Transcriptome sequencing of neonatal thymic epithelial cells

Charles St-Pierre1,2*, Sylvie Brochu1,2*, Juan Ruiz Vanegas1,2, Maude Dumont-Lagacé1,2, Sébastien Lemieux1,3 & Claude Perreault1,2

1Institute for Research in Immunology and Cancer, 2Department of Medicine, 3Department of Computer Science and Operations Research, Université de Montréal, Montreal, QC, Canada H3C 3J7.

In order to gain novel insights into thymus biology, we analysed the whole transcriptome of cortical and medullary thymic epithelial cells (cTECs and mTECs) and of skin epithelial cells (ECs). Consistent with their ability to express ectopic genes, mTECs expressed more genes than other cell populations. Out of a total of 15,069 genes expressed in TECs, 25% were differentially expressed by at least 5-fold in cTECs vs. mTECs. Genes expressed at higher levels in cTECs than mTECs regulate numerous cell functions including cell differentiation, cell movement and microtubule dynamics. Many positive regulators of the cell cycle were overexpressed in skin ECs relative to TECs. Our RNA-seq data provide novel systems-level insights into the transcriptional landscape of TECs, highlight substantial divergences in the transcriptome of TEC subsets and suggest that cell cycle progression is differentially regulated in TECs and skin ECs.

Received 21 March 2013
Accepted 2 May 2013
Published 17 May 2013

Correspondence and requests for materials should be addressed to C.P. (claude.perreault@umontreal.ca)

* These authors contributed equally to this work.

In all vertebrates, the thymus is necessary and sufficient for production of classic adaptive T cells1,2. There are no thymus substitutes in the animal kingdom and T cells generated extrathympically (e.g., in oncostatin M transgenic mice) are poorly functional: they cause severe autoimmunity and are unable to eliminate pathogens2–5. The key components of the thymus are cortical and medullary thymic epithelial cells (cTECs and mTECs) which play several essential functions6. During their intrathymic journey, which takes around 3 weeks, thymocytes undergo numerous reciprocal interactions with TECs located in seven functional zones7–9. TECs produce chemokines that attract bone marrow derived hematopoietic progenitors, as well as interleukin (IL)-7 and the notch ligand DLL4 that induce thymocyte proliferation and differentiation10,11. Furthermore, cTECs and mTECs express unique sets of ligands that mould the repertoire of antigen receptors expressed by thymocytes6,12,13. The cells that induce the positive selection of thymocytes are primarily cTECs, whereas mTECs are instrumental in negative selection.

Recent studies have highlighted several factors that regulate TEC development, maintenance and function including microRNAs, the transcription factor Foxn1 and Wnt signaling14–18. Nonetheless, despite the capital role of TECs, our understanding of TEC biology is quite rudimentary. For instance, it is not yet known whether cTECs and mTECs are maintained by unipotent or bipotent progenitors in postnatal thymi, and what might be the extent of divergence in the functional program of these two TEC populations6,7. In addition, while TECs display considerable proliferative potential19, it remains unclear why the number of TECs decreases rapidly with age, thereby leading to progressive thymic insufficiency5,20,21.

The transcriptome is a critical component of systems-level understanding of cell biology and it can be reliably tackled in its entirety in freshly harvested primary cells. In line with this, microarray analyses of several immune cell populations by the Immunological Genome Project Consortium have yielded fundamental insights into the biology of lymphocytes, dendritic cells and macrophages (http://www.immgen.org/index_content.html). As a first step to gain novel insights into TEC biology, we therefore decided to analyse the whole transcriptome of cTECs, mTECs and skin epithelial cells (ECs). We inferred that including skin ECs in our analyses would enable us to better appreciate the extent of divergence between the transcriptomes of mTECs and cTECs. In addition, we surmised that comparing the transcriptome of skin ECs vs. TECs might yield some clues as to why precocious age-related hypocellularity impinges on TECs but not skin ECs. We elected to analyse gene expression using RNA-seq rather than microarrays because RNA-seq has higher sensitivity and dynamic range coupled to lower technical variations22,23. We report that the transcriptomes of mTECs and cTECs present numerous substantial differences that may have far-reaching biological consequences. In addition, we found that many positive regulators of cell division are repressed in TECs relative to skin ECs. Our RNA-seq data offer a valuable resource to the community that can be mined to explore multiple questions.
Results

