Somatic Genetic Variation in Children: from Mosaicism to Cancer

Valind, Anders

2017

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

ACCGCGCGGGACCGTTAAATTTG
ACTTGGCGGCTAGGGGTGTGGGC
GAGGTGGCCGGTTTGTTAGGGAG
GTGT
AGCT
AAGC
GGAA

Somatic Genetic Variation in Children: from Mosaicism to Cancer

ANDERS VALIND
DEPARTMENT OF LABORATORY MEDICINE, LUND | LUND UNIVERSITY

CCTGAGCCAGGAATGCAGGATGG
GCGGTGAAGAAGGAAGGGGGTGC
TGAGGTAGGTACGGGAGAAAGCT
TGGGGGCTGGGCCTGAGAGGACA
GCCTGGTAGGTAATAGAAGGCTC
CTCGGAGCAGTCGAGGGGGAGAT
GCACCGTCAGAACAGACCCCACC
AGCCTCCTTCCTTTTGGAGTAGA
GCCGGAGGTGGAGGTGAAAACGC
TGGACTGAGGGCACGTGCAATC
GCCGCATCCAGCTTCAAAGCGAA
AGGCTGCGGGACCCCAGCGGAC
AGAGTTTGGAAGTTCAATGTTG
TGTTTGTAACTCAGCTCACCAC

Printed by Media-Tryck, Lund 2017 NORDIC ECOLABEL, 3041 0903

The alternative to thinking in evolutionary terms is not to think at all.

Sir Peter Medawar

LUND UNIVERSITY
Faculty of Medicine

Department of Laboratory Medicine, Lund
Section of Clinical Genetics

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2017:175
ISBN 978-91-7619-557-4
ISSN 1652-8220

Somatic Genetic Variation in Children: from Mosaicism to Cancer

Somatic Genetic Variation in Children: from Mosaicism to Cancer

Anders Valind, MD

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at Belfragesalen, BMC, Lund

On Friday December 15, 2017 at 13:00.

Faculty opponent
Dr Samra Turajlic, MBBS, MRCP, PhD

The Francis Crick Institute and The Royal Marsden Hospital

Organization	Document name
Lund University, Department of Laboratory Medicine	Doctoral Dissertation
Division of Clinical Genetics	Date of issue: December 15th 2017
Author: Anders Valind	Sponsoring organization

Title and subtitle: Somatic Genetic Variation in Children: from Mosaicism to Cancer

Abstract:

This thesis concerns various aspects of somatic mosaicism and genetic intratumor heterogeneity in childhood cancer.

In paper I, I show that aneuploidy in itself does not lead to the level of chromosomal instability that is typically seen in malignant cells. This finding strongly argues against the so called autocatalytic theory of carcinogenesis.

Paper II illustrates that, in rare cases, low-level somatic mosaicism can be unmasked by hitchhiking on the clonal expansion seen in carcinogenesis. Paper III demonstrates that the level of somatic mosaicism at the copy number level in fetuses is lower than in adult humans and that fetal hepatocytes are no more aneuploid than other cells from the fetus. Furthermore, we also detect an organ specific genomic profile in the fetal thymus, due to physiological T-cell receptor rearrangement.

Paper IV highlights that intratumor genetic heterogeneity is a common feature in chemotherapy treated pediatric cancers. In this paper, we also demonstrate that the presence of genetic heterogeneity within single biopsies is associated with lower event free survival and cancer specific overall survival, and that it was a better prognostic predictor than the burden of somatic genetic aberrations.

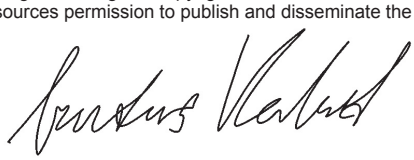
Paper V provides a map of the landscape of intratumor genetic heterogeneity within the primary lesion in Wilms tumor, neuroblastoma and rhabdomyosarcoma. We also discover four different evolutionary trajectories, and show that the presence of some of these evolutionary patterns within the primary tumor predicts inferior survival.

In conclusion, the findings presented in this thesis demonstrate that genetic variation is a rare but significant feature in normal cells of young human tissues. In contrast, such variation is extremely common within childhood solid tumors. Our data suggest that increased knowledge of the evolutionary dynamics within a tumor might lead to improved risk stratification and more personalized treatment.

Key words: Somatic Mosaicism, Childhood Cancer, Tumor Evolution, Wilms Tumor, Neuroblastoma, Rhabdomyosarcoma

Classification system and/or index terms (if any)

Supplementary bibliographical information		Language: English
ISSN and key title: 1652-8220		ISBN: 978-91-7619-557-4
Recipient's notes	Number of pages: 212	Price
	Security classification	

Signature  Date November 9th 2017

Somatic Genetic Variation in Children: from Mosaicism to Cancer

Anders Valind, MD

LUND
UNIVERSITY

Faculty of Medicine
Department of Laboratory Medicine, Lund
Section of Clinical Genetics

ISBN 978-91-7619-557-4
ISSN 1652-8220
Lund University, Faculty of Medicine Doctoral Dissertation Series 2017:175

Printed in Sweden by Media-Tryck, Lund University
Lund 2017

Till Mamma och Pappa

Content

Original Articles

This thesis is based on the following articles:

*= Equal Contribution

Paper I:

Valind A, Jin Y, Baldetorp B, Gisselsson D. **Whole chromosome gain does not in itself confer cancer-like chromosomal instability.** *Proc Natl Acad Sci U S A. 2013 Dec 24;110(52):21119-23.*

Paper II:

Valind A, Pal N, Asmundsson J, Gisselsson D, Holmquist Mengelbier L. **Confined trisomy 8 mosaicism of meiotic origin: a rare cause of aneuploidy in childhood cancer.** *Genes Chromosomes and Cancer.* 2014 Jul;53(7):634-8.

Paper III:

Valind A, Haikal C, Klasson ME, Johansson MC, Gullander J, Soller M, Baldetorp B, Gisselsson D. **The fetal thymus has a unique genomic copy number profile resulting from physiological T cell receptor gene rearrangement.** *Sci Rep.* 2016 Mar 24;6:23500.

Paper VI:

Mengelbier LH*, Karlsson J*, Lindgren D, Valind A, Lilljebjörn H, Jansson C, Bexell D, Braekeveldt N, Ameur A, Jonson T, Kultima HG, Isaksson A, Asmundsson J, Versteeg R, Rissler M, Fioretos T, Sandstedt B, Börjesson A, Backman T, Pal N, Øra I, Mayrhofer M, Gisselsson D. **Intratumoral genome diversity parallels progression and predicts outcome in pediatric cancer.** *Nat Commun.* 2015 Jan 27;6:6125. doi: 10.1038/ncomms7125.

Paper V:

Karlsson J*, Valind A*, Mengelbier LH, Bredin S, Cornmark L, Jansson C, Wali A, Staaf J, Viklund B, Øra I, Börjesson A, Backman T, Braekeveldt N, Sandstedt B, Pal N, Isaksson A, Barbara Gürtl Lackner, Jonson T, Bexell D, & Gisselsson D. **Four evolutionary trajectories underlie genetic intratumor variation in childhood cancer.** *Submitted.*

Papers not included in this thesis:

Karlsson J*, Valind A*, Jansson C, O'Sullivan MJ, Holmquist Mengelbier L, Gisselsson D. **Aberrant epigenetic regulation in clear cell sarcoma of the kidney featuring distinct DNA hypermethylation and EZH2 overexpression.** Oncotarget. 2016 Mar 8;7(10):11127-36

Karlsson J*, Valind A*, Gisselsson D. **BCOR internal tandem duplication and YWHAE-NUTM2B/E fusion are mutually exclusive events in clear cell sarcoma of the kidney.** Genes Chromosomes Cancer. 2016 Feb;55(2):120-3.

Karlsson J, Lilljebjörn H, Holmquist Mengelbier L, Valind A, Rissler M, Øra I, Fioretos T, Gisselsson D. **Activation of human telomerase reverse transcriptase through gene fusion in clear cell sarcoma of the kidney.** Cancer Lett. 2015 Feb 28;357(2):498-501.

Valind A, Gisselsson D. **Reply to Duesberg: Stability of peritriploid and triploid states in neoplastic and nonneoplastic cells.** Proc Natl Acad Sci U S A. 2014 Mar 18;111(11):E975.

Valind A, Gisselsson D. **Reply to Heng: Inborn aneuploidy and chromosomal instability.** Proc Natl Acad Sci U S A. 2014 Mar 18;111(11):E973.

Valind A, Jin Y, Gisselsson D. **Elevated tolerance to aneuploidy in cancer cells: estimating the fitness effects of chromosome number alterations by in silico modelling of somatic genome evolution.** PLoS One. 2013 Jul 24;8(7):e70445.

Abbreviations

ALK = Anaplastic Lymphoma Kinase

AMER1 = APC Membrane Recruitment Protein 1

ARID1A = AT-Rich Interaction Domain 1A

BAF = B-Allele Frequency

BRD4 = Bromodomain Containing 4

ccRCC = Clear Cell Renal Cell Carcinoma

CIN = Chromosomal Instability

CNS = Central Nervous System

CNV = Copy Number Variant

COG = Children's Oncology Group

CRISPR-Cas9 = Clustered Regularly Interspaced Short Palindromic Repeats Associated Protein 9

CTNNB1 = Beta-Catenin

DGCR8 = DiGeorge Syndrome Critical Region Gene 8

DICER1 = Dicer1 Ribonuclease III

DNA = Deoxyribonucleic Acid

DROSHA = Drosha Ribonuclease III

EFS = Event Free Survival

FISH = Fluorescent in situ Hybridization

FOXO1 = Forkhead Box O1

G1 = Gap 1 Phase

HRAS = Harvey Rat Sarcoma Proto-oncogene

HTS = High Throughput Sequencing

IDRF = Image Derived Risk Factor

Indel = Short insertion and/or deletion, a common type of mutation.

INRG-CS = International Neuroblastoma Risk Group Classification System

INRGSS = International Neuroblastoma Risk Group Staging System

INSS = International Neuroblastoma Staging System

ITH = Intratumor Heterogeneity

JQ1 = Jun Qi Bromodomain inhibitor 1

KRAS = Kirsten Rat Sarcoma Proto-Oncogene

LMO1 = LIM Domain Only 1

LOH = Loss of Heterozygosity

MCM2-7 = Minichromosome Maintenance Protein Complex 2-7

MDM2 = Mouse Double Minute 2

MPS = Massively Parallel Sequencing

MYCN = MYCN Proto-Oncogene, bHLH Transcription Factor

MYOD = Myogenic Differentiation 1

MYOG = Myogenin

NB = Neuroblastoma

NF1 = Neurofibromin 1

NGS = Next-Generation Sequencing

NWTS = National Wilms Tumor Study

OS = Overall Survival

PAX3 = Paired Box 3

PAX7 = Paired Box 7

PD1 = Programmed Cell Death Protein 1

PIK3CA = phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha

PTEN = Phosphatase and Tensin Homolog

RMS = Rhabdomyosarcoma

SETD2 = SET Domain Containing 2

SIOP = International Society of Paediatric Oncology

SIX1 = SIX homeobox 1

SIX2 = SIX homeobox 2

SNP = Single Nucleotide Polymorphism

SNV = Single Nucleotide Variant

SV = Structural Variant

TCR = T-Cell Receptor

TERT = Telomerase Reverse Transcriptase

TP53 = Tumor Protein 53

VAF = Variant allele frequency

WGS = Whole Genome Sequencing

WT = Wilms Tumor

Introduction

Genome Stability and Mutation Rates

Life is based on an ever-lasting cycle of cell proliferation, summarized as *omnis cellula e cellula* (all cells come from cells) by the father of modern pathology, Rudolf Virchow. Before each mitotis, the genome of the dividing cell must be duplicated, a process termed replication. Complex regulatory mechanisms have evolved to perform this replication with a high degree of fidelity[1]. Nonetheless, as this apparatus is imperfect, mutations take place. To combat these mutations cells have evolved a myriad of DNA-repair mechanisms. Despite these safe-guards, in each cell division a small number of errors are introduced into the genome of the daughter cells. Such mutations can be subdivided into single nucleotide variants (SNV), small insertions and deletions (indels) as well as larger structural variants (SV) of varying complexity and, finally copy number variants (CNVs), that can range in size from about 100 base-pairs to gains and losses of whole chromosomes, or even whole genome doubling events.

In the last few years multiple approaches have been made to estimate the mutation rate in humans, most of these utilizing high-throughput sequencing (HTS), and focusing on the rate of SNVs and in some studies also indels. It is however imperative to differentiate the per-generation rate of mutations in humans from the per cell division rate, whilst remembering that the per-division germline mutation rate of course is a determinant of the per-generation mutation rate. To further add complexity, recent reports have revealed that the mutation rate in humans depend on which developmental stage the organism is in. On a cellular level, mutation rate varies over different cellular lineages, for instance between the germline and the soma. It is also influenced by local nucleotide context and chromatin state[2]. In addition to this, different classes of mutations vary greatly in their mutation rate as distinct biological processes generate them.

