



RBFOX splicing factors contribute to a broad but selective recapitulation of peripheral tissue splicing patterns in the thymus

Kathrin Jansen, Noriko Shikama-Dorn, Moustafa Attar, et al.

Genome Res. 2021 31: 2022-2034 originally published online October 14, 2021

Access the most recent version at doi:[10.1101/gr.275245.121](https://doi.org/10.1101/gr.275245.121)

References This article cites 76 articles, 18 of which can be accessed free at:
<http://genome.cshlp.org/content/31/11/2022.full.html#ref-list-1>

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

RBFOX splicing factors contribute to a broad but selective recapitulation of peripheral tissue splicing patterns in the thymus

Kathrin Jansen,^{1,2} Noriko Shikama-Dorn,³ Moustafa Attar,^{1,4} Stefano Maio,² Maria Lopopolo,⁴ David Buck,⁴ Georg A. Holländer,^{2,3,5,6} and Stephen N. Sansom^{1,6}

¹The Kennedy Institute of Rheumatology, University of Oxford, Oxford OX3 7FY, United Kingdom; ²Department of Paediatrics and the Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, United Kingdom; ³The University Children's Hospital of Basel and the Department of Biomedicine, University of Basel, 4056 Basel, Switzerland; ⁴Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom; ⁵Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland

Thymic epithelial cells (TEC) control the selection of a T cell repertoire reactive to pathogens but tolerant of self. This process is known to involve the promiscuous expression of virtually the entire protein-coding gene repertoire, but the extent to which TEC recapitulate peripheral isoforms, and the mechanisms by which they do so, remain largely unknown. We performed the first assembly-based transcriptomic census of transcript structures and splicing factor (SF) expression in mouse medullary TEC (mTEC) and 21 peripheral tissues. Mature mTEC expressed 60.1% of all protein-coding transcripts, more than was detected in any of the peripheral tissues. However, for genes with tissue-restricted expression, mTEC produced fewer isoforms than did the relevant peripheral tissues. Analysis of exon inclusion revealed an absence of brain-specific microexons in mTEC. We did not find unusual numbers of novel transcripts in TEC, and we show that *Aire*, the facilitator of promiscuous gene expression, promotes the generation of long “classical” transcripts (with 5' and 3' UTRs) but has only a limited impact on alternative splicing in mTEC. Comprehensive assessment of SF expression in mTEC identified a small set of nonpromiscuously expressed SF genes, among which we confirmed RBFOX to be present with AIRE in mTEC nuclei. Using a conditional loss-of-function approach, we show that *Rbfox2* promotes mTEC development and regulates the alternative splicing of promiscuously expressed genes. These data indicate that TEC recommission a small number of peripheral SFs, including members of the RBFOX family, to generate a broad but selective representation of the peripheral splice isoform repertoire.

[Supplemental material is available for this article.]

T cells are essential for the generation and resolution of an adaptive immune response because they are uniquely able to distinguish between benign self and harmful non-self-antigens. Because T cell antigen receptor (TCR) specificity is generated pseudorandomly, the functional utility and self-reactivity of TCRs must be screened during T cell development in the thymus (Klein et al. 2014). This process is critically dependent on different thymic stromal cells such as thymic epithelial cells (TECs) (Abramson and Anderson 2017). In an initial round of positive selection, cortical TEC (cTEC) positively select developing T cells (thymocytes) that express TCRs of sufficient affinity for self-MHC (Klein et al. 2014). Subsequently, both cTEC and medullary TEC (mTEC) deplete thymocytes that display TCRs with high affinity for self-antigens, a process designated thymocyte negative selection. In addition, mTEC divert a subset of self-reactive T cells to a natural T regulatory cell (nTreg) fate (Stritesky et al. 2012).

To achieve thymocyte negative selection, TEC express and present a molecular mirror of an individual's self-antigens to maturing thymocytes. This process of promiscuous gene expression

(PGE) is essential for the avoidance of autoimmunity and involves transcription of ~89% of protein-coding genes by the TEC population (Sansom et al. 2014; Abramson and Anderson 2017). The mechanisms by which TEC defy developmental and tissue-specific transcriptional controls to achieve PGE are only incompletely deciphered (Abramson and Goldfarb 2016), although the autoimmune regulator (AIRE) (Sansom et al. 2014) and the transcription factor FEZF2 (Takaba et al. 2015) together with the chromatin remodeler CHD4 (Tomofuji et al. 2020) are known to enable PGE. Comprehensiveness of self-representation by TEC cannot, however, be measured by gene number alone because self-peptide diversity is elaborated by alternative mRNA splicing (Anderson et al. 2000; Klein et al. 2000; Wang and Burge 2008), expression of “untranslated” regions (Starck and Shastri 2016), RNA-editing (Danan-Gotthold et al. 2016), proteasome-mediated splicing (Granados et al. 2015), and posttranslational modifications (Raposo et al. 2018). In particular, alternative mRNA splicing greatly increases the complexity of the mammalian proteome: there are approximately three times more annotated protein-coding transcripts than genes in mice and humans (Zerbino et al. 2018).

The generation of alternative splice isoforms is tightly regulated during development and controlled by temporal and

These authors contributed equally to this work.

Corresponding author: stephen.sansom@kennedy.ox.ac.uk, stephen.sansom@cantab.net

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.275245.121>. Freely available online through the *Genome Research* Open Access option.

© 2021 Jansen et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

context-specific expression of SFs. The extent to which TEC reproduce peripheral splice isoforms is unclear, but the transcriptome of TEC has been found to be unusually complex both in terms of isoform number (Keane et al. 2015) and splice junction representation (Danan-Gotthold et al. 2016). These surveys, which compared TEC with limited numbers of peripheral tissues, reported that TEC reproduce a fifth of tissue-restricted isoforms (Keane et al. 2015) or only 20%–60% of tissue-restricted splice junctions (Danan-Gotthold et al. 2016). Even less is known about the mechanisms in TECs that control alternative splicing. Because of its direct interactions with several splicing-related factors (Abramson et al. 2010), it was thought that AIRE might be involved in this process (Keane et al. 2015), but limited evidence supporting this idea suggests only a minor role (Danan-Gotthold et al. 2016). Instead, it is appealing to postulate that mTEC may reuse peripheral SFs to create tissue-specific splice variants. In one plausible model, mTEC might constitutively express a specific subset of peripheral SFs to achieve robust coverage of a limited repertoire of peripheral splice variants. In support of this possibility, an initial analysis identified seven RNA-binding factors with constitutive expression in murine mTEC (St-Pierre et al. 2015). In a second alternative model, mTEC might use promiscuously expressed SFs to achieve a broader stochastic coverage of peripheral isoforms. A possible side effect of such a promiscuous splicing strategy would be the generation of spurious novel isoforms if nascent transcripts interact with SFs in TEC to which they are not exposed in the periphery.

Knowledge of splice isoform representation in the thymus is important for understanding central tolerance because it is known that the absence of tissue-specific splice isoforms can result in the development of pathogenic T cells able to incite autoimmunity (Anderson et al. 2000; Klein et al. 2000). We therefore set out to perform the first comprehensive and unbiased census of isoform representation and SF usage in TEC and peripheral tissues.

Results

TEC express fewer splice isoforms per gene than peripheral tissues

To compare the splicing landscape of TEC with peripheral (i.e., non-thymic) tissues, we performed deep, stranded RNA sequencing of immature and mature mTEC and integrated the data with that from similar sequencing of 21 separate tissues from the mouse ENCODE Project (Pervouchine et al. 2015). To ensure that our investigations were not biased by differential representation of tissue or TEC transcripts in the available reference annotation, we began our analysis by constructing a mouse TEC and Tissue (mT&T) transcriptome assembly. To do so we built a set of reference-guided individual assemblies from each TEC population and peripheral tissue using matched high-depth sequence read pools (200 million reads/pool). To avoid inclusion of assembly artifacts, only transcripts that could be reproducibly detected in biologically replicate sample pools using a refined npIDR approach (Supplemental Fig. 1) were merged into the final mT&T assembly (Fig. 1A). The mT&T assembly included 58,990 novel transcripts in 14,865 protein-coding genes. In agreement with previous reports (Keane et al. 2015; St-Pierre et al. 2015; Danan-Gotthold et al. 2016), we found that mature mTEC expressed a greater number of transcripts from protein-coding genes than any of the surveyed peripheral tissues (60.1%) (Fig. 1B). However, modeling of the relationship between the number of detected genes and transcripts revealed that mature mTEC produced significantly fewer transcripts per

gene than was predicted by the relationship between gene and transcript number in the peripheral tissues (Fig. 1C). We confirmed this finding by repeating the analyses using reference Ensembl annotations, biologically replicate sample pools, and independent data sets (Supplemental Fig. 2A–G).

We next sought to determine the extent to which the splicing of promiscuously expressed genes in TEC mirrors the patterns found in the normal tissue context. To do so we first defined three categories of protein-coding genes: (1) $n = 3889$ Aire-regulated “tissue-restricted antigen” (TRA) genes, (2) $n = 5266$ non-Aire-regulated TRA genes, and (3) $n = 12,885$ non-TRA genes. Aire-regulated genes were defined as those found to be significantly down-regulated in Aire-knockout mTEC ($fc > 2$, BH adjusted $P < 0.05$), whereas TRA genes were identified using the *tau* tissue-specificity metric (Supplemental Methods; Supplemental Fig. 3; Supplemental Table 2; Kryuchkova-Mostacci and Robinson-Rechavi 2017). We reasoned that peripheral tissues are more likely to engage in alternative splicing of the TRA genes that contribute to their specialized functions. We therefore also defined a collection of individual-tissue TRA gene subsets, which we term “iTRA,” by assigning each TRA gene to the tissue in which it was most highly expressed. Comparison of the splicing of iTRA subsets in their cognate peripheral tissues with thymic TRA gene splicing revealed that TEC expressed a significantly smaller fraction of TRA gene isoforms than can be found in the relevant peripheral tissues (Fig. 2A; Supplemental Table 13). These observations did not appear to be a consequence of a lower sequencing coverage of promiscuously expressed genes in TEC (Supplemental Fig. 4A–C), because significantly fewer isoforms were detectable in mTEC regardless of gene expression level (Supplemental Fig. 4D–J). Examination of unique splice junction numbers confirmed that genes more extensively spliced in individual peripheral tissues than in mature mTEC were strongly and significantly overrepresented for the cognate iTRA subsets (19 of 20 tissues, Fisher’s exact test) (Fig. 2B; Supplemental Fig. 5A). For example, in the cerebellum we found $n = 44$ cerebellum iTRA genes with higher splice junction counts and only $n = 1$ such gene with lower numbers of splice junctions ($P = 1.26 \times 10^{-27}$, odds ratio = 171.6). In this tissue we noted that *Plp1*, for which incomplete thymic splicing can contribute to experimental autoimmune encephalomyelitis (EAE), had significantly higher numbers of splice junctions in the periphery (Anderson et al. 2000; Klein et al. 2000). Additionally, we noted that several homologs of genes linked to human autoimmune disease had significantly higher numbers of splice junctions in the peripheral tissues than was found in mature mTEC. These included myelin basic protein, *Mbp* ($n = 112$ junctions in cerebellum, $n = 17$ junctions in mature mTEC, $FDR = 4.066 \times 10^{-15}$), which is a target of autoantibodies in multiple sclerosis (Berger et al. 2003), and *Cyp21a1* ($n = 140$ junctions in adrenal tissue, $n = 5$ junctions in mature mTEC, $FDR = 1.76 \times 10^{-21}$) the homolog of the human *CYP21A2* gene that encodes cytochrome P450 family 21 subfamily A member 2 (also known as P450c21B), a major target of autoantibodies in Addison’s disease (Winqvist et al. 1992). Because previous genomic investigations suggested more promiscuous and extensive programs of splicing in mTEC (Keane et al. 2015; Danan-Gotthold et al. 2016), we sought to verify our observations using the data from those studies. We began by reproducing the original observations of unusually high overall transcriptome complexity in mTEC (Supplemental Fig. 5B,C). However, analysis of these independent data sets with the iTRA approach confirmed that mTEC engage in less alternative splicing of iTRA than can be found in the cognate peripheral tissues (Fig. 2C,D).

