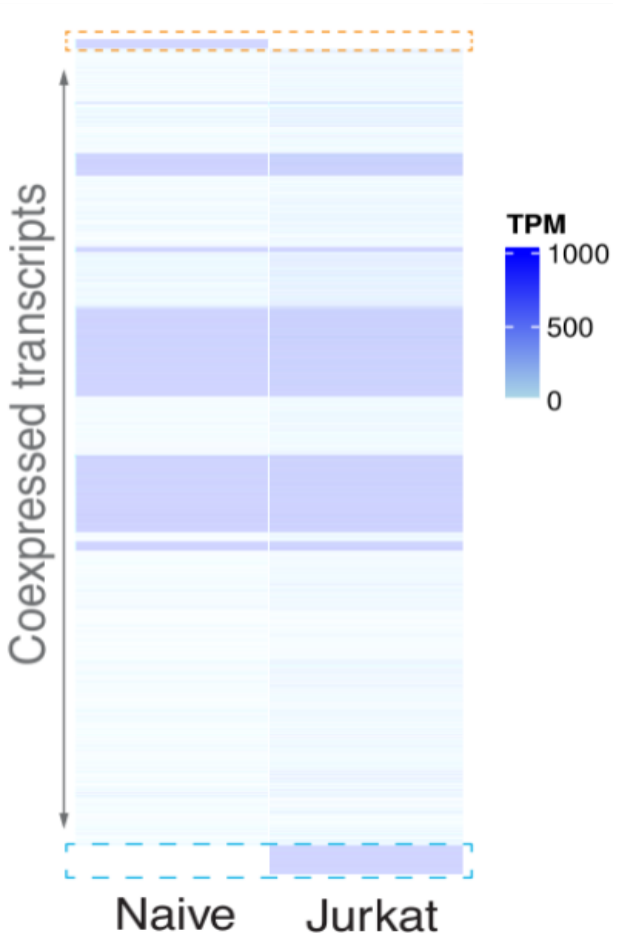
Supporting Information

Direct RNA Nanopore sequencing of primary human T-cells reveals the impact of immortalization on mRNA modifications

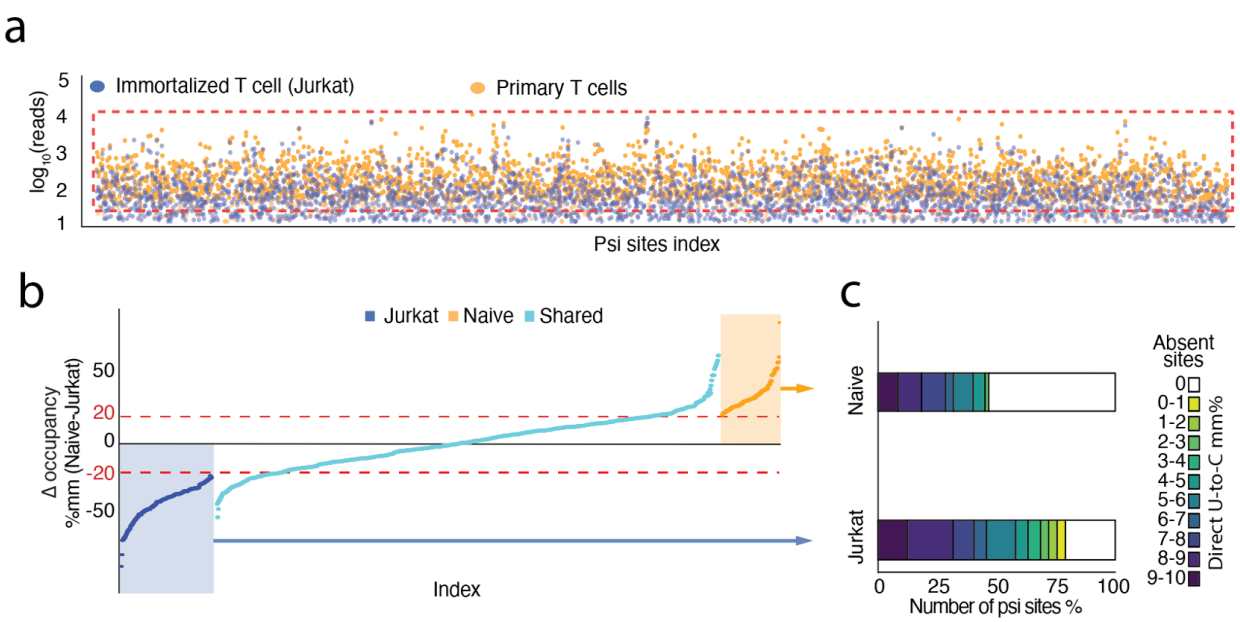
Oleksandra Fanari[1], Dylan Bloch[1], Yuchen Qiu[1], Michele Meseonznik[1], Amr Makhamreh[1], Meni Wanunu[1, 2], and Sara H. Rouhanifard[1,3,*]

