Master's Thesis

Implementation and Comparative Assessment of Diagnostic Cancer Gene Panels in the Molecular Pathology Laboratory

University of Luxembourg

Faculty of Science, Communication and Technology

Master in Integrated Systems Biology

by

Ben Flies

(010081174D)

Abstract

Contents

1 Introduction

Cancer represents a huge burden for health care systems worldwide and is one of the leading death causes. Scientific discoveries in the last decade have had an enormous impact on our understanding of the underlying causes of cancer. The development of omics techniques, in combination with advanced computational power, has lead to an explosion of biological data. It has become clear that cancer is an incredibly complex malignancy, which is affected by genetic, environmental and behavioural factors. The research community is trying to interprete this vast amount of data with the goal to get a deeper understanding of cancer and to cure it eventually. In recent years, several drugs have been approved, which target proteins needed for cancer development, proliferation or metastasis. Molecular testing is employed to check whether these targeted drugs would be of benefit. In that regard, Next-Generation Sequencing (NGS) is an interesting method to gain deep insights into the genetic information of a tumor and to guide personalized therapy.

1.1 The cancer genome

DNA undergoes continuous damage. In cancer cells, the equilibrium between DNA damage and repair systems is dysbalanced [1]. Genetic and epigenetic alterations, in combination with several environmental factors, such as inflammation, enable the hallmarks of cancer [2]. These include replicative immortality, resistance to cell death, ongoing proliferative signaling, invasion and metastasis, growth suppressor evasion, inducement of angiogenesis, energy metabolism reprogramming and evasion of immune destruction.

It is widely accepted that tumors accumulate somatic mutations during their progression in malignancy [3] [4]. DNA can be damaged by endogenous and environmental agents [5]. Carcinogenic substances produced by industry [6] [7] or present in tobacco smoke [8] are known to increase cancer risk. Cellular metabolic processes also produce DNA-damaging products that induce cancer, such as reactive oxygen species [9] [10]. DNA lesions can escape DNA repair mechanisms if the damage happens in an inaccessible region of the DNA or if the DNA repair system is defective [11].

Also, the process of DNA replication has an intrinsic error rate. It has been estimated that DNA replication by DNA plymerase followed by mismatch repair has an overall error rate to 10^{-6} to 10^{-8} [5]. Several DNA polymerases exist. They differ in their error rates, but can be used interchangeably. DNA polymerase β has a much worse error rate than DNA polymerase δ or ϵ. There is evidence that DNA β is increased in some tumors [12], resulting in increased mutagenesis.

Processes affecting chromosomal and microsatellite integrity instability contribute to genomic instability in cancer cells. **Microsatellite instability (MSI)** is caused by inactivation or loss of DNA mismatch repair [13]. Microsatellite elongation or shortening is a consequence of defective or inactive DNA mismatch repair (MMR), which corrects base replication errors [14]. DNA polymerase has a higher error rate in repetitive DNA sequences. When MMR genes are inactivated or defective, the replication mistakes in microsatellites cannot be corrected: MSI is the consequence. In some cancers, MSI can occur despite functional MMR through frameshift mutations at microsatellites. MSI is often associated with cancers harboring mutations in TGFβRII, EGFR, PTEN, and BAX, as they contain such microsatellites [15]. Microsatellite and mismatch errors by DNA polymerase usually result in insertions or deletions and in point mutations, which can be silent, missense or nonsense. These mutations occur at the nucleotide level and might affect the protein structure, leading to defective or overactive enzymes.

Chromosomal instability (CIN) is a common observation in solid tumors, especially in colorectal cancer [14]. Chromosome missegregation plays a crucial role in cancer adaptation [16]. Defects in proteins needed for chromosome segregation lead to chromosome missegregation. This leads to telomere dysfunction, faulty sister chromatid cohesion, loss of heterozygosity (LOH), hypo– or hyperactive spindle assembly checkpoint or defective centrosome duplication and aneuploidy. [14]. About 70% of solid tumors are aneuploid [17]. Another chromosomal instability process has been described recently: chromothripsis happens when chromosomes are fragmented [18] [19] [19]. The cell tries to repair the chromosomes, but this process is far from being perfect, leading to massive chromosomal rearrangements. The question whether CIN is a cause or consequence of tumor development remains unanswered.

Epigenetic changes do not involve changes in the coding sequence, but affect gene expression. Gene expression can be influenced by histone modifications, dysregulation of DNA–binding of transcription factors or altered CpG island methylation. In normal cells, CpG islands in gene promoters are usually unmethylated. This is associated with active transcription. Other CpG islands across the genome are usually methylated. In cancer cells, this situation is often inverted. CpG islands in tumor suppressor promoters are found to be hypermethylated in many tumors. Their gene expression is thus drastically decreased. CpG methylation of tumor suppressor genes is found in 35–40% of colorectal cancers.

1.1.1 Driver and passenger mutations

Cancer progression is a process that recognizes basic Darwinian evolution principles [20] [21] [22] [23]. The population of cancer cells harbors heritable genetic variation. These mutations may be of germline origin or may occur through somatic processes. If the occurring mutations are non–deleterious, they can be passed on to the next generation of cells. The second process, which has to take place in Darwinian evolution, is natural selection. Each cell exhibits a unique combination of genetic and environmental perturbations. Cells are in competition for a variety of resources in their microenvironment, which include space, oxygen and nutrients. Eventually, cells with the best fitness, e.g. with the highest proliferative potential and the lowest death rate, are then selected through natural selection principles and will outlast less fit cells. Additionally, these cells will continue to accumulate new mutations. This results in sequential waves of clonal expansion [20].

Genomic instability in cancerous cells becomes a critical mechanism if it affects oncogenes or tumor suppressor genes, which have the potential to be causative tumor 'driver' mutations. The identification of driver mutations has been a central aim of cancer research. Mutations in at least 350 human genes are found recursively in cancer genomes and are believed to contribute to cancerogenesis [24]. These driver mutations are positively selected during cancer progression and confer a selective advantage to the cells harboring them. Many alterations found in cancer cells are passenger mutations, which occur coincidally or subsequently to driver mutations. These mutations are defined to not contribute to the selective fitness of the cell, even though this conception has been challenged by stochastic tumor progression simulations [25]. Some studies have reported that cancer cells carry 40–80 somatic mutations, and only 5–15 of them are driver mutations [26].

