On the heterogeneity of tumor sequencing

Marlous Hoogstraat

Cover art and design: Marlous Hoogstraat

The research described in this thesis was performed at the University Medical Center Utrecht, within the framework of the Cancer, Stem Cells and Developmental Biology graduate school in Utrecht, the Netherlands

Financial support by Barcode for Life and Amgen B.V. for printing this thesis is gratefully acknowledged.

Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress
ISBN: 978-90-393-6236-5

On the heterogeneity of tumor sequencing

Over de heterogeniteit van tumor sequencen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op donderdag 30 oktober 2014 des middags te 2.30 uur

door

Marie-Louise Hoogstraat

geboren op 28 mei 1984
te Leidschendam

Promotoren: Prof. dr. E.E. Voest
Prof. dr. E.P.J.G. Cuppen

Heterogeneity (ˌhɛtɛrəʊdʒəˈniːətɪ), from the Greek heterogenēs, from heteros 'other' + genos 'a kind'.

1) The fact or state of being composed of unrelated or differing parts or elements
2) The fact or state of being not of the same kind or type

Table of Contents

1

Introduction

On the heterogeneity of tumor sequencing

There is one statement you cannot avoid in cancer research: “Cancer is a disease of the genome”. More than a century ago, even before the discovery of DNA, researchers observed that cancer cells show aberrant chromosomes and cell division[1]. It was discovered later that these aberrant chromosomes and other, smaller damages to the DNA are actually causing the cells to become cancerous, not just being the result thereof. In 1914, Theodor Boveri even described observations now recognized as cell cycle checkpoints, tumor suppressor genes and oncogenes[2]. These aberrations were both found as a result of exposure to toxins or to be naturally occurring. As Boveri already described, not all aberrations will lead to cancer. On the contrary, most cells will die after DNA damage. There are specific genetic events however that are beneficial to the cell, for instance because they induce an increased growth rate and cell proliferation, or enable evasion of natural cell death. Studies on tumor DNA since then revealed several of these events to be recurrent or even frequent in some cancer types, and targeting these events appears to be an attractive approach to cure cancer.

Current technology brings a variety of approaches to the detection of such genetic aberrations. Examples include paired end whole genome sequencing, ultra-deep targeted exon sequencing and genome-wide SNParrays. All techniques have their own benefits and drawbacks, provide other levels of biological information and require different analyses. This ‘data heterogeneity’ poses several difficulties for cancer research: which experimental approach is best suited to answer the research question, and how can these data be integrated with and compared to data from other sources?

On top of that, cancer research from the past decade has presented us with a new challenge: tumor heterogeneity. It was discovered that a single patient does not have a ‘single cancer’, but that tumor cells within a patient can show distinct molecular differences over time (e.g. between a primary tumor and a metastasis at relapse), but also at different tumor locations at the same time and even within a single neoplasm. These differences can be the cause of incomplete or only localized response to targeted therapies.

In this thesis, we use both multiple samples from single patients as well as multiple data types from single samples to study these levels of heterogeneity and its (predicted) effects on targeted treatment outcome. The aim is to increase our understanding of tumor evolution and functioning, which will ultimately lead to an increased knowledge on how to treat cancer.

Next-generation sequencing

What is next-generation sequencing?

In the 1970's, Fred Sanger was one of the first to develop a technique that enabled us to 'read' DNA, i.e. determine the sequence of nucleotides. Initially, this involved four separate reactions, one for each type of nucleotide. The products of these reactions were loaded onto a gel from which the final sequence, usually 200-400 nucleotides, could be determined by eye. The total process was laborious and time-consuming, especially if you realize a complete human genome contains approximately 3 billion nucleotides. Improvement of Sanger sequencing and development of new techniques has been ongoing ever since, accelerated greatly due to the "Human genome project". This project, started in 1990, set out to map the full genetic sequence of every human chromosome. The first draft was published 13 years later and had cost about 2.7 billion dollars. Important to note is that time and money was not only spent on generating the sequence data, but also on fitting the short sequences together to form larger contigs and finally full chromosomes. Refinement of the human genome is also still ongoing: version 38 was released recently.

Improvement of sequencing techniques has greatly increased the speed and amount of throughput, while at the same time reducing costs. The biggest advancement comes from massive parallelization, when multiple sequences from different samples can be sequenced at the same time: next-generation sequencing (NGS). Today, 96 samples can be sequenced simultaneously on one sequence run. Moreover, Illumina® (one of the main sequencer manufacturers) recently announced they have broken the "$1000 genome barrier", meaning they can sequence a person's full genome for less than 1000 US dollars within one day – that is disregarding of the price of the machine itself, which is estimated around 1 million dollars.

NGS and informatics

This progress would not have been possible without advancements in another field: that of computer technology. One genome of 3 billion nucleotides is impossible to analyze without the help of a computer. Furthermore, occasional miscalls occur during sequencing, which means the same stretch of DNA should be sequenced multiple times to distinguish between a sequence error and a correct call. Whole genomes are usually sequenced to at least 10-fold coverage on average, giving an output of over 30 billion basepairs, which is approximately a 30GB file of raw data. Most sequencers also report a quality value for each sequenced base that makes files twice as big. Next, the raw data have to be processed; the sequence output consists of short sequences of 50 to 300 basepairs that are meaningless in itself. Tens of thousands of short sequences

(reads) per sample are aligned to the human reference genome mentioned above. Finally, variation between the sequenced sample and the reference genome has to be assessed, commonly followed by comparisons between two or more samples which means that not only a large amount of disk space is needed to store the sequence files, but also a large amount of computer power to accommodate processing of sequence data.

As each type of sequencer has its own format of raw data and its own specific error profile, different algorithms are required to process the raw data. The specific biological questions that can be answered with, or formulated through, the downstream analyses of NGS data are almost endless. This is where bio-informatics comes in. In order to be able to create scripts, algorithms and software to handle the large datasets, and preferably enable others to perform the same analyses, one would need to have at least some level of computational skills. At the same time, a biological background is needed to be able to understand the data and the research questions, and to anticipate follow-up questions and possible issues one inevitably runs into when working with biological data.

Data types and the biological variation they assess

Fragment sequencing

NGS can be used to assess many types of biological variation. Studies applying NGS to investigate single nucleotide variants (snvs) and small insertions/deletions (indels) are probably the most common (**Figure 1**). When comparing a random genome sample to the human reference genome, approximately 3 million single nucleotide variants can be identified[3]. The vast majority of these variants occur in the 97% of non-protein coding bases of the genome[3,4]. The effect of those variants on an individuals' phenotype is very hard to predict, as they do not encode for an amino acid sequence. Most studies therefore choose to sequence only the coding parts of our genome (the exome), as this also greatly increases the number of samples that can be analyzed using a single sequence run compared to whole genome sequencing. Other studies focus only on genes they know or predict to be associated with the phenotype of interest to be able to sequence even greater numbers of samples, or even just screen parts of genes that contain known variants. Selecting the regions of interest from DNA, so called enrichment, is done using probes on a chip or in a solution complementary to your DNA of interest, or with PCR primers just around the region of interest. The enriched areas can be amplified and subsequently sequenced to a higher coverage, which leads to a higher sensitivity and accuracy for variant calling.

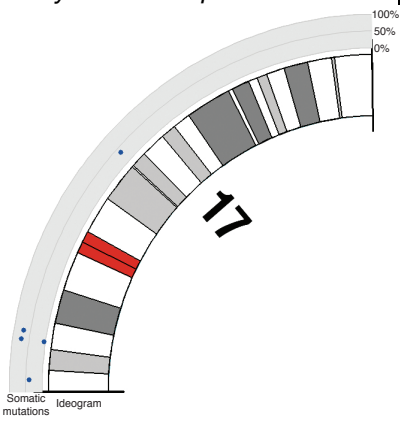
Even after enrichment for genes of interest, interpretation of variants identified through this technique is still a difficult task. Depending on the approach of the

study, the number of identified variants can vary from a handful to a few hundred. For single nucleotide variants, predictions can be made on the effect of the nucleotide change on the protein. Several prediction algorithms have been developed such as SIFT[5] and Polyphen[6], that take into account a.o. conservation of the affected genomic position, conservation of the amino acid sequence in other species or related proteins, the likelihood of the amino acid substitution and the location in the protein structure. They produce a classification ranging from benign to probably damaging / deleterious. Unfortunately the different algorithms do not always agree and variants predicted to be benign have been found to cause disease[7,8]. On top of that, they can't distinguish between protein activating or deleterious mutations. Insertions and deletions usually cause loss of protein function, but in-frame indels of a multitude of 3 basepairs occasionally lead to an activated protein product[9,10].

With or without enrichment, the DNA can be sequenced as a single fragment, or both ends of the read can be sequenced separately. The latter is called paired-end or matepair sequencing and is described below. Both single-end and paired-end sequencing results can not only be used for variant calling, but also to determine copy number status of genes or chromosomes by calculating the relative number of sequence reads in a certain window in your test sample and comparing that to one or more control samples sequenced with the same technique[11]. However, the resolution and quality of the results may vary depending on the type of enrichment (if any), quality of the sequence data and quality of the input material. This type of sequencing is therefore mostly used to detect single nucleotide variants and indels.

***Figure 1**: Circos plot of chromosome 17, of a primary tumor sample of one of the patients described in **chapter 4**.*

The inner band shows the chromosome 17 ideogram with the centromere in red, the outer band shows allele frequencies of tumor specific point mutations detected through exon sequencing of ~2000 genes.



SNParrays

The vast majority of these variants are common and are called single nucleotide polymorphisms (SNPs) if they occur in at least 1% of the population. These polymorphic positions have been utilized to create SNParrays: small chips with short probes specific for the different alleles (A and B) at each SNP position. DNA of a sample is applied on such a chip after being fluorescently labeled. If all DNA fragments bind only to the probe representing allele A or to the probe with allele B, the sample DNA is homozygous for that SNP, while if DNA hybridizes with probes representing allele A and probes with allele B, the sample is heterozygous for this position. B-allele frequencies can be determined by comparing the intensities of the fluorescent signals of both alleles. On top of that, the intensity at each position is representative for DNA copy number at that genomic location: if the sample contains more copies of a fragment of DNA, more fragments will hybridize to the probes representing that fragment thus increasing intensity of the signal (**Figure 2**). So in summary, SNParray data provides information on genotype, homozygosity and copy number status. The downside of this technique is the resolution; only regions containing a polymorphic position can be assessed, which means that some genes or regions will not be interrogated. Array comparative genomic hybridization (aCGH) is a technique similar to SNParray, as it also involves chips with probes representing specific genomic locations and intensity of the fluorescent signal at the probes location indicates copy number status. These probes are strategically placed throughout the genome resulting in a higher resolution than SNParrays, but they can't identify homozygous regions as they do not measure allele frequencies. Although SNParrays and aCGH are not strictly speaking an example of Next-Generation Sequencing as they do not produce sequence reads, they do allow for massive parallel genotyping of hundreds of thousands genomic locations in a sample.

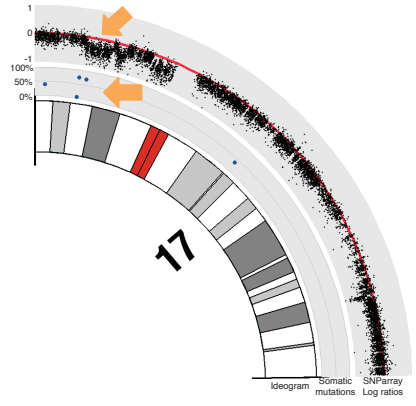


***Figure 2**: Circos plot of chromosome 17, of a primary tumor sample of one of the patients described in **chapter 4**.*

The inner band shows the chromosome 17 ideogram, the second band shows allele frequencies of tumor specific point mutations, and the outer band shows Log R ratios of the intensity of the signal of SNPs on this chromosome from SNParray data.
A deletion at the beginning of the chromosome coincides with more homozygous tumor spe-cific variants in the first tier (orange arrows).

Paired-end or matepair sequencing

Other methods to detect structural variation are matepair and paired-end sequencing[12]. These two methods each have their own underlying technique, but the common operator is that they involve sequencing both ends of a DNA fragment of a fixed length, after which these ends are aligned separately to the reference genome. Discordant read pairs, i.e. reads that do not map to the genome in the right order or at the expected distance from each other (based on the original DNA fragment length), indicate structural variation such as deletions, insertions, inversions and translocations (**Figure 3**). And, similar to regular fragment sequencing, copy number can be estimated based on sequence coverage. These methods have the additional value over the array-based techniques mentioned above of being able to detect copy neutral events, such as double stranded breaks in chromosomes that are erroneously re-assembled. On top of that, the resolution can be much higher than of SNParrays and aCGH as genome coverage is not restricted to probe locations.

RNAseq

Finally, RNAseq or transcriptome sequencing can be used to assess gene expression. The most basic use is the determination of gene expression levels, by isolating messenger RNA (mRNA) from the sample. Normalized counts of single-end sequencing reads over genes or exons can be compared between samples to identify differential expression (**Figure 3**). Paired-end sequencing can supply additional information, by identifying alternative splicing, fusion genes or allele specific expression.

Analysis of gene expression levels presents several important challenges. First of all, it can be difficult to find a reliable control sample. For example, when analyzing blood samples from a child with a congenital disorder, blood samples from the parents cannot be used because a mature person has a different gene expression profile than a developing child, but gene expression differences in blood samples from other children might also be irrelevant because of differences in genetic background and lifestyle. Another example is gene expression analysis from tumor biopsies, where normal tissue from the affected organ may not be available. Second, when analyzing tissue samples, you are usually dealing with a mixture of cell types and expression differences between samples can be due to varying levels of the different cell types within samples. Finally, as RNA is rather instable, sample treatment can affect the outcome of analyses, which makes it difficult to compare results from multiple studies. This is why a well thought-out experimental setup is essential for expression analyses.

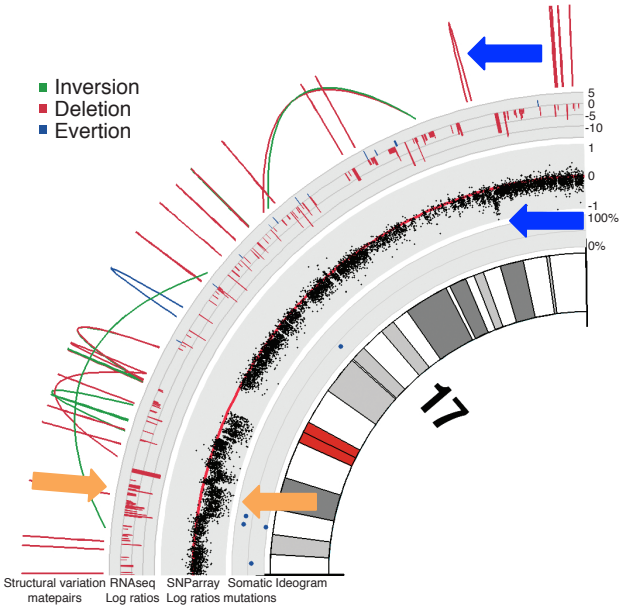
RNAseq can also be used to interpret the effect of genetic events on gene transcription, thereby helping to identify mutations or structural variants that

influence phenotype. For example, a genetic variant identified in a gene that is not expressed in either control or test sample is unlikely to have any effect, while if expression of the altered gene (or genes downstream from it) changes in the altered samples the variant is likely to be functional.

***Figure 3**: Circos plot of chromosome 17, of a primary tumor sample of one of the patients described in **chapter 4**.*

The inner band shows the ideogram of chromosome 17, the second band shows allele frequencies of tumor specific point mutations, the third band shows Log ratios of the SNParray data, the fourth band shows Log ratios of expression levels of the sample measured using RNAseq, compared to a reference sample from the same patient, and the outer band shows structural variation detected through matepair sequencing.

The deletion at the beginning of the chromosome coincides not only with more homozygous somatic vari-ants, but also with a lower gene expression in the RNAseq (orange arrows).
In addition, the small deletion near the end of the chromosome is detected by both matepair and SNP-array data (blue arrows).



What has tumor sequencing brought so far to our understanding of cancer?

Drivers and passengers

Sequencing tumor samples has brought us the concepts of tumor driver and passenger mutations, tumor suppressors and oncogenes[13,14]. As the tumor genome is highly instable, most somatic changes (genetic changes that do occur in the cancer cell, but not in the healthy parental tissue) detected when sequencing a tumor genome are thought to be arbitrary: they occur for instance in non-expressed genes or non-functional regions, they do not influence the protein coding sequence or have no significant effect on protein expression or the cells phenotype that would benefit the tumor. These are called passenger events, and are distributed at random throughout the genome. Tumor drivers are, as the name implies, genetic events that are thought to drive tumorigenesis. They tend to occur in so called cancer genes: genes that are more frequently 'hit' than what would be expected by chance. Examples of these genetic events include frequently observed genetic aberrations such as point mutations or indels at specific genomic locations, gene deletions, amplifications or translocations. COSMIC (the Catalogue of Somatic Mutations in Cancer)[15] is a database that contains information on somatic mutations in cancer from studies all over the world, which can help to distinguish between drivers and passengers.

Oncogenes and tumor suppressors

Cancer genes can be divided into the subclasses of oncogenes and tumor suppressors. Oncogenes are genes that become oncogenic upon activation, for instance through amplification, translocation or a specific amino acid substitution. Well-known examples of oncogenes are growth factor receptors such as *EGFR*, frequently amplified or mutated in lung tumors[16], and *ERBB2*, overexpressed in Her-2 positive breast cancer[17]. Activation of these genes leads to an increased cell growth and proliferation. The distribution of somatic mutations over these genes usually reveals clear 'hotspot regions': single amino acids or short stretches of amino acids that are commonly mutated, while the rest of the gene seems relatively quiet (**Figure 4A**). Alterations in these hotspot regions will render the protein product active, while other mutations will not have this effect or even damage the gene. This activation can be caused by altering the protein structure towards its active form, by increasing the binding affinity of the proteins substrates or by altering its phosphorylation sites. Protein phosphorylation is a common, reversible post-translational modification of either a serine, threonine or tyrosine residue, which is used for example to temporarily activate or inactivate a protein.

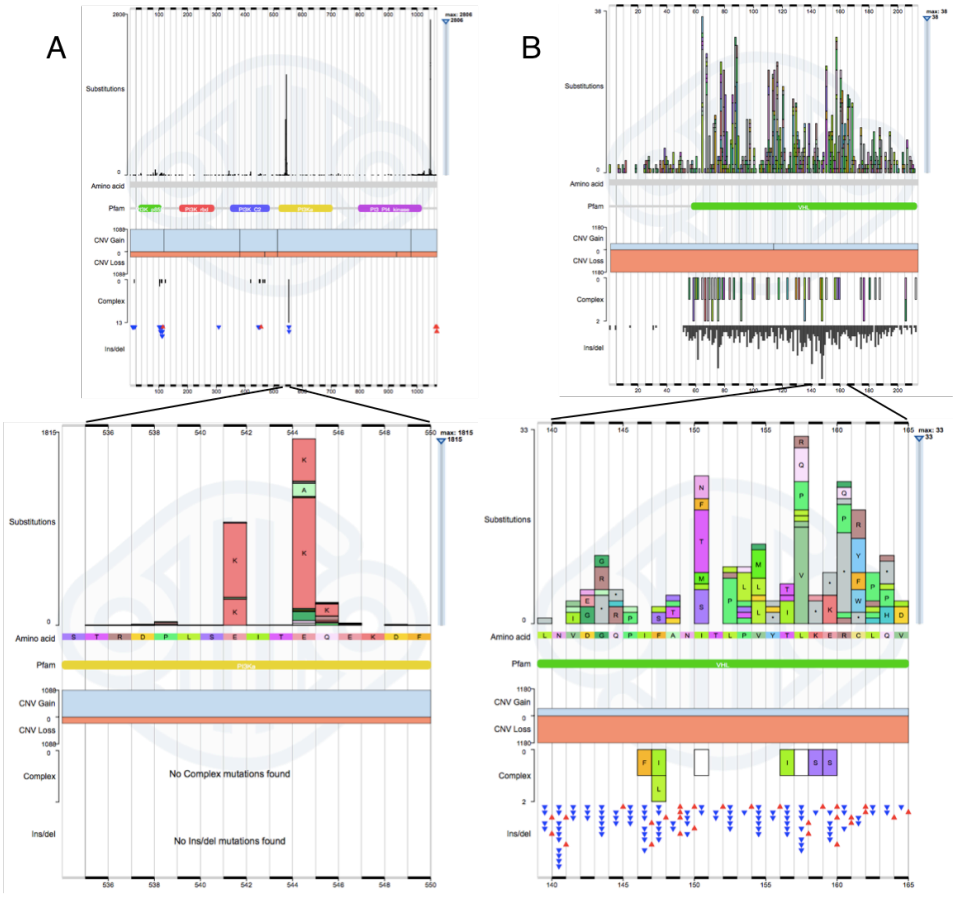


***Figure 4**: COSMIC[15] histograms of a typical tumor suppressor and a typical oncogene.*

***A**: COSMIC histogram of PIK3CA, an oncogene often mutated or amplified in several cancers such as breast, ovarian and colon. The overview in the upper panel displays two very distinct mutational 'hotspots', while the rest of the gene is relatively quiet. Zooming in on one of the hotspots (lower panel) shows a clear preference for substitutions of the negatively charged glutamic acid (E) to positively charged lysine (K).*
***B**: COSMIC histogram of VHL, a tumor suppressor often mutated in kidney tumors. The overview of the full gene in the upper panel shows a distribution of point mutations, insertions/deletions, complex events and copy number loss throughout the gene. The lower panel zooms in on a random region within this gene, and shows there is no clear preference for specific amino acid substitutions at any position.*

When the specific phosphorylation sites or regions flanking these sites are altered, the protein can no longer be regulated through this mechanism[18].

Gene amplification, a strong increase in the number of copies of a complete gene within a single cell, will usually lead to increased transcription and thus expression of the protein product. Gene activation through translocation is a more complex structural event. Genes can for instance be placed downstream of an active promoter, which will lead to increased transcription. Another example involves chromosomal breaks that occur in the middle of two genes that are erroneously reassembled, leading to so-called fusion genes. In most cases, this will not lead to a functional protein, but in several cancers recurrent fusion genes are found that are functional and even involved in tumorigenesis. The first discovered fusion gene is *BCR-ABL*, also known as the Philadelphia chromosome, which occurs through fusion of chromosomes 9 and 22 in chronic myeloid leukemia[19].

Tumor suppressors on the other hand are genes that, when active, help prevent a cell from becoming cancerous. The most well known example is *TP53*. *TP53* becomes active in response to cellular stress such as DNA damage, and will prevent the cell from dividing, induce DNA repair or trigger apoptosis (controlled cell death) if the DNA is beyond repair. This gene is a key player in many cancer types[20]. Another example of a tumor suppressor is *PTEN*, a gene that antagonizes the effect of the *PIK3CA* oncogene on the *AKT/PKB* pathway that leads to cell survival and proliferation[21]. Somatic mutations in these genes are distributed throughout the coding sequence and often include stop-codon introducing mutations and frameshift causing indels (**Figure 4B**). Also (partial) deletions of one or both copies of the gene are frequently observed.

The long tail of somatic variants

Interestingly, while many tumor samples harbor mutations or structural variants in one or more cancer genes, there are also tumor samples that do not contain any known tumor driving events. This indicates that much is still unknown. To discover new tumor driving events and thus increase our understanding of tumor development, a larger sample size is imperative. Recently, international collaborations between major cancer centers have been initiated, such as the International Cancer Genome Consortium (ICGC[22]) and The Cancer Genome Atlas (TCGA). The aim of these consortia is to collect hundreds of samples for all different tumor types and characterize these samples extensively, by generating mutation, copy number variation, expression and methylation data, and ultimately add clinical follow-up data. Moreover, all these data will be freely available for the cancer research community. Publications coming from these consortia and from other large cancer studies[23-25] all reveal a similar conception: per tumor type, only a few genes (usually less than five) are mutated or amplified in more than 5% of all

samples. These are typically the known cancer genes. All other genes are far less frequently hit, leading to a "long tail" of genes often with variants of unknown significance. To distinguish between driver and passenger events, algorithms such as MutSig[26], MuSiC[27] or CADD[28] can be applied on these large datasets to extract genes that are "significantly mutated". However, adding genes discovered through these methods still does not seem to cover the full scope of tumor driving events. It is likely that more drivers can be identified through clustering of gene families or even pathways[29], but these clusters will need to be closely defined. Another approach can be to study the effect of mutations on phenotype on an individual patient-level, for instance by combining genetic data with expression data and (phospho-)proteomics or by conducting additional follow-up experiments. This is however a very time consuming approach, but more feasible when the number of available samples is small.

Implementing NGS in cancer treatment decision making

Targeted treatment strategies

Every year, over 14 million people worldwide get diagnosed with some form of cancer, leading to over 8 million deaths per year. This makes it one of the leading causes of death[30]. On top of that, treating all these patients is expensive: in 2009, estimation of costs related to cancer treatment in Europe alone was 51 billion euros from which approximately 25% was spent on drugs[31]. Traditionally, cancer patients were treated based on their tumor type using rather generic treatments such as chemotherapy. While many patients do benefit from this approach, others do not benefit at all while they do suffer from the side effects of the treatment. By understanding how tumors originate and the disease evolves, new treatment strategies can be developed that increase efficacy of tumor treatment, and at the same time help prevent 'overtreatment' of patients who are unlikely to respond to a certain drug. These strategies include targeted therapy: targeting the tumor driving events discussed in the previous paragraphs. Several successful examples of this approach are already implemented in current cancer care (**Table 1**), such as trastuzumab to target breast cancers overexpressing Her-2[32], vemurafenib to target *BRAF*-mutated melanomas[33] and erlotinib or gefitinib to target *EGFR*-amplified lung cancers[34], but also the use of *KRAS* mutational status to predict non-response to *EGFR*-directed monoclonal antibody therapy in colorectal cancer[35]. The observation that different tumor types can be driven by the same pathways has lead to the more rapid implementation of targeted treatments for several additional cancers. Trastuzumab for example can now also be utilized to treat patients with Her-2 positive gastric cancer[36].

Table 1: Overview of targeted therapies used today.

Category of genomic alteration	Exemplary cancer gene	Type of cancer	Targeted therapeutic agent
Translocation	*BCR-ABL*	CML	Imatinib
	PML-RARα	APL	All-trans-retinoic acid
	EML4-ALK	Breast, colorectal, lung	ALK inhibitor
	ETS gene fusions	Prostate	
	Other	Leukemias, lymphomas, sarcomas	
Amplification	*EGFR*	Lung, colorectal, glioblastoma, pancreatic	Cetuximab, gefitinib, erlotinib, panitumumab, lapatinib
	ERBB2	Breast, ovarian, gastric	Trastuzumab, lapatinib
	KIT, PDGFRA	GISTs, glioma, HCC, RCC, CML	Imatinib, nilotinib, sunitinib, sorafenib
	MYC	Brain, colorectal, leukemia, lung	
	SRC	Sarcoma, CML, ALL	Dasatinib
	PIK3CA	Breast, ovarian, colorectal, endometrial	PI3-kinase inhibitors
Point mutations and small indels	*EGFR*	Lung, glioblastoma	Cetuximab, gefitinib, erlotinib, panitumumab, lapatinib
	KIT, PDGFRA	GISTs, glioma, HCC, RCC, CML	Imatinib, nilotinib, sunitinib, sorafenib
	PIK3CA	Breast, ovarian, colorectal, endometrial	PI3-kinase inhibitors
	BRAF	Melanoma, pediatric astrocytoma	Vemurafenib, dabrafenib
	KRAS	Colorectal, pancreatic, GI tract, lung	Resistance to erlotinib, cetuximab

Abbreviations: APL: Acute Promyelocytic leukemia, GIST: Gastro-intestinal stromal tumor, HCC: Hepatocellular carcinoma, RCC: Renal cell carcinoma, CML: chronic myelogenous leukemia, ALL: Acute lymphoblastic leukemia, PI3: Phosphatidylinositol-3

Adapted from MacConaill L et al, J Clin Oncol Vol. 28, 2010, 5219-5228 [29] Reprinted with permission.

Implementation of targeted therapy has, in several cases, a huge benefit for the patient. The best example is probably found in *BRAF*-mutated melanoma patients, where progression free survival of patients in a large phase III trial increased from 1.6 months with a 5% response rate on chemotherapy, to 5.3 months and a 48% response rate when treated with vemurafenib[33].

Tumor heterogeneity and treatment response

Until recently, the common assumption was that a cell builds up tumor driving events until it becomes cancerous and then gives rise to all other cancer cells within a patient, including topographic or temporal distant metastases. All these daughter cells will then carry the same tumor driving events, and only passenger mutations or mutations facilitating metastasis will be added in the course of time. In the past years, it has become clear that this assumption is often false, and that there are distinct genetic differences between tumor cells within a single patient. One study examining metastatic sites and up to nine different sections from primary clear cell renal cell carcinomas in single patients identified unique mutations in *SETD2* and in tumor suppressors *TP53* and *PTEN* in subsets of samples[37]. Another study using fluorescence in situ hybridization (FISH) on 350 glioblastoma samples observed intermingled subpopulations of cells with mutually exclusive amplifications of growth factor receptors *EGFR*, *MET* and/or *PDGFRA*[38]. Extending this intra-tumor heterogeneity, distinct differences between tumor locations (inter-tumor heterogeneity) were detected in several studies, for instance between primary colorectal tumors and liver, lymph node[39] or lung metastases[40], and between primary non-small cell lung cancer (NSCLC) and corresponding lymph node metastases[41]. These are just some examples of the vast amount of literature describing extensive intra- and inter-tumor heterogeneity.

This heterogeneity is a likely cause of the limited successes of targeted therapy so far[42,43]: if only some of the cells within a tumor contain the targeted event (such as an *EGFR* amplification), it is to be expected that only partial response is observed as the cells without the driving event keep growing unhindered. At the same time, cells that do harbor both the tumor driver as well as a mutation that confers therapy resistance will be selected for under pressure of treatment. In some cases, the resistant cells can be undetectable in a pre-treatment sample if the resistance inducing mutation does not entail any advantage for the untreated tumor cell. Genetic heterogeneity in itself has also been associated with response[42]: if there are many genetically different sub-populations of tumor cells present in a patient, the cancer has more ‘escape routes’ to evade the effects of therapy and will sooner become resistant than a homogeneous tumor.

Approaches to overcome or prevent drug resistance include the sequential administration of several different treatments and drug holidays, where the tumor is left untreated for some time[43]. The last approach is based on the assumption that the resistance inducing mechanism indeed does not have a growth advantage for the tumor, which will lead to suppression of the resistant cells by the sensitive cells if the drug is omitted[43].

Another aspect of the targeted therapy approach is biomarker discovery: the identification of (genetic) characteristics of a tumor that predict response to a certain drug. This is also severely hampered by tumor heterogeneity. Tumor biopsies or slides from resection material only represent a subset of cancer cells present in the patient, and while treatment response is usually determined on the "patient level", biomarkers are identified based on the "sample level". This would mean that, if response of the sampled tumor sub-population is not representative for the overall response of the tumor cell populations in the patient, the wrong genetic information is linked to the treatment outcome. To complicate things even further, a single (genetic) alteration is not sufficient to predict treatment response. If we revisit some of the examples above, we see on average a 50% response rate in vemurafenib treated patients with *BRAF* mutated melanoma[33,44,45], and response rates of approximately 80% for Her-2 positive breast cancer patients treated with trastuzumab in combination with chemotherapy[46], which suggests the presence of modulating factors. These are all important factors to consider in both biomarker discovery and clinical trial design.

Outline of this thesis

In conclusion, NGS has allowed for the discovery of new cancer treatment strategies and altered our conception of the disease irrevocably. In this thesis, we apply diverse methods of NGS to study the genetic makeup of tumors with the ultimate aim to improve cancer treatment. **Chapter 2** describes the development of a diagnostic test to analyze tumor genetics which is directly applicable for patient care: it presents an assay we developed to assess both somatic mutations and gene amplifications in tumor biopsies, which can be used to make informed decisions on subsequent treatment strategies. The other chapters are more focused on the evolution of tumors within patients by comparing samples from various locations and/or time points. In **chapter 3**, we used fragment sequencing, matepair sequencing and SNParrays to study the genetic changes between primary colorectal tumors and matched liver metastases. In **chapter 4**, we add RNAseq to these methods to be able to examine the effect of genetic dissimilarities between samples on the phenotype of the cell. We analyzed up to twelve treatment-naïve ovarian cancer samples per patient in this chapter, all collected at

the same time point but from different tumor sites, to study the extent of tumor heterogeneity without the influence of external factors such as targeted treatment, chemotherapy or radiation. In contrast, **chapter 5** uses the effect of targeted treatment on the genetic makeup of the tumor in order to identify therapy resistance mechanisms in melanoma patients, by comparing somatic mutations in samples before and after treatment with the *BRAF*-inhibitor vemurafenib. This brings us back to our final aim.

These chapters reveal the complexity of the disease and the challenges that still lie ahead, but also offer approaches to use the knowledge we already have to increase the efficiency of cancer treatment and expand our understanding of the mechanisms underlying therapy response.

2

Simultaneous detection of clinically relevant mutations and amplifications for routine cancer pathology

Marlous Hoogstraat[1,4], John W.J. Hinrichs[2], Nicolle J.M. Besselink[1,4], Joyce H. Radersma-van Loon[2], Carmen M.A. de Voijs[2], Ton Peeters[2], Isaac J. Nijman[3,4], Roel A. de Weger[2], Emile E. Voest[1,4], Stefan M. Willems[2,4], Edwin Cuppen[3,4,5], Marco J. Koudijs[1,4]

[1] Department of Medical Oncology, University Medical Center Utrecht, The Netherlands
[2] Department of Pathology, University Medical Center Utrecht, The Netherlands
[3] Center for Molecular Medicine, University Medical Center Utrecht, The Netherlands
[4] Netherlands Center for Personalized Cancer Treatment, Utrecht, The Netherlands
[5] Hubrecht Institute, KNAW and University Medical Center Utrecht, The Netherlands

Abstract

In routine molecular pathology of cancer, various independent experiments are currently required to determine mutation and amplification status of clinically relevant genes. The majority of these tests are designed to identify a limited number of genetic aberrations, most likely to be present in a given tumor type. Here, we present a modified version of a multiplexed PCR and IonTorrent-based sequencing approach that can replace a large number of existing assays. The test allows for the simultaneous detection of point mutations and gene amplifications in 40 genes, including known hotspot regions in oncogenes (e.g. *KRAS*, *BRAF*), inactivating mutations in tumor suppressors (e.g. *TP53*, *PTEN*) and oncogene amplifications (e.g. *ERBB2*, *EGFR*). A validation rate of 100% was obtained for point mutations, and a sensitivity / specificity of 100% and 99% respectively was determined for amplifications in FFPE material.

Implementation of a single assay to effectively detect mutations and amplifications in clinically relevant genes not only improves the efficiency of the workflow within diagnostic labs, but also increases the chance of detecting (rare) actionable variants for a given tumor type, that are typically not assessed in routine pathology. The ability to obtain comprehensive and rapid mutational overviews is key for improving the efficiency of cancer patient care through tailoring treatments based on the genetic characteristics of individual tumors.

Introduction

The implementation of next generation sequencing (NGS) technology in cancer pathology allows for comprehensive mutational profiling of large numbers of genes using only a limited amount of precious material. In addition, the sensitivity to detect variations with a low allele frequency is significantly improved (1-5%[47]) compared to conventional Sanger-based sequencing methods (10-20%), allowing the detection of rare somatic variants or mutations in heterogeneous tumors. Besides point mutations, the detection of small 2-20 bp insertion/deletions (indels) is highly relevant for clinical decision making, as shown for e.g. indels in *EGFR* exon 19 and 20[48], as well as frameshift mutations in *TP53*[49].

