Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrative genomic and sequence-to-function analyses prioritize TNRC6B- linked regulatory variants in erysipelas susceptibility

Jingyi Peng
Guizhou University of Traditional Chinese Medicine
Guanglei Chen
Guizhou University of Traditional Chinese Medicine
Xiaojiao Gao
Guizhou University of Traditional Chinese Medicine
Pantao Zhao
Guizhou University of Traditional Chinese Medicine
Ankangxin Yan
Guizhou University of Traditional Chinese Medicine
Kun Wang
First Affiliated Hospital of Guizhou University of Traditional Chinese Medicine
Qi Yu
Guizhou University of Traditional Chinese Medicine
Weiyi Tian
Guizhou University of Traditional Chinese Medicine
Kun Cai
278527015@qq.com
Guizhou University of Traditional Chinese Medicine

Article

Keywords: erysipelas, TNRC6B, transcriptome-wide association study, enhancer-to-gene mapping, sequence-to-function, regulatory variants

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11343373/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Erysipelas is an acute skin and soft-tissue infection predominantly caused by streptococci, yet the host genetic architecture and non-coding regulatory mechanisms underlying susceptibility remain poorly characterized. We integrated FinnGen R11 erysipelas GWAS data (24,825 cases and 399,149 controls) with GTEx v8 and multiple functional genomic resources. Cross-tissue and tissue-specific TWAS, MAGMA, and Polygenic Priority Score (PoPS) analyses were combined with regional association testing, functional annotation, eQTL/sQTL and colocalization analyses, ENCODE chromatin annotations, rE2G/ABC enhancer-to-gene mapping, and AlphaGenome, Borzoi, and Enformer sequence-to-function models. TNRC6B was the only gene remaining significant after FDR correction in UTMOST ($P = 1.21 \times 10^{-5}$, FDR = 0.045) and was further supported by FUSION, MAGMA ($Z = 3.4697$, $P = 2.61 \times 10^{-4}$), and PoPS (score = 0.7712; genome-wide rank 324/18,383; local rank 2/31). No regional variant reached genome-wide significance. Integrated regulatory analyses prioritized rs4821940 and rs6001827 as Tier A candidates, whereas the 3'UTR variant rs139915 was classified as Tier B. Sequence-to-function models provided context-dependent support, although effect directions varied across models and tissues. Evidence-module integration, leave-one-module-out analyses, and 5,000 random weight perturbations consistently ranked rs4821940, rs6001827, and rs139915 among the top three candidates. These findings prioritize TNRC6B and linked regulatory variants for further investigation but do not establish causal effects.

Introduction

Erysipelas is an acute bacterial infection involving the superficial dermis and lymphatic vessels, with group A *Streptococcus* being the predominant causative pathogen [1, 2]. Although antimicrobial therapy is generally effective in controlling acute infection, recurrence remains common in a subset of patients. Impaired lymphatic drainage, chronic edema, venous insufficiency, obesity, diabetes, and disruption of the skin barrier have all been associated with an increased risk of erysipelas and its recurrence [3]. These observations suggest that, in addition to pathogen exposure, interindividual differences in host genetic background and immune regulation may contribute to disease susceptibility.

Previous studies have provided preliminary evidence for a genetic contribution to susceptibility to erysipelas and related non-necrotizing cellulitis. Family-based linkage analysis identified a disease-associated region at 9q34 and suggested that regulatory variation in *AGTR1* may contribute to susceptibility [4]. Variants in inflammation- and oxidative stress-related genes, including *IL17A*, *CYP1A1*, *SOD1*, *SOD2*, and *CAT*, have also been reported to be associated with erysipelas risk or related immune phenotypes [5–7]. However, these studies have largely focused on individual candidate genes or restricted genomic regions and therefore do not provide a comprehensive view of the genetic architecture underlying erysipelas. Because a substantial proportion of variants associated with complex traits are located in non-coding regions, genomic proximity alone is often insufficient to identify their functional target genes or regulatory mechanisms [8, 9]. Resolving the links between disease-associated variants and gene regulation is therefore essential for interpreting the genetic basis of erysipelas susceptibility.

Transcriptome-wide association studies (TWAS) integrate GWAS summary statistics with genetically predicted gene expression and provide a framework for linking genetic variation to candidate genes [10, 11]. However, TWAS associations may be influenced by linkage disequilibrium (LD), co-regulation, and shared cis-regulatory architecture among neighboring genes, and therefore do not by themselves establish a causal role for the implicated gene [12]. Interpretation of TWAS findings consequently benefits from the integration of complementary evidence, including gene-based association, functional prioritization, quantitative trait locus (QTL) analyses, colocalization, and variant-to-gene mapping [13–16]. Enhancer-to-gene approaches such as regulatory element-to-gene (rE2G) and Activity-by-Contact (ABC) models can further nominate potential target genes of non-coding regulatory elements [17–19]. In parallel, sequence-to-function models, including AlphaGenome, Borzoi, and Enformer, enable prediction of allele-dependent regulatory effects directly from genomic sequence [20, 21]. Because these approaches are methodologically related and are not fully independent, integrating them at the level of evidence classes and evaluating the robustness of the resulting prioritization is preferable to simply counting the number of nominally supportive models.

TNRC6B (trinucleotide repeat containing adaptor 6B) encodes a member of the GW182 protein family and is an important component of microRNA (miRNA)-mediated gene silencing [22, 23]. TNRC6B interacts with Argonaute proteins and participates in RNA-induced silencing complex (RISC)-mediated post-transcriptional regulation, while also recruiting deadenylation machinery such as the CCR4–NOT complex to modulate target mRNA stability and translational efficiency. The established roles of the TNRC6 family in miRNA-dependent repression and viral RNA regulation suggest a potential connection with infection-related and immune-regulatory processes [24]. Nevertheless, there is currently no direct evidence that *TNRC6B* or specific regulatory variants at this locus contribute to erysipelas. Its

potential involvement in erysipelas susceptibility therefore requires systematic evaluation using genetic and functional genomic approaches.

In this study, we integrated FinnGen R11 erysipelas GWAS data with multiple publicly available functional genomic resources. We first evaluated gene-level support for *TNRC6B* using cross-tissue TWAS, tissue-specific TWAS, MAGMA, and the Polygenic Priority Score (PoPS). We then performed regional genetic analyses together with QTL and colocalization analyses, regulatory annotation, rE2G/ABC enhancer-to-gene mapping, and sequence-to-function prediction using AlphaGenome, Borzoi, and Enformer to systematically assess candidate non-coding variants in the *TNRC6B* region. Finally, we applied evidence module-based integration and sensitivity analyses to identify variants with relatively robust TNRC6B-linked regulatory support. An overview of the analytical workflow and the multi-layer evidence integration framework is shown in Fig. 1.

The study integrated FinnGen R11 erysipelas GWAS data with GTEx v8, the eQTL Catalogue, ENCODE functional genomic resources, and the 1000 Genomes reference panel. Gene-level association and prioritization were first performed using UTMOST, FUSION, MAGMA, and PoPS. Variants within the *TNRC6B* ± 1 Mb region were subsequently evaluated by regional variant screening, functional annotation, linkage disequilibrium (LD) analysis, QTL analysis, and colocalization, followed by enhancer-to-gene mapping using rE2G and Activity-by-Contact (ABC). Candidate variants were further assessed using the sequence-to-function models AlphaGenome, Borzoi, and Enformer. Evidence from statistical genetics, functional annotation, QTL analyses, enhancer-to-gene mapping, sequence models, motif/RNA analyses, and target-gene specificity was then integrated at the module level. Robustness of variant prioritization was evaluated using leave-one-module-out analysis, leave-one-model-out analysis, and random weight perturbation.

Results

Multi-layer gene-level analyses prioritize TNRC6B

Using FinnGen R11 erysipelas GWAS data comprising 24,825 cases and 399,149 controls, cross-tissue TWAS with UTMOST identified 265 nominally associated genes ($P < 0.05$). After false discovery rate (FDR) correction, *TNRC6B* was the only gene that remained significant (test score = 10.61, $P = 1.21 \times 10^{-5}$, FDR = 0.045). Tissue-specific FUSION analysis identified 302 Ensembl gene identifiers reaching FDR < 0.05 in at least one GTEx v8 tissue, of which 237 had available gene-symbol annotations. *TNRC6B* reached significance in both Artery Aorta and Brain Cerebellum. Complete UTMOST and FUSION results are provided in Supplementary Tables S1 and S2, respectively.

MAGMA gene-based analysis further supported *TNRC6B*, with a Z score of 3.4697 and $P = 2.61 \times 10^{-4}$, ranking 101st among 18,758 genes. PoPS analysis independently prioritized *TNRC6B* at rank 324 of 18,383 genes genome-wide and rank 2 of 31 genes within the ± 1 Mb regional window. Regional gene-prioritization results, the corresponding MAGMA and PoPS statistics for *TNRC6B*, and chromosome 22 PoPS results are provided in Supplementary Table S3, with extended gene-level visualizations shown in Supplementary Fig. S1.

Collectively, the convergent evidence from TWAS, MAGMA, and PoPS prioritized *TNRC6B* as a key candidate gene for subsequent variant-level regulatory analyses (Fig. 2).

