

Aus dem Institut für Klinische Pharmakologie  
(Prof. Dr. med. J. Brockmöller)  
im Zentrum Pharmakologie und Toxikologie  
der Medizinischen Fakultät der Universität Göttingen

**Genomic variations in the EGFR pathway in relation to skin  
toxicity of EGFR inhibitors analyzed by deep sequencing**

**INAUGURAL-DISSERATION**  
zur Erlangung des Doktorgrades  
der Medizinischen Fakultät  
der Georg-August-Universität zu Göttingen

vorgelegt von

**Sayedmohammad Hasheminasab**

aus

Sary, Iran

Göttingen 2014

Dean: Prof. Dr. rer. nat. H. K. Kroemer

I. Referee: Prof. Dr. Med. J. Brockmüller

II. Referee:

III. Referee:

Date of oral examination

## AFFIDAVIT

Here I declare that my doctoral thesis entitled “**Genomic variations in the EGFR pathway in relation to skin toxicity of EGFR inhibitors analyzed by deep sequencing**” has been written independently with no other source and aids than quoted.

Sayedmohammad Hasheminasab

Göttingen, July 4<sup>th</sup>, 2014

## Contents

Abbreviations .....	vii
1 Introduction .....	1
1.1 Epidermal growth factor receptor inhibitors as anti-cancer drugs .....	2
1.1.1 Epidermal growth factor receptor structure .....	2
1.1.2 EGFR in human cancer .....	2
1.2 Adverse effects of EGFR inhibition therapy; Skin rash.....	4
1.2.1 Diarrhea .....	4
1.2.2 Hypomagnesemia .....	5
1.2.3 Skin rash as the main side effect.....	6
1.3 Current data and hypotheses concerning the skin side effects.....	7
1.3.1 Normal skin layers.....	7
1.3.2 EGFR ligands.....	8
1.3.3 Pathophysiology of skin rash .....	9
1.3.4 Role of inflammation in EGFR inhibition related skin toxicity .....	11
1.4 Possible candidate genes in EGFR inhibition related skin rash.....	12
1.4.1 Ras-MAPK pathway.....	12
1.4.2 PI3K-Akt pathway.....	14
1.4.3 Inflammatory genes involved .....	16
1.5 Known facts about genomic variation in relation to skin side effects of EFGR blocking.....	18
1.6 Genetic variations and genetic testing .....	23
1.6.1 Human genome and its variations .....	23
1.6.2 Importance of genetic testing.....	24
1.6.3 Development of methods for analysis of human genomic variations.....	26
1.6.4 Massively parallel sequencing or deep sequencing.....	28
1.7 Hypotheses and a short work plan of the thesis .....	30
2 Methods.....	32
2.1 Clinical study design and definition of the phenotypes.....	32
2.2 Target region criteria .....	32
2.3 Overview of sample preparation for sequencing .....	33
2.4 DNA isolation and Library preparation .....	34
2.4.1 DNA isolation and genomic DNA fragmentation .....	35
2.4.2 Adaptor ligation .....	36

2.4.3	Size-selection .....	38
2.4.4	Library amplification .....	39
2.5	Analysis of library preparation efficacy .....	39
2.6	Quantifying and pooling of amplified size selected libraries .....	40
2.7	Target enrichment .....	42
2.7.1	First hybridization .....	43
2.7.2	Second hybridization.....	46
2.8	Template preparation .....	46
2.8.1	Enriched pooled library quantification .....	46
2.8.2	Emulsion PCR .....	46
2.8.3	ISP Enrichment.....	48
2.9	Massively Parallel Sequencing .....	50
2.9.1	Addition of sequencing primer and sequencing polymerase .....	50
2.9.2	Loading the sequencing chip.....	51
2.9.3	Sequencing.....	51
2.10	Variant calling .....	52
2.10.1	DNASTAR software.....	52
2.10.2	Torrent Variant Caller .....	53
2.11	Statistical analysis and functional prediction of nonsynonymous variants .....	57
3	Results.....	59
3.1	Patients .....	59
3.2	Target enrichment .....	64
3.3	Template preparation, loading the sequencing chip and massively parallel sequencing .....	65
3.4	Variant detection .....	65
3.4.1	DNASTAR software and Torrent Variant Caller variants .....	65
3.4.2	Comparison of variants found using high stringency, low stringency and DNASTAR software and a description of occurred problems.....	65
3.4.3	Custom parameters as a solution .....	72
3.5	Validation of the consistency of genotypes found using deep sequencing.....	72
3.6	Frequency of SNPs' missing calls in our sample population. ....	73
3.6.1	Hardy-Weinberg equilibrium .....	73
3.6.2	Minor allele frequency (MAF) of the SNPs.....	74
3.7	Functionality of genetic variants.....	75

3.8	Runs of homozygosity .....	79
3.9	Association of SNPs with skin rash.....	81
3.9.1	First analysis phase: Genomic variants predisposing to skin rash .....	81
3.9.2	Second analysis phase: Genomic variants predisposing to high grade skin rash .....	85
3.10	Gene-based and pathway-based analyses.....	88
3.11	Diarrhea as another side effect .....	89
3.12	Survival and effect sizes in carriers of non-synonymous variants .....	90
4	Discussion.....	95
4.1	First analysis phase: Genomic variants predisposing to skin rash .....	95
4.2	Second analysis phase: Genomic variants predisposing to high-grade skin rash .....	100
4.3	Common SNPs in both phases of analysis .....	100
4.4	Previously reported skin rash biomarker rs2227983.....	101
4.5	Novel functionally damaging variations.....	101
4.6	Runs of homozygosity .....	102
4.7	Statistical issues regarding analysis of association of EGFR inhibition related skin toxicity and genetic variations.....	103
4.8	Gene-based and pathway-based analysis.....	103
4.9	The unsolved mystery of association between skin rash and survival .....	104
4.10	Outlook .....	106
4.11	Study limitations .....	107
4.12	Conclusion.....	108
5	Summary .....	109
6	References .....	111

## List of Figures

Figure 1.3.1 Structure of normal epidermis and expression of EGFR in cells in this layer of skin.....	8
Figure 1.4.1 Ras-MAPK pathway.....	13
Figure 1.4.2 PI3K-Akt pathway.....	15
Figure 1.4.3 Inflammatory genes involved in pathogenesis of rash. ....	17
Figure 1.4.4 Overview of EGFR downstream signaling. ....	18
Figure 2.3.1 Overview of sample preparation steps prior to sequencing. ....	34
Figure 2.4.1 Analysis of fragmented genomic DNA. ....	35
Figure 2.5.1 Analysis of library preparation efficacy. ....	40
Figure 2.7.1 Mechanism of action of adaptor blockers and Human Cot-1 DNA. ....	44
Figure 2.8.1 Emulsion PCR. ....	47
Figure 2.8.2 Emulsion PCR difficulties.....	48
Figure 2.8.3 ISP enrichment. ....	49
Figure 2.9.1 Structure of an Ion chip and base calling.....	50
Figure 2.9.2 Ion 316 chip.....	51
Figure 2.10.1 Screen-shot of a Torrent variant caller output. ....	54
Figure 2.10.2 Integrative Genome Viewer (IGV) interface. ....	55
Figure 2.11.1 Explanation of genotypic and allelic models. ....	57
Figure 3.1.1 Spine plot of correlation of skin rash and diarrhea in study population. ....	60
Figure 3.1.2 Correlation between skin rash and survival in patients who passed away during the study period.....	61
Figure 3.1.3 Correlation between diarrhea and survival in patients who passed away during the study period.....	61
Figure 3.1.4 Correlation between gender and skin rash, diarrhea and survival.....	62
Figure 3.1.5 Correlation between age and skin rash, diarrhea and survival. ....	63
Figure 3.2.1 Target enrichment efficacy and coverage analysis.....	64
Figure 3.4.1 Common variants between two Torrent Variant Caller settings and DNASTAR software. ....	66
Figure 3.4.2 The integrative genomics view of a problem occurred due to a homopolymer region.....	69
Figure 3.4.3 Homopolymer as reason for low quality and missing calls.....	70
Figure 3.4.4 The integrative genomics view of a problem occurred due to a low coverage.....	71
Figure 3.6.1 Histograms of missing (or No calls) among the SNPs. ....	73
Figure 3.6.2 Minor allele frequencies (MAF) of SNPs categorized by novelty of SNPs. ....	74
Figure 3.6.3 Minor allele frequency of all exonic SNPs.....	75
Figure 3.7.1 Number of variations in 1000 bp of targeted region categorized by genes. ....	77
Figure 3.9.1 Manhattan plot of association of SNPs found in this study and the presence of skin rash grade $\geq 1$ . ....	81
Figure 3.9.2 Association plot and LD of EGFR SNPs and existence of rash. A) Five sets of EGFR SNPs were associated to rash of grade 1 or higher. ....	84
Figure 3.9.3 Linkage Disequilibrium plot of significant SNPs in MAPK1 gene. ....	85
Figure 3.9.4 Manhattan plot of association of SNPs found in this study and the presence of skin rash grade $\geq 2$ . ....	87
Figure 3.9.5 Association plot of EGFR SNPs and existence of high grade rash.....	87
Figure 3.11.1 Manhattan plot of association of SNPs found in this study and the presence of diarrhea. .	90

Figure 4.1.1 Structure of 5' sequence of EGFR gene. ....	96
Figure 4.1.2. Predicted function of regions of EGFR SNPs with lowest P value in normal human keratinocytes.....	98
Figure 4.1.3 Lack of coverage of EGFR promoter region. ....	99
Figure 4.9.1 Survival plot of the patients in our study (in days) categorized the maximum skin rash grade occurred during the follow-up. ....	106



## List of tables

Table 1.2.1 Adverse effects of EGFR inhibition therapy. ....	4
Table 1.5.1 Summary of polymorphisms studies in EGFR gene by different studies and their association with skin rash. ....	21
Table 2.1.1 Grading of rash according to toxicity criteria of the National Cancer Institute (NCI CTCAE) version 3.0.....	32
Table 2.2.1 The selected genes and their corresponding size.....	33
Table 2.4.1 Adaptor sequences. The IonXpress barcode set <sup>TM</sup> was used.....	37
Table 2.4.2 Reagents used in the ligation reaction of sheared DNA and adaptors.....	38
Table 2.4.3 PCR program of library amplification after size selection.....	39
Table 2.6.1 Genomic library pools' composition and pre-enrichment concentration. ....	42
Table 2.7.1 Amplification of the first capture reaction.....	45
Table 2.8.1 Onetouch ES 8 well strip's wells contents.....	49
Table 2.10.1 Torrent Variant Caller parameter configurations. ....	56
Table 2.11.1 An example output data of Polyphen2 prediction tool. ....	58
Table 3.1.1 Patient characteristics.....	59
Table 3.4.1 Number of common variants between two Torrent Variant Caller settings and DNASTAR software. ....	66
Table 3.4.2 Analysis of discrepancies between Torrent Variant Caller settings high and low stringency..	68
Table 3.5.1 Consistency of genotyping by deep sequencing tested by previously genotyped positions...	72
Table 3.7.1 Number of variants categorized by their effect (noncoding, synonymous, single amino substitution).....	76
Table 3.7.2 Function prediction of effects of nonsynonymous SNPs. ....	78
Table 3.9.1 Germline SNPs associated with presence of skin toxicity of level 1,2 or 3 with P value < 0.05. ....	83
Table 3.9.2 Germline SNPs associated with presence of skin toxicity of level 2 and 3 with P value < 0.05. ....	86
Table 3.10.1 Pathway-based association analysis. ....	88
Table 3.11.1 Ten most significant germline SNPs associated with existence of diarrhea using Fisher's exact test.....	89
Table 3.12.1 Survival time, skin rash and diarrhea in relation to the amino acid substitutions identified in my project. ....	91
Table 3.12.2 Survival, skin rash and diarrhea in carriers of variants predicted to be damaging.....	92
Table 3.12.3 Functional domains and possible function of the protein domain in which the nonsynonymous variations are localized.....	93

## Abbreviations

AKT	v-akt murine thymoma viral oncogene homolog
ARAF	v-raf-1 murine leukemia viral oncogene homolog
ATP	Adenosine triphosphate
BAD	Bcl-2-associated death promoter protein
BCL2	B-cell lymphoma 2
BIM	BCL-2-interacting mediator
BSA	Body surface area
CCL18	CC Chemokine ligand 18
CCL2	CC Chemokine ligand 2
CCL3	CC Chemokine ligand 3
CCL5	CC Chemokine ligand 5
CDCV	Common disease common variations
CDRV	Common disease rare variation
CFTR	Cystic fibrosis transmembrane conductance regulator
CHO	Chinese hamster ovary
CXCL10	CXC chemokine ligand 10
CXCL9	CXC chemokine ligand 9
CXCR	CXC chemokine receptor
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EGFRI	Epidermal growth factor receptor inhibitors
EMSA	Electrophoretic mobility shift assay
ERBB	Erythroblastic Leukemia Viral Oncogene Homolog
ERK	extracellular signal-regulated kinase
FAT	Focal adhesion kinase
FDA	Food and Drug Administration
FGF	Fibroblast growth factors
FOS	Homolog of the retroviral oncogene v-fos.
FOXO	O subclass of the forkhead family of transcription factors
GDP	Guanosine diphosphate
GWAS	Genome wide association study
HER2	Human Epidermal growth factor Receptor 2
HLA	Human leukocyte antigen
HNSCC	Head and neck squamous cell carcinoma
HPV	Human papillomavirus
HWE	Hardy–Weinberg equilibrium

IAP	Inhibitor of apoptosis proteins
IGV	Integrative genomics viewer
IRS	Insulin receptor substrate 1
ISP	Ion Sphere Particle
JUN	Retrovirus-associated DNA sequences (jun)
LMU	Ludwig Maximilian University of Munich
LTE	Low Tris-EDTA buffer
MAF	Minor allele frequency
MAP	Mitogen-activated protein
MAPK	Mitogen-activated protein kinase
MAPKK	Mitogen-activated protein kinase kinase
MAPKKK	Mitogen-activated protein kinase kinase kinase
MEK	MAPK/ERK kinase
MTOR	Mammalian target of rapamycin
MTT	Methylthiazol tetrazolium
NCI	National cancer institute
NFKB	Nuclear factor kappa B
NIH	National institutes of health
NSCLC	Non-small cell lung carcinoma
PCR	Polymerase chain reaction
PDK	phosphatidylinositol-dependent kinase
PED	Phosphoprotein enriched in astrocytes 15
PEG	Polyethylene glycol
PGM	Personal genome machine
PTEN	Phosphatase and tensin homolog
RAF	v-raf murine leukemia viral oncogene homolog
RANTES	Regulated on activation, normal T cell expressed and secreted
RAS	Rat sarcoma
REL	Retrovirus-associated DNA sequences (v-rel)
RNA	Ribonucleic acid
ROH	Runs of homozygosity
RSK	Ribosomal s6 kinase
RTK	Receptor tyrosine kinase
SCC	Squamous cell carcinoma
SNP	Single nucleotide polymorphism
SOS	Son of sevenless
TKI	Tyrosine kinase inhibitor
TNF	Tumor necrosis factor
TSC	Tubular sclerosis complex
TVC	Torrent variant caller
UTR	Untranslated region
WGS	Whole genome sequencing
XCL1	XC chemokine ligand 1

## 1 Introduction

This project is mainly about the impact of genomic variation on one typical side effect and to some extent on the efficacy of an important new class of drugs inhibiting the epidermal growth factor receptor (EGFR). This receptor is a member of receptor tyrosine kinases (RTK) family (Hynes and Lane 2005). It contains an extracellular site including a domain for binding to its ligands, a single membrane region and a cytoplasmic region. The cytoplasmic region contains a tyrosine kinase domain. Binding of the specific ligands will activate the kinase activity of the intracellular region. Its activation initiates EGFR's signaling pathway (Olayioye et al. 2000, Schlessinger 2004, Yarden and Sliwkowski 2001).

The EGFR is important in numerous types of cancer such as breast cancer, head-and-neck cancer, non-small cell lung cancer (NSCLC), renal cancer, ovarian cancer, esophageal cancer and colon cancer (Herbst and Langer 2002, Sizeland and Burgess 1992). These types of cancer for example are still difficult to cure; however, efficacy of therapy and – remarkably – its tolerance are moderately improved using the new class of drugs which target EGFR. As a result of this approach, some of the classical side effects of chemotherapy (e.g. hair loss or mucositis) are not observed or observed with lower incidence and severity. On the other hand, there are certain adverse effects related to this class of targeted therapy including skin toxicity. However, the frequency of adverse drug reactions as well as the efficacy of treatment is not the same in all subjects receiving these drugs.

Genomic variation may explain why some people respond better and other respond worse to this therapy. Interestingly, there is an obvious positive correlation between existence and severity of skin toxicity and survival of the patients. This may underline the role of germline genetic variations considering the individually variable extent of skin toxicity. In the following text the EGFR and its ligands, possible mechanisms in which the skin toxicity might be formed, the impact of genetic variation in developing traits and diseases, and a summary of methods used for detection of these variations are described.

## 1.1 Epidermal growth factor receptor inhibitors as anti-cancer drugs

### 1.1.1 Epidermal growth factor receptor structure

The presence of epidermal growth factor (EGF) binding site on the cell surface was confirmed in 1975 using radio-labeled EGF in fibroblasts (Carpenter et al. 1975). Later on, the function of the EGF receptor (EGFR) was thoroughly investigated and in vivo effects of EGFR targeting was surveyed in mouse models. Knocking out EGFR in different setting resulted in severe abnormalities such as embryonic lethality (Threadgill et al. 1995), or epithelial growth and development failure of the skin, the lung and the gastrointestinal (GI) tract (Miettinen et al. 1995, Sibilina and Wagner 1995, Threadgill et al. 1995). The numerous severe abnormalities associated with absence of EGFR testify for the wide range of roles of EGFR in development of epithelial cells.

The EGFR also known as ERBB-1 (Erythroblastic Leukemia Viral Oncogene Homolog) is a member of ERBB subclass of receptor tyrosine kinases (Hynes and Lane 2005). It has a single membrane region and an extracellular site for binding to its ligands. The cytoplasmic region contains a tyrosine kinase domain. Binding of the specific ligands will activate the kinase activity of the intracellular region. Upon activation, the kinase domain will phosphorylate tyrosine residues on the same molecule (autophosphorylation). Subsequently, homo- (with another EGFR) or hetero- (with another member of ERBB family) dimerization will take place. This dimerization facilitates the beginning of a kinase cascade responsible for activation of EGFR's downstream signaling pathway (Olayioye et al. 2000, Schlessinger 2004, Yarden and Sliwkowski 2001).

### 1.1.2 EGFR in human cancer

Overexpression and deregulation of EGFR were noticed in several epithelial tumors such as breast cancer, NSCLC, renal cancer, ovarian cancer, esophageal cancer and colon cancer (Herbst and Langer 2002, Sizeland and Burgess 1992). Furthermore, several single nucleotide variations (SNVs) were found in the EGFR which could increase the kinase activity of this receptor in human

cancers (Gill et al. 1984, Humphrey et al. 1990). Therefore, the EGFR was targeted as an anticancer strategy (Masui et al. 1984b).

Two classes of EGFR inhibitors are now applied in clinical practice: Monoclonal antibodies, and Small molecule inhibitors.

#### *1.1.2.1 Monoclonal antibody EGFR inhibitors*

These antibodies bind to the extracellular region of the receptor which has the role of ligand binding. This competition between the drug and EGFR ligands will turn off the downstream signaling pathway (Li et al. 2005, Normanno et al. 2003). Cetuximab and panitumumab are two approved drugs from this class. Cetuximab is a human-mouse chimeric monoclonal antibody (IgG1 subtype) which was approved by the FDA for treatment of advanced colorectal cancer and locally advanced squamous cell carcinoma (SCC) of head and neck (HNSCC) (Blick and Scott 2007, Wong 2005). Panitumumab was approved by the FDA in 2006 for treatment of patients with metastatic colorectal cancer having disease progression after chemotherapy; although, later on treatment was limited to *KRAS* wild-type cancers (Hocking and Price 2014).

#### *1.1.2.2 Small-molecule EGFR inhibitors*

Small-molecule inhibitors or tyrosine kinase inhibitors (TKIs) inhibit the activity of the intracellular tyrosine-kinase domain of the receptor by attaching to its ATP binding site. Erlotinib and gefitinib are two examples of this drug category. Erlotinib was approved by FDA in 2004 for treatment of patients with locally advanced or metastatic non-small cell lung carcinoma (NSCLC) (Cataldo et al. 2011). Recently it has also been approved as first-line treatment of pancreatic cancer in combination with gemcitabine (Ciardiello and Tortora 2008). Gefitinib was approved by the FDA already in 2003; however, because of failure in clinical trials its usage is limited to patients who have previously received it and benefited from it (Thatcher et al. 2005).

## 1.2 Adverse effects of EGFR inhibition therapy; Skin rash

In general, adverse effects of EGFR targeted therapy are often less severe and this therapy is better tolerated than cytostatic (cytotoxic) chemotherapy. Classical adverse effects of cytostatic chemotherapy, such as bone marrow suppression, alopecia and mucositis are not observed during EGFR inhibitor monotherapy. Since EGFR is mostly expressed in epidermal or endothelial cells and has important physiological roles, its inhibition results in adverse effects such as skin toxicity, and diarrhea. Additionally hypomagnesemia is found as another side effect (Cunningham et al. 2004, Goldberg et al. 2004, Parmar et al. 2013, Schrag et al. 2005, Van Cutsem et al. 2007). A summary of adverse effect prevalence is displayed in Table 1.2.1.

Table 1.2.1 Adverse effects of EGFR inhibition therapy. Hypomagnesemia is not reported in small molecule tyrosine kinase inhibitors.

	Erlotinib	Gefitinib	Cetuximab	Panitumumab	Control group
Skin rash	<b>50%</b> [monotherapy] (Ciuleanu et al. 2012) <b>63%</b> [monotherapy] (Nakagawa et al. 2012)	<b>49%</b> [monotherapy] (Shi et al. 2013) <b>31%</b> [monotherapy] (Mohamed et al. 2005b)	<b>79%</b> [monotherapy or combination therapy with irinotecan or FOLFIRI] (Ishiguro et al. 2012) <b>81%</b> [combination therapy with perimetrexed or docetaxel] (Kim et al. 2013)	<b>78%</b> [monotherapy or combination therapy with mFOLFOX6, FOLFIRI, and irinotecan] (Boku et al. 2014) <b>53%</b> [monotherapy] (Van Cutsem et al. 2008)	<b>2%</b> [combination therapy with perimetrexed or docetaxel] (Kim et al. 2013)
Diarrhea	<b>18%</b> (Ciuleanu et al. 2012) <b>23%</b> (Nakagawa et al. 2012)	<b>29%</b> (Shi et al. 2013) <b>22%</b> (Mohamed et al. 2005b)	<b>15.1%</b> (Ishiguro et al. 2012) <b>28%</b> (Kim et al. 2013)	<b>9%</b> (Van Cutsem et al. 2008)	<b>13%</b> (Kim et al. 2013)
Hypomagnesemia	–	–	<b>11%</b> (Ishiguro et al. 2012) <b>21%</b> (Kim et al. 2013)	<b>36%</b> (Boku et al. 2014) <b>29%</b> (Van Cutsem et al. 2008)	<b>6%</b> (Kim et al. 2013)

### 1.2.1 Diarrhea

Diarrhea occurs in up to 29% of patients treated with EGFR inhibitors (Table 1.2.1). In case of diarrhea, loperamide was frequently used as a treatment; however, diarrhea remains to be one of the dose-limiting events (Ciardiello et al. 1999, ImClone Systems Incorporated 2006, Kim, E et al. 2008, Shepherd et al. 2005). Diarrhea is a frequent adverse effect of some cytostatic

chemotherapeutic agents such as 5-fluorouracil or irinotecan; thus, not all studies applying EGFR inhibitor combination therapies are informative. However, as an example of EGFR monotherapy, in TITAN study (a randomized multicenter phase 3 of non-small-cell lung cancer with poor prognosis) diarrhea has occurred in 18% of patients with erlotinib monotherapy compared to 2% of patients with docetaxel or perimetrexed chemotherapy regimens (Ciuleanu et al. 2012). . On the other hand, in (Ishiguro et al. 2012) study only 5% of patients treated with cetuximab monotherapy had diarrheas compared to 18% in case of combination of cetuximab with irinotecan or FOLFIRI (folinic acid, fluorouracil and irinotecan chemotherapy regimen). This finding shows that when considering diarrhea, side effects of cytotoxic agents should also be regarded. The underlying mechanisms are not well understood. One suggested mechanism is that EGFR inhibition leads to excessive chloride secretion in the intestine and therefore causes a secretory diarrhea (Uribe et al. 1996).

### 1.2.2 Hypomagnesemia

Hypomagnesemia was not primarily reported as a side effect of EGFR inhibition and was surprisingly only recognized as a frequent problem after approval of these drugs. Retrospective surveys have shown hypomagnesemia in 50–65% of patients receiving anti EGFR monoclonal antibodies (Jean and Shah 2008). This side effect only occurs in case of using monoclonal antibodies against EGFR but not other targeted therapies using monoclonal antibodies. Contrasting to the anti-EGFR antibodies, there is no evidence of an association between the tyrosine kinase inhibitors gefitinib, erlotinib and hypomagnesaemia (Costa et al. 2011). Hypomagnesemia occurs at 6–8 weeks after commencement of EGFR monoclonal antibody therapy (Fakih et al. 2006). This side effect is caused by decreased reabsorption of magnesium in distal tubules of kidney nephrons and therefore renal wasting of magnesium. EGFR activation is necessary for function of TRPM6 (an epithelial Mg<sup>2+</sup> channel) and therefore its inhibition causes hypomagnesemia (Costa et al. 2011).

Other less common side effect are fatigue, asthenia, dyspnea, abdominal pain, nausea, and anorexia (Fakih et al. 2006).



### 1.2.3 Skin rash as the main side effect

Skin rash is clinically the most noteworthy side effect of EGFR therapy. Other skin related toxicities are xerosis, paronychia, hair changes, and telangiectasias (Tsimboukis et al. 2009). Occurrence of rash negatively affects the patients' compliance to the treatment. Particularly if occurring more extensively, skin rash significantly compromises the patient's quality of life and thereby may have a negative impact on treatment's outcome (Nanney et al. 1990a, Nanney et al. 1990b). Minor forms of EGFR inhibition associated rash occur in up to 100% of patients but there are patients in which they occur only to a very minor extent (Parmar et al. 2013).

Although rash can occur at minimum 3 days after start of erlotinib therapy, the median incidence of rash is the 8<sup>th</sup> day (Boeck et al. 2007, Saif et al. 2008). As skin rash occurs during both, monoclonal antibodies and small molecule inhibitors therapies, it is believed to be a class-related side effect. Notably, no difference between the histological pictures of rash in the two sub group has been reported. Rash occurs more frequently and severely in treatment with monoclonal antibodies compared to small-molecule inhibitors. Between the small molecule inhibitors, skin toxicity tends to be more severe with erlotinib compared to gefitinib (Togashi et al. 2011). Rash will ultimately fade out during one month after the end of EGFRi dosing (Li and Perez-Soler 2009).

#### 1.2.3.1 Characterization of rash

Skin rash is referred to as folliculitis (Peuvrel et al. 2012) or more historically, acneiform eruption (Segaert and Van Cutsem 2005). It is often categorized according to National Cancer Institute Common Toxicity Criteria (NCI CTC). The grade of skin rash is found to be dose related (Pérez-Soler and Saltz 2005, Robert et al. 2005, Roé et al. 2006, Segaert et al. 2009, Wolf et al. 2004). In addition, there is a well-confirmed correlation between overall survival and the grade of rash independent of which drug has been used (Peeters et al. 2009, Pérez-Soler et al. 2004, Peuvrel et al. 2012, Shepherd et al. 2005).

Anatomically the rash is more frequent on the face, upper torso (V-shaped region), shoulders, and neck. Less frequently lower parts of the back, gluteal area, arm or legs are affected (Galimont-Collen et al. 2007, Pérez-Soler and Saltz 2005, Robert et al. 2005, Segaert et al. 2009). Thus, there may be an interaction of EGFRi induced skin rash with light exposure. Palms, soles

and mucosa, are always spared. (Agero et al. 2006, Baselga et al. 2002, Robert et al. 2005, Segaert et al. 2009). It has a papulo-pustular appearance; in addition, these lesions (papules and pustules) are monomorphic and aseptic. It is very rare that necrosis happens (Mass 2004, Segaert et al. 2009). A useful distinction from acne is that comedones will never occur. The rash is occasionally painful (Boone et al. 2007, Burtneess et al. 2009, Segaert and Van Cutsem 2005).

### 1.3 Current data and hypotheses concerning the skin side effects

#### 1.3.1 Normal skin layers

Because this is important for understanding of the adverse effects, a brief introduction into skin structure is given here. The human skin is composed of three major layers: epidermis, dermis and hypodermis. The epidermis itself consists of 4 layers: the basal, spinous, granular and the stratum corneum (Figure 1.3.1). Nearly 90% of the cells in epidermis are keratinocytes (other cell types in the skin include: fibroblasts and melanocytes). The basement membrane layer – placed between epidermis and dermis – is responsible for the generation of the epidermal layer (Candi et al. 2005). This layer has the highest expression of the EGFR. Activation of this receptor plays a critical role during the development and migration of keratinocytes in the direction of surface. During this migration, the keratinocytes are differentiated and the expression of EGFR is lowered. The last layer – Stratum corneum – is formed after apoptosis of keratinocytes (Candi et al. 2005, Nanney et al. 1990a).

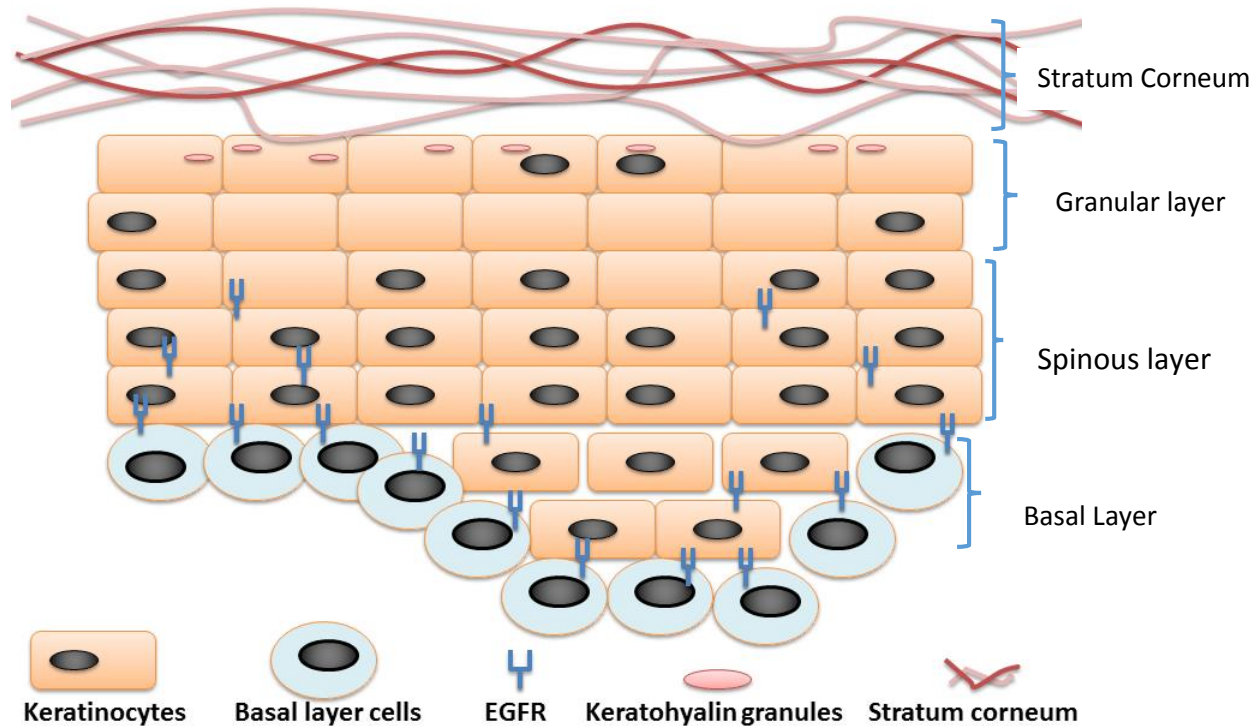


Figure 1.3.1 Structure of normal epidermis and expression of EGFR in cells in this layer of skin. The EGFR is prominently expressed in the basal layer which is responsible for the generation of the epidermis. As the cells migrate to the more superficial layers, the EGFR expression is decreased. In the superficial layers of the epidermis there are keratohyalin granules which contributes to keratinization of the stratum corneum. The stratum corneum is mostly composed of dead keratinocytes, keratin fibers and filaggrin filaments.

### 1.3.2 EGFR ligands

EGFR can be activated upon the presence of several proteins such as EGF, transforming growth factor- $\alpha$  (TGF $\alpha$ ), amphiregulin, heparin-binding EGF (HB-EGF) or epiregulin. All of these proteins contain a region with the amino acid sequence motif  $CX_7CX_{(4-5)}CX_{(10-13)}CXCX_8GXRC$ ; where Cs are cysteine, G is glycine, R is arginine and X could be any other amino acid. The characteristic feature of these molecules is a triple loop in structure, in which 6 cysteine amino acids make 3 intramolecular disulfide bridges. They are all EGFR specific ligands, meaning that they will not activate any other member of ERBB family (Hynes and Lane 2005). EGFR is activated by the auto and/or paracrine secretion of its ligands (by keratinocytes) such as amphiregulin which is the main ligand in the skin.

### 1.3.3 Pathophysiology of skin rash

EGFR is typically expressed on the surface of keratinocytes of the basement membrane layer of the skin. It is involved in several vital functions in human skin including cell growth and proliferation, apoptosis evasion, migration and wound healing. Deterioration of any of these functions may result in a pathologic status which leads to rash. These physiologic roles are described in the following text.

#### 1.3.3.1 EGFR in cell growth and proliferation

EGFR activation increases epidermal thickness by increasing the proliferation of the keratinocytes. Autocrine release of EGFR ligands was seen in keratinocytes both in vitro and in situ which can result in proliferation of these cells. Treatment of human keratinocytes derived from normal skin with EGFR inhibitors blocks the progression from the G1 to the S phase of the cell cycle and also inhibits DNA synthesis (Jost et al. 2000, Kobayashi et al. 1998).

EGFR inhibition also causes the keratinocytes to express terminal differentiation gene markers such as *KRT1* and *KRT10* (keratin molecules which normally exist in outer layers of epidermis). On the other hand, removal of EGFR inhibition facilitates the down regulation of *KRT1* and *KRT10* genes (Peus et al. 1997). In addition, EGF treatment downregulates the expression of other differentiation indicators such as profilaggrin and filaggrin and leads to nuclear retention (undifferentiated keratinocytes at the outer layer stratum corneum), in presence of extracellular matrix adhesion in skin reconstructs. Furthermore, premature hair keratinization was observed in EGFR-null mice (Threadgill et al. 1995).

#### 1.3.3.2 Apoptosis evasion

Activation of EGFR facilitates the survival and evasion of programmed cell death in keratinocytes. Early apoptosis of keratinocytes will lead to thin epithelium and compromised barrier function of skin. On the other hand, neoplasia could develop if the keratinocytes evade the normal apoptosis. If the EGFR is inhibited by either small molecule tyrosine kinase inhibitors or by the anti-EGFR-antibodies, keratinocytes will be prone to death when exposed to ultraviolet irradiation (Peus et

al. 1997) or oxidative stress (Wang et al. 2000) or when undergoing passaging in cell culture (Ravid et al. 2002). Apoptosis rate escalates up to fivefold between days 4 to 12 after EGFR inhibition. This time distance is in accordance with occurrence of rash in EGFR inhibitor (EGFRI) treated patients. Interestingly, treatment of fibroblasts and melanocytes with EGFRI has shown no change in apoptosis rate, that leaves the keratinocytes as the primary suspect for rash after EGFR inhibition (Rodeck et al. 1997). In addition, several anti-apoptotic genes were downregulated after EGFR inhibition such as survivin also known as baculoviral IAP repeat containing 5 (*BIRC5*) and B-cell lymphoma 2-associated athanogene (*BAG1*) (Lacouture 2006).

#### 1.3.3.3 Migration

EGFR activation increases the migration of keratinocytes to the outer layers of the skin and leads to loss of intercellular attachments (Betson et al. 2002). By a similar mechanism, EGFR inhibition may also plays a role in decreasing tumor invasion and metastasis. Increased cell adhesion after EGFRI therapy could be explained by upregulation of cadherin-associated protein CTNND2 and the desmosomal cadherin DSG2. In addition EGFR signaling is required for activation of small GTPases which increase the stability of cadherin regulated intra-cellular contacts (Betson et al. 2002).

#### 1.3.3.4 EGFR in wound healing

EGF receptor activation by its ligands amphiregulin and HB-EGF appears to be vital to wound healing. These mentioned ligands, but not TGF- $\alpha$ , are strongly induced in cultures of human skin biopsy as a model of wound healing. Increased expression of these two ligands was also seen after disruption of mouse skin. HB-EGF is vastly upregulated in healing wounds from biopsies taken 6 days after skin graft removal (Stoll et al. 1997). In EGFR-null mice skin repair was diminished during first 3 days after intervention. An inflammatory process occurred in the wounds of these mice (infiltration of inflammatory cells such as neutrophils was observed). Wound healing in the medium which contains inflammatory cells would be hindered because of ongoing proteolytic processes during inflammation (Repertinger et al. 2004). However, systemic

EGFR inhibition therapy does not impair wound healing in human after full thickness dermal excision (Govindan et al. 2003).

#### 1.3.4 Role of inflammation in EGFR inhibition related skin toxicity

EGFR inhibition by both small molecule TKIs or monoclonal antibodies may lead to the impairment of the mentioned vital EGFR functions such as proliferation, cell migration, wound healing, and differentiation. As a result, skin toxicity may occur caused by diminished skin integrity. On the other hand – possibly as a secondary effect consequent to the disrupted skin integrity – a moderate to severe inflammation is always present in the EGFR associated skin toxicity.

EGFR plays a key role in expression of cytokines in keratinocytes. EGFR activation is inversely associated with expression of chemokines and inflammatory molecules such as TNF $\alpha$ , regulated upon activation normal T cell expressed and presumably secreted protein (RANTES, also known as CCL5) and monocyte chemoattractant protein-1 (MCP1 also known as CCL2). Exposure of epithelial cells to small molecule TKIs elevated the levels of other inflammatory molecules such as CXCL10, and T lymphocytes recruiting chemokines viz. CCL18, XCL1, CXCL9, fractaline and CCL3 (Mascia et al. 2003, Rodeck et al. 1997). Subsequently, the potential chemo-attraction by these molecules will be suppressed after EGFR blockage. EGFR inhibition also stabilizes the chemokine molecules (Pastore et al. 2005).

The pro-inflammatory transcription factor NFKappaB is upregulated after treatment of cervical epithelial cells in an organo-typic culture with EGFRIs (Woodworth et al. 2005). Furthermore, local application of HuMab-10F8 (IL-8 neutralizing antibody) has reversed the skin rash occurred after local injection of Zalutumumab (an anti EGFR monoclonal antibody) (Bangsgaard et al. 2012).

#### 1.4 Possible candidate genes in EGFR inhibition related skin rash

As was outlined above, skin rash is believed to be a class effect of all EGFR inhibitors. We hypothesized that the inhibition of the EGFR pathways is the reason of rash and therefore we analyzed several important proteins from the EGFR pathways. In addition, inflammation is considered to play an important role in formation of EGFR inhibitor induced rash. The transcription factor NFκB is known to stimulate expression of proinflammatory genes and was shown to be upregulated after EGFR inhibition (Woodworth et al. 2005). Furthermore, an antibody neutralizing IL-8 has been shown to be effectively reducing and reversing the rash development. Considering the presented facts we have decided to analyze variation in the EGFR pathways and in selected pro-inflammatory pathways in our study sample. For a better understanding of these molecules a short description is presented in the following text.

Upon stimulation of EGFR two major pathways are activated. These are the Ras-MAPK and the PI3K-Akt pathways. Activation of the EGFR via its ligands results in phosphorylation of tyrosine residues of intracellular protein regions of the EGFR. The phosphotyrosine residue has high affinity for binding to adaptor molecules which after activation, mediate the binding of GTP to Ras and results in its activation (Kolch 2000, Stokoe et al. 1994).

##### 1.4.1 Ras-MAPK pathway

The Ras-MAPK pathway may be activated via the EGFR (Figure 1.4.1). MAPK stands for mitogen-activated protein kinase. The first molecule in this pathway is Ras which is a small molecule GTPase. The Ras family has four members, N-Ras, H-Ras, K-Ras4A and K-Ras4B which the last two are the result of alternative splicing in *KRAS* gene. Conformational changes that occur after binding of GTP to Ras subsequently activates the Raf protein (Barbacid 1987, Jancík et al. 2010, Vakiani and Solit 2011, Wennerberg et al. 2005).

Raf (also known as mitogen-activated protein kinase kinase kinase (MAPKKK), meaning that it phosphorylates MAPKK which in turn phosphorylates MAPK or ERK) is a serine/threonine kinase. Ras-GTP stabilizes Raf by binding to its N terminal and phosphorylating it; besides, it recruits Raf to the membrane. As a result, the C terminal is now made free through the conformational

changes that occur; and therefore can act as a kinase. Since we were not assured by the literature that which of the three raf proteins are relevant in keratinocytes, we analyzed all three, *RAF1*, *ARAF* and *BRAF*. All three kinases have a very similar structure. Raf then continues the kinase cascade by activating MEK (Stokoe et al. 1994, Vakiani and Solit 2011).

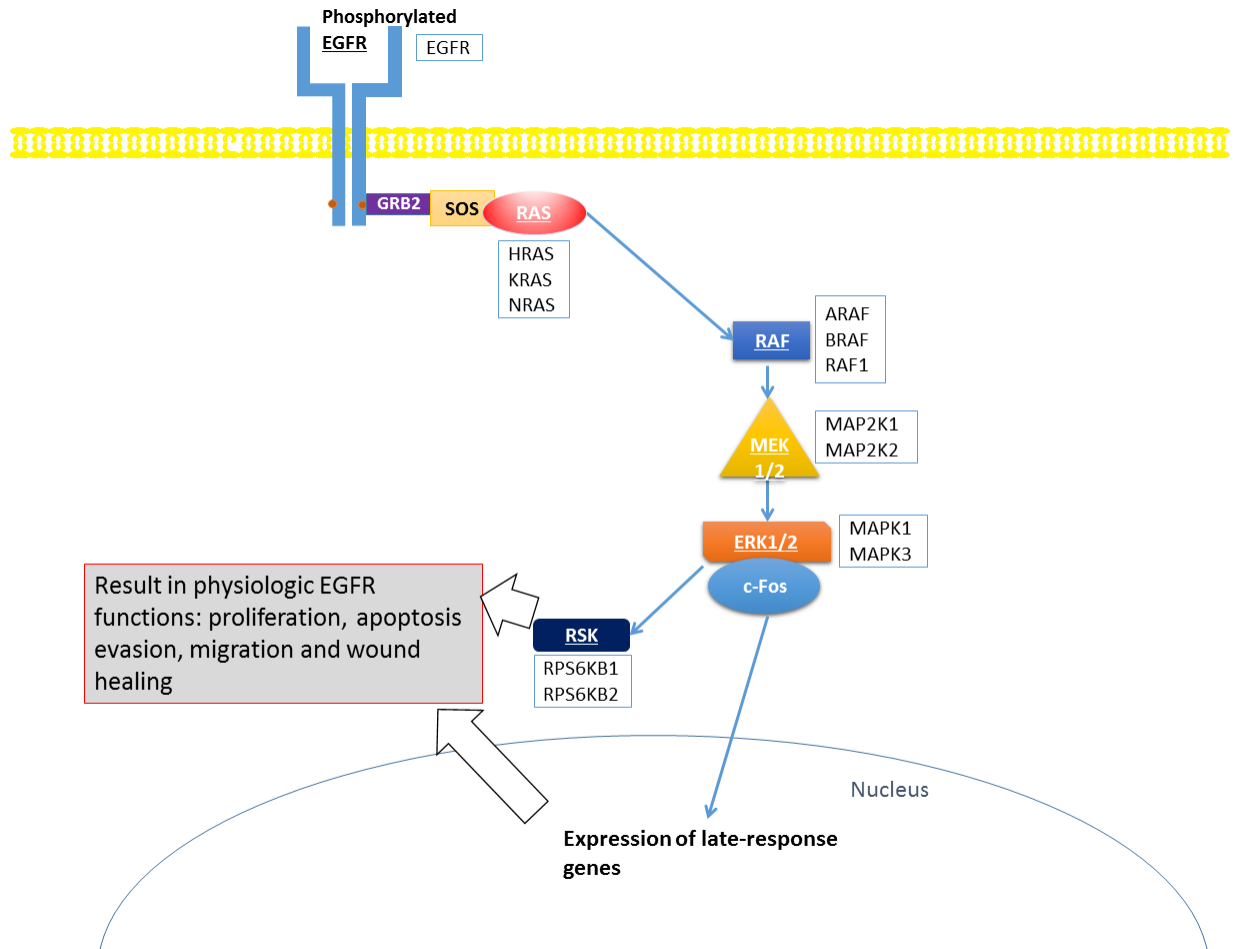


Figure 1.4.1 Ras-MAPK pathway. The underlined molecules are investigated in current study. EGFR, epidermal growth factor receptor; GRB2, growth factor receptor-bound protein 2; SOS, son of sevenless; RAS, rat sarcoma; RAF, v-raf-1 murine leukemia viral oncogene homolog; MEK, MAPK/ERK kinase; ERK, extracellular signal-regulated kinase; RSK, Ribosomal s6 kinase; c-Fos, homolog of the retroviral oncogene v-fos. Names of the genes analyzed corresponding to each protein is shown in boxes.