[1]Dept. of Bioengineering, Northeastern University, Boston, MA
[2]Dept. of Physics, Northeastern University, Boston, MA
[3]*Lead contact*
[*]*Correspondence:* s.rouhanifard@northeastern.edu

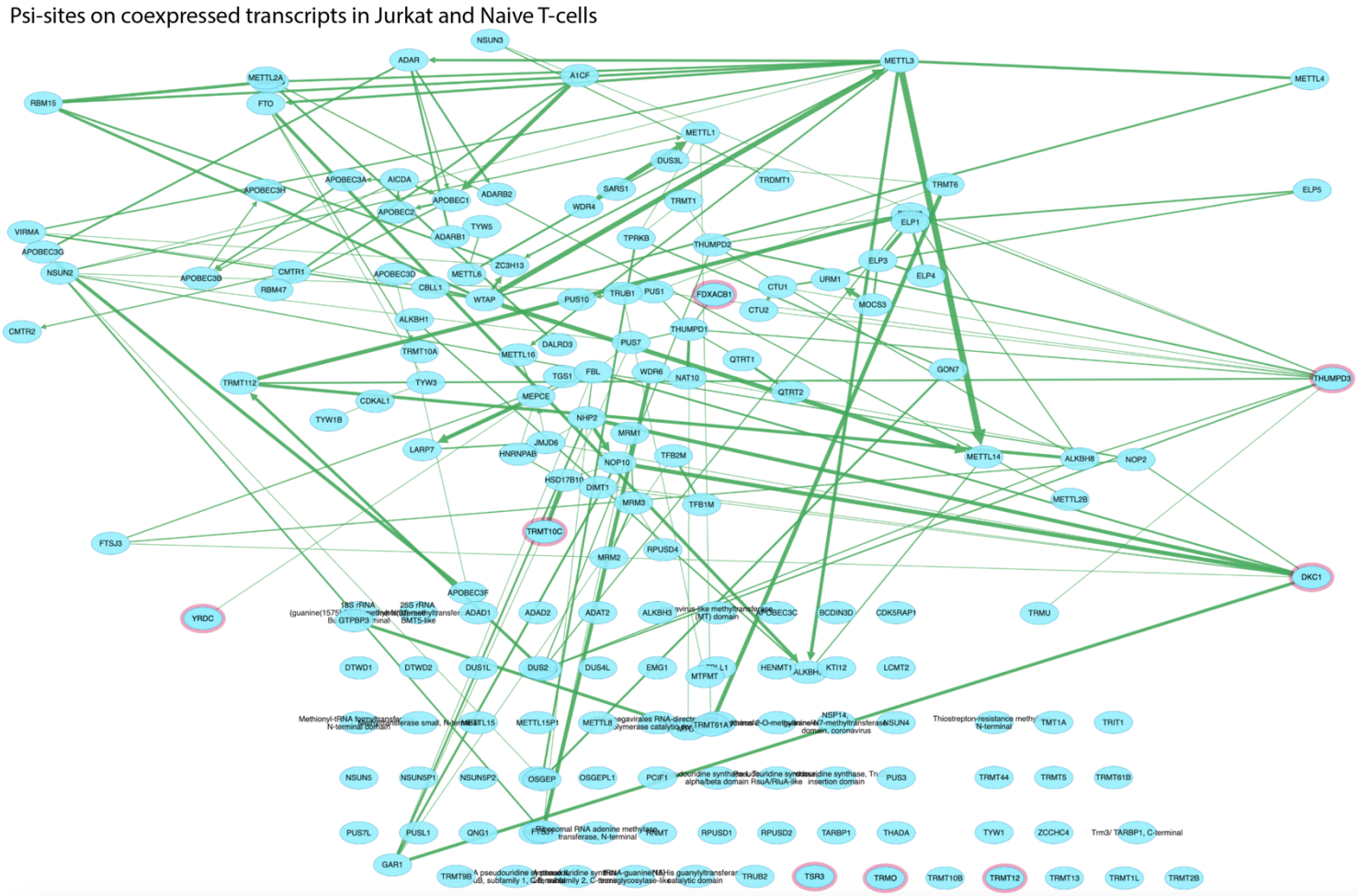
Table of Contents



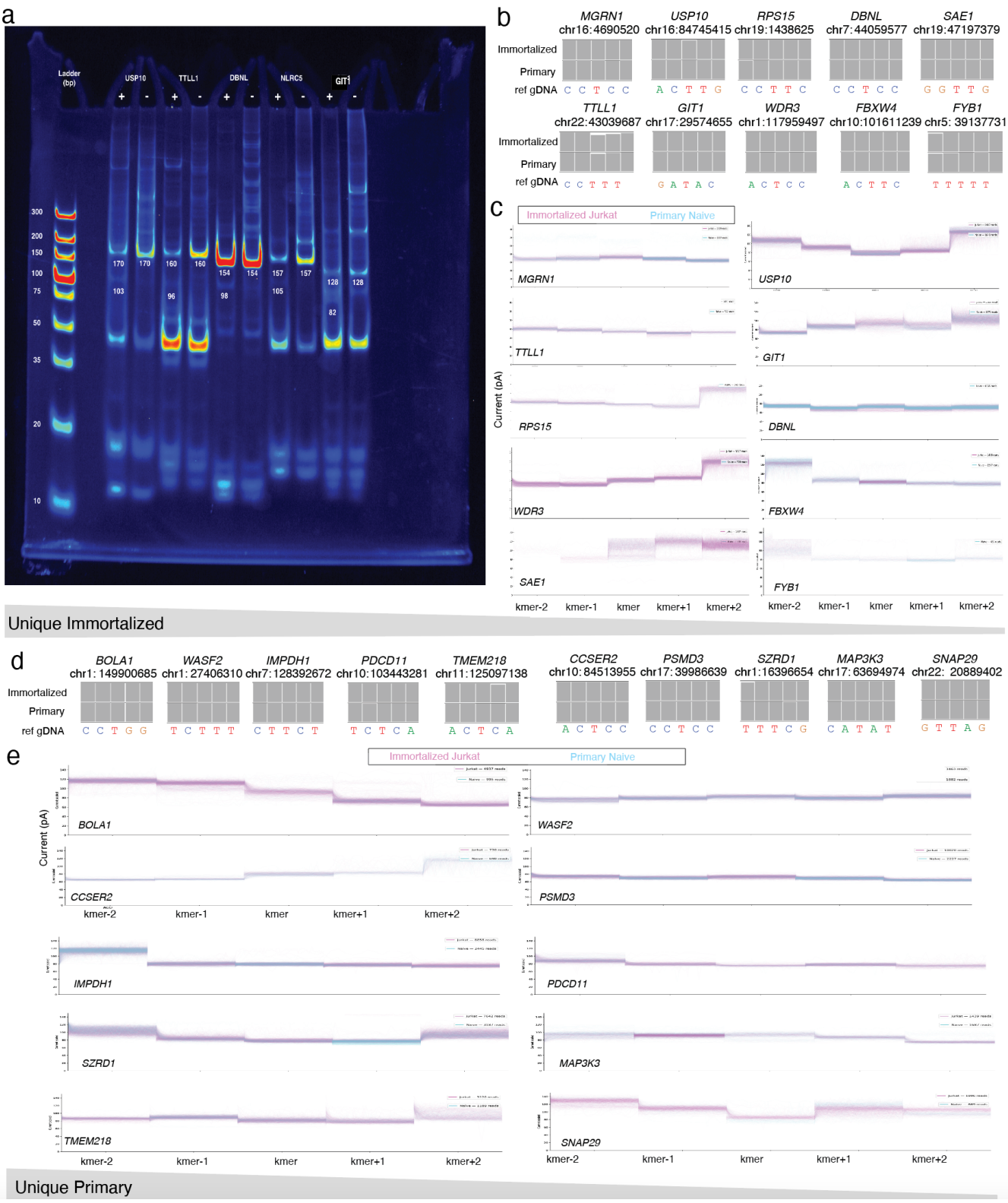
Supplementary Figure 1. Heatmap showing the expression levels across primary and immortalized T-cells. The majority of the transcripts are co-expressed (63%). Uniquely expressed transcripts in primary cells are surrounded by an orange box and constitute 12% of the whole population. Uniquely expressed transcripts in immortalized cells are surrounded by a blue box and constitute 25% of the whole population. The expression levels (TPMs) were assessed using NanoCount.



