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

List of Abbreviations

NGS	Next Generation Sequencing
LNS	Laboratoire National de Sante
SGMB	Service of Genetics and Molecular Biology TST15
Illumina	
TruSight	
Tumor	
15	

List of Figures

List of Tables

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 Targeting Solid Tumors

Melanoma develops from the malignant transformation of melanocytes in the basal epidermal layer of the skin. Melanoma incidence has exploded over the last four decades, with a 15-fold increase in the United States. Both genetic predisposition and environmental factors influence the risk of getting melanoma. Skin cancer often affects fair-skinned individuals. Exposure to UV light, immunosuppression and multiple nevi are risk factors. UV radiation causes cyclobutane pyrimidine dimers (CPDs). By joining adjacent pyrimidine bases, T–T, C–C or C–T dimers (called UV fingerprints) are formed, leading to direct DNA damage. People diagnosed with rare genetic disorders like xeroderma pigmentosum are at great risk. Several susceptibility genes also increase melanoma risk, such as the MC1R gene that caused red hair.

At stage 0, tumor cells are still limited to the epidermis. In the radial growth phase is the earliest stage of melanoma. The tumor has a thickness less than 1mm and the cancer cells have not yet reached blood vessels. The cancer cells then acquire invasive potential: the cancer enters the invasive radial growth phase. In the vertical growth phase, tumor cells enter the blood stream or lymph vessels. The tumor starts to grow into surrounding tissues and is has now a tickness more than 1mm. Survival rates and treatment options decrease drastically with each stage.

Stage 0: Melanoma involves the epidermis but has not reached the underlying dermis. Stages I and II: Melanoma is characterized by tumor thickness and ulceration status. No evidence of regional

lymph node or distant metastasis. Stage III: Melanoma is characterized by lymph node metastasis. No evidence of distant metastasis. Stage IV: Melanoma is characterized by the location of distant metastases and the level of lactate dehydrogenase.

Non-Small Cell Lung Carcinoma: Lung cancer is the most common cancer in developed countries. Smoking is a widely accepted risk factor, as chemical carcinogens in tobacco smoke induce several genetic mutations. Oncogenic triggers cause cells of the normal bronchial epithelium to proliferate, giving rise to meta–, hyper– and dysplastic epithelial lesions. blablabla

Colorectal Cancer

1.1.1 Genomic Instability in Cancers

Virtually all cancers tend to accumulate mutations during their progression. The genetic diversity caused by this instability, the cardinal feature of cancer, in combination with several environmental factors, such as inflammation, enables the hallmarks of cancer (Hallmarks of Cancer: The Next Generation). These include replicative immortality, cell death resistance, ongoing proliferative signaling, invasion and metastasis, growth suppressor evasion, inducement of angiogenesis, energy metabolism reprogramming and immune destruction evasion.

Large-scale studies have demonstrated that cancers are highly heterogeneous. This heterogeneity has been observed both at the inter- and intra-level. Two tumors of similar phenotype often comprise a different subset of mutations that may even have a low overlap. Several clonal subpopulations within the same tumor contribute to the intra-tumor heterogeneity. This is a considerable problem in the clinics, as some subpopulations of a cancer may become resistant to the treatment and may be the source of relapses.

Genomic instabilities: CIN, MSI, CpG.

Many alterations found in cancer cells are passenger mutations, e.g. do not contribute to the selective fitness of the cell. Driver mutations, often happening on oncogenes or tumor suppressor genes, promote the cell's fitness. This concept recognizes Darwinian evolution principles. The heterogeneous population of cancer cells harbors cells with different random somatic and non-deleterious mutations and exhibits different perturbations. Cells with the best fitness, e.g. with the highest proliferative potential and the lowest death rate, are then selected through natural selection principles. These cells will outlast less fit cells. This results in sequential waves of clonal expansion, leading to different subclones within the same tumor that differ in their proliferative, migrative and

invasive potential. The hypothesis that passenger mutations, that occur subsequently or coincidentally to driver mutations, do not influence the cell's fitness at all has been challenged by stochastic tumor evolution simulations (citat). Since then, it has been proposed that, even though the individual effect may be small, the cooperation of multiple accumulated small-scale passenger mutations plays a present role in cancer development and progression.