MEK, also known as MAPKK, is a dual specific kinase – targeting tyrosine and threonine – which is normally inactive. Raf causes a conformational change in MEK and activates it. This is because it phosphorylates ERK which is also known as MAPK. Two isoforms of MEK have been identified; MEK1 (*MAP2K1*) and MEK2 (*MAP2K2*). MEK1 and MEK2 activate ERK1 (*MAPK1*) and ERK2



(*MAPK3*), respectively (Roskoski 2012).

The ERK proteins (extracellular signal-regulated kinase) are serine/threonine kinases, which after being phosphorylated, can phosphorylate its cytoplasmic or nuclear substrates. After being activated, ERK migrates to the nucleus and increases expression of transcription factors or immediate early response genes such as *C-FOS* and *C-MYC*. The nuclear transcription factor targets of ERK1/2, play a major role in inducing the expression of the late-response genes that boost cell survival, cell division, and cell motility. This is believed to be the major result of EGFR stimulation in MAPK pathway (Eferl and Wagner 2003, Murphy et al. 2002, Okazaki and Sagata 1995).

On the other hand, there are more than 50 cytoplasmic substrates for ERK including: RSK family protein kinases, phosphoprotein phosphatases, cytoskeletal proteins, apoptotic proteins, and regulatory and signaling molecules. Ribosomal S6 kinase (RSK) proteins are 90 kDa serine/threonine kinases that control cell growth, motility, proliferation, and survival. ERK1/2 directly activates RSK enzymes when it is stimulated by growth factors, polypeptide hormones, neurotransmitters, or chemokines. RSKs are key components downstream from the Ras-MAPK signaling cascade which phosphorylate 35 proteins. In different types of cancer RSK1 and RSK2 are up-regulated that suggests their involvement in oncogenesis (Anjum and Blenis 2008, Carriere et al. 2008, Ramos 2008).

#### 1.4.2 PI3K-Akt pathway

Another pathway downstream of EGFR is PI3K-Akt pathway. Many of physiologic roles of EGFR are mediated through this pathway (Figure 1.4.2). EGFR binds to and stimulates Phosphoinositide-3-kinase regulatory subunit of Phosphatidylinositol 3 kinases (PI3Ks) via an adaptor proteins, such as IRS-1. PI3Ks are membrane-anchored proteins which facilitate the transfer of a phosphate group of ATP to phosphoinositides. PI3Ks are categorized into four classes: IA, IB, II, III. Class IA interacts with receptor tyrosine kinases and are heterodimers consisting of regulatory and catalytic subunits. Activated PI3K catalyzes the phosphorylation of phosphatidylinositol 3,4-bisphosphate (PIP2) to phosphatidylinositol 3,4,5-triphosphate (PIP3)

(Fresno Vara et al. 2004, Rameh and Cantley 1999, Volinia et al. 1995).

Tumor suppressor phosphatase and tensin homolog (PTEN) can undo the action of PI3Ks by conversion of PIP3 to PIP2 (de-phosphorylation of PIP3) which results in inhibition of the PI3K-Akt pathway (Li et al. 1997, Salmena et al. 2008).

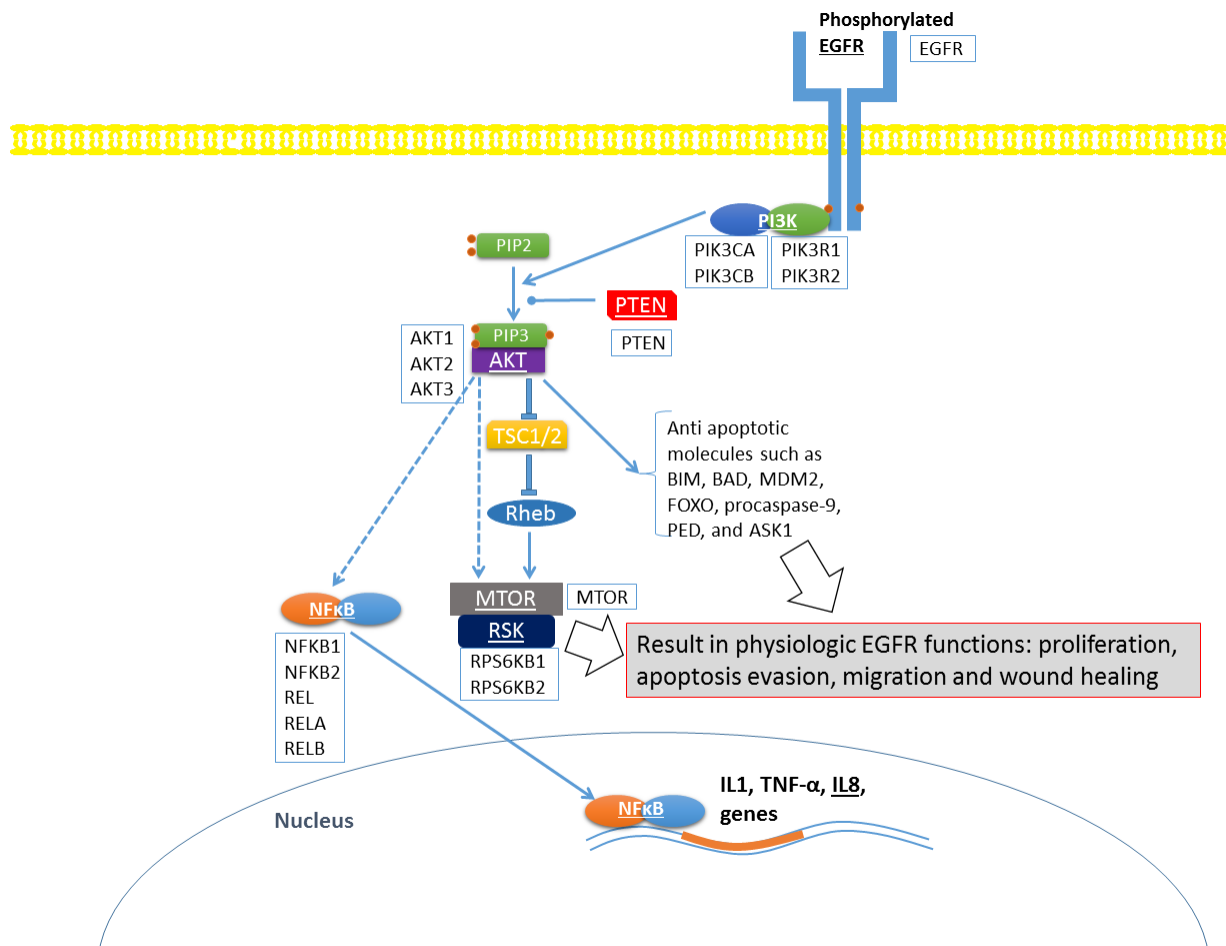


Figure 1.4.2 PI3K-Akt pathway. The underlined molecules are investigated in current study. The large bright-blue arrow indicates that the double inhibition finally results in an activation of the mTOR pathway. IRS-1, Insulin receptor substrate 1; PI3K, phosphatidylinositol 3-kinase; PIP2, phosphatidylinositol biphosphate; AKT, v-akt murine thymoma viral oncogene homolog; TSC, tuberous sclerosis complex; Rheb, Ras homolog enriched in brain; MTOR, mechanistic target of rapamycin. All the factors which have been resequenced in the present study are underlined. Names of the genes analyzed corresponding to each protein is shown in boxes.

Akt protein – a serine/threonine kinase – translocates to the membrane and binds to PIP3 molecule. This positioning then enables the phosphatidylinositol-dependent kinase 1 (PDK-1) to phosphorylate Akt. Phosphorylated Akt (pAkt) directly phosphorylates proteins involved in

avoiding apoptosis, survival and cell growth signaling, including BIM, BAD, MDM2, FOXO, procaspase-9, PED (an inhibitor of caspase 3), ASK1, PRAS40 and TSC2 (Bellacosa et al. 2005, Franke 2008). Akt is involved in several cellular processes via interacting with mechanistic target of rapamycin (mTOR) and NFκB complexes.

mTOR is a protein kinase of PI3K related kinase family. This kinase plays its physiological roles via interactions with other molecules in forms of two complexes named mTOR complex 1 (mTORC1) and 2 (mTORC2) (Laplante and Sabatini 2012). Akt, ERK, and RSK1 kinases directly phosphorylate TSC1/TSC2 proteins to inactivate them. TSC1/2 are GTPase activating proteins that convert the Ras homolog enriched in brain (Rheb) GTPase from GTP-bound state to GDP-bound state. Inhibition of TSC proteins results in an increase of GTP-bound form of Rheb and subsequently induces the mTORC1's kinase activity. The process in which mTORC2 is involved is less understood – specially its upstream activation –; however, it is known that it directly stimulates Akt by phosphorylating it. Activation of Akt results in evading apoptosis and cell growth via mTOR complexes. Major role of mTORC1 is controlling the protein synthesis. mTORC1 serves this role by phosphorylates the translational regulators eukaryotic translation initiation factor 4E (eIF4E) and S6 kinase 1 (S6K1 a member of RSK family proteins) (Laplante and Sabatini 2012).

#### 1.4.3 Inflammatory genes involved

The transcription factor NFκB is known to stimulate expression of proinflammatory genes and is shown to be upregulated after EGFR inhibition (Woodworth et al. 2005). NFκB or “Nuclear factor kappa-light-chain-enhancer of activated B cells” super family of transcription factors in mammals includes the following five transcription factors: p50 (NF-κB1/p105), p52 (NF-κB2/9100), RelA (p65), RelB and c-Rel, which perform their function in dimer form. Homo or hetero dimers of NF-κB transcription factor recognize κB-sites – 9 to 10 base pair DNA regulatory sites – and as a result stimulate the specific target gene expression (Gilmore and Herscovitch 2006, Hoffmann et al. 2006, Moynagh 2005).

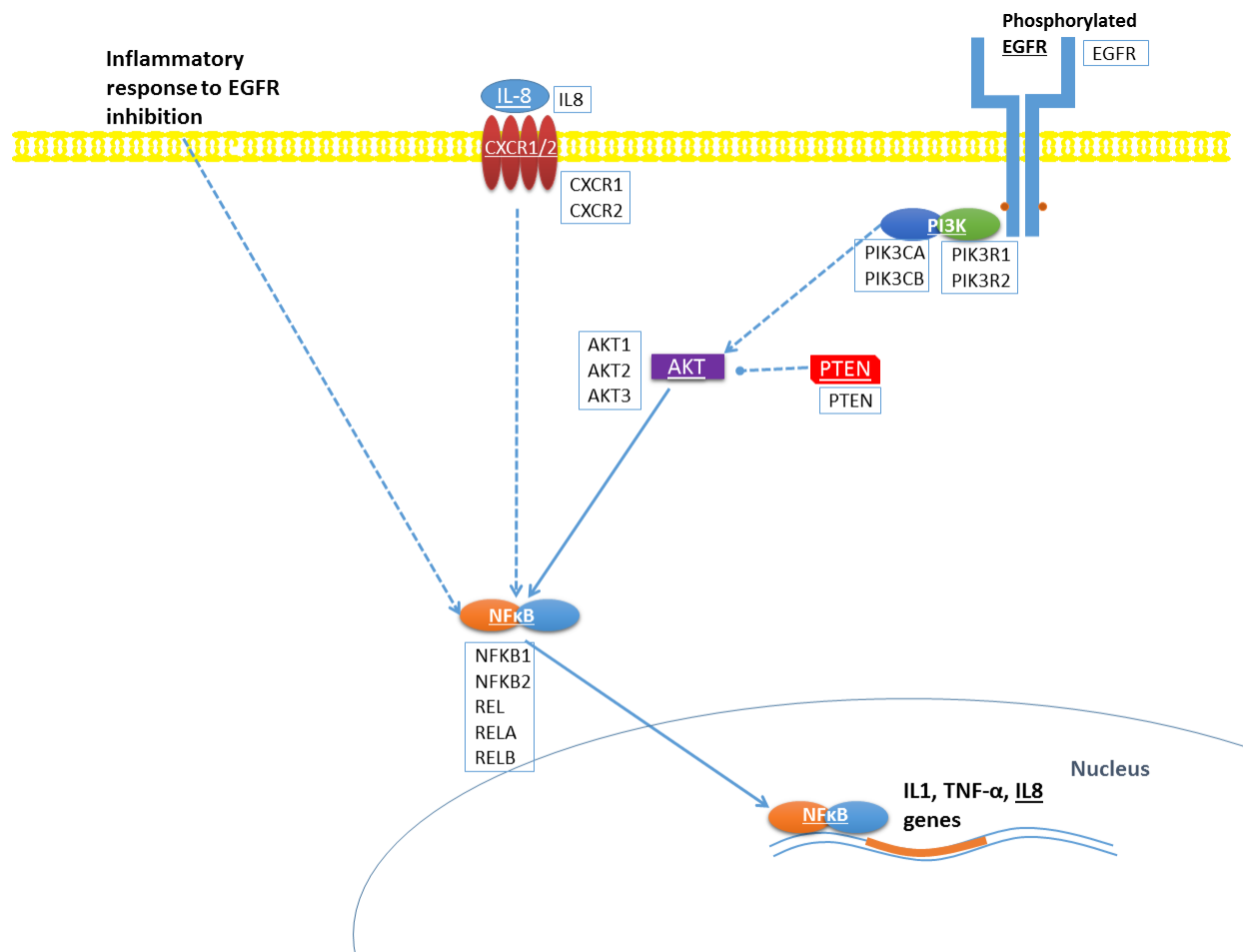


Figure 1.4.3 Inflammatory genes involved in pathogenesis of rash. NFκB, Nuclear factor kappa B, IL8, Interleukin 8; CXCR, CXC chemokine receptor (IL8 receptors); IL1, interleukin 1; TNF-α, tumor necrosis factor alpha. All the factors which have been resequenced in the present study are underlined. Names of the genes analyzed corresponding to each protein is shown in boxes.

Furthermore, in 2012 Bangsgaard's group surveyed the hypothesis that neutrophilic infiltration in rash affected areas is attracted by IL-8. Firstly they induced rash by local injection of zalutumumab (an EGFR inhibitor). Afterwards they applied a human IL-8 neutralizing antibody (HuMab-10F8) locally and reduced the rash. Therefore they concluded that antibodies neutralizing IL-8 may effectively reduce and reverse the rash development. IL8 transcription is also known to be mediated by NFκB (Bangsgaard et al. 2012) (Figure 1.4.3). An overview of stated genes which are involved in downstream signaling of EGFR and hypothetically are responsible for rash development is displayed in Figure 1.4.4.

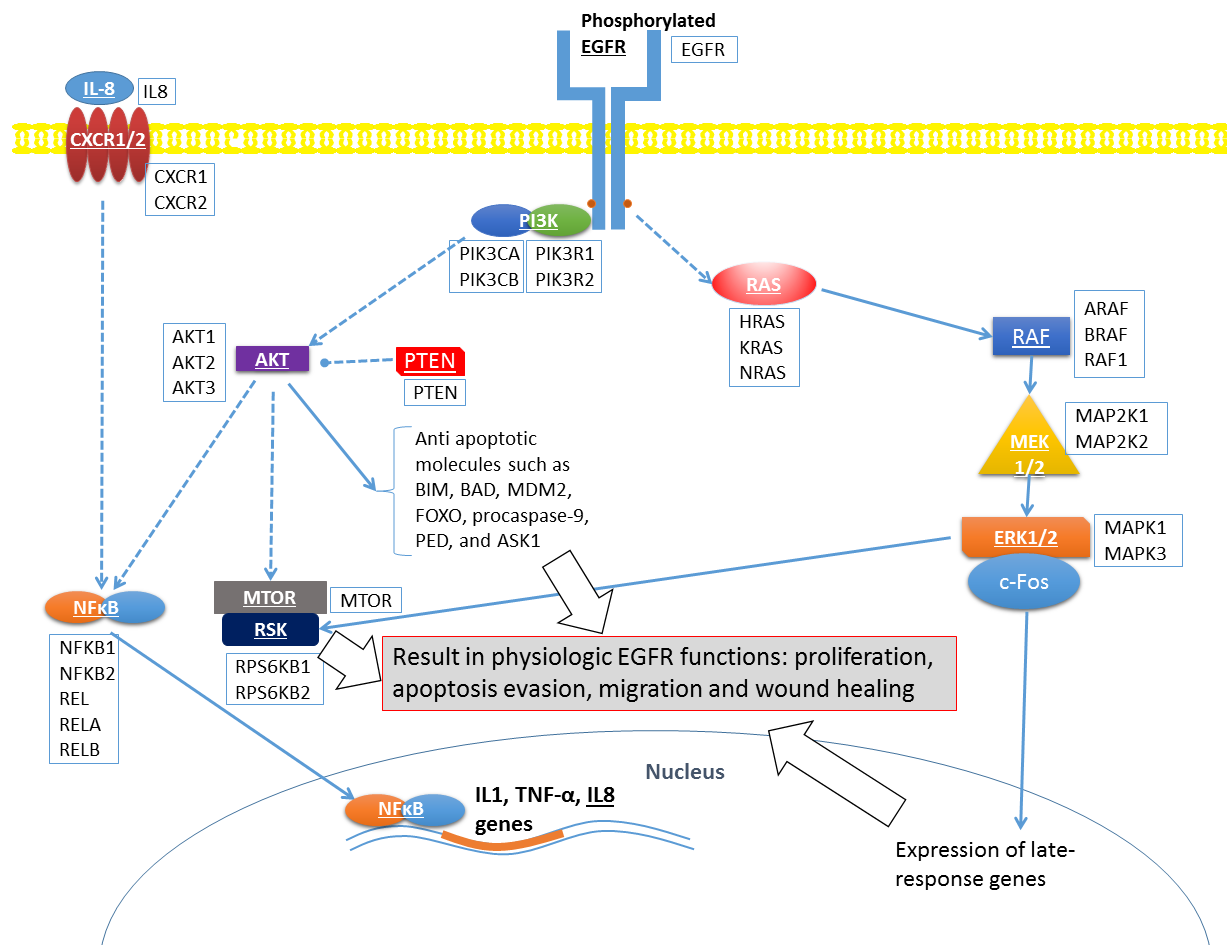


Figure 1.4.4 Overview of EGFR downstream signaling. Names of the genes analyzed corresponding to each protein is shown in boxes.

### 1.5 Known facts about genomic variation in relation to skin side effects of EGFR blocking

Skin toxicity is believed to be a class side effect of EGFRIs, meaning that it can be caused by EGFR inhibition independent of the drug type (small molecule TKIs or monoclonal antibodies). The skin toxicity does not occur in all patients. This fact leaves germline genetic variations in *EGFR* and its downstream signaling pathway as the primary suspects for this difference. Several studies have tested this hypothesis.

In 2009 Klinghammer et al. studied *EGFR* gene polymorphisms in 51 patients with recurrent or metastatic head and neck squamous cell carcinoma (HNSCC) in a single arm phase II study of cetuximab/docetaxel. Forty one percent of their patients had skin toxicity as a treatment side effect. Two variations in *EGFR* gene were analyzed in relation to skin toxicity in their study. Firstly a variation of arginine to lysine in the 521th amino acid position of *EGFR* gene (rs2227983 also known as *EGFR* R497K): allele G to A change at DNA level causes the amino acid change. The G/G genotype was significantly associated with appearance of skin rash in these patients. They have dichotomized patients as follows: patients with rash level 0 and 1 in one group and patients with level 2 and 3 in another group. A dominant model (having or not having a minor allele) categorization was used for analysis. Moreover, a CA microsatellite repeat polymorphism (rs11568315) in intron 1 of *EGFR* gene was surveyed in this study. They have analyzed the correlation between a short (less than or equal to 16) or long (more than 16) CA repeat in the patients. No significant association was found between rs11568315 and skin rash (Klinghammer et al. 2010).

However this association was not confirmed in several other studies. In 2010, Giovannetti et al. studied the correlation between *EGFR* or *AKT1* polymorphisms and skin toxicity in 96 NSCLC patients treated with Gefitinib. The *EGFR* -191 C/A, *EGFR* -216 G/T, *EGFR* R497K, and rs11568315 CA repeat polymorphisms were analyzed, as well as the *AKT1* polymorphisms rs3730350 and rs1130233 were analyzed. They also categorized the patients according to skin toxicity into grade 0-1 and grade 2-3 groups. None of the other studied polymorphisms were significantly associated to skin rash (Giovannetti et al. 2010). In 2007, Li et al. surveyed the *EGFR* polymorphisms -191 C/A, -216 G/T, R497K, and rs11568315 CA repeat in NSCLC patient treated with Gefitinib. The -216 G/T was significantly associated with presence of skin toxicity of grade equal to or more than 1. However, the *EGFR* -191C/A, intron 1 CA dinucleotide repeat polymorphism, or the R497K SNPs had no association with skin rash manifestation (Liu et al. 2008).

In this regard also McKibbin et al. analyzed the *EGFR* polymorphisms in combined data from five pediatric clinical trials with gefitinib. They have found an association between *EGFR* -216 G/T and grade 1 or more skin rash in a recessive model. Intron 1 CA dinucleotide repeat polymorphism, the -191C/A, or R497K did not have a significant association with skin rash (McKibbin et al. 2010).

Another group has surveyed skin toxicity in 52 patients receiving Gefitinib as a first line therapy of NSCLC. They have found an association between short CA repeat alleles and grade 2-3 skin rash. Other 3 frequently studied EGFR polymorphisms (-191 C/A, -216 G/T, R497K) had no significant association to skin rash (Huang et al. 2009). In another study on 175 NSCLC patients receiving gefitinib, EGFR -191 C/A, EGFR -216 G/T, and rs11568315 CA repeat were not found to be significantly associated with skin rash (Gregorc et al. 2008).

EGFR polymorphisms were also analyzed in erlotinib therapies. In a prospective study of 80 NSCLC patients, Rudin et al. have studied EGFR polymorphism in relation to erlotinib skin toxicity. None of EGFR -191 C/A, EGFR -216 G/T, EGFR R497K, and rs11568315 CA repeat were associated with skin rash. They have as well studies a number of SNPs in CYP3A4, CYP3A5, ABCG2 genes (refer to Table 1.5.1). Only ABCG2 16702 G/G genotype was associated with skin rash (Rudin et al. 2008).

EGFR R497K or rs2227983 is an arginine to lysine substitution at the 521th amino acid of EGFR protein (at the time of this functional study they believed that the substitution is at 497 position therefore historically this variation is called R497K). To identify the molecular mechanism in which the rs2227983 is involved, Moriai et al. expressed the wildtype and variant EGFR in Chinese hamster ovary (CHO) cells and compared the binding of EGF and TGF- $\alpha$ . Both receptors have shown two (a high and a low) affinity sites for EGF; however, the variant receptor had no high affinity site for TGF- $\alpha$  but only a low affinity site for TGF- $\alpha$ . The authors also displayed that the cells expressing EGFR R497K had decreased growth response (investigated by MTT assay) to EGF and TGF- $\alpha$ ; and decreased stimulation of FOS, JUN and Myc. Therefore, they hypothesized that substitution of this residue in an extracellular domain modulates the phosphorylation of the intracellular domain of EGFR (Moriai et al. 1994).

Another group investigated the impact of EGFR R497K on EGFR phosphorylation and Myc activation in colorectal cancer tissue. They stained the tumor tissue segments with anti-phosphorylated EGFR and anti-c-Myc antibodies. They have observed a marked decrease in EGFR phosphorylation and c-Myc activation associated with the G/A or A/A genotypes (Wang et al. 2007).

Table 1.5.1 Summary of polymorphisms studied in EGFR gene by different studies and their association with skin rash. Many of the found associations could not be reproduced.

Gene name	Variation functional localization	dbSNP ID (other used identifier)	Results	Study
<b>EGFR</b>	Promoter region	rs712830 (-191 C/A)	No difference	(Giovannetti et al. 2010) (Liu et al. 2008) (McKibbin et al. 2010) (Huang et al. 2009) (Gregorc et al. 2008) (Rudin et al. 2008)
<b>EGFR</b>	Promoter region	rs712829 (-216 G/T)	T allele was associated with skin rash	(Liu et al. 2008)
<b>EGFR</b>	Promoter region	rs712829 (-216 G/T)	T/T genotype associated with skin rash	(McKibbin et al. 2010)
<b>EGFR</b>	Promoter region	rs712829 (-216 G/T)	No difference	(Giovannetti et al. 2010) (Huang et al. 2009) (Gregorc et al. 2008) (Rudin et al. 2008)
<b>EGFR</b>	Exon 13	rs2227983 (R497K)	G/G genotype associated with skin rash	(Klinghammer et al. 2010)
<b>EGFR</b>	Exon 13	rs2227983 (R497K)	A/A genotype associated with skin rash (G/A genotype with absence of skin rash)	(Parmar et al. 2013)
<b>EGFR</b>	Exon 13	rs2227983 (R497K)	No difference	(Giovannetti et al. 2010) (Liu et al. 2008) (McKibbin et al. 2010) (Huang et al. 2009) (Gregorc et al. 2008) (Rudin et al. 2008)
<b>EGFR</b>	Intron 1	rs11568315 (intron 1 CA repeat)	Shorter alleles were associated with skin rash	(Huang et al. 2009)
<b>EGFR</b>	Intron 1	rs11568315 (intron 1 CA repeat)	No difference	(Klinghammer et al. 2010) (Giovannetti et al. 2010) (McKibbin et al. 2010) (Gregorc et al. 2008) (Rudin et al. 2008)
<b>AKT1</b>	Gene upstream	rs3730350	No difference	(Giovannetti et al. 2010)
<b>AKT1</b>	Exon 9	rs1130233	No difference	(Giovannetti et al. 2010)



Gene name	Variation functional localization	dbSNP ID (other used identifier)	Results	Study
<b>AKT1</b>	Intron 5	rs3001371	No difference	(Parmar et al. 2013)
<b>AKT1</b>	Intron 3	rs2498794	No difference	(Parmar et al. 2013)
<b>AKT1</b>	Exon 1	rs1130214	No difference	(Parmar et al. 2013)
<b>CYP3A4</b>	Promoter region	rs2740574 (CYP3A4*1B)	No difference	(Rudin et al. 2008)
<b>CYP3A5</b>	Intron 3	rs776746 (CYP3A5*3)	No difference	(Rudin et al. 2008)
<b>ABCG2</b>	Exon 5	rs2231142 (421 C/A)	No difference	(Rudin et al. 2008)
<b>ABCG2</b>	Exon 2	rs2231137 (34 G/A)	No difference	(Rudin et al. 2008)
<b>ABCG2</b>	Intron 1	rs2622604 (1143 C/T)	No difference	(Rudin et al. 2008)
<b>ABCG2</b>	Intron 13	No_rs (16702 G/A)	G/G genotype was associated with skin rash	(Rudin et al. 2008)
<b>ABCG2</b>	Promoter region	No_rs (-15944 G/A)	No difference	(Rudin et al. 2008)
<b>ABCG2</b>	Promoter region	No_rs (-15622 C/T)	No difference	(Rudin et al. 2008)
<b>PIK3CA</b>	Gene upstream	rs9831234	No difference	(Parmar et al. 2013)
<b>PIK3CA</b>	Intron 1	rs2699905	No difference	(Parmar et al. 2013)
<b>PIK3CA</b>	Intron 1	rs6443624	No difference	(Parmar et al. 2013)
<b>PIK3CA</b>	Intron 1	rs2677760	No difference	(Parmar et al. 2013)
<b>PIK3CA</b>	Intron 9	rs2459693	T/C genotype was associated with skin rash	(Parmar et al. 2013)

No\_rs: no rs number assigned.

In 2011 Parmar *et al.* published the analysis of a number of *EGFR*, *AKT1*, and *PIK3CA* polymorphisms (refer to Table 1.5.1) in the same sample population as the present study in

association with skin rash. They analyzed EGFR -191 C/A, EGFR -216 G/T, EGFR R497K, and rs11568315 CA repeat and found an association between the R497K polymorphism with skin rash. Except for *PIK3CA* rs2459693, which T/C genotype was associated with skin rash; other polymorphisms were not significantly associated with skin rash (Parmar et al. 2013).

As summarized in the Table 1.5.1, few studies could reproduce the presumed consequence of EGFR polymorphisms on skin toxicity manifestation. The functional in vitro effects were not reproduced in the clinical studies. As a conclusion from the introductory review, the data regarding the effect of polymorphisms in the EGFR pathway on skin rash are very controversial and a systematic study to analyze a wider spectrum of genetic variations is needed.

## 1.6 Genetic variations and genetic testing

### 1.6.1 Human genome and its variations

DNA variations are responsible for many variations observed among human populations such as disease susceptibility and the propensity to positive or negative effects of xenobiotics including drugs.

The most common human genetic variation is the single nucleotide polymorphism (SNP) that accounts for 88% of isolated genetic variations (Levy et al. 2007). SNP is referred to as a single-nucleotide variation in a genetic sequence; a common form of variation in the human genome (Feero et al. 2010). These variations occur commonly within a population (e.g. at a rate of 1%). SNPs occur on average at a rate of 5 in 1000 bp and according to recent studies about 15 million of them are found (Genomes Project et al. 2010, Mills et al. 2011). An average individual is estimated to harbor 3.3 million SNPs (Shen et al. 2013).

It is probably necessary to define some genetic terms for the better understanding of the following text. An allele is “one of two or more versions of a genetic sequence at a particular location in the genome” (Feero et al. 2010). Haplotype is “a set of DNA variations, or polymorphisms, that tend to be inherited together. A haplotype can refer to a combination of alleles or to a set of single-nucleotide polymorphisms found on the same chromosome” (Feero

et al. 2010). Haplotypes can reside in one locus, or several loci. Using haplotypes has made it possible to reduce the number of SNPs tested for examining the genome; as a SNP (tagging SNP) can be representative of other SNPs in the same haplotype.

Genetic variations can occur in form of loss or gain of genomic material from 1 kilobase to several megabases in size, by deletions, duplications, inversions, and translocations. Copy number variation (CNV) is “a common variation in this category. CNV is defined as the variation from one person to the next in the number of copies of a particular gene or DNA sequence” (Feero et al. 2010). By the completion of the Human Genome Project, CNV was identified as a widespread and common phenomenon among humans (Sebat et al. 2004). Approximately 0.4% of the genome of unrelated people differ typically with respect to copy number (Kidd et al. 2008).

### 1.6.2 Importance of genetic testing

Genetic variations may make the individuals prone to acquired diseases. Among these diseases, cancers are of bold significance. Genetic variations cause a susceptibility to carcinogens in the environment. Many of these variations affect drug metabolizing enzymes and therefore cause various responses to drugs or xenobiotics. Examples are mentioned in following text.

#### 1.6.2.1 *Kras variations*

Proto-oncogene Ras is a GTPase protein which, upon activation, stimulates many cellular functions especially cell growth and proliferation. In normal situations Ras function is strongly regulated. Ras is frequently mutated in human cancers. As an example, Kras mutations are found in 90% of pancreatic tumors. Many single mutations have been found in *RAS* gene at codons 12, 13 or 61 which contribute to hyperproliferation in tumors (Quinlan and Settleman 2009). These mutations result in ever-active state of this protein even without upstream stimulations. The genetic status of Ras is greatly important for prediction of drug response and survival (Prior et al. 2012).

### 1.6.2.2 Variations in metabolic enzymes

The metabolism of xenobiotics or foreign biologic substances is vital for human body for detoxifying and excretion of these substrates. This metabolism is commonly categorized in three phases: modification, conjugation, and excretion (Omiecinski et al. 2011). Phase I or also known as modification is the addition of functional groups such as -OH, -SH, -NH<sub>2</sub> or -COOH to xenobiotics. In Phase II (conjugation), xenobiotics are conjugated with charged species such as glutathione, sulfate, glycine, or glucuronic acid. Glutathione S-transferases (GSTs) as an example of phase II enzymes are involved in inactivation of oxidative metabolites or carcinogenic compounds. Moreover they may play a role in activation of DNA repair enzymes (Pinarbasi et al. 2005). *GSTM1* and *GSTT1* polymorphisms have been reported to be associated with lung cancer (Hosgood et al. 2007) and prostate cancer (Ashtiani et al. 2011). Finally in phase III (excretion), some xenobiotics may undergo further modification and be excreted. One example for phase III metabolism is the function of membrane transporters to shuttle drugs and other xenobiotics across cellular membranes (Omiecinski et al. 2011).

### 1.6.2.3 Glucose 6 phosphate dehydrogenase

The deficiency of this enzyme was found when 10% of African American soldiers were presented with hemolysis after treatment with primaquine (an antimalarial drug). This enzyme is responsible for performing a reaction which produces NADPH co-enzyme (a reducing agent). NADPH provides the cells with reduced glutathione which in turn clears the free radicals that cause oxidative damage. This oxidative damage causes hemolysis in red blood cells. Decreased glutathione level makes the cells prone to hemolysis in case of administration of certain drugs such as several antimalarial drugs (Schuurman et al. 2009). Nowadays G6PD deficiency or Favism disease is a well know disorder and to date more than 150 causal mutations have been identified (Brockmöller and Tzvetkov 2008).

### 1.6.3 Development of methods for analysis of human genomic variations

Considering the described roles of genetic variations, including their possible roles in contributing to disease prevention, and individualized and optimized choice of drug, the analysis of human genome is substantially vital.

Methods in genotype analysis have undergone huge advancements in recent 30 years. Genotyping by using the restriction site of endonuclease enzymes was historically one of the first methods used. Restriction endonucleases identify a specific sequence and cut DNA at this site. This specificity had been used to identify the genetic variations. Restriction activity has been reported in 1952 for the first time (Luria and Human 1952). Later on it has been shown to be promising in identifying human beta-globin protein genetic variations (Kan and Dozy 1978).

Elucidation of the entire human genome strongly depended on methods to reveal the DNA sequence. One of the first methods was the Maxam-Gilbert method which was not widely used later in DNA sequencing because of its complexity (Maxam and Gilbert 1977). However this method has led to development of Methylation Interference Assay for studying DNA protein interaction (Carey and Smale 2007). In 1977 Sanger et al. reported a new method for determining nucleotide sequence. Sanger and colleagues used dideoxynucleoside triphosphates for termination of polymerase chain reaction (PCR) and then categorized resulting segments according to which of the dideoxynucleoside triphosphates were used (ddATP, ddGTP, ddCTP, or ddTTP) (along with three normal nucleosides). Segments were then typically visualized on 12% acrylamide gels with autoradiography. As a result considering the order of segments by size, determining the exact nucleotide sequence was facilitated (Sanger et al. 1977). This method has been the most widely used method for analysis of genetic variation for 25 years after its invention.

These approaches have made it possible to test the candidate gene variations in association with the specified diseases. Despite the efforts and resources consumed on this approach in the past, only small number of associations has been confirmed. Consequently, the molecular genetic basis of common diseases is still poorly understood (Hirschhorn et al. 2002). In recent years using genome wide association studies (GWAS), the entire genome was scrutinized to find associations

with diseases. GWAS have successfully found over 4000 common variants in which had associations with 200 common complex diseases (Johnson and O'Donnell 2009, Manolio et al. 2008, Visscher et al. 2012, Wellcome Trust Case Control 2007). These variants could have never been identified using the candidate-gene approach, as even several were placed in some gene-depleted regions of genome. But still for most common chronic diseases the known variants do explain only a minor fraction of the variation in disease risk between individuals. One of the exceptions is the association of a common variation in the complement factor H gene (*CFH*) with age-related macular degeneration (Klein et al. 2005). Other successful examples are locus 9p21.3 variants and coronary artery disease and locus 8q24 with cancer (Wellcome Trust Case Control 2007). Up to the present time hundreds of GWAS were performed. Although many were successful; still they do not address a large fraction of disease burden of diseases in population (Schork et al. 2009).

As the GWAS analyze the whole genome regardless of any hypothesis which assumes a pathophysiologic role of specific genes in a common disorder, they are also called “hypothesis-free”. This approach has the advantage of finding novel underlying mechanisms for diseases. Most GWAS analyze the SNPs with the frequency of more than 5% and therefore identifying rare variants is not possible. Consequently the application of GWAS is restricted to discovery of common variants in association to diseases, which fits into the common disease/common variants (CDCV) hypothesis (Lander 1996). On the other hand, most functional or deleterious variants are reported to have a frequency of <1% (Kitsios and Zintzaras 2009). This makes the GWAS to be unable to identify rare variants and make them inappropriate for testing Common disease/rare variants (CDRV) hypotheses according to be common disease are caused by multiple rare variants. Evidence on CDRV hypothesis comes from studies investigating diseases such as cystic fibrosis and BRCA1/2 breast cancer, in which hundreds of rare but disease causing variations have been found (Bobadilla et al. 2002, Iyengar and Elston 2007).

After introduction of next generation sequencing also termed as deep sequencing or massively parallel sequencing, analysis of rare variants was made possible. This technology has made DNA sequencing fast, cost-effective for large targets, and more accurate (e.g. in case of finding rare somatic variations or viral mutations). There are different technologies which the two most

frequently used ones are sequencing-by-synthesis and semi-conductor based sequencing. These technologies facilitated a broad spectrum of applications which can be categorized into DNA, RNA and epigenetic applications.

DNA sequencing applications include whole genome sequencing (WGS), whole exome sequencing, and targeted sequencing. WGS has made it possible to identify every variation in a patient's genome resulting in a detailed catalog of rare genetic variants. As an example 1000 genomes project has largely contributed to this catalog (Genomes Project et al. 2010, Kaiser 2008). However WGS seems to be a very powerful tool for analysis of genomic variations; it is still very expensive and unaffordable to be used for large samples. Whole exome sequencing on the other hand is a more affordable approach which covers all functional variations. Studies using exome sequencing have so far identified causal mutations in several cancers and genetic diseases (Agrawal et al. 2011, Wu et al. 2012, Yan et al. 2011). A more flexible approach is targeted re-sequencing which not only can be used as a primary approach but also can be used as a confirmatory study with larger sample size for variants identified by whole genome sequencing (Mamanova et al. 2010).

#### 1.6.4 Massively parallel sequencing or deep sequencing

High-throughput sequencing technologies have made the sequencing of genome faster, affordable and even more precise. While to this date several high-throughput technologies are introduced, the common quality of all these methods is parallelizing the sequencing reactions. Three platforms for deep sequencing are commonly used at present: Roche/454, Illumina, and Ion torrent (Mardis 2013). In the following text the sequencing-by-synthesis (Illumina <sup>TM</sup>) and semi-conductor based (Ion torrent <sup>TM</sup>) sequencing will be explained.

##### 1.6.4.1 Sequencing-by-synthesis or Illumina sequencing

In sequencing-by-synthesis also termed as Illumina <sup>TM</sup> sequencing approach genomic DNA is fragmented and then ligated to specific adaptors. In the next step a flow cell – a glass

microfabricated device – which allows the genomic DNA fragments to be amplified on its surface is used. This amplification is called bridge amplification. In this process single stranded adaptor ligated DNA is attached to the surface of the flow cell by the adaptor complementary sequences which are already fixed on the surface. Then one cycle of the amplification will be performed using this complementary sequence. The DNA is afterwards denatured and each strand which is attached to a complimentary fixed sequence is straightened (does not have the bridge form). The free adaptor sequence of each fragment is then attached to the complementary sequence fixed on the flow cell surface that forms a bridged DNA molecule. Several cycles are performed to amplify the template molecule.

For sequencing, in each flow the machine adds all four nucleotides simultaneously to the flow cell along with the DNA polymerase. The nucleotides contain a fluorescent label which is unique for each base and a 3'-OH blocking group. The blocking group prevents the addition of more than one nucleotide per reaction. After addition of each base an imaging process will be performed to record that which nucleotide is added in the performed reaction. Subsequently, the 3'-OH blocking group including the fluorescent dye is chemically removed, that allows another nucleotide to be attached in the next sequencing flow. Finally the data will be analyzed with a base-calling algorithm which assigns sequences and corresponding quality values to each read (Bentley et al. 2008).

#### *1.6.4.2 Semi-conductor based or Ion torrent™ sequencing*

In this platform which is explained in details in the methods section, DNA fragmentation of whole genomic DNA is performed as the first step. The fragments then will be ligated to specific adaptor. These adaptors assist the amplification of genomic DNA and also allow the user to barcode different DNA samples from different subjects or tissues. Then an amplification step will be performed using emulsion PCR. In this process a DNA molecule is amplified inside an oil microvesicle in attachment to a specific beads or spheres (Ion Sphere Particle™, ISP). These ISPs already contain specific sequences complimentary to adaptor sequences and the primers used



for amplification at this step. Afterwards sequencing primers are added to ISPs along with the DNA polymerase.

The ISPs are then loaded on a sequencing chip. A sequencing chip contains a million or more of microscopic wells depending on the chip type. These wells are the units of sequencing reaction. Each ISP is then seeded in each well. Nucleotides are then added one by one to the sequencing chip and if they are complementary to the template a PH change will occur. The chip senses this PH change and records it in association with the type of added nucleotide. Then this data will be analyzed (base-calling and quality assessment by the sever computer attached to the sequencing machine (Mardis 2013).

## 1.7 Hypotheses and a short work plan of the thesis

As described above there is a variation in the individual propensity to EGFR inhibitor induced skin rash and it appears to be very likely that genetic variations in proteins from the EGFR-signaling pathway play a role in this variation. Besides the EGFR pathway the EGFR induced skin reaction is apparently highly inflammatory and genetic variations in cytokines, chemokines and their receptors may also be relevant.

We had the possibility to study 126 samples from patients who had been treated with EGFR inhibitors and in whom the adverse effects were characterized in detail. To comprehensively analyze genomic variations we have used deep sequencing on Ion Torrent™ system.

Following tasks were essential to the workplan:

- Definition of the most plausible candidate genes based on a molecular review of the EGFR pathway and based on prior data on genomic variants associated with skin rash
- Bioinformatic selection of the genomic segments to be re-sequenced
- Preparation of DNA samples to be compatible with DNA sequencing
- Performing re-sequencing on the Ion Torrent™ system
- Bioinformatic filtering of reliable versus unreliable variants

- Bioinformatic analysis of the reliable variants concerning their presumed functionality
- Statistical analysis of the variants found in relation to skin rash and as additional exploratory analysis also in relation to diarrhea (another frequent side effect) and in relation to overall survival.

This project should contribute to a better understanding of EGFR induced skin rash and possibly to the development of molecular biomarkers allowing to select individually efficient but better tolerated therapies and doses.

## 2 Methods

### 2.1 Clinical study design and definition of the phenotypes

To analyze the genetic variation in EGFR pathway, patients receiving EGFR inhibitors were enrolled to this study. Before the enrollment, written informed consent concerning pharmacogenomics analysis was obtained from all participating subjects. Inclusion criteria were histologically confirmed cancer disease and first-time treatment with an EGFR inhibitor (either monoclonal antibody or small molecule TKI). Grade of skin rash was evaluated using National Cancer Institute Common Terminology Criteria for Adverse Events, version 3.0 (NCI CTCAE v3.0). This criteria is shown in Table 1.6.4.1. The study protocol was verified by the ethical review boards of the University Ulm and the LMU Munich. Patients were recruited from the stated medical centers.

