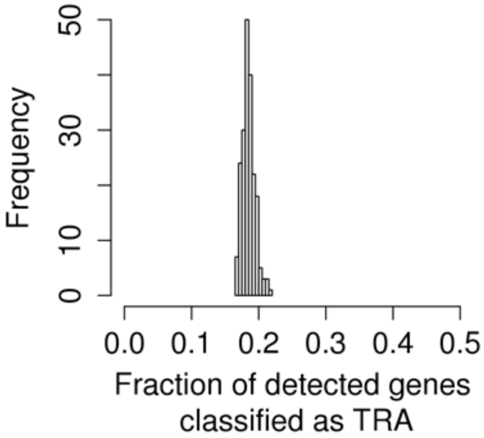
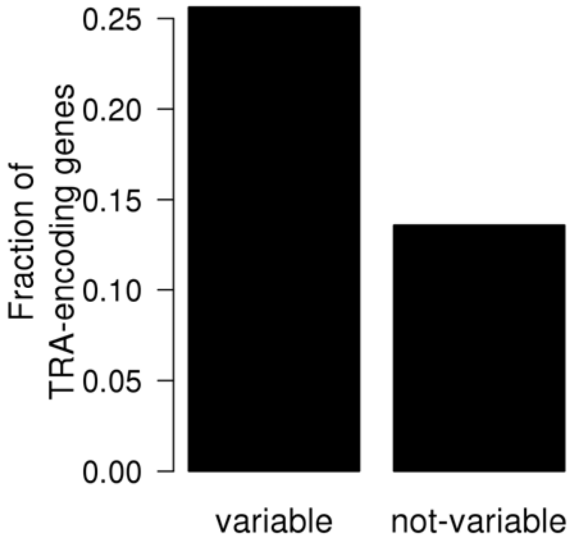


Supplementary Figure 1

Histogram of the fraction of genes detected per mTEC that are classified as TRA-encoding genes.

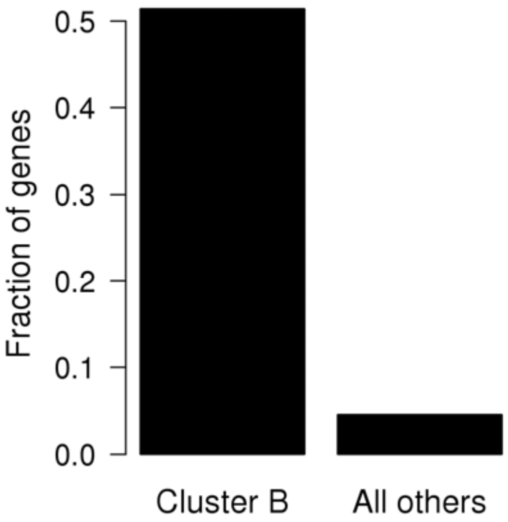
The x-axis shows the fraction of genes detected per cell that are classified as TRA-encoding genes. The y-axis shows the number of cells.



Supplementary Figure 2

Enrichment of TRA-encoding genes among highly variable genes.