Estimating the number of somatic driver and passenger mutations and the rate at which they occur is not well established [27]. Two tumors, even though histologically indistinguishable, might present different subsets of mutations [27] [28]. This observation has been defined as inter-tumor heterogeneity. Additionally, tumors present heterogeneity at the intra–tumor level [29]: subclones of the tumor might present different mutations.

Tumor suppressor genes protect a cell from entering the path to cancer. They comprise genes encoding for cell adhesion proteins, DNA repair proteins, proteins acting in apoptosis pathways, or cell cycle proteins [30]. The action of these proteins inhibits metastasis, excessive cell survival or proliferation. Tumor suppressors mostly follow the two-hit hypothesis [31]: to inactivate the tumor-protecting role of tumor suppressors, two genetic events, often LOH in combination with silencing

point mutations or silencing of both alleles by somatic events, are necessary to inactivate both alleles of the gene. Compared to dominant oncogenes, tumor suppressor genes are often considered to be recessive. Alternatively, tumor progression can be influenced by functional haploinsufficiency of tumor suppressors [32]. According to this conception, a disease state can emerge if a cell / organism has only one functional copy of a given gene and if it cannot produce enough of a gene product to establish a wild-type condition. APC and TP53 are amongst the best known tumor suppressors.

In the canonical Wnt signaling pathway, a destruction complex, including APC, leads to β–catenin phosphorylation, followed by ubiquitination, marking it for degradation in the proteasome. Activation of Wnt signaling inhibits the destruction complex. Consequently, β–catenin is no longer marked for degradation and can translocate to the nucleus, where it acts on gene expression of target genes [33]. In many tumors, loss or dysfunction of APC leads to β–catenin accumulation in the nucleus even in the absence of an extracellular stimulus, resulting in increased cell migration and decreased cell adhesion and apoptosis [34].

TP53 is the master guardian of the genome [35]. In normal situations, p53, the protein encoded by TP53, is targeted for ubiquitination and degradation in the proteasome [36]. In case of cellular stress, p53 is no longer ubiquitinated. p53 can then stop the cell cycle at the G1/S and G2/M transitions, induce DNA repair, and induce apoptosis if the damage cannot be repaired [37]. One mechanism by which p53 acts on cell-cycle arrest is by activating expression of p21. p21 binds to the G1/S transition complex and inhibits its activity, leading to cell-cycle arrest [37]. Inactivation or mutation of TP53 is a crucial step in many cancers, leading to a loss of control over DNA stability [38].

Oncogenes comprise several GTPases, transcription factors, receptor tyrosine kinases and growth factors [39]. Overexpressed or overactive versions of these proteins lead to increased mitogenic signals, causing increased cell growth or proliferation. Two important oncogenic pathways include the RAS–RAF–MEK–ERK and PTEN–PI3K–AKT pathways, which can both be activated by ligand–binding on Epithelial Growth Factor. EGFR is a cell surface tyrosine kinase receptor [40]. It is anchored in the cytoplasmic membrane and is composed of an intracytoplasmic tyrosine kinase domain, a short hydrophobic transmembrane domain and an extracellular ligand-binding domain [41]. Ligand binding causes a conformational change of the receptor, which leads to homo– or heterodimerization, followed by an auto– and cross–phosphorylation of key tyrosine residues on its cytoplasmic domain [41]. This forms docking sites for cytoplasmic adaptor proteins that contain phosphotyrosine-binding and Src homology 2 domains.

Signaling through the PI3K–AKT pathway leads to cell growth, proliferation and survival. The signaling cascade is initiated by integrins, cytokine receptors, G–protein coupled receptors, and receptor tyrosine kinases, such as EGFR [42]. Activation of the receptor results in production of PIP3 by activation of PI3K. PIP3 is anchored in the cell membrane and acts as docking site for proteins containing PH domains, such as PDK1. PIP3-bound PDK1 partially activates Akt by phosphorylation [43]. Full activation of Akt is achieved by phosphorylation of PDK1 by mTORC2 [44]. Activated Akt then acts on a variety of proteins necessary for protein synthesis, glucose metabolism, cell survival / death and proliferation. The phosphatases PP2A and PHLPP can dephosphorylate and thereby inactivate Akt [44]. Additionally, PTEN dephosphorylates PIP3 and indirectly also inactivates Akt [44]. Dysregulation of the PI3K–AKT has been associated with several human diseases including neurological diseases, diabetes and cancer [45]. In cancer, inactivation of PTEN and kinase activity activating mutations on PI3K and Akt are found recursively, leading to enhanced signaling, leading to inhibition of apoptosis and increased proliferation [46].

In the RAS–RAF–MEK–ERK pathway, ligand binding on cell surface receptor tyrosine kinases activates the receptor. One of these receptors is the EGFR. GRB2 binds to Tyr1068 of EGFR through its SH2 domain and recruits SOS, a guanine nucleotide exchange factor [47]. Grb2 and SOS then form a complex with the activated EGFR, which activates SOS [47]. Activated SOS promotes recruition of Ras proteins to the activated EGFR. Through its GEF activity, SOS then induces GDP removal from Ras proteins, which can subsequently bind GTP and become active. Activated Ras recuits Raf proteins to the cell membrane and binds to their N-terminus. The activation of Raf, serine/threonine kinase proteins, is complex. In fact, Raf proteins are considered as gatekeepers of the RAS–RAF–MAPK pathway. In its inactive form, Raf is present in a 'closed' conformation, in which an autoinhibitory domain blocks the catalytic kinase domain [48]. Recruitment to the cell membrane of Raf by Ras results in a conformational change [49], which disrupts the autoinhibitory interaction of Raf. Rafs then form homo– or heterodimers, which leads to partial activation by allostery. Transphorylative events, with optional phosphorylation by other kinases, such as PAK1 [50], then fully activates Raf. Activated Raf can now bind to MEKS, which are tyrosine/threonine kinases. MEKs phosphorylate ERKSs, which are also serine/threonine kinase enzymes. ERKs then translocate to the cell nucleus, where they influence expression of target genes [50]. RAS–RAF–MEK–ERK signaling promotes cell-cycle progression, cell differentiation, growth and survival [50].