The per-generation mutation rate in humans can, in principle, be estimated by whole genome sequencing of parental-child trios, and then finding mutations seen in the offspring but not in any of the parents[3]. One caveat with this method is that depending on the tissue type used for the parental analysis, mutations can be wrongly classified as *de novo* in the offspring when in fact they are mosaic in one

of the parents[4] or somatic in the child; thus inflating the per-generation mutation rate estimate. Nonetheless, recent results have been consistent in reporting a per-generation mutation rate for SNVs of 1.20 x 10^{-8} to 1.29 x 10^{-8} per nucleotide per generation in humans[5,6]. Another important aspect is the fact that spermatogenesis is a continuous process and thus the paternal germline is composed of cells that divide over the lifetime of the organism. This has been illustrated by multiple studies demonstrating that children with older fathers typically have an increased level of *de novo* mutations[3,6,7]. An interesting aspect of this is that mutations that arise during spermatogenesis and confer a selective advantage on the cellular level might be detrimental to fitness (i.e. pathological) at the organismal level[8–10]. Contrary to this, in human females the set of oocytes is fixed by birth, and mutations generated by processes coupled to replication or mitosis are not accumulated in the oocytes during the lifetime of the organism. Despite this, it was recently discovered that maternal age contributes to a small, but significant, effect on the number of *de novo* SNVs in the offspring[6,11]. When it comes to gains and losses of whole chromosomes, the maternal age is a key determinant[12,13]. The germline mutation rates for copy number variation, indels and structural variants are much less explored, but single study estimates for these are approximately 1.2 x 10^{-2} CNVs per genome per generation in humans[14], 2.94 *de novo* indels per genome per generation[15] and 0.16 *de novo* structural variants per genome per generation[15].

In the case of somatic variation, recent studies based on single cell sequencing have estimated the somatic mutation rate for SNVs to be around 2.66 x 10^{-9} per nucleotide per cell division in fibroblasts[16]. When Milholland et al. recalculated the germline rate from a per generation rate to a per cell division rate, this was approximately 3.3 x 10^{-11} mutations per nucleotide per mitosis, approximately two orders of magnitude smaller than their estimate of the somatic rate. This implies that additional genomic safeguards are acting in the germline, keeping down the accumulation of new mutations.

From this we can surmise that throughout the cell divisions that take place during human embryo- and fetogenesis, a significant number of mutations of various classes are introduced in the cells that constitute a human individual. When this affects only a subset of the body's total cells, the phenomenon is denoted mosaicism. Mosaicism is subdivided into somatic mosaicism, when it is present in a subset of cells in the soma, and germline mosaicism that affects a subset of the germcells in the individual in question. It is important to note that somatic mosaicism is not inherited by the offspring, while germline mosaicism may be inherited but will be constitutional in the offspring.

Somatic Mosaicism

The combined development of high throughput genomic technologies together with the availability of genotypes and sequences from large cohorts of patients (typically generated primarily for genome-wide association studies) have over the last five years generated a deluge of data regarding the presence of various types of somatic mosaicism in humans. For instance, presence of mosaic genetic variation at both the copy number level[17–19] and at the nucleotide level[20–22] have been discovered, and connected to increased risk for a broad range of diseases[20,23–26] as well as advanced age[27]. However, these studies have all been focused on adults, typically in the range of 40 to 60 years of age and have also with few exceptions exclusively studied a single organ (i.e. the hematopoietic system), due to the ease of sampling blood.

After the hematopoietic system the second most studied organ within the field of somatic mosaicism is probably the central nervous system (CNS), where the focus mainly has been on somatic retro-transposition events and copy number variation. Here several studies have demonstrated a presence of somatic mosaicism in the human brain[28], and a global network recently was formed to gain further understanding of to what extent somatic mosaicism in the CNS contributes to disease[29]. Somatic variation, including variants seen as drivers of tumorigenesis and malignancy have also been described in non-malignant diseases, such as endometriosis[30], where mutations in *KRAS* and *ARID1A* was seen in deep-infiltrating endometriosis. Even somatic mutations in *TP53* have been found at very low variant allele frequencies (VAF) in healthy subjects[31]. These findings imply an additional layer of complexity present in the malignant transformation and suggests that even mutations seen as strong drivers of oncogenesis act in a context dependent way. Further reinforcing this view, is the recent report that mutations that are typically assumed to be driving tumorigenesis have been detected in sun-exposed, normal skin, and also seem to be under strong positive selection in this tissue[32].

Tumorigenesis

Cancer is an ancient disease; the first written report of cancer in humans come from old Egypt, where the Edwin Smith papyrus written around 3000 BC contains a report of a breast cancer patient, who the author concluded to be untreatable[33]. During the following five-thousand years additional tumor types were described and various approaches to the treatment of cancer were introduced, often based on what general theory of disease that was in vogue at the time.

In 1914 The German zoologist Theodor Boveri postulated that malignant tumors (i.e. cancer) arises through multipolar mitosis and the abnormal karyotypes that follow these divisions[34]. Boveri had thus proposed the foundation of what today is known as the somatic mutation theory of cancer, long before the chemical structure of DNA had been resolved or even the correct chromosome number in man discovered. The somatic mutation theory of cancer states, in brief, that the clonal expansion of cancerous cells is due to somatic changes in the genome of a single cell, which creates a proliferative advantage for that cell, allowing it to expand and over time out-compete adjacent cells. This theory started gaining traction in the 1960's when Peter Nowell together with his graduate student David Hungerford discovered a small and unusual chromosome in white blood cells from patients diagnosed with chronic myeloid leukemia. They discovered the first specific somatic genetic change associated with malignancy, in a time when most researchers did not believe that cancer was caused by somatic genetic aberrations[35]. In 1973 Janet Rowley showed that this peculiar small chromosome was due to a translocation of chromosomes 9 and 22[36].

Much work has since gone into understanding the set of phenotypes that separates malignant cells from normal cells and that underpin the clinical symptoms seen in cancer. In 2000, Hanahan and Weinberg published a seminal paper, outlining the hallmarks of cancer[37], where they defined a set of six phenotypes (evasion of apoptosis, self-sufficiency in growth signals, insensitivity to anti-growth signals, tissue invasion and metastasis, limitless replicative potential and sustained angiogenesis), and suggested that although the mechanistic basis for each of these abilities may differ between different tumor types, they are invariable present in malignant cells. Hanahan and Weinberg also proposed that a tumor should not be seen as a homogenous clump of cells but instead a complex tissue where cancer cells cooperate with and utilize normal cells, such as immune cells, fibroblasts and endothelial cells. About a decade after their first article they published an updated version[38], adding to the six already defined phenotypes, the idea that it is only possible for a tumor cell to acquire any of the six hallmarks if both of the cell-intrinsic characteristic of genome instability and tumor promoting inflammation are present. They also proposed two new emerging hallmarks; deregulating

cellular energetics and avoiding immune destruction. While the hallmarks proposed by Hanahan and Weinberg are obvious simplifications they have been very influential in shaping the conceptual model of tumorigenesis and also in directing experimental cancer research.

Aneuploidy and Chromosomal Instability

During the early 1920s, the cytogeneticist Theophilus Painter reported that the chromosome number in man was 48[39]. Painter arrived at this by studying paraffin embedded sections from human testis that was thinly sliced and then stained in order to visualize the chromosomes. That the number of chromosomes in man was 48 was widely accepted in the scientific community in the decades that followed. In 1956, Albert Levan and Joe Hin Tijo, working at the department of genetics at Lund University, using improved techniques for staining chromosomes before analysis, showed that, in fact, the chromosome number in man is 46. Three years later, in 1959, the French pediatricians Jerome Lejune and Marthe Gauitier demonstrated the first linkage between chromosome abnormalities and diseases, when they discovered that individuals with Down Syndrome have an extra copy of chromosome 21[40], resulting in a total chromosome count of 47.

Aneuploidy denotes that a cell has a chromosome complement that is not a multiple of the haploid set (n=23). It is a common feature in both solid human cancers, as well as leukemias[41], and is typically generated through a process known as chromosomal instability (CIN)[42], which denotes cells that have a tendency to miss-segregate whole chromosomes or acquire large structural variation. Thus, aneuploidy and chromosomal instability are closely related concepts but denote two separate cellular phenotypes and this distinction is crucial when trying to understand them.

CIN is often subdivided into two types, the first is numerical CIN when the process leads to gains or losses of whole chromosomes and the second is structural chromosomal instability, where the resulting genetic aberrations are smaller structural variants, such as inversions or deletions. It is however important to keep in mind that these process typically are intertwined[43,44]. The baseline rate of chromosomal missegregation have been estimated to approximately one missegregation per 50 cell divisions[45]. Chromosomally unstable tumor cell lines typically have much higher rates, for example chromosomally unstable colon cancer cell lines which have segregation defects in approximately 40% of their cell divisions at anaphase[43].

The mechanisms behind an increased miss-segregation rate of whole chromosomes (and thus numerical CIN) are plentiful. They include merotelic kinetochore-microtubuli attachments, in which single kinetochores bind microtubuli from both spindle poles, leading to chromosomes that lag during anaphase[46,47]; centrosomal abberations[48–50]; as well as multipolar mitosis followed by cytokinesis failure[45]. Mechanisms behind structural chromosomal instability in cancer include telomere dysfunction[51]; breakage-fusion-bridge cycles[52]; and defective replication in micronuclei[44,53]. In the latter, segregation errors lead to micronucleus formation in daughter cells, where defects in the nuclear envelope of the micronuclei leads to chromosomal pulverization. Defective replication in micronuclei provides an enlightening example of the complex interplay behind numerical CIN (largely due to mitotic errors) and structural CIN and also illustrates the problem inherent in treating the two as separate processes.

The cellular effects of an extra copy of a whole chromosome have been rather extensively studied, both in human cells as well as in lower eukaryotes. In yeast it has been shown that that the presence of a single extra copy of a chromosome can drive various processes of genomic instability, including increased missegregation of whole chromosomes and recombination defects[54]. Aneuploidy also causes proteotoxic stress[55], and there is a general transcriptomic aneuploidy-response[56]. A number of studies have shown that aneuploidy in general is detrimental to cellular fitness, in both yeast[57], mouse[58] and human[59] cells. In human cells there also seems to be a general aneuploidy response, at least for single chromosome gains, where the mRNA-levels scale with the copy number of the chromosome in question[60]. The picture is more complicated on the protein level[60], implying that there is some sort of protein-level buffering response to the increased transcription that follows from the extra copy of that specific chromosome. It has also been shown that transcriptionally inert chromosomes do not give rise to the typical aneuploidy phenotype[61], illustrating that it is not just the number of chromosomes per se that drive the phenotypes seen in aneuploid cells.

That single chromosome gains can induce specific forms of genomic instability has also been shown in human cells. Passerini et al. showed that aneuploidy sensitizes cells to replication stress, as evident by an accumulation of trisomic and tetrasomic cells in S-phase upon treatment with aphidicolin compared to control cells. This replication stress sensitivity then leads to an accumulation of sub-chromosomal genomic rearrangements[62]. A putative mechanism connecting aneuploidy and replicative stress sensitivity was the downregulation of the *MCM2*-7-complex. This is a DNA-helicase complex that activates replication origins during S-phase, which showed a 25-50% decrease in protein expression in trisomic and tetrasomic cells.

Interestingly, single chromosome aneuploidy have been shown to have a tumor suppressing effect, both in mouse embryonic fibroblasts (MEF) and in transformed human cells[63], in accordance with the epidemiological observation of reduced frequencies of solid tumors in individuals with Down syndrome[64]. This tumor suppressing effect was seen even in cells transduced with oncogenes such as activating mutations in the *HRAS* gene or with dominant-negative variants of *TP53*. Interestingly, when the cells carrying single chromosome gains were exposed to prolonged *in vitro* culturing they evolved and obtained a higher fitness, both by reverting to the euploid state and by accumulation of structural rearrangements (segmental copy number aberrations). When aneuploid cells face environmental hardships *in vitro*, for instance when serum starved or treated with chemotherapeutic drugs, they also grow faster than genetically matched euploid cell lines[65]. Taken together these findings suggest that in mammalian cells, aneuploidy leads to an increased evolutionary plasticity and may create resilience in cells when they grow in more harsh conditions. Recently, it has also been shown that aneuploidy induced non-genetic variation in a population of cells with clonal aneuploidy[66], adding yet another layer of complexity to the role of aneuploidy in tumorigenesis.

A longstanding question is whether aneuploidy (in the absence of *a priori* malignant transformation) also causes whole chromosome instability. If this was the case, it would create a vicious cycle of genome instability, where an initial miss-segregated chromosome would lead to further chromosomal instability, mediated by the gains or losses of additional whole chromosomes. The aneuploidy could also lead to genome instability for structural aberrations. This idea has been termed the autocatalytic theory of carcinogenesis, and have some very vocal proponents[67,68]. As there is a clear causal connection between numerical CIN and aneuploidy, evaluating whether the converse is true or not is very much a challenge.