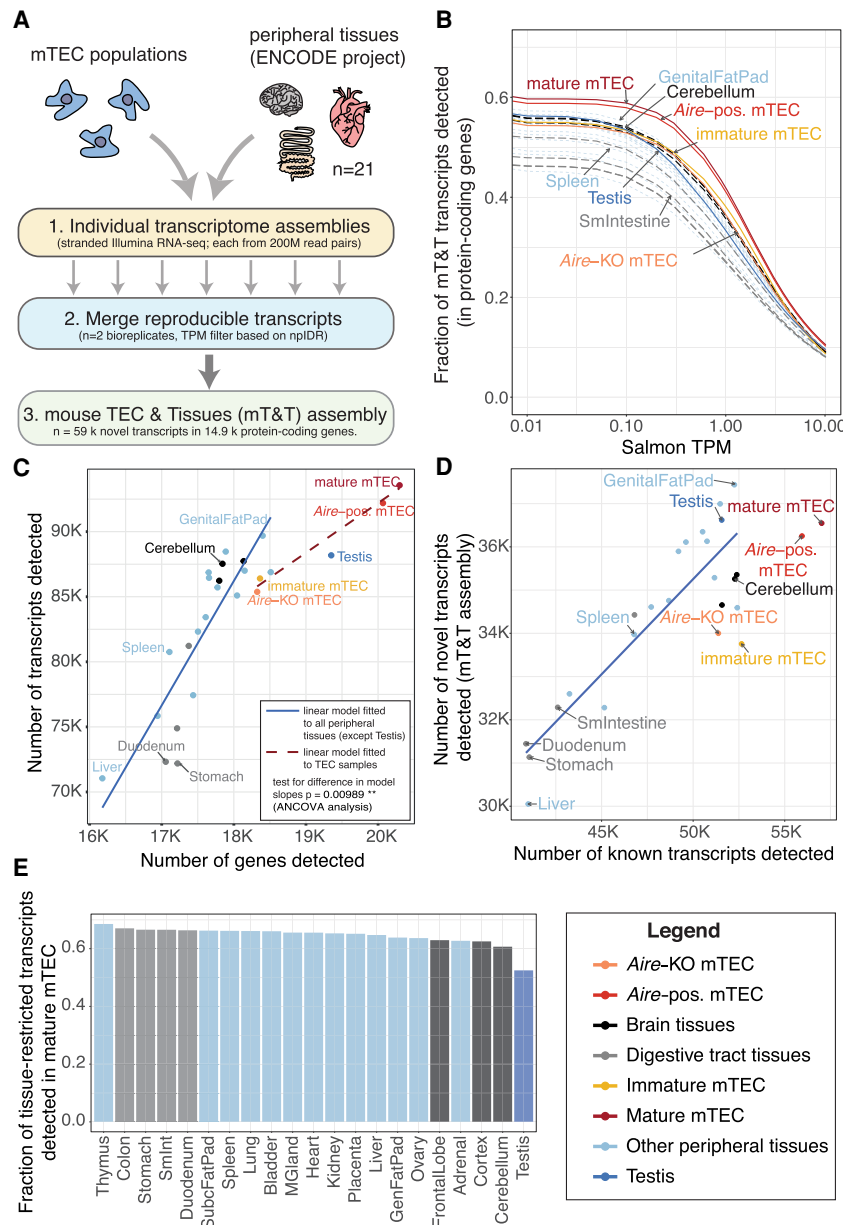


Figure 1. Comparative analysis of transcript expression in mTEC and peripheral tissues. (A) Generation of a common mouse mTEC and peripheral Tissues (mT&T) transcriptome assembly. (B) Fractions of transcripts from protein-coding genes detected in peripheral tissues and mTEC populations across a range of TPM thresholds. The scatter plots (C,D) show the relationships between the number of genes and transcripts (C) and the number of known versus novel transcripts (D) detected in peripheral tissues and mTEC populations. The linear models shown in C and D were fitted to all samples except TEC and testis (solid trend lines) or to the TEC samples (C, dashed trend line). There was a significant difference in the trend line slopes in C ($P=0.00989$, ANCOVA analysis). (E) Fractions of sets of tissue-restricted transcripts ($\tau \geq 0.9$) from peripheral tissues detected in mature mTEC (see also Supplemental Fig. 2I). Analyses shown were restricted to protein-coding genes and performed using a single high-depth sample per tissue. Similar results were obtained using lower-depth biologically replicate sample pools ($n=2$) (Supplemental Fig. 2C–E,H).

As previously reported (Keane et al. 2015; Danan-Gotthold et al. 2016), we found that the mature mTEC population coexpressed transcript isoforms that normally arise in distinct anatomical locations (Supplemental Fig. 6A–C). In addition, we found evidence that such transcripts can be produced by the same cell: single-cell RNA sequencing data revealed that the thyroid and ner-

vous-system specific transcript isoforms of *Calca* were often produced together in individual mTEC (Supplemental Fig. 6D). We also assessed the representation of sets of tissue-restricted isoforms from peripheral tissues in mTEC. Transcripts with testis-restricted expression were most markedly underrepresented in mTEC (52.5% detected), followed by those from the brain (60.7–62.9%), adrenal (62.7%), and ovary (63.6%) (Fig. 1E, with results from biological replicate sample pools shown in Supplemental Fig. 2H).

Finally, although mature mTEC expressed a higher number of known transcripts ($n=57,019$) than was observed in peripheral tissues, we found that they produced a relatively low number of novel transcripts ($n=36,547$) (Fig. 1D). Together our results show that although TEC produce an atypically high absolute number of transcripts, they produce significantly fewer isoforms of TRA genes than can be found in the relevant peripheral tissues.

Genes harboring novel transcripts in mTEC are associated with T cell selection

Alternative splicing events are often associated with the evolution and modification of protein function (Kelemen et al. 2013). Because the transcriptome of TEC is relatively understudied, we reasoned that novel splicing events specific to TEC might be associated with their specialized functions for T cell selection. Using long-read Oxford Nanopore Technologies (ONT) sequencing (Supplemental Fig. 7), we validated the existence 64.3% of the mT&T novel transcript structures that were expressed at moderate or high levels in mature mTEC (>10 counts) (Fig. 3A). Next, we identified novel transcripts that showed TEC or tissue-specific expression. For this analysis we chose a single tissue to represent each group of related tissues (Supplemental Fig. 3A). We found that mature mTEC expressed a higher number of unique novel transcripts than did brain tissue (see also Supplemental Fig. 8), but substantially fewer than were found in the testis (Fig. 3B; Supplemental Table

3). The majority (85%) of the 1572 novel transcripts uniquely expressed in mature mTEC were found for loci that do not require *Aire* for their expression (Fig. 3C).

The 1167 protein-coding genes that harbored TEC-specific novel transcripts displayed significant enrichments for Gene Ontology (GO) biological processes such as “regulation of

