Aalto University
School of Science
Master's Programme in Life Science Technologies

Genetic Risk Discovery in Raynaud's Syndrome through Computational Genomics

Master's Thesis

September 18, 2025

Erika Romppanen

Aalto-yliopisto
Perustieteiden korkeakoulu
Master's Programme in Life Science Technologies

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä:	Erika Romppanen
Työn nimi:	Genetic Risk Discovery in Raynaud's Syndrome through Computational Genomics
Päiväys:	18. syyskuuta 2025
Sivumäärä:	56
Pääaine:	Biomedical Engineering
Koodi:	SCI3059
Valvoja:	Prof. Matias Palva
Ohjaajat:	PhD Hanna M. Ollila, PhD Martin Broberg
Yhteistyötaho:	Suomen Molekyylilääketieteen Instituutti (FIMM)

Raynaud'n tauti on verisuonisairaus, jolle on ominaista kohtauksittainen sormien ja varpaiden pienten verisuonien liiallinen supistuminen, jonka useimmiten laukaisee kylmäaltistus tai emotionaalinen stressi. Vaikka taudin ympäristölliset ja kliiniset riskitekijät tunnetaan, sen geneettinen perusta ei ole täysin selvillä.

Tämän diplomityön tavoitteena oli tunnistaa Raynaud'n taudin geneettisiä ja biologisia riskitekijöitä. Työssä hyödynnettiin uusinta FinnGen R13 -aineistoa genominlaajuisen assosiaatiotutkimuksen (GWAS) avulla, ja tulokset yhdistettiin laajaan meta-analyysiin seitsemän muun tutkimusaineiston kanssa. Jatkoanalyyseinä suoritettiin kausaalien varianttien hienokartoitus (fine-mapping) ja geenien ekspressioon liittyvät kvantitatiiviset lokusanalyysit (eQTL) merkittävien assosiaatioiden tarkentamiseksi ja toiminnallisten merkitysten ymmärtämiseksi.

Meta-analyysi, johon sisältyi 17 793 tapausta ja 1 850 553 verrokkia, on tähän mennessä suurin Raynaud'n tautia koskeva meta-analyysi. Se tunnisti 14 genominlaajuisesti merkittävää lokusta, joista yhdeksän oli uusia löydöksiä. Vahvin assosiaatio havaittiin lähellä *ADRA2A*-geeniä, vahvistaen aiempia tutkimustuloksia adrenergisen signaalijärjestelmän merkityksestä taudissa. Muita merkittäviä lokuksia olivat *IRX1*, *C6orf15/HLA*, *NOS3*, *ADRA1B* ja *SH2B3*, viitaten mahdollisiin yhteyksiin vaskulaarisen tonuksen säätelyssä, immuunimekanismeissa ja kudosten kehityksessä.

Miesten ja naisten väliset analyysit paljastivat eroavaisuuksia *ADRA2A*-geenin johtovarienttien osalta. Hienokartoitus osoitti pienen uskottavan joukon (credible sets) mahdollisia kausaaleja variantteja *ADRA2A*-geenille, kun taas *IRX1*-geenin laajempi uskottava joukko kuvasti suurempaa epävarmuutta kausaalia varianttia kohtaan. eQTL-analyysit yhdistivät useita johtovariantteja geenien ekspressiomuutoksiin säärivaltimossa ja luustolihaksessa, jotka ovat olennaisia kudoksia Raynaud'n taudissa.

Avainsanat:	raynaud'n tauti, gwas, FinnGen, meta-analyysi, fine-mapping, eQTL
Kieli:	Englanti

Aalto University
School of Science
Master's Programme in Life Science Technologies

ABSTRACT OF
MASTER'S THESIS

Author:	Erika Romppanen
Title of thesis:	Genetic Risk Discovery in Raynaud's Syndrome through Computational Genomics
Date:	September 18, 2025
Pages:	56
Major:	Biomedical Engineering
Code:	SCI3059
Supervisor:	Professor Matias Palva
Advisors:	PhD Hanna M. Ollila, PhD Martin Broberg
Collaborative partner:	Institute for Molecular Medicine Finland (FIMM)

Raynaud's syndrome (RS) is a vascular disorder characterized by episodic vasoconstriction, most often affecting the fingers and toes, typically triggered by cold or emotional stress. While environmental and clinical risk factors are known, the genetic basis remains incompletely understood.

This thesis aimed to identify genetic and biological risk factors for RS by performing a genome-wide association study (GWAS) using the latest FinnGen release 13 data, followed by a large-scale meta-analysis across eight cohorts. Downstream analyses, such as fine-mapping and expression quantitative trait locus (eQTL) analyses, were conducted to refine and functionally annotate significant associations.

The meta-analysis, comprising 17,793 cases and 1,850,553 controls, represents the largest meta-analysis of RS conducted to date. It identified 14 genome-wide significant loci, nine of which were novel. The strongest association was observed near *ADRA2A*, reinforcing prior evidence for the role of adrenergic signaling in RS pathophysiology. Additional notable loci included *IRX1*, *C6orf15/HLA*, *NOS3*, *ADRA1B*, and *SH2B3*, suggesting potential roles in vascular tone regulation, immune mechanisms, and tissue development.

Sex-specific analyses revealed distinct lead variants near *ADRA2A* in males and females. Fine-mapping pointed small credible sets at *ADRA2A*, while broader credible sets at *IRX1* indicated greater uncertainty. eQTL analyses linked several lead variants to gene expression changes in tibial artery and skeletal muscle, tissues relevant to RS pathophysiology.

Keywords:	raynaud's syndrome, gwas, FinnGen, meta-analysis, fine-mapping, eQTL
Language:	English

Contents

Abbreviations

A	Adenine
AAF	Alternative Allele Frequency
ALT	Alternative Allele
C	Cytosine
dbSNP	Single-Nucleotide Polymorphism Database
DNA	Deoxyribonucleic Acid
eQTL	Expression Quantitative Trait Loci
FDR	False Discovery Rate
G	Guanine
GSA	Global Screening Array
GWAS	Genome-Wide Association Study
HWE	Hardy-Weinberg Equilibrium
ICD	International Classification of Diseases
LD	Linkage Disequilibrium
MAC	Minor Allele Count
MAF	Minor Allele Frequency
NES	Normalized Effect Size
PIP	Posterior Inclusion Probability
PRS	Polygenic Risk Score
QC	Quality Control
REF	Reference Allele
RS	Raynaud's syndrome
rsID	Reference SNP ID
SNP	Single-Nucleotide Polymorphism
SNV	Single-Nucleotide Variant
T	Thymine
UTR	Untranslated Region
VCF	Variant Call Format

Acknowledgements

I would like to thank my advisor, Hanna Ollila, for giving me the opportunity to pursue this master's thesis and for providing a position in her group that made this work possible.

I am also grateful to my co-advisor, Martin Broberg, for patiently guiding me through the computational aspects of the project and for fostering a low-threshold environment where asking questions was always encouraged and welcomed.

I would like to thank the entire Ollila group for their warm welcome. Working alongside such knowledgeable and driven individuals has been truly inspiring. Your expertise, ambition, and collaborative spirit made this experience remarkably rewarding.

I also wish to thank my Aalto University supervisor, Matias Palva, for his valuable feedback on the thesis.

I gratefully acknowledge the computational resources provided by the Institute for Molecular Medicine Finland (FIMM) Technology Centre, which were essential for completing this work.

Finally, I wish to thank all participants and investigators of the FinnGen study, as well as the contributing biobanks, for granting access to the data used in this thesis.

1 Background

1.1 Raynaud's syndrome

Raynaud's syndrome (RS) is a multifactorial vasospastic disorder that affects the peripheral circulation. It is characterized by recurrent, transient, and reversible episodes of vasoconstriction, typically affecting the fingers and toes [1; 2]. These episodes reduce blood flow and often appear as a three-phase discoloration: first pallor (white) due to ischemia, then cyanosis (blue) from deoxygenation, and finally erythema (red) when blood flow returns. These attacks are usually triggered by exposure to cold or emotional stress and are often painful [2].

RS affects an estimated 3–5% of the general population, though prevalence varies depending on region, gender, and diagnostic criteria [1]. It is more common in females and in colder climates. Furthermore, studies have shown that RS tends to run in families, suggesting a hereditary component to disease susceptibility [1; 2].

RS is commonly classified as either primary or secondary. Primary RS is idiopathic, meaning there is no identifiable underlying cause. Secondary RS, on the other hand, occurs as a comorbidity, especially of autoimmune or connective tissue diseases, such as systemic sclerosis (SSc), systemic lupus erythematosus (SLE), myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), and, more recently, long COVID [3; 4; 5; 6]. RS has a well-defined, international diagnosis code, ICD-10 I73.0 and ICD-9 443.0, which facilitates the identification of cases in electronic health records and biobank data.

The pathophysiology of RS involves both neural and vascular mechanisms. Cold or stress triggers the sympathetic nervous system, leading to the release of norepinephrine and activation of α-adrenergic receptors on vascular smooth muscle, which induces vasoconstriction [2]. Despite frequent attacks, tissue damage is rare in most RS patients [1].

Although RS is generally considered a benign condition, it can affect the quality of life [1]. Treatments include avoiding cold exposure and managing stress, but more severe cases may require medication such as calsium channel blockers, α_1-adrenoreceptor antagonists, or nitric oxide. However, no gold-standard treatment exists, partly due to the complex pathophysiology and the fact that the etiology of the disease is not fully understood [2].

1.2 Genome-wide association study

Genome-wide association study (GWAS) is a powerful computational approach to identify genetic variants associated with complex traits and diseases. These traits, such as RS, often involve the combined effects of multiple genetic and environmental factors [7].

GWAS compares allele frequencies between individuals who share similar ancestry but differ in the trait of interest. The most commonly studied variants are single-nucleotide polymorphisms (SNPs), which are analyzed across the genome to find regions i.e., genomic risk loci, where clusters of variants show statistically significant associations with the phenotype [8].

SNPs are the most common type of variation in the human genome. They represent sequence differences where a single nucleotide (A, T, G, or C) varies between individuals (Figure 1). A common SNP occurs approximately every 400 base pairs. Although most SNPs have no functional effect, some can influence gene expression or protein function, thereby affecting disease risk or other traits [9].

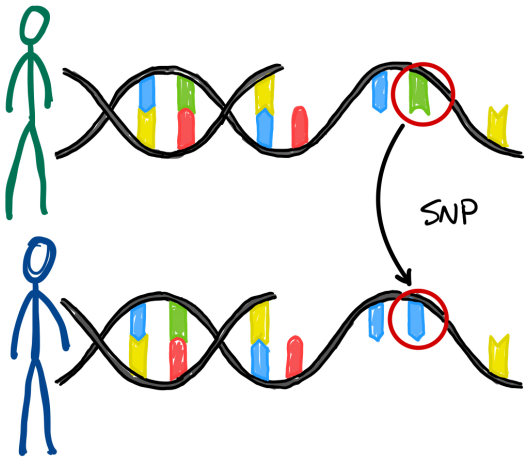


Figure 1: Illustrated concept of a single-nucleotide polymorphism (SNP), showing how a single nucleotide base can vary between individuals at a specific genomic position.

Findings from GWAS have a wide range of applications. For example, genetic variants linked to specific traits can be used to control for genetic differences in epidemiological studies or estimate an individual's risk of developing certain physical or mental conditions based on their genetic profile [8]. Recent advances in cohort-based research have enabled GWAS to leverage large sample sizes from large-scale biobanks, significantly increasing statistical power and enabling a more comprehensive and unbiased discovery of genetic associations [10]. This represents a major improvement over earlier candidate gene studies, which were limited by small sample sizes and often produced underpowered and non-replicable findings [11].

However, interpreting GWAS results can be challenging. Associated variants are often found in regions of high linkage disequilibrium (LD), meaning that they are correlated with nearby genetic variants both causal and non-causal. As a result, the identified variant may not itself be causal, but instead may be inherited together with the true causal variant due to their close proximity on the genome [9; 12]. Moreover, many associated

variants are located in non-coding, intergenic regions of the genome, complicating the interpretation of their biological function [13].

To address these issues, GWAS findings are often followed by downstream analyses such as meta-analysis and fine-mapping. These approaches aim to enhance statistical power and refine association signals by prioritizing variants with the highest probability of being causal.

1.3 Post-GWAS analyses

1.3.1 GWAS meta-analysis

While GWASs have identified many loci associated with complex traits and diseases, the individual effect sizes of common variants are typically small and thus, large sample sizes are often required to detect these associations. Meta-analysis has become a standard method to address this limitation by statistically combining GWAS summary statistics from multiple independent studies, thereby increasing the sample size and power to detect true associations. In recent years, the majority of genetic risk loci have been discovered through large-scale meta-analyses of GWAS data [14].

GWAS meta-analysis enables the identification of more robust and generalizable association signals by integrating data from diverse populations and study designs. It also enables the evaluation of how consistently genetic effects are observed across studies and helps to reduce the influence of random variation or study-specific biases [14]. Different meta-analysis models exist, including effect size-based and sample size-weighted approaches.

The former approach, also known as fixed effects meta-analysis, is the most commonly used and statistically efficient when effect size estimates and their standard errors (SE) are known [14; 15]. It applies inverse-variance weighting, where each study is weighted by the inverse of its variance ($1/\text{SE}^2$) so that the studies with more precise estimates contribute more to the final results of meta-analysis. The combined effect estimate, $\hat{\beta}_F$, is calculated as:

$$\hat{\beta}_F = \frac{\sum w_i \hat{\beta}_i}{\sum_i w_i}, \qquad \text{where} \quad w_i = \frac{1}{\text{SE}^2} = \frac{1}{\text{Var}(\hat{\beta}_i)}. \tag{1}$$

Here, the $\hat{\beta}_i$ and SE_i are the effect size estimate and standard error for the i^{th} study, and the variance of the combined effect estimate and standard error are calculated as:

$$\text{Var}(\hat{\beta}_F) = \frac{1}{\sum_i w_i} \qquad \text{and} \qquad \text{SE}(\hat{\beta}_F) = \sqrt{\text{Var}(\hat{\beta}_F)}. \tag{2}$$

The latter approach, also known as a Z-score meta-analysis, combines p-values and the

direction of effects to generate Z-scores. Each Z-score is then weighted by the square root of the sample size ($\sqrt{N}$) ensuring that studies with larger sample sizes contribute more to the result [15]. The combined Z-score, Z, is calculated as:

$$Z = \frac{\sum_i Z_i w_i}{\sqrt{\sum_i w_i^2}}, \qquad \text{where} \quad w_i = \sqrt{N_i}. \tag{3}$$

Here, Z_i is the Z-score for the i^{th} study, calculated as:

$$Z_i = \Phi^{-1}(1 - \frac{p_i}{2}) \times (\text{effect direction for } i^{\text{th}} \text{ study}), \tag{4}$$

where Φ is the standard normal cumulative distribution function, and p_i the p-value for the i^{th} study.

An important consideration in GWAS meta-analysis is heterogeneity, which refers to variation in effect sizes between studies. Such variation can arise from differences in ancestry, phenotype definitions, environmental exposures, and analytical methods such as genotyping platforms and imputation software [14]. To evaluate heterogeneity, two common metrics are used: Cochran's Q and the I^2 statistic. Cochran's Q test, also referred to as the chi-squared (χ^2) test for heterogeneity, evaluates whether the variability in effect sizes is greater than what would be expected by chance. The test produces an associated heterogeneity p-value, which indicates the statistical significance of the observed heterogeneity, though a more lenient significance threshold (e.g., 0.1 instead of 0.05) can be applied [16; 17]. The I^2 statistic measures how much of the variation in effect estimates is caused by true differences between studies, rather than random chance. I^2 values range from 0–100%, with 0–25% indicating low, 25–50% moderate, 50–75% high, and 75–100% very high heterogeneity [14; 17].

1.3.2 QQ plot

A quantile-quantile (QQ) plot is used to assess whether the distribution of p-values from the GWAS meta-analysis deviates from what is expected under the null hypothesis of no association between genetic variants and the trait. Under the null hypothesis, p-values are expected to follow a uniform distribution between 0 and 1. The QQ plot compares the observed p-values with the expected p-values from this uniform distribution on a $-\log_{10}$ scale. Deviations from the diagonal line may indicate the presence of true associations or potential biases in the data [18].

To further quantify the degree of deviation from the null hypothesis, the genomic control lambda (λ) is calculated. It is defined as the ratio of the median of the observed chi-squared test statistics to the expected median under the null hypothesis:

$$\lambda = \frac{\text{median}(\chi^2_{observed})}{\text{median}(\chi^2_{expected})} \approx \frac{\text{median}(\chi^2_{observed})}{0.455}. \tag{5}$$

Values of $\lambda > 1.1$ suggest inflation, which can result from factors such as population stratification. Conversely, $\lambda < 0.9$ indicates deflation, which may occur when the genetic markers are not truly independent [18].

