PEDIA: Prioritization of Exome Data by Image Analysis

Hsieh, Tzung-Chien[1,2,*]; Mensah, Martin Atta[2,41*]; Pantel, Jean Tori[1,2,41,*]; Krawitz, Peter[1,+]; and the PEDIA consortium

PEDIA consortium: Aguilar, Dione[3]; Bar, Omri[4]; Bayat, Allan[5]; Becerra-Solano, Luis[6]; Bentzen, Heidi Beate[7]; Biskup, Saskia[8]; Borisov, Oleg[1]; Braaten, Oivind[7]; Ciaccio, Claudia[9]; Coutelier, Marie[2]; Cremer, Kirsten[10]; Danyel, Magdalena[2]; Daschkey, Svenja[11]; David-Eden, Hilda[4]; Devriendt, Koenraad[12]; Dölken, Sandra[13]; Douzgou, Sofia[14]; Đukić, Dejan[1]; Ehmke, Nadja[2]; Fauth, Christine[15]; Fischer-Zirnsak, Björn[2]; Fleischer, Nicole[4]; Gabriel, Heinz[16]; Graul-Neumann, Luitgard[2]; Gripp, Karen W.[17]; Gurovich, Yaron[4]; Gusina, Asya[18]; Haddad, Nechama[2]; Hajjir, Nurulhuda[2]; Hanani, Yair[4]; Hertzberg, Jakob[2]; Hoertnagel, Konstanze[8]; Howell, Janelle[19]; Ivanovski, Ivan[20]; Kaindl, Angela[21]; Kamphans, Tom[22]; Kamphausen, Susanne[23]; Karimov, Catherine[24]; Kathom, Hadil[25]; Keryan, Anna[24]; Khalil, Salma-Gamal[2]; Knaus, Alexej[1]; Köhler, Sebastian[26]; Kornak, Uwe[2]; Lavrov, Alexander[27]; Leitheiser, Maximilian[2]; Lyon, Gholson J.[28]; Mangold, Elisabeth[29]; Marín Reina, Purificación[30]; Martinez Carrascal, Antonio[31]; Mitter, Diana[32]; Morlan Herrador, Laura[33]; Nadav, Guy[4]; Nöthen, Markus[10]; Orrico, Alfredo[34]; Ott, Claus-Eric[2]; Park, Kristen[35]; Peterlin, Borut[36]; Pölsler, Laura[15]; Raas-Rothschild, Annick[37]; Revencu, Nicole[38]; Ringmann Fagerberg, Christina[39]; Robinson, Peter Nick[40]; Rosnev, Stanislav[2]; Rudnik, Sabine[15]; Rudolf, Gorazd[36]; Schatz, Ulrich[15]; Schossig, Anna[15]; Schubach, Max[41]; Shanoon, Or[4]; Sheridan, Eamonn[42]; Smirin-Yosef, Pola[43]; Spielmann, Malte[44]; Suk, Eun-Kyung[45]; Sznajer, Yves[46]; Thiel, Christian Thomas[47]; Thiel, Gundula[45]; Verloes, Alain[48]; Vrecar, Irena[36]; Wahl, Dagmar[49]; Weber, Ingrid[15]; Winter, Korina[2]; Wiśniewska, Marzena[50]; Wollnik, Bernd[51]; Yeung, Ming Wai[1]; Zhao, Max[2]; Zhu, Na[2]; Zschocke, Johannes[15]; Mundlos, Stefan[2]; Horn, Denise[2]

1 Institute of Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany
2 Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Medical Genetics and Human Genetics, Berlin, Germany
3 Monterrey Institute of Technology and Higher Education, Mexico
4 FDNA Inc., Boston Massachusetts, United States
5 Rigshospitalet, Department of Neurology, Copenhagen, Denmark
6 Unidad de Investigación Médica en Medicina Reproductiva, Mexico
7 University of Oslo, Oslo, Norway
8 CeGaT GmbH, Tübingen, Germany
9 University of Milan, Milan, Italy
10 Department of Human Genetics, University Hospital of Bonn, Bonn, Germany
11 Heinrich Heine University Düsseldorf, Düsseldorf, Germany
12 Catholic University Leuven, Leuven, Belgium
13 University of Hamburg, Hamburg, Germany
14 University of Manchester, Manchester, United Kingdom
15 Division of Human Genetics, Medical University of Innsbruck, Innsbruck, Austria
16 University of Tübingen, Tübingen, Germany
17 A. I. duPont Hospital for Children, Wilmington, United States
18 National Research and Applied Medicine Centre ‘‘Mother and Child’’, Belarus
19 Lineagen, United States
20 Santa Maria Nuova Hospital, Italy
21 Center for Chronically Sick Children (Sozialpädiatrisches Zentrum, SPZ), Charité - Universitätsmedizin Berlin, Berlin, Germany

