Pulse

Research Article

Analysis of Rare Variants in 470,000 Exome-Sequenced UK Biobank Participants Implicates Novel Genes Affecting Risk of Hypertension

David Curtis

UCL Genetics Institute, University College London, London, UK

Abstract

Introduction: A previous study of 200,000 exome-sequenced UK Biobank participants to test for association of rare coding variants with hypertension implicated two genes at exome-wide significance, *DNMT3A* and *FES*. A total of 42 genes had an uncorrected *p* value <0.001. These results were followed up in a larger sample of 470,000 exome-sequenced participants. ***Methods:*** Weighted burden analysis of rare coding variants in a new sample of 97,050 cases and 172,263 controls was carried out for these 42 genes. Those showing evidence for association were then analysed in the combined sample of 167,127 cases and 302,691 controls. ***Results:*** The association of *DNMT3A* and *FES* with hypertension was replicated in the new sample and they and the previously implicated gene *NPR1*, which codes for a membrane-bound guanylate cyclase, were all exome-wide significant in the combined sample. Also exome-wide significant as risk genes for hypertension were *GUCY1A1*, *ASXL1*, and *SMAD6*, while *GUCY1B1* had a nominal *p* value of <0.0001. *GUCY1A1* and *GUCY1B1* code for subunits of a soluble guanylate cyclase. For two genes, *DBH*, which codes for dopamine beta hydroxylase, and *INPPL1*, rare coding variants predicted to impair gene function were protective against hypertension, again with exome-wide significance. ***Conclusion:*** The findings offer new insights into biological risk factors for hypertension which could be the subject of further investigation. In particular, genetic variants predicted to impair the function of either membrane-bound guanylate cyclase, activated by natriuretic peptides, or soluble guanylate cyclase, activated by nitric oxide, increase risk of hypertension. Conversely, variants impairing the function of dopamine beta hydroxylase, responsible for the synthesis of norepinephrine, reduce hypertension risk.

© 2023 The Author(s).
Published by S. Karger AG, Basel

Introduction

An earlier analysis of the first release of exome sequence data for 200,000 UK Biobank participants identified two genes affecting risk of hypertension at exome-wide significance, *DNMT3A* and *FES* [1]. However, since 20,384 genes were analysed the expected number of genes to be significant at *p* < 0.001 would be 20, whereas in fact 42 genes achieved this threshold, suggesting that some of these might represent true associations. Additionally, a number of these genes had functions which meant that they might plausibly be involved in affecting risk of hypertension.

Correspondence to:
David Curtis, d.curtis@ucl.ac.uk

Karger
OPEN ACCESS

Subsequently, UK Biobank has made available exome sequence data for 470,000 participants. The present study utilised data from the 270,000 participants not previously analysed and restricted attention to only the 42 genes previously achieving a signed log p value (SLP) with absolute value greater than 3. For a subset of these genes which appeared to be of interest, data were analysed for all 470,000 participants to better assess the overall evidence implicating them and to gain insights into the categories of variant contributing to association.

It should be noted that most of this exome sequence data has also been used in two previous studies which tested for gene-wise associations with very large numbers of phenotypes and which included some hypertension-related phenotypes, so results from the present study need to be considered alongside these [2, 3].

Methods

UK Biobank participants are volunteers intended to be broadly representative of the UK population and are not selected on the basis of having any health condition. UK Biobank had obtained ethics approval from the North West Multi-centre Research Ethics Committee which covers the UK (approval number: 11/NW/0382) and had obtained written informed consent from all participants. The UK Biobank approved an application for use of the data (ID 51119) and ethics approval for the analyses was obtained from the UCL Research Ethics Committee (11527/001). The UK Biobank Research Analysis Platform was used to access the Final Release Population level exome OQFE variants in PLINK format for 469,818 exomes which had been produced at the Regeneron Genetics Center based on DNA extracted from stored blood samples and using the protocols described here: https://dnanexus.gitbook.io/uk-biobank-rap/science-corner/whole-exome-sequencing-oqfe-protocol/protocol-for-processing-ukb-whole-exome-sequencing-data-sets [3]. All variants were then annotated using the standard software packages VEP, PolyPhen, and SIFT [4–6]. To obtain population principal components reflecting ancestry, version 2.0 of *plink* (https://www.cog-genomics.org/plink/2.0/) was run with the options −*maf 0.1 − pca 20 approx* [7, 8].

