Face2Gene
Deep learning on imaging data

Prof. Dr. Peter Krawitz, Dipl. Phys.
Director of the Institute for
Genome Statistics and Bioinformatics
University Bonn, Germany

Fachärzte für:

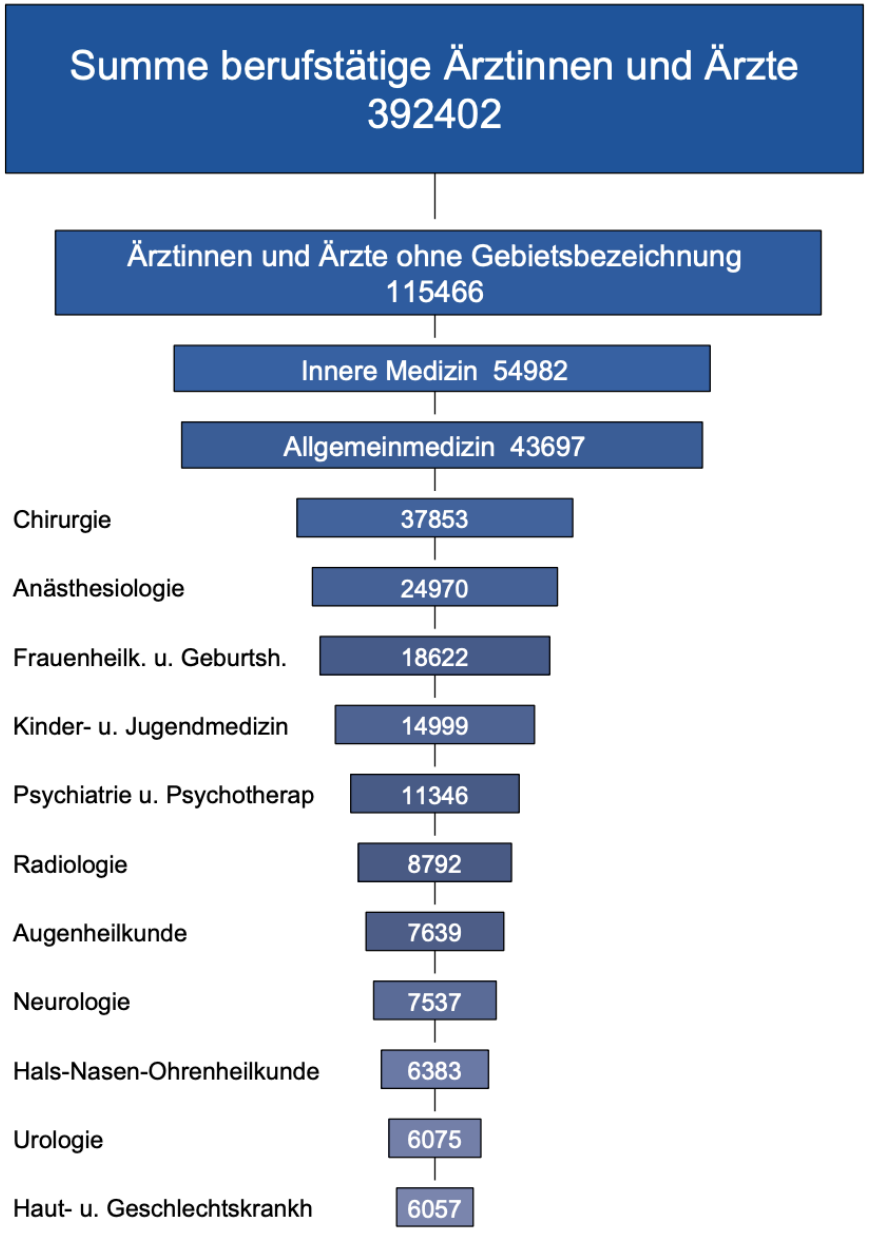
1) Allgemeinmedizin?

2) Pädiatrie?

3) Humangenetik?

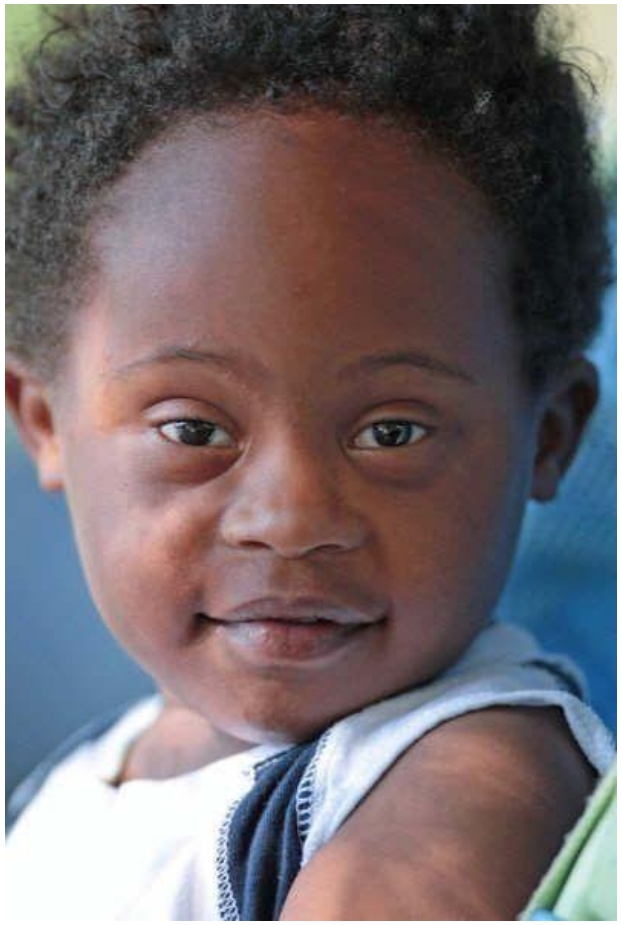
Fachärzte für:

1) Allgemeinmedizin?
2) Pädiatrie?
3) Humangenetik?

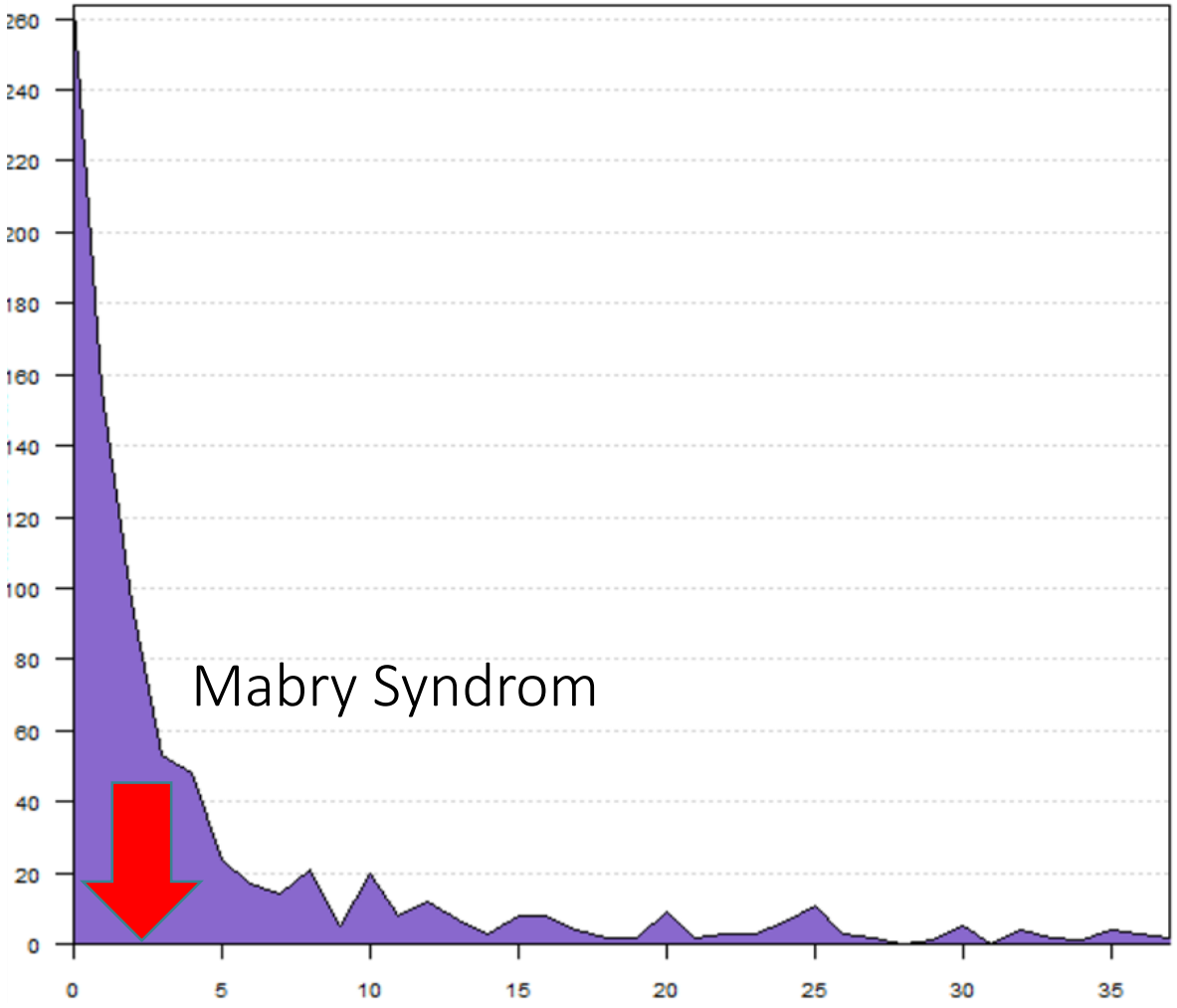


300 Humangenetiker

Haben Sie eine Verdachtsdiagnose?



Prävalenz seltener Erkrankungen



Patienten pro 100.000 Individuen

(...)

Down Syndrom

60

100

Diagnosequote bei syndromalen Erkrankungen?

… bis zu 60% bei Einsatz
moderner Sequenziertechnologie

NovaSeq von Illumina

Brauchen wir aber immer ein Genom und einen Humangenetiker oder kann die geeignete Diagnostik auch durch andere Fachkollegen veranlasst werden?

~~Freigabe 24.05.2011~~

Krankenkasse bzw. Kostenträger

Name, Vorname des Versicherten

geb. am

Kassen-Nr. | Versicherten-Nr. | Status

Betriebsstätten-Nr. | Arzt-Nr. | Datum

Eintrag nur bei Weiterüberweisung!
Betriebsstätten-Nr. des Erstveranlassers | Arzt-Nr. des Erstveranlassers

☐ **Befundübermittlung eilt**, nachrichtlich an | Telefon Nr. | Fax Nr.

Überweisungsschein für Laboratoriums-untersuchungen als Auftragsleistung

10

☐ Kurativ ☐ Präventiv ☐ bei belegärztl. Behandlung ☐ Unfall, Unfallfolgen

Auftragsnummer des Labors

Hier bitte sorgfältig Barcode-Etikett einkleben!

ggf. Kennziffer

Quartal: Q J J

Geschlecht: W M

☐ Kontrolluntersuchung bekannte Infektion

Behandlung gemäß ☐ § 116b SGB V

eingeschränkter Leistungsanspruch ☐ gemäß § 16 Abs. 3a SGB V

Abnahmedatum: T T M M J J

Abnahmezeit: h h m m

☐ Empfängnisregelung, Sterilisation, Schwangerschaftsabbruch

Diagnose/Verdachtsdiagnose

Syndromale, globale Entwicklungsverzögerung

Befund/Medikation

Auftrag

Gen-Panel

Verbindliches Muster

Vertragsarztstempel / Unterschrift überw. Arzt

Nicht zu verwenden bei Arbeitsunfällen, Berufskrankheiten und Schülerunfällen

Muster 10 (1.2012)

Ca. 10-20 Gene
darf jeder Arzt veranlassen

11513 Postnatale Mutationssuche zum Nachweis oder Ausschluss einer krankheitsrelevanten oder krankheitsauslösenden konstitutionellen genomischen Mutation in bis zu 25 Kilobasen kodierender Sequenz einschließlich zugehöriger regulatorischer Sequenzen

Beschreibung

Postnatale Mutationssuche zum Nachweis oder Ausschluss einer krankheitsrelevanten oder krankheitsauslösenden konstitutionellen genomischen Mutation in bis zu 25 Kilobasen kodierender Sequenz einschließlich zugehöriger regulatorischer Sequenzen

Obligater Leistungsinhalt

- Vollständige Sequenzanalyse,
- Bioinformatische Auswertung der erhobenen Sequenzdaten,

Fakultativer Leistungsinhalt

- Untersuchung nicht-kodierender genetischer Elemente,
- Nach- und/oder Bestätigungsdiagnostik zur analytischen Validierung mittels weiterer Verfahren,

Abrechnungsbestimmung

je vollendete 250 kodierende Basen

Anmerkung

Ab der 21. Leistung im Krankheitsfall wird die Gebührenordnungsposition 11513 mit 271 Punkten je vollendeten 250 kodierenden Basen bewertet.
Der Höchstwert für die Untersuchungen der Gebührenordnungsposition 11513 beträgt 24.914 Punkte im Krankheitsfall.
Der Leistungsinhalt ist durch den Umfang der für die Fragestellung auszuwertenden kodierenden Sequenzlänge bestimmt, nicht durch die Sequenzlänge der Rohdaten.