22 GeneTalk, Bonn, Germany
23 University Hospital Magdeburg, Magdeburg, Germany
24 Children's Hospital of Los Angeles, Los Angeles, United States
25 Medical University of Sofia, Sofia, Bulgaria
26 Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, NeuroCure Clinical Research Center, Berlin, Germany
27 Research Institute of Medical Genetics of Russian Academy of Medical Sciences, Russian Federation
28 Cold Spring Harbor Laboratory, Woodbury, United States
29 University of Bonn, Bonn, Germany
30 Hospital General Universitario De Valencia, Valencia, Spain
31 Hospital General De Requena, Servicio Pediatría, Spain
32 University Hospital Leipzig, Leipzig, Germany
33 Hospital Universitario Miguel Servet, Spain
34 Azienda Ospedaliera Universitaria Senese, Siena, Italy
35 Children's Hospital Colorado, United States
36 Clinical Institute of Medical Genetics, University Medical Centre Ljubljana, Ljubljana, Slovenia
37 Sheba Medical Center, Israel
38 Université Catholique de Louvain, Bruxelles, Belgium
39 Odense University Hospital, Odense, Denmark
40 The Jackson Laboratory for Genomic Medicine, Farmington, United States
41 Berlin Institute of Health (BIH), Anna-Louisa-Karsch 2, 10178 Berlin, Germany
42 School of Medicine, University of Leeds, Leeds, United Kingdom
43 Ariel University, Ariel, Israel
44 Department of Genome Sciences, University of Washington, Seattle, United States
45 Center for Prenatal Diagnosis and Human Genetics, Berlin, Germany
46 Cliniques universitaires Saint Luc UCL, Bruxelles, Belgium
47 Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg FAU, Erlangen, Erlangen, Germany
48 Hopital Robert Debré, Paris, France
49 Center for Human Genetics and Laboratory Diagnostics, Germany
50 Poznañ University of Medical Sciences, Poznañ, Poland
51 University Medical Center Göttingen, Göttingen, Germany
* equally contributing first authors
+ corresponding author, pkrawitz@uni-bonn.de

Abstract

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. 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 mutations and simulated multiple exomes of different ethnic backgrounds. With the additional use of similarity scores from computer-assisted analysis of frontal photos, we were able to achieve a top-10-accuracy rate for the disease-causing gene of 99 %. As this performance is significantly higher than without the information from facial pattern recognition, we make gestalt scores available for prioritization via an API.

Rare diseases affect approximately 6% of the population, with genetic syndromes accounting for about 80 %.[1,2] The more than 5,000 entities represent a heterogeneous group of diseases, differing in cause, symptoms, and treatment, making diagnosis an important yet challenging healthcare issue. Due to extensive clinical variability this is true even for well characterized syndromes.[1,3]

Worldwide, more than half a million children born per year have a rare genetic disorder that is suitable for a diagnostic workup by exome sequencing, which has an unprecedented diagnostic yield for many indications such as developmental delay.[4-9] The main remaining concern for the integration of exome sequencing into clinical routine is to increase the efficiency of genetic variant interpretation. Making phenotypic information – the observable, clinical presentation – computer-readable is key in solving this problem, and in providing clinicians with a much-needed tool for diagnosing genetic syndromes.[10]

To date, the most advanced exome prioritization algorithms combine deleteriousness scores for mutations with semantic similarity searches of the clinical description of a patient.[11-15] The human phenotype ontology (HPO) with its extensive vocabulary has become the *lingua franca* for this purpose.[16] However, semantic similarity searches presuppose that facial features can be named. A facial gestalt that is simply described in the literature as *typical* or *characteristic* of a certain disease is of little help for these approaches.

Beyond language, capturing indicative patterns by deep-learning approaches has recently gained attention in assessing facial dysmorphism.[17-21] Artificial neural networks are now able to quantify the similarities of patient photos to hundreds of disease entities and achieve accuracies that match or even surpass the level of dysmorphologists in certain tasks.[22-25] For this reason tools such as Face2Gene are now used in addition to human expertise to guide the molecular testing and to interpret sequence variants. Here we investigate systematically whether facial image analysis can improve the evaluation of exome data and qualifies as a next-generation phenotyping technology for next-generation sequencing.[26]

Results

We first present an overview about the approach to prioritize exome data by image analysis (PEDIA); a detailed description is provided in the Methods.

PEDIA classifier. For the assessment of genetic variants, different sources of evidence have to be considered, from a populational, molecular, and phenotypic level. PEDIA is a Bayesian heuristic, that

can be used to update the probability that a mutation in a gene is disease-causing, given the phenotypic information contained in a frontal photograph.

To build this classifier, we first measured the similarities of the facial gestalt to 216 specific diseases in 679 individuals with the convolutional neural network DeepGestalt.[21] By this means, we were able to acquire scores for disorders with a single genetic etiology that quantify the PP4 criteria of the ACMG guidelines which is used for variant interpretation.[27,28]

In addition to DeepGestalt, we computed further prediction scores that are widely used on clinical features (Phenomizer, Boqa, Feature) and genetic variants (CADD) for all individuals of the PEDIA cohort (Supplemental Table 1). [29,30,31] With this data set we trained and tested a support vector machine that can be used to prioritize the genetic variants in a VCF files from exome sequencing.

