

Genomics

Genomweite Assoziationsstudien (GWAS) – Möglichkeiten und Grenzen

JENNIFER KRIEBEL¹, THOMAS ILLIG^{1, 2}, HARALD GRALLERT¹¹ABTEILUNG FÜR MOLEKULARE EPIDEMIOLOGIE, HELMHOLTZ ZENTRUM MÜNCHEN, NEUHERBERG²MEDIZINISCHE HOCHSCHULE HANNOVER

GWAS have brought many successes to detect DNA-variants for common diseases and other traits. Feasibility to genotype a large amount of SNPs cost efficiently in big populations with various study designs benefits GWAS in contrast to candidate gene studies. However, they only explain a small proportion of estimated heritability so far. It's still a long way to go, to develop methods to discover false negative signals, find other impact e. g. from epigenetics or gene-environment interaction as well as to translate all findings into better understanding of disease and clinical practice.

DOI: 10.1007/s12268-012-0223-7

© Springer-Verlag 2012

Genomweite Assoziationsstudien (GWAS)

■ In den letzten Jahren haben sich DNA-Chip-Technologien zur Genotypisierung großer Zahlen an Einzel-Nukleotid-Polymorphismen (SNPs) rasant weiterentwickelt. Die SNPs wurden dabei über das gesamte Genom mit steigender Abdeckung und unterschiedlichen Schwerpunkten auf bestimmte genomische Regionen selektiert. Bei vielen polygenen Erkrankungen führte diese Art der Analyse zum damals postulierten Durchbruch. Zahlreiche neue Gene wurden entdeckt. Darunter waren auch viele Gene, die aus biologischer Sicht bisher nicht mit der jeweiligen Erkrankung in Zusammenhang gebracht wurden. Genomweite Assoziationsstudien (GWAS) haben somit das Verständnis der genetischen Grundlagen vieler komplexer Merkmale deutlich verbessert.

Neben den Chip-Technologien ist die Grundlage für GWAS eine detaillierte „Karte“ der Genvarianten und Muster des LD (*linkage disequilibrium*), die durch das internationale HapMap-Konsortium durch Sequenzierung erstellt und kürzlich durch das 1.000-Genome-Projekt erweitert wurde. Ebenso sind gut phänotypisierte, große Kohorten und methodische Fortschritte bei der statistischen Auswertung großer Datenmengen essenziell [1, 2]. Derzeit sind über 1.200 GWAS, die über 200 humane Erkrankungen und Eigenschaften untersuchen, publiziert (**Tab. 1**, [3]). Die Schwelle zur genomweiten Signifikanz liegt dabei bei einem P-Wert von 5×10^{-8} [4]. Für eine valide Assoziation wird momentan ein mehrstufiges Design angewandt (**Abb. 1**), ein SNP muss in einer ausreichenden Probengröße untersucht worden sein, dabei genomweite Signifikanz erreichen sowie in einer unabhängigen Population bestätigt repliziert worden sein [4]. Bei komplexen Erkrankungen sind die Effekte der einzelnen Loci meist gering, weshalb eine ausreichende Populationsgröße nur durch Zusammenfassen mehrerer Studien in Meta-Analysen erreicht wird.

Tab. 1: Übersicht der publizierten GWAS bei verschiedenen Krankheiten bzw. Merkmalen.

Krankheit/Merkmal	Publikationen	identifizierte SNPs	Gene
Alzheimer	21	58	<i>APOE, CR1, BIN1</i>
Asthma	18	43	<i>NOTCH4, GSDMB, TSLP</i>
Blutdruck	6	53	<i>RPL6, C10orf107, MTHFR</i>
Blutplättchenzahl	2	46	<i>ARHGFB3, FLI36031, JMJD1C</i>
Body-Mass-Index (BMI)	15	101	<i>FTO, TMEM18, MC4R</i>
Morbus Crohn	11	119	<i>NOD2, IL23R, ATG16L1</i>
diastolischer Blutdruck	3	30	<i>CYP11A1, FGF5, SH2B3</i>
Fettleibigkeit	6	20	<i>FTO, BDNF, MC4R</i>
HDL-Cholesterol	12	96	<i>CETP, LPL, LIPC</i>
kardiovaskuläre Risikofaktoren	4	28	<i>BCHE, WDR1, CETP</i>
koronare Herzerkrankung	17	117	<i>ALDH2, CELSR2, MTAP</i>
LDL-Cholesterol	12	75	<i>CELSR2, APOE, APOB</i>
Lungenfunktion	5	55	<i>GSTCD, MFAP2, THSD4</i>
Nüchternplasmaglukose	6	10	<i>G6PC2, MTNR1B, GCK</i>
systolischer Blutdruck	3	29	<i>CYP17A1, CYP11A1, GNAS</i>
Taille-Hüfte-Verhältnis	1	16	<i>GRB14, LYPLAL1, RSPO3</i>
Triglyceride	11	78	<i>APOA1, GCKR, LPL</i>
Typ-1-Diabetes	10	78	<i>MHC, PTPN22, INS</i>
Typ-2-Diabetes	30	122	<i>TCF7L2, KCNQ1, CDKN2A/B</i>

