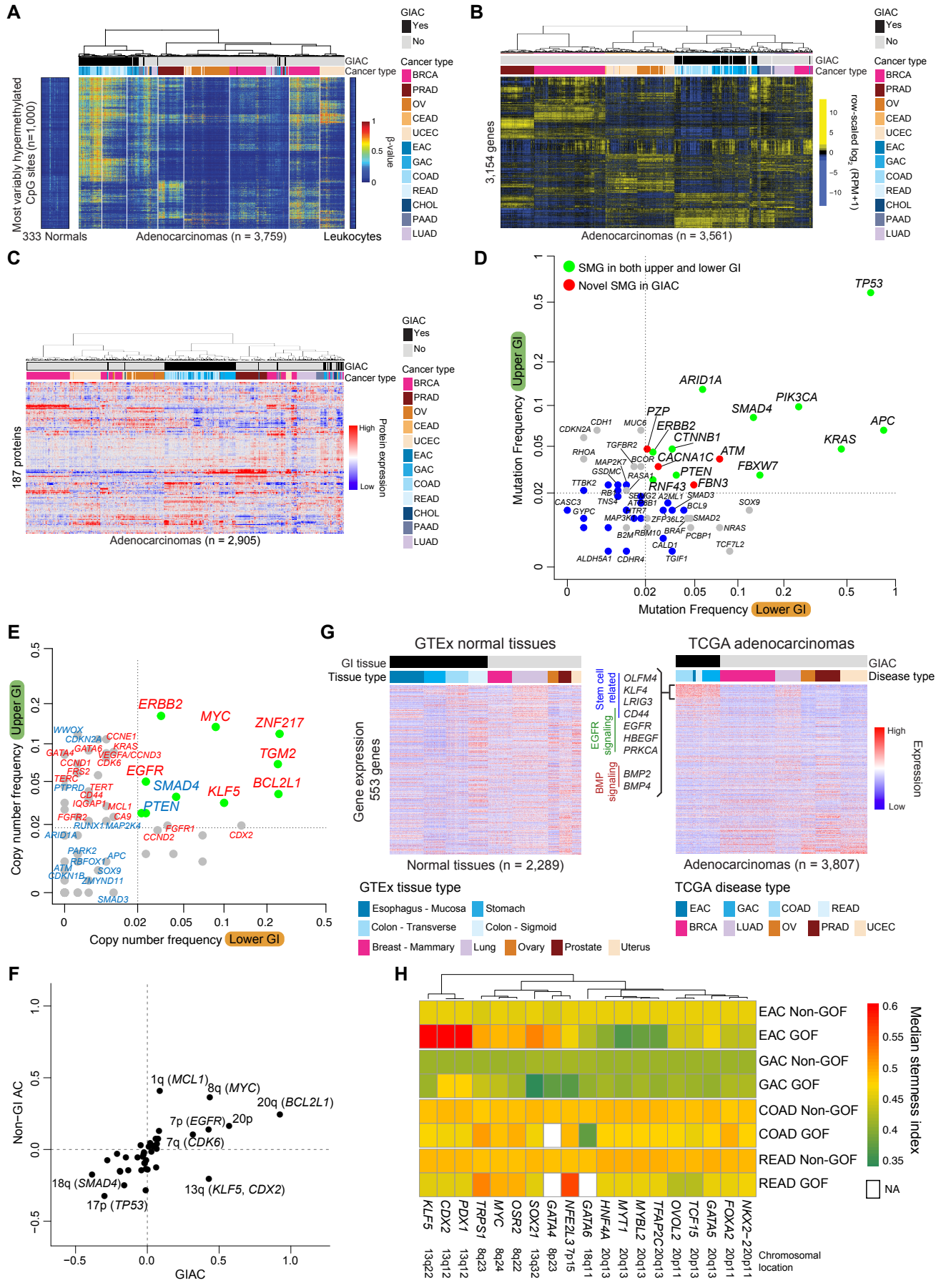
A
GIAC
Yes
No
Cancer type
BRCA
PRAD
OV
CEAD
UCEC
EAC
GAC
COAD
READ
CHOL
PAAD
LUAD
Most variably hypermethylated CpG sites (n=1,000)
β-value
333 Normals
Adenocarcinomas (n = 3,759)
Leukocytes
B
3,154 genes
row-scaled log2 (RPM+1)
Adenocarcinomas (n = 3,561)
C
187 proteins
Protein expression
High
Low
Adenocarcinomas (n = 2,905)
D
SMG in both upper and lower GI
Novel SMG in GIAC
Upper GI
Mutation Frequency
Lower GI
TP53
ARID1A
PIK3CA
SMAD4
APC
PZP
ERBB2
CTNNB1
KRAS
ATM
CACNA1C
FBXW7
PTEN
FBN3
RNF43
E
ERBB2
MYC
ZNF217
TGM2
EGFR
SMAD4
KLF5
BCL2L1
PTEN
Copy number frequency
Lower GI
G
GTEx normal tissues
TCGA adenocarcinomas
GI tissue
Tissue type
GIAC
Disease type
Gene expression
553 genes
Stem cell related
EGFR signaling
BMP signaling
OLFM4
KLF4
LRIG3
CD44
EGFR
HBEGF
PRKCA
BMP2
BMP4
High
Low
Expression
Normal tissues (n = 2,289)
Adenocarcinomas (n = 3,807)
GTEx tissue type
Esophagus - Mucosa
Stomach
Colon - Transverse
Colon - Sigmoid
Breast - Mammary
Lung
Ovary
Prostate
Uterus
TCGA disease type
EAC
GAC
COAD
READ
BRCA
LUAD
OV
PRAD
UCEC
F
Non−GI AC
GIAC
1q (MCL1)
8q (MYC)
20q (BCL2L1)
7p (EGFR)
20p
7q (CDK6)
18q (SMAD4)
13q (KLF5, CDX2)
17p (TP53)
H
EAC Non-GOF
EAC GOF
GAC Non-GOF
GAC GOF
COAD Non-GOF
COAD GOF
READ Non-GOF
READ GOF
Median stemness index
NA
KLF5
CDX2
PDX1
TRPS1
MYC
OSR2
SOX21
GATA4
GATA6
HNF4A
MYT1
MYBL2
TFAP2C
OVOL2
TCF15
GATA5
FOXA2
NKX2−2
Chromosomal location