1.1.2 Tumor Suppressors and Oncogenes in Solid Tumors

Genomic instability in cancerous cells becomes a critical mechanism if it affects oncogenes or tumor suppressor genes. 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 or cell cycle proteins. The action of these proteins inhibits metastasis, excessive cell survival or proliferation. Tumor suppressors mostly follow the two-hit hypothesis, which was first proposed by Knudson for the retinoblastoma protein (pRb): to inactivate the tumor-protecting role of tumor suppressors, two genetic events, often LOH in combination with silencing point mutations, are necessary to inactivate both alleles of the gene. Another possibility of tumor suppressor inactivation is methylation of the gene promoter. Compared to dominant oncogenes, tumor suppressor genes are often considered to be recessive. Alternatively, tumor progression can be influenced by functional haploinsufficiency. 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. Oncogenes comprise several GTPases, transcription factors, receptor tyrosine kinases and growth factors. Overexpressed or overactive versions of these proteins often lead to increased mitogenic signals, causing increased cell growth or proliferation. Mutations in proto–oncogenes can cause a loss of regulation or overactive proteins. Gene duplications or other chromosomal alterations lead to increased protein synthesis. Other mechanisms of importance include post-transcriptional mechanisms as misregulation of protein expression or increase of mRNA / protein stability.

APC: Adenomatous Polyposis Coli (APC) gene codes for a 312 kDa protein. This multi–domain protein has binding sites for microtubules, cytoskeletal regulator proteins (IQGAP1, EB1) and Wnt signaling proteins (b -catenin, axin). 60% of APC mutations in cancer present a C-terminal truncation, resulting in a loss of b–catenin and microtubule binding sites. Wnt signaling regulates, amongst others, cell migration, polarity, differentiation, adhesion and apoptosis. In the canonical Wnt signaling pathway, a destruction complex, including axin, GSK3, CK1a, PP2A and APC, leads to b–catenin phosphorylation, followed by ubiquitination, marking it for degradation in the proteasome. Addition-

ally, transcription factors of the TCF/LEF family form a complex with factors such as Groucho and histone deacetylases. This complex binds to Wnt signaling target genes and thereby represses gene expression. Once Wnt binds to the N-terminus of a cell surface receptor of the Frizzled family of receptors and and a co-receptor of the LRP5/6 family, the destruction complex is inhibited. Consequently, b–catenin is no longer marked for degradation and can translocate to the nucleus. There it displaces the factors binding to TCF/LEF and forms a complex with TCF/LEF, leading to activation of gene expression of target genes. Loss or dysfunction of APC leads to b–catenin accumulation in the nucleus even in the absence of an extracellular stimulus.

APC mutations are suspected to be the initiating event in many CRCs. APC mutations are sufficient for the growth of benign colorectal tumors.

http://www.wormbook.org/chapters/www$_{w}ntsignaling/wntsignaling.htmlhttp://www.ncbi.nlm.nih.gov/pu$
$//jcs.biologists.org/content/120/19/3327.longhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC2634250/$

Also in melanoma: http://www.ncbi.nlm.nih.gov/pubmed/15133491

Present, but less in NSCLC: http://www.ncbi.nlm.nih.gov/pubmed/15072829

TP53: TP53 is one of the master guardians of the genome. In normal situations, p53, the protein encoded by TP53, is regulated by MDM2, MDM4 and E3–ubiquitin ligase, which target p53 for ubiquitination and degradation in the proteasome. In case of cellular stress, p53 is no longer ubiquitinated. p53 becomes activated in several situations, which include DNA damage, osmotic shock, oxidative stress or oncogene expression. In such situations, 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. TP53 thereby maintains genomic stability.

One mechanism by which p53 acts on cell-cycle arrest is by activating expression of p21. p21 binds to the G1/S transition complex, formed by CDK4/CDK6, CDK2, CDK1) and inhibits its activity, leading to cell-cycle arrest. Inactivation or mutation of TP53 is a crucial step in many cancers. A defective p53 does not bind efficiently to DNA, resulting in less p21 expression. As a consequence, p21 cannot act as a cell-cycle stop signal.

The importance of TP53 as tumor suppressor gene becomes evident in the autosomal dominant Li–Fraumeni syndrome. People suffering from this disorder inherit only one functional copy of TP53 and are likely to develop cancer in early ages.

TP53 mutations are found in 50% of CRCs, especially those associated with the methylator phenotype and microsatellite instability. Alterations in TP53 are associated with poor prognosis if

treated with chemotherapy. In fact, wild-type TP53 is required for treatment with chemotherapy based on 5-fluoroacil.

