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.

Host genetic control of gut fungal genera: genus-specific regulation of Malassezia, Saccharomyces and Aspergillus abundance, and overlap with autoimmune liver disease loci

Jian Sun[1]

psbyyffs@126.com

The Affiliated Hospital of Jiangnan University

Minli Liu[1]

The Affiliated Hospital of Jiangnan University

Xiaofan Xu[1]

The Affiliated Hospital of Jiangnan University

Yifan Jin[1]

The Affiliated Hospital of Jiangnan University

Shijing Gao[1]

The Affiliated Hospital of Jiangnan University

Research Article

Keywords: cis-eQTL, colocalization, genetic overlap, MAGMA, Malassezia, mycobiome, primary biliary cholangitis, primary sclerosing cholangitis

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11478534/v1

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

Additional Declarations: No competing interests reported.

Host genetic control of gut fungal genera: genus-specific regulation of *Malassezia*, *Saccharomyces* and *Aspergillus* abundance, and overlap with autoimmune liver disease loci

Jian Sun[1]*, Minli Liu[1], Xiaofan Xu[1], Yifan Jin[1], Shijing Gao[1]

[1]Department of Rheumatology and Immunology, The Affiliated Hospital of Jiangnan University, Wuxi 214122, Jiangsu, China

* Correspondence: Jian Sun, psbyyffs@126.com; ORCID: 0009-0004-2696-6273

Abstract

Background: How much of gut fungal abundance is under host genetic control, and whether the same loci contribute to autoimmune liver disease (AILD), is not well mapped.

Methods: We used gut mycobiome genome-wide association study (GWAS) summary statistics (N = 7,350) for *Malassezia*, *Saccharomyces* and *Aspergillus* abundance, and AILD GWAS for primary biliary cholangitis (PBC, N = 24,510), primary sclerosing cholangitis (PSC, N = 14,890) and autoimmune hepatitis (AIH, N = 485,234). Gene-based, gene-set and cell-type tests were run in MAGMA/FUMA with Bonferroni correction; colocalization, regional-overlap and cis-eQTL matching linked fungal and AILD loci under a triangulation framework.

Results: Host control was genus-specific: the two yeasts carried strong gene-level signal (*Malassezia* 85, *Saccharomyces* 70 significant genes) with non-overlapping

pathway signatures (intracellular processing versus leukocyte recruitment), while *Aspergillus* had none at gene level. On the disease side, PBC returned 381 significant genes and 217 significant gene sets led by adaptive-immunity themes, PSC returned 20 gene sets led by MHC class II antigen presentation, cell-type analysis placed *Malassezia* signals in myeloid and AILD signals in lymphoid compartments; no fungal genus showed gene-level signal in the HLA region. Drug-gene annotation identified SYK (fostamatinib, approved) as the most actionable node of the antifungal-recognition pathway. No fungal-AILD window reached formal colocalization (maximum PP.H4 = 0.16, reflecting limited power). One locus showed single-SNP concordance (chr4:130.2 Mb, C4orf33/SCLT1/PHF17), one strong cis-eQTL support (chr11, AIP cis-eQTL up to $|Z| = 15.0$), and one regional overlap alone (chr14). AIH yielded no testable window (insufficient variant overlap).

Conclusions: Gut fungal abundance exhibits genus-specific host genetic regulation. This first locus-level joint genetic analysis of gut mycobiome and AILD identifies prioritised candidate loci and immune pathways: two prioritised loci (chr4 with single-SNP triangulation, chr11 marked by strong AIP cis-eQTL) and the nominally associated IL-23R immune node; the gene-level structure is robust to ancestry-matched reanalysis, whereas SUMO pathway replication is panel-dependent. These candidates define where fungal and cholangitis genetics meet and are ready for functional follow-up. SYK-fostamatinib is the one annotated axis with an approved drug, untested in AILD; replication in ancestry-matched cohorts is pending.

***Keywords*:** cis-eQTL; colocalization; genetic overlap; MAGMA; *Malassezia*; mycobiome; primary biliary cholangitis; primary sclerosing cholangitis

Background

The gut mycobiome is a component of the intestinal microbiota. Fungi are far less abundant than bacteria but interact directly with the intestinal epithelium and the mucosal immune system [1,2,3,4]. Cohort studies show that gut fungal composition varies between individuals and shifts with disease state: in inflammatory bowel disease, fungal strain diversity and T-cell responses against commensal yeasts contribute to inflammation [5,6,7,8]; in primary sclerosing cholangitis (PSC), increased faecal fungal diversity and a reduced proportion of *Saccharomyces* have been reported [9,10]. These observations raise a question that clinical data alone cannot settle. Is the fungal change a cause or a consequence of disease? Host genetics can test precisely the shared-genetic-basis part of this question, and that is the part this study addresses. Host genetics offers one route to an answer. Evidence that host genotype shapes the gut microbiome is strongest for bacteria: genome-wide association studies (GWAS) and meta-analyses in cohorts from several thousand to tens of thousands of participants have linked the abundance of multiple bacterial taxa to human loci [11,12]. Evidence on the fungal side is much thinner. Existing gut-mycobiome GWAS cover only a few thousand samples [13], and a proof-of-concept fungal GWAS based on 125 participants of the Human Microbiome (HMP) cohort has reported genus-level

fungal-host associations [14]. None has been analysed jointly with the genetics of autoimmune liver disease.

The genetic landscape of autoimmune liver disease (AILD) is well defined. GWAS of primary biliary cholangitis (PBC) established an architecture centred on the MHC region and adaptive-immune genes [15,16]; a multitrait GWAS of PSC added new susceptibility loci to the roughly 19 previously reported and nominated candidate drug directions [17]. PSC and inflammatory bowel disease share genetic components, quantified in a genome-wide genetic-correlation analysis [18] and reviewed in the context of the gut-liver axis [19]. The three AILDs differ in target and pathology — PBC the small intrahepatic ducts, PSC the large ducts with frequent inflammatory bowel disease comorbidity, AIH the hepatocyte — yet all three share an adaptive-immune-centred genetic architecture, which makes them suitable as one spectrum against which to test fungal-side convergence.

Between these two fields lies a gap. Existing microbiota-disease causal studies use Mendelian randomization with bacterial exposures in the microbiota → disease direction [20,21]; this study instead tests host-locus-level genetic overlap. To our knowledge, no study has asked which human loci control both gut fungal colonization and AILD susceptibility; locus-level host-genetic evidence connecting fungi to AILD is missing, and no framework places the fungal-side and disease-side mechanisms in the same analysis.

This study uses gut-mycobiome GWAS summary statistics (N = 7,350; the genera *Malassezia* (yeast), *Saccharomyces* and *Aspergillus*) and AILD GWAS (PBC, N = 24,510; PSC, N = 14,890; AIH, multi-ancestry) with three aims: (1) map host genetic control of the three genera at gene, pathway and cell-type level; (2) test whether fungal and AILD GWAS share signal at the locus level, graded by evidence strength; and (3) localize candidate loci and genes through eQTL bridging and triangulation. Gene and pathway analyses used FUMA/MAGMA [22,23], colocalization used coloc [24], and expression-regulation evidence came from the eQTLGen whole-blood cis-eQTL resource [25]. The fungal cohort is Chinese while the PBC and PSC cohorts are European and the AIH source is multi-ancestry; this cross-ancestry design constrains the analyses throughout and is addressed in Methods and Limitations.

Methods

Study data

Fungal abundance GWAS summary statistics came from a gut mycobiome genome-wide association study of 7,350 individuals of Chinese ancestry [13]. The fungal cohort does not overlap with any of the three AILD GWAS (European-ancestry PBC and PSC; multi-ancestry AIH); cross-dataset matching was by variant position only. Eight taxa were tested; here we report the three genera with the clearest host-genetic signal or relevance: *Malassezia* (767,569 variants passing quality control), *Saccharomyces* and *Aspergillus*. The locus-level screen (colocalization and regional overlap) covered

all eight taxa, so the ranked candidate list in Results additionally includes *Candida* and a Shannon diversity index. AILD GWAS: PBC (N = 24,510, 8,021 cases; GWAS Catalog GCST90061440), PSC (N = 14,890, 2,871 cases; IEU OpenGWAS ieu-a-1112) and AIH (N = 485,234, 821 cases; GWAS Catalog GCST90018785, from a cross-population atlas of 220 phenotypes). Owing to its extreme imbalance (case proportion 0.17%, 821/485,234), AIH was not included in gene-based or gene-set analysis (its effective sample size is comparable to a balanced design of $N \approx 3,300$) and entered the locus-level colocalization and regional-overlap screening (no AIH window met the 50-variant overlap requirement; Results). Fungal files carry position-based variant identifiers; the effect allele is the alternate allele. All coordinates are GRCh37. Upstream sample- and genotype-level quality control follows the source studies; our processing is limited to coordinate unification, effect-allele harmonization and cross-dataset position matching. Whether the source study adjusted for bacterial composition, diet, medication or other host covariates (such as age, sex, body mass index and ancestry principal components) cannot be verified from the summary statistics and is addressed in the Limitations section. Because the fungal GWAS is small (N = 7,350), strict linkage disequilibrium (LD) clumping would leave too few independent loci. We therefore ranked lead SNPs by P value (up to 50 per trait) without clumping and interpreted overlapping windows separately, while acknowledging their correlation. A total of 485 fungal lead SNPs passed the $P < 1 \times 10^{-5}$ screen across the eight taxa; the coloc screen used the top 50 per trait (340 leads

retained after the ≥50-overlapping-variant window filter; each tested against PBC and PSC, giving the 680 windows; Methods, Locus-level colocalization and eQTL). A sensitivity analysis using greedy distance pruning (≥1 Mb within chromosome, per taxon, ranked by fungal P value) applied to the 485 pre-cap leads retained 144; both the chr4:130.2 Mb and chr11:67.2 Mb leads survived pruning (Additional file 10: Table S9). Gene-level results for all eight taxa are summarized in Additional file 9: Table S8.

Table 1. GWAS data sources.

Trait	Source	N (total / cases)	Variants tested	Role in this study
Malassezia abundance	Gut mycobiome GWAS	7,350 / —	767,569	Fungal trait, primary
Saccharomyces abundance	Gut mycobiome GWAS	7,350 / —	≈3.4 M	Fungal trait, comparison
Aspergillus abundance	Gut mycobiome GWAS	7,350 / —	7,811,724	Fungal trait, comparison
PBC	GCST90061440	24,510 / 8,021	Not recorded in this analysis	AILD outcome
PSC	IEU OpenGWAS ieu-a-1112	14,890 / 2,871	Not recorded in this analysis	AILD outcome
AIH	EBI GCST90018785 (PMID 34594039)	485,234 / 821	Not recorded in this analysis	AILD outcome

Note. All fungal traits come from one cohort, so differences between genera more likely reflect biology than genotyping differences. The PBC source is the GCST90061440 European meta-analysis.

Gene-based and gene-set analysis

We ran MAGMA v1.08 through FUMA v1.8.2 with default gene boundaries (35 kb upstream, 10 kb downstream) and the 1000 Genomes EUR reference panel. Gene-based P values were computed as one-sided tests. The number of genes and sets tested differs slightly between traits and analysis batches because FUMA's gene annotation is versioned; the ancestry-panel sensitivity analysis (Discussion) supports the robustness of the gene-level structure. Gene-set analysis used the FUMA default collection (MSigDB curated and GO sets; 16,419 sets for *Malassezia*, 16,447 for PBC). Significance used Bonferroni correction within each analysis: 0.05 divided by the number of genes or sets tested. The LD reference panel is the FUMA default 1000 Genomes EUR; because the fungal GWAS is Chinese, the panel does not fully match the fungal LD structure, a limitation addressed in Limitations.

Locus-level colocalization and eQTL

For each fungal trait we took the lead SNPs with $P < 1 \times 10^{-5}$ (up to 50 per trait; the triangulation threshold for the fungal P leg is stricter, 5×10^{-6}; Methods, Significance thresholds) and ran coloc.abf [26] (R package coloc v5.2.3 [24]) in a ±500 kb window around each lead SNP against each AILD GWAS, matching variants on chromosome and position. Default coloc priors were used ($p1 = p2 = 1 \times 10^{-4}$, $p12 = 1 \times 10^{-5}$). Regional-overlap candidates are defined as lead SNPs with AILD $P < 0.05$ within the same ±500 kb window; candidates are ranked by the colocalization posterior probability PP.H4 (the probability that the two traits share a causal variant within the

window), the ranking serving as priority rather than confirmation. Fungal traits were modelled as quantitative (type = "quant", N = 7,350); AILD traits as case-control (type = "cc") with the sample fraction set to the study case fraction. Windows with fewer than 50 overlapping variants were skipped, leaving 680 tested windows. The conventional colocalization threshold is PP.H4 ≥ 0.8; we regard PP.H4 ≥ 0.75 as supportive, with PP.H3 as the main competing hypothesis. Expression quantitative trait loci were queried from the eQTLGen whole-blood cis-eQTL scan (N = 31,684); candidate-region SNP-gene pairs were retained at false discovery rate (FDR) < 0.05. Because linked lead SNPs can produce correlated windows, the candidate list should be interpreted as a priority ranking of loci rather than a collection of independent tests.

Cell-type specificity analysis

We performed MAGMA cell-type-specific analysis through FUMA using eight blood and bone-marrow single-cell expression reference datasets (GSE89232, Travaglini 2020, Xu 2023 blood and bone marrow, MouseCellAtlas bone marrow and peripheral blood, TabulaMuris marrow FACS and droplet). All reference datasets are peripheral blood/bone-marrow derived, which differ in composition from gut-resident myeloid cells and hepatic immune cells; the enrichment pattern is an approximation of the peripheral immune compartment. Gene-level mean expression per cell compartment (log2-transformed) served as the gene-property covariate in a one-sided test (greater). Correction was applied within each dataset (Bonferroni over the number of cell types tested); results were not combined across datasets.

Drug-gene annotation

Candidate genes were queried against the Drug-Gene Interaction Database (DGIdb 4.0 [27]; queried in September 2026) for interactions with existing clinical or preclinical evidence. Drug lists were used for annotation only and imply no efficacy. Candidate genes were also compared with the candidate drug directions reported by the PSC multitrait GWAS [17].

Significance thresholds

Triangulation grades loci by the number of criteria passed at one SNP: all three criteria at a single SNP constitute tier-1 evidence; regional overlap plus a strong cis-eQTL constitutes tier-2 (the AILD leg may be carried by a linked variant within the same window, as at chr11); regional overlap alone constitutes tier-3. The tiers are interpretive grades calibrated to the available power, not independent confirmatory tests.

Significance thresholds were: Bonferroni correction for gene-based and gene-set analyses (0.05 divided by the number of tests); colocalization posteriors interpreted against the conventional PP.H4 ≥ 0.8 threshold (Methods, Locus-level colocalization and eQTL), with PP.H3 as the main competing hypothesis; triangulation criteria [28] of fungal $P < 5 \times 10^{-6}$, cis-eQTL $|Z| > 5$ and AILD $P < 0.05$ (nominal, uncorrected; a random-position permutation of the nearest-variant criterion is reported in Results (Locus-level overlap), and false-positive control relies on evidence grading rather than any single threshold); and within-dataset Bonferroni correction for cell-type analysis.