Figure S1. Related to Figure 1.

Genomic Features of Gastrointestinal Adenocarcinomas.
(A) Unsupervised DNA methylation clustering of GIAC and non-GIAC adenocarcinomas. Tissues of origin and corresponding tumor designations: Breast, BRCA; Prostate, PRAD; Ovarian, OV; Cervix, CEAD; Endometrial, UCEC; Esophageal, EAC; Gastric, GAC; Colon, COAD; Rectal, READ; Bile duct, CHOL; Pancreas, PAAD; Lung, LUAD. **(B)** Unsupervised mRNA expression clustering of GIAC and non-GIAC. **(C)** Unsupervised RPPA clustering of GIAC and non-GIAC. **(D)** Mutation frequency of the significantly-mutated genes in upper GIAC (vertical axis) plotted against lower GIAC (horizontal axis), excluding MSI and HM-SNV cases. Genes indicated by green dots are mutated in both the upper and lower GI tract (frequency > 0.02). A subset of those that have not been previously reported by similar large-scale genomic studies are marked by red circles. Other previously unreported genes are marked by blue dots. **(E)** Frequency of significant SCNA in upper GIAC (vertical axis) plotted against lower GIAC (horizontal axis). Genes indicated by green dots are common to upper and lower GI tract (frequency > 0.02). **(G)** Gene expression analysis of GTEX normal tissue showing 553 most differentially expressed genes among GI vs non-GI normal tissue (left panel) and TCGA adenocarcinomas, highlighting elevated expression of select genes in GIAC compared to non-GIAC, after excluding differences related to normal tissue expression (right panel). **(H)** Gene expression-based stemness score indicated by color legend corresponding to select genomic loci amplifications stratified by tissue type. **(F)** Mean arm-level copy-number gain or loss in GIAC (horizontal axis) compared to non-GIAC (vertical axis).