TP53 alterations are the early events in lung carcinogenesis

TGF–b: Melanoma: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3662904/ http://www.ncbi.nlm.nih.gov/pubmed/1 http://www.ncbi.nlm.nih.gov/pubmed/21619542

NSCLC: http://www.ncbi.nlm.nih.gov/pubmed/20107423 http://link.springer.com/chapter/10.1007/978-1-4419-6615-5$_2 8 page - 1 http : //www.nature.com/cdd/journal/v21/n8/full/cdd201438a.html$

Colorectal: http://www.ncbi.nlm.nih.gov/pubmed/20517689 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3512565/ http://hmg.oxfordjournals.org/content/16/R1/R14.full

EGFR signaling pathway : EGFR is a protein of the tyrosine kinase receptor family. 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. Ligand binding causes a conformational change if the receptor, which leads to homo– or heterodimerization, followed by an auto– and cross–phosphorylation of key tyrosine residues on its cytoplasmic domain. This forms docking sites for cytoplasmic proteins that contain phosphotyrosine-binding and Src homology 2 domains. These proteins are adaptor molecules for the RAS-RAF-MAPK and PI3K pathways. Both pathways lead to cell survival, proliferation and invasion.

PTEN/PI3K/AKT leads to cell growth, proliferation and survival http://www.nature.com/onc/journal/v27/n41/full/onc20 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3092286/ http://www.hindawi.com/journals/isrn/2013/472432/

Phospholipase C

STATs

Src

In the RAS-RAF-MAPK pathway, GRB2 binds to Tyr1068 of EGFR through their SH2 domain and recruits SOS, a guanine nucleotide exchange factor. Grb2 and SOS then form a complex with the activated EGFR, which activates SOS. Activates SOS, through its GEF activity, then induces GDP removal from Ras proteins, which can subsequently bind GTP and become active. Ras then activates Raf serine/threonine kinase proteins, which phosphorylate and thereby activate MEKs, which are tyrosine/threonine kinases. Activated MEKS then phosphorylate and activate MAPKs, also serine/threonine kinases. MAPKs then act on the expression of target genes that promote cell

survival, cell cycle progression and proliferation.

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3457779/

The RAS-RAF-MAPK pathway is deregulated in many cancers, mainly through activating mutatios on RAS or RAF.

KRAS KRAS has gained interest as negative predictive marker of the successfulness of anti-EGFR targeted therapy. KRAS is mutated in 36–40% of CRCs, 15–25% of NSCLCs and 2% of melanomas. Single nucleotide point mutations in codons 12, 13 and 61 can act as activating mutations. Proteins affected by such mutations are locked in their active GTP-bound state and are consequently constantly active.

BRAF Apart from RAS proteins, RAF proteins are of importance in solid tumors.

https://www.moffitt.org/Filehttp://www.sciencedirect.com/science/article/pii/S0014579301021664 http://www.ncbi.nlm.nih.gov/pubmed/18038764 http://cancerres.aacrjournals.org/content/63/1/1.long

Many cancers have been shown to be dependent on EGFR signaling.Targeting the EGFR signaling pathway is an attractive target blabla and has been of benefit in solid tumors, e.g. melanoma, CRC and NSCLC. Advantages have been made better survival rates. shutting egfr down -¿ apoptosis. but 2 problems: resistances not all mutations are actionable. therefore: molecular testing. the more comprehensible, the better. classical approaches only target some hotspot regions. NGS has the potential to give really deep insights

In recent years, several EGFR targeted anti-cancer drugs have been approved by the FDA. Anti–EGFR targeted monoclonal antibodies and EGFR–specific tyrosine–kinase inhibitors have shown their usefulness in the treatment of solid tumors.

However, solid tumors have a tendency to harbor mutations in proteins acting in the EGFR signaling pathway. Table XXX shows the frequency of tumors harboring mutations in EGFR or downstream proteins. Identifying the mutational status of these proteins is of utmost importance in targeted anti-EGFR therapy. Wild-type or mutated proteins provide either increased sensitivity or resistance to the treatment.

Targeted cancer therapy holds the promise of highly selective tumor cell killing while sparing most of normal proliferating cells, thus avoiding some side effects of conventional cytotoxic therapy.

