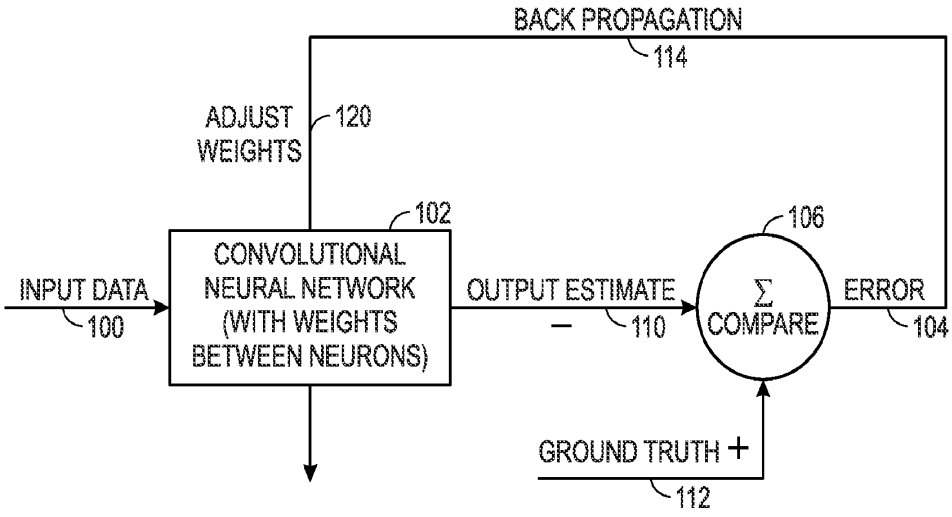
BACK PROPAGATION
114
ADJUST WEIGHTS
120
102
106
CONVOLUTIONAL NEURAL NETWORK (WITH WEIGHTS BETWEEN NEURONS)
INPUT DATA
100
OUTPUT ESTIMATE
110
Σ COMPARE
ERROR
104
GROUND TRUTH
112

FIG. 1

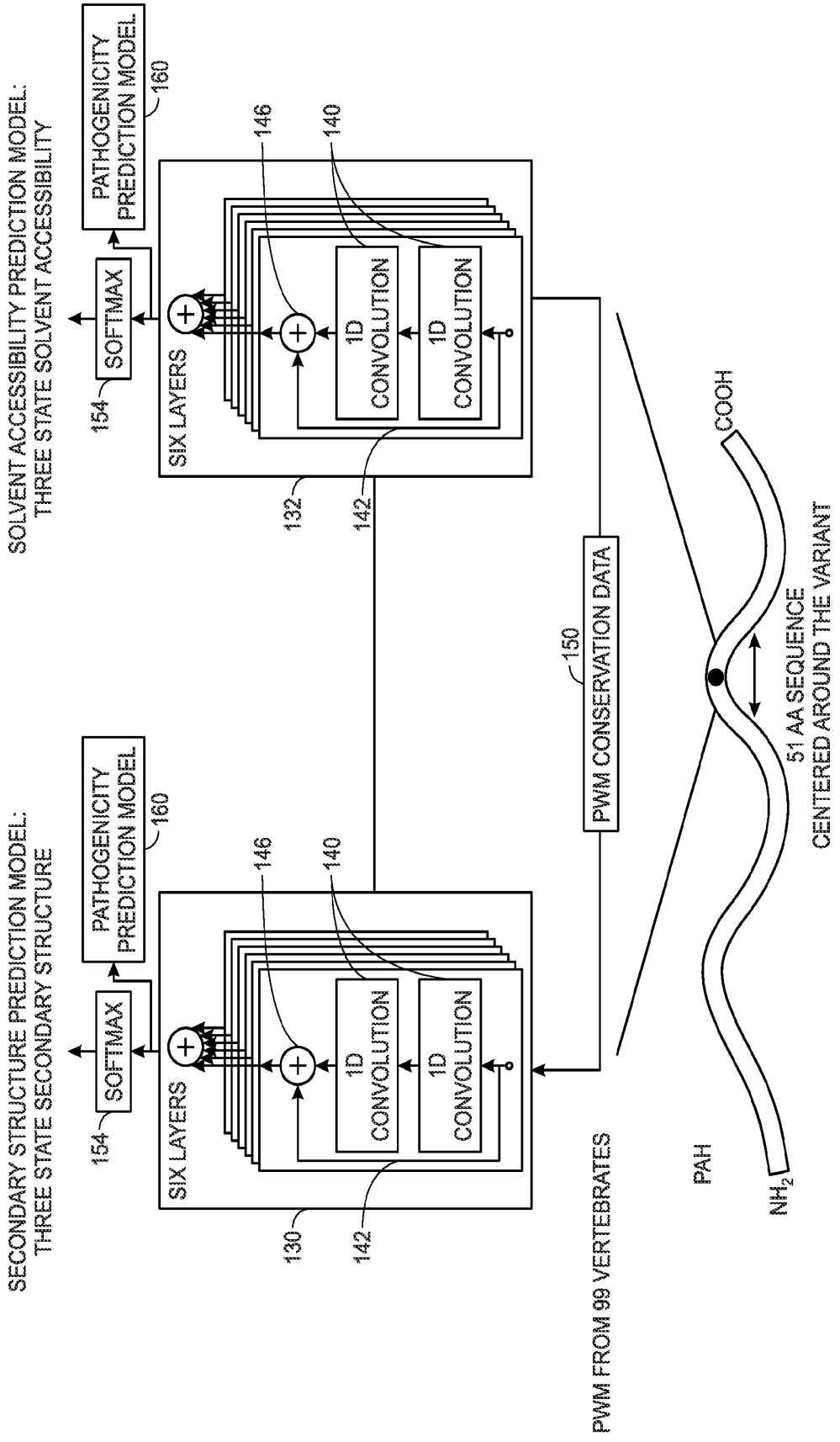
SECONDARY STRUCTURE PREDICTION MODEL:
THREE STATE SECONDARY STRUCTURE
SOLVENT ACCESSIBILITY PREDICTION MODEL:
THREE STATE SOLVENT ACCESSIBILITY
PATHOGENICITY PREDICTION MODEL
160
SOFTMAX
154
SIX LAYERS
1D CONVOLUTION
1D CONVOLUTION
146
140
130
132
142
PWM FROM 99 VERTEBRATES
PWM CONSERVATION DATA
150
COOH
PAH
NH_2
51 AA SEQUENCE
CENTERED AROUND THE VARIANT

FIG. 2

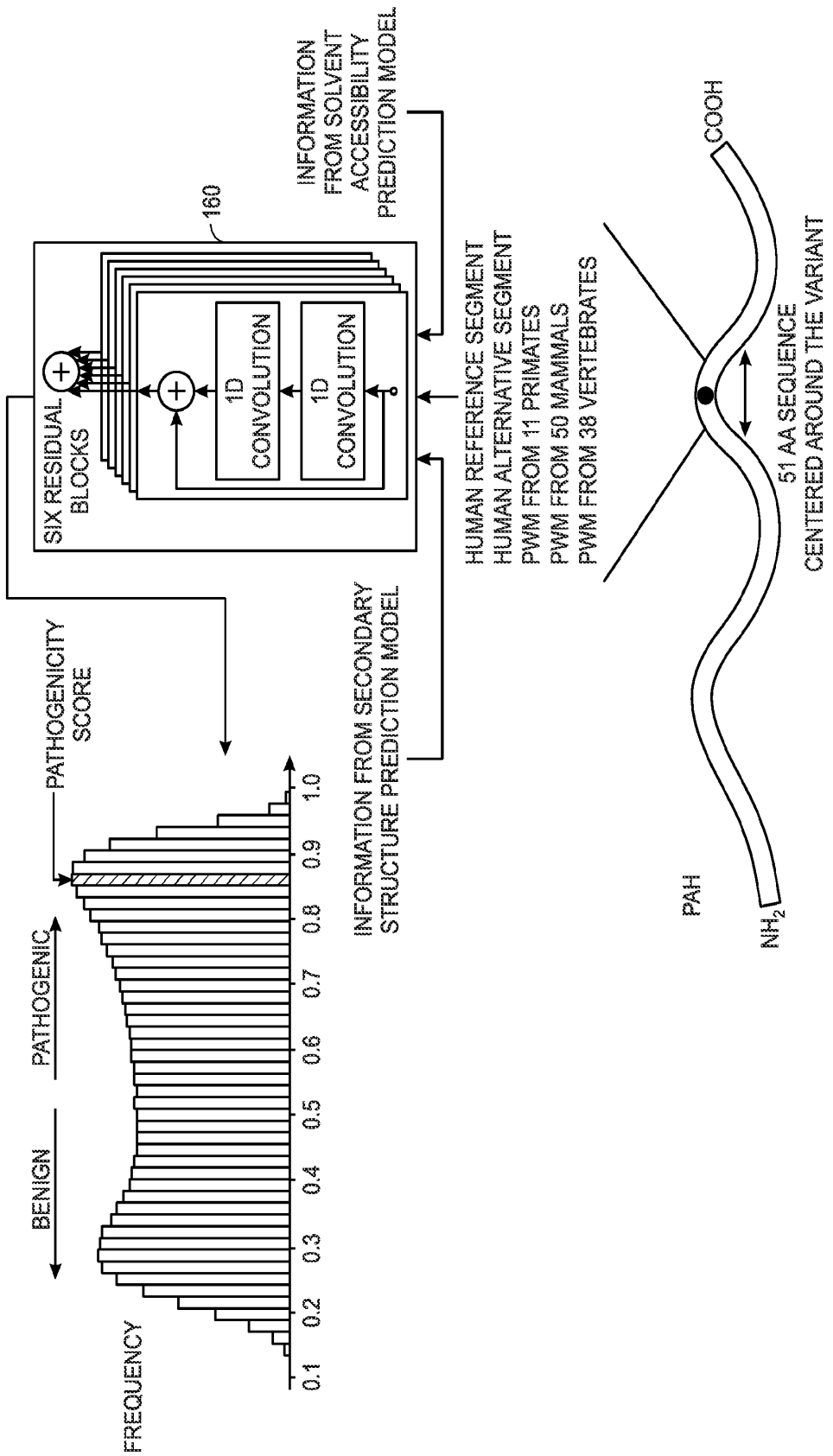
FREQUENCY
BENIGN
PATHOGENIC
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
PATHOGENICITY SCORE
INFORMATION FROM SECONDARY STRUCTURE PREDICTION MODEL
HUMAN REFERENCE SEGMENT
HUMAN ALTERNATIVE SEGMENT
PWM FROM 11 PRIMATES
PWM FROM 50 MAMMALS
PWM FROM 38 VERTEBRATES
INFORMATION FROM SOLVENT ACCESSIBILITY PREDICTION MODEL
160
SIX RESIDUAL BLOCKS
1D CONVOLUTION
1D CONVOLUTION
COOH
51 AA SEQUENCE CENTERED AROUND THE VARIANT
PAH
NH2

FIG. 3

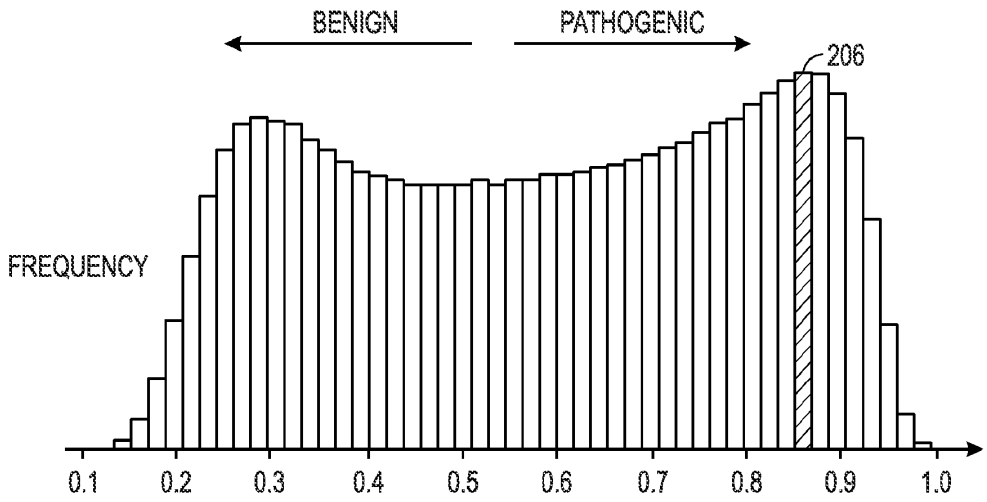

FIG. 4

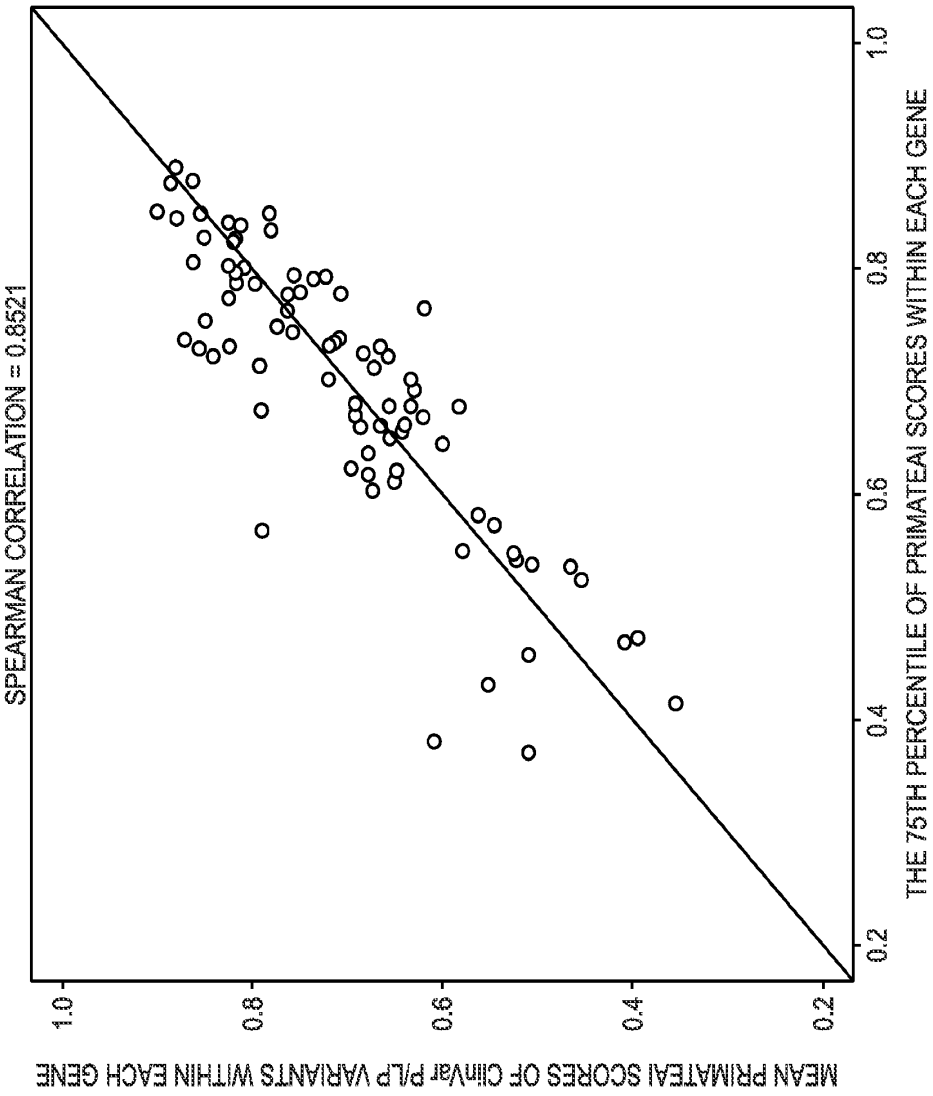
SPEARMAN CORRELATION = 0.8521
MEAN PRIMATEAI SCORES OF ClinVar P/LP VARIANTS WITHIN EACH GENE
1.0
0.8
0.6
0.4
0.2
THE 75TH PERCENTILE OF PRIMATEAI SCORES WITHIN EACH GENE
0.2
0.4
0.6
0.8
1.0