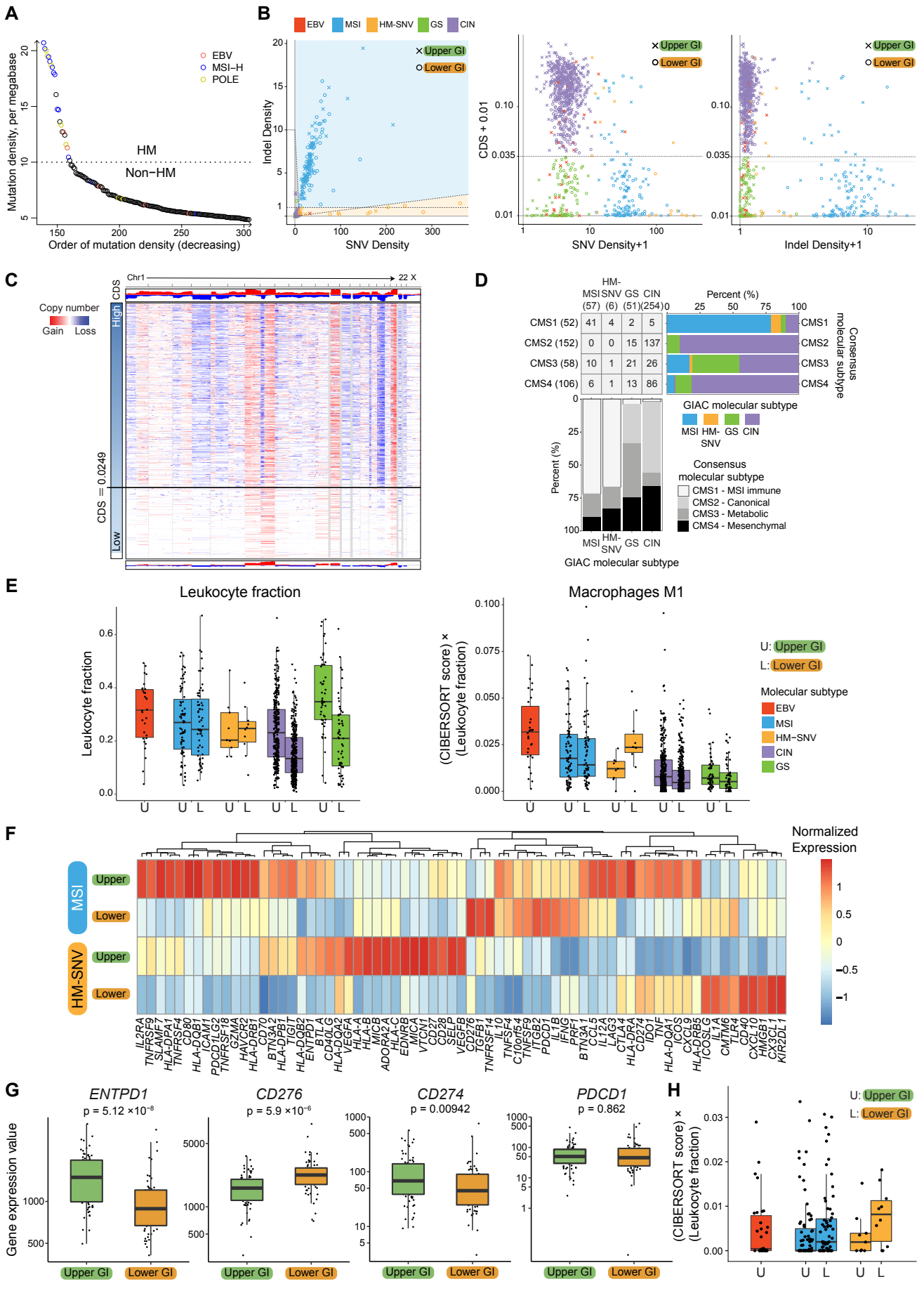
A
Mutation density, per megabase
Order of mutation density (decreasing)
EBV
MSI−H
POLE
HM
Non−HM
B
EBV
MSI
HM-SNV
GS
CIN
Indel Density
SNV Density
Upper GI
Lower GI
CDS + 0.01
SNV Density+1
Indel Density+1
C
Chr1
22 X
Copy number
Gain
Loss
CDS
High
Low
CDS = 0.0249
D
HM-SNV
MSI (57)
HM-SNV (6)
GS (51)
CIN (254)
CMS1 (52) 41 4 2 5
CMS2 (152) 0 0 15 137
CMS3 (58) 10 1 21 26
CMS4 (106) 6 1 13 86
Percent (%)
Consensus molecular subtype
GIAC molecular subtype
CMS1 - MSI immune
CMS2 - Canonical
CMS3 - Metabolic
CMS4 - Mesenchymal
E
Leukocyte fraction
Macrophages M1
(CIBERSORT score) × (Leukocyte fraction)
U: Upper GI
L: Lower GI
Molecular subtype
EBV
MSI
HM−SNV
CIN
GS
F
Normalized Expression
MSI
HM-SNV
Upper
Lower
IL2RA TNFRSF9 SLAMF7 HLA-DPA1 TNFRSF4 CD80 HLA-DQB1 PDCD1LG1 TNFRSF18 GZMA HAVCR2 HLA-DRB1 CD70 BTN3A2 HLA-DPB1 TIGIT HLA-DQB2 ENTPD1 BTLA CD40LG HLA-DQA2 VEGFA HLA-A HLA-B MICB ADORA2A HLA-C EDNRB MICA VTCN1 CD27 CD28 SELP VEGFB CD276 TGFB1 TNFRSF14 IL10 TNFSF4 C10orf54 TNFSF9 ITGB2 PDCD1 IL1B IFNG PRF1 BTN3A1 CCL5 IL12A LAG3 CTLA4 HLA-DRA CD274 IDO1 TNF HLA-DQA1 CCL3 CCL4 HLA-DRB5 ICOSLG IL1A CMTM6 TLR4 CD40 CXCL10 HMGB1 CX3CL1 KIR2DL1
G
ENTPD1
p = 5.12 ×10−8
CD276
p = 5.9 ×10−6
CD274
p = 0.00942
PDCD1
p = 0.862
Gene expression value
Upper GI
Lower GI
H
(CIBERSORT score) × (Leukocyte fraction)
U: Upper GI
L: Lower GI

Figure S2. Related to Figure 2.

Molecular Subtypes of Gastrointestinal Adenocarcinomas.
(A) GIAC samples ranked by mutational density. Horizontal dotted line indicates cutoff of 10 mutations/Mb separating hypermutated (above) from non-hypermutated (below) cancers. **(B)** Two-dimensional scatter plots showing indel vs SNV density (left panel), CDS vs SNV density (middle panel), and CDS vs indel density (right panel) in GIAC tumors, highlighting the separation among molecular subtypes: MSI tumors (blue), HM-SNV (yellow), CIN (purple), GS (blue), and EBV (red). Upper and lower GI tumors are represented by crosses and open circles, respectively. **(C)** Copy-number profiles by chromosomes are shown in the colored heatmap. Degree of amplification or deletion is indicated by intensity of red or blue, respectively. Samples are sorted by CDS as displayed in the left margin, separating CIN (upper) from GS (upper) tumors. **(D)** Contingency table comparing our GIAC molecular subtypes and the CRC molecular subtype assignments (Guinney et al., 2015). **(E)** Differences in leukocyte fraction estimated based on DNA methylation data (left) and M1-Macrophage populations from CIBERSORT adjusted for total leukocytes (right) are shown for the molecular subtypes, stratified by upper vs lower GI tumors. **(F)** Heat map depicting gene expression levels of immunomodulatory mediators in hypermutated tumors divided into MSI and HM-SNV, and further stratified by upper and lower GI tract. **(G)** Box plot indicating gene expression levels of select immunomodulatory mediators in MSI tumors stratified by upper and lower GI tract. Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range. **(H)** Resting NK cell fraction from CIBERSORT adjusted for total leukocytes in indicated molecular subtypes stratified by upper and lower GI tract.

