TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

PHENOTYPIC AND GENETIC SUBTYPING OF HYPERTENSION

Toward personalized hypertension care

Felix Vaura

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1655 | MEDICA – ODONTOLOGICA | TURKU 2022

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

PHENOTYPIC AND GENETIC SUBTYPING OF HYPERTENSION

Toward personalized hypertension care

Felix Vaura

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1655 | MEDICA – ODONTOLOGICA | TURKU 2022

University of Turku

Faculty of Medicine
Department of Internal Medicine
Doctoral Programme in Clinical Research

Supervised by

Professor Teemu Niiranen, MD, PhD
Department of Clinical Medicine
University of Turku
Department of Public Health Solutions
Finnish Institute for Health and Welfare
Turku, Finland

Associate Professor Leo Lahti, DSc
Department of Computing
University of Turku
Turku, Finland

Reviewed by

Assistant Professor Georg Ehret, MD
Division of Cardiology
University Hospitals Geneva
Geneva, Switzerland

Associate Professor Matti Pirinen, PhD
Institute for Molecular Medicine Finland (FIMM)
Department of Mathematics and Statistics
Department of Public Health
University of Helsinki
Helsinki, Finland

Opponent

Professor J. Gustav Smith, MD, PhD
Wallenberg Laboratory
Department of Molecular and Clinical Sciences
Institute of Medicine
Gothenburg University
Gothenburg, Sweden

Cover Image: iStock.com/portfolio/malerapaso

ISBN 978-951-29-8956-0 (PRINT)
ISBN 978-951-29-8957-7 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Painosalama, Turku, Finland 2022

To Ada
いつもありがとう

UNIVERSITY OF TURKU
Faculty of Medicine
Department of Clinical Medicine
Internal Medicine
FELIX VAURA: Phenotypic and Genetic Subtyping of Hypertension –
Toward personalized hypertension care
Doctoral Dissertation, 141 pp.
Doctoral Programme in Clinical Research
September 2022

ABSTRACT

Current knowledge of phenotypic and genotypic hypertension risk factors has not been effectively translated into personalized hypertension care. The aim of this thesis was to explore hypertension subtyping by applying publicly available supervised and unsupervised subtyping algorithms to large datasets with extensive phenotyping and genotyping.

This thesis included participants from two large Finnish studies: 32,442 from FINRISK and 218,792 from FinnGen. FINRISK is a cross-sectional population survey carried out every five years on risk factors for chronic, non-communicable diseases. FinnGen is a public-private partnership research project combining imputed genotype data from biobanks, patient cohorts, and prospective epidemiological surveys. Because every Finnish citizen is linked to health registers via a personal identity code, accurate follow-up is possible for all major end points, including hypertension and cardiovascular disease. In addition, we used publicly available genome-wide association data from several large-scale studies, including the UK Biobank.

In FINRISK, we observed a phenotypic hypertension subgroup characterized by high blood sugar and elevated body mass index, conferring an increased risk for cardiovascular disease. In a genotyped subset of FINRISK, systolic and diastolic blood pressure polygenic risk scores improved the predictive power of an externally validated clinical hypertension risk equation. Using publicly available genetic association data, we observed four genetic hypertension components corresponding to recognizable clinical features and demonstrated their clinical relevance in FINRISK and FinnGen.

In conclusion, data support the existence of a hyperglycemic hypertension subtype and robust genetic hypertension subtypes. Our findings demonstrate the current ability and future potential of genetics together with methodological development to improve personalized hypertension care.

KEYWORDS: hypertension, subtyping, risk factors, genetic risk scores, epidemiology

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Sisätautioppi
FELIX VAURA: Verenpainetaudin alatyypitys fenotyypin ja genotyypin avulla
Väitöskirja, 141 s.
Turun kliininen tohtoriohjelma
Syyskuu 2022

TIIVISTELMÄ

Nykyistä ymmärrystä verenpainetaudin fenotyypillisistä ja geneettisistä riskitekijöistä ei ole tehokkaasti hyödynnetty verenpainetaudin yksilöllisen hoidon mahdollistamiseksi. Tämän väitöskirjatutkimuksen tavoitteena oli tutkia verenpainetaudin alatyypitystä soveltamalla julkisesti saatavilla olevia ohjatun ja ohjaamattoman oppimisen algoritmeja suuriin feno- ja genotyypitettyihin tutkimusaineistoihin.

Tutkimuksessa hyödynnettiin osallistujia kahdesta suuresta suomalaisesta tutkimuksesta: 32,442 FINRISKIstä ja 218,792 FinnGenistä. FINRISKI on viiden vuoden välein toteutettava väestötutkimus kroonisten tarttumattomien tautien riski- ja suojatekijöistä. FinnGen on julkisen ja yksityisen sektorin yhteinen tutkimushanke joka yhdistää imputoitua geneettistä tietoa biopankeista, potilaskohorteista ja prospektiivisista epidemiologisista tutkimuksista. Koska jokainen Suomen kansalainen on yhdistetty terveydenhuollon rekistereihin henkilötunnuksella, pitkäaikaisseuranta on mahdollista kaikkien merkittävien päätepisteiden osalta verenpainetauti ja sydän- ja verisuonitaudit mukaan lukien. Lisäksi tutkimuksessa hyödynnettiin julkisesti saatavilla olevaa tietoa geneettisistä assosiaatioista muun muassa UK Biobank -tutkimuksesta.

FINRISKIssä havaittiin fenotyypityksen perusteella verenpainetaudin alatyyppi, jonka erityispiirteitä olivat korkea verensokeri ja kohonnut painoindeksi. Alatyyppi oli yhteydessä kohonneeseen sydän- ja verisuonitautiriskiin. FINRISKIn genotyypitetyssä alaryhmässä osoitettiin, että systolisen ja diastolisen verenpaineen polygeeniset riskipisteet paransivat verenpainetaudin puhkeamista ennustavan kliinisen riskilaskurin ennustevoimaa. Hyödyntämällä julkista tietoa verenpainetaudin geneettisistä assosiaatioista havaittiin neljä verenpainetaudin geneettistä osatekijää, joiden kliininen merkitys osoitettiin FINRISKIä ja FinnGeniä apuna käyttäen.

Löydökset viittaavat verenpainetaudin hyperglykeemisen fenotyypin ja usean geneettisen alatyypin olemassaoloon. Genetiikkaa ja menetelmäoppia yhdistämällä on nyt ja tulevaisuudessa mahdollista parantaa verenpainetaudin yksilöllistä hoitoa.

AVAINSANAT: verenpainetauti, alatyypitys, riskitekijät, geneettiset riskipisteet, epidemiologia

Table of Contents

Abbreviations

ACC	American College of Cardiology
AHA	American Heart Association
BMI	Body mass index
bNMF	Bayesian non-negative matrix factorization
BP	Blood pressure
CI	Confidence interval
CVD	Cardiovascular disease
DBP	Diastolic blood pressure
eGFR	Estimated glomerular filtration rate
GRS	Genetic risk score
HDL	High-density lipoprotein
HNMF	Hybrid non-negative matrix factorization
HR	Hazard ratio
HyperGEN	Hypertension Genetic Epidemiology Network
MAP	Mean arterial pressure
NMF	Non-negative matrix factorization
PP	Pulse pressure
PREVEND	Prevention of Renal and Vascular End-Stage Disease
PRS	Polygenic risk score
SBP	Systolic blood pressure
SD	Standard deviation
SPRINT	Systolic Blood Pressure Intervention Trial

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Vaura FC, Salomaa VV, Kantola IM, Kaaja R, Lahti L, Niiranen TJ. Unsupervised hierarchical clustering identifies a metabolically challenged subgroup of hypertensive individuals. J Clin Hypertens (Greenwich). 2020;22(9):1546-1553.

II Vaura F, Kauko A, Suvila K, Havulinna AS, Mars N, Salomaa V, FinnGen, Cheng S, Niiranen T. Polygenic risk scores predict hypertension onset and cardiovascular risk. Hypertension. 2021;77(4):1119-1127.

III Vaura F, Kim H, Udler M, FinnGen, Salomaa V, Lahti L, Niiranen T. Multi-trait genetic analysis reveals clinically interpretable hypertension subtypes. Circ Genom Precis Med. 2022;15(4):e003583.

1 Introduction

Known pathophysiological mechanisms of hypertension encompass several organ systems such as the kidney, the cardiovascular system, the nervous system, and even the microbiome [1,2]. However, it is debatable whether this knowledge has been effectively used to benefit individual patients. Indeed, in most cases, the clinician cannot discern why their patient has developed hypertension, and these patients are termed as having essential hypertension [3]. Because of this lack of insight, finding the optimal treatment for most hypertensive patients involves a significant degree of trial and error, which may increase healthcare costs and the frequency of adverse events [4]. Not surprisingly, established guidelines for personalized hypertension care are lacking [5].

One way to approach patient heterogeneity is by examining phenotypic patient characteristics. Based on clinical observations of patient heterogeneity, simple hypertension subtypes such as isolated systolic hypertension, white coat hypertension, and resistant hypertension have been identified [5]. However, it is unclear if these low-resolution subtypes capture the main characteristics of patient heterogeneity in hypertension. To identify patients at high risk for hypertension, clinical risk equations using readily available phenotypic risk factors such as blood pressure (BP) and body mass index (BMI) have been developed [6]. While risk equations provide estimates for individual urgency of prevention, they give limited insight into the potential modes of prevention and treatment because several different pathophysiological mechanisms may underlie high-risk forms of hypertension. Clinical hypertension subtypes based on data-driven methodology could offer better insight into patient heterogeneity than subjective, experience-based subtypes or linear, one-dimensional risk scores.

In addition to phenotypic risk factors, hypertension has many known genetic risk factors. While single mutations can cause hypertension in some rare cases, the disease is primarily polygenic, meaning that most genetic contributions come from thousands of relatively harmless mutations scattered throughout the genome [7]. Several studies have shown that polygenic contributions predict hypertension incidence [8–10]. However, no study has demonstrated that measuring polygenic contributions improves hypertension prediction beyond traditional phenotypic risk

factors. Therefore, approaching personalized healthcare through genetics has had little success. The minor role of genetics in hypertension care is perhaps best demonstrated by the most recent 103-page clinical practice guideline for hypertension by the American College of Cardiology (ACC) and American Heart Association (AHA), which dedicates one paragraph to genetics [5].

Identifying data-driven hypertension subtypes is a necessary first step in designing personalized treatment modalities. This thesis aimed to demonstrate the existence of phenotypic and genetic hypertension subtypes using unsupervised multivariable techniques. An additional aim was to establish the clinical significance of genetics in hypertension prediction. Based on the findings, we discuss future directions for implementing personalized hypertension care.

2 Review of the Literature

2.1 Mosaic theory of hypertension

2.1.1 Original theory

Research in hypertension pathophysiology dates back to observations by Tigerstedt and Volhard, later summarized by Goldblatt: an underperfused kidney releases a BP-raising substance called renin [11–14]. These observations laid the groundwork for subsequent clinical studies by Page and Helmer. They showed that combining renin with plasma containing a "renin activator" formed a BP-raising substance called angiotensin [15]. However, it was clear that vasoconstriction alone could not explain hypertension.

In his 1949 paper, Page summarised available clinical evidence and divided the pathophysiology of hypertension into four components: neural, cardiovascular, endocrine, and renal [16]. For the neural component, Koch and Mies had demonstrated that severing the carotid sinus and aortic depressor nerves led to sustained hypertension in mice, while Grulee and Panos observed that hypertension often followed bulbar poliomyelitis in children [17,18]. For the endocrine component, Selye et al. had shown that administering desoxycorticosterone (a mineralocorticoid) to unilaterally nephrectomized rats, mice, or chicks on a salty diet resulted in moderate hypertension and severe nephrosclerosis [19]. For the cardiovascular component, Steele demonstrated that coarctation of the aorta led to systemic diastolic hypertension [20]. For the renal component, Goldblatt observed that constricting renal arteries in dogs resulted in hypertension [14]. These early clinical observations formed the basis for subsequent hypertension research.

Page further developed his theory of hypertension pathophysiology into a "mosaic theory" of hypertension which he published in 1963 [21]. Along with his theory, he presented the now-famous octagon (Figure 1A), which included the following eight nodes: chemical, reactivity, volume, vascular caliber, viscosity, cardiac output, elasticity, and neural [21]. He further developed the theory to include more general concepts and presented an updated octagon in 1982 (Figure 1B) [1]. The new formulation used the following terms: genetic, environmental, anatomical,

adaptive, neural, endocrine, humoral, and hemodynamic, which continue to be relevant today.

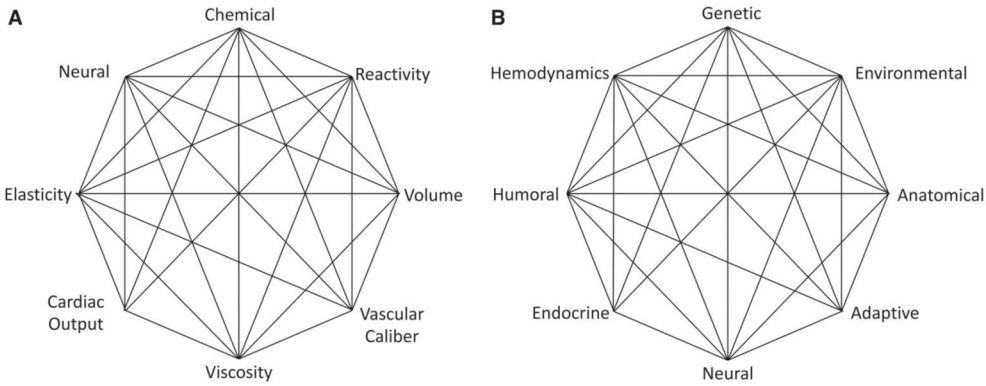

Figure 1. The original (A) and revised (B) Mosaic Theories proposed by Page. The symmetrical arrangement of the eight interconnected nodes depicts their equal importance and interdependence in blood pressure regulation. Some pairs of nodes are missing a connecting line due to a misprint. Reproduced from Circulation Research (Harrison et al., 2021) with permission of Wolters Kluwer Health, Inc.

2.1.2 Revised theory

In their 2021 review, Harrison et al. presented an updated version of Page's octagon based on the current literature on hypertension pathophysiology (Figure 2) [2]. The roles of the kidney, the vasculature, and the central nervous system in hypertension pathophysiology can be summarized as follows: the kidney increases sodium retention, renin release, afferent nerve activity, and serves as a site of immune activation; the vasculature experiences elevated vasomotor tone, smooth muscle hypertrophy, and fibrosis; and the central nervous system improves sympathetic tone, which contributes to hypertension via increased vasoconstriction, renal renin production, sodium retention, and immune activation [2]. The kidney, the vasculature, and the central nervous system continue to be central areas in hypertension research.

We will briefly describe recent insights relating to oxidative stress, the immune system, genetics, salt, and the microbiome. Oxidative stress adversely affects cellular macromolecules via reactive oxygen species and thereby causes cell injury and death [22]. Evidence from studies in animals and cultured cells suggests that reactive oxygen species could also contribute to hypertension across a wide range of regulatory systems, including the vasculature, heart, kidney, brain, and immune system [23]. However, the role of oxidative stress in human hypertension is still unclear. The immune system becomes activated in hypertension and contributes to hypertension via reactive oxygen species, hydrogen peroxide, and cytokines [24].

