is similar to trying to extinguish a species which is both geographically dispersed and extensively heterogeneous, which is very difficult. The species is also blended with species that we do not want to harm. There are several examples from nature where an initial perturbation, leading to a diminished population size, was followed by a series of seemingly independent and non-catastrophic insults, ultimately resulting in a complete extinction of the species. This unfortunate series of events happened to for example, the passenger pigeon, heath hen, the goats on the Galapagos Islands, and possibly also the dinosaurs after the asteroid hit the Earth. [183, 184] If the second strike in a tumor is not successful, a plethora of subclones may start to proliferate through a form of competitive release, similar to what has been seen in nature after mass extinctions. [185]

An ecology-based treatment strategy is the *target the public goods* strategy. Elements of the ecosystem that produce public goods are targeted to suppress the tumor in various ways. This can involve blocking the acquisition or production of growth factors, proangiogenic factors, and metabolites. *Metronomic therapy* is a type of protocol where anti-cancer drugs are administered in a lower dose than the maximum tolerated dose intermittently over a long time. This gives less severe side effects and is more cost effective. [186] Some recent studies have been performed where anoxic bacteria are used to treat cancer. They thrive in the anoxic environments at the center of the tumor and may be used to deliver drugs. [187]

Evolutionary game theory has gained increasing interest in the development of novel antitumoral treatment strategies. The tumor cells are conceptualized as players in a game where they compete for nutrients, survival, and proliferation. The theoretical framework can be used to model strategies and payoffs between different cell types, such as sensitive and resistant cells, in different environments and when subjected to various treatments. Game theory can thus aid in finding optimal drug combinations and developing novel strategies to steer the tumor and cure patients. [188]

Hence, several treatment strategies considering the evolutionary dynamics of cancer have been developed and are under development. Hopefully these strategies hold the possibility to prolong life, increase quality of life during treatment and cure patients that today die of their disease.

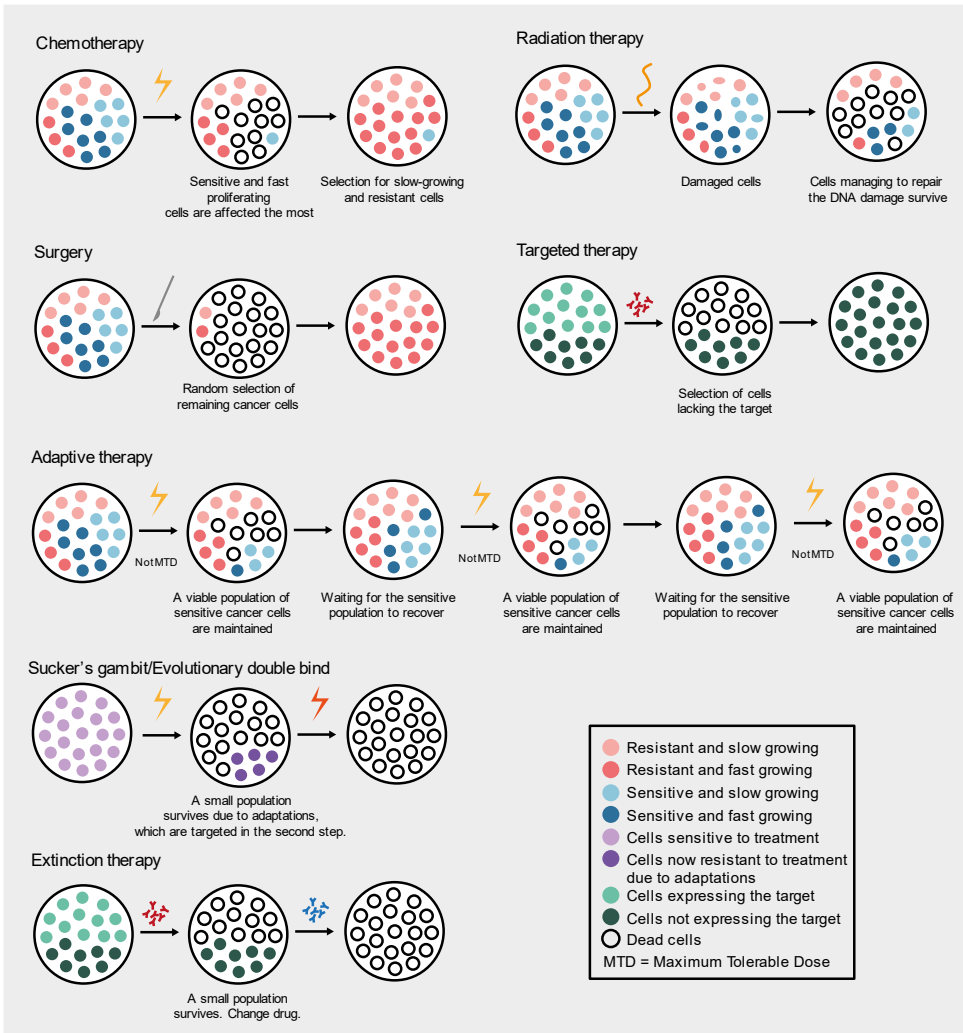


Figure 14. Antitumoral therapies in the context of evolution. Large circles represent the tumor at a certain time point. Each small circle represents 5 % of the tumor cells in that sample and is colored as specified in the legend.

Pediatric tumors

Pediatric tumors differ from adult tumors in several aspects. Firstly, tumor development is considerably less common in children than adults. Each year 350 children are diagnosed with cancer in Sweden, compared to 69 000 cases for adults (cancer i siffror 2023). Secondly, the types of cancers that affect children differ from those occurring in adults. The most common cancer subtypes in children are leukaemia, brain tumors, neuroblastoma, rhabdomyosarcoma, Wilms tumor and germ cell tumor. The latter three are rarely seen in adults. Conversely, the most common tumors in adults are lung cancer, breast cancer and prostate cancer, none of which typically manifest in children. Thirdly, pediatric tumors often exhibit very few somatic pathogenic mutations, and mainly harbour larger chromosomal aberrations. Fourthly, many pediatric tumors are believed to arise due to a faulty embryonal development. In contrast, tumors occurring in adults are thought to arise from an accumulation of somatic genetic aberrations and transcriptional cellular changes, sometimes influenced by environmental factors such as smoking and UV light exposition. For a detailed exploration of embryogenesis and its connection to pediatric cancers, see the following references. [189-194]

Neuroblastoma

Neuroblastoma (NB) is a pediatric tumor believed to result from a faulty embryonal development of the sympathetic nervous system. It often originates in the adrenal medulla or sympathetic ganglia. In Sweden, approximately 20 patients receive a NB diagnosis each year, with a median age at diagnosis of 18 months. Notably, patients under 18 months old at diagnosis exhibit significantly better overall survival [195] compared to children diagnosed later. [195] About 50 % of patients present with disseminated disease at the time of diagnosis, and NB accounts for a total of 15 % of all childhood cancer-related deaths. [196] No clear environmental causes for the development of NB have been identified, and only 1-2 % of the patients have an identified hereditary component, such as germline *ALK* or *PHOX2B* gain-of-function mutations. For the remaining patients, the cause is not identified. [197, 198]

In early studies of metaphase spreads of NB, researchers observed signs of a gene amplification and double minutes in a subset of NB patients. [87] Subsequent investigations revealed that the amplified region at 2p24 encompassed a novel oncogene, coined *MYCN*. [88] It encodes the N-MYC protein, located in the cell nucleus. It is a transcription factor, affecting genes related to cell cycle progression and differentiation. It has a basic helix-loop-helix (bHLH) domain that must bind to another bHLH protein, such as the protein Max, to be able to bind to DNA and initiate transcription of the target genes. [199] While N-MYC is typically highly expressed in the fetal brain and mesenchymal tissues, its expression in adult tissues is minimal. [200] In NB though, *MYCN*-amplification and associated increased protein expression is seen in 20 % of patients and is related to worse prognosis, disease progression and advanced disease stage. [90, 201] The amplicon appears in the form of double minutes (DMs), and studies of NB cell cultures indicate that a higher copy number confers a greater growth advantage compared to neuroblastoma cells with a lower copy number. During cell division the distribution of the DMs follow a skewed binomial distribution, allowing cell-to-cell variation of the copy numbers, and subsequent selection for specific numbers of copies. Interestingly, there appears to be some level of selection against very high amplicon numbers, as NB cells rarely exhibit more than 200 copies. [97]

The *MYCN* locus also harbours an antisense version of *MYCN* denoted *MYCNOS*. It is always co-amplified with *MYCN* and encodes the protein N-CYM which stabilizes N-MYC. In addition, N-MYC induces transcription of N-CYM, resulting in a positive feedback loop. There are other proteins also stabilizing N-MYC such as Aurora kinase A, a protein kinase that plays an important role in cell cycle regulation and mitosis. [202]

Another commonly affected gene in NB is *ALK* (anaplastic lymphoma kinase), a cell-surface receptor tyrosine kinase typically highly expressed in the developing embryonal and neonatal brain. [203] Germline mutations in *ALK* is the major cause of hereditary NB. [197] Somatic *ALK* mutations, cause a ligand-independent activation of the receptor, and is found in 14 % of high-risk NBs. [204] This activation causes increased PI3K signalling, resulting in the stabilization and elevated levels of N-MYC. Additionally, *ALK* mutations cause an upregulation of RET and RAS-signalling, subsequently activating the MAPK pathway, which is often activated in relapsed NB. [205, 206] Due to *ALK*'s proximity to the *MYCN* locus on 2p, it can, with or without mutations, be co-amplified, resulting in an increased expression of the protein and hence upregulation of its downstream pathways. [207]

Another recurrently affected gene is *LIN28B*, encoding a protein that binds small RNA molecules. *LIN28B* is highly expressed and important in developing tissues. Polymorphisms and subsequent misexpression of the *LIN28B* gene product increase N-MYC expression, stabilizes Aurora kinase A, promotes NB tumorigenesis and is related to high-risk NB. [208] [209] Other genes such as *TP53*, which is also a target of N-MYC, is rarely mutated in NB, but mutations in *TP53*

are related to worse prognosis. [210] *PHOX2B*, encoding a master regulator of neural crest development, is sometimes mutated in NB and has an increased expression in NB cells in some patients. It has also been implicated in some cases of hereditary NB [202]

A recent study suggests that *TERT* rearrangements, *ATRX* deletions and *MYCN* amplifications occur in mutually exclusive groups in high-risk NB. [211, 212] *TERT* encodes the protein telomerase reverse transcriptase which lengthens the telomeres, allowing cell division beyond the Hayflick limit. It is also a N-MYC target gene. Rearrangements of *TERT* are seen in 25 % of NB patients, resulting in an increased expression of the corresponding protein product, and is related to more aggressive tumors. [211] *ATRX* encodes a chromatin remodelling protein shown to repress alternative lengthening of telomeres (ALT). Inactivation of *ATRX* results in activation of the ALT program, with a lengthening of the telomeres as a result, allowing the cells to preserve their telomeres and continue dividing. [213] Loss-of-function genetic alterations of this gene is seen in 10 % of patients. Studies have indicated that all NBs require a way to activate *TERT*, either through *TERT* rearrangement, *MYCN* amplification (activating *TERT*) or by bypassing the checkpoint through *ATRX* loss-of-function alterations, resulting in ALT. [202, 212]

NB also harbours recurrent segmental copy number alterations. Gain of 17q is seen in > 50 % of NB patients, and loss of 1p is seen in 33 % of patients, both of which correlate with *MYCN* amplification and poor prognosis. [214] Loss of 11q is on the other hand inversely correlated to *MYCN* amplification and is seen in 33 % of high-risk patients and is associated with worse prognosis. [215] Losses of 17p and 6q have been found to be recurrent in relapsed NB. [206] Other detected alterations in NB are losses of 3p, 4p and 14q as well as gains of 1q and 2p. Approximately 20 % of NB tumors also display chromothripsis events. [216]

Transcriptional and epigenetic profiling of NB cell lines has revealed the presence of two different cell states, denoted as adrenergic (ADR) and mesenchymal (MES) cells. [217] ADR cells are lineage-committed sympathetic noradrenergic cells, while MES cells are immature neural crest cell like or undifferentiated mesenchymal like cells. Neural crest cells are multipotent stem cells usually found in proximity to the neural tube during embryogenesis. Mesenchymal cells are multipotent stromal cells. Cells in the MES cell state are more chemotherapy resistant *in vitro* and are enriched after therapy as well as in relapsed tumors, based on studies on biopsy samples from patients. [218] Recently we also discovered a similar pattern in PDX models. [219] Additionally, it has been shown that the ADR and MES cell states each possess their own super-enhancer (a genomic region with several enhancers in close proximity with a high level of transcription factor binding) landscape and super-enhancer-associated transcription factor network. [218] Each cell type has a core regulatory circuitry (CRC) composed of a set of genes involved in each cell state, identified both *in vitro* and in patient samples. [220] Rather than being two distinct binary cell states, the truth is probably a continuum of cell states, and it

has been shown that transitions between them is possible, hence the cell states are plastic. [218] Some researchers are arguing that there might be neuroblastoma cancer stem cells, [221] but they seem to overlap with the now well characterized ADR and MES cell states.

There are cases of spontaneous regression of neuroblastoma, even in a metastatic setting. [222] This feature is most often observed in small children < 18 months of age. Studies have seen that these tumors have a high expression of nerve growth factor (NGF), encoded by the gene *TRKA*. NGF usually regress as part of normal development. Since the tumor cells become dependent on NGF, they may regress in its absence. [223] It has also been shown that the tumor may regress if there are many immune cells in the tumor or if there is no TERT-expression, resulting in a telomere crisis, since neuroblastoma seems to need *TERT* activation to grow. [224]

In 50 % of the patients, NB presents in the adrenal medulla, but it can occur anywhere in the sympathetic nervous system. Tumors presenting in the adrenal medulla have a worse prognosis compared to those in the thorax. [225] The most common symptoms are pain, hypertension (caused by excretion of adrenalin and noradrenalin from the adrenal gland and tumor cells), diarrhea and a palpable abdominal resistance. Tumors arising in the ganglia next to the spine can cause spinal cord compression, while those in the neck can result in Horner's syndrome. In 2-3 % of patients the paraneoplastic disorder opsoclonus myoclonus syndrome may occur, characterized by rapid involuntary eye movements, muscle contractions, ataxia, and cognitive impairment. Blood plasma levels of normetadrenaline, metadrenaline and methoxythymine can be measured, which are elevated in NB patients. In 90 % of patients with NB dopamine, homovanillic acid (HVA) and vanillylmandelic acid (VMA) can be detected in the urine. [202] Neuroblastoma screening programs have been evaluated, but they resulted in overdiagnosis, i.e., detection of tumors that might have spontaneously regressed and did not decrease the incidence of advanced-stage disease. [226-228]

When neuroblastoma is suspected, the investigation typically involves imaging modalities such as an ultrasound, CT, or MRI. Subsequently, a biopsy of the tumor is taken, followed by histological assessment by a pathologist. The tumors typically present with small round blue cells. Additional analyses may include Fluorescence In Situ Hybridization (FISH) to detect *MYCN*-amplifications and DNA sequencing to identify mutations in e.g., *ALK*. The combination of these results guides the pathologist in establishing the diagnosis and risk assessment. [202]

Patients are subsequently staged according to the International Neuroblastoma Staging System (INSS) into Stage 1, 2A, 2B, 3, 4 and 4S, based on the extent of surgical excision and the presence of metastases. [229] The stage is combined with other prognostic information such as age and genomic alterations to group the patients into risk groups. This is done according to The International Neuroblastoma Risk Group (INRG) Classification system, into very low, low,

intermediate, and high-risk groups, and the stage dictates the treatment schedule. [230] Patients with very low and low-risk tumors make up 50 % of the patient cohort. [230] Patients < 1 year of age, with a tumor < 5 cm in diameter in the adrenal gland only undergo clinical observation since the tumor may spontaneously regress. Otherwise, patients are usually treated with surgery alone, but if the patient presents with symptoms, they are usually administered a short chemotherapy protocol. The 5-year survival for very low and low-risk NB is > 90 %. [231] The intermediate risk group makes up 10 % of the patients [230] and they are treated with surgery and chemotherapy for 2-8 cycles and have a 5-year survival of almost 90 %. [231] Patients with high-risk NB constitute 36 % of the patients [230] and have a dismal 5-year survival of merely 50 %. [231] These patients often obtain a multimodal induction therapy denoted COJEC (cisplatin, vincristine, carboplatin, etoposide, and cyclophosphamide) followed by surgery. This can be followed by *myeloablative therapy*, which is a treatment approach where the cells in the bone marrow are killed before stem cell transplantation. This is achieved using chemotherapeutic agents including melphalan followed by autologous homologous stem cell transplantations (AHSCT). Studies have shown that administering *isoretinoin*, which reduces proliferation and induces differentiation of NB cells, after myeloablative therapy and AHSCT increase the event free, but not the overall survival. [232] Studies have subsequently shown that a combination of both *anti-GD2 antibody*, *IL-2*, *GM-CSF* and isoretinoin after myeloablative therapy and AHSCT improve outcome compared to administering isoretinoin alone. This is therefore given as a standard therapy in North America after myeloablative therapy. [233]

In relapsed NB the 5-year survival is essentially zero. Patients might be treated with additional rounds of chemotherapy or be enrolled in a clinical trial. Examples include 131-I-mIBG, a radiopharmaceutical, which has been used in pilot studies, [234] ALK-inhibitors [235], aurora kinase inhibitors and immune checkpoint inhibitors. [202, 236] Novel treatment strategies are needed to cure the patients that today die of their disease.

Wilms tumor

Wilms tumor (WT), also denoted *nephroblastoma*, is the most common pediatric kidney tumor, accounting for 6 % of all childhood cancer. [237] Approximately 5-10 patients are diagnosed with WT each year in Sweden with a peak at 3 years of age. In 5-10 % of the cases the tumors are bilateral or multifocal, presenting at an earlier age compared to unilateral tumors. [69, 238] Around 15 % of patients with WT have an underlying predisposition such as genetic alterations encompassing *WT1*, *CTR9*, *REST* or *TRIM28*, changes in the familial WT loci *FWT1* (17q12-q21) and *FWT2* (19q13.5) [239] or a genetic syndrome like Wilms tumor-aniridia syndrome

(WAGR) (11p13;*WT1* deletion), Denys-Drash syndrome (defect *WT1*), Beckwith-Wiedemann syndrome (11p15.5;*WT2* deletion) and Frasier syndrome. [240] [238]

WT is thought to develop from an aberrant renal development. *Nephrogenic rests* are defined as groups of normal embryonic metanephric cells persisting beyond the gestational age of 36 weeks. Metanephric cells are cells that are part of the normal development of the fetal kidney. Approximately 1 % of infants still have nephrogenic rests at birth but they usually regress shortly thereafter, while, strikingly, 40 % of WT patients with unilateral and 90 % with bilateral tumors still have them. These nephrogenic rests have consequently been proposed to act as precursor lesions for WT development, suggesting a connection to renal development. [241] Studies have since shown that WT can arise from premalignant clonal expansions in the kidney. [242] As the tumor forms, an extensive branching evolution and intratumoral genetic heterogeneity can develop. [23, 170]

Around 10 % of the sporadic WTs have inactivating mutations of the *WT1* gene at the 11p13 locus, while 25 % instead have an LOH and 9 % a homozygous deletion at this locus. *WT1* encodes several different isoforms of a zinc finger transcription factor, containing four zinc-finger domains. The protein regulates cell growth, differentiation, and apoptosis and is expressed in the kidney, gonads, spleen, and mesothelium. It has been shown to play a central role in renal and gonadal embryogenesis and disruption of the gene results in genitourinary developmental abnormalities, [191] and has also been detected in nephrogenic rests. [243] Mutations in the *CTNNB1* gene is seen in around 15 % of WT patients and a study showed that 19 out of 20 tumors harboring activating mutations in the *CTNNB1* gene on 3p22.1 also had mutations in *WT1*. [244, 245] *CTNNB1* encodes the protein β-catenin, which is a transcriptional co-activator and constitutes an integral part of the Wnt signalling pathway, through promoting target gene transcription. The Wnt signalling pathway is essential for normal renal development. It also results in a dysregulation of the mesenchymal-epithelial transition. [239] Hence, both *WT1* and *CTNNB1* play central roles in normal renal development. Furthermore, it has been shown that genes involved in the first contact between the ureteric bud and the metanephric mesenchyme, an important step in the development of the kidney, are overexpressed in WT and that genes often expressed at later stages in the renal development are conversely downregulated. [246-248]

Another locus that is recurrently affected is *WT2* (11p15). The region is divided into two imprinted domains: *IGF2*/*H19* and *KIP2*/*LIT1*. Imprinting is a type of epigenetic inheritance where the gene expression is dependent on the genetic sex of the individual. *IGF2* encodes a growth factor that is important for renal development and *H19* encodes RNA molecules that may act as a tumor suppressor. [249] It has been shown that this region is affected by LOHs in 29 % of patients and that loss of imprinting of *IGF2* occurs in 40 % of WT patients. [245]

This region is affected in patients with the genetic syndrome Beckwith-Wiedemann, and they display an increased risk of WT development.

AMER1 deletions or mutations are seen in 20 % of WT cases, and typically occur late in the evolution of the tumor. This gene is part of the β-catenin-destruction complex and is involved in kidney development. [239] Also *MYCN* gains can be seen in 6 % of patients and is associated with anaplasia and poor outcome.[250, 251]

TP53 (17p13) mutations are found in 5 % of patients with WT and their presence are associated with tumor progression. Mutations are particularly correlated to *anaplastic* WT where 75 % of patients display disrupted *TP53*. [252] Anaplasia is defined by three criteria: 1) the presence of nuclear enlargement $\geq$ 3-fold, 2) hyperchromasia due to extra chromosomes and, 3) enlarged abnormal mitotic figures. Anaplasia is divided into being *focal*, located only to one or a few localized areas in the primary tumor, or *diffuse*, multifocally spread throughout the tumor. [253] WT with focal anaplasia is considered an intermediate-risk disease while diffuse anaplastic WT (DAWT) is classified as a high-risk histology based on the guidelines from Société Internationale d'Oncologie Pédiatrique Renal Tumours Study Group (SIOP-RTSG). Patients with DAWT have a worse prognosis compared to other subtypes of WT, despite a more intense treatment. [254] It was recently suggested that the chemoresistance observed in these patients may be partly explained by the high proliferative capability of anaplastic cells and the fact that these cells also have a high tolerance to CNAs and double strand breaks (DSBs). [255] Mutations in *TP53* are late events in the evolutionary history of WT, and thus appear regionally. Interestingly, almost all areas with anaplasia have either mutations or LOHs of *TP53*, but these changes are not limited to anaplastic tumors. [255] Anaplastic WTs seem to appear gradually, through an initial LOH of *TP53*, followed by a mutation of the *TP53* gene, after which full-scale anaplasia develops, indicating that it grows more aggressive in time. [255]

Mutations in *SIX1* and *SIX2*, two other genes with importance for renal development, have also been detected in WT. Mutations in these genes are detected more frequently in blastemal tumors compared to necrotic or regressive tumors. [239] Finally, there are several larger chromosomal aberrations that are frequent in WT such as 1p-, 1q+, 2+, 7p-, 7q+, 8+, 12+, 13+, 16q- and 22q-. [256]

Patients with WT may present with various symptoms at the time of diagnosis such as a painless resistance in the abdomen, decreased appetite, weight loss, hematuria (30 % of patients), hypertension (25 % of patients), pain, varicocele, fever, malaise, tumor rupture resulting in instant pain, malformations, hypercalcemia, decreased levels of vWF (von Willebrand Factor) and lung metastases giving rise to symptoms. It is diagnosed using radiological techniques such as a CT-scan or MRI, and the diagnosis is set based on the radiological image alone. No biopsy is needed before the treatment is initiated according to the guidelines used in Sweden.

Patients older than 6 months undergo a preoperative chemotherapeutic treatment protocol including the agents: vincristine weekly and actinomycin-D biweekly for 27 weeks, with the addition of doxorubicin in the metastatic setting. After this the tumor is surgically removed, at which point the pathologist investigate the tumor further through histological staining and genomic analyses. Children younger than 6 months undergo biopsy directly to rule out the possibility of another diagnosis and thereby avoiding unnecessary treatment. [257]

WTs display a characteristic histological pattern containing three different tissue components: blastemal, epithelial, and stromal components. The blastemal component consists of round/oval cells with a large nucleus and a sparse amount of cytoplasm. It often displays mesenchymal features, and patients with a blastemal predominant subtype after surgery have a worse outcome. [258] The epithelial component often represents primitive glomeruli, comma, or s shaped bodies. Finally, stromal cells are similar to fibroblasts in their appearance, but they can display differentiated structures such as smooth muscle, skeletal muscle, or neural elements, raising the suspicion that WT arises from undifferentiated renal mesenchyme. [191, 259]

Depending on how the tumor is risk classified and staged, the patient might receive additional chemotherapy or radiation therapy. [257, 260] Treatment has improved significantly in the past century and the disease today has a 5-year survival over 90 %, but the prognosis of patients with DAWT remains around 58 % (https://nccrexplorer.ccdi.cancer.gov). [254]

Rhabdomyosarcoma

Rhabdomyosarcoma (RMS) is the most common soft tissue sarcoma in children, thought to arise from skeletal myoblast-like cells. It accounts for approximately 7 % of all cancers in children and has a 5-year survival of 60-70 % (https://nccrexplorer.ccdi.cancer.gov/), all subtypes included. [261, 262] The incidence for children aged below 14 years old in Sweden is approximately 4.9 per million, resulting in 5-10 cases each year. [263] There has not been any GWAS studies for RMS, but it has been reported that germline susceptibility syndromes such as the Li-Fraumeni syndrome, neurofibromatosis type 1, Beckwith-Wiedemann syndrome and DICER1 syndrome are associated with an increased risk of developing RMS, accounting for approximately 7 % of cases. [264, 265]

Based on histopathological features, several different subtypes of RMS have been identified. The two major groups are *embryonal rhabdomyosarcoma* (ERMS) and *alveolar rhabdomyosarcoma* (ARMS). [266] ERMS accounts for 60-70 % of all RMS cases and is most frequent in early childhood around 0-4 years of age, but

some data have indicated that the age distribution is bimodal with a second peak in the early teens. [261] ERMS has a better prognosis than ARMS, with a 5-year survival of 66 % compared to 39 %. The tumors most often appear in the head and neck or in the genitourinary areas. [261] Common genetic alterations are whole chromosome gains and losses and 11p LOH. [267] ARMS on the other hand accounts for 20-25 % of all RMS-cases, [261] and has an even incidence across ages. The tumors are more often located to the extremities compared to ERMS. [261, 268] ARMS can be further subdivided into *fusion-positive RMS* (FPRMS) and *fusion-negative RMS* (FNRMS).

FPRMS accounts for 80 % of all ARMS cases and the patients are marginally older than those with FNRMS. [261, 269] The term *fusion-positive* refers to the tumors displaying one of two fusion genes: *PAX3-FOXO1* t(2;13)(q35;q14) or *PAX7-FOXO1* t(1;13)(p36;q14), seen in 60 % and 20 % of ARMS patients respectively. [269] [270, 271] *PAX3* and *PAX7* encode paired box protein 3 and 7 respectively. They are transcription factors usually expressed in skeletal muscle progenitor cells, and their target genes are related to skeletal muscle development. [272] *FOXO1* encodes forkhead box protein O1, which also is a transcription factor playing an important role in skeletal muscle differentiation. [273] The gene fusions result in the transcription of a fusion protein, with enhanced expression and transcriptional activity compared to its wild type proteins separately. The expression of the PAX3-FOXO1 protein is increased due to a copy-number independent process, while PAX7-FOXO1 is increased due to an amplification of the segment.[274] PAX3-FOXO1 is stabilized by the proteins PLK1 and KAT2B, and inhibition of these results in degradation or downregulation of the fusion protein respectively. [275, 276] Both fusion proteins dysregulate several cellular pathways and affect multiple downstream target genes. PAX3-FOXO1 establishes a *myoblastic super enhancer circuit* in collaboration with *MYOD1*, *MYCN*, *MYOG*, *ALK*, *FGFR* as well as chromatin remodelling proteins such as BRD4 and CHD4 driving ARMS cancerogenesis. [277] Some patients with ARMS have other fusion proteins such as PAX3-NCOA1, but their clinical relevance is unclear. [278] FNRMS accounts for 20 % of ARMS cases [269] and is more like ERMS in the sense that it has more whole chromosome changes than FPRMS, point mutations and 11p15.5 losses. [279-281]

There exist additional RMS subtypes that are less common, such as *spindle cell/sclerosing RMS*. It most often arises in paratesticular sites or in the head and neck and has a 5-year survival of 85 % in pediatric patients. [262] *Botryoid RMS*, mainly occurring in children < 1 years of age, frequently appears in mucosa lined tissues such as the vagina, bladder, or nasopharynx. The prognosis is generally good. *Pleomorphic/anaplastic RMS* is a heterogeneous and rare subtype that mainly arises in the extremities in adults. Histologically it displays three types of cells: classic RMS cells, round cells, or spindle cells. It is associated with a poor prognosis with a 5-year survival of 0 % for metastatic disease and 30 % for

localized disease at diagnosis. [282] There are also RMS cases which display a *mixed type* histology or *ganglionic* differentiation.

It is thought that RMS arises from an aberrant myogenesis, but the cell of origin is not well characterized since both myoblasts and non-myogenic cells are able to initiate the disease. [283] Genes implicated in myogenesis such as *MYOD1*, *MYOG* and *MYLG* are frequently involved in RMS pathogenesis, affecting developmental pathways such as causing Wnt activation, Hedgehog activation, Notch activation and Hippo signalling suppression. [284, 285]

Some frequently mutated genes in RMS are *CDKN2A*, *TP53* and *PTEN*. [267] In addition, alterations affecting the RAS-PI3K axis have been shown to be present in 93 % of all RMS patients. [267] There are also various receptor tyrosine kinases that can activate the same pathway, such as *FGFR4*, which has activating mutations in 7 % of patients with FNRMS. [286] FPRMS often has an increased expression of FGFR4 and 2, through increased transcription due to the PAX3-FOXO1.[277] Other RTKs of importance are the IGF1R and PDGF. [264] LOH of 11p15.5 and mutations in the RAS pathway genes in FNRMS and the *PAX3/7-FOXO1* fusion event has been shown through phylogenetic analyses, to occur early in the evolutionary history of the tumors. [170, 287] Together the mutations, copy number, developmental pathway effects and the fusion proteins drive the disease.

RMS usually presents as a painless growing resistance or with symptoms caused by a mass effect due to the tumor growth. [261] Metastases are present in 20 % of patients at diagnosis with the most common sites being the bone marrow, lung, and skeleton. [262] RMS is diagnosed through histopathological assessment and molecular testing for the absence or presence of *FOXO1* gene fusions. ARMS is generally composed of small, densely packed, round cells, with lining septations having some resemblance with pulmonary alveoli. ERMS is on the other hand very stroma-rich, contains rhabdomyoblasts and display no alveolar pattern. Some tumors may display a mixed pattern. The tumor may also express skeletal muscle proteins such as myosin, myoglobin, MYOD1, MYOG, desmin and Z band protein, which may aid the diagnosis but are not pathognomonic. The RMS fusion genes can be detected from clinical biopsies using RT-PCR and FISH, which is clinically relevant to subgroup the tumor. [264]

Most patients are treated with a combination of surgery, radiotherapy, and chemotherapy. In Europe, the treatment regimens are decided by the *European paediatric Soft Tissue Sarcoma Study Group* (EpSSG), stratifying patients as either low, standard, high or very high risk. Low-risk patients are treated with VA (vincristine and actinomycin D), while for patients with standard, high-risk and very high risk obtain the chemotherapy protocol IVA (ifosfamide, vincristine, actinomycin D). [288] Patients with high-risk and very high-risk tumors may receive an addition of doxorubicin, vinorelbine and cyclophosphamide. [289] Recurrent RMS

has a very dismal prognosis and novel treatment strategies are needed for this patient group. So far, CAR-T cell therapy have shown promising results in one case of metastatic, refractory and relapsed rhabdomyosarcoma. [290, 291]

Germ Cell Tumors

Germ cell tumors (GCTs) are a group of neoplasia derived from germ cells, which are the cells that normally give rise to the gametes of an organism. GCTs most often arise in the testes or ovaries, known as gonads, but can also occur at other anatomical sites due to errors in the migration of germ cells during embryogenesis. [292] Examples of such extragonadal sites include the brain (particularly the pineal gland), mediastinum, the oral cavity, neck, and pelvis. [194] There is an ongoing debate regarding whether some extragonadal GCTs may represent metastases from a regressed primary tumor in the gonads. [293] This hypothesis is supported by the observation that patients with extragonadal GCTs generally have a worse 5-year survival than patients with gonadal GCTs, as expected for metastatic disease. [294] GCTs are rare and only account for 3.5 % of all childhood cancer cases in children < 15 years of age, but this proportion rises to 16 % for adolescents aged 15-19 years. [294]

The symptoms of GCTs depend on the subtype and location of the tumor. Intracranial tumors can present with vomiting, nausea, neurological symptoms, and headaches. Tumors in the testes may manifest as a palpable mass, while ovarian tumors may cause abdominal pain, a palpable mass, precocious puberty, hirsutism and in post pubertal patients, amenorrhea, and menorrhagia. If there is suspicion of a GCT the child may undergo a CT scan or ultrasound, followed by a biopsy if a tumor is detected. [295]

GCTs are a heterogeneous group of neoplasms comprising several different subtypes. Diagnosis is based on the histopathological classification of tumor biopsy samples. Each subtype has its own age distribution, treatment plan and prognosis. [296] Some examples of GCT subtypes are: Germinoma (dysgerminoma and seminoma), embryonal carcinoma, endodermal sinus tumor (yolk sac tumor), choriocarcinoma, teratoma, polyembryoma, gonadoblastoma and mixed type (may contain both malignant and benign components). [295] The following sections will cover the subtypes included in paper V of this thesis.

Malignant juvenile granulosa cell tumor

Granulosa cell tumors are most common in patients 50-55 years old, accounting for 95 % of all granulosa cell tumors. There are, however, a juvenile subtype, constituting 5 % of all granulosa cell tumors, which is similar to the adult type but

is histologically distinct. [297, 298] The incidence of juvenile granulosa cell tumor is approximately 0.4-1.7 cases per 100 000 women. [299]

The tumors arise from granulosa cells, excrete the hormone oestrogen and is consequently categorized as a functioning tumor. The most common symptom is abdominal pain. If the patient is prepubertal the patient may present with an early onset of puberty. Patients post puberty may suffer from menometrorrhagia (heavy and irregular menstruation). The tumors most often arise in the ovaries but there are rare cases of it appearing in the testes, in which cases it is usually benign. [300]

The enzyme AKT1 plays a role in these tumors. Duplications of the chromosomal segment covering the corresponding gene has been identified in > 60 % of patients with malignant juvenile granulosa cell tumor < 15 years of age. The tumors not displaying duplications, instead have point mutations of the gene. AKT1 is a serine/threonine kinase that regulates several processes including cell survival, proliferation, and growth. [301] There is a lack of studies of the CNA and mutational landscape of juvenile granulosa cell tumors. One study suggested that trisomy 12 and 14 were frequent in granulosa cell tumors overall. Only three juvenile tumors were included in the study, of which one displayed trisomy 12. The other two tumors were completely diploid. [302]

Diagnosis is based on histopathological assessment of a biopsy from the tumor. The tumors usually stain for the hormone inhibin. The cells have a moderate-abundant eosinophilic vacuolated cytoplasm, increased mitotic activity, and display disorganized follicles. Sometimes there are also tumor giant cells. [297, 303]

Treatment often includes surgical removal of the primary tumor, sometimes including unilateral or bilateral salpingo-oophorectomy. The latter can be performed with or without hysterectomy. This is followed by adjuvant chemotherapy, most often using the chemotherapeutic protocol BEP (bleomycin, etoposide, and cisplatin). [298, 304] Juvenile granulosa cell tumors generally have an excellent prognosis with a 5-year survival rate of 92 % if it is in stage I, but merely 0-22 % in advanced stage. [304]

Teratoma

A teratoma is a tumor that can contain tissues from all three germ layers (ectoderm, endoderm and mesoderm). They are the result of an abnormal embryonal development of pluripotent cells. The tumors may contain epidermis, sweat and sebaceous glands, teeth, hair follicles and neuroectodermal derivatives. The tumors are usually benign but there are cases of malignant transformation or teratomas being part of a mixed germ cell tumor also containing malignant elements. [295, 305]

The tumor may cause pain, a mass effect and vascular steal. Acute symptoms can arise if the tumor haemorrhages or ruptures. Teratomas are treated using surgical removal with the addition of chemotherapy if it is categorised as malignant. The 5-year overall survival rate is > 90 %. [305]

Embryonal carcinoma

Embryonal carcinomas are aggressive germ cell tumors. Around 10-40 % of patients have metastatic disease at diagnosis. The median age at diagnosis is 30 years and it is very rare in children. It most often appears in the ovaries or testis, but can also arise in the brain, primarily in the pineal gland and third ventricle. [306] The embryonal carcinomas are rare and make up 3 % of all ovarian, 10 % of all testicular and 5 % of all intracranial germ cell tumors. [306-308]

The tumors are often surgically removed and followed by chemotherapy. A pure embryonal carcinoma in the brain has a 3-year overall survival rate of 27 %. If it is mixed with germinoma or teratoma it is approximately 70 %. [309] There are currently no overarching studies of the survival rates for ovarian and testicular embryonal carcinoma in children.

Treatment resistance

Despite considerable advancements in cancer therapy, avoiding treatment resistance and relapse continues to pose a substantial challenge in the quest to cure cancer patients. [310] Treatment resistance refers to a reduced sensitivity or complete lack of response of cancer cells to therapeutic interventions, such as chemotherapy, radiation therapy, targeted therapies, or immunotherapies. [311] Understanding the mechanisms underlying the development of treatment resistance is essential for the development of novel treatment strategies that can improve patient outcomes.

