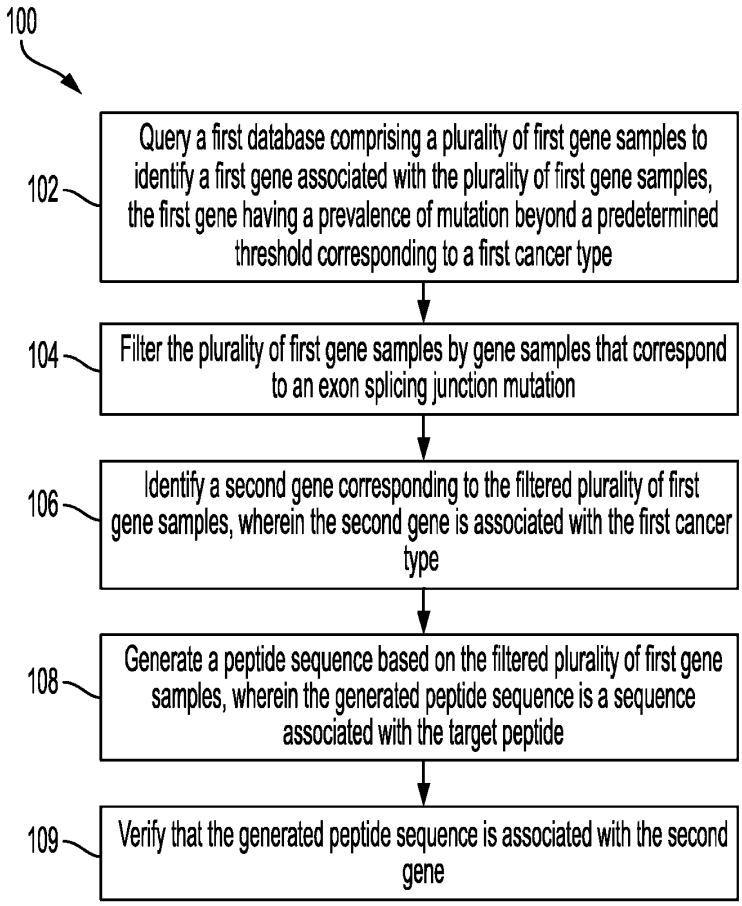

100
Query a first database comprising a plurality of first gene samples to identify a first gene associated with the plurality of first gene samples, the first gene having a prevalence of mutation beyond a predetermined threshold corresponding to a first cancer type
102
Filter the plurality of first gene samples by gene samples that correspond to an exon splicing junction mutation
104
Identify a second gene corresponding to the filtered plurality of first gene samples, wherein the second gene is associated with the first cancer type
106
Generate a peptide sequence based on the filtered plurality of first gene samples, wherein the generated peptide sequence is a sequence associated with the target peptide
108
Verify that the generated peptide sequence is associated with the second gene
109

FIG. 1

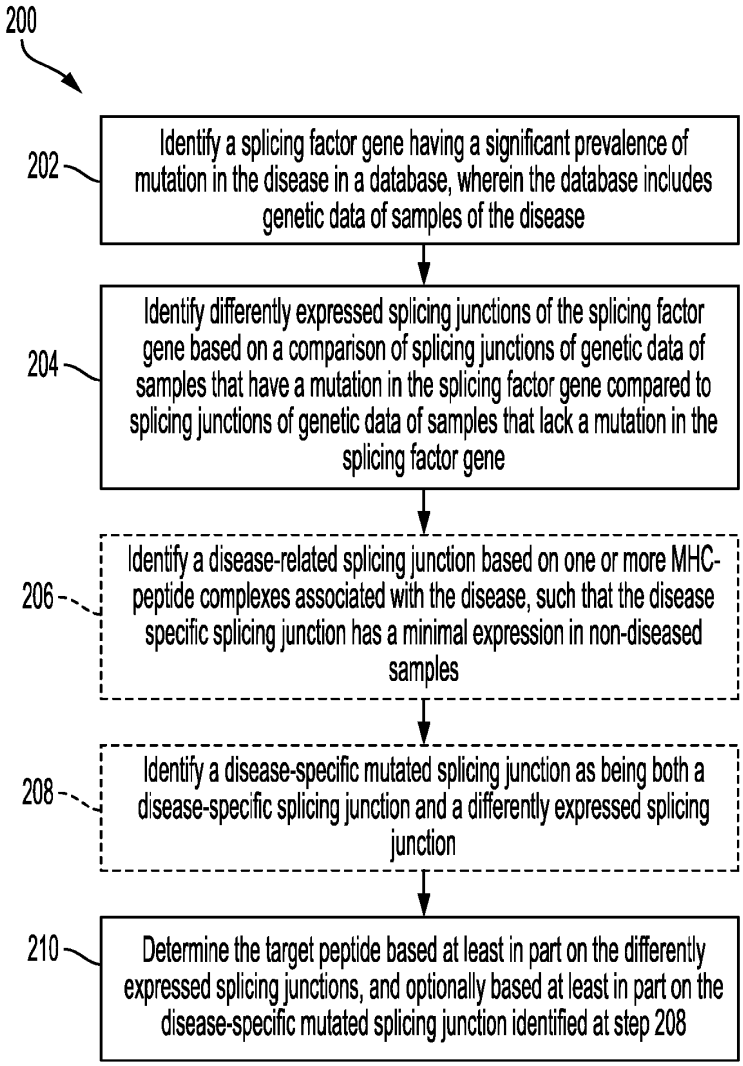

200
202
Identify a splicing factor gene having a significant prevalence of mutation in the disease in a database, wherein the database includes genetic data of samples of the disease
204
Identify differently expressed splicing junctions of the splicing factor gene based on a comparison of splicing junctions of genetic data of samples that have a mutation in the splicing factor gene compared to splicing junctions of genetic data of samples that lack a mutation in the splicing factor gene
206
Identify a disease-related splicing junction based on one or more MHC-peptide complexes associated with the disease, such that the disease specific splicing junction has a minimal expression in non-diseased samples
208
Identify a disease-specific mutated splicing junction as being both a disease-specific splicing junction and a differently expressed splicing junction
210
Determine the target peptide based at least in part on the differently expressed splicing junctions, and optionally based at least in part on the disease-specific mutated splicing junction identified at step 208

FIG. 2

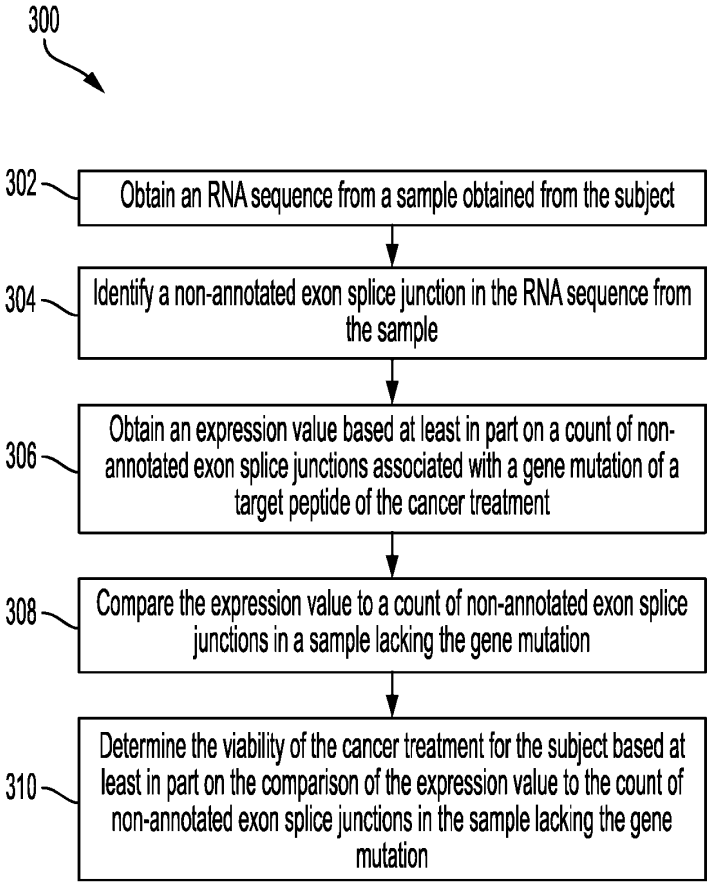

300
302
Obtain an RNA sequence from a sample obtained from the subject
304
Identify a non-annotated exon splice junction in the RNA sequence from the sample
306
Obtain an expression value based at least in part on a count of non-annotated exon splice junctions associated with a gene mutation of a target peptide of the cancer treatment
308
Compare the expression value to a count of non-annotated exon splice junctions in a sample lacking the gene mutation
310
Determine the viability of the cancer treatment for the subject based at least in part on the comparison of the expression value to the count of non-annotated exon splice junctions in the sample lacking the gene mutation

FIG. 3

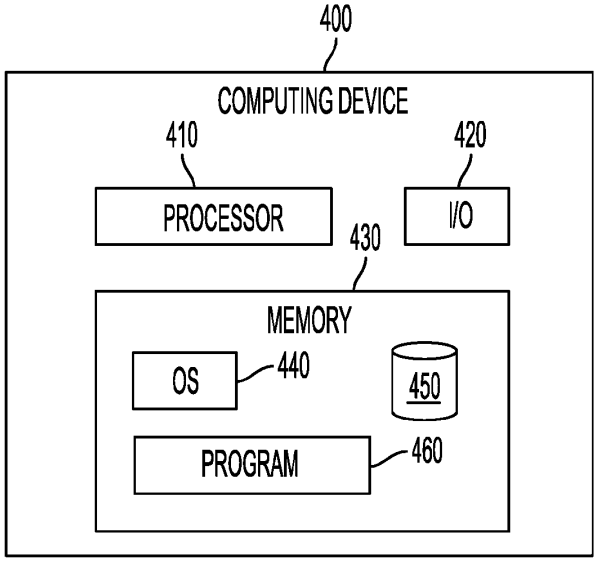

400
COMPUTING DEVICE
410
420
PROCESSOR
I/O
430
MEMORY
OS
440
450
PROGRAM
460

FIG. 4

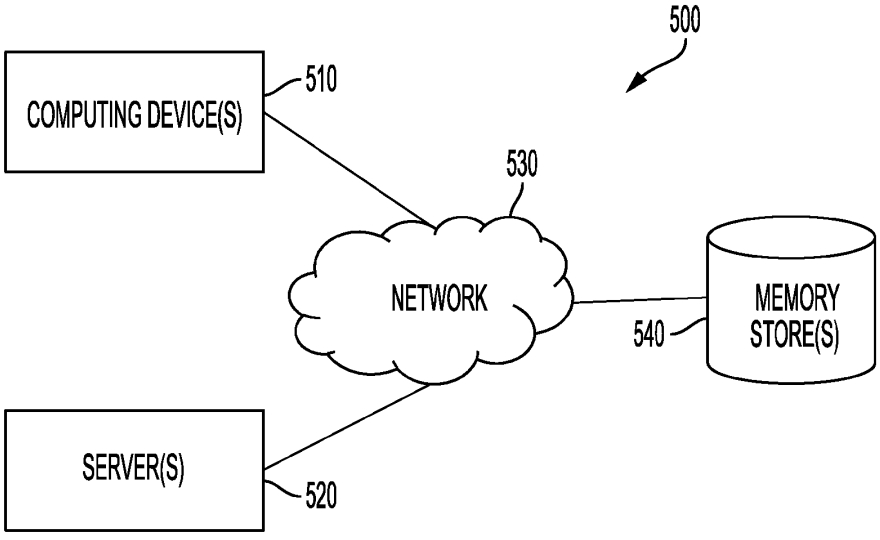

500
COMPUTING DEVICE(S)
510
530
NETWORK
MEMORY
STORE(S)
540
SERVER(S)
520

FIG. 5

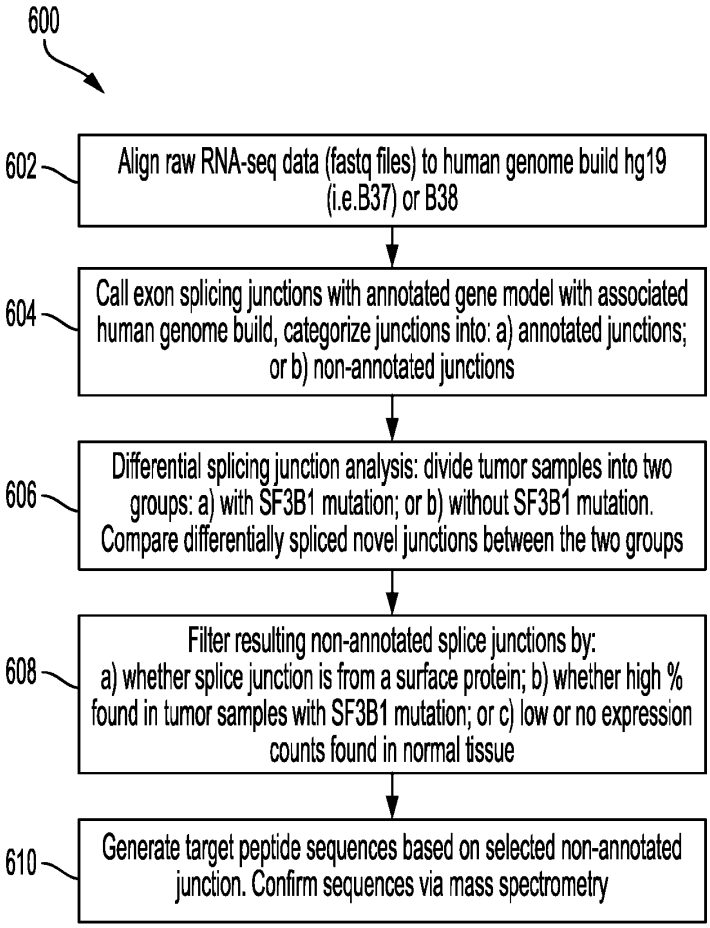

600
602
Align raw RNA-seq data (fastq files) to human genome build hg19 (i.e.B37) or B38
604
Call exon splicing junctions with annotated gene model with associated human genome build, categorize junctions into: a) annotated junctions; or b) non-annotated junctions
606
Differential splicing junction analysis: divide tumor samples into two groups: a) with SF3B1 mutation; or b) without SF3B1 mutation. Compare differentially spliced novel junctions between the two groups
608
Filter resulting non-annotated splice junctions by: a) whether splice junction is from a surface protein; b) whether high % found in tumor samples with SF3B1 mutation; or c) low or no expression counts found in normal tissue
610
Generate target peptide sequences based on selected non-annotated junction. Confirm sequences via mass spectrometry

