GenNet framework: interpretable neural networks for phenotype prediction

Arno van Hilten[1], Steven A. Kushner[2], Manfred Kayser[3], M. Arfan Ikram[4], Hieab H.H. Adams[1,5], Caroline C.W. Klaver[4,6], Wiro J. Niessen*[1,7] and Gennady V. Roshchupkin*[1,4]

[1] Department of Radiology and Nuclear Medicine, Erasmus MC, Medical Center, Rotterdam, the Netherlands.
[2] Department of Psychiatry, Erasmus MC, Medical Center, Rotterdam, the Netherlands.
[3] Department of Genetic Identification, Erasmus MC, Medical Center, Rotterdam, the Netherlands.
[4] Department of Epidemiology, Erasmus MC, Medical Center, Rotterdam, the Netherlands.
[5] Department of Clinical Genetics, Erasmus MC, Medical Center, Rotterdam, the Netherlands.
[6] Department of Ophthalmology, Erasmus MC, Medical Center, Rotterdam, the Netherlands.
[7] Faculty of Applied Sciences, TU Delft, Delft, the Netherlands.

*Jointly supervised this project

Neural networks have been seldomly leveraged in population genomics due to the computational burden and challenge of interpretability. Here, we propose GenNet, a novel open-source deep learning framework for predicting phenotype from genotype. In this framework, public prior biological knowledge is used to construct interpretable and memory-efficient neural network architectures. These architectures obtain good predictive performance for multiple traits and complex diseases, opening the door for neural networks in population genomics.

Introduction

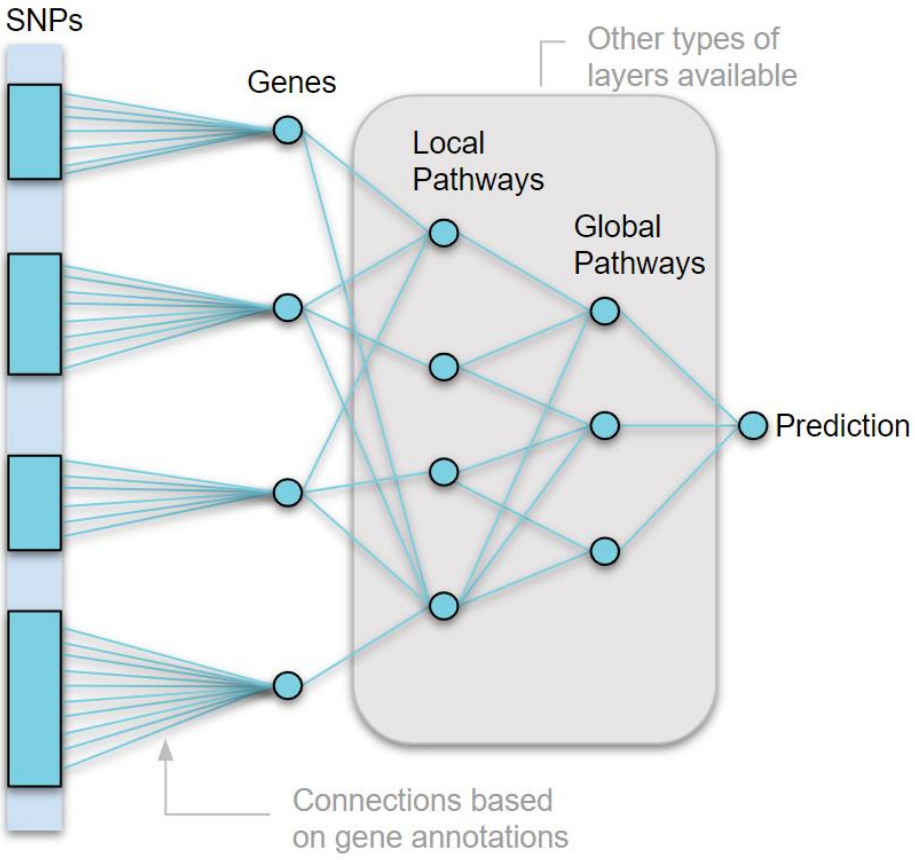
Genome-wide association studies (GWAS) have identified numerous genomic loci associated with complex (polygenic) human traits and diseases. Recent GWAS studies with increasingly larger sample sizes are leading to more significant associations between genotypes and