Tumor Heterogeneity

In 1976 Peter Nowell proposed that a driving force in cancer progression was clonal evolution due to the inherent genomic instability of malignant cells and the Darwinian selection that acted upon this variation[69]. However, it was not possible at the time to experimentally test Dr Nowells hypothesis in a genome wide-fashion.

Advances in sequencing technology have now made it possible and cost effective to probe the global genomic landscape of cancer cells at the nucleotide level as well as the structural- and copy number levels, and thus to test the evolutionary hypothesis of cancer. This development of high throughput genomic assays, however, led to challenges in dealing with the vast amount of data generated through these emerging technologies. Most tumor samples assayed using HTS technology typically consist of a mix of normal and tumor cells. In addition, within the set of tumor cells present in a tissue sample, multiple subclones may co-exist, typically creating a need to first infer a set of clones before constructing a phylogenetic tree.

In 2012 Marco Gerlinger et al published seminal results from a study[70] where they employed whole exome sequencing and SNP-array analysis to multiple biopsies from four patients with clear cell renal cell carcinoma (ccRCC). They showed that there is substantial convergent evolution targeting genes such as *SETD2* and *PTEN*, previously known to be implicated in the pathogenesis of ccRCC. Furthermore, Gerlinger et al discovered that there is substantial genomic heterogeneity within tumors as well as between primary tumors and metastases. This was an important finding demonstrating that genomic heterogeneity can act as a substrate for Darwinian evolution in cancer. In the years since this study a plethora of papers have been published reporting on intratumor heterogeneity and tumor evolution for many different types of tumors[71–73], proposing different evolutionary models for tumor evolution[74]. These studies have been using different assays for interrogating genetic changes and have also used different methods for clonal deconvolution and phylogenetic inference. The surplus of methods available for evolutionary analysis of tumor genomic datasets[75], and conflicting results regarding the role of Darwinian selection, have generated debate on how much of the reported differences in evolutionary processes operating in cancer is model-dependent[76,77].

Another important caveat with studies on intratumor heterogeneity focused on adult cancers is the fact that these tumors typically have very complex genomes with many passenger mutations and ongoing high-grade genome instability, especially in the metastatic setting which have been the focus of most of the published studies. This can potentially be amended by using childhood cancers as

models of tumor evolution, with the assumption that the underlying principles of evolution ought to be, at the very least, very like the ones operating in cancers that affect adults. The fact that childhood cancers typically have lower levels of somatic point mutations[78] and less complex genomes on the copy number level potentially makes it easier to unravel evolutionary principles. Below follows a presentation of the childhood tumor types used to study tumor evolution in this thesis.

Wilms Tumor

Wilms tumor (WT) is an pediatric renal tumor with a distinct histological pattern that may consist of stromal, epithelial, blastemic and anaplastic compartments[79], with similarities to the embryonal kidney both on the morphological as well as on the gene expression level[80]. The relative contribution of its four histological compartments within a single tumor have important clinical implications, in both the European treatment protocols (SIOP)[79], where pre-operative chemotherapy is a defining feature, and in the north American protocols (NWTS/COG)[81], which does not include pre-operative chemotherapy. In the current protocols used in Sweden, SIOP-2001 and SIOP-UMBRELLA, pre-treated WTs are classified into one of three risk groups (Low Risk, Intermediate Risk, and High Risk) based on histopathological examination. The low risk group is composed of cystic partially differentiated WTs and completely necrotic WTs. Tumors that are either epithelial-type, stromal-type, mixed-type, regressive-type or contain focal anaplasia are classified as intermediate risk and the high risk group contain blastemal-type tumors and WTs with diffuse anaplasia.

WT has an incidence of 8.2 cases per 1 million children below the age of 15 per year[82] in Europe, and is the most common renal neoplasm in children. A small subset of cases of WT are associated with various syndromes, such as Beckwith-Wiedermann syndrome and Deny-Drash syndrome, but the clear majority of cases are sporadic. From the 1960's to 2000's the overall five year survival of patients diagnosed with WT in Sweden rose from 36% to approximately 90%[83,84], underscoring the tremendous advances within the field of pediatric oncology during that time period. However, about 15% of the patients still relapse[85], and relapse typically is associated with a poor prognosis, with survival rates after relapse at around 50%[86]. Thus, there is a clear need for 1) novel biomarkers for early identification of patients with a high risk of relapse and 2) novel therapeutic options for treatment, whilst at the same time keeping treatment-related toxicity to a minimum.

In addition to the classical genetic aberrations in WT (mutations affecting *WT1*, methylation changes targeting chromosome 11p15 as well as activating mutations in *CTNNB1* and *AMER1*), recent studies employing high throughput genomic assays have identified recurrent variants in *DROSHA*, *DGCR8*, *DICER1*, *SIX1*, *SIX2* and *TP53*[87,88] as well as *MYCN*[89] and also larger structural aberrations including loss of heterozygosity (LOH) of 11p[90], 1p and 16q[91] as well as loss of 17p[92] and gain of 1q[93]. However, only a subset of the identified recurrent genomic lesions has been tested as predictive or prognostic markers in prospective trials. One such an example is a study on what role 1q gain has in WT cases treated according to SIOP protocols[93], showing that in multivariate-analysis gain of 1q was associated with poorer event free survival (EFS), but not with overall survival (OS). The authors speculated that this lack of significant association was at least in part due to the small number of deaths in the group, which would be a testament to the currently employed salvage therapy protocols.

Another potential confounder is of course intratumor heterogeneity for the biomarker in question, resulting in both false positives and false negatives when using the proposed biomarker for classification, especially in the situation where only a single tumor biopsy is assayed. Before the findings presented in this thesis, this possibility had not been systematically investigated in childhood solid tumors[72,94].

Neuroblastoma

Neuroblastoma (NB) is a childhood tumor arising from neural crest cells of the developing sympathetic nervous system and it is the most common extra-cranial solid malignancy in children, with an incidence rate of 12.0 per million children below the age of 15 per year in Western Europe[95]. Most children with NB are diagnosed before their 5th birthday, with a median age at diagnosis of 19 months[96]. While a small percentage of NB cases are in part due to constitutional genetic changes, for instance polymorphisms in the *LMO1* gene[97] or germline mutations in the *ALK* gene[98,99], the clear majority is sporadic with no apparent genetic risk factor. There is a large degree of variability in the presenting symptoms in children diagnosed with NB, both due to the fact that the primary tumor can be localized anywhere along the developing sympathetic nervous system (even though the most common localization is the abdomen) and because the neuroblastoma clinical course is highly heterogenous[96]. Exemplifying this clinical heterogeneity is the fact that children below 18 months of age are classified as having low risk even when presenting with metastatic disease[100],

while older children presenting with metastatic disease typically have a poor outcome.

The traditional method of both staging and risk stratifying NB has been through the International Neuroblastoma Staging System (INSS). However, this has recently been replaced by the International Neuroblastoma Risk Group Staging System (INRGSS) and the International Neuroblastoma Risk Group classification system (INRG-CS), in order to harmonize the staging of NB patients to facilitate more informative comparisons of clinical trials[100,101]. INRGSS still uses the INSS criteria for diagnosing NB but removes the dependency on surgical procedures for proper staging. Instead it defines a list of Image-Defined Risk Factors (IDRF) which can be interrogated using standard clinical imaging, that map to four different stages; Localized tumor confined to one compartment and not involving vital structures (L1), loco-regional tumor with at least one IDRF (L2), distant metastatic disease (M) and metastatic disease in children below 18 months of age with metastases confined to skin, liver and bone marrow (MS). INRG-CS uses, in addition to the INRGSS stage, information on patient age, histology, tumor differentiation, presence of *MYCN*-amplification, genetic aberrations affecting 11q and tumor ploidy to determine a pre-treatment risk group for each patient. While not at present integrated into the current clinical risk management program in NB, various recurrent somatic genetic aberrations have been detected using high throughput genomic assays and in some cases correlated to prognosis, such as *TERT*-rearrangements[102,103]. Another clinically important somatic genetic aberration in neuroblastoma is activating mutations involving the *ALK* gene, as these can be targeted with protein kinase inhibitors such as crizotinib[104].

Rhabdomyosarcoma

Rhabdomyosarcoma (RMS) is the most common soft tissue sarcoma in children and adolescents, with five year-survival rates of around 70-80% in patients with localized disease[105]. For patients that present with metastatic disease the three year overall survival rate is still only about 30% even when treated with aggressive multimodal therapy[106]. In general, rhabdomyosarcomas are classified into two major histological groups, embryonal- and alveolar rhabdomyosarcoma. Tumors with an alveolar histology typically have a translocation either between chromosomes 2 and 13 or chromosomes 1 and 13 resulting in *PAX3/PAX7-FOXO1* fusion genes[107]. Indeed, in a recent study of the genomic and transcriptomic patterns in childhood rhabdomyosarcoma fusion-negative tumors with an alveolar histology had a transcriptional profile more similar to embryonal rhabdomyosarcomas than to fusion-positive alveolar cases[108], underscoring the

biological importance of transcription factor fusions. Further reinforcing this notion, the clinical course for fusion-negative alveolar RMS is indistinguishable from embryonal RMS but very different from fusion positive cases of alveolar RMS[109]. Interestingly, methylation patterns seem to further stratify the histological subgroups and also impact survival[110]. RMS in Sweden is typically treated using pan-European protocols, currently the CWS-Guidance or CWS-2007-HR[111]. Of note is that the OS for rhabdomyosarcoma decreased in Sweden during the years 2000 to 2010, a decrease that was partly attributed to an increased frequency of metastatic disease at time of diagnosis[112].

Recent investigations on unraveling the intratumor heterogeneity of RMS using whole-genome sequencing (WGS) of single samples have shown that both fusion-positive and fusion-negative RMS cases typically have one or two subclones[113], where the major clone accounts for about 80% of the malignant cells. Based on their WGS-data, Chen et al. also propose two different sequences of genetic events for the two major subtypes of RMS. The formation of the translocation giving rise to the fusion gene is the initial event in fusion-positive cases, often followed by whole genome duplication. Fusion-negative cases, on the other hand, have two cooperating initial events: activating mutations in the *RAS* pathway and loss of heterozygosity of 11p[113].

New functional studies have shown that the *PAX3-FOXO1* fusion leads to creation of *de-novo* super enhancers that cooperate with transcription factors such as *MYOG*, *MYOD* and *MYCN*. This keeps the tumor cells in a primitive, myoblastic state[114] but also to a cellular dependence on *BRD4*, which can be targeted by bromodomain inhibitors such as JQ1[114]. The possibility of targeting the effects of *PAX3-FOXO1* is appealing from a tumor evolutionary point of view, as the translocation giving rise to the fusion gene is seen as the first event in tumorigenesis[113] in the fusion positive cases.

The Present Study

Aims

This thesis focuses on genetic variation among human somatic cells. The primary aims were to test experimentally the so-called autocatalytic theory of cancer (Paper I), to evaluate whether the monoclonal expansion inherent in tumorigenesis can capture a low level of somatic mosaicism (Paper II), to estimate the level of between-organ somatic mosaicism in fetuses and newborns (Paper III) and finally to initialize grand-scale mapping of the intratumor genetic heterogeneity in solid childhood cancers (Papers IV & V).

Methods

Below follow short descriptions and some general remarks regarding the main methods used in the papers that constitute this thesis. For in-depth descriptions of the specific methods used, the reader is directed to the methods section of each paper.

Patients and Tumor Samples

Paper I was based on commercially available human cells. All tumor samples analyzed in papers II, IV and V were collected after written informed consent for genomic analyses had been obtained from patients' parents or guardians. The samples were collected as part of clinical diagnostic procedures. The tissue samples analyzed in paper III were anonymized tissue specimens from clinical sampling procedures post mortem. All studies were approved by the Lund ethics review board. Tumor tissue samples for genomic analysis were frozen upon arrival to the genetics laboratory and stored at -80° C until DNA extraction. Included tumor samples were reviewed by a pathologist prior to genomic analysis to ascertain that they contained representative tumor tissue.

Fluorescent in situ Hybridization

The development of fluorescent in situ hybridization (FISH) in the 1980's afforded scientists to, in a highly specific manner, visualize genomic loci in single cells. This has had broad applications both in clinical analyses and in the research setting. Today, FISH is an indispensable tool for detecting molecular genetic aberrations in tumor cells, thus enabling diagnosis, risk-stratification and, in some cases, targeted therapy. In short, FISH uses selective oligonucleotide probes (e.g. bacterial artificial chromosomes) that hybridize to their genomic complement. These oligonucleotide probes are linked to specific fluorophores, allowing visualization of specific genomic loci in metaphase chromosomes or interphase nuclei through a fluorescence microscope. The technique thereby allows studies of both the structure and number of human chromosomes. An important aspect of FISH is that it allows targeted interrogation of small populations of cells. Drawbacks of using FISH is an inherent background noise that limits the detection of low-frequency anomalies. Also, it cannot detect single nucleotide changes or smaller copy number changes. It should be noted that, in contrast to SNP Array (see below), it is typically a targeted approach where the selection of probes for each experiment determines which genomic regions to be analyzed.

