SNVformer: An Attention-based Deep Neural Network for GWAS Data

Kieran Elmes [1 2] **Diana Benavides-Prado** [3] **Neşet Özkan Tan** [3] **Trung Bao Nguyen** [3] **Nicholas Sumpter** [4] **Megan Leask** [4] **Michael Witbrock** [3] **Alex Gavryushkin** [1]

Abstract

Despite being the widely-used gold standard for linking common genetic variations to phenotypes and disease, genome-wide association studies (GWAS) suffer major limitations, partially attributable to the reliance on simple, typically linear, models of genetic effects. More elaborate methods, such as epistasis-aware models, typically struggle with the scale of GWAS data. In this paper, we build on recent advances in neural networks employing Transformer-based architectures to enable such models at a large scale. As a first step towards replacing linear GWAS with a more expressive approximation, we demonstrate prediction of gout, a painful form of inflammatory arthritis arising when monosodium urate crystals form in the joints under high serum urate conditions, from Single Nucleotide Variants (SNVs) using a scalable (long input) variant of the Transformer architecture. Furthermore, we show that sparse SNVs can be efficiently used by these Transformer-based networks without expanding them to a full genome. By appropriately encoding SNVs, we are able to achieve competitive initial performance, with an AUROC of 83% when classifying a balanced test set using genotype and demographic information. Moreover, the confidence with which the network makes its prediction is a good indication of the prediction accuracy. Our results indicate a number of opportunities for extension, enabling full genome-scale data analysis using more complex and accurate genotype-phenotype association models.

[1]Biological Data Science Lab, School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand [2]Department of Computer Science, University of Otago, Dunedin, New Zealand [3]Strong AI Lab, School of Computer Science, The University of Auckland, Auckland, New Zealand [4]University of Alabama at Birmingham, United States of America. Correspondence to: Alex Gavryushkin <alex@biods.org>.

The 2022 ICML Workshop on Computational Biology. Baltimore, Maryland, USA, 2022. Copyright 2022 by the author(s).

1. Introduction

Genome-wide association studies over large cohorts are the most powerful approach for identifying genotypes associated with phenotypes. As an example, studying urate and gout, GWAS have identified the major urate transporters and transcription factors involved in the kidney (Leask & Merriman, 2021; Boocock et al., 2020). Additionally, estimated effects of causal variants from GWAS can improve prediction of phenotypic outcomes. Thus the availability of large data sets has brought a new era, with the development of predictive models for complex diseases using genotypes and clinical variables. However, with respect to gout and its related traits, only a small handful of studies have attempted to predict the phenotype using genotypes (genetic risk score) and demographics, and with varying accuracy (Tin et al., 2019; Zhang & Lee, 2020; Sun et al., 2022).

Prediction of a trait is limited by two main factors: heritability of the trait and accurate estimation of the underlying genetic effects (Wei et al., 2014). Thus, current methods underlying the generation of GWAS data (e.g. at the UK Biobank (UKBB) (Sudlow et al., 2015)) bring challenges for prediction, i.e. it is typical to opt for simple but biologically unrealistic models, such as the linear model, in data analyses to enable genome-wide screens for single nucleotide variants (SNVs) with significant effects on (disease) phenotypes. As a result, most GWAS assume the independence of individual SNV effects on the phenotype, and therefore cannot model gene-gene interactions (epistasis). Additionally, typical GWAS are done using a simple additive model, that fails to account for potential dominance effects of different genotypes (Zhu et al., 2015; Hivert et al., 2021). They also often simplify complex multi-dimensional phenotypes to a one-dimensional response variable (Lienkaemper et al., 2018). If epistasis contributes to the genetic architecture of a trait, then identifying epistatic variants is important for improving the predictive power of a GWAS (Crona et al., 2017). Current methods for detecting epistatic interactions are limited computationally and statistically, and developing methods to overcome these challenges is difficult (Wei et al., 2014; Elmes et al., 2021b;a).

Deep neural networks, including Transformer-based ones initially designed for natural language processing (NLP)

tasks (Vaswani et al., 2017), have recently been successfully employed to analyse molecular sequence data, the most successful and widely-known example being AlphaFold-2 (Jumper et al., 2021). These approaches have shown great promise in both scalability and potential to support complex models, such as residue-residue interactions in proteins.

