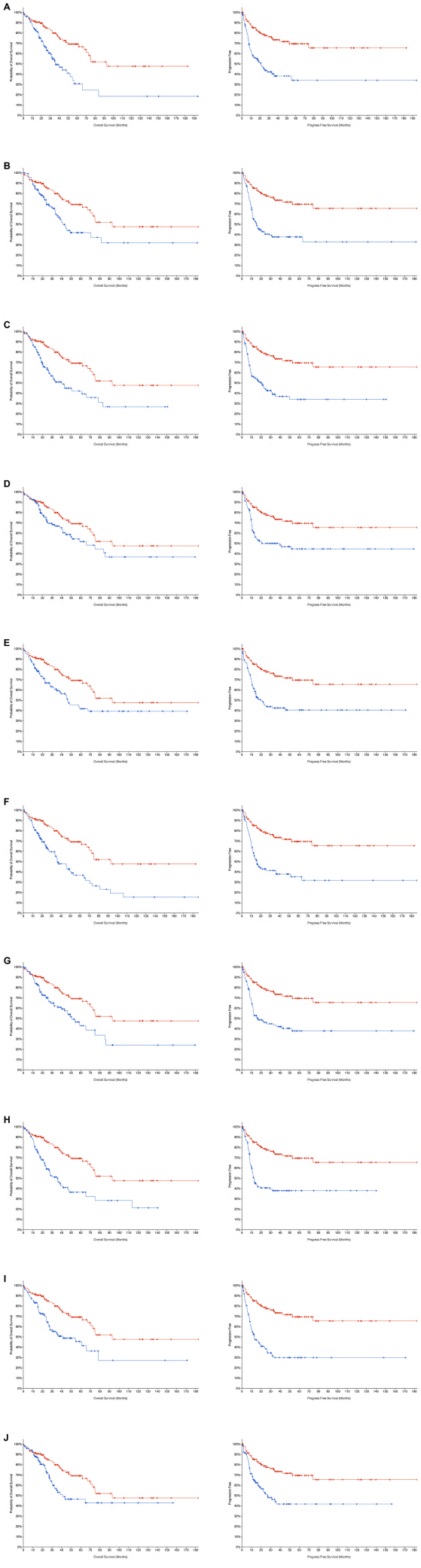
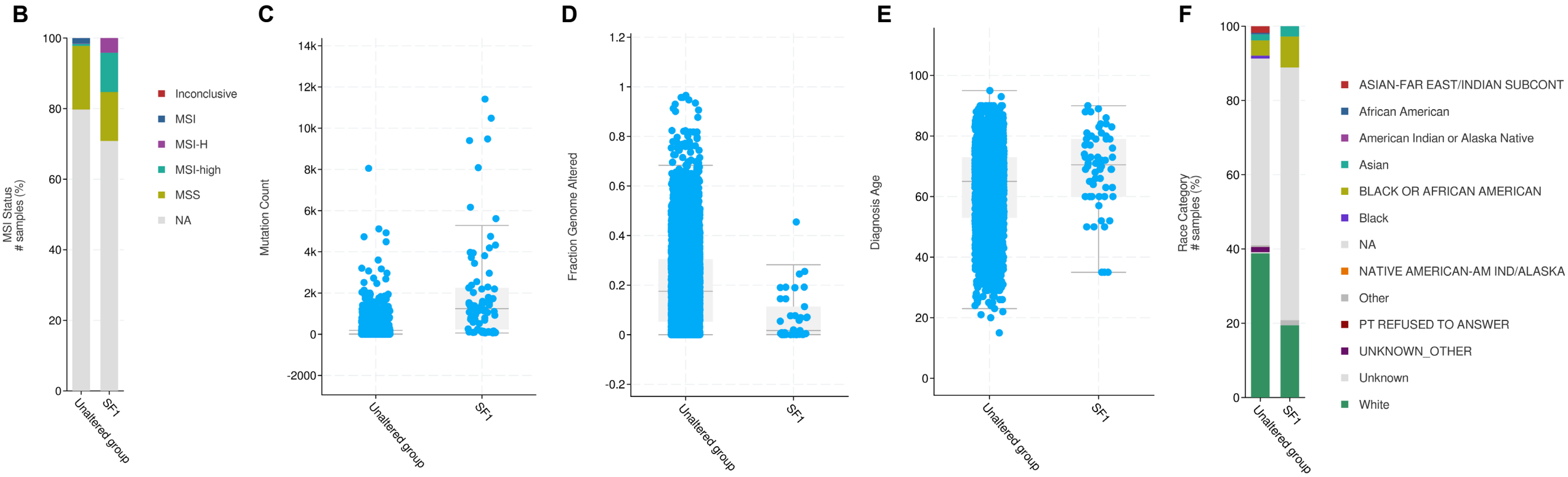
Supplementary Figure 1