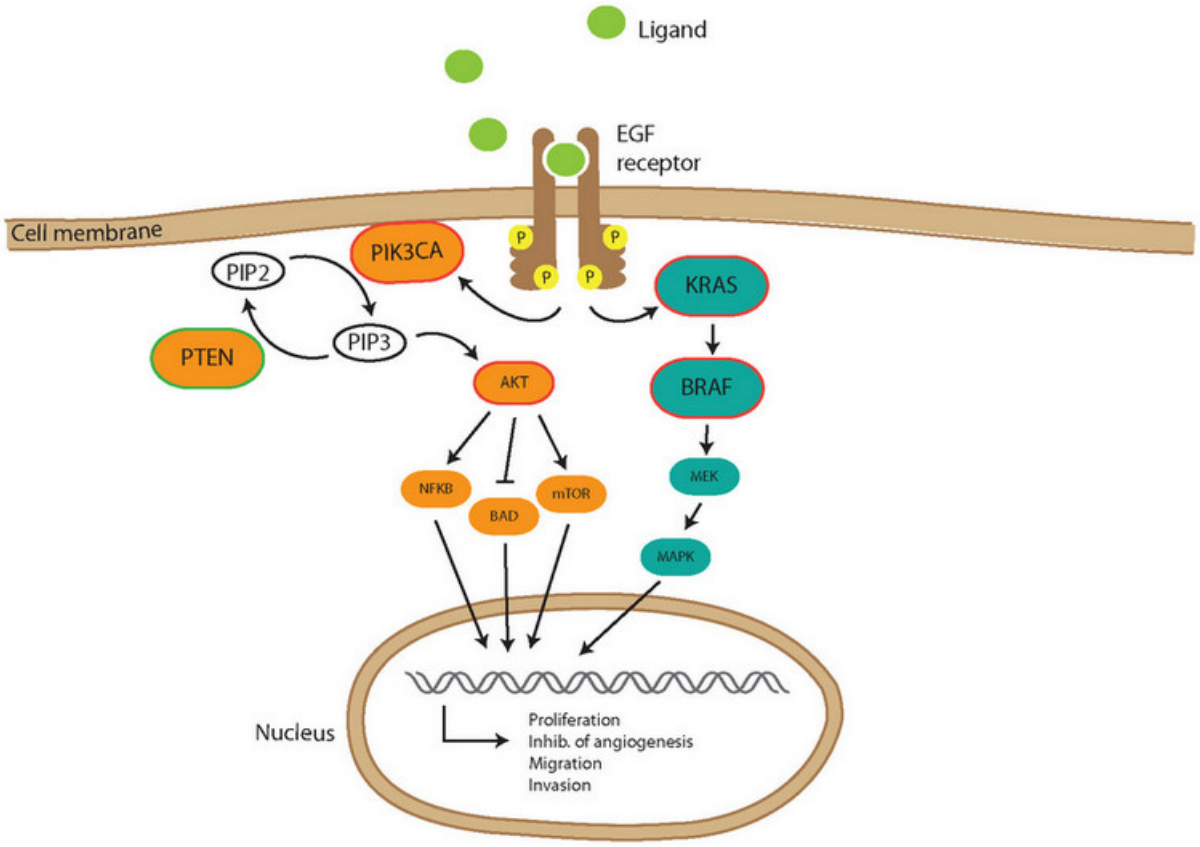

Figure 1: Schematic representation of the EGFR signaling cascade

1.1.2 Molecular Profiling of Solid Tumors

Classical anti-cancer treatments are tailored for the average patient“ and not for the individual. Traditional cytotoxic chemotherapeutic drugs, for instance, are unspecific and have numerous adverse effects: they attack rapidly dividing cells and make no difference between healthy and cancer cells. For a long time, there was no possibility to predict the success of a patient's cancer treatment. In consequence, the clinician had no way to personalize the treatment to the individual patient. Molecular profiling of tumors by several methods has lead to a better understanding of cancer development and progression and to the identification of some recursively found driver mutations, which may be potential anti-cancer targets.

Lung cancer, melanoma and colorectal cancer are amongst the most common cancers worldwide. Classically, diagnosis has been made by observing histologic, anatomic and pathologic alterations. With the success of molecular biology in the last decades, several models describing the molecular progression in the respective cancer types were postulated. This enabled experts to identify potential targets for personalized medicine.

Lung cancer is the most common cancer worldwide, both in terms of new cases (1.8 million) and deaths (1.6 million) (cancer.org). Smoking is a widely accepted risk factor, as chemical carcinogens in tobacco smoke induce several genetic mutations [8]. Lung cancer can be divided into two histological subtypes: small-cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Over the last decade, it has become clear that these subtypes can be classified into additional classes by the mutational status of recurrent driver mutations.

A combination of oncogenic triggers cause cells of the normal bronchial epithelium to proliferate, giving rise to meta–, hyper– and dysplastic epithelial lesions. Genomic events in early stages of lung cancer giving rise to atypical adenomatous hperplasia include LOH on chr.3p, $p16^{INK4a}$ or RB inactivation, as well as mutations in KRAS or in β–catenin. TP53 inactivation and LOH on chr.13q are believed to favor progression into the primary adenocarcinoma stage. After that stage, major chromosomic instability is often detected, giving rise to metastatic adenocarcinoma. These chromosomic events include LOH on chr.2q, chr.9p, chr.18q, and chr.22q. Additionally, the oncogene c-myc can be amplified in late stages. [8]

Frequent mutations in NSCLC affect EGFR (10–35%), KRAS (15–25%), PTEN (4–8%), HER2 (2–4%), DDR2 (4%), PIK3CA (1–3%), BRAF (1–3%), AKT1 (1%), MEK (1%) and NRAS (1%). Additionally, rearrangement of ALK (3–7%), RET (1%) and ROS1 (1%) and amplifications of FGFR1 (20%) and MET (2–4%) are found recursively. These mutations are rarely observed together in the same tumor. (mycancergenome.com)

Melanoma develops from the malignant transformation of melanocytes in the basal epidermal layer of the skin. Exposure to UV light, immunosuppression, fair-skin and multiple nevi are risk factors. UV radiation causes cyclobutane pyrimidine dimers (CPDs) [51]. T–T, C–C or C–T dimers (UV fingerprints) are formed, leading to direct DNA damage. People diagnosed with rare genetic disorders like xeroderma pigmentosum are at great risk [52]. Traditionally, melanoma has been classified based on histological and pathological properties, such as the thickness of the tumor, ulceration or the anatomic location of the tumor.

