

The genetic basis of autoimmunity seen through the lens of T cell functional traits

Received: 16 August 2023

Accepted: 15 January 2024

Published online: 08 February 2024

 Check for updates

Kaitlyn A. Lagattuta^{1,2,3,4,5}, Hannah L. Park^{1,2,3,4,5}, Laurie Rumker^{1,2,3,4,5}, Kazuyoshi Ishigaki^{1,2,3,4,6}, Aparna Nathan^{1,2,3,4,5,7} ✉ & Soumya Raychaudhuri^{1,2,3,4,5,7} ✉

Autoimmune disease heritability is enriched in T cell-specific regulatory regions of the genome. Modern-day T cell datasets now enable association studies between single nucleotide polymorphisms (SNPs) and a myriad of molecular phenotypes, including chromatin accessibility, gene expression, transcriptional programs, T cell antigen receptor (TCR) amino acid usage, and cell state abundances. Such studies have identified hundreds of quantitative trait loci (QTLs) in T cells that colocalize with genetic risk for autoimmune disease. The key challenge facing immunologists today lies in synthesizing these results toward a unified understanding of the autoimmune T cell: which genes, cell states, and antigens drive tissue destruction?

Genetic risk variants for autoimmune disease are enriched in T cell-specific regulatory regions of the genome^{1,2} and are disproportionately close to genes with T cell-specific functions³. These findings imply that T cells are critically important to the development of autoimmune disease. To understand how T cells change to enact autoimmune destruction, we can identify quantitative trait loci (QTLs). A QTL is a location in the genome in which someone's DNA sequence helps to explain some quantifiable characteristic of the individual. These characteristics range from general traits such as birthweight⁴ to detailed molecular traits such as the expression level of a certain gene. By using molecular sequencing data, we can now conduct QTL studies to search for the T cell molecular traits (e.g., gene expression, chromatin accessibility) that mediate risk for autoimmunity.

An expression QTL (eQTL) is a QTL that explains a statistically significant proportion of variance in the expression of a gene. The mechanism for this change in gene expression largely depends on how close the eQTL is to the target gene. Located near their target genes, *cis*-eQTLs impact expression often by altering transcription factor binding in a proximal regulatory element⁵, or altering the rate of mRNA degradation⁶. By contrast, *trans*-eQTLs may regulate distant gene targets by first altering the expression of a nearby transcription factor

(*cis*-eQTL mediation)⁷. However, *trans*-eQTLs can also emerge from unusual mechanisms. For example, *trans*-eQTLs in the major histocompatibility (MHC) locus on chromosome 6 ultimately affect the expression of TCR genes on chromosomes 7 and 14⁸. Most likely, these *trans*-eQTLs alter which antigenic peptides can be bound and presented by HLA, which in turn shapes the thymic selection of TCRs⁹.

eQTLs may help to identify genes that mediate genetic risk for autoimmune disease. For example, rs3087243 is a single-nucleotide polymorphism (SNP) (reference allele: "G", alternate allele: "A") near the gene for *CTLA4*, a vital negative regulator of T cell activation¹⁰. An individual can have 0, 1, or 2 copies of the "G" allele. Since each additional copy of the "G" allele corresponds to a decrease in the expression of the nearby gene *CTLA4*^{11,12}, rs3087243 is a *cis*-eQTL. The "G" allele for rs3087243 not only corresponds to decreased expression of *CTLA4*, but also elevated risk for Graves' disease, rheumatoid arthritis (RA), and type-1 diabetes (T1D)^{13–15}. This overlap at rs3087243 suggests that decreased expression of *CTLA4* is in part responsible for autoimmune destruction in Graves', RA, and T1D. Analogously, *trans*-eQTLs for TCR genes raise the possibility that the MHC locus shapes risk for autoimmunity by promoting the thymic selection of autoreactive TCRs⁹. In this review, we will examine recent approaches to identify T

¹Center for Data Sciences, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ²Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. ³Division of Rheumatology, Inflammation, and Immunity, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. ⁴Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁵Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. ⁶Laboratory for Human Immunogenetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. ⁷These authors contributed equally: Aparna Nathan, Soumya Raychaudhuri.