Table 2.1.1 Grading of rash according to toxicity criteria of the National Cancer Institute (NCI CTCAE) version 3.0.

Grade	Features
1	Macular or papular eruption or erythema without associated symptoms
2	Macular or papular eruption or erythema with pruritus or other associated symptoms; localized desquamation or other lesions covering <50% of body surface area (BSA)
3	Severe, generalized erythroderma or macular, papular or vesicular eruption; desquamation covering ≥50% BSA
4	Generalized exfoliative, ulcerative, or bullous dermatitis

Overall survival (OS) was described as the duration (in days) from treatment start to the date of death. Alive patients at the last available follow-up were censored at the last date they were known to be alive. Follow up was performed for the duration of 360 days.

### 2.2 Target region criteria

Thirty genes were targeted for resequencing. The whole *EGFR* gene along with 2000 bp of its promoter region was one of the targets. The total size of these regions was 196300 base pairs. For the remaining 29 other responsible genes, the exons along with 500 bp of promoter region

were selected. These 29 genes are namely: AKT (*AKT1*, *AKT2*, *AKT3*), RAF (*ARAF*, *BRAF*, *RAF1*), RAS (*HRAS*, *KRAS*, *NRAS*), ERK (*MAPK3*, *MAPK1*), MEK (*MAP2K1*, *MAP2K2*), MTOR (*MTOR*) NFKB (*NFKB1*, *NFKB2*, *REL*, *RELA*, *RELB*), S6K (*RPS6KB1*, *RPS6KB1*), PI3K (*PIK3CA*, *PIK3CB*, *PIK3R1*, *PIK3R2*), PTEN (*PTEN*), Interleukin 8 (*IL8*) and its receptors (*CXCR1*, *CXCR2*) (Table 2.2.1). Total size of the sequencing target is 382400 basepairs.

Table 2.2.1 The selected genes and their corresponding size. The entire EGFR gene along with 2000 bp of its promoter region (total 196300 bp) was also sequenced which is not shown in this table.

Functional Pathway	Gene name	Region size in bp	Gene name	Region size in bp	Gene name	Region size in bp
PI3K/Akt	<i>AKT1</i>	8423	<i>PIK3CA</i>	6201	<i>PTEN</i>	7106
	<i>AKT2</i>	10291	<i>PIK3CB</i>	7060	<i>RPS6KB1</i>	7916
	<i>AKT3</i>	9384	<i>PIK3R1</i>	12762	<i>RPS6KB2</i>	3996
	<i>MTOR</i>	16138	<i>PIK3R2</i>	8439		
Ras/MAPK	<i>ARAF</i>	4444	<i>MAP2K1</i>	4702	<i>NRAS</i>	5554
	<i>BRAF</i>	5146	<i>MAP2K2</i>	4419	<i>RAF1</i>	5825
	<i>HRAS</i>	2228	<i>MAPK1</i>	7330		
	<i>KRAS</i>	7873	<i>MAPK3</i>	4112		
Inflammatory response regulators	<i>NFKB1</i>	8640	<i>REL</i>	4085	<i>RELB</i>	3879
	<i>NFKB2</i>	6538	<i>RELA</i>	4897		
Cytokines and their receptors	<i>CXCR1</i>	3503	<i>CXCR2</i>	4718	<i>IL8</i>	2505

### 2.3 Overview of sample preparation for sequencing

Three major steps will be explained in the following text including: library preparation, target enrichment and template preparation. In the library preparation step the aim is to prepare a size selected DNA library from each patient and as well to barcode the DNA from different study participants with different barcode adaptors so that the libraries would be discriminated by the sequencing machine. The goal of the target enrichment step is to isolate only the sequences which are chosen (targeted) to be sequenced from the gDNA. Finally the prepared sample must be attached to an Ion Sphere Particle <sup>TM</sup> (ISP) and must be amplified in a way that a single molecule of DNA attaches to one ISP. The attached molecule will be then amplified. Figure 2.3.1 displays the mentioned steps.

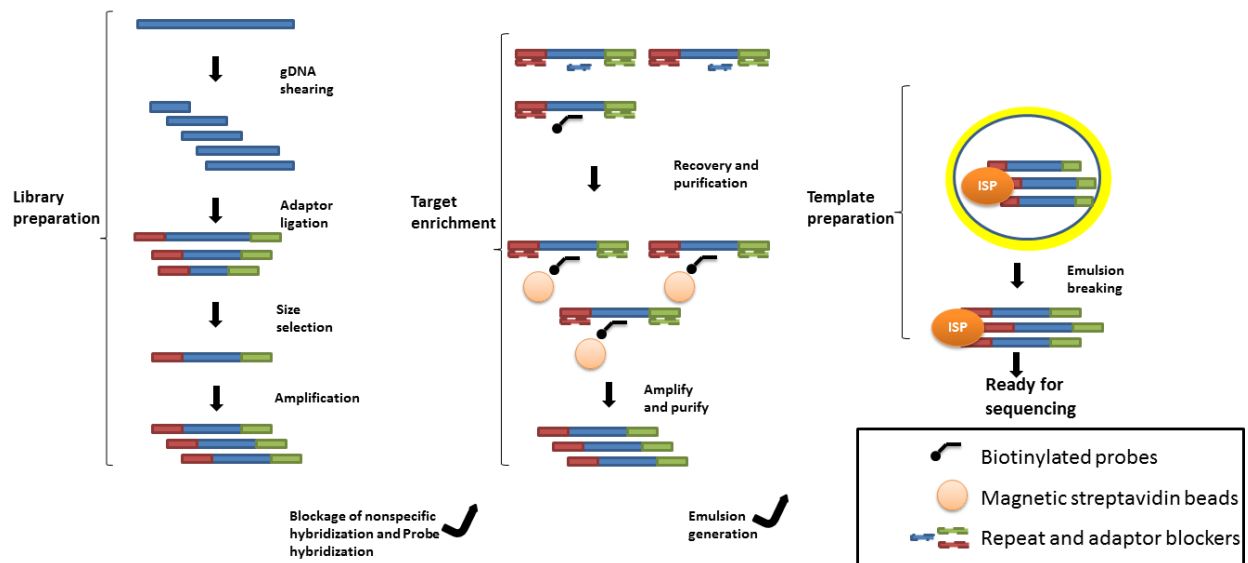


Figure 2.3.1 Overview of sample preparation steps prior to sequencing. During these steps gDNA will be sheared and the fragments are then ligated to one barcode adaptor and one P1 adaptor. Subsequently fragments with specific size (around 350 bp) are selected. After PCR amplification, the targeted regions are then isolated using specific biotinylated probes (shown in black). To evade and block the nonspecific hybridization between adaptor sequences and between genomic repeat regions, adaptor blockers and Human Cot1 DNA are added. Each DNA molecule is then attached to an Ion Sphere Particle (ISP) and an amplification procedure will be performed inside an oil microvesicle. The oil vesicle is then broken and amplified DNA attached to ISP is the result. Figure adopted from Ion Xpress™ Plus Fragment Library Preparation User Guide.

## 2.4 DNA isolation and Library preparation

In order to genotype the gDNA from our sample population, gDNA had to undergo several steps to finally be ready to be analyzed by the Ion Torrent™ sequencing machine. These steps include library preparation which produce gDNA libraries from each individual's DNA sample which have a certain size and are bound to adaptor sequences. The Ion torrent sequencing application used in this study was a 200 bp read application which means that gDNA should be fragmented in a manner to produce fragments of about 200 bp length. The adaptor sequences were sequences of about 40 bp which are recognizable by the machine and contain signals for barcoding gDNA and identification of start point of sequencing. In the following gDNA fragmentation, selection of 200 bp segments and adaptor ligation will be explained.

### 2.4.1 DNA isolation and genomic DNA fragmentation

Genomic DNA was isolated from peripheral whole-blood samples utilizing QIAamp DNA Blood mini kit (Qiagen, Hilden, Germany). The aim of the following process was to prepare libraries from genomic DNA which are compatible with sequencing. Through this process genomic DNA was fragmented using an enzymatic reaction. For this reason, Ion Xpress™ Plus Fragment Library Preparation kit was used. A mixture of restriction enzymes was used for this reason. However the composition of the enzyme mix was not revealed by the company; the result of the enzymatic reaction was blunt-ended DNA fragments. Depending on the duration of the enzymatic reaction, average size of the DNA fragments is different. The longer the reaction runs, the shorter the fragments will be. Figure 2.4.1 demonstrates the relation between time and fragment size analyzed with Bioanalyzer® instrument (not our data).

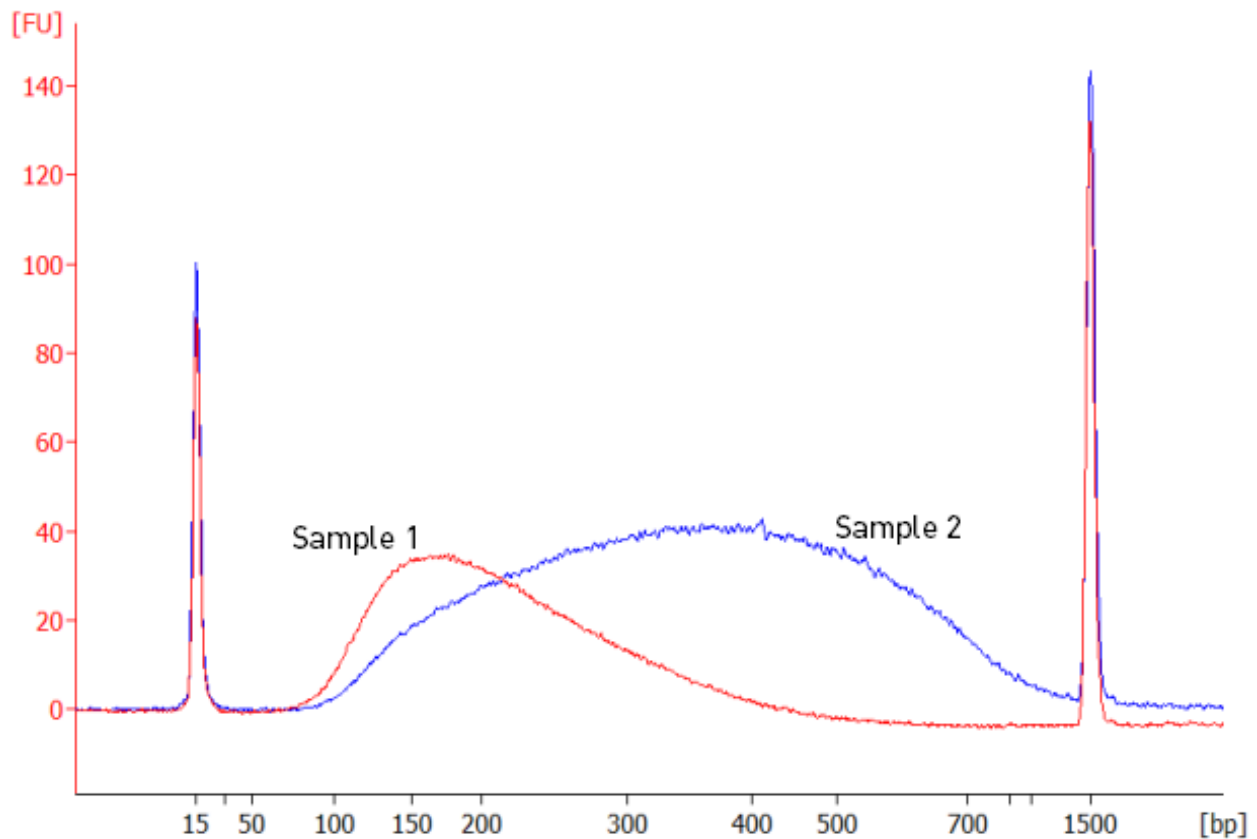


Figure 2.4.1 Analysis of fragmented genomic DNA. Sample 1 was digested for a longer time than Sample 2. The peak of fragment size is shifted from 400 bp in Sample 2 to 150 bp in sample 1. Peaks at 35 and 10380 bp represent weight markers (Image from Ion Xpress™ Plus Fragment Library Preparation User Guide)

In each case, eight gDNA samples were prepared at once. Nuclease free water was added to obtain 500 ng/35 µl starting amount of the input gDNA. Afterwards Ion Shear 10X buffer – the buffer provided by the kit and is necessary for the enzymatic reaction – was added to a 1.5-mL Lo-Bind tube containing 35 µl gDNA.

Afterwards, the mixture was pre-incubated at 37°C in the heat block (Thermomixer 5436, Eppendorf, Hamburg, Germany) for minimum of 5 minutes. Then, 10 µl of the *Ion Shear™ Plus Enzyme Mix II* was added to the mixture and mixed thoroughly by pipetting up and down. Next, the tubes were exactly 8 minutes on the heat block at 37°C. To stop the reaction, 5 µL of Ion Shear™ Stop Buffer was added and the reaction tube was stored on ice for 15 minutes. Finally, the reaction mixture was purified by Agencourt AMPure XP magnetic beads.

These magnetic beads are covered with carboxyl molecules which in presence of polyethylene glycol (PEG) and salt will attach to negatively charged DNA. In order to do so, the DNA fragments were first bound to the magnetic beads, then using a magnetic rack (Dyna, Invitrogen) beads were immobilized. The immobilization facilitated the removal of supernatant. Afterwards, the beads were washed with freshly prepared 70% ethanol. Consequently, the DNA was dissolved in Low TE buffer. The beads were separated using the magnetic rack and were discarded. Fragmented DNA which is dissolved in Low TE could be stored at -20°C.

#### 2.4.2 Adaptor ligation

After fragmentation, a pair of library adapters were ligated to each two sides of the fragmented DNA. Library adapters were ~40 bp double-stranded DNA molecules which contained specific sequences. We have used two types of adapters: Ion P1 Adaptor™ (which contains a sequence which used to attach to Ion Sphere Particles (ISPs)) and Barcoded adapters (which contains "TCAG" key for library identification and a specific). Both adapter types are blunt-ended on one side and have an overhanging single strand end on the other side. There are phosphorothioate bonds (shown with \* in Table 2.4.1) which protect the nuclease activity. The overhanging single strand end ensures the directionality of the adapter, meaning that the adapter is only ligated to the DNA fragment from its blunt end, as the outcome of enzymatic DNA fragmentation is blunt

ended. Barcoded adaptors allow the sequencing of genomic DNA (e.g.) from different subjects or from different tissues in one sequencing run while maintaining the possibility to differentiate between the single subjects.

Table 2.4.1 Adaptor sequences. The IonXpress barcode set <sup>TM</sup> was used. TCAG (red) is the sequencing key which is a signal for the machine. The GAT sequence (green) is an indicator of target sequence initiation. The sequence in between is specific for each barcode.

Name of the adaptor	Adaptor sequence	Length
Ion P1 adaptor	5'-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT-3'	41
	5'-ATCACC GACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGG*T*T-3'	43
<b>Barcode adaptors</b>		
IonXpress_1	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGCTAAGGTAACGAT-3'	43
	5'-ATCGTTACCTTAGCTGAGTCGGAGACACGC-3'	30
IonXpress_2	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGTAAGGAGAACGAT-3'	43
	5'-ATCGTTCTCCTTACTGAGTCGGAGACACGC-3'	30
IonXpress_3	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGAAGAGGATTCGAT-3'	43
	5'-ATCGAATCCTCTTCTGAGTCGGAGACACGC-3'	30
IonXpress_4	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGTACCAAGATCGAT-3'	43
	5'-ATCGATCTTGGTACTGAGTCGGAGACACGC-3'	30
IonXpress_5	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGCAGAAGGAACGAT-3'	43
	5'-ATCGTTCCTTCTGCTGAGTCGGAGACACGC-3'	30
IonXpress_6	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGCTGCAAGTTCGAT-3'	43
	5'-ATCGAACTTGCACTGAGTCGGAGACACGC-3'	30
IonXpress_7	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGTTCGTGATTCGAT-3'	43
	5'-ATCGAATCACGAACTGAGTCGGAGACACGC-3'	30
IonXpress_8	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGTTCCGATAACGAT-3'	43
	5'-ATCGTTATCGGAACTGAGTCGGAGACACGC-3'	30
IonXpress_9	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGTGAGCGGAACGAT-3'	43
	5'-ATCGTTCGGCTCACTGAGTCGGAGACACGC-3'	30
IonXpress_10	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGTGACCGAACGAT-3'	43
	5'-ATCGTTCGGTCAGCTGAGTCGGAGACACGC-3'	30
IonXpress_11	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGTCCCTCGAATCGAT-3'	43
	5'-ATCGATTTCGAGGACTGAGTCGGAGACACGC-3'	30
IonXpress_12	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGTAGGTGGTTCGAT-3'	43
	5'-ATCGAACCACTACTGAGTCGGAGACACGC-3'	30
IonXpress_13	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGTCTAACGGACGAT-3'	43
	5'-ATCGTCCGTTAGACTGAGTCGGAGACACGC-3'	30
IonXpress_14	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGTTGGAGTGTCGAT-3'	43
	5'-ATCGACACTCCACTGAGTCGGAGACACGC-3'	30
IonXpress_15	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGTCTAGAGGTCGAT-3'	43
	5'-ATCGACCTCTAGACTGAGTCGGAGACACGC-3'	30
IonXpress_16	5'-CCATCTCATCCCT*G*CGTGTCTCCGACTCAGTCTGGATGACGAT-3'	43



	5'-ATCGTCATCCAGACTGAGTCGGAGACACGC-3'	<b>30</b>
--	--------------------------------------	-----------

\* represents a phosphorothioate bond

The ligation reaction was performed as follows. Reagents displayed in Table 2.4.2 were added to a 0.2 ml PCR tube.

Table 2.4.2 Reagents used in the ligation reaction of sheared DNA and adaptors.

Reagent	Volume (µl)
Fragmented DNA	28
10X Ligase Buffer	5
Ion P1 Adapter	5
Ion Xpress™ Barcode X	5
dNTP Mix	1
DNA Ligase	2
Nick Repair Polymerase	4
Total	50

Afterwards, the tube was placed in a thermal cycler and the following program was performed not using heated lid option: 25°C for 15 minutes and then 72°C for 5 minutes. Finally the reaction mixture was transferred to a 1.5 Lo-bind tube and purified by Agencourt AMPure XP magnetic beads.

### 2.4.3 Size-selection

The sequencing instrument, the *Ion torrent Personalized Genome Machine*™ (Ion PGM) has a certain read length for each sequencing application. We decided to use the 200 base-read application; therefore, we had to select a specific size of DNA fragments (which were already ligated to adapters). After ligation a size of 350 bp is needed for sequencing. To do so, each sample was run on a 2% agarose gel, and the 350 bp sized fragments were cut out from the gel. The gel piece was then subjected to DNA purification from the gel using Qiacube™ (Qiagen) robot and Qiaquick™ gel extraction kit (Qiagen).

#### 2.4.4 Library amplification

In addition to size selection, amplification is needed to reach the proper yield and proceed further to target enrichment. Amplification was performed using universal primers which anneal to the adaptors; subsequently, an amplification of whole gDNA is possible. To perform the amplification, 200 µl of Platinum® PCR SuperMix High Fidelity, 7.5 µl of Library Amplification Primer Mix and whole amount of purified size-selected library DNA were combined in a 1.5 ml Lo-bind tube and then were divided into two 0.2 ml PCR tubes.

The reaction mixture was placed in a thermal cycler. And the program displayed in Table 2.4.3 was performed.

Table 2.4.3 PCR program of library amplification after size selection.

Stage	Step	Temperature	Time
Holding	Denaturation	95°C	5 min
	Denaturation	95°C	15 sec
14 cycles	Annealing	58°C	15 sec
	Extention	70°C	1 min
Holding		4°C	∞

Subsequently the PCR products were purified using Agencourt AMPure XP magnetic beads.

#### 2.5 Analysis of library preparation efficacy

After fragmentation, adaptors were ligated to the entire fragmented genomic DNA for each patient's DNA sample. Subsequently, adaptor ligated gDNA was analyzed on 2% agarose gel. The 350 bp region was then cut out. After purification of the DNA from the gel, amplification was performed. Figure 2.5.1 shows the documented agarose gel of representative samples selected from the entire process. After sequencing average read length was calculated 152 bp (102-175).

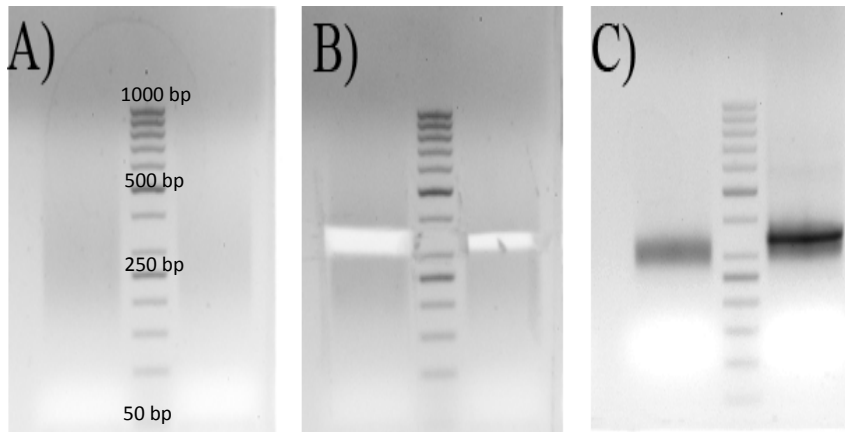


Figure 2.5.1 Analysis of library preparation efficacy. A) Sheared and adaptor-ligated DNA, B) size selection performed by cutting out the gel piece of desired size, C) amplified size selected DNA templates. On the left and right side of the size markers two different samples are illustrated.

## 2.6 Quantifying and pooling of amplified size selected libraries

Considering the capacity of the sequencing chip, read length and the size of our target region, we decided to pool 9 genomic DNA libraries to be analyzed on a single sequencing chip. The Ion 316™ sequencing chip has approximately 6 million wells, within each the polymerase reaction will take place. The size of our targeted region is 382,400 bases. Taking into account a 200 base-read for each well in a perfect condition, each base could be 350 times read (average read depth or coverage of 350x) while pooling 9 samples.

$$\sim 350x \text{ coverage} = \frac{6000000 \text{ wells} * 200 \text{ base read length}}{9 \text{ libraries pooled} * 382400 \text{ size of the target region}}$$

Five hundred nanogram of a DNA pool is needed to start the next step. In order to normalize the libraries in such a way, that each individual's library was equally represented in the pool, a quantitative real time PCR was performed. GeneRead DNaseq Library Quant Array (Qiagen) was used for this reason according the manufacturer's instructions with minor modifications. The concentration is calculated in picomole/liter and in order to convert it to ng/μl an average size of 330 bp and average mass of 660 ng/(nmol × bp) were assumed.

In case there was enough library yield, a mass of  $55 \text{ ng} = \frac{500 \text{ ng}}{9 \text{ samples}}$  from each library were pooled together. If the yield was not enough, equal masses of each library have been used for pooling (while taking the lowest library concentration into account).

Twenty one library pools were made, to sequence the 139 patients' genomic DNA. Pool number 7 was not successfully sequenced due to technical error. Pools' gDNA mass and composition are shown in Table 2.6.1 Eleven pools were made with the recommended 500ng gDNA mass; however, other pools were prepared with less amount due to lack of required DNA concentration. In addition 11 pools were consisted of 9 samples, 1 pool had 10 samples and the other pools had 7 or 8 samples. If a sample would not enough over all coverage (50%), it would be again sequenced. Therefore a number of samples were repeated.

Table 2.6.1 Genomic library pools' composition and pre-enrichment concentration.

Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng
1	500	2	500	3	500	4	500	5	500	6	500	8	238
Barc.	Samp.	Barc.	Samp.	Barc.	Samp.	Barc.	Samp.	Barc.	Samp.	Barc.	Samp.	Barc.	Samp.
1	434	1	292	6	421	1	285	6	436	5	245	1	623
2	297	2	430	7	263	2	302	7	215	9	645	2	219
3	230	3	640	8	419	3	254	10	280	10	228	4	418
4	213	4	433	9	278	4	432	11	634	11	231	5	221
5	209	5	245	10	424	5	644	12	622	12	607	9	286
6	646	13	287	11	626	8	632	13	256	13	222	11	259
8	625	14	435	12	216	9	423	14	240	14	603	12	204
9	627	15	225	13	276	15	416	16	608	15	229	14	260
12	217	16	214	14	232	16	258			16	610		
Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng
9	207	10	130	11	169	12	500	13	500	14	500	15	355
Barc.	Samp.	Barc.	Samp.	Barc.	Samp.	Barc.	Samp.	Barc.	Samp.	Barc.	Samp.	Barc.	Samp.
3	635	2	264	1	275	1	247	4	212	4	284	1	643
6	211	3	609	2	207	2	638	4	413	8	601	3	602
7	234	5	205	3	417	3	629	5	246	9	283	5	273
8	620	6	255	4	268	4	619	6	201	10	220	7	261
10	235	7	252	5	642	9	237	7	621	11	257	11	412
13	270	8	293	6	630	10	611	8	271	13	251	12	233
16	238	10	624	9	242	11	612	11	604	14	223	13	275
		12	203	13	244	12	241	13	279	15	227	14	609
		16	628	14	605	15	206	14	281	16	411	16	236
								16	631				
Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng	Pool No.	DNA ng		
16	147	17	144	18	500	19	500	20	162	21	94		
Barc.	Samp.	Barc.	Samp.	Barc.	Samp.	Barc.	Samp.	Barc.	Samp.	Barc.	Samp.		
1	249	7	616	2	214	1	210	5	270	5	622		
2	633	8	293	9	627	3	249	9	202	6	626		
6	255	9	267	10	282	4	252	10	607	7	634		
7	260	10	641	12	617	8	271	11	415	8	635		
8	210	12	614	13	610	9	603	12	609	13	637		
9	414	13	275	14	240	11	608	14	639	15	645		
10	637	14	218	15	624	12	241	16	625	16	648		
11	243	15	613										
12	288	16	243										

Pool numbers and the corresponding gDNA concentration is displayed in the above table. The composition of each pool (DNA sample numbers) and the corresponding barcode number are also displayed

## 2.7 Target enrichment

Having the gDNA libraries prepared, in the next step we needed to select the target regions and enrich them in our libraries. As only specific target genes were selected, DNA probes were

designed to capture only the specific targets. The probes were single strand DNA biotinylated oligonucleotides. The probe-hybridized target regions of our libraries were isolated using the biotinylated probes using streptavidin-magnet beads. Target enrichment by hybridization actually consisted of two hybridizations: A first hybridization of the library, then a PCR amplification of the enriched sequences and then a second hybridization which ensures the isolation of target regions from the other genomic regions and decreasing the off-target reads.

## 2.7.1 First hybridization

### 2.7.1.1 Adaptor and repeat sequence blocking

Before the first hybridization (target capture), adaptor blockers and Human Cot-1 DNA were added to the pooled DNA libraries. Human Cot-1 DNA<sup>®</sup> (the name Cot or C<sub>0</sub>T stems from the fraction of concentration (or C) of single stranded DNA which remain single stranded after a certain time (or T)) is a placental DNA that is mainly 50 to 300 bp in size and rich in repetitive DNA sequences. These sequences tend to hybridize to each other; as a consequence, probe hybridization to target sequence will be problematic. Human Cot-1 DNA<sup>®</sup> along with adaptor blockers were used to prevent non-specific hybridization. Mechanism of action of adaptor blockers and Human Cot-1 DNA<sup>®</sup> is shown in Figure 2.7.1.

Practically, 5 µl each of Ion TargetSeq™ Blocker BC 1–16, Ion TargetSeq™ Blocker P1 and Human Cot-1 DNA were added to the pooled libraries in a Lo-bind tube. In addition, a hole was made in cap of the tube using a sterile 20 gauge needle. Afterwards, using a vacuum-concentrator (Savant Instruments, Farmingdale, NY) set to 60 °C the mixture was dried for 30 minutes (until it was completely dry).

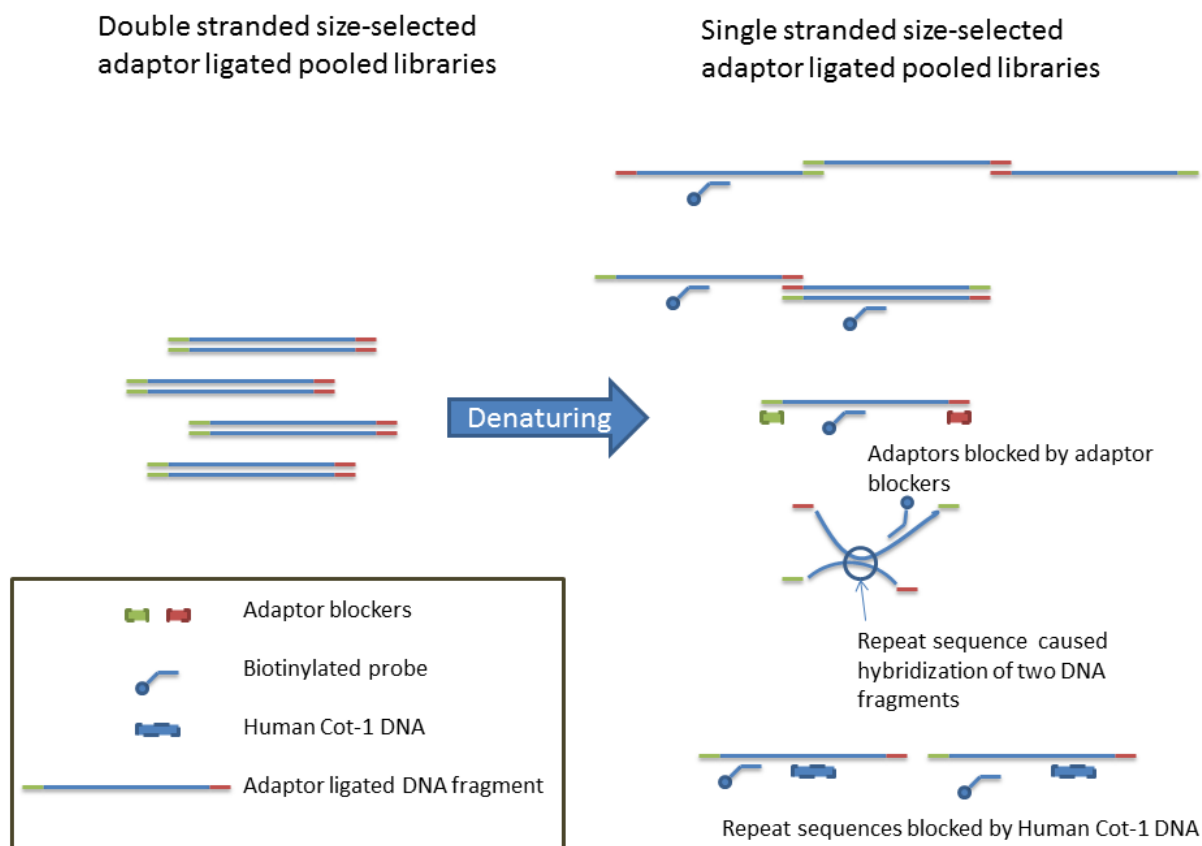


Figure 2.7.1 Mechanism of action of adaptor blockers and Human Cot-1 DNA. Common sequences in adaptors cause attachment of adaptor ligated DNA to attach together and hinder the hybridization process. In addition this nonspecific hybridization may also occur in between repeat sequences of the genome. Figure adopted from Targetseq™ Custom Enrichment User Guide

### 2.7.1.2 First Hybridization procedure

Target enrichment is performed by biotinylated probes which contain complimentary sequences to the target regions. The used Targetseq™ Custom Enrichment kit contains a custom probe pool, in which – according to manufacturers – every base in target region is at least covered by 4 different probes. The biotinylation permits the isolation of these probes by streptavidin beads (Dynabeads M-270 Streptavidin, Life Technologies).

To perform this step, 7.5 µl of the TargetSeq™ Hybridization Solution A (2X) and 3 µl of the TargetSeq™ Hybridization Enhancer B were added to the dried tube. The tube was then put in a 95°C heatblock for 10 minutes to denature the double-stranded DNA. Next, 2.25 µl of custom

probe pool and 2.25 µl nuclease-free water were added to the tube. Afterwards, the tube was placed in a thermal cycler at 47°C for more than 12 hour while using heated lid option.

#### 2.7.1.3 *Washing and recovery of probe hybridized DNA*

The entire amount of PCR product was transferred to the tube containing streptavidin beads (Dynabeads M-270 Streptavidin, Life Technologies). After thoroughly mixing the beads with hybridization mixture, the tube was placed back to the thermal cycler for 45 minutes. This allowed the biotinylated probes to bind to the streptavidin beads. Afterwards, several washing steps were performed using 4 different washing solutions namely “stringent wash solution”, and then wash solutions A, B and C. After washing, the beads were resuspended in 40 µl of LTE buffer.

#### 2.7.1.4 *Amplification and purification of the first capture reaction*

To amplify the captured DNA a PCR reaction was performed using universal primers. The captured DNA, which was still attached to the streptavidin beads, was used as the template. 200 µl of Platinum® PCR SuperMix High Fidelity, 10 µl of Library Amplification Primer Mix and 40 µl of TargetSeq capture beads with DNA were combined and then the PCR reaction described in Table 2.7.1 was executed.

Table 2.7.1 Amplification of the first capture reaction.

Stage	Step	Temperature	Time
<b>Holding</b>	Denaturation	95°C	5 min
	Denaturation	95°C	15 sec
<b>5 cycles</b>	Annealing	58°C	15 sec
	Extension	70°C	1 min
<b>Holding</b>		4°C	∞

Contents of the PCR tube were then added to a Lo-bind tube and the tube was place on a magnetic rack. The supernatant contained our amplified target region and the beads were discarded. The supernatant was subsequently purified using Agencourt AMPure XP magnetic beads.



### 2.7.2 Second hybridization

Second hybridization procedure was performed almost the same as first hybridization process with steps as described in 2.5.2.1 to 2.5.2.4 and the exception of only difference of 14 cycles amplification at the last step.

## 2.8 Template preparation

Previously prepared size selected adaptor ligated libraries enriched for the target region were now available for the next steps. These libraries should undergo a template preparation step to be ready for loading on the sequencing chip and starting the sequencing. These steps include emulsion PCR in which a template is attached to plastic microscopic spheres (Ion Sphere Particles, ISP) and is then amplified. Finally the spheres which were connected to the amplified template will be enriched and prepared for sequencing.

### 2.8.1 Enriched pooled library quantification

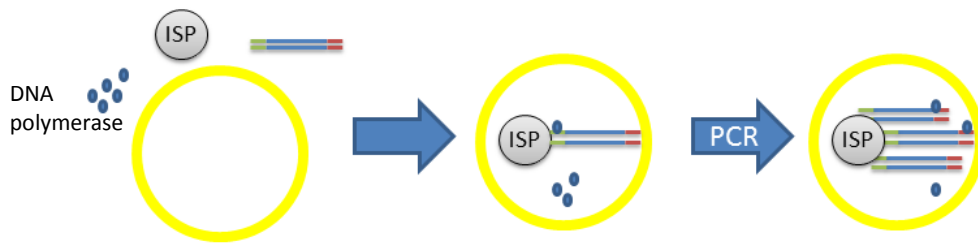
A quantitative real time PCR was performed to quantify the enriched pooled library. GeneRead DNAseq Library Quant Array (Qiagen) was used for this reason according the manufacturer's instructions with minor modifications. The concentration was calculated in picomolar and the pools were diluted to reach a 10pM concentration. The quantification step is necessary for performing emulsion PCR, as the ratio of DNA molecules to number of ISP has a vital role in quality of template.

### 2.8.2 Emulsion PCR

The Ion OneTouch™ System was used to prepare the template for sequencing. Adaptor ligated DNA and ISPs are introduced to Ion OneTouch™ machine. Emulsion PCR (Figure 2.8.1) is the main function of this machine(Dressman et al. 2003). In this process a PCR reaction will take place inside a water-in-oil microemulsion, resulting in amplification of the one DNA molecule attached to a single ISP (Ion Sphere Particle).This PCR reaction is performed using biotinylated primers with the purpose of isolation of the harvest by streptavidin magnet beads. As a result, millions of

copies of a certain DNA piece will be attached to one ISP. Then the emulsion is broken and as a result ISP bound DNA is prepared for sequencing.

This is the optimal scenario that could happen; however, there might also be copies of different DNA pieces attached to one ISP that makes the sequencing for this ISP impossible (Figure 2.8.2). This scenario is called polyclonality. In order to avoid this problem, the ratio of ISP to DNA molecules must be as close as possible to 1:1.



*Figure 2.8.1 Emulsion PCR. In this process one DNA template will be amplified in one microemulsion oil sphere attached to an ion sphere particle (ISP). By quantification of DNA templates a 1 to 1 ratio of template to ISP will be facilitated. PCR primers and DNA polymerase are also added PCR amplification to this reaction.*

For this reason 20 µl of enriched library pool was added to 500 µl of Ion OneTouch™ 2x reagent Mix, 280 µl nuclease-free water, 100 µl Ion OneTouch™ Enzyme Mix and 100 µl Ion OneTouch™ 200 Ion Sphere™ Particles (parts of Ion OneTouch™ 200 Template Kit v2 DL). After mixing thoroughly the reaction mixture was loaded on Ion OneTouch™ Reaction Filter Assembly. Afterwards, 1500 µl of OneTouch™ reaction oil was added to the filter. Then the filter was firmly inserted into OneTouch machine top stage. Finally the reaction was started. The duration of each reaction consisting of 40 cycles of emulsion PCR and breaking of emulsion is 4 hours.

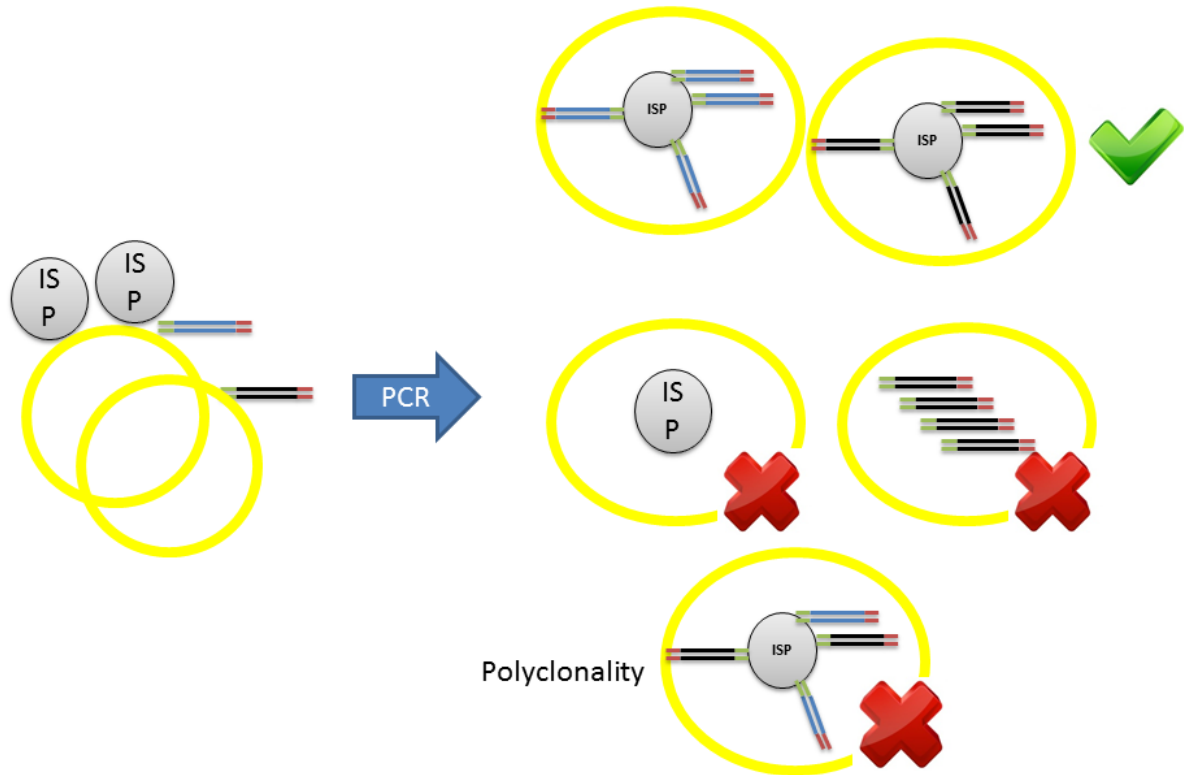


Figure 2.8.2 Emulsion PCR difficulties. Several erroneous scenarios might happen while performing emulsion PCR. An ISP without any DNA template or a vesicle with DNA templates but with no ISP result in an invalid outcome. Also several different templates may be amplified while being attached to one ISP which results in polyclonality (cannot be used for sequencing).

### 2.8.3 ISP Enrichment

ISP enrichment is a process, in which the ISPs that are not attached to DNA are isolated and discarded. To do so, streptavidin magnetic beads were used (Dynabeads® MyOne™ C1, Life Technologies). Magnetic beads attach to the amplified DNA molecules via the biotinylated site. The beads can be isolated by a magnetic field; consequently, ISPs which does not contain DNA, stay in supernatant and can be discarded. Afterwards, the hydrogen bonds between two DNA strands were broken using a highly alkaline solution (melt-off solution). This time using a magnetic field the beads can be separated from ISPs and can be discarded (Figure 2.8.3).

In order to perform this procedure, we have used Ion OneTouch™ ES instrument. First of all, the melt off solution was prepared as following: 10 µl 20% TWEEN was added to 125 µl 0.1M NaOH solution and 865 µl nuclease-free water. Ion OneTouch™ ES instrument 8 well strip was loaded according to Table 2.8.1 . The strip was then placed on Ion OneTouch™ ES instrument. An enrichment procedure lasts about 40 minutes. After completion of the procedure the ISPs were washed twice with 100 µl of Ion OneTouch™ Wash Solution. Finally the template was ready for sequencing.

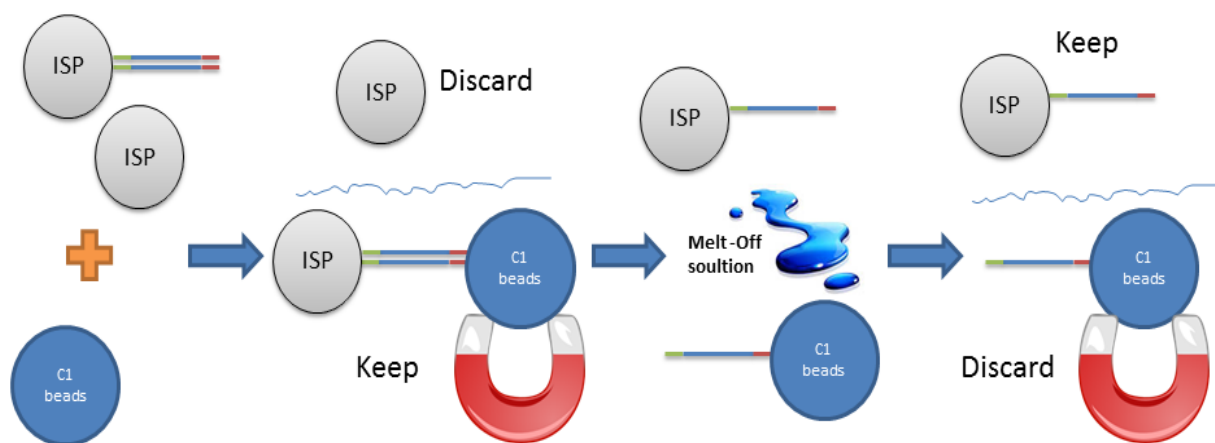


Figure 2.8.3 ISP enrichment. Through this process only ISPs which are bound to DNA templates are selected.

Table 2.8.1 Onetouch ES 8 well strip's wells contents

Well number	Reagent to dispense in well
1	Entire template-positive ISP sample (100 µL)
2	130 µL of Dynabeads® MyOne™ Streptavidin C1
3	300 µL of Ion OneTouch™ Wash Solution
4	300 µL of Ion OneTouch™ Wash Solution
5	300 µL of Ion OneTouch™ Wash Solution
6	Empty
7	300 µL of Melt-Off Solution
8	Empty

## 2.9 Massively Parallel Sequencing

In this study, we have used Ion torrent PGM which is a semi-conductor based sequencing device. The basis of this method is the detection of pH changes following the release of proton after each base is added to the DNA structure by the polymerase enzyme. The prepared templates should be loaded on a chip which contains several wells. Each sequencing reaction happens in a well which is smallest unit of sequencing in this technique. Each well has the capacity of one ISP. Depending on the type of the chip the number of wells that they contain varies. Chips Ion 314, 316 and 318 contain 1.2, 6.1 and 11 million wells respectively. Structure of a well is shown in Figure 2.9.1. During sequencing dNTPs are added one after another and in case they are complementary for the template DNA they will be attached and then a pH change will be detected.

