The transcriptional landscape of mTEC^{hi} and mTEC^{lo}.

Shahan Mamoor
shahanmamoor@gmail.com
San Diego, CA 92101

Medullary thymic epithelial cells, or mTEC, are cells of the thymus that can be sorted and classified based on expression of the class II major histocompatibility complex, MHC-II (1-3). mTEC^{hi} and mTEC^{lo} can be further segregated by expression of the CD80 marker, but there are few systematic analyses of the unique transcriptional behavior of each mTEC cell subset (4-6). We performed global differential gene expression profiling by comparing the transcriptomes of mTEC^{hi} and mTEC^{lo} (7) to determine the most significant transcriptional differences between these two cell subsets of the thymus. We found nearly a dozen groups of gene that distinctly identify these two cell types from each other. These included phospholipase-type enzymes, transcription factors, transcriptional coactivators and epigenetic proteins, cell signaling intermediates, cell surface receptors, molecules involved in ubiquitination, taste receptors, cathepsins, and interleukin-13. mTEC^{hi} and mTEC^{lo} can be discerned with facility as discrete cell types independent of MHC-II and CD80 expression through systematic comparative transcriptional profiling and the molecular descriptions provided here can be used as a resource for future investigations into the organ primarily responsible for providing lymphocyte self-tolerance instruction.

Keywords: thymus, mTEC^{hi}, mTEC^{lo}, differential gene expression analysis, global comparative transcriptional profiling, mTEC transcriptome

Introduction

The thymus is anatomically segregated into the medulla and cortex, with the epithelial cells of the medulla (8) named medullary thymic epithelial cells, or mTEC (9). Expression of the class II major histocompatibility complex antigen receptor MHC-II can be used to sub-classify mTEC into two cell subsets known as mTEC^{hi} and mTEC^{lo} (1,2). CD80 is a reliable marker to independently segregate these two cell types independent of MHC-II expression (10, 11). Aside from CD80 expression, which provides little detail regarding the cell-specific functions and transcriptional behavior of these cell types, there are few available resources to describe the transcriptional landscape of mTEC^{hi} and mTEC^{lo} and how they differ from one another (3-6).

We performed global differential gene expression profiling by comparing the transcriptomes of mTEC^{hi} and mTEC^{lo} using a public microarray dataset (7). This analysis identified a set of molecules whose expression could uniquely identify each cell type by its differential expression. Together, this set of over 50 differentially expressed genes, binned into over a dozen categories, represents an advance in understanding of the transcriptional landscape of each cell mTEC^{hi} and mTEC^{lo} and presents novel insights into the transcriptional behavior of mTEC^{hi} and mTEC^{lo}, which among others, includes the differential expression of taste receptors.

Methods

GEO2R was utilized for this differential gene expression analysis along with dataset GSE89965. This public dataset was previously generated by J. Abramsom (7) using microarray (Affymetrix Mouse Gene 1-ST array) analysis of Aire-GFP Epcam+MHC-II(hi) cells (mTEC^{hi}), and Aire-GFP Epcam+MHC-II(lo) cells (mTEC^{lo}); $n=4$ from mTEC^{hi} and mTEC^{lo}; from C57Bl/6 Adig mice 6 weeks of age (adult). There was no adjustment made to p -values, and no log-transformation was applied to the data. The NCBI generated category of platform annotation

was used. An unpaired two-tailed t-test with Welch's correction was utilized to determine the statistical significance of the RNA expression values between each cell type.

Results

We compared the transcriptomes of mTEC^{hi} and mTEC^{lo} to determine which messages were most different in expression level between each cell type globally using differentially gene expression analysis of a public dataset (7). Out of 34760 total transcripts detected and measured by the microarray dataset utilized for this study, the top 250 transcripts most differentially expressed between mTEC^{hi} and mTEC^{lo} were curated for further analysis. Of these, 54 transcripts (Table 1) were binned into nearly a dozen groups, consisting of transcription factors, epigenetic and nuclear proteins, taste receptors, phospholipases, zinc-finger containing proteins, signaling molecules and intermediates, ubiquitin-related molecules, cathepsins, interleukin and interleukin receptors, innate immune signaling molecules, cytochrome and neutrophilic factors, and DNase/RNases.