Spontaneous mutations in BRAF and NRAS and epigenetic modulations of APC are believed to promote progression of normal epithelium into a dysplastic nevus. Mutations affecting genes PTEN and CDKN2A and altered gene expression of MGMT or RASSF1 then favor invasion of the dermis, which is underlying to the epidermis, during the radial growth phase. Genetic and epigenetic alterations in AKT, MTAP, APAF, CDH1 and CDH2 then result in the vertical growth phase, where the tumor invades surrounding tissues. Finally, loss of TP53 functionality and further epigenetic modulations of MTA2 and MAGE enable metastasis. [53] [54]

The occurrence of the different mutations differs by the anatomic location of the tumor, e.g. whether the specific body part is chronically exposed to the sun. Mutations frequently found in melanoma occur on BRAF (37–50%), NRAS (13–25%), MEK (6–7%), NF1 (11.9%), CTNNB1 (2-4%), GNAQ (1.3%) and GNA11 (1.2%). (mycancergenome.com)

Colorectal cancer (CRC) is one of the best studied cancers. The development of colorectal adenocarcinomas occurs over many years. Caused by the acquisition and accumulation of driver mutations, a normal colorectal epithelium can progress to adenoma, which develops into carcinoma, which can eventually mestastasize.

The molecular progression models in CRC depend on the underlying instability process (chromosomal vs. microsatellite). In CIN CRCs, loss of the tumor suppressor gene APC is often causing the evolution from a normal to a hyperproliferative epithelium. Progression to adenoma stages are associated with DNA hypomethylation, KRAS activation and loss of 18q. Mutations in TGFβRII and PIK3CA and loss of TP53 by LOH on chr.17p then lead to the final carcinoma stage. In MSI CRCs, mutations and hypermethylations in MMR genes result in an hyperproliferative epithelium. BRAF mutations, followed by PIK3CA mutations, loss of TP53 and frameshift mutations affecting TGFβRII, BAX or IGF2R are associated with CRC progression towards the carcinoma stage. [14] [55] [56]

Mutations recursively detected in CRC occur on KRAS (36–40%), SMAD4 (10–35%), PIK3CA (10–30%), BRAF (8–15%), PTEN (5–14%), NRAS (1–6%), and AKT1 (1–6%). (mycancergenome.com)

The mentioned molecular progression profiles are likely to be an oversimplification. These models are based on frequently found alterations in the respective cancers. Due to tumor heterogeneity, these alterations do not always have to be observed in the tumor, and the chronological appearance of these alterations may vary from one tumor to another.

Despite enormous advantages in the understanding of the underlying causes of cancer in the last decades, only a few proteins involved in cancer development and progression are clinically actionable, e.g. can be targeted with specific drugs.

1.2 Targeting the EGFR signaling pathway

1.2.1 Biological Role of EGFR in Solid Tumors

The EGFR pathway is long known to be dysregulated in most solid tumors and thereby presents a rational target for cancer therapy. In normal cells, the tighly regulated EGFR signaling pathway drives

cell-cycle progression, affects differentiation and migration and acts as a survival signal. EGFR ligand binding leads to an activation of several signaling cascades, such as the previously discussed PI3K–AKT and RAS–RAF–MEK–ERK pathways. EGFR levels have been shown to be higher in tumor samples than in surrounding tissues. Also, more of EGFR's ligands EGF and TGFα are found in these locations. EGFR overexpression may result from epigenetic alterations or gene amplification. Increased EGFR mRNA levels in in vitro culture of human CRC cells have also linked EGFR overexpression to tumor progression. EGFR overexpression has been associated with poor prognosis if treated with classical chemotherapy, as this leads to a more aggressive progression.

EGFR signaling as a survival signal Cells communicate with their environment. Without extracellular signals, a cell undergoes apoptosis. Signaling pathways induced by cell–matrix interactions, cell-cell interactions, and soluble survival factors act on a variety of genes and proteins. Loss of matrix attachment leads to cell growth arrest and even cell death in normal epithelial cells, a process called anoikis. The cell–matrix interaction provides important spatial informations to the cell and acts as a safeguard against inappropriate proliferation and migration.

Activation of the EGFR signaling pathway protects normal epithelial cells against anoikis in the suspended state. EGFR blockade sensitizes normal epithelial cells to apoptosis, but the effect is much more pronounced in the suspended state than in the attached state. The redundancy of cell survival signals makes normal epithelial cells relatively resistant to EGFR-blockade in their normal microenvironment.

Tumor cells are often in transit or at sites with inadequate matrix composition. They are often provided with inadequate or missing cell–cell and cell–matrix interactions. They are thereby more dependent on survival signals propagated by soluble mediators, such as EGF or TGFα. Consequently, tumor cells are relatively sensitive against blockade of EGFR or other cell surface receptors. This is counterbalanced by an upregulation of cell surface receptors that activate anti–apoptotic pathways. MAPK activation in the EGFR–activated RAS–RAF–MAPK–ERK pathway is required for high expression of Bcl–XL, an anti-apoptotic protein of the Bcl–2 family. Bcl–2 proteins can be either pro– or anti–apoptotic. They regulate liberation of cytochrome c from the mitochondria, which is essential in the apoptotic caspase pathway. Additionally, EGFR signaling leads to post-transcriptional phosphorylation on the pro-apoptotic Bad protein, which is thereby functionally inactivated.

1.2.2 EGFR-targeted drugs

The observations that EGFR is recursively upregulated in many cancers and that EGFR is such an important mediator of cell-cycle, progression, cell growth, and cell survival, has lead to the development of agents that block this pathway. Pharmacologically, these agents can be classed by their mode of action: EGFR-targeted monoclonal antibodies and EGFR-specific tyrosine kinase inhibitors (TKIs). Several proteins acting downstream of EGFR are often found to be mutated. This lead to the development of other targeted drugs, such as BRAF– or MEK–inhibitors. Table 1 shows a selection of FDA-approved cancer drugs that target components of the EGFR signaling pathway.

