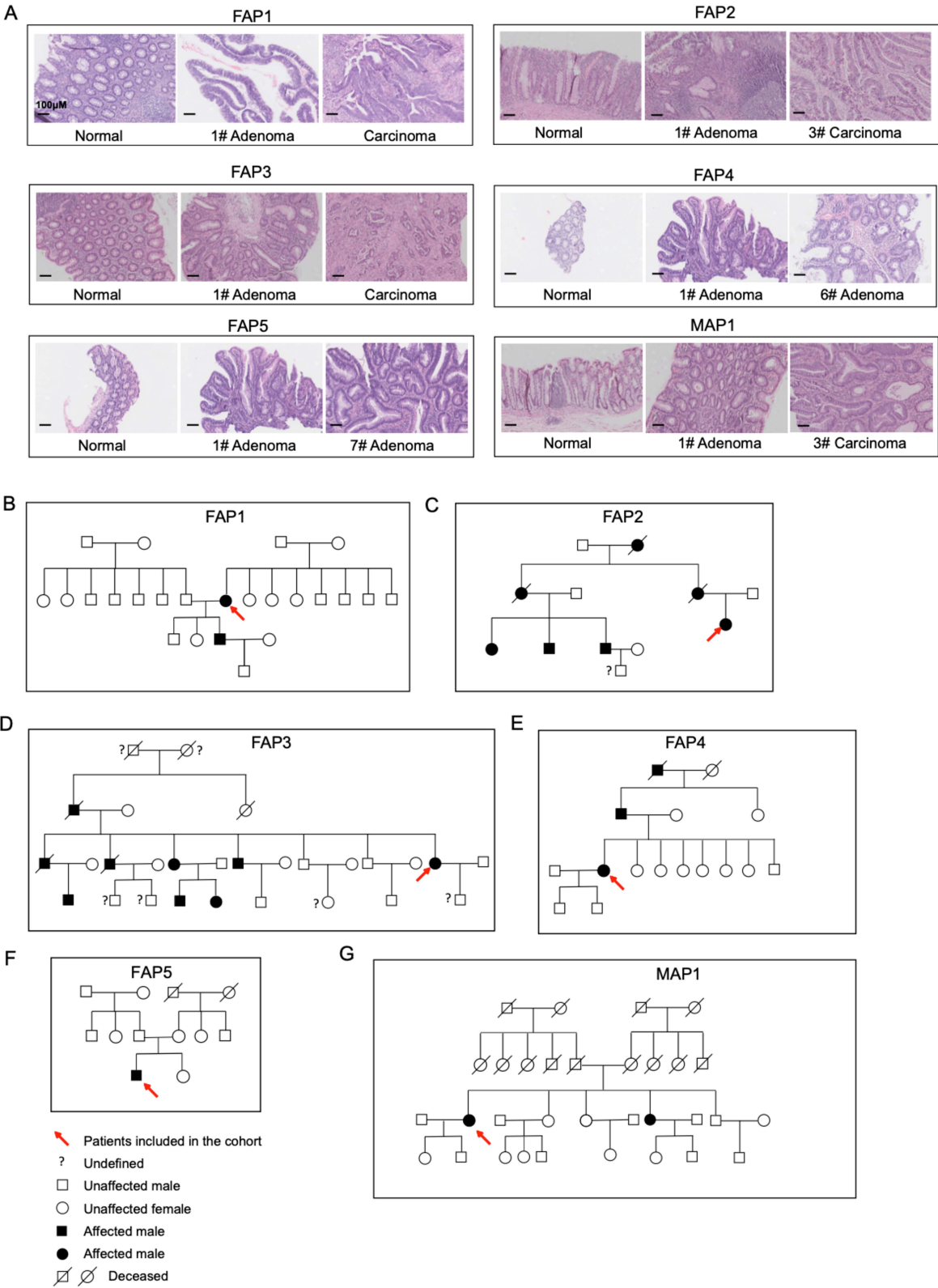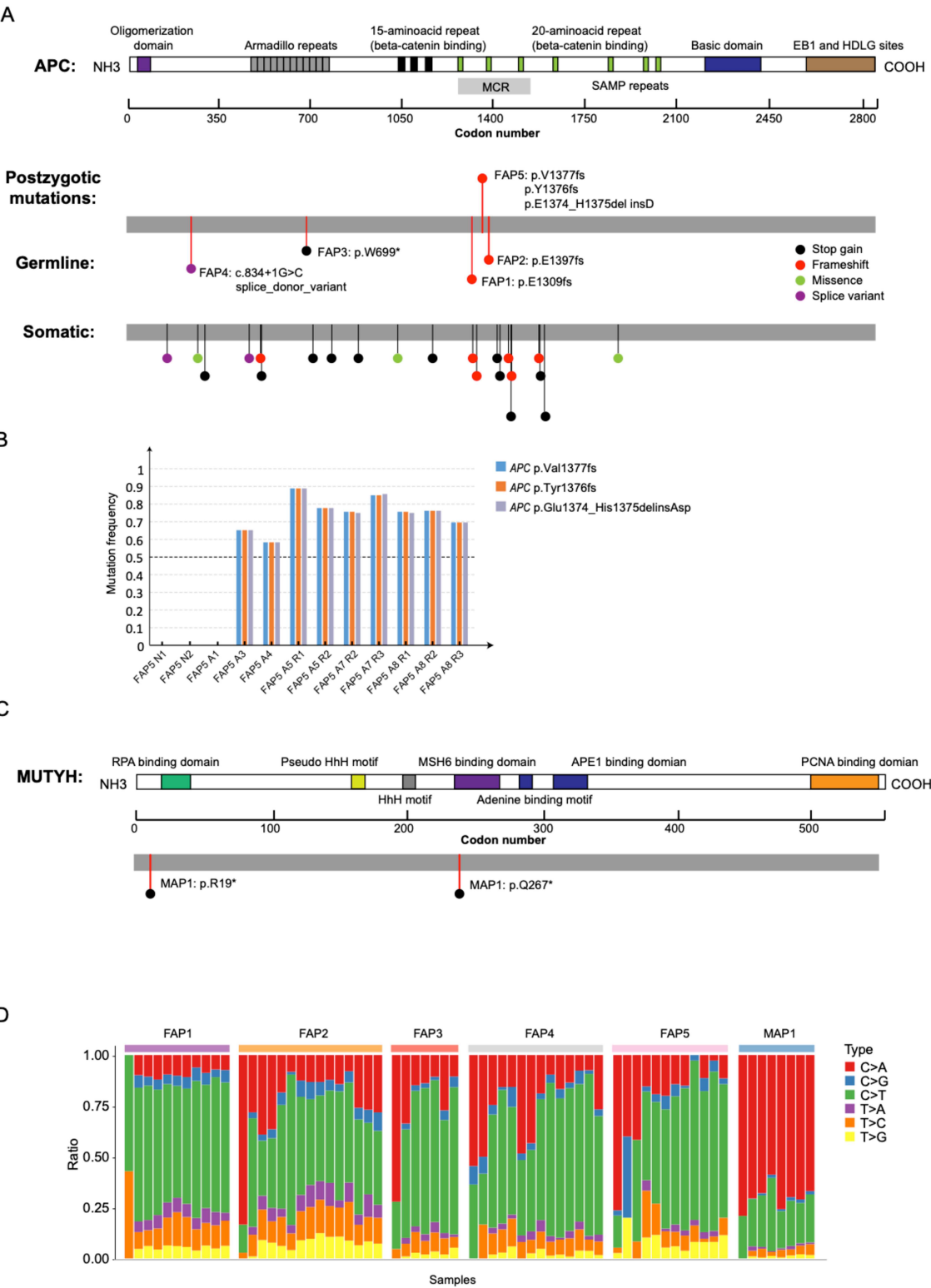## Supplementary Figure 1

Supplementary Figure 1 Hematoxylin and eosin (H&E) staining and family history of the cohort.

