Genome-wide association study identifies *ADRA2A* and *IRX1* as novel risk genes for Raynaud's phenomenon

Sylvia Hartmann[1], Summaira Yasmeen[1], Spiros Denaxas[2,3,4,5], Harry Hemingway[2,3,5], Maik Pietzner[1,6,*], Claudia Langenberg[1,6,7,*]

[1]Computational Medicine, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany
[2]Institute of Health Informatics, University College London, London, UK
[3]Health Data Research UK, London, UK
[4]British Heart Foundation Data Science Centre, London, UK
[5]National Institute of Health Research University College London Hospitals Biomedical Research Centre
[6]MRC Epidemiology Unit, University of Cambridge, Cambridge, UK
[7]Precision Healthcare Institute, Queen Mary University of London, London, UK
*these authors jointly supervised the work

Correspondence

Prof Claudia Langenberg (claudia.langenberg@bih-charite.de)
Dr Maik Pietzner (maik.pietzner@bih-charite.de)

Abstract

Background: Raynaud's phenomenon (RP) is a common vasospastic disorder that causes severe pain and ulcers in fingers and toes triggered by cold or emotional stress. Despite its high reported heritability, no causal genes have been robustly identified, limiting mechanistic understanding and treatment options.

Aim: To investigate the genetic architecture of RP to better understand its aetiology and identify new potential therapeutic targets.

Methods: We conducted a genome-wide association study including 9,084 RP cases and 435,357 controls, based on diagnoses from electronic health records, among participants of the UK Biobank study. We identified candidate causal variants and genes using Bayesian fine-mapping and colocalization with gene expression across 49 tissues. We performed phenome-wide association analyses for all significant RP loci and computed genetic correlations followed by latent causal variable analyses between RP and a total of 205 selected phenotypes.

Results: We identified eight unreported genomic regions associated with the risk of RP at $p<5\times10^{-8}$ and assigned *ADRA2A* (rs7090046, odds ratio (OR) per allele: 1.27; 95%-CI: 1.23-1.31; $p<2.52\times10^{-47}$) and *IRX1* (rs11748327, OR: 1.20; 95%-CI: 1.16-1.23, $p<9.71\times10^{-28}$) as the candidate causal genes at the two strongest loci. Higher expression of *ADRA2A* in tibial artery and *IRX1* in skeletal muscle was thereby associated with a higher RP risk. We identified a likely causal detrimental effect of low fasting glucose levels on RP risk (r_G=-0.12; p-value=0.01), while significant positive genetic correlations with reported comorbidities like migraine, depression, or peripheral artery disease are likely explained by shared risk factors.

Conclusion: Our results provide the first robust evidence for a strong genetic contribution to RP, highlighting a so far underrated role of α_{2A}-adrenoreceptor signalling, encoded at *ADRA2A*, as an important mechanism for hypersensitivity to catecholamine-induced vasospasms even at thermoneutral conditions.

Introduction

Raynaud's phenomenon (RP) is a common episodic, vasospastic disorder that affects 2-5% of the population and can severely affect an individual's quality of life by causing pain or even ulcers[1–3]. RP typically manifests with bi- or triphasic colour change in fingers and toes because of vasospasms in arteriovenous anastomoses responsible for the thermoregulation, that can be triggered by cold or emotional stress[4]. The causes for the more common form of primary RP are largely unknown[4,5], while secondary RP is diagnosed as a consequence of connective tissue diseases, like systemic lupus erythematosus (SLE) or systemic sclerosis (SSc), or triggered by use of drugs such as beta-blockers. Management of RP is predominantly limited to avoidance of triggers and evidence for medical treatment is generally weak. Repurposed vasodilators are the first line treatment if pharmacological intervention is required due to progressive frequency of vasospastic attacks[6], although only calcium channel blockers have so far been shown to lead to a significant and reproducible reduction in the frequency of vasospastic attacks[7]. However, use of systemic drugs for localized symptoms puts patients at risk of generalised adverse effects, such as hypotension. While there are now trials investigating local application of botulinum toxin to mitigate vasospastic effects with early promising results[8], a better understanding of the underlying mechanisms is needed to develop safe and effective treatments.

RP is highly heritable with estimates of 55-64% being reported[9,10], but previous candidate gene studies[11,12], like at serotonin receptors[13], and an early, small (n=640 cases) genome-wide association study (GWAS)[14] failed to provide evidence for any robustly associated regions or genes. In-depth investigation and integration of detailed information from primary and secondary healthcare records with genetic array data now provides the opportunity to study under-investigated diseases with diagnostic specificity at unprecedented scale.

Here, we present the so far largest GWAS for RP including 9,084 cases in the UK Biobank cohort[15] and report eight novel loci. We highlight two independent disease mechanisms supported by the two strongest genetic loci that challenge and advance current understanding of primary RP, with *ADRA2A* highlighting the role of α_{2A}-adrenoreceptor but not α_{2C}-adrenoreceptors and *IRX1* as a putative regulator of prostaglandin responsiveness.

Results

We identified a total of 9,084 RP cases and 435,357 controls of European descent included in the genetic analyses (**Supplementary Tab. 1** and **Supplementary Fig. 1**) based on based on collation and evidence of absence or presence of diagnostic codes from electronic health records (ICD-10 codes: I73.0, I73.00, I73.01; SNOMED-CT: 195295006, 266261006; CTV3/Read2: G730., G7300, G7301, G730z, XE0VQ), including 4,275 prevalent cases and 4,809 incident cases. We followed the recommendation by Wigley *et al.*[4] to summarize all patients with a relevant code under the term RP (Raynaud's phenomenon) rather than Raynaud's syndrome.

Genome-wide association analysis

We identified a total of eight genome-wide significant novel loci ($p<5.0\times10^{-8}$, minor allele frequency (MAF) 0.16%-34.0%) associated with RP (**Fig. 1** and **Tab. 1;** regional association plots in **Supplementary Fig. 2**), two of which had previously been shown to increase the risk of coronary artery disease and myocardial infarction (rs11748327 and rs8051014). The risk of RP differed strongly for carriers of the two less common alleles (MAF = 0.16% and 0.38%, odds ratio (OR) 0.48 and 2.90), with effect sizes being more modest for the six common (MAF>5%) variants (OR 1.10 to 1.27) (**Tab. 1**). All lead variants resided in intergenic regions with no obvious functional variants in LD ($r^2>0.6$) and were of low predicted impact (median CADD score: 0.51, range: 0.17 - 3.02). Despite the established sex-difference of RP, we did not find evidence that effects of any of the identified lead signals differed between men and women (all p-values for interaction > 0.05; **Supplementary Tab. 2**). We estimated a SNP-based heritability on the liability scale of 8.7% (95%-CI: 6.7-10.9%, $p<9.76\times10^{-17}$) with little evidence of genomic inflation (LD-score intercept 1.02; **Supplementary Fig. 3**).

To evaluate whether genetic findings were specific for the development of RP and driven by associations with diseases leading to secondary RP, we performed sensitivity analyses based on a more stringent definition of primary RP (6,185 cases, 68.1%) that excluded potential secondary cases (see **Methods**). Effect estimates for all eight loci were highly consistent in size and direction (**Supplementary Fig. 4, Supplementary Tab. 3**) and remained genome-wide significant for four loci (and $p<1.9\times10^{-4}$ for all), indicating their role in the idiopathic origin of primary RP.