Cases with hypertension were defined in exactly the same way as in the previous study, consisting of participants who met any of the following four criteria: a self-reported diagnosis recorded as hypertension or essential hypertension; self-reported taking medication for high blood pressure; reporting taking any of a list of named medications commonly used to treat high blood pressure (https://www.nhs.uk/conditions/high-blood-pressure-hypertension/); having an ICD10 diagnosis of essential hypertension, hypertensive heart disease or hypertensive renal disease in hospital records or as a cause of death [1]. Participants meeting any of these criteria were taken to be cases whereas all those meeting none of these criteria were taken to be controls. In the primary analyses to implicate specific genes attention was restricted to participants not included in the earlier study, consisting of 97,050 cases and 172,263 controls. For the subsequent analyses using the whole sample, there were 167,127 cases and 302,691 controls.

The same analytic methods as had been used previously were applied, with the description repeated here for convenience. The SCOREASSOC program was used to carry out a weighted burden analysis to test whether, in each gene, sequence variants which were rarer and/or predicted to have more severe functional effects occurred more commonly in cases than controls. Attention was restricted to rare variants with minor allele frequency (MAF) ≤ 0.01 in both cases and controls. As previously described, variants were weighted by overall MAF so that variants with MAF = 0.01 were given a weight of 1 while very rare variants with MAF close to zero were given a weight of 10 [9]. Variants were also weighted according to their functional annotation using the GENEVAR-ASSOC program, which was used to generate the input files for weighted burden analysis by SCOREASSOC [10, 11]. Variants predicted to cause complete loss of function (LOF) of the gene were assigned a weight of 100. Nonsynonymous variants were assigned a weight of 5 but if PolyPhen annotated them as possibly or probably damaging then 5 or 10 was added to this and if SIFT annotated them as deleterious then 20 was added. The full set of weights and categories is displayed in Table 1 of the previous study [1]. As described previously, the weight due to MAF and the weight due to functional annotation were multiplied together to provide an overall weight for each variant. Variants were excluded if there were more than 10% of genotypes missing in the controls and cases or if the heterozygote count was smaller than both homozygote counts in the controls and cases. If a subject was not genotyped for a variant then they were assigned the subject-wise average score for that variant. For each subject, a gene-wise weighted burden score was derived as the sum of the variant-wise weights, each multiplied by the number of alleles of the variant which the given subject possessed.

Analyses were restricted to the 42 genes significant at $p < 0.001$ in the previous study. For each gene, logistic regression analysis was carried out with hypertension as the dependent variable including the first 20 population principal components and sex as covariates and a likelihood ratio test was performed comparing the likelihoods of the models with and without the gene-wise burden score. This is a test for association between the gene-wise burden score and caseness and the statistical significance was summarised as a SLP, which is the log base 10 of the p value given a positive sign if the score is higher in cases and negative if it is higher in controls. Since only 42 genes were analysed, after correcting for multiple testing a gene could be declared statistically significant if it achieved an SLP with absolute value greater than $-\log10(0.05/42) = 2.9$ using the new samples.

Follow-up analyses were performed on all genes individually achieving an absolute value for the SLP of 2 (equivalent to $p = 0.01$) and also *GUCY1B1*. For this subset of genes, the weighted burden analysis described above was carried out using the whole sample of 167,127 cases and 302,691 controls. Additionally, for each subject a count was obtained of the number of variants they carried falling into particular broad categories, such as LOF, protein altering, etc. The full list of these categories is shown in Supplementary Table 1 (for all online suppl. material, see https://doi.org/10.1159/000535157). These counts were entered into a multiple logistic regression analysis with hypertension as the dependent variable and again including sex and 20 principal components as covariates in order to elucidate the contribution of different types of variant to the overall evidence for association. The odds ratios (ORs) associated with each category were estimated along with their

Table 1. Genes with absolute value of SLP exceeding 3 (equivalent to $p < 0.001$) for association with hypertension in previous study showing the SLP obtained in the new sample