(A) Representative H&E staining of samples from five FAP patients (FAP1, FAP2, FAP3, FAP4 and FAP5) and MAP1. (B to G) Pedigrees of five FAP patients and MAP1. Patients in our cohort were indicated with red arrow. Question mark indicates that the diseases history of corresponding person was not recorded.
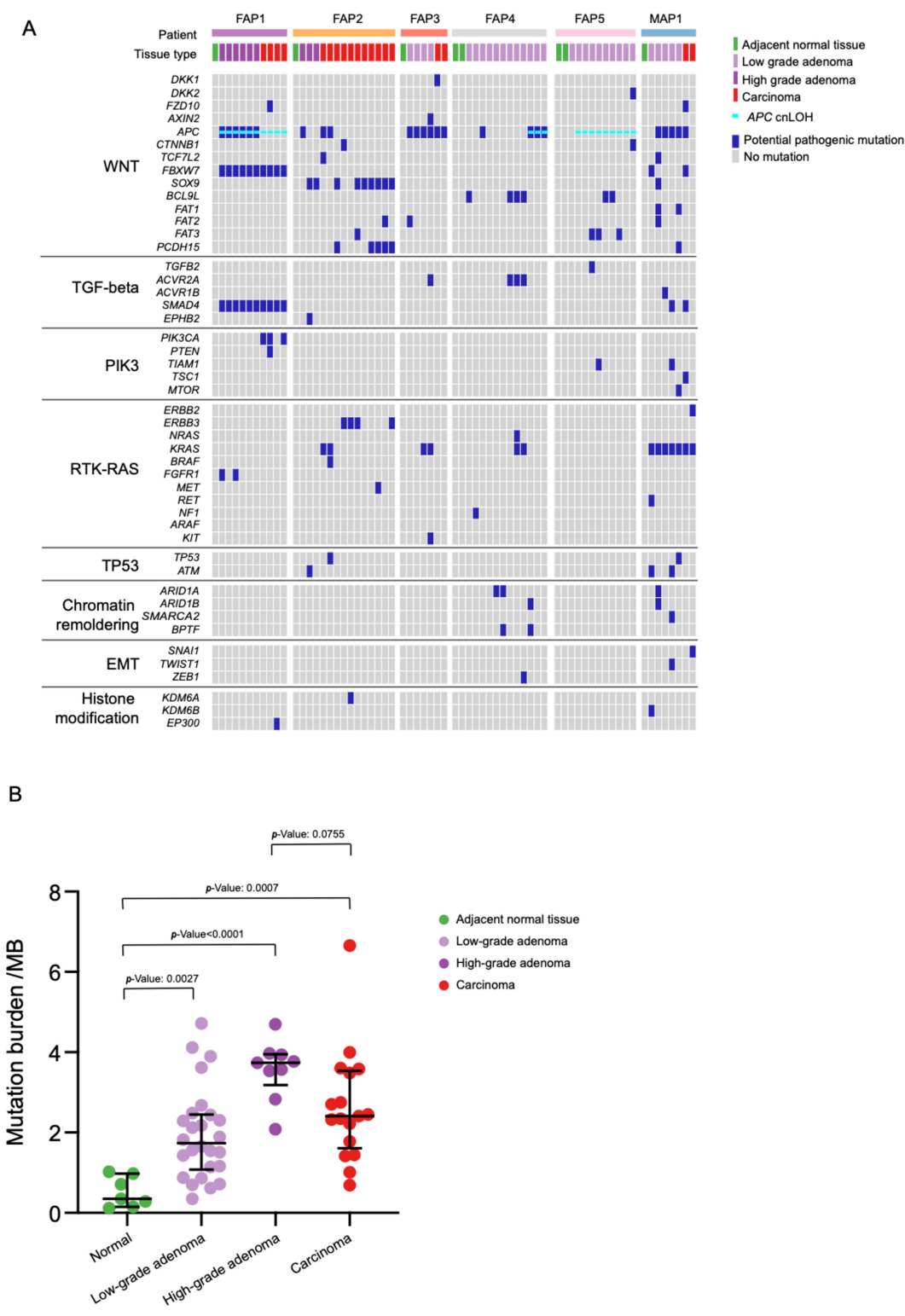
Supplementary Figure 2

Supplementary Figure 2 Summary of germline mutations of the cohort.

(A) Locations of germline mutations, post-zygotic mutations and somatic mutations on *APC* inferred from the lesions of the cohort. (B) Allele frequency of the three post-zygotic mutations on *APC* in each lesion from FAP5. (C) Locations of germline mutations on *MUTYH* inferred from the lesions from MAP1. (D) Percentage distribution of the six types of base-pair substitutions in each sample. The lesions from MAP1 have higher percentage of G:C to A:T transversion mutations compared to FAPs, which is consistent with previous reports.
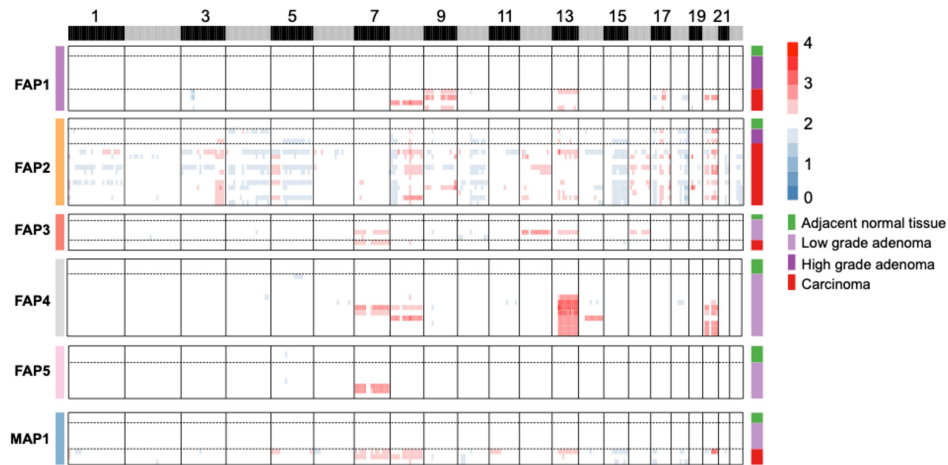
Supplementary Figure 3

Supplementary Figure 3 Mutational landscapes of the cohort.

(A) Mutational landscapes of genes related to recurrently disturbed signal pathways in colorectal cancer. Top, The patient origin and histopathological types for each sample are indicated in different color. The samples from same patients are arranged together. Middle, genes are sorted by the signal pathways they participate in. (B) Scatter diagram showing mutation burdens in adjacent normal tissue, low-grade adenomas (I grade adenomas and II grade adenomas), high-grade adenomas (III grade adenomas) and carcinomas. The samples from MAP1 (*MUTYH*-related polyposis) are not included because of their higher mutation burdens than other patients. For each group, median (center line) and interquartile range are showed with black line.
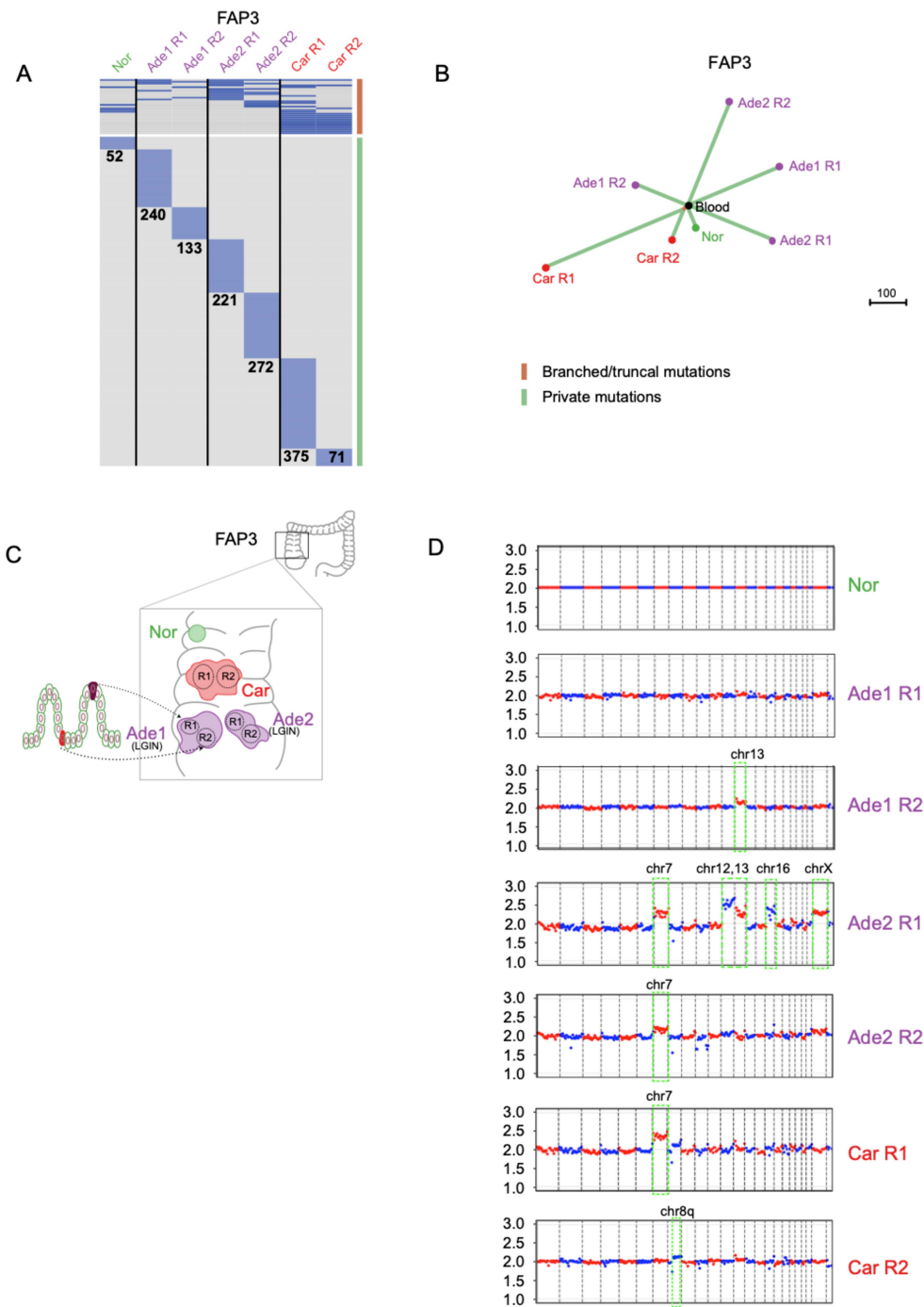
Supplementary Figure 4

Supplementary Figure 4 Summary of CNAs in each samples.

The samples from same patient were arranged together. Adjacent normal tissue, low-grade adenomas, high-grade adenomas and carcinomas were separated by black dot line. Copy number gain and copy number loss were indicated with red and blue respectively.

Supplementary Figure 5

A FAP3

Nor | Ade1 R1 | Ade1 R2 | Ade2 R1 | Ade2 R2 | Car R1 | Car R2

52
240
133
221
272
375 71

Branched/truncal mutations
Private mutations

B FAP3

Ade2 R2
Ade1 R2
Ade1 R1
Blood
Nor
Car R2
Ade2 R1
Car R1

100

C FAP3

Nor
Car
R1 R2
Ade1 (LGIN)
R1 R2
Ade2 (LGIN)
R1 R2

D

Nor

Ade1 R1

chr13
Ade1 R2

chr7 chr12,13 chr16 chrX
Ade2 R1

chr7
Ade2 R2

chr7
Car R1

chr8q
Car R2

Supplementary Figure 5 Somatic mutational and CNAs landscapes of FAP3.

