Aus dem Pathologischen Institut
der Ludwig-Maximilians-Universität München
Direktor: Prof. Dr. Thomas Kirchner

The polymorphic DNA marker rs849142 predicts skin toxicity in anti-EGFR treatment of metastatic colorectal cancer

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von

Matthias Frank Frölich
aus Lingen (Ems)

2018

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter	Prof. Dr. Andreas Jung
Mitberichterstatter	PD Dr. Marlies Michl Prof. Dr. Frank Kolligs Prof. Dr. Georg Enders
Dekan	Prof. Dr. Reinhard Hickel
Tag der mündlichen Prüfung	11.10.2018

Meinen Eltern gewidmet.

Table of Contents

Zusammenfassung

Das Kolonkarzinom stellt sowohl in Deutschland als auch weltweit eine der häufigsten Ursachen für krebsassoziierte Todesfälle dar [1]. Bei Vorliegen von Metastasen ist die Therapie mit gegen den EGF-Rezeptor gerichteten Antikörpern (α–EGFR) wie Cetuximab – neben Chirurgie, Radio- und Chemotherapie – eine wichtige Säule der Therapie [2].

Nebenwirkungen dieser Therapie in der Haut, welche als skin toxicity (ST) bezeichnet werden, können die Lebensqualität durch schwere acneiforme Hautausschläge – Skin Rash – stark beeinträchtigen [3]. Es wurde eine statistische Assoziation zwischen einem erhöhten ST Schweregrad und der Überlebensdauer gezeigt [4]. Bisher existiert keine Möglichkeit das Auftreten von ST bei Patienten sicher vorherzusagen. Daher sind keine gezielten präventiven Maßnahmen oder Dosisanpassungen vor Therapiebeginn möglich. Auch gibt es keine etablierten prognostischen Marker für Überleben und Ansprechen von Patienten unter α–EGFR Therapie.

Auf Basis der Modellerkrankungen wurden im Rahmen dieser Arbeit DNA Polymorphismen (SNPs) in Zusammenhang mit 1) Acne vulgaris, 2) Systemischem Lupus erythematodes und 3) Fc-Immunrezeptoren als mögliche Marker für ST identifiziert und mit Next-Generation-Sequencing Ansätzen analysiert: 64 SNPs wurden in einem ‚training-set' aus 16 Patienten mit ausschließlich hoch- oder geringgradiger ST und in einem ‚validation-set' aus 55 Patienten aller ST-Schweregrade mit Hilfe des Ion Torrent PGM analysiert.

In training- und validation-set konnte der bekannte Zusammenhang von ST und Überlebensdauer bestätigt werden. Eine Analyse der Assoziation der Genotypen mit ST-Schweregraden zeigte eine signifikante Assoziation des SNPs rs849142 mit ST im trainings- ($p=0{,}00395$) und validation-set ($p=0{,}04362$). Ein signifikanter Zusammenhang von rs849142 zum Überleben ließ sich in dem Patientenkollektiv jedoch nicht zeigen.

Damit stellt rs849142 einen validierten potentiell prädiktiven Marker für den Schweregrad von ST bei α–EGFR-Therapie des metastasierten kolorektalen Karzinoms dar.

rs849142 ist ein SNP im JAZF1 (juxtaposed with another zinc finger protein 1) Gen. JAZF1 fungiert als Transkriptionsrepressor [5], spielt eine Rolle im Lipid- und Glucosemetabolismus [6] und wird unter anderem von Blutzellen exprimiert. rs849142 wurde bereits im Zusammenhang mit dem Prostatakarzinom [7], dem systemischen Lupus erythromatodes [8] und der Lupus Nephritis [9] beschrieben.

Summary

Colorectal cancer represents one of the most common causes of cancer-associated death in Germany and worldwide [1]. In metastatic stage, the use of monoclonal antibodies targeted at the EGF receptor (α–EGFR) like Cetuximab is, besides surgery, radiotherapy and chemotherapy, an important therapy strategy [2].

Common adverse events of α–EGFR therapy include severe acneiform rashes in the skin, are summarized as skin toxicity and often decrease patients' quality of life. [3, 10]. A statistical association of severity of skin toxicity (ST) and survival has been found [4]. However, there is no established marker which can predict occurrence of ST. Consequently, preventive treatments or preventive dose-adjustments, which take the severity of ST into account, are difficult to carry out. Furthermore, there is no established prognostic marker for survival and response to α–EGFR therapy.

Based on model-diseases, this work identified DNA polymorphisms (SNPs) associated to 1) acne, 2) systemic lupus erythematosus and 3) Fc-immunoreceptors as potential targets for testing of their association to ST in Next-Generation-Sequencing: 64 SNPs were analyzed in a 'training set' consisting of 16 patients only with high or low grade ST and a 'validation set' consisting of 55 patients of with all ST severities utilizing the Ion Torrent PGM sequencer.

In both sets the known association of ST with survival could be validated. An analysis of association SNPs with ST severity revealed the SNP rs849142 as significantly associated to ST in training ($p=0.00395$) and validation set ($p=0.04362$). However, rs841942 was not found to be significantly associated to survival in this patient collective.

Therefore, rs849142 is a validated, potentially predictive marker for severity of ST in α–EGFR therapy of metastatic colorectal cancer. rs849142 is a SNP in JAZF1 (juxtaposed with another zinc finger protein 1) gene. JAZF1 functions as transcription factor [5], influenced lipid- and glucose metabolism, [6] and is expressed by blood cells. rs849142 has been found to be associated to prostate cancer [7], systemic lupus [8] and lupus nephritis [9].

1. Introduction

1.1 Metastatic Colorectal Cancer is one of the major causes of cancer related death in Germany

Colorectal Cancer is among the first three types of malignant tumors with respect to new cases and cancer-related deaths in both the US [1] and Europe [11]. In its metastatic stage – referred to as metastatic colorectal cancer (mCRC) – pharmacological therapy focuses on chemotherapy (CTX), biologicals like antibodies targeting the epidermal growth factor receptor (EGFR) [2], vascular endothelial growth factor (VEGF) or radiotherapy, which is mostly used for treatment of rectal cancer. [12]

The decision for targeted anti-EGFR-therapy (α–EGFR) is personalized and predictive as efficacy of α–EGFR therapy is limited to tumors without mutations (wild-type, WT) in the rat sarcoma (RAS) genes KRAS (Kirstin RAS) and NRAS (neuronal RAS) [13, 14]. Therefore, the approval authorities, EMA (*European Medicines Agency*) and FDA (*Food and drug administration*) granted the use of the two α–EGFR directed antibodies Cetuximab (CET) and Panitumumab (PAN) only for mCRC characterized by a RAS-WT genotype. Despite improving response and survival rates, α–EGFR agents often cause adverse effects in the skin termed skin toxicity (ST).

1.2 α-EGFR therapy associated skin toxicity decreases patients' quality of life

Adverse events associated with α-EGFR therapies primarily present themselves as skin toxicity (ST) in the form of acneiform skin rash [15]. This results in the reduction of patients' physical [16] and emotional well-being [10] as well as negative financial- and psychosocial effects [17]. Consequently, ST is treated by topical and systemic treatment with antibiotics like tetracycline and immune-response modulating drugs like methylprednisolone [15] dependent on its grade (Table 1).

Grade of ST	Description	Consequences for treatment
1	'Papules and/or pustules covering <10% BSA [body surface area], which may or may not be associated with symptoms of pruritus or tenderness'	No changes
2	'Papules and/or pustules covering 10%-30% BSA, which may or may not be associated with symptoms of pruritus or tenderness; associated with psychosocial impact; limiting instrumental ADL [activities of daily living]'	Topical treatment

3	‘Papules and/or pustules covering >30% BSA, which may or may not be associated with symptoms of pruritus or tenderness; limiting self-care ADL; associated with local superinfection with oral antibiotics indicated’	Temporal interruption of therapy and systemic treatment
4	‘Papules and/or pustules covering any BSA%, which may or may not be associated with symptoms of pruritus or tenderness and are associated with extensive superinfection with IV antibiotics indicated; life- threatening consequences	End of therapy, systemic treatment
5	‘Death’	-

Table 1: Grading of skin toxicity using the Common Terminology Criteria for Adverse Events (CTCAE) 4.0 [3] and grade-dependent treatment of acneiforme skin rash [15], Adopted from [3].

Taking the grade-dependence of α-EGFR treatment into account, a predictive marker for ST would allow personalized preventive treatment of ST. Thus, patients without risk of developing ST would not receive any treatment of the skin thus avoiding interference of treatment with the therapy. In contrast, patients with a risk for high grade ST might receive intensified preventive treatment of ST allowing a more intensive α-EGFR targeted therapy as higher ST-grades were shown to be associated with a better response to α–EGFR therapy [4]. This is translated into improved survival (OS, PFS or both) estimated by Kaplan Meier [18]. Therefore, the identification of a ST predictive biomarker would allow a better adjustment of ST-treatment and also prediction of response on targeted α–EGFR therapy in mCRC. Unfortunately, such a marker has not been established for the clinical practice, yet.

Considering that ST occurs in an organ not related to the primary tumor, and the fact that the molecular mechanism driving ST are not fully understood, it is a useful and unbiased strategy to search for predictive biomarkers by screening for any type of DNA-polymorphisms (SNPs) which are associated with ST. In a second step it might then be helpful to deduce possible molecular mechanisms bases on the gene(s) in which the polymorphisms were found to build a model for α–EGFR related ST. This model can subsequently be tested by translation onto clinical cases giving also an explanation why there are variations of ST grade among different patients (Table 1). This approach can be done by an unbiased, genome-wide-association-study, investigating all variations across the whole genome. However, for this approach hundreds, better thousands of individuals are required [19, 20]. As it is difficult to get hands on such huge numbers of identically treated patients, thus from a clinical study, to exclude co-confounders as much as possible, a targeted approach is a reasonable alternative. For this approach, analysis of selected DNA-markers is done. Thus, fewer cases are in need for a

robust analysis but the difficulty clearly lies in the selection process of appropriate DNA polymorphisms or genes as biomarkers, which are putatively important. From a point of biomedical reasoning it can be assumed that such biomarkers should be related to molecular mechanisms of ST.

1.3 Putative mechanisms of skin toxicity include skin-environment, immune regulation and immune-cell receptors

Several theories on the molecular mechanisms underlying ST have been postulated. [21] As ST occurs in the skin, skin-related effects have to be taken into account: EGFR is expressed in basal keratinocytes [22] and its inhibition causes apoptosis of keratinocytes [23]. Therefore, one possible explanation for the occurrence of ST might be a higher susceptibility for EGFR-inhibition of the skin-environment. This effect may be related to changes in the EGFR as well as EGFR-independent mechanisms.

Another mechanism is related to the immune system: As a part of their mechanism of action, immune cells bind therapeutic antibodies like α-EGFR-antibodies at their Fc-region by Fc γ receptors (FCRG). With their antigen specific part these antibodies make contact with high level EGFR expressing tumor cells thus bringing together the immune- and tumor cells. In case the immune cells have cytotoxic potential, like NK (natural killer) cells, the connected cell, for example the tumor cell, will die. Thus, this process has been termed antibody-dependent-cellular-cytotoxicity (ADCC) [24]. This mechanism might help to explain variances in the severity or grade of ST, which might depend on differences in immune cell regulation and involved immune receptors.

1.4 DNA polymorphisms associated with acne, SLE and Fc receptor γ (FcRγ) are promising biomarkers of skin toxicity

In a next step, potential DNA polymorphisms related to the possible pathogenic mechanisms mentioned have to be selected. Assuming that ST is a tumor independent effect, selection of non-oncological model diseases, which match the putative mechanisms, can be one promising approach:

- First, a common disease linked to deregulation and changes in the skin environment is acne [25] which shares many characteristics with ST, which often presents as acneiform rash. Several SNPs altering susceptibility to acne have been identified [26, 27].

- Second, systemic lupus erythematosus (SLE) is linked to immune dysregulation including skin effects. Several SNPs were found to be associated with SLE [28, 29]. As the skin-related manifestations [30] and part of the pathogenesis in SLE patients might involve severely deregulated immune cells [31], polymorphisms associated with SLE may be used as predictors for ST. In addition, Fc-receptor polymorphisms are also associated with SLE [32].
- Third, SNPs in the family of immunoglobulin receptors (Fc receptors, fragment crystallizable) which bind the therapeutic α-EGFR antibodies may also influence the affinity of the antibody. On the other side, SNPs in the EGFR and its pathways have been described in the context of ST already without identification of clinically used polymorphisms [33, 34]. However, SNPs in Fc-Receptors may vary the intensity of the immune system related effect ADCC. Given CET is an IgG1-antibody [35], it binds to Fc-γ receptors (FcRγ) [36]. In particular, a polymorphism in the FCGR3A gene has been shown to influence the IgG-binding capacity of Fc-receptors [37]. To investigate possible effects of Fcγ-Receptor polymorphisms, common polymorphisms in genes encoding Fc-γ-receptors (FcRγ: FCGR2A, FGR2B, FCGR3A, FCGR3B) might be promising targets.

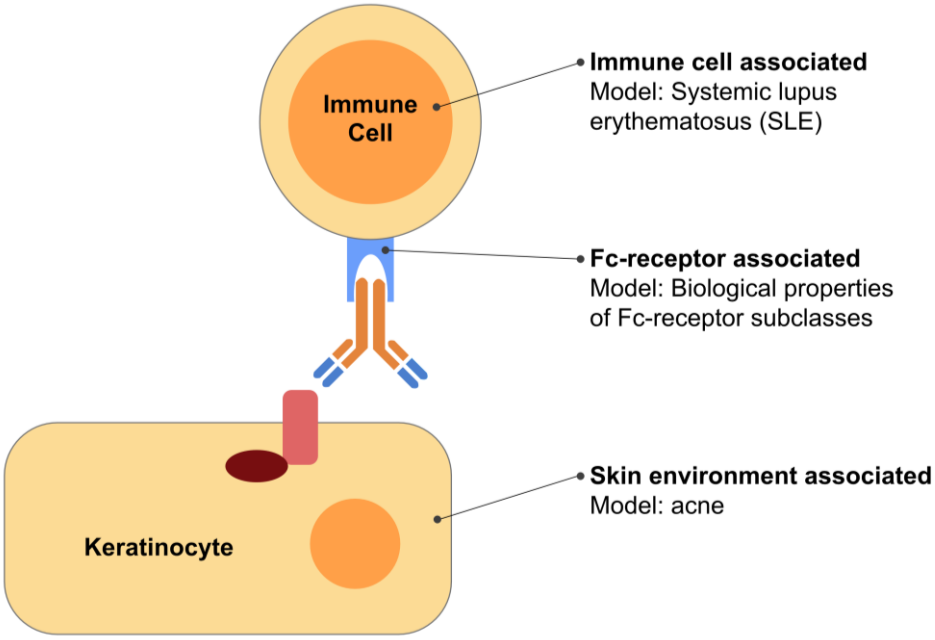

Fig. 1: Groups of promising biomarkers for ST. EGFR downstream and upstream targets were included: Immune cell associated targets known to be associated to systemic lupus erythematosus (SLE), Fc-receptor (fragment crystallizable) polymorphisms and skin environment associated polymorphisms known to be associated with acne were selected based on literature.