Purification of primary cell samples. The first steps of our work involved preparation of pure populations of primary (freshly harvested) cTECs, mTECs and skin ECs. Thymi from fourteen 7 day-old C57BL/6 mice (from 3 different litters) were harvested and stromal cells were enriched as described18,19. Thymic stromal cells were then stained with Ulex Europaeus Lectin 1 (UEA1) and antibodies against CD45, EpCAM and Ly51. Both cTEC (CD45 EpCAM+Ly51−UEA1+) and mTEC (CD45 EpCAM−Ly51+UEA1−) populations were sorted with a FACSAria cell sorter (Supplementary Figure S1). Epidermal cells were harvested from the trunk and dorsum of seven 2 day-old C57BL/6 (from 3 different litters) as described24 and stained with antibodies against EpCAM and CD11c. Skin ECs (EpCAM−CD11c−) were then sorted with a FACSAria cell sorter (Supplementary Figure S1). Post-sort analysis revealed that the purity of the sorted populations was more than 98% for TEC subsets and more than 95% for skin ECs (Supplementary Figure S1). In order to further validate the purity of sorted cell populations we first analysed in our RNA-Seq data the expression of 12 lineage-specific genes found in cTECs (Psmb11, Ly75, Ctsl, Prss16), mTECs (Aire, Cts, Cldn3, TNFrfs11a) and skin ECs (Lhx2, Lgals7, Hr, Krt1)6,13,25–29. The expression profile of these 12 genes was exactly as expected, thereby confirming that our cell populations were highly enriched and presented no substantial cross-contamination (Figure 1).

Overview of differential gene expression in cTECs, mTECs and skin ECs. We quantified transcript levels in reads per kilobase of exon model per million mapped reads (RPKM). The RPKM measure of read density reflects the molar concentration of a transcript in the starting sample by normalizing for RNA length and for the total read number in the measurement30. We found that skin ECs and cTECs expressed similar numbers of genes (circa 11,400; Figure 2a, b). However, consistent with the phenomenon of promiscuous gene expression in mTECs3, we found that mTECs expressed more genes (14,523) than other cell populations (Figure 2a, b). Expression of all transcripts in cTECs, mTECs and skin ECs is presented in Supplementary Table S1.

We used stringent criteria for identification of differentially expressed genes (DEGs). First, since inclusion of transcripts expressed at very low levels increases the risk of false discovery31, we considered only transcripts whose abundance was ≥ 2 RPKM. Second, genes were considered as DEGs only when the RPKM fold-difference between two cell populations was ≥ 5. Skin ECs and TECs diverge early during ontogeny: skin ECs originate from the ectoderm and TECs from the endoderm. Therefore, we were not surprised to see that the mean number of DEGs between skin ECs and TECs was 3,388: more specifically 2,386 and 4,391 in comparison of skin ECs with cTECs and mTECs, respectively (Figure 2c,e). Nonetheless, the number of DEGs between cTECs and mTECs was 3,820 (Figure 2d). Out of a total of 15,069 genes expressed in TECs (Figure 2b), 25% (3,820) were differentially expressed by at least 5-fold in cTECs and mTECs (Figure 2d). Genes overexpressed in mTECs (relative to cTECs and skin ECs) represented the largest category of DEGs. The peculiarity of mTECs was also evident when transcripts were simply categorized as present (≥2 RPKM) or absent, in the three cell populations.
types (Figure 2b). While 2,736 genes were expressed only in mTECs, 322 were uniquely found in cTECs and 297 in skin ECs. Two points can be made from these data. First, mTECs express more genes than other cell populations. When considering genes expressed only or mostly in mTECs (Figure 2b, d, e), mTECs differ roughly equally from cTECs and skin ECs. This can be explained by the unique ability of mTECs to transcribe a host of tissue-restricted genes. This process, termed “promiscuous gene expression”, is controlled in part by the Aire gene and is instrumental in establishing central tolerance13.

Second, even though the transcriptomes of cTECs and mTECs present substantial divergences, cTECs appear more similar to mTECs than to skin ECs. Indeed, only 4.7% (322 / 11,599) of expressed cTEC genes are unique relative to mTECs, while 12% (322 / 1064 / 11,599) of expressed cTEC genes are unique relative to skin ECs (Figure 2b).

