



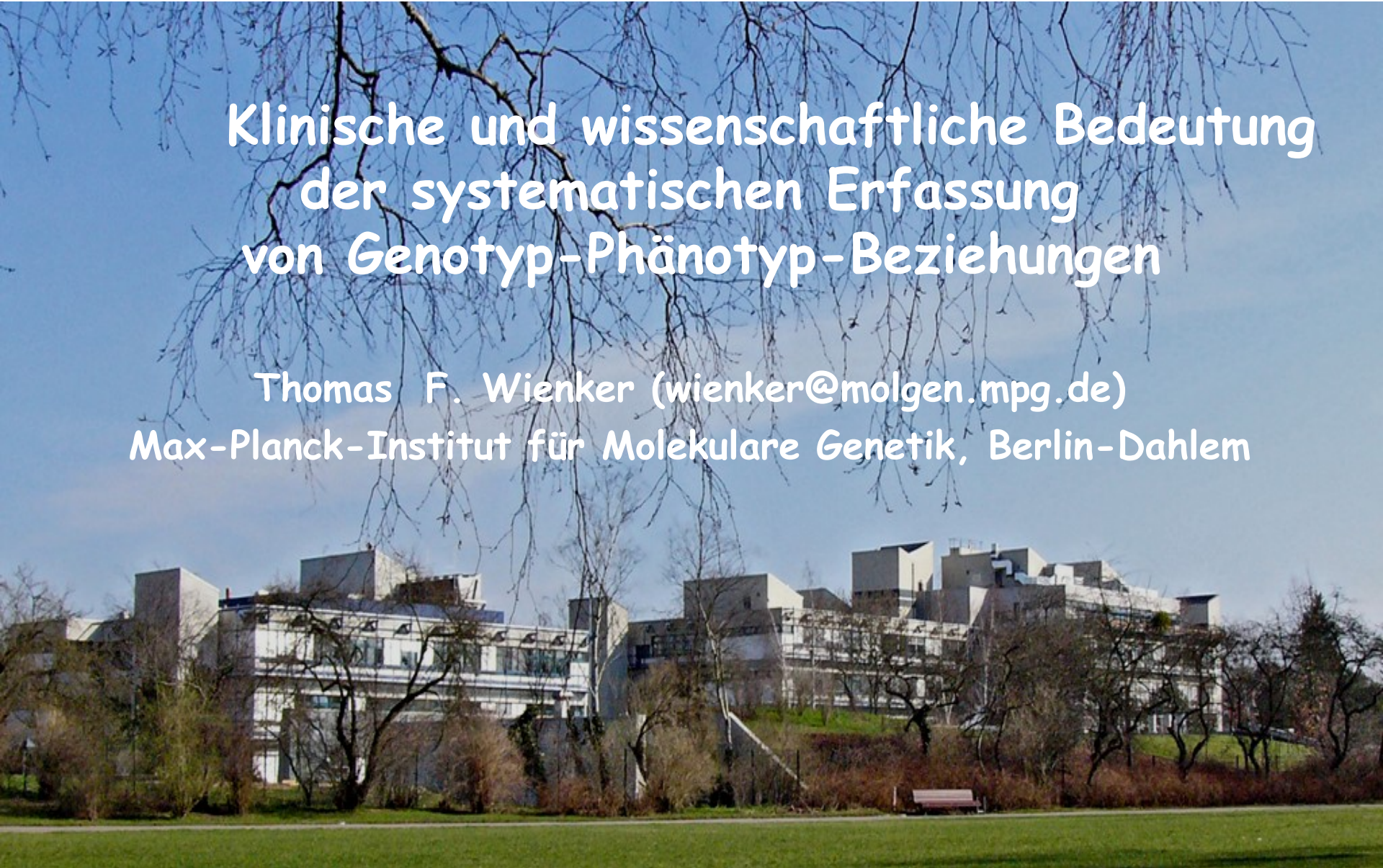
6. TMF-Jahreskongress

2.-3. April, Jena



Klinische und wissenschaftliche Bedeutung der systematischen Erfassung von Genotyp-Phänotyp-Beziehungen

Thomas F. Wienker (wienker@molgen.mpg.de)
Max-Planck-Institut für Molekulare Genetik, Berlin-Dahlem