Nevertheless, the potential of inflammation as a therapeutic target for hypertension is unclear. For example, in a recent randomized controlled trial, a monoclonal antibody blocking interleukin-1 beta reduced inflammatory markers and cardiovascular events but not incident hypertension [25–28]. As genome-wide association studies (GWAS) with large sample sizes, e.g., >100,000, have emerged, genetics has become increasingly relevant in furthering our understanding of hypertension pathophysiology. To date, >30 single gene hypertension mutations and >1000 single nucleotide polymorphisms contributing to BP regulation have been identified [7]. Because hypertension is a dichotomous variable based on arguably arbitrary BP cutoffs, genetic variants contributing to blood pressure regulation must also contribute to hypertension. In particular, BP polygenic risk scores (PRS) have shown promise in predicting incident hypertension [8–10]. While Page had already recognized the role of salt in hypertension, new research has provided insights into how exactly our bodies process salt. In particular, it is now apparent that salt is stored in skeletal muscle and skin and can be measured using magnetic resonance imaging [29]. Finally, the microbiome is a new emerging frontier in hypertension. Changes in microbial richness and diversity have been observed in hypertensive and even in prehypertensive humans, and fecal transfer from hypertensive mice to germ-free mice appears to predispose the latter to hypertension [30,31]. While the hypertension pathways mentioned above can be abstracted as separate entities as we have done here, they are all aspects of a single network of complex interactions. In conclusion, the current literature on hypertension pathophysiology demonstrates that any serious attempt at partitioning patient heterogeneity should at the very least employ a multivariable analysis.

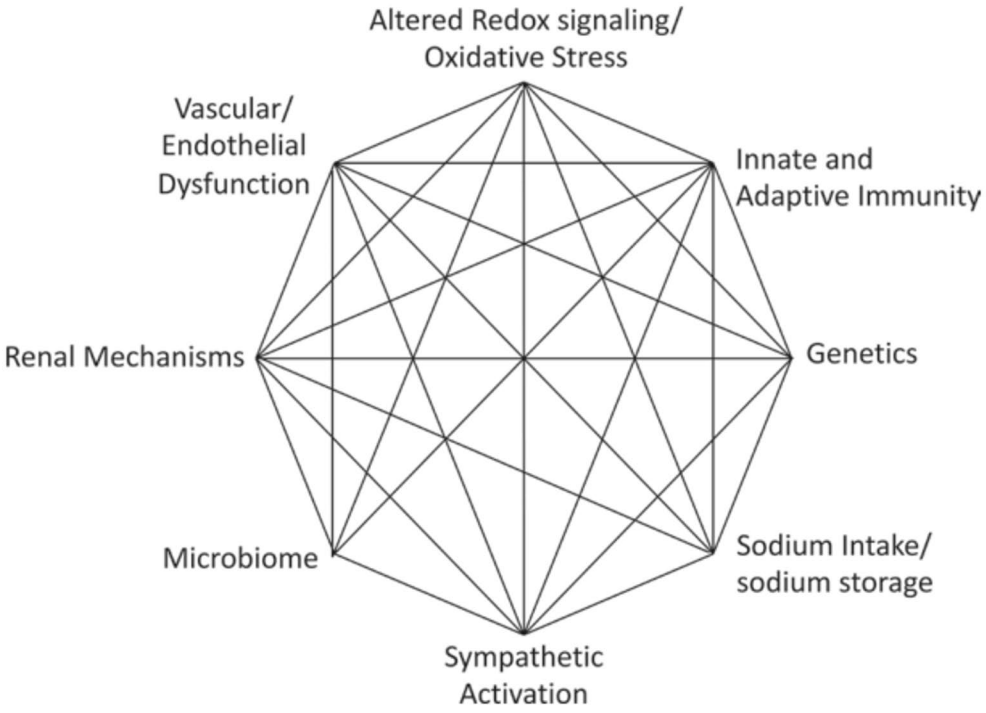

Figure 2. A revised Mosaic Theory incorporating new understanding of cellular, environmental, and genetic mechanisms. The symmetrical arrangement of the eight interconnected nodes depicts their equal importance and interdependence in blood pressure regulation. Some pairs of nodes are missing a connecting line due to a misprint. Reproduced from Circulation Research (Harrison et al., 2021) with permission of Wolters Kluwer Health, Inc.

2.2 Clinical subtypes of hypertension

2.2.1 By etiology

The most common way to clinically subtype hypertension is by etiology. When hypertension has an identifiable cause, it is called secondary hypertension. Secondary hypertension comprises less than 10% of all hypertension cases, and the most common causes of secondary hypertension are obstructive sleep apnea, renal parenchymal disease, renal artery stenosis, and primary aldosteronism [32]. Clinical signs suggesting secondary hypertension include early onset (<30 years), resistance to drug therapy, severe hypertension (>180/110 mmHg), sudden increase of BP, non-dipping or reverse dipping during 24 h ambulatory BP monitoring, and presence of target organ damage such as left ventricular hypertrophy [32]. Although secondary hypertension can often be cured by treating the underlying cause, BP rarely returns to normal, possibly because of underlying primary hypertension [32]. Nevertheless,

from the perspective of this study, secondary hypertension is of limited interest because personalized hypertension care is already possible for these patients due to a single identifiable cause.

On the other hand, in over 90% of hypertension cases, no single identifiable cause can be identified. This type of hypertension is called primary or essential hypertension. Until 2017, most guidelines defined hypertension as systolic BP (SBP) ≥140 mmHg or diastolic BP (DBP) ≥90 mmHg, but the new ACC/AHA definition in 2017 lowered the cutoffs to 130 and 80, respectively [5]. To treat primary hypertension, the ACC/AHA hypertension treatment guidelines recommend a combination of lifestyle changes, such as diet and exercise optimization, and pharmacological intervention with thiazide diuretics, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, and calcium channel blockers [5]. Antihypertensive pharmacological therapy has been shown to lower both BP and the risk of cardiovascular disease (CVD) and death [33–36]. However, little can be said about the specific combination of treatments an individual patient should get. While ACC/AHA guidelines recommend basic laboratory tests for all patients with primary hypertension, their main function is to test for secondary hypertension rather than assign patients to treatment groups [5]. Indeed, finding the right combination of treatments for essential hypertension is still fundamentally random. Identifying subtypes within essential hypertension would be a logical first step toward personalized hypertension care.

2.2.2 By blood pressure reading

Increases in SBP and DBP are associated with stiffening of arteries and increased peripheral resistance, respectively. SBP rises throughout the life course, while DBP rises to approximately 55 years of age and then declines [37]. However, for individuals, SBP and DBP may take different trajectories. Because of this and the difference in pathophysiology, it is common to define hypertension subtypes separately for SBP and DBP, namely isolated systolic hypertension (ISH; high SBP and normal DBP) and isolated diastolic hypertension (IDH; normal SBP and high DBP), and systolic-diastolic hypertension (SDH; high SBP and high DBP). Because CVD risk increases for all BP readings over 115 mmHg SBP and 75 mmHg DBP [38], we expect both ISH and IDH to be associated with increased CVD risk. However, the subtypes differ in their risk factors and clinical significance.

ISH is the most prevalent hypertension subtype in individuals >65 years of age [39]. Moreover, it is independently and robustly associated with future CVD events [40–42], and randomized controlled trials demonstrate that ISH should be treated in older individuals [34,43–45]. Risk factors for ISH include older age, female sex, and increased BMI [46]. On the other hand, IDH is the most prevalent subtype in

individuals <45 years of age [39] and is predicted by younger age, male sex, and increased BMI [46]. The CVD risk associated with IDH appears to depend on the chosen cutoff value: a robust association has been observed using 90 mmHg [47], while 80 mmHg has given mixed results [48–51]. SDH is the most prevalent subtype in individuals aged 45 to 65 years, and individuals with SDH are more likely to be diagnosed with hypertension than those with ISH or IDH [52]. Individuals with ISH and IDH are diagnosed with hypertension less frequently than individuals with SDH. Interestingly, IDH converts to SDH with a hazard ratio (HR) of 23 compared to normotensive individuals, supporting the hypothesis that IDH and SDH share similar pathophysiology [46]. Inspired by this observation, Orias et al. proposed a new subtype called predominantly diastolic hypertension, which includes IDH and a subset of SDH with narrow pulse pressure (PP) [53]. However, this subtype has received little attention. The fact that ISH and IDH appear to differ in pathophysiology and patient characteristics is unsurprising given their strictly non-overlapping definitions. Moreover, because their definitions are based on intuition rather than an objective multivariable analysis, it is unclear how well they capture patient heterogeneity.

2.2.3 Dependence on measurement location

Out-of-office BP monitoring has led to clinical observations that the BP of some individuals is high only in the clinic or only at home. The former is called white-coat hypertension (WCH), and the latter masked hypertension (MH). When BP is high at home and in the clinic, the patient is said to have sustained hypertension (SH). Prevalence estimates of WCH and MH are highly dependent on the exact definitions used. In one multiethnic cohort, the prevalences of WCH and MH were 3.3% and 17.8%, respectively [54].

The metabolic changes and subclinical organ damage associated with WCH are between normotensive individuals and those with sustained hypertension in magnitude. [55,56]. Several meta-analyses have concluded that WCH in individuals not undergoing antihypertensive medication is associated with increased CVD risk [57–59]. The largest meta-analysis to date by Cohen et al. also demonstrated an increased risk for CVD mortality and overall mortality [57]. However, WCH in individuals undergoing antihypertensive treatment does not appear to increase CVD risk or mortality [57]. While out-of-office BP is also elevated in WCH compared to normotensive individuals, likely explaining part of the increased CVD risk, office BP is nevertheless an independent risk factor for the development of SH [60]. Hyperreactivity to stress stimuli has been proposed as the underlying pathophysiological mechanism for WCH. Still, as Mancia et al. point out, the

explanation is not entirely satisfactory and has likely discouraged further investigations into WCH pathophysiology [61].

MH is more elusive to clinicians than WCH because home BP measurement is often not initiated on patients with normal office BP. This delay in diagnosis may explain the increased prevalence of target organ damage in MH [54,62]. Known risk factors for MH include stress, smoking, older age, sedentary lifestyle, and sleep disturbances [63–66]. Moreover, MH has increased prevalence among persons with African heritage, diabetes, and chronic kidney disease [67–69]. As suggested by Yano and Bakris, masked hypertension may further be classified as masked morning, masked daytime, and masked nocturnal hypertension, with heterogeneous risk factors [70]. A systematic review and meta-analysis by Palla et al. (14,739 individuals, mean age 58, follow-up 9.5) demonstrated that compared to WCH and normotension, CVD events and all-cause mortality are higher in patients with MH, regardless of antihypertensive treatment status. MH had fewer CVD events than SH, but the difference was only present in untreated individuals and disappeared in individuals undergoing antihypertensive treatment [71]. Paradoxically, antihypertensive treatment appears to increase the prevalence of both uncontrolled MH and the CVD risk in these patients, as well as the CVD risk in treated normotensive patients [68]. Franklin et al. give a logical explanation for this, proposing that SH converts to MH while MH converts to normotension. If true, this illustrates that lowering BP alone may not eliminate the CVD risk associated with previous MH [72].

WCH and MH undoubtedly represent clinically significant hypertension subtypes that fast-track diagnosis and treatment initiation. However, their value lies in helping clinicians identify individuals with hypertension rather than in comprehensive subtyping. Consequently, WCH and MH offer little insight into overall patient heterogeneity or how the treatment should be initiated.

2.2.4 Pharmacological response

Resistant hypertension (RH) refers to hypertension that persists despite three different antihypertensive medications (including a diuretic) or hypertension that requires at least four medications to be controlled [73]. RH prevalence has been estimated at around 15% in the treated adult population [74]. Risk factors for RH include ethnicity, body mass index, diabetes, systolic BP, estimated glomerular filtration rate (eGFR), and the number of antihypertensive medications [75]. The subtype is associated with increased CVD risk and all-cause mortality [76,77]. However, from a clinician's point of view, this subtype is challenging to assess objectively because of potential adherence issues, and partial or total non-adherence to pharmacological treatment among RH patients has been estimated at 47% [78].

Because the prevalence of obstructive sleep apnea alone is 70 to 90% in individuals with resistant hypertension, secondary hypertension should always be suspected as the underlying reason behind RH [79,80]. For true RH, fluid and sodium retention have been identified as the main underlying mechanisms [81], and spironolactone is recommended as a fourth line of treatment [81–84].

Refractory hypertension refers to treatment-resistant RH, defined as at least five antihypertensive drugs, including a long-acting thiazide diuretic and a mineralocorticoid receptor agonist [85]. While the phenotype appears to be a strict subset of RH, evidence points to differences in pathophysiological mechanisms [86,87], with refractory hypertension characterized by sympathetic hyperactivity [88]. Compared to RH, patients with refractory hypertension are younger, more often female, and have more CVD risk factors [89]. Longitudinal studies examining the CVD risk of patients with refractory hypertension are still lacking [90]. By definition, no effective pharmacological treatment exists for refractory hypertension. New treatment modalities such as renal denervation, carotid sinus activation, and central iliac arteriovenous anastomosis are promising but still lack evidence [87,91,92].

2.3 Clinical subtyping of individuals with hypertension

2.3.1 Clinical risk factors for hypertension

Anthropometric traits can be visually assessed and easily measured. Starting from early insurance industry reports identifying a relationship between body weight and BP [93], several subsequent epidemiological studies have demonstrated a continuous, near-linear relationship between BMI and BP [94,95]. An even stronger relationship has been observed between the waist-to-hip ratio and BP [96]. The attributable risk conveyed by obesity to hypertension has been estimated at 40 to 78% [97,98].

While harder to measure objectively, lifestyle choices such as sodium and potassium intake, alcohol use, and physical activity predict hypertension onset. Using 24-h sodium, the INTERSALT study (n=10,079) demonstrated that every 5.8 grams of salt increased BP by 3-6/0-3 mmHg SBP/DBP, respectively. Longitudinal studies support this finding [99,100]. Conversely, in the Prevention of Renal and Vascular End-Stage Disease (PREVEND) study (n=5511), individuals in the lower tertile of potassium intake had a 20% higher risk of hypertension [101]. While alcohol appears to decrease BP in the short term [102], its long-term effects predispose to hypertension in a dose-dependent manner [103]. Finally, various forms

of physical activity, including endurance, dynamic resistance, and isometric resistance training, lower SBP and DBP [104].

