

Methoden zum Untersuchen von Gen-Strata Interaktionen in genomweiten Assoziations-Metaanalysen am Beispiel der Adipositas¹

Thomas Winkler²

Abstract: Genomweite Assoziations Meta-Analysen (GWAMAs) haben wesentlich zum besseren Verständnis der Genetik von komplexen Krankheiten, wie z.B. Adipositas, beigetragen. Allerdings ist wenig darüber bekannt, ob sich genetische Effekte zwischen Subgruppen, wie z.B. zwischen Männern und Frauen unterscheiden. Die vorliegende Arbeit befasste sich daher mit stratifizierten GWAMA Ansätzen zum Identifizieren von genetischen Varianten, deren Effekte sich signifikant zwischen Subgruppen unterscheiden. Zwei leistungsfähige und einfach zu verwendende Softwarepakete namens EasyQC und EasyStrata wurden entwickelt, welche die Durchführung, Qualitätskontrolle und Auswertung von stratifizierten GWAMAs ermöglichen. Die Softwarepakete werden von verschiedenen genetischen Forschungskonsortien verwendet und haben dadurch bereits erfolgreich zu einem besseren Verständnis der genetischen Effekte auf Krankheiten oder auf krankheitsrelevante Merkmale, sowie deren Gruppenabhängigkeiten, beigetragen.

1 Einführung

Das Ziel von Genetischen Assoziationsstudien ist das Aufspüren von Assoziationen zwischen genetischer Variation und Krankheiten, wie z.B. Typ 2 Diabetes, oder krankheitsrelevanten Parametern, wie z.B. Body Mass Index (BMI, einem Maß für Fettmasse) oder Taille-Hüft-Verhältnis (THV, einem Maß für Fettverteilung).

Die häufigste Form genetischer Variation sind sogenannte Single Nucleotide Polymorphisms (SNPs), welche den Austausch einer einzelnen Base an einer bestimmten Stelle im menschlichen Genom bezeichnen. Technische Fortschritte im Bereich der Hochdurchsatz- und Microarraytechnologie Mitte der 2000er Jahre, haben dazu geführt, dass mehrere Millionen SNPs in einer einzigen Messung erfasst werden können [HD05].

Seither haben genomweite Assoziationsstudien (GWAS), bei denen Millionen von SNPs parallel auf Assoziation getestet werden, wesentlich zum besseren Verständnis der Genetik von komplexen Krankheiten beigetragen [We14]. Da einzelne GWAS durch ihre Stichprobengröße limitiert sind und sehr kleine genetische SNP Effekte nur mit sehr großen Stichprobenzahlen detektiert werden können, haben sich im Laufe der letzten Jahre groß angelegte genetische Forschungskonsortien zusammengefunden. Diese haben sich zum Ziel gesetzt, mehrere GWAS in sogenannten genomweiten Assoziations-Metaanalysen (GWAMAs) zusammen auszuwerten, um somit Stichprobenzahl und Power der Auswertung gleichermaßen zu erhöhen.

¹ Englischer Titel der Dissertation: "Methods to investigate gene-strata interaction in genome-wide association meta-analyses on the example of obesity"

² Institut für Epidemiologie und Präventivmedizin, Universität Regensburg, thomas.winkler@ukr.de

Eines der größten GWAMA Konsortien weltweit ist das Genetic Investigation of ANthropometric Traits (GIANT) Konsortium, ein Zusammenschluss von mehr als 125 einzelnen GWAS für anthropometrische Maße, welche zu einer Gesamtstichprobengröße von mehr als 330,000 Personen beitragen (www.broadinstitute.org/collaboration/giant). Aktuelle GIANT GWAMAs für Adipositasparameter haben eine Vielzahl an assoziierten Genorten für BMI (97 Genorte) und THV (49 Genorte) gefunden [Lo15, Sh15]. Ein Beispiel für die Darstellung eines GWAMA Ergebnisses für THV ist in **Abbildung 1** gezeigt.