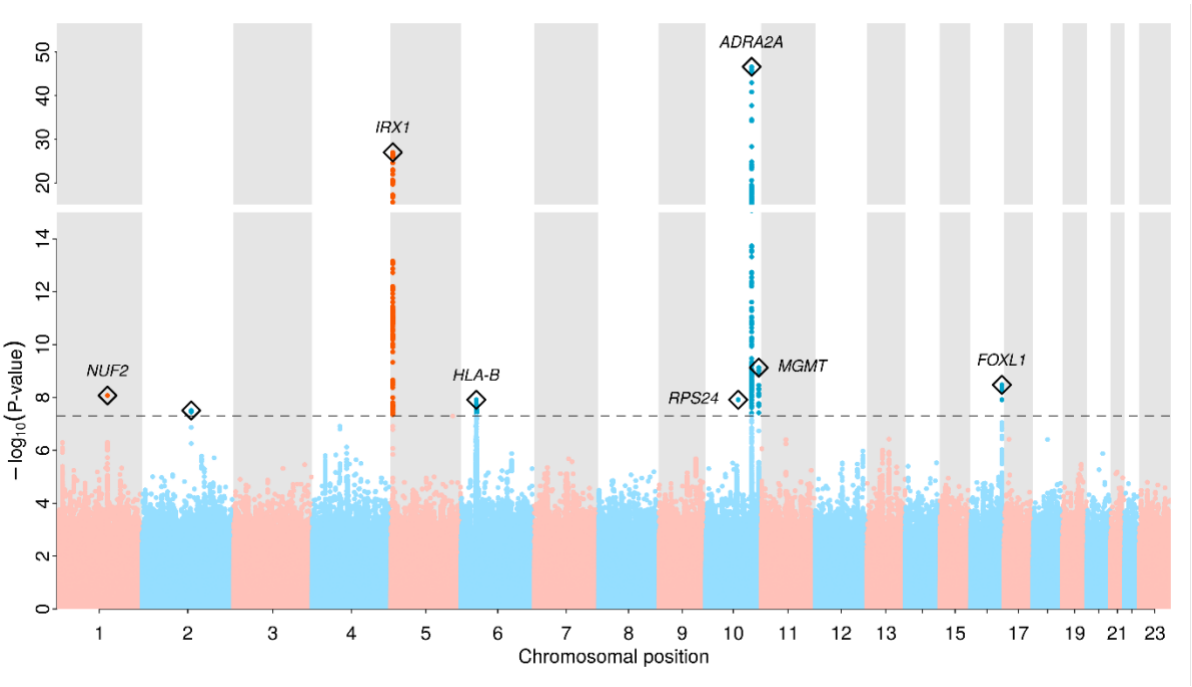


Figure 1 Manhattan plot of genome-wide association results for Raynaud's phenomenon (RP; 9,084 cases, 435,357 controls) Distribution of RP-associated single nucleotide polymorphisms across the genome, -log10(P-values) from logistic regression models are plotted for each variant. The grey dashed line indicates the genome-wide significant threshold at $p=5x10^{-8}$. Regional sentinel variants at significant loci are highlighted with a diamond and the closest gene, if any, is annotated.

SNP	Chr	Position	Alleles	MAF	OR (95%-CI)	P value	Closest gene
rs759480728	1	163231941	CTT/C	0.27	1.12 (1.07, 1.16)	$8.35x10^{-9}$	*NUF2*
rs7601792	2	130081720	G/A	0.0016	2.90 (1.99, 4.22)	$3.07x10^{-8}$	-
rs11748327	5	4029789	C/T	0.30	1.20 (1.16, 1.23)	$9.71x10^{-28}$	*IRX1*
rs2523598	6	31326175	T/G	0.33	1.10 (1.06, 1.13)	$1.22x10^{-8}$	*HLA-B*
rs183746285	10	80032754	C/T	0.0038	0.48 (0.38, 0.62)	$1.22x10^{-8}$	*RPS24*
rs7090046	10	112860930	G/A	0.31	1.27 (1.23, 1.31)	$2.52x10^{-47}$	*ADRA2A*
rs896703	10	130982895	T/C	0.33	1.10 (1.07, 1.14)	$7.34x10^{-10}$	*MGMT*
rs8051014	16	86717254	C/T	0.34	1.10 (1.07, 1.14)	$3.33x10^{-9}$	*FOXL1*

Table 1 Regional lead variants for the 8 loci significantly associated with Raynaud's phenomenon N = 444,441, 9,084 cases and 435,357 controls. Chr = Chromosome; MAF = minor allele frequency; OR = odds ratio from logistic regression models

ADRA2A and *BBIP1* are likely causal genes at 10q25.2

The lead variant at this strongest locus, rs7090046, is located 24.1k bp downstream to *ADRA2A* and 181.9k bases downstream from *BBIP1* (**Fig. 2**) and Bayesian fine-mapping identified a 95%-credible set of only 3 variants (see **Methods**), comprising rs7090046, rs1343449, and rs1343451, explaining 46.6%, 38.0% and 11.7% of the GWAS signal, respectively. Variant rs1343449 is a reported gene expression quantitative trait locus (cis-eQTL) for *ADRA2A* in artery tissue and statistical colocalization confirmed a shared genetic signal with RP with high confidence (tibial artery, posterior probability (PP) > 96%).

124 However, we also observed evidence of colocalization with *BBIP1* expression in the brain
125 (hippocampus, PP> 98%; **Fig. 2** and **Supplementary Tab. 4**), leaving the possibility that variants in the
126 credible set have pleiotropic tissue-dependent regulatory effects. This might be also reflected in terms
127 of direction of gene expression, since the RP risk-increasing A-allele of rs7090046 was associated with
128 higher *ADRA2A* expression in tibial artery (beta=0.28, $p=1.0x10^{-9}$), but lower *BBIP* expression in brain
129 hippocampus (beta=-0.37, $p=9.1x10^{-9}$).

130 *ADRA2A* encodes α_{2A}-adrenoreceptors, responsible for mediating response to stress in the central and
131 peripheral autonomous nervous system[16], and vasoconstriction in response to catecholamine release
132 in the small arteries and arterioles[17]. Its identification suggests that overstimulation or increased
133 expression of α_{2A}-adrenergic receptors underlies the vasospastic effects characteristic of RP and its
134 symptoms[4]. A possible role of *BBIP1*, in contrast, is less obvious. *BBIP1* encodes for a protein in the
135 BBSome complex, which is involved in the signalling from and to cilia. Malfunction of the BBSome
136 complex can lead to primary cilia dysfunction and can result in the Bardet-Biedel-Syndrome[18,19].

137 Despite the strong association with RP, we did not observe any other significant phenotypic
138 associations for rs7090046 or its proxies (r^2>0.6) across thousands of other outcomes or phenome-
139 wide analyses of UK Biobank ($p<10^{-6}$; see **Methods**). We only observed a significant association with
140 "specified peripheral vascular disease" in UK Biobank (OR 1.16, p-value $6.31x10^{-11}$), a non-specific code
141 that comprises, among other diseases, RP. These results therefore indicate an RP-specific mechanism
142 at 10q25.2.