Much research has been dedicated to address this challenge. The adoption of combination chemotherapy has improved survival remarkably since its introduction 50 years ago. Combination chemotherapy was initially inspired by approaches used for tuberculosis treatment. [312] Despite the development of increasingly intricate treatment schedules, survival rates have plateaued in the last two decades for the majority of tumor types. [311] Chemotherapy treatment eliminates rapidly proliferating cells and resistance to these agents can arise through various mechanisms. The cancer cells can utilize drug efflux pumps where the chemotherapeutic drug is actively pumped out of the cell, consequently decreasing its intracellular concentration and effectiveness. [313] Cells may also hinder the drug from entering the cell, [314] or upregulate enzymes causing an increased metabolism of the drug. The cancer cells may also enhance their DNA repair capability or evade death signals, thus hindering apoptosis. [315] Furthermore, an altered expression of certain signalling pathways, epigenetic changes or mutations may result in various resistance mechanisms. [316, 317] Finally, some cells enter a non-proliferating dormant state, evading the effect of the chemotherapeutic agents. [318, 319] Similarly the hypothesized cancer stem cells have been implicated to be treatment resistant, explained by that they are usually not highly proliferative. When the treatment is halted, they may, however, continue to self-renew and start a regrowth of the tumor mass, causing a relapse. [58]

Advancements in cell biology and biotechnology have enabled the exploration of somatic genetic alterations, mRNA expressions, methylation profiles and neo-antigens, spiring a plethora of novel treatment strategies to further increase survival. Targeted therapies against BCR-ABL, HER2 and EGFR as well as

immunotherapy have shown remarkable responses in patients. [178] [179] Treatment resistance does, however, still pose a major problem. [20, 22] Administration of targeted therapies and immunotherapies, eliminates cells having the therapeutic target while selecting for non-target-displaying cells. Since a target almost never exists in all cancer cells, the treatments often have an initial effect, followed by a rapid regrowth. Resistance may also appear through mutations in signalling pathways, making the pathway independent of receptor activation. Hence, blockade of the receptor using an antibody, will not affect the signalling. [20]

A larger tumor size and higher tumor heterogeneity increases the risk of treatment resistance, [320, 321] which could possibly be explained by providing a greater subclonal diversity, setting the stage for resistance and relapse. There is a debate regarding when the resistant cells appear. Some studies suggest that resistant cells are present in the tumor already before administration of the drug, while others believe they are induced by the treatment through the appearance of genetic alterations or through phenotype switching. [176, 311, 322-324]

The tumor microenvironment also plays a role in treatment resistance. It can shield the cancer cells from the therapeutic agents, resulting in a lower concentration reaching the cancer cells and an inferior treatment response. Some tumors are poorly vascularized, such as pancreatic tumors, making it difficult to administer the chemotherapeutic drug to all cells. The poor vascularization may also create a pro-tumorigenic hypoxic environment. [325] Tumor cells can evade the immune system by e.g., downregulating surface molecules that could be recognized by the immune system and cause destruction of the tumor cells. In addition, tumor cells may also reside in anatomical sites where it is harder for both immune cells and certain drugs to reach them. Examples of such anatomical sites are the central nervous system, testes, and the peritoneum. These tumor cells residing here might result in a relapse, if not extra measures are done to eradicate them such as intraperitoneal or intrathecal chemotherapy (administered to the fluid around the brain and spinal cord). [311]

Numerous strategies have been proposed to tackle the issue of treatment resistance. Efflux pump inhibitors have unfortunately, so far, proven ineffective and displayed high toxicity. [313] Cancer stem cells are believed by some researchers to be responsible for treatment resistance and relapse. Efforts are therefore being made to identify, classify, and target these cells to hinder such development. [58] Furthermore, there has been an increasing interest in using combinations of targeted cancer therapies to circumvent treatment resistance, since it is less likely that a cell has developed treatment resistance against several different agents. [326] There has been little attention paid so far to the role of dose scheduling in targeted therapies, posing an attractive way of decreasing toxicity and optimizing treatment response. In addition, mutations in *MYC*, *MYCN*, *RAS*, *KRAS*, *BRAF* and *TP53*, the most affected oncogenes and tumor suppressor genes, remain undruggable. Continuing efforts are being made to target them. [327]

There is also evidence that cancer cells are plastic and can switch between different cell states, denoted phenotype switching. This can allow the cells to e.g., switch between being treatment sensitive and resistant depending on external cues. Targeting the phenotype switching, hindering cancer cells from switching to the treatment resistant state could be used as a treatment strategy. [324, 328]

Right now, we are entering an era of personalized medicine. Traditionally, patients with a specific diagnosis have received standard of care treatment. The shift towards personalized medicine, would, theoretically, enable the design of a treatment protocol customized for an individual patient's tumor. Such a treatment protocol could include a combination of different targeted therapeutic agents based on the genetic and phenotypic profiling of the tumor. Ideally, it would allow for adaptive treatment strategies, where the physician could adjust the treatment plan based on the changing molecular profile of the tumor or to the emergence of resistance. [311]

Finally, early detection of a primary cancer or a relapse remains an attractive strategy to combat treatment resistance. Studies have shown that early detection as well as providing adjuvant treatment decrease the risk of relapse. [311] When the number of cells is small and the diversity low, it is possible to exploit the vulnerabilities of small populations, preventing the development of full-blown resistance and tumor regrowth. There are currently several cancer screening programs such as mammography for breast cancer, HPV tests for cervical cancer, and detection of blood in the stool for colorectal cancer. They have, however, been getting critique for having low sensitivity and specificity. In addition, more patients may be treated for benign tumors, which would not have manifested symptoms. Another critique to screening programs is that if the time points for the screening are too far apart, tumors may arise and give symptoms in between these time points, in which case the screening has not helped to identify the tumor early. Some studies have also indicated that the screening programs do not increase survival as much as was initially expected. [329] Another approach for early detection is to detect circulating tumor cell DNA (ctDNA) in the blood, opening the possibility for easier cancer screening as well as disease monitoring. To detect the ctDNA, there must be a significant amount of DNA present, which might result false negatives. It may also be difficult to decipher from where in the body the ctDNA originated, but there are methods being developed, relying on epigenetic profiles to identify the affected organ. [311] The usage of ctDNA could aid both early detection of cancer before symptoms arise as well as disease monitoring for relapses.

To conclude, treatment resistance remains challenging. There is, however, much ongoing research on how to combat this difficult clinical situation.

Treatment resistance in neuroblastoma

Approximately 50 % of patients with high-risk NB do not respond to first-line treatment or is subject to a relapse within two years. Treatment resistance thus still constitutes a major problem in the quest of curing these patients.

There are several mechanisms used by NB cells to develop resistance. NB cells can increase their expression of the ATP-binding cassette transporter family proteins. This is a group of efflux pumps, associated with resistance to several chemotherapeutic agents across various cancers, including neuroblastoma. [313, 330] Some of these genes are targets of N-MYC, whose corresponding gene, *MYCN*, is amplified in a subset of NB tumors. Inhibiting these efflux pumps might pose an attractive treatment strategy. Efforts creating drugs targeting efflux pumps has, however, so far shown poor effect in preclinical studies and cause extensive toxicity. [331, 332]

Chemotherapy has been shown to induce autophagy, where the cell digests parts of itself, in NB cells both *in vitro* and *in vivo*. By administering hydroxychloroquine, it is possible to inhibit the autophagy, sensitizing NB cells to chemotherapy. [333]

Several targeted treatment strategies have been developed against NB such as ALK and AURORA A inhibitors. They do, however, face the same problem as in adult cancers i.e., that not all cancer cells exhibit the target protein or that some cells obtain mutations resulting in the target protein changing shape, not allowing the drug to bind, thus selecting for non-target displaying cells. While targeting N-MYC is an appealing strategy, its pursuit has been challenging primarily due to its complex chemical structure, with no single drug currently undergoing clinical trials. [334] Continuous research endeavors are ongoing to explore new druggable target and develop new drugs. As we approach the era of personalized medicine, the potential for combining multiple targeted therapies tailored to individual patients' diseases becomes increasingly plausible. [20, 335]

Studies have revealed two distinct cell identities in NB, a sympathetic noradrenergic cell state (ADR) and a neural crest like mesenchymal cell state (MES). [218, 220] The MES cells have been shown to be more resistant to chemotherapy *in vitro* and to be enriched after chemotherapy treatment and in relapses. [218] These states are plastic such that NB cells can transition between the ADR and MES phenotype depending on external selection pressures. [328] Some researchers believe that there are cancer stem cells in neuroblastoma, and that these might explain treatment resistance and relapse. It was shown that NB tumors with poor prognosis harboured a population of cells having markers suggesting an undifferentiated stem cell state. It is somewhat unclear whether these cells are what today is referred to as cells with the MES phenotype, or if it really is a distinct subgroup of cells. [57, 336, 337] Targeting the MES phenotype could be a possible way to tackle treatment resistance in NB.

MicroRNAs can either inhibit translation of mRNA or accelerate its degradation. A misexpression of microRNAs has been detected in NB and is associated with poor outcome. N-MYC can modulate the expression of microRNAs. [338] Another postulated mechanism for treatment resistance is the transfer of small extracellular vesicles (EVs) and exosomes. [339] Exosomes from *MYCN* amplified NB cells can induce migration and resistance in non-*MYCN* amplified NB cells. [340] Exosome-mediated transfer of microRNAs between NB cells and monocytes give rise to cisplatin resistance. [341] Anti-GD2 monoclonal antibody immunotherapy has significantly improved the overall survival in NB. Many patients treated with anti-GD2 monoclonal antibodies do, however, not respond to treatment or develop resistance. In a study it was found that EVs attenuated the efficacy of the treatment *in vivo*. They also modulated the tumor microenvironment to make it more immunosuppressive. By inhibiting these EVs the efficacy of the anti-GD2 treatment could be enhanced *in vivo*. [342]

In conclusion, NB faces many of the same challenges posed by resistance in adult cancer. Significant efforts are being conducted on elucidating the mechanisms by which treatment resistance develops, and how these mechanisms could be targeted to hinder resistance development and increase survival in these patients.

The metastatic process

Metastases are secondary tumors resulting from the migration of cancer cells from the primary tumor, establishing themselves in a different anatomical location. They can manifest in any part of the body, but certain cancer types exhibit a tendency to spread to specific sites, denoted *organotropism*. [343] Metastatic spread is a complex clinical scenario where neither radiation therapy nor surgery are feasible curative treatment options. Furthermore, the situation is often exacerbated by tumors displaying resistance to chemotherapy. Efforts have been made to identify the genetic drivers of metastasis, aiming to target them through precision oncology. However, so far, the search has not been fruitful, and no single recurrent genetic alteration or group of genetic alterations has been shown to universally drive metastasis in patients. [104, 344] Furthermore, there is an extensive intertumoral heterogeneity between different metastases and relative to the primary tumor, further complicating treatment. [11] To cure the patients with metastatic disease, that today die of their disease, novel treatment approaches are needed, necessitating a deeper understanding of the metastatic process.

The formation of a metastasis is a multistep process (**Figure 15**). Initially, cancer cells at the primary site detach from the surrounding cells and invade the surrounding tissue, followed by intravasation into the lymph and blood vessels. Subsequently, the cells must survive the tumultuous journey within the vessels. Upon reaching another organ, the cells need to extravasate, colonize and establish a new tumor at the site. This intricate process has been the subject of extensive research aimed at understanding each step, and to determine whether any of these steps could be targeted for therapeutic intervention.

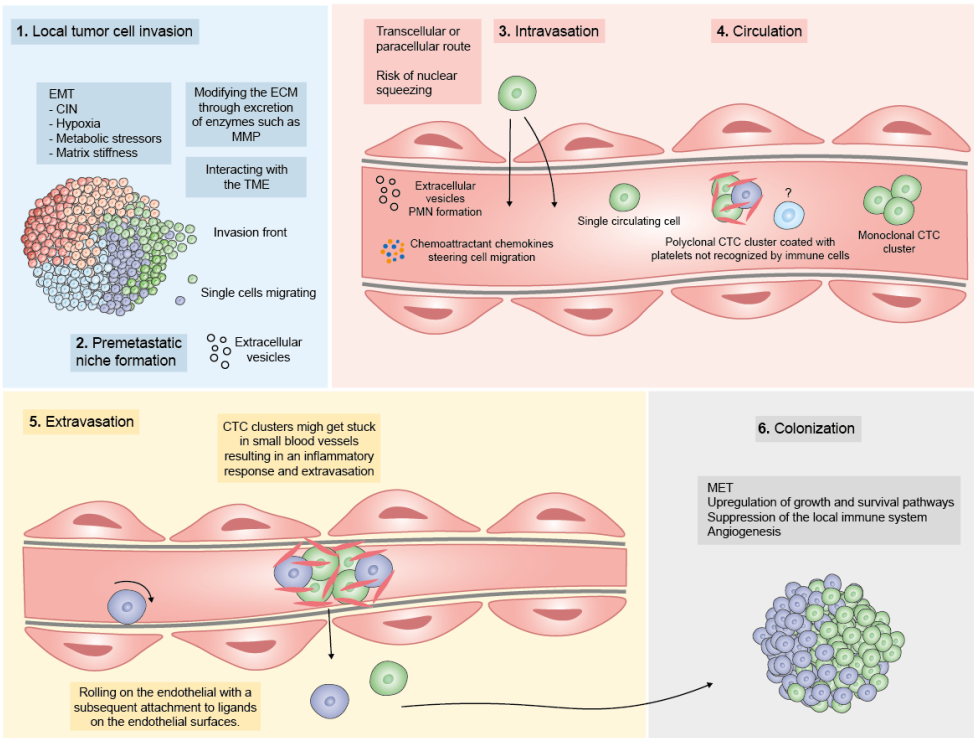
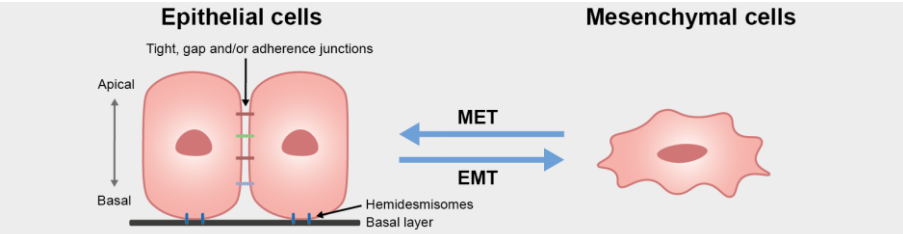


Figure 15. The metastatic cascade. **1)** It begins with a local invasion of the surrounding tissue. Cells undergo an epithelial-to-mesenchymal transition and start to migrate towards the vessels, either as single cells or as part of a tumor invasion front. **2)** The cancer cells can prime a distant site for future colonization, denoted premetastatic niche formation. **3)** The cancer cells enter the blood or lymphatic vessels through either transcellular or paracellular migration and by passing the basement membrane. **4)** In the circulation they travel either as single cells or as part of a circulating tumor cell (CTC) cluster, which can be monoclonal or polyclonal. The cells may even coat themselves with platelets, stromal cells, and immune cells, to evade the immune system in the blood. **5)** As the cells reach the smaller vessels, they can either roll on the endothelial surface and attach to ligands to extravasate, or they can get stuck, causing a local inflammatory response followed by vasodilation and easier extravasation. **6)** The cells may then colonize the site, which requires a mesenchymal-to-epithelial transition, upregulation of certain cellular pathways, suppression of the local immune environment and angiogenesis.

1. Invasion of the surrounding tissue

The initial trigger of invasion is not known, but various mechanisms are thought to be involved. Typically, tumor cells are immobile and tightly bound to surrounding cells and the extracellular matrix (ECM), similar to epithelial cells. To invade the surrounding tissue, the cancer cells have to acquire an increased mobility, which can be achieved through a process called the epithelial-to-mesenchymal transition (EMT). This is a normal cellular process, playing an important role in embryonic development, organogenesis, and wound healing. [345] The EMT is epigenetic and is

hence not dependent upon DNA-sequence alterations, possibly explaining the absence of universal genetic drivers for metastasis. During EMT, tumor cells lose their cell polarity and connections to the surrounding cells and instead gain migratory and invasive properties (**Figure 16**). Mesenchymal cancer cells, resulting from EMT, often display an increased resistance to therapy compared to epithelial cells, making EMT an attractive therapeutic target. [346]



Polarization	Possess apico-basal polarity.	Lack polarization.
Appearance	Cells tightly bound to each other.	Have a spindle-shaped morphology.
Connections	Tight, gap and/or adherence junctions between cells. Bound to a basal lamina at their basal surface via hemidesmosomes and α6β4 integrins.	No physical connections between the cells. They interact with each other only through focal points.
Expression	High levels of E-cadherin. Specific cytokeratins that stabilize desmosomes to ensure the resilience of epithelial cell layers to various physical stresses.	High levels of N-cadherin, fibronectin, and vimentin as well as β1 and β2 integrins.
Migration	Non-migratory	Migratory

Figure 16. Overview of some of the characteristics of epithelial and mesenchymal cells. EMT refers to an epithelial to mesenchymal transition and MET a mesenchymal to epithelial transition.

It has been confirmed across several tumor types both *in vitro* and *in vivo* that there are cancer cells expressing a mixture of both epithelial and mesenchymal markers. Rather than a binary process, with the tumor cells existing in either an epithelial or a mesenchymal state, there consequently is a spectrum of intermediate states. Tumor cells in these intermediate states are more effective in developing drug resistance, reaching the circulation, and forming metastases. [347, 348] Cells in separate EMT states also aggregate in distinct tumor regions, rather than being evenly spread out. The most mesenchymal cells are frequently located close to inflammatory cells and endothelial cells, and they may also secrete chemokines that promote angiogenesis and attract immune cells. [348]

The transition from an epithelial to a mesenchymal state involves various growth factors, signaling pathways and gene regulatory networks. [349, 350] The growth factor TGF-β has e.g., been shown to stimulate the transition, and may serve as a potential target for anti-metastatic therapies. [345, 351] Furthermore, chromosomal instability (CIN) plays a pivotal role in initiating EMT and driving metastatic spread. [352, 353] It has been shown that metastases are enriched in cells with an increased CIN and number of CNAs compared to the primary tumor [354] and cells with high CIN display an enrichment of mesenchymal markers, indicating that they have undergone EMT. [354] Additionally, a high CIN predisposes cells to spill

DNA into the cytosol, activating the cGAS-STING cytosolic DNA-sensing pathway and downstream noncanonical NF-$\kappa\beta$ signaling, promoting metastasis. [354] Notably, administration of STING inhibitors reduce CIN mediated metastasis in malignant melanoma, breast cancer and colorectal cancer *in vivo*. [355]

Other triggers of EMT include hypoxia, metabolic stressors and matrix stiffness. [345] Hypoxia and HIF-dependent signaling has additionally been shown to promote CIN and metastasis. [356, 357] A recent study showed that metastatic subclones in clear cell renal cell carcinoma preferentially originated from the tumor interior, which is often subject to hypoxia caused by rapid tumor growth and inadequate blood supply. The tumor area was also more necrotic and had an increased copy number burden, and proliferation rate compared to peripheral tumor regions. [358]

To reach the blood stream or lymphatic vessels, cancer cells must navigate the tumor microenvironment, interacting with normal cells, including fibroblasts and immune cells. Cancer cells can modify the extracellular matrix (ECM) by excreting enzymes such as matrix metalloproteinases (MMPs), resulting in remodeling or a breaking down of its components, creating gaps and holes that allow the cells to migrate through the tissue. [359] Great efforts have been made to study how cancer cells interact with cancer-associated fibroblasts (CAFs) and the immune microenvironment, interactions that can either promote or inhibit proliferation, providing important clues for possible therapeutic interventions. [166]

The migratory process itself involves the use of protrusions like lamellipodia or filopodia, driven by polymerizing actin filaments, pushing against the cell membrane. Studies have indicated an alternative way to migrate used by cancer cells, called blebbing migration, in which internal pressure creates a bubble on the surface of the cell, which is followed by the formation of actin and myosin fibers within it, moving the cell forwards. [360] Cancer cells can either migrate as single cells or collectively in a coordinated movement with several cells, forming a tumor invasion front. [361]

The most common routes of spread are via the lymphatic or blood vessels, but there are also perineural spread and dissemination via cavities such as the abdominal and pleural spaces. [362] The lymphatic and blood vessels in a tumor are often more fragile than regular vessels, due to the rapid angiogenesis. The fact that the vessels are leakier, may aid the transport of the cancer cells into the fluid stream. [363] To reach the blood or lymphatic vessel cavity, the cells have to migrate through the compact tumor and surrounding tissue, followed by the endothelial layer lining and finally in between two cells (paracellular route) or through (transcellular route) a cell lining the blood or lymphatic vessel walls. [364] When the tumor cells migrate between the cells lining the blood and lymphatic vessel walls, the nucleus might be squeezed. This can challenge the integrity of the nuclear envelope and the DNA, resulting in rupture of nuclear membrane blebs, and rearrangements causing genomic instability, or cell death. The migration hence

results in a substantial physical stress on the nuclear envelope and the cell, requiring an efficient repair of the envelope and DNA for the cells to survive. [365] A high interstitial fluid pressure in the tumor is correlated to a worse therapy response and it might also aid metastatic spread through pushing the cells outwards, but more studies have to be conducted on this area. [366] Cancer cells may furthermore possess a mechanical memory. This means that the cancer cells change their phenotype based on the physical microenvironment and that this phenotype can be retained even after the physical stimulus has been withdrawn. [367]

Thus, utilizing a combination of chromosomal instability, epithelial-to-mesenchymal transition, extracellular matrix remodeling and degradation, tumor microenvironment interaction and blebbing migration, cancer cells can maneuver their way toward the blood and lymphatic vessels.

2. The premetastatic niche (PMN)

There is evidence suggesting that the primary tumor selectively primes specific organ sites for future metastatic spread even before the arrival of any tumor cells. [368] This process, known as *premetastatic niche (PMN) formation*, is believed to be driven by various factors and extracellular vesicles (EVs) secreted by the cancer cells. These EVs are taken up by normal cells in the host tissue, inducing changes such as vascular leakiness, ECM remodeling and immune suppression, all of which are essential for PMN establishment. [368] Cancer cells can also release EVs carrying mRNA which are taken up by non-malignant cells, which subsequently display enhanced migratory behavior and even metastatic capacity, suggesting a transfer of metastatic characteristics between cells. [369]

Furthermore, there is evidence of PMN formation before cells migrate to the lymph nodes, making it easier for the tumor cells travelling through the lymphatic vessels to colonize the lymph nodes. This process involves local repression of immune surveillance and the activation of specific stromal cells. There are also theories that EVs may help to initiate this lymph node remodeling. [370]

In summary, primary tumors may prime specific organ sites, creating a favorable environment for migrating cancer cells to successfully colonize it. Targeting this PMN formation holds the possibility to hinder future metastatic colonization.

3. Circulation

The circulation is a very harsh and turbulent environment, highlighting the importance of the fluid dynamics in the metastatic dissemination and extravasation process. Regions with low hemodynamic flow are at greatest risk of colonization. [371] Cancer cells usually travel as single cells, being exposed to great shear forces, which is likely to induce apoptosis before they have the possibility to lodge in a

capillary in a distant organ. Although rarer, tumor cells can also travel in circulating tumor cell (CTC) clusters. These clusters are more resistant to the shear forces and cytotoxic immune cells. Additionally, the CTC clusters are prone to becoming lodged in small blood vessels, where they may fragment and trigger an inflammatory response, facilitating extravasation. A study showed that CTC clusters merely constitute 2.6 % of all CTC events, with single CTCs making up the remaining 97.4 %. However, CTC clusters exhibit a 23-50-fold increased metastatic potential compared to single CTCs. Notably, knockdown of plakoglobin, an agent which helps maintain CTC cluster integrity, suppresses CTC cluster formation and lung metastases *in vivo*. [372] These clusters can form either during the extravasation or mechanically when single cells adhere to each other in the small capillaries. [373] Furthermore, research suggests that tumor cells may travel in clusters together with stromal cells and immune cells. [374] CTCs also recruit platelets and use them as a physical barrier for protection against elimination by immune cells and it may possibly also aid the adherence of the CTCs to the vessel walls. [375] [376] [377] It has not been studied whether there are "metastases" of TME elements alone without cancer cells, that preconditions the other organs to accept the tumor cells.

Studies are being conducted on using CTCs and cell free DNA (cfDNA) in the blood in cancer patients to monitor disease and detect relapses early. The specificity for detection of CTCs is, however, poor since the number of CTCs is very low in comparison to the number of healthy cells in the blood stream. Moreover, since tumors are heterogeneous, the CTCs may express different surface markers, complicating systematic differentiation between normal cells and tumor cells. Despite these challenges, CTC analysis is a technique that holds great clinical potential for disease monitoring and early relapse detection. [378] Interestingly, for malignant melanoma patients, a photoacoustic method has been developed which can detect CTCs and destroy them using laser pulses. [379]

A study suggested that there were an increased number of CTCs during sleep compared to awake hours in breast cancer patients. Hence, there may be hormonal factors influencing the capability of cancer cells to intravasate into the bloodstream. The cancer cells in the blood during the night also had a higher probability of forming metastases. [380] This does somewhat contradict previous studies which have emphasized the importance of sleep on prognosis. A regular circadian rhythm enhances the immune system, while systematic sleep disruption can promote systemic inflammation and confer a poorer cancer survival rate. [381, 382] Circadian rhythm disruption also increases the risk for breast cancer development *in vivo*. [383]

There are emerging studies showing the importance of chemoattractant chemokines excreted by non-malignant tissues, which in part might explain organotropism. [384-386] Cells at a distant site release protein molecules denoted chemoattractant chemokines, which bind to specific surface receptors on the tumor

cells. This causes the cells to follow the chemokine concentration gradients to the tissue site of emission, a process called *chemotaxis*. [387] Chemokines could pose attractive therapy targets. Breast cancer cells express the chemokine receptors CXCR4 and CCR7, whose ligands CXCL12 and CCL21 are highly expressed in organs that are common sites of metastatic spread. In addition, blocking these interactions, significantly impaired metastasis to regional lymph nodes and the lung *in vivo*. [384]

The *seed and soil hypothesis* was proposed by Stephen Paget in 1889. He noticed that the sites of metastatic spread were nonrandom, and that some organs such as the skeleton, lung and liver, more frequently harbored metastases compared to other sites. He claimed that this organotropism was due to interactions between the soil (the microenvironment in the target organ) and the seed (the metastatic tumor cells). [388] The truth probably lies in a combination between the seed and soil hypothesis and that the organotropism partly can be explained by the anatomy of the vascular and lymphatic drainage. It is today well established that the microenvironment plays an important role in tumorigenesis and metastasis formation. [386, 389]

4. Extravasation into the new tissue

As tumor cells traverse the small capillaries, they can get stuck. This either causes the cell to die, the vessel to rupture or the cell to extravasate. In the bones and the liver, the sinusoidal vessels are highly permeable facilitating extravasation into these organs, possibly explaining why metastases are common in these sites. [390] Cancer cell extravasation may also occur in a similar way as leukocyte transendothelial migration, starting with a rolling on the endothelium and a subsequent attachment to ligands on the endothelial cell surfaces. This is followed by a disruption of the junctions between the endothelial cells, allowing the tumor cells to squeeze between them. The cancer cells then crosses the basement membrane and enter the secondary site. [391]

5. Colonization and growth at the secondary site

A large number of cancer cells are released into the circulation daily in each cancer patient, but most cells die due to a failure of solitary cells to initiate growth and the micrometastases not being able to continue to grow into macroscopic tumors. In a study, merely 2.5 % of extravasated cells formed micrometastases and only 1 % of these progressed to macrometastases in an *in vivo* model for breast cancer, while most micrometastases disappeared. [392, 393] Other studies have indicated that only 1.5 % of cells survived 24 hours in the blood stream and just 0.002 % of all circulating cancer cells survived 14 days. Of these only a subset formed micro- and macrometastases. [394]

The colonization of the metastatic site requires upregulation of growth and survival pathways, suppression of the local immune system, and angiogenesis. The process is possibly also aided by a premetastatic niche formation before the metastatic tumor cells has arrived. [368, 390] Some studies have suggested that cells may undergo a mesenchymal-to-epithelial transition (MET), the reverse process as EMT, resulting in more epithelial like cells. It is not established whether this is necessary for metastasis formation, but the epithelial subtype is more proliferative, which is advantageous when forming a new tumor. [395]

Emerging evidence suggest that disseminated cancer cells also may be lingering in distant organs in a quiescent or dormant state. They may be dormant for a long period of time, perhaps due to the stress posed upon the cell by the migration, just to reawaken and start proliferating. These dormant cancer cells might cause late relapses. Developing drugs targeting the induction of dormancy or reawakening of them poses attractive treatment possibilities. [16, 319, 396]

Metastatic trajectories

Metastatic spread can occur via several different routes and in different directions. *Monoclonal seeding* involves one or several cells from the same subclone colonizing a metastatic site. In *polyclonal seeding* several different subclones colonize the metastatic site together at the same time (direct clonal cooperation) or sequentially (indirect clonal cooperation). The first subclone reaching a distant site could in theory remodel the metastatic niche, making it attractive for additional subclones to colonize it later. Phylogenetically, metastases can be monophyletic or polyphyletic. *Monophyletic* metastases refers to when one phylogenetic branch leads up to all metastases in the patient. In contrast, *polyphyletic metastases* refers to when metastases at different sites originate from different phylogenetic branches in the phylogeny. Furthermore, cancer cells may spread between lymph nodes or between distant metastatic sites, denoted *intermetastatic spread*. Finally, there is also a possibility that cancer cells recolonize the primary tumor site through *self-seeding* (**Paper V: Figure 1A-B**).

One prevailing model of metastatic dissemination is the so-called *sequential or linear progression model* which posits that *metastases arise late* in tumor evolution and the genetic divergence between the primary tumor and metastases is minimal. On the other hand, there is the *parallel progression model*, proposing that *metastases may occur early* in the evolution in the tumor. After dissemination both the primary tumor and metastases diverge from one another genetically. Evidence from several different tumor forms suggests that both linear and parallel metastatic progression, monoclonal, polyclonal, monophyletic, and polyphyletic spread can occur in the same patient. [397, 398]

Traditionally lymph node metastases have been considered precursors of distant metastases. This view is the foundation of the widely used TNM staging system used clinically to group patients and determine treatment schedules. It is widely known that lymph node status is an important prognostic factor, but there is emerging evidence that lymph node removal does not always improve patient survival and it is not known whether lymph node removal influences distant dissemination. In a randomized clinical trial encompassing 891 patients with sentinel node positive breast cancer, sentinel lymph node dissection alone did not result in inferior survival compared to complete axillary lymph node dissection. [399] Studies on malignant melanoma have shown that removal of locoregional metastatic lymph nodes provides no benefit on overall survival. [400, 401] This has raised the question whether the subclones found in the lymph nodes actually are the ones resulting in the distant metastases, or merely an indication that the tumor has metastatic capability. Lymph nodes may merely act as a highly efficient bioassay for catching metastasis-competent cells. [402] In a study on colorectal cancer, it was shown that 65 % of distant metastases arose from subclones in the primary tumor not found in the lymph nodes. [403] In prostate cancer it was suggested that lymph node metastases and skeletal metastases constitute distinct metastatic lineages, and that the lymph node metastasis is not a necessary intermediate step. [404] Tumor cells in the lymph nodes can invade the blood vessels covering the lymph nodes and spread to distant organs rather than via the lymph providing another route for subclones in lymph nodes to reach distant sites. [405, 406]

Much research has been conducted to investigate the evolutionary trajectories of metastatic disease across cancers. In colorectal cancer, a study encompassing 118 biopsies from 23 patients, showed that metastatic spread can occur early. [407] Another study on colorectal cancer elucidated the evolutionary relationship between 94 tumor samples from 10 patients, including samples from the primary tumor, lymph node metastases and liver metastases. Intermetastatic spread was identified between liver metastases as well as between lymph node metastases. The latter could also occur through so-called skip spreading where cancer cells bypassed lymph node layers in their migratory path and that different lymph nodes have different metastatic potentials. They also showed that metastatic spread to the liver could occur either via hematogenous or lymphatic routes. [408] In another study on colorectal cancer it was shown that lymph node metastases were polyphyletic and polyclonal while distant metastases were monophyletic and preferably monoclonal, implicating a wider evolutionary bottleneck for metastatic spread to the lymph nodes compared to distant sites. Many cells from the primary tumor are hence capable of colonizing lymph nodes compared to distant sites, which may partly be explained by a high seeding frequency to locoregional lymph nodes. The

monophyly of distant metastases could partly be explained by a selection pressure. [409]

In a study on 51 samples from 10 prostate cancer patients it was found that 8 out of the 10 patients displayed clear patterns of intermetastatic spread, concluding that this is a common feature in prostate cancer. The intermetastatic spread could occur through either monoclonal or polyclonal seeding. [11] In another study on 26 whole genome sequencing samples from 4 metastatic prostate cancer patients it was found intermetastatic seeding as well as self-seeding from a distant bone metastasis to the surgical bed of the primary tumor site resulting in a local recurrence. They also found that metastatic spread from the primary tumor site could occur several times during its evolutionary history. [410]

In pancreatic cancer, a study based on autopsy samples from seven pancreatic cancer patients showed that metastatic spread is a late event relative to tumor initiation and the metastases had mainly clonal alterations. [325] Pancreatic cancers are poorly vascularized, making it difficult to deliver chemotherapy to the tumor interior. Hypoxia might also create an environment that triggers the formation of additional mutations and metastatic spread. A recent paper showed that metastases in renal cancer often occur from cells in the center of the tumor, which was shown to be more hypoxic. [358] A study encompassing 40 patients across 13 cancer types with metastatic spread encompassing 310 samples (median 4 samples from the primary tumor and 3 samples from metastases for each patient) showed that metastases can arise early and from several different subclones in the primary tumor. [411]

In breast cancer it has been observed that minor subclones in the lymph nodes can give rise to distant metastases, but in other cases the metastatic clone is not found in the lymph nodes. [10] In another study where 299 samples from 170 patients were analyzed concluded that metastatic spread most often occurred late in the evolutionary history of the tumor, [412] while another, encompassing 99 samples from 38 patients, indicated occurrence of both early and late metastatic spread. They also found that metastases were mainly monophyletic. [413]

There are unfortunately few studies on metastatic trajectories in pediatric cancers. A recent study on evolutionary trajectories in neuroblastoma included 470 samples from 283 patients. From this cohort metastatic trajectories could be mapped from 19 patients having at least one sample from the primary tumor site and at least one sample from either a locoregional or distant metastatic site. This analysis resulted in the identification of intermetastatic spread and polyclonal seeding, in this tumor type. [414]

Present investigation

Aim

The overarching aim of this thesis was to develop bioinformatical tools for analysing the phylogenetic genetic relationship between cancer cell populations in the pediatric cancers, neuroblastoma, Wilms tumor, rhabdomyosarcoma, and germ cell tumors. Furthermore, the purpose was to apply these methods to gain insights into the development of local relapse, metastasis, and treatment resistance in these tumor types.

The specific aims of each paper were as follows:

Paper I - Develop a bioinformatical tool capable of reconstructing phylogenetic trees of cancer cell populations based on their genetic profiles. To be applicable to pediatric cancers, the software should be able to handle copy number alterations with or without additional information about point mutations.

Paper II - To investigate treatment resistance and relapse in neuroblastoma, using PDX-models of mice and an adapted treatment protocol containing the drugs encompassed by the COJEC chemotherapy treatment protocol, administered to patients.

Paper III - To decipher the phylogenetic origin of neuroblastoma cells surviving treatment.

Paper IV - To identify key events in the pathogenesis of the typically drug resistant diffuse anaplastic Wilms tumor, compared to blastemal and stromal Wilms tumor, respectively.

Paper V - To identify the most likely ancestral relationship between tumor cells at different metastatic sites and in respect to the primary tumor before and after treatment.

Materials and methods

Sample collection

Biopsies and entire tumors were collected and stored during the years 2000-2021, from pediatric cancer patients with the diagnoses neuroblastoma, Wilms tumor, rhabdomyosarcoma, and germ cell tumors. The archive includes samples collected both from the primary tumor at diagnosis, and for some patients, samples after chemotherapy, from relapses and several metastatic sites. Sometimes entire tumors were stored after surgery. Additionally, for a subset, there are post-mortem samples. The collection has been made with informed consent from the children's parents and an ethics permit has been approved with d.nr. L289-11. It was updated for further data- and material collection in 2017 and renewed in 2023 by the Swedish Ethics Review Authority (2023-01550-01). The projects also includes the handling of genetic and clinical data, which is classified as sensitive information. All raw sequencing data is anonymized and stored at a secure server. Throughout all studies, all clinical data is anonymized.

The families together with the child are informed about the study together with a medical doctor and nurse. They also receive written information about the study, which they can take with them home. They usually get a couple of days to think before they decide on whether to sign up or not.

If they would like to be in the study the caregivers have to sign a form. If the child is at least of an age of 15 years old, the child has to sign as well. Younger children can sign if they want to, but it is not required. The patients are automatically in the study until they actively say that they do not wish to anymore. No new agreement is signed if the child dies, but if an autopsy is to be conducted, the caregivers must be asked for permission.

DNA extraction

To generate the data in paper I, III, IV and V, DNA was extracted from Formalin-Fixed Paraffin-Embedded (FFPE) material. The majority of older clinical tumor specimens are stored in this way, since it holds several advantages, such as preserving tissue architecture and allowing the material to be stored at room temperature. This contrasts with fresh frozen (FF) samples, requiring storage in ultralow temperatures, consequently constituting a higher cost for maintenance of the stored material. Furthermore, storing samples as FF specimens makes it more challenging to assess the material histologically. It does, however, maintain a better DNA quality, aiding subsequent genomic analyses.

Performing genomic analyses on DNA and RNA extracted from FFPE-material is technically challenging. When DNA is extracted from FFPE tumor tissue, the formalin and paraffin must be washed away using chemical agents, some of which might interfere with genomic analyses. The extracted macromolecules are often fragmented, and the total amount of DNA may be low. The total amount of DNA may, however, be low for FF samples as well if the sample is small. Many methods used for sequencing have limits of how fragmented the input DNA can be. There are also limitations on the amount of DNA needed to perform the analysis. The formalin itself might also damage the DNA sequence, affecting the results. In general, FF tissue yields DNA that is less fragmented, and the amount of DNA is greater. Consequently, using FFPE material constitutes a limiting factor regarding which analyses can be performed on the material compared to FF tissue.

Despite samples being stored for many years, it has been shown that both DNA, RNA, and proteins can be extracted from FFPE material while maintaining a quality which in many cases is comparable to FF samples. [415] Several studies have compared the results from targeted deep sequencing (TDS), whole exome sequencing (WES) as well as RNA sequencing (RNAseq), using DNA extracted from either FFPE or FF material, showing strong correlations between the results. [416-419] In one study, whole genome sequencing (WGS) was performed on paired FFPE and FF samples, showing a SNV agreement of 71 % in reliable regions with sufficient coverage, but the copy number data was noisier in the FFPE data set. [420] It has also been shown that the RNA quality is similar from 10 year old samples as from FFPE blocks that are a couple of months old, [421] and there have been advances in the use of single cell RNAseq on FFPE samples. [422]

Papers I-V have included SNP-array to obtain the copy number of each genomic region. To reliably obtain mutational data WES and TDS up to 10000x sequencing depth was used. Papers II and III also included FF samples, allowing single cell whole genome sequencing as well as RNAseq.