Presently, however, such architectures are capable only of handling sequences of modest length, such as those encoding specific proteins. Recently, and in parallel, Transformer-based architectures for NLP have been developed that aim to accommodate extended bodies of text (Wang et al., 2020; Choromanski et al., 2020). Here, we combine these two directions of research and present a Transformer-based deep neural architecture for GWAS data, including a purpose-designed SNV encoder, that is capable of modelling gene-gene interactions and multidimensional phenotypes, and which scales to the whole-genome sequencing data standard for modern GWAS. We trained our model on data from the UK Biobank (UKBB)(Sudlow et al., 2015) to predict gout; that is, we used genotyped SNV data as an input and the patient's gout diagnosis as an output. The training set was composed of 12,108 cases, of which 6,054 had gout and 6,054 did not.

1.1. Transformers background

Neural networks using Transformer blocks have been applied not only to natural language processing tasks (Devlin et al., 2018) but also to the protein folding problem (Jumper et al., 2021) and prediction of nucleotides in DNA sequences based on upstream and downstream context (Ji et al., 2021). The key novelty of the Transformer block is the use of the attention mechanism (Vaswani et al., 2017), where self-attention "heads" assign a relevance score for every token in the sequence with respect to the rest of the tokens via attention calculations. These calculations work by projecting token vectors onto d-dimensional key $\mathbf{K}$, query $\mathbf{Q}$, and value $\mathbf{V}$ vectors, then taking the following dot products of these projections for each head.

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right)\mathbf{V} \quad (1)$$

The time and space complexity of attention calculation in Eq. 1 is quadratic in the input sequence length n. Significant efforts have recently been made to reduce this cost (Zaheer et al., 2020; Rae et al., 2019; Beltagy et al., 2020). One of the proposed solutions is the Linformer model (Wang et al., 2020), which we build on in this study. As a result of a singular value decomposition analysis on the attention calculation, it is demonstrated (Wang et al., 2020) both theoretically and experimentally that the attention calculation can be reduced from quadratic to linear complexity by including two learnable low-dimensional projections over Eq. (1).

2. Methodology

Data. We obtained SNV data for approximately $487,000$ individuals from the UKBB data bank. For each individual, around $870,000$ SNVs were measured, and a further 90 million imputed. We pre-process data in three steps: 1) extract a sequence of SNVs, with corresponding genome positions, for each person, 2) build a mapping from each SNV to an integer as per Table 1, 3) use this mapping to construct a sequence of integers representing the SNV sequence for each person – see the top of Figure 1.

We separate each SNV into two components, its position and its major/minor alleles. SNV positions are preserved as integer positions within the genome sequence. We mapped different variants and their zygosity to 32 possible combinations as follows. We take the SNV: $[major]/[minor]$ and combine the major and minor alleles into a single word depending on the zygosity, encoding two major, two minor, or one major/one minor allele as follows:

Double major	'[major allele]'
Double minor	'[minor allele]
Single major, single minor	'[major allele],[minor allele]'

We used insertion (**ins**) and deletion (**del**) tokens to represent any form of insertion and deletion, with respect to the reference genome, in the given SNV. We follow (Cahyawijaya et al., 2022) in shortening long sequences with an **'I'** token, representing all nucleotides after the first. Doing so allows us to remove the large number of unique long nucleotide sequences from our mapping. A sequence of more than one nucleotide is then encoded as **'AI','CI','GI'** or **'TI'**, depending on the nucleotide in the first position. In any individual, approximately 2% of SNVs are unknown. That is, we do not know whether either chromosome possesses this variant. We record these as the **'nan'** token. Finally, we mapped the resulting 32 possible combinations to 32 integers. For example, if we have an SNV that has **'G, TGAA'** major and minor values, we encode this as **'G, ins'**, which maps to 21. For the complete list of the 32 combinations, see Table 1.

We classify each individual as either having or not having gout following the method in (Cadzow et al., 2017), with the exception being that classification is now based on self-report or urate-lowering therapy (ULT) only, defined as an individual self-reporting use of allopurinol, probenecid, or colchicine, unless diagnosed with leukemia or lymphoma.