SNP Array

SNP Arrays were originally designed and developed to allow high-troughput genotyping of thousands to millions polymorphic markers. However, they have also been extremely useful in mapping both copy number variation and loss of heterozygosity (LOH) in both germline and somatic cells. While the two main manufacturers of SNP-Arrays, Affymetrix and Illumina, differ somewhat in their technologies the underlying principles are the same: each genomic array consists of millions of oligonucleotide probes attached to a surface. The amount of DNA present to hybridize to each probe is quantified using fluorescent or light-absorbing tags attached to it. In general, for each locus represented on the array this procedure yields two values. One corresponds to the allelic composition at that locus, denoted as the B-allele frequency (BAF), and is defined as the relative level of one of the two alleles in the sample assayed. The other signal is termed the Log Ratio (LR), and is a log2-transformed relative measure of the dosage of that specific locus in the sample assayed, compared to its copy number in a normal reference panel. In the setting of tumor samples, SNP array data is typically processed with downstream bioinformatics tools such as TAPS[115] or ASCAT[116], to infer tumor cell purity, allele specific copy numbers and tumor ploidy, or tools such as GISTIC[117,118], to infer significantly gained or lost genomic segments.

High Throughput Sequencing

High throughput sequencing (HTS), also termed massively parallel sequencing (MPS) or next-generation sequencing (NGS), have revolutionized genetics by enabling researchers to sequence both DNA and RNA at an unprecedented scale. This makes it possible to analyze whole genomes, whole exomes (the set of all exons in a genome) or targeted regions for both point mutations / single nucleotide variants (SNVs), structural variants and copy number variants. Furthermore, it allows for *de novo* detection of fusion transcripts and splicing variants. It affords researchers the ability to rather quickly generate high quality genome assemblies for organisms that do not have a reference genome.

Illumina is the predominant HTS platform used in the papers that constitute this thesis. On this platform, clonal clusters of template sequences are generated on a flowcell by a process termed solid phase bridge-amplification. This is followed by a three-step cycle starting with addition of fluorophore-tagged nucleotides, which are terminally blocked to ensure that only one nucleotide hybridizes to each template molecule in each cycle. These tagged nucleotides hybridize to the template sequence attached to the flowcell. The next step in this cycle is imaging of the clusters where the fluorophore of the newly hybridized nucleotide emits a light of a specific wavelength. This is then followed by a cleavage step, where the fluorophore is removed from the nucleotide in question and its 3' OH-group is regenerated to facilitate hybridization to the next nucleotide in the template sequence in the following cycle[119]. Other platforms are the Ion Torrent semiconductor based sequencing from Life Technologies (Thermo Fisher Scientific) and long read sequencing technologies from companies like Pacific Biosciences and Oxford Nanopore. Very briefly, long read sequencing is important both for phasing variants, de-novo assembly of genomes without a reference genome and detection of larger structural variants and complex genomic rearrangements, that typically are hard to detect using technologies with shorter reads lengths[119].

All sequencing technologies in use today have important drawbacks and limitations and the best choice for a certain experiment depends highly on what type of genetic changes that needs to be detected. The very rapid development of these new techniques has also fueled the development of a wide array of new tools to analyze the very large datasets generated by HTS.

Processing of High Throughput Sequencing Data

In the following section a general summary of the principles of analyzing HTS data with the purpose of detecting somatic variants is presented. Typically, the raw reads are converted from a platform specific file format to the fastq file format. A fastq file is a text file that for each read contains a triplet of information; a read name, the sequence of base-called nucleotides and a list of base-call quality scores, one for each nucleotide, which represent the probability of that nucleotide being a sequencing error.

After appropriate quality checks on the raw read data and optional trimming of synthetic adapter sequences, the reads are typically aligned to a reference genome as a first step to detect genetic changes. The process of mapping short reads to a large reference genome is an interesting technical challenge, and very much an active field of research[120]. As the quality of mapping specific reads to the reference genome is dependent on the similarity of that read to the genome in question, reads spanning variants will, on average, map less well to the reference genome than reads that do not. This can be seen by reference allele bias, a problem that predominantly affects short insertions and deletions (Indels)[121]. After the mapping stage is complete, typical workflows perform quality control steps such as removing duplicated sequences and perform statistical analysis of the mapping stage to discover outlier samples with respect to the percentage of mapped reads, mapping quality and coverage of the targeted regions. The computational analysis of MPS data presented in this thesis have largely followed published best practice guidelines, using the bwa mem alignment software[122] and utilizing the Picard[123] and GATK[124,125] toolkits for post-mapping processing and quality control checks. We used published, and well validated tools, such as MuTect[126] and Scalpel[127] for variant detection.

Tumor Phylogenetics

A rising interest in tumor evolution and intratumor heterogeneity has fueled the development of methods for inference of phylogenetic trees representing cancer evolution. There are several important obstacles that need to be tackled in order to properly use phylogenetic methods to reflect tumor evolution. First, when using bulk-sequencing, either on single samples or multiple samples per tumor, it is vital to keep in mind that in general, each sample contains multiple tumor cell clones. This makes it necessary to perform some sort of clonal deconvolution and then use the clone estimates to generate a phylogenetic tree[77]. Furthermore, evidence is accumulating that the infinite-sites model (i.e. that mutations only arise once and never disappears) probably is violated in many tumors[128]. Therefore, evolutionary models distinct from classical species evolution scenarios have been explored, at

least for single cell data[129]. In that setting, methods such as SiFiT[129], which allows mutations to arise multiple times, seem to be an improvement compared to methods based on the infinite-sites assumption. It is also important to realize that the variant calling stage and inference of tumor phylogenies are intertwined and methods that estimate these two at the same time could be an interesting venue to explore. An example of this is variant calling on multiple related tumor samples, where detection of very small subclones could be improved by "sharing" of information between tumor samples[130].

Results

Paper I

Whole chromosome aneuploidy does not lead to cancer like chromosomal instability

A debated issue in the field of tumorigenesis is whether aneuploidy per se induces chromosomal instability, i.e. if an acquired change in chromosome number in turn triggers further chromosomal copy number change. This so called autocatalytic theory has been proposed, with emphasis, by some cancer researchers over many years[67,68]. However, the experimental evidence for or against it has remained scarce[67]. The theory agrees poorly with the fact that some tumors can be highly aneuploid and yet seems to have little chromosomal variability between cells[45,131]. In Paper I we employ and validate a dual-probe FISH strategy to enable precise estimates of low-frequency aneuploidy in cultured cells. The fact that we use two probes per chromosome means that we can estimate the per-probe hybridization error and use this to remove noise from our estimates of the true level of somatic aneuploidy. For validation, we compared the estimates of somatic aneuploidy in euploid fibroblasts using our dual-probe FISH method and the gold standard method in cytogenetics - analysis of G-banded metaphases.

We hypothesized that intercellular variation in chromosome copy number should be a reasonably good proxy of ongoing chromosomal-instability. This led us to interrogate the level of chromosomal variation in a panel of cell lines carrying constitutional trisomies in order to test the central hypothesis of the autocatalytic theory of cancer. If the autocatalytic theory holds true, these cell lines should show a higher rate of chromosome copy variation than euploid cells. We used fibroblasts rather than cancer cells as the former represent a clean system for studying the effects of aneuploidy in isolation from cancer associated sequence mutations and epigenetic changes. As controls, we used two colon cancer cell lines, one "chromosomally stable" (DLD1) and the other with high levels of "chromosomal

instability" (SW480)[132]. Our analysis showed that constitutionally hyperdiploid cells typically had a lower rate of chromosomal variation than even the "chromosomally stable" cancer cell line. This was true for all trisomies investigated; three cases of trisomy 21, one case each of trisomy 8, trisomy 13 and trisomy 18 as well as for a case of double trisomy 2 and 21. We also analyzed the levels of somatic aneuploidy in two triploid cell lines, and showed that triploid cells in fact had an increased level of chromosomal copy number variation compared to euploid fibroblasts. However, this was largely due to *in vitro* accumulation of cells that had reversed from trisomy to disomy for the chromosomes assayed. In order to rule out the very small chance of triploid mixoploidy we also performed DNA densitometry on the triploid cases, and this showed only peaks corresponding to the near-triploid population. In conclusion the data from Paper I strongly argue against the autocatalytic theory of aneuploidy in cancer.

Paper II

Confined trisomy 8 mosaicism of meiotic origin: a rare cause of aneuploidy in childhood cancer

In Paper II we describe a case of trisomy 8 of meiotic origin ascertained through a sporadic WT in a 2.5 year old boy. Tumor and normal tissues were initially investigated by SNP array in the context of a larger study. This revealed two somatic trisomies (+6, +12) that agreed very well with a WT diagnosis. However, the allelic profile for chromosome 8 showed a complex pattern that could only be explained by the presence of three distinct haplotypes. This was only compatible with a meiotic event in the clinical context of the patient, i.e. a situation without any feasible contamination of a foreign cell lineage and where only a single chromosome exhibited the three haplotypes.

The SNP array analysis of the normal kidney showed a normal chromosome 8 profile, with only two haplotypes. This indicated that the detected trisomy 8 mosaic clone was not present in the embryo to an extent where it came to dominate the kidney parenchyma. We also performed FISH on touch preparations from the normal kidney but was unable to find any evidence for cell populations trisomic for chromosome 8. This was also true for peripherial blood samples from the patient. The patient had no developmental delay or malformations raising any suspicion of a mosaic trisomy 8 syndrome. Parental blood samples were not available and we could not discern whether the extra haplotype present in the tumor cells was of maternal or paternal origin. In conclusion, the most likely explanation for the finding was that the fertilized zygote was trisomic for chromosome 8 but that this was corrected during embryogenesis in the vast

majority of cells. Our report illustrates that mutations present in low level mosaic clones may in some cases become detectable by hitchhiking on the clonal expansion of carcinogenesis.

Paper III

The fetal thymus has a unique genomic copy number profile resulting from physiological T cell receptor gene rearrangements

In Paper III we mapped the level of somatic mosaicism at the copy number level in a cohort of five fetuses and newborns using high resolution genomic arrays. We analyzed between 3 and 8 organs per patient, corresponding to 28 organs in total. To minimize the level of false positives we assayed all organs that passed the quality control on the discovery array using another array platform as well. We required CNV calls from both arrays to have at least 50% reciprocal overlap to call the copy number change as a true positive.

Our analysis show that fetuses and newborns have a significantly lower level of somatic mosaicism at the copy number level compared to adults, at least for the organs assayed. An exception from this paucity of mosaic genomic imbalances is the fetal thymus. This organ had deletions that distinguished it from other organs, due to the physiological process of TCR-gene rearrangements during T-cell ontogeny. In addition to looking at the somatic copy number changes between organs we also used FISH to study intra-organ copy number variation at the whole chromosome level. Standard single probe FISH revealed very little difference among different organs in this respect. An exception to this was the liver, where we observed high levels of hepatic aneuploidy, comparable to those reported in earlier studies, which were also based largely on single probe FISH[133,134]. However, FISH by the stringent two-color method developed in Paper I revealed that this finding, was largely due to a high number of false positive signals in hepatic tissue. By the two-color method, hepatocytes showed rates of whole-chromosome number change comparable to more recent single cell sequencing studies[135]. We also used FISH and DNA-densitometry to interrogate the level of polyploidy in the fetal liver. We found no evidence that fetal liver cells should be any different in chromosome copy number status than cells from any other organ. This was in line with polyploidization being a postnatal, differentiation related phenomenon.

Paper IV

Intratumoral genome diversity parallels progression and predicts outcome in pediatric cancer.

In Paper IV we provide the first estimate of intratumor heterogeneity (ITH) in solid childhood cancers using primarily data from whole genome genotyping arrays of samples from WT, NB, hepatoblastoma and malignant rhabdoid tumor. We use the TAPS bioinformatics tool to estimate the clone sizes from the array data, and validate clone sizes using FISH. We first focus on seven patients where multiple samples from the same primary tumor was available after chemotherapy. We show that five of these patients exhibited ITH at the copy number level. Within the setting of ITH as branching evolution, there were also cases of convergent evolution affecting loci known to be important to tumorigenesis in the tumor types in question, such as LOH at 1p or 16q in WT. Using a xenograft system we also show that the level of chromosomal instability *in vitro* was positively correlated to the presence of subclones *in vivo*.

The most common type of ITH was the presence of subclones within a single tumor biopsy, a phenomenon we refer to as microdiversity. We show that the presence of microdiversity in the primary tumor could act as a substrate for heterogeneity among samples in a subsequent relapse or metastatic setting. Finally, using a cohort of 44 cases of WTs treated according to the SIOP-2001 protocol, we show that the presence of microdiversity after chemotherapy was associated with lower event free and overall cancer-specific survival. Microdiversity was a superior predictor than the burden of somatic genetic aberrations or established indices for genome instability.

