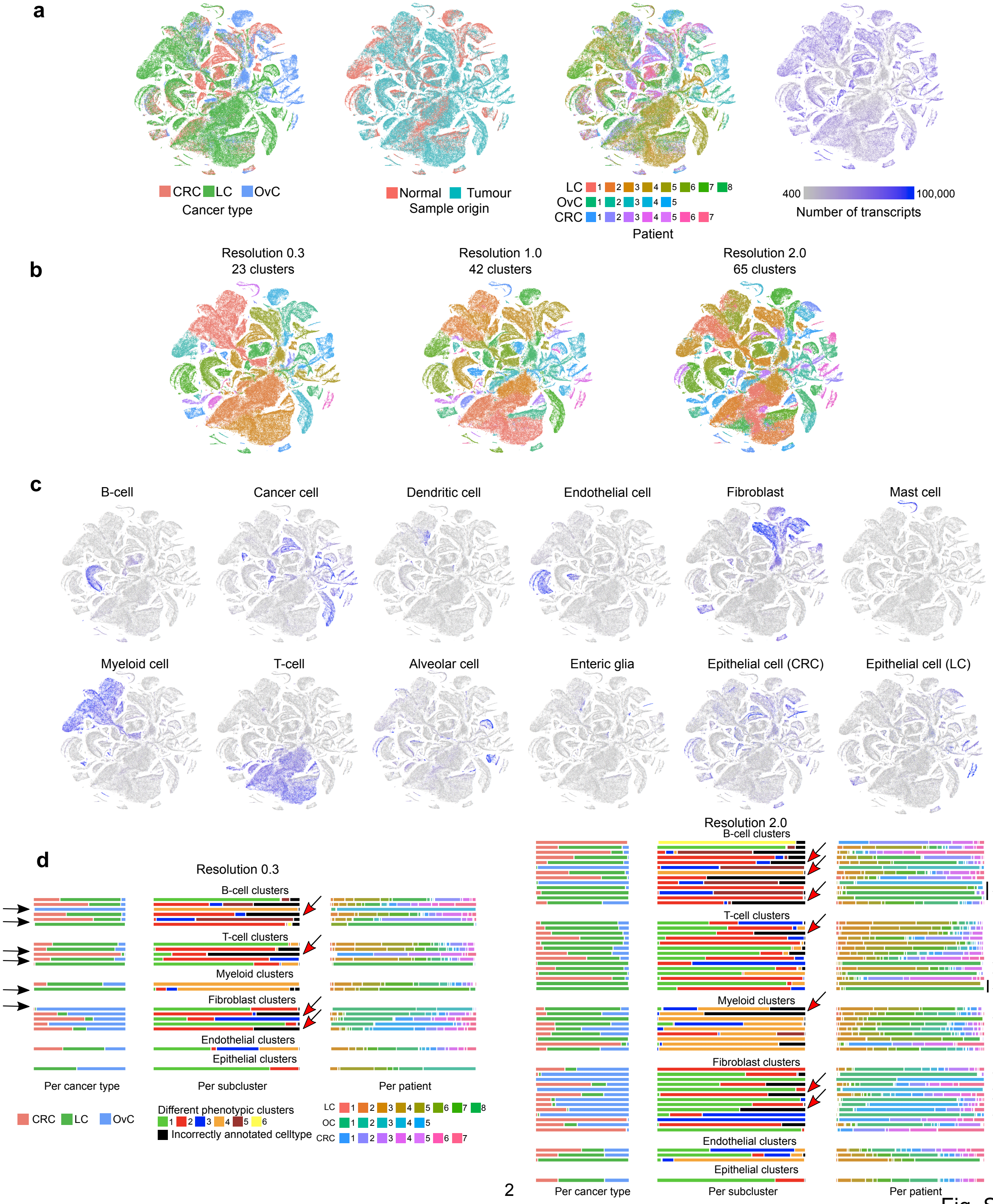
a
CRC LC OvC
Cancer type
Normal Tumour
Sample origin
LC 1 2 3 4 5 6 7 8
OvC 1 2 3 4 5
CRC 1 2 3 4 5 6 7
Patient
400 100,000
Number of transcripts
b
Resolution 0.3
23 clusters
Resolution 1.0
42 clusters
Resolution 2.0
65 clusters
c
B-cell
Cancer cell
Dendritic cell
Endothelial cell
Fibroblast
Mast cell
Myeloid cell
T-cell
Alveolar cell
Enteric glia
Epithelial cell (CRC)
Epithelial cell (LC)
d
Resolution 0.3
B-cell clusters
T-cell clusters
Myeloid clusters
Fibroblast clusters
Endothelial clusters
Epithelial clusters
Per cancer type
Per subcluster
Per patient
CRC LC OvC
Different phenotypic clusters
1 2 3 4 5 6
Incorrectly annotated celltype
LC 1 2 3 4 5 6 7 8
OC 1 2 3 4 5
CRC 1 2 3 4 5 6 7
Resolution 2.0
B-cell clusters
T-cell clusters
Myeloid clusters
Fibroblast clusters
Endothelial clusters
Epithelial clusters
Per cancer type
Per subcluster
Per patient

Fig. S1a-d

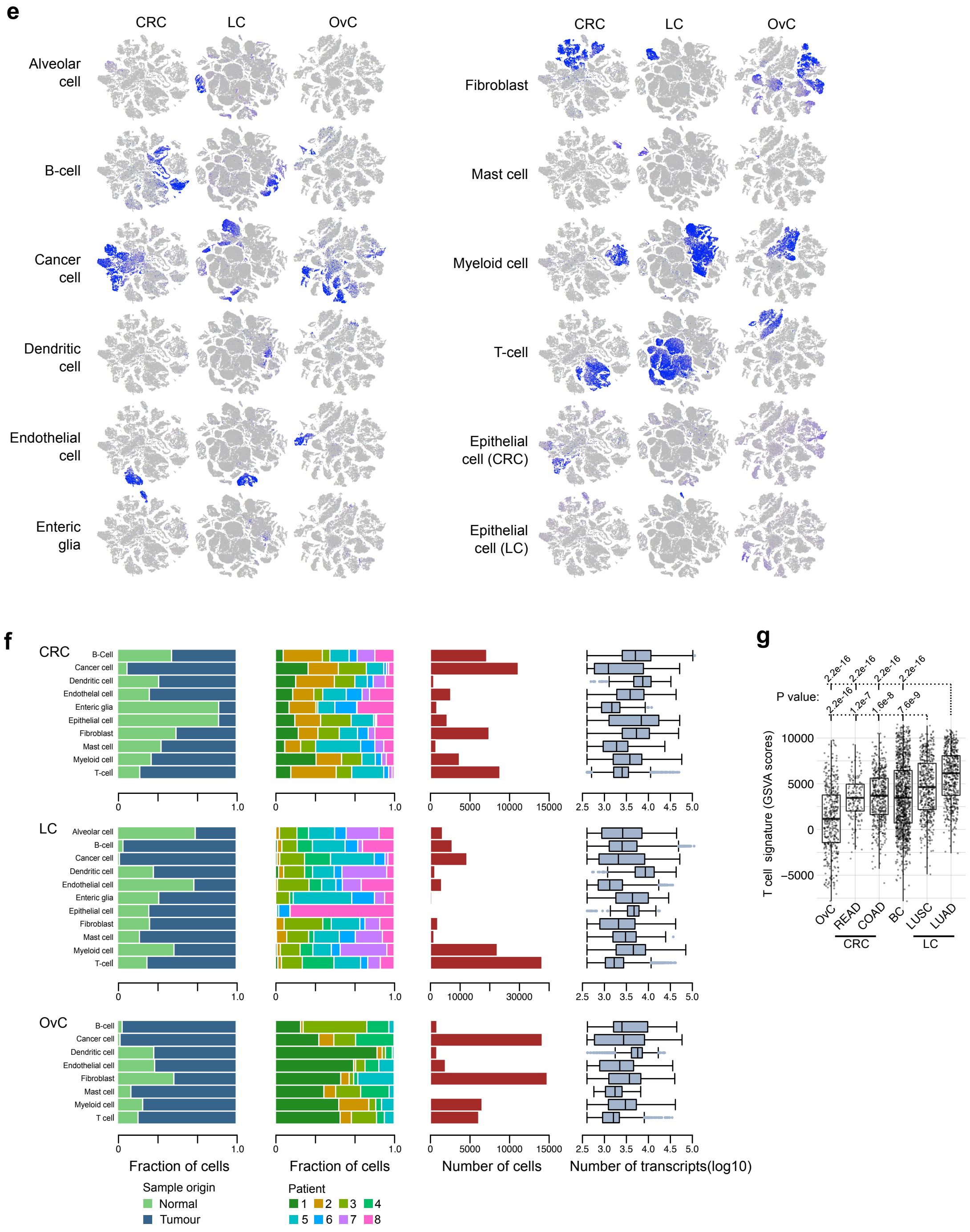
e
CRC
LC
OvC
Alveolar cell
B-cell
Cancer cell
Dendritic cell
Endothelial cell
Enteric glia
Fibroblast
Mast cell
Myeloid cell
T-cell
Epithelial cell (CRC)
Epithelial cell (LC)
f
CRC
B-Cell
Cancer cell
Dendritic cell
Endothelial cell
Enteric glia
Epithelial cell
Fibroblast
Mast cell
Myeloid cell
T-cell
0
1.0
