OPEN

Transcriptome sequencing of neonatal thymic epithelial cells

Charles St-Pierre[1,2]*, Sylvie Brochu[1,2]*, Juan Ruiz Vanegas[1,2], Maude Dumont-Lagacé[1,2], Sébastien Lemieux[1,3] & Claude Perreault[1,2]

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

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

* These authors contributed equally to this work.

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.

In all vertebrates, the thymus is necessary and sufficient for production of classic adaptive T cells[1,2]. There are no thymus substitutes in the animal kingdom and T cells generated extrathymically (e.g., in oncostatin M-transgenic mice) are poorly functional: they cause severe autoimmunity and are unable to eliminate pathogens[2–5]. The key components of the thymus are cortical and medullary thymic epithelial cells (cTECs and mTECs) which play several essential functions[6]. During their intrathymic journey, which takes around 3 weeks, thymocytes undergo numerous reciprocal interactions with TECs located in seven functional zones[7–9]. 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 differentiation[10,11]. Furthermore, cTECs and mTECs express unique sets of ligands that mould the repertoire of antigen receptors expressed by thymocytes[6,12,13]. 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 signaling[14–18]. 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 populations[6,7]. In addition, while TECs display considerable proliferative potential[19], it remains unclear why the number of TECs decreases rapidly with age, thereby leading to progressive thymic insufficiency[5,20,21].