Paper V

Four evolutionary trajectories underlie genetic intratumor variation in childhood cancer

In Paper V we perform a large scale mapping of intratumor genetic heterogeneity in 54 cases of the three most common extracranial malignancies of childhood (WT, NB and RMS). We employ multiregional tumor sampling (median of 4 samples per tumor) followed by high-resolution whole genome genotyping arrays, whole exome sequencing and targeted deep sequencing on multiple samples per patient. We focus primarily on the patterns of ITH present in the primary tumor and generate empirical evolutionary ideograms depicting the evolution of subclones within each primary tumor. We opted to manually construct these evolutionary ideograms to overcome the limitations in current computational tools for tumor evolution inference.

By comparing the clonal composition between all samples originating from each primary tumor we found four different evolutionary patterns that recur in all tumor types: clonal coexistence (COEX) where subclones are detected alongside their parental clone over two or more samples, subclonal variation (VAR) where single samples contain private subclones, clonal sweeps (SWE), where a novel subclone grows to completely dominate a tumor region and finally, clonal explosions (EXP), which is defined as a dramatic difference in the number of genomic imbalances among samples from the same tumor, i.e. extensive branching evolution (Figures 1 & 2).

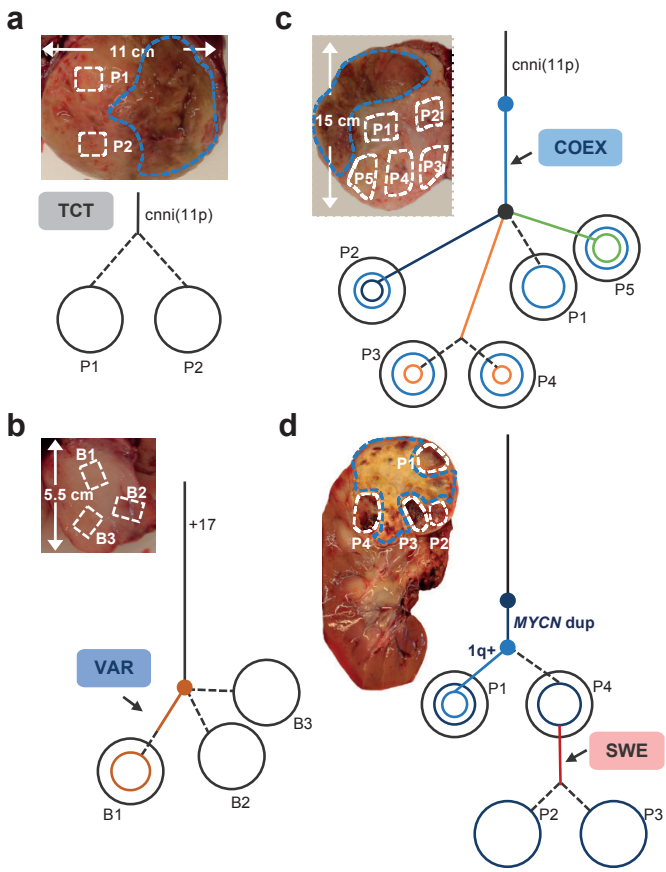


Figure 1: A) Depicts the scenario with no detectable ITH, which we term Tumor Cell Twinning (TCT). B) shows the VAR pattern with a single sample containing a private subclone. C) illustrates the COEX pattern where subclones coexist over large anatomic distances. D) is an example of the SWE pattern where a subclone from one sample completely takes over in another sample from the same primary tumor.

By analyzing branches in the empirical evolutionary ideograms, we demonstrate that some of these patterns are preceded by specific types of mutations. The COEX pattern is typically preceded by whole chromosome gains and losses, while the SWE pattern preferentially follows structural chromosomal aberrations or driver gene mutations, VAR was preceded by a very low number of aberrations, without a specific profile. In contrast the regions showing the highest number of aberrations in the context of EXP all had inactivating mutations of *TP53* or amplification of the p53 inhibitor gene *MDM2*. Furthermore, we demonstrate that *TP53* mutations typically are located in regions with the specific morphological pattern of anaplasia. By knocking out *TP53* in immortalized fibroblasts using *CRISPR-Cas9* we recapitulated the histological hallmarks of anaplasia as well as relatively high levels of branching evolution *in vivo*

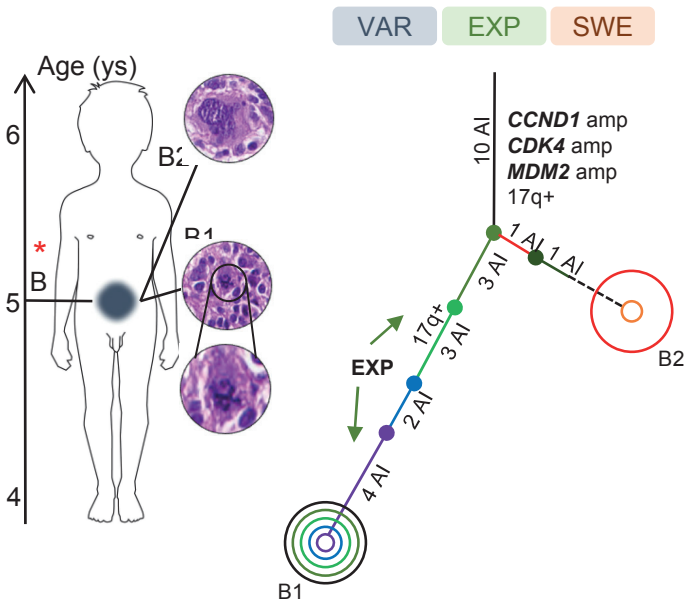


Figure 2: Example of the EXP pattern, where a large number of private genetic aberrations are detected in a single sample, preceded by amplification of the MDM2 gene.

From a clinical point of view, we further reinforce findings from Paper IV that variation of proposed genetic biomarkers is common within a patient's tumor lineage, such as 1p and 11q loss in NB and *TP53* mutation, *MYCN* mutation / gain and 1q gain in WT. Notably, the genetic aberration of longest clinical use as a predictive biomarker –*MYCN* amplification – showed no variation within patients; it was invariably present in the trunks of the evolutionary ideograms. Finally, we show that presence of either the SWE and / or EXP evolutionary pattern predicts inferior survival. While the former showed a strong positive correlation to

established clinical and genetic risk factors, the latter predicted event-free and overall survival in a fashion independent of established risk criteria.

Discussion

The following text highlights some of the most interesting points of discussion from the five papers included in this thesis. For a discussion of minor points, please see the discussion section in each of the papers.

Did we really disprove the autocatalytic theory of cancer?

Since the publication of Paper I, several groups have published papers studying overlapping phenomena. It is especially interesting to put the findings in Paper I in the context of the results from a recent study by Nicholson et al[136], where they show that specific chromosome copy number alterations can induce chromosomal instability by mitotic missegregation. Specifically, they show that DLD1 colon cancer cells with +7 and +13 have an increase of lagging chromosomes at anaphase, compared to non-aneuploid DLD1 cells. They also showed concordant results for this in amniocytes with +13. As it is known that at least some of the anaphase lagging chromosomes do in fact segregate properly in the end[46,47] the rate reported by Nicholson et al might represent an upper bound on missegregation rate in the cell lines assayed. They also combine the cytokinesis block assay[137] with FISH in order to validate their findings of anaphase-lagging. In that analysis, they intriguingly show that there was a chromosome specific increase in missegregation rate, where DLD1 +7 cells showed an increase in missegregation for chromosome 7 and DLD1 +13 displayed increased chromosomal instability for chromosomes 7 and 13. This chromosome specific effect was also seen in amniocytes, where the +13 cells showed an increased rate of missegregation of chromosome 13.

Nicholson et al. claim that one reason for the low rates of aneuploidy induced missegregation in Paper I was the fact that we measured somatic aneuploidy as a marker of CIN, and thus would miss daughter cells with aneuploidy that were eliminated before entering interphase. However, recent functional work have shown that cells that were pharmacologically treated to missegregate chromosomes were still proliferating despite extra copies of whole chromosomes even in the presence of functional *TP53*[138], while only small subsets arrest at G1[138,139]. These recent findings argue that somatic aneuploidy can in fact be used as a proxy for chromosomal instability. The presence of specific aneuploidies in non-neoplastic tissue *in vivo*, for instance +7 in osteoarthritis[140], also argues

against the fact that aneuploidy in the context of non-transformed cells invariably leads to cell cycle arrest. The current state of affairs suggest that our Paper I established the lower bound of chromosomal instability in trisomic cells, while Nicholson established the upper bound. Still neither of the papers show values of chromosomal variation or missegregation in aneuploid non-neoplastic cells that are at the levels observed in prototypical *in vitro* models of chromosomally unstable cancer.

Low-level somatic mosaicism – did we distinguish signal from noise?

Regarding the results from papers II and III, it is pertinent to discuss some of the issues with detecting genetic aberrations present in only a small subset of cells. The detection of the mosaic +8 of meiotic origin in paper II was only possible because of the fact that a cell that carried this genetic aberration underwent the monoclonal expansion inherent in tumorigenesis. Had this aberration not been monoclonally amplified it would never have been detected. Thus to completely rule out somatic mosaicism in, for instance, an organ one would need to assay every single cell that constitutes that organ, and this is of course far from possible today. Hence, it is quite possible that there exists somatic mosaicism at very low frequencies in fetuses and newborns, as the lower detection limit for the method employed in Paper III is approximately a 10% cellular prevalence[18]. This also makes it virtually impossible to answer the question whether the level of somatic mosaicism seen in solid organs[141,142] in adults arise as *de novo* events or simply is the result of expansion of pre-existing clones with low cellular prevalence already in the embryo.

If detection of somatic mosaicism in bulk tissue samples requires a cellular prevalence of at least 10%, it also follows that any aberration that is actually detected as mosaic either conveys a cellular fitness advantage or is fitness neutral but emerged already in early embryogenesis. A confounder here is of course tissue organization, and how stem cell populations contribute to repopulation of organs over time. When and how the presence of somatic mosaicism constitutes a premalignant lesion is an extremely interesting venue for further research using high resolution genomic methods. Using traditional cytogenetic methods, Petersson et al demonstrated the presence of clonal chromosomal aberrations in non-malignant breast tissues[143–145]. Along the same line, Forsberg et al[146] recently showed that histologically normal breast tissue adjacent to breast cancers contains somatic genetic aberrations known to affect the disease course in breast malignancies. Similar results have also been published in the case of prostate cancer[147]. The selection pressures determining whether abnormal clonal expansions in normal tissue occur or not is a field completely unexplored so far.

Does evolutionary genetics have a role in cancer research?

Intratumor heterogeneity has primarily been studied in adult cancers, with some notable exceptions. Chen et al[113] performed whole genome sequencing on single samples from rhabdomyosarcoma patients and tried to deconvolve the clonal composition and infer ordering of mutational events. As the data types and sampling strategies are different there are some pitfalls in comparing our data from papers IV & V with the results from Chen et al. In general, the inter-study variation between sampling strategies, genomic assays and phylogenetic inference-tools hampers cross-study comparisions of ITH. This will hopefully be ameliorated as the field matures, as evident by a recent consensus statement on measures of ITH and tumor evolution[148]. This consensus statement proposes an Evo-index, that includes estimates of both spatial and temporal changes in ITH. The authors also suggest an Eco-Index to classify the tumor microenvironment.

In neuroblastoma, a number of studies have looked at tumor evolution between relapse and diagnostic samples[149–151], with the goal of understanding the clonal dynamics of relapse. They found recurrent, relapse specific mutations targeting the *RAS* pathway and in genes important for cell-cell interactions. There was also significant spatio-temporal heterogeneity for *ALK* mutations. While identifying novel recurrent mutations none of these studies focused on the evolutionary dynamics per se, as we did in papers IV and V.

When it comes to WT, a few papers have looked at different aspect of tumor evolution. Spreafico et al[152] found that mutations in *SIX1* and *DROSHA* can be spatially and temporally heterogeneous events and that co-occurrence of these mutations might be positively selected for in the relapse setting. Cresswell et al[94] used multi-sampled primary tumors analyzed with whole genome genotyping arrays to map evolutionary dynamics in primary WTs. In line with our results from Papers IV and V, they found a significant ITH for 1q gain, and confirmed that multiregional sampling was necessary for determining the 1q status of a tumor. They also found that LOH of 11p15 was invariably a truncal aberration. Interestingly, for one case, they detected potentially differential effects of chemotherapy between different tumor clones.

In the adult setting, there is a multitude of studies of ITH in a wide variety of tumor types. The largest study on ITH to date is the TRACERx study, which is a prospective study focusing on ITH and clinical course in non small cell lung cancer patients in Great Britain[153]. The TRACERx consortium recently published an intermediate analysis of the first 100 patients[154], showing that there exists both a group of driver mutations that are almost invariably present in the trunks of tumor phylogenies, but also a group of heterogeneous drivers, including mutations in *PIK3CA* and *NF1*, that occur later in the evolutionary history of the tumor.