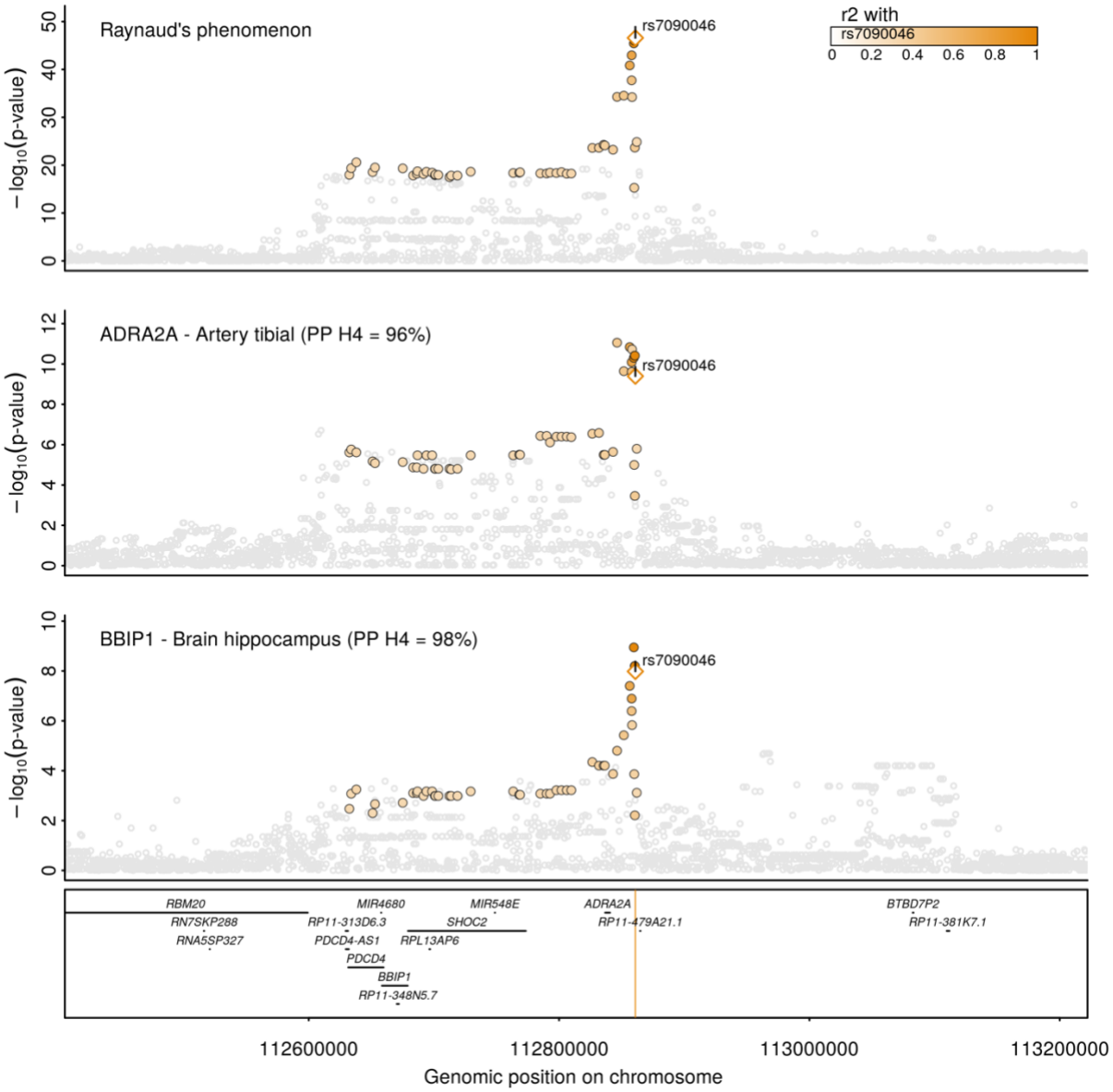
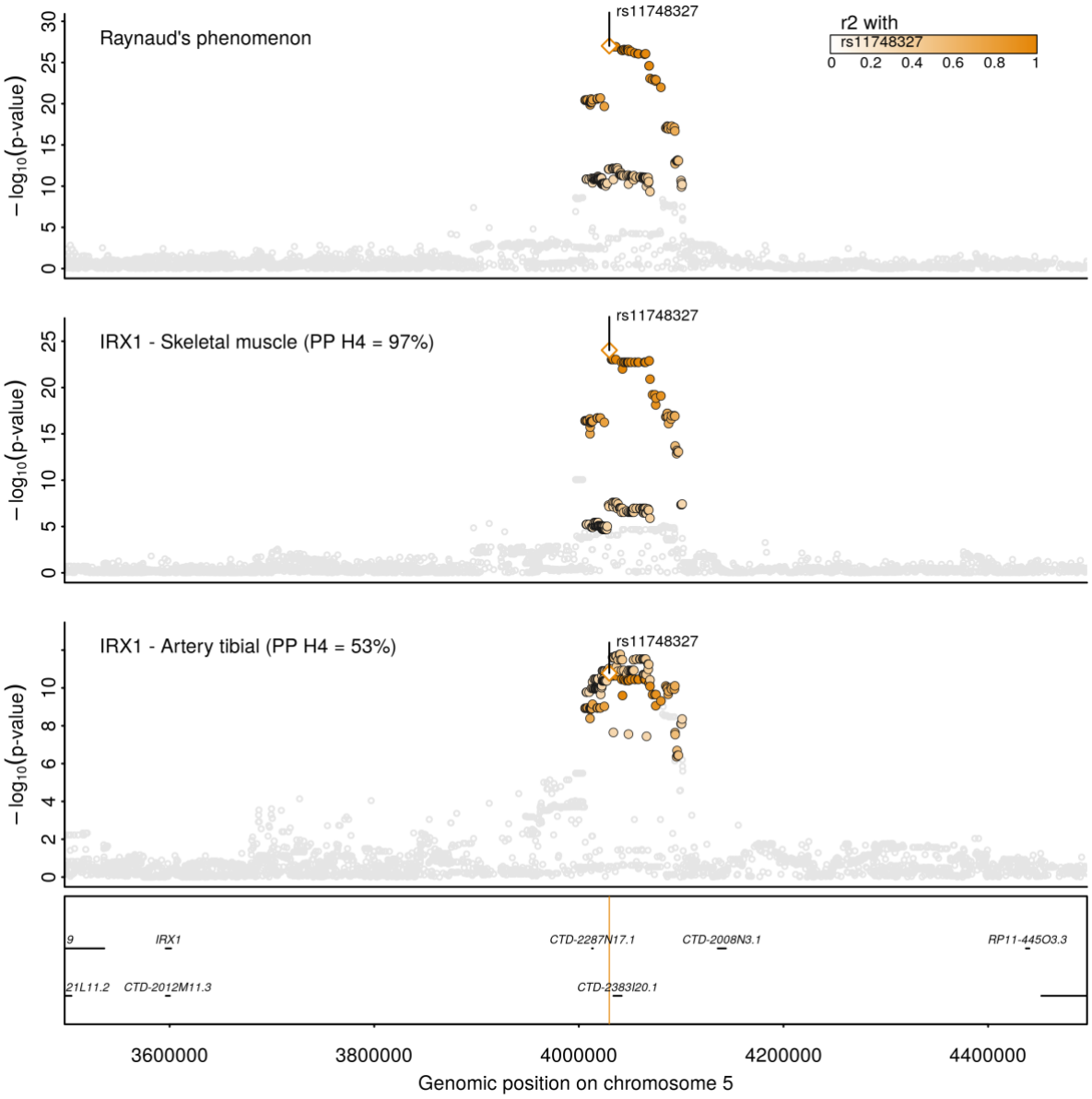


Figure 2 Regional association plot at *ADRA2A/BBIP1*. Regional association plot for Raynaud's phenomenon (top), gene expression of *ADRA2A* in tibial artery (middle), and gene expression of *BBIP1* in hippocampus (bottom). Summary statistics from logistic regression models for RS are form the present study whereas summary statistics from linear regression models were obtained from GTEx v8[20]. Colouring of SNPs is based on linkage disequilibrium with the lead RP variant (rs7090046) at this locus. Numbers in brackets indicate posterior probabilities (PP) for a shared genetic signal with RP based on statistical colocalization.

IRX1 is the likely causal gene at 5p15.33 and links to cardiovascular outcomes

We fine-mapped the second strongest signal at 5p15.33 to 18 variants with rs11748327 explaining 12.4% of the signal. rs11748327 has been reported as a cis-eQTL for *IRX1* in skeletal muscle[20]. Statistical colocalization further confirmed a shared genetic signal between RP and *IRX1* expression in skeletal muscle and tibial artery with high (PP=97%) and moderate (PP=53%; likely due to an unrelated secondary signal) confidence, respectively (**Fig. 3**). The RP-increasing effect-allele of rs11748327 was associated with higher expression of *IRX1* in skeletal muscle (beta=0.57, $p=1.1\times10^{-22}$) and arterial tibial

158 (beta=0.42, p=5.7x10^{-11}), suggesting that higher gene expression confers a higher risk for RS. *IRX1*
159 encodes for members of the homeobox containing genes which have been shown to be involved in
160 embryonic development and cellular differentiation[21]. The lead variant, rs11748327, at this locus tags
161 a cluster of variants (r^2=0.7-1.0) that have previously been reported to be associated with myocardial
162 infarction[22,23] and CAD[24] among East Asian populations and is further a trans-ethnically conserved
163 signal for red blood cell phenotypes[25,26]. In detail, the RP risk-increasing C-allele was also associated
164 with an increased risk for coronary artery disease (beta=-0.06; p<1.8x10^{-8}) in data from the BioBank
165 Japan[24]. The fact that this common locus (MAF_{EUR}=30.0%; MAF_{EAS}=22.3%) has not yet been identified
166 for MI or CAD in Europeans, might point to the possibility of a yet to be identified gene-environment
167 interaction or other characteristics distinct to the East Asian cohorts that account for this observation.



168

169 **Figure 3 Regional association plot at *IRX1*.** Regional association plot for Raynaud's phenomenon (top), gene
170 expression of IRX1 in skeletal muscle (middle), and gene expression of IRX1 in tibial artery (bottom).

Summary statistics from logistic regression models for RS are form the present study whereas summary statistics from linear regression models were obtained from GTEx v8[20]. Colouring of SNPs is based on linkage disequilibrium with the lead RP variant (rs11748327) at this locus. Numbers in brackets indicate posterior probabilities (PP) for a shared genetic signal with RP based on statistical colocalization.

