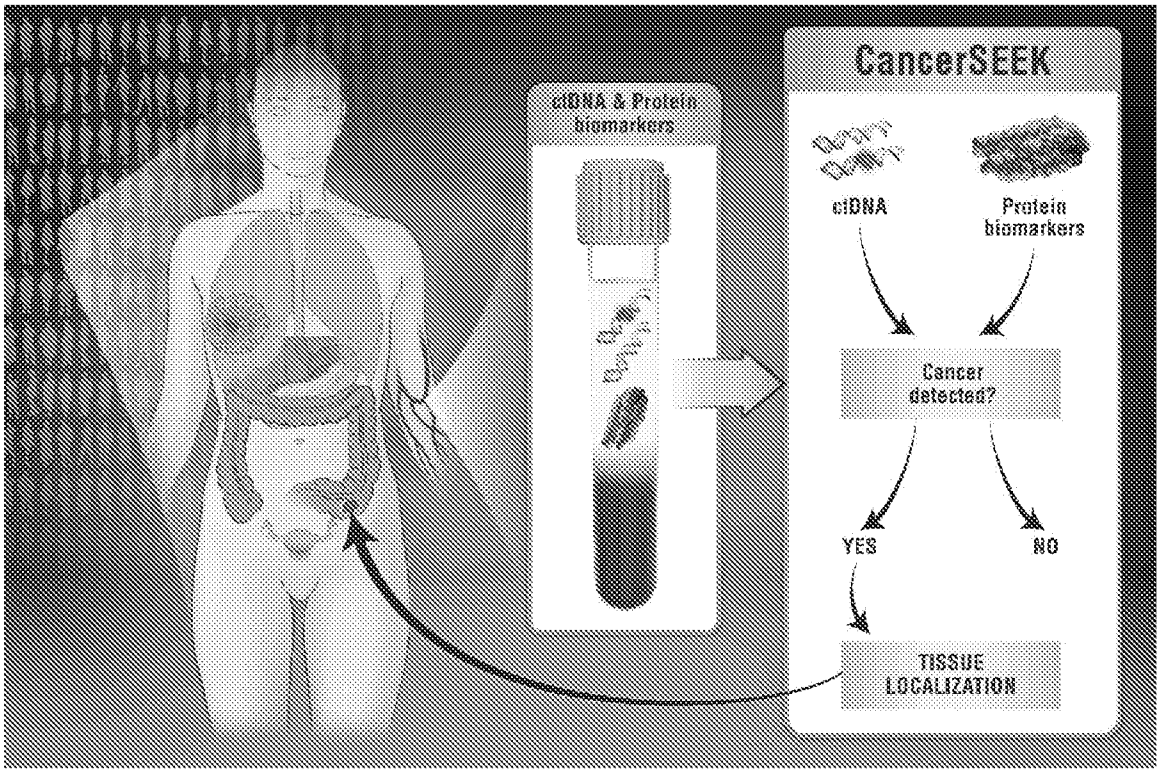

CancerSEEK
ctDNA & Protein biomarkers
ctDNA
Protein biomarkers
Cancer detected?
YES
NO
TISSUE LOCALIZATION

FIG. 1

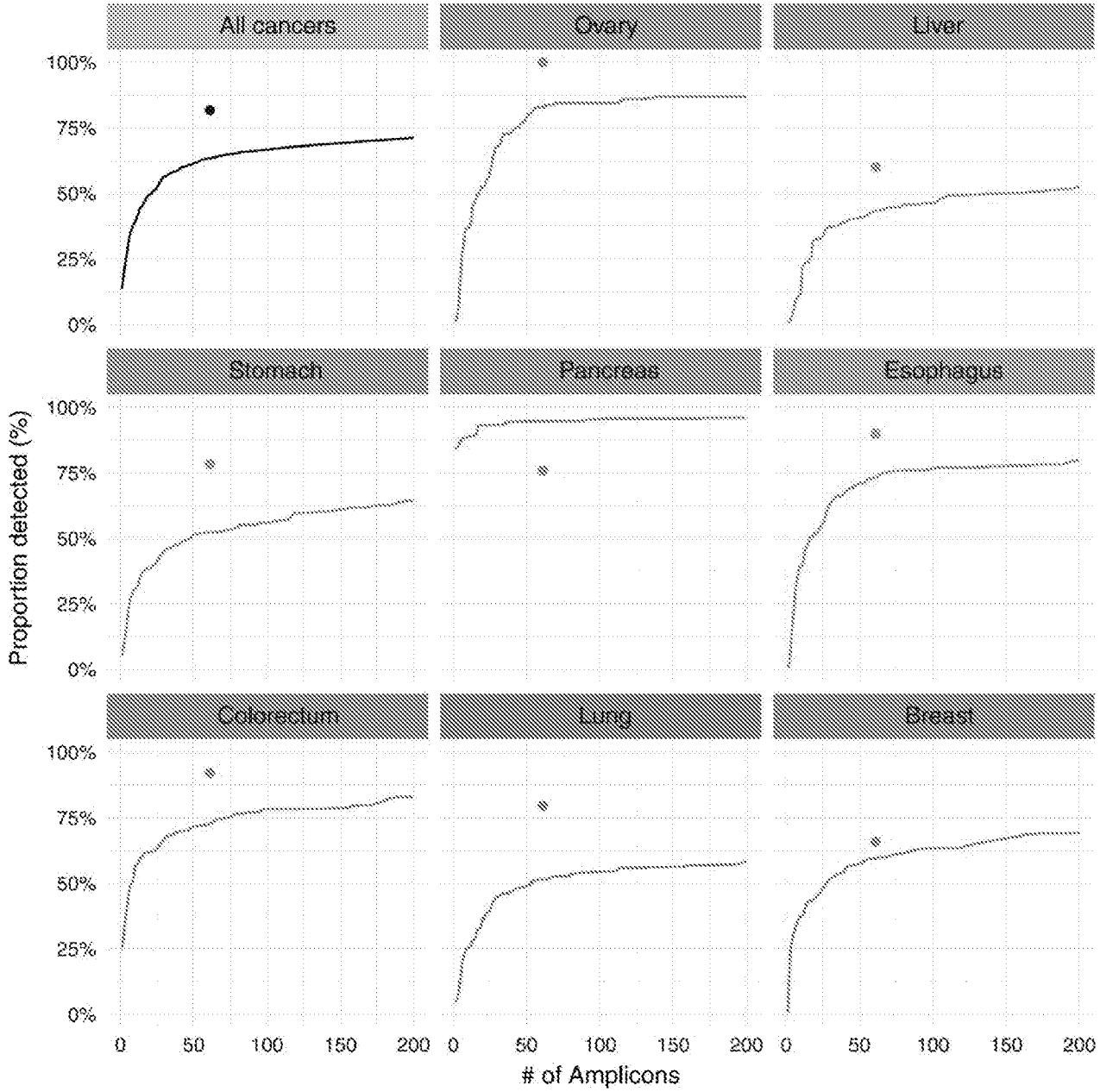

All cancers
Ovary
Liver
Stomach
Pancreas
Esophagus
Colorectum
Lung
Breast
Proportion detected (%)
100%
75%
50%
25%
0%
0
50
100
150
200
# of Amplicons

FIG. 2

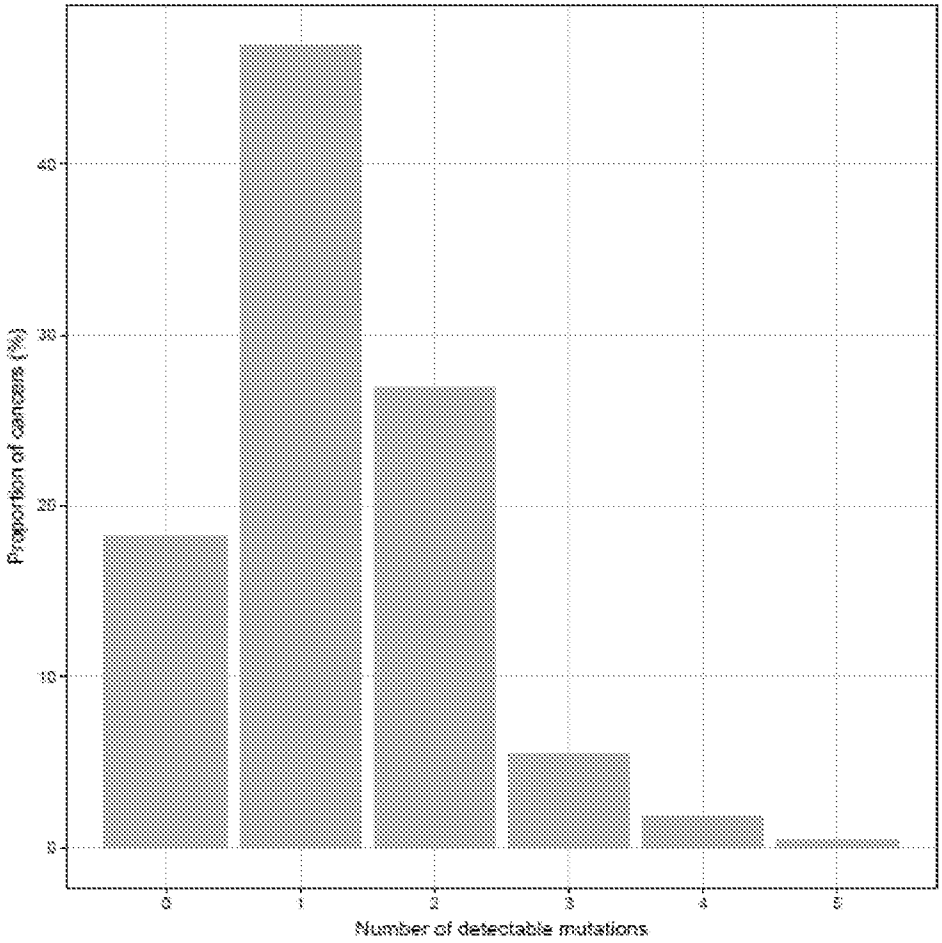

Proportion of cancers (%)
40
30
20
10
0
0
1
2
3
4
5
Number of detectable mutations

FIG. 3

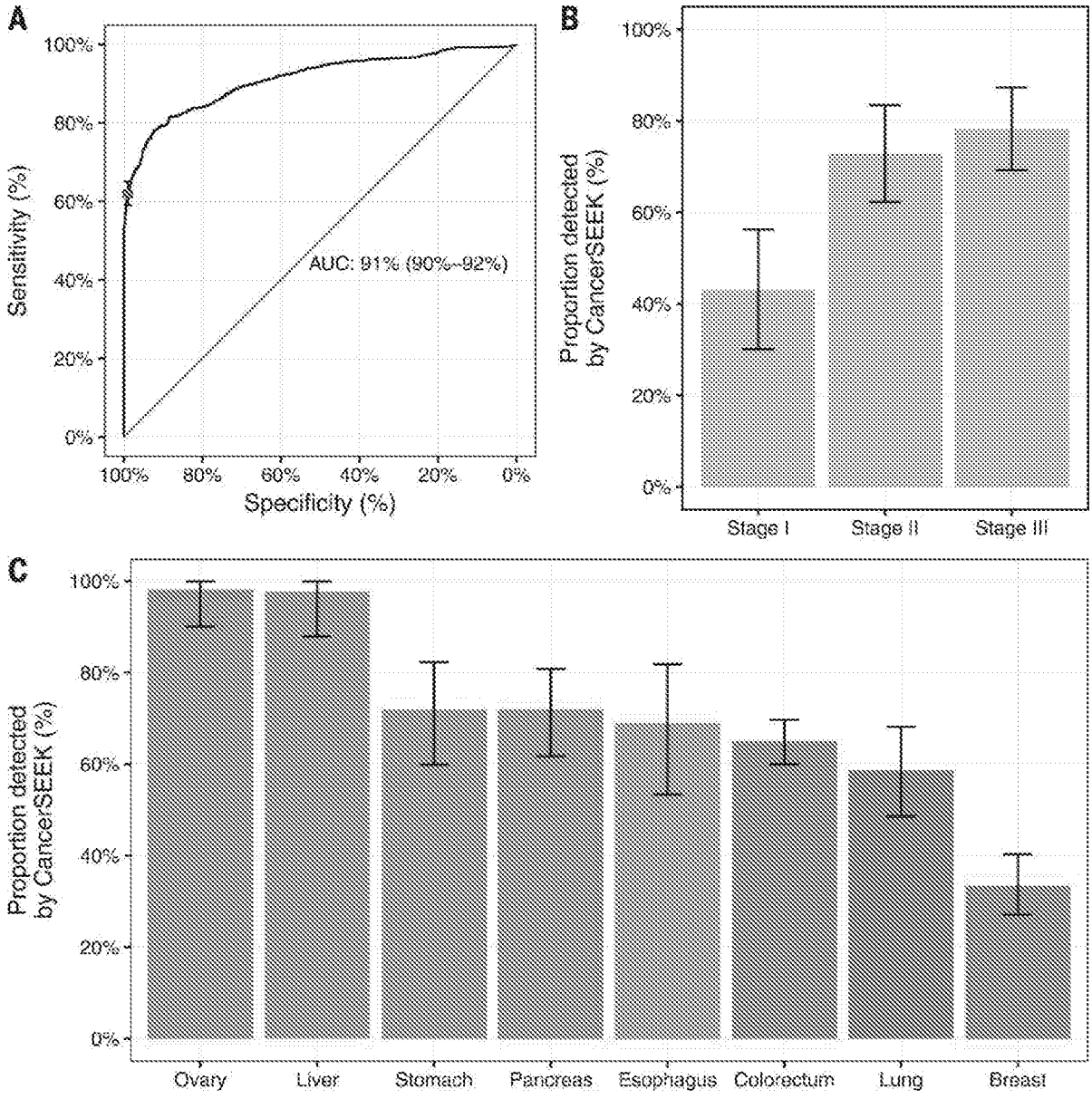

A
100%
80%
60%
40%
20%
0%
Sensitivity (%)
AUC: 91% (90%~92%)
100%
80%
60%
40%
20%
0%
Specificity (%)
B
100%
80%
60%
40%
20%
0%
Proportion detected by CancerSEEK (%)
Stage I
Stage II
Stage III
C
100%
80%
60%
40%
20%
0%
Proportion detected by CancerSEEK (%)
Ovary
Liver
Stomach
Pancreas
Esophagus
Colorectum
Lung
Breast

FIG. 4

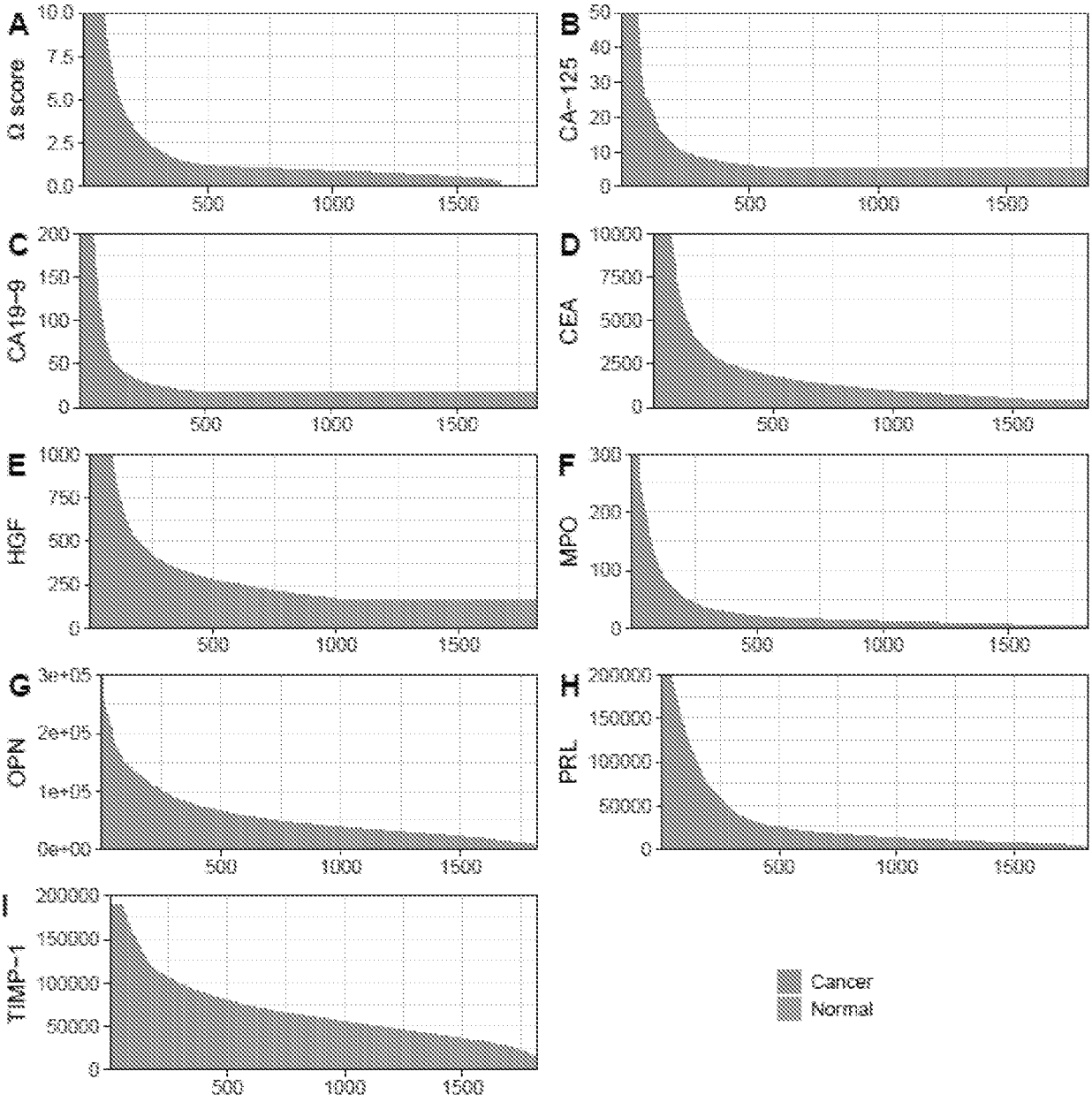

A
Ω score
10.0
7.5
5.0
2.5
0.0
500
1000
1500
B
CA-125
50
40
30
20
10
0
C
CA19-9
200
150
100
50
0
D
CEA
10000
7500
5000
2500
E
HGF
1000
750
500
250
F
MPO
300
200
100
G
OPN
3e+05
2e+05
1e+05
0e+00
H
PRL
200000
150000
100000
50000
I
TIMP-1
Cancer
Normal

FIG. 5

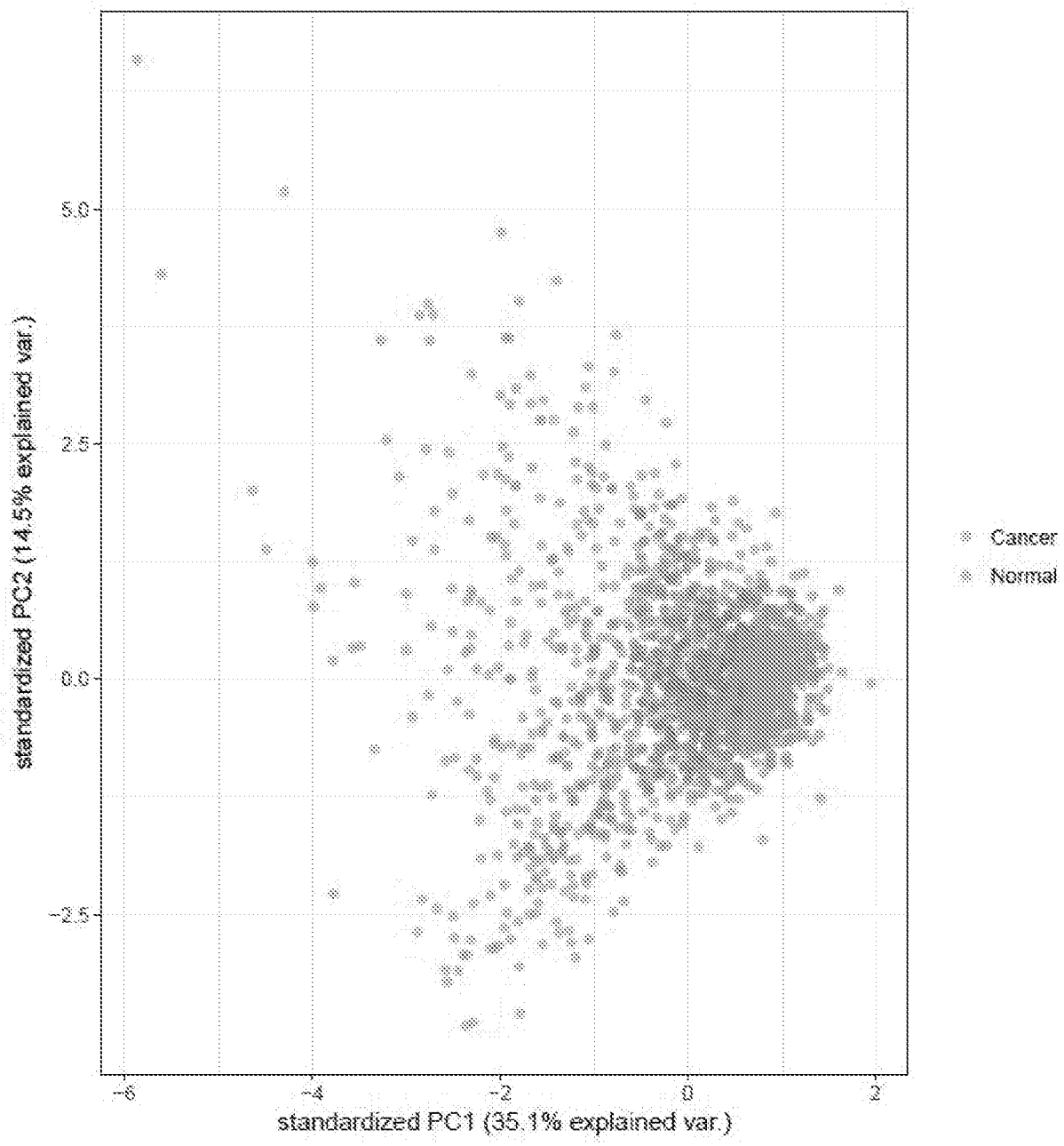

standardized PC2 (14.5% explained var.)
5.0
2.5
0.0
−2.5
−6
−4
−2
0
2
standardized PC1 (35.1% explained var.)
Cancer
Normal

FIG. 6

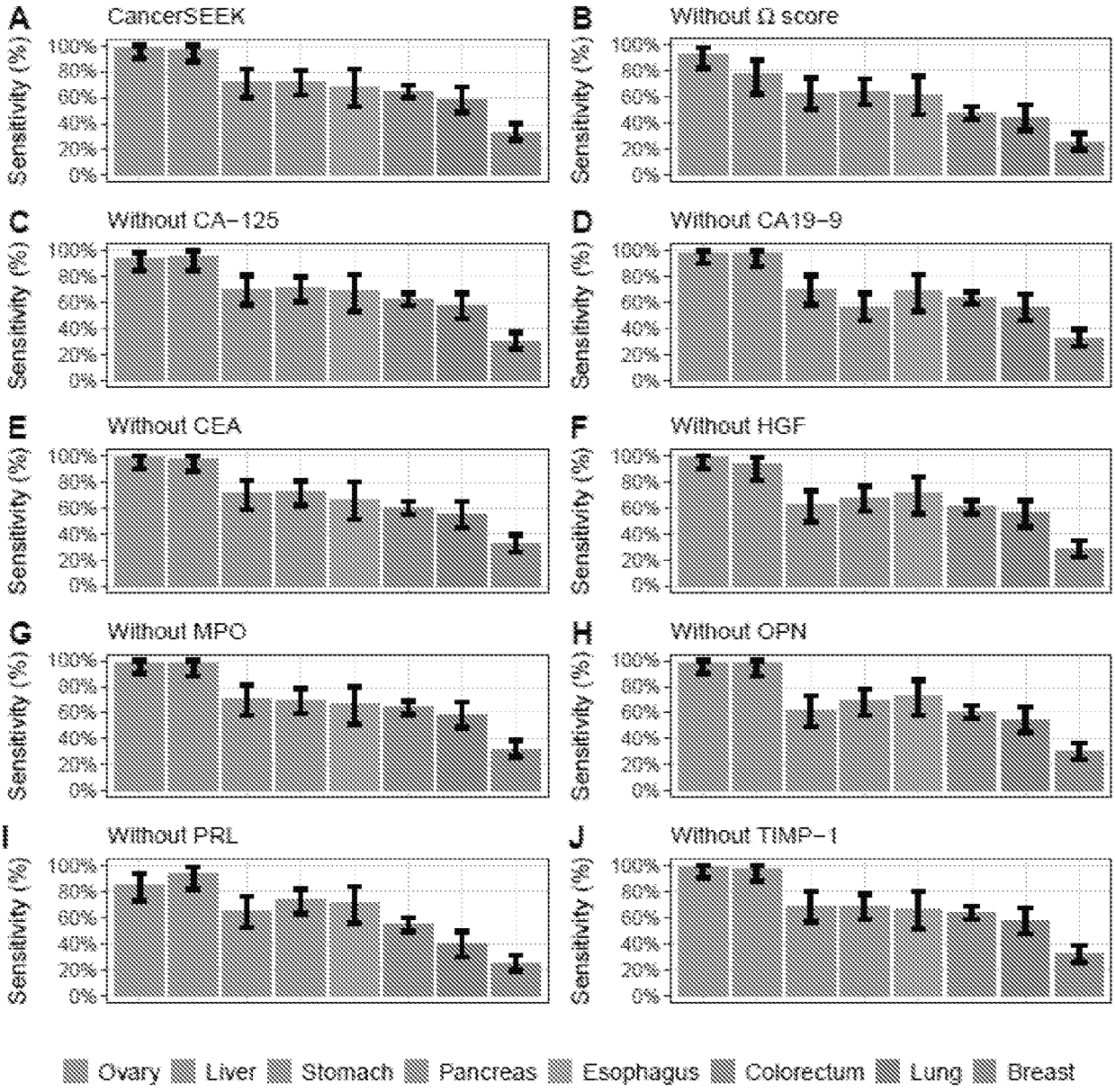

A
CancerSEEK
B
Without Ω score
C
Without CA-125
D
Without CA19-9
E
Without CEA
F
Without HGF
G
Without MPO
H
Without OPN
I
Without PRL
J
Without TIMP-1
Sensitivity (%)
100%
80%
60%
40%
20%
0%
Ovary
Liver
Stomach
Pancreas
Esophagus
Colorectum
Lung
Breast

FIG. 7

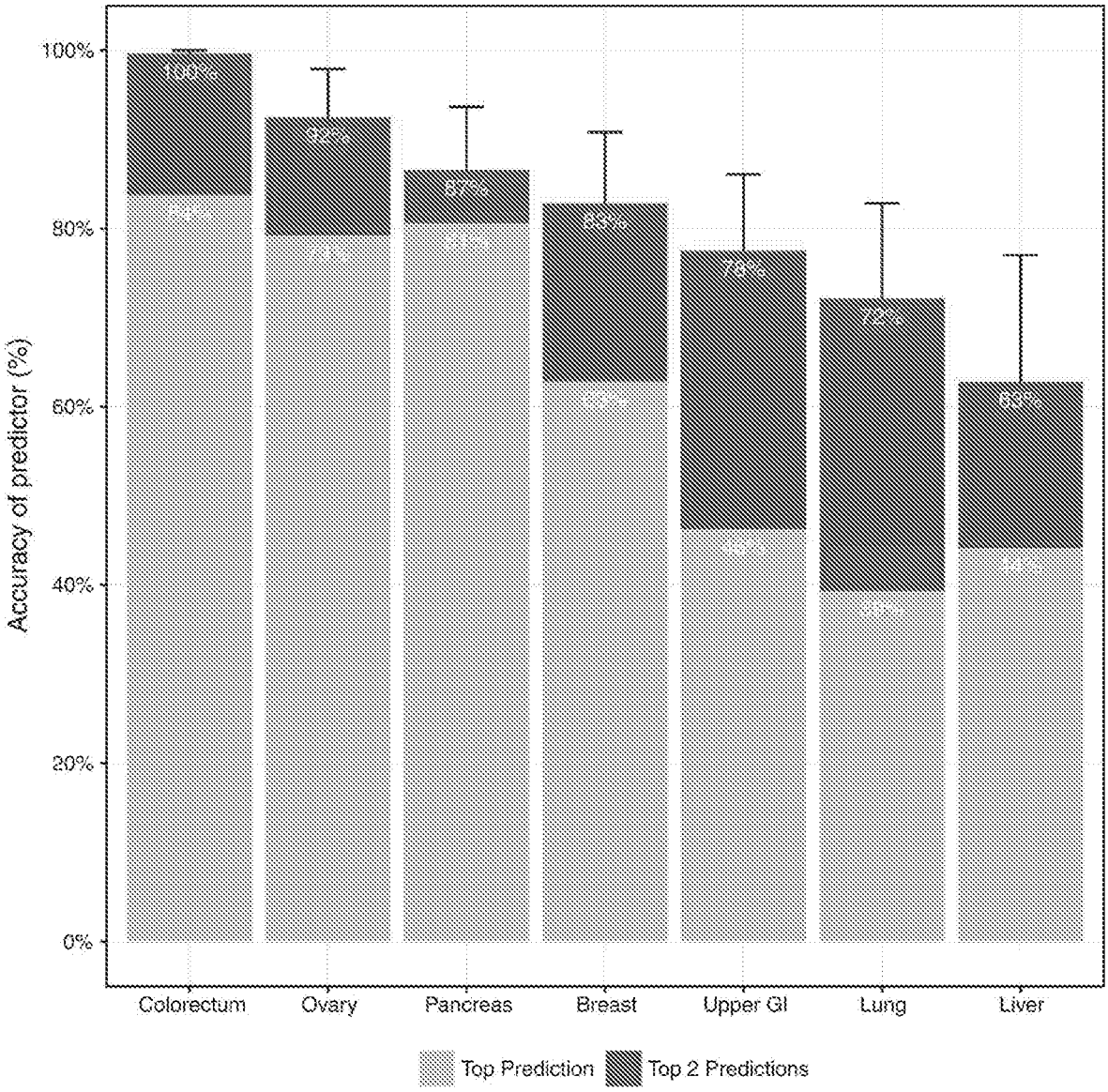

100%
80%
60%
40%
20%
0%
Accuracy of predictor (%)
100%
92%
87%
83%
78%
72%
63%
Colorectum
Ovary
Pancreas
Breast
Upper GI
Lung
Liver
Top Prediction
Top 2 Predictions

FIG. 8

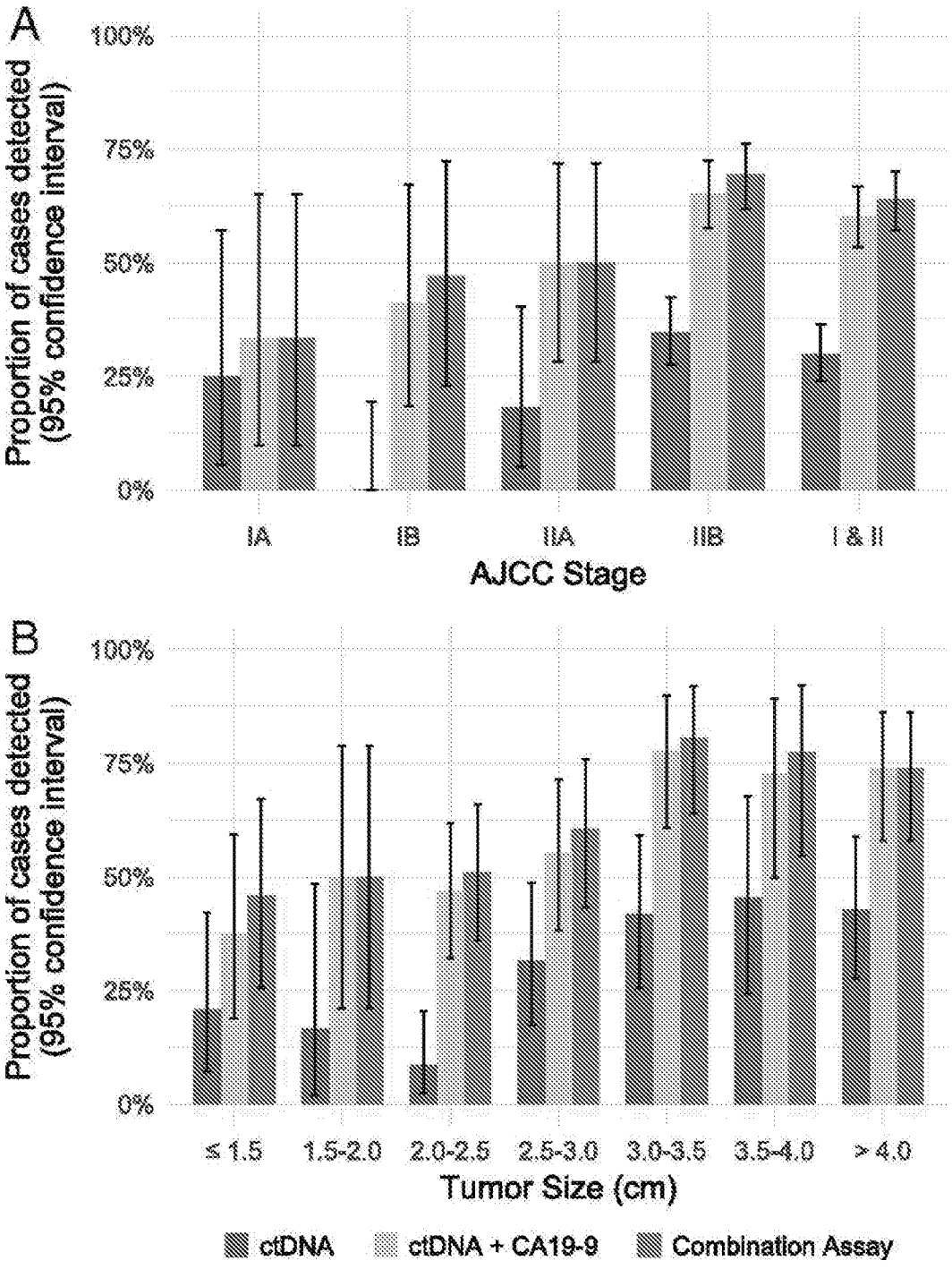

A
Proportion of cases detected (95% confidence interval)
100%
75%
50%
25%
0%
IA
IB
IIA
IIB
I & II
AJCC Stage
B
Proportion of cases detected (95% confidence interval)
100%
75%
50%
25%
0%
≤ 1.5
1.5-2.0
2.0-2.5
2.5-3.0
3.0-3.5
3.5-4.0
> 4.0
Tumor Size (cm)
ctDNA
ctDNA + CA19-9
Combination Assay

