Where five candidate targets are expressed

A GTEx and TCGA expression atlas for MAGEA4, NY-ESO-1, PRAME, SSX2 and CD99

Zatoona Labs · October 2026

Summary

This atlas asks a narrow question of five proteins that are discussed as cancer targets: in which normal tissues, and in which tumours, is each gene actually transcribed? The answer separates the panel into three groups. MAGEA4 and PRAME are close to testis-restricted in normal tissue, with tau values of 1.00 and 0.97, and only 2 of the five genes qualify as tissue specific. CD99 is the opposite: it is expressed in every sampled tissue, at tau 0.19, and its tumour signal is therefore not tumour-specific. CTAG1B (the gene behind the NY-ESO-1 antigen) and SSX2 fall below 1 TPM in every GTEx tissue, which is a limit of the measurement rather than evidence of absence, because the Human Protein Atlas detects both.

Across 52 tissues (GTEx's two cell cultures left out), tau runs from 0.19 (CD99) to 1.00 (MAGEA4): 2 tissue specific (MAGEA4, PRAME); 1 broad (CD99); 2 below 1 TPM everywhere (CTAG1B, SSX2). In tumours, PRAME reaches its highest median in skin cutaneous melanoma, MAGEA4 in lung squamous cell carcinoma, and CD99 in thymoma, where the matched normal tissue is itself high. No condition other than cancer was supplied, so no Expression Atlas or GEO experiment was read; the condition-specific section rests on TCGA and on the atlas's own cancer categories.

1 Data and methods

Expression values are GTEx analysis release GTEx gtex_v10 (GENCODE v39), retrieved 25 October 2026 [1, 2]. The release quantifies 54 tissue sites; two of them are cell cultures rather than tissues, and the atlas reports the remaining 52 tissues; the dataset holds 270 gene-by-site rows for the 5 genes across all 54 sites. Each value is the median TPM over the site's donors, which range from 11 to 818 per site. Gene names were resolved through HGNC: NY-ESO-1 is an HGNC alias of CTAG1B. The atlas therefore reports CTAG1B where the request said NY-ESO-1, and the figures label the gene both ways.

The Human Protein Atlas was read on the same day for its RNA tissue category, its tissue mass-spectrometry result and its antibody-staining reliability, and its nTPM values are quoted only for tissues GTEx does not sample [3, 4]. Tumour expression comes from TCGA through the UCSC Xena Toil recompute, which places 32 tumour types beside GTEx normals processed in the same pipeline [5]; 16 of those tumour types also carry adjacent normal samples. Toil's GTEx samples are an older release than the portal's, so tumours are compared with Toil normals and never with the portal medians in Figure 1.

Specificity is summarised by the tau index, computed on $\log_2(\mathrm{TPM}+1)$ across the 52 tissues: 0 when a gene has the same level everywhere and 1 when one tissue carries all of it [6]. The atlas calls 0.8 or more tissue specific, under 0.5 broad, and the remainder intermediate. Transcript-level medians were read from the same GTEx release and are compared within a gene only, because GTEx quantifies transcripts and genes with different pipelines.

2 Tissue expression

MAGEA4 has one clear site. MAGEA4 is highest in Testis 25.6 TPM; tau 1.00 (tissue specific), at 1 TPM or more in 1 of 52 tissues. Every other tissue sits at or below 1 TPM, and the second-ranked tissue in Table 1 is a zero. The Human Protein Atlas agrees on the tissue and adds the sites GTEx does not sample [4]: placenta reads 4.5 nTPM, and its RNA category is tissue enriched with testis at 20.7 nTPM.

PRAME is the strongest testis signal in the panel. PRAME is highest in Testis 165, Ovary 5.42, Adrenal Gland 2.41 TPM, 30.4 times the next tissue; tau 0.97 (tissue specific), at 1 TPM or more in 3 of 52 tissues. The gap between testis and the next tissue is the largest in the panel, and the atlas adds epididymis at 9.2 nTPM among tissues GTEx does not sample.

CD99 is expressed everywhere. CD99 is highest in Artery - Aorta 404, Artery - Tibial 404, Nerve - Tibial 396 TPM; tau 0.19 (broad), at 1 TPM or more in 52 of 52 tissues. The three highest sites are all large arteries and tibial nerve, and the lowest sampled tissue still clears 1 TPM. The atlas reports low tissue specificity and detection in all tissues [4], with appendix, bone marrow, duodenum, gallbladder, lymph node, placenta, retina, seminal vesicle, thymus and tonsil between 14.1 and 523.3 nTPM in the sites GTEx does not cover.