A shared cardiovascular signal at the *FOXL1-FOXC2-FOXF1* cluster

We identified a credible set comprising ten variants for the signal at 16q24.1, with rs8051014 explaining about 22% of the signal. rs8051014 is located 107,280 bp upstream of the *FOXL1-FOXC2-FOXF1* gene cluster further comprising *MTHFSD* making credible gene assignment difficult. This locus has been previously associated with coronary artery disease among East Asians[24] and our phenome-wide analysis of European-descent UK Biobank participants also showed a significant association with 'coronary atherosclerosis' for rs8051014 ($p<3.7x10^{-5}$) (**Fig. 4**) that was confirmed as a shared genetic signal (PP=96.7%) in colocalization. The RP-risk increasing T-allele of rs8051014 was associated with a mildly increased risk for coronary atherosclerosis (OR 1.04, $p= 4.6 \times 10^{-6}$) highlighting a, at least locally, shared genetic architecture between RS and cardiovascular diseases.

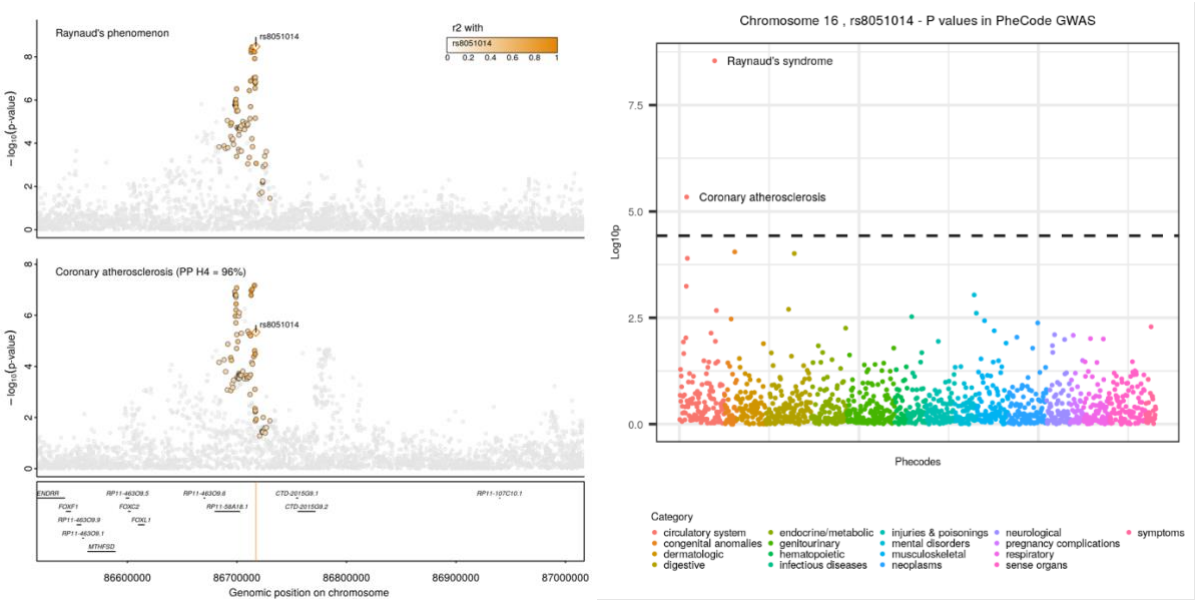


Figure 4 Regional association plot at *FOXL1* and phenome-wide analysis of rs8051014. Left: Regional association plot for Raynaud's phenomenon (top), and coronary atherosclerosis (bottom). Summary statistics from logistic regression models for RP are form the present study and summary statistics for coronary atherosclerosis are from an in-house analysis using UK Biobank. Colouring of SNPs is based on linkage disequilibrium with the lead RP variant (rs11748327) at this locus. **Right:** Phenome-wide analysis of rs8051014: each dot represents one of 1328 different 'phecodes'. Dots are coloured by their clinical category. The -log10(p-values) from logistic regression models associating rs8051014 with the risk to develop any phecode are plotted on the y-axis against the phecode categories (x-axis). The dashed line represents the corrected statistical significance level ($p<3.7x10^{-5}$).

198 ## Genetic correlation between RP and other traits

199 We next tested for a shared genetic architecture (genetic correlation: r_G) between RP and known or
200 suspected risk factors and comorbidities with a vascular or immune origin/component
201 (**Supplementary Tab. 5a**). We observed that possible causes or related disorders for RP, like migraine
202 (r_G= 0.22, p=1.65x10^{-3}), peripheral artery disease (r_G=0.35, p=8.03x10^{-3}), or SLE (r_G=0.27, p=2.19x10^{-3}),
203 only significantly correlated when considering all RP cases but were strongly attenuated once we
204 computed genetic correlations restricted to primary RP cases (**Fig. 5, Supplementary Tab. 5b**). We
205 identified significant inverse correlations with type 2 diabetes (r_G = -0.12, p=4.03x10^{-3}), fasting glucose
206 (r_G = -0.12, p=0.01), body mass index (r_G = -0.11, p=9.47x10^{-3}), and HDL-cholesterol levels (r_G=-0.13,
207 p=1.04x10^{-3}), and a positive correlation with total triglyceride levels (r_G=0.10, p=2.59x10^{-2}) (**Fig. 5**), all
208 of which were consistent in analyses of all as well as primary RP cases only.

209 To distinguish whether genetic correlations represent causal directions from versus towards RP or are
210 due to shared risk factors, we performed latent causal variable analysis[27] (see **Methods**). Out of a total
211 of nine phenotypes with evidence for significant genetic correlations, only SLE (genetic causality
212 proportion (GCP)=0.87; p<1.9x10^{-41}) and fasting glucose (GCP=0.73; p<1.0x10^{-14}) showed strong
213 evidence for a causal genetic effect on RP risk, of which fasting glucose but not SLE persisted in analyses
214 of primary RP cases only (GCP=0.84; p<5.0x10^{-27}). These results indicate that low fasting glucose levels
215 may causally increase the risk for RP.

216 We finally extended the targeted analysis by computing genetic correlations with 185 'phecodes', to
217 gain a more comprehensive view on possible comorbidities (see **Methods**; **Supplementary Tab. 6a**)
218 and observed 13 phecodes with significant robust, although moderate, genetic correlations with RP
219 (**Fig. 5**). This included positive correlations with several gastrointestinal entities including abdominal
220 pain (rG = 0.32, 6.99x10^{-7}), constipation (r_G = 0.27, p=6.75x10^{-3}), and irritable bowel syndrome (r_G =
221 0.24, p=1.87x10^{-3}), as well as disorders of the skin, osteoporosis, or varicose veins (**Fig. 5**). However,
222 none of the phenotypes showed evidence of a direct causal link towards or from RP in latent causal
223 variant analysis (see **Methods**; **Supplementary Tab. 6b**).

