

Medizinische Fakultät
der Universität Duisburg-Essen

Aus dem Institut für Medizinische Informatik, Biometrie und
Epidemiologie

Genetische Effekte auf die Erhöhung von Freien Leichtketten im Serum

Inaugural-Dissertation
zur Erlangung des Doktorgrades der
Naturwissenschaften in der Medizin
durch die Medizinische Fakultät
der Universität Duisburg-Essen

Vorgelegt von
Marina Arendt, M.Sc.
aus Gladbeck

2018

Dekan: Herr Univ.-Prof. Dr. med. J. Buer
1. Gutachter: Herr Priv.-Doz. Dr. med. L. Eisele
2. Gutachter: Herr Prof. Dr. med. J. Dürig
3. Gutachter: Herr Prof. Dr. med. U. Germing, Düsseldorf

Tag der mündlichen Prüfung: 4. Mai 2018

Beiträge

Marina Arendt, Börge Schmidt, Karl-Heinz Jöckel und Lewin Eisele

Genetische Effekte auf die Erhöhung von Freien Leichtketten im Serum

Posterbeitrag im Rahmen des Forschungstages der medizinischen Fakultät, Institut für medizinische Informatik, Biometrie und Epidemiologie, Universitätsklinikum Essen, 2016

Marina Arendt, Börge Schmidt, Susanne Moebus, Jan Dürig, Karl-Heinz Jöckel, Lewin Eisele

Entzündungs-Gensets: Anreicherung genetischer Varianten, die mit monoklonaler und polyklonaler Erhöhung Freier Leichtketten assoziiert sind

Vortrag auf der Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi) Universität zu Lübeck, 2017

Gleichzeitige Veröffentlichung in „Das Gesundheitswesen 8/9 2017“.

Inhaltsverzeichnis

<u>1</u>	<u>EINLEITUNG</u>	<u>6</u>
1.1	FREIE LEICHTKETTEN	8
1.2	PATHOPHYSIOLOGIE FREIER LEICHTKETTEN	9
1.3	MOLEKULARGENETISCHER HINTERGRUND	13
1.4	GENETISCHE UND METHODISCHE ASPEKTE GENOMWEITER ASSOZIATIONSSTUDIEN	15
1.5	GENESET ENRICHMENT ANALYSE	19
1.6	FRAGESTELLUNG	20
<u>2</u>	<u>METHODEN</u>	<u>21</u>
2.1	PROBANDEN- UND LABORDATEN	21
2.2	PROZESSIERUNG DER GENOTYPISIERUNGSDATEN	23
2.3	GENOMWEITE ASSOZIATIONSSTUDIE – GWAS	28
2.4	DARSTELLUNG DER GWAS-ERGEBNISSE	31
2.5	GENESET ENRICHMENT ANALYSE – GSEA	32
2.6	EINGESETZTE SOFTWARE	35
<u>3</u>	<u>ERGEBNISSE</u>	<u>36</u>
3.1	BESCHREIBUNG DER STUDIENPOPULATION	36
3.2	GENOMWEITE ASSOZIATIONSSTUDIE – GWAS	40
3.3	LOCUSZOOM	43
3.4	GENESET ENRICHMENT ANALYSE – GSEA	44

4	<u>DISKUSSION</u>	56
4.1	UNTERSUCHTE POPULATION UND LABORANALYSEN	57
4.2	AUSWAHL DER KOVARIABLEN	59
4.3	GENOMWEITE ASSOZIATIONSSTUDIE UND GENESET ENRICHMENT ANALYSE	62
4.4	GWAS ZUR POLYKLONALEN FLC-ERHÖHUNG	63
4.5	GENSETS AUS DEN ANALYSEN ZUR POLYKLONALEN FLC-ERHÖHUNG	64
4.5.1	GWAS ZUR MONOKLONALEN FLC-ERHÖHUNG	66
4.5.2	GENSETS AUS DEN ANALYSEN ZUR MONOKLONALEN FLC-ERHÖHUNG	68
5	<u>AUSBLICK</u>	73
6	<u>ZUSAMMENFASSUNG</u>	75
7	<u>LITERATURVERZEICHNIS</u>	76
8	<u>ANHANG</u>	88
8.1	ABKÜRZUNGEN	88
8.2	HAUPTKOMPONENTENANALYSE	90
8.3	GENSETS MIT EINER FDR < 25 %	91
8.4	ABBILDUNGSVERZEICHNIS	98
8.5	TABELLENVERZEICHNIS	99
9	<u>DANKSAGUNG</u>	101
10	<u>LEBENS LAUF</u>	102

1 Einleitung

Leichtketten nehmen eine wichtige Rolle im Immunsystem des Menschen ein. Freie Leichtketten (engl. *Free light-chains*, FLC) werden während der Antikörperproduktion ins Blut sezerniert und sind wichtige Parameter im klinischen Alltag bei der Diagnose und Verlaufskontrolle von Plasmazellerkrankungen wie dem Multiplen Myelom (MM). Hierbei ist eine monoklonale Erhöhung der FLC zu beobachten, die zum Nachweis des monoklonalen Zellklons verwendet wird. Das Verhältnis aus den beiden FLC-Konzentrationen von κ und λ wird als $\kappa : \lambda$ -Ratio bezeichnet. Bei der Monoklonalen Gammopathie unklarer Signifikanz (MGUS), die hämatologischen Neoplasien vorangeht, zeigt eine pathologisch veränderte Ratio ein erhöhtes Progressionsrisiko zum MM an. Von polyklonaler FLC-Erhöhung spricht man, wenn beide FLC im gleichen Maße erhöht sind, zum Beispiel im Zuge von chronischen oder akuten Entzündungsreaktionen, sodass die Ratio jedoch im normalen Bereich verbleibt.

In Dispenzieri et al. (2012) wird der Effekt einer FLC-Erhöhung im Serum auf die Gesamtmortalität unabhängig schon vorhandener Lymphoproliferativer Erkrankungen (engl. *Lymphoproliferative disease*, LPD) berichtet. Hierbei bleibt der Mechanismus, über den FLC die Gesamtmortalität beeinflusst, ungewiss. Ebenso ist der negative Einfluss von polyklonaler FLC-Erhöhung auf das Gesamtüberleben in der hier verwendeten Population berichtet worden, bei der die direkte biologische Rolle ebenso unbestimmt bleibt (Eisele et al., 2010). Über die Progression von MGUS zu einer lymphatischen Neoplasie wird unter anderem in Dispenzieri et al. (2010) sowie in Rajkumar et al. (2005) über diese Progression abhängig der pathologischen $\kappa : \lambda$ -Ratio berichtet. In beiden Arbeiten bleibt der Wirkungsmechanismus auf die Progression unbekannt. In einer eigenen Arbeit (bisher unveröffentlicht) wird die Frage untersucht, inwieweit die polyklonale FLC-Erhöhung in Zusammenhang mit kardialer Verkalkung steht und so die Gesamtmortalität und die kardiale Mortalität beeinflusst.

Einleitung

Da zu den biologischen Ursachen noch wenig bekannt ist, ergibt sich aus diesen klinischen Gesichtspunkten die Frage, ob es eine genetische Ursache gibt, die die FLC-Erhö-
hung beeinflusst.

Sowohl für die Entstehung von Plasmazellerkrankungen, als auch für die Empfindlichkeit für systemische Entzündungsreaktionen können genetische Varianten eine Rolle spielen. Zum Beispiel konnten für die Entstehung des MMs eine Reihe von Risikoallelen identifiziert werden (Broderick et al., 2011; Chubb et al., 2013; Greenberg et al., 2012; Johnson et al., 2016), jedoch wurde bisher noch keine Analyse des Einflusses der genetischen Konstitution auf die Erhöhung von FLC durchgeführt.

1.1 Freie Leichtketten

FLC wurden erstmals von Bence Jones (Bence Jones, 1847) entdeckt, als er das gleichzeitige Vorkommen von Harnproteinen und der Knochenerweichung beschrieb. Leichtketten werden im Verlauf der humanen Immunabwehr in Plasmazellen, differenziert aus Stammzellen des Knochenmarks, als Teilstücke von Antikörpern gebildet und stellen so einen Teil der adaptiven Immunantwort dar. Antikörper sind Proteine, deren Grundgerüst aus vier Molekülketten bestehen, zwei Schwereketten und zwei Leichtketten. Es gibt fünf verschiedene Schwereketten, die mit α , γ , μ , δ und ϵ beschrieben werden. Danach werden die zugehörigen Isotypen der Immunglobuline (Ig) als IgA (Dimer), IgM (Pentamer), IgG, IgD und IgE (jeweils als Monomer) bezeichnet, wobei jedes eine andere Rolle im humanen Immunsystem als Abwehrmechanismus einnimmt. Neben diesen Schwereketten existieren zwei Arten von Leichtketten, κ und λ (van den Tweel, Jang G. et al., 1999). Die Bezeichnung als κ und λ haben die Leichtketten von Korngold und Laipiri, die 1965 gezeigt haben, dass Leichtketten Teile der Immunglobulin-Moleküle sind (Bradwell, 2010).

Leichtketten werden von den Plasmazellen in Überproduktion hergestellt, damit eine intakte Zusammensetzung von Schwereketten mit den Leichtketten gewährleistet werden kann (Bradwell, 2010). Überschüssige Leichtketten werden in den Blutkreislauf sezerniert. Hierdurch gibt es neben den an Immunglobulinen gebundenen Leichtketten die Freien Leichtketten im Serum. Obwohl κ doppelt so viel produziert wird wie λ , liegt eine normale $\kappa : \lambda$ -Ratio in gesunden Menschen bei 0,26 - 1,65. Dies resultiert aus der niedrigeren renalen Filtrationsrate von λ , das als Dimer vorliegt. Bei Menschen mit Niereninsuffizienz liegt die normale Ratio bei 0,37 - 3,1. Im Urin ist die verbleibende FLC-Konzentration bei gesunden Personen sehr gering (Hutchison et al., 2008).

1.2 Pathophysiologie Freier Leichtketten

Die Erhöhung der FLC im Serum kann ein Anzeichen hämatologischer Krankheiten wie dem Multiplen Myelom (MM) und anderen Lymphoproliferativen Erkrankungen (LPD) sein, ebenso wie ein Anzeichen akuter oder chronischer Entzündungsreaktionen darstellen (Brebner et al., 2013; Charafeddine et al., 2012).

Pathologische Erhöhungen der FLC-Konzentrationen im Blut können in polyklonale und monoklonale Erhöhungen differenziert werden. Bei einer monoklonalen Erhöhung wird die Überproduktion durch einen einzigen Zellklon verursacht, der dann monoklonales Protein in Form von entweder kompletten Immunglobulinen oder FLC produziert. Diese Überproduktion bewirkt eine Verschiebung der $\kappa : \lambda$ -Ratio außerhalb des normalen Bereichs. Bei einer polyklonalen Erhöhung bleibt die Ratio im Gleichgewicht, da beide Typen an FLC gleichermaßen erhöht sind (Dispenzieri et al., 2009).

Eine monoklonale FLC-Erhöhung im Serum liegt typischerweise bei einer malignen Plasmazellproliferation, also der bösartigen Vermehrung von Immunzellen, vor. Diese erhöhte FLC-Produktion kann durch Neoplasien wie dem MM, einer Amyloidose oder der Vorstufe des Multiplen Myeloms, der MGUS verursacht sein (Dispenzieri et al., 2009; Eisele et al., 2012; Kyle, 2000; Landgren et al., 2009). Bei gesunden Menschen gelangen etwa 1 - 10 mg pro Tag FLC in den Urin. Bei höheren FLC-Konzentrationen im Urin liegt dies entweder an einer Nierenerkrankung oder an einer malignen LPD, bei denen Bence-Jones-Proteine in den Urin abgegeben werden (Bradwell et al., 2001). Bei einer monoklonalen FLC-Erhöhung beispielsweise durch ein MM steigt zu Beginn nur die FLC-Konzentration im Serum, wohingegen die renale Filtrationsrate und der FLC-Metabolismus so hoch bleiben können, dass keine FLC in den Urin gelangen.

Einleitung

Erst wenn die FLC-Produktion 10 - 30 mg pro Tag übersteigt, ist die Reabsorptionskapazität der Nieren erschöpft, wodurch die FLC-Konzentration im Urin stetig – bis zur Überbeanspruchung und dem damit verbundenen Nierenversagen – zunimmt. Bei einer kurzzeitigen polyklonalen Erhöhung, beispielsweise im Zuge einer akuten Entzündungsreaktion, wird die FLC-Konzentration im Urin nicht beeinflusst (Bradwell, 2010).

Neben der MGUS und Plasmazellneoplasien wie dem Multiplen Myelom können auch verschiedene andere Lymphome (zum Beispiel verschiedene Non-Hodgkin-Lymphome oder eine chronisch lymphatische Leukämie) für eine monoklonale FLC-Erhöpfung verantwortlich sein (Tsai et al., 2009).

Monoklonale Gammopathien im Allgemeinen sind durch das Vorkommen von monoklonalen Zellklonen und der damit verbundenen Sekretion von monoklonalem Protein gekennzeichnet. In den meisten Fällen wird die monoklonale Gammopathie von Plasmazellen verursacht, deren physiologische Aufgabe darin besteht Antikörper zu produzieren und zu sekretieren (van den Tweel, Jang G. et al., 1999).

Die MGUS selbst wird definiert über die Konzentration eines monoklonalen Proteins kleiner als $30 \frac{g}{L}$ und weniger als 10 % klonale Plasmazellen im Knochenmark, jedoch ohne die zusätzlichen Diagnosekriterien eines Multiplen Myeloms, anderer B-Zell proliferativen Erkrankungen oder einer Amyloidose (The International Myeloma Working Group, 2002). MGUS Patienten mit einer pathologischen FLC-Ratio haben eine höhere Progressionsrate von MGUS zum Multiplen Myelom als solche ohne pathologische Ratio (Rajkumar et al., 2005). Ebenso ist die Progressionsrate davon abhängig, wie hoch die Konzentration des monoklonalen Proteins ist und welcher Isotyp der Schwereketten vorliegt. Dabei ist eine Progression umso wahrscheinlicher, je höher die Konzentration des monoklonalen Proteins ist, je weiter die pathologische Ratio von dem normalen Bereich abweicht und wenn der Isotyp der Schwereketten ein anderer als IgG ist. Über die verschiedenen Abstufungen, kann der Schweregrad beziehungsweise das Progressionsrisiko der MGUS erhoben werden (Dispenzieri et al., 2009; Rajkumar et al., 2005).

Einleitung

In Abgrenzung zur MGUS, die keine klinischen Symptome zeigt, wird ein Multiples Myelom über mehrere pathologische Veränderungen definiert, die in den CRAB-Kriterien zusammengefasst werden. CRAB steht für *Calcium levels increment, renal insufficiency, anaemia and bone lesions*. Das MM ist damit durch das Vorliegen einer oder mehrerer der Veränderungen Hyperkalzämie, Niereninsuffizienz, Anämie und Knochenläsionen definiert.

2014 wurden die CRAB-Kriterien durch die SLiM-Kriterien erweitert. Bei den SLiM-Kriterien liegt die Plasmazellinfiltration im Knochenmark über 60 % (Sixty, S), die Ratio der erhöhten Freien Leichtkette zur nicht betroffenen Leichtkette beträgt über 100 (Light-chain, Li) und es wird mehr als eine osteolytische Läsion im Ganzkörper-MRT sichtbar (MRT, M) (Rajkumar et al., 2014).

Entzündungsreaktionen oder chronische Infekte, die das Immunsystem stimulieren, ebenso wie Autoimmunkrankheiten (zum Beispiel Systemischer Lupus erythematoses) können zu einer polyklonalen FLC-Erhöhung führen (Draborg et al., 2015). Hier sind beide FLC gleichermaßen erhöht, sodass die $\kappa : \lambda$ -Ratio im normalen Bereich verbleibt. Polyklonale Erhöhungen entstehen meist durch eine akute oder chronische Stimulation des Immunsystems innerhalb der Aktivierung der angeborenen Immunantwort (Brebner et al., 2013). Auch über den Effekt der FLC auf Entzündungsreaktionen ist bereits berichtet worden (Braber et al., 2012; Redegeld et al., 2012).

Einleitung

Weitere Aspekte beeinflussen die FLC-Konzentration im Serum über Veränderungen der Produktion und Ausscheidung der FLC. Dazu zählt die Erhöhung von FLC im Rahmen von Arthritis (Brebner et al., 2013; Nakano et al., 2011). Ebenso wirkt das Alter auf die renale Filtrationsrate ein und kann so eine polyklonale FLC-Erhöpfung hervorrufen (Hutchinson et al., 2012; Hutchison et al., 2008). Körperliche Aktivität hingegen scheint dieser Erhöhung in gewissem Rahmen entgegenwirken zu können (Geffken, 2001).

Auch eine Erniedrigung der FLC kann in einem pathologischen Maßstab vorliegen, zum Beispiel als Reaktion auf eine Chemo- oder Strahlentherapie (Balducci, 2003). Da der Fokus dieser Arbeit auf die Erhöhung der FLC im Serum gerichtet ist, wird hierauf nicht weiter eingegangen.

1.3 Molekulargenetischer Hintergrund

Das humane Genom setzt sich aus 23 Chromosomenpaaren zusammen, wovon ein Chromosomenpaar die Gonosomen (Geschlechtschromosomen) und 22 Chromosomenpaare die Autosomen sind. Diese Chromosomen formen sich aus kondensierter Desoxyribonukleinsäure (engl. *Deoxyribonucleic acid*, DNA). Der Träger der genetischen Information ist kodiert durch die verschiedene Anordnung der vier Nukleinbasen Adenin (A), Thymin (T), Cytosin (C) und Guanin (G). DNA hat die Form einer Doppelhelix, in der Adenin mit Thymin und Guanin mit Cytosin über Wasserstoffbrückenbindungen verbunden sind. Im humanen Genom sorgen komplexe Mechanismen dafür, dass auf demselben DNA-Abschnitt Informationen für mehrere Gene gespeichert werden können. Hierfür sorgen unter anderem die Introne und Exone, wobei diese Bereiche immer genspezifisch sind und die Exone im Zuge der Transkription in Ribonukleinsäure (engl. *Ribonucleic acid*, RNA) ausgeschnitten und die Introne zu dem jeweiligen Gen zusammengefügt werden (Clark, 2006).

Die Struktur und insbesondere die Abfolge der Basen auf den DNA-Strängen sind zum überwiegenden Teil selbst zwischen zwei nicht-verwandten Personen identisch. Unterschiede können durch entweder gravierendere oder weniger einflussreiche Veränderungen entstehen, die sich im Humangenom in drei verschiedene Größenklassen einordnen lassen.

Bei der Genommutation ist die Anzahl der Chromosomen verändert, es sind also Chromosomen verloren gegangen oder hinzugekommen, wie zum Beispiel bei der Down Syndrom hervorrufenden Chromosom 21 Triploidie (Knippers, 1997).

Bei Chromosomenmutationen sind die Chromosomen in ihrer Struktur so verändert, dass dies schon unter dem Lichtmikroskop sichtbar wird. Dabei gehen zum Beispiel ganze Abschnitte eines Chromosoms verloren oder kommen von anderen Chromosomen hinzu. Ein Beispiel hierfür ist das Philadelphia-Chromosom, das den Namen vom Ort seiner Entdeckung trägt. Das Philadelphia-Chromosom ist eine wichtige Ursache für

Einleitung

die Entstehung der chronische myeloische Leukämie (CML). Hierbei gibt es eine Translokation der Gene BCR und ABL, die zusammen ein sogenanntes Fusionsgen bilden und die Transformation von Leukozyten in Leukämiezellen bewirken (Nowell et al., 1960).

Die dritte Form der Mutationen ist die Genmutation, die wiederum in Raster- und Punktmutationen unterteilt werden kann. Bei einer Rastermutation wird durch eine Insertion oder Deletion die Anordnung der Nukleinbasen verändert und somit die Struktur des kodierten Proteins verändert. Da die Proteine über Triplets kodiert werden, kommt es zu meist größeren Proteinveränderungen, wenn die Anzahl der inserierten oder deletierten Nukleinbasen kein Vielfaches von drei ist.

Eine Punktmutation, bei der eine einzelne Base durch eine andere ersetzt wird, wird als Basensubstitution bezeichnet (Knust, E., Janning, W., 2008). Die konkrete Reihenfolge der Basenpaare variiert von Mensch zu Mensch, wodurch verschiedene Formen eines Gens entstehen können. Eine so entstandene konkrete Ausprägung eines Gens wird als Allel bezeichnet.

In verschiedenen Populationen können die konkreten Ausprägungen eines Gens in unterschiedlicher Häufigkeit auftreten, das heißt sie unterscheiden sich in ihrer Allelfrequenz. Die Lokalisation auf dem Genom wird als Genort oder auch Locus bezeichnet. Die genetische Terminologie für einen Einzelbasenaustausch ist SNP (engl. *Single Nucleotide Polymorphism*), sofern diese Variante in der Population eine Häufigkeit von mehr als 1 % aufweist; andernfalls spricht man von SNVs (engl. *Single Nucleotide Variation*) (Ziegler et al., 2008). Befindet sich ein SNP innerhalb eines Gens, so hat dieses Gen mehrere Allele. Im menschlichen Genom befinden sich hunderttausende SNPs in Genen, aber ebenso in nichtkodierenden Bereichen, die dennoch biologische Auswirkungen haben können. Werden solche SNPs zur Entdeckung von Krankheits- oder phänotypisch relevanten Genorten verwendet, so werden sie auch als Marker bezeichnet. Die gleichzeitige Bestimmung tausender Genotypen erfolgt heute auf DNA-Chips durch spezifische Hybridisierungsreaktionen (Clark, 2006).

1.4 Genetische und methodische Aspekte genomweiter Assoziationsstudien

Man spricht von genetischer Assoziation, wenn ein spezifisches Allel in einer betroffenen Gruppe, zum Beispiel Probanden mit einer pathologischen FLC-Ratio, häufiger vorkommt als in einer Vergleichsgruppe. In dieser Arbeit werden biallelische SNPs betrachtet, also solche, von denen es zwei Ausprägungen gibt, zum Beispiel die Varianten $\frac{C}{G}$ und $\frac{A}{T}$ ¹. Diese spezielle Variante würde man als C-A-SNP bezeichnen, da die Nukleinbase C ausgetauscht wurde durch die Nukleinbase A. Einen G-C- oder einen T-A-Austausch bezeichnet man als palindromischen SNP. Da bei diesen Konstellationen die DNA-Stränge nicht unterscheidbar und damit die Haplotypen nicht bestimmbar sind, werden palindromische SNPs aus den GWAS-Analysen ausgeschlossen (Verma et al., 2014). Der Basenaustausch kann auf einem Chromosom oder auf beiden der homologen Chromosomen geschehen. Dadurch ergeben sich drei Varianten, die in einer betrachteten Population vorkommen können.

$$1. \frac{\text{Chr. 1.1}}{\frac{C}{G}} \left| \frac{\text{Chr. 1.2}}{\frac{C}{G}} \right., \quad 2. \frac{\text{Chr. 1.1}}{\frac{C}{G}} \left| \frac{\text{Chr. 1.2}}{\frac{A}{T}} \right. \quad \text{und} \quad 3. \frac{\text{Chr. 1.1}}{\frac{A}{T}} \left| \frac{\text{Chr. 1.2}}{\frac{A}{T}} \right. _2$$

¹ C und G bezeichnen dabei die komplementären Nukleinbasen auf einem DNA-Strang, die an dieser Stelle in dem anderen Allel durch A und T ausgetauscht sind.

²Chr. 1.1 und Chr. 1.2 bezeichnen die beiden homologen Chromosomen. Die Kombinationen C-G und A-T bezeichnen die Basenpaare, die sich auf den komplementären Strängen gegenüber liegen.

Einleitung

Das Allel, das von beiden die geringere Allelfrequenz innerhalb der Population aufweist, wird als *Minor Allele* bezeichnet. Sofern bei dem Beispiel oben A das Risikoallel ist – hier gleich bedeutend mit dem *Minor Allele* – können die drei Ausprägungen auch als die Anzahl von 0, 1 oder 2 Risikoallelen bezeichnet werden. Nach dem Hardy-Weinberg Equilibrium (HWE) bleibt die Verteilung dieser drei Genotypen mit zugehörigen Allelfrequenzen in einer unendlich großen und sich zufällig vermehrenden Population über mehrere Generationen stabil (Ziegler et al., 2008; Ziegler et al., 2010).

Ein Ansatz in der Assoziationsanalyse ist es, einen linearen Effekt über die Anzahl der Risikoallele anzunehmen, was als additives Modell bezeichnet wird. Je nach Skalenniveau des Phänotyps führt man eine logistische oder lineare Regression für den betrachteten SNP durch. Als Ergebnis erhält man den Effektschätzer, der die phänotypische Ausprägung abhängig des Genotyps beschreibt (Ziegler et al., 2010).

In genomweiten Assoziationsstudien (GWAS) führt man die Regressionsanalysen mit genotypisierten, über das gesamte Genom verteilten SNPs durch. Hierbei bedient man sich sogenannter '*tag-SNPs*', also SNPs, die den Haplotypenblock, in dem sie liegen, angemessen repräsentieren (Illumina®, 2010). Als Haplotyp wird die Folge an Basenpaaren der DNA beschrieben, die hintereinander auf der DNA liegen. Diese bilden sogenannte Haplotypenblöcke, die zusammen an die nächste Generation weitergegeben werden.

Einleitung

Die genetischen Loci, die in einem solchen Bereich liegen, sind abhängig voneinander, was man Kopplungsungleichgewicht (engl. *Linkage Disequilibrium*, LD) nennt. Entsprechend ist die Rekombinationsrate innerhalb dieser Regionen besonders gering. Liegen zwei SNPs in hohem LD, werden sie fast immer zusammen vererbt. In GWAS lässt sich dieses Wissen so nutzen, dass immer nur ein SNP eines Haplotypenblockes genotypisiert werden muss, um die Information des gesamten genetischen Bereiches zu erfassen (Ziegler et al., 2010). Das bedeutet auch, dass ein genotypisierter SNP, der mit einem Phänotyp assoziiert ist, nicht direkt in der funktionellen Region eines Gens liegen muss, sondern sich in hohem LD mit einem SNP im eigentlich verursachenden Gen befinden kann (Ziegler et al., 2008). Die genetische Distanz zweier Bereiche auf einem Chromosom kann mit der Einheit Centimorgan (cM) quantifiziert werden. Hierbei beschreibt ein Centimorgan die Wahrscheinlichkeit eines Crossovers in 100 Meiosen, also der Rekombinationsfrequenz von 1 % (Bickeböller et al., 2007).

Eine genetische Region, die ein besonders hohes LD aufweist, ist die des Humanen Leukozyten Antigens (HLA). Innerhalb einer Vererbungslinie weisen diese Regionen nur eine Rekombinationsrate von 2 % auf. Trotzdem ist die HLA-Region ein sehr polymorphe Region, das heißt, dass zu jedem Gen viele allelische Ausprägungen existieren (Mueller-Eckhardt et al., 2004). In genetischen Analysen, wie zum Beispiel der unten beschriebenen *Geneset Enrichment Analyse*, werden diese Regionen ausgeschlossen, da sie eine besonders hohe Gendichte aufweisen (Segrè et al., 2010).