The majority of mutations in oncogenes tested for, result in a constitutive activation of the protein product, as known for the BRAF-V600E, KRAS-G12D, EGFR-L858R and PIK3CA-R1047H point mutations, or the inactivation of tumor suppressors by damaging mutations throughout the transcript as in *TP53*. Besides single nucleotide variants, tumorigenesis is regularly driven by high-level amplifications of genes, as shown for *EGFR* in gliomas or *ERBB2* in mammary carcinoma. The identification of these amplifications was the basis for the development of the *EGFR* and *ERBB2* specific inhibitors cetuximab and trastuzumab, respectively. These treatments significantly improved the overall survival of brain- and breast carcinoma patients[50,51], which underlines the relevance to detect this type of genetic variants. Currently, there are various diagnostic assays to determine *EGFR* or *ERBB2* positivity of tumors, including immunohistochemistry (IHC) to detect increased protein expression, fluorescent or chromogenic in situ hybridization (FISH/CISH)[52,53] and multiplexed ligation-dependent probe amplification (MLPA) to detect gene amplifications[54,55].

In current molecular pathology of cancer, primarily mutations and amplifications are tested that are associated with specific types of cancer, e.g. *KRAS* genotyping for colorectal carcinoma and non-small-cell lung carcinoma (NSCLC). Advances in cancer genome sequencing projects reveal however that tumor driving and clinically relevant variants are often not unique for a single tumor type[56]. By offering comprehensive genetic profiling for all clinically relevant genetic variations for each patient, relevant findings can be expected as shown for *ERBB2* positive gastric carcinomas where HER2 inhibition significantly increases survival[36], which was only recognized 10 years after the initial therapeutic efficacy in mammary carcinoma[57-59]. With the routine methods currently used in cancer pathology, it is not feasible to test each individual patient for all known mutations, indels and amplifications in a cost and time efficient manner. In addition, this is also hampered by the limited amount of tumor material typically obtained, that does not allow mutational profiling of large numbers of genes or exons by conventional

sequencing of PCR fragments, or various FISH/CISH or MLPA assays. Therefore, a single assay to detect the majority of all clinically relevant genetic alterations will be a major improvement for cancer diagnostics.

NGS technology, using sequence enrichment technology and ultra high-throughput sequencing, allows for the detection of mutations and copy number variations[60], but the limited amount of material typically obtained from primary tumors or biopsies, long turn-around times and relatively high costs prevent many cancer pathology labs to introduce these methods into their routine workflow. On the other hand, whole genome sequencing is suitable with limited input, but costs and turnaround times are not compatible with the requisites to detect mutations with low allele frequencies in heterogeneous samples. Here, we describe a modified, two reaction assay based on the IonTorrent AmpliSeq platform. Combined with a tailored bioinformatics framework, this assay allows for the simultaneous detection of clinically relevant mutations, indels and amplifications requiring only 20ng of DNA input and a turnaround time of less than 2 days. Mutations can be detected reliably with allele frequencies as low as 1-5%[47] and amplifications can be identified based on sequence coverage per amplicon with a similar sensitivity as MLPA. This single assay limits the number of analyses typically performed in cancer diagnostics, using only limited amounts of formalin-fixed paraffin embedded (FFPE) material, and simplifies the diagnostic workflow significantly.

Results

The OncoAmp Cancer Panel: a custom designed multiplexed PCR approach to detect clinically relevant genetic variants and high-level amplifications

Flexibility to analyze varying gene sets is an important aspect of a diagnostic assay since the detection of specific genetic variants can vary over time as well as between institutions, where stratification of patients to participate in clinical trial based on targeted anti-cancer therapies can be limited to a single hospital. Commercially available assays are typically designed to identify hotspot mutations or regions most commonly mutated, which is not sufficient to stratify patients to Phase-1 clinical trials using targeted treatments associated with novel gene aberrations. Therefore we aimed for the complete sequencing of clinically relevant (tumor suppressor) genes, (*TP53, CDH1*, *PTEN, VHL)*. In addition, we included a minimum of 5 amplicons per gene for the reliable detection of clinically relevant amplifications, as the original Ion AmpliSeq Cancer Panel design only contained 1 or 2 amplicons for some genes (e.g. *KRAS*). The additional amplicons per gene level the fluctuations in the coverage per amplicon, thereby increasing the sensitivity to detect amplifications, and allow for the detection of partial gene amplifications such as the clinically relevant EGFRvIII variant[61]. The Ion AmpliSeq Designer was used to develop a multiplex PCR assay for a gene panel, based on

the Ion AmpliSeq Cancer Panel® kit, focusing on predictive variants. The custom made OncoAmp Cancer Panel contains 409 amplicons in 40 genes, allowing the detection of clinically relevant gene amplifications and producing full sequence coverage of tumor suppressors and mutation/indel hotspot regions (**Table 1**). By multiplexing 8 samples on a Ion Torrent 318 chip, an average sequence coverage of >1000X per amplicon was obtained allowing the detection of genetic variants with low allele frequencies in heterogeneous tumors; three out of forty-one samples sequenced with OncoAmp version 2 were discarded based on pre-defined criteria due to insufficient coverage.

Sensitive and specific detection of mutations and indels in onco-genes and tumor suppressors.

To determine the performance of the initial design of OncoAmp Cancer Panel, 31 samples, including 8 healthy tissue samples (FFPE) serving as reference samples for amplification detection, were sequenced and analyzed as described. Amplicons with insufficient performance (low coverage or low amount of full length reads) were manually reevaluated and novel primers were designed to cover all regions of interest.

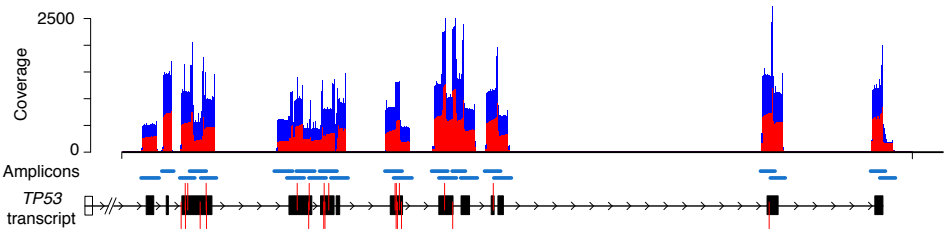
Table 1: Overview of genes in the OncoAmp panel

Type of analysis	Included genes			
Full transcript	*CDH1*	*MDM2*	*TP53*	
	EGFR	*PTEN*	*VHL*	
Mutation hot spot analysis	*AKT1**	*ERCC*	*GNAS*	*NRAS*
	*ALK**	*FGFR1**	*HRAS**	*PDGFRA**
	*BRAF**	*FGFR2**	*KDR**	*PIK3CA*
	CDKN2A	*FGFR3**	*KIT*	*RET*
	CTNNB1	*FLT3**	*KRAS**	
	*ERBB2**	*GNAQ*	*MET**	
Copy number alterations	*AKT1**	*ERBB3*	*FLT3**	*MYCN*
	*ALK**	*ERBB4*	*HRAS**	*PDGFRA**
	*BRAF**	*ESR1*	*KDR**	*RUNX1*
	CCND1	*FGFR1**	*KRAS**	*TOP2A*
	CDK4	*FGFR2**	*MET**	
	CDK6	*FGFR3**	*MTOR*	
	*ERBB2**	*FGFR4*	*MYC*	

*Both copy number status and (hotspot) mutation status are assessed for several genes, which is why they occur twice in this table.

In addition, amplicons showing severe strand bias were also redesigned to improve mutation detection in both directions. Subsequently, 25 additional samples were sequenced on this improved version of the panel, and 16 samples from the previous batch were repeated. Additional samples included 5 fresh frozen tumor samples and 8 blood samples to serve as negative controls for the fresh samples, from which 4 were matched tumor-normal pairs from the same patient. The sequencing data reached an average base coverage depth of >1000x. This allowed reliable detection of variants for the vast majority of positions; on average 91.86% of all bases in our design (including up to 99.11% of *TP53* bases as shown in **Figure 1**, which typically is much lower in commercial designs) were covered with at least 100x. Seventy-eight single nucleotide variants and indels were identified from which nine were previously reported by regular diagnostics, but 1 large deletion in *TP53* previously identified with Sanger sequencing did not pass the variant calling criteria even though it was present in the raw data. An overview of the samples and identified variants can be found in **Supplemental table 1**.

Figure 1: *Full coverage of TP53 using the OncoAmp panel*



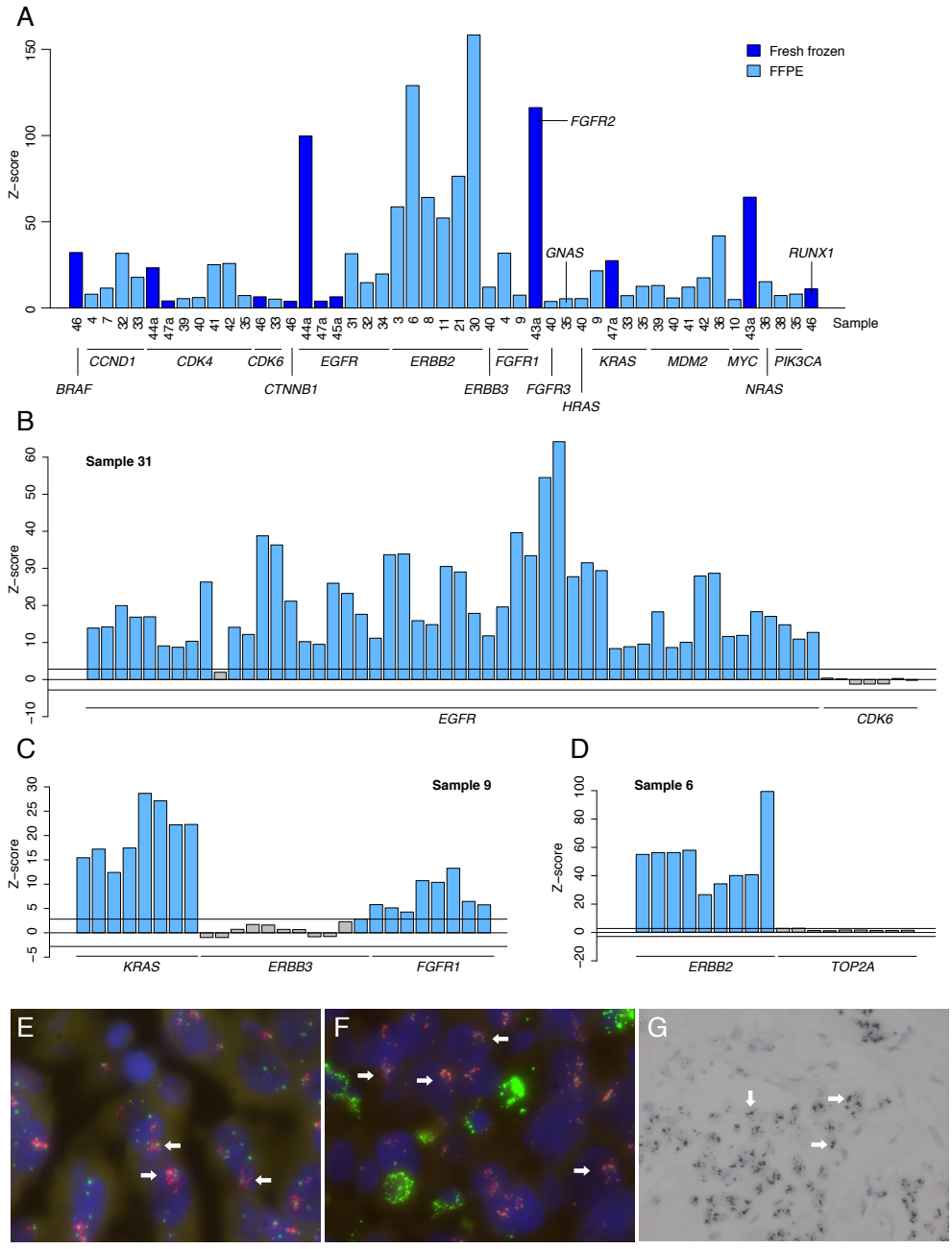
The TP53 transcript is visualized at the bottom, with black boxes indicating the exons; the white box on the left is the untranslated exon 1 in the 5' UTR of the gene. The vertical red lines in these exons indicate mutations detected in several samples sequenced with the OncoAmp panel. The blue horizontal lines above the transcript depict the placement of amplicons. The barplot finally shows the sequence coverage over TP53, divided into forward reads (red) and reverse reads (blue), with the coverage indicated on the Y-axis.

From the sixty-nine single nucleotide variants and insertions / deletions that were detected in addition to the variants previously reported, forty variants for which certified diagnostic assays were available were selected for validation with Sanger sequencing with a true positive rate of 100%. Among the additional variants, hotspot mutations in *PIK3CA*, as well as missense / nonsense mutations and a small frameshift deletion in *TP53* were present, which were previously not tested for in regular diagnostics. From the resequenced samples on the improved version of the AmpliSeq OncoAmp panel we could determine that the reproducibility of this aspect of the assay is very high: on average allele frequencies deviated only 2.15% between version one and two. Finally, the improvement of the panel led to the discovery of a stop-introducing mutation in *CDH1* in one of the repeated samples, which had been filtered out previously due to insufficient coverage.

Z-score analysis identifies high-level amplifications, highly concordant with MLPA and in situ data

Using the OncoAmp panel and the z-score method described here, we initially detected twenty-four high-level gene amplifications in eleven different genes by comparing samples within one (fresh frozen tumor samples) or two (FFPE tumor samples) sequence runs with each other. Subsequently, we compared the same samples with sets of control samples for fresh material or FFPE separately, since a coverage analysis showed a different coverage distribution for both types of material; amplicon coverage appears to be influenced by the length of the PCR fragment (**Supplemental figure 1**), an effect that is stronger in FFPE material compared to fresh (Spearman correlation -0.73 and -0.47, respectively). This second z-score analysis revealed several additional amplifications, leading to a total of fifty gene amplifications in nineteen different genes with z-scores ranging from 3.5 to >100 (**Figure 2A to D**). Moreover, our comprehensive analysis using the NGS-based assay identified several gene amplifications that would not have been screened for in regular pathology, such as *MDM2* in melanoma and *EGFR* in pancreatic cancer.

MLPA analysis of a total of 197 sample/gene combinations on this sample cohort identified a total of thirty-five amplifications of *MYC*, *ERBB2*, *FGFR1*, *FGFR2*, *CCND1, MDM2, EGFR, KRAS* and *CDK4* in the test samples, with values ranging from 2 to 8.73, where a score >2 is diagnostically used as a threshold to call a gene amplified. Four amplifications detected using the OncoAmp panel could not be confirmed with MLPA, and three amplifications that were not detected by our z-score analysis were detected with MLPA values between 2 and 4. This adds up to a sensitivity of 0.92 and a specificity of 0.98 for the OncoAmp panel.


A
Fresh frozen
FFPE
Z−score
FGFR2
GNAS
RUNX1
Sample
CCND1
CDK4
CDK6
EGFR
ERBB2
FGFR1
KRAS
MDM2
MYC
PIK3CA
BRAF
CTNNB1
ERBB3
FGFR3
HRAS
NRAS
B
Sample 31
EGFR
CDK6
C
Sample 9
KRAS
ERBB3
FGFR1
D
Sample 6
ERBB2
TOP2A
E
F
G

Noteworthy, six out of seven discrepancies were seen in the fresh frozen samples, which could also be due to the performance of MLPA on these samples (see discussion). If these measures are re-calculated for FFPE only, a sensitivity of 100% and specificity of 99% are achieved. As a second, independent validation method, amplifications of *EGFR*, *FGFR1* and *ERBB2* in ten samples were subsequently verified using FISH (*EGFR* and *FGFR1*, **Figure 2E, F**) or CISH (*ERBB2*, **Figure 2G**). All amplifications were confirmed and showed clusters with > 10 copies of the targeted gene.

Discussion

The paramount advantages of the assay presented here are the low (20ng) amount of input DNA required, which can be either from fresh frozen samples or FFPE samples, and the ability to get a comprehensive overview of multiple types of clinically relevant aberrations from a single experiment in a time frame suitable for routine screening. This small amount of required input makes the assay very suitable for material of which the DNA yield is usually relatively low, such as cytological samples obtained through lymph node or thyroid punctures. The implementation of a universal test for the majority of cancer samples, allows a more efficient workflow within diagnostic facilities. In addition, the turnaround time from sample retrieval, pathological analysis and DNA isolation to final variant lists and z-score calculations, following standard protocols, is practically feasible in 5 days, which we consider appropriate for routine cancer pathology.

Figure 2: *Gene amplification detection using the OncoAmp panel and in situ hybridization*

(A) *Z-scores of all amplified genes detected in the various samples sequenced with the OncoAmp panel. Genes and samples are indicated on the X-axis, the z-score on the Y-axis. The dark blue bars represent amplifications in fresh frozen samples, while amplifications in FFPE are shown in light blue.* ***(B)*** *Z-scores of EGFR amplification in a brain metastasis of a non-small cell lung carcinoma, CDK6 located on the same chromosome shows no aberrant copy number;* ***(C)*** *Z-scores of amplification of KRAS and FGFR1 in an adenocarcinoma of the stomach, ERBB3 located on the same chromosome as KRAS shows no aberrant copy number;* ***(D)*** *Z-scores of an ERBB2 amplification in an invasive ductulolobular breast carcinoma, TOP2A located nearby shows no aberrant copy number.* ***(E)*** *FISH (400x, DAPI counter stain) showing clusters gene copies for EGFR in (B) confirming EGFR amplification;* ***(F)*** *FISH (400x, DAPI counter stain) on FGFR1, confirming gene amplification in (C);* ***(G)*** *CISH confirming ERBB2 gene amplification in (D).*

The design of the customized multiplex PCR is flexible, and can be tailored or extended further to the needs of individual laboratories, although addition of amplicons to the multiplex PCR will require another round of validation.Although already suitable for implementation in cancer pathology, the Ion AmpliSeq OncoAmp panel can be further improved. As described before[62], regions with high G/C content are difficult to enrich or amplify by PCR resulting in low or variable coverage from amplicons covering specific regions. In the AmpliSeq OncoAmp design, this is the case for example for the first exons of *VHL* and *CDH1*. In addition, amplicons larger than 160bp do not amplify efficiently on fragmented FFPE material (**Supplemental figure 1**), which may also hamper the sensitivity to reliably detect single nucleotide variants in these regions. Iterative redesign of the primer pools may address this issue effectively. Optimization of the variant calling algorithms or settings may also further improve on diagnostic yield. The effect of this was illustrated by a large deletion in *TP53* in sample 83, which was initially called with an older version of the Torrent variant caller (v3.4.51874), but not by the latest release (v4.0-r77897).

This manuscript is primarily focusing on the ability of the OncoAmp assay to detect high-level amplifications with clinical relevance, since the detection of point mutations and indels using the Ion Torrent platform in a diagnostic setting have previously been reported[47,63-65]. These studies report variant allele frequency cutoffs ranging from 4% to 10%. As the 10% cutoff seemed too strict for our data (variants below 10% variant allele could be confirmed), we sought to lower this threshold. The 5% variant allele and 100x coverage cutoff were based on the observation that 8 out of 11 recurrent artifacts in the panel fulfilled these criteria (data not shown). To determine gene amplification, we developed an algorithm based on the modified z-score while several methods are already available to infer copy number state from next-generation sequencing data[12,66-68]. However, most of these algorithms require a single (matched) reference sample and/or are optimized for more comprehensive sequencing data such as whole genome or exome, trying to extract large regions of aberrant copy number. Moreover, these data usually are more evenly distributed across the genome, which is an assumption of some other algorithms that do not require a reference sample. Since initially only sequenced tumor samples were sequenced and the AmpliSeq OncoAmp design has a very small footprint (53 kb) with quite uneven coverage due to PCR amplification of single amplicons, no methods appeared to be suitable for our data. By implementing the modified z-score and analyzing the data on both the gene level and on the level of single amplicons, gene amplification can be determined with a similar sensitivity as MLPA. Interestingly, all but one of the false positives and all false negatives were found in the fresh frozen samples, and three out of four of the MLPA values for the false positives were around 1.8, which is sub threshold but

could indicate a low-level amplification. Since the protocol of MLPA on fresh frozen material has never been validated in our lab, these inconsistencies could also be due to false positives or negatives of the MLPA assay. The minimal tumor percentage of the amplification-positive samples analyzed is 25%, which is a measure for the sensitivity of our method. To increase the sensitivity even more, samples with lower tumor percentages could be enriched for tumor material by laser micro dissection, which is still expected to yield the required input amount of 20ng of DNA.

When using MLPA it is recommended to use samples of similar origin within one run, since fresh samples perform differently than archived material. This also seems to apply here, since the blood and fresh frozen samples performed differently compared to healthy FFPE tissue in terms of coverage per amplicon. Therefore, separate common reference pools for both FFPE and fresh-frozen material are required to reliably detect low-level amplifications using the z-score calculations. Pooling samples from multiple sequencing runs still results in flat copy number profiles for control samples and otherwise 'clean' data of tumor samples, i.e. no larger fluctuations per amplicon are observed when comparing samples from different runs, suggesting high reproducibility of the method. This shows that it is possible to use the results of a single sequencing run with control samples as a common reference pool, without the need to add control samples to each sequence run as is common practice in MLPA analyses, thereby significantly reducing costs. Control germline samples are also not required for variant analysis, as the majority of mutations targeted with this design and relevant for treatment decisions are very rare in germline samples. If germline status needs to be determined e.g. for *TP53* single nucleotide variants, assessing these specific positions with Sanger sequencing is more cost effective than including control samples for each patient on the OncoAmp panel.

We show here that a multiplexed PCR-based enrichment in combination with next-generation sequencing has the potential to be an attractive, efficient alternative to replace multiple diagnostic tests in the near future. This manuscript describes a solid proof of principle for the ability to detect clinically relevant variations in cancer genes using a single assay, which is becoming more and more valuable as the amount of tests required per tumor type increases; when more than two molecular tests are requested for the same sample e.g. *KRAS*, *BRAF* and *NRAS* for colorectal cancer, the assay described here is already more cost effective. For this panel, we included only genes relevant for current treatment options for patients with solid tumors. Therefore, amplification of genes such as *AML1* and *MLL* with prognostic relevance in leukemias[69-71] are not present in this assay and would require an adapted design. Similarly, current progress in clinical cancer research will undoubtedly lead to the identification of new biomarkers for treatment in the

near future. The flexibility of this PCR-based assay allows for adaptation of the design to add or omit genes or regions, adjusting it to any gene set of interest although iterative design and validation of the added regions would be required. These characteristics will be essential for effective implementation of personalized cancer treatment by patient stratification towards a tailored treatment based on genetic characteristics of an individual tumor.

Materials and Methods

Sample collection

FFPE tumor samples were selected from the archive of the department of Pathology of the UMC Utrecht, previously tested for amplifications, large insertions / deletions, activating hotspot mutations or point mutations / small indels, using MLPA or Sanger-based sequencing. In addition, 5 fresh frozen tumor samples previously sequenced using the IonTorrent Cancer Panel® were included to test the panels' performance on fresh frozen tissue. Finally, 8 samples that were histologically determined as healthy tissue and 8 fresh blood samples were selected to serve as a reference set. Overview of the samples can be found in Supplemental table 1.

DNA isolation FFPE samples

Tumor areas specified by pathological analysis on serial H&E sections were harvested from one or two whole 4 μm thick paraffin sections (adding up to approximately 1 square cm tumor tissue) with a scalpel. Tumor percentages of all samples were estimated before further processing and only samples with tumor percentages of at least 10% were used. DNA was isolated from these tissue fragments by using the cobas® DNA sample preparation kit (Roche, Basel, Switzerland) according to the manufacturer's protocol.

DNA isolation fresh samples

Peripheral blood samples (10 mL) were collected in K2EDTA tubes. Biopsies were snap-frozen in liquid nitrogen and stored at -80°C. Histological assessment to confirm the presence of tumor cells and to mark regions with high tumor cellularity for macrodissection was performed by a pathologist. DNA was extracted from 500μl whole blood and from 20 μm macrodissected sections using NorDiag Arrow (Isogen Life Science, De Meern, the Netherlands). DNA was quantified with Qubit 2.0 fluorometer® (Life Technologies, Carlsbad CA, USA).

Gene amplification analysis using Multiplex Ligation-Dependent Probe Amplification (MLPA)

MLPA is a PCR based technique that determines relative copy numbers in a semi-quantitative way and requires only minute quantities of fragmented DNA, which makes it very suitable to analyze DNA isolated from FFPE embedded material[55].

Gene amplification analysis with MLPA was described previously by Moelans et al[72]. In short: 50–100 μl DNA solution was, after centrifugation, used in the MLPA analysis according the manufacturer's instructions, using the P078-B1 kit (MRC Holland, Amsterdam, The Netherlands). All tests were performed in duplicate using an ABI 9700 PCR machine (Life Technologies, Carlsbad, CA, USA). PCR products were analyzed on an ABI 3730 capillary sequencer (Life Technologies). Gene copy numbers were analyzed using Genescan (Life Technologies) and Coffalyser (version 9.4) software (MRC-Holland). For genes with more than one probe present in the kit, the mean of all the probe peaks of this gene in duplicate was calculated. When this mean value was below 0.7 the respective gene was defined as lost, a value between 0.7 and 1.3 was defined as normal, a value between 1.3 and 2.0 as low-level amplification and values >2.0 as high-level amplified.

Gene amplification analysis using in situ hybridization (ISH)

Fresh cut 4 μm sections of formalin-fixed paraffin embedded tissue were submitted to dual-color FISH analysis using two probe sets. The *EGFR* and *FGFR1* probe sets were developed at Cytocell (Cambridge, UK). FISH was performed according to the manufacturer's protocol with some minor modifications. CISH analysis of *ERBB2* was performed using the INFORM HER2 Dual ISH assay on the VENTANA BenchMark ULTRA (Roche diagnostics, Basel, Switzerland) according to the manufacturer's instructions. A positive control was included in each ISH run and consisted of paraffin sections of a case known to be HER-2 amplified by ISH. Normal cells on the same slide, containing 2 copies, served as a "negative" control. ISH scoring was performed according the manufacturer's guidelines.

Screening for and validation of single nucleotide variants and indels using Sanger sequencing

Samples were analyzed by Sanger Sequencing using BigDye® Terminator Cycle Sequencing Kit and ABI 3730 capillary sequencer (Life Technologies).

Design of the OncoAmp panel

The assay presented here is an adaptation of the Ion AmpliSeq Cancer Panel® and primarily contains amplicons to detect currently known actionable mutations and amplifications in solid tumors. Prognostic genes without currently known clinical implications, such as *APC*, are excluded. Specific activating mutations in for example *KRAS* and *PIK3CA* are targeted with single amplicons, while transcripts of tumor suppressors such as *TP53* are targeted completely, including the exons of alternative transcripts and, where possible, with probes extending into the introns to cover variants effecting splicing. In order to detect amplifications, multiple amplicons were designed throughout the transcript, with a minimum of 5 (Supplemental table 2). Amplicons for the multiplexed PCR assay are designed using the Ion Ampliseq Designer tool (Life Technologies) aiming for 150bp

amplicons allowing efficient amplification of fragmented DNA isolated from FFPE specimen. The multiplexed PCR is split up in 2 reactions allowing the amplification and sequencing of overlapping amplicons, required to obtain full sequence coverage of larger exons.

Library preparation and sequencing on the Ion Torrent PGM

Samples analyzed using the Ion Ampliseq Cancer Panel are processed according to manufacturer's protocol. For the OncoAmp Panel, the multiplexed PCR is split up in 2 reactions of 10ul, using 10ng of DNA for each reaction. Each sample is barcoded using the Ion Express barcoded adapters allowing multiplexed sequencing. A total of 16 PCR cycles are performed on fresh frozen samples, FFPE samples undergo an additional 4 cycles. Libraries are quantified using the Qubit fluorometer (Life Technologies), pooled and diluted to a concentration of 0.015ng/ul, which is further processed for sequencing using the Ion OneTouch. Samples are sequenced on a 318v2 chip, allowing the simultaneous sequencing of 8 samples per run, aiming for a minimum average coverage of 500x per sample.

Standard data analysis

Base calling, alignment, coverage analysis and identification of variants and indels are performed using the standard Ion Torrent software (Torrent suite 4.0 and VariantCaller 4.0-r76860, Somatic – High Stringency settings). As ~92% of all target bases were covered with at least 100x (see Results), we discarded samples with < 80% target bases covered > 100x for further analysis. Mutations with coverage depth < 100x or allele frequency < 5% were filtered out to retain only high quality variants. Annotation and effect prediction of variants is performed using the Ensembl (v69) API, after which synonymous variants or variants in non-coding regions are filtered out. All coding variants are stored in an in-house database and common variants in either dbSNP (version 137, minor allele frequency > 0.01) or this in-house database (seen > 10 times) are excluded from this analysis unless they are present in COSMIC, as these are usually not relevant for diagnostics or are caused by reproducible sequencing errors. Sequencing statistics are calculated using the coverageAnalysis plugin v4.0-r77897 of the Torrent suite.

Z-score calculations

To detect gene amplifications, deviations in depth of coverage are determined using the modified or robust z-score according to Iglewicz and Hoaglin[73]. From the control samples, median and median absolute deviation (MAD) of coverage per amplicon and per gene are calculated and normalized on the total number of reads per sample to determine coverage distribution. Based on these values the z-scores are calculated for all amplicons and genes of each tumor sample (briefly, the difference between the coverage of the test sample and the median coverage of

the reference pool, multiplied by a correction factor of 0.6745, is divided by the MAD of the reference pool[73]). Coverage of neighboring amplicons above or below a z-score of 1.5 or -1.5 respectively is aggregated into a potentially affected region, after which the z-score for the total region is calculated. At the same time, a z-score for the complete gene is calculated. If all amplicons of a gene fall within this affected region and the region reaches the significance cutoff of 3.5 as suggested by Iglewicz and Hoaglin, the gene is reported as being amplified. However, as partial amplification of a gene can also occur such as the EGFRvIII variant in glioblastoma[61], genes are also reported if at least 50% of the amplicons within gene, as well as the full gene, reach the significance cutoff, This can occur when several amplicons in a gene do not reach the initial 1.5 cutoff, but all amplicons combined shows a significant deviation from the control samples.

Acknowledgements

SMW is funded by the Dutch Cancer Society (clinical fellowship: 2011-4964). The primer design and synthesis was provided for by Life Technologies. This work was supported by the Center for Personalized Cancer Treatment, which is a collaboration between the University Medical Center Utrecht, Erasmus Medical Center Rotterdam and the Netherlands Cancer Institute Amsterdam and supported by Barcode for Life.

Supplemental table 1: Description of tested samples

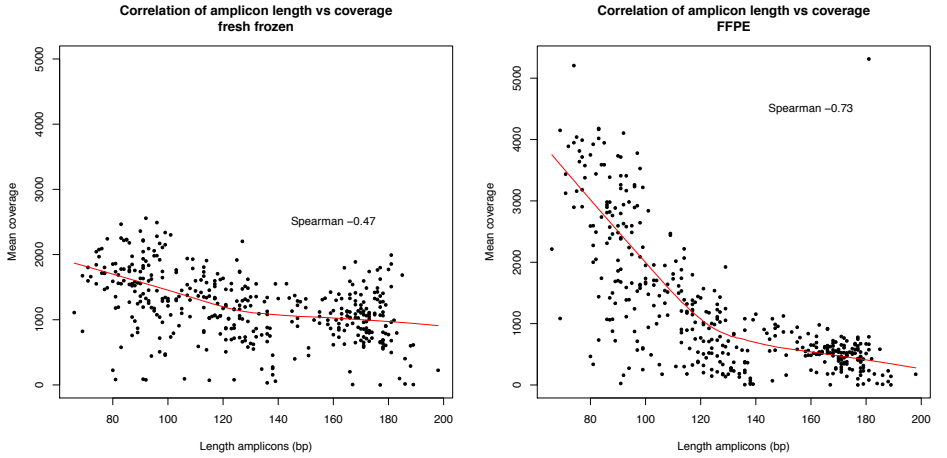
Sample #	Description	Biopsy / macrodissection	Material	Validated Mutations	Confirmed amplifications	tumor cellularity	tumor viability
1	GIST of the intestine	macrodissection	FFPE	*KIT* p.-576-577SL*S	-	80%	100%
2	Adenocarcinoma of the lung	macrodissection	FFPE	*PIK3CA* p.E545K, *PIK3CA* p.D549H, *KRAS* p.G12C	-	10%	80%
3	Metastasis ductal carcinoma of the breast	macrodissection	FFPE	*PIK3CA* p.H1047R	*ERBB2*	30%	95%
4	Invasive lobular carcinoma of the breast	macrodissection	FFPE	*PIK3CA* E542K	*FGFR1, CCND1*	40%	100%
5	Invasive lobular carcinoma of the breast	macrodissection	FFPE	*PIK3CA* p.H1047R, T*P53* p.Q192*	-	60%	100%
6	Invasive ductulolobular carcinoma of the breast	macrodissection	FFPE	*PIK3CA* p.H1047R, *TP53* p.D184fs*63	*ERBB2*	70%	100%
7	Invasive ductulolobular carcinoma of the breast	macrodissection	FFPE	*PIK3CA* p.E542K	*CCND1*	40%	100%
8	Invasive ductal carcinoma of the breast	macrodissection	FFPE	*PIK3CA* p.H1047R	*ERBB2*	70%	100%
9	Adenocarcinoma of the stomach	macrodissection	FFPE	*TP53* p.G245S	*FGFR1, KRAS*	60%	95%
10	Invasive ductulolobular carcinoma of the breast	macrodissection	FFPE	*PIK3CA* p.E545K, *TP53* p.R110P	*MYC*	60%	100%
11	Primary ductolobalar carcinoma of the breast	macrodissection	FFPE	*PIK3CA* p.T1052K, *TP53* p.S241T, *TP53* p.L111Q	*ERBB2*	80%	100%
12a	Metastasis squamous carcinoma of the larynx	macrodissection	FFPE	*PIK3CA* p.H1047R, *TP53* p.E258*, *TP53* p.D208N	-	cytology therefore NA	100%
12b	Squamous carcinoma of the larynx	macrodissection	FFPE	*PIK3CA* p.H1047R, *TP53* p.E258*, *TP53* p.D208N	-	60-70%	95%
13	GIST of the jejunum	macrodissection	FFPE	*KIT* p.VQW555-557V, *KIT* p.T670I, *TP53* p.E339Q	-	90%	100%
14	Livermetastases of a GIST	macrodissection	FFPE	*PIK3CA* p.H1047R, *KIT* p.KVV558-560I	-	80%	90%

15	Normal tissue	whole slides	FFPE	-	-	NA	NA
16	Normal tissue	whole slides	FFPE	-	-	NA	NA
17	Normal tissue	whole slides	FFPE	-	-	NA	NA
18	Invasive ductal carcinoma of the breast	macrodissection	FFPE	-	-	60%	100%
19	Normal tissue	whole slides	FFPE	-	-	NA	NA
20	Invasive ductal carcinoma of the breast	macrodissection	FFPE	*TP53* p.PHHERCS177-183S	-	60%	100%
21	Invasive ductal carcinoma of the breast	macrodissection	FFPE	*TP53* p.M237I	*ERBB2*	25%	100%
22	Invasive ductal carcinoma of the breast	macrodissection	FFPE	*TP53* p.R306*	-	40%	100%
23	Normal tissue	whole slides	FFPE	-	-	NA	NA
24	Invasive ductal carcinoma of the breast	macrodissection	FFPE	*PIK3CA* p.H1047R	-	50%	100%
25	Normal tissue	whole slides	FFPE	-	-	NA	NA
26	Normal tissue	whole slides	FFPE	-	-	NA	NA
27	Normal tissue	whole slides	FFPE	-	-	NA	NA
28	Invasive ductal carcinoma of the breast	macrodissection	FFPE	*PIK3CA* p.E545K, *TP53* p.C242Afs*5	-	40%	100%
29	Apocrine ductal carcinoma of the breast	macrodissection	FFPE	*TP53* p.Q52*	-	70%	100%
30	Invasive lobular carcinoma of the breast	macrodissection	FFPE	*TP53* p.R342*	*ERBB2*	40%	100%
31	Brain metastasis of NSCLC (adenocarcinoma)	biopsy	FFPE	*TP53* p.R110L, *EGFR* p.L858R	*EGFR*	50%	80%
32	Brain metastasis of NSCLC (adenocarcinoma)	biopsy	FFPE	-	*EGFR, CCND1*	60%	80%
33	Squamous cell carcinoma of the esophagus	biopsy	FFPE	*TP53* p.H193L	*KRAS*	40%	95%
34	Lymphnode metastasis of oral squamous cell carcinoma	biopsy	FFPE	-	*EGFR*	60%	90%
35	NSCLC lung (adenocarcinoma)	biopsy	FFPE	-	*KRAS, CDK4*	40%	70%

Supplemental table 1: Description of tested samples (continued)

Sample #	Description	Biopsy / macrodissection	Material	Validated Mutations	Confirmed amplifications	tumor cellularity	tumor viability
36	Pleural carcinomatosis NSCLC (adenocarcinoma)	biopsy	FFPE	*KRAS* p.G12V, *EGFR* p.A840T, *TP53* p.R280I	-	30%	90%
37	Brain metastasis of melanoma	biopsy	FFPE	*NRAS* p.Q61R	*MDM2*	60%	90%
38	Mediastinal lymphnode metastasis of cervical NEC	biopsy	FFPE	*PIK3CA* p.E542K	-	60%	80%
39	Well-differentiated liposarcoma	biopsy	FFPE	-	*MDM2, CDK4*	80%	100%
40	Well-differentiated liposarcoma	macrodissection	FFPE	-	*MDM2*	80%	100%
41	Well-differentiated liposarcoma	macrodissection	FFPE	-	*MDM2, CDK4*	80%	100%
42	Well-differentiated liposarcoma	macrodissection	FFPE	-	*MDM2, CDK4*	80%	100%
43a	Livermetastases of a esophagal carcinoma	biopsy	fresh	*TP53* p.L43Pfs*8	*MYC, FGFR2*	50%	80%
43b	Blood	NA	fresh	-	-	NA	NA
44a	Livermetastasis of pancreatic cancer	biopsy	fresh	-	*EGFR, CDK4*	40%	60%
44b	Blood	NA	fresh	-	-	NA	NA
45a	Lungmetastasis of colorectal cancer	biopsy	fresh	-	*EGFR*	60%	80%
45b	Blood	NA	fresh	-	-	NA	NA
46	Lymphnode metastasis of melanoma	biopsy	fresh	*BRAF* p.V600E	-	80%	95%
47a	Livermetastasis of colorectal cancer	biopsy	fresh	KRAS p.G12S, TP53 p.I251N	-	50%	50%
47b	Blood	NA	fresh	-	-	NA	NA
48	Blood	NA	fresh	-	-	NA	NA
49	Blood	NA	fresh	-	-	NA	NA
50	Blood	NA	fresh	-	-	NA	NA
51	Blood	NA	fresh	-	-	NA	NA

***Supplemental figure 1**: Correlation of amplicon length versus coverage for fresh and FFPE samples*



Scatterplots showing the correlation between amplicon length and mean amplicon coverage calculated from reference samples of each type of material. A LOWESS line has been fitted to the individual datapoints in red.
It is clear from these plots that the coverage per amplicons is influenced by the amplicons size, an effect much stronger in FFPE than in fresh. This is probably due to the fragmented state of the DNA in FFPE samples. As less of the sequencing capacity (reagents, bases) is used for the longer amplicons, the shorter amplicons reach a higher coverage on average in the FFPE samples compared to fresh, where the longer amplicons do reach > 1000x coverage.