(A) Heatmap showing the regional distribution of somatic mutations in all samples from FAP3. The numbers of private mutations for each sample were showed. Samples from different lesions were separated by black line. (B) Phylogenetic tree of lesions from FAP3 by using maximum parsimony algorithm. (C) Schematic diagram showing the representative locations of the biopsies from FAP3. (D) Genome wild copy number of each sample from FAP3, which were derived from whole genome sequencing. The chromosome regions with copy number alterations were highlighted with green dashed box.
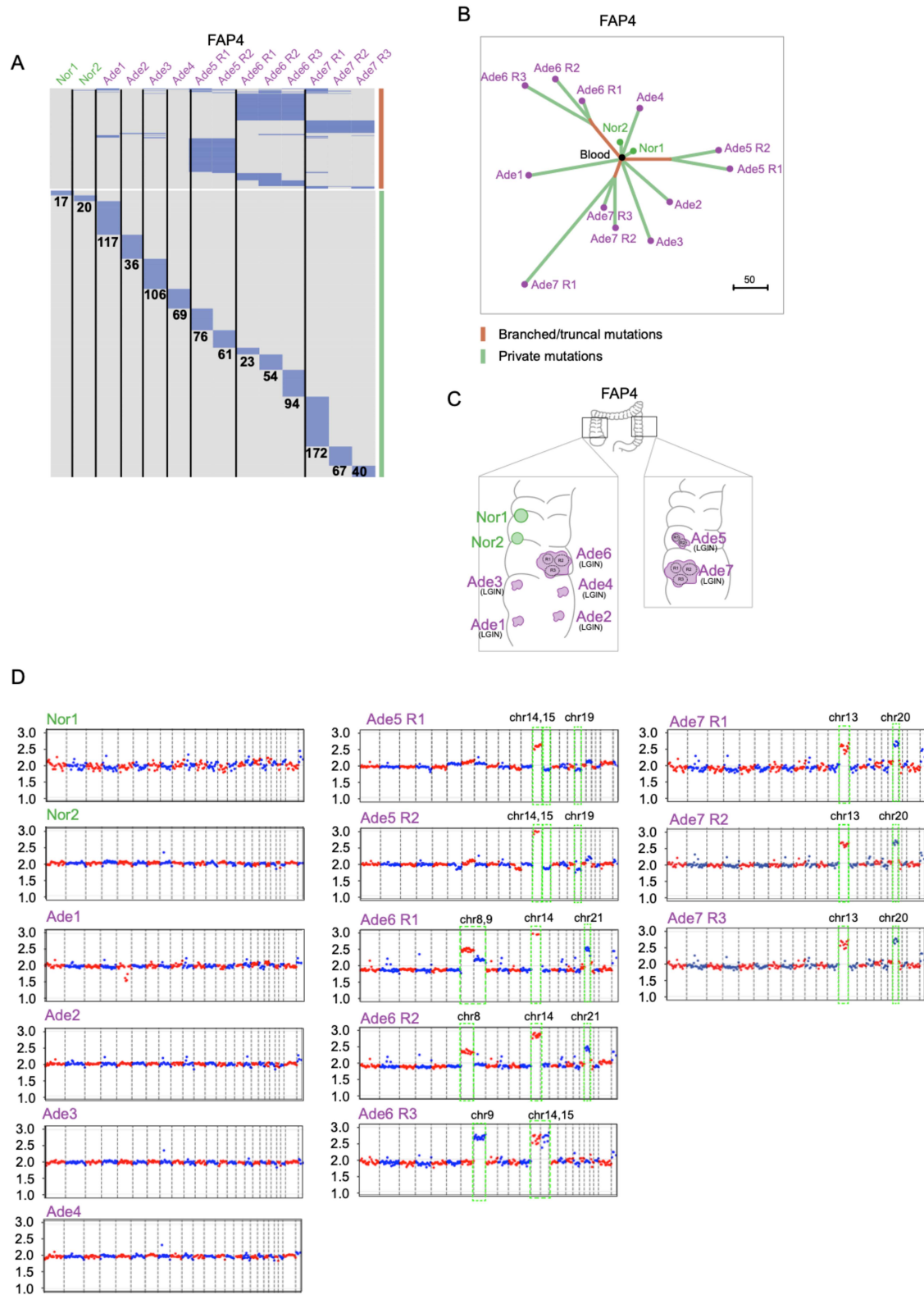
Supplementary Figure 6

Supplementary Figure 6 Somatic mutational and CNAs landscapes of FAP4.

(A) Heatmap showing the regional distribution of somatic mutations in all samples from FAP4. The numbers of private mutations for each sample were showed. Samples from different lesions were separated by black line. (B) Phylogenetic tree of lesions from FAP4 by using maximum parsimony algorithm. (C) Schematic diagram showing the representative locations of the biopsies from FAP4. (D) Genome wild copy number of each sample from FAP4, which were derived from whole genome sequencing. The chromosome regions with copy number alterations were highlighted with green dashed box.
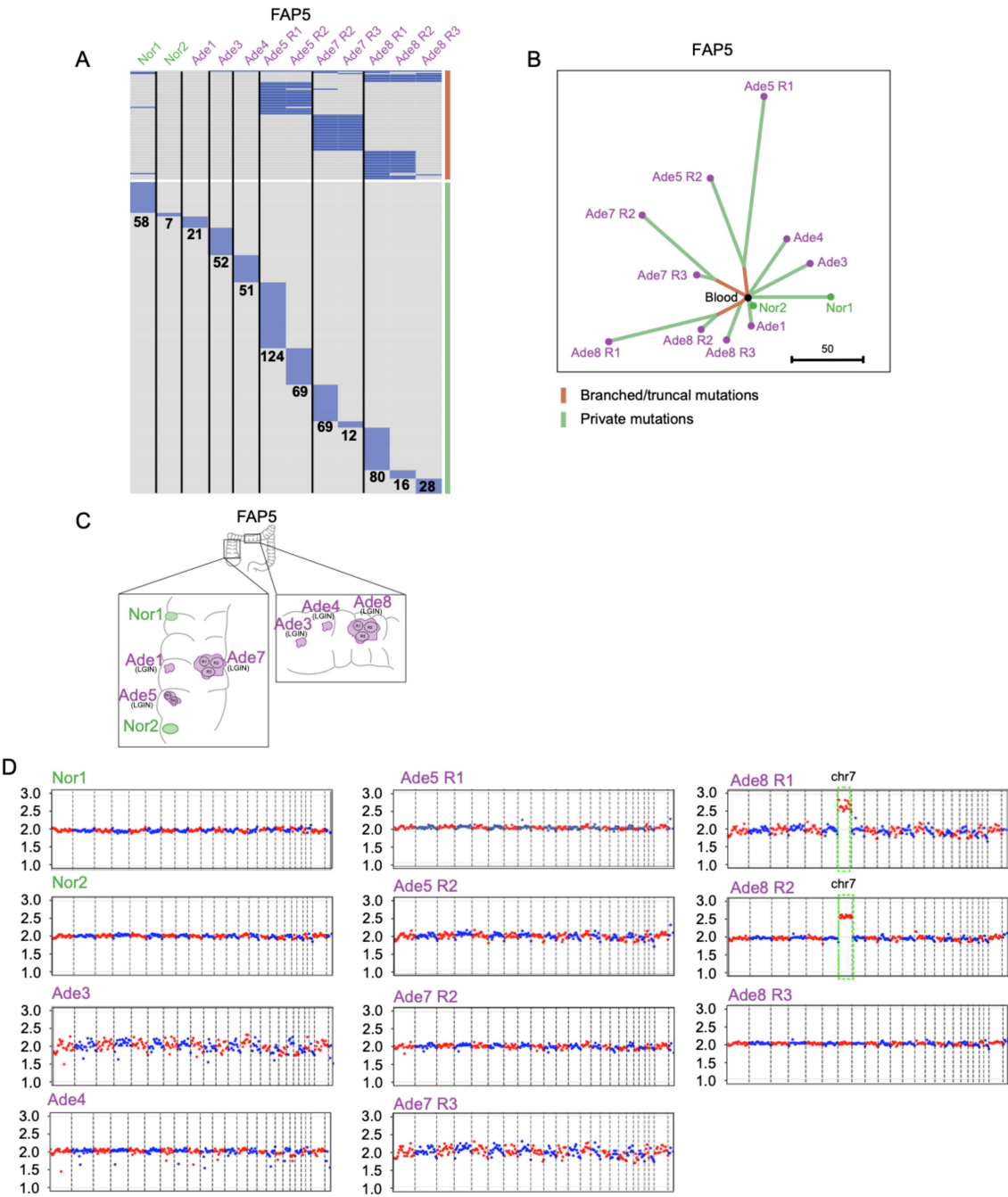
Supplementary Figure 7

Supplementary Figure 7 Somatic mutational and CNAs landscapes of FAP5.