AI-assisted language-editing tools were used during manuscript preparation; all content was reviewed and approved by the authors.

Results

Gene-level association: significant signal in two yeast genera, null in Aspergillus

We tested 8,800, 8,662 and 18,757 protein-coding genes for *Malassezia*, *Saccharomyces* and *Aspergillus*. The two yeasts passed multiple testing correction at many loci (Figure 1): *Malassezia* had 85 significant genes and *Saccharomyces* had 70. *Aspergillus* had none (minimum $P = 5.8 \times 10^{-5}$, threshold 2.7×10^{-6}); no genus showed gene-level signal in the HLA region (chr6:25–36 Mb; strongest $P = 1.7 \times 10^{-4}$). Note: gene-level significance used the 1000 Genomes EUR reference panel, which does not fully match the Chinese-ancestry fungal GWAS. These counts have limited cross-ancestry comparability (see the Limitations section). The small difference in the number of genes tested between the two yeasts reflects FUMA annotation versioning between analysis batches, not variant coverage (Methods, Gene-based and gene-set analysis). Top hits for *Malassezia* included PTPRD ($P = 7.1 \times 10^{-13}$), RBFOX1, MACROD2 and CSMD1. The list also contained immune genes PDCD1 ($P = 4.2 \times 10^{-6}$, Bonferroni-significant, ranked 75th) and IL-23R ($P = 6.9 \times 10^{-5}$, nominal). Top hits for *Saccharomyces* were CSMD1 ($P = 6.8 \times 10^{-18}$), RBFOX1, PTPRD and PRKCE. Twenty-one genes were significant for both yeasts. Most are

very large genes with neuronal functions, which points to shared polygenic architecture rather than shared single loci.

Table 2. Top gene-level hits for each fungus.

Fungus	Gene	CHR	P
Malassezia	PTPRD	9	7.10e-13
Malassezia	RBFOX1	16	1.84e-10
Malassezia	MACROD2	20	2.20e-10
Malassezia	TMEM132D	12	5.30e-10
Malassezia	RPH3AL	17	2.11e-09
Malassezia	PCDH15	10	4.55e-09
Malassezia	CSMD1	8	4.74e-09
Malassezia	MSR1	8	4.97e-09
Saccharomyces	CSMD1	8	6.82e-18
Saccharomyces	RBFOX1	16	8.62e-13
Saccharomyces	PTPRD	9	2.02e-11
Saccharomyces	PRKCE	2	6.26e-09
Saccharomyces	WWOX	16	6.79e-09
Saccharomyces	PARD3B	2	8.41e-09
Saccharomyces	ZNF385D	3	4.61e-08
Saccharomyces	MAGI2	7	5.12e-08
Aspergillus	TMEM45B	11	5.81e-05
Aspergillus	CTD-2203A3.1	5	8.04e-05
Aspergillus	ZFAND6	15	8.70e-05
Aspergillus	TAS2R43	12	1.71e-04
Aspergillus	AADAT	4	2.07e-04
Aspergillus	ZNF19	16	2.75e-04
Aspergillus	TBRG1	11	3.28e-04
Aspergillus	KISS1R	19	4.99e-04

Note. No Bonferroni-significant genes were identified for Aspergillus; all listed Aspergillus entries represent nominal associations only.

Two points stand out in Table 2. First, the yeast top hits are a mixture of very large neuronal genes and immune loci. Second, the *Aspergillus* column reaches only nominal significance, which is the gene-level basis for calling *Aspergillus* a null

result. *Aspergillus*, a trait with limited gene-level power, served as an internal reference for the pipeline. Gene-level analyses of the five additional screened taxa (*Penicillium*, *Candida*, and the phylum-level Ascomycota and Basidiomycota traits) and of the Shannon diversity index returned no Bonferroni-significant genes (Additional file 9: Table S8): across all eight tested traits the gene-level signal is confined to the two yeasts.

Pathway enrichment is genus-specific and mutually exclusive

Malassezia showed enrichment of 13 gene sets at Bonferroni significance (16,419 tested; Figure 2). The strongest were autophagosome membrane docking ($P = 3.5 \times 10^{-11}$), mitochondria-ER membrane tethering ($P = 1.2 \times 10^{-10}$) and SUMO processing ($P = 2.4 \times 10^{-9}$). The theme is intracellular handling: autophagic capture, SUMOylation and organelle contact. *Saccharomyces* showed enrichment of three sets: leukocyte chemoattractant receptor activity ($P = 5.0 \times 10^{-7}$), mitochondrial apoptotic membrane insertion ($P = 1.4 \times 10^{-6}$) and malate dehydrogenase activity ($P = 2.2 \times 10^{-6}$). *Aspergillus* showed enrichment of a single set, arginine metabolism ($P = 2.9 \times 10^{-6}$, passing the Bonferroni threshold of 2.94×10^{-6}; Figure 3). No set was shared across genera within the same FUMA default collection. Because Bonferroni thresholds scale with the number of sets tested per trait, set counts are not directly comparable across genera.

Table 3. *Malassezia* significant gene sets.

Gene set	N genes	BETA	P

GOBP_AUTOPHAGOSOME_MEMBRANE_DOCKING	4	+1.24	3.46e-11
GOBP_MITOCHONDRION_ENDOPLASMIC_RETICULUM_MEMBRANE_TETHERING	2	+1.24	1.18e-10
GOMF_UBIQUITIN_LIKE_PROTEIN_SPECIFIC_ENDOPEPTIDASE_ACTIVITY	3	+1.17	2.36e-09
REACTOME_PROCESSING_AND_ACTIVATION_OF_SUMO	3	+1.17	2.36e-09
REACTOME_SUMO_IS_PROTEOLYTICALLY_PROCESSED	3	+1.17	2.36e-09
GOBP_REGULATION_OF_ATPASE_COUPLED_CALCIUM_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	2	+1.11	7.93e-09
GOBP_POSITIVE_REGULATION_OF_ATPASE_COUPLED_CALCIUM_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	2	+1.11	7.93e-09
GOBP_PROTEIN_DESUMOYLATION	4	+0.95	1.37e-07
GOMF_DESUMOYLASE_ACTIVITY	4	+0.95	1.37e-07
GOBP_SEMI_LUNAR_VALVE_DEVELOPMENT	31	+0.31	7.99e-07
GOBP_AORTIC_VALVE_MORPHOGENESIS	26	+0.33	8.90e-07
GOBP_PROTEIN_SULFATION	6	+0.76	1.12e-06
GOBP_HEART_VALVE_MORPHOGENESIS	37	+0.28	1.28e-06

Note. The SUMO-processing and desumoylation sets share the same genes and therefore have identical statistics, and the three valve-morphogenesis sets share overlapping member genes; we report them for completeness and do not treat them as independent findings.

Driver genes behind the genus-specific pathways

The set-level signals have identifiable gene-level drivers. For the arginine set, 15 of 18 member genes point in the same direction. ASS1 ($P = 9.3 \times 10^{-3}$), ART4 ($P = 1.1 \times 10^{-2}$) and DDAH2 ($P = 1.6 \times 10^{-2}$) lead the set. NOS1 is nominal ($P = 3.2 \times 10^{-2}$) while NOS2 and NOS3 are not members. The signal is therefore arginine synthesis and asymmetric dimethylarginine (ADMA) handling, not inducible nitric oxide synthase itself. For *Saccharomyces*, ACKR4 ($P = 3.6 \times 10^{-6}$, gene-level significant)

and ME2 ($P = 3.7 \times 10^{-6}$, gene-level significant) are the strongest members of the chemoattractant and malate sets; BAX ($P = 1.4 \times 10^{-4}$) leads the mitochondrial set. For *Malassezia*, the gene-level drivers of the intracellular-processing theme are the top hits already listed in Table 2, and its significant gene sets are mostly small (2–6 genes each), apart from the three valve-morphogenesis sets (Table 3), so set membership and top-gene information largely coincide. The full member lists for the other genera are in Additional files 1 and 2: Tables S1 and S2.

PBC gene-based and gene-set analysis

On the disease side, PBC gene-based analysis (N = 24,510; 8,836 genes; Figure 4) returned 381 significant genes. The signal has the known PBC architecture: a dominant MHC class II cluster on chromosome 6, the chr3q13 locus (TIMMDC1, CD80), IL12RB2 ($P = 6.4 \times 10^{-13}$), IL7R ($P = 2.7 \times 10^{-13}$) and CD58 ($P = 3.1 \times 10^{-8}$). IL-23R sits above the Bonferroni threshold ($P = 1.1 \times 10^{-5}$ versus 5.7×10^{-6}) and is therefore a nominal PBC gene-level hit; we return to it in the cross-side comparison below. Gene-set analysis (16,447 sets) returned 217 significant sets; 172 of the 381 significant genes lie in the HLA region (chr6:25–36 Mb). The sets are led by allograft rejection ($P = 1.1 \times 10^{-24}$), lymphocyte-mediated immunity ($P = 2.2 \times 10^{-23}$) and type I diabetes mellitus ($P = 3.6 \times 10^{-21}$), with MHC class II peptide assembly ($P = 5.9 \times 10^{-19}$) and the IL12–IL23 axis represented (IL23 signalling $P = 2.5 \times 10^{-11}$).

Table 4. PBC top significant gene sets (full list in Additional file 4: Table S4).

Gene set	N genes	BETA	P

KEGG_ALLOGRAFT_REJECTION	26	+1.01	1.10e-24
GOBP_LYMPHOCYTE_MEDIATED_IMMUNITY	145	+0.45	2.20e-23
KEGG_TYPE_I_DIABETES_MELLITUS	30	+0.89	3.63e-21
GOBP_PEPTIDE_ANTIGEN_ASSEMBLY_WITH_MHC_CLASS_II_PROTEIN_COMPLEX	9	+2.04	5.93e-19
GOCC_MHC_CLASS_II_PROTEIN_COMPLEX	9	+2.04	5.93e-19
KEGG_ASTHMA	13	+1.26	8.99e-19
GOBP_ADAPTIVE_IMMUNE_RESPONSE_BASED_ON_SOMATIC_RECOMBINATION_OF_IMMUNE_RECEPTORS_BUILT_FROM_IMMUNOGLOBULIN_SUPERFAMILY_DOMAINS	157	+0.38	2.35e-18
GOBP_IMMUNE_EFFECTOR_PROCESS	296	+0.27	2.27e-17
WP_ALLOGRAFT_REJECTION	55	+0.58	2.07e-16
GOBP_LEUKOCYTE_MEDIATED_IMMUNITY	198	+0.31	4.90e-16
KEGG_AUTOIMMUNE_THYROID_DISEASE	27	+0.79	8.59e-16
KEGG_GRAFT_VERSUS_HOST_DISEASE	22	+0.87	1.31e-15
GOBP_B_CELL_MEDIATED_IMMUNITY	75	+0.50	1.61e-15
GOBP_REGULATION_OF_LYMPHOCYTE_MEDIATED_IMMUNITY	93	+0.45	2.79e-15
GOBP_POSITIVE_REGULATION_OF_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	15	+1.03	3.15e-15
GOBP_POSITIVE_REGULATION_OF_LYMPHOCYTE_ACTIVATION	166	+0.35	7.06e-15
GOBP_IMMUNOGLOBULIN_PRODUCTION	62	+0.54	1.15e-14
GOBP_POSITIVE_REGULATION_OF_CELL_ACTIVATION	196	+0.31	2.55e-14
GOBP_POSITIVE_REGULATION_OF_LEUKOCYTE_CELL_CELL_ADHESION	136	+0.37	3.39e-14
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	26	+0.73	3.62e-14

Two contrasts in Table 4 deserve note. First, none of the 217 PBC sets overlaps the 13 *Malassezia* sets, and PBC carries no autophagy or SUMO set. Second, the two sides share individual genes along the IL12–IL23 axis (gene-level signals only; no pathway is shared): IL-23R (nominal on both sides: *Malassezia* $P = 6.9 \times 10^{-5}$, PBC $P = 1.1 \times 10^{-5}$), IL12A (Bonferroni-significant on the PBC side: *Malassezia* $P = 1.2 \times 10^{-2}$, PBC

P = 1.9 × 10^{-12}), CD58 (fungal-side nominal in both yeasts; PBC P = 3.1 × 10^{-8}) and TYK2 (fungal-side nominal in both yeasts; PBC P = 5.3 × 10^{-9}).

PSC gene-based and gene-set analysis

PSC MAGMA (N = 14,890; 18,277 genes) returned 228 Bonferroni-significant genes and recovered the established architecture (Figure 5, Table 5): a dominant MHC signal (HIST1H4A P = 5.2 × 10^{-45}; C6orf10, PSORS1C1, HLA-DOB), with GPX1 (P = 5.7 × 10^{-17}) the strongest non-MHC gene. Gene-set analysis returned 20 Bonferroni-significant sets (17,006 tested), predominantly in the MHC class II antigen processing and presentation theme: peptide antigen assembly with MHC protein complex (P = 2.3 × 10^{-11}), MHC protein complex (5.2 × 10^{-10}), MHC class II receptor activity (9.8 × 10^{-10}), plus adaptive immune response based on somatic recombination (5.0 × 10^{-7}). PSC therefore sits on the adaptive, MHC-centred side of the axis alongside PBC, in contrast to the innate intracellular *Malassezia* signature (Table 5, Figure 5).

Table 5. PSC Bonferroni-significant gene sets (17,006 tested; P < 2.9 × 10^{-6}).