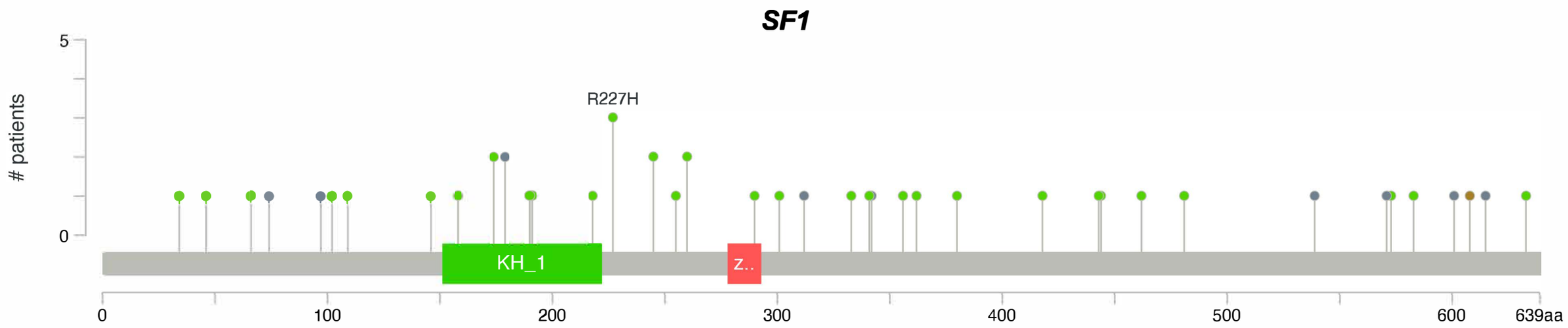
Suppl. Figure 2

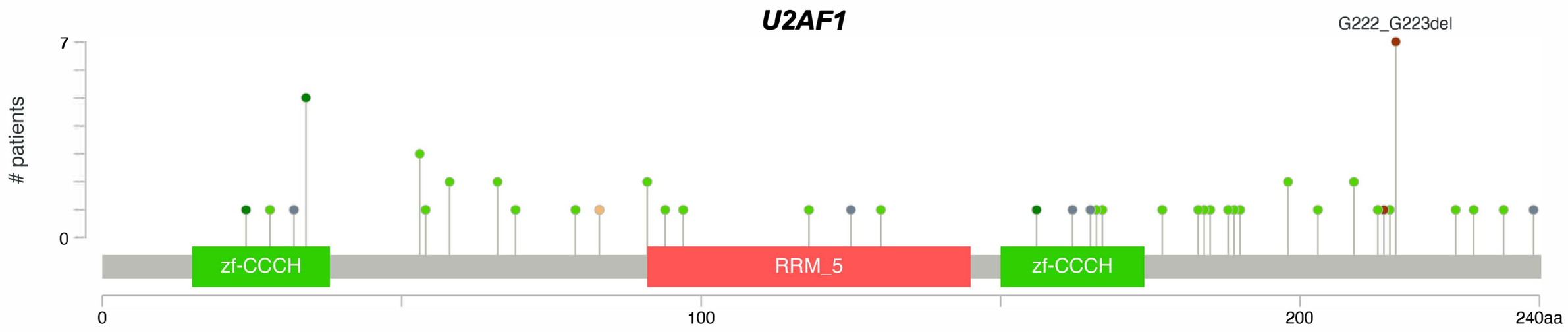
A

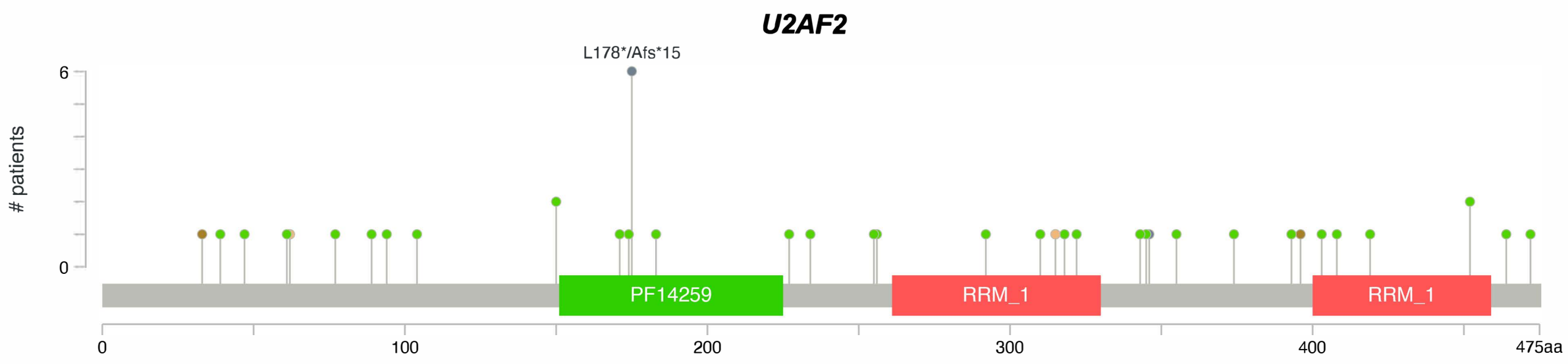
Clinical Attribute	Attribute Type	Statistical Test	p-Value	q-Value
MSI Status	Sample	Chi-squared Test	$< 10^{-10}$	$< 10^{-10}$
Mutation Count	Sample	Wilcoxon Test	$< 10^{-10}$	$< 10^{-10}$
Fraction Genome Altered	Sample	Wilcoxon Test	3.77E-8	1.033E-06
Diagnosis Age	Patient	Wilcoxon Test	9.910E-4	0.0106
Race Category	Patient	Chi-squared Test	$< 10^{-10}$	$< 10^{-10}$