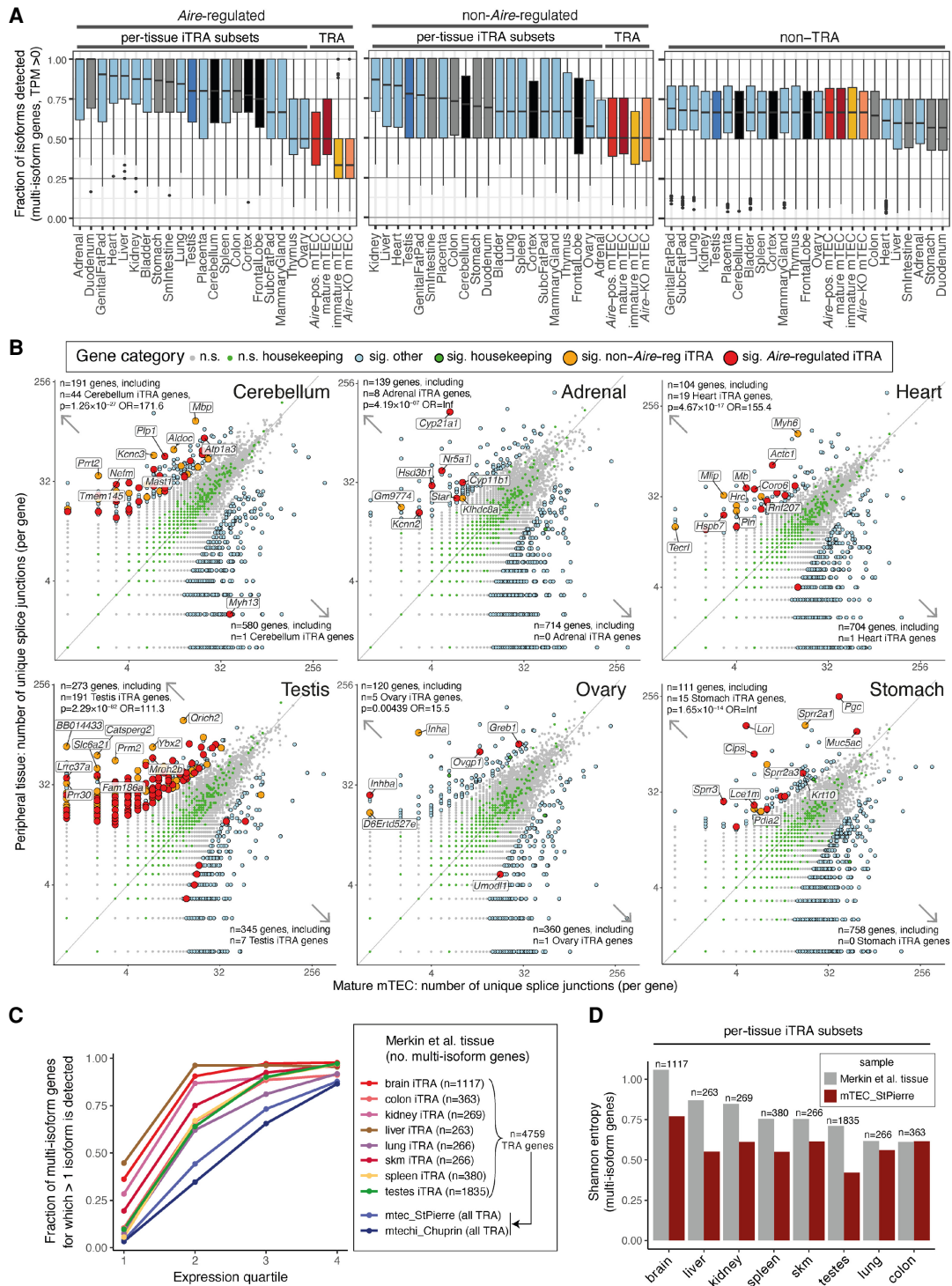


Figure 2. mTEC express fewer isoforms of TRA genes than peripheral tissues. (A) The fraction of isoforms detected (TPM > 0) per multi-isoform protein-coding gene in mTEC and ENCODE peripheral tissues. The box plots show the fractions for *Aire*-regulated TRA genes (*left*), non-*Aire* TRA genes (*middle*), and non-TRA genes (*right*, colors as in Fig. 1) (Supplemental Fig. 3D). (B) The numbers of splice junctions found in protein-coding genes (points) in mature mTEC (this study; x-axis) versus peripheral tissues (ENCODE Project; y-axis). Six selected peripheral tissues are shown, with the remaining 14 displayed in Supplemental Figure 5A. Significant (sig.) differences in junction number were identified using edgeR (BH adjusted $P < 0.05$, $|fc| > 2$). P -values and odds ratios (OR) from Fisher's exact tests for enrichment of iTRA genes among the genes with significantly higher junction counts in the peripheral tissues are reported (*top left*). The top 10 of each tissue's iTRA genes with significant differences in junction counts are labeled (as ranked by edgeR P -value). (C, D) Validation with independent peripheral tissue (Merkin et al. 2012) and TEC data sets (St-Pierre et al. 2013; Chuprin et al. 2015). (C) The fraction (y-axis) of alternatively spliced multi-isoform genes (>1 splice isoform detected) is shown for individual sets of tissue iTRA genes by expression quartile (x-axis). (D) The bar plots show the mean Shannon entropy of splice isoform expression for sets of multi-isoform tissue iTRA genes (x-axis). Alternate versions of C and D made using all genes are shown in Supplemental Figure 5B and 5C. For A and C, the full sets of TRA genes were quantitated in the TEC populations.

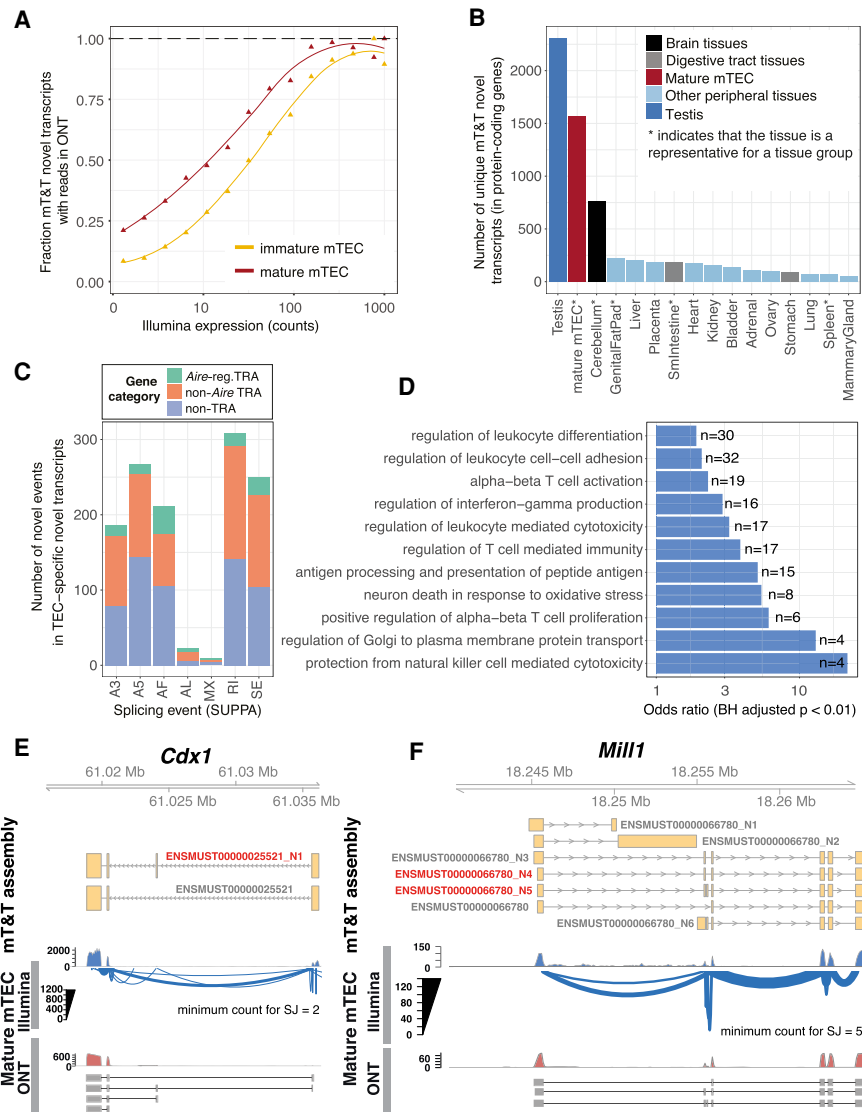


Figure 3. Identification and characterization of TEC-specific novel transcripts. (A) Validation of novel mT&T transcripts using ONT RNA sequencing. The fraction of novel transcripts supported by ONT reads are shown for mature mTEC (red) and immature mTEC (yellow). (B) Number of novel transcripts “uniquely” detected in mature mTEC and representative peripheral tissue samples (Supplemental Methods; Supplemental Figs. 3A, 8). (C) TEC-specific novel splicing events by event type and promiscuous expression status. Individual events may be counted in multiple categories. (SE) skipped exon; (RI) retained intron; (MX) mutually exclusive exon; (A3/A5) alternative 3’/5’ splice site; (AF/AL) alternative first/last exon. (D) Selected GO biological processes significantly overrepresented in the set of genes ($n = 1167$) from which the mTEC-specific novel transcripts were derived (one-sided Fisher’s exact tests; BH adjusted $P < 0.01$). (E,F) *Cdx1* and *Mill1* are displayed as examples to show novel TEC-specific transcripts (red). Existence of the novel transcripts in mature mTEC was supported by both Illumina (Sashimi plots) and long-read ONT (selected reads) data. Novel transcripts are indicated by the “_N” suffix.

leukocyte differentiation,” “regulation of T cell-mediated immunity,” and “antigen processing and presentation of peptide antigen” (Benjamini-Hochberg [BH] adjusted $P < 0.01$, one-sided Fisher’s exact tests) (Fig. 3D; Supplemental Table 4), suggesting that they are likely to encode for nonpromiscuously expressed factors that have a functional role in T cell selection in TEC. Novel transcript structures were detected in genes of known importance for TEC function, including *Foxn1* (Žuklys et al. 2016) and *Aire* (Supplemental Table 3; Abramson and Goldfarb 2016). Genes

that produced transcripts harboring novel exons or exon skipping events in TEC included *Cdx1*, a transcription factor linked with mTEC maturation (Fig. 3E; Handel et al. 2018); *Cd80*, a receptor important for interaction with T cells via CD28 and CTLA4; the MHC class II gene *H2-Aa*; *Mill1*, an MHC class I-like molecule that is known to be expressed on a subpopulation of TEC (Fig. 3F; Kajikawa et al. 2006); as well as *Skint2* gene family members, including *Skint2*, which has been reported to be a novel negative T cell regulator (Yang et al. 2007).

Together these observations suggest that the novel transcript structures detected in thymic epithelial cells may have relevance for TEC function.

Aire contributes to alternative splicing and promotes expression of long transcripts

To clarify the role of *Aire* in alternative splicing in TEC, we also generated deep and stranded Illumina sequencing data from mature *Aire*-positive and mature *Aire*-knockout mTEC. We discovered 492 significant (5% FDR, rMATS) *Aire*-regulated alternative splicing events in transcripts from 459 protein-coding genes (Fig. 4A; Supplemental Tables 5, 14). There was, however, a much larger difference in transcript splicing between immature and mature wild-type mTEC ($n = 2236$ events in $n = 1967$ protein-coding genes, 5% FDR, rMATS analysis) (Fig. 4A; Supplemental Table 5). To confirm that these differences were not linked to the much higher expression of *Aire* in mature mTEC (Supplemental Fig. 9B), we performed an analysis of the differences in splicing events between immature wild-type and mature *Aire*-knockout mTEC. As expected, we found a similar number of differences to that identified for the wild-type comparison, demonstrating that the extensive splicing differences between immature and mature mTEC are controlled by *Aire*-independent factors (Fig. 4A). The majority of the *Aire*-controlled and TEC maturity-related alternative splicing events were found in transcripts encoding non-TRA genes (Fig. 4B). In addition, a large number of retained introns were found in the transcripts of immature mTEC ($n = 760$) (Fig. 4A). Coordinated changes in intron retention during cellular differentiation are not unusual, and transcripts harboring retained introns can arise from genes that have specialized cellular functions (Jacob and Smith 2017). Overall, however, we found that immature and mature TEC had fewer retained introns than did peripheral tissues (Supplemental Fig. 9A).

