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

Sylvia Hartmann1, Summaira Yasmeen1, Spiros Denaxas2,3,4, Harry Hemingway2,3,5, Maik Pietzner1,6,*
Claudia Langenberg1,6,7,*

1Computational Medicine, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany
2Institute of Health Informatics, University College London, London, UK
3Health Data Research UK, London, UK
4British Heart Foundation Data Science Centre, London, UK
5National Institute of Health Research University College London Hospitals Biomedical Research Centre
6MRC Epidemiology Unit, University of Cambridge, Cambridge, UK
7Precision 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)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
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 \( p < 5 \times 10^{-8} \) and assigned \textit{ADRA2A} (rs7090046, odds ratio (OR) per allele: 1.27; 95%-CI: 1.23-1.31; \( p < 2.52 \times 10^{-47} \)) and \textit{IRX1} (rs11748327, OR: 1.20; 95%-CI: 1.16-1.23, \( p < 9.71 \times 10^{-28} \)) as the candidate causal genes at the two strongest loci. Higher expression of \textit{ADRA2A} in tibial artery and \textit{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 (\( r_G = -0.12; \text{p-value}=0.01 \)), 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 \( \alpha_{2A} \)-adrenoreceptor signalling, encoded at \textit{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 \(\alpha_{2A}\)-adrenoreceptor but not \(\alpha_{2C}\)-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 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. 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<5.0×10^{-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<9.76x10^{-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<1.9x10^{-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, -\log_{10}(P-values) from logistic regression models are plotted for each variant. The grey dashed line indicates the genome-wide significant threshold at \( p = 5 \times 10^{-8} \). Regional sentinel variants at significant loci are highlighted with a diamond and the closest gene, if any, is annotated.

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

Table 1 Regional lead variants for the 8 loci significantly associated with Raynaud’s phenomenon

Table 1 Regional lead variants for the 8 loci significantly associated with Raynaud’s phenomenon

\[ \text{Chr} = \text{Chromosome; MAF = minor allele frequency; OR = odds ratio from logistic regression models} \]

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

ADRA2A encodes α₂A-adrenoreceptors, responsible for mediating response to stress in the central and peripheral autonomous nervous system, and vasoconstriction in response to catecholamine release in the small arteries and arterioles. Its identification suggests that overstimulation or increased expression of α₂A-adrenergic receptors underlies the vasospastic effects characteristic of RP and its symptoms. A possible role of BBIP1, in contrast, is less obvious. BBIP1 encodes for a protein in the BBSome complex, which is involved in the signalling from and to cilia. Malfunction of the BBSome complex can lead to primary cilia dysfunction and can result in the Bardet-Biedel-Syndrome.

Despite the strong association with RP, we did not observe any other significant phenotypic associations for rs7090046 or its proxies (r²>0.6) across thousands of other outcomes or phenome-wide analyses of UK Biobank (p<10⁻⁶; see Methods). We only observed a significant association with "specified peripheral vascular disease" in UK Biobank (OR 1.16, p-value 6.31x10⁻¹¹), a non-specific code that comprises, among other diseases, RP. These results therefore indicate an RP-specific mechanism 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\(^2\). 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\(^2\). 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=1.1x10\(^{-22}\)) and arterial tibial...
(beta=0.42, p=5.7x10^{-11}), suggesting that higher gene expression confers a higher risk for RS. IRX1 encodes for members of the homeobox containing genes which have been shown to be involved in embryonic development and cellular differentiation\(^2\). The lead variant, rs11748327, at this locus tags a cluster of variants (\(r^2=0.7-1.0\)) that have previously been reported to be associated with myocardial infarction\(^22,23\) and CAD\(^24\) among East Asian populations and is further a trans-ethnically conserved signal for red blood cell phenotypes\(^25,26\). In detail, the RP risk-increasing C-allele was also associated with an increased risk for coronary artery disease (beta=-0.06; p<1.8x10^{-8}) in data from the BioBank Japan\(^24\). The fact that this common locus (MAF\(_{EUR}=30.0\%\); MAF\(_{EAS}=22.3\%\)) has not yet been identified for MI or CAD in Europeans, might point to the possibility of a yet to be identified gene-environment interaction or other characteristics distinct to the East Asian cohorts that account for this observation.

