

# **Workshop: Datenbank genomischer Varianten für die klinische Anwendung und die medizinische Forschung**

**Essen 19. März 2014**

---

## **Humane Genotyp-Phänotyp-Datenbanken: Ein Statusbericht**

Thomas Bettecken

Max-Planck-Institut für Psychiatrie

München

Interessenkonflikte (coi): Keine



## Sequenz-Varianten im menschlichen Genom

**62.676.337 (44.278.189)**

Varianten mit rs-Nr. (validierte)

**22.508.883**

davon in Genen

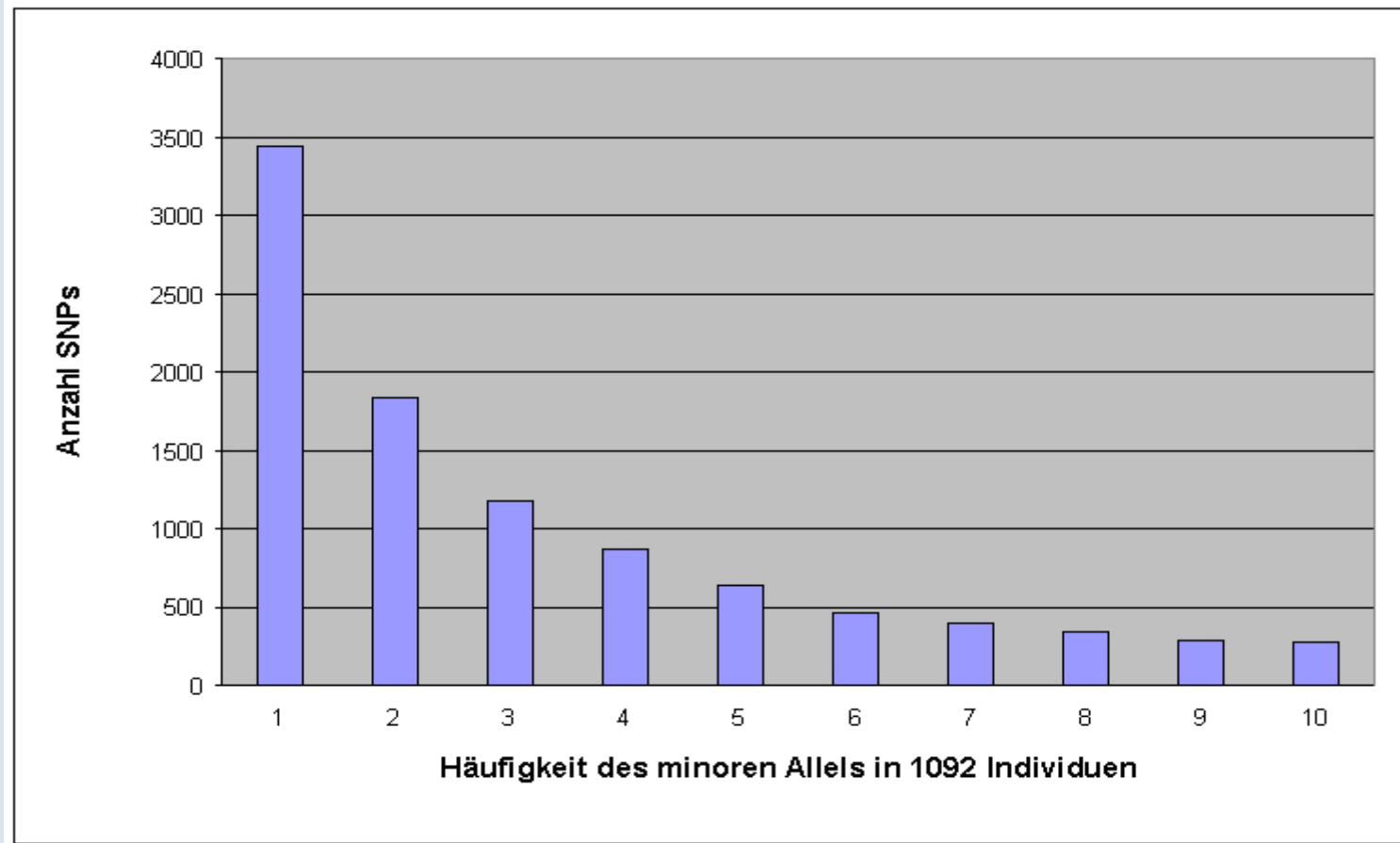
(exemplarisch - dbSNP build 138 April 2013)

Es gibt noch viel mehr Varianten im Genom !