FIG. 9

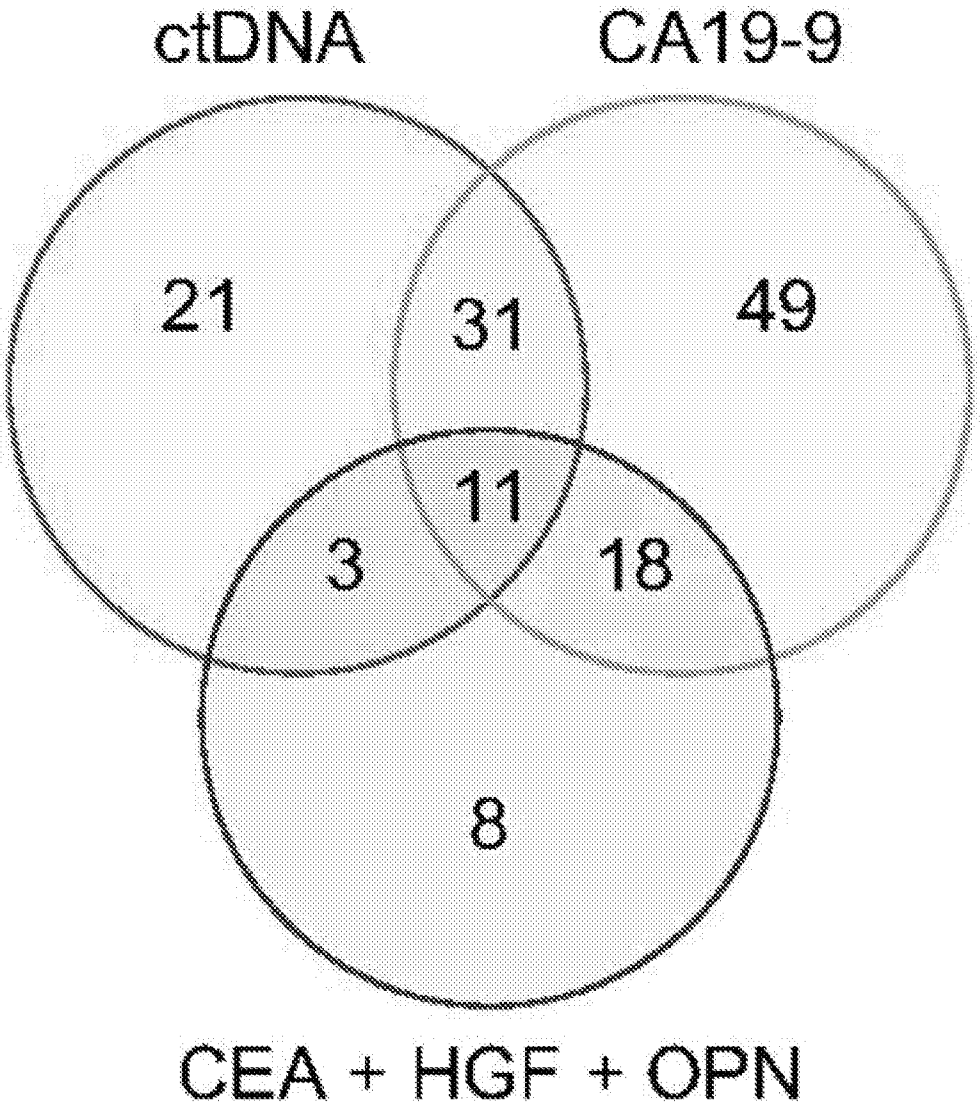


FIG. 10

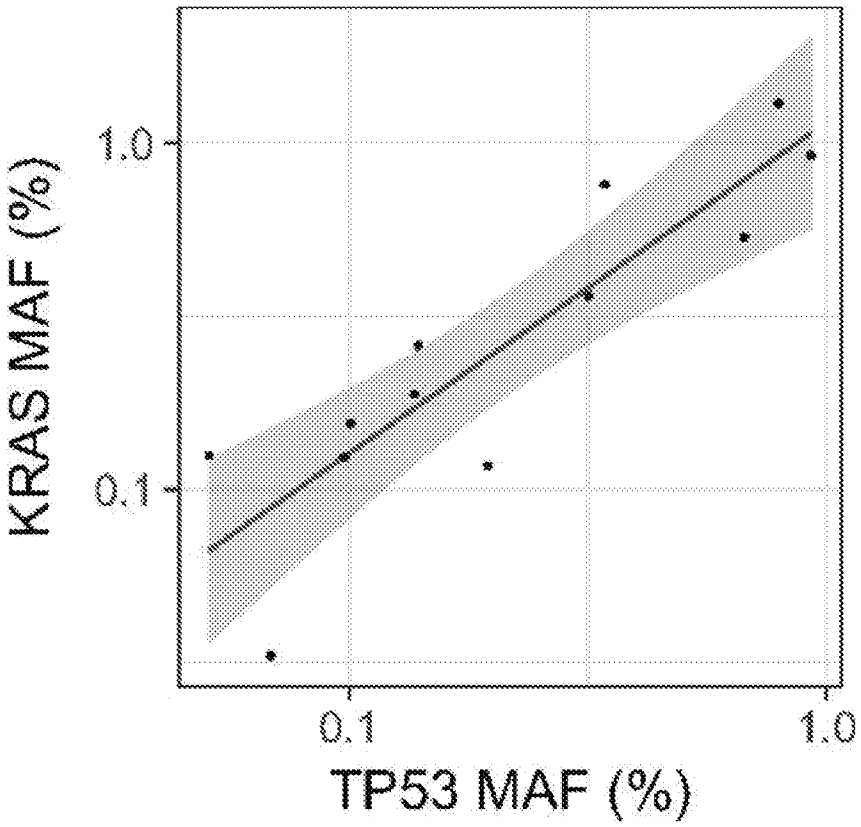


FIG. 11

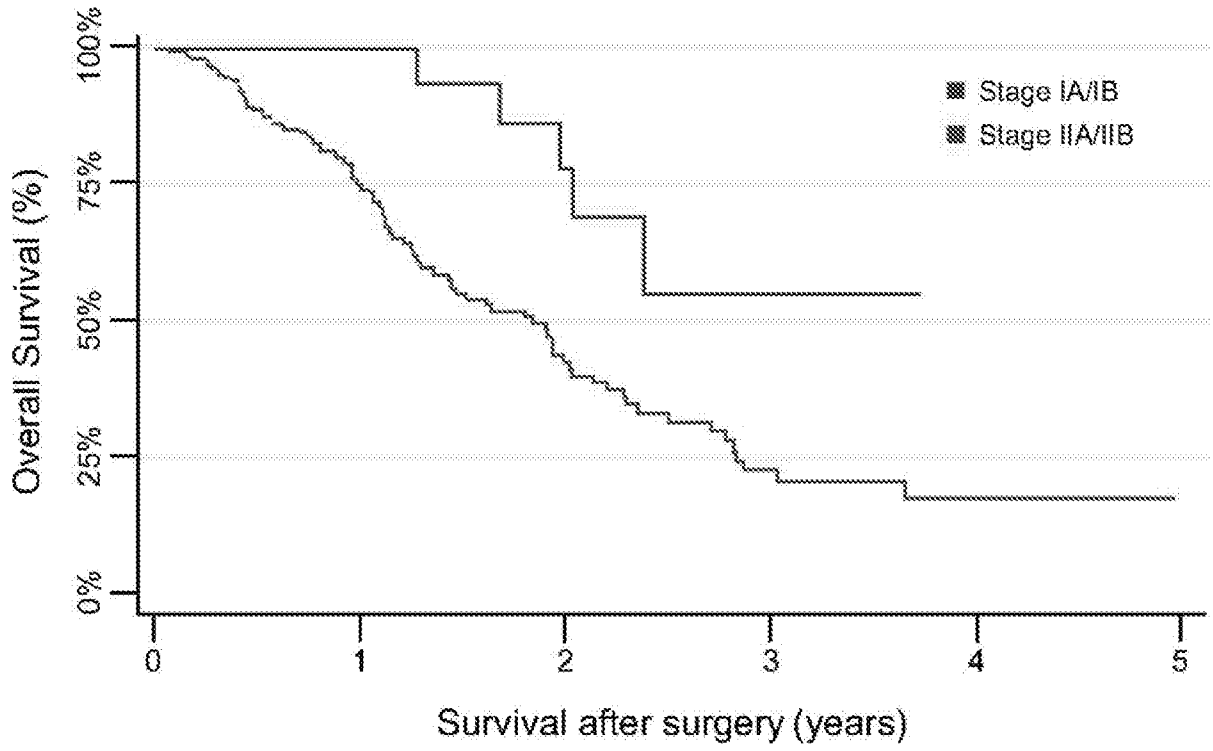

Overall Survival (%)
100%
75%
50%
25%
0%
0
1
2
3
4
5
Survival after surgery (years)
Stage IA/IB
Stage IIA/IIB

FIG. 12

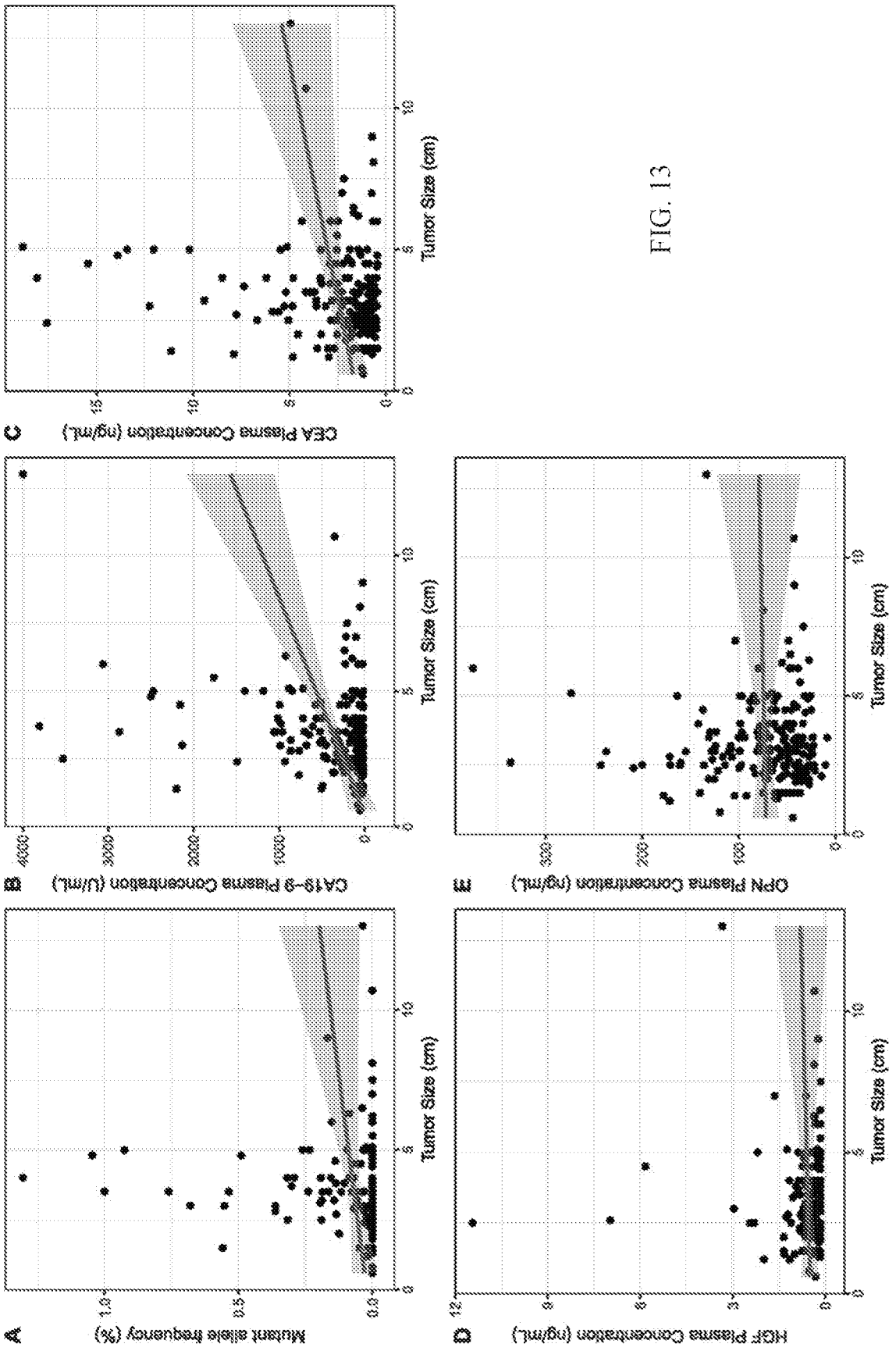

A
Mutant allele frequency (%)
1.0
0.5
0.0
0
5
10
Tumor Size (cm)
B
CA19-9 Plasma Concentration (U/mL)
4000
3000
2000
1000
0
0
5
10
Tumor Size (cm)
C
CEA Plasma Concentration (ng/mL)
15
10
5
0
0
5
10
Tumor Size (cm)
D
HGF Plasma Concentration (ng/mL)
12
9
6
3
0
0
5
10
Tumor Size (cm)
E
OPN Plasma Concentration (ng/mL)
300
200
100
0
0
5
10
Tumor Size (cm)

FIG. 13

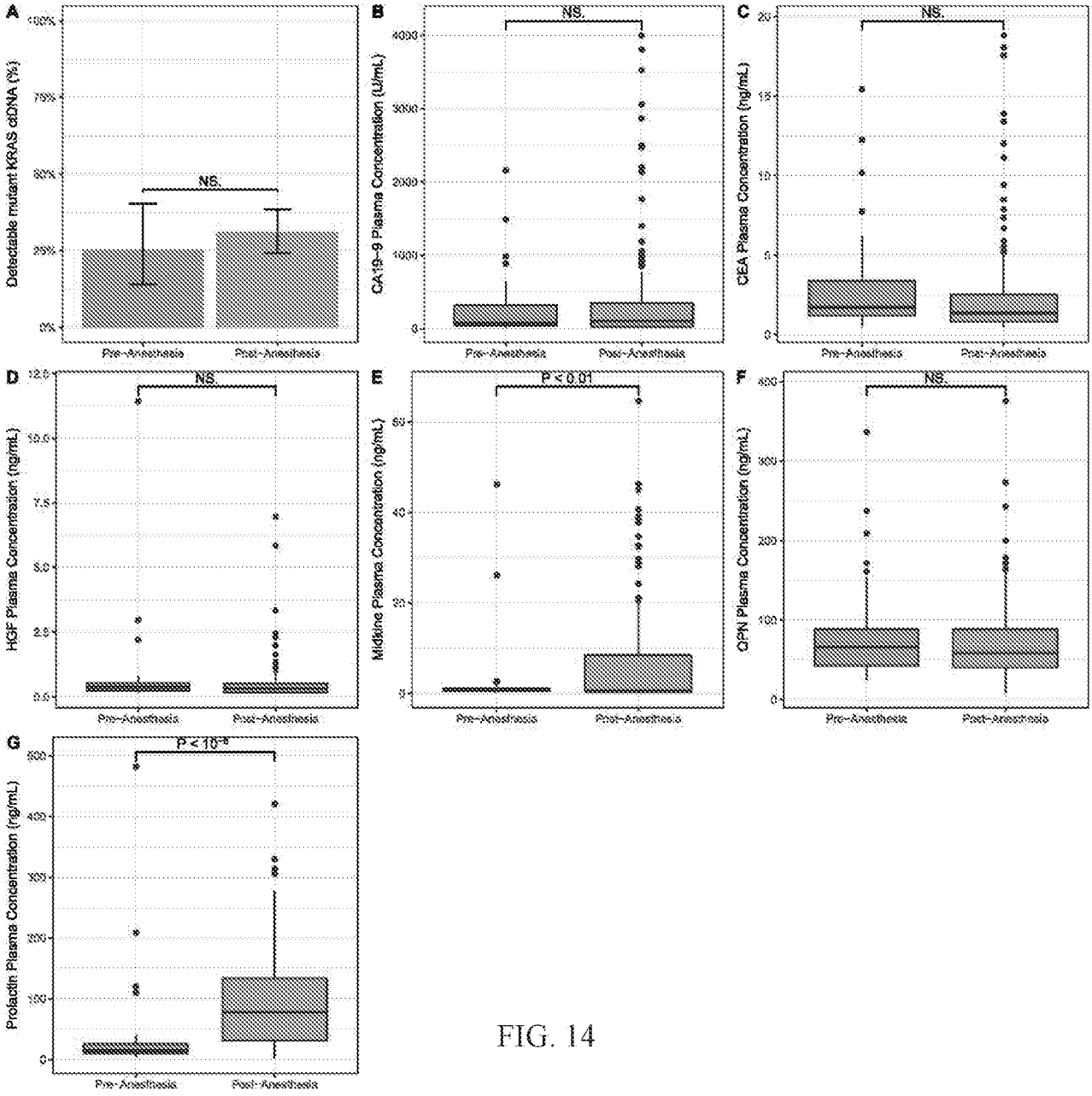

A
Detectable mutant KRAS ctDNA (%)
100%
75%
50%
25%
0%
NS.
Pre-Anesthesia
Post-Anesthesia
B
CA19-9 Plasma Concentration (U/mL)
4000
3000
2000
1000
0
NS.
Pre-Anesthesia
Post-Anesthesia
C
CEA Plasma Concentration (ng/mL)
20
15
10
5
0
NS.
Pre-Anesthesia
Post-Anesthesia
D
HGF Plasma Concentration (ng/mL)
12.5
10.0
7.5
5.0
2.5
0.0
NS.
Pre-Anesthesia
Post-Anesthesia
E
Midkine Plasma Concentration (ng/mL)
60
40
20
0
P < 0.01
Pre-Anesthesia
Post-Anesthesia
F
OPN Plasma Concentration (ng/mL)
400
300
200
100
0
NS.
Pre-Anesthesia
Post-Anesthesia
G
Prolactin Plasma Concentration (ng/mL)
500
400
300
200
100
0
P < 10^-6
Pre-Anesthesia
Post-Anesthesia

FIG. 14

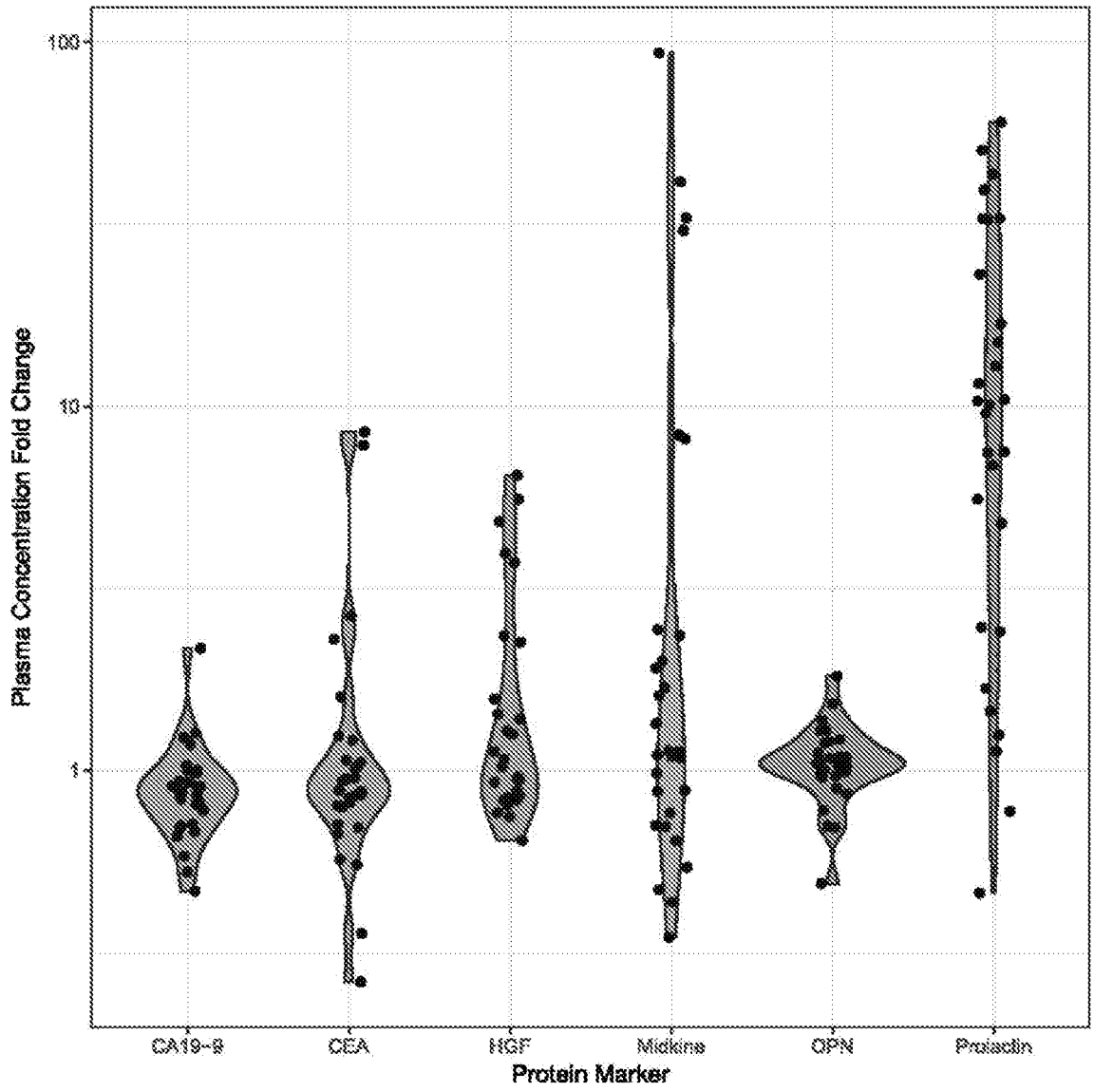

100
10
1
Plasma Concentration Fold Change
CA19-9
CEA
HGF
Midkine
OPN
Prolactin
Protein Marker

FIG. 15

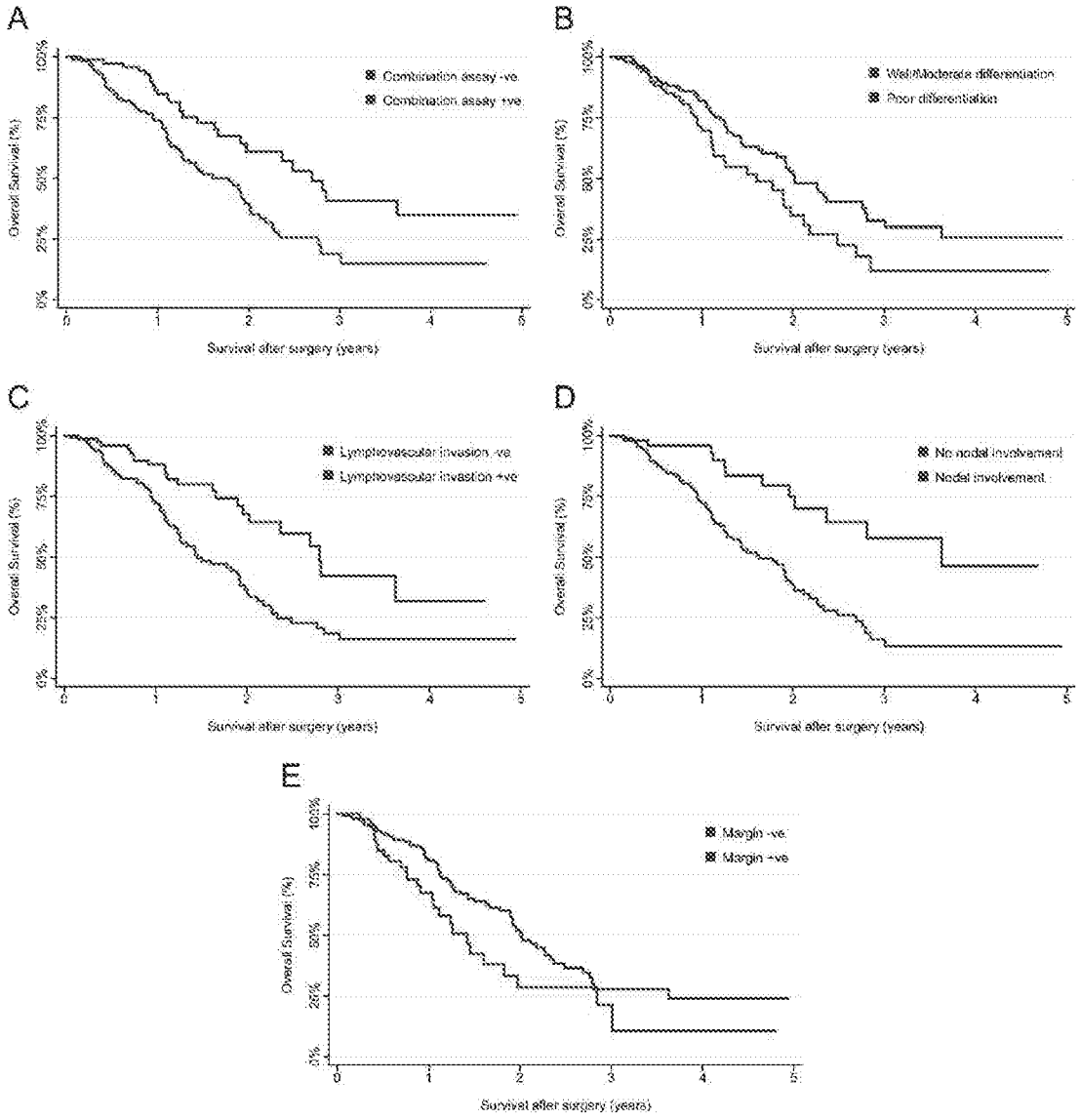

A
Combination assay -ve
Combination assay +ve
Overall Survival (%)
Survival after surgery (years)
B
Well/Moderate differentiation
Poor differentiation
Overall Survival (%)
Survival after surgery (years)
C
Lymphovascular invasion -ve
Lymphovascular invasion +ve
Overall Survival (%)
Survival after surgery (years)
D
No nodal involvement
Nodal involvement
Overall Survival (%)
Survival after surgery (years)
E
Margin -ve
Margin +ve
Overall Survival (%)
Survival after surgery (years)

FIG. 16

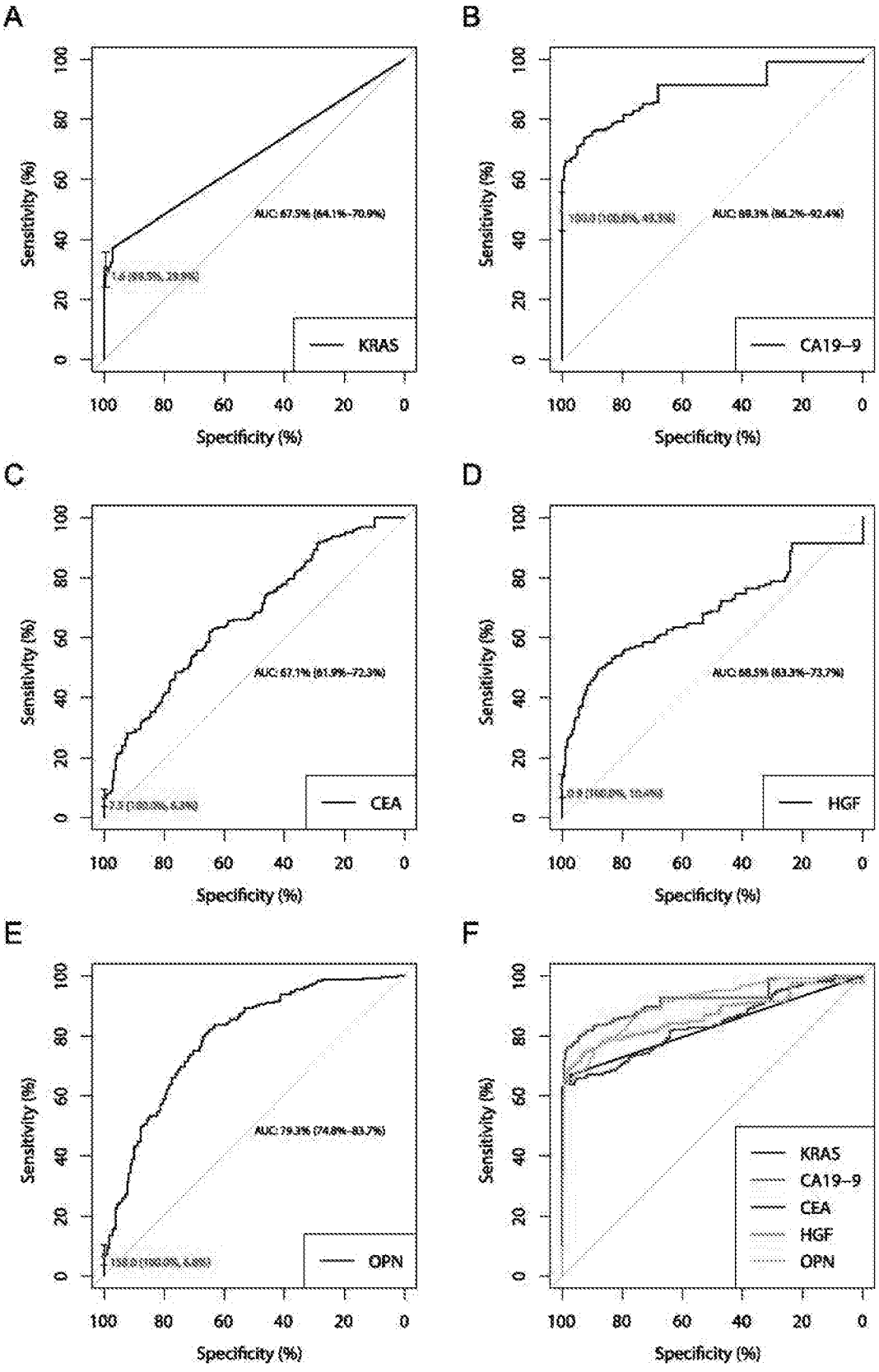

A
Sensitivity (%)
Specificity (%)
AUC: 67.5% (64.1%-70.9%)
KRAS
B
Sensitivity (%)
Specificity (%)
AUC: 89.3% (86.2%-92.4%)
CA19--9
C
Sensitivity (%)
Specificity (%)
AUC: 67.1% (61.9%-72.3%)
CEA
D
Sensitivity (%)
Specificity (%)
AUC: 68.5% (63.3%-73.7%)
HGF
E
Sensitivity (%)
Specificity (%)
AUC: 79.3% (74.8%-83.7%)
OPN
F
Sensitivity (%)
Specificity (%)
KRAS
CA19--9
CEA
HGF
OPN

FIG. 17

A

Tumor Type	SEER Incidence	Total #	# ctDNA Positive	# Protein Positive	# ctDNA + Protein Positive	% ctDNA Positive	% Protein Positive	% ctDNA + Protein Positive
Breast	43.8	150	24	30	47	16%	20%	31%
CRC	10	322	130	168	231	40%	52%	72%
Esophagus	0.63	43	21	28	34	49%	65%	79%
Gastric	1.34	65	20	38	39	31%	58%	60%
Liver	1.5	53	39	43	50	74%	81%	94%
Lung	3.31	109	43	30	61	39%	28%	56%
Ovarian	3.83	86	43	45	59	50%	52%	69%
Pancreas	3.8	412	91	248	271	22%	60%	66%
All Cancers	68.21	1240	411	630	792	33%	51%	64%
All Cancers (Incidence Weighted)	68.21	12239	3102.55	4341.11	5974.96	25%	35%	49%

B

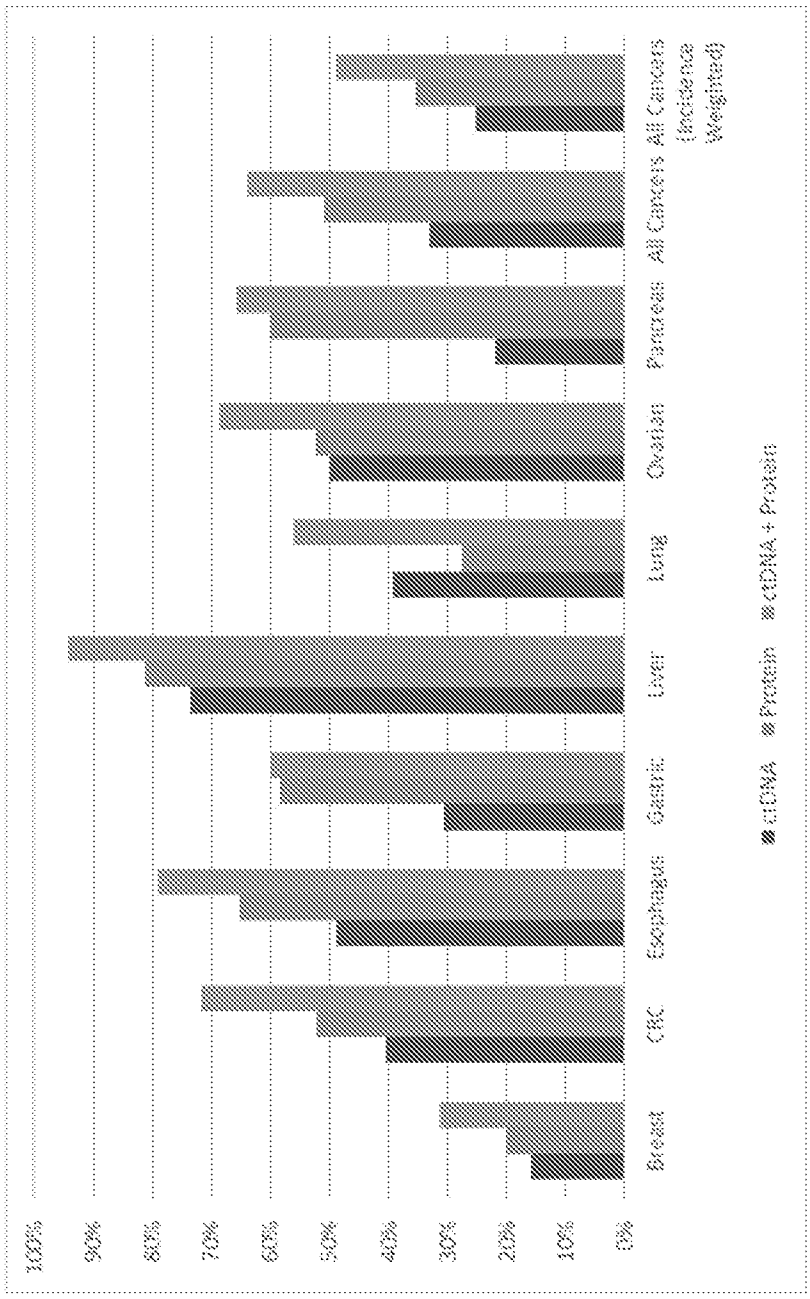


FIG. 18

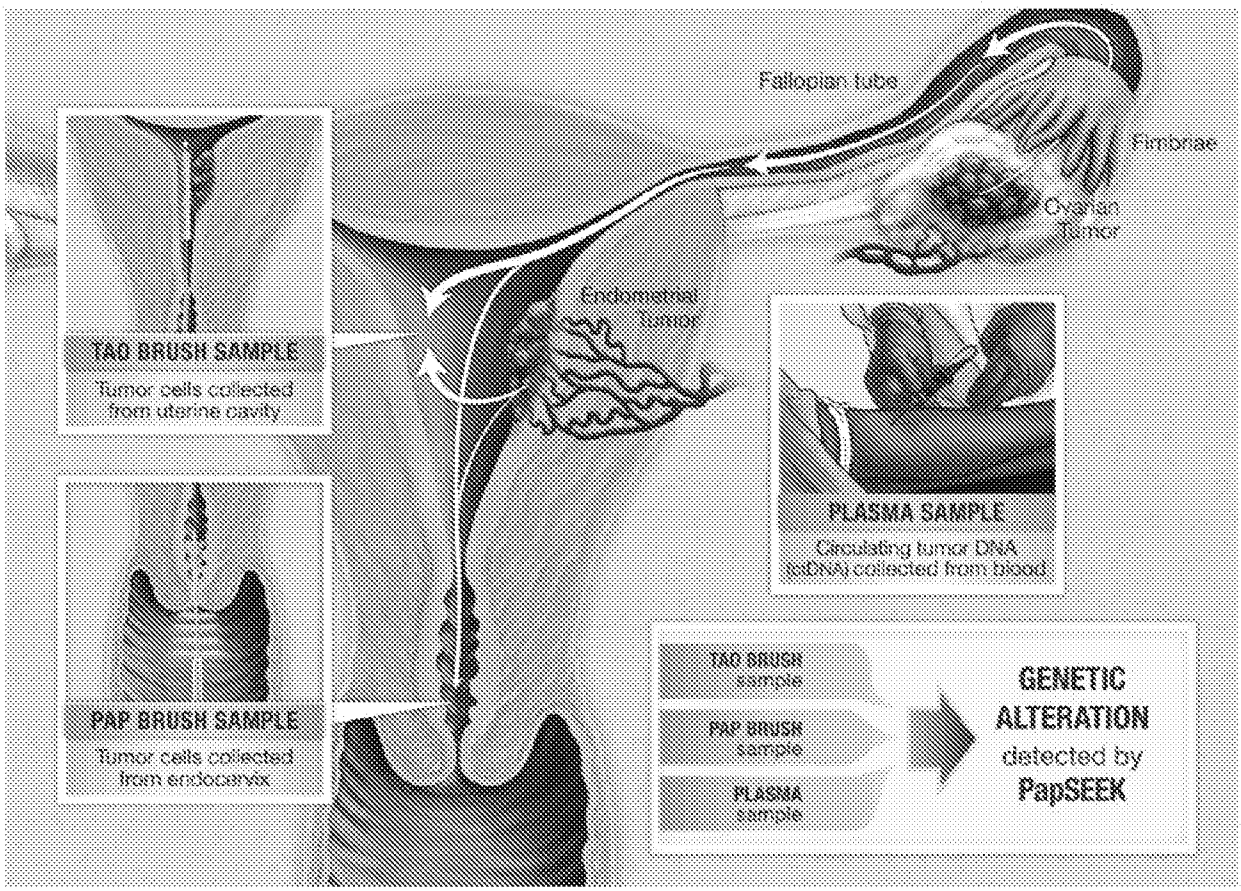

Fallopian tube
Fimbriae
Ovarian Tumor
Endometrial Tumor
TAO BRUSH SAMPLE
Tumor cells collected from uterine cavity
PAP BRUSH SAMPLE
Tumor cells collected from endocervix
PLASMA SAMPLE
Circulating tumor DNA (ctDNA) collected from blood
TAO BRUSH sample
PAP BRUSH sample
PLASMA sample
GENETIC ALTERATION detected by PapSEEK

FIG. 19

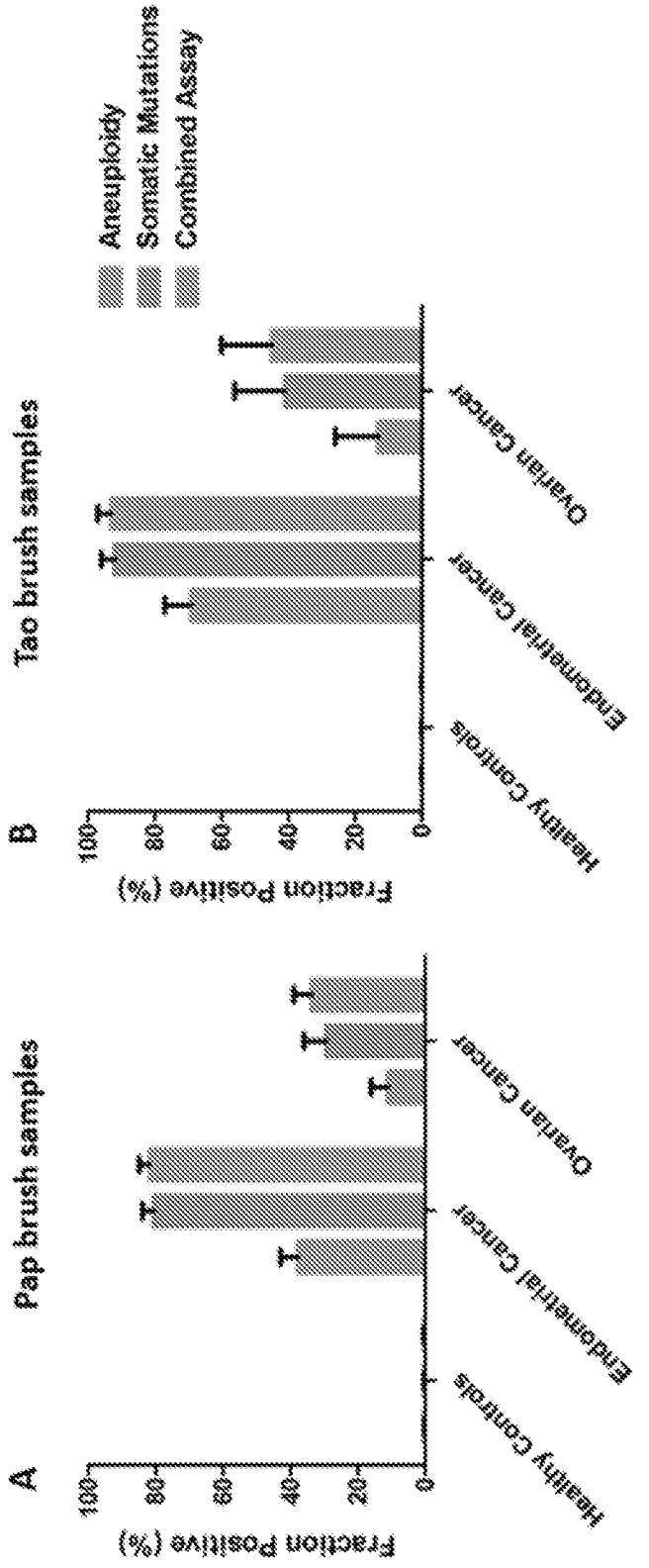

A
Pap brush samples
B
Tao brush samples
Aneuploidy
Somatic Mutations
Combined Assay
Fraction Positive (%)
100
80
60
40
20
0
Healthy Controls
Endometrial Cancer
Ovarian Cancer