✉ e-mail: aparna_nathan@hms.harvard.edu; soumya@broadinstitute.org

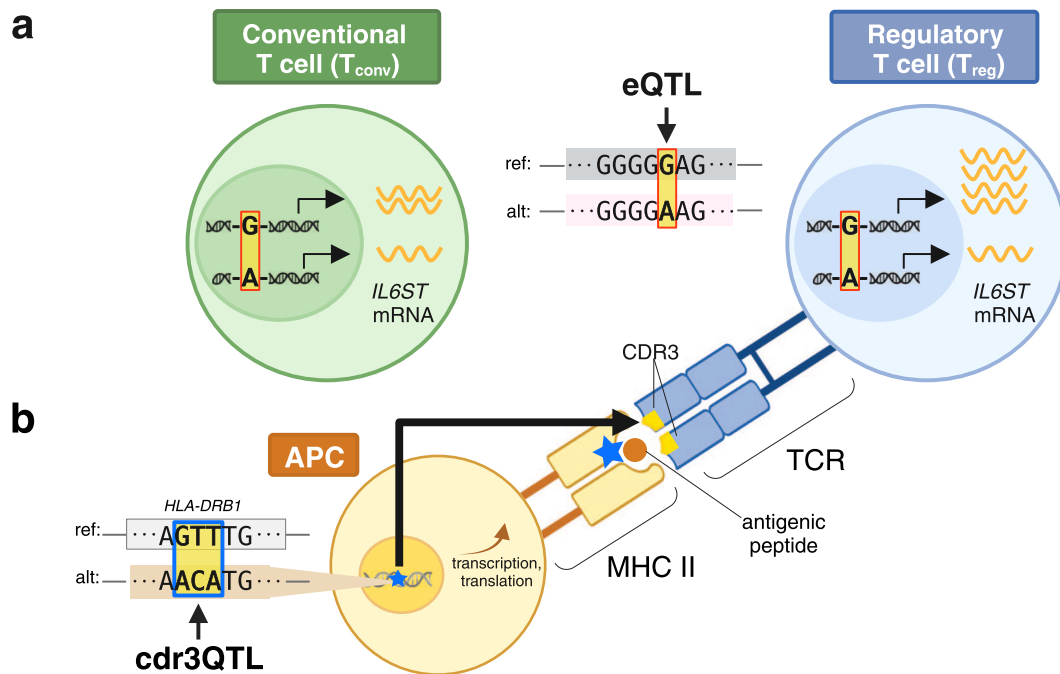


Fig. 1 | A schematic of quantitative trait loci (QTLs) affecting T cell functional traits. a A T cell state-dependent expression QTL (eQTL) at rs7731626 is depicted in two different T cells: one in a conventional T cell state (T_{conv} , left) and one in a regulatory T cell state (T_{reg} , right). Based on eQTL studies, we expect the “G” risk allele for rheumatoid arthritis and multiple sclerosis to increase *IL6ST* expression to a greater extent in the T_{reg} compared to the T_{conv} . **b** The associated molecular trait for a cdr3QTL is amino acid usage in the T cell receptor (TCR). In this antigen

presenting cell (APC), *HLA-DRB1* is transcribed, translated, and loaded with antigen while harboring a genetic variant at amino acid position 13. This genetic variant in *HLA-DRB1* appears to affect amino acid usage in the TCR, by influencing which T cells survive in the thymus. The blue star indicates the location of the cdr3QTL, at the level of *HLA-DRB1* DNA as well as HLA-DRB1 protein. QTL quantitative trait locus, ref reference, alt alternative, CDR3 complementarity determining region 3, MHC major histocompatibility complex. Created with BioRender.com.

cell QTLs, and discuss how to connect numerous types of T cell QTLs to autoimmune disease.

Expression quantitative trait loci (eQTLs): the importance of T cell state

Unfortunately, most risk variants for autoimmune disease do not coincide with the eQTLs identified through bulk RNA sequencing (RNAseq)¹⁶. While rs3087243 clearly implicates *CTLA4* gene expression in RA pathogenesis, cases such as these are quite infrequent. Accounting for linkage disequilibrium (LD), only 25% of neighboring (<100 kb) autoimmune and eQTL associations actually appear to share underlying genetic causes (“statistical colocalization”), using bulk RNAseq data¹⁶.

It is possible that eQTLs relevant to disease only occur in some cells of a given sample. If so, eQTL effects within these pathogenic cells may become diluted and undetectable when all cell types are mixed together for bulk RNAseq. Consistent with this hypothesis, dividing bulk RNAseq tissue samples into constituent cell types by flow cytometry¹⁷ or *in silico* deconvolution¹⁸ has nominated more eQTLs that are significantly enriched for disease loci.

