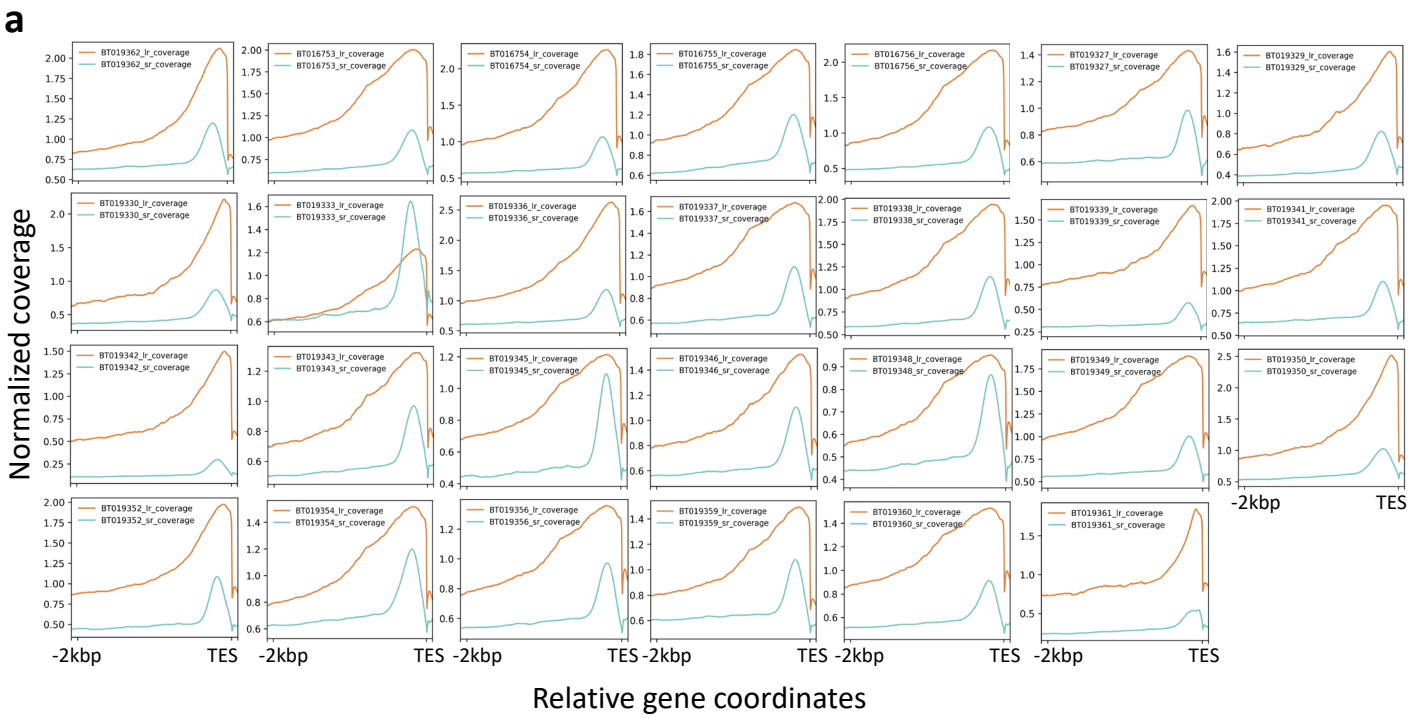
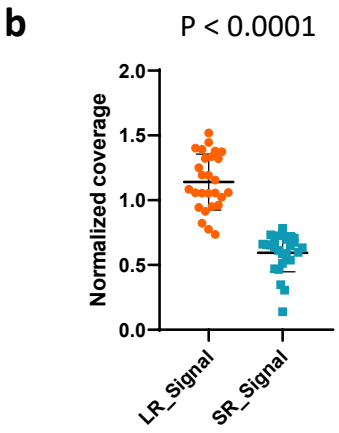
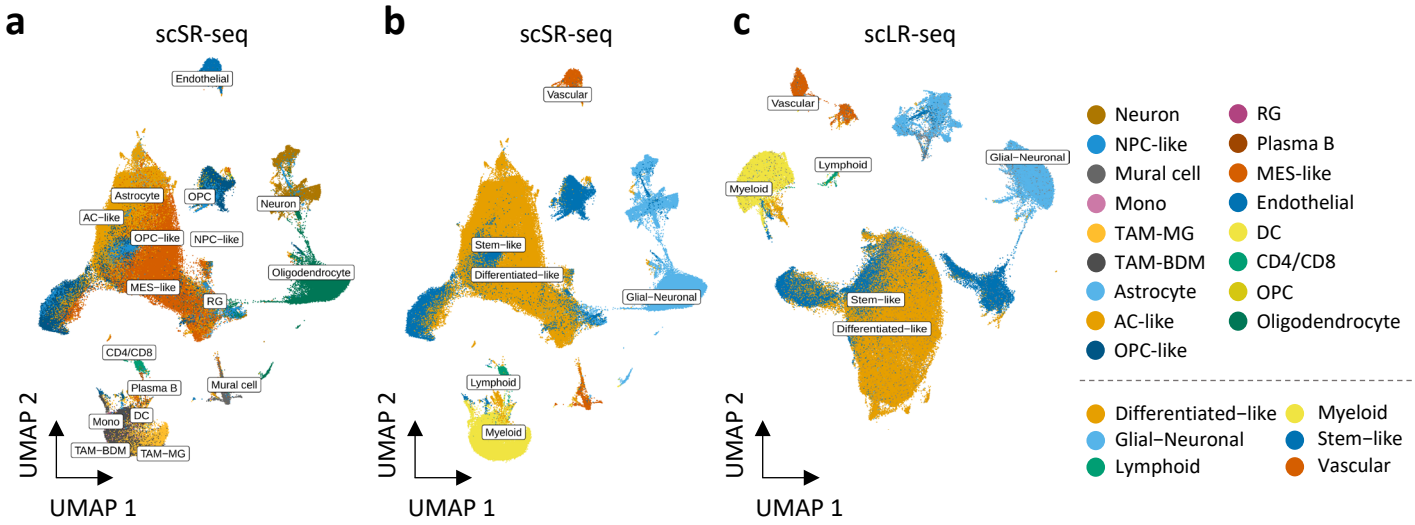
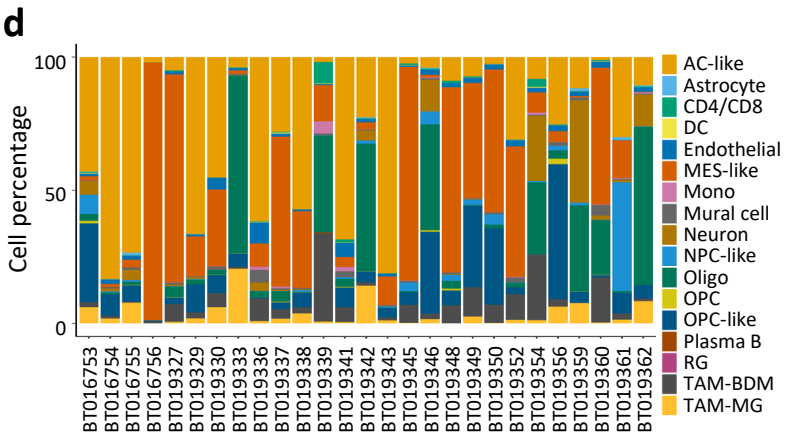
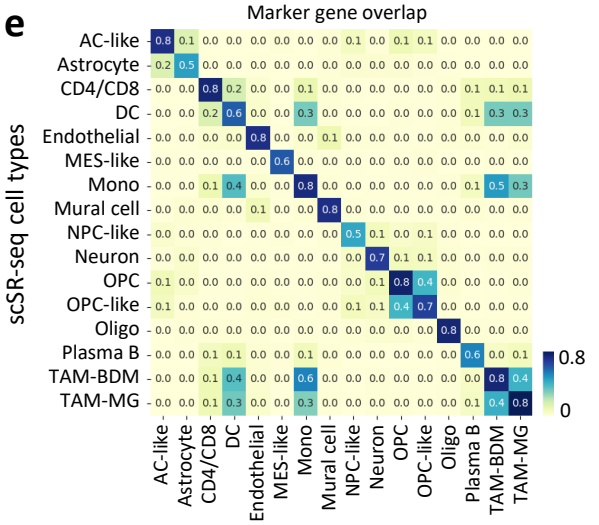