A, Cross-tissue UTMOST and tissue-specific FUSION results for *TNRC6B*. In UTMOST, *TNRC6B* showed a test score of 10.61 ($P = 1.21 \times 10^{-5}$, FDR = 0.045). Tissue-specific FUSION analysis further identified significant associations for *TNRC6B* in Artery Aorta and Brain Cerebellum.

B, Genome-wide MAGMA gene-based association analysis. *TNRC6B* had Z = 3.4697 and $P = 2.61 \times 10^{-4}$, ranking 101st among 18,758 genes. The dashed line indicates the Bonferroni-corrected significance threshold.

C, PoPS prioritization within the *TNRC6B* ± 1 Mb region. *TNRC6B* had a PoPS score of 0.7712 and ranked 2nd among 31 local genes, second only to *EP300*; its genome-wide rank was 324 of 18,383 genes.

Regional analyses indicate predominantly non-coding regulatory signals at the TNRC6B locus

A total of 1,656 variants were located within the *TNRC6B* gene body, whereas 14,486 variants were identified within the ± 1 Mb regional window. The strongest association within the gene body was observed for rs539666509 ($P = 8.20 \times 10^{-6}$), while the lead variant across the ± 1 Mb region was rs1037512591 ($P = 2.80 \times 10^{-6}$). No variant within this region reached genome-wide significance ($P < 5 \times 10^{-8}$).

Within the *TNRC6B* gene body, 146 variants met the prespecified functional-annotation screening threshold of $P < 1 \times 10^{-3}$. VEP annotation identified no missense, synonymous, or clearly splice-related variants among these candidates, whereas rs139915 was

located in the 3' untranslated region (3'UTR) of the *TNRC6B* MANE Select transcript. Gene-body variants meeting $P < 1 \times 10^{-3}$, the fixed set of 27 candidate variants, and the corresponding allele quality-control results are provided in Supplementary Table S4. Integration of LD information with ENCODE candidate cis-regulatory elements (cCREs), chromatin accessibility, and histone-modification datasets indicated that the prioritized signals were predominantly located in regulatory contexts relevant to endothelial, immune, and skin-associated cell types. Chromatin annotations for the candidate variants and their high-LD proxies, together with LD summaries and allele-count information from the reference panels, are provided in Supplementary Table S5.

Among the prioritized candidates, rs4821940 is an intronic *TNRC6B* variant supported by multiple active-chromatin annotations; rs6001827 shows prominent promoter-associated regulatory features; rs139915 is located in the 3'UTR; rs539666509 is supported by enhancer/cCRE annotations but has a low allele frequency; and rs4820399 is characterized primarily by intronic and upstream annotations. These five variants were carried forward for subsequent mechanistic analyses (Fig. 3).

A, FinnGen GWAS association profile across the *TNRC6B* ± 1 Mb region. Key candidate variants, including rs539666509, rs4821940, rs4820399, rs6001827, and rs139915, are highlighted. The dashed line indicates the genome-wide significance threshold ($P = 5 \times 10^{-8}$). No variant within the region reached genome-wide significance.

B, Functional annotation of the five prioritized candidate variants, including GWAS P values, VEP consequences, cCRE annotations, and minor allele counts in the FIN and EUR reference panels.

C, Chromatin regulatory evidence for the candidate variants across biologically relevant cellular contexts, including skin/keratinocyte, fibroblast, endothelial/vascular, monocyte/macrophage, neutrophil, and T-cell contexts.

cCRE, candidate cis-regulatory element; MAC, minor allele count.

QTL and colocalization analyses reveal tissue-dependent regulatory signals for TNRC6B

Across the public QTL datasets examined, the strongest *TNRC6B* cis-eQTL signal was observed in Brain Cerebellum, with a minimum P value of 2.73×10^{-8}. Cis-eQTL signals of varying strength were also detected in Artery Aorta, skin, blood, spleen, and other tissues.

Coloc.abf analysis yielded a posterior probability for a shared causal signal (PP.H4) of 0.797 in Brain Cerebellum, compared with 0.301 in Artery Aorta and 0.087 in Artery Tibial. The Brain Cerebellum colocalization result was sensitive to prior specification, and coloc.susie did not yield a stable and comparable GWAS–QTL signal pair. Candidate-variant eQTL/sQTL lookups, coloc.abf results, prior-sensitivity analyses, and coloc.susie assessments are summarized in Supplementary Table S6.

At the variant level, rs4821940, rs6001827, rs139915, and rs4820399 showed evidence of association with *TNRC6B* expression in available eQTL datasets. Overall, the genetic regulation of *TNRC6B* appeared to be strongly tissue- and context-dependent, with Brain Cerebellum providing the strongest locus-level colocalization support among the tissues examined (Fig. 4A).

rE2G and ABC distinguish regulatory patterns across candidate variants

Among the fixed set of 27 candidate variants, 24 showed an exact overlap with at least one regulatory element represented in the rE2G or ABC maps, and 23 overlapped distal or non-promoter elements. Integration of rE2G and ABC predictions revealed distinct *TNRC6B*-linked regulatory patterns across candidate variants. Complete rE2G/ABC primary edges, target-gene specificity assessments, biosample support, and strict concordance results are provided in Supplementary Table S7, with the overall enhancer-to-gene landscape shown in Supplementary Fig. S2. Importantly, the reported numbers of regulatory links represent aggregated computationally predicted connections across reference maps and biosample contexts and should not be interpreted as independent experimental validation events.

rs4821940 showed the clearest non-promoter variant-to-gene pattern, with 35 predicted non-promoter links to *TNRC6B*. In rE2G, *TNRC6B* was ranked as the top predicted target gene for the corresponding regulatory element across 18 biosample terms, and ABC independently supported a regulatory connection to *TNRC6B*. rs539666509 was likewise supported by both rE2G and ABC, although its low allele frequency limits interpretation based on LD-dependent evidence.

For rs6001827, the predicted *TNRC6B* links arose predominantly from promoter or self-promoter elements, consistent with a promoter-associated regulatory mechanism. rs139915 was linked to *TNRC6B* as well as *ADSL* and *SGSM3*, indicating appreciable target-gene ambiguity. In contrast, rs4820399 showed no *TNRC6B*-specific support from either rE2G or ABC (Fig. 4B,C).

A, Coloc.abf results comparing *TNRC6B* eQTL signals with the erysipelas GWAS in Brain Cerebellum, Artery Aorta, and Artery Tibial. Brain Cerebellum showed the highest PP.H4 (0.797). The dashed line indicates PP.H4 = 0.8, and the panel on the right shows sensitivity of the Brain Cerebellum result to the prior probability of a shared causal variant (p12).

B, Regulatory element-to-gene links predicted by rE2G and ABC across the *TNRC6B* locus. Connections represent computationally predicted links between candidate regulatory elements and *TNRC6B* or alternative target genes.

C, Target-gene specificity and promoter versus non-promoter regulatory patterns for the five prioritized candidate variants. rs4821940 predominantly showed non-promoter regulatory links to *TNRC6B*, whereas rs6001827 was characterized mainly by promoter/self-promoter regulatory support. rs139915 showed additional predicted links to alternative targets, including *ADSL* and *SGSM3*, indicating target-gene ambiguity.

Sequence-to-function models reveal context-dependent effects of candidate variants

All 27 candidate variants were successfully evaluated using REF/ALT predictions from AlphaGenome, Borzoi, and Enformer. Comprehensive sequence-model summaries, model-specific effect rankings, direction-concordance analyses, matched-context comparisons, QTL–prediction comparisons, and polyA- and motif/RNA-related results are provided in Supplementary Table S8. Representative extended sequence-model profiles and cross-model comparisons are shown in Supplementary Figs. S3–S4.

Cross-model analyses showed that the candidate variants exhibited a mixture of concordant and discordant predictions, indicating substantial assay-, tissue-, and model-dependent heterogeneity in sequence-level effects. Complete results for cross-model effect rankings, direction concordance, matched-context comparisons, and QTL–prediction directionality are presented in Supplementary Fig. S4.

Among the prioritized variants, rs139915 showed RNA/polyA-related sequence support, although the predicted direction differed across Borzoi replicates. rs539666509 exhibited relatively strong local sequence-model effects but limited cross-model concordance. rs4820399 also showed local sequence-level effects; however, these predictions were not accompanied by *TNRC6B*-specific enhancer-to-gene support. Overall, the sequence-to-function analyses helped distinguish the predicted regulatory profiles of individual candidate variants while also highlighting the context dependence and model sensitivity of these effects (Fig. 5).

A, Allele-dependent sequence-model predictions surrounding rs4821940, showing REF/ALT differences predicted by AlphaGenome, Borzoi, and Enformer within matched regulatory contexts.

B, RNA/polyA-related predictions for the *TNRC6B* 3'UTR variant rs139915. The predicted direction showed some inconsistency across Borzoi model replicates.

C, Integrated cross-model interpretation of the five prioritized candidate variants. rs4821940 showed predominantly partial and context-dependent support; rs6001827 was characterized mainly by promoter- and transcription-related effects; rs139915 showed RNA/polyA-related but model-sensitive predictions; rs539666509 exhibited cross-model conflict; and rs4820399 showed local sequence-level effects without *TNRC6B*-specific enhancer-to-gene support.