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 variations[22,23]. 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 described[18,19]. 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 described[24] 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*, *Ctss*, *Cldn3*, *TNFrsf11a*) 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).
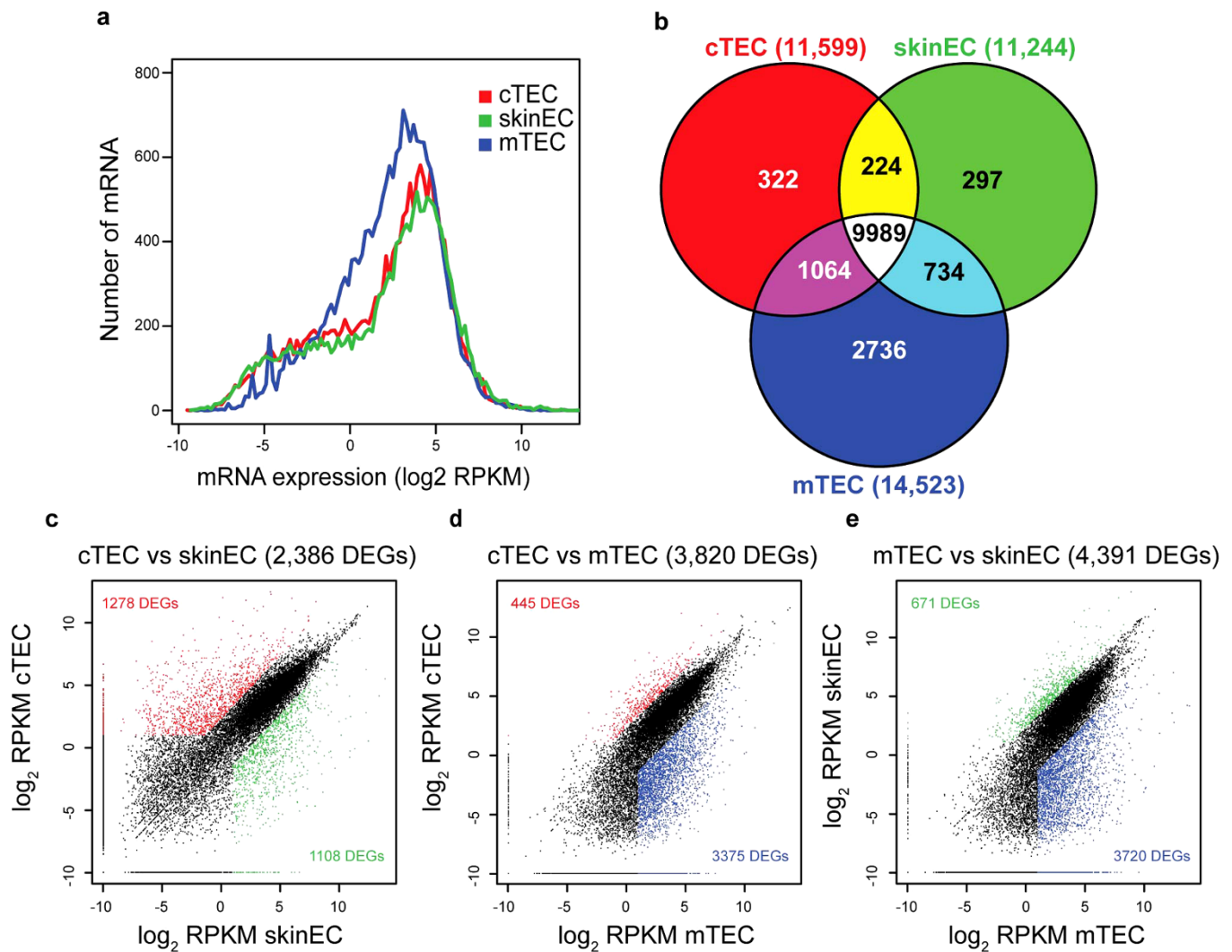
Overview of differential gene expression in cTECs, mTECs and skin ECs. We quantified transcript levels in reads per kilobase of exon model per million mapped reads (RPKM). The RPKM measure of read density reflects the molar concentration of a transcript in the starting sample by normalizing for RNA length and for the total read number in the measurement[30]. 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 mTECs[13], 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 discovery[31], 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
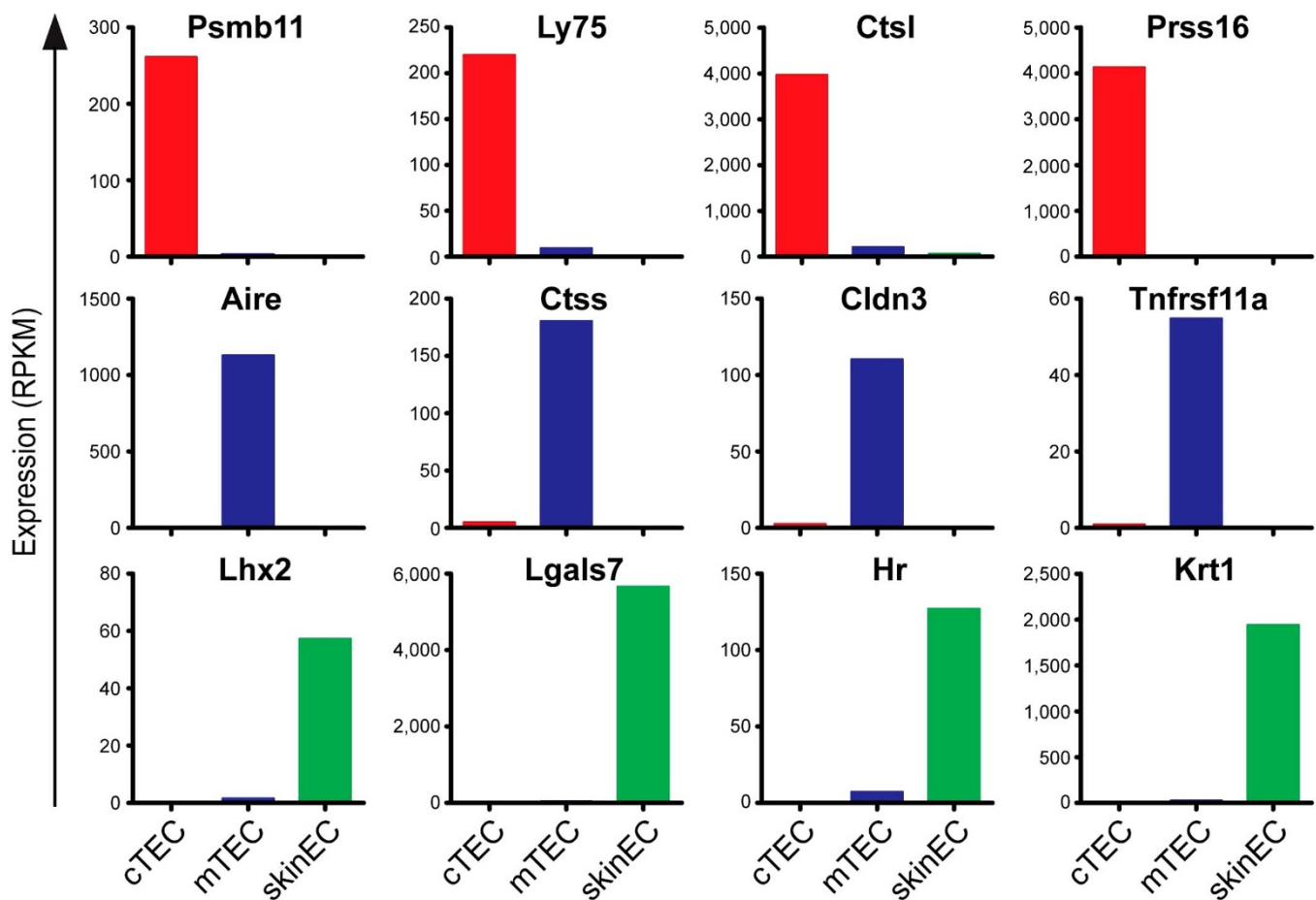
Figure 1 | Expression of lineage-specific genes in primary cTEC, mTEC and skin EC populations. We assessed the abundance of 12 transcripts in freshly sorted cTECs (red histograms), mTECs (blue histograms) and skin ECs (green histograms). These transcripts are known to be selectively expressed in cTECs (*Psmb11*, *Ly75*, *Ctsl*, *Prss16*), mTECs (*Aire*, *Ctss*, *Cldn3*, *TNFrfs11a*) or skin ECs (*Lhx2*, *Lgals7*, *Hr*, *Krt1*). Transcript abundance (in RPKM) was evaluated by RNA-Seq.