Einleitung

Die Verteilungen von Genotypen können sich – wie oben schon beschrieben – zwischen verschiedenen Populationen unterscheiden. Das kann bei der Untersuchung von genetischen Effekten mittels GWAS zu einer Verzerrung der Ergebnisse führen. Dieser Störeffekt wird *Population Stratification* genannt. Darüber hinaus können Familienstrukturen, sowie verborgene Verwandtschaften Fehler hervorrufen, was sich häufig in Form von falsch-positiven Ergebnissen zeigt (Price et al., 2010). Als Ergebnis von GWAS erhält man je nach Regressionsanalyse einen Schätzwert, der die Veränderung der abhängigen Variablen – also der Operationalisierung des Phänotyps – pro Zunahme um ein Risikoallel beschreibt. Wenn der Phänotyp als binäre Variable vorliegt, wird dieser Schätzwert als Odds Ratio (OR) ausgedrückt und kann als Chance interpretiert werden, abhängig des Genotyps den Phänotyp aufzuweisen. Liegt der Phänotyp als kontinuierliche Variable vor, wird dieser Schätzwert als Regressionskoeffizient ausgedrückt. Dieser beschreibt die Zu- oder Abnahme des Phänotyps pro Zunahme um ein Risikoallel. Ob diese ermittelte Assoziation als zufällig betrachtet werden kann, wird über das Signifikanzlevel entschieden. Da über das gesamte Genom getestet wird und somit genauso viele Tests durchgeführt werden wie SNPs genotypisiert sind, führt das nominale Level von $\alpha = 5\%$ zu vielen falsch-positiven Ergebnissen (Multiples Testproblem). Um dafür zu korrigieren, hat sich in GWAS die Verwendung eines genomweiten Signifikanzlevels von $\alpha = 5 \cdot 10^{-8}$ etabliert. Dies entspricht einer Korrektur nach Bonferroni, wenn man vom gleichzeitigen Testen 1.000.000 genetischer Varianten ausgeht. Ebenso ist eine Kontrolle über die *False Discovery Rate* (FDR) möglich (Holden et al., 2008). Die FDR beschreibt den erwarteten Anteil aller falsch-positiven Ergebnisse an allen signifikanten Testergebnissen. Die Einschätzung der Verlässlichkeit der Ergebnisse wird somit über die Wahrscheinlichkeit getroffen, dass die berichteten Ergebnisse falsch-positive Ergebnisse sind (Benjamini et al., 1995; Subramanian et al., 2005).

1.5 Geneset Enrichment Analyse

Geneset Enrichment analysis (GSEA) lässt sich mit „Anreicherungsanalyse auf der Ebene von Gruppen von Genen“ übersetzen. Sie ist beispielsweise als Folgeanalyse einer GWAS sinnvoll.

In einer GSEA werden Sets von Genen betrachtet, die zum Beispiel funktionell verwandt sind oder bei einer biologischen Funktion gemeinsam agieren. Obwohl SNPs in einer GWAS zum Beispiel nicht genomweit signifikant sind, können in der GSEA zusammengehörige Gruppen von SNPs identifiziert werden, die mit dem Phänotyp assoziiert sind. Jeder zu den Genen eines Sets gehörende SNP kann so einen Teil zur Assoziation des Gensets mit dem Phänotyp beitragen, auch wenn dieser Beitrag eventuell nur gering ausfällt (Segrè et al., 2010).

Die Gensets, in denen eine Anreicherung getestet werden soll, sind vorab definiert und in Datenbanken hinterlegt. Dabei werden zum Beispiel Gene zusammengefasst, die gleiche regulatorische Eigenschaften haben oder auf einem Chromosom liegen.

1.6 Fragestellung

Im Rahmen dieser Dissertationsschrift wurde der Einfluss genetischer Varianten, das heißt SNPs, auf die monoklonale und die polyklonale Erhöhung der Freien Leichtketten im Serum untersucht, um das Verständnis für die biologischen Zusammenhänge zu erweitern. Hierfür wurden Daten der Heinz Nixdorf Recall Studie (HNR), einer populationsbasierten Kohortenstudie im Ruhrgebiet, verwendet. Für die Probanden der HNR Studie lagen zum einen Messwerte für die FLC im Serum vor, ebenso wie Genotypisierungsdaten, die mit Hilfe von SNP-Chips der Firma Illumina® ermittelt wurden.

Um genetische Effekte nachzuweisen, wurde eine Genomweite Assoziationsstudie (GWAS) mit FLC als Phänotyp durchgeführt. In getrennten Analysen sollten dabei sowohl genetische Varianten für die polyklonale als auch für die monoklonale FLC-Erhöpfung identifiziert werden. Varianten, die mit einer polyklonalen Erhöhung assoziiert sind, könnten das Verständnis der Physiologie der FLC bei entzündlichen Prozessen verbessern. Marker, die mit einer monoklonalen Erhöhung assoziiert sind, könnten ebenfalls mit dem Auftreten hämatologischer Neoplasien assoziiert sein.

Das Hauptaugenmerk dieser Arbeit liegt auf der Identifizierung von funktionell zusammenhängenden Gensets mit der polyklonalen und monoklonalen FLC-Erhöpfung. Dadurch können biologisch relevante Assoziationen mit der FLC-Erhöpfung aufgedeckt werden. In einem weiteren Schritt wurde deshalb eine *Geneset Enrichment Analyse* (GSEA) durchgeführt, um zu prüfen, ob die SNPs der GWAS' angereichert in bestimmten biologischen Reaktionswegen vorliegen. Eine Anreicherung in einem Signalweg oder einer biologischen Funktion kann so weitere Kenntnisse über die beteiligten biologischen Abläufe der polyklonalen und monoklonalen FLC-Erhöpfung liefern.

2 Methoden

2.1 Probanden- und Labordaten

Als Datengrundlage werden Fragebogen-, Labor- sowie Genotypisierungsdaten der populationsbasierten Heinz Nixdorf Recall Studie (HNR) verwendet. HNR ist eine Kohortenstudie, für die in den Jahren 2000 bis 2003 aus den drei Nachbarstädten Essen, Bochum und Mülheim/Ruhr 4.814 Frauen und Männern im Alter zwischen 45 und 75 Jahren rekrutiert wurden. Der Großteil der Studienpopulation ist zentraleuropäischer Herkunft. Die Studienregion umfasst etwa 600 km² mit ungefähr 1,2 Millionen Einwohnern. Die Probanden wurden zufällig aus den Registern der Einwohnermeldeämter gezogen und haben eine Einverständniserklärung unterschrieben. Die Untersuchung bei Studienbeginn fand von 2000-2003 statt, die 5-Jahres-Nachverfolgung von 2006-2008 und die 10-Jahres-Nachverfolgung von 2011-2015. Weitere Details können den Publikationen von Schmermund et al. und Erbel et al. entnommen werden (Erbel et al., 2012; Schmermund et al., 2002).

Die Serumproben der Untersuchung bei Studienbeginn, bei der 5-Jahres- und der 10-Jahres-Nachverfolgung wurden mit Hilfe von Serumproteinelektrophorese (SPE) auf das Vorhandensein eines monoklonalen Proteins untersucht.

FLC wurden auf einem Dade Behring BNII automatisierten Nephelometer (Siemens, Germany) mit Hilfe eines kommerziell verfügbaren Kits (FREELITE, The Binding Site Ltd, Birmingham, UK) unter Verwendung von hochspezifischen Antiseren gemessen.

MGUS wurde nach den in Kapitel 1.2 vorgestellten Kriterien definiert, einschließlich der Detektion von monoklonalem Protein in der SPE und/oder in der IFE, monoklonaler Proteinkonzentration, Laborergebnisse und Krankheitsgeschichte (Eisele et al., 2012).

Methoden

Als Referenzwerte für die Konzentrationen der Freien Leichtketten werden publizierte Werte für κ von $3,3 - 19,4 \frac{mg}{L}$ und für λ von $5,7 - 26,3 \frac{mg}{L}$ verwendet (Katzmann et al., 2002). Eine pathologische $\kappa : \lambda$ - Ratio (pFLCR) wird durch die in Formel 1 dargestellte Regel definiert (Hutchison et al., 2008).

Formel 1 **Definition einer pathologischen FLC-Ratio abhängig von der Nierenfunktion, die mit der glomerulären Filtrationsrate (eGFR) geschätzt wurde**

$$pFLCR = \begin{cases} \kappa/\lambda < 0,26 & \left| \kappa/\lambda > 1,65 \text{ falls } eGFR > 30 \text{ mL/min} \right. \\ \kappa/\lambda < 0,37 & \left| \kappa/\lambda > 3,1 \text{ falls } eGFR < 30 \text{ mL/min} \right. \end{cases}$$

Die geschätzte glomeruläre Filtrationsrate (engl. *Estimated glomerular filtration rate*, eGFR) wird mit Hilfe der verkürzten MDRD Formel berechnet (Levey et al., 2007). Über die eGFR lässt sich, wie in Formel 1 beschrieben, ein pathologischer Bereich der FLC in Abhängigkeit der Nierenfunktion definieren.

Weitere Informationen wurden aus Fragebögen erhoben, die die Studienteilnehmer jährlich beantworteten. Änderungen im Gesundheitszustand wurden bei den behandelnden Ärzten und Krankenhäusern validiert und weitere medizinische Informationen eingeholt. Die Diagnosen wurden nach dem ICD-10 Code verschlüsselt (International classification of diseases, 10th revision) (World Health Organization, 2016). Für dieses Vorhaben wurde die HNR-Datenbank nach ICD-10 Codes für Lymphoproliferative Erkrankungen (ICD-10: C81-C91) durchsucht.

Für die deskriptive Statistik der Studienpopulation wurde je nach Merkmal die Anzahl mit Prozentangabe der jeweiligen Untergruppe, beziehungsweise der Median mit erstem und drittem Quartil angegeben. Zum Vergleich von kategorialen und stetigen Merkmalen zwischen genotypisierten und nicht-genotypisierten Probanden wurde der exakte Test nach Fisher beziehungsweise der Wilcoxon-Mann-Whitney-Test eingesetzt.

2.2 Prozessierung der Genotypisierungsdaten

Durch die Microarray-Technologie wird es ermöglicht, eine große Anzahl an DNA-Genotypisierungen gleichzeitig zu ermitteln, zum Beispiel um Mutationen wie SNPs aufzudecken (Clark, 2006). In dieser Arbeit wurden die genomweiten SNP-Chip-Arrays HumanOmni1-Quad und HumanOmniExpress von Illumina® verwendet. Der HumanOmni1-Quad umfasst etwa eine Million Varianten und der HumanOmniExpress etwa 700.000 Varianten (Illumina®, 2010).

Die Genotypisierungen der HNR-Probanden wurden am Institut für Humangenetik, Universitätsklinikum Bonn unter der Leitung von Prof. Dr. Nöthen durchgeführt.

Um eine optimale Datenbasis für die späteren Analysen zu schaffen, sind bei einer GWAS verschiedene Qualitätskontrollschritte notwendig (Ziegler et al., 2008). Der HumanOmni1-Quad musste von der Genomversion hg18 auf hg19 konvertiert werden. Dabei werden die SNP-Positionen angeglichen, die sich je nach Version unterscheiden können, was in folgenden Analysen zu Fehlern führen würde.

Methoden

Ein Qualitätsmerkmal für die Reinheit der DNA-Proben ist die Heterozygotierate. Hierüber werden Individuen ermittelt, die einen deutlich erhöhten oder verminderten Anteil an heterozygoten Genotypen aufweisen. In der betrachteten Population wird die mittlere Heterozygotierate über folgende Formel bestimmt (Anderson et al., 2010).

Formel 2 Berechnung der Heterozygotierate pro Individuum. N_{gesamt} ist die Anzahl der nicht-fehlenden Genotypen und $N_{homozygot}$ die Anzahl der homozygoten Genotypen pro Individuum

$$HET = \frac{N_{gesamt} - N_{homozygot}}{N_{gesamt}}$$

Eine sehr hohe oder niedrige Heterozygotierate weist bei humanen DNA-Proben meistens auf eine Kontamination oder auf einen technischen Fehler des Assays hin (Turner et al., 2011).

Die Hauptkomponentenanalyse (engl. *Principal component analysis*, PCA) kann dazu verwendet werden, Populationsstrukturen in genetischen Daten zu ermitteln. Unterschiede in der Abstammung können dann zum Beispiel mittels graphischer Darstellung nachgewiesen werden (vgl. Abbildung 1). Die PCA ist dabei verlässlich in der Detektion von Populationsstrukturen (Novembre et al., 2008).

Methoden

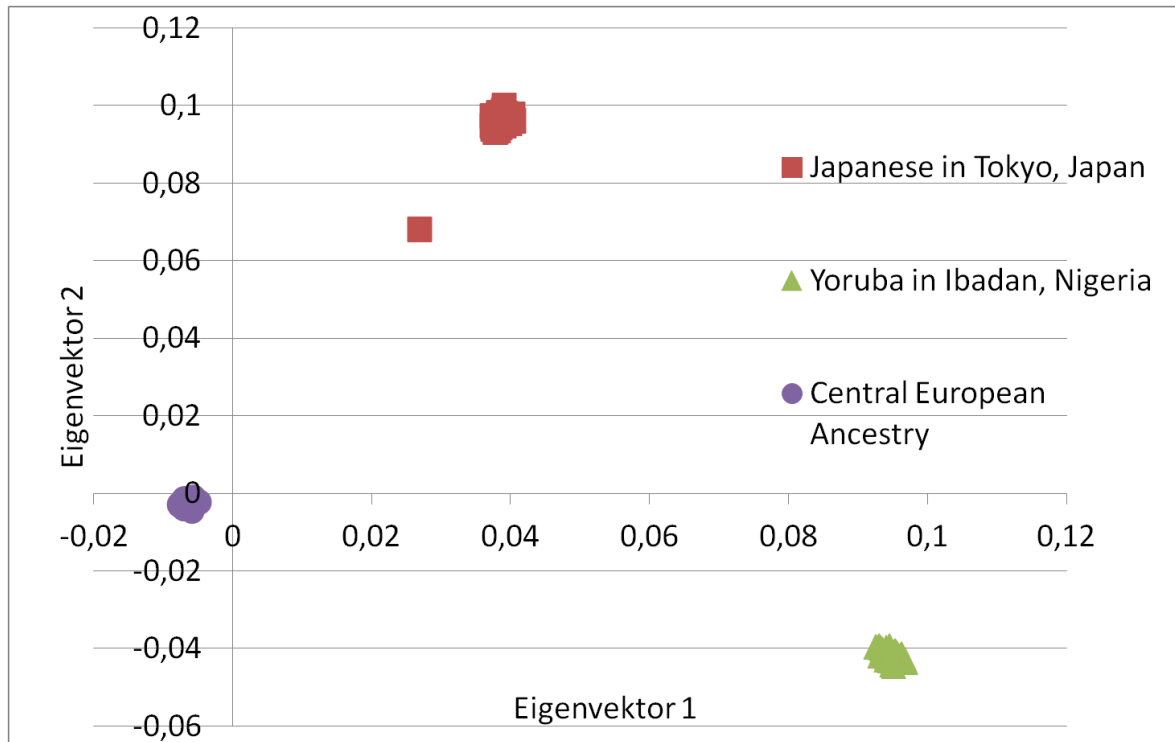


Abbildung 1 Darstellung der ersten beiden Eigenvektoren der HapMap (Thorisson et al., 2005) Vergleichspopulationen, in der erkennbar wird, dass sich die drei Populationen der Europäer, der Japaner und der Nigerianer deutlich voneinander abgrenzen

Als Vergleichspopulationen werden solche aus dem HapMap-Projekt verwendet, welches sich zum Ziel gemacht hat, die Haplotypen des humanen Genoms zu kartographieren. Forschungsgruppen aus verschiedenen Ländern stellen ihre Information kostenfrei zur Verfügung (Thorisson et al., 2005).

Über die Ermittlung von Eigenwerten und den zugehörigen Eigenvektoren werden die genotypischen Informationen strukturiert und in Clustern zusammengefasst. Die ersten beiden Eigenvektoren erklären dabei einen Großteil der Varianz. Daher wird die graphische Darstellung dieser beiden häufig als Darstellung der Populationsstruktur herangezogen (Novembre et al., 2008).

Methoden

Ausreißer eines genetischen Clusters werden als solche definiert, wenn sie sich mehr als fünf Standardabweichungen vom jeweiligen Clustermittelwert entfernen (Price et al., 2006).

Ein Maß dafür, ob *Population stratification* das Ergebnis einer GWAS beeinflusst, indem zum Beispiel zu viele falsch-positive Ergebnisse angezeigt werden, ist der *Genomic control inflation factor* λ_{GC} (Devlin et al., 1999). Ein $\lambda_{GC} \geq 1,1$ wird als Grenze genommen, ab der eine Analyse nochmals überprüft oder durch spezifische Adjustierungen angepasst werden sollte, beispielsweise indem die Hauptkomponenten als Kovariablen in das Regressionsmodell aufgenommen werden (Winkler et al., 2014).

Auf dem Probandenlevel wurden Daten aller Probanden entfernt, die unter 95 % genotypisierte Informationen hatten, nicht von zentraleuropäischer Abstammung waren, miteinander verwandt waren oder sich im dokumentierten Geschlecht vom genetischen Geschlecht unterschieden.

In der PCA wurden mit Hilfe der EIGENSTRAT Software (Price et al., 2006) die Eigenvektoren der genotypischen Daten ermittelt und Ausreißer, die mit mehr als fünf Standardabweichungen vom jeweiligen Gruppenmittelwert abwichen, entfernt. Die Darstellung der ersten beiden Eigenvektoren zusammen mit drei Vergleichspopulationen aus dem HapMap Projekt (Thorisson et al., 2005) sind in Anhang 8.2 in Abbildung 7 dargestellt.

Auf dem SNP-Level wurden alle Varianten entfernt, die in weniger als 95 % der Probanden genotypisiert wurden, vom Hardy-Weinberg-Gleichgewicht mit einem Grenzwert von 10^{-5} abwichen oder eine *Minor Allele Frequency* (MAF) unter 1 % hatten. Zwar sind die DNA-Chips so konzipiert, dass immer ein tagSNP gewählt wird, um einen Haplotypenblock zu repräsentieren, jedoch unterscheiden sich diese tagSNPs je nach Population unter anderem abhängig von der Ethnizität (Teo et al., 2010). Ein Pruning, also das Verringern der Anzahl von Markern, ist also dennoch sinnvoll, da bekannt ist, dass dies zu einer korrekten Detektion von Verwandtschaftsinformationen führt (Verma et al., 2014). In dieser Arbeit wurde das *Pruning* mit Hilfe der Software *plink* durchgeführt (Purcell et al., 2007). In einem Fenster von 50 SNPs wurde paarweise

Methoden

getestet, welches LD die SNPs zueinander aufweisen. Bei einem Wert $\geq 0,2$ wurde einer der beiden Partner entfernt, woraufhin sich das Fenster um 5 SNPs verschob und die Prozedur wiederholt wurde.

Ebenso wurden alle palindromischen SNPs entfernt. Für die hier durchgeführte GWAS selbst wäre es nicht notwendig gewesen, die palindromischen SNPs zu entfernen. Da jedoch die Datensätze intern auch für weitergehende Schritte, wie zum Beispiel dem *Prephasing* vor einer Imputation genutzt werden, ist dieser Schritt trotzdem durchgeführt worden. Für das *Prephasing* ist die Bestimmung der Haplotypen unabdingbar und damit muss die Leserichtung bestimmbar sein, was bei palindromischen SNPs nicht gegeben ist (Verma et al., 2014). Die Qualitätskontrollen wurden mit Hilfe von *plink* durchgeführt.

Nach diesen Qualitätskontrollschritten wurden die Informationen beider DNA-Chips zusammengefügt. Um möglichst viele Probandeninformationen pro SNP zu bekommen, wurden nur SNPs weiter betrachtet, die auf beiden verwendeten Chips genotypisiert wurden, das zum Beispiel dem Vorgehen in Chubb et al. (2013) entspricht. Auf diese Weise wurden möglichst viele Informationen pro Proband, aber auch die Informationen möglichst vieler Probanden der Studie verwendet, da die Genotypisierungsinformationen der beiden DNA-Chips, die derselben Chip-Familie entstammen, zusammengefügt wurden. Die zwei verwendeten Chips wurden dann mit Hilfe von *plink* im merge mode 1 zusammengefügt. Dies bedeutet, dass, sofern sich die Information eines SNPs auf den beiden DNA-Chips unterscheidet, keine der beiden Informationen genutzt wurde.

2.3 Genomweite Assoziationsstudie – GWAS

Die Genomweite Assoziationsstudie (GWAS) wurde mit Hilfe von *plink 1.07* durchgeführt. Als Phänotypen für die lineare Regression wurden die Serumkonzentrationen von κ und λ in logarithmierter Form verwendet, da sie eine schiefe Verteilung aufwiesen. Somit gleichen sie sich der Normalverteilung an und verletzen nicht die Bedingungen für die lineare Regression. Die Summe beider FLC wurde als Maß für die polyklonale Erhöhung verwendet. Für diese Analyse wurden alle Probanden entfernt, die eine MGUS, eine pFLCR oder eine andere Lymphoproliferative Erkrankung haben, um die monoklonale FLC-Erhöpfung auszuschließen. Die Summe aus κ - und λ -Konzentrationen wurde für die Analysen ebenfalls logarithmiert. Als Kovariablen wurden das Geschlecht, das Alter, C-reaktives Protein (CRP) und Kreatinin verwendet. CRP wurde als high-sensitive-CRP (hs-CRP) auf dem BN-II, Dade-Behring gemessen, Kreatinin wurde auf dem ADVIA1650 gemessen und beide Kovariablen als stetige Variable mit der Einheit $\frac{mg}{dL}$ ins Modell aufgenommen.

CRP dient als Entzündungsmarker (Burmeister et al., 2014) und Kreatinin spiegelt die Nierenfunktion wider, also die Fähigkeit FLC zu filtern. Dabei wird der biologische Zusammenhang der Kovariablen mit der FLC-Konzentration in den Mittelpunkt gestellt, um zu vermeiden, dass Störeinflüsse den eigentlichen Effekt der genetischen Varianten auf die FLC-Konzentrationen und der pFLCR verdecken.

Methoden

Die Modelle der linearen Regressionen sind in Formel 3 bis Formel 5 dargestellt.

Formel 3 Lineares Regressionsmodell mit der logarithmierten λ -Konzentration als Phänotyp und den Einflussvariablen der genetischen Varianten (0, 1, 2 bezeichnet die Anzahl der Risikoallele), des Alters, des Geschlechts, sowie der hs-CRP- und der Kreatinin-Konzentration

$$\log([\lambda]) \sim SNP \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} + \text{Alter} + \text{Geschlecht} \begin{pmatrix} 0 \\ 1 \end{pmatrix} + [\text{hs-CRP}] + [\text{Kreatinin}]$$

Formel 4 Lineares Regressionsmodell mit der logarithmierten κ -Konzentration als Phänotyp und den Einflussvariablen der genetischen Varianten (0, 1, 2 bezeichnet die Anzahl der Risikoallele), des Alters, des Geschlechts, sowie der hs-CRP- und der Kreatinin-Konzentration

$$\log([\kappa]) \sim SNP \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} + \text{Alter} + \text{Geschlecht} \begin{pmatrix} 0 \\ 1 \end{pmatrix} + [\text{hs-CRP}] + [\text{Kreatinin}]$$

Formel 5 Lineares Regressionsmodell mit der logarithmierten Summe der λ - und κ -Konzentration als Phänotyp und den Einflussvariablen der genetischen Varianten (0, 1, 2 bezeichnet die Anzahl der Risikoallele), des Alters, des Geschlechts, sowie der hs-CRP- und der Kreatinin-Konzentration

$$\log([\lambda + \kappa]) \sim SNP \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} + \text{Alter} + \text{Geschlecht} \begin{pmatrix} 0 \\ 1 \end{pmatrix} + [\text{hs-CRP}] + [\text{Kreatinin}]$$

Methoden

Für die binäre Variable der pathologischen $\kappa : \lambda$ -Ratio (pFLCR) wurde eine logistische Regression durchgeführt. Eine Erhöhung einer Freien Leichtkette zusammen mit einer pFLCR definiert die monoklonale FLC-Erhöpfung, im Gegensatz zur polyklonalen Erhöhung, bei der beide FLC erhöht sind und es zu keiner Verschiebung der Ratio kommt. Entsprechend gehen in diese Analyse alle Probanden mit ein, auch solche mit einer MGUS oder einem Multiplen Myelom. Dadurch, dass in der Definition von pFLCR über die geschätzte glomeruläre Filtrationsrate die Nierenfunktion schon mit einbezogen wird, werden nur das Alter und das Geschlecht als Kovariablen verwendet, nicht jedoch der Kreatinin-Wert. CRP wurde ebenso nicht als Kovariable berücksichtigt, da CRP als Entzündungsmarker zwar die einzelnen FLC-Konzentrationen beeinflussen kann, jedoch die pFLCR nicht verändert. Das Modell ist in Formel 6 dargestellt.

Formel 6 **Logistisches Regressionsmodell mit pFLCR als Phänotyp und den Einflussvariablen der genetischen Varianten (0, 1, 2 bezeichnet die Anzahl der Risikoallele), des Alters und des Geschlechts**

$$pFLCR \sim SNP \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} + \text{Alter} + \text{Geschlecht} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

2.4 Darstellung der GWAS-Ergebnisse

Um die Ergebnisse einer GWAS in einer Abbildung darstellen zu können, eignen sich Manhattan-Plots. Hierbei sind auf der X-Achse die Basenpositionen über alle Chromosomen aufgetragen und auf der Y-Achse der negative dekadische Logarithmus des p-Wertes aus der GWAS. Hier wurden diese Abbildungen mit der frei verfügbaren Software R und dem Paket qqman erstellt (Turner, 2014).

Zur Darstellung und zur Einordnung in den genetischen Kontext wurde LocusZoom (Pruim et al., 2010) verwendet. Dies ist ein Onlineprogramm, mit dem man regionale Darstellungen von GWAS-Ergebnissen erstellen kann. Hier wurden die Regionen 400 kb um den zu betrachtenden SNP näher dargestellt. Dabei wurden einerseits der negative dekadische Logarithmus der p-Werte der GWAS, sowie die Rekombinationsrate der SNPs untereinander in Centimorgan (cM) aufgetragen.

2.5 Geneset Enrichment Analyse – GSEA

Mit der Software *Meta-analysis gene-set enrichment of variaNT Associations* (MAGENTA) kann eine GSEA durchgeführt werden. Sie wurde von Ayellet Segrè am *Broad Institute of MIT and Harvard* entwickelt (Segrè et al., 2010). MAGENTA läuft in der Matlab Umgebung und kann für eine GSEA basierend auf p-Werten vorangegangener GWAS herangezogen werden.

Als Eingangsdatensatz dient eine Liste von Teststatistiken oder p-Werten der Assoziationsanalysen. Die SNPs werden mit ihren zugehörigen p-Werten den entsprechenden Genen zugeordnet. Dabei wird nicht nur der kodierende Bereich des Gens berücksichtigt, sondern auch ein bestimmter Bereich vor Transkriptionsstart und nach Transkriptionsende.

Im Anschluss erfolgen Korrekturen für mögliche Störvariablen. Typische Störvariablen sind zum Beispiel die Gengröße, die Anzahl der SNPs pro Gen oder pro Kilobase, der Genabstand oder auch die Abhängigkeit der SNPs untereinander. Die Werte der Störvariablen werden zuerst von MAGENTA pro Gen ermittelt, woraufhin eine schrittweise multiple lineare Regression zur Korrektur der p-Werte pro Gen durchgeführt wird. Auf diese Weise werden Effekte verschiedener physikalischer und genetischer Eigenschaften korrigiert.

