Comparative transcriptional profiling of medullary and cortical thymic cell types identifies epigenetic and transcription factors that uniquely define the mTEC^{lo} and cTEC lineages

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

The thymus is a key site of immunological activity, with a medulla and cortex. Epithelial cells within the medulla and cortex play important roles in these immunological processes, including positive and negative selection. In this analysis we compared the transcriptomes of mTEC^{lo} and cTEC (5). We present a series of epigenetic molecules and transcription factors that uniquely define mTEC^{lo} and cTEC, representing the major transcriptional differences between these two cell types of the thymus.

Keywords: thymus, medullary epithelial cells, cortical epithelial cells, mTEC^{lo}, cTEC, Kmt2d, Kmt2a, Kdm4a, Phf20, Phf20l1, L3mbtl3, Mafk, Nr2f2, Zbtb18, Zfp422, autoimmunity, transplant medicine

Introduction

The thymus is subdivided into the medulla and cortex (1, 2), and medullary epithelial cells can be subdivided based on their higher or lower expression of the class II major histocompatibility complex molecules MHC-II (6, 7). mTEC^{hi} and mTEC^{lo} function adjacent to the cortical epithelial cells of the thymus known as cTEC (8). In this study we performed global differential gene expression profiling of mTEC^{lo} and cTEC by comparing their transcriptomes using a public dataset (5). This analysis identified a series of transcription factors and epigenetic molecules that uniquely identify either cell type and represent the most significant transcriptional differences between mTEC^{lo} and cTEC.

Methods

Dataset GSE89965 (5) was utilized for this differential gene expression analysis, performed using GEO2R. This public dataset was generated by investigators (3) using microarray (Affymetrix Mouse Gene 1-ST array) analysis of Aire-GFP, EpCAM+Ly51+ cells (cTEC), and Aire-GFP Epcam+MHC-II(lo) cells (mTEC^{lo}); $n=4$ from mTEC^{lo}; $n=3$ from cTEC, from C57Bl/6 Adig mice 6 weeks of age (adult). No adjustment was made to p -values, and there was no log-transformation 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 performed global differential gene expression profiling of the mTEC^{lo} and cTEC transcriptomes (5) to understand in a systematic fashion the transcriptional behavior of these immunologically important thymic cells types and to determine the most significant transcriptional changes between medullary and cortical thymic epithelium, cells required for the processes of positive and negative selection (1-4).

Kmt2d, Kmt2a, and Kdm4a expression define mTEC^{lo} from cTEC.

We found one lysine demethylase and two lysine specific methyltransferases whose expression was among the most differentially expressed in the mTEC^{lo} transcriptome when compared to the cTEC transcriptome, and epigenetic enzymes were expressed at significantly higher levels in mTEC^{lo}. These included Kmt2a, Kmt2d, and Kdm4a.

Kmt2a, lysine (K)-specific methyltransferase 2A, was the 33rd most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; $p=0.00020684$). Kmt2a expression was significantly higher in mTEC^{lo} than in cTEC (Figure 1; $p=0.0005$).

Kmt2d, lysine (K)-specific methyltransferase 2D, was the 77th most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; $p=0.00044992$). Kmt2d expression was significantly higher in mTEC^{lo} than in cTEC (Figure 1; $p=0.0014$).

Kdm4a, lysine (K)-specific demethylase 4A, was the 158th most differentially expressed gene when determining the genes whose expression was most different between the global gene expression profiles of mTEC^{lo} and cTEC (Table 1; $p=0.00102559$). Kdm4a was expressed at significantly higher levels in mTEC^{lo} than in cTEC (Figure 1; $p=0.0035$).

Phf20 and Phf20l1 expression define mTEC^{lo} from cTEC.

Interestingly, both PHD finger protein 20 (Phf20) and PHD finger protein 20-like 1 (Phf20l1) were among the genes whose expression was most different when comparing the transcriptomes of mTEC^{lo} and cTEC.

Phf20l1, PHD finger protein 20-like 1, was the 193rd most differentially expressed gene when determining the transcripts whose expression levels was most different between the global gene expression profiles of mTEC^{lo} and cTEC (Table 1; $p=0.00124597$). Phf20l1 was expressed at significantly higher levels in mTEC^{lo} than in cTEC (Figure 2; $p=0.0044$).