FIG. 5

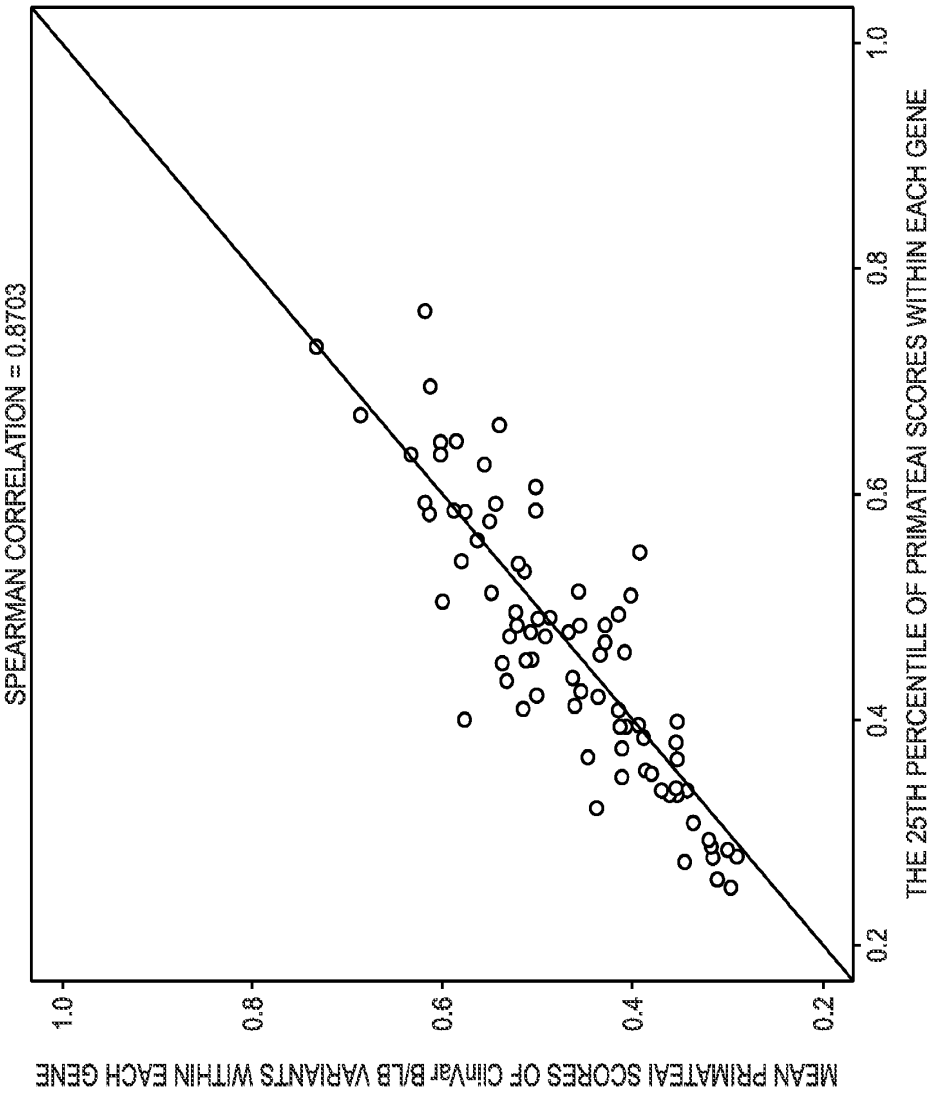
SPEARMAN CORRELATION = 0.8703
MEAN PRIMATEAI SCORES OF ClinVar B/LB VARIANTS WITHIN EACH GENE
THE 25TH PERCENTILE OF PRIMATEAI SCORES WITHIN EACH GENE
0.2
0.4
0.6
0.8
1.0

FIG. 6

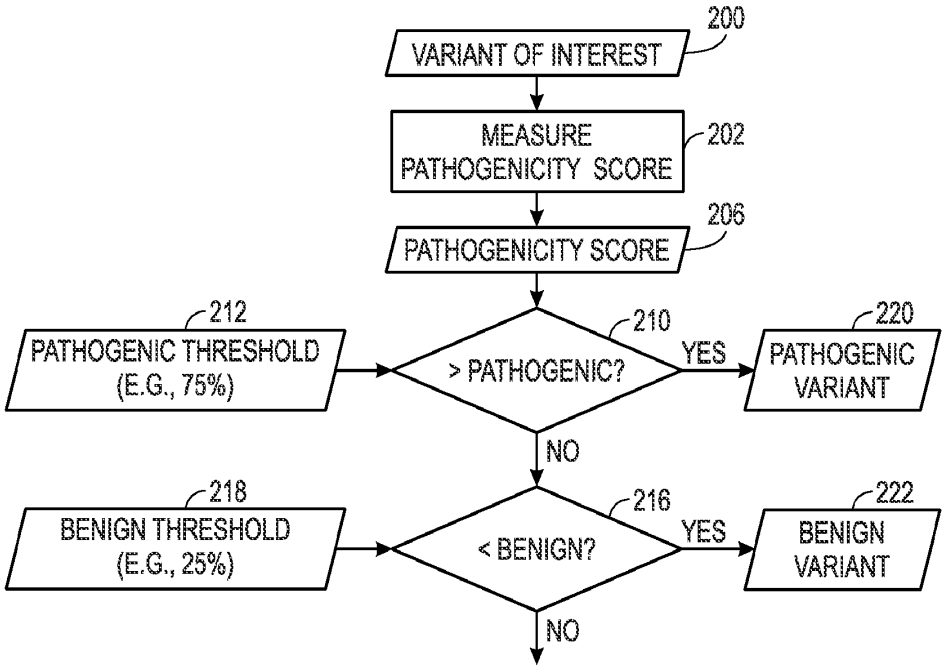
200
VARIANT OF INTEREST
MEASURE PATHOGENICITY SCORE
202
206
PATHOGENICITY SCORE
212
PATHOGENIC THRESHOLD (E.G., 75%)
210
> PATHOGENIC?
YES
220
PATHOGENIC VARIANT
NO
218
BENIGN THRESHOLD (E.G., 25%)
216
< BENIGN?
YES
222
BENIGN VARIANT
NO

FIG. 7

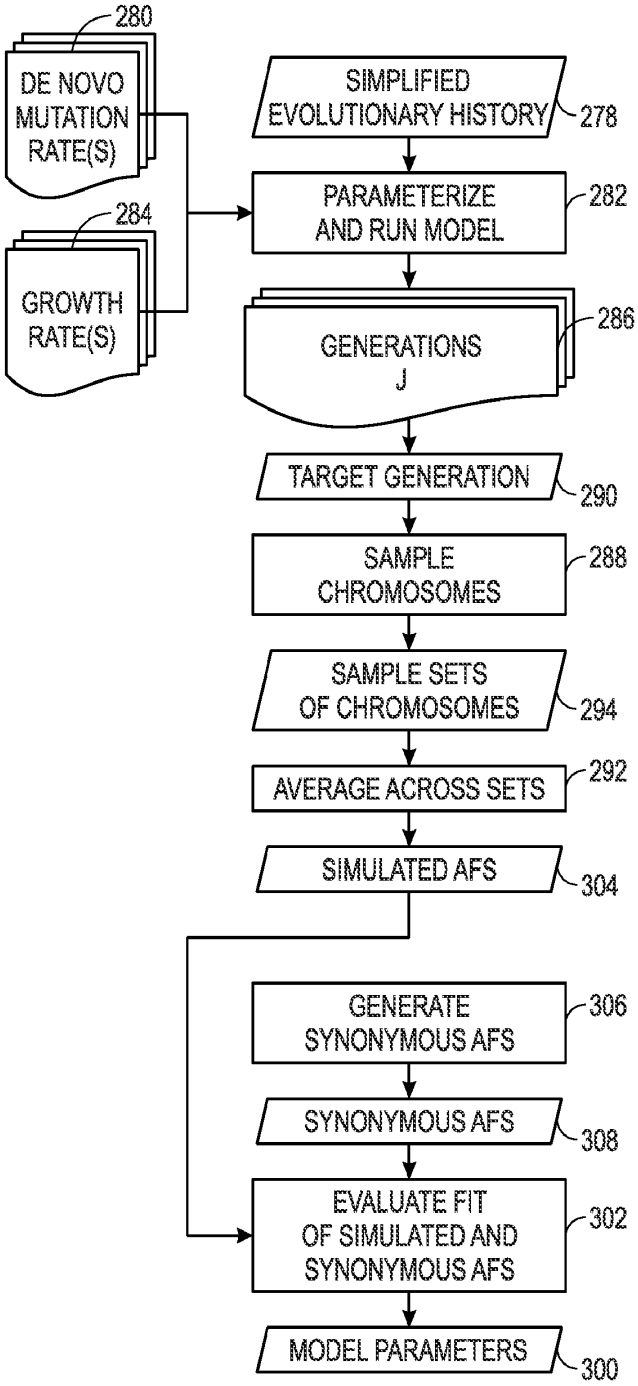
280
DE NOVO MUTATION RATE(S)
284
GROWTH RATE(S)
SIMPLIFIED EVOLUTIONARY HISTORY
278
PARAMETERIZE AND RUN MODEL
282
GENERATIONS J
286
TARGET GENERATION
290
SAMPLE CHROMOSOMES
288
SAMPLE SETS OF CHROMOSOMES
294
AVERAGE ACROSS SETS
292
SIMULATED AFS
304
GENERATE SYNONYMOUS AFS
306
SYNONYMOUS AFS
308
EVALUATE FIT OF SIMULATED AND SYNONYMOUS AFS
302
MODEL PARAMETERS
300

FIG. 8

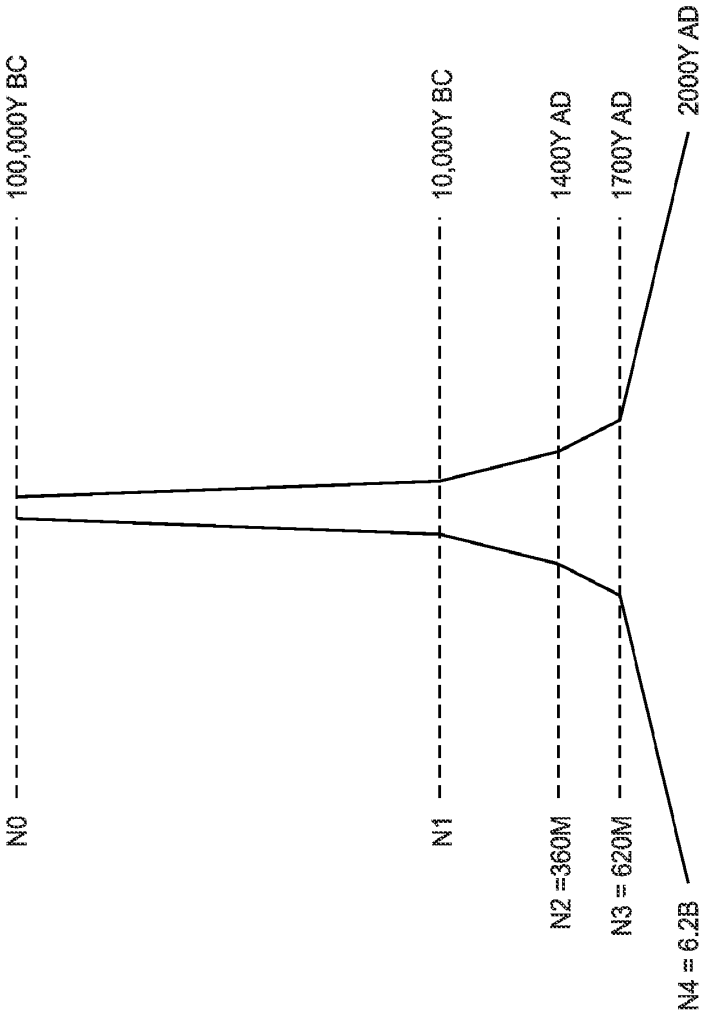
100,000Y BC
10,000Y BC
1400Y AD
1700Y AD
2000Y AD
N0
N1
N2 =360M
N3 = 620M
N4 = 6.2B

FIG. 9

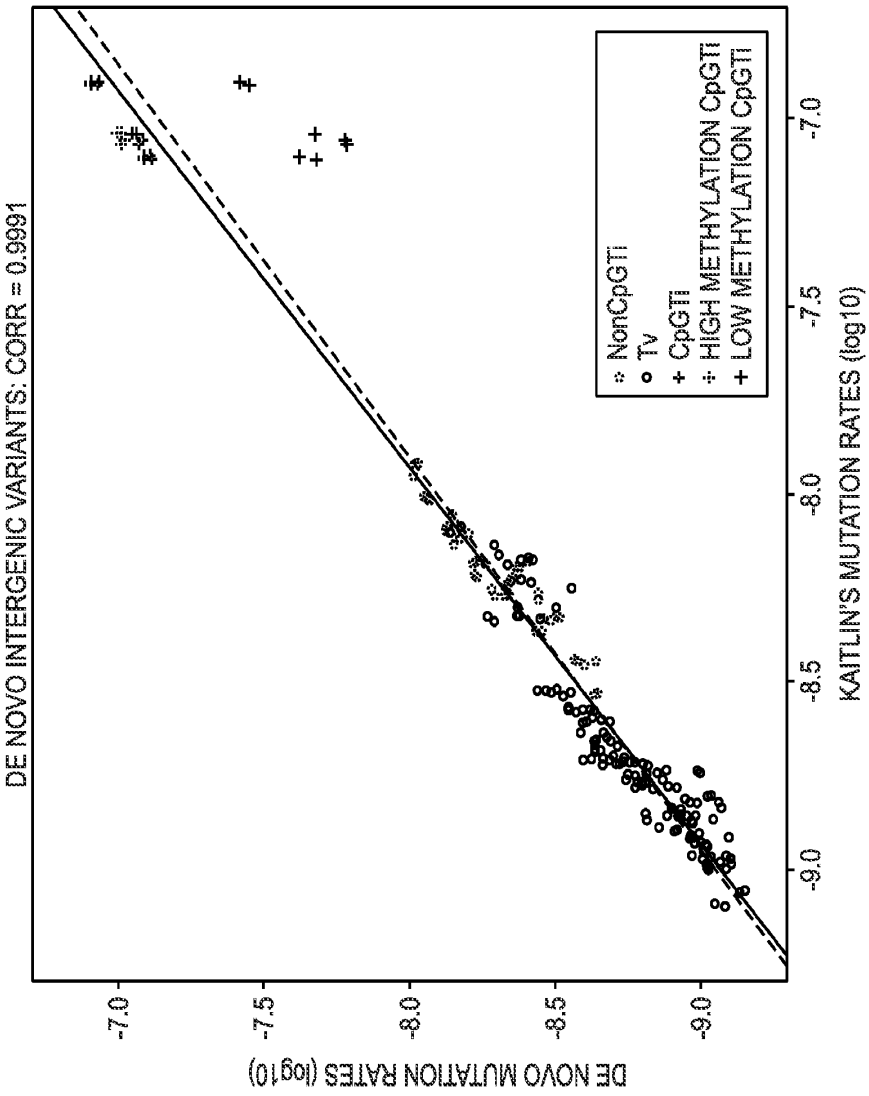
DE NOVO INTERGENIC VARIANTS: CORR = 0.9991
NonCpGTi
Tv
CpGTi
HIGH METHYLATION CpGTi
LOW METHYLATION CpGTi
DE NOVO MUTATION RATES (log10)
KAITLIN'S MUTATION RATES (log10)
-7.0
-7.5
-8.0
-8.5
-9.0

