

Aus dem Institut für Medizinische Biometrie und Statistik
der Universität zu Lübeck
Direktor: Univ.-Prof. Dr. rer. nat. Andreas Ziegler

Adaptive Tests für SNP-Expressions-Assoziationsstudien

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck

– Aus der Sektion Medizin –

vorgelegt von
Markus Oliver Scheinhardt
aus Siegen

Lübeck, 2016

1. Berichterstatter: Prof. Dr. rer. nat. Andreas Ziegler

2. Berichterstatterin: Prof. Dr. rer. nat. Jeanette Erdmann

Tag der mündlichen Prüfung: 08.05.2017

Zum Druck genehmigt. Lübeck, den 08.05.2017

Promotionskommission der Sektion Medizin

Inhaltsverzeichnis

1	Einleitung	1
2	Adaptive lineare Rangtests für eQTL-Studien	5
2.1	Einleitung	5
2.2	Methoden	6
2.2.1	Adaptive lineare Rangtests	6
2.2.2	Alternative Rangtests	7
2.2.3	Simulationsstudien	7
2.2.4	Anwendungen: LIPA bei koronarer Herzerkrankung und KLC1 bei M. Alzheimer	8
2.3	Ergebnisse	9
2.3.1	Entwicklung eines neuen adaptiven Tests	9
2.3.2	Vergleiche mit anderen Tests	9
2.3.3	Anwendungen	10
2.4	Diskussion	11
3	Lagetests für Biomarker-Studien: Ein Vergleich mittels Simulationen für den Fall von zwei Gruppen	13
3.1	Einleitung	13
3.2	Methoden	13
3.2.1	Verwendete adaptive Tests	13
3.2.2	Simulationsstudien	14
3.3	Ergebnisse	15
3.3.1	Simulationsstudie I: Homogene Varianzen	15
3.3.2	Simulationsstudie II: Heterogene Varianzen	16
3.4	Diskussion	16
4	Quantitative Merkmalsloci der Genexpression im Nebennierengewebe ei- ner Kreuzung aus Holstein und Charolais	18
4.1	Einleitung	18
4.2	Methoden	19
4.2.1	Datenerhebung und Vorbereitung	19
4.2.2	Assoziationsanalysen	20
4.2.3	Pfadanalysen	20
4.3	Ergebnisse	20
4.4	Diskussion	21

5 Diskussion und Ausblick	22
Danksagung	iv
Lebenslauf	v
Publikationsverzeichnis	vi

1 Einleitung

Genomweite Assoziationsstudien (engl.: genome-wide association studies, GWAS) haben maßgeblich zur Identifikation vieler Assoziationen zwischen Einzelnukleotidpolymorphismen (engl.: single nucleotide polymorphisms, SNPs) und komplexen genetischen Erkrankungen beigetragen. Auch bei Nutztieren konnten mit GWAS viele SNPs identifiziert werden, die mit wichtigen ökonomischen Merkmalen assoziiert sind. Mehr als 80% der in GWAS identifizierten SNPs liegen jedoch außerhalb von proteinkodierenden oder bekannten regulatorischen Abschnitten des Genoms. Daher sind die zugrundeliegenden molekularen Mechanismen bislang in vielen Fällen unbekannt.

Ein vielversprechender integrativer genomischer Ansatz zur Aufklärung der funktionalen Wirkung dieser Varianten folgt dem zentralen Dogma der Molekularbiologie. Dieses besagt, dass genetisch kodierte Information im wesentlichen von DNA über RNA zu Protein verläuft, wodurch letztlich bestimmte Merkmale, wie phänotypische Risikofaktoren, determiniert werden. Durch tiefere Einblicke in einzelne Schritte dieses Prozesses könnte es bald möglich sein, den kompletten Pfad genetischer Ursachen für einige Krankheiten abzubilden. Ein wichtiger Bestandteil dabei sind Assoziationen zwischen SNPs und Genexpression, da die Stärke der Genexpression direkt durch Variationen auf genomischer Ebene beeinflusst wird. In SNP-Expressions-Assoziationsstudien, die auch als eQTL-Studien (engl.: expression quantitative trait loci studies, eQTL studies) bezeichnet werden, sollen SNPs identifiziert werden, die das Expressionsniveau von mRNA oder nicht-kodierender RNA beeinflussen.

Dazu werden zunächst Genotypen von genomweit verteilten SNPs und die Expressionswerte vieler bekannter Transkripte in relevantem Gewebe bestimmt. Beide Technologien sind im Hochdurchsatzbereich gut etabliert. Nach Qualitätskontrolle und Vorverarbeitung liegen für jedes Individuum diskrete Genotypdaten (z.B. AA, AG, GG bzw. als 0, 1, 2 kodiert) für die SNPs und quantitative (logarithmierte) Expressionsdaten für die Transkripte vor. Bei genomweiten eQTL-Studien werden Milliarden an Kombinationen von SNPs und Transkripten auf Assoziation hin getestet. Aus statistischer Sicht werden dabei durchschnittliche Expressionsniveaus zwischen Genotypgruppen miteinander verglichen. Wenn sich diese signifikant voneinander unterscheiden, ist die Expression offenbar abhängig vom Genotyp, und

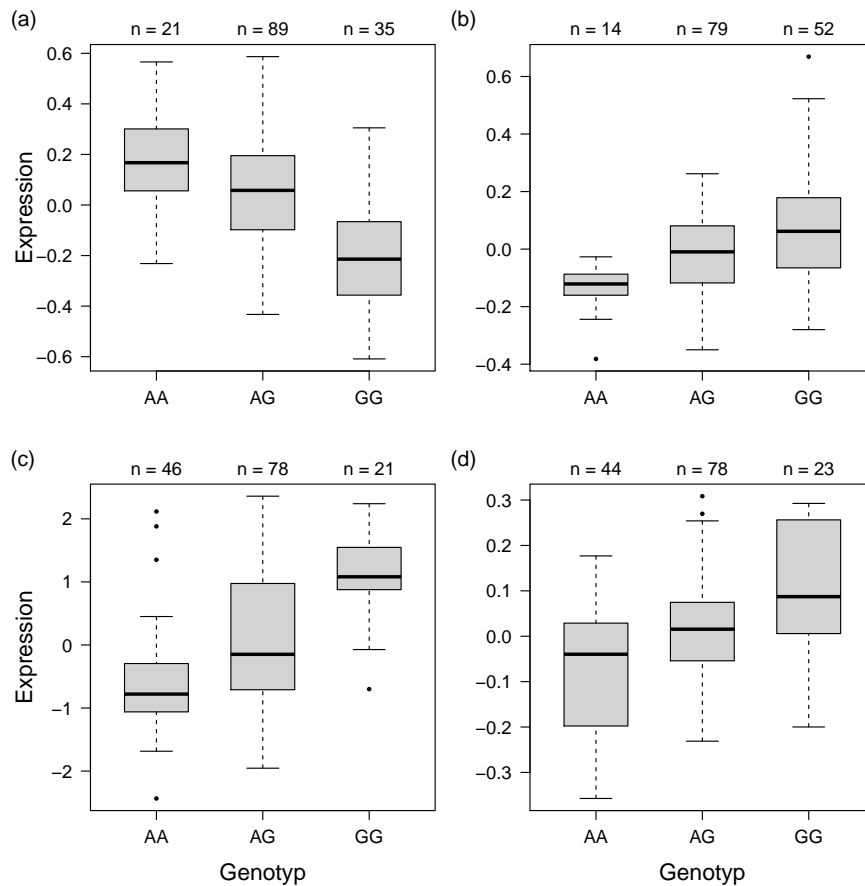


Abbildung 1.1: Expression verschiedener Gene für die drei Genotypgruppen bestimmter Einzelnukleotidpolymorphismen basierend auf Realdaten aus der durchgeführten Studie in Kapitel 4. Die Expressionswerte pro Genotyp können (a) normalverteilt sein mit gleichen Varianzen, (b) normalverteilt sein mit ungleichen Varianzen, (c) schief-verteilt sein oder (d) ganz unterschiedliche Verteilungsformen, z.B. Schiefen in verschiedene Richtungen, haben.

es liegt eine Assoziation vor. In Abbildung 1.1 sind exemplarisch vier zu testende Kombinationen aus Transkripten und SNPs dargestellt.