FIG. 6

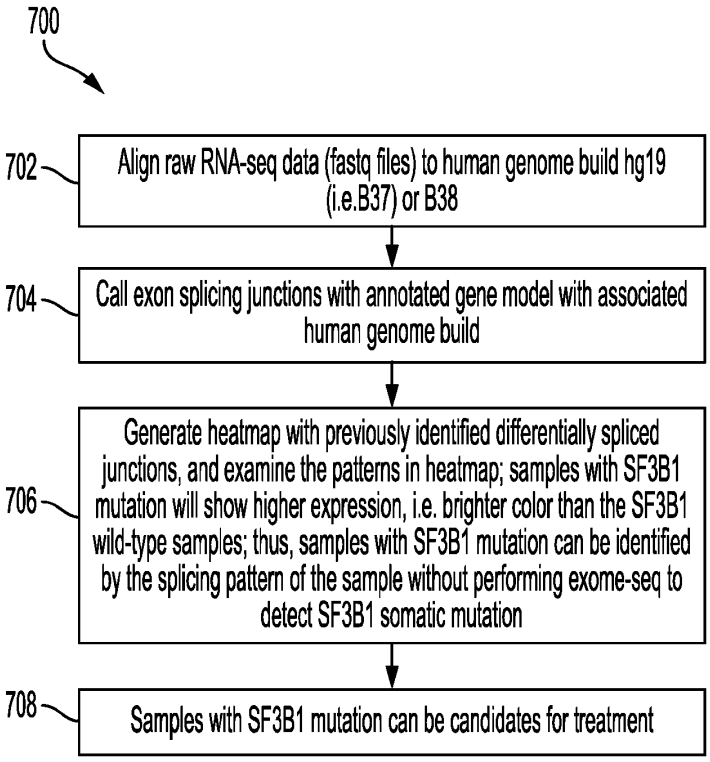

700
702
Align raw RNA-seq data (fastq files) to human genome build hg19 (i.e.B37) or B38
704
Call exon splicing junctions with annotated gene model with associated human genome build
706
Generate heatmap with previously identified differentially spliced junctions, and examine the patterns in heatmap; samples with SF3B1 mutation will show higher expression, i.e. brighter color than the SF3B1 wild-type samples; thus, samples with SF3B1 mutation can be identified by the splicing pattern of the sample without performing exome-seq to detect SF3B1 somatic mutation
708
Samples with SF3B1 mutation can be candidates for treatment

FIG. 7

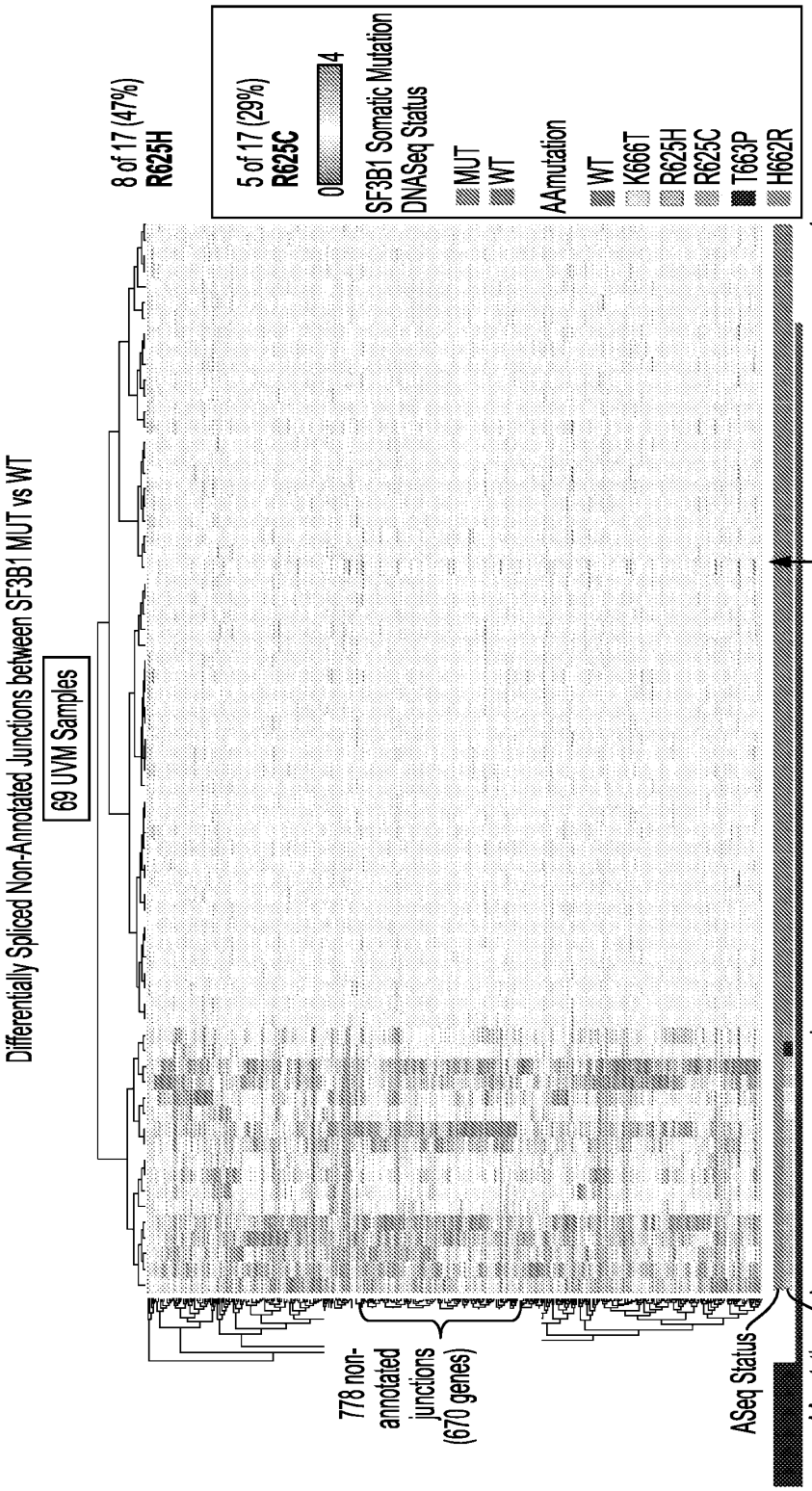

Differentially Spliced Non-Annotated Junctions between SF3B1 MUT vs WT
69 UVM Samples
8 of 17 (47%)
R625H
5 of 17 (29%)
R625C
0
4
SF3B1 Somatic Mutation
DNASeq Status
MUT
WT
AAmutation
WT
K666T
R625H
R625C
T663P
H662R
778 non-annotated junctions (670 genes)
ASeq Status
AAmutation
MUT
WT
LF MUT

FIG. 8

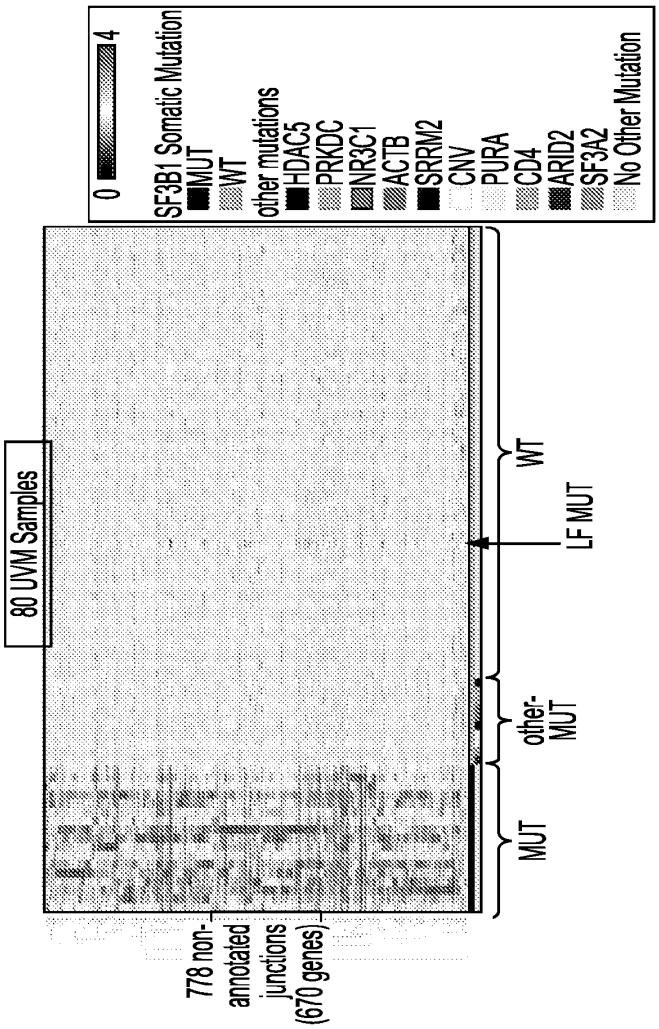

0
4
SF3B1 Somatic Mutation
MUT
WT
other mutations
HDAC5
PRKDC
NR3C1
ACTB
SRRM2
CNV
PURA
CD4
ARID2
SF3A2
No Other Mutation
80 UVM Samples
778 non-annotated junctions (670 genes)
MUT
other-MUT
LF MUT
WT

FIG. 9

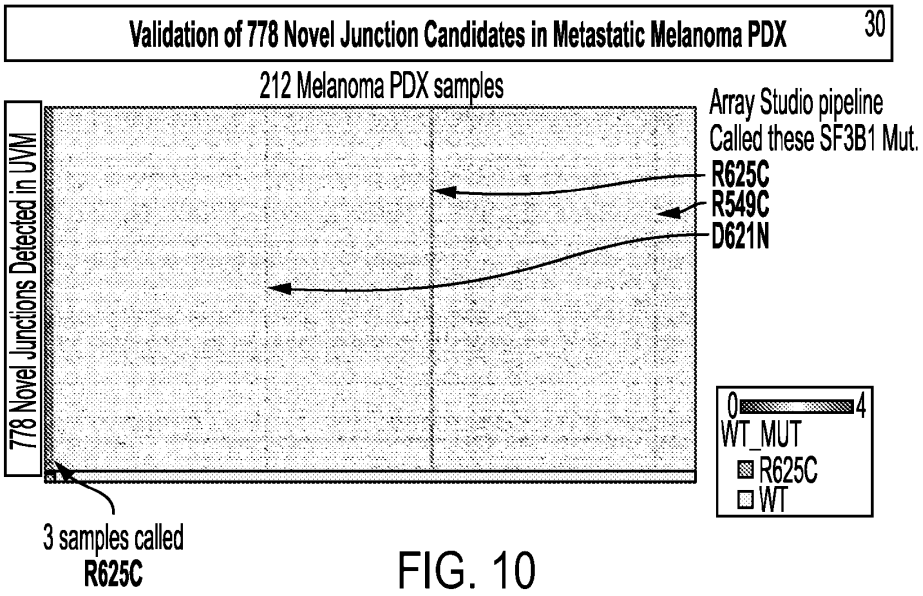

Validation of 778 Novel Junction Candidates in Metastatic Melanoma PDX
30
212 Melanoma PDX samples
778 Novel Junctions Detected in UVM
Array Studio pipeline
Called these SF3B1 Mut.
R625C
R549C
D621N
0
4
WT_MUT
R625C
WT
3 samples called
R625C

FIG. 10

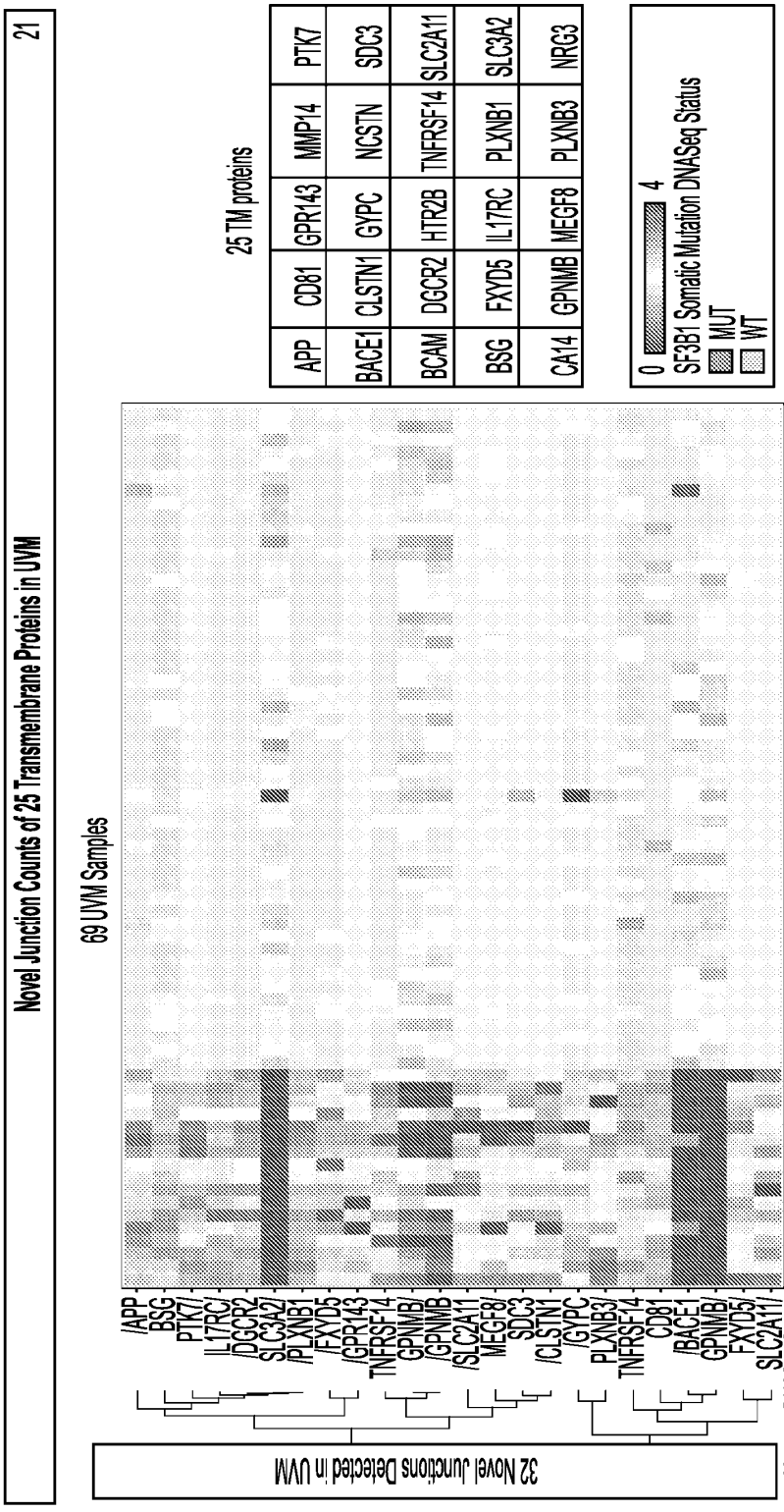

Novel Junction Counts of 25 Transmembrane Proteins in UVM
21
69 UVM Samples
25 TM proteins
APP CD81 GPR143 MMP14 PTK7
BACE1 CLSTN1 GYPC NCSTN SDC3
BCAM DGCR2 HTR2B TNFRSF14 SLC2A11
BSG FXYD5 IL17RC PLXNB1 SLC3A2
CA14 GPNMB MEGF8 PLXNB3 NRG3
0 4
SF3B1 Somatic Mutation DNASeq Status
MUT
WT
32 Novel Junctions Detected in UVM
/APP
BSG
PTK7
IL17RC
/DGCR2
SLC3A2
/PLXNB1
/FXYD5
/GPR143
TNFRSF14
GPNMB
/GPNMB
/SLC2A11
MEGF8
SDC3
/CLSTN1
/GYPC
PLXNB3
TNFRSF14
CD81
/BACE1
GPNMB
FXYD5
SLC2A11
Mutation DNASeq Status