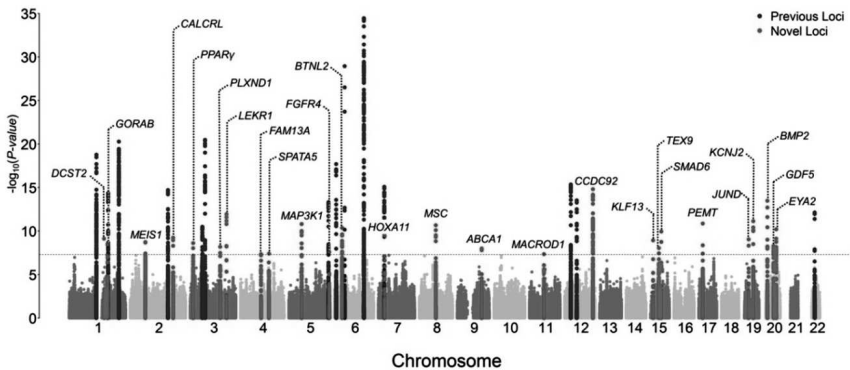


Abb. 1: Manhattan-Plot des GIANT GWAMA Ergebnisses für THV (adjustiert für BMI, aus [Sh15]). Der Plot zeigt Assoziations P-Werte für 2.8M untersuchte SNPs auf einer $-\log_{10}$ Skala über chromosomaler Basenposition des jeweiligen SNPs. Mit THV assoziierte Genorte sind farblich hervorgehoben (blau: bereits bekannte Genorte, rot: neu gefundene Genorte).

Weitere besondere Anforderungen, wie das stratifizierte Auswerten jeder einzelnen GWAS und GWAMA nach Geschlecht, Rauchverhalten, körperlicher Aktivität oder Alter erhöhen die allgemeine Komplexität der GIANT GWAMAs. Ziel der stratifizierten GWAMAs ist, Genorte zu detektieren, an denen sich die genetischen Effekte signifikant zwischen den beiden Gruppen, also z.B. zwischen Männern und Frauen, unterscheiden. Um Effekt-Unterschiede möglichst effektiv zu finden, wurde im Rahmen dieser Doktorarbeit, erstmals eine systematische, methodische Evaluierung von stratifizierten GWAMA Ansätzen zum Identifizieren von Effekt-Unterschieden durchgeführt. Verschiedene Ansätze wurden mittels Simulationen und analytischen Berechnungen hinsichtlich Typ 1 Fehler und Power unter realistischen Szenarien verglichen. Sämtliche statistische Methoden zum Auswerten einer stratifizierten GWAMA wurden in einem R Softwarepaket namens EasyStrata implementiert [Wi14b].

Zusammen wurden in aktuellen GIANT GWAMAs für BMI und THV (allgemeine GWAMAs zur Detektion von assoziierten Genorten [Lo15, Sh15], sowie alters- und geschlechts-stratifizierter GWAMAs zur Detektion von Alters- und Geschlechtsunterschieden [Wi15a]) mehr als 3,000 einzelne GWAS Ergebnisse zusammengetragen und gemeinsam ausgewertet. Da jedes einzelne GWAS Ergebnis 2.8 Millionen Zeilen von SNP-spezifischen Assoziationsergebnissen enthielt, wurden bis zu 8.4 Milliarden Datenpunkte von Assoziationsergebnissen ausgewertet. Um eine valide und effektive Durchführung der hochdimensionaler GWAMAs zu gewährleisten, wurden im Rahmen dieser Doktorarbeit, Prozeduren

zur Qualitätskontrolle (QC) der multiplen big GWAS Datensätze entwickelt und in einem R Softwarepaket namens EasyQC implementiert [Wi14a].

Im Folgenden werden die Entwicklung und die Methodik der beiden R Softwarepakete EasyStrata und EasyQC, sowie deren Anwendung im Rahmen des GIANT Konsortiums beschrieben [Wi15b].

2 Entwicklung der Easy Softwarepakete

Beide Softwarepakete, EasyStrata und EasyQC, wurden in R mittels objektorientierter Methoden programmiert und in das eigens entwickelte Easy-Framework integriert. Das Easy-Framework exportiert eine einzelne R Funktion, entweder EasyStrata() oder EasyQC(), über welche auf die implementierten Methoden zugegriffen werden kann. Zum Starten des Programms wird die jeweilige Funktion in der R Befehlszeile aufgerufen und der Pfad zu einer individuell angepassten ecf-Skriptdatei als Funktionsparameter angegeben. Die ecf-Skriptdatei ist eine einfache Textdatei, welche aus zwei Teilen besteht: Einem Konfigurationsteil, welcher Daten Input/Output bestimmt und einer Skriptschnittstelle, welche die auszuführenden Methoden und Befehle definiert und aneinanderreihet.