phenotypes at more and more independent loci. To illustrate, the latest GWAS for body height based on 700,000 individuals identified more than 3000 near-independent significantly associated single nucleotide polymorphisms (SNPs)[1]. This information, used in combination with annotated biological databases such as: NCBI RefSeq, KEGG, Reactome and GTEx has proven to be highly valuable for understanding the underlying biological mechanisms of complex diseases[2-6]. In this paper, we propose a new framework, GenNet, that integrates these biological data sources for discovery and interpretability in an end-to-end deep learning framework for predicting phenotypes.

Deep learning is the state of the art in many domains such as medical image analysis and natural language processing because of its flexibility and modeling capabilities[7,8]. In many cases, deep learning yields better performance compared to traditional approaches, since it can model highly non-linear relations and scales very well with data size. However, this often comes at the cost of interpretability, since there is a trade-off between complexity and interpretability[9,10].

Additionally, when it comes to genotype data, the number of learnable parameters increases dramatically because of the large input size, making it infeasible to use classical neural networks in this domain. To overcome previous limitations, we propose a new framework, GenNet, in which different types of biological information are used to define biologically plausible neural network architectures, avoiding this trade-off and creating interpretable neural networks for

48 predicting complex phenotypes.



49

50 *Figure 1. Overview of the GenNet Framework. Neural networks are made by combining layers*
51 *made of different sources of prior biological knowledge (i.e. a gene layer from gene annotations,*
52 *a pathway layer from KEGG pathway annotations). These sources define the connections and*
53 *therefore the architecture, creating interpretable networks in the process.*