Gene prioritization.

The term next-generation sequencing (NGS) implies the interrogation of all genes in a single assay. Similarly, the term next-generation phenotyping (NGP) refers to technology enabling similarity searches on a large set of disorders based on clinical patient records and medical imaging data. In order to increase the efficiency in diagnostics, we combined both approaches and benchmarked gene prioritization.

Similar to the performance readout in Gurovich et al., the identification of the disease-gene in exome data also represents a multiclass classification problem and the number of sequence variants in the coding part of the genome illustrates the complexity of the diagnostic assessment. In reference guided-resequencing, about 20,000-30,000 single nucleotide variants and small indels have to be considered. Although the majority of these variants can be removed as benign polymorphisms, rare and potentially disease-causing mutations in more than 100 genes remain in a typical case with a suspected monogenic disorder. When only a deleteriousness score such as CADD is used to rank these mutations, the disease-causing gene is in the top 10 in less than 46 % of the cases of the PEDIA cohort. This performance increases to a top-10-accuracy rate of up to 88 %, when semantic similarity scores are included that are based on HPO feature annotations. These prioritization approaches also represent the current state of the art in diagnostic laboratories for single exomes.[13,14] The additional information contained in frontal photos of dysmorphic cases pushes the correct disease-gene to the top-10 in more than 99 % of the cases in the PEDIA cohort and in the DeepGestalt test set (Figure 1 B).

The value of a frontal photograph can exemplarily be demonstrated by a case with Coffin-Siris syndrome that is shown in Figure 2 A: The characteristic facial features are relatively mild, so the correct diagnosis is only listed as the third suggestion by DeepGestalt. Amongst all the variants encountered in an exome data set, the disease-causing gene *ARID1B* would only achieve rank 24, if scored by the molecular information alone. However, in synopsis with the phenotypic information, the PEDIA approach lists this gene as first candidate by far (Figure 2 C).

Although the syndrome of the case shown in Figure 2 might also be molecularly confirmed by a directed single gene test in other instances where the facial gestalt is more indicative, the high phenotypic variability associated with disease-causing mutations is well-known for genes of syndromic disorders. It has been exhibited in the deciphering developmental disorders (DDD) project, that many such diagnoses were made only after exome sequencing.[6] This finding is also reflected by frontal image analysis of the entire PEDIA cohort with DeepGestalt alone that achieves a top-10-accuracy rate for the disease-causing gene of around 58 %.

The efficiency of a prioritization algorithm can also be measured by the area under the curve (AUC) of the disease-causing mutation versus its ranked position. The higher the AUC, the higher the diagnostic yield in a fixed amount of time that is spend on the analysis of sequence variants (Figure 3). Combining similarity scores from image analysis, phenotypic features and molecular deleteriousness achieves the best AUC on the PEDIA cohort and is therefore suited to speed up diagnostics.

The contribution from the different sources of evidence to the PEDIA score is also reflected by the relative weight of the deleteriousness of the mutation (0.44), all feature-based scores combined (0.25) and the results from image analysis by DeepGestalt (0.31) that can be derived from a linear SVM model. We therefore also conclude that the information contained in a frontal photograph of patient goes beyond, what clinical terms can capture.

Discussion

According to the current version of the Online Mendelian Inheritance of Man Catalog, mutations in about 4000 genes are linked to phenotypes that are often difficult to distinguish and diagnose by clinical features alone, making next-generation sequencing a key technology for their molecular confirmation. However, the size and high variability of the genome as well as the low prevalence of disease-causing variants – many of them occur *de novo* – explain why sequence data analysis of a single individual is still challenging and time consuming.[5,6]

The guidelines for variant classification in the laboratory follow a qualitative heuristic that combines distinct types of evidence (functional, population, phenotype, etc.) and is compatible with Bayesian statistics.[32] The advantage of such a framework is that continuous evidence types can be integrated into the classification system. While *in silico* predictions about a variant's pathogenicity have a relatively long history in bioinformatics and machine learning, the quantification of phenotypic raw data with systems of artificial intelligence just began. Analogous to a score for the deleteriousness of a gene variant, one can include the phenotypic similarity to a distinct syndrome caused by mutations in the respective gene.

We analyzed this approach in the PEDIA cohort, consisting of 679 cases and covering 105 distinct disorders mapping to 181 disease-genes. Among these disorders were 73 phenotypes for which the performance of facial image analysis alone has recently been evaluated.[21] Although the top-10-accuracies for gestalt- and PEDIA-scoring cannot be compared directly, both approaches operate on a similar order of phenotypes and genes, respectively. Adding suitable molecular information to 260 cases from the DeepGestalt publication test set increased the correct disease-gene in the top 10 to about 99%, from 90% with only the phenotypic information. Considering only molecular information and clinical features, but without the results from image analysis, the correct disease gene would have only been placed in the top 10 in 62%. The genetic background, which might correspond to a different number of variant calls or higher load of deleterious mutations, had negligible influence on the performance.