Bei der Entwicklung des Easy-Frameworks wurden in besonderem Maße objektorientierte Programmiermethoden verwendet. Jede über die Skriptschnittstelle zur Verfügung gestellte Methode, wurde als eigene Klasse im Framework verankert. Jede Klasse besitzt Funktionen zum Übersetzen der Skriptparameter in Variablen, zur Validierung der eingegebenen Parameter und zum Ausführen der jeweiligen Methode. Die einheitliche Struktur der verschiedenen Methoden-Klassen ermöglichen dem Entwickler eine einfache Aktualisierung und Erweiterung der Pakete um neue Methoden, welche dem Benutzer über die Skriptschnittstelle voll-parametrisiert zur Verfügung gestellt werden.

Um multiple big GWAS Datensätze möglichst effizient zu verarbeiten, verwendet das Easy-Framework Methoden des data.table R Paketes. Dieses bietet state-of-the-art Funktionen zum Einlesen, Kombinieren und Verarbeiten von sehr großen Datensätzen in R. Das data.table Konzept wurde in einem eigenen Datentyp, namens GWADATA, verankert. Dieser stellt verschiedene Datenverarbeitungsfunktionen bereit und wird innerhalb des Easy-Frameworks als Standard Datentyp für die großen GWAS Datensätze verwendet.

Das ecf-Skriptkonzept erlaubt dem Benutzer verschiedenste zur Verfügung gestellte Grafik- und Statistikmethoden in flexibler Weise aneinanderzureihen und iterativ auf multiple GWAS oder GWAMA Datensätze anzuwenden. Dies vereinfacht nicht nur die Handhabung der multiplen 'big GWAS' Datensätze, sondern ermöglicht zugleich die Entwicklung von individuell angepassten QC Skripten und statistischen Auswertungsroutinen. Eine Reihe an vordefinierten ecf-Skripten zur statistischen Auswertung von stratifizierten GWAMAs oder zur GWAMA-QC werden auf der Website www.genepi-regensburg.de/software zur Verfügung gestellt.

3 Evaluierung statistischer Methoden für stratifizierte GWAMAs und Entwicklung der EasyStrata-Software zur Detektion von Effekt-Unterschieden

3.1 Methoden und Implementierung

Die vorliegende Arbeit befasste sich in besonderem Maße mit Methoden, die basierend auf einer stratifizierten GWAMA (z.B. einer Auswertung getrennt nach Geschlecht) zum Identifizieren von genetischen Varianten verwendet werden können, deren Effekte sich signifikant zwischen zwei Gruppen unterscheiden (z.B. zwischen Männern und Frauen).

Hierfür wurde zunächst eine Reihe an statistischen Tests identifiziert, die auf Ergebnisse einer stratifizierten GWAMA anwendbar sind (**Tabelle 1**). Diese Tests wurden in verschiedener Weise kombiniert und systematisch hinsichtlich Typ 1 Fehler und Power zum Detektieren von signifikanten Effekt-Unterschieden untersucht. Dabei wurden unterschiedliche, realistische Szenarien an Stichproben-, Gruppen- und Effektgrößen betrachtet.

Tab. 1: Betrachtete statistische Tests und deren Implementierung in EasyStrata Funktionen. (β_i SNP-Effektschätzer für Gruppe i , SE_i Standardfehler von β_i , $w_i = 1/SE_i$)

EasyStrata Funktion	Anzahl Strata	Teststatistik	Referenz
CALCLCPDIFF	2	$Z_{Diff} = \frac{\beta_1 - \beta_2}{\sqrt{SE_1^2 - SE_2^2}} \sim N(0, 1)$	Differenz-Test [Ra13]
METAANALYSIS	2...m	$Z = \frac{\sum_i \beta_i w_i}{\sum_i w_i} \sim N(0, 1)$	Meta-Analyse [CH79]
JOINTTEST	2...m	$C_{Joint} = \sum_i \beta_i^2 w_i \sim \mathcal{X}_m^2$	Joint (main + interaction) Test [As10]
CALCPHET	2...m	$C_{Het} = \sum_i [(\beta_i - \frac{\sum_i \beta_i w_i}{\sum_i w_i})^2 w_i] \sim \mathcal{X}_{m-1}^2$	Chochrane's Heterogenitätstest [Co54]