0 5000 10000 15000
2.5 3.0 3.5 4.0 4.5 5.0
LC
Alveolar cell
B-cell
Cancer cell
Dendritic cell
Endothelial cell
Enteric glia
Epithelial cell
Fibroblast
Mast cell
Myeloid cell
T-cell
0 10000 30000
OvC
B-cell
Cancer cell
Dendritic cell
Endothelial cell
Fibroblast
Mast cell
Myeloid cell
T cell
Fraction of cells
Fraction of cells
Number of cells
Number of transcripts(log10)
Sample origin
🟩 Normal
🟦 Tumour
Patient
1 2 3 4
5 6 7 8
g
P value:
2.2e-16
2.2e-16
2.2e-16
2.2e-16
2.2e-16
1.2e-7
1.6e-8
7.6e-9
T cell signature (GSVA scores)
10000
5000
0
−5000
OvC
READ
COAD
BC
LUSC
LUAD
CRC
LC

Fig. S1e-g

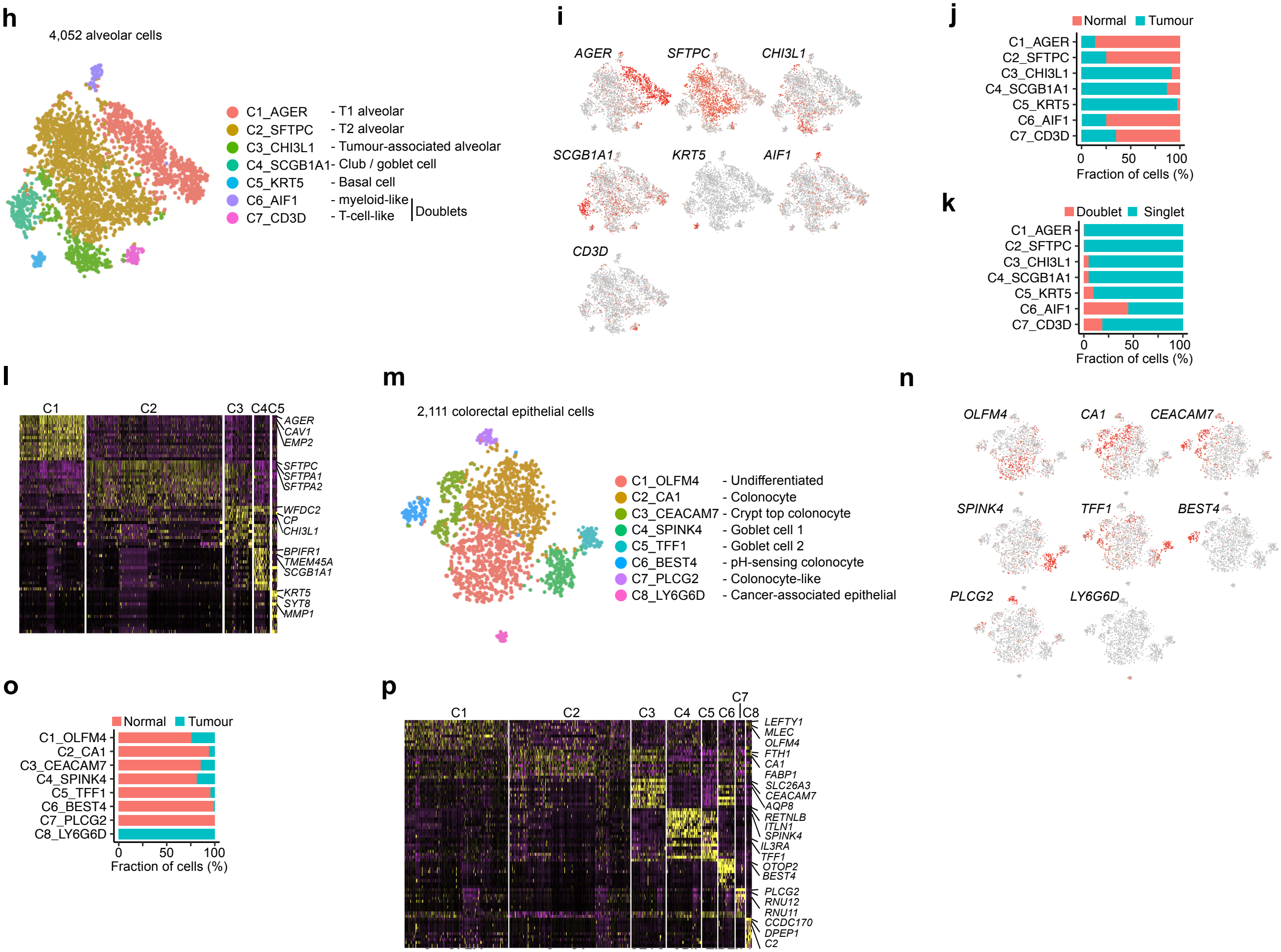
h
4,052 alveolar cells
C1_AGER - T1 alveolar
C2_SFTPC - T2 alveolar
C3_CHI3L1 - Tumour-associated alveolar
C4_SCGB1A1- Club / goblet cell
C5_KRT5 - Basal cell
C6_AIF1 - myeloid-like
C7_CD3D - T-cell-like
Doublets
i
AGER
SFTPC
CHI3L1
SCGB1A1
KRT5
AIF1
CD3D
j
Normal
Tumour
C1_AGER
C2_SFTPC
C3_CHI3L1
C4_SCGB1A1
C5_KRT5
C6_AIF1
C7_CD3D
0
50
100
Fraction of cells (%)
k
Doublet
Singlet
C1_AGER
C2_SFTPC
C3_CHI3L1
C4_SCGB1A1
C5_KRT5
C6_AIF1
C7_CD3D
0
50
100
Fraction of cells (%)
l
C1
C2
C3
C4
C5
AGER
CAV1
EMP2
SFTPC
SFTPA1
SFTPA2
WFDC2
CP
CHI3L1
BPIFR1
TMEM45A
SCGB1A1
KRT5
SYT8
MMP1
m
2,111 colorectal epithelial cells
C1_OLFM4 - Undifferentiated