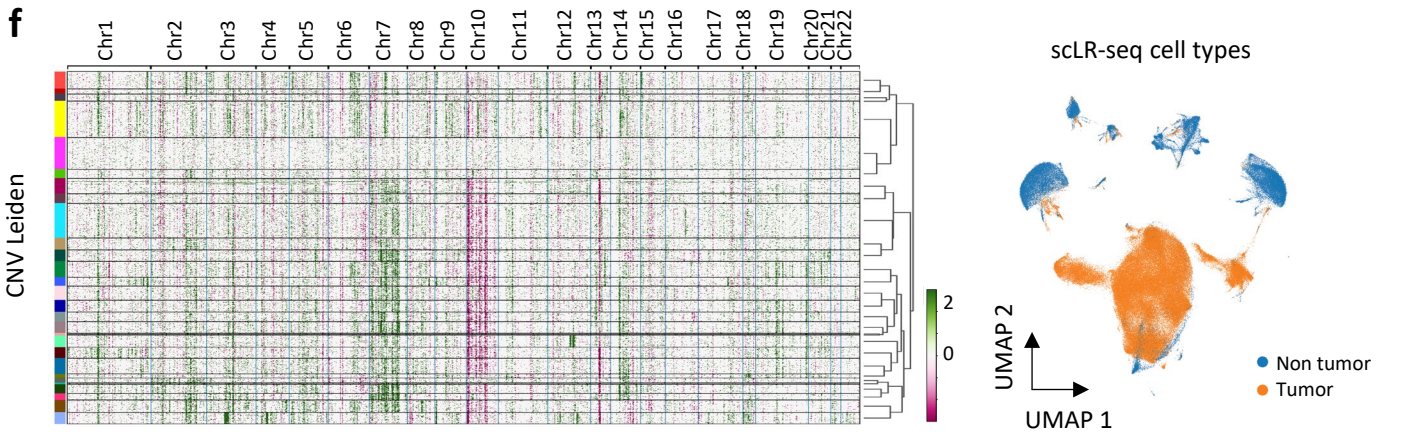


Supplementary Fig. 1. Normalized coverage comparison between scSR-seq and scLR-seq
(a) Read-depth–normalized coverage profiles from short-read and long-read BAM files (27 samples) spanning −2,000 bp upstream to the transcription end site (TES). (b) Dot plot showing normalized coverage distributions from TSS to TES (per 10 bp bin) for scSR-seq and scLR-seq (27 samples). Statistical significance between methods was assessed using a two-sided Mann–Whitney U test (exact P value). Data are presented as mean values ± SD.





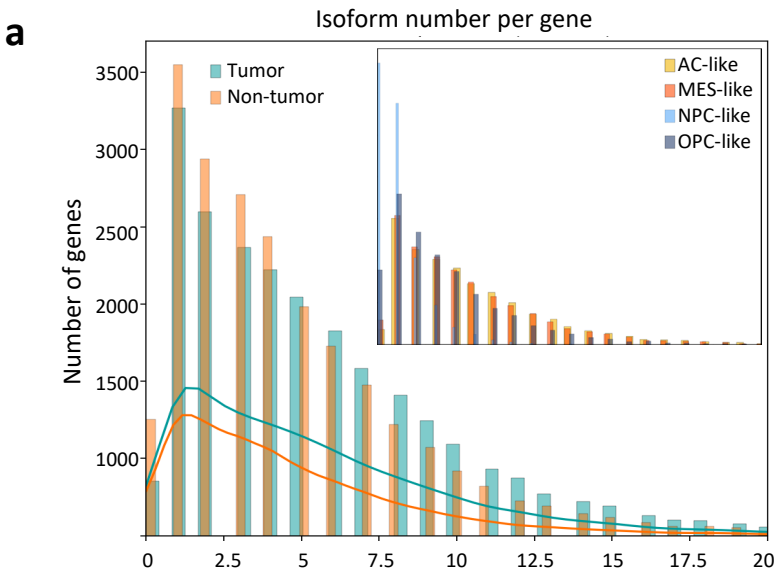


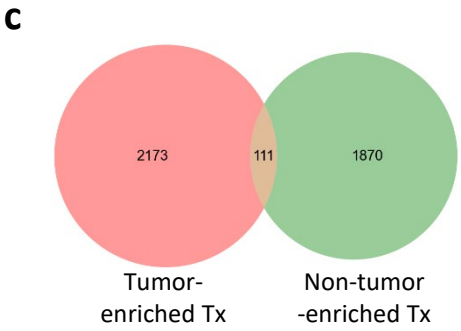


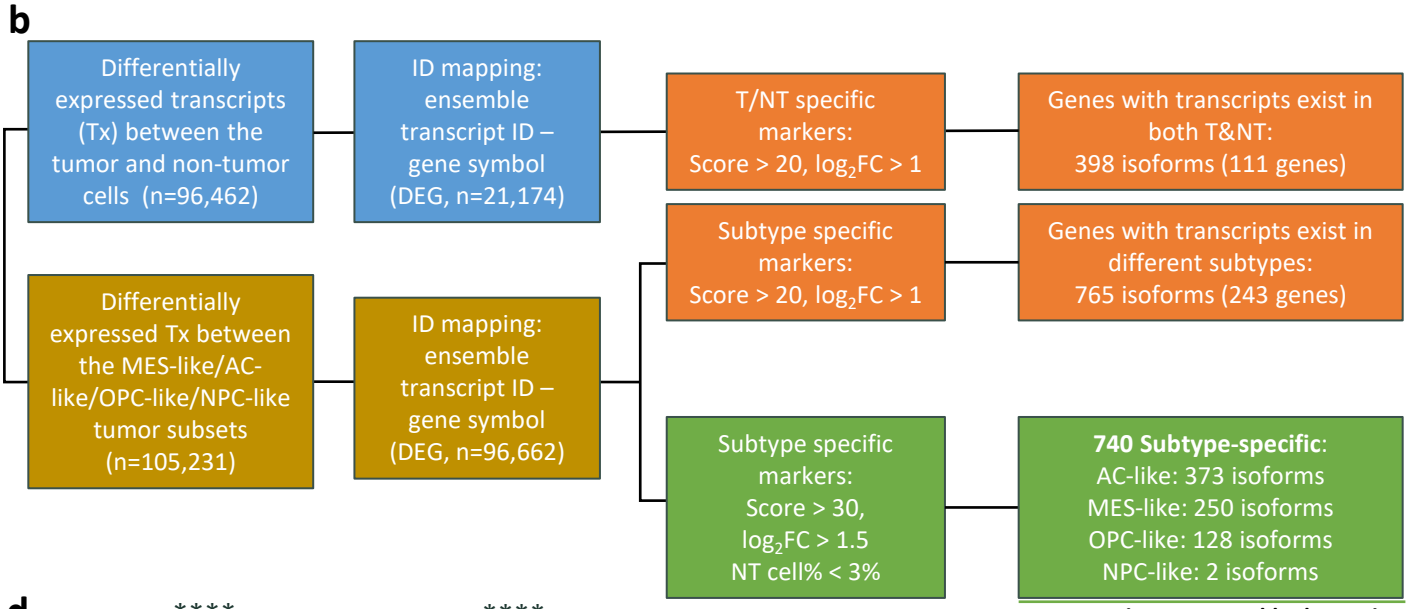
g

Method	Bio conservation					Batch correction					Aggregate score		
	Isolated labels	KMeans NMI	KMeans ARI	Silhouette label	cLISI	Silhouette batch	iLISI	KBET	Graph connectivity	PCR comparison	Batch correction	Bio conservation	Total
X_harmony scSR	0.52	0.54	0.30	0.53	0.99	0.83	0.07	0.26	0.65	0.68	0.50	0.58	0.55
X_harmony scLR	0.60	0.53	0.30	0.53	0.99	0.85	0.08	0.36	0.64	0.71	0.53	0.59	0.56