A series of transcription factors that distinguish mTEC^{hi} from mTEC^{lo}

Utf, undifferentiated embryonic cell transcription factor 1, was the 9th most differentially expressed gene out of 34760 total transcripts detected by the microarray dataset analyzed here (Table 1; $p=3.77E-07$). The expression of Utf1 was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 1; $p=0.0001$).

Irf5, interferon responsive factor 5, was the 10th most differentially expressed gene out of 34760 total transcripts detected by the microarray dataset analyzed here (Table 1; $p=6.75E-07$). The expression of Irf5 was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 1; $p=0.0002$).

Sox7, SRY-box 7, was the 57th most differentially expressed gene out of 34760 total transcripts detected and measured (Table 1; $p=2.5E-05$). The expression of Sox7 was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 1; $p<0.0001$).

Mkx, mohawk homeobox was the 57th most differentially expressed gene out of 34760 total transcripts detected and measured (Table 1; $p=3.14E-05$). The expression of Mkx was significantly higher in mTEC^{lo} than in mTEC^{hi} (Figure 1; $p=0.0002$).

Tcf7l1, or transcription factor 7 like 1 (T cell specific, HMG box), was the 167th most differentially expressed gene out of 34760 total transcripts detected and measured (Table 1; $p=1.22E-04$). The expression of Tcf7l1 was significantly higher in mTEC^{lo} than in mTEC^{hi} (Figure 1; $p=0.0002$).

Tfap2b, or transcription factor AP-2 beta, was the 177th most differentially expressed gene out of 34760 total transcripts detected and measured (Table 1; $p=1.35E-04$). The expression of Tfap2b was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 1; $p=0.0048$).

Gtf2a1, or general transcription factor II A, 1, was the 215th most differentially expressed gene out of 34760 total transcripts detected (Table 1; $p=1.83E-04$). The expression of Gtf2a1 was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 1; $p=0.0017$).

Tbx21, or T-box 21, the murine gene encoding for the T-bet transcription factor required for Th1 cell development and interferon gamma expression by these cells (12), was the 220th most differentially expressed gene when comparing mTEC^{hi} and mTEC^{lo} transcription at the global level (Table 1; $p=1.92E-04$). Tbx21 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 1; $p=.0033$).

Finally, HoxA2, or homeobox A2, was the 225th most differentially expressed gene when comparing mTEC^{hi} and mTEC^{lo} transcriptomes (Table 1; $p=1.97E-04$). The expression of HoxA2 was significantly higher in mTEC^{lo} than in mTEC^{hi} (Figure 1; $p<0.0001$).

Epigenetic molecules, transcriptional coactivators and nuclear DNA-binding molecules that distinguish mTEC^{hi} from mTEC^{lo}

Sirtuin1, or Sirt1 was the 21st most differentially expressed gene globally when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes (Table 1; $p=4.01E-06$). Sirt1 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 2; $p=0.0002$).

Mediator complex subunit 12-like, or Med12l was the 49th most differentially expressed gene globally when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes (Table 1; $p=1.67E-05$). Med12l expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 2; $p<0.0001$).

Lysine (K)-specific demethylase 2B, or Kdm2b was the 139th most differentially expressed gene globally when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes (Table 1; $p=8.42E-05$). Kdm2b expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 2; $p=0.0004$).

Myeloid/lymphoid or mixed-lineage leukemia; translocated to 3, or Mllt3 was the 156th most differentially expressed gene globally when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes (Table 1; $p=1.07E-04$). Mllt3 expression was significantly higher in mTEC^{lo} than in mTEC^{hi} (Figure 2; $p=0.0004$).