Symbol	SLP in original sample	Name	SLP in new sample	SLP in combined sample
DNMT3A	8.21	DNA Methyltransferase 3 Alpha	**6.84**	14.20
FES	6.10	FES Proto-Oncogene, Tyrosine Kinase	**4.66**	9.92
GUCY1A1	5.54	Guanylate Cyclase 1 Soluble Subunit Alpha 1	**5.06**	9.13
NPR1	5.14	Natriuretic Peptide Receptor 1	2.84	7.98
MAG	4.52	Myelin Associated Glycoprotein	0.12	
SMAD6	4.10	SMAD Family Member 6	2.43	6.02
GUCY1B1	3.93	Guanylate Cyclase 1 Soluble Subunit Beta 1	1.88	4.95
SKIV2L	3.87	Ski2 Like RNA Helicase	−0.33	
ASXL1	3.75	ASXL Transcriptional Regulator 1	**5.34**	8.35
MYO10	3.75	Myosin X	−0.40	
PTGER3	3.52	Prostaglandin E Receptor 3	−0.96	
POLR3H	3.47	RNA Polymerase III Subunit H	0.45	
IFT172	3.39	Intraflagellar Transport 172	−0.32	
LITAF	3.37	Lipopolysaccharide Induced TNF Factor	0.54	
CSK	3.37	C-Terminal Src Kinase	1.18	
SNX27	3.34	Sorting Nexin 27	0.10	
TEK	3.31	EK Receptor Tyrosine Kinase	0.88	
PHF21A	3.23	PHD Finger Protein 21A	0.90	
LOC102723675	3.21	Uncharacterized LOC102723675	0.12	
DXO	3.14	Decapping Exoribonuclease	0.37	
RHBDL3	3.12	Rhomboid Like 3	0.08	
ASB7	3.12	Ankyrin Repeat And SOCS Box Containing 7	0.33	
MPP2	3.11	Membrane Palmitoylated Protein 2	0.38	
ZNF549	3.10	Zinc Finger Protein 549	0.55	
MAP3K14-AS1	3.08	MAP3K14 Antisense RNA 1	1.69	
ZNF600	−3.01	Zinc Finger Protein 600	−0.28	
SIRT2	−3.05	Sirtuin 2	−0.67	
PARGP1	−3.05	Poly(ADP-Ribose) Glycohydrolase Pseudogene 1	1.16	
ZNF226	−3.13	Zinc Finger Protein 226	0.15	
LOC105375153	−3.18	Uncharacterized LOC105375153	0.69	
INPPL1	−3.19	Inositol Polyphosphate Phosphatase Like 1	**−4.02**	−7.09
FNDC8	−3.23	Fibronectin Type III Domain Containing 8	−0.43	
KRT82	−3.26	Keratin 82	−0.47	
GOLGA4	−3.36	Golgin A4	−0.51	
DBH	−3.40	Dopamine Beta Hydroxylase	**−5.61**	−9.71
ELOF1	−3.59	Elongation Factor 1 Homolog	0.08	
AGTR1	−3.77	Angiotensin II Receptor Type 1	−0.76	
SIPA1	−3.81	Signal-Induced Proliferation-Associated 1	−1.41	
ZYX	−3.83	Zyxin	−0.31	
ZFAT	−4.00	Zinc Finger And AT-Hook Domain Containing	−0.34	
OSTF1	−4.05	Osteoclast Stimulating Factor 1	0.17	
PREP	−5.03	Prolyl Endopeptidase	−0.11	

SLPs significant after correction for multiple testing are shown in bold. For all genes achieving an absolute value of SLP exceeding 2 and for GUCY1B1, the SLP obtained for combined sample is also shown.

standard errors and the Wald statistic was used to obtain a p value. This p value was converted to an SLP, again with the sign being positive if the OR was greater than 1, indicating that variants in that category tended to increase risk. Data manipulation and statistical analyses were performed using GENEVARASSOC, SCOREASSOC and R [12].