(http://www.nature.com/bjc/journal/v112/n2/full/bjc2014476a.html)

Table 1: EGFR signaling pathway components affected in colorectal cancer, melanoma and non-small cell lung carcinoma

Gene	CRC (%)	Melanoma	NSCLC
EGFR	NA	NA	10–35
KRAS	36–40	2	15–25
NRAS	1–6	13–25	1
BRAF	8–15	37–50	1–3
PTEN	5–14	NA	4–8
PIK3CA	10–30		1–3

Table 2: 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)

1.2 Targeted Sequencing

1.2.1 Target Enrichment Methods

1.2.2 Illumina Sequencing Chemistry

1.3 NGS Data Analysis

1.3.1 GATK Best Practices

1.4 Practical Implications in the Laboratory

1.5 Aims of the Thesis

2 Material and Methods

- How was the data analyzed ?
- Present econometric/statistical estimation method and give reasons why it is suitable to answer the given problem.
- Allows the reader to judge the validity of the study and its findings.
- Depending on the topic this section can also be split up into separate sections.

2.1 Library Preparation

2.1.1 Patients

Melanoma Non-Small Cell Lun Carcinoma (NSCLC) metastatic colorectal cancer (mCRC) Chronic lymphocytic leukemia (CLL) were extracted from blood and did not undergo FFPE treatment -¿ were used as some kind of good quality samples to see if there are really more C¿T variants in FFPE samples.

2.1.2 DNA Extraction, Quantification and Quality Control

DNA Extraction Kit from Qiagen

Quantification Qubit fluorometer, either High Sensitivity kit or Broad Range

Quality Control Illumina Infinium FFPE QC Assay kit

2.1.3 Agilent Haloplex ClearSeq Cancer

2.1.4 Illumina TruSight Tumor 15

2.2 Bioinformatic Analysis

2.2.1 Agilent SureCall

With alignment algorithms installed Windows System, 3.00GHz, 16GB RAM$ modifiable

2.2.2 Illumina BaseSpace TruSight Tumor 15 App

Cloud-based parameters not modifiable

2.2.3 Custom In-House Pipeline (Velona)

Linux System

2.2.4 Variant Calling Algorithms

Tested on Linux Ubuntu 14.04.4 LTS Ťrusty Tahrínstalled on VMware virtual box with of 12GB RAM

GATK HaplotypeCaller

VarScan 2

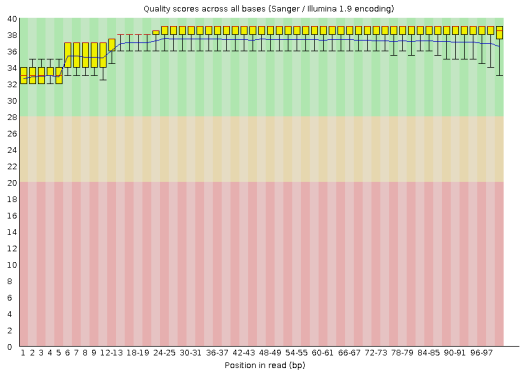
Mutect1.1.7 [2]

SomVarIUS

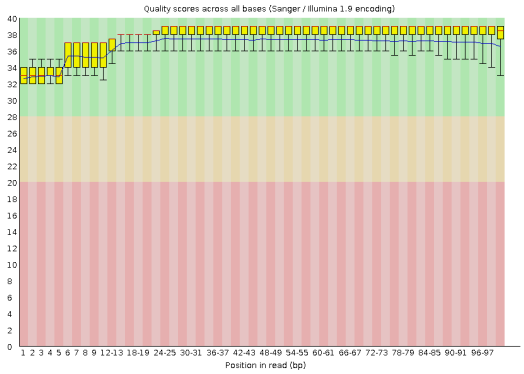
3 Results

3.1 Sample Preparation

Before pooling the adaptor-ligated and indexed sequencing libraries, the sucess of library preparation is validated using the Agilent Bioanalyzer instrument. 1a and 1b show representative electropherograms of a sample that has been processed using both kits. The expected DNA products should be detected at 175-600 bp for Haloplex CSC and 200-400 for TST15.



(a) Agilent Haloplex CSC (High Sensitivity DNA Chip)



(b) Illumina TST15 (MixA) (DNA 1000 DNA Chip)

Figure 1: Electropherograms of representative sequencing libraries prepared by Agilent Haloplex ClearSeq Cancer and Illumina TruSight Tumor 15. (*) represents the lower marker, (**) represents the upper marker

Using the blablabla software, the concentration, molarity and total peak area (TPA) of the expected sequencing libraries were calculated.

Maybe: relationship between dCt and TPA?