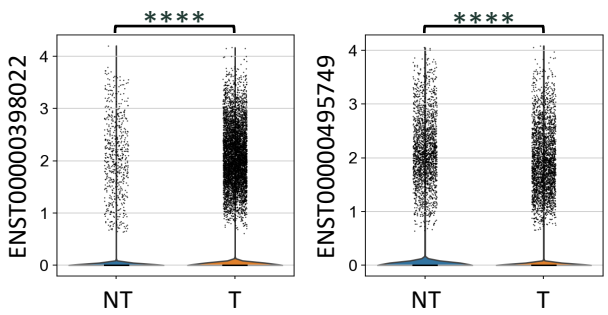
Supplementary Fig. 2. Cell clustering, annotation, and concordance between scSR-seq and scLR-seq
(a,b) UMAP projections of 182,222 single cells clustered using scSR-seq data, colored by fine-grained (a) and broad (b) cell type annotations (n=27 samples). (c) UMAP visualization of 182,222 single cells generated using scLR-seq, colored by broad cell type categories (n=27 samples). (d) Bar plot comparing cell type proportions across individual patients. (e) Heatmap showing Jaccard Similarity Indices for marker genes identified in scSR-seq versus scLR-seq. Scores were calculated based on the top 100 of significantly enriched markers for each cell state (AC-like: 56,795 cells; Astrocyte: 340 cells; CD4/CD8: 940 cells; DC: 142 cells; Endothelial: 3,661 cells; MES-like: 50,287 cells; Mono: 671 cells; Mural: 1,875 cells; NPC-like: 4,565 cells; Neuron: 8,485 cells; OPC: 414 cells; OPC-like: 18,912 cells; Oligodendrocyte: 19,373 cells; Plasma B: 59 cells; TAM-BDM: 9,753 cells; TAM-MG: 5,943 cells). (f) Copy-number variation profiling across individual chromosomes and corresponding UMAP projection of scLR-seq data, with cells (182,222 from 27 samples) classified as normal or malignant based on inferCNV. (g) Dot plot showing Biological Conservation Scores and Batch Correction Scores between scSR-seq and scLR-seq modalities.

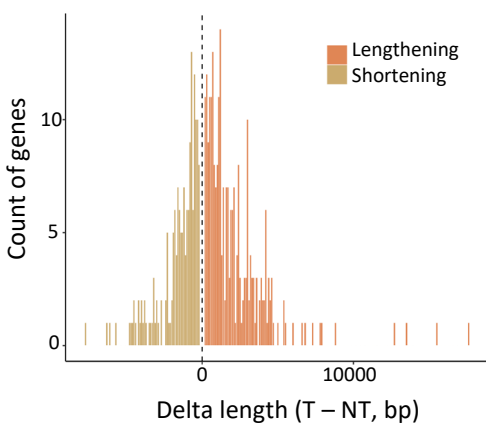


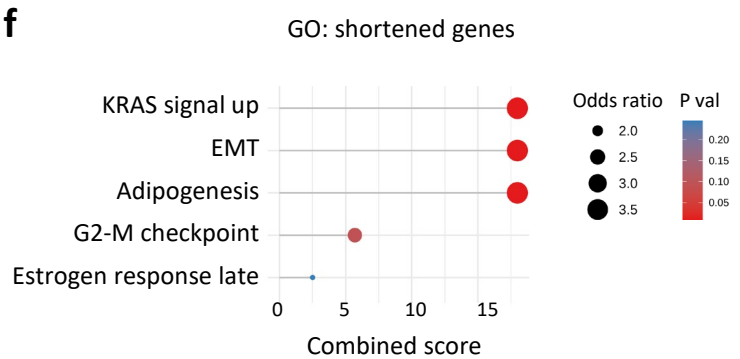




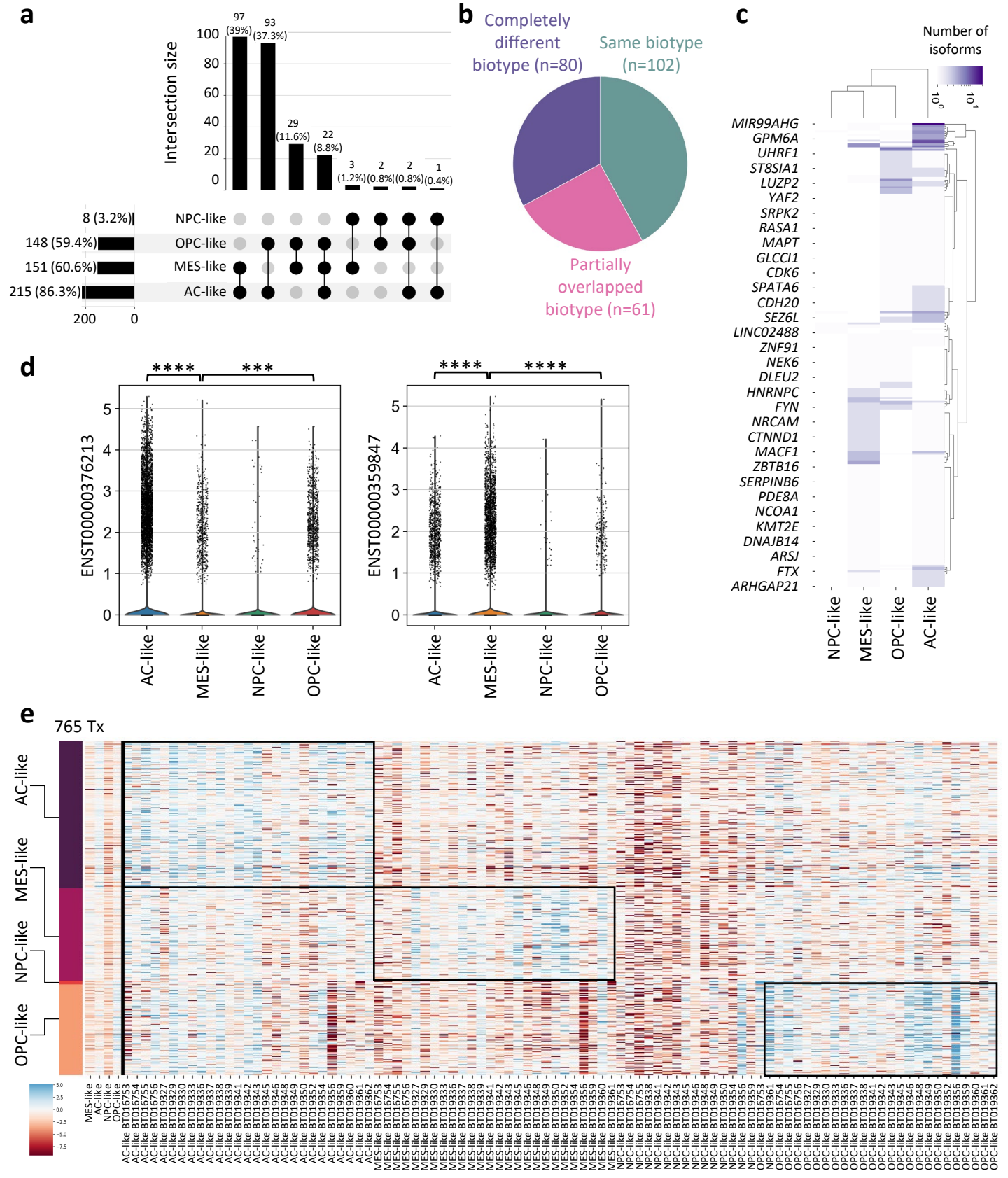
Transcript Type	Ensembl Canonical	Alternative Isoform
Protein-coding	191	213
lncRNA	20	86
Retained intron	---	107
NMD	---	34
Not defined	---	89