Recombination signal binding protein for immunoglobulin kappa J region, or Rbpj, was the 191st most differentially expressed gene globally when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes (Table 1; $p=1.55E-04$). Rbpj expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 2; $p=0.0007$).

Zinc finger, MIZ-type containing 2, or Zmiz2, was the 205th most differentially expressed gene globally when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes (Table 1; $p=1.69E-04$). Zmiz2 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 2; $p=0.0024$).

Chromobox 2, or Cbx2, a polycomb group factor, was the 243rd most differentially expressed gene globally when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes (Table 1; $p=2.38E-04$). Cbx2 expression was significantly higher in mTEC^{lo} than in mTEC^{hi} (Figure 2; $p=0.0019$).

mTEC^{hi} express high levels of IL-13 RNA and IL2 receptor gamma chain RNA relative to mTEC^{lo}

The gamma chain of the interleukin 2 receptor, *Il2rg*, was the 14th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=1.33E-06$). *Il2rg* expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 3; $p<0.0001$).

Interleukin 13, *Il13*, was the 28th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=6.74E-06$). *Il13* expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 3; $p=0.0009$).

Tgfbeta, Notch ligand Dll4, Wnt4 and Bmp2 kinase expression identify mTEC^{hi} from mTEC^{lo}

Wingless-type MMTV integration site family, member 4, *Wnt4*, was the 107th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=5.25E-05$). *Wnt4* expression was significantly higher in mTEC^{lo} than in mTEC^{hi} (Figure 4; $p=0.0010$).

Notch ligand delta-like 4, *Dll4*, was the 188th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=1.51E-04$). *Dll4* expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 4; $p=0.0017$).

Transforming growth factor beta, *Tgfb1*, was the 198th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=1.6E-04$). *Tgfb1* expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 4; $p=0.0004$).

Bmp2 inducible kinase, *Bmp2k*, was the 250th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=2.5E-04$). *Bmp2k* expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 4; $p=0.0012$).

Cell surface receptors that distinguish mTEC^{hi} from mTEC^{lo}: CD80 expression validates faithful segregation of mTEC^{hi} from mTEC^{lo}

CD80 antigen, CD80, was the 104th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=4.91E-05$). CD80 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 5; $p=0.0015$).

Platelet derived growth factor receptor, beta polypeptide, Pdgfrb, was the 127th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=7.16E-05$). Bmp2k expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 5; $p=0.0009$).

Icos ligand, Icosl, was the 129th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=7.24E-05$). Icosl expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 5; $p=0.0029$).

Integrin alpha L, Itgal, was the 217th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=1.86E-04$). Itgal expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 5; $p=0.0021$).

Selection and upkeep of intraepithelial T cells 4, Skint4, was the 230th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=2.06E-04$). Skint4 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 5; $p=0.0036$).

Scavenger receptor class F, member 2, Scarf2, was the 233rd most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=2.12E-04$). Scarf2 expression was significantly higher in mTEC^{lo} than in mTEC^{hi} (Figure 5; $p=0.0019$).

Phospholipase enzymes distinguish mTEC^{hi} from mTEC^{lo}

Phospholipase A1 member A, Pla1a, was the 5th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=1.48\text{E-}07$). Pla1a expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 6; $p<0.0001$).

Phospholipase B1, Plb1, was the 17th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=3.01\text{E-}06$). Plb1 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 6; $p<0.0001$).

Phospholipase A2, group IVA (cytosolic, calcium-dependent), Pla2g4a, was the 207th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=1.71\text{E-}04$). Pla2g4a expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 6; $p=0.0004$).

Caspase 1, Traf1, Gasdermin A3 and Oas1e mark mTEC^{hi}

Gsdma3 A, Gasdermin 3, was the 85th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=3.99\text{E-}05$). Gsdma3 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 7; $p=0.0020$).

2'-5' oligoadenylate synthetase 1E, Oas1e, was the 90th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=4.2\text{E-}05$). Oas1e expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 7; $p=0.0021$).