Statistische Standardverfahren für einen derartigen Vergleich der Lageparameter sind die Varianzanalyse und der Kruskal-Wallis-Test (KW), welche in modernen Software-Paketen laufzeit-effizient implementiert sind. Voraussetzung für die parametrische Varianzanalyse ist eine mindestens annähernde Normalverteilung mit gleichen Varianzen in den Gruppen. Beim KW wird keine Normalverteilung vorausgesetzt, aber die Verteilungsformen in den Gruppen müssen annähernd gleich sein, damit dieser sinnvoll als Lagetest interpretiert werden kann. Expressionswerte sind jedoch trotz Vorverarbeitung häufig nicht-normalverteilt, weisen inhomogene Varianzen auf und haben oft Ausreißer, wie in Abbildung 1.1 zu sehen ist. Solche Verletzungen in den Modellannahmen der Standardverfahren können einerseits zu einem erhöhten Typ-I-Fehler führen, d.h. mehr falsch-positive Ergebnisse, als das vorgegebene Signifikanzniveau eigentlich erlaubt. Andererseits verringert sich da-

durch möglicherweise auch die Güte der Tests, also die Wahrscheinlichkeit wahre Assoziationen zu identifizieren. Da die Genotypverteilung üblicherweise einem Hardy-Weinberg-Gleichgewicht folgen, variieren die Fallzahlen in den Gruppen generell sehr stark, was solche Effekte noch verstärkt. Analog zu eQTL-Analysen können auch Assoziationen von SNPs zu anderen kontinuierlich gemessenen Biomarkern, z.B. Expressionsniveaus von Proteinen oder Metaboliten, untersucht werden. Dabei treten in der Regel die gleichen Herausforderungen auf wie bei Genexpressionsdaten.

Sehr ähnlich sind die Probleme auch bei Fall-Kontroll-Studien in der Frühphase von Biomarker-Studien. Dabei werden die Expressionsniveaus zwischen Fällen und Kontrollen verglichen. Die statistischen Standardverfahren, um auf einen Lageunterschied zwischen zwei Gruppen zu testen, sind der parametrische t -Test und der nicht-parametrische U-Test. Die Voraussetzungen für den t -Test sind analog zu denen der Varianzanalyse und die des U-Tests sind analog zum KW. Auch die Auswirkungen der Abweichungen von den Annahmen sind analog.

Ziel dieser Arbeit war es daher, einen robusteren Lagetest für bis zu drei Stichproben zu finden und auf Gültigkeit und praktische Anwendbarkeit in der Hochdurchsatzsituation von eQTL-Analysen und anderen, ähnlich herausfordernden Studien zu anderen Biomarkern, zu überprüfen. Ein Test wird als robust bezeichnet, wenn er auch bei Abweichungen von den üblichen Voraussetzungen das vorgegebene Signifikanzniveau einhält. Der Test sollte also auch bei nicht-normalverteilten Daten, inhomogenen Varianzen und variierenden Fallzahlen in den Gruppen möglichst viele wahre Assoziationen identifizieren und dabei nicht übermäßig viele falsch-positive Ergebnisse liefern.

Eine generell sehr robuste Methode stellt der Median-Test (MED) dar, der sogar bei ungleichen Verteilungsformen valide Ergebnisse liefert. Doch im Vergleich zu den Standardtests hat dieser eine sehr geringe Güte, falls die Daten normalverteilt sind, und kann für diese und auch andere Situationen nicht empfohlen werden. Permutationstests sind ebenfalls nicht anwendbar, da solche rechenintensiven Verfahren in Anbetracht der immensen Menge an durchzuführenden Tests praktisch nicht durchführbar sind. Bei der großen Menge an untersuchten Biomarkern ist es unmöglich, für jeden potentiellen Biomarker mit seiner individuellen Verteilung manuell den jeweils besten Test auszuwählen. Daher erscheint der Einsatz adaptiver Methoden vielversprechend, die basierend auf Eigenschaften der gemeinsamen Stichprobe automatisiert eine geeignete Teststatistik auswählen bzw. anpassen.

In der Literatur wurden verschiedene infrage kommende Ansätze identifiziert. Diese wurden in der Programmiersprache R effizient implementiert, um in Hochdurchsatzstudien auch tatsächlich anwendbar zu sein. Ihr Verhalten in spezifischen Verteilungssituationen wurde mittels Monte-Carlo Simulationen untersucht.

In *Szymczak et al.* [1] wurde mittels umfangreicher Simulationsstudien ein zweistufiger adaptiver Test für den Einsatz in eQTL-Studien optimiert. In einem ersten Schritt werden durch Selektorstatistiken Schiefe und Länge der Verteilungsschwänze der vorliegenden Daten geschätzt. Dadurch können im zweiten Schritt geeignete Gewichte für eine lineare Rangteststatistik ausgewählt werden. Mittels der g -und- k -Verteilungsfamilie wurden Daten mit unterschiedlichen Verteilungsformen simuliert, um Entscheidungsregeln für die Auswahl der jeweiligen Gewichte zu bestimmen. Der Test wurde in weiteren Simulationsstudien auf Robustheit bei nicht normalverteilten Daten untersucht und mit etablierten adaptiven und nicht-adaptiven Methoden verglichen. Die Methode wurde auf zwei Realdatensätze aus den Bereichen Neurologie und Kardiologie angewendet.

In *Scheinhardt & Ziegler* [2] wurde der neu entwickelte adaptive Test in weiteren Simulationsstudien auf Robustheit und empirische Güte untersucht. Besonders wurde auf den Vergleich der Lageparameter von zwei Gruppen mit inhomogenen Varianzen Wert gelegt. Der neu entwickelte adaptive Test wurde mit weiteren statistischen Methoden verglichen und zeigte sich in vielen simulierten Szenarien gleichwertig oder sogar überlegen gegenüber den Standardmethoden. Allerdings war der Typ-I-Fehler bei schiefen Verteilungen und heterogenen Varianzen zu hoch. Der U-Test erwies sich auch in solchen Szenarien als robust.

In *Brand et al.* [3] wurde der adaptive Test für eine genomweite eQTL-Studie einer F2-Kreuzung aus Milch-Kühen und Fleisch-Rindern eingesetzt. Ziel der Studie war es, Assoziationen von SNPs und Transkripten aus dem Nebennierengewebe zu identifizieren. Es konnten 505 SNPs identifiziert werden, welche mit 129 Transkripten assoziiert waren. Durch Pathway-Analysen konnte gezeigt werden, dass viele der kodierenden Gene tatsächlich eine wichtige Rolle bei der Ausprägung ökonomisch wichtiger Merkmale spielen.

2 Adaptive lineare Rangtests für eQTL-Studien

Szymczak, S., Scheinhardt, M. O., Zeller, T., Wild, P. S., Blankenberg, S. & Ziegler, A. (2013). Adaptive linear rank tests for eQTL studies. *Statistics in Medicine*, 32:524-537.

2.1 Einleitung

Wie in Kapitel 1 beschrieben wurde, sind die Verteilungsannahmen der Standardtests Varianzanalyse und KW in eQTL-Studien oft verletzt, wodurch die damit erhaltenen Ergebnisse nur bedingt valide sind. Daher wird ein statistischer Test benötigt, der den Herausforderungen derartiger Analysen gewachsen ist. In diesem Artikel wurde ein vielversprechender Ansatz eines zweistufigen adaptiven Tests für den Einsatz in eQTL-Analysen optimiert. Diese Art Tests nutzt Informationen der Datenverteilung, um für vorliegende Daten aus einer vorgegebener Schar von Teststatistiken eine geeignete auszuwählen. In einem ersten Schritt werden dazu die Schiefe und die Länge der Verteilungsschwänze geschätzt. Basierend auf diesen Selektorstatistiken wird in einem zweiten Schritt ein geeigneter linearer Rangtest ausgewählt. Diese Strategie scheint ähnlich zum Einsatz eines Vorschalttests zu sein, wodurch gewisse Voraussetzungen des eigentlichen statistischen Tests von Interesse im Voraus geprüft werden. Doch während der Ansatz mit einem p-Wertbasierten Vorschalttest zu einem erhöhten Typ-I-Fehler führt, behält das zweistufige adaptive Verfahren das globale Signifikanzniveau bei, sofern die Selektorstatistiken auf der gemeinsamen Stichprobe berechnet und mit verteilungsfreien Teststatistiken kombiniert werden, da diese dann unabhängig voneinander sind. Zwei Entscheidungsregeln für die Auswahl geeigneter Teststatistiken bei mehr als zwei Gruppen wurden in der Literatur vorgestellt.

Die zur Auswahl stehenden Teststatistiken wurden mittels umfangreichen Simulationen hinsichtlich empirischer Güte und Robustheit für eine Vielzahl von Verteilungsformen miteinander verglichen. Die Familie der g -und- k -Verteilungen wurde verwendet, um zufällige Daten zu generieren, deren Verteilungen auf einem Gitter variierender Schiefen und Längen der Verteilungsschwänze lagen. Anhand der Ergebnisse wurden Entscheidungsregeln für die adaptive Auswahl von Teststatistiken optimiert. Dieser neue adaptive Test wurde hinsichtlich empirischer Güte

mit den beiden etablierten Methoden aus der Literatur verglichen. In einer weiteren Simulationsstudie wurden diese dann mit zwei alternativen Ansätzen linearer Rangtests, lokal mächtigster Rangtest (engl.: locally most powerful rank test, LM-PRT) und maximin Effizienz robuster Rangtest (engl.: maximin efficiency robust rank test, MERT) verglichen. Schließlich wurde die Anwendung aller untersuchten Tests anhand von zwei Realdatensätzen aus den Bereichen Kardiologie und Neurologie aus echten eQTL-Studien demonstriert.