Auszug der insgesamt 1.258 Veröffentlichungen, die 6.400 durch GWAS nachgewiesene krankheitsassoziierte SNPs beschreiben. Beispielhaft sind jeweils drei hochsignifikante Gene genannt, die in der Nähe (bei mehreren das biologisch plausibelste Gen) des assoziierten SNPs liegen (www.genome.gov/26525384; Stand: Mai.2012).

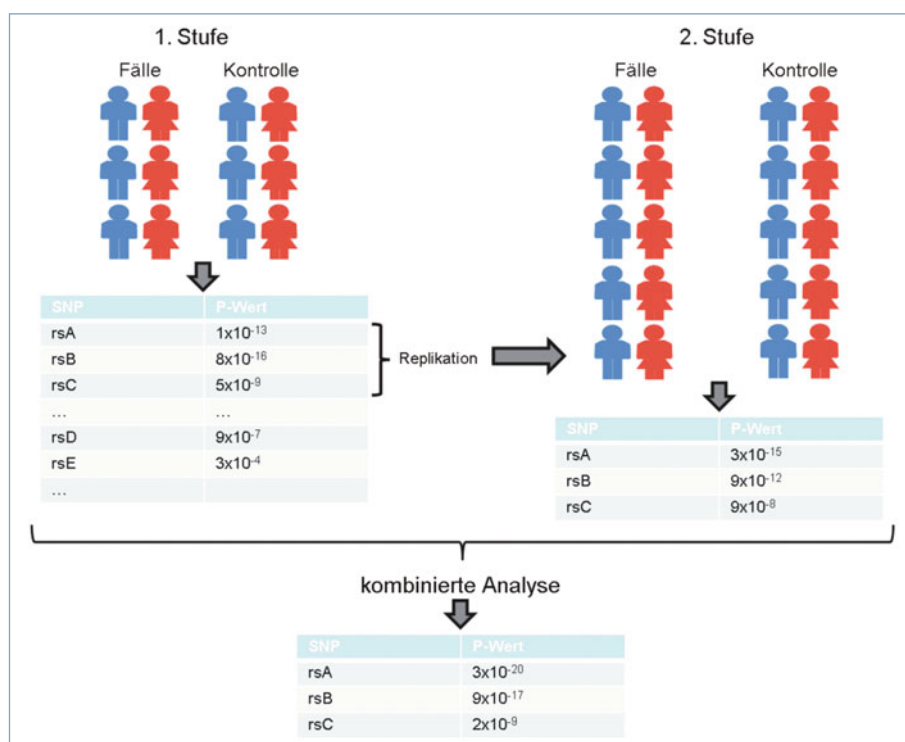
Möglichkeiten von GWAS

Genetische Studien sind unverzichtbar, da sie zum einen kostengünstig durchgeführt werden können, aber auch neben der präzisen Aufklärung der molekularen Pathogenese zur Entdeckung neuer Biomarker beitragen können. Ebenso können neue Angriffspunkte für Pharmaka, neue individuelle Interventionsmöglichkeiten und ein verbessertes Verständnis der Heterogenität einwickelt werden [1]. Generell geben genetische Ansätze durch das Aufzeigen von Verbindungen zwischen Sequenz und Phänotyp starke Hinweise auf die Rolle von Genen oder einen Stoffwechselweg bei humanen Erkrankungen, wodurch die Experimente am Menschen reduziert werden können [5]. Ein Vorteil der GWAS besteht in der hypothesenfreien systematischen Untersuchung häufiger Genvarianten über das gesamte Genom, die im Gegensatz zu Kandidatengenanalysen viel Information auf einmal abdeckt und auch regulatorische Regionen mit einbezieht, die teilweise von Genen weit entfernt liegen. Dadurch kann ein mehr oder weniger vollständiges Bild der genetischen Einflüsse auf ein bestimmtes Merkmal generiert werden. Zudem wurden mit dieser Methode zahlreiche Gene entdeckt, die vorher nicht mit einer bestimmten Krankheit in Verbindung gebracht wurden [6]. Die stetig vereinfachte Durchführung dieses Ansatzes ermöglicht weiterhin die Genotypisierung immer größerer Populationen, wodurch auch kleine Effekte detektiert werden können. Häufige Varianten (Allelfrequenz: über fünf Prozent) scheinen dabei zu höherer Suszeptibilität bei häufigen Krankheiten zu führen (*common disease – common variant*-Hypothese) [7].

Probleme und Grenzen der GWAS

Ein großes Problem bei GWAS ist unser limitiertes Wissen über die Funktion großer Teile des menschlichen Genoms. Für viele Gene, in deren Nähe durch GWAS Varianten assoziiert wurden, ist die Funktion weitgehend unbekannt. Auch regulatorische Mechanismen sind noch nicht vollständig aufgeklärt, was die Verbindung assoziierter Varianten mit der tatsächlichen Funktion, die zu der Assoziation führt, erschwert. Die assoziierten Varianten sind in den seltensten Fällen kausal, was detaillierte Feinkartierungen der im LD befindlichen Varianten nötig macht.