Die genomweite Suche nach Effekt-Unterschieden zeigte sich als am besten geeignet für Varianten deren Effekte zwischen den Gruppen in unterschiedliche Richtungen zeigen. Überraschenderweise zeigte sich der naive Ansatz, zuerst für allgemeine (gruppenkombinierte) Effekte zu filtern und dann auf Effekt-Unterschied zu testen, als am besten geeignet für Varianten, die keinen oder einen kleineren (aber gleichgerichteten) Effekt in einer der beiden Gruppen zeigen. Bemerkenswert war auch, dass das Filtern für Joint Assoziation

eine Erhöhung des Typ 1 Fehlers des Effekt-Unterschied Tests zur Folge hatte, wenn dieser im gleichen Datensatz durchgeführt wird.

Um die diskutierten Ansätze auf hochdimensionale GWAMA Ergebnis-Datensätzen anwenden zu können, wurden sie in das R Paket EasyStrata implementiert [Wi14b] (**Tabelle 1**). Neben einem einfachen Anwenden der Tests, bietet die Software eine erweiterte Funktionalität zum Kombinieren der Tests, zum Extrahieren der signifikanten Ergebnisse sowie zum Erstellen von grafischen Darstellungen der hochdimensionalen Ergebnisse.

3.2 Anwendung auf Daten des GIANT Konsortium

Um geschlechts- und altersabhängige genetische Effekte für Adipositasparameter zu detektieren, wurden im Rahmen des GIANT Konsortiums geschlechts- und alters-stratifizierte (d.h., Männer vs. Frauen und ≥ 50 J vs. < 50 J) GWAMAs durchgeführt [Wi15a]. Hierbei wurden die 330,000 Personen aus verschiedenen GWAS in stratifizierten GWAMAs ausgewertet.

Durch Anwendung der optimalen Ansätze mittels EasyStrata konnten 15 Genorte mit signifikanten Altersunterschieden für den BMI und 44 Genorte mit signifikanten Geschlechtsunterschieden für das THV identifiziert werden (**Abbildung 2**). Dabei zeigte sich sowohl eine Anhäufung von größeren genetischen Effekten auf BMI in der jüngeren Population (an 11 der 15 BMI Genorte), als auch eine Anhäufung von größeren genetischen Effekten auf THV in Frauen (an 28 der 44 THV Genorte).

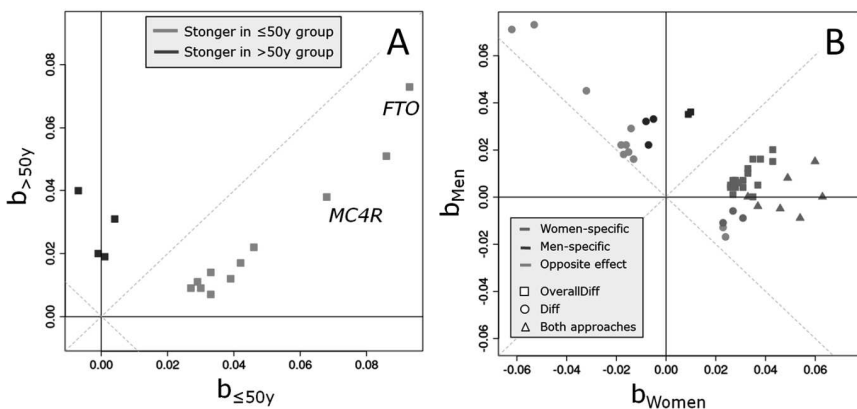


Abb. 2: Vergleich der (A) altersspezifischen genetischen Effekte auf BMI für 15 Genorte mit signifikanten Altersunterschieden, und (B) geschlechtsspezifischen genetischen Effekte auf THV für 44 Genorte mit signifikanten Geschlechtsunterschieden. Die gefundenen BMI Genorte mit Altersunterschied inkludierten Regionen in der Nähe der etablierten Adipositasgene *FTO* und *MC4R*.