The performance for the entire PEDIA cohort is comparable to the DeepGestalt test set. However, there are three important lessons learned from specific subgroups or cases achieving lower PEDIA ranks: 1) Although the convolutional neural network used for image analysis has been pretrained on real-world uncontrolled 2D images, patient photographs that were true frontal, of high resolution, with good lightening and contrast, and few artifacts such as glasses performed better. 2) Particularly rare diseases, or recently described disorders, for which the classifier's representation is based on a smaller training set, show a lower performance, even if experienced dysmorphologists would consider them highly distinguishable.[24,34] 3) Molecular pathway diseases, modeled as a single class, can be biased towards the prevailing gene if there is substructure in the phenotypic series, meaning there actually are gene-specific differences in the gestalt and complete heterogeneity is simply an approximation.[25] This applies also to microdeletion syndromes that can be caused by single gene mutations, such as Smith-Magenis syndrome, or any clinical presentation of a phenotype that is considered atypical.

The only way to overcome the biases of semantic similarity metrics as well as AI-driven image analysis that are due to limited cohort sizes, is sharing of the phenotypic data sets.

In conclusion, the PEDIA study documents that exome variant interpretation benefits from computer-assisted image analysis of facial photographs, particularly if dysmorphism has been stated in the clinical notes. By including similarity scores from DeepGestalt, we improved the top-10-accuracy rate considerably. AI-driven pattern recognition of frontal facial patient photographs is an example of next-generation phenotyping technology with proven clinical value in the interpretation of next-generation sequencing data.

As deep-learning advances in the assessment of other medical imaging data, it will be interesting to study how these classifiers affect variant interpretation separately and in aggregate.[35,36]

Data and Code Availability

PEDIA is freely available for academic use at https://pedia-study.org and the source code is available at https://github.com/PEDIA-Charite.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (KR 3985/7-3, KR 3985/6-1)

Author contributions

Conceived and designed the study and drafted the manuscript: T.C.H., M.A.M., J.T.P., and P.M.K.

Project Coordination: M.A.M., J.T.P., N.F.,

Acquired, analyzed, and interpreted the clinical data: A.D., L.B.S., A.B., S.B., O.B., A.M.C., C.C., M.C, K.C., S.D., M.D., K.D., S.D., S.D., D.D., N.E., C.R.F., B.F.Z., H.G., K.G., Y.G., N.H., N.H., L.M.H., K.H., I.I., J.H., A.K., C.K., H.K., S.K., A.K., A.K., U.K., A.L., M.L., G.L., E.M., D.M., A.O., K.P., B.P., L.P., P.M.R, N.R., S.R., A.R.R., S.R., G.R., U.S., A.S., P.S.Y., E.K.S., M.S., Y.S., C.T., G.T., A.V., I.V., D.W., I.W., K.W., M.W., B.W., M.W.Y., L.G.N., C.E.O.

Chief clinical data review: D.H.

Performed the Bioinformatics and statistical analysis: M.S., J.H., M.A.M., T.C.H., T.K., S.K., M.Z., N.Z., O.B., G.N., Y.G., Y.H., O.S., H.D.E., J.T.P., S.G.K.

Critically revised the manuscript for important intellectual content. H.B.B., P.N.R., S.M., J.Z.,

Competing interests

N.F., H.D.E., Y.G., G.N., O.B., Y.H., are employees of FDNA Inc, T.K. is employee of GeneTalk GmbH.

