

Online-Datenbanken und systematische Metaanalysen komplex- genetischer Erkrankungen

Hintergrund komplex- genetischer Erkrankungen

Die Suszeptibilität komplex-genetischer Erkrankungen wird von einer Vielzahl von Umwelt- und genetischen Faktoren, z. B. in Form von Polymorphismen, bestimmt. Trotz kleiner Effektstärken (Odds Ratios, ORs) der einzelnen Polymorphis-

men ist die Heritabilität, d. h. der Anteil der Summe der genetischen Effekte an der Suszeptibilität, häufig hoch. Die Aufklärung der verantwortlichen Polymorphismen wird jedoch erschwert durch kleine ORs, verminderte Penetranz, phenotypische Variabilität der Erkrankung, Populationsstratifikationen, schwer zu kontrollierende Einflüsse von Umweltfaktoren, Gen-Gen-Interaktionen sowie bei Erkrankungen des höheren Lebensalters durch Komorbiditäten und den Mangel an großen Familien. Demgegenüber stehen Hunderte von publizierten Studien, die die Assoziation bestimmter Polymorphismen mit der entsprechenden Erkrankung beschreiben oder widerlegen, was die adäquate Beurteilung der Studienlage sowie einzelner Studienergebnisse erheblich erschwert bzw. häufig unmöglich macht.

Aus diesem Grund hat unsere Gruppe ein Konzept öffentlich zugänglicher Datenbanken für genetische Assoziationsstudien entwickelt, mit denen wir durch systematische Analyse der publizierten Literatur Hilfestellung bei der Interpretation der Ergebnisse geben. Bisher wurden diese Datenbanken von unserer Gruppe für 4 neuropsychiatrische Erkrankungen generiert: für die Alzheimer-Krankheit (AlzGene; <http://www.alzgene.org>; [1]) das idiopathische Parkinson-Syndrom (PDGene; <http://www.pdgene.org>), Schizophrenie (SZGene; <http://www.szgene.org>; [2]) und seit Kurzem ebenfalls online zu-

gänglich für multiple Sklerose (MSGene; <http://www.msgene.org>). Wir werden im Folgenden das Konzept und die Methodik anhand der Datenbanken für die Alzheimer-Krankheit (AD) sowie für das idiopathische Parkinson-Syndrom (PD) darstellen.

Kandidatengen-Studien und GWAS für AD und PD

Monogenetische mendelische Formen von AD und PD sind mit einem Anteil von 1–5% sehr selten im Vergleich zu den vorherrschenden komplex-genetischen Formen. In den letzten 30 Jahren sind für AD etwa 1200 und für PD etwa 800 Kandidatengen-Studien publiziert worden. Das sind Studien, die –meist auf funktionellen Hypothesen basierend – die Assoziation von bestimmten Polymorphismen mit der Erkrankungssuszeptibilität oder anderen phenotypischen Variablen der komplex-genetischen Form untersuchen. Monatlich erscheinen bis zu 10 neue Studien für die jeweilige Krankheit. Abgesehen von wenigen genetischen Varianten, die als etablierte Risikofaktoren gelten können (z. B. *APOE* für AD, *SNCA* und *MAPT* für PD), ergibt sich aus diesen Studien kein einheitliches Bild. Eine Alternative zu den traditionellen Kandidatengen-Studien besteht seit wenigen Jahren in der Durchführung von weitgehend hypothesenfreien GWAS, die ~300.000–1 Mio. Polymorphismen auf Assoziation testen,

Abkürzungen	
<i>AD</i>	Alzheimer-Krankheit
<i>APOE</i>	„apolipoprotein E“ (Gen-Name)
<i>BDNF</i>	„brain-derived neurotrophic factor“ (Gen-Name)
<i>CLU</i>	„clusterin“ (Gen-Name)
<i>CR1</i>	„complement component (3b/4b) receptor 1“ (Gen-Name)
<i>GBA</i>	„glucosidase, beta, acid“ (Gen-Name)
<i>GWAS</i>	Genomweite Assoziationsstudie
<i>LRRK2</i>	„leucine-rich repeat kinase 2“ (Gen-Name)
<i>MAPT</i>	„microtubule-associated protein tau“ (Gen-Name)
<i>MS</i>	Multiple Sklerose
<i>OR</i>	Odds Ratio (Effektstärke)
<i>PD</i>	Idiopathisches Parkinson-Syndrom
<i>PICALM</i>	„phosphatidylinositol binding clathrin assembly protein“ (Gen-Name)
<i>SNCA</i>	„alpha-synuclein“ (Gen-Name)
<i>SZ</i>	Schizophrenie
<i>TPRG1</i>	„tumor protein p63 regulated 1“ (Gen-Name)
<i>ε4Allel</i>	Epsilon-4-Allel