Phf20, PHD finger protein 20, was the 239th most differentially expressed gene when determining the transcripts whose expression levels was most different between the global gene expression profiles of mTEC^{lo} and cTEC (Table 1; $p=0.00156917$). Phf20 was expressed at significantly higher levels in mTEC^{lo} than in cTEC (Figure 2; $p=0.0023$)

A series of epigenetic factors and nuclear proteins whose differential expression uniquely define mTEC^{lo} from cTEC.

Six epigenetic or otherwise nuclear DNA-binding factors were among the most differentially expressed genes when comparing the global gene expression profiles of mTEC^{lo} and cTEC. All six were expressed at higher levels in mTEC^{lo}. These genes included Mbd6, Cep350, L3mbtl3, Taf4, Chd7, and Atrx.

Mbd6, methyl-CpG binding domain protein 6, was the 53rd most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; $p=0.00029807$). Mbd6 expression was significantly higher in mTEC^{lo} than in cTEC (Figure 3; $p=0.0198$).

Cep350, centrosomal protein 350, was the 92nd most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; $p=0.00055961$). Cep350 expression was significantly higher in mTEC^{lo} than in cTEC (Figure 3; $p=0.0025$).

L3mbtl3, l(3)mbt-like 3, was the 151st most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; $p=0.0009909$). L3mbtl3 expression was significantly higher in mTEC^{lo} than in cTEC (Figure 3; $p=0.0034$).

Taf4, TATA-box binding protein associated factor 4, was the 161st most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; $p=0.00104041$). Taf4 expression was significantly higher in mTEC^{lo} than in cTEC (Figure 3; $p=0.0022$).

Chd7, chromodomain helicase DNA binding protein 7, was the 169th most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1;

p=0.00110514). Chd7 expression was significantly higher in mTEC^{lo} than in cTEC (Figure 3; p=0.0004).

Atrx, alpha thalassemia/mental retardation syndrome X-linked, was the 245th most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; p=0.00020684). Atrx expression was significantly higher in mTEC^{lo} than in cTEC (Figure 3; p=0.0045).

A series of transcription factors whose differential expression uniquely define mTEC^{lo} from cTEC.

Six transcription factors were among the most differentially expressed genes when comparing the global gene expression profiles of mTEC^{lo} and cTEC. Of these, three were expressed at higher levels in cTEC, while three were expressed at higher levels in mTEC^{lo}. These genes included Zfp422, Nfatc2, Mafk, Nr2f2, and Zbtb18.

Zfp422, zinc finger protein 422, was the 15th most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; p=0.00006336). Zfp422 expression was significantly higher in mTEC^{lo} than in cTEC (Figure 4; p=0.0002).

Nfatc2, nuclear factor of activated T cells, cytoplasmic, calcineurin dependent 2, was the 83rd most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; p=0.0005108). Nfatc2 expression was significantly higher in mTEC^{lo} than in cTEC (Figure 4; p=0.0166).

Mafk, v-maf musculoaponeurotic fibrosarcoma oncogene family, protein K (avian), was the 143rd most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; p=0.00091144). Mafk expression was significantly higher in cTEC than in mTEC^{lo} (Figure 4; p=0.0147).

Nr2f2, nuclear receptor subfamily 2, group F, member 2, was the 191st most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; p=0.0012385). Nr2f2 expression was significantly higher in cTEC than in mTEC^{lo} (Figure 4; p=0.0021).

Zbtb18, zinc finger and BTB domain containing 18, was the 201st most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; $p=0.00129713$). Zbtb18 expression was significantly higher in cTEC than in mTEC^{lo} (Figure 4; $p=0.0030$).

Runx2, runt related transcription factor 2, was the 236th most differentially expressed gene globally when comparing the mTEC^{lo} and cTEC transcriptomes (Table 1; $p=0.00155685$). Runx2 expression was significantly higher in mTEC^{lo} than in cTEC (Figure 4; $p=0.0018$).

Discussion

In this study, we compared the transcriptomes of mTEC^{lo} and cTEC (5) by global differential gene expression profiling of a public microarray dataset to understand the most significant transcriptional changes between these two cell types.