FIG. 11

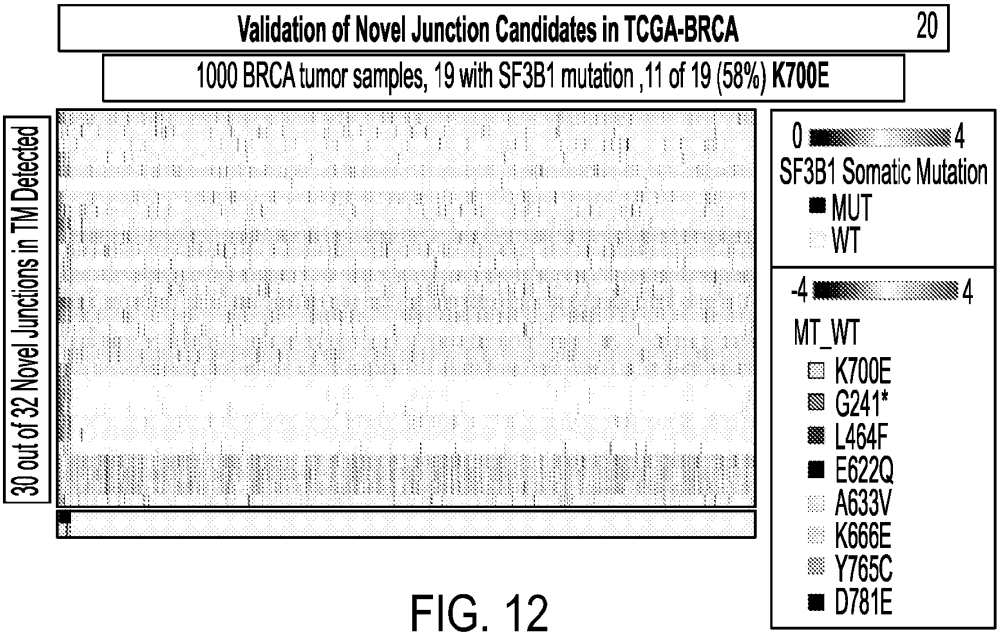

Validation of Novel Junction Candidates in TCGA-BRCA
20
1000 BRCA tumor samples, 19 with SF3B1 mutation ,11 of 19 (58%) K700E
30 out of 32 Novel Junctions in TM Detected
0
4
SF3B1 Somatic Mutation
MUT
WT
-4
4
MT_WT
K700E
G241*
L464F
E622Q
A633V
K666E
Y765C
D781E

FIG. 12

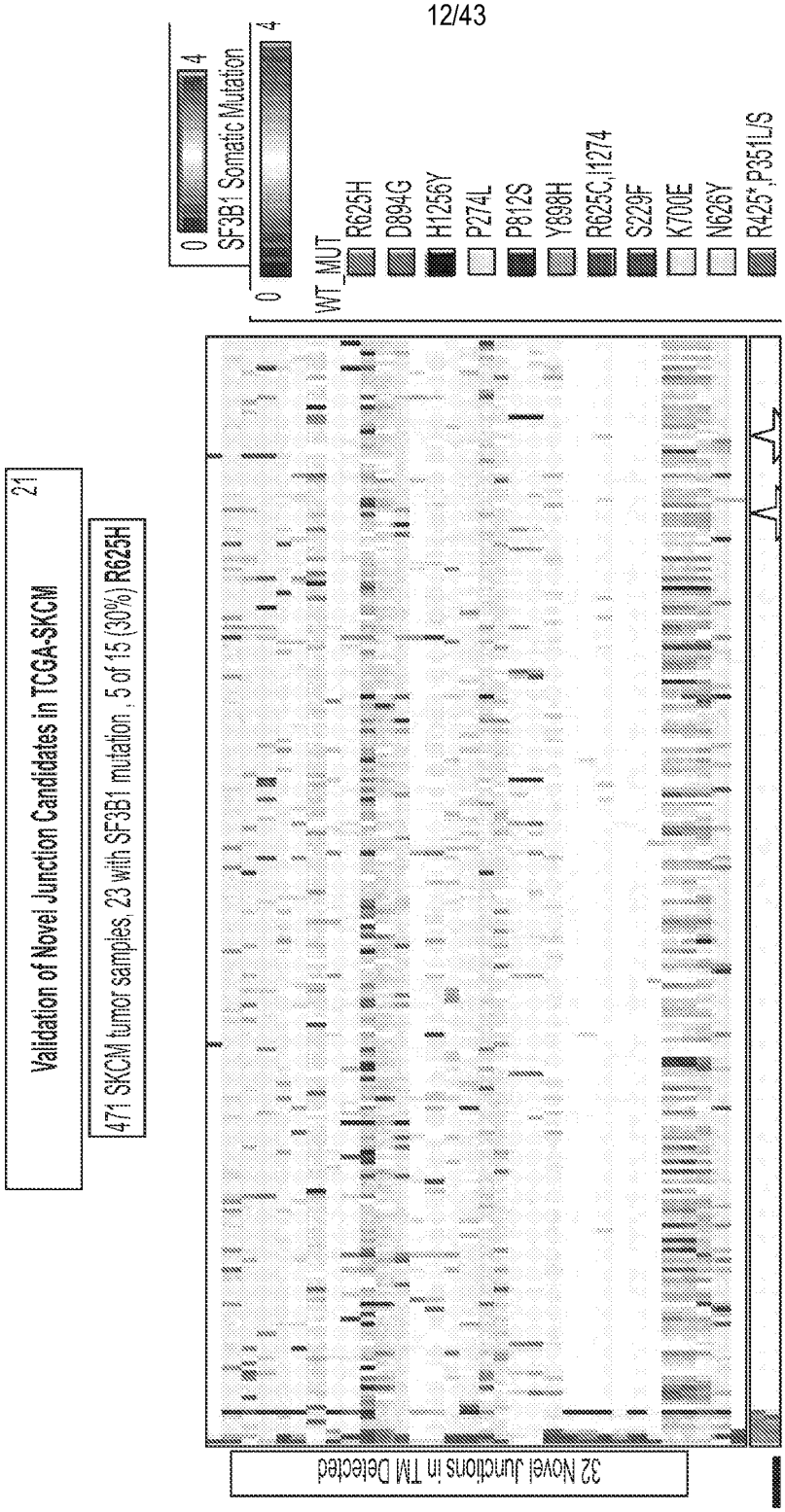

Validation of Novel Junction Candidates in TCGA-SKCM
21
471 SKCM tumor samples, 23 with SF3B1 mutation , 5 of 15 (30%) R625H
32 Novel Junctions in TM Detected
0
4
SF3B1 Somatic Mutation
0
4
WT_MUT
R625H
D894G
H1256Y
P274L
P812S
Y998H
R625C,I1274
S229F
K700E
N626Y
R425*,P351L/S

FIG. 13

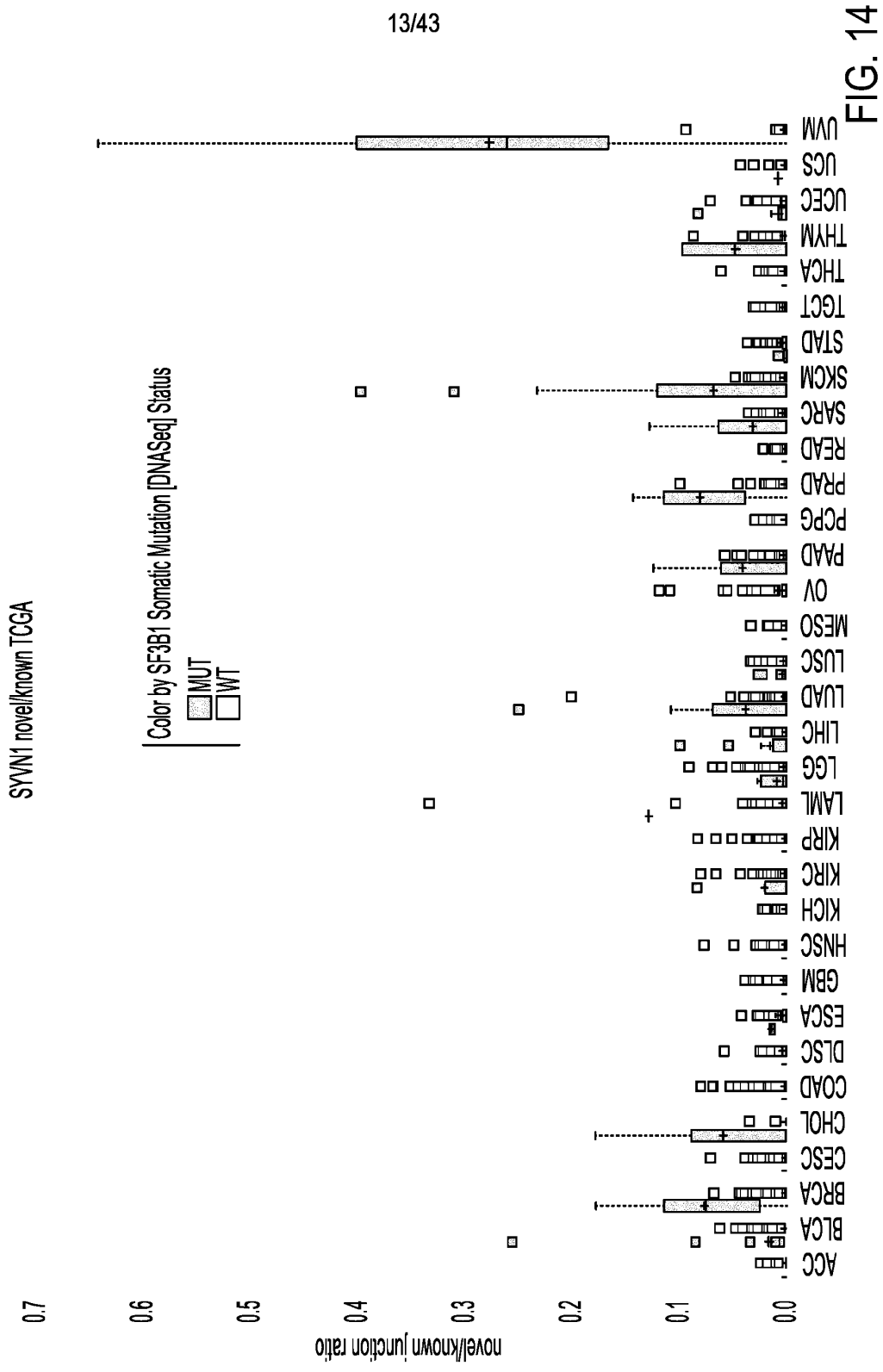

SYVN1 novel/known TCGA
Color by SF3B1 Somatic Mutation [DNASeq] Status
MUT
WT
novel/known junction ratio
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLSC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM

FIG. 14

	SF3B1 Somatic Mutation Status			
Tumor type	missing	MUT	WT	Total
ACC	1	2	76	79
BLCA	20	24	367	411
BRCA	114	19	981	1114
CESC	111	1	194	306
CHOL	1	3	32	36
COAD	102	16	355	473
DLBC		2	46	48
ESCA	1	2	182	185
GBM	6	4	176	186
HNSC	2	6	514	522
KICH			66	66
KIRC	32	5	501	538
KIRP	9	7	275	291
LAML	3	1	175	179
LGG	1	6	527	534
LIHC	7	11	356	374
LUAD	14	16	500	530
LUSC	18	13	470	501
MESO	4	1	82	87
OV	76		354	430
PAAD	1	3	175	179
PCPG			184	184
PRAD	3	4	495	502
READ	18	2	147	167
SARC	11	4	248	263
SKCM	1	23	448	472
STAD	44	9	363	416
TGCT	1		155	156
THCA	11	2	500	513
THYM	1	2	117	120
UCEC	303	15	233	551
UCS		1	56	57
UVM		18	62	80
Grand Total	916	222	9412	10550

FIG. 15A

mean novel junction ratio in MUT	mean novel junction ratio in WT	Fold(MUT/WT)
0.0000	0.0008	0.0
0.0164	0.0022	7.4
0.0770	0.0020	38.5
0.0000	0.0016	0.0
0.0595	0.0014	41.7
0.0000	0.0023	0.0
0.0000	0.0038	0.0
0.0142	0.0030	4.7
0.0000	0.0032	0.0
0.0000	0.0017	0.0
0.0208	0.0014	14.5
0.0000	0.0017	0.0
	0.0018	0.0
0.1286	0.0044	29.5
0.0098	0.0032	3.1
0.0163	0.0008	20.0
0.0387	0.0022	17.6
0.0043	0.0027	1.6
0.0000	0.0014	0.0
	0.0041	0.0
0.0412	0.0029	14.3
	0.0015	0.0
0.0770	0.0015	50.9
0.0000	0.0007	0.0
0.0319	0.0022	14.5
0.0686	0.0018	37.5
0.0015	0.0027	0.5
	0.0016	0.0
0.0000	0.0021	0.0
0.0484	0.0024	20.2
0.0082	0.0014	5.9
0.0081	0.0024	3.3
0.2770	0.0023	122.5

FIG. 15A
CONTINUED

CancerType	inSF3B1_MUT	inSF3B1_WT	Grand Total	% in all MUT	% in all WT
BLCA	2	1	3	8.33	0.27
BRCA	12	1	13	63.16	0.10
CESC		1	1	0.00	0.52
CHOL	1		1	33.33	0.00
COAD		3	3	0.00	0.85
DLBC		1	1	0.00	2.17
HNSC		2	2	0.00	0.39
KIRC	1	2	3	20.00	0.40
KIRP		3	3	0.00	1.09
LAML	1	2	3	100.00	1.14
LGG		3	3	0.00	0.57
LIHC	2		2	18.18	0.00
LUAD	5	2	7	31.25	0.40
OV		4	4	0.00	1.13
PAAD	1	1	2	33.33	0.57
PRAD	3	1	4	75.00	0.20
SARC	1		1	25.00	0.00
SKCM	7		7	30.43	0.00
THCA		1	1	0.00	0.20
THYM	1	1	2	50.00	0.85
UCEC	1	1	2	6.67	0.43
UVM	15	1	16	83.33	1.61
Grand Total	53	31	84		