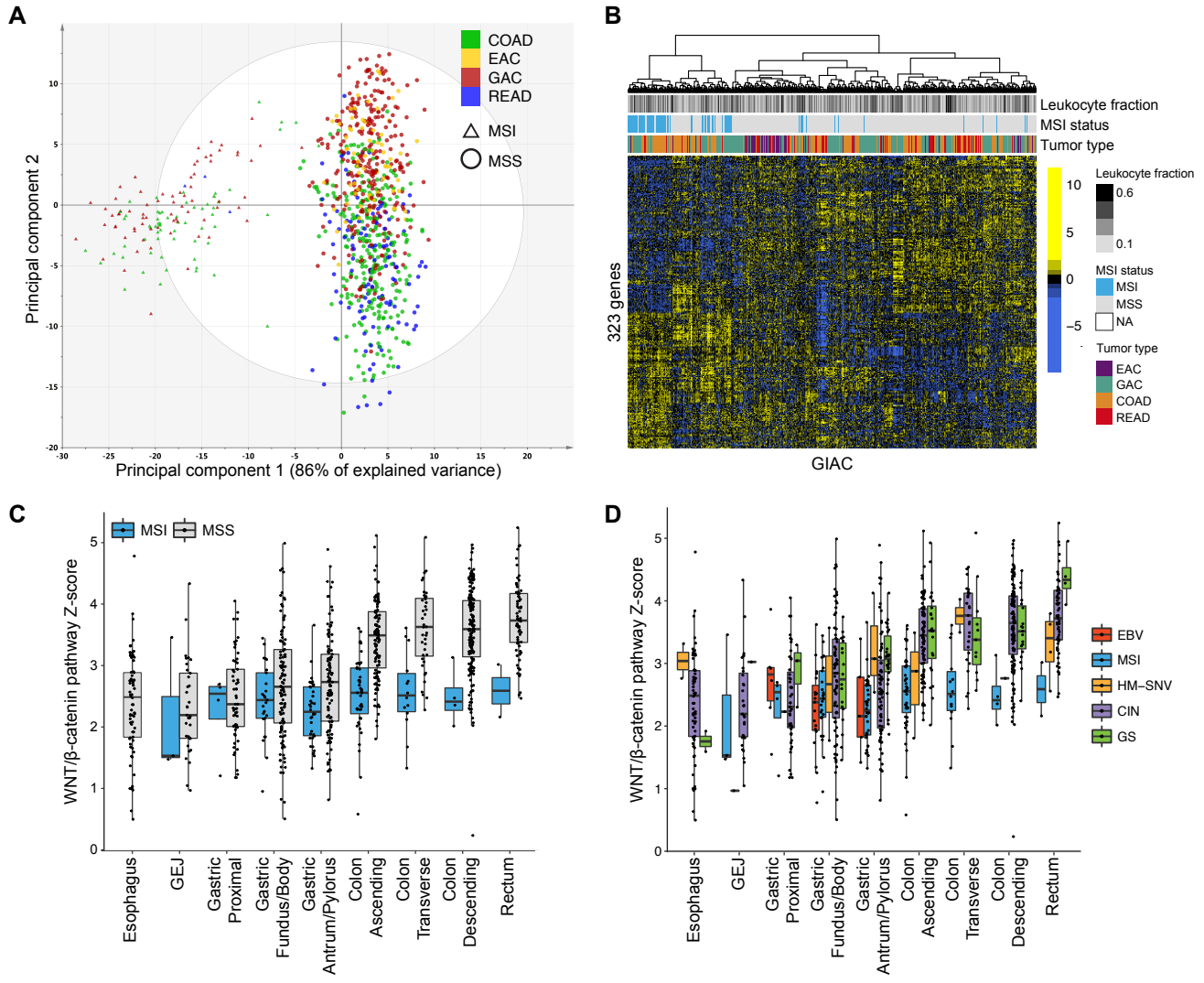
A
COAD
EAC
GAC
READ
MSI
MSS
Principal component 2
Principal component 1 (86% of explained variance)
B
Leukocyte fraction
MSI status
Tumor type
323 genes
GIAC
Leukocyte fraction
0.6
0.1
MSI status
MSI
MSS
NA
Tumor type
EAC
GAC
COAD
READ
C
MSI
MSS
WNT/β-catenin pathway Z-score
Esophagus
GEJ
Gastric Proximal
Gastric Fundus/Body
Gastric Antrum/Pylorus
Colon Ascending
Colon Transverse
Colon Descending
Rectum
D
EBV
MSI
HM−SNV
CIN
GS
WNT/β-catenin pathway Z-score
Esophagus
GEJ
Gastric Proximal
Gastric Fundus/Body
Gastric Antrum/Pylorus
Colon Ascending
Colon Transverse
Colon Descending
Rectum

Figure S3. Related to Figure 3.