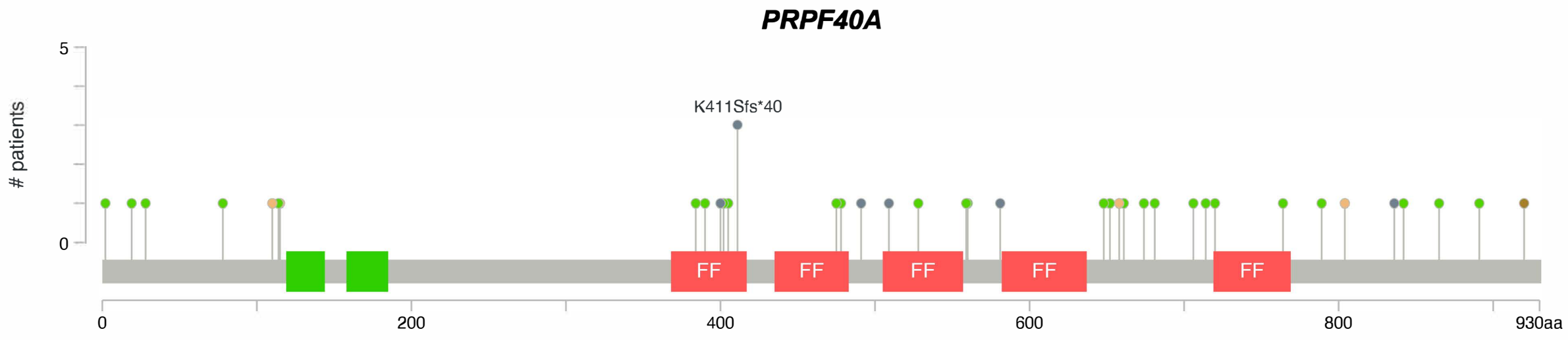


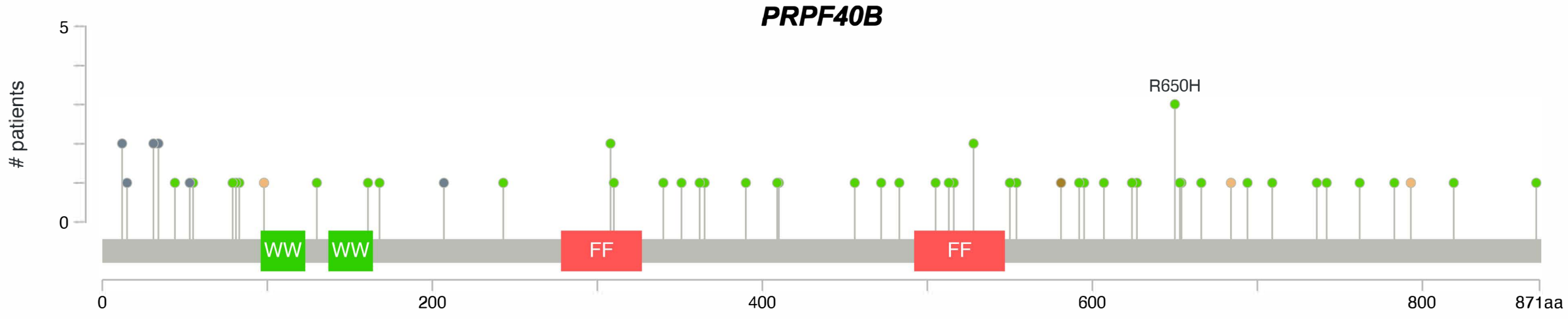
Suppl. Figure 3

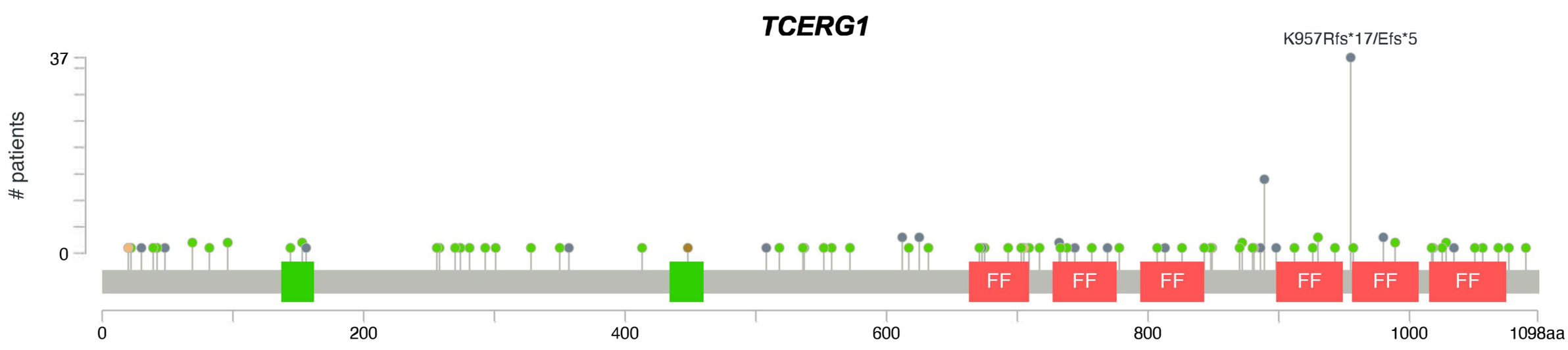






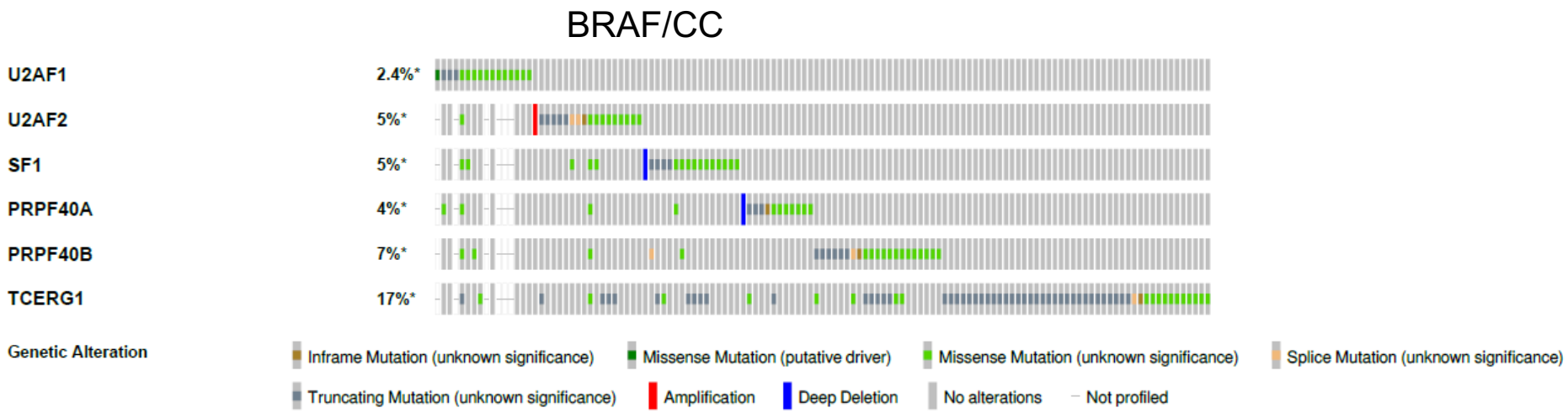




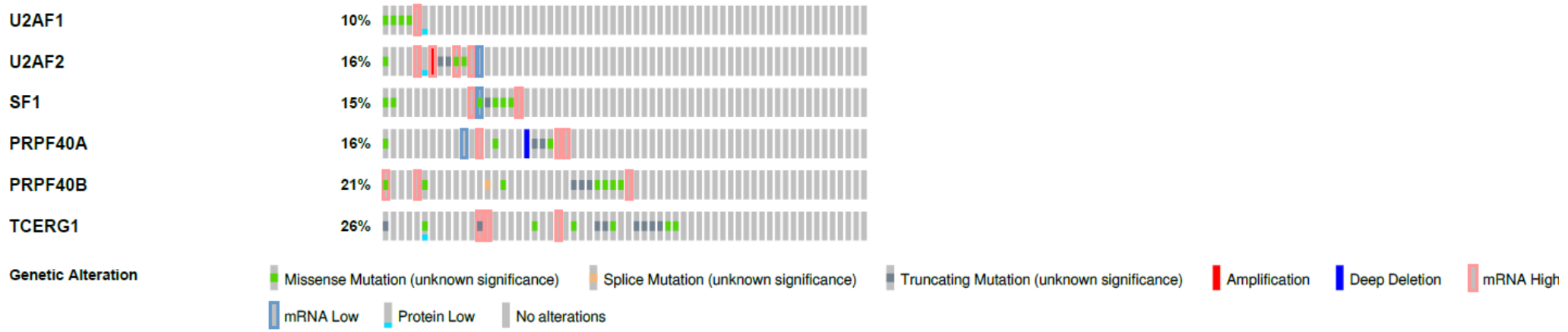


Suppl. Figure 4

A



B



Suppl Figure 5

Suppl. Figure 6

Supplementary Table 1

A	B	Neither	A Not B	B Not A	Both	Log2 Odds Ratio	p-Value	q-Value	Tendency
PRPF40B	TCERG1	1481	41	100	16	2.531	<0.001	<0.001	Co-occurrence
SF1	TCERG1	1488	34	103	13	2.466	<0.001	<0.001	Co-occurrence
U2AF2	TCERG1	1493	29	104	12	2.571	<0.001	<0.001	Co-occurrence
SF1	PRPF40B	1543	38	48	9	2.929	<0.001	<0.001	Co-occurrence
PRPF40A	TCERG1	1494	28	107	9	2.166	<0.001	0.002	Co-occurrence
U2AF2	PRPF40B	1546	35	51	6	2.378	0.002	0.006	Co-occurrence
SF1	PRPF40A	1559	42	32	5	2.536	0.003	0.007	Co-occurrence
SF1	U2AF2	1555	42	36	5	2.362	0.005	0.01	Co-occurrence
U2AF2	PRPF40A	1564	37	33	4	2.357	0.012	0.02	Co-occurrence