For the measurement of several distinct SNPs in different parts of the genome, a targeted next-generation-sequencing (NGS) including bioinformatic evaluation is the standard-approach.

2. Material and Methods

2.1 Metastatic Colorectal Cancer patients treated with Cetuximab from randomized phase 3 study

Stage IV mCRC patients (age between 18 and 75 years) treated with the α-EGFR monoclonal antibody Cetuximab were selected from the FIRE-3 study, which was an open-label randomized phase 3 study including stage IV histologically confirmed mCRC comparing a FOLFIRI CTX backbone together with either CET or bevacizumab (BEV) [13]. As part of the study the parameters gender, age, ECOG performance status, ORR, survival and Skin Rash grade (CTCAE 4.0) were collected besides others.

For this study two subgroups were selected from the FOLFIRI, CET (n=380) study arm:

1. the training set (n=16) consisting of ST grade 1 and 3 patients and
2. the validation set (n=55) consisting of ST grade 1, 2 and 3 patients. Patients with and without RAS and BRAF mutations were included.

Probes and clinical data were procured by Prof. Dr. Sebastian Stintzing and Prof. Dr. Volker Heinemann.

2.2 DNA preparation and isolation

The analysis aimed for genomic alterations, namely individual SNPs of the patient genomic DNA was applicable for this study. Therefore, DNA was isolated from peripheral blood mononuclear cells (PBMNC) by spinning down (10min, 1600xg, room temperature) 5ml of patient's blood taking the buffy coat. This was washed two times in phosphate buffered saline (PBS). Subsequently, the sediment was dissolved in lysis buffer of QIAGEN's blood DNA kits. Finally, DNA was isolated following the instructions given in the manual of the vendor resulting in high quality DNA.

Then the concentration of the DNA containing solution was measured employing UV-photometry determining the optical densities at λ=260nm (OD_{260}), λ=280nm (OD_{280}), and λ=320nm (OD_{320}).

2.3 Targeted Next-Generation-Sequencing using Custom Panels in Ion Torrent PGM

Primer design was done employing Thermo Fisher's Ion Torrent AmpliSeq™ Custom Panel Designer [38] resulting in three Ion AmpliSeq™ Custom DNA Hotspot Panels (Table 2).

Panel	Targets	Coverage	Number of Amplicons	Pipeline Version	Solution Type	DNA Type	Application type
Acne	14	100%	12	4.0	High Specificity	Standard DNA	germline
SLE	36	100%	35	4.0	High Specificity	Standard DNA	germline
FcRγ	20	100%	20	3.6	High Specificity	Standard DNA	germline

Table 2: Custom Panel generation. Input parameters utilized for automated generation of primer panels using Ion Torrent AmpliSeq™ Custom Panel Designer (Thermo Fisher) [38].

For generation of libraries, DNA quality was checked in a first set of steps. On the one hand, the DNA concentration was measured using the Qubit-system (Thermo Fisher, $conc_Q$) following the user's manual. Additionally, the quality of the DNA was determined by running a quantitative (q)-PCR using the patient's DNA together with a detection system specific for the RNase P-gene (Roche Diagnostics). The concentration of amplifiable DNA was calculated on the basis of a standard curve ($conc_{RNaseP}$). Subsequently, the ratio of both concentration values was calculated ($\frac{conc_Q}{conc_{RNaseP}}$). DNAs with a ratio above the factor 2 were not used but instead DNA was re-purified again as inhibitory factors were left over like iron-ions from the heme of the lysed erythrocytes or others. This tedious work was done to save costs in the following expensive library generation step.

Libraries were made using 10ng of DNA based on $conc_{RNaseP}$ together with Ion AmpliSeq™ Library Kit 2.0 in 21 PCR-cycles following the instructions given in the manufacturer's protocol. Subsequently barcodes (Ion Xpress™ Barcode Adapters) were added by the FuPa reaction in the Ion AmpliSeq™ Library Kit protocol to assort for the analysis of data the sequences to individual patients indexed by their individual barcode. After this step, libraries were ready for clonal amplification. Therefore, comparable amounts of the libraries should be put together to reach comparable coverage rates. Thus, another concentration determination was run using the qPCR system Quantitizer (Thermo Fisher) following the user's manual. Each 30pmol of 16 libraries (patients) were put together into a total volume of 70µl which was transferred into the Ion Chef™ pipetting station (Thermo Fisher) together with Ion PGM™ IC-200 Supplies and Ion 318™ Chips v2 BC based on the protocol of the Ion PGM™

IC-200 Kit. The Ion PGM™ IC 200 Kit components include template reagents, nucleotides, enzyme mixes, spike-in sequencing controls, buffers & solutions, consumables, and supplies. In the Ion Chef™ a clonal amplification of individual library molecules was done by emulsion PCR (emPCR) on solid phase represented by primer labeled beads. After this amplification step the Ion Chef™ separated DNA carrying from empty beads and loaded them together with sequencing-primers and Taq-Polymerase onto IonChips-316™ which were now ready to be loaded into the Ion Torrent PGM™ (Personal Genome Machine, Thermo Fisher).
For each run on the Ion Torrent PGM™ a run protocol was written linking the patient's ID with the barcode used. Wash buffers, sodium hydroxide and nucleotides were mounted on the Ion Torrent PGM™. Next a test IonChip-316™ was mounted on the PGM and a pre-run was launched to allow the setting of a neutral pH value of the wash buffer and to get rid of air bubbles in the tubes. After this step the test chip was exchanged with a library/ primer/ DNA-polymerase loaded IonChip-316™ and the run was finally started.

Wet-lab work was carried out and thereby raw-data generated by members of the Jung Lab, especially Gertrud Lenz.

2.4 Quality control and mapping using Ion Reporter

Sequencing results were aligned to the human reference genome built hg19 thereby generating BAM *(binary alignment metafiles)*, which were transferred from the Ion Torrent PGM™ on the local Ion Server System™. With the help of the Ion Torrent Suite™ software BAM files were transferred into the Thermo cloud. This transfer was done by SSL encryption and using the anonymous run names for the cases to protect patient's subgenomic genetic data.
For fine analysis of the data the Ion Reporter™ platform (Version 4.6) was used. In a first step, the reads created by the sequencer had to be aligned to the human reference genome in a mapping process. hg19 built of the human genome was used for subsequent steps. The mapping was carried out on Ion Reporter™ platform utilizing the optimized settings for Ion Torrent PGM™ data.

2.5 Variant Calling and Filtering using Ion Reporter

The information on the status of the target variants had to be determined from the aligned read in the BAM file in a process called variant calling. Ensured by the design of the sequencing process, most variants were covered by several sequencing reads for each patient. Given this

information, the putative status of the variant (reference or non-reference) and statistical errors were determined employing bioinformatics:

Variants were called applying a workflow generated in the Ion Reporter™ including annotation of dbSNP138 based on the genome built hg19 [39]. Hotspot region files matching the sequencing targets were included to enhance calling of reference-SNPs of interest. Filter criteria were set for germline variant detection including minimum SNP variant score of 10.0, minimum SNP allele frequency of 0.1 and minimum SNP coverage 10. These filter settings ensured the necessary sensitivity to detect the variants while limiting false-positive calls to a minimal level.

2.6 Variant data processing

Variant call-results were exported as data-files in the VCF *(*variant call format*)* and TSV (tab separated values) formats and appended in the spreadsheet program Microsoft Excel® version 15. The resulting file allowed determining the number of successful variant calls per panel and patient serving as a quality measure for the panels including information from all patients. As the data sets should be analyzed by the software package PLINK [40], which allowed variant association studies by linkage analysis sequencing calls and ST phenotypes, they had to be reformatted to match PED and MAP format specifications of PLINK. [40]

Caused by unpredictable biological properties of the sequencing process, sequencing quality varies between variants in the process. These result in varying performance of the variant calling process. The proportion of missing genotypes ($\frac{number\ of\ missing\ genotypes_{SNP}}{number\ of\ patients_{SNP}}$) allows an estimation of sequencing quality of each variant in the patient groups. To obtain this estimation, the '--missing' algorithm of PLINK version 1.07 (v1.07) [40] was applied to calculate the number and proportion of missing genotypes per SNP. Results were visualized using Haploview version 4.2 software's plotting functionality. [41]

2.7 Analysis for Deviations from Hardy-Weinberg Equilibrium using PLINK

As a biology-based indicator of consistency, deviations from Hardy-Weinberg-Equilibrium [42] were calculated using the exact Hardy-Weinberg test [43] statistic function of PLINK v1.07 using p=0.001 as threshold. Gender information was not included into the analysis.

2.8 Genotype-Phenotype association analysis using PLINK

The quality-tested results from sequencing representing the genotype information of each patient had to be analyzed with respect to the phenotype information – represented in ST levels. This kind of analysis of the association of phenotypes with genotypes is known as 'association analysis' and is based on regression models. Genotype phenotype association analyses for quantitative markers [44] were performed using PLINK v.1.07. p-values were calculated using Wald statistic [45]. Individuals with no gender information were included into the analysis. Again results were visualized using the plotting functionality of Haploview version 4.2 [41].

2.9 Linkage analysis using Haploview

Variants determined in the association analysis were tested for genetic association with other variants in a process called 'linkage analysis'. It gives information if the variants found are likely to be causal for the genotype-phenotype association found or may be bystanders. Groups of associated variants – called Haplotypes – were analyzed using Haploview v4.2 [41] with block definitions from Gabriel et al. [46]. Linkage Disequilibrium (LD) data was obtained from Hapmap database utilizing the Haploview download function.

2.10 Survival Analysis based on Kaplan Meier estimators

Survival of patient groups was calculated using the Kaplan Meier model [18] in SPSS Statistics version 24. p-values pooled over strata were calculated using the Log-Rank test [47]. Results were visualized using survival diagrams.

3. Results

3.1 Skin toxicity was significantly associated with survival in the FIRE-3 study

First of all, it was tested if there was an association between the ST status and survival in our sub-collection of the FIRE-3 study applying the Kaplan Meier model because this correlation had been described before [4], thus indicating the validity of the FIRE-3 study as well as that these cases were suitable to investigate the hypothesis. As a training-validation approach was aimed for, the two randomly chosen sub-groups of training set and validation set both were tested independently of each other. A statistically significant association between ST status and survival for both subgroups was demonstrated for both PFS ($p_{\text{training set}}$=0.016, $p_{\text{validation set}}$=0.004) and OS ($p_{\text{training set}}$=0.059, $p_{\text{validation set}}$<0.001) (Fig. 2). As these results clearly reflected published data this support was taken to support the idea that the FIRE-3 subgroup of patients was representative for a ST-patient collective.

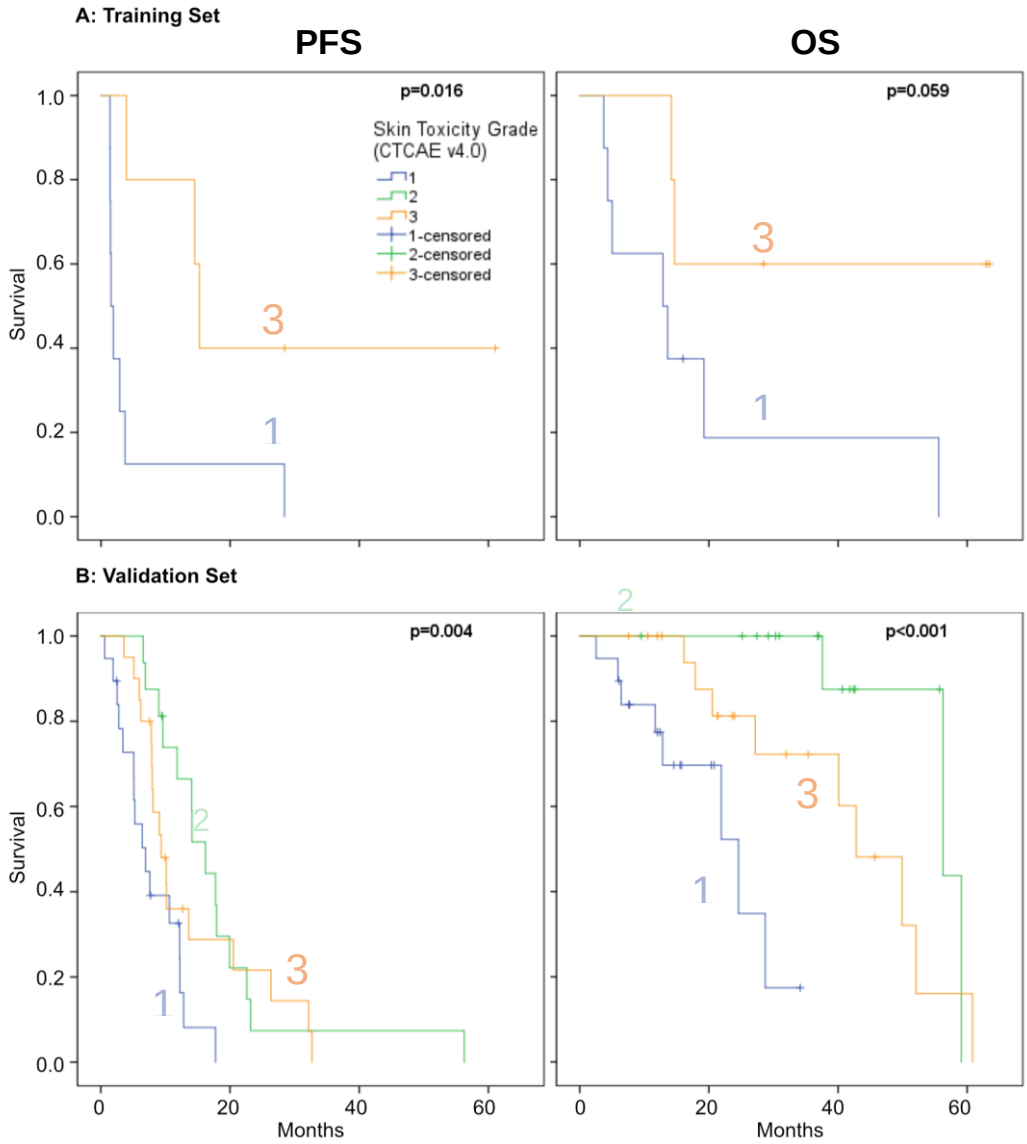

Fig. 2: Severity of skin toxicity is associated to progression free survival (PFS) and overall survival (OS).