4 Entwicklung eines Protokolls und der Software EasyQC zur Qualitätskontrolle von GWAMAs

4.1 Methoden und Implementierung

Die Analysten von GWAMA Konsortien haben es meist mit mehreren hundert [?]big GWAS Datensätzen zu tun, ein Umstand der extensive QC und Software benötigt, die mit einer solch großen Zahl an hochdimensionalen Datensätzen umgehen kann. Um dies zu bewerkstelligen, wurde ein Protokoll zur Durchführung der QC und ein leistungsstarkes Tool namens EasyQC entwickelt, welches die QC von mehreren großen GWAS und GWAMA Datensätzen ermöglicht. Ein typischer Ablauf der QC innerhalb eines GWAMA Projektes ist in **Abbildung 3** dargestellt.

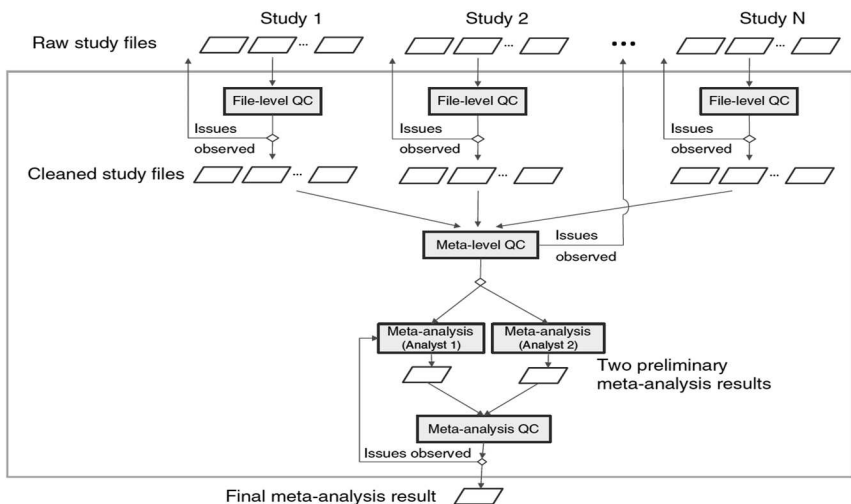


Abb. 3: Ablauf einer typischen GWAMA mit QC (Grafik aus [Wi14a]).

Man nehme an, die beteiligten GWAS wurden seitens der einzelnen Studien bereits durchgeführt und die Ergebnisse auf einem eigens dafür bereit gestellten ftp Server seitens des Konsortiums eingesammelt. Die typische Durchführung der GWAMA kann nun in vier aufeinanderfolgende Schritte aufgeteilt werden: File-level QC, Meta-level QC, Meta-analysis und Meta-Analysis QC.

Beim File-level QC wird jede einzelne, studienspezifische GWAS Ergebnisdatei individuell geputzt und überprüft. Zum Putzen gehören mehrere Schritte, wie z.B. die Harmonisierung der Ergebnisse über alle Studien hinweg oder das Ausschließen von SNPs mit schlechter Qualität: GIANT zum Beispiel, schließt SNPs aus, basierend auf einem zu geringen Vorkommen des selteneren Allels und basierend auf zu geringer Imputations- oder Genotypqualität. Zum Überprüfen gehören detaillierte Plausibilitätsprüfungen, wie z.B. das Überprüfen von allgemeinen Statistiken und Überprüfen der Häufigkeit von SNP Exklusionen. Ziel des File-level QC ist, allgemeine analytische Fehler in den Daten zu finden und Daten schlechter Qualität von vornherein aus der Metaanalyse auszuschließen. Bei

Problemen, wird der jeweilige Studien-Analyst kontaktiert und um eine Validierung oder Korrektur der GWAS Ergebnisse gebeten.

Beim Meta-level QC werden aggregierte Statistiken der geputzten Dateien über mehrere GWAS hinweg verglichen. Ziel dieses Schrittes ist, studienspezifische Probleme und Outlier zu identifizieren, die beim Betrachten eines einzelnen Ergebnisses (File-Level QC) nicht detektiert werden konnten. Zum Meta-level QC gehört das Überprüfen einer Reihe an analytischen Problemen, wie z.B. Problemen mit Phenotyptransformationen (durch Vergleich von Median des Standard Errors in Bezug auf Stichprobengröße, über alle Studien hinweg), mit Allelfrequenzen und Effektrichtungen (durch Plotten der Allelfrequenzen aller SNPs gegen einen Referenzdatensatz), oder mit Populationsstratifizierung (durch Überprüfen von Quantile-Quantile Plots und durch Berechnen des Genomiccontrol Inflationsfaktors [DR54]).