Abrechnungsausschlüsse

	Leistungen	Kapitel
im Krankheitsfall	01793, 11514	

Berichtspflicht

Ja

Ausschluss der Berechnungsfähigkeit der Pauschale für die fachärztliche Grundversorgung

Ja

Gesamt (Punkte)	**542**
Gesamt (Euro)	**58,66**

Exom = 20.000 Gene
Indikationsstellung an Zentren für seltene Erkrankungen

11514 Genehmigungspflichtige postnatale Mutationssuche zum Nachweis od. Ausschluss einer krankheitsrelevanten od. krankheitsauslösenden konstitutionellen genomischen Mutation in mehr als 25 kb kodierender Sequenz einschl. zugehöriger regulatorischer Sequenzen

Beschreibung

Genehmigungspflichtige postnatale Mutationssuche zum Nachweis oder Ausschluss einer krankheitsrelevanten oder krankheitsauslösenden konstitutionellen genomischen Mutation in mehr als 25 Kilobasen kodierender Sequenz einschließlich zugehöriger regulatorischer Sequenzen

Obligater Leistungsinhalt

- Vollständige Sequenzanalyse,
- Bioinformatische Auswertung der erhobenen Sequenzdaten,

Fakultativer Leistungsinhalt

- Untersuchung nicht-kodierender genetischer Elemente,
- Nach- und/oder Bestätigungsdiagnostik mittels weiterer Verfahren,

Abrechnungsbestimmung

einmal im Krankheitsfall

Anmerkung

Die Gebührenordnungsposition 11514 ist nur berechnungsfähig, wenn eine ausführliche Begründung der medizinischen Notwendigkeit im Einzelfall sowie eine vorherige Genehmigung durch die zuständige Krankenkasse vorliegen.

Abrechnungsausschlüsse

	Leistungen	Kapitel
im Krankheitsfall	01793, 11304, 11513	

Berichtspflicht

Ja

Ausschluss der Berechnungsfähigkeit der Pauschale für die fachärztliche Grundversorgung

Ja

Gesamt (Punkte)	**30663**
Gesamt (Euro)	**3.318,53**

Auch bei sehr seltenen Syndromen
Kann die bildgestützte Differentialdiagnostik
ähnlich gut funktionieren wie bei häufigeren

SUGGESTED SYNDROMES (30)

Down Syndrome
HIGH MED LOW
Differential
Clinically Diagnosed
Molecularly Diagnosed
GESTALT FEATURE

Fragile X Syndrome
HIGH MED LOW
Differential
Clinically Diagnosed
Molecularly Diagnosed
GESTALT FEATURE

Hyperphosphatasia ...Tardation Syndrome
HIGH MED LOW
Differential
Clinically Diagnosed
Molecularly Diagnosed
GESTALT FEATURE

CHARGE Syndrome
HIGH MED LOW
Differential
Clinically Diagnosed
Molecularly Diagnosed
GESTALT FEATURE

Turner Syndrome
HIGH MED LOW
Differential
Clinically Diagnosed
Molecularly Diagnosed
GESTALT FEATURE

Neurofibromatosis, Type I; NF1
HIGH MED LOW
Differential
Clinically Diagnosed
Molecularly Diagnosed
GESTALT FEATURE

Lathosterolosis
HIGH MED LOW
Differential
Clinically Diagnosed
Molecularly Diagnosed
GESTALT FEATURE

Craniodiaphyseal Dysplasia; CDD
HIGH MED LOW
Differential
Clinically Diagnosed
Molecularly Diagnosed
GESTALT FEATURE

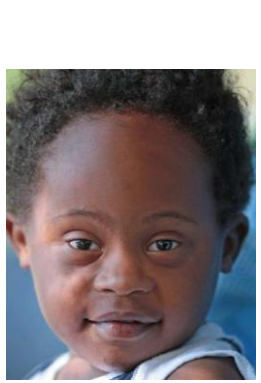
Prader-Willi Syndrome; PWS
HIGH MED LOW
Differential
Clinically Diagnosed
Molecularly Diagnosed
GESTALT FEATURE

Saethre-Chotzen Syndrome; SCS
HIGH MED LOW
Differential
Clinically Diagnosed
Molecularly Diagnosed
GESTALT FEATURE

Robinow Syndrome
HIGH MED LOW
Differential
Clinically Diagnosed
Molecularly Diagnosed
GESTALT FEATURE

Baraitser-Winter Syndrome
HIGH MED LOW
Differential
Clinically Diagnosed
Molecularly Diagnosed
GESTALT FEATURE

Wie könnte es funktionieren?





V.a. Down Syndrom
=> Chromosomenanalyse

V.a. Mabry Syndrom
⇒ GPI Anker Gen-Panel:
⇒ PIGV, PGAP3, etc.

Ein Humangenetiker hilft 100 Kollegen bei der Auswahl des geeigneten Tests, z.B. des 25 kb Gen-Panels

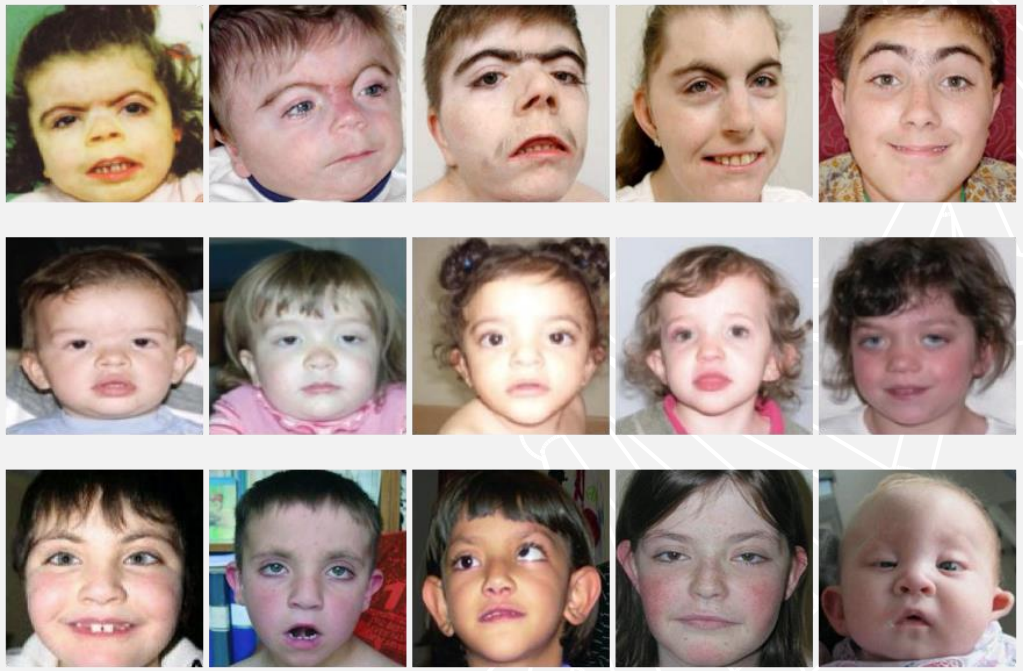
Technische Herausforderung bei der Bildanalyse

- Relativ kleine Datensätze
- Große Variabilität
- Unbekannte Anzahl an Erkrankungen

Types of Face Recognition

Intra-Person

Intra-Syndrome

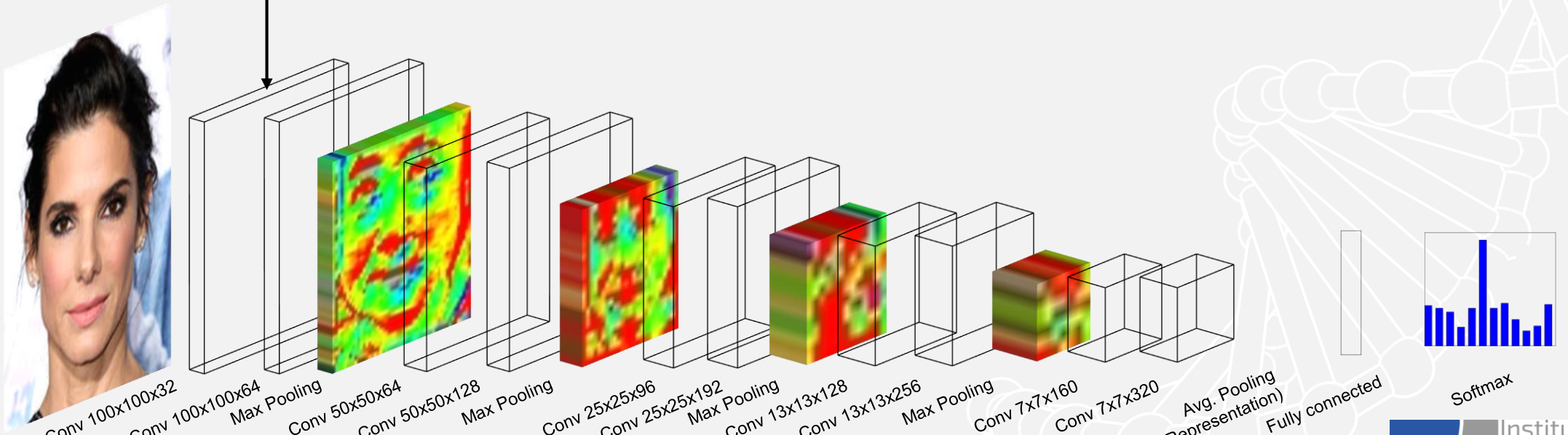


Institut für
Genomische Statistik
und Bioinformatik

Knowledge transfer:
Initiales Training auf Porträts der Allgemeinbevölkerung

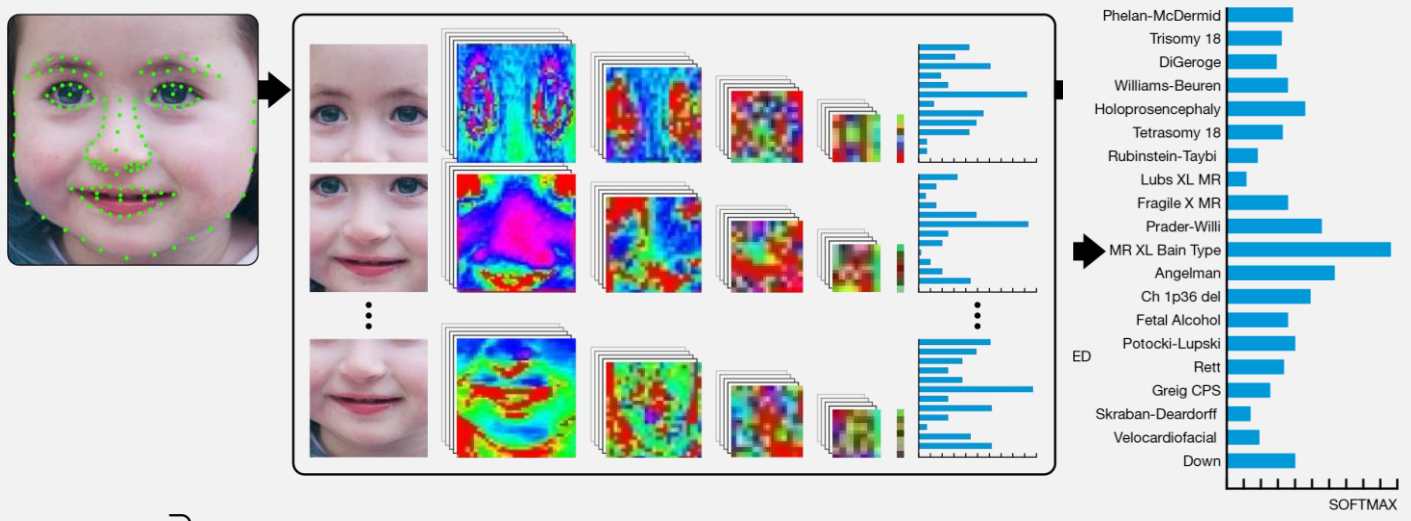
Face Recognition Tasks

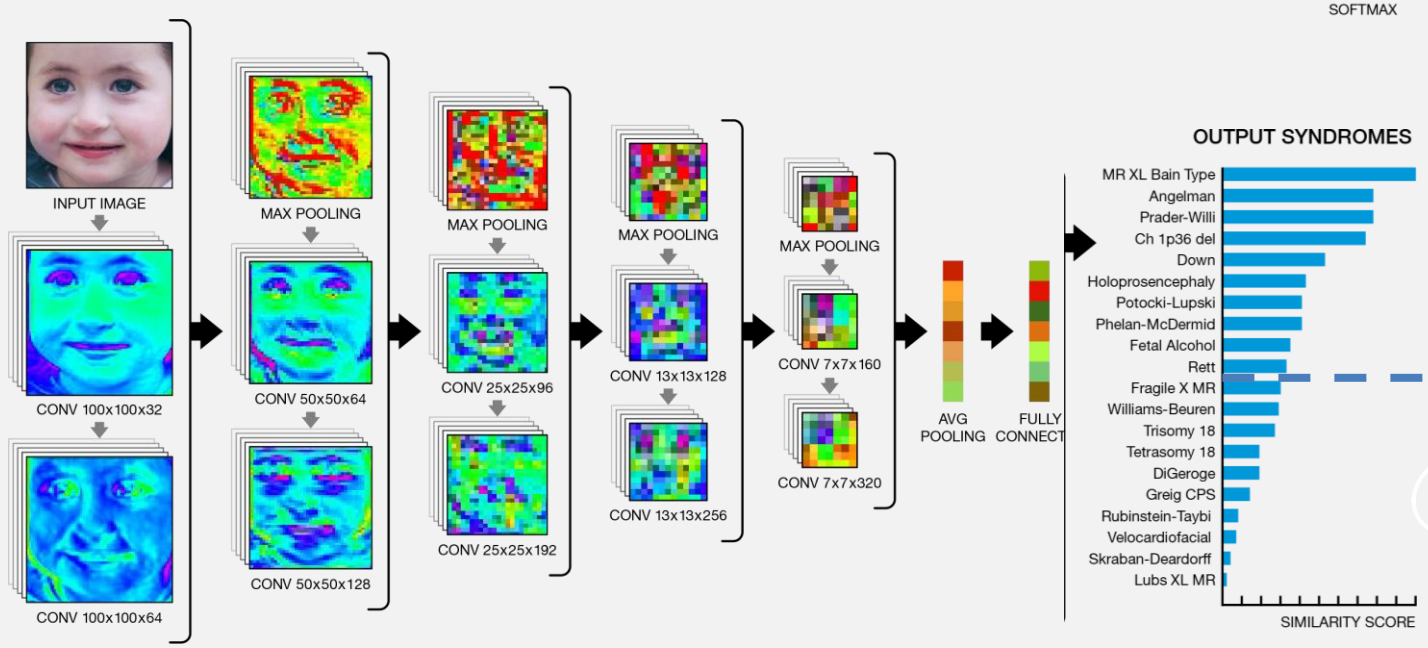
Celebrities Database



Institut für Genomische Statistik und Bioinformatik

DeepGestalt





LETTERS | FOCUS
https://doi.org/10.1038/s41591-018-0279-0
nature medicine

Identifying facial phenotypes of genetic disorders using deep learning

Yaron Gurovich[1*], Yair Hanani[1], Omri Bar[1], Guy Nadav[1], Nicole Fleischer[1], Dekel Gelbman[1], Lina Basel-Salmon[2,3], Peter M. Krawitz[4], Susanne B. Kamphausen[5], Martin Zenker[5], Lynne M. Bird[6,7] and Karen W. Gripp[8]