3.2 Patient collective for training-validation-approach selected

To test the quality of potential biomarkers the statistical training / validation model was used as an approach using the following:

1. The training set consisted exclusively of low (CTCAE 1)- and high (CTCAE 3)- grade ST phenotypes. With this high contrast between the two states an increase in the sensitivity for genetic association analyses was expected. [44]
2. The validation set also included intermediate-grade (CTCAE 2) ST phenotypic patients as this was a better approximation for collectives of real patients.

In addition, as significant SNPs from the training set were evaluated in the analysis of the validation set, adjustment for multiple testing was not necessary [48]. To ensure comparability of patients and data collection within and across both groups, selection of patients from one clinical study has major advantages. Therefore, a clinical study with a adequate number of metastatic colorectal cancer patients receiving α–EGFR therapy was employed to obtain patients for both training and validation set.

To do so, training and validation subgroups were selected from the FIRE-3 study collection resulting in 71 patients totally (16 patients in the training set; 55 patients in the validation set). These two groups did not match perfectly but were reasonably comparable (Table 3). The presented groups were included in the subsequent analyses.

Variable		Training (n=16)	Validation (n=55)	p
Age	Mean (SD)	64,231 (8.89)	62,418 (9.75)	0.543
Sex	Male	6 (37,50%)	43 (78,18%)	
	Female	7 (43,75%)	12 (21,82%)	0.049
	Unknown	3 (18,75%)	0 (0,00%)	
RAS	Wild-type	8 (50,00%)	47 (85,45%)	
	Mutated	1 (6,25%)	8 (14,55%)	1.000
	Unknown	7 (43,75%)	0 (0,00%)	
Skin Toxicity	Grade 1	11 (68,75%)	19 (34,55%)	
	Grade 2	0 (0,00%)	16 (29,09%)	0.017
	Grade 3	5 (31,25%)	20 (36,36%)	
ORR	No	7 (43,75%)	11 (20,00%)	
	Yes	3 (18,75%)	40 (72,68%)	0.007
	unknown	6 (37,50%)	4 (7,32%)	

Table 3: Overview of patients in the training- and validation set. Molecular pathological and clinical data of the patients included in the training- and validation sets. p values are based on Chi-Square tests for categorial variables and t test for continuous variables. (ORR; overall response rate, RAS; rat sarcoma viral oncogene homolog, SD; standard deviation)

3.3 70 Single nucleotide polymorphisms associated to acne, SLE, FcRγ were selected for the investigation of skin toxicity

In parallel, SNPs associated with susceptibility to acne, SLE or FcRγ binding were selected based on screening of literature and online databases (Fig. 3).

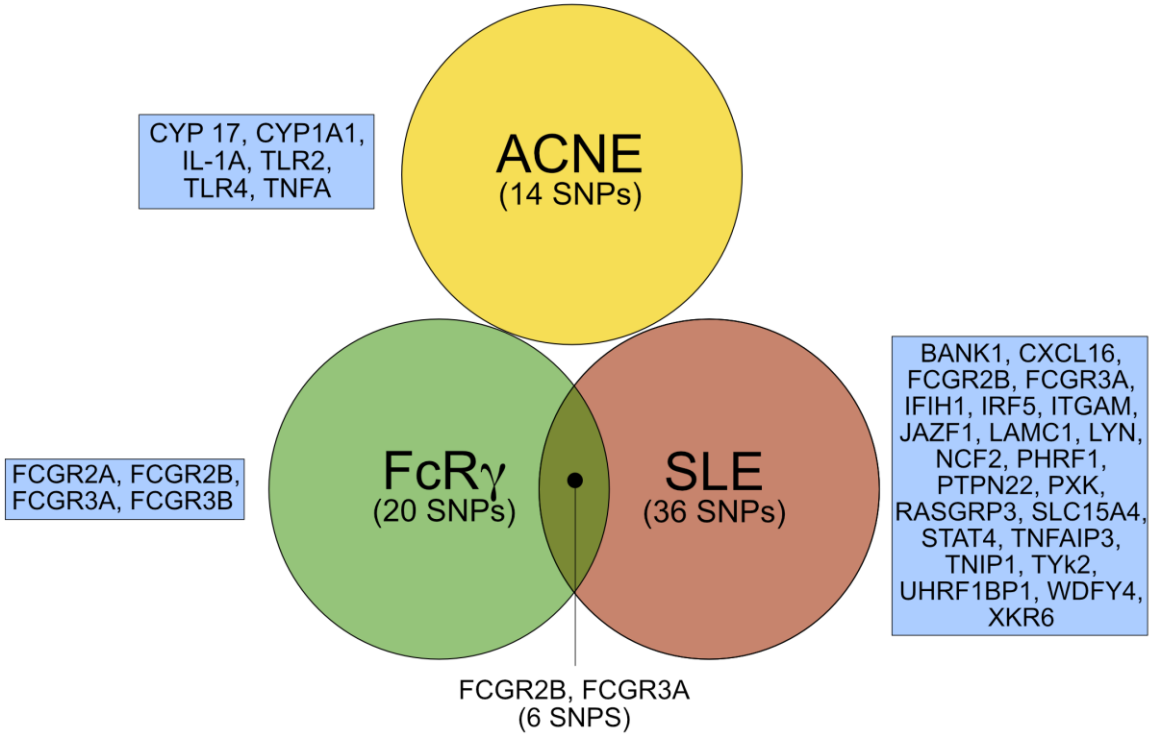

Fig. 3: SNPs and respective genes selected for targeted sequencing (Additional information: Appendix A.2; FcRγ; Fc receptor γ, SNP; single nucleotide polymorphism, SLE; systemic lupus erythematosus)**.**

14 SNPs associated to acne [26], 20 SNPs associated to FcRγ [39] and 36 SNPs associated to SLE [28] were found to be putative biomarkers for ST (Fig. 3). Six of these SNPs were found in both the SLE and FcRγ group as well, because the FcRγ also has a functional role in SLE.

3.4 SNP panel sufficiently covered by Custom AmpliSeq Panels

Next an amplicon panel was generated which covered all previously identified SNPs. Therefore, the commercially available AmpliSeq™ Custom Panel Designer (Ion Torrent) was employed. [38] As some of the SNPs were located on a single amplicon unit the number of amplicons was consequently smaller than that of the SNPs. It was possible to cover all SNPs with three individual panels (Table 4).

Panel	Nº of SNPs	% covered in panel	Nº of amplicons	AmpliSeq™ version
acne	14/14	100%	12	4.0
SLE	36/36	100%	35	4.0
FcRγ	20/20	100%	20	3.6

Table 4: Coverage of targets for sequencing in constructed primer panels (Full list refer to: Appendix A.1)

3.5 Coverage of SNPs by the panels is adequate

Having created panels covering the SNP-targets, sequencing of all patients was performed using the Ion Torrent PGM platform. To obtain the variant information for each case, a bioinformatics pipeline was applied to the resulting data:

1. First, absolute genomic positions of raw data reads from the Ion Torrent PGM™ Sequencer had to be identified. This mapping was done integrating the Ion Reporter™ Software (version 4.6). Low quality reads, which could not be mapped to the reference hg19, were dropped out.
2. Second, from all reads mapped to each target region, the respective variant had to be identified by variant calling. As SNPs are germline variants, fewer reads and thus coverage for each analyzed region were necessary to call the variant compared to when analyzing tumorgenetic mutations from heterogeneous probes like tumors. Variant calling was also done with help of run by the Ion Reporter Software (version 4.6) including BED-files containing the target regions. During this step, preferences for germline variants and quality filters based on coverage, minimum allele frequencies and p-value-based scores were applied.

This procedure resulted in the identification of genotypes for each variant and patient provided that reads in sufficient quality were available.

In a next step, this variant information was used to get an overview of the performance of sequencing for each SNP with regard to the whole patient group. This allowed an estimation of the quality of the panels which was determined by measuring the share of missing genotypes for each SNP:

$$Missing\ genotypes\ (\%) = \frac{Number\ of\ missing\ genotypes_{SNP}}{Total\ number\ of\ patients}$$

The results for non-missing variants of each of the SNPs (colored points) were visualized using the Haploview software (Fig. 4). [41]

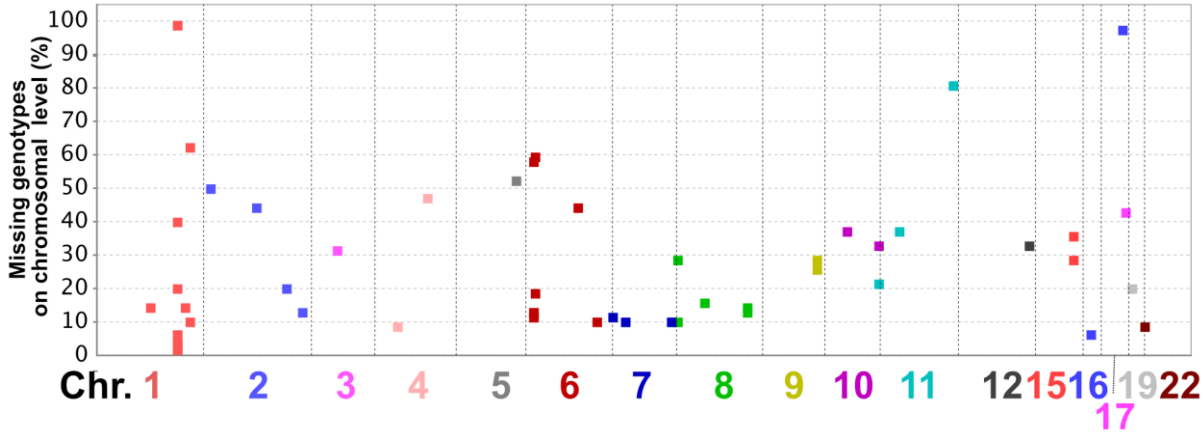

Fig. 4: Proportion of missing genotypes for each SNP – shows the coverage of the individual sequencing targets. A low proportion shows that the respective variant was called in the majority of patients (Full list Appendix A.3)

In conclusion, all but two SNPs showed at least one called genotype (Fig. 4). Furthermore, the majority of SNPs showed a coverage of more than 50%. This resembled a satisfactory coverage and was reached by the majority of SNPs included in the panels. A low genotyping rate is taken into account by the following analyses. Therefore, SNPs with few calls were not excluded from further analyses. Furthermore, some SNPs not targeted in the first place were also sequenced and identified by the variant calling algorithms, because of their close location to the panels' original targets. These variants were included into all further analyses leading to a higher total number of variants.

3.6 Homogeneous, high sequencing quality for DNA probes

Another quality aspect beside the quality of the whole panel, which represents a combined quality measure value, is the sequencing quality of each individual DNA probe, which was checked for in a subsequent step. For this task the number of successfully sequenced genotypes for each patient was investigated (Fig. 5).

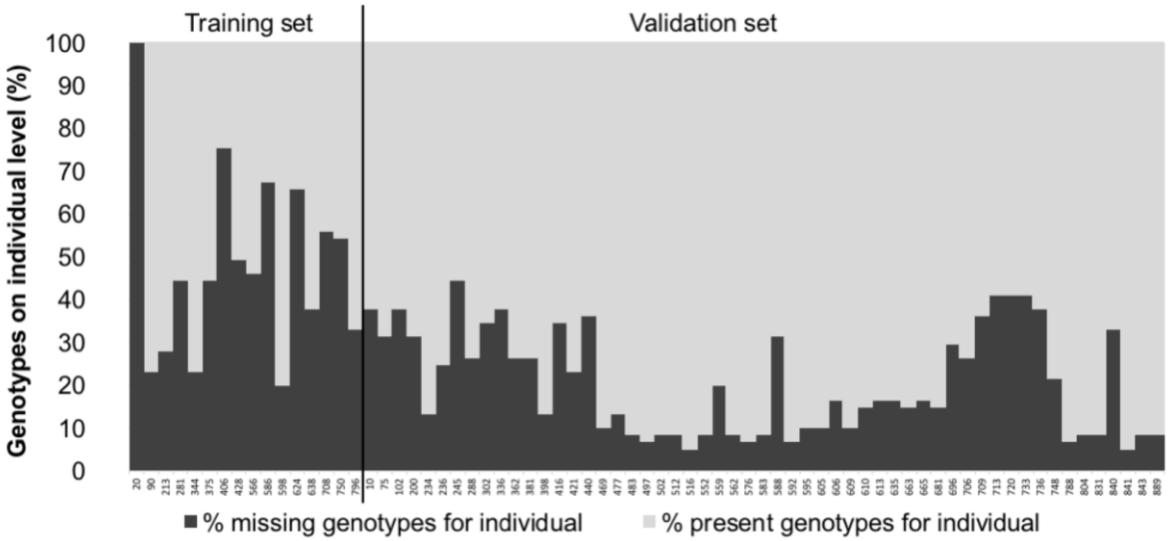

Fig. 5: Number of variants called per patient – shows the completeness of variant calls per patient. For the majority of patients, the panels show an appropriate number of present genotypes (Overview table in Appendix A.3)

To conclude, the majority of patients was analyzed with a good to at least sufficient coverage for all panels. Therefore, this set of data could be utilized for further analyses. In total, variant calls were available for for 61 patients.

3.7 Only a single marker, rs1674761, showed a significant deviation from Hardy-Weinberg equilibrium indicating a systematic error

The quality checks performed up to now only considered the technical quality of the data but did not consider possible systematic errors like population stratification or genotyping errors. As these systematic errors often lead to artificial deviations from the Hardy-Weinberg-equilibrium (HWE) and, as a result, false positive results. Based on a HWE model, these potential false positive results can be detected [19]. For this issue the PLINK software, which allowed a linkage analysis of the different SNPs, was employed. A strict exclusion level was chosen in that the p-value had to be lower than 0.001. Only a single marker (rs1674761, FcRγ-panel, FCGR2B gene) was found to deviate significantly from HWE ($p<0.001$).

Thus, after this final step of quality control a set of 68 SNPs and 61 patients remained for the further functional analysis to figure out association or linkage of SNPs with ST.

3.8 rs849142 significantly associated with Skin Toxicity

To investigate a potential association between SNPs and ST, the genetic association analysis routine of PLINK was employed now. Therefore, an association between genotypes (SNPs) and severity of ST for each patient group was calculated. Here, SNPs showing more than two

alleles were excluded from the analysis, as they cannot be evaluated in a typical quantitative genetic association analysis.