We found the lysine demethylase Kdm4a, the lysine specific methyltransferases Kmt2a and Kmt2d, six transcription factors, including Zfp422, Nfatc2, Mafk, Nr2f2 and Zbtb18, six epigenetic and nuclear DNA-binding factors, including Mbd6, Cep350, L3mbtl3, Taf4, Chd7 and Atrx, and the PHD finger proteins Phf20 and Phf20l1 among the genes whose expression level was most different when comparing the total transcriptomes of mTEC^{lo} to cTEC.

Positive selection occurs in the cortex of the thymus, on the cell surface of cTEC (9-14). Understanding how to perturb these cell types at the level of gene expression, and how to specifically manipulate the most biologically relevant gene products (those whose expression is most unique in each respective cell type) is critical for the ability to manipulate these processes for the discovery of novel pathways and development of novel therapies for the treatment of autoimmune pathologies and of transplant rejection. It is critical to generate mouse models containing specific deletions of these differentially expressed genes in the thymus using B5t-Cre (15-17) to understand their function in thymic development and in homeostatic conditions.

We have performed a similar analysis by comparing the transcriptomes of cTEC and mTEC^{hi}, and together, these studies can serve as a resource for pathway discovery and

specific engineering of targeted therapeutics in autoimmunity and transplant medicine.

References

1. Rodewald, Hans-Reimer, Sabine Paul, Corinne Haller, Horst Bluethmann, and Carmen Blum. "Thymus medulla consisting of epithelial islets each derived from a single progenitor." *Nature* 414, no. 6865 (2001): 763.
2. Rossi, S.W., Jenkinson, W.E., Anderson, G. and Jenkinson, E.J., 2006. Clonal analysis reveals a common progenitor for thymic cortical and medullary epithelium. *Nature*, 441(7096), p.988.
3. Surh, C.D. and Sprent, J., 1994. T-cell apoptosis detected in situ during positive and negative selection in the thymus. *Nature*, 372(6501), p.100.
4. Blackman, M., Kappler, J. and Marrack, P., 1990. The role of the T cell receptor in positive and negative selection of developing T cells. *Science*, 248(4961), pp.1335-1341.
5. Jakub Abramson, Immunology, Weizmann Institute; PRJNA354356; GEO: GSE89965.
6. 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.
7. 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.
8. 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.
9. Laufer, T.M., DeKoning, J., Markowitz, J.S., Lo, D. and Glimcher, L.H., 1996. Unopposed positive selection and autoreactivity in mice expressing class II MHC only on thymic cortex. *Nature*, 383(6595), p.81.
10. Bill, J. and Palmer, E., 1989. Positive selection of CD4+ T cells mediated by MHC class II-bearing stromal cell in the thymic cortex. *Nature*, 341(6243), p.649.
11. Moran, A.E. and Hogquist, K.A., 2012. T-cell receptor affinity in thymic development. *Immunology*, 135(4), pp.261-267.
12. Klein, L., Hinterberger, M., Wirnsberger, G. and Kyewski, B., 2009. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nature Reviews Immunology*, 9(12), p.833.
13. Stefanski, H.E., Mayerova, D., Jameson, S.C. and Hogquist, K.A., 2001. A low affinity TCR ligand restores positive selection of CD8+ T cells in vivo. *The Journal of Immunology*, 166(11), pp.6602-6607.
14. von Boehmer, H., 1994. Positive selection of lymphocytes. *Cell*, 76(2), pp.219-228.

15. Murata, S., Sasaki, K., Kishimoto, T., Niwa, S.I., Hayashi, H., Takahama, Y. and Tanaka, K., 2007. Regulation of CD8+ T cell development by thymus-specific proteasomes. *Science*, 316(5829), pp.1349-1353.
16. Ohigashi, I., Zuklys, S., Sakata, M., Mayer, C.E., Zhanybekova, S., Murata, S., Tanaka, K., Holländer, G.A. and Takahama, Y., 2013. Aire-expressing thymic medullary epithelial cells originate from $\beta 5t$ -expressing progenitor cells. *Proceedings of the National Academy of Sciences*, 110(24), pp.9885-9890.
17. Ohigashi, I., Zuklys, S., Sakata, M., Mayer, C.E., Hamazaki, Y., Minato, N., Hollander, G.A. and Takahama, Y., 2015. Adult thymic medullary epithelium is maintained and regenerated by lineage-restricted cells rather than bipotent progenitors. *Cell reports*, 13(7), pp. 1432-1443.