CTAG1B and SSX2 are the two genes the portal cannot resolve. CTAG1B stays below 1 TPM in every GTEx tissue (highest Testis 0.06), so its tau of 1.00 says nothing about specificity; the Human Protein Atlas detects it, so GTEx's figure is a lower bound (GTEx counts only reads that map to one place, and a gene with a near-identical copy loses them). SSX2 stays below 1 TPM in every GTEx tissue (highest Testis 0.14), so its tau of 1.00 says nothing about specificity; the Human Protein Atlas detects it, so GTEx's figure is a lower bound (GTEx counts only reads that map to one place, and a gene with a near-identical copy loses them). In both cases the atlas detects the RNA and calls the gene tissue enriched [4], so the GTEx near-zero is a lower bound produced by a pipeline that counts only reads mapping to one place, not a statement that the transcript is absent.

Table 1. Highest tissues and specificity summary for each gene. Values are GTEx median TPM; the atlas RNA category is the Human Protein Atlas's own classification and is not interchangeable with tau.

Gene	Highest tissues (median TPM)	Tau	Spread	Atlas RNA category
MAGEA4	Testis 25.6; Adipose - Subcutaneous 0	1.00	tissue specific	Tissue enriched
CTAG1B	Testis 0.06; Adipose - Subcutaneous 0	1.00	below 1 TPM everywhere	Tissue enriched
PRAME	Testis 165; Ovary 5.42	0.97	tissue specific	Tissue enriched
SSX2	Testis 0.14; Thyroid 0.03	1.00	below 1 TPM everywhere	Tissue enriched
CD99	Artery - Aorta 404; Artery - Tibial 404	0.19	broad	Low tissue specificity

3 Condition-specific expression

No condition beyond cancer was supplied with this request, so no Expression Atlas or GEO experiment was read and no disease-versus-control contrast outside cancer is reported here. What can be said about conditions comes from TCGA and from the atlas's cancer classification.

PRAME has the clearest tumour signal. PRAME is highest in TCGA Skin Cutaneous Melanoma 290 TPM (n=102; GTEx Skin - Sun Exposed (Lower Leg) 0.13); Uterine Carcinosarcoma 133 TPM (n=57; GTEx Uterus 0.22); Uterine Corpus Endometrioid Carcinoma 126 TPM (n=180; adjacent normal 5.63 (n=23)). Melanoma and the two uterine tumour types stand far above the normal