References

1. Institute of Medicine, Board on Health Sciences Policy, Committee on Accelerating Rare Diseases Research and Orphan Product Development. Rare Diseases and Orphan Products: Accelerating Research and Development. National Academies Press; 2011.
2. Evans WR, Rafi I. Rare diseases in general practice: recognising the zebras among the horses. Br J Gen Pract 2016;66(652):550–1.
3. Schieppati A, Henter J-I, Daina E, Aperia A. Why rare diseases are an important medical and social issue. Lancet 2008;371(9629):2039–41.
4. de Ligt J, Willemsen MH, van Bon BWM, et al. Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med 2012;367(20):1921–9.
5. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nat Rev Genet 2016;17(1):9–18.
6. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. Nature 2017;542(7642):433–8.
7. Hu H, Kahrizi K, Musante L, et al. Genetics of intellectual disability in consanguineous families. Mol Psychiatry [Internet] 2018;Available from: http://dx.doi.org/10.1038/s41380-017-0012-2
8. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) [Internet]. OMIM. [cited 2018];Available from: https://omim.org/
9. Stark Z, Schofield D, Alam K, et al. Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing with that of usual care overwhelmingly supports early use and reimbursement. Genet Med 2017;19(8):867–74.
10. Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. N Engl J Med 2014;370(25):2418–25.
11. Stark Z, Dashnow H, Lunke S, et al. A clinically driven variant prioritization framework outperforms purely computational approaches for the diagnostic analysis of singleton WES data. Eur J Hum Genet 2017;25(11):1268–72.
12. Pengelly RJ, Alom T, Zhang Z, Hunt D, Ennis S, Collins A. Evaluating phenotype-driven approaches for genetic diagnoses from exomes in a clinical setting. Sci Rep 2017;7(1):13509.
13. Zemojtel T, Köhler S, Mackenroth L, et al. Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. Sci Transl Med 2014;6(252):252ra123.
14. Smedley D, Jacobsen JOB, Jäger M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc 2015;10(12):2004–15.
15. Singleton MV, Guthery SL, Voelkerding KV, et al. Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. Am J Hum Genet 2014;94(4):599–610.
16. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet 2008;83(5):610–5.
17. Loos HS, Wieczorek D, Würtz RP, von der Malsburg C, Horsthemke B. Computer-based recognition of dysmorphic faces. Eur J Hum Genet 2003;11(8):555–60.
18. Boehringer S, Vollmar T, Tasse C, et al. Syndrome identification based on 2D analysis software. Eur J Hum Genet 2006;14(10):1082–9.
19. Boehringer S, Guenther M, Sinigerova S, Wurtz RP, Horsthemke B, Wieczorek D. Automated syndrome detection in a set of clinical facial photographs. Am J Med Genet A 2011;155(9):2161–9.
20. Ferry Q, Steinberg J, Webber C, et al. Diagnostically relevant facial gestalt information from ordinary photos. Elife 2014;3:e02020.
21. Gurovich Y, Hanani Y, Bar O, et al. DeepGestalt - Identifying Rare Genetic Syndromes Using Deep Learning [Internet]. arXiv [cs.CV]. 2018;Available from: http://arxiv.org/abs/1801.07637

22. Basel-Vanagaite L, Wolf L, Orin M, et al. Recognition of the Cornelia de Lange syndrome phenotype with facial dysmorphology novel analysis. Clin Genet 2016;89(5):557–63.
23. Liehr T, Acquarola N, Pyle K, et al. Next generation phenotyping in Emanuel and Pallister-Killian syndrome using computer-aided facial dysmorphology analysis of 2D photos. Clin Genet 2018;93(2):378–81.
24. Pantel JT, Zhao M, Mensah MA, et al. Advances in computer-assisted syndrome recognition and differentiation in a set of metabolic disorders [Internet]. 2017;Available from: http://dx.doi.org/10.1101/219394
25. Knaus A, Pantel JT, Pendziwiat M, et al. Characterization of glycosylphosphatidylinositol biosynthesis defects by clinical features, flow cytometry, and automated image analysis. Genome Med 2018;10(1):3.
26. Hennekam RCM, Biesecker LG. Next-generation sequencing demands next-generation phenotyping. Hum Mutat 2012;33(5):884–6.
27. Richards, S. et al. 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. Genet. Med. 17, 405–424 (2015).
28. Tavtigian SV, Greenblatt MS, Harrison SM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genet Med [Internet] 2018;Available from: http://dx.doi.org/10.1038/gim.2017.210
29. Köhler S, Schulz MH, Krawitz P, et al. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. Am J Hum Genet 2009;85(4):457–64.
30. Bauer S, Köhler S, Schulz MH, Robinson PN. Bayesian ontology querying for accurate and noise-tolerant semantic searches. Bioinformatics 2012;28(19):2502–8.
31. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46(3):310–5.
32. Wright CF, McRae JF, Clayton S, et al. Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. Genet Med [Internet] 2018;Available from: http://dx.doi.org/10.1038/gim.2017.246
33. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 2015;526:68.
34. Dudding-Byth T, Baxter A, Holliday EG, et al. Computer face-matching technology using two-dimensional photographs accurately matches the facial gestalt of unrelated individuals with the same syndromic form of intellectual disability. BMC Biotechnol 2017;17:90.
35. Webb S. Deep learning for biology. Nature 2018;554(7693):555–7.
36. Shen D, Wu G, Suk H-I. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng 2017;19:221–48.

Materials and Methods

Patients

We compiled a cohort of 679 patients with a Mendelian disorder to evaluate Prioritization of Exome Data by Image Analysis (PEDIA). For all cases in this cohort frontal facial photographs were available for analysis and clinical features were documented in HPO terminology.[16] The diagnoses of all individuals have previously been confirmed molecularly and are suitable for analysis by exome sequencing. In total, the cohort covers 105 different monogenic syndromes that are linked to 181 different genes. Of the individuals in this cohort, 375 were published and 309 have not been previously reported (see Supplementary Appendix).

The study was approved by the ethics committees of the Charité - Universitätsmedizin Berlin and of the University of Bonn Medical Center. Written informed consent was provided by the patients or their guardians.