Figure 2 | Differential gene expression in cTECs, mTECs and skin ECs. (a) Frequency distribution of transcript expression levels. Gene expression levels (log2 RPKM) are plotted on the X-axis, whereas the Y-axis represents the frequency distribution of mRNAs calculated for 0.2 bin increments. The transcripts that were not detected by RNA-seq are not displayed in the representation. (b) Venn diagram showing the overlap and discrepancies between genes expressed (RPKM ≥ 2) in the three cell populations. Numbers in parentheses indicate the total number of genes expressed in each cell type. (c–e) Scatter plots representation of gene expression levels in (c) cTECs vs. skin ECs, (d) cTECs vs. mTECs and (e) mTECs vs. skin ECs. Color dots represent DEGs (RPKM ≥ 2 and fold-difference ≥ 5) which were overexpressed in cTECs (red), mTECs (blue) or skin ECs (green).

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 tolerance[13]. 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 + 224/11,599) 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
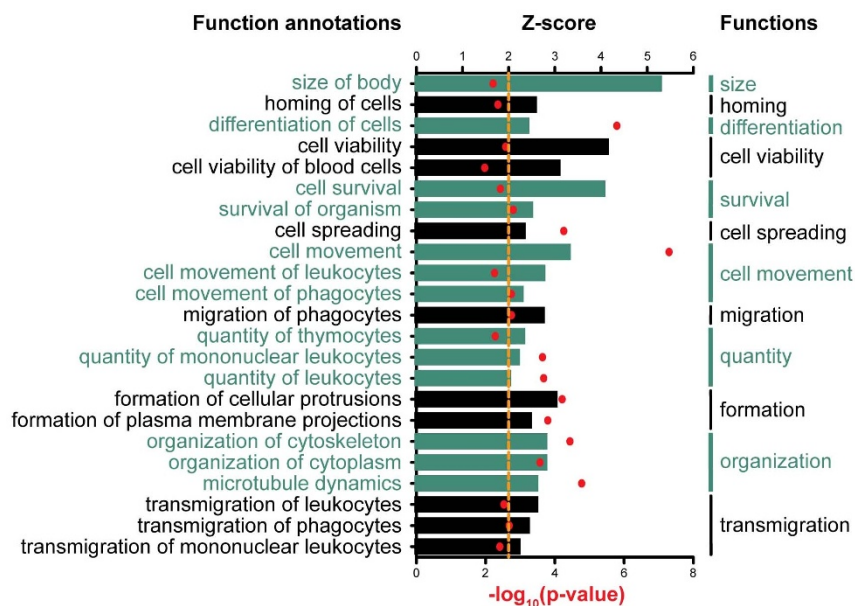
Figure 3 | Downstream effect analysis of genes overexpressed in cTECs relative to mTECs. We used the IPA regulation z-score algorithm to identify biological functions that are expected to be more active in cTECs than mTECs (positive z-score) according to our RNAseq data. In order to enhance the stringency of our analysis, we considered only functions with a z-score ≥ 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., $-\log_{10} \geq 1.3$)].
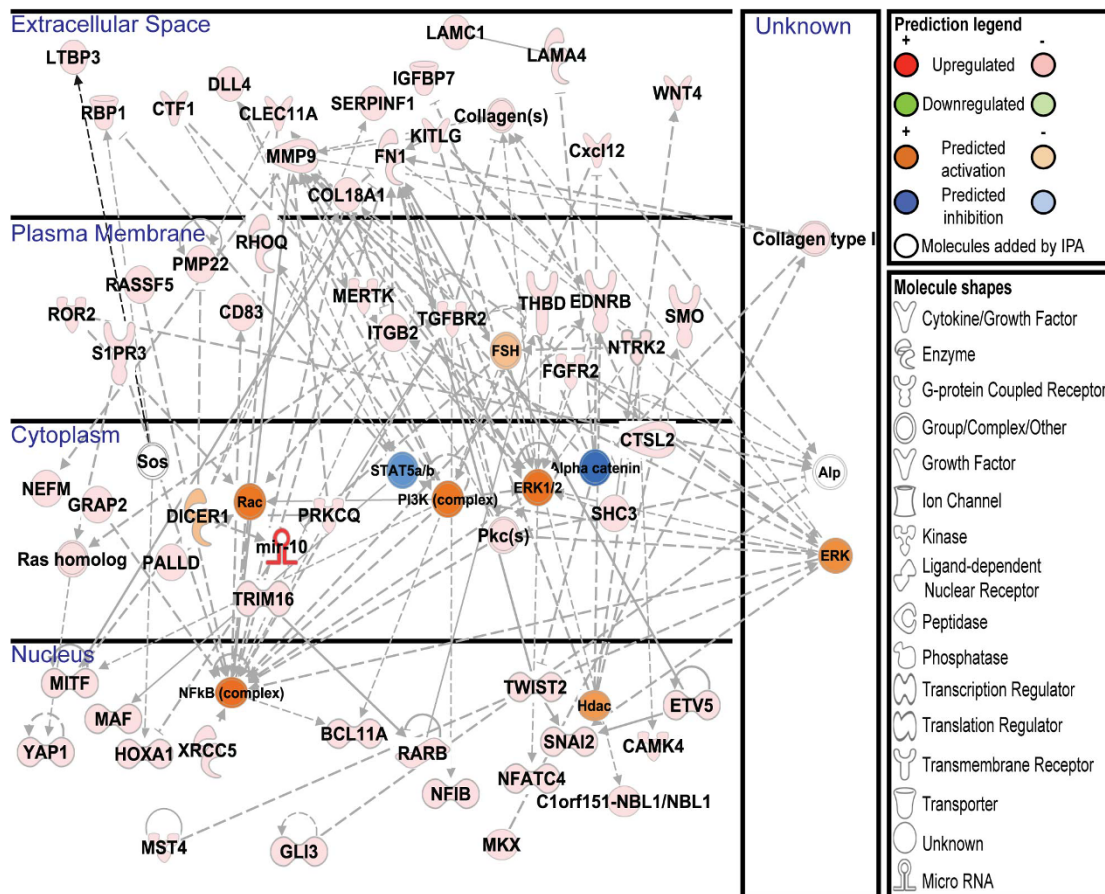
Figure 4 | IPA-based network of cTEC vs. mTEC DEGs involved in cell differentiation. The network displays interactions between cell differentiation related genes that were differentially expressed in cTECs vs. mTECs. Solid and dashed lines indicate direct and indirect interactions, respectively. Genes 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[32,33]. Of note, while NF-κB signals are essential for the development and function of mTECs[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)[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[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[38]. Microtubule dynamics is regulated by a network of proteins including microtubule-associated proteins, microtubule tip complex proteins and kinesins[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[38]. Microtubule-associated