>90%
TOP-10 Accuracy*

Institut für Genomische Statistik und Bioinformatik

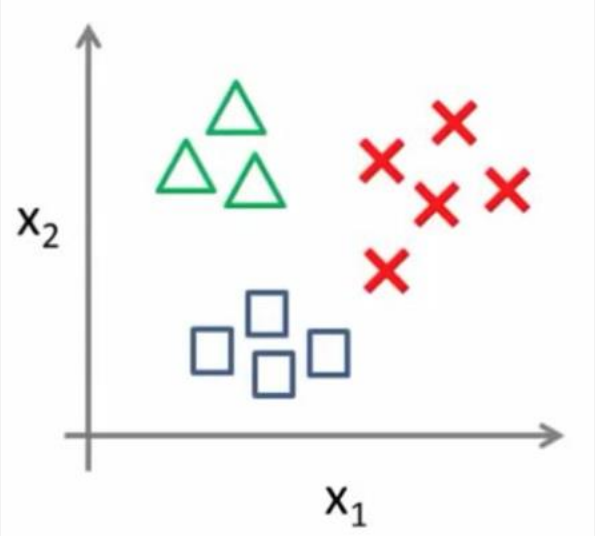
FACE2GENE
Smart Phenotyping. Better Genetics.

Live Demonstration

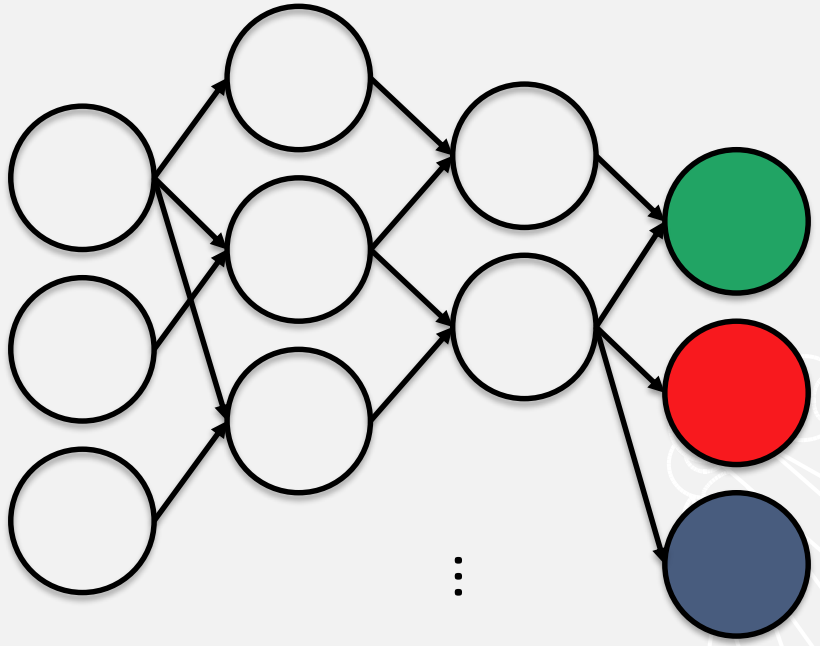
FDNA

Multiclass Classification Problems

Problem



Deep Convolutional Neural Network

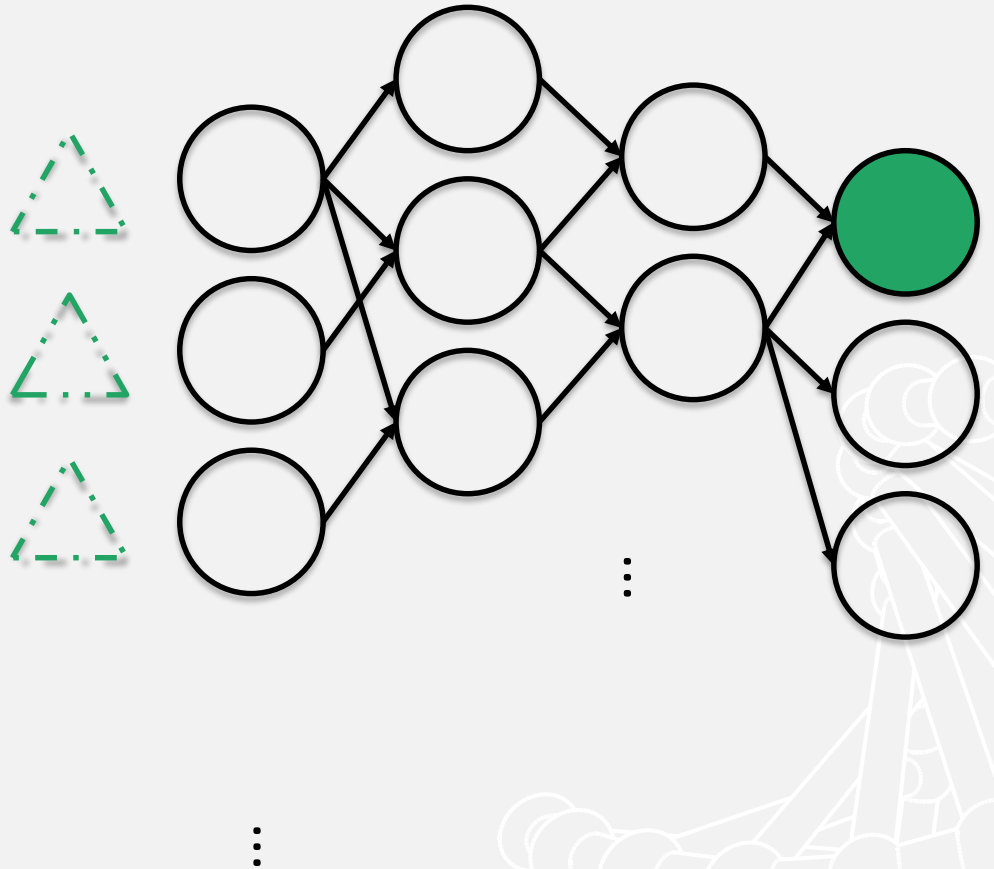


Input - hidden layers - output

Accuracy

$$ACC = \frac{TP + TN}{TP+FP+FN+TN}$$

Deep Convolutional Neural Network

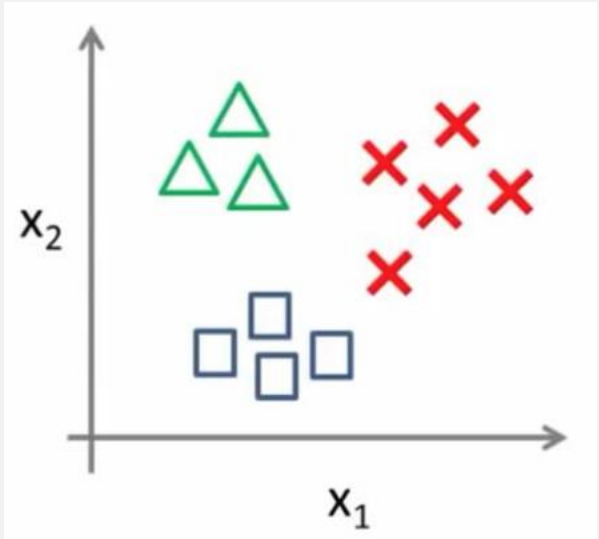


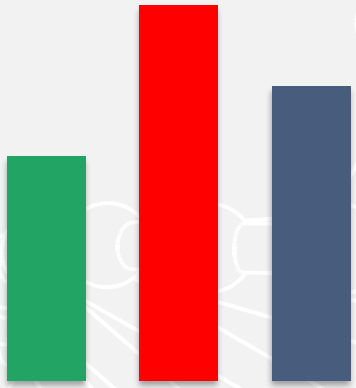
Input - hidden layers - output

Institut für Genomische Statistik und Bioinformatik

Accuracy

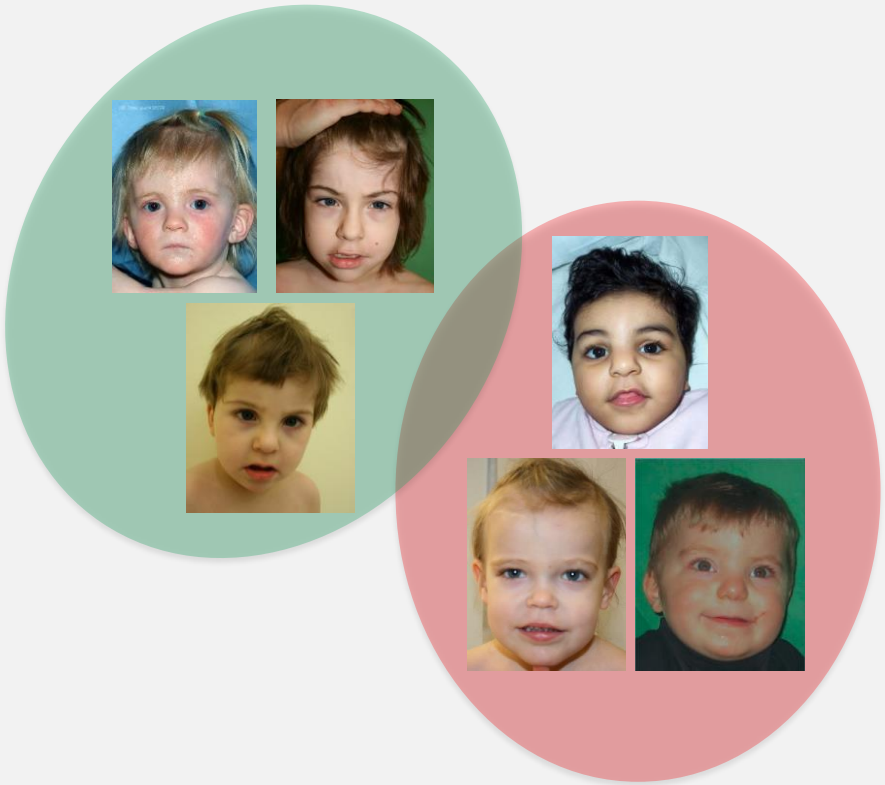
$$ACC = \frac{2 + 4 + 3}{12}$$

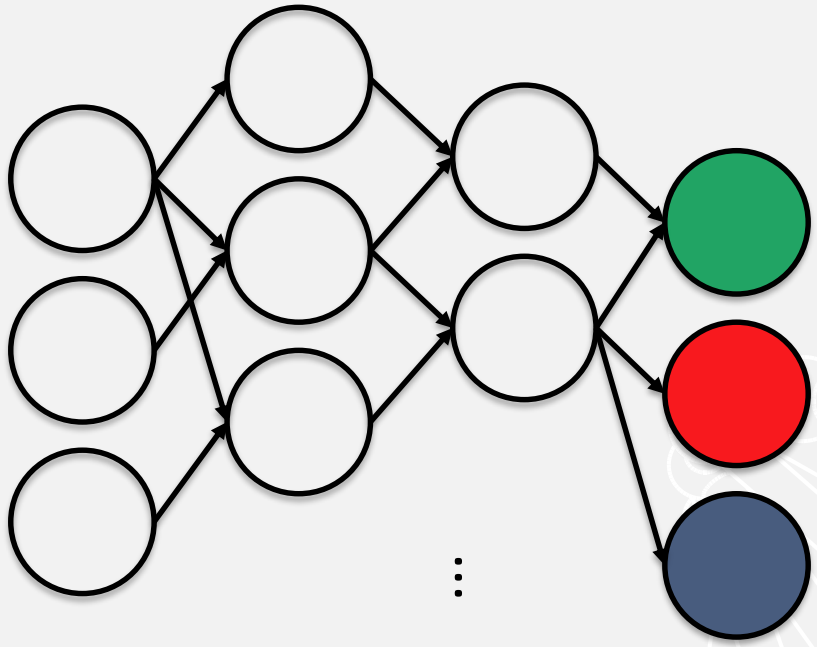




Multiclass Classification Problem: DeepGestalt

Clinical Face Phenotype Space (CFPS)





Syndrome

Coffin Siris

Mabry

Noonan

Input - hidden layers - output

Institut für Genomische Statistik und Bioinformatik

First tool for real world decision support in syndromology

	Number of Genetic Disorders	Number of Training Samples (Syndromic)	Evaluation Method	Accuracy (top-1-accuracy)
Problem 1: Single syndrome vs. other population				
Saraydemir et al. [20]	1	15	3,4-Fold Cross-Validation	97.34%
Burccin et al. [21]	1	10	51 images in a test set	95.30%
Zhao et al. [22]	1	50	Leave-One-Out	96.70%
Basel-Vanagaite et al. [4]	1	134	7 images in test set	94%
Kruszka et al. [23]	1	129	Leave-One-Out	94.30%
Kruszka et al. [24]	1	156	Leave-One-Out	94.90%
Liehr et al. [26]	2	173	10-Fold Cross-Validation	100%
Shukla et al. [19]	6	1126	5-Fold Cross-Validation	94.93 (mAP)[1]
Ferry et al. [3]	8	1363	Leave-One-Out	94.90%
Problem 2: Syndromic vs. normal				
Zhao et al. [22]	14	24	Leave-One-Out	97%
Cerrolaza et al. [27]	15	73	Leave-One-Out	95%
Shukla et al. [19]	6	1126	5-Fold Cross-Validation	98.80%
Problem 3: Multiple syndromes classification				
Loos et al. [29]	5	55	Leave-One-Out	76%
Kuru et al. [28]	15	92	Leave-One-Out	53%
Boehringer et al. [31]	10	147	10-Fold Cross-Validation	75.70%
Boehringer et al. [30]	14	202	91 images in a test set	21%
Ferry et al. [3]	8	1363	Leave-One-Out	75.60%[2]
Shukla et al. [19]	6	1126	5-Fold Cross-Validation	48%[2]