In addition to PEDIA data set, we analyzed a subset of the DeepGestalt study comprising all 260 cases of the publication set with monogenic syndromes which diagnosable by exome sequencing.[21]

Data Preparation

The facial images were analyzed with DeepGestalt (FDNA), a deep convolutional neural network that was trained on more than 17,000 patient images.[21] The results of this analysis are gestalt scores quantifying the similarity to 216 different rare phenotypes per individual. Although DeepGestalt is built as a framework that aims to learn from every additional case, we excluded all data of the PEDIA cohort from the model for benchmarking purposes in a similar manner as described in the original publication.

In addition to the image analysis, we performed semantic similarity searches with the annotated HPO terms by Feature Match (FDNA), Phenomizer and BOQA.[29,30]

We filtered all sequence variants as described by Wright et al. and scored the remaining mutations for deleteriousness with CADD.[31,32] If no exome data was available, we spiked the disease-causing mutation into the exome data of a healthy individual from the 1000 Genomes Project.[33] This exome simulation was applied to the entire PEDIA cohort to assess the influence of the genetic background on the performance of our scoring approach.

For the variants remaining after filtering, we derived the similarity scores from image analysis and semantic similarity searches that were based on HPO feature annotations for the syndromes associated with the respective genes. If there were several syndromes linked to a single gene, the highest gestalt and feature scores were selected. Case data is represented as table with a variable number of lines representing genes and five columns for the different scores (Figure 1 B). All five scores with per line as well as the Boolean label disease gene “true” or “false” were used to train a classifier that yields a single value per gene, the PEDIA score, that can be used for prioritization (Figure 1 C). A detailed description of preprocessing and filtering, as well as all the annotated data, can be found in our code repository.

Gene prioritization

We used a support vector machine (SVM) to prioritize the disease-causing gene in each patient. First, we split the PEDIA cohort into a training and a test set. We used a linear kernel on the five scores to train the SVM and selected the hyperparameter C in the range from 2^{-6} to 2^{12} by performing internal 5-fold cross-validation on the training set. The C with highest top-1 accuracy was selected for training linear SVM. We further benchmarked the performance of each case in the test set with this model. The distance of each gene to the hyperplane - defined as the PEDIA score - was used to rank the genes for the case. If the disease-causing gene was at the first position, we called it a top 1 match, or if it was amongst the first ten genes, we called it a top 10 match.

To evaluate the accuracy, we conducted a 10-fold cross-validation, that is, we split 679 cases into 10 groups to minimize overfitting. For the 260 cases from the DeepGestalt publication test set, where exome diagnostics would be applicable, we randomly selected a patient from the PEDIA cohort with

the same diagnosis and replaced the entire gestalt scores per case. Thus, we were also able to analyze the influence of another large collection of patient images in the exome prioritization. In total, all experiments were conducted ten times and the achieved top-1 and top-10-accuracies were averaged. All training data as well as the classifier are available at https://github.com/PEDIA-Charite and https://pedia-study.org

Performance evaluation in a classification task

Both, DeepGestalt and PEDIA are approaches to solve multiclass classification problems (MCPs), the first tool operating on phenotypes and the second on genes. The difficulty of the task is characterized by the number of classes and the distinguishability of the different entities. For both MCPs the maximum number of classes can be estimated from Online Mendelian Inheritance in Man catalog, that is currently listing around 500 distinct disorders with facial abnormalities and 700 corresponding genes with disease-causing mutations (Figure 1 A).

Learning a phenotype in a neural network requires a certain number of unrelated cases. By the end of 2017, DeepGestalt could distinguish between 216 different entities. Due to more training data, 60 new disorders were added in the last six months and the number is expected to increase further on.

The performance of a prioritization tool can be assessed by the proportion of cases in a test set for which the correct diagnosis or disease-gene is placed at the first position or amongst the first ten suggestions (top-1 and top-10-accuracy). The composition of the test set has an influence on the accuracy because some disease phenotypes are easier to recognize and some gene mutations are more readily identified as deleterious. The setup of the PEDIA cohort, which is comprehensively documented in the Supplementary Appendix, therefore aims at emulating the whole spectrum of cases that could currently be analyzed with DeepGestalt and diagnosed by exome sequencing.