Biomarkers yield direct information about the state of a biological system, and several biomarkers are associated with hypertension. In their comprehensive review, Shere et al. identify 30 such biomarkers and discuss their roles in essential hypertension. The authors note that most of these biomarkers do not have feasible, practical applications due to high costs, lack of technology, and lack of expertise [105]. However, some of these biomarkers are already routinely used in clinical practice. Three meta-analyses for C-reactive protein (CRP) (n=142,640), urate (n=55,607), and urinary albumin-to-creatinine ratio (UACR) (n=27,771) found that all three biomarkers are robustly associated with increased hypertension risk [106–108]. While several studies show a relationship between eGFR and BP [109–114], the evidence is conflicting, possibly due to confounding underlying conditions such as diabetes or chronic kidney disease. Notably, Eriksen et al. investigated the association between baseline BP and eGFR in 1594 individuals without baseline diabetes or kidney or cardiovascular disease [115]. They observed that BP was associated with eGFR decline when calculated with creatinine but not when calculated with iohexol clearance, leaving the authors to conclude a lack of association [115]. Impaired glucose metabolism is closely related to hypertension [116]. In a longitudinal study of 2210 individuals followed over five years, Tatsumi et al. demonstrated that a 1.0 mmol/L increase in fasting blood glucose, a 1.0% increase in glycated hemoglobin, and a 1.0 increment in insulin resistance (defined by homeostasis model assessment of insulin resistance) were associated with 14%, 25%, and 16% increased risk for incident hypertension, respectively [117]. Serum lipids, namely high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, and triglycerides, have been associated with BP and hypertension in certain population studies [118–120]. Recently, a comprehensive biomarker profiling study in 36,985 individuals by Palmu et al. found that these lipids, among other common biomarkers such as glucose, were associated with current or future BP [121].

2.3.2 Supervised clinical subtyping

The main rationale behind treating hypertension is its role as the leading risk factor for CVD, the biggest killer globally [122]. A consequence of this hierarchy is that patient risk profiling is primarily done through CVD risk equations instead of hypertension risk equations. Most notably, the ACC/AHA risk equation is commonly used to estimate atherosclerotic CVD risk based on age, sex, race, SBP, total cholesterol, HDL cholesterol, history of diabetes, antihypertensive medication status, and smoking status [123]. Risk equations are examples of supervised learning,

where the end goal (in this case, predicting atherosclerotic CVD) is known and included as part of the training data for the equation.

Several studies have developed hypertension risk models based on clinical variables [6]. However, most have not been externally validated. Externally validated hypertension risk models include the John Hopkins model, the Framingham score model, and the Jichi Genki hypertension prediction model [124–126]. The John Hopkins model is based on 1130 first-year male medical students, and it models incident hypertension risk in middle age [124]. The model uses age, SBP, BMI, and parental history of hypertension and was intended as an educational tool to demonstrate hypertension risk to students [124]. The model has limitations for wider use, such as its generalizability to only young males. The John Hopkins model has been externally validated in a middle-aged Taiwanese cohort, displaying poor calibration [127]. In contrast, the Framingham score model included 1717 nonhypertensive white individuals aged 20 to 69 (54% women) without diabetes to construct 1-, 2-, and 4-year risk scores for new-onset hypertension [125]. The equation uses variables available in the clinic: SBP, DBP, sex, BMI, age, smoking status, and parental hypertension status. The Framingham score model has been externally validated in white Europeans and Koreans, displaying good calibration [127,128]. More recently, a Japanese biobank study (n=63,495, aged 18 to 83; validation n=14,168, aged 18 to 89 years) incorporated age, sex, BMI, SBP, DBP, low-density lipoprotein cholesterol, urate, proteinuria, current smoking, alcohol intake, and eating rate to construct and validate the Jichi Genki hypertension risk model in Japanese individuals [126]. While linear risk stratification allows one to identify patients at high risk for incident hypertension, it cannot differentiate between qualitatively distinct risk profiles with similar overall risk. Dissecting patient heterogeneity is a multidimensional problem that requires a multidimensional solution, possibly in the form of multiple risk scores, each capturing a different aspect of hypertension risk.

2.3.3 Unsupervised clinical subtyping

In contrast to supervised learning, unsupervised learning does not include the end points as part of the training data. In other words, unsupervised learning tries to identify objective patterns in the data without knowing what the end goal is. Two studies by Guo et al. and Yang et al. have used unsupervised algorithms to dissect phenotypic patient heterogeneity in hypertension [129,130].

Guo et al. used a cohort of 513 individuals with the following baseline variables: age, sex, the prevalence of coronary artery disease (CAD) and cerebral infarction, smoking, diabetes, fasting glucose, carotid plaque thickness, several variables derived from 24-hour ambulatory BP monitoring, and lipids [129]. They first used

principal component analysis (PCA) to reduce the multivariable data into two dimensions and then applied k-means clustering to divide individuals into hypertension clusters. Guo et al. observed four clusters: cluster 1 consisted mainly of young smoking men (n=172), cluster 2 of older diabetic women (n=70), cluster 3 of relatively healthy individuals (n=144), and in cluster 4, everyone had prevalent CAD (n=127). Yang et al. used a larger cohort of 9361 individuals from the randomized, controlled, open-label Systolic Blood Pressure Intervention Trial (SPRINT) to cluster hypertensive individuals based on sex, BMI, SBP, DBP, lipids, kidney markers, history of CVD, various medication use, and the 10-year Framingham CVD risk score [130]. They used a two-step clustering process. The first step estimates the optimal number of clusters, and the second step uses k-means clustering to divide individuals into hypertension clusters. Yang et al. also observed four clusters: cluster 1 consisted of relatively healthy individuals, cluster 2 of individuals with slightly decreased eGFR, cluster 3 of individuals with high BMI, and cluster 4 of individuals with high 10-year Framingham CVD risk score. Unsurprisingly, cluster 4 had the highest CVD risk, with the other clusters displaying similar CVD risk. While these two studies have laid the groundwork for improved classification of hypertension, they are limited in either sample size, lack of longitudinal follow-up, or exclusion of diabetic individuals.

2.4 Genetic subtyping of individuals with hypertension

2.4.1 Genetic risk factors for hypertension

Twin and family studies demonstrate the strong genetic component of hypertension. Monozygotic twins display a higher correlation of BP than dizygotic twins [131], and family studies suggest that parental and grandparental hypertension history increases hypertension risk even after adjusting for environmental effects [132]. However, clinical applications of hypertension genetics have been scarce despite its recognized importance.

Genetic hypertension risk can be divided into monogenic BP disorders and polygenic risk. Monogenic BP disorders refer to those clinical presentations of high or low BP that can be explained by single genetic mutations [7]. We will briefly summarize their main clinical characteristics here. Liddle's syndrome and Gordon's syndrome are BP increasing syndromes resulting from hyperactivity of sodium and chloride transporters. Liddle's syndrome results from elevated sodium absorption and is clinically characterized by hypokalemia and metabolic alkalosis, while Gordon's syndrome results from retention of sodium and potassium and is clinically characterized by hyperkalemia, metabolic acidosis, and normal eGFR [133,134].

Congenital adrenal hyperplasia and apparent mineralocorticoid excess result from disturbances in adrenal steroid metabolism and activity due to enzyme deficiencies. In apparent mineralocorticoid excess, cortisol is not metabolized due to an enzyme deficiency, which through mineralocorticoid receptor agonism leads to hypokalemia and metabolic alkalosis [135]. In congenital adrenal hyperplasia, an enzyme deficiency leads to cortisol deficiency and an androgen excess which typically presents with ambiguous genitalia in female infants [136]. Glucocorticoid remediable aldosteronism, also known as familial hyperaldosteronism type 1, is caused by adrenocorticotropic hormone -mediated hyperaldosteronism, and presents as hyperaldosteronism. Simonetti et al. provide a comprehensive flow chart for clinically discriminating between these monogenic forms of hypertension [137]. BP-lowering monogenic disorders include Bartter's and Gitelman's syndromes, which are both issues in tubular transport in the nephron, both presenting as hypotension and hypokalemic metabolic alkalosis, with additional hypocalciuria and hypomagnesemia characterizing Gitelman's syndrome [138]. Finally, sporadic aldosterone-producing adenomas, multiple endocrine neoplasias, paragangliomas, and the von Hippel-Lindau syndrome can all affect BP regulation and lead to hypertension. While monogenic BP disorders are now well understood, they do not come close to explaining the heritability of hypertension which is estimated at 50 to 60% [139].

More than 1400 genetic associations between single nucleotide polymorphisms (SNP) and BP traits (SBP, DBP, and PP) have been identified in GWASs [140,141]. While high in number and usually common in the population, these SNPs have relatively small effect sizes, with the median around 0.2 mmHg [140]. Together they explain approximately 27% of BP heritability [140]. Curiously, most BP SNPs are also pleiotropic, meaning they also have associations with other traits than BP. Padmanabhan and Dominiczak demonstrated that out of the 1477 known loci, 1302 were pleiotropic with one or more traits: anthropometric, hematological, biochemical, lifestyle, lung-related, heart rate, autoimmune, brain-related, glycemia, and neoplasms [7]. A significant limitation of most BP GWASs is their focus on European individuals, which limits their generalizability to other ancestries. Moreover, despite the large number of known GWAS loci, no study has demonstrated the clinical usefulness of genotyping individuals.

2.4.2 Supervised genetic subtyping

Much like epidemiological risk factors for hypertension can be combined into a single risk equation to stratify individuals by clinical hypertension risk, the effect sizes of several SNPs can be combined into a single, stronger predictor called a genetic risk score (GRS). However, while GRSs successfully predict incident

hypertension, no study has demonstrated that they improve existing clinical hypertension risk equations [8–10]. Lim et al. examined 8556 and 5632 individuals in cross-sectional and longitudinal analyses, respectively [10]. They found that a GRS based on four SNPs was associated with both prevalent and incident hypertension. However, a hypertension risk model containing age, sex, smoking status, SBP, parental history of hypertension, and BMI as predictors did not improve with the addition of the GRS [10]. Fava et al. followed >17,000 middle-aged (mean age 45) Swedes for an average of 23 years [9]. Their GRS based on 29 SNPs was associated with SBP and DBP at baseline and follow-up, as well as prevalent and incident hypertension. However, adding the GRS into a hypertension prediction model using a wide range of traditional risk factors such as anthropometric traits, laboratory markers, lifestyle factors, and socio-economic status did not improve predictive power [9]. Niiranen et al. had a middle-aged (mean age 52) cohort of 5402 and 3266 individuals at baseline and 11-year follow-up, respectively [8]. They constructed a GRS of 32 SNPs and showed that it was associated with BP at baseline and follow-up. However, it did not improve a model containing age, sex, BMI, smoking status, diabetes and hypercholesterolemia status, education level, and the amount of leisure-time exercise [8]. Therefore, because of the lack of additional value compared to clinical variables, the additional cost and trouble associated with genotyping does not currently warrant the inclusion of SNPs into individual risk prediction of hypertension.

2.4.3 Unsupervised genetic subtyping

Two studies by Ma et al. and Luo et al. have assessed unsupervised genetic subtyping of hypertension [142,143]. Ma et al. used a traditional subtyping algorithm, non-negative matrix factorization (NMF), on raw allele counts to compute hypertension components in 1187 hypertensive participants of the Hypertension Genetic Epidemiology Network (HyperGEN) [142,144]. They observed two genetic components characterized by echocardiographic measurements: component 1 had lower septal and lateral left ventricular early diastolic relaxation velocity, lower global longitudinal strain, and lower atrial strain rate than component 2. This approach to genetic subtyping is straightforward but does not consider phenotypic data. Luo et al. proposed a novel hybrid non-negative matrix factorization (HNMF) method for combining phenotypic and genetic information to find hypertension components [143]. They applied HNMF to 660 hypertensive participants of the HyperGEN study and demonstrated superior performance compared to other similar methods (not including bNMF). Notably, HNMF requires the dataset to include both extensive phenotyping and genotyping, which poses challenges in resource allocation. In conclusion, while these studies have pioneered genetic subtyping of

hypertension, neither method uses the publicly available GWAS data linking SNPs and clinical variables.

3 Aims

This thesis was designed to examine the phenotypic and genetic heterogeneity of hypertension using both supervised and unsupervised risk stratification.

The specific aims of this thesis were:

1. To identify phenotypically distinct groups of hypertensive individuals using unsupervised clustering and to assess whether these groups differ in their cardiovascular risk profiles. (Study I)
2. To examine whether BP PRSs have clinical utility in predicting incident hypertension. (Study II)
3. To characterize the genetic basis of hypertension as a mixture of latent components and to assess whether these components differ in their associations to disease end points. (Study III)

4 Materials and Methods

4.1 Study populations

4.1.1 The National FINRISK Study

The National FINRISK Study is a large cross-sectional population survey on risk factors for chronic, non-communicable diseases. Individuals are randomly sampled from five geographical areas of Finland every five years. We used four cohorts from 1997 (n=8446), 2002 (n=9580), 2007 (n=7993), and 2012 (n=6424) with a total of 32,443 individuals aged 24 to 74 years. Because every Finnish citizen is linked to health registers via a personal identity code, accurate follow-up is possible for all major end points, including hypertension and cardiovascular disease. The Coordinating Ethical Committee of the Hospital District of Helsinki and Uusimaa approved FINRISK study protocols. All participants gave informed written consent. Requests to access the data set from qualified researchers trained in human subject confidentiality protocols may be submitted through https://www.thl.fi/biobank/researchers.

4.1.2 FinnGen Study

The FinnGen study is a public-private partnership research project combining imputed genotype data with nationwide longitudinal health registry data to discover genetic associations between SNPs and disease end points. Samples are collected from biobanks, patient cohorts, and prospective epidemiological surveys. By August 2020, genetic samples from 412,000 individuals had been collected and 218,792 analyzed, with >2000 end points available. The Coordinating Ethical Committee of the Hospital District of Helsinki and Uusimaa approved FinnGen study protocols. All participants gave informed written consent. Requests to access the data set from qualified researchers trained in human subject confidentiality protocols may be submitted through the Finnish Biobanks' FinnBB portal (https://finbb.fi/).

4.2 Blood pressure measurement in FINRISK

Trained nurses measured BP three times with a mercury sphygmomanometer using a cuff bladder 14 cm wide and 36 cm long. Measurements were taken from the right arm in a sitting position with at least 5 minutes of rest before the first measurement. The first phase of Korotkoff sounds was recorded as SBP and the fifth phase as DBP. Averages of the three measurements were used in subsequent analyses.

4.3 Biochemical analyses in FINRISK

After obtaining venous blood samples between 11 am and 7 pm after a minimum of four-hour fast, samples were sent to the Finnish Institute for Health and Welfare (Helsinki, Finland) laboratory to be analyzed. In 1997 and 2002, fresh samples were sent daily, while in 2007 and 2012, samples were frozen either at -70 °C or -120 °C and sent in bulk. The biochemical laboratory team at the Finnish Institute for Health and Welfare performed all biochemical analyses relevant to this manuscript. The author has had no part in sample collection, sample handling, or biochemical analyses in FINRISK.