Gurovich, et al.
Nature Medicine
NMED-NT89677D

DeepGestalt. 287 30,000 10-Fold Cross-Validation 90 % top-10-accuracy

FDNA

Exome as gold standard also for pt. with dysmorphism

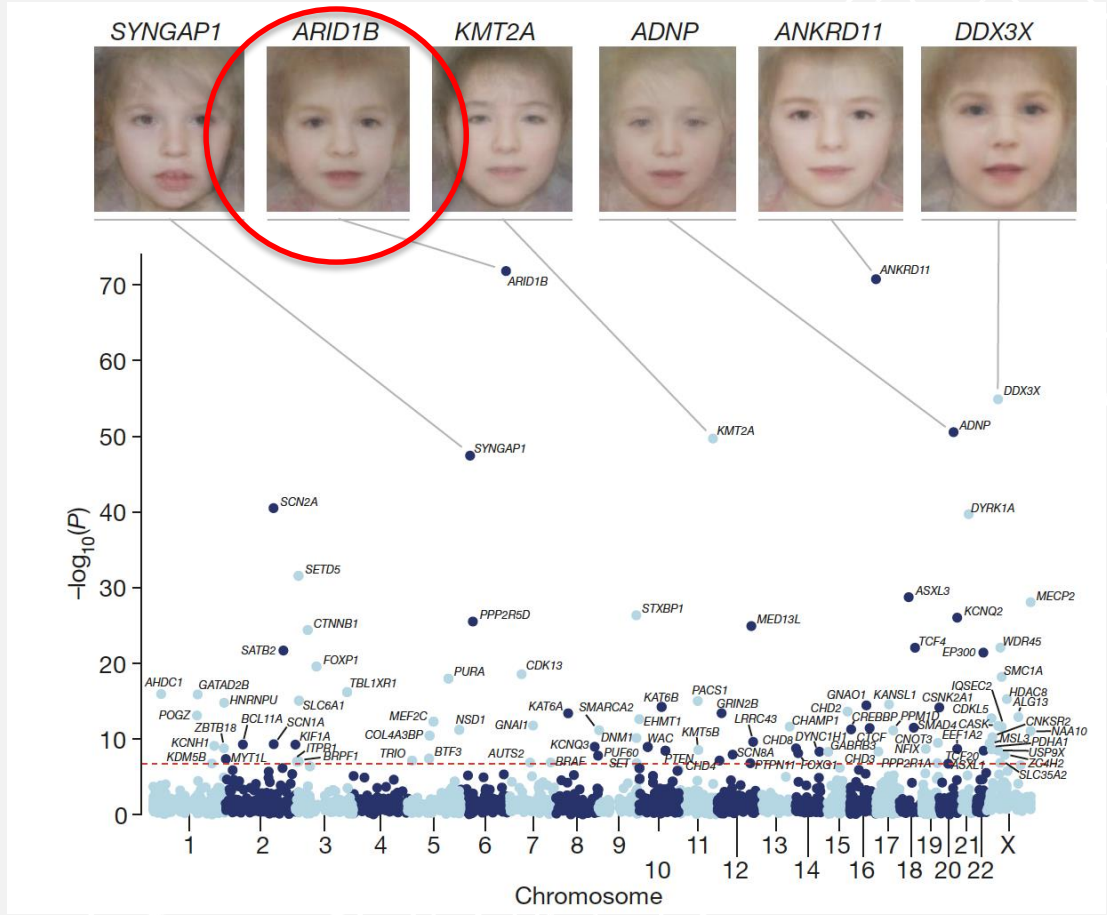
ARTICLE

doi:10.1038/nature21062

Prevalence and architecture of *de novo* mutations in developmental disorders

Deciphering Developmental Disorders Study

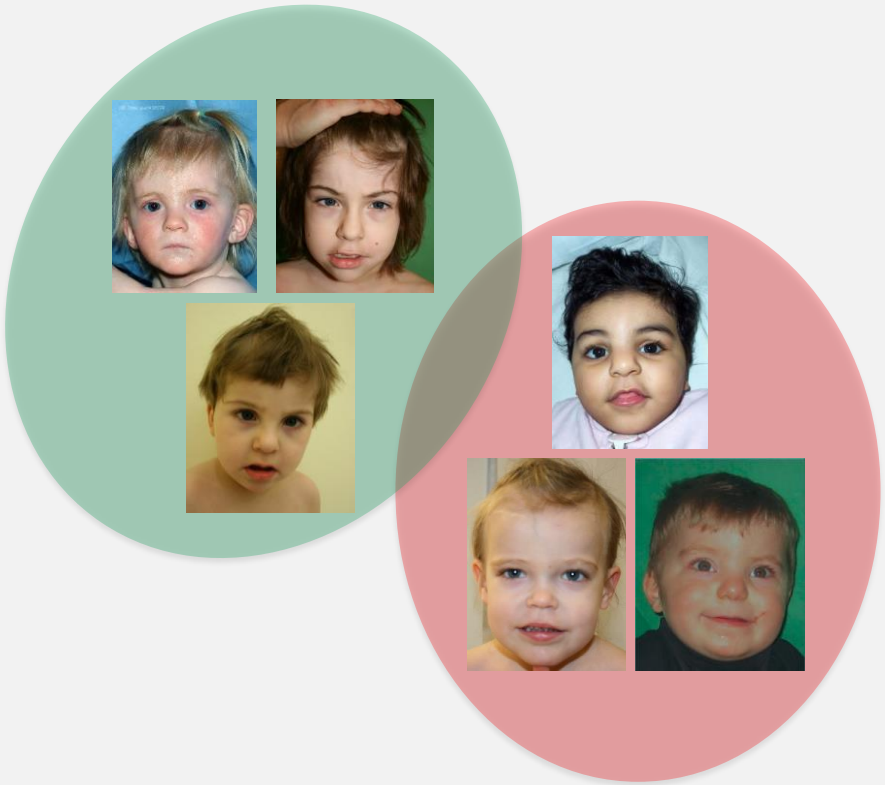
The genomes of individuals with severe, undiagnosed developmental disorders are enriched in damaging *de novo* mutations (DNMs) in developmentally important genes. Here we have sequenced the exomes of 4,293 families containing individuals with developmental disorders, and meta-analysed these data with data from another 3,287 individuals with similar disorders. We show that the most important factors influencing the diagnostic yield of DNMs are the sex of the affected individual, the relatedness of their parents, whether close relatives are affected and the parental ages. We identified 94 genes enriched in damaging DNMs, including 14 that previously lacked compelling evidence of involvement in developmental disorders. We have also characterized the phenotypic diversity among these disorders. We estimate that 42% of our cohort carry pathogenic DNMs in coding sequences; approximately half of these DNMs disrupt gene function and the remainder result in altered protein function. We estimate that developmental disorders caused by DNMs have an average prevalence of 1 in 213 to 1 in 448 births, depending on parental age. Given current global demographics, this equates to almost 400,000 children born per year.

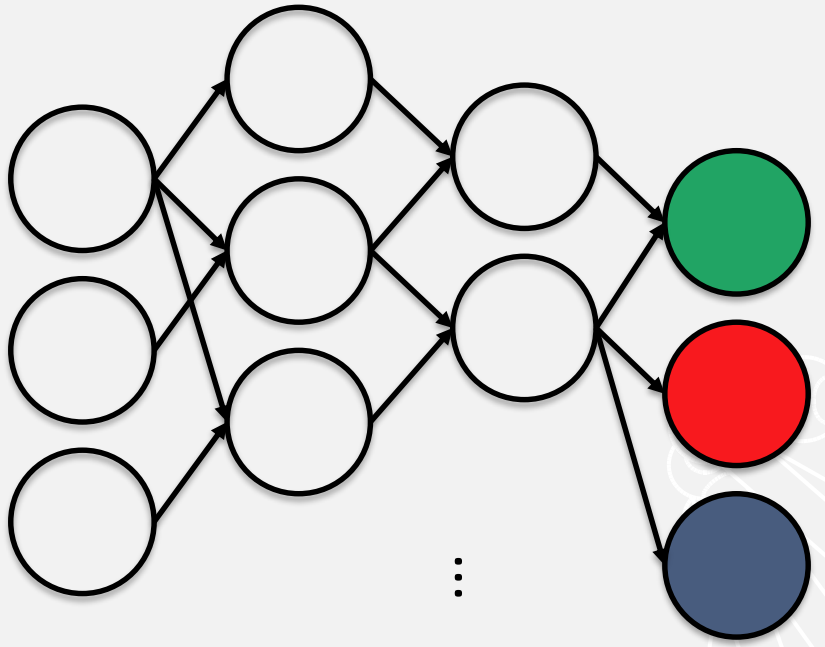


Institut für Genomische Statistik und Bioinformatik

Multiclass Classification Problem: DeepGestalt

Clinical Face Phenotype Space (CFPS)





Syndrome

Coffin Siris

Mabry

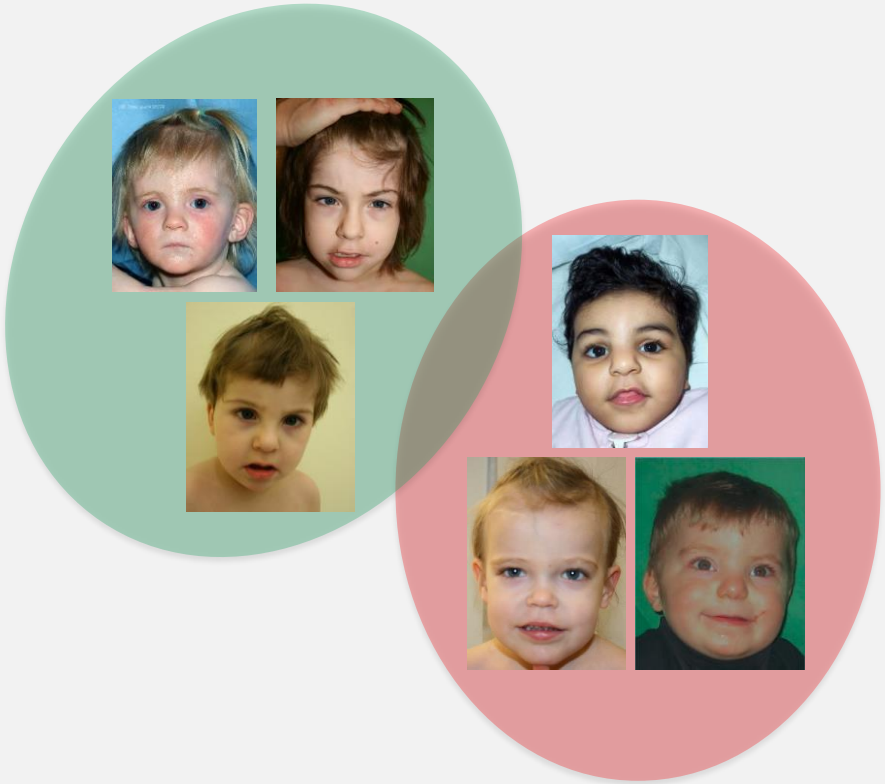
Noonan

Input - hidden layers - output

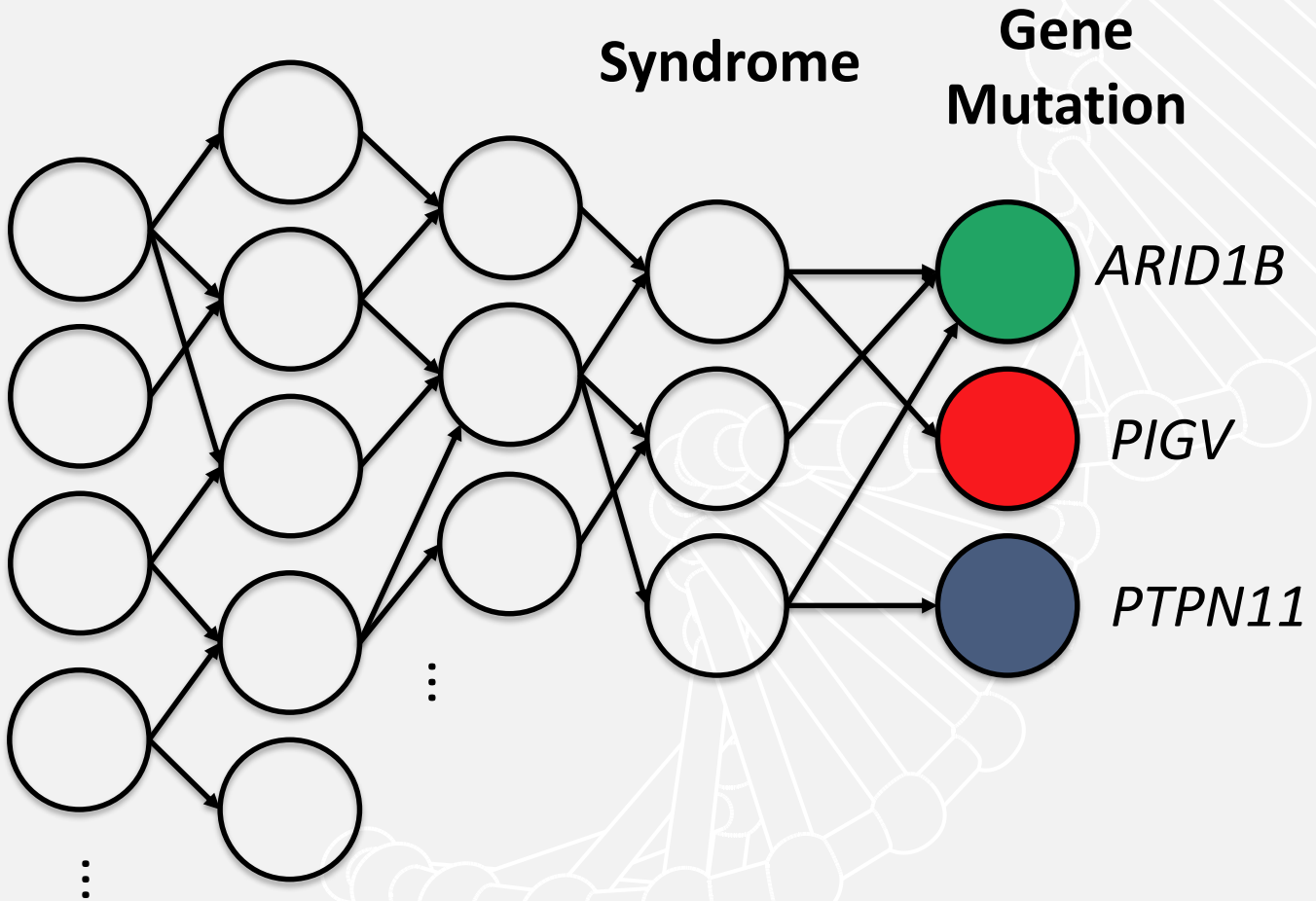
Institut für Genomische Statistik und Bioinformatik

