

Leveraging human genetic variation to uncover the molecular basis of T cell selection and autoimmunity

Caroline Kaiser^{1,2}, MSc, Narutoshi Hibino^{3,4}, MD, Emily Pena⁴, PA-C, Andrew Koh², PhD

¹Department of Human Genetics; ²Department of Pathology; ³Department of Surgery, University of Chicago; ⁴Pediatric Cardiovascular Surgery, Advocate Children's Hospital

Background

How does the immune system distinguish self vs pathogens ?

To respond to a myriad of ever-changing pathogens, the thymus generates T cells each equipped with a unique receptor at their surface to recognize any virus, bacterium, or parasite. To build this comprehensive repertoire, gene segments encoding T cell receptors undergo random rearrangement, yielding specificities against the body's own constituents. To safeguard against harmful auto-reactive T cells, recognition of 'self' triggers cell death for developing T cells in the thymus (Fig1A). Failure in this process leads to the escape of self-reactive T cells that ultimately results in autoimmune diseases (Fig1B).

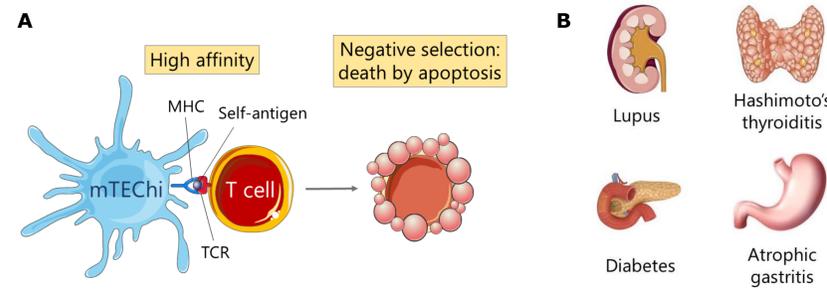


Figure 1. Negative selection and autoimmunity: (A) Presentation of self-antigens in the thymus and recognition by a T cell receptor (TCR). High affinity between the TCR and the self-antigen leads to the death of the T cell. (B) Examples of autoimmune diseases resulting from the escape of auto-reactive T cells.

To 'mirror' all of self in the thymus and promote the selection of a T cell repertoire that can distinguish self vs non-self, medullary thymic epithelial cells (mTECs) possess the remarkable capacity to express virtually any gene, including genes whose expression is normally restricted to highly specialized tissues, e.g insulin (Fig2).

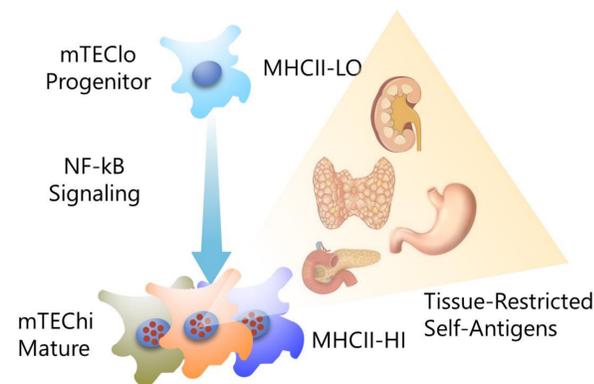


Figure 2. Differentiation of mTECs: mTEC progenitor (expressing low levels of MHCII) differentiate upon NF-kB signaling to mature mTEChi (high levels of MHCII) capable of expressing tissue-restricted self-antigens.

Despite its clinical importance, the precise mechanism driving this tissue-restricted gene (TRG) expression remains unknown. Here, we leverage the natural genetic variation in human populations to uncover the molecular basis of TRG expression and its contribution to autoimmunity

Assembly of a multi-omic human thymi cohort

Genetic sequence

Mutations in the human genome are often single nucleotide changes called SNPs (Fig3A). Whole genome sequencing, or SNP arrays (Fig3B) can estimate the genetic variation among different individuals.