NGS

Next Generation Sequencing



Camille Flammarion, *L'Atmosphère. Météorologique populaire*; Paris 1888

NGS: Next Generation Sequencing

- 1) Die eigentliche Sequenzierung
Erzeugung der ca. 100bp langen **“reads“**
jeder read entspricht einem fastq-File mit je einer Zeichenkette für die Nukleotidsequenz und die read-Qualität
total $200 - 800 \times 10^9$ reads
- 2) Abgleich und Abbildung auf das Referenzgenom, sog. **“alignment“**
Erzeugung der ca. 100bp langen “reads“
jeder read entspricht einem fastq-File mit je einer Zeichenkette für die Nukleotidsequenz und die read-Qualität
Abdeckung (**“coverage“**) und Tiefe (**“depth“**) der Sequenzierung
- 3) Charakterisierung der Abweichungen vom Referenzgenom
“variant calling“
 - single nucleotide variants, SNVs
 - kleine Insertionen oder Deletionen, indels
 - strukturelle Varianten, copy number variants, CNVs
- 4) Aussonderung irrelevanter Varianten (**“filtering“**) und Bewertung hinsichtlich der Auswirkungen auf den Phänotyp bzw. hinsichtlich der krankheitsursächlichen Relevanz (**“annotation“**)

Genome Reference Consortium: GRCh38 [24.12.2013]

Global stats for GRCh38

General Info	
Assembly Type	haploid with alt loci
Release Type	major
Number of Assembly Units	37
Total Bases in Assembly	3,209,286,105
Total Non-N Bases in Assembly	3,049,316,098
Primary Assembly N50	67,794,873
Region Information	
Total number of defined regions	207
Number of Regions with Alternate Loci	178
Number of Regions with Fix Patches	0
Number of Regions with Novel Patches	0
Number of Regions as PAR	4
Alternate Loci/PATCH Information	
Total Number of Alternate Loci scaffolds	261
Number of Alternate Loci scaffolds aligned to the Primary Assembly	261



NGS: Next Generation Sequencing

4) Aussonderung irrelevanter Varianten ("*filtering*") und Bewertung hinsichtlich der Auswirkungen auf den Phänotyp bzw. hinsichtlich der krankheitsursächlichen Relevanz ("*annotation*")

4A) population frequency; if $\text{freq.}(\text{var}) \geq 0.01$, than \pm irrelevant

4B) conservation scores, phylogenetic (ortholog) comparisons

4C) prediction of deleteriousness; consequences of amino acid exchange on structure, activity, and stability



NGS: Next Generation Sequencing

4) Aussonderung irrelevanter Varianten ("*filtering*") und Bewertung hinsichtlich der Auswirkungen auf den Phänotyp bzw. hinsichtlich der krankheitsursächlichen Relevanz ("*annotation*")

4A) population frequency; if $\text{freq.}(\text{var}) \geq 0.01$, than \pm irrelevant

4B) conservation scores, phylogenetic (ortholog) comparisons

4C) prediction of deleteriousness; consequences of amino acid exchange on structure, activity, and stability

4D) functional tests, *in vitro* (cell culture) or *in vivo* (experimental animals)

NGS: Next Generation Sequencing

4) Aussonderung irrelevanter Varianten (*“filtering”*) und Bewertung hinsichtlich der Auswirkungen auf den Phänotyp bzw. hinsichtlich der krankheitsursächlichen Relevanz (*“annotation”*)

4A) population frequency; if $\text{freq.}(\text{var}) \geq 0.01$, than \pm irrelevant

4B) conservation scores, phylogenetic (ortholog) comparisons

4C) prediction of deleteriousness; consequences of amino acid exchange on structure, activity, and stability

4D) cosegregation analysis
i.e. testing both parents and sibs (affected and unaffected, if existing and cooperative)

Pedigree: M 90 00 029 Gene: TLR10 HGNC:15634 OMIM: *606270 Location: 4p14

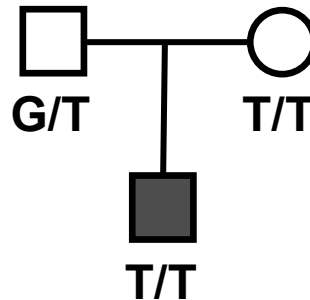
Mutation: Genomic: NC_000004.11:g.38,776,478G>T
 Transcript(s): NM_001017388.2:c.734C>A
 NM_001195106.1:c.734C>A
 NM_001195107.1:c.734C>A
 NM_001195108.1:c.692C>A
 NM_030956.3:c.734C>A
 NM_001017388.1:c.734C>A
 NM_030956.2:c.734C>A
 Protein: NP_001017388.1(TLR10):p.S245X



T/T

Pedigree: M 90 00 029 Gene: TLR10 HGNC:15634 OMIM: *606270 Location: 4p14

Mutation: Genomic: NC_000004.11:g.38,776,478G>T
 Transcript(s): NM_001017388.2:c.734C>A
 NM_001195106.1:c.734C>A
 NM_001195107.1:c.734C>A
 NM_001195108.1:c.692C>A
 NM_030956.3:c.734C>A
 NM_001017388.1:c.734C>A
 NM_030956.2:c.734C>A
 Protein: NP_001017388.1(TLR10):p.S245X