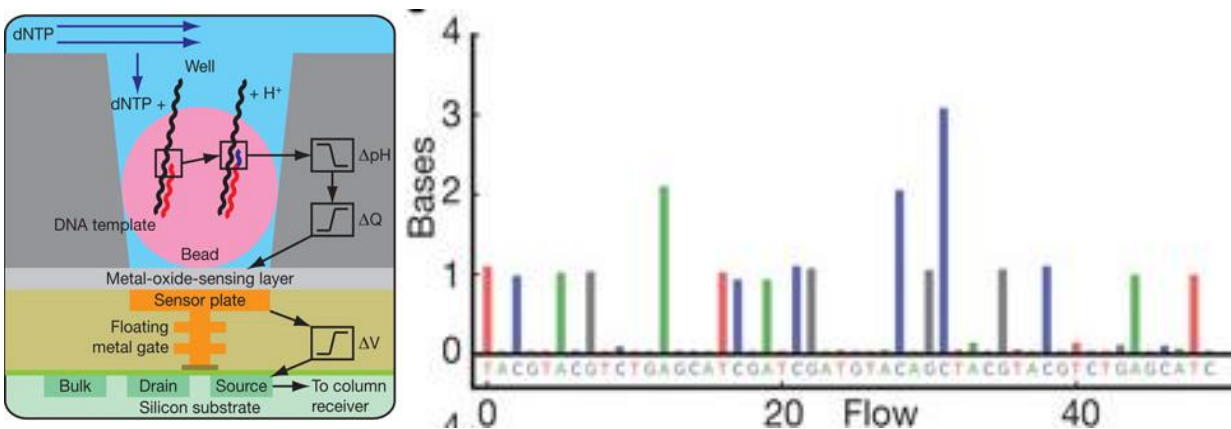


Figure 2.9.1 Structure of an Ion chip and base calling. Figure A shows the cross-section of one sequencing well of the ion chip. Hydrogens released during the sequencing reaction are marked and will be detected by the microarray as indicated. Figure B illustrates the base-calling. The intensity of the signal could indicate the number of nucleotides when there is more than one of a certain nucleotide in a row. Taken from Rothberg et al. Nature **475**, 348-352 (2011) doi:10.1038/nature10242

### 2.9.1 Addition of sequencing primer and sequencing polymerase

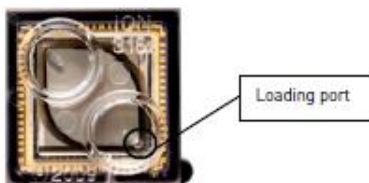
Ion PGM 200 Sequencing kit and Ion 316 chip were used in our study. Five microliters of Control ISPs were added to the enriched template and then 100  $\mu$ l of Annealing buffer was added to the solution. After centrifuging only 15  $\mu$ l of reaction mix was kept and the rest was discarded. Twelve microliters of sequencing primers were added to remaining reaction mix and was placed in a

thermal cycler for annealing process. A program with 2 minutes in 95°C and then 2 minutes at 37°C was performed. After thermal cycling, 3 µl polymerase enzyme was added to the tube to finally reach the volume of 30 µl. The reaction mixture was incubated in room temperature for 5 minutes.

### 2.9.2 Loading the sequencing chip

Ion 316 chips (Figure 2.9.2) were used in this study. The chip contains a loading port, from where the template should be loaded. Before loading the templates, 100 µl isopropanol was loaded on the chip. Next, the excess liquid should be removed from the other port. This step should be repeated 2 times using 100 µl annealing buffer. Subsequently, 30 µl template was loaded on the chip. The pipette tip was fixed firmly on loading port and then the pipette was slowly dialed down to load the liquid on the chip. The chip was then centrifuged to seed the ISPs inside the wells. Finally, the chip was then placed in PGM.

*Figure 2.9.2 Ion 316 chip (Image from Ion PGM™ 200 Sequencing Kit User Guide)*



### 2.9.3 Sequencing

The Ion torrent sequencing machine or PGM is connected to a computer server. The server is responsible for analysis of signals provided by PGM. The software controlling this sever is called Torrent suite. We have used Torrent suite version 3.6.2 for our analysis.

Before running the sequencer, a run should have been planned using Torrent suite software. Run planning requires the info regarding: the kits used for preparation of sequencing template, the barcode set, respective names of each sample for each barcode, sequencing application and targeted regions.

For each run, the predefined plan was selected from the PGM console and matched with the loaded chip. Each run took about 3 hours. After each run, the torrent server starts to analyze the generated sequences and to align them to the selected reference sequence. Torrent suite also includes plug-ins such as Torrent Variant Caller (TVC) and coverage analysis.

Coverage analysis surveys the coverage of selected target region. The file containing coordinates of target regions was uploaded in Torrent suite prior to using the PGM for the first time. TVC is the plug-in which identifies the variations found in gene sequences compared to the reference genome.

## 2.10 Variant calling

To discover the single nucleotide variants (SNVs) two softwares were used: DNASTar and Torrent Variant Caller plug-in of Torrent suite which are described in the following text.

### 2.10.1 DNASTar software

DNASTar (version 11.1) is a commercial software package, in which we have used Seqman NGen, Seqman Pro and Arraystar. Seqman NGen aligns the sequences to the reference genome, which in this case was human genome version 19. Seqman Pro was used to visualize the sequences. Arraystar was used to compare genotypes of all variants in all patients along with the coverage at this position in a single table.

Using Seqman Ngen software the variants were called. The parameters used were as following Mer size(or K-mer size which indicates the number of nucleotides which are considered as a unit in alignments) and minimum matching percentage were set to automatic, genome ploidy was set to Diploid, in the SNP tab of advanced options “calculate SNPs” was checked, minimum SNP percentage was set to 0, SNP confidence threshold was adjusted to 0, minimum SNP-count was 2, minimum base quality score was set to 5 and the Check strands box was checked.

### 2.10.2 Torrent Variant Caller

Torrent Variant Caller identifies single-nucleotide polymorphisms (SNPs), multinucleotide polymorphisms (MNPs), insertions, and deletions in performed sequencing analysis. TVC does this action in two following steps.

Step 1 is to survey the aligned sequences and subsequently, find candidate regions which have evidences of presence of a variant in the examined region. These regions are then marked as probably containing a variant. Candidate regions must pass a minimal set of requirements (such as read quality, minimum coverage, etc.) (Garrison E 2012).

Step 2 is to additionally assess the candidate variant identified in step 1 to assign the confidence rate of each variant and also to assign the genotype. Candidate positions are then evaluated by a statistical model offered from GATK. GATK is a software package developed at the Broad Institute to analyze next-generation resequencing data. Finally, to decide if the assumptions of the model are not in agreement with the data, variants are filtered by several criteria.

All of these actions are controlled by predefined parameters provided for the software. It is possible to use pre-configured parameter settings in accordance with the estimated variant frequency in the sample (Germline or Somatic) and the desired stringency level of the calls. On the other hand, Torrent Variant Caller offers users to define the parameters, depending on their population and aims.

The Torrent Variant Caller has an output file for each individual; in which only the variations are displayed. Using these output files by collecting the variations found a hotspot file containing all variants was made and uploaded to Torrent suite. This file contains a list of chromosomal positions, along with the reference and alternative allele belonging to that position. This file (the hotspot file) obliges the TVC to analyze and report all listed positions in the hotspot file, even if the called allele is not a variant. Consequently, it was determined whether the lack of the call is because of absence of coverage or because the called sequence is the same as reference allele.



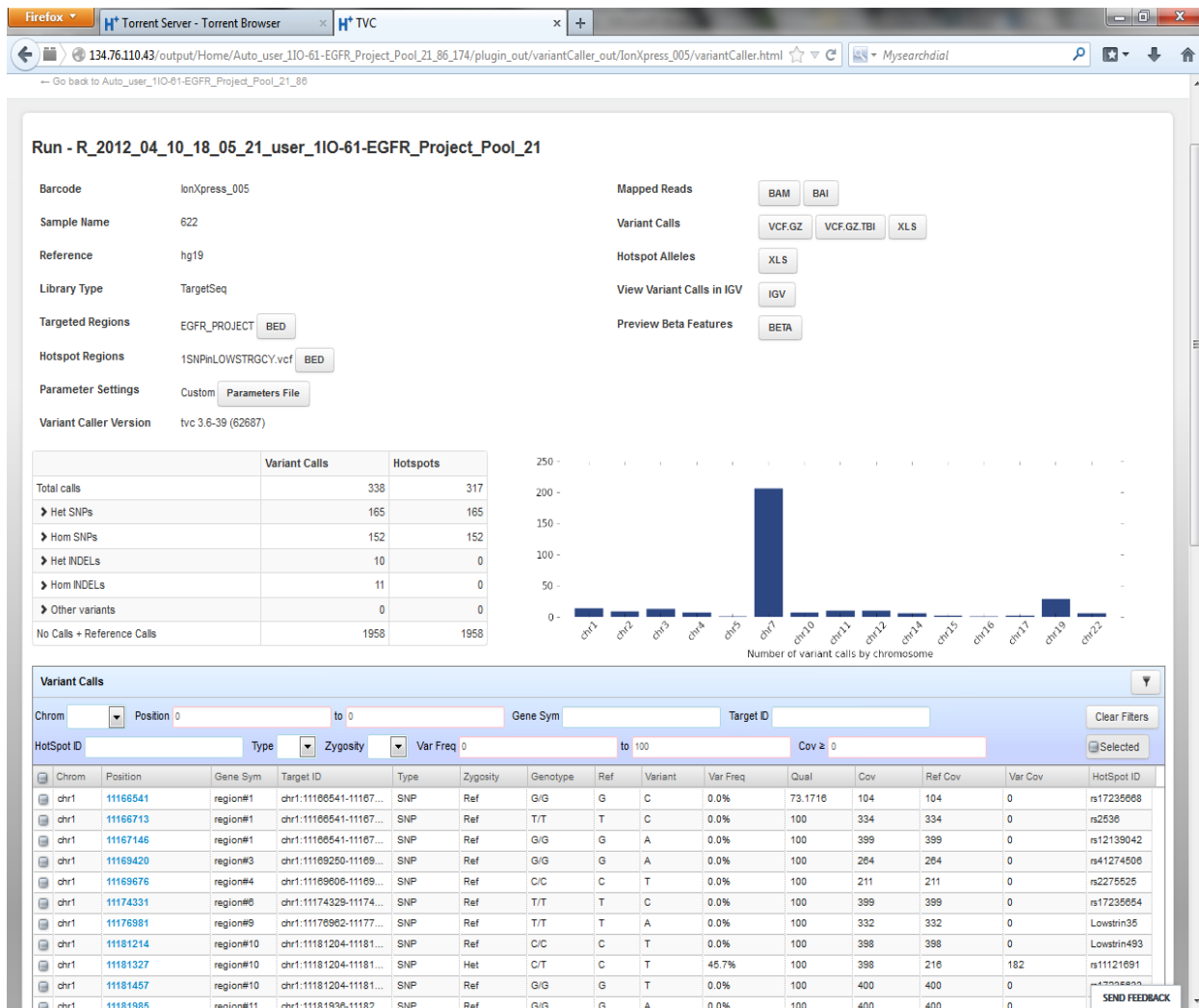


Figure 2.10.1 Screen-shot of a Torrent variant caller output. In the top section information regarding the sequencing run name, application type, barcode name, the parameters used for variant calling along with a summary of the variations found are displayed. A bar chart of variations found categorized by chromosome is shown on the right side. Finally the variants are shown at the bottom. Data regarding variants' chromosomal position, alleles and the quality of sequencing at the position are demonstrated.

Figure 2.10.1 shows a representative sample's Torrent Variant Caller output view. The plug-in output files, can be loaded into Broad Institute's Integrative Genomics Viewer for further visualization or analysis (Thorvaldsdóttir et al. 2013). Upon clicking, the link opens and displays all reads covering the chromosomal position. Figure 2.10.2 shows a representative heterozygote variant call on Integrative Genomics Viewer.

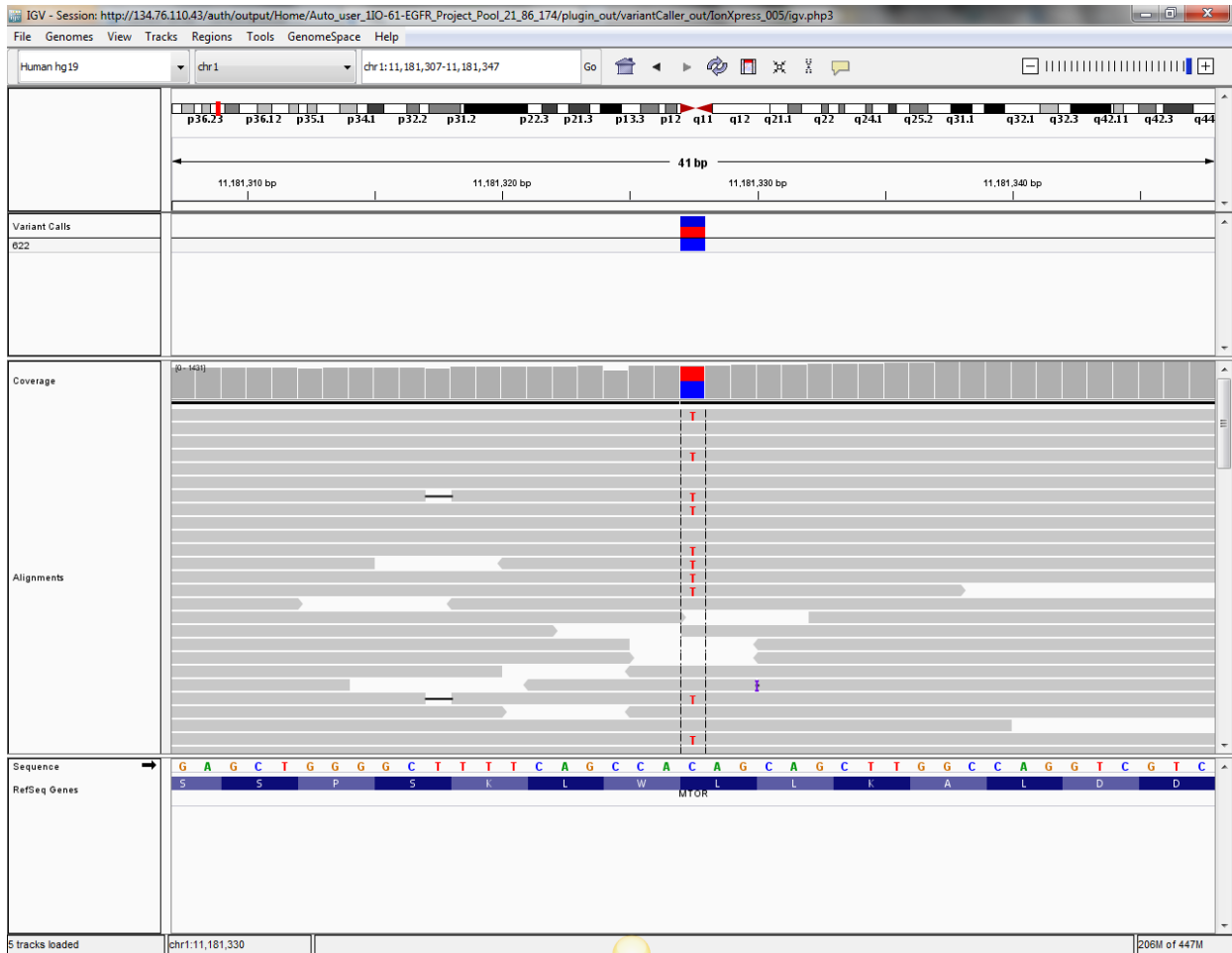


Figure 2.10.2 Integrative Genome Viewer (IGV) interface. Overview of IGV interface for showing a C to T variation. Gray bars represent the reads. Each horizontal grey bar indicates one read. All reads are aligned by the software and the variants are marked. Here only a selected number of the reads at that position are illustrated.

Two pre-configured settings for germline SNPs were applied to the analysis. These configurations were High stringency and Low stringency. Table 2.10.1 displays the settings used in analysis using Torrent Variant Caller. After surveying the results of these analyses, a custom setting was made to better match our desired coverage threshold and confidence of called variants.

Table 2.10.1 Torrent Variant Caller parameter configurations. The steps after which we decided to use custom parameters are stated in paragraph 3.4.3.

Parameter	Germline Stringency	High	Germline Stringency	Low	Custom parameters
data_quality_stringency	8.5		6.5		8.5
hp_max_length	8		8		8
filter_unusual_predictions	0.25		0.3		0.25
filter_insertion_predictions	0.2		0.2		0.2
filter_deletion_predictions	0.2		0.2		0.2
snp_min_cov_each_strand	3		0		3
snp_min_variant_score	10		10		10
snp_min_allele_freq	0.15		0.1		0.15
snp_min_coverage	20		6		10
snp_strand_bias	0.95		0.95		0.95
snp_beta_bias	8		30		8
indel_min_cov_each_strand	3		5		3
indel_min_variant_score	10		10		10
indel_min_allele_freq	0.15		0.1		0.15
indel_min_coverage	20		15		20
indel_strand_bias	0.85		0.85		0.85
indel_beta_bias	8		8		8
hotspot_min_cov_each_strand	3		3		3
hotspot_min_variant_score	10		10		10
hotspot_min_allele_freq	0.15		0.1		0.15
hotspot_min_coverage	20		6		10
hotspot_strand_bias	0.95		0.95		0.95
hotspot_beta_bias	8		30		8
downsample_to_coverage	400		400		400
outlier_probability	0.01		0.01		0.01
prediction_precision	1		1		1
heavy_tailed	3		3		3
kmer_len	19		19		19
min_var_count	5		5		5
short_suffix_match	5		5		5
min_indel_size	4		4		4
max_hp_length	8		8		8
min_var_freq	0.15		0.15		0.15
relative_strand_bias	0.8		0.8		0.8

Parameters in blue color which control the minimum coverage acceptable are modified to prepare the custom parameters from high stringency parameters.

## 2.11 Statistical analysis and functional prediction of nonsynonymous variants

Single marker association analysis of SNPs and skin rash was performed in two phases. In the first and second phases we have dichotomized patients regarding skin rash to respectively grade 0 compared to grade 1 or higher, and grade 0-1 compared to grade 2 or higher. Single marker analysis was performed with Plink software (version v1.07) using Fisher's exact test for genotypic along with allelic models (Purcell et al. 2007). Fisher-Freeman-Halton 2x3 exact test, an extension for Fisher's test which can be used for 2x3 tables, has been used for 2x3 tables. In genotypic model each individuals' genotypes (considering r as reference and A as alternative alleles, rr, rA and AA) are as units which are hypothesized to affect the dependent variable. On the other hand, in the allelic test number of alleles (number of r or A alleles) of each subject's genotype are considered to correspond to the presence of dependent variable (in this text skin rash) (Figure 2.11.1). Correction for multiple testing considering 1437 variants resulted in P value of 0.000035 as the level of significance required for single marker testing.

Genotypic test			Allelic test		
	dependent variable			dependent variable	
	absent	present		absent	present
No. of rr	a	b	No. of r	a	b
No. of rA	c	d	No. of A	c	d
No. of AA	e	f			

Figure 2.11.1 Explanation of genotypic and allelic models. The crosstabulation regarding each of these models is displayed. r : reference allele; A: Alternative allele.

The linkage data was analyzed using haploview software (v 4.2) (Barrett et al. 2005). Association plots were created using integrative genomics viewer software (Thorvaldsdóttir et al. 2013). Prediction of SNPs' functionality was performed using SIFT (Sort Intolerant from Tolerant amino acid substitutions, v1.0.3), Provean (Protein Variation Effect Analyzer, v 1.1.3) and Polyphen2 (Polymorphism Phenotyping v2 v2.2.2) tools (Adzhubei et al. 2010, Choi et al. 2012, Kumar et al. 2009). These tools use algorithms which are trained by known in vitro data for changed functionality. As an example input and output data of Polyphen2 tools is stated in the following. As an input chr1:11190646 C/G was entered and the output can be found in Table 2.11.1.

Table 2.11.1 An example output data of Polyphen2 prediction tool.

<b>Chrom.</b>	<b>Position</b>	<b>Gene</b>	<b>nt1</b>	<b>nt2</b>	<b>pos</b>	<b>aa1</b>	<b>aa2</b>	<b>Based on</b>	<b>pph2 prediction</b>
chr1	11190646	MTOR	C	G	1851	S	R	sequence annotation	deleterious

Gene-based and pathway-based analysis was performed with plink/seq (version 0.08 <http://atgu.mgh.harvard.edu/plinkseq/index.shtml>) using gene burden test, C-alpha test, Frequency-weighted test, Variable threshold test, UNIQ and summary statistics. Rare SNPs that were predicted to be damaging by any of three softwares (SIFT, Provean or Polyphen2) were enrolled in this analysis. Rare SNPs were defined as less than 13 minor allele count which was calculated considering 126 patients and MAF of less than 5%. Minor allele count was used instead of MAF to exclude the missing data effect.

## 3 Results

### 3.1 Patients

One hundred and twenty six patients receiving EGFR inhibition therapy from September 2008 to December 2010 were enrolled in this multicenter study. Patients' demographic data are presented in Table 3.1.1. Median and mean age were 68 and 66.6 years (43-86 years). The majority of patients were receiving erlotinib (73 %) while cetuximab was the second frequent drug (22.2 %). Sixty one percent of patients had bronchial cancer; in addition, 32.5%, 4%, 0.8% and 0.8% were affected with pancreatic, colon, HNSCC and upper GI cancers respectively.

Occurrence and severity of skin rash was surveyed in this study. Skin rash occurred in 80.2 % of the patients taking EGFRIs. The majority of patients with skin toxicity had grade 1 skin rash (38.1%) and a close number had grade 2 rash. Only 6 patients (4.8%) were affected with grade 3 rash. Grade 4 or generalized rash was not observed in this study.

**Table 3.1.1 Patient characteristics**

Characteristic	N	%
<b>Age</b>	Median (range) 68 (43-86)	
<b>Sex</b>	Male	83 65.9
	Female	43 34.1
<b>EGFRI</b>	Erlotinib	92 73.0
	Gefitinib	4 3.2
	Cetuximab	28 22.2
	Panitumumab	2 1.6
<b>Disease</b>	Bronchial Cancer	77 61.1
	Colon Cancer	5 4.0
	Head and neck SCC	1 0.8
	Pancreatic Cancer	41 32.5
	Upper gastrointestinal Cancer	1 0.8
<b>Smoking</b>	Never	38 30.2
	Former	63 50.0
	Present	16 12.7
	NA	9 7.1
<b>Clinical stage</b>	IIIa	2 1.6
	IIIb	16 12.7
	IV	102 81.0
	NA	5 4.0
<b>Skin rash grade</b>	0	25 19.8
	1	48 38.1
	2	47 37.3
	3	6 4.8
<b>Diarrhea grade</b>	0	81 64.3
	1	25 19.8
	2	15 11.9
	3	4 3.2
	4	1 0.8

As shown in the table, patients received either one of the small molecule tyrosin kinase inhibitors or monoclonal antibodies. Clinical staging was according to Classification of Malignant Tumors (TNM) and grading of the adverse effects was done according to NCI CTC version 3.

Since we have chosen skin rash along with diarrhea and survival as parameters for the pharmacogenomics analysis, it may be interesting to know how much these parameters are correlated to each other. If similar processes are underlying, there might be correlations in between these parameters. As a simple approach to this question we illustrate the correlation between the parameters in spine plots and scatterplots.

In Figure 3.1.1 the correlation between diarrhea and skin rash is illustrated. A Spearman's rank correlation coefficient of 0.14 was calculated and there was no significant correlation between skin rash and diarrhea (P value of 0.10).

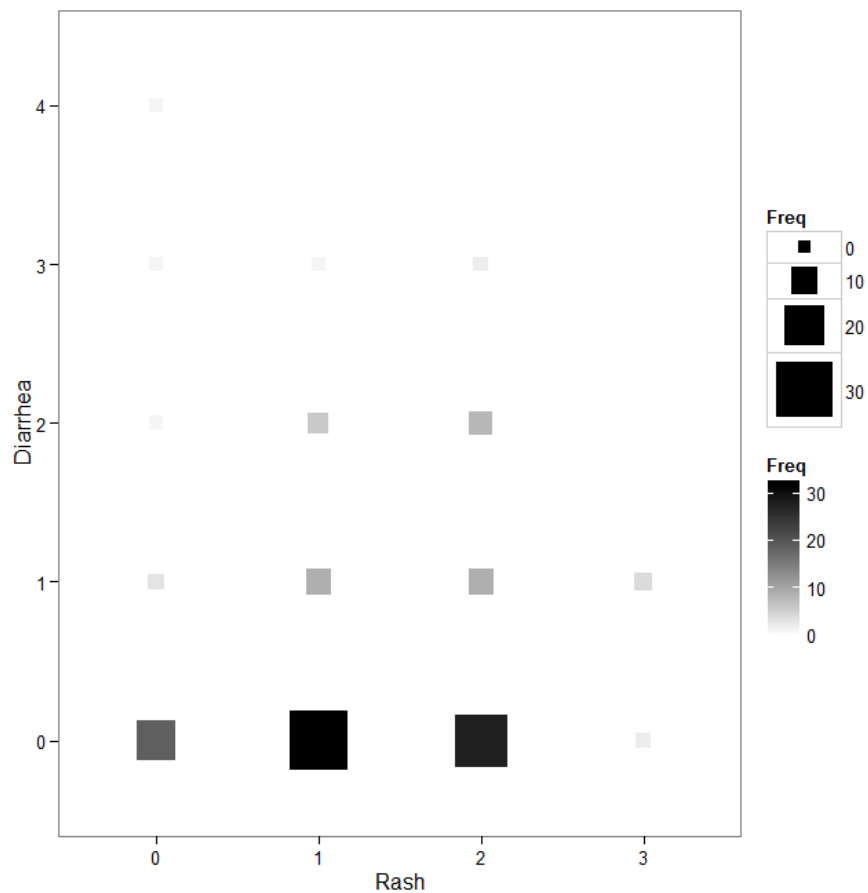


Figure 3.1.1 Spine plot of correlation of skin rash and diarrhea in study population. A correlation was not observed. The size and the color intensity of squares represent the number of patients. Diarrhea did not occur frequently (mainly grade 0).

The correlation between skin rash and overall survival is similarly illustrated in Figure 3.1.2. In this correlation only those patients could be considered who passed away in the observation period. A more complete correlation using Kaplan-Meier-Plots with the present sample has already been published by (Paul et al. 2014) and is in addition illustrated here. The same survey

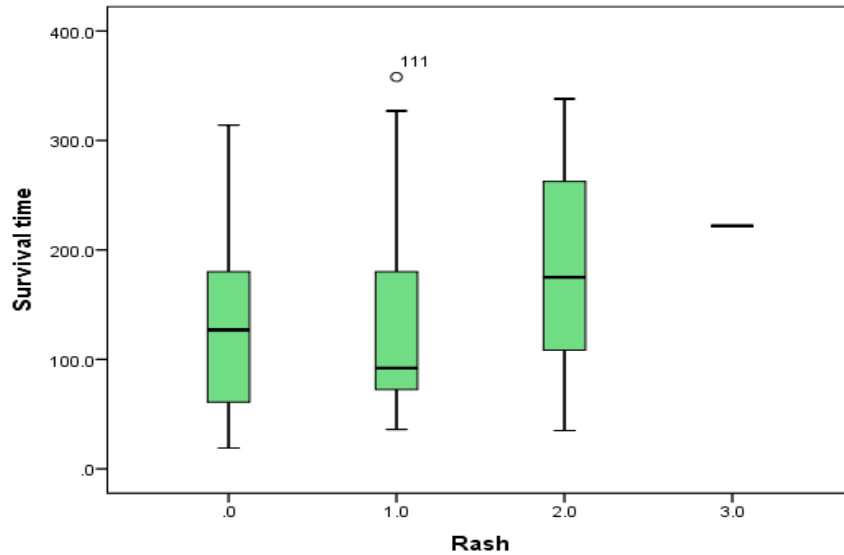


Figure 3.1.2 Correlation between skin rash and survival in patients who passed away during the study period. A trend for a better survival was observed in patients with higher levels of skin rash. Several patients in skin rash grade 3 group were alive at the last day of follow up and therefore this group is small in this survey. Survival time is calculated and displayed in days.

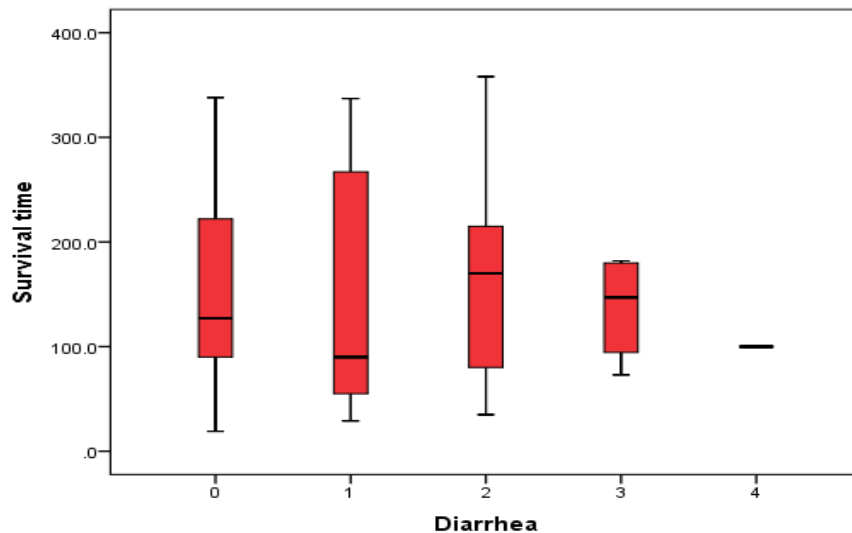


Figure 3.1.3 Correlation between diarrhea and survival in patients who passed away during the study period. A correlation between survival and overall survival was not observed. Several patients in skin rash grade 3 group were alive at the last day of follow up and therefore this group is small in this survey. Survival time is calculated and displayed in days.



was performed for assessing the correlation between diarrhea and overall survival. No correlation was observed.

Next to these questions, an important question is about the confounding factors which would have to be considered in the genotype-phenotype correlation analysis. Therefore, we analyzed whether skin rash, diarrhea and survival (the latter in the subgroup with known time of death only) were correlated with gender (Figure 3.1.4) or age (Figure 3.1.5). No correlations were found in this survey, which to some extent rules out gender or age as a confounding factor.

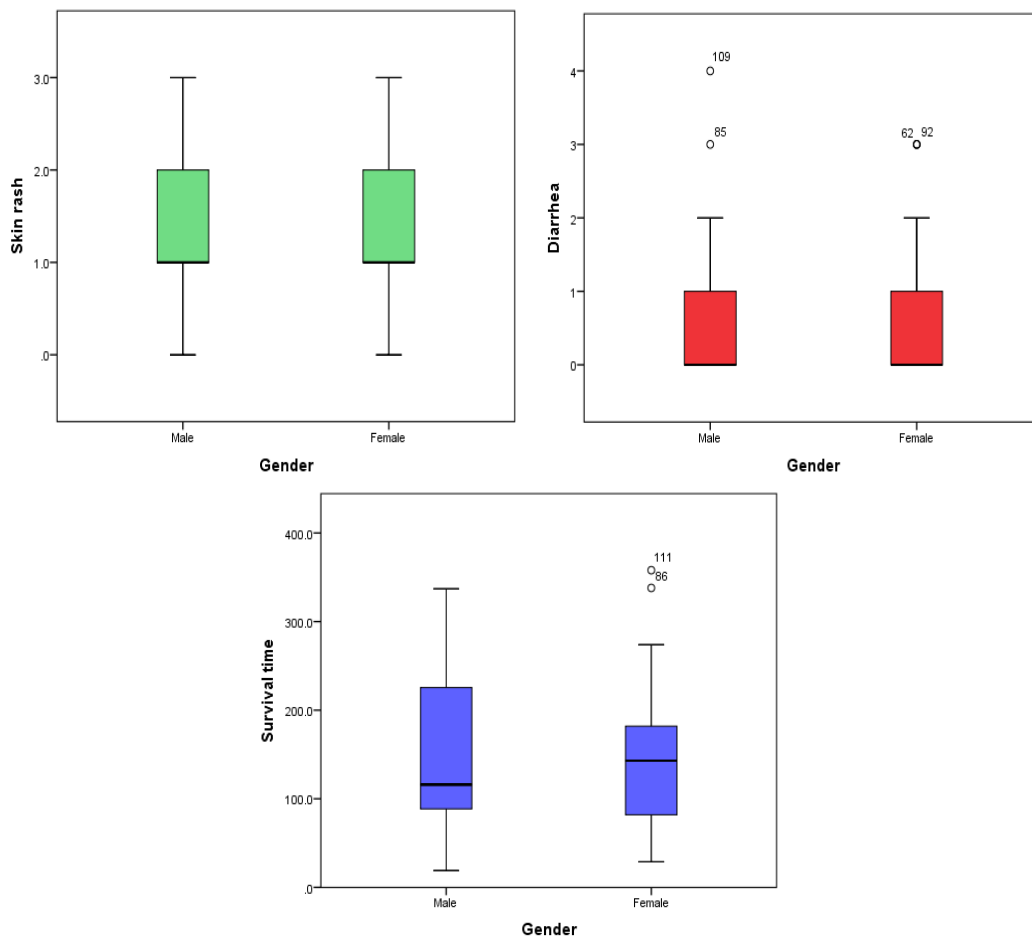


Figure 3.1.4 Correlation between gender and skin rash, diarrhea and survival. Survival was assessed only in patients who passed away during the study period. There was no correlations between gender and phenotypes studied.

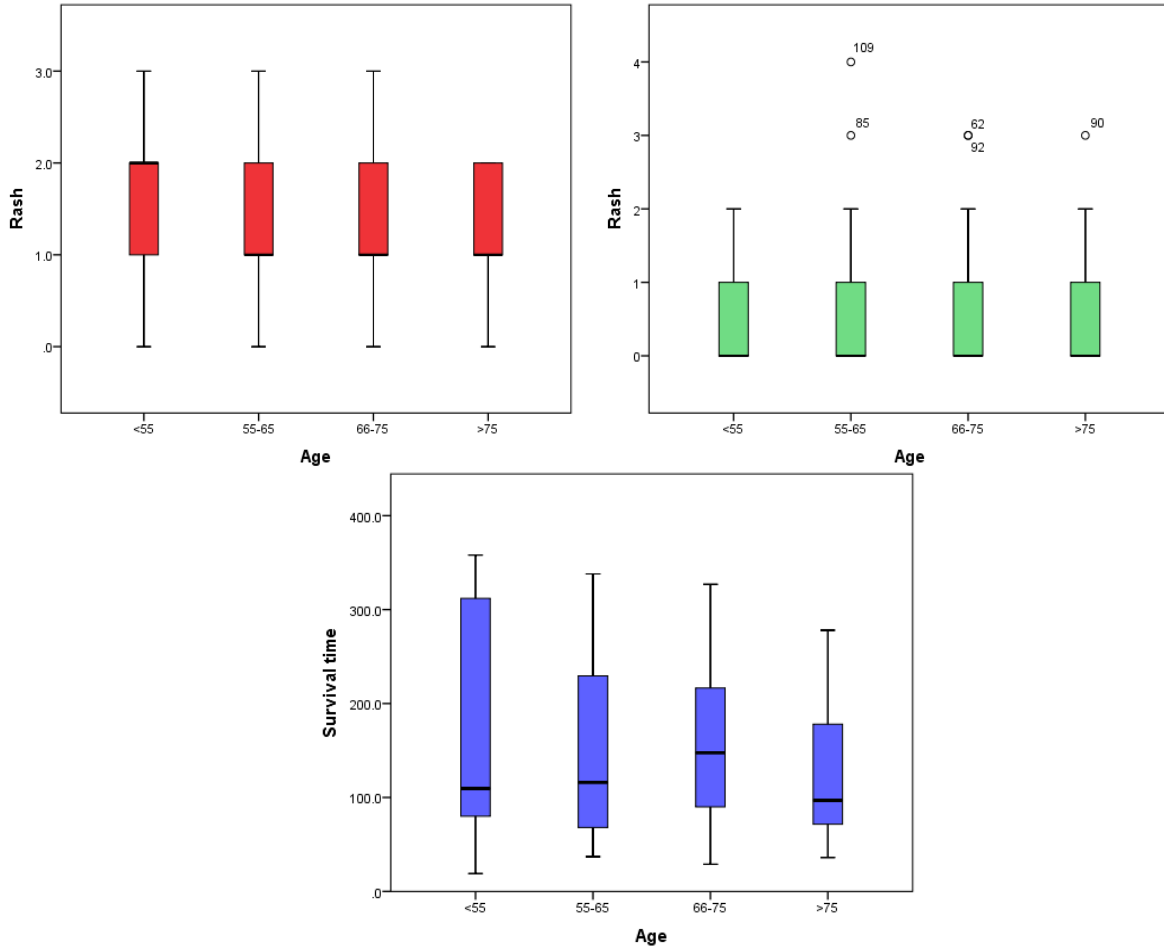


Figure 3.1.5 Correlation between age and skin rash, diarrhea and survival. Survival was assessed only in patients who passed away during the study period. There was no correlations between age and phenotypes studied.

### 3.2 Target enrichment

The file including the target regions' chromosomal positions has been uploaded to the sequencing machine's server computer (see 2.1 in methods). Subsequently, identification of reads which were aligned to the target region or unwantedly to other genomic regions was facilitated. Average percentage of on-target reads (reads which were aligned to target region) was 93% (80%-96%, the lowest percentage through the highest in a sequencing analysis). Moreover, 94% (65%-99%) of target bases were at least one time covered. On average, 85% (52%-98%) of the target region was at least 20 times covered. Figure 3.2.1 displays the visual coverage of EGFR, NFKB2 and MAPK3 in a representative sample.

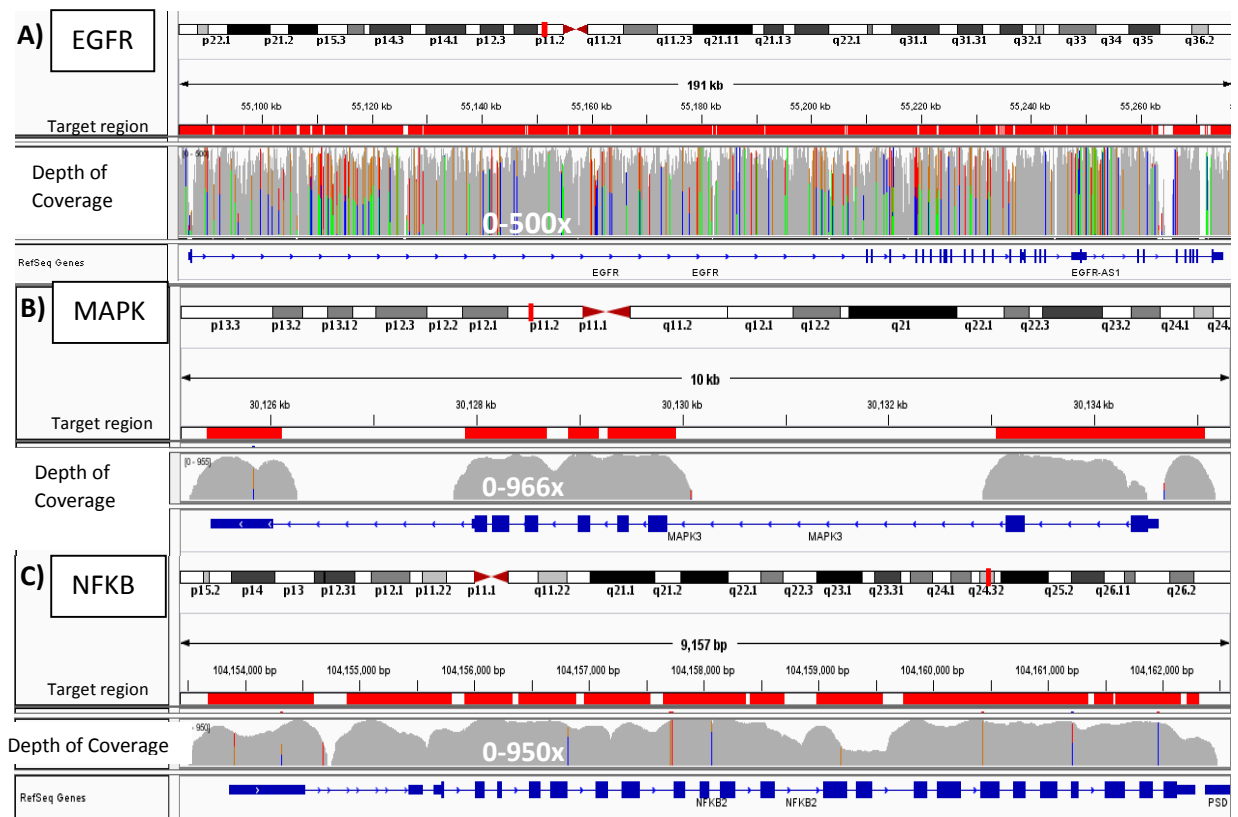


Figure 3.2.1 Target enrichment efficacy and coverage analysis. In all these examples a high coverage was achieved. A) Coverage of EGFR gene. The whole EGFR gene was targeted which is approximately 200 kb long. The variations are shown in form of colored lines. As the region sequenced in EGFR gene was larger than other genes, the frequency of variations was higher. B) and C) NFKB2 and MAPK3 genes (respectively) in a representative sample. Red bars track represent our target region. Coverage is shown in a logarithmic scale. Figures are demonstrated using integrative genome viewer. The blue line displays the genomic region in each figure.

### 3.3 Template preparation, loading the sequencing chip and massively parallel sequencing

After sequencing the 20 pools (180 libraries including several which had to be repeated), an average percentage of 73% of the wells in each chip were loaded and the range between all chips was 36% to 89%. Mean polyclonality was calculated to be 27% (16%-45%). On average, 443 megabases were sequenced in each sequencing run (ranging from 147 to 604 megabases). Overall 8859 megabases (8.859.000.000 bp) were aligned to the human genome version 19 (hg19) by the Torrent suite™ software.

### 3.4 Variant detection

#### 3.4.1 DNASTar software and Torrent Variant Caller variants

Using DNASTar software parameters, 3757 variants were found. On the other hand, using the Torrent Variant Caller with software settings for germline high stringency, 1712 variants were found. Minimum coverage required in this setting was 20, and variants with less coverage have been excluded. The number of variants found with the Torrent Variant Caller software utilizing germline low stringency settings was 2276. Minimum coverage required in this set of parameters was 6x.

#### 3.4.2 Comparison of variants found using high stringency, low stringency and DNASTar software and a description of occurred problems

Table 3.4.1 displays the numbers of common variants between different variant calling procedures. A table of all 1655 common variant calls among all three analyses for every position was made. Variant calls of each procedure were in columns beside each other. Consequently, a comparison of variant calls was facilitated.

Table 3.4.1 Number of common variants between two Torrent Variant Caller settings and DNASTar software.