3

Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer

Wigard P. Kloosterman[1], Marlous Hoogstraat[1,2], Oscar Paling[2], Masoumeh Tavakoli-Yaraki[1], Ivo Renkens[1], Joost Vermaat[2], Markus J. van Roosmalen[2], Stef van Lieshout[1,2], Isaac J. Nijman[3], Wijnand Roessingh[2], Ruben van 't Slot[1], José van de Belt[1], Victor Guryev[3], Marco Koudijs[2], Emile Voest[2], Edwin Cuppen[1,3]

[1]Department of Medical Genetics, University Medical Center Utrecht, The Netherlands
[2]Department of Medical Oncology, University Medical Center Utrecht, The Netherlands
[3]Hubrecht Institute and Cancer Genomics Center, The Royal Dutch Academy of Arts and Sciences and University Medical Center Utrecht, The Netherlands

Abstract

Structural rearrangements form a major class of somatic variation in cancer genomes. Local chromosome shattering, termed chromothripsis, is a mechanism proposed to be the cause of clustered chromosomal rearrangements and was recently described to occur in a small percentage of tumors. The significance of these clusters for tumor development or metastatic spread is largely unclear.

We used genome-wide long mate-pair sequencing and SNP array profiling to reveal that chromothripsis is a widespread phenomenon in primary colorectal cancer and metastases. We find large and small chromothripsis events in nearly every colorectal tumor sample and show that several breakpoints of chromothripsis clusters and isolated rearrangements affect cancer genes, including *NOTCH2*, *EXO1* and *MLL3*. We complemented the structural variation studies by sequencing the coding regions of a cancer exome in all colorectal tumor samples and found somatic mutations in 24 genes, including *APC*, *KRAS*, *SMAD4* and *PIK3CA*. A pairwise comparison of somatic variations in primary and metastatic samples indicated that many chromothripsis clusters, isolated rearrangements and point mutations are exclusively present in either the primary tumor or the metastasis and may affect cancer genes in a lesion-specific manner.

We conclude that chromothripsis is a prevalent mechanism driving structural rearrangements in colorectal cancer and show that a complex interplay between point mutations, simple copy number changes and chromothripsis events drive colorectal tumor development and metastasis.

Introduction

Colorectal cancer develops from a benign adenomatous polyp into an invasive cancer, which can metastasize to distant sites such as the liver[74]. Tumor progression is associated with a variety of genetic changes and chromosome instability often leads to loss of tumor suppressor genes, such as *APC*, *TP53* and *SMAD4*. High-throughput DNA sequencing has indicated that there are between 1,000 and 10,000 somatic mutations in the genomes of adult solid cancers[75-78]. Furthermore, next-generation sequencing has revolutionized our possibilities to profile genetic changes in cancer genomes, yielding important insights into the genes and mechanisms that contribute to cancer development and progression[78,79]. Systematic sequence analysis of coding regions in primary and metastatic tumor genomes has shown that little mutations are required to transform cells from an invasive colorectal tumor into cells that have the capability to metastasize[80]. Similarly, only two new mutations were identified in a brain metastasis compared to a primary breast tumor[81]. These data suggest that essential mutations needed for cancer progression occur predominantly in the primary tumor genome before initiation of metastasis[82]. In line with this hypothesis is the finding that distinct clonal cell populations in primary pancreatic carcinoma can independently seed distant metastases[83]. However, marked genetic differences between primary carcinomas and metastatic lesions do exist[84], and genotyping of rearrangement breakpoints in primary and metastatic pancreatic cancer revealed ongoing genomic evolution at metastatic sites[85].

In particular the impact and contribution of structural genomic changes to cancer development has recently received considerable attention[81,86-88]. Many solid tumor genomes harbour tens to hundreds of genomic rearrangements, which may drive tumor progression by disruption of tumor suppressor genes, formation of fusion proteins, constitutive activation of enzymes or amplification of oncogenes[85-90]. Rearrangements may be complex, involving multiple inter- and intra-chromosomal fusions and often reside in regions of gene amplification[86,91,92]. Recent genome-wide copy number profiling of cancer genomes suggests that 2-3% of all cancers appear to contain very complex rearrangements associated with two copy number states[93,94]. These events involve complete chromosomes or chromosome arms and are proposed to result from massive chromosome shattering, termed chromothripsis[93,94]. The prevalence and impact of such complex rearrangements in heterogeneous clinical specimens of solid tumors as well as their relevance for metastasis formation is currently unclear.

Here, we describe pairwise genomic analyses of matched primary and metastatic colorectal cancer samples from four patients using genome-wide mate-pair sequencing, SNP array profiling and targeted exome sequencing to explore the

genetic changes that constitute colorectal cancer formation and metastasis. We find marked differences between primary and metastatic tumors and show that chromothripsis rearrangements occur frequently in colorectal cancer samples. We conclude that chromothripsis events, along with simple point mutations and structural changes, are major contributors to somatic genetic variation in primary and metastatic colorectal cancer.

Results and Discussion

Patterns of structural variation in primary and metastatic colorectal tumors

Paired-end sequencing has proven a powerful technique to profile genomic rearrangements in cancer genomes[86]. However, there are some limitations associated with the use of short insert paired-end libraries for detecting structural variation[12]. Long-insert paired-end sequencing (also known as long mate-pair sequencing) has the advantage of being able to detect structural changes across repetitive and duplicated sequences[92].

To study the landscape of structural genomic changes in fresh tumor samples, we applied genome-wide long mate-pair sequencing and complementary SNP array profiling to matching primary and metastatic colorectal cancer biopsies from four patients (**Table 1, Materials and Methods**). Parallel analysis of normal tissues allowed us to efficiently detect *de novo* somatic rearrangements in the genomes of primary and metastatic lesions. Per sample, we generated between 10 and 65 million mate-pair sequence reads with an average insert size of 2.5–3kb, resulting in 10x to 48x average physical genome coverage per sample. We identified 352 somatically acquired rearrangements in the four patients, including deletions (177), tandem duplications (39), inversions (58), and interchromosomal rearrangements (78) (**Figures 1A and B**).

Table 1: Patient overview and tumor status

Patient	Gender	Histo-pathology	Primary tumor grade	Metastasis resection[a]	Treatment[b]
Pt 1	Female	Adeno-carcinoma	Moderately differentiated	3 months	No treatment
Pt 2	Male	Adeno-carcinoma	Moderately differentiated	20 months	No treatment
Pt 3	Male	Adeno-carcinoma	Poorly differentiated	10 months	XELOX[c] and Bevacizumab
Pt 4	Female	Adeno-carcinoma	Well differentiated	9 months	5FU, Leucovorin, Oxaliplatin, Bevacizumab

[a]Time between primary resection and metastasis resection, [b]Treatment after primary tumor resection, [c]Capecitabine and oxaliplatin

We independently confirmed the tumor-specific presence of 222 structural changes by PCR across the rearrangement breakpoint. Intrachromosomal rearrangements were particularly prevalent in our colorectal tumor samples, similar to what has been described for other tumor types (**Figure 1B**)[85,87,89]. Deletion-type rearrangements formed the most common class of rearrangements, with small deletions (up to 5 kb) being more common than large deletions. This is in contrast to primary breast cancer genomes, for which tandem duplications form the most common rearrangement class and deletions form the second largest class[87].

Since we sequenced both primary tumor genomes and liver metastases as well as control tissue, we could distinguish between rearrangements that were specific to both or one of these lesions. For all 222 confirmed rearrangements, we performed PCR-based breakpoint sequencing in primary tumor, metastasis and control samples (normal liver and normal colon tissue). The sensitivity of detecting a breakpoint by PCR is below 0.001% and should therefore be a reliable estimate of the presence of a rearrangement in DNA from a highly heterogeneous tumor sample[95]. Based on PCR-based breakpoint sequencing we found that, depending on the patient, between 32 to 95% of all rearrangements were specific to either the primary tumor or the metastasis (**Figure 1C**). There are several potential explanations for the observed differences between primary and metastatic sites: (i) changes could have occurred in the primary tumor and metastasis after dissemination to the liver, (ii) the part of the primary tumor sample that we analyzed did not contain the cells that were giving rise to the metastasis, (iii) metastatic tumor cells may have lost rearrangements that occurred in the primary tumor, and (iv) PCR may not be sensitive enough to detect breakpoints in very low numbers of cells, such as subclones in the primary tumor that may have given rise to the metastasis[83]. Given the significant overlap in somatic structural changes

between primary tumors and corresponding metastases (5%-68%, **Figure 1C**), we reason that many rearrangements arose in the primary tumor before metastatic spread. These overlapping rearrangements within a patient may represent early somatic rearrangements within the primary parental clone[83]. Subsequent genomic instability in the metastatic lesion may have lead to additional structural changes on top of the ones that were found in the primary tumor[85]. The many primary-tumor specific rearrangements likely arose after dissemination to the liver or were present in subclones of the primary tumor that did not have the capability to metastasize. Taken together, our pairwise comparison of structural changes in colorectal tumors shows that primary and metastatic colorectal cancer genomes have rearrangements in common, but also harbour distinct patterns of structural variation.

Chromothripsis is common mechanism driving structural changes in primary and metastatic colorectal tumors

Mate-pair sequencing allows identification of rearrangement breakpoints at nucleotide resolution. Furthermore, mate-pair signatures involved in complex patterns of structural changes may be used to reconstruct rearranged chromosomes by linking chromosomal fragments together based on their relative orientation. We have previously used mate-pair information to resolve a complex chromothripsis event in the germline[96].

***Figure 1**. Rearrangements in colorectal tumors detected by long mate-pair sequencing.*
*(**A**) Circos plots displaying rearrangements and their chromosomal location in primary and metastatic colorectal tumor samples. Rearrangement fusion points and orientations are indicated by coloured links: red, head-head; blue, tail-head; green, head-tail; orange, tail-tail (low coordinate to high coordinate). Chromosome ideograms are shown on the outer ring. The inner two rings show copy number profiles based on log R ratios derived from SNP array analysis. Red copy number plots correspond to the liver metastasis and blue plots correspond to the primary tumor. Copy number variation for matching normal colon and liver tissue are plotted in black. (**B**) Classes of rearrangements identified in tumors of the four patients. Deletion-type rearrangements have tail-head orientation, tandem duplication type rearrangements have head-tail orientation and inverted rearrangements have head-head or tail-tail orientation. (**C**) Lesion-specific presence of rearrangements in primary and metastatic tumors as based on PCR genotyping of samples for primary tumor, metastasis and control tissue.*

Figure 1

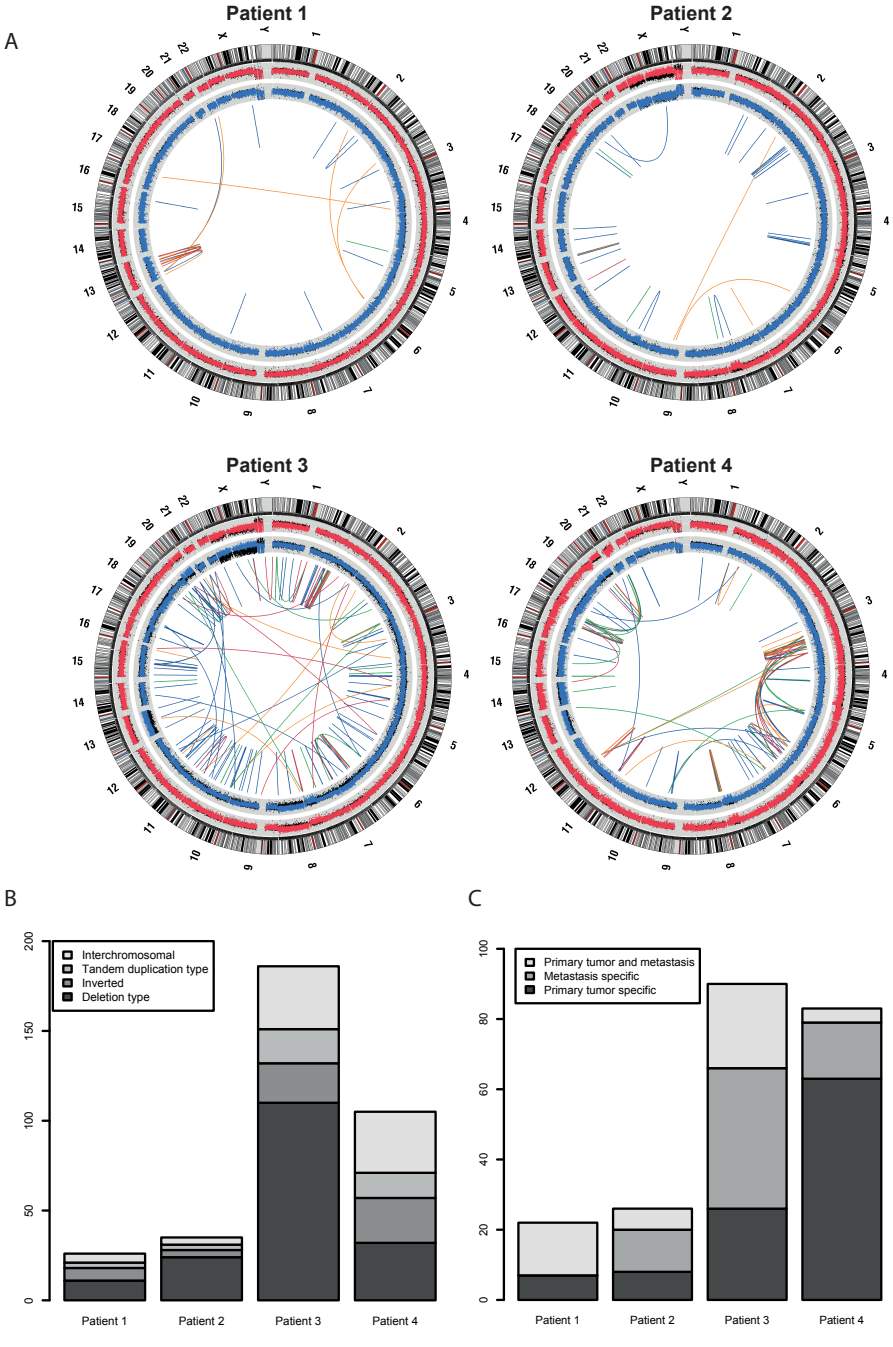

A
Patient 1
Patient 2
Patient 3
Patient 4
B
Interchromosomal
Tandem duplication type
Inverted
Deletion type
0
50
100
150
200
Patient 1
Patient 2
Patient 3
Patient 4
C
Primary tumor and metastasis
Metastasis specific
Primary tumor specific
0
20
40
60
80
100
Patient 1
Patient 2
Patient 3
Patient 4

Figure 2

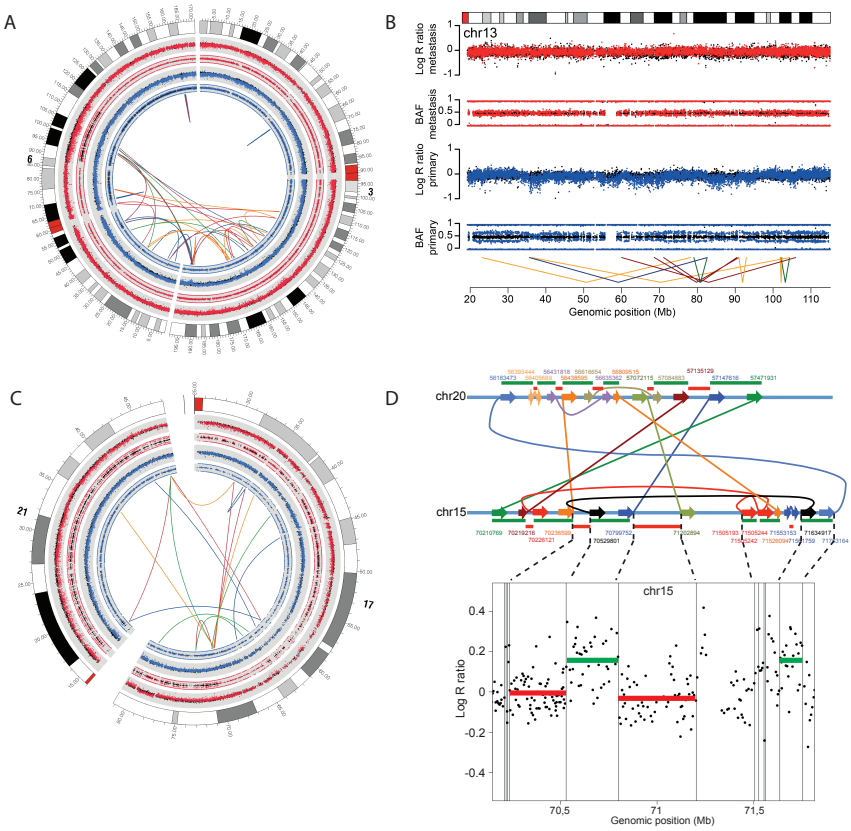


Figure 2. *Examples of rearrangement clusters in primary and metastatic tumor genomes.*
*(**A**) A cluster of rearrangements involving chromosomes 3 and 6 specific for the primary tumor of patient 4. (**B**) A cluster of rearrangements on chromosome 13, which could be found in both the primary tumor and the liver metastasis of patient 1. (**C**) A metastasis-specific cluster of rearrangements involving chromosomes 17 and 21 of patient 4. Orientations of fusions are coloured as in Fig. 1. Red copy number plots and B allele frequencies correspond to the liver metastasis and blue plots correspond to the primary tumor. Copy number variation and B allele frequencies for matching normal colon and liver tissue are plotted in black. (**D**) Breakpoints and copy number changes involving a cluster of rearrangements on chromosomes 15 and 20 in the primary tumor genome of patient 3. The upper panel shows a nucleotide resolution map of fusion points for this cluster. Lines indicate fusions between chromosomal fragments. Genomic coordinates indicate positions of breakpoints. Chromosomal fragments with both head and tail side connected to other fragments are retained, while fragments that lack any link (fusion) are supposed to be deleted. This expected pattern of retained and deleted fragments is reflected by the copy number profile for chromosome 15 (lower panel).*

Close examination of the landscape of genomic rearrangements in primary and metastatic samples, revealed chromosomal locations where breakpoints form complex clusters (**Figure 2**). There are several mechanisms that may account for the occurrence of complex rearrangements in cancer genomes[91,94,97]. Complex rearrangement patterns have been found in cancer amplicons[91], which may result from the breakage-fusion-bridge cycle following telomere dysfunction[97,98]. We do not find evidence for genomic amplification of regions involved in the complex clusters found here. Therefore, we regard it unlikely that these complex rearrangements are a result of the breakage-fusion-bridge cycle. As outlined below, we find that several complex clusters identified here, resemble the chromothripsis rearrangements described recently[94]. Clusters contain short and large chromosomal fragments that have head and tail sides connected to other distant chromosomal fragments as exemplified for the cluster involving chromosomes 15 and 20 in patient 3 (**Figure 2D**).

Furthermore, the inter- and intrachromosomal breakpoints of this cluster and most other clusters (chr 17-21, chr 3-6, chr 13) are associated with copy number changes, leading to two copy numbers states: high for retained fragments (i.e. with head and tail sides connected to other chromosomal fragments) and low for lost fragments (no connection to other fragments) (**Figure 2D**). Such alternated high and low copy number states are a striking feature of chromothripsis clusters identified previously[94]. However, the copy number changes we observed were not always as pronounced as previously reported[94]. This may be due to the fact that we studied heterogeneous tumor biopsies in our study as compared to clonally derived homogeneous cell lines in the previous study. For the clusters on chromosome 1 in patient 3, chromosomes 3 and 6 in patient 4 and chromosomes 17 and 21 in patient 4, we observed that cluster boundaries extend to telomeric regions, representing another characteristic that has been described as a hallmark of chromothripsis[94].

Based on sensitive PCR genotyping of breakpoints, several chromothripsis clusters displayed exclusive presence in either the primary tumor or the metastasis (**Figure 2**), further supporting the notion that they occurred as single simultaneous events, since a progressive model would more likely have resulted in the presence of at least some of the breakpoints in the corresponding lesion.

Capillary sequencing of PCR fragments across breakpoints allowed us to determine sequence characteristics of breakpoint regions. We characterized 159 fusion points at nucleotide resolution, of which 69 fall within complex chromothripsis clusters. There were no major differences in breakpoint characteristics for rearrangements within or outside complex clusters. Overall, we found that 38% were blunt-ended fusions and another 40% contained several

nucleotides of microhomology, the majority of the fusion points having microhomology of 1-3 bp. For 22% of fused segments we observed insertions of short nucleotide stretches, mostly below 6 bp, which likely represent non-templated nucleotides, which are often seen for double-stranded breaks repaired by non-homologous end-joining[99,100]. Next, we determined the overlap of breakpoints with repeat annotation (LINE, SINE, LTR, DNA repeat). However, we could not identify significant association of somatic breakpoints with any of these repeat classes, when compared to a set of randomly sampled positions across the genome (Fisher exact, P=0.5). The sequence characteristics of fusion points that we observed here resemble those that have been detected in various other cancers[87,88,92,101], and are in line with a process of non-homologous end-joining-mediated repair of double-stranded DNA breaks[94,99,100].

Overall, we conclude that small and large chromothripsis events result from massive double-stranded breaks and are frequently occurring in primary and metastatic colorectal cancer.

Chromothripsis cluster contribute to tumorigenesis in conjunction with point mutations, copy number changes and structural rearrangements

Recent studies have shown that complex rearrangements may promote cancer progression through disruption of tumor suppressor genes, or generation of fusion genes[87,88,92,94]. In addition, cancer amplicons frequently center on oncogenes, such as *ERBB2* and *MYC*[91]. To understand the contribution of chromothripsis clusters to tumor growth and metastasis, we analysed the breakpoint regions for the presence of cancer genes. One breakpoint of the cluster on chromosome 1 in patient 3 disrupts the fumarate hydratase gene (*FH*), which is a tumor suppressor frequently mutated in renal cell cancer (**Figure 3A**)[102]. Another rearrangement in the same cluster disrupts *EXO1*, which has tumor suppressor activity and may act together with *APC* to promote gastrointestinal tumor formation[103]. In patient 1, we identified a cluster on chromosome 13, and one of the breakpoints disrupts *MYCBP2* (**Figure 3B**). In addition, there are several cancer related genes from the Cancer Gene Census within the boundaries of this cluster and these may be affected by one of the numerous rearrangements in this cluster[104]. Besides complex clusters, we identified a range of isolated structural rearrangements for which breakpoints affect cancer genes, such as *NOTCH2*, *FHIT*, *MLL3* and *ETV6*[104]. We also detected several genes, which form hotspots of rearrangements in several patients. For example, *PARK2* is a tumor suppressor gene, which is known to contain frequent deletions in colorectal cancers[105]. We identified several independent deletions of *PARK2* in primary and metastatic tumors of patient 3 and 4. Although *PARK2* lies in a common fragile site, which explains the frequent deletions in this gene, it may function as a tumor suppressor and disruption of *Park2* increases adenoma development in *Apc* mutant mice[105,106]. Interestingly,

patient 4 carries two independent *APC* point mutations in the primary tumor and the metastasis respectively (see below and **Table 2**). We also identified several independent rearrangements in *FHIT*, *WWOX, PRKG1* and *MACROD2* in multiple patients. All of these genes are located at common fragile sites and have been found to contain rearrangements in several cancers[85,107].

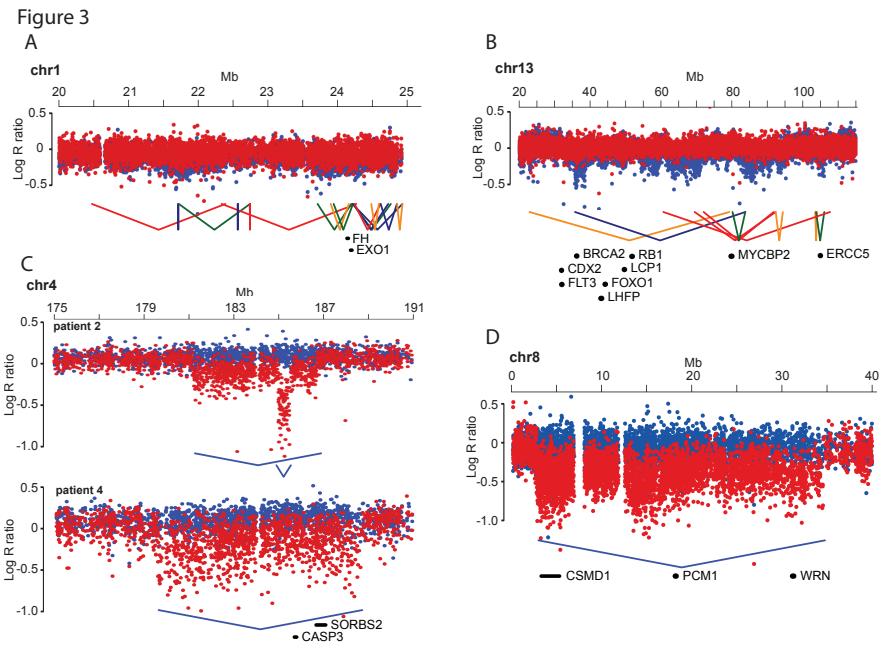


***Figure 3**. Cancer-related genes affected by rearrangements breakpoints. (**A**) Disruption of EXO1 and FH (fumarate hydratase) by rearrangement breakpoints in a metastasis-specific cluster on chromosome 1 in patient 3. (**B**) Disruption of MYCBP2 by a rearrangement breakpoint in a cluster on chromosome 13 in patient 1. Genes from the Cancer Genome Census are also depicted for this cluster. (**C**) Disruption of SORBS2 by metastasis-specific deletions in patients 2 and 4. (**D**) Disruption of CSMD1 by a metastasis-specific deletion in patient 2.*

Table 2. Point mutations identified in the cancer mini-exome of patients 1-4

	Patient 1		Patient 2		Patient 3		Patient 4	
gene	**TC**	**LM**	**TC**	**LM**	**TC**	**LM**	**TC**	**LM**
APC	S1411MfsX4	S1411MfsX4	E1536* R876*	E1536* R876*	Y1376*	Y1376*	A219PfsX74	R499*
DDR2				H340D				
KRAS			G12A	G12A				G12A
PTPRF			D562G	D562G				
SMAD2				R321*				
SMAD4				L495P				
TP53					R273C	R273C	R175H	C275W
MLL3	I155T	I155T						
PARP14	Q1332P	Q1332P						
PIK3CA	E545K	E545K					E545K	
KDR			R1032*					
PRKCD				T419I				
RFC1			P1131P					
EXOC4	K765R							
TSC1					R288C			
FGFR2	R399*	R399*						
NUP98					H1647D	H1647D		
ERBB3			V104M	V104M				
RASA3				V117M				
DNAH9			R4106H					
TAOK1	K484M	K484M						
ATRX							splice site	
TTN	K13350N			H8533Y, E4246K				
EPHA4				F801IfsX6				

TC, colon tumor; LM, liver metastasis. Annotation of the most common CCDS

fsXn; frameshift resulting in stop n residues downstream of indel

To get insight into the contribution of point mutations to tumor development in these and other cancer-relevant genes in our tumor samples, we performed next-generation sequencing based mutational profiling of a cancer mini-exome in all 16 tumor and control samples (1296 genes, **Materials and Methods**). We found canonical disrupting mutations in *APC*, *TP53*, *SMAD2* and *SMAD4* as well as *KRAS* (*G12A*) activation in several patients (**Table 2**)[74]. For patient 2 we identified the same mutations in *KRAS*, *APC* and *PTPRF* in both primary and metastatic tumor. However, mutations in *SMAD2* and *SMAD4* could only be detected in DNA from the metastatic tissue. In contrast, the tumor genomes of patient 4 contained mutation in *APC*, *KRAS* and *TP53*, but both primary tumor and metastasis carried their own private mutations in these genes. These data complement the mate-pair and copy number data, which also show overlapping mutations but also many distinct genetic variations in primary and metastatic samples, which may affect cancer genes in a lesion-specific manner (**Figure 1C**). For example, we identified metastasis-specific recurrent deletions of *CASP3* and *SORBS2* or deletion of *CSMD1* (**Figure 3C and 3D**). Interestingly, *SORBS2*, which is also known as *ArgBP2*, is repressed during oncogenic transformation of the pancreas and the protein was implicated in cell adhesion and migration[108]. Furthermore, *CSMD1* mutations have been found in particularly in advanced colorectal tumors, suggesting a role in metastasis formation[109]. Therefore, the distinct genetic changes in metastatic samples compared to corresponding primary tumors, likely contribute to metastasis formation or provide advantage to tumor growth at metastatic sites (liver). These data emphasize that comprehensive genetic analysis at the nucleotide as well as structural level, of both primary tumor and metastasis is needed to outline an effective targeted treatment strategy for colorectal cancer.

Conclusions

Our data show that clusters of complex genomic rearrangements occur frequently in primary and metastatic colorectal tumors. Based on the features of these complex rearrangement clusters, we find that chromothripsis is a common driver of genetic changes in colorectal cancer. We conclude that complex chromothripsis events in conjunction with simple copy number changes and point mutations shape the dynamic architecture of colorectal cancer genomes and all together provide the genetic basis for tumor growth and metastasis. Therefore, the impact of chromothripsis on tumor development and evolution may be greater than previously anticipated[94]. The molecular mechanisms that drive chromothripsis are unclear, but the characteristics of break points suggest that chromosome shattering occurred randomly, yet regionally, as a result from double stranded breaks and chromosomal fragments are likely repaired by non-homologous end-joining[94,96].

If the reshuffling of genetic information poses any benefit to the cell, chromothripsis clusters may drive tumor formation and metastases. A complex cluster could also be a passive genetic event, for example when coinciding with a growth promoting mutation in the same cell. While the observation that some complex clusters are uniquely present in primary or metastatic lesions could be supportive of this hypothesis, it could also be that chromothripsis events provide a selective advantage specific for the molecular environment of either the primary tumor or the metastasis. The distinct genetic mutation patterns in primary and metastatic tumors, illustrate the need for much more comprehensive screening of cancer genomes than is currently common practice, including profiling of (complex) structural changes along with coding mutations in primary and metastatic lesions.

Materials and Methods

Samples

The research in this study conformed to the Declaration of Helsinki of the World Medical Association concerning human material/data and experimentation. The Medical Ethics Committee (METC) of the University Medical Centre Utrecht, The Netherlands approved the genetic analysis of DNA from tumor and normal tissues of the patients described in this paper. Tissue samples were previously acquired as part of a series of routine diagnostic and pathological analyses in our hospital.

We performed mate-pair sequencing on DNA from tumor biopsies and control samples from 4 patients with colorectal adenocarcinoma attending University Medical Center Utrecht, The Netherlands. For each patient, we obtained DNA from the primary colon tumor, normal colon tissue, liver metastasis and normal liver tissue. We assessed tumor content of biopsies by microscopic analysis of stained cryosections (tumor content >80%).

Preparation of mate-pair libraries and SOLiD sequencing

Mate-paired libraries were generated from 50-100μg DNA isolated from tumor and control samples. Mate-pair library preparation was essentially as described in the SOLiDv3.5 library preparation manual (Applied Biosystems). We performed two genomic DNA size selections per library: one after shearing and one after CAP adaptor ligation. Libraries were cloned and 384 clones per library were picked for capillary sequencing to assess presence of adaptors, insert sizes and chimeric molecules. Chimeric molecules were identified based on a tag distance > 100kb. On average, we observed between 5-15% present chimeric molecules per library. We sequenced 2x 50bp mates for each library on one or two quadrants of a SOLiD V4 sequencing slide.