2.2 Methoden

2.2.1 Adaptive lineare Rangtests

Der zweistufige adaptive Test versucht, eine Teststatistik auszuwählen, die eine möglichst große Güte für die vorliegenden Daten hat. In einem ersten Schritt werden Schiefe und Länge der Verteilungsschwänze der Daten geschätzt. Dazu werden Selektorstatistiken für Schiefe S_q und Verteilungsschwanzlänge T_q als Quotienten von Differenzen bestimmter Quantile der gemeinsamen Stichprobe der Expressionswerte berechnet. Der Logarithmus von S_q ist symmetrisch um 0, gleich 0 für symmetrische Verteilungen, positiv für rechts-schiefe Verteilungen und negativ für links-schiefe Verteilungen. Der kleinst-mögliche Wert von T_q ist 1 und größere Werte sind proportional zur Länge der Verteilungsschwänze. Bei normalverteilten Daten ist $\log(S_q) = 0$ und $T_q = 1.704$.

Auf Grundlage dieser Schätzer wird dann eine lineare Rangteststatistik ausgewählt, um den p-Wert zu berechnen. Diese Rangteststatistiken nutzen Linearkombinationen von gewichteten gruppenweisen Rängen der kontinuierlichen Expressionswerte. Die Teststatistiken sind unter der Nullhypothese, also im Fall dass keine Assoziation vorliegt, verteilungsfrei und asymptotisch χ^2 -verteilt mit 2 Freiheitsgraden. Sie können ohne Modifikation auch für zwei oder mehr als drei Gruppen verwendet werden. Durch die Wahl der Gewichte können lineare Rangtests konstruiert werden, die für bestimmte Verteilungsformen besonders gut geeignet sind. Im Allgemeinen hat KW eine große Güte für Verteilungen mit mittellangen Verteilungsschwänzen, Long Tails Test (LT) für lange und Short Tails Test (ST) für kurze. Für rechts-schiefe Verteilungen ist der Rechtsschiefe Test (RS) gut geeignet und analog der Linksschiefe Test (LS) für links-schiefe Verteilungen. Zusätzlich wurde der MED untersucht.

Die beiden Entscheidungsschemata aus der Literatur von Büning und Beier für die Auswahl der Rangtests (Abbildung 2.1 (a) und (b)) unterscheiden sich kaum voneinander. Für Daten mit symmetrischer Verteilung werden je nach geschätzter Verteilungsschwanzlänge ST, KW oder LT ausgewählt. Bei schief-verteiltern Daten RS beziehungsweise LS. Nur im Fall von schiefen Verteilungen und langen Schwänzen werden von den beiden Methoden verschiedene Tests ausgewählt.

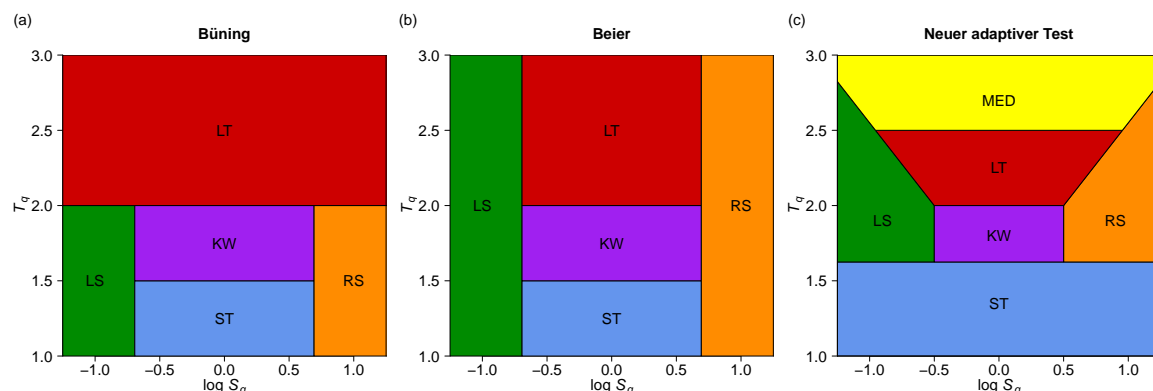


Abbildung 2.1: Entscheidungsschemata der Tests von (a) Büning, (b) Beier und (c) des neu entwickelten adaptiven Tests. Für jedes Wertepaar Schiefe $\log(S_q)$ auf der x-Achse und Länge der Verteilungsschwänze T_q auf der y-Achse ist der auszuwählende lineare Rangtest dargestellt. KW: Kruskal-Wallis-Test; LS: Linksschiefe Test; LT: Long Tails Test; RS: Rechtschiefe Test; ST: Short Tails Test; MED: Median-Test. Adaptiert nach Szymczak et al. (2013), Abbildungen 1 und 4 (Lizenz: 3998131238033).

2.2.2 Alternative Rangtests

Im Gegensatz zu den bereits erwähnten linearen Rangtests werden bei LMPRT und MERT gewisse Verteilungsannahmen getroffen. Entsprechend der vermuteten Verteilung wird deren theoretische Dichte für die Berechnung der Gewichte in der Teststatistik genutzt. Formeln werden bei LMPRT für Normal-, Logistische, Laplace-, Cauchy- und Gumbelverteilung angegeben, bei MERT steht zusätzlich die t -Verteilung zur Verfügung. Bei LMPRT wird nur eine Verteilung angenommen. Wenn die Daten tatsächlich dieser Verteilung unterliegen, so hat LMPRT eine größere Güte als irgendeine verteilungsfreie Methode. Bei MERT werden für die Berechnung der Gewichte eine Mischung von plausiblen Dichten von zwei oder mehr Verteilungen genutzt. Das macht diesen Ansatz flexibler als LMPRT, aber es ist auch eine geringere Güte zu erwarten.

2.2.3 Simulationsstudien

In einer ersten Simulationsstudie wurde ein SNP im Hardy-Weinberg-Gleichgewicht mit variierender Häufigkeit des selteneren Allels (engl.: minor allele fre-

quency, MAF) zwischen 0.5 und 0.1 bei einer Gesamtfallzahl von 600 betrachtet. Zusätzlich wurden Daten mit gleich großen Gruppen erzeugt. Es wurden zufällige Genexpressionsdaten mit einer großen Spanne verschiedener Verteilungsformen generiert. Dafür wurde die Familie der g -und- k -Verteilungen genutzt. Dies sind Transformationen der Quantilfunktion der Standard-Normalverteilung mit Parametern für Lageverschiebung, Streuung, Schiefe und Verteilungsschwanzlänge. Die Parameter wurden so gewählt, dass ein äquidistantes Gitter von Verteilungen verschiedener Kombinationen von Schiefe und Länge der Verteilungsschwänze abgedeckt wird. Unter der Nullhypothese wurden die Daten mit gleichen theoretischen Medianen erzeugt, unter der Alternativhypothese wurden die theoretischen Mediane auf 1, 1.1, und 1.2 für die Genotypen 0, 1, and 2 gesetzt. Die übrigen Parameter blieben gleich. Für jedes Szenario werden 1000 Replikationen durchgeführt um empirische Güte und Typ-I-Fehler zu schätzen. Diese Daten wurden für den Vergleich der Rangtests und der adaptiven Tests verwendet.

In einer zweiten Simulationsstudie wurden die adaptiven Tests mit LMPRT und MERT verglichen. Dazu wurden Expressionswerte für 500 Fälle und 500 Kontrollen simuliert. Es wurden symmetrisch normal, logistisch, Laplace- und Cauchy-verteilte und rechts-schief Gumbel-verteilte Daten generiert. Mittelwertdifferenzen wurden variiert, um empirische Güte und Typ-I-Fehler zu schätzen, bei einem Signifikanzniveau von 0.05 und 10000 Wiederholungen. Für LMPRT wurde jeweils die tatsächlich vorliegende Verteilung für die Berechnung der Gewichte genutzt, und bei MERT wurden verschiedene Mischverteilungen aus jeweils einer Normalverteilung und den sonstigen zur Simulation verwendeten Verteilungen, sowie der t -Verteilung angenommen.

Robustheit wurde nach dem liberalen Kriterium von Bradley definiert. Demgemäß wurde ein Test für ein vorliegendes Szenario als robust betrachtet, wenn der geschätzte Typ-I-Fehler zwischen 0.5α und 1.5α lag, wobei $\alpha = 0.05$ das lokale Signifikanzniveau bezeichnet.