FIG. 15B

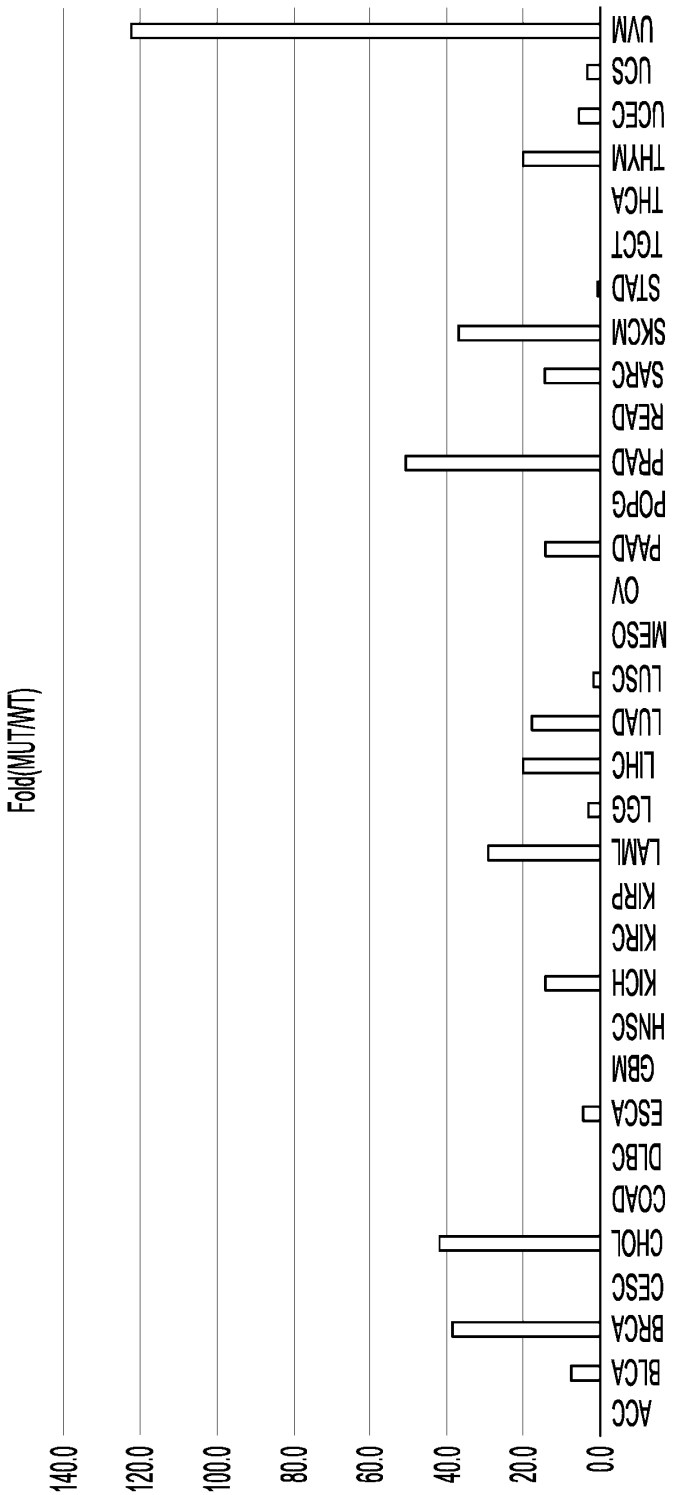

Fold(MUT/WT)
140.0
120.0
100.0
80.0
60.0
40.0
20.0
0.0
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
POPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM

FIG. 15C

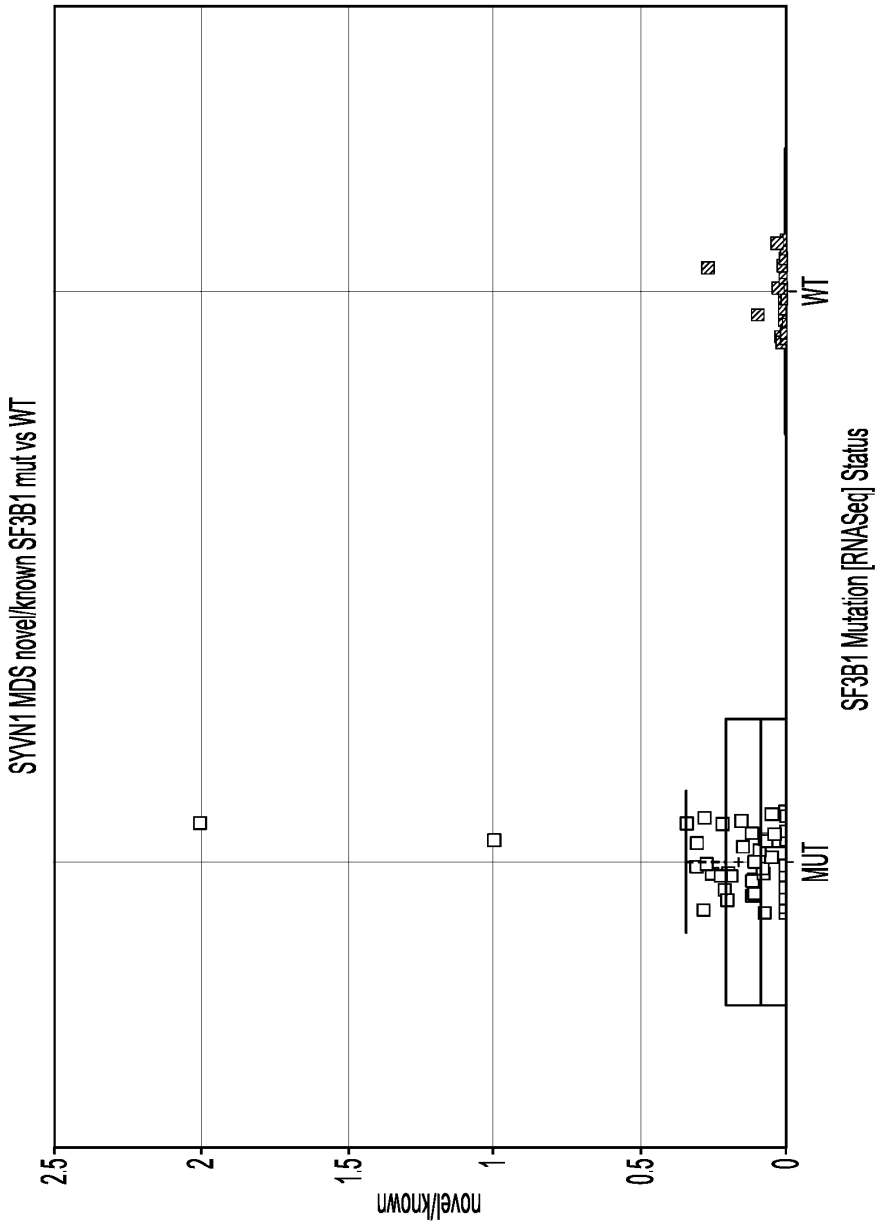

SYVN1 MDS novel/known SF3B1 mut vs WT
novel/known
2.5
2
1.5
1
0.5
0
WT
MUT
SF3B1 Mutation [RNASeq] Status

FIG. 16

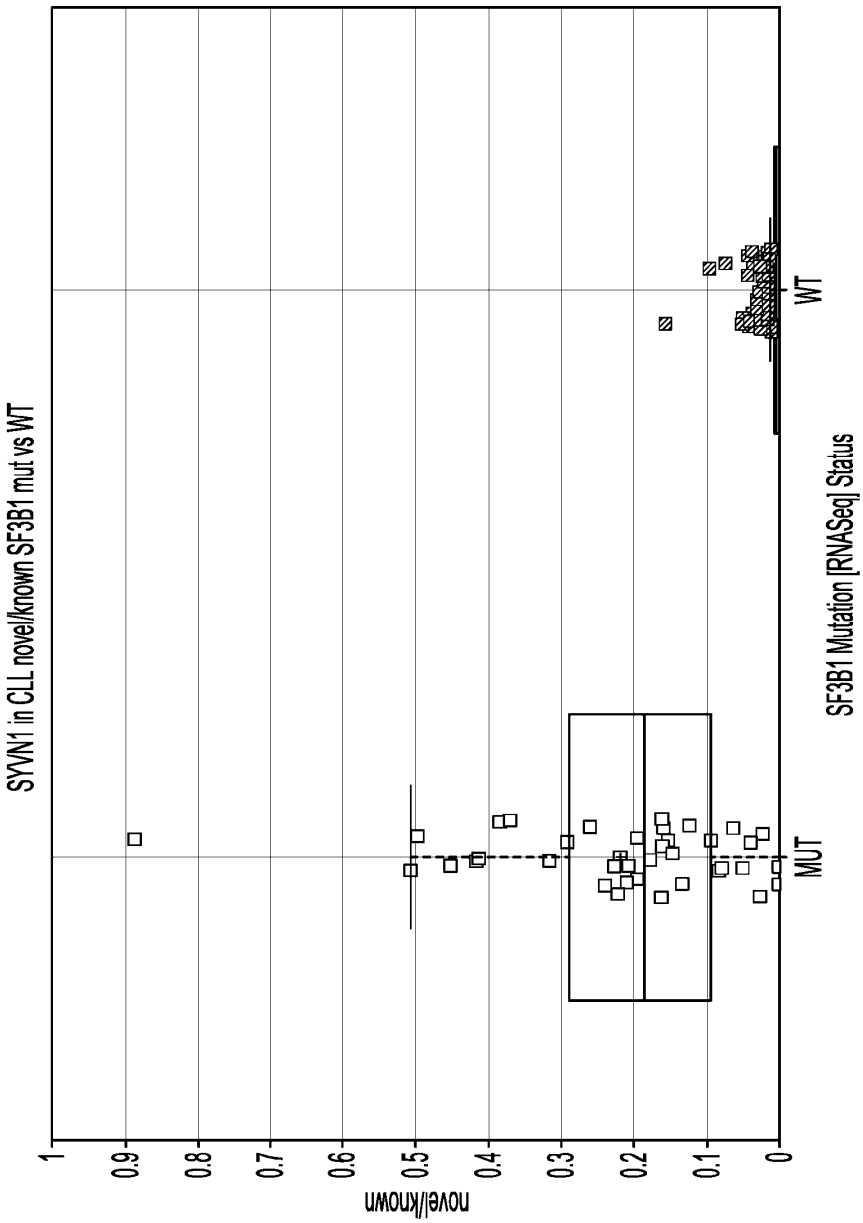

SYVN1 in CLL novel/known SF3B1 mut vs WT
novel/known
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
MUT
WT
SF3B1 Mutation [RNASeq] Status

FIG. 17

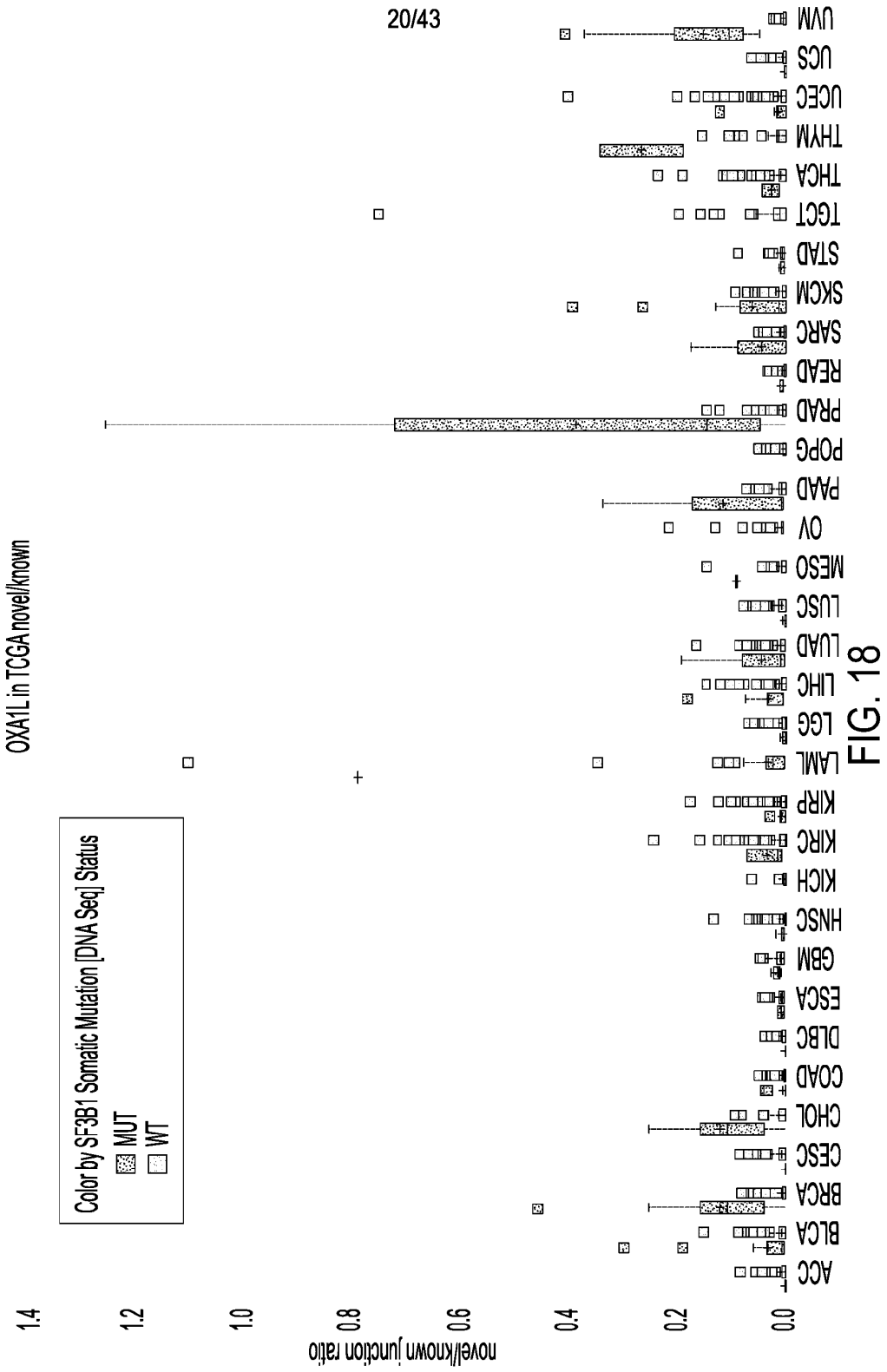

OXA1L in TCGA novel/known
Color by SF3B1 Somatic Mutation [DNA Seq] Status
MUT
WT
novel/known junction ratio
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
POPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM

FIG. 18

	SF3B1 Somatic Mutation Status						
Tumor type	missing	MUT	WT	Total	mean novel juction ratio in MUT	mean novel juction ratio in WT	Fold(MUT/WT)
ACC	1	2	76	79	0.0000	0.0060	0.0
BLCA	20	24	367	411	0.0326	0.0072	4.5
BRCA	114	19	981	1114	0.1201	0.0034	35.1
CESC	111	1	194	306	0.0000	0.0082	0.0
CHOL	1	3	32	36	0.1084	0.0119	9.1
COAD	102	16	355	473	0.0041	0.0023	1.8
DLBC		2	46	48	0.0000	0.0044	0.0
ESCA	1	2	182	185	0.0075	0.0085	0.9
GBM	6	4	176	186	0.0149	0.0097	1.5
HNSC	2	6	514	522	0.0045	0.0037	1.2
KICH			66	66	0.0000	0.0024	0.0
KIRC	32	5	501	538	0.0338	0.0087	3.9
KIRP	9	7	275	291	0.0066	0.0072	0.9
LAML	3	1	175	179	0.7891	0.0320	24.6
LGG	1	6	527	534	0.0024	0.0046	0.5
LIHC	7	11	356	374	0.0295	0.0066	4.5
LUAD	14	16	500	530	0.0414	0.0069	6.0
LUSC	18	13	470	501	0.0023	0.0076	0.3
MESO	4	1	82	87	0.0892	0.0075	12.0
OV	76		354	430	0.0000	0.0069	0.0
PAAD	1	3	175	179	0.1145	0.0082	13.9
PCPG			184	184	0.0000	0.0030	0.0
PRAD	3	4	495	502	0.3850	0.0045	86.4
READ	18	2	147	167	0.0049	0.0025	2.0
SARC	11	4	248	263	0.0432	0.0039	11.2
SKCM	1	23	448	472	0.0598	0.0050	12.1
STAD	44	9	363	416	0.0040	0.0062	0.6
TGCT	1		155	156	0.0000	0.0213	0.0
THCA	11	2	500	513	0.0256	0.0092	2.8
THYM	1	2	117	120	0.2656	0.0115	23.1
UCEC	303	15	233	551	0.0124	0.0126	1.0
UCS		1	56	57	0.0000	0.0050	0.0
UVM		18	62	80	0.1519	0.0025	61.6
Grand Total	916	222	9412	10550			