FIG. 10

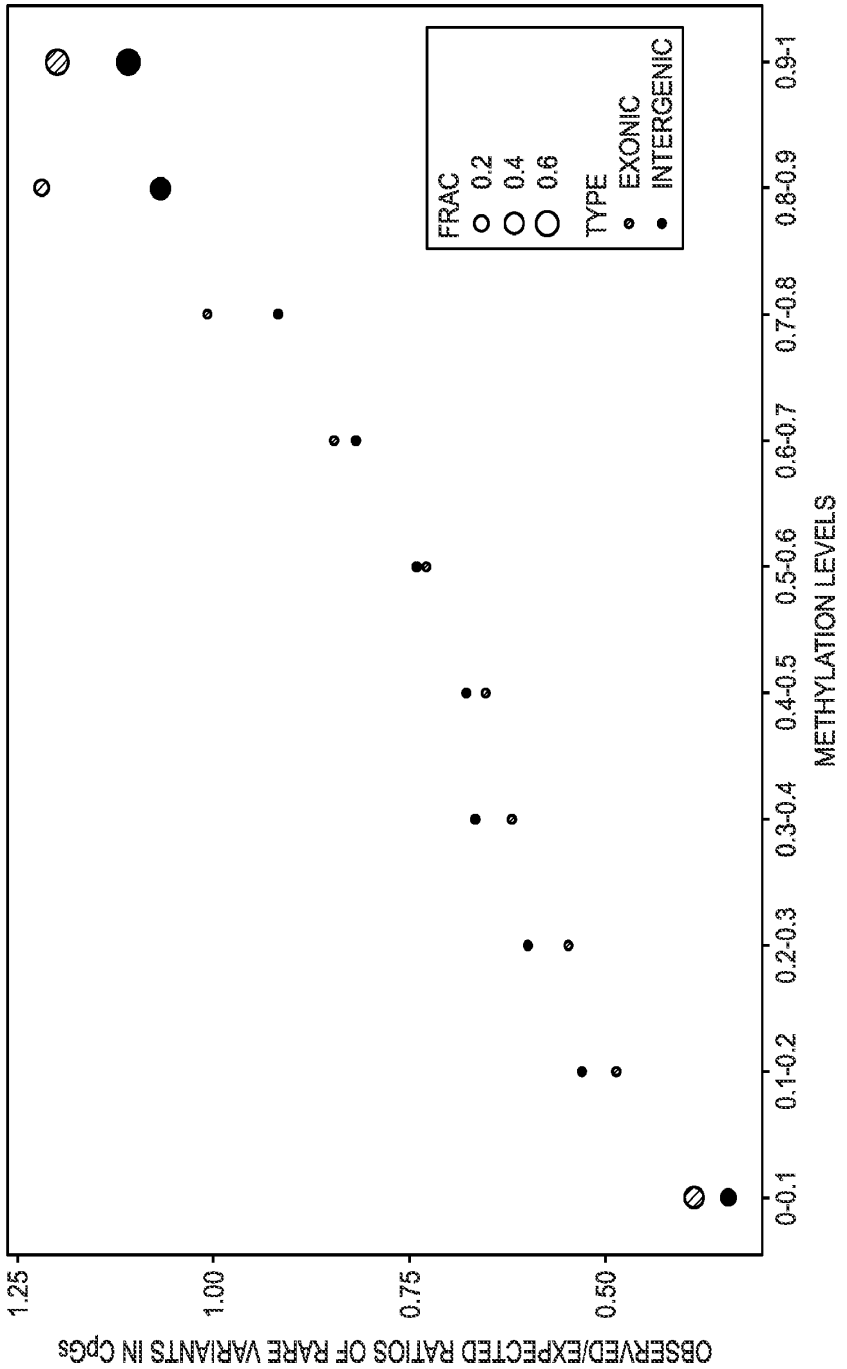
OBSERVED/EXPECTED RATIOS OF RARE VARIANTS IN CpGs
1.25
1.00
0.75
0.50
METHYLATION LEVELS
0-0.1
0.1-0.2
0.2-0.3
0.3-0.4
0.4-0.5
0.5-0.6
0.6-0.7
0.7-0.8
0.8-0.9
0.9-1
FRAC
0.2
0.4
0.6
TYPE
EXONIC
INTERGENIC

FIG. 11

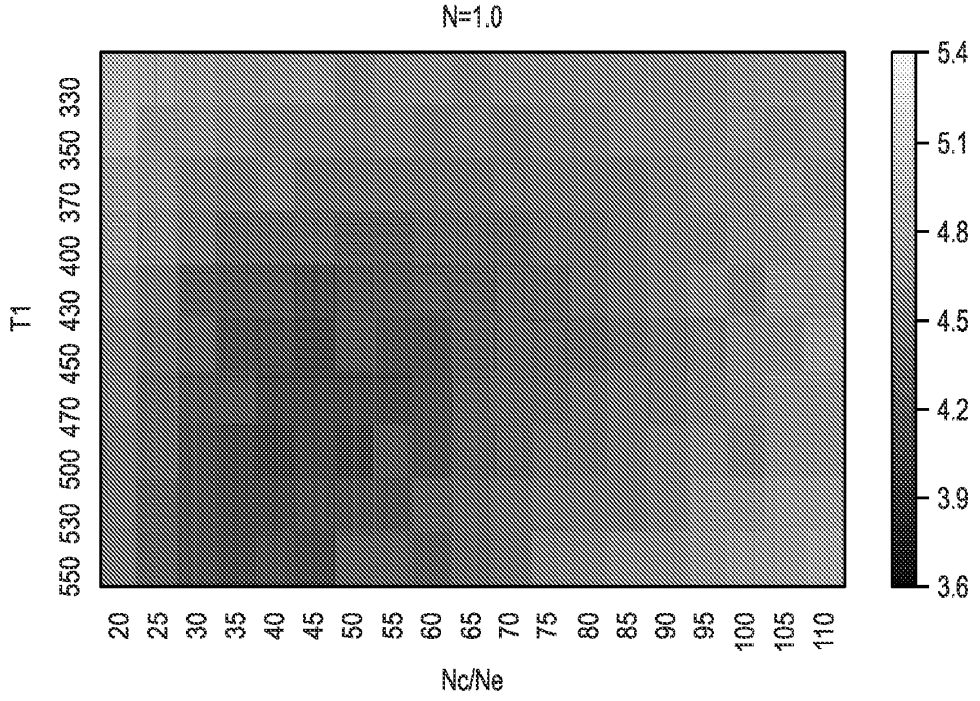

FIG. 12A

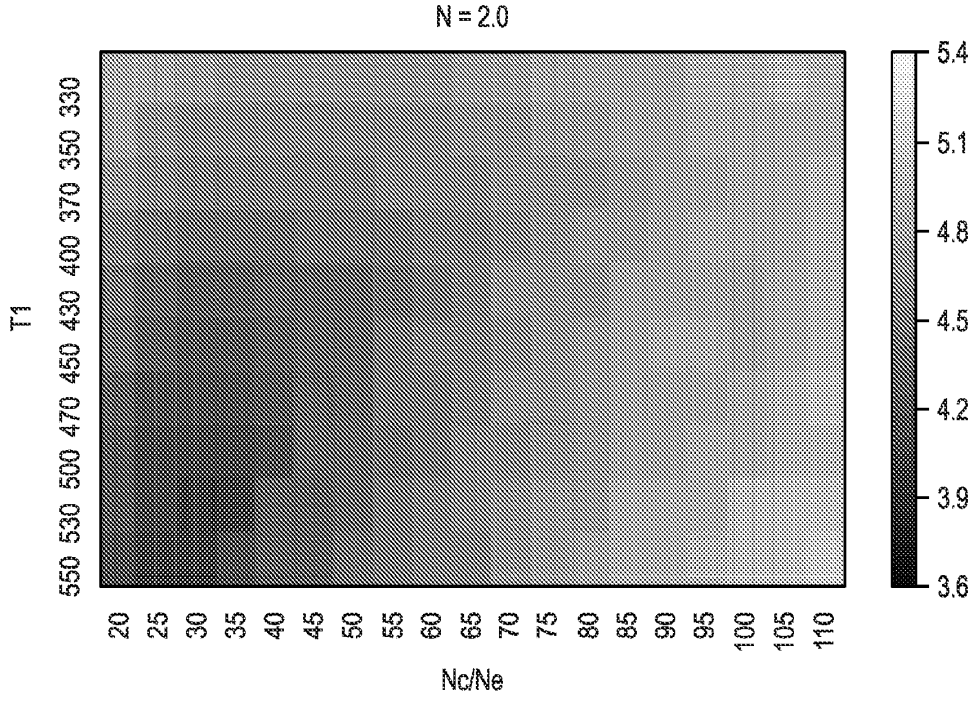

FIG. 12B

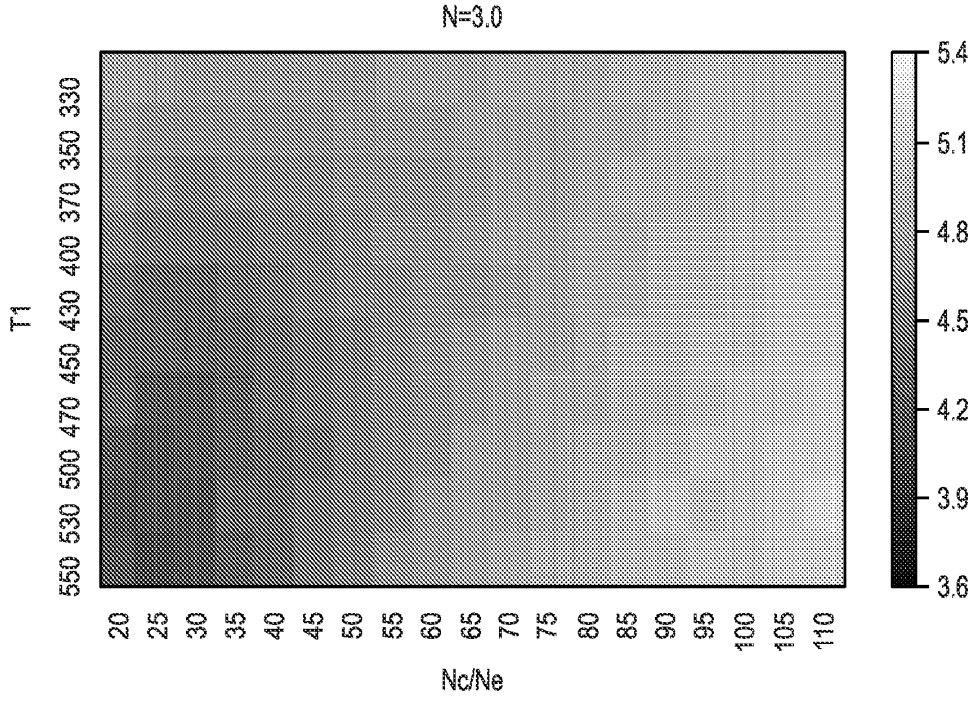

FIG. 12C

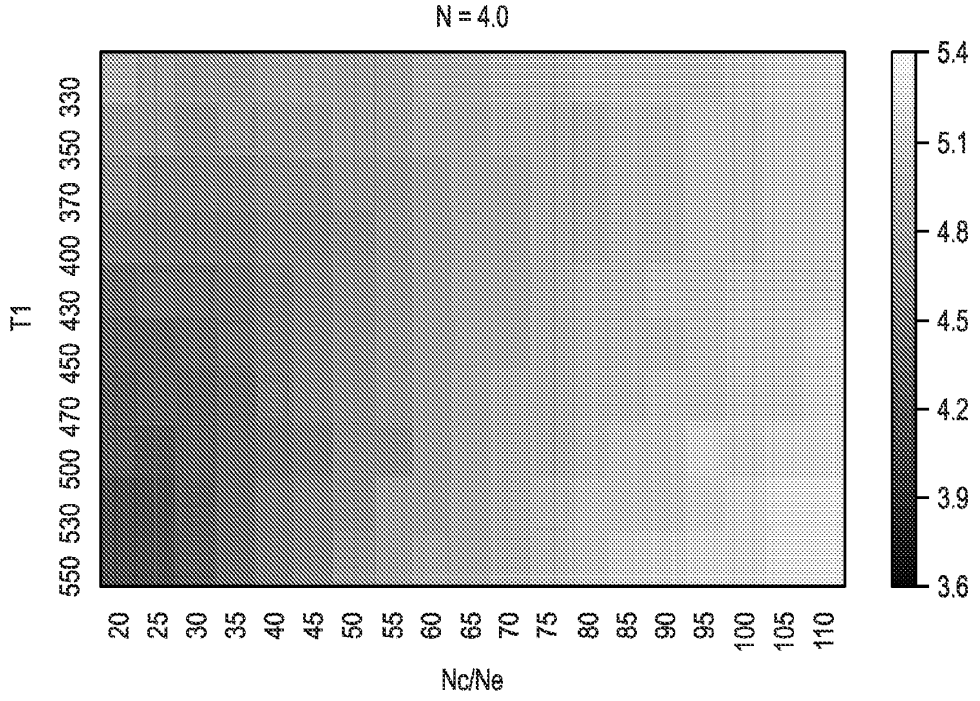

FIG. 12D

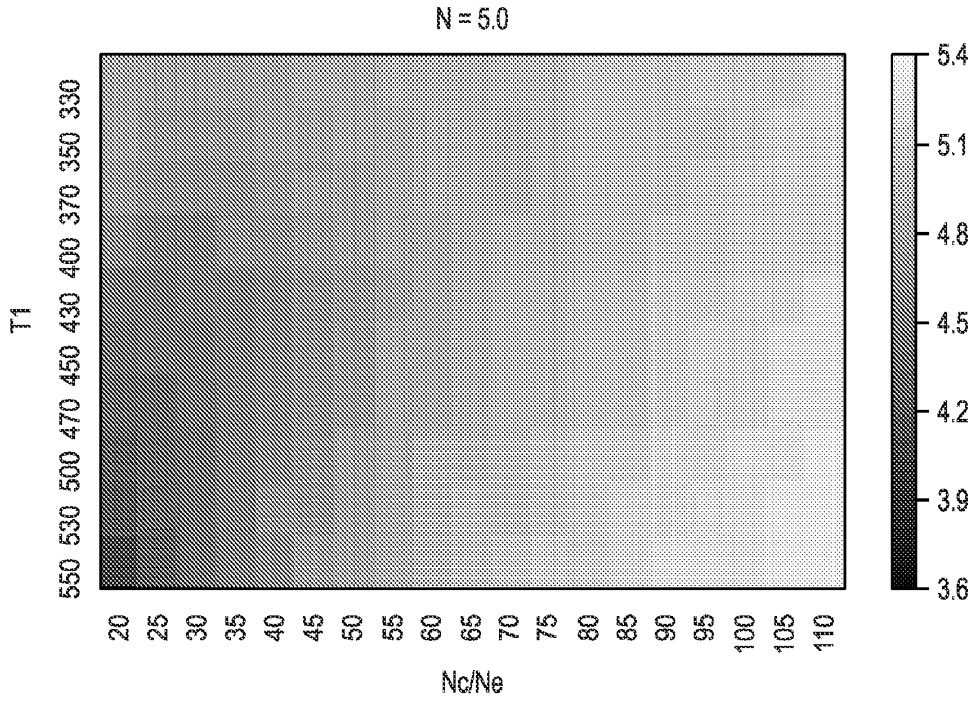
N = 5.0
T1
330 350 370 400 430 450 470 500 530 550
20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110
Nc/Ne
5.4
5.1
4.8
4.5
4.2
3.9
3.6