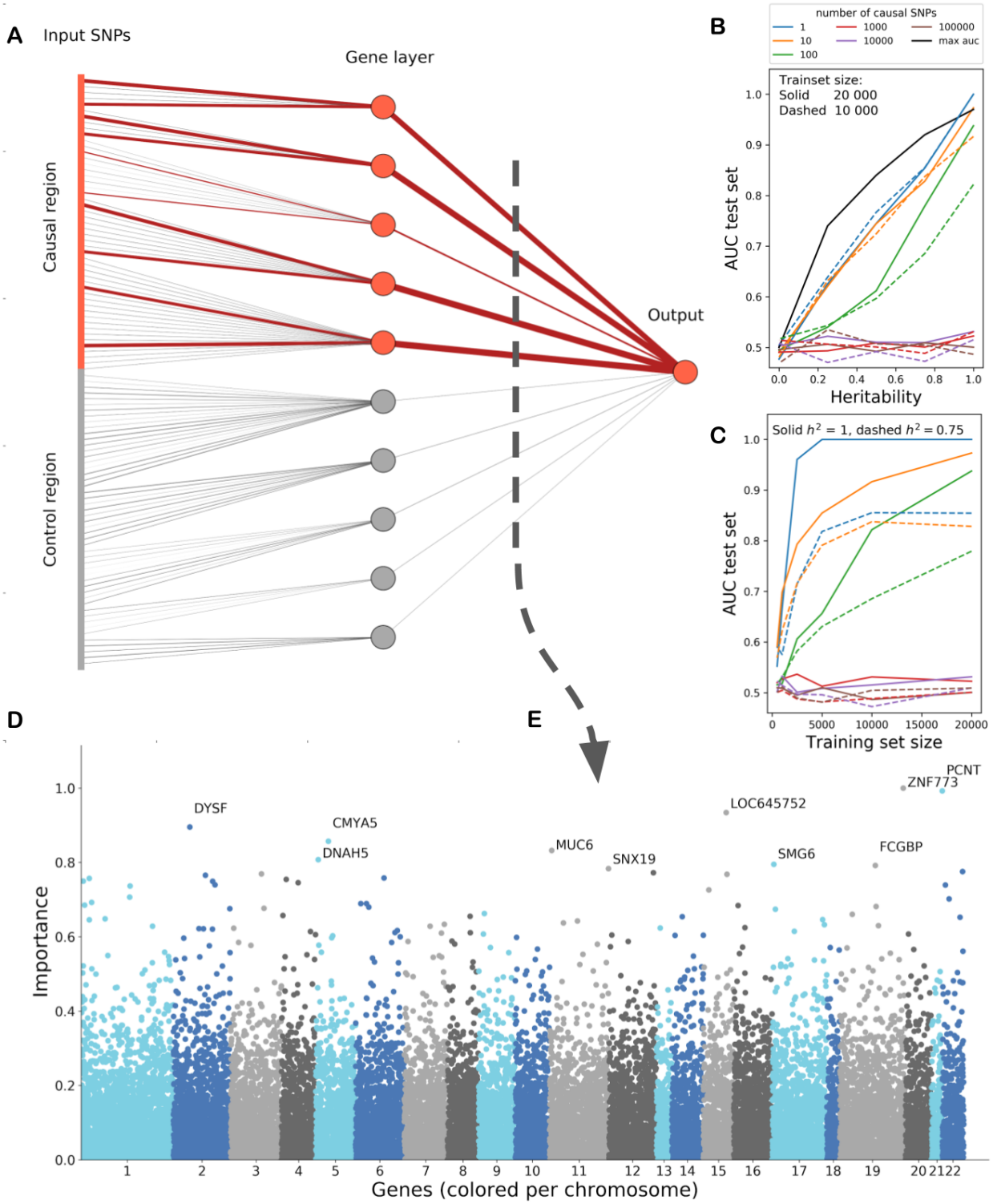
54

55 **Methods**

56 The main concept of the GenNet framework is summarized graphically in Figure 1. In this
57 framework, prior knowledge is used to create groups of connected nodes to reduce the number of
58 learnable parameters in comparison to a fully connected neural network. For example, in the first
59 layer, where biological knowledge in the form of gene annotations, is used to group millions of

single nucleotide polymorphisms (SNPs) and to connect those SNPs to their corresponding genes. The resulting layer retains only meaningful connections, significantly reducing the total number of parameters compared to a classical layer. As a result, these memory-efficient networks are able to handle the millions of inputs needed for genotype-to-phenotype prediction. The biological knowledge is thus used to define only meaningful connections, shaping the architecture of the neural network. Interpretability is inherent to the neural network's architecture. For example, a network that connects SNPs-to-genes and genes-to-output. The learned weights of the connections between layers represent the effect of the SNP on the gene or the effect of the gene on the output. In the network, all neurons represent thus biological entities and weights model the effects between these entities, together forming a biologically interpretable neural network. Each connection in the network is thus based on a biological annotation and the learned weight for this connection represents the importance of this annotation for the predicted outcome.

Many types of layers can be created using this principle. These layers can be used like building blocks to form new network architectures. Apart from gene annotations, our framework provides layers built from exon annotations, pathway annotations, chromosome annotations and cell and tissue type expressions.



77

78 *Figure 2.* ***A)*** *Non-linear simulation showing the basic principle of the network, thickness of the*

79 *connections represents the learned weight (causal, contributing connections in red, control*

connections in grey). This proof of concept can be run online see https://tinyurl.com/y8hh8rul (or see Supplementary 1.1). ***B)*** *A secondary set of simulations show the performance of GenNet, expressed in the area under the curve, for increasing levels of heritability and training set size* ***(C)****. In black the theoretical maximum of the AUC versus heritability*[11]*.* ***D)*** *Manhattan plot of the importance of the genes according to the network for distinguishing between schizophrenia cases and controls.* ***E)*** *This Manhattan plot is a cross section between the gene layer and the outcome of the trained network.*

Results

In order to evaluate the network's performance under a variety of conditions, synthetic data was created with different levels of heritability, number of training samples and polygenicity (see Supplementary Materials 1 for results and detailed description). Figure 2A shows the proof of concept demo that can be run online. Figure 2B and 2C show the main trends in the simulations. As would be expected, the network performs best for traits with high heritability, high number of training samples and low polygenicity, and the performance decreases with decreasing heritability, number of training samples or increase of polygenicity.