Gene set	N genes	P
GOBP_PEPTIDE_ANTIGEN_ASSEMBLY_WITH_MHC_PROTEIN_COMPLEX	17	2.32e-11
GOCC_MHC_PROTEIN_COMPLEX	22	5.18e-10
GOBP_PEPTIDE_ANTIGEN_ASSEMBLY_WITH_MHC_CLASS_II_PROTEIN_COMPLEX	13	5.92e-10
GOMF_MHC_CLASS_II_RECEPTOR_ACTIVITY	9	9.75e-10
GOCC_MHC_CLASS_II_PROTEIN_COMPLEX	14	4.60e-09
GOBP_LYMPHOCYTE_MEDIATED_IMMUNITY	270	5.71e-08
GOCC_LUMENAL_SIDE_OF_ENDOPLASMIC_RETICULUM_MEMBRANE	23	1.08e-07
GOBP_ANTIGEN_PROCESSING_AND_PRESENTATION_OF_PEPTIDE_ANTIGEN	66	1.44e-07

GOBP_ANTIGEN_PROCESSING_AND_PRESENTATION	104	3.65e-07
GOBP_POSITIVE_REGULATION_OF_T_CELL_PROLIFERATION	94	3.70e-07
GOCC_LUMENAL_SIDE_OF_MEMBRANE	30	4.38e-07
GOBP_ADAPTIVE_IMMUNE_RESPONSE_BASED_ON_SOMATIC_RECOMBINATION_OF_IMMUNE_RECEPTORS_BUILT_FROM_IMMUNOGLOBULIN_SUPERFAMILY_DOMAINS	287	4.98e-07
KIM_GLIS2_TARGETS_UP	79	7.15e-07
GOMF_SODIUM_PHOSPHATE_SYMPORTER_ACTIVITY	12	9.67e-07
GOBP_ADAPTIVE_IMMUNE_RESPONSE	422	9.68e-07
GOBP_ALPHA_BETA_T_CELL_ACTIVATION	160	9.86e-07
GOMF_PEPTIDE_ANTIGEN_BINDING	27	1.43e-06
GAURNIER_PSMD4_TARGETS	64	1.50e-06
GOMF_MHC_CLASS_II_PROTEIN_COMPLEX_BINDING	24	2.49e-06
GOBP_POSITIVE_REGULATION_OF_LEUKOCYTE_CELL_CELL_ADHESION	250	2.72e-06

Fungal and AILD loci overlap regionally but do not formally colocalize

We tested 680 fungal-AILD locus pairs (the 340 fungal leads against PBC and PSC; Figure 6, Table 6) with coloc.abf within ±500 kb windows (Methods). No window reached the conventional colocalization threshold of PP.H4 ≥ 0.8; the genome-wide maximum was 0.16, and only 16 of 680 windows reached PP.H4 ≥ 0.1. We therefore report the results as regional-overlap candidates ranked by PP.H4 (Table 6; complete ranked list in Additional file 5: Table S5), not as confirmed shared causal variants. The strongest candidate is *Aspergillus*-PBC at chr14:30.9 Mb (PP.H4 = 0.16, 763 variants). *Aspergillus* carries no gene-level signal, but its low-abundance, noisy measurement means regional overlap can still arise; this candidate should be read with caution. The next candidates are *Saccharomyces*-PSC at chr8:3.1 Mb (PP.H4 = 0.144), *Candida*-PSC at chr3:189.2 Mb (PP.H4 = 0.121) and *Malassezia*-PSC at chr4:130.2 Mb (PP.H4 = 0.068). The rank-4 window (chr1:191.4 Mb, a Shannon

diversity index) and the rank-5 window (chr18:8.6 Mb, *Aspergillus*) are listed in Table 6 for completeness and are not followed up further. The *Malassezia* locus at chr11:67.2 Mb overlaps both PBC (PP.H4 = 0.057; nearest variant P = 0.021) and PSC (PP.H4 = 0.043), and carries independent support from a strong local eQTL for AIP in the eQTLGen whole-blood scan (lead Z = −12.5; cluster SNPs reach Z = −15.0). We retain the chr11 locus as the second *Malassezia*-region candidate because the AIP cis-eQTL signal in this window is strong (Figure 7).

Table 6. Regional-overlap candidate loci between fungal abundance and AILD, ranked by coloc PP.H4.

Rank	Locus	Fungal trait	AILD	Lead SNP (fungal P)	PP.H4	Support
1	chr14:30,918,573	*Aspergillus*	PBC	14:30918573:C:T (2.5×10^{-6})	0.160	Highest PP.H4; tier 3, interpret with caution
2	chr8:3,084,707	*Saccharomyces*	PSC	8:3084707:A:G (8.5×10^{-6})	0.144	Candidate
3	chr3:189,245,593	*Candida*	PSC	3:189245593:A:G (6.2×10^{-6})	0.121	Candidate
4	chr1:191,430,223	Shannon diversity index	PSC	1:191430223:A:G (2.9×10^{-6})	0.110	Listed for completeness
5	chr18:8,637,276	*Aspergillus*	PSC	18:8637276:T:TA (9.5×10^{-6})	0.088	Listed for completeness
6	chr4:130,247,960	*Malassezia*	PSC	4:130247960:A:G (4.2×10^{-6})	0.068	Tier 1 candidate
7	chr11:67,192,555	*Malassezia*	PBC	11:67192555:G:A (1.4×10^{-6})	0.057	Tier 2; + AIP eQTL (Z = −15.0); also overlaps PSC (PP.H4 = 0.043)

Note. None of these candidate loci reached the standard colocalization threshold PP.H4 ≥ 0.8; ranking is for priority purposes only and does not indicate proven shared causal variants. Lead SNPs use position-based identifiers (chr:position:ref:alt, GRCh37), matching the fungal GWAS source files

(Methods, Study data). The two triangulation leads have dbSNP identifiers rs59634627 (chr4:130,247,960) and rs1626067 (chr11:67,192,555).

Evidence tiers: chr4 is a tier-1 locus (single-SNP triangulation), chr11 is a tier-2 locus (regional overlap plus strong cis-eQTL), and chr14 is a tier-3 locus (regional overlap alone).

The ranking of Table 6 reflects limited power: with N = 7,350, windows rarely reach PP.H4 ≥ 0.8 even for genuinely overlapping loci, and the candidates are best read as priorities. Ranks reflect PP.H4 only; the evidence tiers in the main text are independent of this ranking (for example, the tier-1 chr4 locus ranks sixth by PP.H4). Across the three AILDs, only PBC and PSC yielded regional-overlap candidates. Of the 680 tested windows, 340 were against PBC and 340 against PSC. AIH yielded no testable window (none met the 50-variant overlap requirement; Additional file 5: Table S5), so no AIH inference is made. The three evidence components of this triangulation framework were obtained from mutually non-overlapping data sources with no shared study participants. One candidate shows concordant single-SNP-level evidence: at chr4:130.2 Mb the *Malassezia* lead rs59634627 satisfies all three criteria at one variant (fungal $P = 4.2 \times 10^{-6}$; cis-eQTL $|Z| = 5.8–9.0$ for C4orf33, SCLT1 and PHF17; PSC P = 0.039). The chr11 locus sits one tier down: the lead rs1626067 passes the fungal and eQTL legs (AIP Z = −12.5; cluster SNPs −15.0) while its AILD leg is carried by a neighbouring cluster SNP distinct from the nearest PBC variant. That SNP, rs1077256, has PSC P = 0.024; the nearest PBC variant rs3741180 has P = 0.021. The chr14 candidate rests on regional overlap alone. To calibrate the window-

level nominal criterion, we generated an empirical null by sampling 10,000 sets of 340 random genomic positions and scoring each with the same window rules on an ≈27% random subsample of the PSC GWAS (2,072,697 of 7,669,981 variants; seed 42). Under the null, a mean of 19.6 of the 340 leads (range 6–37) had a nearest variant with $P < 0.05$ in the AILD file, matching the 5% nominal expectation; independently, essentially all windows passed the window-minimum criterion (observed 340/340; null median 340/340), because a ±500-kb window containing hundreds of variants almost always includes one with $P < 0.05$ (median per-window minimum P across null windows 2.4×10^{-3}; 99.7% of null windows passed). The observed count of 20 matched the null (empirical $P = 0.498$; observed window statistics in Additional file 11: Table S10). Window-level nominal overlap therefore carries no discriminative weight on its own; the tier-1 status of chr4:130.2 Mb rests on the coincidence of all three legs at a single variant (PSC $P = 0.0391$ at the exact lead in the subsample, identical to the full-data value).

eQTL bridging of candidate SNPs

We queried all fungal candidate SNPs against the eQTLGen whole-blood eQTL scan (Figure 7) and retained 326 SNP-gene pairs (139 SNPs, 94 genes) after FDR filtering. Both *Malassezia* candidates (chr4 with PSC, chr11 with PBC) carry strong cis-eQTL signal. At the chr11 locus the coloc lead rs1626067 regulates seven genes in the cluster, including AIP ($Z = -12.5$; cluster SNPs reach $Z = -15.0$), SSH3 (−17.9), CORO1B (−18.4) and GSTP1 (−5.3). At the chr4 locus the lead rs59634627 regulates

C4orf33 (Z = −9.0), SCLT1 (−7.7) and PHF17 (−5.8). The chr14 *Aspergillus*-PBC lead SNP has no entry in the eQTL scan, so that candidate rests on regional overlap and coloc ranking alone. Independent fine-mapped molecular-QTL evidence supports both loci (Open Targets Platform, accessed October 2026; studies include GTEx v10 and Alasoo 2018 [29]). For chr4, the 95% credible set containing rs59634627 overlaps SCLT1 expression in interferon-γ-stimulated macrophages ($P = 3.3 \times 10^{-13}$) and in classical monocytes ($P = 1.1 \times 10^{-12}$), and C4orf33 expression in unstimulated classical monocytes ($P = 8.1 \times 10^{-6}$), with effect directions concordant with the eQTLGen whole-blood signals. For chr11, the credible sets implicate an AIP eQTL in right atrial appendage ($P = 1.5 \times 10^{-8}$; fine-mapping posterior probability 0.006), a GSTP1 transcript-usage QTL in classical monocytes ($P = 1.1 \times 10^{-7}$) and an RPS6KB2 splicing QTL in interferon-γ-stimulated macrophages ($P = 5.5 \times 10^{-9}$). No liver-tissue credible set overlaps either candidate locus.

Cell-type specificity across the eight reference datasets

We performed MAGMA cell-type analysis using eight blood and bone-marrow scRNA-seq reference datasets (748 cell type-trait tests: eight reference datasets × cell types per dataset × five traits with gene-level results; Figure 8, Table 7). Table 7 is provided at the end of the manuscript document. *Malassezia* abundance showed 11 Bonferroni-significant cell types, all of them myeloid: non-classical and classical monocytes (non-classical: Xu 2023 blood $P = 2.0 \times 10^{-6}$; Travaglini blood $P = 5.73 \times 10^{-6}$), neutrophils, promyelocytes and Pf4-high macrophages. The same myeloid cell

types were enriched across human blood, human bone-marrow and mouse reference datasets, but results were not combined across datasets (Methods, Cell-type specificity analysis). PBC showed 15 significant cell types, predominantly lymphoid (10 of 15: B cells, MAIT, effector-memory CD8, regulatory T and NK) with four macrophage or monocyte hits and one progenitor entry. This is consistent with the MHC class II signal residing in antigen-presenting compartments that span both lineages. Neither *Aspergillus* nor *Saccharomyces* reached significance for any cell type in any dataset, showing that the myeloid enrichment is specific to *Malassezia* rather than a generic property of fungal GWAS. PSC contributed two significant cell types, both T-lineage: T follicular helper cell ($P = 5.6 \times 10^{-4}$) and mucosal-associated invariant T cell (6.0×10^{-4}; Xu 2023 bone marrow). This places PSC on the lymphoid side of the axis together with PBC.

Drug-gene annotation of the antifungal-immunity pathway

We queried DGIdb for druggable nodes among the fungal-recognition machinery connected to the *Malassezia* signature (CARD9, SYK, MALT1, CLEC7A) (Table 8). The most actionable node is SYK, the kinase immediately upstream of CARD9: fostamatinib, an approved SYK inhibitor, is annotated against this axis. MALT1 inhibitors and CLEC7A (Dectin-1) agonists remain in clinical development, and CARD9 itself has no specific drug and would require new chemical matter. No agent in this set has been tested in PSC or PBC; the annotation identifies where pharmacology exists, not where efficacy has been shown.

Table 8. Drug-gene annotation of the antifungal-immunity pathway (DGIdb 4.0).

Gene	Pathway role	Annotated drug (status)	Repositioning feasibility	DGIdb annotation
CARD9	Fungal-sensing adaptor	No specific drug	Low (requires new chemistry)	None (2 curated interactions)
SYK	Kinase upstream of CARD9	Fostamatinib (approved)	High	Approved (4 curated interactions)
MALT1	Downstream of CARD9	MALT1 inhibitors (in development)	Medium	Clinical development (3 curated interactions)
CLEC7A	β-glucan receptor (Dectin-1)	Dectin-1 agonists (in development)	Medium	Clinical development (3 curated interactions)

Note. Feasibility reflects drug availability: fostamatinib is approved; MALT1 and CLEC7A agents are in development; CARD9 has no specific drug.

Discussion

Principal findings

This study maps host genetic control of three gut fungal genera and tests where that control meets autoimmune liver disease. We report three main findings.

First, host control of the mycobiome is genus-specific in both strength and mechanism: the two yeasts show strong gene-level regulation (85 and 70 significant genes) while *Aspergillus* falls below the detection floor of this pipeline. Because all three traits come from the same cohort and pipeline, this contrast is more likely biological than technical. The yeast top hits are dominated by very large neuronal

genes (PTPRD, RBFOX1, CSMD1, MACROD2); whether these reflect shared polygenic architecture, mismatched LD structure or a genuine mucosal-neuronal regulatory axis remains unresolved and is a question for functional follow-up.

Second, the *Malassezia* signature points to intracellular processing: the top gene sets were autophagosome docking, SUMO processing and mitochondria-ER membrane tethering. Autophagy is the standard route by which myeloid cells handle fungi [30]; SUMOylation and organelle contact are named less often in this context. The gene list adds immune weight through PDCD1 and IL-23R; immune and autophagy genes with nominal signal in both yeasts are listed in Additional file 3: Table S3. Clinical observations in Crohn's disease and PSC point the same way [8,9,10]; this study provides a host-genetic entry point for them.

Third, the fungal and disease genetics converge on a short list of graded overlapping loci. One locus shows concordant single-SNP-level evidence (chr4:130.2 Mb, regulating C4orf33, SCLT1 and PHF17). One is supported by regional overlap plus strong cis-eQTL (chr11:67.2 Mb, AIP cluster $Z = -15.0$), and one is a regional-overlap candidate only (chr14:30.9 Mb). Direct colocalization is weak everywhere (PP.H4 0.04–0.16), as expected for a cross-ancestry design; only 16 of 680 tested windows reached PP.H4 ≥ 0.1. PBC genetics is adaptive and MHC-centred (172 significant genes in the HLA region); PSC contributed MHC-II-dominated gene sets on the same side; the *Malassezia* signature is innate and intracellular. By contrast, none of the three fungal genera shows any gene-level signal in the HLA region

(chr6:25–36 Mb; strongest $P = 1.7 \times 10^{-4}$). This suggests that host control of fungal abundance and the HLA-dominated architecture of AILD meet at the locus level rather than inside the antigen-presentation machinery. A high colocalization posterior probability can merely indicate a shared regional variant; it cannot demonstrate pathogenic effects. Our three-tier evidence scheme is calibrated with this limitation in mind. Random permutation testing shows that nominal overlap at the window level has little discriminatory power by itself (empirical $P = 0.498$). Interpretation therefore relies on these graded tiers, anchored upon single-SNP concordance. All gene-level analyses above used the 1000 Genomes EUR reference panel for LD, applied to a Chinese-ancestry GWAS. We therefore re-ran MAGMA for the three fungal traits with the ancestry-matched EAS panel. Coverage was fully preserved: among *Malassezia* significant genes the median SNP count per gene rose from 105 to 113 (*Saccharomyces* 90 to 102); for *Aspergillus*, which has no significant genes under either panel, the relevant observation is the exact replication of the null. No gene lost its entry, and no gene dropped from more than ten SNPs to fewer than three. The gene-level structure was preserved in each trait (Spearman ρ: *Malassezia* 0.85, *Saccharomyces* 0.84, *Aspergillus* 0.97; 16, 11 and 14 of the top 20 genes retained, respectively). At the pathway level the leading autophagosome docking set remained significant under EAS ($P = 6.8 \times 10^{-7}$), whereas the broader SUMO sets did not replicate (best $P = 0.036$). Because these EAS runs used the server's current gene annotation (10,225 genes tested versus 8,800 for the September runs), the pathway

contrast conflates panel with annotation version. We therefore re-ran all three fungal traits under the EUR panel with the current server annotation. The September results reproduced exactly, including the gene counts (85, 70 and 0 significant genes), the top genes, and the pathway results, with the SUMO sets again significant ($P = 2.4 \times 10^{-9}$). The SUMO non-replication under EAS is therefore attributable to LD-panel ancestry mismatch rather than annotation version. The autophagy theme replicated under both panels, whereas the SUMO gene sets replicated only under the EUR panel (Additional file 7: Table S7).