(A) Heatmap showing the regional distribution of somatic mutations in all samples from FAP5. The numbers of private mutations for each sample were showed. Samples from different lesions were separated by black line. (B) Phylogenetic tree of lesions from FAP5 by using maximum parsimony algorithm. (C) Schematic diagram showing the representative locations of the biopsies from FAP5. (D) Genome wild copy number of each sample from FAP5, which were derived from whole genome sequencing. The chromosome regions with copy number alterations were highlighted with green dashed box.
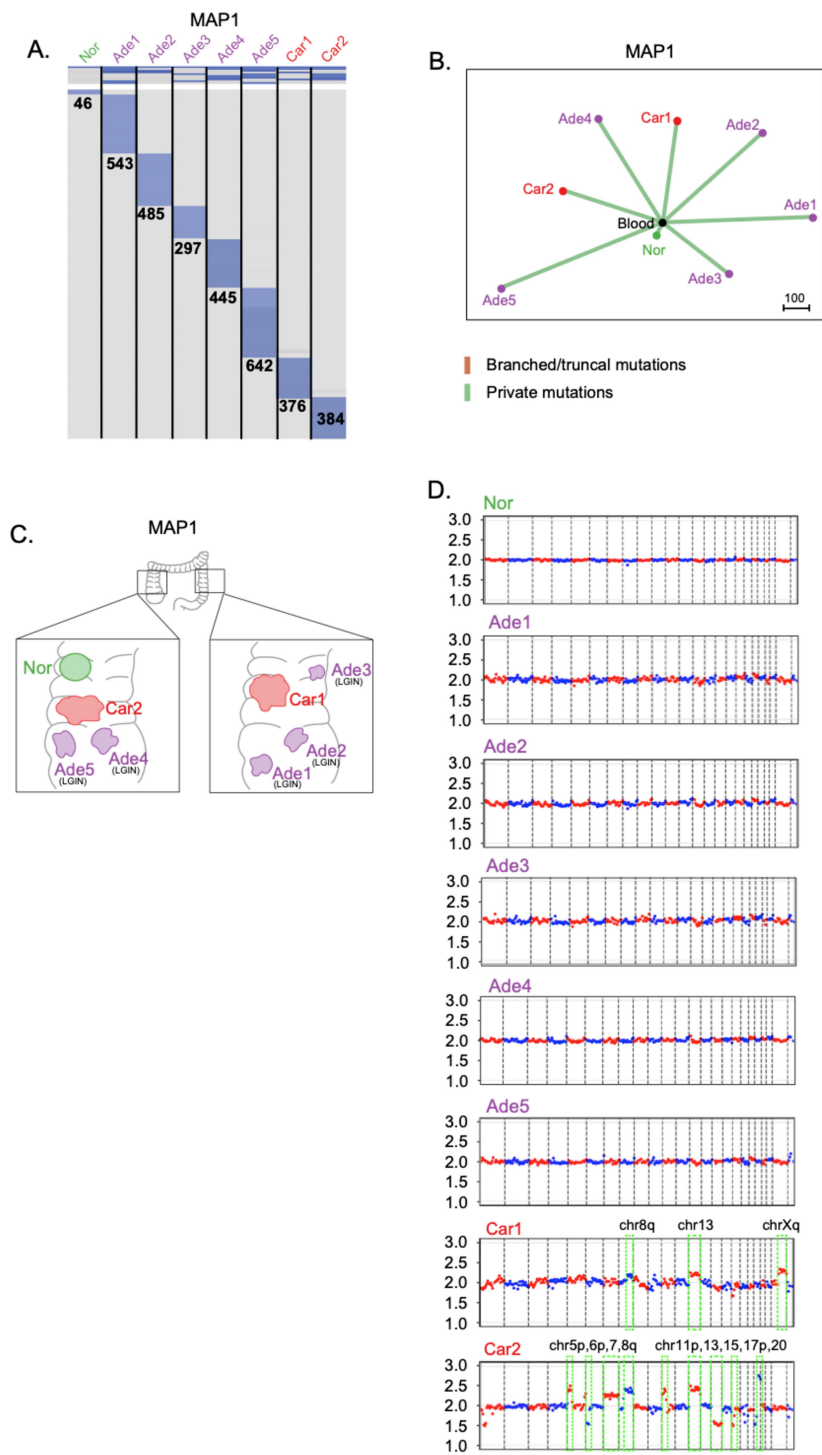
Supplementary Figure 8

Supplementary Figure 8 Somatic mutational and CNAs landscapes of MAP1.

(A) Heatmap showing the regional distribution of somatic mutations in all samples from MAP1. The numbers of private mutations for each sample were showed. Samples from different lesions were separated by black line. (B) Phylogenetic tree of lesions from MAP1 by using maximum parsimony algorithm. (C) Schematic diagram showing the representative locations of the biopsies from MAP1. (D) Genome wild copy number of each sample from MAP1, which were derived from whole genome sequencing. The chromosome regions with copy number alterations were highlighted with green dashed box.
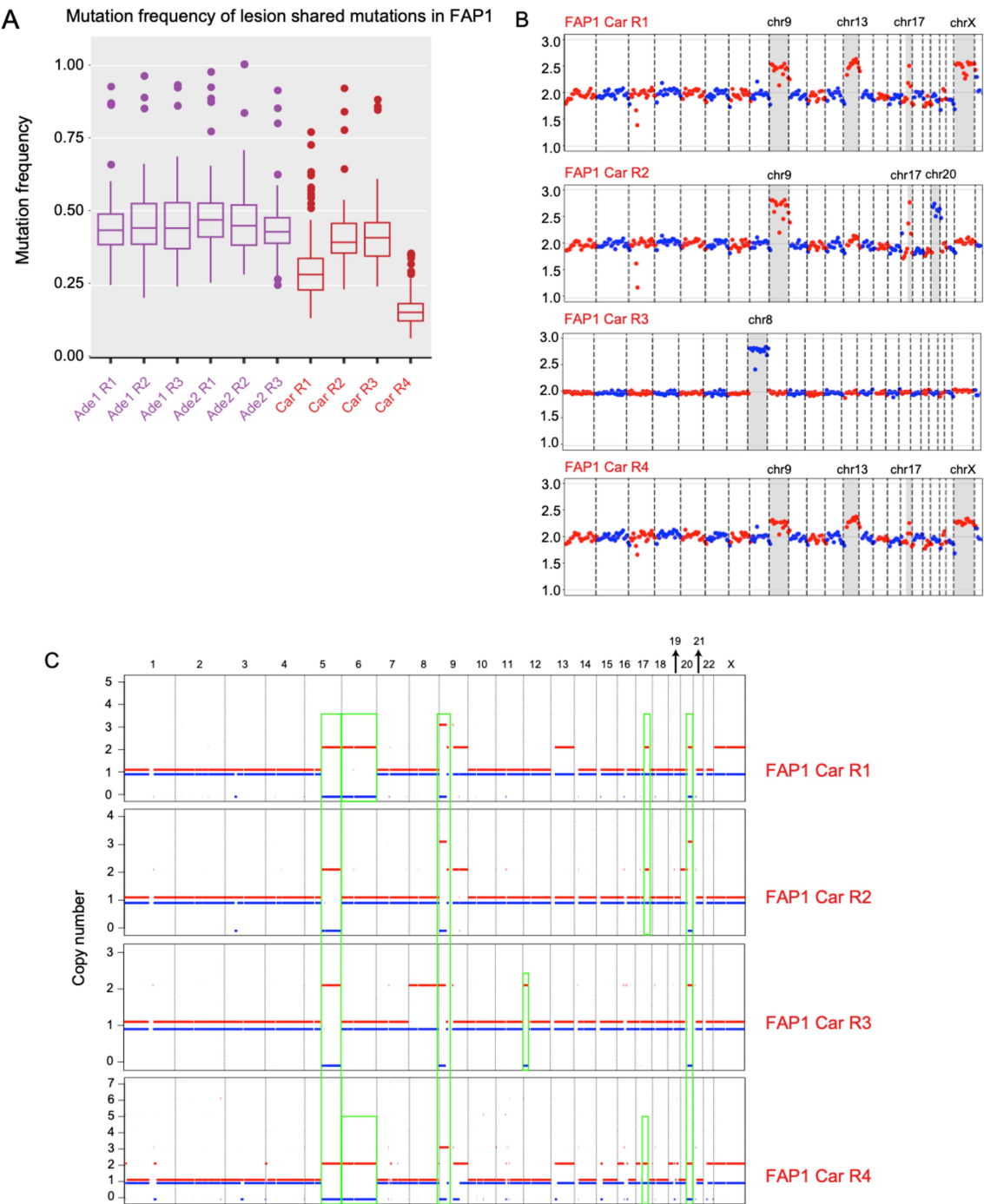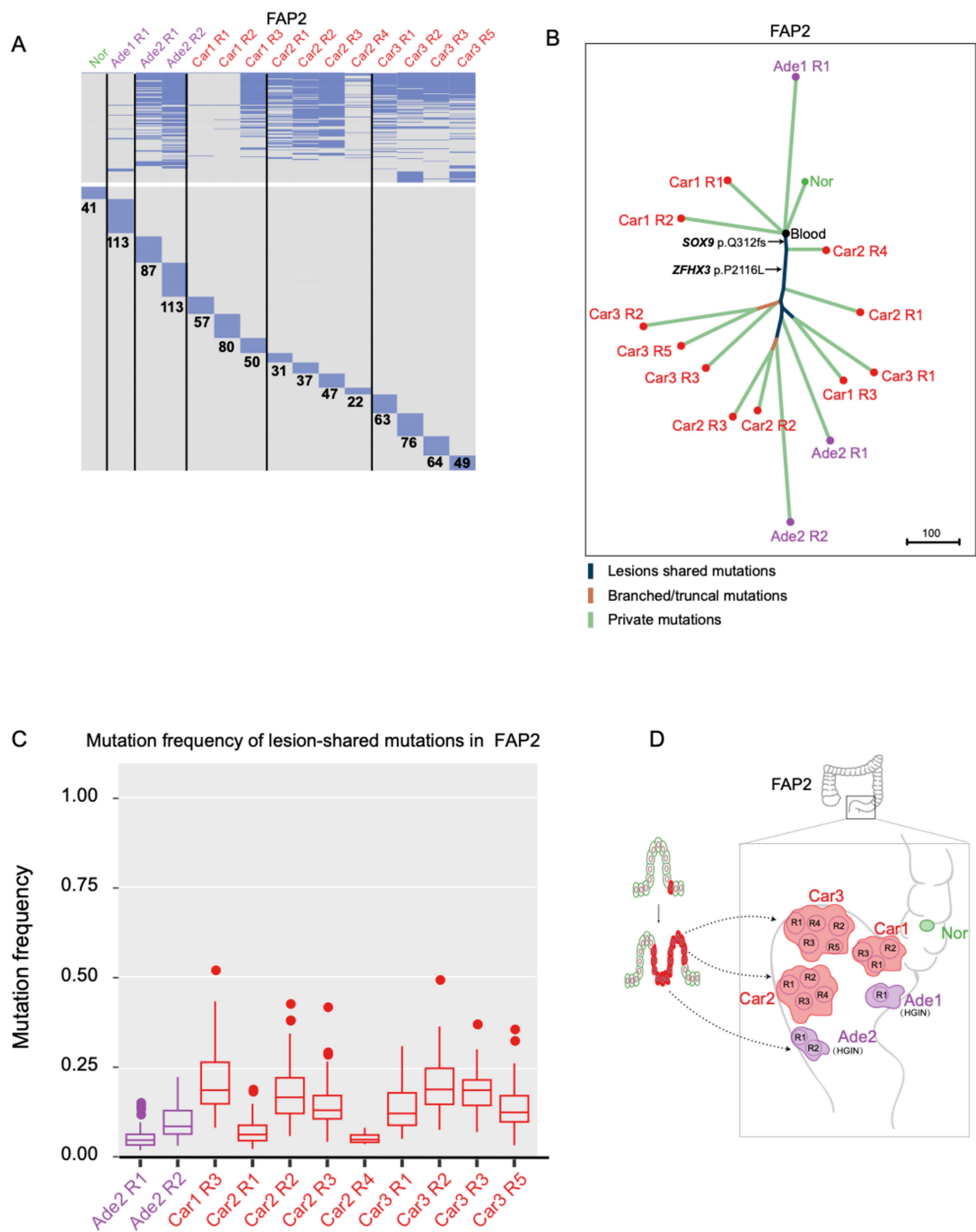
Supplementary Figure 9