	DNASTar	Torrent variant caller	
		High stringency	Low stringency
DNASTar	3757	1655	1757
High stringency	1655	1712	1712
Low stringency	1757	1712	2276

Quantities in the intersection of columns and rows indicate the number of common variants found among the settings represented by the selected column and row

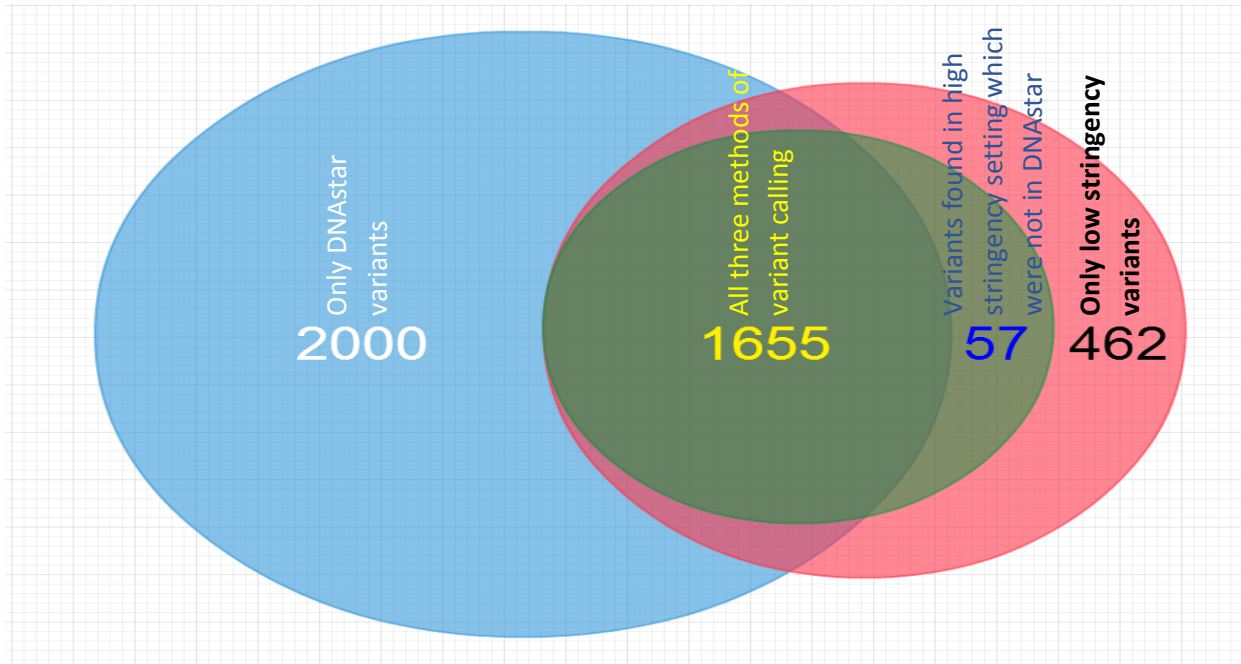


Figure 3.4.1 Common variants between two Torrent Variant Caller settings and DNASTar software. For a better understanding a Venn diagram is depicted. The blue circle represents the DNASTar software variants. The green and the red circles represent variants found by Torrent Variant Caller settings high stringency and low stringency respectively.

At this point considering 1655 positions in 139 samples, 230045 base calls were present for analysis. Approximately 27% of calls were discrepant between DNASTar and either one of the Torrent Variant Caller settings. On the other hand, the percentage of discrepant calls was only 2.6% in between two Torrent Variant Caller settings: high and low stringency. Observation and analysis of the discrepancies between DNASTar and two other settings revealed shortcomings for DNASTar such as wrong calls (including several examples of calling a certain allele in absence of even one read showing the called allele). Therefore, we excluded the DNASTar analysis from further analysis. After comparison, 5849 discrepancies were found between high and low stringency settings.

Several types of problems appeared when the discrepant calls were investigated. We decided to analyze the discrepancies between two settings of Torrent Variant Caller settings starting from the subject number 201 continuing to get a better understanding of discrepancies' origin. After we were observing repetitive problems we stopped the checking process. As a sample, 13.7% (798) of discrepancies were surveyed manually one after another. The integrative genome viewer view for each of these discrepancies was investigated. In 39.5% of the cases the variant call from high stringency setting was more confidently acceptable. These were the cases which confidence of call was low. In 16.7% of cases the coverage was between 10 and 20 (in the high stringency setting minimum coverage acceptable was 20 and in low stringency minimum was 10). Therefore, high stringency setting had no call and on the other hand the low stringency had the call which was without apparent error. In remaining 43.8%, making an acceptable call was not possible considering the quality and depth of coverage at those points (Table 3.4.2). In these instances, high stringency setting had a "no call" result and we decided to consider it the right call in order not to introduce false positive variants.

Table 3.4.2 Analysis of discrepancies between Torrent Variant Caller settings high and low stringency.

<b>Origin of discrepancy</b>	<b>Number (percent) in 798 discrepancies reanalyzed manually</b>	<b>Action taken</b>
<b>Low quality of the reads according to the quality score</b>	316 (39.5%)	Accepting high stringency
<b>Between 10 to 20x coverage</b>	134 (16.7%)	Decided to change the minimum coverage parameter in custom setting
<b>Imbalance between positive and negative reads in heterozygote positions</b>	371 (43.8%)	High stringency calls were accepted to avoid false positive

The following is the description of typical problems encountered. A table of comparison for each position's call by each procedure along with an integrative genomic viewer outlook of the position is provided for each problem.

3.4.2.1 Problem one: Homopolymers

One of the shortcomings of Ion torrent sequencing is the homopolymer problem. This problem occurs when there is a repeat of the same base. Due to the method of base calling in Ion torrent sequencing, distinction between numbers of bases in homopolymeric repeat regions is less confidently possible. In the following example the reference base is G and the alternate one is T. Total read count is 92 times and 14% of the reads show G and 84% show a T in this position. There is a five-time G repetition in the region; therefore, the number of G bases was not well detected due to the mentioned homopolymer problem. Consequently, the T is occasionally placed in the further position. DNASTar software and Torrent Variant Caller setting high stringency have a T/T call; however, the low stringency has the wrong G/T call. In this case variant caller is unable to call the variant correctly. High stringency has more stringent parameters therefore it has a better discrimination in homopolymer cases.

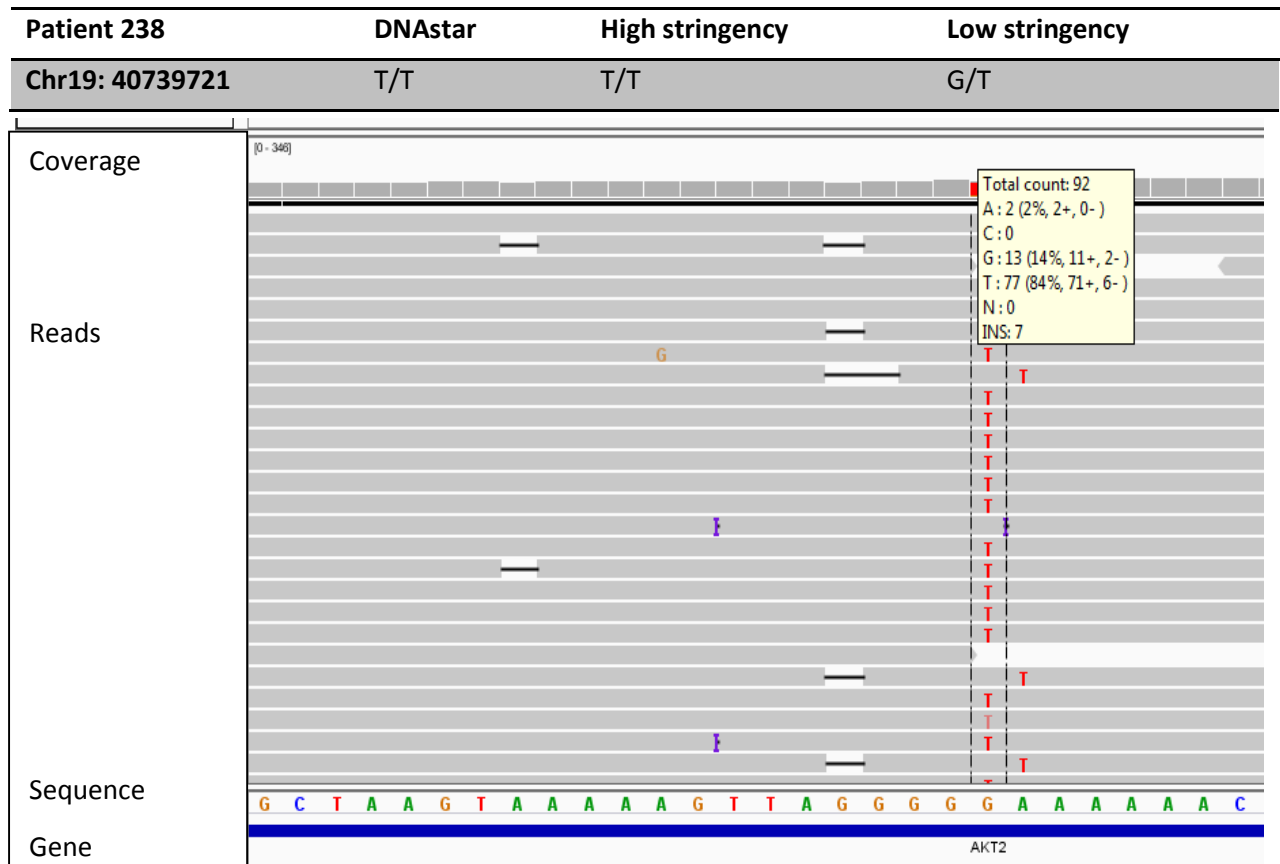


Figure 3.4.2 The integrative genomics view of a problem occurred due to a homopolymer region. Wrong G insertions are detected by the variant caller due to this problem.



Number of “no call” results was significantly higher in five and four homopolymer regions in comparison to regions with less or no homopolymers (P value < 0.001, independent sample T test). Quality score is a confidence level in which the variant calling was performed. It is determined by the torrent suite and the sequencing machine. This score is equal to  $-10 \log_{10}$  (error rate), meaning that in case 1 out of 10 calls are wrong – thus the confidence is 90% – therefore the quality score is equal to 10 (Q10). In case of homo four-mers mean number of no calls and average quality are 37.96 and 59.87 respectively in comparison to 24.63 and 60.10 in regions with less than four homo-mers (numbers and figures are based on custom parameter settings’ calls which will be explained later). Figure 3.4.3 shows the association between existence of regions of five-mer and four-mers and higher number of “no call” results and the lower average quality of the variant calling. However high stringency setting had a right variant calling in the represented problem; we have excluded regions with homopolymers of more than 3 base from our study.

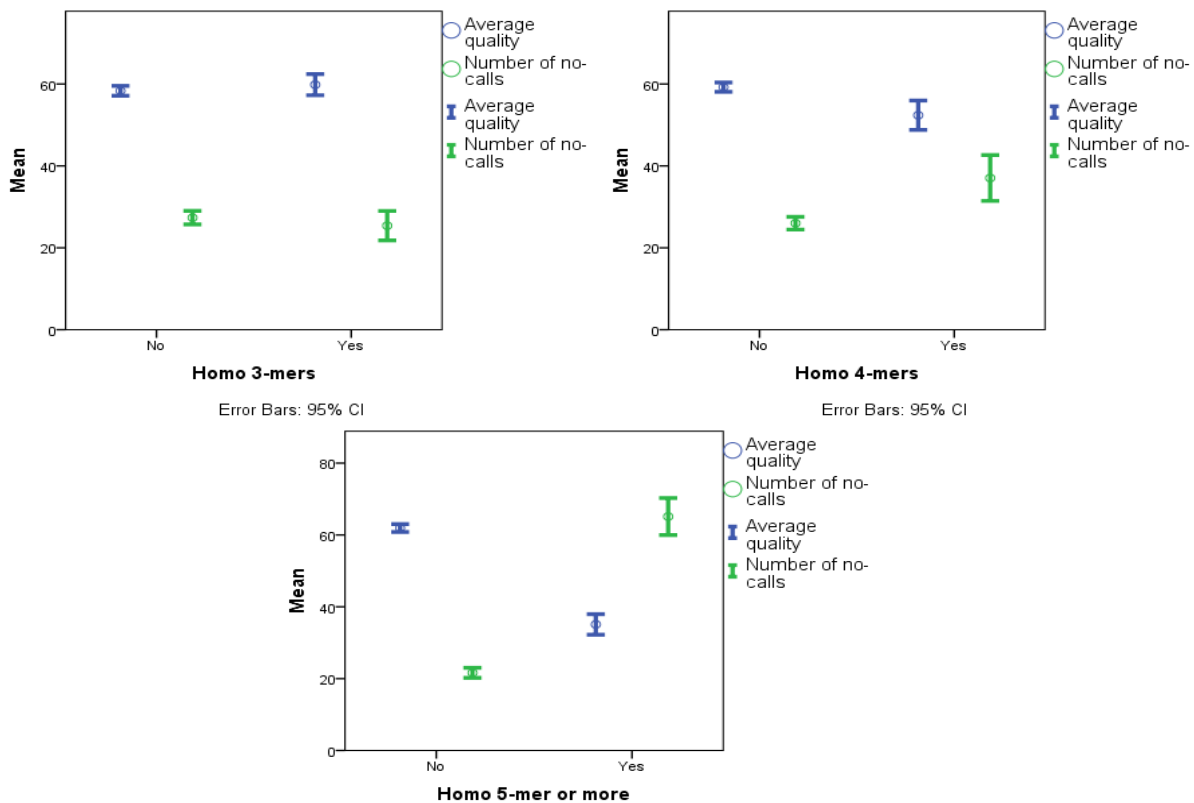


Figure 3.4.3 Homopolymer as reason for low quality and missing calls. This figure displays error bars of association between homopolymer regions and quality of calls along as well as with non-missing calls. As it is displayed in this figure, in case of homo 3-mers number of no calls and average quality are quite similar. But in case of homo 4-mers or more, quality was significantly dropped and the number of no calls was increased.

### 3.4.2.2 Problem two: Low coverage

In the example shown below, depth of coverage is 12x (more than 10 which is the considered limit for an acceptable coverage) (Wheeler et al. 2008). The reference allele was C in this position. All 12 reads were T and the quality of the reads was adequate (judged by the color of reads in integrative genomics viewer). In this case, high stringency setting had no call and other configurations have a T/T call. The reason of not having a call in high stringency setting is the minimum 20x required coverage was set in the parameters. To overcome this problem, minimum coverage was set to 10 in the custom parameter setting which was used later for variant calling.

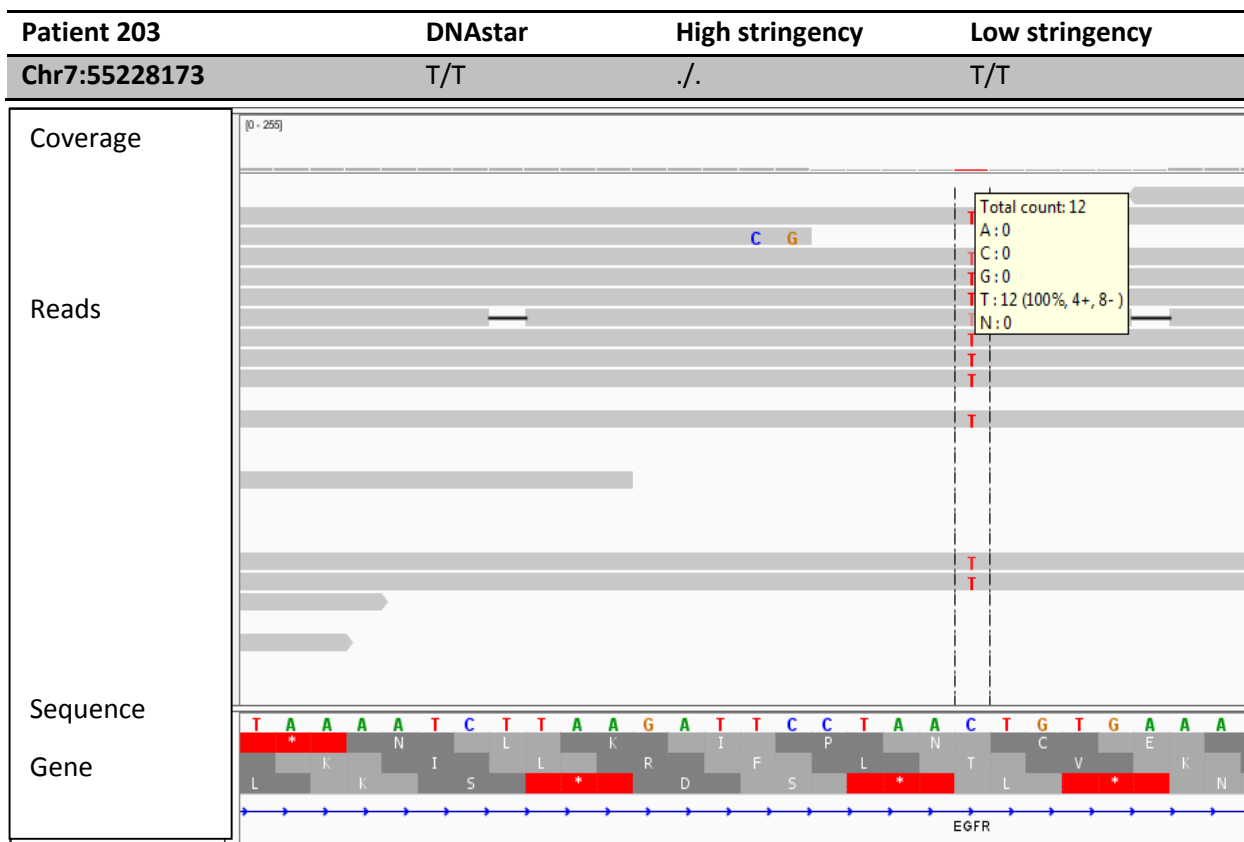


Figure 3.4.4 The integrative genomics view of a problem occurred due to a low coverage. Twelve reads were achieved in this instance, which are enough for variant calling. However the high stringency setting of Torrent Variant Caller do not achieve a result due to its parameter setting and reports no call.

### 3.4.3 Custom parameters as a solution

Considering the mentioned problems, we have changed minimum coverage of Torrent Variant Caller high stringency settings to 10x and made a custom parameter setting. The custom parameters are shown in Table 2.10.1. In addition, as mentioned earlier, the homopolymer region of more than four-mer, were excluded. Using this settings and after exclusion of homopolymer regions, 1538 variants were found. From this number, 353 were newly found and had no entry in dbSNP 138.

### 3.5 Validation of the consistency of genotypes found using deep sequencing

To validate the consistency of variant detection by deep sequencing method, we have used four previously genotyped SNP positions which were also in our target region. This analysis was performed by our collaborators in Bonn (Parmar et al. 2013). These positions were analyzed by KASPar single nucleotide polymorphism genotyping system (KBioscience, Hoddesdon, UK) using a real-time PCR system. These SNPs were rs2227983 in *EGFR* gene and rs1130214, rs1130233 and rs3730358 in *AKT1* gene. After comparison we have found only one discrepancy in rs2227983. A single patient has been detected to be heterozygote by KASPar method but found to be homozygote variant by deep sequencing at this position. The genotypes regarding this position in other samples and other three positions genotyped were identical comparing two genotyping methods.

Table 3.5.1 Consistency of genotyping by deep sequencing tested by previously genotyped positions. Genotypes regarding four SNPs: rs2227983, rs1130214, rs1130233 and rs3730358 were compared.

SNP	rs2227983	rs1130214	rs1130233	rs3730358
Total number of comparisons	126	126	126	126
Number of mismatches	1	0	0	0

In addition, as *ARAF* gene is localized in chromosome X, all the heterozygote variations must have occurred in female patients. We used the 8 genotyped *ARAF* variations for checking the gender of the patients. There were no mismatches found in this survey.

### 3.6 Frequency of SNPs' missing calls in our sample population.

Of 139 genotyped patients' DNA samples, phenotypic data was available for 126 patients. Other samples were excluded from further analysis. To analyze the SNPs for their associations with phenotypes, we chose the SNPs which had more than or equal to 50% present calls (less than 50% missing calls) in the sample population. Of all found SNPs only 1437 were qualified. Histograms of missing genotypes are shown in Figure 3.6.1. Mean number of no calls was 6.49 and the median was 2.38. Eighty percent of SNPs had less than 10% present calls.

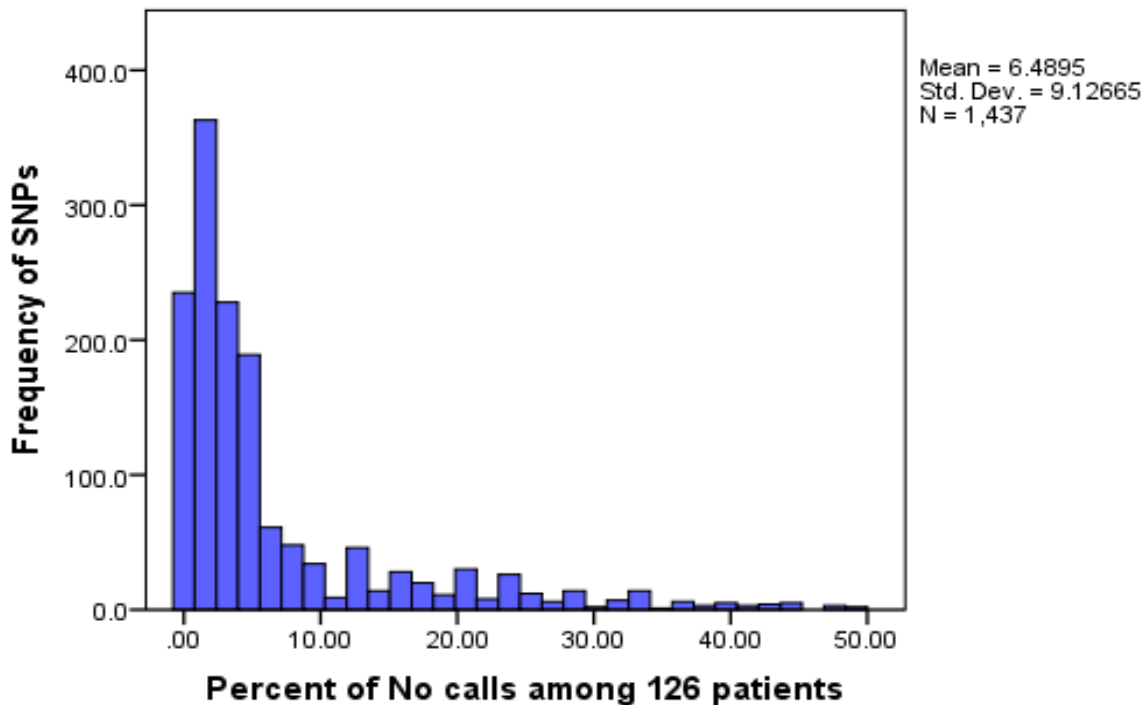


Figure 3.6.1 Histograms of missing (or No calls) among the SNPs. We have analyzed 126 patients and overall, we have identified 1437 Variants. The Y axis represents the number of positions. The X axis represents the percent of no calls in each position. As shown here, in most of the variant positions less than 10% of the calls are absent. And in very few positions 30% of the calls are absent.

#### 3.6.1 Hardy-Weinberg equilibrium

Allele frequencies were then subjected to Hardy-Weinberg equilibrium (HWE) testing. Nineteen SNPs were then marked because their allele distribution was out of HWE. Marking was performed to report it in case of presenting these variations' association with phenotypes.

### 3.6.2 Minor allele frequency (MAF) of the SNPs

In this study average MAF was 8.79%. Figure 3.6.2 shows a histogram of MAF percentages categorized by the novelty of SNPs. Six hundred and fifty one (45%) SNPs had less than 1% MAF (rare SNPs) and the SNPs with less than 5% were 936 (65%). Three hundred and seventeen (22%) novel SNPs were found. Mean MAF of novel SNPs was 0.46% (ranging 0.40%-1.98%). MAF of exonic SNPs categorized by their functionality (data from SIFT software) are displayed in Figure 3.6.3. Total number of coding SNPs was 103 of which 68 were synonymous and 38 were nonsynonymous. Average MAF of noncoding SNPs was 9.02% while average MAF of coding SNPs

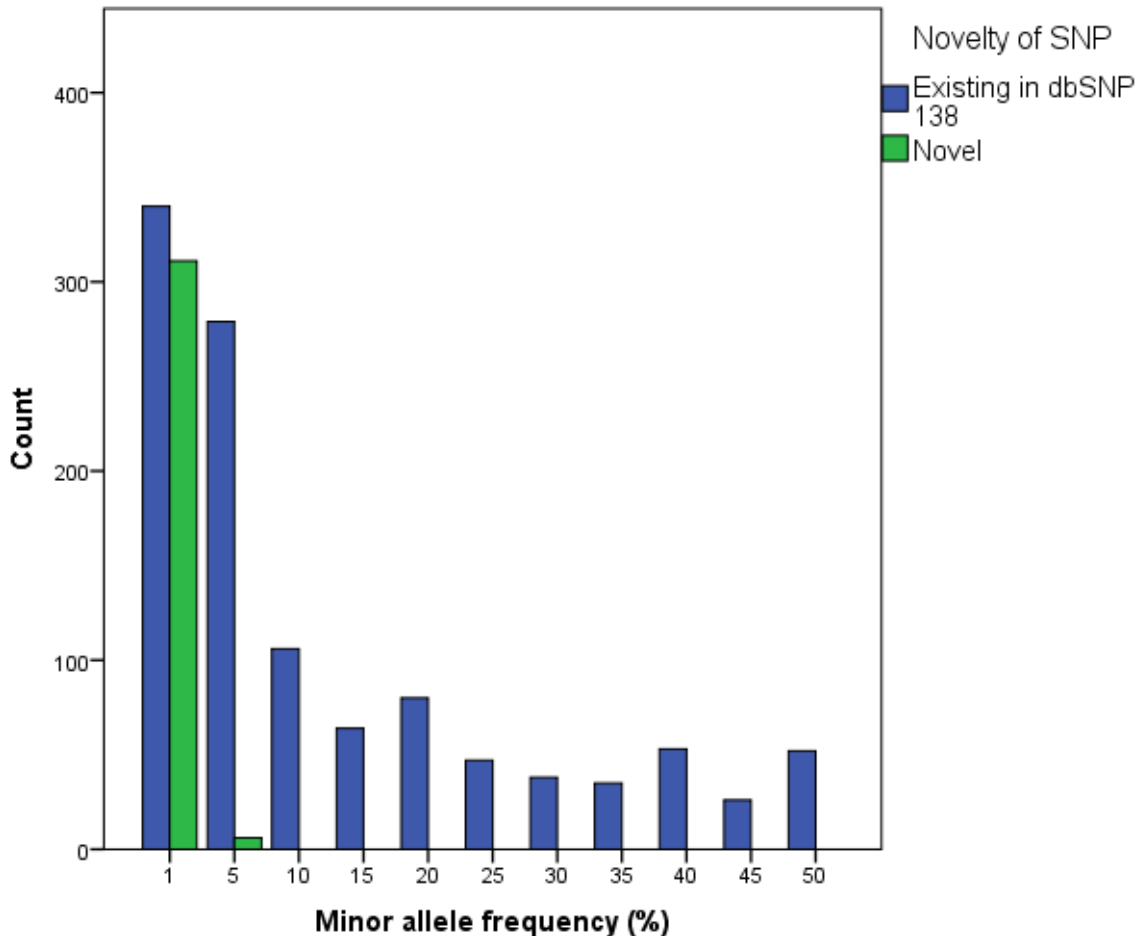


Figure 3.6.2 Minor allele frequencies (MAF) of SNPs categorized by novelty of SNPs. As displayed above, all of the frequent SNPs were already recorded in the databases while the very rare variants were new. The number of singleton variant was 560. The number of variants with less than 1% MAF was 651.

was 6.15%. Average MAF of synonymous and nonsynonymous SNPs were 8.04% and 3.34% respectively. Only one nonsense SNP was found in MTOR with the MAF of 1.60%.

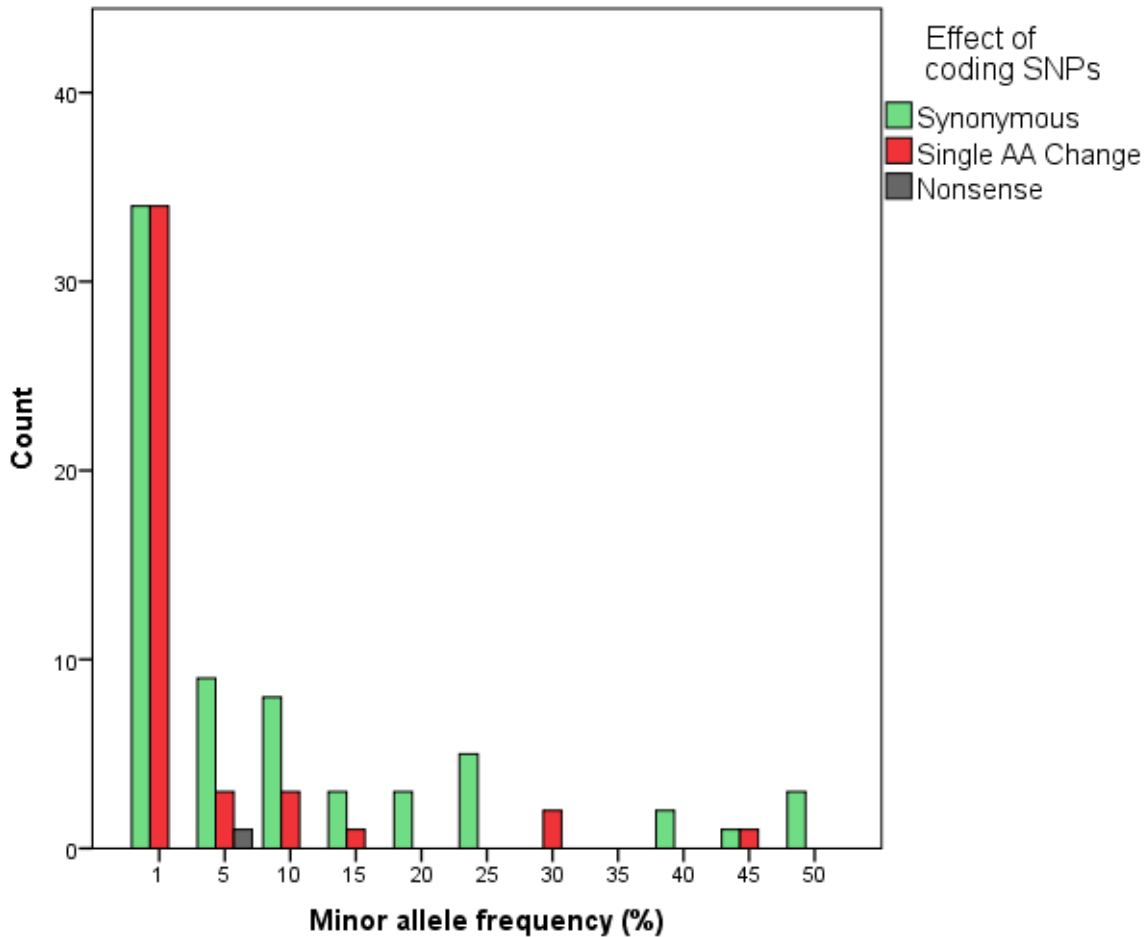


Figure 3.6.3 Minor allele frequency of all exonic SNPs (n= 113, 7.9% of total variations found) categorized by their effect. Categorization is done by SIFT software.

### 3.7 Functionality of genetic variants

Functionality of the variants was predicted by SIFT which is one of the most frequently used bioinformatics tools. According to this prediction, the SNPs were categorized into noncoding, synonymous, single amino acid change and nonsense.

The *EGFR* gene was completely sequenced; therefore the highest number of SNPs was found in this gene. Of 913 variants found in *EGFR* gene region, 212 SNPs were novel. Out of novel SNPs

only 1 caused a single amino acid substitution and had a MAF of 0.4%. *MTOR* had the second highest number of variations followed by *AKT1* gene. Forty eight variants were found in *MTOR*. Four of *MTOR* variations cause amino acid substitution; while 1 deleterious amino acid substitution was predicted by both SIFT and Provean. Prediction of effects of amino acid exchanging variants was performed by Provean and SIFT prediction softwares. The result of this prediction is summarized in Table 3.7.1. Although 38 variants were found in *AKT1*; no amino acid substitution variation has been found. *CXCR2*, *NFKB1* and *RPS6KB2* wielded 4 amino acid substitution variations which was the highest number among genes. Mean minor allele frequencies of these variations were 0.6%, 0.4% and 1.7% in *CXCR2*, *NFKB1* and *RPS6KB2* genes respectively. Number of variations in each 1000 bp of targeted region is shown in Figure 3.7.1. Highest variation ratio was found in *MAP2K2*, followed by *RPS6KB2* and *EGFR*. *MAP2K2* – as the highest – had about 6 variations in 1000 bp. On the other hand, *PTEN* with 0.7 variations had the lowest variation ratio.

Table 3.7.1 Number of variants categorized by their effect (noncoding, synonymous, single amino substitution). Only 1 premature stop-codon variant existing in dbSNP 138 was found in *MTOR* gene which is not shown in this table.

Pathway	Gene name	Target region size	Total variants	Existing in dbSNP 138			Novel		
				Ncod.	NonSyn.	Syn.	Ncod.	NonSyn.	Syn.
The receptor	<i>EGFR</i>	196300	913	688	2	11	209	1	2
PI3K/Akt	<i>AKT1</i>	8423	38	32	0	3	3	0	0
	<i>AKT2</i>	10291	32	22	2	2	6	0	0
	<i>AKT3</i>	9384	20	10	0	0	10	0	0
	<i>MTOR</i>	16138	48	28	2	8	6	2	2
	<i>PIK3CA</i>	6201	15	10	1	3	1	0	0
	<i>PIK3CB</i>	7060	12	8	0	0	3	1	0
	<i>PIK3R1</i>	12762	37	28	2	4	2	1	0
	<i>PIK3R2</i>	8439	17	8	1	5	2	1	0
	<i>PTEN</i>	7106	5	2	0	1	2	0	0
	<i>RPS6KB1</i>	7916	10	8	0	0	2	0	0
<i>RPS6KB2</i>	3996	20	8	3	1	6	1	1	
Ras/MAPK	<i>ARAF</i>	4444	5	5	0	0	0	0	0
	<i>BRAF</i>	5146	9	8	0	1	0	0	0
	<i>HRAS</i>	2228	5	2	0	0	2	0	1
	<i>KRAS</i>	7873	25	21	1	1	2	0	0
	<i>MAP2K1</i>	4702	16	16	0	0	0	0	0
	<i>MAP2K2</i>	4419	27	13	1	3	10	0	0
	<i>MAPK1</i>	7330	24	16	0	2	4	2	0

Pathway	Gene name	Target region size	Total variants	Existing in dbSNP 138			Novel		
				Ncod.	NonSyn.	Syn.	Ncod.	NonSyn.	Syn.
	MAPK3	4112	11	8	0	1	2	0	0
	NRAS	5554	11	7	0	1	2	1	0
	RAF1	5825	22	12	0	1	8	1	0
Inflammatory response regulators	NFKB1	8640	26	16	4	2	3	0	1
	NFKB2	6538	19	12	0	2	3	1	1
	REL	4085	6	5	0	0	1	0	0
	RELA	4897	12	9	2	0	0	1	0
	RELB	3879	16	10	0	2	3	1	0
Cytokines and their receptors	CXCR1	3503	6	2	4	0	0	0	0
	CXCR2	4718	14	7	2	3	1	1	0
	IL8	2505	2	2	0	0	0	0	0

Nonsyn.: nonsynonymous amino acid substitution; Ncod.: noncoding; Syn.: Synonymous.

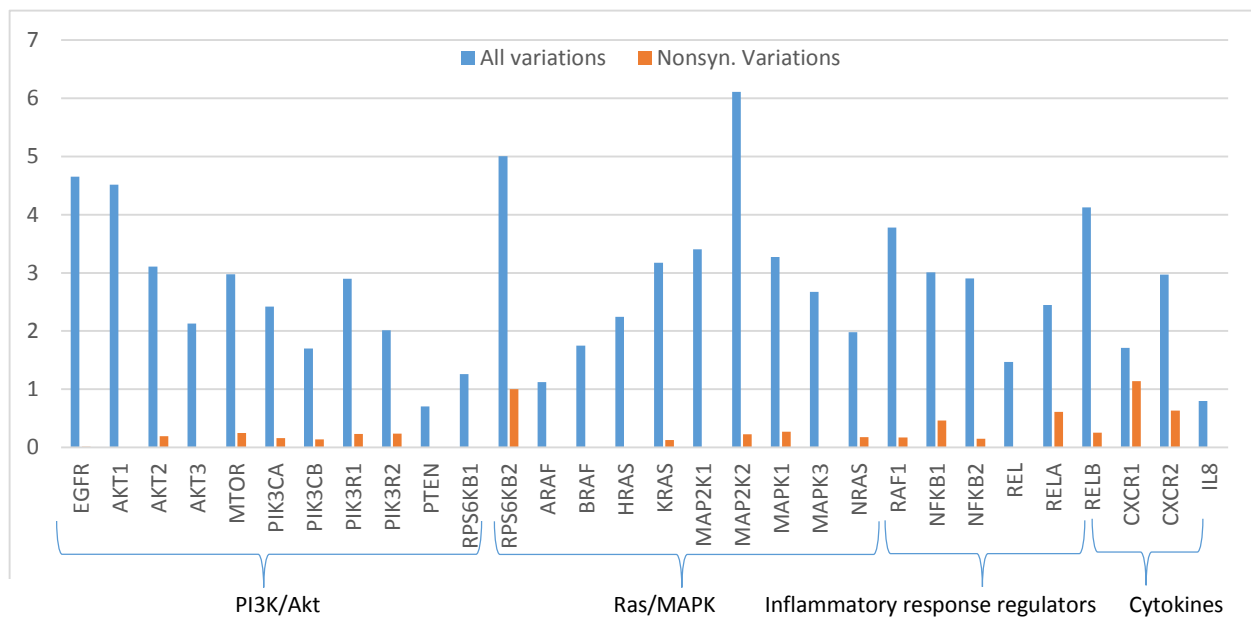


Figure 3.7.1 Number of variations in 1000 bp of targeted region categorized by genes. As well as total number of variations, number of nonsynonymous variations are also shown (in orange).



Table 3.7.2 Function prediction of effects of nonsynonymous SNPs. Prediction tools used are explained in methods section 2.11.

	Chr	Position	Gene name	ID	Codon change	aa substitution	Provean	SIFT	Polyphen2	MAF
The receptor	7	55233024	<i>EGFR</i>	rs144943614	TGC [G/A]TC AAG	V592I	Tol	Tol	Dam	0.004
	7	55229255	<i>EGFR</i>	rs2227983	CCC A[G/A]G GAC	R521K	Tol	Tol	Tol	0.272
	7	55231472	<i>EGFR</i>		CAC [C/A]CA GAG	P560T	Dam	Dam	Tol	0.004
PI3K/Akt	19	40739513	<i>AKT2</i>	rs33933140	GAG A[T/C]T CAG	I136T	Tol	Dam	---	0.420
	19	40745923	<i>AKT2</i>	rs12980468	CAG T[G/A]T GAG	C223Y	Dam	Tol	---	0.004
	3	138384011	<i>PIK3CB</i>		GAT [C/T]GC TCT	R847C	Dam	Dam	Dam	0.004
	3	178927410	<i>PIK3CA</i>	rs2230461	GAT AT[A/G] TAC	I391M	Tol	Tol	Tol	0.056
	5	67569229	<i>PIK3R1</i>	rs367923318	CTC [C/A]CG GAT	P116T	Tol	Dam	Dam	0.004
	5	67588148	<i>PIK3R1</i>	rs3730089	AAT AT[G/A] TCC	M326I	Tol	Tol	Tol	0.122
	5	67588351	<i>PIK3R1</i>		GAA C[T/C]G GAA	L13P	Tol	Dam	---	0.004
	19	18271881	<i>PIK3R2</i>		GAC [G/A]TG GAT	V162M	Tol	Dam	Dam	0.004
	19	18273047	<i>PIK3R2</i>	rs1011320	GCC [T/C]CC ACA	S313P	Tol	Tol	Tol	0.025
	1	11186750	<i>MTOR</i>		ATT C[G/A]C ATT	R2152H	Dam	Dam	Tol	0.004
	1	11199647	<i>MTOR</i>	rs144984370	CCT CA[T/G] GAA	H1647Q	Tol	Tol	Dam	0.004
	1	11308007	<i>MTOR</i>	rs35903812	AAT [G/A]CC TTG	A329T	Tol	Tol	Tol	0.004
	1	11319413	<i>MTOR</i>		AGC AA[T/G] GTG	N18K	Tol	Tol	Tol	0.004
	11	67200123	<i>RPS6KB2</i>		ATA C[G/A]G CCT	R121Q	Tol	Dam	---	0.004
	11	67200812	<i>RPS6KB2</i>	rs55987642	TCG C[C/T]G CCC	P267L	Dam	Tol	Dam	0.056
11	67201490	<i>RPS6KB2</i>	rs200667089	CAG [C/T]GG ATT	R311W	Dam	Dam	Dam	0.004	
11	67202519	<i>RPS6KB2</i>	rs374535834	CCC A[C/T]G GAG	T443M	Tol	Tol	Tol	0.004	
Ras/MAPK	12	25368380	<i>KRAS</i>	rs201170656	ATA [A/C]TG	M189L	Tol	Tol	Tol	0.005
	19	4097323	<i>MAP2K2</i>	rs151133017	AGC C[G/A]G CCT	R313Q	Tol	Tol	Dam	0.004
	22	22123569	<i>MAPK1</i>		TTG [GA/TC]T GAC	D336S	Dam	Dam	Dam	0.004
	1	115251173	<i>NRAS</i>		TTG [C/T]CA TGT	P185S	Tol	Tol	Tol	0.004
	3	12647779	<i>RAF1</i>		ACT [A/G]TT GGT	I201V	Tol	Tol	Tol	0.004
Inflammatory response regulators	4	103518700	<i>NFKB1</i>	rs4648072	GCT [A/G]TG CAG	M507V	Tol	Tol	Tol	0.004
	4	103518782	<i>NFKB1</i>	rs150281816	CAG C[G/A]C CAT	R534H	Dam	Dam	Dam	0.004
	4	103527745	<i>NFKB1</i>	rs149211506	CGC TT[G/T] GGT	L615F	Tol	Tol	Tol	0.004
	4	103533226	<i>NFKB1</i>	rs201233299	AAA C[C/G]A TAT	P793R	Tol	Tol	Dam	0.004
	10	104161516	<i>NFKB2</i>		TCA [C/T]TT GGT	L770F	Tol	Dam	Dam	0.004
	11	65421964	<i>RELA</i>	rs369008511	CCA G[C/G]T CCT	A511G	Tol	Dam	Tol	0.004
	11	65425764	<i>RELA</i>	rs61759893	CCA [G/A]AT ACA	D288N	Dam	Tol	Dam	0.004
	11	65427175	<i>RELA</i>		CTC C[G/A]C CTG	R171H	Tol	Tol	Dam	0.004
	19	45540956	<i>RELB</i>		ACC [G/A]CC AGC	A550T	Tol	Dam	Dam	0.004
Cytokines and their receptors	2	219028932	<i>CXCR1</i>	rs16858808	CAT [C/T]GT GTT	R335C	Dam	Tol	Dam	0.024
	2	219029108	<i>CXCR1</i>	rs2234671	GAG A[G/C]C TGT	S276T	Tol	Tol	Tol	0.061
	2	219029843	<i>CXCR1</i>	rs16858811	TGT A[T/G]G CTA	M31R	Tol	Tol	Tol	0.028
	2	219029932	<i>CXCR1</i>	rs55808468	AT[G/A] TCA	M1I	Tol	Tol	Tol	0.004
	2	218999666	<i>CXCR2</i>		AAC [A/C]AG TAT	K48Q	Tol	Tol	Dam	0.004
	2	219000230	<i>CXCR2</i>	rs10201766	CTG [C/T]GT ACG	R236C	Dam	Tol	Tol	0.008
	2	219000267	<i>CXCR2</i>	rs61733609	CAC C[G/A]G GCC	R248Q	Dam	Dam	Dam	0.008

Tol stands for tolerated, meaning without major effect on protein function expected and Dam stands for damaging with major effects on protein function expected. The algorithms for protein function prediction were Provean (Choi et al. 2012), SIFT (Kumar et al. 2009), and Polyphen2 (Adzhubei et al. 2010).

As shown in Table 3.7.2, from a total of 41 non-synonymous amino acid substitutions, 30 occurred only once in a heterozygous state and 2 occurred twice. Six SNPs occurred less than 15

and more than 5 times in a heterozygous state (MAF of 1% to 5%). Two of these three SNPs occurred once in heterozygote state and one occurred only once.