For quantitative traits like ST, regression models including the phenotype as dependent variable and genotype – coded as number of non-reference alleles – as explanatory variable allowed an evaluation of interdependence of genotype and phenotype. This so-called quantitative association analysis could be performed for multiple SNPs using PLINK (v. 1.07, [40]). The results of this analysis were visualized in Manhattan plots again using Haploview [41]. These plot associated p values for each SNP. Significance levels were visualized as blue (p=0.05) or red (p=0.01) lines. Data points above the respective blue and red lines indicated SNPs which were significantly associated with ST on the given respective significance levels (Fig. 6: training set; Fig. 7: validation set; Fig. 8: whole data set).

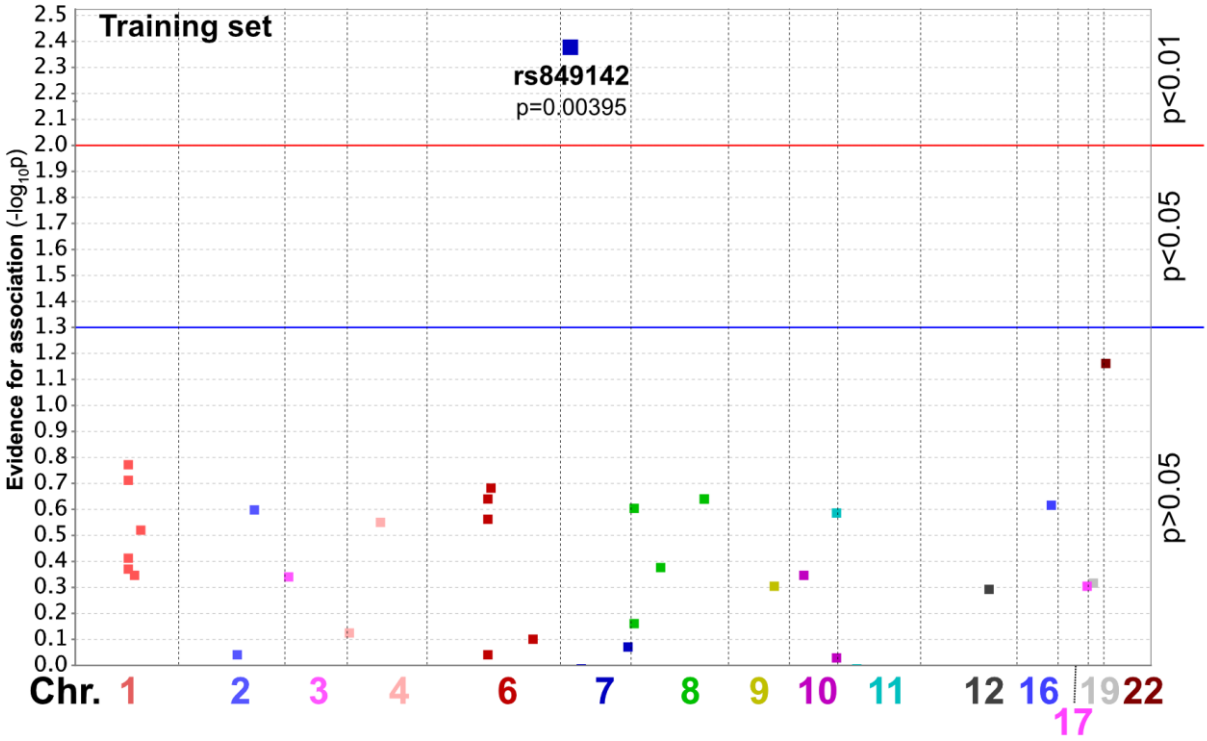

Fig. 6: Manhattan plot for association of SNPs with ST in Training Set: rs849142 shows a significant (p=0.00395) association to ST. The position over the red line indicates a higher significance level than p=0.01. All other SNPs positioned below the blue line are associated with a significance of weakter than p=0.05. (Full list Appendix Table A.4)

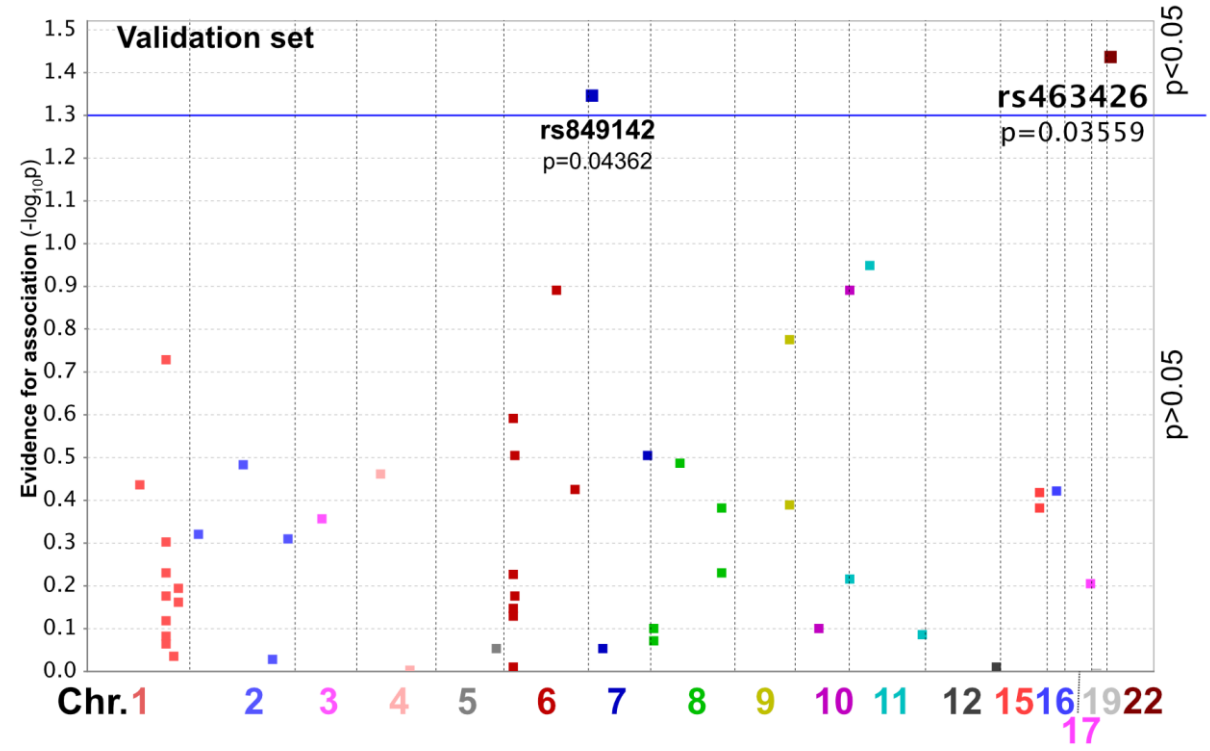

Fig. 7: Manhattan plot for association of SNPs with ST in Validation set: rs849142 and rs463426 show a significant ($p<0.05$) association to ST. Therefore, they are shown above the blue line. All other SNPs are insignificant on $p=0.05$, represented by their position below the blue line. (Full list Appendix Table A.4)

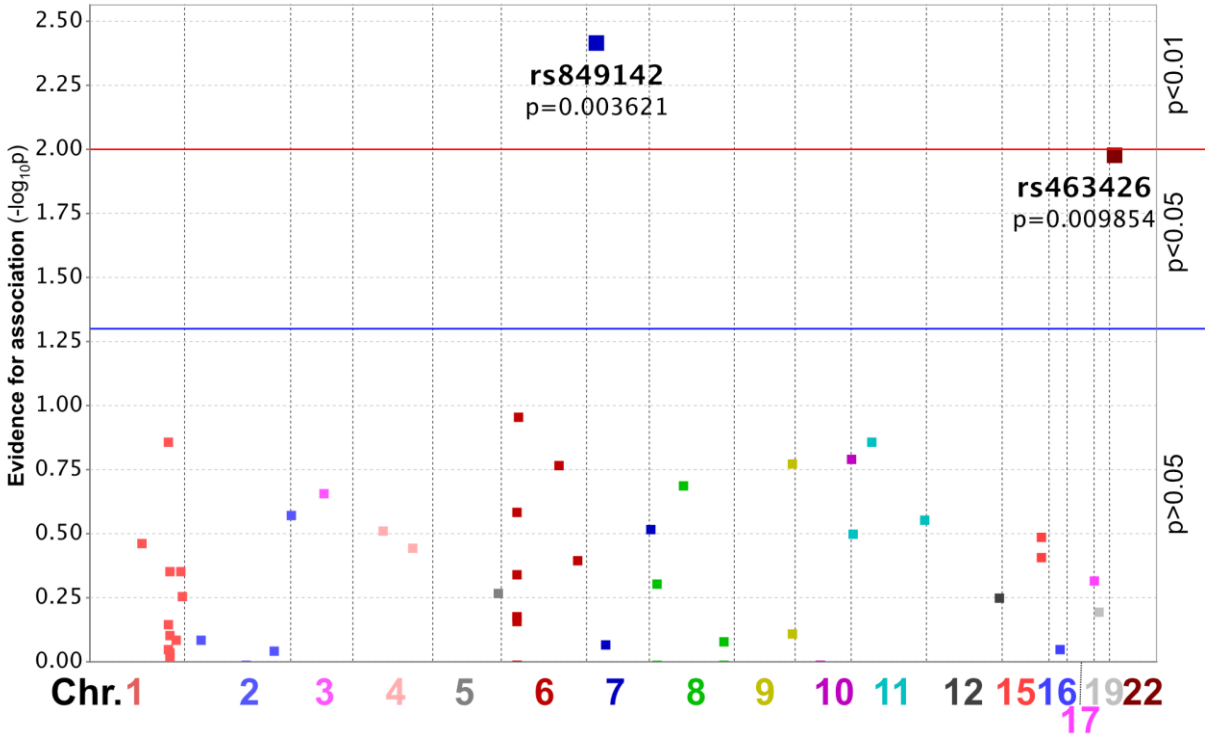

Fig. 8: Manhattan plot for association of SNPs with ST in Both Groups: rs849142 and rs463426 remain significant on $p=0.05$ when analyzing both groups together. No additional SNPs appear as significant. (Full list Appendix Table A.4)

The SNP rs849142 T allele compared to C allele was significantly associated with ST in the training set (p=0.00395, Fig. 6). This result could be confirmed in the validation set (p=0.04362, Fig. 7) though the p value in the validation set showed a less clear significance. This phenomenon is a well-known statistical effect in groups containing intermediate phenotypes: As SNPs do often change the probability of a phenotype but do not show a full penetrance, in case-control approaches including non-extreme phenotypes they are more difficult to unravel statistically [44].
In contrast the SNP rs463426 was found to be significantly associated to ST in the validation and overall groups only. As it failed to reach significance in the training set (p=0.0663) it was not included into further analyses.

Taken together, rs849142 turned out to be a potential, validated candidate predictive biomarker of ST in metastatic colorectal cancer. In particular, an elevated number of T alleles (in contrast to C alleles) was associated with higher ST-grade. Thus, as the major milestone of this work, rs849142 genotyping could be used as a personalized estimator for ST-risk in α-EGFR treatment of metastatic colorectal cancer at least when using Cetuximab, what might turn out as an advantage for patients and physicians.

3.9 rs849142 is not associated with survival

As ST is also associated with survival (Fig. 2) [4] and rs849142 predicts ST, it was now investigated if SNP rs849142 also correlated with survival and may thus also be a useful predictive biomarker for α-EGFR targeted therapy complementing RAS mutations. To investigate this potential relationship, the Kaplan Meier model together with Overall Log-Rank-Estimations were applied following the same scheme of training- and validation set as described already before (Fig. 9).

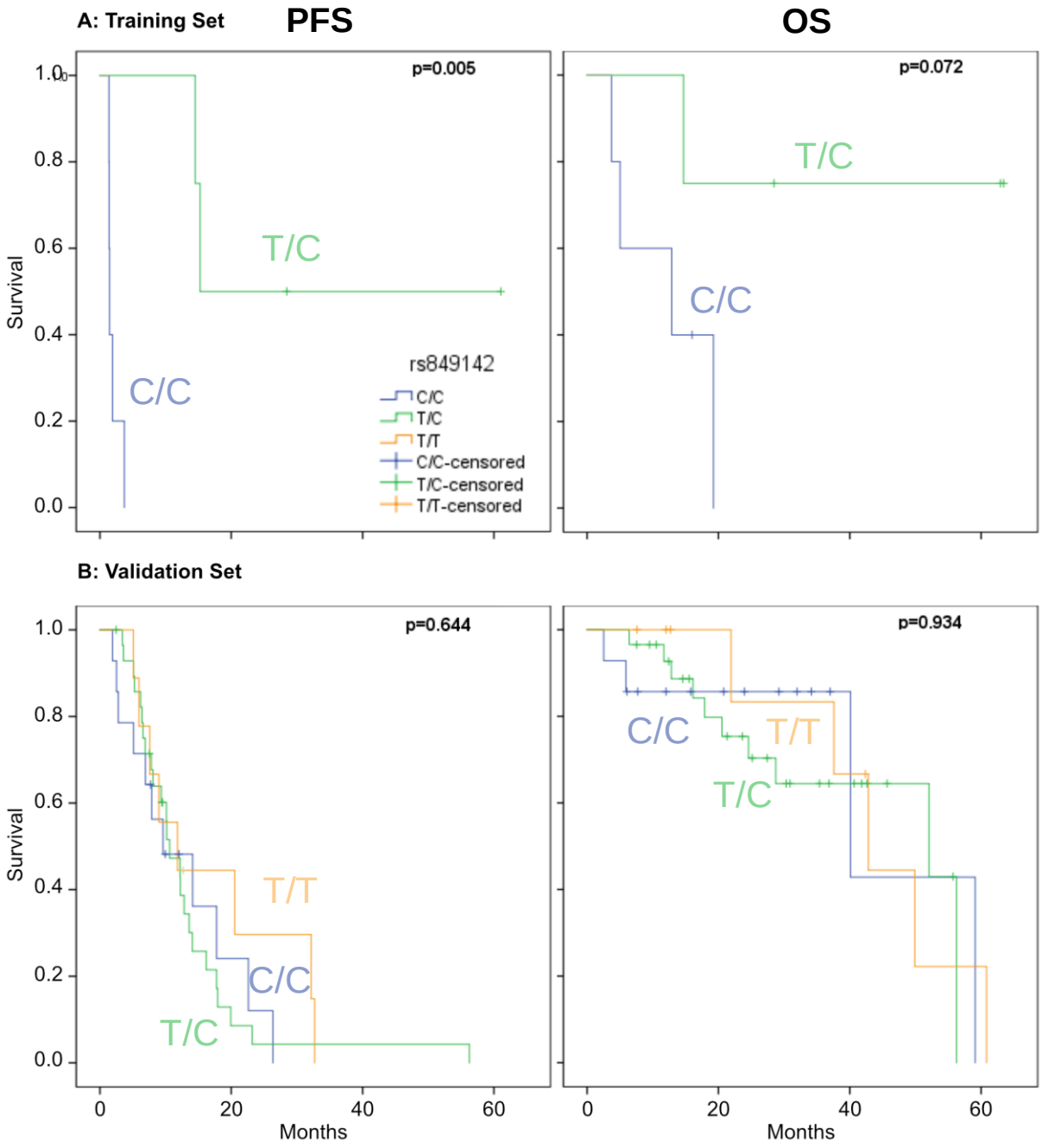

Fig. 9: Kaplan Meier estimators for survival stratified by rs849142 genotypes.