Bei Lungenerkrankungen wurden bisher 22 GWAS veröffentlicht, jedoch bleiben viele Fragen offen, da viele der identifizierten Varianten nur nominale bis allgemeine Krank-



▲ **Abb. 1:** Design der Genomweiten Assoziationsstudie (GWAS). In der Entdeckungsphase wird eine Meta-Analyse einer gewissen Anzahl von Studien durchgeführt (1. Stufe). Signifikante Hits und Hits, die im Grenzbereich der Signifikanzschwelle liegen ($P \geq 5 \times 10^{-8}$), werden in der Replikationsphase in einer größeren Anzahl von unabhängigen Studien untersucht (2. Stufe). Meist werden die beiden Stufen auch gemeinsam analysiert (kombinierte Analyse) und dadurch die statistische Power nochmals erhöht.

heitsrisiken aufweisen. Dies könnte darauf hinweisen, dass es zusätzliche unentdeckte genetische Faktoren gibt bzw. nicht gemessene Gen-Gen-Interaktionen oder bisher schlecht verstandene Umwelteinflüsse [3].

Für manche Erkrankungen, wie z. B. Schizophrenie, scheint der genomweite Ansatz gar komplett der falsche zu sein, da hier keine Varianten mit einer genomweiten Signifikanz detektiert werden konnten [6].

Bei komplexen Merkmalen sind die Effektgrößen einzelner assoziierter genetischer Varianten in der Regel eher gering. Um diese kleinen Effekte zu detektieren, sind große Populationen essenziell. Dies wird vor allem durch Meta-Analysen erreicht. Ein Problem hierbei ist die Verwendung unterschiedlicher Genotypisierungsplattformen. Dies kann zwar durch Imputation – die Vorhersage nicht genotypisierter SNPs auf Basis von LD-Mustern – behoben werden, Plattform-bezogene Qualitätsunterschiede können jedoch weiterhin problematisch sein. Bisherige DNA-Chips decken kaum Varianten im Frequenzbereich von einem bis fünf Prozent ab, wodurch Information verloren geht. Mangels geeigneter Analysemethoden sind außerdem in den meisten GWAS die Gonosomen nicht berück-

sichtigt. Die Menge an statistischen Tests, die bei einer GWA gerechnet werden, kann zu falsch-positiven Ergebnissen führen, wird jedoch streng korrigiert, verschwinden viele Hits im Rauschen. Um dies zu verhindern, werden die meisten GWAS mehrstufig angelegt. In der Entdeckungsphase gefundene Signale werden in groß angelegten Replikationen überprüft und so falsch-positive Assoziationen minimiert sowie einige falsch-negative auf genomweite Signifikanz gebracht [6].