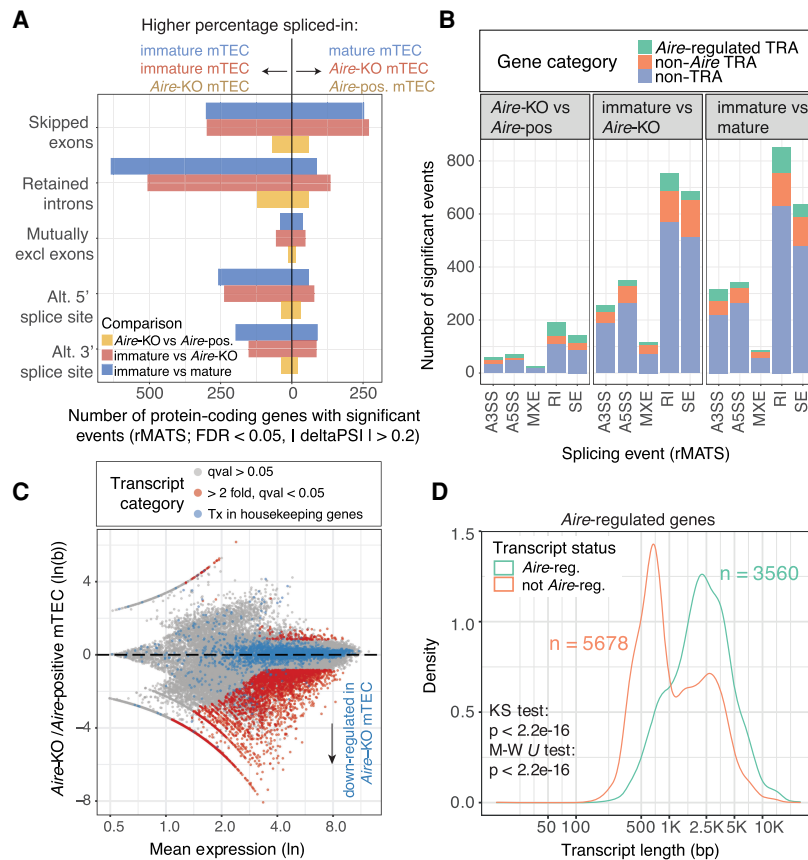


Figure 4. *Aire* promotes the generation of long transcripts in TEC. (A) The numbers of protein-coding genes (x-axis) in which significant differential splicing events (y-axis) were detected; comparisons of immature versus mature mTEC (blue), immature mTEC versus *Aire*-knockout mTEC (red) and *Aire*-knockout versus *Aire*-positive mature mTEC (yellow) ($n = 2$ replicates per sample). (B) Breakdown of identified splicing events by event type and promiscuous expression status. (SE) skipped exon; (RI) retained intron; (MXE) mutually exclusive exon; (A3SS/A5SS) alternative 3'/5' splice site. (C) MA plot of differential transcript expression in *Aire*-knockout compared to *Aire*-positive mature mTEC. Transcripts regulated by *Aire* are shown in red (sleuth, Wald test, $Q_{val} < 0.05$, $f_c \geq 2$, $n = 2$ replicates per sample). Transcripts from housekeeping genes are shown in blue. (D) The length distributions of *Aire*-regulated and non-*Aire*-regulated transcripts in *Aire*-regulated genes (analysis limited to genes that contained at least one significantly *Aire*-regulated transcript as defined in C).

AIRE is known to promote expression of distal exons (Meredith et al. 2015) and the release of stalled polymerases (Giraud et al. 2012). We therefore investigated whether *Aire* might favor the production of long transcripts. Differential transcript usage analysis (Fig. 4C) revealed that *Aire* positively regulated the generation of a large number of transcripts ($n = 4027$, BH adjusted $P < 0.05$, fold change ≥ 2) (Supplemental Table 6). We found that *Aire*-regulated transcripts arising from *Aire*-regulated genes had a significantly different length distribution (Kolmogorov-Smirnov test, $P = 2.2 \times 10^{-16}$) and location (Wilcoxon test, $P = 2.2 \times 10^{-16}$) being on average >1 kb longer than their non-*Aire*-regulated counterparts (Fig. 4D). Analysis of the structures of these transcripts showed that although most (85.3%) of the *Aire*-regulated transcripts comprised “classical” transcripts with both 5' and 3' UTRs, the majority (68.1%) of non-*Aire*-regulated transcripts had nonclassical or unannotated structures (Supplemental Fig. 9C–F). Overall, we found that *Aire*-regulated transcripts had significantly shorter 5' UTRs (–15.5 bp; –7.6%), longer CDSs (+460.8 bp; +38.6%), and longer 3' UTRs (+235.4 bp; 28.4%) (Supplemental Fig. 9G). These findings show that *Aire* plays a major role in pro-

moting the production of long “classical” transcripts (with 5' and 3' UTRs) from its target loci while only having a limited impact on the generation of alternative splicing events. Hence, splicing factors other than *Aire* must be primarily responsible for the alternative splicing of transcripts in mature mTEC.

Medullary TEC express a distinct set of SFs that includes *Rbfox1*

The ability of mTEC to express a large number of peripheral isoforms suggested that they might reuse tissue-specific SFs. After first noting that expression of core spliceosome genes was intact in mTEC (Supplemental Fig. 10A), we systematically identified a set of 146 splicing-related genes that showed tissue-restricted expression (Supplemental Methods; Supplemental Figs. 10B, 11A; Supplemental Table 7). Based on literature searches we further narrowed our focus to $n = 24$ tissue-restricted splicing factors (TRSFs) with known roles in controlling alternative splicing (Fig. 5A; Supplemental Table 8). Of these we noted the frequent (detected in >20% of single mTEC) and nonpromiscuous expression of *Rbfox1*, *Rbm20*, and *Msi1* in mature mTEC. These factors also showed little, if any, expression in skin epithelia (Fig. 5A), suggesting that they may be relevant to the specialized function of mTEC. RBFOX1 and RBM20 control alternative splicing in the brain and myocardium (Guo et al. 2012; Conboy 2017), whereas MS11 is an RNA-binding protein implicated in regulating splicing in photoreceptors (Murphy et al. 2016). Meanwhile, we noted that the majority of the 24 TRSFs

—including those restricted in their expression to the brain (such as *Nova1*, *Nova2*, *Elav2*, *Elav3*, *Elav4*, and *Rbfox3*), testis (*Brdt*), and heart (*Rbm24*)—were expressed at low frequency or showed only promiscuous expression in mature mTEC (Fig. 5A).

In line with the absence of neural and testis-specific SFs, sets of coding exons frequently included in transcripts detected in the brain and testis were found to be excluded from mTEC mRNA ($PSI < 0.1$) (Fig. 5B). In particular, we noted that brain-specific microexons (≤ 30 bp) were rarely included in mTEC transcripts (Fig. 5C), with the PSI distributions of these microexons in TEC being very similar to that found in other non-neuronal tissues (Supplemental Fig. 11B). This observation may be explained by the low expression of *Srrm4* in mTEC (Fig. 5A), because this factor is known to be responsible for promoting the inclusion of neuronal microexons (Quesnel-Vallières et al. 2015, 2016). In summary we found that mTEC constitutively express only a small number of peripheral SFs. Together with the clear absence of a set of neuronal microexons, these results suggest that mTEC may have evolved to selectively represent specific facets of the peripheral tissue isoform repertoire.

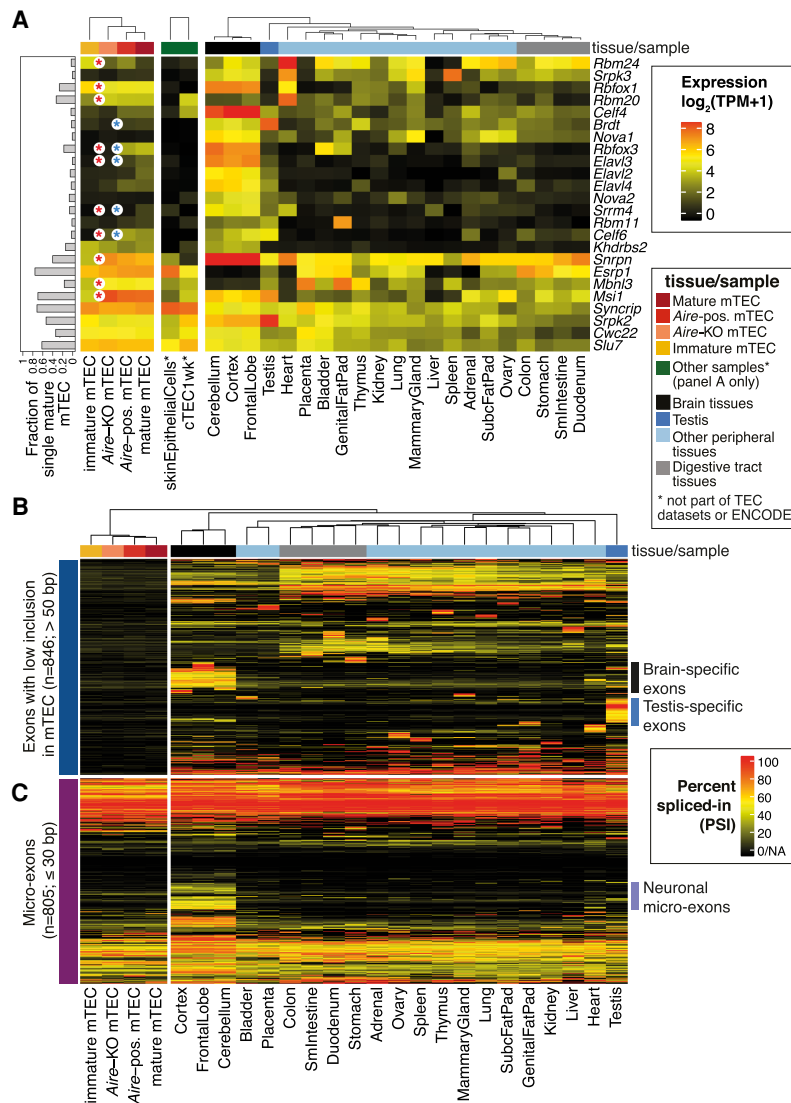


Figure 5. Selective expression of peripheral SFs and exons in mTEC. (A) Expression of a set of tissue-restricted ($\tau > 0.5$) genes that encode for bona fide SFs (Supplemental Table 8) in mTEC populations, skin epithelia, cTEC, and peripheral tissues (mouse ENCODE Project). The bars (left) depict the fraction of a set of single mature mTEC that express each factor. Significant differences in expression between immature versus mature (Aire-KO) mTEC or Aire-KO versus Aire-positive mTEC are indicated by red and blue asterisks, respectively (BH adjusted P -value < 0.05 , $|fc| > 2$, DESeq2 analysis of the population RNA sequencing data, $n = 2$ biological replicates/condition). (B) Patterns of protein-coding exon (>50 bp in length) inclusion for exons with a low inclusion rate in mTEC (mean PSI < 0.1 ; max PSI < 0.2) that were included in at least one of the peripheral tissues (PSI > 0.5). (C) Microexon (≤ 30 bp) inclusion in transcripts from protein-coding genes in the mTEC population and peripheral tissue samples (Supplemental Fig. 11B).

RBFOX is present with AIRE in mTEC nuclei and influences TEC development

Among the TRSFs expressed by mTEC, *Rbfox1* was of particular interest owing to the established roles of RBFOX in controlling alternative splicing during muscle and brain development (Conboy 2017). We noted that *Rbfox1* and *Rbfox2* transcripts in mTEC predominantly included the neuronal B40 exon and excluded the muscle-specific M43 exon (Supplemental Fig. 12A,B; Nakahata and Kawamoto 2005; Damianov and Black 2010). Confocal image analysis of AIRE-expressing mTEC with a pan RBFOX anti-RRM domain antibody revealed that RBFOX factors are present in the

nucleus in proximity to AIRE speckles (Fig. 6A). Although this observation does not imply colocalization, it does provide evidence that RBFOX factors are coexpressed with AIRE and hence may be important for the splicing of *Aire*-regulated transcripts.