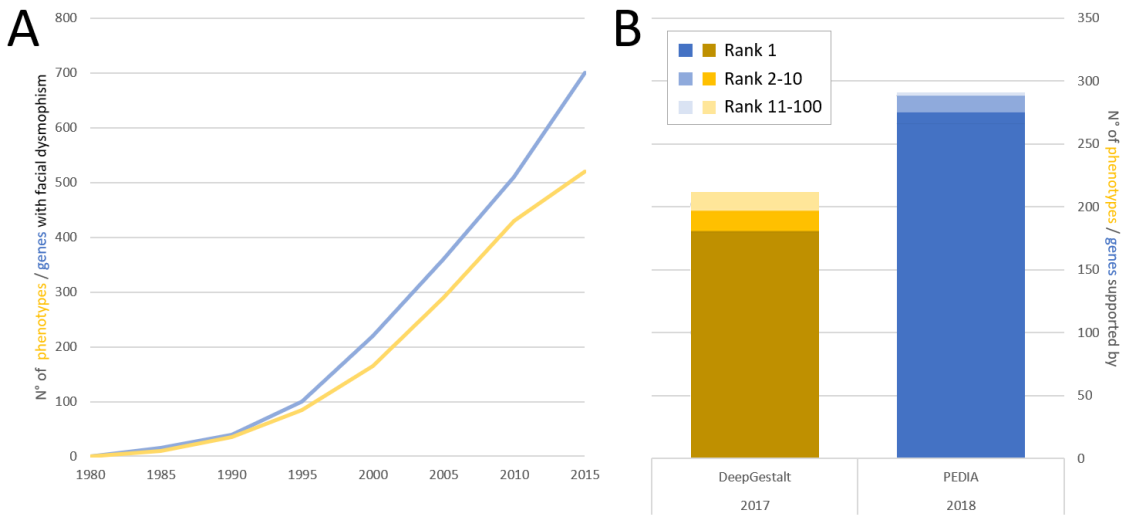

Figure 1: A) Schematic Increase of Mendelian phenotypes with facial abnormalities and associated genes listed in the encyclopedia of Online Mendelian Inheritance in Man over time. B) The next-generation phenotyping tool DeepGestalt could be used to differentiate between 216 disorders in the end of 2017 and achieved a top-10-accuracy rate of 90 %. The subset of Mendelian phenotypes that are suitable for a diagnostic workup by exome sequencing corresponds to 290 genes and in the PEDIA cohort a top-1-accuracy of 98 % was achieved.