4.4 Genotyping and imputation

Samples were genotyped using Illumina and Affymetrix arrays and called with zCall or GenCall (for Illumina) and AxiomGT1 (for Affymetrix) at the Institute for Molecular Medicine Finland. Quality control exclusions were performed both sample-wise: ambiguous gender, missingness >5%, heterozygosity >4 standard deviations (SD), or non-European ancestry; and variant-wise: missingness >2%, Hardy-Weinberg equilibrium $P<1\times10^{-6}$, minor allele count <3 (for zCall) or <10 (for GenCall). After quality control, samples were first prephased with Eagle 2.3.5 and then imputed with Beagle 4.1 using a Finnish population-specific SISu v3 reference panel. Population structure was accounted for with genetic PCA using a pruned set of SNPs of unrelated individuals. The FinnGen sequencing team and FinnGen analysis team performed imputation and PCA computations, respectively, at the Institute for Molecular Medicine Finland. The author has had no part in genotyping, imputation, or genetic PCA in FinnGen.

4.5 Polygenic risk score

A PRS combines the effect sizes from multiple SNPs into a single predictor. We calculated PRSs for SBP and DBP based on 1,098,015 variants using the publicly available PRS-CS pipeline, which computes SNP effect sizes using Bayesian regression and P-values [145]. As PRS-CS parameters, we used default parameters

and a European linkage disequilibrium reference panel derived from the 1000 Genomes Project [146]. For input data to PRS-CS, we used publicly available UK biobank GWAS summary statistics based on 340,000 individuals [147].

4.6 Subtyping methodology

4.6.1 Hierarchical clustering

Hierarchical clustering is an unsupervised algorithm that constructs a hierarchy of nested clusters for a given dataset and can therefore be used to identify subtypes [148]. We performed hierarchical clustering (R package *stats*, function "hclust") with Ward's method (Ward2) and Euclidean distance [149,150]. As variables, we used: mean arterial pressure (MAP), PP, non-HDL cholesterol, glucose, BMI, CRP, eGFR, and alcohol intake. Hierarchical clustering requires the user to choose the optimal number of clusters. To do this, we used the minimum average silhouette width (R function silhouette, package *cluster*) [151,152].

4.6.2 Bayesian non-negative matrix factorization

Non-negative matrix factorization (NMF) is a popular algorithm for finding latent parts-based representations in non-negative data [153]. Unlike hierarchical clustering, which finds groups of similar data points, NMF finds a small set of meaningful properties (also called features or components) that best model the data. Bayesian NMF (bNMF) is a modification of the traditional NMF in which Bayesian statistics is incorporated in the form of a prior distribution imposed on the update step [154]. bNMF has been previously used to identify genetic components of type 2 diabetes [155]. The code for a full bNMF pipeline is available on GitHub [156]. Because bNMF is probabilistic, the number of components in its solution may vary even if input data is kept constant. Hence, to arrive at a single solution, the bNMF pipeline performs the same calculation for a large number (e.g., 1000) of iterations and chooses the solution that appears the most times. In our application, the algorithm takes as input GWAS summary statistics for (1) an end point of interest and (2) related traits and outputs weights that link the original traits to the hypertension components, allowing us to interpret their clinical meaning. We used summary statistics for self-reported hypertension from UKBB and 16 clinical traits: BMI, CRP, eGFR, glucose, glycated hemoglobin, HDL and LDL cholesterol, height, neuroticism, smoking, total cholesterol, triglycerides, UACR, urate, and waist circumference.

4.7 Statistical analyses

We performed all computations using R software [149]. In every study, we estimated longitudinal associations between variables and end points using Cox proportional hazards regression (packages *rms*, *survival*, and *survminer*) [157–159].

Study I

We adjusted Cox models by age and sex and used time-on-study restricted to 10 years as the time scale. We evaluated proportional hazards assumptions using a correlation test based on scaled Schoenfeld residuals (R package *survival*, function "cox.zph") [158,160]. We used R version 3.6.1 for all computations. We considered a two-sided $P<0.05$ statistically significant.

Study II

In FinnGen, we adjusted Cox models by sex, DNA sample collection year, genotyping batch, and the first ten genetic principal components, using age as the time scale. We validated proportional hazards assumptions by inspecting log-minus-log plots because of large sample sizes [161]. In FINRISK, we adjusted Cox models by cohort year, genotyping batch, and the first four genetic principal components, using age as the time scale. We validated proportional hazards assumptions by visually inspecting scaled Schoenfeld residuals (R package *survminer*, function "ggcoxzph") [159,160]. We evaluated improvement in Cox models using Harrell C statistics and the two-category net reclassification improvement with risk categories <7.5% and ≥7.5% (four-year risk for hypertension and ten-year risk for CVD) [162,163]. We defined early-onset hypertension as hypertension onset at <55 years and late-onset hypertension as hypertension onset at ≥55 years. We estimated correlations between SBP and DBP PRSs with Pearson's correlation coefficient (R package *stats*, function "cor.test") [149]. We used R version 3.6.3 for all computations and considered a two-sided $P<0.05$ statistically significant.

Study III

We used linear regression (R package *stats*, function "lm") to estimate cross-sectional associations between cross-sectional variables and hypertension components. We adjusted linear regression and Cox models for sex, DNA sample collection year, genotyping batch, and the first ten (FinnGen) or four (FINRISK) genetic principal components. We used age as the time scale for Cox models. We visually inspected scaled Schoenfeld residuals to validate the proportional hazards assumptions and visually inspected residual plots (R package *stats*, function

"residuals") to validate linearity and homoscedasticity. We used R version 4.0.3 for all computations and considered a two-sided Bonferroni corrected $P<7.8\times10^{-4}$ (0.05/64) statistically significant [164].

5 Results

5.1 Phenotypic hypertension clusters (I)

5.1.1 Cluster characteristics

Out of 14,746 participants, 3726 had grade 2 hypertension and were considered for clustering, while the remaining 11,020 individuals formed a reference group. The optimal number of clusters was two for both sexes, and cluster characteristics were comparable between sexes. The smaller cluster, with 113 individuals, was characterized by a high blood glucose Z-score and elevated BMI. Blood glucose Z-scores were (± one standard deviation) 4.4±1.1 vs. 0.2±0.8 in men and 3.5±1.1 vs. 0.0±0.6 in women. BMI (kg/m^2) was 30.4±4.1 vs. 28.9±4.3 in men and 32.7±4.9 vs. 29.3±5.5 in women. Full cluster characteristics are presented in Article I, Figure 1. We termed the smaller cluster metabolically challenged (MC) and the larger cluster non-MC. The average silhouette widths for MC and non-MC were 0.38 and 0.31 (men) and 0.38 and 0.32 (women). Cluster separation was visible in PCA using components PC4 and PC5 (Article I, Figure 2). In the MC cluster, 110 out of 113 individuals (97%) had metabolic syndrome as defined by the International Diabetes Federation [165].

5.1.2 Cluster associations to CVD

Because cluster characteristics were comparable between men and women, we merged the clusters across sexes. During a median follow-up time of 10 years, 1034 CVD events occurred: 30 in the MC cluster, 504 in the non-MC cluster, and 500 in the reference group. In the unadjusted model, the risk for incident CVD was significantly higher for the MC cluster compared to the non-MC cluster (HR 2.0, 95% confidence interval [CI] 1.4–2.9, P<0.001) and to the reference group (HR 6.7, 95% CI 4.7–9.7, P<0.001). In the age- and sex-adjusted Cox model, the risk for incident CVD was significantly higher for the MC cluster compared to the non-MC cluster (HR 1.6, 95% CI 1.1–2.4, P=0.009) and the reference group (HR 2.5, 95% CI 1.7–3.7, P<0.001).

5.2 Blood pressure polygenic risk scores (II)

5.2.1 Associations to hypertension and cardiovascular disease

Our FinnGen sample comprised 218,754 individuals with 55,917 cases of incident hypertension. Higher BP PRSs were associated with higher hypertension incidence. For one standard deviation (1-SD) increases in PRSs, the HRs were 1.42 (1.41–1.43) for SBP and 1.41 (1.40–1.42) for DBP. The top 2.5% SBP and DBP PRS categories had HRs of 2.19 and 2.26 compared to the 20% to 80% category. Moreover, hypertension was diagnosed 10.6 and 10.5 years earlier in the top categories than in the 20% to 80% category. Early-onset hypertension showed stronger associations to SBP and DBP PRSs than late-onset hypertension (Table 1). Higher SBP PRS was also associated with higher CVD, CHD, and stroke incidences. The HRs per 1-SD increases in SBP PRS were 1.13 (1.12–1.15) for CVD, 1.15 (1.13–1.17) for CHD, and 1.11 (1.09–1.13) for stroke. HRs (for CVD, CHD, and stroke) for the top 2.5% category compared to the 20% to 80% SBP PRS category were 1.30, 1.33, and 1.29, with 2.3, 2.0, and 1.4 years earlier disease onset.

5.2.2 Clinical utility

Our FINRISK sample comprised 9906 individuals with 725 cases of incident hypertension. BP PRSs improved the Framingham clinical risk prediction model of hypertension: the C statistics (%) increased from 79.7 by 0.5 (0.01–0.9; P=0.016) with SBP PRS, by 0.6 (0.3–1.0; $P=8\times10^{-4}$) with DBP PRS, and by 0.7 (0.3-1.1; P=0.0017) with both (Article II, Table 2). The SBP PRS did not change the C statistics in any CVD end points, and the two-category NRI was not statistically significant, either.

Table 1. HR of hypertension onset between BP PRS categories in adjusted Cox models (FinnGen).

	Hypertension			Early-onset hypertension			Late-onset hypertension		
PRS	**HR (95% CI)**	**P value**	**Cases/controls**	**HR (95% CI)**	**P value**	**Cases/controls**	**HR (95% CI)**	**P value**	**Cases/controls**
SBP			55 917/162 837			27 361/191 393			28 556/190 198
<2.5%	0.44 (0.41—0.47)	9×10^{-107}	725/4744	0.39 (0.35—0.44)	2×10^{-52}	269/5200	0.47 (0.43—0.52)	8×10^{-57}	456/5013
2.5-20%	0.64 (0.62—0.65)	5×10^{-255}	6 911/31 367	0.55 (0.52—0.57)	3×10^{-177}	2 603/35 675	0.70 (0.68—0.73)	3×10^{-94}	4308/33 970
20-80%	1 (reference)	...	33 176/98 078	1 (reference)	...	15 658/115 596	1 (reference)	...	17 518/113 736
80-97.5%	1.54 (1.51—1.58)	$<1\times10^{-300}$	12 846/25 437	1.70 (1.66—1.75)	$<1\times10^{-300}$	7348/30 935	1.37 (1.33—1.41)	2×10^{-91}	5498/32 785
>97.5%	2.19 (2.10—2.29)	1×10^{-281}	2259/3211	2.62 (2.48—2.77)	3×10^{-271}	1483/3987	1.68 (1.57—1.81)	3×10^{-45}	776/4694
DBP			55 917/162 837			27 361/191 393			28 556/190 198
<2.5%	0.49 (0.46—0.53)	4×10^{-88}	822/4645	0.37 (0.33—0.42)	3×10^{-55}	256/5211	0.57 (0.53—0.63)	7×10^{-38}	566/4901
2.5-20%	0.67 (0.65—0.69)	2×10^{-203}	7315/30 968	0.56 (0.54—0.59)	7×10^{-162}	2675/35 608	0.75 (0.73—0.78)	2×10^{-64}	4640/33 643
20-80%	1 (reference)	...	32 873/98 375	1 (reference)	...	15 527/115 721	1 (reference)	...	17 346/113 902
80-97.5%	1.55 (1.52—1.58)	$<1\times10^{-300}$	12 662/25 624	1.73 (1.68—1.78)	$<1\times10^{-300}$	7351/30 935	1.35 (1.31—1.39)	5×10^{-80}	5311/32 975
>97.5%	2.26 (2.17—2.36)	$<1\times10^{-300}$	2245/3225	2.78 (2.64—2.93)	$<1\times10^{-300}$	1552/3918	1.60 (1.48—1.73)	3×10^{-33}	693/4777

We defined early-onset and late-onset hypertension as age of onset <55 years and ≥55 years, respectively. We adjusted the Cox proportional hazards models for sex, collection year, genotyping batch, and the first ten genetic principal components. DBP indicates diastolic blood pressure; HR, hazard ratio; PRS, polygenic risk score; and SBP, systolic blood pressure.

5.3 Genetic hypertension components (III)

5.3.1 Component characteristics

The optimal number of latent components given by bNMF was four, present in 510 out of 1000 algorithm iterations. The other two solutions were a three-component solution (462 out of 1000 iterations) and a two-component solution (28 out of 1000 iterations). The highest weighted clinical traits in each component were height (component 1), BMI and waist circumference (component 2), LDL and total cholesterol (component 3), and HDL cholesterol and triglycerides (component 4) (Article III, Table 2). The highest weighted genetic loci in each component were *LCORL* (component 1), *FTO* (component 2), *SH2B3* (component 3), and *KLF14* (component 4). GWAS-based associations between the hypertension components, averaged over SNPs in each component, and the 16 original clinical traits suggested the following clinical interpretations for the hypertension components: Short Stature (component 1), Obesity (component 2), Hypolipidemia (component 3), and Dyslipidemia (component 4) (Article III, Figure 3).

5.3.2 Cross-sectional associations between component GRSs and clinical traits

Our FINRISK sample comprised 21,168 individuals, with 53% women and a mean age of 50 years. All hypertension component GRSs were associated with their respective clinical traits (**Table 2**). A 1-SD increase in the Obesity component GRS was related to 0.29 kg/m^2 (Bonferroni corrected confidence interval [CI], 0.19–0.40; $P=5.0\times10^{-21}$) higher BMI. A 1-SD increase in the Dyslipidemia component GRS was related to 0.013 mmol/L (CI, 0.005–0.022; $P=1.2\times10^{-7}$) lower HDL cholesterol and 1.7% (CI, 0.7–2.8; $P=6.2\times10^{-8}$) higher triglycerides. A 1-SD increase in the Hypolipidemia component GRS was related to 0.026 mmol/L (CI, 0.006–0.047; $P=1.8\times10^{-5}$) lower LDL cholesterol and 0.028 mmol/L (CI, 0.005–0.051; $P=4.5\times10^{-5}$) lower total cholesterol. A 1-SD increase in the Short Stature component GRS was related to a 0.28 cm (CI, 0.14–0.42; $P=1.4\times10^{-9}$) lower height. Associations between the top decile categories of hypertension components and clinical traits agreed with the continuous analysis (Article III, Table VI).

Table 2. Cross-sectional associations between hypertension genetic risk scores and clinical traits.