Multi-evidence integration prioritizes rs4821940, rs6001827, and rs139915

After organizing the available evidence into seven modules—statistical genetics, functional annotation, QTL, enhancer-to-gene mapping, sequence models, motif/RNA evidence, and target-gene specificity—the 27 candidate variants were classified into 2 Tier A, 16 Tier B, 7 Tier C, and 2 Tier D candidates. The complete seven-module evidence states, final tiers, rankings, and major limitations for all 27 variants are provided in Supplementary Table S9.

rs4821940 and rs6001827 were both classified as Tier A and jointly occupied the top two positions, whereas rs139915 was classified as Tier B and consistently ranked third. rs539666509 was assigned to Tier B, while rs4820399 was classified as Tier C. rs4821940 was primarily supported by chromatin, QTL, and enhancer-to-gene evidence and represented the strongest non-promoter *TNRC6B*-linked candidate. By contrast, rs6001827 predominantly reflected promoter-associated regulation, whereas rs139915 defined a distinct 3'UTR/RNA-regulatory pattern.

Across 5,000 random module-weight perturbations, rs4821940, rs6001827, and rs139915 were included among the top three candidates in 100% of iterations. The top-rank frequencies for rs4821940 and rs6001827 were 50.72% and 49.28%, respectively, indicating nearly equivalent overall prioritization despite their distinct regulatory profiles. Multi-layer evidence, major limitations, and final classifications for the five prioritized variants are summarized in Table 1. The complete seven-module evidence matrix for all 27 candidates is provided in Supplementary Table S9, whereas leave-one-module-out, leave-one-model-out, and 5,000-iteration module-weight perturbation results are presented in Supplementary Table S10 and Supplementary Fig. S5.

Taken together, rs4821940 and rs6001827 emerged as two similarly prioritized but mechanistically distinct regulatory candidates, while rs139915 represented a comparatively independent RNA-level regulatory hypothesis (Fig. 6).

Table 1
Integrated evidence and final prioritization of key candidate variants in the *TNRC6B* region

Variant	GWAS P	Genomic context	QTL evidence	Enhancer-to-gene evidence	Sequence-to-function	Key limitation	Final tier / rank
rs4821940	2.16×10^{-5}	Intronic	TNRC6B eQTL support	35 non-promoter TNRC6B links; TNRC6B favored	Partial/context-dependent support	Effect direction varies by context	Tier A, rank 1–2
rs6001827	8.42×10^{-5}	Intronic/promoter-associated	TNRC6B eQTL support	Promoter/self-promoter links	Promoter/transcription-related support	No distal non-promoter TNRC6B links	Tier A, rank 1–2
rs139915	2.64×10^{-4}	3'UTR	TNRC6B eQTL support	TNRC6B links with ADSL/SGSM3 ambiguity	RNA/polyA-related, model-sensitive	Alternative-target ambiguity	Tier B, rank 3
rs539666509	8.20×10^{-6}	Intronic	Limited assessment	143 non-promoter TNRC6B links	Model conflict	Low frequency; FIN MAC = 0, EUR MAC = 1	Tier B
rs4820399	4.59×10^{-5}	Intronic/upstream	TNRC6B eQTL signal	No TNRC6B-specific E2G support	Local sequence effects	Target assignment unsupported	Tier C

Symbols **M**, **L**, **0**, **!**, and **–** denote moderate support, limited support, no support, conflicting evidence, and not assessable, respectively.

A, Support levels for the five prioritized variants across seven evidence modules: statistical genetics, functional annotation, QTL, enhancer-to-gene mapping, sequence models, motif/RNA evidence, and target-gene specificity.

B, Changes in candidate-variant rankings in the leave-one-module-out analysis.

C, Top-1 and top-3 frequencies of rs4821940, rs6001827, and rs139915 across 5,000 random module-weight perturbations.

D, Final integrated prioritization. rs4821940 and rs6001827 were classified as Tier A candidates, representing non-promoter regulatory and promoter-associated regulatory hypotheses, respectively, whereas rs139915 was classified as Tier B and represented a distinct 3'UTR/RNA-level regulatory hypothesis.

Discussion

In this study, we integrated FinnGen R11 erysipelas GWAS data with TWAS, MAGMA, PoPS, QTL and colocalization analyses, regulatory annotation, enhancer-to-gene mapping, and sequence-to-function models to systematically prioritize *TNRC6B* and nearby non-coding variants. *TNRC6B* remained significant after FDR correction in the cross-tissue TWAS and was further supported by tissue-specific TWAS, MAGMA, and PoPS analyses [10, 13–15, 17, 18, 20, 21]. Subsequent variant-level analyses prioritized rs4821940 and rs6001827 as Tier A candidates, whereas rs139915 was classified as a Tier B candidate with a distinct 3'UTR/RNA-regulatory profile. Together, these findings suggest that the *TNRC6B* locus may harbor multiple forms of non-coding regulatory variation rather than a single uniform regulatory mechanism.

Previous studies of host genetic susceptibility to erysipelas have been relatively limited and have mainly focused on the 9q34 linkage signal, *AGTR1*, and selected inflammation- or oxidative stress-related candidate genes [5]. These studies support a contribution of host genetic factors to erysipelas susceptibility, but they have not systematically resolved the regulatory variants underlying disease-associated signals. By extending gene-level prioritization to variant-to-gene mapping and sequence-level functional prediction, our study provides a more detailed framework for interpreting the regulatory architecture of erysipelas susceptibility.

TNRC6B received consistent support across several gene-level approaches, although these methods capture different aspects of genetic evidence. TWAS evaluates associations between genetically predicted gene expression and disease, MAGMA aggregates SNP-level association signals at the gene level, and PoPS prioritizes genes by integrating genetic association with functional features. The consistent prioritization of *TNRC6B* across these analyses supports its designation as a high-priority candidate susceptibility gene. However, these lines of evidence should not be regarded as fully independent. In particular, PoPS incorporates gene-level association information and therefore partly depends on the underlying MAGMA signal. Moreover, other genes within the regional locus also received substantial prioritization, indicating that the locus may contain a more complex multi-gene or multi-regulatory-element architecture rather than a uniquely resolved target gene [12].

At the variant level, rs4821940, rs6001827, and rs139915 displayed distinct predicted regulatory patterns. rs4821940 showed the clearest non-promoter regulatory element-to-*TNRC6B* relationship and was supported by chromatin annotations, QTL evidence, and both rE2G and ABC mapping. This convergence makes rs4821940 the strongest non-promoter *TNRC6B*-linked candidate in the present analysis. By contrast, the evidence for rs6001827 was concentrated primarily in promoter and self-promoter elements, consistent with a promoter-associated regulatory hypothesis. rs139915 is located in the *TNRC6B* 3'UTR and showed RNA/polyA-related evidence, suggesting a potentially distinct post-transcriptional regulatory route [25, 26]. The remaining highlighted variants were less compelling overall: rs539666509 showed substantial regulatory predictions but was limited by its low allele frequency and instability in LD reference panels, whereas rs4820399 lacked *TNRC6B*-specific enhancer-to-gene support despite showing local sequence-level effects. These contrasts illustrate the value of combining target-gene assignment with local functional prediction rather than interpreting either source of evidence in isolation.

QTL and colocalization analyses further indicated that genetic regulation of *TNRC6B* is strongly tissue- and context-dependent [27]. Brain Cerebellum provided the strongest locus-level colocalization support among the tissues examined, but the PP.H4 estimate remained below the prespecified strong-support threshold and was sensitive to prior specification. In addition, coloc.susie did not identify a stable and directly comparable GWAS–QTL signal pair. These results therefore do not support a simple model in which a single shared causal variant robustly explains both *TNRC6B* expression and erysipelas risk across tissues. Instead, they are more consistent with modest and context-dependent regulatory sharing.

Sequence-to-function analyses provided an additional layer of variant interpretation but similarly showed substantial heterogeneity. AlphaGenome, Borzoi, and Enformer predicted allele-dependent effects for several candidate variants, yet effect direction and magnitude varied across assays, tissues, and models [28]. This was particularly evident for rs139915, for which RNA/polyA-related predictions differed across Borzoi replicates, and for rs539666509, which showed relatively strong local effects but limited cross-model concordance. These models should therefore be interpreted as complementary computational predictors rather than as independent functional validations. The combined results support convergence of multiple moderate lines of evidence rather than a single large and uniformly reproducible sequence-level effect.

The established molecular function of *TNRC6B* provides biological plausibility for its potential involvement in host susceptibility. *TNRC6B* belongs to the GW182/TNRC6 family and is a core component of microRNA-mediated gene silencing. Through interactions with Argonaute proteins and deadenylation machinery such as CCR4–NOT, TNRC6 proteins regulate mRNA stability, decay, and translational repression [22]. The involvement of the TNRC6 family in miRNA-dependent repression and host regulation of viral RNA further suggests a potential role in infection-related post-transcriptional control [29]. Nevertheless, these established molecular functions do not demonstrate that the prioritized erysipelas-associated variants alter RISC activity or other TNRC6B-dependent processes. The specific role of *TNRC6B* in streptococcal infection and erysipelas therefore remains to be determined.

Secondary analyses provided additional but limited contextual evidence. MAGMA.Celltyping did not reveal a robust erysipelas-relevant cell-type enrichment pattern [30]. Although random-walk-with-restart analysis placed *TNRC6B* relatively high in the propagated network, most of its propagation score was attributable to its own MAGMA-derived seed, limiting its interpretation as independent network support [31]. MESC suggested a possible expression-mediated component of erysipelas heritability, but the global estimate was imprecise and no tissue-specific estimate remained significant after multiple-testing correction [32]. Detailed results for

MAGMA.Celltyping, network propagation, and MESC are provided in Supplementary Table S11 and Supplementary Figs. S6–S8. These analyses therefore serve primarily to contextualize the main findings rather than to provide independent validation of *TNRC6B*.