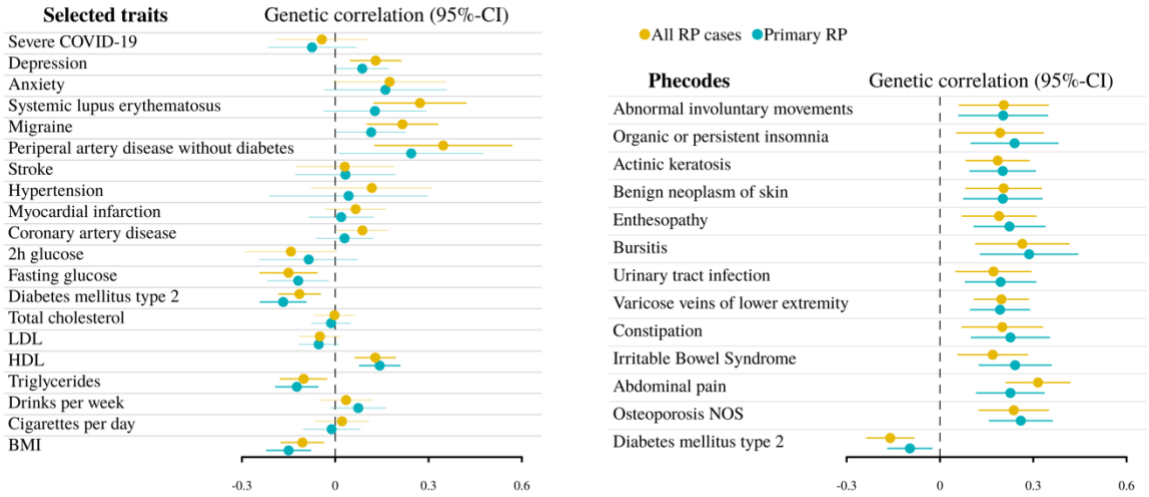


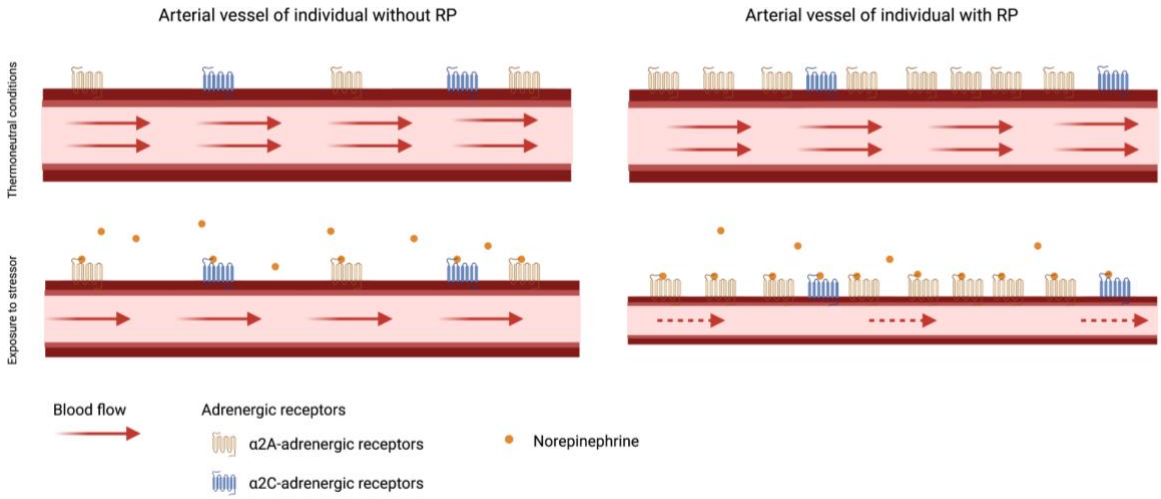
Figure 5 Forest plot summarizing genetic correlation analyses between RP and selected traits and risk factors *Left panel* Results for selected traits as listed in Supplementary Tab. 5a. *Right panel* Significant (q-value<0.05) and robust (persistent in primary RP) genetic correlation with 'phecodes' derived in UK Biobank. Colours indicate results considering all RP cases (yellow) or restricting to primary RP (green). Corresponding estimates and p-values can be found in Supplementary tables 7a and 7b.

Discussion

RP is a common but understudied disease with a high heritable component and identifying responsible genes can advance our understanding of the aetiology and eventually guide the development of treatment strategies. We identified, for the first time, eight robust susceptibility loci. Through integration with gene expression data, we demonstrated that the strongest locus acts via increased expression of α_{2A}-adrenoreceptors in arterial tissue, challenging the current concept of predominantly α_{2C}-adrenoreceptors mediating the vasospasms seen in RP patients[4]. We further highlight a possible role for prostaglandins through increased expression of the transcription factor *IRX1* in muscle and possibly arterial tissue. We observed no strong evidence for a shared genetic architecture with suspected cardiovascular comorbidities for primary RP but did observe a novel putative protective effect of fasting glucose levels that might help to guide primary RP management.

Our results highlight a specific role of α_{2A}-adrenoreceptors overexpression in the aetiology of RP as opposed to α_{2C}-adrenoreceptors (**Fig. 6**), which have long been thought to be the primary cause for cold-induced vasospastic attacks characteristic of RP[4,28–30]. Three highly homologous subtypes of α_2-adrenoreceptors exist: α_{2A}, α_{2B}, and α_{2C}[17], which are encoded on different chromosomes by *ADRA2A*, *ADRA2B,* and *ADRA2C*[31]. They mediate the response to stress in the central and peripheral autonomous nervous system[16] and contribute to vasoconstriction in the small arteries and arterioles[17]. More specifically, α_{2A}-adrenoreceptors mediate arterial vasoconstriction, thrombus stabilization, and hypothermic effects on body temperature[32,33], whereas α_{2C}-adrenoreceptors augment cold-induced

249 vasoconstriction. Our results are in line with a vasoconstrictive role of α_{2A}-adrenoreceptors under
250 thermoneutral conidtions[17] and subsequent RP risk, which is further supported by the fact that a strong
251 cis-eQTL for *ADRA2C* (rs35729104) did not show evidence of an effect on RP risk (OR=1.01, p=0.32).
252 However, it might still be conceivable that both α_2-adrenoreceptors act in a context-specific manner
253 to induce vasospastic effects[34] or another mechanism partakes in cold-induced RP attacks as one study
254 showed that inhibition of α_2-adrenergic receptors did not eliminate cold-induced vasoconstriction[35].



255

256 **Figure 6 Scheme of the key finding.** In people without Raynaud's phenomenon stimulation of α_{2A}- and α_{2C}-
257 adrenoreceptors contributes to vasoconstriction and lower blood flow. In patients with RP, blood flow is
258 further reduced due to overexpression of α_{2A}-adrenoreceptors creating a state of hypersensitivity to
259 catecholamine-release.

260 Currently RP is treated depending on its severity. For mild cases, avoidance of triggers like cold or
261 emotional stress might be sufficient, but severe vasospastic attacks require pharmacological
262 interventions with calcium channel blockers as first line response[6]. However, overall effectiveness of
263 calcium channel blockers[5] or alternative vasodilatory medications such as angiotensin receptor
264 blockers, selective serotonin reuptake inhibitors, or phosphodiesterase-5 inhibitors is limited[6,7] and
265 much hope put in selective α_{2C}-adrenoreceptors inhibitors has not been supported by a recent trial[36].
266 A staggering characteristic of the *ADRA2A* locus was the specificity in phenome-wide screens and
267 comprehensive database lookups, which indicated a RP-specific effect. Such a characteristic may make
268 the inhibition of the gene product, α_{2A}-adrenoreceptors, an interesting pharmacological target. While
269 this does not rule out that systemically administered antagonists might exert adverse effects, inhibiting
270 α_{2A}-adrenoreceptor activity in disease-relevant tissue, for example using topical solutions, might
271 provide a safe and effective treatment option. There are already approved medications that target α_2-
272 adrenoreceptor antagonistically like Yohimbine or Idazoxan, but they lack specificity for α_{2A}-
273 adrenoreceptor[32] and two studies described the occurrence or worsening of RP in patients under
274 Yohimbine treatment[37,38]. Two research compounds, BRL 44408 and BRL 48962, have been shown to
275 specifically target α_{2A}-adrenoreceptors in *in vitro* models but lack any data *in vivo*[39].

276 In addition to hypersensitivity to vasoconstrictive stimuli, missing vasodilation is the primary cause of
277 persistent vasospasms in RP. We identified overexpression of *IRX1*, the second strongest signal for RP,
278 in muscle cells and possibly tibial artery as a putative RP risk increasing mechanism. *IRX1* encodes for
279 the transcription factor iroquois homeobox protein 1, which is involved in neuronal development and
280 remodelling[40–42] and digit development in mice[43]. A transfection study in gastric cancer cell lines
281 identified several putative target genes for IRX1[44], including downregulation of *PTGER1*, encoding the
282 prostaglandin E2 receptor 1 (EP1), and upregulation of *PTGS2*, encoding prostaglandin G/H synthase
283 2. Prostaglandin E2 (PGE_2) can induce both vasodilation and vasoconstriction depending on receptor
284 binding, with binding to EP1 inducing vasoconstriction in smooth muscle cells[45,46]. While the results
285 based on the gastric cell lines would predict a vasodilative effect of increased *IRX1* expression, this
286 effect might be reversed in other tissues. Further experimental work is needed to establish a role of
287 *IRX1* in vasodilation/vasoconstriction, which might also help to refine the application of prostaglandin
288 as a vasodilator in RP.