Supplementary Figure 2. Co-expressed transcripts show common and unique psi-sites.
a. Log_{10} read counts for 3,109 putative ψ sites in primary (orange) and immortalized (blue) libraries; the red dashed box indicates the 20-read cutoff.
b. Mismatch differences (naïve primary - immortalized Jurkat T cells) for unique Ψ-sites detected in immortalized T cells (in blue), unique Ψ-sites detected in primary T cells (in orange) and Ψ-sites common to both cell lines (in turquoise) are ranked in increasing order. Dashed lines at a 20% mismatch difference threshold classify shared sites as either stable or variable in occupancy.
c. U-to-C mismatch percentages for the Ψ-sites unique to primary and immortalized Jurkat T cells span 0-10 %mm range and are shown in color scale. Sites that are completely absent have a mm%=0, shown in white.

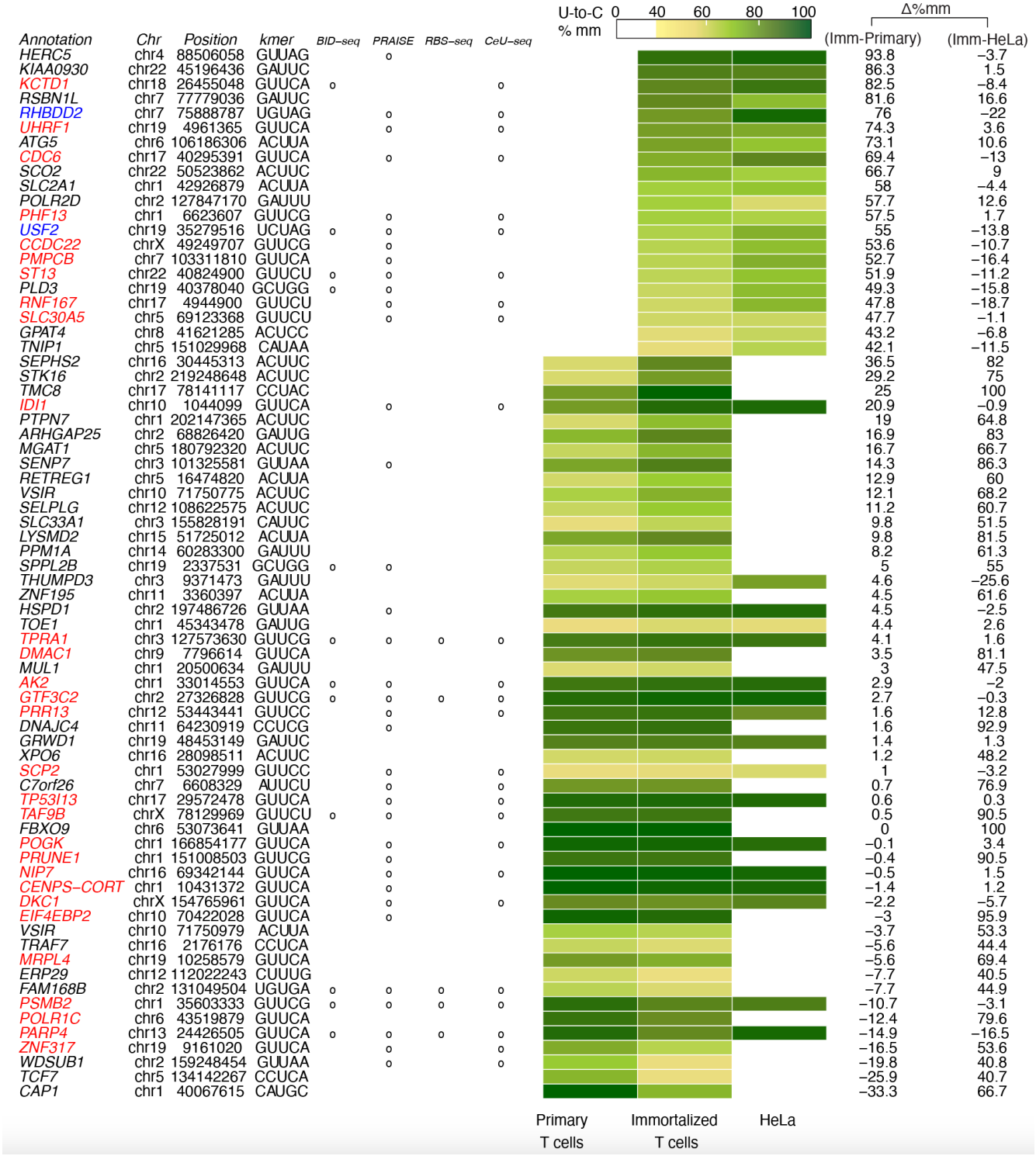


Supplementary Figure 3. Visualization of the RNA modification GO term network with Cytoscape
The network for the RNA modification GO term shows all the proteins known to contribute to the process. Psi-sites common to Jurkat and Naive cells detected on co-expressed transcripts that are located on proteins related to the RNA modification process are highlighted in pink.