Casp1, or caspase 1, was the 223rd most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=1.95\text{E-}04$). Casp1 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 7; $p=0.0006$).

TNF receptor-associated factor 1, Traf1, was the 239th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=2.27\text{E-}04$). Traf1 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 7; $p=0.0007$).

Higher expression of Cathepsins S, H and Z identify mTEC^{hi}

Cathepsin Z 1, Ctsz, was the 31st most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=8.25E-06$). Ctsz expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 8; $p=0.0001$).

Cathepsin S, Ctss, was the 110th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=5.33E-05$). Ctss expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 8; $p=0.0011$).

Cathepsin H, Ctsh, was the 111th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=5.55E-05$). Ctsh expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 8; $p=0.0004$).

Ubiquitin D and the linear ubiquitin deubiquitinase Otulin expression distinguish mTEC^{hi} from mTEC^{lo}

Ubd, or ubiquitin D, was the 227th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=2.02E-04$). Ubd expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 9; $p=0.0057$).

OTU deubiquitinase with linear linkage specificity, Otulin, was the 172nd most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=1.29E-04$). Otulin expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 9; $p=0.0006$).

The expression of two taste receptors, Tas2r138 and Tas2r137, identify mTEC^{lo}

Taste receptor, type 2, member 137, Tas2r137, was the 87th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=4E-05$). Tas2r137 expression was significantly higher in mTEC^{lo} than in mTEC^{hi} (Figure 10; $p=0.0005$).

Taste receptor, type 2, member 138, Tas2r138, was the 210th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=1.78E-04$). Tas2r138 expression was significantly higher in mTEC^{lo} than in mTEC^{hi} (Figure 10; $p=0.0037$).

Dnase1l3, and to a lesser degree, Rnaseh2b expression identify mTEC^{hi}

Deoxyribonuclease 1-like 3, Dnase1l3, was the 106th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=5.15E-05$).

Dnase1l3 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 11; $p=0.0002$).

Ribonuclease H2, subunit B, Rnaseh2b, was the 240th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=2.29E-04$). Rnaseh2b expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 11; $p=0.0015$).

Neutrophil cytosolic factors and Ncf1 and Ncf4, and Cyp1a2 distinguish mTEC^{hi} from mTEC^{lo}

Cytochrome P450, family 1, subfamily a, polypeptide 2, or Cyp1a2, was the 7th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=1.92E-07$). Cyp1a2 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 12; $p<0.0001$).

Neutrophil cytosolic factor 4, Ncf4, was the 44th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=1.31E-05$). Ncf4 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 12; $p=0.0014$).

Neutrophil cytosolic factor 1, Ncf1, was the 155th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=1.05E-04$). Ncf1 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 12; $p=0.0041$).

Transcripts for three zinc finger containing proteins identify mTEChi from mTEClo

Pleiomorphic adenoma gene-like 1 Plagl1, was the 40th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=9.27E-06$). Plagl1 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 13; $p<0.0001$).

Zinc finger protein 318, Zfp318, was the 113th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=5.69E-05$). Zfp318 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 13; $p=0.0002$).

Zinc finger protein 710, Zfp710, was the 133rd most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=7.39E-05$). Zfp710 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 13; $p=0.0007$).

Bcl7a expression is significantly higher in mTEClo than in mTEChi

B cell CLL/lymphoma 7A, Bcl7a, was the 78th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=3.7E-05$). Bcl7a expression was significantly higher in mTEC^{lo} than in mTEC^{hi} (Figure 14; $p=0.0002$).

Wdr59 expression is significantly higher in mTEChi than in mTEClo

WD repeat domain 59, Wdr59, was the 237th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=2.23E-04$). Wdr59 expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 15; $p=0.0007$).

Ttc39a expression is significantly higher in mTEChi than in mTEClo

Tetratricopeptide repeat domain 39A, Ttc39a, was the 93rd most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=4.28E-05$). Ttc39a expression was significantly higher in mTEC^{hi} than in mTEC^{lo} (Figure 16; $p=0.0006$).