289 Little is known about the causes of primary RP, and we identified evidence that genetic liability towards
290 low fasting glucose levels increases RP risk, aligning with our observation that type 2 diabetes are is
291 less common among RP patients (**Supplemental Tab. 1**). A putative mechanism, however, remains
292 elusive and we did not observe evidence that other metabolic risk factors were causally associated
293 with RP risk, including BMI. Overexpression of α_{2A}-adrenoreceptors in pancreatic islet cells has been
294 postulated as a mechanism for diminished insulin secretion and postprandial hyperglycaemia among
295 carriers of risk alleles in *ADRA2A*[47]. However, the reported risk allele, rs553668, is only weakly
296 correlated (r^2=0.15) with our RP lead signal and the same applies for the more recently identified
297 genome-wide significant finding at the same locus[48]. Although likely via distinct genetic variants, both
298 diseases seem to share overexpression of *ADRA2A* as a common, but tissue-specific, disease
299 mechanism, and it is unclear whether overexpression of *ADRA2A* in pancreatic beta-cells also
300 contributes to (high) fasting glucose levels. Conversely, hypoglycaemia exerts profound changes on the
301 vasculature, including vasoconstriction via increased adrenal catecholamine release to maintain blood
302 flow to the brain[49], that may in turn exaggerate hypersensitivity to catecholamines of arteriovenous
303 anastomoses in fingers and toes. However, a drop in body temperature following hypoglycaemia is
304 caused by cutaneous vasodilation, which might itself be a trigger for vasospastic effects. A pragmatic
305 consequence for people at RP risk, or even RP patients, might therefore be to avoid episodes of low
306 plasma glucose levels.

307 While we identify eight loci robustly associated with RP, we did not find support for genes highlighted
308 by previous smaller scale and candidate gene studies. This included subthreshold findings from an
309 earlier GWAS[14] (*NOS1*, smallest p-value in *NOS1* region ± 500kb (p_{min}) = $3x10^{-4}$) as well as candidate
310 gene studies for serotonin 1B (*HTR1B*; p_{min}=$2x10^{-4}$) and 1E (*HTR1E;* p_{min}=$9.1x10^{-5}$) receptors and the

311 beta subunit of the muscle acetylcholine receptor (*CHRNB1*; $p_{min}=1x10^{-4}$)[13]. Our results therefore
312 highlight the need for large sample sizes, which are only now possible with the increased availability
313 of electronic health record linkage.

314 Our study has limitations that need to be considered in the interpretation of the results. Firstly, we
315 defined cases based on billing codes from medical records of various sources which might have missed
316 cases with milder presentations and suffers from potential misclassification if patient symptoms have
317 been wrongly assigned to RP, in particular in the presence of disease that cause RP as a secondary
318 phenomenon. However, our conservative sensitivity analysis, excluding all patients with potential
319 known causes of RP revealed congruent findings and the associated loci have a clear biological link
320 towards RP pathology. Secondly, our study was exclusively comprised of participants of European
321 descent and low case numbers did not permit analysis in other ancestries, which is needed to achieve
322 fair representation of different ancestries in genomic studies and may reveal additional findings.
323 Thirdly, the nature of RP, being primarily diagnosed in a primary care setting, did not allow replication
324 in other large biobanks that miss linkage to primary care records, but will hopefully be available in the
325 future. Finally, while we assign putative candidate genes at two loci, more work is needed to improve
326 gene assignment pipelines to elucidate possible mechanism at other loci including rare variants with
327 strong effect sizes.

328 Our results advance the understanding of RP pathology by shifting the focus from a merely cold-
329 induced phenomenon to a clinical entity it its own right with a distinct genetic architecture. The
330 identification of *ADRA2A* and *IRX1* as candidate causal genes will further inform the development and
331 refinement of treatment strategies although more experimental work is needed.

332 ## ACKNOWLEDGEMENT
333 All research related to UK Biobank has been done under the application number 44448.

334

335 # Methods

336 ## United Kingdom Biobank (UKBB)

337 UK Biobank is a prospective cohort study from the UK which contains more than 500,000 volunteers
338 between 40 and 69 years of age at inclusion. The study design, sample characteristics and genome-
339 wide genotype data have been described in Sudow et al. and Bycroft et al.[15,50]. The UKBB was approved
340 by the National Research Ethics Service Committee North West Multi-Centre Haydock and all study
341 procedures were performed in accordance with the World Medical Association Declaration of Helsinki
342 ethical principles for medical research. We included 444,441 individuals in the GWAS for whom
343 inclusion criteria (given consent to further usage of the data, availability of genetic data and passed
344 quality control of genetic data) applied. Data from the UKBB were linked to death registries, hospital
345 episode statistics (HES) and primary care data.

346 ## RP case ascertainment

347 We used all available electronic health records, including primary and secondary care, death
348 certificates as well as participant's self-report to define RP cases. Primary care records for all UK
349 Biobank participants were released in three batches and we used the data releases from June 2021 for
350 RP case ascertainment. Similarly, we downloaded data tables for hospital admissions and death
351 certificates. We manually curated relevant codes for RP based on the different coding systems used
352 (ICD-10 codes: I73.0, I73.00, I73.01; SNOMED-CT: 195295006, 266261006; CTV3/Read2: G730., G7300,
353 G7301, G730z, XE0VQ) and flagged the first occurrence of a related code as the incidence date. We
354 compared the date of the very first record related to RP with the date of the baseline examination of
355 the same participant to distinguish between prevalent and incident RP. We followed the criteria from
356 Wigley et al.[4] to differentiate between primary and secondary RP. In detail, we classified each RP case
357 as 'secondary' when patients were either under beta-blocker treatment or had been diagnosed with
358 one of the following diseases prior to their RP diagnosis or at the same time (**Supplementary Tab. 7**):
359 arteriosclerosis, carpal tunnel syndrome, systemic lupus erythematosus, systemic sclerosis,
360 polymyositis, dermatomyositis, Sjögren syndrome, hypothyroidism, cryoglobulinemia, frostbite or
361 other connective tissues diseases.

362 ## UKBB medication classification/ beta-blocker usage

363 During the baseline assessment self-reported regular medication were recorded. Medication data
364 were categorized into 6,745 groups (Data field: 20003). 38 categories of the medication data were
365 identified manually to represent medication containing beta-blocker substances.

366 ## Demographic analysis

367 We performed Chi2-tests and analysis of variance to test for statistically significant differences in
368 demographic characteristics between controls and RP cases using standard implementations in R
369 (v4.1.2).

370 ## Genotyping, quality control and participant selection

371 Details on genotyping for UKBB have been reported in detail by Bycroft et al.[15]. Briefly, we used data
372 from the 'v3' release of UKBB containing the full set of Haplotype Reference Consortium (HRC) and
373 1000 Genomes imputed variants. We applied recommended sample exclusions by UKBB including low
374 quality control values, sex mismatch, and heterozygosity outliers. We defined a subset of 'white
375 European' ancestry by clustering participants based on the first four genetic principal component
376 derived from the genotyped data using a k-means clustering approach with k=5. We classified all
377 participants who belonged to the largest cluster and self-identified as of being 'white,' 'British', "Any
378 other white background', or 'Irish' as 'white European'. After application of quality control criteria and
379 dropping participants who have withdrawn their consent, a total of 444,441 UKBB participants were
380 available for analysis with genotype and phenotype data.

381 We used only called or imputed genotypes and short insertions/deletions (here commonly referred to
382 as SNPs for simplicity) with a minor allele frequency (MAF) > 0.001%, imputation score >0.4 for
383 common (MAF≥0.5%) and >0.9 for rare (MAF<0.5%), within Hardy-Weinberg equilibrium ($p_{HWE}>10^{-15}$),
384 and minor allele count (MAC) > 10. This left us with 15,519,342 autosomal and X-chromosomal variants
385 for statistical analysis. GRCh37 was used as reference genome assembly.

386 ## Genome-wide association study

387 We performed genome-wide association studies for RS and primary RS using REGENIE v2.2.4 via a two-
388 step procedure to account for population structure as described in detail elsewhere[51]. We used a set
389 of high-quality genotyped variants (MAF>1%, MAC>100, missingness <10%, $p_{HWE}>10^{-15}$) in the first step
390 for individual trait predictions using the leave one chromosome out (LOCO) scheme. These predictions
391 were used in the second step as offset to run logistic regression models with saddle point
392 approximation to account for case/control imbalance and rare variant associations. Models were
393 adjusted for age, sex, genotyping batch, assessment centre, and the first ten genetic principal
394 components.

395 ## LD score regression and genetic correlation to other common diseases

396 We tested for genomic inflation and calculated the SNP-based heritability using LD-score regression
397 (LDSC v1.0.1 ref. [52]). We further used LDSC to compute genetic correlation between RP as well as
398 primary RP and a set of 20 preselected traits and diseases (Supp**lementary Tab. 6a and 6b**). P-values

399 were corrected fo multiple testing by controlling the false-discovery rate at 5%. To explore genetic
400 correlation with other common diseases, we computed genetic correlation to a set of 185 phecodes
401 (Supp**lementary Tab. 7a**) with at least 1% SNP-based heritability derived from the UK Biobank.