Motivated by the proof of concept and the outcomes of the simulations, the framework was applied to real data from multiple sources, including population-based data from the UK Biobank study and the Rotterdam study, and a case-controls study on schizophrenia from Sweden[12-14]. The analyzed phenotypes vary from traits where high predictive performance can be obtained from a dozen of variants (eye color) to disorders where thousands of variants only explain a small portion of the variance (schizophrenia and bipolar disorder)[15,16]. The genotype data employed included imputed microarray-based GWAS data (eye color in Rotterdam study) as

well as whole exome sequencing (WES) data (hair color, male baldness pattern and bipolar disorder in UK Biobank; schizophrenia in the Swedish study). An overview of the experiments and results can be found in table 1.

Trait	**Dataset (type)**	**Number of input variants**	**Subjects & phenotype**		**Heritability**	**AUC LASSO**	**AUC GenNet**	**GenNet: top three most important genes**
			Class I	**Class II**				
Eye color	Rotterdam (genotype array)	113,241 (exonic) inputs of 16,628 genes	4041 Blue	2250 Other	0.80-0.98	0.68	0.75	*HERC2, OCA2, LAMC1*
Hair color	UK Biobank (exome)	6,986,636 input variants of 15,827 genes	1648 Blond	1656 Red	0.70-0.97	0.78	0.83	*MC1R*, OCA2, TC2N*
			1672 Dark brown	1664 Red	0.70-0.97	0.79	0.88	*MC1R*, OCA2, ZCCHC4*
			4352 Blond	4343 Dark Brown	0.70-0.97	0.64	0.75	*OCA2, TC2N, EXOC2*
Male baldness	UK Biobank (exome)	6,986,636 Input variants of 15,827 genes.	3454 No balding	3454 Severe balding	0.60-0.70	0.57	0.57	*NGEF, NKRD18B, SYNJ2*
Bipolar	UK Biobank (exome)	6,986,636 Input variants of 15,827 genes	343 Cases	347 Controls	0.73-0.93	0.59	0.60	*LINC00266-1, CSMD1, TCERG1L*
Schizophrenia	Sweden (exome)	1,288,701 input variants of 21,390 genes	4969 Cases	6245 Controls	0.80-0.85	0.65	0.74	*ZNF773, PCNT, DYSF*

*Table 1. Summary of the experiments and results in this study for the simplest network in our framework that contains the input SNPs, the gene layer and the output layer. Manhattan plots for gene importance can be found in Supplementary Materials 2,3 & 4. *MC1R was not present in gene annotations but was identified by linkage disequilibrium.*

In general, the framework's predictive performance is in line with trends seen in simulations and literature. Phenotypes with more training samples and phenotypes that require less variants to obtain good predictive performance, such as eye and hair color, performed best. Nonetheless, a good predictive performance, area under curve (AUC) of 0.74 in the held-out test set, was

obtained for schizophrenia, a highly polygenic disorder. All models outperform or match the baseline LASSO logistic regression model (see Methods).

Inspecting the networks, we found that the *OCA2* gene was highlighted as the most important gene to distinguish between brown and blond hair color. *OCA2* is involved in the transport of tyrosine, a precursor of melanin[17]. The signal is probably amplified by the nearby *HERC2*, previous identified via functional genetic studies as harboring a strong, long-distance enhancer regulating *OCA2* gene expression to cause pigmentation variation[17]. *OCA2* and *HERC* are the two most predictive genes according to the network for predicting blue (iris) eye color. Both have been earlier identified by hair and eye color GWASes[18-20].