Starke Diskrepanzen zwischen gefundenen genetischen Heritabilitätsanteilen und geschätzter Heritabilität bringen Ernüchterung. Die bisher erklärte Varianz liegt bei den meisten komplexen Merkmalen im Bereich um zehn Prozent, während beispielsweise die Heritabilität von Typ-2-Diabetes (T2D) bei 25 Prozent [8] oder von der Körpergröße sogar bei 60 bis 80 Prozent [9] liegt. Ein noch extremeres Beispiel ist der Body-Mass-Index (BMI), wo die erklärte Varianz von 1,45 Prozent einer geschätzten Heritabilität von 40 bis 60 Prozent gegenübersteht [10]. Dies weist darauf hin, dass unser Verständnis von komplexen Phänotypen eher gering ist. Mögliche Erklärungen könnten neben seltenen Varianten in Strukturvarianten, die mit den momentanen

Methoden schlecht zu detektieren sind, oder in der Interaktion von Genen und Umweltfaktoren liegen [6].

Eine zentrale Erwartung bei GWAS ist die Verbesserung der Prädiktion. Patienten, die mehrere Risikoallele tragen, haben ein erhöhtes Risiko, eine Krankheit zu entwickeln, gegenüber solchen, die weniger Risikoallele tragen. Im Beispiel von T2D tragen die detektierten Varianten allerdings neben Risikofaktoren wie Alter, Geschlecht und BMI nur geringfügig zur Prädiktion bei [11]. Dies ist auch bei vielen anderen Erkrankungen zu beobachten.

Populationsheterogenität in der Bevölkerung

Derzeit werden immer mehr GWAS in nicht-europäischen Populationen abgeschlossen, wodurch es bald möglich sein wird, genomweite Muster von Assoziationen über eine weite Breite von ethnischen Gruppen zu analysieren [12]. In den unterschiedlichen Populationen haben zum Teil andere Varianten Einfluss auf die Krankheit. So konnte z. B. für Varianten von TCF7L2 eine relativ starke Assoziation mit T2D bei Europäern nicht aber bei Asiaten gezeigt werden [4]. Dies zeigt die Notwendigkeit detaillierter auf GWAS zwischen den Populationen einzugehen.

Fehlende Heritabilität

Die fehlende Heritabilität könnte durch nicht gefundene seltene Varianten oder epigenetische Modifikationen, die in einer veränderten Genexpression resultieren, erklärt werden [3]. Des Weiteren könnte sie durch Methodenprobleme bzw. durch Suszeptibilitätsloci mit niedrigem Effekt bedingt sein oder durch andere DNA-Varianten, die durch SNP-Chips nicht abgedeckt sind (strukturelle Varianten) [5]. Hierbei könnten Resequenzierungsansätze, die gegen gering strukturelle und seltene Varianten gerichtet sind, die Entdeckung der genetischen Basis von interethnischen Unterschieden in der Prävalenz und Präsentation von Krankheiten untersuchen [12].

Monogene Erkrankungen

Ebenso könnten durch *Next-Generation-Sequenzieretechnologien* in Zukunft neue genetische Informationen gefunden werden, die vor allem bei monogenetischen Erkrankungen eine individuelle Prognose bieten könnten. Es wird allerdings weiterhin nicht einfach sein, die pathogenen Mutationen bei der Anzahl an Varianten zu identifizieren [13].

Omics-Techniken

Andere *Omics*-Methoden, wie *Epigenomics*, *Metabolomics* und *Proteomics*, könnten hier von großer Hilfe sein [14]. Epigenetische Veränderungen, die in den GWAS bislang nicht untersucht worden sind, könnten z. B. die genetische Prädisposition beeinflussen. Des Weiteren muss geklärt werden, inwiefern die Risikofaktoren für eine Erkrankung unabhängig vom Lebensstil wirken oder ob sie durch eine Umstellung des Lebensstils positiv beeinflusst werden können.