C2_CA1 - Colonocyte
C3_CEACAM7 - Crypt top colonocyte
C4_SPINK4 - Goblet cell 1
C5_TFF1 - Goblet cell 2
C6_BEST4 - pH-sensing colonocyte
C7_PLCG2 - Colonocyte-like
C8_LY6G6D - Cancer-associated epithelial
n
OLFM4
CA1
CEACAM7
SPINK4
TFF1
BEST4
PLCG2
LY6G6D
o
Normal
Tumour
C1_OLFM4
C2_CA1
C3_CEACAM7
C4_SPINK4
C5_TFF1
C6_BEST4
C7_PLCG2
C8_LY6G6D
0
50
100
Fraction of cells (%)
p
C1
C2
C3
C4
C5
C6
C7
C8
LEFTY1
MLEC
OLFM4
FTH1
CA1
FABP1
SLC26A3
CEACAM7
AQP8
RETNLB
ITLN1
SPINK4
IL3RA
TFF1
OTOP2
BEST4
PLCG2
RNU12
RNU11
CCDC170
DPEP1
C2

Fig. S1h-p

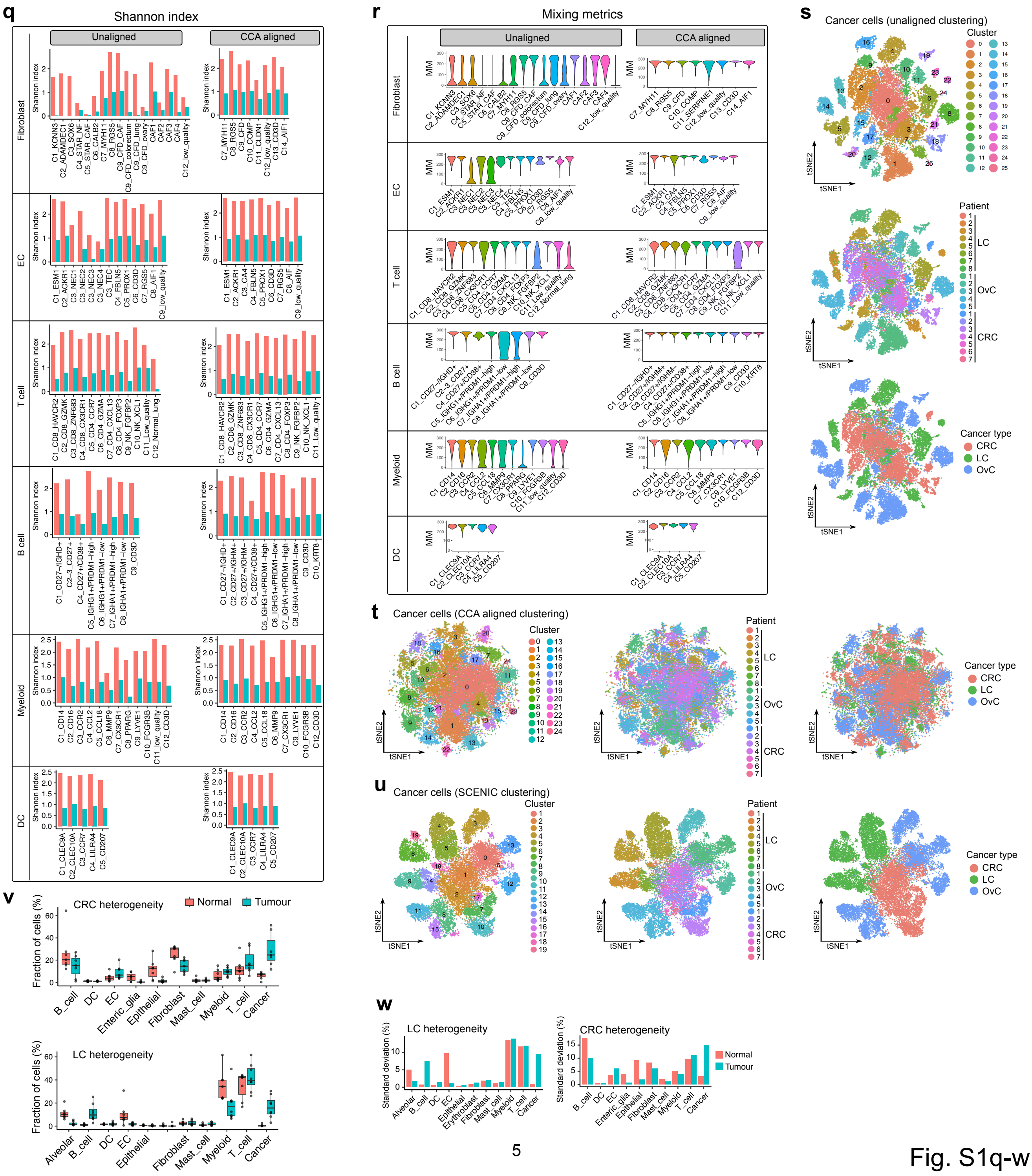
q
Shannon index
Unaligned
CCA aligned
Fibroblast
EC
T cell
B cell
Myeloid
DC
Shannon index
r
Mixing metrics
Unaligned
CCA aligned
MM
s
Cancer cells (unaligned clustering)
Cluster
Patient
LC
OvC
CRC
Cancer type
CRC
LC
OvC
tSNE1
tSNE2
t
Cancer cells (CCA aligned clustering)
u
Cancer cells (SCENIC clustering)
v
CRC heterogeneity
Normal
Tumour
Fraction of cells (%)
LC heterogeneity
B_cell
DC
EC
Enteric_glia
Epithelial
Fibroblast
Mast_cell
Myeloid
T_cell
Cancer
Alveolar
w
LC heterogeneity
CRC heterogeneity
Standard deviation (%)
Erythroblast

Fig. S1q-w

Fig. S1. Major cell type clustering

