Impact of using PRS-CSx and pruning and thresholding for polygenic partitioning of apparent treatment resistant hypertension

Hannah M. Seagle[1-4,9], Jeewoo Kim[1-3,5-7], Alexis T. Akerele[1,5,6,10], VA Million Veteran Program, Adriana Hung[3,8,11], Jacklyn N. Hellwege[1,2,8,11]* and Todd L. Edwards[3,4,11]*

[1]Vanderbilt Genetics Institute, [2]Division of Genetic Medicine, [3]Department of Medicine, [4]Division of Epidemiology, [5]Division of Quantitative Science, [6]Department of Obstetrics and Gynecology, [7]Vanderbilt Medical Scientist Training Program, [8]Division of Nephrology and Hypertension, Vanderbilt University Medical Center, Nashville, TN, 37203, USA,

[9]Joseph Maxwell Cleland Atlanta VA Medical Center, Atlanta, GA, 37203, USA

[10]School of Graduate Studies, Meharry Medical College, Nashville, TN, 37208, USA

[11]VA Tennessee Valley Healthcare System (626), Nashville, TN, 37203, USA

**Contributed equally as co-senior and co-corresponding authors*

**Emails: jacklyn.hellwege@vumc.org and todd.l.edwards@vumc.org*

Apparent treatment-resistant hypertension (aTRH) is a clinically challenging condition with heterogeneous etiologies. Understanding the biological pathways underlying resistance to antihypertensive treatment could inform targeted therapeutic strategies. To evaluate how methodological choices in SNP selection influence biological inference, we applied two approaches to select aTRH-associated variants for clustering: PRS-CSx and pruning and thresholding (P&T). Using k-means clustering, we grouped aTRH-associated variants based on their association profiles across 91 cardiometabolic-related phenotypes. We then performed pathway and tissue enrichment analyses to evaluate the biological processes represented by each cluster. Both methods identified multiple genetic clusters, but the distribution of variants and biological signals differed. Clustering based on PRS-CSx produced unequally distributed clusters of SNPs and yielded limited tissue enrichment, while P&T-based clustering captured more uniform trends across cardiometabolic traits and broader tissue and pathway enrichment. These results demonstrate that methodological choices in SNP selection influence downstream clustering and biological interpretation. Despite some overlap in identified pathways and tissue enrichment, each approach identified unique biological signals, highlighting the potential of pairing polygenic methods and k-means clustering to elucidate the biological heterogeneity of aTRH and guide future mechanistic studies.

Keywords: Disease subtyping, cardiometabolic traits, polygenic clustering, resistant hypertension

1. Introduction

Elevated blood pressure (BP), or hypertension (HTN), is defined by systolic BP (SBP) ≥ 130 mmHg and/or diastolic BP (DBP) ≥ 80 mmHg. HTN contributes to over 13% of global premature deaths and is implicated in many cardiovascular diseases, including coronary artery disease, stroke, and renal damage.[1] In the United States, 34% and 44% of non-Hispanic White (NHW) and non-Hispanic

Black (NHB) adults, respectively, have HTN, which rank among the highest prevalence rates around the world.[2,3] Treatment guidelines include BP-lowering strategies like antihypertensive medications and lifestyle alterations. However, nearly 25% of individuals with HTN do not achieve target blood pressures.[4]

Apparent treatment-resistant HTN (aTRH), defined as failure to achieve BP control on three antihypertensive medications or concurrent prescriptions for four or more medications regardless of achieving control, increases risk for severe health outcomes including coronary heart disease and stroke, particularly in African ancestry populations.[4,5] Due to varying definitions of aTRH and methods of BP assessment, the estimated prevalence of aTRH has a wide range, from 8.4-50%.[6] Individuals with aTRH are at a greater risk for diabetes, chronic kidney disease, cardiovascular disease, and stroke than individuals with controlled HTN.[7]

Despite its clinical significance, the biological mechanisms underlying aTRH remain poorly understood. aTRH is a heterogeneous condition – patients may reach treatment resistance through different biological routes, such as altered renal sodium handling[8], impaired vascular tone regulation[9], or neurohormonal dysregulation.[10] Additionally, compared to NHW individuals, those who identify as Hispanic, Asian, and NHB have 10-12% lower rates of BP control[11], with some antihypertensive medications having lower efficacy in these populations as well.[12] These differences likely reflect multiple genetic and environmental conditions, underscoring the need for more nuanced classification of aTRH beyond conventional clinical definitions.

To begin addressing this complexity, we sought to apply a k-means clustering approach to identify distinct, non-overlapping clusters of genetic variants associated with aTRH. Previous studies[13] have utilized this approach for well-characterized phenotypes, such as type 2 diabetes mellitus, by clustering based on genome-wide significant ($P < 5x10^{-8}$) independent loci. Given the complexities in phenotyping for aTRH, leading to limited sample sizes and power, this strategy may fail to capture the breadth of relevant genetic variation. To address this limitation, we evaluated alternative approaches to SNP selection that capture a broader spectrum of polygenic effects: linkage disequilibrium (LD) pruning and thresholding (P&T[14]) and PRS-CSx[15], a Bayesian shrinkage method that integrates LD information across populations. We then performed k-means clustering using SNPs selected by each method, grouping variants based on their association patterns across 91 cardiometabolic phenotypes.

These clusters, which varied by SNP selection method, may represent candidate mechanistic pathways underlying treatment resistance that can be personalized. For example, one individual may have strong genetic loading for sodium handling mechanisms, while another may have elevated risk through vascular reactivity mechanisms. By performing pathway and tissue enrichment analyses, we can compare how PRS-CSx- and P&T-selected SNPs result in differences in both cluster assignment and the biological processes implicated. This framework allowed us to assess not only the biological heterogeneity of aTRH, but also how methodological choices in SNP selection influence downstream interpretation.

This study evaluates analytic approaches that provide opportunities to further understand the underlying mechanisms of treatment resistance, with the long-term goal of identifying more targeted, biology-informed therapeutic strategies. Further, these approaches may be utilized for SNP selection for other phenotypes where large numbers of genome-wide significant independent loci may be unavailable, allowing for opportunities to further understand biological mechanisms of disease that may not be possible otherwise.