1.3.3 Fine-mapping

While GWAS and meta-analysis can identify genomic risk loci, the associated variants may not be the true causal variants but may instead correlate with them due to LD. As a result, one variant showing a strong association with a trait may be tagging the actual causal variant located nearby. Fine-mapping refers to a set of statistical methods designed to refine these association signals by prioritizing variants that are most likely to have causal effect on the phenotype [12]. These methods leverage GWAS summary statistics and LD information from a reference panel to estimate the probability of each variant being causal.

A key outcome of fine-mapping is the posterior inclusion probability (PIP), representing the probability that a particular variant is causal, given the observed data and model assumptions. Formally, this is defined as:

$$\text{PIP}_j = P(\beta_j \neq 0 | \mathbf{y}, \mathbf{X}), \tag{6}$$

where β_j is the effect size of variant j, $\mathbf{y}$ denotes the phenotype data, and $\mathbf{X}$ the genotype data [19].

In addition to PIPs, the fine-mapping methods also generate credible sets collections of variants that together have a high probability (e.g., 95%) of containing the true causal variant [12].

1.3.4 eQTL

Expression quantitative trait loci (eQTLs) are genomic regions where genetic variants are associated with variation in gene expression levels. In eQTL mapping, gene expression is treated as a quantitative trait, enabling the identification of regulatory loci and the investigation of how polymorphisms influence transcription within these regions [20].

eQTL analysis provides a way to link disease-associated variants to their molecular effects on gene regulation. This connection can help explaining how genetic variation contributes to disease risk. Moreover, eQTLs can reveal regulatory networks that are differentially modulated between healthy and diseased individuals [20].

Large-scale resources such as the Genotype-Tissue Expression (GTEx) project have enabled the systematic identification of eQTLs across a wide range of human tissues, thereby improving our understanding of the regulatory mechanisms underlying complex traits and diseases [21].

1.3.5 Variant annotation

FUMA GWAS

Functional mapping and annotation of genome-wide association studies (FUMA GWAS) is a web-based platform designed to annotate, visualize, and interpret GWAS summary statistics [22]. It integrates multiple genomic data sources to provide a variety of annotations, including predicted functional consequences, regulatory potential, eQTL associations, and chromatin interaction data for each variant.

FUMA analyzes GWAS summary statistics through the SNP2GENE pipeline, which identifies independent genome-wide significant variants based on LD structure and maps them to potential target genes. Functional annotation is performed using metrics such as Combined Annotation-Dependent Depletion (CADD) and RegulomeDB scores.

CADD scores quantify the deleteriousness of single-nucleotide variants (SNVs) and indels. A higher CADD score indicates a greater likelihood that a variant has a harmful effect on gene function or regulation and may contribute to disease [23]. RegulomeDB scores assess the regulatory potential of non-coding variants by integrating a wide range of functional genomic data. A lower RegulomeDB score suggests stronger evidence that the variant is located in a functionally relevant genomic region [24].

FUMA also incorporates multi-marker analysis of genomic annotation (MAGMA) for gene and gene-set analyses. MAGMA aggregates SNP p-values to the gene-level, and evaluates whether predefined biological pathways or gene sets are enriched among the associated genes [25].

LocusZoom

LocusZoom is a visualization tool for displaying GWAS results in a genomic context. It provides interactive plots highlighting association signals, gene annotations, and LD structure across genomic regions of interest [26].

The platform generates regional association plots by plotting GWAS p-values, displayed as $-\log_{10}(\mathrm{p})$, across a 500 kilobase window centered on a variant of interest. These plots integrate information on LD, nearby gene locations, and recombination rates (cM/Mb) based on the selected 1000 Genomes (1000G) reference panel, facilitating interpretation

of the association signal in its genomic location.

LD is typically measured using the squared correlation coefficient r^2, quantifying how often alleles at two loci are inherited together. For example, considering two SNPs: one with alleles A and a (with frequencies p_A and p_a), and another with alleles B and b (with frequencies p_B and p_b). The LD coefficient D is defined as the difference between the observed frequency of the haplotype AB (p_{AB}) and the expected frequency under independence ($p_A p_B$). The r^2 statistic is then calculated as:

$$r^2 = \frac{D^2}{p_A p_a p_B p_b}, \qquad \text{where} \quad D = p_{AB} - p_A p_B. \tag{7}$$

An r^2 value close to 1 indicates strong LD, meaning the alleles are often inherited together, whereas values near 0 suggest independent inheritance [9].

In addition, LocusZoom provides PIPs for variants using the 1000G LD structure as reference.

topr and VannoPortal

`topr` is an R package designed for visualizing, annotating, and exploring GWAS results, particularly for identifying lead variants and generating region-specific and genome-wide plots [27].

Additional functional annotation of lead variants was conducted using VannoPortal (http://www.mulinlab.org/vportal), a variant annotation database that integrates information from 40 genome-wide resources [28]. The information from these resources are organized into five categories including variant basic information, evolutionary annotation, disease or trait association, variant regulatory potential, and variant pathogenicity.

1.4 Current study and objectives

In recent years, GWASs have identified genomic regions associated with RS susceptibility. A single-cohort GWAS conducted in the UK Biobank (UKB) identified significant associations near the *ADRA2A*, *IRX1*, and *MICB* loci, pointing to potential roles in adrenergic signaling and immune regulation [29]. More recently, a large-scale meta-analysis combining four cohorts FinnGen data release 10, UKB, Estonian Biobank (EstBB), and the Mass General Brigham (MGB) Biobank analyzed data from 11,358 cases and over 1.1 million controls [3]. This study by Tervi et al. (2024) confirmed associations near *ADRA2A* and *IRX1*, and identified additional loci, such as

C6orf15/HLA and *NOS3*, reinforcing the robustness and replicability of these findings across diverse populations.

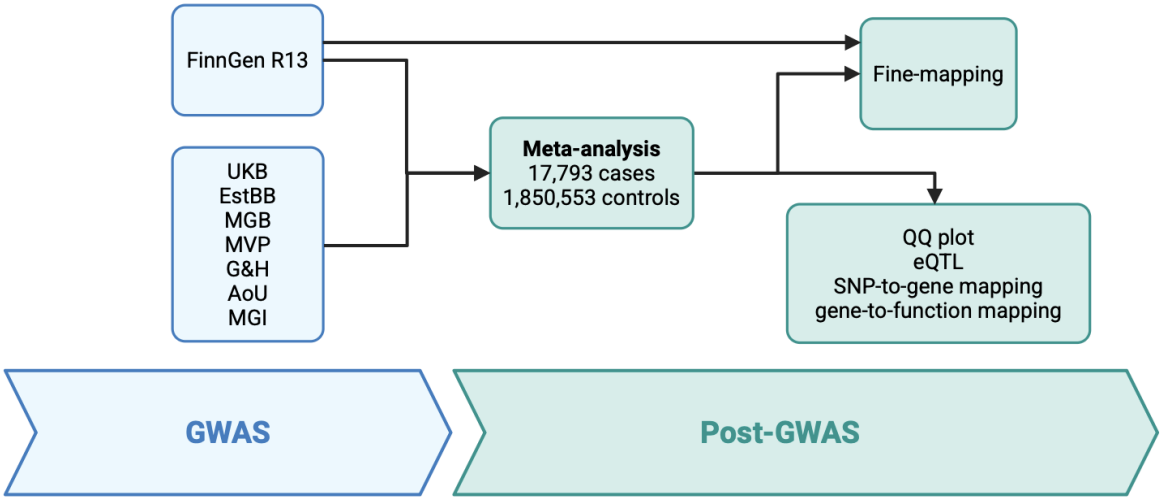


Figure 2: Overview of the study workflow. Genome-wide association study (GWAS) data from FinnGen release 13 and seven additional cohorts were combined through meta-analysis. Post-GWAS downstream analyses included fine-mapping, quantile-quantile (QQ) plotting, expression quantitative trait loci (eQTL) mapping, SNP-to-gene mapping, and gene-to-function mapping to interpret the biological relevance of identified loci.

While these studies have provided important insights into the genetic architecture of RS, the number of large-scale analyses remains limited. The present study leverages genome-wide data from eight cohorts, comprising more than 1.8 million individuals, including the most recent FinnGen data release 13. This represents the largest GWAS meta-analysis of RS performed to date. The primary objective of this thesis was to investigate genetic factors contributing to RS susceptibility using large-scale genomic data, to strengthen and extend previous findings, and to provide more robust evidence for associated loci. In addition to the meta-analysis, post-GWAS analyses, including statistical fine-mapping and eQTL analyses, were employed to refine association signals and explore potential functional mechanisms (Figure 2).

1.5 Translational applications in biomedical engineering

Advances in genomic approaches and methods, such as GWAS and post-GWAS analyses, have opened new opportunities for biomedical engineering by providing insights that can be translated into strategies for disease prevention, diagnosis, and treatment. These approaches are increasingly integrated with computational methods, biosensors, and personalized health technologies, creating applications that connect genetic discoveries with engineering solutions for healthcare.

In personalized and precision medicine, genomic findings are used to tailor prevention strategies or therapies based on an individual's genetic profile. Polygenic risk scores (PRSs) derived from GWAS summary statistics provide estimates of an individual's genetic susceptibility to diseases, including cardiovascular disorders, diabetes, and autoimmune diseases [30; 31]. Such risk estimates can be incorporated into clinical decision support tools to guide earlier interventions.

In drug discovery, genetic insights from GWAS inform the identification of novel therapeutic targets and help predict patient-specific drug responses. For example, loci associated with vascular tone or immune function, such as those implicated in RS, may pinpoint mechanisms that could be targeted to reduce abnormal vasoconstriction [3].

Digital health and wearable technologies represent another key application, where genetic risk data can be combined with multimodal physiological signals (e.g., heart rate, skin temperature, or blood oxygenation) to improve monitoring and prediction of disease-related events [32]. In vasospastic disorders, wearable devices could integrate genetic susceptibility with continuous biometric data to provide personalized feedback, detect episodes early, and support adaptive interventions.

2 Materials and Methods

This section outlines the methodological framework used to identify genetic variants associated with RS, integrating data from multiple large-scale genomic studies. The analysis consists of five main components: study cohorts, GWAS, fine-mapping, GWAS meta-analysis, post-GWAS visualization and annotation, and eQTL analysis.

The GWAS analysis for the FinnGen R13 dataset was conducted as part of this thesis, while GWAS summary statistics for the remaining cohorts were obtained from their respective research initiatives or public repositories. To ensure harmonization across datasets, extensive pre-processing and harmonization steps were implemented prior to meta-analysis. Each of these steps is detailed in the subsections below.

2.1 Study cohorts

This thesis included genotype and phenotype data from eight large-scale cohorts representing both population-based and hospital-based settings. These cohorts were FinnGen R13, UKB, EstBB, MGB, the Million Veteran Program (MVP), Genes & Health (G&H), the All of Us (AoU) Research Program, and the Michigan Genomics Initiative (MGI). RS cases were identified using harmonized definitions across cohorts based on ICD-10 code I73.0, ICD-9 code 443.0, and Phecodes corresponding to RS (e.g., 443.1 and X448.1). Sample sizes per cohort ranged from tens of thousands to over 500,000 participants.

Participants were primarily of European ancestry, with the exception of the G&H cohort, which consists of individuals of British South Asian ancestry. Each cohort applied standard quality control (QC) procedures, including filtering based on minor allele frequency (MAF), and imputation quality metrics.

Genotyping was performed using Illumina or Thermo Fisher arrays specific to each cohort, with genotype imputation conducted using appropriate reference panels. All datasets were aligned to the GRCh38/hg38 reference genome (Genome Reference Consortium Human Build 38), except for the UKB, which was aligned to GRCh37/hg19.

The cohorts represent a mix of general population biobanks and hospital-based registeries. Detailed cohort-level descriptions, including ancestry composition, genotyping and imputation strategies, case-control definitions, and QC thresholds, are provided below. An overview of the number of RS cases, controls, and SNPs included in the GWAS for each cohort is provided in Table 1.

Overview of cohorts included in the study			
Cohort	**RS Cases**	**Controls**	**SNPs**
FinnGen R13	3,278	497,070	19,521,613
UK Biobank	5,152	439,974	15,021,198
Estonian Biobank	2,205	203,345	29,327,699
Mass General Brigham Biobank	1,907	54,452	18,258,115
Million Veteran Program	2,156	447,847	18,763,336
Genes & Health	665	43,109	35,214,096
All of Us Research Program	901	98,010	40,441,507
Michigan Genomics Initiative	1,529	66,746	20,128,245
Total	17,793	1,850,553	

Table 1: Summary of cohorts included in this study, showing the number of Raynauds syndrome (RS) cases and controls, as well as the number of single-nucleotide polymorphisms (SNPs) in the genome-wide association study (GWAS) summary statistics after quality control.

FinnGen

FinnGen is a large-scale public-private partnership project combining genetic data from Finnish biobank participants with longitudinal health registry data, with the aim of gaining deeper understanding of the genetic factors underlying diseases [33]. The project is hosted by the Institute for Molecular Medicine Finland (FIMM) at the University of Helsinki. With over 500,000 participants nearly 10% of Finlands population FinnGen represents a nationally significant resource (https://www.finngen.fi/en). All participants are of Finnish ancestry, a genetically distinct subpopulation of European ancestry due to historical bottlenecks, with a broad age range and both sexes represented.

The FinnGen R13 consists of 500,348 Finnish participants collected from national biobanks. The dataset includes 3,278 individuals diagnosed with RS and 497,070 controls. RS cases were identified using the ICD-10 code I73.0 using nationwide health registry data, including primary care, inpatient, and outpatient hospital records.

Genotyping was performed using a customized ThermoFisher Axiom array (Thermo Fisher Scientific, Santa Clara, CA, USA) and various Illumina GWAS arrays (Illumina Inc., San Diego, CA, USA), and data were lifted over to GRCh38 reference genome build [34; 33]. Genotype imputation was conducted using the Finnish-specific SiSu (Sequence Initiative Suomi) v4.2 reference panel[1].

QC was applied both sample-wise and variant-wise. Sample-wise, individuals with

[1]http://sisuproject.fi

ambiguous sex, genotype missingness greater than 5%, excess heterozygosity (±4 standard deviations), and non-Finnish ancestry were excluded. Variant-wise, variants with genotype missingness greater than 2%, Hardy-Weinberg equilibrium (HWE) p-value less than 10^{-6}, and minor allele count (MAC) less than 3 were excluded [34].

Post-imputation QC was conducted by excluding variants with an imputation INFO score, a measure of imputation quality, below 0.6 or variants with MAF less than 0.0001.

Details on the GWAS performed on FinnGen R13 data are described in Section 2.2.

UK Biobank

The UK Biobank (UKB) is a large-scale population-based cohort study consisting of approximately half a million participants across the United Kingdom. Participants were aged between 40 and 69 years at recruitment between 2006 and 2010, and are primarily of European ancestry. This age group is studied because it is at elevated risk of developing a wide range of diseases (https://www.ukbiobank.ac.uk).

The cohort used in this study includes 445,126 participants, including 5,152 RS cases and 439,974 controls. RS cases were defined using the ICD-10 code I73.0 and ICD-9 code 443.0. The cohort study integrates detailed phenotypic data, health records, and genomic information for long-term follow-up [35].

Genotyping was performed using the UK BiLEVE Axiom and the UK Biobank Axiom arrays (Thermo Fisher Scientific, Santa Clara, CA, USA). The genotypes were aligned to GRCh37 reference genome and genotype imputation was performed using Haplotype Reference Consortium (HRC) as the main reference panel[2], merged with the UK10K and 1000G phase 3 reference panels[3,4] [35].

Post-imputation QC was conducted by excluding variants with INFO <0.3 or MAF <0.001.

Estonian Biobank

The Estonian biobank (EstBB) is a population-based biobank with a current cohort size of more than 200,000 participants maintained by the Estonian Genome Center at the University of Tartu. The cohort consist of adult residents of Estonia aged 18 and older and analysis was restricted to participants of European ancestry [36]. With approximately 20% of Estonias adult population enrolled, the EstBB serves as a highly representative national resource contributing to the development of medical research both in Estonia

[2]https://www.sanger.ac.uk/collaboration/haplotype-reference-consortium/
[3]https://www.uk10k.org
[4]https://www.internationalgenome.org/category/phase-3/

and globally (https://genomics.ut.ee/en).

The cohort used in this study consist of 212,955 participants from the 2023v1 data freeze. RS cases were identified using the ICD-10 code I73.0 and the cohort included 2,205 RS cases and 203,345 controls.

Genotyping was performed using a customized Illumina GSA (Illumina Inc., San Diego, CA, USA) and genotype imputation was conducted using an Estonian-specific reference panel, aligned to the GRCh38 reference genome [37].

Post-imputation QC excluded variants with INFO <0.4 or MAC <2.