Iqce expression is significantly higher in mTEC^{lo} than in mTEC^{hi}

IQ motif containing E, Iqce, was the 114th most differentially expressed gene when comparing the mTEC^{hi} and mTEC^{lo} transcriptomes globally (Table 1; $p=5.75E-05$). Iqce expression was significantly higher in mTEC^{lo} than in mTEC^{hi} (Figure 17; $p=0.0002$).

Discussion

In this systems-level study of the medullary epithelial cells of the thymus, we compared the transcriptomes of mTEC^{hi} and mTEC^{lo} (7) to better understand, in an unbiased fashion, the most significant transcriptional differences between these cell types. As mTEC^{hi} and mTEC^{lo} each have thymic origin, epithelial origin, as well as some level of MHC-II expression in common, this systems-level transcriptional analysis should identify only the most specific differences in gene expression between these two cell types.

A recent study investigating tissue-restricted antigen (TRA) expression and the autoimmune regulator protein Aire utilized single-cell RNA-sequencing to cluster cells based on expression of TRA, and divided mTEC into mTEC^{hi} and mTEC^{lo} types by computational clustering after single-cell RNA-sequencing (4). However, their computational clustering divided cells of the thymus into mTEC^{hi} and mTEC^{lo} groups that did not significantly differ on CD80 expression; this is somewhat problematic as these cell types are defined by their difference in CD80 expression (13). Moreover, after computational clustering, the authors performed differential gene expression profiling as performed here, but only one of the genes Miragaia et al. (4) identified as differentially expressed between mTEC^{hi} and mTEC^{lo} (*Irf5*) were also identified in our study as differentially expression, out of 54 total.

Interestingly, we found that two taste receptors, *Tas2r137* and *Tas2r138* were differentially expressed by mTEC^{lo} and at significantly higher levels than in mTEC^{hi}. The function of these taste receptors in the thymus has not been investigated nor has their differential expression in mTEC been reported but the fact that these receptors were the 87th and 210th most differentially expressed genes out of over 34,000 transcripts detected in these

cell types suggests that they serve some important purpose in housekeeping or immunological functioning in the thymus.

We identified a series of differentially expressed transcription factors, 5 of which are expressed at significantly higher levels in mTEC^{hi} (Utf1, Tbx21, Irf5, Gtf2a1, and Tfap2b) and 4 of which are expressed at significantly higher levels in mTEC^{lo} (Tcf7l1, Mxk, Sox7, and HoxA2). It will be of note to understand how these transcription factors each individually and in concert prime the thymic genome to serve as the immunological stroma for developing lymphocytes, particularly in conjunction with and with respect to the function of the autoimmune regulator molecule Aire (2, 3, 6).

In conclusion, we identified a gene set of 54 molecules that are differentially expressed by either mTEC^{hi} or mTEC^{lo}, are the most significant differences in gene expression between these two cell types, and uniquely define their transcriptional identity.