Maybe: the enrichment of one or the other kit is more affected by bad quality samples

3.2 NGS Data Quality

Sequencing run parameters were calculated by the Illumina Sequencing Viewer software. Table XXX shows the averaged run parameters of runs with Haloplex CSC and TST15 sample preparation.

TST15 has a higher cluster density and therefore a higher total yield, but has lower reads passing a phred-score threshold of Q30 than Haloplex. This is due to the different chemistries used. TST15 uses v3 chemistry, while Haloplex uses v2. v2 generally has lower cluster density and output, but

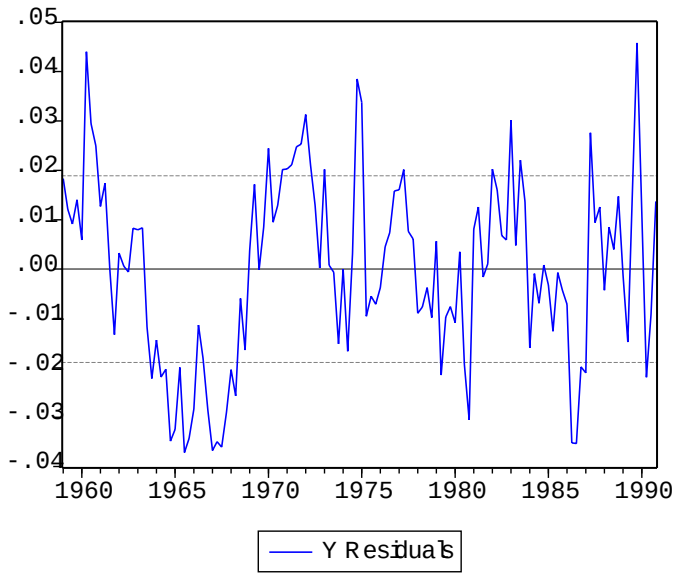
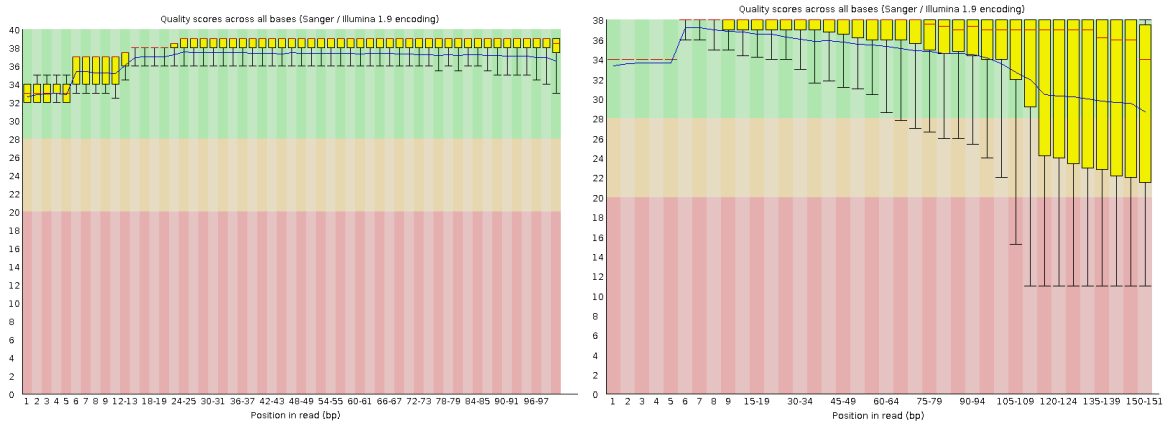


Figure 2: Scatter plot of the corrected peak area (X axis) of the regions corresponding to the sequencing libraries defined in the blabla software and the dCt (Y axis). Agilent Haloplex ClearSeq Cancer data are represented as blue dots, Illumina TruSight Tumor 15 data are represented as red dots.

Table 3: Comparison of Run Parameters (Averaged) of Sequencing Runs with Haloplex CSC & TST15 Sample Preparation

Parameter	Halo CSC	TST15
Yield total (Gb)	3.7	7.37
% >Q30	93.8	82.355
Cluster Density PF (k/mm2)	1084	1180
Cluster Density PF (%)	85.95	79.95



(a) Agilent Haloplex ClearSeq Cancer

(b) Illumina TruSight Tumor 15

Figure 3: Comparison of coverage distributions per amplicon as reported by FastQC

therefore better quality.