Bioinformatic analysis of mate-pair reads

The F3 and R3 mate-pair tags were mapped independently to the human reference genome (GRCh37/hg19) using BWA software V0.5.0 with the following settings: -c -l 25 -k 2 -n 1[110]. Mate-pair tags with unambiguous mapping were combined and split into local (<100kb) and remote (>100kb) mate-pair sets. Local mate-pairs were further split into mate-pairs with normal orientation of the tags relative to each other, mate-pairs with inverted tags and mate-pairs with everted tags[96].

Deletions were called from local mate-pairs with correct orientation and with a mate-pair span in the top 0.5% percentile of the mate-pair size distribution. Tandem duplications were called from local mate-pairs with everted orientation and inversions were called from local mate-pairs with inverted orientation. Mate-pairs were clustered based on overlapping mate-pairs with a maximal tag distance of 2 times the average library insert size. The remote (inter-chromosomal and intra-chromosomal > 100kb) mate-pairs were clustered independently of the relative orientation of the mate-pair tags. The orientation of the different mate-pair tags in a cluster relative to each other is indicated by H (or h for the minus strand) when the tag has its 'head' side (the side that points towards the start of the chromosome) opposed to the pairing tag and T (or t for the minus strand) when a tag has its 'tail' side (the side that points towards the end of the chromosome) opposed to the pairing tag. Mate-pair clustering was performed per patient (4 samples) and tumor-specific rearrangements were selected based on clusters without overlapping mate-pairs derived from normal tissue samples. Tumor-specific rearrangements were confirmed by PCR across the breakpoint in primary tumor, metastasis and normal liver and colon samples. Rearrangement fusion points were visualized by Circos software[111].

SNP-array analysis

DNA from all 16 tumor and control samples was analyzed by Illumina Cyto12 SNP arrays according to standard procedures (Illumina). Copy number changes and allelic profiles were derived from log R ratios and B allele frequencies that are provided by the Illumina Genomestudio package. Since overall copy number changes in the heterogeneous samples that we analyzed are not as marked as in clonally derived cell lines, we used custom scripts to detect areas with low or high log R ratio values (increase in copy number is defined as: a positive shift (> 0.1) in average log R ratio compared to a control sample (healthy colon or liver tissue from the same patient), and a decrease in copy number is defined as a negative shift (> 0.1) in log R ratio compared to the control sample. For both positive and negative changes, we required at least 12 consecutive deviating probes, while allowing a maximum of 2 probes that do not meet the criterion.

Copy number changes were further substantiated by changes in average B allele frequency for heterozygous positions relative to control samples (average B allele frequency shift larger than 0.05, also found in a minimum of 12 sequential probes, including a 2 probe 'mismatch' cut-off). The resulting copy variable regions were manually curated based on B allele frequency plots and log R ratio plots of tumors and matching healthy samples.

Mutational profiling

Mutational analysis of 1296 kinases and cancer-related genes was performed by multiplexed enrichment of barcoded fragment libraries from all 16 samples[112]. Capturing was done using a custom-designed Agilent 244K array with 60-mer tiled probes on both strands[113]. The pool of enriched libraries was sequenced on one slide of a SOLiD3.5 instrument. Data were mapped to the reference genome (GRCh37/hg19) using BWA (-c -l 25 -k 2 -n 1). SNP calling was done using a custom analysis pipeline that identifies mutations with a non-reference allele frequency larger than 15% and a coverage of at least 10x. All identified variants were validated by PCR and capillary sequencing.

Data access

Sequencing data from both mate-pair sequencing and mutational profiling are available from the European Nucleotide Sequence Read Archive (ENA SRA) under the accession number ERP000875. SNP-array data were submitted to the NCBI GEO archive and are available under accession number GSE32711.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

WK conceived and designed the study and performed the experiments and bioinformatic analysis and wrote the paper. MH performed bioinformatic analysis of array data. OP performed the breakpoint sequencing and analyzed the data. MT generated mate-pair libraries. IR performed SOLiD sequencing and generated fragment libraries. JV designed the study and contributed patient material. MR performed analysis of mate-pair sequencing data. SL performed analysis of targeted-exome sequencing data. IN performed analysis of targeted-exome sequencing data and designed the capture array. WR performed breakpoint sequencing. RS performed SNP array analysis. JB generated mate-pair libraries. VG performed analysis of mate-pair sequencing data. MK analyzed breakpoint regions and supervised experiments. EV conceived and supervised the study and wrote the paper. EC conceived, designed and supervised the study and wrote the paper.

Acknowledgements

This work was financially supported by the Cancer Genomics Center (CGC) program of the Netherlands Genomics Initiative (NGI). We thank Martin Poot for critically reading the manuscript.

4

Genomic and transcriptomic plasticity in treatment-naïve ovarian cancer

Marlous Hoogstraat[1,2,*], Mirjam S. de Pagter[3,*], Geert A. Cirkel[1,2,*], Markus J. van Roosmalen[3], Timothy T. Harkins[4], Karen Duran[3], Jennifer Kreeftmeijer[3], Ivo Renkens[3], Petronella O. Witteveen[1], Clarence C. Lee[4], Isaac J. Nijman[2,3], Tanisha Guy[3], Ruben van 't Slot[3], Trudy N. Jonges[5], Martijn P. Lolkema[1,2], Marco J. Koudijs[1,2], Ronald P. Zweemer[6], Emile E. Voest[1,2], Edwin Cuppen[2,3,7,9] and Wigard P. Kloosterman[3,9]

* These authors contributed equally

[1]Department of Medical Oncology, University Medical Center Utrecht, The Netherlands
[2]Center for Personalized Cancer Treatment, Utrecht, The Netherlands
[3]Department of Medical Genetics, , The Netherlands
[4]Life Technologies, Carlsbad, CA, USA
[5]Department of Pathology, University Medical Center Utrecht, The Netherlands
[6]Department of Reproductive Medicine and Gynaecology, Division Woman and Baby, University Medical Center Utrecht, The Netherlands
[7]Hubrecht Institute, KNAW and University Medical Center Utrecht, The Netherlands

Abstract

Intra-tumor heterogeneity is a hallmark of many cancers and may lead to therapy resistance or interfere with personalized treatment strategies. Here, we combined topographic mapping of somatic breakpoints and transcriptional profiling to probe intra-tumor heterogeneity of treatment-naïve stage IIIC/IV epithelial ovarian cancer. We observed that most substantial differences in genomic rearrangement landscapes occurred between metastases in the omentum and peritoneum versus tumor sites in the ovaries. Several cancer genes such as *NF1*, *CDKN2A* and *FANCD2* were affected by lesion-specific breakpoints. Furthermore, the intra-tumor variability involved different mutational hallmarks including lesion-specific kataegis (local mutation shower coinciding with genomic breakpoints), rearrangement classes and coding mutations. In one extreme case, we identified two independent *TP53* mutations in ovary tumors and omentum/peritoneum metastases, respectively. Examination of gene expression dynamics revealed upregulation of key cancer pathways including WNT, integrin, chemokine and hedgehog signalling in only subsets of tumor samples from the same patient. Finally, we took advantage of the multi-level tumor analysis to understand the effects of genomic breakpoints on qualitative and quantitative gene expression changes. We show that intra-tumor gene expression differences are caused by site-specific genomic alterations, including formation of in frame fusion genes. These data highlight the plasticity of ovarian cancer genomes, which may contribute to their strong capacity to adapt to changing environmental conditions and give rise to the high rate of recurrent disease following standard treatment regimes.

Introduction

In recent years, tremendous progress has been made in the understanding of the complexity of the cancer genome[78,114]. Studies including large numbers of patients per tumor type have identified recurrent mutations, copy number variants, epigenetic changes and genomic rearrangements specific for certain cancer types[22,115,116] (http://cancergenome.nih.gov/). Although more than 400 commonly mutated cancer genes have been identified[104,117-119], extensive genetic heterogeneity has been noticed across different cancer types and also within individual tumors[78,120]. Intra-tumor heterogeneity is a result of the action of the evolutionary forces of mutation and selection[78,120]. The traditional linear model of cancer evolution describes multiple, successive cycles of mutations and selection leading to malignant tumor cells, ultimately leading to metastases[82,120,121]. In contrast, parallel evolution describes dissemination of tumor cells from the primary tumor as a continuous process occurring from very early on in tumor development. These disseminated cells may continue to evolve independent of the primary tumor, causing the formation of metastases genetically relatively distinct from the primary tumor and other metastases[82,122]. Several studies have focused on spatial sampling of various cancer types to gain insight into the extent and complexity of tumor evolution[37,83,117,123].

Here we studied intra-tumor heterogeneity in epithelial ovarian cancer. With an annual worldwide incidence of 220,000 and mortality of 140,000, epithelial ovarian cancer is the leading cause of death among women with gynaecological malignancies and a disease in urgent need for improved treatment[124]. Large-scale genomic analysis of ovarian cancer patients has uncovered only a few recurrently mutated genes, such as *TP53* and mutations in *BRCA1*/*BRCA2* [116]. Ovarian cancers show a relatively high number of copy number variations and structural variations (SVs)[116,125,126]. This may be explained by the high incidence of deregulation of genes in the homologous recombination pathway (*BRCA1*/*BRCA2*), which has provided opportunities for successful treatment with PARP inhibitors[126,127]. Expression profiling has been instrumental to classify ovarian cancers and revealed molecular subtypes with prognostic relevance[128,129]. Despite these advances in understanding of ovarian cancer biology, the cure rate has not much improved[130,131].

We set out to understand the intra-tumor dynamics of treatment-naïve epithelial ovarian cancer by high-resolution analysis of genomic rearrangements. Because the effects of genomic rearrangements in tumor development are only poorly understood, we also examined the contribution of genomic rearrangements to intra-tumor differences in gene expression. We found that treatment-naïve epithelial ovarian cancers exhibit remarkably diverse patterns of genomic

rearrangements, which in turn lead to intra-tumor changes in gene expression, including upregulation of major cancer pathways in only subsets of samples from a single patient. These findings provide novel insight in potential mechanisms underlying treatment resistance.

Results

Topographic sampling of treatment-naïve epithelial ovarian cancer

Ovarian cancer is often discovered when the disease is already in an advanced stage, resulting in the presence of a unique metastasis pattern with cancer cells exfoliating throughout the abdominal cavity following the peritoneal fluid circulation route. The tumor mass is often large with metastases spread throughout the abdomen. Standard of care for such advanced ovarian cancer patients involves surgical cytoreduction before starting chemotherapy treatment. We obtained comprehensive tumor and whole blood samples from three treatment-naïve advanced epithelial ovarian cancer patients with high tumor loads (**Table 1)**. Patient 1 and 3 were diagnosed with a serous adenocarcinoma, whereas patient 2 was diagnosed with a carcinosarcoma, which is a less frequently observed (<1-4%) form of epithelial ovarian cancer[132]. Carcinosarcoma is characterized by the mixed histology of carcinomatous and sarcomatous components with a more aggressive behaviour and a poorer prognosis when compared to serous adenocarcinomas[133]. For each patient, tumor biopsies were obtained during surgery from physically separated tumor sites in the abdomen with the final goal to obtain a representative set of samples (**Figure 1A**). The tumor content of each sampling site was generally well above 50% based on histopathological measurements, although computational measurements by ASCAT indicated lower percentages[134]. Particulary, the metastatic tumor biopsies from patient 1 (p1.IV-1, p1.IV-2 and P1.V-1) and the right ovary tumor sample from patient 3 (p3.III) have a relatively low tumor content. However, these samples were included in most of our analysis, because we could compensate for the lower tumor content by deep sequencing of identified genomic changes. A total of 34 samples (27 tumor, 7 reference samples) were obtained and used for the analyses outlined below.

Table 1: Clinical data of epithelial ovarian cancer patients included in this study

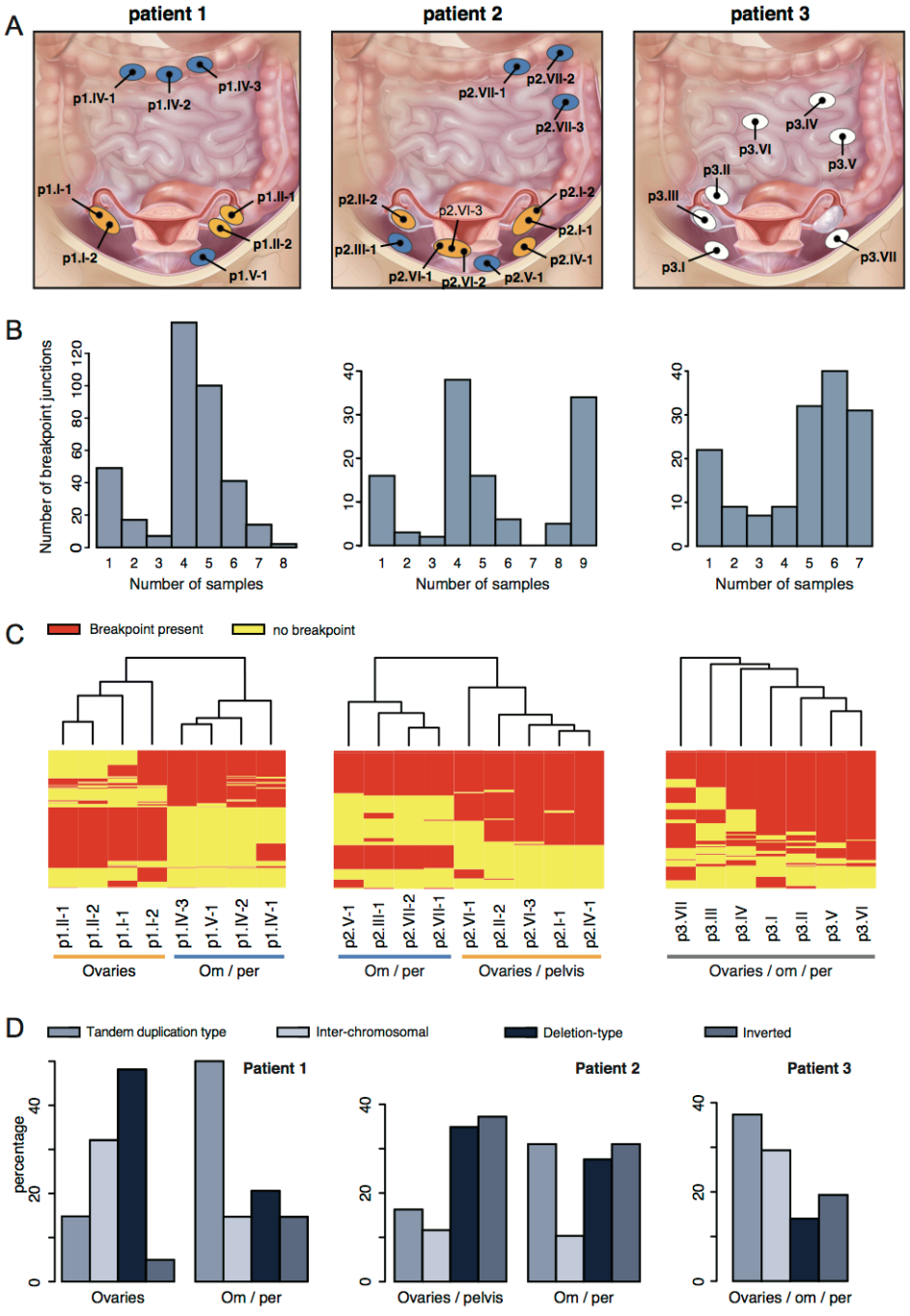
Patient number	Age	Histo-pathology	FIGO stage	Post-operative clinical course
Patient 1	53	Moderate to poorly differentiated serous adeno-carcinoma	IIIC	6 cycles adjuvant combined intra-peritoneal / intravenous chemo-therapy (cisplatin/paclitaxel). No recurrence until 24 months after primary debulking.
Patient 2	71	Carcino-sarcoma	IIIC	6 cycles adjuvant carboplatin mono-therapy. Progressive disease during adjuvant chemo-therapy. Patient died 11 months after primary debulking.
Patient 3	77	Poorly differentiated serous adeno-carcinoma	IV	3 cycles neo-adjuvant carboplatin monotherapy followed by interval debulking. 3 cycles adjuvant carboplatin monotherapy. Disease recurrence at 12 months after primary (incomplete) debulking. Patient died 18 months after primary debulking.

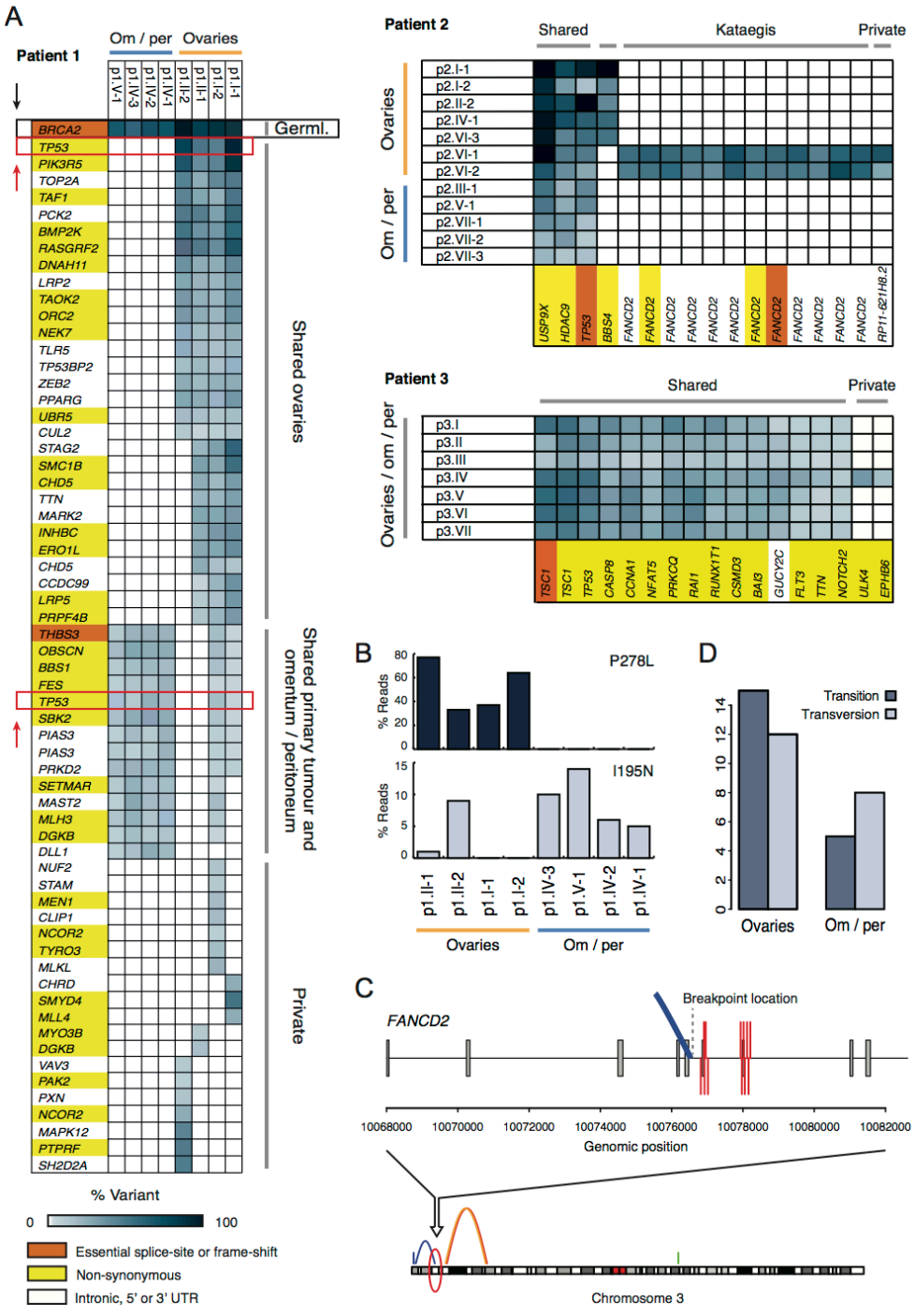
Heterogeneity of structural and copy number variation in treatment-naïve epithelial ovarian cancer

Ovarian cancer is notorious for its frequent genomic instability[116,126]. Whole genome mate-pair sequencing allows direct detection of genomic rearrangement breakpoints based on discordantly oriented and spaced mate-pairs (read pairs)[12]. We performed whole genome mate-pair sequencing using an insert-size of ~3kb for each of the biopsies in order to otain a detailed and comprehensive representation of the genomic instability within tumor samples from three ovarian cancer patients. We used a breakpoint detection algorithm that simultaneously clusters discordant mate-pair sequencing reads from all tumor biopsies per patient[96], allowing us to genotype breakpoints that are present at low frequency with relatively high sensitivity, i.e. based on a single discordant read pair once a robust call is made in another sample of the same patient. For example, given the median physical genomic coverage of ~50x, the data allow us to genotype a heterozygous breakpoint present in at least 5% and 14% of the tumor cells in samples with a tumor percentage of 90% and 30%, respectively. After stringent filtering and removal of SV calls present in matching normal control samples, we found between 120 and 369 somatic genomic rearrangements across the primary and metastatic tumor samples in three patients. We used PCR to validate a set of breakpoint calls and could confirm 95 out of 121 tested (>78% specificity).

For each somatic breakpoint detected by mate-pair sequencing, we determined the number of tumor samples that carried the breakpoint. For patient 1, this revealed that only 2/369 somatic breakpoints are shared between all samples, whereas the majority was found to be shared between only 4 or 5 of the 8 samples from this patient. In patient 2 the largest number of breakpoints is also shared between only 4 samples, but 34/120 breakpoints were shared between all 9 tumor samples. In patient 3 the vast majority of breakpoints was shared between 5-7 of the 7 tumor samples (**Figure 1B**). We then performed unsupervised hierarchical clustering using the breakpoint junctions detected across each of the biopsies per patient. For patient 1 and 2, this revealed two clusters of samples. Particularly patient 1 showed two extremely different branches. For both patients, one cluster contained all biopsies from the omentum and peritoneum, whereas the other cluster contained all biopsies from the ovaries, and for patient 2 also a biopsy from a tumor located in the pelvis. In contrast to the branching patterns detected in patient 1 and 2, a much more homogeneous pattern was detected in patient 3 (**Figure 1C**). Several breakpoints overlap with cancer genes from the Cancer Gene Census, including *NF1, FANCD2 and CDKN2A* and these are all targeted by breakpoints present in only subsets of samples[118]. For *FANCD2* and the cancer-related genes *ERBB4* and *ESR1*, which are targeted by breakpoints in patient 2, we observed a sample-specific effect of the breakpoint on gene expression. Expression of *ESR1* is a prognostic factor for survival in ovarian cancer[135]. Distinct patterns of genomic rearrangement classes are observed among different cancers[87,123,126].

Figure 1: *Somatic genomic rearrangements detected in patient 1 (left), 2 (middle) and 3 (right).*
(A) *Biopsy locations per patient. Ellipses indicate physically separated tumors; black dots represent biopsy locations. Ellipses are not indicative for tumor size. For patient 1 and 2, ellipses are colored according to the corresponding branch derived from the SV analysis (see panel C). Patient 1: ovaries (orange), om/per (blue), patient 2: ovaries/pelvis (orange), om/per (blue). Illustration © 2010 Terese Winslow, U.S. Govt. has certain rights.* ***(B)*** *Bar chart representing the distribution of the frequency of breakpoints per patient.* ***(C)*** *Heat map and clustering analysis of the detected somatic breakpoints per patient. Rows represent breakpoints, red and yellow bars indicate the presence (red) or absence (yellow) of the breakpoint in a sample. Om/per: omentum/peritoneum.* ***(D)*** *Distribution of somatic rearrangement types per branch for patient 1 and 2 and for all patient 3 samples.*


A
patient 1
patient 2
patient 3
p1.IV-1
p1.IV-2
p1.IV-3
p1.I-1
p1.I-2
p1.II-1
p1.II-2
p1.V-1
p2.VII-1
p2.VII-2
p2.VII-3
p2.II-2
p2.VI-3
p2.I-2
p2.I-1
p2.III-1
p2.IV-1
p2.VI-1
p2.VI-2
p2.V-1
p3.IV
p3.VI
p3.V
p3.II
p3.III
p3.I
p3.VII
B
Number of breakpoint junctions
Number of samples
Number of samples
Number of samples
C
Breakpoint present
no breakpoint
p1.II-1
p1.II-2
p1.I-1
p1.I-2
p1.IV-3
p1.V-1
p1.IV-2
p1.IV-1
Ovaries
Om / per
p2.V-1
p2.III-1
p2.VII-2
p2.VII-1
p2.VI-1
p2.II-2
p2.VI-3
p2.I-1
p2.IV-1
Om / per
Ovaries / pelvis
p3.VII
p3.III
p3.IV
p3.I
p3.II
p3.V
p3.VI
Ovaries / om / per
D
Tandem duplication type
Inter-chromosomal
Deletion-type
Inverted
percentage
Patient 1
Patient 2
Patient 3
Ovaries
Om / per
Ovaries / pelvis
Om / per
Ovaries / om / per


A
Patient 1
Om / per
Ovaries
p1.V-1
p1.IV-3
p1.IV-2
p1.IV-1
p1.II-2
p1.II-1
p1.I-2
p1.I-1
Germl.
BRCA2
TP53
PIK3R5
TOP2A
TAF1
PCK2
BMP2K
RASGRF2
DNAH11
LRP2
TAOK2
ORC2
NEK7
TLR5
TP53BP2
ZEB2
PPARG
UBR5
CUL2
STAG2
SMC1B
CHD5
TTN
MARK2
INHBC
ERO1L
CHD5
CCDC99
LRP5
PRPF4B
THBS3
OBSCN
BBS1
FES
TP53
SBK2
PIAS3
PIAS3
PRKD2
SETMAR
MAST2
MLH3
DGKB
DLL1
NUF2
STAM
MEN1
CLIP1
NCOR2
TYRO3
MLKL
CHRD
SMYD4
MLL4
MYO3B
DGKB
VAV3
PAK2
PXN
NCOR2
MAPK12
PTPRF
SH2D2A
Shared ovaries
Shared primary tumour and omentum / peritoneum
Private
% Variant
0
100
Essential splice-site or frame-shift
Non-synonymous
Intronic, 5' or 3' UTR
Patient 2
Shared
Kataegis
Private
Ovaries
Om / per
p2.I-1
p2.I-2
p2.II-2
p2.IV-1
p2.VI-3
p2.VI-1
p2.VI-2
p2.III-1
p2.V-1
p2.VII-1
p2.VII-2
p2.VII-3
USP9X
HDAC9
TP53
BBS4
FANCD2
Patient 3
Shared
Private
Ovaries / om / per
p3.I
p3.II
p3.III
p3.IV
p3.V
p3.VI
p3.VII
TSC1
TSC1
TP53
CASP8
CCNA1
NFAT5
PRKCQ
RAI1
RUNX1T1
CSMD3
BAI3
GUCY2C
FLT3
TTN
NOTCH2
ULK4
EPHB6
B
P278L
I195N
% Reads
p1.II-1
p1.II-2
p1.I-1
p1.I-2
p1.IV-3
p1.V-1
p1.IV-2
p1.IV-1
Ovaries
Om / per
C
FANCD2
Breakpoint location
10068000
10070000
10072000
10074000
10076000
10078000
10080000
10082000
Genomic position
Chromosome 3
D
Transition
Transversion
Ovaries
Om / per

Analysis of breakpoint types per patient revealed that deletions comprised the largest subset (40%) in patient 1, which is consistent with recent findings indicating an excess of deletions in ovarian cancer with germline *BRCA* mutations[126], as was the case for this patient (see below). However, more detailed analysis of branch-specific breakpoints for patient 1 revealed a shift in rearrangement types between the two branches despite their shared BRCA-status (**Figure 1D**). The cluster of four omental/peritoneal metastatic tumors showed a higher percentage of somatic tandem duplication type rearrangements and inversions, and a lower percentage of deletions and interchromosomal rearrangements when compared to the cluster containing the tumors on the ovaries. In line with this, we observed an increase in head-to-tail breakpoint junctions at a cost of tail-to-head junctions for patient 1 in the omental/peritoneal samples when compared to the samples that originated from both ovaries (data not shown). A difference in rearrangement types was, however, not apparent for the branches in patient 2. For patient 3 we observed that the majority (>40%) of somatic SVs comprise tandem duplications, in line with previous studies[126] (**Figure 1D**). Inter-sample differences in rearrangement signatures were not detected for patient 3.

To get further support for the dynamic patterns of heterogeneity revealed by the somatic genomic breakpoints in the ovarian cancers studied here, we used SNP-array genotyping and copy number analysis. We performed unsupervised hierarchical clustering of allele frequencies derived from the SNP-array genotyping data. The analysis includes all SNPs with differences in allele frequencies across each of the biopsies per patient (~10-30k SNPs) (**Methods**). Similar patterns of deviating allele frequencies for specific subsets of samples indicate shared ancestry, whereas diversity of these patterns rather suggests independent evolution. Heat map and clustering analysis of the allele frequencies of included SNPs confirmed the results from the mate-pair analysis for each patient.

Figure 2. *Somatic single nucleotide mutations analysis results for patient 1, 2 and 3.*
(A) *Regional distribution of mutations across tumor samples per patient. Blue gradient indicates the percentage of reads that carried the mutation. Gene colors indicate mutation impact: high, essential splice site or frame-shift (orange); medium, non-synonymous (yellow); or silent, intronic, 5' or 3' UTR (white).* ***(B)*** *Distribution of the two TP53 missense mutations detected in patient 1 (P278L and I195N) across all tumor samples of this patient.* ***(C)*** *Kataegis as detected in patient 2 samples p2.VI-1 and p1.VI-2. The 12 single nucleotide changes in FANCD2 coincide with a genomic breakpoint, which is solely detected in these samples.* ***(D)*** *Transitions versus transversions for patient 1. All ovarian samples (primary tumor (p1.I-1 and p1.I-2) and metastases located in the other ovary (p1.II-1 and p1.II-2)) versus the omentum/peritoneum metastases (p1.IV-1, p1.IV-2, p1.IV-3 and p1.V-1).*

Intra-tumor mutational profiles in ovarian cancer

Next, we screened coding sequences of a total of 2,099 cancer genes across each of the biopsies. We validated all identified mutations on all tumor and matching normal tissue samples using PCR-based resequencing on the MiSeq at >1000X coverage. The MiSeq data were also used to derive or refine mutation frequencies. We detected 63 somatic single nucleotide mutations in patient 1, and considerably fewer mutations in patients 2 and 3 (17 mutations per patient, **Figure 2A**). In patient 1, we also identified a *BRCA2* germline frameshift indel and concomitant LOH of chr 13 in the tumor biopsies.

All patients carried *TP53* mutations. In patient 2 and 3 a single *TP53* mutation was detected in all tumor samples per patient. Interestingly, we identified two different driver *TP53* missense mutations (P278L and I195N) in patient 1, occurring at distinct tumor locations. Both of these mutations have been described in the COSMIC database [136]. Only the samples derived from the right ovary (p1.I-1 and p1.I-2, the presumed primary tumor), contained both *TP53* mutations, albeit I195N was detected at low frequency (1-9% vs 33-77% for P278L) (**Figure 2B)**. We observed 19 mutations that were unique to only a single ovary tumor sample in patient 1 (private mutations). However, none of the four metastases in the omentum and peritoneum (p1.IV-1 to p1.IV-3 and p1.V-1) carried private mutations. In fact, all mutations identified in the omentum and peritoneum cluster of samples were ubiquitous and with the exception of a mutation in *DLL1*, all variants were also identified in the samples at the right ovary.

For patient 2, 12 of the 17 mutations were in *FANCD2* and all 12 occurred in samples p2.VI-1 and p2.VI-2 within a window of 1.2kb and comprising characteristic TpCpX trinucleotides, likely resulting from kataegis [137]. Similar to kataegis described in breast cancer, these mutations coincided with SV uniquely present in these two tumor samples (**Figure 2C**). Most of the other coding variants identified in patient 2 were shared between all samples.

The majority of mutations in patient 3 were ubiquitous. Two mutations occurred in the tumor suppressor gene *TSC1* (one missense, P141R, and one essential splice mutation). The *TSC1* splice site mutation leads to truncation of the *TSC1* transcript, suggesting deregulation of mTOR signalling as a possible contributor to tumor development in patient 3 (**Supplementary Figure 1**) [138]. An additional mutation was found in *CSMD3* in this patient, which is frequently mutated in ovarian cancer and non-small cell lung cancer, although the functional role of this gene in tumor formation is not clear [116,139]. Only two private cancer gene single nucleotide mutations were found in patient 3 (sample p3.IV). Overall, the single nucleotide mutation data revealed a similar pattern of genetic heterogeneity as the structural and copy number variation data for patient 1; one cluster of mutations

occurred at metastatic tumor sites in the omentum and peritoneum and another cluster of mutations was found in the tumors in the left ovary. Both clusters shared mutations with the presumed primary tumor samples (p1.l-1 and p1.l-2). Interestingly, we found a difference in the Transition/Transversion (Ti/Tv) ratio for the two branches in this patient (**Figure 2D**), suggesting that distinct mutational forces acted in different branches of the ovarian tumor in patient 1 [140]. For patients 2 and 3 we found much fewer mutations and several of these were present in all samples. A mutation in *BBS4* further supported the branching pattern observed in patient 2.

Gene expression differences across ovarian cancer biopsies reveal intra-tumor subtypes and branch-specific pathway activation

Gene expression profiling of ovarian cancer enabled classification in distinct subtypes associated with differences in survival and therapy resistance [128,129]. We performed RNA sequencing to measure gene expression across each of the tumor biopsies and we detected between 1000-1300 differentially expressed genes per sample compared to all other samples of the same patient. Hierarchical clustering of gene expression differences for each of the patients revealed two major branches for both patient 1 and patient 2. For patient 3 the clustering of differentially expressed genes across tumor biopsies did not reveal any distinct subgroups. The clustering of samples based on RNA expression differences further substantiated intra-tumor diversity as observed based on genomic breakpoints. We used the normalized coverage for 1500 genes that define six different epithelial ovarian cancer subtypes to classify each of the tumor biopsies from patients 1 to 3 [116,128]. The tumor biopsies from patient 1 fall apart into distinct subtypes following the branching we observed based on clustering of genomic and transcriptomic data: the samples from the omentum and peritoneum clearly display the C1 (high stromal) signature, overlapping with the C2 (high immune) signature as described before [128], whereas samples from the ovaries rather fall into the C4 (low stromal response) category although some expression of genes in the high immune signature can also be observed (**Figure 3A**). The samples from patient 3 display the C2 and C4 gene expression signatures. Patient 2 samples explicitly show the C5 (mesenchymal) signature, as expected from the histological examination, which indicated a carcinosarcoma. The distant metastases of patient 2 are different from the ovary and pelvis samples as they also show overlap with the C2 category.

To determine whether single nucleotide changes observed in the genome were also found in expressed transcripts we analysed the frequency for each of the identified mutations among RNA sequencing reads for patient 1 (**Figure 3B**). This analysis showed that some single nucleotide changes present at the DNA level are not expressed. In addition, we also find that alleles with single nucleotide variants

detected at low frequencies at the DNA level are expressed at very high levels. For example, the shared *TP53* I195N variant and the private *BBS1* and *MEN1* variants are expressed at a much higher frequency in the RNA, suggesting that they are relevant for tumor growth.