Supplementary Table 2

Organ/System	TCGA PanCancer Atlas study	No. of patients	Genetic alteration frequency of CC	Genetic + expression level alteration frequency of CC
Adrenal Gland	Adrenocortical Carcinoma	89	9%	46%
Biliary Tract	Cholangiocarcinoma	36	6%	36%
Bladder/Urinary Tract	Bladder Urothelial Carcinoma	406	15%	39%
Bowel	Colorectal Adenocarcinoma	526	13%	38%
Breast	Breast Invasive Carcinoma	996	7%	34%
CNS/Brain	Brain Lower Grade Glioma	511	7%	39%
	Glioblastoma Multiforme	378	5%	15%
Cervix	Cervical Squamous Cell Carcinoma	278	10%	38%
Esophagus/Stomach	Esophageal Adenocarcinoma	182	9%	32%
	Stomach Adenocarcinoma	434	15%	37%
Eye	Uveal Melanoma	80	3%	36%
Head and Neck	Head and Neck Squamous Cell Carcinoma	496	10%	34%
Kidney	Renal Clear Cell Carcinoma	354	10%	29%
	Kidney Chromophobe	65	2%	31%
	Renal Papillary Cell Carcinoma	274	4%	23%
Liver	Hepatocellular Carcinoma	353	5%	30%
Lung	Lung Adenocarcinoma	507	15%	37%
	Lung Squamous Cell Carcinoma	469	11%	34%
Lymphoid	Diffuse Large B-Cell Lymphoma	37	16%	32%
Myeloid	Acute Myeloid Leukemia	190	7%	25%
Ovary/Fallopian Tube	Ovarian Serous Cystadenocarcinoma	398	10%	23%
Pancreas	Pancreatic Adenocarcinoma	175	6%	34%
Pleura	Mesothelioma	82	2%	29%
Prostate	Prostate Adenocarcinoma	489	7%	27%
Skin	Cutaneous Melanoma	363	21%	38%
Soft Tissue	Pheochromocytoma and Paraganglioma	161	2%	31%
	Sarcoma	253	9%	32%
Testis	Testicular Germ Cell Tumors	144	1%	38%
Thymus	Thymoma	123	2%	19%
Thyroid	Thyroid Carcinoma	482	<1%	16%
Uterus	Uterine Corpus Endometrial Carcinoma	509	23%	46%
	Uterine Carcinosarcoma	56	14%	59%

Supplementary Table 3

Driver Gene	A	B	Neither	A Not B	B Not A	Both	Log2 Odds Ratio	p-Value	q-Value	Tendency
APC	**U2AF2**	**TCERG1**	998	14	49	10	>3	<0.001	**<0.001**	**Co-occurrence**
	SF1	**TCERG1**	998	14	50	9	>3	<0.001	**<0.001**	**Co-occurrence**
	SF1	**PRPF40B**	1022	16	26	7	>3	<0.001	**<0.001**	**Co-occurrence**
	PRPF40A	**TCERG1**	993	19	51	8	>3	<0.001	**<0.001**	**Co-occurrence**
	PRPF40B	**TCERG1**	987	25	51	8	2.631	<0.001	**<0.001**	**Co-occurrence**
	SF1	**U2AF2**	1028	19	20	4	>3	0.001	**0.003**	**Co-occurrence**
	SF1	**PRPF40A**	1025	19	23	4	>3	0.002	**0.004**	**Co-occurrence**
	U2AF2	**PRPF40B**	1018	20	29	4	2.812	0.005	**0.01**	**Co-occurrence**
	PRPF40A	**PRPF40B**	1015	23	29	4	2.606	0.008	**0.013**	**Co-occurrence**
	U2AF2	**PRPF40A**	1023	21	24	3	2.606	0.02	**0.03**	**Co-occurrence**
KRAS	**SF1**	**TCERG1**	541	11	20	6	>3	<0.001	**<0.001**	**Co-occurrence**
	PRPF40A	**TCERG1**	544	8	21	5	>3	<0.001	**<0.001**	**Co-occurrence**
	SF1	**PRPF40B**	554	13	7	4	>3	<0.001	**<0.001**	**Co-occurrence**
	U2AF2	**TCERG1**	542	10	22	4	>3	0.002	**0.008**	**Co-occurrence**
	U2AF2	**PRPF40A**	554	11	10	3	>3	0.003	**0.008**	**Co-occurrence**
	PRPF40B	**TCERG1**	544	8	23	3	>3	0.011	**0.026**	**Co-occurrence**
TP53	**PRPF40A**	**TCERG1**	822	11	31	3	2.854	0.015	0.148	Co-occurrence
	U2AF2	**PRPF40A**	841	12	12	2	>3	0.02	0.148	Co-occurrence
	SF1	**TCERG1**	818	15	31	3	2.4	0.03	0.15	Co-occurrence
BRAF	**PRPF40B**	**TCERG1**	195	15	57	11	1.327	0.033	0.268	Co-occurrence

Supplementary Figure Legends & Table Legends

Supplementary Fig 1. Comparing equal number of unaltered (randomized samples) with altered patient samples. Overall survival (left) (*q* is between 1.79e-05 and 0.0187 for all except D, *q* = 0.0963) and progression free survival (right) (*q* is in between 1.38e-09 and 2.95e-04 for all) of patients with or without genetic alterations in E-complex genes. For each panel, data was obtained from a total of 132 patients with at least one alteration in the 6 queried E-complex genes (Altered group, indicated by the red curve), and a randomized cohort of 132 patients derived from 1219 patients without any alterations in any E-complex factors (Unaltered group, indicated by the blue curve).