Single-cell RNAseq (scRNAseq) data allows researchers to accurately categorize and analyze the different cell types present in a sample. With scRNAseq, recent studies have confirmed that eQTL effects can depend on the type of cell under study. For example, in a study of memory T cells sampled from the peripheral blood of 259 individuals, we observed that additional copies of the “G” allele for the RA-associated SNP rs7731626 corresponded to a larger increase in *IL6ST* expression in regulatory T (T_{reg}) cells compared to non-regulatory T cells¹² (Fig. 1a). Evidence that the functional consequences of rs7731626 may be concentrated in T_{reg} cells could extend to many other autoimmune-associated SNPs, which are known to be enriched in certain naïve T_{reg} -specific regulatory regions of the

genome^{1,2,19}. T_{reg} cells, while highly relevant to immune-mediated disease, constitute a rare cell state (<5% of peripheral T cells²⁰). Thus, it will be important to continue focused isolation and analysis of the T_{reg} cell population for eQTL discovery²¹.

There are multiple ways to represent how eQTL effects depend on transcriptional context. Our research group uses dimensionality reduction to identify groups of genes whose expression changes together, forming a gradient across cells. These transcriptional gradients often approximate the extent of a known T cell state, such as cytotoxicity. Our research group then identifies cell-state-dependent eQTLs, where a genotype’s effect on the expression of a gene varies along the transcriptional gradient¹². However, because transcriptional gradients may correlate with key marker genes, it is sometimes possible to reframe cell-state-dependent eQTLs as co-expression QTLs (co-eQTLs)^{22,23}. In the co-eQTL framing, the correlation in expression between gene A and gene B depends on a genotype. An alternative way to describe this phenomenon is that the genotype’s effect on gene A depends on the expression of gene B. If gene B is expressed in a specific cell state, cell-state-dependent eQTL and co-eQTL are synonymous terms. If gene B does not tag a cell state, the locus is not a cell-state-dependent eQTL, but would still be considered a co-eQTL. The co-eQTL framework, therefore, offers a broader definition of gene expression interaction. The vast number of possible pairings between genes and their regulatory loci precludes comprehensive detection of co-eQTLs, which would be required to estimate the proportion of eQTLs that are co-eQTLs. Alternatively, by focusing on cell states rather than individual genes, we were able to estimate that approximately one third (33%) of eQTLs in T cells depend on transcriptional state¹².

Recent studies have estimated that a substantial proportion of eQTLs depends on cell state, and that this proportion increases when considering autoimmune disease loci. Soskic*, Cano-Gamez* et al.²⁴

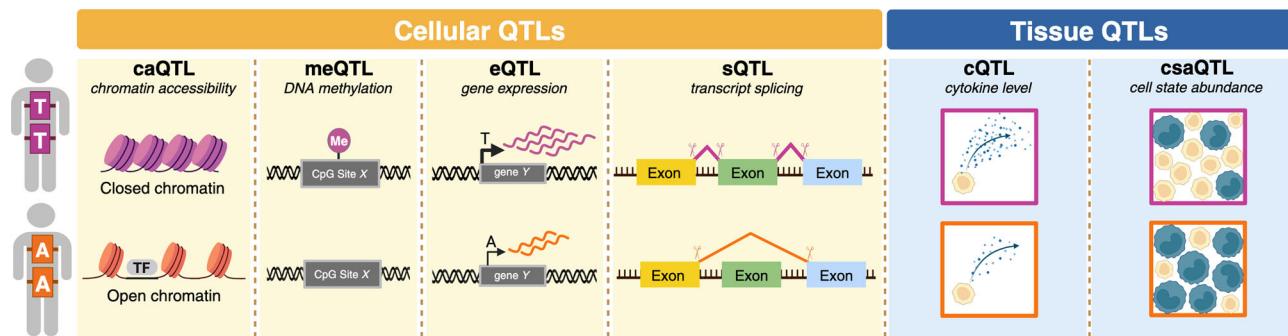


Fig. 2 | Quantitative Trait Loci (QTLs) can regulate diverse molecular phenotypes. Schematic illustrating molecular phenotypes that could be affected by a hypothetical quantitative trait locus (QTL). At this hypothetical locus, an individual may have a “TT” genotype (in purple), an “AA” genotype (in orange), or be heterozygous. Along the top row, we see molecular consequences of genotype “TT”. Along the bottom row, we see molecular consequences of genotype “AA.” We depict six types of QTLs as examples; this set of six is not comprehensive. From left to right: For the hypothetical caQTL, individuals with the “TT” genotype exhibit more closed chromatin than individuals with the “AA” genotype. For the hypothetical meQTL, CpG site X tends to be methylated in individuals with the “TT” genotype, but tends to be unmethylated in individuals with the “AA” genotype. For

the hypothetical eQTL, individuals with the “TT” genotype exhibit greater expression of gene Y compared to individuals with the “AA” genotype. For the hypothetical sQTL, RNA splicing in individuals with the “TT” genotype retains all exons in the transcript of gene Y, while RNA splicing in individuals with the “AA” genotype excludes the middle exon from the transcript of gene Y. For the hypothetical cQTL, individuals with the “TT” genotype exhibit higher levels of cytokine Z (in blue) compared to individuals with the “AA” genotype. For the hypothetical csaQTL, individuals with the “TT” genotype exhibit a greater relative abundance of the yellow cell state compared to individuals with the “AA” genotype. Me methylation, TF transcription factor. Created with Biorender.com.