**Hat eine Variante eine Bedeutung  
im Hinblick auf Gesundheit und Krankheit ?**

**Welche Bedeutung?**

# 17,580 SNPs in einer 1 MBasen-Region (16p)



3,449 SNPs (~20%) kommen in 1092 Individuen nur einmal vor

## Zytogenetik

Erstes Karyogramm des Menschen (Tao C Hsu 1952) / Homo sapiens 46 Chromosomen (Joe Hin Tjio 1956)

Down-Syndrom (Jerome Lejeune et al. 1959)

→ Catalogue of Unbalanced Chromosome Aberrations in Man (Albert Schinzel 1984)

## Molekulargenetik

DNA-Sequenzierung (Frederick Sanger, Allan Maxam und Walter Gilbert 1977)

Erste Mutation auf DNA-Ebene: Im Insulin-Gen eines Diabetes-Patienten (Simon CM Kwok et al. 1983)

→ Human Gene Mutation Database HGMD (David N Cooper und Michael Krawczak 1993)



## Unbekannte Variante / Mutation - Was tun ?

z.B. Stop-Codon- oder Frameshift-Mutation in einem „neuen“ Gen.

Vor Ort kein zweiter Fall bekannt (derselbe Phänotyp - dieselbe Mutation / ähnlicher Phänotyp – gleich gelagerte Mutation).

Welche Vergleichsmöglichkeiten (Datenbanken, ...) gibt es ?

**There is no One-Stop-Shop !**

## Locus-übergreifende Datenbanken mit Phänotyp-Info

Human Gene Mutation Database – HGMD (professionell Version)

136,000+ SNV plus CNV, 5300+ Gene (???), sehr sorgfältig annotiert  
(alles aus der Literatur)

ClinVar

75,299 SNV+CNV, 18,701 Gene, limitierte Phänotyp-Information

OMIM

17,694 SNV + CNV (die meisten davon kausal), 3,174 Gene mit  
ausgewählten kausalen Mutationen, 4,000+ Phänotyp-Beschreibungen  
mit bekannter molekularer Ursache, insgesamt 14,000+ Gene

Exome Variant Server EVS / Exome Sequencing Project ESP

Alle Exom-Varianten von 6503 Individuen,  
Phänotyp-Focus: Blut-, Herz und Lungenerkrankungen,  
aber auch CF und wenige andere seltene Erkrankungen

SNV – Single Nucleotide Variants

CNV – Copy Number Variants

## Locus-übergreifende Datenbanken mit Phänotyp-Info

### Database of Genotypes and Phenotypes - dbGAP

Ergebnisse aus GWAS (Schwerpunkt) und Diagnostischer Sequenzierung,  
z.Zt. 25,121 SNV+CNV

### GeneReviews

Sehr sorgfältige, über Jahre hinweg, erstellte und ausführliche Beschreibungen von 600+ erblichen Erkrankungen, mit Phänotyp-Info. Mutationen sind nur selektiv verzeichnet. Im Stil ähnlich einem Lehrbuch.

### Orphanet / Orphadata

Genzentrierte Datenbank, mit Phänotyp-Beschreibungen, Links zu Locus-spezifischen Datenbanken und zu Einsendern („Broker-Funktion“).

### Gen2Phen

Projekt zur Erstellung/Bereitstellung der Infrastruktur für eine Genotyp-/Phänotyp-Datenbank, mit dezentraler Eintragung (Cafe Variome).

## Locus-Spezifische Datenbanken (LSDB)

### 1. LOVD – Leiden Open Variation Database (Univ. Leiden, NL)

Quelloffenes Datenbanksystem, z.Zt. 2,124,696 Varianten in >22,000 Genen aus 180,656 Individuen, verzeichnet in ~90 dezentralen LOVD Installationen.

### 2. PhenCode (Penn State Univ. & Univ. of California SC, USA)

Links zu krankheitsspezifischen Gen- und Mutationsdatenbanken einschließlich Swissprot, z.Zt. 34,213 Varianten in 1,386 Genen, plus Swissprot: 56,396 Varianten in 11,343 Genen.



## Varianten-Datenbanken ohne Phänotyp-Eintrag

### 1000 Genomes Project

Genomweite Sequenzdaten von z. Zt. 1,092 Individuen  
(Phase 1, geplant sind 2,577 Individuen)

### dbSNP

(z.Zt. 62,676,337 Einträge für genomische Varianten)  
Sammlung von Sequenz-Varianten,  
Locus-abhängige Anzahl an Individuen

Weitere ... (u.a. DmuDB, DECIPHER)

Diverse kommerzielle Datenbanken / Lösungen

Und ... die „Schubladen“ der Kollegen  
 (Varianten „in Auswertung“, nicht-  
 publizierte / ungelöste Fälle, ....)



1. Es existieren eine ganze Reihe großer Genotyp-Phänotyp-Datenbanken.
2. Diese Datenbanken sind leider weitgehend unvernetzt (kein „One-Stop-Shopping“).
3. Sehr viele Varianten/Mutationen befinden sich noch „in Auswertung“, i.e. sie sind zwar erkannt, aber noch nicht abschließend interpretiert und nicht in Datenbanken verzeichnet.
4. Die Zahl der Varianten „in Auswertung“ wächst stark mit zunehmender Anwendung der NGS.



**Vielen Dank für Ihre Aufmerksamkeit!**

---

Diese Präsentation ist demnächst verfügbar:

**[www.tmf-ev.de](http://www.tmf-ev.de)**

[bettecken@mpipsykl.mpg.de](mailto:bettecken@mpipsykl.mpg.de)