The transcriptome of cTECs vs. mTECs. Analysis of genes expressed only in mTECs is particularly complex because they likely encompass two categories of genes: i) genes that play a genuine role in mTEC biology and ii) promiscuously expressed genes that are relevant for thymocyte negative selection but whose importance in mTEC biology per se is unknown. For this reason, we focused herein on transcripts overexpressed in cTECs and will analyse genes overexpressed in mTECs in a separate report. The 445 genes overexpressed in cTECs (Figure 2d) were scrutinised through the use of Ingenuity Pathway Analysis (IPA, Ingenuity® Systems, www.ingenuity.com). We used two metrics to identify the most important downstream effects of these 445 DEGs: activation z-score and p-value (Figure 3). A positive z-score indicates increased functional activity in cTECs relative to mTECs. The p-value, calculated with the Fischer’s exact test, indicates the likelihood that the association between a set of genes in our dataset and a biological function is significant.

Genes overexpressed in cTECs regulate 12 main functions connected to several cell processes (Figure 3). We analysed in more depth the three most robust functional differences (p-value ranging from $10^{-5}$ to $10^{-7}$); they involved cell differentiation, cell movement and microtubule dynamics. Using IPA, we subjected DEGs regulating the three aforementioned functions to IPA gene network analysis. IPA explores the set of input genes to generate networks by using Ingenuity Pathways Knowledge Base for interactions between
Figure 3 | Downstream effect analysis of genes overexpressed in cTECs relative to mTECs. We used the IPA regulation z-score algorithm to identify biological functions that are expected to be more active in cTECs than mTECs (positive z-score) according to our RNAseq data. In order to enhance the stringency of our analysis, we considered only functions with a z-score $\geq 2$ (indicated by an orange dotted line). The p-value (red dots), calculated with the Fischer's exact test, reflects the likelihood that the association between a set of genes in our dataset and a biological function is significant [p-value $\leq 0.05$ (i.e., $-\text{log}_{10} 0.05 = 1.3$)].

Figure 4 | IPA-based network of cTEC vs. mTEC DEGs involved in cell differentiation. The network displays interactions between cell differentiation related genes that were differentially expressed in cTECs vs. mTECs. Solid and dashed lines indicate direct and indirect interactions, respectively. Genes up-regulated in cTECs are coloured in shades of red; genes in white circles were not in our DEG dataset but were inserted by IPA because they are connected to this network. The activity of molecules highly connected to this network (hubs) was assessed with the IPA molecule activity predictor. Blue coloured molecules are predicted to be inhibited (or to have decreased activity) and orange coloured molecules to be activated (or to have increased activity) in cTECs relative to mTECs.
identified DEGs. Cell differentiation and cell movement networks were characterized by activation of RAC, PI3K and MAP kinases (particularly ERK1) culminating in the activation of the NF-κB complex in cTECs (Figure 4–5). Activation of NF-κB by RAC, PI3K and MAP kinases has been reported in several models\textsuperscript{32,33}. Of note, while NF-κB signals are essential for the development and function of mTECs\textsuperscript{34}, NF-κB activity was predicted to be higher in cTECs than mTECs (Figure 4–5). In addition to activation of the NF-κB complex, the cell movement network included two positive regulators of cell migration: HDAC and F-ACTIN (Figure 5)\textsuperscript{35,36}.

Microtubules are polymers of α–tubulins and γ-tubulins that contribute to most cellular functions including cell division, positioning of organelles, cell migration and cell polarity\textsuperscript{37}. In general, epithelial cells are highly polarized with an apical pole usually directed to the lumen of the tissue or the organ and a basal pole in contact with the underlying basal membrane\textsuperscript{38}. Microtubule dynamics is regulated by a network of proteins including microtubule-associated proteins, microtubule tip complex proteins and kinesins\textsuperscript{38}. In our microtubule dynamics network, the activity of three highly connected genes (hubs) was upregulated in cTECs relative to mTECs (Figure 6): Rac, Jnk and Akt. The Rho-like GTPase RAC and its effectors have been shown to participate in microtubule targeting of focal adhesions through interactions with microtubule-associated proteins and microtubule tip complex proteins\textsuperscript{38}. Microtubule-associated kinesin-1 carries JNK to allow its activation and microtubule elongation requires JNK activity throughout the microtubule life cycle\textsuperscript{39}. Akt localizes at the leading edge of migrating cells and enhances establishment of cell polarity and directed cell migration\textsuperscript{40}.