2.2.4 Anwendungen: LIPA bei koronarer Herzerkrankung und KLC1 bei M. Alzheimer

Um die Anwendung der adaptiven Tests bei der Identifikation von eQTL-Assoziationen zu zeigen, wurden zwei Realdatensätze verwendet. Der erste Datensatz umfasst Genotypdaten für den SNP rs1412444, der in GWA-Studien mit koronarer Herzerkrankung assoziiert wurde, und Genexpressionsdaten von Monozyten für

das *LIPA* Gen in 1491 Individuen aus der Gutenberg-Gesundheitsstudie. Für das zweite Beispiel wurden Genotyp- und Genexpressionsdaten zu 193 neuropathologisch gesunden Hirngewebeproben untersucht. Es werden Ergebnisse für eine Assoziation zwischen SNP rs639790 und dem *KLC1* Gen gezeigt, welches eine wichtige Rolle der Entwicklung der Alzheimer-Krankheit spielen könnte.

2.3 Ergebnisse

2.3.1 Entwicklung eines neuen adaptiven Tests

Alle linearen Rangtests hielten das nominale Signifikanzniveau unter der Nullhypothese ein. Die erkennbaren Muster in der empirischen Güte waren bei allen Tests für jede simulierte MAF sehr ähnlich. Generell war für steigende MAF die empirische Güte größer. Sie war für RS und LS bei rechts- bzw. links-schiefe Verteilungen am größten, für ST bei Verteilungen mit kurzen Schwänzen, für KW bei annähernd normalverteilten Daten und für MED bei langen Verteilungsschwänzen. Für LT war sie für alle simulierten Verteilungen ähnlich groß.

Basierend auf den linearen Rangtests mit jeweils größter empirischer Güte für eine MAF von 0.5 auf dem simulierten Gitter von Verteilungen wurde ein neues Entscheidungsschema entwickelt (Abbildung 2.1 (c)), welches einige Unterschiede zu den Tests von Büning und Beier aufweist. Für alle kurzschwänzigen Verteilungen wird ST ausgewählt und für sehr langschwänzige Verteilungen MED statt LT. Für schiefe und langschwänzige Verteilungen verlaufen die Grenzen schräg. KW wird bei Abweichungen von der Form der Normalverteilung zugunsten von ST, RS und LS generell seltener ausgewählt.

2.3.2 Vergleiche mit anderen Tests

Der Vergleich der drei adaptiven Tests ergab, dass der neue Test mit optimiertem Entscheidungsschema auf den simulierten Daten für fast alle Verteilungsformen eine größere empirische Güte von bis zu 20% hat (Abbildung 2.2). Verglichen mit dem Test von Büning ist der neue Test bei kurzschwänzigen und sehr schiefen Verteilungen überlegen. Im Vergleich zum Test von Beier ist er vor allem bei Verteilungen mit kurzen und langen Schwänzen überlegen. Die Unterschiede unter der Nullhypothese sind vernachlässigbar gering.

2.3 Ergebnisse

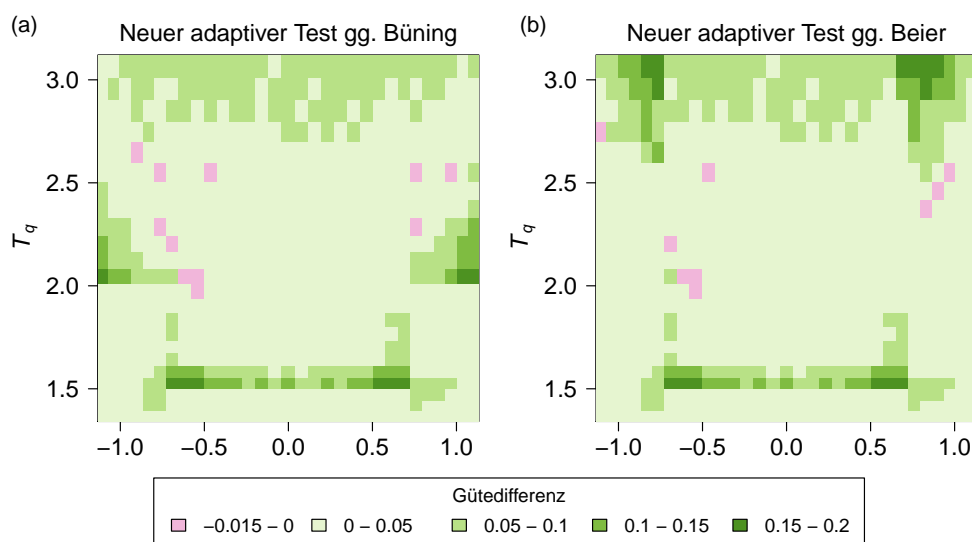


Abbildung 2.2: Gewinn an empirischer Güte des neuen adaptiven Tests beim Vergleich mit den Tests von (a) Bünning und (b) Beier. Die Differenzen für jedes Wertepaar Schiefe $\log(S_q)$ auf der x -Achse und Länge der Verteilungsschwänze T_q auf der y -Achse wurden auf Basis der Ergebnisse für die unter H_1 simulierten Szenarien interpoliert. Adaptiert nach Szymczak et al. (2013), Abbildung 5 (Lizenz: 3998131238033).

In einer zweiten Simulationsstudie wurden die drei adaptiven Tests mit LMPRT und MERT verglichen. Unter der Nullhypothese hielten alle Tests das Signifikanzniveau ein. Die Unterschiede in der empirischen Güte zwischen den adaptiven Tests waren sehr gering. Nur bei Gumbel-verteilten (rechts-schiefen) Daten erwies sich der neue adaptive Test als leicht überlegen. Wie zu erwarten hatte LMPRT mit jeweils optimal gewählten Gewichten die größte empirische Güte von allen Tests. Ebenso hatte MERT eine große Güte bei den simulierten Verteilungen die einer zur Mischung der Gewichte genutzten Verteilungen entsprachen. Der neue adaptive Test hatte dennoch für Cauchy (langschwänzige) Verteilungen eine größere Güte als der jeweils optimale MERT und bei Laplace-verteilten Daten eine ebenso große. In den anderen Szenarien war er nur leicht unterlegen.

2.3.3 Anwendungen

Abbildung 2.3 (a) zeigt, dass Expressionswerte des *LIPA* Gens mit der Anzahl an Risikoallelen T des SNP rs1412144 steigen. Da die Selektorstatistiken ähnlich zu den theoretischen Werten der Normalverteilung waren, wählten die drei adaptiven Tests jeweils KW als lineare Rangteststatistik aus. Die entsprechenden p -Werte waren, so wie bei der Laplace Variante von LMPRT, extrem klein. Bei allen anderen getesteten Varianten von LMPRT und MERT waren sie deutlich größer.

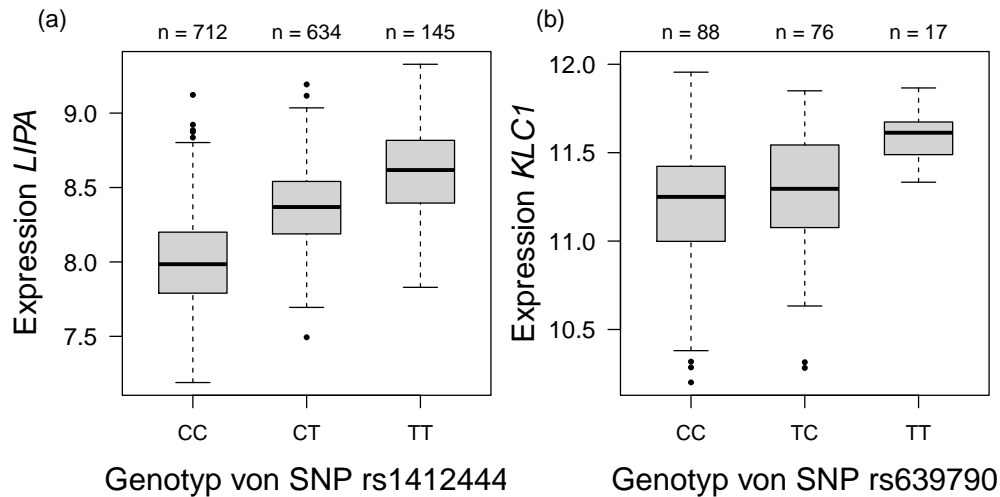


Abbildung 2.3: Expressionswerte (a) des *LIPA* Gens abhängig vom Genotyp des SNP rs1412444, und (b) des *KLC1* Gens abhängig vom Genotyp des SNP rs639790. Adaptiert nach Szymczak et al. (2013), Abbildung 6 (Lizenz: 3998131238033).