Twenty six variants were predicted to be damaging using at least one of the prediction tools Table 3.7.2. According to consistent bioinformatics predictions from all three of Provean, SIFT, and Polyphen2 there were only 4 variants which were predicted to be associated with a loss of function by all three algorithms. Regarding the variants which also occurred in homozygote state, two were predicted to be tolerated by all three of tools and only one was predicted to be damaging by Provean (AKT2, I136T amino acid substitution). Eight amino acid substitutions were predicted to be damaging by two of out of three tools. Of these variations, except for RPS6KB2 P267L which had the MAF of 0.056, others only happened once.

### 3.8 Runs of homozygosity

While screening the minor allele frequencies of rare variants, we surprisingly found several rare variants which occurred in homozygous form. By visualizing these variations we have found that some of these rare SNPs occurred in one patient. Surveying the genotypes of this patient we found a region with runs of homozygosity (ROH) in patient 278 (Figure 3.8.1) at the gene locus coding for the EGFR. In the patient with the ROH we did not find even a single heterozygous variant in the *EGFR* region. A histogram of heterozygous genotype counts are shown in Figure 3.8.2. However other patients had at least 31 heterozygous genotypes at this region (On average, 106 heterozygous variants were found among the patients).

The coverage at this region was also investigated to estimate the quality at this position. Furthermore genotypes of other subjects with similar coverage at this region were investigated. Several heterozygous genotypes were found in patients with similar coverage in the *EGFR* region.

We have determined if there were any particular phenotypes in subject 278 due to the ROH. Maximum skin rash in this patient was grade 1 (apparently not much different from the total sample mean of 1.2). Diarrhea did not occur in this patient (total sample mean 0.6) and the survival time was only 111 days after treatment start date (total sample mean 234). This patient

was enrolled in this study having grade IIIa of bronchial cancer and received erlotinib.

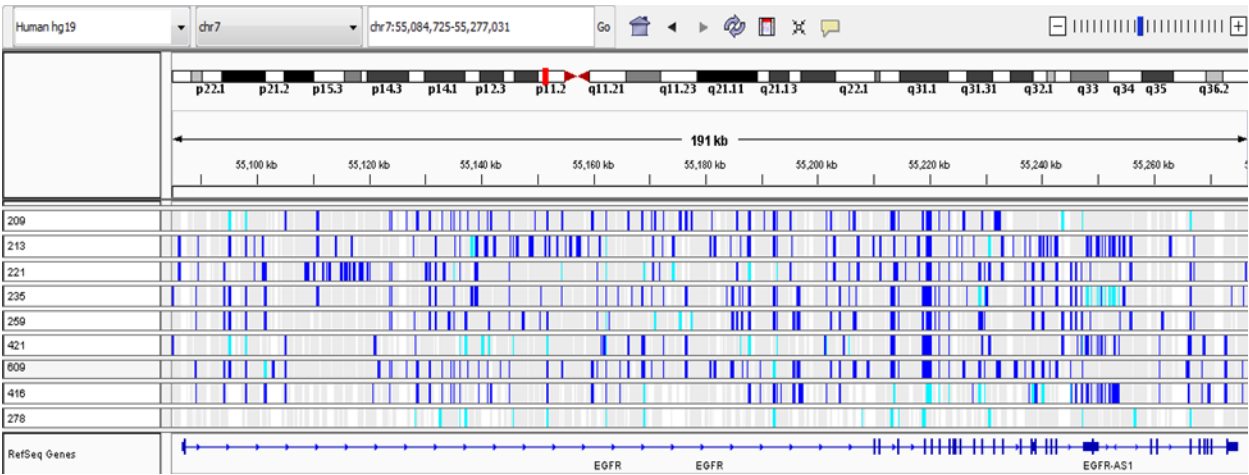


Figure 3.8.1 Runs of homozygosity. Using the integrative genomic viewer it is displayed here that patient number 278 shows no heterozygous genotypes. Gray lines represent regions with homozygous reference genotypes. Dark and light blue lines represent heterozygous and homozygous variant genotypes, respectively. EGFR gene region is displayed.

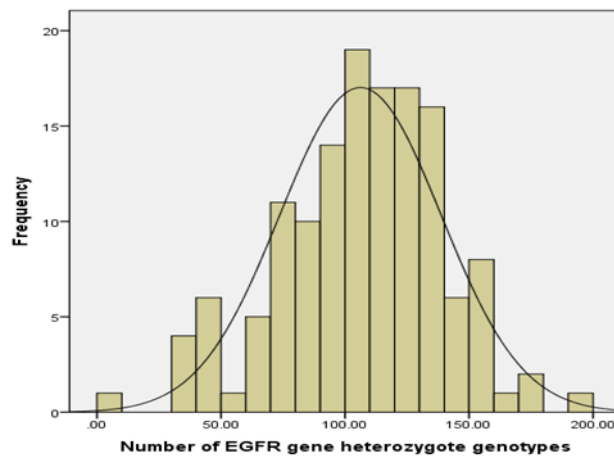


Figure 3.8.2 Histogram of heterozygous genotype counts in the present study. Sample 278 had no variations which is an outlier in this histogram.

### 3.9 Association of SNPs with skin rash

#### 3.9.1 First analysis phase: Genomic variants predisposing to skin rash

Twenty six variants were found to have an association with the existence of level 1 or greater skin rash using a genotypic test, using the conventional 0.05 type-I-error without adjusting for multiple testing. However, none of these SNPs achieved significance after correction for multiple testing (level of significance considering 1437 variation is 0.000035). The Manhattan plot of this association is shown in Figure 3.9.1. The intronic *EGFR* variant rs73420732 had the lowest P value (0.0013) in the genotypic test as well in the allelic test (0.0015). This SNP was in LD with 4 other significant *EGFR* SNPs namely: rs57620875, rs7784637, rs28557040 and rs17335759 while considering a  $r^2 > 0.4$  as cutoff level, all these 4 other SNPs were also in intron 1. The LD plot of these SNPs is shown in Figure 3.9.2 . A SNP in *EGFR-AS1* (*EGFR* antisense RNA 1, rs17290461) which overlaps with the *EGFR* gene was also associated with lack of skin rash.

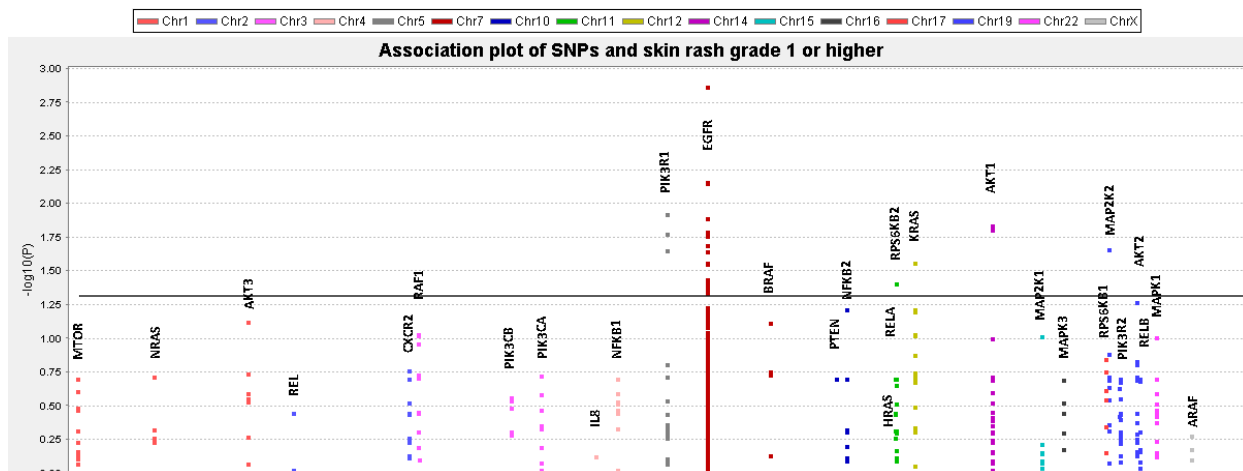


Figure 3.9.1 Manhattan plot of association of SNPs found in this study and the presence of skin rash grade  $\geq 1$ . P value of Fisher's exact test was used for making this plot by Plink software.  $-\log_{10}(0.05)$ , the level of significance equals 1.30 and is shown by the black horizontal line. The required level of significance after correction for multiple testing was 0.000035 ( $-\log_{10}$  of this value equals 4.45)

A zoomed-in association plot for *EGFR* gene is shown in Figure 3.9.2. From the significant *EGFR* SNPs five sets of linked ones were identified with  $r^2 > 0.4$  as cutoff level. First set includes the rs73420732, rs57620875, rs7784637, rs28557040 and rs17335759 localized in intron 1. Other sets were rs2293348, rs940810, and rs9692301 along with SNPs rs1468727 and rs4947984 as another set; as well as two other sets rs6970262 and rs1404908 besides rs7796872 and rs10488141. These SNPs were all localized in noncoding regions.

In the order of decreasing significance, the fourth significant SNP was rs79397497 in *PIK3R1* gene (P value = 0.012). This 3' UTR SNP was linked to the other *PIK3R2* significant SNP rs3729984 which was also in the 3' UTR region.

Two linked SNPs in *AKT1* gene (both in intron 1) were significantly associated to the presence of rash. These two SNPs were rs61757057 and rs61757056 with P values 0.014 and 0.015 respectively.

Table 3.9.1 Germline SNPs associated with presence of skin toxicity of level 1,2 or 3 with P value < 0.05. SNPs with significant allelic test associations are also shown in this table. Colors represent the linkage between variations. Colors of EGFR variant correspond to the colors used in Figure 3.9.2.

SNP Region		Genotype based analysis			Allele based analysis		
Gene name	dbSNP ID	No rash group <sup>a</sup>	Rash group <sup>a</sup>	P value*	No rash group <sup>b</sup>	Rash group <sup>b</sup>	P value
EGFR	rs73420732	21/4/0	99/0/0	1.4E-03	46/4	198/0	1.5E-03
EGFR	rs57620875	21/3/0	98/0/0	6.9E-03	45/3	196/0	7.2E-03
EGFR	rs7784637	22/3/0	101/0/0	7.1E-03	47/3	202/0	7.4E-03
PIK3R1	rs79397497	14/6/0	85/6/1	1.2E-02	34/6	176/8	2.2E-02
EGFR	rs7781264	17/5/1	39/49/10	1.3E-02	39/7	127/69	8.1E-03
AKT1	rs61757057	21/3/1	98/2/0	1.5E-02	45/5	198/2	4.1E-03
AKT1	rs61757056	21/3/1	96/2/0	1.5E-02	45/5	194/2	4.4E-03
EGFR	rs28557040	20/5/0	95/4/0	1.6E-02	45/5	194/4	1.8E-02
PIK3R1	rs3729984	17/6/0	92/6/1	1.7E-02	40/6	190/8	3.0E-02
EGFR	rs6970262	10/6/7	32/55/12	1.7E-02	26/20	119/79	7.4E-01
EGFR	rs41367446	16/3/0	81/1/0	2.0E-02	35/3	163/1	2.2E-02
MAP2K2	rs6630	18/1/2	83/13/0	2.2E-02	37/5	179/13	3.3E-01
PIK3R1	rs3730089	17/1/1	68/24/0	2.2E-02	35/3	160/24	5.9E-01
EGFR	rs17335759	20/3/0	95/1/0	2.3E-02	43/3	191/1	2.4E-02
KRAS	rs61764367	20/4/0	95/3/0	2.7E-02	44/4	193/3	2.9E-02
EGFR	rs2293348	11/7/5	44/49/5	2.8E-02	29/17	137/59	3.8E-01
EGFR	rs1468727	7/13/0	59/31/6	2.8E-02	27/13	149/43	2.2E-01
EGFR	rs17290461	22/2/0	100/0/0	3.6E-02	46/2	200/0	3.7E-02
EGFR	rs78134465	22/2/0	100/0/0	3.6E-02	46/2	200/0	3.7E-02
EGFR	rs1404908	9/7/7	33/50/10	3.8E-02	25/21	116/70	3.2E-01
RPS6KB2	rs201739335	23/2/0	100/0/0	3.9E-02	48/2	200/0	3.9E-02
EGFR	rs4947984	24/0/0	85/16/0	4.0E-02	48/0	186/16	4.7E-02
EGFR	rs4947965	11/6/3	40/42/2	4.1E-02	28/12	122/46	8.5E-01
EGFR	rs10234806	6/12/1	55/31/4	4.2E-02	24/14	141/39	6.1E-02
EGFR	rs940810	13/6/4	45/36/3	4.6E-02	32/14	126/42	4.6E-01
EGFR	rs9692301	9/6/5	37/42/6	4.6E-02	24/16	116/54	3.5E-01
AKT2	rs33933140	12/11/1	29/45/21	5.4E-02	35/13	103/87	2.2E-02
EGFR	rs7796872	15/7/2	80/20/1	6.1E-02	37/11	180/22	3.4E-02
EGFR	rs6947531	5/7/11	32/45/22	6.2E-02	17/29	109/89	3.3E-02
KRAS	rs1137188	1/8/6	18/47/11	6.3E-02	10/20	83/69	4.5E-02
EGFR	rs4140770	13/4/1	37/43/10	6.4E-02	30/6	117/63	3.2E-02
MAPK1	rs13515	17/3/0	51/31/4	9.7E-02	37/3	133/39	2.9E-02
EGFR	rs10488141	12/8/2	75/20/4	1.0E-01	32/12	170/28	4.3E-02
EGFR	rs6593202	23/2/0	70/26/2	1.1E-01	48/2	166/30	3.4E-02

<sup>a</sup>, Homozygote reference genotype/Heterozygotes/Homozygote variant genotype; <sup>b</sup>, Minor allele/reference allele; \* P values were calculated using Fisher's exact test

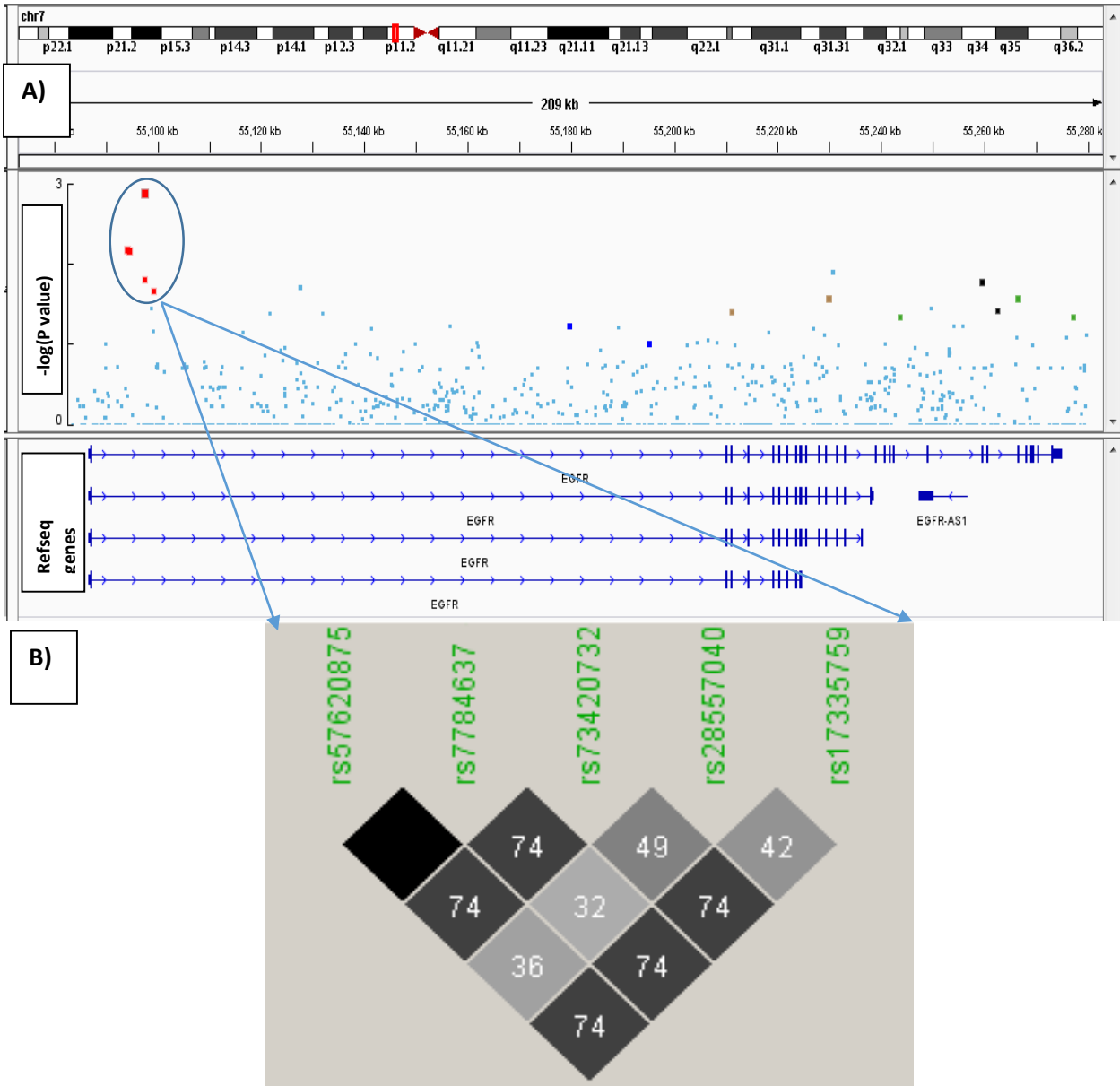


Figure 3.9.2 Association plot and LD of EGFR SNPs and existence of rash. A) Five sets of EGFR SNPs were associated to rash of grade 1 or higher. The lowest P value belonged to rs73420732 which was in LD with rs57620875, rs7784637, rs28557040 and rs17335759. This set is shown in red. Other sets are as follows: rs6970262 and rs1404908: black; rs2293348, rs940810, rs9692301: green; rs1468727 and rs4947984: brown; rs7796872 and rs10488141: blue; B) LD plot of rs73420732; SNP with the lowest P value for association to existence of skin rash. Four other significantly associated SNPs were in LD with this SNP namely rs57620875, rs7784637, rs28557040 and rs17335759.  $r^2$  values are shown. Black square represents  $r^2$  of 100.

### 3.9.2 Second analysis phase: Genomic variants predisposing to high grade skin rash

We have also surveyed the association of high grade rash (grade 2 or higher) with the identified variations. Twenty eight SNPs had P value < 0.05 using Fisher's exact test. The Manhattan plot of SNPs' association to high grade rash is shown in Figure 3.9.4. The lowest P value (0.0012) belonged to rs131515 in 3' UTR region of *MAPK1* gene located on chromosome 22. Four other SNPs in *MAPK1* gene were significantly associated with high grade rash using an allelic test namely: rs6928 (3' UTR), rs9607340 (upstream of the gene), rs9340 (3' UTR), and rs13943 (3' UTR). These 4 SNPs were in LD; however, they were not linked to the SNP with lowest P value (Figure 3.9.3).

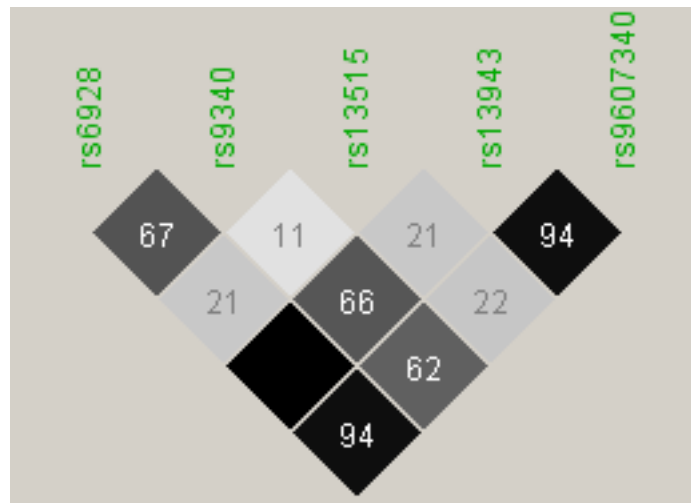


Figure 3.9.3 Linkage Disequilibrium plot of significant SNPs in *MAPK1* gene. rs13515 has the highest association with skin rash grade 2 or higher. rs6928, rs9607340, rs9340, and rs13943 were linked together but not to the most significant SNP.  $r^2$  values. Black square represents  $r^2$  of 100.

Table 3.9.2 demonstrates the SNPs with highest significance level. Two *EGFR* SNPs rs6593202 and rs17335891 had the second and third lowest P values of 0.0042 and 0.0079 respectively. A zoomed-in association plot for *EGFR* gene is shown in Figure 3.9.5. Two sets of *EGFR* gene SNPs were found among the significantly association SNPs: rs4947965, rs11773818, and rs11766798 were linked; as well as SNPs rs62459768 and rs6593207. Both of these sets are localized in the intron 1. Significant SNPs were also found in *AKT1* (rs45598737), *MAP2K2* (rs17851657), *NFKB1* (rs72696119), *AKT3* (rs14403), *BRAF* (rs3789806), *KRAS* (rs1137188 and rs712), *AKT2* (rs33933140), and *PIK3R1* (rs171649) genes.



Table 3.9.2 Germline SNPs associated with presence of skin toxicity of level 2 and 3 with P value < 0.05. SNPs with significant allelic test associations are also shown in this table. Colors represent the linkage between variations. Colors of EGFR variant correspond to the colors used in Figure 3.9.5.

SNP Region		Genotype based analysis			Allele based analysis		
Gene name	dbSNP ID	Low rash group <sup>a</sup>	High rash group <sup>a</sup>	P value	Low rash group <sup>b</sup>	High rash group <sup>b</sup>	P value
MAPK1	rs13515	21/22/3	47/12/1	1.18E-03	64/28	106/14	8.81E-04
EGFR	rs6593202	33/18/2	60/10/0	4.22E-03	84/22	130/10	2.07E-03
EGFR	rs17335891	17/30/1	38/23/6	7.88E-03	64/32	99/35	2.43E-01
AKT1	rs45598737	38/13/0	65/5/1	8.21E-03	89/13	135/7	3.40E-02
EGFR	rs4947965	16/28/1	35/20/4	1.14E-02	60/30	90/28	1.60E-01
EGFR	rs11773818	18/29/1	38/23/6	1.31E-02	65/31	99/35	3.75E-01
EGFR	rs4947488	18/29/1	38/23/6	1.31E-02	65/31	99/35	3.75E-01
EGFR	rs6970262	15/34/4	27/27/15	1.34E-02	64/42	81/57	8.95E-01
EGFR	rs3800827	19/28/1	38/21/6	1.35E-02	66/30	97/33	3.69E-01
EGFR	rs11766798	19/28/1	38/22/6	1.46E-02	66/30	98/34	3.74E-01
EGFR	rs2280653	37/15/0	58/9/5	1.51E-02	89/15	125/19	8.52E-01
EGFR	rs9642564	22/30/1	40/22/6	1.60E-02	74/32	102/34	3.86E-01
MAP2K2	rs17851657	34/11/5	48/22/0	1.77E-02	79/21	118/22	3.10E-01
EGFR	rs62459768	41/12/0	64/5/1	1.77E-02	94/12	133/7	9.03E-02
EGFR	rs6593207	34/19/0	60/12/1	2.03E-02	87/19	132/14	6.00E-02
EGFR	rs12668421	21/30/1	41/25/6	2.33E-02	72/32	107/37	3.92E-01
EGFR	rs79198232	52/1/0	61/10/1	2.39E-02	105/1	132/12	8.74E-03
EGFR	rs6956366	14/34/5	33/28/10	2.42E-02	62/44	94/48	2.33E-01
EGFR	rs56208956	37/10/1	65/5/0	2.44E-02	84/12	135/5	1.84E-02
NFKB1	rs72696119	16/11/4	12/30/11	2.91E-02	43/19	54/52	2.37E-02
EGFR	rs1404908	16/30/3	26/27/14	2.95E-02	62/36	79/55	5.86E-01
EGFR	rs11760406	21/22/1	40/17/6	3.15E-02	64/24	97/29	5.21E-01
AKT3	rs14403	33/18/0	34/27/7	3.17E-02	84/18	95/41	3.35E-02
EGFR	rs6978771	21/27/1	41/22/5	3.33E-02	69/29	104/32	3.65E-01
BRAF	rs3789806	29/18/1	50/11/1	3.82E-02	76/20	111/13	3.75E-02
KRAS	rs1137188	11/25/3	8/30/14	4.06E-02	47/31	46/58	3.67E-02
EGFR	rs76138885	44/6/0	68/1/0	4.06E-02	94/6	137/1	4.36E-02
EGFR	rs6947902	30/19/3	54/12/6	4.11E-02	79/25	120/24	1.96E-01
EGFR	rs17172446	39/7/4	42/24/5	4.39E-02	85/15	108/34	1.05E-01
EGFR	rs12535226	8/32/9	24/30/14	4.58E-02	48/50	78/58	2.32E-01
EGFR	rs723526	31/19/3	54/12/6	4.64E-02	81/25	120/24	1.98E-01
MAPK1	rs6928	21/24/5	18/31/18	4.70E-02	66/34	67/67	1.65E-02
EGFR	rs12673110	15/31/7	34/27/7	4.90E-02	61/45	95/41	5.81E-02
MAPK1	rs9607340	21/22/6	17/31/18	6.17E-02	64/34	65/67	1.62E-02
MAPK1	rs9340	8/27/11	21/26/8	6.46E-02	43/49	68/42	3.44E-02
KRAS	rs712	15/25/5	11/35/16	6.65E-02	55/35	57/67	3.74E-02
AKT2	rs33933140	11/25/12	30/31/10	6.77E-02	47/49	91/51	2.32E-02
EGFR	rs7796872	45/8/0	50/19/3	7.79E-02	98/8	119/25	2.43E-02
EGFR	rs4947963	15/23/5	34/24/3	8.50E-02	53/33	92/30	4.59E-02
MAPK1	rs13943	21/24/6	18/33/17	9.63E-02	66/36	69/67	3.50E-02
PIK3R1	rs171649	21/25/6	19/34/18	1.03E-01	67/37	72/70	3.75E-02
EGFR	rs980653	42/10/0	49/20/4	1.13E-01	94/10	118/28	4.87E-02
AKT2	rs2304189	30/18/3	29/27/11	1.30E-01	78/24	85/49	3.40E-02

<sup>a</sup>, Homozygote reference genotype/Heterozygotes/Homozygote variant genotype; <sup>b</sup>, Minor allele/reference allele; P values were calculated using Fisher's exact test

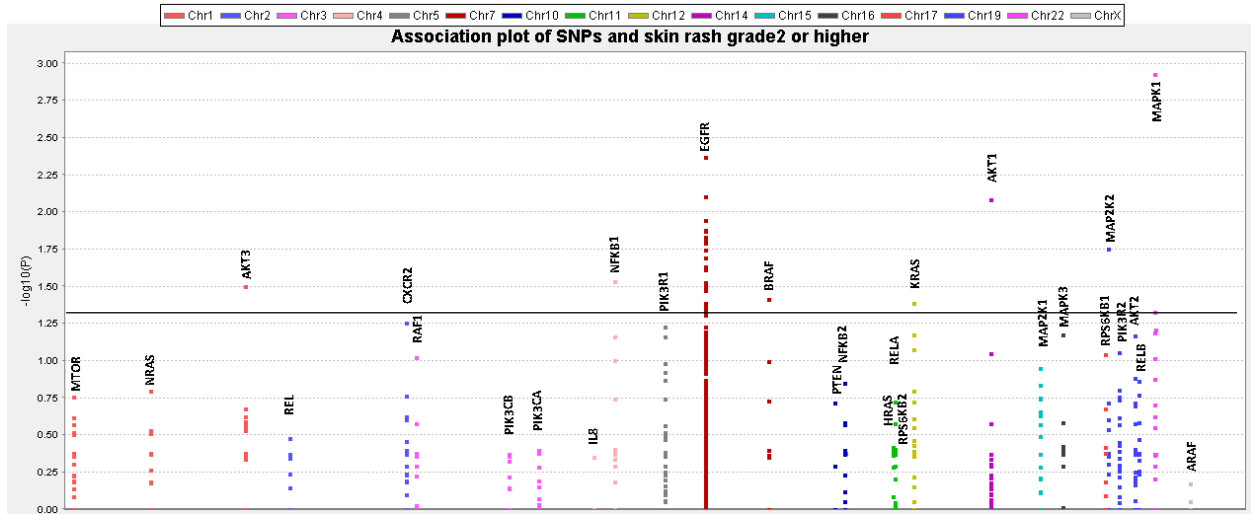


Figure 3.9.4 Manhattan plot of association of SNPs found in this study and the presence of skin rash grade  $\geq 2$ .  $-\log_{10}$  (0.05, level of significance) equals 1.30 and is shown by the black horizontal line. P value of Fisher's exact test was used for making this plot.

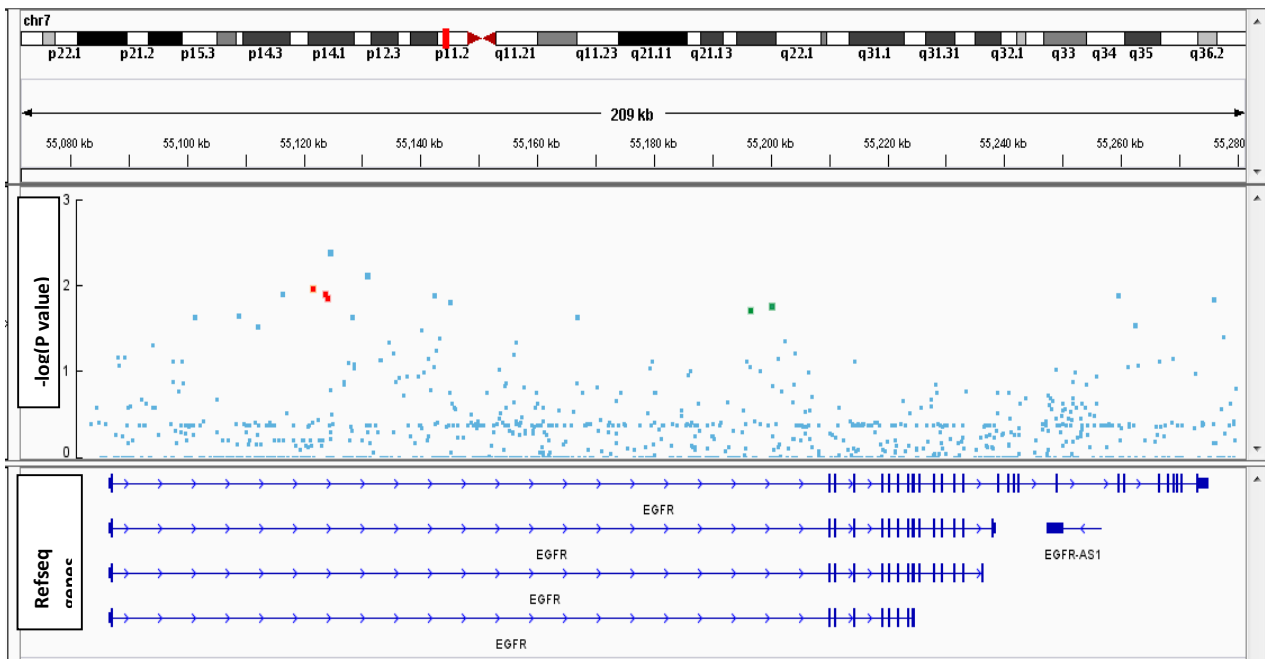


Figure 3.9.5 Association plot of EGFR SNPs and existence of high grade rash. Two sets of EGFR SNPs were associated to rash of grade 2 or higher. These sets are: rs4947965, rs11773818, and rs11766798: red; rs62459768 and rs6593207: green. Both of these sets reside in intron 1.

### 3.10 Gene-based and pathway-based analyses

Gene-based and pathway-based analyses were performed using the PSEQ software. The variations which were analyzed in this analysis were damaging variants found by any of the three (SIFT, Polyphen2 or Provean) variant effect prediction tools and had minor allele frequency of less than 5%. Considering these criteria 23 variations were eligible to this analysis. The gene burden test, the C-alpha test, the frequency-weighted test, the variable threshold test, the UNIQ and summary statistics tests were performed.

In pathway-based analysis with dichotomization of the skin rash phenotype as any grade vs. no rash, an association was found between PI3K/Akt pathway variations using C-alpha and summary statistics tests (P values 0.03 and 0.04, respectively). Among 8 variations considered in this pathway 4 singletons were found in skin rash group and 4 other were found in no rash group. As these tests are permutation-based a minimum P value possible (the I value) is also calculated for each test. In case of PI3K/Akt pathway this value for both C-alpha and summary statistics was 0.002.

In gene-based analysis (using genes as units of comparison) an association has not been found while using either one of skin rash phenotype dichotomization strategies. The significance level was calculated 0.0017 after correction for multiple testing. However, none of the I values (representing the minimum P value possible considering sample size and number of variants) has reached this level.

*Table 3.10.1 Pathway-based association analysis. Both skin reaction related phenotype dichotomizations are shown in this table. Values are reported using gene burden test.*

	<b>Number of variations considered</b>	<b>P value</b>	<b>I value</b>	<b>Description case to control</b>
<b>Skin rash grade 0 compared to 1-3</b>				
<b>PI3K/Akt + EGFR</b>	10	1	0.20	6/4
<b>PI3K/Akt</b>	8	0.83	0.20	4/4
<b>Ras/MAPK +EGFR</b>	4	0.71	0.67	4/0
<b>Ras/MAPK</b>	2	0.36	0.67	2/0
<b>Inflammatory response regulators</b>	7	0.83	0.40	5/2
<b>Chemokines</b>	4	0.58	0.27	8/3

	Number of variations considered	P value	I value	Description case to control
<b>Inflammatory response regulators + Chemokines</b>	11	0.83	0.20	13/5
<b>All pathways</b>	23	0.99	0.01	21/9
<b>Skinrash grades 0-1 compared to 2-3</b>				
<b>PI3K/Akt + EGFR</b>	10	0.83	0.20	4/6
<b>PI3K/Akt</b>	8	1	0.40	2/6
<b>Ras/MAPK +EGFR</b>	4	0.47	0.19	2/2
<b>Ras/MAPK</b>	2	0.83	0.20	0/2
<b>Inflammatory response regulators</b>	7	1	0.20	1/6
<b>Chemokines</b>	4	1	0.40	3/8
<b>Inflammatory response regulators + Chemokines</b>	11	1	0.20	4/14
<b>All pathways</b>	23	0.80	0.07	8/22

### 3.11 Diarrhea as another side effect

Four *EGFR* SNPs had the smallest P values in the analysis for association with existence of diarrhea (equal or more than grade 1) as another side effect. Three of these SNPs (rs11770531 rs11770689 and rs11760524) are positioned in *EGFR* intron 6 and one (rs940806) is located downstream of the *EGFR* gene in an intergenic region. A variation in 3'UTR region of *NRAS* gene (rs14804) and two intronic *NFKB1* SNPs were also significantly associated with diarrhea. However none of these variations reached the significance level when the P value was adjusted by the number of SNPs.

Table 3.11.1 Ten most significant germline SNPs associated with existence of diarrhea using Fisher's exact test.

Gene name	SNP Region				Genotype based analysis			Allele based analysis		
	dbSNP ID	Chr.	Position	Var.	No diarrhea group <sup>a</sup>	Diarrhea group <sup>a</sup>	P value	No diarrhea group <sup>b</sup>	Diarrhea group <sup>b</sup>	P value
<i>EGFR</i>	rs11770531	7p11.2	55220905	T/C	41/2/1	61/19/0	3.56E-03	84/4	141/19	6.79E-02
<i>EGFR</i>	rs11770689	7p11.2	55221469	T/C	37/1/1	57/15/0	6.15E-03	75/3	129/15	1.22E-01
<i>EGFR</i>	rs940806	7p11.2	55279745	A/G	20/17/6	15/49/13	8.52E-03	57/29	79/75	2.98E-02
<i>EGFR</i>	rs11760524	7p11.2	55221167	G/A	41/3/1	61/19/0	1.31E-02	85/5	141/19	1.21E-01
<i>NRAS</i>	rs14804	1p13.2	115249843	A/G	20/10/1	21/33/9	1.36E-02	50/12	75/51	4.97E-03
<i>EGFR</i>	rs62459767	7p11.2	55198347	T/C	40/1/0	64/15/1	1.37E-02	81/1	143/17	7.98E-03
<i>NFKB1</i>	rs3817685	4q24	103534560	G/C	24/19/1	31/31/15	1.58E-02	67/21	93/61	1.62E-02
<i>NFKB1</i>	rs4648050	4q24	103514741	C/T	19/18/1	26/31/15	2.23E-02	56/20	83/61	1.95E-02
<i>EGFR</i>	rs2740764	7p11.2	55267458	G/T	29/12/2	34/41/3	2.31E-02	70/16	109/47	6.58E-02
<i>EGFR</i>	rs11506105	7p11.2	55220177	A/G	12/7/12	17/49/9	2.65E-02	41/41	83/67	4.92E-01

Chr. Chromosome; a, Homozygote reference genotype/Heterozygotes/Homozygote variant genotype; b, Minor allele/reference allele;

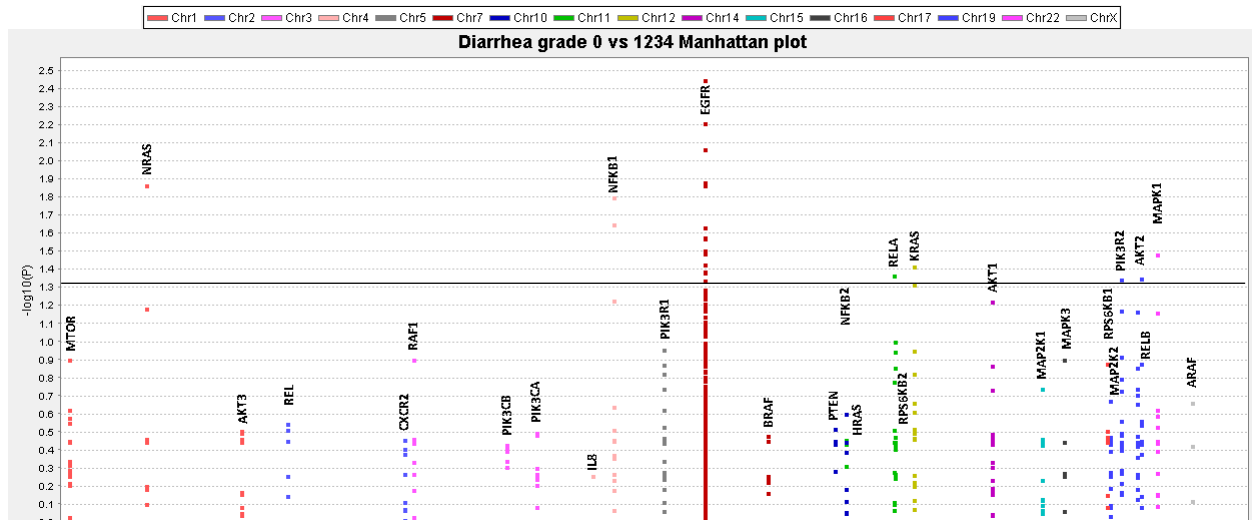


Figure 3.11.1 Manhattan plot of association of SNPs found in this study and the presence of diarrhea.  $-\log_{10}(0.05)$ , level of significance equals 1.30 and is shown by the black horizontal line. P value of Fisher's exact test was used for making this plot.

### 3.12 Survival and effect sizes in carriers of non-synonymous variants

It is well known, that skin rash is related to survival in patients receiving EGFR inhibition therapy. Concerning survival, it was not possible to perform statistical testing for significance with most of the rare variants. However, it may still be informative to present the mean days of survival in carriers of the amino acid substitution variants. This is presented in Table 3.12.1.

The majority of nonsynonymous variations occurred in the PI3K/Akt pathway. Among them two amino acid substitutions in RELA (D288N) and RAF1 (I201V, a novel variation) both occurring in one certain patient were associated with a very low survival (29 days). Other amino acid substitution variations in MTOR (H1647Q) and RPS6KB2 (T443M) were as well associated with a low survival time (19 and 35 days respectively). Nevertheless all of these variations were rare variations and the association with short survival does not proof causality.

Considering survival time, patients with the *PIK3R1* exon 8 M326I variation show a trend for a better survival for carriers of the isoleucine amino acid. While mean survival time was 234 days in our study, heterozygote patients for *PIK3R1* M326I variation (N=25) had 252 days and the homozygote patient had 314 days of survival time. Interestingly, this variation was also

associated to presence of skin rash. In concert with the association of survival and skin rash, 24 out of 26 patients with this variation were affected with skin rash.

Table 3.12.1 Survival time, skin rash and diarrhea in relation to the amino acid substitutions identified in my project. In the second row (total sample mean) mean of survival time, maximum skin rash grade, and maximum diarrhea grade from all patients is shown.

				Provean	SIFT	Polyphen2	N	Survival time	Max skin rash	Max diarrhea	
<b>Total sample mean</b>								<b>126</b>	<b>234</b>	<b>1.2</b>	<b>0.6</b>
	Gene name	aa substitution	Zygosity*								
<b>The receptor</b>	EGFR	V592I	1	Tol	Tol	Dam	1	360	2	2	
	EGFR	R521K	1	Tol	Tol	Tol	55	230	1.1	0.5	
			2				6	226	1.7	0.8	
	EGFR	P560T	1	Dam	Dam	Tol	1	360	2	1	
<b>PI3K/Akt</b>	AKT2	I136T	1	Tol	Dam	---	56	231.7	1.3	0.6	
			2				22	253	1.5	0.5	
	AKT2	C223Y	1	Dam	Tol	---	1	360	2	1	
	PIK3CB	R847C	1	Dam	Dam	Dam	1	360	1	0	
	PIK3CA	I391M	1	Tol	Tol	Tol	11	210.1	1.3	0.5	
	PIK3R1	P116T	1	Tol	Dam	Dam	1	151	2	0	
	PIK3R1	M326I	1	Tol	Tol	Tol	25	252.4	1.7	0.6	
			2				1	314	0	1	
	PIK3R1	L13P	1	Tol	Dam	---	1	360	0	0	
	PIK3R2	V162M	1	Tol	Dam	Dam	1	360	0	1	
	PIK3R2	S313P	1	Tol	Tol	Tol	6	280.7	1.5	1	
	MTOR	R2152H	1	Dam	Dam	Tol	1	360	2	1	
	MTOR	H1647Q	1	Tol	Tol	Dam	1	19	0	0	
	MTOR	A329T	1	Tol	Tol	Tol	1	233	0	0	
	MTOR	N18K	1	Tol	Tol	Tol	1	360	2	0	
	RPS6KB2	R121Q	1	Tol	Dam	---	1	360	1	0	
	RPS6KB2	P267L	1	Dam	Tol	Dam	14	205.5	1.1	0.6	
	RPS6KB2	R311W	1	Dam	Dam	Dam	1	100	0	4	
RPS6KB2	T443M	1	Tol	Tol	Tol	1	35	2	2		
<b>Ras/MAPK</b>	KRAS	M189L	1	Tol	Tol	Tol	1	127	2	0	
	MAP2K2	R313Q	1	Tol	Tol	Dam	1	144	1	0	
	MAPK1	D336S	1	Dam	Dam	Dam	1	90	1	0	
	NRAS	P185S	1	Tol	Tol	Tol	1	360	1	0	
	RAF1	I201V	1	Tol	Tol	Tol	1	29	0	1	
<b>Inflammatory response regulators</b>	NFKB1	M507V	1	Tol	Tol	Tol	1	106	1	0	
	NFKB1	R534H	1	Dam	Dam	Dam	1	358	1	2	
	NFKB1	L615F	1	Tol	Tol	Tol	1	360	1	0	
	NFKB1	P793R	1	Tol	Tol	Dam	1	360	1	2	
	NFKB2	L770F	1	Tol	Dam	Dam	1	358	1	2	
	RELA	A511G	1	Tol	Dam	Tol	1	193	1	0	
	RELA	D288N	1	Dam	Tol	Dam	1	29	0	1	
	RELA	R171H	1	Tol	Tol	Dam	1	278	1	0	
RELB	A550T	1	Tol	Dam	Dam	1	90	2	0		

				Provean	SIFT	Polyphen2	N	Survival time	Max skin rash	Max diarrhea
<b>Total sample mean</b>							<b>126</b>	<b>234</b>	<b>1.2</b>	<b>0.6</b>
	Gene name	aa substitution	Zygoty* <sup>1</sup>							
Cytokines and their receptors	CXCR1	R335C	1	Dam	Tol	Dam	6	185.8	1	0.7
	CXCR1	S276T	1	Tol	Tol	Tol	15	251.3	1.7	0.7
	CXCR1	M31R	1	Tol	Tol	Tol	7	172.1	1	0.6
	CXCR1	M1I	1	Tol	Tol	Tol	1	285	2	0
	CXCR2	K48Q	1	Tol	Tol	Dam	1	360	3	0
	CXCR2	R236C	1	Dam	Tol	Tol	2	220	0.5	1
	CXCR2	R248Q	1	Dam	Dam	Dam	2	227.5	1	0

\* 1 = heterozygous, 2: homozygous; survival time is presented in days, Tol stands for tolerated, meaning without major effect on protein function expected and Dam stands for damaging with major effects on protein function expected. The algorithms for protein function prediction were Provean (Choi et al. 2012), SIFT (Kumar et al. 2009), and Polyphen2 (Adzhubei et al. 2010) .