	Obesity N Loci = 25		**Dyslipidemia** N Loci = 14		**Hypolipidemia** N Loci = 27		**Short Stature** N Loci = 21		**SBP GRS** N Loci = 1,098,015	
Trait	**Beta (CI)**	**P Value**	**Beta (CI)**	**P Value**	**Beta (CI)**	**P Value**	**Beta (CI)**	**P Value**	**Beta (CI)**	**P Value**
BMI (kg/m^2)	**0.2919 (0.1878, 0.3960)**	**5.0×10^{-21}**	-0.0289 (-0.1332, 0.0755)	0.35	-0.0524 (-0.1569, 0.0521)	9.2×10^{-2}	-0.0162 (-0.1207, 0.0884)	0.60	**0.1468 (0.0420, 0.2516)**	**2.6×10^{-6}**
log CRP (mg/L)	0.0239 (-0.0019, 0.0496)	1.8×10^{-3}	-0.0006 (-0.0264, 0.0251)	0.93	0.0031 (-0.0227, 0.0289)	0.68	-0.0081 (-0.0177, 0.0339)	0.29	**0.0301 (0.0043, 0.0560)**	**9.2×10^{-5}**
eGFR (mL/min/ 1.73 m^2)	0.1321 (-0.1756, 0.4398)	0.15	-0.1571 (-0.4649, 0.1507)	8.6×10^{-2}	-0.0592 (-0.3647, 0.2491)	0.52	**0.3282 (0.0200, 0.6364)**	**3.5×10^{-4}**	0.0608 (-0.2484, 0.3701)	0.51
Height (m)	-0.0008 (-0.0022, 0.0006)	4.5×10^{-2}	-0.0010 (-0.0024, 0.0004)	2.0×10^{-2}	-0.0006 (-0.0020, 0.0008)	0.12	**-0.0028 (-0.0042, -0.0014)**	**9.2×10^{-12}**	**-0.0025 (-0.0039, -0.0011)**	**1.3×10^{-9}**
HDLc (mmol/L)	-0.0084 (-0.0169, 0.0000)	8.3×10^{-4}	**-0.0133 (-0.0218, -0.0049)**	**1.2×10^{-7}**	0.0000 (-0.0085, 0.0084)	0.99	-0.0025 (-0.0109, 0.0060)	0.32	**-0.0108 (-0.0193, -0.0023)**	**1.8×10^{-5}**
LDLc (mmol/L)	0.0078 (-0.0127, 0.0283)	0.20	-0.0028 (-0.0234, 0.0177)	0.64	**-0.0263 (-0.0469, -0.0057)**	**1.8×10^{-5}**	-0.0017 (-0.0222, 0.0189)	0.79	0.0005 (-0.0202, 0.0211)	0.94
TC (mmol/L)	0.0065 (-0.0165, 0.0295)	0.34	-0.0050 (-0.0279, 0.0180)	0.47	**-0.0280 (-0.0510, -0.0050)**	**4.5×10^{-5}**	-0.0017 (-0.0248, 0.0213)	0.80	0.0046 (-0.0185, 0.0277)	0.50
log TG (mmol/L)	0.0099 (0.0007, 0.0211)	1.7×10^{-3}	**0.0170 (0.0065, 0.0275)**	**6.2×10^{-8}**	-0.0045 (-0.0151, 0.0061)	0.15	0.0041 (-0.0065, 0.0146)	0.20	**0.0214 (0.0108, 0.0320)**	**1.2×10^{-11}**

The study sample consisted of 21,168 individuals from FINRISK 1997, 2002, 2007, and 2012 cohorts. We calculated betas from linear regression models adjusted for age, sex, cohort year, genotyping batch, and the first four genetic principal components. P values $<7.8\times10^{-4}$ and corresponding betas are bolded, representing a Bonferroni correction of 16 traits × 4 components. BMI indicates body mass index; CI, confidence interval with Bonferroni corrected confidence level 7.8×10^{-4}; CRP, C-reactive protein; eGFR, estimated glomerular filtration rate; GRS, genetic risk score; HDLc, high-density lipoprotein cholesterol; LDLc, low-density lipoprotein cholesterol; TC, total cholesterol; and TG, triglycerides.

5.3.3 Longitudinal associations between component GRSs and end points

Our FinnGen sample comprised 198,148 individuals, with 56% women and a mean age of 58 years at the end of follow-up. All hypertension component GRSs were associated with hypertension and cardiovascular disease (Article III, Table 4). For type 2 diabetes, the HRs per 1-SD increases in GRSs were 1.08 (CI, 1.05–1.10; $P=6.1\times10^{-33}$) for the Obesity component, 1.02 (CI, 1.00–1.05; $P=1.5\times10^{-4}$) for the Dyslipidemia component, and 1.02 (CI, 1.00–1.04; $P=1.8\times10^{-4}$) for the Short Stature component. For chronic kidney disease, the HR was 1.08 (CI, 1.01–1.16; $P=1.9\times10^{-4}$) for the Obesity component. For autoimmune diseases, the HRs were 1.03 (CI, 1.02–1.05; $P=8.3\times10^{-10}$) for the Dyslipidemia component, 1.05 (CI, 1.03–1.07; $P=5.1\times10^{-21}$) for the Hypolipidemia component, and 1.02 (CI, 1.00–1.04; $P=1.4\times10^{-4}$) for the Short Stature component. We observed strong associations between SBP GRS and all end points, except for autoimmune diseases. In addition, all component GRSs were associated with beta blocker, calcium channel blocker, renin-angiotensin-aldosterone system inhibitor, and thiazide diuretic use.

6 Discussion

6.1 Phenotypic hypertension subtypes

Study I suggests that the general hypertensive population contains a small subgroup of metabolically challenged individuals, characterized by elevated blood glucose and BMI, with an elevated CVD risk.

Some important differences between Study I and the study by Yang et al. make the results hard to compare [130]. First, while they had an adequate sample size of 9361 individuals and used several risk factors similar to ours, they excluded diabetic individuals. Our MC cluster consisted solely of individuals with diabetes. Nevertheless, Yang et al. observed a cluster with high BMI, which could share similarities with our MC cluster. Second, including the Framingham risk score for CVD into the clustering was a different methodological choice from ours. Because we attempted to find high CVD risk subgroups in an unsupervised manner, including a CVD risk score in the clustering would make the interpretation of our results difficult.

Interestingly, one of the four clusters of Guo et al. comprised 100% individuals with diabetes, much like our MC cluster [129]. However, we observed only two clusters while Guo et al. observed four. While it is difficult to explain the difference in the number of clusters, as Guo et al. did not report how they chose the number, sample characteristics could offer part of the explanation. Guo et al. excluded individuals with hypertension who were on antihypertensive medication, while 63% of our clustering sample were on antihypertensive medication. Furthermore, 41% of their study sample had had a stroke, which is a high proportion. The variation in results may also point to an inherent instability of phenotypic hypertension subtypes.

Previous research supports classifying the MC cluster as a hypertension subtype. First, it is known that hyperglycemia, the main characteristic of the MC cluster, leads to low-grade inflammation, which causes endothelial dysfunction and vascular stiffening [166]. The detrimental effects of hyperglycemia on vascular health are clearly illustrated by the Framingham study, where diabetes was related to almost 10-fold higher odds for impaired vasculature in old age [167]. Second, the 97% prevalence of the metabolic syndrome in the MC cluster suggests it is a strict subset, or a subtype, of metabolic syndrome. Metabolic syndrome is well-studied and

associated with numerous cardiovascular complications [168]. However, the question of whether the MC cluster represents a novel mechanistic subtype of hypertension is out of scope for Study I. Additional insight might have been gained by clustering the reference group, or an additional "optimal BP" group, in Study I with respect to the same variables as the clustering group. Still, the subtype seems more likely to represent undiagnosed or uncontrolled diabetes resulting from poor lifestyle choices.

6.2 Genetic hypertension subtypes (II and III)

6.2.1 Risk stratification using polygenic risk scores (II)

Study II demonstrates that BP PRSs predict incident hypertension and improve the Framingham model for near-term incidence of hypertension.

Previous research has demonstrated that individuals with early-onset hypertension are at an increased risk of CVD death and organ damage compared to late-onset hypertension [169,170]. While the heritability of hypertension is well established, only early-onset hypertension has been shown to reliably cross generations from parents to offspring [169,171]. Study II demonstrates that hypertension has a substantial genetic component in both early-onset and late-onset hypertension, but the association appears stronger in early-onset hypertension. In particular, 1-SD increases of SBP and DBP PRSs were associated with 54% and 58% greater risks of early-onset hypertension, respectively, while the corresponding risk increases were only 31% and 26% for late-onset hypertension. Therefore, Study II supports early genotyping for high-risk individuals for improved risk assessment.

Previous investigations of BP PRSs have not detected improvements in clinical risk prediction of hypertension [8–10]. In Study II, we used an externally validated Framingham office-based risk prediction model that could be used in any clinic [125]. After including both SBP and DBP PRSs in the risk prediction model, the C statistics increased by 0.7 percentage points from 79.7% to 80.4% (P=0.0017). These results suggest that BP PRSs could be implemented as a complementary tool alongside traditional risk factors to improve hypertension prediction in the office on a population level, although their clinical importance on an individual level still appears limited. However, studies and validation studies are needed because no externally validated hypertension risk score currently combines clinical and genetic information.

Increasing the training sample size is a common way to increase the predictive power of prediction models. However, Study II demonstrates that methodological choices should also be carefully considered. In the largest BP GWAS meta-analysis to date, Evangelou et al. constructed an SBP GRS based on 760,000 individuals and

886 SNPs, demonstrating a 12.9 mmHg difference in SBP between top and bottom GRS deciles in unrelated European individuals from the UK biobank [140]. In Study II, we could replicate this SBP difference at 10.6 mmHg in a comparable FINRISK sample. However, the difference in SBP between top and bottom PRS deciles using a PRS based on only 340,000 individuals from the UK biobank and 1.1 million SNPs was 14.1 mmHg in FINRISK, representing a 33% improvement in performance. This improvement was likely due to PRS-CS, a recently introduced method for constructing PRSs, although differences in GWAS and test data make it difficult to pinpoint the exact source of improvement [145]. As sample sizes in GWASs are already extremely large, methodological developments may be the key to achieving continued improvements in polygenic risk prediction.

6.2.2 Genetic hypertension subtypes (III)

Study III suggests the existence of four genetic hypertension subtypes that correspond to common clinical phenotypes and capture different aspects of hypertension risk.

While genetic risk stratification with an overall GRS is a simple way to stratify individuals into risk-based subtypes, a linear risk predictor cannot capture potential mechanistic phenotypes. To capture qualitatively different phenotypes, a multidimensional analysis is required. In Study III, we performed an unsupervised data-driven analysis that incorporated genetic information about hypertension and about 16 quantitative clinical traits. A recently introduced factorization algorithm allowed us to recognize four data-driven hypertension components: Obesity, Dyslipidemia, Hypolipidemia, and Short Stature. Cross-sectional and longitudinal analyses confirmed that these latent components captured different aspects of genetic hypertension risk. Conceptually, this study demonstrates that GRSs can be used to not only predict hypertension risk but also to give insight into disease heterogeneity and facilitate subtyping.

Remarkably, despite a completely unsupervised framework with no input about the desired outcome, the clinical trait combinations and genetic variants for these components were supported by previous literature. Obesity and dyslipidemia are well-understood comorbidities of hypertension that appear together in metabolic syndrome, which was, perhaps only coincidentally, also captured by Study I at the phenotype level. Two epidemiological observations support the Hypolipidemia component: first, the gene at *SH2B3/LNK* links hypertension with chronic inflammation [172], and second, chronic inflammation is linked with hypolipidemia [173]. Genetic variants at *SH2B3/LNK* and *BANK1* in the Hypolipidemia component have been linked with autoimmune diseases (rheumatoid arthritis and Crohn's disease), and the Hypolipidemia component was associated with autoimmune

diseases in our longitudinal analyses [174,175]. While the Short Stature component may seem unintuitive, it has strong epidemiological support from, for example, Bourgeois et al., who demonstrated a robust association between elevated SBP and short stature in a multi-ethnic cohort of 13,000 individuals [176]. The individual traits and genetic variants in Study III were already well known, as are these four clinical phenotypes. However, our unsupervised, data-driven analysis indicates that they may arise naturally from hypertension genetics.

In Study III, we tested the associations of these components in a sample that was independent of the derivation cohorts. The cross-sectional associations in FINRISK (n=21,168) between the four hypertension component GRSs and clinical traits were as expected (Table 2). Moreover, longitudinal associations between disease end points and hypertension components in FinnGen (n=198,148) further solidified their clinical significance (Table 3). The Obesity component was associated with type 2 diabetes and the Hypolipidemia component with autoimmune diseases, consistent with the hypothesis that chronic inflammation is a mediator between hypertension and the Hypolipidemia component.

Two studies by Ma et al. and Luo et al. genetically subtyped hypertensive participants in the HyperGEN cohort [142,143]. Ma et al. genetically subtyped 1187 hypertensive participants and evaluated their clinical characteristics with nine echocardiographic variables [142]. The main methodological difference to Study III is that Ma et al. directly subtyped individuals with raw allele counts. In contrast, we subtyped hypertension SNPs tagged with multi-trait information about other related traits. Because FINRISK does not contain echocardiographic measurements, directly comparing results is impossible. Moreover, the method used by Ma et al. implicitly assumes an underlying one-to-one correspondence between genetic subtypes and phenotypic subtypes, which is not necessarily the case because the phenotype arises via complex interactions between genes and the environment. In a slightly different approach, Luo et al. genetically subtyped 660 hypertensive participants using their novel method called HNMF [143]. The idea behind HNMF is to factorize the phenotype and genotype simultaneously, producing genetic subtypes that are directly supported by phenotypic traits. This idea is, in principle, identical to ours. However, while Luo et al. used a cohort of 660 individuals to infer associations between phenotype and genotype, we used publicly available GWAS summary statistics based on large samples in the 100,000's independent of our study sample.

Table 3. Longitudinal associations between hypertension genetic risk scores and incident disease end points.