Die Expressionswerte des *KLC1* Gens (Abbildung 2.3 (b)) steigen ebenfalls mit der Anzahl an *T* Allelen für SNP rs639790. Die Selektorstatistiken deuteten auf eine links-schiefe Verteilung mit langen Schwänzen hin. Der neue adaptive Test und der Test von Beier nutzten LS als Teststatistik, was zu den kleinsten *p*-Werten führte, Büning nutzte die weniger gut geeignete Teststatistik LT was zu einem größeren *p*-Wert führte. Wie im ersten Beispiel hatten auch hier alle getesteten Varianten von LMPRT und MERT erheblich größere *p*-Werte als der neue adaptive Test.

2.4 Diskussion

Zweistufige adaptive lineare Rangtests sind eine vielversprechende Alternative zur Identifikation von SNPs, die das Expressionsniveau von Genen beeinflussen. Auf Grundlage einer Simulationsstudie mit Verteilungen auf einem Gitter aus verschiedenen Schiefen und Verteilungsschwanzlängen wurde eine neues Entscheidungsschema für die Auswahl an linearen Rangtests entwickelt. Ein Vergleich mit zwei Methoden aus der Literatur zeigt, dass der neue Test im Allgemeinen eine größere empirische Güte hat und die Vorteile der anderen beiden Methoden in sich vereint.

In einer weiteren Simulationsstudie wurden die drei adaptiven Tests mit einer anderen Klasse von linearen Rangtests (LMPRT und MERT) verglichen. Die empirische Güte des neuen adaptiven Tests war in vielen Szenarien gleich oder sogar größer als die des optimalen MERT und nur geringfügig kleiner als die des optimalen LMPRT. Sowohl LMPRT als auch MERT benötigen aber für optimale Eigenschaften Informa-

tionen über die Dichte der vorliegenden Verteilung. Ist die für die Berechnung der Gewichte genutzte Verteilung nicht korrekt spezifiziert, so ist die Güte im Allgemeinen deutlich geringer. Dagegen wird beim adaptiven linearen Rangtest die Verteilungsform aus den vorliegenden Daten geschätzt, wodurch er gut geeignet für große Expressionsstudien ist, bei denen sich die zu testenden genotyp-spezifischen Verteilungen für jedes betrachtete paar SNP-Transkript ändern können.

In beiden Realdatenbeispielen hat der neue adaptive Test die Rangteststatistik größter Güte ausgewählt und lieferte jeweils die kleinsten p -Werte aller betrachteten Tests.

In diesen Simulationsstudien wurden verschiedene Formen symmetrischer und schiefer Verteilungen bei ungleichen Gruppenstärken untersucht. Eine Untersuchung der Typ-I-Fehler bei ungleichen Varianzen oder sogar unterschiedlichen Verteilungsformen in den Genotypgruppen, die zu extremen Verletzungen der Typ-I-Fehler bei Varianzanalyse und KW führen können, steht noch aus.

3 Lagetests für Biomarker-Studien: Ein Vergleich mittels Simulationen für den Fall von zwei Gruppen

Scheinhardt, M. O. & Ziegler, A. (2013). Location Tests for Biomarker Studies: A Comparison Using Simulations for the Two-sample Case. *Methods of Information in Medicine* 52:351-359.

3.1 Einleitung

Auch bei Biomarker-Studien mit kontinuierlich gemessenen Biomarkern, z.B. Expressionsniveaus von Genen, Proteinen oder Metaboliten, bei denen die Niveaus von Fällen und Kontrollen miteinander verglichen werden, sind die Verteilungsannahmen der Standardmethoden t -Test und U-Test oft nicht erfüllt. Daher wird auch für den Vergleich von zwei Gruppen ein robusterer Test auf einen Lageunterschied mit möglichst großer Güte für verschiedene Verteilungssituationen benötigt.

In der Literatur wurden zwei weitere interessante Ansätze für adaptive Tests identifiziert. Diese wurden in Simulationsstudien mit dem neu entwickelten adaptiven Test aus Kapitel 2 und den entsprechenden Standardmethoden im Hinblick auf Güte und Robustheit für zwei Gruppen untersucht. Zur Simulation wurden ebenfalls die g -und- k -Verteilungen verwendet, um Daten mit variierenden Schiefen und Längen in den Verteilungsschwänzen zu generieren. In einer Simulation waren die Varianzen in den Gruppen identisch, in einer weiteren wurden sie unterschiedlich simuliert.

3.2 Methoden

3.2.1 Verwendete adaptive Tests

Der adaptive Test von Szymczak et. al (Szymczak-Test) mit optimierten Entscheidungsregeln wurde in Kapitel 2 bereits ausführlich beschrieben.

Der adaptive Test von Keselman (Keselman-Test) ist eine robustifizierte Version der Varianzanalyse, bzw. des quadrierten t -Tests im Fall von zwei Gruppen. Die Nullhypothese ist wie bei der klassischen Varianzanalyse die Gleichheit der Gruppenmittelwerte. Die Teststatistik nutzt die Modifikation nach Welch. Die geschätzten Parameter werden überdies adaptiv getrimmt, d.h. ein gewisser Teil der extremsten Beobachtungen werden nicht verwendet, was die Teststatistik robuster macht. Normalverteilung wird vorausgesetzt, aber aufgrund der Welch-Modifikation sind inhomogene Varianzen zulässig. Das Berechnen der getrimmten Teststatistik erfolgt in fünf Schritten. Zunächst wird ein globaler Trimmanteil festgelegt (empfohlen werden 15 %). Dann werden mittels Selektorstatistiken die Längen des oberen und unteren Verteilungsschwanzes der gemeinsamen Stichprobe geschätzt, um damit asymmetrisch den oberen und unteren Trimmanteil zu berechnen. Im dritten Schritt werden die Werte sortiert und die größten und kleinsten Werte getrimmt. Viertens werden auf Grundlage der getrimmten Daten gruppenspezifische Mittelwerte und Varianzen geschätzt. Schließlich wird mit diesen Schätzern eine F Teststatistik nach Welch aufgebaut und die entsprechenden Freiheitsgrade berechnet.

Der adaptive Test von O’Gorman (O’Gorman-Test) ist ebenfalls nicht-parametrisch. Dieser Test setzt identische Verteilungsformen in den Gruppen voraus, damit er als Test auf einen Lageunterschied interpretiert werden kann. In einem ersten Schritt werden die beiden Verteilungsschwänze durch quantil-basierte Selektorstatistiken auf der gemeinsamen Stichprobe geschätzt. Diese Schätzer werden in einem zweiten Schritt verwendet, um kontinuierliche Gewichte für die Teststatistik zu berechnen. Für normalverteilte Daten ergeben sich dabei Gewichte wie beim KW, je nach Grad der Schiefe und der Länge der Verteilungsschwänze werden die Gewichte der größten und kleinsten 25 % der Beobachtungen adaptiv transformiert. Damit ergeben sich je nach der Verteilung der Daten Gewichtsfunktionen, die ähnlich zu den bei Szymczak zur Wahl stehenden Rangtests sind. Daher ist ein ähnliches Verhalten der beiden Methoden zu erwarten. Ein Unterschied besteht allerdings darin, dass die Gewichte der Teststatistik von O’Gorman sowie die daraus resultierenden p -Werte kontinuierlich sind. Die starren Grenzen der Gewichte bei Szymczak könnten bei kleinen Veränderungen zu sprunghaften Veränderungen in den p -Werten führen, sofern dadurch ein anderer linearer Rangtest ausgewählt werden würde.

3.2.2 Simulationsstudien

Die Leistung der drei adaptiven Tests und der Standardtests wurde durch zwei Simulationsstudien verglichen. In allen simulierten Szenarien wurde eine Fall-

Kontroll-Studie mit balancierten Gruppen nachgeahmt. Um die Leistung der Tests bei verschiedenen Stichprobengrößen untersuchen zu können, wurde der Gesamt-Stichprobenumfang zwischen 40 und 200 variiert. Unter der Nullhypothese wurden identische Mediane simuliert und unter der Alternativhypothese wurden sie variiert, um Unterschiede in der Güte der Tests deutlich zu machen. Die Anzahl der Replikationen war 10000.

Wie in den Simulationsstudien in Kapitel 2 wurden auch hier die g -und- k -Verteilungen verwendet, um verschiedene Verteilungsformen erzeugen zu können. Die simulierten Verteilungen unterschieden sich dabei in drei Abstufungen der Schiefe von symmetrisch bis stark rechts-schief und in fünf Abstufungen der Länge der Verteilungsschwänze von sehr kurz bis sehr lang. Das Typ-I-Fehler-Niveau α wurde auf 0.05 gesetzt. In einer ersten Simulationsstudie wurden Daten mit homogenen Varianzen in den Gruppen simuliert. Da in biologischen Daten häufig die Variabilität mit steigendem Mittelwert zunimmt, wurden in einer zweiten Simulationsstudie Daten mit inhomogenen Varianzen betrachtet.