References

1. Aichinger, M., Wu, C., Nedic, J. and Klein, L., 2013. Macroautophagy substrates are loaded onto MHC class II of medullary thymic epithelial cells for central tolerance. *Journal of Experimental Medicine*, 210(2), pp.287-300.
2. Chuprin, A., Avin, A., Goldfarb, Y., Herzig, Y., Levi, B., Jacob, A., Sela, A., Katz, S., Grossman, M., Guyon, C. and Rathaus, M., 2015. The deacetylase Sirt1 is an essential regulator of Aire-mediated induction of central immunological tolerance. *Nature immunology*, 16(7), p.737.
3. Herzig, Y., Nevo, S., Bornstein, C., Brezis, M.R., Ben-Hur, S., Shkedy, A., Eisenberg-Bord, M., Levi, B., Delacher, M., Goldfarb, Y. and David, E., 2017. Transcriptional programs that control expression of the autoimmune regulator gene Aire. *Nature immunology*, 18(2), p.161.
4. Miragaia, R.J., Zhang, X., Gomes, T., Svensson, V., Ilicic, T., Henriksson, J., Kar, G. and Lönnberg, T., 2018. Single-cell RNA-sequencing resolves self-antigen expression during mTEC development. *Scientific reports*, 8(1), p.685.
5. Brennecke, P., Reyes, A., Pinto, S., Rattay, K., Nguyen, M., Küchler, R., Huber, W., Kyewski, B. and Steinmetz, L.M., 2015. Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. *Nature immunology*, 16(9), p.933.
6. Meredith, M., Zemmour, D., Mathis, D. and Benoist, C., 2015. Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nature immunology*, 16(9), p.942.
7. Jakub Abramson, Immunology, Weizmann Institute; PRJNA354356; GEO: GSE89965.
8. Rodewald, H.R., Paul, S., Haller, C., Bluethmann, H. and Blum, C., 2001. Thymus medulla consisting of epithelial islets each derived from a single progenitor. *Nature*, 414(6865), p.763.
9. Derbinski, J., Schulte, A., Kyewski, B. and Klein, L., 2001. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nature immunology*, 2(11), p.1032.
10. Nedic, J., Aichinger, M., Emmerich, J., Mizushima, N. and Klein, L., 2008. Autophagy in thymic epithelium shapes the T-cell repertoire and is essential for tolerance. *Nature*, 455(7211), p.396.
11. Takahama, Y., Ohigashi, I., Baik, S. and Anderson, G., 2017. Generation of diversity in thymic epithelial cells. *Nature Reviews Immunology*, 17(5), p.295.
12. Szabo, S.J., Sullivan, B.M., Stemmann, C., Satoskar, A.R., Sleckman, B.P. and Glimcher, L.H., 2002. Distinct effects of T-bet in TH1 lineage commitment and IFN- γ production in CD4 and CD8 T cells. *Science*, 295(5553), pp. 338-342.
13. Ki, S., Park, D., Selden, H.J., Seita, J., Chung, H., Kim, J., Iyer, V.R. and Ehrlich, L.I., 2014. Global transcriptional profiling reveals distinct functions of thymic stromal subsets and age-related changes during thymic involution. *Cell reports*, 9(1), pp.402-415.

Figure 1: A series of transcription factors that distinguish $mTEC^{hi}$ from $mTEC^{lo}$

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing $mTEC^{hi}$ ($n=4$) and red dots representing $mTEC^{lo}$ ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 2: Epigenetic molecules and transcriptional coactivators that distinguish $mTEC^{hi}$ from $mTEC^{lo}$

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing $mTEC^{hi}$ ($n=4$) and red dots representing $mTEC^{lo}$ ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 3: mTEC^{hi} express high levels of IL-13 RNA and IL2 receptor gamma chain RNA relative to mTEC^{lo}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 4: Tgfbeta, Notch ligand Dll4, Wnt4 and Bmp2 kinase expression identify mTEC^{hi} from mTEC^{lo}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 5: Cell surface receptors that distinguish $mTEC^{hi}$ from $mTEC^{lo}$: CD80 expression validates faithful segregation of $mTEC^{hi}$ from $mTEC^{lo}$

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing $mTEC^{hi}$ (n=4) and red dots representing $mTEC^{lo}$ (n=4). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p-value denoted on each graph).

Figure 6: Phospholipase enzymes distinguish mTEC^{hi} from mTEC^{lo}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 7: Caspase 1, Traf1, Gasdermin A3 and Oas1e mark mTEC^{hi}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 8: Higher expression of Cathepsins S, H and Z identify mTEC^{hi}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 9: Ubiquitin D and the linear ubiquitin deubiquitinase Otulin expression distinguish mTEC^{hi} from mTEC^{lo}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 10: The expression of two taste receptors, Tas2r138 and Tas2r137, identify mTEC^{lo}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 11: Dnase1l3, and to a lesser degree, Rnaseh2b expression identify mTEC^{hi}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 12: Neutrophil cytosolic factors and Ncf1 and Ncf4, and Cyp1a2 distinguish mTEC^{hi} from mTEC^{lo}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 13: Transcripts for three zinc finger containing proteins identify mTEC^{hi} from mTEC^{lo}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 14: Bcl7a expression is significantly higher in mTEC^{lo} than in mTEC^{hi}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 15: Wdr59 expression is significantly higher in mTEC^{hi} than in mTEC^{lo}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 16: Ttc39a expression is significantly higher in mTEC^{hi} than in mTEC^{lo}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Figure 17: Iqce expression is significantly higher in mTEC^{lo} than in mTEC^{hi}