**Figure 3 Regional association plot at IRX1.** Regional association plot for Raynaud’s phenomenon (top), gene 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 from the present study whereas summary statistics from linear regression models were obtained from GTEx v8\textsuperscript{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\textsuperscript{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\textsuperscript{-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 x 10\textsuperscript{-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 from 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\textsuperscript{-5}).
Genetic correlation between RP and other traits

We next tested for a shared genetic architecture (genetic correlation: \( r_G \)) between RP and known or suspected risk factors and comorbidities with a vascular or immune origin/component (Supplementary Tab. 5a). We observed that possible causes or related disorders for RP, like migraine (\( r_G = 0.22, p=1.65\times10^{-3} \)), peripheral artery disease (\( r_G = 0.35, p=8.03\times10^{-3} \)), or SLE (\( r_G = 0.27, p=2.19\times10^{-3} \)), only significantly correlated when considering all RP cases but were strongly attenuated once we computed genetic correlations restricted to primary RP cases (Fig. 5, Supplementary Tab. 5b). We identified significant inverse correlations with type 2 diabetes (\( r_G = -0.12, p=4.03\times10^{-3} \)), fasting glucose (\( r_G = -0.12, p=0.01 \)), body mass index (\( r_G = -0.11, p=9.47\times10^{-3} \)), and HDL-cholesterol levels (\( r_G = -0.13, p=1.04\times10^{-3} \)), and a positive correlation with total triglyceride levels (\( r_G = 0.10, p=2.59\times10^{-2} \)) (Fig. 5), all of which were consistent in analyses of all as well as primary RP cases only.

To distinguish whether genetic correlations represent causal directions from versus towards RP or are due to shared risk factors, we performed latent causal variable analysis (see Methods). Out of a total of nine phenotypes with evidence for significant genetic correlations, only SLE (genetic causality proportion \( GCP \)=0.87; \( p<1.9\times10^{-41} \)) and fasting glucose \( GCP=0.73; p<1.0\times10^{-14} \) showed strong evidence for a causal genetic effect on RP risk, of which fasting glucose but not SLE persisted in analyses of primary RP cases only \( GCP=0.84; p<5.0\times10^{-27} \). These results indicate that low fasting glucose levels may causally increase the risk for RP.

We finally extended the targeted analysis by computing genetic correlations with 185 ‘phecodes’, to gain a more comprehensive view on possible comorbidities (see Methods; Supplementary Tab. 6a) and observed 13 phecodes with significant, robust, although moderate, genetic correlations with RP (Fig. 5). This included positive correlations with several gastrointestinal entities including abdominal pain \( r_G = 0.32, 6.99\times10^{-7} \), constipation \( r_G = 0.27, p=6.75\times10^{-3} \), and irritable bowel syndrome \( r_G = 0.24, p=1.87\times10^{-3} \), as well as disorders of the skin, osteoporosis, or varicose veins (Fig. 5). However, none of the phenotypes showed evidence of a direct causal link towards or from RP in latent causal 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 Table 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 α2A-adrenoreceptors in arterial tissue, challenging the current concept of predominantly α2C-adrenoreceptors mediating the vasospasms seen in RP patients. 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 α2A-adrenoreceptors overexpression in the aetiology of RP as opposed to α2C-adrenoreceptors (Fig. 6), which have long been thought to be the primary cause for cold-induced vasospastic attacks characteristic of RP. Three highly homologous subtypes of α2-adrenoreceptors exist: α2A, α2B, and α2C, which are encoded on different chromosomes by ADRA2A, ADRA2B, and ADRA2C. They mediate the response to stress in the central and peripheral autonomous nervous system and contribute to vasoconstriction in the small arteries and arterioles. More specifically, α2A-adrenoreceptors mediate arterial vasoconstriction, thrombus stabilization, and hypothermic effects on body temperature, whereas α2C-adrenoreceptors augment cold-induced...
vasoconstriction. Our results are in line with a vasoconstrictive role of \( \alpha_{2A} \)-adrenoreceptors under thermoneutral conditions and subsequent RP risk, which is further supported by the fact that a strong cis-eQTL for \( \text{ADRA2C} \) (rs35729104) did not show evidence of an effect on RP risk (OR=1.01, \( p=0.32 \)). However, it might still be conceivable that both \( \alpha_2 \)-adrenoreceptors act in a context-specific manner to induce vasospastic effects or another mechanism partakes in cold-induced RP attacks as one study showed that inhibition of \( \alpha_2 \)-adrenergic receptors did not eliminate cold-induced vasoconstriction.