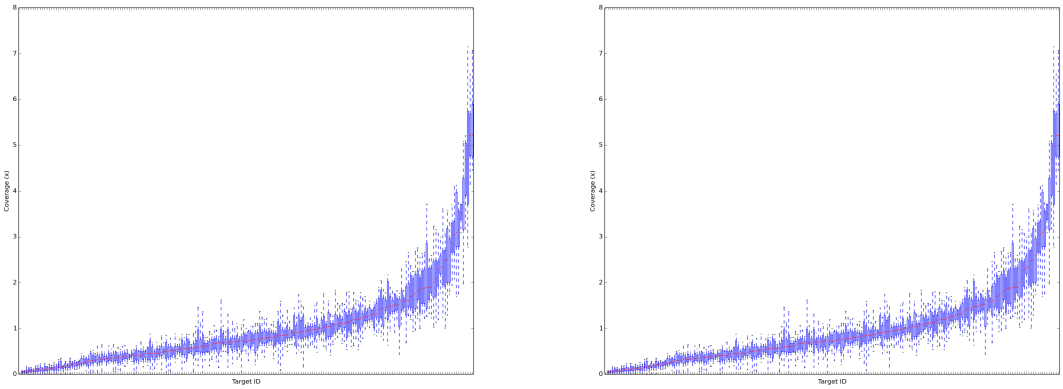
Table XXX shows the boxplot representations of the read qualities per position of two representative FASTQ files as reported by FastQC. Both workflows yield high quality data, yet Haloplex CSC data have more narrow distributions and are of higher quality. This is in direct relationship with the sequencing chemistry kit used.

Table 4: Blablabla

Parameter	Haloplex SureCall A	Haloplex SureCall B	Haloplex Velona	TST15 BaseSpace	TST15 Velona
% mapped	–	–	–	62.6	–
% paired	–	–	–	58.7	–
% singletons	–	–	–	3.8	–

The Samtools Flagstat command was used to determine some basic BAM statistics of BAM files of samples prepared with the respective library preparations and processed with the mentioned bioinformatic pipelines. **??** shows the averaged result of these statistics.

Considering the recommended pipelines, Haloplex CSC data, analyzed with Agilent's SureCall software, has a higher percentage (91%) of mapped reads when compared to Illumina's BaseSpace TruSight Tumor 15 App (62.9%). Data analysis with the recommended SureCall design includes a steps where mates are fixed, but they are not stitched together. Therefore no reads are considered as being paired. The TST15 app in contrast includes a read stitching step and 58% are considered as properly paired. This means that of the 62.6% of mapped reads, 4.2% are not properly paired.



(a) Agilent Haloplex ClearSeq Cancer

(b) Illumina TruSight Tumor 15

Figure 4: Comparison of Coverage Distributions per Amplicon

3.8% of reads processed with the TST15 online App are considered to be singletons, whereas only 1.8% of reads processed with the SureCall software are considered as singletons.

3.3 Coverage Analysis

Coverage Distribution TST15 vs Haloplex

Coverage Distribution per Patient (check if correlation IQR with dCt)

Coverage Distribution per Amplicon (check if some have always lower coverage, check if some failed)

Failed Amplicon Counter

Table 5: Blablabla

Amplicon	1x	50x	250x	500x	1000x
ATM_14	12	0	0	0	0
MAP2K1_2	1	0	0	1	4
ATM_5	0	1	1	0	2
ATM_11	0	1	1	2	3
PTEN_1	0	1	1	0	2
KIT_6	0	1	0	1	2
FGFR3_1	0	0	1	2	2

Table 6: Blablabla

Amplicon	1x	50x	250x	500x	1000x
1METxxxE16TF031SR031	0	11	0	1	3
2KITxxxE09TF003SR003	0	5	6	0	0
2KITxxxE09TF003SR003	0	5	6	0	0
27qxxxxExxTF034SR034	0	0	2	4	5
17qxxxxExxTF018SR018	0	1	1	4	4
1KRASxxE04TF002SR002	0	1	1	3	6
1TP53xxE02TF034SR034	0	0	3	5	9
1EGFRxxE19TF032SR032	0	0	1	5	5
1EGFRxxE21TF035SR035	0	0	1	2	2
2TP53xxE02TF033SR033	0	0	0	1	3

Table XXX and table XXX show how often a given amplicon failed a given coverage threshold. The number of failed amplicons is low for Haloplex CSC as wel las for TST15. Most amplicons were amplified efficiently in most samples, which is also confirmed by figure XXX (amplicon distributions). Some amplicons however fail coverage thresholds in several samples.

