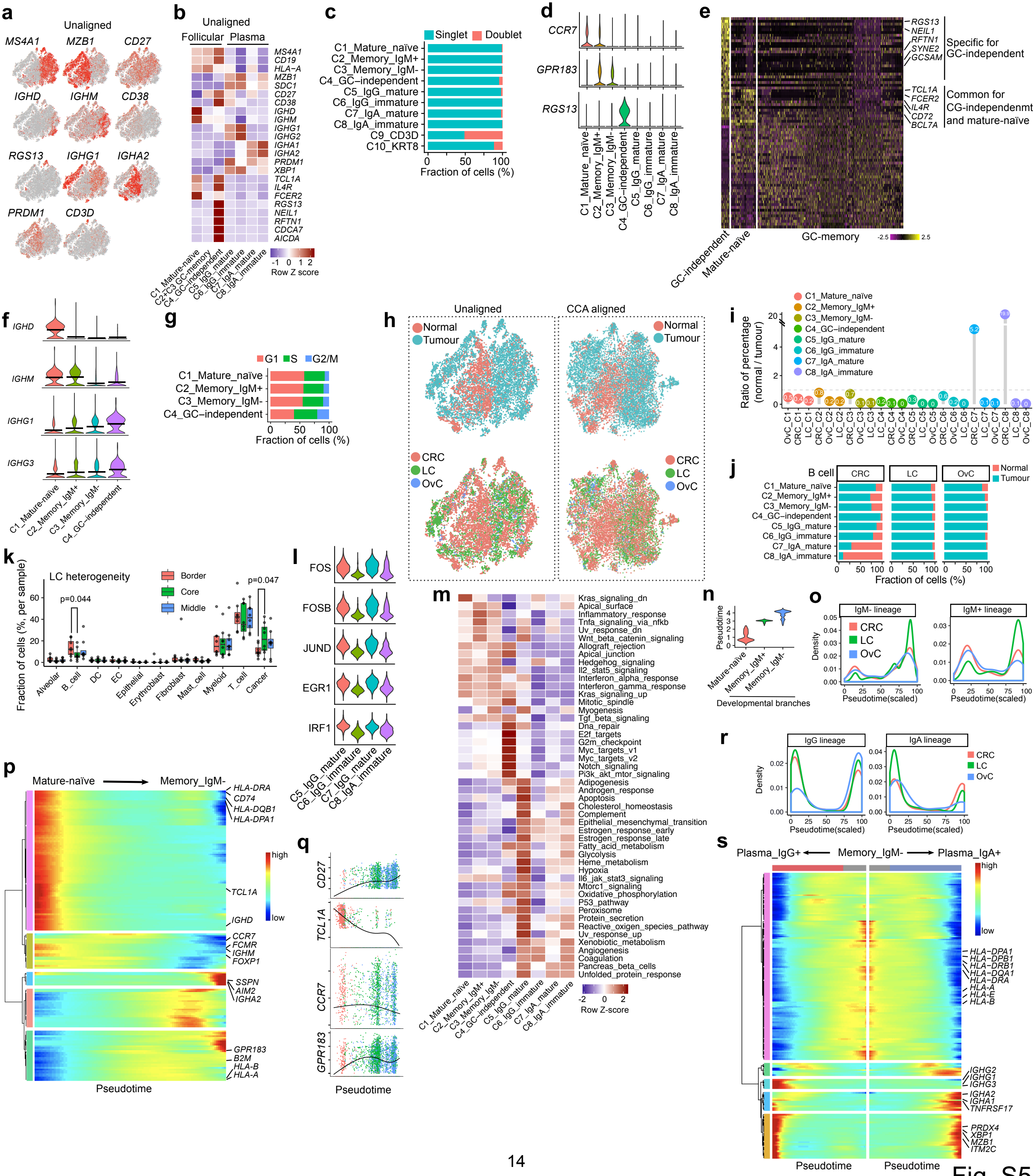
a
Unaligned
MS4A1
MZB1
CD27
IGHD
IGHM
CD38
RGS13
IGHG1
IGHA2
PRDM1
CD3D
b
Unaligned
Follicular
Plasma
MS4A1
CD19
HLA-A
MZB1
SDC1
CD27
CD38
IGHD
IGHM
IGHG1
IGHG2
IGHA1
IGHA2
PRDM1
XBP1
TCL1A
IL4R
FCER2
RGS13
NEIL1
RFTN1
CDCA7
AICDA
C1_Mature-naïve
C2+C3 GC-memory
C4_GC-independent
C5_IgG_mature
C6_IgG_immature
C7_IgA_mature
C8_IgA_immature
-1 0 1 2
Row Z score
c
Singlet
Doublet
C1_Mature_naïve
C2_Memory_IgM+
C3_Memory_IgM-
C4_GC-independent
C5_IgG_mature
C6_IgG_immature
C7_IgA_mature
C8_IgA_immature
C9_CD3D
C10_KRT8
0 50 100
Fraction of cells (%)
d
CCR7
GPR183
RGS13
C1_Mature_naïve
C2_Memory_IgM+
C3_Memory_IgM-
C4_GC-independent
C5_IgG_mature
C6_IgG_immature
C7_IgA_mature
C8_IgA_immature
e
RGS13
NEIL1
RFTN1
SYNE2
GCSAM
Specific for GC-independent
TCL1A
FCER2
IL4R
CD72
BCL7A
Common for CG-independenmt and mature-naïve
GC-independent
Mature-naïve
GC-memory
-2.5 2.5
f
IGHD
IGHM
IGHG1
IGHG3
C1_Mature-naïve
C2_Memory_IgM+
C3_Memory_IgM-
C4_GC-independent
g
G1
S
G2/M
C1_Mature_naïve
C2_Memory_IgM+
C3_Memory_IgM-
C4_GC-independent
0 50 100
Fraction of cells (%)
h
Unaligned
CCA aligned
Normal
Tumour
CRC
LC
OvC
i
C1_Mature_naïve
C2_Memory_IgM+
C3_Memory_IgM-
C4_GC-independent
C5_IgG_mature
C6_IgG_immature
C7_IgA_mature
C8_IgA_immature
Ratio of percentage (normal / tumour)
0 2 4 20
0.5 0.4 0.2 0.8 0.2 0.2 0.7 0.1 0.1 0.2 0.1 0 0.3 0 0 0.6 0.2 0 5.2 0.1 0.1 19.9 0.1 0
OvC_C1 CRC_C1 LC_C1 CRC_C2 OvC_C2 LC_C2 CRC_C3 OvC_C3 LC_C3 LC_C4 CRC_C4 OvC_C4 CRC_C5 LC_C5 OvC_C5 CRC_C6 OvC_C6 LC_C6 CRC_C7 LC_C7 OvC_C7 CRC_C8 LC_C8 OvC_C8
j
B cell
CRC
LC
OvC
Normal
Tumour
C1_Mature_naïve
C2_Memory_IgM+
C3_Memory_IgM-
C4_GC-independent
C5_IgG_mature
C6_IgG_immature
C7_IgA_mature
C8_IgA_immature
0 50 100
Fraction of cells (%)
k
LC heterogeneity
Border
Core
Middle
p=0.044
p=0.047
Fraction of cells (%, per sample)
0 20 40 60
Alveolar
B_cell
DC
EC
Epithelial
Erythroblast
Fibroblast
Mast_cell
Myeloid
T_cell
Cancer
l
FOS
FOSB
JUND
EGR1
IRF1
C5_IgG_mature
C6_IgG_immature
C7_IgG_mature
C8_IgA_immature