Analysis of CIMP and MSI Tumors.
(A) Supervised multivariate analysis of GIAC tumors showing MSI subtype as triangles; colors indicate cancer type: colon (green), esophagus (yellow), stomach (red), and rectum (blue). **(B)** heatmap depicting unsupervised mRNA expression analysis of GIAC after removing tissue-specific effects. **(C)** Box plots showing WNT/β-catenin pathway score as measured by single sample Gene Set Enrichment Analysis (ssGSEA) across the GI tract stratified by MSI and non-MSI tumors. Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range. **(D)** WNT/β-catenin pathway score as derived from ssGSEA across the anatomic gradient, stratified by molecular subtype. Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range

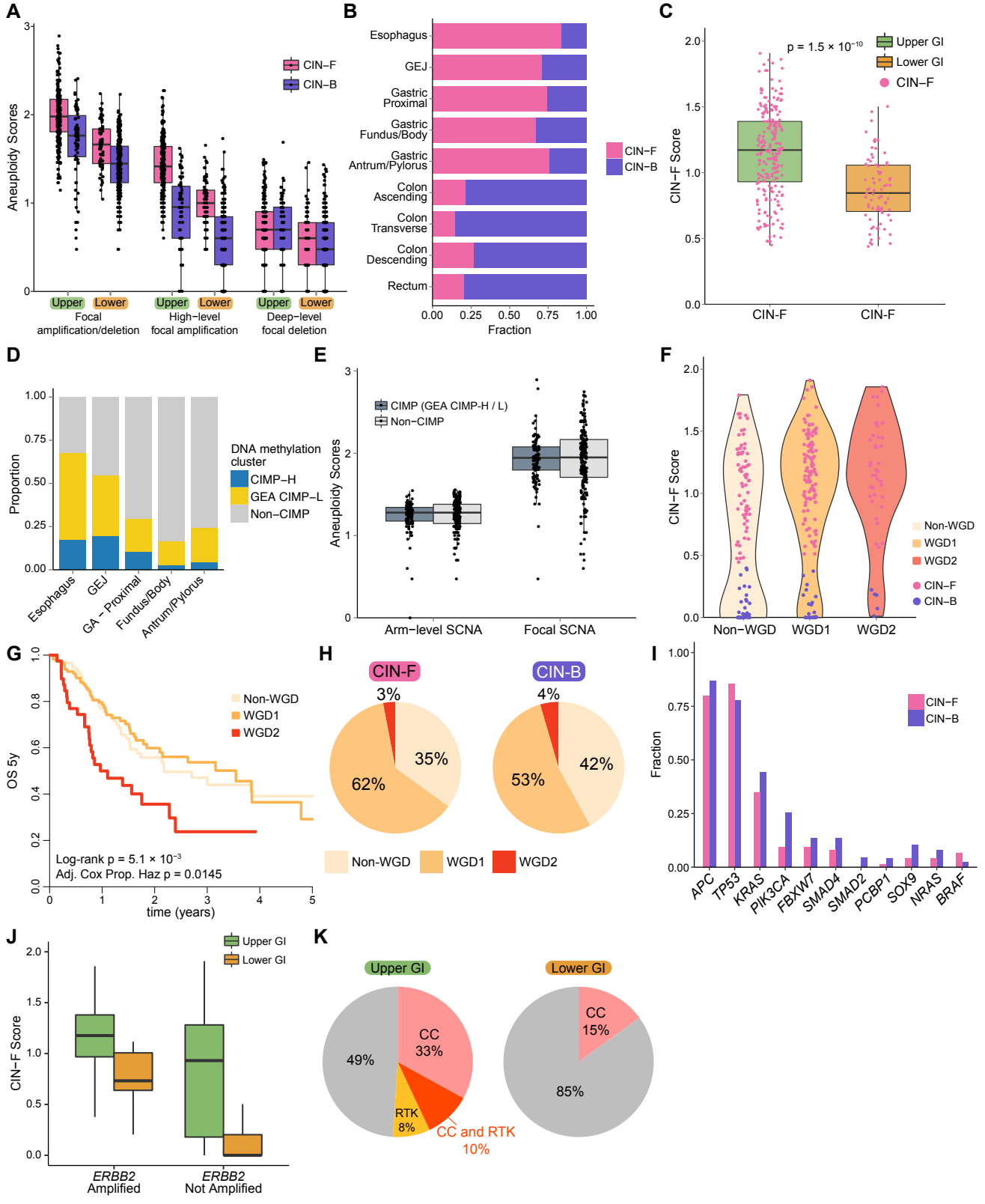
A
Aneuploidy Scores
CIN−F
CIN−B
Upper
Lower
Focal
amplification/deletion
High−level
focal amplification
Deep−level
focal deletion
B
Esophagus
GEJ
Gastric Proximal
Gastric Fundus/Body
Gastric Antrum/Pylorus
Colon Ascending
Colon Transverse
Colon Descending
Rectum
Fraction
C
p = 1.5 × 10−10
Upper GI
Lower GI
CIN−F
CIN−F Score
CIN-F
CIN-F
D
Proportion
DNA methylation cluster
CIMP−H
GEA CIMP−L
Non−CIMP
Esophagus
GEJ
GA − Proximal
Fundus/Body
Antrum/Pylorus
E
CIMP (GEA CIMP-H / L)
Non−CIMP
Aneuploidy Scores
Arm−level SCNA
Focal SCNA
F
CIN−F Score
Non−WGD
WGD1
WGD2
CIN−F
CIN−B
Non−WGD
WGD1
WGD2
G
OS 5y
Non−WGD
WGD1
WGD2
Log-rank p = 5.1 × 10−3
Adj. Cox Prop. Haz p = 0.0145
time (years)
H
CIN-F
3%
35%
62%
CIN-B
4%
42%
53%
Non−WGD
WGD1
WGD2
I
Fraction
CIN−F
CIN−B
APC
TP53
KRAS
PIK3CA
FBXW7
SMAD4
SMAD2
PCBP1
SOX9
NRAS
BRAF
J
Upper GI
Lower GI
CIN−F Score
ERBB2
Amplified
ERBB2
Not Amplified
K
Upper GI
CC
33%
49%
RTK
8%
CC and RTK
10%
Lower GI
CC
15%
85%

Figure S4. Related to Figure 4.