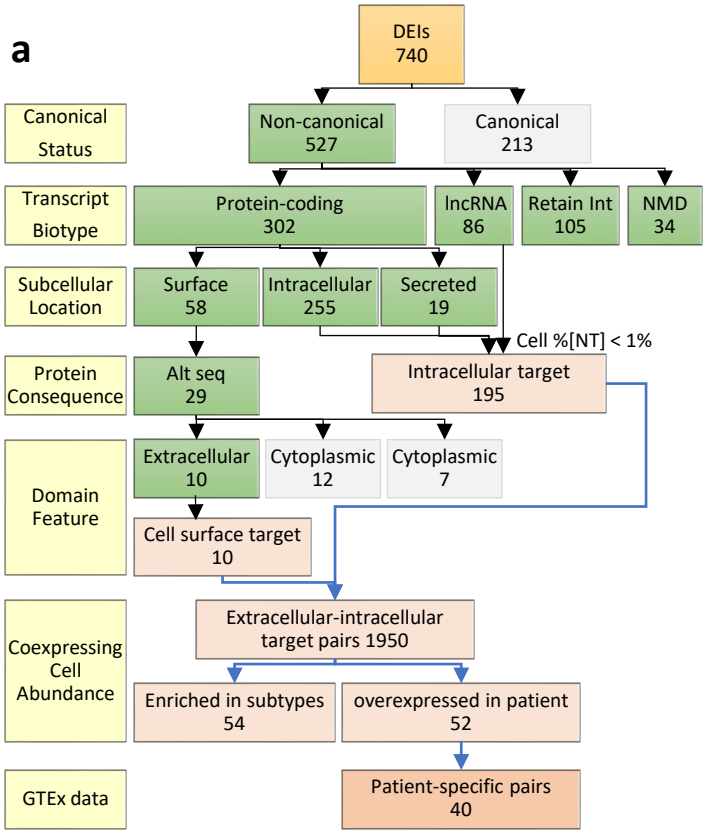


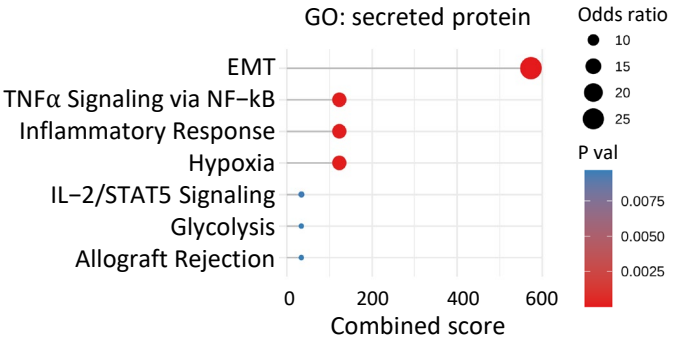


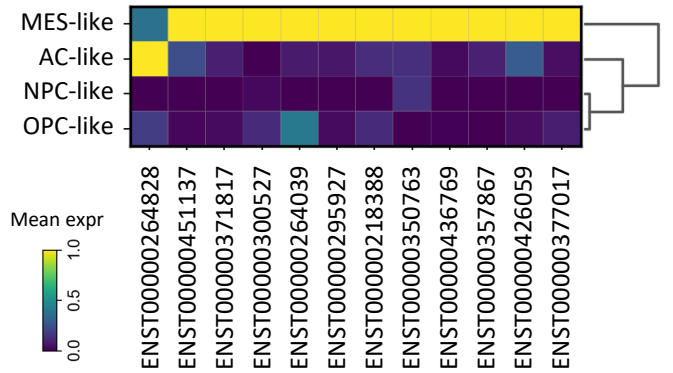
Supplementary Fig. 3. Transcript diversity and differential transcript usage analysis in GBM
(a) Bar plot showing the number of expressed transcripts per gene in tumor versus non-tumor cells (isoforms expressed in at least 10 cells), as well as across malignant subtypes. (b) Schematic workflow for detecting DTU between tumor and non-tumor cells, as well as across malignant subtypes including AC-like, MES-like, NPC-like, and OPC-like cells. The pipeline also identifies subtype-specific marker isoforms. (c) Venn diagram showing genes exhibiting differential transcript usage between tumor and non-tumor cells. The overlapping region indicates genes for which distinct isoforms are detected as differentially expressed in both tumor and non-tumor populations. (d) Dot plot showing the expression of *TM9SF4* isoforms in tumor (130,559) and non-tumor (51,663) cells. (e) Histogram showing the number of genes and corresponding changes in 3′ UTR length between tumor (130,559) and non-tumor (51,663) cells. (f) Top enriched gene ontology terms (MSigDB: Hallmark gene sets) for 171 genes with shortened 3′ UTRs in 27 samples. Statistical significance was assessed using a one-sided Fisher's exact test.