Mass General Brigham Biobank

The Mass General Brigham (MGB) Biobank (formerly Partners HealthCare Biobank) is a hospital-based cohort study including genetic and clinical data from patients within the MGB healthcare network in Boston, Massachusetts, USA. Over 160,000 patients are now participating in the biobank. The study is designed to help researchers understand how peoples health is influenced by their genes, biomarkers, and other factors (https://www.massgeneralbrigham.org/en).

The dataset used in this study includes 54,452 participants, of which 1,907 were identified as RS cases and 52,545 as controls. RS cases were identified using the ICD-10 code I73.0 and ICD-9 code 443.0.

Genotyping was performed using Illumina GSA and Multi-Ethnic Global Array (MEG; Illumina Inc., San Diego, CA, USA). Genotype data were imputed to the Trans-Omics for Precision Medicine (TOPMed) r2 reference panel[5] and aligned to the GRCh38 reference genome. The analysis was restricted to individuals of European ancestry [3; 38].

Post-imputation QC was applied by excluding variants with INFO <0.4 or MAF <0.001.

Million Veteran Program

The Million Veteran Program (MVP) is a nationwide research initiative by the U.S. Department Veterans Affairs aimed at investigating how genetics, lifestyle, military experiences, and environmental exposures influence health and wellness among veterans. Since 2011, one million veterans have joined MVP. Notably, over 25% of enrollees are from minority racial backgrounds, enhancing the diversity and representativeness of the cohort (https://www.research.va.gov/MVP).

The cohort used in this study consists of 450,003 participants, including 2,156 RS cases and 447,847 controls. RS cases were defined using the Phecode 443.1, which corresponds

[5]https://imputation.biodatacatalyst.nhlbi.nih.gov/

to RS and is derived from ICD-9 and ICD-10 codes. Participants included in the analysis were restricted to individuals of European ancestry.

Genotyping was performed using customized Axiom MVP 1.0 Array (Thermo Fisher Scientific, Santa Clara, CA, USA). Genotype imputation was conducted using a hybrid reference panel combining the African Genome Resources (AGR) and the 1000G phase 3 panels, and aligned to the GRCh38 reference genome [39].

Post-imputation QC excluded variants with INFO <0.3, MAC <20, or if their allele frequency deviated more than 10% from the expected frequency based on the 1000G reference data [39].

Genes & Health

Genes & Health (G&H) is a community-based cohort study focusing on individuals of British Pakistani and British Bangladeshi heritage, collectively representing British South Asian Ancestry [40]. The aim of G&H is to enhance health outcomes in these communities by conducting research on health and genetic data from up to 100,000 volunteers (https://www.genesandhealth.org/en).

The cohort used in this study includes 43,774 participants, of which 665 were identified as RS cases and 43,109 as controls using the ICD-10 code I73.0.

Genotyping was performed using Illumina Infinium GSA v3.0 (Illumina Inc., San Diego, CA, USA) and genotypes were imputed using the TOPMed r3 reference panel, aligned to the GRCh38 reference genome [40; 41].

Post-imputation QC was applied to exclude poorly imputed or low-quality variants. Those with INFO <0.3, MAF <0.00001, or a missingness rate greater than 0.2 were removed.

All of Us Research Program

The All of Us (AoU) Research Program is a longitudinal, population-based cohort study initiated by the U.S. National Institutes of Health, aiming to enroll over one million individuals living in the United States. The goal is to advance precision medicine by reflecting the countrys rich diversity and by focusing on groups historically underrepresented in biomedical research (https://allofus.nih.gov).

The cohort used in this study includes 98,911 participants of European ancestry, of which 901 were identified as RS cases and 98,010 as controls. RS cases were defined using the PhecodeX 448.1, which maps to RS, and is derived from ICD codes [42].

Genotyping was performed using Illumina Global Diversity Array (GDA; Illumina Inc.,

San Diego, CA, USA) for sample QC. Variant calls were generated from whole-genome sequencing (WGS) data using a joing calling approach aligned to the GRCh38 reference genome [43].

Michigan Genomics Initiative

The Michigan Genomics Initiative (MGI) is a longitudinal, hospital-based biobank of Michigan Medicine patients, maintained by the University of Michigan. It integrates electronic health record data with genetic data from over 100,000 participants recruited primarily during surgical or diagnostic procedures at Michigan Medicine (https://aidhi.umich.edu).

The freeze 6 cohort used in this study includes 68,275 participants, of whom 1,529 were identified as RS cases and 66,746 as controls, based on Phecode 443.1. Participants consisted primarily of European ancestry.

Genotyping was performed using customized Illumina Infinium CoreExome and Infinium GSA (Illumina Inc., San Diego, CA, USA). Genotype data were aligned to the GRCh38 reference genome and imputed using the TOPMed reference panel [44; 45].

Post-imputation QC was applied by excluding variants with INFO <0.3 or MAF <0.0001.

2.2 GWAS

A genome-wide association study (GWAS) was performed using the FinnGen R13 dataset with REGENIE [46] (version 2.2.4), following the FinnGen REGENIE pipeline[6]. REGENIE employs a two-step procedure. In the first step, a genome-wide prediction model is constructed using Ridge regression on pruned genotype data. To avoid bias from nearby variants, the model is trained using a Leave-One-Chromosome-Out (LOCO) approach, in which the chromosome containing the variant being tested is left out [46].

In the second step, each genetic variant is tested individually for association with RS using logistic regression. For variants showing potential association ($p < 0.01$), REGENIE applies an approximate Firth correction to account for potential bias arising from case-control imbalance or low allele counts [46].

Individuals with RS were identified using the ICD-10 code I73.0. All GWAS analyses conducted on FinnGen R13 were adjusted for the following covariates: sex, age at death or end of follow-up, genotyping batches, and the first ten genetic principal components (PCs) to control for population stratification and other confounding factors.

To examine sex-specific genetic effects, the GWAS was repeated separately in males

[6]https://github.com/FINNGEN/regenie-pipelines

and females (218,284 and 282,064 individuals, respectively). In these analyses, the sex covariate was excluded, and only participants of the corresponding sex were included in each run. All other parameters were kept consistent with the main analysis.

Manhattan plots were generated using R version 4.5.1 with the `qqman` package (version 0.1.9). Plots for each GWAS summary statistic are presented in Appendix C.

2.3 Fine-mapping

To further prioritize likely causal variants at loci identified in the FinnGen R13 GWAS, fine-mapping was conducted using the FinnGen R13 dataset within the FinnGen Sandbox environment.

Fine-mapping was performed using two methods: FINEMAP [47; 48] and SuSiE (Sum of Single Effects) [19], both implemented through FinnGen fine-mapping pipeline[7]. For each genome-wide significant locus ($p < 5 \times 10^{-8}$), a fine-mapping region was specified as a 3 Mb window centered on the lead variant, merging any overlapping regions. GWAS summary statistics were preprocessed and separated into files corresponding to each region. LD patterns between variants within each region were computed using LDstore2 [49], based on FinnGen genotype data.

Using the prepared GWAS summary statistics and the LD information from the FinnGen data, fine-mapping was performed with both FINEMAP and SuSiE, allowing for up to 10 causal variants per locus.

2.4 GWAS meta-analysis

GWAS meta-analysis was performed using METAL, a widely used and computationally efficient software tool for combining GWAS summary statistics across studies [15]. Prior to meta-analysis, GWAS summary statistics were carefully harmonized to ensure compatibility across the study cohorts. These preparatory steps are described in detail in the subsections below.

2.4.1 SNP reference data

The GWAS meta-analysis incorporated SNP data from dbSNP build ID 157, a catalog of genetic variants aligned to the human reference genome assembly GRCh38.p14 (patch release 14), released by the Genome Reference Consortium [50] in February 2022. It integrates primary chromosomal sequences and a set of patch scaffolds, including both

[7] https://github.com/FINNGEN/finemapping-pipeline

error-correcting FIX-patches and new NOVEL-patches. All SNP data were obtained from the National Center for Biotechnology Information (NCBI)[8].

To ensure consistency in chromosome naming conventions between the SNP data and the GWAS summary statistics, the GRCh38.p14 assembly report was employed. For example, the SNP data uses Reference Sequence (RefSeq) accession IDs for chromosomes (such as NC_000001.11) whereas the GWAS summary statistics use more common labels for chromosomes (such as chr1, or simply 1). The relevant columns from the assembly report (`Sequence-Role`, `Assigned-Molecule`, and `RefSeq-Accn`) were used to construct a mapping between the RefSeq accession IDs and conventional chromosome labels. For primary chromosomes, the RefSeq accession IDs were renamed to the standard chromosome number i.e., NC_000001.11 to 1. For patch scaffolds, the renaming convention reflected both the patch type and the chromosome. Specifically, if the `Sequence-Role` indicated a FIX-patch and the `Assigned-Molecule` was chromosome 1, the scaffold was renamed as fix1. This approach was similarly applied to other patch types (e.g., alt-scaffolds and NOVEL-patches) and chromosomes, ensuring that patch information was preserved and easily traceable in the resulting dataset.

Scaffolds were retained for analysis if their `Sequence-Role` corresponded to assembled-molecule (representing the primary chromosome sequences i.e., 1–22, X, and Y) or to patch types such as alt-scaffold, FIX-patch and NOVEL-patch. Unplaced and unlocalized scaffolds and mitochondrial sequences were excluded to focus the analysis solely on the nuclear genome.

Once the assembly report was adapted to the wanted chromosome naming conventions, the original SNP file was annotated using a SAMtools package BCFtools [51] (version 1.2.1) according to the modified assembly report mapping. This process produced a VCF file in which chromosome names matched the desired format and were compatible with GWAS summary statistics. In addition, particular attention was paid to the uniqueness of rsIDs: after renaming the prefixes of alt-scaffolds, FIX-, and NOVEL-patch additions back to the conventional chromosome labels, the dataset was checked to ensure no overlapping rsIDs remained among these entries. Specifically, in cases where duplicates were found, only the variant associated with the assembled-molecule sequence-role was retained. This ensured that each variant maintained a unique identifier within the dataset, thereby preventing potential annotation conflicts.

[8]https://ftp.ncbi.nih.gov/snp/latest_release/VCF/

2.4.2 GWAS summary statistics harmonization

The GWAS summary statistics were first aligned and harmonized using a script provided by the Broad Institute[9]. This harmonization process involved checking each GWAS summary statistic against the human genome reference FASTA file (obtained from https://hgdownload.soe.ucsc.edu/downloads.html; GRCh38 or GRCh37, depending on the summary statistic), which contain the official DNA sequences for each chromosome.

The script verified that all variants were mapped to valid chromosomes, and it identified and corrected any cases where the reported reference and alternative alleles were flipped relative to the reference genome. Additionally, it flagged entries that represented ambiguous indels (insertions or deletions) whose exact position could not be uniquely determined due to repetitive sequence context.

In addition to harmonization, the GWAS summary statistics were annotated with corresponding rsIDs, if those were missing from the original, using a script implemented by FinnGen. This step matched each variants chromosome, position, reference allele, and alternative allele in the GWAS summary statistics to the corresponding entry in the formatted reference SNP VCF file. This rsIDification ensured that each variant could be consistently and unambiguously identified across the cohorts.

Subsequently, any missing columns in GWAS summary statistics were computed and added, for example, p-values, beta coefficients and their standard errors.

2.4.3 GWAS meta-analysis with METAL

After the harmonization and rsIDification, the GWAS meta-analysis was conducted using METAL [15] (version 2011-03-25) with the following settings: `GENOMICCONTROL` option was enabled to adjust for inflation of association test statistics resulting from population stratification. This refers to differences in allele frequencies between subpopulations, and can lead to spurious associations [52]. `AVERAGEFREQ` option was used to track the alternative allele frequency (AAF) across the cohorts. Finally, `ANALYZE HETEROGENEITY` option was employed to assess whether the effect sizes (e.g., beta coefficients) for each variant were consistent throughout the cohorts or showed significant heterogeneity. Heterogeneity between the study cohorts was estimated using Cochrans Q-test [17].

Meta-analysis was run based on both effect size estimates with standard errors (inverse-variance weighting) and sample size. To ensure the robustness of the results, the meta-analysis summary statistics were filtered to include only the SNPs present in more than one study cohort. This filtering step reduces spurious associations that may arise from variants observed in a single cohort only and thus, increases the confidence in findings.

[9] https://www.broadinstitute.org

Incomplete rows were cleaned from the summary statistics and harmonization was rerun to ensure consistency. Before the filtering, the GWAS meta-analysis summary statistics included a total of 74,774,272 SNPs and after the filtering and harmonization 29,375,236 SNPs remained.

Variants were further filtered to include only those with an AAF greater than 0.005 to remove very rare and potentially unreliable variants, and with a heterogeneity p-value greater than 10^{-6}, to ensure consistent effects across cohorts. The final GWAS meta-analysis summary statistics were reported using the GRCh38 reference genome build, with chromosome and base-pair position information added to facilitate downstream analyses.

2.5 Post-GWAS visualization and annotation

To assess the overall quality of the GWAS meta-analysis results, a QQ plot was generated. Following this, lead variants and genomic risk loci were identified and annotated using a combination of tools, including FUMA GWAS (Functional Mapping and Annotation of GWAS; version 1.5.2) [22], LocusZoom (version 0.14.0) [26], the R package `topr` (version 2.0.2) [27], and VannoPortal (Release 1.0.1) [28].

2.5.1 QQ plot analysis

A QQ plot was generated to evaluate the overall distribution of p-values obtained from the GWAS meta-analysis and to assess potential inflation or deflation. The QQ plot is presented in the results section. The plot was generated using R version 4.5.1 with the `qqman` package (version 0.1.9).

2.5.2 FUMA GWAS

FUMA was used to both statistically define and annotate lead variants from the GWAS meta-analysis. The meta-analysis summary statistics were uploaded to FUMA GWAS platform (https://fuma.ctglab.nl), and the SNP2GENE pipeline was applied.

The total sample size was set to 1,868,346, corresponding to the combined sample size across cohorts. Independent significant variants were defined using an r^2 threshold of 0.6, with the 1000G phase 3 (EUR populations) dataset serving as the LD reference panel. All variants were included regardless of MAF (MAF=0), thereby including both common and rare variants.

For functional mapping, variants were filtered using CADD scores (threshold ≥ 12.37) and regulomeDB scores (maximum score of 7), as recommended. Chromatin state filtering was not applied in this analysis.

eQTL mapping was conducted using GTEx version 8 data for Artery Tibial, Muscle Skeletal, and Nerve Tibial tissues, applying a false discovery rate (FDR) threshold of 0.05. Although FUMA provides integrated eQTL annotations, a separate eQTL analysis was conducted using the more recent GTEx v10 data to directly assess regulatory effects in the aforementioned tissues (Section 2.6).

Moreover, MAGMA gene expression analysis was performed within FUMA using GTEx v8 as reference to investigate tissue-specific gene enrichment.

2.5.3 LocusZoom

To visually examine the meta-analysis results and inspect the lead variants, the GWAS meta-analysis summary statistics were uploaded to the LocusZoom web platform (https://my.locuszoom.org). Regional association plots were generated with LD estimates calculated from the 1000G European reference panel to match the ancestry of the study cohorts. These visualizations facilitated the interpretation of association signals by contextualizing their genomic location, thereby making it easier to identify likely causal variants [53].

2.5.4 Variant annotation

To ensure robustness and consistency in the identification of lead variants, the `get_lead_snps()` function from `topr` package was applied with default parameters to detect lead variants based on the GWAS meta-analysis summary statistics. The results obtained from `topr` were then compared with those identified using FUMA and LocusZoom to cross-validate findings and increase confidence in the selected lead variants.

Each lead variant was queried in VannoPortal and primarily assessed for basic information and disease or trait association categories. This included identifying potentially affected genes and transcripts, as well as reviewing credible variants identified by other GWAS and molecular trait QTLs linked to the variant.

To explore broader phenotypic associations, the rsIDs of lead variants were queried in public phenotype association databases (pheWebs), including FinnGen[10] and Open Targets Genetics[11], to assess whether the lead variants had been previously associated with other traits or diseases.

[10]https://results.finngen.fi

[11]https://genetics.opentargets.org

2.6 eQTL analysis

To explore the potential regulatory effects of the lead variants, an independent eQTL analysis was performed using data from the Genotype-Tissue Expression (GTEx) Portal (https://gtexportal.org/, version 10). This version includes gene expression data from 19,788 samples across 54 tissue types collected from 946 donors.

Lead variants were queried across this tissue panel to identify significant eQTL associations. Special attention was given to tissues considered biologically relevant to RS, such as vascular, nervous, and muscular tissues. In GTEx, these tissue categories are represented by Artery Tibial, Nerve Tibial, and Muscle Skeletal, respectively. eQTL associations in these tissues can provide supporting evidence that a variant may influence disease susceptibility by altering gene expression.

3 Results

A GWAS on FinnGen R13 data identified genetic variants associated with RS. To increase statistical power and assess the generalizability of findings, a genome-wide association meta-analysis was subsequently conducted across eight independent cohorts, including a total of 17,793 RS cases and 1,850,553 controls, primarily of European ancestry.