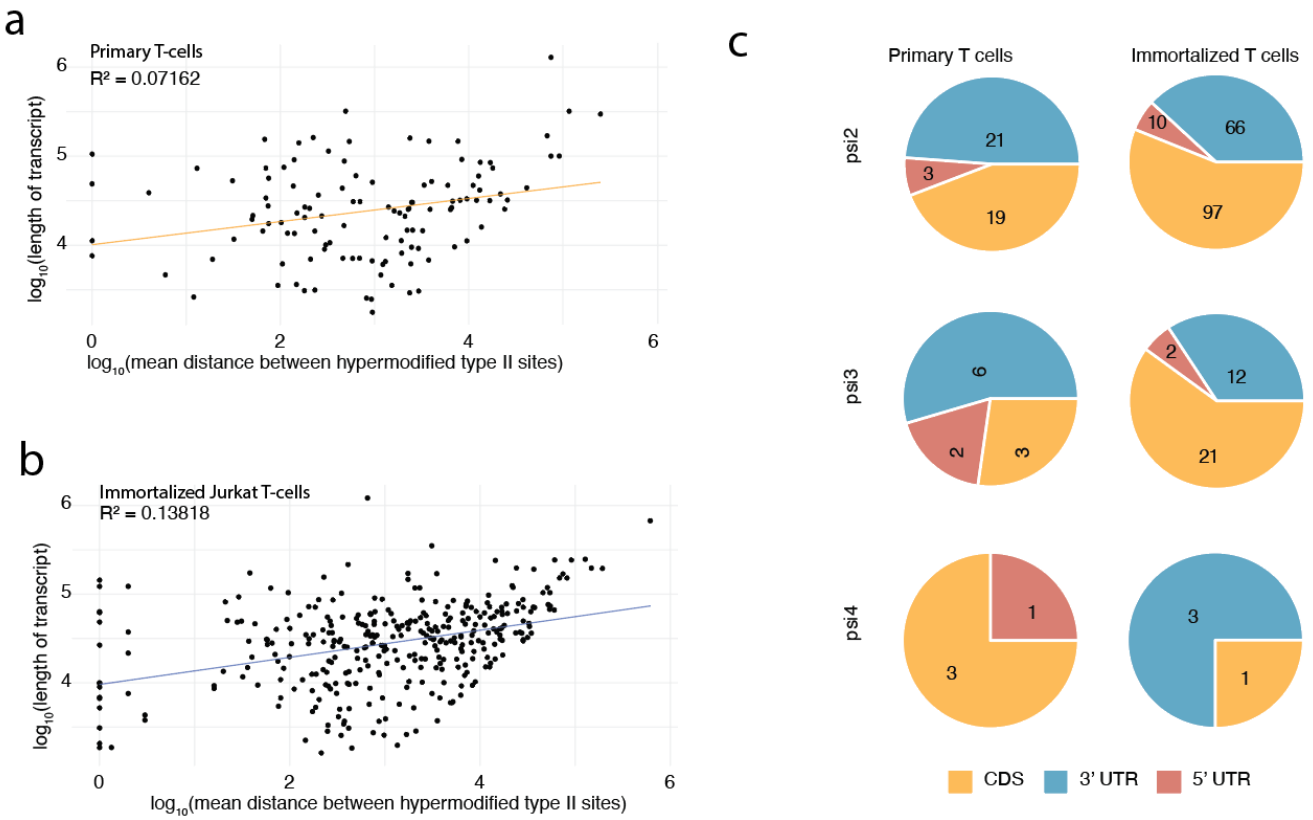


Supplementary Figure 4. Validation of unique psi-sites in primary and immortalized T cells.

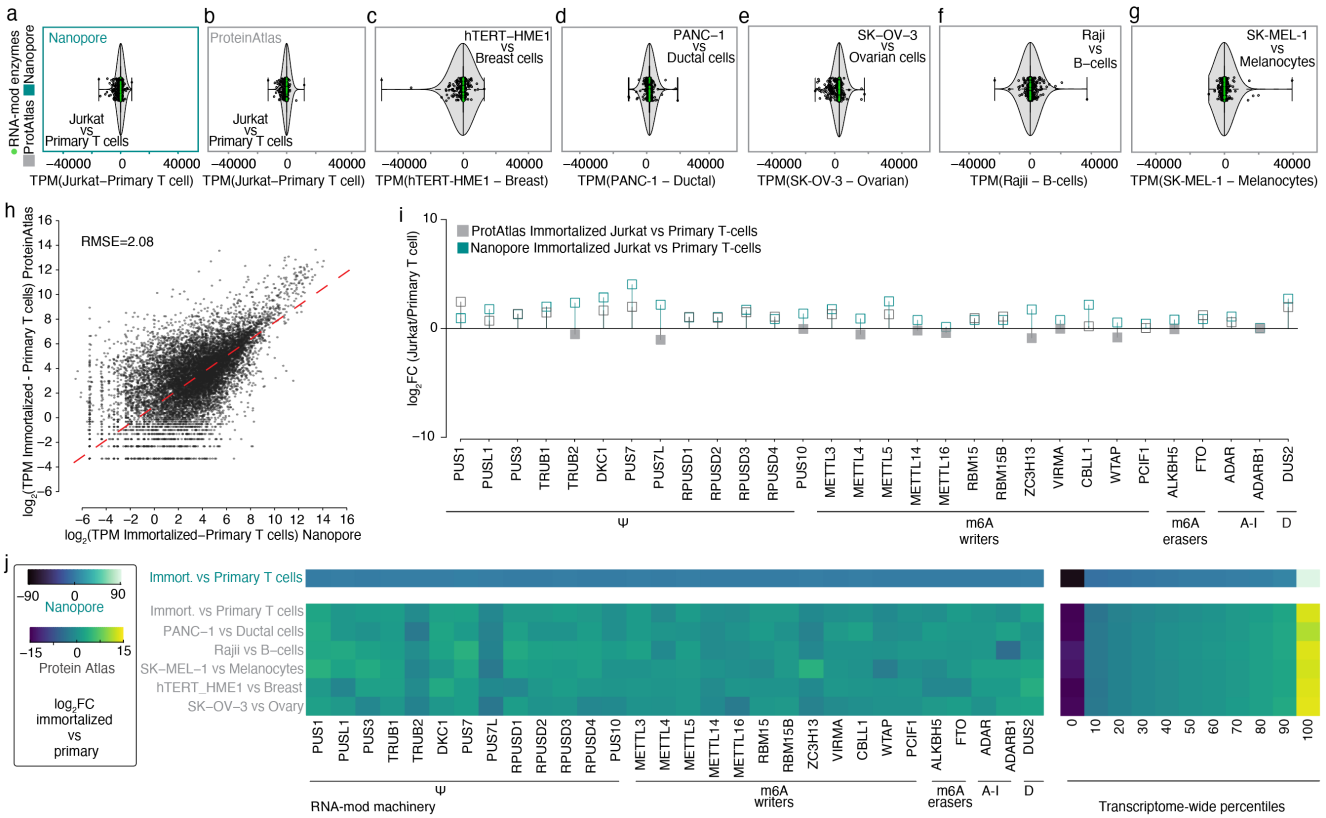
a. CLAP validation of unique psi sites found in Immortalized Jurkat T cells in co-expressed transcripts. **b.** gDNA IGV plots of the unique Immortalized psi-sites show SNVs absence as no mismatch is found.
c. Ionic current traces for the immortalized and primary transcripts centered at the unique psi-sites detected in immortalized T cells.
d. gDNA IGV plots of the unique primary psi-sites show SNVs absence as no mismatch is found.
c. Ionic current traces for the immortalized and primary transcripts centered at the unique psi-sites detected in primary T cells.