Table 1: FDA-approved cancer drugs for solid tumor treatment that target the EGFR pathway

Agent	Target(s)	FDA-approved indication(s)
Afatinib (Gilotrif)	EGFR	NSCLC (with EGFR del19 or L858R)
Cetuximab (Erbitux)	EGFR	Colorectal cancer (KRAS WT)
Cobimetinib (Cotellic)	MEK	Melanoma (with BRAF V600E or V600K
Dabrafenib (Tafinlar)	BRAF	Melanoma (with BRAF V600 mutation)
Erlotinib (Tarceva)	EGFR	NSCLC
Gefitinib (Iressa)	EGFR	NSCLC (with EGFR del19 or L858R)
Necitumumab (Portrazza)	EGFR	Squamous NSCLC
Osimertinib (Tagrisso)	EGFR	NSCLC (with EGFR T790M)
Panitumumab (Vectibix)	EGFR	Colorectal cancer (KRAS WT)
Trametinib (Mekinist)	MEK	Melanoma (with BRAF V600)
Vemurafenib (Zelboraf)	BRAF	Melanoma (with BRAF V600)

Anti–EGFR monoclonal antibodies bind to the extracellular domain of EGFR in its inactive state and thereby compete for EGF or TGFα binding. They thereby inhibit ligand-induced EGFR tyrosine kinase activation. The most popular anti–EGFR monoclonal antibody is cetuximab. Cetuximab binds to EGFR with a higher affinity than the natural ligands EGF or TGFα. Cetuximab binding induces internalization of EGFR, following by its degradation. Cetuximab also binds to EGFRvIII, a constantly active version of EGFR. Cetuximab blocks cell–cycle progression at the G0/G1 boundary, inhibits cell proliferation and induces cancer cell death.

EGFR-specific tyrosine kinase inhibitors comprise three classes that include first generation reversible EGFR-inhibitors (gefitinib, erlotinib), second generation irreversible inhibitors (afatinib, dacomitinib, neratinib) and third generation mutant-selective inhibitors (brigatinib, osimertinib, rocile-

tinib). Third generation agents have a better sensitivity against mutated than wild-type EGFR and have been designed to further decrease treatment-associated side effects. TKIs are low molecular weight molecules that are mainly derived from quinazoline. These compounds bind to EGFR and block ligand-induced receptor phosphorylation by occluding the ATP–binding site. This results in inhibition of cell proliferation, cell–cycle arrest at the G0/G1 boundary and apoptosis.

1.2.3 Predictive markers

Mutations in EGFR and downstream proteins are predictive of the potential success of EGFR-targeted therapy. In many cases the predictive value of a marker, even though theoretically reasonable, has not yet been established. Demonstration of the predictive value of these markers is not trivial and has to be proved in clinical trials. The search for additional predictive markers is ongoing and many EGFR-targeted agents are still in clinical trials.

EGFR is a strong predictive biomarker for the success of the administration of EGFR-specific tyrosine kinase inhibitors. EGFR activating mutations are observed in 10–35% of NSCLC. 90% of EGFR mutations are exon 19 deletions (48%) and exon 21 L585R (c.2573T$>$G) (43%) point mutations. In melanoma and CRC, EGFR mutations are seldom. These mutations confer increased sensitivity to EGFR–specific tyrosine kinase inhibitors. Patients with EGFR–mutated tumors have a longer progression–free survival than those treated with traditional chemotherapy. Also, patients with EGFR–mutated tumors display a better prognosis if treated with EGFR TKIs compared to patients with wild–type EGFR cancers. An interesting demonstration of tumor placticity is the EGFR T790M variant, which occurs in 5% of untreated NSCLCs. Its frequency drastically increases to 50% in NSCLC tumors that have acquired resistance to TKIs. However, a study has shown that this resistance mutation may be lost after an arrest of TKI administration and that patients can be treated again with TKIs after a certain period. EGFR T790M increases sensitivity to third generation TKIs (mutant–specific). The mutation status of EGFR has no predictive value for monoclonal antibody administration. EGFR–specific antibodies have been associated with an increased benefit from cetuximab administration.

KRAS mutations are found in 36–40% of CRCs, 15–25% of NSCLC, and in 2% of melanomas. Critical mutations in the KRAS gene include mutations in codons 12, 13 and 61. Amongst these, the G12C variant is the most common. These mutations lock KRAS in its GTP-bound state, resulting in a constantly active protein. This then leads to a constantly active signal transduction. Blocking

EGFR in that case is useless, as KRAS acts downstream of EGFR. Several KRAS point mutations in codons 12, 13 and 61 have been shown to confer reduced sensitivity to EGFR-targeted monoclonal antibodies in CRC and EGFR–TKIs in NSCLC.

NRAS is an isoform of KRAS. Activating mutations in NRAS codons 12 and 61 are found in 1–6% of CRCs, 13–25% of melanomas and 1% of NSCLCs. NRAS mutations have been associated with reduced sensitivity to EGFR monoclonal antibodies in CRC. The predictive value of the influence of NRAS mutations in NSCLC and melanoma is unknown at this time.

BRAF mutations are very common in melanoma (37–50%) are found in 8–15% of CRCs and 1–4% in NSCLCs. Amongst BRAF-mutated melanomas, the V600E variant is found in 80–90% cases. The V600E variant occurs in the activation segment of the BRAF kinase domain and results in increased kinase activity. BRAF mutations usually confer a resistance to EGFR-targeted therapy in KRAS WT tumors. BRAF V600E mutations have been associated with increased sensitivity to BRAF inhibitors in melanoma and NSCLS and MEK inhibitors in melanoma.

1.3 Tumor DNA Sequencing

The completion of the Human Genome Project in 2001 resulted in a massive boost in molecular medicine. New high-throughput techniques, in combination with advanced computational performance and storage capacities, lead to an explosion of biological data. Amongst the many mutation detection techniques, Next–Generation Sequencing (NGS) constitutes the most powerful method and allows deep insights into the underlying causes of diseases. Today, NGS is used in several disciplines, which include basic molecular biology and pharmacogenomic research, forensics and molecular diagnostics. Even though advances in sequencing technology and computational power and tools have decreased the time and cost of a sequencing experiment, NGS is still mainly used in research, with only a few laboratories using this technique in diagnostics.