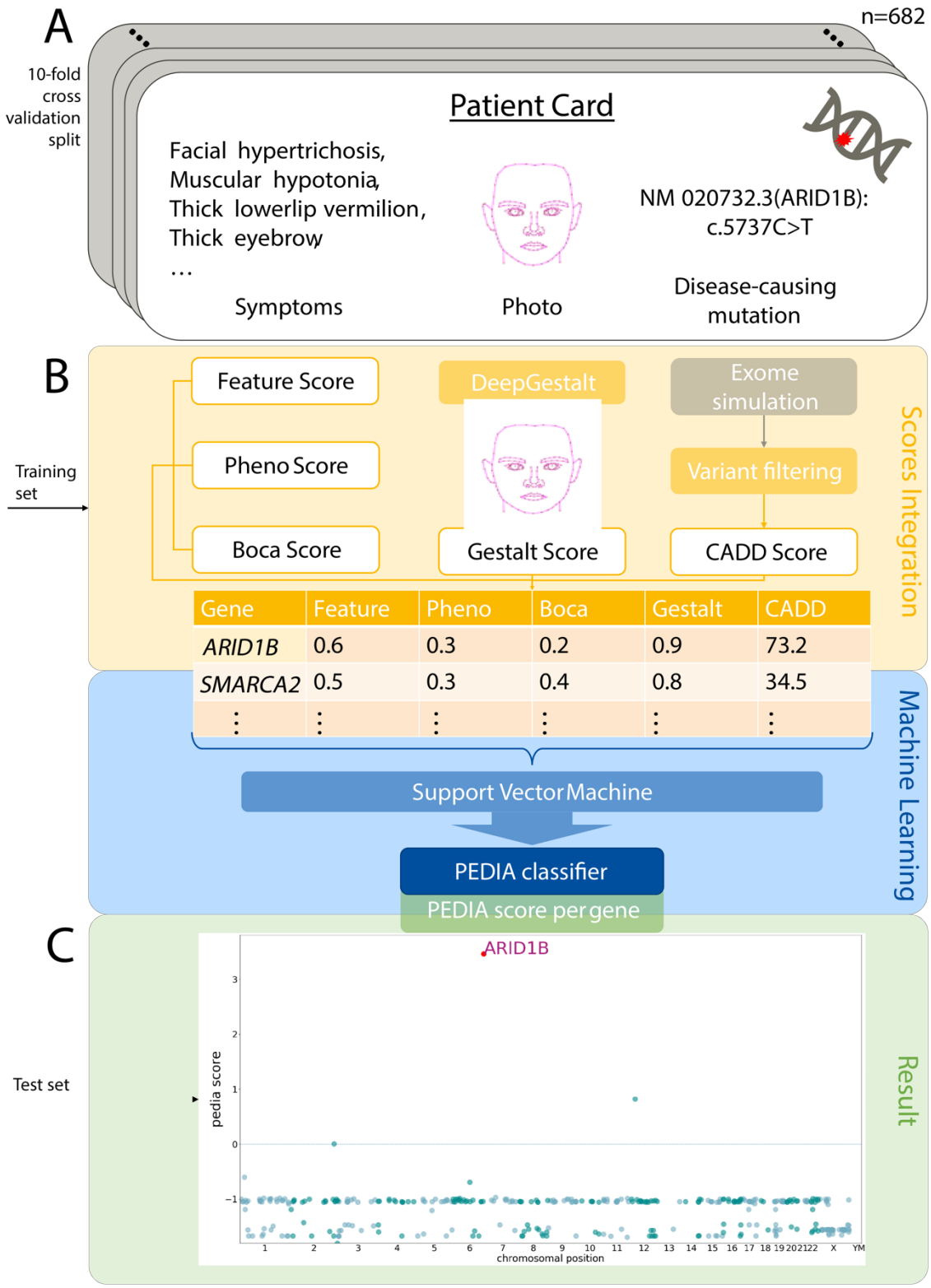

Figure 2: ***Prioritization of Exome Data by Image Analysis.*** *A) Clinical features, facial photograph and disease-causing mutation of one individual of the PEDIA cohort. In total the cohort consists of 679 cases with monogenic disorders that are suitable for a diagnostic workup by exome sequencing. B) Clinical features, images and exome variants were evaluated separately and integrated to a single score by a machine learning approach. C) The disease-causing gene of the case depicted in A achieves the highest PEDIA score and molecularly confirms the diagnosis of Coffin-Siris syndrome. Other genes associated with similar phenotypes such as Nicolaides-Baraitser syndrome, achieved also scores for gestalt but not for variant deleteriousness. This figure has been adapted for bioRxiv by removing the patient photo. The original version with a patient photo is available on request. Also see https://pedia-study.org*

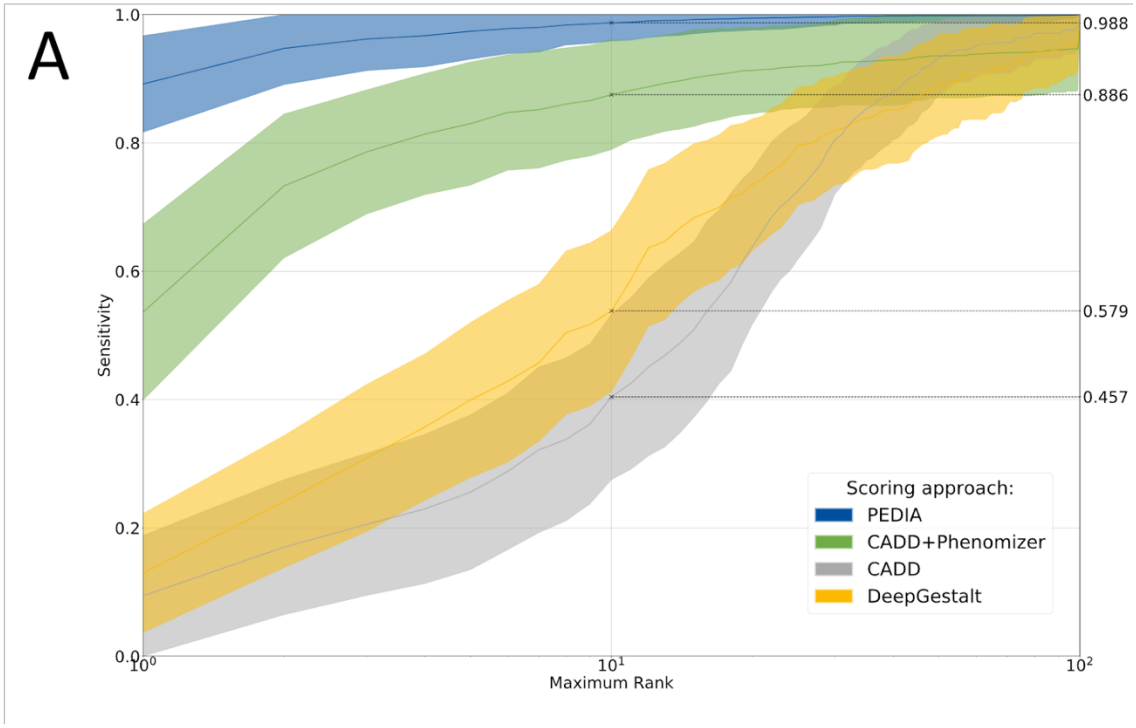

Figure 3: ***Area under the curve for different disease-gene prioritization approaches.*** *For each case the exome variants are ordered according to four different scoring approaches, solely by a molecular deleteriousness score (C), by score from image analysis (DeepGestalt), by a combination of a molecular deleteriousness score and a clinical feature based semantic similarity score (P+C), or the PEDIA score that includes all three levels of evidence. The sensitivity of the prioritization approach depends on the number of genes that are considered in an ordered list. The top 10 accuracy rates of of Figure 1B correspond to the intersect of the curves for PEDIA and DeepGestalt at maximum rank 10^1. Note that for benchmarking DeepGestalt on the gene level, syndrome similarity scores first have to be mapped to the gene level, resulting in a lower performance compared to the readout on a phenotype level, due to heterogeneity. The area under the curve is largest for PEDIA scoring. When e.g. the first ten candidate genes are considered, the syndromic similarity quantified by image analysis increases the sensitivity by about 20 % compared to P+C.*