- Amplicon ATM_14 in Haloplex CSC was never amplified. This amplicon is defined in the BED file but obviously is not part of the kit. (negative control?)
- Amplicons ATM_5, ATM_10, MAP2K1, PTEN_1, KIT_6 and FGFR3_1 in Haloplex CSC data failed a coverage threshold of 1000x in several samples
- In TST15 data, more amplicons fail the respective thresholds. Several amplicons in genes MET, KIT, TP53, KRAS and EGFR fail the 1000x coverage threshold, and often even the required threshold of 500x, which is required by the TruSight Tumor 15 App.

The fact that several amplicons in the genes EGFR and KRAS in TST15 data repetitively fail the required coverage thresholds is problematic. This is especially the case for amplicon 1EGFRxxE21TF035SR035 as it includes the well-known EGFR L858R variant, which confers increases sensitivity for EGFR tyrosine kinase inhibitors.

Fragmentation ¡-¿ Coverage?

(GATK CallableLoci) (GATK CountLoci???) (GATK FindCoveredIntervals)

On-off target; Enrichment Efficiency TST15 vs Haloplex

Coverage across genome, check where there is coverage

Strandedness?

GATK DepthOfCoverage???

3.4 Variant Calling Algorithm Comparison

3.4.1 Detection of Known Single Nucleotide Variants and Deletions

Table XXX shows known variants in the analyzed samples and the variant frequency reported by the recommended pipelines. There is a high concordance between the results of both kits and previously known variants.

TST15 could not detect EGFR del19 in patient F due to low coverage. The corresponding region was inspected in IGV and the deletion was present. The amplicon was not amplified efficiently and has only a coverage of 167x. The TST15 BaseSpace App however applies a coverage threshold at 500x. Variants with a lower depth are not reported. The same sample was sequenced twice and the deletion was never detected. This is probably due to the fragmentation induced by the FFPE fixation.

Haloplex CSC did not find KRAS p.Gly12Val in patient G, also due to low coverage for this amplicon. The corresponding region was inspected in IGV: the region has a coverage of only 180x. Only one read showed the expected C¿A variant. Sample G is also the sample with the worst dCt (2.85). The hig fragmentation in this sample is probably responsible for the bad amplification. The same sample will be re-sequenced in a later run to check is the problem is related to the fragmentation of the sample or if library preparation was bad.

Additional previously unknown variants were found (TODO: put a table into the appendix)

3.4.2 Sensitivity Analysis

BRAF Mut and WT samples from Horizon were analyzed. The BRAF Mut sample was sequenced purely, as well as in a 1/3 and 2/3 dilution with the BRAF WT sample.

The observed variant frequencies as detected by the TST15 BaseSpace App are in line with the expected variant frequencies. D816V variant in cKIT could not be observed as the position of this

Table 7: Blablabla

Sample	Context	Tissue	Known variant	Halo CSC Freq (%)	TST15 Freq (%)
A	NSCLC	FFPE	EGFR L858R	24.7	21.4
B	NSCLC	FFPE	EGFR L858R	24.3	13.2
C	NSCLC	FFPE	EGFR L858R	32.7	29.8
D	Melanoma	FFPE	BRAF V600E	44.4	47.6
E	Melanoma	FFPE	BRAF V600E	18.4	21.4
F	NSCLC	FFPE	EGFR del19	not found	50
G	mCRC	FFPE	KRAS CD 12_13	p.Gly12Val (3.4)	not found
H	mCRC	FFPE	KRAS CD 12_13	p.Gly12Asp (37.4)	G12D (31.4)
I	mCRC	FFPE	KRAS CD 12_13	p.Gly13Asp (9.7)	G13D (8.7)
J	mCRC	FFPE	NRAS p.Gly12Asp	25.9	27.2
K	Melanoma	FFPE	BRAF V600E	66.2	59.5
L	mCRC	FFPE	NRAS p.Gly13Val	6.6	5
M	mCRC	FFPE	KRAS and NRAS WT	WT	WT
N	mCRC	FFPE	KRAS and NRAS WT	WT	WT
O	Melanoma	FFPE	BRAF V600E	37.4	