The RNA expression value of each differentially expressed transcript is graphically depicted, with blue dots representing mTEC^{hi} ($n=4$) and red dots representing mTEC^{lo} ($n=4$). The horizontal black bar denotes the mean expression value, and a two-tailed unpaired t-test with Welch's correction was performed to determine the significance of difference in mean RNA expression for each transcript between cell types (p -value denoted on each graph).

Table 1

Rank	p-value	t	B	Gene Symbol	Gene Name	ID
5	1.48E-07	-22.3542567	-2.82	Pla1a	phospholipase A1 member A	10439442
7	1.92E-07	-21.4842237	-2.83	Cyp1a2	cytochrome P450, family 1, subfamily a, polypeptide 2	10593981
9	3.77E-07	-19.409431	-2.83	Utf1	undifferentiated embryonic cell transcription factor 1	10558580
10	6.75E-07	-17.7698876	-2.84	Irf5	interferon regulatory factor 5	10536898
14	1.33E-06	-16.0368524	-2.86	Il2rg	interleukin 2 receptor, gamma chain	10606016
17	3.01E-06	-14.1462255	-2.88	Plb1	phospholipase B1	10520923
21	4.01E-06	-13.5411983	-2.88	Sirt1	sirtuin 1	10369738
28	6.74E-06	-12.4981379	-2.9	Il13	interleukin 13	10385837
31	8.25E-06	-12.11361	-2.91	Ctsz	cathepsin Z	10490212
40	9.27E-06	-11.8965526	-2.91	Plagl1	pleiomorphic adenoma gene-like 1	10361771
44	1.31E-05	-11.277258	-2.92	Ncf4	neutrophil cytosolic factor 4	10425053
49	1.67E-05	-10.8501038	-2.93	Med12l	mediator complex subunit 12-like	10492231
57	2.5E-05	10.1858395	-2.95	Sox7	SRY (sex determining region Y)-box 7	10415885
71	3.14E-05	9.8280163	-2.96	Mkx	mohawk homeobox	10457323
78	3.7E-05	9.5780231	-2.97	Bcl7a	B cell CLL/lymphoma 7A	10525542
85	3.99E-05	-9.4630522	-2.97	Gsdma3	gasdermin A3	10380956
87	4E-05	9.4603031	-2.97	Tas2r137	taste receptor, type 2, member 137	10537499
90	4.2E-05	-9.3872724	-2.98	Oas1e	2'-5' oligoadenylate synthetase 1E	10533229
93	4.28E-05	-9.3570595	-2.98	Ttc39a	tetratricopeptide repeat domain 39A	10506968
104	4.91E-05	-9.1565013	-2.99	Cd80	CD80 antigen	10435712
106	5.15E-05	-9.0860679	-2.99	Dnase1l3	deoxyribonuclease 1-like 3	10417526
107	5.25E-05	9.0598734	-2.99	Wnt4	wingless-type MMTV integration site family, member 4	10509267
110	5.33E-05	-9.0367713	-2.99	Ctss	cathepsin S	10494271
111	5.55E-05	-8.979505	-2.99	Ctsh	cathepsin H	10587733
113	5.69E-05	-8.9438182	-2.99	Zfp318	zinc finger protein 318	10445531
114	5.75E-05	8.9270507	-2.99	Iqce	IQ motif containing E	10535247

Table 1: A series of genes differentially expressed between mTEC^{hi} and mTEC^{lo} that distinguish each cell type independently of MHC-II and CD80

Rank order with respect to differential expression globally, p-value, t (a moderated t-statistic), B, the log-odds that the gene is differentially expressed between each group, gene symbol, gene name and transcript ID are listed (all provided by NCBI GEO2R).