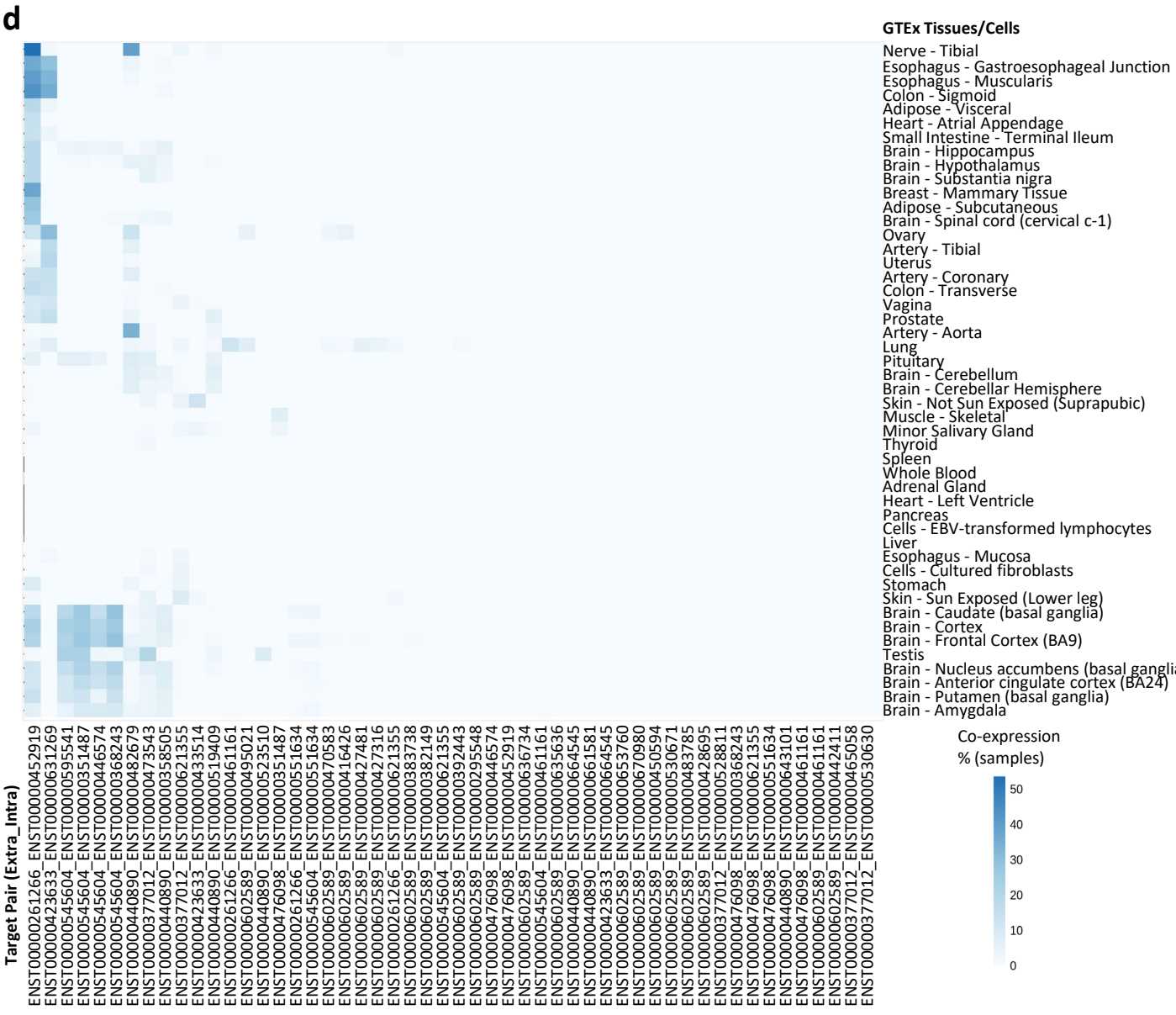


Supplementary Fig. 4. Isoform diversity and subtype-specific transcript usage in malignant GBM cells
(a) UpSet plot showing genes with expressed transcripts shared across different malignant subtypes. (b) Pie chart categorizing these genes based on whether they exhibit completely distinct biotypes, partially overlapping biotypes, or the same biotype across malignant subtypes. (c) Heatmap showing the number of expressed isoforms per gene across tumor subtypes. Each row represents a gene, and the color scale indicates the number of isoforms detected within each subtype. (d) Dot plot showing the expression of *NTRK2* isoforms across tumor subtypes. (e) Heatmap showing mean expression scores of 765 differentially used transcripts across malignant subtypes for each individual patient.

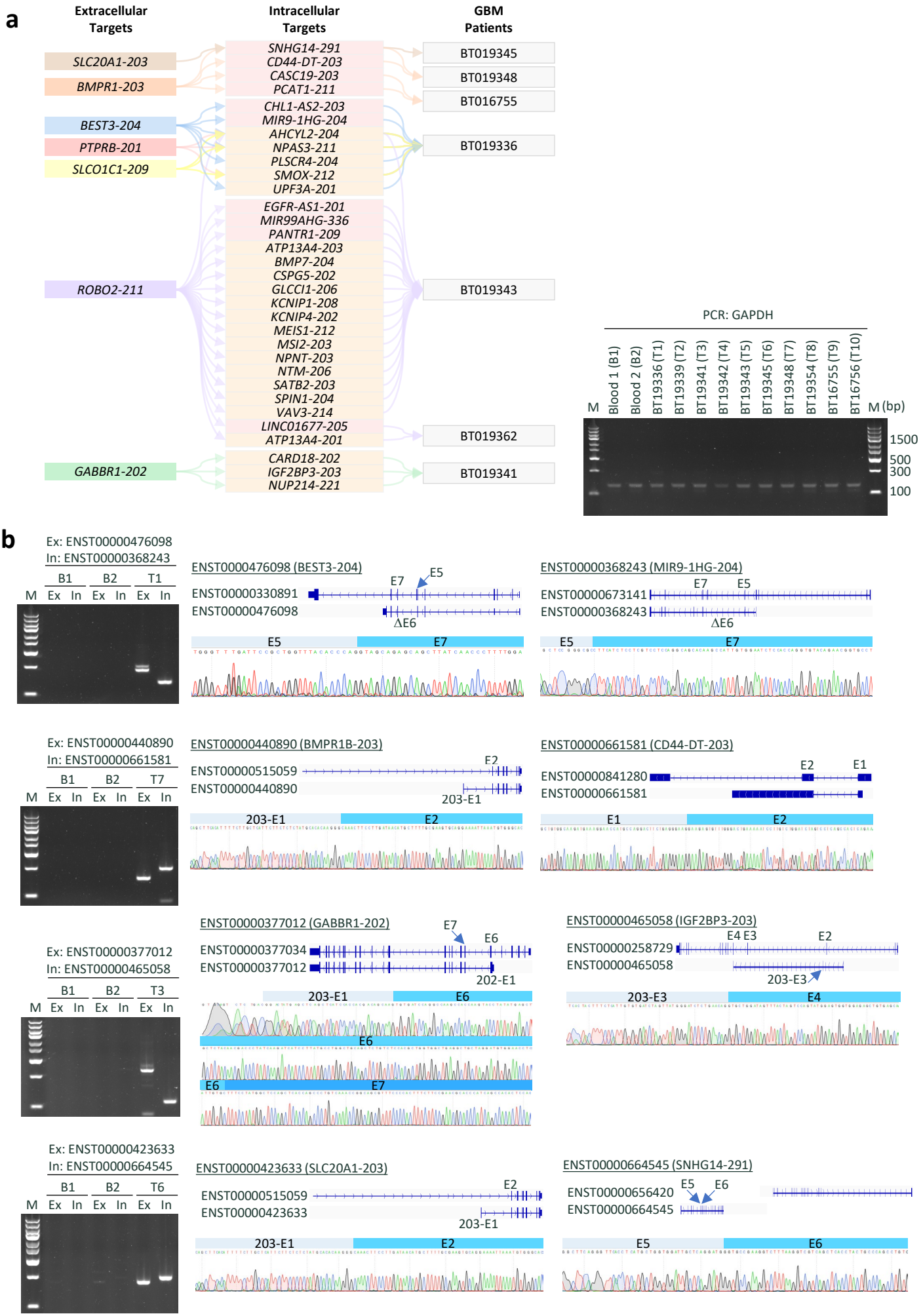

a
DEIs
740
Canonical Status
Non-canonical
527
Canonical
213
Transcript Biotype
Protein-coding
302
lncRNA
86
Retain Int
105
NMD
34
Subcellular Location
Surface
58
Intracellular
255
Secreted
19
Cell %[NT] < 1%
Protein Consequence
Alt seq
29
Intracellular target
195
Domain Feature
Extracellular
10
Cytoplasmic
12
Cytoplasmic
7
Cell surface target
10
Coexpressing Cell Abundance
Extracellular-intracellular target pairs 1950
Enriched in subtypes
54
overexpressed in patient
52
GTEx data
Patient-specific pairs
40


b
GO: secreted protein
EMT
TNFα Signaling via NF−kB
Inflammatory Response
Hypoxia
IL−2/STAT5 Signaling
Glycolysis
Allograft Rejection
0
200
400
600
Combined score
Odds ratio
10
15
20
25
P val
0.0075
0.0050
0.0025


c
MES-like
AC-like
NPC-like
OPC-like
Mean expr
0.0
0.5
1.0
ENST00000264828
ENST00000451137
ENST00000371817
ENST00000300527
ENST00000264039
ENST00000295927
ENST00000218388
ENST00000350763
ENST00000436769
ENST00000357867
ENST00000426059
ENST00000377017