Based on the top 5% most significantly differentially expressed genes we used Cytoscape software to evaluate whether specific cellular pathways or processes have altered expression in any of the branches or samples [114]. All branches showed activation of pathways or processes related to cancer development compared to the total pool of non-tumor samples of all three patients, such as the different aspects of cell division (e.g. cell cycle checkpoints, DNA replication, chromosome segregation) and growth factor and P53 signalling. Specific pathway activation was observed for the two branches in patient 1 and 2 (**Figures 3C**). The samples in the ovarian cluster from patient 1 expressed significantly higher levels of genes involved in ERBB signalling and post-translational protein modification compared to the samples in the omentum/peritoneum cluster. Upregulation of Hedgehog/WNT/Cadherin genes was observed in the ovarian and pelvic samples of patient 2. Interestingly, samples from the omentum and the peritoneum in both patients 1 and 2 had many pathways commonly upregulated compared to the other samples in these patients, including chemokine signalling and cytokine-cytokine receptor interactions, immune response, extracellular matrix organization and integrin signalling. Finally, cell adhesion (CAM) was one of the processes enriched for in the ovary and pelvis samples in patients 1 and 2, although the exact genes upregulated in the samples from these two patients differ.

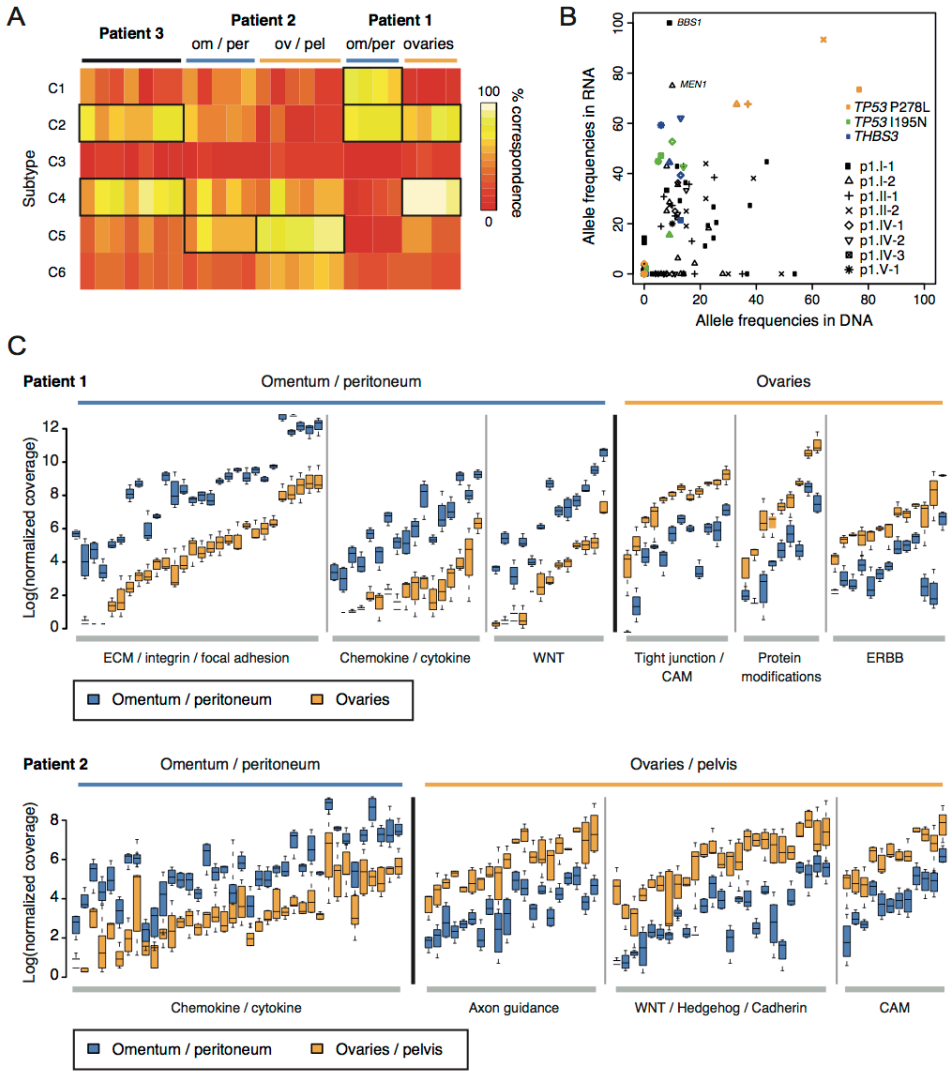


Figure 3. *Gene expression analyses across biopsy sites.*
(A) *Heatmap of the percentage concordance of each sample with the subtypes of ovarian cancer presented by Tothill et al*[104,118,128]*. C1= High stromal response, C2= High immune signature, C3= Low malignant potential (LMP) signature, C4= Low stromal response, C5= Mesenchymal signature, C6= Low grade endometrioid.*
(B) *Allele frequencies of coding mutations in RNA and DNA for patient 1.*
(C) *Branch-specific expression differences of genes involved in major signaling pathways for patient 1 (top) and patient 2 (bottom).*

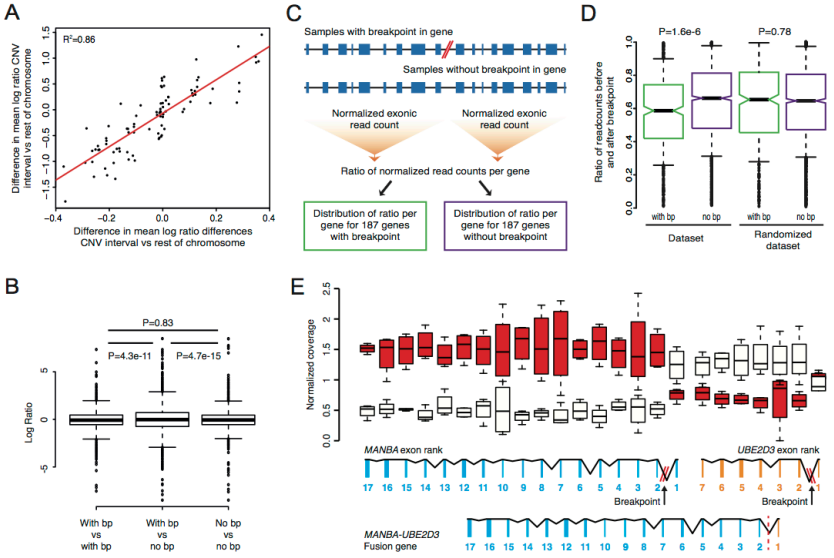


Figure 4. *Intra-tumor differences in gene expression resulting from genomic rearrangements in patient 1.*
(A) *Pairwise comparison of copy number changes and gene expression changes.*
(B) *Boxplot showing log ratios derived from pairwise comparisons of patient 1 samples, categorized in three bins: 1) both samples have a breakpoint, 2) one sample has the breakpoint and the other does not have the breakpoint, 3) both samples do not have a breakpoint. Statistical testing of differences in variance was performed using Levene's test.* ***(C)*** *Schematic representation of a method used to detect expression differences of exons before and after a breakpoint in a gene. Per gene, the ratio of the normalized exonic read count before and after the breakpoint was determined for each of the samples from patient 1. Ratios were separated in two bins: one containing ratios derived from genes with a breakpoint and one containing ratios derived from genes without a breakpoint.*
(D) *Boxplot of the distribution of ratios of the normalized exonic read count before and after a breakpoint for genes that contain a breakpoint (with bp) and genes that do not contain a breakpoint (no bp). The analysis was repeated by randomly assigning breakpoints to samples (randomized dataset). Statistical testing was performed using a Mann-Whitney U test.* ***(E)*** *Changes in gene expression for the exons of the MANBA and UBE2D3 gene exons in patient 1. In the presence of the deletion breakpoint a MANBA-UBE2D3 fusion gene is formed. Red: breakpoint present, white: no breakpoint present.*

Genomic heterogeneity causes intra-tumor differences in gene expression

The marked differences in gene expression observed across biopsies from the same patient prompted us to analyse the contribution of genomic rearrangements to these differences, an aspect of tumor biology which is poorly understood. Based on copy number profiling we identified 14 large copy number gains and losses (range: 0.97-28 Mb) that were only present in subsets of samples from patient 1 and patient 2. For each of these copy number changes we determined the mean of the differences in $\log_2$ ratios from SNP-arrays and compared these to the mean of the $\log_2$ ratios derived from the RNA sequencing. We observed a strong correlation between copy number changes and gene expression changes based on pairwise comparisons (**Figures 4A**), indicating that gene expression is strongly influenced by DNA copy number. However, on a genome-wide basis, only 1.5-1.8% of the differentially expressed genes are within the boundaries of the large copy number changes for patient 1 and 2 respectively. Thus other factors, such as mutations, translocations, epigenetic changes or secondary effects are likely contributing to the intra-tumor gene expression differences.

To further study the effect of genomic breakpoints on gene expression, we utilized the precise breakpoint definition provided by mate-pair sequencing. We reasoned that an expression effect of a breakpoint should result in either a positive, or negative change in gene expression in samples with the breakpoint relative to samples without the breakpoint. To test this, we used the gene expression differences ($\log_2$ ratios) derived from pairwise comparisons based on each of the samples from patient 1 and categorized the comparisons in three bins: 1) both samples have a breakpoint, 2) one sample has the breakpoint and the other does not have the breakpoint and 3) both samples do not have a breakpoint. If genomic breakpoints have an effect on expression of the respective genes, we would anticipate an overall increase in fold changes in bin 2 versus bin 1 and 3. Indeed, we observed a significant increase of the variance of the distribution of fold changes in bin 2, indicating that on average, genomic rearrangement breakpoints affect gene-expression both positively and negatively (**Figure 4B**).

To get a more precise picture of the effects of SVs on gene expression, we measured the normalized read counts for exons located before and after an SV breakpoint in a gene **(Figure 4C)**. A shift in the ratio of the read count before and after the breakpoint would be expected if the expression of the exons before and/or after a breakpoint has changed as a result of the breakpoint. E.g. part of a gene could be upregulated due to fusion with another partner gene or a decrease in expression could be expected if one half of a gene is deleted. Furthermore, measuring this ratio allows us to solely detect the effect of a breakpoint in the gene and exclude influences of other factors on gene expression (e.g. neighbouring breakpoints, promoter methylation). **Figure 4D** shows a boxplot of the distribution

of ratios for the genes that do not contain a breakpoint and the genes that do contain a breakpoint, indicating a marked shift in distribution towards more extreme ratios for all genes containing a breakpoint. We repeated the same analysis by randomly assigning breakpoints to samples. In this case the distribution of ratios is the same for genes with and without breakpoints indicating the specific effects of the breakpoints on gene expression (**Figure 4D**). These results emphasize that expression measurements should not be determined only on a per gene basis, because subtle intra-gene expression differences due to rearrangements will be obscured for whole-gene measurements.

The changes in exon expression for samples with and without a deletion breakpoint in the *MANBA* and *UBE2D3* genes (patient 1) illustrate the sensitivity of the ratio analysis **(Figure 4E)**. Exons at the 3' end of *MANBA* are expressed higher in samples with the breakpoint relative to samples without the breakpoint and the reverse is true for exon 1. Similar effects were found for the *UBE2D3* gene. An *UBE2D3-MANBA* in frame fusion gene resulted from the somatic deletion in the ovary samples and the fusion transcript was expressed as verified by RT-PCR. We systematically searched for genomic rearrangements that are predicted to result in gene fusions and found 28 putative fusions in patient 1. For 12 of the predicted rearrangements, we designed PCR primers for RT-PCR across the breakpoint junctions and we could confirm expression of 7 fusion transcripts of which four were in frame and differentially expressed in tumor biopsies. Among these is one fusion containing the *MAST4* kinase. MAST kinases are involved in recurrent fusions in breast cancer and enhance cell proliferation[115].

Discussion

We here show that treatment-naïve epithelial ovarian cancer, whether serous adenocarcinoma or carcinosarcoma, may display extensive intrinsic genomic and transcriptomic heterogeneity, leading to a broad variety and potentially functional lesion-specific deregulation of cellular pathways. The major genomic and transcriptomic differences were found between distant metastasis located at the omentum or peritoneum versus tumor samples at the ovaries and pelvis, substantiating previous evidence for intra-tumor heterogeneity in serous epithelial ovarian cancer [117,119]. The most striking heterogeneity was found for patient 1, where both the mutation and the SV data supported two subsets of tumor biopsies. This included two independent *TP53* mutations, raising the question as to whether two separately initiated and evolving tumors had occurred in patient 1, with tissue from both tumors interwoven at the right ovary. So-called collision tumors have been described before and these are marked by histologically distinct tumors separated by stroma or basal lamina [141]. Both ovaries are frequently affected in serous ovarian cancer (65%) [142] and in rare cases this involves tumors with bi-focal

origin [143], which is a possibility we cannot fully exclude in our case. Both scenarios were not obvious from the histopathological examination. Interestingly, patient 1 harbored a germline *BRCA2* mutation. BRCA2 functions in DNA repair and disruption of *BRCA2* leads to genomic instability [144]. Therefore, this mutation could possibly have promoted the very early separation of the tumor samples from this patient as opposed to the more coherent evolutionary patterns in patients 2 and 3.

Multi-site profiling of genomic changes allows estimation of the evolutionary course of cancer development, including timing of mutational events [37,120]. *TP53* mutations were present in all samples from the three patients, indicating that these occurred early during tumor evolution. For patient 1, we observed two subsets of samples, which constitute two independent tumors or very early branched subclones (with independent *TP53* mutations). This evolutionary pattern was supported by both mutation data and SV data. Within each of the branches we observed much more coherence than between the branches: we observed no unique mutations and just one unique rearrangement for the omentum and peritoneum samples. For the ovary samples of patient 1 we found several unique changes. This included 19 private mutations and 48 private genomic rearrangements all of which likely occurred late during tumor evolution, demonstrating continuous evolution at both the structural and mutational level in this branch. Furthermore, we observed indications of different mutational mechanisms operating in each of the two subsets in patient 1, both at the level of genomic rearrangements and point mutations. For patient 3, we observed only two private mutations, whereas all other mutations (15) were shared between samples. A large overlap between samples was also supported by the SV data from patient 3. However, we did observe 22 unique rearrangements, suggesting ongoing evolution at separate sites at the level of genomic rearrangements. For patient 2, the SV data showed the presence of two subsets of samples. However, a large fraction (34/120) of somatic SVs were found in all samples indicating a common evolutionary origin as opposed to the very early branching observed for patient 1. The common origin and branching in patient 2 was further supported by the presence of mutations shared by all samples and a unique coding mutation in the ovary/pelvis branch, respectively. In addition, we observed ongoing evolution in two pelvis samples based on the observation of a condensed cluster of mutations in *FANCD2* coinciding with SVs, a mutational process which has been termed 'kataegis' (**Figure 2C**) [137]. These data show that kataegis may act only regionally within the tumor of a single patient. Similarly, we have previously also observed that chromothripsis may exclusively occur in either primary, or metastatic tumor samples from the same patient [145], demonstrating that massive mutation mechanisms may occur late during tumor development and do not necessarily represent an initiating event.

Whole genome sequencing should reveal more single nucleotide changes and provide further insight into possible differences in evolutionary timing relative to structural variations in ovarian cancer.

There is a strong need for improved and targeted therapies for ovarian cancer to increase cure rates [130,131,146]. Several targeted therapies are being tested, but careful selection of patients for targeted treatment is essential [147]. We show here that there is major intra-tumor heterogeneity concerning expression of cellular pathways, some of which are candidates for targeted treatment. For example, overexpression of the Hedgehog pathway was observed in a subset of metastases in the omentum and pelvic region of patient 2, compared to other pelvic lesions and tumor sites in the ovaries. High expression of the hedgehog transcription factor *GLI1* is associated with poor survival in advanced serous ovarian cancer [148] and Hedgehog components are deregulated in various sarcomas, presenting new treatment possibilities such as Hedgehog ligand antagonists and inhibition of Gli transcription activity [149]. Also, we detected strong upregulation of integrin pathway members in peritoneum and omentum metastases in patient 1, as well as elevated expression of inflammatory chemokines and cytokines primarily found in metastases in the omentum of both patients 1 and 2. Expression differences in these genes in metastatic lesions compared to primary tumors have previously been observed [150]. These may provide an attractive target for treatment of ovarian cancer [131,151,152], because the vast majority of patients die as consequence of metastatic disease, while the primary tumor is often completely removed during debulking surgery.

The cancer genome harbours a wide variety of genomic alterations. Particularly the contribution of structural genomic rearrangements to tumor development is poorly understood. We here demonstrate that intra-tumor heterogeneity may involve cancer genes disrupted by genomic breakpoints present in only a subset of tumor masses. Furthermore, we associated the intra-tumor expression differences with genomic rearrangement breakpoints and found that effects may range from altered expression due to copy number changes of entire genes to very subtle effects involving breakpoints affecting only part of a gene, all occurring within a single patient. This data shows that the effects of genomic rearrangements are profound, contribute to intra-tumor heterogeneity and may be equally important as coding mutations for tumor development. Because our study only covered multi-site analysis of three ovarian cancer patients, it remains to be seen how representative the identified genomic and transcriptomic characteristics will be in a larger sample set. Large-scale follow-up studies should be conducted to determine the rate of extreme intra-tumor heterogeneity in ovarian cancer. This aspect of tumor biology requires further attention to fully understand escape routes as a response to treatment and improve survival rates.

Methods

Patient sampling and consent

All patients included are epithelial ovarian cancer FIGO stage III/IV patients undergoing primary cytoreductive debulking according to the standard of care (**Table 1**). During primary debulking, tumor samples were obtained from thoroughly documented locations. Patient 1 and 2 underwent successful optimal debulking. In patient 3, optimal debulking appeared infeasible perioperatively and only partial cytoreduction was achieved. She later underwent successful cytoreduction after neo-adjuvant chemotherapy. We only included the tumor samples obtained from the first debulking for patient 3. In addition, DNA from blood and saliva was obtained from every patient as control samples (Oragene DNA kit, DNA Genotek Inc., Ottawa, ON, Canada). Tumor samples were immediately forwarded from the operating room to the pathology department. If feasible, multiple core and peripheral tumor samples of each individual metastasis or primary tumor were snap frozen in liquid isopentane within one hour. Tissue was processed into frozen sections and stained by Hematoxylin & Eosin (H&E). A pathologist reviewed all slides and confirmed tumor type, estimated tumor cell percentage and amount of necrosis. This study was approved by the ethics committee of the UMC Utrecht, The Netherlands. Patients could indicate in a specific section of the informed consent form that they wanted to be informed about incidental findings in their germline DNA that could affect their health, or the health of their relatives. All patients signed informed consent before debulking.

DNA, RNA isolation

Fresh frozen samples were homogenized and subsequently split for independent DNA and RNA isolation. DNA was isolated using the QIAGEN® Genomic DNA kit (QIAGEN, Hilden, Germany). Total RNA was isolated using TRIzol® reagent (Life Technologies, Carlsbad, California, USA). After isolation, DNA samples were stored at -20°C, RNA samples at -80°C.

Mate-pair sequencing

Mate-paired libraries were generated from 5-10 µg of DNA isolated from tumor and control samples using the 5500 SOLiD™ Mate-Paired library kit (Life Technologies, Carlsbad, California, USA). Samples were sheared to 3kb fragments by Hydroshear DNA shearing (Digilab, Marlborough, Massachusetts, USA). Per library, 2 x 50-bp mates were sequenced on a SOLiD 5500xl or SOLiD WildFire instrument. Forward and reverse tags were mapped independently (samse) to the reference genome (GRCh37) using BWA software and settings –c –l 25 –k 2 –n 10 [110]. Discordant reads were clustered using in-house software as described previously [96]. The software is available from https://github.com/Vityay/1-2-3-SV. In a first step, we estimated the insert size distribution and location of discordant

mate-pairs. This was done separately for each sample. Furthermore, PCR duplicates, reads with mapping quality 0 and non-uniquely mapped reads were removed from further analysis. As a second step clustering of discordant pairs was done for all samples from each patient together. Two pairs are considered to belong to the same cluster when the distance between coordinates of their 5-prime tags together with the distance between 3-prime tags does not exceed the median distance of the library with the largest insert size. The search is continued until no clusters with at least five clones in at least one of the samples can be found. As analysis of discordant read pairs does not give exact breakpoints of structural variants, the output lists genome segments containing each breakpoint along with information about the source of the discordant pairs (samples) and the properties of the cluster. The orientation of the different mate-pair tags in a cluster relative to each other is indicated by H (or h for the minus strand) when the tag has its 'head' side (the side that points towards the start of the chromosome) opposed to the pairing tag and T (or t for the minus strand) when a tag has its 'tail' side (the side that points towards the end of the chromosome) opposed to the pairing tag. The clustering results in calling of intrachromosomal rearrangements (deletion type, inverted, tandem duplication type) and interchromosomal rearrangements. To select for somatic variants, all genomic rearrangement breakpoints were filtered for normal tissue samples (blood, muscle, tuba) and an in-house database of mate-pair sequencing data from healthy individuals. To achieve high-quality calling of somatic structural variants, we required at least 5 independent discordant sequence reads derived from at least one tumor sample [96,145]. For all breakpoint calls consistent with these criteria, presence of the breakpoint in other samples from the same patient was determined based on presence of at least one overlapping discordant read pair with the same orientation. Primers for PCR confirmation of somatic breakpoints were designed based on mate-pair sequencing data. PCRs were performed under standard conditions as described before [96].

Cancer gene resequencing

SNVs and indels were detected by targeted sequencing of a total of 2099 cancer genes. First, samples were interrogated by a designed "Cancer mini-genome" consisting of 1,977 cancer genes. Barcoded fragment libraries were generated from 2 µg of isolated DNA from tumor and control samples as previously described [153]. Pools of libraries were enriched for 1,977 cancer-related genes (Cancer mini-genome [154]) using SureSelect technology (Agilent, USA). Enriched libraries were sequenced on a SOLiD 5500xl or SOLiD WildFire instrument according to the manufacturers protocol. Furthermore, the exons within a subset of 409 oncogenes and tumor suppressor genes were interrogated by the Ion Ampliseq™ Comprehensive Cancer Panel (Life Technologies, Carlsbad, California, USA).

Libraries were constructed from 40 ng of isolated DNA for each sample using standerd AmpliSeq procedures. Barcoded libraries were pooled and sequenced on the Ion PGM™ Sequencer (Life Technologies, Carlsbad, California, USA). Ion Torrent reads were aligned to the human reference genome version 19 (GRCh37) using Tmap. Variant calling on Ion Torrent data was performed using Strelka [155]. SOLiD reads were mapped on the same genome version, using BWA (-c –l 25 –k 2 –n 10) and variant calling was done using a custom pipeline identifying variants with at least 10x coverage, a 15% allele frequency and multiple (>2) occurrences in the seed (the first 25bp most accurately mapped part of the read) as well as support from independent reads (>3). All variant positions identified in either SOLiD or Ion Torrent data were subsequently genotyped in the raw datasets of both techniques for all samples using samtools mpileup, to ensure the presence or absence of possible low-frequency variants. Validation of single nucleotide mutations and indels was performed by PCR amplification of mutation loci followed by Nextera XT library prep and sequencing on MiSeq (Illumina, San Diego, California).

SNP-array analysis

For each sample 200ng DNA was used as in input for copy number profiling using Cyto12 SNP arrays according to standard procedures (Illumina, San Diego, California, USA). Genomic events were identified by applying ASCAT processing (Allele Specific Copy Number Analysis of Tumors) with Nexus Copy Number 6.0 (BioDiscovery, Hawthorne, CA, USA). Briefly, all signals in tumor samples similar to those in the provided reference sample are excluded from analysis, increasing specificity in detecting additional events in tumor samples. Clustering of SNParray data was done by calculating a Euclidean distance matrix based on B-allele frequencies of all SNP positions showing a heterozygous genotype in the reference samples (allele frequency 20% <> 80%), and performing hierarchical clustering on these data using standard R functions.

RNA sequencing

Total RNA was processed with the Poly(A)Purist™ Kit (Life Technologies, Carlsbad, California, USA) to select for poly(A)$^+$ RNA. Next, the mRNA-ONLY Eukaryotic mRNA Isolation kit (Illumina, San Diego, California, USA) was used to select for 5' capped mRNA. Paired-end libraries were constructed from 8-30 µg total RNA per sample, using the SOLiD total RNAseq kit (Life Technologies, Carlsbad, California, USA). Libraries were barcoded and sequenced on a SOLiD 5500xl instrument in paired-end mode (50x35bp). Forward (F3) and reverse (F5) reads were mapped independently to the human reference genome (GRCh37) using BWA (-c –l 25 –k 2 –n 10) [110]. Coverage per gene was determined by adding up read counts of all coding regions as determined by BEDtools 2.16.2 multicov (Quinlan and Hall, Bioinformatics 2010).

Because of the lack of a true reference sample i.e. healthy tissue from the ovaries of each patient, we compared each sample separately against the pool of all other samples of the same patient and determined gene expression differences within patients using the DegSeq R package [156]. To identify the most significant expression differences, DegSeq output was first corrected for multiple testing by selecting genes with a p-value < $4.6399e10^{-08}$ (0.001 divided by 21552 (the number of tested genes)) and extracting the top 5% and bottom 5% of normalized log ratios. Normalized coverages of the genes emerging from this analysis were used to cluster samples based on Poisson Mixture Models generated with the HTSCluster R package [157]. Pathway and GO biological process enrichment of upregulated genes within the thus created clusters was determined by comparing the core samples in these clusters to a pool of all reference samples from our 3 patients and finally to each other by DEGseq and selecting the top 5% genes as described above. Resulting genelists were analyzed through the Reactome FI Cytoscape Plugin [78,114], and pathways or processes containing at least 2 upregulated genes and a false discovery rate (FDR) < 0.1 were reported as being affected. We performed molecular subtyping of samples by calculating the percentage of concordance of up and down regulated genes (compared to the common reference pool) with the profiles presented by Tothill et al [22,116,128].

Data access

The SNP-array data are available from NCBI Gene expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE47633 and the sequencing data are available from the European Nucleotide Archive (ENA; http://www.ebi.ac.uk/ena/) under accession number ERP003455.

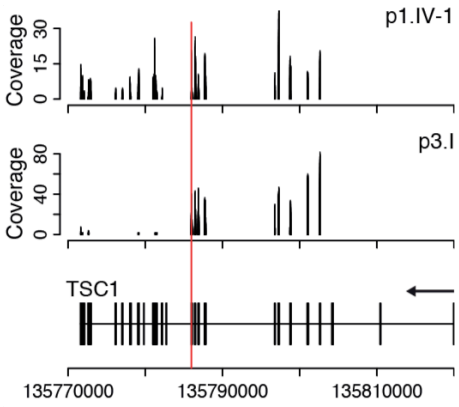
Competing Interests

Tim Harkins and Clarence Lee are full time employees of Life Technologies, a corporation that is commercializing SOLiD and Ion Torrent sequencing technology.

Acknowledgements

We thank Nico Lansu and Ewart the Bruijn for generating sequencing data on SOLiD WildFire, Lars van der Veken for support with analysis of SNP-array data and Maarten van Iterson for help with RNA-seq analysis. This work was financially supported by the Cancer Genomics Centre, which is part of the Netherlands Genomics Initiative, and the personalized cancer care priority program of the University Medical Centre Utrecht. We would like to acknowledge the Netherlands Center for Personalized Cancer Treatment for providing infrastructure access and expertise.

***Supplementary figure 1**. Effect of the splice site mutation in TSC1 on gene expression.*



Normalized read counts for coding regions in TSC1 derived from RNA sequencing data of samples p1.IV-1 (top panel) and p3.I (middle panel). Sample p3.I contains a splice site mutation (red line), which is detected in all patient 3 samples and leads to truncation of the TSC1 transcript.

5

Longitudinal imaging and sequence analysis of melanoma metastases reveals heterogeneity in vemurafenib response driven by diverse resistance mechanisms.

Marlous Hoogstraat[1,6,*], Christa G. Gadellaa-van Hooijdonk[1,6,*], Mary-Ann El Sharouni[5,*], Inge Ubink[1,6], Nicolle J.M. Besselink[1,6], Mark Pieterse[1], Wouter Veldhuis[3], Marijn van Stralen[4], Eelco F.J. Meijer[1], Stefan M. Willems[5,6], Michael A. Hadders[1], Marco J. Koudijs[1,6], Edwin Cuppen[2,6], Emile E. Voest[1,6], Martijn P. Lolkema[1,6].

* These authors contributed equally

[1]Department of Medical Oncology, University Medical Center Utrecht, The Netherlands
[2]Department of Medical Genetics, University Medical Center Utrecht, The Netherlands
[3] Department of Radiology, University Medical Center Utrecht, The Netherlands
[4]Image Sciences Institute, University Medical Center Utrecht, The Netherlands
[5]Department of Pathology, University Medical Center Utrecht, The Netherlands
[6] Netherlands Center for Personalized Cancer Treatment, the Netherlands

Abstract

Vemurafenib has improved the quality of life and life expectancy for many patients with *BRAF* mutant metastatic melanoma. Unfortunately in time the majority of patients become resistant to treatment. We followed ten patients during treatment with vemurafenib, by performing three-dimensional imaging to monitor growth of individual metastatic lesions. Next generation sequencing was performed on sequential biopsies, obtained prior to treatment and upon disease progression in 4 patients to uncover mechanisms of resistance to vemurafenib.

In all patients, only a subset of metastatic lesions developed resistance to vemurafenib while other lesions remained in regression. In two of four patients with repetitive biopsies from progressive lesions we identified mutations that explained resistance to vemurafenib: one patient had an activating mutation in *PIK3CA*, and in another patient we detected a secondary, novel *BRAF L505H* mutation. The functional relevance of this mutation for resistance to vemurafenib was subsequently confirmed in melanoma cells genetically modified to harbor the same mutations.

Our results demonstrate both power and need for sequential biopsies of progressive metastases in melanoma patients to detect known and novel resistance mechanisms. Such knowledge is very important for tailoring treatment and designing novel personalized treatment strategies.

Introduction

Approximately 50% of all melanomas are driven by V600 mutations in *BRAF*. Vemurafenib is the first clinically approved drug that specifically inhibits activated *BRAF* both *in vitro* and *in vivo*[158,159]. This compound produces a response in approximately half the patients and significantly improves progression-free and overall survival. Unfortunately, the majority of patients with metastatic melanoma inevitably develop resistance to vemurafenib after a median time of 6 months[33,44].

Intratumoral heterogeneity and extensive genetic variation between tumors within a single patient[37,40,85,160,161] have been proposed as reasons for treatment failure. Clonal outgrowth of resistant populations of tumor cells is considered the cause of tumor progression. Other studies using sequential samples obtained before treatment and at time of progression show that selective pressure of treatment activates compensatory survival pathways[162,163]. Based on these studies, we hypothesized that separate metastatic lesions within an individual patient can respond differently to vemurafenib due to inter-metastatic genetic heterogeneity and clonal variations, and investigating these genetic differences can lead to an improved understanding of resistance and hence improved treatment. To test this hypothesis, we used a unique longitudinal imaging approach combined with genome analyses. We studied the radiological response of individual metastases and the mutation profiles of pre-treatment samples and biopsy material upon progression from stage IV melanoma patients treated with vemurafenib.

Results and discussion

Disease progression under vemurafenib treatment is a localized event

To better define the tumor's response to vemurafenib, volumetric analysis of single metastatic lesions was performed. Ten patients were included in the volumetric response evaluation. Baseline characteristics of patients included in the volumetric response evaluation are listed in **Table 1**.

Per patient, three to ten separate lesions were measured every eight to twelve weeks during treatment. In total 51 computed tomography (CT) scans were measured; 57 target lesions were identified at baseline, including 25 lymph nodes metastases, 14 liver metastases, and 8 lung metastases. At initial evaluation, after two months of vemurafenib therapy, 45% of patients showed partial remission, 40% had stable disease, and 1 patient had progressive disease (**Figure 1A**). Progression free survival (PFS) ranged from 1.8 months to 21.9 months (median 7.8 months).

Table 1: Baseline characteristics of patients included in volumetric response evaluation.

Characteristic		n=10
Age - median (range)		55 (44-77)
Sex		
	Male - n (%)	8 (80)
	Female - n (%)	2 (20)
M stage at start vemurafenib*		
	M1a - n (%)	1 (10)
	M1b - n (%)	1 (10)
	M1c - n (%)	9 (90)
WHO** performance status		
	WHO 0 - n (%)	4 (40)
	WHO 1 - n (%)	6 (100)
Brain metastases at start vemurafenib		
	Present - n (%)	0 (0)
	Absent - n (%)	8 (80)
	Unknown - n (%)	2 (20)
Baseline LDH - U/L*** median (range)		243 (170-1641)
Previous systemic treatment		
	Yes - n (%)	0 (0)
	No - n (%)	10 (100)
Location primary tumor		
	Trunk	4 (40)
	Lower extremity - n (%)	4 (40)
	Head - n (%)	1 (10)
	Unknown - n (%)	1 (10)

*Stage according to AJCC melanoma staging and classification[164] **Performance status according to World Health Organization ***Lactate dehydrogenase, units per liter

Five patients developed progressive disease based on growth of target lesions identified at baseline. The other patients developed new lesions or showed evident progression of non-target lesions. One patient whose response could not be evaluated by volumetric measurement progressed due to an evidently progressive bone lesion. In each patient, volumetric analysis revealed a heterogeneous pattern of disease progression through tumor growth combined with multiple metastases that continued to respond to treatment. Remarkably, in these progressive patients, several lesions even became undetectable on CT scans (**Figure 1B**). Moreover, by evaluating the volumetric changes per lesion over time it was noticed that both the moment of progression and the growth rate of individual lesions could differ (**Figure 1C**), indicative of large functional heterogeneity and potentially different resistance mechanisms in separate lesions. These data show that a detailed intra-patient response evaluation reveals a fairly homogenous initial response after 2 months, but a heterogeneous growth pattern at progression. For the early detection of progression, two-dimensional RECIST criteria are not ideal since they induce a delay in the recognition of progression of individual sites. Volumetric, three-dimensional assessments are time consuming but have the benefit to identify growing lesions at an earlier stage and their validity has been shown by previous studies[165,166].