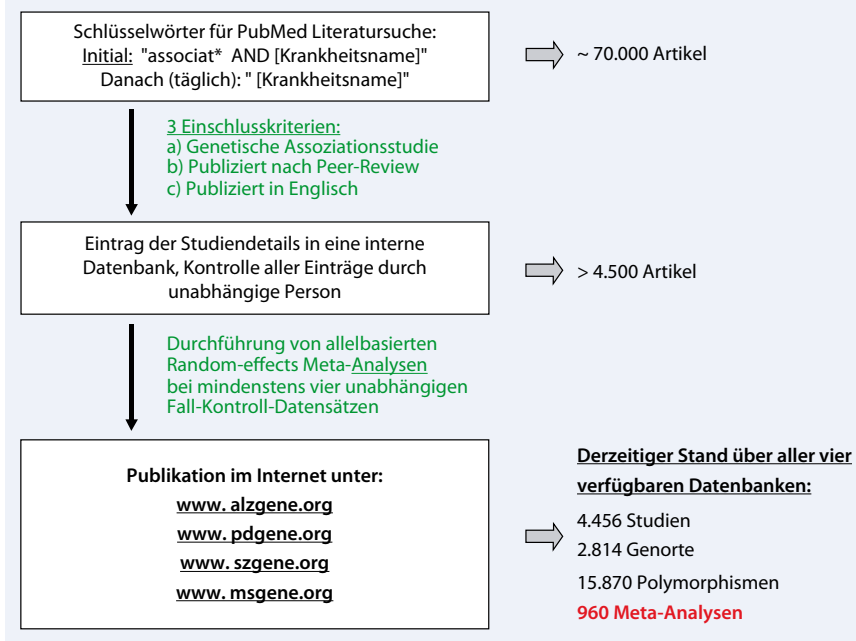


Abb. 1 ▲ Übersicht des Prozedere zur Datenerfassung und -analyse der im Text beschriebenen Datenbanken

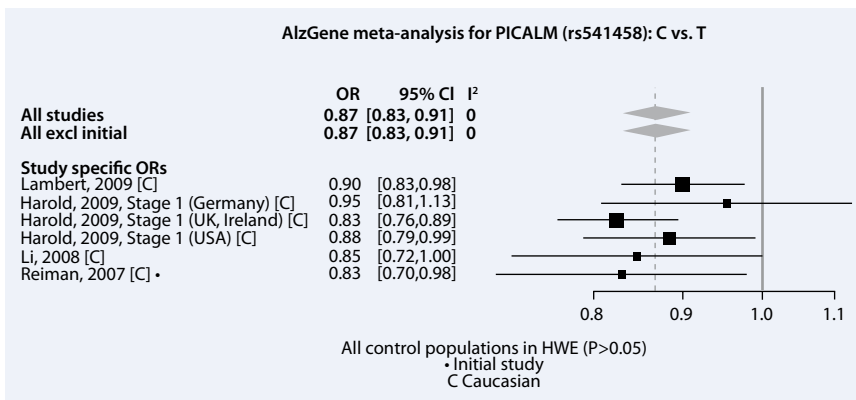


Abb. 2 ▲ Beispiel für Metaanalyse und Forest-Plot zum Gen *PICALM* in der AlzGene-Datenbank. Erstmals als AD-Kandidatengenen hervorgehoben wurde *PICALM* in der Publikation von Harold 2009 [5], während bereits die öffentlich zugänglichen GWAS-Datensätze von Reiman 2007 [7] und Li 2008 [8] nominal signifikante Assoziationen mit erhöhtem AD-Risiko für denselben Polymorphismus zeigten

zumeist gefolgt vom Follow-up der signifikantesten Resultate in unabhängigen Studienpopulationen.

Im Alzheimer-Feld sind bislang ein Dutzend GWAS veröffentlicht worden [3], im Parkinson-Feld sind es derzeit 7 Studien. Während die AD-GWAS immerhin fast alle übereinstimmend eine Assoziation mit Polymorphismen sehen, die sich im Kopplungsungleichgewicht mit dem ε4-Allel von *APOE* befinden, und die PD-GWAS relativ konsistent über die Assoziation mit *SNCA* und *MAPT* berichten, herrscht ansonsten sowohl für AD als auch für PD wenig Konsens bzgl. der Er-

gebnisse. Des Weiteren konnte die überwiegende Zahl der in den individuellen GWAS postulierten „GWAS-Gene“, deren Effektstärken – komplex-genetischen Erkrankungen entsprechend – mit durchschnittlichen ORs von ~1,2 nur moderat sind, bisher nicht oder nur teilweise repliziert werden. Oder sie wurden z. T. in keine nachfolgende Studie eingeschlossen, sodass eine Evaluation dieser Ergebnisse ohne systematischen Ansatz schwierig ist. Eine bemerkenswerte Ausnahme stellen zwei 2009 veröffentlichte AD-GWAS dar, die unabhängig voneinander bei insgesamt über 25.000 Patienten und Kontrol-

len konsistente genomweit signifikante Assoziationen mit einem Polymorphismus in *CLU* fanden [4, 5].