We consider SNVs from the genotyped data only (unimputed), with a minimum allele frequency of 10^{-4}, excluding values with a Hardy-Weinberg equilibrium exact test

Table 1: SNV encoding. Homozygous alleles are encoded as ‘X’, heterozygous as ‘X,Y’, I encodes unique sequences.

‘nan’ : 00	‘ins’ : 01	‘del’ : 02	‘G’ : 03
‘A’ : 04	‘C’ : 05	‘T’ : 06	‘CI’ : 07
‘GI’ : 08	‘TI’ : 09	‘AI’ : 10	‘G,A’ : 11
‘A,G’ : 12	‘G,C’ : 13	‘C,T’ : 14	‘G,T’ : 15
‘C,G’ : 16	‘T,C’ : 17	‘A,ins’ : 18	‘A,C’ : 19
‘CI,del’ : 20	‘G,ins’ : 21	‘GI,del’ : 22	‘T,G’ : 23
‘C,A’ : 24	‘TI,del’ : 25	‘A,T’ : 26	‘C,ins’ : 27
‘T,ins’ : 28	‘AI,del’ : 29	‘T,A’ : 30	‘AI,ins’ : 31

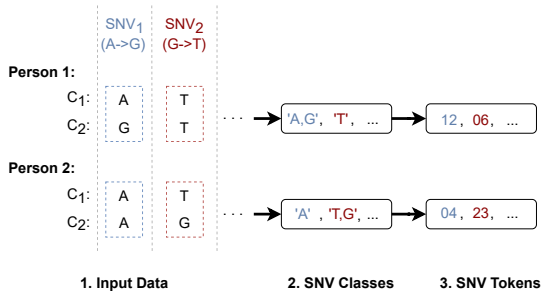


Figure 1: Pre-processing of SNV sequences. Given an input set of SNVs for each person, the combination of SNVs and zygosity is mapped to one of 32 possible cases. Each case is encoded as an integer, and each person’s sequence is encoded as a sequence of these integers.

p-value threshold of 10^{-6}. Among these, we use plink1.9 (Purcell et al., 2007) to find measured or imputed SNVs associated with urate via linear regression, with a p-value $< 10^{-1}$. The resulting data set contains $\approx 66,000$ SNVs. Since the majority of UKBB participants do not have gout, we randomly sub-sample the non-gout cases until we have $\approx 9,000$ of each. 25% of these are reserved for the test set, with remaining 70% used for training, and 5% saved for later verification. As well as genotype information, we include age, sex, and BMI for each individual.

2.1. Transformer Network Architecture

We embed each SNV as a one-hot encoding of the token (Table 1), concatenated with a sine/cosine positional encoding. This is then passed into a Transformer encoder followed by a binary output block for gout/non-gout classification. Self-attention blocks within the encoder use the Linformer architecture (Wang et al., 2020).

The output takes the classification token (first token of the encoder output), reduces it to two cases with a fully connected layer, and classifies with binary Softmax. Our architecture is summarised in Figure 2.

We used a 96 element embedding (where the input uses 32 elements for the token and 64 for the position), 4 attention heads, 6 layers, and a Linformer k-value of 96. We trained using the AMSGrad variant of the AdamW algorithm (Reddi et al., 2019) for 60 epochs with a learning rate of 10^{-7} using batches of size 10 on one 48GB Nvidia Quadro RTX 8000 GPU.

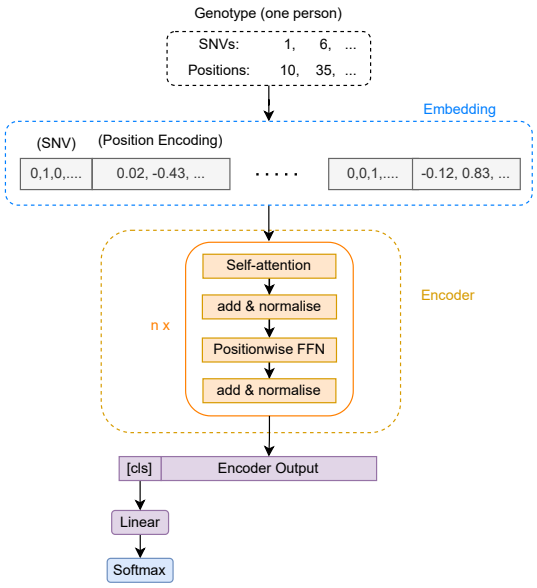
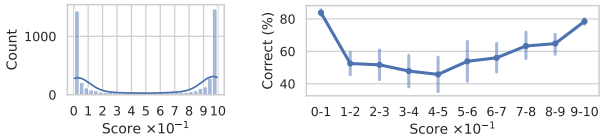