Results

Table 1 shows the results of the primary analysis. The two genes which were exome-wide significant in the previous study, *DNMT3A* and *FES*, again show

convincing evidence of association in the new sample with SLPs of 6.84 and 4.66 respectively. Additionally, four genes which were not exome-wide significant in the previous sample produce results which are significant after correction for multiple testing in the new sample: *GUCY1A1* (SLP = 5.06), *ASXL1* (SLP = 5.34), *INPPL1* (SLP = −4.02), and *DBH* (SLP = −5.61). These genes were taken forward for secondary analysis in the combined sample and it is noteworthy that, as also shown in Table 1, all six of them then produced SLPs with absolute value exceeding the threshold of 5.61 to be regarded as exome-wide significant. All other genes producing an SLP with absolute value exceeding 2 were also analysed in the combined sample, along with *GUCY1B1* which produced an SLP of only 1.88 but which codes for a subunit of the same guanylate cyclase as *GUCY1A1*. Two of these genes produced SLPs in the combined sample exceeding the criterion for exome-wide significance, *NPR1* (SLP = 7.98) and *SMAD6* (SLP = 6.02). In the combined sample, *GUCY1B1* produced an SLP of 4.95.

In order to gain insights into the effects of different categories of variant within these genes of interest, counts for variants of each category in each subject were entered into multiple logistic regression analysis along with sex and 20 principal components as covariates. Table 2 shows results for all genes and categories where the category was significant at $p < 0.05$, with an absolute value of SLP >1.3. Additionally, for any gene in which the SIFT or PolyPhen annotations were significant the results for the protein-altering category are also shown because the annotations only apply to protein-altering variants (full results for all genes and all categories are shown in online supplementary Table 2).

A number of findings are of interest. For every gene, LOF variants produce the largest effect. For the genes in which impaired functioning increases hypertension risk, the OR for a LOF variant ranges from 1.33 for *SMAD6* to 1.96 for *FES*. For the genes in which impaired function is protective against hypertension, the OR for a LOF variant is 0.75 for *INPPL1* and 090 for *DBH*. Although LOF variants are associated with the largest effect size they are also very rare and for most genes, the mean number of LOF variants carried per subject is fewer than 0.001. By contrast, other categories of variant tend to be commoner but to be associated with smaller effect sizes. For example, in *DNMT3A* a LOF variant has an OR of 1.68 but a nonsynonymous variant annotated as probably damaging by SIFT has an OR of $1.13 \times 1.23 = 1.39$. The table shows that there is a lack of consistency across genes in terms of which categories of variant show evidence of association. It is generally the case that variants in categories asso-ciated with an effect have low cumulative frequencies, but it is perhaps noteworthy that nonsynonymous variants in *DBH* which are annotated as deleterious by SIFT have OR of 0.9, and the average number carried per subject is greater than 0.035. This means that about 1 person in 30 carries a *DBH* variant of a category which on average slightly reduces their risk of hypertension.

Discussion

This study confirms that rare coding variants which impair the function of the two genes identified in the earlier study, *DNMT3A* and *FES*, do indeed increase the risk of hypertension. Coding variants in *NPR1* have previously been reported to be associated with hypertension risk and this gene is also exome-wide significant in the combined sample [13, 14]. Four genes are newly implicated at conventional levels of significance, whether one corrects for the multiple testing of 42 genes in this study or the 20,384 originally considered, these being *GUCY1A1*, *ASXL1*, *INPPL1*, and *DBH*. Variants damaging the function of *GUCY1A1* or *ASXL1* increase hypertension risk while variants in *INPPL1* and *DBH* are protective. For two other genes, the results are somewhat more ambiguous. *SMAD6* failed to achieve statistical significance in the primary analysis of the new sample and barely reached the criterion of exome-wide significance in the combined sample. *GUCY1B1* was not statistically significant in the primary analysis and nor is it exome-wide significant overall, but it is such a biologically plausible candidate that it does not seem sensible to simply ignore it.

As mentioned above, this dataset has been used for previous two studies testing for association between exome sequence variance with a very large number of phenotypes, which for convenience we can refer to as the Regeneron and AstraZeneca studies [2, 3]. The Regeneron study carried out a variety of single variant and gene-wise burden tests on 3,994 health-related traits to produce a total of about 2.3 billion tests, yielding a critical p value of 2.18×10^{-11} (corresponding to SLP = 10.66) and reported 8,865 significant associations which are presented in their online supplementary Data 2 [3]. This contains no gene-wise associations for any of these genes with any hypertension-related phenotype but does include one single variant association of a nonsynonymous variant in *DBH*, 9:133636634:G:C, with reduced diastolic blood pressure at 5.48×10^{-14} (SLP = −13.26). For the AstraZeneca study, all gene-wise and variant-wise associations with 17,361 binary and 1,419 quantitative