Candidate loci and the IL-23R intersection

The chr4:130.2 Mb locus is the only tier-1 candidate and requires independent validation. The lead SNP rs59634627 passes all three triangulation criteria at a single variant (Results, Locus-level overlap); even though the PSC leg is nominal, three independent measurements coincide at one base. We found no prior report linking C4orf33, SCLT1 or PHF17 to PSC, so this is a new candidate locus. The three lines of evidence are directionally consistent (the allele increases *Malassezia* abundance and PSC risk while lowering expression of the three genes), compatible with a single causal variant. The chr4 credible set is furthermore a fine-mapped eQTL for SCLT1 in interferon-γ-stimulated macrophages (the same myeloid compartment in which the *Malassezia* cell-type signal resides), linking the candidate locus to the cell-level partition. At chr11 the credible sets again reach into immune compartments (GSTP1 in classical monocytes and RPS6KB2 in stimulated macrophages), while the AIP

eQTL itself is strongest in cardiac tissue; the AIP-to-aryl-hydrocarbon-receptor hypothesis therefore remains a whole-blood-led inference.

The chr11:67.2 Mb locus points to AIP, a regulator of the aryl hydrocarbon receptor pathway, which handles environmental compounds and modulates mucosal immunity [31,32]. Microbial metabolites, including fungal products, are a known source of aryl-hydrocarbon-receptor ligands [33], so the path from SNP through AIP and the aryl hydrocarbon receptor to the fungus-immune interface is a biologically grounded hypothesis. The evidence remains statistical. The cis-eQTL signal is strong (cluster Z = −15.0) but the colocalization posterior does not support a shared causal variant, and we place this locus in the second evidence tier.

IL-23R is the only named gene that appears on both the fungal side (the signal is *Malassezia*-specific, because IL-23R's window carried no quality-controlled SNPs in the *Saccharomyces* GWAS) and the disease side (PBC). On the PBC side it is a nominal gene-level hit ($P = 1.1 \times 10^{-5}$, above the Bonferroni threshold; listed in Additional file 6: Table S6); on the fungal side it is likewise nominal (Results, Gene-level association; Figure 1). IL-23/IL-23R signalling is known to maintain inflammation in cholestatic liver disease [34]. IL-23R is expressed on MAIT cells [35] and T helper 17 cells; MAIT cells are one of the two T-lineage compartments enriched in the PSC signal, which makes IL-23R the one node where the gene-level and cell-level intersections coincide. The IL-23/IL-17 axis is pharmacologically mature, although no drug in it is established for AILD. Our data add no causal

evidence for this pathway; one gene recurring in two independent trait lists gives follow-up experiments a concrete intersection: a gene-list-level hint, not pathway-level causal evidence. At the cellular level the same division of labour is visible. The *Malassezia* signal sits exclusively in myeloid compartments, the PBC signal predominantly in lymphoid compartments, and the two comparator fungal genera show no cellular enrichment at all (Results, Cell-type specificity). PSC contributed two T-lineage cell types, including T follicular helper cells — a compartment previously implicated in PBC [36] (Results, Cell-type specificity). The candidate loci, pathways and cell types thus point to one pattern: innate recognition of *Malassezia* on one side, adaptive immune damage on the other. We did not perform a formal heterogeneity test across the three diseases. PBC and PSC are both adaptive and MHC-centred, AIH yielded no passing candidate, and the three diseases are placed side by side to cover the AILD spectrum rather than to be compared. This split also cautions against reading the axis as a simple drug target: dampening fungal recognition (for example, SYK inhibition) could impair clearance, whereas the disease side calls for immune regulation rather than broad suppression. Which side to push, and in which patients, is for functional experiments to settle. External comparison supports the novelty of these candidates. An independent fungal GWAS based on 125 HMP participants, with validation in independent cohorts, reported 148 fungi-associated variants across nine taxa and seven chromosomes (chr1, chr4, chr5, chr10, chr11, chr16 and chr18), including *Aspergillus* and *Candida* [14]. Direct lookup

of our lead SNPs and all five candidate windows in their S1 Table returned zero overlapping variants; our chr4:130.2 Mb and chr11:67.2 Mb candidates therefore add to, rather than replicate, the published map at the variant level.

Boundaries: MR comparisons and causal direction

Two recent studies tested microbiota-AILD causality with Mendelian randomization [20,21]. Both used bacterial exposures and one-directional designs, whereas we test host loci against both fungi and disease. The conclusions complement rather than compete. The boundary on causal direction needs stating separately. The exposure side of this study is host genotype, fixed at conception and not subject to reverse shaping by fungal abundance or disease state, so the host-genetic leg is immune to reverse causation. The study does not, however, test the direction between the two traits: the *Malassezia*-PSC association could reflect fungi affecting disease, disease (or its treatment) altering fungi, or shared causes, and the data cannot distinguish these. What this framework tests is shared genetic components: the chr4 and chr11 candidates suggest that at least part of the clinically observed fungus–disease association may be driven by shared genetic causes. The shared-cause explanation now has candidate-level support. Microbiome-as-exposure Mendelian randomization is constrained by fungal GWAS sample size and instrument strength; bidirectional designs, as an independent statistical check, still require fungal GWAS samples in the tens of thousands with instrument strength $F > 10$, which the current $N = 7,350$ does not reach. This study takes a different route and leaves the direction question to

functional experiments: knockdown or overexpression of the candidate genes can test the causal order of gene change → fungal burden → immune phenotype.

The two comparison genera

Beyond the *Malassezia*-centred findings, the other two analysed genera sharpen the genus-specificity claim. *Saccharomyces* carries a complementary signal: its pathway hits name leukocyte chemoattractant receptors (ACKR4, CCR6, CMKLR1), where CCR6 marks gut T helper 17 cells and dendritic cells and ACKR4 controls immune-cell trafficking [37], and its malate set matches the biochemistry of the genus itself, which ferments malate. *Saccharomyces* is therefore under host control through leukocyte recruitment rather than intracellular handling.

Aspergillus is the useful negative control. Its single significant set, arginine metabolism, is carried by ASS1, ART4 and DDAH2 and is about arginine supply and ADMA clearance, which modulate nitric oxide flux in macrophages [38]. That a genus outside host gene control still yields one coherent set argues that the pipeline detects biology at the pathway level even where gene-level power is absent. As a negative control it rules out pipeline-wide false positives; genuinely weak but real control of *Aspergillus* is covered by the detection-floor statement. A 125-participant fungal GWAS reported nominal *Aspergillus*-associated variants [14], consistent with this detection-floor interpretation. It is also the genus behind the top-ranked regional-overlap candidate with PBC (chr14:30.9 Mb, PP.H4 = 0.16), which argues against dismissing low-abundance fungi from disease follow-up. Notably, its single arginine-

metabolism set has no member gene passing Bonferroni correction — pathway-level signal without a gene-level driver — supporting the genus-level null call.

The pattern extends beyond the three reported genera: *Penicillium*, *Candida*, the phylum-level Ascomycota and Basidiomycota traits, and the Shannon diversity index all returned zero significant genes, against 85 and 70 for the yeasts (Additional file 9: Table S8). The phylum-level Ascomycota trait is null while its constituent genus *Saccharomyces* carries 70 significant genes, indicating that phylum aggregation dilutes genuine genus-level signal. Host genetic control of gut fungal abundance is thus not a property of fungi in general but a genus-specific feature concentrated in the two yeasts.

Translational implications

These results carry three levels of translational implication, of decreasing evidence strength.

Level 1 is a prioritised candidate list. The chr4:130.2 Mb locus (C4orf33/SCLT1/PHF17) and the chr11:67.2 Mb locus (AIP) carry replicable statistical evidence and translate directly into functional experiments. For example, knockdown or overexpression of the candidate genes in macrophage-fungus co-culture or gut-organoid systems, with fungal burden, cytokine output and barrier phenotypes as readouts, prioritised by evidence tier.

Level 2 is drug-repositioning clues. The antifungal-recognition pathway contains one immediately actionable node, SYK-fostamatinib (Table 8). The PSC multitrait GWAS

independently nominated candidate drug directions [17]. The IL-12/IL-23 pathway has approved drugs. Genetic association, however, is not drug response: simtuzumab, an antifibrotic with a clear mechanistic rationale, showed no clinical benefit in a randomised phase 2 trial in PSC [39], illustrating the gap between statistical association and clinical efficacy. We therefore treat the drug annotations as a source of hypotheses, not as efficacy evidence.

Level 3 is a biomarker hypothesis — whether carriers of specific alleles form a patient subgroup more responsive to fungal intervention. No data in this study support this layer. We raise it only as a question for future work.

Limitations

Several limitations apply. First, the cross-ancestry design. The fungal GWAS is Chinese while the PBC and PSC GWAS and the MAGMA reference panel are European (the AIH source is multi-ancestry); this is an exploratory cross-ancestry analysis whose gene-level results await replication in ancestry-matched data [40]. A reference-panel sensitivity analysis using the EAS panel supported the gene-level structure and the leading autophagy gene set, but not the broader SUMO sets. A version-matched EUR-panel rerun reproduced the September results exactly, including the SUMO sets (Discussion). Second, colocalization power is limited: the maximum PP.H4 is 0.16, so negative results do not exclude shared causal loci and only indicate that current data are insufficient to support colocalization. A formal

power calculation for the colocalization screen was not performed. The statement that negative results reflect limited power is therefore qualitative, resting on the small fungal sample size (N = 7,350) rather than a calculated power curve. Third, triangulation and regional overlap rely on nominal P-value thresholds. A random-position permutation (10,000 iterations against the 340 PSC windows on an ≈27% random subsample of the PSC GWAS; seed 42) estimated that a mean of 19.6 of the 340 leads (range 6–37) would have a nearest variant with $P < 0.05$ by chance, matching the nominal 5% expectation; the observed count (20) was indistinguishable from the null (empirical $P = 0.498$), and the window-minimum criterion was passed by essentially all windows under the null and is non-informative. False-positive control therefore rests on evidence grading and on the single-SNP coincidence of the three triangulation legs, not on any single window-level threshold; the candidate list is graded by tier. Fourth, eQTL evidence comes primarily from eQTLGen whole blood (European ancestry), distant from the target tissues. Fine-mapped myeloid-cell QTLs (Alasoo 2018 [29]; GTEx v10, via the Open Targets Platform) partially offset this for both candidate loci, but no liver- or gut-mucosa-specific QTL data are available for them, and tissue-specific regulation may be missed. Fifth, whether the source study adjusted for bacterial composition, diet and medication cannot be verified from summary statistics alone; until verified, these factors remain potential unmeasured confounders and fungal signals may partly reflect them. Sixth, causal direction was not tested; the AIH case proportion is only 0.17% (Methods, Study data), and AIH

contributed no tested window (Additional file 5: Table S5), so no AIH conclusion is drawn; whether this zero overlap reflects variant coverage of the AIH file or a genome-build mismatch was not investigated. Seventh, all evidence is statistical association without functional experiments; the chr4 and chr11 loci require replication in independent fungal GWAS and same-ancestry AILD data. Eighth, technical sensitivity analyses leave residual uncertainty: no formal power assessment was performed for *Aspergillus* (its negative result may partly reflect the detection floor); the gene-level P-value distribution is left-shifted genome-wide (every one of the 8,800 tested *Malassezia* genes is nominal), consistent with the panel-ancestry mismatch, for which Bonferroni calling was retained; gene-based tests were one-sided, although a two-sided approximation from the gene Z-statistics retained 65 of the 85 *Malassezia* and 47 of the 70 *Saccharomyces* significant genes (and none for *Aspergillus*), leaving the genus-specific pattern unchanged; and a full permutation of the joint three-leg triangulation rule was not performed.

Conclusions

Overall, the evidence supports three calibrated statements. Host genetic control of gut fungal abundance is genus-specific, with the two yeasts under the strongest influence. The genetic basis of autoimmune liver disease remains centred on adaptive immunity and the MHC axis. The two may share a small number of loci and immune-network intersection points — most suggestively, *Malassezia*-related signals overlap PSC at

chr4 and both PBC and PSC at chr11 — but the current evidence is insufficient to confirm shared causal variants, and wholly insufficient to infer any causal direction between fungal change and disease. This study provides testable candidate loci and mechanistic directions. Moving from statistical overlap to a mechanistic chain will require: (1) larger, ancestry-matched fungal GWAS; (2) multi-tissue expression-QTL resources covering liver and gut mucosa; (3) functional experiments at the candidate genes; and (4) bidirectional analyses once instrument strength permits (Figure 9).

Abbreviations

AILD, autoimmune liver disease; AIH, autoimmune hepatitis; AIP, aryl-hydrocarbon-receptor-interacting protein; DGIdb, Drug-Gene Interaction Database; EAS, East Asian; eQTL, expression quantitative trait locus; EUR, European; FDR, false discovery rate; FUMA, Functional Mapping and Annotation; GO, Gene Ontology; GRCh37, Genome Reference Consortium human build 37; GWAS, genome-wide association study; HLA, human leukocyte antigen; LD, linkage disequilibrium; MAGMA, Multi-marker Analysis of GenoMic Annotation; MAIT, mucosal-associated invariant T; MHC, major histocompatibility complex; MSigDB, Molecular Signatures Database; PBC, primary biliary cholangitis; PP.H4, colocalization posterior probability (shared causal variant); PSC, primary sclerosing cholangitis; scRNA-seq, single-cell RNA sequencing; SNP, single nucleotide polymorphism; SUMO, small ubiquitin-like modifier; SYK, spleen tyrosine kinase.

Declarations

Ethics approval and consent to participate

This study uses only published GWAS summary statistics obtained from public databases (Table 1) and involves no new human participants; ethics approval and informed consent for each source study were obtained by the original studies.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the EBI GWAS Catalog, https://www.ebi.ac.uk/gwas/ (PBC, accession GCST90061440; AIH, accession GCST90018785) and IEU OpenGWAS, https://gwas.mrcieu.ac.uk/ (PSC, accession ieu-a-1112); the colocalization results generated in this study are included within the article and its additional files. The eQTLGen whole-blood cis-eQTL summary statistics are publicly available at https://www.eqtlgen.org [25]; fine-mapped molecular-QTL evidence was obtained from GTEx v10 (dbGaP accession phs000424.v10) and Alasoo 2018 [28] via the Open Targets Platform (https://www.opentargets.org/, accessed October 2026). Fungal abundance GWAS summary statistics were obtained from the source mycobiome study [13]; restrictions apply to their redistribution, and they are available from the source study or its authors upon reasonable request. Gene-based, gene-set and cell-type results underlying the

figures are provided in Additional files 1–7 and 9 (Tables S1–S10). Permutation, pruning and external-lookup outputs are provided in Additional file 10 (Table S9), Additional file 11 (Table S10) and Additional file 12 (source_data.zip). Analysis code is available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors received no specific funding for this work.