A Mutation frequency of lesion shared mutations in FAP1

B

C

Supplementary Figure 9 CNAs landscape of lesions from FAP1.

(A) Mutation frequencies of lesion shared mutations in samples from FAP1. (B) Genome wild copy number of the four regions of carcinoma from FAP1. The copy number was derived from whole genome sequencing. The chromosome regions with copy number alterations were highlighted with gray shadow. (C) Genome wild copy number derived from whole exome sequencing by using "Sequenza". The chromosome regions with CNAs and LOHs were highlighted with dashed green box.
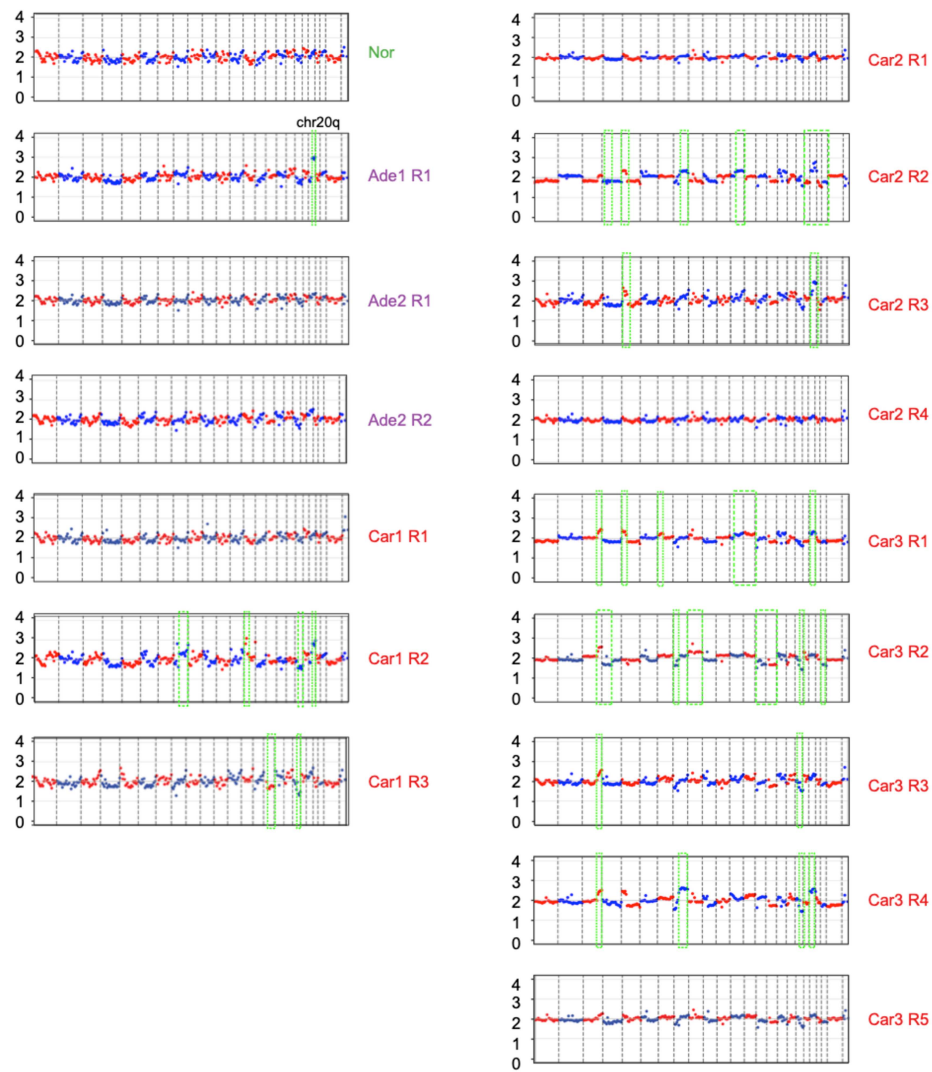
Supplementary Figure 10

Supplementary Figure 10 Somatic mutational landscapes of FAP2.

(A) Heatmap showing the regional distribution of somatic mutations in all samples from FAP2. The numbers of private mutations for each sample were showed. Samples from different lesions were separated by black line. (B) Phylogenetic tree of lesions at different evolutionary stages from FAP2 by using maximum parsimony algorithm. (C) Mutation frequencies of lesion shared mutations in samples from FAP2. (D) Schematic diagram showing the representative locations of the biopsies from FAP2.
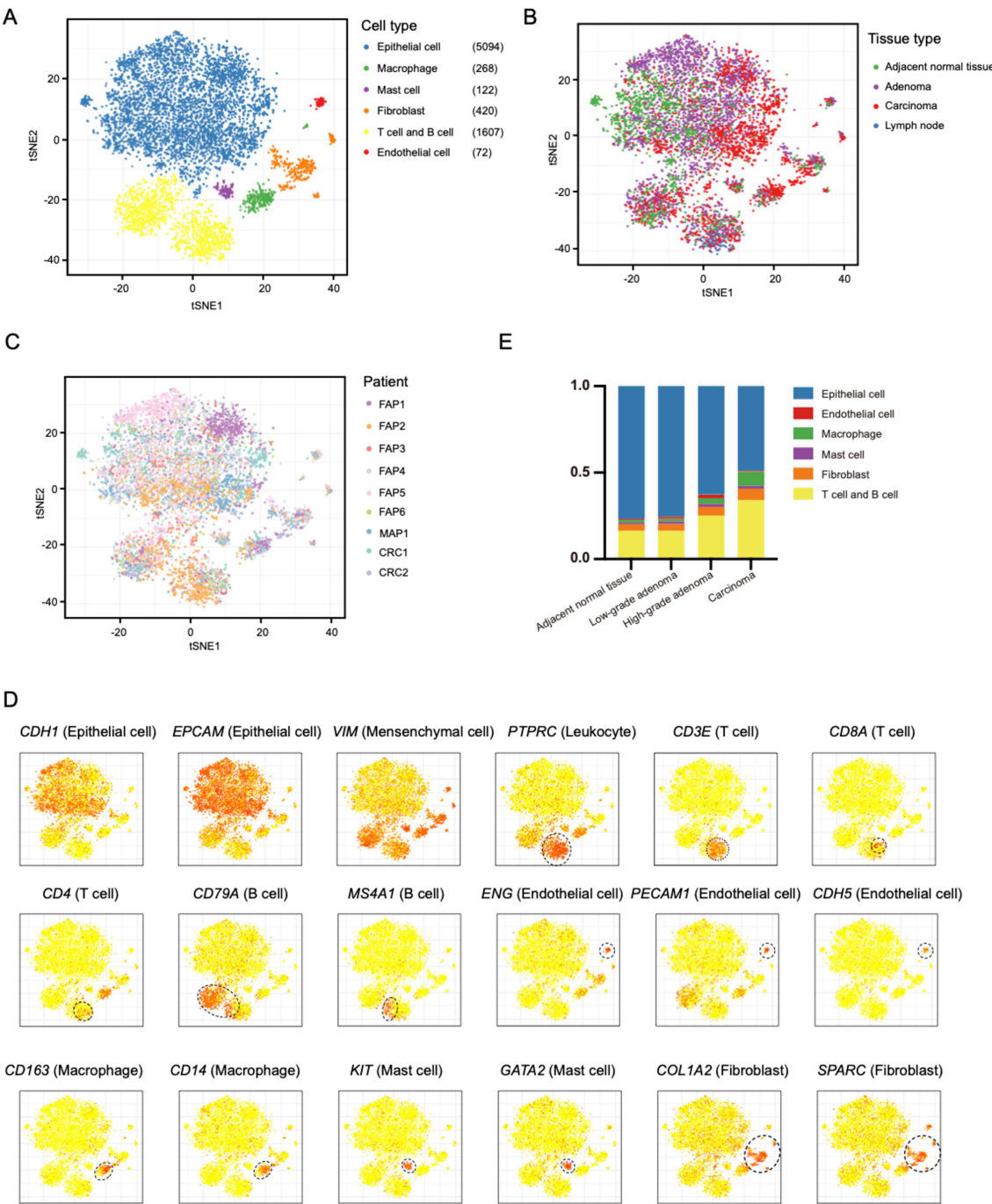
Supplementary Figure 11

Supplementary Figure 11 CNAs landscapes of FAP2.