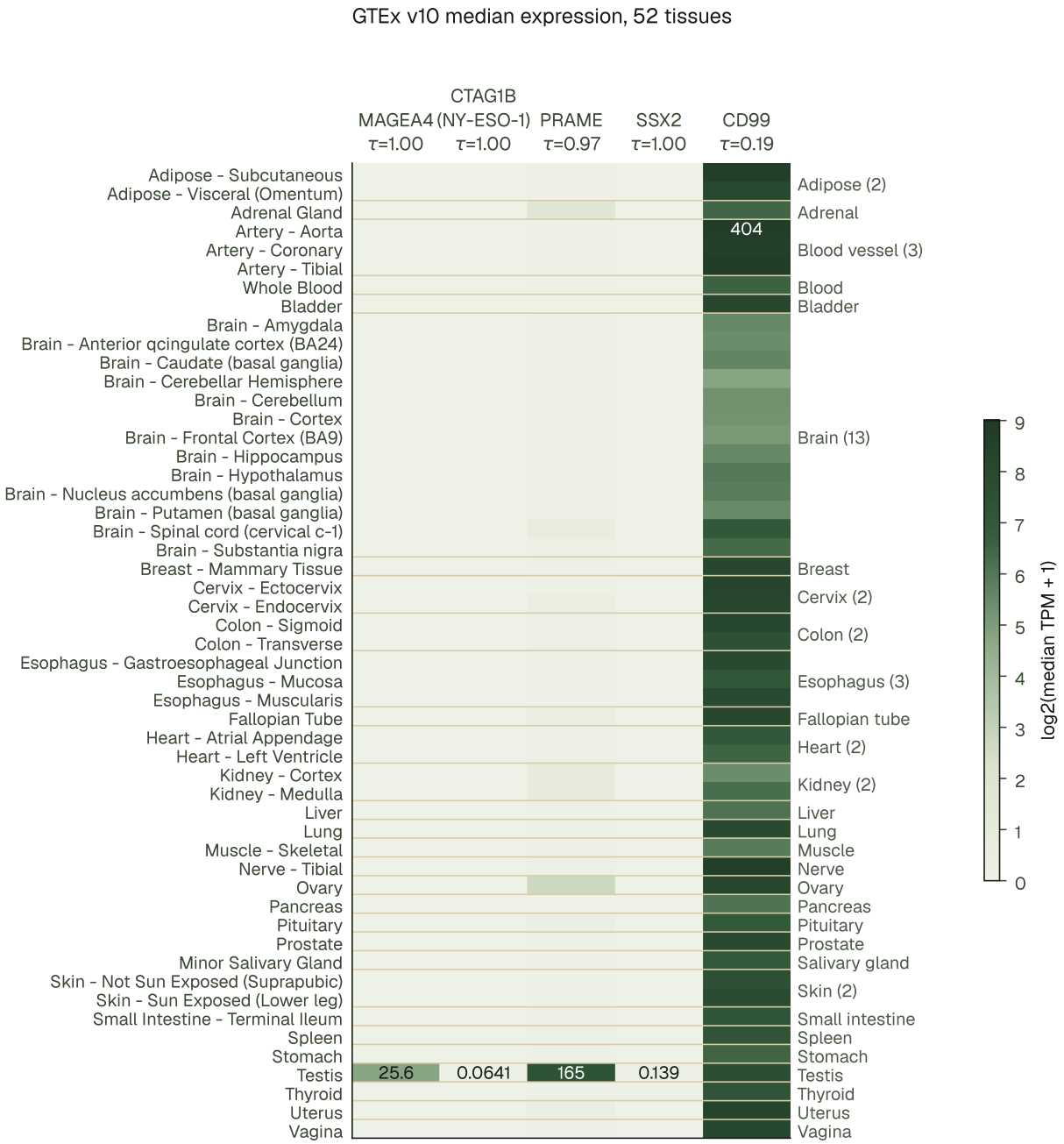


Figure 1. Median expression of the five genes across the 52 GTEx tissues, on a $\log_2(\mathrm{TPM}+1)$ scale. Rows are grouped by organ system and the number of sites in each group is given at the right. Tau is printed under each gene name. MAGEA4 and PRAME light up only in testis; CD99 is dark in every row; CTAG1B and SSX2 stay at the floor of the scale, which is the measurement limit discussed in the text, not an absence call. The maximum of each column is printed in its cell.

tissues in Figure 2, and the atlas classifies PRAME RNA as cancer enhanced, detected in many tumour types [4].

MAGEA4 follows. MAGEA4 is highest in TCGA Lung Squamous Cell Carcinoma 8.38 TPM (n=498; adjacent normal 0.05 (n=50), GTEx Lung 0); Testicular Germ Cell Tumor 0.54 TPM (n=148; GTEx Testis 29.2); Head & Neck Squamous Cell Carcinoma 0.38 TPM (n=518; adjacent normal 0.02 (n=44)). The lung squamous signal rises clearly above its adjacent normal of 0.05 TPM, and the atlas also calls MAGEA4 cancer enhanced. Testicular germ cell tumours do not show it at the level normal testis does, which is consistent with the gene being a germ-cell programme rather than a general tumour programme.

CD99 is high in tumours but not selectively so: CD99 is highest in TCGA Thymoma 856 TPM (n=119); Glioblastoma Multiforme 750 TPM (n=153; GTEx Brain - Spinal Cord (Cervical C-1) 163); Sarcoma 630 TPM (n=258). Thymoma is the highest, and normal thymus is 523.3 nTPM in the atlas, so the tumour value reflects a tissue that already expresses the gene at high level. The atlas classifies CD99 RNA as low cancer specificity and detected in all cancers [4].