Die adjustierten p-Werte pro Gen werden anschließend diesen Gensets zugeordnet. Gene, denen innerhalb der Gensets kein p-Wert zugeordnet wurde, werden ebenso entfernt wie Gene, die denselben SNP zugeordnet bekommen haben. Dies kann auftreten, wenn sich eine Upstream-Region eines Gens und die Downstream-Region eines anderen Gens überlappen. Die p-Werte innerhalb eines Gensets werden daraufhin der Größe nach sortiert.

Der p-Wert-Grenzwert wird aus der Gesamtverteilung der adjustierten p-Werte bestimmt, wobei zum Beispiel das 95. Perzentil der Verteilung als Grenzwert verwendet werden kann. Dieses Perzentil kann in MAGENTA als Parameter eingestellt werden.

Methoden

Üblicherweise werden die anschließenden Rechnungen mit dem 95. und dem 75. Perzentil durchgeführt.

Die Gene, die unterhalb dieser Grenze liegen, werden als *leading edge fraction* bezeichnet (Segrè et al., 2010). Diese *leading edge fraction* wird pro Genset bestimmt und der Anteil bezogen auf die Anzahl aller Gene im Genset berechnet. Für die Beurteilung, ob eine Anreicherung in einem bestimmten Genset vorliegt, wird eine Nullverteilung aus 10.000 zufällig permutierten Gensets identischer Größe generiert. Der nominale p-Wert für die Anreicherung ist der Anteil von zufällig erzeugten Gensets mit einer *leading edge fraction* gleich oder größer als die des betrachteten Gensets.

Ist der nominale p-Wert kleiner als 10^{-4} werden zur Sensitivitätsanalyse bis zu 1.000.000 Gensets permutiert. Als Ergebnis liefert MAGENTA einen nominalen, das heißt einen nicht für das multiple Testen korrigierten, *Genset-Anreicherungs* p-Wert sowie die Grenzen, die zum Beispiel für das 95. Perzentil benutzt wurden, ebenso wie die zugehörige FDR (Segrè et al., 2010).

Die p-Werte, die aus der GWAS wie im Kapitel 2.3 beschrieben resultierten, wurden zusammen mit den SNP-Informationen und Regressionskoeffizienten der linearen Regressionen, beziehungsweise Odds Ratios (OR) der logistischen Regression, in die Analysen der GSEA gegeben.

Die Geschlechtschromosomen, sowie die Informationen zu pseudoautosomalen Regionen des X-Chromosoms wurden entfernt, da sie fehlende Werte beinhalten aufgrund fehlender Genotypisierung auf den beiden hier verwendeten SNP-Chips. Die GSEA wurde mit allen in MAGENTA (Version 2, Juli 2011) hinterlegten Genset-Datenbanken durchgeführt. Hierzu zählt die Gene Ontology (GOTERM), die ihre Gensets anhand molekularer Aktivitäten der Genprodukte, entweder in zellulären Komponenten oder ganzen Signalwegen, zusammenstellt (The Gene Ontology). Die Panther-Datenbank beinhaltet Genfamilien, die jeweils einen phylogenetischen Stammbaum hinterlegt haben (Mi et al., 2009). Die *Kyoto Encyclopedia of Genes and Genomes* (KEGG) ist eine Zusammenstellung aus Signalwegen molekularer Interaktion und Reaktionsnetzwerken für Metabolismen, genetischer und umweltbezogener Informationen, zellulärer Prozesse,

Methoden

menschlicher Krankheiten und Medikamentenentwicklung (Kyoto Encyclopedia of Genes and Genomes, 2016). Die Ingenuity-Datenbank wird ebenfalls aus ontologischen Informationen zusammengestellt, hat ihre Hauptaufgabe aber im Feld der Krankheitserforschung und damit zusammenhängend die Ermittlung von Angriffspunkten für Medikamente (Felciano et al., 2013).

Die Reactome-Datenbank ist aus biologischen Signalwegen konstruiert, die unter anderem mit der KEGG- und der Gene Ontology-Datenbank validiert wurde (Reactome Database, 2016). BioCarta stellt Gensets zur Verfügung, die Geninteraktionen innerhalb von Signalwegen für zelluläre Prozesse wie zum Beispiel der Apoptose und der Signalweiterleitung darstellen (BioCarta LLC, San Diego, USA, 2016). Zusätzlich kann man eine Datenbank mit mitochondrialen Gensets (Pagliarini et al., 2008) hinzuziehen, was hier nicht anwendbar war, da keine mitochondriale DNA genotypisiert wurde. Diese Gensetinformatoren sind so in Tabellen hinterlegt, dass sie innerhalb der MAGENTA-Analyse eingelesen werden konnten.

Neben der Auswahl der Datenbanken mussten weitere Parameter angegeben werden, nach deren Werten verschiedene Analysen und Einstellungen gemacht wurden. Hier wurden die Gennamen den jeweiligen Gensets zugewiesen, HLA-Regionen entfernt und 10.000 permutierte Gensets generiert, um den Anreicherungs-p-Wert zu berechnen. Regionen, die 110.000 Basenpaare upstream oder 40.000 downstream liegen, wurden noch zu einem Gen gezählt. Wie groß dieses Genfenster im optimalen Fall sein soll, ist nicht genau geklärt, es sollte jedoch regulatorische Abschnitte mit einschließen (Holmans 2010). Basierend auf einer Studie über regulatorische Genelemente wurden die hier verwendeten Gengrenzen empfohlen (Veyrieras et al., 2008). Jedem Gen wurde so ein bestimmtes Set an SNPs zugeordnet. Aus diesem Set an SNPs wählt MAGENTA den SNP aus, der den niedrigsten p-Wert erreicht hat.

2.6 Eingesetzte Software

Zur Durchführung der GWAS wurde ein Linux basiertes Betriebssystem Ubuntu 12.04.2 LTS verwendet. Die Software *plink* Version 1.07 wurde für die Verarbeitung genetischer Daten und deren Analyse verwendet. Genauere Informationen zu *plink* finden sich bei (Purcell et al., 2007).

Für die Datenverarbeitung vor den genetischen Analysen, ebenso wie zur Anfertigung der Deskriptionen, wurde die Software SAS Version 9.4. verwendet (The SAS Institute, Cary, NJ). Matlab® kam in der Version R2013b zum Einsatz, MAGENTA in der Version 2, 2011. Die Manhattan Plots wurden mit der freiverfügbaren Software R Version 3.1.0 und dem R Paket qqman erstellt (Turner, 2014).

Die Hauptkomponentenanalyse wurde mit Hilfe des EIGENSTRAT Software durchgeführt (Price et al., 2010).

Als webbasierte Anwendung wurde der Genome-Browser des NCBI (<http://www.ncbi.nlm.nih.gov/genome/>) verwendet, um detaillierte Informationen zu Genen und deren Funktionen zu erhalten; ebenso wie das Onlineprogramm LocusZoom, das zur genaueren Betrachtung genetischer Regionen verwendet wurde (Pruim et al., 2010).

3 Ergebnisse

3.1 Beschreibung der Studienpopulation

Von den 4.814 Probanden der Heinz Nixdorf Recall Studie (HNR) gibt es mittlerweile genetische Informationen von 3.874 Probanden, wovon wiederum 3.832 komplette Informationen zu den in dieser Arbeit untersuchten Variablen haben. Auf den hier verwendeten Illumina® SNP-Chips sind 2.109 Probanden genotypisiert, wie in Abbildung 2 dargestellt.

Eine Übersicht der Studienpopulation gibt die Tabelle 1, in der nicht nur die Informationen der verwendeten SNP-Chips Human-Omni1-Quad und HumanOmniExpress aufgeführt sind, sondern zum Vergleich auch die Charakteristika der nicht genotypisierten Probanden. Zwei Probanden wurden aus Datenqualitätsgründen entfernt, da sie eine zu hohe beziehungsweise zu niedrige Heterozygotierate ($\geq 5 \cdot$ Standardabweichung) aufwiesen. Ein Proband wurde entfernt, da er – wie in Abbildung 7 (Anhang 8.2) erkennbar – eher dem Cluster der japanischen Population zuzuordnen wäre, obwohl die HNR-Kohorte zentraleuropäischer Abstammung ist.

Nach den Qualitätskontrollen befanden sich 779 Probanden mit Informationen zu 628.618 SNPs auf dem HumanOmni1-Quad und 1.348 Probanden mit 646.304 SNPs auf dem HumanOmniExpress. Nachdem diese beiden Chips vereint wurden entstanden ein Analysedatensatz mit 2.127 Probanden und 545.020 SNPs (Abbildung 3). Von diesen hatten 2.109 Probanden vollständige phänotypische Daten vorliegen.

Ergebnisse

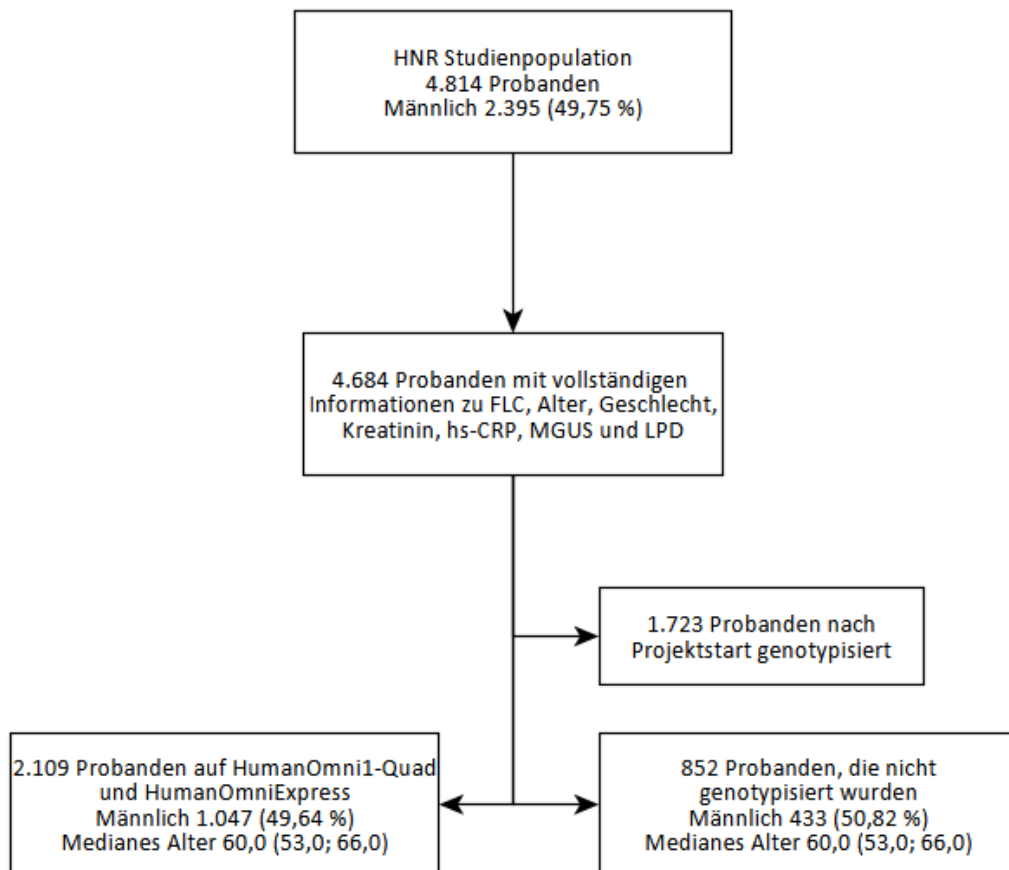


Abbildung 2 Studienübersicht der HNR-Studie aufgeteilt nach Genotypisierungen

Ergebnisse

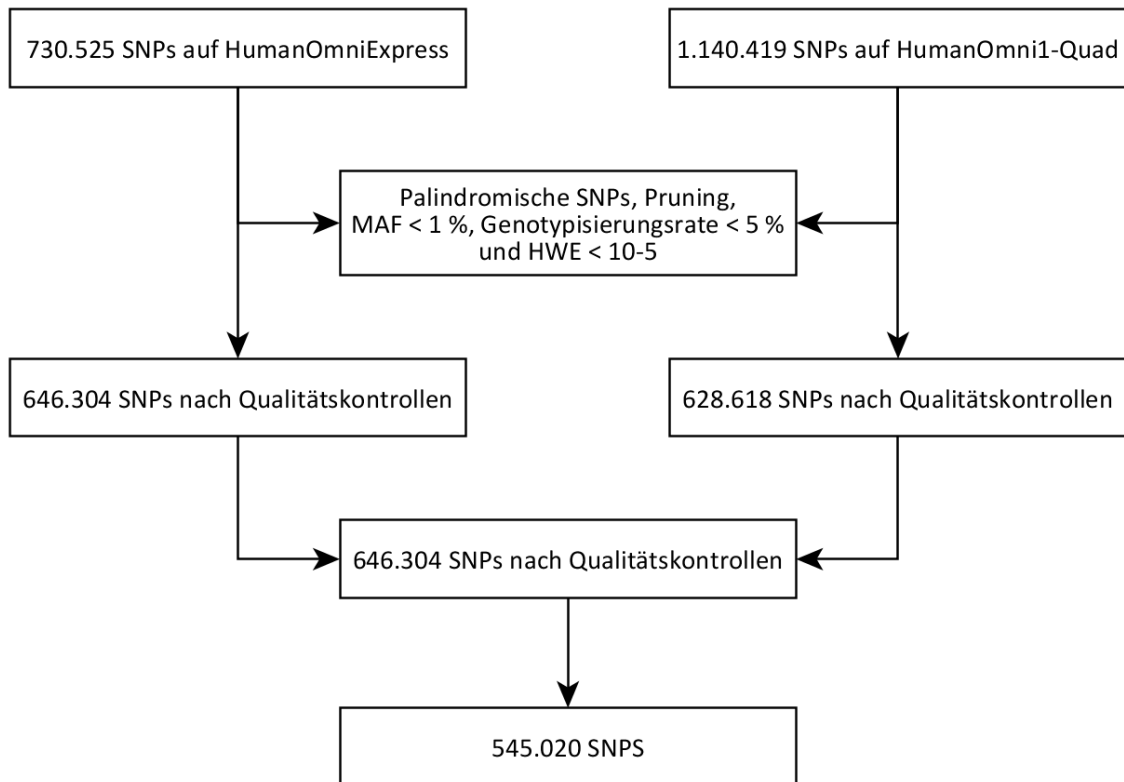


Abbildung 3 Übersicht der Qualitätskontrollschritte der zwei verwendeten DNA-Chips HumanOmni1-Quad und HumanOmniExpress, ebenso wie das Zusammenfügen beider Chips, um möglichst viele Probandeninformationen pro SNP zu erhalten

Ergebnisse

Tabelle 1 Beschreibung der HNR-Studienpopulation genotypisiert auf HumanOmni1-Quad und HumanOmniExpress zur Basiserhebung. Hierbei wurden nur vollständige Daten verwendet. Vergleichende Tests wurden je nach Skalenniveau mit dem exakten Test nach Fisher beziehungsweise dem Wilcoxon-Mann-Whitney Test durchgeführt

	HumanOmni1-Quad und HumanOmniExpress	Nicht genotypisiert	p-Wert
Gesamtanzahl	2.109	852	
Untersuchte Variablen			
κ - Median (q1; q3) [mg/L]	14,8 (12,3; 18,2)	15,1 (12,3; 18,4)	0,4014
λ - Median (q1; q3) [mg/L]	14,9 (12,1; 17,9)	15,0 (12,1; 18,4)	0,4807
tFLC - Median (q1; q3) [mg/L]	29,7 (24,8; 36,0)	30,1 (25,2; 36,7)	0,4456
pFLCR - N (%)	163 (7,7 %)	54 (6,3 %)	0,2126
Kovariablen			
Männlich- N (%)	1.047 (49,6 %)	433 (50,8 %)	0,5700
Alter - Median (q1; q3) [Jahren]	60,0 (53,0; 66,0)	60,0 (53,0; 66,0)	0,7319
LPD - N (%)	22 (1,0 %)	8 (0,9 %)	1,0000
MGUS - N (%)	123 (5,8 %)	43 (5,1 %)	0,4281
Kreatinin - Median (q1; q3) [mg/dL]	0,9 (0,8; 1,0)	0,9 (0,8; 1,0)	0,7719
hs - CRP Median (q1; q3) [mg/dL]	0,1 (0,1; 0,3)	0,2 (0,1; 0,3)	0,0040

q1: 1. Quartil

q3: 3. Quartil

N: Anzahl

3.2 Genomweite Assoziationsstudie – GWAS

Wie in Kapitel 2.3 beschrieben, wurde eine Genomweite Assoziationsstudie (GWAS) mit den Genotypen der Illumina® SNP-Chips HumanOmni1-Quad und HumanOmniExpress durchgeführt. Als abhängige Variablen wurden dabei in der linearen Regression die logarithmierten Werte der κ - und λ -Konzentration sowie deren Summe verwendet, in der logistischen Regression die dichotomisierten Werte in Form der pathologischen Freien Leichtketten-Ratio (pFLCR) als binäre Variable. Als Kovariablen wurden das Alter, das Geschlecht, sowie in der linearen Regression zusätzlich die hs-CRP- und die Kreatinin-konzentration verwendet. In Abbildung 4 sind die Ergebnisse der GWAS als Manhattan-Plots dargestellt. SNPs mit einem nominalen p-Wert kleiner als 10^{-5} sind in Tabelle 2 dargestellt, das heißt alle genetischen Varianten, die sich in Abbildung 4 oberhalb der horizontalen Linie befinden werden in Tabelle 2 im Detail aufgeführt.

In einer Sensitivitätsanalyse (Ergebnisse nicht explizit aufgeführt), in der zusätzlich auf die verwendeten SNP-Chip-Arrays adjustiert wurde, wurde keine Abweichung der Beta-schätzer in den Analysen zur polyklonalen FLC-Erhöhung gefunden. In den Analysen zur monoklonalen FLC-Erhöhung wurde eine Abweichung von $< 1\%$ gefunden.

Ergebnisse

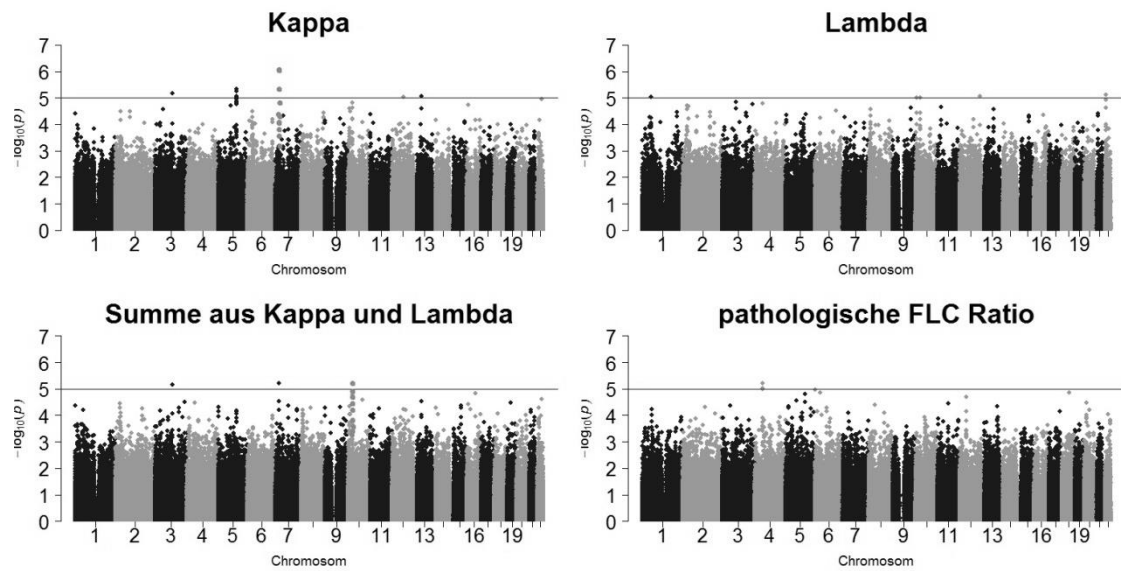


Abbildung 4 **Manhattan-Plots der GWAS-Ergebnisse der vier verschiedenen Analysen zur κ - und λ -Konzentration, deren Summe und der pathologischen FLC-Ratio**

Ergebnisse

Tabelle 2 Ergebnisse aus der GWAS, die mit Daten der HNR-Studie ($N_{\text{ges}} = 2.109$) zu Studienbeginn durchgeführt wurde. Es sind jene genetischen Varianten aufgeführt, die einen p-Wert $< 10^{-5}$ erreichten

Phänotyp	Chr	SNP ³	Position	Allel	Effektmaß ⁴	p-Wert
Serumkonzentration κ	3	<u>rs13063911</u>	106700416	G	1,0453	$6,93 \cdot 10^{-6}$
	5	rs988700	114284436	T	0,9485	$5,58 \cdot 10^{-6}$
	5	rs11749427	114289334	A	0,9535	$9,23 \cdot 10^{-6}$
	5	rs9326939	114294204	A	0,9534	$8,78 \cdot 10^{-6}$
	5	rs10055057	114383501	T	0,9481	$4,76 \cdot 10^{-6}$
	7	rs2042500	24075809	C	1,0519	$4,86 \cdot 10^{-6}$
	7	<u>rs10272500</u>	24076563	A	1,0572	$8,34 \cdot 10^{-7}$
	12	rs10879011	70101689	A	0,9521	$9,08 \cdot 10^{-6}$
	13	rs12584920	47465037	T	1,0549	$8,82 \cdot 10^{-6}$
Serumkonzentration λ	1	rs17785382	53707953	G	1,0438	$9,29 \cdot 10^{-6}$
	10	<u>rs12569910</u>	28995563	G	1,0435	$9,70 \cdot 10^{-6}$
	12	rs10847142	126882153	A	1,0537	$8,58 \cdot 10^{-6}$
	22	rs738791	24117525	C	1,0428	$7,56 \cdot 10^{-6}$
Summe aus den Serumkonzentrationen κ und λ	3	<u>rs13063911</u>	106700416	G	1,0420	$6,87 \cdot 10^{-6}$
	7	<u>rs10272500</u>	24076563	A	1,0487	$6,01 \cdot 10^{-6}$
	10	<u>rs12569910</u>	28995563	G	1,0413	$6,66 \cdot 10^{-6}$
Odds Ratios						
pathologische FLC Ratio	4	rs2289433	47839929	G	1,96	$6,32 \cdot 10^{-6}$
	4	rs2053404	47887536	A	1,932	$9,85 \cdot 10^{-6}$

Der λ_{GC} -Wert erreicht bei allen vier Analysen einen Wert von 1 mit den zugehörigen mittleren χ^2 -Statistiken von 0,99 für κ , 1,00 für λ , 1,00 für die Summe aus beiden FLC und 0,97 für die pathologische FLC-Ratio.

³ Unterstrichen sind Varianten, die in zwei Analysen mit dem jeweiligen Phänotyp assoziiert sind.

⁴ Die Effektmaße sind transformiert, also die Ergebnisse mit der Basis e exponiert, was aus dem Logarithmieren der ursprünglichen abhängigen Variablen resultiert. Hieraus ergibt sich, dass das Effektmaß den prozentualen Zuwachs pro Risikoallel beschreibt. Für die erste Zeile ergibt sich zum Beispiel eine Zunahme der κ -Konzentration um 4,53 % pro Risikoallel des rs13063911 SNPs. Ein Effektmaß kleiner als 1 beschreibt analog die prozentuale Abnahme pro Risikoallel.

3.3 LocusZoom

Wie in Kapitel 2.4 beschrieben, kann das Onlineprogramm LocusZoom verwendet werden, um genetische Regionen lokal genauer darzustellen. Die nähere Betrachtung der beiden genetischen Varianten rs2289433 und rs2053404 aus der GWAS zur pathologischen FLC-Erhöhung ist in Abbildung 5 zu finden. Diese SNPs liegen beide auf Chromosom 4 und zeigen mit Odds Ratios um 1,9 vergleichsweise hohe Effektstärken.

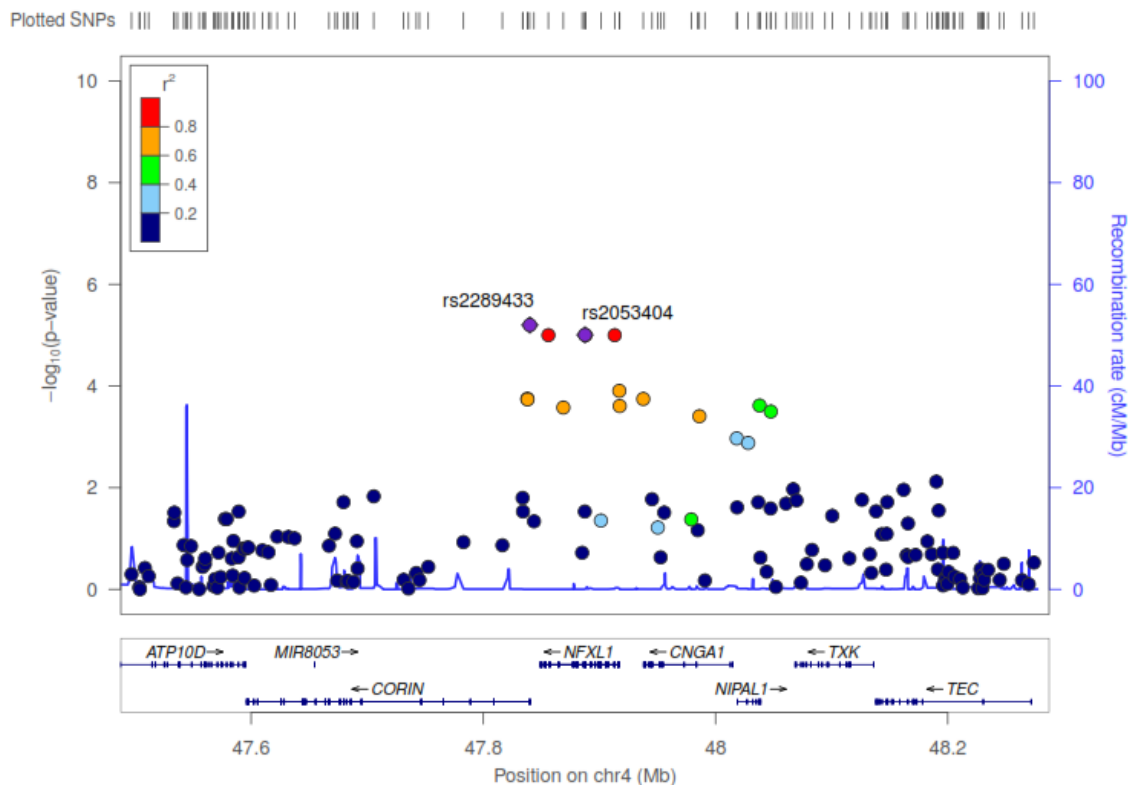


Abbildung 5 LocusZoom der genetischen Region um die zwei Varianten rs2289433 und rs2053404, die in der GWAS zu pFLCR assoziiert sind. In LocusZoom ist immer nur die Darstellung eines SNPs möglich. Die beiden SNPs sind daher für diese Darstellung graphisch zusammengefügt worden

3.4 Geneset Enrichment Analyse – GSEA

Wie in Kapitel 2.5 beschrieben, wurde die GSEA mit Hilfe des Programms MAGENTA in der Matlab-Umgebung ausgeführt. In Tabelle 3, Tabelle 7 und Tabelle 11 sind die Genesets dargestellt, die mit einer FDR < 5 % angereichert sind. In den fünf Spalten sind dabei die Datenbank, aus der das Geneset stammt, die Bezeichnung des Genesets und die im Geneset enthaltenen Gene dargestellt, ebenso wie der Anreicherungs-p-Wert und die *False Discovery Rate* (FDR). Wie auch zum Beispiel in Subramanian et al. (2005) empfohlen, werden die Genesets mit einer FDR < 25 % in Anhang 8.3 aufgelistet.