Multiclass Classification Problem: PEDIA

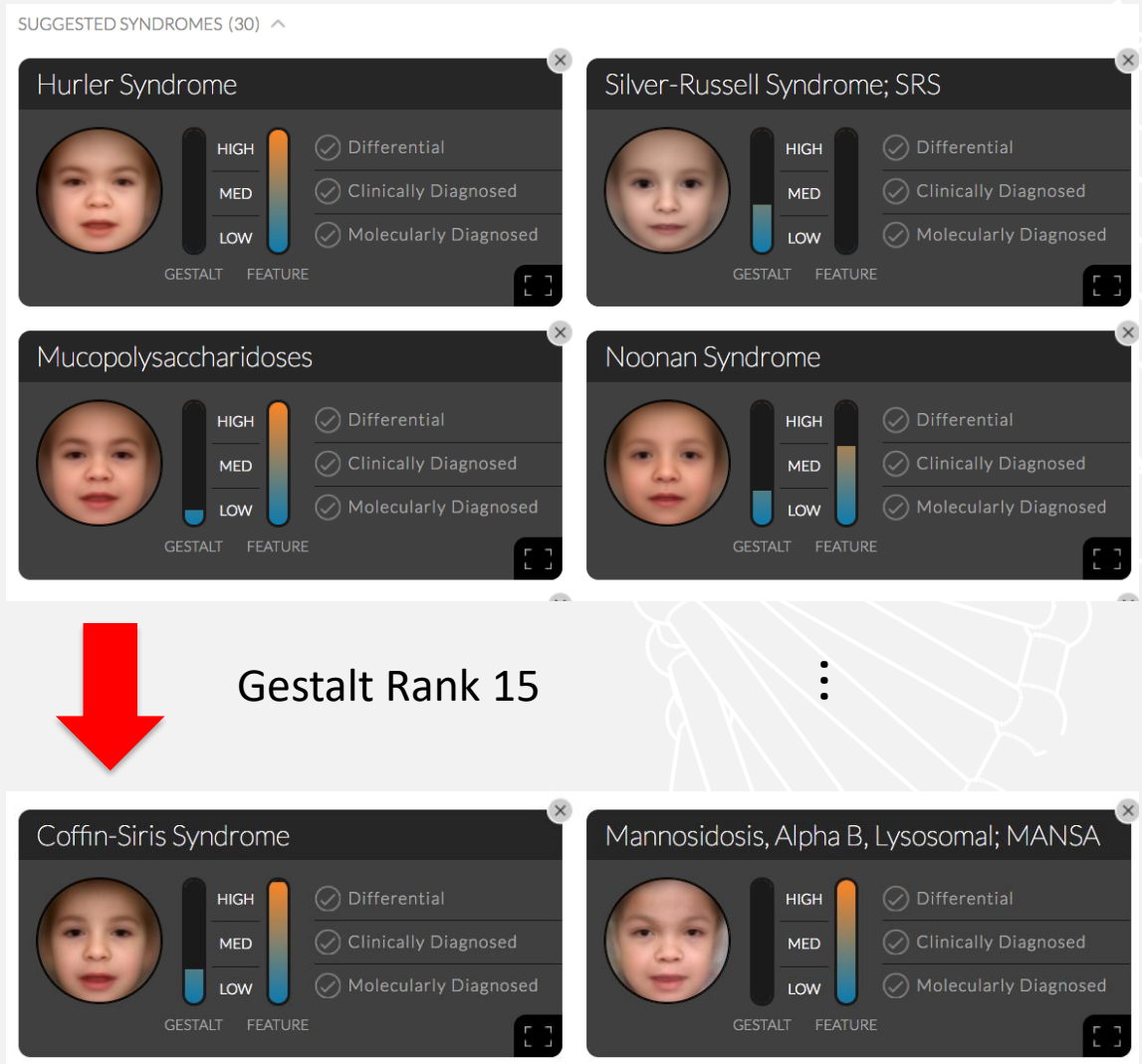
Clinical Face Phenotype Space (CFPS)



Photos + Features + Exome



A case from DeepGestalt/PEDIA cohort with Coffin Siris

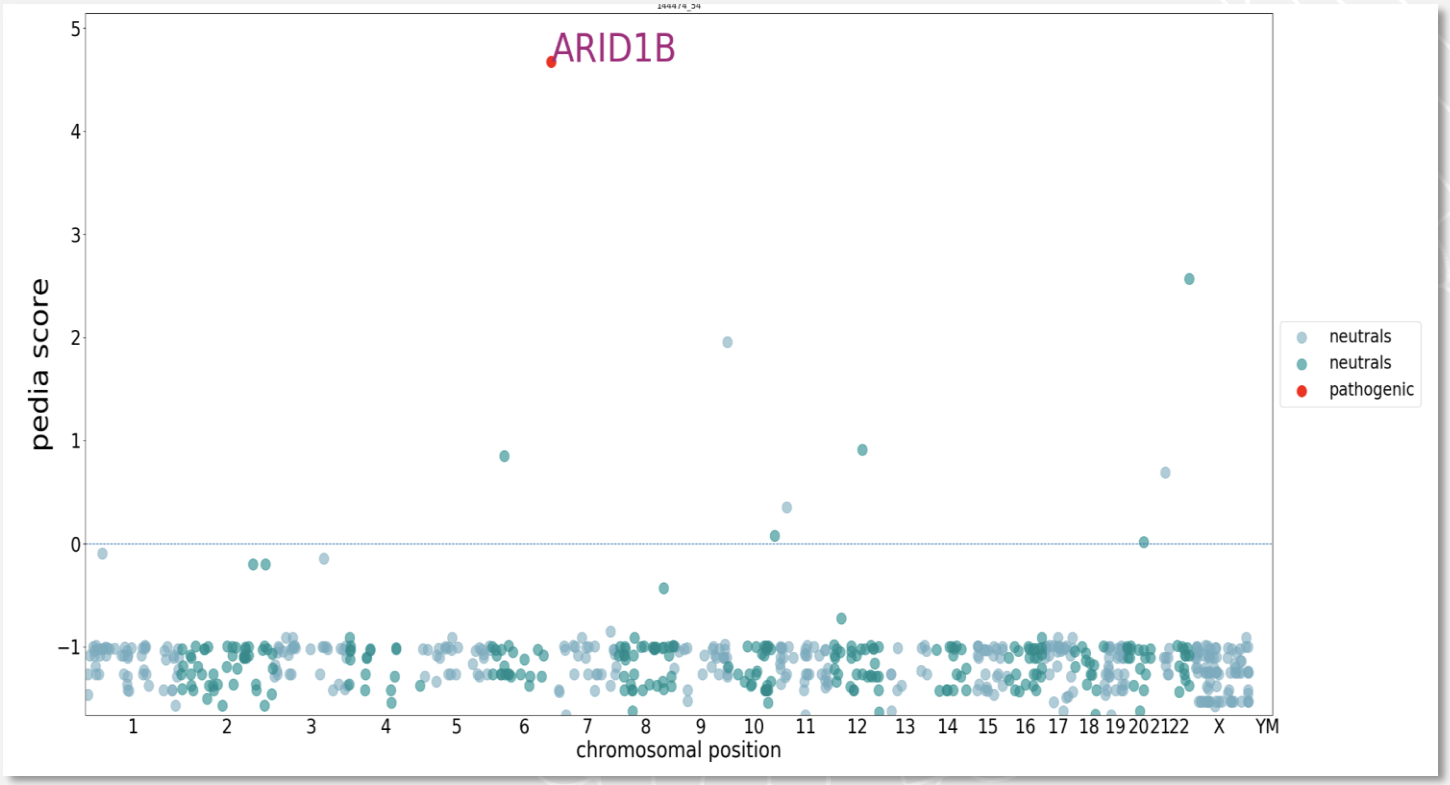


A case from DeepGestalt/PEDIA cohort with Coffin Siris

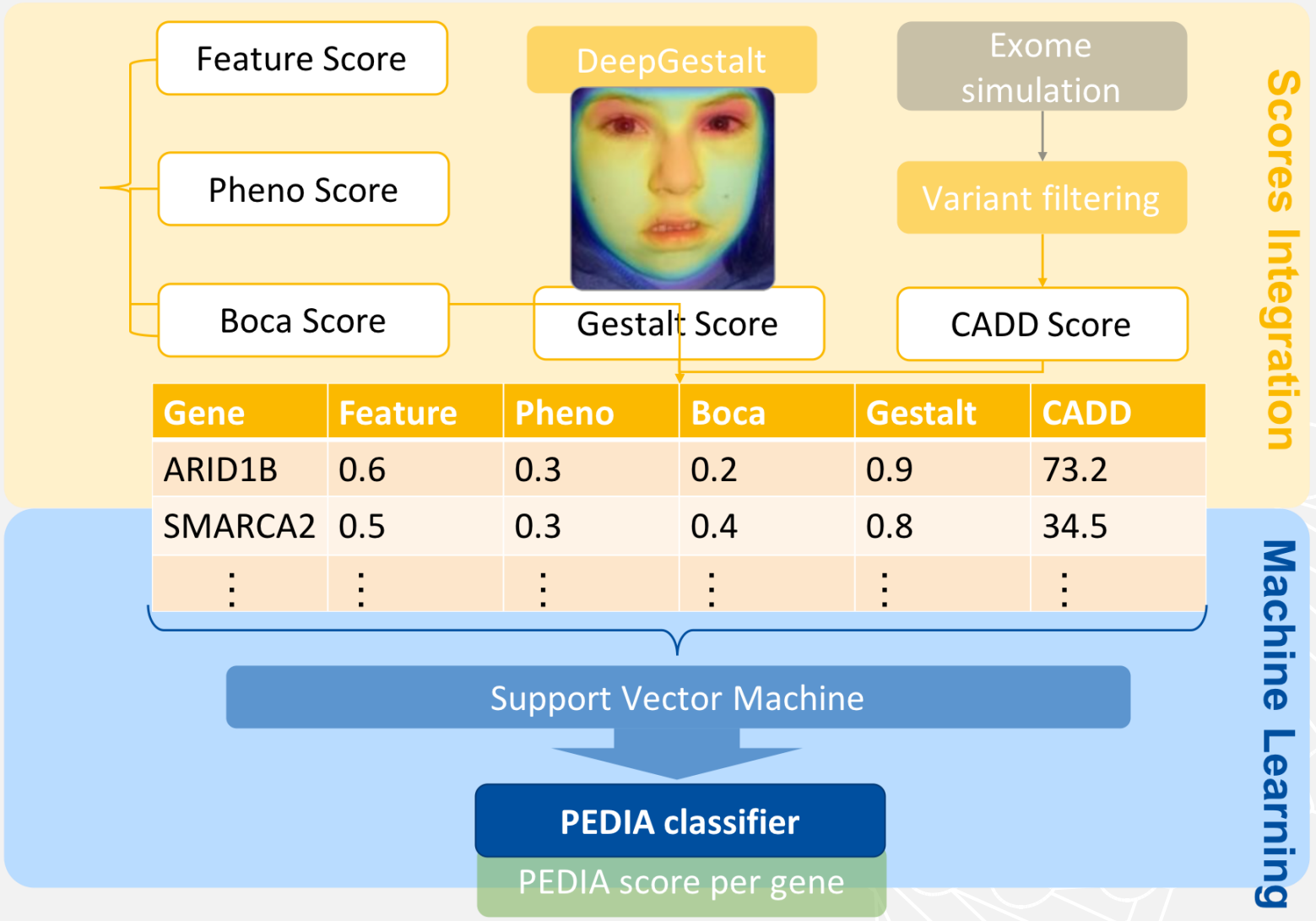
With phenotypic and Molecular information combined (PEDIA), the disease causing mutation NM_020732.3:c.2228C>T, p.Pro743Leu in *ARID1B* is readily identified

PEDIA result

Rank	Gene	PEDIA Score
1	ARID1B	4.67418
2	SHANK3	2.56548
3	EHMT1	1.95629
4	OTOGL	0.913388
5	COL11A2	0.849759
6	CLDN14	0.689384
7	USH1C	0.354995
8	FGFR2	0.0784148
9	SLC12A5	0.014466
10	SZT2	-0.0954363