		Obesity N Loci = 25		**Dyslipidemia** N Loci = 14		**Hypolipidemia** N Loci = 27		**Short Stature** N Loci = 21		**SBP GRS** N Loci = 1,098,015	
End Point	**Cases / Controls**	**Hazard Ratio (CI)**	**P Value**	**Hazard Ratio (CI)**	**P Value**	**Hazard Ratio (CI)**	**P Value**	**Hazard Ratio (CI)**	**P Value**	**Hazard Ratio (CI)**	**P Value**
Hypertension	44,472 / 153,676	**1.07 (1.05-1.08)**	**1.0×10^{-40}**	**1.05 (1.04-1.07)**	**4.9×10^{-27}**	**1.15 (1.13-1.16)**	**8.2×10^{-178}**	**1.09 (1.07-1.11)**	**9.9×10^{-74}**	**1.43 (1.41-1.46)**	**$<10^{-300}$**
Cardiovascular disease	23,562 / 174,586	**1.03 (1.01-1.05)**	**2.2×10^{-6}**	**1.04 (1.01-1.06)**	**1.4×10^{-7}**	**1.07 (1.04-1.09)**	**5.3×10^{-22}**	**1.05 (1.03-1.07)**	**1.8×10^{-14}**	**1.14 (1.11-1.16)**	**2.1×10^{-80}**
Coronary heart disease	16,419 / 181,729	**1.03 (1.01-1.06)**	**3.3×10^{-5}**	**1.04 (1.01-1.07)**	**2.0×10^{-7}**	**1.07 (1.04-1.10)**	**5.5×10^{-17}**	**1.05 (1.03-1.08)**	**4.6×10^{-11}**	**1.16 (1.13-1.19)**	**9.7×10^{-74}**
Stroke	9,819 / 188,329	1.03 (1.00-1.07)	1.2×10^{-3}	1.03 (0.99-1.06)	8.5×10^{-3}	**1.06 (1.03-1.10)**	**4.6×10^{-9}**	**1.05 (1.01-1.08)**	**8.5×10^{-6}**	**1.10 (1.06-1.14)**	**1.1×10^{-20}**
Heart failure	11,454 / 186,694	**1.05 (1.02-1.09)**	**8.4×10^{-8}**	1.03 (0.99-1.06)	4.9×10^{-3}	**1.05 (1.02-1.09)**	**2.2×10^{-8}**	**1.04 (1.00-1.07)**	**2.1×10^{-4}**	**1.15 (1.11-1.18)**	**2.1×10^{-45}**
Type 2 diabetes	26,404 / 171,744	**1.08 (1.05-1.10)**	**6.1×10^{-33}**	**1.02 (1.00-1.05)**	**1.5×10^{-4}**	0.99 (0.97-1.01)	0.21	**1.02 (1.00-1.04)**	**1.8×10^{-4}**	**1.12 (1.10-1.15)**	**1.7×10^{-76}**
Chronic kidney disease	2,362 / 195,786	**1.08 (1.01-1.16)**	**1.9×10^{-4}**	1.04 (0.97-1.12)	4.2×10^{-2}	1.06 (0.99-1.14)	2.8×10^{-3}	1.04 (0.97-1.11)	7.6×10^{-2}	**1.12 (1.04-1.21)**	**5.0×10^{-8}**
Autoimmune diseases	34,312 / 163,836	1.01 (0.99-1.02)	0.35	**1.03 (1.02-1.05)**	**8.3×10^{-10}**	**1.05 (1.03-1.07)**	**5.1×10^{-21}**	**1.02 (1.00-1.04)**	**1.4×10^{-4}**	1.01 (0.99-1.03)	8.9×10^{-2}

The study sample consisted of 198,148 individuals from FinnGen. We calculated hazard ratios per one standard deviation increases from Cox proportional hazards models with age as the time scale. We adjusted the models for sex, DNA sample collection year, genotyping batch, and the first ten genetic principal components. P values $<7.8\times10^{-4}$ and corresponding hazard ratios are bolded, representing a Bonferroni correction of 16 traits × 4 components. CI indicates confidence interval with Bonferroni corrected confidence level 7.8×10^{-4}; SBP GRS, genetic risk score for systolic blood pressure.

6.3 Limitations of the study

In Study I, combining Z-scores from glycated hemoglobin and fasting glucose to quantify blood glucose is not optimal, but it was supported by a moderately high Pearson correlation of 0.58 between the variables. Second, the clustering observed was not particularly strong, and the separation in PCA plots was incomplete. Direct data grouping with unsupervised clustering algorithms may not be optimal for identifying substructures in hypertension. Other methods, such as unsupervised non-linear dimension reduction algorithms, could yield better results [177]. Finally, direct measurements of the cardiovascular system, e.g., via the echocardiogram, could have improved cluster quality. In Study II, parental history of hypertension could not be included in the Framingham hypertension equation because it was not available in our sample. However, the Framingham equation is based on ascertained family history, while in the clinical setting, it is usually self-reported and may not be accurate. Both Study II and III used GWAS results from only 340,000 individuals (UK Biobank) instead of the large meta-analysis of 760,000 individuals from the International Consortium for Blood Pressure used by Evangeou et al. [140]. We did this to avoid overfitting because some FINRISK cohorts are part of the International Consortium for Blood Pressure. Moreover, both Study II and Study III were based on individuals of European ancestry and, therefore, the results might not be generalizable to other ancestries. Finally, in Study III, we accounted for linkage disequilibrium by only including independent hypertension SNPs in the analysis. Adjusting effect sizes for linkage disequilibrium instead could allow more variants to be included, which would increase the power to detect genetic components.

6.4 Clinical implications

Previous studies and our results suggest that the data poorly support well-defined phenotypic subtypes of hypertension. The strongest support is given for a small hyperglycemic phenotype, corresponding clinically to uncontrolled diabetes. While this is an interesting observation, the need to control diabetes in hypertensive individuals has long been recognized, and the importance of glycemic control is already emphasized in hypertension guidelines. Moreover, while it is unclear whether this hypothetical hyperglycemic subtype results from lifestyle choices or underlying metabolic impairment, current literature already provides good physiological explanations for poor glycemic control in hypertensive individuals. In conclusion, we recommend that clinicians continue to pay particular attention to blood glucose levels in hypertensive individuals.

Knowledge of hypertension genetics has been rapidly expanding in recent years, yet demonstrations of clinical relevance have been lacking. Our results demonstrate that including genetic information in clinical risk prediction models for hypertension

would improve their predictive ability. Furthermore, it appears that individual genetic makeup is particularly strongly linked to hypertension that is diagnosed before midlife. Therefore, young individuals with a family history of hypertension could benefit from polygenic risk assessment. Finally, genetics has the potential to not only place individuals on a linear risk scale from low risk to high risk but also to identify qualitatively different hypertension components that predict distinct aspects of hypertension risk, such as obesity and autoimmune disease.

6.5 Future prospects

Commonly used clinical variables appear to be inadequate for identifying phenotypic hypertension clusters. Future studies should investigate hypertension clustering via hypertension-specific phenotyping of the cardiovascular system and attempt to discover optimal therapies for these clusters. Methodological alternatives to hierarchical clustering, such as non-linear dimension reduction algorithms, should also be considered. BP PRSs improve clinical risk prediction of hypertension, but validated risk equations combining PRSs with clinical variables are lacking. Future studies should develop and validate risk equations that include both phenotypic and genetic information and further develop PRS methodology. Moreover, studies examining the effect of PRS-based genetic risk counseling on health behavior and hypertension prevention are needed. Planning information systems and policies for storing large quantities of genetic information in hospitals and health centers may soon become necessary as genetic information becomes commonplace in the clinical setting. Finally, our results demonstrate a component-based representation of the genetic basis of hypertension. Future refinements of these components could benefit personalized risk prediction and genetic risk counseling, but advancements in genetic subtyping methodology are likely needed to achieve this.

7 Summary/Conclusions

This thesis aimed to examine the phenotypic and genetic heterogeneity of hypertension from a clinical perspective. To achieve this, we applied publicly available supervised and unsupervised subtyping algorithms to large datasets. Our data comprised genotyped cohort and biobank studies with long-term follow-up for cardiovascular end points, as well as publicly available genetic association data. This approach permitted us to reliably estimate cross-sectional and longitudinal associations between the observed heterogeneity and clinically relevant features such as anthropometric measures, biomarkers, and cardiovascular end points. Because data-analytic methods are highly dependent on input data and methodology, the results should be viewed as rough characterizations of hypertension heterogeneity that require further investigation.

We observed a phenotypic hypertension subgroup characterized by hyperglycemia and elevated BMI. The subgroup is supported by pathophysiological connections between hyperglycemia and blood pressure and the comorbidity between hypertension and metabolic disease. However, the subtype was small and incompletely separated from the main cohort. Moreover, it is unclear whether this subtype represents a mechanistic subtype, lifestyle choices, or a mixture of both.

BP PRSs can stratify individuals from low to high hypertension risk. This stratification adds clinical utility over traditional risk equations based on clinical variables. Because of decreasing genotyping costs, genetic data will likely become more commonplace in the clinical setting. However, the current literature lacks validated risk equations that combine clinical and genetic variables. Therefore, future research should conduct training and validation studies for such equations. Moreover, while we demonstrated clinical utility, it is still unclear whether BP PRSs have added utility over parental history of hypertension, including age of onset. Finally, continued methodological development of both PRS and genetic subtyping algorithms must not be ignored.

We observed four genetic hypertension components corresponding to recognizable clinical features: obesity, dyslipidemia, hypolipidemia, and short stature. An individual's hypertension risk can be understood as a mixture of these components, and all four components may be estimated for any given individual

using only 73 genetic variants. High-resolution refinements of these low-resolution genetic hypertension components could benefit personalized healthcare in the future via personalized risk prediction and genetic risk counseling.

Acknowledgements

The work for this thesis was carried out in the Department of Medicine, University of Turku, Finland. I was privileged to work with data collected and curated by the investigators of two important Finnish studies: FINRISK and FinnGen. The staff and participants of these two studies deserve an enormous amount of credit for their contributions to this thesis and science in general.

I wish to express my deepest gratitude to my supervisor Professor Teemu Niiranen who introduced me to a dynamic field of research and an outstanding team of colleagues. You provided me with the tools, resources, and opportunities to grow into a scientist, and your mentorship extended well beyond this thesis. I admire your writing and leadership skills, and I am constantly trying to learn from you.

Likewise, I wish to express my sincere gratitude to my second supervisor Associate Professor Leo Lahti. In addition to sharing your mathematical and methodological expertise, you have encouraged me to think back to the fundamentals. Your perspective on science has influenced the type of research questions I wish to tackle in the future.

I am enormously indebted to Assistant Professor Georg Ehret and Associate Professor Matti Pirinen for reviewing this thesis. To say that your comments were helpful would be a massive understatement. I extend my gratitude to Professor Jussi Hernesniemi and Dr. Kimmo Porthan for being on my follow-up committee. I am highly honored that Professor J. Gustav Smith agreed to travel from Sweden to be my opponent in the public examination of this thesis – I am nervously waiting to hear your thoughts.

I wish to express my warmest thanks to all my co-authors in the papers that make up this thesis: Dr. Anni Kauko, Dr. Karri Suvila, Docent Aki Havulinna, Docent Nina Mars, Ms. Hyunkyung Kim, Assistant Professor Miriam Udler, Docent Ilkka Kantola, Professor Susan Cheng, Professor Risto Kaaja, and Professor Veikko Salomaa. Our joint efforts elevated the quality of every single paper and taught me the value of collaboration. I also wish to thank all the fantastic people at Turku Hypertension Center and Turku Data Science Group for the shared moments in science and leisure. Special thanks go to Dr. Joonatan Palmu, Mr. Jouko Nurkkala, and Docent Jenni Aittokallio for our scientific collaborations and adventures abroad.

These studies were financially supported by personal grants from the University of Turku and the Finnish Hypertension Society, which are gratefully acknowledged. I also wish to thank the Faculty of Medicine staff for being patient and accommodating whenever my research activities clashed with my schedule in medical school.

I have been blessed with several academic teachers and mentors that I wish to acknowledge. I thank Kari Vahtera for nurturing my early interest in learning, Ulpu Valo and Ari Lius for introducing me to science, Dr. Simon Wadsley for teaching me rigour, Professor Enrico Coen for demonstrating the importance of stories, Dr. Hugh Woolfenden for sparking my interest in biology and medicine, and Professor Risto Kaaja and Docent Maija Hollmén for introducing me to medical research.

Life exists outside of research, and I want to thank my friends and family for always being there for me. In particular, I wish to thank my parents for their unconditional support in whatever I do and my sister Sofia for helping me keep up with the times.

Finally, I want to thank the light of my life and partner in crime, Ada, for making life fun and my son, Christopher, for reminding me what's important.

September 2022
Felix Vaura

References

1. Page IH. The mosaic theory 32 years later. Hypertension. 1982;4:177.
2. Harrison DG, Coffman TM, Wilcox CS. Pathophysiology of Hypertension: The Mosaic Theory and Beyond. Circ Res. 2021;128:847–863.
3. Staessen JA, Wang J, Bianchi G, Birkenhäger WH. Essential hypertension. Lancet. 2003;361:1629–1641.
4. Savoia C, Volpe M, Grassi G, Borghi C, Agabiti Rosei E, Touyz RM. Personalized medicine-a modern approach for the diagnosis and management of hypertension. Clin Sci. 2017;131:2671–2685.
5. Whelton PK, Carey RM, Aronow WS, Casey DE Jr, Collins KJ, Dennison Himmelfarb C, 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: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension. 2018;71:e13–e115.
6. Echouffo-Tcheugui JB, Batty GD, Kivimäki M, Kengne AP. Risk models to predict hypertension: a systematic review. PLoS One. 2013;8:e67370.
7. Padmanabhan S, Dominiczak AF. Genomics of hypertension: the road to precision medicine. Nat Rev Cardiol. 2021;18:235–250.
8. Niiranen TJ, Havulinna AS, Langén VL, Salomaa V, Jula AM. Prediction of Blood Pressure and Blood Pressure Change With a Genetic Risk Score. J Clin Hypertens. 2016;18:181–186.
9. Fava C, Sjögren M, Montagnana M, Danese E, Almgren P, Engström G, et al. Prediction of blood pressure changes over time and incidence of hypertension by a genetic risk score in Swedes. Hypertension. 2013;61:319–326.
10. Lim N-K, Lee J-Y, Lee J-Y, Park H-Y, Cho M-C. The Role of Genetic Risk Score in Predicting the Risk of Hypertension in the Korean population: Korean Genome and Epidemiology Study. PLoS One. 2015;10:e0131603.
11. Tigerstedt R, Bergman PQ. Niere und Kreislauf. Skand Arch Für Physiol. 1898;8:223–271.
12. Sarre H. [Franz Volhard--the red and pale high pressure in the modern viewpoint. 1]. Med Klin. 1973;68:1309–1312.
13. Luft FC, Dietz R. Franz Volhard in historical perspective. Hypertension. 1993;22:253–256.
14. Goldblatt H, Lynch J, Hanzal RF, Summerville WW. Studies on experimental hypertension: I. The production of persistent elevation of systolic blood pressure by means of renal ischemia. J Exp Med. 1934;59:347–379.
15. Page IH, Helmer OM. A crystalline pressor substance (angiotonin) resulting from the reaction between renin and renin-activator. J Exp Med. 1940;71:29–42.
16. Page IH. Pathogenesis of arterial hypertension. J Am Med Assoc. 1949;140:451–458.
17. Koch E, Mies H. Chronischer arterieller Hochdruck durch experimentelle Dauerausschaltung des Blutdruckzugler. Krankheitsforschung. 1929;7:241–256.
18. Grulee CG Jr, Panos TC. Epidemic poliomyelitis in children; clinical study, with special reference to symptoms and management of bulbar polioencephalitis. Am J Child. 1948;75:24–39.
19. Selye H. Production of Nephrosclerosis by Overdosage with Desoxycorticosterone Acetate. Can Med Assoc J. 1942;47:515–519.
20. Steele JM. Evidence for general distribution of peripheral resistance in coarctation of the aorta: report of three cases. J Clin Invest. 1941;20:473–480.