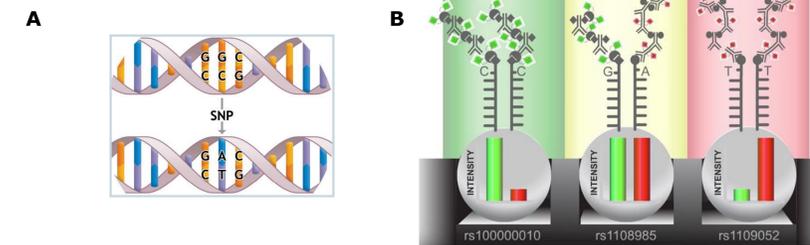


Figure 3. Genotyping of human variants: (A) Single Nucleotide Polymorphism (G vs A) in two different DNA samples. (B) Hybridization of human DNA on SNP arrays indicate the different SNPs present in the genome (i.e. the genotype). The Illumina Multi-Ethnic array covers 1.8 million SNPs in the genome.

Isolation of thymic epithelial cells

mTECs are sorted into four different populations based on MHCII expression (Fig4A), and processed for downstream applications such as RT-qPCR or RNA-seq (Fig4B).

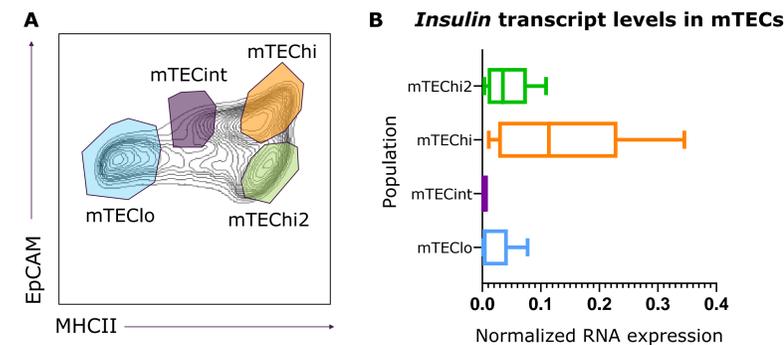


Figure 4. Isolation of different mTEC populations: (A) Contour plot showing the gating strategy for human mTECs (representative ID #39). Live CD45- EpCAM+ Ly51- are further separated according to their MHCII expression: mTEClo, mTECint, mTEChi, and an additional uncharacterized mTEChi2 population. (B) RT-qPCR profiles for Insulin transcripts averaged across 2-5 individuals. RNA levels are normalized to housekeeping Hprt expression.

Accessibility profiles

The accessibility of DNA is an important indicator of gene activity and binding of transcription factors. ATAC-seq enables the mapping of accessible DNA (Fig5).

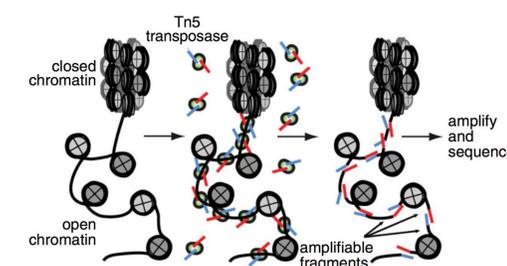


Figure 5. ATAC-seq pipeline: Tn5 transposase enzyme can cut the genome and inserts sequencing barcodes at accessible loci. Sequencing then reveals 'open' regions of chromatin.