kinesin-1 carries JNK to allow its activation and microtubule elongation requires JNK activity throughout the microtubule life cycle[39]. AKT localizes at the leading edge of migrating cells and enhances establishment of cell polarity and directed cell migration[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[8,10,11,41]. 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 metallopeptidase (*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
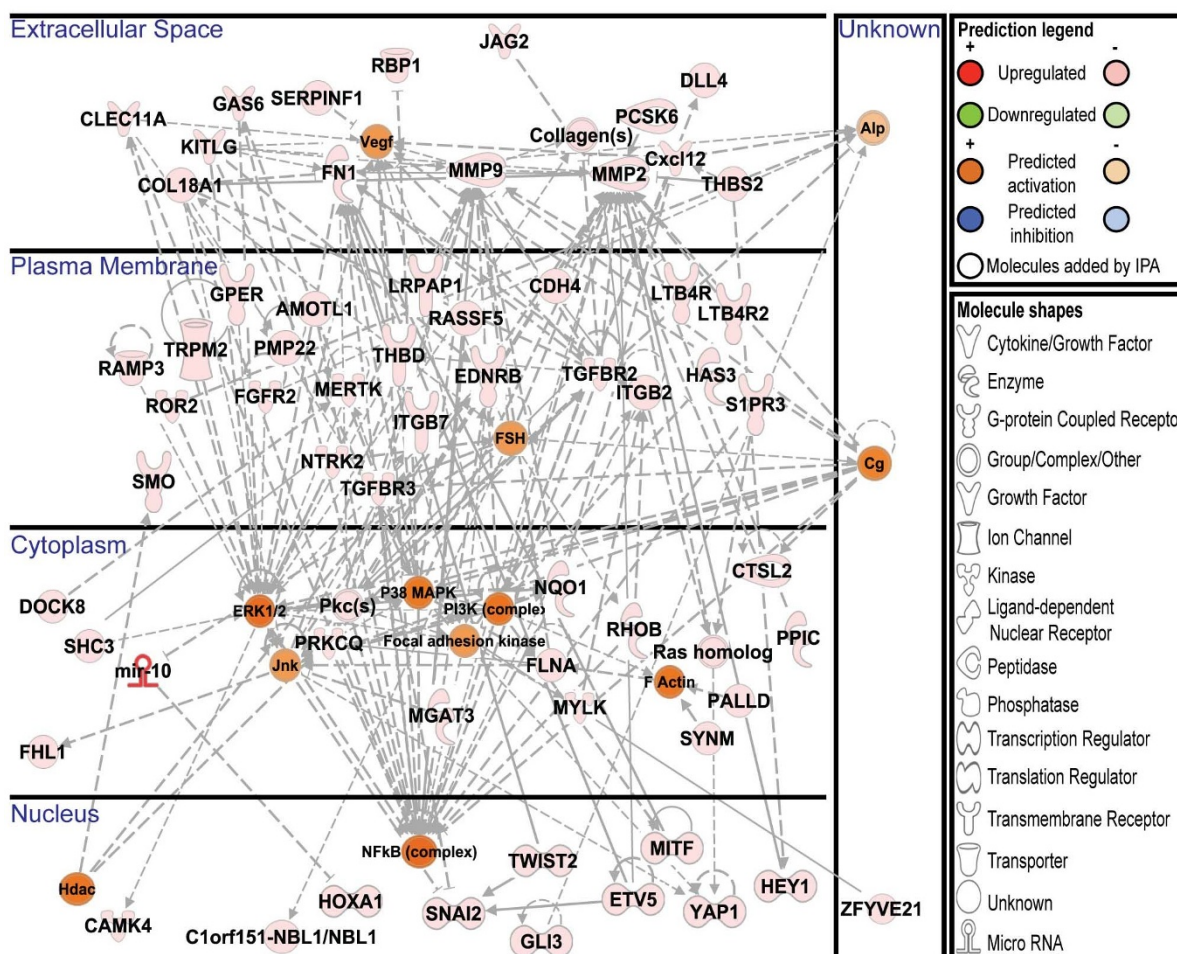
Figure 5 | IPA-based network of cTEC vs. mTEC DEGs involved in cell movement. The network displays interactions between cell movement related genes that were differentially expressed in cTECs vs. mTECs. Solid and dashed lines indicate direct and indirect interactions, respectively. Genes upregulated 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.
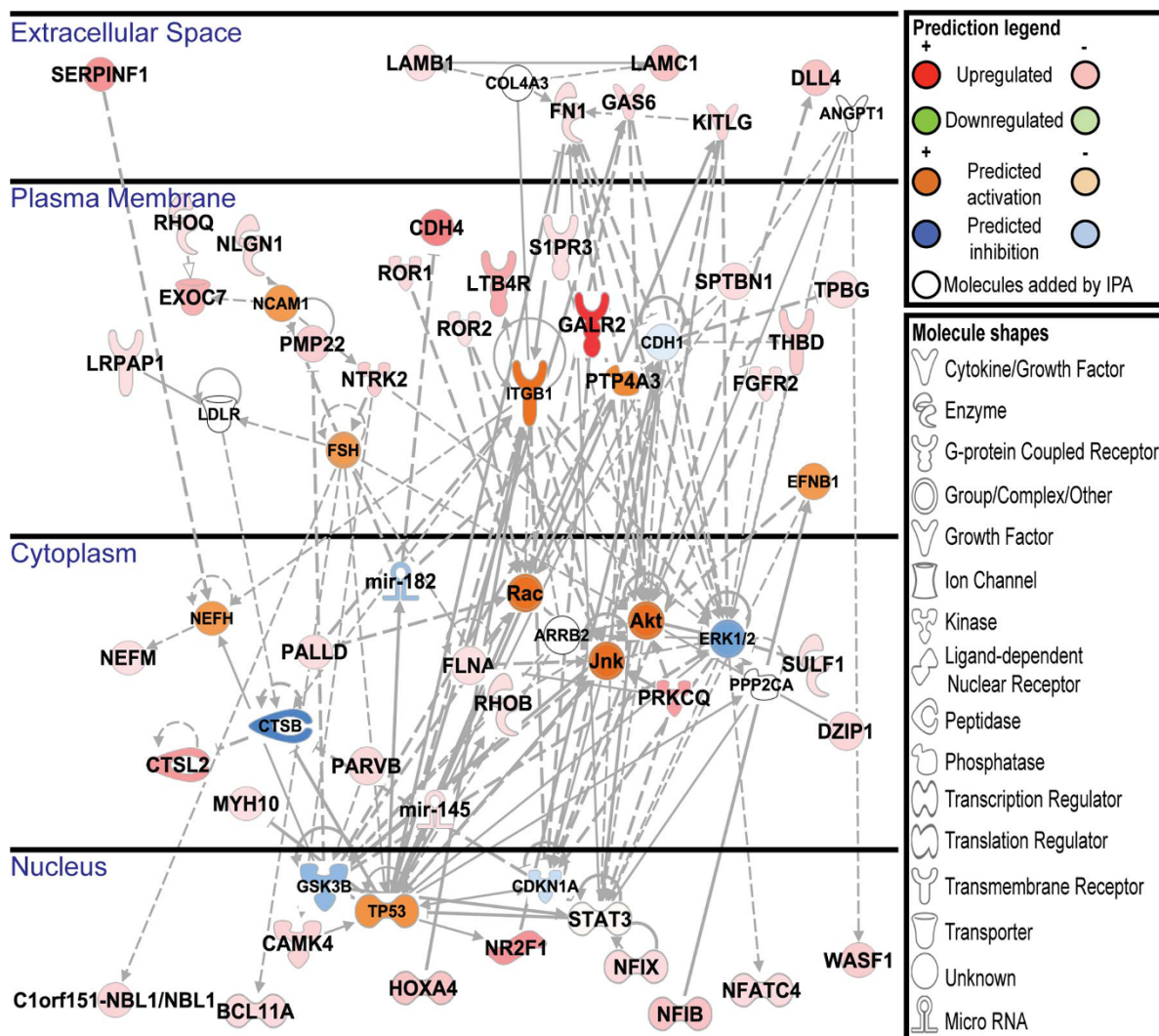
Figure 6 | IPA-based network of cTEC vs. mTEC DEGs involved in microtubule dynamics. The network displays interactions between genes regulating microtubule dynamics that were differentially expressed in cTECs vs. mTECs. Solid and dashed lines indicate direct and indirect interactions, respectively. Genes 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.