In the training set a significant association of rs849142 with survival was found for PFS (p=0.005) but not OS (p=0.072). However, no association neither with PFS nor OS were confirmed in the validation set (PFS p=0.644, OS p=0.934). Therefore, rs849142 was not associated with survival and is thus not an appropriate predictive biomarker for survival in anti-EGFR targeted therapy.

3.10 Deciphering a functional role for rs849142 using linkage disequilibrium

In a final approach it was investigated if the genomic context of rs849142 is associated to specific genes which might help to model to understand the molecular mechanism underlying ST or at least the association of rs849142 and ST. DNA fragments located next to each other are often inherited together as they build up a coupling group except in cases a breakage-point is located between them. Thus, SNPs especially close to each other often show a highly positive correlation in their co-occurrence. This is referred to as linkage disequilibrium. Even in cases where linkage disequilibrium is found between two loci on then DNA like a SNP and a gene it is difficult to differentiate between causal and bystander effects (Fig. 10) [49]. Nevertheless, such an approach is at least worth to be tried. Therefore, two approaches where used here:

1. Is rs849142 located in a gene and is this gene known to be associated with ST or ST-like effects?
2. Might other genes flanking rs849142 on the same coupling group be responsible for ST or ST-like effects?

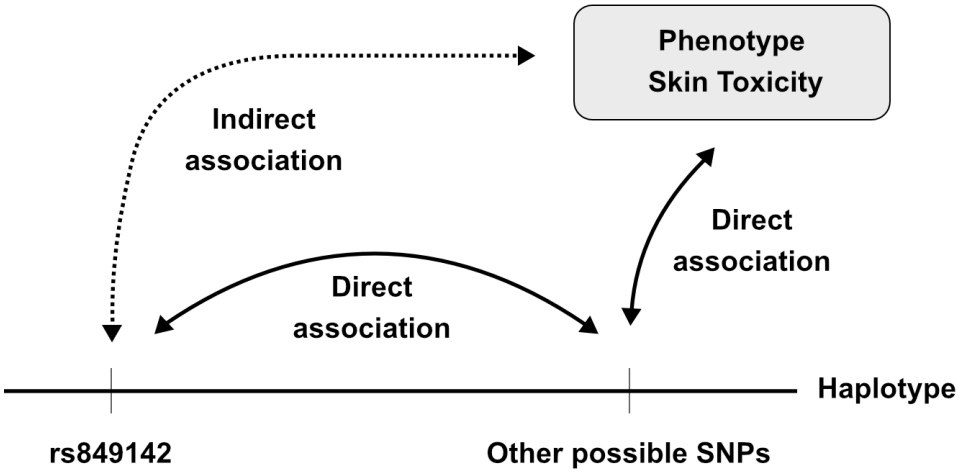

Fig. 10: Biology of a haplotype: rs849142 may be associated with another causal SNP and therefore only indirectly associated to skin toxicity. Adopted from Balding 2006 [49]

3.11 rs849142 is located in the JAZF1 gene

rs849142 is located on chromosome 7 (position 28146272 in GRCh38) which is part of the intronic region of the JAZF1 gene (Juxtaposed with another zinc finger protein 1 gene) also known as TIP27 (27 kDa protein TAK1-interacting protein), a repressor of transcriptional activation by TAK1 [5]. TAK1 was shown to influence glucose homeostasis related to PI3-Kinase/AKT in mice [6] and lipid metabolism in adipocytes [50]. Thus, this gene does not

seem to have obvious effects in ST. To investigate the importance in ST-related tissues, the mRNA expression profile of JAZF1 was recovered from the GTEx-portal (http://www.gtexportal.org) and was subsequently analyzed (Fig. 11). The highest expression profiles for JAZF1 are found in whole blood followed by sigmoid colon and also relatively high expression in adipose tissue but less prominent in the skin.

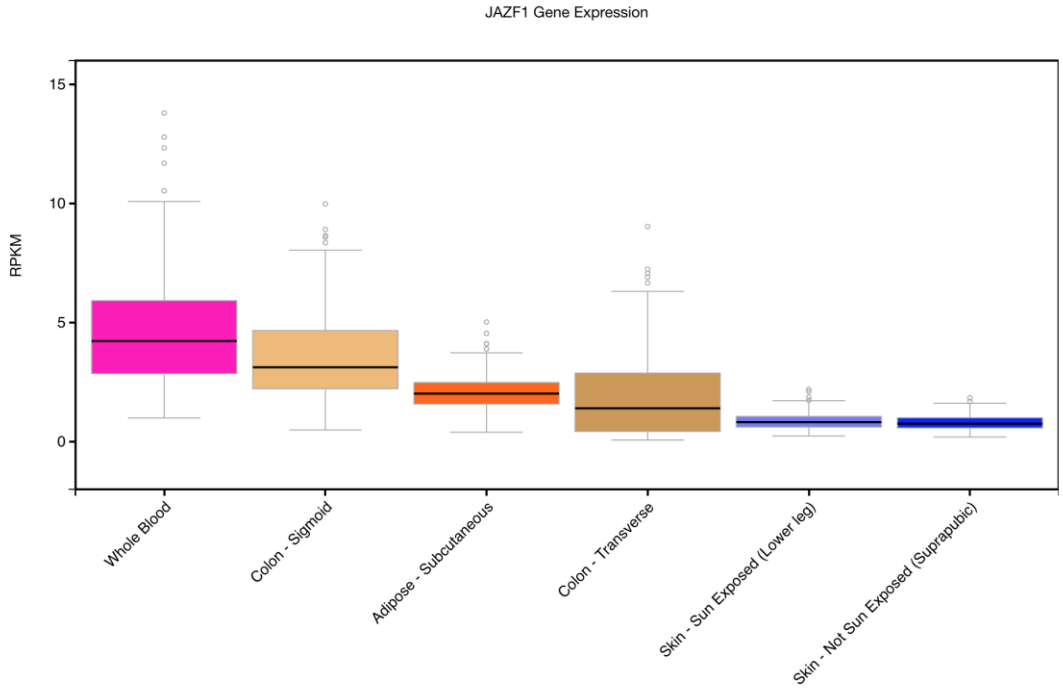

Fig. 11: Expression of JAZF1 gene in several normal tissues: JAZF1 is expressed in blood, colon and skin as well (GTEx Analysis Release V6 (dbGaP Accession phs000424.v6.p1); http://www.gtexportal.org/). Estimated in 'Reads Per Kilobase of transcript per Million mapped reads'.

3.12 rs849142 shows linkage disequilibrium with other SNPs but not other genes

Thus, in a next step the linkage disequilibrium rs849142 was analyzed by linkage analysis of SNPs neighboring rs849142 employing the Haploview software again [41]. Genomic variants near each other tend to be statistically associated (Fig. 10): New variants occur in the context of preexisting other variants – referred to as haplotype. Therefore, the new variants are associated to variants in the respective haplotype. This is described as variants being in linkage disequilibrium (LD) within the population. As long as no recombination or new mutation occurs, the variant remains associated to the Haplotype. These variants can be

pooled in haplotype blocks using linkage disequilibrium measures*. Thus, variants 100 kilobases up- and downstream of rs849142 were analyzed for LD and haplotype blocks.

Linkage analysis revealed several SNPs within in linkage disequilibrium with rs849142 which were visualized in LD- (linkage disequilibrium) plots (Fig. 12). Importantly, only smaller blocks (coupling groups) were found (Fig. 13).

rs849142 was found in a block together with the two other SNPs rs864745 and rs849140, respectively (Fig.13). However, rs849142 was also in high LD with several other SNPs (Fig. 12).

Therefore, the differentiation between a causal and a bystander effects of rs849142 remained unclear. But such a differentiation of markers is irrelevant for clinical purposes: A non-causal marker in linkage disequilibrium with a causal marker would also serve as a surrogate measurement.

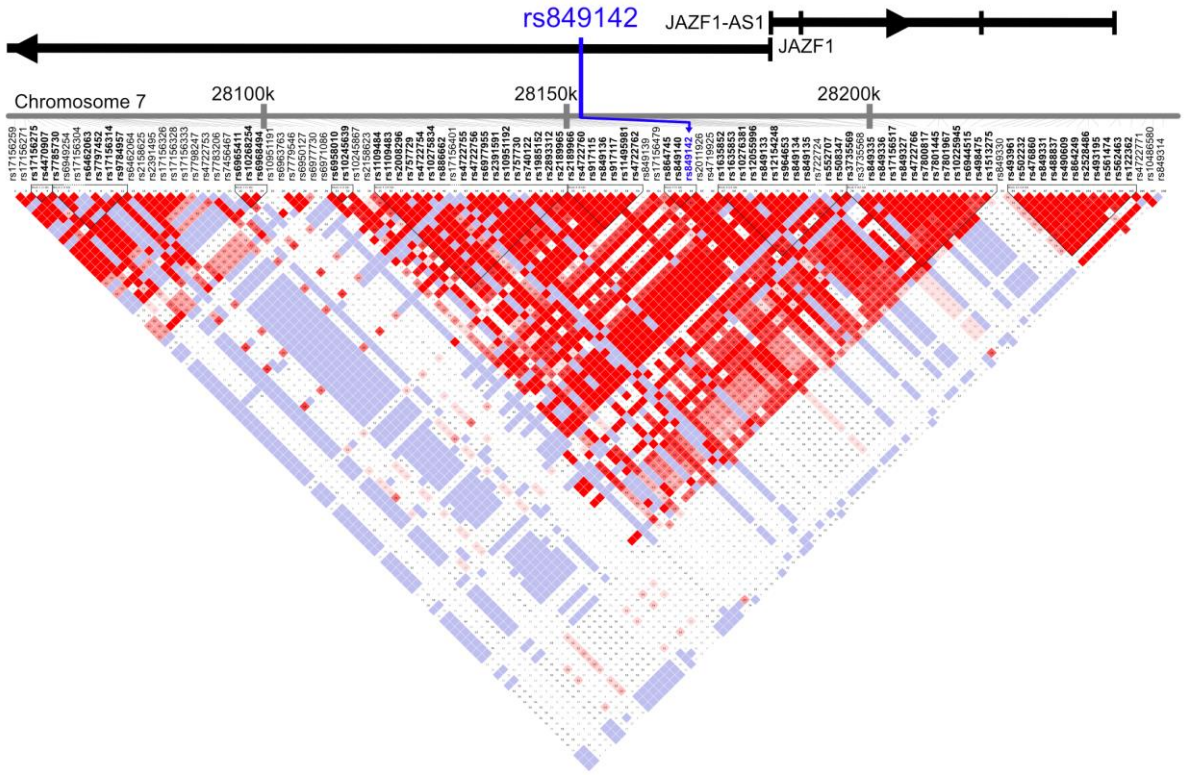

Fig. 12: Linkage Disequilibrium Plot of SNPs in JAZF1 gene using Version 2 release 24 CEU (CEPH, Utah residents with ancestry from northern and western Europe), chromosome 7 28152+/-100 kilobases: Variants are represented by grey lines on top. As an overview it shows that many SNPs form groups, which are in linkage equilibrium with each other, represented by red triangles under the respective variants. Coupling groups based on Gabriel Block definition are encircled by black triangles. rs849142 is indicated by a blue arrow (detailed view: Fig. 13, 14; full-size figure: Appendix A.5)

* Haplotype blocks are regions with little evidence of historical recombination. Gabriel et al. (2008) define them as 'as a region over which a very small proportion (<5%) of comparisons among informative SNP pairs show strong evidence of historical recombination'

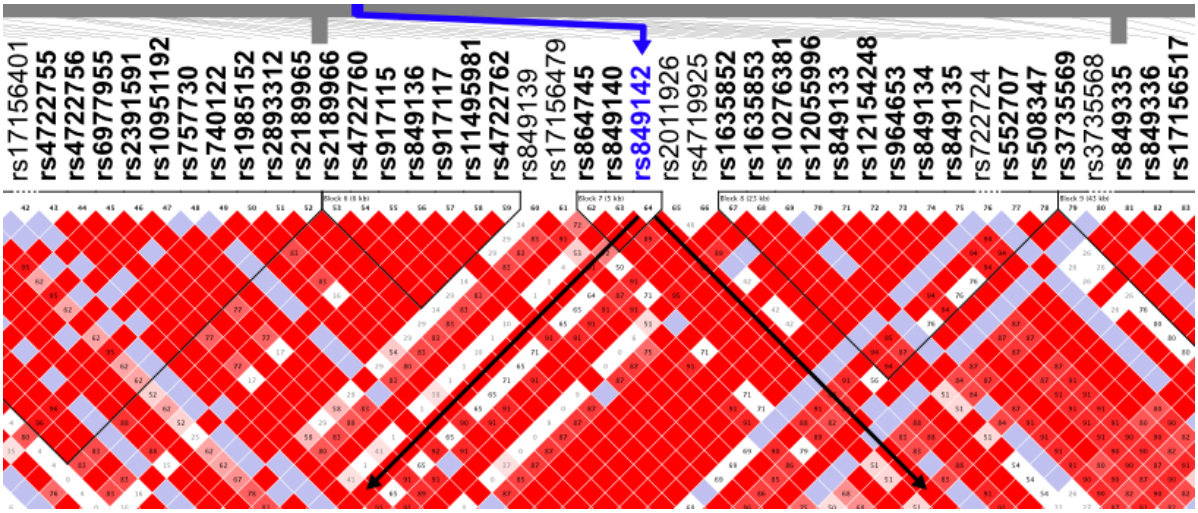

Fig. 13: rs849142 Haplogroup: Each rectangle represents LD between two SNPs; red indicates a strong LD. Rectangles indicating LD values for rs849142 are indicated by arrows: Many SNPs are in LD with rs849142. Using the block definition from Gabriel et al. [46], two other SNPs rs864745 as well as rs849140 form a coupling group with rs841942: (Fig. 14).