Supplementary Fig 2. Selected clinical attributes significantly associated with SF1 mutations. Altered group (patients with *SF1* mutations) and Unaltered group (patients with no mutations in any E-complex factor) are compared for MSI (microsatellite instability) profile (A and B), Mutation Count (A and C), Fraction Genome Altered (A and D), Diagnosis Age (A and E) and Race Category or demographic profile (A and F). Patients with *SF1* mutations (Altered group) had lower mean levels of genome alterations, higher mutation count and diagnosis age and different MSI (microsatellite instability) and demographic profile.

Supplementary Fig 3. Location and frequency of genetic mutations in SF1, U2AF1, U2AF2, PRPF40A, PRPF40B and TCERG1. The horizontal bars represent the amino acid length of each protein while red and green boxes indicate specific functional motifs such as KH-1, zf-CCCH or RRM domains. The height of each lollipop indicates the frequency of mutations in patients. Green lollipops indicate missense mutations, gray indicates truncating mutations, brown indicates in-frame mutations, orange indicates splice mutations and purple indicates fusion mutations.

Supplementary Fig 4. Oncoprint showing incidence of genetic alterations of U2AF1, U2AF2, SF1, PRPF40A, PRPF40B and TCERG1 in (A) BRAF cohorts from all bowel cancer studies and (B) BRAF cohort from TCGA PanCancer Atlas database.

Supplementary Fig 5. Heat map derived from QIAGEN Ingenuity Pathway Analysis (IPA) using gene list of co-expressed mRNAs associated with U2AF1, U2AF2, SF1, PRPF40A, PRPF40B and TCERG1 in cancers of the esophagus, stomach and bowel (left) and pancreas and liver (right). Compare pathways with Fig 5C and Supplementary Fig 6.

Supplementary Fig 6. IPA analysis using protein and mRNA co-expression data associated with U2AF1, U2AF2, SF1, PRPF40A, PRPF40B and TCERG1 in bowel cancers. The full length of the heat map (same heat map as in Fig 5C) is shown. Lower portion of IPA heat map (marked with *) indicates upregulated pathways with lower z-scores. Left 5 lanes are derived from gene lists of proteins co-expressed with SF1, U2AF2, U2AF1, PRPF40A and TCERG1. Right 6 lanes are derived from mRNA gene lists co-expressed with SF1, PRPF40A, TCERG1, PRPF40B, U2AF1 and U2AF2.

Supplementary Table 1. Co-occurrence tendency of two Commitment Factor (CC) gene mutations in tumors. A q value of < 0.05 indicates significant co-occurrence in tumor tissue of mutated genes listed under column A and B.

Supplementary Table 2. Incidence of U2AF1, U2AF2, SF1, PRPF40A, PRPF40B and TCERG1 alterations in different cancers listed in cBioPortal. Genetic alterations are from all non-redundant studies of each type of cancer. Genetic plus expression level alteration frequency is derived from TCGA, PanCancer Atlas study of each type of cancer.

Supplementary Table 3. Co-occurrence tendency of two Commitment Factor (CC) gene mutations in tumors from patients in the *APC*, *KRAS*, *TP53* or *BRAF* cohorts. A *q* value of < 0.05 indicates significant co-occurrence (in bold) in tumor tissue of mutated genes listed under column A and B.

Supplementary Methods

(1) Survival curves with equal number of patients in altered and unaltered groups.

Survival data was downloaded from cBioPortal by clicking the download icon and Data tab on the right side of the survival curves, and lists of case ID for both the altered (132 cases) and unaltered group (1219 cases) were obtained. Sample ID was downloaded separately from the Download tab of the query. Subsequently, ten sub-lists from the sample ID list of the unaltered group were generated using the randomize function in Excel, each containing 132 sample IDs. The 6 E-complex genes were then queried specifically in patient sets defined by these 132 altered and 132 randomly generated unaltered sample IDs, generating overall and progression free survivals of patients with or without any genetic alterations in E-complex genes. *P* and *q*-values were computed for each of these survival curves.

(2) TCGA PanCancer Atlas Study; Bowel Cancer

cBioPortal also reports changes in mRNA and protein levels in human cancer tissues in specific TCGA PanCancer Atlas studies. Thus, for each type of GI cancer, we separately examined their cognate TCGA PanCancer Atlas study. Each TCGA PanCancer Atlas study was queried for CC factor alteration in genetic (mutation/structural variant/putative copy-number alterations from GISTIC), expression level (mRNA expression z-scores relative to diploid samples (RNA Seq V2 RSEM)) and protein/phosphoprotein level (Protein expression z-scores (mass spectrometry by CPTAC)) changes.

