

Herausforderung: Archivierung genomischer Hochdurchsatzdaten



**PROF. DR. MED. JOCHEN HAMPE
KLINIK FÜR INNERE MEDIZIN I
UNIVERSITÄTSKLINIKUM SCHLESWIG-HOLSTEIN
CHRISTIAN-ALBRECHTS-UNIVERSITÄT ZU KIEL**

Genomisches Datenspektrum



Keimbahn-DNA

Somatische
DNA

mRNA

Translation

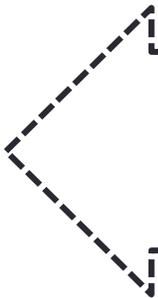
Proteine

Zellstruktur

Metabolismus

Biologische Relevanz

Technologische Reife
„ease of global analysis“



Projektumfeld I



Datentypen

Genomweite Genotypdaten:

Für einen Datensatz:

N=700,000 – 2,500,000

Polymorphismen

- Rohdaten (color-codes/intensities)
- Kontextabhängiger Genotype-call (im Gesamtdatensatz)
- Imputierter Genotypdatensatz auf N=5,500,000 Varianten

Klinische Daten: Phänotyp, Überleben, epidemiologische Variablen

Projektumfeld

Kartierung von Risikovarianten:

Gallensteinleiden (DFG)

Gallenblasenkarzinom (DFG)

Kolonkarzinom (NGFN/BmBF)

Internationale Dimension:

Datensätze aus UK / Chile /

Belgien

Projektumfeld II



Datentypen

Hochdurchsatzsequenzdaten:
Für einen Datensatz:
Gigabasen bis Terrabasen von
„finished“ Sequence

- Rohdaten (color-codes/intensities)
- Kontextabhängige Sequenz-calls
- Technologiewechsel 2nd auf 3rd Generation steht an.

Klinische Daten: Phänotyp,
Überleben, epidemiologische
Variablen, Gewebetyp,
Qualitätskriterien

Projektumfeld

Funktionell Genomische Projekte:
SNP-abhängige Splicen (DFG)
Translationale Regulation des
ER-Stresses bei Leberverfettung
(BmBF)

Virtual Liver (SysBio/BmBF)

IHEC – Leberdatensätze
(beantragt)

Zusammenfassung / Ausblick



- Komplexe Datenstrukturen wie beschrieben
- In keinem der Projekte Mittel oder Infrastruktur zur Archivierung
- Klinische / Personenbezogene Daten
 - Anbindung an Pseudonymisierungsservice POPGEN (external ID?)
 - Datenreduktion – separate Datenhaltung?
- Genomische Daten
 - „finished Data“ versus Rohdaten
 - Rohdaten: Binär/Bilddatensätze, häufig proprietäre Formate
 - Enddaten: Einfacher zu standardisieren, annotierte Textformate
 - Enddaten: Genotype-Call existiert nur im Kontext eines Experiments.

Projektgruppe



UKSH Kiel - Statistics: Michael Krawczak, Michael Nothnagel
UKSH Kiel – Genomic Gastroenterology/Hepatology, IKMB
FLI Jena: Karol Szafranski, Matthias Platzer