3.3 Ergebnisse

3.3.1 Simulationsstudie I: Homogene Varianzen

In der ersten Simulationsstudie hielten alle Tests das nominale Typ-I-Fehler-Niveau ein. Die Unterschiede in der empirischen Güte zwischen den Tests waren bei der größeren Fallzahl von 100 pro Gruppe deutlicher ausgeprägt als bei 20 pro Gruppe. Die Güte war für normalverteilte Daten bei allen Tests außer dem MED ähnlich hoch. Am größten war sie beim t -Test, dicht gefolgt von U-Test und Keselman-Test. Die Güte der parametrischen Verfahren (t -Test und Keselman-Test) war umso kleiner, je schief und langschwänziger die Daten verteilt waren. MED hatte eine große Güte bei Verteilungen mit langen Schwänzen, doch nur geringe Güte bei solchen mit kurzen Schwänzen und größerer Schiefe. Die anderen nicht-parametrischen Tests zeigten generell eine große Güte bei nicht-normalverteilten Daten.

Im Fall sehr langer und sehr kurzer Verteilungsschwänze waren die beiden adaptiven Tests von Szymczak und O’Gorman dem U-Test überlegen. Beim Vergleich dieser beiden Tests hatte der von O’Gorman-Test eine etwas größere Güte bei Verteilungen mit mittleren Schwänzen und der Szymczak-Test bei solchen mit sehr kurzen oder sehr langen Schwänzen.

3.3.2 Simulationsstudie II: Heterogene Varianzen

In der zweiten Simulationsstudie konnte in einigen Szenarien keiner außer dem U-Test den nominalen Typ-I-Fehler einhalten. Dabei war im Allgemeinen der Typ-I-Fehler der anderen Tests größer, je mehr die Daten von der Normalverteilung abweichen, besonders für die größere Fallzahl. Die empirische Güte kann nur für diejenigen Tests sinnvoll verglichen werden, die für ein spezifisches Szenario robust sind, also den Typ-I-Fehler einhalten.

Der t -Test war für sehr schiefe Verteilungen nicht robust. Im Fall langer Verteilungsschwänze war die Güte relativ klein, in den anderen Szenarien groß. Der Test von Keselman war nicht nur bei sehr schiefen, sondern auch bei langschwänzigen Verteilungen nicht robust. In den anderen Szenarien hatte er eine ähnlich große Güte wie der t -Test. Der MED war im Fall sehr langer Verteilungsschwänze nicht valide und hatte nur eine geringe Güte für annähernd normal-verteilte Daten. Die adaptiven Tests von Szymczak und O’Gorman waren insbesondere bei schiefen Verteilungen nicht robust. In den Szenarien wo sie valide waren, hatte der U-Test dennoch eine höhere empirische Güte. Dieser zeigte bei diesen Daten als einziges keine Abweichungen vom Typ-I-Fehler und hatte generell eine große empirische Güte, außer für sehr schiefe Verteilungen.

3.4 Diskussion

In diesem Artikel wurden drei adaptive Tests mit Standardmethoden für den Vergleich der Lage bei zwei Gruppen in Hinblick auf empirische Güte und Robustheit miteinander verglichen. Mittels der g -und- k -Verteilungsfamilie wurden viele relevante Verteilungsformen simuliert, um ein großes Spektrum potentiell auftretender Daten aus Biomarker-Studien abzudecken.

Der parametrische adaptive Test von Keselman, bei dem extreme Beobachtungen getrimmt werden, scheint für nicht-normalverteilten Daten mit homogenen Varianzen besser geeignet zu sein als der t -Test, aber nicht so gut wie die nicht-parametrischen Tests. Im Fall von heterogenen Varianzen sollte dieser Test nur bei normalverteilten Daten verwendet werden, da er in anderen Fällen nicht robust ist. In den Szenarien, wo er valide war hatte er eine größer Güte als die meisten anderen Tests.

Die Ergebnisse für die nicht-parametrischen adaptiven Tests von Szymczak und O’Gorman waren sehr ähnlich. Im Fall von homogenen Varianzen hielten beide Tests den Typ-I-Fehler ein und hatten eine größere Güte für sehr schiefe oder sehr langschwänzige Verteilungen als alle anderen Tests. Auch bei normalverteilten Daten war sie nur wenig geringer als beim t -Test. Bei kleiner Fallzahl hatte der Szymczak-Test eine etwas geringere empirische Güte bei symmetrischen Verteilungen als der O’Gorman-Test. Der Grund ist vermutlich, dass durch den größeren Median in der zweiten Gruppe gelegentlich statt des optimalen KW als Teststatistik RS ausgewählt wurde. Im Fall von heterogenen Varianzen zeigten beide Tests eine auffällige Schwäche. Beide konnten für schiefe Verteilungen nicht das vorgegebene Typ-I-Fehler-Niveau einhalten. Ursächlich dafür sind offenbar die asymmetrischen Gewichtsfunktionen der beiden Tests für schiefe Daten. So werden beispielsweise beim RS nur die kleinsten 50 % der Werte für die Berechnung der Teststatistik verwendet. Dieses Weglassen kann bei inhomogenen Varianzen zu großen Unterschieden in den geschätzten Gruppenmittelwerten unter der Nullhypothese führen, was dann fälschlicherweise zu signifikanten p -Werten führt. Durch diese systematische Schwäche nicht-parametrischer adaptiver Tests mit asymmetrischen Gewichtsfunktionen sind sie für die Anwendung auf Daten mit potentiell heterogenen Varianzen trotz der großen Stärken im Fall homogener Varianzen nicht uneingeschränkt empfehlenswert.

Der U-Test hielt in diesen Simulationsstudien in allen Szenarien das Typ-I-Fehler-Niveau ein und hatte eine recht große Güte außer bei sehr schiefen oder bei sehr langschwänzigen Verteilungen. Wenn die Verteilungsformen sich nicht grundsätzlich unterscheiden und die Unterschiede in den Varianzen nicht zu groß sind dann kann der U-Test empfohlen werden auch wenn sonst keine Annahmen über die Verteilung der Daten angenommen werden können.

Die hier durchgeführten Simulationsstudien haben verschiedene Einschränkungen. Zum Beispiel wurde nur die Situation von zwei balancierten Gruppen untersucht. Ergebnisse für links-schiefe Daten wurden hier nicht gezeigt. Diese sind im Wesentlichen analog zu rechts-schiefen Verteilungen, weshalb die Schlussfolgerungen ebenfalls analog sind. Des weiteren wird davon ausgegangen, dass die Erweiterungen von t -Test und U-Test für mehr als zwei Gruppen, Varianzanalyse und KW, in entsprechenden Verteilungssituationen die gleichen Stärken und Schwächen haben, wie diese.

4 Quantitative Merkmalsloci der Genexpression im Nebennierengewebe einer Kreuzung aus Holstein und Charolais

Brand, B., Scheinhardt, M. O., Friedrich, J., Zimmer, D., Reinsch, N., Ponsuksili, S., Schwerin, M. & Ziegler, A. (2016). Adrenal cortex expression quantitative trait loci in a German Holstein \times Charolais cross. *BMC Genetics*, 17.1:135.

4.1 Einleitung

Ergebnisse aus eQTL-Studien sowie die Integration von Proteom- und Metabolomdaten in genetische Analysen haben viel zum Verständnis von Genen und molekularen Netzwerken bei Menschen und Nutztieren beigetragen. Bei Rindern wurde von den meisten bisherigen QTL und GWA Studien der genetische Hintergrund komplexer Merkmale wie Milchleistung oder die Reproduktion beleuchtet. Aufgrund des wachsenden Interesses an Tierschutz und dem Einfluss von Stress auf wichtige ökonomische Merkmale bei Nutztieren ist die Nebenniere mit seinen Funktionen für die Reaktion auf Stress zunehmend wichtiger geworden. Stressabhängige Effekte auf die Produktivität, die Gesundheit und die Fortpflanzungsfähigkeit von Rindern sind noch immer nicht vollständig verstanden. Ein besseres Verständnis von molekularen Netzwerken die einen Einfluss auf solche Merkmale bei Rindern haben, könnte sogar Licht auf die Beziehung von Stress und Fettleibigkeit bzw. von stressabhängigen Effekten auf die Gesundheit des Menschen werfen.

Die Nebenniere ist ein vielversprechendes Ziel für die Untersuchung solcher Effekte, da sie sowohl einen großen Einfluss auf die Reaktion des Körpers auf Stress als auch auf wichtige Stoffwechselprozesse hat. Menschliche Krankheiten wie die Addisonsche Krankheit, das Cushing-Syndrom und Primärer Hyperaldosteronismus, die alle mit abnormalen Niveaus an Glucocorticoiden oder Mineralocorticoiden in Verbindung stehen, sind ebenfalls ein Indiz dafür. Jüngere Studien in Rindern und Schweinen zeigten zum Beispiel Assoziationen zwischen genetischen Varianten aus dem Glucocorticoidrezeptor und dem Mineralocorticoidrezeptor mit dem Cortisolspiegel und der Fleischqualität.