Systematische Synopsen und Metaanalysen

Beispiel der AlzGene- und PDGene-Datenbanken

Um die Interpretationen der Kandidatengen- und GWAS-Assoziationsstudien bei AD und PD zu erleichtern, haben wir die öffentlich zugänglichen „AlzGene-“ [1] und „PDGene-Datenbanken“ konzipiert. Diese Datenbanken listen alle publizierten Assoziationsstudien der jeweiligen Erkrankung inklusive GWAS, fassen die wichtigsten Studiendetails zusammen und bieten als integralen Bestandteil systematische Metaanalysen aller Polymorphismen mit ausreichenden Daten. Nach regelmäßigen Literatursuchen werden alle Assoziationsstudien eingeschlossen, die in einer englischsprachigen Zeitschrift nach Peer-Review veröffentlicht wurden. Hauptcharakteristika der Studien (z. B. Population, Ethnizität, Diagnosekriterien, Studiengröße, Erkrankungsalter) sowie Genotypedistribution und Allelfrequenzen der getesteten Polymorphismen werden in die Datenbank eingegeben. Bei Polymorphismen, für die 4 oder mehr unabhängige Fall-Kontroll-Datensätze vorliegen, werden allelische Random-Effects-Metaanalysen kalkuliert und deren Resultate auf den jeweiligen Datenbanken publiziert (■ Abb. 1).

Die Methodik der Metaanalysen folgt allgemein anerkannten Standardkriterien (z. B. den PRISMA Guidelines, s. <http://www.prisma-statement.org/>) und wird im Detail in den Primärpublikationen der Datenbanken beschrieben (s. [1, 2]). Durch diesen Ansatz können so die publizierten Genotypdaten quantitativ erfasst werden, und unabhängig vom Resultat, das die Autoren der einzelnen Studien nach jeweils unterschiedlichen Analysen beschreiben, kann ein zusammenfassendes Urteil gebildet werden. Darüber hinaus werden zusätzliche Metaanalysen nach gewissen Stratifikationskriterien durchgeführt: In einer Subanalyse wird die initiale (zumeist positive) Studie ausgeschlossen; des Weiteren werden alle

Datensätze, die eine signifikante Deviation ($p < 0,05$) vom Hardy-Weinberg-Äquilibrium (engl. „Hardy-Weinberg Equilibrium“, HWE) in den Kontrollpopulationen als Hinweis auf mögliche Genotypisierungsfehler zeigen, ausgeschlossen. Außerdem führen wir bei ausreichender Datenlage ethnizitätsspezifische Metaanalysen durch, um ethnischen Diversitäten Rechnung zu tragen. Die in den Metaanalysen signifikant positiven Polymorphismen werden dann nach Kriterien, die von dem Human Genome Epidemiology Network (HuGENet) entwickelt worden sind [6], auf epidemiologische Validität geprüft und klassifiziert. Schließlich werden sie auf den respektiven Internetseiten in einer separaten Sektion („Top Results“) besonders hervorgehoben.

Einige der genetischen Datenbanken (AlzGene und SZGene) sind integriert in Online-Kommunikationsforen der jeweiligen Erkrankungen, wo sich außer Informationen zu den genetischen Studien auch allgemeine Informationen zu den Krankheitsbildern finden. PDGene verweist auf äquivalente Informationen unter <http://www.pdonlineresearch.org/>.

Integration von GWAS-Daten

Die systematische Integration von GWAS-Daten und anderen ähnlich umfangreichen Studien stellt eine konzeptionelle und praktische Herausforderung für jede genetische Datenbank dar. Zum einen sind die Hauptbefunde der größtenteils hypothesenfrei generierten GWAS-Daten besonders interessant, zum anderen stellen auch die nichtassoziierten genetischen Marker einer GWAS eine wertvolle Ressource in weiterführenden Analysen dar, z. B. solchen, die im Rahmen unserer Datenbanken durchgeführt werden. Da allerdings die massive Akkumulation von größtenteils negativen Daten weder sinnvoll noch informativ ist, haben wir ein dreistufiges Protokoll für den GWAS-Dateneinschluss entwickelt:

Im 1. Schritt werden Daten für die Gene und Polymorphismen eingeschlossen, die von den Autoren als assoziiert hervorgehoben werden („featured genes“). Diese sind häufig durch bestimmte Filterpro-

medgen 2010 · 22:235–241 DOI 10.1007/s11825-010-0225-0
© Springer-Verlag 2010

C.M. Lill · L. Bertram

Online-Datenbanken und systematische Metaanalysen komplex-genetischer Erkrankungen

Zusammenfassung

Die Suszeptibilität komplex-genetischer Erkrankungen wird durch eine unbekannte Anzahl genetischer Faktoren bestimmt. In den letzten Jahrzehnten sind Hunderte von Suszeptibilitätsloci für die Alzheimer-Krankheit (AD), das idiopathische Parkinson-Syndrom (PD), Schizophrenie (SZ) und multiple Sklerose (MS) beschrieben worden. Diese schwierig zu überschauende Datenlage für jede dieser Erkrankungen wird kompliziert durch die ansteigende Zahl genomweiter Assoziationsstudien. Um die Evaluation und Interpretation der Ergebnisse zu erleichtern, haben wir Internet-Datenbanken für genetische Assoziationsstudien in AD, PD, SZ und MS generiert.

Neben einer detaillierten Synopse aller verfügbaren Studien bieten die Datenbanken allel-basierte Metaanalysen der Polymorphismen, zu denen ausreichende Daten vorliegen. In diesem Review stellen wir anhand der Datenbanken für AD (AlzGene) und PD (PDGene) den Hintergrund und die Implikationen dieses Ansatzes vor.