A major strength of this study is the integration of gene-level association, regulatory genomics, variant-to-gene mapping, and sequence-to-function prediction within a unified analytical framework. Importantly, related computational approaches were grouped into evidence modules rather than treated as independent votes, reducing inflation from correlated methods. The leave-one-module-out, leave-one-model-out, and random module-weight perturbation analyses further showed that rs4821940, rs6001827, and rs139915 remained among the most highly prioritized variants under a broad range of analytical perturbations. This robustness does not establish causality, but it increases confidence that their prioritization is not driven solely by any single computational method.

Several limitations should be considered. First, the genetic data were derived predominantly from the Finnish population, and differences in allele frequency and LD structure may limit generalizability to other ancestries. Second, no variant in the *TNRC6B* region reached conventional genome-wide significance, and the regional associations should therefore be interpreted as sub-genome-wide signals requiring replication. Third, available QTL and epigenomic resources do not fully reproduce the cellular and inflammatory environment of acute erysipelas [33]. Fourth, low-frequency variants such as rs539666509 are poorly represented in commonly used reference panels, reducing confidence in LD-dependent analyses. Fifth, enhancer-to-gene and sequence-to-function results remain computational predictions, and some candidate effects were model- or context-dependent. Finally, no allele-specific functional experiments were performed. Thus, the present study prioritizes regulatory hypotheses but does not establish *TNRC6B* or any individual variant as causal.

Overall, rs4821940, rs6001827, and rs139915 represent three mechanistically distinct regulatory hypotheses at the *TNRC6B* locus: a non-promoter regulatory element-to-gene model, a promoter-associated model, and a 3'UTR/RNA-level model, respectively. Their consistent prioritization across multi-layer evidence and robustness analyses makes them the principal candidates for future mechanistic investigation. However, experimental studies will be required to determine whether these variants directly alter *TNRC6B* regulation and whether such effects contribute to erysipelas susceptibility.

Conclusion

In this study, we integrated FinnGen R11 erysipelas GWAS data with TWAS, MAGMA, PoPS, QTL and colocalization analyses, regulatory annotation, rE2G/ABC enhancer-to-gene mapping, and sequence-to-function analyses to systematically prioritize *TNRC6B* and nearby non-coding variants. *TNRC6B* remained significant after FDR correction in the cross-tissue TWAS and received consistent support from tissue-specific TWAS, MAGMA, and PoPS, indicating a high gene-level priority.

Further analyses suggested that the principal candidate signals at the *TNRC6B* locus are predominantly regulatory rather than protein-coding. rs4821940 and rs6001827 received the highest integrated prioritization and represented non-promoter and promoter-associated regulatory hypotheses, respectively, whereas rs139915 supported a distinct 3'UTR/RNA-level regulatory hypothesis. Module-based evidence integration and robustness analyses further supported these three variants as the leading candidates.

Overall, *TNRC6B* emerges as a high-priority candidate gene for erysipelas susceptibility, with rs4821940, rs6001827, and rs139915 representing the principal variants for future functional investigation. However, because no variant in the region reached genome-wide significance and several lines of evidence showed tissue- and model-dependent heterogeneity, these findings should be interpreted as computational prioritization rather than evidence of established causality or experimentally validated regulatory mechanisms.

Methods

Data sources and overall study design

GWAS summary statistics for erysipelas were obtained from FinnGen Release 11 for phenotype AB1_ERYSIPELAS, comprising 24,825 cases and 399,149 controls. All genomic analyses were performed using GRCh38 coordinates. The conventional genome-wide significance threshold was set at $P < 5 \times 10^{-8}$.

The study was conducted in a stepwise framework. FinnGen GWAS data were first integrated with GTEx v8 expression-prediction resources for gene-level prioritization. Variants within the *TNRC6B* gene body and a ± 1 Mb regional window were subsequently screened and functionally annotated. QTL and colocalization analyses, enhancer-to-gene mapping, and sequence-to-function predictions were then incorporated to characterize potential regulatory mechanisms. Finally, evidence from different analytical domains was grouped into

predefined modules for variant tiering and robustness assessment. Software and resource versions, genome builds, recorded parameters, and documentation status for each analysis are summarized in Supplementary Table S12.

Gene-level association and prioritization

Cross-tissue transcriptome-wide association analysis was performed using UTMOST by integrating genetically predicted gene expression across 49 GTEx v8 tissues. False discovery rate (FDR) < 0.05 was used as the multiple-testing-adjusted significance threshold. Tissue-specific TWAS was subsequently performed using FUSION with GTEx v8 tissue-specific expression weights and GWAS Z statistics to evaluate associations between genetically predicted expression and erysipelas susceptibility.

Genome-wide gene-based association analysis was performed using MAGMA v1.10 with Ensembl release 116 gene annotations. MAGMA gene-level statistics were subsequently used as input for PoPS v0.2, which prioritizes genes by integrating genetic association evidence with functional features, including expression, pathway, and protein–protein interaction information. *TNRC6B* was evaluated according to both its genome-wide PoPS rank and its rank among genes within the local ± 1 Mb region.

Regional variant screening and regulatory annotation

FinnGen GWAS variants within the *TNRC6B* gene body and the surrounding ± 1 Mb region were extracted for regional analyses. $P < 1 \times 10^{-4}$ and $P < 1 \times 10^{-3}$ were used as screening thresholds for functional prioritization and were not interpreted as GWAS significance thresholds. Subsequent functional analyses were based on a prespecified set of 27 candidate variants.

Candidate variants were annotated using Ensembl Variant Effect Predictor (VEP) v116, with additional integration of CADD, SpliceAI, and MANE Select transcript information. Linkage disequilibrium was evaluated using 1000 Genomes Project Phase 3 European (EUR) and Finnish (FIN) reference populations, with $r^2 \geq 0.8$ defining high-LD proxies.

Candidate variants and their high-LD proxies were further intersected with ENCODE SCREEN candidate cis-regulatory elements (cCREs), DNase-seq/ATAC-seq chromatin-accessibility data, H3K27ac, H3K4me1/H3K4me3 histone-modification tracks, and transcription factor ChIP-seq datasets. Regulatory evidence was evaluated with particular attention to skin/keratinocyte and fibroblast, endothelial/vascular, and major immune-cell contexts.

QTL and colocalization analyses

TNRC6B cis-eQTL and sQTL data were obtained primarily from eQTL Catalogue release 7, including GTEx v8 and multiple immune-cell and stimulation-condition datasets. QTL and FinnGen GWAS datasets were harmonized to the GRCh38 reference genome, and effect alleles were aligned before directional comparisons.

Single-signal colocalization was performed using coloc.abf implemented in coloc v5.2.3. The primary prior probability for a shared causal variant was set to $p12 = 1 \times 10^{-5}$, with p12 values of 1×10^{-6} and 1×10^{-4} examined in prior-sensitivity analyses. $PP.H4 \geq 0.8$ was predefined as relatively strong support for colocalization, whereas $0.5 \leq PP.H4 < 0.8$ was considered moderate support. Potential multi-signal colocalization was further evaluated using coloc.susie together with susieR v0.14.2.

Enhancer-to-gene mapping

The fixed set of 27 candidate variants was mapped against ENCODE-rE2G v1.0.0 and ABC v1.1.2 reference maps. Exact genomic overlap between each variant and annotated regulatory elements was required, and corresponding target genes, prediction scores, and biosample information were retained.

Promoter/self-promoter and non-promoter regulatory elements were evaluated separately. Promoter/self-promoter links were excluded when specifically assessing distal regulatory mechanisms. Concordant support between rE2G and ABC was defined when the variant, regulatory element, and predicted target gene were consistent across both methods. Alternative predicted target genes were also recorded to evaluate target-gene specificity.

Sequence-to-function analyses

The same 27 candidate variants were evaluated using REF/ALT sequence predictions from AlphaGenome, Borzoi, and Enformer.

AlphaGenome analyses were performed using the official SDK v0.9.0 and the GRCh38 reference genome to evaluate predicted effects of candidate variants on chromatin, transcriptional, and RNA-related molecular phenotypes. Borzoi analyses used four official pretrained model replicates, with REF and ALT predictions compared for each candidate variant. Enformer analyses used the official v1 model to assess local regulatory effects and *TNRC6B*-associated transcriptional predictions.

Directional comparisons across models were performed only when the assay type, tissue or cellular context, and evaluated genomic region could be reasonably matched. For 3'UTR candidates, additional analyses incorporated ENCODE eCLIP data, miRNA annotations, evolutionary conservation, and TF/RNA motif information to evaluate potential RNA-level regulatory effects.

Multi-evidence integration and robustness analyses

Evidence for each candidate variant was organized into seven predefined modules: statistical genetics, functional annotation, QTL, enhancer-to-gene mapping, sequence models, motif/RNA evidence, and target-gene specificity. AlphaGenome, Borzoi, and Enformer were grouped within a single sequence-model module, whereas rE2G and ABC were grouped within a single enhancer-to-gene module to reduce overcounting of methodologically related evidence.