Patient derived xenograft models

A xenograft is a heterologous transplant of living cells, tissues, or organs from one species to another. A patient derived xenograft (PDX) is a model of cancer where entire parts of a tumor, including the tissue around the tumor encompassing e.g., stromal cells, are implanted into immunodeficient or humanized animals, most often mice. PDX models are used to create an environment that allows for a more natural growth of the cancer cells, compared to culturing the cells *in vitro*. [423] This allows monitoring of the tumor size and evaluation of for example different treatment protocols. The tissue can either be placed heterotopically (subcutaneously) or orthotopically i.e., at the same anatomical site as it was extracted from the patient. In the latter case it is called a patient derived orthotopic

xenograft (PDOX) model. [424] When the tumor burden becomes large the tumor is passaged over to a new mouse. In this way it is possible to follow tumor evolution over time.

PDX models have been used since 1969 when colon adenocarcinoma fragments were removed from patients and planted into *nude mice* (lacks a thymus and have a T-cell deficiency). [425] The transplantation success rate did however remain low for many years. Instead, cell line-based models have been widely used for drug development. They are cheaper, easier to use and do not cause direct harm to any animals. In 2001 it was, however, shown that there is a great inconsistency in drug response between cell line derived models and clinical trials. [426] PDX models have, on the other hand, been shown to excel in reflecting tumor characteristics, conserving the genetic patterns from the primary tumors as well as paralleling the patients' clinical outcomes. [427] They simulate the tumor progression and evolution seen in the patient, and provide promising models for identifying prognosis biomarkers, investigating tumor evolution and for evaluating new drugs. [424, 428] In paper II and III we employ neuroblastoma PDX models in collaboration with Daniel Bexell's research group at Lund University. These PDX models have been shown to retain the genotype and phenotype of the patient tumors and may also display infiltrative growth and metastasis to distant organs such as the bone marrow. [429-431]

To avoid transplantation rejection, mice which are immunocompromised to varying degree are used. In our studies *nude mice* and *NSG mice* (lacks mature T cells, B cells and NK cells) were used. Since the immune environment has been shown to play an important role in tumorigenesis, this is a serious shortcoming of these model systems. [424] There are ongoing studies where attempts are being made to reconstitute the human immune system in the animals by injection of human hematopoietic stem cells. [432] Another limitation is that the interaction between the murine stroma and human tumor cells may be different than that between human stroma and human tumor cells. Finally, establishing neuroblastoma PDXs based on the implantation of single tumor biopsies might not cover the entirety of the wide genomic changes found in high-risk patient tumors. It could lack the mutation or mutations possibly driving tumor progression and/or treatment resistance. Implantation of tissue from several tumor areas might mitigate this issue but could still disturb the balance regarding the proportions of each subclone.

There have also been discussions concerning the usage of PDX-models as patient avatars with the aim to try drugs to support clinical decision-making for the corresponding patient. Besides the ethical aspect concerning harm to the animals, one additional limitation is the time frame since it may take several months to establish a PDX model. Furthermore, there are pharmacokinetic differences between mice and humans making direct comparisons difficult. It is also labor-

intensive, expensive, low throughput, and it currently lacks the flexibility required for target screening. [433] Despite these limitations, PDXs have been shown to parallel clinical outcome in various tumor types and is an excellent model system for trying new drugs and treatment schedules. [424, 427-429]

Organoids

Organoids are three-dimensional (3D) tissue cultures derived from cells. In contrast to 2D cell cultures, 3D organoids more closely resemble tissues and organs *in vivo*. They also simulate the spatial organization of cells and cellular interactions, making them widely used in cancer research for study of tumorigenesis and *in vitro* drug testing. They have been shown to maintain the genetic and phenotypic heterogeneity, morphology, and treatment resistance observed in the original tumor. [434, 435] Additionally, studies suggest that tumor organoids can predict patient responses to drugs. [436, 437] However, it is important to note that organoids, as model systems also have limitations. They are simplified representations of tumors and lack for example vasculature. While capturing certain aspects of tumor heterogeneity they do not fully replicate the cellular diversity and does not recapitulate the tumor microenvironment observed *in vivo*, but they provide valuable tools for cancer research and drug development. [435]

Single nucleotide polymorphism microarray

A *single nucleotide polymorphism microarray* (SNP-array) protocol uses SNP probes to compare the signal intensity of the tumor sample to a reference sample created from a population of many normal samples. This allows for the identification of SNPs, copy number changes, CNNIs and LOHs.

In project I, II and III the *CytoscanHD assay* was used to analyze DNA extracted from FF-material and in paper IV and V the *OncoScan FFPE assay* was used, the latter of which is preferred for genomic copy number analysis of FFPE samples. OncoScan relies on the *Molecular Inversion Probe* (MIP) technology. In the OncoScan method 335 000 probes are used, each being 20 base pairs long. Each probe is designed to bind to a certain region of the DNA using two homology regions located at the two ends of the probe (**Figure 17**). The probes are mixed with the DNA in a test tube. When a probe comes in close proximity to a location of the genome to where it fits, the homology regions will bind to the DNA. There will, however, be a one base pair gap in between the two homology regions, resulting in the probe forming an incomplete circle. The solution with the probes and DNA is then split into two tubes: one containing A and T nucleotides and one containing G and C nucleotides.

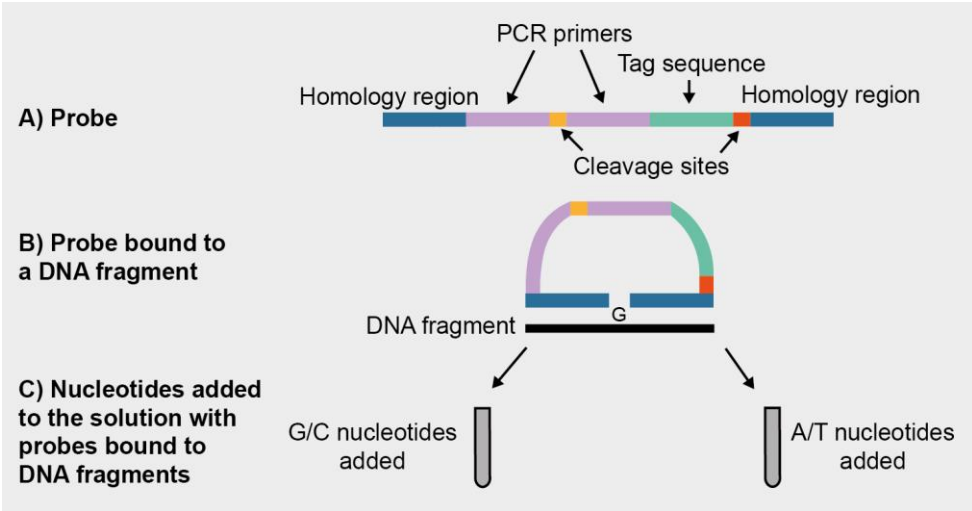


Figure 17. A) The design of the probes used in the OncoScan FFPE assay. The homology regions bind to a specific part of the DNA chain. The PCR primers are needed for PCR amplification. The cleavage sites are areas at which enzymes will cleave the probes. The tag sequence is a unique barcode for the particular probe. **B)** The probes are added to a solution with the DNA fragments to be analyzed. The homology regions of the probes bind to the fragmented DNA. There is a gap in between the two homology regions, resulting in the probe forming an incomplete circle. **C)** The solution with the probes, probes bound to DNA, and unbound DNA is split into two tubes. To one of the tubes, G and C nucleotides are added, and to the other, A and T nucleotides. Adapted from. [438]

The nucleotide that fit the gap binds to fill it, making the probe forming a complete circle (**Figure 18**). Exonucleases are subsequently added that digest incomplete circular forms of the probes and the DNA, leaving only circular probes. In the next step, enzymes cleave the circular probes, resulting in linear forms of the probes, which are amplified using PCR. This is followed by cleavage of the probe again. The remaining fragment is composed of a tag sequence, which acts as a barcode for a particular genomic DNA region, and a PCR primer site. The fragments are then applied into two different arrays, one for each of the two tubes. The tag sequences hybridize at a specific position in the array that matches the tag sequence. A fluorescent signal is generated from the tag sequences that has hybridized in the array and the light is detected. The intensity data along with information about where the locus is, and which tube it originated from is stored. The Cytoscan HD method operates in a similar manner but does not apply the MIP technology but uses straight probes.

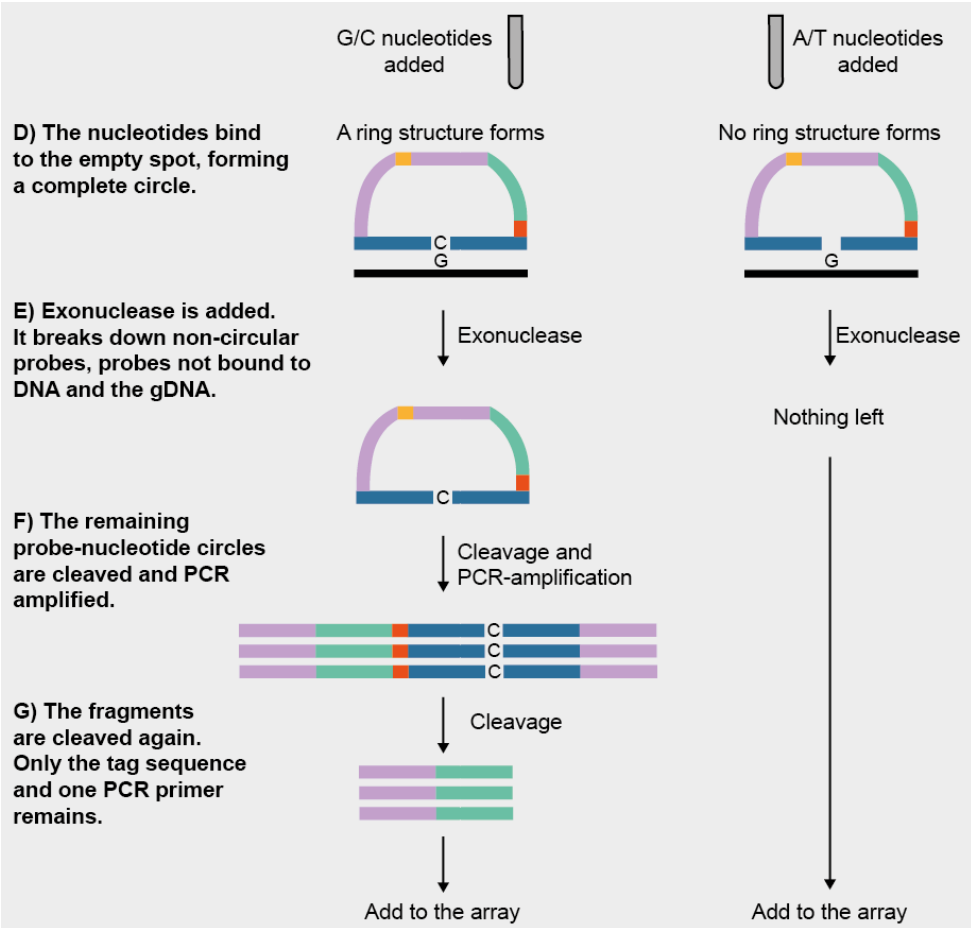


Figure 18. D) Nucleotides bind to the empty spot if possible, forming a complete circle. **E)** Exonucleases remove the probes that have not formed complete circles, not bound to DNA and the genomic DNA (gDNA). **F)** The circles that remain are cleaved and PCR amplified. **G)** The amplified fragments are cleaved again. The remaining fragments, containing only the tag sequence and one PCR primer is added to the array. Adapted from.[438]

For each analyzed sample, CEL-files are obtained containing the fluorescence intensity data. It can be used to calculate the log2-ratio (log2R) of the intensity relative to a reference signal intensity.

$$\log2R = \log2\left(\frac{\text{A/T signal intensity} + \text{G/C signal intensity}}{\text{Reference signal intensity}}\right) \qquad (24)$$

Where a log2R of 0 indicates that the region has the same copy number as the reference sample, a value > 0 indicates that there is a gain and a value < 0 that there is a loss. The B-allele frequency (BAF) can be calculated as

$$\mathrm{BAF} = \frac{\mathrm{G/C\ signal\ intensity}}{\mathrm{A/T\ signal\ intensity\ +\ G/C\ signal\ intensity}} \quad (25)$$

The values obtained, if the tumor cell fraction is 100 %, will be 0/2, 1/2 and 2/2 for a diploid segment (1+1), 0, 1/3, 2/3 and 1 for a region with a gain (2+1), 0/1 and 1/1 for a region with a loss (1+0) and finally 0/2 and 2/2 for a region with a CNNI (2+0). Normally a tumor sample will also contain some normal cells. The proportion of cells in the sample being tumor cells is denoted the *tumor cell fraction* (TCF). Depending on the TCF, the BAF will deviate from these expected values. It will also deviate if not all tumor cells have the CNA. Hence it is, from this data, possible to calculate the proportion of cancer cells having each CNA.

The software Chromosome Analysis Suite (ChAS) and Nexus were used in the studies to normalize the CEL-files, as well as for visualizing the copy numbers across each chromosome through plotting the log2R for each genomic location as well as providing the BAF for each location. This allowed detection of the copy numbers present in each sample. The log2R and BAF could be used to calculate the proportion of cancer cells that have a particular genetic alteration.

Next-generation sequencing (NGS)

Next-generation sequencing (NGS) encompasses the new technologies used to identify the nucleotide sequence of DNA or RNA molecules. For DNA the methods involve whole genome sequencing (WGS; the entire DNA is sequenced), whole exome sequencing (WES; sequencing is focused on the protein coding parts of the genome) and targeted sequencing (TDS; sequencing is focused on a certain number of genes and genomic regions).

There is a plethora of different techniques used to perform sequencing. The NovaSeq 6000 platform was used in study III, IV and V. It constitutes the following general steps, most of which are shared with several other sequencing methods. *Library preparation*: (1) Fragmentation of the DNA (unless it already is fragmented) and ligation of the fragments to oligonucleotide adaptors. By adding different adaptors to different samples, they can be sequenced simultaneously (multiplexing), since the reads can be demultiplexed later. (2) Quality control and quantification of the DNA. (3) Indexing, normalization, and pooling of the sample. *Cluster generation*: (4) The fragments are added to a so-called flow cell where the adaptors bind. (5) Amplification of the sequences using bridge PCR. *Sequencing*: (6) A primer is added that binds to the beginning of the strands. (7) Fluorescently labeled nucleotides are added. (8) These will bind to the DNA-sequence one at a time. Each nucleotide subsequently get excited by a laser and when it deexcites it

will emit a photon corresponding to that base pair. (9) The light pulses are detected one at a time by the machine. Different base pairs will emit light pulses with different frequencies. A software assigns each frequency to a base pair and outputs the DNA sequence. In paired-end sequencing, DNA fragments are sequenced in both directions, which aids the identification of true variants. This is followed by data analysis, covered in the next section.

Bioinformatical processing of high throughput sequencing data

The major steps in bioinformatical pipelines used to analyze sequencing data are similar for WGS, WES and TDS. There are, however, some important differences regarding data size, the need of parallelization and how the variants are filtered. The data files from WGS are significantly larger than those from TDS, often requiring parallelization during the computational steps. The section below mainly highlights the steps used for TDS analysis in paper V.

As stated in the previous section, several samples can be sequenced together, due to the unique adapter sequences used for each sample. This process is known as multiplexing. When multiple samples are sequenced together, the raw data obtained after sequencing needs to be demultiplexed before further analysis. This is necessary because the dataset contains a mixture of reads from several samples. After *demultiplexing*, the samples are converted into fastq files, containing information about all reads. For each read the file includes a read name, the called genomic sequence and the base call quality score for each nucleotide, indicating the probability of that nucleotide being a sequencing error. Subsequently, *quality control* is performed on the raw fastq-files, most often using the software FastQC and MultiQC. [439] This is followed by *mapping/aligning the reads to a reference genome*, in our case the hg38/GRCh38 reference genome, using the Burrows-Wheeler aligner, implemented by the software bwamem2. [440] If the reads are small, as is common in FFPE-material, they could theoretically map to several different regions of the genome, which might be a problem. Also, indels may cause the read to map poorly to the reference genome and may wrongly be filtered away. Sequencing reads can also be *paired-end sequenced*, meaning that sequencing has been performed in both directions. In these cases, there will be two fastq-files for each sample, both of which are used as an input to the bwamem2 software for alignment to the reference genome. This increases the probability that each read will be mapped correctly.

The *read depth*, also denoted *sequencing depth*, refers to how many reads have detected a specific nucleotide. If a region has 40 reads mapping to it, it is said to have a 40x read depth. The deeper the sequencing is, the more reads you have, and thus the read depth increases. How deep the average sequencing should be, is determined before sequencing is performed, and is most often limited by the cost.

The *coverage* is the proportion of an area that is covered by reads. It does not provide information about how many reads have mapped to the area, just the fraction covered (**Figure 19**).

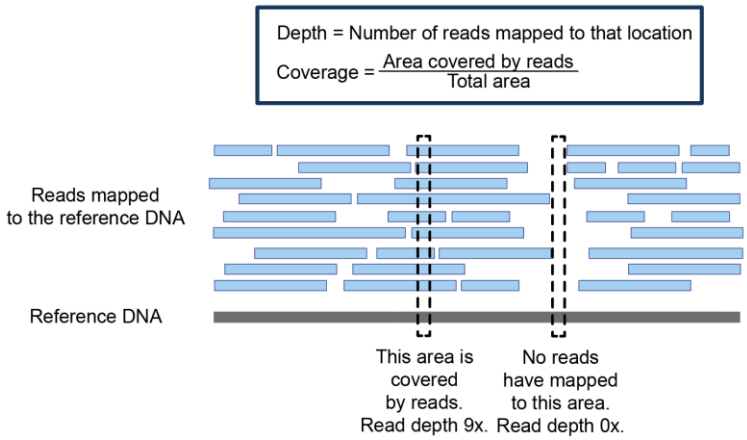


Figure 19. Illustration of read depth and coverage. The read depth is the number of reads that have mapped to a specific genomic location. The coverage is the fraction of a specified area that is covered by reads.

For WGS the average read depth is usually around 10-100x, while that for WES typically is 30-200x. TDS can, however, be performed up to 10000x. Note that all sequencing techniques theoretically can be sequenced deep, but as a larger fraction of the DNA is to be sequenced, it will require a higher number of total reads and will consequently be more costly. Sequencing deeper allows for the identification of rare subclonal variants, and the possibility to calculate the proportion of cancer cells harboring a specific variant more reliably, by decreasing the number of false positives and negatives.

After the mapping to the reference genome, Sequence Alignment Map files (SAM-files) are obtained. A SAM-file contains a header and information about the alignment of each of the reads, such as where it was mapped and the mapping quality. The SAM-files are very large, and the information is consequently usually stored as a Binary Alignment Map (BAM) file instead. A BAM-file consists of a compressed binary representation of the SAM-file. The BAM-files are sorted and indexed to make subsequent steps faster. During library preparation duplicate reads might be generated. These *duplicate reads* can be identified and marked using the MarkDuplicates software from Picard tools. As mentioned, the sequencing machine also gives information about the base quality score, which is a measure of how certain the machine is about the nucleotide it has called. To detect and correct for systematic errors made by the sequencing machine when

estimating the accuracy of each call, *base quality score recalibration* can be performed using Picard tools (http://broadinstitute.github.io/picard).

This is followed by *variant calling*. This process involves detecting differences between the sample's genome and a reference genome using software collectively denoted *variant callers*. There is a plethora of variant callers, specialized at detecting certain genetic alterations. Each software has its own advantages and flaws. [441] It is becoming common practice to perform the analysis using several variant callers and only selecting variants detected by at least two. It is also highly recommended to include a normal sample from the same patient to filter out somatic variants from germline variants. In study V, Manta, Strelka2, Varscan2 and Mutect2 were used for variant calling, also including a normal sample from each patient in the analysis. [442-445] The variant calling results in the generation of vcf-files (variant calling format files). These are text files that store all detected genetic variants in a standardized manner. During variant calling, some software such as Mutect2, also allows correction for strand bias and orientation bias.

After the variants have been called, they must be filtered. Most variant callers have tools specifically produced to filter variants detected by it, to reduce the number of false positives. In addition, it is recommended to filter variants found in the genome aggregation database (gnomAD) and exome aggregation consortium (ExAC) database. These are comprehensive catalogues of genetic variants in the human population. Variants found in a relevant custom panel of normal are also filtered out, to further reduce the number of false positives. After filtering, the variants can be annotated using Annovar, generating a txt file with information about each detected variant. [446]

When analysing WES or TDS data it is also recommended to use the accompanying Browser Extensible Data (BED) file during variant calling. The BED-file contains information on which genomic regions were sequenced by the sequencing platform.

In our studies we subsequently perform *post filtering* using customized scripts to fuse data from several variant callers to filter variants detected by several software as well as to remove variants with low sequencing depth or poor mapping quality. The variants can also be confirmed by visualizing them in the Integrative Genomics Viewer. [447]

Each variant has a variant allele frequency (VAF) which is the proportion of variant alleles in a that specific locus. This value can be used to calculate the proportion of cancer cells having each variant.

Single cell low pass WGS

Multiple techniques are available for analyzing the karyotype of individual cells. Fluorescence In Situ Hybridization (FISH) uses fluorescently labeled DNA probes that bind specific target sequences on chromosomes. FISH is, however, limited by the number of probes that can be utilized simultaneously. Multicolor FISH enables the use of several different fluorescent probes that are each labeled with a different color. It allows the visualization of more chromosomal alterations than traditional FISH, but it is still limited by the number of probes that can be used. [448] Spectral karyotyping (SKY), which is also a FISH based technique, stands out by allowing staining of all 24 different (22+X+Y) human chromosomes simultaneously without a priori knowledge of any abnormalities, but is limited to larger chromosomal aberrations. [449] All these methods require the cell to be in metaphase and relies on the hybridization quality. Other methods, such as array comparative genomic hybridization (aCGH) measure the average DNA content, discarding the single cell resolution, but there is a possibility to perform subclonal deconvolution based on such data in some cases. [450]

In paper II and III, we adopted a modified single cell whole genome sequencing (scWGS) protocol for estimating the copy number profile of individual cells. [451] This involves using fresh frozen tissue followed by nuclei isolation and sorting of single nuclei using Fluorescence-Activated Cell Sorting (FACS). [451, 452] Both the sample of interest and controls are included, where the latter constitutes at least one sample with normal reference cell nuclei (positive control) and one sample without any nuclei (negative control). Libraries are prepared without whole genome preamplification to reduce PCR amplification induced bias using a modified version [451] of a previously published protocol. [453] Each library is barcoded allowing pooling (multiplexing) of libraries for sequencing using the Illumina HiSeq2500 sequencing platform.

The reads are subsequently demultiplexed followed by alignment to the human reference genome. The resulting BAM-files are sorted, and duplicate reads are marked. Duplicate reads and reads with a poor mapping quality are filtered out. The copy number is estimated using the software Aneufinder. [454] It bins the mapped reads into approximately 1 Mb large, non-overlapping areas. Aneufinder subsequently uses a Hidden Markow Model (HMM) to predict the copy number state of each bin. The absence of pre-sequencing PCR amplification does, however, result in a poor coverage, since the total number of reads will be lower than it would with PCR amplification. Losses during library preparation will be random and the distribution of reads across the genome will thus be rather even. Since it is the distribution of reads across the chromosomes that is used to estimate the copy number, the absence of PCR amplification, consequently, still allows estimation of the copy number. More reads are expected in regions with a higher

copy number and fewer in a region with a lower copy number. The software can estimate copy numbers ranging from 0 to 10 copies. All copy number states are evaluated for each bin by fitting a binomial distribution or a delta distribution (for nullisomy) to the aligned reads and the Baum-Welch algorithm is used to find the best distribution parameters, transition probabilities and posterior probabilities for each. The copy number state resulting in the highest posterior probability is the copy number assigned to that bin. Aneufinder also includes quality control steps since the single cell sequencing data can be noisy. [454]

Despite its advantages, this method has limitations, for example, it requires FF material. Furthermore, since the analysis is performed on single cells, degradation and contamination might have a pronounced effect on the quality of the sequencing data. Another risk is that two cells might be sequenced together, mistakenly classifying the cell as tetraploid. The data can also be noisy and thus requires robust filtering steps. On the other hand, since amplification is not performed, the risk of false positives and negatives is reduced. It also allows a detailed analysis of individual cells' karyotypes, which has important applications in cancer research, given that the tumors are highly aneuploid and heterogeneous. [5]

For bulk DNA data, it is often necessary to perform *subclonal deconvolution*. This process involves deducing the subclones present in a sample using various mathematical methods. In scWGS data, each copy number profile represents a single cell. Hence for scWGS data it is not necessary to perform subclonal deconvolution to identify genomic subclones of cancer cells, significantly aiding downstream phylogenetic analyses.

Phylogenetics

Papers I-V involved phylogenetic analyses to track tumor subclones across space and time. When using bulk DNA data, it is important to note that each sample, usually contains several different tumor cell subclones. This makes it necessary to perform subclonal deconvolution and the phylogenetic reconstruction should ideally be based on these subclones as taxa. There are, however, many studies where each sample is assumed to only contain a single subclone, and the phylogenies consequently display so called sample trees, not representing true phylogenies. [455] Paper I involved the development of a software we called *DEVOLUTION* that performs subclonal deconvolution. It creates an event matrix illustrating the genetic alterations present in each identified subclone, which can be used as input for phylogenetic reconstruction. See the section below for more details. Paper II involved the development of a software we named *Single Cell Event Matrix (SCEM)*. SCEM generates an event matrix based on scWGS data, which can be used as input for phylogenetic reconstruction.

For downstream phylogenetic tree reconstruction, well established phylogenetic methods were used such as the maximum likelihood and maximum parsimony method, described in the introduction. For maximum parsimony the Hamming distance was used to obtain the distances between each of the biological entities followed by usage of the pratchet algorithm with the TBR search method with 2000 iterations to search for the most parsimonious tree. For maximum likelihood the Hamming distance was used to obtain the distances between the biological entities. To reconstruct the maximum likelihood tree, a neighbour joining tree was used as a starting point and the Jukes and Cantor model as a substitution model.

Paper IV involved the development of a novel method för phylogenetic reconstruction we named *the modified maximum parsimony (MMP) method*, whose basic principle is similar to the maximum parsimony method, but it additionally has a criterion to minimize instances of backmutations by integrating information about the prevalence of each subclone in each sample in the phylogenetic reconstruction.

Software: DEVOLUTION

Because of the extensive intratumoral genetic heterogeneity, each tumor is composed of a wide variety of *different* subclones with distinct genetic profiles. Single cell sequencing has opened up new possibilities to identify which subclones are present in a tumor, but it is, unfortunately, too costly to implement on the scales needed in the clinic. Hence, bulk sampling analysis methods such as WGS, WES and TDS are still being widely used. This poses a set of problems. Firstly, it means that millions of cells are analyzed at once. Secondly, each biopsy will encompass several different subclones, whose relative proportion will vary across samples. Thirdly, it is not known whether the mutations detected in a sample are present in the same subclone, or in different ones since the data constitutes a mixture of the signal from several different subclones (**Figure 20**). *Subclonal deconvolution* defines the process of using the information available from each biopsy to try to deduce, or deconvolve, which subclones are present in the sample, often using various mathematical methods. In this way, it is possible to elucidate which genetic alterations are located in which cells and hence which subclones are present in the tumor.

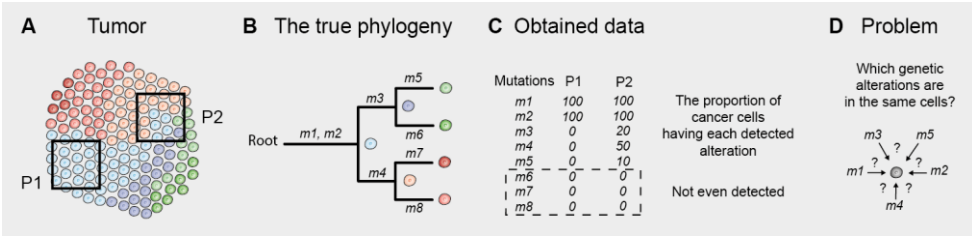


Figure 20. Illustration of the problem with bulk genotyping data. **A)** A simplified tumor. Differently colored circles represent different subclones. Imagine taking two biopsies, P1 and P2. These samples will encompass different proportions of and types of subclones. **B)** The true evolutionary relationship between the subclones in the tumor. This is not known by the researcher but reflects what we are trying to reconstruct with the data at hand. **C)** The data obtained through bulk DNA genotyping. The proportion of cancer cells having each detected genetic alteration can be calculated. Note that some mutations are not detected since the biopsies did not cover them. **D)** To figure out which subclones are present in the tumor, based on the available data, which genetic alterations are in the same cells, and which are in different ones has to be deduced.

There are several different software tools that use bulk DNA genotyping data to perform subclonal deconvolution. Some even integrate subclonal deconvolution with the phylogenetic reconstruction. Most methods are limited to the usage of somatic mutational data and do not allow the integration of CNAs. This is a serious shortcoming since almost all tumors have CNAs, which have been shown to be of integral importance for the understanding of tumorigenesis and cancer evolution. [48, 106]

PyClone and SciClone are two popular methods for subclonal deconvolution, but they require sequencing data to operate and are mainly clustering algorithms. SciClone focuses exclusively on regions of the DNA that are copy number and LOH neutral. PyClone may integrate information regarding copy numbers but assumes that the CNA is clonal in the sample, i.e. that all cells display the aberration, which is, in many cases, not true. [456, 457] MAGOS has been shown to outperform both PyClone and SciClone, but it requires sequencing data and only includes variants found across all samples. [458] Our group published the software CRUST which can segregate clonal and subclonal variants, while also adjusting the VAFs for copy number changes during the subclonal deconvolution. It was also shown to outperform both MAGOS and SciClone. It does, however, also require sequencing data. The output from PyClone, SciClone and CRUST may be used with DEVOLUTION, described below. [459]

Besides the popular PyClone and SciClone software, there are several other ones. Clomial requires sequencing data and assumes that the variants are located at diploid segments. [460] LiCHeE requires deep sequencing data and does not incorporate CNAs. [461] SCHISM require sequencing data and only consider variants in diploid regions. [462] SPRUCE on the other hand can create phylogenetic trees jointly from SNVs and CNAs but require sequencing data to operate. [463]

REVOLVER requires sequencing data and is mainly designed to detect repeated cancer evolution and common evolutionary trajectories across many patients. [464] Methods such as TITAN and THetA are also limited to sequencing data, and additionally can only integrate information from a single biopsy at a time. [465, 466]

PhyloWGS can include both sequencing data and CNAs to create phylogenetic trees, but it requires WGS-data. It does allow input data from several samples but does not integrate the information between samples during the inference procedure, which is a serious shortcoming. [467] HATCHet also allows CNAs and even handle them jointly across multiple tumor samples from the same patient, but it also requires WGS data to operate. The output is a file indicating the proportion of cells in each sample having a particular alteration. [468]

TuMult can create phylogenetic trees based on CNAs, but it is limited to sample trees and cannot integrate point mutations. [469] MEDICC2 is a software that is specifically designed for CNA evolution, but it does not allow inclusion of point mutations. [470] Hence, there is a lack of tools that allows the integration of both CNAs and point mutations in the subclonal deconvolution procedure for multiregional sampling data from the same patient.

To fill this methodological gap, we created DEVOLUTION, a software that mitigates some of these problems. It allows the user to create phylogenetic trees using CNA data alone, SNV data alone, or by combining the two modalities in the same phylogeny. The input of the software is a so-called segment file or mutation list. It is a table in which, for each sample, all genetic alterations are present, along with their genomic location, the type of alteration and in which proportion of tumor cells it was detected. During the subclonal deconvolution the algorithm initially clusters genetic alterations using the dbscan algorithm. [471] Subsequently it nests the clusters based on the pigeonhole principle, while integrating information available from all samples during the inference procedure, trying to minimize occurrences of back mutations and parallel evolution (**Figure 21**). The output file contains information about which subclones are present in each sample along with the proportion of cells having each genotype. This information is further used to reconstruct phylogenetic trees using the MP and ML method. The software can also generate several phylogenetic trees when there are multiple possible solutions.

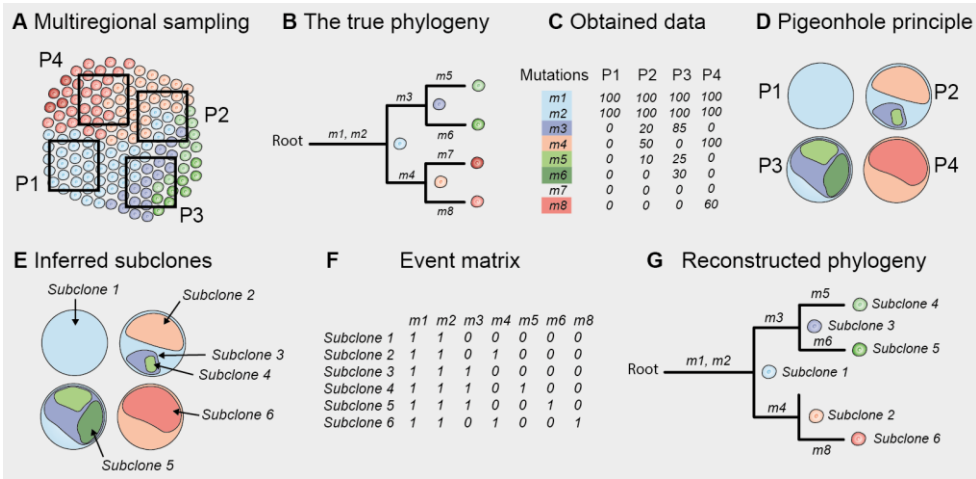


Figure 21. An illustration of the basic principle of the DEVOLUTION algorithm. **A)** A simplified tumor. Each shade of color represents a unique subclone. P1-P4 represents four biopsies collected from the tumor. **B)** The true evolutionary relationship between the subclones. This is not known. **C)** Using the data obtained from bulk DNA genotyping it is possible to calculate the proportion of cells having each genetic alteration in each sample. The numbers in the matrix correspond to the percentage of cells having each of the genetic alterations (y-axis) in each sample (x-axis). **D)** By using the pigeonhole principle, it is possible to deduce which subclones are present in each sample and to what extent. In this example there is only one possible solution for the nesting that does not result in back mutations or instances of parallel evolution. **E)** Based on the nesting we can infer which subclones are present across samples. **F)** An event matrix is generated, illustrating which genetic alterations are present in each subclone. **G)** Based on the event matrix, a phylogenetic tree can be reconstructed. Note here that the subclone having mutation m7 is not included in the phylogeny, since no cell having this alteration was sampled.

Software: Single cell event matrix (SCEM)

To create phylogenetic trees based on scWGS data for paper II and III, a software named *Single Cell Event Matrix (SCEM)* was developed. It is freely available at github: https://github.com/NatalieKAndersson/SCEM.

It uses the raw single cell copy number data obtained after using Aneufinder as input. [454] In the data matrix, each column corresponds to a single cell, with each row representing a genomic bin covering approximately 5 Mbp. The value of each matrix element indicates the copy number at the corresponding genomic position. Before running the algorithm, the user specifies the cutoff for a copy number being considered a *stem event*, both within and between groups. A stem event is defined as a copy number alteration that all cancer cells have had. If the cutoff is set to e.g., 90 %, it means that if a CNA is present in at least 90 % of the analyzed cells, it is inferred in the remaining cells as well, and has subsequently been lost in these cells. The algorithm also allows the user to specify whether some cells are grouped in some ways e.g., have been collected at the same time point or area, or

if the dataset involves several patients, each patient can be designed to a group. This is taken into consideration throughout the algorithm when inferring alterations. The user can also specify the cutoff for stem events considered within such groups.

The algorithm begins by merging columns having identical genomic profiles. The number of cells represented by each column is annotated in row two in the matrix. Subsequently the algorithm loops through each row for each column (representing a single cell or a group of cells with identical genomic profiles) in the matrix. The start and end positions along with the copy number of each CNA for each column is stored in a separate matrix (**Figure 22**).

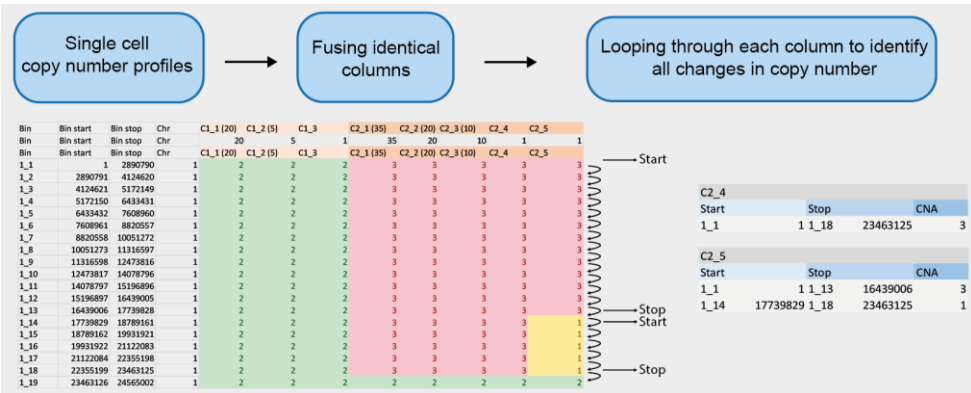


Figure 22. An overview of the first steps of the SCEM algorithm. The algorithm takes single cell copy number profiles as input and first fuses columns with identical genomic profiles. The second row in the matrix indicates the number of cells with each genetic profile. For example, C1_1 represents 20 cells and C1_5 represents 5 cells. The algorithm then loops through each row for each column. It identifies all changes in copy number and stores the information in a separate matrix, exemplified by cell C2_4 and C2_5 here.

Subsequent steps of the algorithm handle various situations that can appear in the data set (**Figure 23**). Further details can be seen in the extensive manual in the github page: https://github.com/NatalieKAndersson/SCEM.

(1) *Overlapping CNAs:* After a CNA, additional CNAs can occur, covering the same region.

(2) *Isochromosome events:* It involves the loss of one arm and a duplication of the other. If there are cells having an isochromosome event it might need to be inferred for other cells being subject to consecutive CNAs covering the same region.

(3) *Duplications or losses of chromosomes:* Duplications or losses of chromosomes will cause an integer copy number increase or decrease of all bins covering the duplicated or lost chromosomal segment. The previous copy number state thus need to be inferred in the event matrix.

(4) *Consecutive events:* There might be CNAs having the same break points but with different copy number states. In these cases, the one with the lowest copy number is set to be an inferred event in the cells with consecutive alterations.

The algorithm outputs an event matrix which can be used to reconstruct phylogenetic trees using e.g., the MP and ML method. See the github page for a flow chart of the workflow along with examples.