FIG. 20

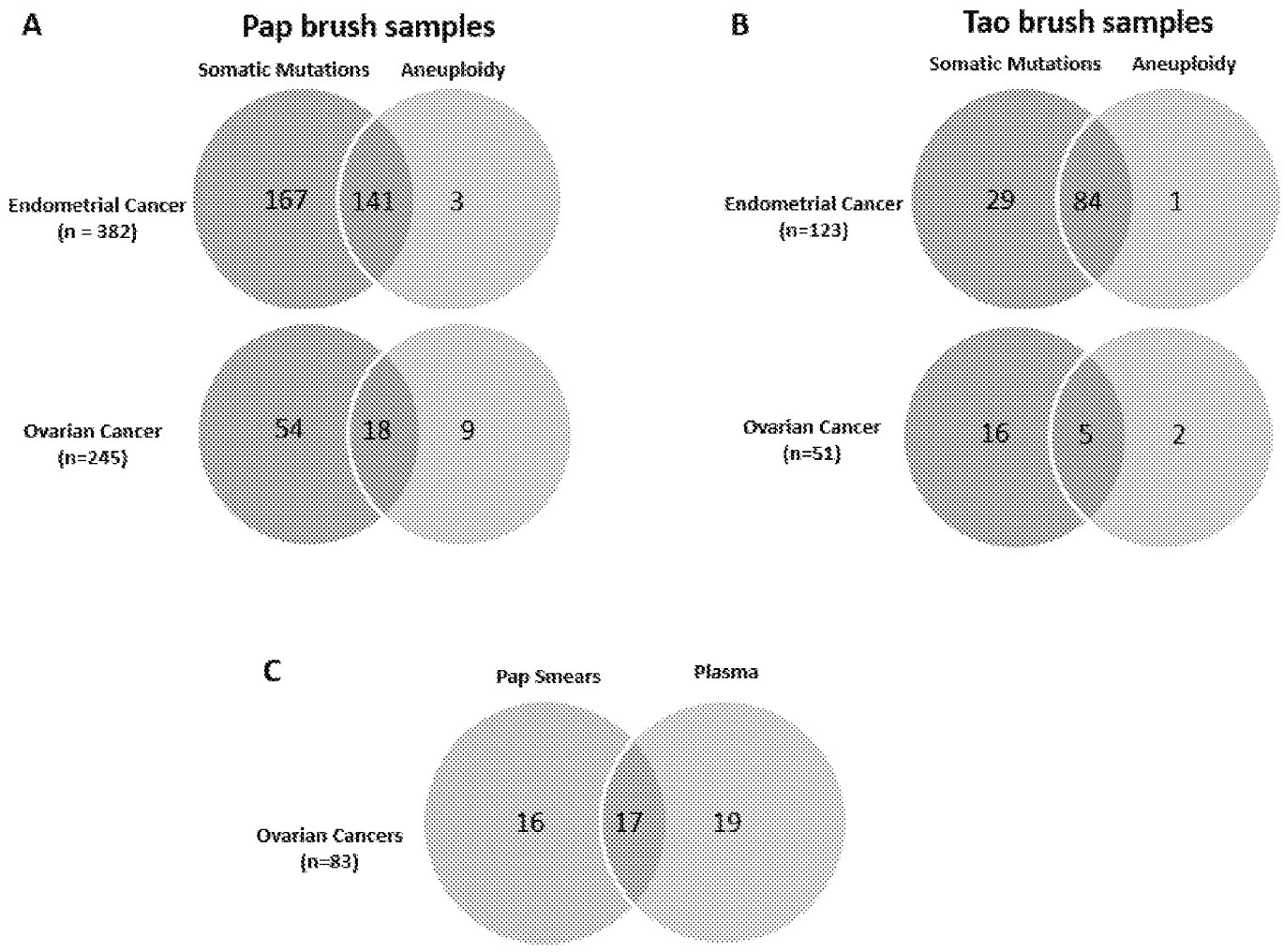

A
Pap brush samples
Somatic Mutations
Aneuploidy
Endometrial Cancer
(n = 382)
167
141
3
Ovarian Cancer
(n=245)
54
18
9
B
Tao brush samples
Somatic Mutations
Aneuploidy
Endometrial Cancer
(n=123)
29
84
1
Ovarian Cancer
(n=51)
16
5
2
C
Pap Smears
Plasma
Ovarian Cancers
(n=83)
16
17
19

FIG. 21

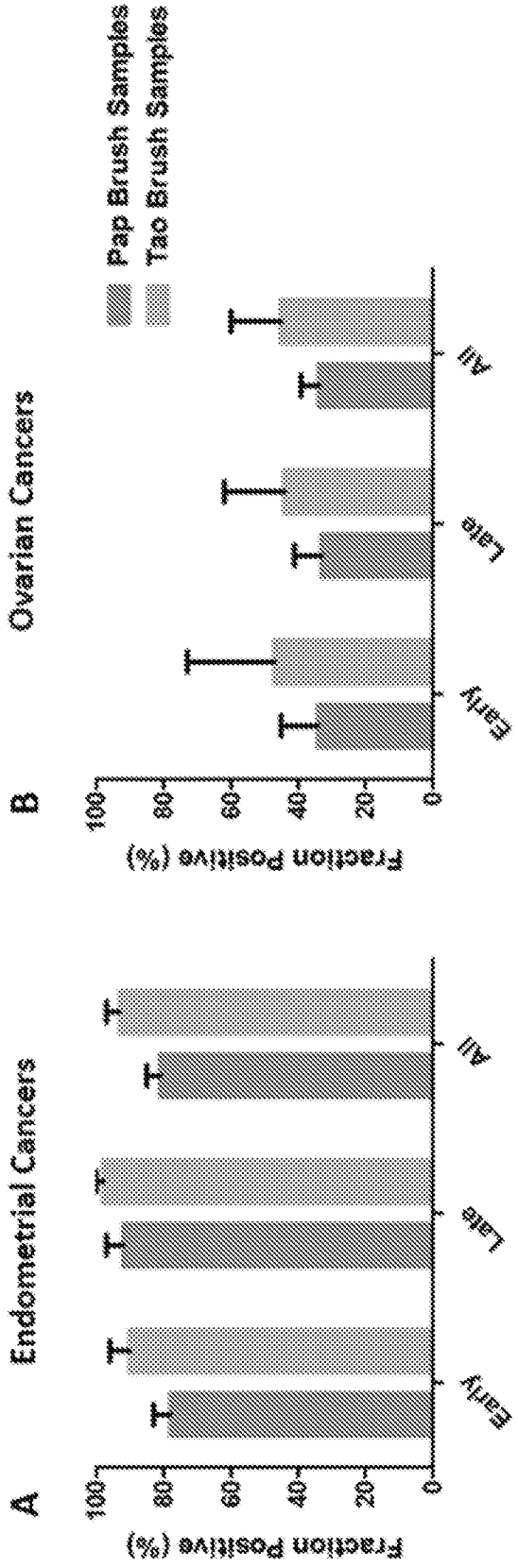

A
Endometrial Cancers
Fraction Positive (%)
100
80
60
40
20
0
Early
Late
All
B
Ovarian Cancers
Pap Brush Samples
Tao Brush Samples
Fraction Positive (%)
100
80
60
40
20
0
Early
Late
All

FIG. 22

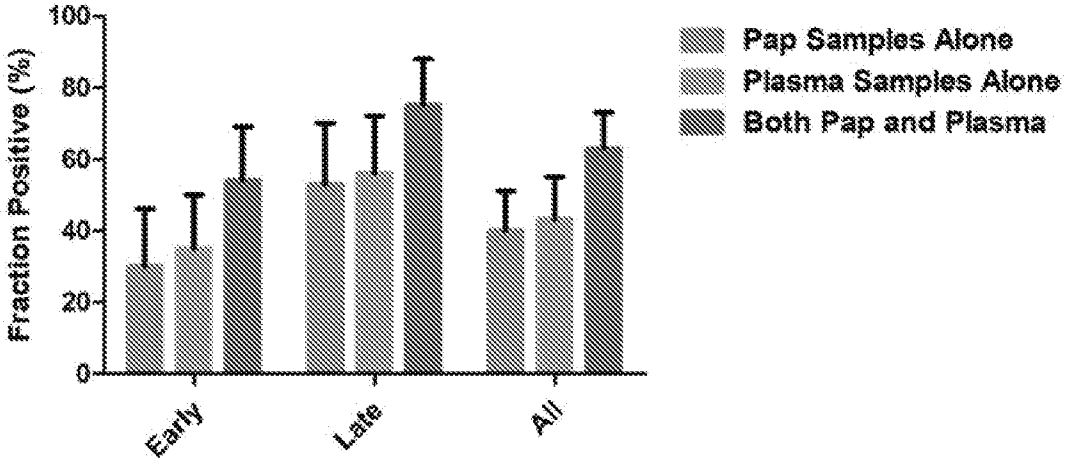

Fraction Positive (%)
100
80
60
40
20
0
Early
Late
All
Pap Samples Alone
Plasma Samples Alone
Both Pap and Plasma

FIG. 23

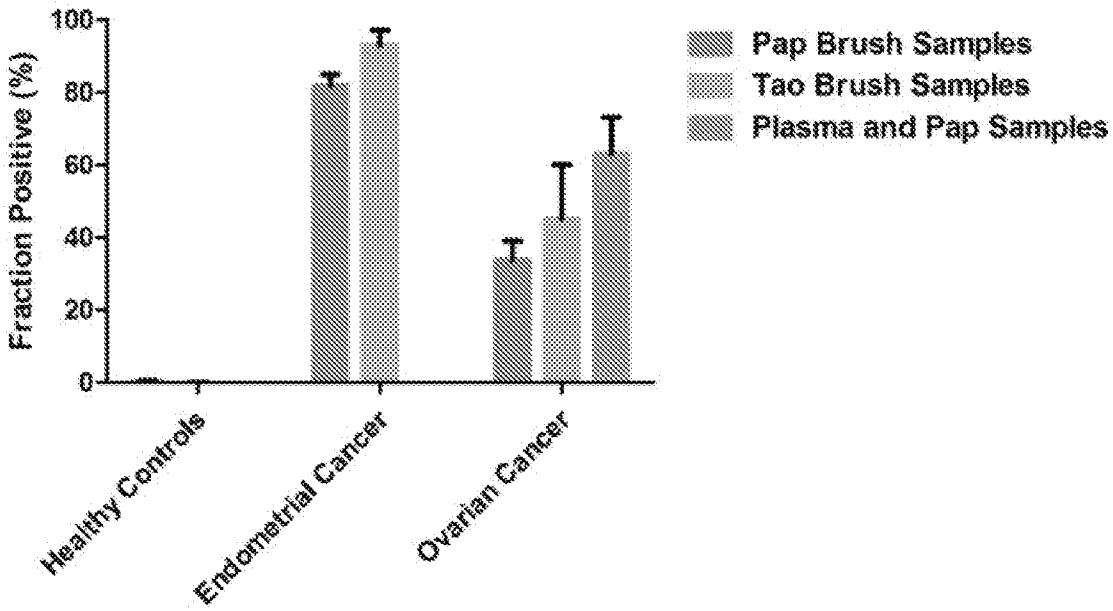

Fraction Positive (%)
100
80
60
40
20
0
Healthy Controls
Endometrial Cancer
Ovarian Cancer
Pap Brush Samples
Tao Brush Samples
Plasma and Pap Samples

FIG. 24

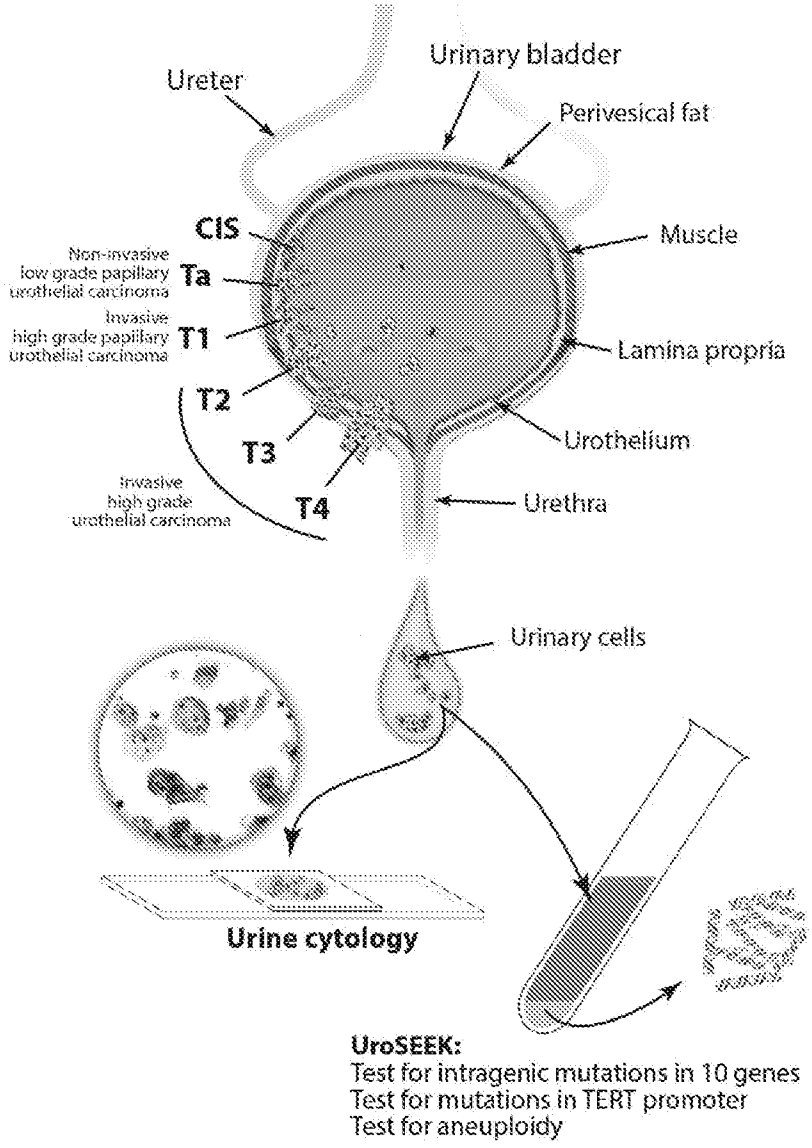

Ureter
Urinary bladder
Perivesical fat
CIS
Muscle
Non-invasive
low grade papillary
urothelial carcinoma
Ta
Invasive
high grade papillary
urothelial carcinoma
T1
Lamina propria
T2
T3
Urothelium
Invasive
high grade
urothelial carcinoma
T4
Urethra
Urinary cells
Urine cytology
UroSEEK:
Test for intragenic mutations in 10 genes
Test for mutations in TERT promoter
Test for aneuploidy

FIG. 25

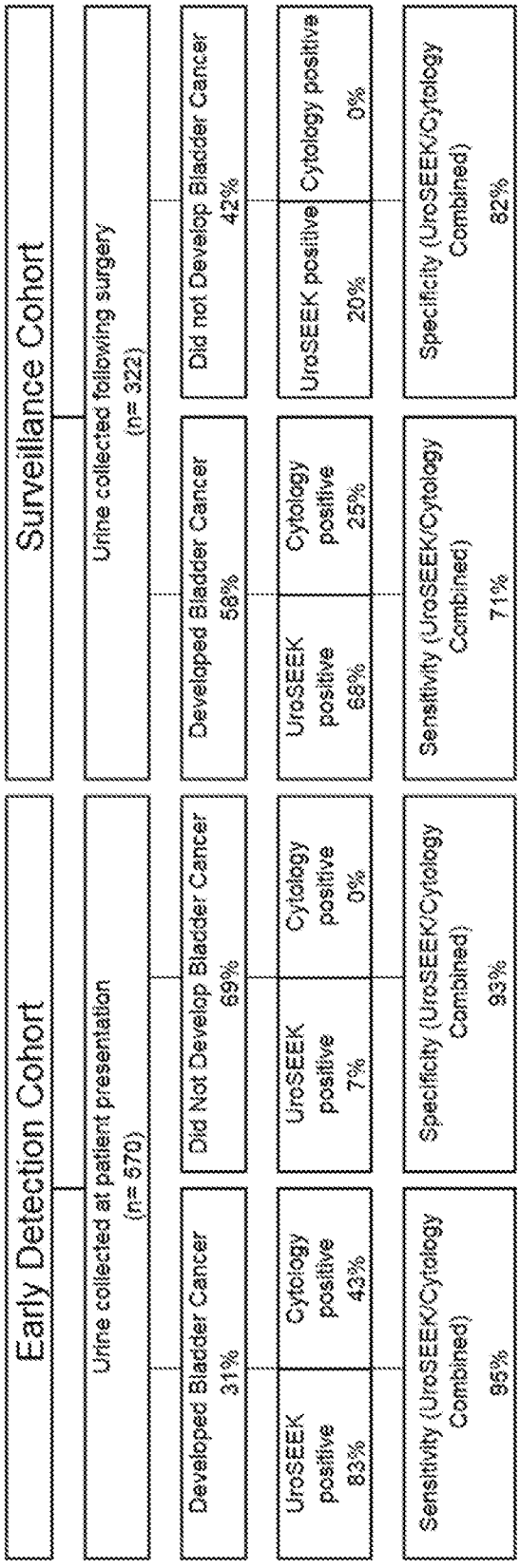

Early Detection Cohort
Urine collected at patient presentation (n= 370)
Developed Bladder Cancer 31%
UroSEEK positive 83%
Cytology positive 43%
Sensitivity (UroSEEK/Cytology Combined) 95%
Did Not Develop Bladder Cancer 69%
UroSEEK positive 7%
Cytology positive 0%
Specificity (UroSEEK/Cytology Combined) 93%
Surveillance Cohort
Urine collected following surgery (n= 322)
Developed Bladder Cancer 58%
UroSEEK positive 68%
Cytology positive 25%
Sensitivity (UroSEEK/Cytology Combined) 71%
Did not Develop Bladder Cancer 42%
UroSEEK positive 20%
Cytology positive 0%
Specificity (UroSEEK/Cytology Combined) 82%

FIG. 26

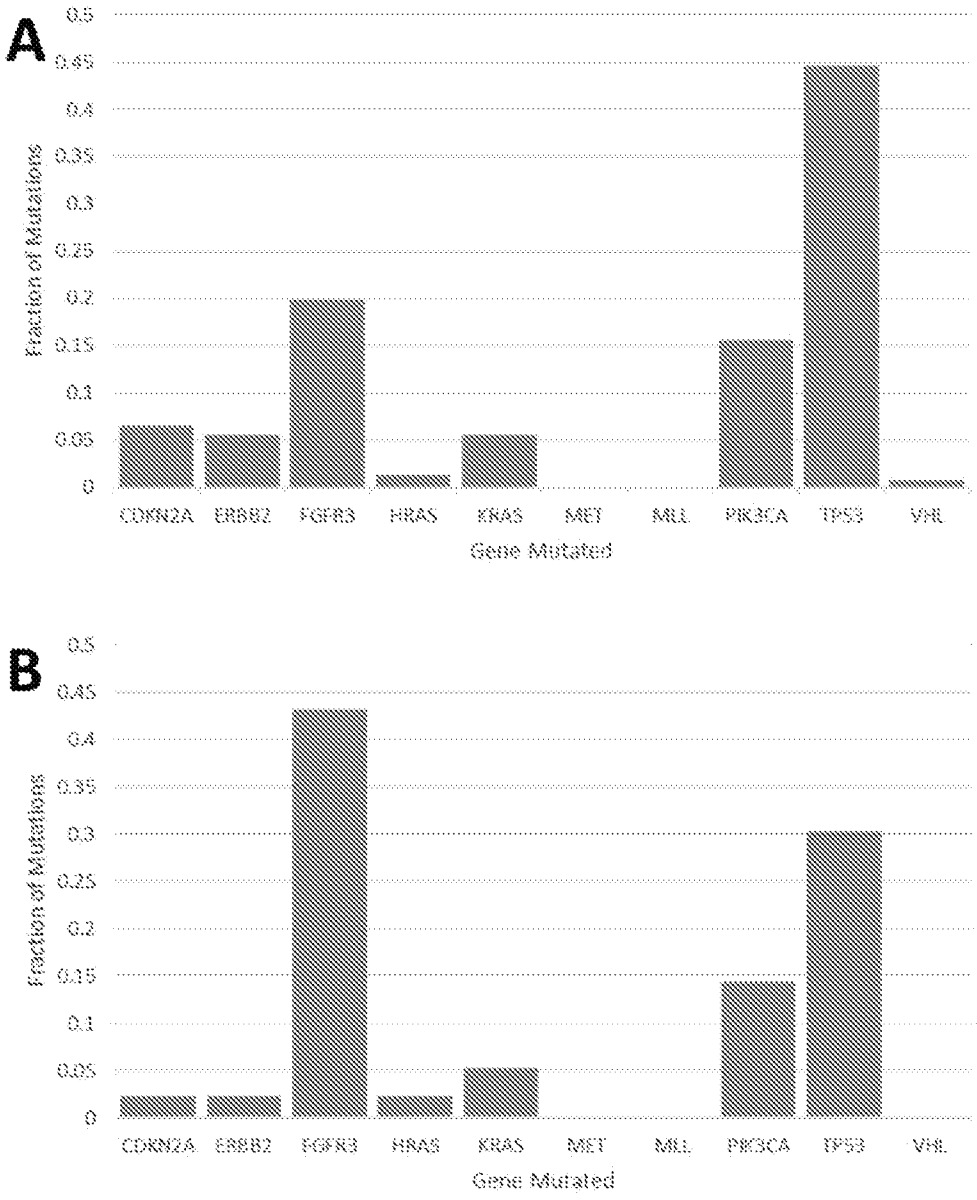

A
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Fraction of Mutations
CDKN2A
ERBB2
FGFR3
HRAS
KRAS
MET
MLL
PIK3CA
TP53
VHL
Gene Mutated
B
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Fraction of Mutations
CDKN2A
ERBB2
FGFR3
HRAS
KRAS
MET
MLL
PIK3CA
TP53
VHL
Gene Mutated

FIG. 27

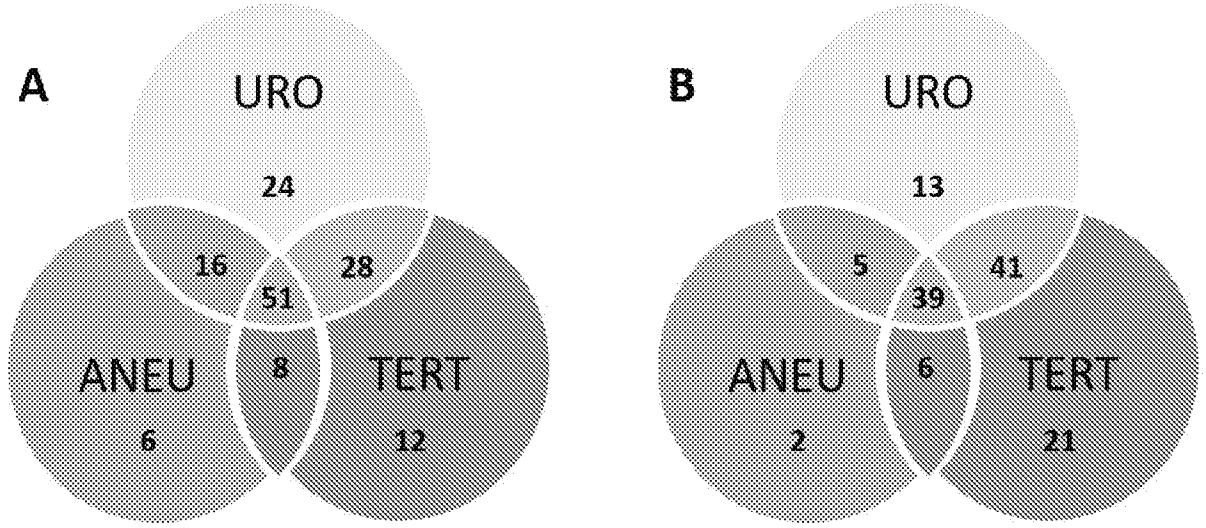


FIG. 28

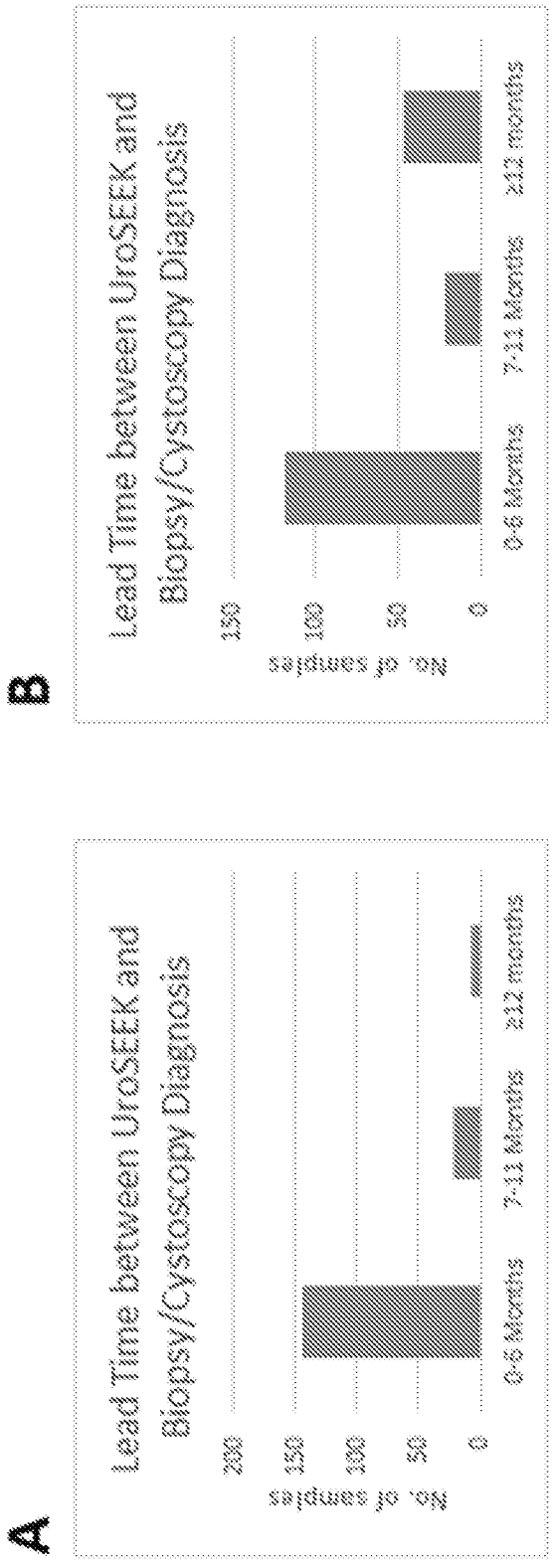

A
Lead Time between UroSEEK and Biopsy/Cystoscopy Diagnosis
No. of samples
200
150
100
50
0
0-6 Months
7-11 Months
≥12 months
B
Lead Time between UroSEEK and Biopsy/Cystoscopy Diagnosis
No. of samples
150
100
50
0
0-6 Months
7-11 Months
≥12 months

FIG. 29

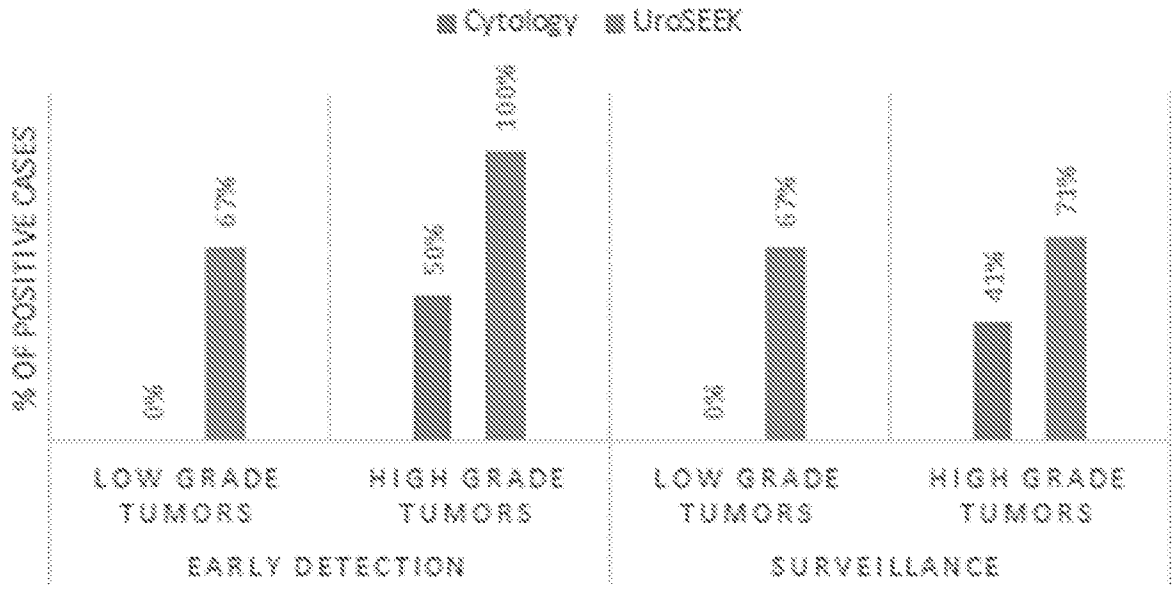

Cytology
UroSEEK
% OF POSITIVE CASES
0%
67%
50%
100%
0%
67%
41%
71%
LOW GRADE TUMORS
HIGH GRADE TUMORS
LOW GRADE TUMORS
HIGH GRADE TUMORS
EARLY DETECTION
SURVEILLANCE

FIG. 30

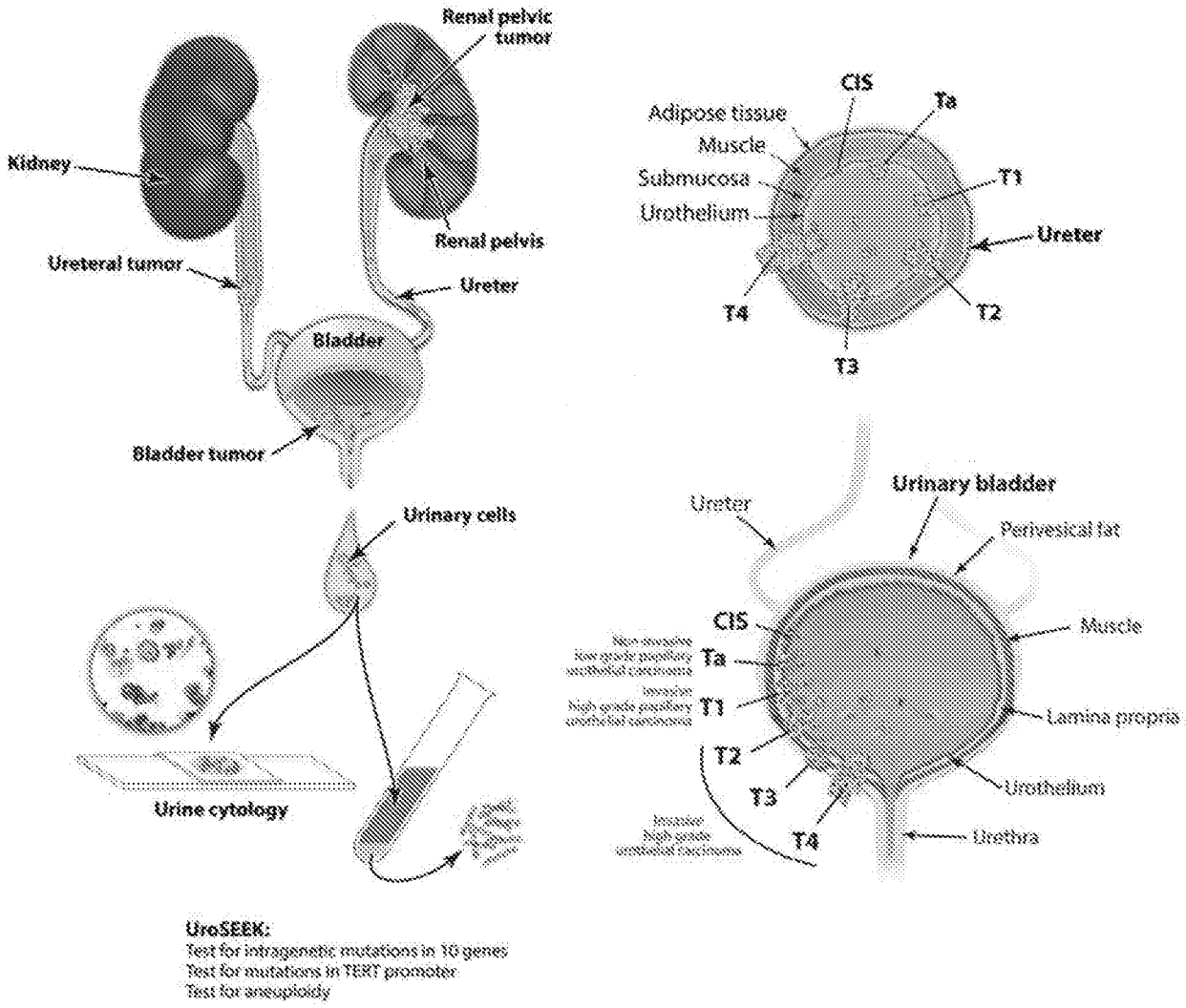

Renal pelvic tumor
Kidney
Renal pelvis
Ureteral tumor
Ureter
Bladder
Bladder tumor
Urinary cells
Urine cytology
UroSEEK:
Test for intragenetic mutations in 10 genes
Test for mutations in TERT promoter
Test for aneuploidy
CIS
Ta
Adipose tissue
Muscle
Submucosa
Urothelium
T1
Ureter
T4
T2
T3
Urinary bladder
Ureter
Perivesical fat
CIS
Muscle
Ta
T1
Lamina propria
T2
T3
Urothelium
T4
Urethra

FIG. 31

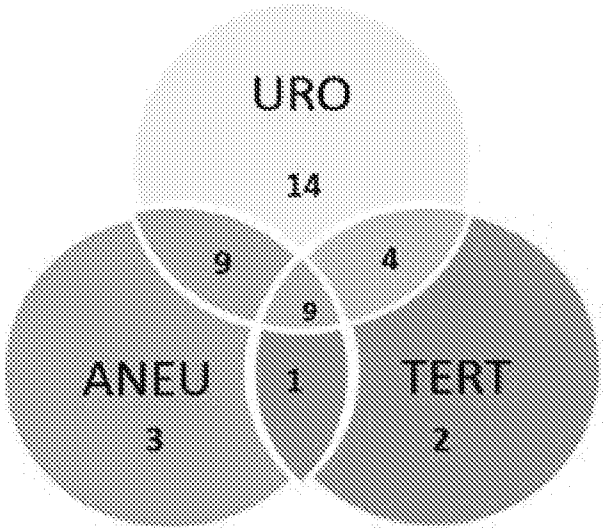


FIG. 32

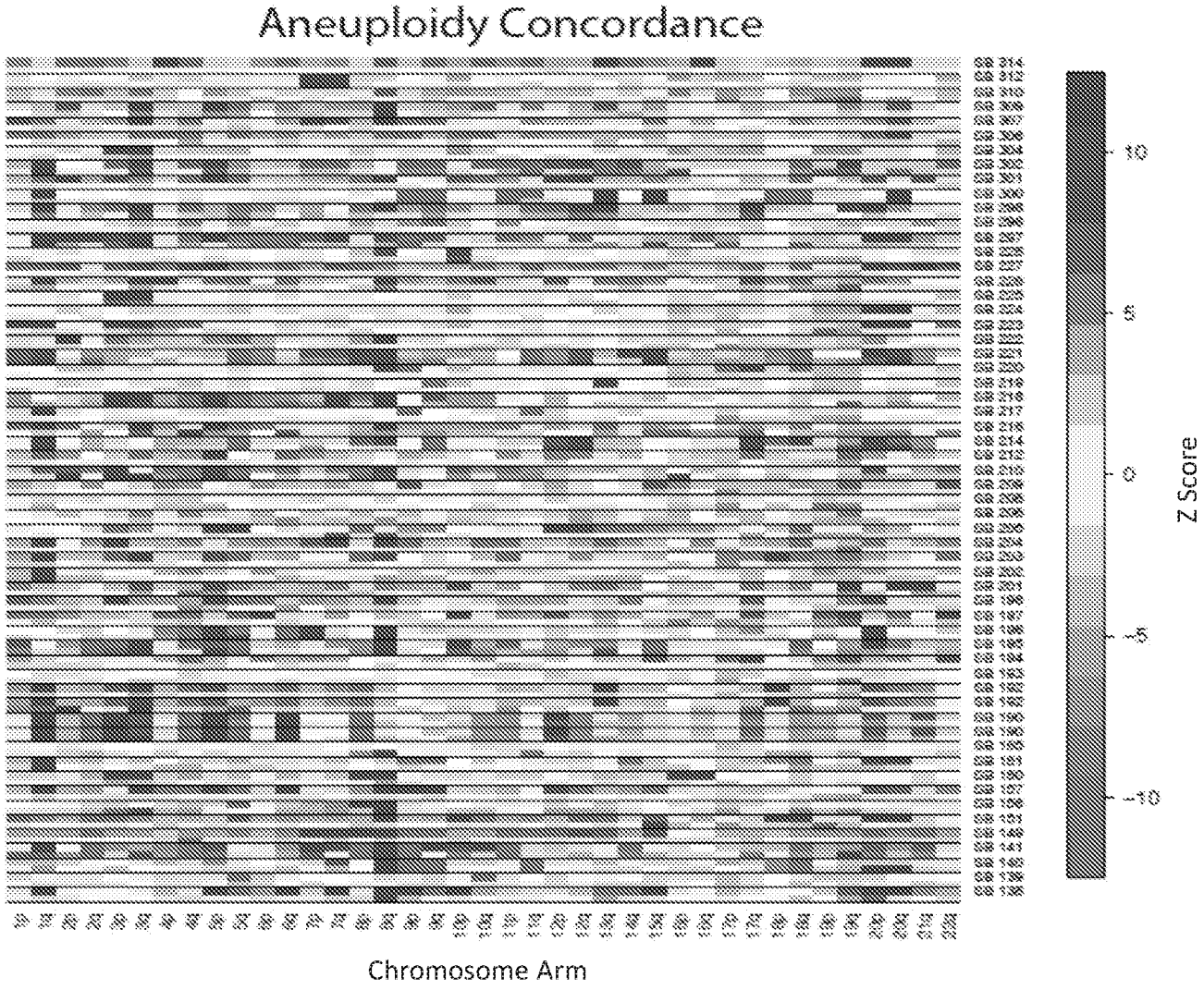


FIG. 33

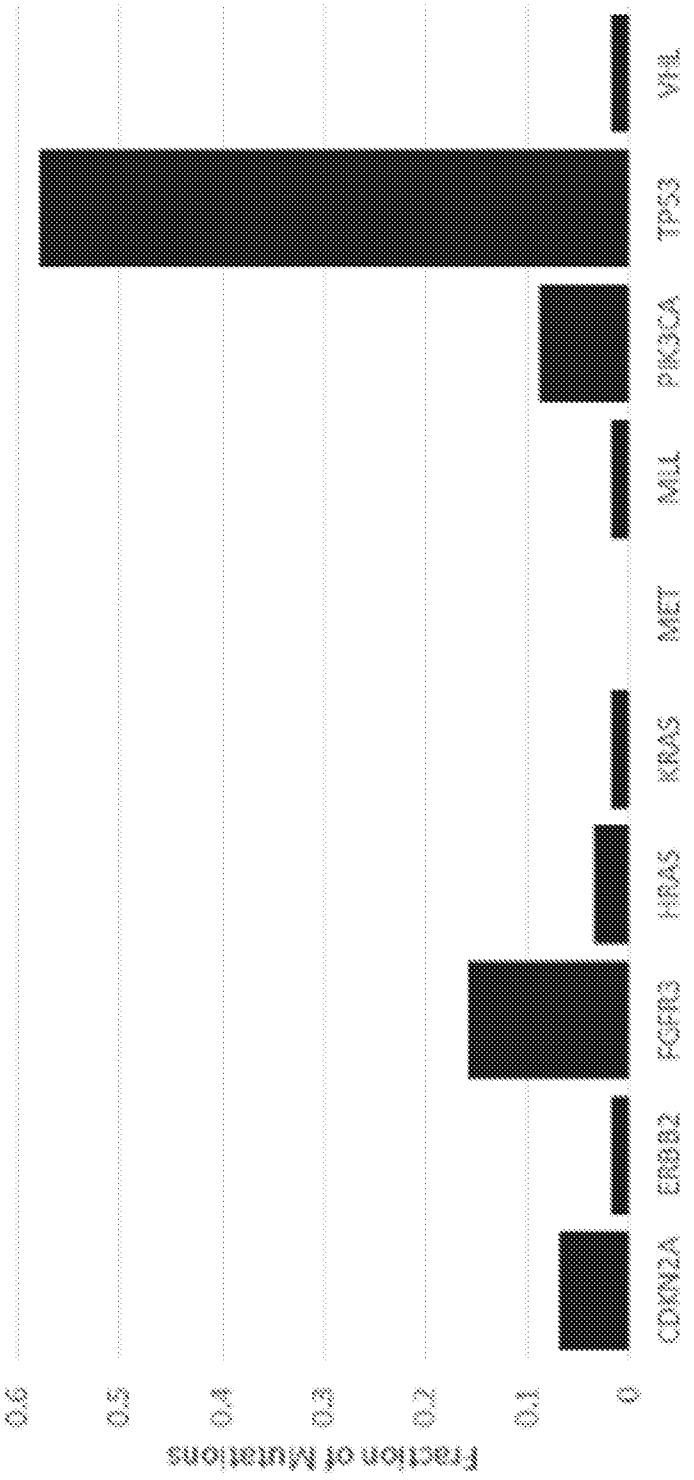

Fraction of Mutations
0.6
0.5
0.4
0.3
0.2
0.1
0
COL2A
ERBB2
FGFR3
HRAS
KRAS
MET
MLL
PIK3CA
TP53
VHL
Gene Mutated