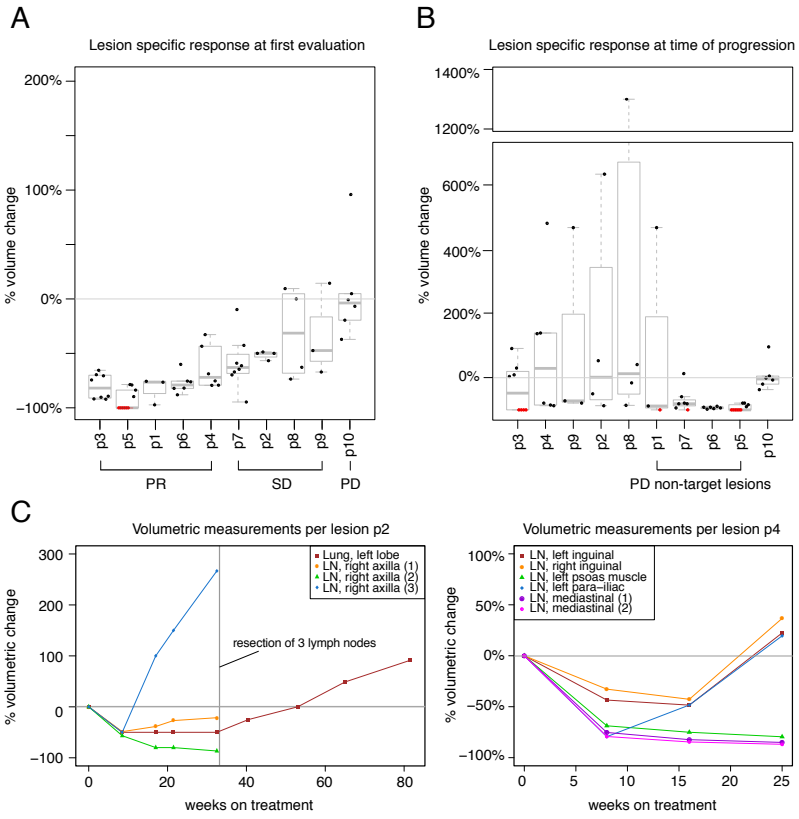


Figure 1: *Topographic diversity in the development of resistance to vemurafenib.* ***(A)*** *Volumetric measurements of all different lesions in 10 patients treated with vemurafenib at the first evaluation (after approximately 8 weeks of treatment), compared to baseline volume. Grey boxplots outline the dispersion of measurements per patient, while the black dots represent individual lesions. Data points highlighted in red indicate lesions undetectable on the CT-scans. Patients are indicated on the X-axis and sorted by initial response.* ***(B)*** *Volumetric measurements of the same patients at time of progression, compared to smallest volume measured for progressive lesions. All other lesions are compared to baseline. For each patient, multiple lesions were still smaller than baseline volume, indicating sensitivity of these lesions to vemurafenib. The lesion with increased volume shown for patient 1 was too small to account for progressive disease in itself, but progression was determined in combination with new lesions not shown in this graph.* ***(C)*** *Dynamic changes per lesion over time in 2 patients. Volume per lesion was measured every 8 to 12 weeks. In patient #2, a progressive lymph node was surgically removed after 32 weeks, together with two other lymph nodes nearby. After surgery a remaining lesion in the lung started to grow. Three lymph node lesions in patient #4 continued to respond to vemurafenib until the end of treatment, while three other lesions showed progression at the fourth evaluation at 25 weeks.*

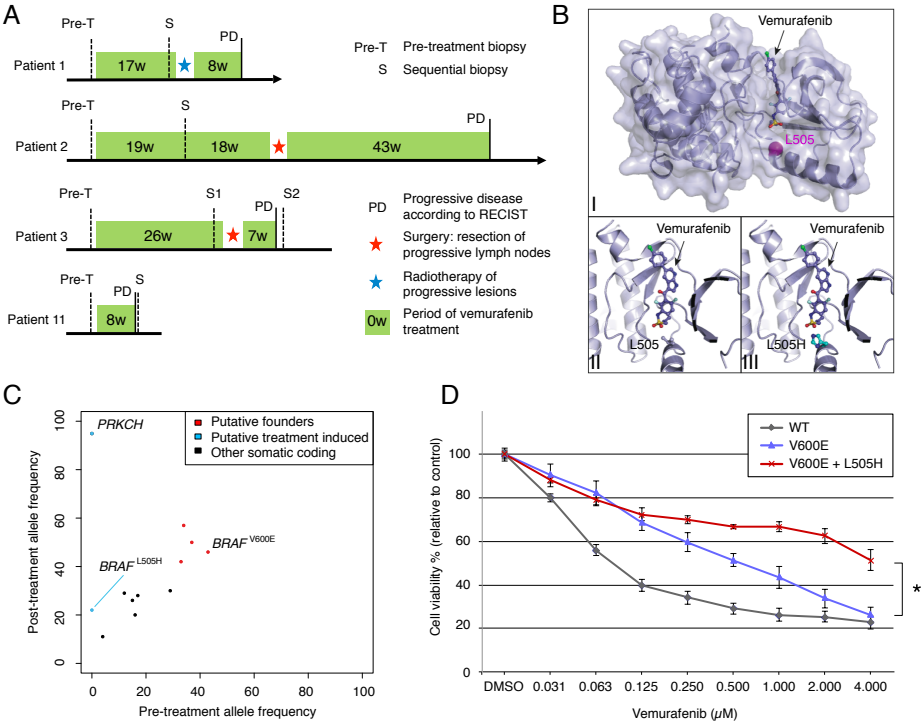


Figure 2: *Biopsies of progressive lesions within one patient revealed distinct mechanisms of therapy resistance.**(A)** Timing of biopsies and vemurafenib treatment in all 4 sequenced patients. Dotted lines indicate taking of biopsies. **(B) i)** Overview of BRAFV600E bound to vemurafenib [PDB:3OG7]. BRAF is depicted as a cartoon while vemurafenib is depicted in ball and stick representation. Leucine 505 is highlighted in magenta. **ii)** Close-up view of the ATP binding pocket of BRAFV600E with vemurafenib and leucine 505 highlighted in ball and stick representation. **iii)** Close-up view of the ATP binding pocket of BRAFV600E with vemurafenib and residue 505 highlighted in ball and stick representation. The wild type leucine has been mutated in silico to histidine depicted in cyan. The model reveals a putative clash between the histidine side chain and vemurafenib. **(C)** Comparison of allele frequencies of variants in the pre- and post treatment biopsy of patient #3. Putative founder mutations based on high allele frequency in both pre-treatment and sequential biopsy are highlighted in red; de novo mutations and mutations with increased allele frequency in the sequential biopsy are highlighted in blue. **(D)** Survival assay for A375 cells transduced with BRAF wildtype, BRAF V600E or BRAF V600E+L505H. A one-way ANOVA showed a significant difference (P= 0.0132) in cell viability between the cell lines with BRAF V600E constructs under treatment with vemurafenib compared to BRAF V600E+L505H.*

Mutational profiling of sequential biopsies reveals distinct mechanisms of resistance

Based on the volumetric measurements progressive lesions were identified and sequential biopsies were obtained in four patients (**Figure 2A**). In three out of the four patients, the sequential biopsy was taken from the same lesion as the pre-treatment biopsy (**Table 2**). All specimens contained 70-95% vital tumor cells based on histopathological examination. To identify the genetic defects in each tumor, the coding regions of 1,977 cancer related genes were sequenced to a median coverage of 150x from both pre- and post-treatment biopsies and a matched blood sample was analyzed as a reference.

Table 2: Overview of sites of sequential melanoma biopsies.

Patient	Time between biopsies (weeks)	Biopsy location		Tumor percentage	
		Baseline	Sequential	Baseline	Sequential
#1	16	Subcutaneous lesion thorax	Subcutaneous lesion thorax[a]	95	70
#2	19	Axillary lymph node	Axillary lymph node[a]	80	80
#3	24	Inguinal lymph node	Inguinal lymph node[a]	90	90
#11	8	Subcutaneous lesion breast	Soft tissue mass thoracic verte-bra 10-11	90	90

a: biopsy from same lesion as pre-treatment biopsy

Between 8 and 100 coding somatic variants could be identified per tumor sample. *BRAF* mutations were seen in all pre-treatment and sequential biopsies and mutations in DNA repair genes were found in 3 patients, including *MSH3*, *RAD50, ERCC3* and *ERCC4*. Both *ERCC3* and *ERCC4* are associated with melanoma and other skin cancers, according to the Sanger Cancer Gene Census[118].

To investigate whether there were any genetic changes between baseline and sequential biopsies, allele frequencies of variants in either sample were compared. While subsets of variant frequencies were stable, including the *BRAF* V600 mutation, unique mutations were also detected in samples from 2 patients, as well as large shifts in allele frequencies. By exploring mutations unique for the progressive lesions and mutations with increased allele frequencies, we sought to identify the resistance mechanisms in each tumor while the mutations with stable frequencies were considered as tumor driver mutations. Recent studies describing resistance against BRAF inhibition indicate several recurrent genetic mechanisms, including reactivation of the MAPK pathway through *MAP2K1*, *MAP2K2* or *NRAS* mutations[167-170], as well as activating mutations in *PIK3CA*[171,172] or other PI3K pathway alterations[173] and loss of function or aberrant splicing mutations in *RB1*,

PTEN[174] or *NF1*[175]. In one patient (#11), sequential biopsies showed three mutations that were not detected in the pre-treatment biopsy, including an activating mutation in *PIK3CA E545K* readily explaining the resistance. As the second biopsy was taken from a different lesion as the pre-treatment biopsy and the patient showed progression at the first evaluation, it is highly likely this mutation was present in the progressive lesion at the start of treatment. *PIK3CA* mutations in melanoma appear to be rare; data generated by the TCGA Research Network (http://cancergenome.nih.gov/) shows known activating *PIK3CA* mutations *H1047L* and *E545K* in three out of 342 cases, and van Allen *et al.*[173] discovered a *PIK3CA H1047R* mutation in one resistant tumor biopsy out of the 45 tested samples. Sanger sequencing of an in-house cohort consisting of 192 treatment-naïve samples from 158 patients revealed no additional *PIK3CA* or *BRAF L505H* mutations. If activating mutations in *PIK3CA* and additional mutations in *BRAF* mostly confer growth advantage to a melanoma cell under treatment stress, it is plausible that these events will not be detected frequently in treatment-naïve sample sets such as our cohort or the TCGA dataset, but observation frequencies might increase when more sequential biopsies are analyzed.

In contrast, while the sequential biopsy from patient #1 did not show any unique mutations compared to the pre-treatment sample, a large number of mutations with increased allele frequency were detected. This included multiple mutations in PI3K genes, such as *PIK3C2G*, *PIK3C2B* and *PIK3CG*, as well as mutations in transcription factors *TP73* and *TP63* belonging to the p53 family, a novel missense mutation in *CTNNB1* and a known activating mutation in *MAP2K1* (*MEK1 P124S*). While reactivation of the MAPK-pathway has been connected with resistance against BRAF inhibition, literature associating the *P124S* mutation described here with resistance is contradictory[163,167,176]. As endogenous beta-catenin seems to be required for effective BRAF inhibition[177], loss of function of *CTNNB1* could induce resistance. Another possible suspect in this patient is *CNKSR2*, member of a regulatory complex that facilitates activation of RAF[178,179]. This also comprises *ARAF* and *RAF1*, thus disruption of the regulatory complex could be a mechanism for the tumor to circumvent BRAF inhibition. However, no known resistance associated variants could be identified for this patient or patient #2.

Vemurafenib resistance through a de novo, secondary *BRAF* mutation

In patient #3, a secondary *BRAF* mutation (*L505H*) was detected in the sequential biopsy. Secondary mutations have rarely been reported although amplification of *BRAF V600E* has been documented as a known resistance mechanism[180]. The affected amino acid is located in the vicinity of the ATP-binding pocket, which is where vemurafenib engages *BRAF*. Modeling indicated a putative steric clash between the histidine imidazole ring and the propane-1-sulonamide moiety of

vemurafenib (**Figure 2B**). We therefore hypothesized that the *L505H* mutation leads to vemurafenib resistance by altering the binding affinity of the compound to *BRAF*. Our hypothesis was supported by a recent mutagenesis screen of *BRAF* where *BRAF V600E+L505H* was identified as secondary mutation that conferred resistance to vemurafenib in vitro[181]. Sanger sequencing of cDNA showed that both *BRAF* mutations were located on the same allele: 53% (n=33) of traces showed the *V600E* only, 23% (n=14) showed both the *V600E* and the *L505H* mutation, in 24% (n=15) *BRAF* was wild type, and in 0% (n=0) of the traces the *L505H* was found without presence of the *V600E*. Furthermore, a missense mutation in *PRKCH* was not present in the pre-treatment tumor sample but was detected in the post-treatment sample with an allele frequency of 95% (**Figure 2C**). This gene encodes protein kinase C (PKC) eta, and is involved in keratinocyte differentiation through activation of the MAPK pathway[182]. Moreover, PKCs have been shown to be involved in malignant melanoma upon activation through the Wnt signaling pathway[183].

Patient #3 underwent surgical resection, where not only the resistant, *BRAF V600E+L505H* mutation containing lymph node was removed, but also two other lymph nodes that were still sensitive to vemurafenib. Samples from these lymph nodes were interrogated using ultra-deep sequencing, and neither *PRKCH* nor the secondary *BRAF L505H* mutation could be detected while sequence coverage on these positions reached 17,000 and 40,000x respectively, which supported the model of the secondary *BRAF* mutation as a mechanism of resistance. Unfortunately, tissue from the primary tumor was not available for testing. Intriguingly, neither of these mutations was detected in a fourth lymph node lesion that became resistant to vemurafenib seven weeks later, suggesting yet another mechanism of resistance in this patient.

To establish if the *BRAF L505H* mutation can lead to resistance, *BRAF V600E* and *BRAF V600E+L505H* were overexpressed in the A375 *BRAF V600E* mutant melanoma cell lines, after which these cells were exposed to increasing concentrations of vemurafenib. Overexpression of *BRAF V600E+L505H* significantly increased cell viability upon vemurafenib exposure compared to overexpressed *BRAF V600E* (p=0.01, **Figure 2D**).

Clinical implications and future recommendations

There are two emerging concepts of resistance to targeted agents: clonal outgrowth of pre-existing resistant cell populations and the activation of compensatory survival pathways[173,184,185]. We were unable to generate convincing genetic evidence for negative selection. This suggests that sensitivity to BRAF inhibition seems to depend on the presence of a *BRAF* mutation whereas resistance could be based upon outgrowth of either insusceptible subclones or a

subpopulation of cells with a de novo mutation. The volumetric measurements make a significant contribution at this point: in some patients, such as patient #2, the lesions that developed resistance over time did not reach a volumetric decrease of > 50%, which would suggest that the resistance mechanism was already present at the start of treatment. In other patients, all lesions continued to respond to treatment for a long time after which a new metastasis developed, such as in patient #6 who was progression free for one year when a new metastasis was detected, rather indicating a de novo mutation. These clinical differences should be taken into account when treating patients with metastatic *BRAF* mutant melanoma. Incorporating detailed imaging modalities, such as volumetric assessments, detecting the rate of tumor shrinkage could be used to properly select early resistant sites for biopsy. As genetics alone does not reveal all potential resistance mechanisms the inclusion of RNA sequencing or (phospho)proteomics to probe expression of genetic variants would be a valuable addition to future studies. This would assist in exploring the downstream effects of genetic variants on gene expression and protein function.

The inherent heterogeneity of resistance mechanisms poses a challenge for rational design of treatment strategies to circumvent therapy resistance. It can be concluded that patients become only partially resistant to vemurafenib, which implies that some tumor lesions are still suppressed by the drug. Adding drugs that target the resistance pathways may therefore be more effective than completely replacing vemurafenib. The combination of BRAF and MEK inhibitors improves the outcome for *BRAF* mutant melanoma patients[45,174] and supports this premise. Furthermore, for those patients with overt genetic activation of the PI3K pathway, such as our patient #11 with an activating *PIK3CA* mutation, a combination of a BRAF and a PI3K inhibitor seems logical.

Conclusion

These results re-emphasize the challenges of personalized or precision medicine in patients with metastatic melanoma. The intra-patient heterogeneity in response to treatment and the diversity of genetically induced resistance mechanisms to *BRAF* inhibition, being either re-activation of the RAS/RAF pathway, activation of the PI3K/Akt or MAPK pathway, or other mechanisms yet to be defined, require a novel approach on how to best treat these patients. Our study has several important aspects for clinical decision-making. First, it demonstrates heterogeneity of treatment response in metastatic melanoma patients. Early detection of growing lesions allows timely biopsies to potentially identify new druggable genetic events in this generally rapidly growing disease. Secondly, sequential biopsies are most informative when obtained from sites of progressive disease rather than from sites that are easily accessible but not growing. Finally, sequential biopsies may be very

informative in uncovering mechanisms of drug resistance which will ultimately lead to new (combinations of) targeted therapies to either prevent or overcome resistance. In conclusion, whereas significant clinical benefit comes from the identification of mutations that drive metastatic melanoma and drugs that target these mutations, the development of resistance poses a major clinical problem. Our study shows the promise of combining timely, volumetric imaging parameters and selective sampling of progressive lesions with next generation sequencing (NGS) based genetic data to improve treatment outcome in individual patients with metastatic *BRAF* mutant melanoma.

Materials and methods

Patient accrual and sample acquisition

After study approval by the institutional review board, eleven patients with treatment naïve, *BRAF* V600-mutated stage IV melanoma eligible for treatment with vemurafenib were included. The patients were included in the CPCT-02 umbrella biopsy protocol (ClinicalTrials.gov; NCT01855477) and gave written informed consent for repeated research biopsies and the use of their clinical data to correlate with sequencing results from their biopsies. We obtained baseline scans before start of vemurafenib treatment and volumetrically evaluated the response of metastatic lesions during the course of treatment. Ten out of eleven patients had volumetrically evaluable lesions, and we obtained peripheral blood samples and paired pre-treatment and progression biopsies of metastatic lesions from four of the patients. Peripheral blood samples (10mL) were collected in K2EDTA tubes. Biopsies were snap-frozen in liquid nitrogen and stored at -80°C. Histological assessment to confirm the presence of tumor cells and to mark regions with high tumor cellularity for macrodissection was performed by a pathologist. DNA was extracted from 500μl whole blood and from 20 μm macrodissected sections using NorDiag Arrow (Isogen Life Science, De Meern, the Netherlands). DNA was quantified with Qubit 2.0 fluorometer® (Life Technologies, Carlsbad CA, USA).

Therapy response measurements

Pre-treatment baseline CT scans and follow-up CT scans until end of treatment were acquired on a 64-slice Brilliance CT (Philips, Best, The Netherlands) with 0.625 mm collimation, and reconstructed in the transversal and coronal plane at 5 mm slice thickness with 4 mm increment. Images were anonymized and stored as DICOM images at the central facility. Three-dimensional (3D) volumetric measurements were performed in the transversal plane, using the semi-automatic software tool EncoreUnFoie v5.0 (Image Sciences Institute, Utrecht, the Netherlands).

For each patient, volumetrically evaluable target lesions were selected on the baseline scan. Although RECIST 1.1[175,186] guidelines advice a maximum of five target lesions and two lesions per organ, we selected up to ten lesions per patient and five per organ (according to RECIST 1.0) because we were interested in differences in response between lesions within one patient and not only in overall response. Volumetric measurements were performed by semi-automatically contouring the lesion on all axial slices. Measurements of the individual lesions and the sum of volumes of the lesions were recorded. The percentage change in volume compared to baseline volume and to the smallest volume measured in this study was calculated for each target lesion individually and for each patient's cumulative volume of target lesions. RECIST 1.1 criteria were used to classify response for diameter measurements as progressive disease (PD; evident progression of non-target lesions or occurrence of new lesions, though SD based on target lesions), stable disease (SD), partial response (PR) or complete response (CR)[186]. The cut-off values for these categories were mathematically extrapolated for volumetric measurements, assuming a spherical tumor mass (**Table 3**)[187].

Table 3: Response categories as defined by RECIST 1.1[36], extrapolated to volume measurements[37].

Response category	Sum of diameters	Sum of volumes
Progressive disease	>20% increase or new lesions	>73% increase or new lesions
Stable disease	<30% decrease and <20% increase	<65% decrease and <73% increase
Partial response	>30% decrease	>65% decrease
Complete response	Disappearance target lesions	Disappearance target lesions

Cancer gene re-sequencing

Single nucleotide variants (SNVs) and small insertions/deletions (indels) were detected in a designed "Cancer mini-genome" consisting of 1,977 cancer genes, based on Vermaat et al., Hoogstraat et al.[40,188]. Barcoded fragment libraries were generated from 600ng DNA from tumor and control blood samples as previously described[153]. Pools of libraries were enriched for this gene set using SureSelect technology (Agilent Technologies, Santa Clara California, USA). Enriched libraries were sequenced to an average coverage of 150x on a SOLiD 5500xl instrument according to the manufacturers protocol. Mapping, variant calling and annotation was done as previously described[188]. Somatic variants were extracted by comparing variant lists of both tumor and control samples and subsequently genotyping discordant positions in the raw datasets of all three samples using sam tools mpileup[189] to ensure the absence or presence of the variant in a given sample. Validation of a subset of single nucleotide mutations and indels was

performed by PCR amplification of mutation loci followed by Ion Plus fragment library preparation and sequencing on the Ion PGM™ to an average coverage of 20.000x. Sequencing data from both SOLiD 5500xl and IonTorrent PGM are available from ENA-SRA (ERP005163).

Sanger sequencing of cDNA

2μg total RNA was used for cDNA synthesis using both locus specific (BRAF_1450FW and BRAF_1927RV; primer sequences listed in **additional data file 4(A)**) and oligo-dT oligonucleotides (Fermentas, Thermo Scientific, Landsmeer, the Netherlands). After cDNA synthesis, a PCR was performed using AmpliTaq Gold (Life Technologies, Carlsbad, California, USA) and primers BRAF_1450FW and BRAF_1927RV. After purification, 25ng of PCR product was sub-cloned using the CloneJET™ PCR Cloning Kit (Fermentas, Thermo Scientific, Landsmeer, the Netherlands) with pJET1.2/blunt Cloning Vector in competent *E. coli* TOP10 cells (Life Technologies, Carlsbad, California, USA) according to manufacturer's protocol. In total, 48 colonies from each PCR reaction were picked after which colony PCR was performed using BRAF_1450FW and BRAF_1927RV primers, and products were analyzed by capillary sequencing.

Cell culture and mutagenesis

A375 V600E mutant melanoma cells, kindly provided by the Clevers lab (Hubrecht Institute, Utrecht, The Netherlands), were cultured in Dulbecco's modified Eagle's medium (Lonza, Breda, the Netherlands) containing 4,5g/L glucose and supplemented with 10% fetal bovine serum (Lonza, Breda, the Netherlands) and 2mM Ultraglutamine 1 (Lonza, Breda, the Netherlands).

The pBABE-BRAF-V600E vector was kindly provided by the lab of René Bernards (NKI, Amsterdam, The Netherlands). We introduced specific mutations using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara California, USA) according to the manufacturers' protocol. Phoenix-ampho retroviral packaging cells (kindly provided by the Bos lab, UMC Utrecht, The Netherlands) were transfected using Xtreme-gene 9 (Roche, Basel, Switzerland) following the manufacturer's protocol. Twenty-four hours before transfection, $4x10^6$ cells were plated in 10-cm dishes. The cells were transfected with the retroviral expression vector pBABE-puro-BRAF-V600E, pBABE-puro-BRAF 'wild type' or pBABE-puro-BRAF-V600E-L505H mutant constructs. After twenty-four hours the medium was replaced and forty-eight hours after transfection the viral particles containing supernatant was spun at 1500rpm for 5 minutes and the supernatant snap frozen by submersion in liquid nitrogen. Twenty-four hours before transduction, $1x10^5$ of A375 cells were plated in a 6-well plate. On the day of transduction, the medium was replaced by 10-times diluted virus containing medium supplemented with 5ug/ml Polybrene (Sigma-Aldrich, Saint Louis

Missouri, USA). Fourty-eight hours after transduction, medium was replaced with selective medium containing 0,5 μg/ml puromycin (Sigma-Aldrich, Saint Louis Missouri, USA). Cells were selected with puromycin for over a week, which resulted in approximate 50% survival compared to non-selected cells.

Cell viability assay

The various strains of A375 cells were seeded in triplicate at 3000 cells per well in 96-well plates, avoiding the use of the outer wells. Ten milligrams of vemurafenib (Selleckchem, Houston Texas, USA) were diluted in 204,1 μL DMSO giving a stock solution of 100mM. After 24 hours, the medium was replaced by medium supplemented with various concentrations of vemurafenib. After seventy-two hours of treatment with vemurafenib, cell viability was evaluated using an MTT assay. Samples were measured with a spectrophotometer at a wavelength of 572 nm Values were corrected for background absorbance by subtraction of absorption at 690nm.

Authors' contributions

MH carried out the data analysis and participated in the data interpretation and manuscript preparation, CGGH supplied clinical and volumetric data and participated in the manuscript preparation, IU supplied clinical and volumetric data and participated in the manuscript preparation, NJMB performed all sequencing experiments, MP carried out cell culturing and the cell viability assays, WV participated in volumetric data acquisition, MvS participated in volumetric data acquisition, EFJM participated in the cell culturing and cell viability assays, SMW performed pathological data acquisition, MAH carried out the protein modeling and interpreted these data, MJK participated in the design and coordination of the study and interpretation of the data, EC participated in the design of the study and revision of the manuscript, EEV participated in the design of the study and revision of the manuscript, MPL conceived of the study and participated in the design and coordination of the study, the interpretation of the data and preparation of the manuscript.

Acknowledgements

SMW is funded by the Dutch Cancer Society (clinical fellowship: 2011-4964). This work was supported by the Center for Personalized Cancer Treatment, which is a collaboration between the University Medical Center Utrecht, Erasmus Medical Center Rotterdam and the Netherlands Cancer Institute Amsterdam.

6

Summarizing discussion

Summary of the discussion

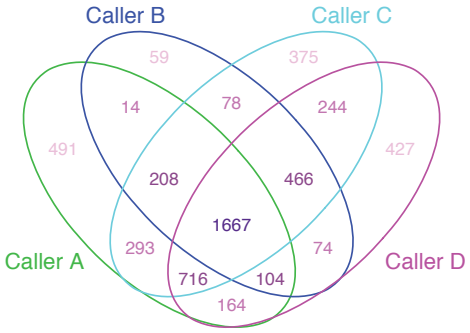
One could draw several conclusions from this thesis. First of all, the analysis of just one type of somatic variation is not enough to get a complete image of a specific tumor. Instead, integration of structural variation, copy number alterations, expression data and point mutations and indels provides valuable insight into the mechanisms driving tumor development and evolution. But even then we're still not rendering the full map of the tumor landscape. Extra information is bound to be discovered in relatively uncharted terrain, such as regulatory regions. Second, interpretation of somatic variation is still a daunting task. If we include structural variation, hundreds of somatic events can be detected per sample, and separating drivers from passengers with or without phenotypic data is not always feasible. Finally, and most importantly, sampling only one tumor location at a specific point in time will not render representative data for the complete population of tumor cells within a patient, as we even observe differences between samples taken from the same neoplasm. If we want to understand cancer and therapy response in all their complexity, we will need to look both deeper, e.g. by extensively studying single patients, and broader, by analyzing as many samples as possible from all different tumor types. Below, I will discuss these points in more detail, including possible approaches and implications on future research.

Capturing heterogeneity

In **chapter 4**, we observe two distinct tumor populations within single samples of patient #1, one being more pronounced than the other. In **chapter 5**, we detect increased allele frequencies of several somatic mutations between pre- and post-treatment samples from two of our patients, also indicative of the presence of sub-populations of tumor cells. These sub-populations are clinically highly relevant, as they potentially contain therapy resistance mechanisms that can be intercepted by combining or altering treatment strategies[42,190]. Importantly, most variants in the minor sub-population of these samples had allele frequencies of less than 10%, even less than 5%, and were not detected using the standard procedure for somatic variant calling on the 150x coverage sequence data. Instead, they were detected in another sample of the same patient and either confirmed by revisiting the raw sequence data, or detected through deep sequencing (>1000x coverage) of the specific variant location. If a variant could be detected in the raw 150x sequence data, the use of more lenient settings for somatic variant calling would have identified it. However, this inevitably leads to more false positive variant calls. Finding the balance between too many false positives (low specificity) or too many false negatives (low sensitivity) is an ongoing struggle. New algorithms for somatic variant calling using different approaches are released on a regular basis, and

overlap between results of these algorithms is often small[191,192] (**Figure 1**). To improve these results, we will need to understand where the false positives and negatives come from for each method, and probably combine methods to accomplish high confidence variant calls with high sensitivity.

***Figure 1**: Comparison of somatic variant callers, applied to the same set of 16 samples of lung squamous cell carcinomas from TCGA.*



The Venn diagram shows the differences and overlap of variants between four different somatic callers, detected in the 80x coverage Illumina whole-exome sequencing data. Approximately 10-15% of variants reported by callers A, C and D were only identified with that specific caller.

Figure adapted from Kim and Speed, BMC Bioinformatics 2013[191].

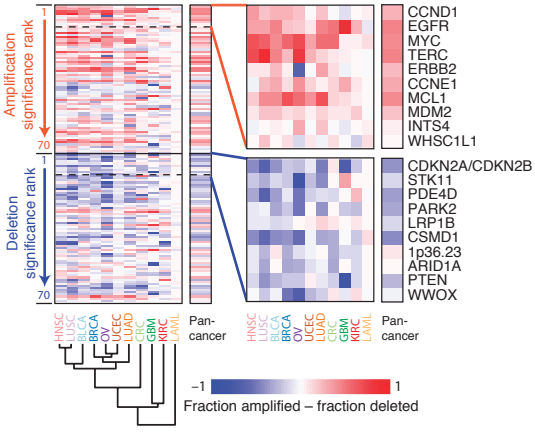
If a variant could only be detected in the deep sequencing data, but not in the raw 150x sequence data, improvement of variant calling algorithms would not have made a difference. Only deeper sequencing (leading to higher costs) can solve this issue, which may also be required to successfully study samples with a low tumor cell percentage. In some cases, tumor biopsies or slides contain a lot of 'normal cells', for instance from healthy tissue or stromal cells. If the somatic cells and normal cells are easily distinguishable, the sample can be enriched for tumor cells using e.g. rough macrodissection or more specific laser microdissection[193,194]. However, especially immune cells can have infiltrated the tumor and be distributed throughout the sample, which makes it very difficult to separate the different cell types. This was the case in four of the metastatic samples from patient #1 in **chapter 4**. The presence of these cells actually plays an important role in tumor progression; infiltrating T cells for instance significantly increase long-term survival in advanced ovarian cancer[195] and colorectal cancer[196], while tumor-associated macrophages appear to promote tumor invasion, growth, angiogenesis, and metastasis[195,197]. Increased understanding of the complex interactions between immune cells in the tumor microenvironment and cancer cells has lead to the implementation of immunotherapeutic drugs such as Ipilimumab[198], indicating that even though these non-cancerous cells hamper the genetic analyses of tumor samples, their presence alone is clinically relevant.

Both tumor sub-populations and differences in the tumor microenvironment can result in varying responses within single patients. Combinations of several targeted therapies at once are proposed to engage all different aspects and to prevent or postpone therapy resistance, either through combined inhibition of multiple pathways (e.g. Raf/Mek and Akt/PI3K[199-201]), by targeting a single pathway on different levels (e.g. combinations of *BRAF* and *MEK* inhibitors[45] or dual *PIK3CA* and *MTOR* inhibitors[202]), or by combining targeted therapy with chemotherapy or immune therapy[50,198,203].

Rare events… or are they?

Deep analysis of single patients may lead to rare, but relevant findings, which would not have been discovered through broader, shallower analysis of large cohorts. For example, localized, dense shattering of a chromosome called chromothripsis was first discovered in a single patient with chronic lymphocytic leukemia (CLL)[94]. Further analysis of larger groups of patients and cell lines revealed a prevalence of 1-3%[93,94,204] in all cancer types, and up to 18-25%[94,205] in neuroblastoma and bone tumors, respectively. Similarly, we detected chromothripsis in all four colorectal patients we studied in **chapter 3**. Interestingly, chromothripsis is now associated with poor outcome in patients with multiple myeloma[93], neuroblastoma[205] and malignant melanoma[206], showing the relevance of this discovery.

***Figure 2**: Significantly recurrent focal somatic copy number alterations (SCNAs) in eleven tumor types.*
The color scale is determined by subtracting the fraction of samples with deletions from the fraction samples with amplifications. Several genes show amplifications or deletions in multiple tumor types, such as MYC, MCL1 and CSMD1, while other events are more exclusive for a single tumor type like ERBB2 deletions in ovarian cancer or PTEN loss in glioblastoma.



Abbreviations: HNSC: head and neck squamous cell carcinoma, LUSC: lung squamous cell carcinoma, BLCA: bladder cancer, BRCA: breast cancer, OV: ovarian cancer, UCEC: uterine cervical cancer, LUAD: lung adeno-carcinoma, CRC: colorectal cancer, GBM: glioblastoma, KIRC: kidney renal cell carci-noma, LAML: acute myeloid leukemia.
Figure adapted from Zack et al, Nature Genetics 2013[207]

This might also prove to be true for other rare observations described in this thesis, such as the secondary mutation in *BRAF* described in **chapter 5** or distinctly different tumor cell populations with their own driving events in single patients, as described in **chapter 3**, **chapter 4** and by Gerlinger et al[37]. Large-scale studies that make their data publicly available such as ICGC[22] and TCGA allow for an estimation of the population frequency of specific (genetic) events. To increase sample size even more, one can consider combining datasets from several tumor types. So-called "pan-cancer" analyses presented a great amount of overlap in tumor driving events throughout tumor types[207,208] (**Figure 2**). These studies have resulted in newly discovered copy number alterations of RNAs and amplification of genes related to epigenetic and mitochondrial regulation[207], and significantly mutated genes involved in splicing, transcription regulation, metabolism, proteolysis and histones[208]. Pan-cancer studies can give more insight into mechanisms underlying copy number variation[207] and tumor clonal architecture[208]. However, they also show that several genes are exclusively mutated in a single tumor type, indicative of a different underlying biology of the tumorigenesis[208], and cross-tumor models predicting patient survival showed very limited predictive power[209]. Furthermore, these sample sets usually consist of single samples per patient, obtained from primary, treatment-naïve tumors, while sequential sampling from multiple locations would be required to capture the aforementioned events.

Separating the wheat from the chaff

Large scale studies and broad analysis of tumor genomes leads to a wealth of data: per sample, hundreds of point mutations and structural variants can be detected, and with the little overlap we observe, addition of large numbers of patients will increase the number of observations almost linear. Interpretation of this many variants is very challenging, which is why many studies focus on genes known to be involved in cancer like the "cancer mini genome" used in **chapters 3, 4** and **5**, or even on specific mutations and copy number alterations with known clinical relevance such as the panel described in **chapter 2**. Methods are being developed to discriminate between driver and passenger events[26-28], but the biological mechanism or clinical relevance of a (newly discovered) driver is not always clear. Moreover, variants that are assumed to be 'passengers' can turn out to have clinical relevance in the right context. For example, the activating *PIK3CA* mutation in **chapter 5** is a well-known tumor driver, but its implication for therapy response in melanoma was still unclear. Even when an association between a genetic variant and therapy response is known for specific types of cancer, this association may be different or non-existent for other tumor types. This has been shown for example for *BRAF* V600E mutations, that confer sensitivity to vemurafenib in melanoma[33,44] but not in colorectal cancer[184], and for *EGFR*

inhibitors successfully targeting *EGFR* amplified lung cancer[210], but not glioblastoma[211]. As a consequence, conclusions on the function and relevance of genetic variation can only be made after thorough testing of the hypothesis in relevant model systems or patient cohorts. Cancer cell lines are widely used for this purpose. However, a recent publication[212] has shown inconsistent results between studies using the same cell lines. Moreover, it is very difficult (if not impossible) to find cell lines with a genetic background and somatic profile similar to the tumor in which the tested mutation is observed. Alternatives for cell lines such as patient derived xenografts (PDX) or tumor organoids are currently being explored. Organoids are vital, complex multicellular cultures derived from patient tumors that resemble the original tumor more than traditional cell lines[213]. PDX are also derived from patient tumors and implanted into immune deficient mice. They allow researchers to study the interaction between tumor cells and surrounding tissue, such as stroma and immune cells[214]. However, PDX have the disadvantage that they are not readily available but can take months to be available for testing[215], which is a delay not every cancer patient can afford.