The three networks depicted in Figures 4–6 show that several genes which are crucial to early thymocytes development are expressed at higher levels in cTECs than mTECs. That is notably the case for Cxcl12, Dll4, Kitl and Wnt4 which regulate the recruitment of thymocyte progenitors, as well as the expansion and T-lineage commitment of immature thymocytes\textsuperscript{41,42,43,44}. In addition, components of several families of genes involved in cell-cell or cell-matrix interactions were overexpressed in cTECs: cadherin (Cdh4), ephrin receptor (Efnb1), fibronectin (Fn1), laminin (Lama4, Lamb1, Lamc1), integrin (Itgb1, Itgb2) and matrix metallo-enzymes (Mmp2, Mmp9).

The transcriptome of cTECs vs. skin ECs. We analysed genes differentially expressed in cTECs vs. skin ECs using IPA downstream analysis in order to identify biological functions affected by the 2,386 DEGs depicted in Figure 2c. Almost all functions predicted to be activated in cTECs (positive z-score) were related to the immune system: recruitment and proliferation of haematopoietic cells, movement and quality of leukocytes, etc (Figure 7a). Upregulation of these functions in cTECs is interesting.
and supports the validity of our approach, but was somewhat predictable. In contrast, most functions repressed in cTECs compared to skin ECs (negative z-score) were connected to the cell cycle and particularly to the S phase. To the best of our knowledge, differences in the proliferation potential of cTECs vs. skin ECs have not been reported previously. We therefore analysed more in depth cell cycle regulation in TECs and skin ECs. Based on a comprehensive literature review, we identified DEGs from our dataset that play a key role as positive regulators of the cell cycle. We found that almost all positive regulators in our list were upregulated in skin ECs compared to cTECs (Figure 7b). More specifically, we noted an upregulation by at least 5-fold of transcripts coding for the following proteins implicated in the G1/S phase transition, DNA synthesis and formation of the prereplication complexes:

- Cyclins D1 and D2, cyclin E1, cyclin-dependant-kinase-6, E2F transcription factors 1 and 2, and four minichromosome-maintenance proteins (22, 23, 25 and 27). Of note, the expression profile of mTECs was similar to that of cTECs (Figure 7b).

To gain further insights into the machinery regulating the cell cycle in cTECs and skin ECs, we generated a protein-protein interaction network including DEGs that were S phase-related according to IPA database (Figure 8). IPA network analysis points to a decrease functional activity of three key molecules in cTECs relative to skin ECs: cyclins, histone H3 and HDAC. Cyclins play a critical role in exit from cell quiescence and initiate DNA replication in response to mitogenic signals. Histone H3, and in particular its CENP-A variant which resides at centromeres, is fundamental to centromere function and chromosome segregation during mitosis. HDAC modulates cell proliferation via regulation of gene transcription and by preventing premature sister chromatid separation. At least two other S phase-related DEGs could impinge on the proliferation potential of EC populations. Indeed, compared with cTECs, skin ECs expressed higher levels of the epidermal growth factor receptor Erbb2 (9-fold) and the transcription factor Mycn (60-fold) than cTECs (Figure 8 and Supplementary Table S1). Both Erbb2 and Mycn are potent stimulators of cell proliferation.

**Discussion**

Our RNA-seq data provide novel insights into the transcriptional landscape of TECs and offer a valuable resource to the community.
Figure 7 | Downstream effect analysis of genes differentially expressed in cTECs vs. skin ECs. (a) IPA of DEGs was based on two metrics: z-score and p-value. The biological functions that are expected to be increased or decreased according to the gene expression changes in our dataset were identified using the IPA regulation z-score algorithm. A positive or negative z-score value indicates that a function is predicted to be increased or decreased in cTECs relative to skin ECs. In order to enhance the stringency of our analysis, we considered only functions with a z-score $\geq 2$ or $\leq -2$ (represented by orange and blue dotted lines). The p-value (red dots), calculated with the Fischer’s exact test, reflects the likelihood that the association between a set of genes in our dataset and a related biological function is significant [p-value $\leq 0.05$ (i.e., $-\log_{10} p \geq 1.3$)]. (b) Heat map representation depicting the expression of positive regulators of the cell cycle in cTECs, skin ECs and mTECs. The heat map was generated with the Multi Experiment Viewer v4.8 software.