Schlüsselwörter

Neuropsychiatrische Erkrankungen · Alzheimer-Krankheit · Idiopathisches Parkinson-Syndrom · Genetische Assoziation · Metaanalyse

Online databases and systematic meta-analyses of genetically complex diseases

Abstract

Susceptibility to genetically complex disorders is determined by an unknown number of genetic determinants, and decades of intensive research have yielded hundreds of such potential susceptibility loci for Alzheimer's disease (AD), Parkinson's disease (PD), schizophrenia (SZ), and multiple sclerosis (MS). The results of genome-wide association studies are now adding to an already vast and complicated body of data. To facilitate the evaluation and interpretation of these findings, we have recently created online databases for genetic association studies in AD, PD, SZ, and MS. In addition to pro-

viding detailed summaries for each eligible study, the databases present the results of allele-based meta-analyses for all polymorphisms with sufficient genotype data. In this review, we discuss the background and implications of the database approach developed by our group, using current findings from the AD (AlzGene) and PD (PDGene) databases as examples.

Keywords

Neuropsychiatric traits · Alzheimer's disease · Idiopathic Parkinson's disease · Genetic association · Meta-analysis

Tab. 1 Zusammenfassung der derzeitigen „Top Results“ der AlzGene-Datenbank

Genort	Polymorphismus	Stratum	Studien (n)	Stichprobe (n)	OR (95%-KI)	p-Wert
ACE	rs1800764	C	4	1565	0,79 (0,68–0,92)	0,002
ADAM10	rs17269348	All	4	10982	1,15 (1,03–1,28)	0,009
APOE	APOE_e2/3/4	All	36	9413	3,69 (3,30–4,12)	<1×10 ⁻¹⁶
CALHM1	rs2986017	C	10	8441	1,18 (1,03–1,35)	0,016
CCR2	rs1799864	C	3	1426	0,73 (0,56–0,97)	0,024
CH25H	rs13500	All	7	3413	1,44 (1,07–1,94)	0,014
CHRN2	rs4845378	All	4	1363	0,67 (0,50–0,90)	0,007
CLU	rs11136000	All	9	26690	0,86 (0,82–0,69)	3,3×10 ⁻¹⁶
CR1	rs6656401	All	7	17181	1,19 (1,09–1,29)	3,1×10 ⁻⁰⁵
CST3	rs1064039	C	8	3014	1,16 (1,00–1,33)	0,043
DAPK1	rs4878104	All	7	5789	0,88 (0,82–0,96)	0,002
ECE1	rs213045	All	5	3691	0,86 (0,75–0,99)	0,044
ENTPD7	rs911541	All	4	7172	1,10 (1,01–1,21)	0,034
GAB2	rs2373115	All	12	12577	0,85 (0,76–0,94)	0,003
GAPDHS	rs4806173	All	4	3462	0,87 (0,75–0,99)	0,049
GRN	rs5848	All	4	3711	1,13 (1,02–1,25)	0,017
GWA_14q32.13	rs11622883	All	5	3568	0,84 (0,77–0,93)	0,0003
hCG2039140	rs1903908	All	4	2865	1,23 (1,06–1,44)	0,007
ILIA	rs1800587	C	19	8212	1,09 (1,01–1,18)	0,039
ILIB	rs1143634	C	6	2720	1,16 (1,02–1,32)	0,021
IL33	rs7044343	All	4	9865	0,85 (0,73–0,98)	0,021
IL8	rs4073	All	4	1593	1,27 (1,08–1,50)	0,004
LDLR	rs5930	All	4	1501	0,85 (0,72–0,99)	0,039
LOC439999	rs498055	All	7	5285	1,15 (1,02–1,30)	0,016
LOC651924	rs6907175	All	6	4882	0,89 (0,82–0,97)	0,004
MTHFR	rs1801133	All	26	8604	1,13 (1,04–1,23)	0,003
NEDD9	rs760678	All	8	6697	0,89 (0,81–0,98)	0,012
OTC	rs5963409	All	5	2864	1,17 (1,04–1,33)	0,008
PGBD1	rs3800324	All	7	6670	1,21 (1,02–1,45)	0,028
PICALM	rs541458	All	6	21915	0,87 (0,83–0,91)	3,5×10 ⁻¹⁰
PRNP	rs1799990	C	11	6301	0,92 (0,85–0,99)	0,045
SORCS1	rs600879	All	4	2856	1,24 (1,04–1,48)	0,017
SORL1	rs2282649	A	3	1666	1,30 (1,13–1,50)	0,00017
TF	rs1049296	All	16	9233	1,18 (1,06–1,31)	0,003
TFAM	rs2306604	All	6	2336	0,86 (0,75–0,99)	0,037
THRA	rs939348	All	6	5841	1,10 (1,01–1,19)	0,03
TNF	rs4647198	A	3	771	1,37 (1,05–1,79)	0,019
TNK1	rs1554948	All	5	3712	0,84 (0,76–0,93)	0,0006