In the experiments with schizophrenia as outcome, the network was able to classify cases and controls with a maximum accuracy of 68.4% (mean of 66.3 ± 1.37 over 10 runs). We estimate the theoretical upper limit for classification, including all genetic aspects, to be an accuracy of 72% (supplementary methods 5). The model obtains an area under the receiver operating curve of 0.74 (ranging 0.72-0.74) in the held-out test set, thereby considerably outperforming the LASSO logistic regression baseline (AUC of 0.64). The GenNet AUC compares favorably to polygenic risk scoring for schizophrenia, which have AUC values on the order of 0.70 (ranging 0.49-0.85)[15] . This is noteworthy since in this study the schizophrenia predictions are based on whole exome sequencing data as opposed to GWAS arrays spanning the whole genome.

Discussion

Here, we present a novel framework to train interpretable neural networks for phenotype prediction from genotype. The proposed neural networks have connections defined by prior biological knowledge only, reducing the number of connections and therefore the number of

trainable parameters. Consequently, the networks are interpretable and overcome computational limitations. All experiments were run on a single GPU (Nvidia GeForce GTX 1080) and converged within 48 hours. Simulations show the network's performance when varying the degree of heritability, polygenicity and sample size. The suggested sample size, heritability and polygenicity are conservative. When applying the framework to UK Biobank, Rotterdam study and Swedish Schizophrenia WES data, good predictive performance was achieved with smaller sample sizes and higher number of inputs than suggested by the simulations. In these experiments, widely different phenotypes are predicted with generally good performance based exclusively on WES SNPs.

For traits with a known etiology, well-replicated genes such as *HERC2* and *OCA2* for eye color and *OCA2* and *TCN2* for hair color were found to be important by the network[18-21]. For schizophrenia, a disorder with an unclear etiology, the network identified previously unimplicated genes, including *ZNF773* and *PCNT*. However, it is important to note that the importance captured by the network bears more similarity to effect size rather than statistical significance.

In general, these experiments indicate that neural networks in our framework can overcome computational limitations while still obtaining good predictive performance, opening the door for genetic risk prediction by neural networks. Aside from computational benefits, the architecture offers interpretability, alleviating one the most important shortcomings of neural networks. In this study, WES data and exonic variants from microarrays have been used, however the principles in the GenNet framework can be leveraged to handle diverse types of input, including genotype, gene expression and methylation data or combinations thereof)

In conclusion, we developed a freely-available framework, which can be used to build interpretable neural networks for genotype data by incorporating prior biological knowledge. We have demonstrated the effectiveness of this novel framework across multiple datasets and for multiple phenotypes. Given that each network node is interpretable, we anticipate this approach to have wide applicability for uncovering novel insights into the genetic architecture of complex traits and diseases.

GenNet is an open-source framework, providing code and tutorials (https://github.com/arnovanhilten/GenNet/). This includes tutorials for applying the networks as well as creating new layers and networks from prior knowledge.

Online Methods

Sweden Schizophrenia

Sweden-Schizophrenia Population-Based Case-Control Exome Sequencing study (dbGaP phs000473.v2.p2), is a case control study with 4969 cases and 6245 controls[14]. All individuals aged 18-65, have parents born in Sweden and provided written informed consent. Cases were at least 2 times hospitalized with schizophrenia discharge diagnosis and do not have a hospital register diagnosis consistent with a medical or other psychiatric disorder that mitigates the schizophrenia diagnosis. Cases do not have a relationship closer than 2nd degree relative with any other case. Controls do not have any relation to either case or control and all controls have never been hospitalized with a discharge diagnosis of schizophrenia.

The .bim, .bam and .bed files were converted using HASE[22] to .h5 format, a format that allows fast and parallel data reading. After conversion, the data is transposed and SNPs without any variance are removed (~1.2 million SNPs remain). The data is split in a training, validation and test set (ratio of 60/20/20), while preserving the ratio cases and controls. All SNPs with standard deviation greater than zero are used as input to the network after z-score normalization (based on the mean and standard deviation of the training set).