Fazit

Es besteht die Hoffnung, dass die Ergebnisse aus genetischen Analysen zu einer verbesserten, klinischen Vorsorge (*care*) für die steigende Anzahl an Patienten führen könnte. Trotz Bedenken durch Probleme mit dem multiplen Testen waren GWAS im Gegensatz zu Kandidatengenstudien sehr erfolgreich bei der Detektion neuer Loci für komplexe Merkmale. Dadurch hat sich das Verständnis der molekularen Grundlagen vieler Merkmale deutlich verbessert. Die Erklärung der gesamten vorhergesagten Heritabilität dieser Merkmale konnte damit allerdings bisher nicht erreicht werden. Die Aufgabe der Forschung in den nächsten Jahren wird neben der funktionellen Charakterisierung der zahlreichen neuen Risikoloci die Entwicklung neuer Methoden sein, um aus den vorhandenen Daten weitere Informationen zu ziehen, sowie die Suche nach der unentdeckten Heritabilität auf andere Bereiche auszuweiten. Dabei bieten sich momentan epigenetische Untersuchungen und die Integration weiterer *Omics*-Daten an. Auch auf Gen-Gen- und Gen-Umwelt-Interaktionen

muss in Zukunft stärkeres Augenmerk gelegt werden. ■

Literatur

- [1] Herder C (2010) Genetische Studien zum Typ-2-Diabetes. *Der Diabetologe* 6:203–209
- [2] Kriebel J, Grallert H, Illig T (2012) Typ-2-diabetes-assoziierte Gene. *Der Diabetologe* 8:26–34
- [3] Todd J L, Goldstein D B, Ge D et al. (2011) The state of genome-wide association studies in pulmonary disease: a new perspective. *Am J Respir Crit Care Med* 184:873–880
- [4] Imamura M, Maeda S (2011) Genetics of type 2 diabetes: the GWAS era and future perspectives [Review]. *Endocr J* 58:723–739
- [5] Sham P C, Cherny S S (2011) Genetic Architecture of Complex Diseases. In: *Analysis of Complex Disease Association Studies: A Practical Guide*. Zeggini E, Morris A (Hrsg). Elsevier, London, Burlington, San Diego, 1–13.
- [6] Manolio T A (2010) Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 363:166–176
- [7] Adhikari K, Alchawa T, Ludwig K et al. (2012) Is It Rare or Common? *Genet Epidemiol* (im Druck)
- [8] Morris A, Voight B F, Teslovich T M et al. (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* (im Druck)
- [9] Perola M (2011) Genetics of human stature: Lessons from genome-wide association studies. *Horm Res Paediatr* 76 Suppl 3:10–11
- [10] Speliotes E K, Willer C J, Berndt S I et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42:937–948
- [11] Lango H, Palmer C N, Morris A D et al. (2008) Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. *Diabetes* 57:3129–3135
- [12] McCarthy M I (2011) The importance of global studies of the genetics of type 2 diabetes. *Diabetes Metab J* 35:91–100
- [13] McCarthy M I (2010) Genomics, type 2 diabetes, and obesity. *N Engl J Med* 363:2339–2350
- [14] Herder C, Karakas M, Koenig W (2011) Biomarkers for the prediction of type 2 diabetes and cardiovascular disease. *Clin Pharmacol Ther* 90:52–66

Korrespondenzadresse:

Dr. Harald Grallert
Abteilung für molekulare Epidemiologie
Helmholtz Zentrum München
Ingolstädter Landstraße 1
D-85764 Neuherberg
Tel.: 089-3187-1195
Fax: 089-3187-4567
harald.grallert@helmholtz-muenchen.de

AUTOREN



Jennifer Kriebel

Jahrgang 1986. 2005–2010 Biologiestudium an der Universität Düsseldorf. Seit 2011 Doktorandin am Helmholtz Zentrum München (HMGU), Abteilung für Molekulare Epidemiologie.



Harald Grallert

Jahrgang 1978. 1998–2004 Biologiestudium an der TU München. 2004–2007 Doktorand am Helmholtz Zentrum München (HMGU), Institut für Epidemiologie. 2008 Promotion an der TU München. Seit 2008 Postdoktorand am HMGU, seit 2011 Arbeitsgruppenleiter in der Abteilung Molekulare Epidemiologie.



Thomas Illig

Jahrgang 1965. 1986–1991 Biologiestudium an der Universität Regensburg, 1995 Promotion. 1996–1998 Postdoktorand an der Universität Regensburg. 1998–2001 Postdoktorand am Helmholtz Zentrum München (HMGU). 2001–2010 Gruppenleiter am HMGU. 2006 Habilitation an der LMU München. Seit 2010 Professur an der LMU. 2011–2012 Abteilungsleiter am HMGU. Seit 2012 Leiter der Hannover Unified Biobank.