Authors' contributions

JS conceptualized and designed the study, performed the formal analysis and wrote the original draft. ML contributed to study design and manuscript revision. XX, YJ and SG contributed to data acquisition, data curation, analysis and manuscript revision. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Professor Ju-Sheng Zheng, Professor An Pan and their teams for generating and sharing the gut mycobiome GWAS summary statistics (N = 7,350) that made this study possible. The authors acknowledge the use of AI-assisted language-editing tools during manuscript preparation; all content was reviewed and approved by the authors.

References

[1] Li XV, Leonardi I, Putzel GG, Semon A, Fiers WD, Kusakabe T, et al. Immune regulation by fungal strain diversity in inflammatory bowel disease. Nature. 2022;603:672-678.

[2] Carrillo-Serradell L, Liu-Tindall J, Planells-Romeo V, Aragon-Serrano L, Isamat M, Gabaldon T, et al. The human mycobiome: composition, immune interactions, and impact on disease. Int J Mol Sci. 2025;26(15):7281.

[3] Zhang Z, Zhang Y, Yuan Q, Wang Z, Hu S, Yin P, et al. A meta-analysis of the human gut mycobiome using internal transcribed spacer data. Microorganisms. 2024;12(12):2567.

[4] Schöler D, Schnabl B. The role of the microbiome in liver disease. Curr Opin Gastroenterol. 2024;40(3):134-142.

[5] Leonardi I, Gao IH, Lin WY, Allen M, Li XV, Fiers WD, et al. Mucosal fungi promote gut barrier function and social behavior via type 17 immunity. Cell. 2022;185:831-846.

[6] Martini GR, Tikhonova E, Rosati E, DeCelie MB, Sievers LK, Tran F, et al. Selection of cross-reactive T cells by commensal and food-derived yeasts drives cytotoxic TH1 cell responses in Crohn's disease. Nat Med. 2023;29(10):2602-2614.

[7] Sokol H, Leducq V, Aschard H, Pham HP, Jegou S, Landman C, et al. Fungal microbiota dysbiosis in IBD. Gut. 2017;66(6):1039-1048.

[8] Limon JJ, Tang J, Li D, Wolf AJ, Michelsen KS, Funari V, et al. *Malassezia* is associated with Crohn's disease and exacerbates colitis in mouse models. Cell Host Microbe. 2019;25(3):377-388.e6.

[9] Lemoinne S, Kemgang A, Ben Belkacem K, Straube M, Jegou S, Corpechot C, et al. Fungi participate in the dysbiosis of gut microbiota in patients with primary sclerosing cholangitis. Gut. 2020;69(1):92-102.

[10] Rühlemann MC, Solovjeva MEL, Zenouzi R, Liwinski T, Kummen M, Lieb W, et al. Gut mycobiome of primary sclerosing cholangitis patients is characterised by an increase of Trichocladium griseum and Candida species. Gut. 2020;69(10):1890-1892.

[11] Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet. 2021;53(2):156-165.

[12] Bubier JA, Chesler EJ, Weinstock GM. Host genetic control of gut microbiome composition. Mamm Genome. 2021;32(4):263-281.

[13] Shuai M, Xu F, Pan XF, Zhang Z, Zhou D, Zhang K, et al. Host genetic regulation of gut mycobiome offers insights into the gut–brain interaction. Vita. 2026;1(2):202-221. https://doi.org/10.15302/vita.2026.04.0024

[14] Van Syoc EP, Davenport ER, Bordenstein SR. Gut fungi are associated with human genetic variation and disease risk. PLoS Biol. 2025;23(9):e3003339.

[15] Cordell HJ, Fryett JJ, Ueno K, Darlay R, Aiba Y, Hitomi Y, et al. An international genome-wide meta-analysis of primary biliary cholangitis: novel risk loci and candidate drugs. J Hepatol. 2021;75(3):572-581.

[16] Hitomi Y, Nakamura M. The genetics of primary biliary cholangitis: a GWAS and post-GWAS update. Genes (Basel). 2023;14(2):405.

[17] Han Y, Byun J, Zhu C, Sun R, Roh JY, Cordell HJ, et al. Multitrait genome-wide analyses identify new susceptibility loci and candidate drugs to primary sclerosing cholangitis. Nat Commun. 2023;14(1):1069.

[18] Ji SG, Juran BD, Mucha S, Folseraas T, Jostins L, Melum E, et al. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. Nat Genet. 2017;49(2):269-273.

[19] Hov JR, Karlsen TH. The microbiota and the gut-liver axis in primary sclerosing cholangitis. Nat Rev Gastroenterol Hepatol. 2023;20(3):135-154.

[20] Zhang J, Hu Y, Xu J, Shao H, Zhu Q, Si H. Genetically predicted immune cells mediate the association between gut microbiota and autoimmune liver diseases. Front Microbiol. 2024;15:1442506.

[21] Fu Y, Li J, Zhu Y, Chen C, Liu J, Gu S, et al. Causal effects of gut microbiome on autoimmune liver disease: a two-sample Mendelian randomization study. BMC Med Genomics. 2023;16:232.

[22] de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11(4):e1004219.

[23] Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.

[24] Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. PLoS Genet. 2021;17(9):e1009440.

[25] Võsa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet. 2021;53(9):1300-1310.

[26] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383.

[27] Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, et al. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Res. 2021;49(D1):D1144-D1151.

[28] Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol. 2016;45(6):1866-1886.

[29] Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, et al. Shared genetic effects on chromatin and gene expression indicate a role for fine-tuning regulation in macrophage immune response. Nat Genet. 2018;50(3):424-431.

[30] Kanayama M, Shinohara ML. Roles of autophagy and autophagy-related proteins in antifungal immunity. Front Immunol. 2016;7:47.

[31] Bahman F, Choudhry K, Al-Rashed F, Al-Mulla F, Sindhu S, Ahmad R. Aryl hydrocarbon receptor: current perspectives on key signalling partners and immunoregulatory role in inflammatory diseases. Front Immunol. 2024;15:1421346.

[32] Han H, Safe S, Jayaraman A, Chapkin RS. Diet-host-microbiota interactions shape aryl hydrocarbon receptor ligand production to modulate intestinal homeostasis. Annu Rev Nutr. 2021;41:455-478.

[33] Zelante T, Iannitti RG, Cunha C, De Luca A, Giovannini G, Pieraccini G, et al. Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. Immunity. 2013;39(2):372-385.

[34] Cornillet M, Geanon D, Bergquist A, Björkström NK. Immunobiology of primary sclerosing cholangitis. Hepatology. 2025;82(4):911-926.

[35] Azad AI, Gutierrez F, Gores GJ. MAIT cells in liver disease. Cells. 2026;15(1):69.

[36] Ma H, Liu Y, Zhu Y, Xu G, Xiang X, Xue S, et al. Profiling of circulating T follicular helper and T peripheral helper cell subsets in patients with primary biliary cholangitis. Immunol Lett. 2026;277:107100.

[37] Naser IH, Hamza AA, Alhili A, Faisal AN, Ali MS, Kadhim NA, et al. Atypical chemokine receptor 4 (ACKR4/CCX-CKR): a comprehensive exploration across

physiological and pathological landscapes in contemporary research. Cell Biochem Funct. 2024;42(3):e4009.

[38] Karadima E, Chavakis T, Alexaki VI. Arginine metabolism in myeloid cells in health and disease. Semin Immunopathol. 2025;47(1):11.

[39] Muir AJ, Levy C, Janssen HLA, Montano-Loza AJ, Shiffman ML, Caldwell S, et al. Simtuzumab for primary sclerosing cholangitis: phase 2 study results with insights on the natural history of the disease. Hepatology. 2019;69(2):684-698.

[40] Miao J, Guo H, Song G, Zhao Z, Hou L, Lu Q. Quantifying portable genetic effects and improving cross-ancestry genetic prediction with GWAS summary statistics. Nat Commun. 2023;14:832.

Figure legends

Figure 1. Gene-based association from MAGMA for the two yeast traits (N = 7,350). (A) *Malassezia* abundance (8,800 genes tested; 85 Bonferroni-significant). (B) *Saccharomyces* abundance (8,662 genes; 70 significant). Each point is one gene; the dashed line marks the Bonferroni threshold. PDCD1 and IL-23R (red) are annotated for their cross-side relevance: PDCD1 passes Bonferroni correction (ranked 75th of the 85 significant genes) while IL-23R is nominal; neither gene has a *Saccharomyces* gene-level entry, so both appear only in (A).

Figure 2. Cross-genus comparison of host-process signatures; dashed lines mark the Bonferroni threshold for each genus. (A) The 13 Bonferroni-significant gene sets for *Malassezia*. (B) The three

Saccharomyces and one *Aspergillus* significant sets. No set is shared across genera: the mycobiome is not one trait but several.

Figure 3. *Aspergillus* gene and pathway MAGMA results. (A) Gene-based scan: no gene passes Bonferroni correction (minimum $P = 5.8 \times 10^{-5}$). (B) Gene-set scan: arginine metabolism is the only significant set among 17,009 tested.

Figure 4. Gene-based and gene-set MAGMA analysis for PBC (N = 24,510). (A) Genome-wide gene-based scan (8,836 genes; 381 Bonferroni-significant); the MHC tower and named non-MHC loci (TIMMDC1, GPSM3, DLEU1) match the established PBC architecture. (B) Top 15 of 217 significant gene sets (16,447 tested), all adaptive-immune in theme; dashed line, Bonferroni threshold.

Figure 5. PSC gene and pathway MAGMA results (N = 14,890; 18,277 genes; 17,006 sets). (A) Gene-based scan: the MHC tower dominates (HIST1H4A $P = 5.2 \times 10^{-45}$; y-axis truncated at $-\log_{10}(P) = 22$); GPX1 is the top non-MHC gene. (B) Pathway scan: top 10 of 20 Bonferroni-significant sets; MHC class II antigen processing and adaptive immune response dominate.

Figure 6. Regional association tracks at the three candidate loci (±500 kb, GRCh37). Fungal GWAS (upper track) versus AILD GWAS (lower track). Shaded band: lead position. Black diamonds mark the lead variant; the open diamond in (B) marks the nearest PBC variant because the lead variant is absent there. (A) chr14:30.9 Mb, *Aspergillus* × PBC (the PBC track carries no variant at the lead); this candidate rests on the window-level nominal criterion alone (tier 3). (B) chr11:67.2 Mb, *Malassezia* × PBC (nearest nominal variant rs3741180, 1.2 kb from the lead, P = 0.021; the locus also overlaps PSC, whose leg is carried by rs1077256, P = 0.024). (C) chr4:130.2 Mb, *Malassezia* × PSC (PSC P = 0.039 at the lead; single-SNP triangulation passed).

Figure 7. eQTL support at the two *Malassezia* candidate loci. Each point is one SNP-gene pair from the eQTLGen whole-blood scan (326 SNP-gene pairs retained genome-wide after false discovery rate filtering; 105 fall within the two ±500 kb candidate windows shown). x-axis: *Malassezia* GWAS $-\log_{10}(P)$; y-axis: |eQTL Z|. Black diamonds mark the coloc lead SNPs. (A) chr11:67.2 Mb: the lead rs1626067 is a cis-eQTL for AIP (Z = −12.5; neighbouring cluster SNPs reach AIP Z = −15.0), SSH3 (−17.9), CORO1B (−18.4) and GSTP1 (−5.3), among seven cluster genes. (B) chr4:130.2 Mb: the lead rs59634627 regulates C4orf33 (Z = −9.0), SCLT1 (−7.7) and PHF17 (−5.8).

Figure 8. MAGMA cell-type enrichment across eight blood/bone-marrow single-cell reference datasets. (A–C) Top cell types for *Malassezia*, PBC and PSC in the best-powered dataset for each trait; coloured bars pass within-dataset Bonferroni correction. (D) PSC enrichment across the human reference datasets. *Malassezia* signals are exclusively myeloid; PBC signals are predominantly lymphoid and PSC signals are T-lineage.

Figure 9. Working model of the candidate fungus-immune genetic convergence. The fungal side (blue: *Malassezia*, intracellular processing, myeloid compartment) and the disease side (red: AILD, adaptive immunity, lymphoid compartment) converge on IL-23R and on the chr4/chr11 candidate loci, with evidence graded from concordant single-SNP evidence (chr4) to regional overlap alone (tier-3 locus omitted). The antifungal-recognition pathway offers the most immediate pharmacological entry point (SYK-fostamatinib); CARD9, MALT1 and CLEC7A require new chemistry. All convergence points are hypothesis-generating.

Additional files

The following items use real data and are provided as additional files because they support but do not carry the main claims. Each is explained where it appears.

Additional file 1: Table S1. GOBP arginine metabolic process member genes (*Aspergillus*).

Additional file 2: Table S2. Member genes of the three *Saccharomyces* significant sets.

Additional file 3: Table S3. Nominal immune genes shared direction across the two yeasts.

Additional file 4: Table S4. Full list of PBC significant gene sets (n = 217).

Additional file 5: Table S5. Summary of the 680 colocalization tests (full result file in Data availability).

Additional file 6: Table S6. Top gene-level hits for PBC.

Additional file 7: Table S7. Reference-panel sensitivity and version-matched replication.

Additional file 8: coloc_fungi_aild_results.tsv. Full colocalization results for all 680 tested windows (fungal trait, AILD, lead SNP, fungal P, chromosome, base pair, overlapping variant count and posterior probabilities PP.H0–PP.H4).

Additional file 9: Table S8. Gene-level MAGMA summary for all eight screened fungal traits.

Additional file 10: Table S9. Distance-pruned lead SNPs (144 of 485 retained; the chr4:130.2 Mb and chr11:67.2 Mb leads are retained).

Additional file 11: Table S10. Observed window statistics for the 340 PSC windows used in the permutation analysis.