UK Biobank

We applied the framework to multiple phenotypes in the UK Biobank using the first release of the WES data, providing whole exome sequencing for 50,000 UK Biobank participants[23]. Phenotypes are self-reported (touchscreen questions UK Biobank Assessment Centre). Similar to the Sweden cohort all variants without variance were removed, data was converted to hierarchical data format (.h5), and transposed. For every phenotype an equal number of cases

and controls were sampled. The resulting dataset is split in a train, validation and a test set (ratio of 60/20/20). Related cases, and cases with related controls, (kinship > 0.0625) are all in the training set. This is done under the assumption that related cases and controls could ease training, the shared genetic information could steer the network towards the discriminatory features. The validation and test sets contain only unrelated cases and controls within and between sets. Unrelated controls are randomly sampled and added to gain an even distribution between cases and controls in all sets. Misaligned SNPs and sex chromosomes were masked in the first layer and therefore not included in the study.

Rotterdam Study

The Rotterdam study is a prospective cohort study, in the first cohort 6291 participants were genotyped using the Illumina 550K and 550K duo arrays. Samples with low call rate (<97.5%), with excess autosomal heterozygosity (>0.336) or with sex-mismatch were excluded, as were outliers identified by the identity-by-state clustering analysis (outliers were defined as being >3 standard deviation (SD) from population mean or having identity-by-state probabilities >97%). For imputation the Markov Chain Haplotyping (MACH) package version 1.0 software (Imputed to plus strand of NCBI build 37, 1000 Genomes phase I version 3) and minimac version 2012.8.6 were used (call rate >98%, MAF >0.001 and Hardy–Weinberg equilibrium P-value > 10^{-6}). From here on processing steps are identical as described for Sweden Schizophrenia obtaining 113,241 exonic variants.

Eyes were examined by an ophthalmological medical researcher and eye (iris) color was categorized into three categories; blue, intermediate and brown using standard images and based on the predominant color and pigmentation [24].

Prior Knowledge

All SNPs were annotated using Annovar[25]. A sparse connectivity matrix is generated connecting the SNPs to their corresponding genes. The intron-exon annotations of Annovar were used as a proxy to create the connectivity matrices between SNPs, exons and genes. The pathway masks were built by using GeneSCF[26] and the KEGG database[3]. GTEx tissue-expression masks were made using the fully processed, filtered and normalized gene expression matrices for each tissue directly obtainable from the GTEx website[5] and from derived t-score statistics[27]. Single cell expression masks are available made using data from FUMA[28]. Expression masks are continuous and various thresholds can be used to create connectivity matrices, the threshold should be chosen with care to ensure unique nodes and thus interpretability.

All available networks in the framework are trait independent but trait specific neural networks can be created with more specific prior knowledge (e.g. brain cell expression to select genes in the network for predicting neurological disorders). The GenNet framework is quite flexible, any information that groups data uniquely can be used to create layers.

Neural Network Architecture

In the GenNet framework, layers are available built from biological knowledge such as; exon annotations, gene annotations, pathway annotations, cell expression and tissue expression. Information from these resources are used to define only meaningful connections, shaping an interpretable and lightweight neural network, allowing evaluation of millions of input variants together. These networks bear similarities to the first generation of neural networks and recently interest for these networks has rekindled for biological applications[29-32], In neural networks the

dimensionality of the data is reduced, resulting in a loss of information every layer. Networks with the GenNet architecture aim to reduce this loss by using prior biological knowledge to create layers in the network, compressing the network in a biologically sensible way. These sources are used to define connections between nodes/neurons in the network (see implementation). The network itself will learn how important the connections are for the prediction of the outcome. Thus, giving additional information for the sources used to define the network. For example, in most networks, gene annotations are used to group SNPs. During training the network will figure out which gene is important and which SNPs in the gene are important for predicting the outcome.