Cosegregation: negative, discard

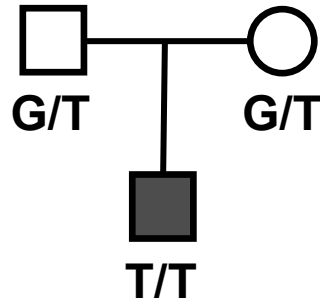
NGS: Next Generation Sequencing

*There is no genetic diagnosis
(sensu strictu)
outside the framework of the
Mendelian paradigm.*

cf. Ken Weiss, TMF-School extra, Berlin, 24. bis 28. September 2012

Pedigree: M 90 00 029 Gene: TLR10 HGNC:15634 OMIM: *606270 Location: 4p14

Mutation: Genomic: NC_000004.11:g.38,776,478G>T
 Transcript(s): NM_001017388.2:c.734C>A
 NM_001195106.1:c.734C>A
 NM_001195107.1:c.734C>A
 NM_001195108.1:c.692C>A
 NM_030956.3:c.734C>A
 NM_001017388.1:c.734C>A
 NM_030956.2:c.734C>A
 Protein: NP_001017388.1(TLR10):p.S245X



Cosegregation: positive, go on

NGS: Next Generation Sequencing

4) Aussonderung irrelevanter Varianten ("*filtering*") und Bewertung hinsichtlich der Auswirkungen auf den Phänotyp bzw. hinsichtlich der krankheitsursächlichen Relevanz ("*annotation*")

4A) population frequency; if $\text{freq.}(\text{var}) \geq 0.01$, than \pm irrelevant

4B) conservation scores, phylogenetic (ortholog) comparisons

4C) prediction of deleteriousness; consequences of amino acid exchange on structure, activity, and stability

4D) cosegregation analysis, i.e. testing parents and sibs

4E) evidence based, database searches for similar findings with respect to genes and/or variants

NGS: Next Generation Sequencing

Nachdem **alle** Schritte der Filterung und Bewertung der Gene bzw. Varianten (4A bis 4D) durchlaufen sind, bleibt als letzter und entscheidender Schritt der

empirische (evidenzbasierte) Vergleich

d.h. die Suche nach

**“genotypisch und phänotypisch
gleichgelagerten Fällen“**

in einschlägigen Datenbanken (**Referenzdatenbanken**).

Referenzdatenbanken enthalten zu Genen und Varianten gesicherte Information.

NGS: Next Generation Sequencing

Referenzdatenbanken (variation resources) gibt es für

- neutrale Varianten, d.h. Varianten, die keine Abweichung vom normalen Phänotyp bewirken, und insbesondere nicht krankheitsverursachend sind,
z.B. Exome Variant Server (EVS, University of Washington),
dbSNP (URL: <http://www.ncbi.nlm.nih.gov/snp>)
- pathogenetisch relevante Varianten (~ Mutationen), die gesichert krankheitsverursachend sind,
z.B. OMIM (Online Mendelian Inheritance in Man)
HGMD (Human Gene Mutation Database)
OrphaNet/OrphaData, URL: <http://www.orphadata.org/>
- Varianten beider Typs, aber mit verlässlichen Angaben zur phänotypischen Relevanz (“clinical significance“, “functional significance“, “clinical phenotype“; cave: keine standardisierte Terminologie),
z.B. ClinVar, URL: <https://www.ncbi.nlm.nih.gov/clinvar/>

NGS: Next Generation Sequencing

Referenzdatenbanken (variation resources) sind auch danach zu unterscheiden, ob sie Varianten versammeln zu

- einem oder vielen Genen (Genorten, Loci), sog. locus-spezifische Datenbanken, z.B. Leiden Open Variation Database (LOVD), UK–CysticFibrosis Database, UMD–CFTR (France)
- zu einer bestimmten Krankheit oder einem Syndrom oder Krankheitsgruppe (OMIM: *“phenotypic series“*), Register (z.B. *“multicentre international registry of familial mediterranean fever“*)
- zu einem bestimmten Typ von Varianten, z.B. strukturelle Varianten (*“copy number variants“*, CNVs, z.B. DECIPHER) oder chromosomale Aberrationen



March 12, 2014

Despite Hurdles, NGS is Changing Clinical Practice for Rare Disease, Cancer

By [Monica Heger](#)

Now that a number of institutions and companies have launched clinical exome and whole-genome pipelines for diagnosing rare disorders, the diagnostic rate has plateaued at around 25 percent, according to researchers that presented at last week's conference.