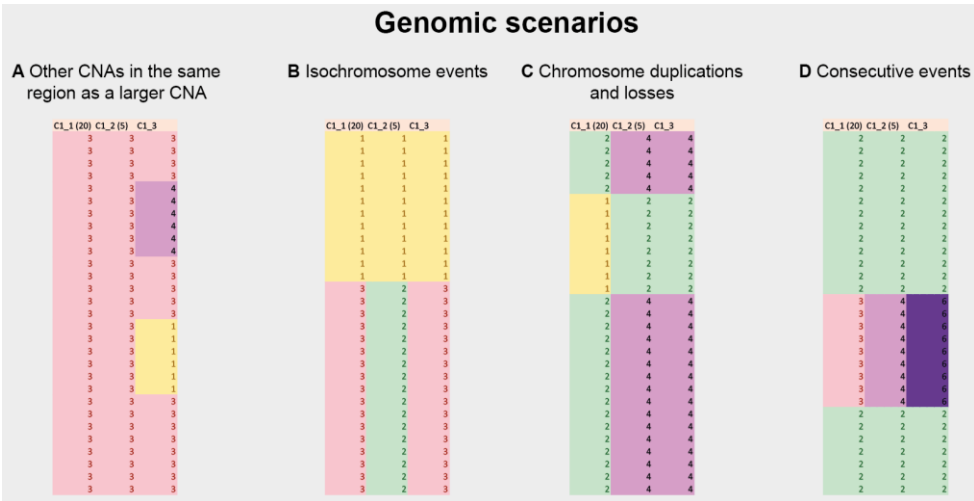


Figure 23. An overview of some of the situations that can arise in the scWGS dataset. The SCEM algorithm handles all of these to not underestimate the evolutionary relationship between the cells. There is also the possibility to add information about whether some of the cells are grouped in some way, for example by being collected at the same time point or belonging to the same patient, if several patients are included in the input data set. This is taken into consideration during the inference procedure when stem alterations are being identified. **A)** Cell C1_3 can be interpreted as having five separate CNAs in the area. It is, however, evident from the information in C1_1 (encompassing 20 cells) and C1_2 (encompassing 5 cells) that there most likely has been a copy number of 3 initially, followed by two smaller CNAs resulting in a small region with copy number 4 and another with copy number 1. Hence, C1_3 is considered having three separate CNAs in the visualized region, by the SCEM algorithm. **B)** The cells encompassed by column C1_2 have an isochromosome event followed by a loss the second part of the event. **C)** If a cell has undergone a duplication or loss of a region already encompassing a CNA, it will cause a shift in the copy numbers of the corresponding bins. In these cases, the previous copy number state have to be deduced, using information from other cells in the data set, and subsequently inferred in the event matrix. Here Cells C1_2 and C1_3 most likely have had the profile seen in the 20 cells represented by C1_1 followed by a duplication of the entire visualized region. **D)** Some cells might undergo additional copy number gains of a segment already affected by a gain. In these cases, the one with the lowest copy number state is inferred in the cells having consecutive alterations.

Software: The modified maximum parsimony method (MMP)

In paper IV a novel method for phylogenetic reconstruction named *the modified maximum parsimony* (MMP) method was developed. This software integrates an

event matrix with information regarding the proportion of cells in each sample having each genetic alteration. The goal is to minimize the number of events in the phylogeny, similar to the MP method, while also adhering to the pigeon principle to minimize occurrences of backmutations.

The input to the MMP algorithm is the output of the DEVOLUTION algorithm. Specifically, MMP requires an event matrix indicating which genetic alterations are present in each subclone, pie sizes representing the proportion of cells in each sample with a particular genotype, and a heatmap depicting the distribution of genetic alterations across samples (**Figure 24**).

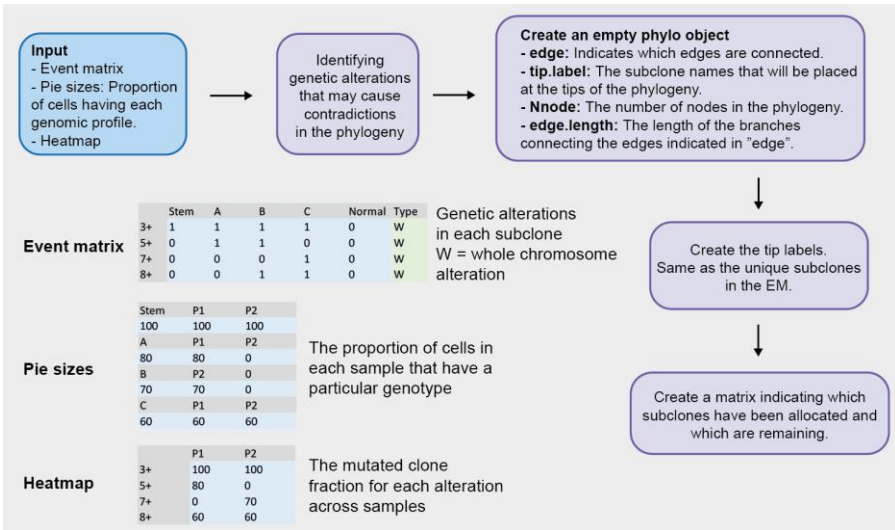


Figure 24. An overview of the initial steps of the MMP algorithm. It takes three matrices as input: 1) An event matrix illustrating the presence or absence of genetic alterations in each subclone. The column "Type" indicates whether the CNA is a whole chromosome alteration. These will be allowed to be parallel in the phylogeny. 2) The proportion of cells in each sample having a particular genotype. 3) A heatmap illustrating the MCF of each genetic alteration in each sample. Before running the main MMP algorithm the software scans through the dataset to identify possible instances of parallel or back mutations and alerts the user if that is the case. The user can thus choose whether to continue running the code or halting. The main MMP algorithm code begins with the creation of an empty phylo object (the main object used for storing information about a tree structure). Furthermore, a matrix is created, indicating which subclone have been allocated to the phylo object and which ones remains.

The software analyzes the event matrix and alerts the user of whether or not there are contradictions in the dataset, that might lead to back mutations or parallel evolution in the phylogeny. If the user wants to proceed, an empty phylo object is created. It is an R object that will contain the information needed to plot the phylogeny. The software also creates two matrices that will be used in the main loop: (1) *allocated*: Indicates which subclones have been allocated to the phylo object. (2) *remaining*: Indicates which subclones have not yet been allocated to the

phylo object. The algorithm loops from 1 to the total number of subclones, including the ones denoted stem and normal in the event matrix (**Figure 25**).

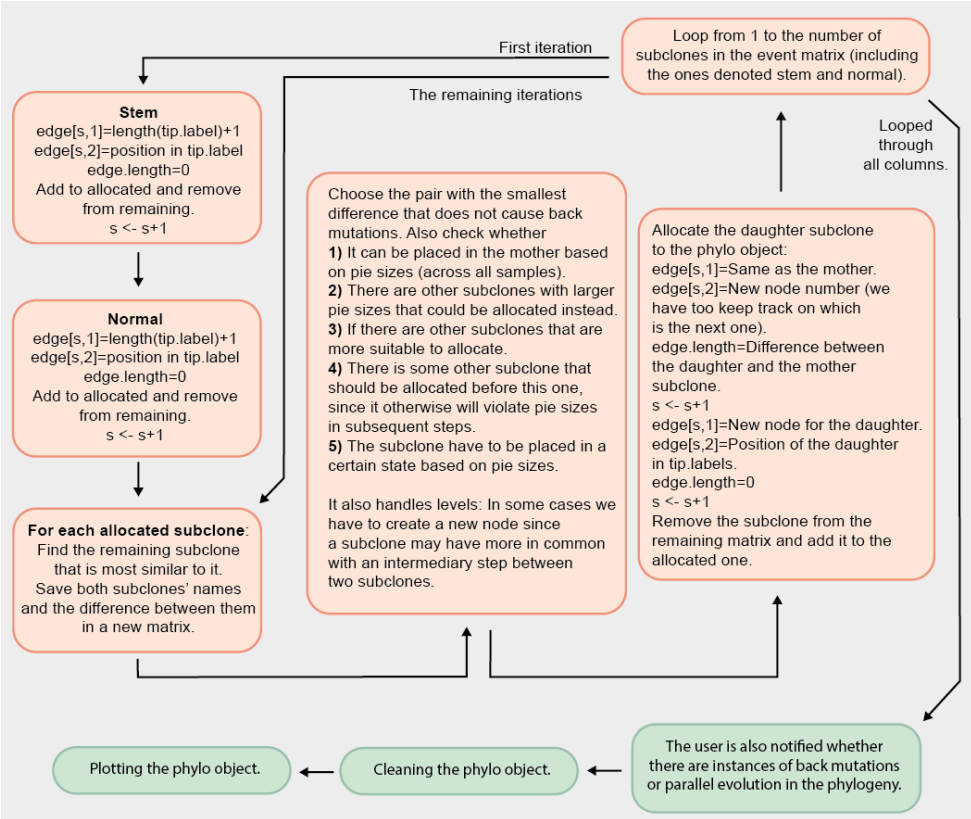


Figure 25. Illustration of the main loop of the MMP algorithm and the last three steps to obtain the final phylogenetic tree. Subclones are iteratively added to the phylo object, providing information of how to plot the phylogenetic tree. In the first iteration of the loop, the column denoted stem, encompassing the stem alterations, and the normal cell, are considered. They are allocated to the phylo object i.e., information is stored about how to plot their evolutionary relationship. For the remaining subclones the algorithm uses a set of strategies to choose which subclone to allocate to the phylo object. When all subclones have been allocated, the loop is ended, and the phylogenetic tree is plotted. The user is also informed about whether there are instances of back mutations or parallel evolution in the phylogeny.

For the first iteration information about where the stem and normal should be plotted are stored. These are now said to be "allocated" to the tree structure. For each subclone that is allocated to the tree structure, the most similar not allocated subclone is identified. The software considers several situations when deciding on which subclone should be allocated to the tree structure. When a subclone has been selected, information about it is added to the phylo object. The subclone name is also added to the *allocated* and removed from the *remaining matrix*. When all subclones have been allocated, the loop is ended. The user is informed

about whether there are instances of back mutations or parallel evolution in the final phylogeny, followed by plotting the phylo object, resulting in a MMP tree.

Results

Paper I

DEVOLUTION – A method for phylogenetic reconstruction of aneuploid cancer based on multiregional genotyping data

This paper involved the development of a software for subclonal deconvolution, denoted DEVOLUTION (see the methods section for more details). It was validated using data from 56 pediatric tumors including 22 neuroblastomas, 20 Wilms tumors and 8 rhabdomyosarcomas, comprising a total of 253 biopsies. For each patient, phylogenetic trees were reconstructed using the maximum parsimony method and the maximum likelihood method. For 20 patients there was both SNP-array data as well as sequencing data, in which cases phylogenies were reconstructed using the SNP-array data alone, sequencing data alone, as well as with both datasets conjointly.

We showed that the phylogenies represented plausible biological scenarios and illustrated key events in the evolution of these tumors. For example, genetic alterations known to arise early in the evolutionary history of NB, WT and RMS were found in the stem of the phylogenies in almost all patients. DEVOLUTION also turned out to be a useful tool for tracking metastatic spread, e.g., in one patient denoted NB5 (**Manuscript Figure 2a**) there were signs of polyclonal seeding from the NB primary tumor to the bone marrow, indicating that several subclones from the primary tumor colonized the same distant site. A similar scenario was seen in patient NB22 (**Manuscript Figure 2b**), where signs of polyclonal seeding was seen in several metastases, suggesting that it might be a common feature in these tumors. Additionally, there were also tendences towards subclonal variation being higher among lymph node metastases compared to distant sites across patients, suggesting a weaker threshold for metastatic spread to the lymph nodes compared to distant sites.

To assess the accuracy of DEVOLUTION, the algorithm was tested on simulated data. Three tumors with increasing mutation frequency was simulated. Virtual biopsies were taken at random sites across these tumors. Each virtual biopsy contained a certain number of cells, each harboring a set of genetic alterations. This allowed calculation of the proportion of cells in each sample having each genetic alteration, providing the input data for the DEVOLUTION algorithm. Subsequently, the DEVOLUTION algorithm computed which subclones were

likely present in what sample and to which proportion. This was followed by reconstruction of phylogenetic trees based on this data. The generated phylogenetic trees and subclonal deconvolution was compared to the true evolutionary relationship and subclonal composition, known via the simulations. The data revealed a robust performance of DEVOLUTION. As expected, the *total absolute number* of correctly timed genetic alterations increased with the number of biopsies at hand, but surprisingly the *proportion* of correctly timed events did not. The reason for this is that when increasing the number of samples, the chance of finding an area with a late genetic alteration that is hard to allocate correctly in the phylogeny, also increases, due to limited information. Upon excluding genetic alterations present in only one biopsy, the result was almost perfect for this simulated dataset.

Finally, an external WES dataset from the TRACERx project on non-small-cell lung cancer was used to benchmark the software against a similar bioinformatic tool called MAGOS. [458, 472] DEVOLUTION provided a more detailed picture of the tumor, since MAGOS could only include somatic point mutations present in all biopsies analyzed.

In conclusion, DEVOLUTION, provides a standardized framework for subclonal deconvolution and phylogenetic analysis of multiregional data, aiding tree-to-tree comparison during downstream analysis. It additionally allows reconstructing phylogenetic trees using both CNAs and SNVs. By analyzing a tumor's phylogenetic tree, an overview of the genetic heterogeneity can be obtained, and it allows the user to identify early genetic alterations. These mutations pose as attractive therapy targets since they are present in most, or all cancer cells in the tumor. The phylogeny can also be used to dynamically track subclones in time and space. DEVOLUTION makes it possible to identify which subclones survive therapy, which ones make up relapses, which ones metastasize and when as well as identify how different metastases are related to one another. Hence, by using DEVOLUTION, it is possible to unlock the evolutionary history of the tumor and gain increased insights into the evolutionary dynamics of the tumor during treatment and tumor progression, which may have important clinical implications.

Paper II

Clinically relevant treatment of PDX models reveals patterns of neuroblastoma chemoresistance

This study sought to uncover COJEC treatment resistance in high-risk NB, through analysis of genomic and transcriptomic changes during treatment and at relapse using high-risk NB PDX models.

Tumor samples from three patients were dissociated, and organoids were formed which were implanted subcutaneously in nude and NSG mice. One high-risk NB

patient derived xenograft (NB-PDX) mice model [429, 431] was established for each patient, labeled as PDX1, PDX2 and PDX3. All three PDX models had a *MYCN*-amplification, 1p deletion and 17q gain (**Manuscript Figure 1A**). PDX1 encompassed a total of 33 mice, PDX 22 mice and PDX3 41 mice. When the tumor reached a size of 500 mm^3 the mouse was randomized into one of three treatment arms. The mice were either treated with 1) a COJEC like *in vivo* protocol, 2) cisplatin to simulate a poor treatment choice or 3) saline solution to simulate free tumor growth. All agents were administered intraperitoneally. Samples were collected from the parental tumor organoids before injection, from the tumor after therapy and at relapse. RNAseq was performed on all collected samples while SNP-array and single cell low-pass WGS was performed on a subset.

Response to treatment

PDX1 was derived from a COJEC refractory tumor, while PDX2 and PDX3 were from tumors that showed an initial response followed by a later relapse. As expected, all mice treated with saline solution rapidly progressed in their disease. In PDX1, all mice treated with cisplatin displayed progressive disease, while there was a trend of mice treated with COJEC surviving a bit longer, but this difference was not statistically significant compared to treatment with saline solution alone. A subset of mice in the PDX1 group were subjected to high-dose COJEC (HD-COJEC). It significantly increased the survival but still resulted in progressive disease in 7/8 mice and stable disease in 1/8 mice, thus mirroring the lack of response seen in the corresponding patient. PDX2 displayed a significant tumor size reduction and increased survival in response to either cisplatin or COJEC compared to control mice. In 5/8 tumors, a partial response was seen, in 2/8 stable disease and in 1/8 a complete response with a relapse 110 days after treatment initiation. Finally, PDX3 displayed a significant tumor size reduction and increased survival in response to either cisplatin or COJEC, but 5/7 rapidly progressed after treatment was halted and 2/7 had a complete response but later relapsed. To resemble the clinical regimen in patients, 9 mice underwent surgical removal of the remaining tumor after COJEC treatment and tumor shrinkage < 200 mm^3. Of these merely 2 relapsed, while the others were classified as cured (**Manuscript Figure 1B**). Collectively, the PDX models mirrored the diversity in responses seen in the corresponding patient group.

Histology

COJEC treated PDX1 tumors were the least differentiated and PDX3 the most. All tumors were positive for PHOX2B and had an intense staining for Ki67. In addition, PDX1 and PDX2 had large and collapsed blood vessels, a feature related to a worse prognosis, [473] while the vessels were smaller and open in PDX3.

Bulk whole genome copy number profiling

Whole genome copy number profiling was performed on the parental organoids as well as on samples after therapy and from relapses. This was followed by phylogenetic analysis using DEVOLUTION, developed in Paper I. [474] For each PDX model, one phylogenetic tree was reconstructed using the ML method. All PDXs displayed a high number of subclones and branching evolution.

In PDX1 the parental organoid had undergone a whole genome doubling (WGD), giving rise to a vast number of subsequent subclones. The copy number aberration burden (CNAB) and genetic diversity was, surprisingly, comparable between the group treated with COJEC and the group treated with saline solution, suggesting that the COJEC treatment did not induce additional CNAs. The fraction of private aberrations was, however, significantly higher in the treated samples, indicating diverse evolutionary trajectories despite arising from the same parental material. In addition, two of the treated tumors displayed a selective sweep.

PDX2 had fewer private clones and a lower CNAB compared to PDX1. The genetic diversity and CNAB were also comparable between COJEC treated samples and controls. No selective sweeps were seen, but treatment specific copy number losses were identified on chromosomes 22q (*LARGE1*) and 8q (*RABA2A* and *CDH7*).

Finally, PDX3 had additional copies of an already gained 17q+ and 1pq, both of which were detected in the majority of samples. The 17q+ and 1pq+ gains were also subject to parallel evolution. These gains were seen both in the controls and the treated samples, suggesting that they are selected for by intrinsic evolutionary forces rather than by the chemotherapy.

The CNAB was increased in relapsed tumors and in regrown tumors compared to the other treatment groups. There was, however, a linear correlation between high CNAB and how long time the tumor had grown, hence again suggesting that the COJEC treatment did not give rise to additional genetic alterations. The CNAB increased over time, while the genetic diversity remained stable.

Across the PDX models, small deletions of certain genes were detected. Interestingly, deletions of *MACROD2* and *LSAMP* were found in both controls and treated samples across all three PDX models. *MACROD2* has been suggested to cause CIN, [475] and it was associated with heavy branching and accumulation of CNAs. *LSAMP* encodes a neuronal surface glycoprotein and has been suggested to be a tumor suppressor. [476] *PTPRD* was deleted in PDX1 and 3 and is a tumor suppressor in NB. [477] A low transcriptional expression of the regions covered by the small deletions, correlated with poor prognosis both in the PDX3 model and in patients (**Manuscript Figure S9 G-I**).

In conclusion, no specific genetic subclone was found to be responsible for chemotherapy resistance and relapse in any of the PDX models.

Single cell DNA sequencing

ScWGS copy number analysis was performed on samples from the intrinsically resistant PDX1 and the responsive PDX3. The software SCEM, developed as part of this paper (see the methods section for more details), was used to analyze the scWGS data. Interestingly, PDX1 displayed an extensive genetic diversity with no subclone being detected in more than one sample, possibly contributing to an evolutionary mechanism for treatment resistance in this model. In PDX3 two major subclones denoted A (1p-, MNA, 5p+, 17q+) and C (1p-, MNA, 5p+, 17q+, 1pq+, 17q++) were identified across all tumors. Subclone C was enriched after treatment but was also found in the controls. Individual NB cells also seem to be able to accumulate high numbers of CNAs. Consistent with the results from bulk DNA analysis, no significant enrichment of distinct subclones after treatment were found that could explain the treatment resistance. Hence, the single cell data confirmed the absence of resistance specific subclones.

RNAseq analysis – Identification of specific cell states

Bulk RNAseq was performed on all tumors. The samples clustered, as expected, strongly by PDX model. Tumors that were cured displayed increased expression of nervous system development genes and reduced expression of cell cycle and DNA repair genes. In contrast, treatment-resistant or relapsed tumors displayed an increased expression of genes related to early embryonic development and downregulation of genes related to nervous system development. COJEC treated relapsed NB PDXs were shown to resemble human immature Schwann cell precursors, while the COJEC treated tumors that were surgically removed and the mice later cured, had a high expression of genes correlated to neuroblasts and chromaffin cells.

We subsequently sought to investigate whether there were signs of the ADR and MES cell state in the PDX cohort. The presence of these states have been confirmed *in vitro* [218] and in patient samples, [478] but not yet *in vivo*. Since there is some debate regarding which gene signature most correctly represents the ADR and MES cell states, six publicly available ADR- and MES-like gene signatures were used in this study. The mice that were COJEC treated, underwent surgery, and were later cured showed an enriched ADR signature, while relapsed tumors showed an enriched MES-like signature for all six gene signature datasets individually.

A merged gene signature was constructed by combining all the six individual ADR and MES gene signatures, respectively, together with clusters specific for cured and resistant tumors in our cohort, creating the merged datasets “merged ADR” and “merged MES”. The genes involved in the merged ADR dataset was enriched in tumors treated with COJEC followed by surgery where the mouse was cured, while the merged MES dataset was enriched in relapsed tumors. A low expression of the merged ADR signature correlated to poor prognosis in NB

patients. Surprisingly, a high expression of the merged or individual MES signatures did not correlate to prognosis. To investigate this further, the merged ADR and merged MES signatures were subdivided further into an ADR-cured, ADR-relapse, MES-cured and MES-relapse subsignature, based on the PDX data. The ADR-cured and MES-relapse subsignatures were further filtered through a cohort of 219 stage 3 and 4 NB patients to generate an integrated ADR signature and an integrated MES signature. The ADR-cured subsignature was significantly enriched in mice that underwent surgery after COJEC treatment and did not get a relapse (Sur-cured) and the MES-relapse subsignature was enriched in relapse and regrown tumors. To conclude, the ADR- and MES-like cell states do occur in *MYCN*-amplified NB PDXs and a mesenchymal phenotype might confer treatment resistance in NB.

MYCN

The *MYCN* gene is essential in NB pathogenesis. PDX1 both had a higher number of copies of *MYCN* as well as an increased RNA expression of it, compared to PDX3. Relapsed tumors had higher levels of *MYCN* RNA as well as fraction of N-MYC protein expressing cells compared to COJEC treated tumors that were surgically removed and cured.

Organoids

3D NB organoids were established from control as well as COJEC treated tumors from all three PDX models. By administering COJEC it was shown that organoids derived from COJEC treated and relapsed tumors retained their chemoresistance as well as transcriptional signature *ex vivo*. Relapsed tumor organoid cells were also re-implanted back into mice and did maintain their tumorigenic capacity *in vivo*. Consequently, these organoids provide a powerful model system to try novel treatments for resistant NB in future studies.

Conclusions

In this study a COJEC-like treatment protocol was developed, which was applied to three *MYCN*-amplified NB PDX models. We found an extensive intra- and intertumoral genetic and transcriptional heterogeneity. Mice displaying upfront progression during treatment or relapse had an increased clonal diversity, higher copy number and transcriptional levels of *MYCN*, and displayed an immature MES-like cell state. In contrast, mice that were cured after COJEC and subsequent surgery displayed an ADR-like cell state and had a lower number of copies of *MYCN* and transcriptional levels. Hence COJEC treatment seems to select for cells with a MES-like phenotype, and these cells might contribute to treatment resistance in NB. A predominant ADR signature after treatment were on the other hand correlated to a good overall prognosis in mice. Some cells displayed both ADR- and MES-like markers, indicating that they are in an intermediate state, as has previously proposed to exist. [328, 347] The overall finding was that NB

chemoresistance seems to primarily be mediated by transcriptional, rather than, genomic changes. The 3D organoids from relapsed tumors retained their chemoresistance and transcriptional features *ex vivo* (**Manuscript Figure 9**) and may hence be used for drug development in future studies.

Paper III

Early evolutionary branching across spatial domains predisposes to clonal replacement under chemotherapy in neuroblastoma

Little is known about how the NB subclones surviving chemotherapy are related to the ones present at diagnosis. This information is essential for the development of new treatment strategies for these patients.

In this study, DNA was extracted from FFPE tissue samples collected prior to and after chemotherapy treatment in 12 patients, encompassing a total of 89 tumor areas. Whole genome genotyping was performed on all samples and WES on a subset. Variants detected with WES were confirmed with TDS using a customized panel. This was followed by subclonal deconvolution using DEVOLUTION developed in Paper I. [474] The identified subclones could furthermore be mapped across anatomical sites, creating so-called clonal geographies. In addition, scWGS data was available before and after treatment. The software SCEM, developed in paper II was used to analyze this data. [219] This setup was used to compare the subclones detected before and after therapy for each patient.

Spatiotemporal genomic profiling revealed two contrasting evolutionary patterns in response to chemotherapy

All 7 patients whose tumor responded significantly to chemotherapy, and whose tumors could be evaluated before and after therapy (Patients 3-9), showed a surprisingly similar ancestral relationship between the subclones detected before and after treatment. Subclones present prior to chemotherapy were depleted and replaced by subclones with a different most recent common ancestor (MRCA), a pattern we denoted *collateral clonal replacement* (CCR). This pattern suggests intense selection by the chemotherapy treatment for other subclones than the major ones at diagnosis. The pattern appeared irrespective of whether the sample was collected from the primary tumor, from a metastasis at the time of diagnosis (Patients 8-9), or when comparing a sample from the primary tumor to a later metastatic relapse (Patients 4, 9-12). CCR could also develop gradually over time (Patient 8). There was no overall increase in the number of genetic aberrations after treatment compared to before, but an increased genetic intratumoral diversity. An example of CCR can be seen in **Manuscript Figure S5g-i**.

In contrast, tumors that showed a poor treatment response with inferior killing of cancer cells, sometimes even progressing during treatment, displayed a pattern of *linear evolution*. The subclones after treatment were linear descendants to the ones present before treatment indicating that the major subclones just continued to evolve, without being affected by the chemotherapy. An example of a tumor presenting linear evolution can be seen **Manuscript Figure S5f**.

In vivo mimicking of progressive growth under chemotherapy and effective chemotherapy treatment

Since the number of patients was limited, a *MYCN*-amplified NB PDX model system was designed. To mimic progressive tumor growth under chemotherapy, mice were treated with cisplatin as monotherapy (n=6). To mimic progression after incomplete surgery, resulting in a bottleneck, mice underwent partial surgical resection of their non-chemotherapy exposed tumors, followed by regrowth (n=4). Mice in the control group were treated with saline solution (n=4). Two samples were collected from each of the tumors after treatment. The genomic profile was compared to the parental cell culture. Phylogenetic analysis revealed linear evolution under ineffective chemotherapy treatment, just as was seen in the patient cohort.

To mimic progression under multimodal chemotherapy treatment a different PDX model was used, known to be resistant to chemotherapy treatment with COJEC. [219] Samples were collected after adequate treatment with COJEC (n=3) or after free growth (n=3) for comparison. Both the treated and untreated tumors exhibited linear evolution. Hence, the PDX models confirmed the pattern of linear evolution under ineffective treatment.

To mimic the situation of effective chemotherapy tumor response a PDX model system known to be sensitive to COJEC treatment was used. [219] Mice were randomized into either being treated with COJEC (n=9), followed by surgical resection, or saline solution (n=9). Two samples were collected from each tumor after treatment. Just as in the patient cases there was no increase in the total number of genetic alterations, but an increase in the genetic intratumor diversity after treatment. Two treated mice had a later relapse, in which the same pattern of CCR, as seen in the patients, could be seen. Hence, the PDX models confirmed the pattern of CCR observed during effective treatment.

Collateral clonal replacement in vitro

To further investigate the subclonal landscape before and after chemotherapy the cell line IMR-32 was used. It is derived from a high-risk *MYCN*-amplified NB.

Three biological replicates were allowed to grow freely, three were subjected to low-dose cisplatin and three to high-dose cisplatin. Free growth resulted in fixation on a certain subclone. Low-dose cisplatin did not alter the subclonal landscape. When cells were subjected to an adequate treatment with high-dose cisplatin killing 94 % of cells, CCR was seen. Creating a mechanical bottleneck by randomly removing 94 % of all cells did not result in CCR. Neither did creating three successive mechanical bottlenecks, reducing the size by 50 % each, induce CCR. In contrast, administering cisplatin at three different time points, also causing a 50 % population size reduction each, did. Hence, this supports that CCR is caused by Darwinian selection under chemotherapy, and not just the reduction of the number of cells.

We also used the SK-N-SH cell line, which is known to have two NB cell types, of which one is more resistant to chemotherapy than the other. These two cell types display gene expression signatures corresponding to the ADR (sensitive) and MES (resistant) cell states. [328, 479] Cells were either treated with cisplatin or allowed to grow freely. In the latter experiment the subclone present at initiation was retained. The treated cells displayed CCR, and the cells shifted to an epithelioid morphology, corresponding to a selection of resistant MES-like cells.

Spatial territories

Since the extracted DNA was derived from FFPE-blocks it was possible to map subclones to their corresponding anatomical areas. In untreated tumors clonal territories showed a dense pattern with no obvious anatomical barriers between them. In the primary tumor after chemotherapy on the other hand, there were vastly dispersed islands of surviving tumor cells surrounded by necrosis, hemorrhage, and other reactive changes. Each such compartment often contained a clonal population of cancer cells and was represented by a unique branch in the phylogeny. This stresses the importance to, when feasible, sample such distinct tumor compartments for genomic analyses to fully characterize the genomic landscape of the tumor (**Manuscript Figure 5k**).

MYCN amplicon diversity

An in-depth analysis of the amplicon copy number architecture was performed for 6 patients with *MYCN*-amplification. An amplified state was identified across all biopsied areas. Surprisingly, spatiotemporal heterogeneity in the *MYCN* amplicon breakpoints was seen in all 6 patients. All these patients had at least one shared breakpoint across all samples. In total 76 % of samples obtained after chemotherapy exhibited variant (non-stem) breakpoints while merely 30 % of the untreated primary tumor samples did. This indicates that there is a common breakpoint and that it was followed by further modifications.

Single cell analysis

Low-pass single cell WGS was performed on 9 NB primary tumors, totaling 505 single cells. The software produced in project II (SCEM) [219] was used to create phylogenetic trees, revealing a surprisingly extensive phylogenetic branching in all analyzed tumors. Most cases had one or several groups of cells with identical profiles, but also a plethora of cells with unique genomic profiles. In one case no two cells had the same CNA profile, revealing a surprisingly extensive diversity. In 5/6 cases, *MYCN* copy number gains preceded 17q gain. In addition, amplicon numbers were significantly higher in cells having acquired 17q+ and 1p-. The presence of *MYCN* both with and without 17q+ was also confirmed using FISH where cells with *MYCN* amplification but lacking the 17q+ alteration were found. Single cell analysis thus identified early ancestral cells not ascertained by bulk sequencing.

To further investigate the extensive phylogenetic branching, scWGS was performed on samples from three PDX models of *MYCN*-amplified NB monitored for 7 generations, followed by establishment, monitoring, and passaging of tumor organoids for 7 generations. This analysis confirmed that branching evolution was a consistent feature *in vivo* as well as *in vitro* in NB.

Conclusions

In this study we comprehensively compared the subclonal landscape of NBs before and after chemotherapy. Tumors which significantly responded to treatment, showed a pattern of CCR where the subclones present at diagnosis were depleted by the chemotherapy, selecting for other subclones in the tumor, eventually regrowing and resulting in a treatment resistant relapse. In contrast, tumors showing poor treatment response had subclones after chemotherapy that were linear descendants to the ones present before treatment. These two patterns were confirmed both *in vivo* in PDX models and *in vitro* in NB cell cultures. CCR suggests a chemotherapy induced selection of pre-existing, possibly resistant, NB cell populations. Treatment induced selection of pre-existing resistant cancer cell subclones has been proposed for several other malignancies. [322, 480] The scWGS analysis revealed extensive phylogenetic branching to be a consistent feature of NB both in patients, *in vivo* and *in vitro*, emerging already at the ancestral states, forming the substrate for CCR.

Our study has important implications for how to sample these tumors for genomic analyses in the clinic. During upfront progression, new genetic alterations are added to the pre-existing ones while tumor shrinkage followed by a relapse implies a radical shift in the subclonal landscape, necessitating resampling of the

tumor. The spatial analysis revealed the importance of sampling distinct anatomical compartments to characterize the genomic landscape.

Paper IV

Resolving the pathogenesis of anaplastic Wilms tumors through spatial mapping of cancer cell evolution

This study aimed at identifying key events in the evolutionary history of DAWT compared to other WT subtypes. It was performed via spatial mapping of 189 tumor areas from 20 WTs, including 10 intermediate risk tumors, 5 blastemal type tumors and 5 DAWTs. When possible, samples were taken from whole mount sections (WMS), each comprising 0.5 cm thick and up to a decimetre in diameter large FFPE samples of the patient's tumor. Grids composed of 1x1 cm large squares were marked and DNA was extracted from each such square, excluding areas encompassed by necrosis, regression, normal tissue, or debris, resulting in a median of 9 samples per tumor (**Manuscript Figure 1**). SNP-array as well as deep sequencing of the tumor suppressor gene *TP53* was performed on each sample along with microscopic assessment of the corresponding area.

Based on the SNP-array data, for which 169 samples was successfully analysed, phylogenetic trees could be reconstructed for 16 patients, using DEVOLUTION [474] in combination with the MP and ML method as well as our developed method for phylogenetic reconstruction MMP (details described in the methods section). When there were no contradictions in the dataset, MP, ML and MMP generated identical phylogenies. When there were contradictions, such as crossovers of clone sizes in the data set, the MMP method produced the most biologically reasonable scenarios. In total, 14 out of the 16 tumors, from which we could reconstruct phylogenetic trees, displayed branching evolution or a combination of linear and branching evolution, confirming that intratumoral heterogeneity of CNAs is common in WT.

In all cases with classical DAWT (patients DA_1-4), anaplasia was only seen in areas with at least one *TP53*-mutation. Studies have indicated that anaplasia starts focally and then spreads within the tumor. In this study, the phylogenetic trees suggested that the evolutionary process leading up to anaplasia starts with a single hit of *TP53*. This hit seemed to trigger an extensive branching and saltatory evolution along with additional mutations of *TP53*. These mutations were seen in parallel across different anatomical compartments in the tumor, eventually resulting in multifocal anaplastic lesions.

To assess the complexity of the phylogenetic trees three different measures were used: Phylogenetic species richness (PSR), divergence and irregularity. *PSR* is the sum of the distances between all subclones in the phylogeny, representing the

degree of diversity between the subclones. The *divergence* is defined as the PSR divided by the number of comparisons made, representing the mean phylogenetic richness. The *irregularity* is a measure of how uniformly distant the subclones are from the stem i.e. how much the phylogeny differs from a star phylogeny. It is the variance of the phylogenetic distances from the stem to the subclones (**Figure 26**). DAWT displayed a significantly increased PSR, divergence and irregularity as well as number of CNAs and genetically distinct subpopulations, compared to the other subtypes. We also found an accumulation of CNAs and regressive changes as anaplastic features increased in DAWT, while the number of CNAs only increased with proliferation in the other subtypes.

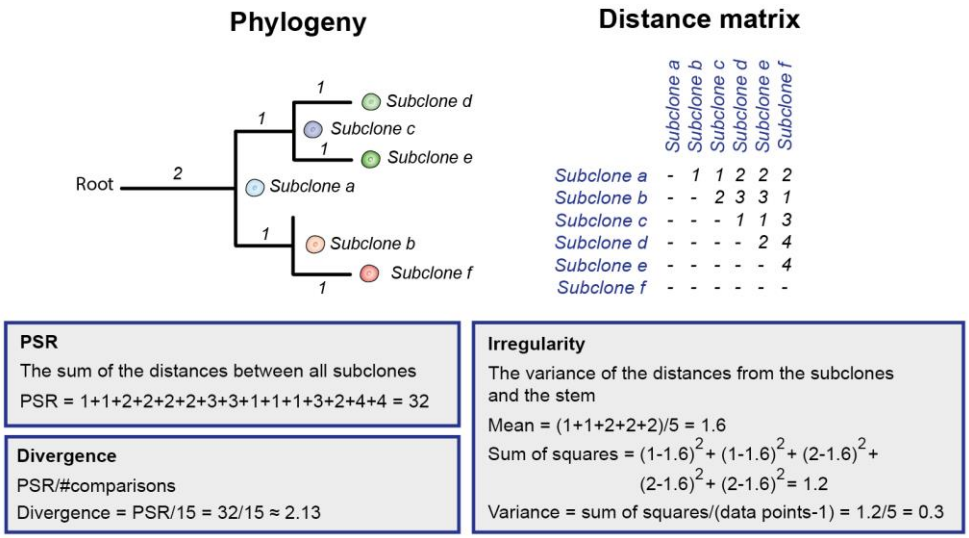


Figure 26. An example of how to assess the complexity of a phylogenetic tree through calculating the PSR, divergence and irregularity.

WT usually displays several macroscopic compartments separated by areas of necrosis, fibrous septae and regressive tissue. We therefore sought to investigate how the subclones varied, both within the compartments, but also between compartments in the same tumor. A total of 170 areas within compartments and 133 between compartments were compared for the 18 patients for which such comparisons were possible. Each comparison was classified according to the four evolutionary trajectories as previously described by our group, encompassing subclonal variation (VAR), tumor cell twinning (TCT), clonal coexistance (COEX) and clonal sweeps (SWE) (**Figure 27**). [481]

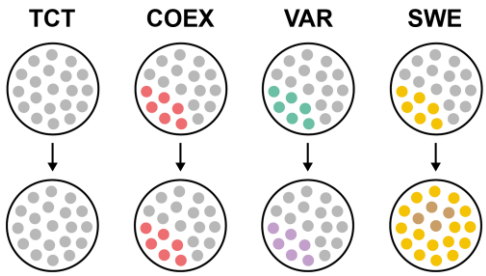


Figure 27. Description of the four evolutionary trajectories, as described in [481], when doing a comparison of two areas in the same tumor. Tumor cell twinning (TCT) refers to a homogenous clonal landscape indicating that there is no genetic intratumoral variation between these two areas. If at least one subclone is prevalent together with its mother subclone in the two compared areas it is referred to as clone coexistence (COEX). Subclonal variation (VAR) describes the presence of different subclones between the two areas. If a subclone in one sample expands and sweeps the other area it is referred to as a regional clonal sweep (SWE).

Within compartments, TCT was observed in 73 % of the comparisons and merely 15% were VAR, 10 % COEX, and 2 % SWE, implying that individual compartments are mainly homogenous in their subclonal composition. In contrast, 76 % of the comparisons between compartments within the same tumor classified as SWE, 20 % as TCT and 4 % as VAR. No comparison classified as COEX. This implied that new compartments form by budding from an already present compartment, rarely encompassing more than a single subclone (**Manuscript Figure 5**).