GWAS summary statistics from the additional cohorts were obtained from publicly available resources or through research collaborations. These additional cohorts included participants from the UKB, EstBB, MGB Biobank, MVP, G&H, AoU, and MGI.

3.1 GWAS results from FinnGen R13

In the FinnGen R13 dataset, two loci reached genome-wide significance ($p < 5 \times 10^{-8}$), with lead SNPs identified near the α_{2A}-adrenergic receptor (*ADRA2A*) and the Iroquois Homeobox 1 (*IRX1*) genes (Figure 3). The FinnGen R13 GWAS included 3,278 RS cases and 497,070 controls of Finnish ancestry. RS diagnoses were based on ICD-10 code I73.0.

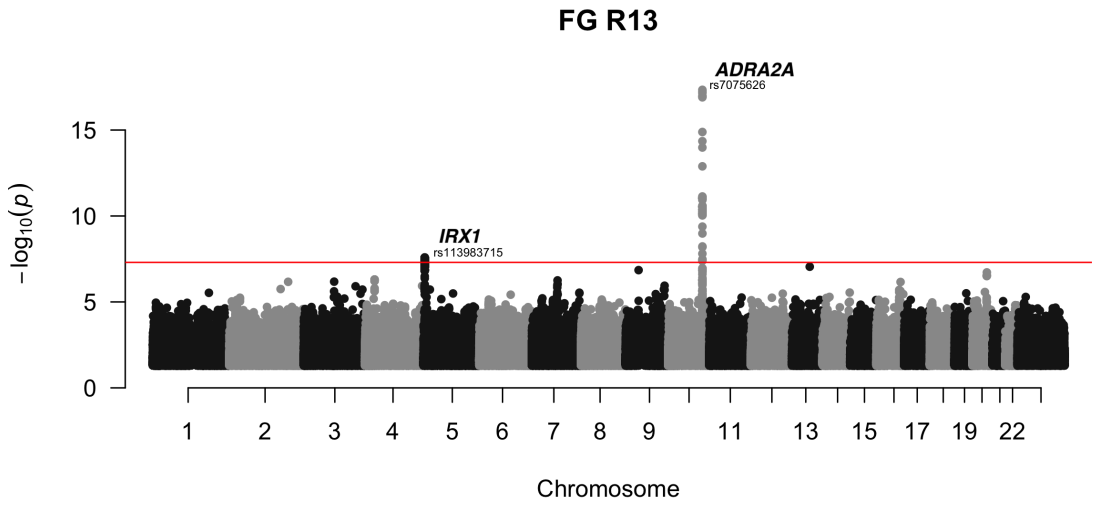


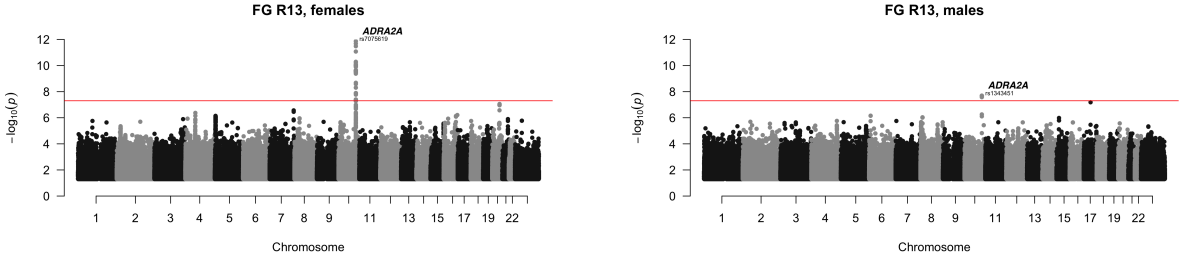
Figure 3: Manhattan plot of genome-wide association results for Raynaud's syndrome in the FinnGen (FG) R13 dataset, including 3,278 cases and 497,070 controls. The red horizontal line indicates the genome-wide significance threshold ($p < 5 \times 10^{-8}$).

The lead variant near the *ADRA2A* locus was rs7075626, located on chromosome 10, showing the strongest association signal with a p-value of 4.59×10^{-18}. The reference allele was G and the alternative allele C, with an AAF of 31.1%. Near the *IRX1* locus on chromosome 5, the lead variant was rs113983715 with a p-value of 2.59×10^{-8}. The reference allele was T, the alternative allele G, and the AAF was 22.6%.

To explore the sex-specific genetic effects, separate GWAS analyses were performed in females (n = 282,064; 2,368 cases) and males (n = 218,284; 910 cases) from the FinnGen R13 dataset. Both analyses identified genome-wide significant associations near the *ADRA2A* locus on chromosome 10, although different lead variants were observed in each sex.

In females, the lead variant was rs7075619, with a p-value of 1.36×10^{-12}. The reference allele was G and the alternative allele T, with an AAF of 21.5%. In males, the lead variant was rs1343451, with a p-value of 2.02×10^{-8}. The reference allele was A and the alternative allele G, with an AAF of 24.8%.

These findings support a robust and consistent association near the *ADRA2A* locus across sexes. Manhattan plots illustrating the sex-specific results are presented in Figure 4, and summary statistics for FinnGen lead variants are provided in Appendix D.



(a) FinnGen R13 females

(b) FinnGen R13 males

Figure 4: Manhattan plots of sex-specific genome-wide association results for Raynaud's syndrome in FinnGen (FG) R13 dataset: **(a)** females (2,368 cases, 279,696 controls), **(b)** males (910 cases, 217,374 controls). The red horizontal line indicates the genome-wide significance threshold ($p < 5 \times 10^{-8}$).

3.2 Fine-mapping results from FinnGen R13

Fine-mapping was conducted on the two genome-wide significant loci identified in FinnGen R13 GWAS: *ADRA2A* on chromosome 10 and *IRX1* on chromosome 5. Both loci were successfully fine-mapped using two complementary methods: SuSiE and FINEMAP.

Near the *ADRA2A* locus, SuSiE converged on a 99% credible set consisting of five variants. The top three variants with highest posterior inclusion probabilities (PIPs) were rs7090046 (PIP = 0.33), rs1343451 (PIP = 0.28), and rs1343449 (PIP = 0.19). FINEMAP produced a similar 95% credible set, also prioritizing the same three variants with PIPs of 0.32, 0.27, and 0.18, respectively.

In contrast, near the *IRX1* locus, the fine-mapping results diverged between the two methods. SuSiE identified a broader 95% credible set comprising 155 variants. The top

three variants were rs521666, rs113983715, and rs2005071 (PIP = 0.0106, 0.0104, and 0.0102, respectively). FINEMAP, on the other hand, identified rs17723577, rs187285896, and rs533256615 (PIP = 0.016, 0.015, and 0.014, respectively) as the most likely causal variants.

3.3 GWAS meta-analysis results

A genome-wide association meta-analysis combining summary statistics from eight cohorts identified 14 independent loci that achieved genome-wide significance ($p < 5 \times 10^{-8}$). Lead variants were identified using LocusZoom and `topr` R package. Of the 14 loci, five had been previously reported by Tervi et al. [3], while nine represent novel associations. See Appendix A for the complete list of significant independent GWAS meta-analysis lead variants.

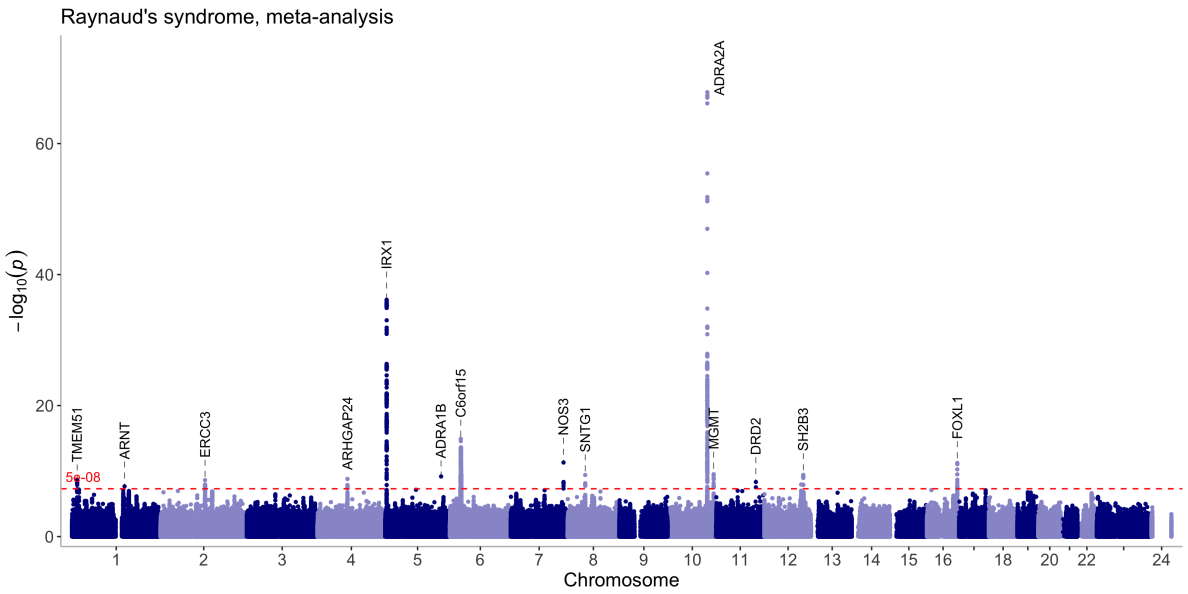


Figure 5: A manhattan plot of Raynaud's syndrome genome-wide association meta-analysis (17,793 cases and 1,850,553 controls) combining FinnGen R13, the UK Biobank, the Estonian Biobank, the Mass General Brigham Biobank, the Million Veteran Program, Genes & Health, the All of Us Research Program, and the Michigan Genomics Initiative data. Each point represents a single nucleotide polymorphism (SNP) with its genomic position on the x-axis and the negative $\log_{10}$ of its p-value on the y-axis. The red horizontal line indicates the genome-wide significance threshold ($p < 5 \times 10^{-8}$).

The strongest association in the meta-analysis was identified near the *ADRA2A* locus on chromosome 10. The lead variant, rs7075626, is an intergenic SNP located downstream of *ADRA2A*. It reached genome-wide significance in four individual cohorts (FinnGen, UKB, EstBB, MVP), and showed the most significant association overall ($p = 1.45 \times 10^{-68}$,

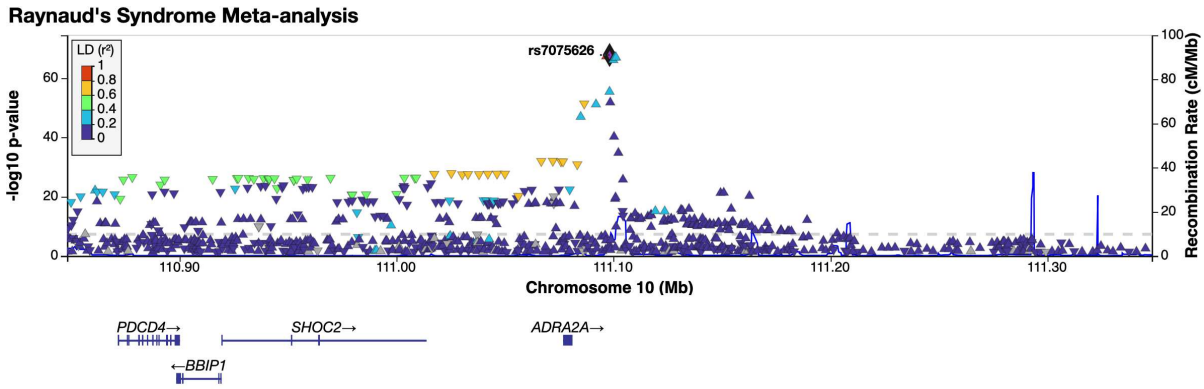


Figure 6: LocusZoom regional association plot for the lead variant rs7075626 near the *ADRA2A* locus, showing the strength of the association, local linkage disequilibrium (LD), recombination rates, and the positions of nearby genes.

beta [SE] = 0.20 [0.01]; Figures 5 and 6). Heterogeneity analysis indicated moderate but non-significant variation across cohorts ($I^2 = 44.7\%$, $\chi^2 = 12.6$, p = 0.08, see Appendix A). Association statistics for rs7075626 across cohorts are presented in Table 2.

Comparison of the lead variant rs7075626	
SNP	rs7075626
FinnGen R13	$p = 4.59 \times 10^{-18}$ Beta = 0.223 SE = 0.026
UK Biobank	$p = 1.39 \times 10^{-24}$ Beta = 0.214 SE = 0.021
Estonian Biobank	$p = 1.17 \times 10^{-9}$ Beta = 0.223 SE = 0.026
Mass General Brigham Biobank	$p = 1.26 \times 10^{-7}$ Beta = 0.176 SE = 0.033
Million Veteran Program	$p = 1.56 \times 10^{-12}$ Beta = 0.240 SE = 0.034
Genes & Health	p = 0.649 Beta = 0.027 SE = 0.060
All of Us Research Program	$p = 2.92 \times 10^{-5}$ Beta = 0.214 SE = 0.051
Michigan Genomics Initiative	$p = 7.62 \times 10^{-5}$ Beta = 0.157 SE = 0.040
Meta-analysis	$p = 1.45 \times 10^{-68}$ Beta = 0.202 SE = 0.012

Table 2: Comparison of the GWAS meta-analysis lead variant rs7075626 near the *ADRA2A* locus across the eight cohorts.

Several other loci demonstrated strong evidence of association. The second strongest signal was observed for rs7706161 ($p = 7.07 \times 10^{-37}$, beta [SE] = 0.15 [0.01]), an intergenic variant located near the *IRX1* gene on chromosome 5. Another notable finding was rs3130968 ($p = 1.21 \times 10^{-15}$, beta [SE] = 0.14 [0.02]), located near Chromosome 6 Open Reading Frame 15 (*C6orf15*) in the human leukocyte antigen (HLA) region on chromosome 6.

Additional lead variants were located in various genomic regions. These included intergenic variants near genes such as *FOXL1*, *MGMT*, and *DRD2*; intronic variants within coding transcripts of genes including *ARHGAP24*, *SH2B3*, *SNTG1*, and *ARNT*; as well as variants within untranslated regions (UTRs), including both 5t' UTR and 3t' UTR intronic variants, such as those found in *NOS3*, *ADRA1B*, *ERCC3*, and *TMEM51*. The significant independent lead variants from GWAS meta-analysis, along with their association statistics, are presented in Table 3.

Independent genetic loci from GWAS meta-analysis (17,793 cases and 1,850,553 controls)									
SNP	**CHR**	**POS (b38)**	**Ref.**	**Alt.**	**Comb.Alt.AF**	**p-value**	**Beta**	**SE**	**Nearest gene**
rs7075626	10	111,098,213	G	C	0.333	1.45×10^{-68}	0.202	0.012	*ADRA2A*
rs7706161	5	4,047,779	A	G	0.273	7.07×10^{-37}	0.152	0.012	*IRX1*
rs3130968	6	31,097,294	C	T	0.142	1.21×10^{-15}	0.143	0.018	*C6orf15/HLA*
rs3918226	7	150,993,088	C	T	0.076	4.95×10^{-12}	0.141	0.020	*NOS3*
rs526882	16	86,679,425	A	C	0.351	5.85×10^{-12}	0.078	0.011	*FOXL1*
rs2209592	10	129,165,236	A	G	0.672	3.03×10^{-10}	-0.074	0.012	*MGMT*
rs7310615	12	111,427,245	C	G	0.543	3.89×10^{-10}	-0.070	0.011	*SH2B3*
rs13263870	8	50,583,718	T	C	0.735	3.95×10^{-10}	-0.080	0.013	*SNTG1*
rs6556458	5	159,867,336	C	T	0.580	6.63×10^{-10}	0.073	0.012	*ADRA1B*
rs17010957	4	85,798,012	T	C	0.145	1.58×10^{-9}	0.094	0.016	*ARHGAP24*
rs4150495	2	127,272,049	G	A	0.358	2.39×10^{-9}	-0.086	0.014	*ERCC3*
rs10803364	1	15,191,937	A	G	0.422	2.44×10^{-9}	-0.066	0.011	*TMEM51*
rs569240936	11	113,536,934	T	A	0.011	4.62×10^{-9}	1.017	0.174	*DRD2*
rs10305667	1	150,859,632	C	T	0.260	2.13×10^{-8}	-0.070	0.013	*ARNT*

Table 3: Independent GWAS meta-analysis lead variants for Raynaud's syndrome, combining FinnGen R13, the UK Biobank, the Estonian Biobank, the Mass General Brigham Biobank, the Million Veteran Program, Genes & Health, the All of Us Research Program, and the Michigan Genomics Initiative data (17,793 cases and 1,850,553 controls). Base-pair positions are based on the GRCh38 genome build. The reference allele (Ref.) is the non-effect allele, and the alternative allele (Alt.) is the effect allele. Combined alternative allele frequency (Comb. Alt. AF) is the frequency of the Alt. across all cohorts. Beta indicates the estimated effect size of Alt., and SE is the standard error of Beta. p-values were calculated using METAL [15]. See Appendix A for complete list of meta-analysis lead variants.