FIG. 12E

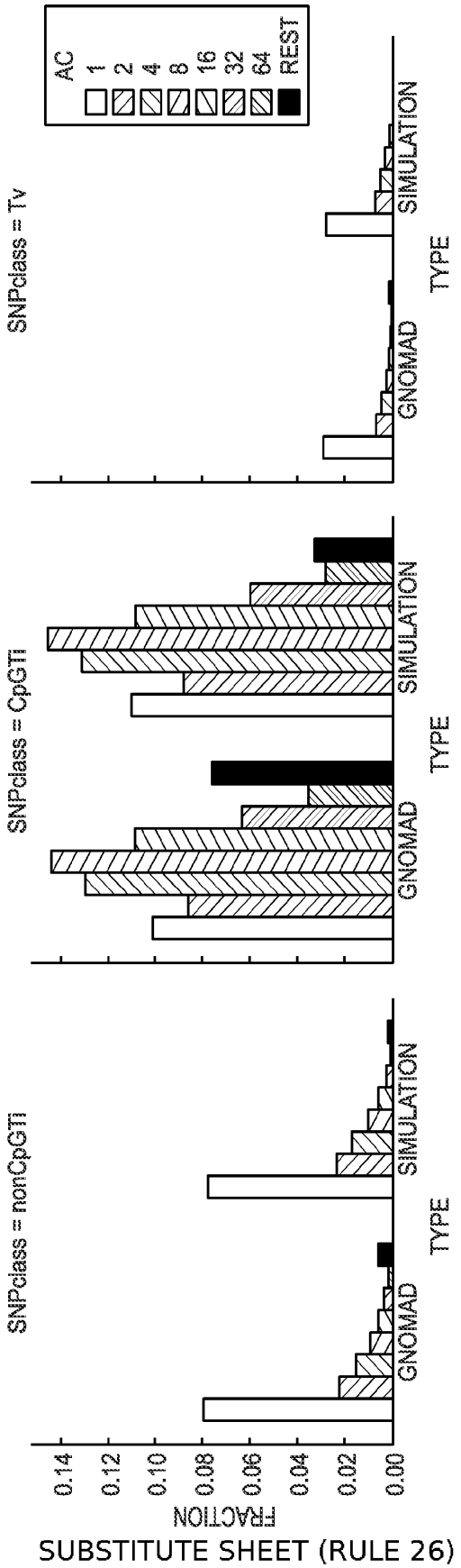
AC
1
2
4
8
16
32
64
REST
SNPclass = Tv
SNPclass = CpGTi
SNPclass = nonCpGTi
GNOMAD
SIMULATION
TYPE
FRACTION
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00

FIG. 13

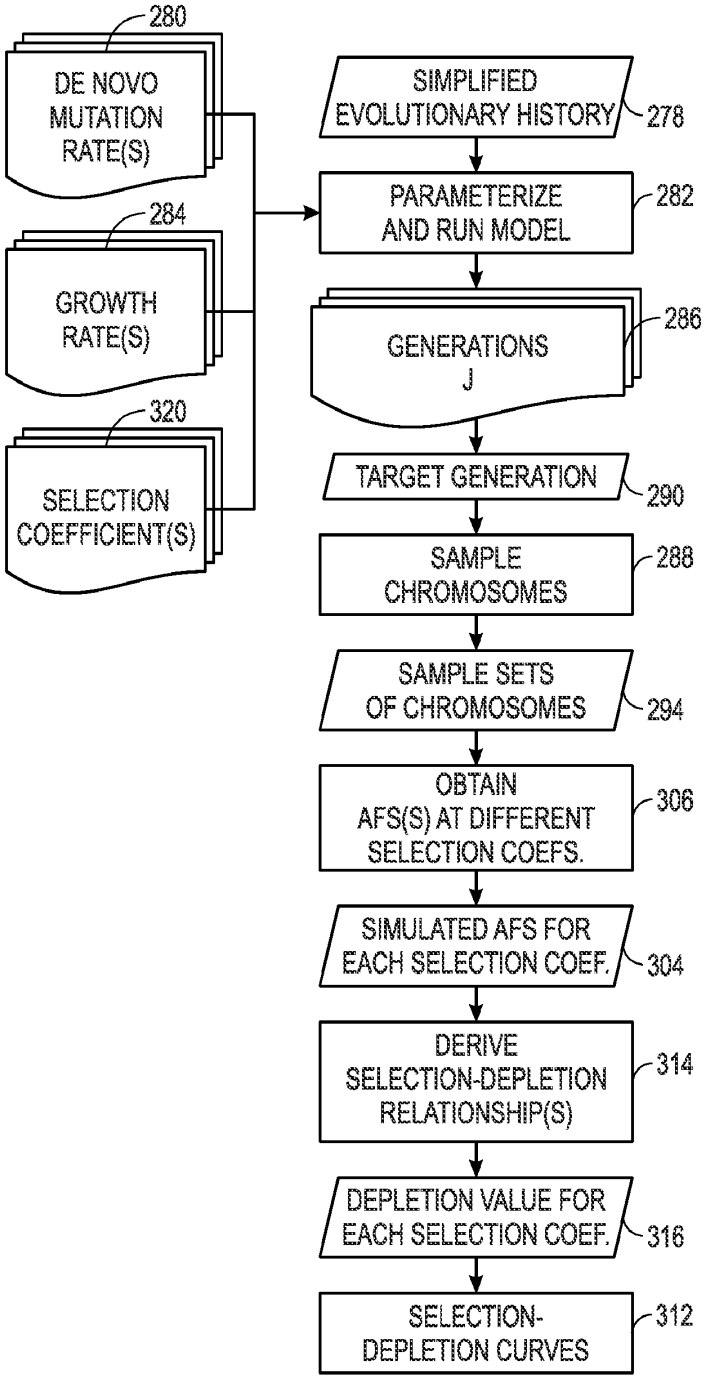
280
DE NOVO MUTATION RATE(S)
284
GROWTH RATE(S)
320
SELECTION COEFFICIENT(S)
SIMPLIFIED EVOLUTIONARY HISTORY
278
PARAMETERIZE AND RUN MODEL
282
GENERATIONS J
286
TARGET GENERATION
290
SAMPLE CHROMOSOMES
288
SAMPLE SETS OF CHROMOSOMES
294
OBTAIN AFS(S) AT DIFFERENT SELECTION COEFS.
306
SIMULATED AFS FOR EACH SELECTION COEF.
304
DERIVE SELECTION-DEPLETION RELATIONSHIP(S)
314
DEPLETION VALUE FOR EACH SELECTION COEF.
316
SELECTION-DEPLETION CURVES
312

FIG. 14

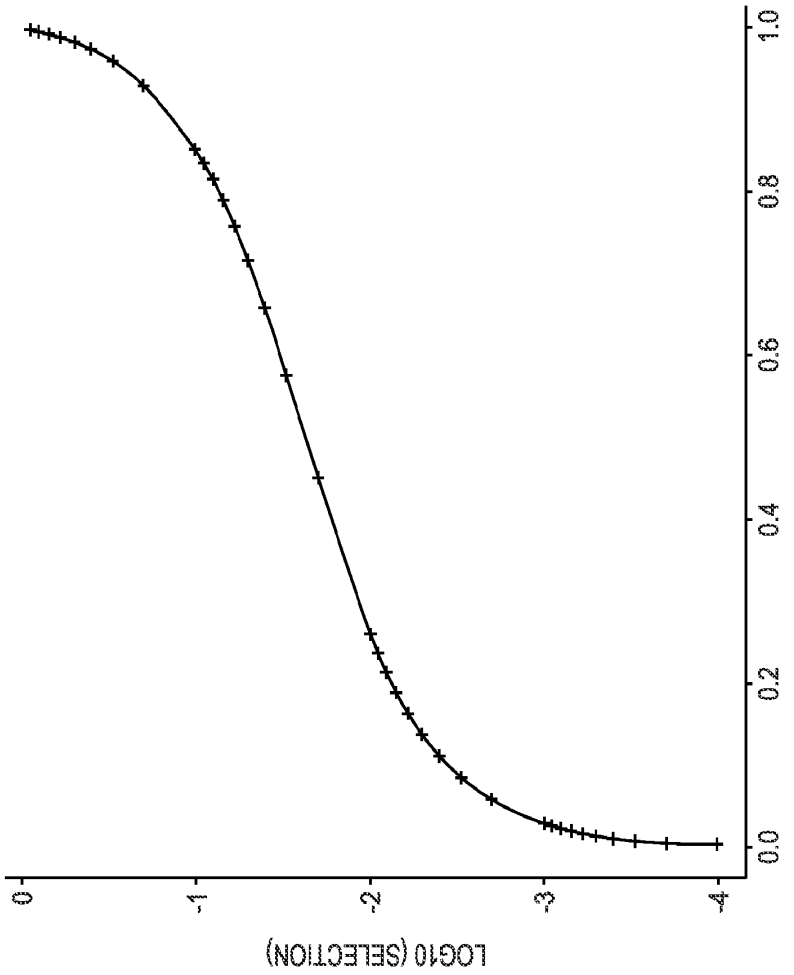

FIG. 15

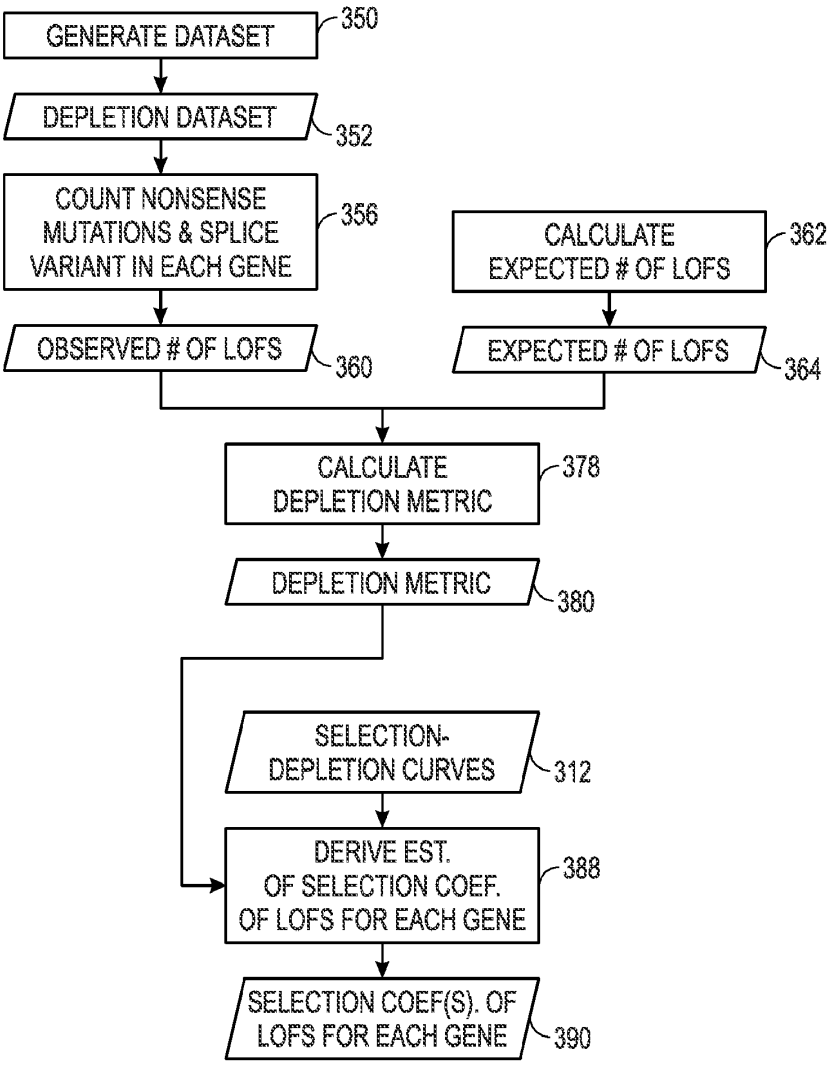
GENERATE DATASET
350
DEPLETION DATASET
352
COUNT NONSENSE MUTATIONS & SPLICE VARIANT IN EACH GENE
356
OBSERVED # OF LOFS
360
CALCULATE EXPECTED # OF LOFS
362
EXPECTED # OF LOFS
364
CALCULATE DEPLETION METRIC
378
DEPLETION METRIC
380
SELECTION-DEPLETION CURVES
312
DERIVE EST. OF SELECTION COEF. OF LOFS FOR EACH GENE
388
SELECTION COEF(S). OF LOFS FOR EACH GENE
390

FIG. 16

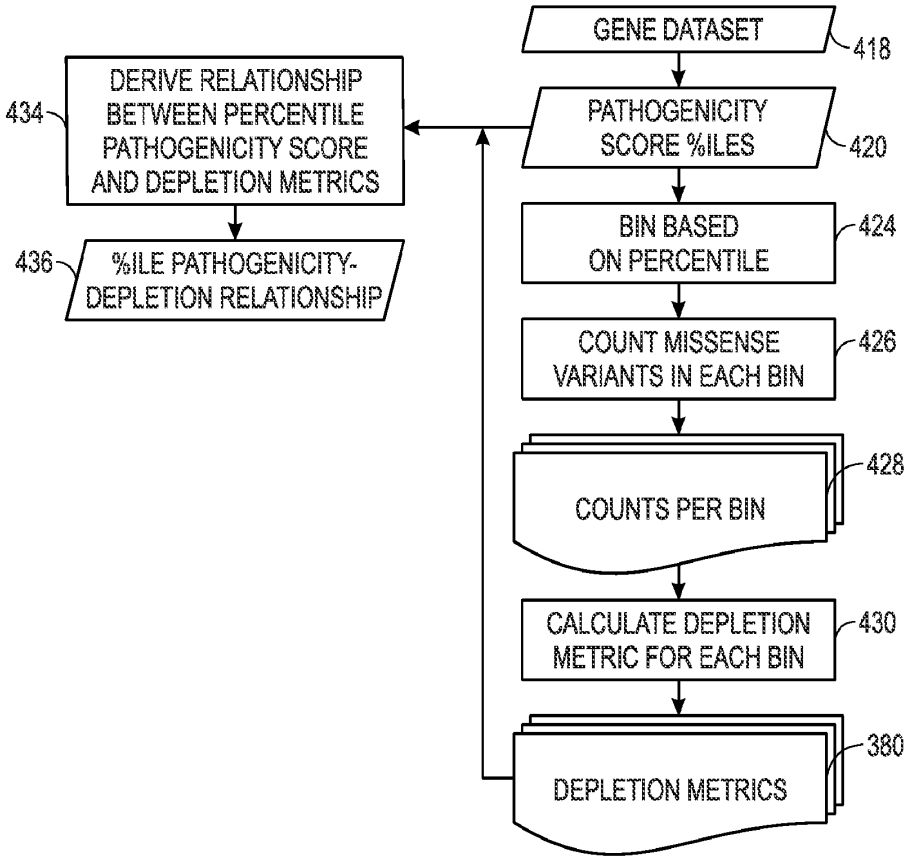
GENE DATASET
418
PATHOGENICITY SCORE %ILES
420
BIN BASED ON PERCENTILE
424
COUNT MISSENSE VARIANTS IN EACH BIN
426
COUNTS PER BIN
428
CALCULATE DEPLETION METRIC FOR EACH BIN
430
DEPLETION METRICS
380
DERIVE RELATIONSHIP BETWEEN PERCENTILE PATHOGENICITY SCORE AND DEPLETION METRICS
434
%ILE PATHOGENICITY-DEPLETION RELATIONSHIP
436