To describe the possible effect of rare amino acid substitution variations, we have surveyed the survival, skin rash and diarrhea in the damaging variations with less than 5% minor allele frequency. A SNP was considered damaging when it was predicted to be damaging by any of the three prediction algorithms. As it is displayed in Table 3.12.2, when summarizing over all proteins we have not found a remarkable difference of phenotypes in carriers of these variations.

Table 3.12.2 Survival, skin rash and diarrhea in carriers of variants predicted to be damaging.

	Number	Survival time	Max skin rash	Max diarrhea
<b>Total group</b>	126	234.3	1.2	0.6
<b>All carriers with a variant predicted to be loss-of-function by Provean</b>	12	240.9	1.05	1.06
<b>All carriers with a variant predicted to be loss-of-function by SIFT</b>	14	266.2	1.07	0.79
<b>All carriers with a variant predicted to be loss-of-function by Polyphen2</b>	16	225.3	1.06	0.86
<b>All carriers with a variant predicted to be loss-of-function by all three tools</b>	4	261.4	0.75	1.5

In addition, for better displaying the function of these variations we have identified the protein localization of nonsynonymous variations in Table 3.12.3. For this reason we have surveyed the domain in the InterPro database (Hunter et al. 2012) and cross-checked the literature.

Table 3.12.3 Functional domains and possible function of the protein domain in which the nonsynonymous variations are localized. Conservative changes are defined as when an amino acid is replaced by another amino acid with the same polarity or electric charge. CXCR proteins are chemokine receptors which have seven membrane spanning regions. The position of variations on CXCR proteins is reported considering their position in this structure.

	Gene name	Substitution	aa substitution	Functional domain or protein region, possible effects
The receptor	EGFR	Cons.	V592I	Furin-like cysteine rich domain, EGFR dimerization and signal transduction
	EGFR	Cons.	R521K	Furin-like cysteine rich domain, EGFR dimerization and signal transduction
	EGFR	Noncons.	P560T	Furin-like cysteine rich domain, EGFR dimerization and signal transduction
PI3K/Akt pathway	AKT2	Noncons.	I136T	Linker region between PH(Pleckstrin Homology) domain and kinase domain, unknown effect
	AKT2	Noncons.	C223Y	Kinase domain, substrate phosphorylation
	PIK3CB	Noncons.	R847C	Kinase domain, substrate phosphorylation
	PIK3CA	Cons.	I391M	Membrane-binding domain (C2), Membrane recognition
	PIK3R1	Noncons.	P116T	Rho GTPase activating protein (RhoGAP), catalyzing the hydrolysis of GTP
	PIK3R1	Cons.	M326I	SH2(Src homology 2) domain, interacting with phosphotyrosine-containing target peptides
	PIK3R1	Noncons.	L13P	SH3 (Sarc homology 3) domain, assembly of kinase Subunit
	PIK3R2	Cons.	V162M	RhoGAP, catalyzing the hydrolysis of GTP
	PIK3R2	Noncons.	S313P	Linker region between RhoGAP and SH2 domain
	MTOR	Cons.	R2152H	Kinase domain, substrate phosphorylation
	MTOR	Noncons.	H1647Q	Focal adhesion kinase (FAT) domain
	MTOR	Noncons.	A329T	Armadillo type fold domain, Interaction with the rictor protein in the mTORC2 complex (regulation of Akt)
	MTOR	Noncons.	N18K	---
	RPS6KB2	Noncons.	R121Q	Kinase domain, substrate phosphorylation
	RPS6KB2	Noncons.	P267L	Kinase domain, substrate phosphorylation
	RPS6KB2	Noncons.	R311W	Kinase domain, substrate phosphorylation
RPS6KB2	Noncons.	T443M	Proline rich domain, binding to SH3 domains	
Ras/MAPK	KRAS	Cons.	M189L	The last amino acid, unknown effect
	MAP2K2	Noncons.	R313Q	Kinase domain, phosphorylation of ERK protein
	MAPK1	Noncons.	D336S	---
	NRAS	Noncons.	P185S	---
	RAF1	Cons.	I201V	Linker region between two kinase domains
Inflammatory response regulators	NFKB1	Cons.	M507V	Linker region between glycin rich domain and Ankyrin repeat domain
	NFKB1	Cons.	R534H	Linker region between glycin rich domain and Ankyrin repeat domain
	NFKB1	Cons.	L615F	Ankyrin repeat domain, NFKB inhibitor (IkBa) binds to this domain
	NFKB1	Noncons.	P793R	Linker region between Ankyrin repeat domain and death domain
	NFKB2	Cons.	L770F	Linker region between Ankyrin repeat domain and death domain



	<b>Gene name</b>	<b>Substitution</b>	<b>aa substitution</b>	<b>Functional domain or protein region, possible effects</b>
	<i>RELA</i>	Noncons.	A511G	---
	<i>RELA</i>	Noncons.	D288N	NF-kappa-B/Rel/Dorsal domain, protein dimerisation, DNA binding and nuclear localisation
	<i>RELA</i>	Cons.	R171H	NF-kappa-B/Rel/Dorsal domain, protein dimerisation, DNA binding and nuclear localisation
	<i>RELB</i>	Noncons.	A550T	---
<b>Cytokines and their receptors</b>	<i>CXCR1</i>	Noncons.	R335C	Cytoplasmic region of Chemokine receptor, G protein binding and catalyzing
	<i>CXCR1</i>	Cons.	S276T	Extracellular region of Chemokine receptor, acts as specific receptors for the CXCL8
	<i>CXCR1</i>	Noncons.	M31R	Extracellular region of Chemokine receptor, acts as specific receptors for the CXCL8
	<i>CXCR1</i>	Cons.	M1I	Extracellular region of Chemokine receptor, acts as specific receptors for the CXCL8
	<i>CXCR2</i>	Noncons.	K48Q	Extracellular region of Chemokine receptor, acts as specific receptors for the CXCL8
	<i>CXCR2</i>	Noncons.	R236C	Cytoplasmic region of Chemokine receptor, G protein binding and catalyzing
	<i>CXCR2</i>	Noncons.	R248Q	Cytoplasmic region of Chemokine receptor, G protein binding and catalyzing

\*The functional domain of some of variations were not identified because they were very close to either C or N terminal of the protein and their possible function has not been reported.

As summarized in Table 3.12.3, 26 of the 41 amino acid substitutions were non-conservative replacements. Most of these variants could be assigned to known functional domains of the respective proteins. Of the 41 variations, 7 occurred in kinase domains and may cause the stoppage of the kinase cascade downstream of EGFR signaling pathway. However, which of these variants really has functional effects in the respective pathway needs further experimental elucidation.

## 4 Discussion

The development of targeted therapy has promised a new era in cancer therapy by improving response rates with less adverse effects on normal cells. The estrogen receptor (ER) in ER positive breast cancer was probably the first target for targeted cancer therapy and nowadays individualized therapy targeting the ER is widely used in ER positive breast cancer (Ward 1973). Typical side effects of chemotherapy such as bone marrow suppression, alopecia and mucositis are often less prevalent in targeted therapies and may even be absent in case of monotherapy. EGFR was targeted for the first time in 1983 by Mendelsohn's group. They have developed an anti-EGFR antibody and successfully reduced the proliferation of EGFR bearing cancer cells both in vitro and in situ (against human tumor xenografts of squamous carcinoma cells and breast adenocarcinoma cells in athymic mice, both expressing EGFR) (Kawamoto et al. 1983, Masui et al. 1984a). However EGFR inhibition seems to have its own class of side effects such as skin toxicity. These side effects remarkably occur only in a subgroup of patients. Interestingly, the occurrence of skin toxicity has been proved to be correlated to drug response and survival rate in treated patients, but how much of this correlation is true or only due to confounding is still controversial (Petrelli et al. 2012).

### 4.1 First analysis phase: Genomic variants predisposing to skin rash

After exclusion of low quality sequences and erroneous variations, 1437 variants were included in the analysis in relation to skin toxicity. For the first analysis phase we dichotomized the patients by existence of skin rash phenotype into two groups: group of skin rash grade 0 and group of skin rash grades 1-3. As we had only 25 patients with no rash in the study population, number of variations in this group was relatively small. Therefore we tested the associations by Fisher's exact test. In this phase of analysis, we found 27 associations (from 1437 tests) with a significance level of less than 0.05. However, correction for multiple testing considering 1437 variants would result in a level of 0.000035 which was reached by none of the variations.

The most significant associations found in this phase were between 3 SNPs in EGFR gene and existence of skin rash. These 3 SNPs along with 2 other skin rash associated EGFR SNPs were

linked. All of them were positioned in intron 1 of EGFR gene on chromosome 7 between positions 55094331 and 55099394. Intron 1 of EGFR gene is believed to be an important regulatory region for EGFR transcription (Gebhardt et al. 1999). EGFR transcription is tightly regulated and its overexpression is reported in several types of cancer and is associated with a poor prognosis. The promoter region of EGFR does not contain a TATA box. Upstream of the ATG start codon, there are four transcription factor SP1 binding sites positioned between -500 to -100 bases. There are two enhancer regions which are found upstream of the promoter region. In addition to these two upstream enhancer regions, two other enhancer regions are found in intron 1. One of them is located at +1788- +2318 position (Maekawa et al. 1989). The other enhancer region is localized 1.1 kb downstream of the former mentioned enhancer region (McInerney et al. 2001) (Figure 4.1.1). This finding highlights the role of intron 1 in transcription regulation.

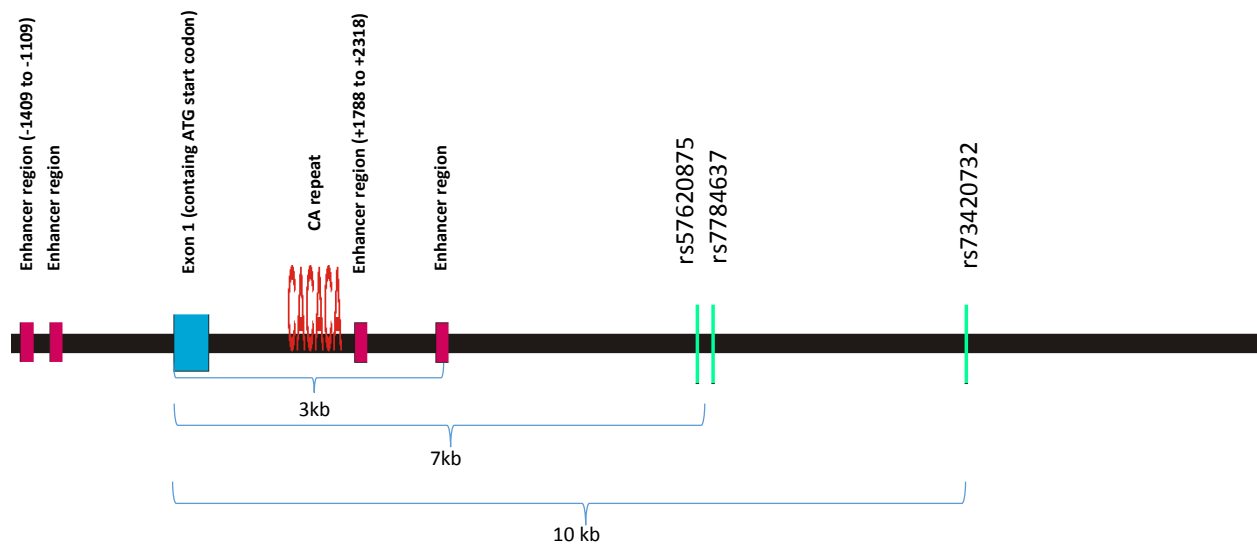


Figure 4.1.1 Structure of 5' sequence of EGFR gene. Four enhancer regions are shown as red boxes (Brandt et al. 2006). The CA repeat region which is believed to have a regulatory role in EGFR transcription is localized between exon 1 and first downstream enhancer region (Gebhardt et al. 1999). First 3 EGFR SNPs with the lowest P values are located 7kb (two SNPs and 10 kb downstream of transcription start site. The exact location of the enhancer regions without noted position are not determined in the literature.

Another feature in intron 1 is a CA sequence repeat which is also reported to have a regulatory role in transcription of *EGFR* (Gebhardt et al. 1999). The shorter repeat length was associated with higher transcription of *EGFR* gene. Up to 5 times lower levels of *EGFR* pre-RNA have been shown in vitro in samples with higher number of CA repeats (Mohamed et al. 2005a). Concerning the association with skin rash, unfortunately, the target capture method used by myself could not reliably capture repeat sequences and thus the analysis of the CA repeat length was not achievable. Therefore, in our study this region was even excluded from our target region. But shorter repeat length was reported in a few studies to be associated with occurrence of skin rash and was nominated as a biomarker for prediction of the skin rash and the outcome (Huang et al. 2009). However, further studies including the first report of our study (Parmar et al. 2013) and a prospective study (Loupakis et al. 2014) could not reproduce these results.

In addition, the structure of chromatin in intron 1 has convinced the scientists to analyze the role of intron 1 in *EGFR* gene transcription (Chrysogelos 1993). Intron 1 has DNase I hypersensitive sites, which are believed to be associated with important regulatory elements. These DNA sequences are exposed and accessible in chromatin structure, which makes them possible regulatory elements. Many transcription factors are found to interact with these regions (Chrysogelos 1993).

Furthermore, *EGFR* SNP rs73420732 along with rs28557040 (another highly skin rash associated SNP) occurs at the region predicted to be important for transcription initiation and elongation. These predictions are based on the histone modifications found in chromatin structure of normal human keratinocytes. A study published in 2011 analyzed the chromatin structure of 9 human cell lines and consequently predicted their probable functions (Ernst et al. 2011). Among these cell lines there was a normal human keratinocyte line which we assume to be the primary suspect in formation of *EGFR* inhibition mediated skin rash. Histone modifications are changes in histone proteins which are specific to cell types and are believed to determine the function of different segments of genome. These modifications include: methylation, acetylation, and phosphorylation (Bannister and Kouzarides 2011). Methylation is a modification on histone proteins which tends to have regulatory properties. In the methylation process, methyl-groups are added to lysine or arginine residues of these proteins. Methylation of histone proteins makes

them targets for proteins with specific domains for identification of these sites. As an example trimethylation of lysine 4, which is a character of promoter regions which are found by transcription factors (Roidl and Hacker 2014). The region in which rs73420732 and rs28557040 exist, is predicted to be the beginning of *EGFR* transcription elongation (transcription transition site) (Figure 4.1.2). Second and third most significant SNPs in this phase, rs57620875 and rs7784637, were predicted to be located in a region with strong enhancer activity. Although the data is newly introduced and needs to be more extensively surveyed, the regulatory role of these regions and the impact of genetic variations on this role could not be ruled out. As these variations occurred mostly in the no skin rash group, we hypothesized that changes in these regulatory regions result in the impairment of EGFR expression and therefore lower sensitivity of these cells to the EGFR inhibition.

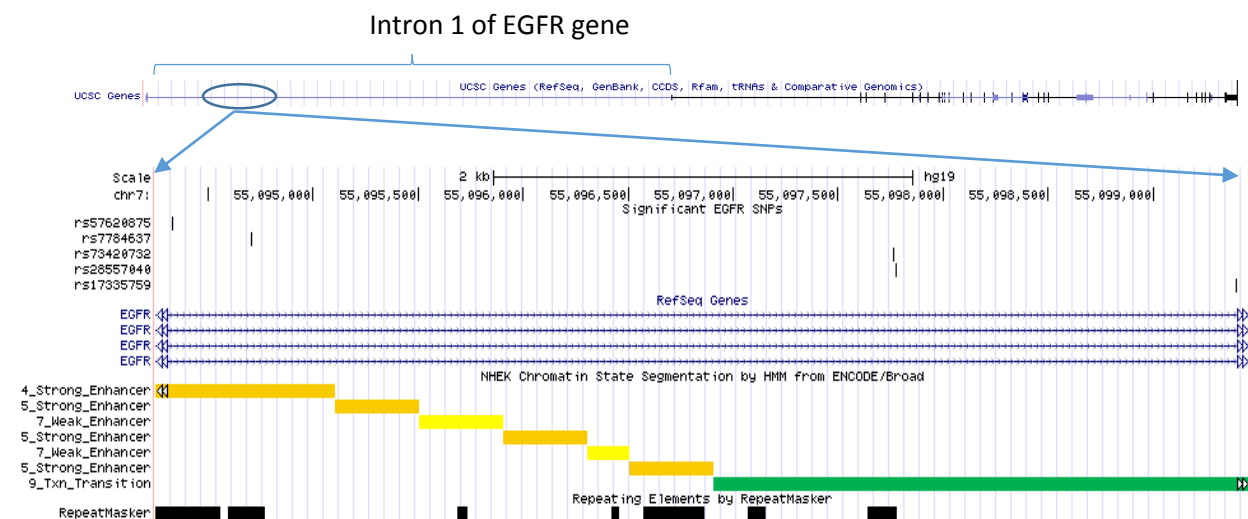


Figure 4.1.2. Predicted function of regions of EGFR SNPs with lowest *P* value in normal human keratinocytes. Black lines present the position of EGFR SNPs which were found to have significant association with presence of skin rash; orange, yellow and green bars show strong enhancer regions, weak enhancer regions and transcription transition regions, respectively (Ernst et al. 2011). These predictions are based on histone modifications found in these regions' histones in normal human keratinocytes. As an example 4\_strong\_enhancer refers to sites with histone 3 lysine 4 mono- and di-methylation along with histone 3 lysine 27 acetylation. Transcription transitional region is the position from where attachment of polymerase complex starts the elongation process. These functions have been suggested by considering chromatin modifications found at these positions in human keratinocyte cell. For further description in this regard please refer to (Ernst et al. 2011).

Two common variations upstream of transcription start site were suspected to be predictors of skin rash after therapy. These two are *EGFR* -191 C/A and *EGFR* -216 G/T polymorphisms. However the promoter region was included in our target region, we had no coverage at this region for unknown reasons (Figure 4.1.3).

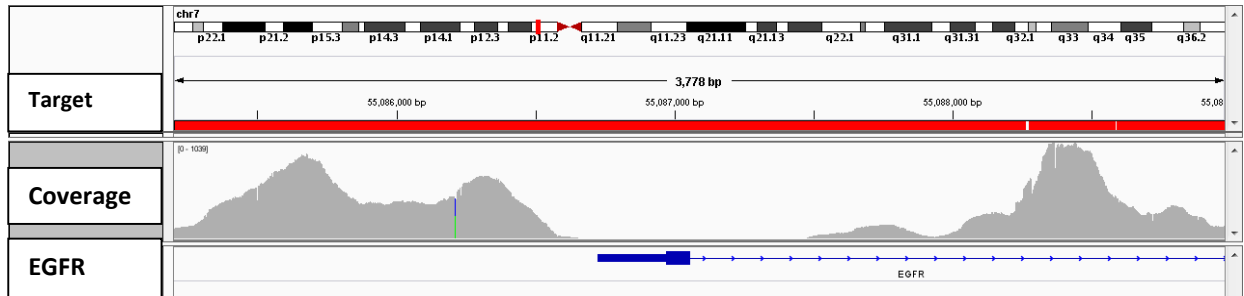


Figure 4.1.3 Lack of coverage of *EGFR* promoter region. Although the promoter region was targeted, no reads were found. Presented sample has one of the highest number of reads

Among SNPs in genes other than *EGFR* which were highly associated with skin rash, there is *PIK3R1* rs3730089 which caused a M326I amino acid substitution in regulatory subunit 1 of phosphatidylinositol 3-kinase (*PIK3R1*). Following *EGFR* activation, *PIK3* is activated via an adaptor protein (*IRS-1*). *PIK3* regulatory subunits bind to this protein which results in activation of the downstream signaling. M326I amino acid substitution was more prevalent among the group with skin rash (26% vs 10%) in this study. Functional studies found decreased expression of *PIK3R1* and yet on the other hand increased attachment to *IRS-1* which activates the *PIK3* downstream signaling following this variation (Almind et al. 2002, Luo and Cantley 2005). M326I was also found to be associated with colon cancer (Li et al. 2008); however studies performed afterwards could not find a significant association of this SNP with prostate cancer (Paradis et al. 2003) or endometrial cancer (Wang et al. 2012). The mechanism in which this variation may cause higher sensitivity to *EGFR* therapy, are remained to be understood. Higher activity of *PIK3* downstream signaling which possibly occurs in the skin cells with this variation, may make them more prone to *EGFR* inhibition.

## 4.2 Second analysis phase: Genomic variants predisposing to high-grade skin rash

In the second phase of analysis, patients were grouped into high grade skin rash (grades 2-3) and low grade rash (grades 0-1) to survey the association of variations with severity of rash. In this phase, 34 SNPs have reached significance level of 0.05. The most significant SNP was *MAPK1* (*ERK2*) rs13515. This SNP which was not in LD with other SNPs which were found in this gene. Other four *MAPK1* SNPs were associated with severe skin rash and were linked considering  $r^2 > 0.4$ .

Rs13515 is the substitution from C to T in 3' UTR of *ERK2* gene. No report has been found about the function of this variation. In our study 54% of patients with no or low grade skin rash were carriers of this variation and on the other hand 21% of patients with high level skin rash were carriers of this variation. The 3' UTR region is characterized with its role in gene expression by altering the mRNA's processing, stability, and translation efficiency (Zhao et al. 2011). It contains several sequences that are involved in gene expression, including microRNA response elements which are responsible for binding of miRNAs and affect the post transcriptional gene expression. The 3' UTR region of *MAPK1* or *ERK2* has been found to be targeted by a number of miRNAs. The MEK-ERK pathway has been shown to be regulated by miRNAs (Perry et al. 2009, Zheng et al. 2012). In addition miRNA miR-199a has been shown to target *MAPK1* mRNA and lower the expression of ERK protein (Kim, S et al. 2008). However no miRNA has been reported to target the position of rs13515 in *MAPK1* RNA transcript (3975<sup>th</sup> position in the 5594 base transcript) to the best of our knowledge. The mechanism in which a possible binding miRNA could lower the sensitivity of skin cells to EGFR inhibition is yet to be understood.

## 4.3 Common SNPs in both phases of analysis

Eight SNPs: rs6970262, rs1404908, rs4947965, rs33933140, rs7796872, rs1137188, rs13515, and rs6593202 were associated with skin rash in both analysis phases. All of these SNPs were relatively frequent (MAF > 20%). Among these SNPs, *AKT2* rs33933140 (the only functional variation) which causes I136T amino acid substitution was more common among the patients with skin rash. As mentioned earlier, Tan et al. have suggested that baseline level of pAkt is probably a predictor of appearance of rash in patients receiving EGFR inhibition. They compared

the baseline pAkt level in 18 breast cancer patients prior to receiving erlotinib. Baseline pAkt was significantly higher in patients with no rash compared to patients with grade 1-2 rash (Tan et al. 2008). However regarding rs33933140, a search in the literature did not result in a study which has investigated the effect of this variation. Though this SNP is localized in the linker region between pleckstrin homology (PH) domain and the kinase domain of Akt2 protein, it has been predicted to be damaging for protein function by the prediction tools. Higher skin rash prevalence among the patients with this variation is probably the result of decreased Akt phosphorylation (lower pAkt) subsequent to this amino acid substitution. It is also important to mention that this variation caused amino acid exchange in the protein coded by one of the 42 *AKT2* transcripts.

#### 4.4 Previously reported skin rash biomarker rs2227983

An amino acid substitution variation rs2227983 in EGFR was previously reported to be associated with skin rash (Parmar et al. 2013). This variation substitutes arginine to lysine at 521th amino acid of 1210 amino acid EGFR protein. In this study this variation was not found to be associated with skin rash. This result is compatible with results from several studies. McKibbin et al. surveyed this polymorphism in pooled data from five pediatric clinical trials of gefitinib. They dichotomized their patients similar to the phase two of our analysis (McKibbin et al. 2010). They could not find an association between this SNP and skin rash. As well, a phase 3 trial of erlotinib in pancreatic cancer patients could not find an association between this SNP and skin toxicity (Boeck et al. 2013). They had a dichotomization similar to our first phase of study. Rudin et al (Rudin et al. 2008) and Huang et al. (Huang et al. 2009) have reached similar results in NSCLC patients .

#### 4.5 Novel functionally damaging variations

We found a previously not reported variation in *EGFR* (as our major cancer predisposing gene) which substitutes proline to threonine at amino acid 560. The codon change is CAC [C/A]CA GAG at position 55231472 in chromosome 7. This substitution, which resides in a furin-like cysteine rich domain, is involved in the process of signal transduction by receptor tyrosine kinases. This region is believed to play a role in EGFR dimerization which happens immediately after EGFR



activation (Raz et al. 1991). We suggest that this variation can be a causal rare mutation, which can eventually contribute to cancer formation. Other novel variations which cause amino acid substitution and are predicted to be damaging, are MTOR R2152H and MAPK1 D336S. These variations are located in protein kinase domains of the corresponding proteins and result in loss of function. It is also important to state that the predictions are not 100% true and many found to be damaging variations are fully functional and vice versa.

#### 4.6 Runs of homozygosity

We observed a patient who had no heterozygous genotypes in *EGFR* region. Subsequently, we surveyed the depth of coverage at this position which might have been a reason for not calling heterozygote genotypes. However, the coverage at this subject's resequencing was not optimal (20x coverage for 52% of target region); it was sufficient for calling 721 genotypes at this position (including 82 homozygote variant genotypes).

Regions of runs of homozygosity are possibly caused by autozygosity or copy number variation. Autozygosity is defined as inheriting two chromosomal segments that are identical (one from each parent) from a common ancestor. On the other hand, small chromosomal deletions may also cause a similar genotype. Chromosomal deletions can result in regions with loss of heterozygosity due to losing one copy of a gene locus. When performing SNP arrays with microarray, copy number variation can be simply distinguished from autozygosity. When doing deep sequencing, however – if the application is not designed for identification of copy number variation – distinguishing between these two processes is not always possible. In the present study, gene copy number variation taqman assays can be used to distinguish between the stated processes.

In both cases of autozygosity or copy number variation this patient is susceptible for cancer formation. We surveyed the available phenotypes of this patient to identify any extreme phenotypes. This patient had a low survival after treatment start date (111 days) compared to mean survival rate. However, regarding skin rash (grade 1 occurred in this patient) or diarrhea

(grade 0) no remarkable phenotype was observed. No variations occurred in exon regions of this patient within the germline genome.

#### 4.7 Statistical issues regarding analysis of association of EGFR inhibition related skin toxicity and genetic variations

Study design and statistical analysis to find the genetic basis of skin rash variability seems to be challenging as the rash is developed in about 60-80% of patients receiving EGFRIs. In this situation an association study has to be performed in a sample population with three to four times case size compared to control size. In the present study, 19.8% of patients were affected with rash. A significant association would only be achievable in case of a strong underlying association of a common variation and not multiple rare variations. In addition, correction for multiple testing causes a further power loss in analysis. Another measure which has been taken for overcoming this problem is dichotomizing patients in two groups of rash grade  $\leq 1$  and  $>1$ , which will result in a relatively similar case-control sample size. However, this dichotomization may or may not reflect the reality considering the causality of variations in formation of rash; as the underlying mechanisms are not well-known.

#### 4.8 Gene-based and pathway-based analysis

An additional analysis which best suits the study of rare variations is gene-based analysis. Rare variations have a small but effective impact on a disease or phenotype. Each of these variations may not reach a significant level when studying their association with a phenotype, specially when the sample size is small. The common disease/rare variations hypothesis considers the additive effect of these variations in contributing to a disease or phenotype. To perform this analysis only predicted to be damaging rare SNPs should be enrolled in the analysis (Do et al. 2012). After Bonferroni's correction for multiple testing considering the 30 genes, P value of 0.0017 was found as the level of significance. Kirino et al. also used correction by the number of targeted genes in their study (targeted deep sequencing of loci found by GWAS to be associated with Behcet's disease) (Kirino et al. 2013). Performing this analysis for our study – considering damaging rare variants predicted by any of SIFT, Provean or PolyPhen2 – we could not find an

association between skin rash phenotype and genes with loss-of-function variants. In gene-based tests the I value is reported which represents the lowest P value in this analysis. However this value did not reach the corrected level of significance, which highlights the small sample size as a reason for the low power of finding statistical significance. To overcome this problem studies with higher sample size are recommended.

In pathway-based analysis an association was found between high grade skin rash, and PI3K/Akt pathway variations. This finding in addition to the other *PIK3R1* SNP rs3730089, which was found to be associated with presence of skin rash, introduces *PIK3R1* as an important skin rash determinant. Additionally, presence of skin rash in patients receiving PI3K inhibitors is another proof for involvement of PI3K in skin rash pathology (Mayer et al. 2014). In Mayer et al.'s study 40% of patients receiving the buparlasib (a PI3K inhibitor) developed skin rash. The fact that only a fraction of patients developed rash, highlights the possible role of PI3K genetic variations in these patients.

#### 4.9 The unsolved mystery of association between skin rash and survival

Survival of patients receiving EGFR inhibitors is associated with the skin rash. Figure 4.9.1 displays this association in the present study. The reasons behind this association are not well understood. Shorter intron 1 CA repeat polymorphism was reported to increase the expression of EGFR in vitro (Mohamed et al. 2005b). To explain the association of CA repeat with skin rash and also survival, scientists have hypothesized that the higher expression of EGFR causes the susceptibility of both keratinocytes and cancer cells to EGFR inhibition (Mohamed et al. 2005a). However several studies could not reproduce the association of CA repeat and skin rash or survival (Parmar et al. 2013). In addition – knowing that inflammatory response is an important part of skin rash pathology – a more extensive inflammatory response against tumor cells has also been suggested to be a reason for a better survival in patients with skin rash (Albanell et al. 2002).

The role of confounding factors in the correlation found between these agents' skin toxicity and survival cannot be overlooked. Firstly the rash is known to be a dose-dependent side effect. Difference in the bioavailability of the drugs in between the patients may cause this correlation.

To clarify, higher plasma levels of the drug in the patients – caused by individual difference in the drug metabolizing enzymes – may cause better efficacy and development of skin rash. There are recent studies which use a dose-to-rash approach, meaning that they escalate the dose of EGFRIs until the patients develop skin rash (Ho et al. 2011, Renouf et al. 2014). However the dose escalation was well tolerated; these studies did not find a benefit in efficacy for this approach. As another confounding, relatively low survival of some patients may not provide enough time for the development of rash. Therefore the association between no rash and low survival may be falsely identified. Nevertheless, the trend towards better survival with higher grades of skin rash cannot be attributed to the mentioned probable confounding.

As the majority of variations were rare in the present study, testing the association of these variations and survival was not possible. Though a descriptive analysis was performed on nonsynonymous variations (the functionally most important variations) regarding their effects on patients' survival, and average of patient's maximum skin rash and diarrhea in the observation period. These variations were mostly rare; however, there were a few common variations. Regarding the common variations, carriers of alternative allele of rs3730089 in *PIK3R1* gene (M326I) had a better survival (homozygous variant genotype 314 days, heterozygous genotype 260 days) compared to non-carriers (231 days). Interestingly, 24 out of 26 carriers developed skin rash after EGFR inhibitor administration. The function of this variation has been discussed in paragraph 4.1.

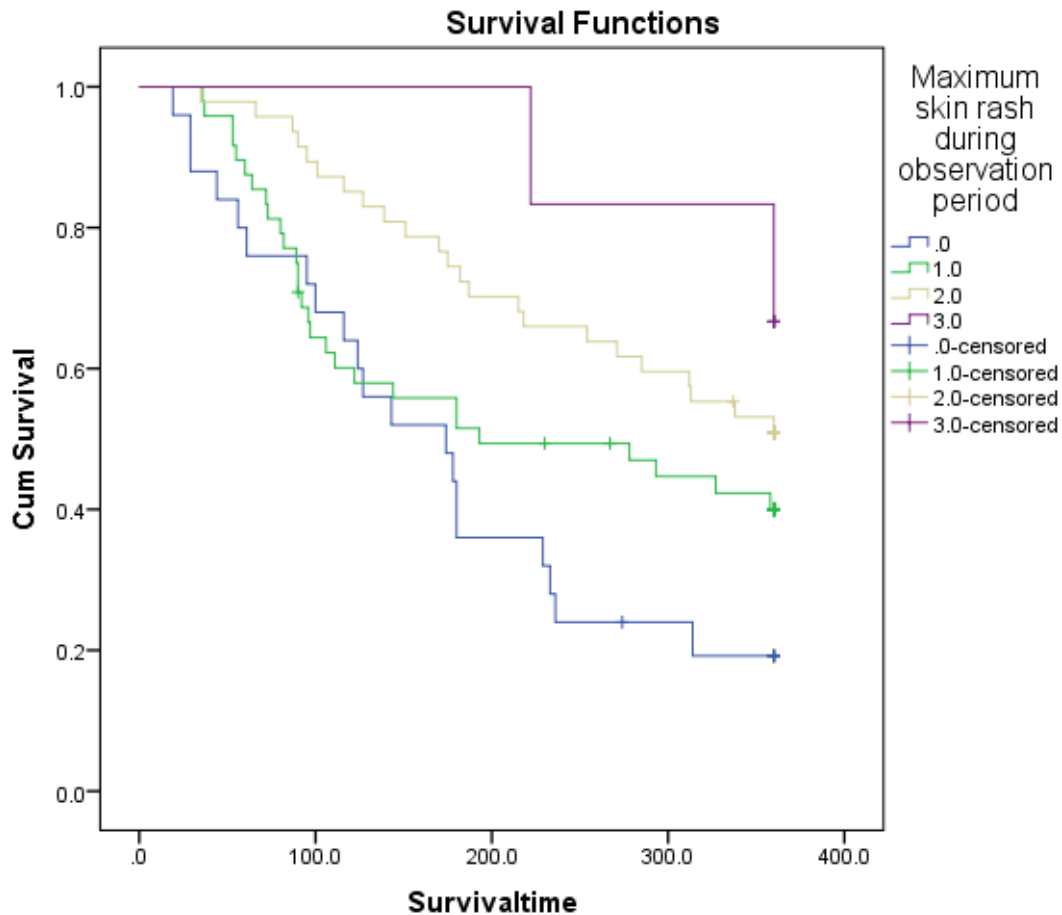


Figure 4.9.1 Survival plot of the patients in our study (in days) categorized the maximum skin rash grade occurred during the follow-up. Patients with higher rash grades had a better overall survival rate.

#### 4.10 Outlook

The strongest associations were found between a number of *EGFR* intron 1 SNPs and skin rash in this study. Several approaches can be used to survey the effects of these SNPs. To investigate whether these positions are transcription factor binding sites, electrophoretic mobility shift assay (EMSA) can be utilized. In this method, protein binding of the intronic DNA sequence can be tested for both reference and variant sequences. In addition, as intron 1 was found to have an important role in gene expression, effects of these intronic variants can be determined by reporter gene assays.

An Akt2 nonsynonymous SNP (rs33933140, I136T) was associated with both presence and high grade skin rash. This variation has been predicted to be damaging for protein function by the

prediction tools. Effects of this variation (such as chemokine expression or apoptosis) can be investigated using siRNAs for silencing Akt2 protein expression in normal human keratinocyte cell culture. Additionally, the kinase activity of wildtype and variant Akt2 can be investigated. In order to survey this activity, the wildtype and the variant protein can be expressed in a cell line which has low or no baseline expression of the selected kinase. Subsequently, using antibodies for phosphorylated Akt2 substrates this activity can be investigated. Human Akt2 knockout haploid cell line is also available for further investigation (Burckstummer et al. 2013).

To investigate the effects of variations in PIK3R1 (or more generally PI3K), the variant proteins can be expressed in yeast and then affinity of PIK3R1 to IRS-1 (adaptor protein between EGFR and PI3K) or to the kinase subunit PIK3CA or PIK3CB (the catalytic subunits of PI3K) can be analyzed. Moreover, to analyze the effects of these variations on downstream signaling, the pAkt level and also kinase activity of Akt protein can be examined (Almind et al. 2002).

To the best of our knowledge, a model for skin rash has not been produced. Potentially immune cell infiltration via chemoattraction can be investigated using keratinocyte cell cultures and chemotaxis assays. The most well-known chemotaxis assay is the two-chamber assay. In this assay keratinocytes and immune cells are placed in two different chambers for culture, but they share the culture medium (Boyden 1962). Between two chambers, there is a filter only allowing a certain size of motile cells to migrate to the keratinocyte chamber. Along with chemokine expression after EGFR inhibition, chemoattraction can be investigated. Using this model a wide screening for relevant chemokines can be performed.

#### 4.11 Study limitations

A major pitfall while performing Ion torrent sequencing is the presence of homopolymer issue which also existed in pyrosequencing (Michils et al. 2012). Using Ion torrent sequencing nucleotides are added one after another and the alteration of PH is recorded. PH alterations are indicators of a reaction in which the added nucleotide firstly has matched the template and secondly is added to the under-construction strand. In case of a nucleotide repeat, multiple reactions happen after addition of one nucleotide since the following template nucleotide in the

template is also complimentary to the added nucleotide. This phenomenon results in higher PH change compared to the PH change when only one reaction has occurred (in case of no-repeat). Distinction of number of nucleotides is not confidently made in a repeat sequence using this method. Therefore, the presence of homopolymer regions produce artifacts which resemble insertions and/or deletions (Tarabeux et al. 2014). To resolve this problem, studies have taken several measures such as accepting higher sequencing qualities or filtering them out from the study (Abou Tayoun et al. 2013, Ross et al. 2013, Tarabeux et al. 2014). Accordingly, we filtered out the homopolymer regions and all of the insertion/deletions from the study which results in lower false positive results but less coverage of target region.

#### 4.12 Conclusion

In conclusion, we have found 1437 single nucleotide variations in 380kb targeted region of 126 patients who underwent EGFR inhibitor therapy. Three SNPs in a regulatory region of intron 1 of *EGFR* were found to be strongly associated with the presence of rash. Furthermore, an *ERK2* (*MAPK1*) 3' UTR variation was found to be associated with high levels of skin rash. Pathway-based analysis of rare variations suggested the PI3K pathway to be involved in the rash etiology. In addition to this finding, rs3730089 in *PIK3R1* was found to have an association with skin rash and also a trend towards a better survival was found in patients having this variation.

A number of deleterious variations in EGFR pathway were announced for the first time including EGFR P560T, MTOR R2152H, and MAPK1 D336S which can be causal rare variations for cancer formation. Previously reported candidate biomarker variations for EGFR inhibition related skin toxicity, rs2227983 which causes a single amino acid substitution was not significantly associated with skin toxicity and other biomarkers such as intron 1 CA repeat, -216G/T, and -191C/A could not be investigated.

## 5 Summary

The development of targeted therapy has promised a new era in cancer therapy by improving response rates and fewer side effects of the therapeutics on normal cells. As the involvement of EGFR was noticed in several epithelial tumors, this receptor was identified as a target for cancer therapy already in 1983. However, EGFR inhibition seems to have its own class of side effects such as skin toxicity and diarrhea. These side effects remarkably occur only in a subgroup of patients with remarkable severity. Interestingly, the occurrence of skin toxicity was in almost all studies correlated with drug response and survival rate in treated patients. In this study we aimed to analyze the genetic variations in EGFR signaling pathway in patients receiving EGFR inhibition therapy using a deep sequencing method. If variants could be identified which significantly predicted skin toxicity, analysis for such variants would be used as biomarkers for individual drug dosing. In addition, the relationship between variants and phenotypes might contribute to our understanding of the skin rash with EGFR inhibition therapy.

Thirty genes reflecting the EGFR downstream pathway (PI3K/Akt and Ras/MAPK) and other genes involved such as IL8 and its receptors were targeted for resequencing. The whole *EGFR* gene along with 2000 bp of its promoter region was one of the targets. The total size of the resequenced *EGFR* gene was 196300 base pairs. For the remaining 29 other genes involved or presumed to be involved in rash physiopathology, the exons along with about 500 bp of the respective promoter regions were selected. We studied DNA samples from 126 patients who had been treated with EGFR inhibitors and in whom the adverse effects were characterized in detail. Before deep sequencing analysis was performed, three major steps: library preparation, target enrichment and template preparation were performed. Deep sequencing was executed using Ion Torrent Personal Genomics Machine™ (PGM). Variant calling was done by Torrent Variant Caller software. Statistical analysis was performed to survey the association of the identified variants with skin toxicity, diarrhea and overall survival.

Total 1437 single nucleotide variations were identified. These variations were present in more than 50% of subjects. From all variants found, 45% had less than 1% minor allele frequency and 65% of the variations had less than 5% minor allele frequency. Average minor allele frequency



was 8.79% in this study. Three hundred and seventeen SNPs (22%) were not listed in the dbSNP 138. Three SNPs in intron 1 of EGFR were found to be strongly associated with the presence of rash. Intron 1 of *EGFR* was already shown earlier to be an important regulatory region for EGFR transcription. Furthermore, an *ERK2* variation was found to be associated with high levels of skin rash. Pathway-based analysis of rare variations suggested the PI3K pathway to be involved in rash etiology. In addition to this finding, rs3730089 in *PIK3R1* was found to have an association with skin rash and also a trend towards the better survival was found in patients having this variation. These two findings suggest the involvement of PIK3R1 in EGFR inhibitor induced skin toxicity.

A number of deleterious variations in EGFR pathway were announced for the first time including EGFR P560T, MTOR R2152H, and MAPK1 D336S which can be causal rare variations in cancer pathology. Previously reported biomarker candidate variations for EGFR inhibition related skin toxicity such as intron 1 CA repeat, -216G/T, and -190C/A were not studied and the rs2227983 single amino acid substitution was not significantly associated with skin toxicity.