The QQ plot comparing the observed versus expected $-\log_{10}$(p-values) is presented in Figure 7. The observed values deviate sharply from the expected distribution, especially

in the upper tail. This suggests that there are more significant findings than would be expected by chance, pointing to potential true associations. The genomic control lambda λ, calculated using the median of the test statistics, was 1.046, indicating minimal inflation and suggesting that confounding factors are likely well controlled in the meta-analysis.

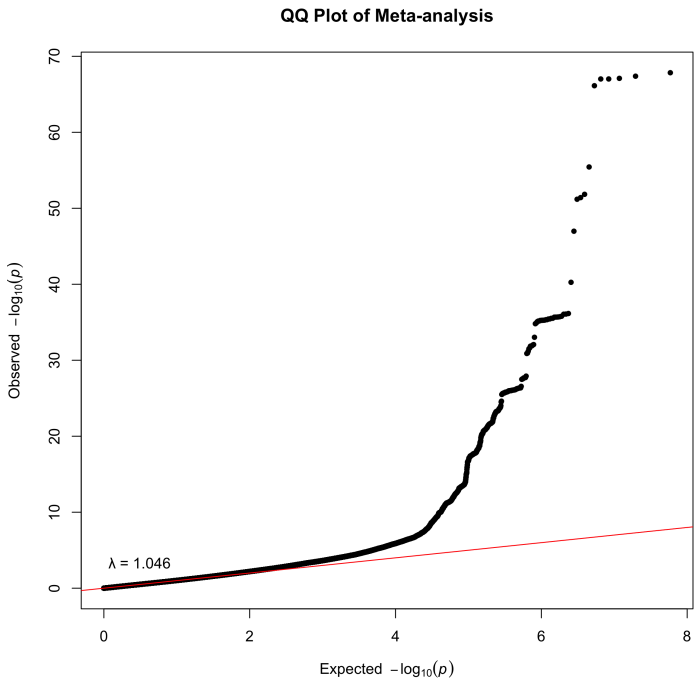


Figure 7: Quantile-Quantile (QQ) plot of observed and expected $-\log_{10}$ (p-values) from the GWAS meta-analysis of Raynaud's syndrome across eight cohorts with 17,793 cases and 1,850,553 controls. Genomic control lambda (λ) is the ratio between the median of the observed and expected chi-squared statistics under the null hypothesis.

To further refine the genome-wide significant loci identified in the primary meta-analysis, PIPs were obtained from the LocusZoom website using the GWAS meta-analysis summary statistics. At the *ADRA2A* locus on chromosome 10, a credible set of five variants was identified. The lead variant, rs7075626, showed the highest probability of being causal (PIP = 0.54), followed by rs7908645 (PIP = 0.19). For the *IRX1* locus on chromosome 5, a larger credible set of 20 variants was identified, with the top variants being rs7706161, rs72731435, and rs10512704 (PIP = 0.15, 0.13, and 0.12, respectively), suggesting multiple potential causal variants at this locus. The credible sets and corresponding PIPs for both loci are presented in Table 4.

Using the FUMA GWAS platform, the input SNPs from the GWAS meta-analysis summary statistics were mapped to a total of 1,427 protein-coding genes. FUMA identified one genome-wide significant locus, with a lead variant rs7067540 near *PDCD4*.

Credible set variants for *ADRA2A* and *IRX1* loci		
Locus	**SNP**	**PIP**
ADRA2A	rs7075626	0.54
	rs7908645	0.19
	rs1343449	0.10
	rs1343451	0.08
	rs7090046	0.08
IRX1	rs7706161	0.15
	rs72731435	0.13
	rs10512704	0.12
	rs56324263	0.07
	rs11743330	0.06
	rs2005071	0.05
	rs57367478	0.05
	rs59774773	0.05
	rs57592121	0.04
	rs59799921	0.04

Table 4: Credible set variants and posterior inclusion probabilities (PIPs) for the *ADRA2A* and *IRX1* loci based on GWAS meta-analysis summary statistics using LocusZoom. The top 10 variants for *IRX1* ranked by PIP are shown.

MAGMA gene-set analysis revealed one significantly enriched gene set associated with liver cancer recurrence (`KUROKAWA_LIVER_CANCER_EARLY_RECURRENCE_UP`), originally described by Kurokawa et al. [54], including two genes. This gene set remained significant after multiple testing correction, with a Bonferroni-adjusted p-value of 9.7×10^{-22}. However, its relevance to RS remains unclear and may reflect pleiotropic regulatory mechanisms. Furthermore, tissue-specific expression analysis using MAGMA and GTEx v8 did not identify any significantly enriched tissues.

In addition to the primary fixed effects meta-analysis, a sample size-weighted meta-analysis was conducted. The results supported the findings from the primary meta-analysis, with consistent signals among the top associations. The total weighted sample size for this meta-analysis was 1,804,692. In total, 15 lead variants were identified using the LocusZoom and `topr` tools. The top 10 independent lead variants from sample size-weighted meta-analysis are summarized in Table 5.

3.4 eQTL analysis results

To explore the regulatory effects of lead variants, single-tissue eQTL associations were assessed using data from GTEx Analysis Release v10. Normalized effect size (NES) is the effect of the alternative allele relative to the reference allele in the human genome reference.

Top 10 independent genetic loci from sample size-weighted GWAS meta-analysis									
SNP	**CHR**	**POS (b38)**	**Ref.**	**Alt.**	**p-value**	**Weight**	**Z-score**	**Het.p-value**	**Nearest gene**
rs7075626	10	111,098,213	G	C	3.25×10^{-67}	1,804,692	17.32	0.012	*ADRA2A*
rs7706161	5	4,047,779	A	G	2.30×10^{-36}	1,804,692	-12.59	0.18	*IRX1*
rs3130968	6	31,097,294	C	T	1.91×10^{-12}	1,751,395	7.04	0.002	*C6orf15/HLA*
rs526882	16	86,679,425	A	C	1.58×10^{-11}	1,804,692	-6.74	0.74	*FOXL1*
rs6556458	5	159,867,336	C	T	4.02×10^{-10}	1,804,692	6.25	0.87	*ADRA1B*
rs13263870	8	50,583,718	T	C	5.27×10^{-10}	1,760,918	6.21	0.28	*SNTG1*
rs4150495	2	127,272,049	G	A	3.45×10^{-9}	1,045,497	-5.91	0.72	*ERCC3*
rs7310615	12	111,427,245	C	G	3.71×10^{-9}	1,804,692	5.90	0.019	*SH2B3*
rs3918226	7	150,993,088	C	T	4.28×10^{-9}	1,804,692	5.87	0.003	*NOS3*
rs17010957	4	85,798,012	T	C	4.28×10^{-9}	1,804,692	-5.82	0.51	*ARHGAP24*

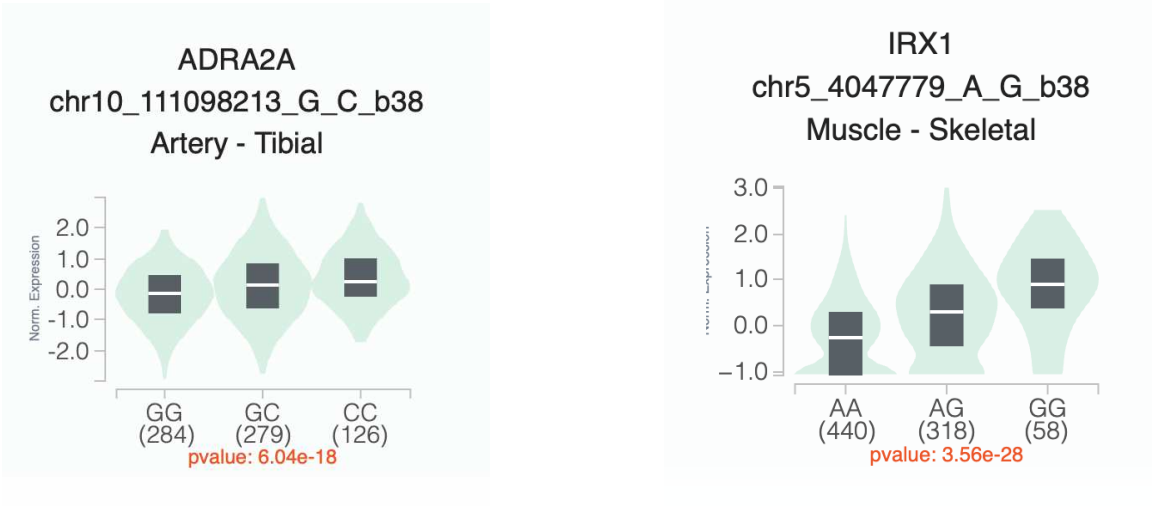
Table 5: Top 10 independent sample size-weighted GWAS meta-analysis lead variants for Raynaud's syndrome, combining FinnGen R13, the UK Biobank, the Estonian Biobank, the Mass General Brigham Biobank, the Million Veteran Program, Genes & Health, the All of Us Research Program, and the Michigan Genomics Initiative data. Base-pair positions are based on the GRCh38 genome build. The reference allele (Ref.) is the non-effect allele, and the alternative allele (Alt.) is the effect allele. Weight is the sample size-based weighting used in the meta-analysis. Z-score is the standardized effect size of Alt. Het.p-value is the heterogeneity test p-value across cohorts. p-values were calculated using METAL [15]. See Appendix B for complete list of sample size-weighted meta-analysis lead variants.

Two lead variants rs7075626 near *ADRA2A* and rs7706161 near *IRX1* demonstrated significant associations with gene expression in selected tissues (Table 6). Notably, rs7075626 was associated with increased *ADRA2A* expression in tibial arteries. In contrast, no significant association was observed in coronary arteries (NES = 0.027, p = 0.51) or the aorta (NES = 0.072, p = 0.13).

GTEx eQTL associations for lead variants				
SNP	**Gene**	**Tissue**	**NES**	**p-value**
rs7075626	*ADRA2A*	Artery Tibial	0.31	6.0×10^{-18}
rs7706161	*IRX1*	Muscle Skeletal	0.54	3.6×10^{-28}
		Artery Tibial	0.41	1.5×10^{-14}
rs3130968	*HLA-C*	Muscle Skeletal	0.75	8.9×10^{-39}
		Artery Tibial	0.39	5.9×10^{-20}
		Nerve Tibial	0.42	1.0×10^{-14}
rs17010957	*ARHGAP24*	Artery Tibial	0.69	1.0×10^{-55}
		Nerve Tibial	0.15	7.0×10^{-5}

Table 6: Expression quantitative trait locus (eQTL) associations from GTEx v10 for selected lead variants across relevant tissues. **NES** = Normalized Effect Size.

Additionally, rs3130968 near *HLA-C* and rs17010957 near *ARHGAP24* also showed strong tissue-specific eQTL associations, particularly in skeletal muscle and tibial arteries. However, interpreting eQTL signals in regions such as *HLA-C* is challenging due to extensive LD and high gene density, which complicate the identification of causal variants [55]. Expression levels stratified by genotype for *ADRA2A* and *IRX1* are visualized in violin plots (Figure 8).



(a) rs7075626 (*ADRA2A*)

(b) rs7706161 (*IRX1*)

Figure 8: Expression quantitative trait locus (eQTL) violin plots for meta-analysis lead variants in relevant tissues, showing gene expression levels stratified by genotype: **(a)** rs7075626 and *ADRA2A* expression in Artery Tibial, and **(b)** rs7706161 and *IRX1* expression in Muscle Skeletal, based on GTEx v10 data.

4 Discussion

This thesis investigated the genetic and biological risk factors underlying Raynaud's syndrome (RS) through a genome-wide association study (GWAS) of FinnGen R13 data, followed by a large-scale meta-analysis across eight cohorts. The results revealed 14 genome-wide significant loci, including nine novel associations. These findings strengthen the evidence that RS has a genetic component and highlight potential biological pathways involved in disease susceptibility.

4.1 Key findings

The strongest association was observed near *ADRA2A* locus, which encodes the α_{2A}-adrenergic receptor, a key modulator of sympathetic nervous system activity and cold-induced vasoconstriction [56; 3]. This locus produced the top signal in both FinnGen R13 GWAS and the GWAS meta-analysis, underscoring its potential central role in RS pathophysiology.

The second strongest association was found near *IRX1*, which encodes a homeobox transcription factor involved in tissue development and cellular differentiation [57; 58]. Although its specific role in RS is not fully understood, its consistent signal across analytic methods suggest potential regulatory role in thermogenic response or adipose tissue function [59].

The HLA region, near *C6orf15*, also emerged among the top associations, reinforcing the contribution of immune mechanisms to RS particularly in secondary RS, which is often linked to underlying autoimmune conditions [60]. The associated variant, rs3130968, has previously been associated with peripheral vascular disease and several autoimmune diseases [61; 62].

The *NOS3* locus, encoding endothelial nitric oxide synthase (eNOS), also showed a strong association. *NOS3* is a well-known regulator of vascular dilation through nitric oxide production [63], and its link to RS is consistent with known impairments in vasodilation among affected individuals [60]. The variant rs3918226 has been previously associated with coronary artery disease and hypertension [64; 65], and has also been identified as a novel susceptibility marker for peripheral artery disease [66].

Other loci of interest include *ADRA1B* and *SH2B3*. *ADRA1B*, encoding the α_{1B}-adrenergic receptor, contribute to vasoconstriction, complementing the role of *ADRA2A* [67]. *SH2B3* (also known as *LNK*) is a negative regulator of cytokine signaling and has been linked to both immune-mediated and cardiovascular traits, including hypertension [68; 69]. The *SH2B3* variant rs7310615 has been linked to blood cancers such as myeloproliferative neoplasms and JAK2 V617F clonal hematopoiesis [70].

4.2 GWAS and fine-mapping in FinnGen

In the FinnGen R13 GWAS, two loci near *ADRA2A* and *IRX1* reached genome-wide significance for association with RS. Notably, sex-specific analyses revealed different lead variants near *ADRA2A* in females and males, suggesting possible sex-specific genetic effects. Such differences may arise from variation in LD structures between sexes, for example due to differences in recombination rates [71; 9]. Additionally, differences in sample size and statistical power could affect the identification of lead variants in sex-specific analyses. Given that RS is more common in females, with an approximate female-to-male ratio of 4:1, sex remains the most consistent risk factor for the disease [72]. These findings highlight the importance of considering sex as a biological variable in genetic studies of RS.

Fine-mapping further refined the signal at both loci. Near *ADRA2A*, both methods produced a compact credible set with relatively high PIPs, supporting a strong and localized association at the locus. In contrast, fine-mapping near *IRX1* resulted in a broader credible set with lower PIPs, indicating greater uncertainty in identifying likely causal variants at this locus.

4.3 GWAS meta-analysis

The GWAS meta-analysis across eight cohorts identified a robust association near the *ADRA2A* locus, where the lead variant rs7075626 showed consistent association across multiple cohorts. While moderate heterogeneity was observed ($I^2 = 44\%$), the non-significant heterogeneity p-value of 0.08 suggests that variation in effect estimates was likely due to chance rather than true biological or methodological differences. These results support the reliability of this association across the diverse study cohorts.

Fine-mapping further refined the signal near *ADRA2A*, resulting in a compact credible set with high PIPs. These findings are consistent with those from the FinnGen R13 GWAS and provide strong evidence for a causal role of variants at this locus.

The QQ plot showed slight inflation in the upper tail of the distribution, suggesting the presence of true polygenic associations. However, the genomic inflation factor ($\lambda = 1.046$) was close to the expected value of 1, indicating that the potential confounding factors, such as population stratification, are unlikely to have had a major impact. Furthermore, the consistency of top associations in both effect size-based and sample size-weighted meta-analyses supports the robustness of the findings.

MAGMA analysis, on the other hand, did not reveal any significant enriched pathways or tissues.

4.4 eQTL

The eQTL findings provide functional support for the GWAS meta-analysis signals by linking lead variants to regulatory effects in biologically relevant tissues. For instance, increased *ADRA2A* expression in tibial arteries may enhance α_{2A}-adrenergic receptor activity, potentially contributing to the vasoconstriction characteristics of RS.

Lead variants near the *ADRA2A* (rs7075626), *IRX1* (rs7706161), *HLA-C* (rs3130968), and *ARHGAP24* (rs17010957) loci were associated with gene expression in relevant vascular and muscular tissues, specifically the tibial artery and skeletal muscle.

Genotype-stratified expression levels, visualized with violin plots, demonstrated allele-dependent effects on gene expression. For rs7075626 (*ADRA2A*), individuals with the CC genotype exhibited the highest expression in tibial artery tissue. Similarly, for rs7706161 (*IRX1*) the GG genotype showed the highest expression in skeletal muscle.