Figure 2: Architecture of SNVformer. Input (genotype) is shown at the top, followed by the embedding, network’s layers, and output at the bottom.



(a) (b)

Figure 3: (a) Distribution of output scores produced by our network. Scores can be interpreted as network’s confidence indicator. (b) Comparison between accuracy (vertical) and output score (horizontal) produced by our network. Includes the 95% confidence interval from 1,000 bootstrap samples.

3. Results

We evaluate our network using a test set of $5,364$ individuals, exactly half of whom have gout. For each sample, the network outputs a score in the range $[0, 1]$, which we interpret as a prediction that the individual has gout if it is greater than 0.5, and that they do not have gout if it is less than 0.5. The exact score can be interpreted as a confidence indicator for that prediction, and we see in Figure 3b that as scores deviate from 0.5, the accuracy of the prediction

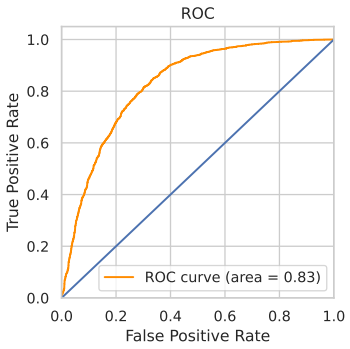


Figure 4: ROC curve when trainied on 66k genotyped SNVs combined with Age, Sex, and BMI.

increases significantly. Moreover, we see in Figure 3a that scores less than 0.1 or greater than 0.9, where accuracy is around 80%, are not uncommon. We will build on this to perform an analysis of which features are contributing to the certainty and uncertainty of predictions. Based on the output score of the network, we achieve an area under the receiver operating characteristic curve (AUROC) of 0.83 (Figure 4).

4. Discussion and Work in Progress

Although our preliminary attempt described in this paper has already achieved prediction accuracy similar to the best performing methods (0.84, see below), there are paths to architectural and data-driven improvements.

Other Transformer architectures have been proposed for the problem of learning over long sequences (in NLP). Memorizing Transformers learn to use a fixed-length "short-term" memory to store tokens expected to be useful later (Wu et al., 2022), while Jaegle et al. iterate over read-process-write tasks. The latter approach identifies and queries missing information from the input, processes and accumulates it in a latent space, and uses that latent representation both for prediction and to further query the input. Radically different from Linformer in their approaches to distant dependencies, these architectures may further improve the identification of distant interacting SNVs in long sequences.

Although GWAS analysis provides some explanatory power by identifying SNVs most predictive of gout, it is commonly limited to occurrence or co-occurrence; predictive power could be substantially improved if relevant broader gene-gene (including higher-order) interactions could be identified. A popular method for explaining predictions in Transformers is by analysing the values of the attention matrix, often by visualisation (Vig, 2019), since these reflect the task-dependent *importance* of each token (Wiegreffe & Pinter, 2019). Alternatives include methods to compute and propagate trained attention-based token relevancy scores (Chefer et al., 2021), or to generate higher-level conceptual explanations (Rigotti et al., 2021). We plan to apply such methods to identify which specific SNVs and combinations support phenotype (gout) prediction. We shall also seek to improve the SNV sequence representation for prediction by pretraining embeddings using sequences from the whole UKBB, using large-scale language-modelling techniques, and, based on preliminary experiments, by investigating other encodings for the SNVs themselves.