a t-SNE plots of 183,376 single cells colour-coded per cancer type, sample of origin, patient, and number of transcripts. **b** t-SNE plots colour-coded by different clusters generated by PCA based clustering at different resolutions. **c** Expression of metagenes projected on t-SNE plots. Composition of the metagenes: B-cell (*CD79A*, *CD79B*), cancer (*EPCAM*, *KRT7*, *KRT18*), DC (*CD1C*, *CD207*, *CLEC9A*, *LILRA4*, *CCL17*), EC (*CLDN5*, *PECAM1*, *VWF*), fibroblast (*COL1A*, *BGN*, *DCN*), mast cell (*MS4A2*, *TPSAB1*, *CPA3*), myeloid (*CD68*, *LYZ*, *AIF1*), T-cell (*CD3D*, *CD3E*, *CD3G*), alveolar cell (*CLDN18*, *SFTPA1*, *SFTPA2*, *SFTPC*), enteric glia (*S100B*, *PLP1*), epithelial cell (colorectal, *MT1E*, *MT1G*, *ITLN1*, *ZG16*), epithelial cell (lung, *CAPS*, *TPP3*). **d** Barplots ordered per major cell type at resolutions 0.3 (left column) and 2.0 (right column). The barplots represent the composition per cancer type (left panel), the annotation per subcluster, as previously described (middle panel) [2] and the contribution per patient (right panel). Cancer and tissue specific clusters (alveolar) are excluded. At the lower resolution arrows indicate the tissue specific clusters (black arrows), at the higher resolution arrows indicate patient specific clusters (black arrows). Cells clustered into an incorrect major cell