Table 2. Results obtained from multiple logistic regression analyses including all variant categories along with sex and 20 principal components as covariates

Gene/category	Number of separate variants	Total count in controls	Mean count in controls	Total count in cases	Mean count in cases	OR (95% confidence interval)	SLP
DNMT3A							
Three prime UTR	35	92	0.000304	31	0.000186	0.60 (0.40–0.91)	−1.84
Protein altering	404	3,191	0.010543	1,963	0.011746	1.13 (0.99–1.29)	1.24
LOF	141	319	0.001054	287	0.001717	1.68 (1.43–1.98)	9.70
PolyPhen probably damaging	138	513	0.001695	383	0.002292	1.23 (1.03–1.48)	1.68
FES							
Protein altering	521	8,325	0.027503	4,803	0.028739	0.93 (0.89–0.97)	−2.78
LOF	65	74	0.000244	79	0.000473	1.96 (1.42–2.71)	4.50
SIFT deleterious	269	1,639	0.005415	1,150	0.006881	1.32 (1.16–1.50)	4.68
GUCY1A1							
Three prime UTR	28	143	0.000473	111	0.000665	1.49 (1.15–1.93)	2.74
LOF	49	113	0.000373	113	0.000676	1.85 (1.42–2.42)	5.37
NPR1							
Protein altering	589	10,020	0.033102	6,178	0.036963	0.93 (0.89–0.98)	−2.27
LOF	54	142	0.000469	118	0.000706	1.52 (1.19–1.96)	3.10
SIFT	311	3,783	0.012497	2,577	0.015417	1.13 (1.01–1.26)	1.55
PolyPhen probably damaging	151	687	0.002270	486	0.002908	1.22 (1.06–1.42)	2.25
SMAD6							
Protein altering	694	7,467	0.024669	4,473	0.026764	1.02 (0.94–1.11)	0.23
Indel, etc.	24	488	0.001612	263	0.001574	0.96 (0.83–1.12)	−0.20
LOF	186	225	0.000743	165	0.000987	1.33 (1.08–1.64)	2.27
PolyPhen probably damaging	305	885	0.002924	569	0.003405	1.13 (1.00–1.27)	1.36
GUCY1B1							
Protein altering	293	2,167	0.007158	1,309	0.007835	1.02 (0.94–1.11)	0.17
LOF	40	63	0.000208	53	0.000317	1.47 (1.01–2.14)	1.41
PolyPhen probably damaging	102	219	0.000724	194	0.001161	1.42 (1.06–1.91)	1.77
ASXL1							
Intronic etc.	799	32,186	0.106334	17,554	0.105036	0.98 (0.96–1.00)	−1.60
Synonymous	418	14,901	0.049228	9,056	0.054186	1.04 (1.01–1.06)	2.02
LOF	79	291	0.000960	326	0.001953	1.88 (1.60–2.22)	14.17
INPPL1							
Five prime UTR	56	5,549	0.018331	3,194	0.019114	1.06 (1.01–1.11)	2.05
Protein altering	828	8,594	0.028392	4,534	0.027130	1.00 (0.95–1.04)	−0.07
LOF	106	233	0.000770	96	0.000575	0.75 (0.59–0.96)	−1.72
SIFT	433	3,295	0.010886	1,587	0.009496	0.90 (0.81–1.00)	−1.37
DBH							
Synonymous	220	4,367	0.014427	2,437	0.014582	0.93 (0.88–0.98)	−2.22
Protein altering	492	18,379	0.060719	9,602	0.057453	0.99 (0.95–1.04)	−0.17
LOF	50	1,443	0.004768	717	0.004291	0.90 (0.82–0.99)	−1.55
SIFT deleterious	269	12,166	0.040193	6,000	0.035901	0.90 (0.83–0.97)	−2.21

Results are shown for each category of variant for which the category-wise results are nominally significant at $p < 0.05$ (absolute value of SLP >1.3). Where the results of SIFT or PolyPhen were significant, the results are also shown for all protein altering variants.

phenotypes are reported on the AstraZeneca PheWAS Portal at https://azphewas.com/ [2]. This was accessed to find the most significant p value for any analysis of each of these genes with the phenotype "ICD10 I10 Essential (primary) hypertension," and Table 3 shows the results obtained compared with those for the current study. It