We intend to compare our method to polygenic risk-scores (Richardson et al., 2019), which for gout and its related traits have achieved prediction accuracy (AUROC) of between 0.64 and 0.91 (Zhang & Lee, 2020; Tin et al., 2019; Sun et al., 2022) depending on whether the model includes additional factors such as demographics and/or clinical factors. For example, Sun et al. have developed a prediction model for clustering the renal underexcretion of urate phenotype using genetic and clinical variables, achieving AUROC of 0.91 when 7 clinical measures (age, hypertension, nephrolithiasis, blood glucose, serum urate, urea nitrogen, and creatinine) are added. For gout, prediction accuracy (AUROC) improved from 0.68 to 0.84 with the addition of age and sex (Tin et al., 2019). It will be important to assess how these phenotypic and other variables impact the prediction accuracy of our approach given that certain environmental factors might also modify the effect of genetic risk.

Due to the facility of Transformer models in representing complex language, they have the potential to use a far wider range of supporting data than that used in previous work on gout prediction. In particular, a major focus of our future work will involve enabling our models to access existing well-established evidence of interactions directly from the scientific literature, providing context to the Transformer learner itself. We will apply existing techniques from our labs and others for open domain Transformer-based question-answering against the biomedical literature and therefore extract evidence of genomic interactions, including interactions supported by multi-step protein-protein interaction pathways.

Acknowledgements

AG acknowledges support from the Royal Society Te Apārangi through a Rutherford Discovery Fellowship (RDF-UOC1702). This work was partially supported by the Ministry of Business, Innovation, and Employment of New Zealand through a Data Science Programmes grant (UOAX1932).

References

Beltagy, I., Peters, M. E., and Cohan, A. Longformer: The Long-Document transformer. April 2020.

Boocock, J., Leask, M., Okada, Y., Asian Genetic Epidemiology Network (AGEN) Consortium, Matsuo, H., Kawamura, Y., Shi, Y., Li, C., Mount, D. B., Mandal, A. K., Wang, W., Cadzow, M., Gosling, A. L., Major, T. J., Horsfield, J. A., Choi, H. K., Fadason, T., O'Sullivan, J., Stahl, E. A., and Merriman, T. R. Genomic dissection of 43 serum urate-associated loci provides multiple insights into molecular mechanisms of urate control. *Hum. Mol. Genet.*, 29(6):923–943, April 2020.

Cadzow, M., Merriman, T. R., and Dalbeth, N. Performance of gout definitions for genetic epidemiological studies: analysis of UK biobank. *Arthritis Res. Ther.*, 19(1):181, August 2017.

Cahyawijaya, S., Yu, T., Liu, Z., Mak, T. T. W., Zhou, X., and others. SNP2Vec: Scalable Self-Supervised Pre-Training for Genome-Wide association study. *arXiv preprint arXiv*, 2022.

Chefer, H., Gur, S., and Wolf, L. Transformer interpretability beyond attention visualization. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2021.

Choromanski, K., Likhosherstov, V., Dohan, D., Song, X., Gane, A., Sarlos, T., Hawkins, P., Davis, J., Mohiuddin, A., Kaiser, L., Belanger, D., Colwell, L., and Weller, A. Rethinking attention with performers. *arXiv [cs.LG]*, September 2020.

Crona, K., Gavryushkin, A., Greene, D., and Beerenwinkel, N. Inferring genetic interactions from comparative fitness data. *Elife*, 6, December 2017.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. October 2018.

Elmes, K., Heywood, A., Huang, Z., and Gavryushkin, A. A fast Lasso-Based method for inferring Higher-Order interactions. December 2021a.

Elmes, K., Schmich, F., Szczurek, E., Jenkins, J., Beerenwinkel, N., and Gavryushkin, A. Learning epistatic gene interactions from perturbation screens. *PLoS One*, 16(7): e0254491, July 2021b.

Hivert, V., Sidorenko, J., Rohart, F., Goddard, M. E., Yang, J., Wray, N. R., Yengo, L., and Visscher, P. M. Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals. *Am. J. Hum. Genet.*, 108(5):962, May 2021.

Jaegle, A., Borgeaud, S., Alayrac, J. B., Doersch, C., and others. Perceiver io: A general architecture for structured inputs & outputs. *arXiv preprint arXiv*, 2021.