Genome wild copy number of samples from FAP2. The copy number was derived from whole genome sequencing. The chromosome regions with copy number alterations were highlighted with green dashed box.
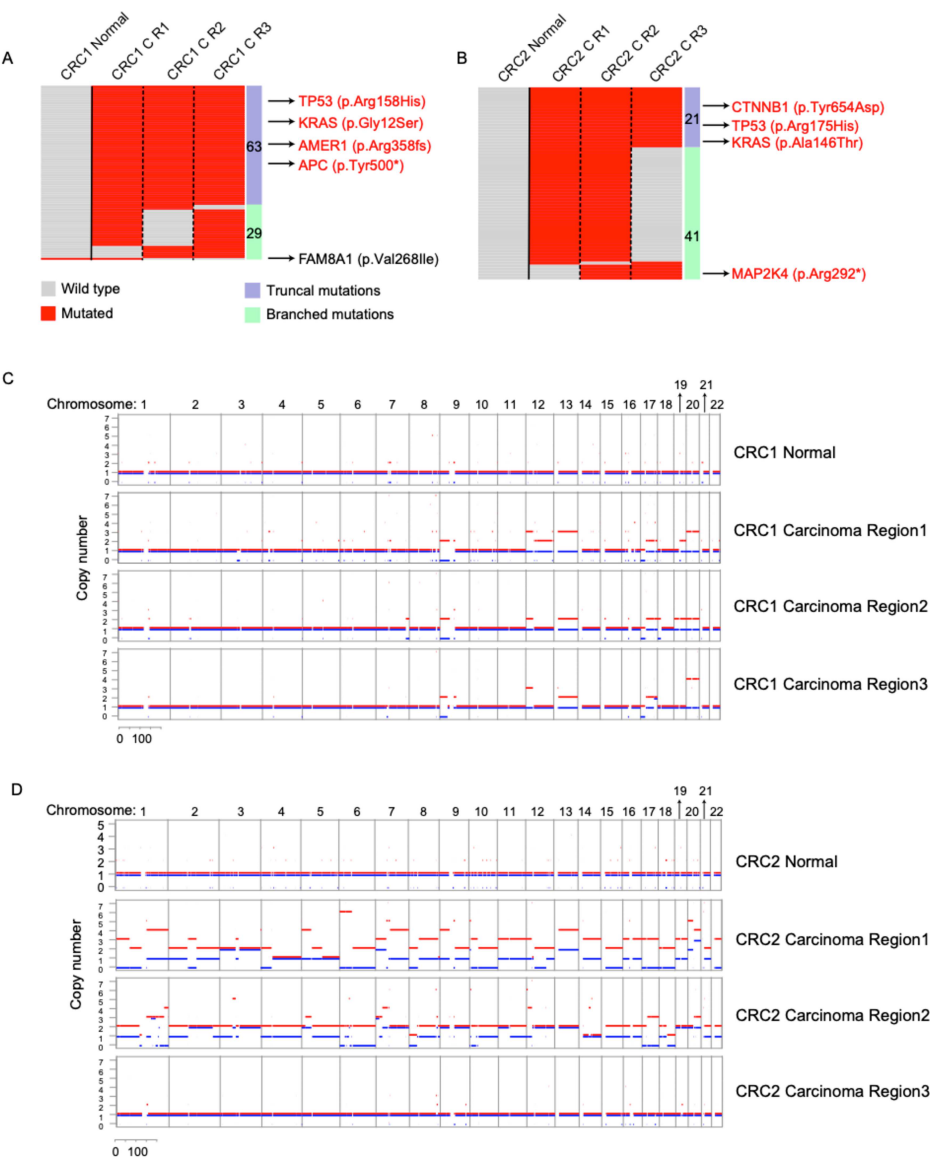
Supplementary Figure 12

Supplementary Figure 12 Cell type classification of all single cells from the cohort.
(A) Clustering analysis of all single cells using the unsupervised clustering of regulation network signatures of transcription factors with tSNE. Six types of cells were identified. The numbers of cells from each cell type were showed. (B) Tissue type origin for each single cell was indicated with different color. (C) Patient origin for each single cell was indicated on the tSNE with different color. (D) Expression patterns of known cell type markers on tSNE map. (E) The histogram shows the percentage distribution of each cell type in adjacent normal tissues, low-grade adenomas, high-grade adenomas and carcinomas.
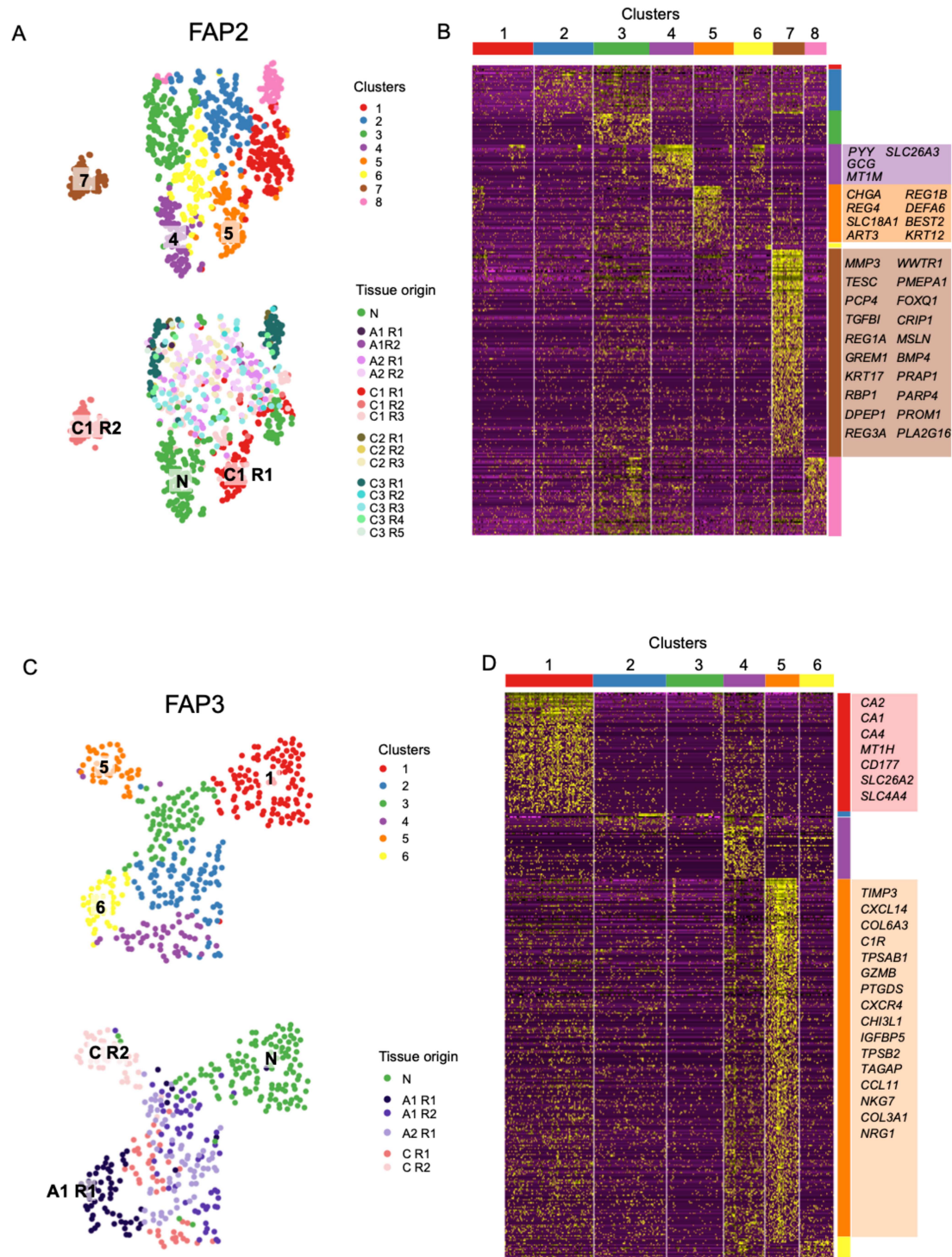
Supplementary Figure 13

Supplementary Figure 13 Somatic mutational and CNAs landscapes of CRC1 and CRC2.