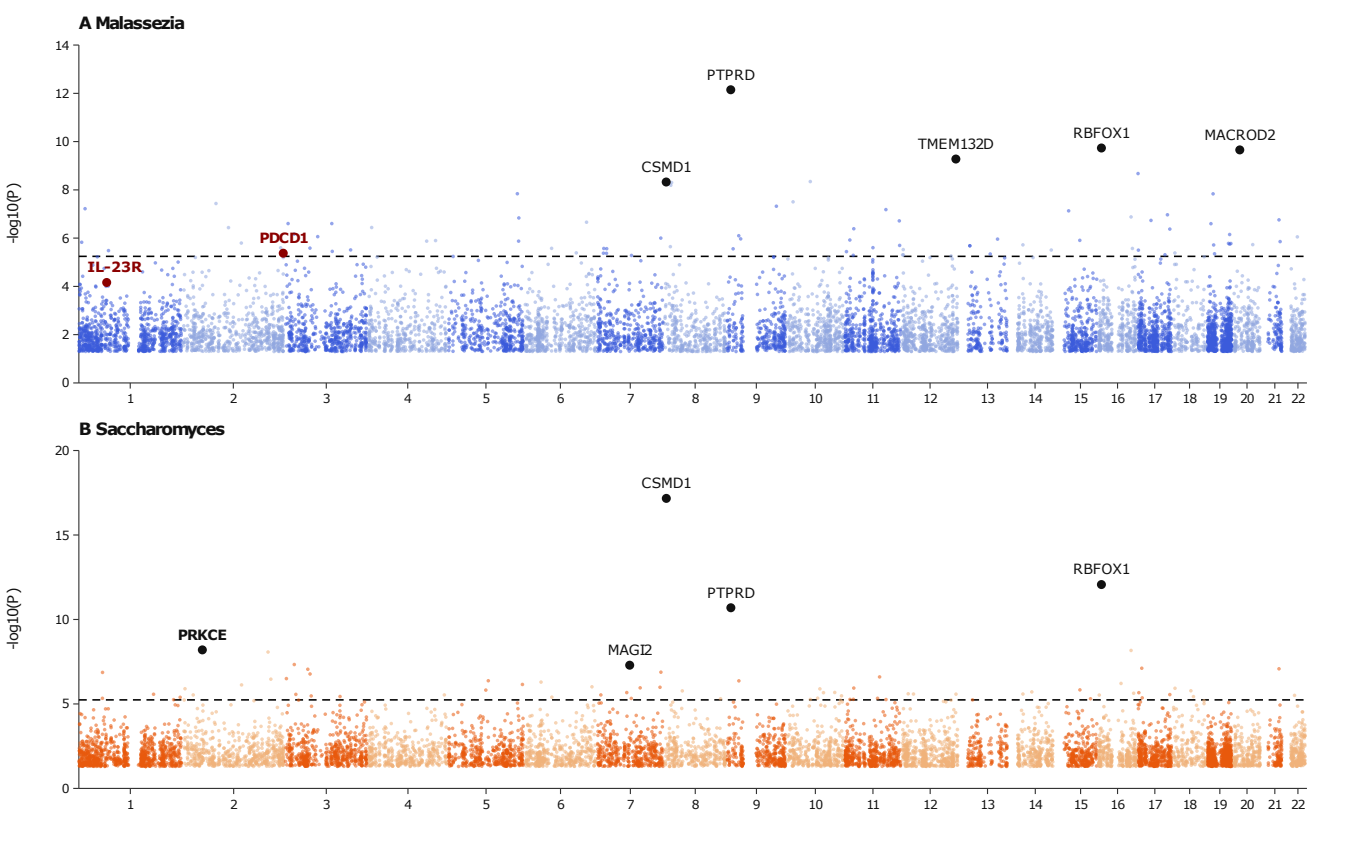
Additional file 12: source_data.zip. Machine-readable outputs underlying the new analyses (permutation null and observed summaries; fine-mapped QTL credible-set lookups; external FAV lookup).

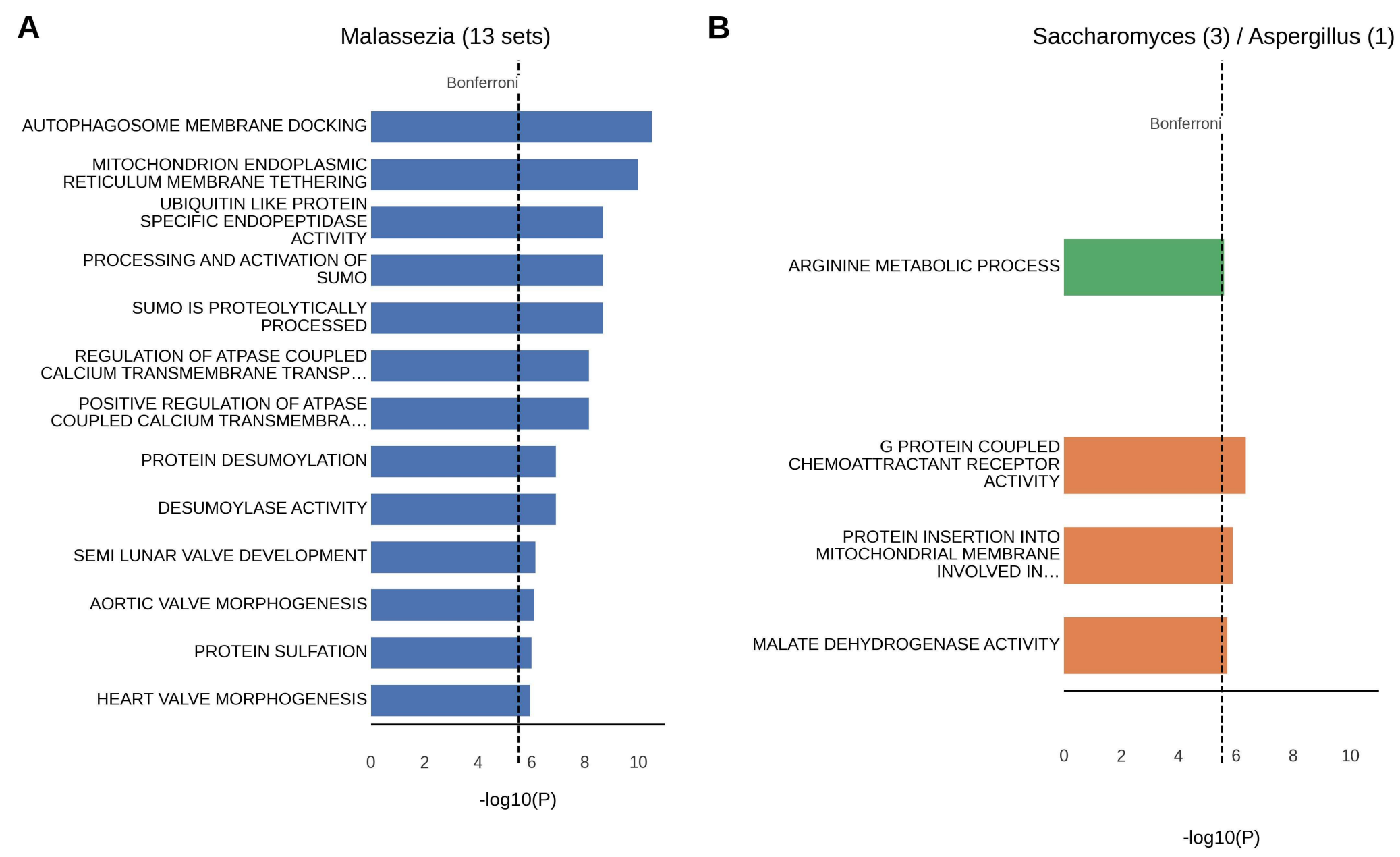
Table 7. Bonferroni-significant cell types (eight scRNA-seq reference datasets).

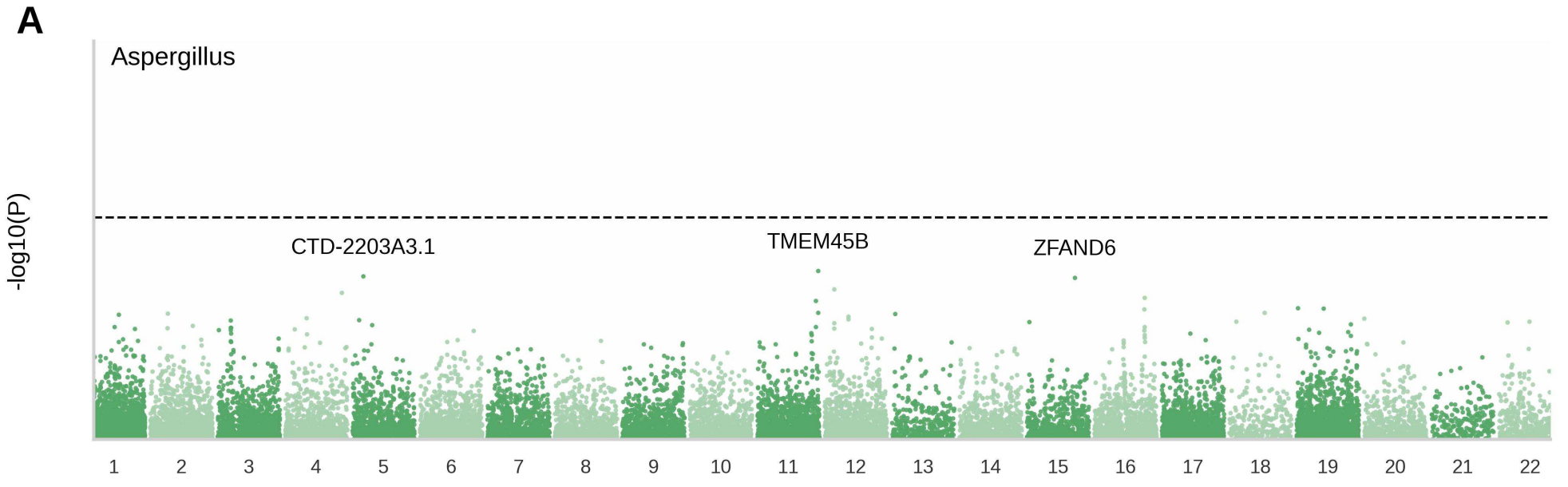
Trait	Dataset	Cell type	P	Bonferroni α
Malassezia	541 Xu Human 2023 Blood level1	non_classical_monocyte	2.00e-06	2.08e-03
Malassezia	539 Travaglini 2020 Blood level1	non_classical_monocyte	5.73e-06	2.63e-03
Malassezia	541 Xu Human 2023 Blood level1	classical_monocyte	1.43e-05	2.08e-03
Malassezia	542 Xu Human 2023 Bone marrow level1	classical_monocyte	1.14e-04	1.32e-03
Malassezia	542 Xu Human 2023 Bone marrow level1	non_classical_monocyte	1.40e-04	1.32e-03
Malassezia	TabulaMuris FACS Marrow	monocyte	1.51e-04	2.38e-03
Malassezia	542 Xu Human 2023 Bone marrow level1	neutrophil	3.45e-04	1.32e-03
Malassezia	MouseCellAtlas Bone Marrow	Bone_Marrow.Neutrophil_Mmp8_high	3.61e-04	9.62e-04
Malassezia	542 Xu Human 2023 Bone marrow level1	promyelocyte	5.69e-04	1.32e-03
Malassezia	MouseCellAtlas Peripheral Blood	Macrophage_Pf4_high	8.06e-04	2.38e-03
Malassezia	539 Travaglini 2020 Blood level1	classical_monocyte	9.05e-04	2.63e-03
PBC	MouseCellAtlas Peripheral Blood	B_cell_Ly6d_high	3.48e-06	2.38e-03
PBC	MouseCellAtlas Bone Marrow	Bone_Marrow.Hematopoietic_stem_progenitor_cell	1.20e-05	9.62e-04
PBC	542 Xu Human 2023 Bone marrow level1	mucosal_invariant_T_cell	2.02e-05	1.32e-03
PBC	MouseCellAtlas Bone Marrow	Bone_Marrow_c_kitMacrophage_Cd74_high	1.19e-04	9.62e-04

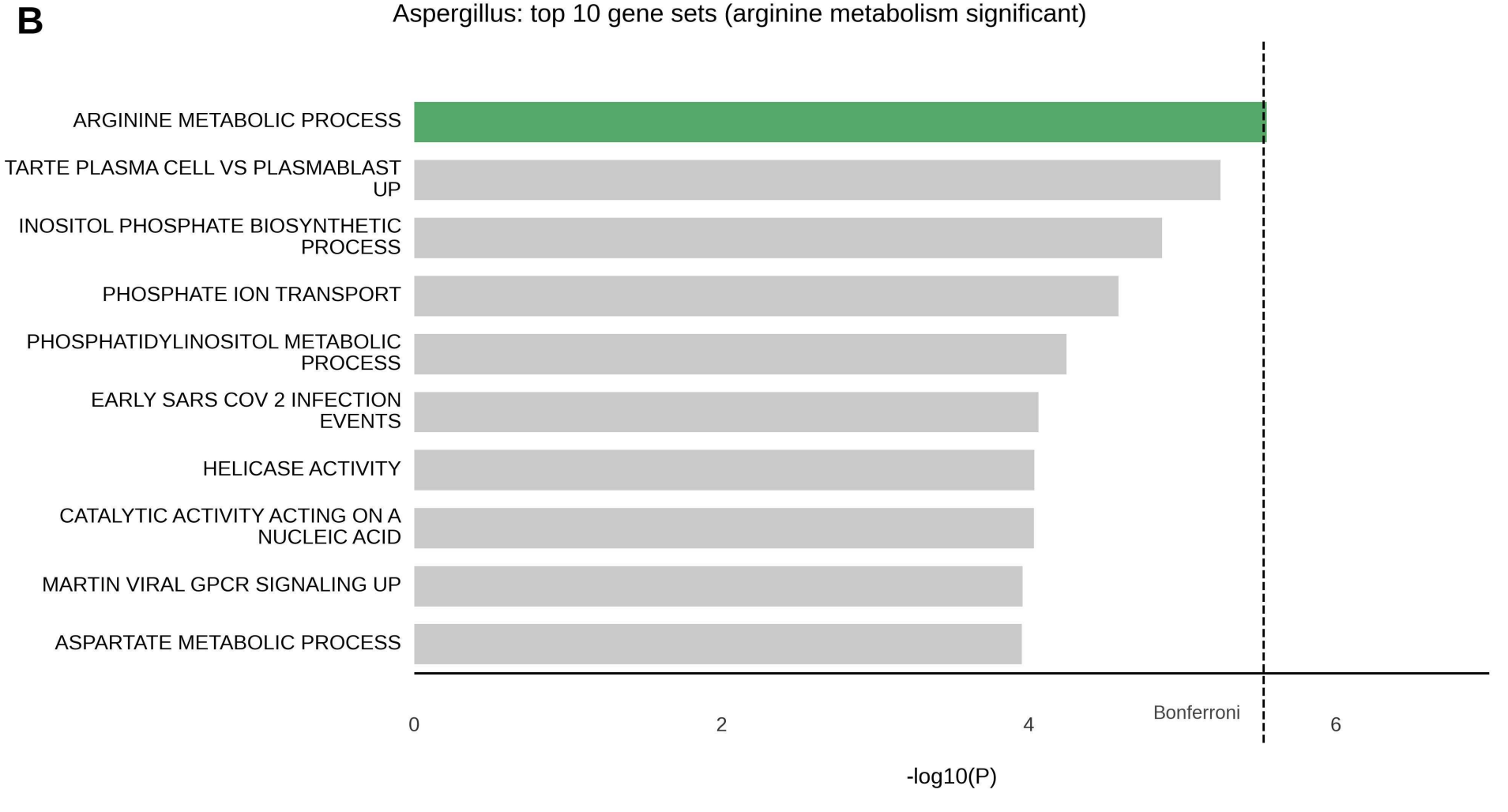
PBC	MouseCellAtlas Bone Marrow	Bone_Marrow.B_cell_Igkc_high	1.50e-04	9.62e-04
PBC	542 Xu Human 2023 Bone marrow level1	effector_memory_CD8_positive__alpha_beta_T_cell__terminally_differentiated	3.30e-04	1.32e-03
PBC	542 Xu Human 2023 Bone marrow level1	memory_B_cell	3.46e-04	1.32e-03
PBC	MouseCellAtlas Peripheral Blood	Macrophage_Flt.ps1_high	4.30e-04	2.38e-03
PBC	TabulaMuris FACS Marrow	immature_B_cell	4.39e-04	2.38e-03
PBC	MouseCellAtlas Peripheral Blood	Macrophage_S100a4_high	4.44e-04	2.38e-03
PBC	542 Xu Human 2023 Bone marrow level1	regulatory_T_cell	6.06e-04	1.32e-03
PBC	542 Xu Human 2023 Bone marrow level1	effector_memory_CD8_positive__alpha_beta_T_cell	6.41e-04	1.32e-03
PBC	TabulaMuris FACS Marrow	naive_B_cell	6.77e-04	2.38e-03
PBC	MouseCellAtlas Peripheral Blood	NK_cell_Gzma_high	1.28e-03	2.38e-03
PBC	TabulaMuris droplet Marrow	monocyte	4.22e-03	6.25e-03
PSC	Xu Human 2023 Bone marrow level1	T follicular helper cell	5.63e-04	1.32e-03
PSC	Xu Human 2023 Bone marrow level1	mucosal invariant T cell	5.95e-04	1.32e-03