Figure 8 | IPA-based network of cTEC vs. skin EC DEGs involved in S phase. The network displays interactions between genes regulating S phase that were differentially expressed in cTECs vs. skin ECs. Solid and dashed lines indicate direct and indirect interactions, respectively. Genes up-regulated in cTECs are coloured in shades of red; genes down-regulated in cTECs are coloured in shades of green. Genes in white circles were not in our DEG dataset but were inserted by IPA because they are connected to this network. The activity of molecules highly connected to this network (hubs) was assessed with the IPA molecule activity predictor. Blue coloured molecules are predicted to be inhibited (or to have decreased activity) and orange coloured molecules to be activated (or to have increased activity) in cTECs relative to skin ECs.
transcripts were detected for 1,686 genes. Deep sequencing of specific antigens. Corollary, MAGE peptides might not represent genuine tumor-specific expression program. Using a unique method combining microdissection, microarrays and computational analyses, Griffith et al. found that 2,162 stromal genes were expressed only in the medulla. However, their method did not allow to segregate genes expressed by TECs vs. other stromal cells. Consistent with these reports, we found mTECs expressed more transcripts than cTECs and skin ECs (Figure 2). Of particular relevance, our RNA-seq analyses revealed that out of 3,820 genes differentially expressed by at least 5-fold in mTECs vs. cTECs, 445 were overexpressed in cTECs while 3,375 were overexpressed in mTECs (Figure 2d). It has been shown that the bulk of promiscuously expressed genes are only turned on in CD80+ mTECs, and that stromal cells from different cortical and medullary subregions display unique gene expression signatures. We therefore anticipate that further RNA-seq analyses of discrete cTEC and mTEC subsets will reveal further complexity and heterogeneity in expression profiles.

We believe that further systems-level analysis of the thousands of genes expressed only in mTECs should be most useful to investigate the breadth and mechanisms of promiscuous gene expression in mTECs. For instance, it can be seen that genes of the MAGE-A and -B families (Magea1, a2, a3, a5, a6, a8, a10, b4 and b18) are absent in cTECs and skin ECs but are expressed (albeit at low levels) in mTECs (Supplementary Table S1). This observation is intriguing considering that MAGE-A and -B genes are receiving a lot of attention as targets for cancer immunotherapy because it is widely assumed that they code for tumor-specific antigens. These “cancer-germline” or “cancer-testis” genes are expressed in many tumors but not in normal tissues (with the exception of two tissues that do not express MHC class I molecules: placental trophoblast and testicular germ cells). That we detected Magea and Mageb transcripts in mTECs suggests that T-cells specific for MAGE peptides may be submitted to negative selection in the thymic medulla. As a corollary, MAGE peptides might not represent genuine tumor-specific antigens.

Overall, we detected transcripts from 20,977 genes in our mTECs: 14,523 genes had RPKM ≥ 2 and 6,454 were expressed at lower levels (0 < RPKM < 2) (Figure 2 and Supplementary Table S1). Hence, no transcripts were detected for 1,686 genes. Deep sequencing of transcriptomes from mTEC subsets (e.g., CD80hi vs. CD80lo mTECs) and thymic dendritic cell subsets would be needed to precisely estimate the repertoire of genes expressed in cells responsible for negative selection of thymocytes. Ideally, these studies should be complemented by mass spectrometry sequencing of MHC-associated peptides expressed on mTECs and thymic dendritic cells because gene expression at the transcript level does not always correlate with expression at the peptide level. That is particularly true for transcripts expressed at low levels since MHC-associated peptides derive preferentially from highly abundant mRNAs.