Furthermore, they discovered that copy number ITH, rather than ITH at the sequence level, was associated with an increased risk of recurrence or death, in both univariate and multivariate analyses. These findings are also in line with our data from paper V where we find that the number of non-truncal allelic imbalances independently predicted relapse at multivariate analysis. However, only stage and the number of evolutionary patterns present in the primary tumor remained as significant predictors when overall survival was analyzed (Paper V, Supplementary Figure 17a-b).

Where do we go from here?

In the years to come, a central theme in cancer biology and clinical oncology is understanding the role of evolution when a monoclonal mutated cell population transits from being part of benign somatic mosaicism, to being premalignant, to finally become invasive and present as overt cancer. What forces determine if a cell lineage regresses or continues along this road to malignancy? With increased knowledge of the dynamics of intratumor evolution it seems plausible that therapies exploiting the principles underlying this evolution could benefit patients, perhaps using adaptive therapy[155] or by evolutionary trapping of highly heterogeneous populations[156]. However, in order to fully elucidate these very complex mechanisms, and employ that knowledge in the clinical setting, advances in both sequencing technology and in software processing tools are needed. We also need to improve evolutionary models that are applicable to cancer. A very interesting venue for further research is the combination of mathematical modelling with more traditional cancer research, for instance using quantitative models to understand the evolutionary dynamics of tumor response to combination therapies[157] or to examine how spatial constraints shape patterns of ITH[158].

In the last two years several studies have been published that deal with targeted sequencing of known oncogenes in tumor samples to inform clinical management of various forms of cancers[159]. However, this approach has generally not led to increased survival in patients or improved patient quality of life[160,161]. Using targeted sequencing of known oncogenes to inform treatment in the setting of a highly advanced, metastatic cancer is of course a daunting task. Even if the logistics are perfectly set up to do the sequencing in a clinically relevant time-frame, there is no guarantee that 1) the sequencing detects any alterations in genes or pathways that we know are amendable to targeted therapy and 2) there exists a currently recruiting trial that includes drugs targeting that specific genetic aberration. This is hard even at major cancer centers, as evident by the recent publication of the first 10 000 patients that underwent clinical sequencing using the MSK-IMPACT platform at the Memorial Sloan Kettering Cancer Center, where about 11% of the patients were enrolled into clinical trials based on the

sequencing results[162]. It is probable that the limiting factor for targeted therapies is not the sequencing in itself, but our inability to interpret sequence data and act upon it. Even if we had perfect knowledge of the driver mutations in a cancer and had therapies to target them, the specter of intratumor evolution would still be an issue, as precise targeted therapy always carries the risk of selection of resistant subclones that spawn the next relapse.

Another emerging therapeutic option in cancer is immunotherapy, exemplified by the use of monoclonal antibodies such as the programmed cell death protein 1 (PD1) targeting antibody nivolumab in the treatment of melanoma, lung cancer and some renal cell carcinomas[163–165]. The efficacy of this drug seem to be related to the number of neo-epitopes that tumor cells present to the host immune system and on the mutational burden of the malignant cells[166]. In addition to this, certain types of mutations are more immunogenic than others[167]. In childhood solid tumors, with their quiescent genome at the sequence level, the predicted level of neo-antigens is smaller than in adult cancers[168]. This indicates that antibodies targeting PD1 would be less effective in childhood cancers than in adult ones. A notable exception to this might be microsatellite-unstable tumors, which typically have a high mutational burden, regardless of patient age.

To further complicate matters, the immune response generated by a neo-epitope depends, at least in part, of whether the mutation that generated it is clonal or subclonal in the tumor[169]. Investigations have also shown that there is a complex interplay between immune response and tumor evolution[170,171]. As with all targeted therapy, current immunotherapies may select for the outgrowth of resistant clones[172].

Thus, both therapies targeting activated oncogenes and immunotherapies, at least in their current form, seem to be hampered by the rules of evolution that govern tumorigenesis. Whether increased understanding of tumor evolution will lead to improved patient outcome remains to be seen. If we can change the rules of the evolutionary game between the patient and her tumor, rather than keep playing this ancient game, it might well be that we can turn cancer into a chronic disease or even cure it all together.

Conclusions

To summarize, the following broad conclusions can be drawn from the work presented in this thesis:

- Extra copies of whole chromosomes do not, in general, generate chromosomal instability at the level seen in cancer cells
- In rare cases the monoclonal expansion of tumorigenesis may reveal low-grade somatic mosaicism
- Newborns and fetuses have significantly lower levels of somatic mosaicism at the copy number level than adults
- The fetal liver is a euploid organ and fetal hepatocytes are no more aneuploid than other somatic cells in the fetus
- ITH is a common phenomenon in malignant pediatric tumors and may have prognostic impact
- Childhood solid tumors may contain multiple evolutionary patterns and the repertoire of such patterns correlate with patient outcome
- Different evolutionary patterns seem to be preceded by specific classes of mutations
- Proposed genetic biomarkers in childhood cancers may be missed when only single tumor samples are analyzed
- Genome profiles of primary tumors correspond poorly with later metastatic relapses in the same patient

Populärvetenskaplig Sammanfattning

Våra kroppar består av miljarder celler som alla kommer från ett och samma befruktade ägg. Denna första cell innehåller också originalritningen till vår arvsmassa. För att arvsmassans information skall kunna föras vidare till alla kroppens celler, måste dess DNA-molekyler dela sig så att dottercellerna kan få var sin komplett kopia. Trots de intrikata verktyg som evolutionen gett oss för att se till att denna process går felfritt så kommer förändringar i arvsmassan att introduceras vid varje celldelning. De senaste åren har den tekniska utvecklingen inom DNA-sekvensering gjort det möjligt för forskare att uppskatta hur ofta sådana mutationer uppstår och upptäcka dem även i frisk vävnad, där de endast finns i ett litet antal celler. Samma tekniska utveckling har också gjort det möjligt att studera hur delar av enskilda tumörer skiljer sig genetiskt från varandra. Därmed kan man nu för första gången börja förstå de evolutionära regler som styr tumörutveckling.

Min avhandling baseras på detaljerad analys av hur olika områden i kroppen hos foster och nyfödda skiljer sig åt genetiskt. Variationen jämförs sedan med hur olika delar av tumörer hos barn skiljer sig åt. Vi visar att foster har signifikant lägre frekvens av genetiska skillnader mellan organ än vad vuxna har. Detta faktum stämmer bra överens med två olika modeller av hur genetisk variation inom en person uppkommer; antingen finns den i oerhört låg frekvens (under vår detektionsgräns) redan hos nyfödda eller så är det ett fenomen som uteslutande beror på individens ålder. Vi visar att det förstnämnda kan förklara vissa märkliga fynd vid genetisk analys av tumörceller såsom förekomst av en kromosomkopia med en genprofil som saknas i resten av patientens celler. I motsats till situationen i normal vävnad fann vi att elakartade tumörer hos barn uppvisar en mycket stor genetisk variation inom sig. Än större var skillnaden mellan ursprungliga tumörer och metastaser vid återfall i tumörsjukdom. Vi upptäckte fyra olika evolutionära mönster som kan förekomma samtidigt inom en och samma tumör. Två av dessa avspeglade en omfattande förmåga hos cancercellen att förändra sin arvsmassa.

Just dessa två mönster var också tydligt kopplade till tumörformer med hög risk för återfall och död. Vi fördjupade oss även i hur sådan genetisk flexibilitet i cancerceller kan uppstå. Närmare bestämt testade vi en kontroversiell hypotes om hur genetiska förändringar drivs fram och får normala celler att bli cancerceller. Hypotesen hävdar att den avgörande faktorn bakom om en cell blir elakartad eller

inte är ett felaktigt antal kromosomer. Detta felaktiga kromosomtal skulle i sin tur leda till att cellen får allt svårare att hålla reda på sina kromosomer och därför riskerar att få ett allt mer avvikande kromosomtal. Genom att studera celler från individer födda med ett avvikande antal kromosomer visar vi att dessa celler inte har svårare att hålla ordning på sin kromosomuppsättning än vad celler med rätt antal har. Vi kunde därmed motbevisa en teori som florerat länge i den vetenskapliga litteraturen, trots att den saknat grund i experimentella data.

Acknowledgements

En lista över personer som jag har att tacka för stöd, direkt eller indirekt, rörande min avhandling blir av nödvändighet summarisk och det finns en definitiv risk att jag missat att nämna någon som borde funnits med nedan. Med det sagt vill jag tacka följande personer:

Min enastående handledare **David Gisselsson Nord**, vars entusiasm för forskning verkligen är smittsam, för att du introducerade mig till cancerforskningens värld och framförallt för att du lät mig bli en del av din forskargrupp. All empiri talar för att det är omöjligt att lämna ett möte med Dig utan att tycka att forskning är det roligaste en människa kan ägna sig åt.

Mina fantastiska bihandledare, **Jenny Karlsson** och **Linda Holmqvist-Mengelbier**, för att ni sett till att jag blivit en inte helt oäven forskare, och för att alla gånger jag fått inkräkta i ert lilla rum (Prio 1 vid nästa flytt borde vara att ni får ett större rum!) med mer eller mindre (oftast mindre) genomtänkta frågor!

Anette Welin, utan Dig hade hela avdelningen för klinisk genetik slutat fungera direkt! **Yuesheng** för hjälp med karyotypering. Seniora forskare och forskningsgruppsledare på klinisk genetik, **Nisse**, **Felix**, **Thoas**, **Bertil**, **Figge**, **Karolin**, **Marcus**, **Anna** och **Kajsa** för en extremt trevlig och tillåtande forskningsmiljö. **Linda M**, **Jenny N**, **Andrea**, **Marianne**, **Calle H**, **Helena S** och **Carro** för att jag ibland inte trott mina öron under lunchdiskussionerna och för all hjälp.

Doktoranderna, nuvarande (**Jakob**, **Karim**, **Ram**, **Naveen**, **Mattias**, **Mia**, **Pablo**, **Niklas** och **Sofia**) samt före detta (**Linda O**, **Setareh**, **Elsa**, **Axel**, **Charles**, **Daniel**, **Nils** och **Gisela**) för en fantastisk gemenskap. Särskilt tack till dom av er som jag delat basturummet med.

Alla postdocs vid avdelningen, **Calle S**, **Christina**, **Eleanor**, **Henrik**, **Helena Å**, **Kristina**, **Louise**, **Maria**, **Marion**, **Minjun** och **Rebeqa**, för en fantastiskt trevlig arbetsmiljö.

Läkarkollektivet på klinisk genetik, för att stämningen på kliniken är så bra att man ibland inte tror att man arbetar i offentlig sjukvård.

Julia, För kulbo-skapet och för att du introducerade mig till Parks and Recreation och tusen andra saker. **Henrik** för alla bullarna (Även om jag tror att vi är bull-

even nu!) **David**, för att ditt egenpåtagna epitet ”bästa killen” nästan alltid stämmer. **Mimmi**, för att du ser till att vår umgängeskrets inte bara pratar medicin. **Clemens** för tågresorna, speciellt den där vi diskuterade vad som utgör kristet beteende. **Oskar** för att du är irriterande vettig (och numera har byxor på sig!) och för att du gick med på att vara toastmaster. **Eric**, för all fotodokumentation och för din förmåga att lämna knäskålarna hemma. **Ola**, för din gästfrihet och för att du är den enda hyresvärd jag haft som kan gastroskopera. **Alexander**, för åtskilliga lektioner i vad som utgör historiska nödvändigheter (det är oftast att stanna till stängningsdags på svartklubbar) och **Johanna**, för ditt prisvinnande göteborgska manér och för alla gånger jag fått hänga hos er. **Viktor**, för bakverken (trots att du alltid tycker att dom är misslyckade!) och ett av de finaste sms jag någonsin fått. **Marcus**, för ständiga påminnelser om att medicin i grunden borde räknas till humaniora. **Nils** och **Ida**, för att ni är bland dom snällaste människor jag träffat och för den privata fritidsgård ni drev i Ronneby. **Pontus** och **Andrée**, för trams, folköl och konserter.

Jenny, Kristian och **Peder**, för att ni är mina syskon.

Mamma, för att du är den starkaste och snällaste person jag vet. **Pappa**, trots att du gick bort innan jag påbörjade min forskarutbildning så har lärdomar jag fått av Dig varit centrala under avhandlingsarbetet och jag är övertygad om att Du är stolt över mig. Om jag har lyckats med saker i livet har det varit för att jag haft er två som föräldrar.