profiled CD4⁺ T cells over a time course of anti-CD3/anti-CD28 stimulation, and found that 2265 of 6407 (35%) eQTLs depended on the activation state of the T cell. They then queried whether each eQTL statistically colocalized with genetic risk for immune-mediated disease. 60% of the colocalizing eQTLs were specific to T cell activation, and would have been missed if they did not account for this cell state dependence²⁴. In a similar study, Yazar*, Alquicira-Hernandez*, Wing* et al.²⁵ applied scRNAseq to PBMCs at steady state, and scanned for eQTLs that colocalized with genetic risk for autoimmune disease. They observed that 68% of the T cell eQTLs colocalizing with disease loci were detected in only one of the T cell states²⁵. Evidently, accounting for T cell states has substantially enhanced disease-relevant eQTL discovery in T cells. The majority of autoimmune GWAS associations, however, still remain unexplained.

Genetic regulation of T cell functions

While gene expression is an important molecular trait, any quantity that varies across people and is measured by molecular assays can be studied as a quantitative molecular trait. Expansion of multimodal single cell technologies has prompted an ever-growing list of quantitative molecular traits and genomic loci that regulate them. Other types of QTLs are steadily gaining recognition, such as chromatin accessibility QTLs (caQTLs), histone modification QTLs (hQTLs), DNA methylation QTLs (meQTLs), splicing QTLs (sQTLs), and protein QTLs (pQTLs) (Fig. 2). These QTLs are conceptually analogous to eQTLs, with gene expression substituted for some other molecular trait; particular approaches are comprehensively reviewed elsewhere²⁶. We and others have identified cell morphology QTLs (cmQTLs)²⁷ by measuring thousands of morphological phenotypes (e.g. size of mitochondria, granularity of endoplasmic reticulum) in vitro through multiplexed staining and imaging. cmQTLs can connect genetic variants to cellular function in human-derived, disease-relevant cell types.

In general, QTL studies query thousands of molecular traits simultaneously. Such investigations are relatively unbiased, data-driven, and can be readily applied to any cell type. However, the number of hypotheses considered imposes a substantial statistical multiple testing burden. As a result, studies often limit eQTL searches to SNPs within 1 Mb of the target gene (*cis*-eQTLs), and are only powered to detect eQTLs with relatively large effects²⁸.

For T cells, can we search for QTLs in a way that takes advantage of known T cell functions? In the 1970’s, the observation that activated

lymphocytes secrete proteins which markedly impact other lymphocytes spurred a series of supernatant studies, identifying soluble cytokines²⁹. These formative studies led to a common framework, in which the immune system can be understood through cytokine signaling. Cytokine signaling facilitates important T cell functions, including the attack of foreign antigens—or self-antigens, in the case of autoimmunity. Thus, identifying cytokine QTLs (cQTLs) may be an efficient and biologically interpretable way to search for disease-relevant molecular traits.

In a recent demonstration of the power of cQTLs, Nath et al.³⁰ analyzed 11 circulating cytokines, each of which may functionally represent the coordinated activity of thousands of genes. By starting with a curated set of molecular traits ($n=11$), Nath et al. had enough statistical power to scan the entire genome and find genetic regulation in *trans*. Rather than applying a univariate GWAS framework to each of the 11 cytokines separately, Nath et al. considered the 11 cytokines jointly through a multivariate GWAS framework. The multivariate framework finds an optimal linear combination of input traits (cytokine levels) regulated by each locus, a powerful strategy to detect QTLs from potentially redundant trait measurements. This approach identified eight cQTLs, one of which significantly colocalizes with genetic risk for ulcerative colitis (UC)³⁰.