Of its first 1,012 cases, 265, or around 26 percent, have been solved, Beaudet said. "In the majority of those [unsolved] cases, the mutation is in the data, but we're not able to identify it."

Transitorische Genotyp–Phänotyp Datenbank

Problem:

- NGS (WGS, WES oder Panel-Diagnostik) erfolgreich durchgeführt
- alle seltenen/privaten Varianten sind annotiert und bewertet
- alle Computer-gestützten Analysen sprechen für Pathogenität
- kein Match in einer Datenbank für neutrale/polymorphe Varianten
- kein Match in einer Datenbank für krankheitsverursachende Varianten
- reguläre Segregation entspr. genetischem Modell in der Kernfamilie

1. Eingrenzung & Beschreibung des Problems

2. Mittel & Wege zur Problemlösung

Transitorische Genotyp–Phänotyp Datenbank

Was ist eine “diagnostische Situation“ ?

gegeben: Patient/Familie mit einer klinisch-genetisch definierten Krankheit

gesucht: die pathogenetisch relevante Variante (Mutation)
im ursächlich relevanten Genlocus

D \Rightarrow **G**

nicht:

G \Rightarrow **D**

Untersuchung von Gesunden, z.B. Screening,
oder nebenläufige Befunde, “incidental/unsolicited“ findings

Transitorische Genotyp–Phänotyp Datenbank

Gene oder Varianten sind in ihrer diagnostischen Wertigkeit nicht abschliessend zu beurteilen, weil gleichgelagerte Vergleichsfälle fehlen; die Beurteilung ist bestenfalls “tentativ“, jedoch nicht definitiv.

Dieser vorläufige Schwebeszustand wird in Anlehnung an den in der Genkartierung üblichen Sprachgebrauch als ***in limbo*** bezeichnet.

Der “in limbo“ Beurteilung einer Beobachtung wird aufgelöst durch eine zweite Beobachtung :

- neutrale Variante, keine pathogenetische Relevanz
- pathogenetisch relevante, kausal relevante Variante

(Evidenzbasierte Beurteilung)

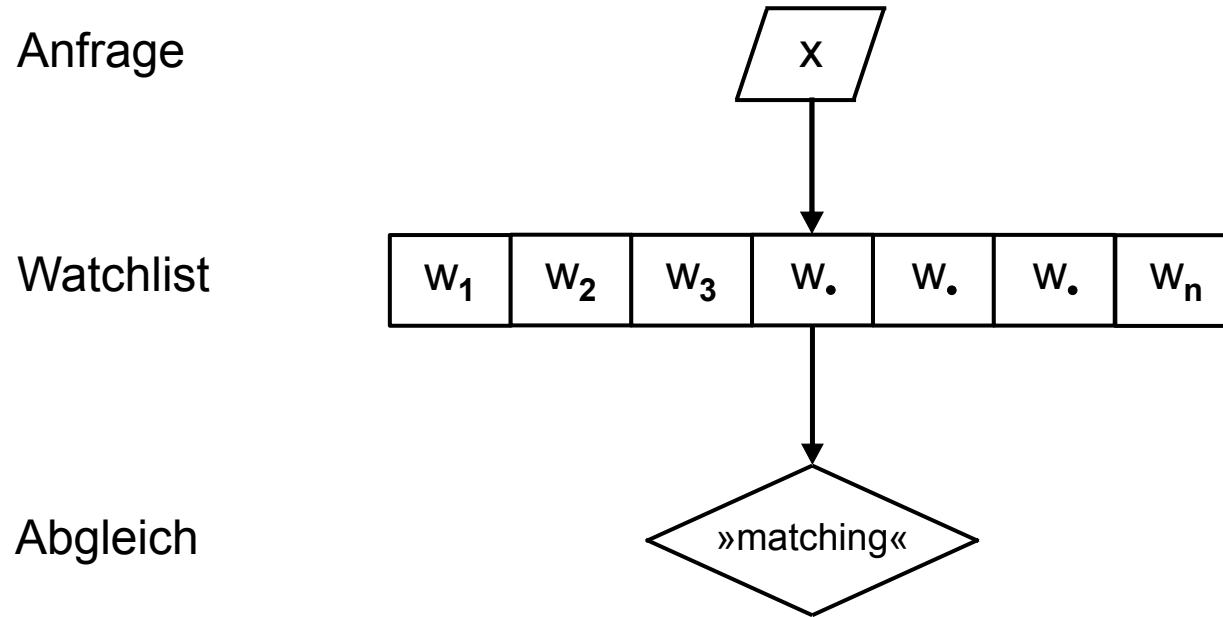
Die rasche Auflösung des “*in limbo*“ Zustandes ist Ziel dieser Datenbank.