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[42]: cyclins D1 and D2, cyclin E1, cyclin-dependant-kinase-6, E2F transcription factors 1 and 2, and four minichromosome-maintenance proteins (−2, −3, −5 and −7). 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[43]. Histone H3, and in particular its CENP-A variant which resides at centromeres, is fundamental to centromere function and chromosome segregation during mitosis[44,45]. HDAC modulates cell proliferation via regulation of gene transcription and by preventing premature sister chromatid separation[46]. 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[47,48].

Discussion

Our RNA-seq data provide novel insights into the transcriptional landscape of TECs and offer a valuable resource to the community.
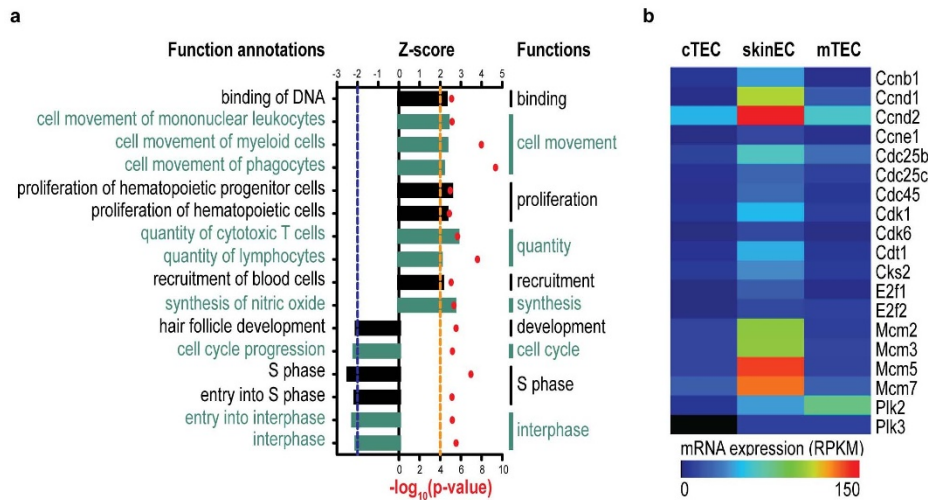
Figure 7 | Downstream effect analysis of genes differentially expressed in cTECs vs. skin ECs. (a) IPA of DEGs was based on two metrics: z-score and p-value. The biological functions that are expected to be increased or decreased according to the gene expression changes in our dataset were identified 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 ≥ 2 or ≤ −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 ≤ 0.05 (i.e., −log$_{10}$ ≥ 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[60].
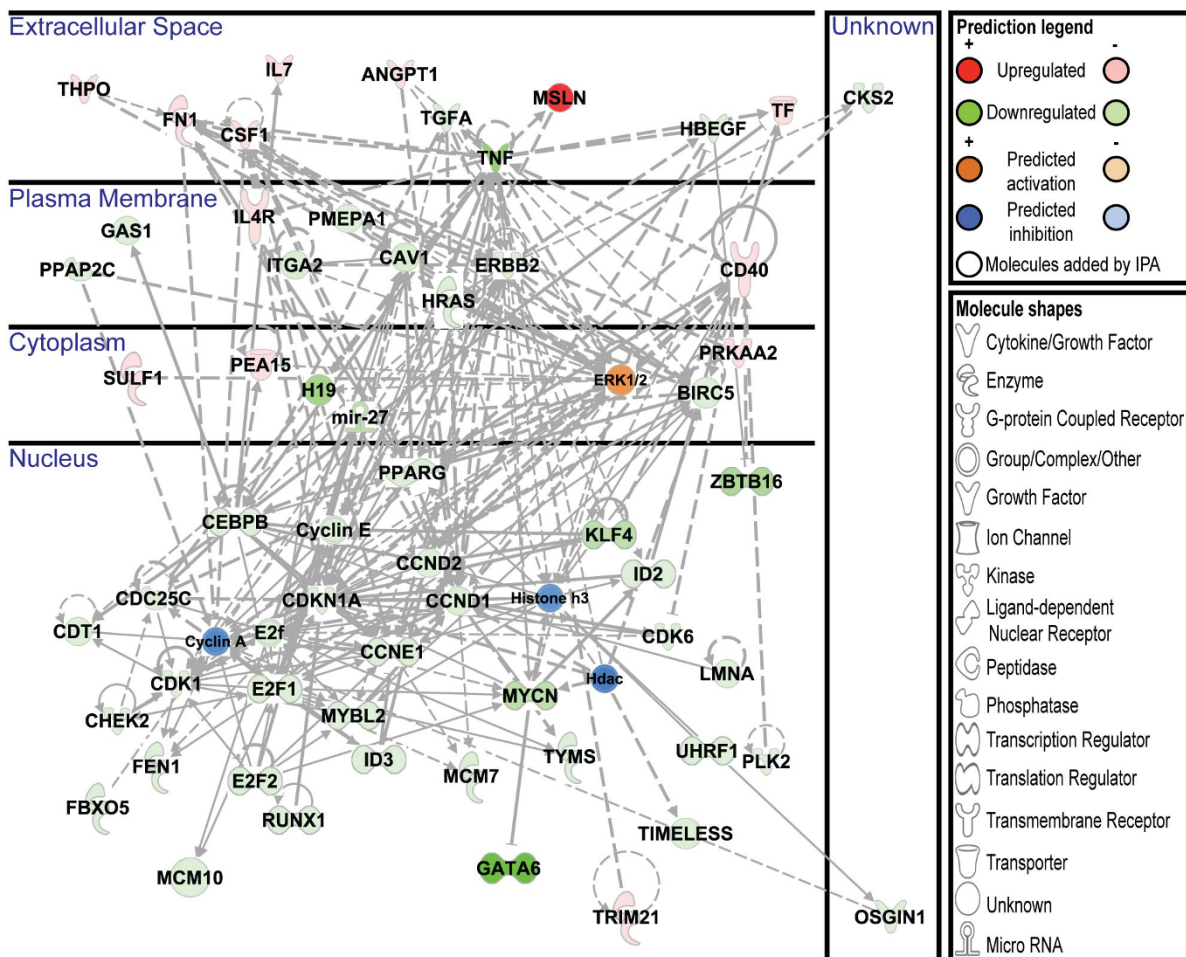
Figure 8 | IPA-based network of cTEC vs. skin EC DEGs involved in S phase. The network displays interactions between genes regulating S phase that were differentially expressed in cTECs vs. skin ECs. Solid and dashed lines indicate direct and indirect interactions, respectively. Genes up-regulated in 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.

RNA-seq data presented in Supplementary Table S1 furnish a platform for future studies on TEC biology as these datasets can be mined to explore multiple questions. We conclude from our data that the transcriptomes of mTECs and cTECs present numerous substantial differences and that mTECs express more genes than other cell types analysed herein (cTECs and skin ECs). Based on RT-PCR and microarray analyses, Derbinski et al. reported that numerous genes were overexpressed in mTECs relative to cTECs, macrophages and dendritic cells[49]. Notably, they estimated that the pool of genes overexpressed in mTECs versus cTECs was probably >545 genes[49]. Extrapolating from their gene array analysis, they suggested that 5–10% of all mouse genes might be transcribed in mTECs, in addition to their cell lineage–specific expression program[49]. 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[9]. However, their method did not allow to segregate genes expressed by TECs vs. other stromal cells[9]. 