Diffuseness: the chosen panel, rule and CTA impact

Use HPA v23 consensus RNA and a 10% cutoff for adult/placental Reproductive concentration. The current local replay moves **MTCP1 from CTA to Neither**. Safety criteria and Unsafe membership are unchanged. These are research identity criteria, not clinical safety estimates. The boundary is applied to every protein, without named-target exceptions.

Category	Before	After
CTA	523	522
Neither	19	20
Unsafe	18031	18031
Unassigned	1290	1290

Every changed assignment · All source-specific scores · Complete tissue roles and weights · Full curation report

The fixed panel

Use HPA v23 consensus RNA because it covers placenta and the adult reproductive organs used by the positive-expression rule. This avoids mixing a GTEx donor median with an HPA maximum or a single-cell embryonic mean. HPA consensus is a curated cross-dataset summary, not a single donor cohort; this score cannot resolve donor heterogeneity or antigen presentation.

There are **38 comparator contexts representing 29 anatomical units**, 11 Reproductive reference contexts and one excluded **Permissive** context: thymus. Brain subdivisions share one organ weight, as do colon/rectum and small intestine/duodenum. The full comparator comprises adipose, adrenal, appendix, bone marrow, brain, colon, esophagus, gallbladder, heart, kidney, liver, lung, lymph node, pancreas, parathyroid, pituitary, retina, salivary gland, skeletal muscle, skin, small intestine, smooth muscle, spinal cord, spleen, stomach, thyroid, tongue, tonsil and urinary bladder.

Disposable tissues count for identity. Thymus counts in neither the comparator nor the Reproductive reference; its RNA remains visible. This Permissive role is separate from the safety tier and from the legacy scope alias named `permissive`. Thymic expression of peripheral antigens has experimental support in Projection of an Immunological Self Shadow Within the Thymus by the Aire Protein. It does not imply every thymic cell expresses every protein. Resolved thymic anatomy uses the same profile role in human, dog and mouse; quantitative scores here are human-only. Explicit embryonic/fetal stage retains Reproductive precedence.

Continuous score and cutoff

For each comparator context, use **N / (N + R)**, where N is that context's nTPM and R is the highest of the 11 Reproductive nTPM values in this same HPA matrix. Average within each anatomical unit, then average the 29 unit scores. Multiply by 100 to display a percentage. It is an average RNA similarity, **not the percentage of tissues expressing the protein**.

There is no per-tissue on/off expression threshold. If all normals are tenfold below R the score is 9.09%; if all equal R it is 50%. One arbitrarily high organ contributes at most 1/29 = 3.45 percentage points. A broad profile accumulates contributions across organs. Scores **above 10%** fail the adult/placental identity route; equality passes. This custom boundary expresses our preference, not a published biological or safety threshold. The benchmark of tissue-specificity metrics motivates attention to tissue sampling; our bounded Reproductive-anchored score is not its Tau statistic.

Which route can this reject?

Apply the identity gate only with a **qualifying adult/placental reference of at least 5 nTPM**, complete HPA comparator/reference measurements and exact same-source protein-member coverage. Missing members are not zero and missing tissues are not removed from the denominator. Scores for weaker anchors remain descriptive.

A supported embryonic route remains independent. An adult profile cannot establish embryonic diffuseness without compatible developmental/adult measurements. Among the 523 proteins entering this change, **154 have embryonic-only qualifying evidence**; their developmental concentration remains explicitly unassessed. **7 adult-positive proteins lack a complete HPA contrast**: CIMIP3, H2AL1Q, PTTG1IP2, SMIM42, TMEM247, TSPY9, ZC3H11C. Their required GTEx core remains complete, and absence of this additional contrast is disclosed rather than imputed as either passing or failing. This

limitation should be resolved with compatible panels, not by borrowing a favorable ratio from a different source.

A failed identity route with no remaining qualifying route becomes **Neither** when required coverage is complete and normal safety has no unresolved hold. Independent safety exclusions still produce Unsafe. Missing safety evidence does not become Neither. IHC confidence changes the safety assessment, not this RNA score.

Why MTCP1 changes

MTCP1 has testis 5.1 nTPM, skin 4.1, marrow 2.8, pancreas and tonsil 2.5, liver/retina/lymph node 2.4, and widespread values around 1-2 nTPM. Its **24.53%** HPA score indicates diffuse expression despite passing the existing normal-tissue safety gates. Direct GTEx independently yields a descriptive **16.82%** score; the sources are not independent cohorts because HPA incorporates GTEx. Its supported embryonic summary is 1.32 TPM, below the 5 TPM association floor. Every MTCP1 HPA measurement.

Established targets and cases needing more data

Protein	Category	HPA reference nTPM	Embryonic TPM	Adult contrast	Diffuseness (%)
CTAG2	CTA	10.1	7.151	Concentrated	5.147
MAGEA4	CTA	21.5	203.6	Concentrated	0.1515
CT55	CTA	3.7	10.85	Weak adult reference	0.334
ZNF730	Unsafe	4.9	41.65	Weak adult reference	9.532
PRAME	CTA	102.2	22.89	Concentrated	0.304
CT83	CTA	62.2	4.286	Concentrated	0.1494
MTCP1	Neither	5.1	1.321	Diffuse	24.53
FAM72D	CTA	2.1	10.39	Weak adult reference	23.21
ZNF878	CTA	1	21.26	Weak adult reference	19.82
NY-ESO-1	CTA	15.5	0.6505	Concentrated	0.0718
XAGE1A/B	CTA	98.8	2.237	Concentrated	0.2247

PRAME, MAGEA4, NY-ESO-1, XAGE1A/B and CTAG2 remain CTA. FAM72D and ZNF878 retain embryonic evidence; their weak adult anchors cannot settle developmental specificity. ZNF730's HPA score falls from 10.34% to 9.53% after excluding thymus, but its independent normal-expression exclusions keep it Unsafe. Diffuseness does not rescue unsafe proteins.

Sensitivity around the chosen cutoff

Cutoff (%)	Adult routes failing	CTA to Neither	Proteins moved
5	4	2	GOLGA8K, MTCP1
7.5	1	1	MTCP1
10	1	1	MTCP1
12.5	1	1	MTCP1
15	1	1	MTCP1
20	1	1	MTCP1
25	0	0	None

At 5%, CTAG2 and GDF9 lose the adult concentration route but retain supported embryonic evidence; GOLGA8K and MTCP1 would move to Neither. Between 7.5% and 20%, only MTCP1 moves. At 25%, it remains CTA. We choose 10% for an interpretable broad contrast, not to force a requested number of removals. Full sensitivity table.

GTEx and reproduction

GTEx v11 remains the mandatory 52-context coverage core and a separately reported donor-median comparison. It is not an interchangeable score: it lacks placenta and has different reproductive values for some familiar antigens. No score takes the minimum or maximum across sources.

Rebuild with `scripts/regenerate_cta_walkthrough.py`, followed by `scripts/audit_cta_profile_concentration.py`. Native tables use the uniform cache readers. The report manifest pins code, source checksums, before/after assignments and every output. `oncoref.expression_diffuseness.assess()` exposes numeric terms; `oncoref.cta_selection_policy.concentration_evidence()` applies the route policy.