Both *Rbfox1* and *Rbfox2* showed a robust nonpromiscuous and tissue-restricted expression in mTEC (Fig. 5A; Supplemental Fig. 12C,D). To investigate their functions in mTEC, we generated conditional knockouts (Gehman et al. 2011, 2012) using the expression of Cre recombinase under the transcriptional control of the *Foxn1* locus (Žuklys et al. 2009). Total thymus cellularity of *Rbfox1* thymus knockout (tKO) (*Rbfox1*^{lox/lox}; *Foxn1*^{cre/+}) mice was 15% increased, whereas that of age-matched *Rbfox2* tKO (*Rbfox2*^{lox/lox}; *Foxn1*^{cre/+}) mice remained unchanged when compared to controls (Supplemental Figs. 13C, 14C). However, the phenotype of cTEC, mTEC, immature mTEC, and mature mTEC was unaffected in *Rbfox1* tKO mice (marker phenotypes in Supplemental Table 9; Supplemental Fig. 13). In contrast, *Rbfox2* tKO mice showed an overall reduction in TEC frequency (CD45⁺ EPCAM⁺, *Rbfox2* tKO 0.12% vs. WT 0.17% of all thymic cells recovered, $P < 0.05$, two-sided Welch two sample t -test) (Supplemental Fig. 14D,E). The *Rbfox2* tKO mice also had proportionally more cTEC (10.4% vs. 6.7% of all TEC, $P = 0.02$, Welch two sample t -test) and proportionally fewer mTEC (86.2% vs. 90.8% of all TEC, $P = 0.03$, Welch two sample t -test) present in their epithelial scaffolds (Fig. 6B,C). *Rbfox1/2* tKO (*Rbfox1*^{lox/lox}; *Rbfox2*^{lox/lox}; *Foxn1*^{cre/+}) animals showed changes in cTEC and mTEC frequency that were quantitatively similar to that of *Rbfox2* tKO mice (Supplemental Fig. 15) suggesting that, as in other tissues (Gehman et al. 2011), *Rbfox2* can compensate for the loss of *Rbfox1* in TEC or that *Rbfox1* might not play a role in this aspect of TEC development.

Thymocyte development and positive selection were quantitatively normal in the *Rbfox1/2* tKO animals (Supplemental Fig. 16). In summary these data show that *Rbfox1* contributes to the regulation of thymic cellularity, and *Rbfox2* increases the relative frequency of mTEC and decreases the relative frequency of cTEC within the TEC compartment.

RBFOX contributes to the alternative splicing of self-antigen transcripts in mTEC

To investigate the impact of RBFOX on the mTEC transcriptome we performed RNA sequencing of immature and mature mTEC

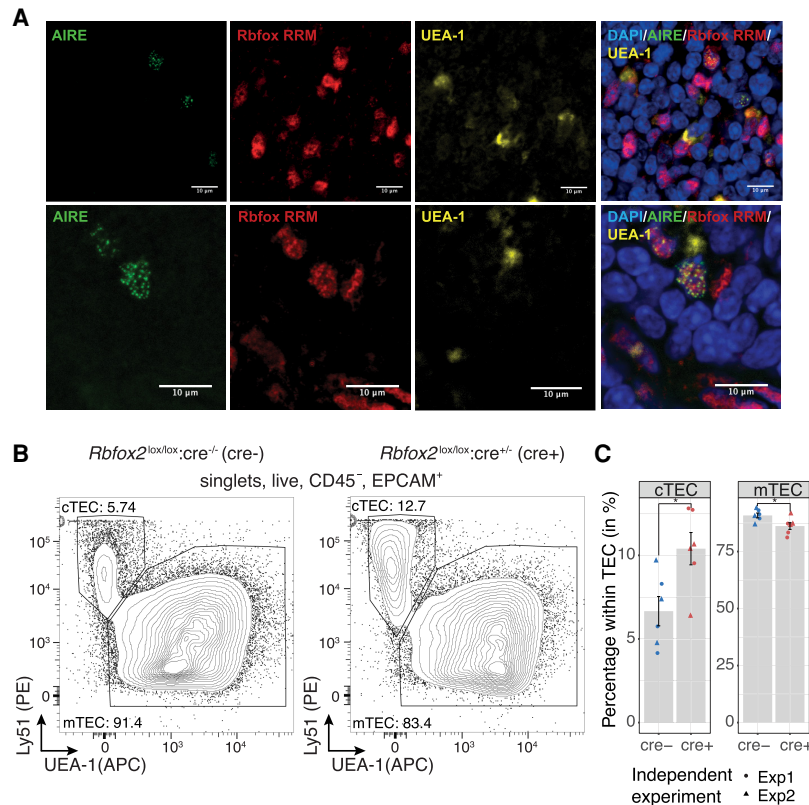


Figure 6. RBFOX is present with AIRE in mTEC nuclei and influences TEC development. (A) Confocal immunofluorescence analysis of the localization of AIRE (green) and RBFOX (red) in the thymic medulla. RBFOX was detected using an anti-RRM domain-specific antibody, mTEC were identified by reactivity with the lectin UEA-1 (yellow), and nuclei were labeled using DAPI (blue). Two representative sections are shown in the upper and lower panels. (B) Representative flow cytometric analysis of cTEC and mTEC frequencies among thymic epithelial cells extracted from *Rbfox2* tKO animals. (C) Quantification of cTEC and mTEC frequency in *Rbfox2* tKO animals. The mean \pm SE from $n = 2$ independent experiments is shown in bar graphs; (*) P -value < 0.05 for two-sided Welch two sample t -test.

isolated from *Rbfox1* tKO, *Rbfox2* tKO, and control mice. In mature mTEC, loss of *Rbfox1* or *Rbfox2* had no obvious effect on the expression of known maturation or subpopulation markers (Supplemental Fig. 17) and induced only minor changes in gene expression (Supplemental Fig. 18A) but caused hundreds of alterations in alternative splicing (*Rbfox1* tKO: $n = 559$ events in 535 genes; *Rbfox2* tKO: $n = 668$ events in 624 genes; $FDR < 0.05$, $|\Delta PSI| > 0.2$) (Supplemental Fig. 18; Supplemental Tables 10, 11). In mature mTEC, loss of *Rbfox1* and *Rbfox2* resulted in $n = 123$ significant alternative splicing events in 104 TRA genes and $n = 122$ events in 110 TRA genes, respectively. In immature mTEC, loss of *Rbfox1* and *Rbfox2* resulted in the alternative splicing of a smaller number of TRA genes (72 events in 51 TRA genes and 43 events in 36 TRA genes, respectively) in keeping with the lower level of promiscuous gene expression in these cells (Supplemental Fig. 19A–C). Analysis of the overlap of the sets of genes alternatively spliced in *Rbfox1* and *Rbfox2* tKO mice revealed that the splicing of 10 TRA genes was affected by both factors in mature mTEC (odds ratio [OR] = 4.64, $P = 1.71 \times 10^{-4}$, two-sided Fisher's exact test), whereas the splicing of six TRA genes was affected by both factors in immature mTEC (OR = 18.5, $P = 3.23 \times 10^{-6}$, two-sided Fisher's exact test) (Supplemental Fig. 19D). We did not find an obvious overlap between the splicing events identified in the *Rbfox1* and *Rbfox2* tKO mTEC

and those found in *Aire*-knockout mTEC or that were associated with mTEC maturation (Supplemental Fig. 20).

We next sought to establish whether the genes that showed modified splicing patterns in the *Rbfox1* and *Rbfox2* tKO mature mTEC included those previously predicted to be RBFOX targets in mouse brain (Weyn-Vanhentenryck et al. 2014) and muscle tissues (Pedrotti et al. 2015; Singh et al. 2018). In mature mTEC, we identified large and significant overlaps between predicted RBFOX brain target genes and those for which loss of *Rbfox2* induced (1) exon skipping in mTEC (OR = 7.8, BH adjusted $P = 5.4 \times 10^{-19}$) or (2) the use of mutually exclusive exons (OR = 13.8, BH adjusted $P = 1.4 \times 10^{-5}$) in mTEC (Fig. 7A). Similarly large and significant overlaps were identified for the RBFOX muscle target genes (Supplemental Fig. 18D). In keeping with this observation, many of the TRA genes alternatively spliced upon loss of *Rbfox2* showed specific expression in neuronal tissues (Supplemental Fig. 18E). GO biological processes overrepresented among the genes alternatively spliced upon loss of *Rbfox2* in mature mTEC included “muscle contraction” and “neuron migration” in line with the known roles of *Rbfox* gene family members in muscle and neuronal tissues (Fig. 7B; Conboy 2017). Examples of genes harboring alternative splicing events in the *Rbfox2* tKO mature mTEC included *Fn1* and *Insr*, two known RBFOX2 target genes (Chen and Manley 2009; Weyn-Vanhentenryck et al. 2014) as well as the TRA *Myom2* (Fig. 7C).

Previous studies have identified an enriched RBFOX recognition motif in proximity to RBFOX-regulated exons (Weyn-Vanhentenryck et al. 2014). We observed a significant enrichment of the conserved RBFOX RRM WGCAUGM motif (Ray et al. 2013) upstream of exons repressed and downstream from exons enhanced by the presence of *Rbfox2* in both mature and immature mTEC (Fig. 7D; Supplemental Fig. 19E), following the patterns previously described for this factor (Weyn-Vanhentenryck et al. 2014). In summary these data provide evidence that RBFOX SFs directly regulate the splicing of both promiscuously and nonpromiscuously expressed genes in mature mTEC.