m
Kras_signaling_dn
Apical_surface
Inflammatory_response
Tnfa_signaling_via_nfkb
Uv_response_dn
Wnt_beta_catenin_signaling
Allograft_rejection
Apical_junction
Hedgehog_signaling
Il2_stat5_signaling
Interferon_alpha_response
Interferon_gamma_response
Kras_signaling_up
Mitotic_spindle
Myogenesis
Tgf_beta_signaling
Dna_repair
E2f_targets
G2m_checkpoint
Myc_targets_v1
Myc_targets_v2
Notch_signaling
Pi3k_akt_mtor_signaling
Adipogenesis
Androgen_response
Apoptosis
Cholesterol_homeostasis
Complement
Epithelial_mesenchymal_transition
Estrogen_response_early
Estrogen_response_late
Fatty_acid_metabolism
Glycolysis
Heme_metabolism
Hypoxia
Il6_jak_stat3_signaling
Mtorc1_signaling
Oxidative_phosphorylation
P53_pathway
Peroxisome
Protein_secretion
Reactive_oxigen_species_pathway
Uv_response_up
Xenobiotic_metabolism
Angiogenesis
Coagulation
Pancreas_beta_cells
Unfolded_protein_response
C1_Mature_naïve
C2_Memory_IgM+
C3_Memory_IgM-
C4_GC-independent
C5_IgG_mature
C6_IgG_immature
C7_IgA_mature
C8_IgA_immature
-2 0 2
Row Z-score
n
Pseudotime
0 1 2 3 4
Mature-naïve
Memory_IgM+
Memory_IgM-
Developmental branches
o
IgM- lineage
IgM+ lineage
CRC
LC
OvC
Density
Pseudotime(scaled)
0 25 50 75 100
p
Mature-naïve
Memory_IgM-
HLA-DRA
CD74
HLA-DQB1
HLA-DPA1
TCL1A
IGHD
CCR7
FCMR
IGHM
FOXP1
SSPN
AIM2
IGHA2
GPR183
B2M
HLA-B
HLA-A
high
low
Pseudotime
q
CD27
TCL1A
CCR7
GPR183
Pseudotime
r
IgG lineage
IgA lineage
CRC
LC
OvC
Density
Pseudotime(scaled)
0 25 50 75 100
s
Plasma_IgG+
Memory_IgM-
Plasma_IgA+
high
low
HLA-DPA1
HLA-DPB1
HLA-DRB1
HLA-DQA1
HLA-DRA
HLA-A
HLA-E
HLA-B
IGHG2
IGHG1
IGHG3
IGHA2
IGHA1
TNFRSF17
PRDX4
XBP1
MZB1
ITM2C
Pseudotime
Pseudotime

Fig. S5

Fig. S5. Delineation of B-cell taxonomy

a t-SNE plots showing marker gene expressions for unaligned clusters. **b** Heatmap of functional gene sets for unaligned clusters. **c** Fraction of singlet and doublet cells predicted by DoubletFinder. **d** Violin plot showing expressions of GC-migration related genes. **e** Heatmap of differentially expressed genes between