FIG. 34

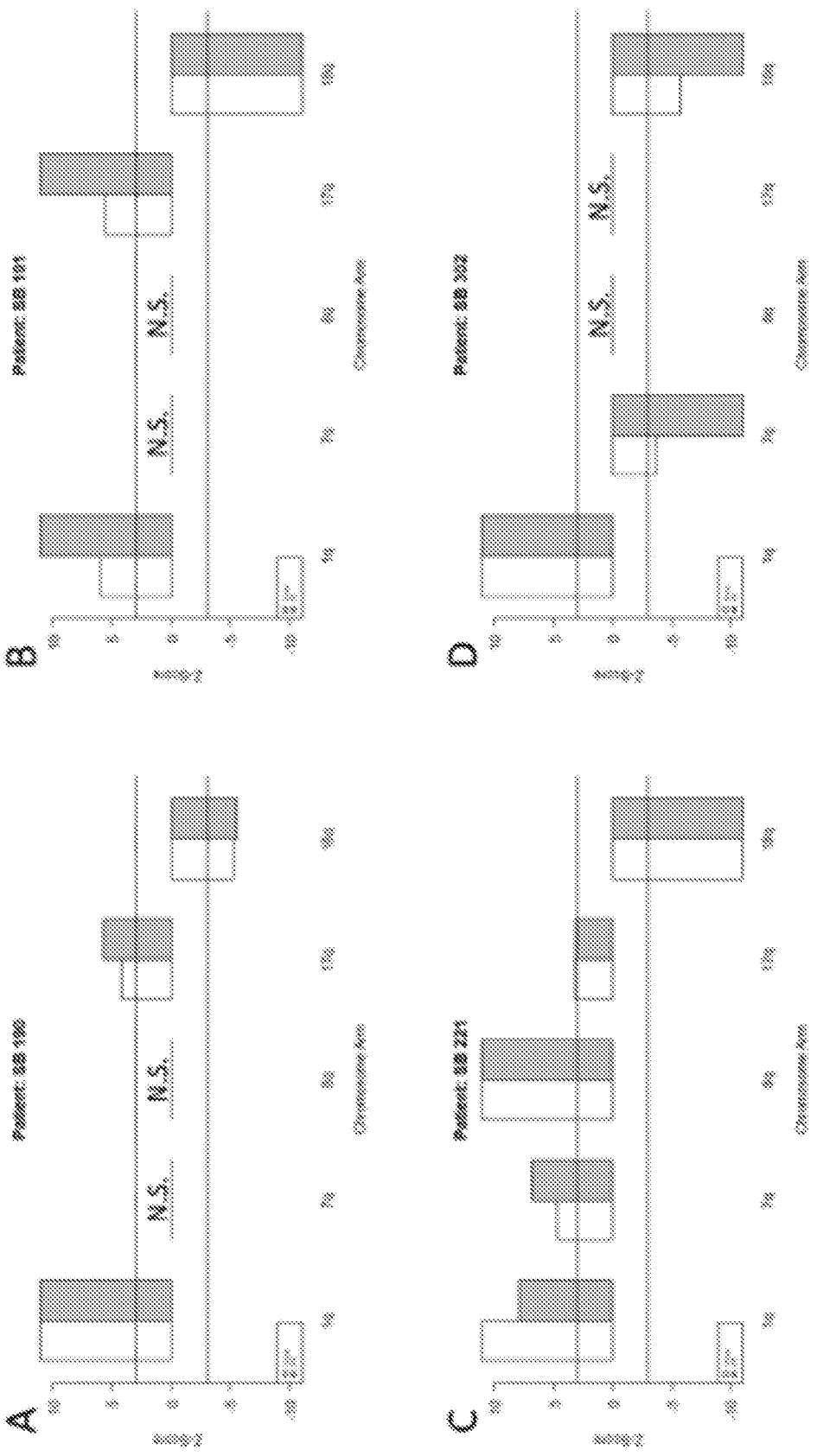


FIG. 35

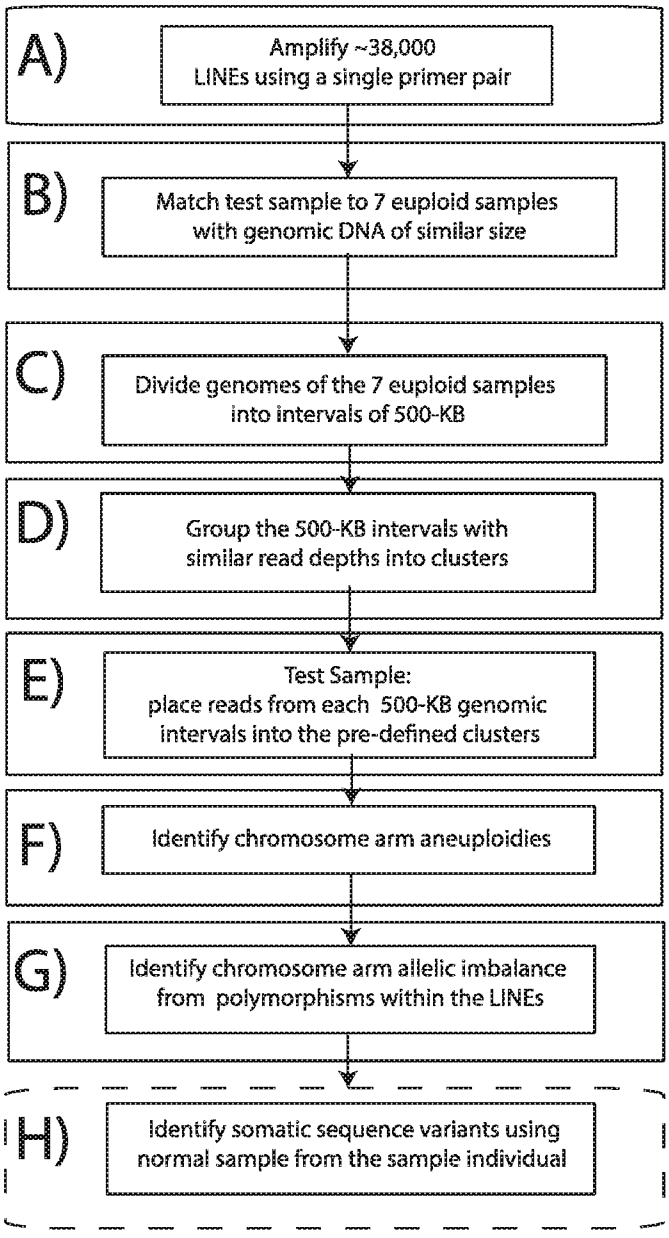

A)
Amplify ~38,000 LINEs using a single primer pair
B)
Match test sample to 7 euploid samples with genomic DNA of similar size
C)
Divide genomes of the 7 euploid samples into intervals of 500-KB
D)
Group the 500-KB intervals with similar read depths into clusters
E)
Test Sample: place reads from each 500-KB genomic intervals into the pre-defined clusters
F)
Identify chromosome arm aneuploidies
G)
Identify chromosome arm allelic imbalance from polymorphisms within the LINEs
H)
Identify somatic sequence variants using normal sample from the sample individual

FIG. 36

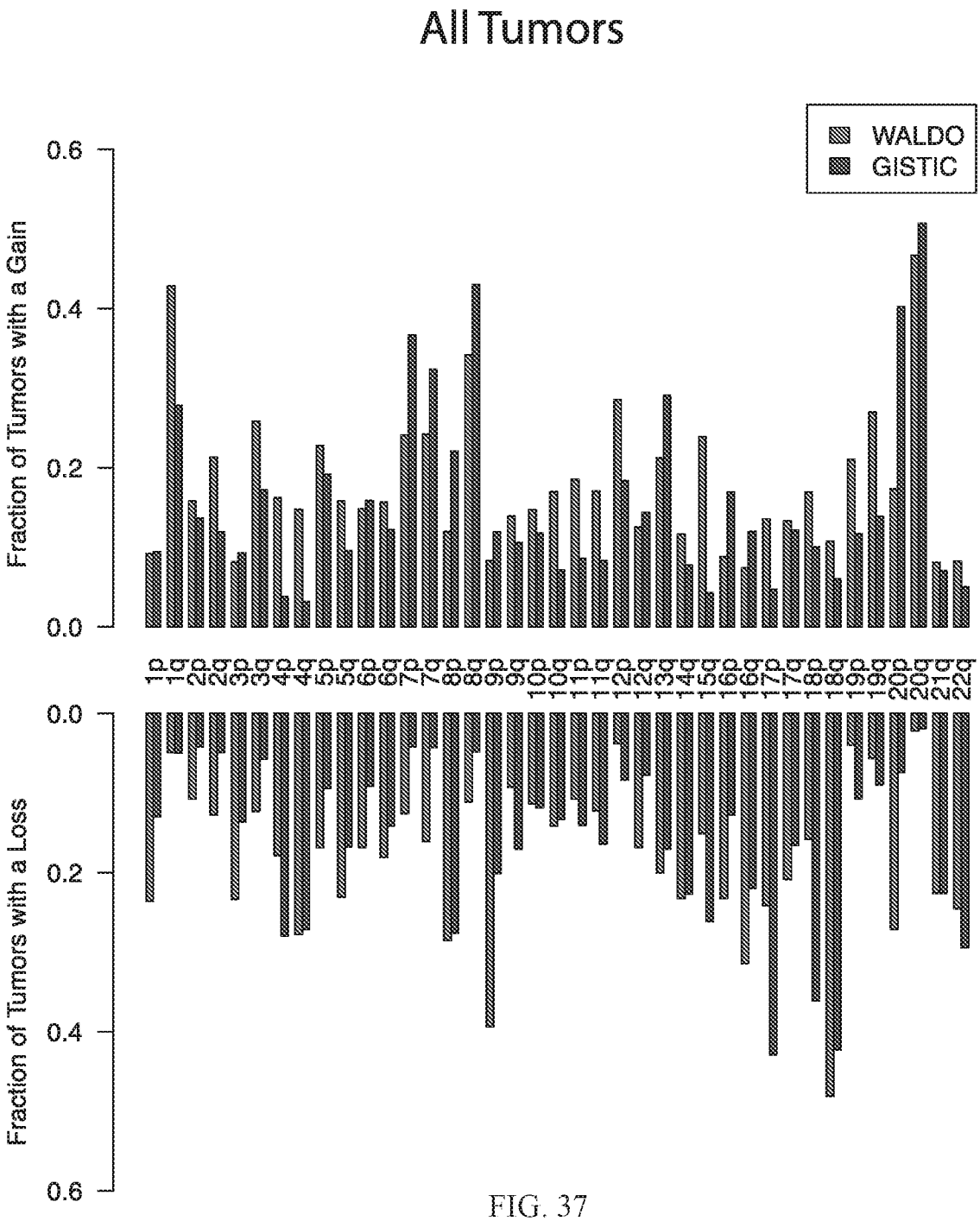

All Tumors
WALDO
GISTIC
Fraction of Tumors with a Gain
0.6
0.4
0.2
0.0
Fraction of Tumors with a Loss
0.0
0.2
0.4
0.6
1p
1q
2p
2q
3p
3q
4p
4q
5p
5q
6p
6q
7p
7q
8p
8q
9p
9q
10p
10q
11p
11q
12p
12q
13q
14q
15q
16p
16q
17p
17q
18p
18q
19p
19q
20p
20q
21q
22q

FIG. 37

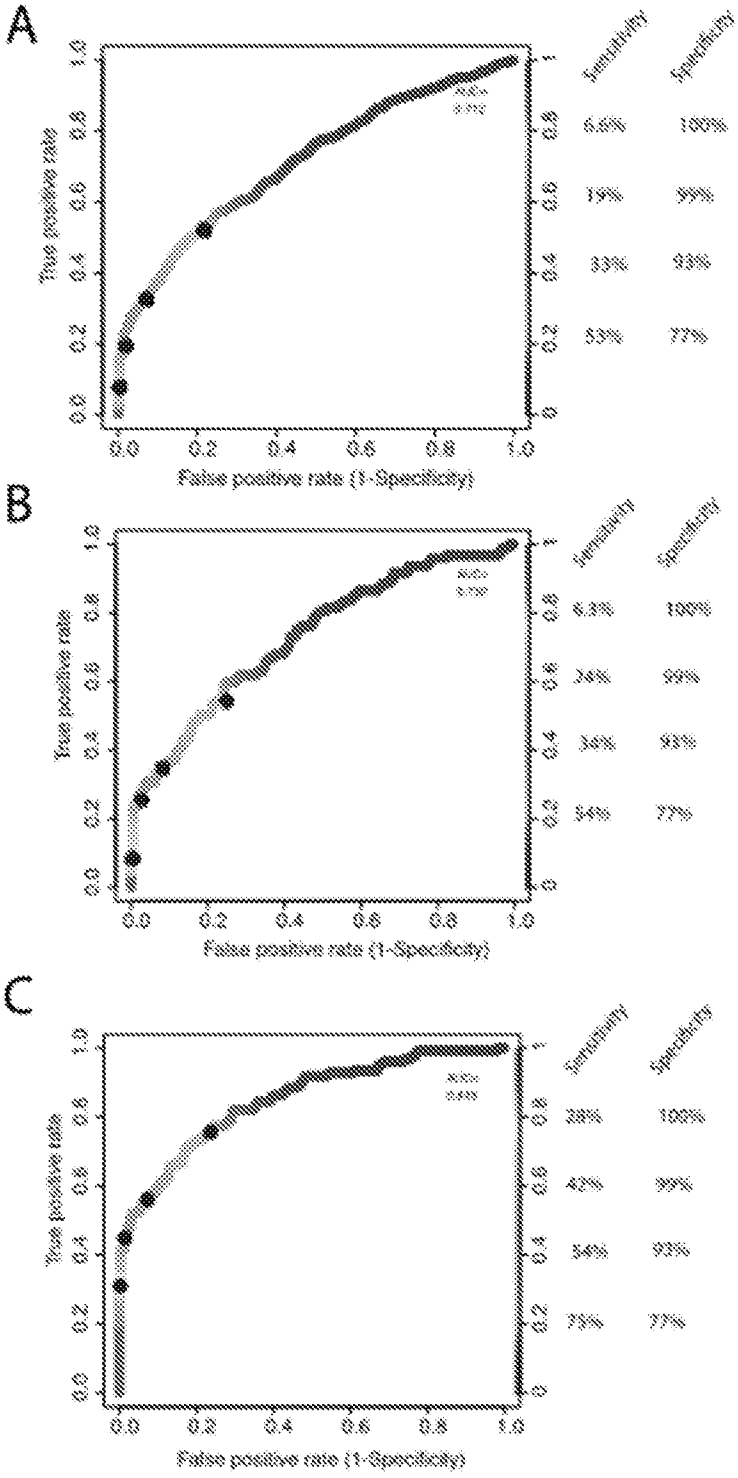

A
True positive rate
False positive rate (1-Specificity)
Sensitivity
Specificity
6.6%
100%
19%
99%
33%
93%
53%
77%
B
True positive rate
False positive rate (1-Specificity)
Sensitivity
Specificity
6.3%
100%
24%
99%
34%
93%
54%
77%
C
True positive rate
False positive rate (1-Specificity)
Sensitivity
Specificity
28%
100%
42%
99%
54%
93%
75%
77%

FIG. 38

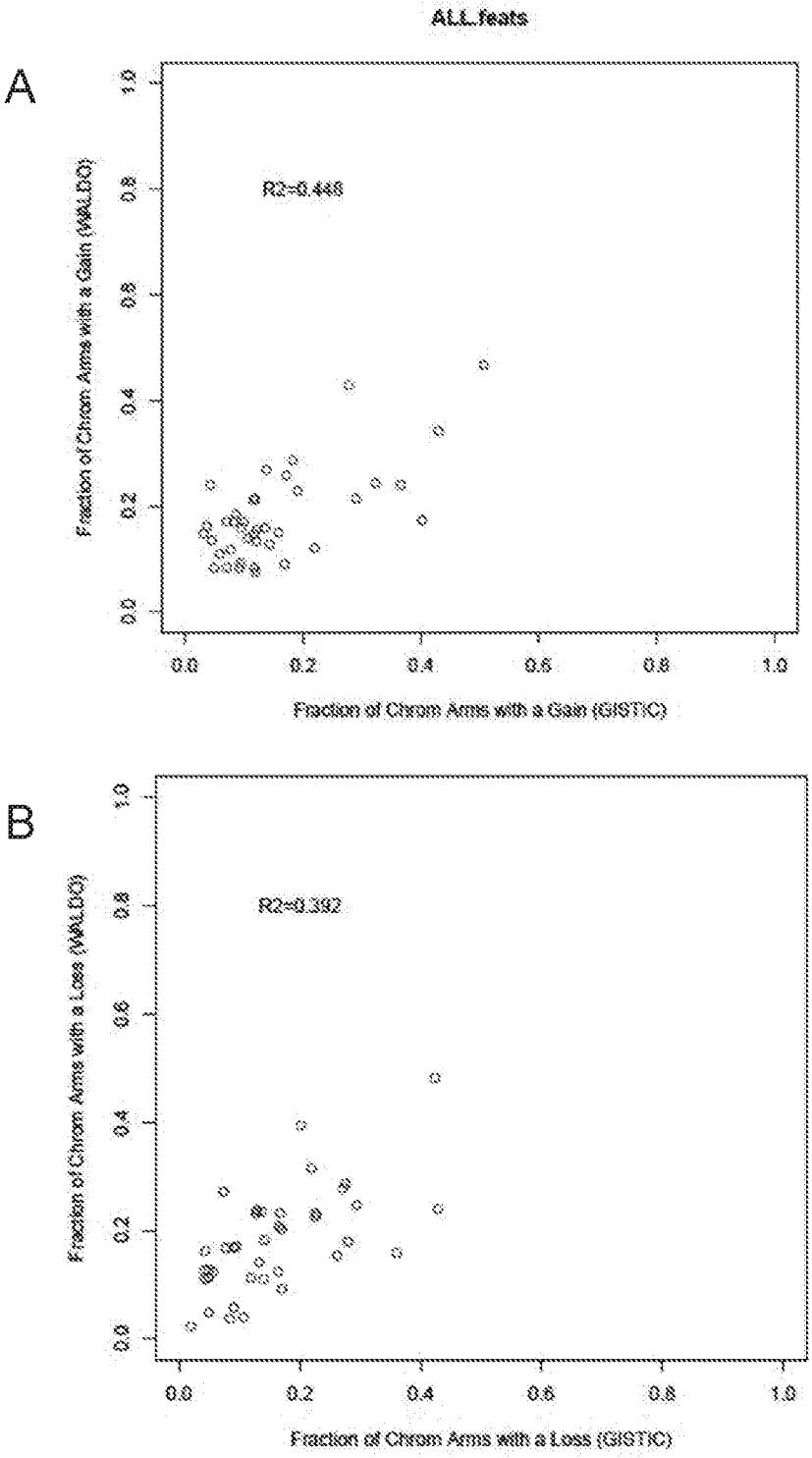

A
ALL.feats
R2=0.446
Fraction of Chrom Arms with a Gain (WALDO)
Fraction of Chrom Arms with a Gain (GISTIC)
0.0
0.2
0.4
0.6
0.8
1.0
B
R2=0.392
Fraction of Chrom Arms with a Loss (WALDO)
Fraction of Chrom Arms with a Loss (GISTIC)

FIG. 39

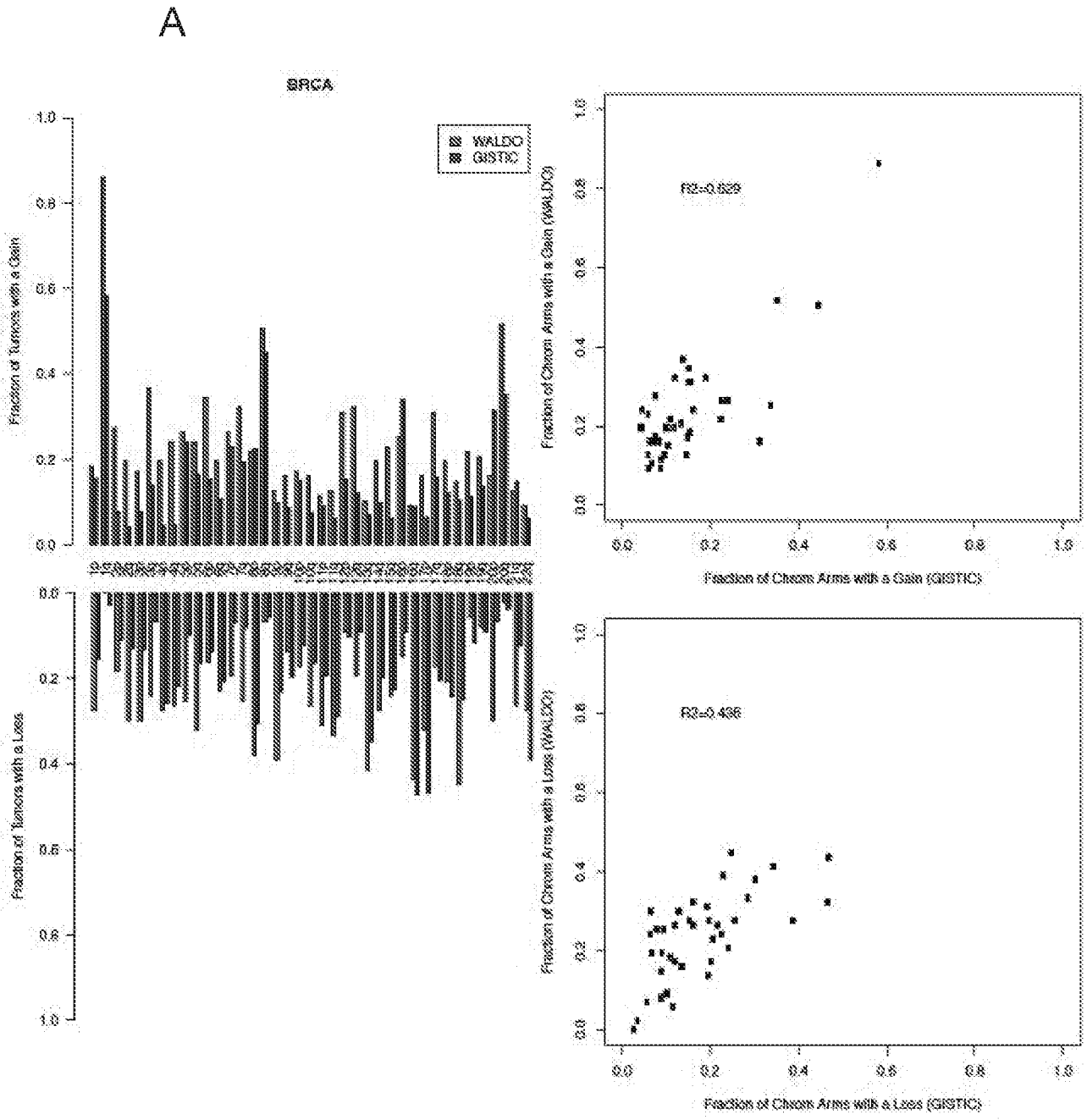

A
BRCA
WALDO
GISTIC
Fraction of Tumors with a Gain
Fraction of Tumors with a Loss
Fraction of Chrom Arms with a Gain (WALDO)
Fraction of Chrom Arms with a Gain (GISTIC)
R2=0.629
Fraction of Chrom Arms with a Loss (WALDO)
Fraction of Chrom Arms with a Loss (GISTIC)
R2=0.496

FIG. 40

B

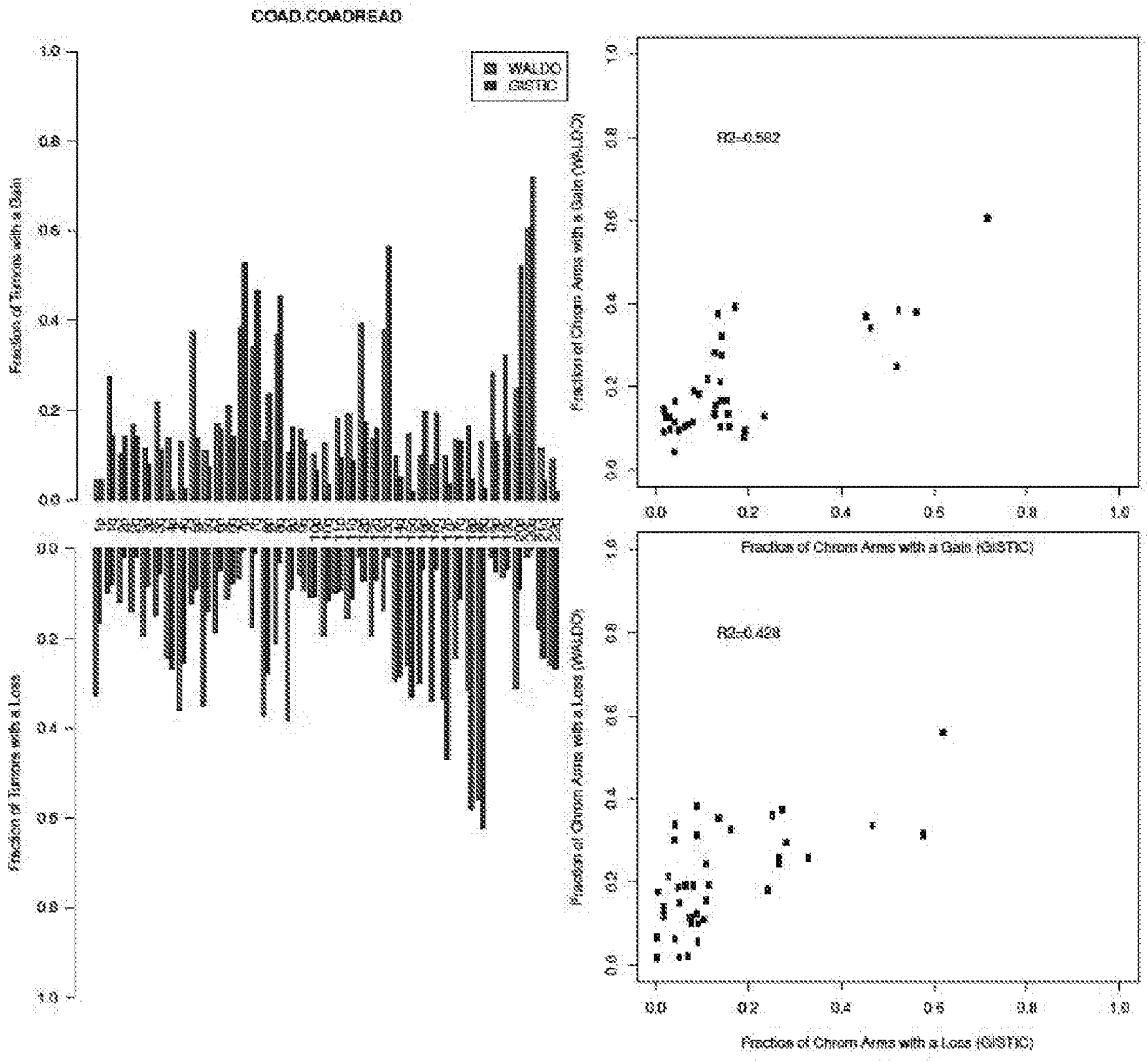


FIG. 40 (continued)

C

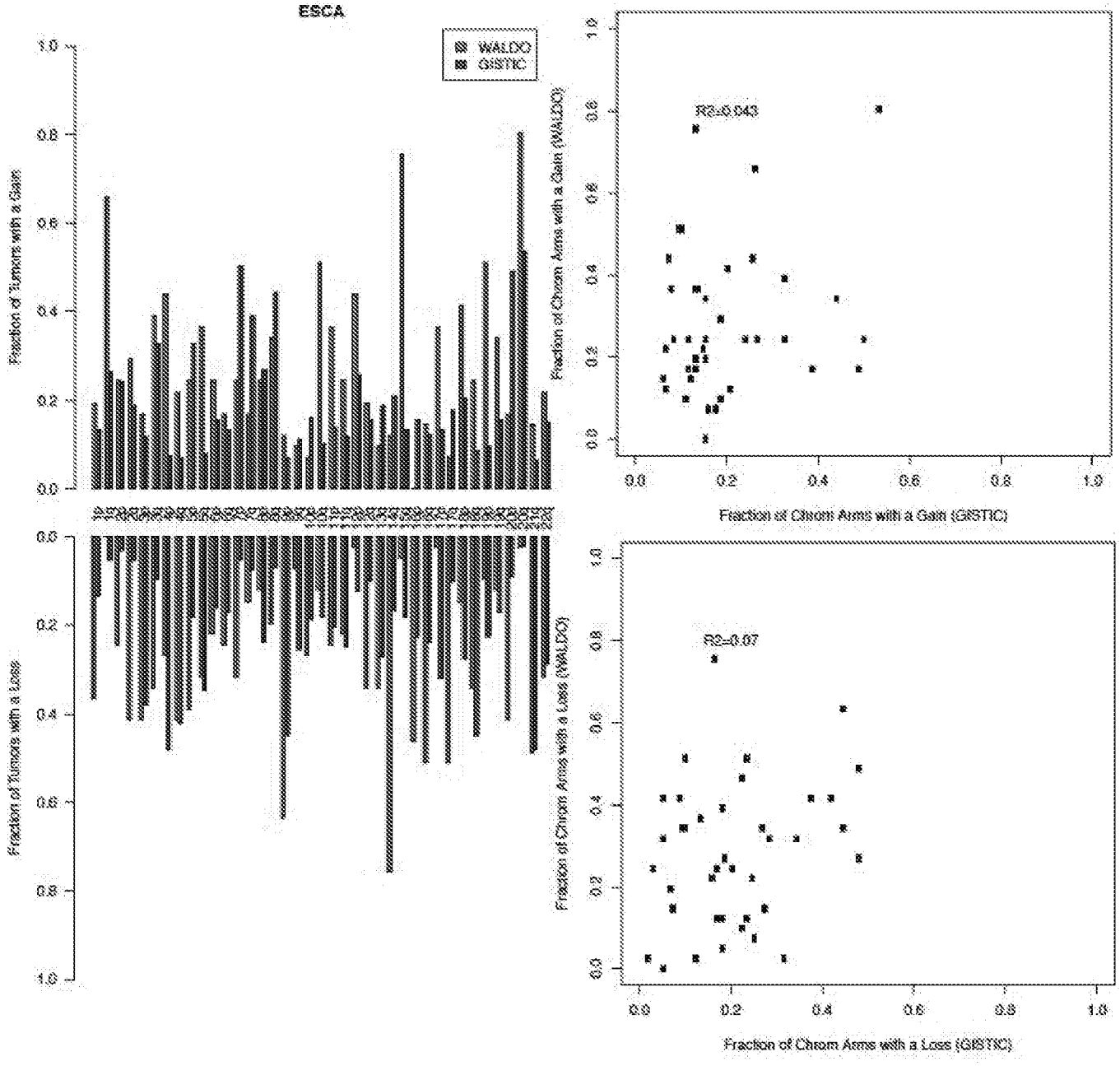


FIG. 40 (continued)

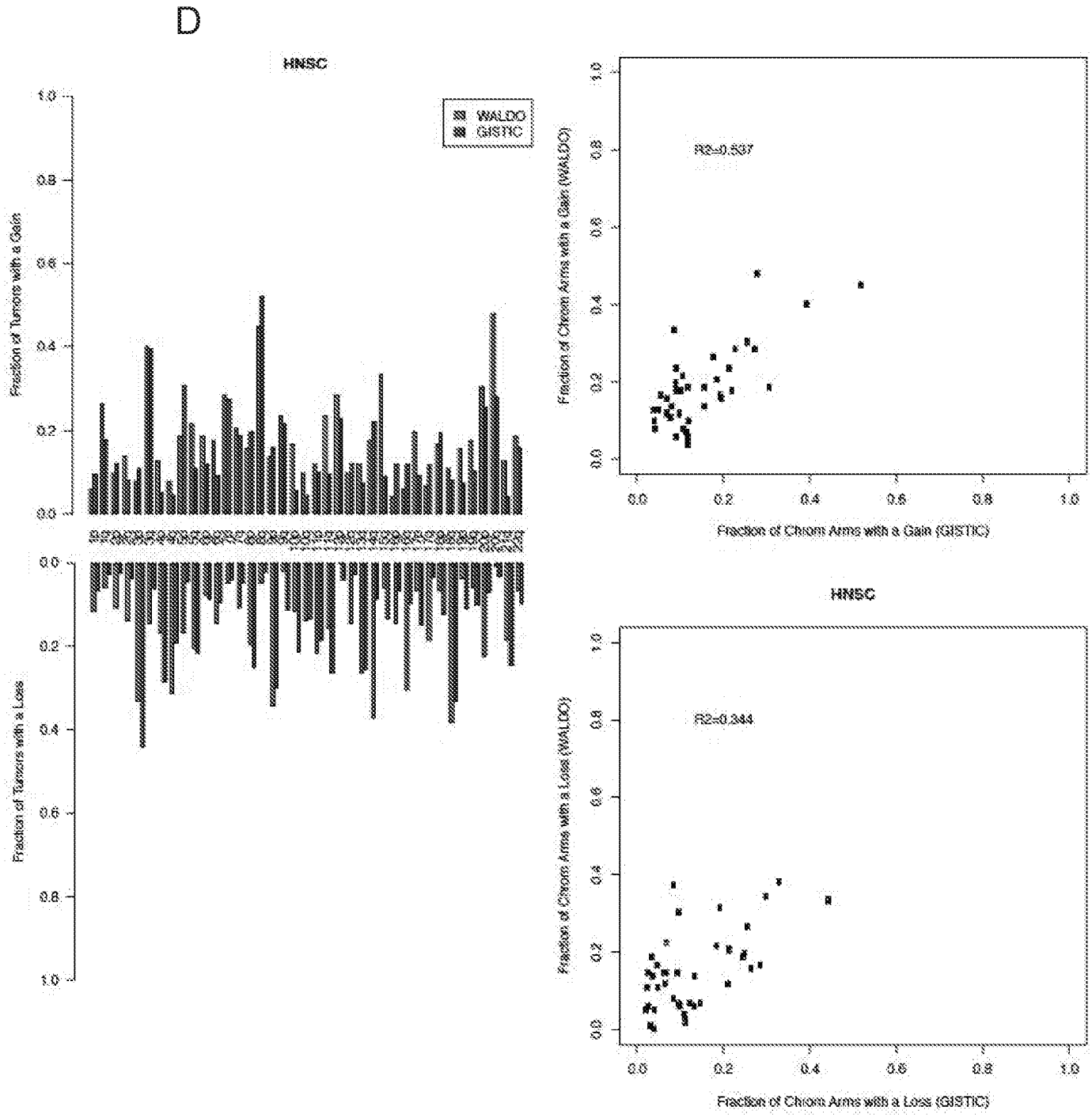


FIG. 40 (continued)

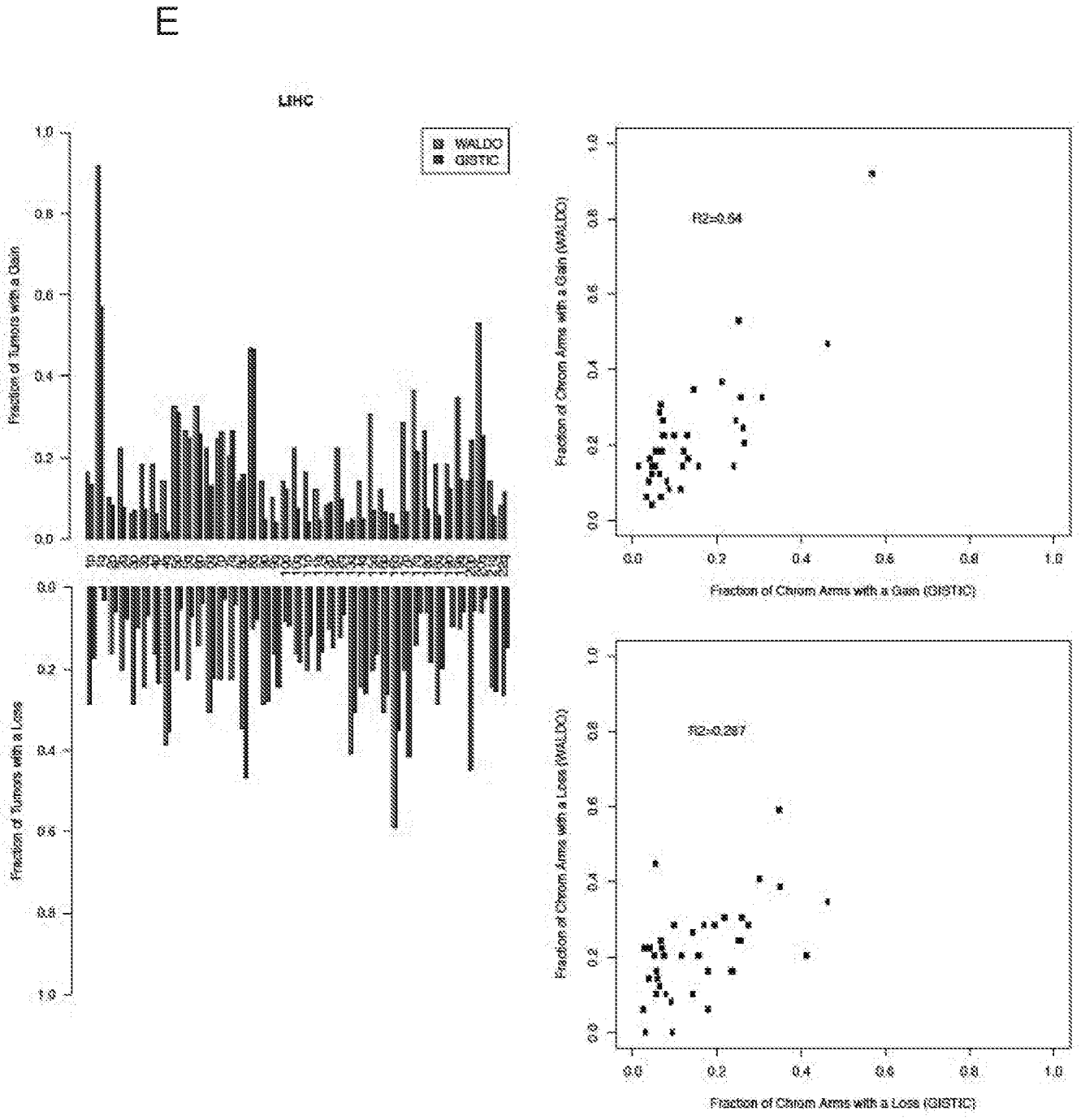


FIG. 40 (continued)