Supplementary Figure 5. Hypermodified type I sites in Immortalized T cell, primary T cells and HeLa.
Heatmap of hypermodified type I sites (% U-to-C mismatch >40) found in the three cell lines validated with orthogonal methods. For clarity we only show the top 5 hypermodified type I sites found in 2 out of 3 conditions. Sites surrounded by a TRUB1 motif are highlighted in red, sites surrounded by a PUS7 motif are highlighted in blue.



Supplementary Figure 6. Hypermodified type II modification on primary and immortalized T cells.
a. Correlation plots show the mean distance between psi-sites found on the same transcript vs the length of the transcript in log scale for primary naïve T cells.
b. Correlation plots show the mean distance between psi-sites found on the same transcript vs the length of the transcript in log scale for immortalized T cells.
c. Distribution of hypermodified type II sites across gene regions are shown for primary and immortalized T cells.



Supplementary Figure 7. RNA-modification machinery-associated gene expression levels are similar across primary and immortalized cell lines.

a. Transcriptome-wide TPM differences between immortalized and primary T cells using Nanopore DRS; RNA-modification enzymes are highlighted in green and TPM values are boxed in turquoise. **b.** TPM differences between immortalized and primary T cells from The Human Protein Atlas RNA-seq data (TPM values boxed in grey). **c.** TPM differences between immortalized and primary breast cells from The Human Protein Atlas RNA-seq data. **d.** TPM differences between immortalized and primary ductal cells from The Human Protein Atlas RNA-seq data. **e.** TPM differences between immortalized and primary ovarian cells from The Human Protein Atlas RNA-seq data. **f.** TPM differences between immortalized and primary B cells from The Human Protein Atlas RNA-seq data. **g.** TPM differences between immortalized and primary melanocytes from The Human Protein Atlas RNA-seq data. **h.** Correlation of Nanopore and Human Protein Atlas data for immortalized vs. primary T cells (relative abundance in log scale). **i.** Log2 fold changes of RNA modification machinery genes in immortalized vs. primary T cells (light grey: RNA-seq; turquoise: Nanopore; unfilled markers: >0; filled markers: <0). **j.** Expression levels of RNA-modification machinery relative to whole-transcriptome percentiles from Nanopore and Protein Atlas data.