An alternative way to leverage immunological knowledge in the search for disease-relevant QTLs is to focus on cell state abundance QTLs (csaQTLs). A recent large-scale csaQTL study³¹ analyzed 118 immune cell populations isolated by canonical surface markers in 3357 individuals. Orrù et al. queried csaQTLs as well as pQTLs and cmQTLs, resulting in 731 molecular traits. Of the 122 significant association signals, 51 colocalized with genetic risk for at least one autoimmune disease. For example, rs72928038 near *BACH2* was associated with increased CD28 expression on CD45RA⁺ cells (predominantly naïve T cells) and was in high LD with risk alleles for autoimmune thyroiditis, vitiligo, multiple sclerosis (MS) and T1D. Focusing on T cell subpopulations rather than gene expression helps to reduce the number of molecular traits under investigation, thereby achieving enough statistical power to detect some QTLs in *trans*. In fact, the majority of significant associations reported by Orrù et al. corresponded to SNPs and cell state protein markers from different chromosomes³¹. However, this sort of study design is restricted to pre-specified T cell states and functions, since protein markers and gates need to be defined a priori.

It is now possible to define cell states in the context of high-dimensional single-cell data. For example, RNA sequencing has enabled data-driven strategies to refine our understanding of T cell states. We have found canonical correlation analysis between mRNA and protein to be effective in identifying T cell states that correspond to immunologically relevant surface proteins^{12,32}. Szabo et al.³³ used a variant of non-negative matrix factorization (NMF) to identify seven gene expression modules in T cells (e.g. cytotoxicity, proliferation, and IFN response). In a recent study, Jagadeesh et al.³⁴ identified gene programs through NMF, and found significant heritability enrichment for celiac disease in the putative enhancers of a T cell program characterized by *ETS1*, *CD247*, and *CD28*. Close collaboration between experimental, statistical, and computational scientists will be essential to interpret gene modules suggested by modern scRNAseq datasets in the context of long-established T cell functions. Using these gene modules to represent T cell state may uncover more eQTLs that are context-dependent. Ongoing work in our group seeks to define the genetic regulation of these functional modules, and to what extent they colocalize with autoimmune disease risk.

Genetic regulation of the immunological synapse (HLA-TCR)

One of the most critical T cell functions is antigen recognition through the immunological synapse³⁵. Fragments of antigen presented on HLA molecules are recognized by the TCR expressed on the surface of T cells. On both sides of the immunological synapse, genetic variation has important consequences. On the antigen presentation side, the MHC locus, which encodes HLA proteins, is among the most polymorphic in the human genome. Germline genetic variation in the MHC locus alters the sequence of HLA proteins, which in turn constrains which antigenic peptides are presented to the immune system. On the T cell side, the gene segments that encode the TCR somatically rearrange separately in each T cell. Newly rearranged TCRs are screened in the thymus, where their affinity to peptides presented on the host's HLA molecules must fall within an optimal range that allows recognition of foreign antigen whilst limiting self-reactivity³⁶.

How do HLA variants influence thymic selection? Recent work by our group⁹ cast TCR amino acid usage as a quantitative molecular trait, and used multivariate regression to search for QTLs in the MHC locus.

Due to extreme linkage disequilibrium, the MHC locus is routinely excluded from QTL studies. However, with careful statistical approaches designed to capture the effects of HLA haplotypes³⁷, the MHC locus can be robustly investigated for QTLs³⁸. In our cdr3-QTL study⁹, the strongest associations linked amino acid usage in CDR3, the antigen-recognizing region of the TCR, to HLA-DRB1 amino acid position 13 ("cdr3QTL", Fig. 1b). HLA-DRB1 amino acid position 13 confers a large fraction of genetic risk for both RA, T1D, and other autoimmune diseases. cdr3QTLs raise the intriguing possibility that the thymic selection of autoreactive TCRs plays a critical role in autoimmune disease.

As studies from our group and others identify the effect of disease alleles on TCR amino acids, it becomes important to understand how those TCR amino acid changes alter antigen recognition, the likelihood of T cell activation, and the resultant T cell transcriptional state. Predicting cognate antigen from TCR sequence is a well-recognized problem, currently limited by available training data. Progress in this space will depend critically on whether new TCR de-orphanization technologies³⁹ can expand the available training data to capture the range of antigens recognized by each TCR. Our recent work^{40,41} has demonstrated a role for TCR amino acid features in guiding differentiation of memory and regulatory T cell states. Perhaps, genetic variants in the MHC region that confer risk for autoimmune disease promote the thymic selection of TCRs inclined toward an effector memory T cell state. Consistent with this hypothesis, Orrù et al.³¹ reported a csaQTL for the relative proportion of effector memory CD4⁺ T cells near *HLA-DRB1* (rs9271536).