402 Signal selection and fine-mapping

403 We used regional clumping (±500kb) to select independent genomic regions associated with RS
404 treating the extended MHC region as a whole (chr6:25.5-34.0Mb) and collapsed neighbouring regions
405 using BEDtools v1.5.

406 Within each region, apart from the MHC region, we performed statistical fine-mapping using "Sum of
407 single effects" model (SuSiE)[53] as implemented in the R package *susieR* (v.0.11.92). Briefly, SuSiE
408 employs a Bayesian framework for variable selection in a multiple regression problem with the aim to
409 identify sets of independent variants each of which likely contain the true causally underlying genetic
410 variant. We implemented the workflow using default prior and parameter settings, apart from the
411 minimum absolute correlation, which we set to 0.1. Since SuSiE is implemented in a linear regression
412 framework, we used the GWAS summary statistics with a matching correlation matrix of dosage
413 genotypes instead of individual level data to implement fine-mapping (*susie_rss()*) as recommended
414 by the authors[53]. We computed the correlation matrix with LDstore v1.0.

415 Sex-specific SNP effect

416 We tested for a potential modulating effect of sex on SNP – RP association using an additional
417 interaction term in logistic regression models with the same adjustment set as described above. We
418 implemented this analysis in R v.4.1.2 using a set of 361,781 unrelated white European individuals.

419 eQTL mapping to tissues/ Functional annotation

420 We systematically tested for a shared genetic signal between RS loci and gene expression levels (eQTL)
421 in 49 tissues from the GTEx project (v8)[20]. Briefly, we considered all protein coding genes or processed
422 transcripts encoded in a 1Mb window around RP loci and performed statistical colocalization[54] to test
423 for a shared genetic signal between RP and gene expression. We only report RP loci – gene – tissue
424 triplets with a posterior probability >80% for a shared genetic signal as candidate causal genes. We
425 adopted a recently recommended prior setting[55] with $p_{12}=5x10^{-6}$. All GTEx variant-gene cis-eQTL
426 associations from each tissue were downloaded in January 2020 from
427 https://console.cloud.google.com/storage/browser/gtex-resources.

428 Phenome wide association studies and variant look-up

429 We performed phenome-wide association studies for each RP locus identified using the
430 comprehensive electronic health record linkage to generate a set of 1,448 'phecodes', from which we
431 used 1,328 'phecodes' with a case number of more than 200. To generate phecode-based outcome

variables, we mapped ICD-10, ICD-9, Read version 2, Clinical Terms Version 3 (CTV3) terms, and SNOMED-CT codes from self-report or medical health records, including cancer registry, death registry, hospitalization (e.g., Hospital Episode Statistics for England), and primary care, to a set of summarized clinical entities called phecodes[56,57]. For example, more than 90 ICD-10 codes can indicate participants with type 1 diabetes that are here collectively summarized under the phecode 'type 1 diabetes'[58]. We used any code that was recorded, irrespective if it contributed to the primary cause of death or hospital admission, to define phecodes. We adjusted all analyses for test centre to account for regional differences in coding systems and case ascertainment. For each participant and phecode, we kept only the first entry irrespective of the original data set, generating a first occurrence data set. We dropped codes that were before or in the participants' birth year to minimize coding errors from electronical health records. We implemented the same logistic regression framework using REGENIE as described above to test for associations. We applied a stringent Bonferroni correction to account for multiple testing ($p < 3.7 \times 10^{-5}$).

We further performed look-ups for all independent lead variants and proxies ($r^2 > 0.6$) associated with RP in the OpenTargets database[59] and the GWAS catalog (https://www.ebi.ac.uk/gwas/docs/about) to search for possibly previously reported traits and to establish novelty.

LCV analysis

In addition to MR studies, to assess the evidence for a causal relationship between RP and each of the genetically correlated phenotypes (**Supplementary Tab. 5a and 6a**) we implemented a latent causal variable (LCV) model[27]. The LCV model estimates a parameter termed genetic causality proportion (GCP). The values of GCP ranges from -1 to 1, quantifying both magnitude and direction of genetic causality between two traits. Briefly, the LCV model fits an unobserved latent variable, L, which mediates the genetic correlation between two traits, and compares the correlation of each trait with L to estimate the GCP. A |GCP| value of one indicates full genetic causality (suggestive for vertical pleiotropy) whereas GCP value of zero implies no evidence for genetic causality (suggesting that the genetic correlation between two traits is likely to be mediated by horizontal pleiotropy) and |GCP| < 1 indicates partial genetic causality between two traits[ref]. Negative values of GCP estimate suggest that RP lies downstream to the other trait and interventions on other trait are likely to affect RP, while positive GCP values indicate that the other trait lies downstream to RP. We implemented LCV analysis using standard parameters with version 1.0. We used the same LD-scores as in the genetic correlation analysis and excluded the extended MHC regions from any calculations. We considered the number of significant genetic correlation as testing burden for LCV results.

Data availability

All individual level data is publicly available to bona fade researchers from the UK Biobank (https://www.ukbiobank.ac.uk/). GWAS summary statistics will be deposited at the GWAS catalog upon publication.

Author contributions

Conceptualization: MP, CL
Data curation/Software: SH, SY, SD, MP
Formal Analysis: SH, YS, MP
Methodology: SD, MP
Visualization: SH, SY, MP
Funding acquisition: CL, HH
Project administration: CL, HH
Supervision: MP, CL
Writing – original draft: SH, MP, CL
Writing – review & editing: YS, SD, HH

Competing interests

All authors declare that they have no competing interests.

References

1. Hughes, M. *et al.* Prediction and impact of attacks of Raynaud'S phenomenon, as judged by patient perception. *Rheumatol. (United Kingdom)* **54**, 1443–1447 (2015).

2. De Angelis, R., Salaffi, F. & Grassi, W. Health-Related Quality of Life In Primary Raynaud Phenomenon. *JCR J. Clin. Rheumatol.* **14**, 206–210 (2008).

3. Fábián, B., Csiki, Z. & Bugán, A. Raynaud-szindrómás betegek életminőségének jellemzői. *Orv. Hetil.* **159**, 636–641 (2018).

4. Wigley, F. M. & Flavahan, N. A. Raynaud's Phenomenon. *N. Engl. J. Med.* **375**, 556–565 (2016).

5. Rirash, F. *et al.* Calcium channel blockers for primary and secondary Raynaud's phenomenon. *Cochrane Database Syst. Rev.* **2017**, (2017).

6. Belch, J. *et al.* ESVM guidelines – the diagnosis and management of Raynaud's phenomenon. *Vasa* **46**, 413–423 (2017).

7. Su, K. Y. C. *et al.* Vasodilators for primary Raynaud's phenomenon. *Cochrane Database Syst. Rev.* **2021**, (2021).

8. Ennis, D., Ahmad, Z., Anderson, M. A. & Johnson, S. R. Botulinum toxin in the management of primary and secondary Raynaud's phenomenon. *Best Pract. Res. Clin. Rheumatol.* **35**, 101684 (2021).

9. Cherkas, L. F. *et al.* Heritability of Raynaud's phenomenon and vascular responsiveness to cold: A study of adult female twins. *Arthritis Care Res.* **57**, 524–528 (2007).

10. Hur, Y. M. *et al.* Feeling of cold hands and feet is a highly heritable phenotype. *Twin Res. Hum. Genet.* **15**, 166–169 (2012).

11. Smyth, A. E., Hughes, A. E., Bruce, I. N. & Bell, A. L. A case-control study of candidate vasoactive mediator genes in primary Raynaud's phenomenon. *Rheumatology* **38**, 1094–1098 (1999).

12. Shemirani, A. H. *et al.* Polymorphism of clotting factors in Hungarian patients with Raynaud's. *Blood Coagul. Fibrinolysis* **22**, 56–59 (2011).

13. Susol, E. *et al.* A two-stage, genome-wide screen for susceptibility loci in primary Raynaud's phenomenon. *Arthritis Rheum.* **43**, 1641–1646 (2000).

14. Munir, S., Freidin, M. B., Brain, S. & Williams, F. M. K. Association of raynaud's phenomenon with a polymorphism in the NOS1 gene. *PLoS One* **13**, 1–8 (2018).

15. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

16. Zhang, X. & Sun, Y. The predictive role of adra2a rs1800544 and htr3b rs3758987 polymorphisms in motion sickness susceptibility. *Int. J. Environ. Res. Public Health* **18**, (2021).

