## **Supporting Information**

## Pinto et al. 10.1073/pnas.1308311110



**Fig. S1.** Defining coexpression groups. (A) Enrichment of the sorted antigen-positive vs. antigen-negative medullary thymic epithelial cells (mTECs) by quantitative PCR. Shown are Mucin (*MUC*)1 (#97, #96), carcinoembryonic antigen-related cell adhesion molecule (*CEACAM*)5 (CEA) (#116, #118), and sodium/glucose co-transporter (*SGLT*)1 (#177, #174). n.d., not detectable. (*B*) Additional example of autoimmune regulator (*AIRE*) mRNA levels within each mTEC subset. Shown MUC1 (#96), CEA, (#118), and SGLT1 (#177). (C) Example of two to three individuals sorted for MUC1<sup>+</sup>, CEA<sup>+</sup>, and SGLT1<sup>+</sup> mTECs used for array hybridization. Numbers within circles refer to genes. (*D*) Additional example of protein coexpression patterns of the three antigens defining the mTEC subsets in one individual (#153).



**Fig. 52.** Tissue-restricted self-antigen (TRA) enrichment and clustering of genes up-regulated in MUC1<sup>+</sup>, CEA<sup>+</sup>, and SGLT1<sup>+</sup> sorted human mTECs. (*A*) TRA content within up-regulated genes of MUC1<sup>+</sup>, CEA<sup>+</sup>, and SGLT1<sup>+</sup> mTECs. The different subsets showed a significant TRA enrichment (based on Fisher's exact test *P* value; MUC1<sup>+</sup> mTECs:  $P = 3.831 \times e^{-16}$ ; CEA<sup>+</sup> and SGLT1<sup>+</sup> mTECs:  $P = 2.2 \times e^{-16}$ ) compared with the basal TRA content in the human genome of ~20%. (*B*–*D*) Clustering of genes up-regulated in MUC1<sup>+</sup>, CEA<sup>+</sup>, and SGLT1<sup>+</sup> sorted human mTECs. The 10-gene window clustering algorithm was used to assess clustering of up-regulated genes in the different antigen-expressing mTEC subsets. In all three subsets, the up-regulated genes showed significant clustering above a cluster size of 4 compared with 1,000 random simulations. Two of the largest clusters consisting of 10 genes were observed in CEA sorted mTECs, one of which was the *CEACAM* gene family on chromosome 19. k indicates number of genes. The TRA content and clustering data for each antigen was averaged among individuals; MUC1, *n* = 3 (#94, #96, #97); CEA, *n* = 2 (#116, #118); SGLT1, *n* = 3 (#169, #174, #177).



**Fig. S3.** Coexpression patterns of single MUC1<sup>+</sup>/MUC1<sup>-</sup> and CEA<sup>+</sup>/CEA<sup>-</sup> mTECs. (A) Each row represents a single MUC1<sup>+</sup> mTEC (*Left*) or MUC1<sup>-</sup> mTEC (*Right*) from one individual (#103), as analyzed for expression of the genes listed above the columns; the last column shows the total number of genes expressed by each cell. Black stripes denote detected expression; gray stripes denote lack of expression. Expression patterns were arranged from top to bottom according to increasing numbers of genes expressed per cell. (*B*) Each row represents a single CEA<sup>+</sup> mTEC (*Left*) or CEA<sup>-</sup> mTEC (*Right*) (#143), as described for A.



**Fig. 54.** Coexpressed genes are colocalized. Additional examples of the distribution of distances between the *MUC1* and *CEA* gene loci using DNA-FISH in MUC1<sup>+</sup> vs. MUC1<sup>-</sup> mTECs (A) (#133) and CEA<sup>+</sup> vs. CEA<sup>-</sup> mTECs (B and C) (#134, #137). In all samples studied, the antigen-positive fraction showed significant colocalization of the two loci compared with antigen-negative mTECs. Each plot represents cells from one individual; a minimum of 50 cells was analyzed using the fully automatic image analysis for each sample. (*x* axis: distances in 0.2- $\mu$ m increments.)

Table S1.	Correlation analysis of genes ranked at different			
positions of	on the MUC1 <sup>+</sup> vs. MUC1 <sup>-</sup> microarray in terms of fold			
change in MUC1-positive and -negative mTECs as detected by				
single-cell	PCR			

	MUC1	PSCA	APOA2	GALNTL2
MUC1	_	0.5030	0.0520	-0.1199
PSCA	_	_	0.0264	-0.1497
APOA2	_	_	_	-0.0368
GALNTL2	—	—	_	_

Coexpression analysis #105, MUC1<sup>+</sup> and MUC1<sup>-</sup> mTECs ( $\kappa$  index).  $\kappa$  index of correlation between the genes studied by single-cell PCR.  $\kappa < 0.1$ , no concordance;  $\kappa = 0.10-0.40$ , weak concordance;  $\kappa = 0.41-0.60$ , clear concordance;  $\kappa = 0.61-0.80$ , strong concordance;  $\kappa = 0.81-1.00$ , nearly complete concordance. PSCA, prostate stem cell antigen; APO, apolipoprotein A2; GALNTL2, UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase-like protein 2.

Table S2.Correlation analysis of gene expression in MUC1-positive and -negative mTECs as detected by single-cell PCR

	MUC1	MUC4	CEA	CEACAM6
MUC1	_	0.6329	0.3158	0.6067
MUC4	_	—	0.5110	0.7669
CEA	_	—	—	0.5447
CEACAM6	—	—	—	—

Coexpression analysis #103, MUC1<sup>+</sup> and MUC1<sup>-</sup> mTECs ( $\kappa$  index).  $\kappa$  index of correlation between the genes studied by single-cell PCR.  $\kappa$  index of correlation between the genes studied by single-cell PCR.  $\kappa < 0.1$ , no concordance;  $\kappa = 0.10-0.40$ , weak concordance;  $\kappa = 0.41-0.60$ , clear concordance;  $\kappa = 0.61-0.80$ , strong concordance;  $\kappa = 0.81-1.00$ , nearly complete concordance.

Table S3. Correlation analysis of gene expression in CEApositive and -negative mTECs as detected by single-cell PCR

	MUC1	MUC4	CEA	CEACAM6
MUC1	_	0.5181	0.3421	0.4378
MUC4	—	_	0.5502	0.6496
CEA	—	_	—	0.7623
CEACAM6	—	—	—	_

Coexpression analysis #143, CEA<sup>+</sup> and CEA<sup>-</sup> mTECs ( $\kappa$  index).  $\kappa$  index of correlation between the genes studied by single-cell PCR.  $\kappa < 0.1$ , no concordance;  $\kappa = 0.10-0.40$ , weak concordance;  $\kappa = 0.41-0.60$ , clear concordance;  $\kappa = 0.61-0.80$ , strong concordance;  $\kappa = 0.81-1.00$ , nearly complete concordance.

SANG SANG