„Top Results“, d. h. Genorte mit mindestens einem Polymorphismus mit nominal signifikanter (p-Wert $\leq 0,05$) Metaanalyse in AlzGene (<http://www.alzgene.org>) am 25.04. 2010. Für Genorte mit mehr als einem nominal signifikanten Polymorphismus ist nur das Resultat mit der größten statistischen Signifikanz gelistet. **Stratum:** All schließt alle ethnischen Subgruppen ein, wobei C und A sich nur auf kaukasische bzw. asiatische Studien beschränken. **Studien (n)** und **Stichprobe (n)** beziffert die Anzahl der in die Metaanalysen eingeschlossenen unabhängigen Datensätze bzw. die Anzahl der Fälle und Kontrollen. Es ist zu beachten, dass die jeweiligen Datenbanken regelmäßig aktualisiert werden, sodass die hier gezeigten Ergebnisse von denen der Online-Datenbanken abweichen können. **OR** Odds Ratio, **95%-KI** 95%-Konfidenzintervall.

zesse, z. B. Korrektur für multiples Testen, gegangen und/oder in vielen Fällen in unabhängigen Populationen repliziert worden und stellen die wichtigsten Ergebnisse der jeweiligen GWAS dar. Darüber hinaus werden von GWAS, die ihre Datensätze öffentlich zugänglich machen, auch die nichtassoziierten Genotypdaten verwendet:

Im 2. Schritt werden die Genotypdaten für alle Polymorphismen übernommen, für die bereits Datenbankeinträge vorliegen. So zeigte sich in den AlzGene-Metaanalysen der durch die beiden bereits oben erwähnten AD-GWAS-Studien identifizierten Polymorphismen in *CR1* und *PICALM*, dass wenigstens nominale Assoziationssignale bereits in sehr viel früher

veröffentlichten GWAS vorlagen [7, 8], die somit sozusagen die AD-Assoziationen mit diesen beiden Loci bereits „repliziert“ hatten, noch bevor diese als Top-GWAS-Signale identifiziert wurden (■ **Abb. 2**). Diese unmittelbare Replikation, die der Glaubwürdigkeit der neuen GWAS-Ergebnisse gleich ein sehr viel größeres Gewicht gibt, kann besonders effizient durch

Tab. 2 Zusammenfassung der derzeitigen „Top Results“ der PDGene-Datenbank

Genort	Polymorphismus	Stratum	Studien (n)	Stichprobe (n)	OR (95%-KI)	p-Wert
ABCB1	rs1045642	All	11	4205	0,90 (0,82–0,99)	0,04
APOE	rs7412	All	39	15783	1,15 (1,03–1,28)	0,012
BDNF	rs11030104	All	4	8178	1,12 (1,04–1,22)	0,003
BST1	rs11931532	A	3	20392	0,81 (0,76–0,87)	2,7×10 ⁻⁰⁹
CALB1	rs1805874	All	6	6553	0,89 (0,80–0,99)	0,033
CYP2D6	CYP2D6*4	All	34	12223	1,13 (1,01–1,25)	0,025
DRD2	rs1079597	All	5	4049	1,17 (1,01–1,36)	0,043
DRD3	rs6280	All	9	10936	1,08 (1,02–1,15)	0,012
FARP1	rs2282048	All	11	6580	0,91 (0,84–0,99)	0,033
FGF20	rs12718379	All	5	6326	0,92 (0,85–0,99)	0,02
GBA	GBA_N3705	All	18	16037	3,17 (2,21–4,55)	1,6×10 ⁻¹⁰
GLIS1	rs797906	All	4	8189	1,08 (1,01–1,16)	0,028
GSTM1	GSTM1*O	C	11	5832	0,89 (0,82–0,98)	0,01
GWA_2q36.3	rs10200894	C	18	13675	0,87 (0,78–0,97)	0,013
GWA_7p14.2	rs17329669	All	24	15658	1,13 (1,03–1,24)	0,01
LRP8	rs5174	All	4	8187	0,93 (0,87–0,99)	0,04
LRRK2	rs1994090	A	3	20392	1,39 (1,24–1,57)	1,2×10 ⁻⁰⁸
MAOB	rs1799836	All	11	3606	1,20 (1,02–1,42)	0,028
MAPT/STH	MAPT_H1H2	C	21	16846	0,75 (0,71–0,80)	<1×10 ⁻¹⁶
MTHFR	rs1801133	A	3	569	1,45 (1,11–1,90)	0,006
NUCKS1	rs823128	All	5	32300	0,73 (0,67–0,80)	5,1×10 ⁻¹²
PARK2	rs1801582	All	16	4812	0,79 (0,64–0,98)	0,027
PDXK	rs2010795	All	11	16295	1,09 (1,02–1,16)	0,008
PINK1	rs1043424	C	8	9456	0,91 (0,83–0,99)	0,019
PITX3	rs2281983	All	5	4285	1,10 (1,01–1,20)	0,041
PM20D1	Rs11240572	A	3	20392	0,76 (0,69–0,85)	1,9×10 ⁻⁰⁷
SLC41A1	rs947211	A	3	20392	0,76 (0,70–0,83)	8,5×10 ⁻¹⁰
SLC45A3	rs16856139	A	3	20392	0,74 (0,66–0,84)	9,3×10 ⁻⁰⁷
SLC6A3	rs72527228	C	6	2059	0,85 (0,74–0,98)	0,023
SNCA	rs356165	All	8	5696	1,43 (1,33–1,55)	<1×10 ⁻¹⁶
UCHL1	rs5030732	All	21	17560	0,92 (0,86–0,99)	0,014
USP24	rs13312	All	4	3072	0,76 (0,66–0,86)	2,6×10 ⁻⁰⁵