17. Chotani, M. A., Flavahan, S., Mitra, S., Daunt, D. & Flavahan, N. A. Silent α(2C)-adrenergic receptors enable cold-induced vasoconstriction in cutaneous arteries. *Am. J. Physiol. - Hear. Circ. Physiol.* **278**, 1075–1083 (2000).

18. Scheidecker, S. *et al.* Exome sequencing of Bardet-Biedl syndrome patient identifies a null mutation in the bbsome subunit BBIP1 (BBS18). *J. Med. Genet.* **51**, 132–136 (2014).

19. Klink, B. U., Gatsogiannis, C., Hofnagel, O., Wittinghofer, A. & Raunser, S. Structure of the human BBSome core complex. *Elife* **9**, 1–22 (2020).

524 20. Consortium, G. The GTEx Consortium atlas of genetic regulatory effects across human tissues
525 The GTEx Consortium* Downloaded from. (2021).

526 21. Küster, M. M. *et al.* Epigenetic inactivation of the tumor suppressor IRX1 occurs frequently in
527 lung adenocarcinoma and its silencing is associated with impaired prognosis. *Cancers (Basel).*
528 **12**, 1–20 (2020).

529 22. Aoki, A. *et al.* SNPs on chromosome 5p15.3 associated with myocardial infarction in Japanese
530 population. *J. Hum. Genet.* **56**, 47–51 (2011).

531 23. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes.
532 *Nat. Genet.* **53**, 1415–1424 (2021).

533 24. Koyama, S. *et al.* Population-specific and trans-ancestry genome-wide analyses identify
534 distinct and shared genetic risk loci for coronary artery disease. *Nat. Genet.* **52**, 1169–1177
535 (2020).

536 25. Vuckovic, D. *et al.* The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* **182**,
537 1214-1231.e11 (2020).

538 26. Chen, M. H. *et al. Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals*
539 *from 5 Global Populations*. *Cell* vol. 182 (2020).

540 27. O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52
541 diseases and complex traits. *Nat. Genet.* **50**, 1728–1734 (2018).

542 28. Freedman, R. R., Sabharwal, S. C., Desai, N., Wenig, P. & Mayes, M. Increased α-Adrenergic
543 responsiveness in idiopathic raynaud's disease. *Arthritis Rheum.* **32**, 61–65 (1989).

544 29. Edwards, J. M., Phinney, E. S., Taylor, L. M., Keenan, E. J. & Porter, J. M. α2-Adrenergic
545 receptor levels in obstructive and spastic Raynaud's syndrome. *J. Vasc. Surg.* **5**, 38–45 (1987).

546 30. Flavahan, N. A. A vascular mechanistic approach to understanding Raynaud phenomenon.
547 *Nat. Rev. Rheumatol.* **11**, 146–158 (2015).

548 31. Kim, J. O. *et al.* Association between common genetic variants of α2A-, α2B- And α2C-
549 adrenoceptors and the risk of silent brain infarction. *Mol. Med. Rep.* **9**, 2459–2466 (2014).

550 32. Gyires, K., Zádori, Z. S., Török, T. & Mátyus, P. α2-Adrenoceptor subtypes-mediated
551 physiological, pharmacological actions. *Neurochem. Int.* **55**, 447–453 (2009).

552 33. Hunter, J. C. *et al.* Assessment of the role of α2-adrenoceptor subtypes in the antinociceptive,
553 sedative and hypothermic action of dexmedetomidine in transgenic mice. *Br. J. Pharmacol.*
554 **122**, 1339–1344 (1997).

555 34. Freedman, R. R., Moten, M., Migaly, P. & Mayes, M. Cold-induced potentiation of α2-
556 adrenergic vasoconstriction in primary raynaud's disease. *Arthritis Rheum.* **36**, 685–690
557 (1993).

558 35. Cooke, J. P. *et al.* Role of digital artery adrenoceptors in Raynaud's disease. *Vasc. Med.* **2**, 1–7
559 (1997).

560 36. Herrick, A. L. *et al.* A double-blind, randomized, placebo-controlled crossover trial of the α2c-
561 adrenoceptor antagonist orm-12741 for prevention of cold-induced vasospasm in patients
562 with systemic sclerosis. *Rheumatol. (United Kingdom)* **53**, 948–952 (2014).

563 37. Johnson, S., Iazzetta, J. & Dewar, C. Severe Raynaud's Phenomenon with Yohimbine Therapy
564 for Erectile Dysfunction. *J. Rheumatol.* **30**, 2503–2505 (2003).

565 38. Sandler, B. & Aronson, P. Yohimbine-induced cutaneous drug eruption, progressive renal

566 failure, and lupus-like syndrome. *Urology* **41**, 343–345 (1993).

567 39. Cobos-Puc, L., Aguayo-Morales, H., Ventura-Sobrevilla, J., Luque-Contreras, D. & Chin-Chan,
568 M. Further analysis of the inhibition by agmatine on the cardiac sympathetic outflow: Role of
569 the α2-adrenoceptor subtypes. *Eur. J. Pharmacol.* **805**, 75–83 (2017).

570 40. Cheng, C. W., Hui, C. C., Strähle, U. & Cheng, S. H. Identification and expression of zebrafish
571 Iroquois homeobox gene irx1. *Dev. Genes Evol.* **211**, 442–444 (2001).

572 41. Bosse, A. *et al.* Identification of the vertebrate Iroquois homeobox gene family with
573 overlapping expression during early development of the nervous system. *Mech. Dev.* **69**, 169–
574 181 (1997).

575 42. He, S. *et al.* Transcriptional control of synaptic remodeling through regulated expression of an
576 immunoglobulain superfamily protein. *Curr. Biol.* **25**, 2541–2548 (2015).

577 43. Zülch, A., Becker, M. B. & Gruss, P. Expression pattern of Irx1 and Irx2 during mouse digit
578 development. *Mech. Dev.* **106**, 159–162 (2001).

579 44. Guo, X. *et al.* Homeobox gene IRX1 is a tumor suppressor gene in gastric carcinoma. *Oncogene*
580 **29**, 3908–20 (2010).

581 45. Gomez, I., Foudi, N., Longrois, D. & Norel, X. The role of prostaglandin E2 in human vascular
582 inflammation. *Prostaglandins Leukot. Essent. Fat. Acids* **89**, 55–63 (2013).

583 46. Ren, J., Karpinski, E. & Benishin, C. G. Prostaglandin E2 contracts vascular smooth muscle and
584 inhibits potassium currents in vascular smooth muscle cells of rat tail artery. *J. Pharmacol.*
585 *Exp. Ther.* **275**, 710–9 (1995).

586 47. Rosengren, A. H. *et al.* Overexpression of alpha2A-adrenergic receptors contributes to type 2
587 diabetes. *Science* **327**, 217–20 (2010).

588 48. Vujkovic, M. *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular
589 outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet.* **52**,
590 680–691 (2020).

591 49. Morrison, S. F. Differential control of sympathetic outflow. *Am. J. Physiol. Regul. Integr. Comp.*
592 *Physiol.* **281**, R683-98 (2001).

593 50. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide
594 Range of Complex Diseases of Middle and Old Age. *PLoS Med.* **12**, 1–10 (2015).

595 51. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and
596 binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).

597 52. Bulik-sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in
598 genome-wide association studies. *Nat. Publ. Gr.* **47**, (2015).

599 53. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable
600 selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Ser. B Stat.*
601 *Methodol.* **82**, 1273–1300 (2020).

602 54. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic
603 Association Studies Using Summary Statistics. *PLoS Genet.* **10**, (2014).

604 55. Wallace, C. Eliciting priors and relaxing the single causal variant assumption in colocalisation
605 analyses. *PLoS Genet.* **16**, 1–20 (2020).

606 56. Denaxas, S. *et al.* Mapping the Read2/CTV3 controlled clinical terminologies to Phecodes in
607 UK Biobank primary care electronic health records: implementation and evaluation. *AMIA ...*

608 *Annu. Symp. proceedings. AMIA Symp.* **2021**, 362–371 (2021).

609 57. Wu, P. *et al.* Mapping ICD-10 and ICD-10-CM Codes to phecodes: Workflow development and
610 initial evaluation. *JMIR Med. Informatics* **7**, 1–13 (2019).

611 58. Bastarache, L. Using Phecodes for Research with the Electronic Health Record: From PheWAS
612 to PheRS. *Annu. Rev. Biomed. Data Sci.* **4**, 1–19 (2021).

613 59. Ochoa, D. *et al.* Open Targets Platform: Supporting systematic drug-target identification and
614 prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).

615