Tabelle 3 Genesets, aus den GSEA-Ergebnissen zu den logarithmierten κ -Werten, die mit einer FDR < 5 % angereichert sind

Datenbank	Geneset	p-Wert	FDR	Gennamen
Reactome	G Protein Activation	$3,00 \cdot 10^{-4}$	$3,11 \cdot 10^{-2}$	GNAI1, GNAI2, GNAI3, GNAO1, GNAT1, GNAT2, GNAZ, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, OPRM1, PDYN, POMC, GNB5, GNG13, GNG2, GNG12, GNB4, GNG8, GNAT3
Reactome	Inhibition of Insulin secretion by adrenaline noradrenaline	$4,00 \cdot 10^{-4}$	$3,31 \cdot 10^{-2}$	ADCY5, ADCY6, ADRA2A, ADRA2C, CACNA1C, CACNA1D, CACNB2, CACNB3, GNAI1, GNAI2, GNAO1, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, INS, CACNA2D2, GNB5, GNG13, GNG2, GNG12, GNB4, GNG8
GOTERM	Clathrin coated vesicle membrane	$1,07 \cdot 10^{-4}$	$3,89 \cdot 10^{-2}$	AP1B1, AP1G1, DAB2, HIP1, MYO6, AP3B2, AP3B1, AP1M1, HIP1R, AP1M2, DENND1A

Zur detaillierten Betrachtung sind in Tabelle 4 bis Tabelle 6 die drei Genesets aus den Analysen zu den logarithmierten κ -Werten genauer dargestellt, indem zu jedem Gen der ihm zugewiesene SNP mit dessen adjustiertem GWAS-p-Wert und dem Effektschätzer dargestellt ist.

Ergebnisse

Tabelle 4 Details des Gensets *G Protein Activation* mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu den logarithmierten κ -Werten⁵

Gene	SNP	⁶ Adjustierte GWAS-p-Werte	GWAS-Effektschätzer ⁷
G Protein Activation			
GNAI1	rs6948782	$6,85 \cdot 10^{-1}$	0,9819
GNAI2 ⁸	rs739983	$1,40 \cdot 10^{-5}$	0,9591
GNAI3	rs6537837	$8,74 \cdot 10^{-1}$	0,9834
GNAO1	rs3790112	$2,21 \cdot 10^{-1}$	0,9738
GNAT1 ⁸	rs739983	$1,64 \cdot 10^{-5}$	0,9591
GNAT2	rs668413	$6,14 \cdot 10^{-1}$	0,9816
GNAZ	rs6003461	$1,79 \cdot 10^{-3}$	1,0562
GNB1	rs6661597	$1,01 \cdot 10^{-1}$	0,9752
GNB2	rs221795	$5,31 \cdot 10^{-1}$	0,9829
GNB3	rs870849	$3,08 \cdot 10^{-1}$	1,0258
GNG3	rs693698	$3,50 \cdot 10^{-1}$	1,0214
GNG4	rs10159416	$4,44 \cdot 10^{-1}$	1,0249
GNG5	rs7524576	$9,26 \cdot 10^{-3}$	1,0541
GNG7	rs12973643	$2,15 \cdot 10^{-2}$	0,9642
GNG10	rs10817197	$7,66 \cdot 10^{-3}$	0,9661
GNG11	rs1541335	$2,90 \cdot 10^{-1}$	0,9641
GNGT1	rs1541335	$3,10 \cdot 10^{-1}$	0,9641
GNGT2	rs850522	$4,17 \cdot 10^{-1}$	1,0239
OPRM1	rs483481	$1,29 \cdot 10^{-1}$	0,9711
PDYN	rs16979691	$8,85 \cdot 10^{-1}$	0,9719
POMC	rs10460566	$4,88 \cdot 10^{-2}$	0,9687
GNB5	rs4776002	$7,86 \cdot 10^{-1}$	0,9849
GNG13	rs556179	$3,07 \cdot 10^{-2}$	1,0306
GNG2	rs1953866	$1,22 \cdot 10^{-1}$	0,9635
GNG12	rs344929	$1,85 \cdot 10^{-1}$	0,9545
GNB4	rs1345445	$1,03 \cdot 10^{-2}$	0,9680
GNG8	rs313840	$5,78 \cdot 10^{-2}$	0,9691
GNAT3	rs799964	$6,82 \cdot 10^{-1}$	1,0205

⁵ Die fett gedruckten Gene befinden sich auch im Genset *Inhibition of Insulin secretion by adrenaline noradrenaline* und sind dort ebenfalls markiert.

⁶ Im Zuge der GSEA werden die GWAS-p-Werte auf Störgrößen korrigiert. Aus den hier angegebenen adjustierten p-Werten wird die 95. Perzentilgrenze bestimmt, die als Grenzwert für die Anreicherungsanalyse fungiert.

⁷ GWAS-Effektschätzer sind analog zu Fußnote 4, Seite 42 transformiert.

⁸ Die markierten Gene gehören zur *leading edge fraction* (vgl. Kapitel 2.5), da sie einen adjustierten GWAS-p-Wert kleiner $6,39 \cdot 10^{-4}$ aufweisen

Ergebnisse

Tabelle 5 Details des Gensets *Inhibition of Insulin secretion by adrenaline noradrenaline* mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu den logarithmierten κ -Werten

Gene	SNP	Adjustierte GWAS-p-Werte	GWAS-Effektschätzer ⁹
<i>Inhibition of Insulin secretion by adrenaline noradrenaline</i>			
ADCY5	rs1055137	$6,26 \cdot 10^{-1}$	0,9775
ADCY6	rs2453469	$4,77 \cdot 10^{-1}$	1,0200
ADRA2A	rs7096359	$3,13 \cdot 10^{-1}$	0,9634
ADRA2C	rs16844725	$7,01 \cdot 10^{-1}$	0,9752
CACNA1C	rs2239091	$6,79 \cdot 10^{-1}$	1,0276
CACNA1D	rs3821846	$2,55 \cdot 10^{-1}$	1,0298
CACNB2	rs12219393	$6,33 \cdot 10^{-1}$	1,0445
CACNB3	rs2453469	$4,12 \cdot 10^{-1}$	1,0200
GNAI1	rs6948782	$6,85 \cdot 10^{-1}$	0,9819
GNAI2 ¹⁰	rs739983	$1,40 \cdot 10^{-5}$	0,9591
GNAO1	rs3790112	$2,21 \cdot 10^{-1}$	0,9738
GNB1	rs6661597	$1,01 \cdot 10^{-1}$	0,9752
GNB2	rs221795	$5,31 \cdot 10^{-1}$	0,9829
GNB3	rs870849	$3,08 \cdot 10^{-1}$	1,0258
GNG3	rs693698	$3,50 \cdot 10^{-1}$	1,0214
GNG4	rs10159416	$4,44 \cdot 10^{-1}$	1,0249
GNG5	rs7524576	$9,26 \cdot 10^{-3}$	1,0541
GNG7	rs12973643	$2,15 \cdot 10^{-2}$	0,9642
GNG10 ¹⁰	rs10817197	$7,66 \cdot 10^{-3}$	0,9661
GNG11	rs1541335	$2,90 \cdot 10^{-1}$	0,9641
GNGT1	rs1541335	$3,10 \cdot 10^{-1}$	0,9641
GNGT2	rs850522	$4,17 \cdot 10^{-1}$	1,0239
INS	rs10743181	$2,60 \cdot 10^{-2}$	0,9573
CACNA2D2	rs1540293	$4,29 \cdot 10^{-1}$	0,9695
GNB5	rs4776002	$7,86 \cdot 10^{-1}$	0,9849
GNG13	rs556179	$3,07 \cdot 10^{-2}$	1,0306
GNG2	rs1953866	$1,22 \cdot 10^{-1}$	0,9635
GNG12	rs344929	$1,85 \cdot 10^{-1}$	0,9545
GNB4	rs1345445	$1,03 \cdot 10^{-2}$	0,9680
GNG8	rs313840	$5,78 \cdot 10^{-2}$	0,9691

⁹ GWAS-Effektschätzer sind analog zu Fußnote 4, Seite 42 transformiert.

¹⁰ Die markierten Gene gehören zur *leading edge fraction* (vgl. Kapitel 2.5), da sie einen adjustierten GWAS-p-Wert kleiner $8,38 \cdot 10^{-3}$ aufweisen.

Ergebnisse

Tabelle 6 Details des Gensets *Clathrin coated vesicle membrane* mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu den logarithmierten κ -Werten

Gene	SNP	Adjustierte GWAS-p-Werte	GWAS-Effektschätzer ¹¹
<i>Clathrin coated vesicle membrane</i>			
AP1B1 ¹²	rs11090564	$1,15 \cdot 10^{-2}$	0,9684
AP1G1	rs1392487	$3,29 \cdot 10^{-2}$	0,9716
DAB2	rs4957476	$4,07 \cdot 10^{-1}$	1,0230
HIP1	rs3757598	$1,34 \cdot 10^{-2}$	1,0401
MYO6	rs7766781	$8,30 \cdot 10^{-1}$	0,9744
AP3B2	rs11635044	$1,68 \cdot 10^{-1}$	0,9551
AP3B1	rs4588572	$1,64 \cdot 10^{-2}$	0,9629
AP1M1	rs11086023	$5,94 \cdot 10^{-1}$	1,0198
HIP1R	rs1696320	$5,00 \cdot 10^{-2}$	1,0406
AP1M2	rs2569507	$7,19 \cdot 10^{-1}$	1,0155
DENND1A	rs969585	$3,45 \cdot 10^{-2}$	1,0334

¹¹ GWAS-Effektschätzer sind analog zu Fußnote 4, Seite 42 transformiert.

¹² Das markierte Gen gehört zur *leading edge fraction* (vgl. Kapitel 2.5), da es einen adjustierten GWAS-p-Wert kleiner $1,25 \cdot 10^{-2}$ aufweist.

Ergebnisse

Die in Tabelle 7 dargestellten Gensets sind in der GSEA-Analyse der λ -Werte angereichert und deren Details in Tabelle 8 bis Tabelle 10 dargestellt.

Tabelle 7 Gensets aus den GSEA-Ergebnissen zu den logarithmierten λ -Werten, die mit einer FDR < 5 % angereichert sind

Datenbank	Genset	p-Wert	FDR	Gennamen
Panther Molecular Functions	HMG box transcription factor	$1,19 \cdot 10^{-4}$	$1,14 \cdot 10^{-2}$	SOX1, SOX2, SOX3, SOX4, SOX5, SOX9, SOX10, SOX11, SOX15, SOX12, SRY, SSRP1, UBTF, SOX14, SOX13, TOX4, SOX30, SOX21, HBP1, TOX3, SOX8, SOX18, SOX6, SOX17, SOX7, ACRC, UBTF1
Panther Molecular Functions	DNA glycosylase	$2,70 \cdot 10^{-3}$	$3,77 \cdot 10^{-2}$	MPG, MUTYH, NTHL1, OGG1, TDG, UNG, CCNO, SMUG1, NEIL3, NEIL1, MCEE, NEIL2
Panther	Interleukin signaling pathway	$2,00 \cdot 10^{-4}$	$4,75 \cdot 10^{-2}$	CDKN1A, CDKN1B, CHUK, CSF2RB, ELK3, IL1A, IL2, IL2RA, IL2RB, IL4, IL4R, IL5, IL5RA, IL6, IL6R, IL6ST, IL7, CXCR1, CXCR2, IL10, IL10RA, IL10RB, IL11RA, IL12A, IL12RB1, IL12RB2, IL13, IL13RA1, IL13RA2, IL15, IL16, IL17A, IRS1, MAP3K3, MAP3K4, MAPK4, MAPK6,

Ergebnisse

Tabelle 8 Details des Gensets *HMG box transcription factor* mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu den logarithmierten λ -Werten

Gene	SNP	Adjustierte GWAS-p-Werte	GWAS-Effektschätzer ¹³
HMG box transcription factor			
SOX1	rs9549830	$7,46 \cdot 10^{-1}$	0,9778
SOX2	rs4434184	$3,62 \cdot 10^{-1}$	1,0288
SOX3	NaN ¹⁵	NaN	NaN
SOX4	rs9356777	$5,58 \cdot 10^{-1}$	1,0168
SOX5	rs7305773	$5,26 \cdot 10^{-1}$	0,9586
SOX9	rs17246175	$9,25 \cdot 10^{-1}$	1,0222
SOX10	rs8135665	$2,53 \cdot 10^{-2}$	1,0360
SOX11	rs17363466	$7,81 \cdot 10^{-1}$	1,0244
SOX15	rs4602096	$1,67 \cdot 10^{-2}$	1,0410
SOX12	rs6082956	$6,59 \cdot 10^{-1}$	0,9800
SRY	NaN	NaN	NaN
SSRP1 ¹⁴	rs2439451	$6,45 \cdot 10^{-4}$	1,0438
UBTF	rs9910055	$5,59 \cdot 10^{-1}$	0,9788
SOX14	rs1372733	$1,35 \cdot 10^{-1}$	0,9703
SOX13	rs7528560	$1,78 \cdot 10^{-2}$	0,9492
TOX4	rs8013500	$4,77 \cdot 10^{-3}$	0,9567
SOX30	rs13155976	$6,10 \cdot 10^{-1}$	0,9687
SOX21	rs9516466	$4,32 \cdot 10^{-1}$	1,0240
HBP1	rs2301801	$1,31 \cdot 10^{-2}$	0,9655
TOX3	rs2335	$2,16 \cdot 10^{-1}$	1,0341
SOX8	rs12595929	$5,94 \cdot 10^{-1}$	0,9778
SOX18	rs13039817	$3,72 \cdot 10^{-1}$	0,9790
SOX6	rs12360975	$6,02 \cdot 10^{-1}$	1,0469
SOX17	rs4738915	$3,62 \cdot 10^{-1}$	1,0297
SOX7 ¹⁴	rs4503064	$3,75 \cdot 10^{-3}$	1,0363
ACRC	NaN	NaN	NaN
UBTFL1	rs10830403	$9,53 \cdot 10^{-1}$	0,9909

¹³ GWAS-Effektschätzer sind analog zu Fußnote 4, Seite 42 transformiert.

¹⁴ Die markierten Gene gehören zur *leading edge fraction* (vgl. Kapitel 2.5), da sie einen adjustierten GWAS-p-Wert kleiner $3,90 \cdot 10^{-3}$ aufweisen.

¹⁵ NaN wird für die Gene ausgegeben, die in dem definierten Genfenster keine genetische Variante zugeordnet bekommen haben oder wenn zwei Genen – durch Überlappung der Downstream Region des einen mit der Upstream Region des anderen – derselbe SNP zugeordnet wurde (vgl. Kapitel 2.5).

Ergebnisse

Tabelle 9 Details des Gensets *DNA Glycosylase* mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu den logarithmierten λ -Werten

Gene	SNP	Adjustierte GWAS-p-Werte	GWAS-Effektschätzer ¹⁶
<i>DNA Glycosylase</i>			
MPG	rs12927713	$2,33 \cdot 10^{-1}$	0,9794
MUTYH	rs3219474	$7,71 \cdot 10^{-1}$	1,0249
NTHL1	rs8059880	$6,59 \cdot 10^{-1}$	0,9626
OGG1	rs787839	$3,24 \cdot 10^{-1}$	1,0263
TDG	rs12815289	$1,91 \cdot 10^{-1}$	1,0282
UNG	rs7305045	$1,41 \cdot 10^{-3}$	1,0478
CCNO	rs163098	$2,57 \cdot 10^{-1}$	1,0322
SMUG1	rs2124528	$1,60 \cdot 10^{-1}$	1,0358
NEIL3	rs17677777	$8,23 \cdot 10^{-1}$	1,0242
NEIL1	rs34879829	$3,39 \cdot 10^{-2}$	1,0434
MCEE	rs13016081	$1,60 \cdot 10^{-2}$	1,0439
NEIL2 ¹⁷	rs804267	$3,54 \cdot 10^{-4}$	0,9594

¹⁶ GWAS-Effektschätzer sind analog zu Fußnote 4, Seite 42 transformiert.

¹⁷ Das markierte Gen gehört zur *leading edge fraction* (vgl. Kapitel 2.5), da es einen adjustierten GWAS-p-Wert kleiner $9,35 \cdot 10^{-4}$ aufweist.

Ergebnisse

Tabelle 10 Details des Gensets *Interleukin signaling pathway* mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu den logarithmierten λ -Werten

Gene	SNP	Adjustierte GWAS-p-Werte	GWAS-Effektschätzer ¹⁸
<i>Interleukin signaling pathway</i>			
CDKN1A	rs183541	$7,16 \cdot 10^{-1}$	0,9735
CDKN1B	rs4555	$6,19 \cdot 10^{-1}$	1,0191
CHUK	rs12265333	$6,29 \cdot 10^{-1}$	0,9827
CSF2RB	rs4821540	$6,17 \cdot 10^{-1}$	1,0289
ELK3 ¹⁹	rs11108403	$9,55 \cdot 10^{-3}$	1,0464
IL1A	rs2853550	$9,85 \cdot 10^{-1}$	1,0144
IL2	rs10857092	$7,49 \cdot 10^{-1}$	0,9767
IL2RA	rs12722602	$8,98 \cdot 10^{-2}$	1,0545
IL2RB	rs11704001	$6,99 \cdot 10^{-1}$	0,9703
IL4	rs2243248	$5,79 \cdot 10^{-1}$	1,0313
IL4R	rs3024614	$8,28 \cdot 10^{-1}$	1,0361
IL5	rs2237060	$9,85 \cdot 10^{-1}$	0,9946
IL5RA	rs6808378	$7,50 \cdot 10^{-1}$	0,9743
IL6	rs2069827	$6,80 \cdot 10^{-2}$	1,0481
IL6R	rs3103309	$3,51 \cdot 10^{-1}$	0,9776
IL6ST	rs16884734	$2,10 \cdot 10^{-1}$	1,0330
IL7	rs4739111	$8,90 \cdot 10^{-1}$	1,0131
CXCR1	rs17462354	$6,32 \cdot 10^{-1}$	1,0301
CXCR2	rs876961	$3,20 \cdot 10^{-1}$	0,9783
IL10	rs4073250	$6,06 \cdot 10^{-1}$	0,9739
IL10RA	rs2243935	$6,49 \cdot 10^{-1}$	1,0255
IL10RB	rs2284550	$2,93 \cdot 10^{-1}$	0,9741
IL11RA ¹⁹	rs2772554	$3,99 \cdot 10^{-3}$	0,9586
IL12A	rs2886666	$8,84 \cdot 10^{-1}$	0,9789
IL12RB1 ¹⁹	rs3761041	$8,41 \cdot 10^{-3}$	1,0539
IL12RB2	rs3762314	$1,97 \cdot 10^{-1}$	0,9752
IL13	rs2243248	$5,66 \cdot 10^{-1}$	1,0313
IL13RA1	NaN	NaN	NaN
IL13RA2	NaN	NaN	NaN
IL15	rs4254850	$2,94 \cdot 10^{-1}$	1,0205
IL16	rs8034928	$1,33 \cdot 10^{-2}$	0,9618
IL17A	rs17578851	$3,62 \cdot 10^{-1}$	1,0321
IRS1	rs13386125	$7,44 \cdot 10^{-1}$	0,9731
MAP3K3	rs9944483	$1,30 \cdot 10^{-1}$	0,9753
MAP3K4	rs2277091	$9,59 \cdot 10^{-3}$	1,0358

¹⁸ GWAS-Effektschätzer sind analog zu Fußnote 4, Seite 42 transformiert.

¹⁹ Die markierten Gene gehören zur *leading edge fraction* (vgl. Kapitel 2.5), da sie einen adjustierten GWAS-p-Wert kleiner $9,58 \cdot 10^{-3}$ aufweisen.

Ergebnisse

MAPK4	rs1370480	$1,44 \cdot 10^{-1}$	0,9670
MAPK6	rs4776002	$8,00 \cdot 10^{-1}$	0,9867
MAPK7	rs739830	$8,73 \cdot 10^{-1}$	0,9879
SPI1	rs2290148	$8,60 \cdot 10^{-1}$	1,0152
SPIB	rs2293456	$2,13 \cdot 10^{-1}$	1,0264
MAP3K7	rs1231690	$3,49 \cdot 10^{-1}$	0,9737
IRS2	rs9521509	$2,38 \cdot 10^{-2}$	1,0327
MAD2L1BP	rs9472084	$2,97 \cdot 10^{-1}$	0,9598
MAP3K2	rs1568277	$9,07 \cdot 10^{-1}$	1,0148
IL17C	rs12446521	$1,22 \cdot 10^{-1}$	0,9465
IL17B	rs443402	$6,30 \cdot 10^{-1}$	1,0356
IFT81	rs2302689	$5,27 \cdot 10^{-1}$	0,9774
IL23A	rs3816804	$5,39 \cdot 10^{-2}$	0,9533
IL17D	rs9579937	$3,10 \cdot 10^{-2}$	1,0406
IL20RA	rs6936698	$7,60 \cdot 10^{-1}$	0,9774
SLC48A1	rs7135791	$4,30 \cdot 10^{-1}$	0,9703
TEX11	NaN	NaN	NaN
C21orf7	rs8133819	$7,53 \cdot 10^{-1}$	1,0221
IL21	rs2893008	$8,23 \cdot 10^{-1}$	1,0274
IL25	rs3210043	$3,18 \cdot 10^{-1}$	0,9663
YSK4	rs2874739	$4,23 \cdot 10^{-1}$	0,9748
SLA2	rs4812383	$1,11 \cdot 10^{-2}$	1,0421
IFT20	rs11080054	$3,45 \cdot 10^{-1}$	1,0211
IL17F	rs7750841	$2,50 \cdot 10^{-1}$	0,9733
SPIC	rs17500203	$3,47 \cdot 10^{-2}$	1,0356
MRPL10	rs2325751	$2,60 \cdot 10^{-1}$	1,0245
CCDC122	rs9533705	$2,69 \cdot 10^{-2}$	1,0438

Aus der GSEA mit der logarithmierten Summe beider FLC konnten keine Gensets mit einer FDR < 5 % ermittelt werden. Die Gensets mit einer FDR < 25 % sind in Anhang 8.3 dargestellt.

Ergebnisse

Die in Tabelle 11 dargestellten Gensets sind in der GSEA-Analyse der pathologischen Ratio angereichert und deren Details in Tabelle 12 und Tabelle 13 dargestellt.

Tabelle 11 Gensets aus den GSEA-Ergebnissen zu der pathologischen FLC-Ratio, die mit einer FDR < 5 % angereichert sind

Datenbank	Genset	p-Wert	FDR	Gennamen
BIOCARTA	EIF2 Pathway	$1,40 \cdot 10^{-3}$	$2,19 \cdot 10^{-2}$	EIF2S1, EIF2S3, EIF5, GSK3B, PPP1CA, EIF2AK2, EIF2B5, EIF2S2, EIF2AK3, EIF2AK1, EIF2AK4
BIOCARTA	Fibrinolysis Pathway	$8,00 \cdot 10^{-4}$	$3,72 \cdot 10^{-2}$	CPB2, F2, F2R, F13A1, FGA, FGB, FGG, SERPINE1, SERPINB2, PLAT, PLAU, PLG

Tabelle 12 Details des Gensets *EIF2 pathway* mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu der pathologischen FLC-Ratio

Gene	SNP	Adjustierte GWAS-p-Werte	Odds Ratios
EIF2 pathway			
EIF2S1	rs9743912	$2,58 \cdot 10^{-2}$	0,3291
EIF2S3	NaN	NaN	NaN
EIF5	rs7401414	$6,97 \cdot 10^{-1}$	0,6347
GSK3B	rs4261913	$6,01 \cdot 10^{-1}$	0,7423
PPP1CA	rs10896172	$4,78 \cdot 10^{-2}$	1,88
EIF2AK2	rs4648186	$5,73 \cdot 10^{-1}$	1,398
EIF2B5	rs12493550	$7,58 \cdot 10^{-1}$	0,5493
EIF2S2	rs1015363	$5,78 \cdot 10^{-1}$	1,257
EIF2AK3 ²⁰	rs13411467	$7,01 \cdot 10^{-4}$	1,852
EIF2AK1	rs2286680	$1,04 \cdot 10^{-2}$	0,3797
EIF2AK4	rs12442713	$3,80 \cdot 10^{-3}$	1,672

²⁰ Das markierte Gen gehört zur *leading edge fraction* (vgl. Kapitel 2.5), da es einen adjustierten GWAS-p-Wert kleiner $2,10 \cdot 10^{-3}$ aufweist.

Ergebnisse

Tabelle 13 Details des Gensets *Fibrinolysis Pathway* mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu der pathologischen FLC-Ratio

Gene	SNP	Adjustierte GWAS-p-Werte	Odds Ratios
Fibrinolysis Pathway			
CPB2	rs2296122	$4,07 \cdot 10^{-1}$	1,396
F2	rs3136456	$4,91 \cdot 10^{-1}$	0,5132
F2R	rs250728	$1,59 \cdot 10^{-2}$	0,5762
F13A1	rs381061	$1,87 \cdot 10^{-1}$	1,598
FGA	rs1044291	$2,11 \cdot 10^{-2}$	1,524
FGB	rs1044291	$4,38 \cdot 10^{-2}$	1,524
FGG	rs1044291	$2,19 \cdot 10^{-2}$	1,524
SERPINE1	rs1048365	$1,80 \cdot 10^{-2}$	1,694
SERPINB2	rs1243039	$7,49 \cdot 10^{-1}$	1,321
PLAT ²¹	rs2070713	$1,27 \cdot 10^{-2}$	0,6248
PLAU	rs2242258	$8,15 \cdot 10^{-1}$	0,784
PLG	rs4252117	$4,10 \cdot 10^{-1}$	0,6942

Bei den Gensets, die eine FDR $\geq 5\%$ und $< 25\%$ aufweisen, gibt es Überschneidungen bei den verschiedenen Analysen. Diese sind im Einzelnen:

- Chromatin chromatin-binding protein
- Opioid proopiomelanocortin pathway
- Non-receptor tyrosine protein kinase

Jeweils zwischen der κ -Analyse und der Summe aus κ - und λ -Analyse.

- P13 kinase pathway
- Interleukin signaling pathway
- DNA glycosylase
- Gamma-aminobutyric acid synthesis

²¹ Das markierte Gen gehört zur *leading edge fraction* (vgl. Kapitel 2.5), da es einen adjustierten GWAS-p-Wert kleiner $1,45 \cdot 10^{-2}$ aufweist.

Ergebnisse

Jeweils zwischen der λ -Analyse und der Summe aus κ - und λ -Analyse. Kein Genset wird in allen drei Analysen (κ -, λ - und der Summe aus κ - und λ -Analyse) mit einer FDR $< 25\%$ gefunden.

Bei den angereicherten Gensets mit einer FDR $\geq 5\%$ und $< 25\%$ für die Analysen mit pFLCR sind die Gensets *NFKB Signaling*, *VEGF Signaling* und der *CD40 pathway* wegen ihrer besonderen biologischen Funktion interessant (vgl. 8.3). NF- κ B ist ein Transkriptionsfaktor, der vor allem in der Immunantwort und der damit verbundenen Zellproliferation von großer Bedeutung ist (Karin et al., 2000). VEGF ist ein Wachstumsfaktor, der neben Bedeutung in der Vaskulogenese sowie der Angiogenese, die Migration von Monozyten und Makrophagen stimuliert (Yla-Herttuala et al., 2007). CD40 nimmt eine Rolle als Kostimulant ein und wird als Tumornekrosefaktor auf B-Zellen exprimiert (Elgueta et al., 2009).