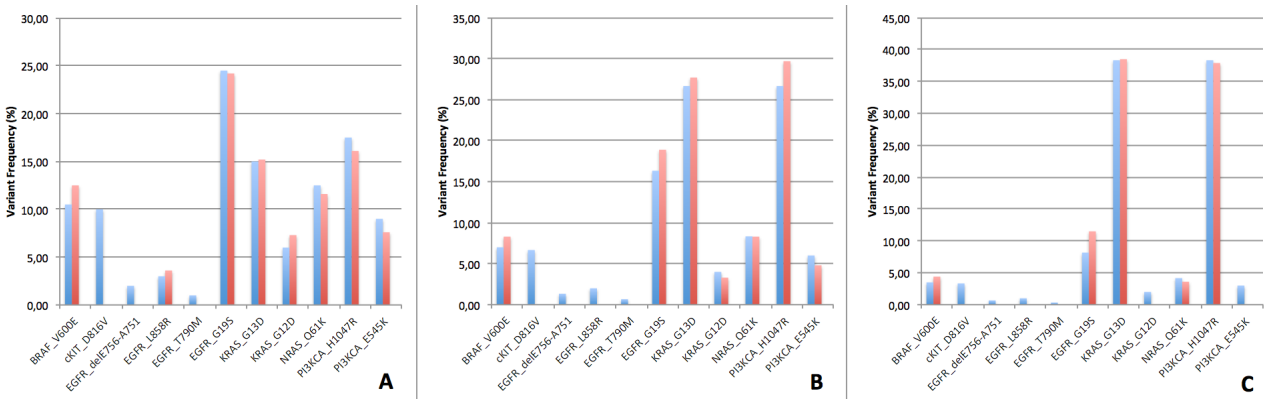


Figure 5: Blablabla

Table 8: Blablabla

Gene	Variant	Exp 100%	Obs Halo CSC	Obs TST 15	Exp at 66%	Obs Halo CSC	Obs TST15	Exp at 33%	Obs Halo CSC	Obs TST15
BRAF	V600E	10.5	–	12.5	7	–	8.3	3.5	–	4.4
cKIT	D816V	10	–	–	6.67	–	–	3.33	–	–
EGFR	delE756-A751	2	–	–	1.33	–	–	0.67	–	–
EGFR	L858R	3	–	3.6	2	–	–	1	–	–
EGFR	T790M	1	–	–	0.67	–	–	0.33	–	–
EGFR	G719S	24.5	–	24.2	16.33	–	18.9	8.17	–	11.5
KRAS	G13D	15	–	15.2	26.67	–	27.7	38.33	–	38.5
KRAS	G12D	6	–	7.3	4	–	3.3	2	–	–
NRAS	Q61K	12.5	–	11.6	8.33	–	8.3	4.17	–	3.6
PIK3CA	H1047R	17.5	–	16.1	26.67	–	29.7	38.33	–	37.9
PIK3CA	E545K	9	–	7.6	6	–	4.8	3	–	–

variant is not covered by the TST15 kit.

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3986649/

TODO: do the same with Haloplex CSC

TODO: call variants with MuTect 1.1.7, VarScan 2, GATK HaplotypeCaller, SomVarIUS?, Freebayes?, Vardict??????? TODO: compare results

Among the variants detected by MuTect, 40–50 % are C¿T variants. This has been reported in several studies. TODO: do this for all samples, check if this is really statistically significant or only happened in a few samples

Tools that may be of use somehow: GATK SelectVariants; GATK VariantFiltration; GATK VariantEval; GATK ValidateVariants

4 Conclusions

References

[1] M. Berg, “EGFR and downstream genetic alterations in KRAS/BRAF and PI3K/AKT pathways in colorectal cancer – implictaions for targeted therapy,” *Discovery Medicine*, vol. 14, no. 76, pp. 207–214, 2012.

[2] K. Cibulskis, M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E. S. Lander, and G. Getz, “Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples,” *Computational Biology*, 2013.

[3] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray, “Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012,” *International Journal of Cancer*, 2015.

[4] R. d. d. Direction de la Santé, Service des statistiques, “Statistiques des causes de décès pour l’année 2014,” 2014.

[5] F. Mertes, A. E. Sharawy, S. Sauer, J. M. L. M. van Helvoort, P. van der Zaag, A. Franke, M. Nilsson, H. Lehrach, and A. J. Brookes, “Targeted enrichment of genomic dna regions for next-generation sequencing,” *Briefings in Functional Genomics*, vol. 10, no. 6, pp. 374–386, 2011.