Discussion

Our work shows that although mature mTEC produce an exceptionally high proportion (~60%) of peripheral splice variants, they are unable to recreate the full diversity of TRA gene isoforms that are present in the periphery. The concept of selective representation of peripheral isoforms in TEC is supported by a recent qPCR-based study that estimated that a quarter of the genes studied contained epitopes not expressed in the thymus (Shilov et al. 2019). Using our refined iTRA approach, we noticed that genes that have far fewer splice junctions in murine mTEC included

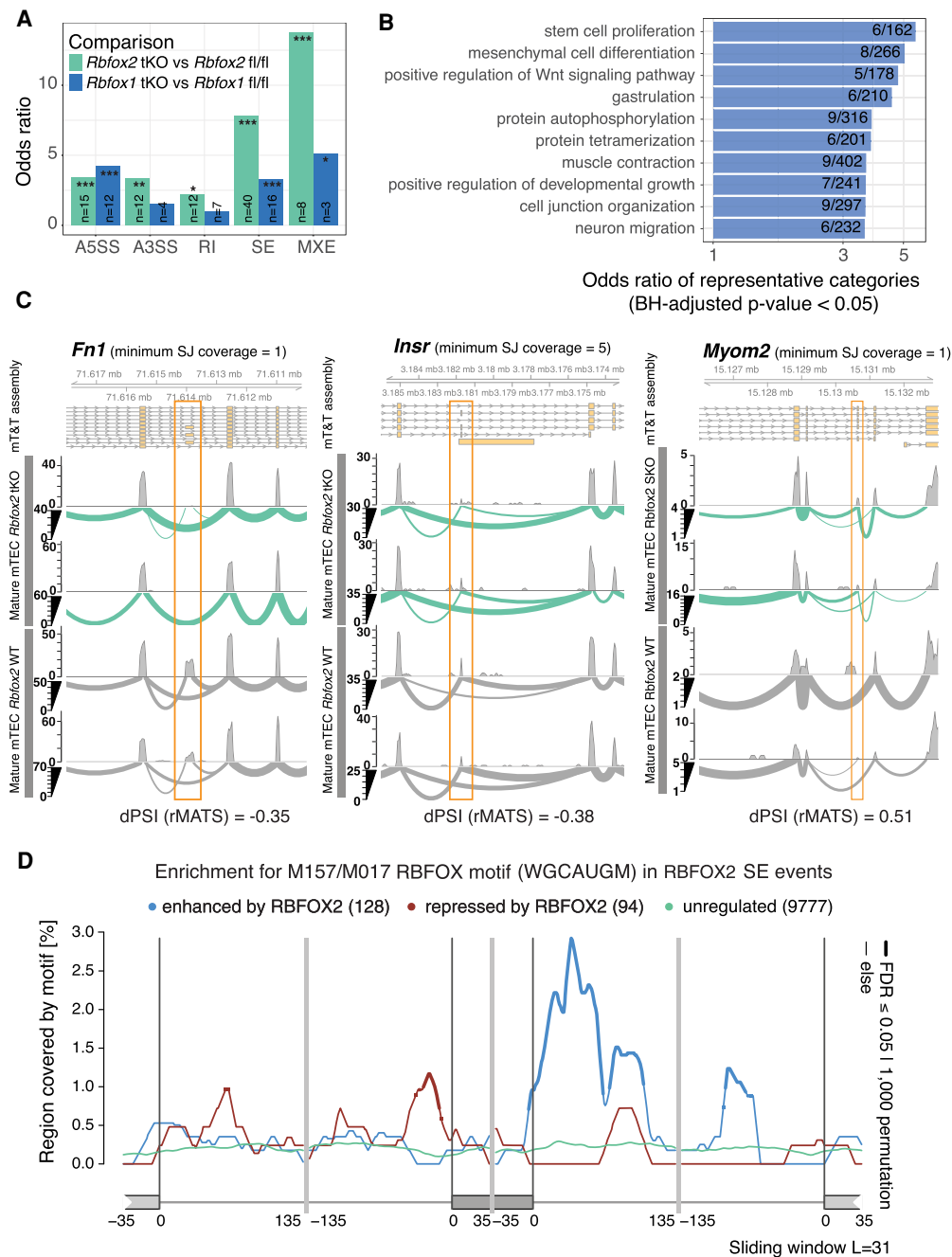


Figure 7. RBFOX regulates alternative splicing in TEC. (A) The bar plots show the enrichments (ORs) of previously predicted *Rbfox* target genes (Weyn-Vanhentenryck et al. 2014) in the sets of genes differentially spliced in the *Rbfox1* tKO or *Rbfox2* tKO mature mTEC (two-sided Fisher's exact tests, BH adjusted *P*-values): (*) *P* < 0.05, (**) *P* < 0.01, (***) *P* < 0.001; abbreviations as for Figure 4B. (B) Selected GO biological processes significantly overrepresented (one-sided Fisher's exact tests, BH adjusted *P*-values < 0.05) in genes that contained significantly RBFOX2-regulated SE events in mature mTEC. (C) Three examples of significantly (FDR < 0.05) RBFOX2-regulated splicing events in mature mTEC. *Fn1* and *Insr* are known RBFOX target genes (Chen and Manley 2009; Weyn-Vanhentenryck et al. 2014). *Myom2* is an example of a non-*Aire* TRA gene. (D) Enrichment of the RBFOX recognition motif (M159/M017) (Ray et al. 2013) in the sequences surrounding exons that were found to be significantly regulated by RBFOX2 in mature mTEC (Supplemental Fig. 18B). The lines show enrichments for sets of exons found to be enhanced (blue), repressed (red), or not significantly regulated (green) by RBFOX2. Thicker lines indicate regions of statistically significant enrichment (FDR ≤ 0.05, *n* = 1000 permutations).

those encoding homologs of known autoantibody targets in human autoimmune disease (such as PLP1, MBP, and CYP21A2). Although confirmation is needed by analysis of human mTEC, these findings suggest, as is understood to be the case in a mouse model of EAE, that absence of TRA isoforms from the thymus may lead to an incomplete central screening of peripheral tissue

epitopes and thus contribute to the development human autoimmune disease (Anderson et al. 2000; Klein et al. 2000). We found for TRA genes that very substantive numbers of splice junctions were absent in TEC, but although we consider it to be unlikely, we cannot formally exclude the possibility that peptides from non-thymically expressed splice isoforms may be redundant or

excluded from the peripheral MHC peptidome. Overall, our rigorous transcriptome assembly-based census of transcript structures showed an even representation of tissue-restricted transcripts from the 21 surveyed peripheral tissues in mTEC. Most notably we found a lower representation of transcripts from the testis and brain together with a corresponding absence (or only very weak) expression of brain and testis-specific SFs in mTEC. It is possible that it may be functionally less important to educate T cells against splice isoform epitopes specific to these tissues, which constitute so-called immune privileged sites. Further, the unusually high complexity of splicing in the testes and brain suggests the possibility that immune surveillance may act as a constraint on the evolution of splicing complexity in non-immune privileged peripheral tissues. Our finding that neuronal microexons are not frequently spliced into transcripts in mTEC is likely explained by the absence of the microexon SF *Srrm4* in these cells (Quessel-Vallières et al. 2015). The limited inclusion of neuronal microexons in mTEC provides a possible link between the observations that such exons are misregulated in the brains of patients with autism (Irimia et al. 2014) and growing evidence that autoimmunity may be involved in autism (Hughes et al. 2018). Aside from the testis and cerebellum, adrenal and liver tissues showed a weaker isoform representation in mature TEC. In humans, these tissues are known sites of autoimmunity: autoimmune adrenalitis is the most common cause of Addison's disease, and liver diseases such as autoimmune hepatitis, primary biliary cirrhosis, and sclerosing cholangitis are all thought to be the consequence of autoimmunity (Decock et al. 2009). Alternative splicing has been implicated in such diseases (Webster 2017), and our data suggest that lack of thymic representation of isoforms specific to these tissues might contribute to a higher susceptibility for autoimmunity.

Our investigations revealed a large number of splicing differences between immature and mature mTEC. One possible explanation for this observation is the expression of AIRE in mature mTEC, but our analysis of *Aire*-knockout mTEC confirms that this molecule plays only a small role in alternative splicing in mTEC. Rather, we found that *Aire* promotes the use of longer "classical" transcripts containing both 5' and 3' UTRs in agreement with and extending the existing understanding that AIRE releases stalled RNA polymerases that otherwise produce "short" transcripts (Giraud et al. 2012). A comprehensive survey of splicing factor expression revealed that mTEC complement typical epithelial splicing factors such as *Esrp1/2* with a small number of peripheral restricted splicing factors. Unlike skin epithelia, mature mTEC showed nonpromiscuous expression of *Rbm20*, *Msi1*, and *Rbfox1*. This observation, along with the apparently incomplete representation of peripheral structures and absence of excessive numbers of novel splice isoforms in mTEC, suggests that TEC undertake a selective program of alternative splicing to ensure the accurate representation of a subset of the peripheral splice isoform repertoire. We hence conclude that transcript structures in mTEC are primarily shaped by a small number of splicing and mRNA processing factors (Guyon et al. 2020) and that mTEC do not rely on "promiscuous" mechanisms to promote and increase splice isoform diversity, as previously suggested (Keane et al. 2015). Rather (and assuming that our observations hold in humans), a limited thymic representation of peripheral splice isoforms is consistent with the concept that tissue-specific isoforms are relevant sources of autoantigens in immune-mediated diseases (Evsukova et al. 2010; Juan-Mateu et al. 2016; Newman et al. 2017). Furthermore, characterization of transcript structures in human thymic epithelial cells would be expected to aid the identi-

cation of autoantigen transcripts encoding epitopes against which central tolerance has not been achieved (Ng et al. 2004). Knowledge of such epitopes is of great value for development of antigen-specific therapies for autoimmune disease such as those based on the use of tolerizing peptides or tolerogenic dendritic cells (Pozsgay et al. 2017; Mosanya and Isaacs 2019).

Our discovery of nonpromiscuous *Rbfox1* expression in TEC was of particular interest because this factor is otherwise restricted to muscle and neural tissues in which it plays important developmental roles (Conboy 2017). We detected *Rbfox1* in medullary but not cortical TEC, suggesting that it may be important for the development or function of this subpopulation. We therefore performed functional analysis of the role of *Rbfox1* and its homolog *Rbfox2* in mTEC, excluding *Rbfox3* from our investigations because it showed only weak and promiscuous expression in these cells. Phenotypic analysis of TEC in animals with a selective loss of one or both of these factors showed that *Rbfox2* increases mTEC frequency and decreases cTEC frequency within the TEC population. Recently, it has been shown that although in the embryonic and newborn thymus, cortical and medullary TEC arise from a common bipotent progenitors, in the adult mouse, mTEC are mostly replenished by lineage-restricted cells (Ohigashi et al. 2015). Our data suggest that members of the *Rbfox* gene family might influence the cell fate choices of the bipotent progenitors or act later to promote the differentiation within the mTEC lineage, but other possibilities cannot be excluded.

In mature mTEC we found that RBFOX factors shape the splicing of both promiscuously and nonpromiscuously expressed genes. It is possible that RBFOX-mediated splicing of promiscuously expressed genes may involve a molecular interaction with AIRE because we found them to be coexpressed in mTEC nuclei, and both proteins are known to partner with DDX5 (Abramson et al. 2010; Damianov et al. 2016). The larger role identified for *Rbfox2* in splicing in mTEC was not unexpected, because in the cerebellum it is known that *Rbfox2* can largely compensate for loss of *Rbfox1* but *Rbfox1* is less well able to ameliorate an absence of *Rbfox2* (Gehman et al. 2012). In addition, the changes in alternative splicing identified following loss of *Rbfox2* may involve *Rbfox1* because *Rbfox1* was itself differentially spliced in the absence of *Rbfox2*. We found examples of both *Aire*-dependent and *Aire*-independent TRA that were alternatively spliced by *Rbfox2* in mTEC. Given the weak population-level expression of TRA in mTEC and the relatively small amount of biological material sequenced for this analysis, we expect the actual number of TRA regulated by RBFOX factors in mTEC to be substantially higher than is reported here. Ultimately, phenotypic investigation of TEC-specific *Rbfox* knockout animals will be important to determine the extent to which their contribution to the splicing of TRA in mTEC is functionally important for the establishment of central tolerance. In summary, our data show that mTEC reuse a small set of peripheral SFs that includes RBFOX to selectively reproduce a broad but ultimately limited subset of the peripheral splice isoform repertoire.