Variants were classified into Tiers A–D according to module-level support. Tier A required moderate or strong support from at least three distinct evidence modules, including evidence from either the QTL or enhancer-to-gene module. Tier B variants were supported by at least two modules but showed missing or conflicting evidence in other domains. Tier C variants were supported primarily by a single evidence category, whereas Tier D variants lacked clear *TNRC6B*-specific support.

Leave-one-module-out and leave-one-model-out analyses were performed to assess whether variant prioritization depended disproportionately on any single evidence source. Robustness was further evaluated using 5,000 random module-weight perturbations, from which the frequencies of each candidate ranking within the top 1 and top 3 were calculated. FDR correction was applied to large-scale association analyses, with FDR < 0.05 used as the principal multiple-testing threshold where applicable. All directional comparisons were performed only after harmonization of REF/ALT alleles and effect alleles.

Declarations

Ethics statement

This study used only publicly available, de-identified GWAS, QTL, and functional genomic data and involved no new recruitment of human participants or collection of biological specimens. Ethical approval and informed consent for the original FinnGen, GTEx, and other source studies were obtained by the respective study investigators.

Competing interests

The authors declare no competing interests.

Author Contribution

J.P. conceived the study, performed computational analyses, interpreted the results, and drafted the manuscript. G.C., X.G., P.Z., A.Y., K.W., Q.Y., and W.T. contributed to data interpretation and manuscript revision. K.C. supervised the study, acquired funding, and critically revised the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgements

We acknowledge the participants and investigators of the FinnGen study, the GTEx Consortium, the eQTL Catalogue, the ENCODE Consortium, and the 1000 Genomes Project for making the genomic and functional genomic resources used in this study publicly available. This work was supported by the Guizhou Provincial Key Laboratory of Microbiome and Infectious Disease Prevention and

Control (ZDSYS[2023]004); the Basic and Applied Research on Important Pathogenic Microorganisms Program (Qian Ke He Zhong Yin Di [2025]024); the Guizhou Provincial Science and Technology Innovation Talent Workstation for Traditional Chinese and Ethnic Medicines (Qian Ke He Ping Tai KXJZ [2024]034); and the National Natural Science Foundation of China (Grant No. 82661001).

Data Availability

The GWAS summary statistics for erysipelas used in this study were obtained from FinnGen Release 11 for phenotype AB1_ERYSIPELAS. Gene-expression and QTL resources were obtained from GTEx v8 and the eQTL Catalogue. Functional genomic annotations were obtained from ENCODE, and linkage disequilibrium reference data were derived from the 1000 Genomes Project Phase 3 reference panels. Additional regulatory and enhancer-to-gene resources used in the analyses, including rE2G and Activity-by-Contact (ABC) reference maps, are publicly available from their respective repositories or project websites. All datasets analyzed in this study are publicly available and are described in the Methods and Supplementary Table S12. Processed result tables supporting the findings of this study are provided in the Supplementary Information.

References

1. Raff, A. B. & Kroshinsky, D. Cellulitis: a review. *JAMA* **316**, 325–337 (2016).
2. Stevens, D. L. & Bryant, A. E. Severe group a streptococcal infections. in Streptococcus pyogenes: basic biology to clinical manifestations (eds (eds Ferretti, J. J., Stevens, D. L. & Fischetti, V. A.) (University of Oklahoma Health Sciences Center, (2016). Oklahoma City (OK).
3. Quirke, M. et al. Risk factors for nonpurulent leg cellulitis: a systematic review and meta-analysis. *Br. J. Dermatol.* **177**, 382–394 (2017).
4. Hannula-Jouppi, K. et al. Genetic susceptibility to non-necrotizing erysipelas/cellulitis. *PLOS One*. **8**, e56225 (2013).
5. Emene, C. C., Kravchenko, I. E., Aibatova, G. I. & Rizvanov, A. A. Analysis of serum cytokines and single-nucleotide polymorphisms of SOD1, SOD2, and CAT in erysipelas patients. *J. Immunol. Res.* 2157247 (2017). (2017).
6. Kadkina, V. V. et al. The role of CYP1A1 gene polymorphism in patients with erysipelas. *BioNanoScience* **7**, 648–653 (2017).
7. Stappers, M. H. T. et al. TLR1, TLR2, and TLR6 gene polymorphisms are associated with increased susceptibility to complicated skin and skin structure infections. *J. Infect. Dis.* **210**, 311–318 (2014).
8. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
9. Gallagher, M. D. & Chen-Plotkin, A. S. The post-GWAS era: from association to function. *Am. J. Hum. Genet.* **102**, 717–730 (2018).
10. Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
11. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
12. Wainberg, M. et al. Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* **51**, 592–599 (2019).
13. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
14. Weeks, E. M. et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat. Genet.* **55**, 1267–1276 (2023).
15. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLOS Genet.* **10**, e1004383 (2014).
16. Mountjoy, E. et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021).
17. Gschwind, A. R. et al. An encyclopedia of human enhancer-gene regulatory interactions. *Nature* **657**, 179–190 (2026).
18. Fulco, C. P. et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
19. Avsec, Ž. et al. Advancing regulatory variant effect prediction with AlphaGenome. *Nature* **649**, 1206–1218 (2026).
20. Linder, J., Srivastava, D., Yuan, H., Agarwal, V. & Kelley, D. R. Predicting RNA-seq coverage from DNA sequence as a unifying model of gene regulation. *Nat. Genet.* **57**, 949–961 (2025).

21. Avsec, Ž. et al. Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods*. **18**, 1196–1203 (2021).
22. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16**, 421–433 (2015).
23. Tritschler, F., Huntzinger, E. & Izaurralde, E. Role of GW182 proteins and PABPC1 in the miRNA pathway: a sense of déjà vu. *Nat. Rev. Mol. Cell. Biol.* **11**, 379–384 (2010).
24. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
25. Gruber, A. J. & Zavolan, M. Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.* **20**, 599–614 (2019).
26. Mayr, C. What are 3′ UTRs doing? *Cold Spring Harbor Perspect. Biol.* **11**, a034728 (2019).
27. GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
28. Zhou, J. et al. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat. Genet.* **50**, 1171–1179 (2018).
29. Turner, M. & Petkau, G. RNA-binding proteins and ribonucleoproteins as determinants of immunity. *Nat. Rev. Immunol.* **26**, 382–403 (2026).
30. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
31. Cowen, L., Ideker, T., Raphael, B. J. & Sharan, R. Network propagation: a universal amplifier of genetic associations. *Nat. Rev. Genet.* **18**, 551–562 (2017).
32. Yao, D. W., O'Connor, L. J., Price, A. L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat. Genet.* **52**, 626–633 (2020).
33. Fairfax, B. P. et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).

Figures

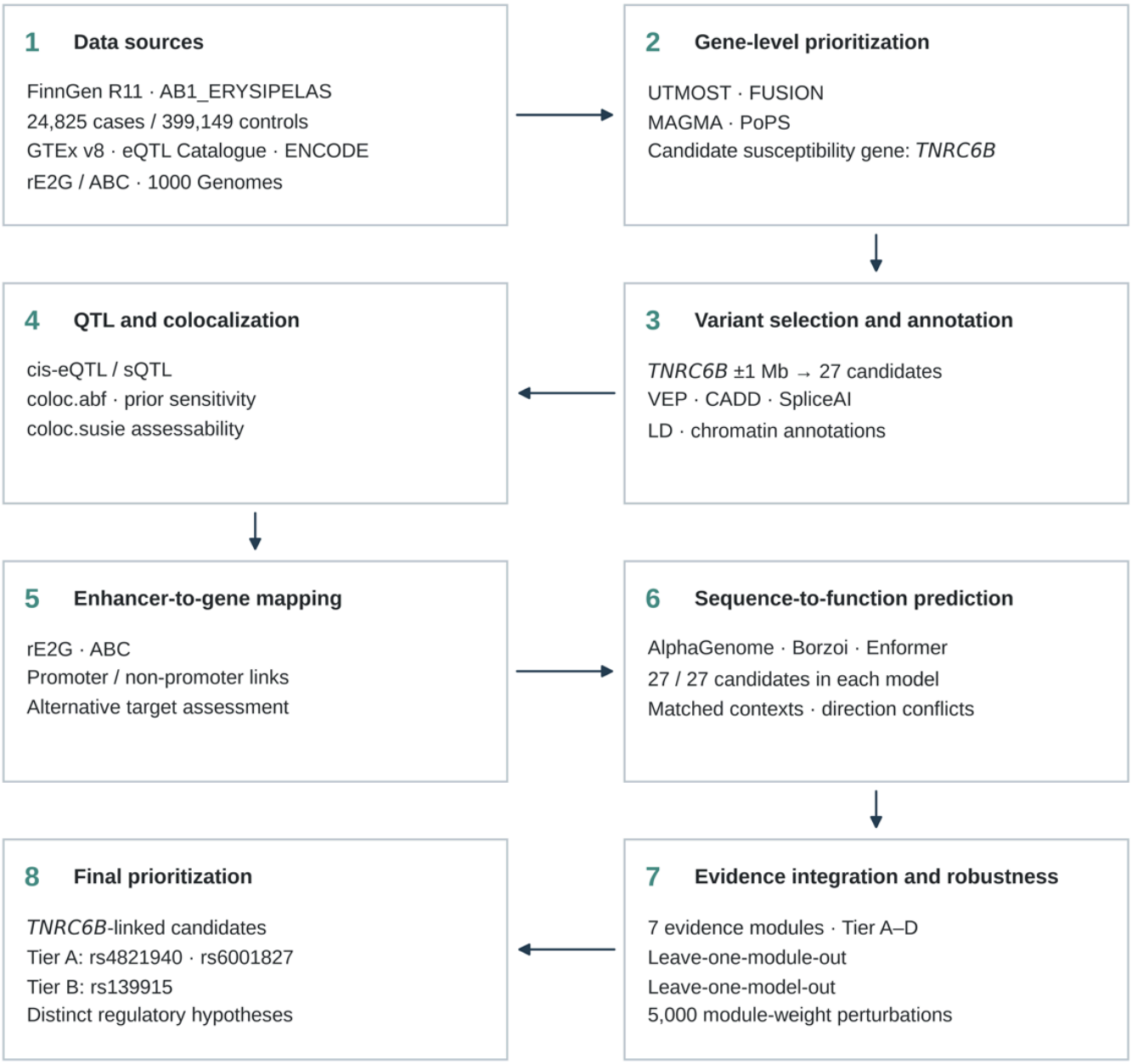


Figure 1

Overall analytical workflow for prioritizing TNRC6B-linked regulatory variants in erysipelas susceptibility.