Nach abgeschlossenen File-level und Meta-level QC, folgt die eigentliche Metaanalyse der geputzten GWAS Datensätze. Um etwaige Fehler in Bezug auf Studien In-/ oder Exklusionen sowie andere metaanalytische Fehler auszuschließen, wird diese in der Regel unabhängig von zwei Analysten parallel durchgeführt.

Metaanalysis QC meint das Überprüfen und Validieren des Metaanalyseergebnisses. Dazu gehört der Vergleich der Ergebnisse der beiden Meta-Analysten, sowie das Überprüfen der Ergebnisse auf potentielle Zwischenstudien-Heterogenitäten und auf Einhaltung des Typ 1 Fehlers.

Sämtliche Methoden der File-level, Meta-level und Metaanalysis QC wurden in EasyQC implementiert. Vordefinierte ecf-Skripte für die unterschiedlichen QC Schritte stehen auf www.genepi-regensburg.de/software zum Download bereit.

4.2 Anwendung auf Daten des GIANT Konsortium

Die beschriebenen QC Schritte wurden im Rahmen aktueller GIANT Metaanalysen entwickelt, protokolliert und angewandt [Lo15, Sh15, Wi15a]. Insgesamt wurden bis zu 3,000 einzelne GWAS Ergebnisdateien qualitätskontrolliert und in unterschiedlichen Metaanalysen zusammen ausgewertet.

Hierbei hat sich die enorme Wichtigkeit einer gründlichen Qualitätskontrolle der einzelnen GWAS Ergebnisse gezeigt. Bei circa der Hälfte aller zu Beginn eingesammelten GWAS Ergebnisse wurden Probleme festgestellt, die ein Kontaktieren des jeweiligen Studienanalysten zur Folge hatten. Am häufigsten waren Probleme mit Phenotyptransformationen oder Allelfrequenzen. **Abbildung 4** zeigt verschiedene Muster in Allelfrequenzplots, wie sie im Laufe der GIANT Meta-level QC beobachtet wurden. Abweichungen von Referenzfrequenzen wurden vor allem festgestellt für Studien, deren Ethnizität nicht zur Referenz passte und für Studien, bei denen Allele, Allelfrequenzen oder Strangannotation falsch kodiert wurden. Da solche Fehler ein Umdrehen der Effektrichtung nach sich ziehen, würden sich diese Probleme negativ auf Power und Validität der Metaanalyse auswirken und einen systematischen Fehler im geschätzten Effekt ergeben.

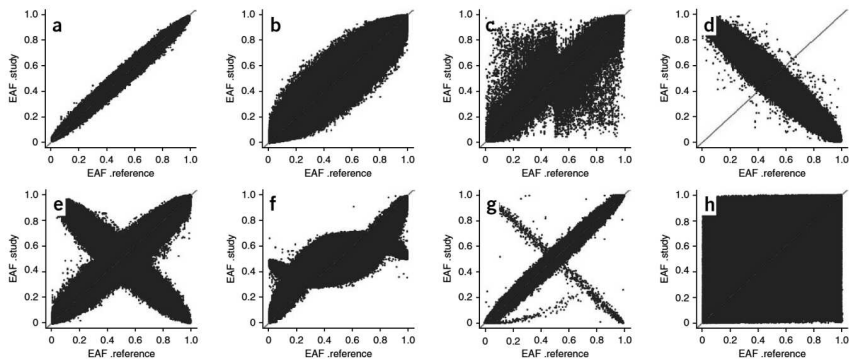


Abb. 4: Qualitätskontrolle von studienspezifischen GWAS Ergebnissen mittels Allelfrequenzplots (from [Wi14a]). Die Graphen zeigen die Häufigkeit des Effektallels (EAF, Effect Allele Frequency) in einer bestimmten Studie über der EAF einer europäischen Referenzpopulation. (a)-(c): Kodierung wurde richtig durchgeführt, wobei die Ethnizität der Studie (a) sehr gut, (b) eher weniger und (c) gar nicht zur Referenzpopulation passt. (d)+(e): Fehlerhafte Kodierung von (d) allen, oder (e) einem Teil der Allele, Allelfrequenzen oder Strangannotationen. (f)-(h): Andere systematische Fehler in der Analyse oder in der Datenverarbeitung.