F

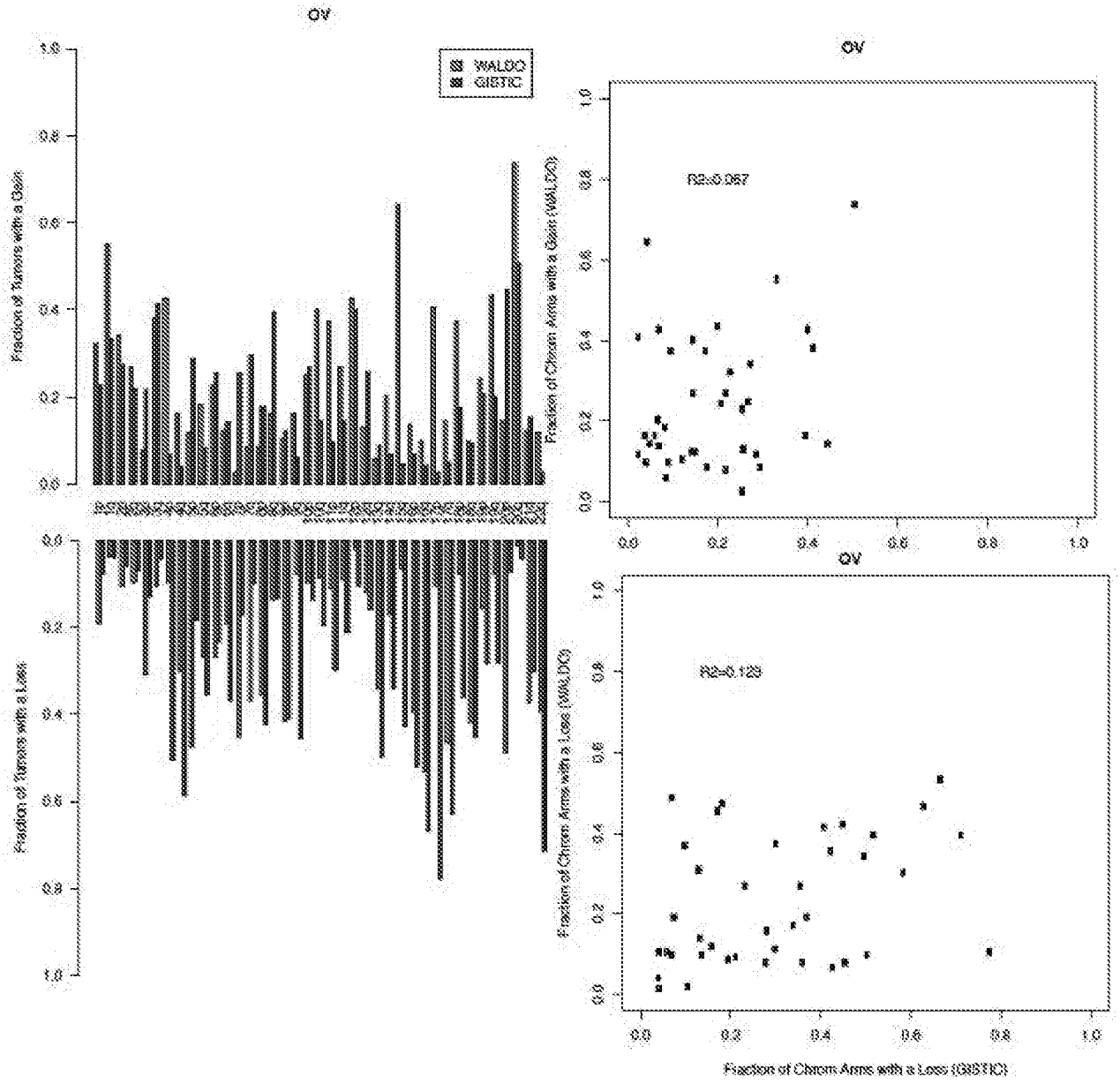


FIG. 40 (continued)

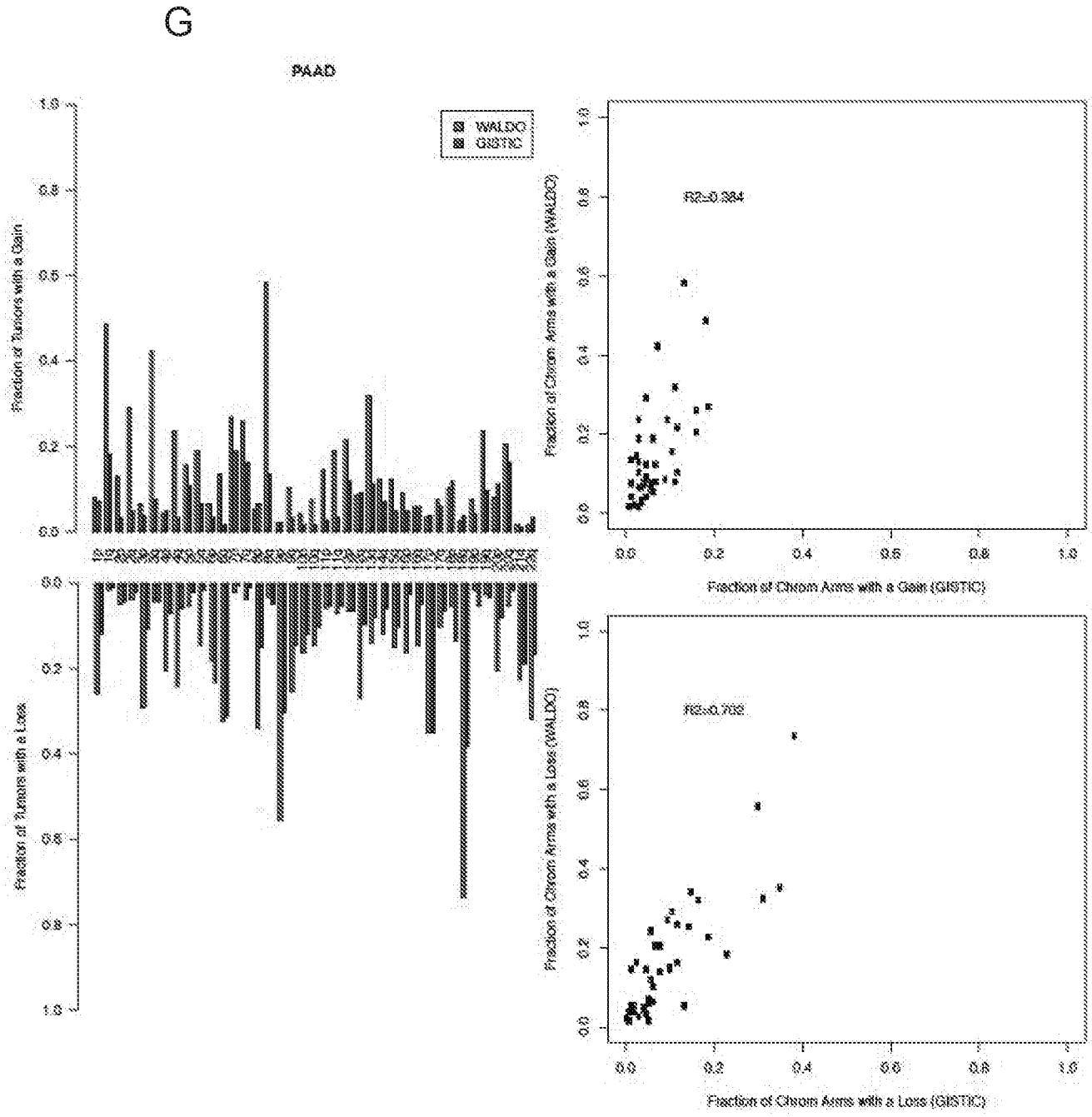


FIG. 40 (continued)

H

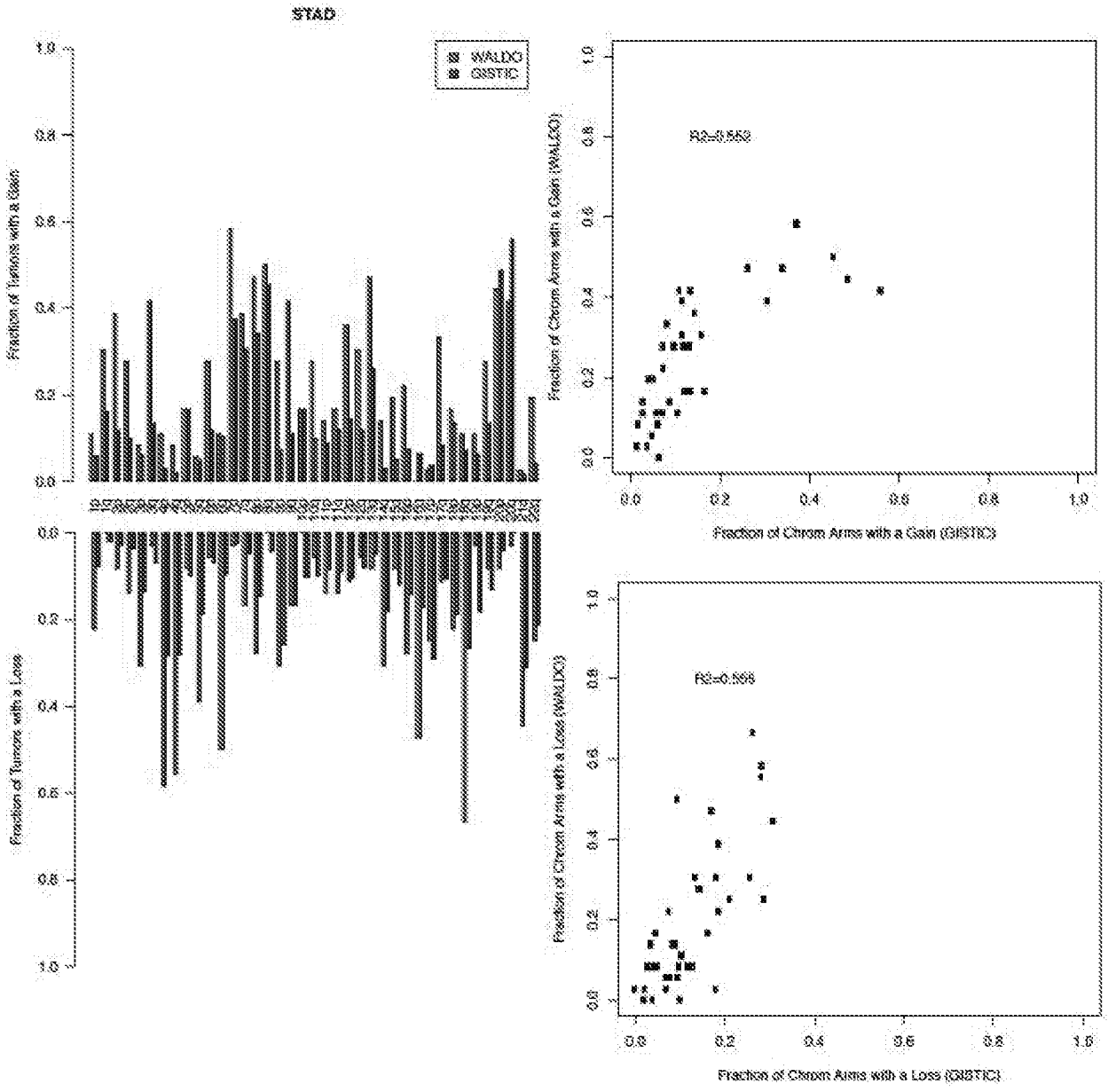


FIG. 40 (continued)

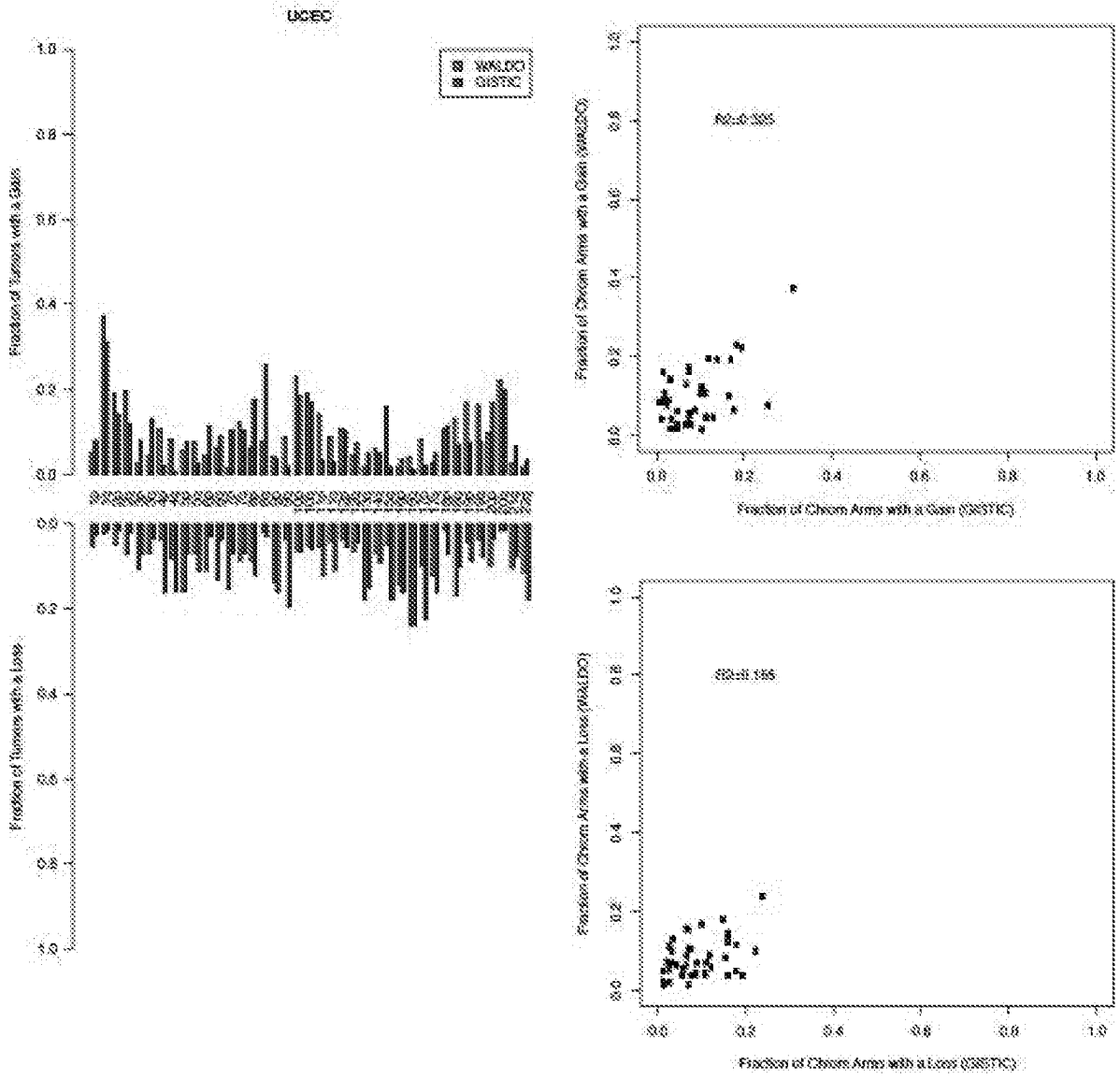
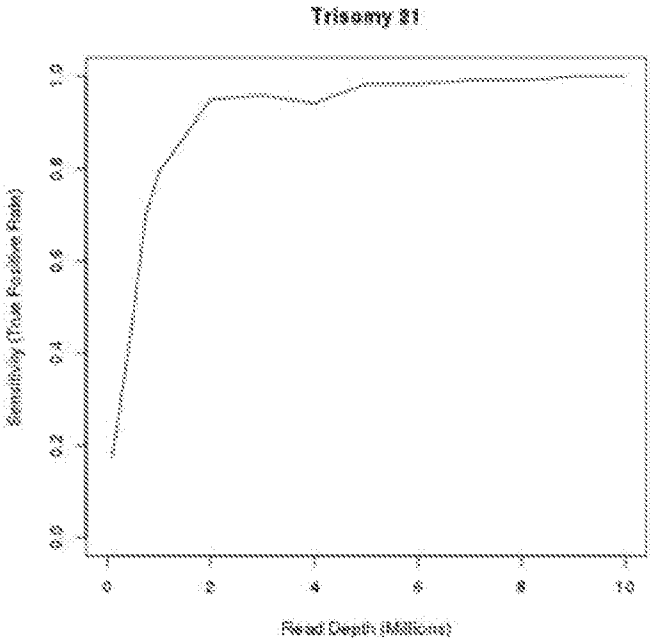


FIG. 40 (continued)

A


Trisomy 18
Sensitivity (True Positive Rate)
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Read Depth (Millions)

B

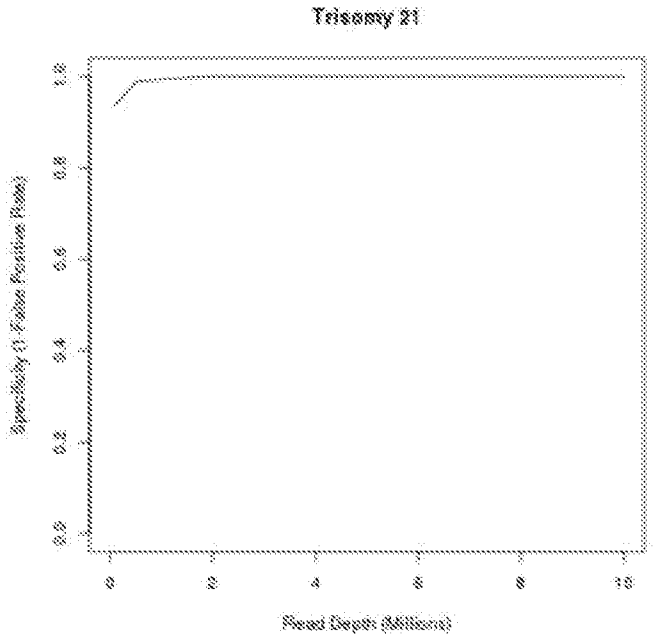

Trisomy 21
Specificity (1-False Positive Rate)
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Read Depth (Millions)

FIG. 41

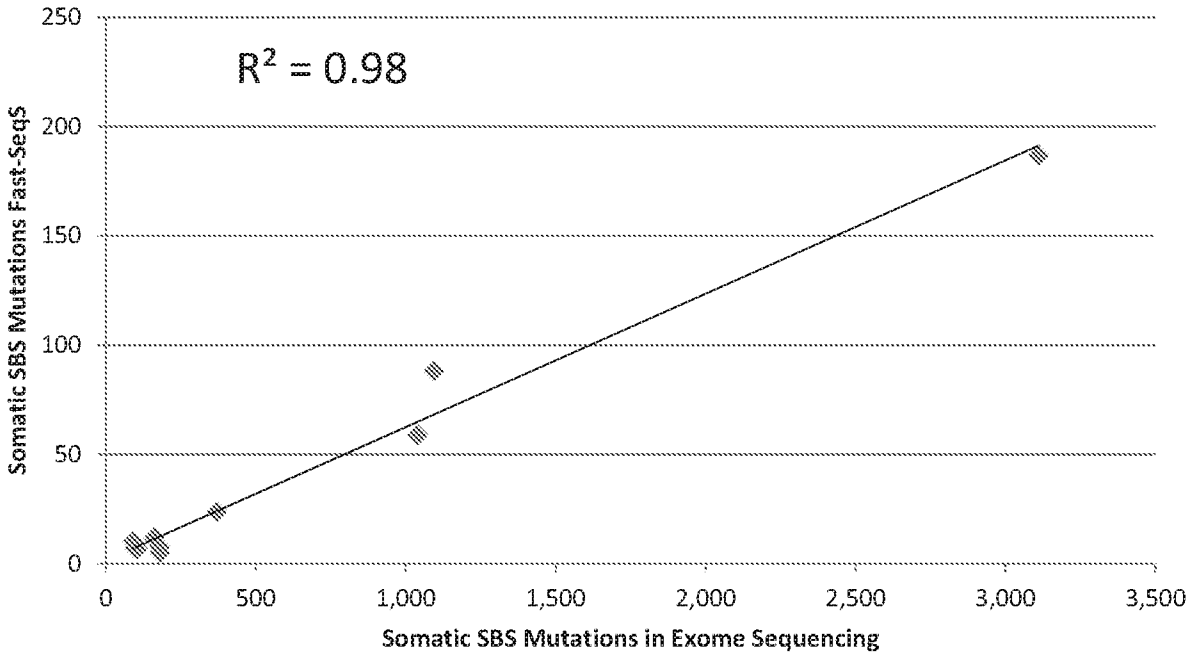

250
200
150
100
50
0
$R^2 = 0.98$
Somatic SBS Mutations Fast-SeqS
0
500
1,000
1,500
2,000
2,500
3,000
3,500
Somatic SBS Mutations in Exome Sequencing

FIG. 42

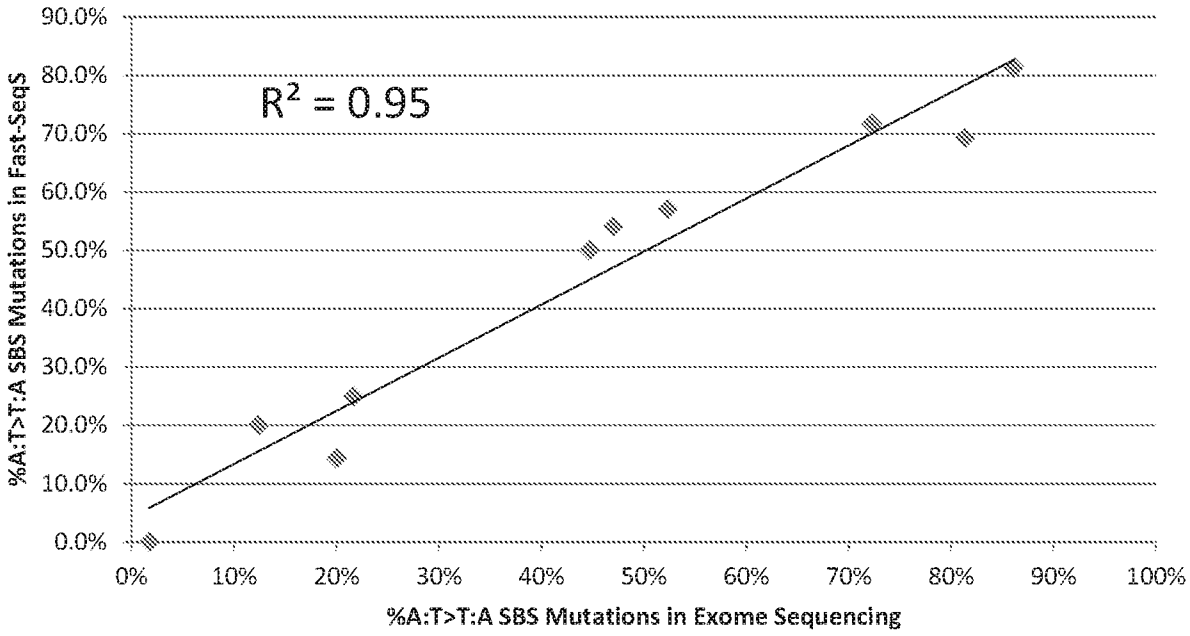
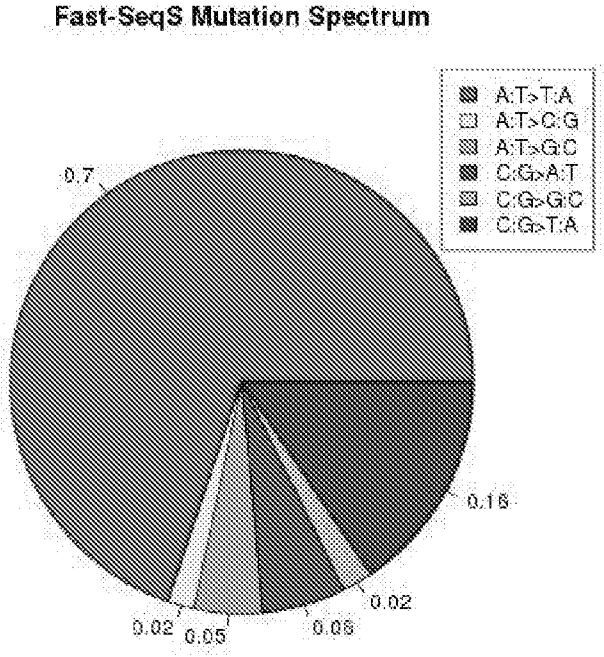

90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
%A:T>T:A SBS Mutations in Fast-SeqS
R² = 0.95
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
%A:T>T:A SBS Mutations in Exome Sequencing

FIG. 43

A


Fast-SeqS Mutation Spectrum
A:T>T:A
A:T>C:G
A:T>G:C
C:G>A:T
C:G>G:C
C:G>T:A
0.7
0.16
0.02
0.06
0.05
0.02

B

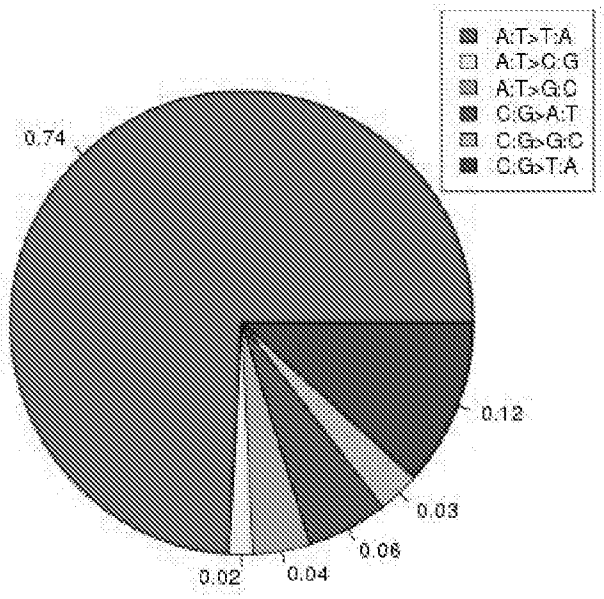

Exome Sequencing Mutation Spectrum
A:T>T:A
A:T>C:G
A:T>G:C
C:G>A:T
C:G>G:C
C:G>T:A
0.74
0.12
0.03
0.06
0.04
0.02

FIG. 44

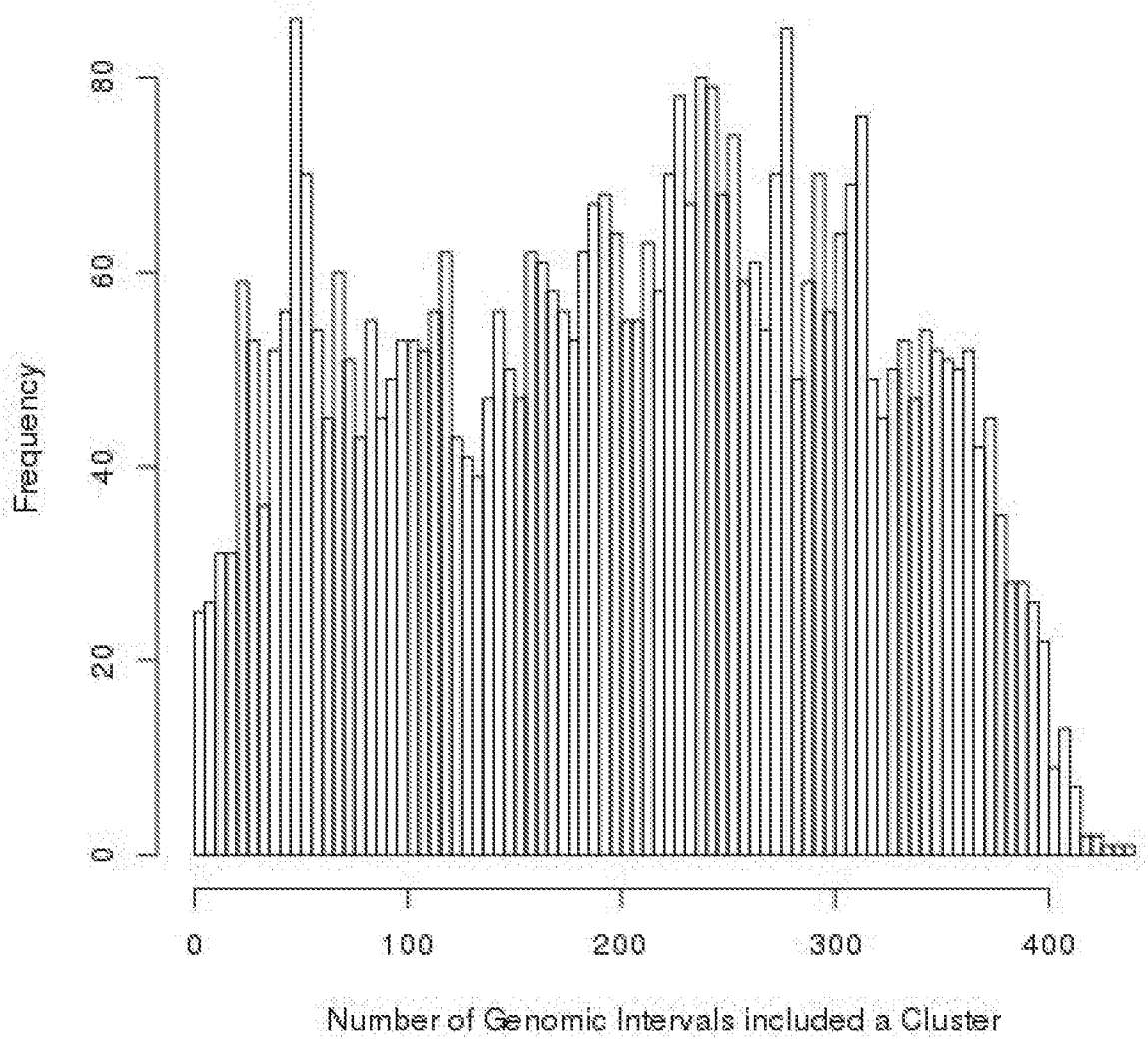

Frequency
0
20
40
60
80
0
100
200
300
400
Number of Genomic Intervals included a Cluster

FIG. 45

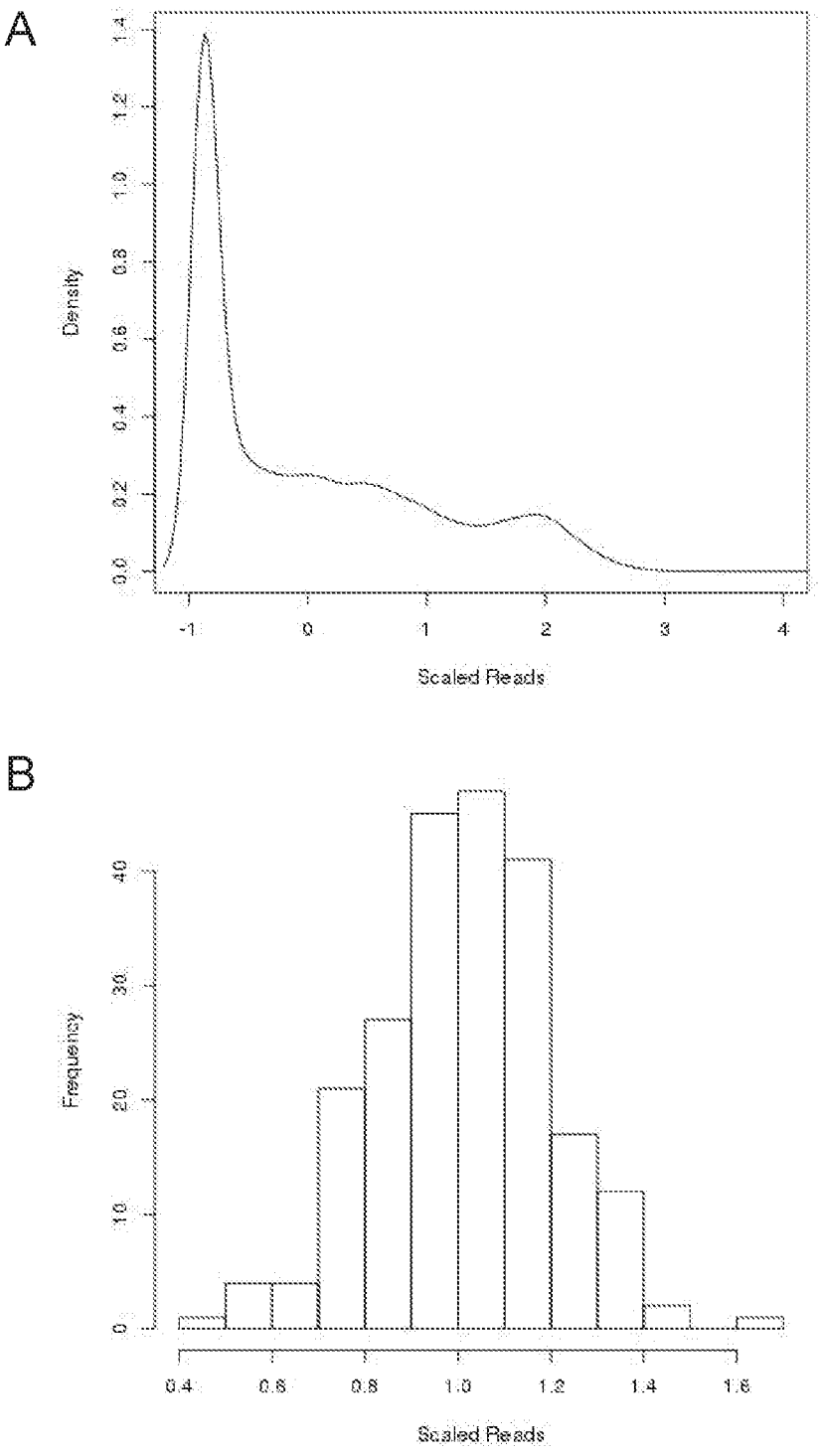

A
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
-1
0
1
2
3
4
Scaled Reads
B
Frequency
0
10
20
30
40
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Scaled Reads

FIG. 46

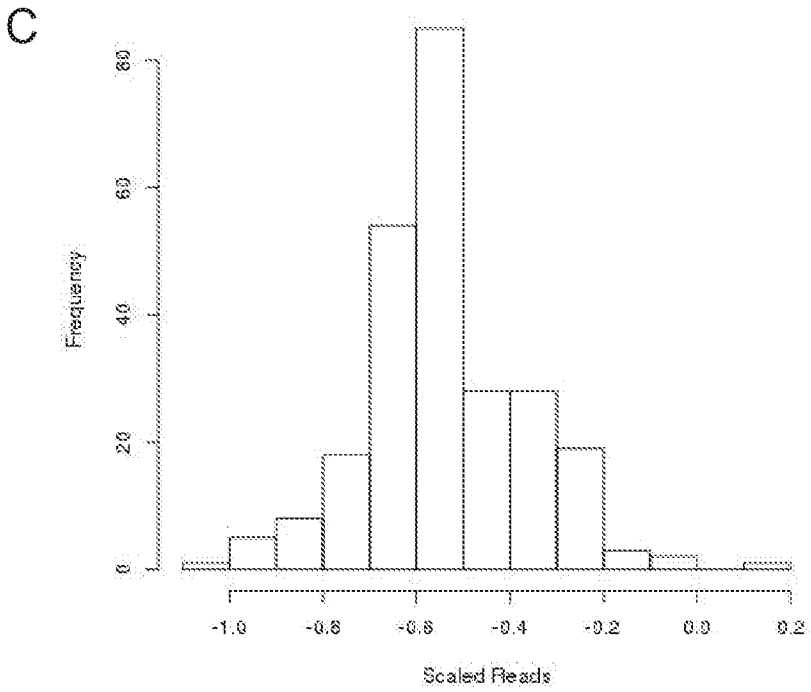


FIG. 46

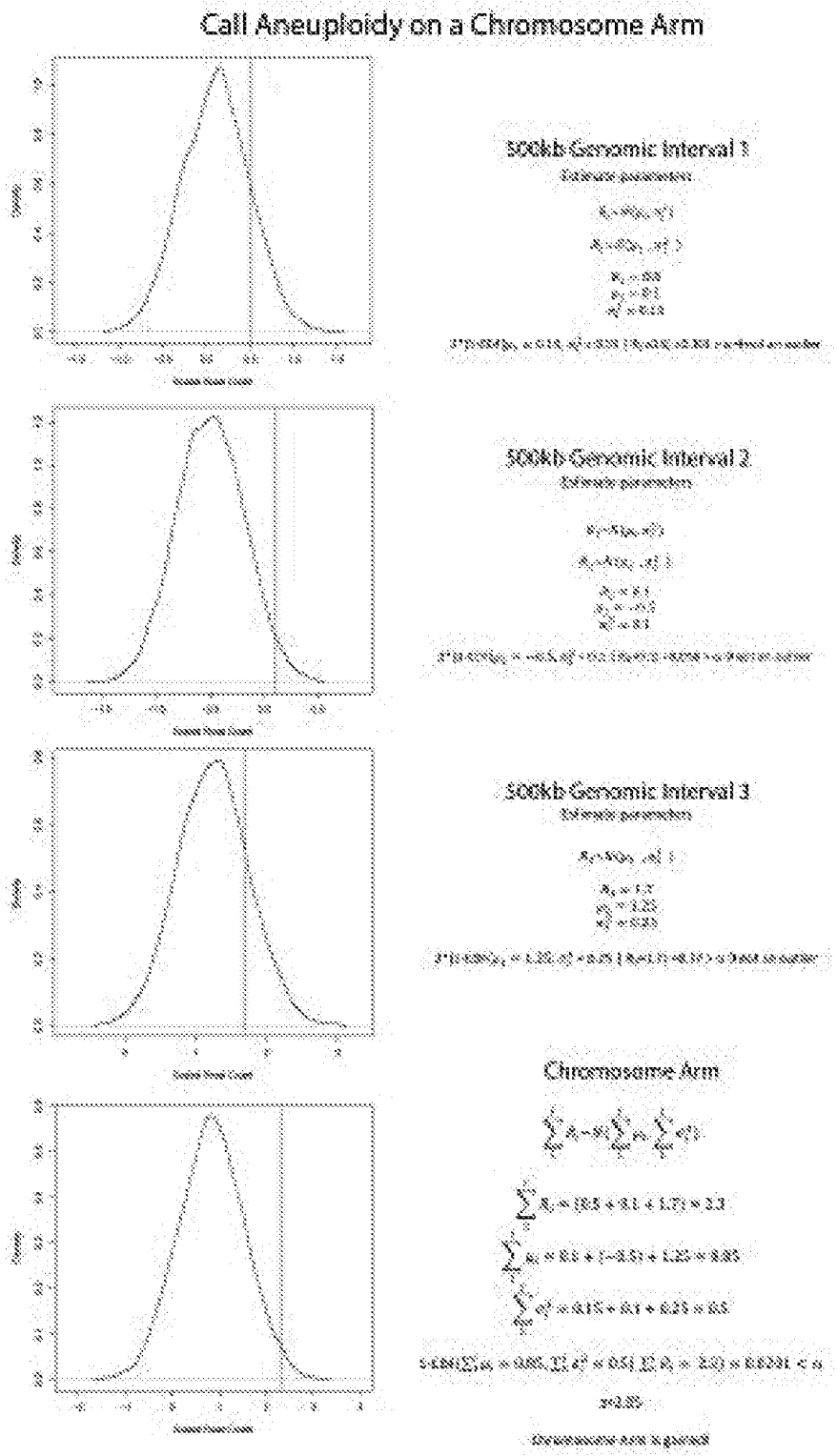

Call Aneuploidy on a Chromosome Arm
500kb Genomic Interval 1
500kb Genomic Interval 2
500kb Genomic Interval 3
Chromosome Arm