Prioritization of Exome Data by Image Analysis: PEDIA workflow



Prioritization of Exome Data by Image Analysis: PEDIA cohort

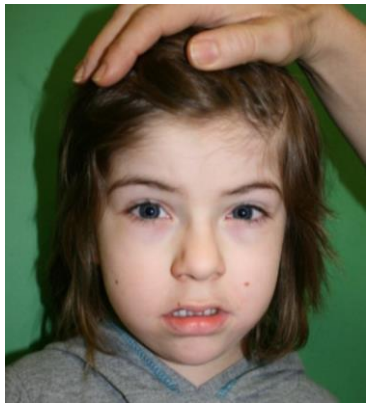
10-fold cross validation split

n>1000

Patient Card

Facial hypertrichosis,
Muscular hypotonia,
Thick lower lip vermilion,
Thick eyebrow,
…

Symptoms

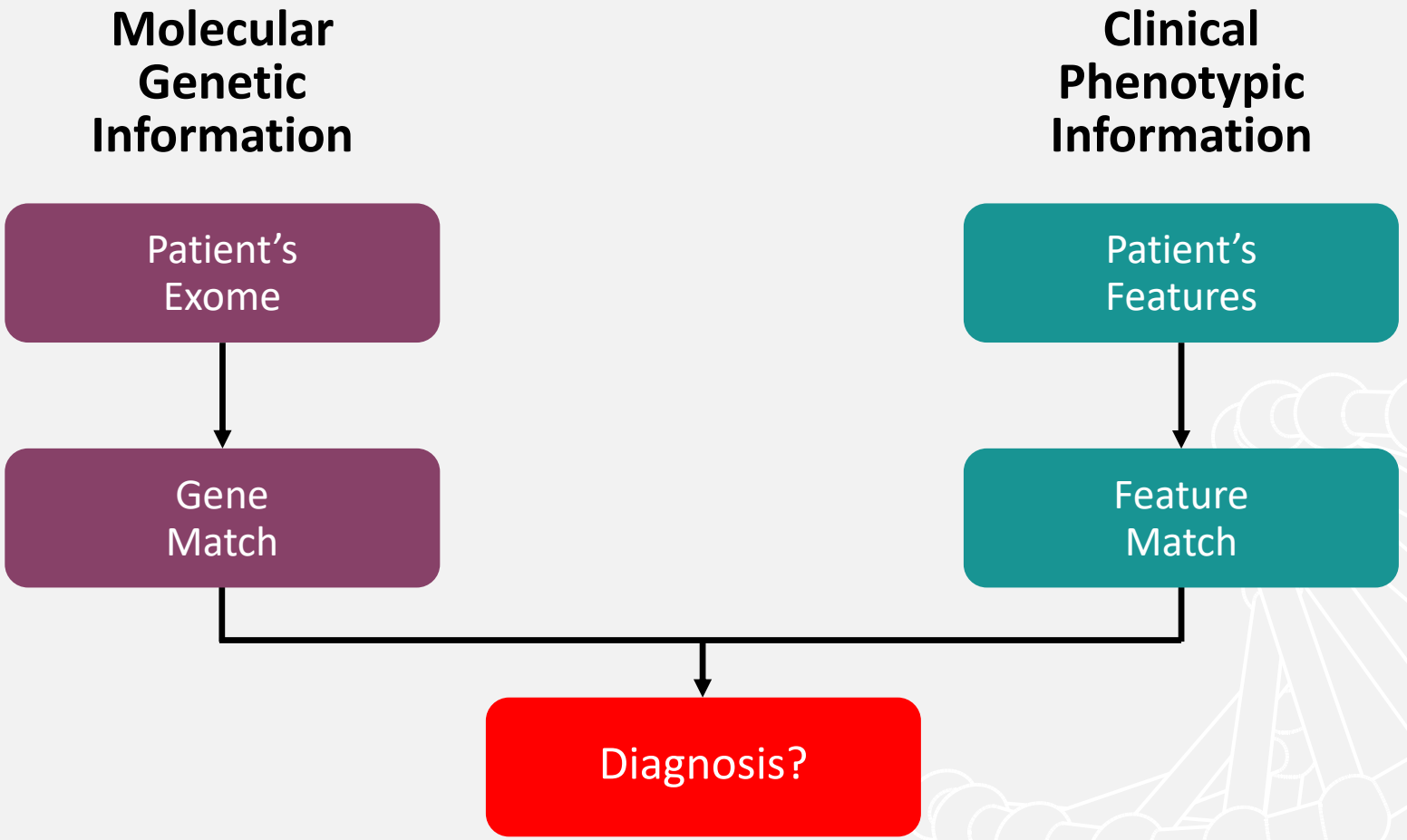


Photo

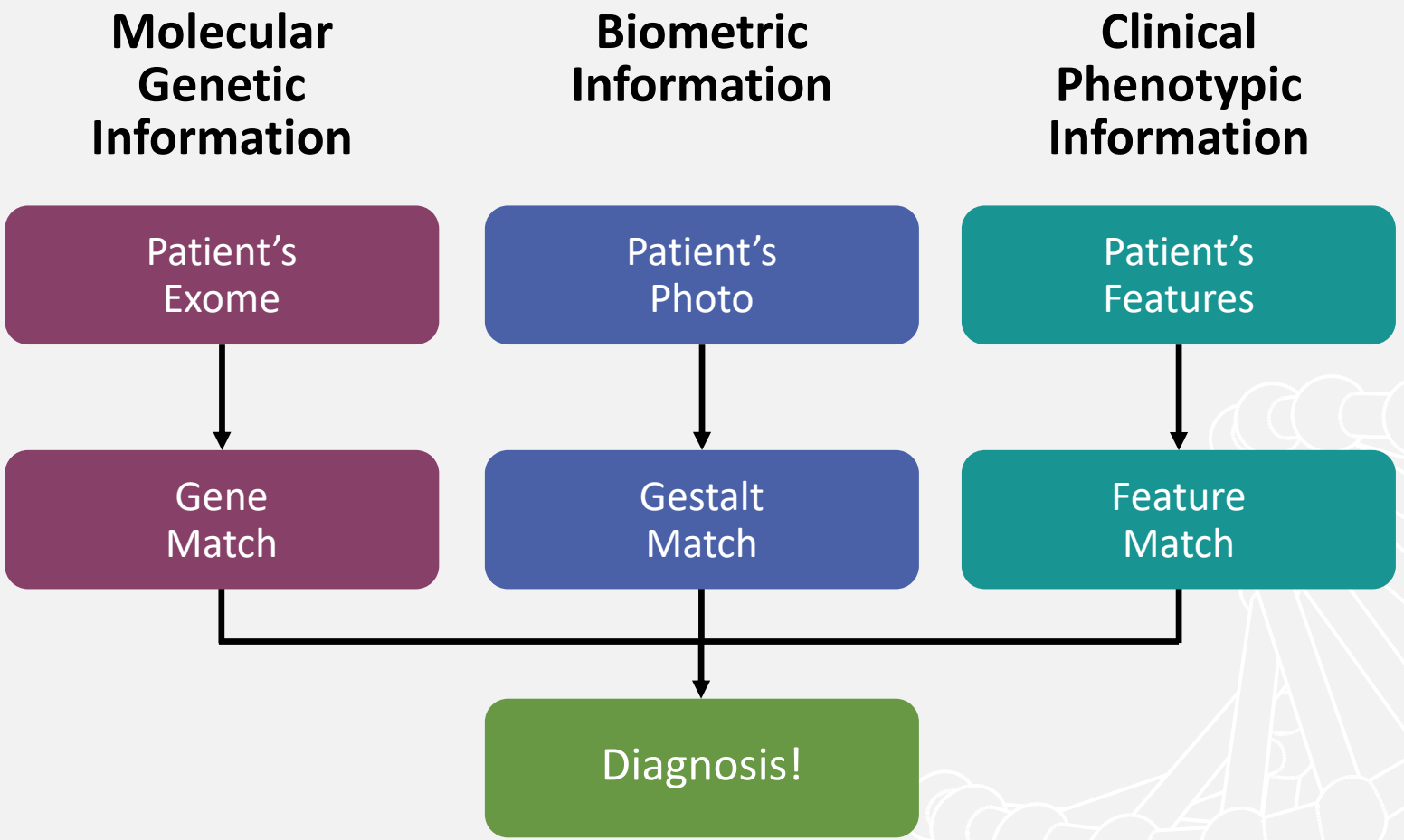
NM 033419.3:
c.402G>A

Disease-causing mutation

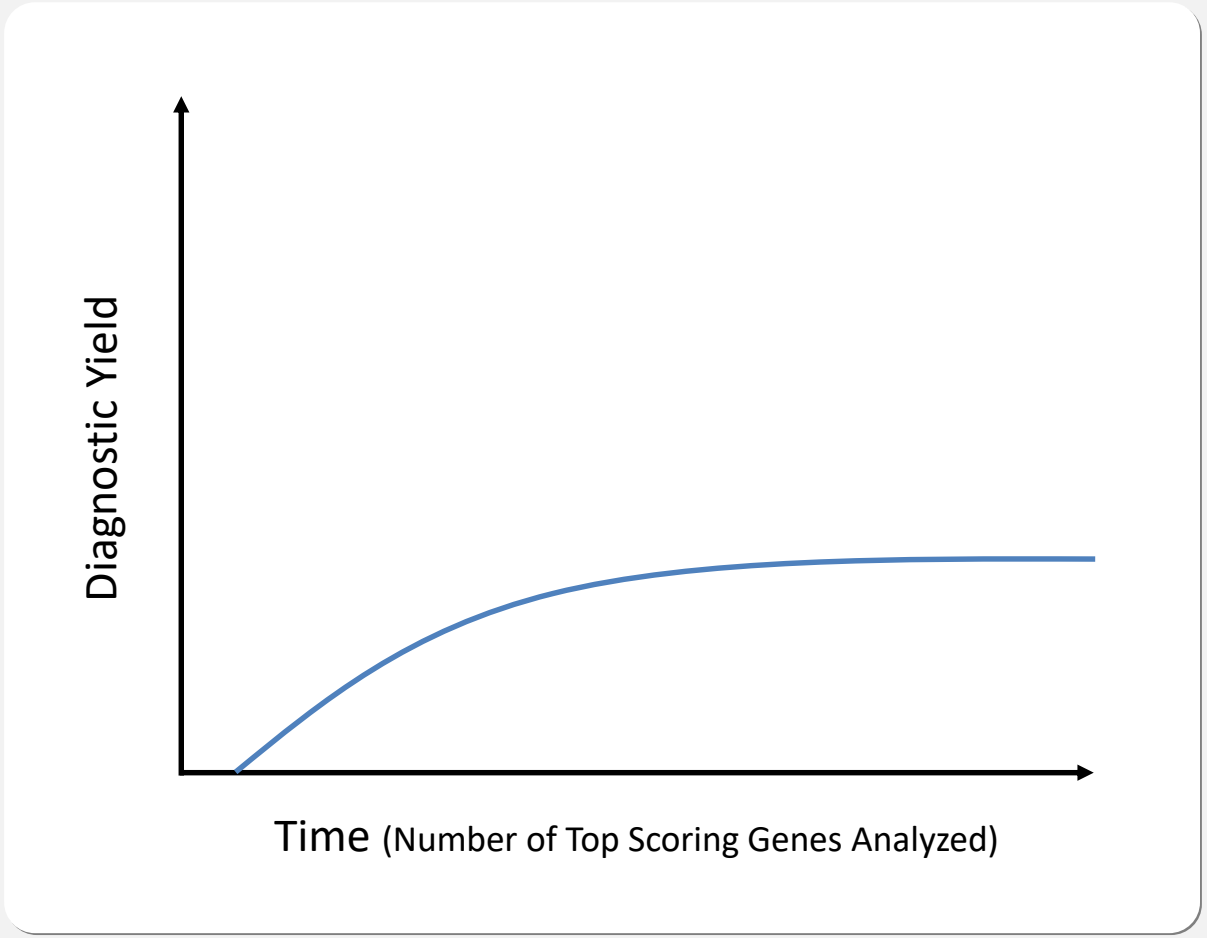
PEDIA = **P**rioritization of **E**xome **D**ata by **I**mage **A**nalysis



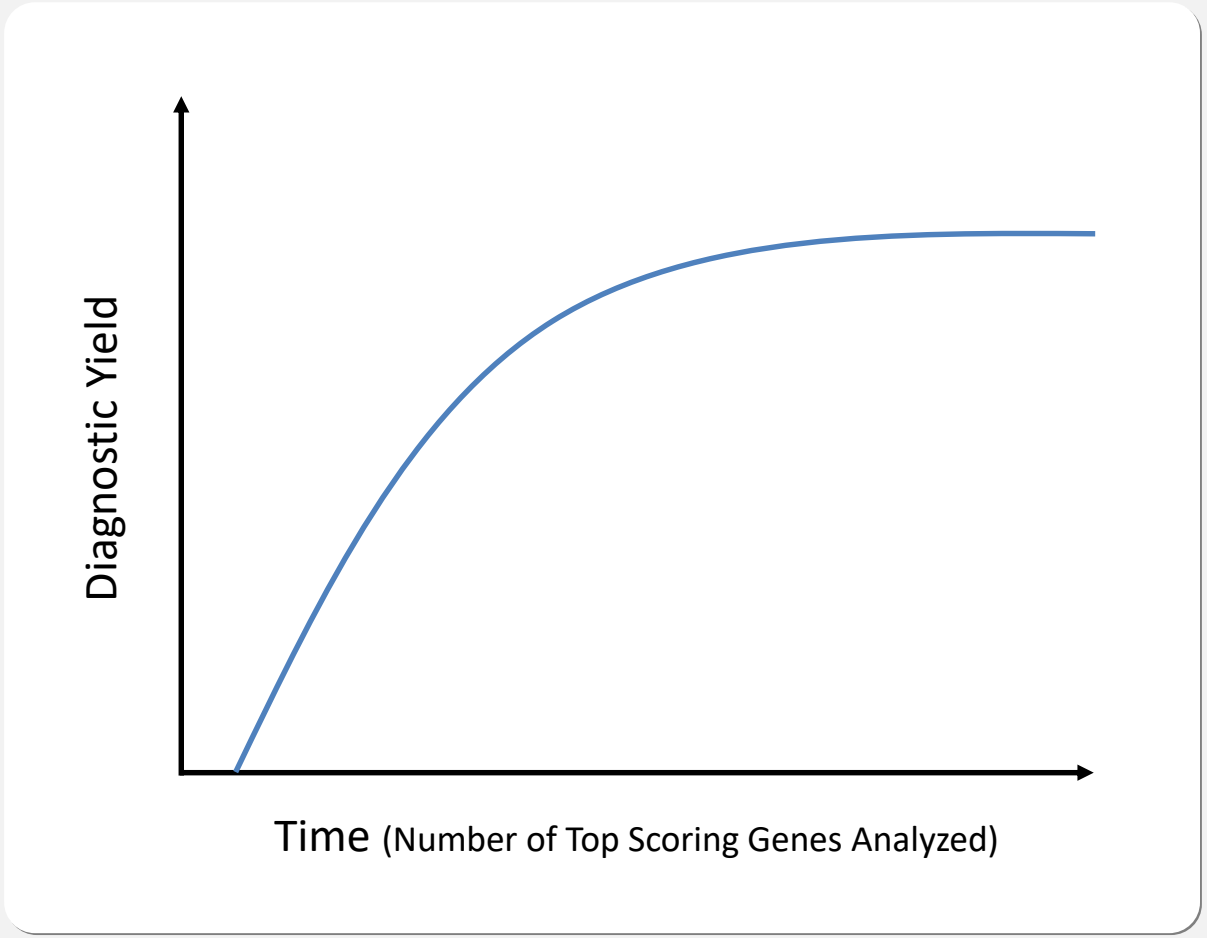
PEDIA = **P**rioritization of **E**xome **D**ata by **I**mage **A**nalysis