4 Diskussion

In der vorliegenden Dissertationsschrift wurden genetische Effekte auf die Erhöhung von Freien Leichtketten im Serum untersucht. Dabei wurde einerseits die polyklonale Erhöhung, die mit Entzündungsprozessen einhergeht und andererseits die monoklonale Erhöhung, die ein Anzeichen hämatologischer Neoplasien sein kann, unterschieden. Ausgangspunkt für die Fragestellung waren Beobachtungen, dass die polyklonale Erhöhung mit einer erhöhten Gesamt- und kardialen Mortalität in der Allgemeinbevölkerung assoziiert ist (Dispenzieri et al., 2012; Eisele et al., 2010). Zudem ist eine pathologisch veränderte FLC-Ratio, das heißt eine monoklonale Erhöhung der FLC, mit einem erhöhten Progressionsrisiko bei MGUS assoziiert (Rajkumar et al., 2005). Die zugrundeliegenden biologischen Zusammenhänge sind jedoch wenig bekannt. Ein Ziel der vorliegenden Untersuchung war es deshalb, das Verständnis dieser Zusammenhänge zu erweitern. Für die Untersuchung der genetischen Effekte wurde eine GWAS mit genetischen Informationen von 2.109 Probanden der populationsbasierten Heinz Nixdorf Recall Studie durchgeführt, der eine GSEA nachgeschaltet wurde, um die Ergebnisse der GWAS zu aggregieren und so auch auf der Ebene funktionell zusammenhängender Gensets interpretieren zu können.

4.1 Untersuchte Population und Laboranalysen

In Kapitel 3.1 ist beschrieben, wie sich die hier untersuchte Population aus der Studienpopulation zusammensetzt.

Um etwaige Selektionseffekte zwischen den auf HumanOmni1-Quad und HumanOmni-Express genotypisierten und den nicht-genotypisierten Probanden zu ermitteln, wurden diese zwei Gruppen miteinander verglichen. Dabei wurden nur die Probanden verglichen, deren vollständige Daten zu Freien Leichtketten, den Kovariablen Alter, Geschlecht, Kreatinin und C-reaktivem Protein, ebenso wie zu den Ausschlusskriterien MGUS und Lymphoproliferativen Erkrankungen der Analysen zur polyklonalen FLC-Erhöpfung, vorlagen.

Hierbei zeigt sich, dass sich diese beiden Gruppen bei den betrachteten Merkmalen nur in der durchschnittlichen hs-CRP-Konzentration unterscheiden (vgl. Tabelle 1). Der Unterschied resultiert aus der höheren Konzentration der nicht genotypisierten Probanden. Medizinisch ist der Median beider Gruppen jedoch nicht als pathologisch einzuordnen (Pearson et al., 2003). Da CRP bei der polyklonalen FLC-Erhöpfung als Kovariable mit ins Modell aufgenommen wurde, wird hier für den Effekt – auch wenn er als zufällig betrachtet werden kann – adjustiert.

Da die Auswahl der Probanden, die genotypisiert wurden, nach keinem spezifischen Muster durchgeführt wurde, kann die Auswahl als zufällig betrachtet werden. Ebenso werden in den Sensitivitätsanalysen (vgl. 3.2) Effektunterschiede von < 1 % gefunden, was die Validität unterstützt.

Die Konzentrationen der Freien Leichtketten (FLC) sind über den Dade Behring BNII automatisierten Nephelometer (Siemens, Germany) mit Hilfe eines kommerziell verfügbaren Kits (FREELITE, The Binding Site Ltd, Birmingham, UK) unter Verwendung von hochspezifischen Antiseren gemessen worden. Die Präzision und Reproduzierbarkeit sind über CLSI Richtlinien (Clinical and Laboratory Standards Institute) geprüft

Diskussion

(The Binding Site Ltd, 2016). Die Bestimmungsgrenze liegt bei $0,6 - 120.000 \frac{mg}{L}$ bei κ - und $0,3 - 139.000 \frac{mg}{L}$ bei λ -Leichtketten. Die unteren Grenzen liegen somit etwa um den Faktor 10 unterhalb der normalen Bereiche der Freien Leichtketten im Serum mit κ von $3,3 - 19,4 \frac{mg}{L}$ und mit λ von $5,7 - 26,3 \frac{mg}{L}$ (Katzmann et al., 2002). Zusätzlich ist das FREELITE Kit für Freie Leichtketten so konzipiert, dass eine Vielzahl der hoch variablen FLC-Epitope mit Hilfe der polyklonalen Antikörper erkannt werden können (The Binding Site Ltd, 2016). Alle Laboranalysen wurden im Zentrallabor des Universitätsklinikums Essen mit demselben Analysegerät durchgeführt. Daher sind etwaige Verzerrungen, verursacht durch technische Gegebenheiten, eher auszuschließen und die Messungen als valide anzusehen.

4.2 Auswahl der Kovariablen

Die Auswahl der Kovariablen für die Regressionsmodelle erfolgte unter Beachtung der biologischen Zusammenhänge mit FLC, um potentielle Störfaktoren ausschließen zu können.

Um einen möglichst unverzerrten Einfluss der genetischen Varianten auf die FLC-Erhöpfung ermitteln zu können, wurde für verschiedene Variablen adjustiert. Das Alter und das Geschlecht wurden in das Modell aufgenommen, wie auch in Dispenzieri et al. (2012), die über die Altersabhängigkeit ebenso wie das Gesamtüberleben abhängig von FLC berichten. Zusätzlich wurden in den Analysen zur polyklonalen FLC-Erhöpfung jene Effekte mit aufgenommen, die die Entzündungsreaktionen oder die Nierenkapazität widerspiegeln und dadurch die Assoziationsergebnisse hätten beeinflussen können.

Das CRP gehört zum angeborenen Immunsystem und ist dabei im Zuge von Entzündungsreaktionen an der Vermittlung der Immunantwort beteiligt. FLC gehören zum adaptiven Immunsystem und werden ebenfalls bei Entzündungsreaktionen in erhöhtem Maße produziert. Durch die Herkunft aus unterschiedlichen Kompartimenten des Immunsystems sehen einige Autoren hier die Möglichkeit, dass sich CRP und FLC komplementär als Entzündungsmarker verhalten (Hutchison et al., 2011) und entsprechend eigenständige Marker für die Stimulation des jeweiligen Kompartiments sein könnten.

Um einen potentiellen komplementären Effekt der CRP-Erhöpfung zu kontrollieren, wurde deshalb bei der Analyse der polyklonalen FLC-Erhöpfung das CRP im statistischen Modell berücksichtigt. Die berechneten Effektmaße für die polyklonale FLC-Erhöpfung können so als unabhängig von der Stimulation des angeborenen Immunsystems angesehen werden.

Diskussion

Andererseits war eine Kontrolle für die Stimulation des angeborenen Immunsystems bei der Analyse der monoklonalen Erhöhung nicht notwendig, da Entzündungsreaktionen zwar die absoluten FLC-Konzentrationen im Serum ändern können, aber keinen Einfluss auf die FLC-Ratio haben.

Der Kreatininwert spiegelt die Nierenfunktion wider und somit die Fähigkeit, FLC aus dem Blut zu filtern. Eine erhöhte FLC-Konzentration kann folglich auch aus einer schlechteren Nierenkapazität resultieren. Um dies in den Analysen zur polyklonalen FLC-Erhö-
hung zu berücksichtigen, wurde der Kreatininwert in die Modelle aufgenommen. In den Analysen zur monoklonalen FLC-Erhö-
hung ist die Nierenfunktion bereits bei der Defini-
tion der pathologischen FLC-Ratio durch die von der eGFR abhängigen Referenzbereiche für die pathologische Ratio berücksichtigt. Ähnliche Adjustierungsansätze finden sich zum Beispiel in Johnson et al. (2016), die den genetischen Einfluss auf das Gesamtüber-
leben beim Multiplen Myelom untersuchten, ebenso wie bei Huang et al. (2013), die die genetische Regulation von Blutgerinnungsfaktoren in einer GWAS untersuchten.

Als Ausdruck der internen Validität können die übereinstimmenden Ergebnisse der Ana-
lysen angesehen werden. Unterstrichen sind in Tabelle 2 diejenigen SNPs, für die in meh-
reren Analysen eine Assoziation mit der polyklonalen FLC-Erhö-
hung gefunden wurde. Dabei fällt auf, dass die SNPs, die in der linearen Regression mit der Summe aus κ und λ ,
auch in den einzelnen Analysen assoziiert sind. Ebenso verhält es sich mit dem *Interleu-
kin signaling pathway*, der einerseits in der GSEA der λ -Serumkonzentrationen, anderer-
seits in der GSEA zur Summe beider Freien Leichtketten angereichert ist.

Diskussion

Zu diskutieren ist, ob es noch weitere Aspekte gibt, die mit einbezogen werden müssten. Heaney et al. (2016) beschreiben einen Zusammenhang der Höhe der FLC und dem Trainingseffekt eines Menschen. Es wäre somit denkbar, dass man die Lebensgewohnheiten einbeziehen muss. Die Vermutung dabei ist, dass basale Entzündungen seltener und schwächer auftreten, wenn ein gesünderer und agilerer Lebensstil geführt wird. In Heaney et al. (2016) wurde die RPE Skala (engl. *Rating of Perceived Exertion*, Bewertung des Anstrengungsgefühls) (Borg 1998) zur Ermittlung der wöchentlichen Bewegung der Probanden herangezogen. Da solch ein Instrument in der hier verwendeten Kohorte nicht zur Verfügung steht, konnte dieser Aspekt nicht in die Analysen einbezogen werden.

4.3 Genomweite Assoziationsstudie und Geneset Enrichment Analyse

In Kapitel 3.2 sind die einzelnen Ergebnisse der Genomweiten Assoziationsstudie (GWAS) beschrieben. Dargestellt werden alle genetischen Varianten mit einem p-Wert kleiner als 10^{-5} , auch wenn die für das multiple Testen über die Korrektur nach Bonferroni angepasste genomweite Signifikanzgrenze bei $\alpha = \frac{0,05}{545.020} = 9,2 \cdot 10^{-8}$ läge. Bei der

hier vorliegenden – zumindest für GWAS – geringen Populationsgröße waren im Verhältnis zu den typischerweise kleinen Effektgrößen bei genetischen Assoziationsstudien kaum geringere Signifikanzlevel zu erreichen. Häufig werden heutzutage daher Metaanalysen durchgeführt, die die Ergebnisse mehrerer Studien zusammenfassen und so das Probenkollektiv deutlich erhöhen. Als Beispiel ist hier das CHARGE (engl. *Cohorts for Heart and Aging Research in Genomic Epidemiology*) Konsortium zu nennen, das genetische Effekte auf Herz-Kreislaufkrankungen und Alterungsprozesse untersucht (Psaty et al., 2009). Damit können dann auch Varianten mit kleiner Effektgröße und hoher Variabilität aufgedeckt werden.

Die *Geneset Enrichment Analyse* (GSEA) stellt über die Möglichkeit der Konsortien hinweg eine Analyse dar, trotz fehlender genomweiter Signifikanz in einer GWAS, signifikante Ergebnisse auf Geneset-Ebene zu erhalten und so biologische Schlussfolgerungen zu ermöglichen. In einer GSEA werden SNPs Genen zugeordnet, die wiederum in biologische Signalwege zusammengefasst werden können, um so herauszufinden, ob es eine Anreicherung mehrerer genetischer Varianten aus einer GWAS in einem biologischen Kontext gibt. So können funktionell zusammenhängende Genesets ermittelt werden. Hier werden solche Genesets als angereichert im ersten Schritt definiert, die eine $FDR < 5\%$ aufweisen, um einen konservativeren und konzentrierteren Blick auf die Genesets zu werfen. Da es bei solchen Analysen häufig um die Generierung von Hypothesen geht, gibt es Veröffentlichungen, in denen auch weniger stringente FDR Schwellenwerte verwendet werden (Subramanian et al., 2005). Im Anhang werden deshalb auch Genesets mit einer $FDR < 25\%$ aufgeführt.

4.4 GWAS zur polyklonalen FLC-Erhöhung

Zu den drei Varianten rs13063911, rs10272500 und rs12569910, die in der GWAS mit der Summe aus beiden Freien Leichtketten assoziiert sind, konnten im Genome-Browser des NCBI (<http://www.ncbi.nlm.nih.gov/>) keine spezifischen Gene identifiziert werden. Außerdem erbrachte eine Literaturrecherche keine Publikationen mit Informationen zu diesen SNPs. Insgesamt kann so keine Aussage zu möglichen biologischen Zusammenhängen zwischen diesen Varianten und FLC-Konzentrationen gemacht werden.

4.5 Gensets aus den Analysen zur polyklonalen FLC-Erhöhung

Wie im Kapitel 3.4, den Ergebnissen der GSEA, dargestellt, ähneln sich die beiden in der Analyse zu den logarithmierten κ -Werten angereicherten Gensets *G Protein Activation* und *Inhibition of Insulin secretion by adrenaline noradrenalin* durch die große Anzahl gemeinsamer Gene, vor allem der verschiedenen G-Proteine. G-Proteine sind solche, die genregulatorische Wirkungen darüber erzielen, dass sie am Guanin der DNA binden können (Siffert et al., 1995). G-Proteine werden unter anderem zur Signalweiterleitung in B-Zell-Lymphoblasten exprimiert. Diese Expression ist bei Patienten mit Bluthochdruck besonders stark ausgeprägt, unabhängig, ob sie eine Lymphoproliferative Erkrankung aufweisen (Siffert et al., 1995). In Frey et al. (2014) wird eine Assoziation von *GNB3* innerhalb der HNR-Kohorte mit dem Auftreten von koronaren Ereignissen beschrieben. Als Verbindung zur polyklonalen FLC-Erhöhung und damit verbunden mit der Reaktion des Körpers auf kurzzeitige oder chronische Entzündungen, kann das Auftreten von koronarer Verkalkung gesehen werden. Inflammatorische Biomarker wie das CRP oder auch Interleukine sind mit dem Nachweis von asymptomatischem Koronarkalk und dem erhöhten Risiko des Auftretens koronarer Ereignisse assoziiert (Jenny et al., 2010; Möhlenkamp et al., 2011).

Der *Interleukin signaling pathway* ist einerseits in der Analyse der λ -Serumkonzentrationen, andererseits in der Analyse zur Summe beider Freien Leichtketten angereichert.

Interleukine fungieren als Botenstoffe zwischen Leukozyten, aber auch anderen an der Immunantwort beteiligten Zellen. Je nach Art des Interleukins haben sie höchst unterschiedliche, teils entgegengewirkende Aufgaben, wie zum Beispiel das Wachstum, die Hemmung oder auch die Differenzierung verschiedener Immunzellen. Interleukine sind den humoralen Bestandteilen des angeborenen Immunsystems zuzuordnen (Janeway et al., 1997).

Für die einzelnen SNPs, die den Genen im Interleukin-Signalweg zugeordnet wurden (vgl. Tabelle 10), findet sich zum Beispiel für die zwei Varianten rs2243248 und

Diskussion

rs3024614 eine Assoziation mit der Reaktivierung von Hepatitis B in Lymphom-Patienten (Hsiao et al., 2016). Die Variante rs8034928 ist assoziiert mit dem vermehrten Auftreten von kardiovaskulären Erkrankungen mit einer gleichzeitigen Erhöhung der IL-16 Konzentration, was in weiteren Arbeiten mit dem oben erwähnten Auftreten von kardialer Verkalkung untersucht werden sollte (Huang et al., 2013). rs2853550 trat vermehrt in rheumatisch bedingten Herzerkrankungen auf, was in einer Arbeit von Gupta et al. (2016) in einer indischen Population gezeigt wurde. Jedoch existiert hierzu keine Untersuchung in einer zentraleuropäischen Population.

Zusammenfassend zeigt sich, dass genetische Varianten, die in Gensets mit G-Proteinen angereichert sind, in Zusammenhang mit Entzündungsreaktionen und koronaren Herzerkrankungen stehen, die wiederum mit Interleukinen und CRP assoziiert sind. Im angereicherten Genset mit Interleukinen findet sich ebenso der Zusammenhang von genetischen Varianten und kardiovaskulären Erkrankungen.

Denkbar wäre hier also, dass in Übereinstimmung mit der oben dargestellten Sichtweise von CRP und FLC als komplementäre Entzündungsmarker, auch die polyklonale FLC Erhöhung – möglicherweise über ähnliche molekular-biologische Mechanismen – mit Koronarverkalkung und koronaren Ereignissen assoziiert ist.

4.5.1 GWAS zur monoklonalen FLC-Erhöhung

In Kapitel 3.2 sind die Ergebnisse der GWAS zur monoklonalen FLC-Erhöhung beschrieben.

Auffällig sind bei der logistischen Regression zur pFLCR die für eine GWAS relativ großen Odds Ratios (OR) von 1,96 für die Variante rs2289433 und 1,93 für die Variante rs2053404. Träger des entsprechenden Risikoallels haben damit eine fast zweifach erhöhte Chance, eine pathologische Ratio aufzuweisen. Hierbei wäre es denkbar, dass genau die Probanden, die an einer LPD leiden oder eine MGUS haben, häufiger das Risikoallel aufweisen; dies ist bei beiden genetischen Varianten nicht der Fall (Daten nicht explizit aufgeführt). Führt man eine lineare Regression mit der logarithmierten Ratio aus $\kappa : \lambda$ aus, erhält man Effektschätzer, die ähnliche Effektgrößen wie die lineare Regression der polyklonalen FLC-Erhöhung aufweisen (Daten nicht explizit aufgeführt).

Um die Regionen, in denen die SNPs liegen, genauer zu betrachten, wurde das Onlineprogramm LocusZoom verwendet. Die Darstellung dieser beiden Varianten findet sich in Abbildung 5. Wie hierin erkennbar, sind beide Varianten dem Gen *Corin* zugeordnet. Corin kodiert für Proteine, die im Zusammenhang mit atrialem natriuretischem Peptid stehen, was wiederum ein kardiales Hormon darstellt und so das Blutvolumen sowie den Blutdruck reguliert (<https://www.ncbi.nlm.nih.gov/gene/10699>). Über diesen Zusammenhang wäre es denkbar, dass Menschen mit pFLCR auch häufiger an kardialen Erkrankungen leiden, beziehungsweise in den kardiovaskulären Risikofaktoren, wie zum Beispiel dem B-type natriuretisches Peptid (BNP), erhöhte Werte aufweisen. In der hier vorliegenden Population liegen die durchschnittlichen BNP-Werte jedoch bei $27,8 \pm 26,2 \frac{\text{pg}}{\text{mL}}$ für die Probanden mit pFLCR und bei $26,6 \pm 29,5 \frac{\text{pg}}{\text{mL}}$ für die Probanden ohne pFLCR.

Tatsächlich wird in Kristinsson et al. (2009) eine erhöhte Sterblichkeit von MGUS-Patienten im Vergleich zur Normalbevölkerung beschrieben. Lymphoproliferative Erkrankungen gelten als Haupttodesursache. Darüber hinaus findet sich in MGUS-Patienten

Diskussion

eine erhöhte kardiale Sterblichkeit im Vergleich zur Normalbevölkerung, vor allem das Versterben an ischämischen Herzkrankheiten, Herzfehlern, Herzklappenfehlern, Kardiomyopathien und Arrhythmien. Als mögliche Ursache wird hier eine Kombination aus den pathogenen Mechanismen, die mit MGUS in Verbindung stehen, und vorhandenen Krankheiten oder Faktoren, die dazu geführt haben, dass auf MGUS getestet wurde, angesehen.

MGUS ist eine Vorstufe des Multiplen Myeloms – einer hämatologischen Neoplasie – bei der eine Verschiebung der FLC-Ratio in den pathologischen Bereich zu beobachten ist. Es ist somit denkbar, dass mit der pathologischen FLC-Ratio und den assoziierten genetischen Varianten auch eine erhöhte kardiale Sterblichkeit assoziiert ist.

4.5.2 Gensets aus den Analysen zur monoklonalen FLC-Erhöhung

In Kapitel 3.4 sind die Ergebnisse der GSEA zur monoklonalen FLC-Erhöhung dargestellt.

Eukaryotic Initiation Factor 2 ist ein Initiationsfaktor der Translation während der Proteinbiosynthese (Kimball, 1999). Die Fibrinolyse ist ein Prozess, in dem aggregierte Blutplättchen wieder aufgelöst werden und eine weitere Koagulation verhindert wird (Cesarman-Maus et al., 2005). Es sind somit beides Prozesse, die zunächst nicht unmittelbar in den Kontext der monoklonalen FLC-Erhöhung eingestuft werden können.

Jedoch findet sich der AMI-Signalweg unter den Ergebnissen mit einer FDR < 25 % im Anhang 8.3. AMI steht für *Acute myocardial infarction* (dt. Akuter Myokardinfarkt). In diesem Genset finden sich einerseits Gene, die als Koagulationsfaktoren im Blut fungieren, aber auch solche, die den Prozess der Fibrinolyse unterstützen und damit den Zusammenhang zum *Fibrinolysis*-Genset herstellen. In Tabelle 13 sind die detaillierten Ergebnisse dieses Gensets dargestellt. Erkennbar wird dort, dass SNPs, bei denen pro Allel die Chance steigt, dass eine pFLCR nachweisbar ist, Genen zugeordnet werden, die an der Bildung von Blutgerinnseln beteiligt sind (zum Beispiel der *Fibrinogen gamma chain* (FGG)). Hingegen SNPs, bei denen die Chance sinkt, eine pFLCR nachzuweisen, Genen zugeordnet werden, die mit der Auflösung solcher Gerinnsel in Verbindung stehen, wie zum Beispiel das PAT (Plasminogen Aktivator) sowie das Plasminogen (PLG) selbst. Dabei ist zu beachten, dass keine Aussage über die Expression des jeweiligen Gens getroffen werden kann. Weiterführend wäre hierbei also interessant, ob die Richtung der Expressionsstärke der assoziierten und angereicherten Gene mit der Richtung der Effektgröße der Assoziation der jeweiligen SNPs mit pFLCR übereinstimmt.

Diskussion

Hiermit lässt sich eine Verbindung zu den Ergebnissen der GWAS zur monoklonalen FLC-Erhöpfung finden – in der es Hinweise auf eine Assoziation der pathologischen FLC-Ratio und den assoziierten genetischen Varianten mit einer erhöhten kardialen Sterblichkeit gibt – worauf die Anreicherung im Genset zum Akuten Myokard Infarkt schließen lässt.

Drei weitere Gensets, die mit einer FDR < 25 % angereichert sind, können in den Zusammenhang mit dem Auftreten hämatologischer Neoplasien gebracht werden.

NF- κ B (engl. *nuclear factor 'kappa-light-chain-enhancer' of activated B-cells*) ist in nahezu allen Zellen im Zytosol zu finden. In B-Lymphozyten befindet es sich jedoch im Zellkern, wo es ohne äußere Einflüsse auf die Expression bestimmter Gene wirken kann. Es nimmt eine Schlüsselrolle nicht nur im Zuge der Immunantwort bei Entzündungen ein, sondern auch bei der Induktion von Apoptose zum Beispiel bei einer Zellentartung (Sen et al., 1986). Eine Stimulation erfolgt über Tumornekrosefaktoren oder Interleukine, aber auch eine direkte Stimulation von bakteriellen Proteinen ist möglich. Eine Hemmung seiner Wirkung ist durch Vitamin E-Derivate und Cortisol bekannt (Ahn et al., 2007).

VEGF (engl. *Vascular Endothelial Growth factor*) ist ein Wachstumsfaktor, der insbesondere die Neuanlage von Blutgefäßen während der Embryogenese, aber auch die Gefäßneubildung stimuliert (Yla-Herttuala et al., 2007). VEGF ist einer der Hauptfaktoren für die Metastasierung von Tumoren (Izzedine et al., 2007).

In hämatologischen Neoplasien steht VEGF in starkem Zusammenhang mit dem Risiko und der Aggressivität des Multiplen Myeloms. Hierbei spielen vor allem genetische Veränderungen der SNPs rs699947, rs1570360, rs2010963, rs2071559 und rs2305948 die Hauptrolle, die sich jedoch nicht unter den in dieser Arbeit dargestellten Varianten befinden. Der Vorgang der Gefäßneubildung ist ein Hauptaspekt für die Manifestierung eines Multiplen Myeloms (Brito et al., 2014).

Der CD40-Signalweg lässt sich in den Kontext der adaptiven Immunantwort einordnen. CD40 fungiert als Kostimulant und somit als Rezeptor der Familie der Tumornekrosefaktoren, der auf der Oberfläche von B-Zellen, Tumoren und Nicht-Immunzellen präsentiert

Diskussion

wird. Neben vielen Aufgaben nimmt CD40 eine Hauptrolle in der humoralen sowie der zellulären Immunantwort ein. Darüber hinaus ist aus Tumorzellen bekannt, dass sie sich das CD40-System zu Nutze machen, indem antiapoptotische Agentien hergestellt werden, die der Tumorzelle einen Schutz bieten und so ein schnelles Wachstum begünstigen (Elgueta et al., 2009).

In einer kürzlich erschienenen Arbeit von Law et al. (2017), die als Kontrollen Probanden der Heinz Nixdorf Recall Studie verwenden, wurde eine GWAS mit anschließender *Geneset Enrichment* Analyse zur chronischen lymphatischen Leukämie (CLL), dem Hodgkin Lymphom und dem Multiplen Myelom durchgeführt. Zwar finden sich in dieser Arbeit keine übereinstimmenden Gensets mit denen aus der GSEA der hier vorliegenden Arbeit, vermutlich da hier die pathologische FLC-Ratio als Phänotyp genutzt wurde und nicht eine definierte Ausprägung einer hämatologischen Neoplasie. Jedoch zeigen sich inhaltliche Übereinstimmungen in den angereicherten Gensets. Law et al. (2017) berichten acht Gensets, die mit Entzündungsreaktionen und Antigenprozessierung in Verbindung stehen, was sich mit den hier präsentierten Gensets inhaltlich deckt.

Zusammenfassend zeigt sich in der GSEA zur monoklonalen FLC-Erhöhung eine Anreicherung in Gensets, die mit kardiovaskulären Erkrankungen einhergehen. Dabei deckt sich dies vor allem mit den Ergebnissen der GWAS der monoklonalen Erhöhung, da auch dort eine Assoziation mit Corin – einem in Herzzellen produzierten und exprimierten Enzym – gefunden wird.

Da auch Gensets gefunden wurden, die mit Entzündungsreaktionen in Verbindung stehen, liegt die Vermutung nahe, dass der Effekt der koronaren Verkalkung, die damit verbundenen systemischen Entzündungsreaktionen, sowie die kardiovaskulären Ereignisse und Erkrankungen mit einer monoklonalen FLC-Erhöhung zusammenhängen.