Table 3. Comparison of results from current study to those reported for the AstraZeneca study

Gene	SLP for combined sample in current study	SLP for AstraZeneca study
DNMT3A	14.20	8.19
FES	9.92	5.04
GUCY1A1	9.13	6.61
NPR1	7.98	4.12
SMAD6	6.02	2.25
GUCY1B1	4.95	1.70
ASXL1	8.35	10.17
INPPL1	−7.09	−3.45
DBH	−9.71	−4.12

The results for the AstraZeneca study are displayed as the equivalent SLP for the most significant result reported for that gene with the phenotype ICD10 Essential (primary) hypertension.

can be seen that for each gene except *ASXL1*, the AstraZeneca results provide less support for association than the present study.

There are a number of differences between the approach used here and that used in the Regeneron and AstraZeneca studies which might account for differences in the results obtained. One is that the Regeneron study analyses are based on 430,998 participants with European ancestry and the AstraZeneca study used 394,695 exomes which were of high quality from participants who were predominantly unrelated and of European ancestry while results for the current study are based on all 469,818 exomes made available in the final release. Another difference is that the other studies used multiple phenotypes and multiple methods of single variant and gene-based analyses, necessitating a rigorous correction for multiple testing, meaning that some results potentially of interest might not be identified. Another difference is that the present study uses a definition of caseness which combines information from a number of sources rather than using a single measure. Finally, the present study implements a weighted burden analysis in which all variants are analysed together but variants which are rare and/or predicted to have more serious consequences are accorded higher weights. The other studies implemented burden analyses using different sets of variant category. Thus, one analysis might only include a small number of LOF variants and another might include LOF variants along with a much larger number of nonsynonymous variants. However, without weighting the LOF variants more strongly there could be a risk that their signal would be swamped out by the other variants. Finally, it should be clearly stated that although the present study arguably highlights some genes of interest more effectively than the other studies it also misses associations which were picked up by the others. The Regeneron study did report a risk-lowering association of *SLC9A3R2* with hypertension and measured blood pressure (SLP = −18.70) while the AstraZeneca study found a significant association of *PKD1* with hypertension (SLP = 12.28) but neither of these genes were picked up in the current investigation.

Another report has recently been published which used the same dataset to test for association between LOF variants and systolic or diastolic blood pressure [15]. They report single variant associations with ten genes (*ANKDD1B*, *ENPEP*, *PNCK*, *BTN3A2*, *C1orf145*, *CASP9*, *DBH*, *KIAA1161*, *OR4X1*, and *TMC3*) and associations with burden of LOF variants for five genes (*TTN*, *NOS3*, *FES*, *SMAD6*, and *COL21A1*). Thus, the only gene-wise results which overlap with this study are for *FES* and *SMAD6*. The variant association they report for *DBH* is with the splice site variant 9:133636712:T:C but the minus log10 p value they report for this is only 6.79, which arguably should not be regarded as statistically significant given that 515,198 LOF variants were tested. Their approach differed from the current study in that they considered only LOF variants, that they carried out single variant analyses as well as gene-based burden tests, and that they used measured blood pressure rather than a clinical hypertension phenotype.

For at least some of the implicated genes, there are plausible biological mechanisms which might contribute to the observed association. As discussed previously, DNMT3A methylates DNA conditional on the associated H3K4 residue being unmethylated and hence has effects on gene expression. In mice *Dnmt3a* knockdown has been shown to lead to reduced methylation of the gene for angiotensin receptor type 1a, *Agtr1a*, leading to increased *Agtr1a* expression and salt-induced hypertension [16]. It has been reported that in human children and adolescents

Curtis

DNMT3A levels are positively associated with obesity and with diastolic blood pressure [17]. However, the present study indicates that impaired functioning of *DNMT3A* increases rather than decreases the risk of hypertension. Another implicated gene, *ASXL1*, is also a transcriptional regulator and is involved in the deubiquitination of histone H2A lysine 119 which is frequently mutated in cancers but the mechanism whereby impaired function might increase hypertension risk remains to be elucidated [18]. In Bohring-Opitz syndrome, characterised by a prominent metopic suture, hypertelorism, exophthalmos, cleft lip and palate, limb anomalies, as well as difficulty feeding with severe developmental delays, almost 50% have de novo LOF variants in *ASXL1* [19, 20]. However, the finding that over 600 UK Biobank participants carry *ASXL1* LOF variants shows that Bohring-Opitz syndrome is by no means an inevitable consequence and further work could address whether specific subsets of LOF variants, perhaps affecting particular transcripts, are relevant.