Interpretation

Interpretation of the network is straightforward due to the simplicity of the concept, the higher the weight is the more important it is for the network. A network built by gene annotations can be seen as ~20 000 (number of genes) parallel regressions followed by a single logistic regression. The learned weights in these regressions are similar to the coefficients in logistic regression. Especially the last node, a single neuron with a sigmoid activation, $P = Sigmoid(\sum_{i=0}^{n} x_i w_i + B)$ is similar to logistic regression $Y = Sigmoid(\sum_{i=0}^{n} x_i \beta_i + B)$. To compare coefficients (β) the inputs are normalized in logistic regression. In the neural network this is achieved by batch normalization (without center and scaling), normalizing the weights (w) after every activation. Since batch normalization is a batch-wise approximation the learned weights can be multiplied with the standard deviation of the activations for a more accurate estimate, resulting in the importance measure. We noticed it might be beneficial to add an L1 penalty on the weights in the last dense layer, as used in LASSO logistic regression. This

regularizer constraints the search space which might lead to a better generalization and reduced noise in the signal for easier interpretation.

Implementation

Technically, the computational performance of the implemented Keras/Tensorfow[33,34] layer should be on par or an improvement over similar layers. It is implemented using sparse matrix multiplication, making it faster than the slice-wise locallyconnected1D layer and more memory efficient than dense matrix multiplication. The layer is friendly to use, with only one extra input compared to a normal dense layer. This extra input, the sparse connectivity matrix, is made with prior knowledge and describes how neurons are connected between layers.

The networks behave similar to normal fully connected artificial neural networks but is pruned by removing irrelevant connection ($Out = Activation(\sum_{i=0}^{n} x_i w_i + b)$) With w as a sparse matrix with learnable weights, initialized with a sparse connectivity matrix defining connections. The networks are optimized using the ADAM or Adadelta optimizers [35,36], using weighted binary cross entropy with weights depending on the imbalance of the classes, and are all trained on a single GPU.

Baseline

As a baseline method, LASSO logistic regression was implemented in Tensorflow by using a dense layer of a single neuron with a sigmoid activation function and L1 regularization on weights.

Upper Bound

Population characteristics can be used to calculate the upper bound of performance for a classifier for any trait. This can be done by creating a confusion matrix. The accuracy between true and false positives for a perfect classifier, based solely on genetic inputs, is given by the concordance rate between monozygotic twins. It is impossible to predict better based solely on genetic code than the rate a trait occurs in people with virtually the same genetic code. The chance of misclassifying a control should be better than the prevalence, which is often close to zero for most diseases. Creating a confusion matrix can give insights in the upper bound for accuracy, sensitivity and specificity in the dataset. An example for schizophrenia in our dataset can be found in Supplementary 5.

Code Availability

The code, tutorials and trained networks can be found on github.com/arnovanhilten/GenNet/ in the form of Jupyter notebooks. The code has been made with an emphasis on easy to use, with comments and tutorials.

Data Availability

Code to run and generate data for the simulations are publicly available on GitHub. The genetic and phenotypic UK Biobank data are available upon application to the UK Biobank (https://www.ukbiobank.ac.uk/). Access to the Sweden-Schizophrenia Exome Sequencing study can be requested on DBGaP (https://www.ncbi.nlm.nih.gov/gap/) (dbGaP phs000473.v2.p2).