References

1. Kang, S., Kang, M. S., Ryu, E. & Myung, K. Eukaryotic DNA replication: Orchestrated action of multi-subunit protein complexes. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* 0–1 (2017). doi:10.1016/j.mrfmmm.2017.04.002
2. Chen, C., Qi, H., Shen, Y., Pickrell, J. & Przeworski, M. Contrasting determinants of mutation rates in germline and soma. *bioRxiv* 117325 (2017). doi:10.1101/117325
3. Rahbari, R. *et al.* Timing, rates and spectra of human germline mutation. *Nat. Genet.* **48,** 1–11 (2015).
4. Acuna-Hidalgo, R. *et al.* Post-zygotic Point Mutations Are an Underrecognized Source of De Novo Genomic Variation. *Am. J. Hum. Genet.* **97,** 67–74 (2015).
5. Kong, A. *et al.* Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488,** 471–475 (2012).
6. Jonsson, H. *et al.* Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* (2017). doi:10.1038/nature24018
7. Consortium:, T. G. of the N. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46,** 1–95 (2014).
8. Giannoulatou, E. *et al.* Contributions of intrinsic mutation rate and selfish selection to levels of de novo HRAS mutations in the paternal germline. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 20152–7 (2013).
9. Goriely, A., McVean, G., Röjmyr, M., Ingemarsson, B. & Wilkie, A. O. M. Evidence for Selective Advantage of Pathogenic FGFR2 Mutations in the Male Germ Line. *Science (80-.).* **301,** 643–646 (2003).
10. Goriely, A. & Wilkie, A. O. M. Paternal age effect mutations and selfish spermatogonial selection: Causes and consequences for human disease. *Am. J. Hum. Genet.* **90,** 175–200 (2012).
11. Wong, W. S. W. *et al.* New observations on maternal age effect on germline de novo mutations. *Nat. Commun.* **7,** 10486 (2016).
12. Kuliev, A., Zlatopolsky, Z., Kirillova, I., Spivakova, J. & Cieslak Janzen, J. Meiosis errors in over 20,000 oocytes studied in the practice of preimplantation aneuploidy testing. *Reprod. Biomed. Online* **22,** 2–8 (2011).
13. Fragouli, E. *et al.* The cytogenetics of polar bodies: Insights into female meiosis and the diagnosis of aneuploidy. *Mol. Hum. Reprod.* **17,** 286–295 (2011).

14. Itsara, A. *et al.* De novo rates and selection of large copy number variation De novo rates and selection of large copy number variation. *Genome Res.* 1469–1481 (2010). doi:10.1101/gr.107680.110

15. Kloosterman, W. P. *et al.* Characteristics of de novo structural changes in the human genome. *Genome Res.* **25,** 792–801 (2015).

16. Milholland, B. *et al.* Differences between germline and somatic mutation rates in humans and mice. *Nat. Commun.* **8,** 15183 (2017).

17. Forsberg, L. a *et al.* Age-related somatic structural changes in the nuclear genome of human blood cells. *Am. J. Hum. Genet.* **90,** 217–28 (2012).

18. Laurie, C. C. *et al.* Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* **44,** 642–50 (2012).

19. Jacobs, K. B. *et al.* Detectable clonal mosaicism and its relationship to aging and cancer. *Nat. Genet.* **44,** 651–8 (2012).

20. Genovese, G. *et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N. Engl. J. Med.* 141126135947008 (2014). doi:10.1056/NEJMoa1409405

21. Jaiswal, S. *et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N. Engl. J. Med.* 1–11 (2014). doi:10.1056/NEJMoa1408617

22. Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20,** 1472–1478 (2014).

23. Jaiswal, S. *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N. Engl. J. Med.* **377,** 111–121 (2017).

24. Forsberg, L. a *et al.* Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat. Genet.* **46,** 624–8 (2014).

25. Bonnefond, A. *et al.* Association between large detectable clonal mosaicism and type 2 diabetes with vascular complications. *Nat. Genet.* **45,** 1040–1043 (2013).

26. Gibson, C. J. *et al.* Clonal Hematopoiesis Associated With Adverse Outcomes After Autologous Stem-Cell Transplantation for Lymphoma. *J. Clin. Oncol.* **35,** 1598–1605 (2017).

27. Acuna-Hidalgo, R. *et al.* Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *Am. J. Hum. Genet.* **101,** 50–64 (2017).

28. McConnell, M. J. *et al.* Mosaic Copy Number Variation In Human Neurons. *Science (80-.).* 632–637 (2013).

29. McConnell, M. J. *et al.* Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. *Science (80-.).* **356,** eaal1641 (2017).

30. Anglesio, M. S. *et al.* Cancer-Associated Mutations in Endometriosis without Cancer. *N. Engl. J. Med.* **376,** 1835–1848 (2017).

31. Krimmel, J. D. *et al.* Ultra-deep sequencing detects ovarian cancer cells in peritoneal fluid and reveals somatic *TP53* mutations in noncancerous tissues. *Proc. Natl. Acad. Sci.* **113,** 6005–6010 (2016).

32. Martincorena, I. *et al.* High burden and pervasive positive selection of somatic mutations in normal human skin. *Science (80-.).* **348,** 880–886 (2015).

33. Hajdu, S. I., Vadmal, M. & Tang, P. A note from history: Landmarks in history of cancer, part 7. *Cancer* **121,** 2480–2513 (2015).

34. Boveri, T. Concerning the Origin of Malignant Tumours by Theodor Boveri. Translated and annotated by Henry Harris. *J. Cell Sci.* **121,** 1 LP-84 (2007).

35. Nowell, P. C. Review series personal perspective Discovery of the Philadelphia chromosome : a personal perspective. *J. Clin. Invest.* **117,** 2033–2035 (2007).

36. Rowley, J. A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. *Nature* **243,** 290–293 (1973).

37. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100,** 57–70 (2000).

38. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).

39. Painter, T. Studies in mammalian spermatogenesis, II. The spermatogenesis of man. *J Exp Zool* **37,** 291–338 (1923).

40. Lejeune, J., Gauthier, M. & Turpin, R. Les chromosomes humains en culture de tissus. *CR Hebd Seances Acad Sci* **248,** 602–603 (1959).

41. Holland, A. J. & Cleveland, D. W. Losing balance: the origin and impact of aneuploidy in cancer. *EMBO Rep.* **13,** 501–14 (2012).

42. van Jaarsveld, R. H. & Kops, G. J. P. L. Difference Makers: Chromosomal Instability versus Aneuploidy in Cancer. *Trends in Cancer* **2,** 561–571 (2016).

43. Burrell, R. a. *et al.* Replication stress links structural and numerical cancer chromosomal instability. *Nature* **494,** 492–496 (2013).

44. Crasta, K. *et al.* DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482,** 53–8 (2012).

45. Gisselsson, D. *et al.* Generation of trisomies in cancer cells by multipolar mitosis and incomplete cytokinesis. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 20489–93 (2010).

46. Thompson, S. L. & Compton, D. a. Chromosome missegregation in human cells arises through specific types of kinetochore-microtubule attachment errors. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 17974–8 (2011).

47. Thompson, S. L. & Compton, D. a. Examining the link between chromosomal instability and aneuploidy in human cells. *J. Cell Biol.* **180,** 665–72 (2008).

48. Ganem, N. J., Godinho, S. a & Pellman, D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460,** 278–82 (2009).

49. Silkworth, W. T., Nardi, I. K., Scholl, L. M. & Cimini, D. Multipolar spindle pole coalescence is a major source of kinetochore mis-attachment and chromosome mis-segregation in cancer cells. *PLoS One* **4,** e6564 (2009).

50. Cosenza, M. R. & Krämer, A. Centrosome amplification, chromosomal instability and cancer: mechanistic, clinical and therapeutic issues. *Chromosom. Res.* **24,** 105–126 (2016).

51. Gisselsson, D. *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.* **98,** 12683–12688 (2001).

52. Gisselsson, D. *et al.* Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5357–62 (2000).

53. Hatch, E. M., Fischer, A. H., Deerinck, T. J. & Hetzer, M. W. Catastrophic Nuclear Envelope Collapse in Cancer Cell Micronuclei. *Cell* **154,** 47–60 (2013).

54. Sheltzer, J. M. *et al.* Aneuploidy drives genomic instability in yeast. *Science* **333,** 1026–30 (2011).

55. Oromendia, A. B., Dodgson, S. E. & Amon, A. Aneuploidy causes proteotoxic stress in yeast. *Genes Dev.* 2696–2708 (2012). doi:10.1101/gad.207407.112

56. Sheltzer, J. M., Torres, E. M., Dunham, M. J. & Amon, A. Transcriptional consequences of aneuploidy. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 12644–9 (2012).

57. Torres, E. M. *et al.* Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast. *Science (80-.).* **317,** 916–924 (2007).

58. Williams, B. R. *et al.* Aneuploidy Affects Proliferation and Spontaneous Immortalization in Mammalian Cells. *Science (80-.).* **322,** 703–709 (2008).

59. Segal, D. J. & Mccoy, E. E. Studies on Down's Syndrome in Tissue Culture. *J. Cell. Physiol.* **83,** 85–90 (1974).

60. Stingele, S. *et al.* Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Mol. Syst. Biol.* **8,** 608 (2012).

61. Sears, D. D., Hegemann, J. H. & Hieter, P. Meiotic recombination and segregation of human-derived artificial chromosomes in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 5296–300 (1992).

62. Passerini, V. *et al.* The presence of extra chromosomes leads to genomic instability. *Nat. Commun.* **7,** 10754 (2016).

63. Sheltzer, J. M. *et al.* Single-chromosome Gains Commonly Function as Tumor Suppressors. *Cancer Cell* 1–16 (2017). doi:10.1016/j.ccell.2016.12.004

64. Ganmore, I., Smooha, G. & Izraeli, S. Constitutional aneuploidy and cancer predisposition. *Hum. Mol. Genet.* **18,** R84-93 (2009).

65. Rutledge, S. D. *et al.* Selective advantage of trisomic human cells cultured in non-standard conditions. *Sci. Rep.* **6,** 22828 (2016).

66. Beach, R. R. *et al.* Aneuploidy Causes Non-genetic Individuality. *Cell* **169,** 229–242.e21 (2017).

67. Duesberg, P. & Rausch, C. Genetic instability of cancer cells is proportional to their degree of aneuploidy. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 13692–13697 (1998).

68. Nicholson, J. M. & Duesberg, P. On the karyotypic origin and evolution of cancer cells. *Cancer Genet. Cytogenet.* **194,** 96–110 (2009).

69. Nowell, P. C. The Clonal Evolution of Tumor Cell Populations. *Science (80-.).* **194,** 23–28 (1976).

70. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* 689–698 (2015). doi:10.1056/NEJMoa1500987

71. Zhang, J. *et al.* Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science (80-.).* **346,** 256–259 (2014).

72. Mengelbier, L. H. *et al.* Intratumoral genome diversity parallels progression and predicts outcome in pediatric cancer. *Nat. Commun.* **6,** 6125 (2015).

73. Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci. U. S. A.* (2013). doi:10.1073/pnas.1219747110

74. Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47,** 209–216 (2015).

75. Schwartz, R. & Schäffer, A. A. The evolution of tumour phylogenetics: principles and practice. *Nat. Rev. Genet.* **18,** 213–229 (2017).

76. Kostadinov, R., Maley, C. C. & Kuhner, M. K. Bulk Genotyping of Biopsies Can Create Spurious Evidence for Hetereogeneity in Mutation Content. *PLoS Comput. Biol.* **12,** 1–15 (2016).

77. Alves, J. M., Prieto, T. & Posada, D. Multiregional Tumor Trees Are Not Phylogenies. *Trends in Cancer* **3,** 546–550 (2017).

78. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science (80-.).* **339,** 1546–1558 (2013).

79. Vujanić, G. M. & Sandstedt, B. The pathology of Wilms' tumour (nephroblastoma): the International Society of Paediatric Oncology approach. *J. Clin. Pathol.* **63,** 102–9 (2010).

80. Gadd, S. *et al.* Clinically Relevant Subsets Identified by Gene Expression Patterns Support a Revised Ontogenic Model of Wilms Tumor: A Children's Oncology Group Study. *Neoplasia* **14,** 742–IN21 (2012).

81. Pietras, W. Advances and Changes in the Treatment of Children with Nephroblastoma. 809–820 (2012).

82. Pastore, G. *et al.* Malignant renal tumours incidence and survival in European children (1978-1997): Report from the Automated Childhood Cancer Information System project. *Eur. J. Cancer* **42,** 2103–2114 (2006).

83. Wåhlin, N., Christofferson, R., Pal, N. & Kullendorff, C.-M. Wilms ' tumör hos barn – avsevärt förbättrad överlevnad. *Lakartidningen* 1440–1443 (2006).

84. Gatta, G. *et al.* Childhood cancer survival in Europe 1999-2007: Results of EUROCARE-5-a population-based study. *Lancet Oncol.* **15,** 35–47 (2014).

85. Brok, J., Treger, T. D., Gooskens, S. L., van den Heuvel-Eibrink, M. M. & Pritchard-Jones, K. Biology and treatment of renal tumours in childhood. *Eur. J. Cancer* **68,** 179–195 (2016).

86. Spreafico, F. *et al.* Treatment of relapsed Wilms tumors: lessons learned. *Expert Rev. Anticancer Ther.* **9,** 1807–1815 (2009).