(3) Generation of driver gene APC, TRP53, KRAS or BRAF cohorts.

cBioPortal database was used to investigate how genetic changes in CC factors associate with specific driver genes in colorectal cancer patients. Four different cohorts of bowel cancer patients, each having genetic alterations in the driver genes *APC, TRP53, KRAS* or *BRAF* were isolated and examined for CC alterations.

To generate the APC cohort, all non-redundant bowel cancer studies were selected from cBioPortal [leaving out the 2 overlapping studies on Colorectal Adenocarcinoma (TCGA, Firehose Legacy and TCGA, Nature 2012 studies)] and queried for genetic alterations in *APC*. 68% of bowel cancer patients (or 4422 patients) had genetic changes in their *APC* gene (Table 2). Of these, a cohort of 4393 patients were selected that only carried *APC* gene mutations (excluding gene amplifications or deletions). The 4393 patient samples were queried for genetic alterations in the six CC factors. CC factors were altered in 4% (177) of the patients carrying *APC* mutations.

Similarly, we selected for a cohort of patients with *KRAS* mutations. 43% (2796 patients) of bowel cancer patients had *KRAS* mutation (Table 1B). Of this, a *KRAS* cohort of 2737 patients, 4% (106 patients) had mutations in CC factors. 66% (4302 patients) with bowel cancers had *TP53* mutations. Of these 3% (123 patients) also had CC factor mutations (Table 1B).

BRAF gene was altered in a lower proportion, 10% (670 patients), of bowel cancer patients but of these, a higher fraction, 19% (127 patients), carried CC mutations (Table 1B). We also queried the TCGA PanCancer Atlas dataset (colorectal adenocarcinoma) for *BRAF* alterations (Table 1B and Figure 4*E*). In this study of 594 patients, *BRAF* was altered (with mutated *BRAF* plus expression level changes) in 19% (112 patients) of the patients. Of these 112 patients, 62 patients carried only *BRAF* gene mutations and constituted the *BRAF*/TCGA cohort (Table 1B). 61% (38 patients) of the *BRAF*/TCGA cohort, also had alterations in CC factors, with MACR patients having the highest alteration rates in CC genes (Figure 4*E*).

***(4) Genes co-expressed with CC factors and IPA*.**

The bowel TCGA PanCancer Atlas database in cBioPortal was queried for RNAs and proteins whose expression levels correlate with each CC factor. For example, to identify genes in bowel cancers that are significantly co-expressed with *SF1,* the TCGA PanCancer Atlas dataset

(Colorectal adenocarcinoma) was queried for SF1 followed by co-expression data. The RNA co-expression data in cBioPortal is derived from RNA sequencing from 592 patient samples. A list of 19880 genes whose RNA levels positively or negatively correlated with *SF1* was obtained, out of which the expression correlation of 9576 genes were significant (q-value < 0.05). This list of 9576 genes was designated co-expressed with *SF1*. Similar co-expression gene lists (q-value < 0.05) were obtained for each CC factor: U2AF2, U2AF1, PRPF40A, PRPF40B and TCERG1.

The Colorectal adenocarcinoma (TCGA PanCancer Atlas) dataset also allowed for the query for proteins that are co-expressed with individual CC factors (example of NUDT21 co-expression with SF1, q = 6.53e-21, Figure 5*B*). The protein co-expression data is derived from CPTAC (NCI's Clinical Proteomic Tumor Analysis Consortium) mass spectrometry of 84 samples. Mass spectroscopy CPTAC sample data is available for SF1, U2AF1, U2AF2, PRPF40A and TCERG1 but not for PRPF40B. We generated a list of 5520 proteins that co-expressed with SF1, of which 3413 proteins were significantly co-expressed (q-value < 0.05). Similar queries generated gene lists for protein co-expression for each of the other CC factors.

We utilized the 5 lists of proteins co-expressed with each of the CC factors (SF1, U2AF1, U2AF2, PRPF40A and TCERG1) together with the 6 lists of mRNAs co-expressed with each of the CC factors (SF1, U2AF1, U2AF2, PRPF40A, PRPF40B and TCERG1) to simultaneously perform QIAGEN Ingenuity Pathway Analysis (IPA).