Ji, Y., Zhou, Z., Liu, H., and Davuluri, R. V. DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. *Bioinformatics*, February 2021.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature*, July 2021.

Leask, M. P. and Merriman, T. R. The genetic basis of urate control and gout: Insights into molecular pathogenesis from follow-up study of genome-wide association study loci. *Best Pract. Res. Clin. Rheumatol.*, 35(4):101721, December 2021.

Lienkaemper, C., Lamberti, L., Drain, J., Beerenwinkel, N., and Gavryushkin, A. The geometry of partial fitness orders and an efficient method for detecting genetic interactions. *J. Math. Biol.*, 77(4):951–970, May 2018.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., and Sham, P. C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, 81(3):559–575, September 2007.

Rae, J. W., Potapenko, A., Jayakumar, S. M., and Lillicrap, T. P. Compressive transformers for Long-Range sequence modelling. November 2019.

Reddi, S. J., Kale, S., and Kumar, S. On the convergence of adam and beyond. *arXiv preprint arXiv:1904.09237*, 2019.

Richardson, T. G., Harrison, S., Hemani, G., and Davey Smith, G. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *Elife*, 8, March 2019.

Rigotti, M., Miksovic, C., Giurgiu, I., Gschwind, T., and Scotton, P. Attention-based interpretability with concept transformers. In *International Conference on Learning Representations*, 2021.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., and Collins, R. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.*, 12(3):e1001779, March 2015.

Sun, M., Sun, W., Zhao, X., Li, Z., Dalbeth, N., Ji, A., He, Y., Qu, H., Zheng, G., Ma, L., Wang, J., Shi, Y., Fang, X., Chen, H., Merriman, T. R., and Li, C. A machine learning-assisted model for renal urate underexcretion with genetic and clinical variables among chinese men with gout. *Arthritis Res. Ther.*, 24(1):67, March 2022.