Figure 1: Kmt2d, Kmt2a, and Kdm4a expression define mTEC^{lo} from cTEC.

The expression of each differentially expressed gene is represented graphically, with the mean RNA expression value denoted by a horizontal black bar: mTEC^{lo}, $n=4$; cTEC, $n=3$. The statistical significance of the difference between the mean values between each cell type was calculated using a two-tailed unpaired t-test with Welch's correction, shown as a p -value on each graph.

Figure 2: Phf20 and Phf20l1 expression define mTEC^{lo} from cTEC.

The expression of each differentially expressed gene is represented graphically, with the mean RNA expression value denoted by a horizontal black bar: mTEC^{lo}, n=4; cTEC, n=3. The statistical significance of the difference between the mean values between each cell type was calculated using a two-tailed unpaired t-test with Welch's correction, shown as a *p*-value on each graph.

Figure 3: A series of epigenetic factors and nuclear proteins whose differential expression uniquely define mTEC^{lo} from cTEC.

The expression of each differentially expressed gene is represented graphically, with the mean RNA expression value denoted by a horizontal black bar: mTEC^{lo}, n=4; cTEC, n=3. The statistical significance of the difference between the mean values between each cell type was calculated using a two-tailed unpaired t-test with Welch's correction, shown as a *p*-value on each graph.

Figure 4: A series of transcription factors whose differential expression uniquely define $mTEC^{lo}$ from $cTEC$.

The expression of each differentially expressed gene is represented graphically, with the mean RNA expression value denoted by a horizontal black bar: $mTEC^{lo}$, $n=4$; $cTEC$, $n=3$. The statistical significance of the difference between the mean values between each cell type was calculated using a two-tailed unpaired t-test with Welch's correction, shown as a p -value on each graph.

Table 1

Rank	p-value	t	B	Gene Symbol	Gene Name	ID
15	0.00006336	-10.1970	-3.71	Zfp422	zinc finger protein 422	10449989
33	0.00020684	-8.2167	-3.75	Kmt2a	lysine (K)-specific methyltransferase 2A	10592942
53	0.00029807	-7.6774	-3.76	Mbd6	methyl-CpG binding domain protein 6	10373143
77	0.00044992	-7.1058	-3.78	Kmt2d	lysine (K)-specific methyltransferase 2D	10432298
83	0.0005108	-6.9368	-3.78	Nfatc2	nuclear factor of activated T cells, cytoplasmic, calcineurin dependent 2	10489961
92	0.00055961	-6.8174	-3.79	Cep350	centrosomal protein 350	10359086
143	0.00091144	6.2066	-3.81	Mafk	v-maf musculoaponeurotic fibrosarcoma oncogene family, protein K (avian)	10526961
151	0.0009909	-6.1063	-3.82	L3mbtl3	l(3)mbt-like 3	10368380
158	0.00102559	-6.0654	-3.82	Kdm4a	lysine (K)-specific demethylase 4A	10515590
161	0.00104041	-6.0485	-3.82	Taf4	TATA-box binding protein associated factor 4	10490352
169	0.00110514	-5.9774	-3.82	Chd7	chromodomain helicase DNA binding protein 7	10503180
191	0.0012385	5.8449	-3.83	Nr2f2	nuclear receptor subfamily 2, group F, member 2	10564527
193	0.00124597	-5.8380	-3.83	Phf20l1	PHD finger protein 20-like 1	10424485
201	0.00129713	-5.7918	-3.83	Zbtb18	zinc finger and BTB domain containing 18	10352092
236	0.00155685	5.5857	-3.84	Runx2	runt related transcription factor 2	10451061
239	0.00156917	-5.5769	-3.84	Phf20	PHD finger protein 20	10477815
245	0.00163129	-5.5338	-3.85	Atrx	alpha thalassemia/mental retardation syndrome X-linked	10606263

Table 1: Transcription factors, epigenetic molecules, DNA and RNA binding factors that are among the most differentially expressed genes when comparing the transcriptomes of mTEC^{lo} and cTEC.

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 ID are listed (all provided by NCBI GEO2R).