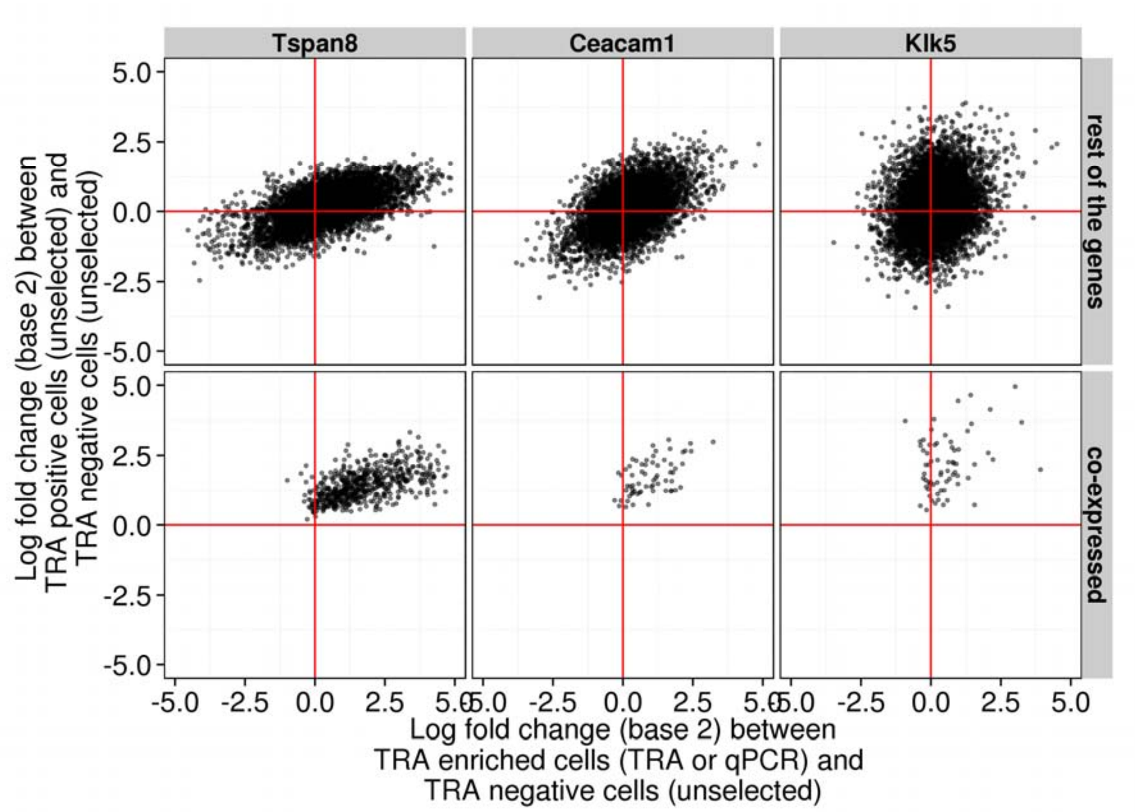
Barplot showing the fraction of TRA-encoding genes (y-axis) in two subsets of genes (x-axis), those that were detected to be highly variable and the rest of the genes that were not detected to be highly variable (See Methods for details). The difference in the fraction of genes overlapping with TRA-encoding genes was detected to be significantly different between the two gene subsets (p-value $< 2.2 \times 10^{-16}$, Fisher’s exact test).



Supplementary Figure 3

Enrichment of genes in the *Tspan8* co-expressed gene set among genes from cluster B resulting from the k-medoids clustering.

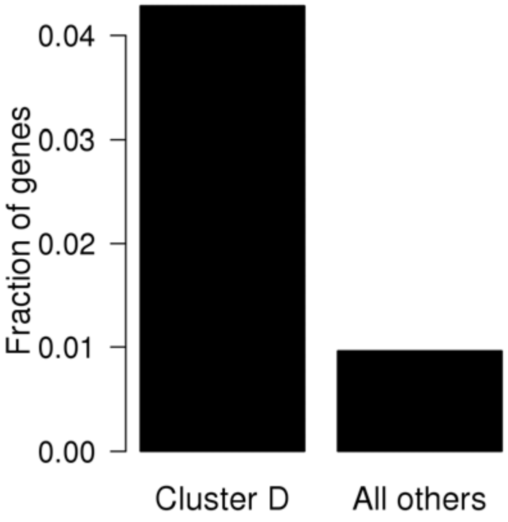
Barplot showing the fraction of genes from the *Tspan8* co-expressed gene set (y-axis) in two subsets of genes (x-axis), genes belonging to cluster B resulting from the k-medoids clustering and genes grouped to the rest of the k-medoids clusters (See Methods for details). The difference in the fraction of genes overlapping with the *Tspan8* co-expressed gene set was detected to be significantly different between the two gene subsets (p-value < 2.2×10^{-16}, Fisher's exact test).



Supplementary Figure 4

Validation of co-expressed gene sets using independent experimental approaches.