2. Methods

2.1. *Study Population and Phenotype Definition*

2.1.1. *The Million Veteran Program (MVP)*

We used MVP cross-population summary statistics from an aTRH genome-wide association study (GWAS) in this analysis to select aTRH-associated loci.[16] The MVP cohort is comprised of Veteran Health Administration users, 35% of whom identify as a racial or ethnic minority, making it one of the most heterogeneous research cohorts. Genotypes for approximately 750,000 genetic markers have been measured with the Affymetrix Axiom Biobank array. Genetic variants not covered by the array were imputed using the Trans-Omics for Precision Medicine (TOPMed) reference panel.[17] The case definition for this analysis has been previously described.[18] Briefly, aTRH cases were defined using a previously validated algorithm[6], based on failure to achieve controlled BP on three antihypertensive medications, one of which must be a thiazide diuretic, or prescribed four or more medications regardless of achieving control, excluding patients with chronic kidney disease (stages 4 and 5) as well as causes of secondary HTN. Controls were defined as individuals who achieved control (<140/90 mmHg) on 1 or 2 medication classes. The total number of samples included 16,833 cases (11,762 NHW and 5,071 NHB) and 53,931 controls (42,850 NHW and 11,091 NHB). The GWAS was performed race-stratified and then meta-analyzed.

2.1.2. *Vanderbilt University Medical Center's (VUMC) BioVU*

BioVU is an institutional resource that supports large-scale genomic research at VUMC. A description of human subjects' protection employed for BioVU has been previously described.[19] BioVU genotype data of NHW and NHB adults was used for polygenic score (PS) development and p-value thresholding (described in 2.2).[20] BioVU participant DNA samples were genotyped on a custom Illumina Multi-Ethnic Genotyping Array (MEGA-ex; Illumina Inc., San Diego, CA, USA). Phenotyping of BioVU participants used the same algorithm for defining aTRH cases and controls as above. The total number of samples included 4,286 cases (3,541 NHW and 745 NHB) and 39,356 controls (33,552 NHW and 5,804 NHB).

2.2. *Defining aTRH Loci with PS Development and P-Value Thresholding*

An overview of the methods can be found in Figure 1. Due to the small number of independent genome-wide Bonferroni-corrected significant SNPs ($P < 5x10^{-8}$) from the MVP aTRH GWAS meta-analysis, we compared two methods which could be used to identify the best threshold reflecting genetic contribution to phenotypic variation, prior to clustering. The first method combined PS development with PRS-CSx[15] followed by optimal p-value thresholding.[21] The optimal p-value threshold determined by this step was then used as a p-value cutoff for SNPs that would be used downstream in the clustering analysis (described in 2.3).

PS weights were generated using the MVP NHW and NHB summary statistics with PRS-CSx[15] software and --meta flag. Using those weights, PSs at variable p-value thresholds were built using BioVU genotype data with PLINK2[22], followed by p-value thresholding (range: P=1 to $5x10^{-8}$, 16 thresholds), as described previously.[21] We then performed logistic regression analyses of the PSs

and aTRH, adjusted for age, age^2, sex, body mass index (BMI), and the first ten principal components, and determined which p-value threshold resulted in the maximal variance explained by the PS. Optimal p-value thresholds were P < 1 in the NHW population and P < 0.01 in the NHB population, as determined by maximal variance explained (0.05% and 0.14%, respectively). After filtering to SNPs with P < 0.01 in the MVP cross-population meta-analysis, we had 17,361 SNPs (aTRH SNPs) that were used for the PRS-CSx-based clustering analysis.

The second method utilized the pruning and thresholding approach (P&T[14]). We performed linkage disequilibrium (LD) pruning of SNPs at the same 16-variable p-value thresholds, using an r^2 threshold of 0.1 within a 250 kb window. Independent SNPs at each of the 16-variable p-value thresholds were used to construct PSs as described above. Logistic regression and testing for maximal variance explained were performed as above. The optimal p-value threshold was P < 0.01 for the NHW and NHB populations (maximal variance explained: 0.58% and 3.30%, respectively). We then extracted the 5,479 independent SNPs with P < 0.01 from the MVP aTRH GWAS meta-analysis summary statistics to be used for the P&T-based clustering analysis.

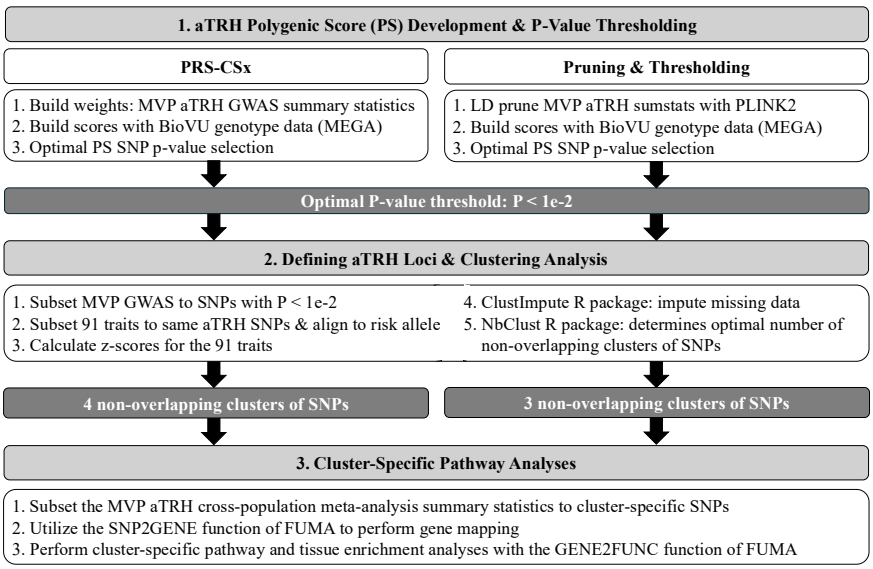

Fig. 1. Schematic overview of methods. MVP GWAS refers to the cross-population meta-analysis of aTRH. BioVU data refers to the genotype data from BioVU (MEGA). PS: Polygenic scores. LD: Linkage disequilibrium.