„Top Results“, d. h. Genorte mit mindestens einem Polymorphismus mit nominal signifikanter (p-Wert ≤0,05) Metaanalyse in PDGene (<http://www.pdgene.org>) am 25.04. 2010. Für Genorte mit mehr als einem nominal signifikanten Polymorphismus ist nur das Resultat mit der größten statistischen Signifikanz gelistet. **Stratum:** All schließt alle ethnischen Subgruppen ein, wobei C und A sich nur auf kaukasische bzw. asiatische Studien beschränken. **Studien (n)** und **Stichprobe (n)** beziffert die Anzahl der in die Metaanalysen eingeschlossenen unabhängigen Datensätze bzw. die Anzahl der Fälle und Kontrollen. Es ist zu beachten, dass die jeweiligen Datenbanken regelmäßig aktualisiert werden, sodass die hier gezeigten Ergebnisse von denen der Online-Datenbanken abweichen können. **OR** Odds Ratio; **95%-KI** 95%-Konfidenzintervall.

eine Datenbankstruktur wie etwa die von AlzGene realisiert werden. Darüber hinaus werden durch den 2. Schritt der Integration von GWAS-Daten für viele Polymorphismen, für die 3 oder weniger Datensätze aus Kandidatengen-Studien vorliegen, überhaupt erst Metaanalysen möglich. Ein solches Beispiel ist der Polymorphismus rs11030104 des Gens *BDNF* auf PDGene: Hier waren vor Einschluss der GWAS-Daten 2 Fall-Kontroll-Datensätze verfügbar. Nach Einschluss von 2 weiteren GWAS-Datensätzen konnte eine Metaanalyse durchgeführt werden, die nominal signifikante Assoziation zeigt. Mehr Stu-

dien sind notwendig um zu klären, ob sich diese Befunde bestätigen.

Im 3. Schritt werden Metaanalysen von GWAS-Datensätzen aufgenommen: Die von den Autoren herausgestellten Polymorphismen werden ähnlich wie im 1. Schritt in die Datenbank eingetragen. Zusätzlich werden die verwendeten Daten aus den betreffenden GWAS eingetragen, sofern sie verfügbar sind. Ein Beispiel dafür ist der Eintrag für *TPRG1* auf PDGene (s. <http://www.pdgene.org> für weitere Details).

Aktueller Stand von AlzGene und PDGene

Momentan umfasst AlzGene mehr als 1350 individuelle Studien, die etwa 2800 Polymorphismen in rund 700 Genen getestet haben; die Zahlen sind vergleichbar für PDGene mit ~900 Studien, ~2300 Polymorphismen und ~500 Genen. Darüber hinaus präsentieren AlzGene und PDGene augenblicklich jeweils über 300 bzw. 200 individuelle Metaanalysen – mehr als 5-mal so viele, wie in den jeweiligen Krankheitsfeldern im Laufe der letzten 30 Jahre

auf konventionelle Art und Weise durchgeführt und veröffentlicht wurden.

Darunter zeigen auf AlzGene momentan 37 vom *APOE-ε4*-Allel unabhängige Gene nominal signifikante Risikoeffekte mit mindestens einem Polymorphismus (■ **Tab. 1**). Für PDGene liegen für insgesamt 32 Gene nominal signifikante Metaanalysen vor (■ **Tab. 2**). Etwa ein Viertel dieser Gene war ursprünglich durch eine der publizierten GWAS als Risikogen postuliert worden. Wie in den ■ **Tab. 1 und 2** sichtbar wird, sind die Risikoeffekte der meisten Polymorphismen moderat mit einer durchschnittlichen OR von ~1,2. Die Ausnahme bei AD ist das *APOE-ε4*-Allel mit einer OR von 3,8 im allelischen Modell. Bei PD zeigen *SNCA*, *GBA* und *LRRK2* mit manchen Polymorphismen z. T. deutlich überdurchschnittliche Effektstärken.

Die relativ kleinen Effektstärken der Mehrzahl der Metaanalysen entsprechen den Ergebnissen von Studien anderer komplex-genetischer Erkrankungen [2, 9] und haben wichtige und bekannte Implikationen: Die benötigte Studiengröße für Assoziationsstudien bei AD und PD sowie anderen komplex-genetischen Erkrankungen ist weitaus größer, als es das Design vieler Studien, die heute noch publiziert werden, berücksichtigt: Um etwa eine OR von 1,25 mit 80% Trennschärfe („Power“) und einem p-Wert von 0,05 zu erkennen, sind für Allelfrequenzen von 0,5 bzw. 0,05 zusammen etwa 1400 bzw. 6000 Fälle und Kontrollen nötig [10]. Um genomweite signifikante p-Werte von $<5 \times 10^{-8}$ für dieselbe OR zu erreichen, müsste diese Studiengröße noch etwa um das 5-Fache angehoben werden.