Future directions

Building a functional interpretation of a GWAS largely consists of two tasks: linking loci to genes, and identifying the critical cell state(s). Identifying disease-critical cell states based on GWAS is a complex challenge, but recent progress has been made with multimodal scATAC-RNAseq⁴². Efforts to link loci to genes have largely focused on genetic variation between individuals, but limited sample sizes restrict the power of these approaches. As a promising alternative approach, we and others have recently leveraged variation between cells in terms of chromatin accessibility (multimodal scATAC-RNAseq) to construct

Table 1 | Useful resources for the study and interpretation of immune cell QTLs

Name	Description	URL
ImmPort ⁵¹	Data sharing portal for published data, encompassing transcriptomic, metabolomic, proteomic, and flow cytometry-based profiling of immune cells	https://www.immport.org/home
dbSNP ⁵²	Broad collection of genetic polymorphisms, documenting the rsID number, genomic position, reference and alternate alleles, and possible clinical significance of each polymorphism	https://www.ncbi.nlm.nih.gov/snp/
GWAS Catalog ⁵³	Database with interactive browser for GWAS summary statistics, aggregating results from over 50,000 GWA studies.	https://www.ebi.ac.uk/gwas/
eQTL Catalogue ⁵⁴	eQTL and sQTL summary statistics from uniform processing of 32 published datasets, including bulk RNAseq and scRNAseq.	https://www.ebi.ac.uk/eql/
GTE ⁵⁵ Portal	Database with interactive browser for the Genotype-Tissue Expression project, which applied bulk RNAseq to 54 non-diseased tissue sites from nearly 1000 individuals. Includes scRNAseq samples from 16 donors for 8 tissues. Open access to eQTL summary statistics and gene count data; access to sequencing data requires application through dbGaP	https://gtportal.org/home/
DICE ¹⁷ Database	Database with interactive browser for the DICE project, which applied bulk RNAseq to 15 FACS-sorted immune cell states (including 11 T cell states) from 91 individuals. Interactive browser allows users to query by gene, rsID, or cell state to collect, visualize and download cell-state-dependent eQTL summary statistics.	https://dice-database.org
IMGT ⁵⁶	The International Immunogenetics Information System. Documents international nomenclature, nucleotide and amino acid sequences for HLA, TCR and Immunoglobulin genes across 38 species	https://www.imgt.org/
QTLbase ⁵⁷	Curated summary statistics from 377 independent QTL studies and 22 types of molecular QTLs, including eQTLs, pQTLs, caQTLs, and sQTLs.	http://mulinlab.org/qtlbase
scQTLbase ⁵⁸	Database with interactive browser for eQTLs identified in single-cell datasets. Includes a web tool that runs R package "coloc" ⁵⁹ to test for colocalization between the existing single-cell QTLs and user-supplied GWAS summary statistics for a trait of interest.	http://bioinfo.szbl.ac.cn/scQTLbase

gene-enhancer maps^{43,44} that can link GWAS variants to causal genes and cell types.

High-throughput screening technologies may drastically expand our understanding of genetic risk variants. Massively parallel reporter assays (MPRA) can simultaneously screen thousands of GWAS variants for regulatory activity, and has recently pinpointed which RA-associated variant in the *BACH2* locus actually reduces expression of *BACH2*, promoting effector T cell differentiation⁴⁵. However, MPRA assays do not account for chromatin inaccessibility, leading to potential false positive results. CRISPR-Cas9 screening approaches, on the other hand, can incorporate chromatin accessibility profiling, and have recently been applied to primary T cells⁴⁶. Future work should extend these genome editing approaches to introduce disease-relevant genetic variants in primary T cells in a high-throughput manner, and characterize the molecular traits that result. This would avoid the complexity of linkage between potential causal variants. Unconstrained by natural selection, these technologies could open the door to molecular characterization of highly pathogenic variants.

Molecular characterizations of disease-associated loci will continue to gain complexity (Table 1). Synthesizing results across different types of QTLs presents a new challenge that will soon become critical. For example, how should we interpret a disease-associated locus that appears to regulate the chromatin accessibility of a gene (caQTL), but not the gene's expression (cis-eQTL)? Bossini-Castillo et al. suggest that we have not yet identified the relevant cell state for the cis-eQTL²¹. Another possibility, however, is that the locus is a cis-eQTL, but its association test just falls short of statistical significance due to limited power. Statistical methods that boost colocalization power by integrating evidence across multiple '-omic' layers are starting to emerge. The Bayesian method OPERA⁴⁷, for example, identified 58% more genes relevant to complex traits after considering six types of QTLs (e.g. caQTLs, mQTLs, pQTLs) in conjunction with eQTLs. We are eager to see this approach extended to T cell functional traits, with appropriate modeling of cell state dependence.