Functional interpretation of genetic variants can also be partially assessed by the generation of additional data from the patient sample, such as gene expression and (phoshpo)proteomics data. We have used this approach in **chapter 4**, by adding RNAseq data and overlapping that with the matepair sequencing results and mutational analysis. In part, we could study the influence of breakpoints and mutations on gene expression, as we show for instance for the *TSC1* splice site point mutation. However, as the downstream pathway of *TSC1* is regulated through phosphorylation, effect of the mutation on the phenotype of the tumor could not be determined. This would have been possible through addition of phospho-proteomics: the (semi-)quantative analysis of phosphorylated and non-phosphorylated peptides in a sample. Phospho-proteomics can also be very valuable in the study of resistance mechanisms against kinase inhibitors, such as our study on vemurafenib resistance in **chapter 5**. There, we could have used it to investigate whether re-activation of the Ras/Raf/ERK pathway was achieved, or the activation of alternative pathways e.g. activation of Akt through *PIK3CA*. Phospho-proteomic profiles can be generated for instance globally with the use of mass spectrometry (MS) or targeted with reverse-phase protein arrays (RPPA)[216]. The MS approach will also detect novel phosphorylation sites, but requires at least 500 micrograms input material in order to reach sufficient sensitivity. This is not always available from tumor biopsies. In contrast, the targeted RPPA approach can be used on a limited amount of starting material[217], but supplies limited information.

In some cases, recurrent observations do not involve single events such as point mutations or amplifications, but more complex events such as chromothripsis described in **chapter 3** or kataegis observed in **chapter 4**. Even though underlying mechanisms have been proposed[94,137] it is still unclear whether these events are tumor driving, or only an effect of some earlier (genetic) defect in the tumor. For example, in **chapter 4** we noticed that multiple point mutations in the region affected by kataegis fall within *FANCD2*, a gene of the Fanconi anemia complementation group involved in the repair of double-stranded breaks[218]. Similarly, in **chapter 3** we observed that chromothripsis clusters can affect several cancer-associated genes like *EXO1*, a gene involved in DNA mismatch repair[219]. Both the breakpoint clusters in chromothripsis and the mutation clusters kataegis appear to be a result of a single event, as all mutations within the cluster lie on the same DNA strand[137] and DNA fragments in chromothripsis alternate between two copy number states[94]. Considering these observations, one can imagine that the underlying mechanisms of such complex events actually drive tumor formation as they increase the chance of damage to essential tumor suppressor genes, but more thorough analysis of samples exhibiting these events should pinpoint the exact cause to test this hypothesis.

Uncaptured variation

Because the interpretation of most somatic variation detected in cancer samples is still complicated, more straightforward approaches are used to evaluate samples in a timeframe relevant for clinical decision-making. It is relatively easy today to routinely assess clinically actionable point mutations and copy number alterations in tumor biopsies using a single assay, e.g. through the panel described in **chapter 2** or commercially available “cancer panels” such as TruSeq Cancer Panel from Illumina® or Life Technologies Ion AmpliSeq™ Cancer Hotspot Panel. These panels are far from complete, but other types of somatic variation can be more challenging to detect or require separate assays. For instance, recurrent fusion genes resulting from chromosomal breaks have been reported, but these rearrangements can occur in many variants. In some cases, the location of the chromosomal break differs, such as in the breakpoint cluster region (BCR) from the *BCR-ABL* fusion[220] and *EML4-ALK* rearrangements in non-small cell lung cancer[221] (**Figure 3**), but in other cases fusion partners can even vary as reported for MAST kinase and Notch fusions in breast cancer[115], and again for *ALK* in lung cancer[222] (**Figure 3**). This variation makes it difficult to systematically screen for fusion genes, while their presence can provide treatment options[222,223].

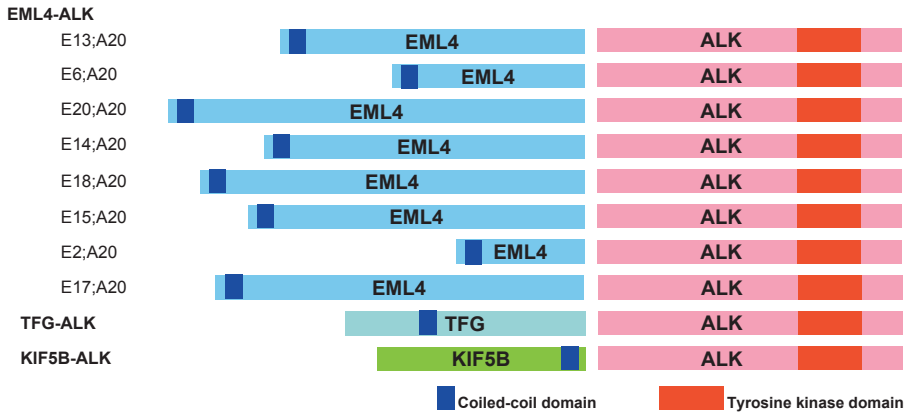


***Figure 3**: Different variants of EML4-ALK and non-EML4 fusion partners. The nomenclature refers to the exon in EML4 translocated to the exon in ALK. Figure adapted from Sasaki et al, Eur J Cancer 2010[224] with permission from Elsevier.*

Promoter hypermethylation is a well-known mechanism for the inactivation of tumor-suppressor genes and has been described in almost every tumor type[225]. Examples include *BRCA1* in breast and ovarian cancer[226], *RB1* in gliomas[227] and *VHL* in renal cell carcinomas[228]. In contrast, genome-wide hypomethylation is associated with increased mutation rates and chromosomal instability[229]. Methylation status cannot be determined using regular DNA sequencing. Instead, other methods such as microarrays or bisulfite sequencing are used to assess promoter methylation[230], either in a targeted manner (e.g. using methylation-specific PCR) or whole genome[229]. A more unexplored type of variation also involves regulatory regions: point mutations in promoter or enhancer regions. The best-studied example is *TERT*, a gene responsible for telomere maintenance, where point mutations in the promoter are associated with increased expression[231,232]. Specific mutations of cytosine to adenine or thymine in the ETS transcription factor binding site[231] were detected in a large variety of tumor types[233]. It is to be expected that mutation of promoter or enhancer regions is quite common in tumor samples. However, not all regulatory regions in the human genome are fully mapped. On top of that, promoter regions have a high GC-content and low sequence complexity, which impedes the design of specific capture probes and makes systematic screening problematic.

In conclusion, (clinically) relevant information can be obtained through the addition of new types of data. The decreasing amount of input material required for sequencing allows for the generation of this data, but a systematic approach would need to be developed first.

Solving the sampling-issue? CTC & ctDNA

The ever-increasing capabilities of NGS allow us now to sequence deeper to detect mutations present in less than 1% of tumor cells, and start with minimal amounts of input material. This enables us to capture most of the genetic variation in a given sample. However, the sampling itself is still a major issue. We need to analyze more samples per patient, starting with multiple samples at initial diagnosis and repeating this every time a tumor becomes resistant to capture all potentially relevant tumor cell populations. This is currently not feasible at all, as the taking of multiple tissue biopsies is extremely demanding on the patient and often not safe. Both the research and the clinical community are therefore searching for alternatives.

One promising alternative, also made possible by advances in NGS techniques, is called “liquid biopsies”: extracting the tiny amount of tumor material found in the bloodstream. These liquid biopsies are easy to perform compared to tissue biopsies, significantly safer for the patients and easily repeatable over time[234]. Currently, two approaches are being explored: circulating tumor cells (CTCs) and cell-free tumor DNA (ctDNA). The first approach is based on the observation that tumors shed single cells into the bloodstream, which is the route of metastasizing of most tumor types. It involves enrichment of those tumor cells from the total population of cells in the blood such as leukocytes, typically using cell-surface markers[235] or other tumorcell-specific characteristics[236]. Presence of circulating tumor cells has been used for several years as a biomarker for prognosis and treatment response monitoring[237-240], but an increasing amount of studies report on generating aCGH profiles and even single nucleotide variants[241-243] detected in these CTCs. Unfortunately, CTCs are very rare, estimated at fewer than 1 CTC per million leukocytes[234] with average CTC counts per 7.5 mL blood ranging from 1 to ~100 in different tumor types[244]. On top of that, CTCs manifest a great amount of heterogeneity in size, cytomorphology and biomarker presence, both between cancers and within single patients[245]. This means that tumor profiling using CTCs is not available for every patient; indeed, Heitzer et al.[241] reported finding no CTCs in 29.7% of patients in his study. Moreover, considering the complexity and extensive heterogeneity of the disease, one can question whether the small amount of tumor cells obtained through this approach is representative at all for the total population of tumor cells in the patient. As the yield of CTCs also depends on the manner of isolation[244], optimization of isolation or enrichment methods may hopefully lead to a higher efficiency.

The second approach, using cell-free tumor DNA or plasma DNA, does not involve enrichment for tumor material. Instead, the total amount of cell-free DNA in the blood is sequenced and the fraction of tumor DNA is determined through allele frequencies of somatic mutations. Levels of ctDNA are highly dependent on the tumor load in the patient, and can range from less than 1% to over 40%[246]. They can therefore be used to monitor therapy response and tumor progression, and have been successfully used to detect residual disease in patients who had undergone resection with curative intent[247]. When the tumor load (and thus the fraction of ctDNA) is high enough, whole genome or exome sequencing can be performed to search for somatic variants, but with small fractions such as can be expected after surgery or successful treatment, other approaches are more practical. For instance, Murtaza et al.[246] used digital PCR and tagged-amplicon deep sequencing (TAm-seq) to monitor allele frequencies of somatic mutations in serial samples, first identified using whole exome sequencing of a pre-treatment sample. Besides the potentially low levels of plasma DNA, there are some additional difficulties to be taken into consideration: first of all, the DNA is fragmented, so several regular enrichment methods can't be applied. Also, the DNA must be extracted from the plasma as fast as possible, as degradation of leukocytes increases the amount of total cell-free DNA and thus decreases the fraction of ctDNA.

Even though liquid biopsies provide an attractive approach for non-invasive genetic screening, it is not likely that they can ever fully replace tissue samples. For one thing, tumor histology cannot be determined from single cells or DNA fragments and the composition and influence of the tumor environment can't be assessed. When using ctDNA, it's not possible to perform expression and proteomics analyses, which is also highly improbable with the low amount of material coming from CTCs. However, liquid biopsies do allow for more frequent and safer sampling, enable genetic screening of patients who would not be eligible for tissue biopsies and provide very valuable additional information when analyzed in addition to tissue biopsies or resection material.

Concluding remarks: on the heterogeneity of translational cancer research

If we consider all of the above, we can conclude that cancer can only be beaten if we join forces globally. Not only do we need large patient cohorts to increase our understanding of mechanisms underlying the disease, but also deep multi-level analyses together with sequential sampling to identify clinically relevant events. This means we need heterogeneous groups of experts in disciplines such as clinical oncology, pathology, systems biology, genetics, proteomics and molecular biology to work together and to be able to understand each other. In order to capture tumor heterogeneity, we need to develop and improve sampling methods, wet-lab techniques and bioinformatics algorithms, and learn which treatments can be combined safely to prevent therapy resistance caused by this heterogeneity. On top of that, I believe we need to redefine “treatment response” by shifting from response on the patient level towards response on the sample level, thus associating the right response to the right genetic profile, in order to increase the efficiency of biomarker discovery. Accomplishing any or all of these points will give us new weapons in the battle against cancer.

7

Addendum

Nederlandse samenvatting:
Over de heterogeniteit van tumor sequencing

DNA, het script van je lichaam

Alle levende wezens op aarde, van bacteriën, schimmels en het onkruid tussen de stoeptegels tot walvissen, aardappelen en mensen, worden gemaakt tot wat ze zijn door DNA. DNA is een molecuul dat is opgebouwd uit vier verschillende bouwstenen, ook wel basen of nucleotiden genoemd: Adenosine (A), Thymine (T), Cytosine (C) en Guanine (G). Deze bouwstenen worden aan elkaar geregen tot lange strengen DNA. Bij bacteriën is dit één lange, circulaire streng, bij mensen zijn dit 46 losse strengen: de chromosomen. Hiervan krijgen mensen er 23 van hun moeder (1 tot en met 22, en X), en 23 van de vader (ook 1 tot en met 22, en X of Y). De volgorde van deze bouwstenen is niet willekeurig, sterker nog: het bepaalt alles wat er met de cellen in je lichaam gebeurt. Dit werkt door middel van genen: stukken van het DNA die vertaald kunnen worden naar bepaalde eiwitten, die op hun beurt weer zorgen voor specifieke eigenschappen. Hier heb je er als mens ongeveer 20.000 van. Het DNA van mensen die geen familie van elkaar zijn, komt voor ongeveer 99,9% overeen. Het verschil van 0,1% heeft in sommige gevallen invloed op wie je bent, bijvoorbeeld of je blond met blauwe ogen bent, goed tegen alcohol kan, makkelijk bruin wordt en aanleg hebt om dik te worden, of dat je rood haar hebt en sproeten, zowel van alcohol als van de zon heel snel knalrood wordt en kan eten wat je wil zonder aan te komen. De verschillen kunnen echter ook van invloed zijn op gezondheid, bijvoorbeeld op aanleg voor hart- en vaatziekten, diabetes, alzheimer en kanker. Meestal bestaan deze verschillen uit de verandering van een enkele nucleotide, maar de verschillen kunnen ook groter zijn. Het syndroom van Down wordt bijvoorbeeld veroorzaakt doordat er een compleet chromosoom te veel is.

DNA en kanker

In de 20^{e} eeuw heeft men ontdekt wat DNA is, hoe het werkt en hoe we het kunnen aflezen (oftewel sequencen). Dit heeft ervoor gezorgd dat de genetische achtergrond van veel aandoeningen nu bekend is. Ook is hierdoor bekend geworden dat kanker wordt veroorzaakt door veranderingen in het DNA, die voornamelijk tijdens je leven ontstaan. Dit zijn zogenaamde somatische varianten. Deze veranderingen kunnen ontstaan doordat je DNA van buitenaf beschadigd raakt zoals door zonlicht, asbest of virussen, maar ook door ongelukkig toeval. Gedurende je leven blijven de meeste cellen zich namelijk delen. Darm- en huidcellen delen zich bijvoorbeeld heel vaak: de cellen aan de binnenkant van je darm leven maar een paar dagen, de cellen in je opperhuid een paar weken. De afgestorven cellen worden vervolgens vervangen door ‘verse’ cellen. Tijdens een celdeling wordt het DNA in de cel verdubbeld, en vervolgens verdeeld over de

twee dochtercellen. Tijdens de verdubbeling van het DNA en het delen van de cel wil er wel eens iets mis gaan. De cel heeft diverse checkpoints en reparatie-mechanismen ingebouwd, die er voor zorgen dat de cel niet deelt zolang het DNA niet in orde is. Over het algemeen werken deze mechanismen heel goed, maar soms deelt een cel toch door, vooral als er een gen beschadigd is wat juist voor deze checks moet zorgen. Veranderingen aan het DNA van dit soort genen zien we dus ook heel vaak terug in diverse soorten kanker. Andere veranderingen die we vaak terug zien zorgen er juist voor dat een gen veel harder werkt dan zou moeten. Dit zijn bijvoorbeeld genen die doorgeven aan een cel dat hij moet gaan delen of groeien. Als er maar genoeg van dit soort veranderingen in het DNA optreden krijg je uiteindelijk een cel die veel harder groeit dan zou moeten, en die steeds meer beschadigingen in zijn DNA oploopt: een tumorcel.

Verschil in variatie

Zoals in de eerste alinea al genoemd is, zijn er verschillende soorten variatie die op kunnen treden. Er kunnen verschillen van enkele nucleotiden zijn en verdubbeling of verlies van volledige chromosomen, maar in tumoren gebeurt het ook heel vaak dat chromosomen breken en vervolgens weer verkeerd aan elkaar geplakt worden en/of ongelijk verdeeld worden over de twee dochtercellen tijdens de celdeling. Hierdoor kunnen er hele complexe structurele variaties ontstaan, waarbij er meerdere chromosomen als het ware door elkaar gehusseld zijn. Als de breuken in het DNA in, of in de buurt van, een gen liggen, kan het gebeuren dat een gen daardoor niet meer werkt of juist veel harder gaat werken. Iets anders wat vaak gezien wordt en in eerste instantie ook veroorzaakt wordt door een breuk in het DNA is gen amplificatie: een kleine regio van het DNA waar maar één of een klein aantal genen op ligt, wordt steeds verdubbeld waardoor er uiteindelijk veel meer dan 2 kopieën van het gen aanwezig zijn. Dit leidt over het algemeen ook tot over-activering.

Doelgerichte behandeling

Tegenwoordig kunnen we al deze typen variatie oppikken met behulp van DNA sequencing, waardoor we per patiënt het ‘tumor genoom’ in kaart kunnen brengen. Dit is zeer waardevol, omdat er voor steeds meer somatische varianten een gerichte behandeling ontwikkeld wordt. De traditionele behandeling van kanker is chemotherapie, waarbij er stoffen die giftig zijn voor cellen aan de patiënt worden toegediend. Deze stoffen zijn niet alleen giftig voor kankercellen, maar voor alle cellen in het lichaam waardoor de patiënt vaak veel nare bijwerkingen ervaart. Het kan zelfs leiden tot gevaarlijke infecties en hartschade. Daarom wordt er hard gewerkt om behandelingen te ontwikkelen, die specifiek de kankercellen aanvallen maar de gezonde cellen sparen. Dit kan gedaan worden door processen te remmen die voornamelijk belangrijk zijn voor de tumor, zoals celdeling en -groei, of aanmaak van nieuwe bloedvaten die de tumor nodig heeft om voedingsstoffen en

zuurstof te krijgen. Hoewel alle tumoren in de basis gedreven worden door dezelfde processen, kan de manier waarop deze processen geactiveerd worden per patiënt verschillen en dus de aanpak om de processen te remmen ook. Om te bepalen wat de beste aanpak is, kan in het DNA worden gezocht naar aanwijzingen.

Omdat reparatie van DNA in tumorcellen vaak defect is, kan de hoeveelheid breuken en nucleotide veranderingen blijven oplopen. Niet al deze veranderingen zullen bijdragen aan groei of overleving van de tumor, daarom wordt er onderscheid gemaakt tussen tumor aandrijvende mutaties en zogenaamde passagiers. Er zijn ondertussen meerdere genen bekend die gezien worden als aandrijvers, omdat er vaak somatische mutaties in gevonden worden in meerdere tumor typen. Dit zijn aantrekkelijke opties om een behandeling tegen te ontwikkelen, wat in sommige gevallen ook al gelukt is. Deze behandelingen kunnen de levensduur en -kwaliteit van de patiënt sterk verbeteren, al zijn de meeste behandelingen nog steeds niet vrij van bijwerkingen. Ook zijn er terugkerende DNA veranderingen bekend, die er voor zorgen dat een tumor juist niet op een bepaalde gerichte behandeling reageert. Door deze informatie mee te nemen, kan er voor elke patiënt een persoonlijk behandelingsplan opgesteld worden. Hierop is de bepaling gebaseerd die in **hoofdstuk 2** gepresenteerd wordt. Deze bepaling kijkt tegelijkertijd naar gen-amplificaties en nucleotide veranderingen in een veertigtal bekende kanker genen, waaraan een respons op een specifieke behandeling geassocieerd is. Hiermee vervangen we meerdere bestaande testen, zodat er goedkoper en efficiënter een profiel opgesteld kan worden.

Tumor heterogeniteit

Doordat de schade aan het DNA van tumorcellen blijft oplopen, is niet elke cel binnen een tumor genetisch hetzelfde. In tegendeel, we zien regelmatig grote verschillen. In **hoofdstuk 3** onderzoeken we deze heterogeniteit in vier patiënten, door nucleotide veranderingen en DNA breuken te vergelijken tussen de oorspronkelijke tumor in de dikke darm en een uitzaaiing in de lever. We zien zelfs in één patiënt dat er verschillende tumor aandrijvende veranderingen aanwezig zijn in de twee tumoren. Ook beschrijven we in dit hoofdstuk dat zeer complexe structurele variatie niet zo zeldzaam is als men eerst verwachtte, omdat we dit in allevier de patiënten terugzien. In **hoofdstuk 4** breiden we dit onderzoek verder uit, door tot twaalf aparte samples te onderzoeken van drie patiënten met eierstokkanker. Hierin bekijken we de verschillen in het DNA tussen twee of drie samples van dezelfde tumor, en tussen de oorspronkelijke tumor en meerdere uitzaaiingen. Bovendien kijken we hier niet alleen naar DNA, maar ook naar RNA, het molecuul dat gebruikt wordt om DNA te vertalen naar eiwit. Hiermee kunnen we vaststellen wat het effect van genetische veranderingen is op het uiteindelijke

eiwit. Ook in deze studie zagen we grote verschillen tussen de samples: in één patiënt zagen we twee ‘subpopulaties’ bestaande uit een deel van de oorspronkelijke tumor in de eierstok en een aantal uitzaaiingen, elk met zijn eigen tumor aandrijvende mutaties.

De genetische verschillen tussen tumoren van dezelfde patiënt zijn zeer belangrijk voor de kliniek. Ze kunnen er namelijk voor zorgen dat de tumor niet reageert op één behandeling, maar dat er combinaties van behandelingen nodig zijn om alle subpopulaties effectief te bestrijden. De huidige theorie is dat er in de tumor heel veel kleine subpopulaties zijn, of zelfs maar enkele cellen, die niet gevoelig zijn voor een specifieke doelgerichte behandeling. Als de patiënt vervolgens deze behandeling krijgt, zullen de meeste tumorcellen afsterven, maar de ongevoelige cellen doorgroeien. Uiteindelijk ontstaat er dan een nieuwe tumor die geheel ongevoelig (resistent) is. In **hoofdstuk 5** onderzoeken we deze theorie, door DNA van huidtumoren voor behandeling te vergelijken met DNA van dezelfde tumoren nadat ze ongevoelig zijn geworden. Hier vinden we inderdaad genetische veranderingen die de resistentie kunnen verklaren, en stellen we combinaties van behandelingen voor die, in deze patiënten, resistentie zou kunnen voorkomen. Ook ontdekten we in deze studie een somatische variant die nog nooit eerder in een huidkanker patiënt gezien was, maar waarvan we aan konden tonen dat het de resistentie veroorzaakte. Dit toont aan dat we nog lang niet alle variaties gezien hebben die binnen een tumor op kan treden.

In het laatste hoofdstuk worden de bovengenoemde resultaten bediscussieerd; wat is de relevantie van onze bevindingen voor de kliniek, en wat er volgens mij nog moet gebeuren om onze kennis van kanker te vergroten zodat de ziekte beter bestreden kan worden.

References

1. Hansemann, D. *Ueber asymmetrische Zelltheilung in Epithelkrebsen und deren biologische Bedeutung*. (1890).
2. Boveri, T. *Zur Frage der Entstehung maligner Tumoren*. (1914).
3. 1000 Genomes Project Consortium *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
4. Bernstein, B. E. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
5. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4,** 1073–1081 (2008).
6. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* **Chapter 7,** Unit7–Uni20 (2012).
7. Thusberg, J., Olatubosun, A. & Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum. Mutat.* **32,** 358–368 (2011).
8. Castellana, S. & Mazza, T. Congruency in the prediction of pathogenic missense mutations: state-of-the-art web-based tools. *Briefings in Bioinformatics* **14,** 448–459 (2013).
9. Growney, J. D. *et al.* Activation mutations of human c-KIT resistant to imatinib mesylate are sensitive to the tyrosine kinase inhibitor PKC412. *Blood* **106,** 721–724 (2005).
10. Gazdar, A. F. Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene* **28 Suppl 1,** S24–31 (2009).
11. Abel, H. J. & Duncavage, E. J. Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. *Cancer Genet* **206,** 432–440 (2013).
12. Medvedev, P., Stanciu, M. & Brudno, M. Computational methods for discovering structural variation with next-generation sequencing. *Nat. Methods* **6,** S13–20 (2009).
13. Bozic, I. *et al.* Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 18545–18550 (2010).
14. Haber, D. A. & Settleman, J. Cancer: drivers and passengers. *Nature* **446,** 145–146 (2007).
15. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39,** D945–D950 (2010).
16. Bethune, G., Bethune, D., Ridgway, N. & Xu, Z. Epidermal growth factor receptor (EGFR) in lung cancer: an overview and update. *Journal of thoracic disease* **2,** 48–51 (2010).
17. Slamon, D. J. *et al.* Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235,** 177–182 (1987).
18. Reimand, J. & Bader, G. D. Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers. *Mol Syst Biol* **9,** 637–637 (2013).
19. Rowley, J. D. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243,** 290–293 (1973).
20. Harris, C. C. Structure and function of the p53 tumor suppressor gene: clues for rational cancer therapeutic strategies. *JNCI Journal of the National Cancer Institute* **88,** 1442–1455 (1996).
21. Leslie, N. R. & Downes, C. P. PTEN function: how normal cells control it and tumour cells lose it. *Biochem J* **382,** 1–11 (2004).
22. Hudson, T. J. *et al.* International network of cancer genome projects. *Nature* **464,** 993–998 (2010).
23. Li, M. *et al.* Whole-exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway. *Nat Genet* (2014). doi:10.1038/ng.3030
24. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155,** 462–477 (2013).
25. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489,** 519–525 (2012).
26. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
27. Dees, N. D. *et al.* MuSiC: identifying mutational significance in cancer genomes. *Genome Research* **22,** 1589–1598 (2012).
28. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human

genetic variants. *Nat Genet* **46,** 310–315 (2014).
29. Macconaill, L. E. & Garraway, L. A. Clinical implications of the cancer genome. *J Clin Oncol* **28,** 5219–5228 (2010).
30. Ferlay, J. *et al.* GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. at <http://globocan.iarc.fr>
31. Luengo-Fernandez, R., Leal, J., Gray, A. & Sullivan, R. Economic burden of cancer across the European Union: a population-based cost analysis. *Lancet Oncol.* **14,** 1165–1174 (2013).
32. Slamon, D. J. *et al.* Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* **344,** 783–792 (2001).
33. Chapman, P. B. *et al.* Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med* **364,** 2507–2516 (2011).
34. Carlson, J. J., Garrison, L. P., Ramsey, S. D. & Veenstra, D. L. Epidermal growth factor receptor genomic variation in NSCLC patients receiving tyrosine kinase inhibitor therapy: a systematic review and meta-analysis. *J Cancer Res Clin Oncol* **135,** 1483–1493 (2009).
35. Allegra, C. J. *et al.* American Society of Clinical Oncology provisional clinical opinion: testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *J Clin Oncol* **27,** 2091–2096 (2009).
36. Bang, Y.-J. *et al.* Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet* **376,** 687–697 (2010).
37. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366,** 883–892 (2012).
38. Snuderl, M. *et al.* Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell* **20,** 810–817 (2011).
39. Miranda, E. *et al.* Genetic and epigenetic alterations in primary colorectal cancers and related lymph node and liver metastases. *Cancer* **119,** 266–276 (2013).
40. Vermaat, J. *et al.* Primary colorectal cancers and their subsequent hepatic metastases are genetically different: implications for selection of patients for targeted treatment. *Clin Cancer Res* **18,** 688–699
41. Wang, F., Fang, P., Hou, D.-Y., Leng, Z.-J. & Cao, L.-J. Comparison of Epidermal Growth Factor Receptor Mutations between Primary Tumors and Lymph Nodes in Non-small Cell Lung Cancer: a Review and Meta-analysis of Published Data. *Asian Pac. J. Cancer Prev.* **15,** 4493–4497 (2014).
42. Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: implications for targeted therapeutics. *Br. J. Cancer* **108,** 479–485 (2013).
43. Crockford, A., Jamal-Hanjani, M., Hicks, J. & Swanton, C. Implications of intratumour heterogeneity for treatment stratification. *J. Pathol.* **232,** 264–273 (2014).
44. Sosman, J. A. *et al.* Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *N Engl J Med* **366,** 707–714 (2012).
45. Flaherty, K. T. *et al.* Combined BRAF and MEK inhibition in melanoma with BRAF V600 mutations. *N Engl J Med* **367,** 1694–1703 (2012).
46. Livingston, R. B. & Esteva, F. J. Chemotherapy and herceptin for HER2(+) metastatic breast cancer: the best drug? *Oncologist* **6,** 315–316 (2001).
47. Hadd, A. G. *et al.* Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens. *J Mol Diagn* **15,** 234–247 (2013).
48. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 13306–13311 (2004).
49. Said, R. *et al.* P53 Mutations in Advanced Cancers: Clinical Characteristics, Outcomes, and Correlation between Progression-Free Survival and Bevacizumab-Containing Therapy. *Oncotarget* (2013).
50. Romond, E. H. *et al.* Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* **353,** 1673–1684 (2005).
51. Lv, S. S. *et al.* Correlation of EGFR, IDH1 and PTEN status with the outcome of patients

with recurrent glioblastoma treated in a phase II clinical trial with the EGFR-blocking monoclonal antibody cetuximab. *Int J Oncol* **41,** 1029–1035 (2012).
52. Kato, N. *et al.* Evaluation of HER2 gene amplification in invasive breast cancer using a dual-color chromogenic in situ hybridization (dual CISH). *Pathol. Int.* **60,** 510–515 (2010).
53. Shao, M.-M. *et al.* Epidermal growth factor receptor gene amplification and protein overexpression in basal-like carcinoma of the breast. *Histopathology* **59,** 264–273 (2011).
54. Moelans, C. B., de Weger, R. A., Van der Wall, E. & Van Diest, P. J. Current technologies for HER2 testing in breast cancer. *Crit. Rev. Oncol. Hematol.* **80,** 380–392 (2011).
55. Schouten, J. P. *et al.* Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30,** e57 (2002).
56. Dancey, J. E., Bedard, P. L., Onetto, N. & Hudson, T. J. The Genetic Basis for Cancer Treatment Decisions. *Cell* **148,** 409–420 (2012).
57. Gravalos, C. & Jimeno, A. HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. *Annals of Oncology* **19,** 1523–1529 (2008).
58. Pazo Cid, R. A. & Antón, A. Advanced HER2-positive gastric cancer: Current and future targeted therapies. *Crit. Rev. Oncol. Hematol.* **85,** 350–362 (2013).
59. Tanner, M. *et al.* Amplification of HER-2 in gastric carcinoma: association with Topoisomerase IIalpha gene amplification, intestinal type, poor prognosis and sensitivity to trastuzumab. *Ann. Oncol.* **16,** 273–278 (2005).
60. Han, S.-W. *et al.* Targeted sequencing of cancer-related genes in colorectal cancer using next-generation sequencing. *PLoS ONE* **8,** e64271 (2013).
61. Gan, H. K., Kaye, A. H. & Luwor, R. B. The EGFRvIII variant in glioblastoma multiforme. *J Clin Neurosci* **16,** 748–754 (2008).
62. Duncavage, E. J., Abel, H. J., Szankasi, P., Kelley, T. W. & Pfeifer, J. D. Targeted next generation sequencing of clinically significant gene mutations and translocations in leukemia. *Mod Pathol* **25,** 795–804 (2012).
63. Becker, K. *et al.* Deep ion sequencing of amplicon adapter ligated libraries: a novel tool in molecular diagnostics of formalin fixed and paraffin embedded tissues. *J. Clin. Pathol.* (2013). doi:10.1136/jclinpath-2013-201549
64. Beadling, C. *et al.* Combining highly multiplexed PCR with semiconductor-based sequencing for rapid cancer genotyping. *J Mol Diagn* **15,** 171–176 (2013).
65. Singh, R. R. *et al.* Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes. *J Mol Diagn* **15,** 607–622 (2013).
66. Wood, H. M. *et al.* Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens. *Nucleic Acids Res.* **38,** e151 (2010).
67. Hayes, J. L. *et al.* Diagnosis of copy number variation by Illumina next generation sequencing is comparable in performance to oligonucleotide array comparative genomic hybridisation. *Genomics* **102,** 174–181 (2013).
68. Duan, J., Zhang, J.-G., Deng, H.-W. & Wang, Y.-P. Comparative studies of copy number variation detection methods for next-generation sequencing technologies. *PLoS ONE* **8,** e59128 (2013).
69. Michaux, L. *et al.* MLL amplification in myeloid leukemias: A study of 14 cases with multiple copies of 11q23. *Genes Chromosom. Cancer* **29,** 40–47 (2000).
70. Rayeroux, K. C. & Campbell, L. J. Gene amplification in myeloid leukemias elucidated by fluorescence in situ hybridization. *Cancer Genet. Cytogenet.* **193,** 44–53 (2009).
71. Barber, K. E. *et al.* Amplification of the ABL gene in T-cell acute lymphoblastic leukemia. *Leukemia* **18,** 1153–1156 (2004).
72. Moelans, C. B. *et al.* HER-2/neu amplification testing in breast cancer by multiplex ligation-dependent probe amplification in comparison with immunohistochemistry and in situ hybridization. *Cell. Oncol.* **31,** 1–10 (2009).
73. Iglewicz, B. & Hoaglin, D. C. *How to detect and handle outliers*. **16,** 1–87 (ASQC Quality Press, 1993).
74. Markowitz, S. D. & Bertagnolli, M. M. Molecular origins of cancer: Molecular basis of colorectal cancer. *N Engl J Med* **361,** 2449–2460 (2009).
75. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446,** 153–158 (2007).
76. Sjöblom, T. *et al.* The consensus coding sequences of human breast and colorectal

cancers. *Science* **314,** 268–274 (2006).
77. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318,** 1108–1113 (2007).
78. Stratton, M. R. Exploring the genomes of cancer cells: progress and promise. *Science* **331,** 1553–1558 (2011).
79. Mcdermott, U., Downing, J. R. & Stratton, M. R. Genomics and the continuum of cancer care. *N Engl J Med* **364,** 340–350 (2011).
80. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 4283–4288 (2008).
81. Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464,** 999–1005 (2010).
82. Klein, C. A. Parallel progression of primary tumours and metastases. *Nat Rev Cancer* **9,** 302–312 (2009).
83. Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467,** 1114–1117 (2010).
84. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461,** 809–813 (2009).
85. Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467,** 1109–1113 (2010).
86. Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* **40,** 722–729 (2008).
87. Stephens, P. J. *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462,** 1005–1010 (2009).
88. Berger, M. F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470,** 214–220 (2011).
89. Totoki, Y. Y. *et al.* High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet* **43,** 464–469 (2011).
90. Myllykangas, S. & Knuutila, S. Manifestation, mechanisms and mysteries of gene amplifications. *Cancer Lett.* **232,** 79–89 (2005).
91. Bignell, G. R. *et al.* Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution. *Genome Research* **17,** 1296–1303 (2007).
92. Hillmer, A. M. *et al.* Comprehensive long-span paired-end-tag mapping reveals characteristic patterns of structural variations in epithelial cancer genomes. *Genome Research* **21,** 665–675 (2011).
93. Magrangeas, F., Avet-Loiseau, H., Munshi, N. C. & Minvielle, S. Chromothripsis identifies a rare and aggressive entity among newly diagnosed multiple myeloma patients. *Blood* **118,** 675–678 (2011).
94. Stephens, P. J. *et al.* Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* **144,** 27–40 (2010).
95. Leary, R. J. *et al.* Development of personalized tumor biomarkers using massively parallel sequencing. *Sci Transl Med* **2,** 20ra14–20ra14 (2010).
96. Kloosterman, W. P. *et al.* Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. *Human Molecular Genetics* **20,** 1916–1924 (2011).
97. O'Hagan, R. C. *et al.* Telomere dysfunction provokes regional amplification and deletion in cancer genomes. *Cancer Cell* **2,** 149–155 (2001).
98. Artandi, S. E. & DePinho, R. A. Telomeres and telomerase in cancer. *Carcinogenesis* **31,** 9–18 (2009).
99. Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* **79,** 181–211 (2009).
100. Simsek, D. & Jasin, M. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat Struct Mol Biol* **17,** 410–416 (2010).
101. Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467,** 1109–1113 (2010).
102. Alam, N. A. *et al.* Missense Mutations in Fumarate Hydratase in Multiple Cutaneous and Uterine Leiomyomatosis and Renal Cell Cancer. *J Mol Diagn* **7,** 437–443 (2004).
103. Kucherlapati, M. *et al.* Tumor progression in Apc(1638N) mice with Exo1 and Fen1

deficiencies. *Oncogene* **26,** 6297–6306 (2007).
104. Santarius, T., Shipley, J., Brewer, D., Stratton, M. R. & Cooper, C. S. A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* **10,** 59–64 (2010).
105. Poulogiannis, G. *et al.* PARK2 deletions occur frequently in sporadic colorectal cancer and accelerate adenoma development in Apc mutant mice. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 15145–15150 (2010).
106. Drusco, A. *et al.* Common fragile site tumor suppressor genes and corresponding mouse models of cancer. *J. Biomed. Biotechnol.* **2011,** 984505–984505 (2010).
107. Smith, D. I., McAvoy, S., Zhu, Y. & Perez, D. S. Large common fragile site genes and cancer. *Semin Cancer Biol* **17,** 31–41 (2006).
108. Taieb, D. *et al.* ArgBP2-dependent signaling regulates pancreatic cell migration, adhesion, and tumorigenicity. *Cancer Res* **68,** 4588–4596 (2008).
109. Farrell, C. *et al.* Somatic mutations to CSMD1 in colorectal adenocarcinomas. *Cancer Biol Ther* **7,** 609–613 (2008).
110. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
111. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Research* **19,** 1639–1645 (2009).
112. Nijman, I. J. *et al.* Mutation discovery by targeted genomic enrichment of multiplexed barcoded samples. *Nature Publishing Group* **7,** 913–915 (2010).
113. Mokry, M. *et al.* Accurate SNP and mutation detection by targeted custom microarray-based genomic enrichment of short-fragment sequencing libraries. *Nucleic Acids Res.* **38,** e116–e116 (2010).
114. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13,** 2498–2504 (2003).
115. Robinson, D. R. *et al.* Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. *Nat Med* **17,** 1646–1651 (2011).
116. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474,** 609–615 (2011).
117. Bashashati, A. *et al.* Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J. Pathol.* **231,** 21–34 (2013).
118. Futreal, P. A. *et al.* A census of human cancer genes. *CORD Conference Proceedings* **4,** 177–183 (2004).
119. Khalique, L. *et al.* Genetic intra-tumour heterogeneity in epithelial ovarian cancer and its implications for molecular diagnosis of tumours. *J. Pathol.* **211,** 286–295 (2007).
120. Yates, L. R. & Campbell, P. J. Evolution of the cancer genome. *Nat. Rev. Genet.* **13,** 795–806 (2012).
121. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).
122. Gray, J. W. J. Evidence emerges for early metastasis and parallel evolution of primary and metastatic tumors. *Cancer Cell* **4,** 4–6 (2003).
123. Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467,** 1109–1113 (2010).
124. Ferlay, J. *et al.* GLOBOCAN 2008 v2.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet]. at <Lyon, France: International Agency for Research on Cancer; 2010. Available from: http://globocan.iarc.fr, accessed on day/month/year.>
125. Malek, J. A. *et al.* Copy number variation analysis of matched ovarian primary tumors and peritoneal metastasis. *PLoS ONE* **6,** e28561 (2011).
126. Mcbride, D. J. *et al.* Tandem duplication of chromosomal segments is common in ovarian and breast cancer genomes. *J. Pathol.* **227,** 446–455 (2012).
127. Banerjee, S., Kaye, S. B. & Ashworth, A. Making the best of PARP inhibitors in ovarian cancer. *Nat Rev Clin Oncol* **7,** 508–519 (2010).
128. Tothill, R. W. *et al.* Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res* **14,** 5198–5208 (2008).
129. Verhaak, R. G. W. *et al.* Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J. Clin. Invest.* **123,** 517–525 (2013).
130. Ledermann, J. A. & Kristeleit, R. S. Optimal treatment for relapsing ovarian cancer. *Ann. Oncol.* **21 Suppl 7,** vii218–22 (2010).
131. Vaughan, S. *et al.* Rethinking ovarian cancer: recommendations for improving outcomes. in