Stärken und Limitationen des systematischen Datenbankansatzes

Die Stärken des Datenbankansatzes liegen auf der Hand: In der Annahme, dass die Literatursuchen, Einschlusskriterien, Dateneingabe und -analysen korrekt und hinreichend sind, sind sie die einzigen verfügbaren Datenbanken, die eine umfassende und anspruchsvolle Quelle über den aktuellen Status der genetischen Forschung in dem Bereich bieten. Des Weiteren hat unser Ansatz im Vergleich zu

konventionell publizierten Metaanalysen einen entscheidenden Vorteil: Konventionelle Metaanalysen sind durch die sehr hohe Publikationsfrequenz in diesen Forschungsgebieten sehr schnell, nicht selten schon vor ihrer Publikation, überholt. Unser dynamischer, internetbasierter Datenbankansatz erlaubt eine regelmäßige Aktualisierung der Metaanalysen, die somit jederzeit den aktuellen Forschungsstand widerspiegeln.

Außerdem konzentrieren sich die im Rahmen der Datenbankerfassung eingeschlossenen Polymorphismen nicht auf bestimmte Gene, chromosomale Regionen oder genetische Netzwerke, sondern umfassen alle getesteten Loci und ermöglichen dadurch und durch die einheitliche Methodik der Metaanalysen die Vergleichbarkeit der Ergebnisse zwischen den Studien, Genen, chromosomalen Regionen, bestimmten Pathways und selbst zwischen den 4 Datenbanken und damit unterschiedlichen Krankheiten. Auf der Homepage der Datenbanken zeigen wir außerdem eine Liste der Gene, für die die Metaanalyse wenigstens eines Polymorphismus einen nominal signifikanten Effekt zeigt. Diese Sektion wird „Top Results“ genannt und zeigt dem Benutzer der Datenbank sozusagen eine „Echtzeitaufnahme“ der vielversprechendsten Kandidatengene, die auf der Analyse von Hunderten von Studien und Tausenden von Daten beruht. Dies kann bei der Planung und Priorisierung zukünftiger Assoziationsstudien (z. B. weiterer Replikationsstudien, Fine-Mapping) oder funktioneller Studien zu diesen Kandidatengenen helfen.

Trotz der etablierten Rolle der Datenbanken im Management, der Präsentation und Interpretation der genetischen Forschung der 4 komplexen Erkrankungen gibt es in dem Datenbankkonzept natürlich auch Limitationen, die im Folgenden kurz erläutert werden sollen: Selbst mit einer umfassenden und systematischen Literatursuche und Dateneingabe wie oben beschrieben kann nicht ausgeschlossen werden, dass einige publizierten Studien übersehen oder dass bestimmte Daten falsch wiedergegeben werden. Dies kann durch die Datenbanknutzer selbst vermindert werden, die explizit auffordert werden, uns bei Fehlern oder

Versäumnissen zu kontaktieren. Außerdem schließen wir jede genetische Assoziationsstudie ein, die nach Peer-Review in englischer Sprache veröffentlicht worden ist, d. h. wir nehmen selbst keine Wertung der Studienqualität vor. Durch dieses objektive und liberale Vorgehen kommt es möglicherweise in Einzelfällen zu dem Einschluss fehlerhaft durchgeführter Studien. In diesem Kontext wird in unsere Metaanalysen auch jede Studie eingeschlossen, die die untersuchte Krankheitsentität nach anerkannten Standarddiagnosekriterien definiert. Dies führt möglicherweise in manchen Fällen zu einem Verlust von Trennschärfe durch den Einschluss verwandter, aber nicht identischer Erkrankungen, die unter denselben Diagnosekriterien zusammengefasst werden, jedoch nicht zu einer systematischen Verzerrung der Metaanalyse-Ergebnisse.

Eine andere Limitation liegt in der Beschränkung der Metaanalysen auf ausschließlich allelische Kontraste. Während dieses einheitliche Modell die Vergleichbarkeit zwischen einzelnen Metaanalysen oft überhaupt erst ermöglicht und ein hohes Maß an Objektivität erhält, erlaubt es keine Inferenz des Vererbungsmodus und hat weniger Trennschärfe als genotypbasierte Modelle, die die eigentliche Transmission besser charakterisieren. Andere Limitationen bestehen in dem Ausschluss von haplotypbasierten und berechneten („imputed“) Genotypdaten, womit wir möglicherweise wichtige Assoziationen nicht erfassen, sowie in der Fokussierung auf Haupteffekte (und dadurch bedingter fehlender Berücksichtigung von Gen-Gen- und Gen-Umwelt-Interaktionen) und der fehlenden Berücksichtigung von Kovariablen wie Alter und Geschlecht, die ohne Zugang zu Individualdaten nicht erfasst werden können. Der wichtigste Aspekt, der hier genannt werden sollte, ist jedoch, dass die Anzahl der „wahren“ Assoziationen zu jedem Zeitpunkt mit großer Wahrscheinlichkeit kleiner ist als die Anzahl der nominal signifikanten Metaanalysen der Datenbanken („Top Results“). Dies ist z. B. bedingt durch multiples Testen, Kopplungsungleichgewicht zwischen den assoziierten Varianten, undetektierter Bias sowie unbemerkte Studienartefakte.