Entzündungsprozesse scheinen also einer monoklonalen Erhöhung voran zu gehen, da zum Beispiel ähnliche molekulare Mechanismen eingebunden sind. Es ist bekannt, dass chronische Entzündungen vielen Neoplasien vorangehen. Eine Entzündung ist eine Reaktion des Immunsystems auf schädliche Reize, bei der Zellen des Immunsystems attrahiert werden und so die Auslöser des Reizes bekämpfen. In chronischen Entzündungen

Diskussion

hingegen nehmen Immunzellen teilweise entgegengesetzte Rollen ein. Dabei beschleunigen sie die Krebsentstehung sowie die Metastasierung unter anderem durch entstehende DNA-Schäden (Zhang et al., 2017). In Groot Kormelink et al. (2014) wird der Effekt von FLC auf die Krebspathologie solider Tumore untersucht. Dort wird vermutet, dass FLC tumorassoziierte Mastzellen aktivieren, die in Entzündungen nahe von Tumorgeweben vorkommen und dort das Tumorstadium anregen. Dieses Modell könnte so auf weitere Tumorarten ausgeweitet werden (Groot Kormelink et al., 2014).

Denkbar wäre, dass chronische Entzündungen, bei denen Plasmazellen ständig FLC produzieren, die Tumorneubildung sowie das Tumorstadium begünstigen. Einerseits könnten in Plasmazellen, die sich aufgrund der vielen FLC-Produktion häufig teilen, DNA-Fehler entstehen, was zu einer Malignität dieser Zelle führen kann. Andererseits wäre denkbar, dass in hämatologischen Neoplasien ähnliche Mechanismen wirken, wie in Groot Kormelink et al. (2014) beschrieben und so die vermehrte FLC-Produktion auch zur Tumorentwicklung und zum Tumorstadium in hämatologischen Krebserkrankungen beiträgt.

Ein weiteres internes Projekt geht in eine ähnliche Forschungsrichtung. Hier wurde der Effekt von Luftverschmutzungen im Rahmen von Feinstäuben auf die Entwicklung von MGUS, aber auch die Level von FLC untersucht. Feinstäube führen zu vermehrten Entzündungen, deren Effekt auf hämatologische Neoplasien weiter untersucht werden soll. Erste Ergebnisse zeigen, dass es eine Assoziation zwischen chronischer Exposition zu Feinstäuben und der Entwicklung einer MGUS gibt (Orban et al., 2017).

Diskussion

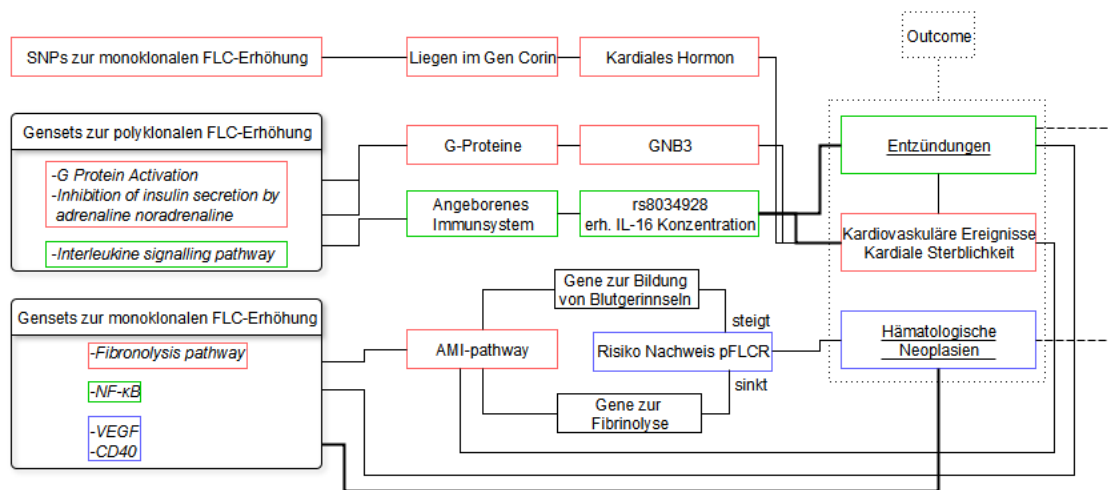


Abbildung 6 Zusammenfassende Übersicht der Ergebnisse

Abbildung 6 gibt den Überblick der Ergebnisse dieser Dissertationsschrift. Als Hauptaspekte dieser Arbeit sind die Assoziationen von monoklonaler und polyklonaler FLC-Erhöhung auf Entzündungen, kardiovaskuläre Ereignisse und kardiale Sterblichkeit, sowie hämatologische Neoplasien zu nennen. Die Zusammenhänge sind vor allem über die Anreicherung von SNPs der GWAS in Gensets gegeben, die mit den genannten Hauptaspekten in Verbindung stehen. So scheinen polyklonale FLC-Erhöhlungen nicht nur mit Entzündungen, sondern auch mit kardiovaskulären Ereignissen in Verbindung zu stehen. Ebenso wie monoklonale FLC-Erhöhlungen nicht nur einen direkten Zusammenhang zu hämatologischen Neoplasien haben, sondern ebenso mit kardiovaskulären Ereignissen sowie Entzündungen, die einer Neoplasie voran gehen könnten.

5 Ausblick

Die Ergebnisse der vorliegenden Arbeit können Ausgangspunkt für verschiedene weitere Untersuchungen sein.

Auf zwei weiteren DNA-Chips von Illumina® sind nochmals etwa 1.700 Probanden nach Beginn dieser Arbeit genotypisiert worden. Mit Hilfe dieser genetischen Informationen sollten die Analysen erweitert werden. In Vorversuchen zeigte sich, dass für die Varianten, die in dieser Arbeit gefunden wurden, etwa dieselben Betaschätzer der Analyse zur polyklonalen FLC-Erhöhung erreicht wurden und die Effekte in dieselbe Richtung zeigten, jedoch erreichte keine der Varianten ein Signifikanzlevel kleiner als $\alpha = 0,05$. Darüber hinaus wäre vor allem von Interesse, ob die GSEA zu vergleichbaren Ergebnissen führt, was in einem nächsten Schritt getestet werden sollte. Die Varianten der Analysen zur monoklonalen FLC-Erhöhung sind auf den neueren Chips nicht genotypisiert.

Die GSEA dient als Aggregationsanalyse der GWAS, aus der in dieser Analyse keine genomweit signifikanten Ergebnisse aufgrund der für GWAS geringen Fallzahl mit gleichzeitig kleinen Effekten genetischer Varianten zu erwarten sind. Eine weitere Möglichkeit die Power zu erhöhen ist, eine genetische Imputation durchzuführen. Bei der Imputation werden fehlende SNPs aller Probanden durch statistische Modelle mit Informationen aus Referenzgenomen auf Grundlage der Haplotypenstruktur hinzugefügt, das heißt die SNP-Dichte einzelner Probanden erhöht, ebenso wie gar nicht genotypisierte SNPs hinzugefügt (Marchini et al., 2010). Auch kann die Zahl der Fälle und der Kontrollen erhöht werden, indem sich verschiedene Studien in Konsortien zusammenschließen und ihre Ergebnisse in Metaanalysen zusammenfügen, was ebenfalls die Power einer GWAS erhöht (Cohn et al., 2003).

Ableitend aus den Analysen zur polyklonalen FLC-Erhöhung sollten weitere Analysen folgen, die den Zusammenhang von FLC und koronarer Verkalkung in den Fokus nehmen. In einer eigenen Arbeit, die noch unveröffentlicht ist, wird der Effekt von FLC auf kardiale Ereignisse untersucht. Hier wird auch die asymptomatische koronare Verkalkung

berücksichtigt, die ebenso einen Effekt auf FLC und auf kardiale Ereignisse hat. Die Untersuchung der asymptomatischen koronaren Verkalkung und ihre prognostische Bedeutung hinsichtlich kardialer Ereignisse war eines der Hauptziele der Heinz Nixdorf Recall Studie (Erbel et al., 2012; Schmermund et al., 2002). Zur Messung der asymptomatischen koronaren Verkalkung wurden bildgebende Verfahren verwendet und mittels des Kalkscores (coronary calcification score, CAC score) quantifiziert. Diese Daten stehen also für die Probanden der HNR-Studie zur Verfügung. In einzelnen Modellen kann nun überprüft werden, ob es eine Assoziation von FLC und dem Kalkscore gibt. Interessant wird dann auch sein zu untersuchen, ob genetische Varianten in den hier berichteten Genesets mit einem erhöhten Kalkscore assoziiert sind.

Darüber hinaus scheint auch die monoklonale FLC-Erhöpfung mit kardiovaskulären Ereignissen und damit verbunden auch mit dessen Risiken wie beispielsweise der koronaren Verkalkung in Verbindung zu stehen, da sich Anreicherungen in Genesets einerseits zu kardiovaskulären Erkrankungen und andererseits zu Entzündungsreaktionen zeigen. In der Arbeit von Law et al. (2017) wird die Anreicherung von genetischen Varianten, die assoziiert sind mit hämatologischen Neoplasien, in Genesets zu Entzündungsreaktionen und Antigenprozessierung beschrieben. Eine Analyse der hier verwendeten Daten mit definierten Ausprägungen von hämatologischen Neoplasien wie dem Multiplen Myelom und nicht nur der Aggregation in Form der pathologischen FLC-Ratio sollte folgen. Jaiswal et al. berichten eine Assoziation von *Clonal hematopoiesis of indeterminate potential* (CHIP), einer somatischen Mutation in Blut- oder Knochenmarkzellen ohne Nachweis einer hämatologischen Neoplasie, und dem Auftreten einer Atherosklerose in Mäusen sowie ein erhöhtes Risiko einer koronaren Herzerkrankung im Menschen. Eine Assoziation von CHIP mit dem erhöhten Risiko einer hämatologischen Neoplasie ist ebenfalls bekannt (Jaiswal et al., 2017).

Die Rolle von erhöhten FLC-Werten und der Entstehung von hämatologischen Neoplasien sollte weiter erforscht werden, um die zugrundeliegenden molekularen Mechanismen aufzudecken.

6 Zusammenfassung

Die polyklonale Erhöhung Freier Leichtketten (FLC) kann ein Anzeichen akuter oder chronischer Entzündungsreaktionen darstellen. Die monoklonale Erhöhung kann ein Anzeichen hämatologischer Krankheiten wie der Vorstufe der Monoklonalen Gammopathie unklarer Signifikanz (MGUS) oder Lymphoproliferativen Erkrankungen (LPD) sein. In einer Genomweiten Assoziationsstudie (GWAS) sowie einer nachgeschalteten *Geneset Enrichment Analyse* (GSEA) wurde die Assoziation und die Anreicherung von genetischen Varianten mit der monoklonalen und polyklonalen FLC-Erhöpfung ermittelt.

Bei der polyklonalen FLC-Erhöpfung zeigt sich in der GWAS keine Assoziation mit genetischen Varianten, die in den Kontext von systemischen Entzündungen eingestuft werden kann. In Gensets mit G-Proteinen und Interleukinen, sind jedoch genetische Varianten angereichert, die in Zusammenhang mit Entzündungsreaktionen und koronaren Herzerkrankungen stehen. Diese sind wiederum mit Interleukinen und CRP assoziiert. Da CRP und FLC komplementäre Entzündungsmarker sind, könnte die polyklonale FLC-Erhöpfung – zum Beispiel über ähnliche molekularbiologische Mechanismen – mit Koronarverkalkung und koronaren Ereignissen assoziiert sein. Die monoklonale FLC-Erhöpfung ist mit genetischen Varianten assoziiert, die sich in der Nähe des Genortes für Corin befinden. Corin ist wiederum mit einem kardialen Hormon assoziiert, das das Blutvolumen und den Blutdruck reguliert. In der GSEA spiegelt sich diese Verbindung wider, als dass genetische Varianten in Gensets angereichert sind, die mit Entzündungen und kardiovaskulären Ereignissen in Verbindung stehen (AMI-pathway, NF- κ B, VEGF und CD40-Signalweg). Einerseits zeigt sich hier die bekannte Verbindung von Entzündungen und koronarer Verkalkung mit kardiovaskulären Ereignissen über zum Beispiel der koronaren Verkalkung. Andererseits scheinen Entzündungen molekulare Mechanismen auszulösen oder zu nutzen, die auch in Verbindung mit hämatologischen Neoplasien stehen, sodass entzündliche Reaktionen auch einer hämatologischen Krebsentstehung voranzugehen scheinen.

7 Literaturverzeichnis

- [1] Ahn, K. S., Sethi, G., Krishnan, K., Aggarwal, B. B. (2007): Gamma-tocotrienol inhibits nuclear factor-kappaB signaling pathway through inhibition of receptor-interacting protein and TAK1 leading to suppression of antiapoptotic gene products and potentiation of apoptosis. *J Biol Chem.* 282 (1), 809–820.
- [2] Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., Zondervan, K. T. (2010): Data quality control in genetic case-control association studies. *Nat protoc.* 5 (9), 1564–1573.
- [3] Balducci, L. (2003): Myelosuppression and its consequences in elderly patients with cancer. *Oncology.* 17, 27–32.
- [4] Bence Jones, H. (1847): Papers on chemical pathology. *Lancet.* 50 (1247), 88–92.
- [5] Benjamini, Y.; Hochberg, Y. (1995): Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *J R Stat Soc.* 57 (1), 289–300.
- [6] Bickeböller, H., Fischer, C. (2007): Einführung in die Genetische Epidemiologie. Berlin, Heidelberg: Springer.
- [7] BioCarta LLC, San Diego, USA (2016): BioCarta. Online-Publikation; http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways, zuletzt geprüft am 01.07.2017.
- [8] Braber, S., Thio, M., Blokhuis, B. R., Henricks, P. A. J., Koelink, P. J., Groot Kormelink, T., Bezemer, G. F. G., Kerstjens, H. A. M., Postma, D. S., Garssen, J., Kraneveld, A. D., Redegeld, F. A., Folkerts, G. (2012): An Association between neutrophils and immunoglobulin free Light Chains in the pathogenesis of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 185 (8), 817–824.
- [9] Bradwell, A. R. (2010): Serum free Light Chain analysis (plus Hevylite). 6th ed. Birmingham: Binding Site.
- [10] Bradwell, A. R., Carr-Smith, H. D., Mead, G. P., Tang, L. X., Showell, P. J., Drayson, M. T., Drew, R. (2001): Highly sensitive, automated immunoassay for immunoglobulin free light chains in serum and urine. *Clin Chem.* 47 (4), 673–680.

- [11] Brebner, J. A.; Stockley, R. A. (2013): Polyclonal free light chains: a biomarker of inflammatory disease or treatment target?. *F1000 Med Rep.* 5:4.
- [12] Brito, A. B. C., Lourenco, G. J., Oliveira, G. B., Souza, C. A. de, Vassallo, J., Lima, C. S. P. (2014): Associations of VEGF and VEGFR2 polymorphisms with increased risk and aggressiveness of multiple myeloma. *Ann Hematol.* 93 (8), 1363–1369.
- [13] Broderick, P., Chubb, D., Johnson, D. C., Weinhold, N., Försti, A., Lloyd, A., Oliver, B., Ma, Y. P., Dobbins, S. E., Walker, B. A., Davies, F. E., Gregory, W. A., Child, J. A., Ross, F. M., Jackson, G. H., Neben, K., Jauch, A., Hoffmann, P., Mühleisen, T. W., Nöthen, M. M., Moebus, S., Tomlinson, I. P., Goldschmidt, H., Hemminki, K., Morgan, G. J., Houlston, R. S. (2011): Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat Genet.* 44 (1), 58–61.
- [14] Burmeister, A., Assi, L. K., Ferro, C. J., Hughes, R. G., Barnett, A. H., Bellary, S., Cockwell, P., Pratt, G., Hutchison, C. A. (2014): The relationship between high-sensitivity CRP and polyclonal free Light Chains as markers of inflammation in chronic disease. *Int J Lab Hematol.* 36 (4), 415–424.
- [15] Cesarman-Maus, G.; Hajjar, K. A. (2005): Molecular mechanisms of fibrinolysis. *Br J Haematol.* 129 (3), 307–321.
- [16] Charafeddine, K. M., Jabbour, M. N., Kadi, R. H., Daher, R. T. (2012): Extended use of serum Free Light Chain as a biomarker in lymphoproliferative disorders: A comprehensive review. *Am J Clin Pathol.* 137 (6), 890–897.
- [17] Chubb, D., Weinhold, N., Broderick, P., Chen, B., Johnson, D. C., Försti, A., Vijaykrishnan, J., Migliorini, G., Dobbins, S. E., Holroyd, A., Hose, D., Walker, B. A., Davies, F. E., Gregory, W. A., Jackson, G. H., Irving, J. A., Pratt, G., Fegan, C., Fenton, James A L, Neben, K., Hoffmann, P., Nöthen, M. M., Mühleisen, T. W., Eisele, L., Ross, F. M., Straka, C., Einsele, H., Langer, C., Dörner, E., Allan, J. M., Jauch, A., Morgan, G. J., Hemminki, K., Houlston, R. S., Goldschmidt, H. (2013): Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. *Nat Genet.* 45 (10), 1221–1225.
- [18] Clark, D. P. (2006): *Molecular Biology: das Original mit Übersetzungshilfen. Understanding the Genetic Revolution.* 1. Aufl. Heidelberg: Elsevier, Spektrum Akad. Verl.

- [19] Cohn, L. D.; Becker, B. J. (2003): How meta-analysis increases statistical power. *Psych meth.* 8 (3), 243–253. [20] Devlin, B.; Roeder, K. (1999): Genomic control for association studies. *Biomet-rics.* 55, 997–1004.
- [21] Dispenzieri, A., Katzmann, J. A., Kyle, R. A., Larson, D. R., Melton, L. J., Colby, C. L., Therneau, T. M., Clark, R., Kumar, S. K., Bradwell, A., Fonseca, R., Jelinek, D. F., Rajkumar, S. V. (2010): Prevalence and risk of progression of light-chain monoclonal gammopathy of undetermined significance. A retrospective population-based cohort study. *Lancet.* 375 (9727), 1721–1728.
- [22] Dispenzieri, A., Katzmann, J. A., Kyle, R. A., Larson, D. R., Therneau, T. M., Colby, C. L., Clark, R. J., Mead, G. P., Kumar, S., Melton, L. J., Rajkumar, S. V. (2012): Use of nonclonal serum immunoglobulin free Light Chains to predict overall survival in the general population. *Mayo Clin Proc.* 87 (6), 517–523.
- [23] Dispenzieri, A., Kyle, R., Merlini, G., Miguel, J. S., Ludwig, H., Hajek, R., Palumbo, A., Jagannath, S., Blade, J., Lonial, S., Dimopoulos, M., Comenzo, R., Einsele, H., Barlogie, B., Anderson, K., Gertz, M., Harousseau, J. L., Attal, M., Tosi, P., Sonneveld, P., Boccadoro, M., Morgan, G., Richardson, P., Sezer, O., Mateos, M. V., Cavo, M., Joshua, D., Turesson, I., Chen, W., Shimizu, K., Powles, R., Rajkumar, S. V., Durie, B G M (2009): International Myeloma Working Group guidelines for serum-free light chain analysis in multiple myeloma and related disorders. *Leukemia.* 23 (2), 215–224.
- [24] Draborg, A. H., Lydolph, M. C., Westergaard, M., Olesen Larsen, S., Nielsen, C. T., Duus, K., Jacobsen, S., Houen, G., Assassi, S. (2015): Elevated concentrations of serum immunoglobulin free Light Chains in systemic lupus erythematosus patients in relation to disease activity, inflammatory status, B cell activity and Epstein-Barr Virus Antibodies. *PLoS One.* 10 (9), e0138753.
- [25] Eisele, L., Dürig, J., Huttmann, A., Dührsen, U., Führer, A., Kieruzel, S., Bröcker-Preuss, M., Bokhof, B., Erbel, R., Mann, K., Jöckel, K.-H., Moebus, S. (2010): Polyclonal free Light Chain elevation and mortality in the German Heinz Nixdorf Recall study. In: American Society of Hematology (Hg.): 52nd ASH Annual Meeting Abstracts. ASH. Orlando, 04.-07.12.2010: *blood* (116).

- [26] Eisele, L., Dürig, J., Hüttmann, A., Dührsen, U., Assert, R., Bokhof, B., Erbel, R., Mann, K., Jöckel, K.-H., Moebus, S. (2012): Prevalence and progression of monoclonal gammopathy of undetermined significance and light-chain MGUS in Germany. *Ann Hematol.* 91 (2), 243–248.
- [27] Elgueta, R., Benson, M. J., Vries, V. C. de, Wasiuk, A., Guo, Y., Noelle, R. J. (2009): Molecular mechanism and function of CD40/CD40L engagement in the immune system. *Immunol Rev.* 229 (1).
- [28] Erbel, R., Eisele, L., Moebus, S., Dragano, N., Möhlenkamp, S., Bauer, M., Kälisch, H., Jöckel, K.-H. (2012): The Heinz Nixdorf Recall study. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz.* 55 (6-7), 809–815.
- [29] Felciano, R. M., Bavari, S., Richards, D. R., Billaud, J.-N., Warren, T., Panchal, R., Kramer, A. (2013): Predictive systems biology approach to broad-spectrum, host-directed drug target discovery in infectious diseases. *Pac Symp Biocomput.*, 17–28.
- [30] Frey, U. H., Moebus, S., Möhlenkamp, S., Kälisch, H., Bauer, M., Lehmann, N., Nöthen, M., Mühleisen, T. W., Stang, A., Erbel, R., Jöckel, K.-H., Peters, J., Siffert, W. (2014): GNB3 gene 825 TT variant predicts hard coronary events in the population-based Heinz Nixdorf Recall study. *Atherosclerosis.* 237 (2), 437–442.
- [31] Geffken, D. F. (2001): Association between physical activity and markers of inflammation in a healthy elderly population. *Am J Epidemiol.* 153 (3), 242–250.
- [32] Greenberg, A. J., Lee, A. M., Serie, D. J., McDonnell, S. K., Cerhan, J. R., Liebow, M., Larson, D. R., Colby, C. L., Norman, A. D., Kyle, R. A., Kumar, S., Rajkumar, S. V., Diasio, R. B., Slager, S. L., Vachon, C. M. (2012): Single-nucleotide polymorphism rs1052501 associated with monoclonal gammopathy of undetermined significance and multiple myeloma. *Leukemia.* 27 (2), 515–516.
- [33] Groot Kormelink, T., Powe, D. G., Kuijpers, S. A., Abudukelimu, A., Fens, M. H. A. M., Pieters, E. H. E., van der Kassing Ven, W. W., Habashy, H. O., Ellis, I. O., Blokhuis, B. R., Thio, M., Hennink, W. E., Storm, G., Redegeld, F. A., Schiffelers, R. M. (2014): Immunoglobulin free light chains are biomarkers of poor prognosis in basal-like breast cancer and are potential targets in tumor-associated inflammation. *Oncotarget.* 5 (10), 3159–3167.

- [34] Gupta, U., Mir, S. S., Garg, N., Agarwal, S. K., Pande, S., Mittal, B. (2016): Association study of inflammatory genes with rheumatic heart disease in North Indian population: A multi-analytical approach. *Immunol Lett.* 174, 53–62.
- [35] Heaney, J. L., Phillips, A. C., Drayson, M. T., Campbell, J. P. (2016): Serum free light chains are reduced in endurance trained older adults. Evidence that exercise training may reduce basal inflammation in older adults. *Exp Gerontol.* 77, 69–75.
- [36] Holden, M., Deng, S., Wojnowski, L., Kulle, B. (2008): GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies. *Bioinformatics.* 24 (23), 2784–2785.
- [37] Hsiao, L.-T., Wang, H.-Y., Yang, C.-F., Chiou, T.-J., Gau, J.-P., Yu, Y.-B., Liu, H.-L., Chang, W.-C., Chen, P.-M., Tzeng, C.-H., Chan, Y.-J., Yang, M.-H., Liu, J.-H., Huang, Y.-H. (2016): Human cytokine genetic variants associated with HBsAg reverse seroconversion in Rituximab-treated Non-Hodgkin Lymphoma patients. *Medicine.* 95 (11), e3064.
- [38] Huang, H., Zeng, Z., Zhang, L., Liu, R., Li, X., Qiang, O., Zhang, Q., Chen, Y. (2013): The association of interleukin-16 gene polymorphisms with susceptibility of coronary artery disease. *Clin Biochem.* 46 (3), 241–244.
- [39] Hutchison, A. T., Jones, D. R., Raison, R. L. (2012): The ability to interact with cell membranes suggests possible biological roles for free light chain. *Immunol Lett.* 142 (1-2), 75–77.
- [40] Hutchison, C. A.; Landgren, O. (2011): Polyclonal Immunoglobulin Free Light Chains as a potential biomarker of immune stimulation and inflammation. *Clin Chem.* 57 (10), 1387–1389.
- [41] Hutchison, C. A., Plant, T., Drayson, M., Cockwell, P., Kountouri, M., Basnayake, K., Harding, S., Bradwell, A. R., Mead, G. (2008): Serum free light chain measurement aids the diagnosis of myeloma in patients with severe renal failure. *BMC Nephrol.* 9 (11), 1–8.
- [42] Illumina® (2010): The Omni family of microarrays. Online-Publikation; <http://www.illumina.com/applications/genotyping/human-genotyping-arrays/omni-arrays.ilmn>, zuletzt geprüft am 01.07.2017.

- [43] Izzedine, H., Rixe, O., Billemont, B., Baumelou, A., Deray, G. (2007): Angiogenesis inhibitor therapies: focus on kidney toxicity and hypertension. *Am J Kidney Dis.* 50 (2), 203–218.
- [44] Jaiswal, S., Natarajan, P., Silver, A.J., Gibson, C.J., Bick, A.G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., Ardissino, D., Baber, U., Mehran, R., Fuster, V., Danesh, J., Frossard, P., Saleheen, D., Melander, O., Sukhova, G.K., Neuberg, D., Libby, P., Kathiresan, S., Ebert, B.L. (2017): Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med* 377 (2), 111–121.
- [45] Janeway, C., Travers, P. (1997): *Immunologie*. 2. Aufl. Heidelberg, Berlin, Oxford: Spektrum, Akad. Verl.
- [46] Jenny, N. S., Brown, E. R., Detrano, R., Folsom, A. R., Saad, M. F., Shea, S., Szklo, M., Herrington, D. M., Jacobs, D. R., JR (2010): Associations of inflammatory markers with coronary artery calcification: results from the multi-ethnic study of atherosclerosis. *Atherosclerosis.* 209 (1), 226–229.
- [47] Johnson, D. C., Weinhold, N., Mitchell, J. S., Chen, B., Kaiser, M., Begum, D. B., Hillengass, J., Bertsch, U., Gregory, W. A., Cairns, D., Jackson, G. H., Forsti, A., Nickel, J., Hoffmann, P., Nöthen, M. M., Stephens, O. W., Barlogie, B., Davis, F. E., Hemminki, K., Goldschmidt, H., Houlston, R. S., Morgan, G. J. (2016): Genome-wide association study identifies variation at 6q25.1 associated with survival in multiple myeloma. *Nat Commun.* 7.
- [48] Karin, M.; Ben-Neriah, Y. (2000): Phosphorylation meets ubiquitination: the control of NF-kappaB activity. *Annual review of immunology.* 18, 621–663.
- [49] Katzmann, J. A., Clark, R. J., Abraham, R. S., Bryant, S., Lymp, J. F., Bradwell, A. R., Kyle, R. A. (2002): Serum reference intervals and diagnostic ranges for free and free immunoglobulin Light Chains: Relative sensitivity for detection of monoclonal Light Chains. *Clin Chem.* 48 (9), 1437–1444.
- [50] Kimball, S. R. (1999): Eukaryotic initiation factor eIF2. *Int J Biochem Cell Biol.* 31 (1), 25–29
- [51] Knippers, R. (1997): *Molekulare Genetik*. 7., durchges. und korr. Aufl. Stuttgart: Thieme.