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[hi] mTECs[49], and that stromal cells from different cortical and medullary subregions display unique gene expression signatures[9]. 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)[50]. 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 $\geq$ 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., CD80[hi] vs. CD80[lo] 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[51].

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 *Foxn1*)[52] as well as genes required for attraction of thymic seeding thymocyte progenitors (*Ccl25, Cxcl12*)[53,54] and for expansion of early thymic progenitors (*Dll4, Kitl, Il7* and *Wnt4*)[8,18] 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 *Galr2*, 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.[19] revealed that TECs proliferate actively, their self renewal potential is seemingly inferior to that of other populations such as epidermal and intestinal ECs[5,20,21]. 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[55,56]. 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[57]. 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 Expérimentation 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.[19] which yields adequate numbers of TECs and minimises changes in expression of cell surface markers[58]. 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 isotype controls were used: APC-Cy7 rat IgG2aκ (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[24], 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[22,23]. In order to deal with subject-to-subject variation, we therefore used the pooling approach proposed by Kendziorski et al.[59]. 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 chips (Agilent). Transcriptome libraries were generated from 1 μg of total RNA using the TruSeq RNA Sample Prep Kit (v2) (Illumina) following the manufacturer's protocols. Briefly, poly-A mRNA was purified using poly-T oligo-attached magnetic beads using two rounds of purification. During the second elution, RNA was fragmented and primed for cDNA synthesis. Reverse transcription of the first strand was done using random primers and SuperScript II (InvitroGene). A second round of reverse transcription was performed to generate a double-stranded cDNA, purified

using Ampure XP beads method (Beckman). End repair of fragmented cDNA, adenylation of the 3′ ends and ligation of adaptors were done following the manufacturer's protocol. Enrichment of DNA fragment with adapter molecules on both ends was done using 15 cycles of PCR amplification using the Illumina PCR mix and primers cocktail.