5 Conclusion

In der vorliegenden Doktorarbeit sind Methoden und Software zur Qualitätskontrolle von GWAS und GWAMAs im Allgemeinen, als auch zur statistischen Auswertung von stratifizierten GWAMAs, untersucht und entwickelt worden. Die Softwarepakete EasyQC (zur GWAMA QC) und EasyStrata (zur Auswertung von stratifizierten GWAMAs) sind im Rahmen von Arbeiten des GIANT Konsortiums entstanden und konnten dort bereits erfolgreich in diversen Projekten eingesetzt werden.

Beide Pakete basieren auf dem eigens entwickelten Easy Framework, welches effiziente Datenverarbeitungsmethoden für multiple hochdimensionale GWAS und GWAMA Datensätze mit einer Skriptschnittstelle verbindet. Über die Skriptschnittstelle können die zur Verfügung gestellten Methoden vom Benutzer flexibel und in einfacher Weise aneinandergereiht und ausgeführt werden. Bereits angepasste ecf-Skripte zur GWAMA-QC und zur statistischen Auswertung von stratifizierten GWAMAs stehen auf www.genepi-regensburg.de/software zum Download zur Verfügung.

Die Software EasyQC und die darin enthaltenen Methoden wurden in GIANT unter anderem zur QC von GWAS und GWAMAs für Adipositasparameter eingesetzt. Insgesamt wurden in den aktuellen GIANT GWAMA Projekten bis zu 330,000 Personen von 125 Studien gemeinsam ausgewertet und eine Vielzahl an assoziierten Genorten gefunden: 97 assoziierte Genorte für den BMI [Lo15] (einem Maß für allgemeine Adipositas, Fettmasse) und 49 assoziierte Genorte für das THV [Sh15] (einem Maß für zentrale Adipositas, Fettverteilung). Besonders interessant war eine daraus resultierende Trennung der Wirkweise der Genetik für Fettmasse und Fettverteilung: Während die gefundenen BMI Genorte vor allem im Zentralnervensystem und im Gehirn wirken (viele davon liegen direkt in der Ap-

petitsignalkette), spielen die gefundenen THV Genorte eher eine Rolle bei der Entstehung von Fettzellen und bei der Insulinregulation.

Auch die optimierten Methoden zur gezielten Detektion von gruppenspezifischen Effekt-Unterschieden und die Software EasyStrata wurden bereits erfolgreich für Adipositasparameter angewandt. Hierzu wurden GIANT GWAMAs stratifiziert nach Alter (≤ 50 Jahre vs. > 50 Jahre) und Geschlecht (Männer vs. Frauen) durchgeführt und ausgewertet [Wi15a]. Für den BMI konnten 15 Genorte identifiziert werden, an denen sich die genetischen Effekte zwischen jüngeren und älteren Probanden signifikant unterscheiden. Für das THV konnten 44 Genorte gefunden werden, an denen sich die genetischen Effekte signifikant zwischen Männern und Frauen unterscheiden. Umgekehrt konnten keine Geschlechtsunterschiede für den BMI und keine Altersabhängigkeit für das THV identifiziert werden. Die Hypothese einer unterschiedlichen Wirkweise der Genetik für Fettmasse (BMI) und der Genetik für Fettverteilung (THV) konnte weiter verstärkt werden: Während BMI Genetik eher vom Alter und nicht vom Geschlecht abhängt, scheint THV Genetik in besonderem Maß vom Geschlecht, nicht aber vom Alter, abzuhängen.

Die Ergebnisse haben enorm zum besseren Verständnis von biologischen Mechanismen der Entstehung von Adipositas beigetragen. Sie bieten neue Ansatzpunkte für die Entwicklung von Medikamenten und Therapien, sowie Potential für eine verbesserte Prädiktion von Adipositas und verwandter Krankheiten, wie z.B. Typ 2 Diabetes. Das Wissen, dass manche genetische Effekte lediglich in bestimmten Gruppen wirken, kann helfen, Medikamente und Therapien im Rahmen einer individualisierten Medizin besser auf die jeweilige Personengruppe anzupassen.