2.3. *Defining Clusters of aTRH SNPs with Distinct Cardiometabolic Profiles*

We adapted our approach based on a previously published unsupervised "hard clustering" method with imputation of missing traits.[13] Briefly, we identified 91 cardiometabolic-related traits[23-50] that are associated with aTRH, HTN, or BP. Publicly available summary statistics were obtained and the effect estimates were aligned to the aTRH risk allele from the cross-population MVP meta-analysis. Z-scores were then calculated for each trait. K-means clustering of the aTRH SNPs and 91 traits was performed using the ClustImpute[51] and NbClust[52] R packages. The ClustImpute package imputes missing data and the NbClust package determines the optimal number of clusters. ClustImpute was performed with three clusters (minimum). NbClust parameters were based on those used previously[13] which included a minimum of three and maximum of 15 clusters, distance was set to Euclidean, all indices were used, and the significance value for Beale's index was set to 0.1. Clustering parameters were identical for both the PRS-CSx and P&T approaches.

2.4. *Cluster-Specific Pathway Analyses*

To investigate the biological pathways and tissue-specific expression patterns associated with each cluster, we used the SNP2GENE and GENE2FUNC functions of FUMA[53] (Functional Mapping and Annotation of GWAS). We input cluster-specific subsets of the aTRH MVP cross-population meta-analysis summary statistics into the SNP2GENE function of FUMA to perform gene mapping and then used the GENE2FUNC function for functional enrichment analyses, including pathway and tissue enrichment based on Genotype-Tissue Expression (GTEx) version 8 profiles. Parameters for FUMA analyses were identical for both SNP selection methods.

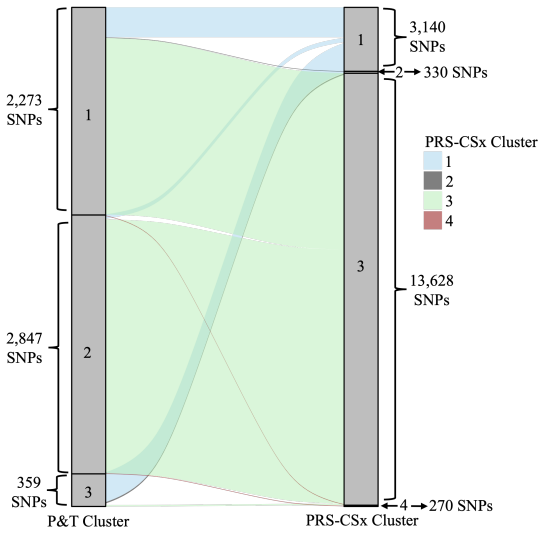

Fig. 2. Alluvial plot showing the differences in distribution of SNPs between P&T Clusters and PRS-CSx Clusters. The number of SNPs per cluster is shown next to each bracket. SNPs that were not shared between both methods were excluded for plotting.

3. Results

We compared the PRS-CSx and P&T methods as two different SNP selection approaches to k-means clustering for phenotypes with limited numbers of independent genome-wide significant loci, with the overarching goal of determining the genetic underpinnings of phenotypic heterogeneity in aTRH. For PRS-CSx, the optimal p-value thresholds were $P < 1$ in the NHW population and $P < 0.01$ in the NHB population (maximal variance explained: 0.05% and 0.14%, respectively). For P&T, the optimal threshold was $P < 0.01$ for the NHW and NHB populations (maximal variance explained: 0.58% and 3.30%, respectively). We then performed two sets of clustering analyses, using the 17,367 SNPs selected from PRS-CSx and 5,479 from P&T (SNPs with $P < 0.01$), aligned to the aTRH risk allele, across 91 cardiometabolic traits. The optimal number of clusters was determined to be four for the PRS-CSx-based method and three for P&T-based method, which represented non-overlapping subsets of SNPs with similar cardiometabolic profiles through application of an unsupervised "hard clustering" method with imputation of missing trait associations.[13] The distribution of SNPs by cluster and method is shown in Figure 2. Clustering based on P&T resulted in two near-equally proportioned clusters out of a total of three, whereas PRS-CSx-based clustering resulted in four clusters, with one including over four times as many SNPs than the next largest cluster (Figure 2). Most SNPs in P&T-based Clusters 1 and 2 were captured in the PRS-CSx-based Cluster 3, where P&T Cluster 3 mostly overlapped with PRS-CSx Cluster 1 (Figure 2).

3.1. *Identifying Biologically Meaningful Clusters of aTRH SNPs*

Each of the non-overlapping SNP clusters derived from both PRS-CSx and P&T showed a unique pattern of association across the 91 cardiometabolic traits (Figure 3), reflecting the genetic and mechanistic heterogeneity of aTRH.

PRS-CSx Cluster 1 showed mixed positive and negative associations across traits, with strong positive associations for SBP, DBP, and pulse pressure, as well as traits related to metabolic health, including fasting glucose, two-hour glucose, and most of the anthropometric and central adipose tissue volume traits. Moderately negative associations were observed for certain cardiac structural/functional traits and some blood cell traits, including hematocrit, hemoglobin concentration, mean corpuscular hemoglobin, and mean corpuscular volume (Figure 2A). PRS-CSx Cluster 2 was characterized by increased associations of all the anthropometric measures, all but one liver-related trait, and many of the lifestyle, sleep and BP traits, as well as cardiac structural/functional phenotypes (Figure 2A). Negative associations appeared for multiple blood cell traits, circulating plasma lipids, women's health and endocrine traits, including age at menarche, sex hormone binding globulin, and plasma renin. PRS-CSx Cluster 3 showed consistently weak or near-zero associations across most traits, except for strong negative associations with BP traits. Cluster 4 displayed nearly inverse patterns of Cluster 2, with negative associations with anthropometric, liver, lifestyle, BP, and cardiac traits, but strong positive associations with blood cell traits and moderate positive associations with women's health and lipid traits (Figure 2A).

P&T Cluster 1 displayed broadly increased associations across nearly all glycemic, anthropometric, adiposity, BP and heart rate, liver, blood cell, lifestyle and women's health traits, with only a few exceptions: decreased associations for HDL, select blood cell traits, testosterone, and sex hormone binding globulin (Figure 2B). P&T Cluster 2 showed the opposite pattern, with negative associations across nearly all glycemic, adiposity, lipid, BP, heart rate, and blood cell traits, while cardiac structural/functional traits remained weakly associated (Figure 2B). Cluster 3 was characterized by strong associations with cardiac structural/functional traits, largely in the negative direction, coupled with positive associations for BP traits and negative associations for adiposity, lipid, and blood cell traits.