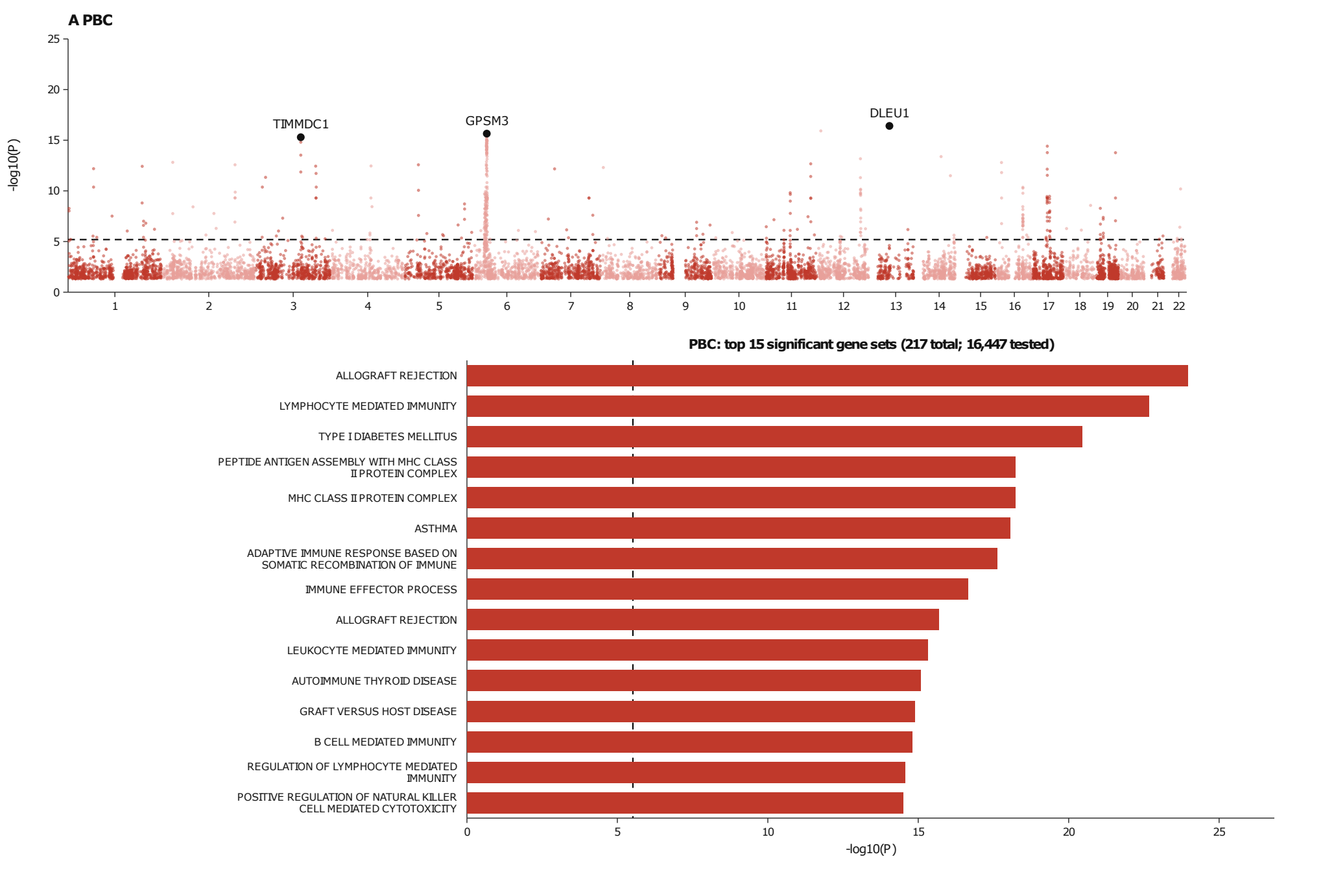
Note. Bonferroni α denotes the significance threshold after correction for the number of cell-type tests performed within each individual reference dataset. Cell-type labels are reproduced verbatim from the source reference datasets.

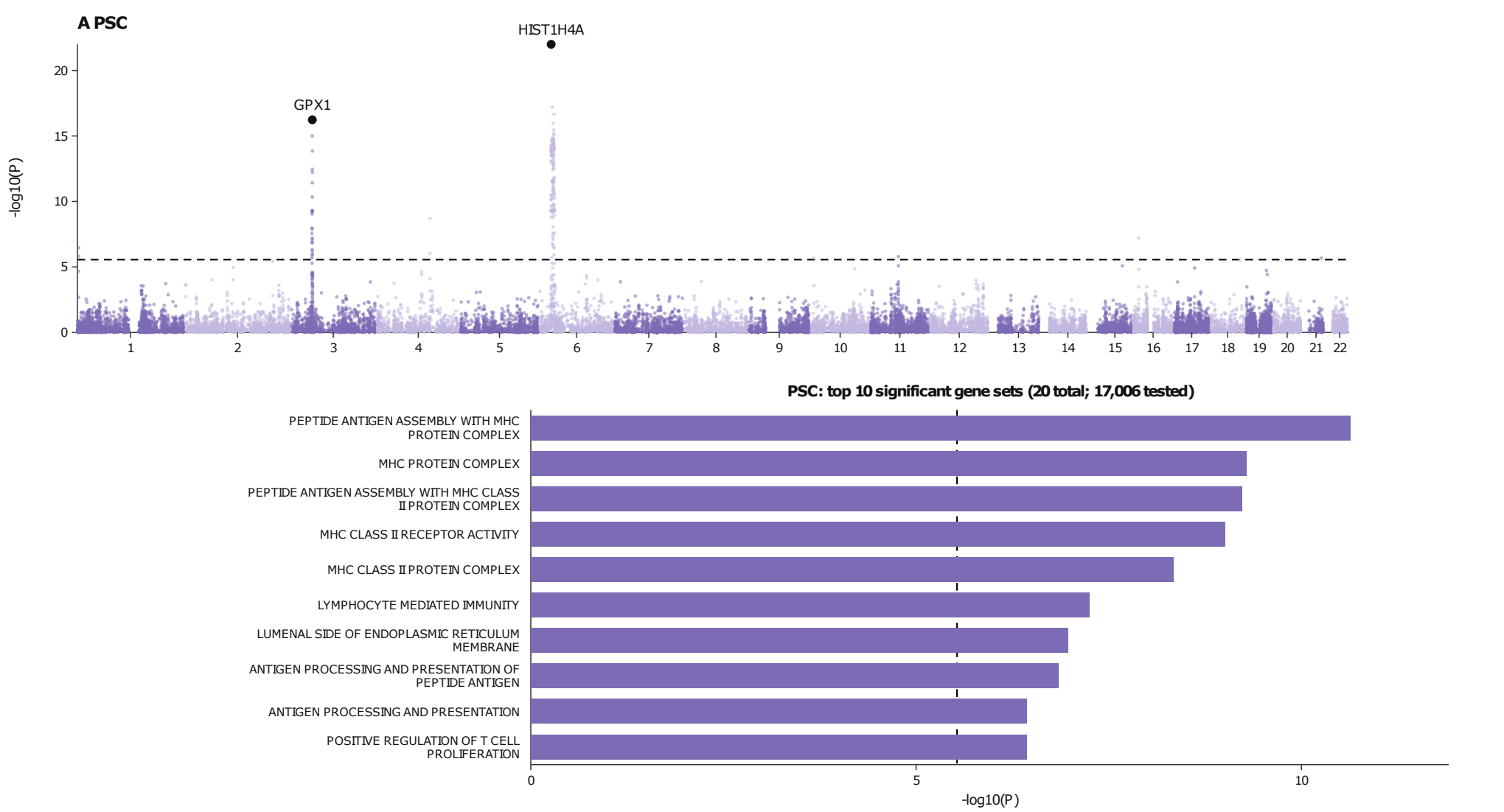

A Malassezia
-log10(P)
PTPRD
CSMD1
TMEM132D
RBFOX1
MACROD2
PDCD1
IL-23R
B Saccharomyces
-log10(P)
CSMD1
RBFOX1
PTPRD
PRKCE
MAGI2

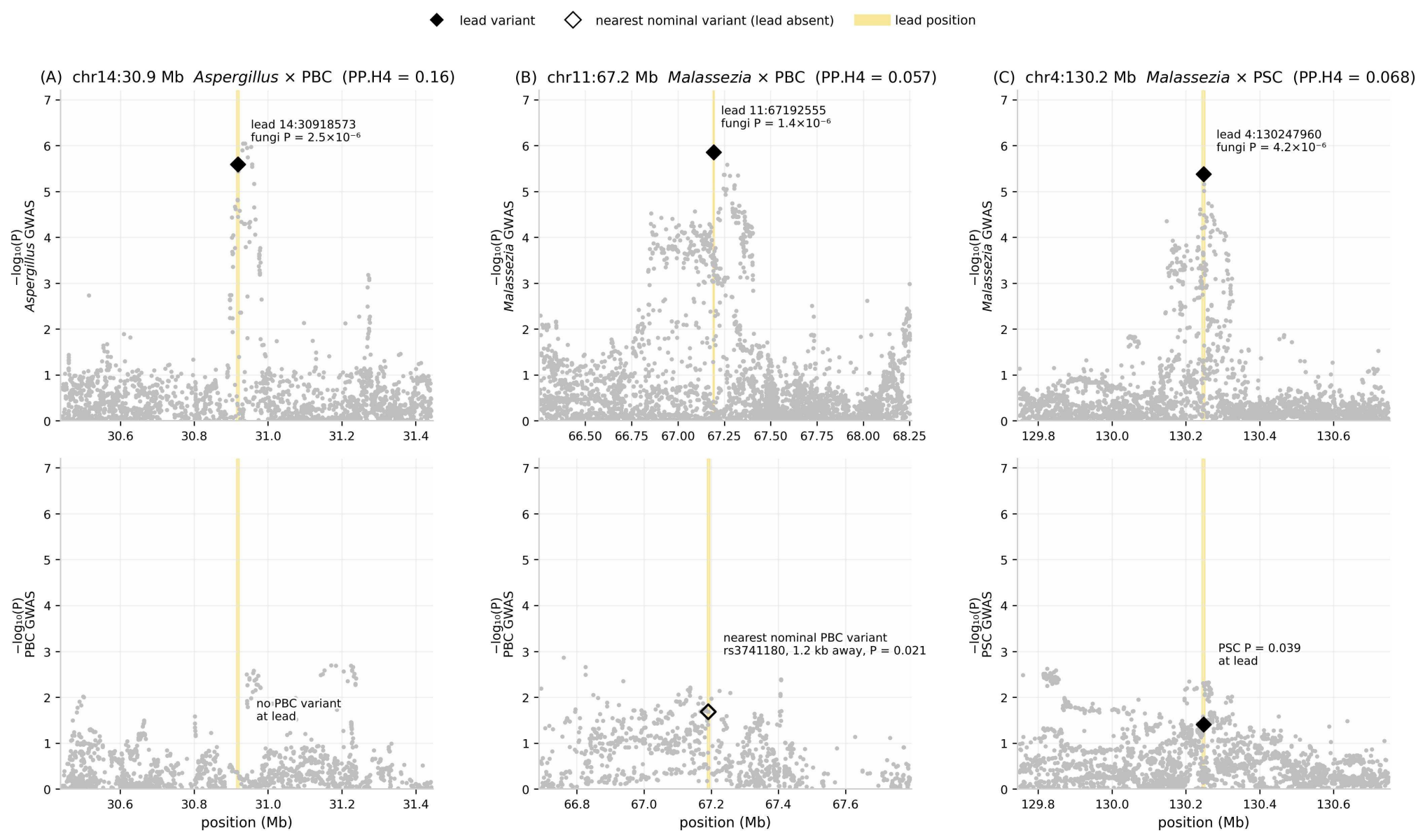

A
Malassezia (13 sets)
Bonferroni
AUTOPHAGOSOME MEMBRANE DOCKING
MITOCHONDRION ENDOPLASMIC RETICULUM MEMBRANE TETHERING
UBIQUITIN LIKE PROTEIN SPECIFIC ENDOPEPTIDASE ACTIVITY
PROCESSING AND ACTIVATION OF SUMO
SUMO IS PROTEOLYTICALLY PROCESSED
REGULATION OF ATPASE COUPLED CALCIUM TRANSMEMBRANE TRANSP…
POSITIVE REGULATION OF ATPASE COUPLED CALCIUM TRANSMEMBRA…
PROTEIN DESUMOYLATION
DESUMOYLASE ACTIVITY
SEMI LUNAR VALVE DEVELOPMENT
AORTIC VALVE MORPHOGENESIS
PROTEIN SULFATION
HEART VALVE MORPHOGENESIS
0
2
4
6
8
10
-log10(P)
B
Saccharomyces (3) / Aspergillus (1)
Bonferroni
ARGININE METABOLIC PROCESS
G PROTEIN COUPLED CHEMOATTRACTANT RECEPTOR ACTIVITY
PROTEIN INSERTION INTO MITOCHONDRIAL MEMBRANE INVOLVED IN…
MALATE DEHYDROGENASE ACTIVITY
0
2
4
6
8
10
-log10(P)