FIG. 47

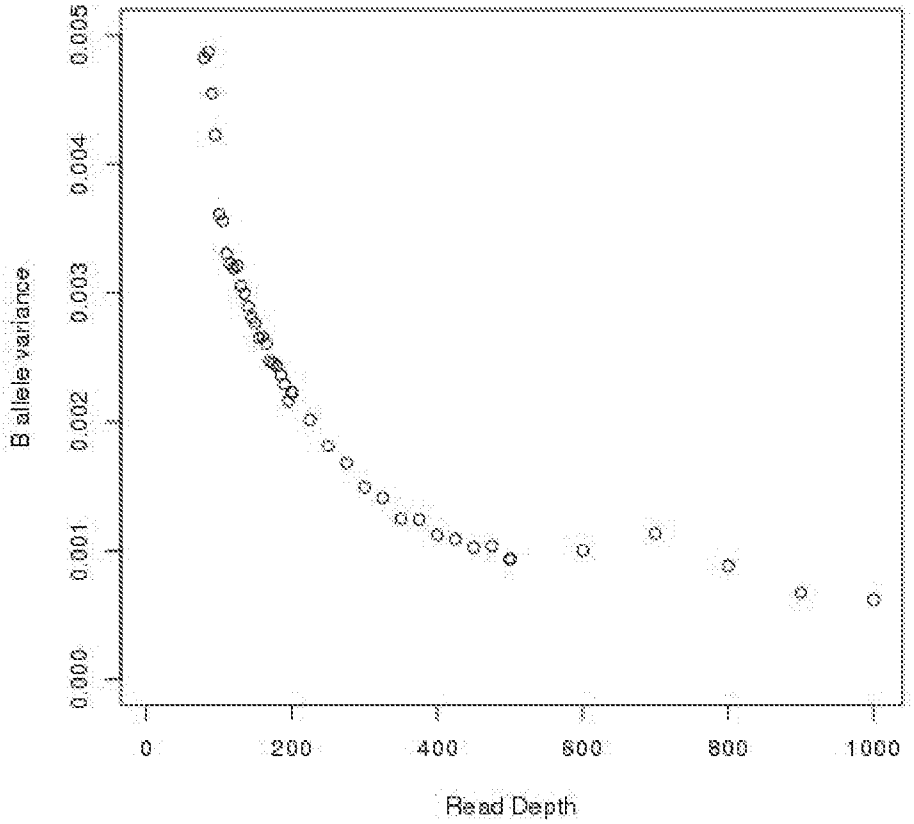

B allele variance
0.000
0.001
0.002
0.003
0.004
0.005
0
200
400
600
800
1000
Read Depth

FIG. 48

```
Synthetics with one arm alterations

N <-  number of normal WBC samples used
f <- desired neoplastic cell fraction
r <- chromosome arm (1p..22q)
r_type <- alteration type (gain or loss)
p(r) <- probability of arm gain (0.5 default)
rho <- desired unique read depth of synthetic sample (9M default)
j<-desired number of repeats

For j=1:5
        For i=1:N
                s <- sample[i]
                U <-  get UIDs from s
                For f in 1:F
                        For r in 1p:22q    #select alteration types and spike-in to t
                                For r_type in gain loss
                                        h <- new synthetic sample
                                        t <- copy of s
                                        u_ref <- get all reference allele UIDs that that map
                                                to chr arm r
                                        if  r_type== gain
                                                t  <- add  duplicate u_ref reads
                                        else (r_type == loss)
                                                t  <- subtract u_ref reads
                                        Endif
                                        h_normal<-random select (1-f)*rho UIDs from s
                                        h_aneuploid<-random select f*rho UIDs from t
                                        h<-h_normal+h_aneuploid
                                End
                        End
                End
```

FIG. 49

```
N <-  number of normal WBC samples used
d <- desired degree of aneuploidy (number arms altered)
f <- desired neoplastic cell fraction
r <- chromosome arm (1p..22q)
r_type<-arm alteration type (gain or loss)
p(r) <- probability of arm gain (0.5 default)
rho <- desired unique read depth of synthetic sample (9M default)
j<- desired number of repeats

For j=1:5
        For i=1:N
                s <- sample[i]
                U <-   get UIDs from s
                For f in 1:F
                        h <- new synthetic sample
                        For d in (5,10,15,20,25)  #desired numbers of altered arms
                                t <- copy of s
                                For a in 1:d    #select alteration types and spike-in to t
                                        r <- random select arm
                                        r_type <- alteration type with p(r)
                                        u_ref <- get all reference allele UIDs that that map
                                                to chr arm r
                                        if  r_type== gain
                                                t  <- add  duplicate u_ref reads
                                        else (r_type == loss)
                                                t  <- subtract u_ref reads
                                        Endif
                                        h_normal<-random select (1-f)*rho UIDs from s
                                        h_aneuploid<-random select f*rho UIDs from t
                                        h<-h_normal+h_aneuploid
                                End
                        End
                End
        End
End
```

FIG. 50

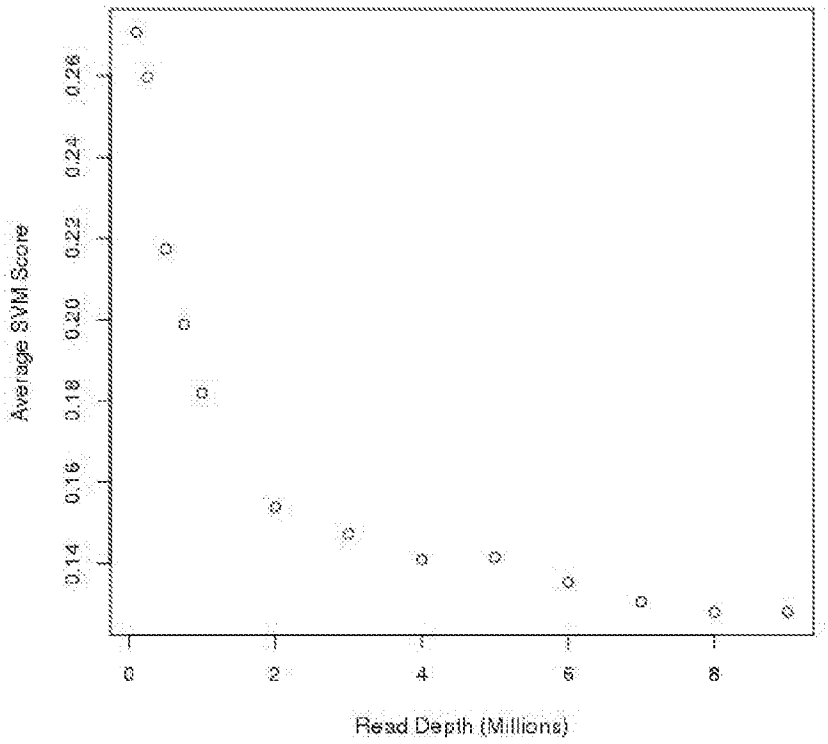

Average SVM Score
0.14
0.16
0.18
0.20
0.22
0.24
0.26
0
2
4
6
8
Read Depth (Millions)

FIG. 51

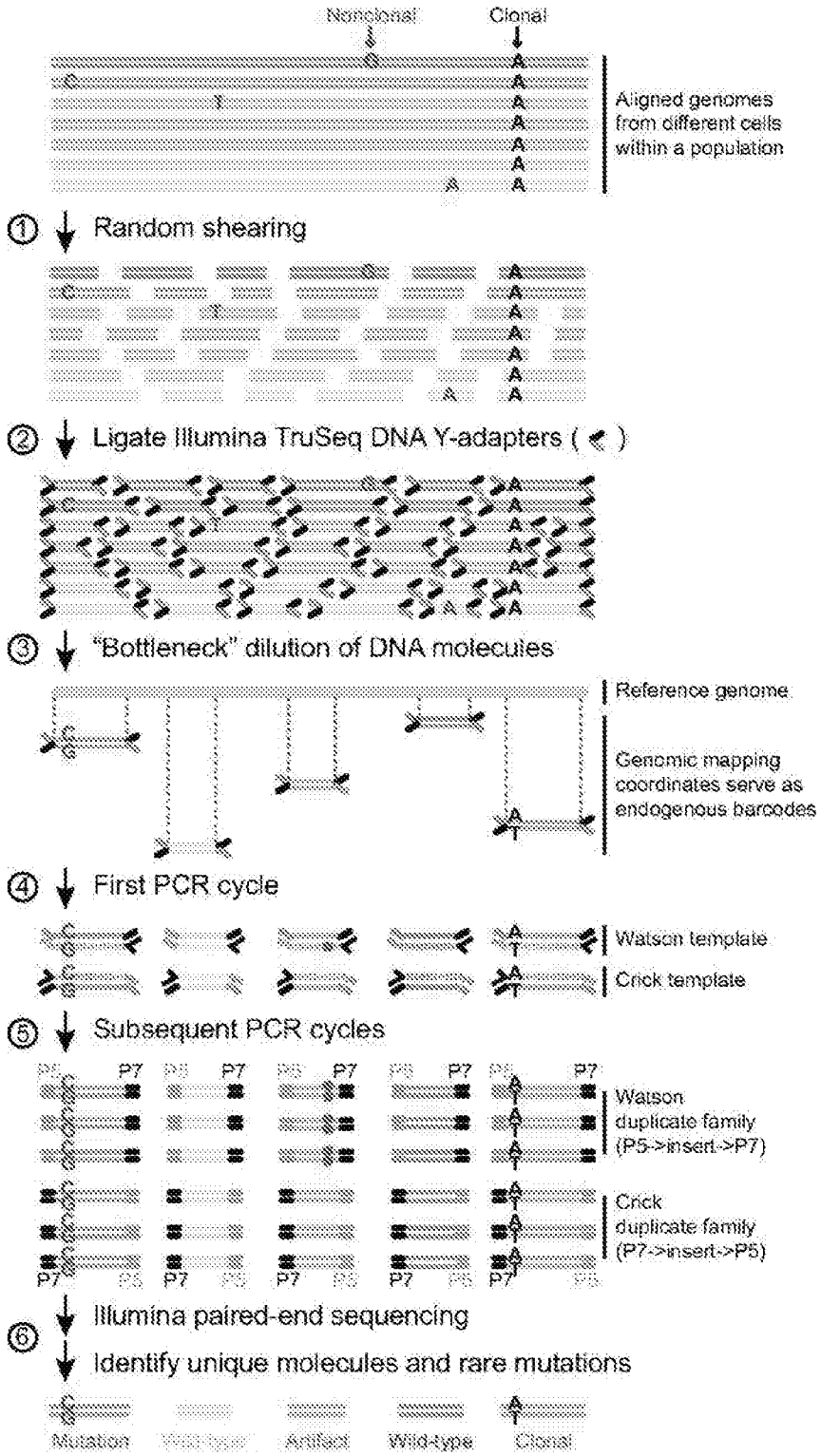

Nonclonal
Clonal
Aligned genomes from different cells within a population
① Random shearing
② Ligate Illumina TruSeq DNA Y-adapters ( )
③ "Bottleneck" dilution of DNA molecules
Reference genome
Genomic mapping coordinates serve as endogenous barcodes
④ First PCR cycle
Watson template
Crick template
⑤ Subsequent PCR cycles
P5
P7
Watson duplicate family (P5->insert->P7)
Crick duplicate family (P7->insert->P5)
⑥ Illumina paired-end sequencing
Identify unique molecules and rare mutations
Mutation
Artifact
Wild-type
Clonal

FIG. 52

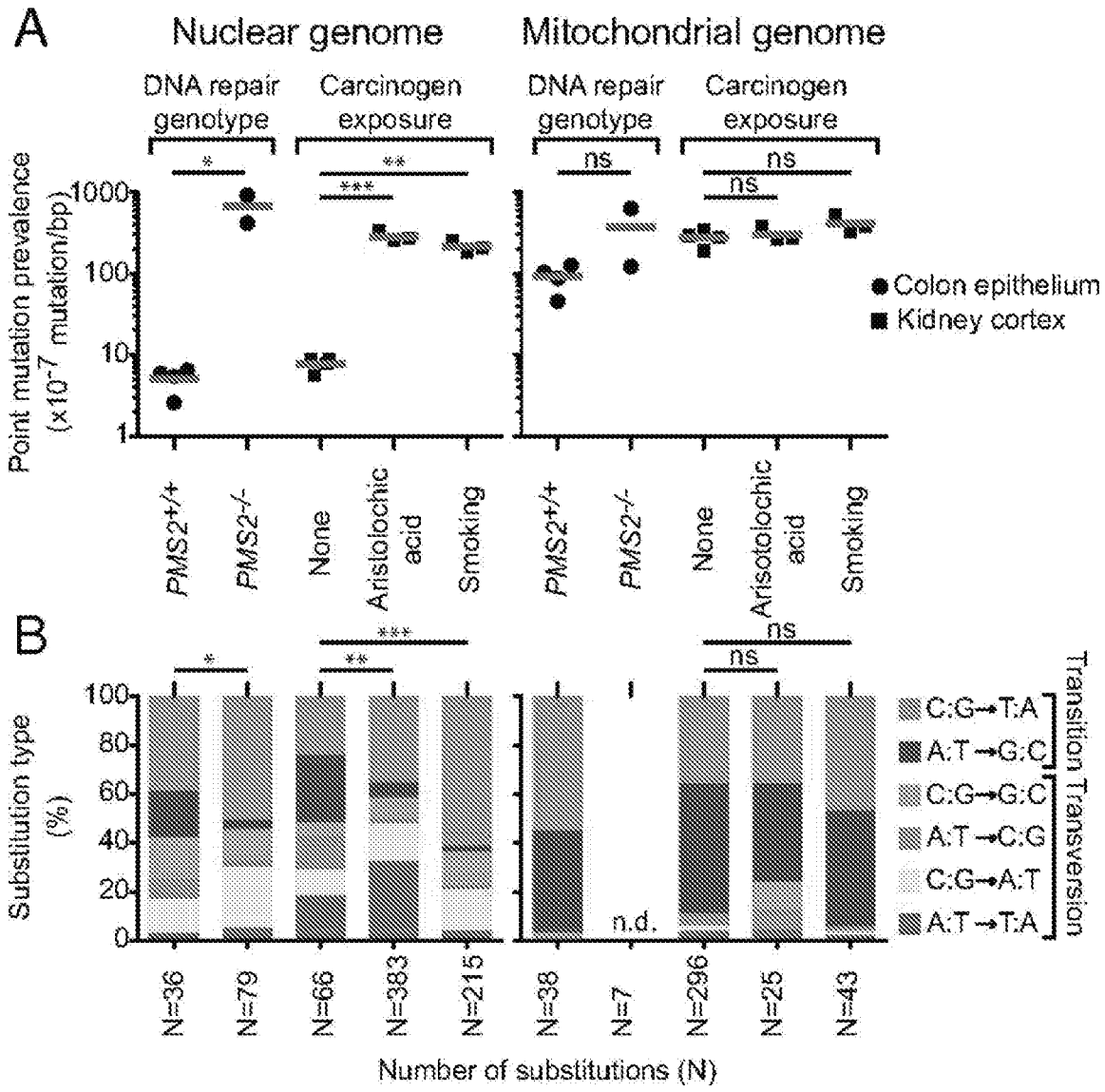

A
Nuclear genome
Mitochondrial genome
DNA repair genotype
Carcinogen exposure
DNA repair genotype
Carcinogen exposure
Point mutation prevalence (x10-7 mutation/bp)
1000
100
10
1
*
**
***
ns
ns
ns
Colon epithelium
Kidney cortex
PMS2+/+
PMS2-/-
None
Aristolochic acid
Smoking
PMS2+/+
PMS2-/-
None
Arisotolochic acid
Smoking
B
Substitution type (%)
100
80
60
40
20
0
*
**
***
ns
ns
n.d.
C:G→T:A
A:T→G:C
C:G→G:C
A:T→C:G
C:G→A:T
A:T→T:A
Transition
Transversion
N=36
N=79
N=66
N=383
N=215
N=38
N=7
N=296
N=25
N=43
Number of substitutions (N)

FIG. 53

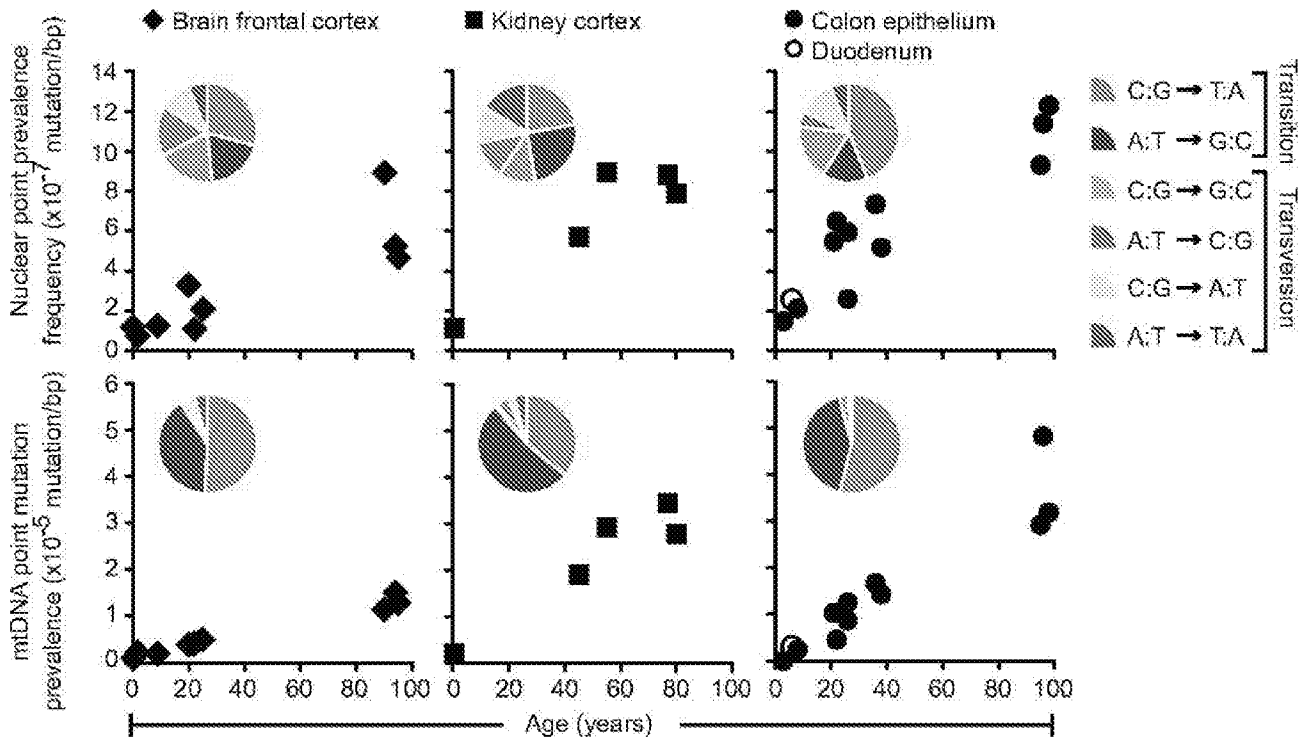

Brain frontal cortex
Kidney cortex
Colon epithelium
Duodenum
Nuclear point prevalence frequency (x10⁻⁷ mutation/bp)
mtDNA point mutation prevalence (x10⁻⁵ mutation/bp)
C:G → T:A
A:T → G:C
C:G → G:C
A:T → C:G
C:G → A:T
A:T → T:A
Transition
Transversion
Age (years)

FIG. 54

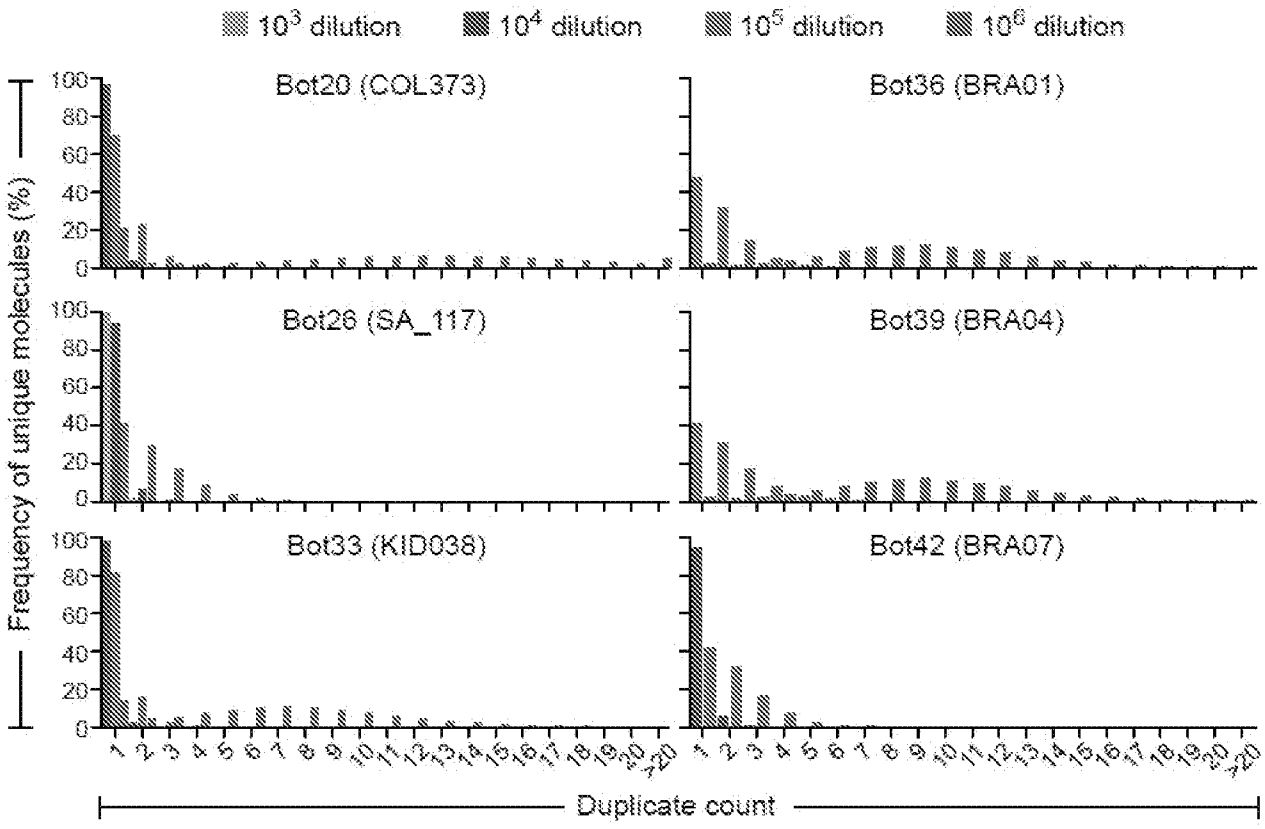

10^3 dilution
10^4 dilution
10^5 dilution
10^6 dilution
Bot20 (COL373)
Bot36 (BRA01)
Bot26 (SA_117)
Bot39 (BRA04)
Bot33 (KID038)
Bot42 (BRA07)
Frequency of unique molecules (%)
Duplicate count

FIG. 55

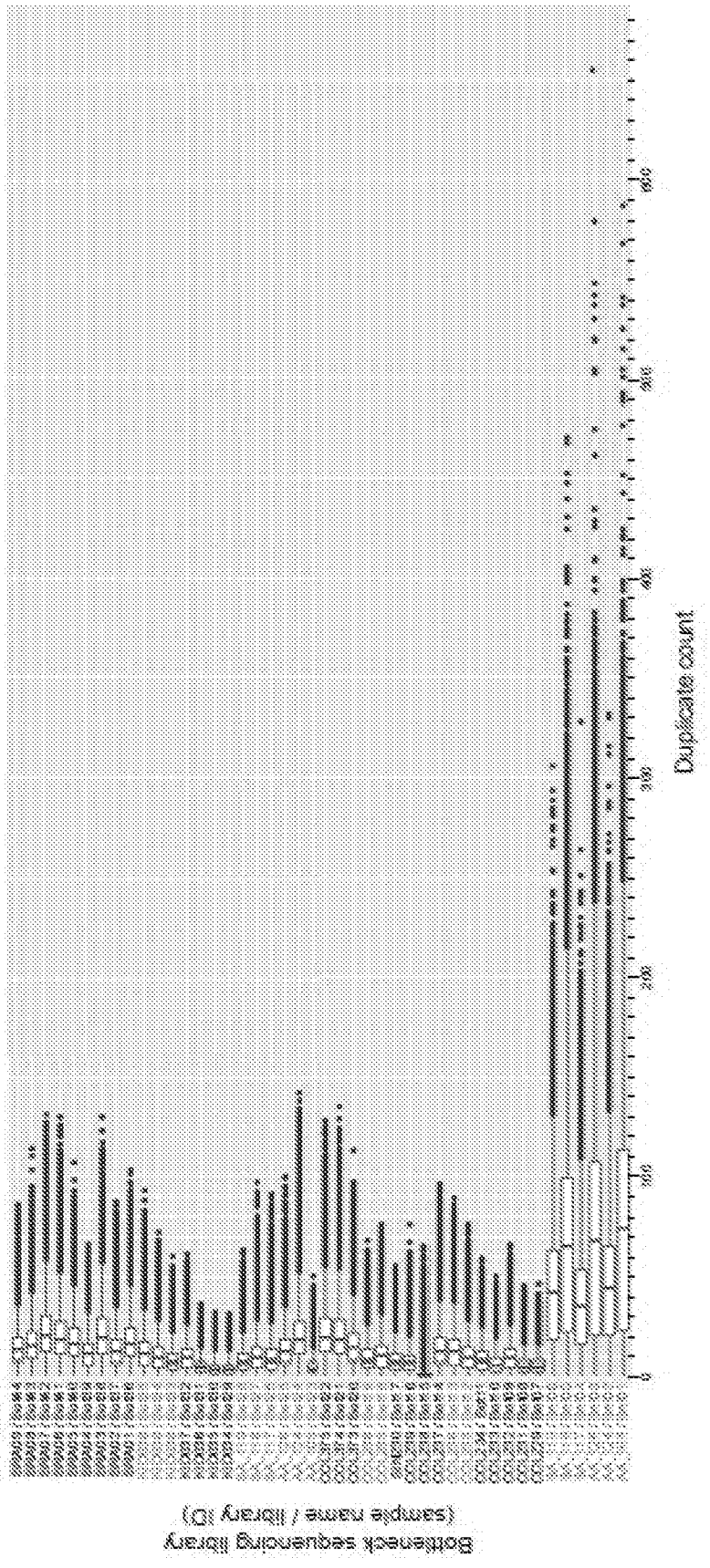

Duplicate count
Bottleneck sequencing library
(sample name / library ID)

FIG. 56

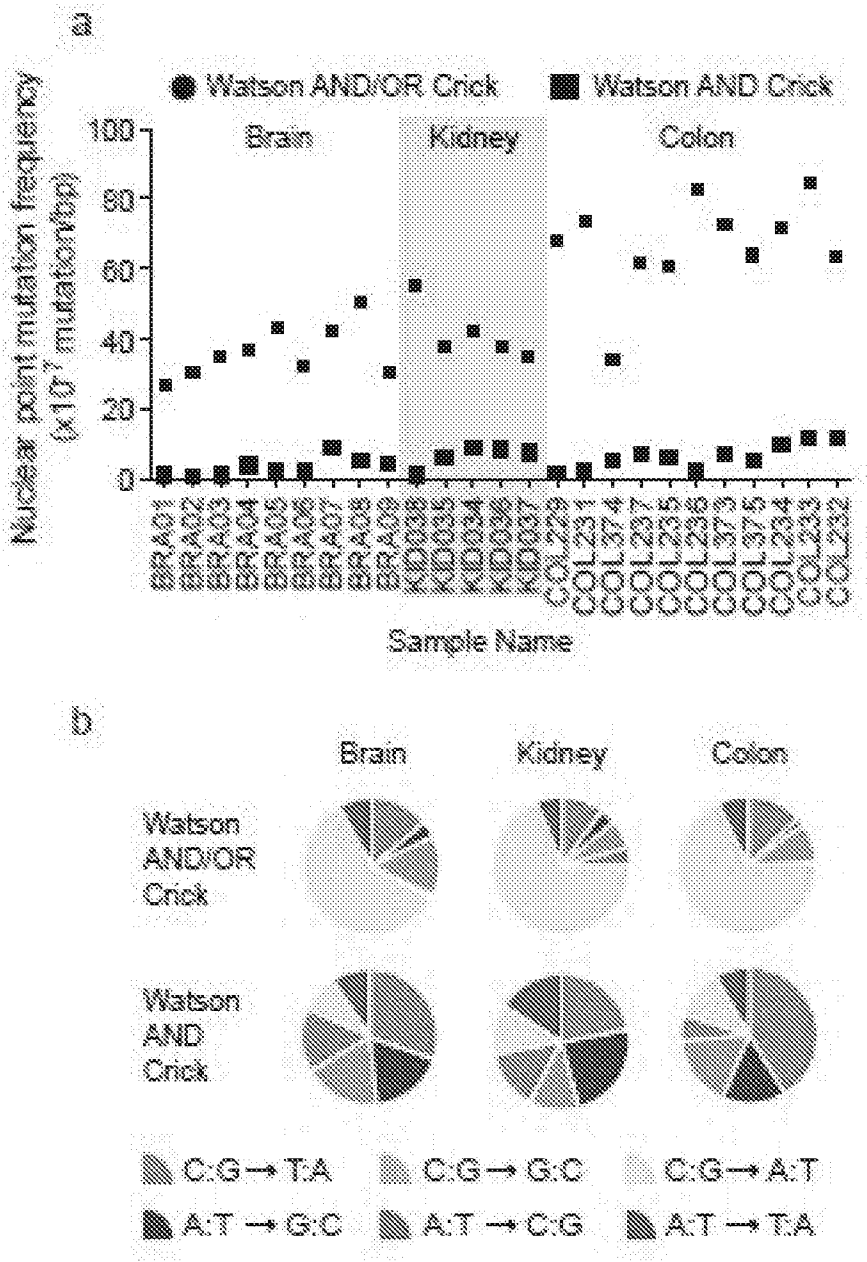

a
Watson AND/OR Crick
Watson AND Crick
Brain
Kidney
Colon
Nuclear point mutation frequency (x10⁻⁷ mutation/bp)
100
80
60
40
20
0
BRA01
BRA02
BRA03
BRA04
BRA05
BRA06
BRA07
BRA08
BRA09
KID035
KID035
KID034
KID036
KID037
COL229
COL231
COL374
COL237
COL235
COL236
COL373
COL375
COL234
COL233
COL232
Sample Name
b
Brain
Kidney
Colon
Watson AND/OR Crick
Watson AND Crick
C:G → T:A
C:G → G:C
C:G → A:T
A:T → G:C
A:T → C:G
A:T → T:A

FIG. 57

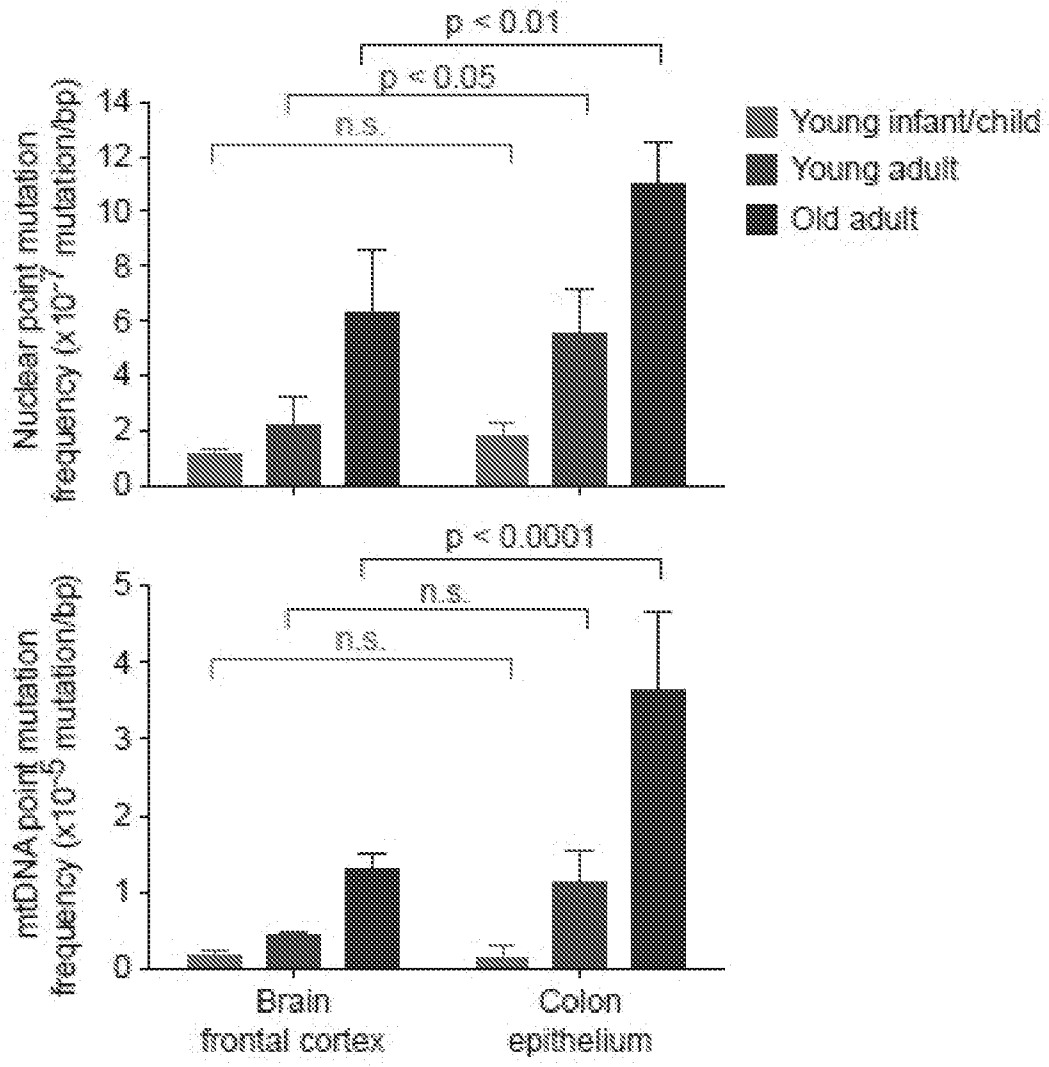

p < 0.01
p < 0.05
n.s.
Nuclear point mutation frequency (x 10^-7 mutation/bp)
14
12
10
8
6
4
2
0
Young infant/child
Young adult
Old adult
p < 0.0001
n.s.
n.s.
mtDNA point mutation frequency (x10^-5 mutation/bp)
5
4
3
2
1
0
Brain frontal cortex
Colon epithelium

FIG. 58

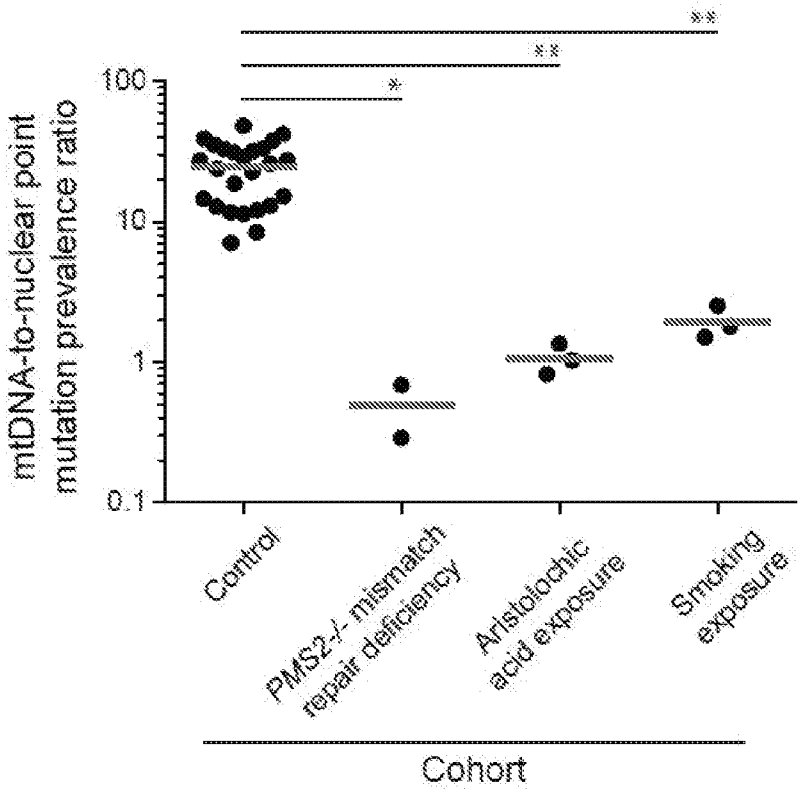

**
**
*
100
10
1
0.1
mtDNA-to-nuclear point mutation prevalence ratio
Control
PMS2-/- mismatch repair deficiency
Aristolochic acid exposure
Smoking exposure
Cohort