NGS has profoundly impacted the field of oncology. A wide variety of NGS applications have been applied to study the genetics and epigenetics of cancer. ChIP–Seq and FAIRE–Seq allow determination of DNA–protein interactions and identification of DNA regulatory elements, respectively. The cancer transcriptome can be studied with RNA–Seq experiments. NGS has accelerated discovery of genetic and epigenetic alterations in tumors. Also, the time and cost of an NGS experiment are rapidly decreasing.

Targeted NGS is the method of choice in molecular pathology laboratories. This method allows deeper insights into oncogenes and tumor suppressor genes than array–based and single–gene approaches, which are currently used in most laboratories. Multiple genes can be studied in a single experiment, while classical methods are much more restricted in that regard. Targeted NGS experiments differs from whole–genome or whole–exome sequencing, as they capture and sequence only a selection of regions of interest (ROIs). This approach increases the efficiency of the experiment and allows to sequence more samples in the same period of time. Additionally, coverage, e.g. the number of sequencing reads that align to a specific base of the reference genome, is drastically increased in targeted NGS. This offers the sensitivity to detect low–frequency mutations in the tumor sample. Targeted NGS methods can be applied to study insertions, deletions and point mutations in genes of interest. NGS can thereby guide personalized cancer therapy by identifying the mutational status of genetic markers, which are predictive of the potential success of targeted cancer treatment.

Several factors are responsible for the fact that NGS is not yet widely used in clinical molecular diagnostics laboratores. The implementation of NGS requires a substantial initial financial investment. Several NGS bench-top devices have become available in the last decade. These instrumentations differ in their underlying chemistry that influences the instrument's performance, accuracy, output and time per run. Common sequencing principles include pyrosequencing (454), sequencing by ligation (SOLiD), ion semiconductor sequencing (Ion Torrent) and sequencing by synthesis (Illumina). Trained experts are required to set up and maintain a workflow in the laboratory and a bioinformatic data analysis pipeline. Also, the implementation of new techniques into the workflow of molecular diagnostics laboratories requires a carefull assessment of the sensitivity and sensibility of the method. Therefore, each step that is performed from the initial starting material to sample processing, sequencing library preparation, sequencing assay and bioinformatic processing has to be checked for sources of potential errors or variability.

1.3.1 Practical implications in the laboratory

The quality of the genetic testing of the tumor is affected by several factors. These include the content of tumor cells in the sample, the quality of the tissue material, sequencing library preparation and the the bioinformatic pipeline.

The biopsy usually consists of an admixture of normal and cancer cells. The sensitivity of tumor variant detection is linked to the tumor cell content of the specimen. In addition, cancers are highly heterogenous, e.g. a small subpopulation might present mutations that provide resistance

to targeted treatment. Detecting these low-frequency mutations and clearly delineating them from possible sample processing or sequencing induced artifacts presents an important challenge.

Tumor biopsies usually yield a limited amount of tissue, therefore it is important to optimize sample usage by multiplexing analysis. In Luxembourg, all relevant tumor biopsies are usually sent to the Laboratoire National de Santé (LNS) to the Service of Pathologic Anatomy where the biopsy is fixed in formalin and embedded in paraffin (FFPE). FFPE preserves the tissue morphology and thereby allows histological analysis. In addition, it allows specimen storage for decades. Sample quality, however, is influenced by this fixation method, but also by the size of the biopsy, and its fixation time. DNA extraction from FFPE samples is difficult and yields low amounts of DNA; formaldehyde leads to cross-linking of nucleic acids and proteins; FFPE introduces fixation artifacts into DNA sequences, for instance C¿T transitions. These circumstances complicate sample processing as well as NGS data interpretation. Though, FFPE samples have been shown to be still suitable for downstream analyses.

Sequencing library preparation also affects the final NGS result. Several technologies for target enrichment exist and are available for different sequencing instruments. Essential for all these enrichment methods are the enrichement of target regions and sample multiplexing, which requires the incorporation of a unique index adaptor combination for each sample. Target enrichment methods can be separated into three basic groups: targeted circularization, hybrid capture of target fragments and PCR-based enrichment methods. In contrast to uniplex long-range PCR, short-range multiplex PCR produces short DNA fragments of target regions. There is thereby no need of DNA shearing. Hybridization-based methods require a so-called shotgun library construction before target regions can be captured. During this process, genomic DNA is sheared randomly into small fragments and an adapter- and index-linked library is produced. Biotinylated baits are added that bind to target regions. Target regions can then be captured using streptadivin coated magnetic beads. Targeted circularization methods rely on a digestion of DNA by restriction enzymes. The produced DNA fragments are then circularized and uncircularized DNA fragments are removed by exonucleases. Only circularized target regions are then amplified by PCR.

The establishment and validation of a bioinformatic NGS data analysis pipeline still constitutes a challenge in diagnostics. After generation of FASTQ files of the sequencer, data generally undergo quality control, followed by trimming of low quality bases, alignment to the reference genome, variant calling and variant annotation. For each of these steps, several bioinformatic algorithms and tools exist. The computational pipeline of the molecular pathology laboratory has to incorporate the tools that allow the most sensitive and sensible analysis of data. For instance, quality trimming influences

the mapping to the reference genome. The mapping, in turn, strongly affects the variant calling. In fact, variant calling is a critical step in NGS data analysis. Several tool kits as SAMtools, SPLINTER, VarScan2 or GATK allow variant annotation, but vary in their false-positive and false-negative detection rates. These tools have to be carefully assessed, as false-positives or false-negatives should absolutely be avoided when it comes to the subscription of a targeted chemotherapeutic agent.

To facilitate interpretation of NGS data, variants have to be annotated and their clinical actionability has to be identified. Several databases have emerged in this field (such as mycancergenome.org) and numerous tools allow to automatize variant annotation. Here again, the choice of the database and the variant annotator is important.

Finally, the sample-to-results time is a very pragmatic, but important factor. The time from the biopsy to the potential start of an administration of a targeted chemotherapeutic drug should be reduced to a minimum. For instance, in case of late-stage cancer patients, it would be unacceptable if analysis would take several weeks. To reduce the sample-to-results time to under two weeks, the sample processing workflow should be as short as possible, while still yielding high quality sequencing libraries. The bioinformatic pipeline should not only incorporate the best tools, but should also be automatized to further reduce the time of analysis.