(A) Heatmap showing the regional distribution of truncal/branched somatic mutations in all samples from CRC1. The numbers of truncal mutations (purple) and branched mutations (green) for each sample were showed. Samples from different lesions were separated by solid black line. Different regions from same lesion were separated by dashed black line. The potential driver mutations in carcinomas from CRC1 were not detected in the corresponding adjacent normal tissue. (B) Heatmap showing the regional distribution of truncal/branched somatic mutations in all samples from CRC2. The numbers of truncal mutations (purple) and branched mutations (green) for each sample were showed. Samples from different lesions were separated by solid black line. Different regions from same lesion were separated by dashed black line. The potential driver mutations in carcinomas from CRC2 were not detected in the corresponding adjacent normal tissue. (C) Genome wild copy number of samples from CRC1. The genome wild copy numbers were derived from whole exome sequencing by using "Sequenza". (D) Genome wild copy number of samples from CRC2. The genome wild copy numbers were derived from whole exome sequencing by using "Sequenza".

Supplementary Figure 14

A FAP2

Clusters
1 2 3 4 5 6 7 8

Clusters
1
2
3
4
5
6
7
8

Tissue origin
N
A1 R1
A1R2
A2 R1
A2 R2
C1 R1
C1 R2
C1 R3
C2 R1
C2 R2
C2 R3
C3 R1
C3 R2
C3 R3
C3 R4
C3 R5

PYY   SLC26A3
GCG
MT1M

CHGA   REG1B
REG4   DEFA6
SLC18A1   BEST2
ART3   KRT12

MMP3   WWTR1
TESC   PMEPA1
PCP4   FOXQ1
TGFBI   CRIP1
REG1A   MSLN
GREM1   BMP4
KRT17   PRAP1
RBP1   PARP4
DPEP1   PROM1
REG3A   PLA2G16

C FAP3

Clusters
1 2 3 4 5 6

Clusters
1
2
3
4
5
6

Tissue origin
N
A1 R1
A1 R2
A2 R1
C R1
C R2

CA2
CA1
CA4
MT1H
CD177
SLC26A2
SLC4A4

TIMP3
CXCL14
COL6A3
C1R
TPSAB1
GZMB
PTGDS
CXCR4
CHI3L1
IGFBP5
TPSB2
TAGAP
CCL11
NKG7
COL3A1
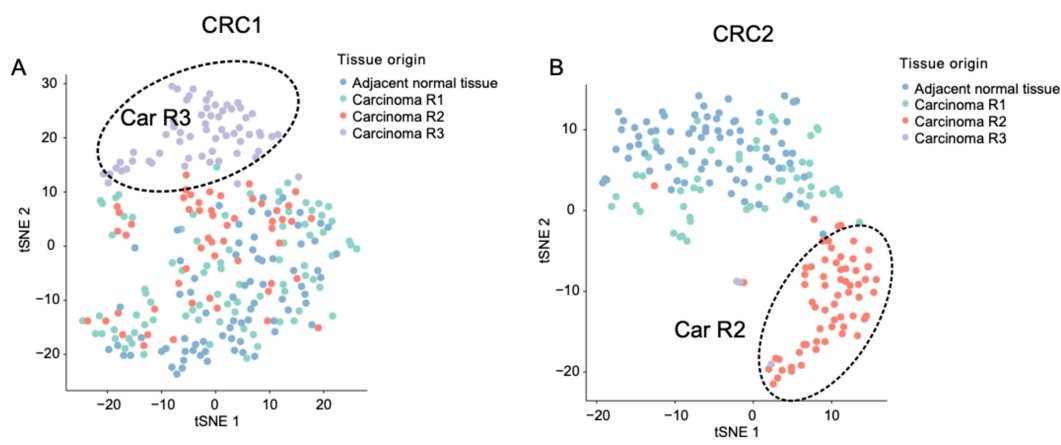NRG1

Supplementary Figure 14 Transcriptome heterogeneity of lesions at different evolutionary stages from FAP2 and FAP3.

(A) Clustering analysis of all epithelial cells from FAP2 by using tSNE. Up, single cells in different clusters were indicated with different colors. Down, the tissue origin for each single cell was indicated with different colors. (B) Heatmap showing the specifically highly expressed genes in each cluster of cells from FAP2. Representative genes of cluster 4, 5 and 7 were showed. (C) Clustering analysis of all epithelial cells from FAP3 by using tSNE. Up, single cells in different clusters were indicated with different colors. Down, the tissue origin for each single cell was indicated with different colors. (D) Heatmap showing the specifically highly expressed genes in each cluster of cells from FAP3. Representative genes of cluster 1 and 5 were showed.
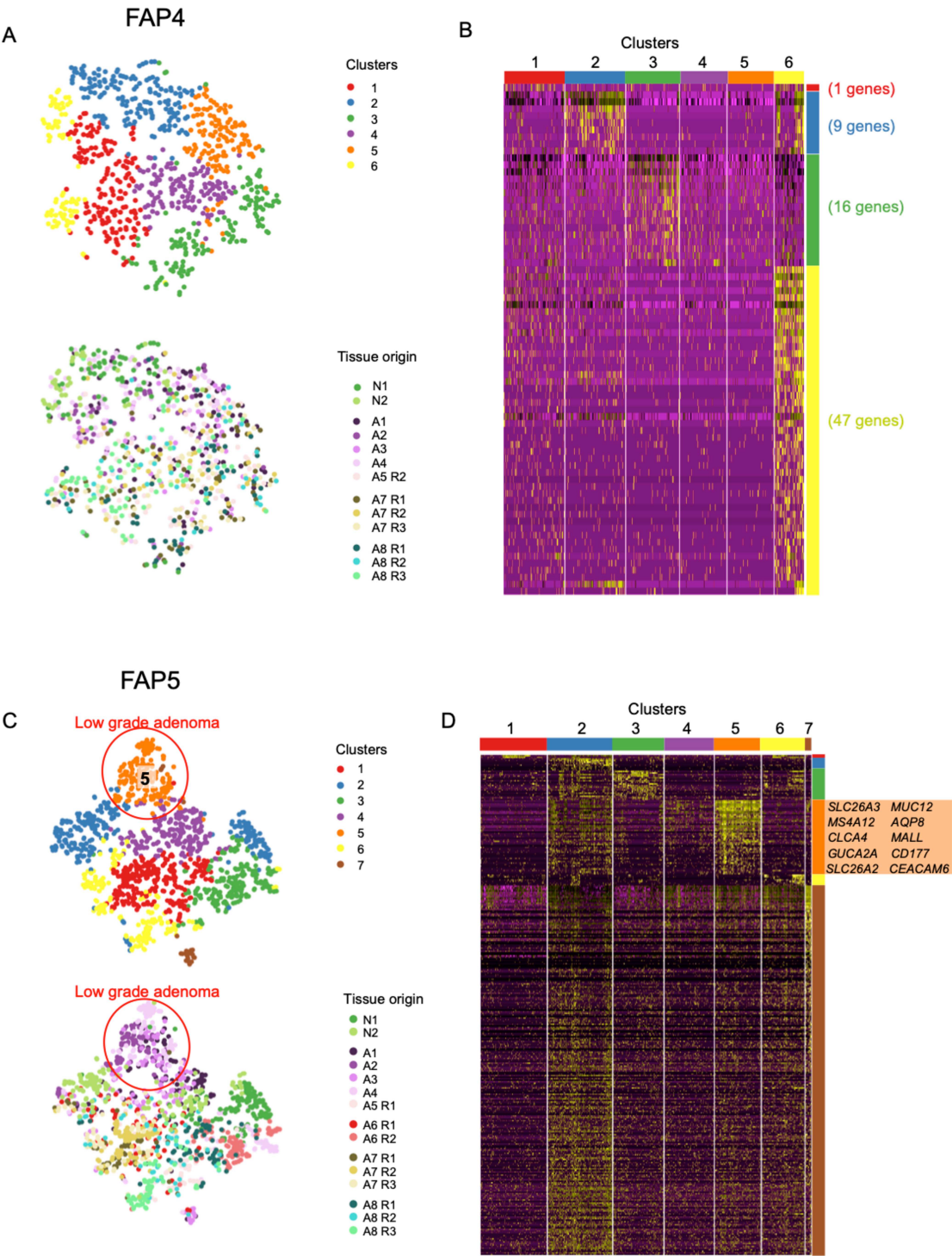
Supplementary Figure 15

Supplementary Figure 15 Transcriptome heterogeneity of lesions from CRC1 and CRC2.