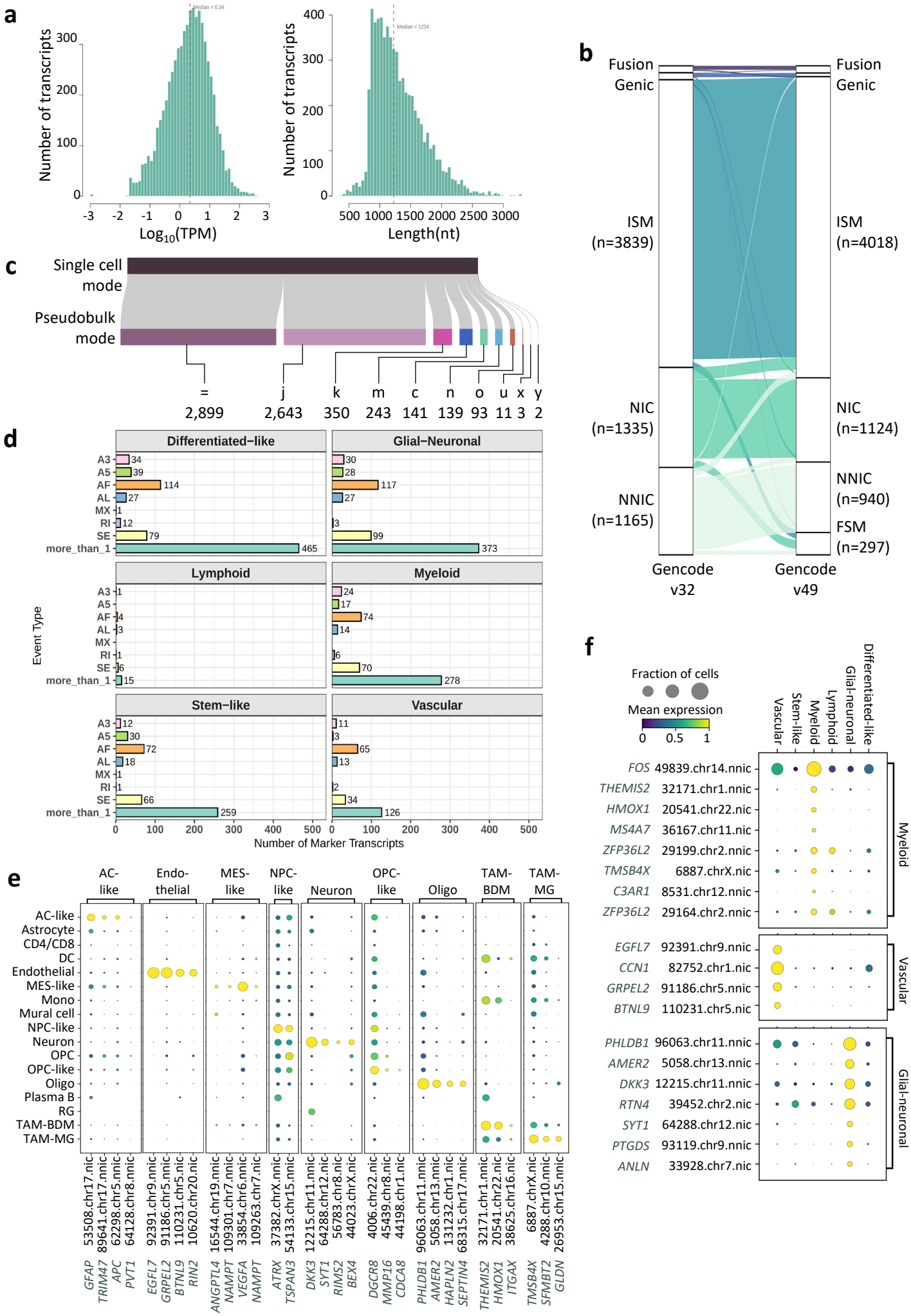

d
GTEx Tissues/Cells
Nerve - Tibial
Esophagus - Gastroesophageal Junction
Esophagus - Muscularis
Colon - Sigmoid
Adipose - Visceral
Heart - Atrial Appendage
Small Intestine - Terminal Ileum
Brain - Hippocampus
Brain - Hypothalamus
Brain - Substantia nigra
Breast - Mammary Tissue
Adipose - Subcutaneous
Brain - Spinal cord (cervical c-1)
Ovary
Artery - Tibial
Uterus
Artery - Coronary
Colon - Transverse
Vagina
Prostate
Artery - Aorta
Lung
Pituitary
Brain - Cerebellum
Brain - Cerebellar Hemisphere
Skin - Not Sun Exposed (Suprapubic)
Muscle - Skeletal
Minor Salivary Gland
Thyroid
Spleen
Whole Blood
Adrenal Gland
Heart - Left Ventricle
Pancreas
Cells - EBV-transformed lymphocytes
Liver
Esophagus - Mucosa
Cells - Cultured fibroblasts
Stomach
Skin - Sun Exposed (Lower leg)
Brain - Caudate (basal ganglia)
Brain - Cortex
Brain - Frontal Cortex (BA9)
Testis
Brain - Nucleus accumbens (basal ganglia)
Brain - Anterior cingulate cortex (BA24)
Brain - Putamen (basal ganglia)
Brain - Amygdala
Co-expression
% (samples)
50
40
30
20
10
0
Target Pair (Extra_Intra)

Supplementary Fig. 5. Workflow and validation for prioritizing extracellular–intracellular target pairs in GBM
(a) Schematic illustrating the multi-step workflow used to identify and prioritize DEIs for patient-specific and subtype-specific target pairs. (b) Top enriched gene ontology terms (MSigDB Hallmark gene sets) for the 61 identified secreted proteins. Statistical significance was assessed using a one-sided Fisher's exact test. (c) Heatmap showing the expression of EMT-associated secreted proteins across four cell groups (AC-like: 56,795 cells; MES-like: 50,287 cells; OPC-like: 18,912 cells; NPC-like: 4,565 cells).
(d) The heatmap presents a descriptive analysis of gene co-occurrence frequencies across human tissues. It shows the percentage of GTEx samples co-expressing each of the 52 prioritized target pairs across normal human tissues, based on bulk RNA-sequencing data. The heatmap was generated using hierarchical clustering of the proportion of samples per tissue exhibiting concurrent expression of both genes, defined as TPM > 1. Tissues with fewer than 100 samples in the GTEx v8 dataset were excluded to ensure robust percentage estimates.