1.4 Aims of the Thesis

References

[1] M. Berwick and P. Vineis, "Markers of dna repair and susceptibility to cancer in humans: an epidemiologic review," *JNCI J Natl Cancer Inst*, vol. 92, no. 2, pp. 874–897, 2000.

[2] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: The next generation," *Cell*, vol. 114, no. 5, pp. 646—-674, 2011.

[3] D. Hao, L. Wang, and L. Di, "Distinct mutation accumulation rates among tissues determine the variation in cancer risk," *Scientific Reports*, vol. 6, 2016.

[4] I. Tomlinson, P. Sasieni, and W. Bodmer, "How many mutations in a cancer?," *Am J Pathol.*, vol. 160, pp. 755–758, 2002.

[5] K. R. Loab and L. A. Loeb, "Significance of multiple mutations in cancer," *Carcinogenesis*, vol. 21, pp. 376–385, 2000.

[6] T. Kauppinena, J. Toikkanena, D. Pedersenb, R. Youngb, W. Ahrensc, P. Boffettad, J. Hansene, H. Kromhoutf, J. M. Blascog, D. Mirabellih, V. Orden-Riverag, B. Pannetti, N. Platoj, A. Savelaa, R. Vincentk, and l. M Kogevinasd, "Occupational exposure to carcinogens in the european union," *Occup Environ Med*, vol. 57, pp. 10–18, 2000.

[7] M. Kogevinas, M. Sala, P. Boffetta, N. Kazerouni, H. Kromhout, and S. Hoar-Zahm, "Cancer risk in the rubber industry: a review of the recent epidemiological evidence.," *Occup Environ Med*, vol. 55, pp. 1–12, 1998.

[8] J. Cornfield, W. Haenszel, E. C. Hammond, A. M. Lilienfeld, M. B. Shimkin, and E. L. Wynder, "Smoking and lung cancer: recent evidence and a discussion of some questions," *Int. J. Epidemiol.*, vol. 38, pp. 1175–1191, 2009.

[9] P. T. Schumacker, "Reactive oxygen species in cancer cells: Live by the sword, die by the sword," *Cancer Cell*, vol. 10, pp. 241–252, 2006.

[10] G. Waris and H. Ahsan, "Reactive oxygen species: role in the development of cancer and various chronic conditions," *Journal of Carcinogenesis*, vol. 5, 2006.

[11] F. Dietlein, L. Thelen, and H. C. Reinhardt, "Cancer-specific defects in dna repair pathways as targets for personalized therapeutic approaches," *Trends in Genetics*, vol. 30, no. 8, 2014.

[12] V. Bergoglio, M. Pillaire, M. Lacroix-Triki, B. Raynaud-Messina, Y. Canitrot, A. Bieth, M. Garès, M. Wright, G. Delsol, L. A. Loeb, C. Cazaux, and J. Hoffmann, "Deregulated dna polymerase β induces chromosome instability and tumorigenesis," *Cancer Res*, vol. 62, 2002.

[13] C. Boland and A. Goel, "Microsatellite instability in colorectal cancer," *Gastroenterology*, vol. 138, pp. 2073–2087, 2010.

[14] R. Kanthan, J. Senger, and S. C. Kanthan, "Molecular events in primary and metastatic colorectal carcinoma: A review," *Pathology Research International*, 2012.

[15] V. Deschoolmeester, M. Baay, P. Specenier, F. Lardon, and J. B. Vermorken, "A review of the most promising biomarkers in colorectal cancer: one step closer to targeted therapy," *Oncologist*, vol. 15, pp. 699–731, 2010.

[16] P. V. Jallepalli and C. Lengauer, "Chromosome segregation and cancer: cutting through the mystery," *Nature Reviews Cancer*, vol. 1, pp. 109–117, 2001.

[17] M. Giam and G. Rancati, "Aneuploidy and chromosomal instability in cancer: a jackpot to chaos," *Cell Div*, vol. 10, 2015.

[18] J. Forment, A. Kaidi, and S. Jackson, "Chromothripsis and cancer: causes and consequences of chromosome shattering.," *Nature Reviews Cancer*, vol. 12, pp. 663–670, 2012.

[19] A. Rode, K. Maass, K. Willmund, P. Lichter, and A. Ernst, "Chromothripsis in cancer cells: An update.," *Int J Cancer*, vol. 138, pp. 2322–2333, 2016.

[20] M. Greaves and C. C. Maley, "Clonal evolution in cancer," *Nature*, vol. 481, pp. 306–313, 2012.

[21] M. Gerlinger and C. Swanton, "How darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine," *British Journal of Cancer*, vol. 103, pp. 1139—1143, 2010.

[22] J. Breivik, "Don't stop for repairs in a war zone: Darwinian evolution unites genes and environment in cancer development," *PNAS*, vol. 98, no. 10, pp. 5379–5381, 2001.

[23] M. P. Little, "Cancer models, genomic instability and somatic cellular darwinian evolution," *PNAS*, vol. 5, no. 19, 2010.

[24] M. R. Stratton, P. J. Campbell, and P. A. Futreal, "The cancer genome," *Nature*, vol. 458, pp. 719–724, 2009.

[25] C. D. McFarland, K. S. Korolev, G. V. Kryukov, S. R. Sunyaev, and L. A. Mirny, "Impact of deleterious passenger mutations on cancer progression," *Proc Natl Acad Sci USA*, vol. 110, pp. 2910—-2915, 2013.

[26] I. Bozic, T. Antal, H. Ohtsuki, H. Carter, D. Kim, S. Chen, R. Karchin, K. W. Kinzler, B. Vogelstein, and M. A. Nowak, "Accumulation of driver and passenger mutations during tumor progression," *Proc Natl Acad Sci USA*, vol. 107, pp. 18545—18550, 2010.

[27] S. K. Merid, D. Goranskaya, and A. Alexeyenko, "Distinguishing between driver and passenger mutations in individual cancer genomes by network enrichment analysis," *BMC Bioinformatics*, vol. 15, 2014.