Zusammenfassend lässt sich sagen, dass die methodischen Ergebnisse, bestätigt durch die Anwendung auf Daten des GIANT Konsortiums und die zur Verfügung gestellte Softwareimplementierung, die Durchführung zukünftiger GWAMA Auswertungen zum Detektieren von assoziierten Genorten und Effekt-Unterschieden unterstützen und anleiten können. Komplexität und Dimension der GWAMAs werden dabei aufgrund einer steten Erweiterung des SNP-Panels und einer steten Erhöhung der Stichprobenzahlen, weiter steigen. Die Softwarepakete werden nicht nur von Analysten des GIANT-Konsortiums, sondern auch von anderen international anerkannten genetischen Forschungskonsortien verwendet, wie z.B. dem Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Konsortium für Blutdruck und Lipidparameter, dem Social Science Genetic Association Consortium (SSGAC) für sozialwissenschaftliche Phänotypen, oder dem Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) für Glukose- und Insulinparameter. Neue Erkenntnisse über genetische Effekte auf Krankheiten und krankheitsrelevante Merkmale, sowie deren Gruppenabhängigkeiten, werden letztendlich zu einem besseren Verständnis der Ursachen und Entstehung von Krankheiten beitragen.

Literaturverzeichnis

- [As10] Aschard, H., Hancock D.B. et al.: Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. *Human Heredity* 70, 292-300, 2010.

- [CH79] Cox, D.R.; Hinkley, D.V.: Theoretical statistics. London New York: Chapman and Hall ; distributed in U.S. by Halsted Press, 1979.
- [Co54] Cochran, W.G.: The Combination of Estimates from Different Experiments. *Biometrics* 10, 101-129, 1954.
- [DR54] Devlin, B.; Roeder, K.: Genomic control for association studies. *Biometrics* 55, 997-1004, 1954.
- [HD05] Hirschhorn, J.N.; Daly, M.J.: Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6, 95-108, 2005.
- [Lo15] Locke, A.E., Kahali B. et al.: Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518, 197-206, 2015.
- [Ra13] Randall, J.C., Winkler T.W. et al.: Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genetics* 9, e1003500, 2013.
- [Sh15] Shungin, D., Winkler T.W. et al.: New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518, 187-196, 2015.
- [We14] Welter, D., MacArthur J. et al.: The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research* 42, D1001-1006, 2014.
- [Wi14a] Winkler, T.W., Day F.R. et al.: Quality control and conduct of genome-wide association meta-analyses. *Nature Protocols* 9, 1192-1212, 2014.
- [Wi14b] Winkler, T.W., Kutalik Z. et al.: EasyStrata: Evaluation and Visualization of stratified genome-wide association meta-analysis data. *Bioinformatics*, 2014.
- [Wi15a] Winkler, T.W., Justice A.E. et al.: The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genetics* 11, e1005378, 2015.
- [Wi15b] Winkler, T.W.: Methods to investigate gene-strata interaction in genome-wide association meta-analyses on the example of obesity. Dissertation, University of Regensburg, 2015.



Thomas Winkler wurde geboren am 5. Oktober 1981, in Regensburg. Von 2004 bis 2009 absolvierte er ein Mathematikstudium an der Fachhochschule Regensburg, welches er mit einer sehr guten Gesamtnote, erfolgreich als Diplom-Mathematiker (FH) abgeschlossen hat. Seit 2009 ist er als wissenschaftlicher Mitarbeiter am Lehrstuhl für Genetische Epidemiologie an der Universität Regensburg angestellt. Seine Doktorarbeit im Bereich der Biomedizinischen Wissenschaften (Dr.rer.physiol) begann er in 2010 und hat diese in 2015 erfolgreich mit summa cum laude abgeschlossen. In 2012 absolvierte er einen dreimonatigen Forschungsaufenthalt am Charles Bronfman Institut für Personalisierte Medizin, an der Mount Sinai School of Medicine, in New York, USA.

Für seine Tätigkeit im Rahmen von verschiedenen, internationalen, genetischen Forschungskonsortien wurde ihm 2014 der CHARGE (Cohorts for Heart and Aging Research in Genetic Epidemiology) Konsortium "Golden Tiger Award" verliehen.