The study integrated FinnGen R11 erysipelas GWAS data with GTEx v8, the eQTL Catalogue, ENCODE functional genomic resources, and the 1000 Genomes reference panel. Gene-level association and prioritization were first performed using UTMOST, FUSION, MAGMA, and PoPS. Variants within the *TNRC6B* ±1 Mb region were subsequently evaluated by regional variant screening, functional annotation, linkage disequilibrium (LD) analysis, QTL analysis, and colocalization, followed by enhancer-to-gene mapping using rE2G and Activity-by-Contact (ABC). Candidate variants were further assessed using the sequence-to-function models AlphaGenome, Borzoi, and Enformer. Evidence from statistical genetics, functional annotation, QTL analyses, enhancer-to-gene mapping, sequence models, motif/RNA analyses, and target-gene specificity was then integrated at the module level. Robustness of variant prioritization was evaluated using leave-one-module-out analysis, leave-one-model-out analysis, and random weight perturbation.

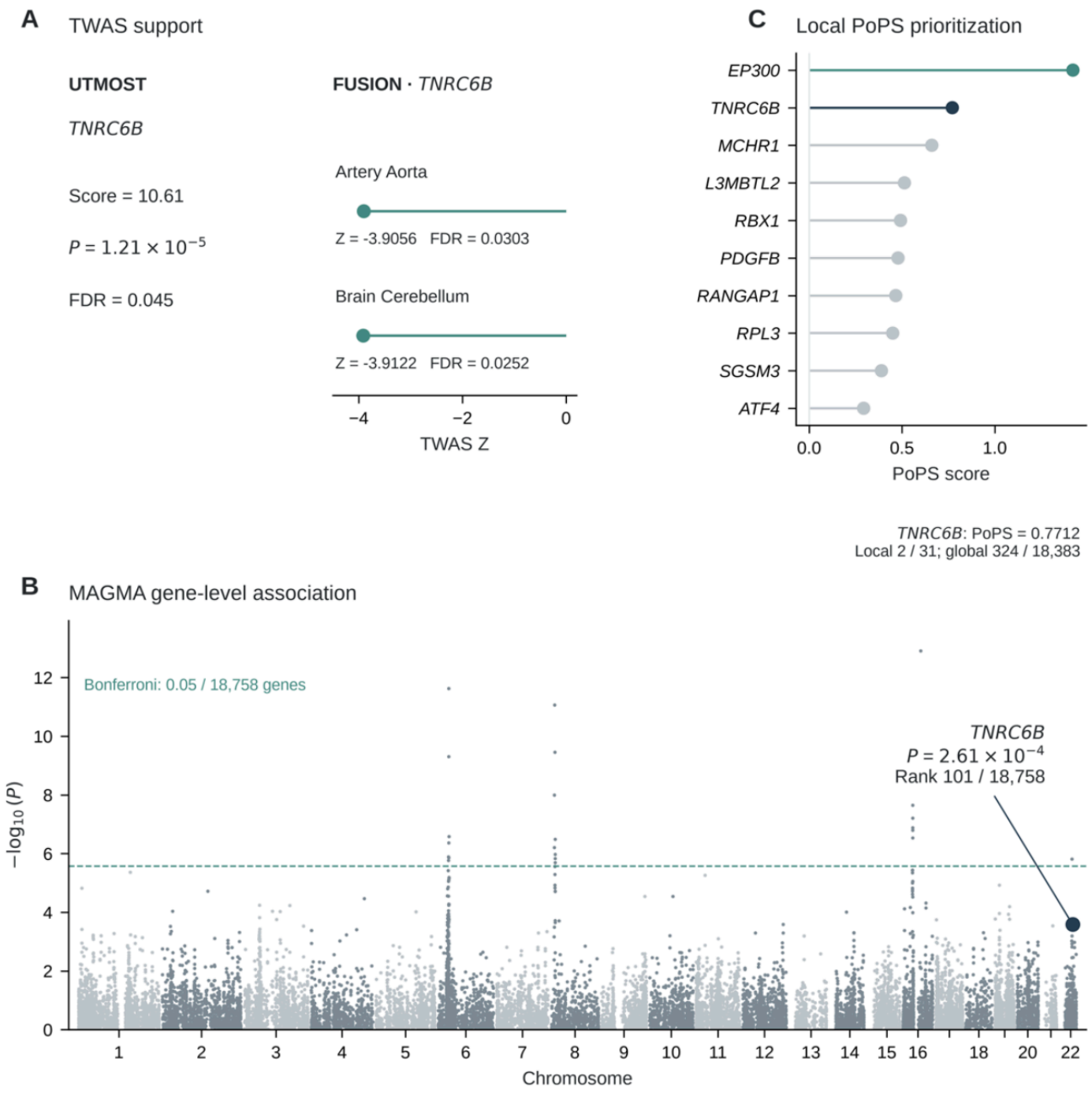


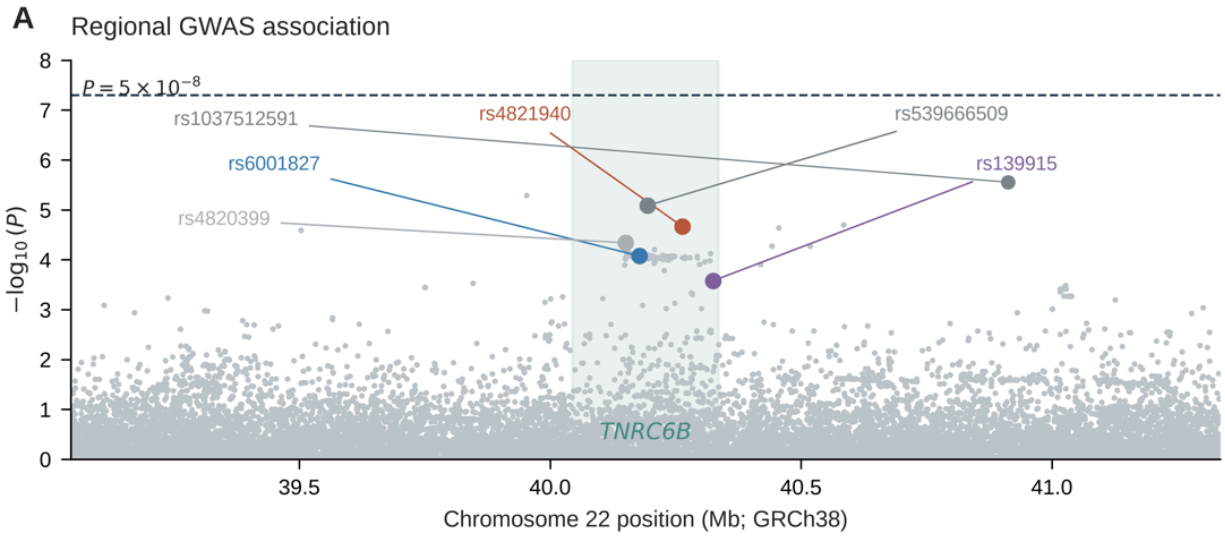
Figure 2

Multi-layer gene-level analyses prioritize *TNRC6B*.

A,Cross-tissue UTMOST and tissue-specific FUSION results for *TNRC6B*. In UTMOST, *TNRC6B* showed a test score of 10.61 ($P = 1.21 \times 10^{-5}$, FDR = 0.045). Tissue-specific FUSION analysis further identified significant associations for *TNRC6B* in Artery Aorta and Brain Cerebellum.

B,Genome-wide MAGMA gene-based association analysis. *TNRC6B* had Z = 3.4697 and $P = 2.61 \times 10^{-4}$, ranking 101st among 18,758 genes. The dashed line indicates the Bonferroni-corrected significance threshold.

C, PoPS prioritization within the *TNRC6B* ±1 Mb region. *TNRC6B* had a PoPS score of 0.7712 and ranked 2nd among 31 local genes, second only to *EP300*; its genome-wide rank was 324 of 18,383 genes.