Perhaps the genes for which a biological mechanism is most obvious are *NPR1*, *GUCY1A1*, and *GUCY1B1*. *NPR1* codes for a membrane-bound guanylate cyclase which acts as a receptor for natriuretic peptides and variants in this gene associated with blood pressure have been shown to modulate guanylate cyclase activity [14]. *GUCY1A1* and *GUCY1B1* code for subunits of a soluble guanylate cyclase which responds to nitric oxide signalling and has a well-established role in blood pressure regulation [21]. Guanylate cyclase produces cyclic GMP which acts as an intracellular messenger to mediate responses such as vasodilation and the results reported here further reinforce the notion that impaired guanylate cyclase activity is a risk factor for hypertension.

Another gene for which the role in hypertension susceptibility is fairly obvious is *DBH*. Dopamine beta hydroxylase catalyses the oxidative hydroxylation of dopamine to norepinephrine in the adrenal medulla and the synaptic vesicles of postganglionic sympathetic neurons, and variants affecting both copies of the gene can act recessively to cause norepinephrine deficiency syndrome, characterised by orthostatic hypotension among other symptoms [22]. The present results demonstrate that variants affecting a single copy of the gene can be protective against hypertension, presumably by leading to a more modest reduction in norepinephrine levels. In animal studies, inhibition of dopamine beta hydroxylase using nepicastat reduces behaviours associated with use of cocaine and morphine [23, 24]. Nepicastat was trialled as a treatment for post-traumatic stress disorder and for cocaine dependence, but although both

trials were completed no results have been published (https://clinicaltrials.gov/study/NCT00641511, https://clinicaltrials.gov/study/NCT00656357). It is unclear whether dopamine beta hydroxylase might be a suitable drug target for the management of hypertension. In order to avoid central effects, it might be productive to investigate therapeutic agents which did not cross the blood-brain barrier. As for *ASXL1*, current knowledge does not allow us to understand by what mechanisms *FES*, *SMAD6*, or *INPPL1* might impact risk of hypertension and if the results of the present study appear convincing then this could act as a stimulus to carrying out further research into the function of these genes.

Overall, this study implicates a small number of genes in which variants predicted to impair gene function can increase or decrease risk of hypertension. The cumulative allele frequency of these variants is low and hence these results do not have clinical or public health implications. For example, exome sequencing could identify a very small number of people at moderately increased risk of developing hypertension, but doing so would not clearly lead to any therapeutic intervention if they were not, in fact, hypertensive. Rather, the findings may focus attention on particular pathways important to moderating the risk of hypertension and this might ultimately lead to the identification of novel drug targets and treatment strategies.

Statement of Ethics

UK Biobank had obtained ethics approval from the North West Multi-centre Research Ethics Committee which covers the UK (approval number: 11/NW/0382) and had obtained written informed consent from all participants. The UK Biobank approved an application for use of the data (ID 51119) and ethics approval for the analyses was obtained from the UCL Research Ethics Committee (11527/001).

Conflict of Interest Statement

The author declares that he has no conflict of interest.

Funding Sources

Author Contributions

D.C. carried out the analyses and prepared the manuscript.

Data Availability Statement

The raw data are available on application to UK Biobank. Detailed results with variant counts cannot be made available because they might be used for subject identification. Scripts and relevant derived variables will be deposited in UK Biobank. Software and scripts used to carry out the analyses are also available at https://github.com/davenomiddlenamecurtis. Further enquiries can be directed to the corresponding author.

References

1 Curtis D. Analysis of 200,000 exome-sequenced UK Biobank subjects implicates genes involved in increased and decreased risk of hypertension. Pulse. 2021;9(1–2):17–29.

2 Wang Q, Dhindsa RS, Carss K, Harper AR, Nag A, Tachmazidou I, et al. Rare variant contribution to human disease in 281,104 UK Biobank exomes. Nature. 2021;597:527–32.