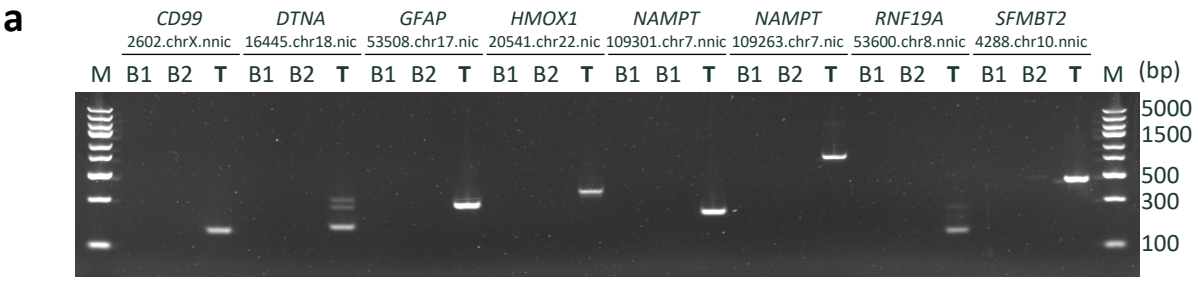

a
Extracellular Targets
Intracellular Targets
GBM Patients
SLC20A1-203
BMPR1-203
BEST3-204
PTPRB-201
SLCO1C1-209
ROBO2-211
GABBR1-202
SNHG14-291
CD44-DT-203
CASC19-203
PCAT1-211
CHL1-AS2-203
MIR9-1HG-204
AHCYL2-204
NPAS3-211
PLSCR4-204
SMOX-212
UPF3A-201
EGFR-AS1-201
MIR99AHG-336
PANTR1-209
ATP13A4-203
BMP7-204
CSPG5-202
GLCCI1-206
KCNIP1-208
KCNIP4-202
MEIS1-212
MSI2-203
NPNT-203
NTM-206
SATB2-203
SPIN1-204
VAV3-214
LINC01677-205
ATP13A4-201
CARD18-202
IGF2BP3-203
NUP214-221
BT019345
BT019348
BT016755
BT019336
BT019343
BT019362
BT019341
PCR: GAPDH
M
Blood 1 (B1)
Blood 2 (B2)
BT19336 (T1)
BT19339 (T2)
BT19341 (T3)
BT19342 (T4)
BT19343 (T5)
BT19345 (T6)
BT19348 (T7)
BT19354 (T8)
BT16755 (T9)
BT16756 (T10)
M (bp)
1500
500
300
100
b
Ex: ENST00000476098
In: ENST00000368243
B1
B2
T1
M Ex In Ex In Ex In
ENST00000476098 (BEST3-204)
ENST00000330891
ENST00000476098
E7
E5
ΔE6
E5
E7
ENST00000368243 (MIR9-1HG-204)
ENST00000673141
ENST00000368243
E7
E5
ΔE6
E5
E7
Ex: ENST00000440890
In: ENST00000661581
B1
B2
T7
M Ex In Ex In Ex In
ENST00000440890 (BMPR1B-203)
ENST00000515059
ENST00000440890
E2
203-E1
203-E1
E2
ENST00000661581 (CD44-DT-203)
ENST00000841280
ENST00000661581
E2
E1
E1
E2
Ex: ENST00000377012
In: ENST00000465058
B1
B2
T3
M Ex In Ex In Ex In
ENST00000377012 (GABBR1-202)
ENST00000377034
ENST00000377012
E7
E6
202-E1
203-E1
E6
E6
E7
ENST00000465058 (IGF2BP3-203)
ENST00000258729
ENST00000465058
E4 E3
E2
203-E3
203-E3
E4
Ex: ENST00000423633
In: ENST00000664545
B1
B2
T6
M Ex In Ex In Ex In
ENST00000423633 (SLC20A1-203)
ENST00000515059
ENST00000423633
E2
203-E1
203-E1
E2
ENST00000664545 (SNHG14-291)
ENST00000656420
ENST00000664545
E5
E6
E5
E6

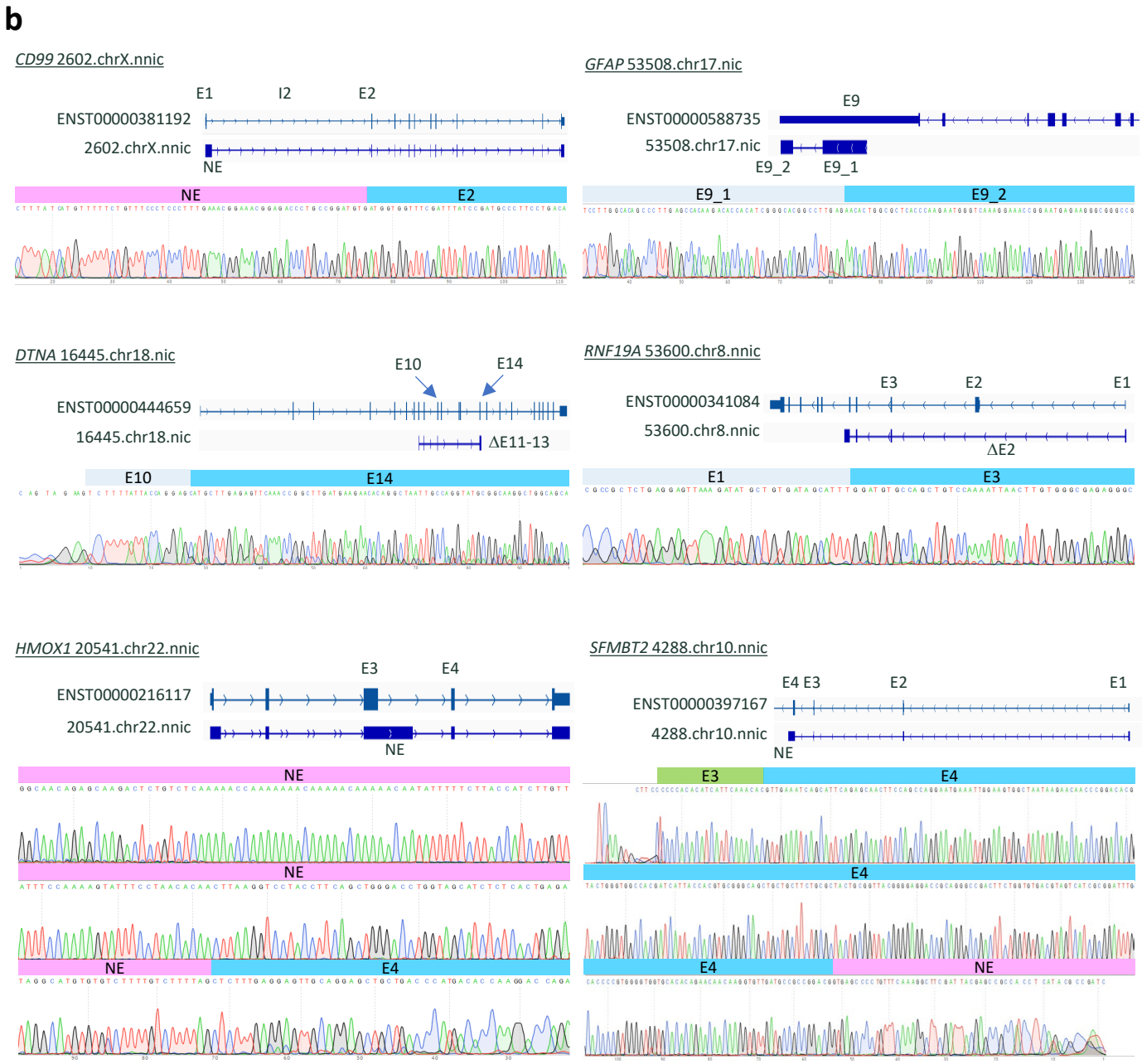
Supplementary Fig. 6. PCR and Sanger sequencing validation of selected extracellular–intracellular target pairs in tumor samples and control blood

(a) Schematic representation of extracellular-intracellular target pair combinations identified across GBM patients. The diagram illustrates patient-specific pairing patterns between extracellular targets (left) and intracellular targets (middle), with connecting lines indicating the pairing relationships. (b) Sanger sequencing was performed as an orthogonal validation of selected junctions of extracellular–intracellular target pairs in tumor samples and control blood. The experiment was performed once, and the chromatograms and Sanger images are shown. E, exon; Ex, extracellular; In, intracellular; B, blood control; T: tumor sample.