PEDIA Approach for the Lab



PEDIA Approach for the Lab



Increase diagnostic yield by
quantifying phenotype information
(PP4 in ACMG guidelines*):

"Patient's phenotype or family history
is highly specific for
a disease with a single genetic etiology"

** Sue Richards, Standards and Guidelines for the Interpretation of Sequence Variants, Genet Med. 2015*

FDNA

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

The American College of Medical Genetics and Genomics (ACMG) previously developed guidance for the interpretation of sequence variants.[1] In the past decade, sequencing technology has evolved rapidly with the advent of high-throughput next-generation sequencing. By adopting and leveraging next-generation sequencing, clinical laboratories are now performing an ever-increasing catalogue of genetic testing spanning genotyping, single genes, gene panels, exomes, genomes, transcriptomes, and epigenetic assays for genetic disorders. By virtue of increased complexity, this shift in genetic testing has been accompanied by new challenges in sequence interpretation. In this context the ACMG convened a workgroup in 2013 comprising representatives from the ACMG, the Association for Molecular Pathology (AMP), and the College of American Pathologists to revisit and revise the standards and guidelines for the interpretation of sequence variants. The group consisted of clinical laboratory directors and clinicians. This report represents expert opinion of the workgroup with input from ACMG, AMP, and College of American Pathologists stakeholders. These recommendations primarily apply to the breadth of genetic tests used in clinical laboratories, including genotyping, single genes, panels, exomes, and genomes. This report recommends the use of specific standard terminology—"pathogenic," "likely pathogenic," "uncertain significance," "likely benign," and "benign"—to describe variants identified in genes that cause Mendelian disorders. Moreover, this recommendation describes a process for classifying variants into these five categories based on criteria using typical types of variant evidence (e.g., population data, computational data, functional data, segregation data). Because of the increased complexity of analysis and interpretation of clinical genetic testing described in this report, the ACMG strongly recommends that clinical molecular genetic testing should be performed in a Clinical Laboratory Improvement Amendments–approved laboratory, with results interpreted by a board-certified clinical molecular geneticist or molecular genetic pathologist or the equivalent.

Genet Med advance online publication 5 March 2015

Key Words: ACMG laboratory guideline; clinical genetic testing; interpretation; reporting; sequence variant terminology; variant reporting

[1]Department of Molecular and Medical Genetics, Knight Diagnostic Laboratories, Oregon Health & Science University, Portland, Oregon, USA; [2]College of American Pathologists, Chicago, Illinois, USA; [3]GeneDx, Gaithersburg, Maryland, USA; [4]Department of Pediatrics, Section of Genetics, Medical College of Wisconsin, Milwaukee, Wisconsin, USA; [5]Department of Human Genetics, Clinical Molecular Genetics Laboratory, The University of Chicago, Chicago, Illinois, USA; [6]Cytogenetics/Molecular Genetics Laboratory, Nationwide Children's Hospital, Columbus, Ohio, USA; [7]Department of Pathology, Ohio State University College of Medicine, Columbus, Ohio, USA; [8]Department of Pediatrics, Ohio State University College of Medicine, Columbus, Ohio, USA; [9]Department of Pathology and Laboratory Medicine, University of California Los Angeles School of Medicine, Los Angeles, California, USA; [10]Department of Pediatrics, University of California Los Angeles School of Medicine, Los Angeles, California, USA; [11]Department of Human Genetics, University of California Los Angeles School of Medicine, Los Angeles, California, USA; [12]Department of Human Genetics, Emory Genetics Laboratory, Emory University, Atlanta, Georgia, USA; [13]Department of Pathology, ARUP Institute for Clinical and Experimental Pathology, University of Utah, Salt Lake City, Utah, USA; [14]Department of Pediatrics, Molecular Genetics Laboratory, Children's Hospital Colorado, University of Colorado Anschutz Medical School, Denver, Colorado, USA; [15]Partners Laboratory for Molecular Medicine and Department of Pathology, Brigham & Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; [16]Current affiliation: Phoenix Children's Hospital, Phoenix, Arizona, USA. Correspondence: Sue Richards (richarsu@ohsu.edu)

Approved by the ACMG Board of Directors on 15 December 2014 and the AMP Board of Directors on 9 January 2015.

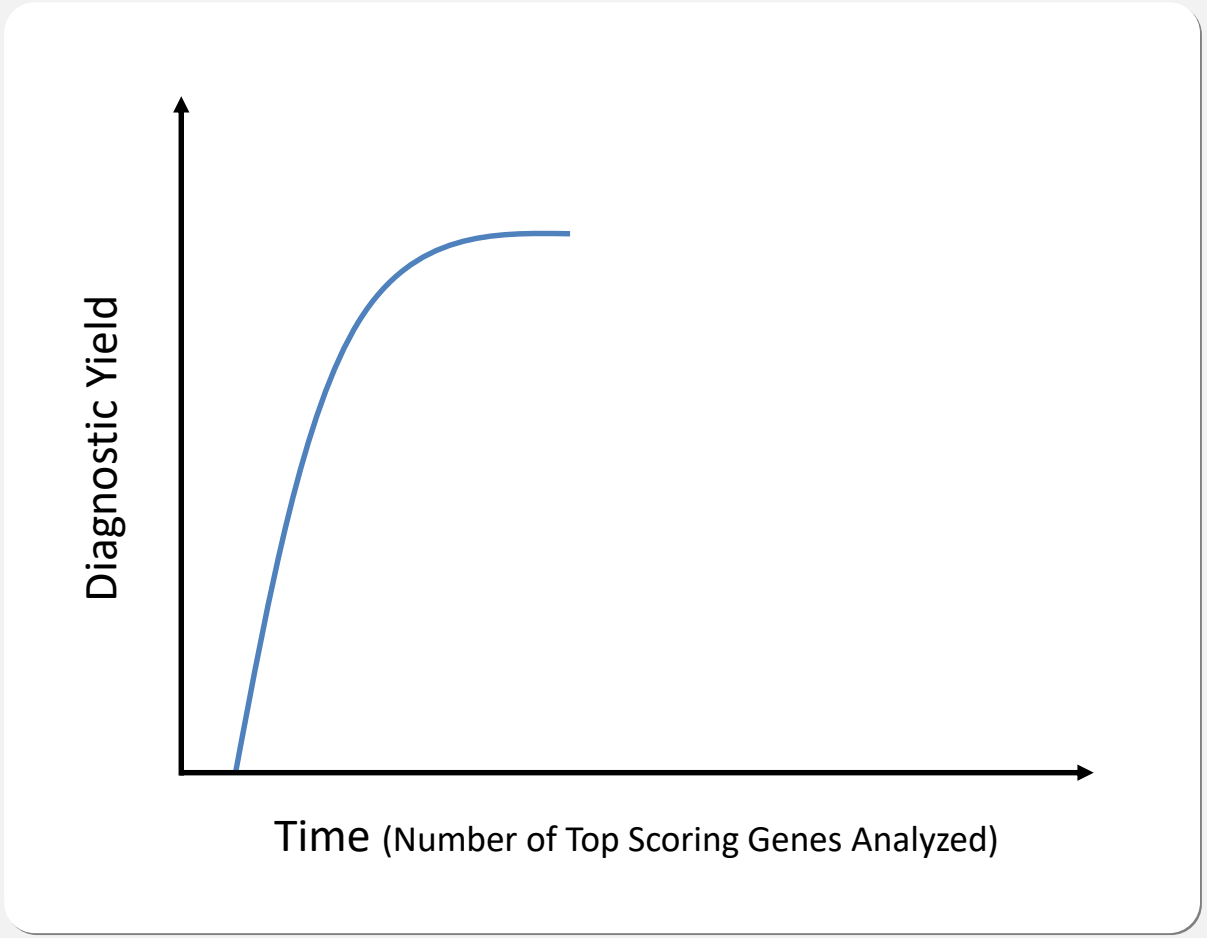
Submitted 28 January 2015; accepted 28 January 2015; advance online publication 5 March 2015. doi:10.1038/gim.2015.30

Table 3 Criteria for classifying pathogenic variants

Evidence of pathogenicity	Category
Very strong	PVS1 null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease Caveats: • Beware of genes where LOF is not a known disease mechanism (e.g., *GFAP*, *MYH7*) • Use caution interpreting LOF variants at the extreme 3′ end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity. PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established. PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case–control studies, is >5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance. Note 2: In instances of very rare variants where case–control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.
Moderate	PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation PM2 Absent from controls (or at extremely low frequency if recessive) (**Table 6**) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing. PM3 For recessive disorders, detected in *trans* with a pathogenic variant Note: This requires testing of parents (or offspring) to determine phase. PM4 Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before Example: Arg156His is pathogenic; now you observe Arg156Cys Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level. PM6 Assumed de novo, but without confirmation of paternity and maternity
Supporting	PP1 Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease Note: May be used as stronger evidence with increasing segregation data PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant. PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

LOF, loss of function; OR, odds ratio.

PEDIA Approach for the Lab

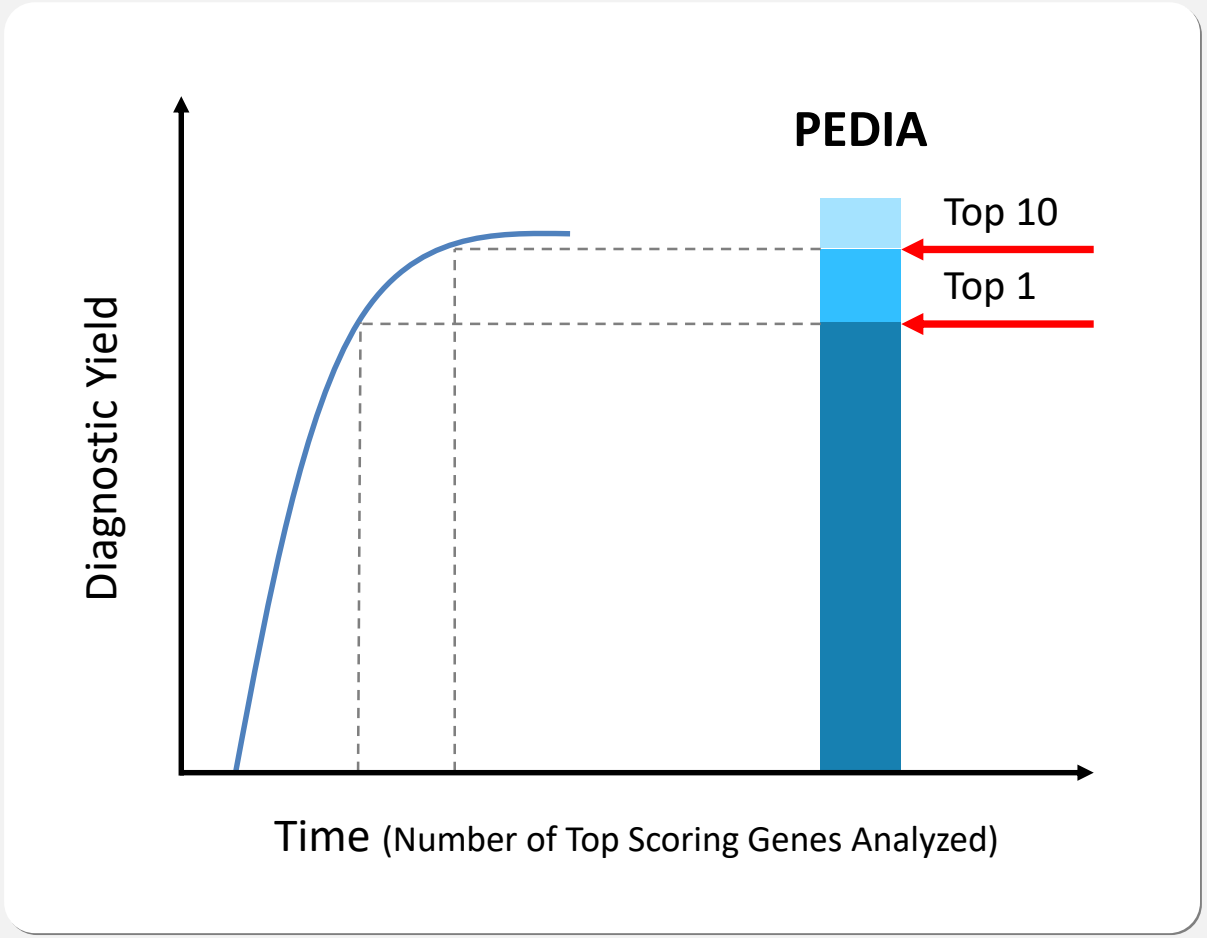


Reach the same diagnostic yield in less time by automatizing the transfer of NGP data by the Face2Gene Lab API