FIG. 17

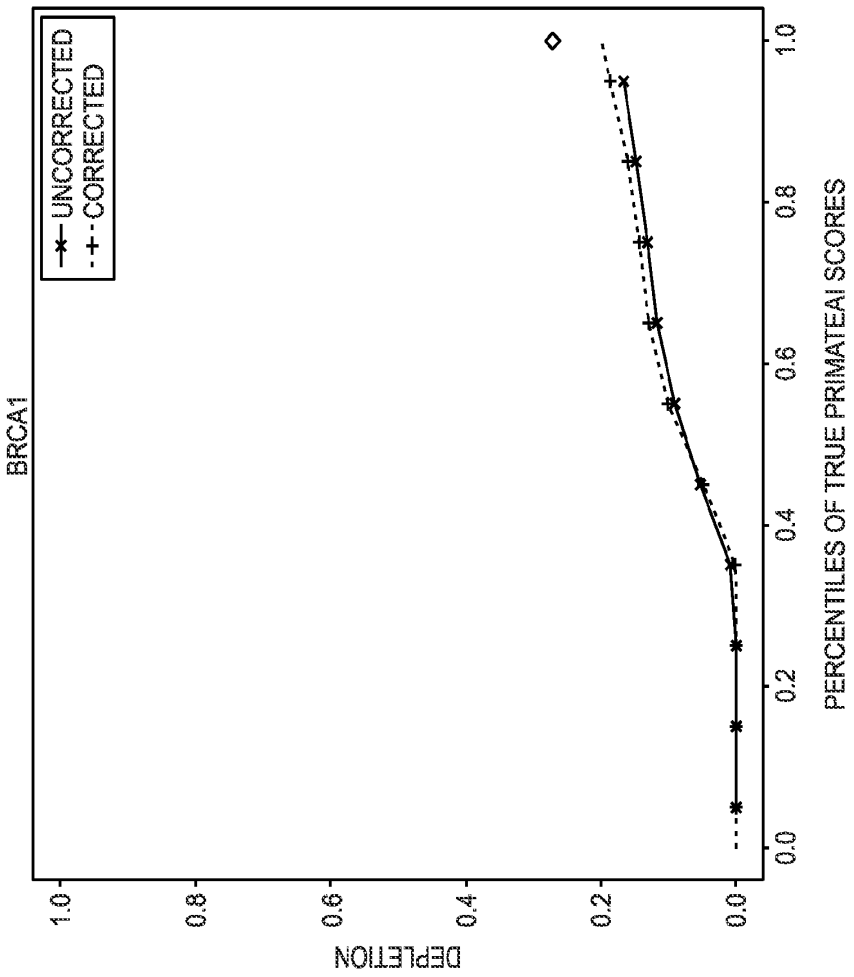
BRCA1
UNCORRECTED
CORRECTED
DEPLETION
1.0
0.8
0.6
0.4
0.2
0.0
PERCENTILES OF TRUE PRIMATEAI SCORES
0.0
0.2
0.4
0.6
0.8
1.0

FIG. 18

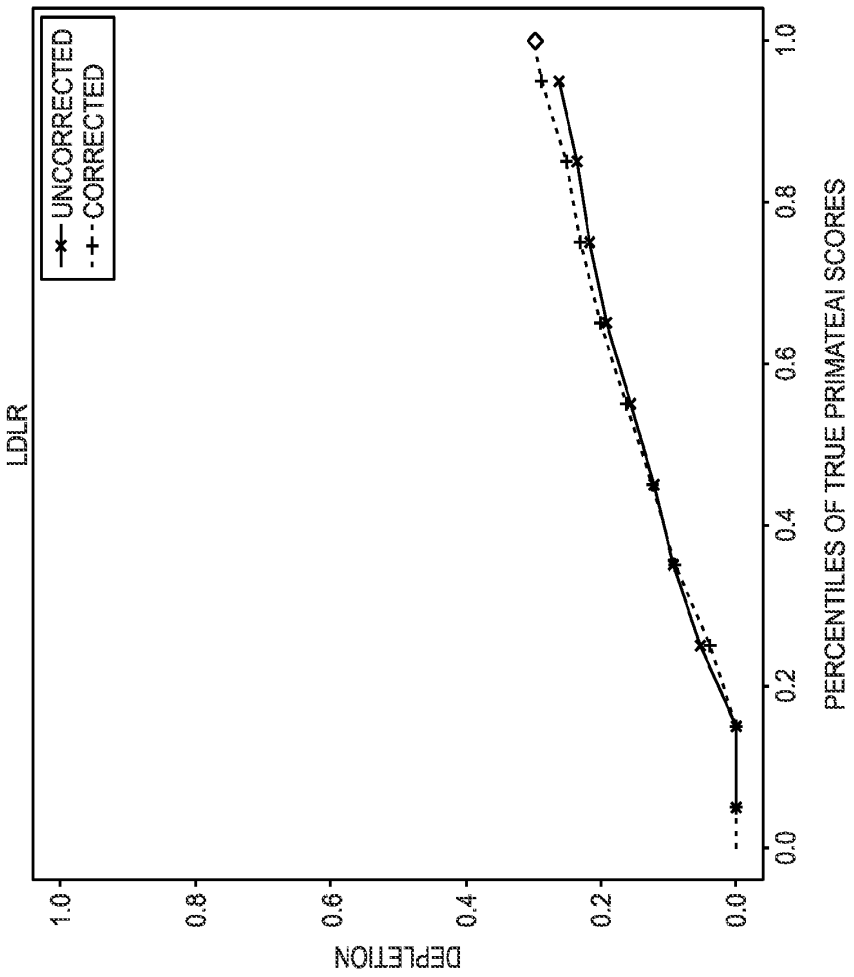
LDLR
UNCORRECTED
CORRECTED
DEPLETION
1.0
0.8
0.6
0.4
0.2
0.0
PERCENTILES OF TRUE PRIMATEAI SCORES
0.0
0.2
0.4
0.6
0.8
1.0

FIG. 19

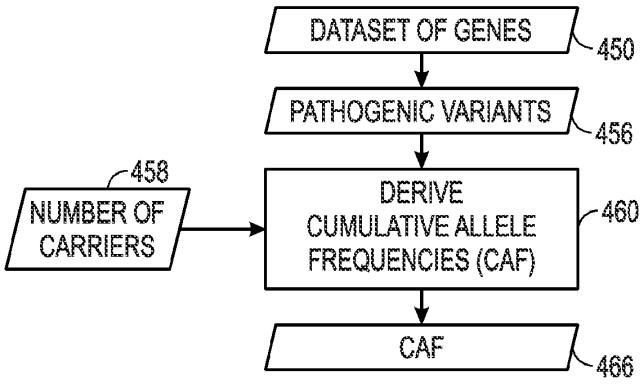
DATASET OF GENES
450
PATHOGENIC VARIANTS
456
458
NUMBER OF CARRIERS
DERIVE CUMULATIVE ALLELE FREQUENCIES (CAF)
460
CAF
466

FIG. 20

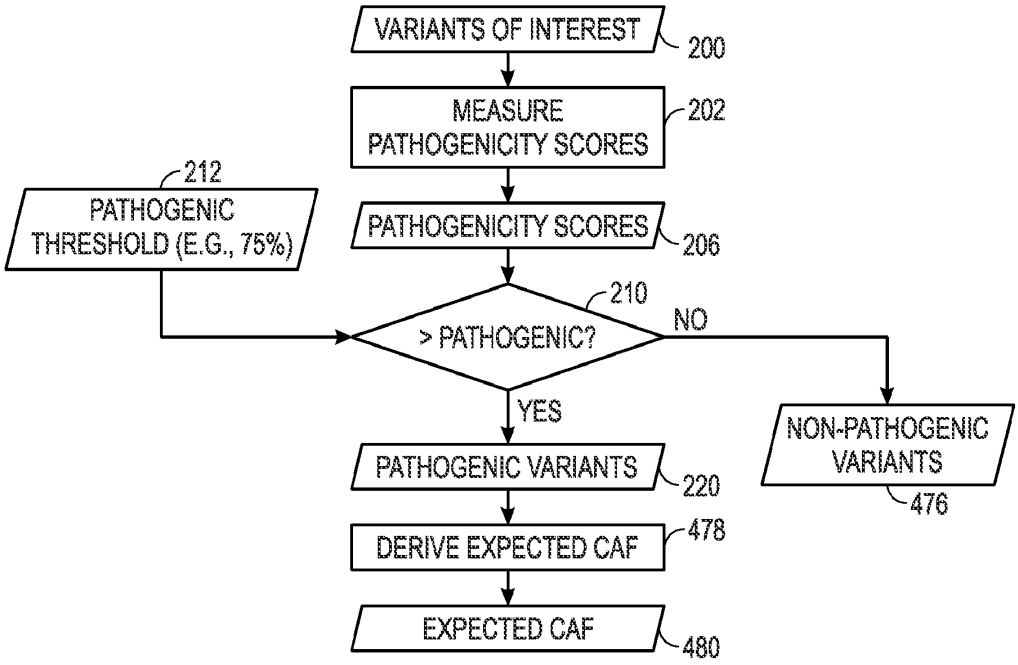
VARIANTS OF INTEREST
200
MEASURE PATHOGENICITY SCORES
202
212
PATHOGENIC THRESHOLD (E.G., 75%)
PATHOGENICITY SCORES
206
210
> PATHOGENIC?
NO
YES
NON-PATHOGENIC VARIANTS
476
PATHOGENIC VARIANTS
220
DERIVE EXPECTED CAF
478
EXPECTED CAF
480

FIG. 21

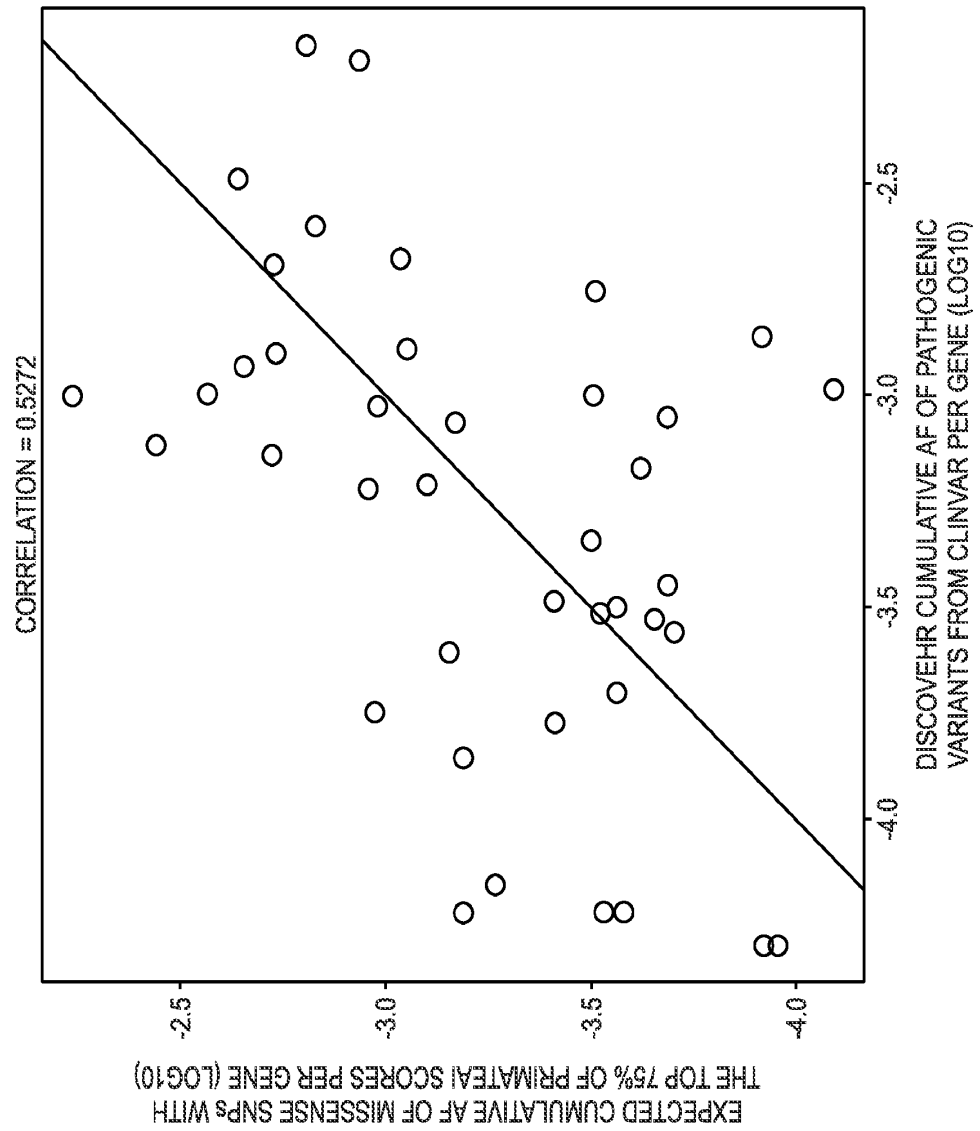
CORRELATION = 0.5272
DISCOVEHR CUMULATIVE AF OF PATHOGENIC VARIANTS FROM CLINVAR PER GENE (LOG10)
-2.5
-3.0
-3.5
-4.0
EXPECTED CUMULATIVE AF OF MISSENSE SNPs WITH THE TOP 75% OF PRIMATEAI SCORES PER GENE (LOG10)
-2.5
-3.0
-3.5
-4.0