Transitorische Genotyp-Phänotyp Datenbank

Watchlist

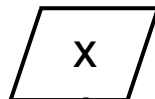
W_1	W_2	W_3	W_{\cdot}	W_{\cdot}	W_{\cdot}	W_n
-------	-------	-------	-------------	-------------	-------------	-------

Transitorische Genotyp-Phänotyp Datenbank

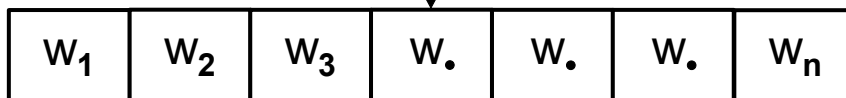


Transitorische Genotyp-Phänotyp Datenbank

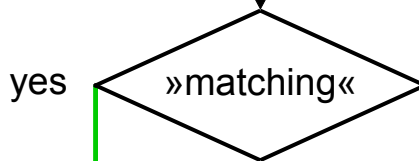
Anfrage



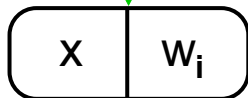
Watchlist



Abgleich



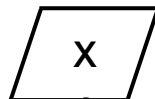
Aktion



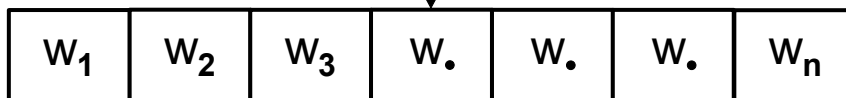
*automatische
Benachrichtigung*

Transitorische Genotyp-Phänotyp Datenbank

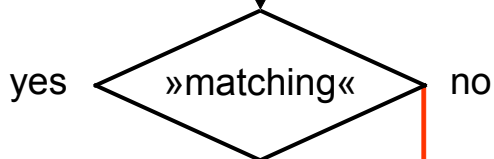
Anfrage



Watchlist



Abgleich

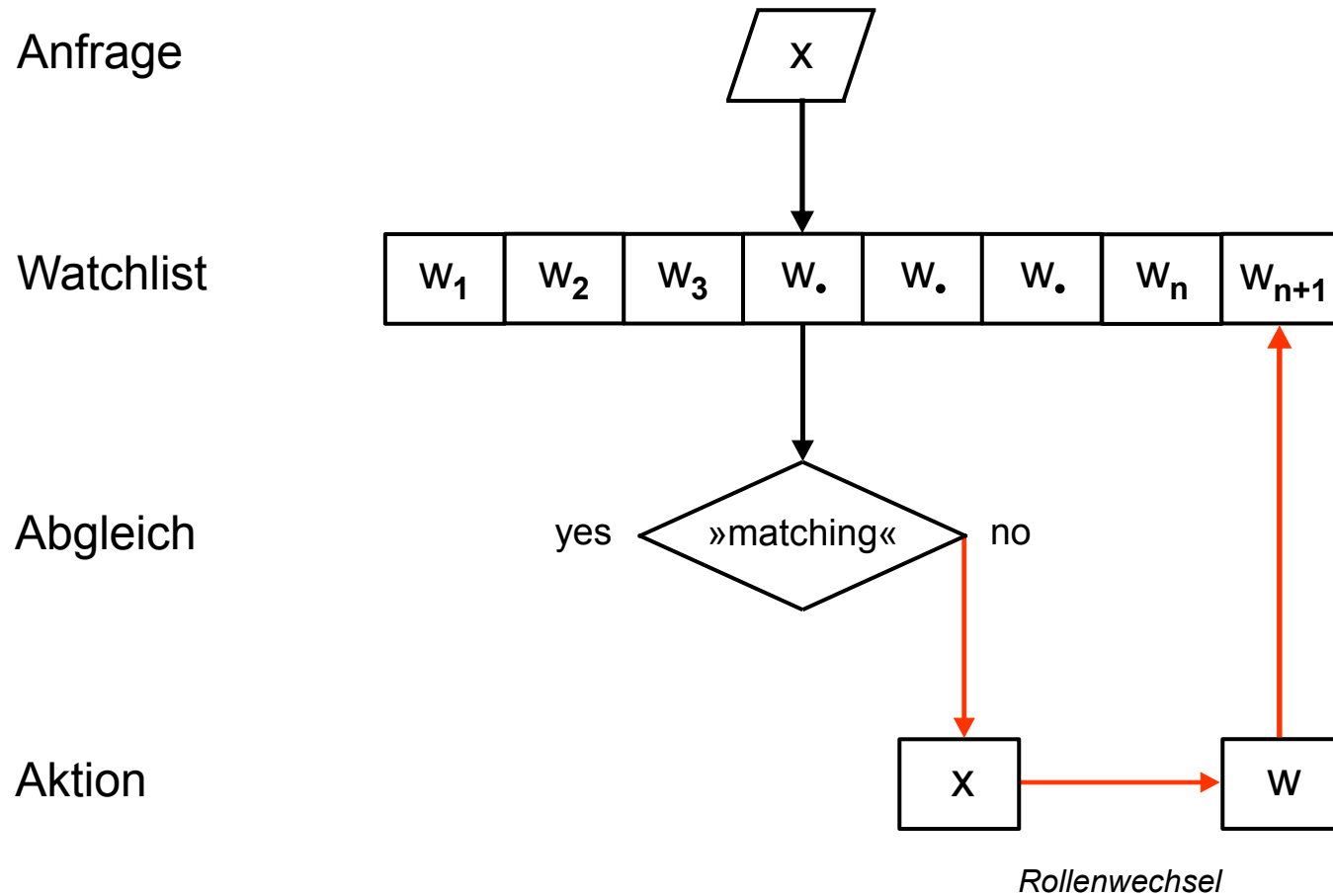


Aktion



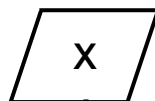
Rollenwechsel

Transitorische Genotyp-Phänotyp Datenbank

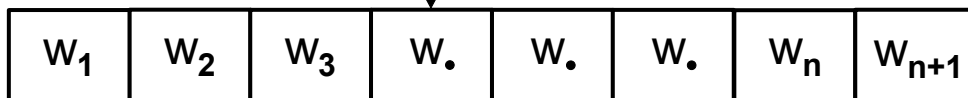


Transitorische Genotyp-Phänotyp Datenbank

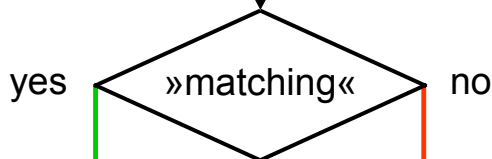
Anfrage



Watchlist



Abgleich



Aktion



automatische Benachrichtigung



Rollenwechsel

Transitorische Genotyp–Phänotyp Datenbank

Typische Szenarien für eine Anfrage

A) “novel gene“

Ein Genlocus, der bisher nicht als Krankheitsgen identifiziert wurde (OMIM, OrphaNet/OrphaData, HGMD, ClinVar, DmuDB, LOVD u.a.m.) mit einer seltenen/privaten, molekular-pathophysiologisch pathogenen Variante (by computational methods, z.B. Mutation Taster)