In conclusion, DAWT displayed complex evolutionary histories compared to the other WT subtypes, defined through a significantly higher PSR, divergence, irregularity, number of subpopulations and CNAs. Hence, extensive phylogenetic tree complexity seems to be a hallmark of DAWT, and *TP53*-mutations appears to be the initiating event for this complexity. All areas with classic anaplasia displayed *TP53*-mutations. The *TP53* mutations were often regional and occurred in parallel, indicating that they are late events, in line with previous studies. [170, 251] Furthermore, WTs were shown to be intratumorally heterogeneous with respect to CNAs. Compartments of cancer cells within the same tumor are mainly homogenous, while the birth process of new compartments involves, in most cases, the budding of a single subclone that subsequently clonally sweeps forming a new macroscopic compartment. The latter stresses the need for cross-compartmental sampling to better characterize the genetic landscape of the patient's disease if testing is to be done for targeted therapy.

Paper V

Elucidating the metastatic trajectories across several pediatric tumors

Since the most common cause of death in pediatric cancer patients is a metastatic and chemotherapy resistant tumor, both surgery, radiation therapy and chemotherapy are unfeasible as curative treatment options for these patients. How metastatic spread occurs in pediatric tumors has to a large extent been unexplored. To cure the patients that today die of their disease, novel treatment strategies are needed, which requires increased knowledge about the metastatic process across pediatric cancers.

In this study, we therefore aimed at elucidating the evolutionary trajectories of metastatic disease in the pediatric tumors, NB, WT and GCT. More specifically the aim was to decipher when in the evolutionary history of the tumor, metastases appear, whether there were signs of monoclonal or polyclonal spread (one or several subclones from the primary tumor colonizing the same metastatic site), mono- or polyphyletic spread (one subclone or several phylogenetic branches being responsible for metastatic spread) and intermetastatic spread (subclones from one metastatic site spreading to another).

Children with metastatic cancer treated at the pediatric oncology center in southern Sweden during the years 2000-2020 were included in the study. Histological assessment was performed on hundreds of samples. Selection of tumor areas was based on the presence of tumor cells of a sufficient amount to be included for genomic analysis. Due to the spatiotemporal genetic heterogeneity of tumors, DNA was extracted from multiple sites from the primary tumor and several metastatic sites for each patient to assess the subclonal landscape. A total of 171 samples from 17 patients were included with a median of 8 (range 3 to 22) samples per patient including a median of 3 (range 1 to 10) samples from the primary and 3 (range 1 to 19) samples from metastases per patient. All samples underwent whole genome copy number profiling and 54 samples from 8 patients also targeted deep sequencing, with a matched normal sample included for bioinformatics analyses. Phylogenetic trees were reconstructed, enabling spatiotemporal tracking of subclones for each patient.

Since the study included a median of 8 samples per patient, it enabled a detailed exploration of the genetic heterogeneity across tumors. Extensive intratumoral and intertumoral genetic heterogeneity was identified both within the primary tumor and metastases, as well as between primary tumor and metastases and between metastases within the same patient. The metastases often also had additional genetic changes that could not be detected in the primary tumor. Similarly, there were several mutations in the primary tumor that could not be detected in any of the metastases.

When metastatic dissemination occurs in the evolution of the tumor is not known. In this study, early dissemination was defined as a metastasis originating from a subclone present before the MRCA. Late dissemination was defined as a metastasis originating from a subclone that are arose after the MRCA of the primary tumor. In 13 out of 18 patients, signs of early metastatic spread could be identified. In 3 of 18 patients there were instances of both early and late metastatic spread.

A total of 6 out of 18 patients displayed both polyphyletic and polyclonal seeding, indicating that metastases can arise several times during the evolution of the tumor and that several distinct subclones can colonize the same metastatic site. It also implies that several subclones in the primary tumor possess metastatic capability.

In total 9 out of 18 patients had metastases in at least two different sites. Strikingly, 8 out of these 9 patients displayed intermetastatic spread. Hence implying that metastases can give rise to new metastases, without involvement of the primary tumor.

In conclusion, our results provide further knowledge of the evolutionary dynamics of metastatic spread in pediatric cancer patients. Often treatment decision for targeted therapy is based on sampling the primary tumor. Metastases often display different genomic alterations than the primary tumor, stressing the need to sample metastases for assessment of therapeutic targets, since metastases may completely lack targets present in the primary, resulting in progression at distant sites despite the primary tumor responding to treatment. Interestingly, we found that metastatic spread can occur early and several times during tumor evolution. Finally, intermetastatic spread is identified as a common feature in pediatric tumors, indicating that a single metastasis can act as a hub for further metastatic spread despite the primary tumor being removed, having important clinical implications.

Discussion

Did we develop new tools for the analysis of the evolutionary history of a tumor?

Indeed, as part of this thesis, three software tools were created: DEVOLUTION, SCEM and MMP.

In paper I, the primary objective was to develop a new bioinformatical tool to reconstruct phylogenetic trees based on multiregional copy number data, with or without the addition of point mutations. At the time, there was a lack in tools capable of doing this. Existing methods either necessitated sequencing data, could only incorporate information from a single sample, or assumed clonality for all genetic alterations, resulting in biopsy trees. DEVOLUTION addressed these limitations by enabling the inference of subclones based on multiregional copy number data alone, mutational data alone, or a combination of both. Evaluation using data from pediatric tumors and simulated data showed a robust performance in depicting subclones having genetic alterations present in more than one sample.

The algorithm employs the pigeon-hole principle to nest clusters of genetic alterations within each sample, which is not especially controversial. Additionally, it aimed to minimize occurrences of back-mutations and parallel evolution by integrating information across samples. Hence, the software assumes that these situations are rare, which might not always be true. Whole chromosome and whole chromosome arm alterations can occur in parallel or undergo back-mutations. Nevertheless, given that there is no way to know what the correct solution is, it is reasonable to assume that the preferred solution is the one minimizing the number of genetic events needed to explain the data, akin to Occam's razor.

One potential limitation of the software is its reliance on the mutated clone fraction (MCF) as input. The MCF is defined as the proportion of cells in the sample harbouring a specific genetic alteration. The calculation of the MCF relies on either the log2 ratio for whole genome copy number profiling data, or the VAFs for SNVs. These values are inherently noisy, and consequently the MCF values used in DEVOLUTION will be as well. This might hamper the deduction of possible subclones since DEVOLUTION heavily depends on these values.

For instance, if an alteration A has a slightly larger MCF than alteration D, the algorithm may not nest alteration A into B, even though, due to the noise, they could potentially have a reversed size relationship. Thus, it becomes necessary to round the MCF values before running the algorithm on the data. In paper I we conducted simulations to assess the accuracy of the algorithm. These simulations did, however, assume perfect MCF-values based on the simulated dataset. Nonetheless it allowed us to show that *if* the MCF-values are accurate, the

algorithm provides robust nesting and estimation of the subclones. In real-word data, however, the MCF values tend to be noisy, thereby potentially affecting the performance of DEVOLUTION. Yet, it is important to note that this challenge is not unique to our software; rather, it is a limitation inherent to any software attempting similar nesting approaches.

The issue becomes more profound with very large datasets, where MCF values might cross over several times in the dataset. This might induce longer branches and instances of parallel evolution and back mutations. A rounding of the MCF-values before analysis, and a stronger clustering using dbscan in DEVOLUTION might be needed in these cases to circumvent this problem. A way to improve the software could be to remove the rigid constraints during the nesting process and allow a certain degree of uncertainty in the MCF values.

Another challenge lies in the inherent ambiguity of subclonal deconvolution solutions. In many cases, there exist several possible solutions to the nesting problem. This ambiguity it not unique to DEVOLUTION but is a challenge for all software trying to perform subclonal deconvolution. DEVOLUTION addresses the problem by allowing generation of multiple solutions. Moreover, the software establishes a standardized framework for deducing subclones, based on preset rules applied to all patients in the study, thereby facilitating tree-to-tree comparisons. In addition, we used the maximum parsimony and maximum likelihood methods for phylogenetic reconstruction. These are well established and robust methods that have been used for decades, providing further robustness to the analyses in paper I-V. It is worth noting that the general tree structure seldom changes between different solutions. Smaller subclones present in single sample can, in many cases, be nested in several different ways. This can give rise to minor differences in the tree structure, by altering the position of some of the smaller, distal, branches in the phylogeny. For larger subclones the evolutionary relationship is often robust despite crossovers in the dataset. Importantly, small subclones that are difficult to allocate, rarely alter the overarching conclusions drawn from the phylogenetic trees.

Additionally, DEVOLUTION offers features such as the ability to color-code the phylogenetic tree to highlight which branches are robust (there are only one solution for the nesting of the subclones encompassed by these branches) and which ones could be placed in another location. It also worth noting that the pie sizes in the trees do not represent clone sizes but is the proportion of cells in each sample harbouring that genomic profile. This could be updated in a new version of the software.

Overall, the application of DEVOLUTION in our studies has increased our understanding of the subclonal evolutionary dynamics in pediatric cancer. Specifically, we have gained insights into treatment resistance and relapse in

neuroblastoma, the pathogenesis of anaplastic Wilms tumor, and elucidated the evolutionary trajectories of metastatic pediatric cancer.

Paper II involved the development of the software SCEM. It takes single cell copy number profiles, obtained after using Aneufinder, [454] as input and outputs an event matrix which can be used for phylogenetic reconstruction. However, determining the sequential order of some CNAs is challenging, since multiple plausible solutions often exist. For example, if some cells have copy number 4 of a certain segment and some cells have 3 copies of the same segment with the same break points, it is not known whether they all had 4 copies initially and some cells lost one copy, or if all had 3 copies followed by some cells gaining one copy. To address this, SCEM relies on a predefined set of rules to systematically manage such scenarios. During the analysis of the phylogenies in paper II, we furthermore conducted manual assessments to ensure the biological plausibility of SCEM's handling of each CNA.

Furthermore, scWGS data for copy number inference can only robustly detect CNAs larger than 5 Mbp. It is therefore not possible to determine the exact break point of CNAs across cells. Since the bins are so large it may look as if they have the same breakpoints, but they may differ, making the software infer ancestral similarity between cell populations, which may not true. This is, however, nothing that is an inherent flaw of SCEM, but of the sequencing itself. Nevertheless, SCEM allowed systematic inference of phylogenetic trees based on single cell copy number profiles, thereby providing deepened understanding of the evolutionary dynamics during treatment resistance in paper II and III.

In paper III we introduced the MMP method, which is a novel method for phylogenetic reconstruction. One possible limitation is that it relies on the input clone sizes being correct. If they are, it will, however, produce a phylogeny that does not result in back mutations. This is in many cases biologically sound, especially for point mutations. Currently the MMP method merely allows creation of a single phylogeny based on the input event matrix. In future improvements of the software, the capability to generate multiple possible phylogenetic trees could be incorporated.

In conclusion, yes, we did develop tools for the analysis of the evolutionary history of a tumor. There is a plethora of different bioinformatical tools and selecting the appropriate one is crucial for accurately interpreting the data. Which ones can be chosen does, however, depend on the available data. Moving forward, efforts should focus on developing software that integrates CNAs and SNVs in the phylogenetic reconstruction while also correcting the VAFs based on the CNAs directly.

The importance of multiregional and temporal sampling

Since tumors are genetically heterogeneous, samples should ideally be taken from multiple positions and time points and the bioinformatical tools used should be able to handle this type of data. Multiregional and temporal sampling provides a better assessment of the subclonal landscape of the tumor, thereby increasing the possibility to learn more about tumor evolution, resistance development, and metastatic dissemination.

While the importance of multiregional and temporal sampling, along with subclonal deconvolution, has been recognized for several years, many studies still rely on a single sample per patient. The studies are most likely limited by samples collected many years ago or by clinical routine. Having a single sample will, however, unfortunately not capture the landscape of different subclones that may be present in different parts of the tumor, possibly leading to faulty conclusions. Moreover, neglecting subclonal deconvolution, further impedes the understanding of tumor evolution. In many studies, the genetic alterations identified in a sample are assumed to be clonal when performing the phylogenetic reconstruction, resulting in biopsy trees, not constituting true phylogenies. [455]

While sample collection must prioritize patient well-being and not cause unnecessary harm to the patient, it is essential to recognize the potential benefits of multiregional and temporal sampling for treatment decisions. Basing the entire treatment decision on a single sample, may actually be harmful for the patient. Different parts of the tumor might display different mutations. A mutation, whose protein product can be targeted, could be clonal in one part of the tumor, and completely absent in the other parts. Administering a drug against this target, will thus only cause killing of the cells in the sampled area, while the rest of the tumor continues to grow. Furthermore, sampling several regions allows identification of additional possible treatment targets, possibly allowing a treatment plan incorporating a combination of several targeted therapeutic drugs tailored for that patient's tumor. In addition, different parts of a tumor can display varying histology which could categorize the tumor as more aggressive, than it would have based on a biopsy from another part. This might have a major impact on the initial treatment the patient receives, and possibly also the prognosis.

However, sample collection practices is strictly guided by clinical praxis, limiting the feasibility of multiregional and temporal sampling Analyzing more samples from each tumor may also increase the discomfort of the patient, as well as genomic analysis costs and workload. Nonetheless, it is in the best interest of both the patient and healthcare providers to better characterize each patient's disease as it may have a major impact on the prognosis. Efforts should be directed toward exploring the feasibility of implementing multiregional and temporal sampling.

Did we learn more about the development of chemoresistance?

Paper II aimed at elucidating COJEC treatment resistance in high-risk *MYCN* amplified NB.

A limitation of most studies investigating treatment resistance is the use of a single agent, which does not recapitulate the multidrug regimens patients are administered in the clinic. In paper II, however, all five chemotherapeutic drugs encompassed in the COJEC treatment protocol was used, consequently representing a study setting that more closely resembles the clinical scenario. There are, however, some important differences and limitations to the developed protocol. The drugs were administered intraperitoneally in the mice while it is given intravenously in patients. The drug is, however, quickly absorbed by the blood vessels in the abdominal cavity and should thus in theory quickly reach the desired concentrations in the blood with a not too long delay compared to intravenous administration. Another aspect is that it may be difficult to calibrate the amount of each drug and their relative amounts, so that the effect in the mouse corresponds to that of the patient. This requires meticulous calibration. Furthermore, to aid close monitoring of tumor size, the tumors were placed subcutaneously instead of in the adrenal gland, where the tumors were growing in the corresponding patients. There is a possibility that the concentration of drug reaching the subcutaneous tissue versus the adrenal gland differ, possibly affecting the selection pressure the tumors undergo. In addition, the drugs were not administered in the exact same interval and combinations as in the clinical setting, which possibly could affect tumor evolution. Finally, the mice were immunocompromised, not recapitulating the immune microenvironment in a normal tumor growth setting.

Despite these differences, the model constitutes a significant advancement compared to most studies, which use cell cultures containing similar cancer cells and administer a single drug to model treatment resistance, which is far from the clinical setting.

What roles do blood vessels play in NB treatment resistance?

Histological assessment showed that both PDX1 and PDX2 displayed large and collapsed blood vessels, which has been shown to be correlated to worse prognosis, [473] while PDX3 displayed smaller and open blood vessels.

Larger vessels are theoretically expected to provide the tumor with a high blood flow, facilitating the delivery of oxygen and nutrients. This could promote tumor growth but should also result in a higher concentration of chemotherapeutic agents within the tumor tissue, thereby killing more cancer cells. Tumors with large blood vessels may therefore experience a stricter evolutionary bottleneck, favouring the selection of resistant cells compared to tumors with smaller blood vessels. Some studies have, however, shown that aggressive tumors such as pancreatic tumors,

have few blood vessels and that cancer cells often are subject to a hypoxic environment. [325, 482] Furthermore, metastatic subclones have been shown to originate from hypoxic areas in renal cancer primary tumors. [358]

Provided that the large blood vessels are not malfunctional, which they very well could be, the tumor possibly leverages between two or more environmental conditions: (1) *Large blood vessels*: Resulting in a strict evolutionary bottleneck due to high doses of chemotherapy reaching the tumor cells, selecting for cells with resistance mechanisms. It also provides the resistant cells with oxygen and nutrients, allowing the resistant cells to grow. This could possibly result in CCR which is the evolutionary pattern seen in paper III. (2) *Small blood vessels*: Resulting in hypoxia and nutrient depletion, selecting for slowly proliferating cells with metastatic potential. The cells in this environment may be exposed to low drug concentrations, possibly allowing both sensitive and resistant cells to proliferate, resulting in linear evolution.

Consequently, there should be a heterogeneity in which cells are present after treatment. Some cells survive therapy because they have developed resistance mechanisms, which have been selected for by treatment, while other cells survive because they have not been exposed to high enough drug concentrations. Higher administered drug doses should consequently kill the cells that otherwise would have survived due to not enough drug reaching them. This might explain why the COJEC-HD therapy resulted in a better tumor shrinkage and survival in PDX1 in paper II. More studies are needed to investigate this.

Future investigations could also explore the metastatic capability of subclones located near versus far from blood vessels in tumors harbouring large or small blood vessels. Targeting the larger blood vessels could possibly act as a treatment strategy, but might result in less drug reaching the cells, which could be detrimental. Studies on VEGF-inhibitors have, indeed, so far shown disappointing results. [21] The reason for this might be that the inhibitor induces a hypoxic environment, which as stated above, might result in slow proliferation, avoidance of antitumorigenic drugs and drive metastatic spread instead. Hence, targeting blood vessels may not be the way to go.

Do genetic alterations play a role in NB treatment resistance?

Not so surprisingly, we found no recurrent CNAs that could explain treatment resistance across tumors in neither paper II nor III. These findings were confirmed both with SNP-array and single cell copy number profiling.

We did, however, find small deletions in *MACROD2*, *LSAMP* and *PTPRD*, in the NB PDXs tumors in paper II, but they were not treatment specific, and it was unclear, at that time, whether they were recurrently affected genes in patients or merely PDX specific. Interestingly, *LSAMP* has in a recent study, published after ours, been shown to act as a tumor suppressor in NB. WGS of 35 patients with

high-risk NB showed that 17 % displayed CNAs affecting *LSAMP*, and low expression of the gene was associated with poor overall and event-free survival. Additionally, SNP-array of 16 NB cell lines detected CNAs affecting *LSAMP* in 44 %. Furthermore, knockdown of *LSAMP* in cell lines increased proliferation and overexpression decreased proliferation and viability. [483] *PTPRD* has also been shown to be a tumor suppressor in NB. [477] *MACROD2* has so far not been extensively studied in NB, but it is a known tumor suppressor in other cancer forms. [475]

In study II we limited our investigation to CNAs. In a future study, sequencing to detect SNVs, SVs and rearrangements could be used to specifically identify mutations affecting efflux pumps, metabolism, DNA damage repair mechanisms, evasion of apoptosis, proliferation etc. providing information about which cellular processes cause these cells to be treatment resistant. These mechanisms could potentially be targeted to enhance the effect of the chemotherapeutic treatment given as standard of care. Previous studies have indicated that mutations in the RAS-MAPK pathway and *ALK* mutations are more common after treatment, but not all patients express them, and no mutations have been identified that alone can explain resistance. [206, 484, 485] In paper III we did, however, perform WES and TDS, and did not find any SNVs that could explain treatment resistance.

These findings collectively suggest that while certain genetic alterations may confer advantageous traits to cells in resistance development, CNAs and SNVs are unlikely to fully explain chemotherapy resistance in NB.

Mesenchymal cell states – Drivers of resistance?

Several studies have identified the presence of an ADR and MES like state *in vitro* and a few studies even in patient samples. [220, 478] Using RNAseq in paper II we could, for the first time, confirm the presence of cells being in an ADR and MES cell state in neuroblastoma PDX models. Mice that were cured displayed an ADR-like phenotype, while mice whose tumors progressed during treatment or that developed a relapse, displayed an immature MES-like cell state. Notably, the COJEC treatment appeared to select for cells with a MES-like phenotype, possibly contributing to treatment resistance. Some PDX tumor cells were even in an intermediate state between ADR and MES, which has been proposed previously. [220] A study came out shortly after ours, where the researchers did scRNAseq of 18 biopsies from 15 PDX models. They could confirm the presence of an ADR and MES cell state in neuroblastoma PDXs, further strengthening our findings. They also found cells having a mixture of adrenergic and mesenchymal features. [328] This collectively suggests that the ADR and MES cell states are not binary, but rather represent a spectrum of states and possibly that there is a plasticity where ADR cells can develop into MES cells, and vice versa. Interestingly, studies have found that noradrenergic NB cells, which are similar to ADR cells, become more mesenchymal after single agent treatment with cisplatin or ALK inhibitors *in vitro*.

[486] [487] This suggests that the MES cell state can be induced by antitumoral treatment.

Future studies should focus on determining *what* it is that makes the NB MES cells more resistant to chemotherapy treatment compared to ADR cells. Interestingly, the MES cells seem to be resistant to such a wide range of different chemotherapeutic agents. Possibly the cells are less proliferative, or easily can enter a hibernating state during chemotherapy treatment. It is also not known how the tumor cells behave *during* treatment, when the drug is there. Samples are usually collected when treatment has been halted and the phenotype might thus have changed somewhat. It is possible that the intermediate cells are fully in the MES state when the drug is present but slowly transitions to an ADR state when it is halted. More studies are needed on how plastic these states are and whether this plasticity could be targeted.

Furthermore, since the states seem to be plastic, it should be investigated whether there is some way to hinder the development into a MES cell state. Targeting that mechanism could open up the possibility to administer a drug during chemotherapy treatment to hinder the cancer cells from developing resistance. Another treatment possibility would be to reprogram the MES cells back to ADR cells. This is something that could be further investigated using the tumor organoids established as part of paper II. These organoids retained both the transcriptional MES state as well as its COJEC treatment resistance and thus provide powerful tools to test novel drugs for treatment resistant NB and to study the characteristics of the MES state.

Mesenchymal like cells have been identified across several cancer types have been shown to confer both treatment resistance and metastatic dissemination. [345, 347, 351] [346, 395] It is unclear whether the latter is also true for NB MES cells compared to ADR cells. Performing RNAseq on matched samples from the primary tumor and metastases from the same patient could confirm whether such a relationship exists or not. It also remains to be investigated if a transition to a mesenchymal phenotype is a general feature, not just for cancer cells, but whether normal cells in the surrounding tissue also are steered towards this phenotype by chemotherapy. The MES state might be a general way in which cells can protect themselves to intense stress, which might also explain why it is seen across so many cancer types in response to cytostatic treatment.

There is, however, a debate regarding how to define the mesenchymal phenotype. Consequently, we used six different datasets in paper II, and also constructed new lists based on all these six datasets as well as the RNAseq data from the PDXs. Hopefully, future studies can reach a consensus on how to define the cell states. It is also possible that there is a plethora of *different* MES states depending on the antitumoral treatment the cells have been subjected to, which might explain why there are discrepancies in how different research groups define the state. This has

not been investigated. It would be surprising if such a wide variety of drugs would converge to one phenotype, so probably there is some heterogeneity in the MES-like state. A way to possibly identify such a diversity could be to prepare several Petri dishes with NB cell lines and treat each with a different single antitumoral drug followed by analysis of the transcriptional state after treatment. If there are different transcriptional signatures based on the drugs given, it should be detectable. It could, furthermore, be possible to identify common features across all these signatures, posing attractive targets to hinder the development of the MES state. By administering drugs of different concentrations or take samples at different time points it could also be possible to elucidate the transcriptional trajectory the cell takes to reach the MES cell state. In such an experimental setup, a similar study could be performed using human cells to see how non-malignant cells in the same tissue react on the same treatment. It is also not clear how MES cells present in an untreated tumor, differs from those in a treated one.

In paper II only three PDX models were used, thus representing three patients, possibly not accounting for the great intertumoral heterogeneity between NB patients' tumors. [170] The study was also limited to *MYCN* amplified tumors, since there is a lack of models for non-*MYCN* amplified tumors due to the difficulties to establish the PDX models for these tumors. In future studies it would be valuable to analyze a wider range of NB tumors. A further limitation is the fact that the mice are immunodeficient, which does not recapitulate the clinical setting where the immune microenvironment might shape the evolution of the tumor. [166] On the other hand, we did, show that the MES phenotype could appear in an immunodeficient setting. Hence, its appearance might be independent on immune editing. There are also some studies indicating that the ADR and MES states are present in patients' tumors, [478] where the immune system does play a role. Thus, the ADR and MES cell states seem to be able to develop both in the presence and in the absence of an immune system, which is interesting. It remains to be investigated whether there are any differences in the ADR and MES states identified in the presence versus absence of immune cells.

To conclude, cells with a MES-like phenotype seem to play an essential role in conferring treatment resistance in high-risk NB PDX models.

An evolutionary path to resistance in neuroblastoma – a general feature across cancers?

In paper III, we observed two contrasting evolutionary pathways as a response to chemotherapy treatment, depending on whether the tumor responded significantly to treatment or if it progressed.

A tumor responding significantly to treatment, indicated by shrinking, suggests that the major subclone or subclones in the tumor are sensitive to the administered treatment. If there are no chemotherapy resistant cells in the tumor, the tumor is

expected to continue shrinking until it is eradicated, leading to patient cure. However, if chemotherapy resistant cells exist, they will be selected for, resulting in an initial tumor shrinkage followed by regrowth as resistant cells become dominant – a pattern termed collateral clonal replacement (CCR), as observed in paper III for patients having an initial significant tumor response to treatment.

If the subclones constituting the bulk of the tumor are treatment resistant, or quickly switches from a sensitive to a resistant phenotype, significant tumor shrinkage does not occur compared to responsive tumors. The subclones making up a resistant tumor after treatment will be the same subclones as the ones in the primary tumor before treatment along with linear descendants to them. This is also what was seen in NB tumors having a poor treatment response in paper III.

CCR is consistent with selection of pre-existing resistant clones, the presence of which has been confirmed in several studies. [322, 480] Interestingly, the *in vivo* and *in vitro* analyses in paper III showed that the CCR pattern does not appear due to merely a mechanical bottleneck. Consequently, it must be due to a selection by the chemotherapy treatment for cells with certain characteristics, most likely treatment resistance. When creating a mechanical bottleneck *in vitro*, the proportions between sensitive and resistant cell populations will typically not change. The regrown population after treatment should thus be of similar subclonal composition as it was before treatment. However, a mechanical bottleneck *in vivo* induced by debulking, such as partial tumor removal, may disrupt the relative prevalence of each subclone, thereby altering the evolutionary trajectory.

In contrast, administration of chemotherapy will selectively kill sensitive cells to a higher degree than resistant cells, disturbing the balance between sensitive and resistant cells. This disturbance allows resistant cells to proliferate, unimpeded by the inhibitory effects of sensitive cells. This might explain why the mechanical bottleneck does not give the same evolutionary pattern as a chemotherapy induced reduction of cells. This also implies that the relationship and interaction between the cells is important for the development of a treatment resistant tumor.

Further investigation is needed to understand some tumors seem to be inherently resistant to treatment i.e. do not significantly respond to treatment, despite it being the first time the cells are exposed to the treatment. Furthermore, analysing RNAseq data before and after therapy, could be performed to assess whether this inherent resistance appears because of MES cells already being present in the primary tumor before treatment. It is not known whether cells in the MES state are present already in the untreated primary tumor. Using a NB cell line with ADR and MES like cells, we could however confirm in paper III that the treatment selected for MES-like cells *in vitro*. In paper II we could also confirm that treatment selected for MES-like cells *in vivo*. Whether this selection for MES-like cells is seen in patients as well remains to be investigated. Possibly the presence of such cells could be used as a resistance marker already at diagnosis.

In paper III, scWGS copy number analysis underscored extensive clonal branching as a common feature in NB. Consequently, there is a plethora of subclones present in the tumor already at diagnosis, many of which are not even detectable using bulk DNA methods. Moreover, in paper IV, we found that a combination of both linear and branching evolution, was common across WT subtypes. This confirms that intratumoral heterogeneity of CNAs is common in WT as well, in line with previous studies. [23, 170] Furthermore, we found that DAWT, which is often treatment resistant, had more complex phylogenies with increased PSR, divergence, irregularity and number of CNAs compared to other WT subtypes. This substantial subclonal variation observed within primary tumors, observed both in high-risk NB and DAWT, seems to set the stage for natural selection of resistant subclones during chemotherapy treatment.

In conclusion, we identified two distinct evolutionary paths to treatment resistance, denoted CCR and linear evolution. Both pathways imply selection of pre-existing resistant subclones, which is in line with findings from various studies in other cancer forms. [322, 480] Notably, the presence of extensive intratumoral heterogeneity appears to lay the groundwork for this selection process. In NB, the resistant subclones are likely derived from NB tumor cells in a MES like cell state, present in the primary tumor already before treatment initiation. Importantly, the potential universality of CCR and linear evolution as evolutionary routes across different cancer types remains unexplored.

Monoclonal tumor compartments – cribs for the development of resistant subclones?

In paper III, we found that in untreated NB tumors, the subclonal landscape was dense and the subclones were intermixed, lacking clear anatomical barriers between them. However, post-treatment there was instead small islands of surviving subclones, separated by necrosis, haemorrhage, and reactive changes. Each compartment often harboured a single subclone type. This indicates a strong selection for specific subclones capable of withstanding the administered therapy and cramped conditions of the small surviving islands of cells. This is highly clinically relevant. When feasible, distinct tumor areas should be sampled for genomic analyses to better characterize the genomic landscape of the patient's tumor. This finding aligns with our observations in WT in paper IV, where the tumor cells after treatment often are confined to compartments comprising homogenous subclone populations. However, it is possible that smaller subclones may exist within these compartments, necessitating scWGS for their detection.

The observed compartmentalization may represent a defensive response to the intense treatment, enabling cells to shield themselves and evade the effects of chemotherapy. Treatment induced selection pressure favours specific subclones, fostering the development of a homogeneous population within each compartment. By being packed closely together, the cells within these compartments can

increase the intratumoral fluid pressure in the region, [366] creating a pressure gradient directed outwards, thereby hindering chemotherapeutic agents from reaching the interior of the compartment. An increased intratumoral fluid pressure in the tumor is associated with a lower recurrence-free survival in lung cancer, can promote metastasis and may act as a prognostic biomarker. [366, 488] It remains to be investigated whether this is also true in NB and WT.

To conclude, tumor cell compartmentalization and an associated increased intratumoral fluid pressure seems to be a common feature across cancers and might contribute to development of resistance. [366]

How are new tumor compartments born?

In the previous section, we highlighted the potential role of tumor cell compartmentalization in conferring treatment resistance. This underscores the importance of unravelling the mechanisms behind their formation. In paper IV, we consequently delved into the process of tumor compartment genesis. We found that in WT, the emergence typically involved the budding of a single subclone, which migrated from the original compartment and underwent a clonal expansion in a new location.

Hence, monoclonal spread followed by a clonal sweep is the most common way for a new tumor compartment to be formed in WT. Further investigations are needed to determine the factors that differentiate budding compartments from those that do not bud. If cells migrate out of the tumor compartment due to a high pressure within, budding subclones may potentially harbour additional genetic alterations, presenting a promising avenue for exploration in future studies.

Does *MYCN* play a role in treatment resistance?

Phylogenetic analysis conducted on both SNP-array and scWGS data confirmed that *MYCN* amplification occurs early and can even precede 17q gains in NB, a result also confirmed using FISH analysis. Intriguingly, spatiotemporal heterogeneity was observed in the *MYCN* breakpoints in all 6 patients with *MYCN*-amplification analysed. In each case there was at least one shared breakpoint, indicating that there is an initial *MYCN* amplification event, followed by additional structural changes of the amplicon. Notably, these structural changes were more prevalent after treatment compared to before. However, it remains uncertain whether the increased number of breakpoints simply increases as a function of time. Moreover, it is also unclear whether the different amplicon variants are selected for, or whether they are neutral alterations, and increases in prevalence due to selection of other subclonal features.

Nevertheless, it is clear both from this study and from many others, that *MYCN* plays an important role in NB pathogenesis. [200] *MYCN* amplification is

consistently associated with poorer prognosis and treatment resistance, but it is unclear whether this correlation also implies a causation.

What role does *TP53* have in the development of anaplasia and resistance?

Interestingly, *TP53* mutations were detected in all areas displaying classical anaplasia in the DAWTs in paper IV. The phylogenetic trees indicated that *TP53* mutations were followed by extensive branching and saltatory evolution, and sometimes even additional mutations of *TP53*. The initial *TP53* mutations also occurred in parallel across several anatomical compartments in the same tumor, resulting in multifocal anaplasia.

It remains to be investigated *why TP53* mutations appear so often, or rather, why *TP53* mutations are so strongly selected for in some WTs. The fact that some tumors do not display any detectable *TP53* mutations, while others have several parallel *TP53* mutations across the tumor area, suggests that there is something in the microenvironment in these tumors that either selects for or against the mutations.

Future studies could investigate whether it is dense cellular growth conditions in tumor compartments that selects for *TP53* mutated cells. Both stromal WT, blastemal WT and DAWT display distinct tumor compartments separated by fibrous septae, as does many other cancers. Possibly, the dense conditions in these compartments cause nuclear squeezing, inducing DNA damage and adventuring the integrity of the nuclear envelope. [365] This might result in an upregulation of the transcription of the *TP53* gene. Studies have shown that increased transcription of a gene also results in an increased number of mutations in that gene, due to the repeated quantum measurements. [75-77] The *TP53* mutations might allow the cell to survive despite accumulating additional genetic alterations, while cells not having it are successively depleted.

Using RNAseq or IHC it should be possible to investigate whether an upregulation of *TP53* precedes a mutation in the gene, and whether the mutation in turn precedes anaplasia, further unravelling the path to anaplastic WT. In that case there should be: 1) Small subclones having an upregulation of *TP53*, but no mutation in it. 2) Small subclones having an upregulation and mutation of *TP53*, with or without anaplastic features. 3) Larger subclones having an upregulation and mutation of *TP53* in combination with displaying anaplastic features.

We also observed that *TP53* mutations were followed by extensive phylogenetic branching. It is, however, not evident that the *TP53* mutations *cause* this branching. It could be that the mutation *allows* the cell to survive despite acquiring additional CNAs. A cell that would otherwise have died. By combining the sequencing information in this study with scWGS copy number profiling of the same sample, this could be investigated. Perhaps compartments not displaying *TP53* mutations also contain small subclones with many CNAs that remain minor

or undergo apoptosis due to having a functional *TP53*. I therefore propose that it could be that tight cellular growth conditions cause nuclear squeezing and CNAs to accumulate resulting in the following two scenarios: 1) If the cell already has or acquire a *TP53* mutation, the subclone can expand. 2) In the absence of a *TP53* mutation, the subclone remains small, or undergo apoptosis due to the accumulation of CNAs.

Nuclear squeezing might also occur when the cancer cells are budded out from a compartment, resulting in additional genetic alterations, possibly also promoting CIN. If the cell also has a *TP53*-mutation it can still survive and may form a new compartment. The high pressure within the compartments might also make it difficult for drugs to reach the cells, since the pressure gradient is directed away from the compartment center. Studies have shown that an increased interstitial fluid pressure promotes proliferation and invasion in oral squamous carcinoma. [489]

Using RNAseq it would be possible to compare the phenotype of the WT cells having *TP53* mutations, but where the cells are not anaplastic, to anaplastic areas. Our group showed in a subsequent study after paper IV, that these cells, harbouring *TP53* mutations but not displaying anaplastic features histologically, might resemble preanaplastic cell populations. This indicates that there might be a gradual transition towards anaplasia. The anaplastic cells have a high proliferative index, dividing more rapidly than they are dying, possibly explaining part of the chemoresistance seen in DAWT. [255]

There is also a lack of studies investigating which subclones metastasizes in WT. It is not known whether it is the anaplastic areas, or whether they merely confer treatment resistance. It also remains to be investigated whether the anaplastic state could be reversible, which would be an intriguing treatment strategy considering the poor prognosis of DAWT.

In conclusion, *TP53* seems to be important for anaplasia formation, but its role in chemotherapy resistance is still unclear. In an ongoing study in the group, we are searching for variants in other genes that might be involved in anaplasia and resistance development in DAWT.

Did we gain new knowledge about metastatic spread across pediatric tumors?

In paper V, we elucidated the metastatic trajectories in patients with NB, WT or GCT. We could show that metastases can arise early, as well as several times during tumor evolution. A metastasis was defined as having arisen early if the metastasizing subclone or subclones were ancestral to the MRCA in the primary tumor. Whether the spread really is *chronologically* early relative to tumor initiation, is, however, hard to know. The patient's tumor might have been growing at the site for a long time before detection. The metastasis could hence be

late relative to tumor initiation, but early relative to the phylogenetic definition. This could possibly be investigated by utilizing clock like mutations to time the genetic alterations.

Early metastatic spread has important clinical implications since it makes it difficult to cure the patient, consequently resulting in a worse prognosis for the patient. It is therefore essential to learn more about why some tumors metastasize early, some late and why some tumors do not metastasize at all. It remains to be investigated what characterizes subclones spreading early from those spreading late and how they both differ from the primary tumor. There might be certain conditions in the primary tumor, that selects for cells to migrate. Since the subclones metastasizing early, by definition, cannot be detected in the primary tumor anymore, it does, however, make it difficult to perform such an investigation.

As described in the section on metastatic spread, the cancer cell must both make it into the circulation, extravasate and then colonize the distant site to form a metastasis. It is possible that patients not presenting with early metastatic spread, also have circulating tumor cells, but that these subclones have not yet acquired the ability to *colonize* distant sites. In a future study, ctDNA or CTCs could be compared between patients having distant metastases, with those without, to identify whether there are some transcriptional or genomic markers that might increase the risk that the CTCs colonize.

In paper V we also found that several different subclones within the primary tumor can possess metastatic capability. It is, however, not known whether these tumors gained metastatic capability early, such that all or most of the descendent subclones maintain this characteristic, or whether the metastatic capability arose via parallel evolution within the tumor. Using RNAseq it would be possible to investigate the similarities and differences between the distinct subclones having metastatic capability, as well as to compare this to other subclones in the primary tumor. In metastases that has arisen early, the subclone is, by definition, no longer detectable in the primary tumor. It would be interesting to investigating whether the subclones now making up the bulk of the tumor, and which are descendent to the metastasizing subclone, still possess metastatic capability.

Strikingly, we also identified intermetastatic spread to be a common feature across pediatric cancer types. Besides one recent study showing it occurring in a small cohort of NB patients, [414] this has not been shown before in pediatric tumors. We here confirm that this is true not only for NB, but also for WTs and GCTs. We also find instances where it happens in several steps i.e., that a metastasis gives rise to a new metastasis, that in turn gives rise to a new metastasis, further enlightening the complexity of metastatic disease. This stresses the need for novel treatment strategies taking this process into account, since it implies that metastases can arise despite the primary tumor being removed.