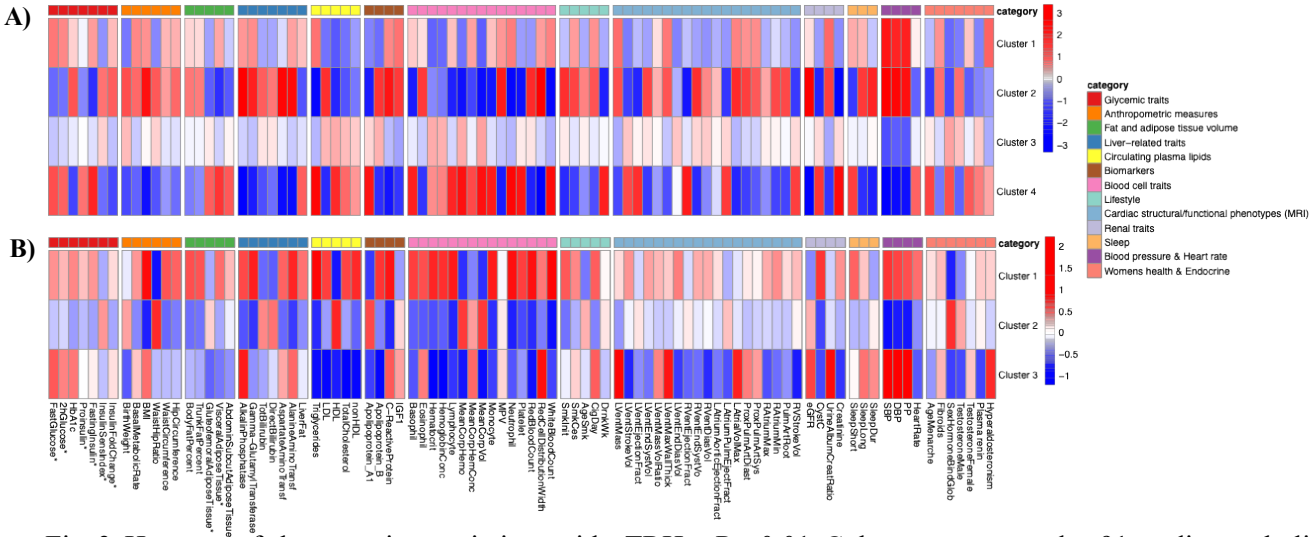

Fig. 3. Heatmap of cluster-trait associations with aTRH at $P < 0.01$. Columns correspond to 91 cardiometabolic traits, grouped by category. Rows correspond to the clusters. Color intensity reflects the direction and magnitude of association (z-score aligned to the aTRH risk allele), with blue indicating negative and red indicating positive associations. Breaks for the color scale were calculated using a quantile-based method. *Traits adjusted for BMI. **A)** PRS-CSx-based results. **B)** P&T-based results.

Overall, both methods produced clusters with nearly inverse association patterns. Compared to PRS-CSx-derived clusters, P&T clusters tended to capture broader and more uniformly positive or

negative trends. These differences indicate that SNP selection strategy can shape both cluster composition and the biological pathways implicated, with PRS-CSx highlighting more subtle contrasts between variant groups and P&T emphasizing broader cardiometabolic signatures.

3.2. *Cluster-Specific Pathways*

The GENE2FUNC tissue specificity analyses revealed significant enrichment ($P_{Bonferroni} < 0.05$) for both PRS-CSx- and P&T-based Clusters, though the patterns differed between methods (Table 1). PRS-CSx Cluster 1 showed enrichment in the bladder, adrenal gland, and heart (left ventricle), while Cluster 3 was enriched for the bladder and adrenal gland (Table 1). P&T Clusters revealed tissue enrichment for more total tissues, including whole blood for Cluster 1 and the bladder, colon, uterus, prostate, and kidney for Cluster 2.

Table 1. Cluster-specific FUMA GENE2FUNC results for tissue specificity. Only significantly ($P_{Bonferroni} < 0.05$) enriched tissues are shown.

Method	Cluster	Tissue	$P_{Bonferroni}$
PRS-CSx	1	Bladder	5.15×10^{-7}
		Adrenal gland	7.82×10^{-7}
		Heart left ventricle	2.20×10^{-4}
	3	Bladder	5.90×10^{-5}
		Adrenal gland	1.10×10^{-3}
P&T	1	Whole blood	2.46×10^{-3}
	2	Bladder	4.98×10^{-9}
		Colon	2.92×10^{-7}
		Uterus	8.94×10^{-4}
		Prostate	4.66×10^{-2}
		Kidney	4.70×10^{-2}

Table 2. Cluster-specific GENE2FUNC results for significantly enriched gene sets ($P_{FDR} < 0.05$). The top 4 of over 20 significant pathways are presented for PRS-CSx Cluster 1 and P&T Cluster 2.

Method	Cluster	Gene Ontology Biological Process	P_{FDR}
PRS-CSx	1	Anterior-posterior pattern specification	7.33×10^{-3}
		Glucocorticoid biosynthetic process	1.87×10^{-2}
		Embryonic skeletal system development	1.87×10^{-2}
		Membrane repolarization during cardiac muscle cell action potential	3.61×10^{-2}
	3	Homophilic cell adhesion via plasma membrane adhesion molecules	1.32×10^{-3}
		Embryonic skeletal system development	3.08×10^{-2}
		Glucocorticoid biosynthetic process	4.33×10^{-2}
P&T	1	Ubiquitin dependent endoplasmic reticulum-associated degradation	2.91×10^{-2}
		Endoplasmic reticulum-associated degradation	3.40×10^{-2}
	2	Embryonic skeletal system development	3.92×10^{-8}
		Embryonic skeletal system morphogenesis	2.41×10^{-7}
		Skeletal system development	2.41×10^{-7}
		Anterior-posterior pattern specification	3.50×10^{-7}

The GENE2FUNC analysis also revealed both convergences and divergences between methods for significant enrichment ($P_{FDR} < 0.05$) of gene ontology biological pathways (Table 2). The top four PRS-CSx-based Cluster 1 pathways were enriched for developmental and regulatory processes, including anterior-posterior pattern specification ($P=7.33 \times 10^{-3}$), glucocorticoid biosynthetic process ($P=1.87 \times 10^{-2}$), embryonic skeletal system development ($P=1.87 \times 10^{-2}$), and membrane repolarization during cardiac muscle cell action potential ($P=3.61 \times 10^{-2}$). PRS-CSx Cluster 3 was enriched for homophilic cell adhesion via plasma membrane adhesion molecules ($P=1.32 \times 10^{-3}$),

embryonic skeletal system development ($P=3.08x10^{-2}$), and glucocorticoid biosynthetic process ($P=4.33x10^{-2}$). In contrast, P&T-based Cluster 1 was enriched for endoplasmic reticulum-associated degradation pathways ($P=2.91x10^{-2}$ and $P=3.40x10^{-2}$), while the top four significant pathways for P&T Cluster 2 were strongly enriched for a broader set of developmental processes, including embryonic skeletal system development and morphogenesis ($P=3.92x10^{-8}$ and $P=2.41x10^{-7}$), skeletal system development ($P=2.41x10^{-7}$), and anterior-posterior pattern specification ($P=3.50x10^{-7}$).