B Focal-variant annotation

Variant	GWAS P	VEP context	cCRE	FIN / EUR MAC
rs4821940	2.16×10^{-5}	Intronic	Not detected	MAC 95 / 494
rs6001827	8.42×10^{-5}	Intronic	pELS + CTCF	MAC 76 / 436
rs139915	2.64×10^{-4}	3′UTR	Not detected	MAC 71 / 452
rs539666509	8.20×10^{-6}	Intronic	dELS + CTCF	Low AF; MAC 0 / 1
rs4820399	4.59×10^{-5}	Intronic	Not detected	MAC 76 / 441

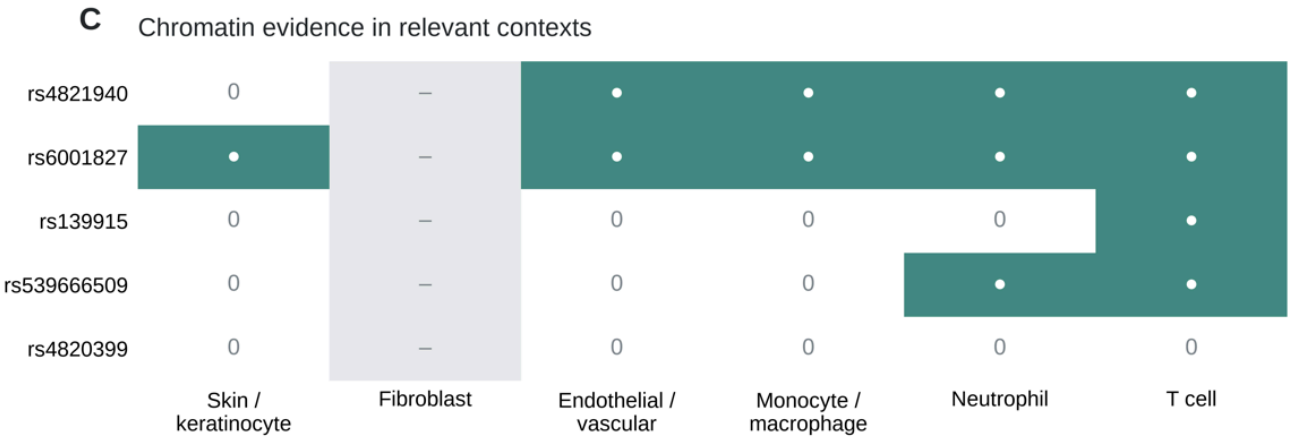


Figure 3

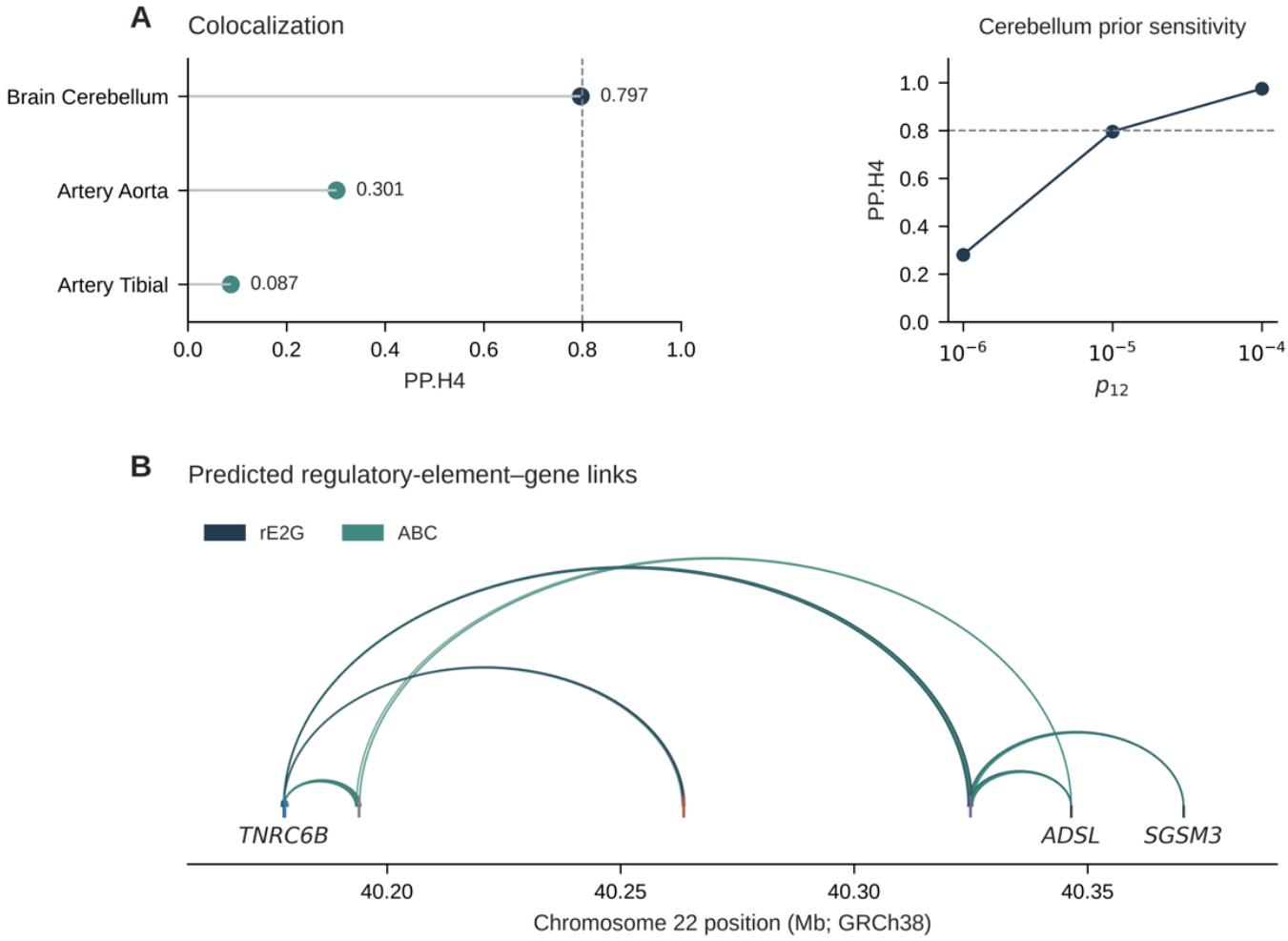
Genetic association and regulatory annotation across the *TNRC6B*locus.

A, FinnGen GWAS association profile across the *TNRC6B* ±1 Mb region. Key candidate variants, including rs539666509, rs4821940, rs4820399, rs6001827, and rs139915, are highlighted. The dashed line indicates the genome-wide significance threshold (P = 5 × 10⁻⁸). No variant within the region reached genome-wide significance.

B, Functional annotation of the five prioritized candidate variants, including GWAS P values, VEP consequences, cCRE annotations, and minor allele counts in the FIN and EUR reference panels.

C, Chromatin regulatory evidence for the candidate variants across biologically relevant cellular contexts, including skin/keratinocyte, fibroblast, endothelial/vascular, monocyte/macrophage, neutrophil, and T-cell contexts.

cCRE, candidate cis-regulatory element; MAC, minor allele count.



C Target specificity and mechanism

Variant	Non-promoter	Promoter	Interpretation
rs4821940	35	0	*TNRC6B* favored; non-promoter
rs6001827	0	2829	*TNRC6B* promoter / self-promoter
rs139915	388	0	Alternative targets: *ADSL*, *SGSM3*
rs539666509	143	0	*TNRC6B* links; low AF / MAC
rs4820399	0	0	No *TNRC6B*-specific E2G support