Unsurprisingly, we could not identify any genetic alterations or point mutations that alone could explain metastatic spread across patients. [104] The cause of metastatic dissemination seems to be at a transcriptional level. In a future study, RNAseq or IHC, hold the possibility to unravel more about what makes the metastasizing subclones different from other subclones in the primary tumor. By using the phylogenetic trees from paper V, it is possible to deduce from which histological area each metastasis arose in the primary tumor. It would be possible to deduce whether there are phenotypic differences between the metastatic subclones in the primary tumor, the metastasis and non-metastatic subclones in the primary tumor, by either doing RNAseq on the different areas, or performing immunohistochemistry staining of the histological sections.

The RNAseq data additionally allows for the identification of whether there are any transcriptional cell states responsible for metastatic spread. Studies have indicated that metastasizing cells undergo EMT, suggesting that the metastasizing subclones should be in a state resembling a mesenchymal cell, possibly similar to the mesenchymal cells identified as key players in treatment resistance. Mesenchymal cells have been shown to have increased metastatic capability in several adult cancers, but it has not yet been shown in pediatric cancers. [347, 351] RNAseq also holds the possibility to identify whether subclones responsible for intermetastatic spread are different from the other subclones that does not spread again.

It is also possible that the metastatic cells undergo MET in the distant site. If they do not, or if they undergo EMT again, this could possibly explain intermetastatic spread. This could also be explored using RNAseq or IHC. Cells responsible for intermetastatic spread are of particular interest. By analysing the surfaceome of these cells, potential surface markers could be identified that possibly could be targeted, thereby hindering intermetastatic spread.

How can metastatic spread be hindered, and metastases eradicated?

One of the major challenges in cancer is how to cure patients with metastatic disease. Parallels can be drawn to attempts to exterminate a species that is both heterogeneous and geographically dispersed, without harming other species in the same environment. This is extremely difficult and requires identifying something targetable that unifies all tumor cells at all metastatic sites as well as the primary tumor from all normal cells. Since the tumor arises from the normal cells, this is very difficult. It is further complicated by the vast intratumoral genetic and phenotypical heterogeneity in combination with the continuing subclonal evolution. So far, no identified single mutation or other genetic alteration has been shown to cause metastatic spread across patients. In Paper V, we also found that there was a significant genetic heterogeneity between the primary tumor and the metastases as well as between metastases in the same patient. Hence, either a personalized combination of different targeted therapies should be given to each

patient, requiring sampling of all metastatic sites, since they can harbour different mutations, or novel ways are needed to tackle metastatic disease.

Immunotherapy has proved useful for several cancers but suffers from similar problems as targeted therapies do, namely that of only focusing on a single target. Efforts are, however, being conducted to create multi-targeted CAR-T therapies, holding great promise for future cancer patients. [490] Surprisingly, there is a particular lack of studies investigating tumor cell negative locoregional lymph nodes, especially, what makes the immune cells in these lymph nodes different from the ones in cancer cell positive lymph nodes. It might be that cancer cells simply haven't reached the lymph nodes, but it may also be because the immune microenvironment in these lymph nodes does not allow primary tumor cells to colonize them. In a future study, one could compare the surfaceome and RNA profile of immune cells in the negative to those in positive lymph nodes. This could hold the possibility to identify potential novel targets for immunotherapy. It could also be possible to culture the immune cells found in these negative lymph nodes. Possibly they could be expanded and reinjected into the patient repeatedly as a treatment, as a form of autologous immunotherapy. These immune cells should, however, already be able to spread physiologically from the lymph nodes but maybe they are too few in numbers to eradicate the tumor. More studies are needed to investigate these possibilities.

Another approach to tackle metastatic spread could be to hinder metastases from forming in the first place. Efforts have been done to learn more about what makes cancer cells undergo EMT, how this process works and how cancer cells start migrating towards the blood vessels, with special focus on finding treatment targets. It is, however, difficult to target EMT since it is a normal cellular process. Cells intravasating and extravasating are also normal, at least for immune cells, so it may be hard to specifically target these processes. Especially since it seems as if the tumor cells use similar methods as immune cells do.

Targeting circulating tumor cell clusters is an intriguing thought to hinder metastatic extravasation. These clusters have been shown to be much more prone to colonization compared to single circulating tumor cells. Hence, splitting up these clusters holds the possibility to make it harder for the cells to extravasate and colonize. [491] Furthermore, a device has been developed that can kill melanoma tumor cell clusters with lasers, denoted the Cytophone. It scans the blood for melanoma cells using laser and sound waves. It does not require any needles or blood to be drawn and it can scan the entire blood volume in a couple of hours. The device could possibly also be used for early detection. [379] Possibly similar devices could be created for other tumor cell types in the future.

The colonization process itself could be targeted. Very few of the circulating tumor cells manage to colonize and grow at the new site. It would be interesting to investigate the clusters that do colonize and compare them to those that do not. No

study has been performed on pediatric cancer for CTC cluster detection and analysis. In a future study, blood could be collected at diagnosis, at some selected time points during treatment and at follow ups, when blood samples are collected anyway, thus not posing any extra harm to the child. There are many techniques available for CTC detection. The collection of samples from many patients allows the identification of similarities and differences in the amount and proportions of single and clustered tumor cells. The method could be used for early detection of relapses. One could also identify which molecules hold the CTCs together and do tests *in vitro* to try to separate them. Possibly, this could result in the development of a drug that could be administered before, during and after chemotherapy treatment to avoid colonization of the CTC clusters present in the blood. One problem is, however, that most clusters to not circulate especially long in the blood, making it difficult to time the CTC cluster spread with treatment. If the drug is not harmful for the patient, i.e., have few if any side effects, one possibility would be to take the drug daily during the entire treatment course.

There are currently great efforts being done to include knowledge based on cancer evolutionary theories into treatment protocols. Adaptive therapy is, however, difficult in the metastatic setting since the primary tumor, and the metastases, may behave differently and may have varying proportions of sensitive and resistant cells.

Can leukemia metastasize? There is a widely held belief that hematological cancers may be less geographically heterogeneous than carcinomas, but this has not been systematically proved. Future studies should consider how two bone marrow samples from respective crista looks. Leukemias can also colonize the mediastinum and the brain. No study has investigated how these subclones are different from the ones found in the bone marrow. Leukemic cells are also very motile, compared to metastatic solid tumors, most likely due to the motile nature of the immune cells to begin with, but what makes immature cancer cells moving out in the blood, while immature normal cells in the bone marrow does not is unclear.

The ultimate treatment strategy

In the best of worlds, the ultimate cancer treatment strategy would be a tailored treatment plan based on the specific genetic and phenotypic makeup of each patient's tumor. For this to be possible, it is essential to, when feasible, obtain several biopsies from the primary tumor and from metastases to assess the subclonal landscape of the disease.

1. **Personalized:** There are today a countless number of different targeted therapeutic drugs. Based on the genetic makeup of the tumor and knowledge about which subclones are present in the primary tumor, and

possibly also at metastatic sites, it should, theoretically, be possible to administer a personalized combination of these drugs based on that person's tumor's subclonal composition.

2. **Orthogonal:** Administered treatments should be orthogonal to each other i.e., they should be designed to work in different ways, making it unlikely that the tumor is resistant to all. Examples of treatments that are orthogonal with respect to one another are targeted therapies, chemotherapy, immunotherapy, surgery, and radiation therapy.
3. **Evolutionary steering and extinction therapy:** Strategic drug combinations could potentially steer cancer cell evolution in a desired direction, facilitating susceptibility to other treatments. Furthermore, when the tumor is small, the therapy should also be switched to exploit the vulnerabilities of a small population. In this way, the tumor could be pushed into an extinction vortex, from which there is no return.
4. **Avoiding administering maximum tolerable dose:** If possible, it is also important to not over-treat the patient. This is especially important for pediatric cancer patients, since they may suffer from a vast range of long-term effects affecting their entire adult life. By not administering maximum tolerable dose (MTD) for chemotherapeutic drugs it is possible to minimize side effects and long-term effect. A lower dose of chemotherapeutic drugs will also pose less stress on the cancer cells. Less stress might result in a lower selection pressure for the selection of pre-existing resistant subclones, compared to administering MTD. In addition, the pre-existing resistant cell population is kept in check by keeping a steady amount of sensitive cancer cells. Administering lower doses of chemotherapeutic drugs can also lower the risk of cells utilizing survival strategies such as going into a hibernating or dormant state, possibly decreasing the risk of a later relapse.
5. **Agile:** Ideally, the treatment should be agile or adaptive in the sense that it should be possible to switch between treatments based on real-time monitoring of the tumor and in response to changes in the tumor's molecular landscape. In the future this might be possible to do, by analysing ctDNA in the blood.
6. **Prophylactic drugs against resistance:** Drugs that hinder development of resistance or usage of resistance mechanisms. E.g., drugs that hinder phenotype switching into a resistant phenotype or drugs that reprogram cells with a resistant phenotype into a sensitive phenotype.
7. **Prophylactic drugs against metastasis:** Drugs that hamper colonization of circulating tumor cells. E.g., drugs that split CTC clusters in the blood.

Overarching limitations of the studies

Both NB, WT, RMS, and GCT are rare diagnoses and have around 10 cases per year each in Sweden. It hence takes long time to accumulate sufficient material to conduct a study and those conducted are limited by the cohort size. In addition, just as in adult tumors, pediatric tumors are both intratumorally and intertumorally genetically and phenotypically heterogeneous, making it even hard to draw general conclusions.

These limitations in the number of patients and amount of tumor material, requires smart design of the projects such as to thoroughly use the material available in the archives. Most of these samples are, however, FFPE material. In addition, some samples may be up to 20 years old. The DNA is often fragmented and scarce. This significantly limits the types of genomic analyses that can be performed, as discussed in the methods section. Pediatric tumors mainly harbour larger chromosomal aberrations. Consequently SNP-array has been used throughout our studies with the addition of TDS and RNAseq when feasible. Bulk DNA methods require subclonal deconvolution, which brings various problems with it, as discussed previously. In study II and III we additionally used single cell whole genome sequencing, in which case no subclonal deconvolution is needed. We could see that many of the subclones deduced through subclonal deconvolution of bulk DNA data, was in fact present when analysing single cells, providing strength to the subclonal deconvolution performed.

Since patients with childhood cancer are few compared to adult cancer types, autopsy studies hold the possibility to provide additional information about the metastatic process in pediatric cancers.

Conclusions

To summarize, the following broad conclusions could be drawn from the work presented in this thesis.

Paper I

- We developed DEVOLUTION, a software which provides a standardized framework for subclonal deconvolution and phylogenetic analysis of multiregional data, aiding tree-to-tree comparison during downstream analysis. It additionally allows phylogenetic reconstruction using both CNAs and SNVs.
- The software was benchmarked against 56 pediatric tumors, simulated data as well as an external data set while comparing it to similar tools.
- DEVOLUTION could be used to dynamically track subclones in time and space. This makes it possible to identify which subclones survive therapy, which ones make up relapses, which ones metastasize and when, as well as to identify how different metastases are related to one another.

Paper II

- We developed a COJEC-like treatment protocol which was applied to three *MYCN*-amplified NB PDX models.
- There is an extensive intra- and intertumoral genetic and transcriptional heterogeneity in NB tumors.
- Mice displaying upfront progression during treatment or relapse had an increased clonal diversity, higher copy number and transcriptional levels of *MYCN*, and displayed an immature MES-like cell state.
- Mice cured after COJEC, and subsequent surgery displayed an ADR-like cell state and had a lower number of copies of *MYCN* and transcriptional levels.
- COJEC treatment seems to select for cells with a MES-like phenotype, and these cells might contribute to treatment resistance. A predominant ADR signature after treatment were on the other hand correlated to a good overall prognosis.
- Some NB cells displayed both ADR- and MES-like markers, indicating that they are in an intermediate state, as previously proposed.
- 3D organoids were created from control and COJEC treated tumors.
- Organoids from relapsed tumors retained their chemoresistance and transcriptional features *ex vivo*.
- Collectively, these results indicate that NB chemoresistance is primarily mediated by transcriptional changes.

Paper III

- NB tumors which significantly responded to treatment, showed a pattern of CCR where the subclones present at diagnosis were depleted by the chemotherapy, selecting for other subclones in the tumor, eventually regrowing and resulting in a treatment resistant relapse.
- NB tumors showing poor treatment response had subclones after chemotherapy that were linear descendants to the ones present before treatment.
- These two evolutionary patterns were confirmed both *in vivo* in PDX models and *in vitro* in NB cell cultures.
- CCR suggests a chemotherapy induced selection of pre-existing, possibly resistant, NB cell populations.
- ScWGS analysis revealed extensive phylogenetic branching to be a consistent feature of NB both in patients, *in vivo* and *in vitro*, emerging already at the ancestral states, forming the substrate for CCR.
- The CCR and linear evolution pattern has important implications of how to sample these tumors for genomic analyses in the clinic. During upfront progression, new genetic alterations are merely added to the pre-existing ones while tumor shrinkage followed by a relapse implies a radical shift in the subclonal landscape, necessitating resampling of the tumor. The spatial analysis revealed the importance of sampling distinct anatomical compartments to characterize the genomic landscape.

Paper IV

- DAWT display more complex evolutionary histories compared to other WT subtypes. Extensive phylogenetic tree complexity seems to be a hallmark of DAWT.
- *TP53*-mutations seems to be the initiating event for this complexity.
- All WT areas displaying classic anaplasia displayed *TP53*-mutations, which often occurred in parallel, indicating that they are late events.
- WTs are intratumorally heterogeneous with respect to CNAs.
- Compartments of cancer cells within the same tumor mainly are homogenous.
- The birth process of new compartments in WT involves, in most cases, the budding of a single subclone that subsequently clonally sweeps forming a new macroscopic compartment.
- Cross-compartmental sampling of WT is needed to characterize the genetic landscape of the patient's disease, which is of considerable interest for targeted therapy.

Paper V

- Our results provide further knowledge of the evolutionary dynamics of metastatic spread in pediatric cancer patients.
- Metastases often have a different genomic profile than the primary tumor, stressing the need to sample metastases for assessment of therapeutic targets, since metastases may completely lack targets present in the primary, resulting in progression at distant sites despite the primary tumor responding to treatment.
- Metastatic spread can occur early.
- Metastatic spread can occur several times during tumor evolution.
- Intermetastatic spread is a common feature in pediatric tumors, indicating that a single metastasis can act as a hub for further metastatic spread.

Acknowledgements

As many have said before me, there is no “I” in science. Therefore, I would like to thank the truly remarkable people I have had the privilege to work with during these past years.

Firstly, a big thank you to my main supervisor **David Gisselsson**. I could not imagine a better supervisor than you. You are inspiring, encouraging and a great role model. Your energy, enthusiasm, and interest in research spread to everyone around you. Furthermore, thank you for letting me get the freedom to explore things on my own while always being available to offer guidance if needed. I am immensely grateful for the opportunity to work with you over the past 6 years, throughout my bachelor thesis, master thesis, and now the PhD. You are a truly amazing researcher and appreciated group leader. Thank you for giving me the opportunity to be part of your group. Without you I would not have been even close to where I am today. Anyone having you as a supervisor is lucky!

I would also like to thank my co-supervisors. **Subhayan Chattopadhyay**, thank you for your expertise in mathematics and bash, and taking time to answer all my confused questions on slack. Furthermore, thank you for taking time to give input to this chonky thesis. You are a very kind and fun person and an excellent researcher. I’m sure you have a very bright future ahead of you! You are creative and innovative, which is important in research, and you always come up with interesting ideas and new angles on things. You also have a talent for meme creation, so that could always be a backup plan. I also hope I will meet your cat someday! **Anders Valind**, thank you for giving input and having interesting and well thought questions on the manuscripts. I’m sure you will be a great pediatric oncologist!

I also want to thank the current and previous members of our research group.

Jenny Karlsson: Thank you for the support during all these years. You are a very kind and caring person, that it is fun to be around. You always have good feedback, clever input, and well thought through ideas. I especially thank you for the support during the metastasis project and for taking time to go give input on this rather hefty thesis. I also hope *you-know-who* is happy soon. Thank you for all the work you have put in.

Caroline Jansson: You have an immense and truly invaluable lab expertise. Thank you for all the support and effective lab work, especially during the metastasis project. You are easy to work with and always have a smile on your face, which makes it fun to be around you. I hope we will continue to work together for many years. In addition, you are almost from Alvesta, so that is a big plus as well.

Michele Ferro: Thank you for all the help with the endless DNA extractions and dilutions, and for keeping track of all the boxes with different numbers and my name on for the metastasis project. You are a very kind, caring and funny person. I really hope you are having a good time doing research in Chicago, and that we meet soon again.

Linda Holmquist Mengelbier: You are a great researcher and always have thoughtful input and questions. You are also a very kind and fun person who spreads joy around you. In addition, thank you for supporting my descent on skis in the Rocky Mountains. If it weren't for you, I'd probably still be up there, and this thesis would never have been produced. So, I guess we should all thank you!

Bahar Rastegar: You are an amazing and warm person with an energy and positivity that spreads to the people around you. It was a pleasure to work with you on the DAWT paper. I hope we will continue to collaborate in future projects. Also, thank you for organizing the yearly research group barbeque evenings! In addition, congratulations to becoming a mother!

Geoffroy Durand: You are a kind and funny person! It was a pleasure having you in the group and working with you! I wish you all luck on your new job (or maybe it is not as new any longer) and being a father!

Kaname Uno: You are kind, friendly, a good fisherman and a dedicated researcher. I'm sure you are also a great doctor. Thank you for teaching me about life in Japan and that there are edible fish in the waters of Malmö. I really hope we will meet this autumn when you begin your post doc in Lund, and possibly also collaborate in the future.

Hiroaki Yasui: Thank you for teaching me about life in Japan. It was nice working with you for paper II and III. I always enjoyed your questions on phylogenetics. I hope you are having a good time in Japan and that we meet again.

I also want to thank my very dear friends, **Hanna Thorsson, Valeria Difilippo, Louise Ahlgren** and **Noelia Puente Moncada**. Thank you for the amazing parties, travels, laughs, experiences, tea times, concerts, food, drinks, gym sessions, saunas, after works, and discussions. I can't imagine how these years would have been without you. You are the best! #TheSpiritNuts

Hanna Thorsson: Thank you for being so welcoming when I first began as a student at the department, and for encouraging me to do a PhD in the first place. You are a very warm person who always makes sure everyone is included and are feeling well. You are also a very dedicated researcher, and I'm certain you will have great career whatever path you choose after your PhD. Also, thank you for putting up with sharing office with me, despite me unknowingly "ignoring you" on a daily basis when you talk due to my noise cancelling headphones. Thank you.

Valeria Difilippo: We have had so many fun moments during these past years. You are a kind, fun and warm person. You also make really amazing food. I appreciate that. I also have to admit that your balcony actually is better than mine and that it is unreasonable for me to not have olive oil in my kitchen. It has also been great that we have had each other the last couple of months, both being in the process to defend. We could sit and write at LUX and have panic and cappuccino together. I liked that. There really should be a panic room at the office. When this is over, I'm sure you will continue to have an amazing future as a bioinformatician!

Louise Ahlgren: We first met, many years ago, when you were my fadder in medical school. I was so happy when I discovered that you were working here at the department as well! You are a very kind and super funny gal who brings joy to people around you. Furthermore, thank you for teaming up with Hanna to encourage me to pursue a PhD. I wish you the best of luck with your AT (congratulations again!) and the remaining part of your PhD. I have no doubt you will have a brilliant future as a medical doctor!

Noelia Puente Moncada: Thank you for all these amazing years. Also, thank you for the easter coffee cup. I use it every day. You are an extremely kind and fun person. You always have something interesting to tell and care about the people around you. I hope you will move back here in Sweden soon! I miss you a lot!

I also want to express my gratitude to everyone at the department that makes it as an amazing workplace as it is! Thank you for all the interesting, engaging and funny discussions during lunch, the after works, parties and of course also the scientific discussions. I wish you all the best! **Saskia, Jakob, Hannah, Carl S, Karim, Mattias, Ton, Valeriia, Vendela, Rasmus, Larissa, Josephine, Efe, Gladys, Ram, Somadri, Varsha, Maria, Giulia, Andrea, Carro, Marianne, Linda, Jenny, Marianne, Henrik, Tina, Helena, Axel, Rebeqa, Nils, Katrin, Pablo, Charlotte, Catharina, Philip, Minjun, Sofia, Elenor, Helena Å, Jan Köster, Kristina Lundin, Linda, Christina O Å, Elenor, Niklas, Ludvig, Petr** and the students **Jakob and Carl.**

Karim Saba: It was great sharing office with you together with Hanna, Louise, and Valeria. Good old times! It was also really fun at the ESHG in Glasgow together with Valeria and Jakob and the subsequent trip to Edinburgh all together. Furthermore, it has been a pleasure to plan parties together with you together with

the other party pixies. Furthermore, thank you for generating the raw counts for the RNA analyses in the knockout manuscript. I'm sure you have an amazing future as a bioinformatician. Last, but absolutely not least, thank you for letting me meet your cats and giving me the honour to get one of your cat calendars.

Mattias Pilheden: With you at the lunch table, there are always super interesting discussions. Continue with that. Also, thank you for the collaboration in the party pixies.

Saskia R. Sydow: Thank you for a really great time during CARES. You are kind and fun to be around. I really hope we will take a rowing session sometime soon (when the weather allows) on the canal in Malmö!

Jakob Hofvander: Thank you for a great time at ESHG in Glasgow and the detour to Edinburgh together with Valeria and Karim. It was super fun. I wish you good luck with your future research career and being a father!

Hannah Åbacka: We had a great time at CARES together. Congratulations again to your PhD. I wish you all luck in the future!

Andrea, Carro, Marianne, Linda & Jenny/Jenny & Linda: Thank you for creating a great atmosphere in the lunchroom!

I especially want thank the PIs for creating this work environment and taking the very good decision to hire all these amazing people stated above: **David Gisselsson, Karolin Hansén Nord, Thoas Fioretos, Anna Hagström, Bertil Johansson, Marcus Järås, Nils Mandahl, Felix Mittelman, Kajsa Paulsson** and **Fredrik Mertens**.

Anette Welin: Thank you for keeping the department afloat.

Gustav Christensson: You are a very kind and clever person. I'm sure you have a bright future ahead of you, both in research as well as a medical doctor.

Fredrik Mertens: Thank you for offering to lend me your PhD hat.

Ulf Kristoffersson: Thank you for offering to lend me your PhD hat.

I also want to thank the opponent and examination board for the PhD defence. **Alexander Anderson**, thank you for accepting to be the opponent for the thesis defence. **Emma Hammarlund, Lars Hjorth and Jakob Stenman,** thank you for accepting to be in the examination board. **Axel Hyrenius Wittsten,** thank you for accepting to be the chair of the thesis defence. I wish you all luck with running your own research group at the department!

Throughout the projects our research group have also had several collaborators, and I would like to give a specific thanks to them.

Daniel Bexell, Adriana Mañas, Kristina Aaltonen, Aleksandra Adamska, Alexandra Seger, Katarzyna Radke, Javanshir Esfandyari and Madhura Satish Bhave: Thank you for your amazing work with the PDX models for paper II and III. Also thank you for a nice time at the NBCNS meeting.

Alexandra Seger - Lexi: Thank you for a great time during CARES and showing me your work in Bexell's mouse lab. Also thank you for the cat calendar and letting me meet your cats.

Karin Jirström's research group: Thank you for the collaboration.

Alexandra Petersson: Thank you for the good collaboration for your pancreatic cancer evolution project. It was very fun collaborating with you on that project! I wish you all luck with the rest of your PhD! Also, thank you for a great time at the conferences in Banff and Tampa. Furthermore, congratulations to becoming a mother!

Karin Jirström: Thank you for the good collaboration and a nice time at the conference in Banff.

Floris Foier's group: Thank you for the collaboration regarding single cell copy number profiling.

Elna Dickson, Rana Soylu-Kucharz, Amoolya Sai Dwijesha Sofia Lundh, Maria Björkqvist, Åsa Petersén: Thank you for the collaboration on the Huntington manuscript.

Scilifelab and CTG: Thank you for the collaboration.

Barncancerfonden, Cancerfonden, Vetenskapsrådet, Markussens stipendier, Fysiografiska sällskapet, Lund university and John och Augusta Perssons stiftelse: Thank you for the financial support.

Finally, I would like to thank my childhood friends, friends from the university study years, my partner and my family, all of whom have all supported me.

Siri Rydman, Lisa Löfgren, Fanny Bruer, Jeanette (Nette) Samuelsson. Mina kära barndomsvänner. Det är helt otroligt att vi har känt varandra i över 20 år och fortfarande har daglig kontakt, trots att vi bor i olika delar av landet. Ni har alltid funnits där, både när det varit som svårast och som roligast. Jag kan inte föreställa mig vem jag vore idag om det inte vore för er. Ni är de bästa vännerna man kan tänka sig. Jag hoppas att vi får chansen att fortsätta vara vänner i resten av våra liv.

Jake, Shwan och Mattias: Så tacksam att jag fått lära känna er! Hoppas vi får många år till att skapa minnen tillsammans med Siri, Lisa, Fanny och Nette!

Annika Mannerberg and Emilia Palmér: Now all of us are finally 100 % doctors. Thank you for all the amazing years with physics studies, work in the radiation therapy department, travels, Lundakarnevalen, beer festivals, and interesting discussions about everything from julkulor to UV-glasses. #partyvoice

Jonas Scherman and Karl Henriksson: Thank you for all the fun gatherings and interesting discussions.

Lea Styrmisdottir, Matilda Widén, Astrid Öberg, Felicia Feuerstein, Lovisa Sahlén, Freja Franzén, Emilia Blomberg, Alexandra Johansson, Elin Åkesson, Amanda Hörberg, Basem Aloumar and other persons I met and hanged out with in medical school. Thank you for making the years in medical school amazing!

Stina Oleandersson, Mattias Sjö, Emil Boman, Emil Andersson, Markus Ernstsson, Cornelia Säll, Tim Almqvist, Emilia Palmér, Annika Mannerberg, David Boholm, Oscar Erlandsson, Hanna Karlsson, Sebastian Pfaff, Fredrik Bertilsson, Johannes Sandberg, Jill Wiberg, Daniel Persson, Daniel Johansson, Joakim Alnefjord, Nellie Edvardsson, Sofia Borgå, Thomas K Jensen, Jacob Taxén, Daniel Mikkola, Pernilla Helmér, Rasmus Göransson, Lisa Rämisch, Truls Sassner and all other amazing people I've had the privilege to hang out with during and after my physics studies: Thank you for making the physics studies as well as the years after that to an amazing and fun time. Thank you all for all the wonderful memories and long-lasting friendships we've made. Additionally, my heartfelt thanks to the weekly physics pub (**Rydbergs' pub**) held in the cellar at fysicum, which still fuses some of us together each Wednesday evening. These nights have sparked countless interesting discussions on everything from quantum dots to planning of wine tastings.

Katedralskolan Växjö: Thank you for an excellent education and inspirational and supportive teachers.

Mamma och Pappa: Tack för allt stöd genom åren. Jag är så tacksam att jag har haft just er som föräldrar. Utan er hade jag aldrig varit där jag är idag. Tack!

Fredrik Bertilsson: My dearest partner and second half. Thank you for being by my side throughout the entire process, always being there to support me. I look forward to many more years spent together!

References

1. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976;194(4260):23-8.
2. Komarova NL, Burger JA, Wodarz D. Evolution of ibrutinib resistance in chronic lymphocytic leukemia (CLL). Proc Natl Acad Sci U S A. 2014;111(38):13906-11.
3. Dexter DL, Kowalski HM, Blazar BA, Fligiel Z, Vogel R, Heppner GH. Heterogeneity of tumor cells from a single mouse mammary tumor. Cancer research. 1978;38(10):3174-81.
4. Heppner GH. Tumor heterogeneity. Cancer research. 1984;44(6):2259-65.
5. Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. Nat Med. 2016;22(1):105-13.
6. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883-92.
7. Martelotto LG, Ng CKY, Piscuoglio S, Weigelt B, Reis-Filho JS. Breast cancer intra-tumor heterogeneity. Breast Cancer Research : BCR. 2014;16(3):210.
8. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science. 2014;346(6206):256-9.
9. 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.
10. Barry P, Vatsiou A, Spiteri I, Nichol D, Cresswell GD, Acar A, et al. The Spatiotemporal Evolution of Lymph Node Spread in Early Breast Cancer. Clin Cancer Res. 2018;24(19):4763-70.
11. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, et al. The evolutionary history of lethal metastatic prostate cancer. Nature. 2015;520(7547):353-7.
12. Vincent MD. The animal within: carcinogenesis and the clonal evolution of cancer cells are speciation events sensu stricto. Evolution. 2010;64(4):1173-83.
13. Gatenby RA, Brown JS. Integrating evolutionary dynamics into cancer therapy. Nature reviews Clinical oncology. 2020;17(11):675-86.
14. Chabner BA, Roberts TG, Jr. Timeline: Chemotherapy and the war on cancer. Nat Rev Cancer. 2005;5(1):65-72.
15. Tiede S, Kalathur RKR, Luond F, von Allmen L, Szczerba BM, Hess M, et al. Multi-color clonal tracking reveals intra-stage proliferative heterogeneity during mammary tumor progression. Oncogene. 2021;40(1):12-27.

16. Dhimolea E, de Matos Simoes R, Kansara D, Al'Khafaji A, Bouyssou J, Weng X, et al. An Embryonic Diapause-like Adaptation with Suppressed Myc Activity Enables Tumor Treatment Persistence. Cancer Cell. 2021;39(2):240-56 e11.
17. Coffey JC, Wang JH, Smith MJ, Bouchier-Hayes D, Cotter TG, Redmond HP. Excisional surgery for cancer cure: therapy at a cost. Lancet Oncol. 2003;4(12):760-8.
18. Tohme S, Simmons RL, Tsung A. Surgery for Cancer: A Trigger for Metastases. Cancer research. 2017;77(7):1548-52.
19. Baskar R, Dai J, Wenlong N, Yeo R, Yeoh KW. Biological response of cancer cells to radiation treatment. Front Mol Biosci. 2014;1:24.
20. Gillies RJ, Verduzco D, Gatenby RA. Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. Nat Rev Cancer. 2012;12(7):487-93.
21. Sitohy B, Nagy JA, Dvorak HF. Anti-VEGF/VEGFR therapy for cancer: reassessing the target. Cancer research. 2012;72(8):1909-14.
22. El-Sayes N, Vito A, Mossman K. Tumor Heterogeneity: A Great Barrier in the Age of Cancer Immunotherapy. Cancers (Basel). 2021;13(4).
23. Mengelbier LH, Karlsson J, Lindgren D, Valind A, Lilljebjorn H, Jansson C, et al. Intratumoral genome diversity parallels progression and predicts outcome in pediatric cancer. Nat Commun. 2015;6:6125.
24. Vincze O, Colchero F, Lemaitre JF, Conde DA, Pavard S, Bieuville M, et al. Cancer risk across mammals. Nature. 2022;601(7892):263-7.
25. Doonan JH, Sablowski R. Walls around tumours - why plants do not develop cancer. Nat Rev Cancer. 2010;10(11):794-802.
26. Maley CC, Shibata D. Cancer cell evolution through the ages. Science. 2019;365(6452):440-1.
27. Aktipis CA, Boddy AM, Jansen G, Hibner U, Hochberg ME, Maley CC, et al. Cancer across the tree of life: cooperation and cheating in multicellularity. Philos Trans R Soc Lond B Biol Sci. 2015;370(1673).
28. Trigos AS, Pearson RB, Papenfuss AT, Goode DL. How the evolution of multicellularity set the stage for cancer. British journal of cancer. 2018;118(2):145-52.
29. Casas-Selves M, Degregori J. How cancer shapes evolution, and how evolution shapes cancer. Evolution (N Y). 2011;4(4):624-34.
30. Davies PC, Lineweaver CH. Cancer tumors as Metazoa 1.0: tapping genes of ancient ancestors. Phys Biol. 2011;8(1):015001.
31. Merlo LM, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. Nature reviews cancer. 2006;6(12):924-35.
32. Maley CC, Aktipis A, Graham TA, Sottoriva A, Boddy AM, Janiszewska M, et al. Classifying the evolutionary and ecological features of neoplasms. Nat Rev Cancer. 2017;17(10):605-19.
33. Pienta KJ, Hammarlund EU, Axelrod R, Amend SR, Brown JS. Convergent Evolution, Evolving Evolvability, and the Origins of Lethal Cancer. Mol Cancer Res. 2020;18(6):801-10.
34. Caulin AF, Maley CC. Peto's Paradox: evolution's prescription for cancer prevention. Trends Ecol Evol. 2011;26(4):175-82.
35. Tollis M, Boddy AM, Maley CC. Peto's Paradox: how has evolution solved the problem of cancer prevention? BMC Biol. 2017;15(1):60.
36. Sonnenschein C, Soto AM. Theories of carcinogenesis: an emerging perspective. Semin Cancer Biol. 2008;18(5):372-7.

37. Vaux DL. In defense of the somatic mutation theory of cancer. Bioessays. 2011;33(5):341-3.
38. Haas D, Ablin AR, Miller C, Zoger S, Matthay KK. Complete pathologic maturation and regression of stage IVS neuroblastoma without treatment. Cancer. 1988;62(4):818-25.
39. Huggins C. Endocrine-induced regression of cancers. Cancer research. 1967;27(11):1925-30.
40. Mintz B, Illmensee K. Normal genetically mosaic mice produced from malignant teratocarcinoma cells. Proc Natl Acad Sci U S A. 1975;72(9):3585-9.
41. Soto AM, Sonnenschein C. The tissue organization field theory of cancer: a testable replacement for the somatic mutation theory. Bioessays. 2011;33(5):332-40.
42. Sonnenschein C, Soto AM. Somatic mutation theory of carcinogenesis: why it should be dropped and replaced. Mol Carcinog. 2000;29(4):205-11.
43. Sonnenschein C, Soto AM, Rangarajan A, Kulkarni P. Competing views on cancer. J Biosci. 2014;39(2):281-302.
44. Knudson AG. Hereditary cancer: two hits revisited. J Cancer Res Clin Oncol. 1996;122(3):135-40.
45. MacPherson D, Dyer MA. Retinoblastoma: from the two-hit hypothesis to targeted chemotherapy. Cancer research. 2007;67(16):7547-50.
46. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971;68(4):820-3.
47. Paduch R. Theories of cancer origin. European Journal of Cancer Prevention. 2015;24(1):57-67.
48. Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. Nature. 2020;578(7793):82-93.
49. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, et al. Accumulation of driver and passenger mutations during tumor progression. Proc Natl Acad Sci U S A. 2010;107(43):18545-50.
50. Ben-David U, Amon A. Context is everything: aneuploidy in cancer. Nature Reviews Genetics. 2020;21(1):44-62.
51. Gordon DJ, Resio B, Pellman D. Causes and consequences of aneuploidy in cancer. Nat Rev Genet. 2012;13(3):189-203.
52. Chunduri NK, Storchova Z. The diverse consequences of aneuploidy. Nat Cell Biol. 2019;21(1):54-62.
53. Vasudevan A, Schukken KM, Sausville EL, Girish V, Adebambo OA, Sheltzer JM. Aneuploidy as a promoter and suppressor of malignant growth. Nature Reviews Cancer. 2021;21(2):89-103.
54. Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization. Nat Rev Cancer. 2018;18(1):19-32.
55. Rubin H. Fields and field cancerization: the preneoplastic origins of cancer: asymptomatic hyperplastic fields are precursors of neoplasia, and their progression to tumors can be tracked by saturation density in culture. Bioessays. 2011;33(3):224-31.
56. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer. 1953;6(5):963-8.
57. Rowan K. Are cancer stem cells real? After four decades, debate still simmers. J Natl Cancer Inst. 2009;101(8):546-7.

58. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature. 2001;414(6859):105-11.
59. Gupta PB, Chaffer CL, Weinberg RA. Cancer stem cells: mirage or reality? Nat Med. 2009;15(9):1010-2.
60. Valent P, Bonnet D, De Maria R, Lapidot T, Copland M, Melo JV, et al. Cancer stem cell definitions and terminology: the devil is in the details. Nature Reviews Cancer. 2012;12(11):767-75.
61. Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature. 1994;367(6464):645-8.
62. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003;100(7):3983-8.
63. Singh SK, Clarke ID, Terasaki M, Bonn VE, Hawkins C, Squire J, et al. Identification of a cancer stem cell in human brain tumors. Cancer research. 2003;63(18):5821-8.
64. Chaffer CL, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, et al. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proc Natl Acad Sci U S A. 2011;108(19):7950-5.
65. Tomasetti C, Vogelstein B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science. 2015;347(6217):78-81.
66. Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science. 2017;355(6331):1330-4.
67. Rozhok AI, Wahl GM, DeGregori J. A Critical Examination of the "Bad Luck" Explanation of Cancer Risk. Cancer Prev Res (Phila). 2015;8(9):762-4.
68. Jassim A, Rahrmann EP, Simons BD, Gilbertson RJ. Cancers make their own luck: theories of cancer origins. Nature Reviews Cancer. 2023;23(10):710-24.
69. Knudson AG, Jr., Strong LC. Mutation and cancer: a model for Wilms' tumor of the kidney. J Natl Cancer Inst. 1972;48(2):313-24.
70. Deng N, Zhou H, Fan H, Yuan Y. Single nucleotide polymorphisms and cancer susceptibility. Oncotarget. 2017;8(66):110635-49.
71. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017;101(1):5-22.
72. Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. Synonymous mutations frequently act as driver mutations in human cancers. Cell. 2014;156(6):1324-35.
73. Jung H, Lee KS, Choi JK. Comprehensive characterisation of intronic mis-splicing mutations in human cancers. Oncogene. 2021;40(7):1347-61.
74. Castiglia D, Zambruno G. Mutation mechanisms. Dermatol Clin. 2010;28(1):17-22.
75. Wang W, Hellinga HW, Beese LS. Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. Proc Natl Acad Sci U S A. 2011;108(43):17644-8.
76. Watson JD, Crick FH. Genetical implications of the structure of deoxyribonucleic acid. Nature. 1953;171(4361):964-7.
77. Cui P, Ding F, Lin Q, Zhang L, Li A, Zhang Z, et al. Distinct contributions of replication and transcription to mutation rate variation of human genomes. Genomics Proteomics Bioinformatics. 2012;10(1):4-10.

78. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94-101.
79. Nowell P, Hungerford D. A minute chromosome in human chronic granulocytic leukemia. Landmarks in medical genetics: classic papers with commentaries. 2004;132(51):103.
80. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature. 1973;243(5405):290-3.
81. Stam K, Heisterkamp N, Grosveld G, de Klein A, Verma RS, Coleman M, et al. Evidence of a new chimeric bcr/c-abl mRNA in patients with chronic myelocytic leukemia and the Philadelphia chromosome. N Engl J Med. 1985;313(23):1429-33.
82. Yu YP, Liu P, Nelson J, Hamilton RL, Bhargava R, Michalopoulos G, et al. Identification of recurrent fusion genes across multiple cancer types. Sci Rep. 2019;9(1):1074.
83. Pos O, Radvanszky J, Buglyo G, Pos Z, Rusnakova D, Nagy B, et al. DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. Biomed J. 2021;44(5):548-59.
84. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nat Rev Genet. 2009;10(8):551-64.
85. Thompson SL, Bakhoum SF, Compton DA. Mechanisms of chromosomal instability. Curr Biol. 2010;20(6):R285-95.
86. Yan Y, Guo G, Huang J, Gao M, Zhu Q, Zeng S, et al. Current understanding of extrachromosomal circular DNA in cancer pathogenesis and therapeutic resistance. J Hematol Oncol. 2020;13(1):124.
87. Cox D, Yuncken C, Spriggs AI. Minute Chromatin Bodies in Malignant Tumours of Childhood. Lancet. 1965;1(7402):55-8.
88. Kohl NE, Kanda N, Schreck RR, Bruns G, Latt SA, Gilbert F, et al. Transposition and amplification of oncogene-related sequences in human neuroblastomas. Cell. 1983;35(2 Pt 1):359-67.
89. Schwab M, Alitalo K, Klempnauer K-H, Varmus HE, Bishop JM, Gilbert F, et al. Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. Nature. 1983;305(5931):245-8.
90. 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(4653):1121-4.
91. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011;144(1):27-40.
92. Ilic M, Zaalberg IC, Raaijmakers JA, Medema RH. Life of double minutes: generation, maintenance, and elimination. Chromosoma. 2022;131(3):107-25.
93. Yi E, Chamorro Gonzalez R, Henssen AG, Verhaak RGW. Extrachromosomal DNA amplifications in cancer. Nat Rev Genet. 2022;23(12):760-71.
94. Koche RP, Rodriguez-Fos E, Helmsauer K, Burkert M, MacArthur IC, Maag J, et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. Nat Genet. 2020;52(1):29-34.
95. Lange JT, Rose JC, Chen CY, Pichugin Y, Xie L, Tang J, et al. The evolutionary dynamics of extrachromosomal DNA in human cancers. Nat Genet. 2022;54(10):1527-33.

96. Yi E, Gujar AD, Guthrie M, Kim H, Zhao D, Johnson KC, et al. Live-Cell Imaging Shows Uneven Segregation of Extrachromosomal DNA Elements and Transcriptionally Active Extrachromosomal DNA Hubs in Cancer. Cancer Discov. 2022;12(2):468-83.
97. Lundberg G, Rosengren AH, Hakanson U, Stewenius H, Jin Y, Stewenius Y, et al. Binomial mitotic segregation of MYCN-carrying double minutes in neuroblastoma illustrates the role of randomness in oncogene amplification. PloS one. 2008;3(8):e3099.
98. deCarvalho AC, Kim H, Poisson LM, Winn ME, Mueller C, Cherba D, et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. Nat Genet. 2018;50(5):708-17.
99. Chapman OS, Luebeck J, Sridhar S, Wong IT, Dixit D, Wang S, et al. Circular extrachromosomal DNA promotes tumor heterogeneity in high-risk medulloblastoma. Nat Genet. 2023;55(12):2189-99.
100. Albertson DG. Gene amplification in cancer. Trends Genet. 2006;22(8):447-55.
101. Albertson DG, Collins C, McCormick F, Gray JW. Chromosome aberrations in solid tumors. Nat Genet. 2003;34(4):369-76.
102. Matsui A, Ihara T, Suda H, Mikami H, Semba K. Gene amplification: mechanisms and involvement in cancer. Biomol Concepts. 2013;4(6):567-82.
103. Storlazzi CT, Lonoce A, Guastadisegni MC, Trombetta D, D'Addabbo P, Daniele G, et al. Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure. Genome Res. 2010;20(9):1198-206.
104. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575(7781):210-6.
105. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature. 1998;396(6712):643-9.
106. Drews RM, Hernando B, Tarabichi M, Haase K, Lesluyes T, Smith PS, et al. A pan-cancer compendium of chromosomal instability. Nature. 2022;606(7916):976-83.
107. Gisselsson D, Jonson T, Petersen A, Strombeck B, Dal Cin P, Hoglund M, et al. Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. Proc Natl Acad Sci U S A. 2001;98(22):12683-8.
108. Ganem NJ, Godinho SA, Pellman D. A mechanism linking extra centrosomes to chromosomal instability. Nature. 2009;460(7252):278-82.
109. Fei F, Zhang D, Yang Z, Wang S, Wang X, Wu Z, et al. The number of polyploid giant cancer cells and epithelial-mesenchymal transition-related proteins are associated with invasion and metastasis in human breast cancer. J Exp Clin Cancer Res. 2015;34:158.
110. Dewhurst SM, McGranahan N, Burrell RA, Rowan AJ, Gronroos E, Endesfelder D, et al. Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. Cancer Discov. 2014;4(2):175-85.
111. Kuznetsova AY, Seget K, Moeller GK, de Pagter MS, de Roos JA, Durrbaum M, et al. Chromosomal instability, tolerance of mitotic errors and multidrug resistance are promoted by tetraploidization in human cells. Cell Cycle. 2015;14(17):2810-20.
112. Lin KC, Torga G, Sun Y, Axelrod R, Pienta KJ, Sturm JC, et al. The role of heterogeneous environment and docetaxel gradient in the emergence of polyploid, mesenchymal and resistant prostate cancer cells. Clin Exp Metastasis. 2019;36(2):97-108.
113. Pienta KJ, Hammarlund EU, Axelrod R, Brown JS, Amend SR. Poly-aneuploid cancer cells promote evolvability, generating lethal cancer. Evol Appl. 2020;13(7):1626-34.

114. 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(1):116-28.
115. Zhang D, Wang Y, Zhang S. Asymmetric cell division in polyploid giant cancer cells and low eukaryotic cells. Biomed Res Int. 2014;2014:432652.
116. Zhang X, Yao J, Li X, Niu N, Liu Y, Hajek RA, et al. Targeting polyploid giant cancer cells potentiates a therapeutic response and overcomes resistance to PARP inhibitors in ovarian cancer. Sci Adv. 2023;9(29):eadf7195.
117. Bhatia S, Khanna KK, Duijf PHG. Targeting chromosomal instability and aneuploidy in cancer. Trends Pharmacol Sci. 2024;45(3):210-24.
118. de Queiroz K. Nodes, branches, and phylogenetic definitions. Syst Biol. 2013;62(4):625-32.
119. Black JRM, McGranahan N. Genetic and non-genetic clonal diversity in cancer evolution. Nat Rev Cancer. 2021;21(6):379-92.
120. Felsenstein J. Inferring phylogenies. Inferring phylogenies2004. p. 664-.
121. Vaz C, Nascimento M, Carrico JA, Rocher T, Francisco AP. Distance-based phylogenetic inference from typing data: a unifying view. Brief Bioinform. 2021;22(3).
122. Hamming RW. Error detecting and error correcting codes. Bell Labs Technical Journal. 1950;29(2):147-60.
123. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406-25.
124. Gascuel O, Steel M. Neighbor-joining revealed. Mol Biol Evol. 2006;23(11):1997-2000.
125. RR S. A statiscal method for evaluating systematic relationships. Univ Kans sci bull. 1958;38:1409-38.
126. Bromham L, Penny D. The modern molecular clock. Nat Rev Genet. 2003;4(3):216-24.
127. Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J, et al. A Big Bang model of human colorectal tumor growth. Nat Genet. 2015;47(3):209-16.
128. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006;4(5):e88.
129. Rieux A, Eriksson A, Li M, Sobkowiak B, Weinert LA, Warmuth V, et al. Improved calibration of the human mitochondrial clock using ancient genomes. Mol Biol Evol. 2014;31(10):2780-92.
130. Sun JX, Mullikin JC, Patterson N, Reich DE. Microsatellites are molecular clocks that support accurate inferences about history. Mol Biol Evol. 2009;26(5):1017-27.
131. Zhang Y, Kohrn BF, Yang M, Nachmanson D, Soong TR, Lee IH, et al. PolyG-DS: An ultrasensitive polyguanine tract-profiling method to detect clonal expansions and trace cell lineage. Proc Natl Acad Sci U S A. 2021;118(31).
132. Cavalli-Sforza L, Edwards A. The reconstruction of evolution. Ann Hum Genet. 1963;27:105-6.
133. Farris JS. Methods for computing Wagner trees. Systematic Biology. 1970;19(1):83-92.
134. Fitch WM. Toward defining the course of evolution: minimum change for a specific tree topology. Systematic Biology. 1971;20(4):406-16.
135. Page RD. On islands of trees and the efficacy of different methods of branch swapping in finding most-parsimonious trees. Systematic Biology. 1993:200-10.

136. Land AH, Doig AG. An automatic method for solving discrete programming problems: Springer; 2010.
137. Aldrich J, Fisher R. the making of maximum likelihood 1912-22. Department of Economics, University of Southampton, United Kingdom. 1995.
138. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 1981;17(6):368-76.
139. Arenas M. Trends in substitution models of molecular evolution. Front Genet. 2015;6:319.
140. Jukes TH, Cantor CR. Evolution of protein molecules. Mammalian protein metabolism. 1969;3(21):132.
141. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980;16(2):111-20.
142. Kimura M. Estimation of evolutionary distances between homologous nucleotide sequences. Proc Natl Acad Sci U S A. 1981;78(1):454-8.
143. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 1985;22(2):160-74.
144. Tavaré S. Some probabilistic and statistical problems on the analysis of DNA sequence. Lecture of Mathematics for Life Science. 1986;17:57.
145. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10(3):512-26.
146. O'Reilly JE, Puttick MN, Pisani D, Donoghue PCJ. Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. Palaeontology. 2018;61(1):105-18.
147. Nascimento FF, Reis MD, Yang Z. A biologist's guide to Bayesian phylogenetic analysis. Nat Ecol Evol. 2017;1(10):1446-54.
148. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. The journal of chemical physics. 1953;21(6):1087-92.
149. Gregory TR. Understanding natural selection: essential concepts and common misconceptions. Evolution: Education and outreach. 2009;2(2):156-75.
150. Wallace A. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. III Tendency Var. Depart Indefinitely Orig Type J Proc Linn Soc Lond. 1858.
151. Darwin C. On the Origin of Species: Harvard University Press; 1964.
152. Vendramin R, Litchfield K, Swanton C. Cancer evolution: Darwin and beyond. EMBO J. 2021;40(18):e108389.
153. Liu Y. A new perspective on Darwin's Pangenesis. Biol Rev Camb Philos Soc. 2008;83(2):141-9.
154. Liu Y. Like father like son. A fresh review of the inheritance of acquired characteristics. EMBO reports. 2007;8(9):798-803.
155. Huxley J. Evolution. The modern synthesis. Evolution The Modern Synthesis. 1942.
156. Fisher RA. XV.—The correlation between relatives on the supposition of Mendelian inheritance. Earth and Environmental Science Transactions of the Royal Society of Edinburgh. 1919;52(2):399-433.
157. Spencer H. The Principles of Biology: Volume 1: Outlook Verlag; 2020.

158. Wright S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. 1932.
159. de Visser JA, Krug J. Empirical fitness landscapes and the predictability of evolution. Nat Rev Genet. 2014;15(7):480-90.
160. Fragata I, Blanckaert A, Dias Louro MA, Liberles DA, Bank C. Evolution in the light of fitness landscape theory. Trends Ecol Evol. 2019;34(1):69-82.
161. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458(7239):719-24.
162. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007;8(8):610-8.
163. Parker TM, Gupta K, Palma AM, Yekelchyk M, Fisher PB, Grossman SR, et al. Cell competition in intratumoral and tumor microenvironment interactions. EMBO J. 2021;40(17):e107271.
164. Stern DL. The genetic causes of convergent evolution. Nat Rev Genet. 2013;14(11):751-64.
165. Agrawal AA, Zhang X. The evolution of coevolution in the study of species interactions. Evolution. 2021;75(7):1594-606.
166. Anderson NM, Simon MC. The tumor microenvironment. Curr Biol. 2020;30(16):R921-R5.
167. Davis A, Gao R, Navin N. Tumor evolution: Linear, branching, neutral or punctuated? Biochim Biophys Acta Rev Cancer. 2017;1867(2):151-61.
168. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61(5):759-67.
169. Waliszewski P. Controversies about the genetic model of colorectal tumorigenesis. Pol J Pathol. 1995;46(4):239-43.
170. Andersson N, Bakker B, Karlsson J, Valind A, Holmquist Mengelbier L, Spierings DCJ, et al. Extensive Clonal Branching Shapes the Evolutionary History of High-Risk Pediatric Cancers. Cancer research. 2020;80(7):1512-23.
171. Cortes-Ciriano I, Lee JJ, 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(3):331-41.
172. Shen MM. Chromoplexy: a new category of complex rearrangements in the cancer genome. Cancer Cell. 2013;23(5):567-9.
173. Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. Identification of neutral tumor evolution across cancer types. Nat Genet. 2016;48(3):238-44.
174. Fluegen G, Avivar-Valderas A, Wang Y, Padgen MR, Williams JK, Nobre AR, et al. Phenotypic heterogeneity of disseminated tumour cells is preset by primary tumour hypoxic microenvironments. Nat Cell Biol. 2017;19(2):120-32.
175. Grant PR. Convergent and divergent character displacement. Biological journal of the Linnean Society. 1972;4(1):39-68.
176. West J, Ma Y, Newton PK. Capitalizing on competition: An evolutionary model of competitive release in metastatic castration resistant prostate cancer treatment. J Theor Biol. 2018;455:249-60.
177. Abshire D, Lang MK. The Evolution of Radiation Therapy in Treating Cancer. Semin Oncol Nurs. 2018;34(2):151-7.

178. Siveen KS, Prabhu KS, Achkar IW, Kuttikrishnan S, Shyam S, Khan AQ, et al. Role of Non Receptor Tyrosine Kinases in Hematological Malignances and its Targeting by Natural Products. Mol Cancer. 2018;17(1):31.
179. Knight A, Karapetyan L, Kirkwood JM. Immunotherapy in Melanoma: Recent Advances and Future Directions. Cancers (Basel). 2023;15(4).
180. Gatenby RA, Silva AS, Gillies RJ, Frieden BR. Adaptive therapy. Cancer research. 2009;69(11):4894-903.
181. Zhang J, Cunningham JJ, Brown JS, Gatenby RA. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. Nat Commun. 2017;8(1):1816.
182. Gatenby RA, Brown J, Vincent T. Lessons from applied ecology: cancer control using an evolutionary double bind. Cancer research. 2009;69(19):7499-502.
183. Gatenby RA, Zhang J, Brown JS. First Strike-Second Strike Strategies in Metastatic Cancer: Lessons from the Evolutionary Dynamics of Extinction. Cancer research. 2019;79(13):3174-7.
184. Artzy-Randrup Y, Epstein T, Brown JS, Costa RLB, Czerniecki BJ, Gatenby RA. Novel evolutionary dynamics of small populations in breast cancer adjuvant and neoadjuvant therapy. NPJ Breast Cancer. 2021;7(1):26.
185. Hull P. Life in the Aftermath of Mass Extinctions. Curr Biol. 2015;25(19):R941-52.
186. Mutsaers AJ. Metronomic chemotherapy. Top Companion Anim Med. 2009;24(3):137-43.
187. Duong MT, Qin Y, You SH, Min JJ. Bacteria-cancer interactions: bacteria-based cancer therapy. Exp Mol Med. 2019;51(12):1-15.
188. Wolfl B, Te Rietmole H, Salvioli M, Kaznatcheev A, Thuijsman F, Brown JS, et al. The Contribution of Evolutionary Game Theory to Understanding and Treating Cancer. Dyn Games Appl. 2022;12(2):313-42.
189. Kattner P, Strobel H, Khoshnevis N, Grunert M, Bartholomae S, Pruss M, et al. Compare and contrast: pediatric cancer versus adult malignancies. Cancer Metastasis Rev. 2019;38(4):673-82.
190. Johnsen JI, Dyberg C, Wickstrom M. Neuroblastoma-A Neural Crest Derived Embryonal Malignancy. Front Mol Neurosci. 2019;12:9.
191. Rivera MN, Haber DA. Wilms' tumour: connecting tumorigenesis and organ development in the kidney. Nat Rev Cancer. 2005;5(9):699-712.
192. Dressler GR. The cellular basis of kidney development. Annu Rev Cell Dev Biol. 2006;22:509-29.
193. Saab R, Spunt SL, Skapek SX. Myogenesis and rhabdomyosarcoma the Jekyll and Hyde of skeletal muscle. Curr Top Dev Biol. 2011;94:197-234.
194. Oosterhuis JW, Looijenga LHJ. Human germ cell tumours from a developmental perspective. Nat Rev Cancer. 2019;19(9):522-37.
195. London WB, Castleberry RP, Matthay KK, Look AT, Seeger RC, Shimada H, et al. Evidence for an age cutoff greater than 365 days for neuroblastoma risk group stratification in the Children's Oncology Group. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2005;23(27):6459-65.
196. Liu S, Yin W, Lin Y, Huang S, Xue S, Sun G, et al. Metastasis pattern and prognosis in children with neuroblastoma. World J Surg Oncol. 2023;21(1):130.

197. Mosse YP, Laudenslager M, Longo L, Cole KA, Wood A, Attiyeh EF, et al. Identification of ALK as a major familial neuroblastoma predisposition gene. Nature. 2008;455(7215):930-5.
198. Trochet D, Bourdeaut F, Janoueix-Lerosey I, Deville A, de Pontual L, Schleiermacher G, et al. Germline mutations of the paired-like homeobox 2B (PHOX2B) gene in neuroblastoma. Am J Hum Genet. 2004;74(4):761-4.
199. Wenzel A, Schwab M. The mycN/max protein complex in neuroblastoma. Short review. Eur J Cancer. 1995;31A(4):516-9.
200. Huang M, Weiss WA. Neuroblastoma and MYCN. Cold Spring Harb Perspect Med. 2013;3(10):a014415.
201. Seeger RC, Brodeur GM, Sather H, Dalton A, Siegel SE, Wong KY, et al. Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas. N Engl J Med. 1985;313(18):1111-6.
202. Matthay KK, Maris JM, Schleiermacher G, Nakagawara A, Mackall CL, Diller L, et al. Neuroblastoma (Primer). Nature Reviews: Disease Primers. 2016;2(1).
203. Iwahara T, Fujimoto J, Wen D, Cupples R, Bucay N, Arakawa T, et al. Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. Oncogene. 1997;14(4):439-49.
204. Bresler SC, Weiser DA, Huwe PJ, Park JH, Krytska K, Ryles H, et al. ALK mutations confer differential oncogenic activation and sensitivity to ALK inhibition therapy in neuroblastoma. Cancer Cell. 2014;26(5):682-94.
205. Lambertz I, Kumps C, Claeys S, Lindner S, Beckers A, Janssens E, et al. Upregulation of MAPK Negative Feedback Regulators and RET in Mutant ALK Neuroblastoma: Implications for Targeted Treatment. Clin Cancer Res. 2015;21(14):3327-39.
206. Eleveld TF, Oldridge DA, Bernard V, Koster J, Colmet Daage L, Diskin SJ, et al. Relapsed neuroblastomas show frequent RAS-MAPK pathway mutations. Nat Genet. 2015;47(8):864-71.
207. De Brouwer S, De Preter K, Kumps C, Zabrocki P, Porcu M, Westerhout EM, et al. Meta-analysis of neuroblastomas reveals a skewed ALK mutation spectrum in tumors with MYCN amplification. Clin Cancer Res. 2010;16(17):4353-62.
208. Molenaar JJ, Domingo-Fernandez R, Ebus ME, Lindner S, Koster J, Drabek K, et al. LIN28B induces neuroblastoma and enhances MYCN levels via let-7 suppression. Nat Genet. 2012;44(11):1199-206.
209. Schnepp RW, Khurana P, Attiyeh EF, Raman P, Chodosh SE, Oldridge DA, et al. A LIN28B-RAN-AURKA Signaling Network Promotes Neuroblastoma Tumorigenesis. Cancer Cell. 2015;28(5):599-609.
210. Wang H, Wang X, Xu L, Zhang J. TP53 and TP53-associated genes are correlated with the prognosis of paediatric neuroblastoma. BMC Genom Data. 2022;23(1):41.
211. Valentijn LJ, Koster J, Zwijnenburg DA, Hasselt NE, van Sluis P, Volckmann R, et al. TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. Nat Genet. 2015;47(12):1411-4.
212. Peifer M, Hertwig F, Roels F, Dreidax D, Gartlgruber M, Menon R, et al. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. Nature. 2015;526(7575):700-4.
213. Napier CE, Huschtscha LI, Harvey A, Bower K, Noble JR, Hendrickson EA, et al. ATRX represses alternative lengthening of telomeres. Oncotarget. 2015;6(18):16543-58.

214. Bown N, Cotterill S, Lastowska M, O'Neill S, Pearson AD, Plantaz D, et al. Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma. N Engl J Med. 1999;340(25):1954-61.
215. Attiyeh EF, London WB, Mosse YP, Wang Q, Winter C, Khazi D, et al. Chromosome 1p and 11q deletions and outcome in neuroblastoma. N Engl J Med. 2005;353(21):2243-53.
216. Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D, et al. The genetic landscape of high-risk neuroblastoma. Nat Genet. 2013;45(3):279-84.
217. Boeva V, Louis-Brennetot C, Peltier A, Durand S, Pierre-Eugene C, Raynal V, et al. Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. Nat Genet. 2017;49(9):1408-13.
218. Van Groningen T, Koster J, Valentijn LJ, Zwijnenburg DA, Akogul N, Hasselt NE, et al. Neuroblastoma is composed of two super-enhancer-associated differentiation states. Nature genetics. 2017;49(8):1261-6.
219. Manas A, Aaltonen K, Andersson N, Hansson K, Adamska A, Seger A, et al. Clinically relevant treatment of PDX models reveals patterns of neuroblastoma chemoresistance. Sci Adv. 2022;8(43):eabq4617.
220. Gartlgruber M, Sharma AK, Quintero A, Dreidax D, Jansky S, Park YG, et al. Super enhancers define regulatory subtypes and cell identity in neuroblastoma. Nat Cancer. 2021;2(1):114-28.
221. Garner EF, Beierle EA. Cancer Stem Cells and Their Interaction with the Tumor Microenvironment in Neuroblastoma. Cancers (Basel). 2015;8(1).
222. Diede SJ. Spontaneous regression of metastatic cancer: learning from neuroblastoma. Nat Rev Cancer. 2014;14(2):71-2.
223. Nakagawara A, Arima-Nakagawara M, Scavarda NJ, Azar CG, Cantor AB, Brodeur GM. Association between high levels of expression of the TRK gene and favorable outcome in human neuroblastoma. N Engl J Med. 1993;328(12):847-54.
224. Brodeur GM, Bagatell R. Mechanisms of neuroblastoma regression. Nature reviews Clinical oncology. 2014;11(12):704-13.
225. Vo KT, Matthay KK, Neuhaus J, London WB, Hero B, Ambros PF, et al. Clinical, biologic, and prognostic differences on the basis of primary tumor site in neuroblastoma: a report from the international neuroblastoma risk group project. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2014;32(28):3169-76.
226. Sawada T, Sugimoto T, Matsumura T, Tunoda A, Takeda T, Yamamoto K, et al. Mass screening for neuroblastoma in infancy. The Japanese Mass Screening Study Group of Neuroblastoma. Prog Clin Biol Res. 1988;271:525-34.
227. Schilling FH, Spix C, Berthold F, Erttmann R, Sander J, Treuner J, et al. Children may not benefit from neuroblastoma screening at 1 year of age. Updated results of the population based controlled trial in Germany. Cancer Lett. 2003;197(1-2):19-28.
228. Hisashige A, Group NBSE. Effectiveness of nationwide screening program for neuroblastoma in Japan. Glob J Health Sci. 2014;6(4):94-106.
229. Brodeur GM, Pritchard J, Berthold F, Carlsen NL, Castel V, Castelberry RP, et al. Revisions of the international criteria for neuroblastoma diagnosis, staging and response to treatment. Prog Clin Biol Res. 1994;385:363-9.
230. Cohn SL, Pearson AD, London WB, Monclair T, Ambros PF, Brodeur GM, et al. The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task

Force report. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(2):289-97.
231. Park JR, Bagatell R, London WB, Maris JM, Cohn SL, Mattay KK, et al. Children's Oncology Group's 2013 blueprint for research: neuroblastoma. Pediatr Blood Cancer. 2013;60(6):985-93.
232. Matthay KK, Villablanca JG, Seeger RC, Stram DO, Harris RE, Ramsay NK, et al. Treatment of high-risk neuroblastoma with intensive chemotherapy, radiotherapy, autologous bone marrow transplantation, and 13-cis-retinoic acid. Children's Cancer Group. N Engl J Med. 1999;341(16):1165-73.
233. Yu AL, Gilman AL, Ozkaynak MF, London WB, Kreissman SG, Chen HX, et al. Anti-GD2 antibody with GM-CSF, interleukin-2, and isotretinoin for neuroblastoma. N Engl J Med. 2010;363(14):1324-34.
234. Matthay KK, Yanik G, Messina J, Quach A, Huberty J, Cheng SC, et al. Phase II study on the effect of disease sites, age, and prior therapy on response to iodine-131-metaiodobenzylguanidine therapy in refractory neuroblastoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2007;25(9):1054-60.
235. Umapathy G, Mendoza-Garcia P, Hallberg B, Palmer RH. Targeting anaplastic lymphoma kinase in neuroblastoma. APMIS. 2019;127(5):288-302.
236. Pathania AS, Prathipati P, Murakonda SP, Murakonda AB, Srivastava A, Avadhesh, et al. Immune checkpoint molecules in neuroblastoma: A clinical perspective. Semin Cancer Biol. 2022;86(Pt 2):247-58.
237. Miller RW, Young JL, Jr., Novakovic B. Childhood cancer. Cancer. 1995;75(1 Suppl):395-405.
238. Murphy AJ, Cheng C, Williams J, Shaw TI, Pinto EM, Dieseldorff-Jones K, et al. Genetic and epigenetic features of bilateral Wilms tumor predisposition in patients from the Children's Oncology Group AREN18B5-Q. Nat Commun. 2023;14(1):8006.
239. Perotti D, Williams RD, Wegert J, Brzezinski J, Maschietto M, Ciceri S, et al. Hallmark discoveries in the biology of Wilms tumour. Nat Rev Urol. 2023.
240. Maciaszek JL, Oak N, Nichols KE. Recent advances in Wilms' tumor predisposition. Hum Mol Genet. 2020;29(R2):R138-R49.
241. Beckwith JB, Kiviat NB, Bonadio JF. Nephrogenic rests, nephroblastomatosis, and the pathogenesis of Wilms' tumor. Pediatr Pathol. 1990;10(1-2):1-36.
242. Coorens THH, Treger TD, Al-Saadi R, Moore L, Tran MGB, Mitchell TJ, et al. Embryonal precursors of Wilms tumor. Science. 2019;366(6470):1247-51.
243. Park S, Bernard A, Bove KE, Sens DA, Hazen-Martin DJ, Garvin AJ, et al. Inactivation of WT1 in nephrogenic rests, genetic precursors to Wilms' tumour. Nat Genet. 1993;5(4):363-7.
244. Maiti S, Alam R, Amos CI, Huff V. Frequent association of beta-catenin and WT1 mutations in Wilms tumors. Cancer research. 2000;60(22):6288-92.
245. Satoh Y, Nakadate H, Nakagawachi T, Higashimoto K, Joh K, Masaki Z, et al. Genetic and epigenetic alterations on the short arm of chromosome 11 are involved in a majority of sporadic Wilms' tumours. Br J Cancer. 2006;95(4):541-7.
246. Li W, Kessler P, Williams BR. Transcript profiling of Wilms tumors reveals connections to kidney morphogenesis and expression patterns associated with anaplasia. Oncogene. 2005;24(3):457-68.

247. Li CM, Guo M, Borczuk A, Powell CA, Wei M, Thaker HM, et al. Gene expression in Wilms' tumor mimics the earliest committed stage in the metanephric mesenchymal-epithelial transition. Am J Pathol. 2002;160(6):2181-90.
248. Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, Ferdinand JR, et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. Science. 2018;361(6402):594-9.
249. Yoshimizu T, Miroglio A, Ripoche MA, Gabory A, Vernucci M, Riccio A, et al. The H19 locus acts in vivo as a tumor suppressor. Proc Natl Acad Sci U S A. 2008;105(34):12417-22.
250. Williams RD, Chagtai T, Alcaide-German M, Apps J, Wegert J, Popov S, et al. Multiple mechanisms of MYCN dysregulation in Wilms tumour. Oncotarget. 2015;6(9):7232-43.
251. Wegert J, Vokuhl C, Ziegler B, Ernestus K, Leuschner I, Furtwangler R, et al. TP53 alterations in Wilms tumour represent progression events with strong intratumour heterogeneity that are closely linked but not limited to anaplasia. J Pathol Clin Res. 2017;3(4):234-48.
252. Bardeesy N, Falkoff D, Petruzzi MJ, Nowak N, Zabel B, Adam M, et al. Anaplastic Wilms' tumour, a subtype displaying poor prognosis, harbours p53 gene mutations. Nat Genet. 1994;7(1):91-7.
253. Faria P, Beckwith JB, Mishra K, Zuppan C, Weeks DA, Breslow N, et al. Focal versus diffuse anaplasia in Wilms tumor--new definitions with prognostic significance: a report from the National Wilms Tumor Study Group. Am J Surg Pathol. 1996;20(8):909-20.
254. Vujanic GM, Mifsud W, Chowdhury T, Al-Saadi R, Pritchard-Jones K, Renal Tumour Special Interest Group of the Children's C, et al. Characteristics and outcomes of preoperatively treated patients with anaplastic Wilms tumors registered in the UK SIOP-WT-2001 and IMPORT study cohorts (2002-2020). Cancer. 2022;128(8):1666-75.
255. Uno K, Rastegar B, Jansson C, Durand G, Valind A, Chattopadhyay S, et al. A Gradual Transition Toward Anaplasia in Wilms Tumor Through Tolerance to Genetic Damage. Mod Pathol. 2023;37(1):100382.
256. Szychot E, Apps J, Pritchard-Jones K. Wilms' tumor: biology, diagnosis and treatment. Transl Pediatr. 2014;3(1):12-24.
257. van den Heuvel-Eibrink MM, Hol JA, Pritchard-Jones K, van Tinteren H, Furtwangler R, Verschuur AC, et al. Position paper: Rationale for the treatment of Wilms tumour in the UMBRELLA SIOP-RTSG 2016 protocol. Nat Rev Urol. 2017;14(12):743-52.
258. Reinhard H, Semler O, Burger D, Bode U, Flentje M, Gobel U, et al. Results of the SIOP 93-01/GPOH trial and study for the treatment of patients with unilateral nonmetastatic Wilms Tumor. Klin Padiatr. 2004;216(3):132-40.
259. Bhutani N, Kajal P, Sharma U. Many faces of Wilms Tumor: Recent advances and future directions. Ann Med Surg (Lond). 2021;64:102202.
260. Vujanic GM, Gessler M, Ooms A, Collini P, Coulomb-l'Hermine A, D'Hooghe E, et al. The UMBRELLA SIOP-RTSG 2016 Wilms tumour pathology and molecular biology protocol. Nat Rev Urol. 2018;15(11):693-701.
261. Ognjanovic S, Linabery AM, Charbonneau B, Ross JA. Trends in childhood rhabdomyosarcoma incidence and survival in the United States, 1975-2005. Cancer. 2009;115(18):4218-26.

262. Hawkins DS, Spunt SL, Skapek SX, Committee COGSTS. Children's Oncology Group's 2013 blueprint for research: Soft tissue sarcomas. Pediatr Blood Cancer. 2013;60(6):1001-8.
263. Lychou SE, Gustafsson GG, Ljungman GE. Higher rates of metastatic disease may explain the declining trend in Swedish paediatric rhabdomyosarcoma survival rates. Acta Paediatr. 2016;105(1):74-81.
264. Skapek SX, Ferrari A, Gupta AA, Lupo PJ, Butler E, Shipley J, et al. Rhabdomyosarcoma. Nat Rev Dis Primers. 2019;5(1):1.
265. Li H, Sisoudiya SD, Martin-Giacalone BA, Khayat MM, Dugan-Perez S, Marquez-Do DA, et al. Germline Cancer Predisposition Variants in Pediatric Rhabdomyosarcoma: A Report From the Children's Oncology Group. J Natl Cancer Inst. 2021;113(7):875-83.
266. Agaram NP. Evolving classification of rhabdomyosarcoma. Histopathology. 2022;80(1):98-108.
267. Shern JF, Chen L, Chmielecki J, Wei JS, Patidar R, Rosenberg M, et al. Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. Cancer Discov. 2014;4(2):216-31.
268. Tsokos M, Webber BL, Parham DM, Wesley RA, Miser A, Miser JS, et al. Rhabdomyosarcoma. A new classification scheme related to prognosis. Arch Pathol Lab Med. 1992;116(8):847-55.
269. Sorensen PH, Lynch JC, Qualman SJ, Tirabosco R, Lim JF, Maurer HM, et al. PAX3-FKHR and PAX7-FKHR gene fusions are prognostic indicators in alveolar rhabdomyosarcoma: a report from the children's oncology group. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2002;20(11):2672-9.
270. Barr FG, Galili N, Holick J, Biegel JA, Rovera G, Emanuel BS. Rearrangement of the PAX3 paired box gene in the paediatric solid tumour alveolar rhabdomyosarcoma. Nat Genet. 1993;3(2):113-7.
271. Davis RJ, D'Cruz CM, Lovell MA, Biegel JA, Barr FG. Fusion of PAX7 to FKHR by the variant t(1;13)(p36;q14) translocation in alveolar rhabdomyosarcoma. Cancer research. 1994;54(11):2869-72.
272. Buckingham M, Relaix F. The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. Annu Rev Cell Dev Biol. 2007;23:645-73.
273. Xu M, Chen X, Chen D, Yu B, Huang Z. FoxO1: a novel insight into its molecular mechanisms in the regulation of skeletal muscle differentiation and fiber type specification. Oncotarget. 2017;8(6):10662-74.
274. Bennicelli JL, Edwards RH, Barr FG. Mechanism for transcriptional gain of function resulting from chromosomal translocation in alveolar rhabdomyosarcoma. Proc Natl Acad Sci U S A. 1996;93(11):5455-9.
275. Thalhammer V, Lopez-Garcia LA, Herrero-Martin D, Hecker R, Laubscher D, Gierisch ME, et al. PLK1 phosphorylates PAX3-FOXO1, the inhibition of which triggers regression of alveolar Rhabdomyosarcoma. Cancer research. 2015;75(1):98-110.
276. Bharathy N, Suriyamurthy S, Rao VK, Ow JR, Lim HJ, Chakraborty P, et al. P/CAF mediates PAX3-FOXO1-dependent oncogenesis in alveolar rhabdomyosarcoma. J Pathol. 2016;240(3):269-81.

277. Gryder BE, Yohe ME, Chou HC, Zhang X, Marques J, Wachtel M, et al. PAX3-FOXO1 Establishes Myogenic Super Enhancers and Confers BET Bromodomain Vulnerability. Cancer Discov. 2017;7(8):884-99.
278. Wachtel M, Dettling M, Koscielniak E, Stegmaier S, Treuner J, Simon-Klingenstein K, et al. Gene expression signatures identify rhabdomyosarcoma subtypes and detect a novel t(2;2)(q35;p23) translocation fusing PAX3 to NCOA1. Cancer research. 2004;64(16):5539-45.
279. Weber-Hall S, Anderson J, McManus A, Abe S, Nojima T, Pinkerton R, et al. Gains, losses, and amplification of genomic material in rhabdomyosarcoma analyzed by comparative genomic hybridization. Cancer research. 1996;56(14):3220-4.
280. Shukla N, Ameur N, Yilmaz I, Nafa K, Lau CY, Marchetti A, et al. Oncogene mutation profiling of pediatric solid tumors reveals significant subsets of embryonal rhabdomyosarcoma and neuroblastoma with mutated genes in growth signaling pathways. Clin Cancer Res. 2012;18(3):748-57.
281. Anderson J, Gordon A, McManus A, Shipley J, Pritchard-Jones K. Disruption of imprinted genes at chromosome region 11p15.5 in paediatric rhabdomyosarcoma. Neoplasia. 1999;1(4):340-8.
282. Noujaim J, Thway K, Jones RL, Miah A, Khabra K, Langer R, et al. Adult Pleomorphic Rhabdomyosarcoma: A Multicentre Retrospective Study. Anticancer Res. 2015;35(11):6213-7.
283. Drummond CJ, Hanna JA, Garcia MR, Devine DJ, Heyrana AJ, Finkelstein D, et al. Hedgehog Pathway Drives Fusion-Negative Rhabdomyosarcoma Initiated From Non-myogenic Endothelial Progenitors. Cancer Cell. 2018;33(1):108-24 e5.
284. Roma J, Almazan-Moga A, Sanchez de Toledo J, Gallego S. Notch, wnt, and hedgehog pathways in rhabdomyosarcoma: from single pathways to an integrated network. Sarcoma. 2012;2012:695603.
285. Crose LE, Galindo KA, Kephart JG, Chen C, Fitamant J, Bardeesy N, et al. Alveolar rhabdomyosarcoma-associated PAX3-FOXO1 promotes tumorigenesis via Hippo pathway suppression. Journal of Clinical Investigation. 2014;124(1):285-96.
286. Taylor JGt, Cheuk AT, Tsang PS, Chung JY, Song YK, Desai K, et al. Identification of FGFR4-activating mutations in human rhabdomyosarcomas that promote metastasis in xenotransplanted models. Journal of Clinical Investigation. 2009;119(11):3395-407.
287. Chen L, Shern JF, Wei JS, Yohe ME, Song YK, Hurd L, et al. Clonality and evolutionary history of rhabdomyosarcoma. PLoS Genet. 2015;11(3):e1005075.
288. Bisogno G, Jenney M, Bergeron C, Gallego Melcon S, Ferrari A, Oberlin O, et al. Addition of dose-intensified doxorubicin to standard chemotherapy for rhabdomyosarcoma (EpSSG RMS 2005): a multicentre, open-label, randomised controlled, phase 3 trial. Lancet Oncol. 2018;19(8):1061-71.
289. Gartrell J, Pappo A. Recent advances in understanding and managing pediatric rhabdomyosarcoma. F1000Res. 2020;9.
290. Heske CM, Mascarenhas L. Relapsed Rhabdomyosarcoma. J Clin Med. 2021;10(4).
291. Hegde M, Joseph SK, Pashankar F, DeRenzo C, Sanber K, Navai S, et al. Tumor response and endogenous immune reactivity after administration of HER2 CAR T cells in a child with metastatic rhabdomyosarcoma. Nat Commun. 2020;11(1):3549.
292. Pierce JL, Frazier AL, Amatruda JF. Pediatric Germ Cell Tumors: A Developmental Perspective. Adv Urol. 2018;2018:9059382.