- [52] Knust, E., Janning, W. (2008): Genetik: Allgemeine Genetik - Molekulare Genetik - Entwicklungsgenetik: Thieme.
- [53] Kristinsson, S. Y., Bjorkholm, M., Andersson, T. M.-L., Eloranta, S., Dickman, P. W., Goldin, L. R., Blimark, C., Mellqvist, U.-H., Wahlin, A., Turesson, I., Landgren, O. (2009): Patterns of survival and causes of death following a diagnosis of monoclonal gammopathy of undetermined significance: a population-based study. *Haematologica*. 94 (12), 1714–1720.
- [54] Kyoto Encyclopedia of Genes and Genomes (2016): KEGG Database. Hg. v. Kanehisa Laboratories. Online-Publikation; <http://www.genome.jp/kegg/>, zuletzt geprüft am 01.07.2017.
- [55] Kyoto Encyclopedia of Genes and Genomes: KEGG Database, 2016 Online-Publication; <http://www.genome.jp/kegg/>.
- [56] Landgren, O., Kyle, R. A., Pfeiffer, R. M., Katzmann, J. A., Caporaso, N. E., Hayes, R. B., Dispenzieri, A., Kumar, S., Clark, R. J., Baris, D., Hoover, R., Rajkumar, S. V. (2009): Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. *Blood*. 113 (22), 5412–5417.
- [57] Law, P. J., Sud, A., Mitchell, J. S., Henrion, M., Orlando, G., Lenive, O., Broderick, P., Speedy, H. E., Johnson, D. C., Kaiser, M., Weinhold, N., Cooke, R., Sunter, N. J., Jackson, G. H., Summerfield, G., Harris, R. J., Pettitt, A. R., Allsup, D. J., Carmichael, J., Bailey, J. R., Pratt, G., Rahman, T., Pepper, C., Fegan, C., Strandmann, E. P. von, Engert, A., Forsti, A., Chen, B., Filho, M. I. d. S., Thomsen, H., Hoffmann, P., Nöthen, M. M., Eisele, L., Jöckel, K.-H., Allan, J. M., Swerdlow, A. J., Goldschmidt, H., Catovsky, D., Morgan, G. J., Hemminki, K., Houlston, R. S. (2017): Genome-wide association analysis of chronic lymphocytic leukaemia, Hodgkin lymphoma and multiple myeloma identifies pleiotropic risk loci. *Scientific reports*. 7.
- [58] Levey, A. S., Coresh, J., Greene, T., Marsh, J., Stevens, L. A., Kusek, J. W., van Lente, F. (2007): Expressing the modification of diet in renal disease study equation for estimating glomerular filtration rate with standardized serum creatinine values. *Clin Chem*. 53 (4), 766–772.

- [59] Marchini, J.; Howie, B. (2010): Genotype imputation for genome-wide association studies. *Nat Rev Genet.* 11 (7), 499–511.
- [60] Mi, H.; Thomas, P. (2009): PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol Biol.* 563, 123–140.
- [61] Möhlenkamp, S., Lehmann, N., Moebus, S., Schmermund, A., Dragano, N., Stang, A., Siegrist, J., Mann, K., Jöckel, K.-H., Erbel, R. (2011): Quantification of coronary atherosclerosis and inflammation to predict coronary events and all-cause mortality. *JACC.* 57 (13), 1455–1464.
- [62] Mueller-Eckhardt, C., Kiefel, V. (2004): *Transfusionsmedizin. Grundlagen - Therapie - Methodik. 3. überarbeitete und erweiterte Auflage.* Berlin, Heidelberg, s.l.: Springer Berlin Heidelberg.
- [63] Nakano, T., Matsui, M., Inoue, I., Awata, T., Katayama, S., Murakoshi, T. (2011): Free immunoglobulin light chain. Its biology and implications in diseases. *Clin Chim Acta.* 412 (11-12), 843–849.
- [64] Novembre, J.; Stephens, M. (2008): Interpreting principal component analyses of spatial population genetic variation. *Nat Genet.* 40 (5), 646–649.
- [65] Nowell, P. C.; Hungerford, D. A. (1960): Chromosome studies on normal and leukemic human leukocytes. *J Natl Cancer Inst.* 25, 85–109.
- [66] Orban, E., Arendt, M., Hennig, F., Lucht, S., Eisele, L., Jakobs, H., Dürig, J., Hoffmann, B., Jöckel, K.-H., Moebus, S. (2017): Is long-term particulate matter and nitrogen dioxide air pollution associated with incident monoclonal gammopathy of undetermined significance (MGUS)? An analysis of the Heinz Nixdorf Recall study. *Env int*, 237–245.
- [67] Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S.-E., Walford, G. A., Sugiana, C., Boneh, A., Chen, W. K., Hill, D. E., Vidal, M., Evans, J. G., Thorburn, D. R., Carr, S. A., Mootha, V. K. (2008): A mitochondrial protein compendium elucidates complex I disease biology. *Cell.* 134 (1), 112–123.

- [68] Pearson, T. A., Mensah, G. A., Alexander, R. W., Anderson, J. L., Cannon, R. O., Criqui, M., Fadl, Y. Y., Fortmann, S. P., Hong, Y., Myers, G. L., Rifai, N., Smith, S. C., Taubert, K., Tracy, R. P., Vinicor, F. (2003): Markers of inflammation and cardiovascular disease. *Circulation*. 107 (3), 499–511.
- [69] Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., Reich, D. (2006): Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 38 (8), 904–909.
- [70] Price, A. L., Zaitlen, N. A., Reich, D., Patterson, N. (2010): New approaches to population stratification in genome-wide association studies. *Nat Rev Genet*. 11 (7), 459–463.
- [71] Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., Boehnke, M., Abecasis, G. R., Willer, C. J. (2010): LocusZoom. Regional visualization of genome-wide association scan results. *Bioinformatics*. 26 (18), 2336–2337.
- [72] Psaty, B. M., O'Donnell, C. J., Gudnason, V., Lunetta, K. L., Folsom, A. R., Rotter, J. I., Uitterlinden, A. G., Harris, T. B., Witteman, J. C., Boerwinkle, E. (2009): Cohorts for heart and aging research in genomic epidemiology (CHARGE) Consortium. *Circulation: Cardiovascular Genetics*. 2 (1), 73–80.
- [73] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., Bakker, P. I. de, Daly, M. J., Sham, P. C. (2007): PLINK. A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81 (3), 559–575.
- [74] Rajkumar, S. V., Dimopoulos, M. A., Palumbo, A., Blade, J., Merlini, G., Mateos, M.-V., Kumar, S., Hillengass, J., Kastiris, E., Richardson, P., Landgren, O., Paiva, B., Dispenzieri, A., Weiss, B., LeLeu, X., Zweegman, S., Lonial, S., Rosinol, L., Zamagni, E., Jagannath, S., Sezer, O., Kristinsson, S. Y., Caers, J., Usmani, S. Z., Lahuerta, J. J., Johnsen, H. E., Beksac, M., Cavo, M., Goldschmidt, H., Terpos, E., Kyle, R. A., Anderson, K. C., Durie, B. G. M., Miguel, J. F. S. (2014): International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol*. 15 (12), e538–e548.

- [75] Rajkumar, S. V., Kyle, R. A., Therneau, T. M., Melton, L. J., Bradwell, A. R., Clark, R. J., Larson, D. R., Plevak, M. F., Dispenzieri, A., Katzmann, J. A. (2005): Serum free light chain ratio is an independent risk factor for progression in monoclonal gammopathy of undetermined significance. *Blood*. 106 (3), 812–817.
- [76] Reactome Database (2016). Online-Publikation; <http://www.reactome.org>, zuletzt geprüft am 01.07.2017.
- [77] Redegeld, F. A., Thio, M., Groot Kormelink, T. (2012): Polyclonal immunoglobulin free Light Chain and chronic inflammation. *Mayo Clin Proc*. 87 (10), 1032–1033.
- [78] Schmermund, A., Möhlenkamp, S., Stang, A., Grönemeyer, D., Seibel, R., Hirche, H., Mann, K., Siffert, W., Lauterbach, K., Siegrist, J., Jöckel, K.-H., Erbel, R. (2002): Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: Rationale and design of the Heinz Nixdorf Recall study. *Am Heart J*. 144 (2), 212–218.
- [79] Segrè, A. V., Groop, L., Mootha, V. K., Daly, M. J., Altshuler, D., Visscher, P. M. (2010): Common inherited variation in mitochondrial genes is not enriched for associations with Type 2 diabetes or related glycemic traits. *PLoS Genet*. 6 (8), e1001058.
- [80] Sen, R.; Baltimore, D. (1986): Inducibility of K immunoglobulin enhancer-binding protein NF- κ B by a posttranslational mechanism. *Cell*. (47), 921–928.
- [81] Siffert, W., Roskopf, D., Moritz, A., Wieland, T., Kaldenberg-Stasch, S., Kettler, N., Hartung, K., Beckmann, S., Jakobs, K. H. (1995): Enhanced G protein activation in immortalized lymphoblasts from patients with essential hypertension. *J Clin Invest*. 96 (2), 759–766.
- [82] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B., Gillette, M., Paulovich, A., Pomeroy, S., Golub, T., Lander, E., Mesirov, J. (2005): Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 102 (43), 15545–15550.
- [83] Teo, Y.-Y.; Sim, X. (2010): Patterns of linkage disequilibrium in different populations: implications and opportunities for lipid-associated loci identified from genome-wide association studies. *Curr Opin Lipidol*. 21 (2), 104–115.

- [84] The Binding Site Ltd (2016): FREELITE. Birmingham, UK. Online-Publikation; <http://www.bindingsite.com/de-de/discover/freelite-and-hevylite/freelite/laboratory-information/precision>, zuletzt geprüft am 01.07.2017.
- [85] The Gene Ontology: Gene ontology database. Online-Publikation; <http://www.geneontology.org/>, zuletzt geprüft am 01.07.2017.
- [86] The International Myeloma Working Group (2002): Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. *Br J Haematol.* **(121)**, 749–757.
- [87] Thorisson, G. A., Smith, A. V., Krishnan, L., Stein, L. D. (2005): The International HapMap project Web site. *Genome Res.* **15 (11)**, 1592–1593.
- [88] Tsai, H.-t., Caporaso, N. E., Kyle, R. A., Katzmann, J. A., Dispenzieri, A., Hayes, R. B., Marti, G. E., Albitar, M., Ghia, P., Rajkumar, S. V., Landgren, O. (2009): Evidence of serum immunoglobulin abnormalities up to 9.8 years before diagnosis of chronic lymphocytic leukemia: a prospective study. *Blood.* **114 (24)**, 4928–4932.
- [89] Turner, S., Armstrong, L. L., Bradford, Y., Carlson, C. S., Crawford, D. C., Crenshaw, A. T., Andrade, M. de, Doheny, K. F., Haines, J. L., Hayes, G., Jarvik, G., Jiang, L., Kullo, I. J., Li, R., Ling, H., Manolio, T. A., Matsumoto, M., McCarty, C. A., McDavid, A. N., Mirel, D. B., Paschall, J. E., Pugh, E. W., Rasmussen, L. V., Wilke, R. A., Zuvich, R. L., Ritchie, M. D. (2011): Quality control procedures for genome-wide association studies. *Curr Protoc Hum Genet.*, Unit1.19.
- [90] Turner, S. D. (2014): qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots.
- [91] van den Tweel, Jang G., Gleichmann, E. (1999): *Immunologie. Das menschliche Abwehrsystem.* Heidelberg, Berlin: Spektrum, Akad. Verl.
- [92] Verma, S. S., Andrade, M. de, Tromp, G., Kuivaniemi, H., Pugh, E., Namjou-Khales, B., Mukherjee, S., Jarvik, G. P., Kottyan, L. C., Burt, A., Bradford, Y., Armstrong, G. D., Derr, K., Crawford, D. C., Haines, J. L., Li, R., Crosslin, D., Ritchie, M. D. (2014): Imputation and quality control steps for combining multiple genome-wide datasets. *Front Genet.* **5**.

- [93] Veyrieras, J.-B., Kudaravalli, S., Kim, S. Y., Dermitzakis, E. T., Gilad, Y., Stephens, M., Pritchard, J. K. (2008): High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* 4 (10), e1000214.
- [94] Winkler, T. W., Day, F. R., Croteau-Chonka, D. C., Wood, A. R., Locke, A. E., Maggi, R., Ferreira, T., Fall, T., Graff, M., Justice, A. E., Luan, J., Gustafsson, S., Randall, J. C., Vedantam, S., Workalemahu, T., Kilpelainen, T. O., Scherag, A., Esko, T., Kutalik, Z., Heid, I. M., Loos, R. J. F. (2014): Quality control and conduct of genome-wide association meta-analyses. *Nat protoc.* 9 (5), 1192–1212.
- [95] World Health Organization (2016): International Classification of Diseases (ICD). Online-Publikation; <http://www.who.int/classifications/icd/en/>, zuletzt geprüft am 01.07.2017.
- [96] Yla-Herttuala, S., Rissanen, T. T., Vajanto, I., Hartikainen, J. (2007): Vascular endothelial growth factors: biology and current status of clinical applications in cardiovascular medicine. *JACC.* 49 (10), 1015–1026.
- [97] Zhang, Q., Zhu, B., Li, Y. (2017): Resolution of cancer-promoting inflammation. A new approach for anticancer therapy. *Front. Immunol.* 8 (6).
- [98] Ziegler, A., König, I. R. (2010): A statistical approach to genetic epidemiology. 2. Aufl. Weinheim: Wiley-VCH.
- [99] Ziegler, A., König, I. R., Thompson, J. R. (2008): Biostatistical aspects of genome-wide association studies. *Biom J.* 50 (1), 8–28.

8 Anhang

8.1 Abkürzungen

AMI	Acute myocardical infarction, dt. Akuter Myokardinfarkt
BNP	B-type natriuretisches Peptid
bp	<i>base pair</i> , engl. für Basenpaar innerhalb der DNA
CAC	Coronary artery calcification - dt. Koronare Verkalkung
CHARGE	Cohorts for Heart and Aging research in genomic epidemiology
CHIP	<i>Clonal hematopoiesis of indeterminate potential</i>
CLL	chronische lymphaitische Leukämie
CLSI	Clinical and Laboratory Standards Institute
cM	centi Morgan - Einheit für den Genabstand
CML	chronische myeloische Leukämie
CRAB	<i>Calcium levels increment, renal insufficiency, anaemia and bone lesions</i>
dL	Deziliter
DNA	<i>Desoxyribonucleic acid</i> , engl. für Desoxyribonukleinsäure
eGFR	<i>estimated glomerular filtration rate</i> , engl. für die geschätzte glomeruläre Filtrationsrate
FDR	<i>False-Discovery rate</i> , engl. für Rate der falsch-positiven Ergebnisse eines Tests
FGG	Fibrinogen Gamma Chain
FLC	<i>Free light-chain</i> , engl. für Freie Leichtketten
G-A-T-C	Die vier DNA-Basen: Guanin, Adenin, Thymin und Cytosin
GOTERM	Gene Ontology
GSEA	<i>Geneset Enrichment Analysis</i> , engl. für Genset Anreicherungsanalyse
GWAS	Genomweite Assoziationsstudie
HLA	<i>Human Leukocyte Antigen</i> , engl. für Humanes Leukozyten Antigen
HNR	Heinz Nixdorf Recall Studie
(hs-)CRP	(hoch-sensitives) C-reaktives Protein
HWE	Hardy-Weinberg Equilibrium
ICD	International classification of diseases

Anhang

IFE	Immunfixationselektrophorese
Ig	Immunglobulin
kb	Kilo Basenpaare
KEGG	Kyoto Encyclopedia of Genes und Genomes
km ²	Quadratkilometer
L	Liter
LD	<i>Linkage Disequilibrium</i> , engl. für Kopplungsungleichgewicht
LPD	<i>Lymphoproliferative disease</i> , engl. für Lymphoproliferative Erkrankung
MAF	<i>Minor Allele Frequency</i> , engl. für Häufigkeit des weniger häufigen Allels
MAGENTA	Meta-analysis gene-set enrichment of variaNT Associations
MB	Megabasenpaare
MDRD	Modification of Diet in Renal Disease
mg	Milligramm
mL	Milliliter
MGUS	Monoklonale Gammopathie unklarer Signifikanz
MM	Multiples Myelom
MRT	Magnetresonanztomographie
NF-κB	Nuclear factor 'kappa-light-chain-enhancer' of activated B-cells
OR	Odds ratio
PCA	<i>Principal Component Analysis</i> , engl. für Hauptkomponentenanalyse
pg	Pikogramm
pFLCR	pathologische FLC Ratio
PLAT	Plasminogen Aktivator
PLG	Plasminogen
RNA	<i>Ribonucleic Acid</i> , engl. für Ribonukleinsäure
SLiM	S= Mehr als 60 % Plasmazellinfiltration im Knochenmark (sixty) Li = Verhältnis der betroffenen zur nicht-betroffenen Freien Leichtkette muss über 100 liegen, ebenso wie die Konzentration der betroffenen Freien Leichtkette über 100 mg/L liegen muss M = Mehr als eine fokale Läsion im Ganzkörper-MRT mit einer Größe größer als 5 mm
SNP	<i>Single Nucleotide Polymorphism</i> , engl. für Einzelbasenaustausch
SNV	<i>Single Nucleotide Variation</i> , engl. für Einzelbasenvariation
SPE	Serumproteinelektrophorese
VEGF	engl. <i>Vascular Endothelial Growth Factor</i>

8.2 Hauptkomponentenanalyse

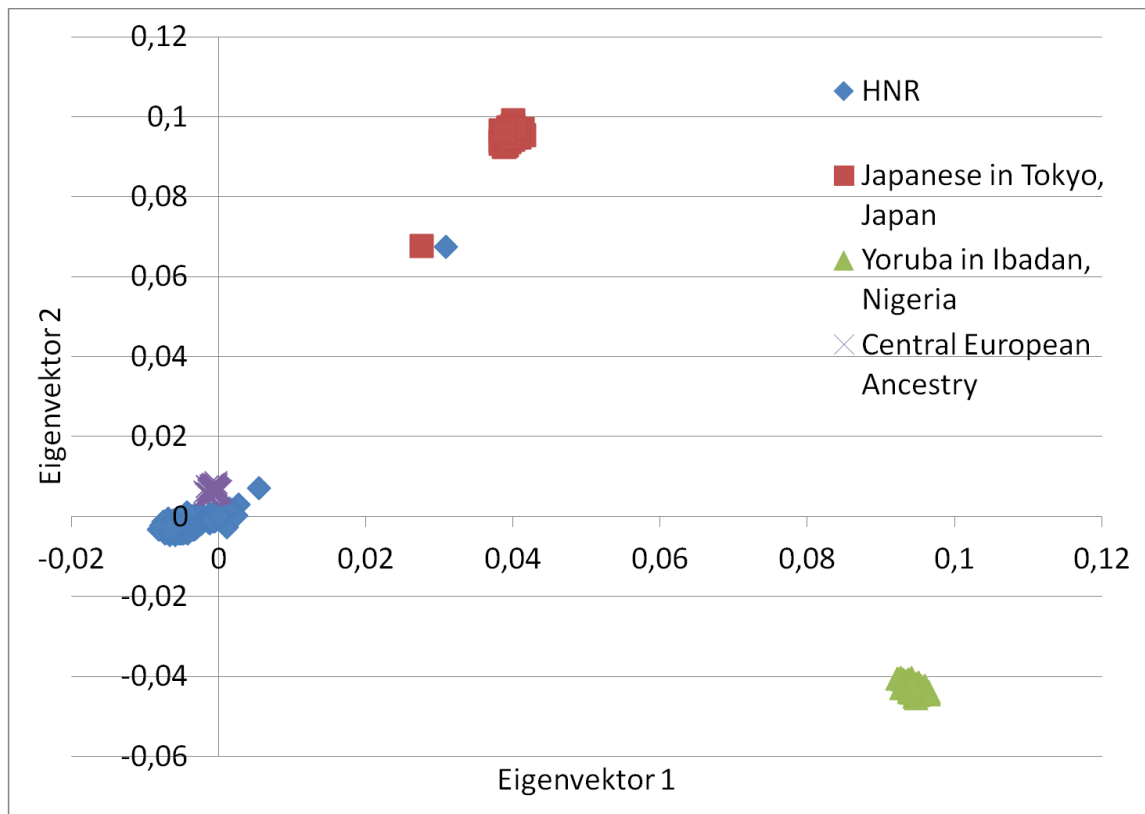


Abbildung 7 Darstellung der ersten beiden Eigenvektoren der Hauptkomponentenanalyse, in der die hier verwendete HNR-Population in Vergleich gesetzt wird mit einem Kollektiv aus Europa (CEU), einem aus Japan und einem aus Afrika

8.3 Gensets mit einer FDR < 25 %

Gensets mit einer FDR $\geq 5\%$ und < 25 %:

Tabelle 14 Gensets aus der GSEA mit den logarithmierten κ -Werten mit einer FDR $\geq 5\%$ und < 25 %

Datenbank	Genset	p-Wert	FDR	Gene
Reactome	ADP signalling through P2Y purinoceptor 12	$1,00 \cdot 10^{-4}$	$5,02 \cdot 10^{-2}$	GNAI1, GNAI2, GNAI3, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, GNB5, GNG13, GNG2, GNG12, GNB4, P2RY12, GNG8
Reactome	Platelet activation triggers	$9,00 \cdot 10^{-4}$	$7,09 \cdot 10^{-2}$	AKT1, AKT2, COL1A1, COL1A2, MAPK14, F2, F2R, F2RL2, FCER1G, FYN, GNA11, GNA12, GNA15, GNAI1, GNAI2, GNAI3, GNAQ, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, LCP2, LYN, P2RY1, PDPK1, PIK3CA, PIK3CB, PIK3CG, PIK3R1, PLA2G4A, PLCG2,
Reactome	GS Alpha mediated events in glucagon signalling	$1,40 \cdot 10^{-3}$	$7,09 \cdot 10^{-2}$	ADCY1, ADCY2, ADCY3, ADCY5, ADCY6, ADCY7, ADCY8, ADCY9, GCG, GCGR, GNAS, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNG13, GNG2, ADCY10, GNG12, GNB4, GNG8, ADCY4
Reactome	G Beta Gamma signalling through PLC Beta	$2,30 \cdot 10^{-3}$	$7,18 \cdot 10^{-2}$	GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, PLCB2, PLCB3, GNB5, PLCB1, GNG13, GNG2, GNG12, GNB4, GNG8
Reactome	Thromboxane signalling through TP receptor	$2,60 \cdot 10^{-3}$	$7,74 \cdot 10^{-2}$	GNA11, GNA15, GNAQ, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, TBXA2R, GNA14, GNA13, GNB5, GNG13, GNG2, GNG12, GNB4, GNG8
Reactome	Signal amplification	$2,90 \cdot 10^{-3}$	$7,78 \cdot 10^{-2}$	MAPK14, GNA11, GNA15, GNAI1, GNAI2, GNAI3, GNAQ, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, P2RY1, PLA2G4A, SRC, TBXA2R, GNA14, GNA13, GNB5, GNG13, GNG2, GNG12, GNB4, P2RY12, GNG8

Anhang

Fortsetzung Tabelle 14

Datenbank	Genset	p-Wert	FDR	Gene
Reactome	Glucagon signalling in metabolic regulation	$8,00 \cdot 10^{-4}$	$7,92 \cdot 10^{-2}$	ADCY1, ADCY2, ADCY3, ADCY5, ADCY6, ADCY7, ADCY8, ADCY9, GCG, GCGR, GNAS, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, PRKACA, PRKACB, PRKACG, PRKAR1A, PRKAR1B, PRKAR2A, PRKAR2B, GNG13, GNG2, ADCY10, GNG12, GNB4, GNG8, ADCY4
Reactome	Regulation of insulin secretion by glucagon like peptide 1	$2,80 \cdot 10^{-3}$	$7,95 \cdot 10^{-2}$	ADCY5, ADCY6, ADCY8, CACNA1A, CACNA1C, CACNA1D, CACNA1E, CACNB2, CACNB3, GCG, GLP1R, GNAI1, GNAI2, GNAO1, GNAS, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, INS, ITPR1, ITPR2, ITPR3, KCNB1, KCNC2, KCNJ11, KCNS3, PRKACA, PRKACB, PR
Reactome	Opioid signalling	$1,10 \cdot 10^{-3}$	$8,59 \cdot 10^{-2}$	ADCY1, ADCY2, ADCY3, ADCY5, ADCY6, ADCY7, ADCY8, ADCY9, ADRBK1, CALM1, CALM2, CALM3, CAMK4, CDK5, CREB1, GNA11, GNA15, GNAI1, GNAI2, GNAI3, GNAL, GNAO1, GNAQ, GNAT1, GNAT2, GNAZ, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, ITPR1,
Reactome	ADP signalling through P2Y purinoceptor 1	$6,80 \cdot 10^{-3}$	$8,59 \cdot 10^{-2}$	MAPK14, GNA11, GNA15, GNAQ, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, P2RY1, PLA2G4A, SRC, GNA14, GNB5, GNG13, GNG2, GNG12, GNB4, GNG8
Reactome	G Beta Gamma signalling through P13KGamma	$6,00 \cdot 10^{-3}$	$9,11 \cdot 10^{-2}$	AKT1, AKT2, RHOA, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, PDPK1, PIK3CG, AKT3, GNB5, PIK3R5, GNG13, GNG2, GNG12, GNB4, GNG8, PIK3R6
Reactome	Glucagon type ligand receptors	$4,00 \cdot 10^{-3}$	$9,25 \cdot 10^{-2}$	ADCYAP1, ADCYAP1R1, GCG, GCGR, GHRH, GHRHR, GIP, GIPR, GLP1R, GNAS, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, SCT, SCTR, VIP, VIPR1, VIPR2, GLP2R, GNB5, GNG13, GNG2, GNG12, GNB4, GNG8
Reactome	Thrombin signalling through proteinase activated receptors	$7,80 \cdot 10^{-3}$	$9,32 \cdot 10^{-2}$	F2, F2R, F2RL2, GNA11, GNA12, GNA15, GNAQ, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, F2RL3, GNA14, GNA13, GNB5, GNG13, GNG2, GNG12, GNB4, GNG8
Panther Molecular Functions	Large G-protein	$2,90 \cdot 10^{-3}$	$1,01 \cdot 10^{-1}$	GBP1, GBP2, GBP3, GNA11, GNA12, GNA15, GNAI1, GNAI2, GNAI3, GNAL, GNAO1, GNAQ, GNAS, GNAT1, GNAT2, GNAZ, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, GNA14, GNA13, GNB5, ATL3, ATL1, GNG13, GNG2, GNG12, GNB4, ATL2, GNG8, GBP4, GBP5