Molecular Features of CIN.
(A) Box plot showing aneuploidy scores for focal SCNA categories including general focal SCNA events (amplifications/deletions), high-level focal amplifications, and deep-level focal deletions in CIN-F and CIN-B tumors in the upper and lower GI tract (see methods for score details). Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range. **(B)** Proportion of CIN-F and CIN-B tumors across the anatomic gradient in GIAC CIN. **(C)** Distribution of CIN-F tumors in upper and lower GI tract. Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range. **(D)** Proportion of CIN GEA tumors stratified by CIMP-H, GEA CIMP-L, and non-CIMP status across the upper GI anatomic gradient. **(E)** Box plot showing aneuploidy scores (arm-level, and focal events) in upper GI CIN tumors stratified as CIMP-H/L and non-CIMP. Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range. **(F)** Distribution of CIN-F and CIN-B tumors in the upper GI tract stratified by whole genome doubling (WGD) status. **(G)** Kaplan-Meier curve indicating proportion of overall survival in patients with upper GI CIN tumors stratified by WGD status. **(H)** Proportion of WGD scored as Non-WGD, one WGD (WGD1), or more than one WGD (WGD2) in CIN-F and CIN-B tumors in the lower GI tract. **(I)** Frequency of somatic mutation in indicated genes between CIN-F and CIN-B tumors in the lower GI tract. **(J)** Box plot displaying CIN-F score in ERBB2-amplified and non-amplified CIN tumors in the upper and lower GI tract. Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range. **(K)** Distribution of co-amplified genomic events in ERBB2-amplified CIN tumors in the upper compared to lower GI tract. "RTK" indicates Receptor Tyrosine Kinase genes; "CC" indicates Cell Cycle regulator genes. Gray indicates the fraction of tumors with neither RTK, nor CC gene amplification among *ERBB2* amplified tumors.

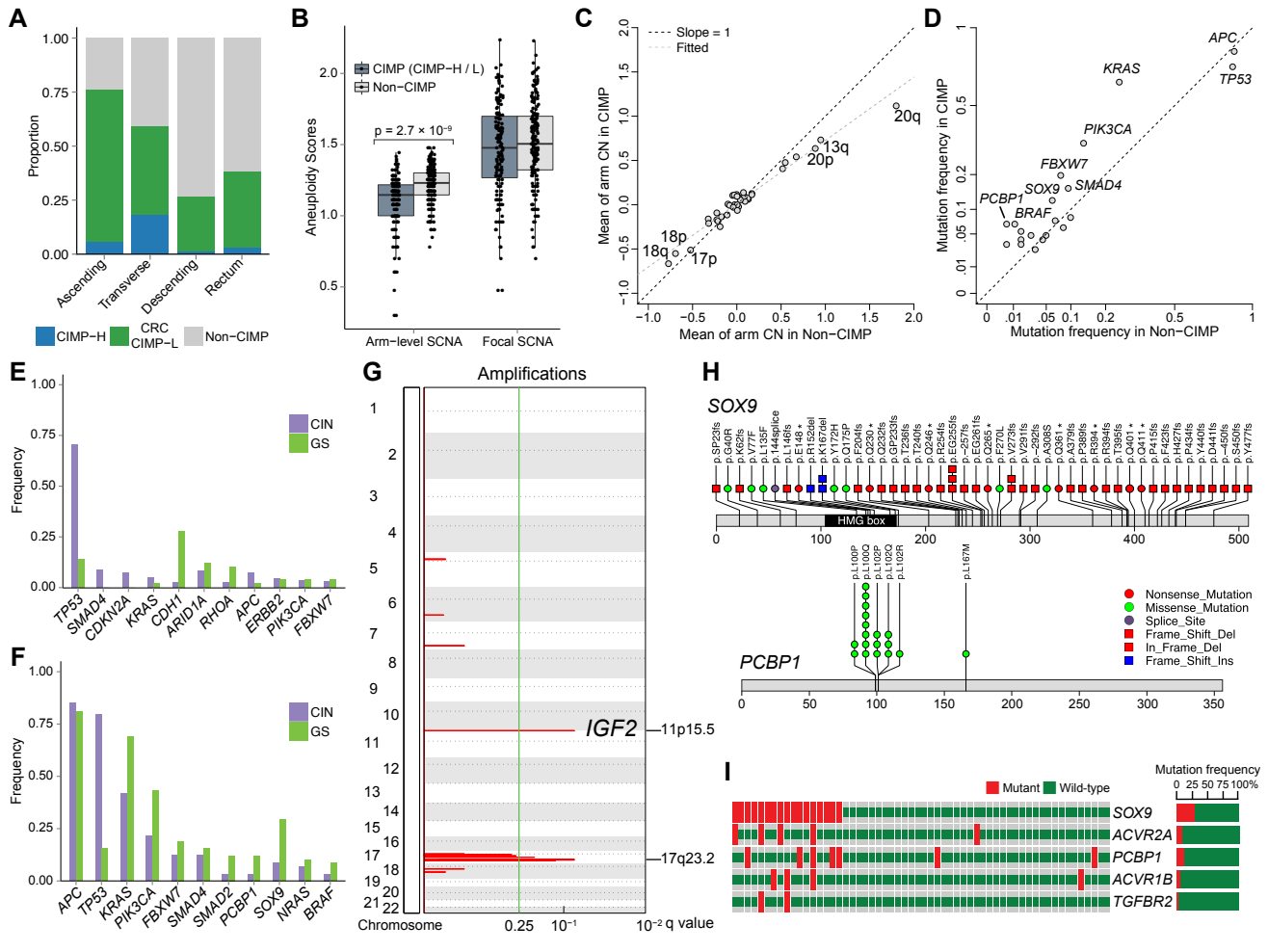
A
Proportion
1.00
0.75
0.50
0.25
0.00
Ascending
Transverse
Descending
Rectum
CIMP−H
CRC CIMP−L
Non−CIMP
B
Aneuploidy Scores
CIMP (CIMP−H / L)
Non−CIMP
p = 2.7 × 10−9
Arm−level SCNA
Focal SCNA
C
Slope = 1
Fitted
Mean of arm CN in CIMP
Mean of arm CN in Non−CIMP
20q
13q
20p
18p
18q
17p
D
Mutation frequency in CIMP
Mutation frequency in Non−CIMP
APC
TP53
KRAS
PIK3CA
FBXW7
SOX9
SMAD4
PCBP1
BRAF
E
Frequency
CIN
GS
TP53
SMAD4
CDKN2A
KRAS
CDH1
ARID1A
RHOA
APC
ERBB2
PIK3CA
FBXW7
F
Frequency
CIN
GS
APC
TP53
KRAS
PIK3CA
FBXW7
SMAD4
SMAD2
PCBP1
SOX9
NRAS
BRAF
G
Amplifications
IGF2
11p15.5
17q23.2
Chromosome
q value
H
SOX9
HMG box
Nonsense_Mutation
Missense_Site
Splice_Site
Frame_Shift_Del
In_Frame_Del
Frame_Shift_Ins
PCBP1
I
Mutant
Wild-type
Mutation frequency
SOX9
ACVR2A
PCBP1
ACVR1B
TGFBR2