FIG. 19A

CancerType	in_SF3B1MUT	in_SF3B1WT	Grand Total	% in all MUT	% in all WT
ACC		2	2	0.00	2.63
BLCA	3	11	14	12.50	3.00
BRCA	14	4	18	73.68	0.41
CESC		7	7	0.00	3.61
CHOL	1	2	3	33.33	6.25
HNSC		3	3	0.00	0.58
KICH		1	1		1.52
KIRC	2	12	14	40.00	2.40
KIRP		9	9	0.00	3.27
LAML	1	23	24	100.00	13.14
LGG		2	2	0.00	0.38
LIHC	3	9	12	27.27	2.53
LUAD	5	10	15	31.25	2.00
LUSC		10	10	0.00	2.13
MESO	1	1	2	100.00	1.22
OV		3	3		0.85
PAAD	1	5	6	33.33	2.86
PRAD	3	5	8	75.00	1.01
SARC	1		1	25.00	0.00
SKCM	9	5	14	39.13	1.12
STAD		1	1	0.00	0.28
TGCT		11	11		7.10
THCA		14	14	0.00	2.80
THYM	2	5	7	100.00	4.27
UCEC	1	13	14	6.67	5.58
UCS		1	1	0.00	1.79
UVM	17		17	94.44	0.00
Grand Total	64	169	233		

FIG. 19B

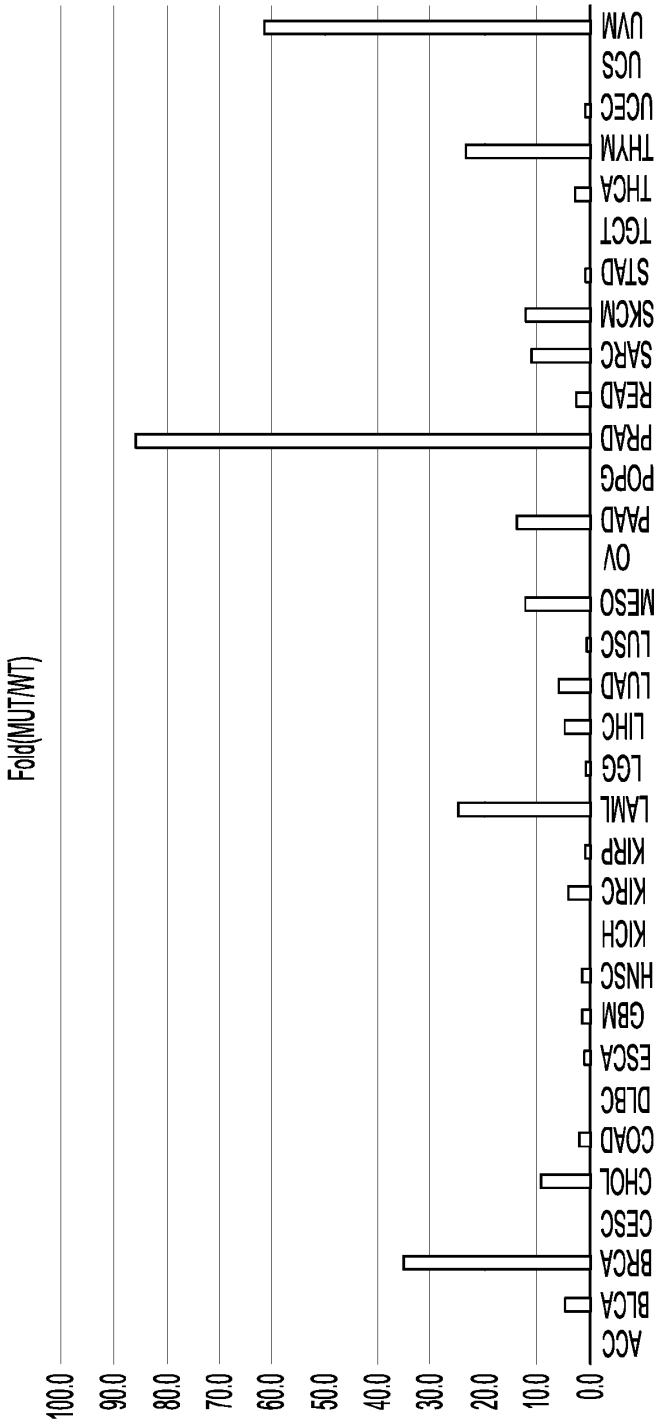

Fold(MUT/WT)
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
POPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM

FIG. 19C

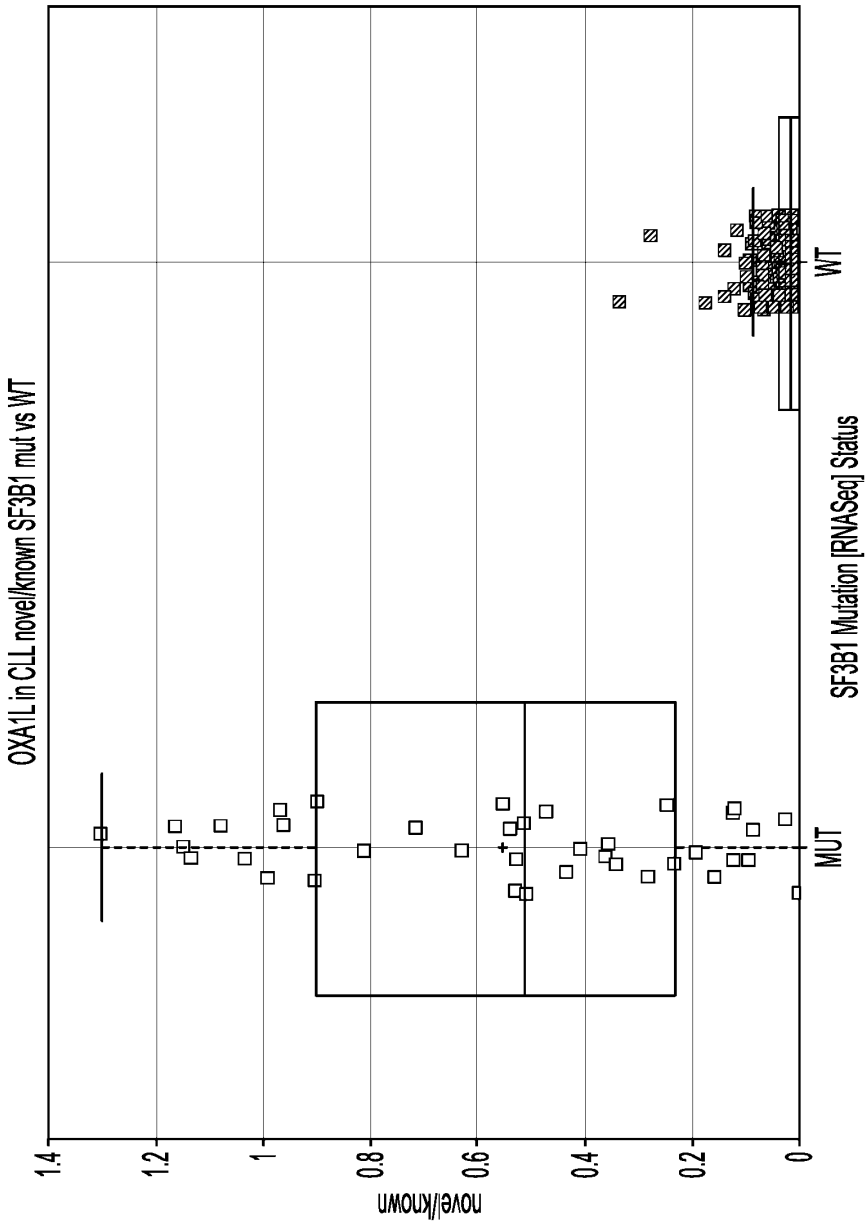

OXA1L in CLL novel/known SF3B1 mut vs WT
novel/known
1.4
1.2
1
0.8
0.6
0.4
0.2
0
MUT
WT
SF3B1 Mutation [RNASeq] Status

FIG. 20

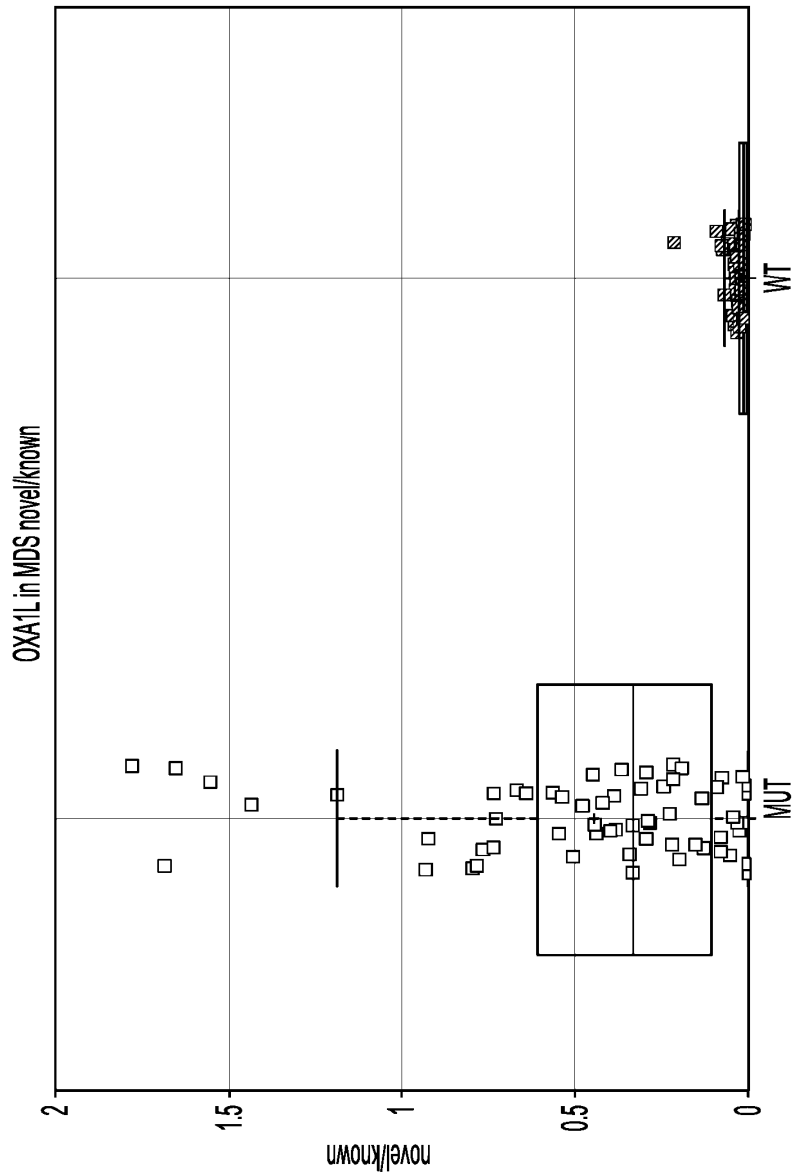

OXA1L in MDS novel/known
novel/known
2
1.5
1
0.5
0
SF3B1 Mutation [RNASeq] Status
WT
MUT

FIG. 21

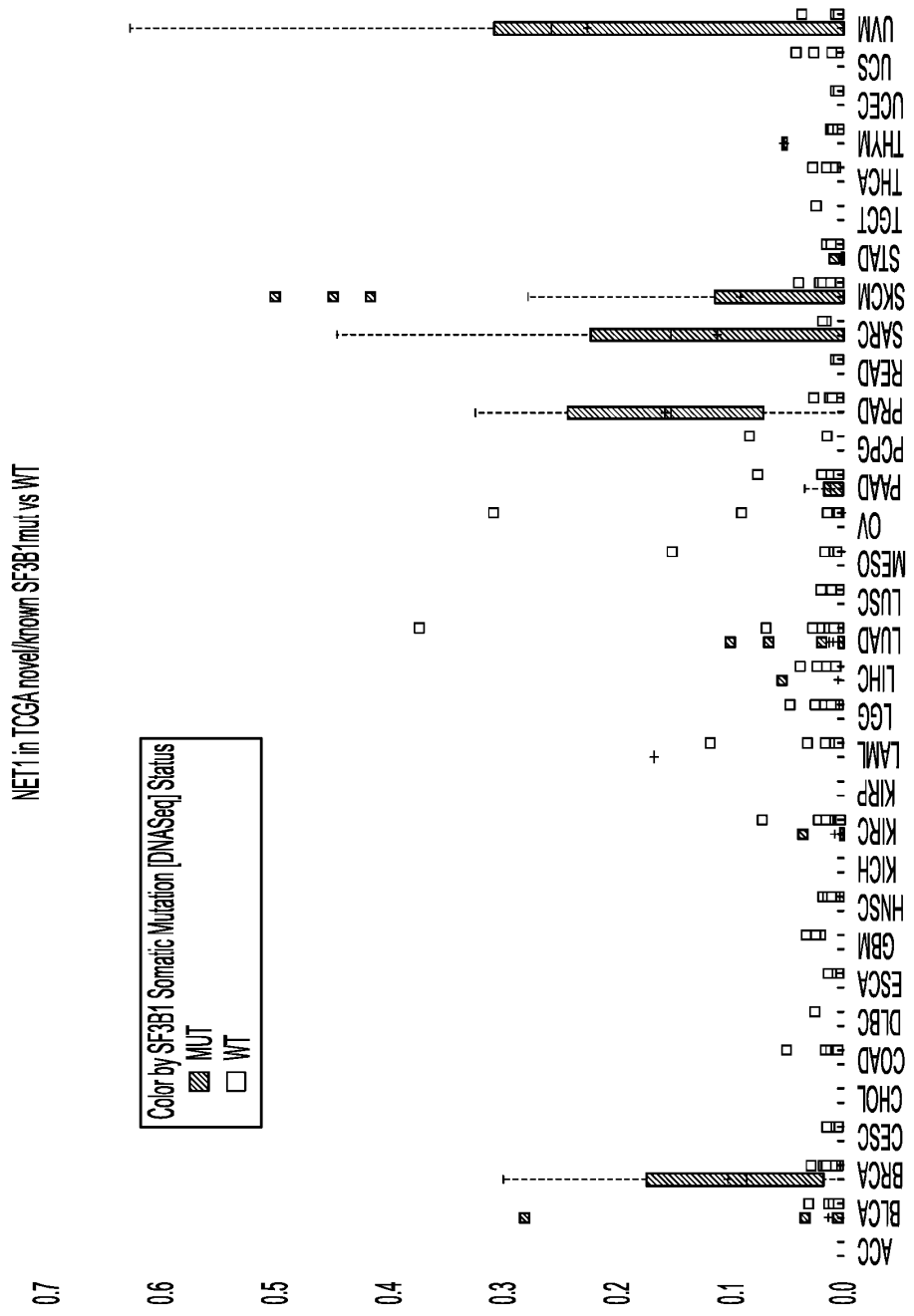
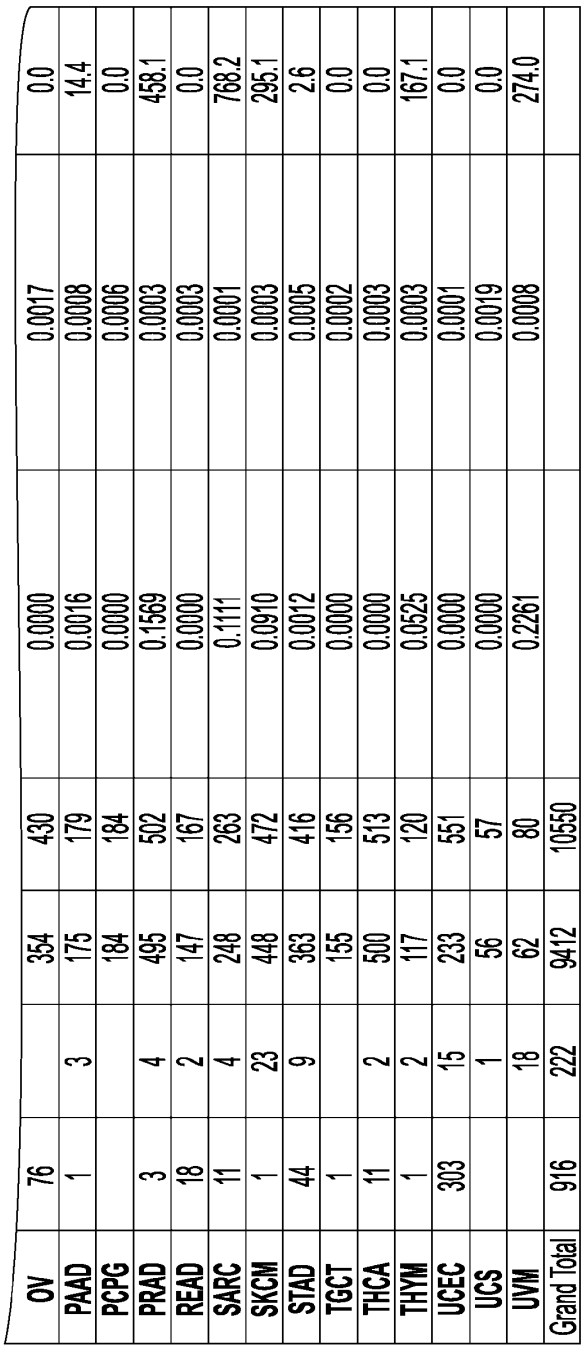