FIG. 59

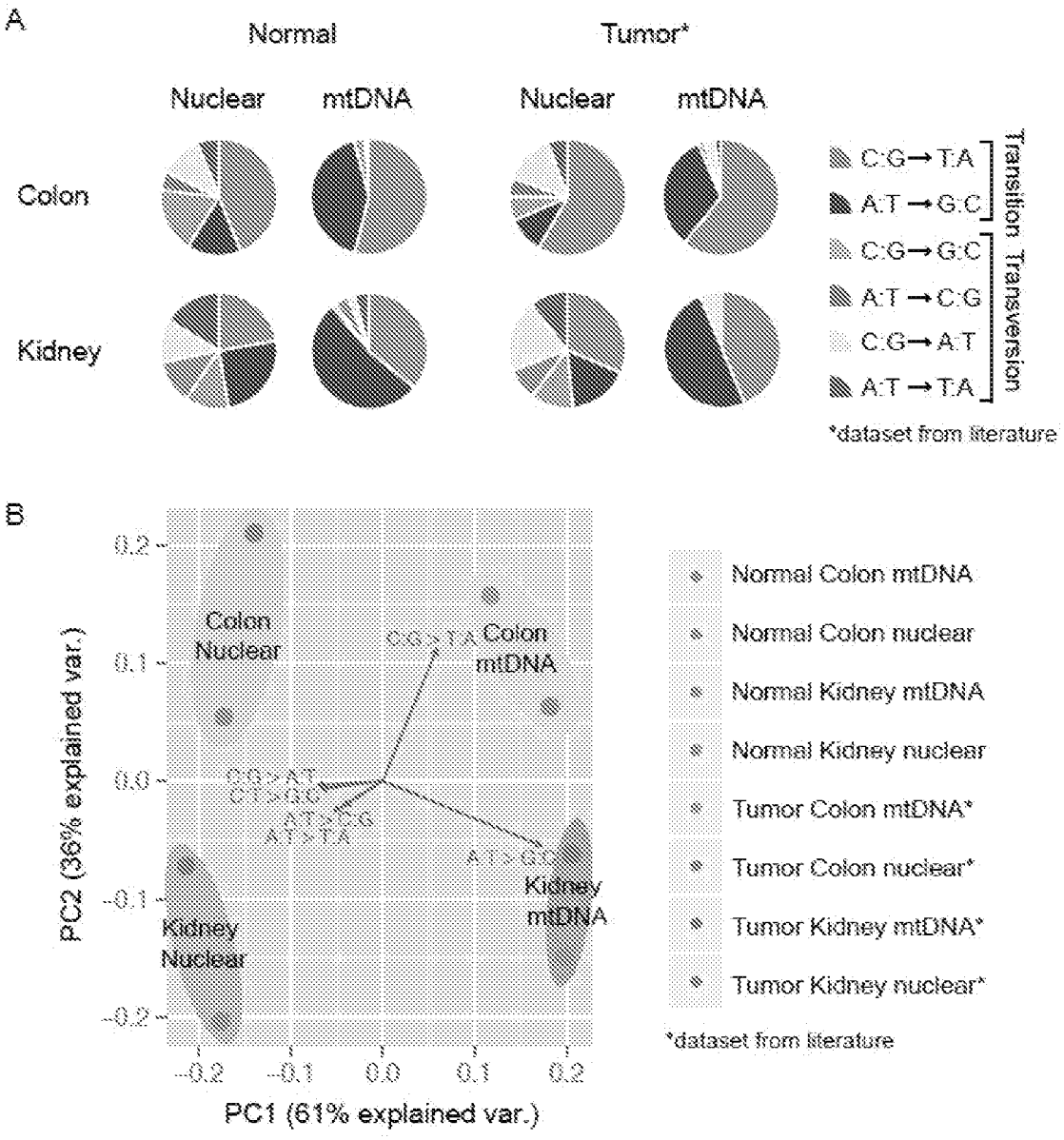

A
Normal
Tumor*
Nuclear
mtDNA
Nuclear
mtDNA
Colon
Kidney
C:G→T:A
A:T→G:C
C:G→G:C
A:T→C:G
C:G→A:T
A:T→T:A
Transition
Transversion
*dataset from literature
B
PC2 (36% explained var.)
PC1 (61% explained var.)
0.2
0.1
0.0
-0.1
-0.2
Colon Nuclear
Colon mtDNA
Kidney Nuclear
Kidney mtDNA
C:G>T:A
C:G>A:T
C:G>G:C
A:T>C:G
A:T>T:A
A:T>G:C
Normal Colon mtDNA
Normal Colon nuclear
Normal Kidney mtDNA
Normal Kidney nuclear
Tumor Colon mtDNA*
Tumor Colon nuclear*
Tumor Kidney mtDNA*
Tumor Kidney nuclear*
*dataset from literature

FIG. 60

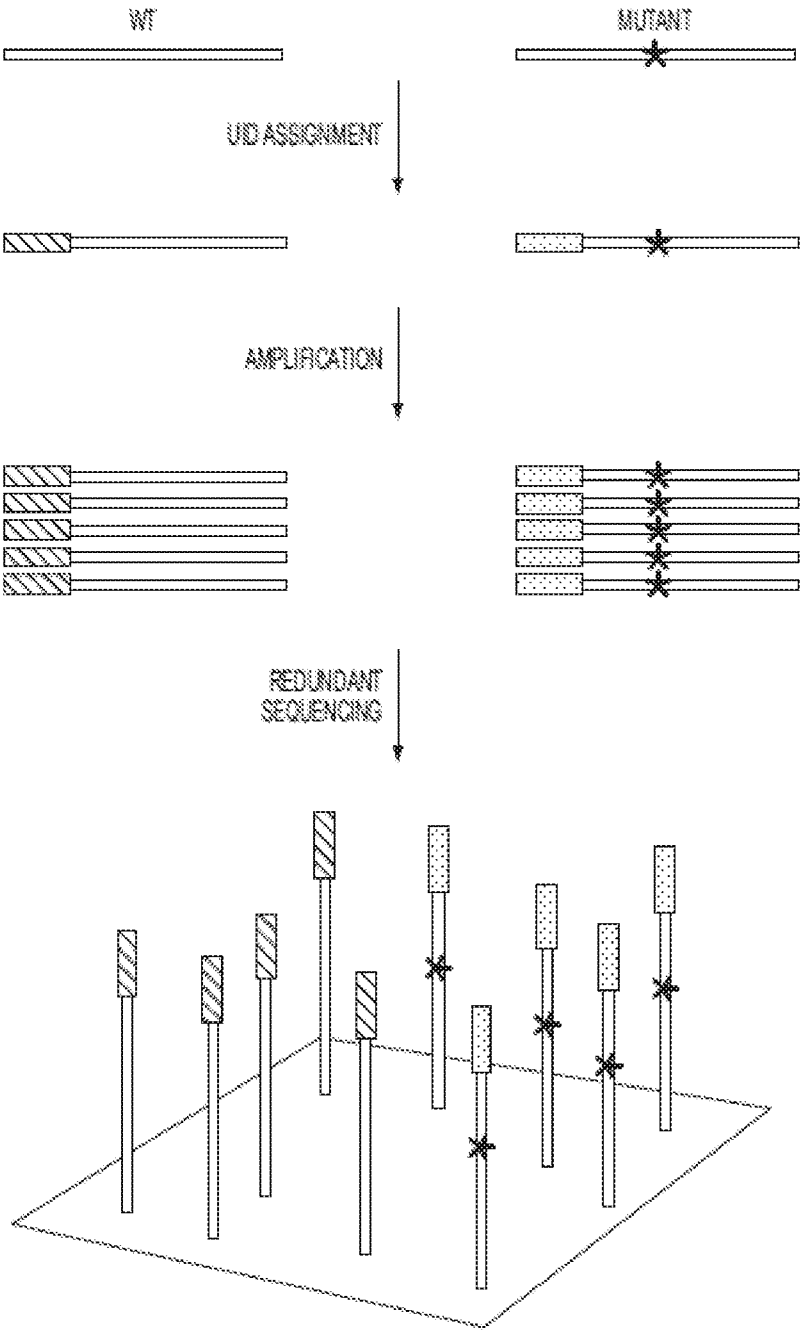

WT
MUTANT
UID ASSIGNMENT
AMPLIFICATION
REDUNDANT SEQUENCING

FIG. 61

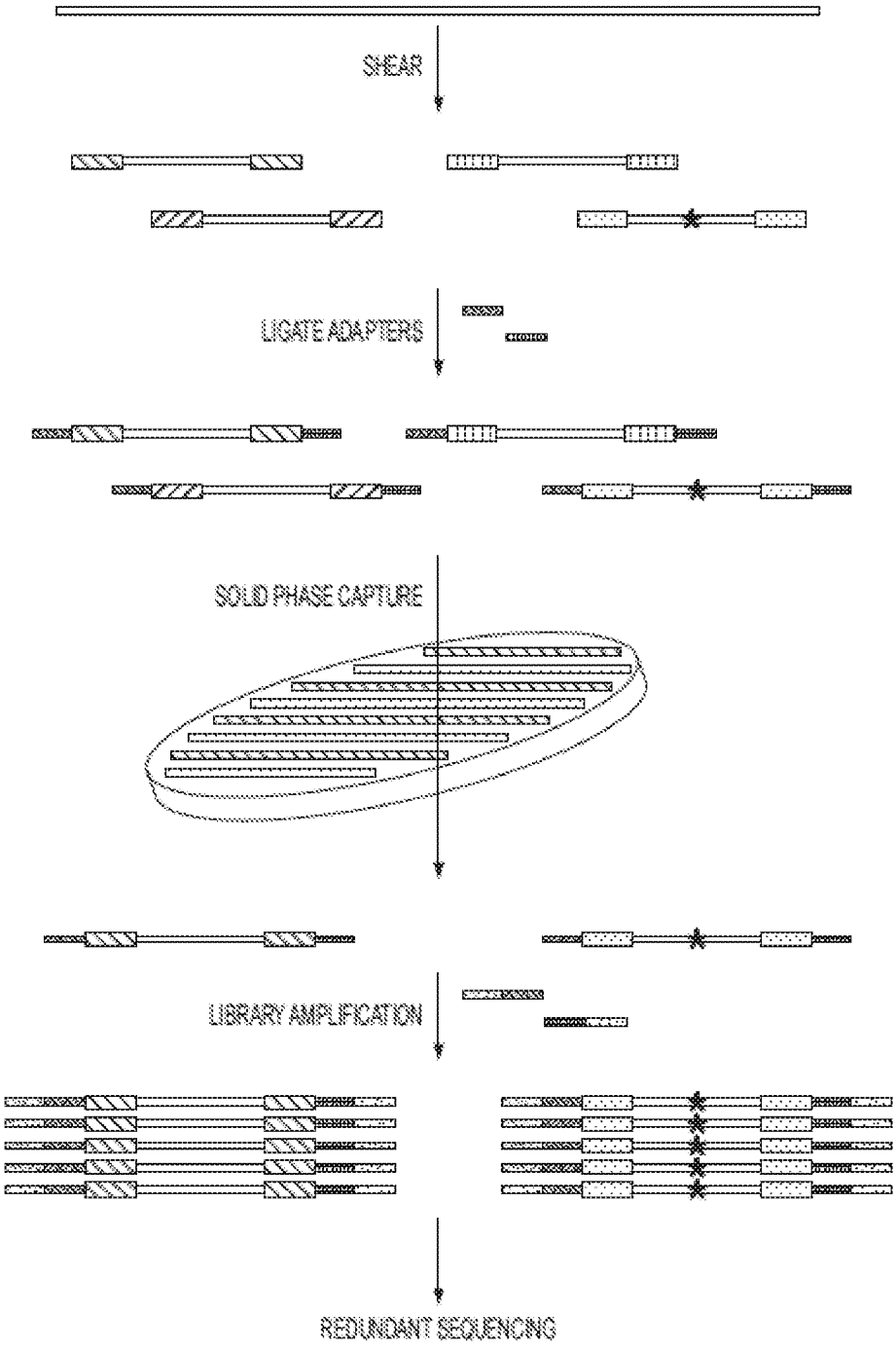

SHEAR
LIGATE ADAPTERS
SOLID PHASE CAPTURE
LIBRARY AMPLIFICATION
REDUNDANT SEQUENCING

FIG. 62

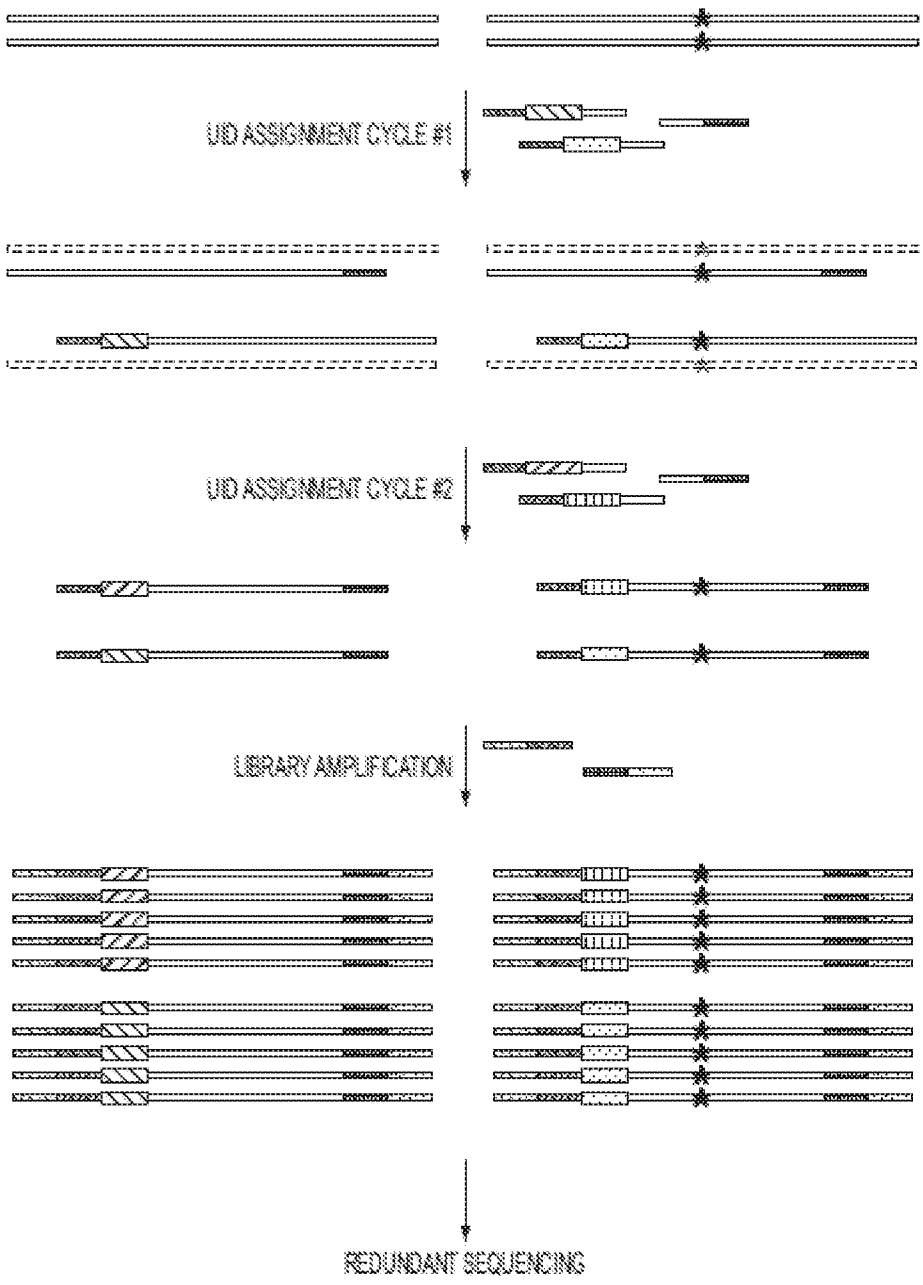
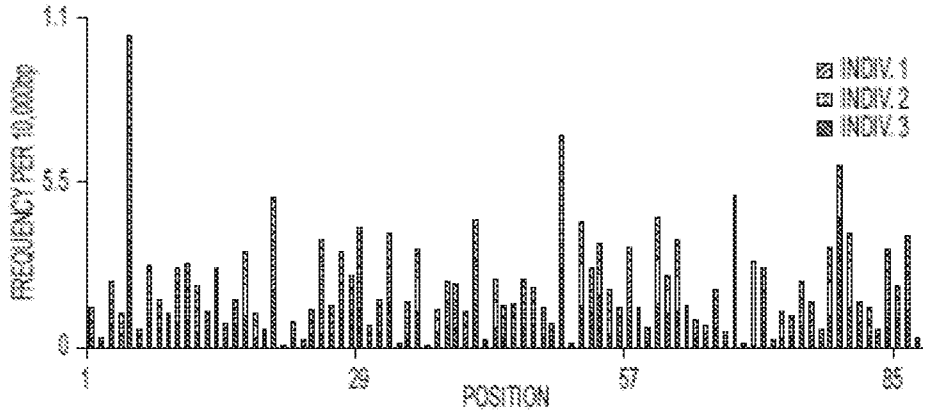

UID ASSIGNMENT CYCLE #1
UID ASSIGNMENT CYCLE #2
LIBRARY AMPLIFICATION
REDUNDANT SEQUENCING

FIG. 63

A


1.1
0.55
0
FREQUENCY PER 10,000bp
INDV. 1
INDV. 2
INDV. 3
1
29
57
85
POSITION

B

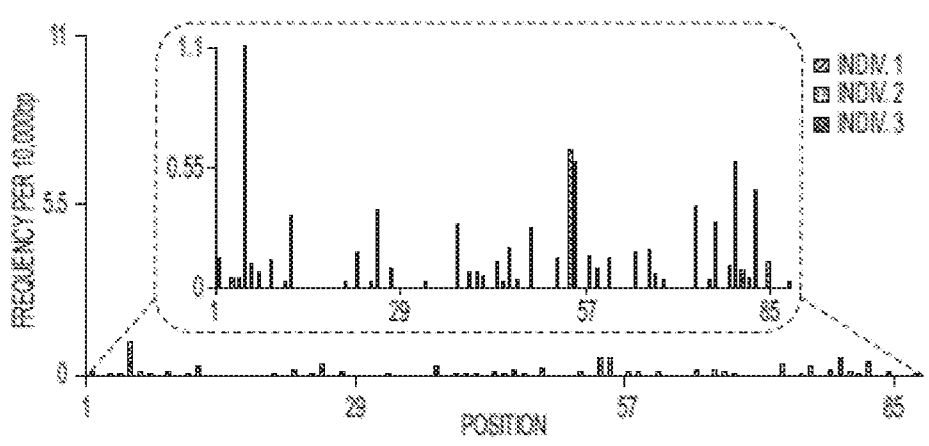

11
5.5
0
FREQUENCY PER 10,000bp
1.1
0.55
0
INDV. 1
INDV. 2
INDV. 3
1
29
57
85
POSITION

FIG. 64

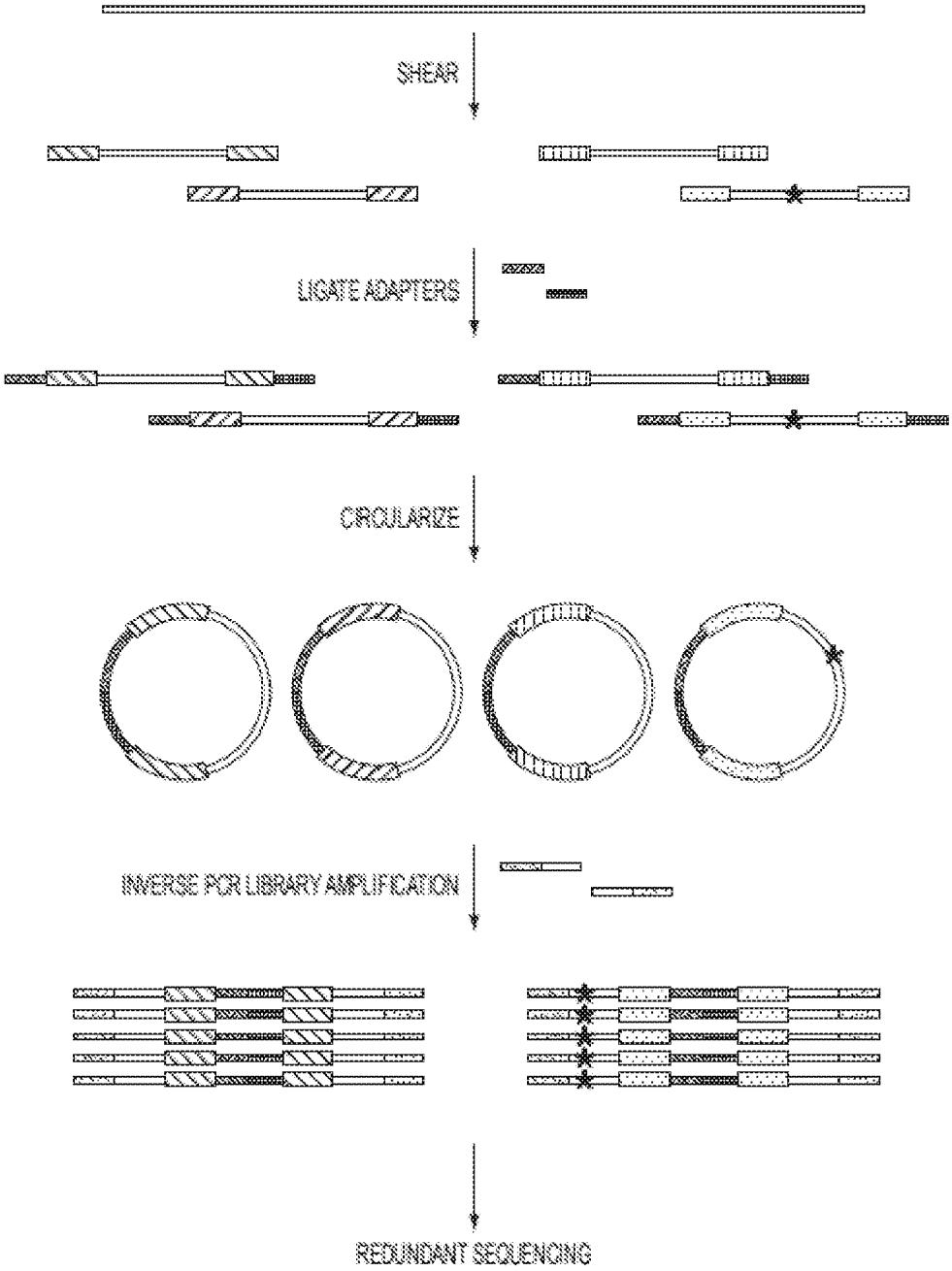

SHEAR
LIGATE ADAPTERS
CIRCULARIZE
INVERSE PCR LIBRARY AMPLIFICATION
REDUNDANT SEQUENCING

FIG. 65

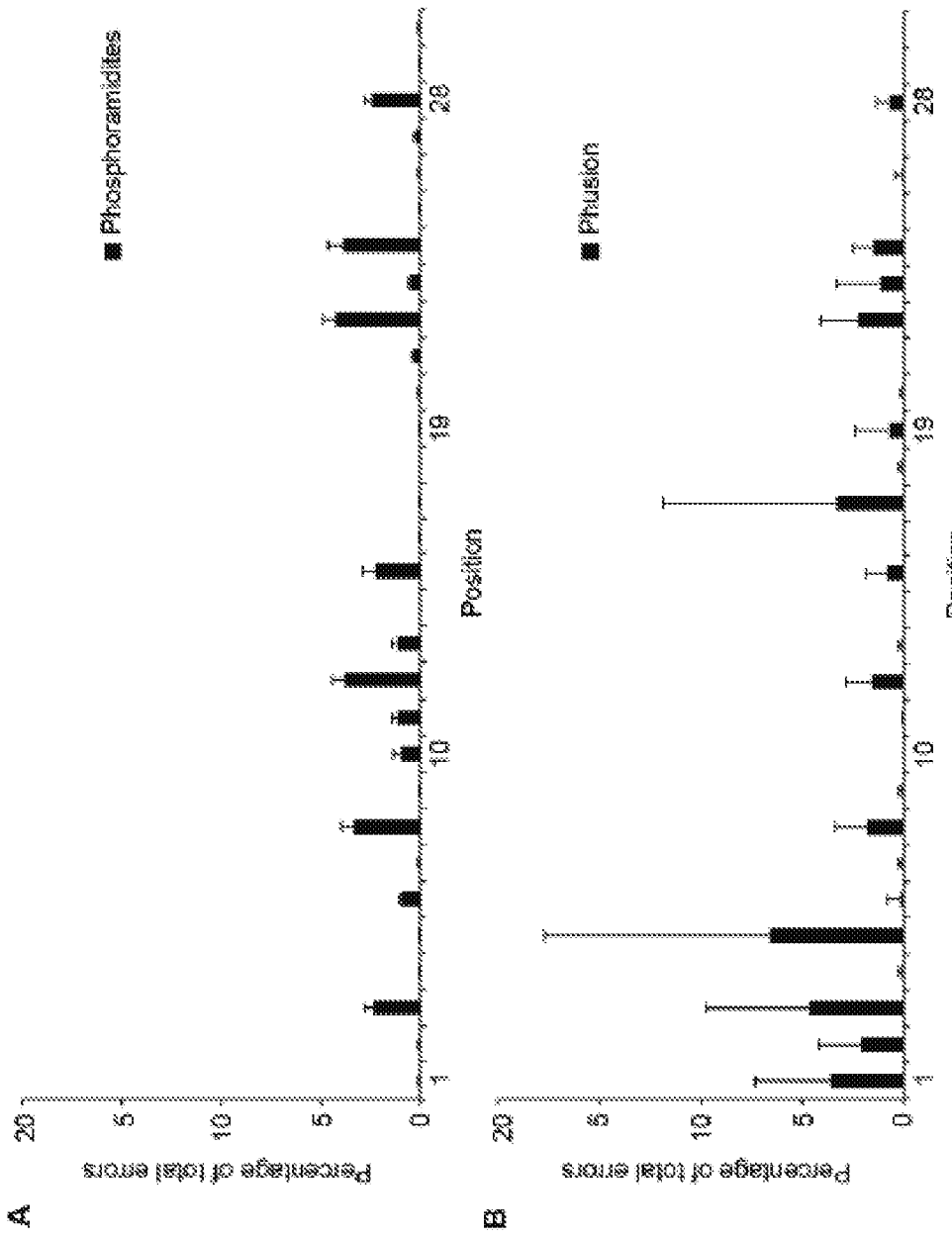

A
Phosphoramidites
Percentage of total errors
Position
B
Phusion
Percentage of total errors
Position

FIG. 66

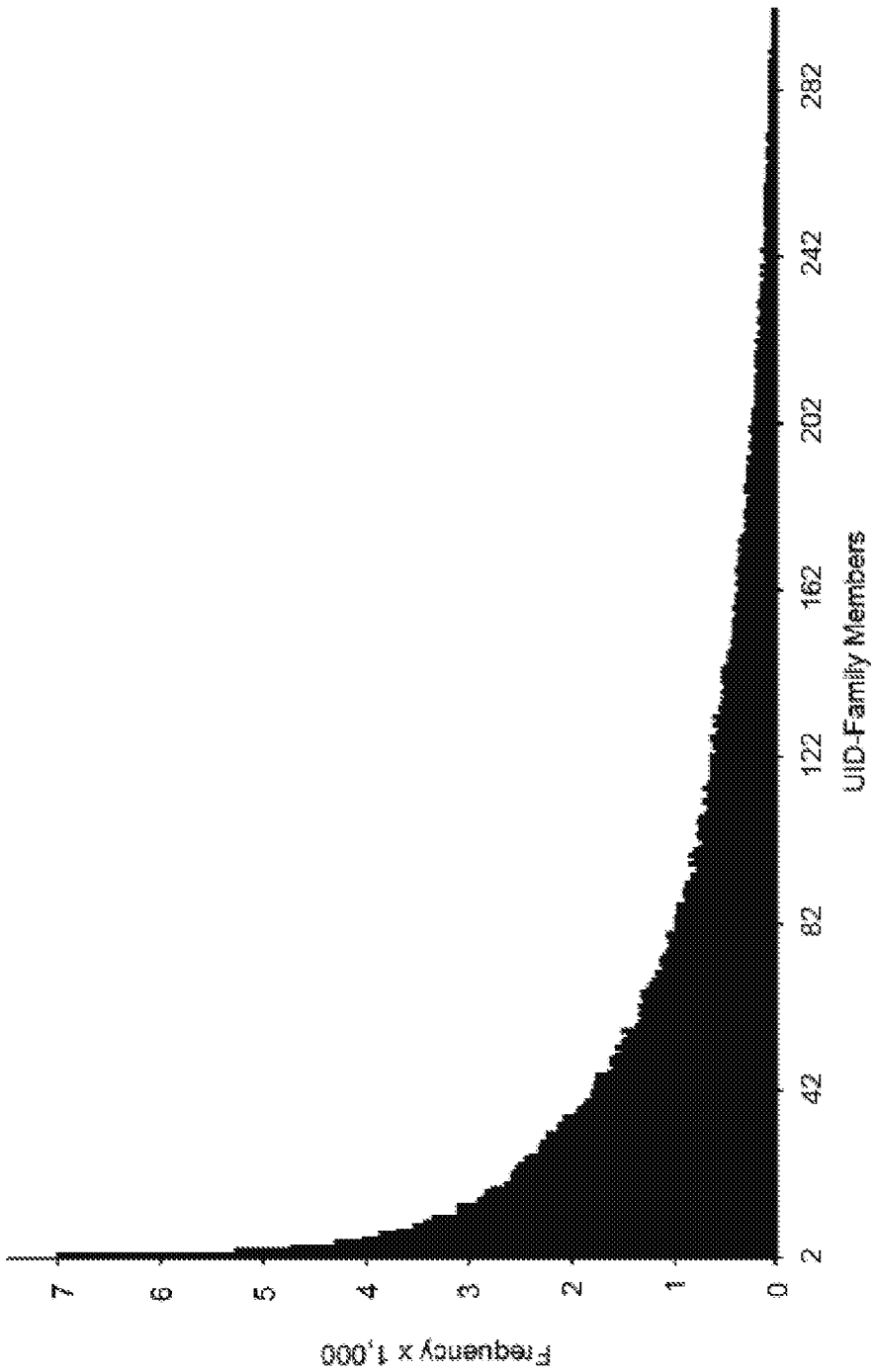

Frequency x 1,000
0
1
2
3
4
5
6
7
2
42
82
122
162
202
242
282
UID-Family Members

FIG. 67

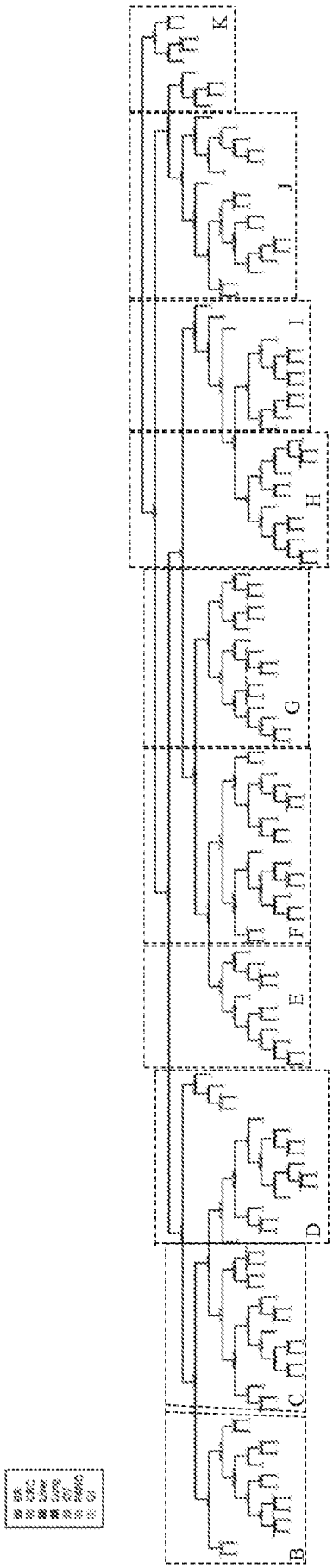


FIG. 68A

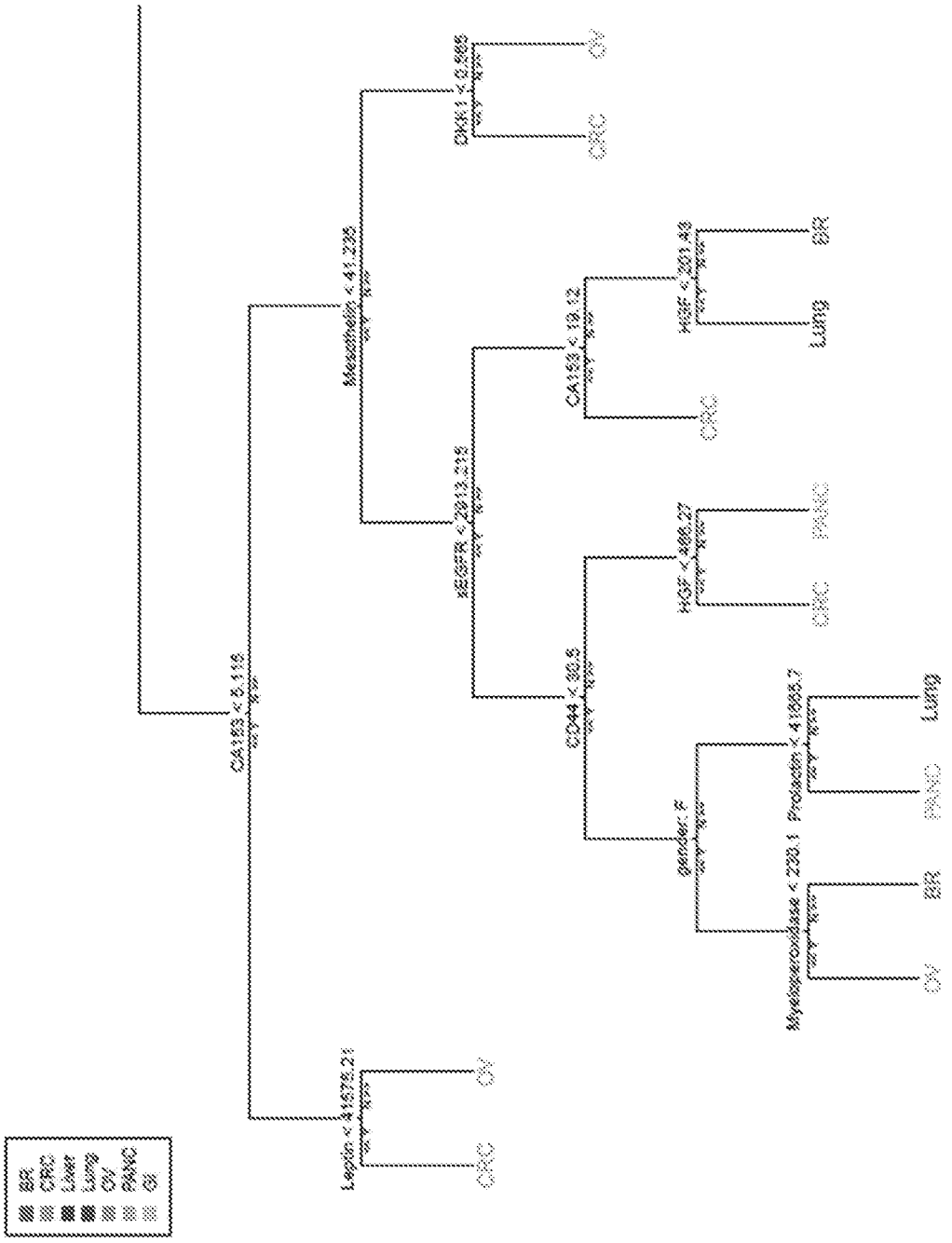

BR
CRC
Liver
Lung
OV
PANC
GI
CA153 < 5.116
Leptin < 41575.21
CRC
OV
Mesothelin < 41.235
DKK1 < 0.585
CRC
OV
sEGFR < 2913.215
CA153 < 19.12
CRC
HGF < 201.43
Lung
BR
CD44 < 30.5
HGF < 466.27
CRC
PANC
gender: F
Myeloperoxidase < 230.1
Prolactin < 41565.7
OV
BR
PANC
Lung

FIG. 68B

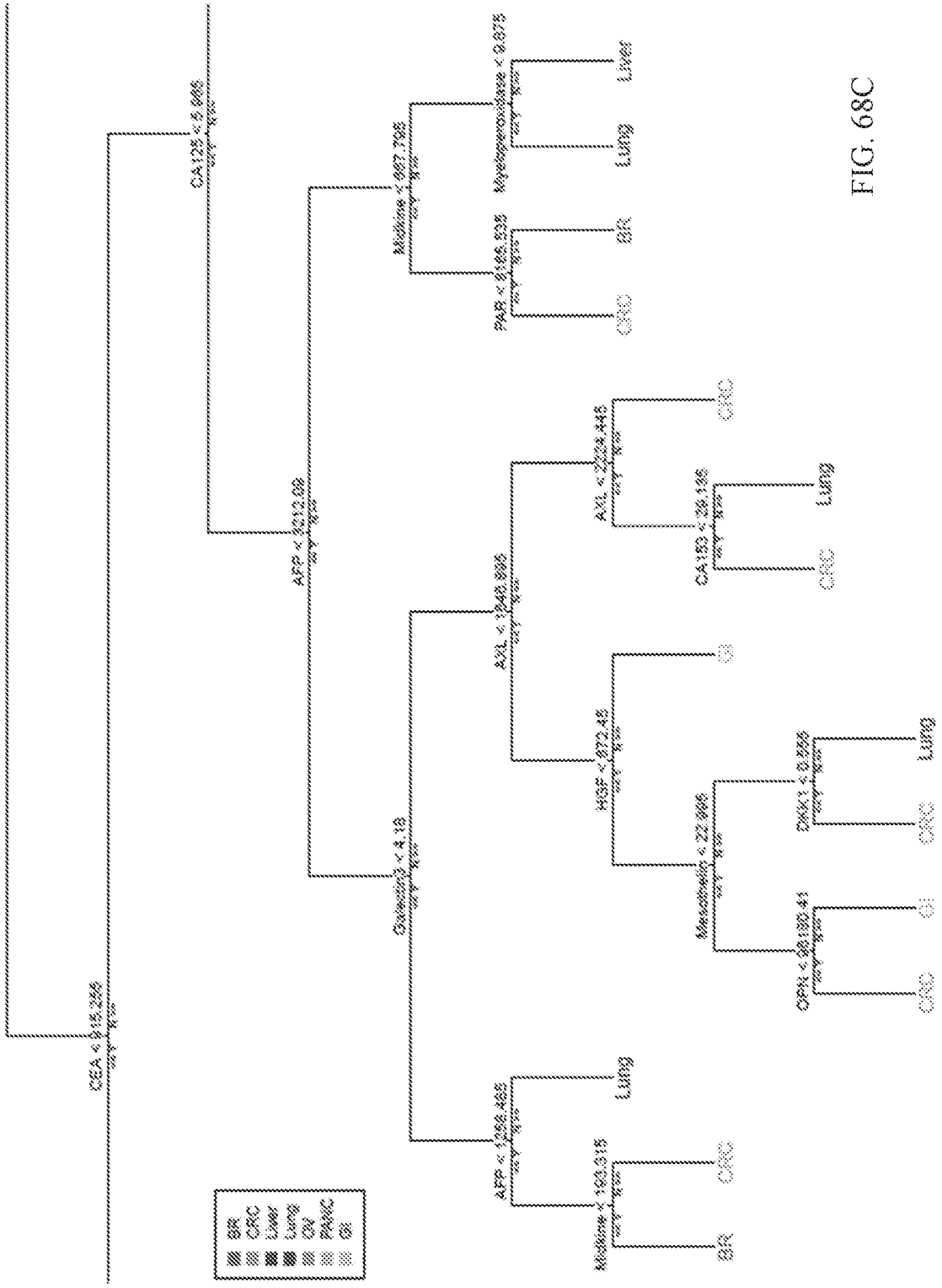

CEA < 9.15.255
CA125 < 5.965
AFP < 3312.09
Galectin3 < 4.18
Midkine < 667.795
PAR < 8165.535
Myeloperoxidase < 9.875
AXL < 1848.995
AXL < 2224.445
CA153 < 28.135
HGF < 672.45
Mesothelin < 22.995
OPN < 98180.41
DKK1 < 0.555
AFP < 1258.465
Midkine < 193.315
BR
CRC
Liver
Lung
OV
PANC
GI