Nonetheless, numerous transcripts are more abundant in cTECs than mTECs (Figure 2d). Thus, the two key transcription factors that regulate thymus organogenesis and TEC differentiation (Hoxa3 and Fosn1) as well as genes required for attraction of thymic seeding thymocyte progenitors (Gli2, Cxcl12) and for expansion of early thymic progenitors (Dld4, Kitl, Il2 and Wnt4) are all expressed at higher levels in cTECs than mTECs (Supplementary Table S1). Our bioinformatic analyses suggest that genes overexpressed in cTECs regulate numerous cell functions (Figure 3), and in particular cell differentiation, cell movement and microtubule dynamics (Figures 4–6). We posit that the role of these genes in TEC biology should now be assessed in animal models involving TEC-specific gene deletion or overexpression. Priority might be given to genes such as G alc2, a component of the microtubule dynamics network (Figure 6) that is overexpressed by 35-fold in cTECs relative to mTECs (Supplementary Table 1) but whose role in thymus biology is unknown.

While the seminal work of Gray et al. revealed that TECs proliferate actively, their self renewal potential is seemingly inferior to that of other populations such as epidermal and intestinal ECs. Our comparison of the transcriptome of TECs and skin ECs suggests that one major handicap of TECs is that they express relatively low levels of positive regulators of the cell cycle such as cyclins and E2F transcription factors (Figure 7–8). This idea is consistent with studies showing that transgenic overexpression of E2F2 or of members of the cyclin D family increases TEC proliferation and thymic cellularity. However, proper interpretation of our findings is limited by our ignorance as regards the nature and properties of post-natal TEC progenitors and the strategy used for TEC self-renewal: asymmetric cell division or population asymmetry. Study of TEC self-renewal is further complicated by the fact that, in contrast to other types of tissue-specific progenitor and stem cells, we have no markers for TEC progenitors and TECs do not proliferate or form colonies in vitro. Unbiased exploration of a system using high-throughput methods such as transcriptome sequencing leads to generation of novel hypotheses. Accordingly, we believe that in-depth exploration of cell cycle regulation in TECs (Figure 7–8) may provide a conceptual framework for understanding the rules governing TEC homeostasis and self-renewal.

Methods

Mice. C57BL/6 mice purchased from The Jackson Laboratory (Bar Harbor) were bred and housed under specific-pathogen-free conditions in sterile ventilated racks at the Institute for Research in Immunology and Cancer. All procedures were in accordance with the Canadian Council on Animal Care guidelines and approved by the Comité de Déontologie et Experime ntation Animale de l’Université de Montréal.

Flow cytometry and cell sorting. Enrichment of thymic stromal cells was performed using the protocol of Gray et al. which yields adequate numbers of TECs and minimises changes in expression of cell surface markers. Thymic stromal cells from fourteen 7 day-old C57BL/6 mice were stained with biotinylated Ulex Europaeus Lectin 1 (UEA1; Vector Laboratories), PE-Cy7 conjugated streptavidin (BD Biosciences) and the following antibodies from BD Biosciences: Alexa700 anti-CD45, APC-Cy7 anti-EpCAM and Alexa647 anti-Ly51. The following antibodies were used: APC-Cy7 rat IgG2a (Biolegend) and Alexa647 rat IgG2a (AbD Serotec). Both cTEC (CD45 EpCAM+ Ly51+ UEA1+ ) and mTEC (CD45 EpCAM+ Ly51+ UEA1+ ) populations were sorted on a three laser FACSAria (BD Biosciences). Epidermal cells were harvested from the trunk and dorsum of 2 day-old C57BL/6 mice (n = 7) as described and stained with the following antibodies (BD Biosciences): APC-Cy7 anti-EpCAM and FITC anti-CD11c. EpCAM+ CD11c+ skin ECs were sorted on a three laser FACSAria.