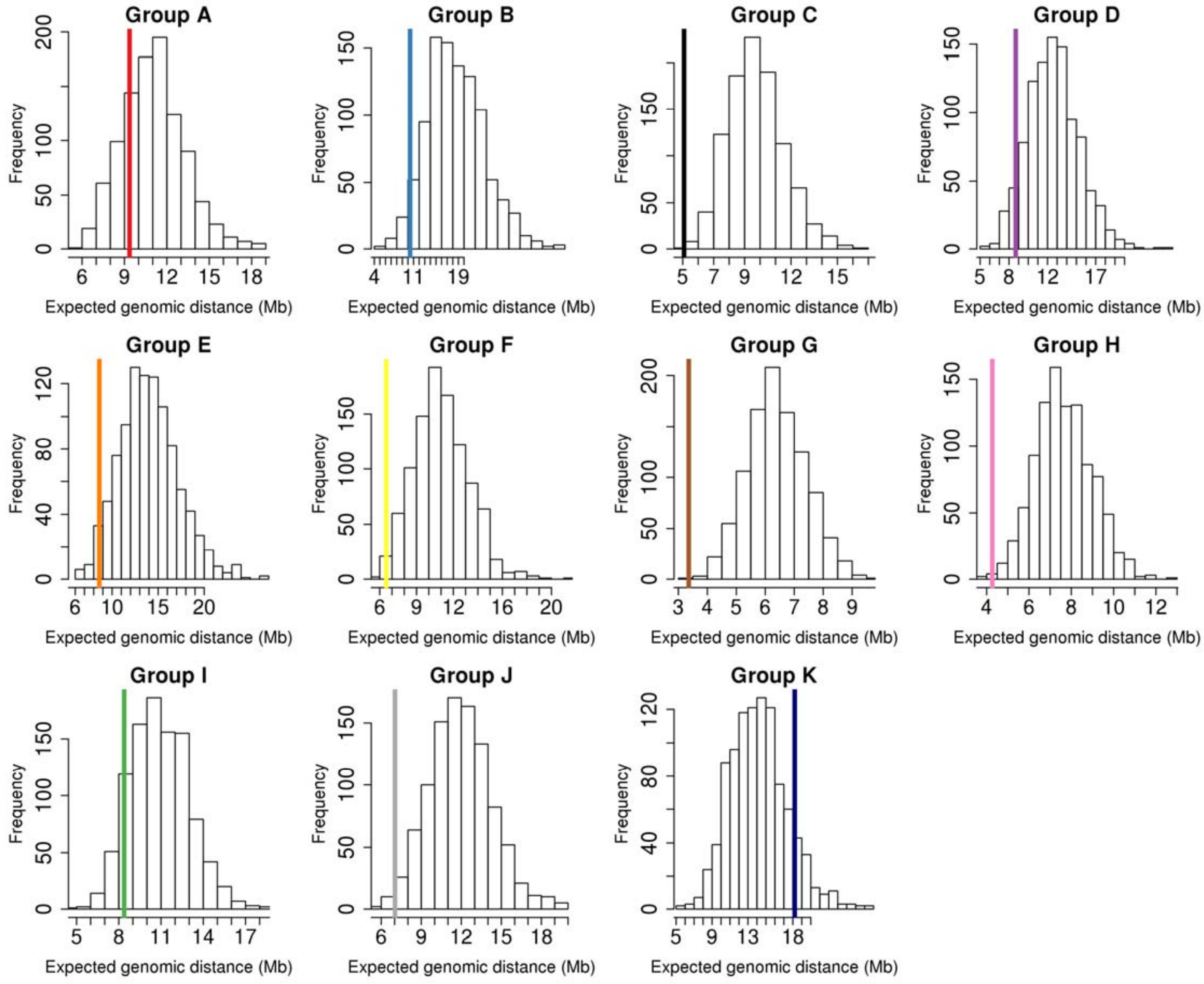
Each point depicts one gene. The y-axis shows the logarithmic fold change (base 2) of the unselected mature mTECs where TRA expression was detected by the scRNA-seq assay with respect to the unselected mature mTECs where the scRNA-seq assay did not detect the expression of the TRA. The x-axis shows the logarithmic fold change (base 2) between the mature mTECs selected for the expression of the TRA (by either qPCR or flow cytometry) and the unselected mature mTECs where the scRNA-seq assay did not detect the expression of the TRA. Each column panel shows the data for one TRA. The row panels split the genes according to whether they were detected to be co-expressed with the specific TRA or not.