CTAG1B and SSX2 show no tumour signal in this read. CTAG1B is highest in TCGA Testicular Germ Cell Tumor 0.26 TPM (n=148; GTEx Testis 1.9); Lung Squamous Cell Carcinoma 0.03 TPM (n=498; adjacent normal 0 (n=50), GTEx Lung 0.03); Bladder Urothelial Carcinoma 0.02 TPM (n=407; adjacent normal 0.02 (n=19)). SSX2 is highest in TCGA Thyroid Carcinoma 0.37 TPM (n=504; adjacent normal 0.61 (n=59)); Testicular Germ Cell Tumor 0.05 TPM (n=148; GTEx Testis 27.1); Adrenocortical Cancer 0 TPM (n=77). The atlas calls CTAG1B low cancer specificity and SSX2 not detected in cancer [4]. Both genes have near-identical paralogues, and Toil's RSEM shares reads between them: the same pipeline reads CTAG1B at 1.9 TPM and SSX2 at 27.1 TPM in testis, against 0.06 and 0.14 on the GTEx portal. The tumour values for these two genes should therefore be read as lower bounds as well.

4 Tissue specificity

Across 52 tissues (GTEx's two cell cultures left out), tau runs from 0.19 (CD99) to 1.00 (MAGEA4): 2 tissue specific (MAGEA4, PRAME); 1 broad (CD99); 2 below 1 TPM everywhere (CTAG1B, SSX2). Tau is a summary of shape, not of level: a gene that is absent everywhere and a gene that is present in one tissue both score near 1, which is why the two near-zero genes are counted separately rather than called specific.

MAGEA4 and PRAME are the tissue-specific members of the panel, and both are specific for testis. That is a restricted normal distribution in the sense that matters for a target: outside testis, the sampled tissues are at or near the floor. It is not a claim that the proteins are absent from every other cell type, because bulk tissue medians dilute rare expressing cells, and because testis itself is an immune-privileged site whose antigens are the classical cancer-testis group.

CD99 is the ubiquitous member. It clears 1 TPM in all 52 tissues and 10 TPM in all of them as well, so no tissue in the panel is a spared site. A target with this distribution has no expression-defined therapeutic window in normal tissue; any selectivity would have to come from something other than tissue-level transcription.

CTAG1B and SSX2 cannot be classified from GTEx. Their tau values of 1.00 and 1.00 are computed on values below 1 TPM and are noise. The atlas's categories, tissue enriched with de-