Whole transcriptome sequencing (RNA-Seq). Two factors can influence gene expression profiling: technical and biological variation. Since RNA-seq data show much less technical variation than microarrays, multiple technical replicates are not needed for RNA-seq. In order to deal with subject-to-subject variation, we therefore used the pooling approach proposed by Kendziorzki et al. We pooled RNA from sorted cells harvested from newborn mice derived from 3 different litters: 14 mice for TECs and 7 mice for skin ECs. Total RNA was isolated using Trizol as recommended by the manufacturer (Invitrogen), and then further purified using RNeasy columns (Qiagen). Sample quality was assessed using Bioanalyzer RNA Nano Kitl, and fragmented and primed for cDNA synthesis. Reverse transcription of the first strand was done using random primers and SuperScript II (Invitrogen). A second round of reverse transcription was performed to generate a double-stranded cDNA, purified.
Sequencing. Paired-end (2 × 100 bp) sequencing was performed using the Illumina HiSeq2000 machine running TruSeq v3 chemistry. Cluster density was targeted at around 600–800 k clusters/mm². Two RNA-Seq libraries were sequenced per lane (8 lanes per slide). Detail of the Illumina sequencing technologies can be found at http://www.illumina.com/applications/sequencing.illumina. Briefly, DNA libraries are incorporated into a fluidic flow cell design with 8 individual lanes. The flow cell surface is populated with capture oligonucleotide anchors, which hybridize the appropriately modified DNA segments of a sequencing library generated from a genomic DNA sample. By a process called “bridge amplification,” captured DNA templates are amplified in the flow cell by “arching” over and hybridizing to an adjacent anchor oligonucleotide primer. Sequencing reaction is performed by hybridizing a primer complementary to the adapter sequence, then cyclically adding DNA polymerase and a mixture of 4 differently colored fluorescent reversible dye terminators to the captured DNA in the flow cell. By using this approach, nonmodiﬁed DNA fragments and unincorporated nucleotides are washed away, while captured DNA fragments are extended 1 nucleotide at a time. After each nucleotide-coupling cycle takes place, the flow cell is scanned, and digital images are acquired to record the locations of ﬂuorescently labeled nucleotide incorporations. Following imaging, the ﬂuorescent dye and the terminal 3’ blocker are chemically removed from the DNA before the next nucleotide coupling cycle.

Samples were prepared into 250 bp paired-end libraries and sequenced on the Illumina HiSeq1000 to obtain an average of 10.2 giga base pair of transcript sequences. Sequencing adapters were masked from the end of both sets of 100 bp reads prior to alignment. Data was mapped to the Mus musculus (mm9) reference genome using the ELANDv2 alignment tool from the CASAVA 1.8.2 software (Illumina). The frequency distribution depicted in Figure 2a was generated using the publicly available statistical software package ‘R’ (http://www.r-project.org/). RNA-Seq data have been deposited in GEO archives (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE44945.

Ingenuity pathway analysis. DEGs (RPKM ≥ 3; fold change ≥ 3) were tabulated and uploaded into the IPA software (Ingenuity Systems, http://www.ingenuity.com/). To determine the top biological functions associated with the observed gene expression proﬁles, we performed a downstream effect analysis from which we extracted the most drastically affected functions (p-value ≤ 0.05; z-score ≥ 2). The right-tailed Fisher’s exact test was used to estimate the probability that association between a set of molecules and a function or pathway might be due to random chance. The IPA z-score algorithm was used to predict the direction of change for a given function (increase or decrease). A z-score ≥ 2 means that a function is signiﬁcantly increased whereas a z-score ≤ 2 indicates a signiﬁcantly decreased function. We also generated protein-protein interaction networks to depict the DEGs which are involved in speciﬁc functional deregulations within cell types. The network view is an interactive representation that shows the molecular relationship between molecules from the dataset based on Ingenuity Knowledge Database.
Acknowledgements

CSP is supported by a studentship from the Natural Sciences and Engineering Research Council of Canada. CP holds a Canada Research Chair in Immunobiology. This work was supported by grant # MOP-42384 from the Canadian Institutes for Health Research (CP). The Institute for Research in Immunology and Cancer (IRIC) receives infrastructure support from the Canadian Center of Excellence in Commercialization and Research, the Canadian Foundation for Innovation, and the Fonds de la Recherche en Santé du Québec. The authors thank the personnel of the following IRIC core facilities for assistance: animal research facility, bioinformatics, flow cytometry and genomics. We are grateful to Pierre Chagnon, Patrice Gendron and Brian Wilhelm for advice and thoughtful comments on analysis of RNA-seq data.

Author contributions

C.S.P. and S.B. analysed the data, discussed results and prepared figures. M.D.L. and J.R.V. carried out experiments. S.L. analysed data and discussed results. C.P. designed the study, discussed results and wrote the first draft of the manuscript. All authors reviewed the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

License: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/

How to cite this article: St-Pierre, C. et al. Transcriptome sequencing of neonatal thymic epithelial cells. Sci. Rep. 3, 1860; DOI:10.1038/srep01860 (2013).