4. Discussion

This study provides insights into the biological heterogeneity of aTRH by applying a clustering framework to genetic variants using two different SNP selection strategies: PRS-CSx and P&T. Each approach identified distinct, non-overlapping clusters of variants with unique cardiometabolic association profiles, supporting the hypothesis that aTRH is a disorder influenced by multiple pathophysiological processes.

Despite differences in SNP selection, both methods revealed clusters enriched for multiple groups of cardiometabolic traits. However, notable divergences were observed. PRS-CSx selected approximately five times the number of SNPs selected by P&T and yielded four clusters with more nuanced traits profiles, whereas P&T produced three clusters characterized by broader and more interpretable trends. These methodological differences highlight how SNP selection strategy can influence both the composition of clusters and downstream biological interpretations.

Across methods, clusters enriched for adiposity and glycemic traits (PRS-CSx Cluster 1 and P&T Cluster 1) suggest a metabolic energy-driven pathway to treatment resistance. These findings are supported by the results of another study that reported a gradual increase in the prevalence of resistant HTN with BMI in overweight patients.[54] Increased adipose tissue is related to increased activity of the renin-angiotensin-aldosterone system (RAAS), which is crucial in BP regulation. In particular, increased adiposity can lead to increased expression of angiotensin type 1 and 2 receptors, elevated circulating angiotensin II, angiotensin-converting enzyme, and aldosterone levels.[55] In obesity specifically, the increased activity of the RAAS can increase sodium retention, leading to volume expansion and HTN.[56]

Positive associations between multiple biomarkers and aTRH are well supported by current literature. One study reported increased risk of aTRH with biomarkers such as interleukin-6 (IL-6), tumor necrosis factor-α (TNF-α), and transforming growth factor-β (TGF-β).[57] These results align with the strong positive association of multiple biomarkers observed in PRS-CSx Cluster 2 and P&T Cluster 3. Other positive associations shared between these clusters include sleep-related traits. Sleep disturbances, obstructive sleep apnea in particular, are highly prevalent (60-71%) among individuals with aTRH.[58] In particular, one study found sleep apnea was significantly more common in individuals with resistant HTN compared to individuals with controlled HTN.[59]

Lastly, positive associations with blood cell traits (PRS-CSx Cluster 4 and P&T Cluster 1) may indicate a vascular-reactivity related mechanism. In a large Japanese cohort over a 40-year follow-up period, absolute neutrophil and white blood cell count have been shown to predict new onset HTN.[60] Further, one study found that compared to both normotensive and controlled hypertensive patients, individuals with aTRH had a significantly higher neutrophil count and neutrophil/lymphocyte ratio.[61] Positive associations with white blood cell, neutrophil and lymphocyte count for both clusters support these findings.

Pathway and tissue enrichment analyses reinforced the unique biological profiles of clusters identified by each method. For PRS-CSx, Cluster 1 was significantly enriched in gene sets related to anterior-posterior pattern specification, glucocorticoid biosynthesis, and embryonic skeletal system development, suggesting early development and hormonal regulation pathways. Glucocorticoid receptor signaling has been implicated in HTN, particularly through sodium and potassium handling and BP regulation.[62] One study reported 51% of patients who carried a heterozygous or homozygous glucocorticoid receptor mutation had HTN.[62] Tissue specificity analysis of PRS-CSx Cluster 1 further revealed enrichment in the adrenal gland, heart, and bladder, aligning with traits tied to volume regulation and hormonal control. PRS-CSx Cluster 3 was enriched for biological processes related to homophilic cell adhesion, embryonic development, and glucocorticoid biosynthesis. One study[63] examined the relationship between adhesion molecules and aTRH and found that endothelial-selectin was a significant indicator of aTRH.

In contrast, P&T identified clusters enriched in a broader range of tissues, including whole blood, bladder, colon, uterus, prostate, and kidney. Functionally, P&T Cluster 1 was linked to protein quality-control pathways including ubiquitin-dependent and endoplasmic reticulum-associated degradation. P&T Cluster 2 was highly significantly enriched for skeletal system development and morphogenesis pathways, overlapping partially with the PRS-CSx findings, but with stronger statistical support, and implicated bladder tissue as well. Together, these enrichment patterns indicate that while both methods highlight developmental and hormonal pathways, PRS-CSx emphasized neurohormonal and cardiac regulation, whereas P&T uncovered stronger signals in developmental and protein homeostasis processes, as well as a broader set of tissues, including reproductive and gastrointestinal systems.

These results demonstrate that both methods recover biologically meaningful clusters but may emphasize different aspects of underlying biology. In this analysis, PRS-CSx tended to highlight pathways related to developmental, endocrine, and cardiac processes, whereas P&T captured a wider range of tissue and pathway enrichments, including immune, reproductive, and protein degradation pathways. The partial overlap between enriched skeletal-related processes across both methods suggests stronger biological themes, but each approach also identified distinct biological signals. These results reinforce that the choice of SNP selection method influences both cluster composition and downstream enrichment analyses. Despite the differences, the overlap in developmental and hormonal pathways across methods provides converging evidence for their role in aTRH, while method-specific findings may reflect complementary aspects of the underlying biology. Taken together, the findings of this study underscore the biological heterogeneity underlying aTRH, suggesting that distinct clusters of genetic variants act through different cardiometabolic routes to contribute to HTN that is refractory to standard treatments.

The identification of mechanistically distinct genetic clusters has implications for precision therapeutics in aTRH. By understanding the unique pathway through which a patient becomes resistant, for example, metabolic overload vs. vascular stiffness vs. hematologic dysfunction, clinicians may be able to tailor interventions to target the underlying biology. Stratification by genetic cluster could test whether pathway-informed interventions improve BP control, which may have potential to overcome the limitations of current treatment guidelines and address population differences in HTN control and efficacy of antihypertensives.