Figure 6 Scheme of the key finding. In people without Raynaud’s phenomenon stimulation of \( \alpha_{2A} \) and \( \alpha_{2C} \)-adrenoreceptors contributes to vasoconstriction and lower blood flow. In patients with RP, blood flow is further reduced due to overexpression of \( \alpha_{2A} \)-adrenoreceptors creating a state of hypersensitivity to catecholamine-release.

Currently RP is treated depending on its severity. For mild cases, avoidance of triggers like cold or emotional stress might be sufficient, but severe vasospastic attacks require pharmacological interventions with calcium channel blockers as first line response. However, overall effectiveness of calcium channel blockers or alternative vasodilatory medications such as angiotensin receptor blockers, selective serotonin reuptake inhibitors, or phosphodiesterase-5 inhibitors is limited and much hope put in selective \( \alpha_{2C} \)-adrenoreceptors inhibitors has not been supported by a recent trial.

A staggering characteristic of the \( \text{ADRA2A} \) locus was the specificity in phenome-wide screens and comprehensive database lookups, which indicated a RP-specific effect. Such a characteristic may make the inhibition of the gene product, \( \alpha_{2A} \)-adrenoreceptors, an interesting pharmacological target. While this does not rule out that systemically administered antagonists might exert adverse effects, inhibiting \( \alpha_{2A} \)-adrenoreceptor activity in disease-relevant tissue, for example using topical solutions, might provide a safe and effective treatment option. There are already approved medications that target \( \alpha_2 \)-adrenoreceptor antagonistically like Yohimbine or Idazoxan, but they lack specificity for \( \alpha_{2A} \)-adrenoreceptor and two studies described the occurrence or worsening of RP in patients under Yohimbine treatment. Two research compounds, BRL 44408 and BRL 48962, have been shown to specifically target \( \alpha_{2A} \)-adrenoreceptors in \textit{in vitro} models but lack any data in \textit{vivo}. 
In addition to hypersensitivity to vasoconstrictive stimuli, missing vasodilation is the primary cause of persistent vasoconstrictions in RP. We identified overexpression of *IRX1*, the second strongest signal for RP, in muscle cells and possibly tibial artery as a putative RP risk increasing mechanism. *IRX1* encodes for the transcription factor irquiso homeobox protein 1, which is involved in neuronal development and remodelling and digit development in mice. A transfection study in gastric cancer cell lines identified several putative target genes for *IRX1*, including downregulation of *PTGER1*, encoding the prostaglandin E2 receptor 1 (EP1), and upregulation of *PTGS2*, encoding prostaglandin G/H synthase 2. Prostaglandin E2 (PGE) can induce both vasodilation and vasoconstriction depending on receptor binding, with binding to EP1 inducing vasoconstriction in smooth muscle cells. While the results based on the gastric cell lines would predict a vasodilative effect of increased *IRX1* expression, this effect might be reversed in other tissues. Further experimental work is needed to establish a role of *IRX1* in vasodilation/vasoconstriction, which might also help to refine the application of prostaglandin as a vasodilator in RP.