87. Wegert, J. *et al.* Mutations in the SIX1/2 Pathway and the DROSHA/DGCR8 miRNA Microprocessor Complex Underlie High-Risk Blastemal Type Wilms Tumors. *Cancer Cell* **27,** 298–311 (2015).

88. Walz, A. L. *et al.* Recurrent DGCR8, DROSHA, and SIX Homeodomain Mutations in Favorable Histology Wilms Tumors. *Cancer Cell* **27,** 286–297 (2015).

89. Williams, R. D. *et al.* Multiple mechanisms of MYCN dysregulation in Wilms tumour. *Oncotarget* **6,** 7232–7243 (2015).

90. Satoh, Y. *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* **95,** 541–7 (2006).

91. Grundy, P. E. *et al.* Loss of heterozygosity for chromosomes 1p and 16q is an adverse prognostic factor in favorable-histology Wilms tumor: A report from the National Wilms Tumor Study Group. *J. Clin. Oncol.* **23,** 7312–7321 (2005).

92. Maschietto, M. *et al.* TP53 mutational status is a potential marker for risk stratification in Wilms tumour with diffuse anaplasia. *PLoS One* **9,** 1–8 (2014).

93. Chagtai, T. *et al.* Gain of 1q as a prognostic biomarker in Wilms Tumors (WTs) treated with preoperative chemotherapy in the International Society of Paediatric Oncology (SIOP) WT 2001 trial: A SIOP renal tumours biology consortium study. *J. Clin. Oncol.* **34,** 3195–3203 (2016).

94. Cresswell, G. D. *et al.* Intra-Tumor Genetic Heterogeneity in Wilms Tumor: Clonal Evolution and Clinical Implications. *EBioMedicine* **9,** 120–129 (2016).

95. Heck, J. E., Ritz, B., Hung, R. J., Hashibe, M. & Boffetta, P. The epidemiology of neuroblastoma: A review. *Paediatr. Perinat. Epidemiol.* **23,** 125–143 (2009).

96. Whittle, S. B. *et al.* Overview and recent advances in the treatment of neuroblastoma. *Expert Rev. Anticancer Ther.* **17,** 369–386 (2017).

97. Oldridge, D. A. *et al.* Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. *Nature* **528,** 418–421 (2015).

98. Mossé, Y. P. *et al.* Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature* **455,** 930–935 (2008).

99. Janoueix-Lerosey, I. *et al.* Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature* **455,** 967–970 (2008).

100. Cohn, S. L. *et al.* The International Neuroblastoma Risk Group (INRG) classification system: An INRG task force report. *J. Clin. Oncol.* **27,** 289–297 (2009).

101. Monclair, T. *et al.* The International Neuroblastoma Risk Group (INRG) staging system: An INRG Task Force report. *J. Clin. Oncol.* **27,** 298–303 (2009).

102. Peifer, M. *et al.* Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **526,** 700–704 (2015).

103. Valentijn, L. J. *et al.* TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. *Nat. Genet.* **47,** 1411–1414 (2015).

104. Mosse, Y. P. *et al.* Safety and activity of crizotinib for paediatric patients with refractory solid tumours or anaplastic large-cell lymphoma: a Children's Oncology Group phase 1 consortium study. *Lancet Oncol* **14,** 472–480 (2013).

105. Stevens, M. C. G. *et al.* Treatment of nonmetastatic rhabdomyosarcoma in childhood and adolescence: Third study of the International Society of Paediatric Oncology-SIOP malignant mesenchymal tumor 89. *J. Clin. Oncol.* **23,** 2618–2628 (2005).

106. Breneman, J. C. *et al.* Prognostic factors and clinical outcomes in children and adolescents with metastatic rhabdomyosarcoma-a report from the intergroup rhabdomyosarcoma study IV. *J. Clin. Oncol.* **21,** 78–84 (2003).

107. Anderson, J. *et al.* Detection of the PAX3-FKHR fusion gene in paediatric rhabdomyosarcoma: a reproducible predictor of outcome? *Br. J. Cancer* **85,** 831–5 (2001).

108. Shern, J. F. *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.* **4,** 216–231 (2014).

109. Williamson, D. *et al.* Fusion gene-negative alveolar rhabdomyosarcoma is clinically and molecularly indistinguishable from embryonal rhabdomyosarcoma. *J. Clin. Oncol.* **28,** 2151–2158 (2010).

110. Seki, M. *et al.* Integrated genetic and epigenetic analysis defines novel molecular subgroups in rhabdomyosarcoma. *Nat. Commun.* **6,** 7557 (2015).

111. VSTB. Aktuella behandlingsprotokoll för barntumörer utom NHL och CNS. (2017).

112. Lychou, S. E., Gustafsson, G. G. & Ljungman, G. E. Higher rates of metastatic disease may explain the declining trend in Swedish paediatric rhabdomyosarcoma survival rates. *Acta Paediatr. Int. J. Paediatr.* **105,** 74–81 (2016).

113. Chen, L. *et al.* Clonality and Evolutionary History of Rhabdomyosarcoma. *PLoS Genet.* **11,** 1–25 (2015).

114. Gryder, B. E. *et al.* PAX3-FOXO1 establishes myogenic super enhancers and confers BET bromodomain vulnerability. *Cancer Discov.* **7,** 884–899 (2017).

115. Rasmussen, M. *et al.* Allele-specific copy number analysis of tumor samples with aneuploidy and tumor heterogeneity. *Genome Biol.* **12,** R108 (2011).

116. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci.* **107,** 16910–16915 (2010).

117. Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 20007–12 (2007).

118. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12,** R41 (2011).

119. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17,** 333–351 (2016).

120. Ye, H., Meehan, J., Tong, W. & Hong, H. Alignment of short reads: A crucial step for application of next-generation sequencing data in precision medicine. *Pharmaceutics* **7,** 523–541 (2015).

121. Lunter, G. & Goodson, M. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21,** 936–939 (2011).

122. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Prepr. arXiv* **0,** 3 (2013).

123. Picard Toolkit. at <http://broadinstitute.github.io/picard/>

124. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–498 (2011).

125. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

126. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31,** 213–9 (2013).

127. Narzisi, G. *et al.* Accurate de novo and transmitted indel detection in exome-capture data using microassembly. *Nat. Methods* 1–7 (2014). doi:10.1038/nmeth.3069

128. Davis, A. & Navin, N. E. Computing tumor trees from single cells. *Genome Biol.* **17,** 113 (2016).

129. Zafar, H., Tzen, A., Navin, N., Chen, K. & Nakhleh, L. SiFit: inferring tumor trees from single-cell sequencing data under finite-sites models. *Genome Biol.* **18,** 178 (2017).

130. Josephidou, M., Lynch, A. G. & Tavar??, S. multiSNV: a probabilistic approach for improving detection of somatic point mutations from multiple related tumour samples. *Nucleic Acids Res.* **43,** e61 (2015).

131. Paulsson, K. *et al.* Formation of trisomies and their parental origin in hyperdiploid childhood acute lymphoblastic leukemia. *Blood* **102,** 3010–3015 (2003).

132. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386,** 623–627 (1997).

133. Duncan, A. W. *et al.* Frequent aneuploidy among normal human hepatocytes. *Gastroenterology* **142,** 25–8 (2012).

134. Duncan, A. W. *et al.* The ploidy conveyor of mature hepatocytes as a source of genetic variation. *Nature* **467,** 707–10 (2010).

135. Knouse, K. a., Wu, J., Whittaker, C. a. & Amon, A. Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. *Proc. Natl. Acad. Sci.* **111,** 13409–13414 (2014).

136. Nicholson, J. M. *et al.* Chromosome mis-segregation and cytokinesis failure in trisomic human cells. *Elife* **4,** 1–23 (2015).

137. Fenech, M. The cytokinesis-block micronucleus technique and its application to genotoxicity studies in human populations. *Environ. Health Perspect.* **101,** 101–107 (1993).

138. Soto, M. *et al.* p53 Prohibits Propagation of Chromosome Segregation Errors that Produce Structural Aneuploidies. *Cell Rep.* **19,** 2423–2431 (2017).

139. Santaguida, S. *et al.* Chromosome Mis-segregation Generates Cell-Cycle-Arrested Cells with Complex Karyotypes that Are Eliminated by the Immune System. *Dev. Cell* **41,** 638–651.e5 (2017).

140. Broberg, K. *et al.* Clonal chromosome aberrations are present in vivo in synovia and osteophytes from patients with osteoarthritis. *Hum. Genet.* **101,** 295–298 (1997).

141. Piotrowski, A. *et al.* Somatic mosaicism for copy number variation in differentiated human tissues. *Hum. Mutat.* **29,** 1118–24 (2008).

142. Huallachain, M. O., Karczewski, K. J., Weissman, S. M. & Eckehart, A. Extensive genetic variation in somatic human tissues. *Proc. Natl. Acad. Sci.* **109,** 18018–18023 (2012).

143. Petersson, C. *et al.* Chromosome aberrations in prophylactic mastectomies from women belonging to breast cancer families. *Genes Chromosom. Cancer* **16,** 185–188 (1996).

144. Petersson, C. *et al.* Karyotypic abnormalities in fibroadenomas of the breast. *Int. J. Cancer* **70,** 282–286 (1997).

145. Lundin, C. P. *et al.* Cytogenetic changes in benign proliferative and nonproliferative lesions of the breast. *Cancer Genet. Cytogenet.* **107,** 118–120 (1998).

146. Forsberg, L. A. *et al.* Signatures of post-zygotic structural genetic aberrations in the cells of histologically normal breast tissue that can predispose to sporadic breast cancer. *Genome Res.* **25,** 1521–1535 (2015).

147. Cooper, C. S. *et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat. Genet.* **47,** 689–689 (2015).

148. Virginia, C. C. M. & Piper, G. Classifying the evolutionary and ecological features of neoplasms. *Nat. Rev. Cancer* (2017). doi:10.1038/nrc.2017.69

149. Schramm, A. *et al.* Mutational dynamics between primary and relapse neuroblastomas. *Nat. Genet.* **47,** 872–877 (2015).

150. Eleveld, T. F. *et al.* Relapsed neuroblastomas show frequent RAS-MAPK pathway mutations. *Nat. Genet.* **47,** 864–871 (2015).

151. Schleiermacher, G. *et al.* Emergence of new ALK mutations at relapse of neuroblastoma. *J. Clin. Oncol.* **32,** 2727–2734 (2014).

152. Spreafico, F. *et al.* Chromosomal anomalies at 1q, 3, 16q, and mutations of SIX1 and DROSHA genes underlie Wilms tumor recurrences. *Oncotarget* **7,** (2016).

153. Jamal-Hanjani, M. *et al.* Tracking Genomic Cancer Evolution for Precision Medicine: The Lung TRACERx Study. *PLoS Biol.* **12,** 1–7 (2014).

154. Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **376,** 2109–2121 (2017).

155. Gillies, R. J., Verduzco, D. & Gatenby, R. a. Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nat. Rev. Cancer* **12,** 487–93 (2012).

156. Chen, G. *et al.* Targeting the Adaptability of Heterogeneous Aneuploids. *Cell* **160,** 771–784 (2016).

157. Bozic, I. *et al.* Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife* **2,** e00747–e00747 (2013).

158. Waclaw, B. *et al.* A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* **525,** 261–264 (2015).

159. Tannock, I. F. & Hickman, J. A. Limits to Personalized Cancer Medicine. *N. Engl. J. Med.* **362,** 567–571 (2016).

160. Le Tourneau, C. *et al.* Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *Lancet Oncol.* **16,** 1324–1334 (2015).

161. Stockley, T. L. *et al.* Molecular profiling of advanced solid tumors and patient outcomes with genotype-matched clinical trials: the Princess Margaret IMPACT/COMPACT trial. *Genome Med.* **8,** 109 (2016).

162. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23,** 703–713 (2017).

163. Brahmer, J. *et al.* Nivolumab versus Docetaxel in Advanced Squamous-Cell Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **373,** 123–135 (2015).

164. Motzer, R. J. *et al.* Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N. Engl. J. Med.* **373,** 1803–1813 (2015).

165. Robert, C. *et al.* Nivolumab in Previously Untreated Melanoma without *BRAF* Mutation. *N. Engl. J. Med.* **372,** 320–330 (2015).

166. Rizvi, N. A. *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science (80-.).* **348,** 124–128 (2015).

167. Turajlic, S. *et al.* Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol.* **18,** 1009–1021 (2017).

168. Chang, T.-C. *et al.* The neoepitope landscape in pediatric cancers. *Genome Med.* **9,** 78 (2017).

169. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science (80-.).* **351,** 1463–1470 (2016).

170. Riaz, N. *et al.* Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* **0,** 1–16 (2017).

171. Jiménez-Sánchez, A. *et al.* Heterogeneous Tumor-Immune Microenvironments among Differentially Growing Metastases in an Ovarian Cancer Patient. *Cell* **170,** 927–938.e20 (2017).

172. Zaretsky, J. M. *et al.* Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *N. Engl. J. Med.* **375,** 819–829 (2016).