Anhang

Fortsetzung Tabelle 14

Datenbank	Genset	p-Wert	FDR	Gene
Reactome	G Protein Beta Gamma Signalling	$9,00 \cdot 10^{-3}$	$1,30 \cdot 10^{-1}$	AKT1, AKT2, RHOA, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, PDPK1, PIK3CG, PLCB2, PLCB3, AKT3, GNB5, PLCB1, PIK3R5, GNG13, GNG2, GNG12, GNB4, GNG8, PIK3R6
Panther Molecular Functions	Kinase	$2,60 \cdot 10^{-3}$	$1,30 \cdot 10^{-1}$	DGKA, DGKB, DGKG, DGKQ, ITPKA, ITPKB, DGKZ, DGKE, DGKD, DGKI, IP6K1, CIB1, TPK1, IP6K2, PXX, PI4K2B, RFK, PI4K2A, AGK, SNX16, NADK, DCAKD, ITPKC, COASY, NTPCR, IP6K3, CIB4, DGKK, DGKH, IPMK
Reactome	Xenobiotics	$1,95 \cdot 10^{-2}$	$1,42 \cdot 10^{-1}$	CYP1A1, CYP1A2, CYP2A6, CYP3A7, CYP2A13, CYP2B6, CYP2C19, CYP2C8, CYP2C9, CYP2C18, CYP2D6, CYP2E1, CYP2F1, CYP3A4, CYP3A5
Reactome	CD28 Co-stimulation	$1,33 \cdot 10^{-2}$	$1,66 \cdot 10^{-1}$	AKT1, AKT2, CD28, CD80, CD86, CDC42, MAP3K8, MTOR, FYN, GRB2, LCK, LYN, PAK1, PAK2, PDPK1, PIK3CA, PIK3R1, RAC1, SRC, VAV1, YES1, MAP3K14, GRAP2, AKT3, TRIB3, MLST8, MAPKAP1, THEM4, RICTOR
Reactome	Amine derived hormones	$2,26 \cdot 10^{-2}$	$1,68 \cdot 10^{-1}$	AANAT, DBH, DDC, DIO1, DIO2, DIO3, PNMT, SLC5A5, TH, TPH1, TPO, IYD
Panther Molecular Functions	Esterase	$5,20 \cdot 10^{-3}$	$1,78 \cdot 10^{-1}$	ACHE, BCHE, CES1, ESD, FASN, PAFAH1B1, PDE6G, PDE6H, PON1, PON2, PON3, PPT1, CES2, PAN2, ACOT8, RPP38, POP4, PNPLA6, TREX2, ACOT7, PMPCA, ACOT9, ACOT11, EXOSC3, NT5C3, PPME1, GDE1, PTRH2, SIAE, GDPD2, NEIL3, ACOT13, MTMR14, MPPE1, GDPD3, NEIL1, GDPD5, PNP
Reactome	Activation of kainate receptors upon glutamate binding	$1,65 \cdot 10^{-2}$	$1,84 \cdot 10^{-1}$	CALM1, CALM2, CALM3, DLG1, DLG3, DLG4, GNB1, GNB2, GNB3, GNG3, GNG4, GNG5, GNG7, GNG10, GNG11, GNGT1, GNGT2, GRIK1, GRIK2, GRIK3, GRIK4, GRIK5, PLCB2, PLCB3, GNB5, PLCB1, GNG13, GNG2, GNG12, GNB4, NCALD, GNG8
Reactome	Platelet adhesion to exposed collagen	$2,38 \cdot 10^{-2}$	$1,86 \cdot 10^{-1}$	CD36, COL1A1, COL1A2, FCER1G, FYN, GP1BA, GP5, GP9, ITGA2, ITGB1, LYN, VWF, GP6
Panther Molecular Functions	Non-receptor tyrosine protein kinase	$1,50 \cdot 10^{-3}$	$1,89 \cdot 10^{-1}$	ABL1, ABL2, BLK, BMX, BTK, CSK, DYRK1A, PTK2B, FER, FES, FGR, FRK, FYN, HCK, ITK, JAK1, JAK2, JAK3, LCK, LYN, MATK, PTK2, PTK6, SRC, SRMS, SYK, TEC, TXK, TYK2, YES1, ZAP70, DYRK3, DYRK2, TNK1, DYRK4, DYRK1B, TNK2, SH3BP4
Panther Molecular Functions	Hsp70 family chaperone	$2,02 \cdot 10^{-2}$	$1,90 \cdot 10^{-1}$	HSPA2, HSPA4, HSPA5, HSPA6, HSPA8, HSPA9, HSPA13, HYOU1, HSPH1, HSPA4L, HSPA12B, HSPA12A

Anhang

Fortsetzung Tabelle 14

Datenbank	Genset	p-Wert	FDR	Gene
Panther Biological Functions	Other apoptosis	$1,18 \cdot 10^{-2}$	$2,04 \cdot 10^{-1}$	SCARB1, SCARB2, PSEN2, SNCG, NAE1, NOL3, PPP1R15A, GULP1, PSMG2, CARD6
BIOCARTA	Asbcell Pathway	$8,00 \cdot 10^{-3}$	$2,14 \cdot 10^{-1}$	FAS, FASLG, CD4, CD28, CD80, CD40, CD40LG, IL2, IL4, IL10
Panther	Adrenaline and noradrenaline biosynthesis	$4,50 \cdot 10^{-3}$	$2,20 \cdot 10^{-1}$	DDC, HDC, PAH, PNMT, SLC6A2, SLC6A3, TH, TPH1, SLC6A16, SLC6A20, SLC6A15, TPH2, SLC6A19, SLC6A18, SLC6A17
Panther Molecular Functions	Actin and actin related protein	$1,24 \cdot 10^{-2}$	$2,24 \cdot 10^{-1}$	ACTA1, ACTA2, ACTB, ACTC1, ACTG1, ACTG2, ACTL6A, ACTR3, ACTR2, ACTR1B, ACTR1A, ACTL7B, ACTL7A, PARVB, ACTL6B, PARVA, ACTR10, ACTR3B, PARVG, ACTR6, ACTR5, ACTL8, ARPM1, ACTR8, ACTRT1, ACTRT2, C20orf134, ACTL9, ACTBL2, POTE, ACTR3C
BIOCARTA	TCapoptosis pathway	$4,90 \cdot 10^{-3}$	$2,29 \cdot 10^{-1}$	FAS, FASLG, CD3D, CD3E, CD3G, CD247, CD4, CD28, CCR5
Panther Molecular Functions	Chromatin chromatin-bindingprotein	$8,10 \cdot 10^{-3}$	$2,45 \cdot 10^{-1}$	BRDT, RCC1, CREBBP, PHC1, PHC2, EP300, HMGN1, HMGN2, DPF2, SAFB, SCML1, SMARCC1, SMARCC2, SMARCD1, SMARCD2, SMARCD3, SMARCE1, SSRP1, SS18, SUPT4H1, SUPT5H, SUPT6H, SYCP1, TADA2A, TNP2, HIRA, UBTF, MYST3, BRD3, DPF3, DPF1, CHAF1B, SMC1A, ARID1A, CBX4, SAP3
Panther	Opioid proopiomelanocortin pathway	$1,01 \cdot 10^{-2}$	$2,47 \cdot 10^{-1}$	GNAI1, GNAI2, GNAI3, GNAO1, GNG3, GNG4, GNG5, GNG7, GNG10, GNGT2, OPRD1, OPRM1, POMC, GNB5, GNG2, GNG12, GNG8, DNAJC25-GNG10

Tabelle 15 Gensets aus der GSEA mit den logarithmierten λ -Werten mit einer FDR $\geq 5\%$ und $< 25\%$

Datenbank	Genset	p-Wert	FDR	Gene
KEGG	Glycerophospho-lipid metabolism	$6,00 \cdot 10^{-4}$	$8,35 \cdot 10^{-2}$	ACHE, CDS1, CHAT, CHKA, CHKB, DGKA, DGKB, DGKG, DGKQ, GPD1, GPD2, LCAT, PCYT1A, PLA2G1B, PLA2G2A, PLA2G4A, PLA2G5, PLD1, PLD2, PCYT2, TAZ, PLA2G6, PLA2G10, GNPAT, DGKZ, DGKE, DGKD, PPAP2A, PPAP2C, PPAP2B, JMJD7-PLA2G4B, CDS2, DGKI, PCYT1B, PGS1, PTDSS1
Panther	PI3 kinasepathway	$3,00 \cdot 10^{-3}$	$8,50 \cdot 10^{-2}$	GNAT1, GNAT2, HRAS, IGF1BP1, JAK1, KRAS, NRAS, PIK3R1, PIK3R2, PTGER1, PTGER3, RRAS, PIK3R3
Panther	Gamma-aminobutyric acidsynthesis	$2,18 \cdot 10^{-2}$	$1,28 \cdot 10^{-1}$	GAD1, GAD2, ALDH5A1, CSAD, GADL1
Panther Biological process	Fatty acid biosynthesis	$5,10 \cdot 10^{-3}$	$1,91 \cdot 10^{-1}$	ACACA, ACACB, CYP8B1, FASN, LTA4H, NDUFB1, PCCB, PTGDS, PTGIS, RNPEP, TBXAS1, MCAT, OXSM, RNPEPL1
Panther Molecular Functions	Acyltransferase	$5,40 \cdot 10^{-3}$	$2,48 \cdot 10^{-1}$	AUH, AUP1, CHAT, CHM, CHML, COX10, CPT1A, CPT1B, CPT2, CRAT, DBT, DCI, DLST, ECH1, ECHS1, EPB42, FASN, FDFT1, FDPS, FNTA, FNTB, GDI1, GDI2, GGT1, GGT7, GGT5, LCAT, ME1, ME2, PGGT1B, RABGGTB, SOAT1, ELOVL4, TAF1, TAZ, TGM4, TGM1, TGM2, TGM3, PNPLA4

Anhang

Tabelle 16 Gensets aus der GSEA mit der logarithmierten Summe der λ - und κ -Werte mit einer FDR $\geq 5\%$ und $< 25\%$

Datenbank	Genset	p-Wert	FDR	Gene
Panther Molecular Functions	Chromatin chromatin-binding protein	$1,50 \cdot 10^{-3}$	$1,61 \cdot 10^{-1}$	BRDT, RCC1, CREBBP, PHC1, PHC2, EP300, HMG1, HMG2, DPF2, SAFB, SCML1, SMARCC1, SMARCC2, SMARCD1, SMARCD2, SMARCD3, SMARCE1, SSRP1, SS18, SUPT4H1, SUPT5H, SUPT6H, SYCP1, TADA2A, TNP2, HIRA, UBTf, MYST3, BRD3, DPF3, DPF1, CHAF1B, SMC1A, ARID1A, CBX4, SAP3
Panther	Opioid proopiomelanocortin pathway	$7,30 \cdot 10^{-3}$	$1,61 \cdot 10^{-1}$	GNAI1, GNAI2, GNAI3, GNAO1, GNG3, GNG4, GNG5, GNG7, GNG10, GNGT2, OPRD1, OPRM1, POMC, GNB5, GNG2, GNG12, GNG8, DNAJC25-GNG10
Panther Molecular Functions	Non-receptor tyrosine protein kinase	$1,90 \cdot 10^{-3}$	$1,62 \cdot 10^{-1}$	ABL1, ABL2, BLK, BMX, BTK, CSK, DYRK1A, PTK2B, FER, FES, FGR, FRK, FYN, HCK, ITK, JAK1, JAK2, JAK3, LCK, LYN, MATK, PTK2, PTK6, SRC, SRMS, SYK, TEC, TXK, TYK2, YES1, ZAP70, DYRK3, DYRK2, TNK1, DYRK4, DYRK1B, TNK2, SH3BP4
Panther Molecular Functions	Damaged DNA-binding protein	$5,70 \cdot 10^{-3}$	$1,74 \cdot 10^{-1}$	BRCA2, DDB1, DDB2, FEN1, MSH6, MPG, MSH2, MSH3, MSH4, NBN, RAD23A, RAD23B, XPA, XPC, AIMP2, RAD50, APTX, NEIL3, UTP6, XAB2, MCPH1, NEIL1, HINT3, NEIL2
Panther	PI3 kinase pathway	$2,31 \cdot 10^{-2}$	$1,77 \cdot 10^{-1}$	GNAT1, GNAT2, HRAS, IGF1R, JAK1, KRAS, NRAS, PIK3R1, PIK3R2, PTGER1, PTGER3, RRAS, PIK3R3
Panther Biological process	Oogenesis	$7,70 \cdot 10^{-3}$	$1,87 \cdot 10^{-1}$	BLK, C18orf1, CGB, ERBB4, ESR1, ESR2, FGR, FRK, FSHB, FYN, GDF9, GNRH1, GNRH2, GNRHR, HCK, LCK, LDLR, LHB, LRP1, LRP2, LRP3, LRP4, LRP6, LRP5, LYN, MAGOH, MOS, OXTR, SRC, STAR, VLDLR, YES1, LRP8, BMP15, PIWIL1, STARD3, SSPO, SH3BP4, FSCN2, LRP10, LRP12, C
Panther Molecular Functions	Phosphatase modulator	$1,01 \cdot 10^{-2}$	$2,01 \cdot 10^{-1}$	PPP1R12A, PPP1R3A, PPP1R3C, PPP1R3D, PPP1R7, PPP6R2, PHACTR2, PPP6R1, PPP1R16B, PPP1R12C, PPP6R3, PHACTR4, PPP1R3B, PPP1R16A, PPP1R3F, PPP1R3E, PHACTR3, LRGUK, PHACTR1
Panther	Interleukin signaling pathway	$7,70 \cdot 10^{-3}$	$2,12 \cdot 10^{-1}$	CDKN1A, CDKN1B, CHUK, CSF2RB, ELK3, IL1A, IL2, IL2RA, IL2RB, IL4, IL4R, IL5, IL5RA, IL6, IL6R, IL6ST, IL7, CXCR1, CXCR2, IL10, IL10RA, IL10RB, IL11RA, IL12A, IL12RB1, IL12RB2, IL13, IL13RA1, IL13RA2, IL15, IL16, IL17A, IRS1, MAP3K3, MAP3K4, MAPK4, MAPK6,

Anhang

Fortsetzung Tabelle 16

Datenbank	Genset	p-Wert	FDR	Gene
Panther Biological process	Carbohydrate transport	$6,10 \cdot 10^{-3}$	$2,18 \cdot 10^{-1}$	BLK, FGR, FRK, FYN, SLC37A4, HCK, INS, LCK, LYN, RSC1A1, SLC2A1, SLC2A2, SLC2A3, SLC2A4, SLC2A5, SRC, SLC35A2, YES1, REEP5, ADIPOQ, SLC35B1, SORBS1, SLC2A6, SH3BP4, SLC2A8, SLC45A1, SLC35B3, SLC45A2, REEP2, SVOP, SLC2A9, SLC45A4, HIAT1, REEP1, SLC2A11, RE
Panther Biological process	Developmental processes	$1,70 \cdot 10^{-3}$	$2,19 \cdot 10^{-1}$	ABL1, ABL2, ADM, ADRA1D, ADRA1B, ADRA1A, AES, ANXA13, CXCR5, BMI1, BSG, CAMK2A, CAMK2B, CAMK2D, CAMK2G, CAPS, CEBPB, CISH, CRYBA1, CRYBA2, CRYBA4, CRYBB1, CRYBB2, CRYBB3, CRYGA, CRYGB, CRYGC, CRYGD, CRYGS, CUX1, DMRT1, DPP6, ARID3A, E2F2, E2F3, E2F4, E2F5
Panther Molecular Functions	DNA glycosylase	$2,01 \cdot 10^{-2}$	$2,24 \cdot 10^{-1}$	MPG, MUTYH, NTHL1, OGG1, TDG, UNG, CCNO, SMUG1, NEIL3, NEIL1, MCEE, NEIL2
Reactome	Eicosanoid ligand binding receptors	$2,50 \cdot 10^{-3}$	$2,32 \cdot 10^{-1}$	LTB4R, GPR17, PTGDR, PTGER1, PTGER2, PTGER3, PTGER4, PTGFR, PTGIR, TBXA2R, CYSLTR1, GPR44, LTB4R2, CYSLTR2, OXER1
Panther	Gamma-aminobutyric acid synthesis	$2,30 \cdot 10^{-2}$	$2,38 \cdot 10^{-1}$	GAD1, GAD2, ALDH5A1, CSAD, GADL1
Panther Biological process	Protein-lipid modification	$1,13 \cdot 10^{-2}$	$2,44 \cdot 10^{-1}$	FNTA, FNTB, NMT1, PGGT1B, PIGH, PPT1, RABGGTB, GPAA1, NMT2, PIGK, LYPLA1, LYPLA2, PIGG, HHAT, HHATL, PIGO, PIGM, PTAR1
Panther Biological process	Gluconeogenesis	$8,90 \cdot 10^{-3}$	$2,49 \cdot 10^{-1}$	ACACA, ACACB, GPI, PC, PCCA, PCK1, PCK2, PRKAA1, PRKAA2, PRKAB1, PRKAB2, ADIPOQ, MAEA, SDS, MCCC1, ACN9, SDSL

Tabelle 17 Gensets aus der GSEA mit der pathologischen Ratio mit einer FDR $\geq 5\%$ und $< 25\%$

Datenbank	Genset	p-Wert	FDR	Gene
BIOCARTA	CD40 pathway	$2,70 \cdot 10^{-3}$	$5,41 \cdot 10^{-2}$	CD40, CD40LG, CHUK, DUSP1, IKBKB, MAP3K1, NFKB1, NFKBIA, RELA, TNFAIP3, TRAF3, TRAF6, IKBKG, IKBKAP, MAP3K14
BIOCARTA	AMI pathway	$3,40 \cdot 10^{-3}$	$5,92 \cdot 10^{-2}$	SERPINC1, COL4A1, COL4A2, COL4A3, COL4A4, COL4A6, F2, F2R, F7, F10, FGA, FGB, FGG, PLAT, PLG, PROC, PROS1, TFPI, AHSP
BIOCARTA	Plateletapp Pathway	$8,30 \cdot 10^{-3}$	$7,29 \cdot 10^{-2}$	APP, COL4A1, COL4A2, COL4A3, COL4A4, COL4A6, F2, F9, F11, SERPINE1, PLAT, PLAU, PLG
Panther Molecular process	Protein biosynthesis	$1,00 \cdot 10^{-4}$	$1,12 \cdot 10^{-1}$	MRPL49, CYLD, EEF1B2, EEF1D, EEF1G, EEF2, EIF1AX, EIF2S1, EIF2B1, EIF2S3, EIF4A1, EIF4A2, EIF4B, EIF4E, EIF4EBP1, EIF4EBP2, EIF4G1, EIF4G2, EIF5, EIF5A, ETF1, GCLM, GSPT1, ICT1, EIF3E, EIF6, MTIF2, RPL10A, NHP2L1, PA2G4, PIGH, EIF2AK2, RPL3, RPL3L, RPL4,
BIOCARTA	Extrinsic pathway	$1,25 \cdot 10^{-2}$	$1,25 \cdot 10^{-1}$	SERPINC1, F2, F2R, F3, F5, F7, F10, FGA, FGB, FGG, PROC, PROS1, TFPI
Ingenuity	NFKB Signaling	$2,60 \cdot 10^{-3}$	$1,55 \cdot 10^{-1}$	TNFRSF17, CD40, CD40LG, CHUK, MAP3K8, EGF, GH1, GSK3B, IKBKB, LCK, LTBR, MYD88, NFKB1, NFKB2, PLCG2, PRKCZ, MAPK8, EIF2AK2, RELA, RELB, MAP3K7, TGFA, TNFAIP3, TRAF2, TRAF6, UBE2N, UBE2V1, ZAP70, IKBKG, TNFSF11, RIPK1, BCL10, BTRC, TAB1, TNFSF13B, MALT1, I
Ingenuity	VEGF Signaling	$3,97 \cdot 10^{-2}$	$2,18 \cdot 10^{-1}$	ARNT, BAD, BCL2, BCL2L1, ELAVL1, PTK2B, SFN, GRB2, HIF1A, NOS3, PTK2, PXN, RAF1, SHC1, VCL, YWHAE
KEGG	Endocytosis	$5,00 \cdot 10^{-4}$	$2,48 \cdot 10^{-1}$	ADRB1, ADRB2, ADRB3, ADRBK1, ADRBK2, AP2A1, AP2A2, AP2B1, ARF6, ARRB1, ARRB2, CBL, CBLB, CDC42, AP2M1, AP2S1, CLTA, CLTB, CLTC, CCR5, CSF1R, DAB2, DNMT1, DNMT2, EGF, EGFR, EPS15, ERBB3, ERBB4, F2R, FGFR3, FGFR2, FGFR4, FLT1, GRK4, GRK5, GRK6, HLA-F, HRAS, H
Ingenuity	Huntingtons Disease Signaling	$2,78 \cdot 10^{-2}$	$2,49 \cdot 10^{-1}$	AP2A2, AKT1, APAF1, BAX, BCL2L1, BDNF, CACNA1B, CASP8, CASP9, CDK5, CLTC, CTSD, DCTN1, DLG4, DYNC1I2, EGF, EGFR, MTOR, GLS, GRB2, GRIN2B, HTT, HIP1, HRAS, DNAJB1, IGF1, IGF1R, ITPR1, JUN, MAP3K10, NEUROD1, NTRK1, PDPK1, PENK, MAPK8, MAPK9, RASA1, REST, SG

8.4 Abbildungsverzeichnis

Abbildung 1	Darstellung der ersten beiden Eigenvektoren der HapMap (Thorisson et al., 2005) Vergleichspopulationen, in der erkennbar wird, dass sich die drei Populationen der Europäer, der Japaner und der Nigerianer deutlich voneinander abgrenzen	25
Abbildung 2	Studienübersicht der HNR-Studie aufgeteilt nach Genotypisierungen	37
Abbildung 3	Übersicht der Qualitätskontrollschritte der zwei verwendeten DNA-Chips HumanOmni1-Quad und HumanOmniExpress, ebenso wie das Zusammenfügen beider Chips, um möglichst viele Probandeninformationen pro SNP zu erhalten	38
Abbildung 4	Manhattan-Plots der GWAS-Ergebnisse der vier verschiedenen Analysen zur κ - und λ -Konzentration, deren Summe und der pathologischen FLC-Ratio	41
Abbildung 5	LocusZoom der genetischen Region um die zwei Varianten rs2289433 und rs2053404, die in der GWAS zu pFLCR assoziiert sind. In LocusZoom ist immer nur die Darstellung eines SNPs möglich. Die beiden SNPs sind daher für diese Darstellung graphisch zusammengefügt worden	43
Abbildung 6	Zusammenfassende Übersicht der Ergebnisse	72
Abbildung 7	Darstellung der ersten beiden Eigenvektoren der Hauptkomponentenanalyse, in der die hier verwendete HNR-Population in Vergleich gesetzt wird mit einem Kollektiv aus Europa (CEU), einem aus Japan und einem aus Afrika	90

8.5 Tabellenverzeichnis

Tabelle 1	Beschreibung der HNR-Studienpopulation genotypisiert auf HumanOmni1-Quad und HumanOmniExpress zur Basiserhebung. Hierbei wurden nur vollständige Daten verwendet. Vergleichende Tests wurden je nach Skalenniveau mit dem exakten Test nach Fisher beziehungsweise dem Wilcoxon-Mann-Whitney Test durchgeführt	39
Tabelle 2	Ergebnisse aus der GWAS, die mit Daten der HNR-Studie ($N_{\text{ges}} = 2.109$) zu Studienbeginn durchgeführt wurde. Es sind jene genetischen Varianten aufgeführt, die einen p-Wert $< 10^{-5}$ erreichten	42
Tabelle 3	Gensets, aus den GSEA-Ergebnissen zu den logarithmierten κ -Werten, die mit einer FDR $< 5\%$ angereichert sind	44
Tabelle 4	Details des Gensets <i>G Protein Activation</i> mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu den logarithmierten κ -Werten	45
Tabelle 5	Details des Gensets <i>Inhibition of Insulin secretion by adrenaline noradrenaline</i> mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu den logarithmierten κ -Werten	46
Tabelle 6	Details des Gensets <i>Clathrin coated vesicle membrane</i> mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu den logarithmierten κ -Werten	47
Tabelle 7	Gensets aus den GSEA-Ergebnissen zu den logarithmierten λ -Werten, die mit einer FDR $< 5\%$ angereichert sind	48
Tabelle 8	Details des Gensets <i>HMG box transcription factor</i> mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu den logarithmierten λ -Werten	49
Tabelle 9	Details des Gensets <i>DNA Glycosylase</i> mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu den logarithmierten λ -Werten	50

Anhang

Tabelle 10	Details des Gensets <i>Interleukin signaling pathway</i> mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu den logarithmierten λ -Werten	51
Tabelle 11	Gensets aus den GSEA-Ergebnissen zu der pathologischen FLC-Ratio, die mit einer FDR < 5 % angereichert sind	53
Tabelle 12	Details des Gensets <i>EIF2 pathway</i> mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu der pathologischen FLC-Ratio	53
Tabelle 13	Details des Gensets <i>Fibrinolysis Pathway</i> mit den zugehörigen Genen, den zugeordneten SNPs, den adjustierten GWAS-p-Werten und Effektschätzern zu der pathologischen FLC-Ratio	54
Tabelle 14	Gensets aus der GSEA mit den logarithmierten κ -Werten mit einer FDR $\geq 5\%$ und $< 25\%$	91
Tabelle 15	Gensets aus der GSEA mit den logarithmierten λ -Werten mit einer FDR $\geq 5\%$ und $< 25\%$	94
Tabelle 16	Gensets aus der GSEA mit der logarithmierten Summe der λ - und κ -Werte mit einer FDR $\geq 5\%$ und $< 25\%$	95
Tabelle 17	Gensets aus der GSEA mit der pathologischen Ratio mit einer FDR $\geq 5\%$ und $< 25\%$	97

9 Danksagung

Ich danke allen Mitarbeitern vom IMIBE, die mich auf irgendeine Weise in meiner Promotionszeit unterstützt, gefördert und gefordert haben. Mein Dank gilt Professor Jöckel, der mich nach meiner Masterarbeit als Mitarbeiterin eingestellt hat. Im Besonderen möchte ich mich bei meinem Betreuer und Doktorvater Lewin Eisele bedanken, der mit mir zusammen dieses spannende Thema bearbeitet hat. Ich bedanke mich bei der Heinz Nixdorf Stiftung (Vorstand: M. Nixdorf, Essen) für die finanzielle Unterstützung der HNR Studie. Ebenso bei der *Investigative group* sowie den Mitarbeitern des Studienzentrums. Teile dieser Arbeit wurden durch einen internen Forschungszuschuss an Lewin Eisele der Medizinischen Fakultät des Universitätsklinikums Essen (IFORES) gefördert. Die Untersuchungen zu MGUS und FLC in der HNR Studie wurden durch eine Forschungsförderung der Firma Celgene GmbH, München, D, unterstützt. Reagenzien für die FLC-Bestimmung wurden von der Firma The Binding Site, Birmingham, UK, kostenfrei zur Verfügung gestellt.

Bei meinen lieben Kolleginnen Carolin, Henrike, Mirjam, Sonali, Steffi, Denise und Jessy möchte ich mich nicht nur für die fachlichen Hilfen, sondern auch für die stets gute Unterhaltung in den Pausen und für die Gründung der *AG Mittag* bedanken. Danke auch an Jan, meinen persönlichen SAS-Joker. Ebenso ein Dank an Börge Schmidt für die Aufnahme in seine AG. Meinen Eltern möchte ich für Alles danken; dass ihr immer an mich geglaubt und meine Ziele, egal wie verrückt sie auch klangen, unterstützt habt. Meinem kleinen großen Bruder, der einen der wichtigsten Teile meines Lebens ausmacht und mir immer zur Seite steht. Meiner lieben Charlotte, die mich nun schon seit so vielen Jahren als konstante Einheit begleitet, danke für *Immer-da-sein*. Ebenso meiner P(K/M)A Anke, ohne deren motivierenden Nachrichten und aufmunternden Worte ich wohl deutlich später fertig geworden wäre.

Zu guter Letzt, aber nicht weniger wertvoll, danke ich Jan-Philipp; dass du mich am Ende bei Allem unterstützt hast und mein Leben in so vielen Facetten bereicherst!

10 Lebenslauf

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.