[28] M. Cusnir and L. Cavalcante, "Inter-tumor heterogeneity," *Human Vaccines and Immunotherapeutics*, vol. 8, 2012.

[29] F. Michor and K. Polyak, "The origins and implications of intratumor heterogeneity," *Cancer Prev Res*, vol. 3, pp. 1361–1364, 2010.

[30] W. Sun and J. Yang, "Functional mechanisms for human tumor suppressors," *J Cancer*, vol. 1, pp. 136–140, 2010.

[31] H. Chial, "Tumor suppressor (ts) genes and the two-hit hypothesis," *Nature Education*, vol. 1, 2010.

[32] M. Santarosa and A. Ashworth, "Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way," *Biochim Biophys Acta.*, vol. 1654, pp. 105–122, 2004.

[33] B. T. MacDonald, K. Tamai, and X. He, "Wnt/β-catenin signaling: components, mechanisms, and diseases," *Dev Cell*, vol. 17, no. 1, pp. 9–26, 2009.

[34] P. Polakis, "Wnt signaling and cancer," *Genes Dev*, vol. 14, pp. 1837–1851, 2000.

[35] M. Farnebo, V. J. Bykov, and K. G. Wiman, "The p53 tumor suppressor: a master regulator of diverse cellular processes and therapeutic target in cancer," *Biochem Biophys Res Commun.*, vol. 396, no. 1, pp. 85–89, 2010.

[36] Y. Yang, C. H. Li, and A. M. Weissman, "Regulating the p53 system through ubiquitination," *Oncogene*, vol. 23, pp. 2096–2106, 2004.

[37] J. T. Zilfou and S. W. Lowe, "Tumor suppressive functions of p53," *Cold Spring Harb Perspect Biol.*, vol. 1, no. 5, 2009.

[38] M. Olivier, M. Hollstein, and P. Hainaut, "Tp53 mutations in human cancers: Origins, consequences, and clinical use," *Cold Spring Harb Perspect Biol.*, vol. 2, no. 1, 2010.

[39] C. M. Croce, "Oncogenes and cancer," *N Engl J Med*, vol. 358, pp. 502–511, 2008.

[40] R. S. Herbst, “Review of epidermal growth factor receptor biology,” *Int J Radiat Oncol Biol Phys.*, vol. 59, pp. 21–26, 2004.

[41] K. M. Ferguson, “Structure-based view of epidermal growth factor receptor regulation,” *Annu Rev Biophys.*, vol. 37, pp. 353–373, 2008.

[42] J. LoPiccolo, G. M. Blumenthal, W. B. Bernstein, and P. A. Dennis, “Targeting the pi3k/akt/mtor pathway: effective combinations and clinical considerations,” *Drug Resist Updat*, vol. 11, pp. 32–50, 2008.

[43] B. A. Hemmings and D. F. Restuccia, “Pi3k-pkb/akt pathway,” *Cold Spring Harb Perspect Biol*, 2012.

[44] J. R. Hart and P. K. Vogt, “Phosphorylation of akt: a mutational analysis,” *Oncotarget*, vol. 2, no. 6, pp. 467—-476, 2011.

[45] I. Hers, E. E. Vincent, and J. M. Tavaré, “Akt signalling in health and disease,” *Cell Signal*, vol. 23, no. 10, pp. 1515–1527, 2011.

[46] J. F. Vara, A. Casado, J. de Castro, P. Cejas, C. Belda-Iniesta, and M. González-Barón, “Pi3k/akt signalling pathway and cancer.,” *Cancer Treat Rev.*, vol. 30, pp. 193–204, 2004.

[47] E. H. Lowenstein, R. J. Daly, A. G. Batzer, W. Li, B. Margolis, R. Lammers, A. Ullrich, E. Y. Skolnik, D. Bar-Sagi, and J. Schlessinger, “The sh2 and sh3 domain-containing protein grb2 links receptor tyrosine kinases to ras signaling,” *Cell*, vol. 70, no. 3, pp. 431–442, 1992.

[48] N. H. Tran, X. Wu, and J. A. Frost, “B-raf and raf-1 are regulated by distinct autoregulatory mechanisms,” *J Biol Chem.*, vol. 280, no. 6, pp. 16244–16253, 2005.

[49] K. Hibino, T. Shibata, T. Yanagida, and Y. Sako, “A rasgtp-induced conformational change in c-raf is essential for accurate molecular recognition,” *Biophys J.*, vol. 97, no. 5, p. 1277–1287, 2009.

[50] M. Zang, C. Hayne, and Z. Luo, “Interaction between active pak1 and raf-1 is necessary for phosphorylation and activation of raf-1,” *J Biol Chem.*, vol. 277, no. 6, pp. 4395–4405, 2002.

[51] S. Mouret, C. Baudouin, M. Charveron, A. Favier, J. Cadet, and T. Douki, “Cyclobutane pyrimidine dimers are predominant dna lesions in whole human skin exposed to uva radiation,” *PNAS*, vol. 103, no. 37, pp. 13765–13770, 2011.

[52] K. Paszkowska-Szczur, R. J. Scott, P. Serrano-Fernandez, A. Mirecka, P. Gapska, B. Górski, C. Cybulski, R. Maleszka, M. Sulikowski, L. Nagay, J. Lubinski, and T. D. T, "Xeroderma pigmentosum genes and melanoma risk," *Int J Cancer*, vol. 133, no. 5, pp. 1094–1100, 2013.

[53] R. A. Anvekar, J. J. Asciolla, D. J. Missert, and J. E. Chipuk, "Born to be alive: a role for the bcl-2 family in melanoma tumor cell survival, apoptosis, and treatment," *Frontiers in Oncology*, vol. 1, no. 34, 2011.

[54] B. Bandarchi, L. Ma, R. Navab, A. Seth, and G. Rasty, "From melanocyte to metastatic malignant melanoma," *Dermatol Res Pract.*, 2010.

[55] S. D. Markowitz and M. M. Bertagnolli, "Molecular basis of colorectal cancer," *The new england journal of medicine*, vol. 361, no. 55, pp. 2449–2460, 2009.

[56] R. J. Davies, R. Miller, and N. Coleman, "Colorectal cancer screening: prospects for molecular stool analysis," *Nature Reviews Cancer*, vol. 5, pp. 199–209, 2005.