Supplementary Figure 5

Enrichment of genes in the *Klk5* co-expressed gene set among genes from cluster D resulting from the k-medoids clustering.

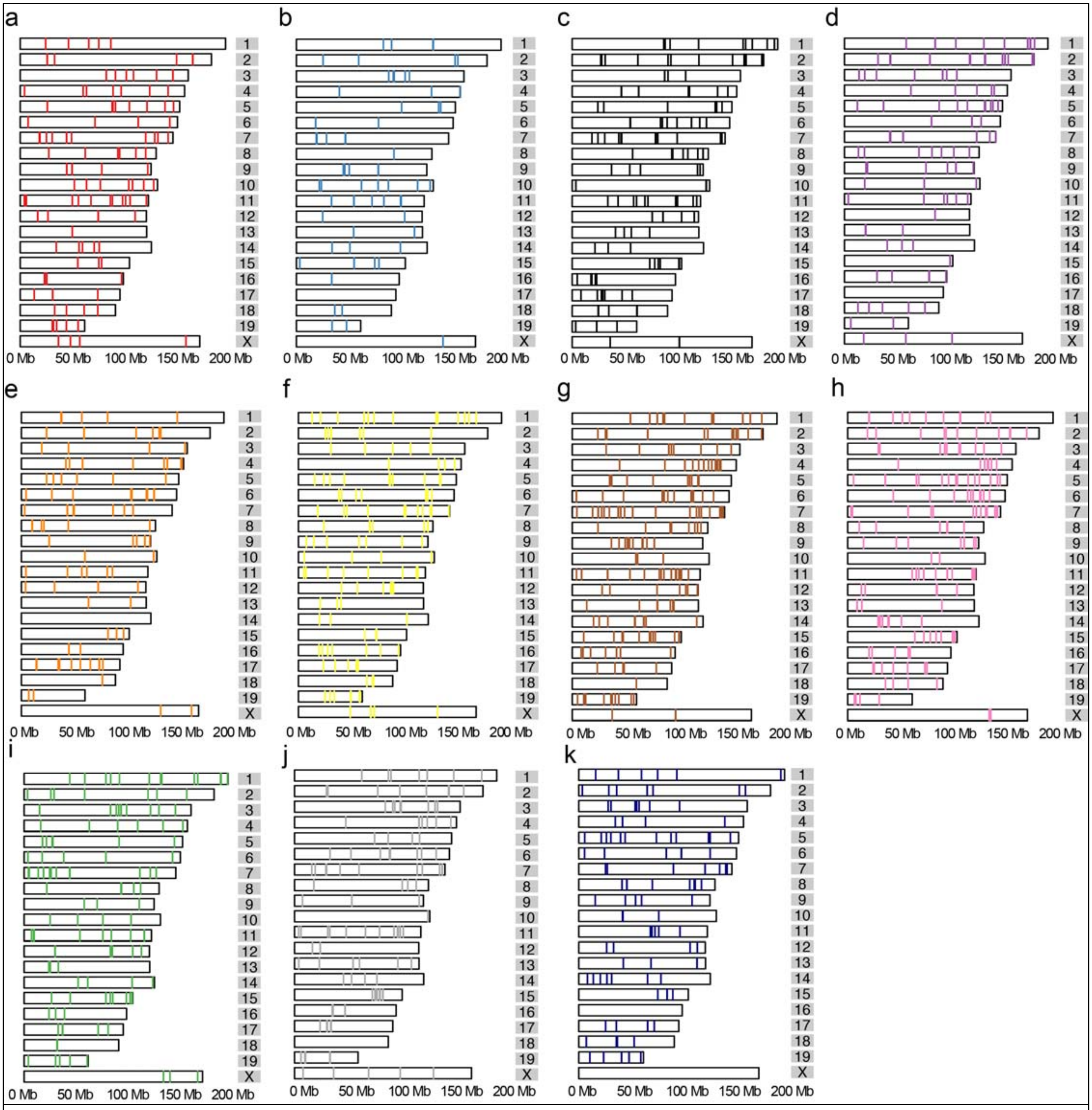
Barplot showing the fraction of genes from the *Klk5* co-expressed gene set (y-axis) in two subsets of genes (x-axis), genes belonging to cluster D resulting from the k-medoids clustering and genes grouped to the rest of the k-medoids clusters (See Methods for details). The difference in the fraction of genes overlapping with the *Klk5* co-expressed gene set was detected to be significantly different between the two gene subsets (p-value < 9.6×10^{-4}, Fisher's exact test).