While promising, this work has some limitations. First, the analyses were performed in NHW and NHB cohorts, limiting generalizability to other populations. Given known differences in HTN prevalence and treatment response, population-specific clustering approaches should be further

explored. Additionally, the clustering approach is based solely on genetic association profiles. Future studies including transcriptomic or proteomic data may provide additional biologically relevant information. Finally, SNP selection methods clearly influenced results, underscoring the need for continued methodological evaluation in clustering-based approaches. The difference in number of SNPs along with the SNPs selected could both contribute to these influences.

Together, our findings underscore the potential of genetic clustering to disentangle the biological and mechanistic heterogeneity underlying aTRH and demonstrate that SNP clustering provides a powerful framework to identify distinct mechanistic pathways. Both PRS-CSx and P&T produced biologically interpretable clusters but emphasized different aspects of the genetic architecture. In this analysis, P&T demonstrated clearer interpretability and stronger tissue and pathway signals, indicating that it may be a preferred SNP selection strategy. This work also lays the foundation for use of clustering methods for phenotypes with a limited number of independent genome-wide significant variants. This study represents a step toward precision medicine in aTRH by linking genetic architecture to putative mechanistic pathways, and ultimately, a shift from broad, population-level HTN treatment strategies toward more tailored interventions.

5. Acknowledgements

Vanderbilt University Medical Center's BioVU is supported by institutional funding and by the Vanderbilt CTSA grant UL1 TR000445 from NCATS/NIH. HMS and JK were funded by TL1TR002244. HMS was also funded by T32GM145734-01. ATA was funded by T32HL007737. We thank the Veterans who participated in the Million Veteran Program and the participants of Vanderbilt University Medical Center's BioVU. This research is based on data from the Million Veteran Program, Office of Research and Development, and Veterans Health Administration (MVP000 and MVP001/002). This publication does not represent the views of the Department of Veterans Affairs or the United States Government.

VA Million Veteran Program
Core Acknowledgements for Publications
June 2025

MVP Program Office

- Sumitra Muralidhar, Ph.D., Program Director
  US Department of Veterans, 810 Vermont Avenue NW, Washington, DC 20420
- Jennifer Moser, Ph.D., Associate Director, Scientific Programs
  US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420
- Jennifer E. Deen, B.S., Associate Director, Cohort & Public Relations
  US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

MVP Executive Committee

- Co-Chair: Philip S. Tsao, Ph.D., Program Director
  VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304
- Co-Chair: Sumitra Muralidhar, Ph.D.
  US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420
- J. Michael Gaziano, M.D., M.P.H
  VA Boston Healthcare System, 105 S. Huntington Avenue, Boston, MA 02130
- Elizabeth Hauser, Ph.D.

Durham VA Medical Center, 508 Fulton Street, Durham, NC 27705
- Amy Kilbourne, Ph.D., M.P.H.
  VA HSR&D, 2215 Fuller Road, Ann Arbor, MI 48105
- Michael Matheny, M.D., M.S., M.P.H
  VA Tennessee Valley Healthcare System, 1310 24th Ave. South, Nashville, TN 37212
- Dave Oslin, M.D
  Philadelphia VA Medical Center, 3900 Woodland Avenue, Philadelphia, PA 19104
- Deepak Voora, M.D.
  Durham VA Medical Center, 508 Fulton Street, Durham, NC 27705

MVP Co-Principal Investigators
- J. Michael Gaziano, M.D., M.P.H
  VA Boston Healthcare System, 105 S. Huntington Avenue, Boston, MA 02130
- Philip S. Tsao, Ph.D.
  VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304

MVP Core Operations
- Jessica V. Brewer, M.P.H., Director, MVP Cohort Operations
  VA Boston Healthcare System, 105 S. Huntington Avenue, Boston, MA 02130
- Mary T. Brophy M.D., M.P.H., Director, VA Central Biorepository
  VA Boston Healthcare System, 105 S. Huntington Avenue, Boston, MA 02130
- Kelly Cho, M.P.H., Ph.D., Director, MVP Phenomics
  VA Boston Healthcare System, 105 S. Huntington Avenue, Boston, MA 02130
- Lori Churby, B.S., Director, MVP Regulatory Affairs
  VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304
- Scott L. DuVall, Ph.D., Director, VA Informatics and Computing Infrastructure (VINCI)
  VA Salt Lake City Health Care System, 500 Foothill Drive, Salt Lake City, UT 84148
- Saiju Pyarajan, Ph.D., Director, Data and Computational Sciences
  VA Boston Healthcare System, 105 S. Huntington Avenue, Boston, MA 02130
- Robert Ringer, Pharm.D., Director, VA Albuquerque Central Biorepository
  New Mexico VA Health Care System, 1501 San Pedro Drive SE, Albuquerque, NM 87108
- Luis E. Selva, Ph.D., Director, MVP Biorepository Coordination
  VA Boston Healthcare System, 105 S. Huntington Avenue, Boston, MA 02130
- Shapoor (Alex) Shayan, M.S., Director, MVP PRE Informatics
  VA Boston Healthcare System, 105 S. Huntington Avenue, Boston, MA 02130
- Brady Stephens, M.S., Principal Investigator, MVP Cohort Development and Management
  VA Boston Healthcare System, 105 S. Huntington Avenue, Boston, MA 02130
- Stacey B. Whitbourne, Ph.D., Director, MVP Cohort Development and Management
  VA Boston Healthcare System, 105 S. Huntington Avenue, Boston, MA 02130

References

1 Lawes, C. M., Vander Hoorn, S., Rodgers, A. & International Society of, H. Global burden of blood-pressure-related disease, 2001. *Lancet* **371**, 1513-1518 (2008). https://doi.org:10.1016/S0140-6736(08)60655-8

2 Roger, V. L. *et al.* Executive summary: heart disease and stroke statistics--2012 update: a report from the American Heart Association. *Circulation* **125**, 188-197 (2012). https://doi.org:10.1161/CIR.0b013e3182456d46

3 Whelton, P. K. *et al.* 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension* **71**, 1269-1324 (2018). https://doi.org:10.1161/HYP.0000000000000066