293. Ronchi A, Cozzolino I, Montella M, Panarese I, Zito Marino F, Rossetti S, et al. Extragonadal germ cell tumors: Not just a matter of location. A review about clinical, molecular and pathological features. Cancer Med. 2019;8(16):6832-40.
294. Program S. Cancer incidence and survival among children and adolescents: United States SEER program, 1975-1995: National Cancer Institute; 1999.
295. Lanzkowsky P. Chapter 27 - Germ Cell Tumors. In: Lanzkowsky P, editor. Manual of Pediatric Hematology and Oncology (Fifth Edition). San Diego: Academic Press; 2011. p. 776-95.
296. Fonseca A, Frazier AL, Shaikh F. Germ Cell Tumors in Adolescents and Young Adults. J Oncol Pract. 2019;15(8):433-41.
297. Young RH, Dickersin GR, Scully RE. Juvenile granulosa cell tumor of the ovary. A clinicopathological analysis of 125 cases. Am J Surg Pathol. 1984;8(8):575-96.
298. Bin Naeem S, Baloch NU, Jhatial MA, Abbas M, Fasih S, Masood Sheikh R, et al. Clinicopathological Features and Outcomes of Granulosa Cell Tumor of the Ovaries - A Retrospective Study. Cureus. 2023;15(5):e38892.
299. Merras-Salmio L, Vettenranta K, Mottonen M, Heikinheimo M. Ovarian granulosa cell tumors in childhood. Pediatr Hematol Oncol. 2002;19(3):145-56.
300. Dudani R, Giordano L, Sultania P, Jha K, Florens A, Joseph T. Juvenile granulosa cell tumor of testis: case report and review of literature. Am J Perinatol. 2008;25(4):229-31.
301. Bessiere L, Todeschini AL, Auguste A, Sarnacki S, Flatters D, Legois B, et al. A Hot-spot of In-frame Duplications Activates the Oncoprotein AKT1 in Juvenile Granulosa Cell Tumors. EBioMedicine. 2015;2(5):421-31.
302. Mayr D, Kaltz-Wittmer C, Arbogast S, Amann G, Aust DE, Diebold J. Characteristic pattern of genetic aberrations in ovarian granulosa cell tumors. Mod Pathol. 2002;15(9):951-7.
303. Parikshaa G, Ariba Z, Pranab D, Nalini G, Manish R, Vanita S, et al. Juvenile granulosa cell tumor of the ovary: A comprehensive clinicopathologic analysis of 15 cases. Ann Diagn Pathol. 2021;52:151721.
304. Li J, Chu R, Chen Z, Meng J, Yao S, Song K, et al. Progress in the management of ovarian granulosa cell tumor: A review. Acta Obstet Gynecol Scand. 2021;100(10):1771-8.
305. Moraru L, Mitranovici MI, Chiorean DM, Coros M, Moraru R, Oala IE, et al. Immature Teratoma: Diagnosis and Management-A Review of the Literature. Diagnostics (Basel). 2023;13(9).
306. Jiang T, Raynald, Yang H, Wang J, Du J, Zhang W, et al. Primary intracranial embryonal carcinoma in children: report of two cases with review of the literature. Int J Clin Exp Pathol. 2017;10(11):10700-10.
307. DiSaia PJ, Creasman WT, Mannel RS, McMeekin DS, Mutch DG. Clinical gynecologic oncology: Elsevier Health Sciences; 2017.
308. Tahri Y, Moueqqit O, Mokhtari M, Ramdani M, Nadir M, Bennani A, et al. Unusual Presentation of Embryonal Carcinoma of the Testis: A Case Report. Cureus. 2023;15(2):e35175.
309. Matsutani M, Sano K, Takakura K, Fujimaki T, Nakamura O, Funata N, et al. Primary intracranial germ cell tumors: a clinical analysis of 153 histologically verified cases. J Neurosurg. 1997;86(3):446-55.
310. Hanahan D. Rethinking the war on cancer. Lancet. 2014;383(9916):558-63.

311. Vasan N, Baselga J, Hyman DM. A view on drug resistance in cancer. Nature. 2019;575(7782):299-309.
312. Crofton J. Chemotherapy of pulmonary tuberculosis. Br Med J. 1959;1(5138):1610-4.
313. Gottesman MM, Pastan IH. The Role of Multidrug Resistance Efflux Pumps in Cancer: Revisiting a JNCI Publication Exploring Expression of the MDR1 (P-glycoprotein) Gene. J Natl Cancer Inst. 2015;107(9).
314. Muley H, Fado R, Rodriguez-Rodriguez R, Casals N. Drug uptake-based chemoresistance in breast cancer treatment. Biochem Pharmacol. 2020;177:113959.
315. Mansoori B, Mohammadi A, Davudian S, Shirjang S, Baradaran B. The Different Mechanisms of Cancer Drug Resistance: A Brief Review. Adv Pharm Bull. 2017;7(3):339-48.
316. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. 2017;168(4):613-28.
317. Delou JMA, Souza ASO, Souza LCM, Borges HL. Highlights in Resistance Mechanism Pathways for Combination Therapy. Cells. 2019;8(9).
318. Mitola G, Falvo P, Bertolini F. New Insight to Overcome Tumor Resistance: An Overview from Cellular to Clinical Therapies. Life (Basel). 2021;11(11).
319. Endo H, Inoue M. Dormancy in cancer. Cancer Sci. 2019;110(2):474-80.
320. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21.
321. Goldie JH, Coldman AJ. The genetic origin of drug resistance in neoplasms: implications for systemic therapy. Cancer research. 1984;44(9):3643-53.
322. Prieto-Vila M, Usuba W, Takahashi RU, Shimomura I, Sasaki H, Ochiya T, et al. Single-Cell Analysis Reveals a Preexisting Drug-Resistant Subpopulation in the Luminal Breast Cancer Subtype. Cancer research. 2019;79(17):4412-25.
323. Shi ZD, Pang K, Wu ZX, Dong Y, Hao L, Qin JX, et al. Tumor cell plasticity in targeted therapy-induced resistance: mechanisms and new strategies. Signal Transduct Target Ther. 2023;8(1):113.
324. Gunnarsson EB, De S, Leder K, Foo J. Understanding the role of phenotypic switching in cancer drug resistance. J Theor Biol. 2020;490:110162.
325. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature. 2010;467(7319):1114-7.
326. Jin H, Wang L, Bernards R. Rational combinations of targeted cancer therapies: background, advances and challenges. Nat Rev Drug Discov. 2023;22(3):213-34.
327. Xie X, Yu T, Li X, Zhang N, Foster LJ, Peng C, et al. Recent advances in targeting the "undruggable" proteins: from drug discovery to clinical trials. Signal Transduct Target Ther. 2023;8(1):335.
328. Thirant C, Peltier A, Durand S, Kramdi A, Louis-Brennetot C, Pierre-Eugene C, et al. Reversible transitions between noradrenergic and mesenchymal tumor identities define cell plasticity in neuroblastoma. Nat Commun. 2023;14(1):2575.
329. Shieh Y, Eklund M, Sawaya GF, Black WC, Kramer BS, Esserman LJ. Population-based screening for cancer: hope and hype. Nature reviews Clinical oncology. 2016;13(9):550-65.
330. Fletcher JI, Williams RT, Henderson MJ, Norris MD, Haber M. ABC transporters as mediators of drug resistance and contributors to cancer cell biology. Drug Resist Updat. 2016;26:1-9.

331. Yu DM, Huynh T, Truong AM, Haber M, Norris MD. ABC transporters and neuroblastoma. Adv Cancer Res. 2015;125:139-70.
332. Porro A, Haber M, Diolaiti D, Iraci N, Henderson M, Gherardi S, et al. Direct and coordinate regulation of ATP-binding cassette transporter genes by Myc factors generates specific transcription signatures that significantly affect the chemoresistance phenotype of cancer cells. J Biol Chem. 2010;285(25):19532-43.
333. Belounis A, Nyalendo C, Le Gall R, Imbriglio TV, Mahma M, Teira P, et al. Autophagy is associated with chemoresistance in neuroblastoma. BMC Cancer. 2016;16(1):891.
334. Wolpaw AJ, Bayliss R, Buchel G, Dang CV, Eilers M, Gustafson WC, et al. Drugging the "Undruggable" MYCN Oncogenic Transcription Factor: Overcoming Previous Obstacles to Impact Childhood Cancers. Cancer research. 2021;81(7):1627-32.
335. Greengard EG. Molecularly Targeted Therapy for Neuroblastoma. Children (Basel). 2018;5(10).
336. Veschi V, Verona F, Thiele CJ. Cancer Stem Cells and Neuroblastoma: Characteristics and Therapeutic Targeting Options. Front Endocrinol (Lausanne). 2019;10:782.
337. Aravindan N, Jain D, Somasundaram DB, Herman TS, Aravindan S. Cancer stem cells in neuroblastoma therapy resistance. Cancer Drug Resist. 2019;2(4):948-67.
338. Buechner J, Einvik C. N-myc and noncoding RNAs in neuroblastoma. Mol Cancer Res. 2012;10(10):1243-53.
339. Zhou X, Wang X, Li N, Guo Y, Yang X, Lei Y. Therapy resistance in neuroblastoma: Mechanisms and reversal strategies. Front Pharmacol. 2023;14:1114295.
340. Fonseka P, Liem M, Ozcitti C, Adda CG, Ang CS, Mathivanan S. Exosomes from N-Myc amplified neuroblastoma cells induce migration and confer chemoresistance to non-N-Myc amplified cells: implications of intra-tumour heterogeneity. J Extracell Vesicles. 2019;8(1):1597614.
341. Challagundla KB, Wise PM, Neviani P, Chava H, Murtadha M, Xu T, et al. Exosome-mediated transfer of microRNAs within the tumor microenvironment and neuroblastoma resistance to chemotherapy. J Natl Cancer Inst. 2015;107(7).
342. Liu X, Wills CA, Chen L, Zhang J, Zhao Y, Zhou M, et al. Small extracellular vesicles induce resistance to anti-GD2 immunotherapy unveiling tipifarnib as an adjunct to neuroblastoma immunotherapy. J Immunother Cancer. 2022;10(4).
343. Gao Y, Bado I, Wang H, Zhang W, Rosen JM, Zhang XH. Metastasis Organotropism: Redefining the Congenial Soil. Dev Cell. 2019;49(3):375-91.
344. Christensen DS, Ahrenfeldt J, Sokac M, Kisistok J, Thomsen MK, Maretty L, et al. Treatment Represents a Key Driver of Metastatic Cancer Evolution. Cancer research. 2022;82(16):2918-27.
345. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol. 2014;15(3):178-96.
346. De Las Rivas J, Brozovic A, Izraely S, Casas-Pais A, Witz IP, Figueroa A. Cancer drug resistance induced by EMT: novel therapeutic strategies. Arch Toxicol. 2021;95(7):2279-97.
347. Jolly MK, Somarelli JA, Sheth M, Biddle A, Tripathi SC, Armstrong AJ, et al. Hybrid epithelial/mesenchymal phenotypes promote metastasis and therapy resistance across carcinomas. Pharmacol Ther. 2019;194:161-84.

348. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. Nature. 2018;556(7702):463-8.
349. Fares J, Fares MY, Khachfe HH, Salhab HA, Fares Y. Molecular principles of metastasis: a hallmark of cancer revisited. Signal Transduct Target Ther. 2020;5(1):28.
350. De Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. Nat Rev Cancer. 2013;13(2):97-110.
351. Hao Y, Baker D, Ten Dijke P. TGF-beta-Mediated Epithelial-Mesenchymal Transition and Cancer Metastasis. Int J Mol Sci. 2019;20(11).
352. Tijhuis AE, Johnson SC, McClelland SE. The emerging links between chromosomal instability (CIN), metastasis, inflammation and tumour immunity. Mol Cytogenet. 2019;12:17.
353. Chromosomal Instability Drives Metastasis Independent of Aneuploidy. Cancer Discov. 2018;8(3):OF7.
354. Bakhoum SF, Ngo B, Laughney AM, Cavallo JA, Murphy CJ, Ly P, et al. Chromosomal instability drives metastasis through a cytosolic DNA response. Nature. 2018;553(7689):467-72.
355. Li J, Hubisz MJ, Earlie EM, Duran MA, Hong C, Varela AA, et al. Non-cell-autonomous cancer progression from chromosomal instability. Nature. 2023;620(7976):1080-8.
356. Rankin EB, Giaccia AJ. Hypoxic control of metastasis. Science. 2016;352(6282):175-80.
357. Tang M, Bolderson E, O'Byrne KJ, Richard DJ. Tumor Hypoxia Drives Genomic Instability. Front Cell Dev Biol. 2021;9:626229.
358. Zhao Y, Fu X, Lopez JI, Rowan A, Au L, Fendler A, et al. Selection of metastasis competent subclones in the tumour interior. Nat Ecol Evol. 2021;5(7):1033-45.
359. Winkler J, Abisoye-Ogunniyan A, Metcalf KJ, Werb Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. Nat Commun. 2020;11(1):5120.
360. Charras G, Paluch E. Blebs lead the way: how to migrate without lamellipodia. Nat Rev Mol Cell Biol. 2008;9(9):730-6.
361. Friedl P, Gilmour D. Collective cell migration in morphogenesis, regeneration and cancer. Nat Rev Mol Cell Biol. 2009;10(7):445-57.
362. Moreira MCS, Dos Santos AC, Cintra MB. Perineural spread of malignant head and neck tumors: review of the literature and analysis of cases treated at a teaching hospital. Radiol Bras. 2017;50(5):323-7.
363. McDonald DM, Baluk P. Significance of blood vessel leakiness in cancer. Cancer research. 2002;62(18):5381-5.
364. Herman H, Fazakas C, Hasko J, Molnar K, Meszaros A, Nyul-Toth A, et al. Paracellular and transcellular migration of metastatic cells through the cerebral endothelium. J Cell Mol Med. 2019;23(4):2619-31.
365. Denais CM, Gilbert RM, Isermann P, McGregor AL, te Lindert M, Weigelin B, et al. Nuclear envelope rupture and repair during cancer cell migration. Science. 2016;352(6283):353-8.
366. Salavati H, Debbaut C, Pullens P, Ceelen W. Interstitial fluid pressure as an emerging biomarker in solid tumors. Biochim Biophys Acta Rev Cancer. 2022;1877(5):188792.

367. Cambria E, Coughlin MF, Floryan MA, Offeddu GS, Shelton SE, Kamm RD. Linking cell mechanical memory and cancer metastasis. Nat Rev Cancer. 2024;24(3):216-28.
368. Peinado H, Zhang H, Matei IR, Costa-Silva B, Hoshino A, Rodrigues G, et al. Pre-metastatic niches: organ-specific homes for metastases. Nat Rev Cancer. 2017;17(5):302-17.
369. Zomer A, Maynard C, Verweij FJ, Kamermans A, Schafer R, Beerling E, et al. In Vivo imaging reveals extracellular vesicle-mediated phenocopying of metastatic behavior. Cell. 2015;161(5):1046-57.
370. du Bois H, Heim TA, Lund AW. Tumor-draining lymph nodes: At the crossroads of metastasis and immunity. Sci Immunol. 2021;6(63):eabg3551.
371. Follain G, Osmani N, Azevedo AS, Allio G, Mercier L, Karreman MA, et al. Hemodynamic Forces Tune the Arrest, Adhesion, and Extravasation of Circulating Tumor Cells. Dev Cell. 2018;45(1):33-52 e12.
372. Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell. 2014;158(5):1110-22.
373. Goetz JG. Metastases go with the flow. Science. 2018;362(6418):999-1000.
374. Duda DG, Duyverman AM, Kohno M, Snuderl M, Steller EJ, Fukumura D, et al. Malignant cells facilitate lung metastasis by bringing their own soil. Proc Natl Acad Sci U S A. 2010;107(50):21677-82.
375. Lim M, Park S, Jeong HO, Park SH, Kumar S, Jang A, et al. Circulating Tumor Cell Clusters Are Cloaked with Platelets and Correlate with Poor Prognosis in Unresectable Pancreatic Cancer. Cancers (Basel). 2021;13(21).
376. Jiang X, Wong KHK, Khankhel AH, Zeinali M, Reategui E, Phillips MJ, et al. Microfluidic isolation of platelet-covered circulating tumor cells. Lab Chip. 2017;17(20):3498-503.
377. Gay LJ, Felding-Habermann B. Contribution of platelets to tumour metastasis. Nat Rev Cancer. 2011;11(2):123-34.
378. Ju S, Chen C, Zhang J, Xu L, Zhang X, Li Z, et al. Detection of circulating tumor cells: opportunities and challenges. Biomark Res. 2022;10(1):58.
379. Galanzha EI, Menyaev YA, Yadem AC, Sarimollaoglu M, Juratli MA, Nedosekin DA, et al. In vivo liquid biopsy using Cytophone platform for photoacoustic detection of circulating tumor cells in patients with melanoma. Sci Transl Med. 2019;11(496).
380. Diamantopoulou Z, Castro-Giner F, Schwab FD, Foerster C, Saini M, Budinjas S, et al. The metastatic spread of breast cancer accelerates during sleep. Nature. 2022;607(7917):156-62.
381. Houshyari M, Taghizadeh-Hesary F. The Metastatic Spread of Breast Cancer Accelerates during Sleep: How the Study Design can Affect the Results. Asian Pac J Cancer Prev. 2023;24(2):353-5.
382. Berisha A, Shutkind K, Borniger JC. Sleep Disruption and Cancer: Chicken or the Egg? Front Neurosci. 2022;16:856235.
383. Van Dycke KC, Rodenburg W, van Oostrom CT, van Kerkhof LW, Pennings JL, Roenneberg T, et al. Chronically Alternating Light Cycles Increase Breast Cancer Risk in Mice. Curr Biol. 2015;25(14):1932-7.
384. Muller A, Homey B, Soto H, Ge N, Catron D, Buchanan ME, et al. Involvement of chemokine receptors in breast cancer metastasis. Nature. 2001;410(6824):50-6.

385. Zlotnik A. Chemokines in neoplastic progression. Semin Cancer Biol. 2004;14(3):181-5.
386. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. Nat Rev Cancer. 2009;9(4):239-52.
387. Roussos ET, Condeelis JS, Patsialou A. Chemotaxis in cancer. Nat Rev Cancer. 2011;11(8):573-87.
388. Langley RR, Fidler IJ. The seed and soil hypothesis revisited--the role of tumor-stroma interactions in metastasis to different organs. Int J Cancer. 2011;128(11):2527-35.
389. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nat Med. 2013;19(11):1423-37.
390. Massague J, Obenauf AC. Metastatic colonization by circulating tumour cells. Nature. 2016;529(7586):298-306.
391. Reymond N, d'Agua BB, Ridley AJ. Crossing the endothelial barrier during metastasis. Nat Rev Cancer. 2013;13(12):858-70.
392. Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, et al. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. Am J Pathol. 1998;153(3):865-73.
393. Weiss L. Metastatic inefficiency. Adv Cancer Res. 1990;54:159-211.
394. Fidler IJ. Metastasis: quantitative analysis of distribution and fate of tumor emboli labeled with 125 I-5-iodo-2'-deoxyuridine. J Natl Cancer Inst. 1970;45(4):773-82.
395. Bakir B, Chiarella AM, Pitarresi JR, Rustgi AK. EMT, MET, Plasticity, and Tumor Metastasis. Trends Cell Biol. 2020;30(10):764-76.
396. Phan TG, Croucher PI. The dormant cancer cell life cycle. Nat Rev Cancer. 2020;20(7):398-411.
397. Hu Z, Curtis C. Looking backward in time to define the chronology of metastasis. Nat Commun. 2020;11(1):3213.
398. Turajlic S, Swanton C. Metastasis as an evolutionary process. Science. 2016;352(6282):169-75.
399. Giuliano AE, Ballman KV, McCall L, Beitsch PD, Brennan MB, Kelemen PR, et al. Effect of Axillary Dissection vs No Axillary Dissection on 10-Year Overall Survival Among Women With Invasive Breast Cancer and Sentinel Node Metastasis: The ACOSOG Z0011 (Alliance) Randomized Clinical Trial. JAMA. 2017;318(10):918-26.
400. Leiter U, Stadler R, Mauch C, Hohenberger W, Brockmeyer N, Berking C, et al. Complete lymph node dissection versus no dissection in patients with sentinel lymph node biopsy positive melanoma (DeCOG-SLT): a multicentre, randomised, phase 3 trial. Lancet Oncol. 2016;17(6):757-67.
401. Faries MB, Thompson JF, Cochran AJ, Andtbacka RH, Mozzillo N, Zager JS, et al. Completion Dissection or Observation for Sentinel-Node Metastasis in Melanoma. N Engl J Med. 2017;376(23):2211-22.
402. Markowitz SD. Cancer bypasses the lymph nodes. Science. 2017;357(6346):35-6.
403. Naxerova K, Reiter JG, Brachtel E, Lennerz JK, van de Wetering M, Rowan A, et al. Origins of lymphatic and distant metastases in human colorectal cancer. Science. 2017;357(6346):55-60.
404. Mangiola S, Hong MK, Cmero M, Kurganovs N, Ryan A, Costello AJ, et al. Comparing nodal versus bony metastatic spread using tumour phylogenies. Sci Rep. 2016;6:33918.

405. Brown M, Assen FP, Leithner A, Abe J, Schachner H, Asfour G, et al. Lymph node blood vessels provide exit routes for metastatic tumor cell dissemination in mice. Science. 2018;359(6382):1408-11.
406. Pereira ER, Kedrin D, Seano G, Gautier O, Meijer EFJ, Jones D, et al. Lymph node metastases can invade local blood vessels, exit the node, and colonize distant organs in mice. Science. 2018;359(6382):1403-7.
407. Hu Z, Ding J, Ma Z, Sun R, Seoane JA, Scott Shaffer J, et al. Quantitative evidence for early metastatic seeding in colorectal cancer. Nat Genet. 2019;51(7):1113-22.
408. Zhang C, Zhang L, Xu T, Xue R, Yu L, Zhu Y, et al. Mapping the spreading routes of lymphatic metastases in human colorectal cancer. Nat Commun. 2020;11(1):1993.
409. Reiter JG, Hung WT, Lee IH, Nagpal S, Giunta P, Degner S, et al. Lymph node metastases develop through a wider evolutionary bottleneck than distant metastases. Nat Genet. 2020;52(7):692-700.
410. Hong MK, Macintyre G, Wedge DC, Van Loo P, Patel K, Lunke S, et al. Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. Nat Commun. 2015;6:6605.
411. Zhao ZM, 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 U S A. 2016;113(8):2140-5.
412. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer Cell. 2017;32(2):169-84 e7.
413. Fimereli D, Venet D, Rediti M, Boeckx B, Maetens M, Majjaj S, et al. Timing evolution of lobular breast cancer through phylogenetic analysis. EBioMedicine. 2022;82:104169.
414. Gundem G, Levine MF, Roberts SS, Cheung IY, Medina-Martinez JS, Feng Y, et al. Clonal evolution during metastatic spread in high-risk neuroblastoma. Nat Genet. 2023;55(6):1022-33.
415. Kokkat TJ, Patel MS, McGarvey D, LiVolsi VA, Baloch ZW. Archived formalin-fixed paraffin-embedded (FFPE) blocks: A valuable underexploited resource for extraction of DNA, RNA, and protein. Biopreserv Biobank. 2013;11(2):101-6.
416. Astolfi A, Urbini M, Indio V, Nannini M, Genovese CG, Santini D, et al. Whole exome sequencing (WES) on formalin-fixed, paraffin-embedded (FFPE) tumor tissue in gastrointestinal stromal tumors (GIST). BMC Genomics. 2015;16:892.
417. Spencer DH, Sehn JK, Abel HJ, Watson MA, Pfeifer JD, Duncavage EJ. Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. J Mol Diagn. 2013;15(5):623-33.
418. Graw S, Meier R, Minn K, Bloomer C, Godwin AK, Fridley B, et al. Robust gene expression and mutation analyses of RNA-sequencing of formalin-fixed diagnostic tumor samples. Sci Rep. 2015;5:12335.
419. Hedegaard J, Thorsen K, Lund MK, Hein AM, Hamilton-Dutoit SJ, Vang S, et al. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. PloS one. 2014;9(5):e98187.
420. Robbe P, Popitsch N, Knight SJL, Antoniou P, Becq J, He M, et al. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. Genet Med. 2018;20(10):1196-205.

421. Ribeiro-Silva A, Zhang H, Jeffrey SS. RNA extraction from ten year old formalin-fixed paraffin-embedded breast cancer samples: a comparison of column purification and magnetic bead-based technologies. BMC Mol Biol. 2007;8:118.
422. Xu Z, Zhang T, Chen H, Zhu Y, Lv Y, Zhang S, et al. High-throughput single nucleus total RNA sequencing of formalin-fixed paraffin-embedded tissues by snRandom-seq. Nat Commun. 2023;14(1):2734.
423. Kopetz S, Lemos R, Powis G. The promise of patient-derived xenografts: the best laid plans of mice and men. Clin Cancer Res. 2012;18(19):5160-2.
424. Liu Y, Wu W, Cai C, Zhang H, Shen H, Han Y. Patient-derived xenograft models in cancer therapy: technologies and applications. Signal Transduct Target Ther. 2023;8(1):160.
425. Rygaard J, Povlsen CO. Heterotransplantation of a human malignant tumour to "Nude" mice. Acta Pathol Microbiol Scand. 1969;77(4):758-60.
426. Johnson JI, Decker S, Zaharevitz D, Rubinstein LV, Venditti JM, Schepartz S, et al. Relationships between drug activity in NCI preclinical in vitro and in vivo models and early clinical trials. Br J Cancer. 2001;84(10):1424-31.
427. Izumchenko E, Paz K, Ciznadija D, Sloma I, Katz A, Vasquez-Dunddel D, et al. Patient-derived xenografts effectively capture responses to oncology therapy in a heterogeneous cohort of patients with solid tumors. Ann Oncol. 2017;28(10):2595-605.
428. Gao H, Korn JM, Ferretti S, Monahan JE, Wang Y, Singh M, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat Med. 2015;21(11):1318-25.
429. Braekeveldt N, von Stedingk K, Fransson S, Martinez-Monleon A, Lindgren D, Axelson H, et al. Patient-Derived Xenograft Models Reveal Intratumor Heterogeneity and Temporal Stability in Neuroblastoma. Cancer research. 2018;78(20):5958-69.
430. Braekeveldt N, Bexell D. Patient-derived xenografts as preclinical neuroblastoma models. Cell Tissue Res. 2018;372(2):233-43.
431. Braekeveldt N, Wigerup C, Gisselsson D, Mohlin S, Merselius M, Beckman S, et al. Neuroblastoma patient-derived orthotopic xenografts retain metastatic patterns and geno- and phenotypes of patient tumours. Int J Cancer. 2015;136(5):E252-61.
432. Bareham B, Georgakopoulos N, Matas-Cespedes A, Curran M, Saeb-Parsy K. Modeling human tumor-immune environments in vivo for the preclinical assessment of immunotherapies. Cancer Immunol Immunother. 2021;70(10):2737-50.
433. Malaney P, Nicosia SV, Dave V. One mouse, one patient paradigm: New avatars of personalized cancer therapy. Cancer Lett. 2014;344(1):1-12.
434. Kopper O, de Witte CJ, Lohmussaar K, Valle-Inclan JE, Hami N, Kester L, et al. An organoid platform for ovarian cancer captures intra- and interpatient heterogeneity. Nat Med. 2019;25(5):838-49.
435. Xu H, Jiao D, Liu A, Wu K. Tumor organoids: applications in cancer modeling and potentials in precision medicine. J Hematol Oncol. 2022;15(1):58.
436. Vlachogiannis G, Hedayat S, Vatsiou A, Jamin Y, Fernandez-Mateos J, Khan K, et al. Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. Science. 2018;359(6378):920-6.
437. Wensink GE, Elias SG, Mullenders J, Koopman M, Boj SF, Kranenburg OW, et al. Patient-derived organoids as a predictive biomarker for treatment response in cancer patients. NPJ Precis Oncol. 2021;5(1):30.

438. Jung H-S, Lefferts JA, Tsongalis GJ. Utilization of the oncoscan microarray assay in cancer diagnostics. Applied Cancer Research. 2017;37(1):1.
439. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-8.
440. Vasimuddin M, Misra S, Li H, Aluru S, editors. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. 2019 IEEE international parallel and distributed processing symposium (IPDPS); 2019: IEEE.
441. Koboldt DC. Best practices for variant calling in clinical sequencing. Genome Med. 2020;12(1):91.
442. Reble E, Castellani CA, Melka MG, O'Reilly R, Singh SM. VarScan2 analysis of de novo variants in monozygotic twins discordant for schizophrenia. Psychiatr Genet. 2017;27(2):62-70.
443. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nature methods. 2018;15(8):591-4.
444. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2015;32(8):1220-2.
445. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling Somatic SNVs and Indels with Mutect2. bioRxiv. 2019:861054.
446. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
447. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.
448. Speicher MR, Gwyn Ballard S, Ward DC. Karyotyping human chromosomes by combinatorial multi-fluor FISH. Nat Genet. 1996;12(4):368-75.
449. Schrock E, du Manoir S, Veldman T, Schoell B, Wienberg J, Ferguson-Smith MA, et al. Multicolor spectral karyotyping of human chromosomes. Science. 1996;273(5274):494-7.
450. Bakker B, van den Bos H, Lansdorp PM, Foijer F. How to count chromosomes in a cell: An overview of current and novel technologies. Bioessays. 2015;37(5):570-7.
451. van den Bos H, Spierings DC, Taudt AS, Bakker B, Porubsky D, Falconer E, et al. Single-cell whole genome sequencing reveals no evidence for common aneuploidy in normal and Alzheimer's disease neurons. Genome biology. 2016;17(1):116.
452. Matevossian A, Akbarian S. Neuronal nuclei isolation from human postmortem brain tissue. J Vis Exp. 2008(20).
453. Falconer E, Hills M, Naumann U, Poon SS, Chavez EA, Sanders AD, et al. DNA template strand sequencing of single-cells maps genomic rearrangements at high resolution. Nat Methods. 2012;9(11):1107-12.
454. Bakker B, Taudt A, Belderbos ME, Porubsky D, Spierings DC, de Jong TV, et al. Single-cell sequencing reveals karyotype heterogeneity in murine and human malignancies. Genome biology. 2016;17(1):115.
455. Alves JM, Prieto T, Posada D. Multiregional Tumor Trees Are Not Phylogenies. Trends Cancer. 2017;3(8):546-50.
456. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. Nat Methods. 2014;11(4):396-8.

457. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. PLoS Comput Biol. 2014;10(8):e1003665.
458. Ahmadinejad N, Troftgruben S, Maley C, Wang J, Liu L. MAGOS: Discovering Subclones in Tumors Sequenced at Standard Depths. bioRxiv. 2019:790386.
459. Chattopadhyay S, Karlsson J, Valind A, Andersson N, Gisselsson D. Tracing the evolution of aneuploid cancers by multiregional sequencing with CRUST. Brief Bioinform. 2021;22(6).
460. Zare H, Wang J, Hu A, Weber K, Smith J, Nickerson D, et al. Inferring clonal composition from multiple sections of a breast cancer. PLoS Comput Biol. 2014;10(7):e1003703.
461. Popic V, Salari R, Hajirasouliha I, Kashef-Haghighi D, West RB, Batzoglou S. Fast and scalable inference of multi-sample cancer lineages. Genome biology. 2015;16:91.
462. Niknafs N, Beleva-Guthrie V, Naiman DQ, Karchin R. SubClonal Hierarchy Inference from Somatic Mutations: Automatic Reconstruction of Cancer Evolutionary Trees from Multi-region Next Generation Sequencing. PLoS Comput Biol. 2015;11(10):e1004416.
463. El-Kebir M, Satas G, Oesper L, Raphael BJ. Inferring the mutational history of a tumor using multi-state perfect phylogeny mixtures. Cell systems. 2016;3(1):43-53.
464. Caravagna G, Giarratano Y, Ramazzotti D, Tomlinson I, Graham TA, Sanguinetti G, et al. Detecting repeated cancer evolution from multi-region tumor sequencing data. Nat Methods. 2018;15(9):707-14.
465. Oesper L, Mahmoody A, Raphael BJ. THetA: inferring intra-tumor heterogeneity from high-throughput DNA sequencing data. Genome biology. 2013;14(7):R80.
466. Ha G, Roth A, Khattra J, Ho J, Yap D, Prentice LM, et al. TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. Genome Res. 2014;24(11):1881-93.
467. Deshwar AG, Vembu S, Yung CK, Jang GH, Stein L, Morris Q. PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. Genome biology. 2015;16:35.
468. Zaccaria S, Raphael BJ. Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. Nat Commun. 2020;11(1):4301.
469. Letouzé E, Allory Y, Bollet MA, Radvanyi F, Guyon F. Analysis of the copy number profiles of several tumor samples from the same patient reveals the successive steps in tumorigenesis. Genome biology. 2010;11(7):1-19.
470. Kaufmann TL, Petkovic M, Watkins TBK, Colliver EC, Laskina S, Thapa N, et al. MEDICC2: whole-genome doubling aware copy-number phylogenies for cancer evolution. Genome biology. 2022;23(1):241.
471. Ester M, Kriegel H-P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining; Portland, Oregon: AAAI Press; 1996. p. 226–31.
472. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the Evolution of Non–Small-Cell Lung Cancer. New England Journal of Medicine. 2017;376(22):2109-21.

473. Tadeo I, Berbegall AP, Castel V, Garcia-Miguel P, Callaghan R, Pahlman S, et al. Extracellular matrix composition defines an ultra-high-risk group of neuroblastoma within the high-risk patient cohort. Br J Cancer. 2016;115(4):480-9.
474. Andersson N, Chattopadhyay S, Valind A, Karlsson J, Gisselsson D. DEVOLUTION-A method for phylogenetic reconstruction of aneuploid cancers based on multiregional genotyping data. Commun Biol. 2021;4(1):1103.
475. Sakthianandeswaren A, Parsons MJ, Mouradov D, MacKinnon RN, Catimel B, Liu S, et al. MACROD2 Haploinsufficiency Impairs Catalytic Activity of PARP1 and Promotes Chromosome Instability and Growth of Intestinal Tumors. Cancer Discov. 2018;8(8):988-1005.
476. Kresse SH, Ohnstad HO, Paulsen EB, Bjerkehagen B, Szuhai K, Serra M, et al. LSAMP, a novel candidate tumor suppressor gene in human osteosarcomas, identified by array comparative genomic hybridization. Genes Chromosomes Cancer. 2009;48(8):679-93.
477. Meehan M, Parthasarathi L, Moran N, Jefferies CA, Foley N, Lazzari E, et al. Protein tyrosine phosphatase receptor delta acts as a neuroblastoma tumor suppressor by destabilizing the aurora kinase A oncogene. Mol Cancer. 2012;11:6.
478. Bedoya-Reina OC, Li W, Arceo M, Plescher M, Bullova P, Pui H, et al. Single-nuclei transcriptomes from human adrenal gland reveal distinct cellular identities of low and high-risk neuroblastoma tumors. Nat Commun. 2021;12(1):5309.
479. Biedler JL, Helson L, Spengler BA. Morphology and growth, tumorigenicity, and cytogenetics of human neuroblastoma cells in continuous culture. Cancer research. 1973;33(11):2643-52.
480. Bhang HE, Ruddy DA, Krishnamurthy Radhakrishna V, Caushi JX, Zhao R, Hims MM, et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. Nat Med. 2015;21(5):440-8.
481. Karlsson J, Valind A, Holmquist Mengelbier L, Bredin S, Cornmark L, Jansson C, et al. Four evolutionary trajectories underlie genetic intratumoral variation in childhood cancer. Nat Genet. 2018;50(7):944-50.
482. Harris AL. Hypoxia--a key regulatory factor in tumour growth. Nat Rev Cancer. 2002;2(1):38-47.
483. Martinez-Monleon A, Gaarder J, Djos A, Kogner P, Fransson S. Identification of recurrent 3q13.31 chromosomal rearrangement indicates LSAMP as a tumor suppressor gene in neuroblastoma. Int J Oncol. 2023;62(2).
484. Schleiermacher G, Javanmardi N, Bernard V, Leroy Q, Cappo J, Rio Frio T, et al. Emergence of new ALK mutations at relapse of neuroblastoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2014;32(25):2727-34.
485. Fransson S, Martinez-Monleon A, Johansson M, Sjoberg RM, Bjorklund C, Ljungman G, et al. Whole-genome sequencing of recurrent neuroblastoma reveals somatic mutations that affect key players in cancer progression and telomere maintenance. Sci Rep. 2020;10(1):22432.
486. Piskareva O, Harvey H, Nolan J, Conlon R, Alcock L, Buckley P, et al. The development of cisplatin resistance in neuroblastoma is accompanied by epithelial to mesenchymal transition in vitro. Cancer Lett. 2015;364(2):142-55.

487. Debruyne DN, Bhatnagar N, Sharma B, Luther W, Moore NF, Cheung NK, et al. ALK inhibitor resistance in ALK(F1174L)-driven neuroblastoma is associated with AXL activation and induction of EMT. Oncogene. 2016;35(28):3681-91.
488. Mori T, Koga T, Shibata H, Ikeda K, Shiraishi K, Suzuki M, et al. Interstitial Fluid Pressure Correlates Clinicopathological Factors of Lung Cancer. Ann Thorac Cardiovasc Surg. 2015;21(3):201-8.
489. Yu T, Liu K, Wu Y, Fan J, Chen J, Li C, et al. High interstitial fluid pressure promotes tumor cell proliferation and invasion in oral squamous cell carcinoma. Int J Mol Med. 2013;32(5):1093-100.
490. Shah NN, Maatman T, Hari P, Johnson B. Multi Targeted CAR-T Cell Therapies for B-Cell Malignancies. Front Oncol. 2019;9:146.
491. Rozenberg JM, Buzdin AA, Mohammad T, Rakitina OA, Didych DA, Pleshkan VV, et al. Molecules promoting circulating clusters of cancer cells suggest novel therapeutic targets for treatment of metastatic cancers. Front Immunol. 2023;14:1099921.

Printed by Media-Tryck, Lund 2024 NORDIC SWAN ECOLABEL 3041 0903

”Nothing in biology makes sense except in the light of evolution”

Theodosius Dobzhansky

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of laboratory medicine

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2024:46
ISBN 978-91-8021-539-8
ISSN 1652-8220