(A) tSNE map of all epithelial cells from CRC1. The black circle indicates that the cells from Car R3 were tend to cluster together. (B) tSNE map of all epithelial cells from CRC2. The black circle indicates that the cells from Car R2 were tend to cluster together.
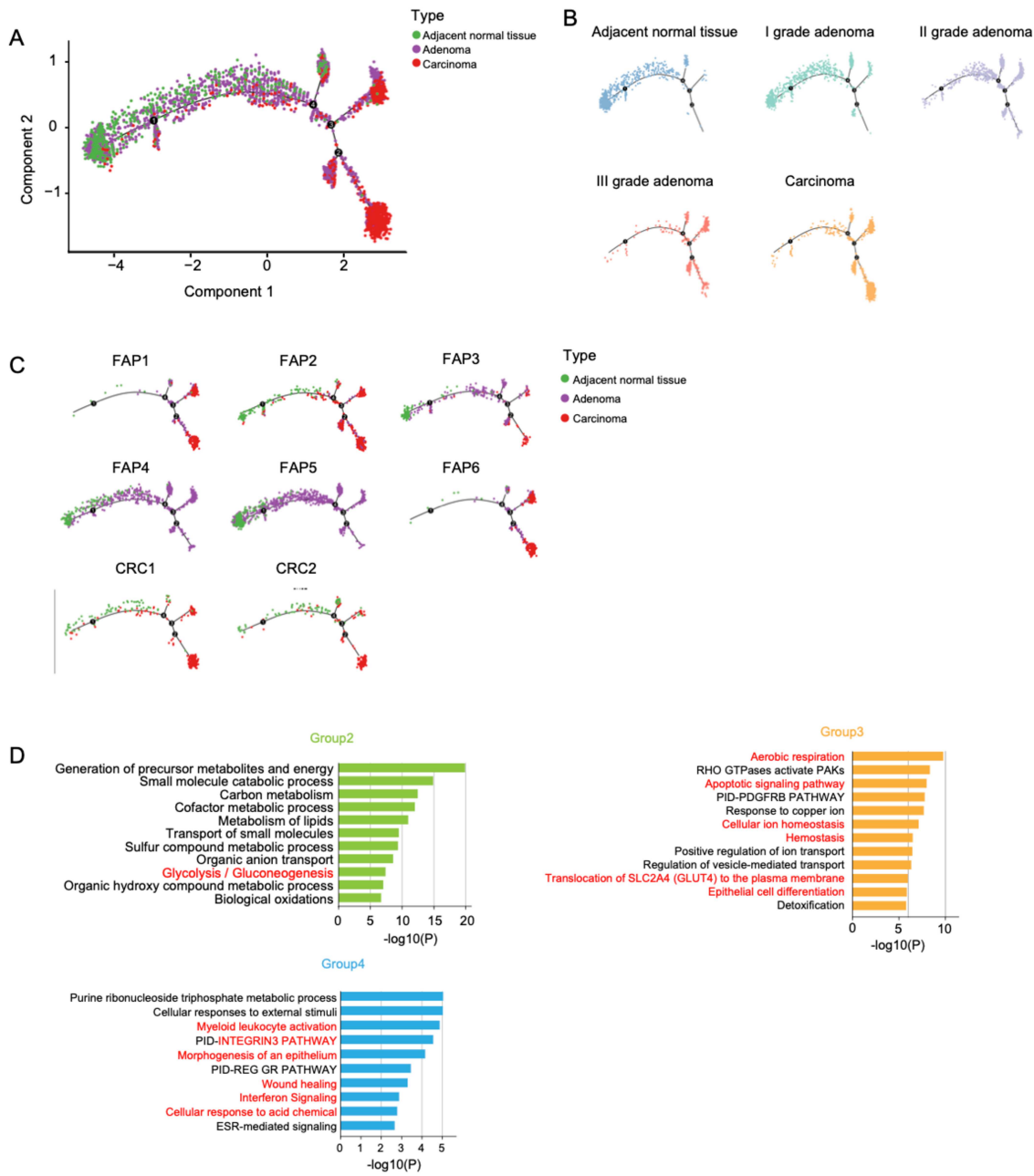
Supplementary Figure 16

FAP4

A

Clusters
- 1
- 2
- 3
- 4
- 5
- 6

Tissue origin
- N1
- N2
- A1
- A2
- A3
- A4
- A5 R2
- A7 R1
- A7 R2
- A7 R3
- A8 R1
- A8 R2
- A8 R3

B

Clusters

(1 genes)
(9 genes)
(16 genes)
(47 genes)

FAP5

C

Low grade adenoma

5

Clusters
- 1
- 2
- 3
- 4
- 5
- 6
- 7

Low grade adenoma

Tissue origin
- N1
- N2
- A1
- A2
- A3
- A4
- A5 R1
- A6 R1
- A6 R2
- A7 R1
- A7 R2
- A7 R3
- A8 R1
- A8 R2
- A8 R3

D

Clusters

SLC26A3   MUC12
MS4A12   AQP8
CLCA4   MALL
GUCA2A   CD177
SLC26A2   CEACAM6

Supplementary Figure 16 Transcriptome heterogeneity of lesions at different evolutionary stages from FAP4 and FAP5.

(A) Clustering analysis of all epithelial cells from FAP4 by using tSNE. Up, single cells in different clusters were indicated with different colors. Down, the tissue origin for each single cell was indicated with different colors. (B) Heatmap showing the specifically highly expressed genes in each cluster of cells from FAP4. The total gene numbers of cluster 1, 2, 3, 6 were showed. No highly expressed genes for cluster 4 and 5 were found. (C) Clustering analysis of all epithelial cells from FAP5 by using tSNE. Up, single cells in different clusters were indicated with different colors. Down, the tissue origin for each single cell was indicated with different colors. The cells in cluster 5 were mainly from low grade adenomas. (D) Heatmap showing the specifically highly expressed genes in each cluster of cells from FAP5. Representative genes of cluster 5 were showed.
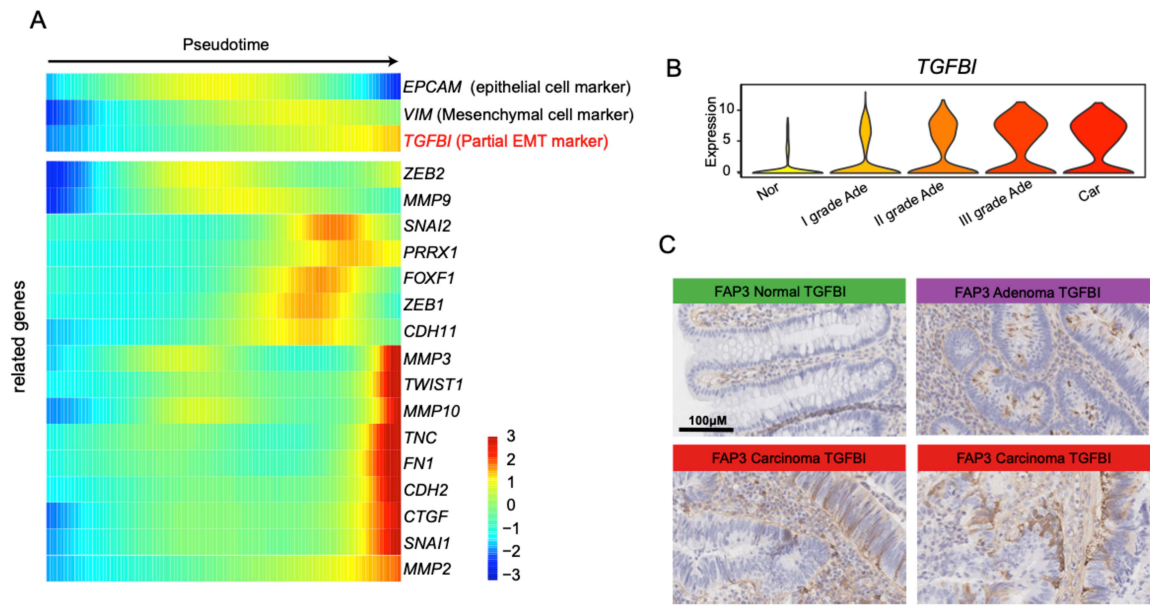
Supplementary Figure 17

Supplementary Figure 17 Construction of pseudotime trajectory of carcinogenesis.

(A) Monocle 2 was used to construct the pseudotime trajectory of all epithelial cells. Cells on the map are colored by their tissue type identity. (B) Cells on the map were colored by their pathological types. (C) Cells from different patients were separately plotted on the pseudotime map. Cells on the maps were colored by their tissue type identity. (D) Gene ontology analysis of genes from group 2, 3 and 4 in figure 5A.
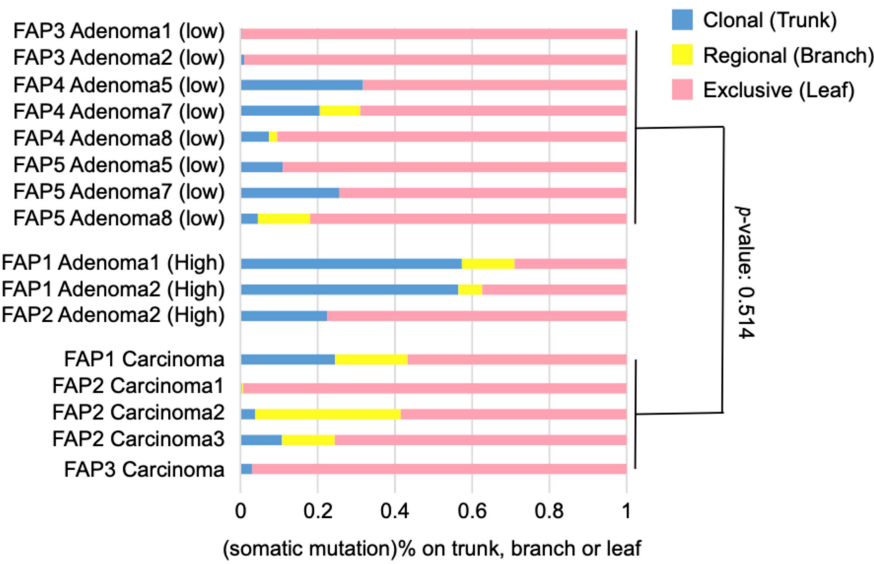
Supplementary Figure 18

Supplementary Figure 18 Transcriptome dynamics of genes related to EMT.

(A) Heatmap showing scaled expression of EMT related genes along the pseudotime.

(E) The violin plot indicating the gene expression level of *TGFBI* in adjacent normal

tissue and lesions at different evolutionary stages. (F) Immumohistochemical staining

of TGFBI in adjacent normal tissue and lesions at different evolutionary stages from

FAP3. Scale bars, 100 µm.

Supplementary Figure 19

Supplementary Figure 19 Percentage of trunk, branched and private exclusive mutations in each multiregionally sampled lesion.

Clonal (trunk) mutations were mutations that were shared by all sampled regions from same lesion. Regional (Branch) mutations were mutations that were shared by at least two but not all of the sampled regions from same lesion. Exlusive (Leaf) were mutations that detected in only one of the sampled regions from same lesion.