Figure S5. Related to Figure 5.

Molecular Features of CIN and GS Colorectal Cancer.
(A) Proportion of lower GI CIN tumors stratified by CIMP-H, CRC CIMP-L, and non-CIMP as represented by fraction per anatomic site. **(B)** Box plot showing aneuploidy scores (arm-level and focal events) in lower GI CIN tumors stratified by CIMP-H/L and non-CIMP Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range. **(C)** Mean arm-level SCNA of indicated chromosome arms in CRC CIN tumors plotted by CIMP-H/L versus non-CIMP status. Distribution along the reference slope (y = x; dotted-line) indicates equal mean arm-level SCNA in CIMP-H/L versus non-CIMP groups. **(D)** Frequency of indicated mutated genes in lower GI CIN tumors plotted by CIMP-H/L or non-CIMP status. Distribution along the reference slope (x/y =1; dotted-line) indicates equal mean arm-level SCNA in CIMP-H/L and non-CIMP groups. **(E)** Frequency of somatic mutations in indicated genes in upper GI tract stratified by CIN and GS molecular subtypes. **(F)** Frequency of somatic mutations in indicated genes in lower GI tract stratified by CIN and GS molecular subtypes. **(G)** GISTIC plot showing amplifications in lower GI GS, highlighting a peak at 11p15.5 containing a gene that encodes *IGF2.* **(H)** Schematic showing distribution of specific somatic mutations (subdivided into nonsense, missense, splice site, frameshift deletion, in-frame deletion, and frameshift insertion) along the *SOX9* and *PCBP1* gene locus in all lower GI tumors. **(I)** CoMut plot showing significant association between mutations in *SOX9* and indicated genes of the TGFβ pathway in lower GI GS tumors.