a
Number of transcripts
Median = 0.34
Log10(TPM)
Median = 1224
Length(nt)
b
Fusion
Genic
ISM
(n=3839)
NIC
(n=1335)
NNIC
(n=1165)
ISM
(n=4018)
NIC
(n=1124)
NNIC
(n=940)
FSM
(n=297)
Gencode
v32
Gencode
v49
c
Single cell
mode
Pseudobulk
mode
= j k m c n o u x y
2,899 2,643 350 243 141 139 93 11 3 2
d
Differentiated−like
Glial−Neuronal
Lymphoid
Myeloid
Stem−like
Vascular
Event Type
A3 A5 AF AL MX RI SE more_than_1
Number of Marker Transcripts
e
AC-like
Endo-thelial
MES-like
NPC-like
Neuron
OPC-like
Oligo
TAM-BDM
TAM-MG
AC-like
Astrocyte
CD4/CD8
DC
Endothelial
MES-like
Mono
Mural cell
NPC-like
Neuron
OPC
OPC-like
Oligo
Plasma B
RG
TAM-BDM
TAM-MG
GFAP TRIM47 APC PVT1
EGFL7 GRPEL2 BTNL9 RIN2
ANGPTL4 NAMPT VEGFA NAMPT
ATRX TSPAN3
DKK3 SYT1 RIMS2 BEX4
DGCR8 MMP16 CDCA8
PHLDB1 AMER2 HAPLN2 SEPTIN4
THEMIS2 HMOX1 ITGAX
TMSB4X SFMBT2 GLDN
f
Fraction of cells
Mean expression
0 0.5 1
Vascular
Stem-like
Myeloid
Lymphoid
Glial-neuronal
Differentiated-like
FOS 49839.chr14.nnic
THEMIS2 32171.chr1.nnic
HMOX1 20541.chr22.nic
MS4A7 36167.chr11.nic
ZFP36L2 29199.chr2.nnic
TMSB4X 6887.chrX.nic
C3AR1 8531.chr12.nnic
ZFP36L2 29164.chr2.nnic
EGFL7 92391.chr9.nnic
CCN1 82752.chr1.nic
GRPEL2 91186.chr5.nnic
BTNL9 110231.chr5.nic
PHLDB1 96063.chr11.nnic
AMER2 5058.chr13.nnic
DKK3 12215.chr11.nnic
RTN4 39452.chr2.nnic
SYT1 64288.chr12.nic
PTGDS 93119.chr9.nnic
ANLN 33928.chr7.nic
Myeloid
Vascular
Glial-neuronal

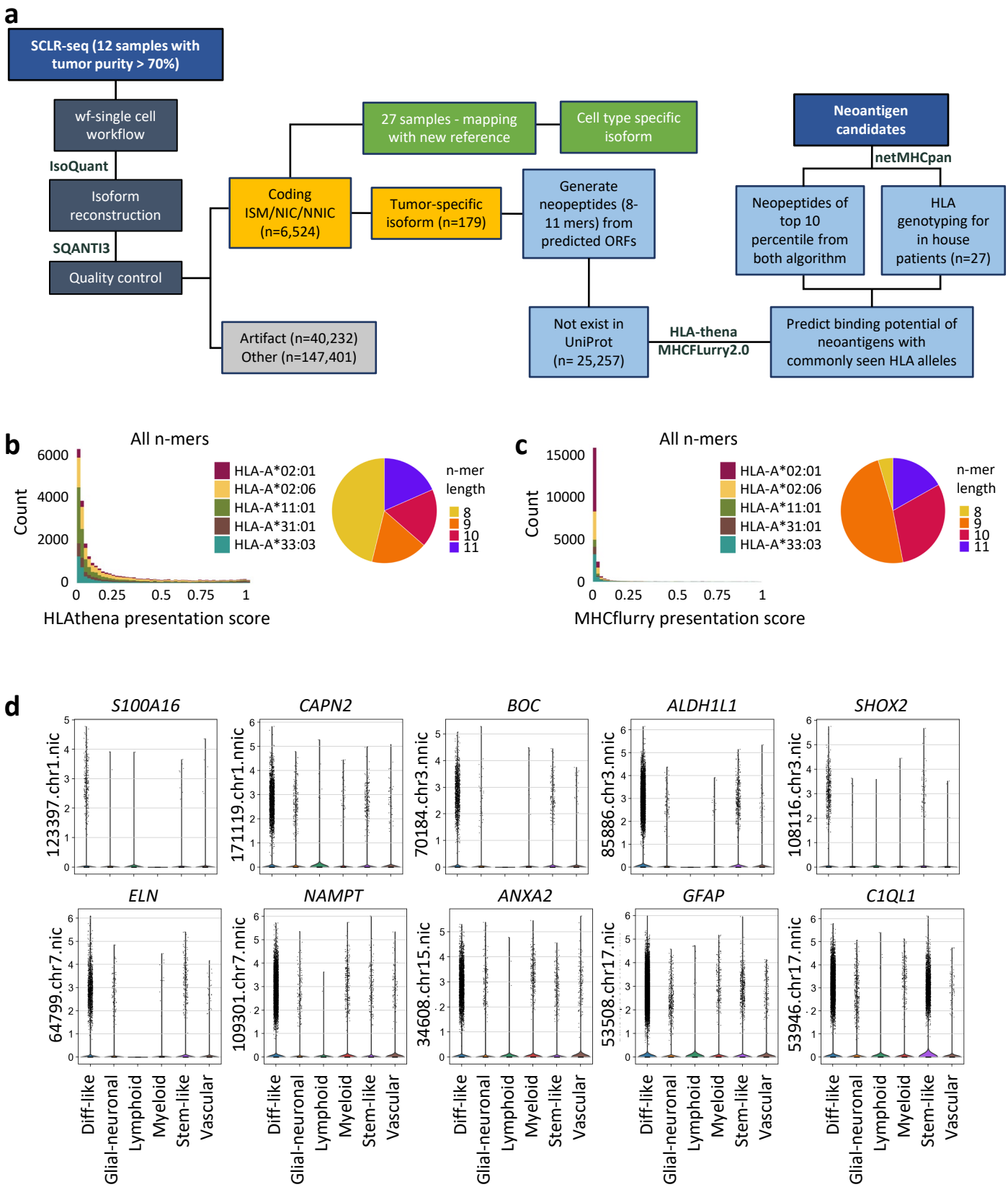
Supplementary Fig. 7. Characterization of isoforms not present in existing annotations and their cell-type-specific expression (a) Histograms showing the distribution of transcript abundance (TPM, left) and transcript length (right) across all isoforms in 27 samples. (b) Sankey plot illustrating changes in transcript structural annotation for 6,524 transcripts between GENCODE v32 and v49. (c) Sankey plot summarizing classification of 6,524 isoforms reconstructed in barcode-aware single-cell mode relative to isoforms reconstructed from pseudobulk (sample-level aggregated) reads. Class codes indicate the structural relationship between single-cell–derived isoforms and pseudobulk models: "=" denotes exact intron-chain matches; "j" indicates isoforms sharing at least one splice junction match; "k" denotes containment of reference; "m" indicates retained matched intron(s); "c" represents containment within a larger transcript; "n" denotes retained novel intron structure; "o" indicates generic exon overlap on the same strand; "u" represents unknown/intergenic transcripts; "x" denotes overlap with exons on the opposite strand; and "y" indicates transcripts containing a reference within its intron. (d) Bar plots showing the distribution of splicing event types stratified by major cell types (Differentiated-like: 108,300; Glial-Neuronal: 28,049; Stem-like: 22,831; Myeloid: 16,535; Vascular: 5,511; Lymphoid: 996 cells) in 27 samples. (e,f) Dot plots highlighting the top differentially expressed isoforms across fine-resolution cell states (e) and broader non-malignant cell populations, including myeloid, vascular, and glial–neuronal cells in 27 samples (f). Differential expression was evaluated using a two-sided t-test with Benjamini–Hochberg correction, and isoforms with adjusted $P < 0.05$ are shown.





Supplementary Fig. 8. Validation of unannotated isoforms by PCR and Sanger sequencing

(a,b) Sanger sequencing was performed as an orthogonal validation of selected unannotated splice junction in another six newly identified isoforms. The experiment was performed once, and the chromatograms (a) and Sanger images (b) are shown. E, exon; NE: neo exon; B, blood control; T: tumor sample.



Supplementary Fig. 9. Prediction and characterization of peptides derived from isoforms not present in existing annotations (a) Schematic workflow outlining generation of peptides from isoforms and prediction of their MHC-binding affinities. (b,c) Histograms and pie charts showing the distribution of peptide presentation scores by HLA allele (left) and peptide length (right) as predicted by HLAthena (b) and MHCflurry 2.0 (c). (d) Dot plot displaying expression patterns of select isoforms across major GBM cell types.