21. Page IH. The nature of arterial hypertension. Arch Intern Med. 1963;111:103–115.
22. Ghezzi P, Jaquet V, Marcucci F, Schmidt HHHW. The oxidative stress theory of disease: levels of evidence and epistemological aspects. Br J Pharmacol. 2017;174:1784–1796.
23. Griendling KK, Camargo LL, Rios FJ, Alves-Lopes R, Montezano AC, Touyz RM. Oxidative Stress and Hypertension. Circ Res. 2021;128:993–1020.
24. Madhur MS, Elijovich F, Alexander MR, Pitzer A, Ishimwe J, Van Beusecum JP, et al. Hypertension: Do Inflammation and Immunity Hold the Key to Solving this Epidemic? Circ Res. 2021;128:908–933.
25. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. N Engl J Med. 2017;377:1119–1131.
26. Everett BM, Cornel JH, Lainscak M, Anker SD, Abbate A, Thuren T, et al. Anti-Inflammatory Therapy With Canakinumab for the Prevention of Hospitalization for Heart Failure. Circulation. 2019;139:1289–1299.
27. Ridker PM, MacFadyen JG, Everett BM, Libby P, Thuren T, Glynn RJ, et al. Relationship of C-reactive protein reduction to cardiovascular event reduction following treatment with canakinumab: a secondary analysis from the CANTOS randomised controlled trial. Lancet. 2018;391:319–328.
28. Ridker PM, Libby P, MacFadyen JG, Thuren T, Ballantyne C, Fonseca F, et al. Modulation of the interleukin-6 signalling pathway and incidence rates of atherosclerotic events and all-cause mortality: analyses from the Canakinumab Anti-Inflammatory Thrombosis Outcomes Study (CANTOS). Eur Heart J. 2018;39:3499–3507.
29. Kopp C, Linz P, Wachsmuth L, Dahlmann A, Horbach T, Schöfl C, et al. (23)Na magnetic resonance imaging of tissue sodium. Hypertension. 2012;59:167–172.
30. Li J, Zhao F, Wang Y, Chen J, Tao J, Tian G, et al. Gut microbiota dysbiosis contributes to the development of hypertension. Microbiome. 2017;5:14.
31. Ferguson JF, Aden LA, Barbaro NR, Van Beusecum JP, Xiao L, Simmons AJ, et al. High dietary salt-induced dendritic cell activation underlies microbial dysbiosis-associated hypertension. JCI Insight. 2019;5: e126241.
32. Rimoldi SF, Scherrer U, Messerli FH. Secondary arterial hypertension: when, who, and how to screen? Eur Heart J. 2014;35:1245–1254.
33. Kostis JB, Cabrera J, Cheng JQ, Cosgrove NM, Deng Y, Pressel SL, et al. Association between chlorthalidone treatment of systolic hypertension and long-term survival. JAMA. 2011;306:2588–2593.
34. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. Final results of the Systolic Hypertension in the Elderly Program (SHEP). SHEP Cooperative Research Group. JAMA. 1991;265:3255–3264.
35. Five-year findings of the hypertension detection and follow-up program. I. Reduction in mortality of persons with high blood pressure, including mild hypertension. Hypertension Detection and Follow-up Program Cooperative Group. JAMA. 1979;242:2562–2571.
36. Effects of treatment on morbidity in hypertension. Results in patients with diastolic blood pressures averaging 115 through 129 mm Hg. JAMA. 1967;202:1028–1034.
37. Cheng S, Xanthakis V, Sullivan LM, Vasan RS. Blood Pressure Tracking Over the Adult Life Course. Hypertension. 2012;60:1393–1399.
38. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R, Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet. 2002;360:1903–1913.
39. Huang J, Wildman RP, Gu D, Muntner P, Su S, He J. Prevalence of isolated systolic and isolated diastolic hypertension subtypes in China. Am J Hypertens. 2004;17:955–962.
40. Antikainen R, Jousilahti P, Tuomilehto J. Systolic blood pressure, isolated systolic hypertension and risk of coronary heart disease, strokes, cardiovascular disease and all-cause mortality in the middle-aged population. J Hypertens. 1998;16:577–583.
41. Wilking SVB, Van B. Wilking S. Determinants of Isolated Systolic Hypertension. JAMA J Am Med Assoc. 1988;260: 3451.

42. Garland C, Barrett-Connor E, Suarez L, Criqui MH. Isolated systolic hypertension and mortality after age 60 years: a prospective population-based study. Am J Epidemiol. 1983;118:365–376.
43. Staessen JA, Fagard R, Thijs L, Celis H, Arabidze GG, Birkenhäger WH, et al. Randomised double-blind comparison of placebo and active treatment for older patients with isolated systolic hypertension. The Systolic Hypertension in Europe (Syst-Eur) Trial Investigators. Lancet. 1997;350:757–764.
44. Liu L, Wang JG, Gong L, Liu G, Staessen JA. Comparison of active treatment and placebo in older Chinese patients with isolated systolic hypertension. Systolic Hypertension in China (Syst-China) Collaborative Group. J Hypertens. 1998;16:1823–1829.
45. Beckett NS, Peters R, Fletcher AE, Staessen JA, Liu L, Dumitrascu D, et al. Treatment of Hypertension in Patients 80 Years of Age or Older. N Engl J Med. 2008;358:1887–1898.
46. Franklin SS, Pio JR, Wong ND, Larson MG, Leip EP, Vasan RS, et al. Predictors of New-Onset Diastolic and Systolic Hypertension. Circulation. 2005;111:1121–1127.
47. Li Y, Wei F-F, Wang S, Cheng Y-B, Wang J-G. Cardiovascular risks associated with diastolic blood pressure and isolated diastolic hypertension. Curr Hypertens Rep. 2014;16:489.
48. Li F-R, He Y, Yang H-L, Liu H-M, Zhou R, Chen G-C, et al. Isolated systolic and diastolic hypertension by the 2017 American College of Cardiology/American Heart Association guidelines and risk of cardiovascular disease: a large prospective cohort study. J Hypertens. 2021;39:1594–1601.
49. Kaneko H, Itoh H, Yotsumoto H, Kiriyama H, Kamon T, Fujiu K, et al. Association of Isolated Diastolic Hypertension Based on the Cutoff Value in the 2017 American College of Cardiology/American Heart Association Blood Pressure Guidelines With Subsequent Cardiovascular Events in the General Population. J Am Heart Assoc. 2020;9:e017963.
50. McGrath BP, Kundu P, Daya N, Coresh J, Selvin E, McEvoy JW, et al. Isolated Diastolic Hypertension in the UK Biobank: Comparison of ACC/AHA and ESC/NICE Guideline Definitions. Hypertension. 2020;76:699–706.
51. Jacobsen AP, Al Rifai M, Arps K, Whelton SP, Budoff MJ, Nasir K, et al. A cohort study and meta-analysis of isolated diastolic hypertension: searching for a threshold to guide treatment. Eur Heart J. 2021;42:2119–2129.
52. Conell C, Flint AC, Ren X, Banki NM, Chan SL, Rao VA, et al. Underdiagnosis of Isolated Systolic and Isolated Diastolic Hypertension. Am J Cardiol. 2021;141:56–61.
53. Orias M, Tabares AH, Peixoto AJ. Hypothesis: it is time to reconsider phenotypes in hypertension. J Clin Hypertens. 2010;12:350–356.
54. Tientcheu D, Ayers C, Das SR, McGuire DK, de Lemos JA, Khera A, et al. Target Organ Complications and Cardiovascular Events Associated With Masked Hypertension and White-Coat Hypertension: Analysis From the Dallas Heart Study. J Am Coll Cardiol. 2015;66:2159–2169.
55. Mancia G, Bombelli M, Cuspidi C, Facchetti R, Grassi G. Cardiovascular Risk Associated With White-Coat Hypertension: Pro Side of the Argument. Hypertension. 2017;70:668–675.
56. Sega R, Trocino G, Lanzarotti A, Carugo S, Cesana G, Schiavina R, et al. Alterations of cardiac structure in patients with isolated office, ambulatory, or home hypertension: Data from the general population (Pressione Arteriose Monitorate E Loro Associazioni [PAMELA] Study). Circulation. 2001;104:1385–1392.
57. Cohen JB, Lotito MJ, Trivedi UK, Denker MG, Cohen DL, Townsend RR. Cardiovascular Events and Mortality in White Coat Hypertension: A Systematic Review and Meta-analysis. Ann Intern Med. 2019;170:853–862.
58. Huang Y, Huang W, Mai W, Cai X, An D, Liu Z, et al. White-coat hypertension is a risk factor for cardiovascular diseases and total mortality. J Hypertens. 2017;35:677–688.
59. Briasoulis A, Androulakis E, Palla M, Papageorgiou N, Tousoulis D. White-coat hypertension and cardiovascular events: a meta-analysis. J Hypertens. 2016;34:593–599.
60. Mancia G, Bombelli M, Brambilla G, Facchetti R, Sega R, Toso E, et al. Long-term prognostic value of white coat hypertension: an insight from diagnostic use of both ambulatory and home blood pressure measurements. Hypertension. 2013;62:168–174.

61. Mancia G, Facchetti R, Bombelli M, Cuspidi C, Grassi G. White-Coat Hypertension: Pathophysiological and Clinical Aspects: Excellence Award for Hypertension Research 2020. Hypertension. 2021;78:1677–1688.
62. Hänninen M-RA, Niiranen TJ, Puukka PJ, Kesäniemi YA, Kähönen M, Jula AM. Target organ damage and masked hypertension in the general population: the Finn-Home study. J Hypertens. 2013;31:1136–1143.
63. Landsbergis PA, Dobson M, Koutsouras G, Schnall P. Job strain and ambulatory blood pressure: a meta-analysis and systematic review. Am J Public Health. 2013;103:e61-71.
64. Ohkubo T, Kikuya M, Metoki H. Prognosis of "Masked" Hypertension and "White-Coat" Hypertension Detected by 24-h Ambulatory Blood Pressure Monitoring. 10-Year Follow-up From the Ohasama Study. J Am Coll Cariol. 2005;46:508–515.
65. Cacciolati C, Tzourio C, Hanon O. Blood pressure variability in elderly persons with white-coat and masked hypertension compared to those with normotension and sustained hypertension. Am J Hypertens. 2013;26:367–372.
66. Mezick EJ, Hall M, Matthews KA. Sleep duration and ambulatory blood pressure in black and white adolescents. Hypertension. 2012;59:747–752.
67. Ogedegbe G, Spruill TM, Sarpong DF, Agyemang C, Chaplin W, Pastva A, et al. Correlates of isolated nocturnal hypertension and target organ damage in a population-based cohort of African Americans: the Jackson Heart Study. Am J Hypertens. 2013;26:1011–1016.
68. Franklin SS, Thijs L, Li Y, Hansen TW, Boggia J, Liu Y, et al. Masked hypertension in diabetes mellitus: treatment implications for clinical practice. Hypertension. 2013;61:964–971.
69. Agarwal R, Sinha AD, Light RP. Toward a definition of masked hypertension and white-coat hypertension among hemodialysis patients. Clin J Am Soc Nephrol. 2011;6:2003–2008.
70. Yano Y, Bakris GL. Recognition and management of masked hypertension: a review and novel approach. J Am Soc Hypertens. 2013;7:244–252.
71. Palla M, Saber H, Konda S, Briasoulis A. Masked hypertension and cardiovascular outcomes: an updated systematic review and meta-analysis. Integr Blood Press Control. 2018;11:11–24.
72. Franklin SS, O'Brien E, Staessen JA. Masked hypertension: understanding its complexity. Eur Heart J. 2017;38:1112–1118.
73. Calhoun DA, Jones D, Textor S, Goff DC, Murphy TP, Toto RD, et al. Resistant hypertension: diagnosis, evaluation, and treatment: a scientific statement from the American Heart Association Professional Education Committee of the Council for High Blood Pressure Research. Circulation. 2008;117:e510-26.
74. Achelrod D, Wenzel U, Frey S. Systematic review and meta-analysis of the prevalence of resistant hypertension in treated hypertensive populations. Am J Hypertens. 2015;28:355–361.
75. Buhnerkempe MG, Botchway A, Nolasco Morales CE, Prakash V, Hedquist L, Flack JM. Predicting the risk of apparent treatment-resistant hypertension: a longitudinal, cohort study in an urban hypertension referral clinic. J Am Soc Hypertens. 2018;12:809–817.
76. Daugherty SL, Powers JD, Magid DJ, Tavel HM, Masoudi FA, Margolis KL, et al. Incidence and prognosis of resistant hypertension in hypertensive patients. Circulation. 2012;125:1635–1642.
77. Sim JJ, Bhandari SK, Shi J, Reynolds K, Calhoun DA, Kalantar-Zadeh K, et al. Comparative risk of renal, cardiovascular, and mortality outcomes in controlled, uncontrolled resistant, and nonresistant hypertension. Kidney Int. 2015;88:622–632.
78. Strauch B, Petrák O, Zelinka T, Rosa J, Somlóová Z, Indra T, et al. Precise assessment of noncompliance with the antihypertensive therapy in patients with resistant hypertension using toxicological serum analysis. J Hypertens. 2013;31:2455–2461.
79. Lloberes P, Lozano L, Sampol G, Romero O, Jurado MJ, Ríos J, et al. Obstructive sleep apnoea and 24-h blood pressure in patients with resistant hypertension. J Sleep Res. 2010;19:597–602.
80. Muxfeldt ES, Margallo VS, Guimaraes GM, Salles GF. Prevalence and Associated Factors of Obstructive Sleep Apnea in Patients with Resistant Hypertension. Am J Hypertens. 2014;27:1069–1078.
81. Graves JW, Bloomfield RL, Buckalew VM Jr. Plasma volume in resistant hypertension: guide to pathophysiology and therapy. Am J Med Sci. 1989;298:361–365.