FIG. 68C

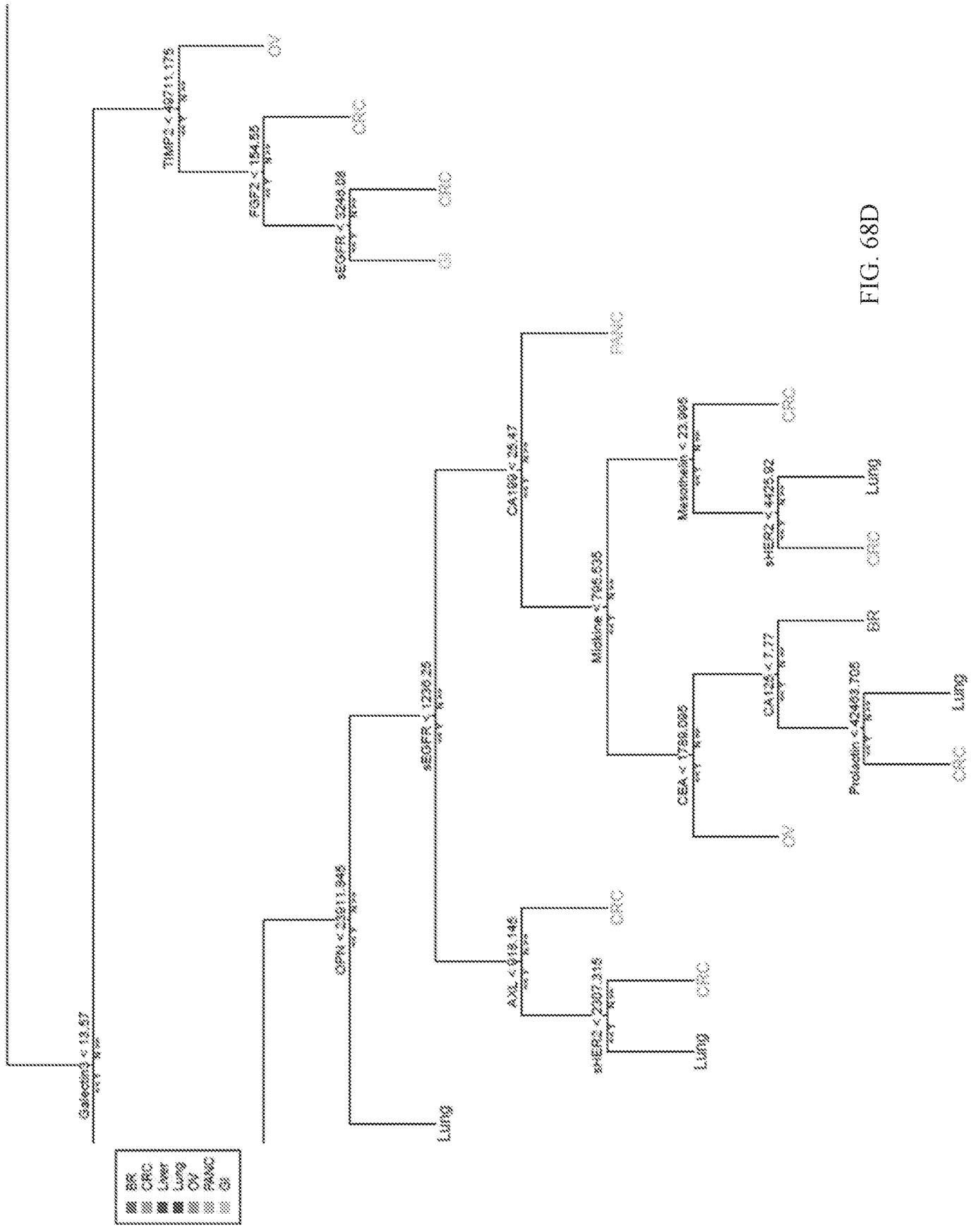

BR
CRC
Liver
Lung
OV
PANC
GI
Galectin3 < 13.57
TIMP2 < 49711.175
FGF2 < 154.55
sEGFR < 3248.08
OPN < 23911.945
sEGFR < 1236.25
AXL < 919.145
sHER2 < 2307.315
CA199 < 25.47
Midkine < 795.535
CEA < 1789.095
CA125 < 7.77
Proladin < 42453.705
Mesothelin < 23.995
sHER2 < 4425.92
Lung
CRC
OV
BR
GI
PANC

FIG. 68D

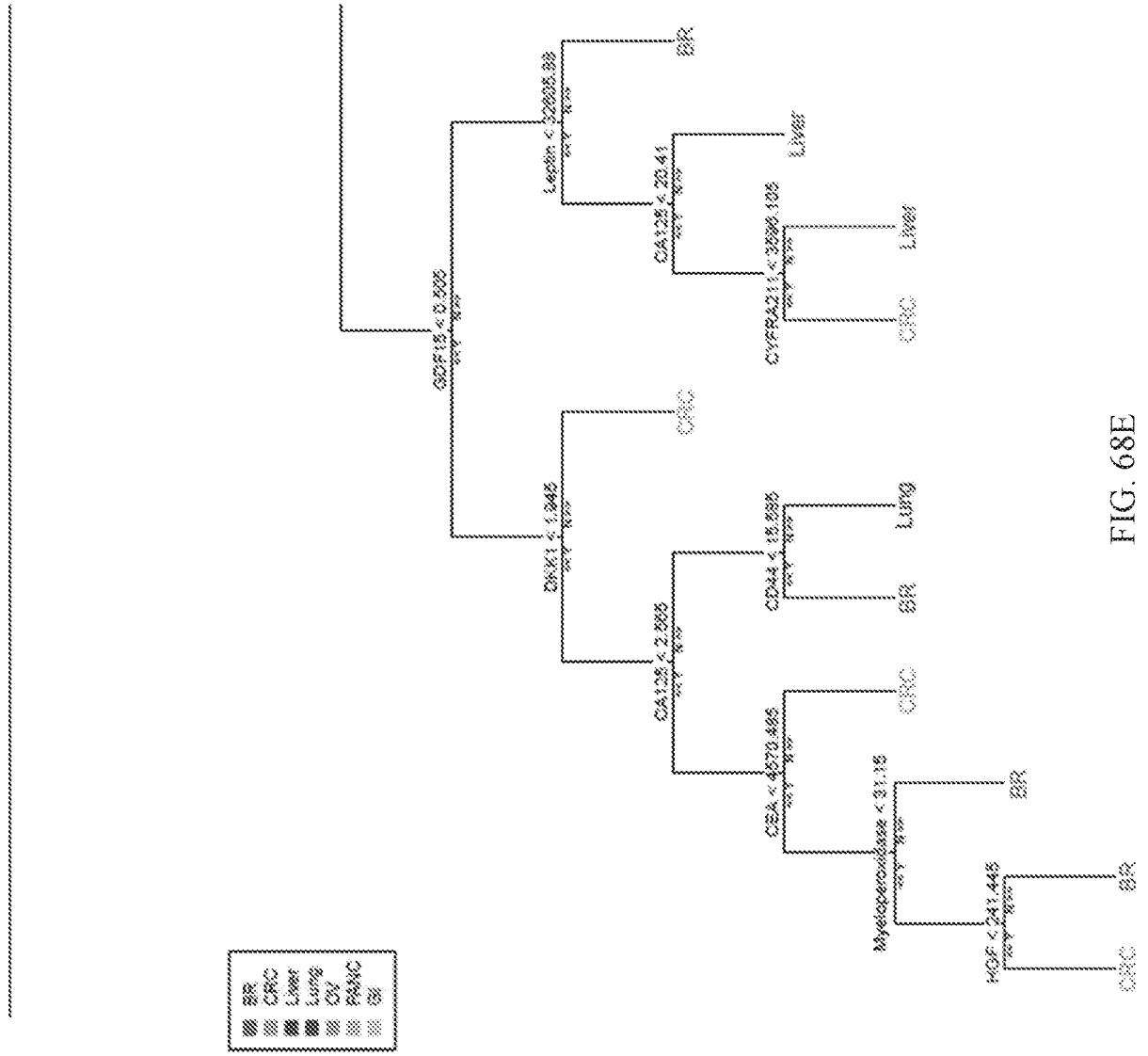

BR
CRC
Liver
Lung
OV
PANC
GDF15 < 0.805
Lectin < 32605.88
BR
CA125 < 20.41
Liver
CYFRA21-1 < 3286.105
Liver
CRC
DKK1 < 1.845
CRC
CD44 < 15.885
Lung
BR
CA125 < 2.805
CRC
CEA < 4670.465
CRC
Myeloperoxidase < 31.15
BR
HGF < 241.445
BR
CRC

FIG. 68E

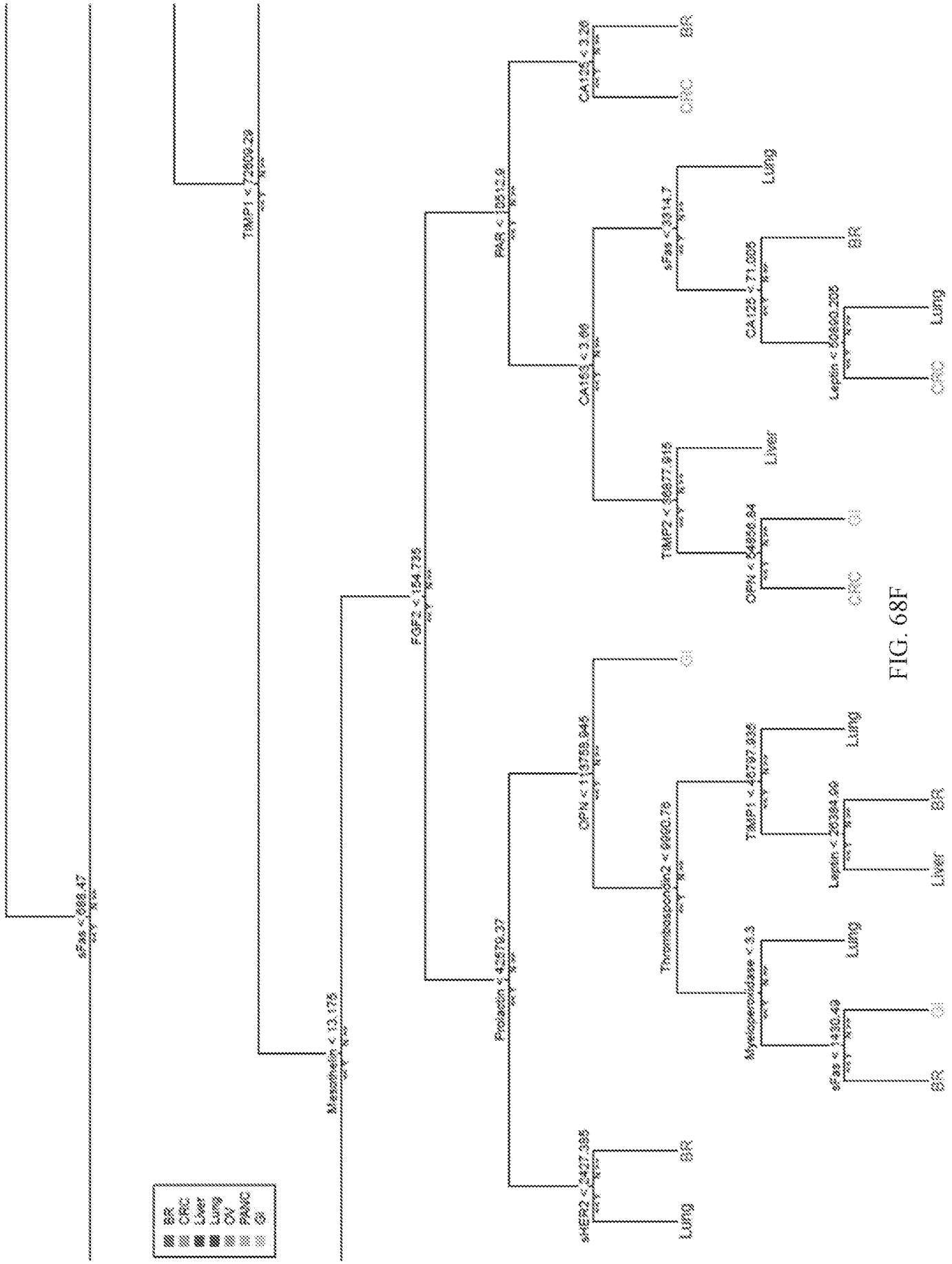

BR
CRC
Liver
Lung
OV
PANC
GI
sFas < 688.47
TIMP1 < 72609.29
Mesothelin < 13.175
FGF2 < 154.735
Prolactin < 42579.37
sHER2 < 2427.385
OPN < 113759.945
Thrombospondin2 < 9990.76
Myeloperoxidase < 3.3
sFas < 1430.49
TIMP1 < 46797.936
Leptin < 26384.99
PAR < 10512.9
CA153 < 3.66
TIMP2 < 36877.915
OPN < 54858.84
sFas < 3614.7
CA125 < 71.005
Leptin < 50690.205
CA125 < 3.26
Lung
BR
CRC
GI
Liver

FIG. 68F

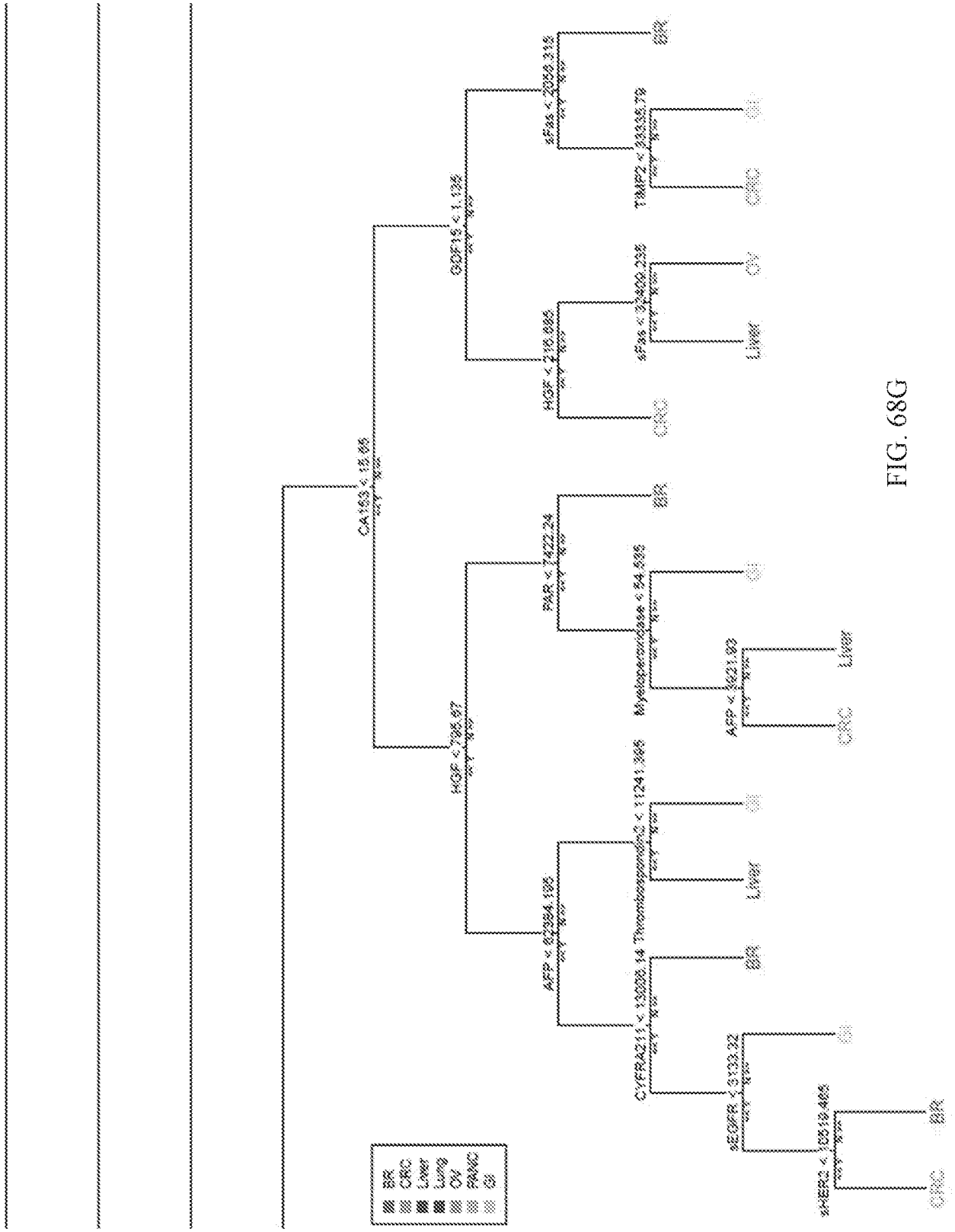

BR
CRC
Liver
Lung
OV
PANC
GI
CA153 < 15.65
HGF < 795.67
GDF15 < 1.135
AFP < 62384.195
PAR < 7422.24
HGF < 216.885
sFas < 2056.315
CYFRA211 < 13006.14
Thrombospondin2 < 11241.395
Myeloperoxidase < 54.585
sEGFR < 3133.32
sHER2 < 10519.465
BR
Liver
CRC
GI

FIG. 68G

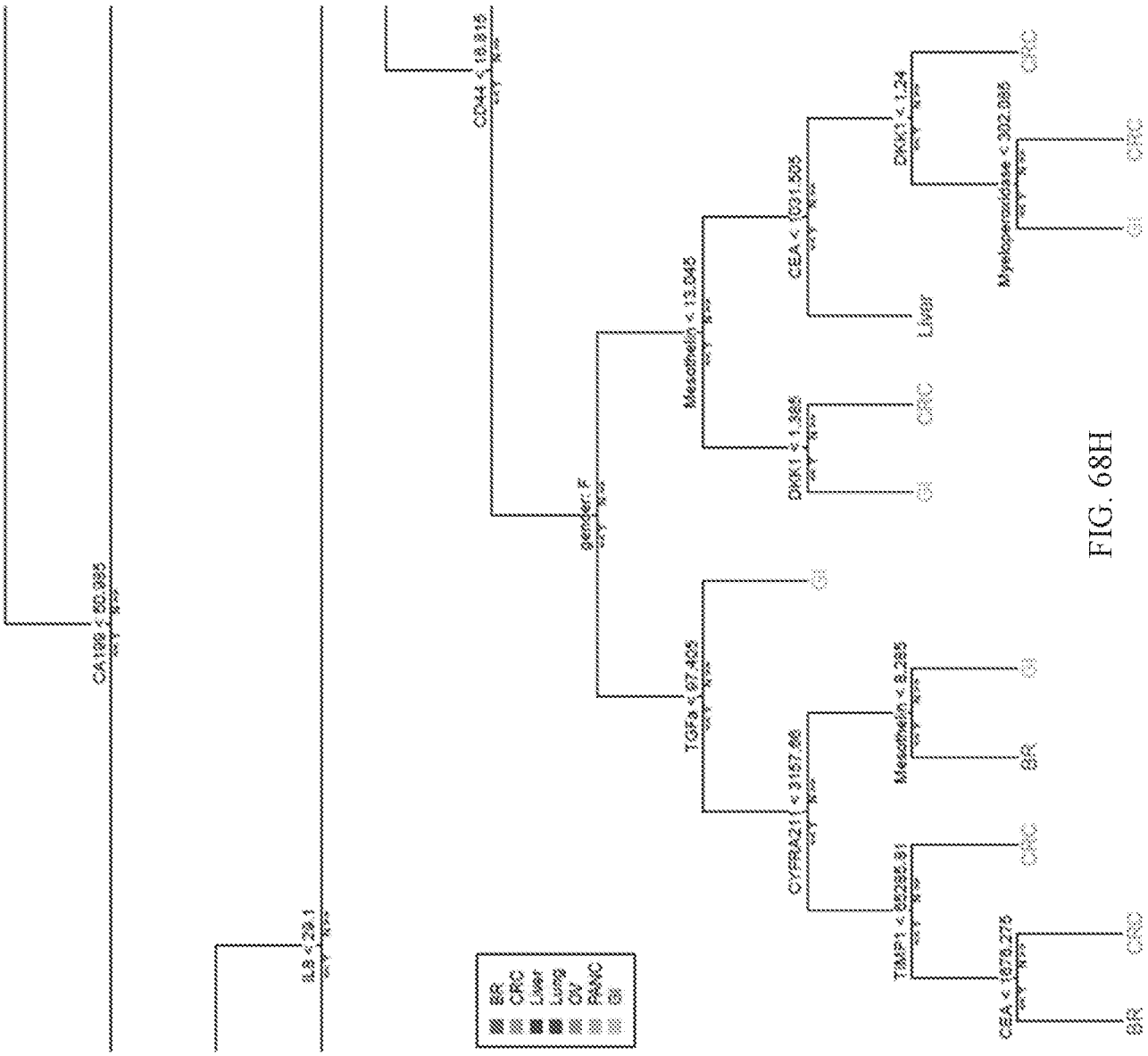

BR
CRC
Liver
Lung
OV
PANC
GI
CA199 < 50.995
ILB < 29.1
CD44 < 18.815
gender: F
TGFa < 97.405
Mesothelin < 13.045
CYFRA21 < 3157.86
DKK1 < 1.385
CEA < 1031.505
DKK1 < 1.24
Myeloperoxidase < 302.085
Mesothelin < 8.285
TIMP1 < 65285.91
CEA < 1878.275
Liver

FIG. 68H

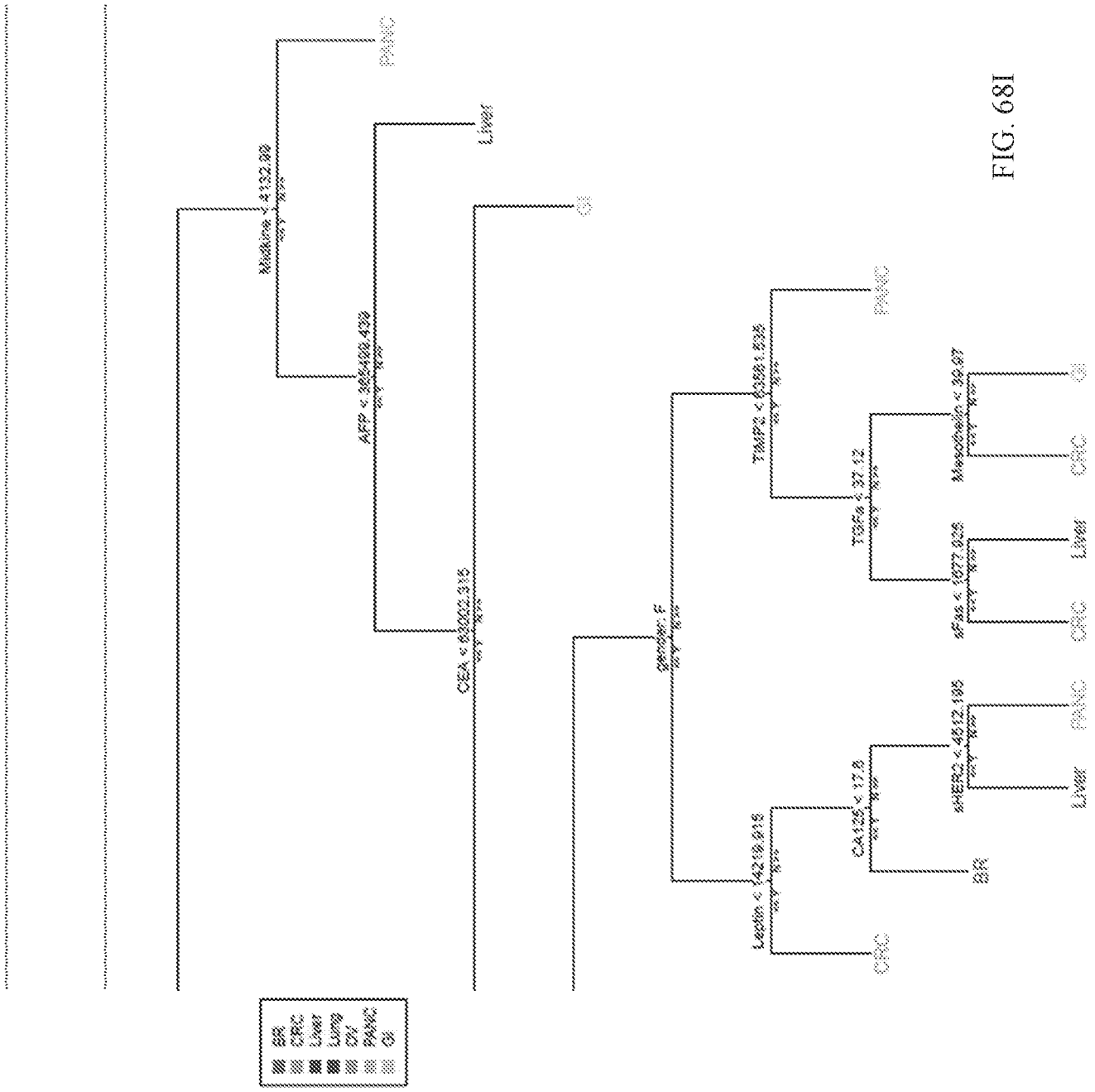

BR
CRC
Liver
Lung
OV
PANC
GI
Midkine < 4132.99
AFP < 365499.439
CEA < 63002.315
gender: F
Leptin < 14219.915
CA125 < 17.8
sHER2 < 4512.195
sFas < 1677.925
TGFa < 37.12
TIMP2 < 63561.535
Mesothelin < 39.97
CRC
BR
Liver
PANC
CRC
Liver
CRC
GI
PANC

FIG. 68I

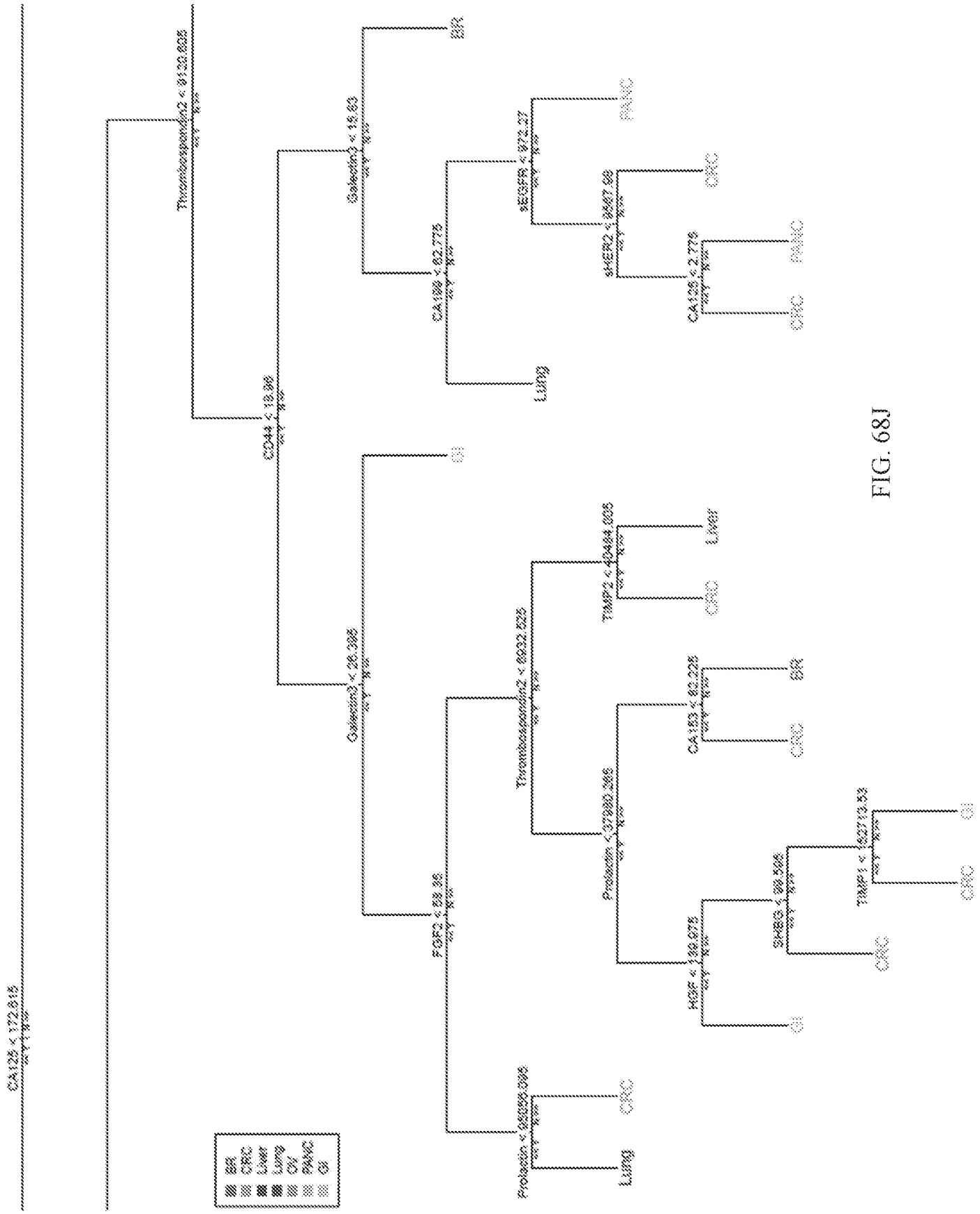

CA125 < 172.615
Thrombospondin2 < 9130.605
CD44 < 18.96
Galectin3 < 15.63
Galectin3 < 26.395
CA199 < 62.775
sEGFR < 972.27
sHER2 < 9567.98
CA125 < 2.775
Lung
BR
PANC
CRC
GI
FGF2 < 58.35
Thrombospondin2 < 6932.525
TIMP2 < 40484.005
Liver
Prolactin < 37980.365
CA153 < 62.225
HGF < 139.975
SHBG < 99.595
TIMP1 < 152713.53
Prolactin < 95056.095
BR
CRC
Liver
Lung
OV
PANC
GI

FIG. 68J

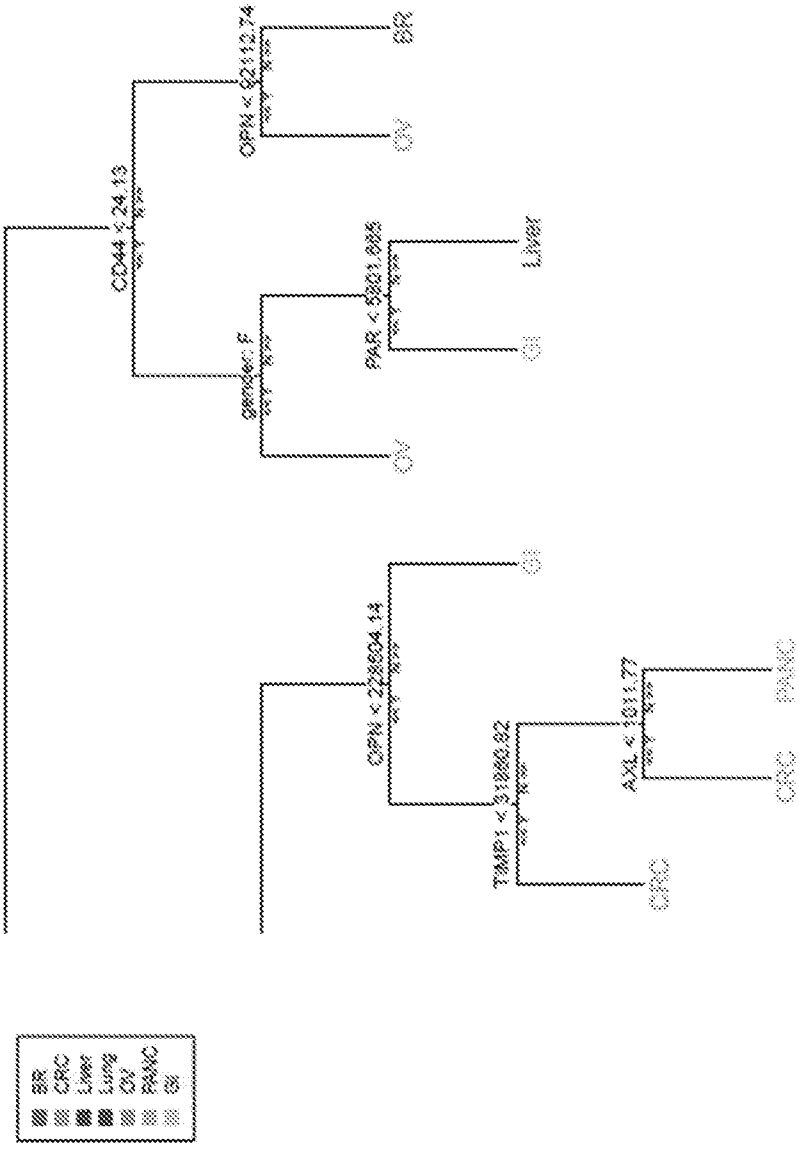

BR
CRC
Liver
Lung
OV
PANC
GI
CD44 < 24.13
OPN < 92112.74
BR
OV
gender: F
PAR < 5801.685
Liver
GI
OV
OPN < 228504.14
GI
TIMP1 < 31880.82
AXL < 1011.77
PANC
CRC
CRC

FIG. 68K

```
## R version 3.5.0 (2018-04-23)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 17134)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] caret_6.0-80        ggplot2_2.0.0       lattice_0.20-35
## [4] inTrees_1.2         DT_0.4              randomForest_4.6-14
##
## loaded via a namespace (and not attached):
##  [1] magic_1.5-8         ddalpha_1.3.4       tidyr_0.8.1
##  [4] sfsmisc_1.1-2       splines_3.5.0       foreach_1.4.4
##  [7] prodlim_2018.04.18  assertthat_0.2.0    stats4_3.5.0
## [10] arules_1.6-1        DRR_0.0.3           yaml_2.1.19
## [13] robustbase_0.93-1   ipred_0.9-6         pillar_1.2.3
## [16] backports_1.1.2     glue_1.2.0          digest_0.6.15
## [19] RRF_1.8             colorspace_1.2-2    recipes_0.1.3
## [22] gbm_2.1.3           htmltools_0.3.6     Matrix_1.2-14
## [25] plyr_1.8.4          psych_1.8.4         timeDate_3043.102
## [28] pkgconfig_2.0.1     CVST_0.2-2          broom_0.4.5
## [31] pcrrr_0.2.5         xtable_1.5-2        scales_0.5.0
## [34] gower_0.1.2         lava_1.6.2          tibble_1.4.2
## [37] xgboost_0.71.2      withr_2.1.2         nnet_7.3-12
## [40] lasyeval_0.2.1      mnormt_1.5-5        survival_2.42-4
## [43] magrittr_1.5        evaluate_0.11       nlme_3.1-137
## [46] MASS_7.3-50         dimRed_0.1.0        foreign_0.8-70
## [49] class_7.3-14        tools_3.5.0         data.table_1.11.4
## [52] stringr_1.3.1       kernlab_0.9-36      munsell_0.5.0
## [55] bindrcpp_0.2.2      pls_2.6-0           compiler_2.5.0
## [58] RcppRoll_0.3.0      rlang_0.2.1         grid_3.5.0
## [61] iterators_1.0.9     htmlwidgets_1.2     rmarkdown_1.10
## [64] geometry_0.3-6      gtable_0.2.0        ModelMetrics_1.1.0
## [67] codetools_0.2-15    abind_1.4-5         reshape2_1.4.3
## [70] R6_2.2.2            lubridate_1.7.4     knitr_1.20
## [73] dplyr_0.7.6         bindr_0.1.1         rprojroot_1.3-2
## [76] stringi_1.2.3       parallel_3.5.0      Rcpp_0.12.17
## [79] rpart_4.1-13        DEoptimR_1.0-8      tidyselect_0.2.4
```

FIG. 69

```
## [1] "Tumor samples: n = 628"
```

```
## [1] "Proteins: AFP, CA125, CA153, CA199, CEA, CYFRA211, FGF2, HE4, HGF, IL6, IL8, Leptin, OPN, Prolactin, sFa
s, TGFa, TIMP1, TIMP2, Galectin3, Myeloperoxidase, SHBG, DKK1, GDF15, NSE, OPG, CD44, Mesothelin, Midkine, AXL, P
AR, sEGFR, sHER2, Thrombospondin2"
```

```
## [1] "Predictions from randomForest:"
```

```
##
##   Breast      CRC    Liver     Lung  Ovarian Pancreas   st_eso
##       48      392       17       25       41       62       41
```

```
## 18635 rules (length<=6) were extracted from the first 500 trees.
```

```
## [1] "Predictions from inTrees classifier:"
```

```
## predictions
##   Breast      CRC    Liver     Lung  Ovarian Pancreas   st_eso
##       57      352       19       36       45       50       67
```

```
## [1] "Accuracies from the paper:"
```

```
##   Breast      CRC    Liver     Lung  Ovarian Pancreas   st_eso
##     0.63     0.54     0.44     0.39     0.79     0.81     0.46
```

```
## [1] "Accuracies from inTrees classifier:"
```

```
##    Breast       CRC     Liver      Lung   Ovarian  Pancreas    st_eso
## 0.6376812 0.8892281 0.3862636 0.4098361 0.8112208 0.7058824 0.6025641
```

FIG. 69 (continued)