References

1. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in~ 700000 individuals of European ancestry. *Human Molecular Genetics.* 2018;27(20):3641-3649.
2. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research.* 2007;35(suppl_1):D61-D65.
3. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research.* 2000;28(1):27-30.
4. Croft D, Mundo AF, Haw R, et al. The Reactome pathway knowledgebase. *Nucleic Acids Research.* 2014;42(D1):D472-D477.
5. Lonsdale J, Thomas J, Salvatore M, et al. The genotype-tissue expression (GTEx) project. *Nature Genetics.* 2013;45(6):580.
6. Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function. *The American Journal of Human Genetics.* 2018;102(5):717-730.
7. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Medical Image Analysis.* 2017;42:60-88.
8. Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing. *ieee Computational Intelligence Magazine.* 2018;13(3):55-75.
9. Edwards L, Veale M. Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke L & Tech Rev.* 2017;16:18.
10. Došilović FK, Brčić M, Hlupić N. Explainable artificial intelligence: A survey. Paper presented at: 2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)2018.
11. Wray NR, Yang J, Goddard ME, Visscher PM. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS genetics.* 2010;6(2).
12. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203-209.
13. Ikram MA, Brusselle G, Ghanbari M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. *European Journal of Epidemiology.* 2020:1-35.
14. Genovese G, Fromer M, Stahl EA, et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nature neuroscience.* 2016;19(11):1433.
15. Ripke S, Neale BM, Corvin A, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014;511(7510):421-427.
16. Stahl EA, Breen G, Forstner AJ, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics.* 2019;51(5):793-803.
17. Visser M, Kayser M, Palstra R-J. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Research.* 2012;22(3):446-455.
18. Han J, Kraft P, Nan H, et al. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS genetics.* 2008;4(5).
19. Hysi PG, Valdes AM, Liu F, et al. Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nature Genetics.* 2018;50(5):652-656.
20. Candille SI, Absher DM, Beleza S, et al. Genome-wide association studies of quantitatively measured skin, hair, and eye pigmentation in four European populations. *PloS One.* 2012;7(10).
21. Liu F, van Duijn K, Vingerling JR, et al. Eye color and the prediction of complex phenotypes from genotypes. *Current Biology.* 2009;19(5):R192-R193.
22. Roshchupkin GV, Adams H, Vernooij MW, et al. HASE: Framework for efficient high-dimensional association analyses. *Scientific Reports.* 2016;6:36076.
23. Van Hout CV, Tachmazidou I, Backman JD, et al. Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *BioRxiv.* 2019:572347.
24. Kayser M, Liu F, Janssens ACJ, et al. Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *The American Journal of Human Genetics.* 2008;82(2):411-423.
25. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research.* 2010;38(16):e164-e164.
26. Subhash S, Kanduri C. GeneSCF: a real-time based functional enrichment tool with support for multiple organisms. *BMC Bioinformatics.* 2016;17(1):365.
27. Finucane HK, Reshef YA, Anttila V, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature genetics.* 2018;50(4):621-629.
28. Watanabe K, Mirkov MU, de Leeuw CA, van den Heuvel MP, Posthuma D. Genetic mapping of cell type specificity for complex traits. *Nature communications.* 2019;10(1):1-13.

29. Gazestani VH, Lewis NE. From genotype to phenotype: Augmenting deep learning with networks and systems biology. *Current Opinion in Systems Biology.* 2019.
30. Ma J, Yu MK, Fong S, et al. Using deep learning to model the hierarchical structure and function of a cell. *Nature Methods.* 2018;15(4):290.
31. Michael KY, Ma J, Fisher J, Kreisberg JF, Raphael BJ, Ideker T. Visible machine learning for biomedicine. *Cell.* 2018;173(7):1562-1565.
32. Wang D, Liu S, Warrell J, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science.* 2018;362(6420):eaat8464.
33. *Keras* [computer program]. 2015.
34. Abadi M, Barham P, Chen J, et al. Tensorflow: A system for large-scale machine learning. Paper presented at: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)2016.
35. Zeiler MD. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:12125701.* 2012.
36. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980.* 2014.

Acknowledgements

This work was funded by the Dutch Technology Foundation (STW) through the 2005 Simon Steven Meester grant 2015 to W.J. Niessen. This research has been conducted using the UK Biobank Resource under Application Number 23509. This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative (application number 17610). The Rotterdam study is supported by the Netherlands Organization for Scientific Research (NWO, 91203014, 175.010.2005.011, 91103012).

Author contributions

A.H and G.R conceived and designed the method. A.H performed experiments and implemented the method. G.R and W.N supervised the work. M.A.I, C.K, H.A, M.K and S.K. provided or gave access to data. A.H, G.R, W.N, M.K., M.A.I., S.K, wrote, re-vised and approved the manuscript.

Competing interests

W. N. is co-founder and shareholder of Quantib BV. Other authors declare no competing interests.

406 **Additional Information**

407 Supplementary is available for this paper at: https://www.

408 Correspondence and requests should be addressed to A.H or G.R.

409