Supplementary Figure 6

Expected vs. observed genomic proximity of the 11 groups of co-expressed genes resulting from the k-medoids clustering.

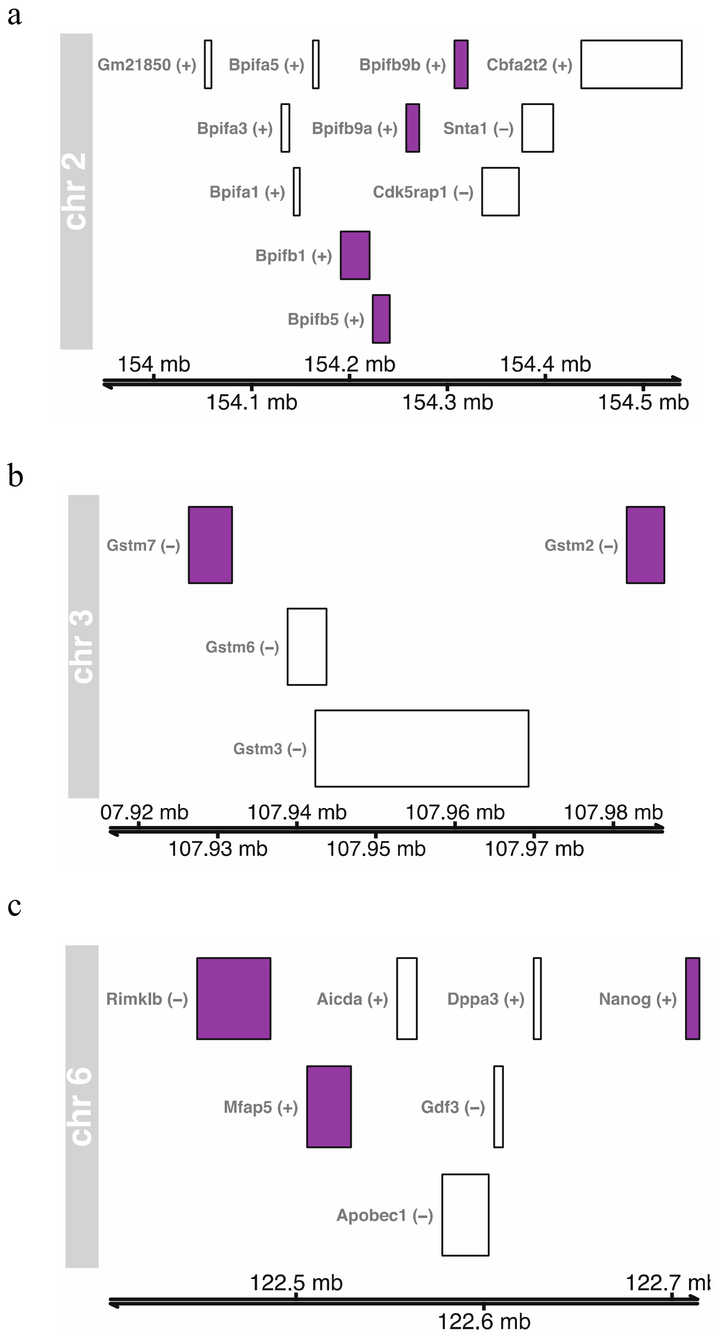
Each panel shows the data for one of the 11 clusters from Fig. 2. The histograms show the expected distribution of the median genomic distance between pairs of genes in the genome according to the size of each gene cluster. This expected distribution was estimated by sampling random genes of the same size of the gene cluster for 1,000 times. The observed median distance for each cluster is depicted with vertical solid lines.