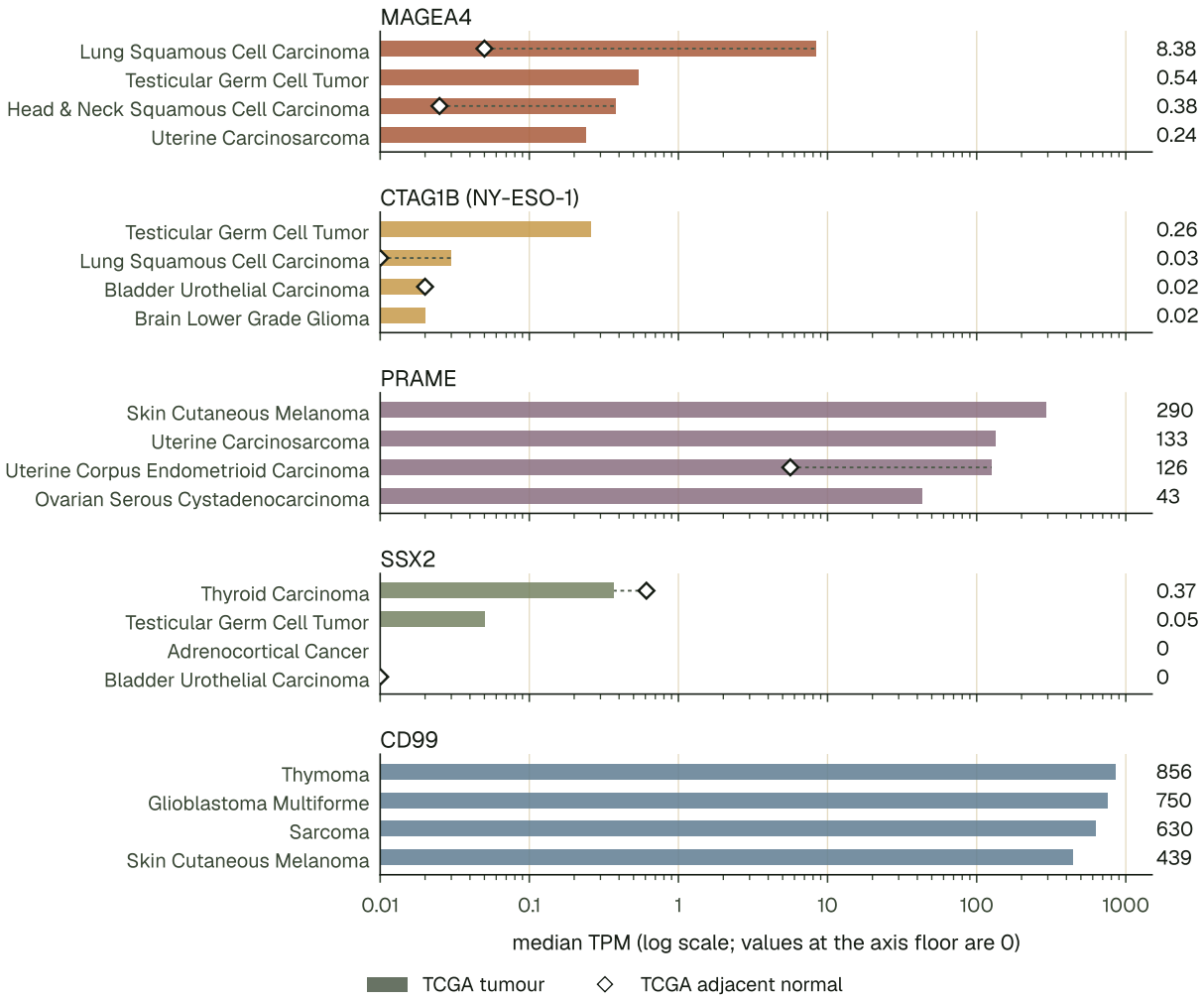


Figure 2. The four highest TCGA tumours for each gene, with the adjacent normal median where the cohort has one (open diamond, joined by a dotted line). The axis is logarithmic; a value of zero is drawn at the axis floor and printed as 0. PRAME separates melanoma and uterine tumours from their normal tissue, MAGEA4 separates lung squamous carcinoma from adjacent normal, and CD99 is high in every tumour shown, including tumours of tissues that normally express it.

tection in a single tissue [4], are the better summary for these two, and they point to testis as well.

5 Isoform-level differences

GTEx quantifies between 2 and 18 transcripts per gene in this panel, and the transcript medians show one dominant form per gene in testis. For MAGEA4, 18 transcripts are quantified and the testis signal is led by MAGEA4-218 at 35%, so no single isoform carries the gene. CTAG1B is the opposite: CTAG1B-201 carries 89% of its 2 quantified transcripts.

PRAME is led by PRAME-202 at 48%, SSX2 by SSX2-202 at 100%, and CD99 by CD99-205 at 75%.

Among tissues with clear expression, at or above 10 TPM, no gene's two top tissues lead with different transcripts. The one tissue-level difference sits below that threshold: in ovary, where PRAME reaches 5.42 TPM, the leading transcript is PRAME-214 at 53%, while testis leads with