Table 1-1

Rank	p-value	t	B	Gene Symbol	Gene Name	ID
127	7.16E-05	-8.6192677	-3.01	Pdgfrb	platelet derived growth factor receptor, beta polypeptide	10456046
129	7.24E-05	-8.6038945	-3.01	Icosl	icos ligand	10364361
133	7.39E-05	-8.5754526	-3.01	Zfp710	zinc finger protein 710	10554370
139	8.42E-05	-8.3975648	-3.02	Kdm2b	lysine (K)-specific demethylase 2B	10533569
155	1.05E-04	-8.0973496	-3.03	Ncf1	neutrophil cytosolic factor 1	10534202
156	1.07E-04	8.0743458	-3.04	Mllt3	myeloid/lymphoid or mixed-lineage leukemia; translocated to, 3	10514255
167	1.22E-04	7.907121	-3.04	Tcf7l1	transcription factor 7 like 1 (T cell specific, HMG box)	10545458
172	1.29E-04	-7.8340636	-3.05	Otulin	OTU deubiquitinase with linear linkage specificity	10427910
177	1.35E-04	-7.7748779	-3.05	Tfap2b	transcription factor AP-2 beta	10345016
188	1.51E-04	-7.6393579	-3.06	Dll4	delta-like 4 (Drosophila)	10474958
191	1.55E-04	-7.6051234	-3.06	Rbpj	recombination signal binding protein for immunoglobulin kappa J region	10521913
198	1.6E-04	-7.5609766	-3.07	Tgfb1	transforming growth factor, beta 1	10551185
205	1.69E-04	-7.4983985	-3.07	Zmiz2	zinc finger, MIZ-type containing 2	10374151
207	1.71E-04	-7.4807231	-3.07	Pla2g4a	phospholipase A2, group IVA (cytosolic, calcium-dependent)	10358434
210	1.78E-04	7.4300185	-3.07	Tas2r138	taste receptor, type 2, member 138	10544284
215	1.83E-04	-7.3997199	-3.08	Gtf2a1	general transcription factor II A, 1	10401882
217	1.86E-04	-7.3779506	-3.08	Itgal	integrin alpha L	10557591
220	1.92E-04	-7.3419876	-3.08	Tbx21	T-box 21	10390328
223	1.95E-04	-7.3239131	-3.08	Casp1	caspase 1	10582985
225	1.97E-04	7.3127239	-3.08	Hoxa2	homeobox A2	10544751
227	2.02E-04	-7.2790611	-3.08	Ubd	ubiquitin D	10445112
230	2.06E-04	-7.2568171	-3.09	Skint4	selection and upkeep of intraepithelial T cells 4	10507062
233	2.12E-04	7.2211595	-3.09	Scarf2	scavenger receptor class F, member 2	10434105
237	2.23E-04	-7.1604392	-3.09	Wdr59	WD repeat domain 59	10581835
239	2.27E-04	-7.138576	-3.09	Traf1	TNF receptor-associated factor 1	10481949
240	2.29E-04	-7.1282053	-3.1	Rnaseh2b	ribonuclease H2, subunit B	10415791
243	2.38E-04	7.0854954	-3.1	Cbx2	chromobox 2	10383055
250	2.5E-04	-7.0245873	-3.1	Bmp2k	BMP2 inducible kinase	10523468

Table 1: A series of genes differentially expressed between mTEC^{hi} and mTEC^{lo} that distinguish each cell type independently of MHC-II and CD80

Rank order with respect to differential expression globally, p-value, t (a moderated t-statistic), B, the log-odds that the gene is differentially expressed between each group, gene symbol, gene name and transcript ID are listed (all provided by NCBI GEO2R).