Tin, A., Marten, J., Halperin Kuhns, V. L., Li, Y., Wuttke, M., Kirsten, H., Sieber, K. B., Qiu, C., Gorski, M., Yu, Z., Giri, A., Sveinbjornsson, G., Li, M., Chu, A. Y., Hoppmann, A., O'Connor, L. J., Prins, B., Nutile, T., Noce, D., Akiyama, M., Cocca, M., Ghasemi, S., van der Most, P. J., Horn, K., Xu, Y., Fuchsberger, C., Sedaghat, S., Afaq, S., Amin, N., Ärnlöv, J., Bakker, S. J. L., Bansal, N., Baptista, D., Bergmann, S., Biggs, M. L., Biino, G., Boerwinkle, E., Bottinger, E. P., Boutin, T. S., Brumat, M., Burkhardt, R., Campana, E., Campbell, A., Campbell, H., Carroll, R. J., Catamo, E., Chambers, J. C., Ciullo, M., Concas, M. P., Coresh, J., Corre, T., Cusi, D., Felicita, S. C., de Borst, M. H., De Grandi, A., de Mutsert, R., de Vries, A. P. J., Delgado, G., Demirkan, A., Devuyst, O., Dittrich, K., Eckardt, K.-U., Ehret, G., Endlich, K., Evans, M. K., Gansevoort, R. T., Gasparini, P., Giedraitis, V., Gieger, C., Girotto, G., Gögele, M., Gordon, S. D., Gudbjartsson, D. F., Gudnason, V., German Chronic Kidney Disease Study, Haller, T., Hamet, P., Harris, T. B., Hayward, C., Hicks, A. A., Hofer, E., Holm, H., Huang, W., Hutri-Kähönen, N., Hwang, S.-J., Ikram, M. A., Lewis, R. M., Ingelsson, E., Jakobsdottir, J., Jonsdottir, I., Jonsson, H., Joshi, P. K., Josyula, N. S., Jung, B., Kähönen, M., Kamatani, Y., Kanai, M., Kerr, S. M., Kiess, W., Kleber, M. E., Koenig, W., Kooner, J. S., Körner, A., Kovacs, P., Krämer, B. K., Kronenberg, F., Kubo, M., Kühnel, B., La Bianca, M., Lange, L. A., Lehne, B., Lehtimäki, T., Lifelines Cohort Study, Liu, J., Loeffler, M., Loos, R. J. F., Lyytikäinen, L.-P., Magi, R., Mahajan, A., Martin, N. G., März, W., Mascalzoni, D., Matsuda, K., Meisinger, C., Meitinger, T., Metspalu, A., Milaneschi, Y., V. A. Million Veteran Program, O'Donnell, C. J., Wilson, O. D., Gaziano, J. M., Mishra, P. P., Mohlke, K. L., Mononen, N., Montgomery, G. W., Mook-Kanamori, D. O., Müller-Nurasyid, M., Nadkarni, G. N., Nalls, M. A., Nauck, M., Nikus, K., Ning, B., Nolte, I. M., Noordam, R., O'Connell, J. R., Olafsson, I., Padmanabhan, S., Penninx, B. W. J. H., Perls, T., Peters, A., Pirastu, M., Pirastu, N., Pistis, G., Polasek, O., Ponte, B., Porteous, D. J., Poulain, T., Preuss, M. H., Rabelink, T. J., Raffield, L. M., Raitakari, O. T., Rettig, R., Rheinberger, M., Rice, K. M., Rizzi, F., Robino, A., Rudan, I., Krajcoviechova, A., Cifkova, R., Rueedi, R., Ruggiero, D., Ryan, K. A., Saba, Y., Salvi, E., Schmidt, H., Schmidt, R., Shaffer, C. M., Smith, A. V., Smith, B. H., Spracklen, C. N., Strauch, K., Stumvoll, M., Sulem, P., Tajuddin, S. M., Teren, A., Thiery, J., Thio, C. H. L., Thorsteinsdottir, U., Toniolo, D., Tönjes, A., Tremblay, J., Uitterlinden, A. G., Vaccargiu, S., van der Harst, P., van Duijn, C. M., Verweij, N., Völker, U., Vollenweider, P., Waeber, G., Waldenberger, M., Whitfield, J. B., Wild, S. H., Wilson, J. F., Yang, Q., Zhang, W., Zonderman, A. B., Bochud, M., Wilson, J. G., Pendergrass, S. A., Ho, K., Parsa, A., Pramstaller, P. P., Psaty, B. M., Böger, C. A., Snieder, H., Butterworth, A. S., Okada, Y., Edwards, T. L., Stefansson, K., Susztak, K., Scholz, M., Heid, I. M., Hung, A. M., Teumer, A., Pattaro, C., Woodward, O. M., Vitart, V., and Köttgen, A. Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat. Genet.*, 51(10):1459–1474, October 2019.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. U., and Polosukhin, I. Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 30*, pp. 5998–6008. Curran Associates, Inc., 2017.

Vig, J. A multiscale visualization of attention in the transformer model. June 2019.

Wang, S., Li, B. Z., Khabsa, M., Fang, H., and Ma, H. Linformer: Self-Attention with linear complexity. June 2020.

Wei, W.-H., Hemani, G., and Haley, C. S. Detecting epistasis in human complex traits. *Nat. Rev. Genet.*, 15(11): 722–733, November 2014.

Wiegreffe, S. and Pinter, Y. Attention is not not explanation. August 2019.

Wu, Y., Rabe, M. N., Hutchins, D., and Szegedy, C. Memorizing transformers. March 2022.

Zaheer, M., Guruganesh, G., Dubey, A., Ainslie, J., Alberti, C., Ontanon, S., Pham, P., Ravula, A., Wang, Q., Yang, L., and Ahmed, A. Big bird: Transformers for longer sequences. July 2020.

Zhang, Y. and Lee, M. T. M. Serum urate polygenic risk score can improve gout risk prediction: A Large-Scale cohort study. *Front. Genet.*, 11:604219, 2020.

Zhu, Z., Bakshi, A., Vinkhuyzen, A. A. E., Hemani, G., Lee, S. H., Nolte, I. M., van Vliet-Ostaptchouk, J. V.,

Snieder, H., LifeLines Cohort Study, Esko, T., Milani, L., Mägi, R., Metspalu, A., Hill, W. G., Weir, B. S., Goddard, M. E., Visscher, P. M., and Yang, J. Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am. J. Hum. Genet.*, 96(3):377–385, March 2015.