** Sue Richards, ... , Hedi Rehm, Standards and Guidelines for the Interpretation of Sequence Variants, Genet Med. 2015*

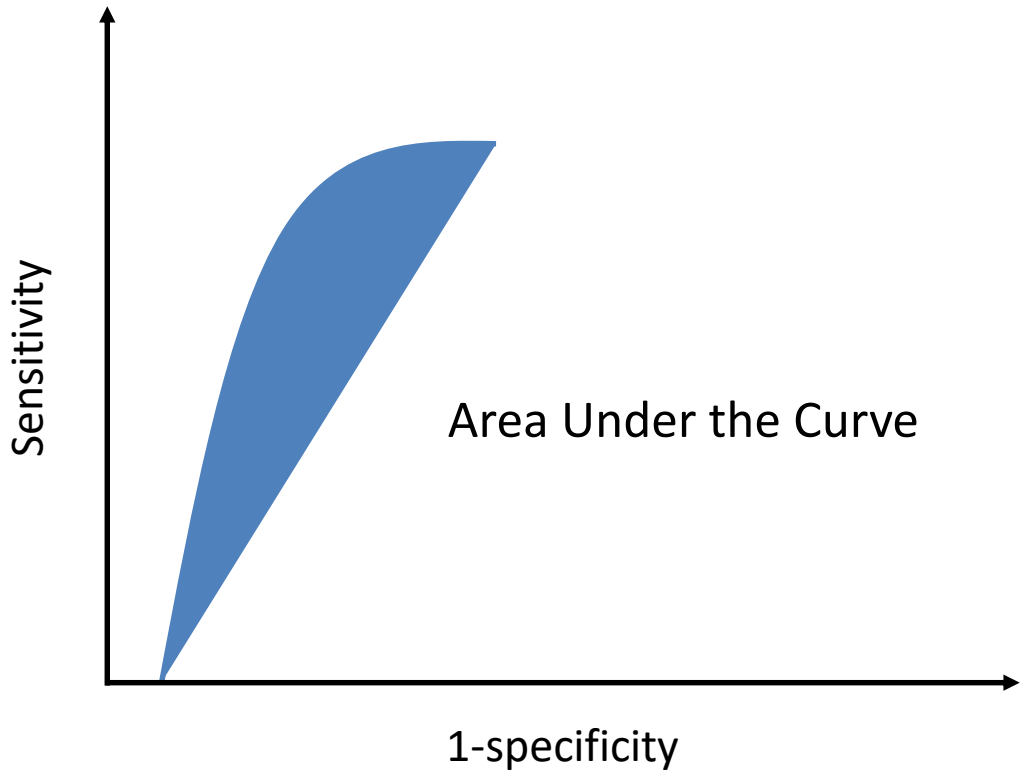
FDNA

PEDIA Approach for the Lab

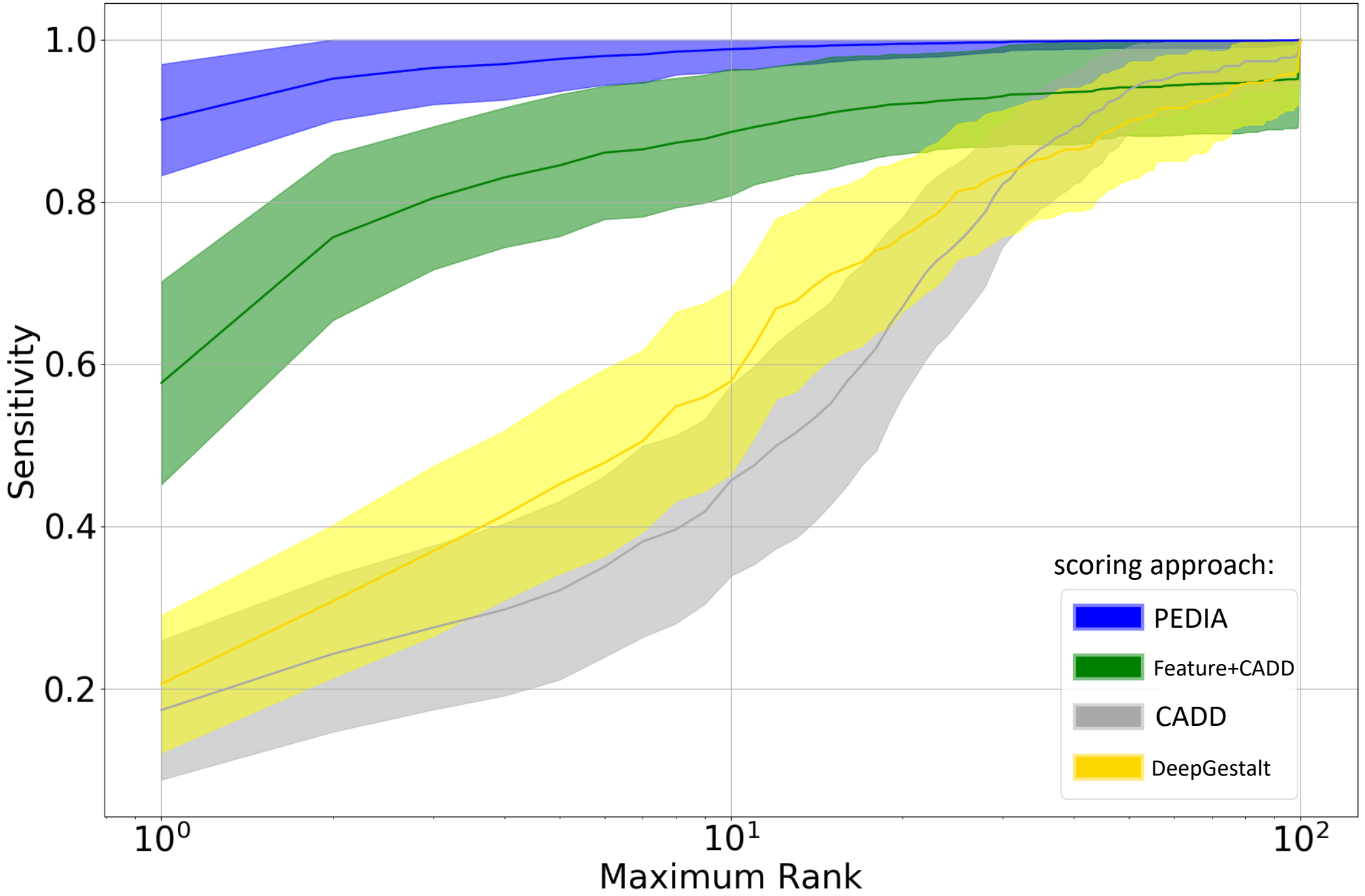


Accuracy rates:
the proportion of cases in which the correct disease gene is listed at the first position (top 1) or amonst the first ten genes (top 10 accuracy)

PEDIA Approach for the Lab



The value that DeepGestalt adds to any existing bioinformatics workflow can also be measured by the AUC.


1.0
0.8
0.6
0.4
0.2
Sensitivity
10^0
10^1
10^2
Maximum Rank
scoring approach:
PEDIA
Feature+CADD
CADD
DeepGestalt

Acknowledgements

Peter Krawitz
Martin Mensah
Max Schubach
Nurulhuda Hajjir
Sebastian Köhler
Marten Jäger
Tori-Jean Pantel
Fabian Brand
Carlo Maj
Oleg Borisov
Dejan Đukić
Nadja Ehmke
Tom Kamphans
Marie Coutelier

Ricarda Flöttmann
Magdalena Danyel
Uwe Kornak
Claus-Eric Ott
Denise Horn
Luitgard Graul-Neumann
Alexej Knaus
Max Zhao
Irena Vrecar
Gholson Lyon
Malte Spielmann
Stefan Mundlos

Alfredo Orrico
Diana Mitter
Manuel Corpas
Nicole Fleischer
Yaron Gurovich
Sharon Wolfson
Karin Boss
Bernd Wollnik
Ivan Ivanovski
Gorazd Rudolf
Susanne Kamphausen
Karen Gripp
Angelika Kaindl
Dagmar Wahl

Sandra Willson
Franz Alisch
Koenraad Devriendt
Sofia Annas
Hilda David Eden
Omri Bar
Yair Hanani
Marzena Wiśniewska
Antonio Martinez Carrascal
IGSB group members

The patients & their families

and many more

Thank you!

UNIVERSITÄT BONN

Institut für Genomische Statistik und Bioinformatik

FACE2GENE

CHARITÉ UNIVERSITÄTSMEDIZIN BERLIN

DeepGestalt

Was ist NGP?

Wie kann ich schwere Fälle schneller knacken?

Was mach ich mit den ungelösten?

PEDIA

GestaltMatcher

LETTERS | FOCUS
https://doi.org/10.1038/s41591-018-0279-0
nature medicine

Identifying facial phenotypes of genetic disorders using deep learning

Yaron Gurovich[1*], Yair Hanani[1], Omri Bar[1], Guy Nadav[1], Nicole Fleischer[1], Dekel Gelbman[1], Lina Basel-Salmon[2,3], Peter M. Krawitz[4], Susanne B. Kamphausen[5], Martin Zenker[5], Lynne M. Bird[6,7] and Karen W. Gripp[8]

Syndromic genetic conditions, in aggregate, affect 8% of the population[1]. Many syndromes have recognizable facial features[2] that are highly informative to clinical geneticists[3–5]. Recent studies show that facial analysis technologies measured up to the capabilities of expert clinicians in syndrome identification[6–9]. However, these technologies identified only a few disease phenotypes, limiting their role in clinical settings, where hundreds of diagnoses must be considered. Here we present a facial image analysis framework, DeepGestalt, using computer vision and deep-learning algorithms, that quantifies similarities to hundreds of syndromes. DeepGestalt outperformed clinicians in three initial experiments, two with the goal of distinguishing subjects with a target syndrome from other syndromes, and one of separating different genetic subtypes in Noonan syndrome. On the final experiment reflecting a real clinical setting problem, DeepGestalt achieved 91% top-10 accuracy in identifying the correct syndrome on 502 different images. The model was trained on a dataset of over 17,000 images representing more than 200 syndromes, curated through a community-driven phenotyping platform. DeepGestalt potentially adds considerable value to phenotypic evaluations in clinical genetics, genetic testing, research and precision medicine.

ARTICLE Genetics inMedicine

Open

PEDIA: prioritization of exome data by image analysis

A full list of authors and affiliations appears at the end of the paper.

Purpose: Phenotype information is crucial for the interpretation of genomic variants. So far it has only been accessible for bioinformatics workflows after encoding into clinical terms by expert dysmorphologists.

Methods: Here, we introduce an approach driven by artificial intelligence that uses portrait photographs for the interpretation of clinical exome data. We measured the value added by computer-assisted image analysis to the diagnostic yield on a cohort consisting of 679 individuals with 105 different monogenic disorders. For each case in the cohort we compiled frontal photos, clinical features, and the disease-causing variants, and simulated multiple exomes of different ethnic backgrounds.

Results: The additional use of similarity scores from computer-assisted analysis of frontal photos improved the top 1 accuracy rate

REPORT

The Discovery of a *LEMD2*-Associated Nuclear Envelopathy with Early Progeroid Appearance Suggests Advanced Applications for AI-Driven Facial Phenotyping

Felix Marbach,[1,2,19] Cecilie F. Rustad,[3,19] Angelika Riess,[4,19] Dejan Đukić,[5] Tzung-Chien Hsieh,[6] Itamar Jobani,[7] Trine Prescott,[8] Andrea Bevot,[9] Florian Erger,[1,2] Gunnar Houge,[10,11] Maria Redfors,[12,13] Janine Altmueller,[14] Tomasz Stokowy,[11] Christian Gilissen,[15] Christian Kubisch,[16] Emanuela Scarano,[17] Laura Mazzanti,[17] Torunn Fiskerstrand,[10,11,18] Peter M. Krawitz,[6] Davor Lessel,[16,20] and Christian Netzer[1,2,20,*]

Over a relatively short period of time, the clinical geneticist's "toolbox" has been expanded by machine-learning algorithms for image analysis, which can be applied to the task of syndrome identification on the basis of facial photographs, but these technologies harbor potential beyond the recognition of established phenotypes. Here, we comprehensively characterized two individuals with a hitherto unknown genetic disorder caused by the same *de novo* mutation in *LEMD2* (c.1436C>T;p.Ser479Phe), the gene which encodes the nuclear envelope protein LEM domain-containing protein 2 (LEMD2). Despite different ages and ethnic backgrounds, both individuals share a progeria-like facial phenotype and a distinct combination of physical and neurologic anomalies, such as growth retardation; hypoplastic jaws crowded with multiple supernumerary, yet unerupted, teeth; and cerebellar intention tremor. Immunofluorescence analyses of patient fibroblasts revealed mutation-induced disturbance of nuclear architecture, recapitulating previously published data in LEMD2-deficient cell lines, and additional experiments suggested mislocalization of mutant LEMD2 protein within the nuclear lamina. Computational analysis of facial features with two different deep neural networks showed phenotypic proximity to other nuclear envelopathies. One of the algorithms, when trained to recognize syndromic similarity (rather than specific syndromes) in an unsupervised approach, clustered both individuals closely together, providing hypothesis-free hints for a common genetic etiology. We show that a recurrent *de novo* mutation in *LEMD2* causes a nuclear envelopathy whose prognosis in adolescence is relatively good in comparison to that of classical Hutchinson-Gilford progeria syndrome, and we suggest that the application of artificial intelligence to the analysis of patient images can facilitate the discovery of new genetic disorders.