NET1 in TCGA novel/known SF3B1mut vs WT
Color by SF3B1 Somatic Mutation [DNASeq] Status
MUT
WT
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM

FIG. 22

Tumor type	missing	MUT	WT	Total	mean novel junction ratio in MUT	mean novel junction ratio in WT	Fold(MUT/WT)
ACC	1	2	76	79	0.0000	0.0000	0.0
BLCA	20	24	367	411	0.0135	0.0003	43.6
BRCA	114	19	981	1114	0.1025	0.0003	343.5
CESC	111	1	194	306	0.0000	0.0003	0.0
CHOL	1	3	32	36	0.0000	0.0000	0.0
COAD	102	16	355	473	0.0000	0.0004	0.0
DLBC		2	46	48	0.0000	0.0006	0.0
ESCA	1	2	182	185	0.0000	0.0007	0.0
GBM	6	4	176	186	0.0000	0.0004	0.0
HNSC	2	6	514	522	0.0000	0.0003	0.0
KICH			66	66	0.0000	0.0000	0.0
KIRC	32	5	501	538	0.0079	0.0005	15.7
KIRP	9	7	275	291	0.0000	0.0000	0.0
LAML	3	1	175	179	0.1667	0.0010	159.5
LGG	1	6	527	534	0.0000	0.0006	0.0
LIHC	7	11	356	374	0.0049	0.0002	20.1
LUAD	14	16	500	530	0.0123	0.0013	9.5
LUSC	18	13	470	501	0.0000	0.0004	0.0
MESO	4	1	82	87	0.0000	0.0023	0.0

FIG. 23A



OV	76		354	430	0.0000	0.0017	0.0
PAAD	1	3	175	179	0.0016	0.0008	14.4
PCPG			184	184	0.0000	0.0006	0.0
PRAD	3	4	495	502	0.1569	0.0003	458.1
READ	18	2	147	167	0.0000	0.0003	0.0
SARC	11	4	248	263	0.1111	0.0001	768.2
SKCM	1	23	448	472	0.0910	0.0003	295.1
STAD	44	9	363	416	0.0012	0.0005	2.6
TGCT	1		155	156	0.0000	0.0002	0.0
THCA	11	2	500	513	0.0000	0.0003	0.0
THYM	1	2	117	120	0.0525	0.0003	167.1
UCEC	303	15	233	551	0.0000	0.0001	0.0
UCS		1	56	57	0.0000	0.0019	0.0
UVM		18	62	80	0.2261	0.0008	274.0
Grand Total	916	222	9412	10550			

FIG. 23A
CONTINUED

CancerType	in-SF3B1MUT	in_SF3B1WT	Grand Total	% in all MUT	% in all WT
BLCA	1		1	4.17	
BRCA	12		12	63.16	
COAD		1	1		0.28
KIRC		1	1		0.20
LAML	1	1	2	100.00	0.57
LIHC	1		1	9.09	
LUAD	2	2	4	12.50	0.40
MESO		1	1		1.22
OV		2	2		0.56
PAAD		1	1		0.57
PCPG		1	1		0.54
PRAD	3		3	75.00	
SARC	1		1	25.00	
SKCM	7		7	30.43	
THYM	2		2	100.00	
UVM	13		13	72.22	
Grand Total	**43**	**10**	**53**		

FIG. 23B

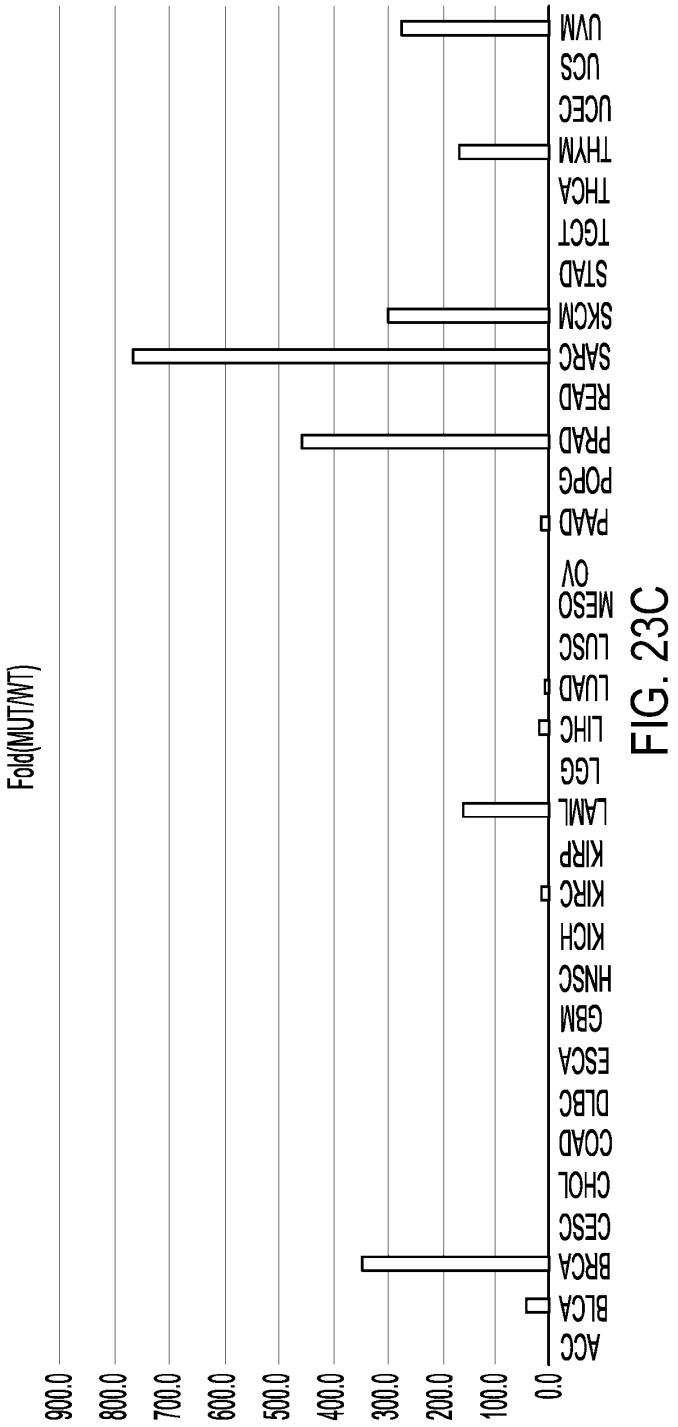

Fold(MUT/WT)
900.0
800.0
700.0
600.0
500.0
400.0
300.0
200.0
100.0
0.0
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM

FIG. 23C

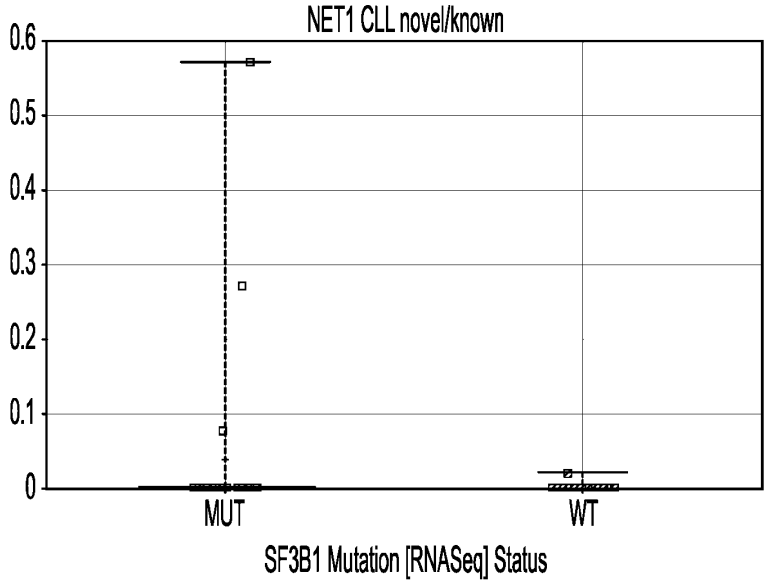

NET1 CLL novel/known
0.6
0.5
0.4
0.3
0.2
0.1
0
MUT
WT
SF3B1 Mutation [RNASeq] Status