82. Václavík J, Sedlák R, Plachy M, Navrátil K, Plásek J, Jarkovsky J, et al. Addition of spironolactone in patients with resistant arterial hypertension (ASPIRANT): a randomized, double-blind, placebo-controlled trial. Hypertension. 2011;57:1069–1075.
83. Oxlund CS, Henriksen JE, Tarnow L, Schousboe K, Gram J, Jacobsen IA. Low dose spironolactone reduces blood pressure in patients with resistant hypertension and type 2 diabetes mellitus: a double blind randomized clinical trial. J Hypertens. 2013;31:2094–2102.
84. Václavík J, Sedlák R, Jarkovský J, Kociánová E, Táborský M. Effect of spironolactone in resistant arterial hypertension: a randomized, double-blind, placebo-controlled trial (ASPIRANT-EXT). Medicine (Baltimore). 2014;93:e162.
85. Acelajado MC, Pisoni R, Dudenbostel T, Dell'Italia LJ, Cartmill F, Zhang B, et al. Refractory hypertension: definition, prevalence, and patient characteristics. J Clin Hypertens. 2012;14:7–12.
86. Modolo R, de Faria AP, Almeida A, Moreno H. Resistant or Refractory Hypertension: Are They Different? Curr Hypertens Rep. 2014;16:485.
87. Muxfeldt ES, Chedier B, Rodrigues CIS. Resistant and refractory hypertension: two sides of the same disease? J Bras Nefrol. 2019;41:266–274.
88. Dudenbostel T, Acelajado MC, Pisoni R, Li P, Oparil S, Calhoun DA. Refractory Hypertension: Evidence of Heightened Sympathetic Activity as a Cause of Antihypertensive Treatment Failure. Hypertension. 2015;66:126–133.
89. Calhoun DA, Booth JN 3rd, Oparil S, Irvin MR, Shimbo D, Lackland DT, et al. Refractory hypertension: determination of prevalence, risk factors, and comorbidities in a large, population-based cohort. Hypertension. 2014;63:451–458.
90. Muxfeldt ES, Chedier B. Refractory hypertension: what do we know so far? J Hum Hypertens. 2020;35:181–183.
91. Carey RM, Calhoun DA, Bakris GL, Brook RD, Daugherty SL, Dennison-Himmelfarb CR, et al. Resistant Hypertension: Detection, Evaluation, and Management: A Scientific Statement From the American Heart Association. Hypertension. 2018;72:e53–e90.
92. Acelajado MC, Hughes ZH, Oparil S, Calhoun DA. Treatment of Resistant and Refractory Hypertension. Circ Res. 2019;124:1061–1070.
93. Bray GA. Life insurance and overweight. Obes Res. 1995;3:97–99.
94. Huang Z, Willett WC, Manson JE, Rosner B, Stampfer MJ, Speizer FE, et al. Body weight, weight change, and risk for hypertension in women. Ann Intern Med. 1998;128:81–88.
95. Hubert HB, Feinleib M, McNamara PM, Castelli WP. Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study. Circulation. 1983;67:968–977.
96. Mertens IL, Van Gaal LF. Overweight, obesity, and blood pressure: the effects of modest weight reduction. Obes Res. 2000;8:270–278.
97. Forman JP. Diet and Lifestyle Risk Factors Associated With Incident Hypertension in Women. JAMA. 2009;302:401–411.
98. Garrison RJ, Kannel WB, Stokes J 3rd, Castelli WP. Incidence and precursors of hypertension in young adults: the Framingham Offspring Study. Prev Med. 1987;16:235–251.
99. Stamler J. The INTERSALT Study: background, methods, findings, and implications. Am J Clin Nutr. 1997;65:626S-642S.
100. Takase H, Sugiura T, Kimura G, Ohte N, Dohi Y. Dietary Sodium Consumption Predicts Future Blood Pressure and Incident Hypertension in the Japanese Normotensive General Population. J Am Heart Assoc. 2015;4:e001959.
101. Kieneker LM, Gansevoort RT, Mukamal KJ, de Boer RA, Navis G, Bakker SJL, et al. Urinary potassium excretion and risk of developing hypertension: the prevention of renal and vascular end-stage disease study. Hypertension. 2014;64:769–776.
102. Tasnim S, Tang C, Musini VM, Wright JM. Effect of alcohol on blood pressure. Cochrane Database Syst Rev. 2020;7:CD012787.
103. Roerecke M, Tobe SW, Kaczorowski J, Bacon SL, Vafaei A, Hasan OSM, et al. Sex-Specific Associations Between Alcohol Consumption and Incidence of Hypertension: A Systematic Review and Meta-Analysis of Cohort Studies. J Am Heart Assoc. 2018;7:e008202.

104. Cornelissen VA, Smart NA. Exercise Training for Blood Pressure: A Systematic Review and Meta-analysis. J Am Heart Assoc. 2013;2:e004473.
105. Shere A, Eletta O, Goyal H. Circulating blood biomarkers in essential hypertension: a literature review. J Lab Precis Med. 2017;2:99–99.
106. Jayedi A, Rahimi K, Bautista LE, Nazarzadeh M, Zargar MS, Shab-Bidar S. Inflammation markers and risk of developing hypertension: a meta-analysis of cohort studies. Heart. 2019;105:686–692.
107. Grayson PC, Kim SY, LaValley M, Choi HK. Hyperuricemia and incident hypertension: a systematic review and meta-analysis. Arthritis Care Res. 2011;63:102–110.
108. Ren F, Li M, Xu H, Qin X, Teng Y. Urine albumin-to-creatinine ratio within the normal range and risk of hypertension in the general population: A meta-analysis. J Clin Hypertens. 2021;23:1284–1290.
109. Lindeman RD, Tobin JD, Shock NW. Association between blood pressure and the rate of decline in renal function with age. Kidney Int. 1984;26:861–868.
110. Wang Q, Xie D, Xu X, Qin X, Tang G, Wang B, et al. Blood pressure and renal function decline: a 7-year prospective cohort study in middle-aged rural Chinese men and women. J Hypertens. 2015;33:136–143.
111. Peralta CA, Jacobs DR, Katz R, Ix JH, Madero M, Duprez DA, et al. Association of Pulse Pressure, Arterial Elasticity, and Endothelial Function With Kidney Function Decline Among Adults With Estimated GFR > 60 mL/min/1.73 m2: The Multi-Ethnic Study of Atherosclerosis. Am J Kidney Dis Off J Natl Kidney Found. 2012;59:41–49.
112. Palatini P, Dorigatti F, Saladini F, Benetti E, Mos L, Mazzer A, et al. Factors associated with glomerular hyperfiltration in the early stage of hypertension. Am J Hypertens. 2012;25:1011–1016.
113. Kotchen TA, Piering AW, Cowley AW, Grim CE, Gaudet D, Hamet P, et al. Glomerular hyperfiltration in hypertensive African Americans. Hypertens Dallas Tex 1979. 2000;35:822–826.
114. Okada R, Yasuda Y, Tsushita K, Wakai K, Hamajima N, Matsuo S. Glomerular hyperfiltration in prediabetes and prehypertension. Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc. 2012;27:1821–1825.
115. Eriksen BO, Stefansson VTN, Jenssen TG, Mathisen UD, Schei J, Solbu MD, et al. Elevated blood pressure is not associated with accelerated glomerular filtration rate decline in the general non-diabetic middle-aged population. Kidney Int. 2016;90:404–410.
116. Ferrannini E, Cushman WC. Diabetes and hypertension: the bad companions. Lancet. 2012;380:601–610.
117. Tatsumi Y, Morimoto A, Asayama K, Sonoda N, Miyamatsu N, Ohno Y, et al. Fasting Blood Glucose Predicts Incidence of Hypertension Independent of HbA1c Levels and Insulin Resistance in Middle-Aged Japanese: The Saku Study. Am J Hypertens. 2019;32:1178–1185.
118. Cicero AFG, Rosticci M, Baronio C, Morbini M, Parini A, Grandi E, et al. Serum LDL cholesterol levels and new onset of arterial hypertension: an 8-year follow-up. Eur J Clin Invest. 2014;44:926–932.
119. Paynter NP, Sesso HD, Conen D, Otvos JD, Mora S. Lipoprotein subclass abnormalities and incident hypertension in initially healthy women. Clin Chem. 2011;57:1178–1187.
120. Otsuka T, Takada H, Nishiyama Y, Kodani E, Saiki Y, Kato K, et al. Dyslipidemia and the Risk of Developing Hypertension in a Working-Age Male Population. J Am Heart Assoc. 2016;5:e003053.
121. Palmu J, Tikkanen E, Havulinna AS, Vartiainen E, Lundqvist A, Ruuskanen MO, et al. Comprehensive biomarker profiling of hypertension in 36 985 Finnish individuals. J Hypertens. 2022;40:579–587.
122. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388:1659–1724.
123. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College

of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129:S49-73.
124. Pearson TA, LaCroix AZ, Mead LA, Liang K-Y. The Prediction of Midlife Coronary Heart Disease and Hypertension in Young Adults: The Johns Hopkins Multiple Risk Equations. Am J Prev Med. 1990;6:23–28.
125. Parikh NI, Pencina MJ, Wang TJ, Benjamin EJ, Lanier KJ, Levy D, et al. A risk score for predicting near-term incidence of hypertension: the Framingham Heart Study. Ann Intern Med. 2008;148:102–110.
126. Kanegae H, Oikawa T, Suzuki K, Okawara Y, Kario K. Developing and validating a new precise risk-prediction model for new-onset hypertension: The Jichi Genki hypertension prediction model (JG model). J Clin Hypertens. 2018;20:880–890.
127. Chien K-L, Hsu H-C, Su T-C, Chang W-T, Sung F-C, Chen M-F, et al. Prediction models for the risk of new-onset hypertension in ethnic Chinese in Taiwan. J Hum Hypertens. 2011;25:294–303.
128. Kivimäki M, Batty GD, Singh-Manoux A, Ferrie JE, Tabak AG, Jokela M, et al. Validating the Framingham Hypertension Risk Score: results from the Whitehall II study. Hypertension. 2009;54:496–501.
129. Guo Q, Lu X, Gao Y, Zhang J, Yan B, Su D, et al. Cluster analysis: a new approach for identification of underlying risk factors for coronary artery disease in essential hypertensive patients. Sci Rep. 2017;7:43965.
130. Yang D-Y, Nie Z-Q, Liao L-Z, Zhang S-Z, Zhou H-M, Sun X-T, et al. Phenomapping of subgroups in hypertensive patients using unsupervised data-driven cluster analysis: An exploratory study of the SPRINT trial. Eur J Prev Cardiol. 2019;26:1693–1706.
131. Luft FC. Twins in cardiovascular genetic research. Hypertension. 2001;37:350–356.
132. Niiranen TJ, McCabe EL, Larson MG, Henglin M, Lakdawala NK, Vasan RS, et al. Risk for hypertension crosses generations in the community: a multi-generational cohort study. Eur Heart J. 2017;38:2300–2308.
133. Enslow BT, Stockand JD, Berman JM. Liddle's syndrome mechanisms, diagnosis and management. Integr Blood Press Control. 2019;12:13–22.
134. Healy JK. Pseudohypoaldosteronism Type II. Hypertension. 2014;63:648–654.
135. Morineau G, Sulmont V, Salomon R, Fiquet-Kempf B, Jeunemaître X, Nicod J, et al. Apparent mineralocorticoid excess: report of six new cases and extensive personal experience. J Am Soc Nephrol. 2006;17:3176–3184.
136. El-Maouche D, Arlt W, Merke DP. Congenital adrenal hyperplasia. Lancet. 2017;390:2194–2210.
137. Simonetti GD, Mohaupt MG, Bianchetti MG. Monogenic forms of hypertension. Eur J Pediatr. 2012;171:1433–1439.
138. Bhandari S. The pathophysiological and molecular basis of Bartter's and Gitelman's syndromes. Postgrad Med J. 1999;75:391–396.
139. Kupper N, Ge D, Treiber FA, Snieder H. Emergence of novel genetic effects on blood pressure and hemodynamics in adolescence: the Georgia Cardiovascular Twin Study. Hypertension. 2006;47:948–954.
140. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet. 2018;50:1412–1425.
141. Giri A, Hellwege JN, Keaton JM, Park J, Qiu C, Warren HR, et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. Nat Genet. 2019;51:51–62.
142. Ma Y, Jiang H, Shah SJ, Arnett D, Irvin MR, Luo Y. Genetic-Based Hypertension Subtype Identification Using Informative SNPs. Genes. 2020;11:1265.
143. Luo Y, Mao C, Yang Y, Wang F, Ahmad FS, Arnett D, et al. Integrating hypertension phenotype and genotype with hybrid non-negative matrix factorization. Bioinformatics. 2018;35:1395–1403.
144. Williams RR, Rao DC, Ellison RC, Arnett DK, Heiss G, Oberman A, et al. NHLBI family blood pressure program: methodology and recruitment in the HyperGEN network. Hypertension genetic epidemiology network. Ann Epidemiol. 2000;10:389–400.
145. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019;10:1776.

146. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.
147. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, 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. 2015;12:e1001779.
148. Zhang Z, Murtagh F, Van Poucke S, Lin S, Lan P. Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R. Ann Transl Med. 2017;5:75.
149. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
150. Murtagh F, Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? J Classif. 2014;31:274–295.
151. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. 2021. cluster: Cluster Analysis Basics and Extensions. R package version 2.1.2.
152. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53–65.
153. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature. 1999;401:788–791.
154. Tan VYF, Févotte C. Automatic relevance determination in nonnegative matrix factorization with the β-divergence. IEEE Trans Pattern Anal Mach Intell. 2013;35:1592–1605.
155. Udler MS, Kim J, von Grotthuss M, Bonàs-Guarch S, Cole JB, Chiou J, et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. PLoS Med. 2018;15:e1002654.
156. Kim C, Westerman K, Smith K, Kim J, von Grotthuss M, Udler M (2020). Pipeline for GWAS clustering using Bayesian non-negative matrix factorization (bNMF). GitHub repository. https://github.com/gwas-partitioning/bnmf-clustering.
157. Harrell FE (2021). rms: Regression Modeling Strategies. R package version 6.2-0. https://CRAN.R-project.org/package=rms.
158. Therneau T (2021). A Package for Survival Analysis in R. R package version 3.2-11. https://CRAN.R-project.org/package=survival.
159. Alboukadel Kassambara, Marcin Kosinski and Przemyslaw Biecek (2021). survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.9. https://CRAN.R-project.org/package=survminer.
160. Schoenfeld D. Partial residuals for the proportional hazards regression model. Biometrika. 1982;69:239–241.
161. Schemper M. Cox Analysis of Survival Data with Non-Proportional Hazard Functions. The Statistician. 1992;41:455–465.
162. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA. 1982;247:2543–2546.
163. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008;27:157–72.
164. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ. 1995;310:170.
165. Alberti KGMM, Zimmet P, Shaw J, IDF Epidemiology Task Force Consensus Group. The metabolic syndrome--a new worldwide definition. Lancet Lond Engl. 2005;366:1059–1062.
166. Volpe M, Battistoni A, Savoia C, Tocci G. Understanding and treating hypertension in diabetic populations. Cardiovasc Diagn Ther. 2015;5:353–363.
167. Niiranen TJ, Lyass A, Larson MG, Hamburg NM, Benjamin EJ, Mitchell GF, et al. Prevalence, Correlates, and Prognosis of Healthy Vascular Aging in a Western Community-Dwelling Cohort: The Framingham Heart Study. Hypertension. 2017;70:267–274.
168. Tune JD, Goodwill AG, Sassoon DJ, Mather KJ. Cardiovascular consequences of metabolic syndrome. Transl Res. 2017;183:57–70.

169. Niiranen TJ, McCabe EL, Larson MG, Henglin M, Lakdawala NK, Vasan RS, et al. Heritability and risks associated with early onset hypertension: multigenerational, prospective analysis in the Framingham Heart Study. BMJ. 2017;357:j1949.
170. Suvila K, McCabe EL, Lehtonen A, Ebinger JE, Lima JAC, Cheng S, et al. Early Onset Hypertension Is Associated With Hypertensive End-Organ Damage Already by MidLife. Hypertension. 2019;74:305–312.
171. Miall WE, Oldham PD. The hereditary factor in arterial blood-pressure. Br Med J. 1963;1:75–80.
172. Dale BL, Madhur MS. Linking inflammation and hypertension via LNK/SH2B3. Curr Opin Nephrol Hypertens. 2016;25:87–93.
173. Elmehdawi R. Hypolipidemia: A Word of Caution. Libyan J Med. 2008;3:84–90.
174. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet. 2010;42:508–514.
175. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 2015;47:979–986.
176. Bourgeois B, Watts K, Thomas DM, Carmichael O, Hu FB, Heo M, et al. Associations between height and blood pressure in the United States population. Medicine (Baltimore). 2017;96:e9233.
177. Anowar F, Sadaoui S, Selim B. Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). Comput Sci Rev. 2021;40:100378.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2022

ISBN 978-951-29-8956-0 (PRINT)
ISBN 978-951-29-8957-7 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)