4.5 Previous studies

The findings of this thesis build on and expand previous genetic investigations of RS. Notably, prior research by Tervi et al. (2024) identified eight genome-wide significant loci associated with RS: *ADRA2A*, *IRX1*, *NOS3*, *C6orf15/HLA*, *ACVR2A*, *TMEM51*, *PCDH10-DT*, and *RAB6C*. Similarly, Hartmann et al. (2023) reported three loci, *ADRA2A*, *IRX1*, and *MICB*, associated with RS risk. In line with these findings, this thesis replicated several of these signals, including *ADRA2A*, *IRX1*, *NOS3*, *C6orf15/HLA*, and *TMEM51*, increasing the confidence and robustness of their involvement in RS pathophysiology.

Earlier research has emphasized the role of α_{2C}-adrenergic receptor (*ADRA2C*) in cold-induced vasoconstriction [56]. However, the recent genetic analyses, including the present thesis, point an additional and independent role of *ADRA2A* in modulating vascular tone under cold exposure [3; 29; 59].

4.6 Limitations

Although a GWAS is a powerful and widely used tool for identifying genetic loci associated with complex diseases and traits, interpreting these associations in a biological context remains challenging [8]. Most complex diseases and traits are influenced by thousands of common variants with small effect sizes. In addition, most variants identified in GWAS lie in non-coding regions making it difficult to interpret direct functional impacts [22]. Furthermore, genetic associations can vary across populations with different ancestries, limiting the generalizability of findings [8].

Regarding the phenotype definition, RS cases were identified as either primary or secondary RS based on diagnostic codes from medical records. While this approach leverages large-scale data, it may not capture all relevant cases. For example, mild or benign RS cases may be underreported in healthcare records or overlapping symptomps may be misclassified. The phenotype definitions were based on ICD-10 codes, a combination of ICD-10 and ICD-9 codes, or phecodes such as 443.1 and X448.1 all derived from health registries. These registries include hospital and primary care records (FinnGen), health records (UKB), hospital-based (MGB, MGI), and community-based or population-based cohorts (EstBB, G&H, AoU, MVP).

Another limitation concerns the diversity in QC procedures across cohorts. The GWAS summary statistics employed different genotyping arrays, imputation reference panels, MAF cut-offs, and quality filters. This may introduce variability that complicates direct comparisons across cohorts. Additionally, the analysis was mostly limited to individuals of European ancestry. While this minimizes the confounding factors due to population stratification, it limits the generalizability of findings to populations beyond European ancestry.

4.7 Translational potential and applications

The findings of this thesis have practical implications for research, clinical management, and biomedical engineering applications in RS. The robust genetic loci identified, particularly those involved in adrenergic signaling (*ADRA2A*, *ADRA1B*) and vascular regulation (*NOS3*), highlight biological pathways that could be directly targeted to reduce abnormal vasoconstriction. In a biomedical engineering context, these insights may guide the development of devices, sensors, or algorithms that modulate or monitor vascular responses in real time.

The discovered loci and variants could also contribute to risk stratification and precision medicine. PRSs derived from these results could identify individuals with higher susceptibility to RS, enabling personalized preventive measures, early monitoring, or tailored lifestyle recommendations.

Finally, integration of genetic risk information with digital health and wearable technologies presents additional translational possibilities. Individuals at higher genetic risk could benefit from personalized alerts during cold exposure or stress, allowing timely interventions to prevent or reduce the severity of vasospastic attacks.

5 Conclusions

This thesis presents the largest genome-wide association study and meta-analysis of Raynaud's syndrome (RS) conducted to date, leveraging data from over 17,000 cases and 1.8 million controls across eight study cohorts. The results revealed 14 genome-wide significant loci, of which nine were novel, implicating pathways in vascular tone regulation (*ADRA2A, ADRA1B, NOS3*), transcriptional regulation of developmental and metabolic pathways (*IRX1*), and immune function (*SH2B3, C6orf15/HLA*).

Fine-mapping and eQTL analyses refined candidate causal variants and highlighted tissue-specific regulatory effects, particularly in tibial arteries and skeletal muscle. Sex-specific analyses revealed stronger associations in females, consistent with the higher prevalence of RS in women, while shared loci across sexes suggest common biological mechanisms.

Strengths of this study include the scale of the analysis, combining eight large cohorts to achieve high statistical power and identify nine novel associations for RS. Replication of previously reported associations, along with sex-specific analyses and fine-mapping, enhanced the robustness of the results. However, limitations exists, such as the predominance of European ancestry among participants and potential phenotype misclassification, which may limit the generalizability of the results. Addressing these limitations through increased population diversity and integration of multi-omic data could provide a more comprehensive understanding of the underlying biology of RS.

Future analyses could incorporate colocalization approaches to further validate the relationship between eQTL findings and GWAS signals. In addition, LD score regression (LDSC) may be employed to estimate genetic correlations between RS and related conditions, such as migraine, autoimmune diseases, or other vascular disorders, potentially revealing shared biological pathways.

References

[1] Herrick, A. L. 2019. Raynauds phenomenon. *Journal of Scleroderma and Related Disorders*, 4(2), pp. 89–101. https://doi.org/10.1177/2397198319826467.

[2] Nawaz, I., Nawaz, Y., Nawaz, E., Manan, M. R. and Mahmood, A. 2022. Raynauds phenomenon: Reviewing the pathophysiology and management strategies. *Cureus*, 14(1). http://dx.doi.org/10.7759/cureus.21681.

[3] Tervi, A., Ramste, M., Abner, E., Cheng, P., Lane, J. M. et al. 2024. Genetic and functional analysis of Raynauds syndrome implicates loci in vasculature and immunity. *Cell genomics*, 4(9). https://doi.org/10.1016/j.xgen.2024.100630.

[4] Koenig, M., Joyal, F., Fritzler, M. J., Roussin, A., Abrahamowicz, M. et al. 2008. Autoantibodies and microvascular damage are independent predictive factors for the progression of Raynaud's phenomenon to systemic sclerosis: A twenty-year prospective study of 586 patients, with validation of proposed criteria for early systemic sclerosis. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, 58(12), pp. 3902–3912. https://doi.org/10.1002/art.24038.

[5] Barbacki, A., Rached-d'Astous, N., Pineau, C. A., Vinet, E., Grenier, L.-P. et al. 2022. Clinical significance of raynaud phenomenon in systemic lupus erythematosus. *JCR: Journal of Clinical Rheumatology*, 28(2), pp. e488–e490. https://doi.org/10.1097/rhu.0000000000001773.

[6] Giuggioli, D., Spinella, A., de Pinto, M., Mascia, M. T. and Salvarani, C. 2022. From Raynaud phenomenon to systemic sclerosis in COVID-19: A case report. *Advances in Skin & Wound Care*, 35(2), pp. 123–124. https://doi.org/10.1097/01.asw.0000795240.63966.53.

[7] Bush, W. S. and Moore, J. H. 2012. Chapter 11: Genome-wide association studies. In *Translational Bioinformatics*, Lewitter, F. (ed.), volume 8. PLOS Computational Biology. https://doi.org/10.1371/journal.pcbi.1002822.

[8] Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y. et al. 2021. Genome-wide association studies. *Nature Reviews Methods Primers*, 1(59). https://doi.org/10.1038/s43586-021-00056-9.

[9] Pritchard, J. 2023. *An owner's guide to the human genome*. Stanford University. https://web.stanford.edu/group/pritchardlab/HGbook.html.

[10] Abdellaoui, A., Yengo, L., Verweij, K. J. and Visscher, P. M. 2023. 15 years of GWAS discovery: Realizing the promise. *The American Journal of Human Genetics*, 110 (2), pp. 179–194. https://doi.org/10.1016/j.ajhg.2022.12.011.

[11] Duncan, L. E., Ostacher, M. and Ballon, J. 2019. How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. *Neuropsychopharmacology*, 44(9), pp. 1518–1523. https://doi.org/10.1038/s41386-019-0389-5.

[12] Schaid, D. J., Chen, W. and Larson, N. B. 2018. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics*, 19 (8), pp. 491–504. https://doi.org/10.1038/s41576-018-0016-z.

[13] Tak, Y. G. and Farnham, P. J. 2015. Making sense of GWAS: Using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics & chromatin*, 8(57), pp. 1–18. https://doi.org/10.1186/s13072-015-0050-4.

[14] Evangelou, E. and Ioannidis, J. P. 2013. Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*, 14(6), pp. 379–389. https://doi.org/10.1038/nrg3472.

[15] Willer, C. J., Li, Y. and Abecasis, G. R. 2010. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17), pp. 2190–2191. https://doi.org/10.1093/bioinformatics/btq340.

[16] Carlson, R. B., Martin, J. R. and Beckett, R. D. 2023. Ten simple rules for interpreting and evaluating a meta-analysis. *PLOS Computational Biology*, 19(9). https://doi.org/10.1371/journal.pcbi.1011461.

[17] Neupane, B., Loeb, M., Anand, S. S. and Beyene, J. 2012. Meta-analysis of genetic association studies under heterogeneity. *European journal of human genetics*, 20(11), pp. 1174–1181. https://doi.org/10.1038/ejhg.2012.75.

[18] Wang, J., Yu, J., Lipka, A. E. and Zhang, Z. 2022. Interpretation of manhattan plots and other outputs of genome-wide association studies. In *Genome-Wide Association Studies*, Torkamaneh, D. and Belzile, F. (editors), volume 2481, pp. 63–80. Springer, New York, NY. https://doi.org/10.1007/978-1-0716-2237-7_5.

[19] Wang, G., Sarkar, A., Carbonetto, P. and Stephens, M. 2020. A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(5), pp. 1273–1300. https://doi.org/10.1111/rssb.12388.

[20] Gilad, Y., Rifkin, S. A. and Pritchard, J. K. 2008. Revealing the architecture of gene regulation: The promise of eQTL studies. *Trends in genetics*, 24(8), pp. 408–415. https://doi.org/10.1016/j.tig.2008.06.001.

[21] GTEx Consortium. 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, 369(6509), pp. 1318–1330. https://www.science.org/doi/abs/10.1126/science.aaz1776.

[22] Watanabe, K., Taskesen, E., Van Bochoven, A. and Posthuma, D. 2017. Functional mapping and annotation of genetic associations with FUMA. *Nature communications*, 8(1826). https://doi.org/10.1038/s41467-017-01261-5.

[23] Kircher, M., Witten, D. M., Jain, P., O'roak, B. J., Cooper, G. M. and Shendure, J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics*, 46(3), pp. 310–315. https://doi.org/10.1038/ng.2892.

[24] Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A. et al. 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Genome research*, 22(9), pp. 1790–1797. https://doi.org/10.1101/gr.137323.112.

[25] De Leeuw, C. A., Mooij, J. M., Heskes, T. and Posthuma, D. 2015. MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology*, 11(4). https://doi.org/10.1371/journal.pcbi.1004219.

[26] Boughton, A. P., Welch, R. P., Flickinger, M., VandeHaar, P., Taliun, D. et al. 2021. LocusZoom.js: Interactive and embeddable visualization of genetic association study results. *Bioinformatics*, 37(18), pp. 3017–3018. https://doi.org/10.1093/bioinformatics/btab186.

[27] Juliusdottir, T. 2023. topr: An R package for viewing and annotating genetic association results. *BMC bioinformatics*, 24(268). https://doi.org/10.1186/s12859-023-05301-4.

[28] Huang, D., Zhou, Y., Yi, X., Fan, X., Wang, J. et al. 2022. VannoPortal: Multiscale functional annotation of human genetic variants for interrogating molecular mechanism of traits and diseases. *Nucleic acids research*, 50(D1), pp. D1408–D1416. https://doi.org/10.1093/nar/gkab853.

[29] Hartmann, S., Yasmeen, S., Jacobs, B. M., Denaxas, S., Pirmohamed, M. et al. 2023. ADRA2A and IRX1 are putative risk genes for Raynauds phenomenon. *Nature communications*, 14(6156). https://doi.org/10.1038/s41467-023-41876-5.

[30] Lewis, C. M. and Vassos, E. 2020. Polygenic risk scores: From research tools to clinical instruments. *Genome medicine*, 12(44). https://doi.org/10.1186/s13073-020-00742-5.

[31] Saurabh, R., Fouodo, C. J., König, I. R., Busch, H. and Wohlers, I. 2022. A survey of genome-wide association studies, polygenic scores and UK Biobank highlights

resources for autoimmune disease genetics. *Frontiers in immunology*, 13. https://doi.org/10.3389/fimmu.2022.972107.

[32] Bayoumy, K., Gaber, M., Elshafeey, A., Mhaimeed, O., Dineen, E. H. et al. 2021. Smart wearable devices in cardiovascular care: Where we are and how to move forward. *Nature Reviews Cardiology*, 18(8), pp. 581–599. https://doi.org/10.1038/s41569-021-00522-7.

[33] Kurki, M. I., Karjalainen, J., Palta, P., Sipilä, T. P., Kristiansson, K. et al. 2023. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*, 613(7944), pp. 508–518. https://doi.org/10.1038/s41586-022-05473-8.

[34] FinnGen. Documentation of R13 release, 2025. https://finngen.gitbook.io/documentation/. Accessed: March 22, 2025.

[35] Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T. et al. 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726), pp. 203–209. https://doi.org/10.1038/s41586-018-0579-z.

[36] Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.-L., Alavere, H. et al. 2015. Cohort profile: Estonian biobank of the Estonian genome center, University of Tartu. *International journal of epidemiology*, 44(4), pp. 1137–1147. https://doi.org/10.1093/ije/dyt268.

[37] Milani, L., Alver, M., Laur, S., Reisberg, S., Haller, T. et al. 2024. From biobanking to personalized medicine: The journey of the Estonian Biobank. *medRxiv*. https://doi.org/10.1101/2024.09.22.24313964.

[38] Boutin, N. T., Schecter, S. B., Perez, E. F., Tchamitchian, N. S., Cerretani, X. R. et al. 2022. The evolution of a large biobank at Mass General Brigham. *Journal of Personalized Medicine*, 12(8). https://doi.org/10.3390/jpm12081323.

[39] Verma, A., Huffman, J. E., Rodriguez, A., Conery, M., Liu, M. et al. 2024. Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science*, 385(6706). https://www.science.org/doi/10.1126/science.adj1182.

[40] Finer, S., Martin, H. C., Khan, A., Hunt, K. A., MacLaughlin, B. et al. 2020. Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *International journal of epidemiology*, 49(1), pp. 20–21. https://doi.org/10.1093/ije/dyz174.

[41] Genes & Health. Health data, 2025. https://www.genesandhealth.org/researchers/data/. Accessed: May 23, 2025.

[42] Shuey, M. M., Stead, W. W., Aka, I., Barnado, A. L., Bastarache, J. A. et al. 2023. Next-generation phenotyping: Introducing phecodeX for enhanced discovery research in medical phenomics. *Bioinformatics*, 39(11). https://doi.org/10.1093/bioinformatics/btad655.

[43] Bick, A. G., Metcalf, G. A., Mayo, K. R., Lichtenstein, L., Rura, S. et al. 2024. Genomic data in the All of Us Research Program. *Nature*, 627, pp. 340–346. https://doi.org/10.1038/s41586-023-06957-x.

[44] University of Michigan. Michigan Genomics Initiative freeze 6 genome-wide genotypes v1.1, 2023. https://sites.lsa.umich.edu/phanalyticsdocumentation/wp-content/uploads/sites/1044/2023/09/freeze6_white_paper_v_1_1.pdf. Accessed: May 23, 2025.

[45] Zawistowski, M., Fritsche, L. G., Pandit, A., Vanderwerff, B., Patil, S. et al. 2023. The Michigan Genomics Initiative: A biobank linking genotypes and electronic clinical records in Michigan Medicine patients. *Cell Genomics*, 3(2). https://doi.org/10.1016/j.xgen.2023.100257.

[46] Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J. A. et al. 2021. Computationally efficient whole-genome regression for quantitative and binary traits. *Nature genetics*, 53(7), pp. 1097–1103. https://doi.org/10.1038/s41588-021-00870-7.

[47] Benner, C., Spencer, C. C., Havulinna, A. S., Salomaa, V., Ripatti, S. and Pirinen, M. 2016. FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics*, 32(10), pp. 1493–1501. https://doi.org/10.1093/bioinformatics/btw018.

[48] Benner, C., Havulinna, A. S., Salomaa, V., Ripatti, S. and Pirinen, M. 2018. Refining fine-mapping: Effect sizes and regional heritability. *BioRxiv*, pp. 318–618. https://doi.org/10.1101/318618.

[49] Benner, C., Havulinna, A. S., Järvelin, M.-R., Salomaa, V., Ripatti, S. and Pirinen, M. 2017. Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *The American Journal of Human Genetics*, 101(4), pp. 539–551. https://doi.org/10.1016/j.ajhg.2017.08.012.

[50] Schneider, V. A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C. et al. 2017. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome research*, 27(5), pp. 849–864. https://doi.org/10.1101/gr.213611.116.