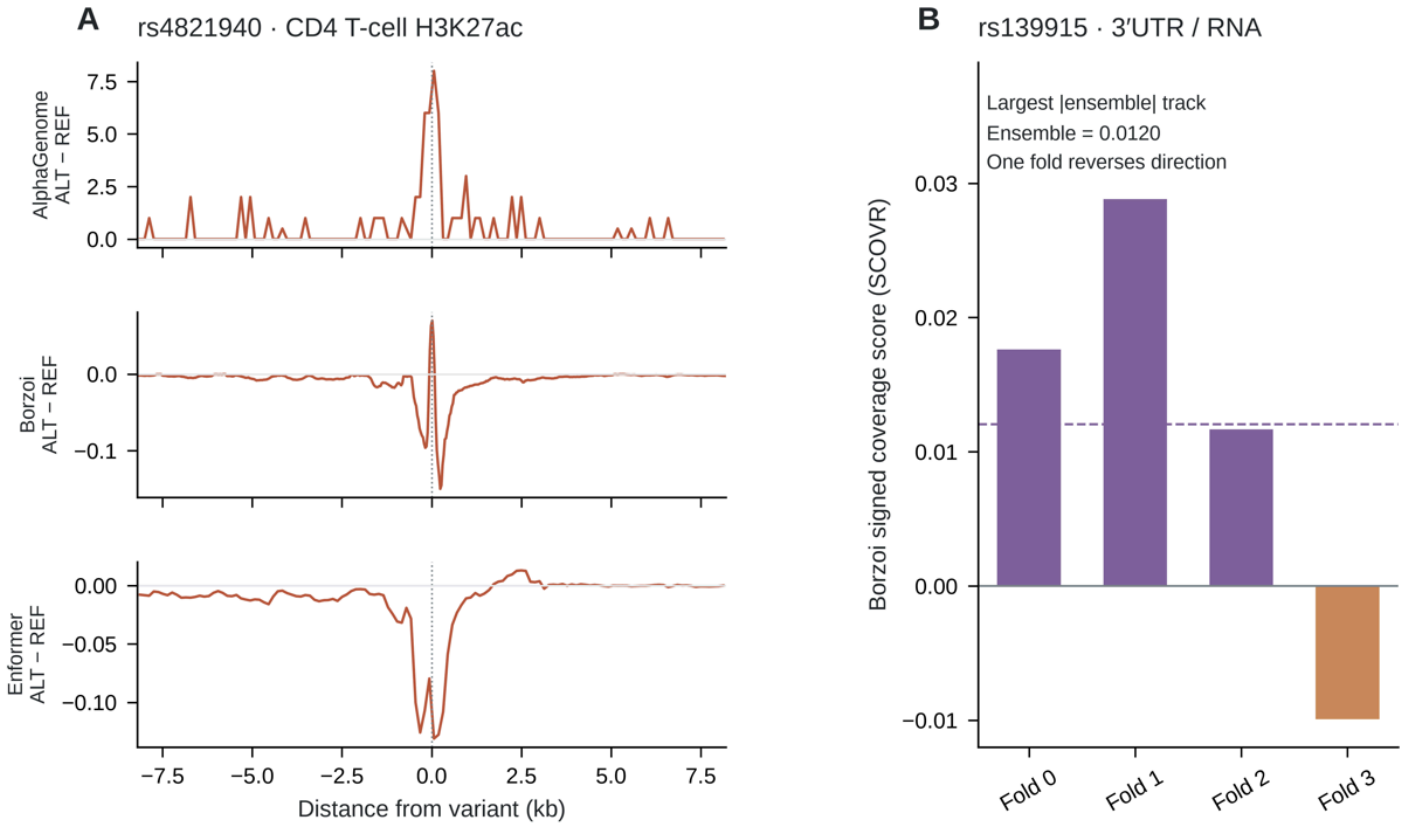
Figure 4

QTL and enhancer-to-gene analyses reveal context-dependent regulatory links involving *TNRC6B*.

A,Coloc.abf results comparing *TNRC6B* eQTL signals with the erysipelas GWAS in Brain Cerebellum, Artery Aorta, and Artery Tibial. Brain Cerebellum showed the highest PP.H4 (0.797). The dashed line indicates PP.H4 = 0.8, and the panel on the right shows sensitivity of the Brain Cerebellum result to the prior probability of a shared causal variant (p12).

B,Regulatory element-to-gene links predicted by rE2G and ABC across the *TNRC6B*locus. Connections represent computationally predicted links between candidate regulatory elements and *TNRC6B* or alternative target genes.

C,Target-gene specificity and promoter versus non-promoter regulatory patterns for the five prioritized candidate variants. rs4821940 predominantly showed non-promoter regulatory links to *TNRC6B*, whereas rs6001827 was characterized mainly by promoter/self-promoter regulatory support. rs139915 showed additional predicted links to alternative targets, including *ADSL*and *SGSM3*, indicating target-gene ambiguity.



C Focal-variant model summary

Variant	AlphaGenome	Borzoi	Enformer	Overall interpretation
rs4821940	Present	Present	Present	Partial; context-dependent
rs6001827	Present	Present	Present	Promoter / transcription-related
rs139915	Present	Present	Present	RNA / polyA; model-sensitive
rs539666509	Present	Present	Present	Model conflict; low AF / MAC
rs4820399	Present	Present	Present	Local effects; no *TNRC6B* E2G

Figure 5

Sequence-to-function models reveal context-dependent regulatory effects of prioritized candidate variants.

A,Allele-dependent sequence-model predictions surrounding rs4821940, showing REF/ALT differences predicted by AlphaGenome, Borzoi, and Enformer within matched regulatory contexts.

B,RNA/polyA-related predictions for the *TNRC6B* 3'UTR variant rs139915. The predicted direction showed some inconsistency across Borzoi model replicates.

C,Integrated cross-model interpretation of the five prioritized candidate variants. rs4821940 showed predominantly partial and context-dependent support; rs6001827 was characterized mainly by promoter- and transcription-related effects; rs139915 showed RNA/polyA-related but model-sensitive predictions; rs539666509 exhibited cross-model conflict; and rs4820399 showed local sequence-level effects without *TNRC6B*-specific enhancer-to-gene support.

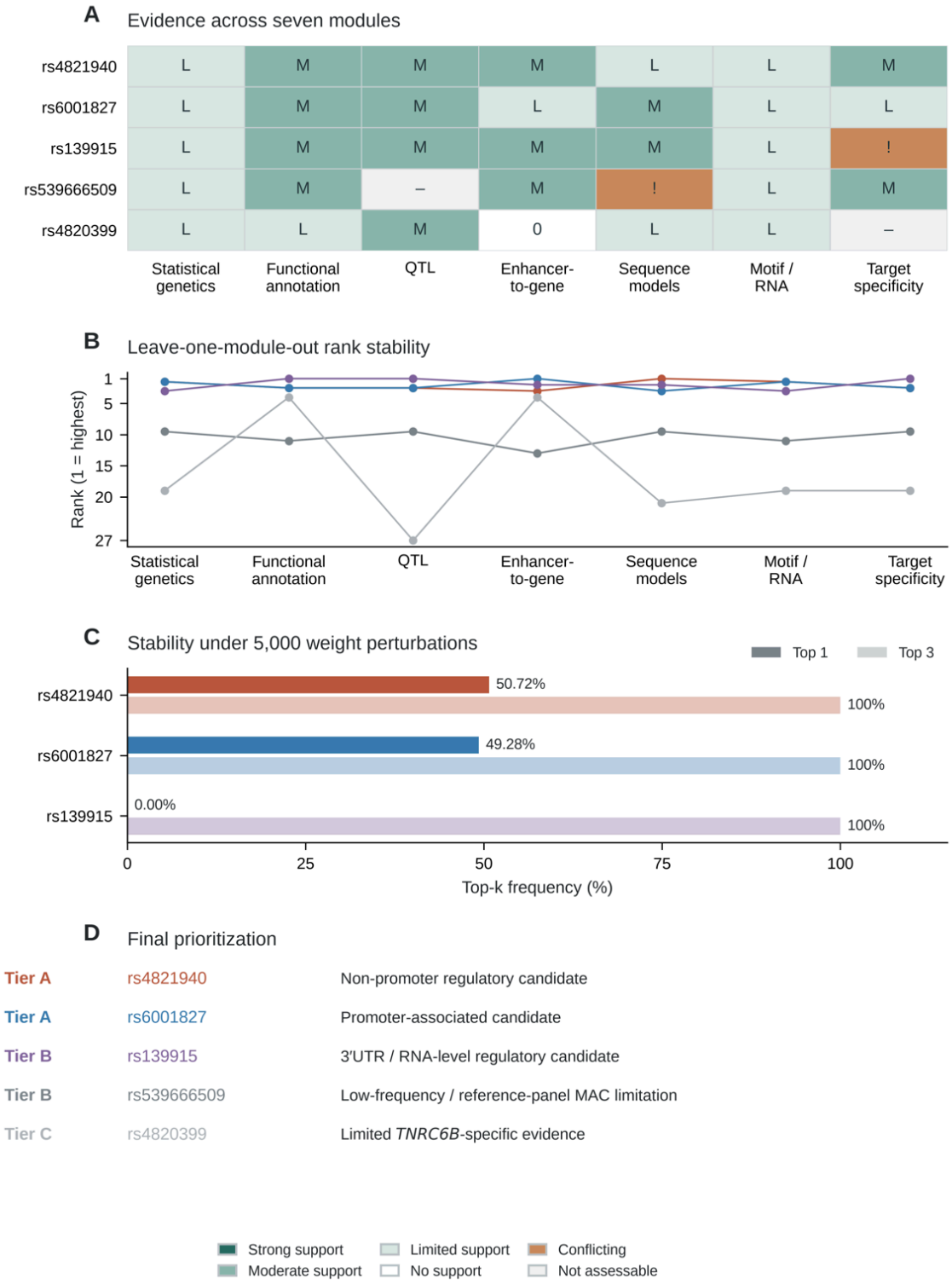


Figure 6

Multi-evidence integration and robustness analyses prioritize TNRC6B-linked candidate variants.

Symbols **M**, **L**, **0**, **!**, and **–** denote moderate support, limited support, no support, conflicting evidence, and not assessable, respectively.

A, Support levels for the five prioritized variants across seven evidence modules: statistical genetics, functional annotation, QTL, enhancer-to-gene mapping, sequence models, motif/RNA evidence, and target-gene specificity.

B, Changes in candidate-variant rankings in the leave-one-module-out analysis.

C, Top-1 and top-3 frequencies of rs4821940, rs6001827, and rs139915 across 5,000 random module-weight perturbations.

D, Final integrated prioritization. rs4821940 and rs6001827 were classified as Tier A candidates, representing non-promoter regulatory and promoter-associated regulatory hypotheses, respectively, whereas rs139915 was classified as Tier B and represented a distinct 3'UTR/RNA-level regulatory hypothesis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFiles.zip
- SupplementaryTables.xlsx
- SupplementaryFigure.zip
- FigS1extendedgenelevel.tiff
- FigS2fulle2glandscape.tiff
- FigS3extendedsequencemodels.tiff
- FigS4crossmodelconsistency.tiff
- FigS5fullintegrationrobustness.tiff
- FigS6MAGMAcelltyping.tiff
- FigS7RWRsensitivity.tiff
- FigS8MESC.tiff