follicular B-cell subclusters. **f** Differentially expressed immunoglobulin genes between follicular B-cell subclusters, mean expression indicated as horizontal bars. **g** Fraction of cells for follicular B-cell subclusters with predicted cell cycle phases. **h** t-SNE plots for unaligned (left panels) and CCA-aligned (right panels) methods colour-coded by sample origin (upper panels), or cancer type (lower panels). **i** Ratio of relative % for B-cell subclusters by comparing normal with tumour tissues, <1 indicates tumour enrichment. Most B-cells (C1-C6) were enriched in tumour ($FDR < 5.2x10^{-8}$), while IgA^+ plasma cells (C7-C8) were enriched in normal colon ($FDR < 8.9x10^{-118}$). **j** Fraction of cells derived from normal (red) or tumour tissue (green) for each B-cell subcluster in each cancer type. **k** Boxplot showing the prevalence of each cell type stratified for tissue taken from the border, middle or core tumour region in LC. P-values < 0.05 comparing border versus core regions are shown. **l** Violin plot showing transcription factor activities (AUC score) calculated by SCENIC for plasma cell subclusters. **m** Differences in cancer hallmark pathway activities scored per cell by AUCell for B-cell clusters. **n** Violin plot showing pseudotime distribution of different branches indicated in Fig. 5g. **o** Density plots for CRC, LC and OvC along the trajectories of B memory cells of the IgM^+ and IgM^- lineages. **p** Heatmap of expression dynamics of differentially expressed genes along the pseudotime from mature-naïve to class-switched memory B-cell (IgM-). **q** Marker gene dynamics along the pseudotime from mature-naïve to class-switched memory B-cells (IgM^-) colour-coded as in Fig. 5g. **r** Density plots for CRC, LC and OvC along the differentiation trajectories of IgM^- to plasma cells (both the IgG^+ and IgA^+ lineages are shown). **s** Heatmap of expression dynamics of differentially expressed genes along the pseudotime from IgM^- memory B-cells to IgG^+ or IgA^+ plasma cells.