11, 719–725 (2011).
132. Rauh-Hain, J. A. *et al.* Carcinosarcoma of the ovary compared to papillary serous ovarian carcinoma: A SEER analysis. *Gynecologic oncology* (2013). doi:10.1016/j.ygyno.2013.07.097
133. Harris, M. A. *et al.* Carcinosarcoma of the ovary. *Br. J. Cancer* **88,** 654–657 (2003).
134. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 16910–16915 (2010).
135. Zamagni, C. *et al.* Oestrogen receptor 1 mRNA is a prognostic factor in ovarian cancer patients treated with neo-adjuvant chemotherapy: determination by array and kinetic PCR in fresh tissue biopsies. *Endocrine-related cancer* **16,** 1241–1249 (2009).
136. Forbes, S. A. *et al.* COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res.* **38,** D652–7 (2010).
137. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149,** 979–993 (2012).
138. Dobbin, Z. C. & Landen, C. N. The Importance of the PI3K/AKT/MTOR Pathway in the Progression of Ovarian Cancer. *International journal of molecular sciences* **14,** 8213–8227 (2013).
139. Liu, P. *et al.* Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. *Carcinogenesis* **33,** 1270–1276 (2012).
140. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Reports* **3,** 246–259 (2013).
141. Bige, O. *et al.* Collision tumor: serous cystadenocarcinoma and dermoid cyst in the same ovary. *Archives of gynecology and obstetrics* **279,** 767–770 (2009).
142. Cotran, R. S., Kumar, V., Collins, T. & Robbins, S. L. *Robbins pathologic basis of disease*. (W.B. Saunders Company, 1999). at <http://books.google.com/books?id=k9hrAAAAMAAJ&q=cotran+(kumar+(collins+pathologic+basis+of+disease))&dq=cotran+(kumar+(collins+pathologic+basis+of+disease))&hl=&cd=1&source=gbs_api>
143. Abeln, E. C. *et al.* Molecular genetic evidence for unifocal origin of advanced epithelial ovarian cancer and for minor clonal divergence. *Br. J. Cancer* **72,** 1330–1336 (1995).
144. O'Donovan, P. J. & Livingston, D. M. BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair. *Carcinogenesis* **31,** 961–967 (2010).
145. Kloosterman, W. P. *et al.* Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer. *Genome Biol.* **12,** R103 (2011).
146. Banerjee, S. & Kaye, S. B. New strategies in the treatment of ovarian cancer: current clinical perspectives and future potential. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19,** 961–968 (2013).
147. Smolle, E. *et al.* Targeting signaling pathways in epithelial ovarian cancer. *International journal of molecular sciences* **14,** 9536–9555 (2013).
148. Ciucci, A. *et al.* Expression of the glioma-associated oncogene homolog 1 (gli1) in advanced serous ovarian cancer is associated with unfavorable overall survival. *PLoS ONE* **8,** e60145 (2013).
149. Kelleher, F. C., Cain, J. E., Healy, J. M., Watkins, D. N. & Thomas, D. M. Prevailing importance of the hedgehog signaling pathway and the potential for treatment advancement in sarcoma. *Pharmacol. Ther.* **136,** 153–168 (2012).
150. Davidson, B. Anatomic site-related expression of cancer-associated molecules in ovarian carcinoma. *Current cancer drug targets* **7,** 109–120 (2007).
151. Sawada, K., Ohyagi-Hara, C., Kimura, T. & Morishige, K.-I. Integrin inhibitors as a therapeutic agent for ovarian cancer. *J Oncol* **2012,** 915140 (2012).
152. Mantovani, A., Allavena, P., Sica, A. & Balkwill, F. Cancer-related inflammation. *Nature* **454,** 436–444 (2008).
153. Harakalova, M. M. *et al.* Multiplexed array-based and in-solution genomic enrichment for flexible and cost-effective targeted next-generation sequencing. *Nat Protoc* **6,** 1870–1886 (2011).
154. Vermaat, J. S. *et al.* Primary Colorectal Cancers and Their Subsequent Hepatic Metastases Are Genetically Different: Implications for Selection of Patients for Targeted Treatment.

Clinical Cancer Research **18,** 688–699 (2012).
155. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28,** 1811–1817 (2012).
156. Wang, L., Feng, Z., Wang, X., Wang, X. & Zhang, X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics (Oxford, England)* **26,** 136–138 (2010).
157. Rau, A., Celeux, M., Martin-Magniette, M. L. & Maugis-Rabusseau, C. Clustering high-throughput sequencing data with Poisson mixture models. *Technical report RR-7786, Inria Saclay--Ile-de-France* (2011).
158. Yang, H. H. *et al.* RG7204 (PLX4032), a selective BRAFV600E inhibitor, displays potent antitumor activity in preclinical melanoma models. *Cancer Res* **70,** 5518–5527 (2010).
159. Joseph, E. W. E. *et al.* The RAF inhibitor PLX4032 inhibits ERK signaling and tumor cell proliferation in a V600E BRAF-selective manner. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 14903–14908 (2010).
160. Swoboda, A. *et al.* Identification of genetic disparity between primary and metastatic melanoma in human patients. *Genes Chromosom. Cancer* **50,** 680–688 (2011).
161. Stoecklein, N. H. & Klein, C. A. Genetic disparity between primary tumours, disseminated tumour cells, and manifest metastasis. *Int J Cancer* **126,** 589–598 (2010).
162. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481,** 506–510 (2012).
163. Trunzer, K. *et al.* Pharmacodynamic effects and mechanisms of resistance to vemurafenib in patients with metastatic melanoma. *J Clin Oncol* **31,** 1767–1774 (2013).
164. Balch, C. M. *et al.* Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol* **27,** 6199–6206 (2009).
165. Tirumani, S. H. *et al.* Radiologic assessment of earliest, best, and plateau response of gastrointestinal stromal tumors to neoadjuvant imatinib prior to successful surgical resection. *Eur J Surg Oncol* (2013). doi:10.1016/j.ejso.2013.10.021
166. Schiavon, G. *et al.* Tumor volume as an alternative response measurement for imatinib treated GIST patients. *PLoS ONE* **7,** e48372 (2012).
167. Emery, C. M. C. *et al.* MEK1 mutations confer resistance to MEK and B-RAF inhibition. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 20411–20416 (2009).
168. Greger, J. G. J. *et al.* Combinations of BRAF, MEK, and PI3K/mTOR inhibitors overcome acquired resistance to the BRAF inhibitor GSK2118436 dabrafenib, mediated by NRAS or MEK mutations. *Molecular Cancer Therapeutics* **11,** 909–920 (2012).
169. Nazarian, R. R. *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468,** 973–977 (2010).
170. Wagle, N. *et al.* MAP Kinase Pathway Alterations in BRAF-Mutant Melanoma Patients with Acquired Resistance to Combined RAF/MEK Inhibition. *Cancer Discovery* (2013). doi:10.1158/2159-8290.CD-13-0631
171. Romano, E. *et al.* Identification of Multiple Mechanisms of Resistance to Vemurafenib in a Patient with BRAFV600E-Mutated Cutaneous Melanoma Successfully Rechallenged after Progression. *Clin Cancer Res* **19,** 5749–5757 (2013).
172. Mao, M. M. *et al.* Resistance to BRAF inhibition in BRAF-mutant colon cancer can be overcome with PI3K inhibition or demethylating agents. **19,** 657–667 (2013).
173. Van Allen, E. M. *et al.* The Genetic Landscape of Clinical Resistance to RAF Inhibition in Metastatic Melanoma. *Cancer Discovery* (2013). doi:10.1158/2159-8290.CD-13-0617
174. Lito, P., Rosen, N. & Solit, D. B. Tumor adaptation and resistance to RAF inhibitors. *Nat Med* **19,** 1401–1409 (2013).
175. Whittaker, S. R. *et al.* A genome-scale RNA interference screen implicates NF1 loss in resistance to RAF inhibition. *Cancer Discovery* **3,** 350–362 (2013).
176. Shi, H. H. *et al.* Preexisting MEK1 exon 3 mutations in V600E/KBRAF melanomas do not confer resistance to BRAF inhibitors. *Cancer Discov* **2,** 414–424 (2012).
177. Biechele, T. L. T. *et al.* Wnt/β-catenin signaling and AXIN1 regulate apoptosis triggered by inhibition of the mutant kinase BRAFV600E in human melanoma. *Sci Signal* **5,** ra3–ra3 (2012).
178. Liu, L. *et al.* Proteomic characterization of the dynamic KSR-2 interactome, a signaling scaffold complex in MAPK pathway. *ACTA-BIOENERG* **1794,** 1485–1495 (2009).
179. Rajakulendran, T. *et al.* CNK and HYP form a discrete dimer by their SAM domains to

mediate RAF kinase signaling. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 2836–2841 (2008).
180. Shi, H. H. *et al.* Melanoma whole-exome sequencing identifies (V600E)B-RAF amplification-mediated acquired B-RAF inhibitor resistance. *Nat Commun* **3,** 724–724 (2011).
181. Wagenaar, T. R. *et al.* Resistance to vemurafenib resulting from a novel mutation in the BRAFV600E kinase domain. *Pigment Cell Melanoma Res* (2013). doi:10.1111/pcmr.12171
182. Efimova, T. T., Deucher, A. A., Kuroki, T. T., Ohba, M. M. & Eckert, R. L. R. Novel protein kinase C isoforms regulate human keratinocyte differentiation by activating a p38 delta mitogen-activated protein kinase cascade that targets CCAAT/enhancer-binding protein alpha. *J Biol Chem* **277,** 31753–31760 (2002).
183. Dissanayake, S. K. S. & Weeraratna, A. T. A. Detecting PKC phosphorylation as part of the Wnt/calcium pathway in cutaneous melanoma. *Methods Mol Biol* **468,** 157–172 (2007).
184. Prahallad, A. *et al.* Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* **483,** 100–103 (2012).
185. Wagle, N. N. *et al.* Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J Clin Oncol* **29,** 3085–3096 (2011).
186. Eisenhauer, E. A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European Journal of Cancer* **45,** 228–247 (2008).
187. van Kessel, C. S. C. *et al.* Semi-automatic software increases CT measurement accuracy but not response classification of colorectal liver metastases after chemotherapy. *Eur J Radiol* **81,** 2543–2549 (2012).
188. Hoogstraat, M. *et al.* Genomic and transcriptomic plasticity in treatment-naive ovarian cancer. *Genome Research* (2013). doi:10.1101/gr.161026.113
189. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
190. Chin, L., Andersen, J. N. & Futreal, P. A. Cancer genomics: from discovery science to personalized medicine. *Nat Med* **17,** 297–303 (2011).
191. Kim, S. Y. & Speed, T. P. Comparing somatic mutation-callers: beyond Venn diagrams. *BMC Bioinformatics* **14,** 189–189 (2012).
192. Rashid, M., Robles-Espinoza, C. D., Rust, A. G. & Adams, D. J. Cake: a bioinformatics pipeline for the integrated analysis of somatic variants in cancer genomes. *J Gerontol* **29,** 2208–2210 (2013).
193. Decarlo, K., Emley, A., Dadzie, O. E. & Mahalingam, M. Laser capture microdissection: methods and applications. *Methods Mol Biol* **755,** 1–15 (2011).
194. de Bruin, E. C. *et al.* Macrodissection versus microdissection of rectal carcinoma: minor influence of stroma cells to tumor cell gene expression profiles. *BMC Genomics* **6,** 142 (2005).
195. Colvin, E. K. Tumor-associated macrophages contribute to tumor progression in ovarian cancer. *Front Oncol* **4,** 137–137 (2013).
196. Roxburgh, C. S., Salmond, J. M., Horgan, P. G., Oien, K. A. & McMillan, D. C. Tumour inflammatory infiltrate predicts survival following curative resection for node-negative colorectal cancer. *European Journal of Cancer* **45,** 2138–2145 (2008).
197. Demaria, S. *et al.* Cancer and inflammation: promise for biologic therapy. *J. Immunother.* **33,** 335–351 (2010).
198. Robert, C. *et al.* Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med* **364,** 2517–2526 (2011).
199. Engelman, J. A. *et al.* Effective use of PI3K and MEK inhibitors to treat mutant Kras G12D and PIK3CA H1047R murine lung cancers. *Nat Med* **14,** 1351–1356 (2008).
200. Dumble, M. *et al.* Discovery of Novel AKT Inhibitors with Enhanced Anti-Tumor Effects in Combination with the MEK Inhibitor. *PLoS ONE* **9,** e100880 (2014).
201. Greger, J. G., Eastman, S. D., Zhang, V. & Bleam, M. R. Combinations of BRAF, MEK, and PI3K/mTOR inhibitors overcome acquired resistance to the BRAF inhibitor GSK2118436 dabrafenib, mediated by NRAS or MEK …. *Molecular cancer …* (2012).
202. Courtney, K. D., Corcoran, R. B. & Engelman, J. A. The PI3K Pathway As Drug Target in Human Cancer. *Journal of Clinical Oncology* **28,** 1075–1083 (2010).
203. Wong, Y.-N. *et al.* Phase II trial of cetuximab with or without paclitaxel in patients with advanced urothelial tract carcinoma. *Journal of Clinical Oncology* **30,** 3545–3551 (2012).
204. Kim, T.-M. *et al.* Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. *Genome Research* **23,** 217–227 (2013).

205. Molenaar, J. J. *et al.* Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature* **483,** 589–593 (2012).
206. Hirsch, D. *et al.* Chromothripsis and focal copy number alterations determine poor outcome in malignant melanoma. *Cancer Res* **73,** 1454–1460 (2013).
207. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat Genet* **45,** 1134–1140 (2013).
208. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502,** 333–339 (2013).
209. Yuan, Y. *et al.* Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nature Biotechnology* **32,** 644–652 (2014).
210. Wang, F. *et al.* High EGFR copy number predicts benefits from tyrosine kinase inhibitor treatment for non-small cell lung cancer patients with wild-type EGFR. *J Transl Med* **11,** 90 (2013).
211. Yung, W. K. A. *et al.* Safety and efficacy of erlotinib in first-relapse glioblastoma: a phase II open-label study. *Neuro-oncology* **12,** 1061–1070 (2010).
212. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504,** 389–393 (2013).
213. Sato, T. *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141,** 1762–1772 (2011).
214. Kopetz, S., Lemos, R. & Powis, G. The promise of patient-derived xenografts: the best laid plans of mice and men. *Clin Cancer Res* **18,** 5160–5162 (2012).
215. Jin, K. *et al.* Patient-derived human tumour tissue xenografts in immunodeficient mice: a systematic review. *Clin Transl Oncol* **12,** 473–480 (2010).
216. Schmelzle, K. & White, F. M. Phosphoproteomic approaches to elucidate cellular signaling networks. *Curr. Opin. Biotechnol.* **17,** 406–414 (2006).
217. Boja, E. S. & Rodriguez, H. Proteogenomic convergence for understanding cancer pathways and networks. *Clin Proteomics* **11,** 22 (2014).
218. Hussain, S. *et al.* Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. *Human Molecular Genetics* **13,** 1241–1248 (2004).
219. Tishkoff, D. X., Amin, N. S., Viars, C. S., Arden, K. C. & Kolodner, R. D. Identification of a human gene encoding a homologue of Saccharomyces cerevisiae EXO1, an exonuclease implicated in mismatch repair and recombination. *Cancer Res* **58,** 5027–5031 (1998).
220. Melo, J. V. BCR-ABL gene variants. *Baillieres Clin Haematol* **10,** 203–222 (1996).
221. Solomon, B., Varella-Garcia, M. & Camidge, D. R. ALK gene rearrangements: a new therapeutic target in a molecularly defined subset of non-small cell lung cancer. *J Thorac Oncol* **4,** 1450–1454 (2009).
222. Ou, S.-H. I., Bartlett, C. H., Mino-Kenudson, M., Cui, J. & Iafrate, A. J. Crizotinib for the Treatment of ALK-Rearranged Non-Small Cell Lung Cancer: A Success Story to Usher in the Second Decade of Molecular Targeted Therapy in Oncology. *Oncologist* **17,** 1351–1375 (2011).
223. O'Brien, S. G. *et al.* Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *N Engl J Med* **348,** 994–1004 (2003).
224. Sasaki, T., Rodig, S. J., Chirieac, L. R. & Jänne, P. A. The biology and treatment of EML4-ALK non-small cell lung cancer. *Eur. J. Cancer* **46,** 1773–1780 (2010).
225. Esteller, M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* **21,** 5427–5440 (2002).
226. Catteau, A., Harris, W. H., Xu, C. F. & Solomon, E. Methylation of the BRCA1 promoter region in sporadic breast and ovarian cancer: correlation with disease characteristics. *Oncogene* **18,** 1957–1965 (1999).
227. Bello, M. J., Gonzalez-Gomez, P. & Alonso, M. E. Methylation analysis of cell cycle control genes RB1, p14ARF and p16INK4a in human gliomas. *Cancer* (2004).
228. Herman, J. G. *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 9700–9704 (1994).
229. Shanmuganathan, R. *et al.* Conventional and nanotechniques for DNA methylation profiling. *J Mol Diagn* **15,** 17–26 (2013).
230. Clark, S. J., Statham, A., Stirzaker, C., Molloy, P. L. & Frommer, M. DNA methylation: bisulphite modification and analysis. *Nat Protoc* **1,** 2353–2364 (2005).
231. Huang, F. W. *et al.* Highly Recurrent TERT Promoter Mutations in Human Melanoma.

Science **339,** 957–959 (2013).
232. Vinagre, J. *et al.* Frequency of TERT promoter mutations in human cancers. *Nat Commun* **4,** – (2013).
233. Killela, P. J., Reitman, Z. J. & Jiao, Y. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. in 1–6 (2013). doi:10.1073/pnas.1303607110/-/DCSupplemental
234. Friedlander, T. W., Premasekharan, G. & Paris, P. L. Looking back, to the future of circulating tumor cells. *Pharmacol. Ther.* **142,** 271–280 (2014).
235. Lowes, L. E., Hedley, B. D., Keeney, M. & Allan, A. L. User-defined protein marker assay development for characterization of circulating tumor cells using the CellSearch® system. *Cytometry A* **81,** 983–995 (2012).
236. Onstenk, W., Gratama, J. W., Foekens, J. A. & Sleijfer, S. Towards a personalized breast cancer treatment approach guided by circulating tumor cell (CTC) characteristics. *CANCER TREATMENT REVIEWS* **39,** 691–700 (2013).
237. de Bono, J. S. *et al.* Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin Cancer Res* **14,** 6302–6309 (2008).
238. Cristofanilli, M. *et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med* **351,** 781–791 (2004).
239. Botteri, E. *et al.* Modeling the relationship between circulating tumour cells number and prognosis of metastatic breast cancer. *Breast Cancer Res Treat* **122,** 211–217 (2010).
240. Cohen, S. J. *et al.* Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *J Clin Oncol* **26,** 3213–3221 (2008).
241. Heitzer, E. *et al.* Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. *Cancer Res* **73,** 2965–2975 (2013).
242. Magbanua, M. J. M. *et al.* Genomic profiling of isolated circulating tumor cells from metastatic breast cancer patients. *Cancer Res* **73,** 30–40 (2012).
243. Lohr, J. G. *et al.* Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nature Biotechnology* (2014). doi:10.1038/nbt.2892
244. Allard, W. J. *et al.* Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin Cancer Res* **10,** 6897–6904 (2004).
245. van de Stolpe, A., Pantel, K., Sleijfer, S., Terstappen, L. W. & Toonder, den, J. M. J. Circulating tumor cell isolation and diagnostics: toward routine clinical use. in *Cancer Res.* **71,** 5955–5960 (2011).
246. Murtaza, M. *et al.* Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* – (2013). doi:10.1038/nature12065
247. Esposito, A. *et al.* Monitoring tumor-derived cell-free DNA in patients with solid tumors: clinical perspectives and research opportunities. *CANCER TREATMENT REVIEWS* **40,** 648–655 (2014).

Dankwoord

Er was eens, lang geleden, in een klein stadje niet zo heel ver van Utrecht vandaan, een meisje dat achter de toonbank van een dierenwinkel stond en droomde van grootsere dingen dan het verkopen van hondenbrokken en kattenbaksteentjes. Ze besloot die droom te volgen en de grote stap te zetten om terug naar school te gaan, om het wonderbaarlijke vak van bioinformagicus te leren. Wat ze nooit had verwacht, was dat dit ertoe zou leiden dat ze acht jaar later voor een committee van professoren zou staan om haar proefschrift te verdedigen. Dit heeft ze echter niet helemaal zelf gedaan, maar met de hulp van de vele geweldige mensen en andere wezens die ze onderweg tegenkwam, zoals…

Once upon a time, in a small town not so very far from Utrecht, stood a girl at the counter of a petstore who dreamt of bigger things than selling dog chow and kitty litter. She decided to follow those dreams and took the big step of going back to school, to learn the wonderous trade of bioinformagician. What she never had expected, was that this would lead her to stand in front of a committee of professors eight years later to defend her thesis. However, she did not accomplish this all by herself, but with the help of many amazing people and other creatures she met along the way. In order of appearance-ish, there are…

Papi y **Mami**, jullie waren er natuurlijk het eerst! Jullie hebben mij al vanaf de basisschool mijn eigen keuzes laten maken, zelfs toen ik op het laatste moment besloot toch niet te gaan studeren maar met dieren wilde gaan werken, en later na een paar jaar full-time werken toch weer wel ging studeren… Ik heb niet altijd de juiste keuzes gemaakt, maar van deze twee zal ik nooit spijt hebben. Beide keuzes waren niet mogelijk geweest zonder jullie steun, hier ben ik jullie dan ook enorm dankbaar voor. **Broertje,** als ik jou een tip mag geven… Blijf zoals je bent!

Meneer Ploeger, mijn leraar biologie en decaan op de middelbare school. U was de eerste die mij vertelde dat ik vooral moest doen wat ik wilde doen, niet wat ik dacht dat er van mij verwacht werd. Dit heeft er mede voor gezorgd dat ik dit pad heb gevolgd en ben waar ik nu ben, ook al is het toch een heel andere richting geworden dan die ik eerst gekozen had.

Hans, **Marja**, **Aldo**, **André** en **Elizabeth**, jullie hebben de basis gelegd van alles wat ik nu weet en mijn interesse weten te wekken en vast te houden, zelfs als het ging om chemie, statistiek of aminozuur eigenschappen. Die laatste kennis gebruik ik overigens nog regelmatig. Mede dankzij jullie heb ik niet alleen veel geleerd, maar ook een enorm leuke tijd gehad in Leiden.

Daar hebben natuurlijk ook mijn **klasgenoten van het HBO** aan bijgedragen. Er was altijd een hele relaxte, gezellige sfeer in onze groep en ik vind het echt heel leuk dat we zo nu en dan nog contact hebben, ook al zijn sommige mensen een heel andere kant op gegaan of zelfs naar het buitenland vertrokken. We moeten toch echt maar een keer naar die strandtent in Noordwijk gaan!

Emile, eigenlijk staat alles wat ik zeggen wilde al in het boek wat je bij je afscheid van het UMC hebt gekregen, maar toch wil ik nog eens benadrukken dat ik het heel erg op prijs stel dat ik van jou de kans heb gekregen om te promoveren. Deze kans had ik niet overal gekregen.

Edwin, ook een hele grote “dank je” voor jou, dat je me zo op hebt genomen in je groep. Dit heeft een wereld aan mogelijkheden geopend die ik anders niet had gehad. Bovendien waren de werkbesprekingen en symposia van zowel het AMG als de Cuppen-groep Hubrecht interessant en leerzaam, en de retraites en borrels onvergetelijk!

Marco, zonder jou had ik het sowieso niet gered. Enorm bedankt voor de begeleiding, je betrokkenheid en je (bijna) nooit aflatende positiviteit en enthousiasme, voor het feit dat ik altijd bij je kon binnenlopen met vragen of om ‘even te relativeren’. Dit boekje is net zo goed van jou als van mij. **Ies**, jij ook zeker bedankt voor al je begeleiding, voor de vele brainstorm-momenten en voor de mogelijkheid om stoom bij je af te kunnen blazen.

De helft van dit boekje was er niet geweest als **Wigard** mij niet de kans had gegeven om mee te werken aan zijn projecten. Dus Wigard, hartelijk dank voor de mogelijkheid en voor alles wat ik met deze projecten geleerd heb, zowel qua data-analyse als qua artikelen schrijven. **Martijn**, jij ook bedankt voor de mogelijkheden om mee te werken aan een meer klinisch project en om de eerste opzet van het artikel volledig zelf te schrijven.

Mark en **Stef**, my partners in crime! Jullie zijn allebei fantastische mensen, zowel om mee te werken als om mee in één kantoor te zitten (of in de auto naar Terschelling of zo). Bedankt voor alles en vooral voor de gezelligheid. Ik ga jullie missen! **Martin**, **Robert** en **Annelies**, jullie ook. Keep up the good vibe!

Mirjam, Ilse, Ivo, Karen, Ruben, Mark en vooral **Nicolle**, aangezien ik er zelf helemaal niks van bak in het lab was er ook nooit een boekje gekomen zonder jullie bijdrage. Mijn dank is groot!

Monique en **Alice**, heel erg bedankt voor alles wat jullie in de afgelopen paar jaren voor mij geregeld en gepland hebben en dat ik altijd bij jullie terecht kon met wat voor vragen dan ook. **Ida**, we hebben maar kort een kamer gedeeld, maar ook jou wil ik heel graag bedanken voor je goede adviezen en prettige gesprekken.

All other colleagues of the **Medical Genetics** department, in particular **Nayia**, **Kirsten**, **Terry**, **Glen**, **Magdalena**, **Jessica**, **Sara**, **Mariëlle, Ellen** and **Lars**, for all your help, advice, entertainment en gezelligheid.

Alle collega's van de **Medische Oncologie** natuurlijk ook, in het bijzonder **Geert**, **Christa**, **Ilse** en **Fleur**. Wat ik misschien nog wel het leukst vond aan onze samenwerkingen en gesprekken is het 'leren praten' met mensen uit een complete andere discipline, zowel om uit te kunnen leggen wat ik nu de hele dag achter de computer doe als om te begrijpen wat jullie doen. **Geert**, je bent meer dan welkom om ook op het NKI nog een keer bams te komen mergen als je je verveelt…

Very importantly, the whole **Cuppen-groep** at the Hubrecht. Just like Edwin you accepted and included me in your group like I belonged there, and some of my best memories of the past 4 years are with you. You guys are awesome. **Sander**, hodor hodor, hodor HODOR hodor! Hodor. **Ewart** en **Marieke**, voor alle adviezen en borrel-gesprekken. Waarschijnlijk zegt altijd iedereen dit en gebeurt het veel te weinig, maar we'll keep in touch…

Iedereen met wie ik verder heb samengewerkt aan projecten, zoals van de pathologie**: Stefan, Roel, John, Carmen, Joyce, Ton** en **Marja**, NKI: **Magali**, **Esther**, **Daniël** en **Lodewyk**, Erasmus: **John**, **Els** en **Jozien**, en iedereen die ik ongetwijfeld vergeten ben.

Querido, ook al snap je vaak niet waar ik me nu toch weer zo enorm druk over maak (en heb je daar waarschijnlijk gelijk in), toch steun je mij wel hier in en accepteer je dit. Je hebt me vrij gelaten om in weekenden en avonden achter mijn laptop te duiken om 'toch nog even wat te doen', maar je hebt er ook voor gezorgd dat ik dat regelmatig niet deed door voor de nodige afleiding te zorgen. Dus thank you too!

Stay tuned for more adventures of the girl from the petstore, when she continues her journey in Amsterdam. Scheduled for fall 2014…

List of publications

Peer reviewed Pubmed publications

Hoogstraat M*, de Pagter MS*, Cirkel GA*, van Roosmalen MJ, Harkins TT, Duran K, Kreeftmeijer J, Renkens I, Witteveen PO, Lee CC, Nijman IJ, Guy T, van 't Slot R, Jonges TN, Lolkema MP, Koudijs MJ, Zweemer RP, Voest EE, Cuppen E, Kloosterman WP. Genomic and transcriptomic plasticity in treatment-naïve ovarian cancer. Genome Res. 2014 Feb;24(2):200-11. doi: 10.1101/gr.161026.113.

Nijman IJ, van Montfrans JM, **Hoogstraat M**, Boes ML, van de Corput L, Renner ED, van Zon P, van Lieshout S, Elferink MG, van der Burg M, Vermont CL, van der Zwaag B, Janson E, Cuppen E, Ploos van Amstel JK, van Gijn ME. Targeted next-generation sequencing: a novel diagnostic tool for primary immunodeficiencies. J Allergy Clin Immunol. 2014 Feb;133(2):529-34. doi: 10.1016/j.jaci.2013.08.032.

Kloosterman WP, **Hoogstraat M**, Paling O, Tavakoli-Yaraki M, Renkens I, Vermaat JS, van Roosmalen MJ, van Lieshout S, Nijman IJ, Roessingh W, van 't Slot R, van de Belt J, Guryev V, Koudijs M, Voest E, Cuppen E. Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer. Genome Biol. 2011 Oct 19;12(10):R103. doi: 10.1186/gb-2011-12-10-r103.

Beaufort CM, Helmijr JCA, Piskorz AM, **Hoogstraat M**, Ruigrok-Ritstier K, Besselink NJM, Murtaza M, van IJcken WFJ, Heine AAJ, Smid M, Koudijs MJ, Brenton JD, Berns EMJJ, Helleman J. Ovarian Cancer Cell line Panel (OCCP): clinical importance of in vitro morphological subtypes. PLoS One. 2014 Sep 17;9(9):e103988. doi: 10.1371/journal.pone.0103988.

In press

Hoogstraat M, Hinrichs JWJ, Besselink NJM, Radersma-van Loon JH, de Voijs CMA, Peeters A, Nijman IJ, de Weger RA, Voest EE, Willems SM, Cuppen E, Koudijs MJ. Simultaneous detection of clinically relevant mutations and amplifications for routine cancer pathology. Journal of Molecular Diagnostics

In preparation

Weeber F*, Koudijs MJ*, **Hoogstraat M**, Besselink NJM, van Lieshout S, Nijman IJ, Cuppen E, Offerhaus GJ, Voest EE. Effective therapeutic intervention and comprehensive genetic analysis of mTOR signaling in PEComa: a case report.

Hoogstraat M*, Gadellaa-van Hooijdonk CG*, Ubink I, Besselink NJM, Pieterse M, Veldhuis W, van Stralen M, Meijer EFJ, Willems SM, Hadders MA, Koudijs MJ, Cuppen E, Voest EE, Lolkema MPJ. Longitudinal imaging and sequence analysis of melanoma metastases reveals heterogeneity in vemurafenib response driven by diverse resistance mechanisms.

Lips EH*, Michaut M*, **Hoogstraat M**, Mulder L, Besselink NJM, Koudijs MJ, Voest EE, Bernards R, Nederlof PM, Wesseling J, Rodenhuis S, Wessels LFA. Next generation sequencing of triple negative breast cancer to find predictors for chemotherapy response.

Curriculum Vitae

Marlous Hoogstraat was born on May 28th, 1984 in Leidschendam, the Netherlands. She graduated from high-school at Veurs College in Leidschendam in 2002, with a thesis on “DNA and medical science”. In 2006, she started her bachelor study on bioinformatics at University college Leiden, and graduated *cum laude* in 2010 after an internship at the Universitat Autonoma of Barcelona in the lab of Leonardo Pardo, and a second internship at Plant Research International in Wageningen. The second internship resulted in a publication of a web-based tool for the analysis of plant ChIP-seq data in Nucleic Acid Research.

After finishing her bachelor, Marlous joined the groups of Emile Voest at the department of Medical Oncology and Edwin Cuppen at the department of Human Genetics at the University Medical Center Utrecht, as a PhD student in the graduate school of Cancer, Genetics & Developmental Biology. The results of her research there are presented in this thesis. In September 2014, Marlous continued her career as a postdoctoral researcher in the lab of Lodewyk Wessels on Computational Biology at the Dutch Cancer Institute.