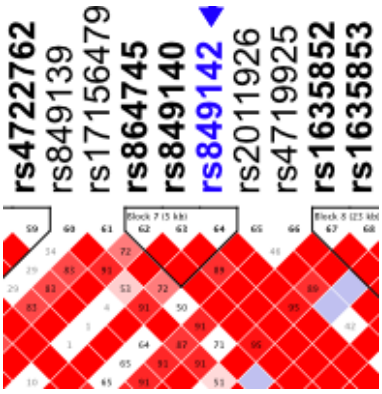

Fig. 14: Block definition from Gabriel et al [46]

4. Discussion

4.1 rs849142 might be a validated biomarker predicting ST in mCRC which may improve patients' treatment

Aim of this research project was to find a predictive marker for skin toxicity in targeted α-EGFR treatment of mCRC. Besides skin or receptor-specific mechanisms, immune dysregulation was proposed as a potential mechanism of ST.

This research project discovered the rs849142 SNP T allele as a validated biomarker for skin-related adverse effects in targeted α-EGFR treatment of mCRC. rs849142 is a common SNP in caucasian populations [39] which is known to be associated with SLE [8], lupus nephritis [9, 51], or prostate cancer [7]. rs849142 is located in the JAZF1 gene which functions as a repressor of transcriptional activation by TAK1 [5] and has an important role in lipid and glucose metabolism. [6, 50] Interestingly, the gene JAZF1 was shown to have a high mRNA expression in blood – a hint for its role in immune cells. However, expression within the skin was found to be low. This was taken as an indicator that the function of JAZF1 might not take over a mechanistic role in ST. But it has to be considered that no expression data from skin samples with ST is yet available. JAZF1 might be expressed in the skin of patients developing ST in the presence of α-EGFR directed antibodies like Cetuximab. Moreover, it might also be considered that JAZF1 is expressed only in a certain subset of cells like stem cells which are in a minority thus explaining overall low expression levels.

From a biological point of view, these findings suggest that immune cell regulation is an important mechanism of ST.

Irrespective of its biological function, rs849142 might be utilized among other markers in clinical practice as predictor of ST. In this context, the reliable prediction of ST using rs849142 would be a major advantage for both patients and physicians:

1) On the one hand, patients at risk of high-grade ST may be treated preventively for the adverse effect. This might increase quality of life as well as acceptance of treatment. In case a preventive treatment is not applicable to the patient, a different treatment like anti-VEGF antibodies might be chosen in order to prevent severe adverse events.
2) On the other hand, the use of preventive agents for ST might be limited to the cases where necessary: Patients in which the biomarker predicts a lower risk of ST could skip the treatment for ST. The agents used for prevention ofr ST, e.g. glucocorticoids, retinoids and antibiotics, might cause severe adverse events or might interfere with the

biological function in an unknown manner. Therefore, apart from economic reasons, their use should be limited to cases where necessary for medical reasons.

Yet, there are some obstacles for broad adoption of rs849142 as a predictive marker for ST:

1) Surely, the occurrence of adverse events is not solely dependent on genetics: Additional factors like performance status or age may also play a major role. A possible way to take them also into account would be a rating system based on genetic as well as clinical factors.
2) Additionally, for treatment of stage IV cancer patients, subjective and personal aspects should be considered. In this scenario, genetic tests might be of supportive value in the final decision of treatment.

To review the value of rs849142 in a clinical setting and to gain further information on its reliability, in a next step, it should be tested in a prospective clinical study.

Given that ST has been shown to be associated with survival of mCRC patients treated with α-EGFR antibodies, a predictive biomarker for ST might also predict response and survival of mCRC patients. As ST occurs after treatment start, rs849142 may be a predictive biomarker and allow better treatment planning. In contrast, ST itself is retrospective and is only a prognostic but not predictive marker.

The training set showed some association of rs849142 to survival (PFS p=0.005; OS p=0.072). The high-grade-ST-associated T allele is associated with longer survival, which is in line with theoretical assumptions. However, this association could not be confirmed in the validation set.

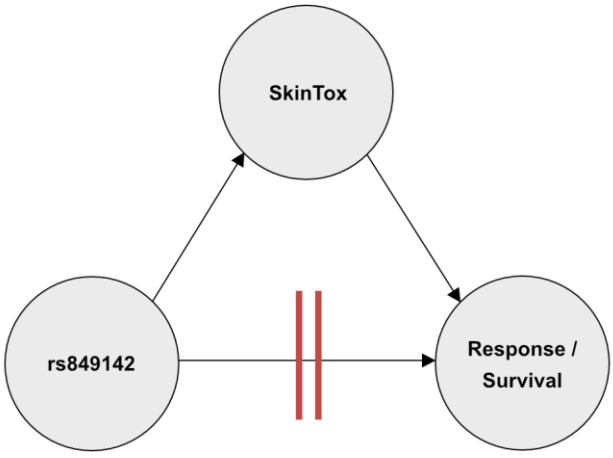

Fig. 15: Lack of association between rs849142 and survival may be due to insufficient statistical power

One explanation is lack of statistical power: On the one hand, rs849142 and ST did not show a perfect association. On the other hand, ST is only one of several potential predictors for survival (Fig. 15). Therefore, the association from a ST-specific marker to ST-related changes in survival may be influenced by several additional variables including patient-specific characteristics. Subsequently, these

confounders might make the association to survival difficult to unravel. Therefore, more patients may be necessary to show a significant effect on survival.

Alternatively, this lack of association to survival can have biological reasons: rs849142 might only reflect ST-specific aspects, as it has been selected with focus on ST, and not on response. To find predictive markers for response, other selection processes and therefore other SNPs may be necessary. It further points on the importance of molecular genetic analyses specific for adverse events.

4.2 Limitations of this work

As this work has focused on a predefined set of SNPs, it does not provide a comprehensive overview like genome wide association studies (GWAS) does. Yet, the theoretical foundation of SNP selection and the training-validation approach give a certain validity to the results. Statistical adjustments of multiple testing were not applied, because a validation set was included and only used to validate results from the training set. As a consequence, SNPs significantly associated to ST in the validation but not in the training set were not further investigated. Due to the limited number of tested markers, multi-marker analyses were difficult to carry out, too. Use of the DNA sequencing methods in the training as well as the validation set might yield some problems, but systematic errors for sequencing within high or low ST groups seem unlikely, because sequencing quality strongly depends on the DNA quality. Systematic differences of DNA quality between high and low ST groups seem unlikely.

Additionally, given that the analysis was carried out in a non-randomized, retrospective approach selection effects with regard to other variables than ST cannot be ruled out fully. For this reason, a prospective clinical trial using rs849142 as a predictor for ST would be the next step before application to clinical practice.

Yet, clinical application of a single genomic marker like rs849142 may have structural weaknesses: rs849142 is no exception concerning the fact that the predictive value of SNPs is often limited to probabilities [19]. Therefore, deterministic predictions for individual patients remain difficult. This limitation may lead to false positive as well as false negative predictions by the biomarker. In fact, this could harm patients. In order to estimate this risk, calculations on positive and negative predictive value in a prospective clinical trial would allow physicians better judgments on the value of rs849142 in this setting. An additional way to increase the predictive value would be the multi-marker approach mentioned above.

In addition, the approach to identify markers associated to both ST and response is problematic. Biological mechanisms leading to response to α-EGFR-antibodies are not necessarily the same as those mechanisms which cause ST. Targets for sequencing were selected on the basis of putative ST-related mechanisms, but without taking mechanisms for response and survival into account. This may be one reason why this work was able to find a SNP associated to ST but not to survival. Therefore, studies to find predictive markers for mCRC patients treated with α-EGFR-antibodies should be based on other sequencing panels than those focused on adverse events.

4.3 Further research on the genetic foundation of α-EGFR-associated ST should be have a broader focus than EGFR and downstream pathways

Research on molecular mechanisms in targeted α-EGFR treatment have been focused on EGFR and downstream targets like KRAS or NRAS, as these have a major importance for survival and response [52]. This has resulted in some success. SNPs associated with α-EGFR efficacy have been revealed in the EGFR [33]. The EGFR pathway [33] and EGFR-turnover [53] are also associated with α-EGFR efficacy. Additional targets involving the immune system like Fcγ-receptors [54] have also been shown to be associated with α-EGFR efficacy. Research on SNPs as predictors for adverse skin effects, has focused on the same targets. Few SNPs associated to both survival and ST have been unveiled yet [33].

Starting from theoretical concept, this work also tested an EGFR-independent approach: Immune cell regulation, immune receptors and skin-specific aspects were identified as important theoretical concepts explaining the occurrence of ST. Consequently, it was focused on upstream targets like Fc-receptors, skin and immune regulation related targets. Thus, this work provides a new view on potential markers for ST. The successful identification of rs849142 based on a relatively small number of target SNPs shows the high potential of this approach. Therefore, future research on the genetic foundation of ST should take immunological and skin-related mechanisms into account.

To further investigate the biological properties of rs849142 in the context of ST, ST-affected skin probes will be necessary. Yet, such specimens are not collected routinely and are very difficult to obtain. Analyses based on cell-models are also challenging, because a systemic condition like ST cannot be modeled in wet-lab conditions easily due to the complexity of the system.

4.4 Outlook

rs849142 is a validated predictive marker for ST in α-EGFR mCRC treatment. It shows the importance of genetic variations for practical clinical treatment. However, translation of research results into clinical practice remains a major obstacle. The broad availability of next generation sequencing technology might make genetic analyses more common and increase their utilization to estimate response and adverse events. Particularly broader application of whole genome sequencing approaches may allow identification of further markers by genome wide association analyses. A problem in association studies with SNPs is their partly weak association to phenotypes. Therefore, large patient collectives are necessary to reach significance. Consequently, recruiting a sufficient number of patients – especially for rare diseases or not often used treatments – may be only possible in concentrated clinical centers.
Given that costs per base sequenced are declining rapidly, further studies should include SNPs to gain a more comprehensive overview on the topic. However, this will require a large number of patients and careful adjustments for multiple testing. A genome-wide association study should also allow the selection of additional SNPs for a screening panel to improve reliability and predictive value for individual patients under treatment α–EGFR for mCRC.

The results of this work were partly published in a peer-reviewed journal by the author after first submission of this thesis [55].

5. References

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2015.* CA Cancer J Clin, 2015. 65(1): p. 5-29.
2. Van Cutsem, E., et al., *Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Annals of Oncology, 2014. 25(suppl 3): p. iii1-iii9.
3. Chen, A.P., et al., *Grading dermatologic adverse events of cancer treatments: the Common Terminology Criteria for Adverse Events Version 4.0.* J Am Acad Dermatol, 2012. 67(5): p. 1025-39.
4. Stintzing, S., et al., *Prognostic value of cetuximab-related skin toxicity in metastatic colorectal cancer patients and its correlation with parameters of the epidermal growth factor receptor signal transduction pathway: results from a randomized trial of the GERMAN AIO CRC Study Group.* Int J Cancer, 2013. 132(1): p. 236-45.
5. Nakajima, T., et al., *TIP27: a novel repressor of the nuclear orphan receptor TAK1/TR4.* Nucleic Acids Research, 2004. 32(14): p. 4194-4204.
6. Yuan, L., et al., *Transcription factor TIP27 regulates glucose homeostasis and insulin sensitivity in a PI3-kinase/Akt-dependent manner in mice.* Int J Obes (Lond), 2015. 39(6): p. 949-58.
7. Prokunina-Olsson, L., et al., *Refining the prostate cancer genetic association within the JAZF1 gene on chromosome 7p15.2.* Cancer Epidemiol Biomarkers Prev, 2010. 19(5): p. 1349-55.
8. Cunninghame Graham, D.S., et al., *Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus.* PLoS Genet, 2011. 7(10): p. e1002341.
9. de Zubiria Salgado, A. and C. Herrera-Diaz, *Lupus nephritis: an overview of recent findings.* Autoimmune Dis, 2012. 2012: p. 849684.
10. Joshi, S.S., et al., *Effects of epidermal growth factor receptor inhibitor-induced dermatologic toxicities on quality of life.* Cancer, 2010. 116(16): p. 3916-23.
11. Ferlay, J., et al., *Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012.* Eur J Cancer, 2013. 49(6): p. 1374-403.
12. Sermeus, A., et al., *Advances in radiotherapy delivery for rectal cancer: a European perspective.* Expert Rev Gastroenterol Hepatol, 2015. 9(4): p. 393-7.
13. Heinemann, V., et al., *FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab as first-line treatment for patients with metastatic colorectal cancer (FIRE-3): a randomised, open-label, phase 3 trial.* Lancet Oncol, 2014. 15(10): p. 1065-75.
14. Douillard, J.Y., et al., *Final results from PRIME: randomized phase III study of panitumumab with FOLFOX4 for first-line treatment of metastatic colorectal cancer.* Ann Oncol, 2014. 25(7): p. 1346-55.
15. Pinto, C., et al., *Management of skin toxicity associated with cetuximab treatment in combination with chemotherapy or radiotherapy.* Oncologist, 2011. 16(2): p. 228-38.
16. Eilers, R.E., Jr., et al., *Dermatologic infections in cancer patients treated with epidermal growth factor receptor inhibitor therapy.* J Natl Cancer Inst, 2010. 102(1): p. 47-53.
17. Balagula, Y., S.T. Rosen, and M.E. Lacouture, *The emergence of supportive oncodermatology: the study of dermatologic adverse events to cancer therapies.* J Am Acad Dermatol, 2011. 65(3): p. 624-35.
18. Kaplan, E.L. and P. Meier, *Nonparametric estimation from incomplete observations.* Journal of the American statistical association, 1958. 53(282): p. 457-481.
19. Lewis, C.M. and J. Knight, *Introduction to genetic association studies.* Cold Spring Harb Protoc, 2012. 2012(3): p. 297-306.