Protein coding genes: ~ 22.000 (RefSeqGene, hg18)
Disease genes: ~ 3.000 (OMIM, OrphaData)

Transitorische Genotyp–Phänotyp Datenbank

Typische Szenarien für eine Anfrage

A) “novel gene“

Ein Genlocus, der bisher nicht als Krankheitsgen identifiziert wurde (OMIM, OrphaNet/OrphaData, HGMD, ClinVar, DmuDB, LOVD u.a.m.) mit einer seltenen/privaten, molekular-pathophysiologisch pathogenen Variante (by computational methods, z.B. Mutation Taster)

B) “novel variant“

Eine molekular-pathophysiologisch überzeugende oder verdächtige Variante in einem bekannten Krankheitsgen, bei kompatiblem Phänotyp (variant of unknown significance, VUS)

C) “novel phenotype“

Bekanntes Krankheitsgen, aber Phänotyp ausserhalb des bekannten Spektrums (OMIM: phenotypic series)

Transitorische Genotyp–Phänotyp Datenbank

Minimal-suffiziente Daten für die typischen Szenarien

A) “novel gene“

- gene symbol (HGNC)
- genetisches Modell (mode of inheritance, MOI)
- Phänotyp-Kategorie (top level HPO term, u.U. auch second level)

B) “novel variant“

- gene symbol (HGNC, OMIM genotype [*] and phenotype [#])
- genetisches Modell, MOI
- genomische Variante, HGVS mit Referenzsequenz (*“assembly“*)
z.B. **NC_000015.9:g.66995761G>T**

C) “novel phenotype“

Dieses Szenario wird überführt nach A) oder B), oder A) und B)