FIG. 22

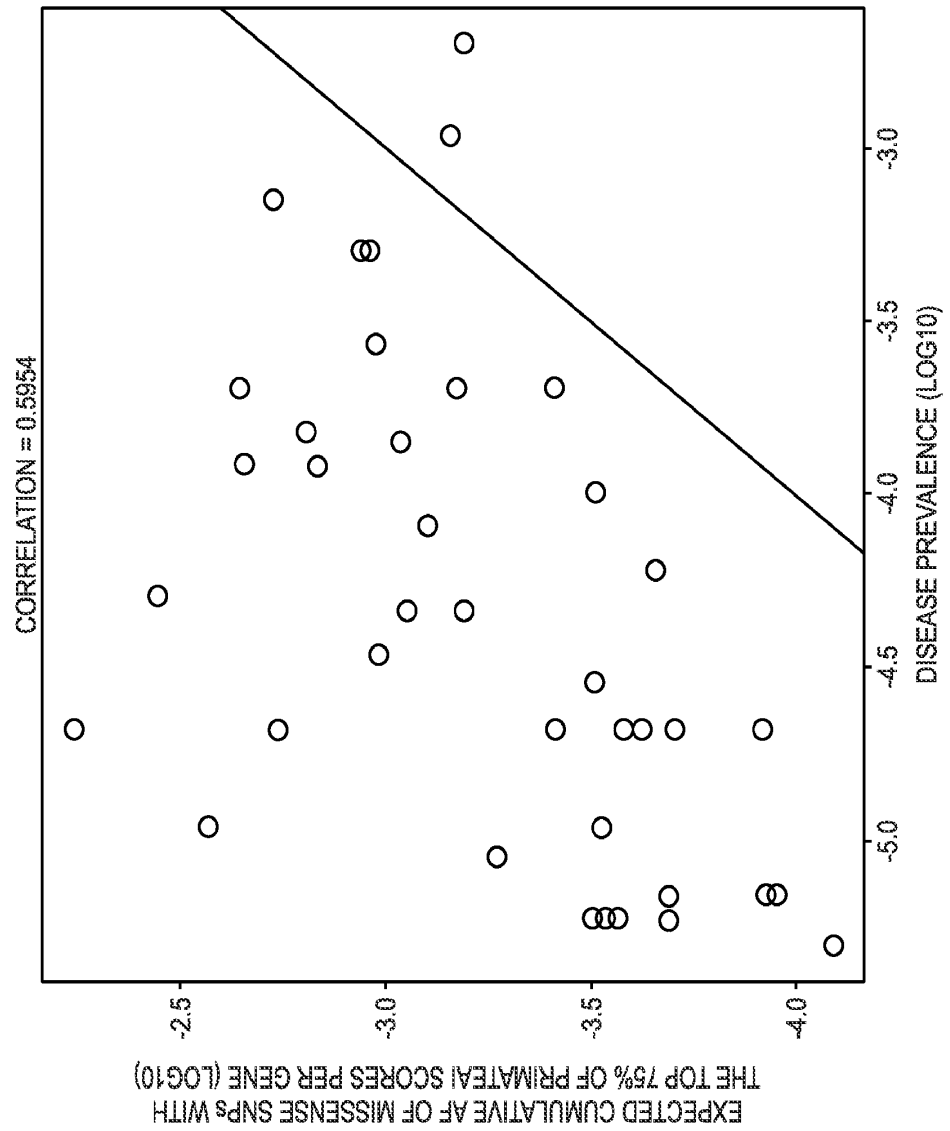
CORRELATION = 0.5954
DISEASE PREVALENCE (LOG10)
-3.0
-3.5
-4.0
-4.5
-5.0
EXPECTED CUMULATIVE AF OF MISSENSE SNPs WITH THE TOP 75% OF PRIMATEAI SCORES PER GENE (LOG10)
-2.5
-3.0
-3.5
-4.0

FIG. 23

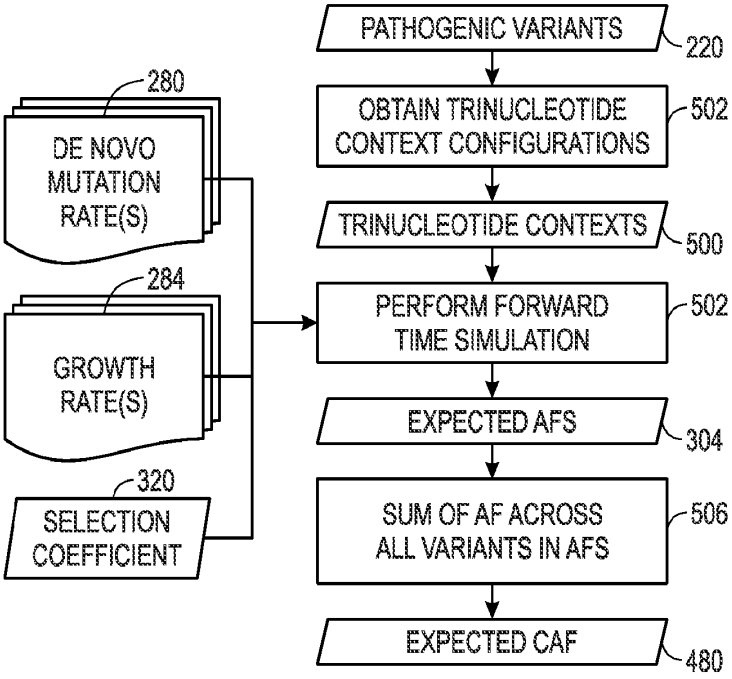
PATHOGENIC VARIANTS
220
OBTAIN TRINUCLEOTIDE CONTEXT CONFIGURATIONS
502
TRINUCLEOTIDE CONTEXTS
500
PERFORM FORWARD TIME SIMULATION
502
EXPECTED AFS
304
SUM OF AF ACROSS ALL VARIANTS IN AFS
506
EXPECTED CAF
480
280
DE NOVO MUTATION RATE(S)
284
GROWTH RATE(S)
320
SELECTION COEFFICIENT

FIG. 24

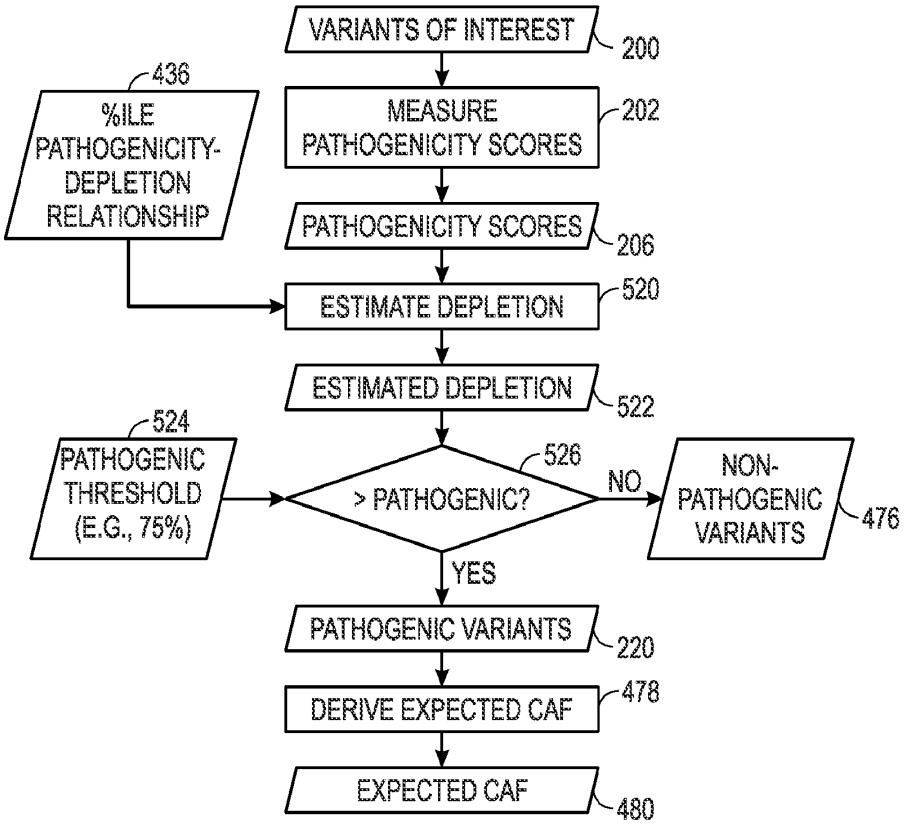
VARIANTS OF INTEREST
200
436
%ILE PATHOGENICITY-DEPLETION RELATIONSHIP
MEASURE PATHOGENICITY SCORES
202
PATHOGENICITY SCORES
206
ESTIMATE DEPLETION
520
ESTIMATED DEPLETION
522
524
PATHOGENIC THRESHOLD (E.G., 75%)
526
> PATHOGENIC?
NO
NON-PATHOGENIC VARIANTS
476
YES
PATHOGENIC VARIANTS
220
DERIVE EXPECTED CAF
478
EXPECTED CAF
480

FIG. 25

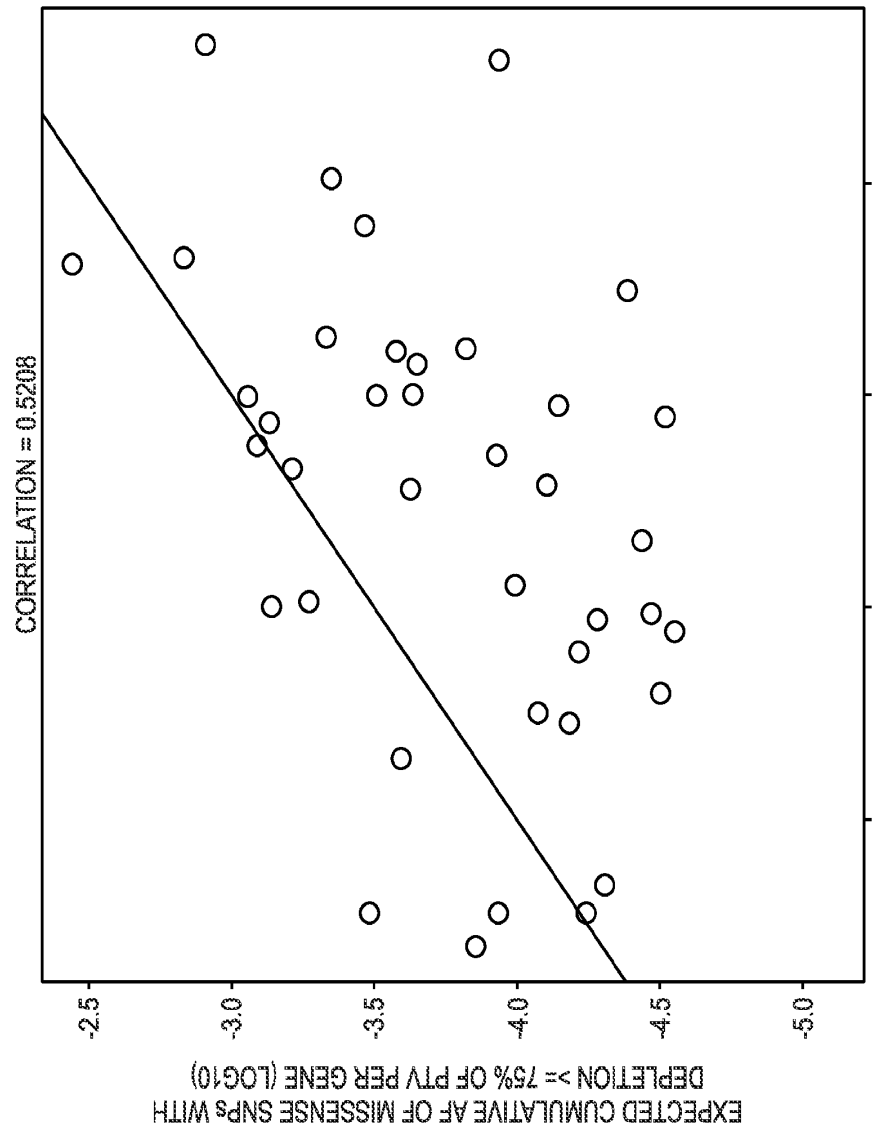
CORRELATION = 0.5208
DISCOVEHR CUMULATIVE AF OF PATHOGENIC VARIANTS FROM CLINVAR PER GENE (LOG10)
EXPECTED CUMULATIVE AF OF MISSENSE SNPs WITH DEPLETION >= 75% OF PTV PER GENE (LOG10)
-2.5
-3.0
-3.5
-4.0
-4.5
-5.0

FIG. 26

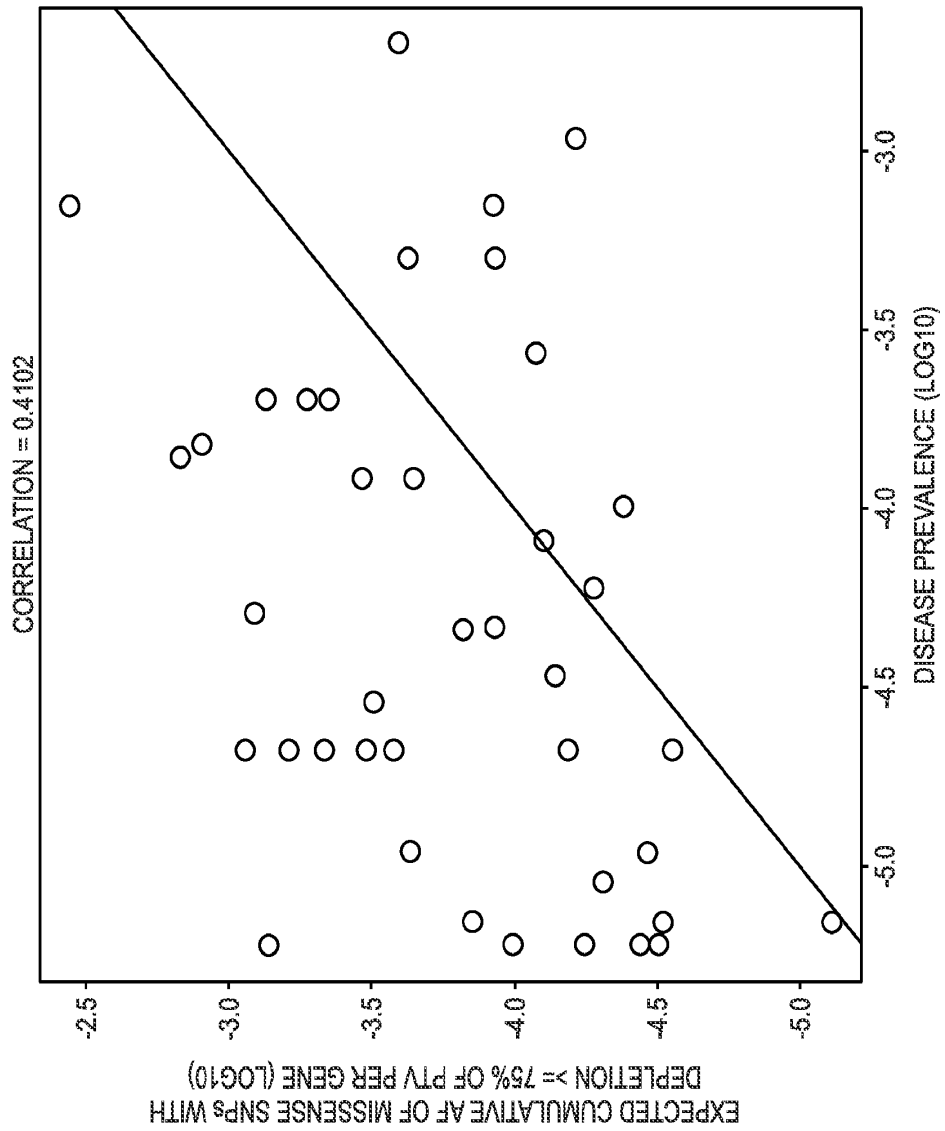
CORRELATION = 0.4102
EXPECTED CUMULATIVE AF OF MISSENSE SNPs WITH DEPLETION >= 75% OF PTV PER GENE (LOG10)
-2.5
-3.0
-3.5
-4.0
-4.5
-5.0
DISEASE PREVALENCE (LOG10)
-3.0
-3.5
-4.0
-4.5
-5.0

FIG. 27

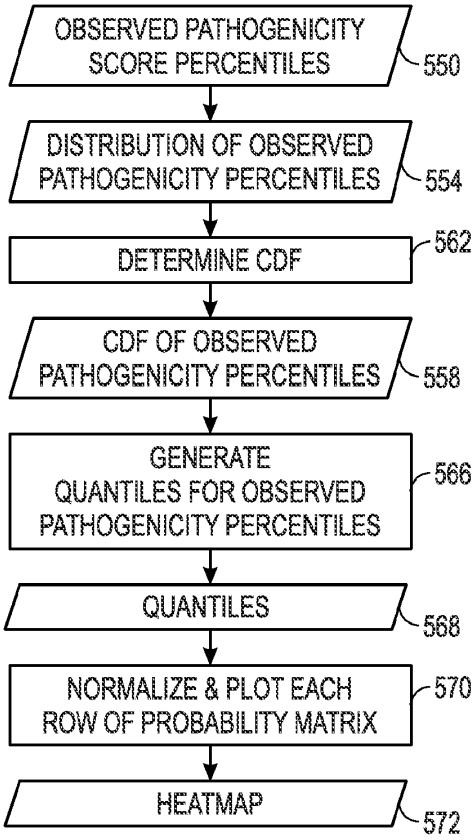
OBSERVED PATHOGENICITY SCORE PERCENTILES
550
DISTRIBUTION OF OBSERVED PATHOGENICITY PERCENTILES
554
DETERMINE CDF
562
CDF OF OBSERVED PATHOGENICITY PERCENTILES
558
GENERATE QUANTILES FOR OBSERVED PATHOGENICITY PERCENTILES
566
QUANTILES
568
NORMALIZE & PLOT EACH ROW OF PROBABILITY MATRIX
570
HEATMAP
572