Identify genetic loci impacting TRG expression

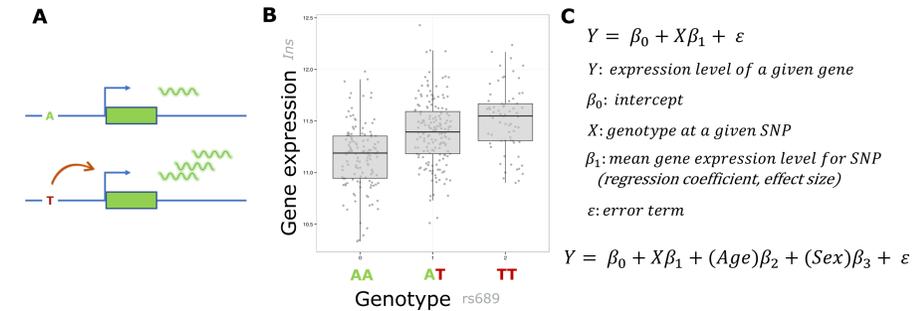
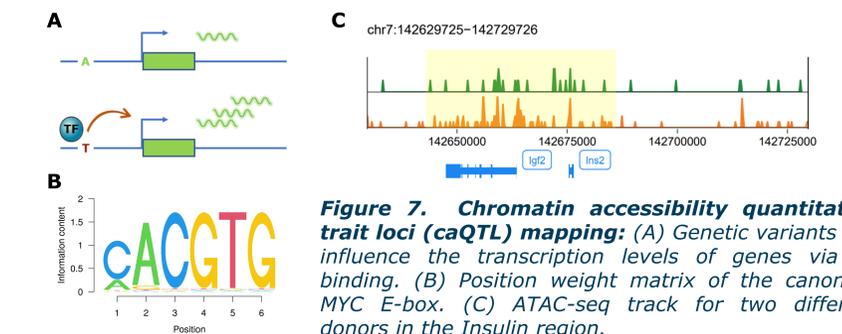


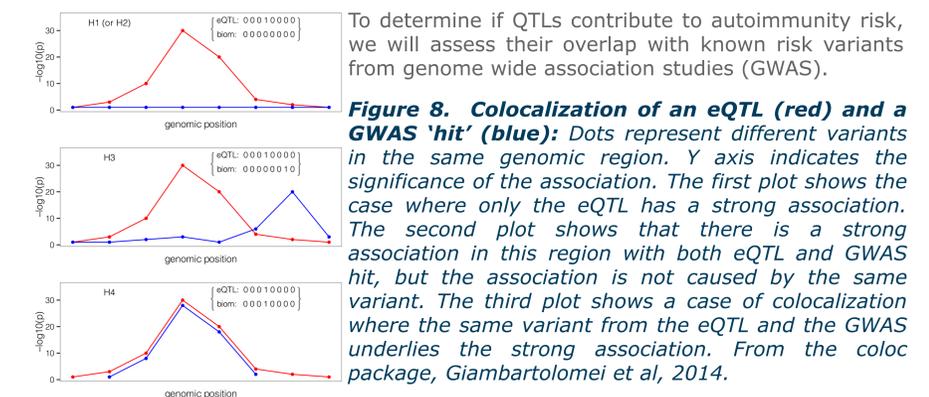
Figure 6. Expression quantitative trait loci (eQTL) mapping: (A) Genetic variants can influence the transcription levels of genes. (B) Plot showing the effect of genotype on gene expression (rs689 is an eQTL for the Insulin gene). (C) Linear model used to compute the effect (β_1) seen in (B). The final formula accounts for age and sex confounders from the donors.

Characterize accessibility and binding of transcription factors



The reduced accessibility and transcription at the *Ins* locus can be linked to a disruption of the TF binding motif (T to A mutation at position 5).

Determine the contribution to autoimmune diseases



To determine if QTLs contribute to autoimmunity risk, we will assess their overlap with known risk variants from genome wide association studies (GWAS).

Figure 8. Colocalization of an eQTL (red) and a GWAS 'hit' (blue): Dots represent different variants in the same genomic region. Y axis indicates the significance of the association. The first plot shows the case where only the eQTL has a strong association. The second plot shows that there is a strong association in this region with both eQTL and GWAS hit, but the association is not caused by the same variant. The third plot shows a case of colocalization where the same variant from the eQTL and the GWAS underlies the strong association. From the coloc package, Giambartolomei et al, 2014.