[51] Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), pp. 2987–2993. https://doi.org/10.1093/bioinformatics/btr509.

[52] Devlin, B. and Roeder, K. 1999. Genomic control for association studies. *Biometrics*, 55(4), pp. 997–1004. https://doi.org/10.1111/j.0006-341X.1999.00997.x.

[53] Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S. et al. 2010. LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics*, 26(18), pp. 2336–2337. https://doi.org/10.1093/bioinformatics/btq419.

[54] Kurokawa, Y., Matoba, R., Takemasa, I., Nagano, H., Dono, K. et al. 2004. Molecular-based prediction of early recurrence in hepatocellular carcinoma. *Journal of hepatology*, 41(2), pp. 284–291. https://doi.org/10.1016/j.jhep.2004.04.031.

[55] Matzaraki, V., Kumar, V., Wijmenga, C. and Zhernakova, A. 2017. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome biology*, 18(76). https://doi.org/10.1186/s13059-017-1207-1.

[56] Chotani, M. A., Flavahan, S., Mitra, S., Daunt, D. and Flavahan, N. A. 2000. Silent α2C-adrenergic receptors enable cold-induced vasoconstriction in cutaneous arteries. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(4), pp. H1075–H1083. https://doi.org/10.1152/ajpheart.2000.278.4.H1075.

[57] Küster, M. M., Schneider, M. A., Richter, A. M., Richtmann, S., Winter, H. et al. 2020. Epigenetic inactivation of the tumor suppressor IRX1 occurs frequently in lung adenocarcinoma and its silencing is associated with impaired prognosis. *Cancers*, 12 (12). https://doi.org/10.3390/cancers12123528.

[58] Zülch, A., Becker, M.-B. and Gruss, P. 2001. Expression pattern of IRX1 and IRX2 during mouse digit development. *Mechanisms of development*, 106(1-2), pp. 159–162. https://doi.org/10.1016/S0925-4773(01)00411-7.

[59] Tervi, A., Ramste, M. and Ollila, H. M. 2025. Response to Dr. Nicholas A. Flavahans and Dr. Ali H. Eids letters regarding our recent publication in Cell Genomics. *Cell Genomics*, 5(5). https://doi.org/10.1016/j.xgen.2025.100885.

[60] Flavahan, N. A. 2015. A vascular mechanistic approach to understanding Raynaud phenomenon. *Nature Reviews Rheumatology*, 11(3), pp. 146–158. https://doi.org/10.1038/nrrheum.2014.195.

[61] De Bakker, P. I., McVean, G., Sabeti, P. C., Miretti, M. M., Green, T. et al. 2006. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nature genetics*, 38(10), pp. 1166–1172. https://doi.org/10.1038/ng1885.

[62] Klarin, D., Lynch, J., Aragam, K., Chaffin, M., Assimes, T. L. et al. 2019. Genome-wide association study of peripheral artery disease in the Million Veteran Program. *Nature medicine*, 25(8), pp. 1274–1279. https://doi.org/10.1038/s41591-019-0492-5.

[63] Ahmad, A., Dempsey, S. K., Daneva, Z., Azam, M., Li, N. et al. 2018. Role of nitric oxide in the cardiovascular and renal systems. *International journal of molecular sciences*, 19(9). https://doi.org/10.3390/ijms19092605.

[64] Salvi, E., Kutalik, Z., Glorioso, N., Benaglio, P., Frau, F. et al. 2012. Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of endothelial NO synthase. *Hypertension*, 59(2), pp. 248–255. https://doi.org/10.1161/HYPERTENSIONAHA.111.181990.

[65] Salvi, E., Kuznetsova, T., Thijs, L., Lupoli, S., Stolarz-Skrzypek, K. et al. 2013. Target sequencing, cell experiments, and a population study establish endothelial nitric oxide synthase (eNOS) gene as hypertension susceptibility gene. *Hypertension*, 62(5), pp. 844–852. https://doi.org/10.1161/HYPERTENSIONAHA.113.01428.

[66] Zhabin, S., Lazarenko, V., Azarova, I., Klyosova, E., Bashkatov, D. et al. 2024. The promoter polymorphism rs3918226 of the endothelial nitric oxide synthase gene as a novel susceptibility marker for peripheral artery disease. *Annals of Vascular Surgery*, 108, pp. 557–563. https://doi.org/10.1016/j.avsg.2024.05.037.

[67] Adefurin, A., Ghimire, L. V., Kohli, U., Muszkat, M., Sofowora, G. G. et al. 2017. Genetic variation in the alpha1B-adrenergic receptor and vascular response. *The pharmacogenomics journal*, 17(4), pp. 366–371. https://doi.org/10.1038/tpj.2016.29.

[68] Dale, B. L. and Madhur, M. S. 2016. Linking inflammation and hypertension via LNK/SH2B3. *Current opinion in nephrology and hypertension*, 25(2), pp. 87–93. https://doi.org/10.1097/mnh.0000000000000196.

[69] Devallière, J. and Charreau, B. 2011. The adaptor LNK (SH2B3): An emerging regulator in vascular cells and a link between immune and inflammatory signaling. *Biochemical pharmacology*, 82(10), pp. 1391–1402. https://doi.org/10.1016/j.bcp.2011.06.023.

[70] Hinds, D. A., Barnholt, K. E., Mesa, R. A., Kiefer, A. K., Do, C. B. et al. 2016. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. *Blood, The Journal of the American Society of Hematology*, 128(8), pp. 1121–1128. https://doi.org/10.1182/blood-2015-06-652941.

[71] Kong, A., Thorleifsson, G., Gudbjartsson, D. F., Masson, G., Sigurdsson, A. et al. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature*, 467(7319), pp. 1099–1103. https://doi.org/10.1038/nature09525.

[72] Fraenkel, L. 2002. Raynauds phenomenon: Epidemiology and risk factors. *Current rheumatology reports*, 4(2), pp. 123–128. https://doi.org/10.1007/s11926-002-0007-z.

A Appendix: Independent genetic loci identified in the GWAS meta-analysis of Raynaud's syndrome

GWAS meta-analysis of Raynaud's syndrome															
	SNP ID	CHROM	POS	REF	ALT	AAF	FreqSE	Effect	StdErr	P	Direction	HetISq	HetChiSq	HetPVal	Nearest gene
1	rs7075626	10	111,098,213	G	C	0.3329	0.0516	0.2018	0.0115	1.447e-68	++++++++	44.7	12.648	0.08116	ADRA2A
2	rs7706161	5	4,047,779	A	G	0.2733	0.0294	0.1518	0.012	7.073e-37	++++++++	14.8	8.213	0.3142	IRX1
3	rs3130968	6	31,097,294	C	T	0.1419	0.0276	0.1432	0.0179	1.21e-15	++++++?–	20.1	7.509	0.2763	C6orf15/HLA
4	rs3918226	7	150,993,088	C	T	0.0758	0.0071	0.1412	0.0204	4.95e-12	++++++++	25.3	9.374	0.2269	NOS3
5	rs526882	16	86,679,425	A	C	0.3513	0.0179	0.078	0.0113	5.848e-12	++++++++	0	2.295	0.9417	FOXL1
6	rs2209592	10	129,165,236	A	G	0.6722	0.0266	-0.074	0.0118	3.025e-10	−−−−−−−−	0	3.249	0.8611	MGMT
7	rs7310615	12	111,427,245	C	G	0.543	0.0485	-0.0695	0.0111	3.89e-10	−−−−−−−−	44.1	12.533	0.08432	SH2B3
8	rs13263870	8	50,583,718	T	C	0.735	0.0437	-0.0801	0.0128	3.952e-10	−−−−?−−−	13.6	6.941	0.3264	SNTG1
9	rs6556458	5	159,867,336	C	T	0.5799	0.0112	0.0725	0.0117	6.628e-10	++++++++	0	3.989	0.781	ADRA1B
10	rs17010957	4	85,798,012	T	C	0.1454	0.0123	0.0937	0.0155	1.579e-09	++++++++	0	3.717	0.8117	ARHGAP24
11	rs4150495	2	127,272,049	G	A	0.3584	0.0194	-0.0857	0.0144	2.388e-09	−−??−−?–	0	1.248	0.8702	ERCC3
12	rs10803364	1	15,191,937	A	G	0.4223	0.0445	-0.0662	0.0111	2.437e-09	−−−−−−−−	0	1.239	0.99	TMEM51
13	rs569240936	11	113,536,934	T	A	0.0106	0.0015	1.0168	0.1735	4.622e-09	+???+?–?	0	0.425	0.8086	DRD2
14	rs10305667	1	150859632	C	T	0.2599	0.0454	-0.0703	0.0126	2.127e-08	−−−−+−−−	0	5.548	0.5933	ARNT

	MVP				MGB				FG				EstBB			
	AAF	Effect	StdErr	P	AAF	Effect	StdErr	P	AAF	Effect	StdErr	P	AAF	Effect	StdErr	P
1	0.3153	0.240417	0.0339367	1.56E-12	0.361681	0.175913	0.0329482	1.26E-07	0.3111	0.223387	0.0257856	4.59E-18	0.268638	0.209441	0.0344256	1.17E-09
2	0.2889	0.206041	0.0330194	4.21E-10	0.278827	0.1543	0.0339173	6.89E-06	0.227263	0.158535	0.0285206	2.72E-08	0.234948	0.130628	0.0359236	0.000276624
3	0.1372	0.210844	0.0665918	0.00161	0.129022	0.218078	0.0431789	8.65E-07	0.101551	0.101566	0.0397346	0.0105847	0.123509	0.111677	0.069665	0.108920594
4	0.079	0.0910194	0.0625233	0.1451	0.075905	0.25109	0.0535218	5.49E-06	0.0701592	0.123572	0.0464888	0.00785833	0.065915	0.0574201	0.0621441	0.355494593
5	0.3408	0.0905814	0.0322804	0.004833	0.36829	0.0597912	0.033018	0.07016168	0.341321	0.0824498	0.0257102	0.00134181	0.371514	0.0538964	0.0317156	0.089250365
6	0.6635	-0.0372958	0.0367828	0.3134	0.669238	-0.0881615	0.0338998	0.009687462	0.652144	-0.0695237	0.0257508	0.00693681	0.728203	-0.0492028	0.0342479	0.150813077
7	0.5099	-0.103459	0.0308161	0.000838	0.537135	-0.0915016	0.032171	0.004484252	0.587249	-0.0408157	0.0253298	0.107099	0.553298	-0.0382994	0.0307399	0.212793325
8	0.6952	-0.0962189	0.0354985	0.006786	0.678617	-0.0177235	0.0348259	0.6108112	0.816706	-0.0759932	0.0322213	0.0183502	0.761803	-0.0400219	0.0360696	0.267181879
9	0.5509	0.137769	0.0382291	0.0002894	0.577805	0.0786428	0.034968	0.024336605	0.582319	0.0653138	0.0253752	0.0100554	0.583222	0.06877	0.0325457	0.034598717
10	0.1416	0.0782859	0.0459863	0.08884	0.137023	0.0504868	0.0464399	0.27697336	0.152489	0.0801988	0.0346948	0.0208022	0.125474	0.118178	0.0467599	0.011493439
11	0.3642	-0.108854	0.0330802	0.0009628	0.362374	-0.0775227	0.0330595	0.018715007	NA	NA	NA	NA	NA	NA	NA	NA
12	0.3947	-0.0554347	0.0325868	0.08861	0.424265	-0.0820979	0.0322314	0.010759944	0.487598	-0.068358	0.0246179	0.00549035	0.388425	-0.0733536	0.0313673	0.019359316
13	6.00E-04	0.297328	1.14601	0.7953	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
14	0.2407	-0.07139	0.0366276	0.05114	0.256822	-0.0501495	0.0361923	0.165857074	0.274032	-0.056147	0.0278317	0.0436556	0.251747	-0.124982	0.0351828	0.000381812

	G&N				UKB				AoU				MGI			
	AAF	Effect	StdErr	P	AAF	Effect	StdErr	P	AAF	Effect	StdErr	P	AAF	Effect	StdErr	P
1	0.560344	0.0273298	0.0599898	0.648696	0.333002	0.213501	0.0208616	1.39E-24	0.31929	0.2143	0.051268	2.92E-05	0.357619	0.157028	0.0396931	7.62E-05
2	0.247978	0.094217	0.0686808	0.170123	0.303129	0.166783	0.0211494	3.12E-15	0.29171	0.074696	0.052435	0.15429	0.273362	0.0943385	0.0416925	0.02365283
3	0.00723032	0.172962	0.333034	0.603515	0.168706	0.138661	0.0257729	7.44E-08	NA	NA	NA	NA	0.129769	-0.000274568	0.0948277	0.9976898
4	0.00951478	0.299834	0.309579	0.332784	0.0812126	0.124545	0.0356411	0.000475084	0.084066	0.26741	0.085612	0.0017868	0.0737948	0.13024	0.0708213	0.06591565
5	0.406234	0.0605625	0.0604403	0.316333	0.334992	0.0808787	0.0209106	0.000109812	0.34935	0.13126	0.049785	0.0083756	0.367225	0.0768635	0.0391217	0.04944582
6	0.599294	-0.0341478	0.0611859	0.576777	0.680996	-0.0845956	0.0212988	7.13E-05	0.67934	-0.10887	0.051028	0.032876	0.653197	-0.101238	0.04061	0.012669
7	0.904772	-0.231909	0.101643	0.0225123	0.516622	-0.0623776	0.0200607	0.001874347	0.50793	-0.0025785	0.047552	0.95676	0.550716	-0.146645	0.0384489	0.000136721
8	NA	NA	NA	NA	0.738256	-0.114421	0.0223248	2.97E-07	0.73104	-0.094147	0.053564	0.078808	0.690016	-0.0827573	0.0418692	0.04809063
9	0.599717	0.0638286	0.0610856	0.296067	0.586702	0.0597869	0.0211972	0.004794795	0.58458	0.098487	0.048124	0.040705	0.564292	0.0438851	0.0433241	0.3110843
10	0.190547	0.189249	0.0765498	0.013427	0.148509	0.107883	0.0276143	9.35E-05	0.14174	0.11435	0.068563	0.095342	0.140199	0.0578728	0.0539607	0.2834966
11	0.27128	-0.129077	0.0669703	0.0539324	0.359844	-0.0814241	0.0209044	9.82E-05	NA	NA	NA	NA	0.370082	-0.0657788	0.0388758	0.0906417
12	0.558448	-0.0859069	0.0603596	0.154663	0.392109	-0.0691821	0.0205221	0.000748669	0.39826	-0.043134	0.048417	0.37299	0.416059	-0.0396747	0.0380605	0.297221
13	0.0108398	1.03353	0.175469	3.31E-07	NA	NA	NA	NA	1.01E-05	-1.0084	10.084	0.92035	NA	NA	NA	NA
14	0.462832	0.00537592	0.0599304	0.928524	0.238773	-0.0617625	0.0236171	0.008918463	0.24137	-0.11768	0.055392	0.033623	0.242446	-0.0864797	0.0441229	0.04999923

Table A1: GWAS meta-analysis lead variants for Raynaud's syndrome, combining FinnGen R13, the UK Biobank, the Estonian Biobank, the Mass General Brigham Biobank, the Million Veteran Program, Genes & Health, the All of Us Research Program, and the Michigan Genomics Initiative data (17,793 cases and 1,850,553 controls). Meta-analysis was performed using METAL [15].

SNP ID: Variant ID; **CHROM**: Chromosome number; **POS**: Base-pair position (b38); **REF**: Reference allele; **ALT**: Alternative allele; **AAF**: Alternative allele frequency; **FreqSE**: Standard error of AAF; **Effect**: Estimated effect size; **StdErr**: Standard error of Effect; **P**: P-value; **Direction**: Direction of Effect across cohorts; **HetISq**: Heterogeneity I-squared statistic; **HetChiSq**: Heterogeneity chi-square statistic; **HetPval**: P-value for heterogeneity; **Nearest gene**: Nearest gene to the variant.