4 Muntner, P. *et al.* Potential U.S. Population Impact of the 2017 ACC/AHA High Blood Pressure Guideline. *J Am Coll Cardiol* **71**, 109-118 (2018). https://doi.org:10.1016/j.jacc.2017.10.073

5 Noubiap, J. J. *et al.* Global prevalence of resistant hypertension: a meta-analysis of data from 3.2 million patients. *Heart* **105**, 98-105 (2019). https://doi.org:10.1136/heartjnl-2018-313599

6 Shuey, M. M. *et al.* Characteristics and treatment of African-American and European-American patients with resistant hypertension identified using the electronic health record in an academic health centre: a case-control study. *BMJ Open* **8**, e021640 (2018). https://doi.org:10.1136/bmjopen-2018-021640

7 Irvin, M. R. *et al.* Apparent treatment-resistant hypertension and risk for stroke, coronary heart disease, and all-cause mortality. *J Am Soc Hypertens* **8**, 405-413 (2014). https://doi.org:10.1016/j.jash.2014.03.003

8 Khawaja, Z. & Wilcox, C. S. Role of the kidneys in resistant hypertension. *Int J Hypertens* **2011**, 143471 (2011). https://doi.org:10.4061/2011/143471

9 Ma, J. *et al.* Signaling pathways in vascular function and hypertension: molecular mechanisms and therapeutic interventions. *Signal Transduct Target Ther* **8**, 168 (2023). https://doi.org:10.1038/s41392-023-01430-7

10 Tsioufis, C. *et al.* Pathophysiology of resistant hypertension: the role of sympathetic nervous system. *Int J Hypertens* **2011**, 642416 (2011). https://doi.org:10.4061/2011/642416

11 Abrahamowicz, A. A., Ebinger, J., Whelton, S. P., Commodore-Mensah, Y. & Yang, E. Racial and Ethnic Disparities in Hypertension: Barriers and Opportunities to Improve Blood Pressure Control. *Curr Cardiol Rep* **25**, 17-27 (2023). https://doi.org:10.1007/s11886-022-01826-x

12 Jamerson, K. & DeQuattro, V. The impact of ethnicity on response to antihypertensive therapy. *Am J Med* **101**, 22S-32S (1996). https://doi.org:10.1016/s0002-9343(96)00265-3

13 Suzuki, K. *et al.* Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* **627**, 347-357 (2024). https://doi.org:10.1038/s41586-024-07019-6

14 Prive, F., Vilhjalmsson, B. J., Aschard, H. & Blum, M. G. B. Making the Most of Clumping and Thresholding for Polygenic Scores. *Am J Hum Genet* **105**, 1213-1221 (2019). https://doi.org:10.1016/j.ajhg.2019.11.001

15 Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat Genet* **54**, 573-580 (2022). https://doi.org:10.1038/s41588-022-01054-7

16 Irvin, M. R. *et al.* Genome-Wide Association Study of Apparent Treatment-Resistant Hypertension in the CHARGE Consortium: The CHARGE Pharmacogenetics Working Group. *Am J Hypertens* **32**, 1146-1153 (2019). https://doi.org:10.1093/ajh/hpz150

17 Gaziano, J. M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* **70**, 214-223 (2016). https://doi.org:10.1016/j.jclinepi.2015.09.016

18 Breeyear, J. H., Shuey, M. M., Edwards, T. L. & Hellwege, J. N. Blood Pressure Polygenic Scores Are Associated With Apparent Treatment-Resistant Hypertension. *Circ Genom Precis Med* **15**, e003554 (2022). https://doi.org:10.1161/CIRCGEN.121.003554

19 Pulley, J., Clayton, E., Bernard, G. R., Roden, D. M. & Masys, D. R. Principles of human subjects protections applied in an opt-out, de-identified biobank. *Clin Transl Sci* **3**, 42-48 (2010). https://doi.org:10.1111/j.1752-8062.2010.00175.x

20 Armstrong, N. D. *et al.* Whole genome sequence analysis of apparent treatment resistant hypertension status in participants from the Trans-Omics for Precision Medicine program. *Front Genet* **14**, 1278215 (2023). https://doi.org:10.3389/fgene.2023.1278215

21 Hellwege, J. N. *et al.* Predictive models for abdominal aortic aneurysms using polygenic scores and PheWAS-derived risk factors. *Pac Symp Biocomput* **28**, 425-436 (2023).

22 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015). https://doi.org:10.1186/s13742-015-0047-8

23 Chen, J. *et al.* The trans-ancestral genomic architecture of glycemic traits. *Nat Genet* **53**, 840-860 (2021). https://doi.org:10.1038/s41588-021-00852-9

24 Broadaway, K. A. *et al.* Loci for insulin processing and secretion provide insight into type 2 diabetes risk. *Am J Hum Genet* **110**, 284-299 (2023). https://doi.org:10.1016/j.ajhg.2023.01.002

25 Williamson, A. *et al.* Genome-wide association study and functional characterization identifies candidate genes for insulin-stimulated glucose uptake. *Nat Genet* **55**, 973-983 (2023). https://doi.org:10.1038/s41588-023-01408-9

26 Warrington, N. M. *et al.* Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat Genet* **51**, 804-814 (2019). https://doi.org:10.1038/s41588-019-0403-1

27 Sudlow, C. *et al.* 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**, e1001779 (2015). https://doi.org:10.1371/journal.pmed.1001779

28 Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415-1424 (2021). https://doi.org:10.1038/s41588-021-00931-x

29 Sinnott-Armstrong, N. *et al.* Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat Genet* **53**, 185-194 (2021). https://doi.org:10.1038/s41588-020-00757-z

30 Liu, Y. *et al.* Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning. *Elife* **10** (2021). https://doi.org:10.7554/eLife.65554

31 Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675-679 (2021). https://doi.org:10.1038/s41586-021-04064-3

32 Keaton, J. M. *et al.* Genome-wide analysis in over 1 million individuals of European ancestry yields improved polygenic risk scores for blood pressure traits. *Nat Genet* **56**, 778-791 (2024). https://doi.org:10.1038/s41588-024-01714-w

33 Chen, M. H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* **182**, 1198-1213 e1114 (2020). https://doi.org:10.1016/j.cell.2020.06.045

34 Saunders, G. R. B. *et al.* Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature* **612**, 720-724 (2022). https://doi.org:10.1038/s41586-022-05477-4