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.ilmn. 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 HiSeq 1000 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 ≥ 2; fold change ≥ 5) 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 regulation 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 implicated in specific functional discrepancies between cell types. The network view is an interactive representation that shows the molecular relationship between molecules from the dataset based on Ingenuity Knowledge Database.

1. Boehm, T. Design principles of adaptive immune systems. *Nat. Rev. Immunol.* **11**, 307–317 (2011).
2. Blais, M. E. *et al*. Why T cells of thymic versus extrathymic origin are functionally different. *J. Immunol.* **180**, 2299–2312 (2008).
3. Boehm, T., Hess, I. & Swann, J. B. Evolution of lymphoid tissues. *Trends Immunol* **33**, 315–321 (2012).
4. Blais, M. E. *et al*. Do thymically and strictly extrathymically developing T cells generate similar immune responses? *Blood* **103**, 3102–3110 (2004).
5. Heinonen, K. M. & Perreault, C. Development and functional properties of thymic and extrathymic T lymphocytes. *Crit. Rev. Immunol.* **28**, 441–466 (2008).
6. Anderson, G. & Takahama, Y. Thymic epithelial cells: working class heroes for T cell development and repertoire selection. *Trends Immunol* **33** (2012).
7. Anderson, G., Lane, P. J. & Jenkinson, E. J. Generating intrathymic microenvironments to establish T-cell tolerance. *Nat Rev Immunol* **7**, 954–963 (2007).
8. Petrie, H. T. & Zuniga-Pflucker, J. C. Zoned out: functional mapping of stromal signaling microenvironments in the thymus. *Annu. Rev. Immunol.* **25**, 649–679 (2007).
9. Griffith, A. V. *et al*. Spatial mapping of thymic stromal microenvironments reveals unique features influencing T lymphoid differentiation. *Immunity* **31**, 999–1009 (2009).
10. Ciofani, M. & Zuniga-Pflucker, J. C. The thymus as an inductive site for T lymphopoiesis. *Annu. Rev Cell Dev. Biol* **23**, 463–493 (2007).
11. Love, P. E. & Bhandoola, A. Signal integration and crosstalk during thymocyte migration and emigration. *Nat. Rev. Immunol.* **11**, 469–477 (2011).
12. Liston, A. *et al*. Gene dosage--limiting role of Aire in thymic expression, clonal deletion, and organ-specific autoimmunity. *J. Exp. Med.* **200**, 1015–1026 (2004).
13. Klein, L., Hinterberger, M., Wirnsberger, G. & Kyewski, B. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nat. Rev. Immunol.* **9**, 833–844 (2009).
14. Papadopoulou, A. S. *et al*. The thymic epithelial microRNA network elevates the threshold for infection-associated thymic involution via miR-29a mediated suppression of the IFN-alpha receptor. *Nat. Immunol.* **13**, 181–187 (2011).
15. Zuklys, S. *et al*. MicroRNAs control the maintenance of thymic epithelia and their competence for T lineage commitment and thymocyte selection. *J Immunol.* **189**, 3894–3904 (2012).
16. Corbeaux, T. *et al*. Thymopoiesis in mice depends on a Foxn1-positive thymic epithelial cell lineage. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16613–16618 (2010).
17. Nowell, C. S. *et al*. Foxn1 regulates lineage progression in cortical and medullary thymic epithelial cells but is dispensable for medullary sublineage divergence. *PLoS. Genet.* **7**, e1002348 (2011).
18. Heinonen, K. M. *et al*. Wnt4 regulates thymic cellularity through the expansion of thymic epithelial cells and early thymic progenitors. *Blood* **118**, 5163–5173 (2011).
19. Gray, D. H. *et al*. Developmental kinetics, turnover and stimulatory capacity of thymic epithelial cells. *Blood* **108**, 3777–3785 (2006).
20. Dooley, J. & Liston, A. Molecular control over thymic involution: From cytokines and microRNA to aging and adipose tissue. *Eur. J Immunol.* **42**, 1073–1079 (2012).
21. Griffith, A. V., Fallahi, M., Venables, T. & Petrie, H. T. Persistent degenerative changes in thymic organ function revealed by an inducible model of organ regrowth. *Aging Cell* **11**, 169–177 (2012).
22. Oshlack, A., Robinson, M. D. & Young, M. D. From RNA-seq reads to differential expression results. *Genome Biol.* **11**, 220 (2010).
23. Wilhelm, B. T. *et al*. RNA-seq analysis of 2 closely related leukemia clones that differ in their self-renewal capacity. *Blood* **117**, e27–e38 (2011).
24. Hakkinen, L., Koivisto, L. & Larjava, H. An improved method for culture of epidermal keratinocytes from newborn mouse skin. *Methods Cell Sci.* **23**, 189–196 (2001).
25. Hamazaki, Y. *et al*. Medullary thymic epithelial cells expressing Aire represent a unique lineage derived from cells expressing claudin. *Nat. Immunol.* **8**, 304–311 (2007).
26. Rhee, H., Polak, L. & Fuchs, E. Lhx2 maintains stem cell character in hair follicles. *Science* **312**, 1946–1949 (2006).
27. Gendronneau, G. *et al*. Galectin-7 in the control of epidermal homeostasis after injury. *Mol. Biol. Cell* **19**, 5541–5549 (2008).
28. Kim, H. *et al*. Loss of hairless confers susceptibility to UVB-induced tumorigenesis via disruption of NF-kappaB signaling. *PLoS One* **7**, e39691 (2012).
29. Roth, W. *et al*. Keratin 1 maintains skin integrity and participates in an inflammatory network in skin through interleukin-18. *J Cell Sci.* **125**, 5269–5279 (2012).
30. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621–628 (2008).
31. Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. Differential expression in RNA-seq: a matter of depth. *Genome Res.* **21**, 2213–2223 (2011).
32. Williams, L. M. *et al*. Rac mediates TNF-induced cytokine production via modulation of NF-kappaB. *Mol. Immunol* **45**, 2446–2454 (2008).
33. Hutti, J. E. *et al*. Oncogenic PI3K mutations lead to NF-kappaB-dependent cytokine expression following growth factor deprivation. *Cancer Res.* **72**, 3260–3269 (2012).
34. Akiyama, T., Shinzawa, M. & Akiyama, N. TNF receptor family signaling in the development and functions of medullary thymic epithelial cells. *Front Immunol* **3**, 278 (2012).
35. Maruthamuthu, V., Aratyn-Schaus, Y. & Gardel, M. L. Conserved F-actin dynamics and force transmission at cell adhesions. *Curr. Opin. Cell Biol.* **22**, 583–588 (2010).
36. Lin, K. T. *et al*. HDAC inhibitors augmented cell migration and metastasis through induction of PKCs leading to identification of low toxicity modalities for combination cancer therapy. *Clin. Cancer Res.* **18**, 4691–4701 (2012).
37. Etienne-Manneville, S. From signaling pathways to microtubule dynamics: the key players. *Curr. Opin. Cell Biol.* **22**, 104–111 (2010).
38. de Forges, H., Bouissou, A. & Perez, F. Interplay between microtubule dynamics and intracellular organization. *Int. J Biochem. Cell Biol.* **44**, 266–274 (2012).
39. Daire, V. *et al*. Kinesin-1 regulates microtubule dynamics via a c-Jun N-terminal kinase-dependent mechanism. *J Biol. Chem.* **284**, 31992–32001 (2009).
40. Buttrick, G. J. & Wakefield, J. G. PI3-K and GSK-3: Akt-ing together with microtubules. *Cell Cycle* **7**, 2621–2625 (2008).
41. Louis, J. *et al*. The signaling protein Wnt4 enhances thymopoiesis and expands multipotent hematopoietic progenitors through β-catenin-independent signaling. *Immunity* **29**, 57–67 (2008).
42. Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer* **9**, 153–166 (2009).
43. Sherr, C. J. & Roberts, J. M. Living with or without cyclins and cyclin-dependent kinases. *Genes Dev.* **18**, 2699–2711 (2004).
44. Luo, J. *et al*. Histone h3 exerts a key function in mitotic checkpoint control. *Mol. Cell Biol.* **30**, 537–549 (2010).
45. Lagana, A. *et al*. A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A. *Nat Cell Biol.* **12**, 1186–1193 (2010).
46. Magnaghi-Jaulin, L., Eot-Houllier, G., Fulcrand, G. & Jaulin, C. Histone deacetylase inhibitors induce premature sister chromatid separation and override the mitotic spindle assembly checkpoint. *Cancer Res.* **67**, 6360–6367 (2007).