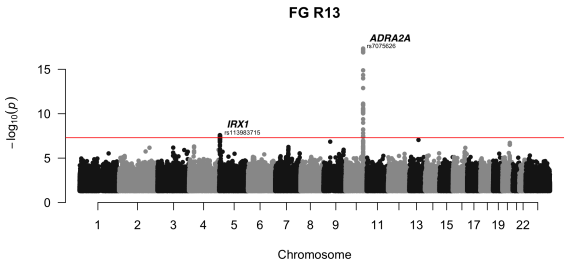
B Appendix: Independent genetic loci identified in the sample size-weighted GWAS meta-analysis of Raynaud's syndrome

Sample size-weighted GWAS meta-analysis of Raynaud's syndrome															
	SNP ID	**CHROM**	**POS**	**REF**	**ALT**	**AAF**	**FreqSE**	**Weight**	**Zscore**	**P**	**Direction**	**HetISq**	**HetChiSq**	**HetPVal**	**Nearest gene**
1	rs7075626	10	111,098,213	G	C	0.3219	0.0431	1804692	17.321	3.246e-67	++++++++	61.3	18.089	0.01158	ADRA2A
2	rs7706161	5	4,047,779	A	G	0.2675	0.0321	1804692	-12.593	2.298e-36	++++++++	31.1	10.166	0.1794	IRX1
3	rs3130968	6	31,097,294	C	T	0.1297	0.0317	1751395	7.041	1.906e-12	++++++?–	70.9	20.636	0.002132	C6orf15/HLA
4	rs526882	16	86,679,425	A	C	0.3464	0.0153	1804692	-6.74	1.582e-11	++++++++	0	4.308	0.7436	FOXL1
5	rs6556458	5	159,867,336	C	T	0.5755	0.0149	1804692	6.253	4.022e-10	++++++++	0	3.172	0.8687	ADRA1B
6	rs13263870	8	50,583,718	T	C	0.7488	0.0485	1760918	6.211	5.265e-10	––––?–––	19.8	7.483	0.2785	SNTG1
7	rs4150495	2	127,272,049	G	A	0.3587	0.0185	1045497	-5.909	3.452e-09	?––?––?–	0	2.061	0.7245	ERCC3
8	rs7310615	12	111,427,245	C	G	0.5495	0.0646	1804692	5.897	3.708e-09	––––––––	58.3	16.807	0.01869	SH2B3
9	rs3918226	7	150,993,088	C	T	0.0738	0.0116	1804692	5.873	4.284e-09	++++++++	67.1	21.267	0.003395	NOS3
10	rs17010957	4	85,798,012	T	C	0.1455	0.0109	1804692	-5.818	5.948e-09	++++++++	0	6.296	0.5056	ARHGAP24
11	rs7601684	2	129,324,037	G	A	0.0066	0.0211	1804692	5.704	1.171e-08	++++++–+	62.3	18.564	0.009668	PPIAP65
12	rs10803364	1	15,191,937	A	G	0.4247	0.0468	1804692	5.682	1.335e-08	––––––––	0	4.57	0.7123	TMEM51
13	rs10861399	12	105,336,497	C	T	0.2894	0.0132	1804692	5.627	1.839e-08	++–+–+++	78.1	32.036	4e-05	C12orf75
14	rs2209592	10	129,165,236	A	G	0.6708	0.0254	1804692	5.625	1.85e-08	––––––––	37.4	11.179	0.131	MGMT
15	rs7125196	11	61,505,093	T	C	0.1411	0.0339	1804692	5.487	4.092e-08	––+–––––	0	4.374	0.7359	PPP1R32

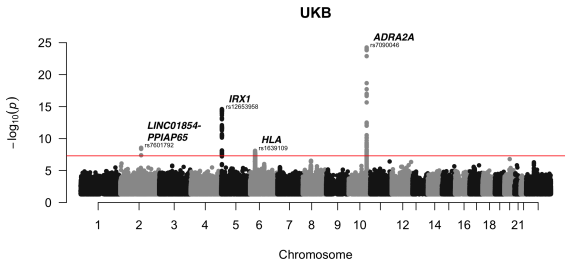
Table B1: Sample size-weighted GWAS meta-analysis lead variants for Raynaud's syndrome, combining FinnGen R13, the UK Biobank, the Estonian Biobank, the Mass General Brigham Biobank, the Million Veteran Program, Genes & Health, the All of Us Research Program, and the Michigan Genomics Initiative data. Meta-analysis was performed using METAL [15].

SNP ID: Variant ID; **CHROM**: Chromosome number; **POS**: Base-pair position (b38); **REF**: Reference allele; **ALT**: Alternative allele; **AAF**: Alternative allele frequency; **FreqSE**: Standard error of AAF; **Weight**: Sample size; **Zscore**: Combined z-score; **P**: P-value; **Direction**: Direction of Effect across cohorts; **HetISq**: Heterogeneity I-squared statistic; **HetChiSq**: Heterogeneity chi-square statistic; **HetPval**: P-value for heterogeneity; **Nearest gene**: Nearest gene to the variant.

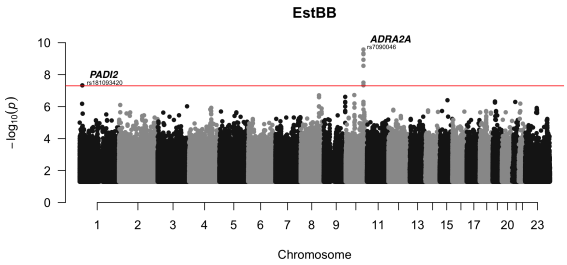
C Appendix: Manhattan plots of GWAS summary statistics from all cohorts



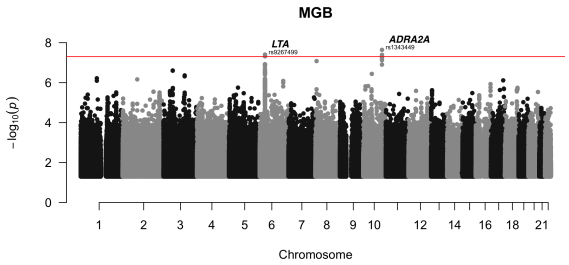
(a) FinnGen R13



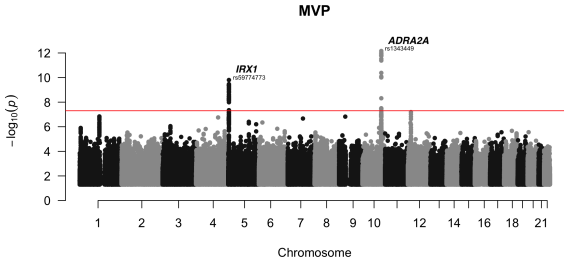
(b) UK Biobank



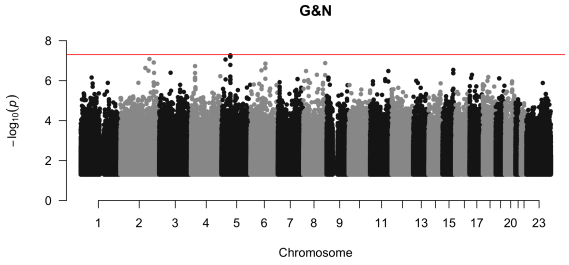
(c) Estonian Biobank



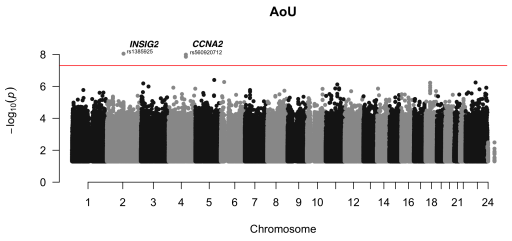
(d) Mass General Brigham Biobank



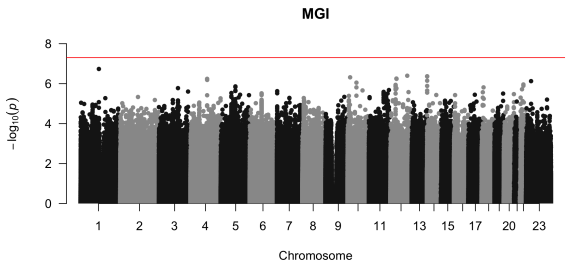
(e) Million Veteran Program



(f) Genes & Health



(g) All of Us Research Program



(h) Michigan Genomics Initiative

Figure C1: Manhattan plots of individual genome-wide association study (GWAS) summary statistics. Each subfigure (ah) represents results from a distinct cohort. Top variants exceeding the genome-wide significance ($p < 5 \times 10^{-8}$) are annotated with the corresponding single-nucleotide polymorphism (SNP) ID and nearest gene.

D Appendix: GWAS lead variants from all cohorts

FinnGen R13 top 5 variants (3,278 cases and 497,070 controls)									
SNP	**CHR**	**POS (b38)**	**Ref.**	**Alt.**	**Alt.AF**	**p-value**	**Beta**	**SE**	**Nearest gene**
rs7075626	10	111,098,213	G	C	0.3111	4.58564e-18	0.223387	0.0257856	*ADRA2A*
rs7090046	10	111,101,172	G	A	0.247112	5.813e-18	0.23614	0.0273432	*ADRA2A*
rs1343451	10	111,100,633	A	G	0.247725	7.08435e-18	0.235094	0.0272936	*ADRA2A*
rs7908645	10	111,096,667	T	G	0.322353	1.04665e-17	0.219372	0.0256016	*ADRA2A*
rs1343449	10	111,100,768	A	G	0.247926	1.12331e-17	0.233566	0.0272841	*ADRA2A*

Table D1: The FinnGen R13 top 5 variants.

FinnGen R13 females top 5 variants (2,368 cases and 279,696 controls)									
SNP	**CHR**	**POS (b38)**	**Ref.**	**Alt.**	**Alt.AF**	**p-value**	**Beta**	**SE**	**Nearest gene**
rs7075619	10	111,098,192	G	T	0.215259	1.3627e-12	0.237818	0.0335531	*ADRA2A*
rs7075626	10	111,098,213	G	C	0.310556	2.07205e-12	0.213004	0.0303012	*ADRA2A*
rs7908645	10	111,096,667	T	G	0.321707	3.29306e-12	0.209545	0.030087	*ADRA2A*
rs61862975	10	111,098,529	C	T	0.209708	8.43529e-12	0.231781	0.033931	*ADRA2A*
rs7090046	10	111,101,172	G	A	0.246645	5.13807e-11	0.211871	0.0322637	*ADRA2A*

Table D2: The FinnGen R13 top 5 variants for females.

FinnGen R13 males top 5 variants (910 cases and 217,374 controls)									
SNP	**CHR**	**POS (b38)**	**Ref.**	**Alt.**	**Alt.AF**	**p-value**	**Beta**	**SE**	**Nearest gene**
rs1343451	10	111,100,633	A	G	0.24831	2.02283e-08	0.280758	0.0500457	*ADRA2A*
rs7090046	10	111,101,172	G	A	0.247716	2.03606e-08	0.28119	0.0501327	*ADRA2A*
rs1343449	10	111,100,768	A	G	0.248488	2.3093e-08	0.279543	0.050034	*ADRA2A*
rs7084501	10	111,099,941	A	G	0.249035	2.65112e-08	0.278614	0.0500829	*ADRA2A*
rs146525259	17	43,645,760	C	T	0.000436363	6.52259e-08	2.69401	0.498535	*MEOX1*

Table D3: The FinnGen R13 top 5 variants for males.

UKB top 5 variants (5,152 cases and 439,974 controls)									
SNP	**CHR**	**POS (b37)**	**Ref.**	**Alt.**	**Alt.AF**	**p-value**	**Beta**	**SE**	**Nearest gene**
rs7090046	10	112,860,930	G	A	0.310478	5.58727e-25	0.216857	0.0210086	*ADRA2A*
rs7084501	10	112,859,699	A	G	0.314374	8.64570e-25	0.215686	0.0209805	*ADRA2A*
rs1343449	10	112,860,526	A	G	0.31589	1.26503e-24	0.213614	0.0208535	*ADRA2A*
rs7075626	10	112,857,971	G	C	0.333002	1.39380e-24	0.213501	0.0208616	*ADRA2A*
rs1343451	10	112,860,391	A	G	0.314425	1.49865e-24	0.213717	0.020897	*ADRA2A*

Table D4: The UK Biobank (UKB) top 5 variants.

EstBB top 5 variants (2,205 cases and 203,345 controls)									
SNP	**CHR**	**POS (b38)**	**Ref.**	**Alt.**	**Alt.AF**	**p-value**	**Beta**	**SE**	**Nearest gene**
rs7090046	10	111,101,172	G	A	0.234032	2.68380e-10	0.226733	0.0358981	*ADRA2A*
rs1343449	10	111,100,768	A	G	0.234627	4.72291e-10	0.222923	0.0357934	*ADRA2A*
rs7084501	10	111,099,941	A	G	0.238634	4.95176e-10	0.222154	0.0357125	*ADRA2A*
rs1343451	10	111,100,633	A	G	0.234809	5.25255e-10	0.222123	0.0357607	*ADRA2A*
rs7075626	10	111,098,213	G	C	0.268638	1.17303e-09	0.209441	0.0344256	*ADRA2A*

Table D5: The Estonian Biobank (EstBB) top 5 variants.

MGB top 5 variants (1,907 cases and 54,452 controls)									
SNP	**CHR**	**POS (b38)**	**Ref.**	**Alt.**	**Alt.AF**	**p-value**	**Beta**	**SE**	**Nearest gene**
rs1343449	10	111,100,768	A	G	0.328876	2.28881e-08	0.185572	0.0328098	*ADRA2A*
rs9267499	6	31,569,519	G	C	0.105365	3.95804e-08	-0.297404	0.0559452	*LTA*
rs1343451	10	111,100,633	A	G	0.334053	4.05630e-08	0.18127	0.0326456	*ADRA2A*
rs4947324	6	31,560,353	C	T	0.105287	4.82248e-08	-0.295356	0.0559053	*LTA*
rs7084501	10	111,099,941	A	G	0.331393	4.96661e-08	0.181331	0.0328741	*ADRA2A*

Table D6: The Mass General Brigham (MGB) Biobank top 5 variants.

MVP top 5 variants (2,156 cases and 447,847 controls)									
SNP	**CHR**	**POS (b38)**	**Ref.**	**Alt.**	**Alt.AF**	**p-value**	**Beta**	**SE**	**Nearest gene**
rs1343449	10	111,100,768	A	G	0.30230	6.905e-13	0.232689	0.0322929	*ADRA2A*
rs745557	10	111,086,540	G	A	0.71080	9.868e-13	-0.256965	0.0359617	*ADRA2A*
rs7908645	10	111,096,667	T	G	0.31480	1.108e-12	0.242836	0.0340913	*ADRA2A*
rs7075619	10	111,098,192	G	T	0.23460	1.377e-12	0.264356	0.0370694	*ADRA2A*
rs7075626	10	111,098,213	G	C	0.31530	1.555e-12	0.240417	0.0339367	*ADRA2A*

Table D7: The Million Veteran Program (MVP) top 5 variants.

G&H top 5 variants (665 cases and 43,109 controls)									
SNP	**CHR**	**POS (b38)**	**Ref.**	**Alt.**	**Alt.AF**	**p-value**	**Beta**	**SE**	**Nearest gene**
rs1296028096	5	55,265,085	A	C	0.000514004	5.3853e-08	2.9355	0.423459	*DHX29*
rs543944605	5	55,417,368	G	C	0.000502581	5.50123e-08	2.93223	0.423435	*SKIV2L2*
rs556289815	5	55,362,446	C	T	0.000514004	5.88613e-08	2.92033	0.422273	*SKIV2L2*
rs1407765436	5	55,365,582	C	G	0.000514004	5.88613e-08	2.92033	0.422273	*SKIV2L2*
rs551557728	5	55,171,459	G	A	0.000525426	6.31059e-08	2.90888	0.421465	*CDC20B*

Table D8: The Genes&Health (G&H) top 5 variants.

AoU top 5 variants (901 cases and 98,010 controls)									
SNP	**CHR**	**POS (b38)**	**Ref.**	**Alt.**	**Alt.AF**	**p-value**	**Beta**	**SE**	**Nearest gene**
rs1385925	2	118,051,875	C	T	7.7766e-01	8.9195e-09	-3.3388e-01	5.8066e-02	*INSIG2*
rs560920712	4	121,820,220	C	T	5.9650e-04	1.0256e-08	8.5668e+00	1.4960e+00	*CCNA2*
rs535666815	4	121,944,783	T	C	6.2177e-04	1.3842e-08	8.3098e+00	1.4642e+00	*TRPC3*
rs191321845	5	133,741,616	C	T	5.5606e-04	3.9447e-07	7.7998e+00	1.5379e+00	*C5orf15*
rs142409243	6	23,401,014	A	G	1.2638e-04	5.3039e-07	2.2217e+01	4.4301e+00	*AL139231.1*

Table D9: The All of Us (AoU) Research Program top 5 variants.

MGI top 5 variants (1,529 cases and 66,746 controls)									
SNP	**CHR**	**POS (b38)**	**Ref.**	**Alt.**	**Alt.AF**	**p-value**	**Beta**	**SE**	**Nearest gene**
rs111892769	1	114,691,937	A	T	0.0108889	1.845213E-07	0.781429	0.149863	*AMPD1*
rs10861399	12	105,336,497	C	T	0.279772	4.020403E-07	0.203107	0.0400764	*C12orf75*
rs141891028	14	20,840,739	C	A	0.00128458	4.275466E-07	1.61327	0.319063	*RNASE1*
rs147957889	14	20,842,577	C	T	0.00128802	4.359339E-07	1.61234	0.319113	*RNASE1*
rs72783829	10	15,979,521	T	A	0.0449835	4.829047E-07	0.436322	0.0866926	*MINDY3*

Table D10: The Michigan Genomics Initiative (MGI) top 5 variants.