20. Purcell, S., S.S. Cherny, and P.C. Sham, *Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits.* Bioinformatics, 2003. 19(1): p. 149-50.
21. Lacouture, M.E., *Mechanisms of cutaneous toxicities to EGFR inhibitors.* Nat Rev Cancer, 2006. 6(10): p. 803-12.
22. Nanney, L.B., et al., *Immunolocalization of epidermal growth factor receptors in normal developing human skin.* J Invest Dermatol, 1990. 94(6): p. 742-8.
23. Rodeck, U., et al., *EGF-R dependent regulation of keratinocyte survival.* J Cell Sci, 1997. 110 (Pt 2): p. 113-21.
24. Taylor, R.J., et al., *Ex vivo antibody-dependent cellular cytotoxicity inducibility predicts efficacy of cetuximab.* Cancer Immunol Res, 2015. 3(5): p. 567-74.
25. Eichenfield, L.F., et al., *Evolving perspectives on the etiology and pathogenesis of acne vulgaris.* J Drugs Dermatol, 2015. 14(3): p. 263-72.
26. Szabo, K. and L. Kemeny, *Studying the genetic predisposing factors in the pathogenesis of acne vulgaris.* Hum Immunol, 2011. 72(9): p. 766-73.
27. Zhang, M., et al., *A genome-wide association study of severe teenage acne in European Americans.* Human genetics, 2014. 133(3): p. 259-264.
28. Guerra, S.G., T.J. Vyse, and D.S. Cunninghame Graham, *The genetics of lupus: a functional perspective.* Arthritis Res Ther, 2012. 14(3): p. 211.
29. Pan, F., et al., *Genetic susceptibility and haplotype analysis between Fcgamma receptor IIB and IIIA gene with systemic lupus erythematosus in Chinese population.* Lupus, 2008. 17(8): p. 733-8.
30. Parodi, A. and E. Cozzani, *Cutaneous manifestations of lupus erythematosus.* G Ital Dermatol Venereol, 2014. 149(5): p. 549-54.
31. Gottschalk, T.A., E. Tsantikos, and M.L. Hibbs, *Pathogenic Inflammation and Its Therapeutic Targeting in Systemic Lupus Erythematosus.* Front Immunol, 2015. 6: p. 550.
32. Li, X., et al., *Fcgamma receptors: structure, function and role as genetic risk factors in SLE.* Genes Immun, 2009. 10(5): p. 380-9.
33. Hasheminasab, S.M., et al., *High-throughput screening identified inherited genetic variations in the EGFR pathway contributing to skin toxicity of EGFR inhibitors.* Pharmacogenomics, 2015. 16(14): p. 1605-19.
34. Stintzing, S., et al., *Polymorphisms in Genes Involved in EGFR Turnover Are Predictive for Cetuximab Efficacy in Colorectal Cancer.* Molecular Cancer Therapeutics, 2015. 14(10): p. 2374-2381.
35. *Cetuximab Patent.* European Patent Office.
36. Bruhns, P., et al., *Specificity and affinity of human Fcgamma receptors and their polymorphic variants for human IgG subclasses.* Blood, 2009. 113(16): p. 3716-25.
37. Dong, C., et al., *Fcgamma receptor IIIa single-nucleotide polymorphisms and haplotypes affect human IgG binding and are associated with lupus nephritis in African Americans.* Arthritis Rheumatol, 2014. 66(5): p. 1291-9.
38. Life-Technologies. *Ampliseq Custom panel Designer*. Available from: https://www.ampliseq.com/.
39. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res, 2001. 29(1): p. 308-11.
40. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. 81(3): p. 559-75.
41. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. 21(2): p. 263-5.
42. Hardy, G.H., *Mendelian Proportions in a Mixed Population.* Science, 1908. 28(706): p. 49-50.

43. Wigginton, J.E., D.J. Cutler, and G.R. Abecasis, *A note on exact tests of Hardy-Weinberg equilibrium.* Am J Hum Genet, 2005. 76(5): p. 887-93.
44. Slatkin, M., *Disequilibrium mapping of a quantitative-trait locus in an expanding population.* Am J Hum Genet, 1999. 64(6): p. 1764-72.
45. Wald, A., *Tests of statistical hypotheses concerning several parameters when the number of observations is large.* Transactions of the American Mathematical society, 1943. 54(3): p. 426-482.
46. Gabriel, S.B., et al., *The structure of haplotype blocks in the human genome.* Science, 2002. 296(5576): p. 2225-2229.
47. Mantel, N., *Evaluation of survival data and two new rank order statistics arising in its consideration.* Cancer Chemother Rep, 1966. 50(3): p. 163-70.
48. Armstrong, R.A., *When to use the Bonferroni correction.* Ophthalmic Physiol Opt, 2014. 34(5): p. 502-8.
49. Balding, D.J., *A tutorial on statistical methods for population association studies.* Nat Rev Genet, 2006. 7(10): p. 781-91.
50. Ming, G.F., et al., *JAZF1 can regulate the expression of lipid metabolic genes and inhibit lipid accumulation in adipocytes.* Biochem Biophys Res Commun, 2014. 445(3): p. 673-80.
51. Gateva, V., et al., *A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus.* Nature genetics, 2009. 41(11): p. 1228-1233.
52. De Roock, W., et al., *KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer.* Lancet Oncol, 2011. 12(6): p. 594-603.
53. Stintzing, S., et al., *Polymorphisms in Genes Involved in EGFR Turnover Are Predictive for Cetuximab Efficacy in Colorectal Cancer.* Molecular cancer therapeutics, 2015. 14(10): p. 2374-2381.
54. Lopez-Albaitero, A., et al., *Role of polymorphic Fc gamma receptor IIIa and EGFR expression level in cetuximab mediated, NK cell dependent in vitro cytotoxicity of head and neck squamous cell carcinoma cells.* Cancer Immunol Immunother, 2009. 58(11): p. 1853-64.
55. Froelich M.F. et al., *The DNA-polymorphism rs849142 is associated with skin toxicity induced by targeted anti-EGFR therapy using cetuximab.* Oncotarget, 2018. 9(54): p. 30279-88.

6. Acknowledgements

At this point, I would like to acknowledge the great support, which made this work possible.

First, I would like to express my gratefulness to Prof. Dr. Thomas Kirchner and the Institute of Pathology of LMU Munich, which offered an extraordinary positive environment with both warm and helpful employees and a highly sophisticated technical equipment to carry out the analyses necessary for this work. Also, I am very grateful for the opportunity to visit congresses related to the topic of this thesis.

Furthermore, I would like to express my deepest thanks to Prof. Dr. Andreas Jung. With his great support regarding the strategy to carry out all milestones necessary for this thesis, accompanied by his deep theoretical understanding and practical experience, complemented with his availability to all my questions and his great personal support, he made this work possible. Thank you very much.

I would also like acknowledge the work of the members of the Jung Lab very much, especially Gertrud Lenz who did the wet-lab work with a very high quality and generated thereby the raw-data which were used as the basis of this work. The high quality of the raw data made the analyses presented in this thesis possible.

The assessment of clinic-associated questions like the one in this thesis, is often limited by the availability of clinical data and appropriate patient collectives. To my great contentment, Prof. Dr. Sebastian Stintzing and Prof. Dr. Volker Heinemann from Medical Clinic III of LMU Munich gave me the opportunity using the probes and their data-set of the FIRE-3 study. This high-quality data was an exceptional foundation to asses my study question.

Besides many members of the Institute of Pathology of LMU Munich who remain unmentioned, I would like to highlight Dr. Thomas Grünewald and Dr. Jörg Kumbrink. Concerning bioinformatics, Dr. Grünewald gave me decisive advice on statistical methods and corresponding software for my analyses. Discussions with Dr. Kumbrink helped me a lot in preparation of results for congresses and other occasions.

Thank you very much to all of you.

7. Eidesstattliche Versicherung

Ich, Matthias Frölich, erkläre hiermit an Eides statt,
dass ich die vorliegende Dissertation mit dem Thema

The polymorphic DNA marker rs849142 predicts skin toxicity in anti-EGFR treatment of metastatic colorectal cancer

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München 11.10.2018,

Matthias F. Frölich

A. Appendix

A.1 Sequencing targets

panel	type	rs-code	Chr	start	end	gene
acne	SNP	rs5743708	4	154626316	154626317	TLR2
acne	SNP	rs4986790	9	120475301	120475302	TLR4
acne	SNP	rs4986791	9	120475601	120475602	TLR4
acne	SNP	rs361525	6	31543100	31543101	TNFA
acne	SNP	rs1800629	6	31543030	31543031	TNFA
acne	SNP	rs1799724	6	31542481	31542482	TNFA
acne	SNP	rs1800630	6	31542475	31542476	TNFA
acne	SNP	rs1799964	6	31542307	31542308	TNFA
acne	SNP	rs17561	2	113537222	113537223	IL-1A
acne	SNP	rs1048943	15	75012984	75012985	CYP1A1
acne	SNP	rs4646903	15	75011640	75011641	CYP1A1
acne	SNP	rs743572	10	104597151	104597152	CYP17
acne	SNP	rs4133274	8	128676130	128676131	Chr 8q24
acne	SNP	rs13248513	8	128691211	128691212	Chr 8q24
SLE	SNP	rs9888739	16	31313252	31313253	ITGAM
SLE	SNP	rs548234	6	106568033	106568034	-
SLE	SNP	rs2230926	6	138196065	138196066	TNFAIP3
SLE	SNP	rs7708392	5	150457484	150457485	TNIP1
SLE	SNP	rs463426	22	21809184	21809185	-
SLE	SNP	rs6590330	11	128311058	128311059	-
SLE	SNP	rs4917014	7	50305862	50305863	-
SLE	SNP	rs507230	11	35129171	35129172	-
SLE	SNP	rs10516487	4	102751075	102751076	BANK1
SLE	SNP	rs7812879	8	11340180	11340181	-
SLE	SNP	rs7829816	8	56849385	56849386	LYN
SLE	SNP	rs13385731	2	33701889	33701890	RASGRP3
SLE	SNP	rs10911363	1	183549756	183549757	NCF2
SLE	SNP	rs7574865	2	191964632	191964633	STAT4
SLE	SNP	rs2476601	1	114377567	114377568	AP4B1-AS1, PTPN22
SLE	SNP	rs2205960	1	173191474	173191475	-
SLE	SNP	rs3135394	6	32408496	32408497	-
SLE	SNP	rs10847697	12	129299384	129299385	SLC15A4
SLE	SNP	rs2070197	7	128588999	128589000	IRF5
SLE	SNP	rs4963128	11	589563	589564	PHRF1
SLE	SNP	rs2280381	16	86018632	86018633	-
SLE	SNP	rs1990760	2	163124050	163124051	IFIH1
SLE	SNP	rs280519	19	10472932	10472933	TYK2
SLE	SNP	rs1913517	10	50119053	50119054	LRRC18, WDFY4
SLE	SNP	rs6445975	3	58370176	58370177	PXK
SLE	SNP	rs849142	7	28185890	28185891	JAZF1
SLE	SNP	rs11755393	6	34824635	34824636	UHRF1BP1
SLE	SNP	rs6985109	8	10761584	10761585	XKR6
SLE	SNP	rs396716	1	161514595	161514596	FCGR3A
SLE	SNP	rs396991	1	161514541	161514542	FCGR3A
SLE	SNP	rs445509	1	161512347	161512348	FCGR3A
SLE	SNP	rs1042206	1	161512958	161512959	FCGR3A
SLE	SNP	rs148181339	1	161518213	161518214	FCGR3A
SLE	SNP	rs1050501	1	161643797	161643798	FCGR2B

SLE	SNP	rs1051009	17	4637885	4637886	CXCL16
SLE	SNP	rs10752900	1	183086237	183086238	LAMC1
FcRγ	SNP	rs10127939	1	161518332	161518333	FCGR3A
FcRγ	SNP	rs145557772	1	161518233	161518234	FCGR3A
FcRγ	SNP	rs396991	1	161514541	161514542	FCGR3A
FcRγ	SNP	rs396716	1	161514595	161514596	FCGR3A
FcRγ	SNP	rs443082	1	161514627	161514628	FCGR3A
FcRγ	SNP	rs445509	1	161512347	161512348	FCGR3A
FcRγ	SNP	rs1042206	1	161512958	161512959	FCGR3A
FcRγ	SNP	rs148181339	1	161518213	161518214	FCGR3A
FcRγ	SNP	rs1801274	1	161479744	161479745	FCGR2A
FcRγ	REGION	rs428888	1	161548496	161548497	FCGR3A
FcRγ	REGION	rs403016	1	161548632	161548633	FCGR3A
FcRγ	SNP	rs1050501	1	161643797	161643798	FCGR2B
FcRγ	SNP	rs6427598	1	161484563	161484564	FCGR2A
FcRγ	SNP	rs368433	1	161484209	161484210	FCGR2A
FcRγ	REGION	rs3219018	1	161581052	161581058	FCGR2B
FcRγ	SNP	rs844	1	161647532	161647533	FCGR2B
FcRγ	SNP	rs1674761	1	161645470	161645471	FCGR2B
FcRγ	SNP	rs5017567	1	161641295	161641296	FCGR2B
FcRγ	SNP	rs12118043	1	161646823	161646824	FCGR2B
FcRγ	SNP	rs5030738	1	161599653	161599654	FCGR3B

A.2 Hotspot files

Acne

Track name="CHP2_HotSpots" description="CHP2_COSMIC_Mutations_v60" type=bedDetail

Chr	Start	End	SNP	REF	Panel
chr4	154626316	154626317	rs5743708	REF=G;OBS=A;ANCHOR=C	acne
chr9	120475301	120475302	rs4986790	REF=A;OBS=G;ANCHOR=G	acne
chr9	120475601	120475602	rs4986791	REF=C;OBS=T;ANCHOR=T	acne
chr6	31543100	31543101	rs361525	REF=G;OBS=A;ANCHOR=T	acne
chr6	31543030	31543031	rs1800629	REF=G;OBS=A;ANCHOR=T	acne
chr6	31542481	31542482	rs1799724	REF=C;OBS=T;ANCHOR=T	acne
chr6	31542475	31542476	rs1800630	REF=C;OBS=A;ANCHOR=C	acne
chr6	31542307	31542308	rs1799964	REF=T;OBS=C;ANCHOR=T	acne
chr2	113537222	113537223	rs17561	REF=C;OBS=A;ANCHOR=T	acne
chr15	75012984	75012985	rs1048943	REF=T;OBS=C;ANCHOR=T	acne
chr15	75011640	75011641	rs4646903	REF=A;OBS=G;ANCHOR=C	acne
chr10	104597151	104597152	rs743572	REF=A;OBS=G;ANCHOR=T	acne
chr8	128676130	128676131	rs4133274	REF=A;OBS=G;ANCHOR=T	acne
chr8	128691211	128691212	rs13248513	REF=T;OBS=C;ANCHOR=T	acne

SLE

Track name="CHP2_HotSpots" description="CHP2_COSMIC_Mutations_v60" type=bedDetail