47. Baselga, J. & Swain, S. M. Novel anticancer targets: revisiting ERBB2 and discovering ERBB3. *Nat Rev Cancer* **9**, 463–475 (2009).

48. Bell, E. *et al.* MYCN oncoprotein targets and their therapeutic potential. *Cancer Lett.* **293**, 144–157 (2010).

49. Derbinski, J. *et al.* Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. *J. Exp. Med.* **202**, 33–45 (2005).

50. van der Bruggen, P. *et al.* Tumor-specific shared antigenic peptides recognized by human T cells. *Immunol. Rev.* **188**, 51–64 (2002).

51. de Verteuil, D., Granados, D. P., Thibault, P. & Perreault, C. Origin and plasticity of MHC I-associated self peptides. *Autoimmun. Rev.* **11**, 627–635 (2012).

52. Manley, N. R. & Condie, B. G. Transcriptional regulation of thymus organogenesis and thymic epithelial cell differentiation. *Prog. Mol. Biol. Transl. Sci.* **92**, 103–120 (2010).

53. Calderon, L. & Boehm, T. Synergistic, context-dependent, and hierarchical functions of epithelial components in thymic microenvironments. *Cell* **149**, 159–172 (2012).

54. Zlotoff, D. A. *et al.* CCR7 and CCR9 together recruit hematopoietic progenitors to the adult thymus. *Blood* **115**, 1897–1905 (2010).

55. Rodriguez-Puebla, M. L. *et al.* Cyclin D2 overexpression in transgenic mice induces thymic and epidermal hyperplasia whereas cyclin D3 expression results only in epidermal hyperplasia. *Am. J Pathol.* **157**, 1039–1050 (2000).

56. Scheijen, B., Bronk, M., van der Meer, T., De, J. D. & Bernards, R. High incidence of thymic epithelial tumors in E2F2 transgenic mice. *J. Biol. Chem.* **279**, 10476–10483 (2004).

57. Simons, B. D. & Clevers, H. Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell* **145**, 851–862 (2011).

58. Seach, N., Wong, K., Hammett, M., Boyd, R. L. & Chidgey, A. P. Purified enzymes improve isolation and characterization of the adult thymic epithelium. *J Immunol Methods* **385**, 23–34 (2012).

59. Kendziorski, C., Irizarry, R. A., Chen, K. S., Haag, J. D. & Gould, M. N. On the utility of pooling biological samples in microarray experiments. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4252–4257 (2005).

60. Saeed, A. I. *et al.* TM4 microarray software suite. *Methods Enzymol.* **411**, 134–193 (2006).

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.

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).