FIG. 24

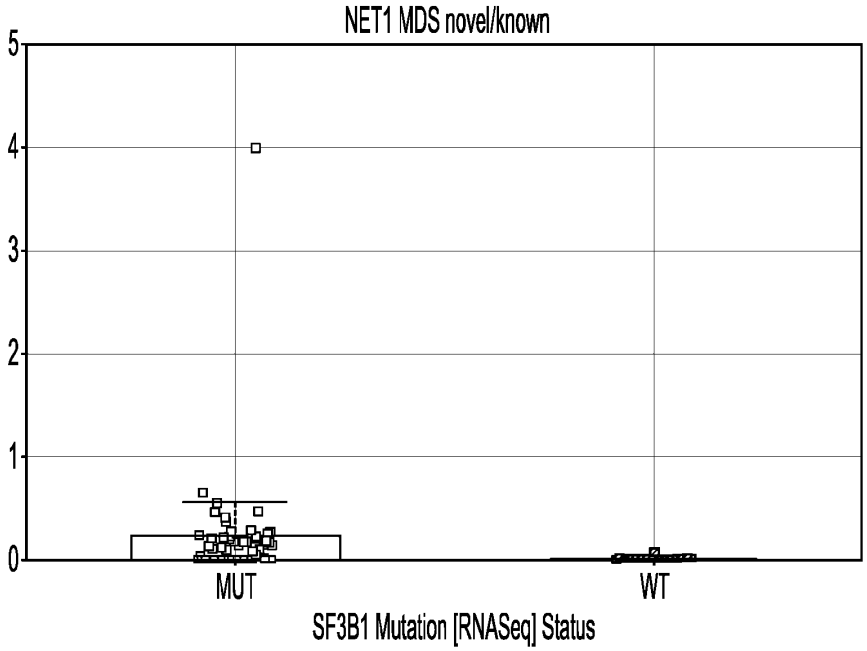

NET1 MDS novel/known
5
4
3
2
1
0
MUT
WT
SF3B1 Mutation [RNASeq] Status

FIG. 25

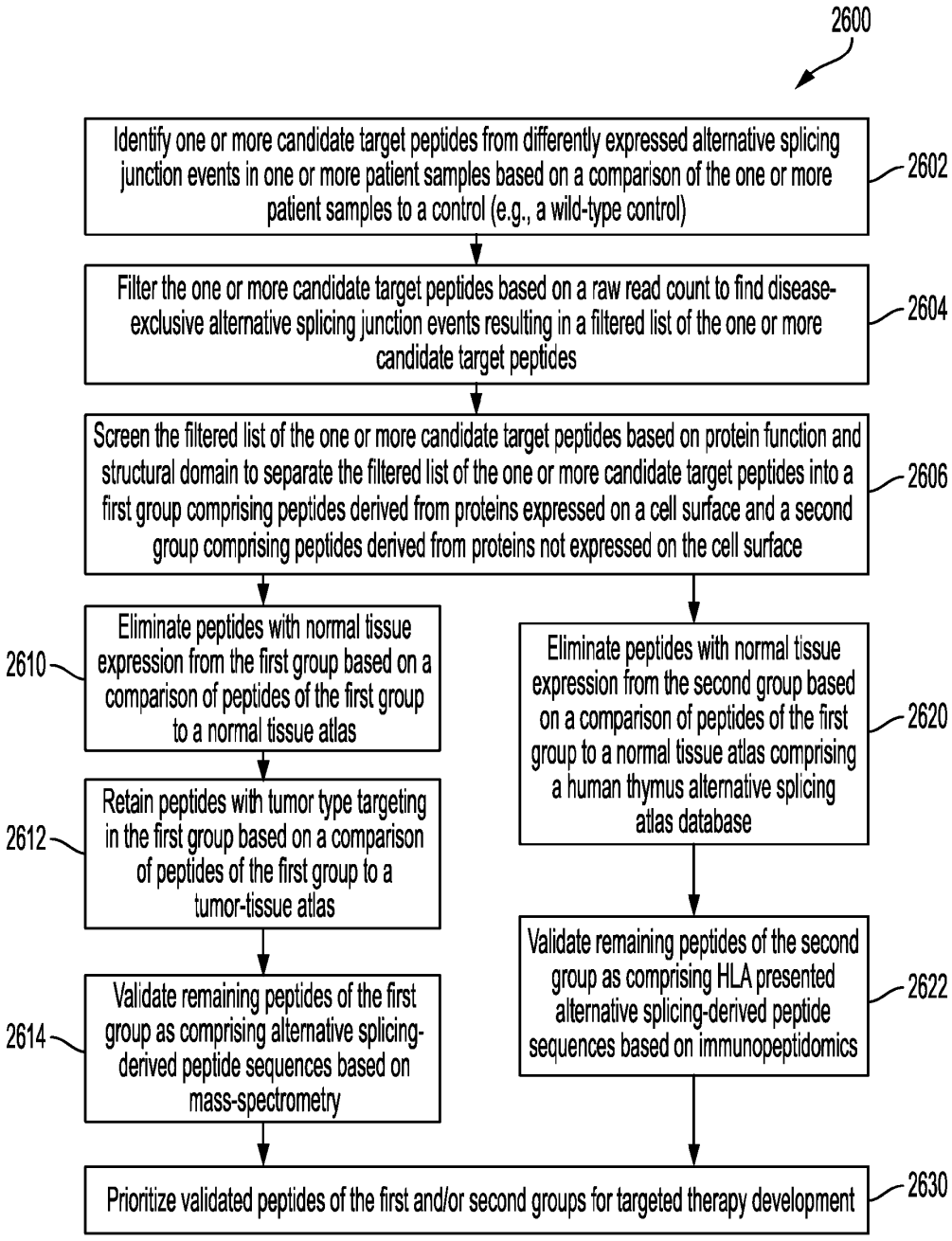

2600
Identify one or more candidate target peptides from differently expressed alternative splicing junction events in one or more patient samples based on a comparison of the one or more patient samples to a control (e.g., a wild-type control)
2602
Filter the one or more candidate target peptides based on a raw read count to find disease-exclusive alternative splicing junction events resulting in a filtered list of the one or more candidate target peptides
2604
Screen the filtered list of the one or more candidate target peptides based on protein function and structural domain to separate the filtered list of the one or more candidate target peptides into a first group comprising peptides derived from proteins expressed on a cell surface and a second group comprising peptides derived from proteins not expressed on the cell surface
2606
Eliminate peptides with normal tissue expression from the first group based on a comparison of peptides of the first group to a normal tissue atlas
2610
Retain peptides with tumor type targeting in the first group based on a comparison of peptides of the first group to a tumor-tissue atlas
2612
Validate remaining peptides of the first group as comprising alternative splicing-derived peptide sequences based on mass-spectrometry
2614
Eliminate peptides with normal tissue expression from the second group based on a comparison of peptides of the first group to a normal tissue atlas comprising a human thymus alternative splicing atlas database
2620
Validate remaining peptides of the second group as comprising HLA presented alternative splicing-derived peptide sequences based on immunopeptidomics
2622
Prioritize validated peptides of the first and/or second groups for targeted therapy development
2630

FIG. 26A

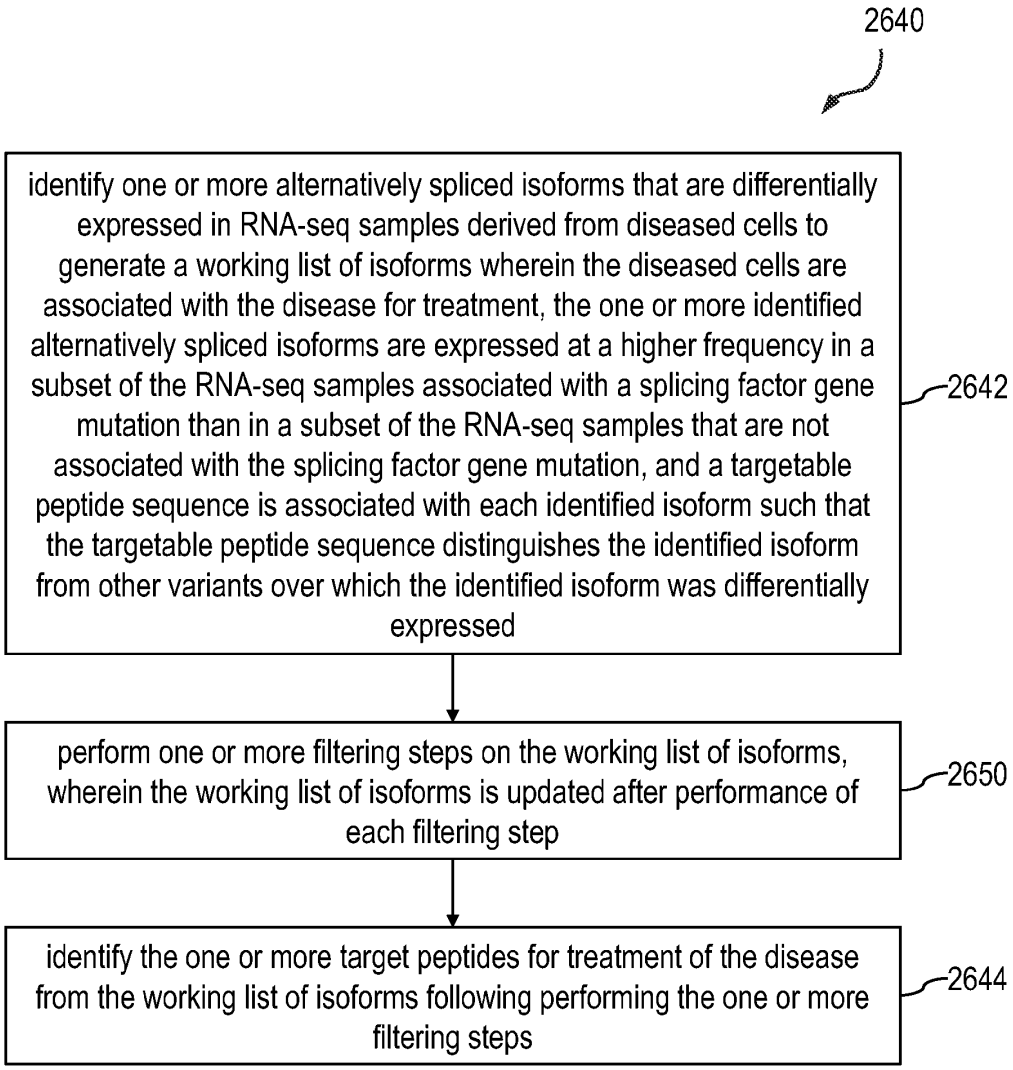

2640
identify one or more alternatively spliced isoforms that are differentially expressed in RNA-seq samples derived from diseased cells to generate a working list of isoforms wherein the diseased cells are associated with the disease for treatment, the one or more identified alternatively spliced isoforms are expressed at a higher frequency in a subset of the RNA-seq samples associated with a splicing factor gene mutation than in a subset of the RNA-seq samples that are not associated with the splicing factor gene mutation, and a targetable peptide sequence is associated with each identified isoform such that the targetable peptide sequence distinguishes the identified isoform from other variants over which the identified isoform was differentially expressed
2642
perform one or more filtering steps on the working list of isoforms, wherein the working list of isoforms is updated after performance of each filtering step
2650
identify the one or more target peptides for treatment of the disease from the working list of isoforms following performing the one or more filtering steps
2644

FIG. 26B

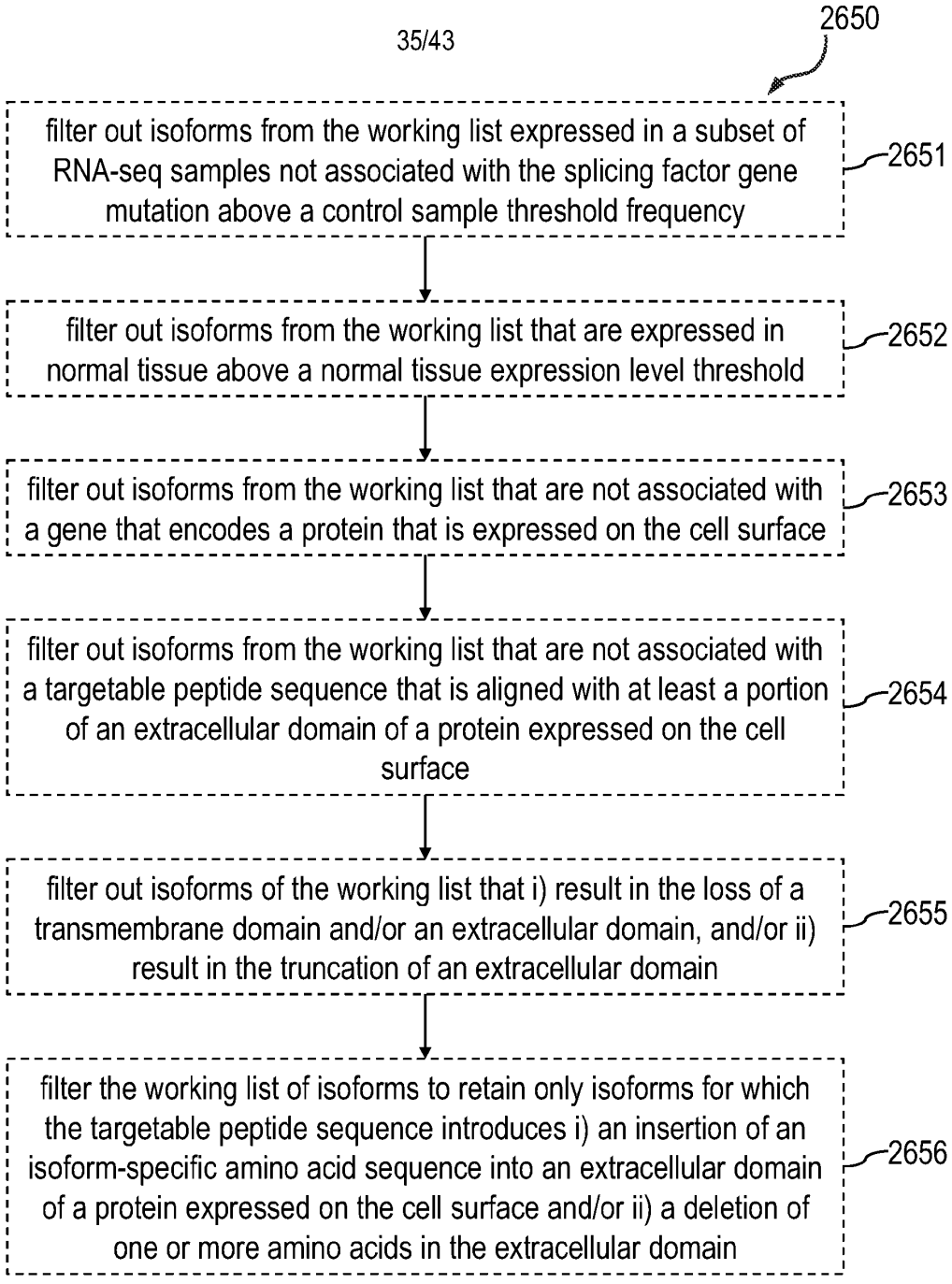

2650
filter out isoforms from the working list expressed in a subset of RNA-seq samples not associated with the splicing factor gene mutation above a control sample threshold frequency
2651
filter out isoforms from the working list that are expressed in normal tissue above a normal tissue expression level threshold
2652
filter out isoforms from the working list that are not associated with a gene that encodes a protein that is expressed on the cell surface
2653
filter out isoforms from the working list that are not associated with a targetable peptide sequence that is aligned with at least a portion of an extracellular domain of a protein expressed on the cell surface
2654
filter out isoforms of the working list that i) result in the loss of a transmembrane domain and/or an extracellular domain, and/or ii) result in the truncation of an extracellular domain
2655
filter the working list of isoforms to retain only isoforms for which the targetable peptide sequence introduces i) an insertion of an isoform-specific amino acid sequence into an extracellular domain of a protein expressed on the cell surface and/or ii) a deletion of one or more amino acids in the extracellular domain
2656

FIG. 26C

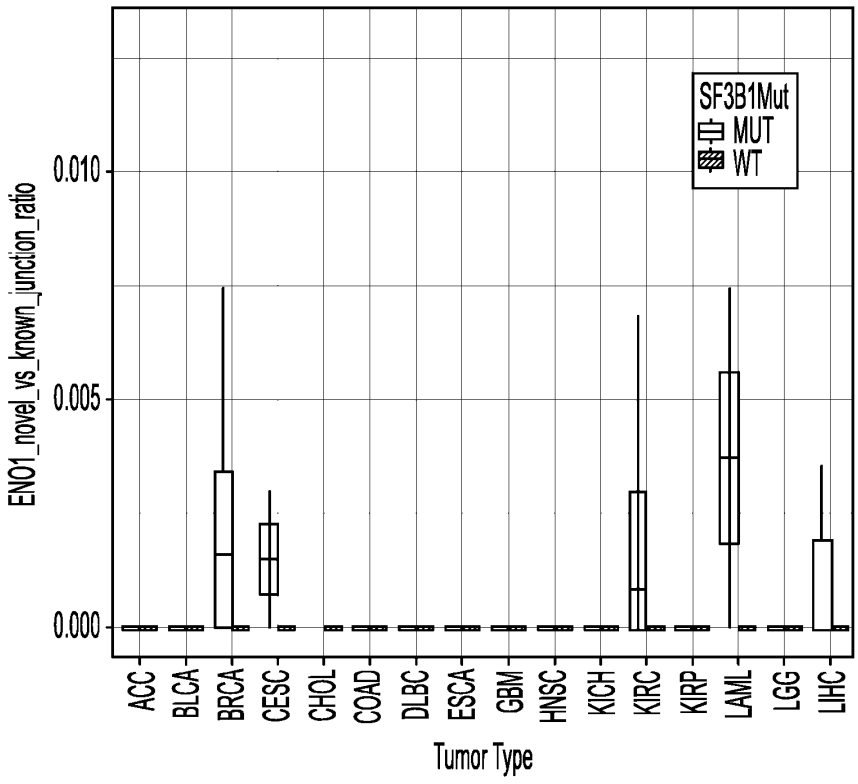

SF3B1Mut
MUT
WT
0.010
0.005
0.000
ENO1_novel_vs_known_junction_ratio
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
Tumor Type