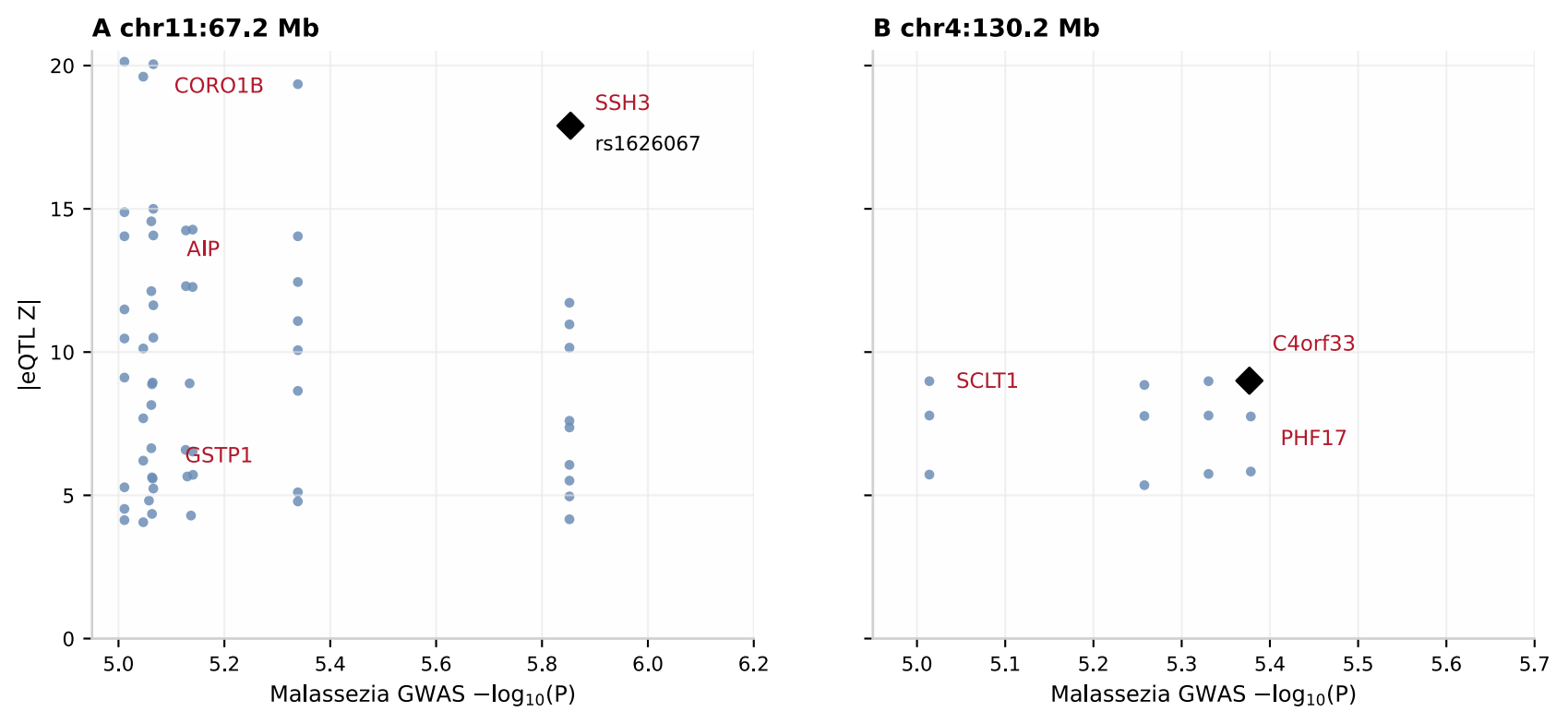

A
Aspergillus
-log10(P)
CTD-2203A3.1
TMEM45B
ZFAND6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

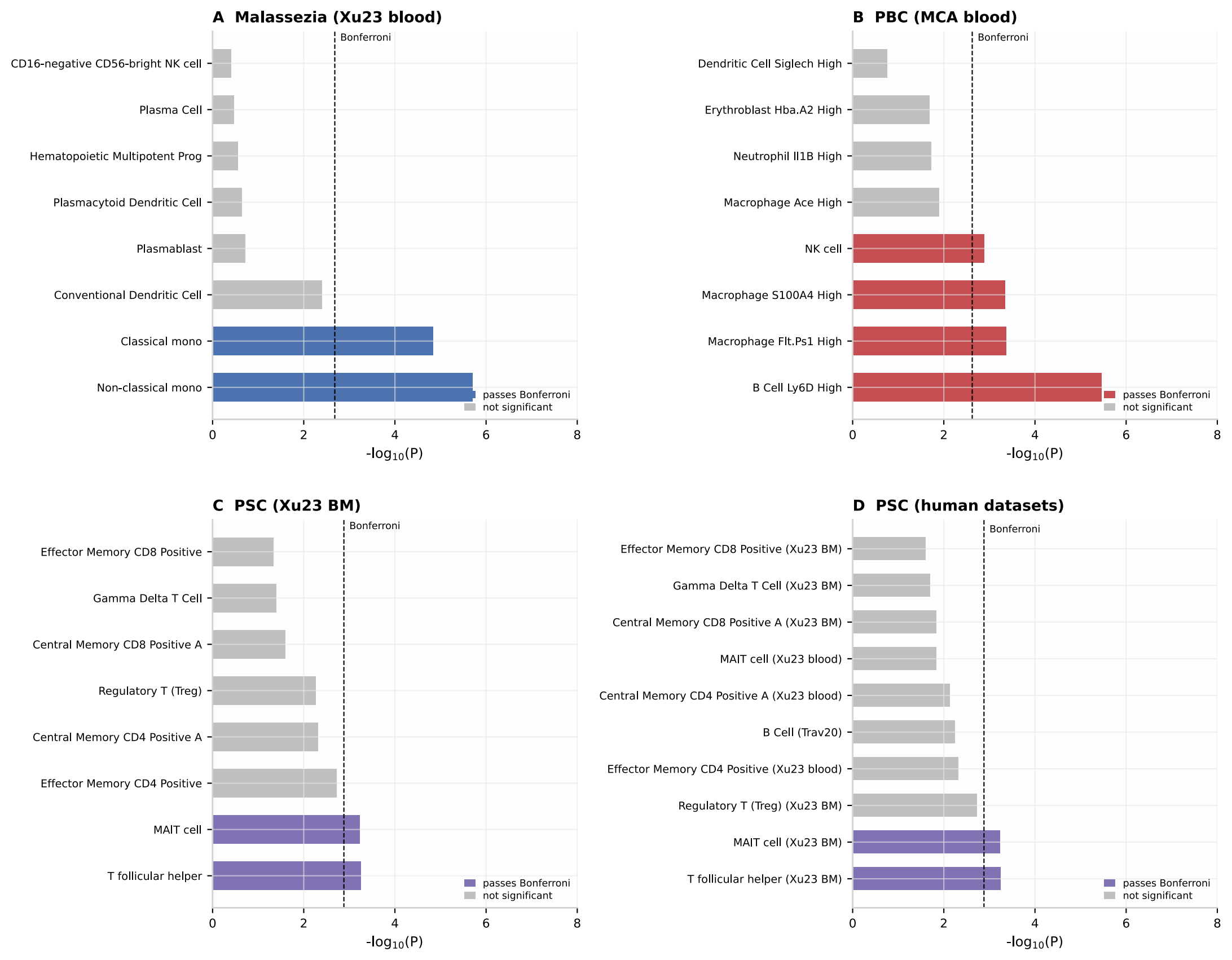

B
Aspergillus: top 10 gene sets (arginine metabolism significant)
ARGININE METABOLIC PROCESS
TARTE PLASMA CELL VS PLASMABLAST UP
INOSITOL PHOSPHATE BIOSYNTHETIC PROCESS
PHOSPHATE ION TRANSPORT
PHOSPHATIDYLINOSITOL METABOLIC PROCESS
EARLY SARS COV 2 INFECTION EVENTS
HELICASE ACTIVITY
CATALYTIC ACTIVITY ACTING ON A NUCLEIC ACID
MARTIN VIRAL GPCR SIGNALING UP
ASPARTATE METABOLIC PROCESS
0
2
4
6
Bonferroni
-log10(P)

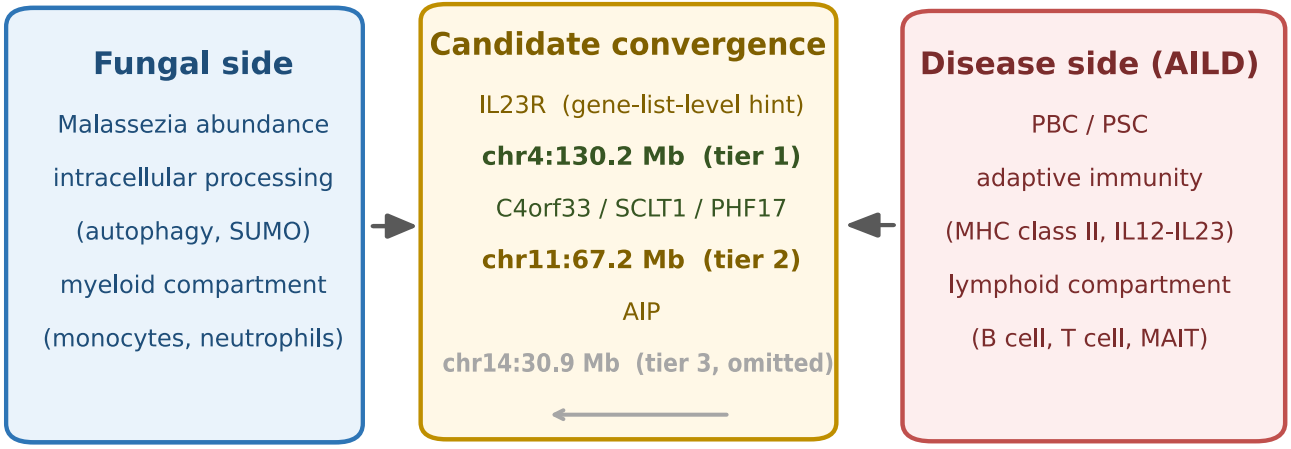

A PBC
-log10(P)
TIMMDC1
GPSM3
DLEU1
PBC: top 15 significant gene sets (217 total; 16,447 tested)
ALLOGRAFT REJECTION
LYMPHOCYTE MEDIATED IMMUNITY
TYPE I DIABETES MELLITUS
PEPTIDE ANTIGEN ASSEMBLY WITH MHC CLASS II PROTEIN COMPLEX
MHC CLASS II PROTEIN COMPLEX
ASTHMA
ADAPTIVE IMMUNE RESPONSE BASED ON SOMATIC RECOMBINATION OF IMMUNE
IMMUNE EFFECTOR PROCESS
ALLOGRAFT REJECTION
LEUKOCYTE MEDIATED IMMUNITY
AUTOIMMUNE THYROID DISEASE
GRAFT VERSUS HOST DISEASE
B CELL MEDIATED IMMUNITY
REGULATION OF LYMPHOCYTE MEDIATED IMMUNITY
POSITIVE REGULATION OF NATURAL KILLER CELL MEDIATED CYTOTOXICITY
-log10(P)


A PSC
HIST1H4A
GPX1
-log10(P)
0
5
10
15
20
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
PSC: top 10 significant gene sets (20 total; 17,006 tested)
PEPTIDE ANTIGEN ASSEMBLY WITH MHC PROTEIN COMPLEX
MHC PROTEIN COMPLEX
PEPTIDE ANTIGEN ASSEMBLY WITH MHC CLASS II PROTEIN COMPLEX
MHC CLASS II RECEPTOR ACTIVITY
MHC CLASS II PROTEIN COMPLEX
LYMPHOCYTE MEDIATED IMMUNITY
LUMENAL SIDE OF ENDOPLASMIC RETICULUM MEMBRANE
ANTIGEN PROCESSING AND PRESENTATION OF PEPTIDE ANTIGEN
ANTIGEN PROCESSING AND PRESENTATION
POSITIVE REGULATION OF T CELL PROLIFERATION
0
5
10
-log10(P)

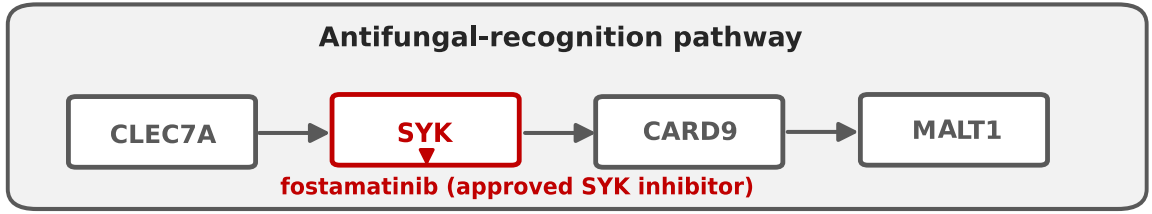

lead variant
nearest nominal variant (lead absent)
lead position
(A) chr14:30.9 Mb Aspergillus × PBC (PP.H4 = 0.16)
(B) chr11:67.2 Mb Malassezia × PBC (PP.H4 = 0.057)
(C) chr4:130.2 Mb Malassezia × PSC (PP.H4 = 0.068)
lead 14:30918573
fungi P = 2.5×10⁻⁶
lead 11:67192555
fungi P = 1.4×10⁻⁶
lead 4:130247960
fungi P = 4.2×10⁻⁶
−log10(P) Aspergillus GWAS
−log10(P) Malassezia GWAS
−log10(P) Malassezia GWAS
no PBC variant at lead
nearest nominal PBC variant
rs3741180, 1.2 kb away, P = 0.021
PSC P = 0.039
at lead
−log10(P) PBC GWAS
−log10(P) PBC GWAS
−log10(P) PSC GWAS
position (Mb)
position (Mb)
position (Mb)


A chr11:67.2 Mb
B chr4:130.2 Mb
|eQTL Z|
Malassezia GWAS −log10(P)
CORO1B
SSH3
rs1626067
AIP
GSTP1
SCLT1
C4orf33
PHF17


A Malassezia (Xu23 blood)
Bonferroni
CD16-negative CD56-bright NK cell
Plasma Cell
Hematopoietic Multipotent Prog
Plasmacytoid Dendritic Cell
Plasmablast
Conventional Dendritic Cell
Classical mono
Non-classical mono
passes Bonferroni
not significant
0
2
4
6
8
-log10(P)
B PBC (MCA blood)
Bonferroni
Dendritic Cell Siglech High
Erythroblast Hba.A2 High
Neutrophil Il1B High
Macrophage Ace High
NK cell
Macrophage S100A4 High
Macrophage Flt.Ps1 High
B Cell Ly6D High
passes Bonferroni
not significant
0
2
4
6
8
-log10(P)
C PSC (Xu23 BM)
Bonferroni
Effector Memory CD8 Positive
Gamma Delta T Cell
Central Memory CD8 Positive A
Regulatory T (Treg)
Central Memory CD4 Positive A
Effector Memory CD4 Positive
MAIT cell
T follicular helper
passes Bonferroni
not significant
0
2
4
6
8
-log10(P)
D PSC (human datasets)
Bonferroni
Effector Memory CD8 Positive (Xu23 BM)
Gamma Delta T Cell (Xu23 BM)
Central Memory CD8 Positive A (Xu23 BM)
MAIT cell (Xu23 blood)
Central Memory CD4 Positive A (Xu23 blood)
B Cell (Trav20)
Effector Memory CD4 Positive (Xu23 blood)
Regulatory T (Treg) (Xu23 BM)
MAIT cell (Xu23 BM)
T follicular helper (Xu23 BM)
passes Bonferroni
not significant
0
2
4
6
8
-log10(P)



evidence grade: concordant single SNP (tier 1) → regional overlap alone (tier 3)



All convergence points are hypothesis-generating.

Supplementary Files

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

- Additionalfile1.docx
- Additionalfile2.docx
- Additionalfile3.docx
- Additionalfile4.docx
- Additionalfile5.docx
- Additionalfile6.docx
- Additionalfile7.docx
- Additionalfile9TableS8.docx
- Additionalfile8.tsv
- Additionalfile12sourcedata.zip
- TableS9.docx
- TableS10.docx