FIG. 28

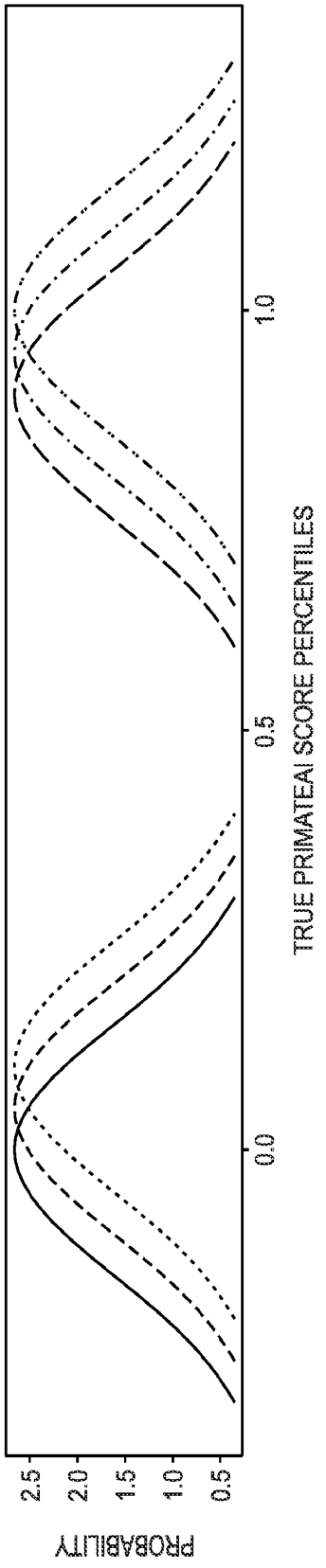
PROBABILITY
2.5
2.0
1.5
1.0
0.5
0.0
0.5
1.0
TRUE PRIMATEAI SCORE PERCENTILES

FIG. 29

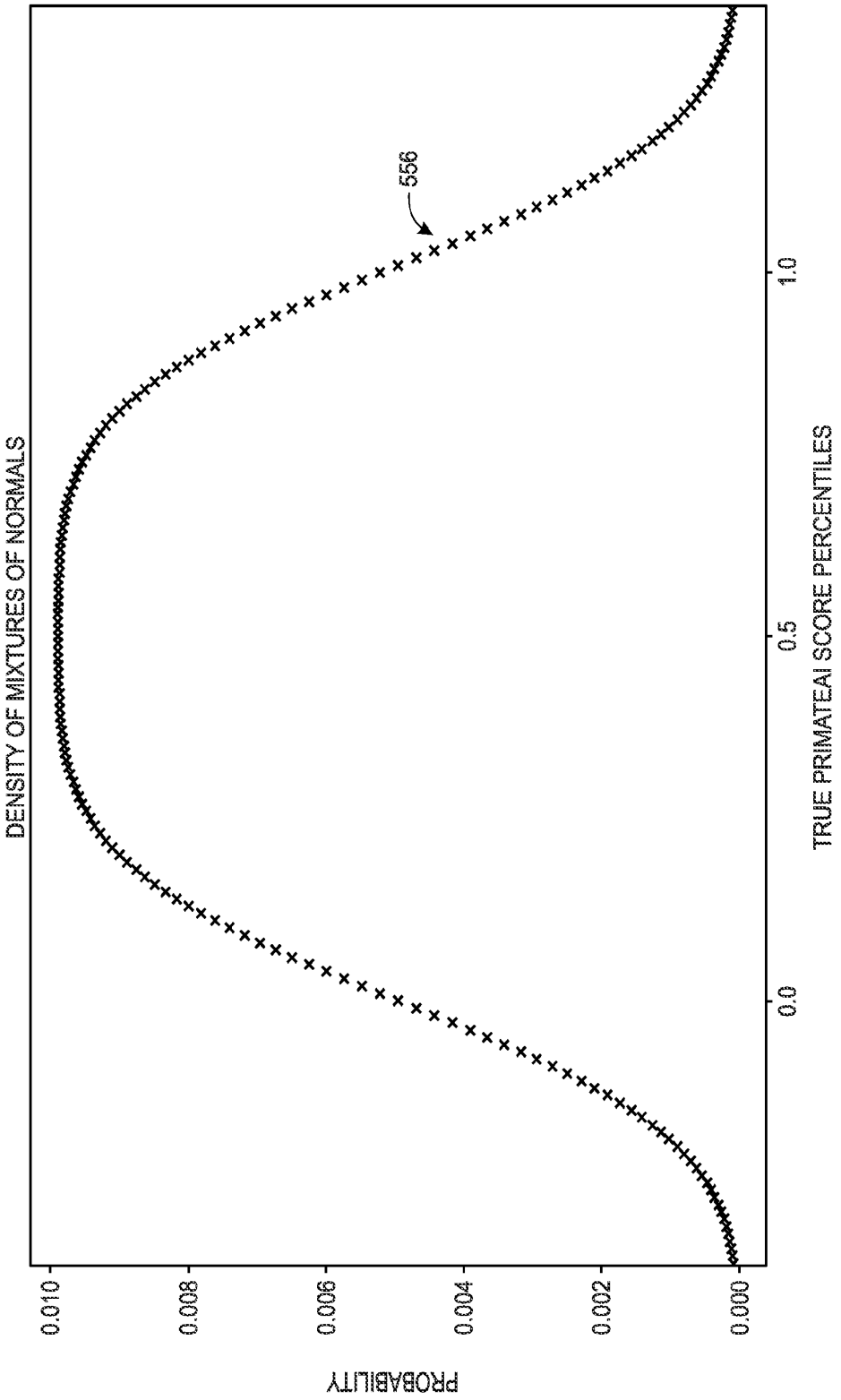
DENSITY OF MIXTURES OF NORMALS
PROBABILITY
0.010
0.008
0.006
0.004
0.002
0.000
556
0.0
0.5
1.0
TRUE PRIMATEAI SCORE PERCENTILES

FIG. 30

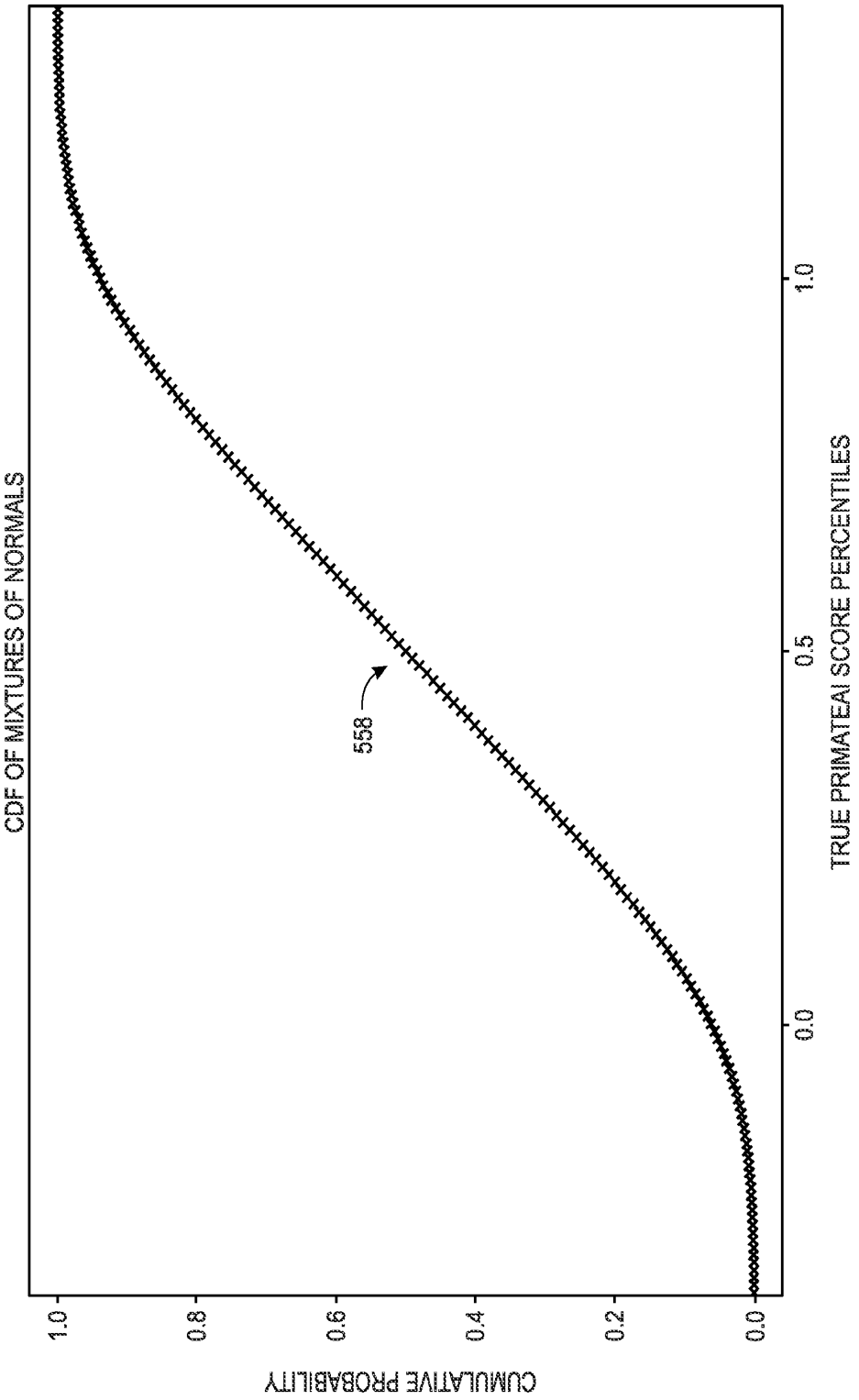
CDF OF MIXTURES OF NORMALS
CUMULATIVE PROBABILITY
1.0
0.8
0.6
0.4
0.2
0.0
558
0.0
0.5
1.0
TRUE PRIMATEAI SCORE PERCENTILES

FIG. 31

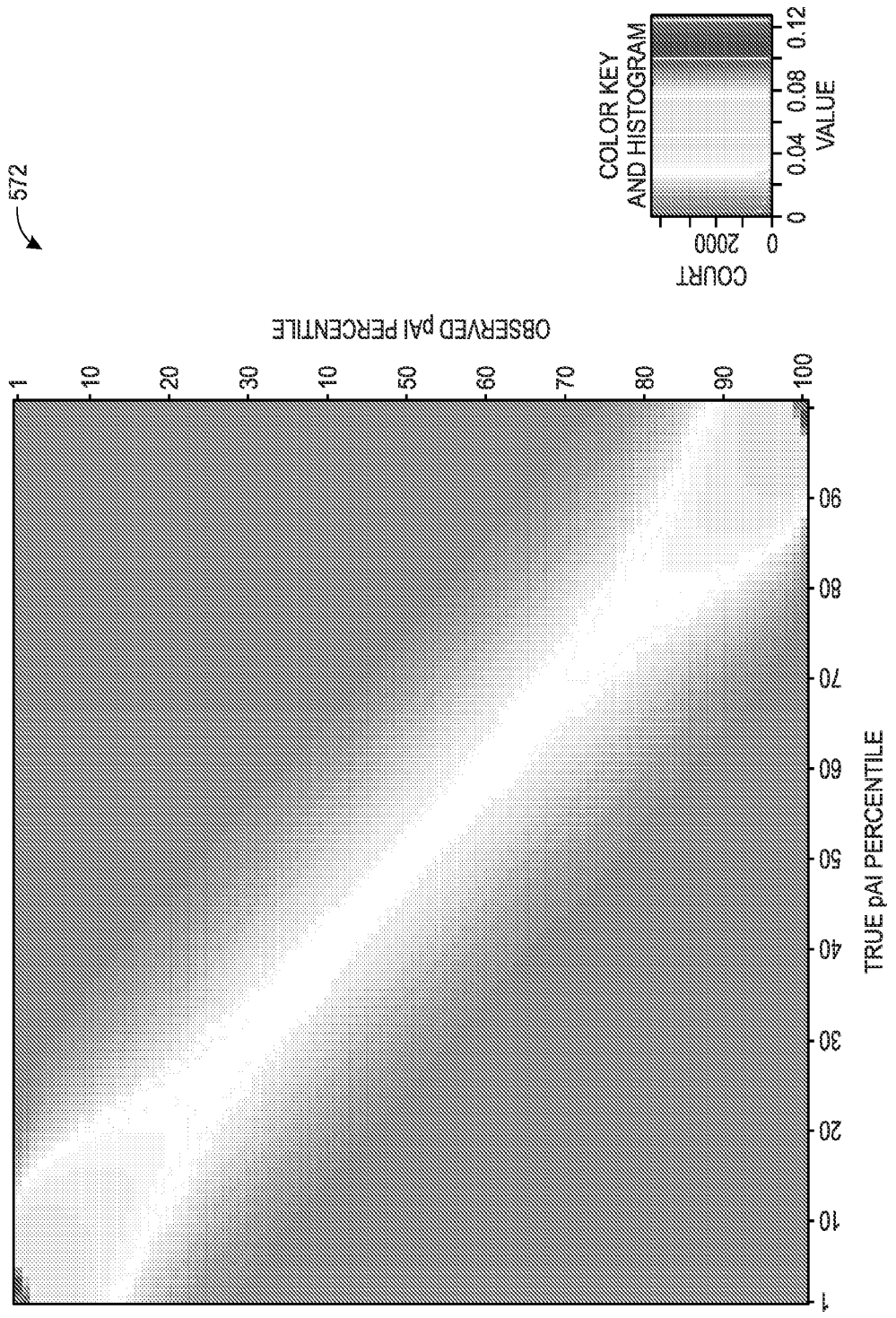
572
COLOR KEY AND HISTOGRAM
COUNT
0
2000
VALUE
0
0.04
0.08
0.12
OBSERVED pAI PERCENTILE
1
10
20
30
40
50
60
70
80
90
100
TRUE pAI PERCENTILE