FIG. 27A

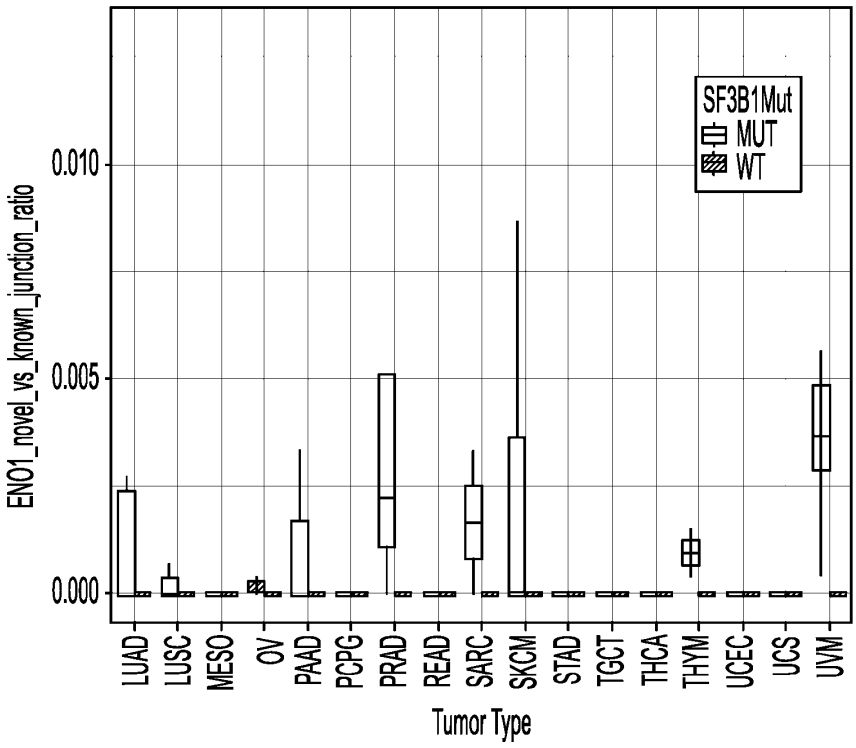

SF3B1Mut
MUT
WT
ENO1_novel_vs_known_junction_ratio
0.010
0.005
0.000
LUAD
LUSC
MESO
OV
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM
Tumor Type

FIG. 27B

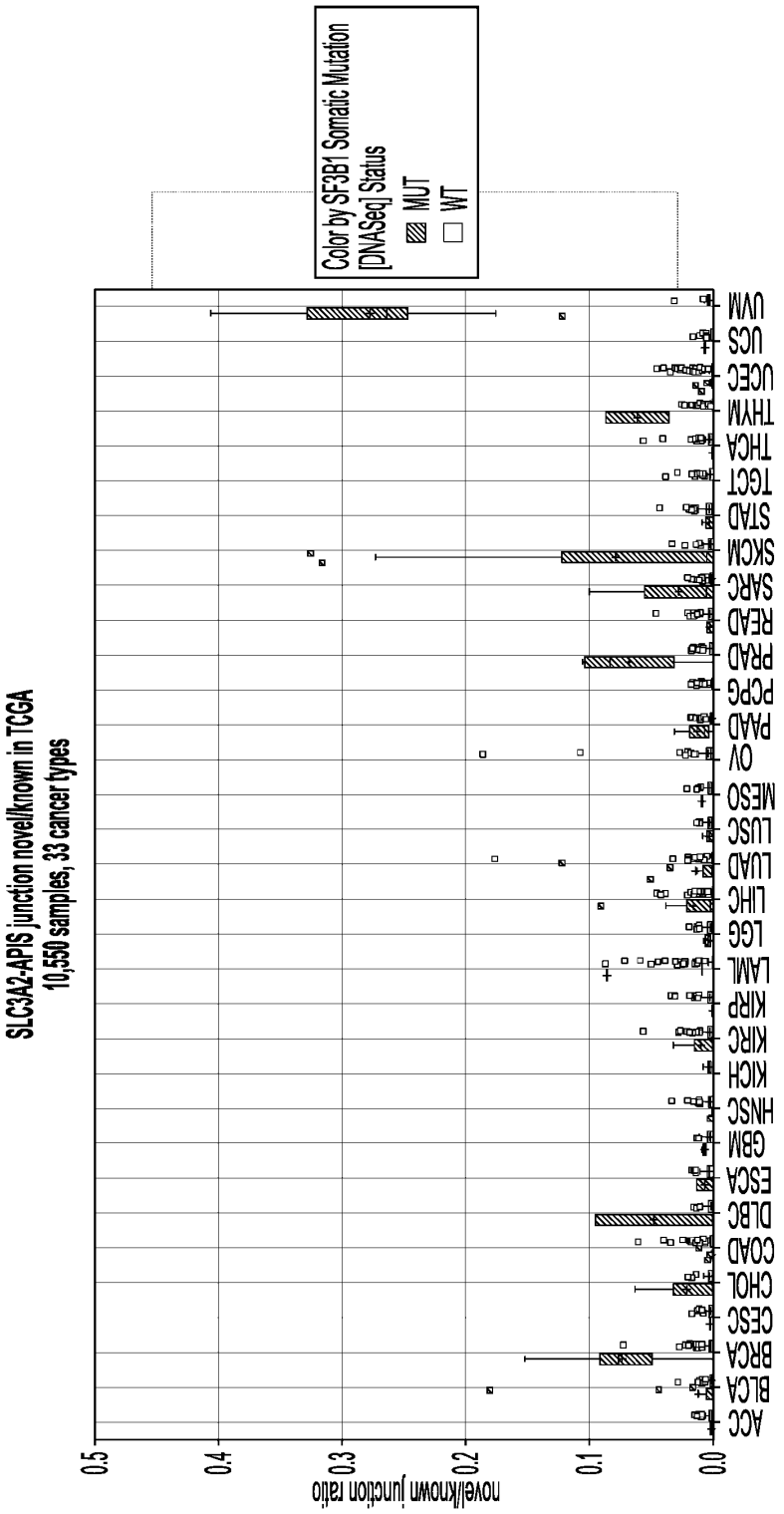

SLC3A2-APIS junction novel/known in TCGA
10,550 samples, 33 cancer types
novel/known junction ratio
0.0
0.1
0.2
0.3
0.4
0.5
Color by SF3B1 Somatic Mutation [DNASeq] Status
MUT
WT
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM

FIG. 28

Summary of SLC3A2-APIS in TCGA

Tumor type	missing	MUT	WT	Grand Total
ACC	1	2	76	79
BLCA	20	24	367	411
BRCA	114	19	981	1114
CESC	111	1	194	306
CHOL	1	3	32	36
COAD	102	16	355	473
CLBC		2	46	48
ESCA	1	2	182	185
GBM	6	4	176	186
HNSC	2	6	514	522
KICH			66	66
KIRC	32	5	501	538
KIRP	9	7	275	291
LAML	3	1	175	179
LGG	1	6	527	534
LIHC	7	11	356	374
LUAD	14	16	500	530
LUSC	18	13	470	501
MESO	4	1	82	87
OV	76		354	430
PAAD	1	3	175	179
PCPG			184	184
PRAD	3	4	495	502
READ	18	2	147	167
SARC	11	4	248	263
SKCM	1	23	448	472
STAD	44	9	363	416
TGCT	1		155	156
THCA	11	2	500	513
THYM	1	2	117	120
UCEC	303	15	233	551
UCS		1	56	57
UVM		18	62	80
Grand Total	916	222	9412	10550

FIG. 29

mean APIS ratio in MUT	mean APIS ratio in WT	Fold(MUT/WT)
0.0012	0.002	0.6
0.0114	0.0016	7.1
0.0732	0.002	36.6
0.0024	0.0018	1.3
0.0211	0.0027	7.8
0.0011	0.0021	0.5
0.0476	0.0018	26.4
0.0066	0.0031	2.1
0.0068	0.0025	2.7
0.0003	0.0023	0.1
	0.0028	0.0
0.0095	0.0026	3.7
0	0.0026	0.0
0.0857	0.0037	23.2
0.0039	0.0023	1.7
0.0161	0.0018	8.9
0.0141	0.0017	8.3
0.0032	0.002	1.6
0.0088	0.0022	4.0
	0.0039	0.0
0.0125	0.0019	6.6
	0.0017	0.0
0.0678	0.0019	35.7
0.0024	0.0025	1.0
0.0275	0.0016	17.2
0.0786	0.0023	34.2
0.0034	0.003	1.1
	0.002	0.0
0	0.0022	0.0
0.0613	0.0019	32.3
0.0019	0.0026	0.7
0.0063	0.0015	4.2
0.2772	0.004	69.3

FIG. 29
CONTINUED

Samples with SLC3A2-APIS ratio >5%

tumor type	SF3B1 MUT	SF3B1 WT	Grand Total	% in all MUT	% in all WT
BLCA	1		1	5	
BRCA	14	1	15	58.3	0.27
CHOL	1		1	33.3	
COAD		1	1		0.3
DLBC	1		1	50	
KIRC		1	1		0.2
LAML	1	4	5	100	2.3*
LIHC	1		1	9	
LUAD	2	1	3	12.5	0.6
OV		2	2		0.56
PRAD	3		3	75	
SARC	1		1	24	
SKCM	8		8	34.8	
THCA		1	1		0.2
THYM	1		1	50	
UVM	18		18	100	
Grand Total	**52**	**11**	**63**		

FIG. 29
CONTINUED

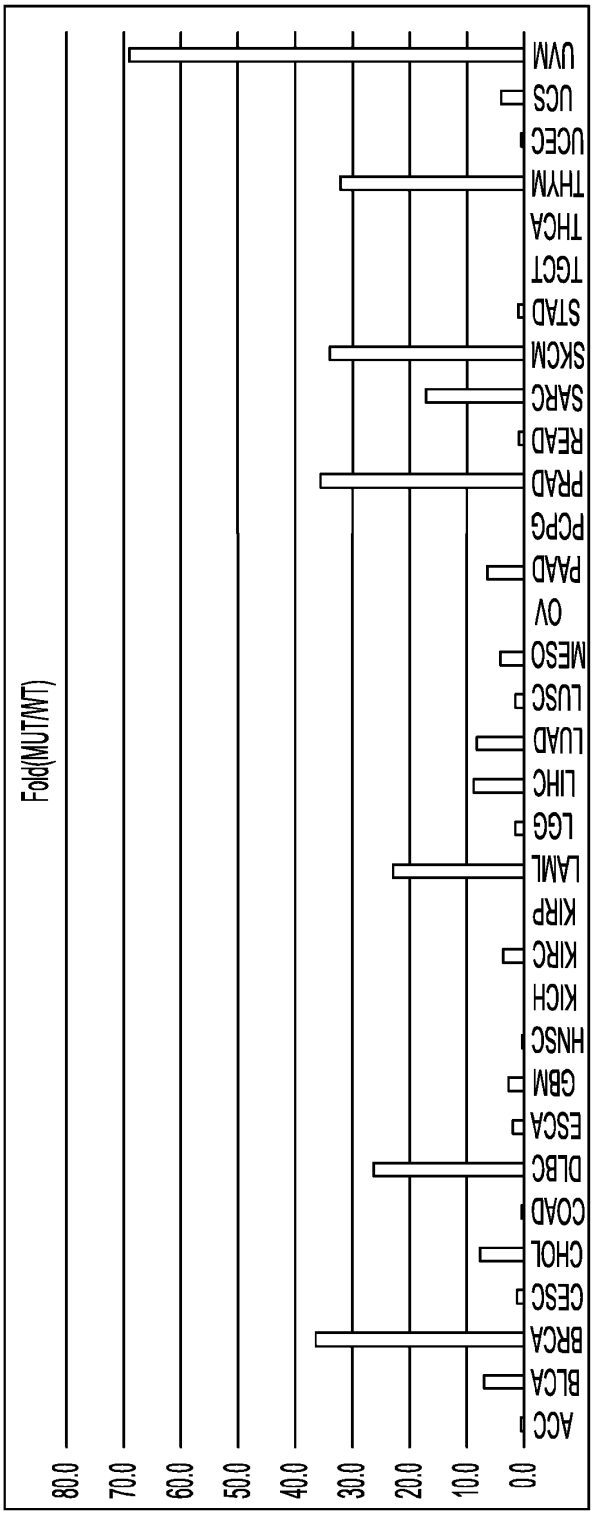


FIG. 29
CONTINUED

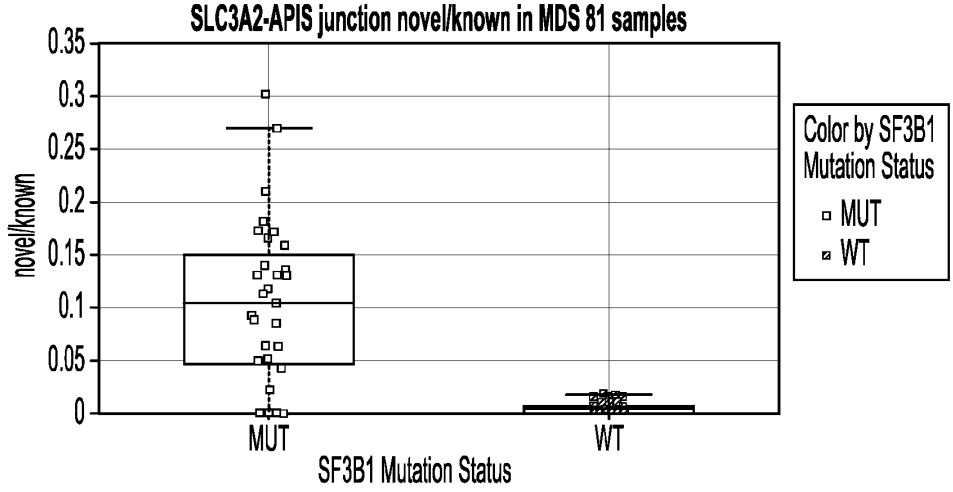

SLC3A2-APIS junction novel/known in MDS 81 samples
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
novel/known
MUT
WT
SF3B1 Mutation Status
Color by SF3B1 Mutation Status
MUT
WT

FIG. 30

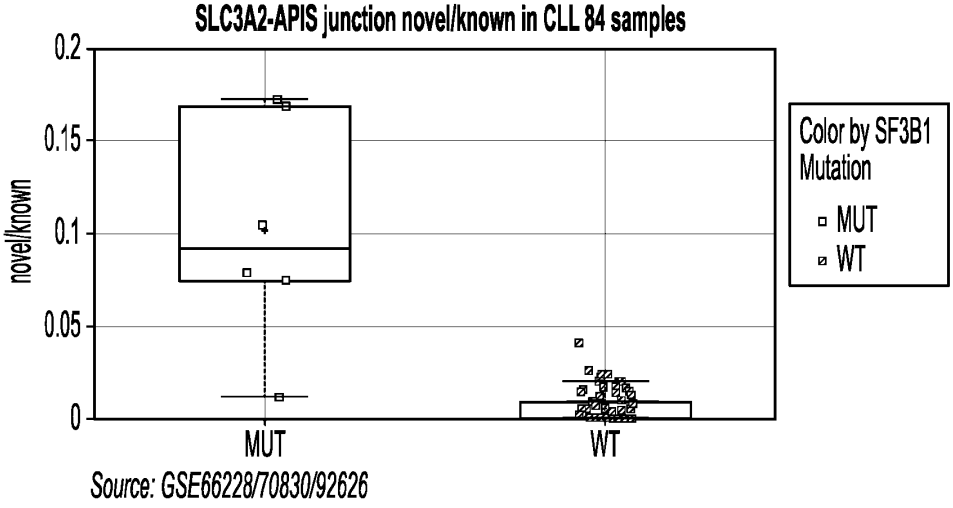

SLC3A2-APIS junction novel/known in CLL 84 samples
0.2
0.15
0.1
0.05
0
novel/known
MUT
WT
Source: GSE66228/70830/92626
Color by SF3B1 Mutation
MUT
WT

FIG. 31