Methods

Mice

Wild-type C57BL/6 mice were obtained from Harlan Laboratories or Janvier and maintained as a laboratory in-house colony. *Aire*^{GFP/+} mice were previously described (Sansom et al. 2014). *Rbfox1* and *Rbfox2* mutant mice (Gehman et al. 2011, 2012) were maintained on a mixed 129S2/Sv x C57BL/6J background. All

animals were kept under specific pathogen-free conditions, and work performed was covered by a UK Home Office Project Licence (G.A.H.).

Isolation, sorting, and immunostaining of thymic epithelial cells

TEC were isolated from multiple thymi and sorted by surface phenotype (Supplemental Tables 9, 12; Supplemental Methods; Supplemental Fig. 21; Dhalla et al. 2020). Sections were prepared and stained for confocal microscopy as described in the Supplemental Methods.

RNA sequencing data

For analysis of immature, mature, and *Aire*-knockout mTEC poly(A)⁺ RNA-seq libraries were prepared from cells pooled from multiple female mice (1 μg of total RNA; 4 wk old; n = 2 biological replicates) (Supplemental Table 1) and subjected to 101 bp paired-end stranded Illumina RNA-seq. For analysis of *Rbfox1* tKO and *Rbfox2* tKO animals, immature and mature mTEC were isolated from individual mice and sex-matched littermate controls (10,000–15,000 cells per animal, n = 2 biological replicates, 4 wk old). Stranded RNA-seq libraries were prepared (NEB) and subjected to 150 bp paired-end Illumina sequencing.

For long-read sequencing, CD80⁺86⁺ or CD80⁺86⁻ mTEC were obtained from 4- to 6-wk-old female wild-type mice and processed as above. Details of the ONT sequencing can be found in the Supplemental Methods.

Computational methods

To construct the mT&T assembly, reads were mapped with HISAT2 (Kim et al. 2019), and reference-guided assembly was performed with StringTie (Genome assembly GRCm38.p5, Ensembl release 91 annotations) (Pertea et al. 2015). To identify reproducibly detected transcripts, we used a modified npIDR approach (Supplemental Fig. 1; Dobin et al. 2013; Pervouchine et al. 2015). ONT reads were mapped with minimap2 (Li 2018). We quantified expression level with Salmon (TPMs) or featureCounts (counts) for the Illumina data sets and with featureCounts for the ONT data sets (Liao et al. 2014; Patro et al. 2017). Splicing analyses were performed using SJCounts (Pervouchine et al. 2013), rMATS (for identification of differential events) (Shen et al. 2014), and SUPPA (for computation of PSI values and transcript annotation) (Alamancos et al. 2015). Differential expression analyses were performed using DESeq2 (genes) (Love et al. 2014), edgeR (splice junctions) (Robinson et al. 2010), or kallisto/sleuth (transcript usage) (Bray et al. 2016; Pimentel et al. 2017). For each analysis, reads were pre-processed, trimmed, de-duplicated, and down-sampled as appropriate to avoid bias. For details of other analyses, method references, and algorithm parameters and versions, please see the “Computational methods” section of the Supplemental Methods.

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO); <https://www.ncbi.nlm.nih.gov/geo/> under accession number GSE145931.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Prof. Chris Ponting (Univ. Edinburgh) for support and advice. The anti-RRM antibody and *Rbfox* mutant mice were kind gifts of Prof. Douglas Black and Julia Nikolic (Univ. California, Los Angeles). We thank Prof. Black for helpful comments and advice. This work was initiated with funds from the Medical Research Council (MRC) Computational Genomics Analysis and Training (CGAT) program (G1000902). S.N.S. and G.A.H. were supported by Wellcome Trust funding (#066521). S.N.S. and M.A. are supported by funding from the Kennedy Trust for Rheumatology Research (KTRR) (KENN171803/Core support). K.J. was supported by a Wellcome Trust PhD studentship and funds from the MRC (MR/S025308/1, MR/S035850/1).

Author contributions: K.J. performed the experiments and computational analyses. S.N.S. performed additional analyses of splice junction counts and transcript structure. N.S.-D. generated RNA sequencing data from the wild-type and *Aire*-knockout mTEC. S.M. assisted with the flow cytometry experiments. M.A. and M.L. performed, and D.B. was responsible for, the ONT library construction and sequencing. S.N.S. conceived the study with input from G.A.H. K.J., G.A.H., and S.N.S. designed the experiments and analyses, interpreted results, and wrote the manuscript. G.A.H. and S.N.S. supervised the study.

References

- Abramson J, Anderson G. 2017. Thymic epithelial cells. *Annu Rev Immunol* **35**: 85–118. doi:10.1146/annurev-immunol-051116-052320
- Abramson J, Goldfarb Y. 2016. AIRE: from promiscuous molecular partnerships to promiscuous gene expression. *Eur J Immunol* **46**: 22–33. doi:10.1002/eji.201545792
- Abramson J, Giraud M, Benoist C, Mathis D. 2010. Aire’s partners in the molecular control of immunological tolerance. *Cell* **140**: 123–135. doi:10.1016/j.cell.2009.12.030
- Alamancos GP, Pagès A, Trincado JL, Bellora N, Eyraes E. 2015. Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA* **21**: 1521–1531. doi:10.1261/rna.051557.115
- Anderson AC, Nicholson LB, Legge KL, Turchin V, Zaghoulani H, Kuchroo VK. 2000. High frequency of autoreactive myelin proteolipid protein-specific T cells in the periphery of naive mice: mechanisms of selection of the self-reactive repertoire. *J Exp Med* **191**: 761–770. doi:10.1084/jem.191.5.761
- Berger T, Rubner P, Schautzer F, Egg R, Ulmer H, Mayringer I, Dilitz E, Deisenhammer F, Reindl M. 2003. Antimyelin antibodies as a predictor of clinically definite multiple sclerosis after a first demyelinating event. *N Engl J Med* **349**: 139–145. doi:10.1056/NEJMoa022328
- Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**: 525–527. doi:10.1038/nbt.3519
- Chen M, Manley JL. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10**: 741–754. doi:10.1038/nrm2777
- Chuprin A, Avin A, Goldfarb Y, Herzig Y, Levi B, Jacob A, Sela A, Katz S, Grossman M, Guyon C, et al. 2015. The deacetylase Sirt1 is an essential regulator of *Aire*-mediated induction of central immunological tolerance. *Nat Immunol* **16**: 737–745. doi:10.1038/ni.3194
- Conboy JG. 2017. Developmental regulation of RNA processing by *Rbfox* proteins. *Wiley Interdiscip Rev RNA* **8**: e1398. doi:10.1002/wrna.1398
- Damianov A, Black DL. 2010. Autoregulation of fox protein expression to produce dominant negative splicing factors. *RNA* **16**: 405–416. doi:10.1261/rna.1838210
- Damianov A, Ying Y, Lin CH, Lee JA, Tran D, Vashisht AA, Bahrami-Samani E, Xing Y, Martin KC, Wohlschlegel JA, et al. 2016. *Rbfox* proteins regulate splicing as part of a large multiprotein complex LASR. *Cell* **165**: 606–619. doi:10.1016/j.cell.2016.03.040
- Danan-Gotthold M, Guyon C, Giraud M, Levanon EY, Abramson J. 2016. Extensive RNA editing and splicing increase immune self-representation diversity in medullary thymic epithelial cells. *Genome Biol* **17**: 219. doi:10.1186/s13059-016-1079-9
- Decock S, McGee P, Hirschfield GM. 2009. Autoimmune liver disease for the non-specialist. *BMJ* **339**: b3305. doi:10.1136/bmj.b3305
- Dhalla F, Baran-Gale J, Maio S, Chappell L, Holländer GA, Ponting CP. 2020. Biologically indeterminate yet ordered promiscuous gene expression in