Human Phenotype Ontology, HPO: <http://www.human-phenotype-ontology.org/>

HPO:0001871	Abnormality of blood and blood-forming tissues
HPO:0003549	Abnormality of connective tissue
HPO:0000152	Abnormality of head and neck
HPO:0001939	Abnormality of metabolism/homeostasis
HPO:0001197	Abnormality of prenatal development or birth
HPO:0001438	Abnormality of the abdomen
HPO:0000769	Abnormality of the breast
HPO:0001626	Abnormality of the cardiovascular system
HPO:0000598	Abnormality of the ear
HPO:0000818	Abnormality of the endocrine system
HPO:0000478	Abnormality of the eye
HPO:0000119	Abnormality of the genitourinary system
HPO:0002715	Abnormality of the immune system
HPO:0001574	Abnormality of the integument
HPO:0003011	Abnormality of the musculature
HPO:0000707	Abnormality of the nervous system
HPO:0002086	Abnormality of the respiratory system
HPO:0000924	Abnormality of the skeletal system
HPO:0001507	Growth abnormality
HPO:0002664	Neoplasm

REPORT

The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease

Peter N. Robinson,^{1,2,*} Sebastian Köhler,^{1,2} Sebastian Bauer,¹ Dominik Seelow,^{1,3} Denise Horn,¹ and Stefan Mundlos^{1,2,4}

There are many thousands of hereditary diseases in humans, each of which has a specific combination of phenotypic features, but computational analysis of phenotypic data has been hampered by lack of adequate computational data structures. Therefore, we have developed a Human Phenotype Ontology (HPO) with over 8000 terms representing individual phenotypic anomalies and have annotated all clinical entries in Online Mendelian Inheritance in Man with the terms of the HPO. We show that the HPO is able to capture phenotypic similarities between diseases in a useful and highly significant fashion.

¹Institute for Medical Genetics, ²Berlin-Brandenburg Center for Regenerative Therapies, ³Department of Neuropaediatrics, Charité-Universitätsmedizin Berlin, 13353 Berlin, Germany; ⁴Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany

*Correspondence: peter.robinson@charite.de

DOI 10.1016/j.ajhg.2008.09.017. ©2008 by The American Society of Human Genetics. All rights reserved.

The American Journal of Human Genetics 83, 610–615, November 7, 2008

Transitorische Genotyp–Phänotyp Datenbank

3 leitende Prinzipien für die Datenerfassung und Datenhaltung

- beobachtungsnahe Daten, d.h. primäre molekulargenetische Befunde, weitestgehender Verzicht auf Daten der Interpretationsebenen
- Beachtung des Sparsamkeitsprinzips (*parsimony principle*), es werden **nur** die Daten erhoben, die für das Match-making **unbedingt** erforderlich sind
- Sonder- und Spezialfälle werden ggf. suboptimal oder gar nicht abgebildet

KISS

Keep it simple, stupid!

... there is no better way to advance the proper practice of medicine than to give our minds to the discovery of the usual laws of nature by careful investigation of cases of rarer forms of diseases.

William Harvey (1578–1657)
in a letter to the physician Jan Vlackfeld
Haarlem, Netherlands, April 24th, 1657

cf. Groft SC and Posada M (2010)
in: Rare Disease Epidemiology, chpt. 1
Adv. Exp. Med. Biol. Vol. 686, pp. 3 –14



About TMF

Our Mission

We bring together researchers in different disciplines and develop concepts, infrastructures, and methods that promote medical research. We thus increase the safety, quality, and efficiency of medical research, ensure the necessary standardization, and offer researchers, irrespective of their scientific issues, assistance in dealing with the increasingly complex legal and regulatory infrastructure of medical research.

Vielen Dank fürs Zuhören!

Fragen werden gerne beantwortet!

Kommentare sind erwünscht!

TMF - Jahrestagung in JENA: Wienker_2014apr03

6. TMF-Jahrestagung
2.-3. April 2014

1

NGS

Next Generation Sequencing

2

NGS: Next Generation Sequencing

- Die genotypische Sequenzierung ermöglicht die schnelle Identifizierung von Mutationen in Genen oder in nicht-codierenden Regionen des Genoms.
- Die Identifizierung von Mutationen in Genen ermöglicht die Identifizierung von Krankheitsursachen.
- Die Identifizierung von Mutationen in Genen ermöglicht die Identifizierung von Krankheitsursachen.