Des Weiteren erreichen viele der positiven Metaanalysen keine sehr hohe statistische Signifikanz (■ Tab. 1, 2). Obwohl dies häufig durch kleine Studiengrößen und damit ungenügende Trennschärfe selbst in der Kombination aller verfügbaren Daten erklärt werden kann, muss auch hier in Betracht gezogen werden, dass diese Metaanalysen falsch-positiv sein können. Dasselbe gilt für hochsignifikante Metaanalysen mit wenigen Studien, deren Signifikanz durch eine einzige großangelegte Studie, etwa eine GWAS, bedingt ist. Letztlich können die wahren Risikofaktoren nur erkannt werden durch die Akkumulation von genügend hochqualitativen, verzerrungsfreien Genotypdaten (aus familienbasierten und Fall-Kontroll-Studien), die in konsistenter Weise die Assoziation belegen, in Verbindung mit funktionellen und biologischen Evidenzen.

Fazit für die Praxis

Da die Datenbanken durch die systematische Erfassung und Analyse der Literatur eine Synopse der vielversprechendsten Genorte für die jeweiligen Erkrankungen bieten, stellt sich die Frage, ob diese Liste der „Top Results“ zur individuellen Prädiktion der jeweiligen Erkrankung genutzt werden kann. Obwohl dies derzeit aus verschiedenen Gründen nicht möglich ist, gab es in den letzten Monaten Bestrebungen, z. B. im Bereich der multiplen Sklerose [11] und für verschiedene kardiovaskuläre Phänotypen [12], bestimmte Prädiktionsalgorithmen für die Kombination der bisher „etablierten“ Risikovarianten für Fall-Kontroll-Populationen zu entwickeln. Bisher war dieses Vorgehen allerdings von beschränktem Erfolg, was höchstwahrscheinlich an der insgesamt nur unvollständig erfassten genetischen und nichtgenetischen Kausalität der untersuchten Phänotypen liegt. Die zukünftige Identifizierung und Evaluierung neuer Risikovarianten dürfte zu einer progressiven Verbesserung dieser Algorithmen beitragen, und in diesem Kontext bieten die Datenbanken durch die kontinuierliche Integration der publizierten Daten eine adäquate Quelle für die Entwicklung verbesserter Algorithmen.

Zusammenfassend heben die Genorte und Varianten in den „Top Results“ der jeweiligen krankheitsspezifischen Datenbanken die zum aktuellen Zeitpunkt jeweils vielversprechendsten Kandidaten aus Hunderten untersuchter Gene hervor, deren weiterführende Analyse, z. B. anhand funktioneller und pharmakogenomischer Studien, mit hoher Priorität indiziert ist.

Korrespondenzadresse

Dr. L. Bertram
Neuropsychiatrische Genetik
Max-Planck-Institut für Molekulare Genetik
Ihnestraße 63, Raum 204.1
14195 Berlin
lbertram@molgen.mpg.de

Danksagung. Unser Dank gilt allen Mitarbeitern und Kollaborateuren, die zur Erstellung und Aktualisierung der besprochenen Datenbanken beigetragen haben (s. hierzu die „Credits-Sektionen“ auf den Internetseiten der jeweiligen Datenbanken), insbesondere den Mitarbeitern der Alzheimer- und Schizophrenia-Research-Foren, die die Datenbanken im Internet darstellen.

Interessenkonflikt. Der korrespondierende Autor weist auf folgende Beziehung(en) hin: AlzGene wird gefördert durch den Cure Alzheimer Fund (Waltham, USA), PDGene durch die Michael J. Fox Foundation for Research in Parkinson's disease (New York, USA), und SZGene durch die National Alliance on Research in Schizophrenia and Depression (NARSAD; New York, USA), alle Förderungen jeweils an L.B. als Principal Investigator.

Literatur

- Bertram L, McQueen MB, Mullin K et al (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet* 39:17–23
- Allen NC, Bagade S, McQueen MB et al (2008) Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* 40:827–834
- Bertram L, Tanzi RE (2009) Genome-wide association studies in Alzheimer's disease. *Hum Mol Genet* 18:R137–R145
- Lambert JC, Heath S, Even G et al (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 41:1094–1099
- Harold D, Abraham R, Hollingworth P et al (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 41:1088–1093
- Ioannidis JP, Boffetta P, Little J et al (2008) Assessment of cumulative evidence on genetic associations: interim guidelines. *Int J Epidemiol* 37:120–132
- Reiman EM, Webster JA, Myers AJ et al (2007) GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron* 54:713–720

- Li H, Wetten S, Li L et al (2008) Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* 65:45–53
- Lohmueller KE, Pearce CL, Pike M et al (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 33:177–182
- Bertram L, Tanzi RE (2008) Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. *Nat Rev Neurosci* 9:768–778
- De Jager PL, Chibnik LB, Cui J et al (2009) Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol* 8:1111–1119
- Ioannidis JP (2009) Prediction of cardiovascular disease outcomes and established cardiovascular risk factors by genome-wide association markers. *Circ Cardiovasc Genet* 2:7–15