35 Khurshid, S. *et al.* Clinical and genetic associations of deep learning-derived cardiac magnetic resonance-based left ventricular mass. *Nat Commun* **14**, 1558 (2023). https://doi.org:10.1038/s41467-023-37173-w

36 Pirruccello, J. P. *et al.* Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nat Commun* **11**, 2254 (2020). https://doi.org:10.1038/s41467-020-15823-7

37 Schmidt, A. F. *et al.* Druggable proteins influencing cardiac structure and function: Implications for heart failure therapies and cancer cardiotoxicity. *Sci Adv* **9**, eadd4984 (2023). https://doi.org:10.1126/sciadv.add4984

38 Aung, N. *et al.* Genome-Wide Analysis of Left Ventricular Maximum Wall Thickness in the UK Biobank Cohort Reveals a Shared Genetic Background With Hypertrophic Cardiomyopathy. *Circ Genom Precis Med* **16**, e003716 (2023). https://doi.org:10.1161/CIRCGEN.122.003716

39 Pirruccello, J. P. *et al.* Genetic analysis of right heart structure and function in 40,000 people. *Nat Genet* **54**, 792-803 (2022). https://doi.org:10.1038/s41588-022-01090-3

40 Ahlberg, G. *et al.* Genome-wide association study identifies 18 novel loci associated with left atrial volume and function. *Eur Heart J* **42**, 4523-4534 (2021). https://doi.org:10.1093/eurheartj/ehab466

41 Liu, H. *et al.* Epigenomic and transcriptomic analyses define core cell types, genes and targetable mechanisms for kidney disease. *Nat Genet* **54**, 950-962 (2022). https://doi.org:10.1038/s41588-022-01097-w

42 Agrawal, S. *et al.* Inherited basis of visceral, abdominal subcutaneous and gluteofemoral fat depots. *Nat Commun* **13**, 3771 (2022). https://doi.org:10.1038/s41467-022-30931-2

43 Teumer, A. *et al.* Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun* **10**, 4130 (2019). https://doi.org:10.1038/s41467-019-11576-0

44 Dashti, H. S. *et al.* Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *Nat Commun* **10**, 1100 (2019). https://doi.org:10.1038/s41467-019-08917-4

45 van de Vegte, Y. J. *et al.* Genetic insights into resting heart rate and its role in cardiovascular disease. *Nat Commun* **14**, 4646 (2023). https://doi.org:10.1038/s41467-023-39521-2

46 Day, F. R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet* **49**, 834-841 (2017). https://doi.org:10.1038/ng.3841

47 Kim, J. *et al.* Genome-wide meta-analysis identifies novel risk loci for uterine fibroids within and across multiple ancestry groups. *Nat Commun* **16**, 2273 (2025). https://doi.org:10.1038/s41467-025-57483-5

48 Leinonen, J. T. *et al.* Genetic analyses implicate complex links between adult testosterone levels and health and disease. *Commun Med (Lond)* **3**, 4 (2023). https://doi.org:10.1038/s43856-022-00226-0

49 Folkersen, L. *et al.* Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab* **2**, 1135-1148 (2020). https://doi.org:10.1038/s42255-020-00287-2

50 Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat Genet* **50**, 906-908 (2018). https://doi.org:10.1038/s41588-018-0144-6

51 Pfaffel, O. CLUSTIMPUTE: AN R PACKAGE FOR K-MEANS CLUSTERING WITHBUILD-IN MISSING DATA IMPUTATION. *Preprint* (2020). https://doi.org:10.13140/RG.2.2.20143.36007

52 Charrah, M., Ghazzali N., Boiteau, V., Niknafs A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software* **61**, 1-36 (2014). https://doi.org:https://doi.org/10.18637/jss.v061.i06

53 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017). https://doi.org:10.1038/s41467-017-01261-5

54 Haddadin, F. *et al.* The prevalence and predictors of resistant hypertension in high-risk overweight and obese patients: A cross-sectional study based on the 2017 ACC/AHA guidelines. *J Clin Hypertens (Greenwich)* **21**, 1507-1515 (2019). https://doi.org:10.1111/jch.13666

55 Hall, J. E., do Carmo, J. M., da Silva, A. A., Wang, Z. & Hall, M. E. Obesity-induced hypertension: interaction of neurohumoral and renal mechanisms. *Circ Res* **116**, 991-1006 (2015). https://doi.org:10.1161/CIRCRESAHA.116.305697

56 Lohmeier, T. E. & Iliescu, R. The sympathetic nervous system in obesity hypertension. *Curr Hypertens Rep* **15**, 409-416 (2013). https://doi.org:10.1007/s11906-013-0356-1

57 Chen, J. *et al.* Inflammation and Apparent Treatment-Resistant Hypertension in Patients With Chronic Kidney Disease. *Hypertension* **73**, 785-793 (2019). https://doi.org:10.1161/HYPERTENSIONAHA.118.12358

58 Carnethon, M. R. & Johnson, D. A. Sleep and Resistant Hypertension. *Curr Hypertens Rep* **21**, 34 (2019). https://doi.org:10.1007/s11906-019-0941-z

59 Bhandari, S. K. *et al.* Comparisons of sleep apnoea rate and outcomes among patients with resistant and non-resistant hypertension. *Respirology* **21**, 1486-1492 (2016). https://doi.org:10.1111/resp.12840

60 Tian, N., Penman, A. D., Mawson, A. R., Manning, R. D., Jr. & Flessner, M. F. Association between circulating specific leukocyte types and blood pressure: the atherosclerosis risk in communities (ARIC) study. *J Am Soc Hypertens* **4**, 272-283 (2010). https://doi.org:10.1016/j.jash.2010.09.005

61 Belen, E., Sungur, A., Sungur, M. A. & Erdogan, G. Increased Neutrophil to Lymphocyte Ratio in Patients With Resistant Hypertension. *J Clin Hypertens (Greenwich)* **17**, 532-537 (2015). https://doi.org:10.1111/jch.12533

62 Verouti, S., Hummler, E. & Vanderriele, P. E. Role of glucocorticoid receptor mutations in hypertension and adrenal gland hyperplasia. *Pflugers Arch* **474**, 829-840 (2022). https://doi.org:10.1007/s00424-022-02715-6

63 de Faria, A. P. *et al.* Deregulation of Soluble Adhesion Molecules in Resistant Hypertension and Its Role in Cardiovascular Remodeling. *Circ J* **80**, 1196-1201 (2016). https://doi.org:10.1253/circj.CJ-16-0058