Daher wurde der genetische Hintergrund für die Genexpression der Nebennierenrinde in der experimentellen F2 Rinderpopulation SEGFAM untersucht. SEGFAM ist eine Kreuzung aus reinrassigen Deutschen Holstein-Kühen und ebenfalls reinrassigen Charolais-Bullen. Holstein ist eine Rasse, die für hohe Milcherträge gezüchtet wird, wohingegen bei der Züchtung von Charolais das Fleisch im Fokus steht. In der F2 Generation einer Kreuzung so unterschiedlicher Züchtungen ist eine große genetische und phenotypische Variabilität zu erwarten. Von der Identifikation von eQTL-Assoziationen in dieser Population sind daher neue Einblicke in die genetische Regulierung der Genexpression im Zusammenhang mit ökonomisch wichtigen Merkmalen und stressabhängigen Stoffwechselprozessen zu erwarten. Die beteiligten Gene könnten interessante Objekte für weiterführende Studien darstellen.

In dieser Studie wurden zunächst mittels des Szymczak-Tests eQTL-Assoziationen im Gewebe der Nebennierenrinde in Rindern identifiziert. Danach wurde überprüft, inwiefern etwas über den Einfluss der assoziierten Gene und molekularen Signalwege auf ökonomisch wichtige Merkmale in Rindern bekannt ist.

4.2 Methoden

4.2.1 Datenerhebung und Vorbereitung

Der für die Analyse genutzte Datensatz umfasste Genotypen von 37204 SNPs und Expressionswerte von 10986 Transkripten aus 145 F2 Kühen der experimentellen SEGFAM Population, die am Leibniz-Institut für Nutztierbiologie (FBN) gezüchtet wurde. Die DNA wurde aus Milchdrüsengewebe extrahiert und mit einem speziell für Rinder designten SNP-Chip gemessen. Aus Gewebe der Nebennierenrinde wurde RNA extrahiert und mit einem Expressions-Chip, der ebenfalls speziell für Rinder designt wurde, gemessen. Sowohl Genotyp- als auch Expressionsdaten wurden nach relativ konservativen Regeln qualitätskontrolliert und vorverarbeitet.

Aufgrund der familiären Struktur waren die Expressionsdaten zunächst korreliert. Da der statistische Test jedoch unkorrelierte Daten voraussetzt, wurden die Expressionsdaten im Rahmen der Vorverarbeitung mittels einer Serie von univariaten Analysen unter Anwendung eines Sire-Dam-Modells um die systematischen Einflussgrößen Schlachtalter, Schlachtzeitpunkt und Familienstruktur bereinigt. Die Residuen aus diesem Modell enthalten für jedes Tier eine unkorrelierte genetische Komponente sowie eine Zufallsabweichung. Die Residuen können damit als statistisch

unabhängig betrachtet werden und wurden für die Assoziationsanalysen anstelle der rohen Expressionsdaten verwendet.

4.2.2 Assoziationsanalysen

Für die genomweite eQTL-Analyse wurde der adaptive Szymczak-Test verwendet, der speziell für derartige Analysen optimiert wurde. Aufgrund der großen Menge an durchzuführenden Tests (374 Millionen) wurden die resultierenden p -Werte mit der Methode von Benjamini-Hochberg adjustiert und ein globales Signifikanzniveau von 0.1 festgelegt. Dieser Schwellenwert ist für solche Assoziationen, bei denen sich SNP und Transkript in genomischer Nähe zueinander befinden, wesentlich konservativer als bei vielen anderen Studien üblich.

4.2.3 Pfadanalysen

Die in der eQTL-Analyse identifizierten, assoziierten Gene wurden mit Hilfe der Software Ingenuity[®] Pathway Analysis (IPA) näher untersucht, um ihre biologische Relevanz in ausgewählten biologischen Netzwerken aufzuklären. In einer funktionalen Analyse wurde überprüft, ob diese Gene im Zusammenhang mit spezifischen molekularen oder zellulären Funktionen im Vergleich mit einer Referenzmenge von Genen über-repräsentiert sind. In einer ähnlichen Upstream Analyse wurde überprüft, ob es eine Überlappung der assoziierten Gene mit solchen Genen gibt, die von spezifischen Transkriptionsfaktoren reguliert werden.

4.3 Ergebnisse

In der eQTL-Analyse wurden insgesamt 505 SNPs identifiziert, die mit 129 verschiedenen Transkripten assoziiert waren. Bei den meisten dieser Assoziationen (482 von 505) liegen SNP und Gen auf demselben Chromosom und nur bei 17 auf unterschiedlichen. Bei sechs Assoziationen war die Position der SNPs nicht bekannt. Aufgrund des hohen Kopplungs-Ungleichgewichts in dieser untersuchten F2 Population ist die sonst übliche Einteilung in cis- und trans-Effekte hier nicht sinnvoll.

Wie die funktionale Pfadanalyse zeigte, sind die Gene der 129 regulierten Transkripte an verschiedenen molekularen und zellulären Funktionen beteiligt. Dazu gehören Einfangen freier Radikale, zellulärer Ausgleich, Zellmorphologie, Fettstoffwechsel und Biochemie kleiner Moleküle. Ebenso wurden einige Krankheiten und Biofunktionen identifiziert, bei denen diese Gene eine Rolle spielen. Die wichtigs-

ten davon waren Fettstoffwechsel, Entzündungsreaktion sowie Entwicklung und Funktion des Fortpflanzungssystems.

Die Upstream-Analyse identifizierte einige wichtige vorgeschaltete und vier Haupt-Regulierer. Die meisten der in dieser Analyse identifizierten Krankheiten, Biofunktionen, kausalen Netzwerke und vorgeschalteten Regulierer standen mit relativ wenigen Genen im Zusammenhang.

4.4 Diskussion

In dieser Studie wurde der genetische Hintergrund der Genexpression in der Nebennierenrinde einer experimentellen F2 Population einer Kreuzung der Rinderrassen Holstein und Charolais untersucht, um molekulare Netzwerke zu identifizieren, welche potentiell die phänotypische Variation ökonomisch wichtiger Merkmale steuern. Dabei wurden 505 SNPs identifiziert, welche die Expression von 129 Genen beeinflussen. Eine Pfadanalyse ergab, dass ein Teil dieser Gene tatsächlich eine wichtige Rolle in molekularen Netzwerken spielen, die einen Einfluss auf ökonomisch wichtiger Merkmale in Rindern haben. Andere Studien belegen dies ebenfalls für einen Teil der Gene im Zusammenhang mit Fortpflanzungsfähigkeit, Milchleistung sowie Wachstums- und Gesundheitsmerkmalen in Rindern und anderen Nutztieren. In anderen Studien in der SEGFAM Population wurde gezeigt, dass einige der 129 Gene das Verhalten, das Temperament und das Wachstum der Tiere zu Pupertätsbeginn beeinflussen. Für andere Gene konnten immerhin Zusammenhänge mit krankheitsspezifischen Phänotypen oder spezifischen Stoffwechselnetzwerken in Menschen und Mäusen festgestellt werden, was nahelegt, dass ähnliche Mechanismen auch bei Rindern existieren.

Auch wenn diese Hinweise auf den Einfluss der identifizierten Gene auf wichtige ökonomische Merkmale plausibel sind, muss dennoch in Betracht gezogen werden, dass die unterschiedlichen Expressionsniveaus in der Nebennierenrinde nicht die Ursache dafür sein müssen. Aber es mag in anderen Organen ähnliche Effekte geben, wodurch die Merkmale tatsächlich beeinflusst werden. Insgesamt sprechen diese Ergebnisse dafür, dass die eQTL-Analyse ein nützlicher Ansatz ist, um bessere Einblicke in den molekularen und genetischen Hintergrund komplexer Merkmale zu erlangen. Natürlich sind weitere Studien notwendig, um die tatsächlich kausalen SNPs zu identifizieren und die zugrundeliegenden molekularen Mechanismen und Effekte von genetischer Variation auf Genexpression in anderen Geweben aufzuklären.

5 Diskussion und Ausblick

In Biomarker-Studien mit kontinuierlich gemessenen Biomarkern, wie z.B. eQTL-Studien, stellen die Verteilungen der Daten besondere Herausforderungen an statistische Tests. Aus praktischen Gründen spielt jedoch die benötigte Rechenzeit bei der Wahl der Methode für manche Analysen eine Rolle. Bei genomweiten eQTL-Analysen z.B. kann dies je nach Implementierung durchaus einen Unterschied von mehreren Wochen ausmachen. Da die Standardmethoden besonders effizient implementiert sind, werden diese überwiegend verwendet. Doch die Ergebnisse der Standardmethoden wie t -Test und Varianzanalyse sind zum Teil nicht valide, da ihre Verteilungsannahmen oft nicht erfüllt sind. Daher wird eine robustere Alternative benötigt.