3 Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. Nature. 2021;599(7886):628–34.

4 McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016;17(1):122.

5 Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7.20.

6 Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073–81.

7 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4(1):7.

8 Galinsky KJ, Bhatia G, Loh PR, Georgiev S, Mukherjee S, Patterson NJ, et al. Fast principal-component analysis reveals convergent evolution of ADH1B in europe and east asia. Am J Hum Genet. 2016;98(3):456–72.

9 Curtis D. Multiple linear regression allows weighted burden analysis of rare coding variants in an ethnically heterogeneous population. Hum Hered. 2020;85(1):1–10.

10 Curtis D. A rapid method for combined analysis of common and rare variants at the level of a region, gene, or pathway. Adv Appl Bioinform Chem. 2012;5:1–9.

11 Curtis D. Pathway analysis of whole exome sequence data provides further support for the involvement of histone modification in the aetiology of schizophrenia. Psychiatr Genet. 2016;26:223–7.

12 R Core Team. R. A language and environment for statistical computing. [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2014. Available from: http://www.r-project.org.

13 Liu C, Kraja AT, Smith JA, Brody JA, Franceschini N, Bis JC, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. Nat Genet. 2016;48(10):1162–70.

14 Vandenwijngaert S, Ledsky CD, Lahrouchi N, Khan MAF, Wunderer F, Ames L, et al. Blood pressure-associated genetic variants in the natriuretic peptide receptor 1 gene modulate guanylate cyclase activity. Circ Genom Precis Med. 2019;12(8):e002472.

15 Lecluze E, Lettre G. Association analyses of predicted loss-of-function variants prioritized 15 genes as blood pressure regulators. Can J Cardiol. 2023;S0828–282X(23)01516-7.

16 Kawakami-Mori F, Nishimoto M, Reheman L, Kawarazaki W, Ayuzawa N, Ueda K, et al. Aberrant DNA methylation of hypothalamic angiotensin receptor in prenatal programmed hypertension. JCI Insight. 2018;3(21):e95625.

17 Salah N, Salem L, Taha S, Youssef M, Annaka L, Hassan S, et al. DNMT3A and TET2; potential estimates of generic DNA methylation in children and adolescents with obesity; relation to metabolic dysregulation. Horm Res Paediatr. 2022;95(1):25–34.

18 Thomas JF, Valencia-Sánchez MI, Tamburri S, Gloor SL, Rustichelli S, Godínez-López V, et al. Structural basis of histone H2A lysine 119. deubiquitination by Polycomb repressive deubiquitinase BAP1/ASXL1. Sci Adv [Internet]. 2023 Aug 9 [cited 2023 Sep 1]; 9(32). Available from: https://pubmed.ncbi.nlm.nih.gov/37556531/.

19 Dangiolo SB, Wilson A, Jobanputra V, Anyane-Yeboa K. Bohring-Opitz syndrome (BOS) with a new ASXL1 pathogenic variant: review of the most prevalent molecular and phenotypic features of the syndrome. Am J Med Genet. 2015;167A(12):3161–6.

20 Hoischen A, van Bon BWM, Rodríguez-Santiago B, Gilissen C, Vissers LELM, de Vries P, et al. De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. Nat Genet. 2011;43(8):729–31.

21 Buys E. Sips P. New insights into the role of soluble guanylate cyclase in blood pressure regulation. Curr Opin Nephrol Hypertens. 2014;23(2):135–42.

22 Kim CH, Zabetian CP, Cubells JF, Cho S, Biaggioni I, Cohen BM, et al. Mutations in the dopamine β-hydroxylase gene are associated with human norepinephrine deficiency. Am J Med Genet. 2002;108(2):140–7.

23 Frankowska M, Surówka P, Suder A, Pieniążek R, Pukło R, Jastrzębska J, et al. Treatment with dopamine β-hydroxylase (DBH) inhibitors prevents morphine use and relapse-like behavior in rats. Pharmacol Rep. 2021;73(6):1694–711.

24 Schroeder JP, Epps SA, Grice TW, Weinshenker D. The selective dopamine β-hydroxylase inhibitor nepicastat attenuates multiple aspects of cocaine-seeking behavior. Neuropsychopharmacology. 2013 38:1032 8.

Curtis