## 6 References

- Abou Tayoun A, Tunkey C, Pugh T, Ross T, Shah M, Lee C, Harkins T, Wells W, Tafe L, Amos C et al. (2013): A comprehensive assay for CFTR mutational analysis using next-generation sequencing. *Clin. Chem.* 59(10), 1481-1488
- Adzhubei I, Schmidt S, Peshkin L, Ramensky V, Gerasimova A, Bork P, Kondrashov A, Sunyaev S (2010): A method and server for predicting damaging missense mutations. *Nat. Met.* 7(4), 248-249
- Agero A, Dusza S, Benvenuto-Andrade C, Busam K, Myskowski P, Halpern A (2006): Dermatologic side effects associated with the epidermal growth factor receptor inhibitors. *J. Am. Acad. Dermatol.* 55(4), 657-670
- Agrawal N, Frederick M, Pickering C, Bettegowda C, Chang K, Li R, Fakhry C, Xie T-X, Zhang J, Wang J et al. (2011): Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* 333(6046), 1154-1157
- Albanell J, Rojo F, Averbuch S, Feyereislova A, Mascaro J, Herbst R, LoRusso P, Rischin D, Sauleda S, Gee J et al. (2002): Pharmacodynamic studies of the epidermal growth factor receptor inhibitor ZD1839 in skin from cancer patients: histopathologic and molecular consequences of receptor inhibition. *J. Clin. Oncol.* 20(1), 110-124
- Almind K, Delahaye L, Hansen T, Van Obberghen E, Pedersen O, Kahn C (2002): Characterization of the Met326Ile variant of phosphatidylinositol 3-kinase p85alpha. *Proc. Natl. Acad. Sci. U. S. A.* 99(4), 2124-2128
- Anjum R, Blenis J (2008): The RSK family of kinases: emerging roles in cellular signalling. *Nat. Rev. Mol. Cell Biol.* 9(10), 747-758
- Ashtiani Z, Hasheminasab S-M, Ayati M, Goulian B, Modarressi M (2011): Are GSTM1, GSTT1 and CAG repeat length of androgen receptor gene polymorphisms associated with risk of prostate cancer in Iranian patients? *Pathol. Oncol. Res.* 17(2), 269-275
- Bangsgaard N, Houtkamp M, Schuurhuis D, Parren P, Baadsgaard O, Niessen H, Skov L (2012): Neutralization of IL-8 prevents the induction of dermatologic adverse events associated with the inhibition of epidermal growth factor receptor. *PLoS One* 7(6),
- Bannister A, Kouzarides T (2011): Regulation of chromatin by histone modifications. *Cell Res.* 21(3), 381-395
- Barbacid M (1987): ras genes. *Annu. Rev. Biochem.* 56 779-827

Barrett J, Fry B, Maller J, Daly M (2005): Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2), 263-265

Baselga J, Rischin D, Ranson M, Calvert H, Raymond E, Kieback D, Kaye S, Gianni L, Harris A, Bjork T et al. (2002): Phase I safety, pharmacokinetic, and pharmacodynamic trial of ZD1839, a selective oral epidermal growth factor receptor tyrosine kinase inhibitor, in patients with five selected solid tumor types. *J. Clin. Oncol.* 20(21), 4292-4302

Bellacosa A, Kumar CC, Cristofano AD, Testa JR (2005): Activation of AKT kinases in cancer: implications for therapeutic targeting. *Adv. Cancer Res.* 94 29-86

Bentley D, Balasubramanian S, Swerdlow H, Smith G, Milton J, Brown C, Hall K, Evers D, Barnes C, Bignell H et al. (2008): Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456 (7218), 53-59

Betson M, Lozano E, Zhang J, Braga VM (2002): Rac activation upon cell-cell contact formation is dependent on signaling from the epidermal growth factor receptor. *J. Biol. Chem.* 277 (40), 36962-36969

Blick SKA, Scott LJ (2007): Cetuximab: a review of its use in squamous cell carcinoma of the head and neck and metastatic colorectal cancer. *Drugs* 67 (17), 2585-2607

Bobadilla J, Macek M, Fine J, Farrell P (2002): Cystic fibrosis: a worldwide analysis of CFTR mutations—correlation with incidence data and application to screening. *Hum. Mutat.* 19(6) 575-606

Boeck S, Hausmann A, Reibke R, Schulz C, Heinemann V (2007): Severe lung and skin toxicity during treatment with gemcitabine and erlotinib for metastatic pancreatic cancer. *Anticancer Drugs* 18 (9), 1109-1111

Boeck S, Jung A, Laubender R, Neumann J, Egg R, Goritschan C, Vehling-Kaiser U, Winkelmann C, Fischer von Weikersthal L, Clemens M et al. (2013): EGFR pathway biomarkers in erlotinib-treated patients with advanced pancreatic cancer: translational results from the randomised, crossover phase 3 trial AIO-PK0104. *Br. J. Cancer* 108 (2), 469-476

Boku N, Sugihara K, Kitagawa Y, Hatake K, Gemma A, Yamazaki N, Muro K, Hamaguchi T, Yoshino T, Yana I et al. (2014): Panitumumab in Japanese patients with unresectable colorectal cancer: a post-marketing surveillance study of 3085 patients. *Jpn. J. Clin. Oncol.* 44 (3), 214-223

Boone S, Rademaker A, Liu D, Pfeiffer C, Mauro D, Lacouture M (2007): Impact and management of skin toxicity associated with anti-epidermal growth factor receptor therapy: survey results. *Oncology* 72 (3-4), 152-159

Boyden S (1962): The chemotactic effect of mixtures of antibody and antigen on polymorphonuclear leucocytes. *J. Exp. Med.* 115 453-466

Brandt B, Meyer-Staeckling S, Schmidt H, Agelopoulos K, Buerger H (2006): Mechanisms of egfr gene transcription modulation: relationship to cancer risk and therapy response. *Clin. Cancer Res.* 12(24), 7252-7260

Brockmüller J, Tzvetkov M (2008): Pharmacogenetics: data, concepts and tools to improve drug discovery and drug treatment. *Eur. J. Clin. Pharmacol.* 64(2), 133-157

Burckstummer T, Banning C, Hainzl P, Schobesberger R, Kerzendorfer C, Pauler FM, Chen D, Them N, Schischlik F, Rebsamen M et al. (2013): A reversible gene trap collection empowers haploid genetics in human cells. *Nat. Met.* 10(10), 965-971

Burtness B, Anadkat M, Basti S, Hughes M, Lacouture M, McClure J, Myskowski P, Paul J, Perlis C, Saltz L et al. (2009): NCCN Task Force Report: Management of dermatologic and other toxicities associated with EGFR inhibition in patients with cancer. *J. Natl. Compr. Canc. Netw.* 7 Suppl 1 5-21

Candi E, Schmidt R, Melino G (2005): The cornified envelope: a model of cell death in the skin. *Nature reviews. Molecular cell biology* 6(4), 328-340

Carey M, Smale S (2007): Methylation interference assay. *CSH Protoc* 2007

Carpenter G, Lembach KJ, Morrison MM, Cohen S (1975): Characterization of the binding of 125-I-labeled epidermal growth factor to human fibroblasts. *J. Biol. Chem.* 250(11), 4297-4304

Carriere A, Ray H, Blenis J, Roux P (2008): The RSK factors of activating the Ras/MAPK signaling cascade. *Front. Biosci.* 13 4258-4275

Cataldo V, Gibbons D, Pérez-Soler R, Quintás-Cardama A (2011): Treatment of non-small-cell lung cancer with erlotinib or gefitinib. *N. Engl. J. Med.* 364(10), 947-955

Choi Y, Sims G, Murphy S, Miller J, Chan A (2012): Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 7(10), e46688

Chrysogelos S (1993): Chromatin structure of the EGFR gene suggests a role for intron 1 sequences in its regulation in breast cancer cells. *Nucleic Acids Res.* 5736-5741

Ciardiello F, Bianco R, Damiano V, De Lorenzo S, Pepe S, De Placido S, Fan Z, Mendelsohn J, Bianco A, Tortora G (1999): Antitumor activity of sequential treatment with topotecan and anti-epidermal growth factor receptor monoclonal antibody C225. *Clin. Cancer Res.* 5(4), 909-916

Ciardiello F, Tortora G (2008): EGFR antagonists in cancer treatment. *N. Engl. J. Med.* 358 (11), 1160-1174

Ciuleanu T, Stelmakh L, Cicenias S, Miliauskas S, Grigorescu A, Hillenbach C, Johannsdottir H, Klughammer B, Gonzalez E (2012): Efficacy and safety of erlotinib versus chemotherapy in second-line treatment of patients with advanced, non-small-cell lung cancer with poor prognosis (TITAN): a randomised multicentre, open-label, phase 3 study. *Lancet Oncol.* 13 (3), 300-308

Costa A, Tejpar S, Prenen H, Van Cutsem E (2011): Hypomagnesaemia and targeted anti-epidermal growth factor receptor (EGFR) agents. *Target. Oncol.* 6 (4), 227-233

Cunningham D, Humblet Y, Siena S, Khayat D, Bleiberg H, Santoro A, Bets D, Mueser M, Harstrick A, Verslype C et al. (2004): Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N. Engl. J. Med.* 351 (4), 337-345

Do R, Kathiresan S, Abecasis G (2012): Exome sequencing and complex disease: practical aspects of rare variant association studies. *Hum. Mol. Genet.* 21 (R1), 9

Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B (2003): Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. U. S. A.* 100 (15), 8817-8822

Eferl R, Wagner E (2003): AP-1: a double-edged sword in tumorigenesis. *Nat. Rev. Cancer* 3 (11), 859-868

Ernst J, Kheradpour P, Mikkelsen T, Shores N, Ward L, Epstein C, Zhang X, Wang L, Issner R, Coyne M et al. (2011): Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473 (7345), 43-49

Fakih M, Wilding G, Lombardo J (2006): Cetuximab-induced hypomagnesemia in patients with colorectal cancer. *Clin. Colorectal Cancer* 6 (2), 152-156

Feero W, Guttmacher A, Collins F (2010): Genomic medicine--an updated primer. *N. Engl. J. Med.* 362 (21), 2001-2011

Franke T (2008): PI3K/Akt: getting it right matters. *Oncogene* 27 (50), 6473-6488

Fresno Vara J, Casado E, de Castro J, Cejas P, Belda-Iniesta C, González-Barón M (2004): PI3K/Akt signalling pathway and cancer. *Cancer Treat. Rev.* 30 (2), 193-204

Galimont-Collen A, Vos L, Lavrijssen A, Ouwerkerk J, Gelderblom H (2007): Classification and management of skin, hair, nail and mucosal side-effects of epidermal growth factor receptor (EGFR) inhibitors. *Eur. J. Cancer* 43 (5), 845-851

Gebhardt F, Zanker KS, Brandt B (1999): Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. *J. Biol. Chem.* 274 (19), 13176-13180

Genomes Project C, Abecasis G, Altshuler D, Auton A, Brooks L, Durbin R, Gibbs R, Hurles M, McVean G (2010): A map of human genome variation from population-scale sequencing. *Nature* 467 (7319), 1061-1073

Gill GN, Kawamoto T, Cochet C, Le A, Sato JD, Masui H, McLeod C, Mendelsohn J (1984): Monoclonal anti-epidermal growth factor receptor antibodies which are inhibitors of epidermal growth factor binding and antagonists of epidermal growth factor binding and antagonists of epidermal growth factor-stimulated tyrosine protein kinase activity. *J. Biol. Chem.* 259 (12), 7755-7760

Gilmore T, Herscovitch M (2006): Inhibitors of NF-kappaB signaling: 785 and counting. *Oncogene* 25 (51), 6887-6899

Giovannetti E, Zucali P, Peters G, Cortesi F, D'Incecco A, Smit E, Falcone A, Burgers J, Santoro A, Danesi R et al. (2010): Association of polymorphisms in AKT1 and EGFR with clinical outcome and toxicity in non-small cell lung cancer patients treated with gefitinib. *Mol. Cancer Ther.* 9 (3), 581-593

Goldberg R, Sargent D, Morton R, Fuchs C, Ramanathan R, Williamson S, Findlay B, Pitot H, Alberts S (2004): A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *J. Clin. Oncol.* 22 (1), 23-30

Govindan R, Behnken D, Read W, McLeod H (2003): Wound healing is not impaired by the epidermal growth factor receptor-tyrosine kinase inhibitor gefitinib. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 14 (8), 1330-1331

Gregorc V, Hidalgo M, Spreafico A, Cusatis G, Ludovini V, Ingersoll R, Marsh S, Steinberg S, Viganò M, Ghio D et al. (2008): Germline polymorphisms in EGFR and survival in patients with lung cancer receiving gefitinib. *Clin. Pharmacol. Ther.* 83 (3), 477-484

Herbst R, Langer C (2002): Epidermal growth factor receptors as a target for cancer treatment: the emerging role of IMC-C225 in the treatment of lung and head and neck cancers. *Semin. Oncol.* 29 (1 Suppl 4), 27-36

Hirschhorn J, Lohmueller K, Byrne E, Hirschhorn K (2002): A comprehensive review of genetic association studies. *Genet. Med.* 4 (2), 45-61

Ho C, Sangha R, Beckett L, Tanaka M, Lau D, Eisen D, Burich R, Luciw P, Khan I, Mack P et al. (2011): Escalating weekly doses of cetuximab and correlation with skin toxicity: a phase I study. *Invest. New Drugs* 29(4), 680-687

Hocking C, Price T (2014): Panitumumab in the management of patients with KRAS wild-type metastatic colorectal cancer. *Therap. Adv. Gastroenterol.* 7(1), 20-37

Hoffmann A, Natoli G, Ghosh G (2006): Transcriptional regulation via the NF-kappaB signaling module. *Oncogene* 25(51), 6706-6716

Hosgood H, Berndt S, Lan Q (2007): GST genotypes and lung cancer susceptibility in Asian populations with indoor air pollution exposures: a meta-analysis. *Mutat. Res.* 636(1-3), 134-143

Huang C-L, Yang C-H, Yeh K-H, Hu F-C, Chen K-Y, Shih J-Y, Lin Z-Z, Yu C-J, Cheng A-L, Yang P-C (2009): EGFR intron 1 dinucleotide repeat polymorphism is associated with the occurrence of skin rash with gefitinib treatment. *Lung Cancer* 64(3), 346-351

Humphrey PA, Wong AJ, Vogelstein B, Zalutsky MR, Fuller GN, Archer GE, Friedman HS, Kwatra MM, Bigner SH, Bigner DD (1990): Anti-synthetic peptide antibody reacting at the fusion junction of deletion-mutant epidermal growth factor receptors in human glioblastoma. *Proc. Natl. Acad. Sci. U. S. A.* 87(11), 4207-4211

Hunter S, Jones P, Mitchell A, Apweiler R, Attwood T, Bateman A, Bernard T, Binns D, Bork P, Burge S et al. (2012): InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* 40(Database issue), 12

Hynes NE, Lane HA (2005): ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat. Rev. Cancer* 5(5), 341-354

Ishiguro M, Watanabe T, Yamaguchi K, Satoh T, Ito H, Seriu T, Sakata Y, Sugihara K (2012): A Japanese post-marketing surveillance of cetuximab (Erbix<sup>®</sup>) in patients with metastatic colorectal cancer. *Jpn. J. Clin. Oncol.* 42(4), 287-294

Iyengar S, Elston R (2007): The genetic basis of complex traits: rare variants or "common gene, common disease"? *Methods Mol. Biol.* 376 71-84

Jancík S, Drábek J, Radzioch D, Hajdúch M (2010): Clinical relevance of KRAS in human cancers. *J. Biomed. Biotechnol.* 2010 150960

Jean G, Shah S (2008): Epidermal growth factor receptor monoclonal antibodies for the treatment of metastatic colorectal cancer. *Pharmacotherapy* 28 (6), 742-754

Johnson A, O'Donnell C (2009): An open access database of genome-wide association results. *BMC Med. Genet.* 10 6

Jost M, Kari C, Rodeck U (2000): The EGF receptor - an essential regulator of multiple epidermal functions. *Eur. J. Dermatol.* 10 (7), 505-510

Kaiser J (2008): DNA sequencing. A plan to capture human diversity in 1000 genomes. *Science* 319 (5862), 395

Kan Y, Dozy A (1978): Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proc. Natl. Acad. Sci. U. S. A.* 75 (11), 5631-5635

Kawamoto T, Sato J, Le A, Polikoff J, Sato G, Mendelsohn J (1983): Growth stimulation of A431 cells by epidermal growth factor: identification of high-affinity receptors for epidermal growth factor by an anti-receptor monoclonal antibody. *Proc. Natl. Acad. Sci. U. S. A.* 80 (5), 1337-1341

Kidd J, Cooper G, Donahue W, Hayden H, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F et al. (2008): Mapping and sequencing of structural variation from eight human genomes. *Nature* 453 (7191), 56-64

Kim E, Hirsh V, Mok T, Socinski M, Gervais R, Wu Y-L, Li L-Y, Watkins C, Sellers M, Lowe E et al. (2008): Gefitinib versus docetaxel in previously treated non-small-cell lung cancer (INTEREST): a randomised phase III trial. *Lancet* 372 (9652), 1809-1818

Kim E, Neubauer M, Cohn A, Schwartzberg L, Garbo L, Caton J, Robert F, Reynolds C, Katz T, Chittoor S et al. (2013): Docetaxel or pemetrexed with or without cetuximab in recurrent or progressive non-small-cell lung cancer after platinum-based therapy: a phase 3, open-label, randomised trial. *Lancet Oncol.* 14 (13), 1326-1336

Kim S, Lee U, Kim M, Lee E-J, Kim J, Lee M, Choung S, Kim Y, Choi Y-C (2008): MicroRNA miR-199a\* regulates the MET proto-oncogene and the downstream extracellular signal-regulated kinase 2 (ERK2). *J. Biol. Chem.* 283 (26), 18158-18166

Kirino Y, Bertsias G, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I, Seyahi E, Ozyazgan Y, Sacli F, Erer B, Inoko H et al. (2013): Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-B\*51 and ERAP1. *Nat. Genet.* 45 (2), 202-207



Kitsios G,Zintzaras E (2009): Genome-wide association studies: hypothesis-"free" or "engaged"? *Transl. Res.* 154 (4), 161-164

Klein R, Zeiss C, Chew E, Tsai J-Y, Sackler R, Haynes C, Henning A, SanGiovanni J, Mane S, Mayne S et al. (2005): Complement factor H polymorphism in age-related macular degeneration. *Science* 308 (5720), 385-389

Klinghammer K, Knödler M, Schmittel A, Budach V, Keilholz U,Tinhofer I (2010): Association of epidermal growth factor receptor polymorphism, skin toxicity, and outcome in patients with squamous cell carcinoma of the head and neck receiving cetuximab-docetaxel treatment. *Clin. Cancer Res.* 16 (1), 304-310

Kobayashi T, Hashimoto K, Okumura H, Asada H,Yoshikawa K (1998): Endogenous EGF-family growth factors are necessary for the progression from the G1 to S phase in human keratinocytes. *J. Invest. Dermatol.* 111 (4), 616-620

Kolch W (2000): Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. *The Biochemical journal* 351 Pt 2 289-305

Kumar P, Henikoff S,Ng P (2009): Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4 (7), 1073-1081

Lacouture M (2006): Mechanisms of cutaneous toxicities to EGFR inhibitors. *Nat. Rev. Cancer* 6 (10), 803-812

Lander E (1996): The new genomics: global views of biology. *Science* 274 (5287), 536-539

Laplante M,Sabatini D (2012): mTOR signaling in growth control and disease. *Cell* 149 (2), 274-293

Levy S, Sutton G, Ng P, Feuk L, Halpern A, Walenz B, Axelrod N, Huang J, Kirkness E, Denisov G et al. (2007): The diploid genome sequence of an individual human. *PLoS Biol.* 5 (10),

Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, Puc J, Miliarensis C, Rodgers L,McCombie R (1997): PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* 275 (5308), 1943-1947

Li L, Plummer S, Thompson C, Tucker T,Casey G (2008): Association between phosphatidylinositol 3-kinase regulatory subunit p85alpha Met326Ile genetic polymorphism and colon cancer risk. *Clin. Cancer Res.* 14 (3), 633-637

Li S, Schmitz KR, Jeffrey PD, Wiltzius JJW, Kussie P, Ferguson KM (2005): Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* 7(4), 301-311

Li T, Perez-Soler R (2009): Skin toxicities associated with epidermal growth factor receptor inhibitors. *Target. Oncol.* 4(2), 107-119

Liu G, Gurubhagavatula S, Zhou W, Wang Z, Yeap B, Asomaning K, Su L, Heist R, Lynch T, Christiani D (2008): Epidermal growth factor receptor polymorphisms and clinical outcomes in non-small-cell lung cancer patients treated with gefitinib. *Pharmacogenomics J* 8(2), 129-138

Loupakis F, Antoniotti C, Cremolini C, Zhang W, Yang D, Wakatsuki T, Bohanes P, Schirripa M, Salvatore L, Masi G et al. (2014): Prospective study of EGFR intron 1 (CA)<sub>n</sub> repeats variants as predictors of benefit from cetuximab and irinotecan in chemo-refractory metastatic colorectal cancer (mCRC) patients. *Pharmacogenomics J*

Luo J, Cantley L (2005): The negative regulation of phosphoinositide 3-kinase signaling by p85 and its implication in cancer. *Cell Cycle* 4(10), 1309-1312

Luria S, Human ML (1952): A nonhereditary, host-induced variation of bacterial viruses. *J. Bacteriol.* 64(4), 557

Maekawa T, Imamoto F, Merlino G, Pastan I, Ishii S (1989): Cooperative function of two separate enhancers of the human epidermal growth factor receptor proto-oncogene. *J. Biol. Chem.* 264(10), 5488-5494

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ (2010): Target-enrichment strategies for next-generation sequencing. *Nat. Met.* 7(2), 111-118

Manolio T, Brooks L, Collins F (2008): A HapMap harvest of insights into the genetics of common disease. *J. Clin. Invest.* 118(5), 1590-1605

Mardis E (2013): Next-generation sequencing platforms. *Annu. Rev. Anal. Chem. (Palo Alto Calif.)* 6 287-303

Mascia F, Mariani V, Girolomoni G, Pastore S (2003): Blockade of the EGF receptor induces a deranged chemokine expression in keratinocytes leading to enhanced skin inflammation. *The American journal of pathology* 163(1), 303-312

Mass R (2004): The HER receptor family: a rich target for therapeutic development. *Int. J. Radiat. Oncol. Biol. Phys.* 58(3), 932-940

Masui H, Kawamoto T, Sato J, Wolf B, Sato G, Mendelsohn J (1984a): Growth inhibition of human tumor cells in athymic mice by anti-epidermal growth factor receptor monoclonal antibodies. *Cancer Res.* 44 (3), 1002-1007

Masui H, Kawamoto T, Sato JD, Wolf B, Sato G, Mendelsohn J (1984b): Growth inhibition of human tumor cells in athymic mice by anti-epidermal growth factor receptor monoclonal antibodies. *Cancer Res.* 44 (3), 1002-1007

Maxam A, Gilbert W (1977): A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U. S. A.* 74 (2), 560-564

Mayer I, Abramson V, Isakoff S, Forero A, Balko J, Kuba M, Sanders M, Yap J, Van den Abbeele A, Li Y et al. (2014): Stand up to cancer phase Ib study of pan-phosphoinositide-3-kinase inhibitor buparlisib with letrozole in estrogen receptor-positive/human epidermal growth factor receptor 2-negative metastatic breast cancer. *J. Clin. Oncol.* 32 (12), 1202-1209

McInerney JM, Wilson MA, Strand KJ, Chrysogelos SA (2001): A strong intronic enhancer element of the EGFR gene is preferentially active in high EGFR expressing breast cancer cells. *J. Cell. Biochem.* 80 (4), 538-549

McKibbin T, Zhao W, Tagen M, Daw N, Furman W, McGregor L, Geyer J, Allen J, Stewart C (2010): Epidermal growth factor receptor polymorphisms and risk for toxicity in paediatric patients treated with gefitinib. *Eur. J. Cancer* 46 (11), 2045-2051

Michils G, Hollants S, Dehaspe L, Van Houdt J, Bidet Y, Uhrhammer N, Bignon Y-J, Vermeesch J, Cuppens H, Matthijs G (2012): Molecular analysis of the breast cancer genes BRCA1 and BRCA2 using amplicon-based massive parallel pyrosequencing. *J Mol Diagn* 14 (6), 623-630

Miettinen PJ, Berger JE, Meneses J, Phung Y, Pedersen RA, Werb Z, Derynck R (1995): Epithelial immaturity and multiorgan failure in mice lacking epidermal growth factor receptor. *Nature* 376 (6538), 337-341

Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK (2011): Mapping copy number variation by population-scale genome sequencing. *Nature* 470 (7332), 59-65

Mohamed M, Ramalingam S, Lin Y, Gooding W, Belani C (2005a): Skin rash and good performance status predict improved survival with gefitinib in patients with advanced non-small cell lung cancer. *Ann. Oncol.* 16 (5), 780-785

Mohamed M, Ramalingam S, Lin Y, Gooding W, Belani C (2005b): Skin rash and good performance status predict improved survival with gefitinib in patients with advanced non-small cell lung cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 16(5), 780-785

Moriai T, Kobrin M, Hope C, Speck L, Korc M (1994): A variant epidermal growth factor receptor exhibits altered type alpha transforming growth factor binding and transmembrane signaling. *Proc Natl Acad Sci U S A* 91(21), 10217-10221

Moynagh P (2005): The NF-kappaB pathway. *J. Cell Sci.* 118 (Pt 20), 4589-4592

Murphy L, Smith S, Chen R-H, Fingar D, Blenis J (2002): Molecular interpretation of ERK signal duration by immediate early gene products. *Nat. Cell Biol.* 4(8), 556-564

Nakagawa K, Kudoh S, Ohe Y, Johkoh T, Ando M, Yamazaki N, Seki A, Takemoto S, Fukuoka M (2012): Postmarketing surveillance study of erlotinib in Japanese patients with non-small-cell lung cancer (NSCLC): an interim analysis of 3488 patients (POLARSTAR). *J. Thorac. Oncol.* 7(8), 1296-1303

Nanney L, Stoscheck C, King L, Underwood R, Holbrook K (1990a): Immunolocalization of epidermal growth factor receptors in normal developing human skin. *J. Invest. Dermatol.* 94(6), 742-748

Nanney L, Stoscheck C, King L, Underwood R, Holbrook K (1990b): Immunolocalization of epidermal growth factor receptors in normal developing human skin. *The Journal of investigative dermatology* 94(6), 742-748

Normanno N, Maiello MR, De Luca A (2003): Epidermal growth factor receptor tyrosine kinase inhibitors (EGFR-TKIs): simple drugs with a complex mechanism of action? *J. Cell. Physiol.* 194(1), 13-19

Okazaki K, Sagata N (1995): The Mos/MAP kinase pathway stabilizes c-Fos by phosphorylation and augments its transforming activity in NIH 3T3 cells. *The EMBO journal* 14(20), 5048-5059

Olayioye M, Neve R, Lane H, Hynes N (2000): The ErbB signaling network: receptor heterodimerization in development and cancer. *The EMBO journal* 19(13), 3159-3167

Omicinski C, Vanden Heuvel J, Perdew G, Peters J (2011): Xenobiotic metabolism, disposition, and regulation by receptors: from biochemical phenomenon to predictors of major toxicities. *Toxicol. Sci.* 120 Suppl 1 75

Paradis A, Kantoff P, Giovannucci E, Stampfer M, Ma J (2003): Association between the Met326Ile polymorphism of the p85alpha regulatory subunit of phosphatidylinositol 3-kinase and prostate cancer risk: a prospective study. *Cancer Epidemiol. Biomarkers Prev.* 12(2), 172-173

Parmar S, Schumann C, Rüdiger S, Boeck S, Heinemann V, Kächele V, Seeringer A, Paul T, Seufferlein T, Stingl J (2013): Pharmacogenetic predictors for EGFR-inhibitor-associated skin toxicity. *Pharmacogenomics J* 13(2), 181-188

Pastore S, Mascia F, Mariotti F, Dattilo C, Mariani V, Girolomoni G (2005): ERK1/2 regulates epidermal chemokine expression and skin inflammation. *Journal of immunology (Baltimore, Md. : 1950)* 174(8), 5047-5056

Paul T, Schumann C, Rüdiger S, Boeck S, Heinemann V, Kächele V, Steffens M, Scholl C, Hichert V, Seufferlein T et al. (2014): Cytokine regulation by epidermal growth factor receptor inhibitors and epidermal growth factor receptor inhibitor associated skin toxicity in cancer patients. *Eur. J. Cancer* 50(11), 1855-1863

Peeters M, Siena S, Van Cutsem E, Sobrero A, Hendlisz A, Cascinu S, Kalofonos H, Devercelli G, Wolf M, Amado R (2009): Association of progression-free survival, overall survival, and patient-reported outcomes by skin toxicity and KRAS status in patients receiving panitumumab monotherapy. *Cancer* 115(7), 1544-1554

Peréz-Soler R, Saltz L (2005): Cutaneous adverse effects with HER1/EGFR-targeted agents: is there a silver lining? *J. Clin. Oncol.* 23(22), 5235-5246

Pérez-Soler R, Chachoua A, Hammond L, Rowinsky E, Huberman M, Karp D, Rigas J, Clark G, Santabárbara P, Bonomi P (2004): Determinants of tumor response and survival with erlotinib in patients with non-small-cell lung cancer. *J. Clin. Oncol.* 22(16), 3238-3247

Perry M, Williams A, Tsitsiou E, Larner-Svensson H, Lindsay M (2009): Divergent intracellular pathways regulate interleukin-1beta-induced miR-146a and miR-146b expression and chemokine release in human alveolar epithelial cells. *FEBS Lett.* 583(20), 3349-3355

Petrelli F, Borgonovo K, Cabiddu M, Lonati V, Barni S (2012): Relationship between skin rash and outcome in non-small-cell lung cancer patients treated with anti-EGFR tyrosine kinase inhibitors: a literature-based meta-analysis of 24 trials. *Lung Cancer* 78(1), 8-15

Peus D, Hamacher L, Pittelkow M (1997): EGF-receptor tyrosine kinase inhibition induces keratinocyte growth arrest and terminal differentiation. *J. Invest. Dermatol.* 109(6), 751-756

Peuvrel L, Bachmeyer C, Reguiat Z, Bachet J, André T, Bensadoun R, Bouché O, Ychou M, Dréno B (2012): Semiology of skin toxicity associated with epidermal growth factor receptor (EGFR) inhibitors. *Support. Care Cancer* 20(5), 909-921

Pinarbasi H, Silig Y, Gurelik M (2005): Genetic polymorphisms of GSTs and their association with primary brain tumor incidence. *Cancer Genet. Cytogenet.* 156(2), 144-149

Prior I, Lewis P, Mattos C (2012): A comprehensive survey of Ras mutations in cancer. *Cancer Res.* 72 (10), 2457-2467

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, Sklar P, de Bakker P, Daly M et al. (2007): PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81(3), 559-575

Quinlan MP, Settleman J (2009): Isoform-specific ras functions in development and cancer. *Future Oncol.* 5(1), 105-116

Rameh L, Cantley L (1999): The role of phosphoinositide 3-kinase lipid products in cell function. *J. Biol. Chem.* 274(13), 8347-8350

Ramos J (2008): The regulation of extracellular signal-regulated kinase (ERK) in mammalian cells. *Int. J. Biochem. Cell Biol.* 40(12), 2707-2719

Ravid T, Sweeney C, Gee P, Carraway K, Goldkorn T (2002): Epidermal growth factor receptor activation under oxidative stress fails to promote c-Cbl mediated down-regulation. *J. Biol. Chem.* 277(34), 31214-31219

Raz E, Schejter E, Shilo B (1991): Interallelic complementation among DER/flb alleles: implications for the mechanism of signal transduction by receptor-tyrosine kinases. *Genetics* 129(1), 191-201

Renouf D, Tang P, Hedley D, Chen E, Kamel-Reid S, Tsao M, Tran-Thanh D, Gill S, Dhani N, Au H et al. (2014): A phase II study of erlotinib in gemcitabine refractory advanced pancreatic cancer. *Eur. J. Cancer* 50(11), 1909-1915

Repertinger S, Campagnaro E, Fuhrman J, El-Abaseri T, Yuspa S, Hansen L (2004): EGFR enhances early healing after cutaneous incisional wounding. *The Journal of investigative dermatology* 123(5), 982-989

Robert C, Soria J-C, Spatz A, Le Cesne A, Malka D, Pautier P, Wechsler J, Lhomme C, Escudier B, Boige V et al. (2005): Cutaneous side-effects of kinase inhibitors and blocking antibodies. *Lancet Oncol.* 6(7), 491-500

Rodeck U, Jost M, Kari C, Shih D, Lavker R, Ewert D, Jensen P (1997): EGF-R dependent regulation of keratinocyte survival. *J. Cell Sci.* 110 ( Pt 2) 113-121

Roé E, García Muret M, Marcuello E, Capdevila J, Pallarés C, Alomar A (2006): Description and management of cutaneous side effects during cetuximab or erlotinib treatments: a prospective study of 30 patients. *J. Am. Acad. Dermatol.* 55(3), 429-437

Roidl D, Hacker C (2014): Histone methylation during neural development. *Cell Tissue Res.* 356(3), 539-552

Roskoski R (2012): ERK1/2 MAP kinases: structure, function, and regulation. *Pharmacol. Res.* 66(2), 105-143

Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB (2013): Characterizing and measuring bias in sequence data. *Genome Biol.* 14(5),

Rudin C, Liu W, Desai A, Karrison T, Jiang X, Janisch L, Das S, Ramirez J, Poonkuzhali B, Schuetz E et al. (2008): Pharmacogenomic and pharmacokinetic determinants of erlotinib toxicity. *J. Clin. Oncol.* 26(7), 1119-1127

Saif MW, Merikas I, Tsimboukis S, Syrigos K (2008): Erlotinib-induced skin rash. Pathogenesis, clinical significance and management in pancreatic cancer patients. *Jop* 9(3), 267-274

Salmena L, Carracedo A, Pandolfi P (2008): Tenets of PTEN tumor suppression. *Cell* 133(3), 403-414

Sanger F, Nicklen S, Coulson A (1977): DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74(12), 5463-5467

Schlessinger J (2004): Common and distinct elements in cellular signaling via EGF and FGF receptors. *Science* 306(5701), 1506-1507

Schork N, Murray S, Frazer K, Topol E (2009): Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* 19(3), 212-219

Schrag D, Chung K, Flombaum C, Saltz L (2005): Cetuximab therapy and symptomatic hypomagnesemia. *J. Natl. Cancer Inst.* 97(16), 1221-1224

Schuurman M, van Waardenburg D, Da Costa J, Niemarkt H, Leroy P (2009): Severe hemolysis and methemoglobinemia following fava beans ingestion in glucose-6-phosphatase dehydrogenase deficiency: case report and literature review. *Eur. J. Pediatr.* 168(7), 779-782

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M et al. (2004): Large-scale copy number polymorphism in the human genome. *Science* 305(5683), 525-528

Segaert S, Van Cutsem E (2005): Clinical signs, pathophysiology and management of skin toxicity during therapy with epidermal growth factor receptor inhibitors. *Ann. Oncol.* 16(9), 1425-1433

Segaert S, Chiritescu G, Lemmens L, Dumon K, Van Cutsem E, Tejpar S (2009): Skin toxicities of targeted therapies. *Eur. J. Cancer* 45 Suppl 1 295-308

Shen H, Li J, Zhang J, Xu C, Jiang Y, Wu Z, Zhao F, Liao L, Chen J, Lin Y et al. (2013): Comprehensive characterization of human genome variation by high coverage whole-genome sequencing of forty four Caucasians. *PLoS One* 8(4),

Shepherd F, Rodrigues Pereira J, Ciuleanu T, Tan E, Hirsh V, Thongprasert S, Campos D, Maoleekoonpiroj S, Smylie M, Martins R et al. (2005): Erlotinib in previously treated non-small-cell lung cancer. *N. Engl. J. Med.* 353(2), 123-132

Shi Y, Zhang L, Liu X, Zhou C, Zhang L, Zhang S, Wang D, Li Q, Qin S, Hu C et al. (2013): Icotinib versus gefitinib in previously treated advanced non-small-cell lung cancer (ICOGEN): a randomised, double-blind phase 3 non-inferiority trial. *Lancet Oncol.* 14(10), 953-961

Sibilia M, Wagner EF (1995): Strain-dependent epithelial defects in mice lacking the EGF receptor. *Science* 269(5221), 234-238

Sizeland AM, Burgess AW (1992): Anti-sense transforming growth factor alpha oligonucleotides inhibit autocrine stimulated proliferation of a colon carcinoma cell line. *Mol. Biol. Cell* 3(11), 1235-1243

Stokoe D, Macdonald S, Cadwallader K, Symons M, Hancock J (1994): Activation of Raf as a result of recruitment to the plasma membrane. *Science* 264(5164), 1463-1467

Stoll S, Garner W, Elder J (1997): Heparin-binding ligands mediate autocrine epidermal growth factor receptor activation in skin organ culture. *The Journal of clinical investigation* 100(5), 1271-1281

Tan A, Steinberg S, Parr A, Nguyen D, Yang S (2008): Markers in the epidermal growth factor receptor pathway and skin toxicity during erlotinib treatment. *Ann. Oncol.* 19(1), 185-190

Tarabeux J, Zeitouni B, Moncoutier V, Tenreiro H, Abidallah K, Lair S, Legoix-Né P, Leroy Q, Rouleau E, Golmard L et al. (2014): Streamlined ion torrent PGM-based diagnostics: BRCA1 and BRCA2 genes as a model. *European journal of human genetics : EJHG* 22(4), 535-541

Thatcher N, Chang A, Parikh P, Rodrigues Pereira J, Ciuleanu T, von Pawel J, Thongprasert S, Tan EH, Pemberton K, Archer V et al. (2005): Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet* 366(9496), 1527-1537

Thorvaldsdóttir H, Robinson J, Mesirov J (2013): Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14(2), 178-192



Threadgill D, Dlugosz A, Hansen L, Tennenbaum T, Lichti U, Yee D, LaMantia C, Mourton T, Herrup K, Harris R (1995): Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. *Science* 269 (5221), 230-234

Togashi Y, Masago K, Fujita S, Hatachi Y, Fukuhara A, Nagai H, Sakamori Y, Kim Y, Mio T, Mishima M (2011): Differences in adverse events between 250 mg daily gefitinib and 150 mg daily erlotinib in Japanese patients with non-small cell lung cancer. *Lung Cancer* 74 (1), 98-102

Tsimboukis S, Merikas I, Karapanagiotou E, Saif M, Syrigos K (2009): Erlotinib-induced skin rash in patients with non-small-cell lung cancer: pathogenesis, clinical significance, and management. *Clin. Lung Cancer* 10 (2), 106-111

Uribe JM, Gelbmann CM, Traynor-Kaplan AE, Barrett KE (1996): Epidermal growth factor inhibits Ca(2+)-dependent Cl- transport in T84 human colonic epithelial cells. *Am. J. Physiol.* 271 (3 Pt 1), C914-922

Vakiani E, Solit D (2011): KRAS and BRAF: drug targets and predictive biomarkers. *J Pathol* 223 (2), 219-229

Van Cutsem E, Peeters M, Siena S, Humblet Y, Hendlisz A, Neyns B, Canon J-L, Van Laethem J-L, Maurel J, Richardson G et al. (2007): Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. *J. Clin. Oncol.* 25 (13), 1658-1664

Van Cutsem E, Siena S, Humblet Y, Canon JL, Maurel J, Bajetta E, Neyns B, Kotasek D, Santoro A, Scheithauer W et al. (2008): An open-label, single-arm study assessing safety and efficacy of panitumumab in patients with metastatic colorectal cancer refractory to standard chemotherapy. *Ann. Oncol.* 19 (1), 92-98

Visscher P, Brown M, McCarthy M, Yang J (2012): Five years of GWAS discovery. *Am. J. Hum. Genet.* 90 (1), 7-24

Volinia S, Dhand R, Vanhaesebroeck B, MacDougall L, Stein R, Zvelebil M, Domin J, Panaretou C, Waterfield M (1995): A human phosphatidylinositol 3-kinase complex related to the yeast Vps34p-Vps15p protein sorting system. *The EMBO journal* 14 (14), 3339-3348

Wang L-E, Ma H, Hale K, Yin M, Meyer L, Liu H, Li J, Lu K, Hennessy B, Li X et al. (2012): Roles of genetic variants in the PI3K and RAS/RAF pathways in susceptibility to endometrial cancer and clinical outcomes. *J. Cancer Res. Clin. Oncol.* 138 (3), 377-385

Wang W-S, Chen P-M, Chiou T-J, Liu J-H, Lin J-K, Lin T-C, Wang H-S, Su Y (2007): Epidermal growth factor receptor R497K polymorphism is a favorable prognostic factor for patients with colorectal carcinoma.

*Clinical cancer research : an official journal of the American Association for Cancer Research* 13 (12), 3597-3604

Wang X, McCullough K, Franke T, Holbrook N (2000): Epidermal growth factor receptor-dependent Akt activation by oxidative stress enhances cell survival. *J. Biol. Chem.* 275 (19), 14624-14631

Ward H (1973): Anti-oestrogen therapy for breast cancer: a trial of tamoxifen at two dose levels. *Br. Med. J.* 1 (5844), 13-14

Wellcome Trust Case Control C (2007): Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447 (7145), 661-678

Wennerberg K, Rossman K, Der C (2005): The Ras superfamily at a glance. *J. Cell Sci.* 118 (Pt 5), 843-846

Wheeler D, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen Y-J, Makhijani V, Roth G et al. (2008): The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452 (7189), 872-876

Wolf M, Swaisland H, Averbuch S (2004): Development of the novel biologically targeted anticancer agent gefitinib: determining the optimum dose for clinical efficacy. *Clin. Cancer Res.* 10 (14), 4607-4613

Wong S-F (2005): Cetuximab: an epidermal growth factor receptor monoclonal antibody for the treatment of colorectal cancer. *Clin. Ther.* 27 (6), 684-694

Woodworth C, Michael E, Marker D, Allen S, Smith L, Nees M (2005): Inhibition of the epidermal growth factor receptor increases expression of genes that stimulate inflammation, apoptosis, and cell attachment. *Mol. Cancer Ther.* 4 (4), 650-658

Wu C-H, Fallini C, Ticozzi N, Keagle P, Sapp P, Piotrowska K, Lowe P, Koppers M, McKenna-Yasek D, Baron D et al. (2012): Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature* 488 (7412), 499-503

Yan X-J, Xu J, Gu Z-H, Pan C-M, Lu G, Shen Y, Shi J-Y, Zhu Y-M, Tang L, Zhang X-W et al. (2011): Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat. Genet.* 43 (4), 309-315

Yarden Y, Sliwkowski MX (2001): Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.* 2 (2), 127-137

Zhao W, Blagev D, Pollack JL, Erle DJ (2011): Toward a systematic understanding of mRNA 3' untranslated regions. *Proc Am Thorac Soc* 8 (2), 163-166

Zheng D, Radziszewska A, Woo P (2012): MicroRNA 497 modulates interleukin 1 signalling via the MAPK/ERK pathway. *FEBS Lett.* 586(23), 4165-4172

## Acknowledgement

I would like to gratefully thank Prof. Dr. Jürgen Brockmöller for his supervision, support and his kind help for precisely reviewing the current manuscript.

My special thanks are to PD Dr. Mladen Tzvetkov who taught me how to do a systematic research, systematic thinking and many aspects of biology.

I thank our kind and patient technical assistant Mrs. Karoline Jobst and my fellow lab mates in the institute of clinical pharmacology and institute of pharmacology: Nawar, Sina, Kate, Claudia, Kristin, Tina, Sebastian, Joao, Robert, Jiayin, Konrad, and Ali.

## CURRICULUM VITAE

Sayed Mohammad Hasheminasab

I was born on June 4<sup>th</sup> in 1983 in Sary which is located between the Alborz mountain range and the Caspian Sea in the north of Iran. I have attended the Beheshti high school which is under the supervision of the National Organization for Developing Exceptional Talents (NODET) in Sary. I graduated in 2001 from the school. In the same year I have participated in general university entrance examination and achieved the 90<sup>th</sup> place in experimental sciences category.

Later in 2001 I started the human medicine major in Tehran University of Medical Sciences. During that time I was performing research projects at the department of Human genetics under supervision of Dr. Modarressi and Dr. Ashtiani. The research projects were mainly on cancer genetics and I have contributed to a number of publications. I was graduated in December 2009.

I joined the Institute of Clinical Pharmacology for scientific work in November 2011. Since then I was working on the present project and some other scientific projects such as functional studying of a plasma monoamine transporter under the supervision of Dr. M. Tzvetkov and Professor J. Brockmüller.

### Publications

Aminian K, Farsad F, Ghanbari A, Fakhreih S, Hasheminasab SM (2010): A randomized trial comparing four *Helicobacter pylori* eradication regimens: standard triple therapy, ciprofloxacin based triple therapy, quadruple and sequential therapy. *Trop. Gastroenterol.* 31(4), 303-307

Ashtiani ZO, Hasheminasab SM, Ayati M, Goulian BS, Modarressi MH (2011): Are GSTM1, GSTT1 and CAG repeat length of androgen receptor gene polymorphisms associated with risk of prostate cancer in Iranian patients? *Pathol. Oncol. Res.* 17(2), 269-275

Ashtiani ZO, Heidari M, Hasheminasab SM, Ayati M, Rakhshani N (2012): Mitochondrial D-Loop polymorphism and microsatellite instability in prostate cancer and benign hyperplasia patients. *Asian Pac. J. Cancer Prev.* 13(8), 3863-3868

Ghafouri-Fard S, Ousati Ashtiani Z, Sabah Golian B, Hasheminasab SM, Modarressi MH (2010): Expression of two testis-specific genes, SPATA19 and LEMD1, in prostate cancer. *Arch. Med. Res.* 41(3), 195-200

Motevaseli E, Shirzad M, Raoofian R, Hasheminasab SM, Hatami M, Dianatpour M, Modarressi MH (2013): Differences in vaginal lactobacilli composition of Iranian healthy and bacterial vaginosis infected women: a comparative analysis of their cytotoxic effects with commercial vaginal probiotics. *Iranian Red Crescent medical journal* 15(3), 199-206

Seraj JM, Yazdani N, Ashtiani ZO, Seraj SM, Hasheminasab SM, Memar B, Mirashrafi F, Borghei H, Yazdani J, Mostaan LV (2011): TP53 gene expression in HPV-positive oral tongue SCC and its correlation with nodal metastasis. *Pathol. Res. Pract.* 207 (12), 758-761