In dieser Arbeit wurden verschiedene adaptive und nicht-adaptive Ansätze in Simulationsstudien miteinander verglichen. Dazu wurde ein R-Paket implementiert, das es ermöglicht, die statistischen Eigenschaften infrage kommender Tests unter unterschiedlichen Bedingungen zu prüfen. Die Anzahl der Gruppen und die Fallzahlen in den einzelnen Gruppen können damit flexibel variiert werden. Die Verwendung der g -und- k -Verteilungsfamilie ermöglicht es, dass sich die simulierten Daten in den zu vergleichenden Gruppen im Grad und der Richtung der Schiefe, der Länge ihrer Verteilungsschwänze, in ihren Varianzen und natürlich auch im Mittel unterscheiden können.

In mehreren Simulationen zeigte sich, dass adaptive Tests generell eine gute Alternative zu üblicherweise eingesetzten Standardmethoden darstellen. Der für eQTL-Analysen optimierte adaptive Szymczak-Test zeigte bei simulierten Daten in verschiedenen Szenarien eine größere empirische Güte als die meisten anderen überprüften Tests. Auch die Anwendung auf mehrere Realdatensätze lieferte biologisch plausible Ergebnisse. Es zeigten sich jedoch auch deutliche Schwächen im Hinblick auf Robustheit bei schief-verteiltern Daten mit inhomogenen Varianzen. Daher ist der Szymczak-Test in dieser Form nicht uneingeschränkt empfehlenswert.

Eine einfache Möglichkeit zur Behebung dieses Problems wäre, die zur Verfügung stehenden linearen Rangtests ausschließlich auf solche mit symmetrischen Gewichtsfunktionen (ST, KW, LT und MED) zu beschränken. Dieser modifizierte adaptive Test wäre vermutlich nach wie vor den nicht-parametrischen Standardmethoden U-Test und KW für lang- und kurz-schwänzige Verteilungen überlegen,

die sich selbst angesichts inhomogener Varianzen bei balancierten Gruppen als robust erwiesen. Weitere Modifikationen des Szymczak-Tests wie z.B. die Verwendung von LMPRT oder MERT im zweiten Schritt erscheinen verlockend, da diese für viele spezifische Verteilungen eine größere Güte haben als die hier verwendeten linearen Rangteststatistiken. Doch da diese Rangtests theoretische Dichten für die Berechnung der Gewichte nutzen, sind sie nicht verteilungsfrei. Deshalb wäre es auf theoretischer Basis nicht mehr gewährleistet, dass das zweistufige Verfahren das globale Signifikanzniveau einhält. Es ist allerdings schwer einzuschätzen, ob diese Einschränkung tatsächlich auch von praktischer Relevanz ist.

Der U-Test erschien in den Simulationsstudien aus Kapitel 3 besonders robust. In später durchgeführten Simulationen zeigte sich jedoch, dass auch dieser nicht mehr das Typ-I-Fehler-Niveau einhält, wenn Daten mit inhomogenen Varianzen in unbalancierten Gruppen vorliegen. Wie in Kapitel 1 beschrieben, ist das bei eQTL-Analysen jedoch sehr häufig der Fall. Sowohl der KW als auch alle hier untersuchten adaptiven Tests haben diesbezüglich dasselbe Problem. Es konnte inzwischen in der Literatur ein vielversprechender nicht-adaptiver Test identifiziert werden. Dieser Test ähnelt dem Keselman-Test, mit dem Unterschied, dass die Schätzer dabei nicht adaptiv getrimmt werden und der Trimmanteil etwas höher ist. In Simulationen war dieser sogar für Daten in unbalancierten Gruppen mit inhomogenen Varianzen robust und hatte dabei eine größere Güte als der MED. Für Daten mit unterschiedlichen Schiefen in den einzelnen Gruppen war jedoch das Typ-I-Fehler-Niveau etwas zu hoch. Dieser Test wurde effizient implementiert und die Veröffentlichung der entsprechenden Simulations-Ergebnisse ist geplant.

Moderne Verfahren zur Messung von Genotypen und Genexpression, DNaseq und RNAseq, führen zu ganzzahligen Daten. Da bei DNaseq die Anzahl an Genotypen direkt geschätzt wird, können die hier beschriebenen Methoden ohne Modifikation verwendet werden. Bei RNAseq wird statt eines kontinuierlichen Niveaus die Anzahl an Transkript-Kopien gezählt. Die in dieser Arbeit verwendeten nicht-parametrischen Methoden können direkt mit Zähldaten verwendet werden, wohingegen die Expressionsdaten für den Einsatz mit parametrischen Methoden zunächst im Rahmen der Vorverarbeitung in kontinuierliche Werte überführt werden müssen.

Danksagung

An dieser Stelle möchte ich mich für die großartige Hilfe bedanken, die mir von vielen Seiten zuteil wurde. Insbesondere möchte ich meinem Doktorvater Herrn Univ.-Prof. Dr. rer. nat. Andreas Ziegler für die ausgezeichnete wissenschaftliche Anleitung danken. Durch konstruktive Vorschläge und praktische Hilfe hat er maßgeblich zum Gelingen der Arbeit beigetragen. Danke auch für die vielen guten Hinweise zum Schreiben und Präsentieren, sowie die angenehme Gesellschaft beim Betriebssport. Auch bei Frau Univ.-Prof. Dr. rer. biol. hum. Inke R. König möchte ich mich dafür bedanken, dass sie immer ein offenes Ohr und gute Tipps für mich hatte.

Im weiteren danke ich allen Co-Autoren für die angenehme Zusammenarbeit. Dies gilt insbesondere für Frau Dr. rer. nat. Silke Szymczak und Herrn Dr. rer. nat. Bodo Brand.

Besonderer Dank geht an meine Kollegen für die angenehme Zusammenarbeit und die netten Gespräche, allen voran Herrn Dr. rer. hum. biol. Arne Schillert. Herrn Marvin Wright, M.Sc. danke ich außerdem für die umfangreiche praktische Hilfe beim Erstellen dieser Arbeit und Herrn Dominik Thiele, M.Sc. für die Korrektur der Einleitung.

Schließlich danke ich meiner Frau Muriel fürs Korrekturlesen und ihre persönliche Unterstützung.

Lebenslauf

Persönliche Daten

Name Markus Oliver Scheinhardt

Geburtsjahr 1983

Berufliche Tätigkeit

seit 2010 **Wissenschaftlicher Mitarbeiter**
Institut für Medizinische Biometrie und Statistik
Universität zu Lübeck
Universitätsklinikum Schleswig-Holstein, Campus Lübeck

Studium

seit 2010 **Promotionsstudium**
Biostatistik
Universität zu Lübeck

2006 – 2010 **Master of Science**
Computational Life Science
Universität zu Lübeck

2003 – 2006 **Bachelor of Science**
Computational Life Science
Universität zu Lübeck

Publikationsverzeichnis

Zeitschriftenartikel

- [1] Szymczak, S., **Scheinhardt, M. O.**, Zeller, T., Wild, P. S., Blankenberg, S. & Ziegler, A. (2013). Adaptive linear rank tests for eQTL studies. *Statistics in Medicine*, 32:524-537.
- [2] **Scheinhardt, M. O.** & Ziegler, A. (2013). Location Tests for Biomarker Studies: A Comparison Using Simulations for the Two-sample Case. *Methods of Information in Medicine* 52:351-359.
- [3] Brand, B., **Scheinhardt, M. O.**, Friedrich, J., Zimmer, D., Reinsch, N., Ponsuksilli, S., Schwerin, M. & Ziegler, A. (2016). Adrenal cortex expression quantitative trait loci in a German Holstein \times Charolais cross. *BMC Genetics*, 17.1:135.
- [4] Andlauer, T. et.al (2016). Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Science Advances*, 2(6):e1501678.

Kongressbeiträge

Vorträge

- [5] Ziegler, A., **Scheinhardt, M.O.** & Szymczak, S. (2012). Adaptive linear rank tests for eQTL studies *European Mathematical Genetics Meeting, 2012*. Göttingen, Deutschland.
- [6] **Scheinhardt, M.O.**, Szymczak, S. & Ziegler, A. (2012). Adaptive Tests für Biomarker-Studien. *Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V., 2012*. Braunschweig, Deutschland.

Poster

- [7] **Scheinhardt, M.O.**, Brand, B., Zimmer, D., Reinsch, N., Schwerin, M. & Ziegler, A. (2014). EQTL and pathway analysis on expression profiles of a cattle cross *Annual Scientific International Genetic Epidemiology Society Meeting, 2014*. Wien, Österreich.