Supplementary Figure 7

Karyogram depicting the genomic position of the 11 groups of co-expressed genes resulting from the k-medoids clustering.

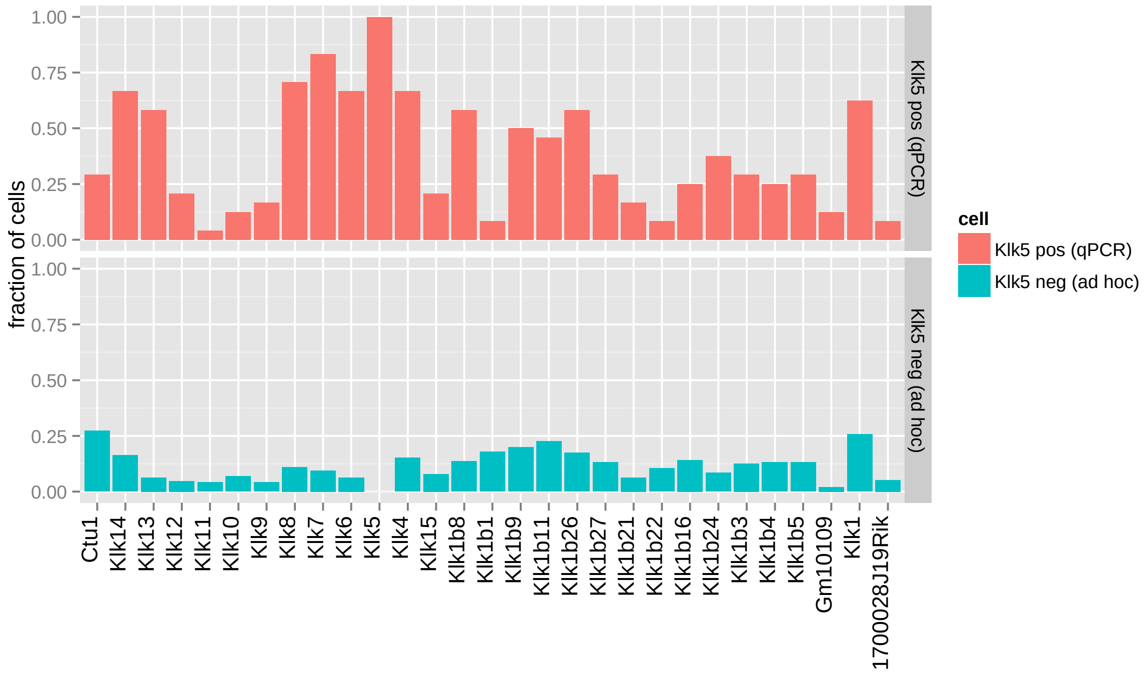
Each panel corresponds to one gene cluster. The colors correspond to the colors from Fig. 2. The colored vertical lines in the chromosomes mark the genomic position of genes.



Supplementary Figure 8

Examples of co-expressed genomic loci and their locations in the genome.

(a) The genomic locus of the BPI fold-containing B gene family is shown. The x-axis represents the genomic positions from chromosome 2 of the mouse genome. Protein coding genes are shown in boxes. Genes colored in purple belong to cluster “D”. **(b)** The genomic locus of the Glutathione S-transferase Mu gene family is shown. The x-axis represents the genomic positions from chromosome 3 of the mouse genome. Protein coding genes are shown in boxes. Genes colored in purple belong to cluster “D”. **(c)** Example of co-expressed genes from unrelated gene families that are clustered in the genome. The x-axis represents the genomic positions from chromosome 6 of the mouse genome. Protein coding genes are shown in boxes. Genes colored in purple belong to cluster “D”.



Supplementary Figure 9

Fraction of cells detecting the expression of genes from the Kallikrein protease gene family.

The y-axis shows the fraction of cells for which the scRNA-seq assay detected gene expression. The x-axis shows genes ordered by genomic location. The upper panel shows the data for the qPCR-selected Klk5+ cells. The lower panel shows the data for the unselected mature mTECs for which Klk5 expression was not detected by the scRNA-seq assay.