3

Genome Reference Consortium: GRCh38 (24.12.2013)

4

NGS: Next Generation Sequencing

(A) population frequency if frequent (not for rare variants)
(B) conservation score, phylogenetic (orthologous comparison)
(C) prediction of deleteriousness, consequence of amino acid exchange on structure, activity and stability
???
(D) conservation scores, phylogenetic (orthologous comparison)
(E) prediction of deleteriousness, consequence of amino acid exchange on structure, activity and stability

5

NGS: Next Generation Sequencing

(A) population frequency if frequent (not for rare variants)
(B) conservation score, phylogenetic (orthologous comparison)
(C) prediction of deleteriousness, consequence of amino acid exchange on structure, activity and stability
(D) functional tests, in vitro (cell culture) or in vivo (experimental animals)

6

NGS: Next Generation Sequencing

(A) population frequency if frequent (not for rare variants)
(B) conservation score, phylogenetic (orthologous comparison)
(C) prediction of deleteriousness, consequence of amino acid exchange on structure, activity and stability
(D) conservation analysis
i.e. testing both parents and child (affected and unaffected, if existing and coexisting)

7

Phylo: HBB, HBA, HBD, HBE, HBI, HBJ, HBL, HBM, HBN, HBP, HBR, HBS, HBT, HBU, HBV, HBY, HBZ, HCA, HCB, HCC, HCD, HCE, HCF, HCG, HCH, HCI, HCL, HCM, HCN, HCO, HCP, HCR, HCS, HCT, HCU, HCV, HCY, HCA, HCB, HCC, HCD, HCE, HCF, HCG, HCH, HCI, HCL, HCM, HCN, HCO, HCP, HCR, HCS, HCT, HCU, HCV, HCY

8

NGS: Next Generation Sequencing

There is no genetic diagnosis (sensu stricto) outside the framework of the Mendelian paradigm.

© Genetik, TMF-School 2014, Berlin, 24. bis 26. September 2014

9

Phylo: HBB, HBA, HBD, HBE, HBI, HBJ, HBL, HBM, HBN, HBP, HBR, HBS, HBT, HBU, HBV, HBY, HBZ, HCA, HCB, HCC, HCD, HCE, HCF, HCG, HCH, HCI, HCL, HCM, HCN, HCO, HCP, HCR, HCS, HCT, HCU, HCV, HCY

10

NGS: Next Generation Sequencing

(A) population frequency if frequent (not for rare variants)
(B) conservation score, phylogenetic (orthologous comparison)
(C) prediction of deleteriousness, consequence of amino acid exchange on structure, activity and stability
(D) conservation scores, phylogenetic (orthologous comparison)
(E) prediction of deleteriousness, consequence of amino acid exchange on structure, activity and stability

11

NGS: Next Generation Sequencing

Nachdem alle Schritte der Planung und Bewertung der Genetik (Genomik) abgeschlossen sind, ist die Bewertung der Ergebnisse zu erfolgen. Diese Bewertung ist in der Regel ein gemeinsames Ziel der beteiligten Akteure.

12

NGS: Next Generation Sequencing

Referenzdatenbanken (variation resources) allow to compare the results of NGS with known variations in the human genome. These variations are organized in different ways, such as by gene, by chromosome, or by population.

13

NGS: Next Generation Sequencing

Referenzdatenbanken (variation resources) allow to compare the results of NGS with known variations in the human genome. These variations are organized in different ways, such as by gene, by chromosome, or by population.

14

genomeweb

DeepTie Handles NGS is Changing Clinical Practice for Rare Disease, Cancer

15

Transcriptional Genotyp-Phänotyp Datenbank

1. Eingrenzung & Beschreibung des Problems
2. Mittel & Wege zur Problemlösung

16

Transcriptional Genotyp-Phänotyp Datenbank

Wie ist eine "diagnostische Situation"?

Genotyp: $D = G$
Phänotyp: $G = D$

17

Transcriptional Genotyp-Phänotyp Datenbank

Die "diagnostische Situation" ist eine Situation, in der ein Patient mit einer bestimmten Genotyp-Phänotyp-Beziehung konfrontiert wird.

18

Transcriptional Genotyp-Phänotyp Datenbank

Wachstumsstörungen

19

Transcriptional Genotyp-Phänotyp Datenbank

Typische Szenarien für eine Anamnese

20

Transcriptional Genotyp-Phänotyp Datenbank

Mitochondriale Daten für die typischen Szenarien

21

Human Phenotype Ontology (HPO)

Standardized terminology for describing human phenotypes

22

REPORT

The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Phenotypic Data

23

Transcriptional Genotyp-Phänotyp Datenbank

Zentrale Ressourcen für die Diagnosefindung und Datenanalyse

24

KISS

Keep it simple, stupid!

25

... there is no better way to advance the progress of medicine than to give our minds to the discovery of the usual laws of nature by careful investigation of cases of rare forms of disease.

26

About TMF

27

Vielen Dank fürs Zuhören!
Fragen werden gerne beantwortet!
Kommentare sind erwünscht!

28

29

TMF - Jahrestagung in JENA, Weicker 2014apr03

30