- single medullary thymic epithelial cells. *EMBO J* **39**: e101828. doi:10.15252/embj.2019101828
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21. doi:10.1093/bioinformatics/bts635
- Evsyukova I, Somarelli JA, Gregory SG, Garcia-Blanco MA. 2010. Alternative splicing in multiple sclerosis and other autoimmune diseases. *RNA Biol* **7**: 462–473. doi:10.4161/rna.7.4.12301
- Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, Shiue L, Ares M Jr, Mody I, Black DL. 2011. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat Genet* **43**: 706–711. doi:10.1038/ng.841
- Gehman LT, Meera P, Stoilov P, Shiue L, O'Brien JE, Meisler MH, Ares M Jr, Otis TS, Black DL. 2012. The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. *Genes Dev* **26**: 445–460. doi:10.1101/gad.182477.111
- Giraud M, Yoshida H, Abramson J, Rahl PB, Young RA, Mathis D, Benoist C. 2012. Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. *Proc Natl Acad Sci* **109**: 535–540. doi:10.1073/pnas.1119351109
- Granados DP, Laumont CM, Thibault P, Perreault C. 2015. The nature of self for T cells—a systems-level perspective. *Curr Opin Immunol* **34**: 1–8. doi:10.1016/j.coi.2014.10.012
- Guo W, Schafer S, Greaser ML, Radke MH, Liss M, Govindarajan T, Maatz H, Schulz H, Li S, Parrish AM, et al. 2012. *RBM20*, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat Med* **18**: 766–773. doi:10.1038/nm.2693
- Guyon C, Jmari N, Padonou F, Li YC, Ucar O, Fujikado N, Couplier F, Blanchet C, Root DE, Giraud M. 2020. Aire-dependent genes undergo Clp1-mediated 3'UTR shortening associated with higher transcript stability in the thymus. *eLife* **9**: e52985. doi:10.7554/eLife.52985
- Handel AE, Shikama-Dorn N, Zhanybekova S, Maio S, Graedel AN, Zuklys S, Ponting CP, Holländer GA. 2018. Comprehensively profiling the chromatin architecture of tissue restricted antigen expression in thymic epithelial cells over development. *Front Immunol* **9**: 2120. doi:10.3389/fimmu.2018.02120
- Hughes HK, Mills Ko E, Rose D, Ashwood P. 2018. Immune dysfunction and autoimmunity as pathological mechanisms in autism spectrum disorders. *Front Cell Neurosci* **12**: 405. doi:10.3389/fncel.2018.00405
- Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M, Quesnel-Vallières M, Tapial J, Raj B, O'Hanlon D, et al. 2014. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**: 1511–1523. doi:10.1016/j.cell.2014.11.035
- Jacob AG, Smith CWJ. 2017. Intron retention as a component of regulated gene expression programs. *Hum Genet* **136**: 1043–1057. doi:10.1007/s00439-017-1791-x
- Juan-Mateu J, Villate O, Eizirik DL. 2016. MECHANISMS IN ENDOCRINOLOGY: alternative splicing: the new frontier in diabetes research. *Eur J Endocrinol* **174**: R225–R238. doi:10.1530/EJE-15-0916
- Kajikawa M, Baba T, Tomaru U, Watanabe Y, Koganei S, Tsuji-Kawahara S, Matsumoto N, Yamamoto K, Miyazawa M, Maenaka K, et al. 2006. MHC class I-like MILL molecules are β_2 -microglobulin-associated, GPI-anchored glycoproteins that do not require TAP for cell surface expression. *J Immunol* **177**: 3108–3115. doi:10.4049/jimmunol.177.5.3108
- Keane P, Ceredig R, Seoighe C. 2015. Promiscuous mRNA splicing under the control of AIRE in medullary thymic epithelial cells. *Bioinformatics* **31**: 986–990. doi:10.1093/bioinformatics/btu785
- Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, Stamm S. 2013. Function of alternative splicing. *Gene* **514**: 1–30. doi:10.1016/j.gene.2012.07.083
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**: 907–915. doi:10.1038/s41587-019-0201-4
- Klein L, Klugmann M, Nave KA, Tuohy VK, Kyewski B. 2000. Shaping of the autoreactive T-cell repertoire by a splice variant of self protein expressed in thymic epithelial cells. *Nat Med* **6**: 56–61. doi:10.1038/71540
- Klein L, Kyewski B, Allen PM, Hogquist KA. 2014. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol* **14**: 377–391. doi:10.1038/nri3667
- Kryuchkova-Mostacci N, Robinson-Rechavi M. 2017. A benchmark of gene expression tissue-specificity metrics. *Brief Bioinform* **18**: 205–214. doi:10.1093/bib/bbw008
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100. doi:10.1093/bioinformatics/bty191
- Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930. doi:10.1093/bioinformatics/btt656
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. doi:10.1186/s13059-014-0550-8
- Meredith M, Zemmour D, Mathis D, Benoist C. 2015. Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nat Immunol* **16**: 942–949. doi:10.1038/ni.3247
- Merkin J, Russell C, Chen P, Burge CB. 2012. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**: 1593–1599. doi:10.1126/science.1228186
- Mosanya CH, Isaacs JD. 2019. Tolerising cellular therapies: what is their promise for autoimmune disease? *Ann Rheum Dis* **78**: 297–310. doi:10.1136/annrheumdis-2018-214024
- Murphy D, Cieply B, Carstens R, Ramamurthy V, Stoilov P. 2016. The Musashi 1 controls the splicing of photoreceptor-specific exons in the vertebrate retina. *PLoS Genet* **12**: e1006256. doi:10.1371/journal.pgen.1006256
- Nakahata S, Kawamoto S. 2005. Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. *Nucleic Acids Res* **33**: 2078–2089. doi:10.1093/nar/gki338
- Newman JRB, Conesa A, Mika M, New FN, Onengut-Gumusc S, Atkinson MA, Rich SS, McIntyre LM, Concannon P. 2017. Disease-specific biases in alternative splicing and tissue-specific dysregulation revealed by multi-tissue profiling of lymphocyte gene expression in type 1 diabetes. *Genome Res* **27**: 1807–1815. doi:10.1101/gr.217984.116
- Ng B, Yang F, Huston DP, Yan Y, Yang Y, Xiong Z, Peterson LE, Wang H, Yang XF. 2004. Increased noncanonical splicing of autoantigen transcripts provides the structural basis for expression of intolerized epitopes. *J Allergy Clin Immunol* **114**: 1463–1470. doi:10.1016/j.jaci.2004.09.006
- Ohigashi I, Zuklys S, Sakata M, Mayer CE, Hamazaki Y, Minato N, Hollander GA, Takahama Y. 2015. Adult thymic medullary epithelium is maintained and regenerated by lineage-restricted cells rather than bipotent progenitors. *Cell Rep* **13**: 1432–1443. doi:10.1016/j.celrep.2015.10.012
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**: 417–419. doi:10.1038/nmeth.4197
- Pedrotti S, Giudice J, Dagnino-Acosta A, Knoblauch M, Singh RK, Hanna A, Mo Q, Hicks J, Hamilton S, Cooper TA. 2015. The RNA-binding protein Rbfox1 regulates splicing required for skeletal muscle structure and function. *Hum Mol Genet* **24**: 2360–2374. doi:10.1093/hmg/ddv003
- Perteau M, Perteau GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. Stringtie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**: 290–295. doi:10.1038/nbt.3122
- Pervouchine DD, Knowles DG, Guigo R. 2013. Intron-centric estimation of alternative splicing from RNA-seq data. *Bioinformatics* **29**: 273–274. doi:10.1093/bioinformatics/bts678
- Pervouchine DD, Djebali S, Breschi A, Davis CA, Barja PP, Dobin A, Tanzer A, Lagarde J, Zaleski C, See LH, et al. 2015. Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. *Nat Commun* **6**: 5903. doi:10.1038/ncomms6903
- Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods* **14**: 687–690. doi:10.1038/nmeth.4324
- Pozsgay J, Szekanecz Z, Sármay G. 2017. Antigen-specific immunotherapies in rheumatic diseases. *Nat Rev Rheumatol* **13**: 525–537. doi:10.1038/nrnheum.2017.107
- Quesnel-Vallières M, Irimia M, Cordes SP, Blencowe BJ. 2015. Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes Dev* **29**: 746–759. doi:10.1101/gad.256115.114
- Quesnel-Vallières M, Dargaei Z, Irimia M, Gonatopoulos-Pournatzis T, Ip JY, Wu M, Sterne-Weiler T, Nakagawa S, Woodin MA, Blencowe BJ, et al. 2016. Misregulation of an activity-dependent splicing network as a common mechanism underlying autism spectrum disorders. *Mol Cell* **64**: 1023–1034. doi:10.1016/j.molcel.2016.11.033
- Raposo B, Merky P, Lundqvist C, Yamada H, Urbonaviciute V, Niaudet C, Viljanen J, Kihlberg J, Kyewski B, Ekwall O, et al. 2018. T cells specific for post-translational modifications escape intrathymic tolerance induction. *Nat Commun* **9**: 353. doi:10.1038/s41467-017-02763-y
- Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, et al. 2013. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**: 172–177. doi:10.1038/nature12311
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140. doi:10.1093/bioinformatics/btp616
- Sansom SN, Shikama-Dorn N, Zhanybekova S, Nusspaumer G, Macaulay IC, Deadman ME, Heger A, Ponting CP, Holländer GA. 2014. Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. *Genome Res* **24**: 1918–1931. doi:10.1101/gr.171645.113
- Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. 2014. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci* **111**: E5593–E5601. doi:10.1073/pnas.1419161111

- Shilov ES, Gorshkova EA, Minnegaliev AR, Potashnikova DM. 2019. [Splicing pattern of mRNA in thymus epithelial cells limits the transcriptome available for negative selection of autoreactive T cells]. *Mol Biol (Mosk)* **53**: 109–119. doi:10.1134/S0026898419010154
- Singh RK, Kolonin AM, Fiorotto ML, Cooper TA. 2018. Rbfox-splicing factors maintain skeletal muscle mass by regulating Calpain3 and proteostasis. *Cell Rep* **24**: 197–208. doi:10.1016/j.celrep.2018.06.017
- Starck SR, Shastri N. 2016. Nowhere to hide: unconventional translation yields cryptic peptides for immune surveillance. *Immunol Rev* **272**: 8–16. doi:10.1111/imr.12434
- St-Pierre C, Brochu S, Vanegas JR, Dumont-Lagacé M, Lemieux S, Perreault C. 2013. Transcriptome sequencing of neonatal thymic epithelial cells. *Sci Rep* **3**: 1860. doi:10.1038/srep01860
- St-Pierre C, Trofimov A, Brochu S, Lemieux S, Perreault C. 2015. Differential features of AIRE-induced and AIRE-independent promiscuous gene expression in thymic epithelial cells. *J Immunol* **195**: 498–506. doi:10.4049/jimmunol.1500558
- Stritesky GL, Jameson SC, Hogquist KA. 2012. Selection of self-reactive T cells in the thymus. *Annu Rev Immunol* **30**: 95–114. doi:10.1146/annurev-immunol-020711-075035
- Takaba H, Morishita Y, Tomofuji Y, Danks L, Nitta T, Komatsu N, Kodama T, Takayanagi H. 2015. Fezf2 orchestrates a thymic program of self-antigen expression for immune tolerance. *Cell* **163**: 975–987. doi:10.1016/j.cell.2015.10.013
- Tomofuji Y, Takaba H, Suzuki HI, Benlaribi R, Martinez CDP, Abe Y, Morishita Y, Okamura T, Taguchi A, Kodama T, et al. 2020. Chd4 choreographs self-antigen expression for central immune tolerance. *Nat Immunol* **21**: 892–901. doi:10.1038/s41590-020-0717-2
- Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**: 802–813. doi:10.1261/rna.876308
- Webster NJG. 2017. Alternative RNA splicing in the pathogenesis of liver disease. *Front Endocrinol (Lausanne)* **8**: 133. doi:10.3389/fendo.2017.00133
- Weyn-Vanhenyryck SM, Mele A, Yan Q, Sun S, Farny N, Zhang Z, Xue C, Herre M, Silver PA, Zhang MQ, et al. 2014. HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep* **6**: 1139–1152. doi:10.1016/j.celrep.2014.02.005
- Winqvist O, Karlsson FA, Kämpe O. 1992. 21-Hydroxylase, a major autoantigen in idiopathic Addison's disease. *Lancet* **339**: 1559–1562. doi:10.1016/0140-6736(92)91829-W
- Yang Y, Liu XK, Nguyen T, Bishop C, Graf D, Dong C. 2007. Characterization of B7S3 as a novel negative regulator of T cells. *J Immunol* **178**: 3661–3667. doi:10.4049/jimmunol.178.6.3661
- Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Giron CG, et al. 2018. Ensembl 2018. *Nucleic Acids Res* **46**: D754–D761. doi:10.1093/nar/gkx1098
- Zuklys S, Gill J, Keller MP, Hauri-Hohl M, Zhanybekova S, Balciunaite G, Na KJ, Jeker LT, Hafen K, Tsukamoto N, et al. 2009. Stabilized β -catenin in thymic epithelial cells blocks thymus development and function. *J Immunol* **182**: 2997–3007. doi:10.4049/jimmunol.0713723
- Žuklys S, Handel A, Zhanybekova S, Govani F, Keller M, Maio S, Mayer CE, Teh HY, Hafen K, Gallone G, et al. 2016. Foxn1 regulates key target genes essential for T cell development in postnatal thymic epithelial cells. *Nat Immunol* **17**: 1206–1215. doi:10.1038/ni.3537

Received January 11, 2021; accepted in revised form August 17, 2021.