type are highlighted by red arrows at both resolutions. **e** Expression of metagenes projected on t-SNE plots per cancer type. Same metagenes were used as in (**c**). **f** Barplots representing per cell type in different cancers from left to right: fraction of cells per origin, fraction of cells per patient, number of cells, and the total number of detected transcripts. **g** Boxplot showing the prevalence of T-cells based on GSVA scores in bulk RNA-seq of different TCGA cancer types, including LC (both for LUAD and LUSC with 541 and 502 patients respectively), BC (1,119 patients), CRC (rectal cancer or READ and colorectal cancer or COAD with 167 and 483 patients respectively) and OvC (430 patients). *P*-values for LUAD and LUSC versus all other cancer types are depicted. **h** t-SNE plot showing 4,052 alveolar cells colour-coded by clusters. **i** t-SNE plots showing marker gene expressions of alveolar clusters. **j** Fraction of cells derived from normal (red) or tumour tissue (green) for each alveolar cluster. **k** Fraction of singlet and doublet cells in each alveolar cluster calculated by DoubletFinder. C6 and C7 have higher doublet percentage than other clusters. **l** Heatmap for differential gene expression of alveolar clusters. **m** t-SNE plot showing 2,111 colorectal epithelial cells colour-coded by clusters. **n** t-SNE plots showing marker gene expressions of colorectal epithelial cell clusters. **o** Fraction of cells derived from normal (red) or tumour tissue (green) for each colorectal epithelial cluster. **p** Heatmap for differential gene expression of colorectal epithelial cell clusters. **q-r** Shannon index (**q**) and mixing metrics (**r**) in all cell type subclusters after unaligned and CCA-aligned clustering; Red and green denotes the Shannon index for patient and cancer type, respectively. **s-u** t-SNEs of cancer cells colour-coded for each separate cluster (left), per individual patient (middle) and cancer type (right) by unaligned clustering (**s**), CCA aligned clustering (**t**) and SCENIC clustering that leveraging transcription factor activity calculated by SCENIC pipeline (**u**). **v** Boxplots showing the relative percentage of cancer cells and each stromal subcluster derived from normal (red) or tumour (green) in CRC and LC, respectively. **w** Standard deviation of the relative percentage of cancer cells and each stromal subcluster in normal and tumour tissue in CRC and LC, respectively. The standard deviation was considered a measure of inter-tumour heterogeneity.