Tissue-resident T cells may play a crucial role in the development of autoimmunity. Studies examining T cells from inflamed tissues have identified disease-relevant T cell states that were not previously appreciated: for example, *IL17*⁺ CD8 T cells in UC⁴⁸ and *GZMK*⁺ T cells in RA⁴⁹. Single-cell profiling of systemic lupus erythematosus (SLE) patient samples recently identified specific regulation of *ORMDL3* at a previously difficult to annotate SLE locus⁵⁰. Hence additional emphasis on collecting tissue samples from patients with autoimmune disease and directly sampling tissue sites of inflammation is critical. Similarly, new studies should query genomic variants that modify the cellular response to disease-relevant stimuli. Identifying multiple types of QTLs (eQTLs, pQTLs, cdr3QTLs) relevant to T cell function in inflamed tissue will bring us closer to understanding how genetic variation shapes risk for autoimmune disease.

References

1. Trynka, G. et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* **45**, 124–130 (2013).
2. Amariuta, T. et al. IMPACT: Genomic annotation of cell-state-specific regulatory elements inferred from the epigenome of bound transcription factors. *Am. J. Hum. Genet.* **104**, 879–895 (2019).
3. Hu, X. et al. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am. J. Hum. Genet.* **89**, 496–506 (2011).
4. Arya, R. et al. A quantitative trait locus (QTL) on chromosome 6q influences birth weight in two independent family studies. *Hum. Mol. Genet.* **15**, 1569–1579 (2006).
5. Wong, E. S. et al. Interplay of cis and trans mechanisms driving transcription factor binding and gene expression evolution. *Nat. Commun.* **8**, 1092 (2017).
6. Pai, A. A. et al. The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. *PLoS Genet.* **8**, e1003000 (2012).
7. Westra, H.-J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
8. Sharon, E. et al. Genetic variation in MHC proteins is associated with T cell receptor expression biases. *Nat. Genet.* **48**, 995–1002 (2016).
9. Ishigaki, K. et al. HLA autoimmune risk alleles restrict the hyper-variable region of T cell receptors. *Nat. Genet.* **54**, 393–402 (2022).
10. Walunas, T. L. et al. CTLA-4 can function as a negative regulator of T cell activation. *Immunity* **1**, 405–413 (1994).
11. Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune. *Cell* **167**, 1398–1414.e24 (2016).
12. Nathan, A. et al. Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature* **606**, 120–128 (2022).
13. Ueda, H. et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* **423**, 506–511 (2003).
14. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
15. Onengut-Gumuscu, S. et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–386 (2015).
16. Chun, S. et al. Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat. Genet.* **49**, 600–605 (2017).
17. Schmiedel, B. J. et al. Impact of genetic polymorphisms on human immune cell gene expression. *Cell* **175**, 1701–1715.e16 (2018).
18. Kim-Hellmuth, S. et al. Cell type-specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020).
19. Ohkura, N. et al. Regulatory T cell-specific epigenomic region variants are a key determinant of susceptibility to common autoimmune diseases. *Immunity* **52**, 1119–1132.e4 (2020).
20. Baecher-Allan, C., Viglietta, V. & Hafler, D. A. Human CD4+CD25+ regulatory T cells. *Semin. Immunol.* **16**, 89–98 (2004).
21. Bossini-Castillo, L. et al. Immune disease variants modulate gene expression in regulatory CD4+ T cells. *Cell Genom.* **2** (2022).
22. van der Wijst, M. G. P. et al. Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat. Genet.* **50**, 493–497 (2018).
23. Li, S. et al. Identification of genetic variants that impact gene co-expression relationships using large-scale single-cell data. *Genome Biol.* **24**, 80 (2023).
24. Soskic, B. et al. Immune disease risk variants regulate gene expression dynamics during CD4+ T cell activation. *Nat. Genet.* **54**, 817–826 (2022).
25. Yazar, S. et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376**, eabf3041 (2022).
26. Zhu, Z. et al. Leveraging molecular quantitative trait loci to comprehend complex diseases/traits from the omics perspective. *Hum. Genet.* **142**, 1543–1560 (2023).
27. Tegtmeier, M. et al. High-dimensional phenotyping to define the genetic basis of cellular morphology. *Nat. Commun.* **15**, 347 (2024).
28. Huang, Q. Q., Ritchie, S. C., Brozynska, M. & Inouye, M. Power, false discovery rate and Winner's Curse in eQTL studies. *Nucleic Acids Res.* **46**, e133–e133 (2018).
29. Zlotnik, A. Perspective: Insights on the nomenclature of cytokines and chemokines. *Front. Immunol.* **11**, 908 (2020).
30. Nath, A. P. et al. Multivariate genome-wide association analysis of a cytokine network reveals variants with widespread immune, haematological, and cardiometabolic pleiotropy. *Am. J. Hum. Genet.* **105**, 1076–1090 (2019).