Little is known about the causes of primary RP, and we identified evidence that genetic liability towards low fasting glucose levels increases RP risk, aligning with our observation that type 2 diabetes are less common among RP patients. A putative mechanism, however, remains elusive and we did not observe evidence that other metabolic risk factors were causally associated with RP risk, including BMI. Overexpression of α2a-adrenoreceptors in pancreatic islet cells has been postulated as a mechanism for diminished insulin secretion and postprandial hyperglycaemia among carriers of risk alleles in *ADRA2A*. However, the reported risk allele, rs553668, is only weakly correlated with our RP lead signal and the same applies for the more recently identified genome-wide significant finding at the same locus. Although likely via distinct genetic variants, both diseases seem to share overexpression of *ADRA2A* as a common, but tissue-specific, disease mechanism, and it is unclear whether overexpression of *ADRA2A* in pancreatic beta-cells also contributes to (high) fasting glucose levels. Conversely, hypoglycaemia exerts profound changes on the vasculature, including vasoconstriction via increased adrenal catecholamine release to maintain blood flow to the brain, that may in turn exaggerate hypersensitivity to catecholamines of arteriovenous anastomoses in fingers and toes. However, a drop in body temperature following hypoglycaemia is caused by cutaneous vasodilation, which might itself be a trigger for vasoconstrictive effects. A pragmatic consequence for people at RP risk, or even RP patients, might therefore be to avoid episodes of low plasma glucose levels.

While we identify eight loci robustly associated with RP, we did not find support for genes highlighted by previous smaller scale and candidate gene studies. This included subthreshold findings from an earlier GWAS (*NOS1*, smallest p-value in *NOS1* region ± 500kb (*p* = 3x10^-4) as well as candidate gene studies for serotonin 1B (*HTR1B*; *p* = 2x10^-4) and 1E (*HTR1E*; *p* = 9x10^-5) receptors and the
beta subunit of the muscle acetylcholine receptor (CHRNB1; \(p_{\text{min}}=1\times10^{-4}\))\textsuperscript{13}. Our results therefore highlight the need for large sample sizes, which are only now possible with the increased availability of electronic health record linkage.

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

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

\textbf{ACKNOWLEDGEMENT}

All research related to UK Biobank has been done under the application number 44448.
Methods

United Kingdom Biobank (UKBB)

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

RP case ascertainment

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

UKBB medication classification/ beta-blocker usage

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

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

Genotyping, quality control and participant selection

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

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

Genome-wide association study

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

LD score regression and genetic correlation to other common diseases

We tested for genomic inflation and calculated the SNP-based heritability using LD-score regression (LDSC v1.0.1 ref. 52). We further used LDSC to compute genetic correlation between RP as well as primary RP and a set of 20 preselected traits and diseases (Supplementary Tab. 6a and 6b). P-values
were corrected for multiple testing by controlling the false-discovery rate at 5%. To explore genetic correlation with other common diseases, we computed genetic correlation to a set of 185 phecodes (Supplementary Tab. 7a) with at least 1% SNP-based heritability derived from the UK Biobank.

Signal selection and fine-mapping

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

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

Sex-specific SNP effect

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

eQTL mapping to tissues/ Functional annotation

We systematically tested for a shared genetic signal between RS loci and gene expression levels (eQTL) in 49 tissues from the GTEx project (v8). Briefly, we considered all protein coding genes or processed transcripts encoded in a 1Mb window around RP loci and performed statistical colocalization to test for a shared genetic signal between RP and gene expression. We only report RP loci – gene – tissue triplets with a posterior probability >80% for a shared genetic signal as candidate causal genes. We adopted a recently recommended prior setting with $p_{12} = 5 \times 10^{-6}$. All GTEx variant-gene cis-eQTL associations from each tissue were downloaded in January 2020 from https://console.cloud.google.com/storage/browser/gtex-resources.

Phenome wide association studies and variant look-up

We performed phenome-wide association studies for each RP locus identified using the comprehensive electronic health record linkage to generate a set of 1,448 ‘phecodes’, from which we 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\textsuperscript{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'\textsuperscript{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< 3.7x10\textsuperscript{-5}).

We further performed look-ups for all independent lead variants and proxies (r\textsuperscript{2}>0.6) associated with
RP in the OpenTargets database\textsuperscript{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\textsuperscript{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 fide 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


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


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


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