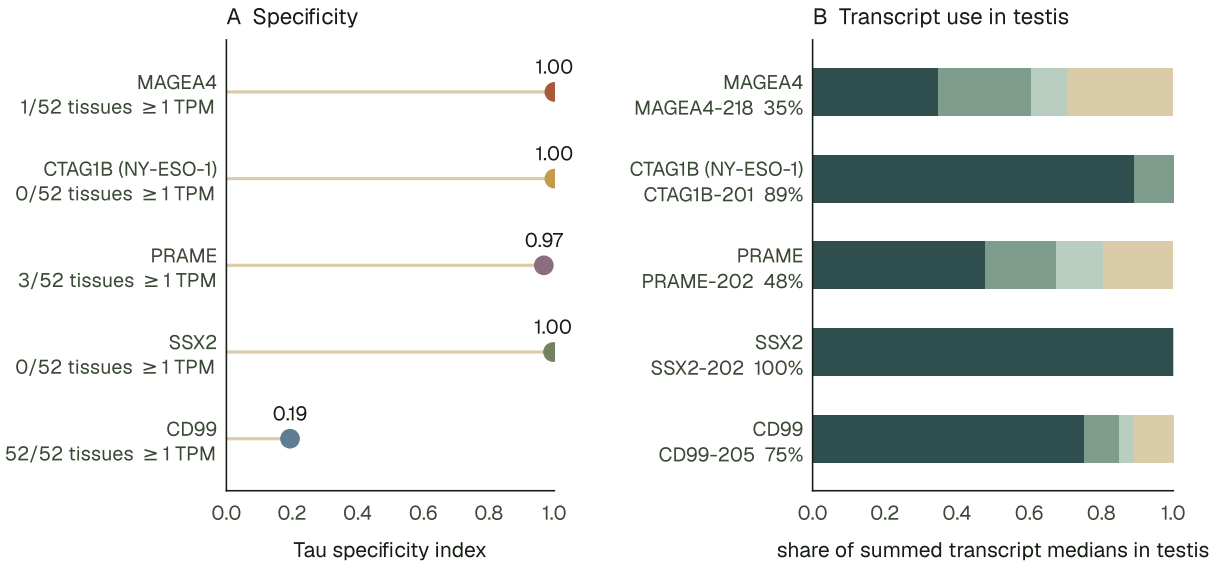


Figure 3. (A) Tau for each gene, with the number of tissues at or above 1 TPM. CD99 is the only broad gene; the two genes below 1 TPM everywhere have tau values that carry no specificity information. (B) Share of the summed transcript medians in testis, with the leading transcript and its share printed under each gene. The darkest segment is the leading transcript, the middle shade the second and the lightest the third; the pale remainder is all other transcripts.

PRAME-202. At that level the difference is a hypothesis about alternative promoter use, not a finding, and it would need a transcript-resolved dataset with more depth in ovary to test.

Two cautions apply to all of these shares. GTEx quantifies transcripts and genes with different pipelines, so a transcript's median is comparable with its siblings and not with the gene's TPM; the shares in Figure 3B are within-gene proportions for that reason. And the transcript models themselves are GENCODE calls, so a share is the abundance of an annotated model, not a direct measurement of protein isoform.

6 Limitations

The values are bulk post-mortem tissue medians. A median mixes every cell type in the sample, so a tissue that reads low can still contain a rare population that expresses the gene, and a tissue that reads high can owe its signal to one abundant cell type. Nothing here resolves expression to a cell type.

RNA is not protein. The atlas's tissue mass spectrometry does not detect MAGEA4, CTAG1B, PRAME or SSX2 protein, and reports CD99 as tissue enhanced; antibody-staining reliability is enhanced for the first four and supported for CD99 [4]. Those are the atlas's own categories, and a protein-level claim beyond them would need the staining record for the specific tissue.

Several values are limits of the measurement rather than biology. A gene below 1 TPM everywhere is not absent: CTAG1B and SSX2 are detected by the atlas, and GTEx's pipeline drops reads that map to more than one place, which penalises genes with near-identical paralogues. CD99 is pseudoautosomal, and GTEx also lists a Y copy that reads 0 TPM everywhere by construction; the dataset holds the X copy, and the Y copy should not be read as silence. Medians from fewer than 50 donors move with one donor: kidney medulla has 11 donors, cervix endocervix 23, cervix ectocervix 24 and fallopian tube 29. Tissues GTEx does not sample, including

placenta, thymus, bone marrow, lymph nodes and retina, are unknowns here, not zeros; the atlas readings quoted above are the only values for them.

The tumour comparison has its own limits. Toil's GTEx samples are an older release than the portal's, so tumours must be compared with Toil normals, as in Figure 2, and not with the medians in Figure 1. TCGA has no small-cell lung cancer and few cases of most rare tumours, and groups of fewer than 10 samples were left out. Adjacent normal is available for only 16 of the 32 tumour types, so many tumour values have no matched normal in this read.

Finally, expression suggests where a drug against a target would act; it does not prove a toxicity or its absence. This report holds no drug facts, approvals or warnings, and no condition data beyond cancer.

References

[1] GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318–1330. PMID 32913098 10.1126/science.aaz1776

[2] GTEx Portal. GTEx Analysis Release V10, GENCODE v39. Broad Institute of MIT and Harvard. https://gtexportal.org (accessed 25 October 2026).

[3] Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419. PMID 25613900 10.1126/science.1260419

[4] Human Protein Atlas. Tissue expression, cancer classification and antibody-staining records for MAGEA4, CTAG1B, PRAME, SSX2 and CD99. https://www.proteinatlas.org (accessed 25 October 2026).

[5] Vivian J, Rao AA, Nothaft FA, Ketchum C, Armstrong J, Novak A, et al. Toil enables reproducible, open source, big biomedical data analyses. Nat Biotechnol. 2017;35(4):314–316. PMID 28398314 10.1038/nbt.3772

[6] Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 2005;21(5):650–659. PMID 15388519 10.1093/bioinformatics/bti042