31. Orrù, V. et al. Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. *Nat. Genet.* **52**, 1036–1045 (2020).
32. Nathan, A. et al. Multimodally profiling memory T cells from a tuberculosis cohort identifies cell state associations with demographics, environment and disease. *Nat. Immunol.* **22**, 781–793 (2021).
33. Szabo, P. A. et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 4706 (2019).
34. Jagadeesh, K. A. et al. Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat. Genet.* **54**, 1479–1492 (2022).
35. Grakoui, A. et al. The immunological synapse: a molecular machine controlling T cell activation. *Science* **285**, 221–227 (1999).
36. Ashby, K. M. & Hogquist, K.A. A guide to thymic selection of T cells. *Nat. Rev. Immunol.* **24**, 103–117 (2023).
37. Sakaue, S. et al. Tutorial: a statistical genetics guide to identifying HLA alleles driving complex disease. *Nat. Protoc.* **18**, 2625–2641 (2023).
38. Kang, J. B. et al. Mapping the dynamic genetic regulatory architecture of HLA genes at single-cell resolution. *Nat. Genet.* **55**, 2255–2268 (2023).
39. Dobson, C. S. et al. Antigen identification and high-throughput interaction mapping by reprogramming viral entry. *Nat. Methods* **19**, 449–460 (2022).
40. Lagattuta, K. A. et al. Repertoire analyses reveal T cell antigen receptor sequence features that influence T cell fate. *Nat. Immunol.* **23**, 446–457 (2022).
41. Lagattuta, K. A. et al. The T cell receptor sequence influences the likelihood of T cell memory formation. *bioRxiv* <https://doi.org/10.1101/2023.07.20.549939> (2023).
42. Gupta, A. et al. Dynamic regulatory elements in single-cell multimodal data implicate key immune cell states enriched for autoimmune disease heritability. *Nat. Genet.* **55**, 2200–2210 (2023).
43. Dey, K. K. et al. SNP-to-gene linking strategies reveal contributions of enhancer-related and candidate master-regulator genes to autoimmune disease. *Cell Genom.* **2**, 100145 (2022).
44. Sakaue, S. et al. Tissue-specific enhancer-gene maps from multimodal single-cell data identify causal disease alleles. *bioRxiv* <https://doi.org/10.1101/2022.10.27.22281574> (2022).
45. Mouri, K. et al. Prioritization of autoimmune disease-associated genetic variants that perturb regulatory element activity in T cells. *Nat. Genet.* **54**, 603–612 (2022).
46. Legut, M. et al. A genome-scale screen for synthetic drivers of T cell proliferation. *Nature* **603**, 728–735 (2022).
47. Wu, Y. et al. Joint analysis of GWAS and multi-omics QTL summary statistics reveals a large fraction of GWAS signals shared with molecular phenotypes. *Cell Genom.* **3**, 100344 (2023).
48. Smillie, C. S. et al. Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell* **178**, 714–730.e22 (2019).
49. Jonsson, A. H. et al. Granzyme K+ CD8 T cells form a core population in inflamed human tissue. *Sci. Transl. Med.* **14**, eabo0686 (2022).
50. Perez, R. K. et al. Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. *Science* **376**, eabf1970 (2022).
51. Bhattacharya, S. et al. ImmPort: disseminating data to the public for the future of immunology. *Immunol. Res.* **58**, 234–239 (2014).
52. Sherry, S. T. et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
53. MacArthur, J. et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
54. Kerimov, N. et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.* **53**, 1290–1299 (2021).
55. GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
56. Lefranc, M.-P. IMGT, the international ImMunoGeneTics information system. *Cold Spring Harb. Protoc.* **2011**, 595–603 (2011).
57. Huang, D. et al. QTLbase2: an enhanced catalog of human quantitative trait loci on extensive molecular phenotypes. *Nucleic Acids Res.* **51**, D1122–D1128 (2023).
58. Ding, R. et al. scQTLbase: an integrated human single-cell eQTL database. *Nucleic Acids Res.* **52**, D1010–D1017 (2024).
59. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-45170-w>.

Correspondence and requests for materials should be addressed to Aparna Nathan or Soumya Raychaudhuri.

Peer review information *Nature Communications* thanks Ferhat Ay, Mulin Li and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024