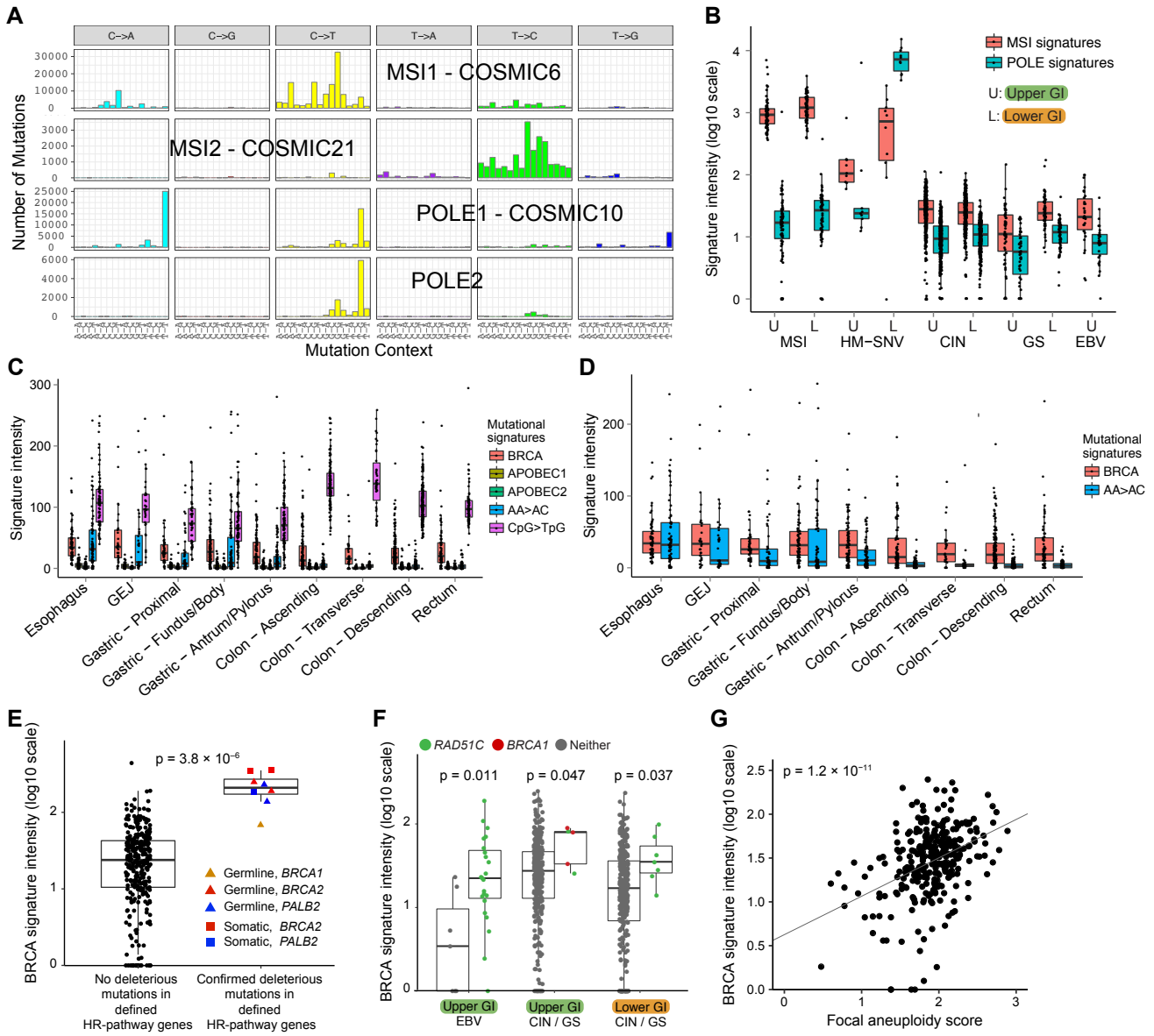
A
C->A
C->G
C->T
T->A
T->C
T->G
MSI1 - COSMIC6
MSI2 - COSMIC21
POLE1 - COSMIC10
POLE2
Number of Mutations
Mutation Context
B
MSI signatures
POLE signatures
U: Upper GI
L: Lower GI
Signature intensity (log10 scale)
MSI
HM-SNV
CIN
GS
EBV
C
Signature intensity
Mutational signatures
BRCA
APOBEC1
APOBEC2
AA>AC
CpG>TpG
Esophagus
GEJ
Gastric - Proximal
Gastric - Fundus/Body
Gastric - Antrum/Pylorus
Colon - Ascending
Colon - Transverse
Colon - Descending
Rectum
D
Signature intensity
Mutational signatures
BRCA
AA>AC
E
BRCA signature intensity (log10 scale)
p = 3.8 × 10^-6
Germline, BRCA1
Germline, BRCA2
Germline, PALB2
Somatic, BRCA2
Somatic, PALB2
No deleterious mutations in defined HR-pathway genes
Confirmed deleterious mutations in defined HR-pathway genes
F
RAD51C
BRCA1
Neither
p = 0.011
p = 0.047
p = 0.037
BRCA signature intensity (log10 scale)
Upper GI EBV
Upper GI CIN / GS
Lower GI CIN / GS
G
p = 1.2 × 10^-11
BRCA signature intensity (log10 scale)
Focal aneuploidy score

Figure S6. Related to Figure 6.

Gastrointestinal Adenocarcinoma Mutational Signatures.
(A) Significant enrichment of MSI-1, MSI-2, POLE-1, and POLE-2 COSMIC mutational signatures in GIAC displayed using the 96-substitution classification, which is defined by the substitution class and sequence context immediately 3' and 5' to the mutated base. **(B)** Box plot showing MSI and POLE mutational signature activity in GIAC molecular subtypes stratified by upper and lower GI tract. Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range. **(C)** BRCA, APOBEC1, APOBEC2, AA>AC, and CpG>TpG mutational signature activities in CIN and GS tumors across the anatomic gradient. Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range. **(D)** Box plot showing BRCA and AA>AC mutational signature activities in CIN tumors across the anatomic gradient. Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range. **(E)** Box plot showing BRCA signature intensity in relation to confirmed deleterious mutations in defined HR-pathway genes. Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range. **(F)** Box plot showing BRCA signature intensity in RAD51C and BRCA1 epigenetically silenced upper EBV, upper CIN/GS, and lower CIN/GS tumors. Horizontal bars indicate median values, boxes represent interquartile range, and whiskers indicate values within 1.5 times interquartile range. **(G)** Scatter plot displaying BRCA signature intensity plotted against focal aneuploidy score in upper GI CIN tumors.