Chr	Start	End	SNP	REF	Panel
chr16	31313252	31313253	rs9888739	REF=C;OBS=T;ANCHOR=T	sle
chr6	106568033	106568034	rs548234	REF=C;OBS=T;ANCHOR=T	sle
chr6	138196065	138196066	rs2230926	REF=T;OBS=G;ANCHOR=T	sle
chr5	150457484	150457485	rs7708392	REF=G;OBS=C;ANCHOR=T	sle
chr22	21809184	21809185	rs463426	REF=T;OBS=C;ANCHOR=T	sle
chr11	128311058	128311059	rs6590330	REF=G;OBS=A;ANCHOR=T	sle
chr7	50305862	50305863	rs4917014	REF=T;OBS=G;ANCHOR=T	sle
chr11	35129171	35129172	rs507230	REF=G;OBS=A;ANCHOR=T	sle
chr4	102751075	102751076	rs10516487	REF=G;OBS=A;ANCHOR=T	sle
chr8	11340180	11340181	rs7812879	REF=T;OBS=C;ANCHOR=T	sle
chr8	56849385	56849386	rs7829816	REF=A;OBS=G;ANCHOR=T	sle
chr2	33701889	33701890	rs13385731	REF=T;OBS=C;ANCHOR=T	sle
chr1	183549756	183549757	rs10911363	REF=G;OBS=T;ANCHOR=T	sle
chr2	191964632	191964633	rs7574865	REF=T;OBS=G;ANCHOR=T	sle
chr1	114377567	114377568	rs2476601	REF=A;OBS=G;ANCHOR=T	sle
chr1	173191474	173191475	rs2205960	REF=G;OBS=T;ANCHOR=T	sle
chr6	32408496	32408497	rs3135394	REF=A;OBS=G;ANCHOR=T	sle
chr12	129299384	129299385	rs10847697	REF=G;OBS=A;ANCHOR=T	sle
chr7	128588999	128589000	rs2070197	REF=T;OBS=C;ANCHOR=T	sle
chr11	589563	589564	rs4963128	REF=T;OBS=C;ANCHOR=T	sle
chr16	86018632	86018633	rs2280381	REF=C;OBS=T;ANCHOR=T	sle
chr2	163124050	163124051	rs1990760	REF=C;OBS=T;ANCHOR=T	sle
chr19	10472932	10472933	rs280519	REF=A;OBS=G;ANCHOR=T	sle
chr10	50119053	50119054	rs1913517	REF=A;OBS=G;ANCHOR=T	sle
chr3	58370176	58370177	rs6445975	REF=G;OBS=T;ANCHOR=T	sle
chr7	28185890	28185891	rs849142	REF=T;OBS=C;ANCHOR=T	sle
chr6	34824635	34824636	rs11755393	REF=A;OBS=G;ANCHOR=T	sle
chr8	10761584	10761585	rs6985109	REF=G;OBS=A;ANCHOR=T	sle
chr1	161514595	161514596	rs396716	REF=A;OBS=G;ANCHOR=T	sle
chr1	161514541	161514542	rs396991	REF=A;OBS=C;ANCHOR=T	sle
chr1	161512347	161512348	rs445509	REF=C;OBS=T;ANCHOR=C	sle
chr1	161512958	161512959	rs1042206	REF=A;OBS=C;ANCHOR=A	sle
chr1	161518213	161518214	rs148181339	REF=T;OBS=C;ANCHOR=A	sle
chr1	161643797	161643798	rs1050501	REF=T;OBS=C;ANCHOR=A	sle
chr17	4637885	4637886	rs1051009	REF=G;OBS=A;ANCHOR=T	sle
chr1	183086237	183086238	rs10752900	REF=G;OBS=A;ANCHOR=T	sle

FcRγ

Track name="CHP2_HotSpots" description="CHP2_COSMIC_Mutations_v60" type=bedDetail

Chr	Start	End	SNP	REF	Panel
chr1	161518332	161518333	rs10127939	REF=A;OBS=C;ANCHOR=G	FCGR
chr1	161518233	161518234	rs145557772	REF=G;OBS=A;ANCHOR=C	FCGR
chr1	161514541	161514542	rs396991	REF=A;OBS=C;ANCHOR=A	FCGR
chr1	161514595	161514596	rs396716	REF=A;OBS=G;ANCHOR=T	FCGR
chr1	161514627	161514628	rs443082	REF=C;OBS=T;ANCHOR=G	FCGR
chr1	161512347	161512348	rs445509	REF=C;OBS=T;ANCHOR=C	FCGR

chr1	161512958	161512959	rs1042206	REF=A;OBS=C;ANCHOR=A	FCGR
chr1	161518213	161518214	rs148181339	REF=T;OBS=C;ANCHOR=A	FCGR
chr1	161479744	161479745	rs1801274	REF=A;OBS=G;ANCHOR=C	FCGR
chr1	161548496	161548497	rs428888	REF=T;OBS=C;ANCHOR=T	FCGR
chr1	161548632	161548633	rs403016	REF=C;OBS=G;ANCHOR=C	FCGR
chr1	161643797	161643798	rs1050501	REF=T;OBS=C;ANCHOR=A	FCGR
chr1	161484563	161484564	rs6427598	REF=T;OBS=A;ANCHOR=G	FCGR
chr1	161484209	161484210	rs368433	REF=T;OBS=C;ANCHOR=A	FCGR
chr1	161581052	161581058	rs3219018	REF=G;OBS=C;ANCHOR=T	FCGR
chr1	161647532	161647533	rs844	REF=A;OBS=G;ANCHOR=T	FCGR
chr1	161645470	161645471	rs1674761	REF=A;OBS=C;ANCHOR=T	FCGR
chr1	161641295	161641296	rs5017567	REF=A;OBS=C;ANCHOR=C	FCGR
chr1	161646823	161646824	rs12118043	REF=C;OBS=A;ANCHOR=C	FCGR
chr1	161599653	161599654	rs5030738	REF=G;OBS=T;ANCHOR=A	FCGR

A.3 Sequencing summary statistics

Sequenced genotypes per SNP

Chr	SNP	n missing genotype	n patients	% missing genotype
1	rs2476601	11	71	15.49%
1	rs1801274	3	71	4.23%
1	rs368433	5	71	7.04%
1	rs6427598	3	71	4.23%
1	rs445509	2	71	2.82%
1	rs1042206	1	71	1.41%
1	rs396991	3	71	4.23%
1	rs396716	1	71	1.41%
1	rs443082	3	71	4.23%
1	rs148181339	1	71	1.41%
1	rs145557772	3	71	4.23%
1	rs10127939	71	71	100.00%
1	rs5030738	15	71	21.13%
1	rs5017567	3	71	4.23%
1	rs1050501	71	71	100.00%
1	rs1674761	29	71	40.85%
1	rs12118043	5	71	7.04%
1	rs844	3	71	4.23%
1	rs2205960	11	71	15.49%
1	rs10752900	8	71	11.27%
1	rs10911363	45	71	63.38%
2	rs13385731	36	71	50.70%
2	rs17561	32	71	45.07%
2	rs1990760	15	71	21.13%
2	rs7574865	10	71	14.08%
3	rs6445975	23	71	32.39%
4	rs10516487	7	71	9.86%
4	rs5743708	34	71	47.89%
5	rs7708392	38	71	53.52%
6	rs1799964	42	71	59.15%
6	rs1800630/rs4645836	10	71	14.08%
6	rs1799724	10	71	14.08%
6	rs1800629	9	71	12.68%

6	rs361525	9	71	12.68%
6	rs3135394	43	71	60.56%
6	rs11755393	14	71	19.72%
6	rs548234	32	71	45.07%
6	rs2230926	8	71	11.27%
7	rs849142	9	71	12.68%
7	rs4917014	8	71	11.27%
7	rs2070197	8	71	11.27%
8	rs6985109	21	71	29.58%
8	rs7812879	8	71	11.27%
8	rs7829816	12	71	16.90%
8	rs4133274	10	71	14.08%
8	rs13248513	11	71	15.49%
9	rs4986790	19	71	26.76%
9	rs4986791	21	71	29.58%
10	rs1913517	27	71	38.03%
10	rs743572	24	71	33.80%
11	rs4963128	16	71	22.54%
11	rs507230	27	71	38.03%
11	rs6590330	58	71	81.69%
12	rs10847697	24	71	33.80%
15	rs79812015/rs4646903	26	71	36.62%
15	rs1048943	21	71	29.58%
16	rs9888739	5	71	7.04%
16	rs2280381	70	71	98.59%
17	rs1051009	31	71	43.66%
19	rs280519	15	71	21.13%
22	rs463426	7	71	9.86%

Sequenced genotypes per patient

ID	n missing genotypes	n sequenced genotypes	% missing genotypes
20	61	61	100.00%
90	14	61	22.95%
213	17	61	27.87%
281	27	61	44.26%
344	14	61	22.95%
375	27	61	44.26%
406	46	61	75.41%
428	30	61	49.18%
566	28	61	45.90%
586	41	61	67.21%
598	12	61	19.67%
624	40	61	65.57%
638	23	61	37.70%
708	34	61	55.74%
750	33	61	54.10%
796	20	61	32.79%
10	23	61	37.70%
75	19	61	31.15%
102	23	61	37.70%
200	19	61	31.15%
234	8	61	13.11%

236	15	61	24.59%
245	27	61	44.26%
288	16	61	26.23%
302	21	61	34.43%
336	23	61	37.70%
362	16	61	26.23%
381	16	61	26.23%
398	8	61	13.11%
416	21	61	34.43%
421	14	61	22.95%
440	22	61	36.07%
469	6	61	9.84%
477	8	61	13.11%
483	5	61	8.20%
497	4	61	6.56%
502	5	61	8.20%
512	5	61	8.20%
516	3	61	4.92%
552	5	61	8.20%
559	12	61	19.67%
562	5	61	8.20%
576	4	61	6.56%
583	5	61	8.20%
588	19	61	31.15%
592	4	61	6.56%
595	6	61	9.84%
605	6	61	9.84%
606	10	61	16.39%
609	6	61	9.84%
610	9	61	14.75%
613	10	61	16.39%
635	10	61	16.39%
663	9	61	14.75%
665	10	61	16.39%
681	9	61	14.75%
696	18	61	29.51%
706	16	61	26.23%
709	22	61	36.07%
713	25	61	40.98%
720	25	61	40.98%
733	25	61	40.98%
736	23	61	37.70%
748	13	61	21.31%
788	4	61	6.56%
804	5	61	8.20%
831	5	61	8.20%
840	20	61	32.79%
841	3	61	4.92%
843	5	61	8.20%
889	5	61	8.20%

A.4 Association analysis results

Chr	SNP	p Hardy-Weinberg	p all patients	p training set	p validation set
1	rs2476601	0.3953	0.3346	NA	0.3578
1	rs1801274	0.2965	0.1337	0.4146	0.183
1	rs368433	1.0000	NA	NA	NA
1	rs6427598	0.3298	0.8631	1.0000	0.8075
1	rs445509	1.0000	NA	NA	NA
1	rs1042206	1.0000	NA	NA	NA
1	rs396991	0.1769	0.6883	0.1638	0.6517
1	rs396716	1.0000	NA	NA	NA
1	rs443082	1.0000	NA	NA	NA
1	rs148181339	1.0000	NA	NA	NA
1	rs145557772	1.0000	NA	NA	NA
1	rs10127939	1.0000	NA	NA	NA
1	rs5030738	1.0000	0.7656	NA	0.5762
1	rs5017567	1.0000	NA	NA	NA
1	rs1050501	1.0000	NA	NA	NA
1	rs1674761	1,51E-08	0.8936	NA	0.8445
1	rs12118043	1.0000	0.4318	0.1877	0.7451
1	rs844	1.0000	0.9228	0.3728	0.4894
1	rs2205960	1.0000	0.7963	0.4366	0.8999
1	rs10752900	0.8005	0.4304	0.2924	0.6770
1	rs10911363	0.1303	0.5341	NA	0.6244
2	rs13385731	1.0000	0.7914	NA	0.4698
2	rs17561	0.6911	1.0000	NA	0.3211
2	rs1990760	0.5525	0.8752	0.8801	0.9198
2	rs7574865	0.5010	0.2592	0.2437	0.4797
3	rs6445975	1.0000	0.2125	0.4397	0.4290
4	rs10516487	0.5785	0.2985	0.7245	0.3371
4	rs5743708	1.0000	0.3493	0.2722	0.9700
5	rs7708392	1.0000	0.5232	NA	0.8659
6	rs1799964	1.0000	0.4411	NA	0.9601
6	rs1800630/rs4645836	0.5806	1.0000	0.2199	0.5821
6	rs1799724	1.0000	0.6686	0.2666	0.6985
6	rs1800629	1.0000	0.2511	0.8754	0.2504
6	rs361525	1.0000	0.6478	NA	0.7264
6	rs3135394	1.0000	0.6445	NA	0.6516
6	rs11755393	0.7853	0.1068	0.1996	0.3059
6	rs548234	0.3124	0.166	0.7698	0.1256
6	rs2230926	1.0000	0.3891	NA	0.3691
7	**rs849142**	0.4356	**0.003621**	**0.00395**	**0.04362**
7	rs4917014	0.5478	0.835	1.0000	0.8655
7	rs2070197	1.0000	0.2922	0.8154	0.3069
8	rs6985109	0.5629	1.0000	0.6667	0.7741
8	rs7812879	0.5965	0.4808	0.2413	0.8298
8	rs7829816	1.0000	0.1979	0.4071	0.3193
8	rs4133274	1.0000	1.0000	0.2199	0.5762
8	rs13248513	1.0000	0.8089	0.2199	0.4076
9	rs4986790	1.0000	0.1637	NA	0.1635
9	rs4986791	1.0000	0.7494	0.4816	0.3997
10	rs1913517	0.7621	1.0000	0.4366	0.7776

10	rs743572	1.0000	0.1561	0.9001	0.1262
11	rs4963128	0.0413	0.3063	0.2509	0.5945
11	rs507230	1.0000	0.1334	1.0000	0.1097
11	rs6590330	0.2092	0.2706	NA	0.8028
12	rs10847697	0.09936	0.5425	0.495	0.9548
15	rs79812015/rs4646903	1.0000	0.3138	NA	0.3746
15	rs1048943	1.0000	0.3765	NA	0.4071
16	rs9888739	0.6566	0.8668	0.2345	0.3694
16	rs2280381	1.0000	NA	NA	NA
17	rs1051009	0.3535	0.465	0.4778	0.6102
19	rs280519	0.05934	0.6209	0.4636	0.9892
22	**rs463426**	1.0000	**0.009854**	0.0663	**0.03559**

A.5 Linkage Disequilibrium Plot of SNPs in JAZF1 gene