FIG. 32

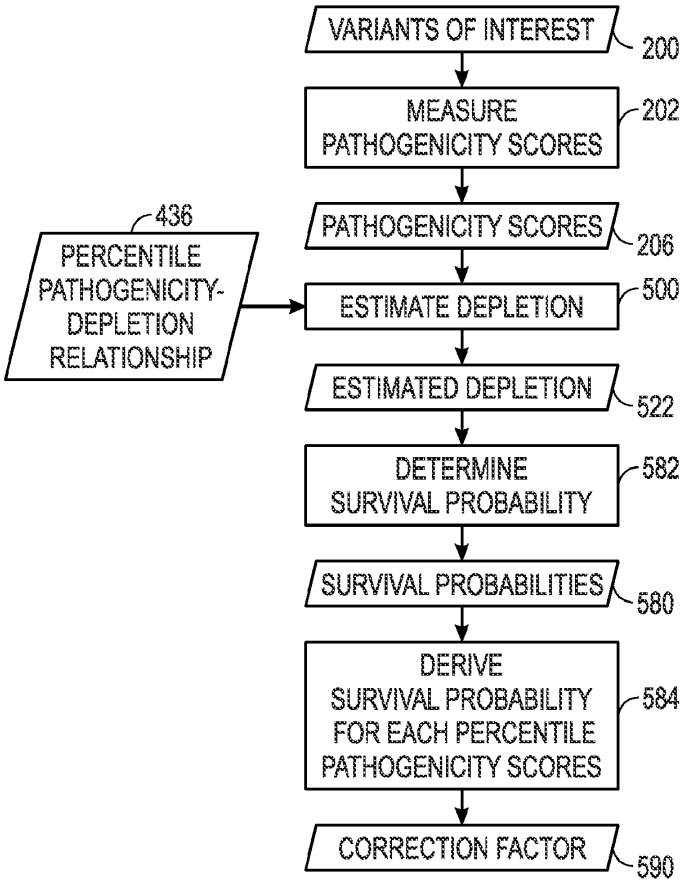
VARIANTS OF INTEREST
200
MEASURE PATHOGENICITY SCORES
202
PATHOGENICITY SCORES
206
436
PERCENTILE PATHOGENICITY-DEPLETION RELATIONSHIP
ESTIMATE DEPLETION
500
ESTIMATED DEPLETION
522
DETERMINE SURVIVAL PROBABILITY
582
SURVIVAL PROBABILITIES
580
DERIVE SURVIVAL PROBABILITY FOR EACH PERCENTILE PATHOGENICITY SCORES
584
CORRECTION FACTOR
590

FIG. 33

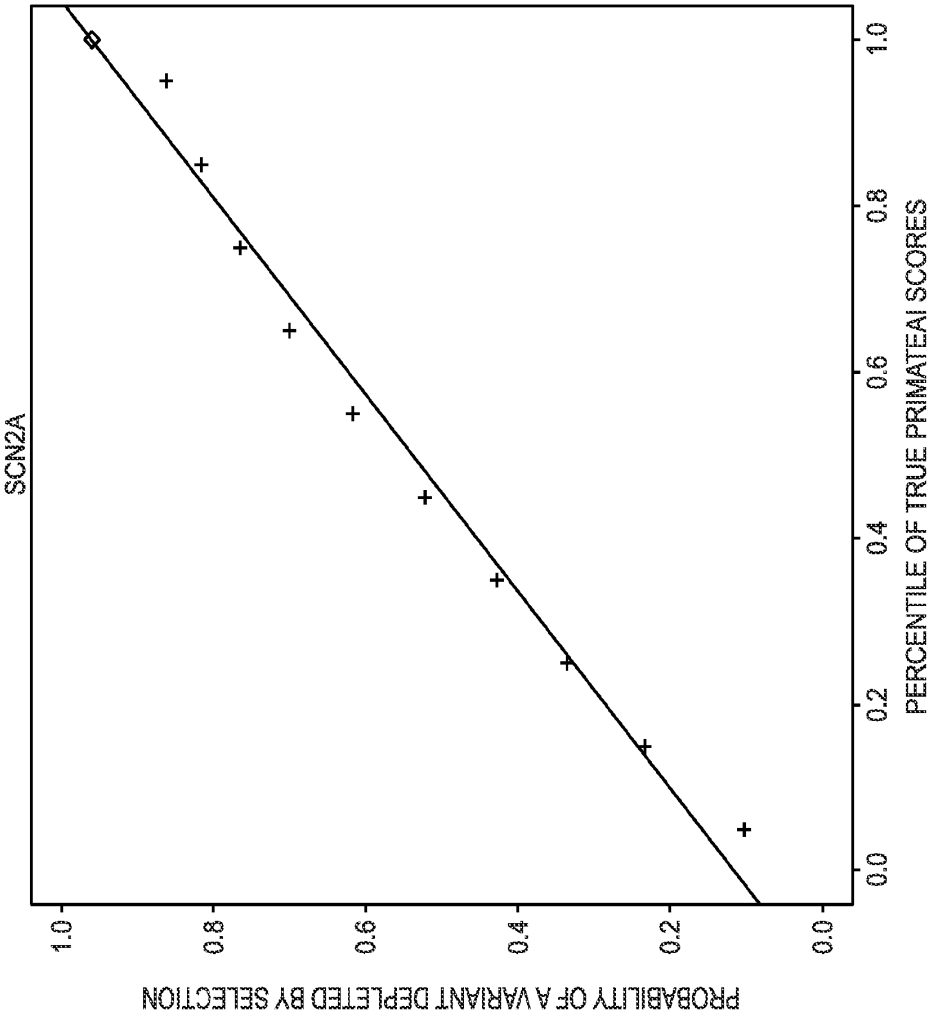
SCN2A
PROBABILITY OF A VARIANT DEPLETED BY SELECTION
PERCENTILE OF TRUE PRIMATEAI SCORES
0.0
0.2
0.4
0.6
0.8
1.0

FIG. 34

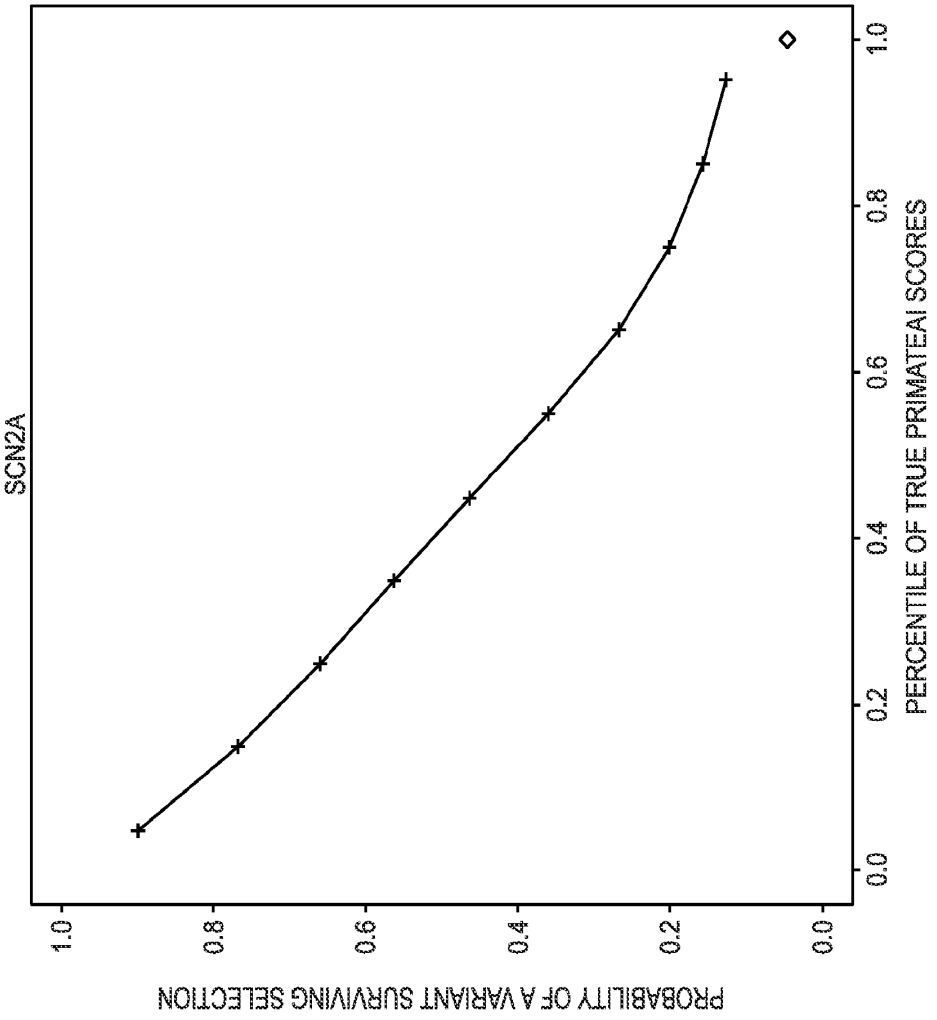
SCN2A
PROBABILITY OF A VARIANT SURVIVING SELECTION
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
PERCENTILE OF TRUE PRIMATEAI SCORES

FIG. 35

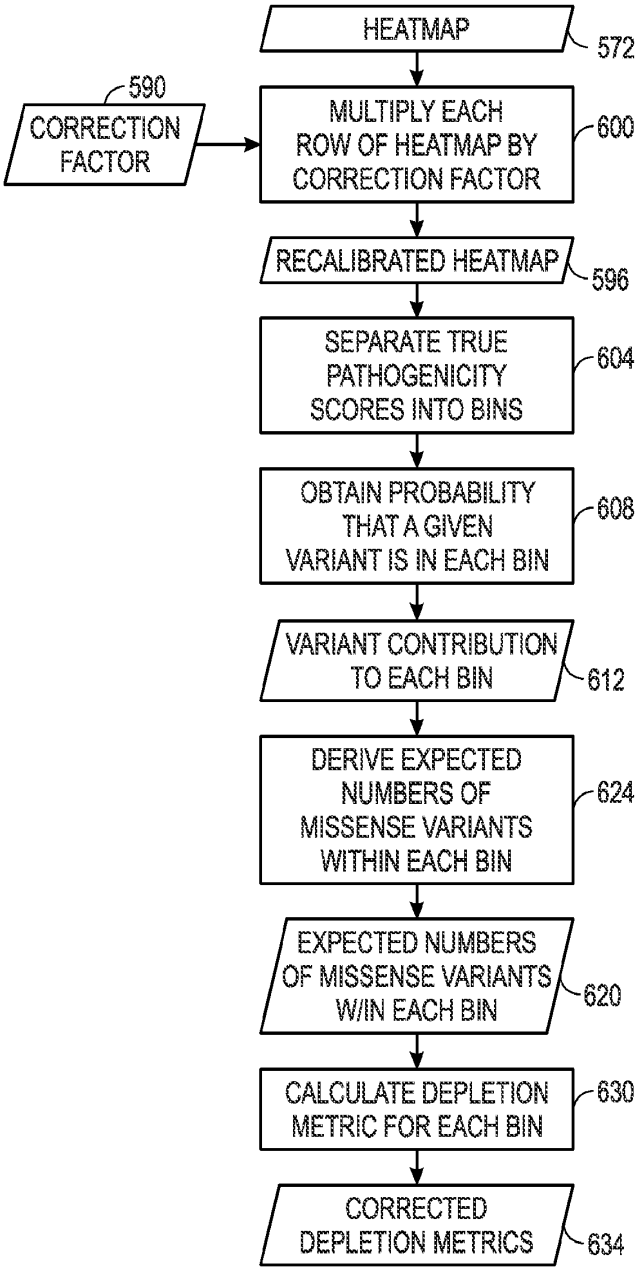
HEATMAP
572
590
CORRECTION FACTOR
MULTIPLY EACH ROW OF HEATMAP BY CORRECTION FACTOR
600
RECALIBRATED HEATMAP
596
SEPARATE TRUE PATHOGENICITY SCORES INTO BINS
604
OBTAIN PROBABILITY THAT A GIVEN VARIANT IS IN EACH BIN
608
VARIANT CONTRIBUTION TO EACH BIN
612
DERIVE EXPECTED NUMBERS OF MISSENSE VARIANTS WITHIN EACH BIN
624
EXPECTED NUMBERS OF MISSENSE VARIANTS W/IN EACH BIN
620
CALCULATE DEPLETION METRIC FOR EACH BIN
630
CORRECTED DEPLETION METRICS
634

FIG. 36

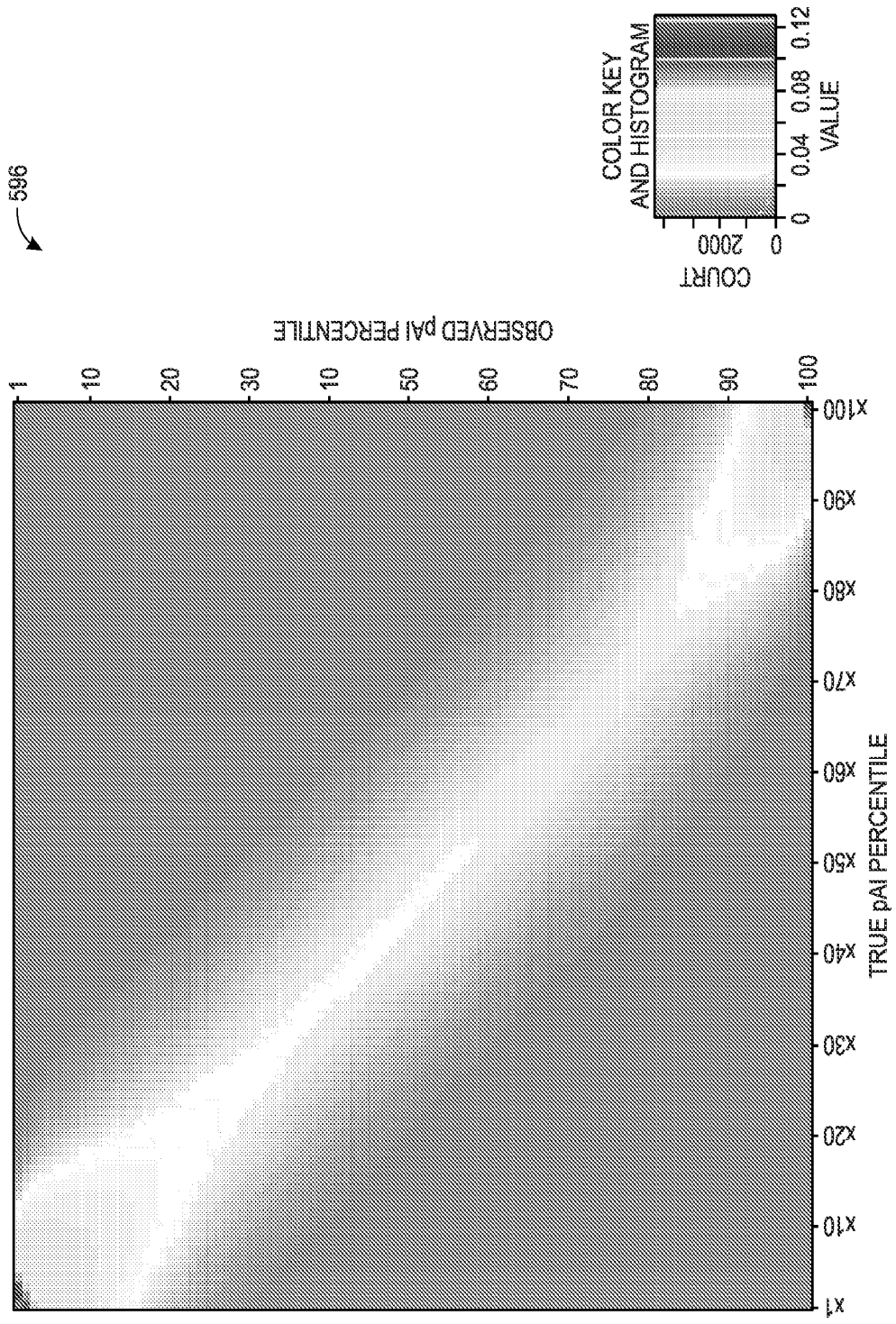
596
COLOR KEY AND HISTOGRAM
COUNT
0
2000
VALUE
0
0.04
0.08
0.12
OBSERVED pAI PERCENTILE
1
10
20
30
40
50
60
70
80
90
100
TRUE pAI PERCENTILE
x1
x10
x20
x30
x40
x50
x60
x70
x80
x90
x100

FIG. 37

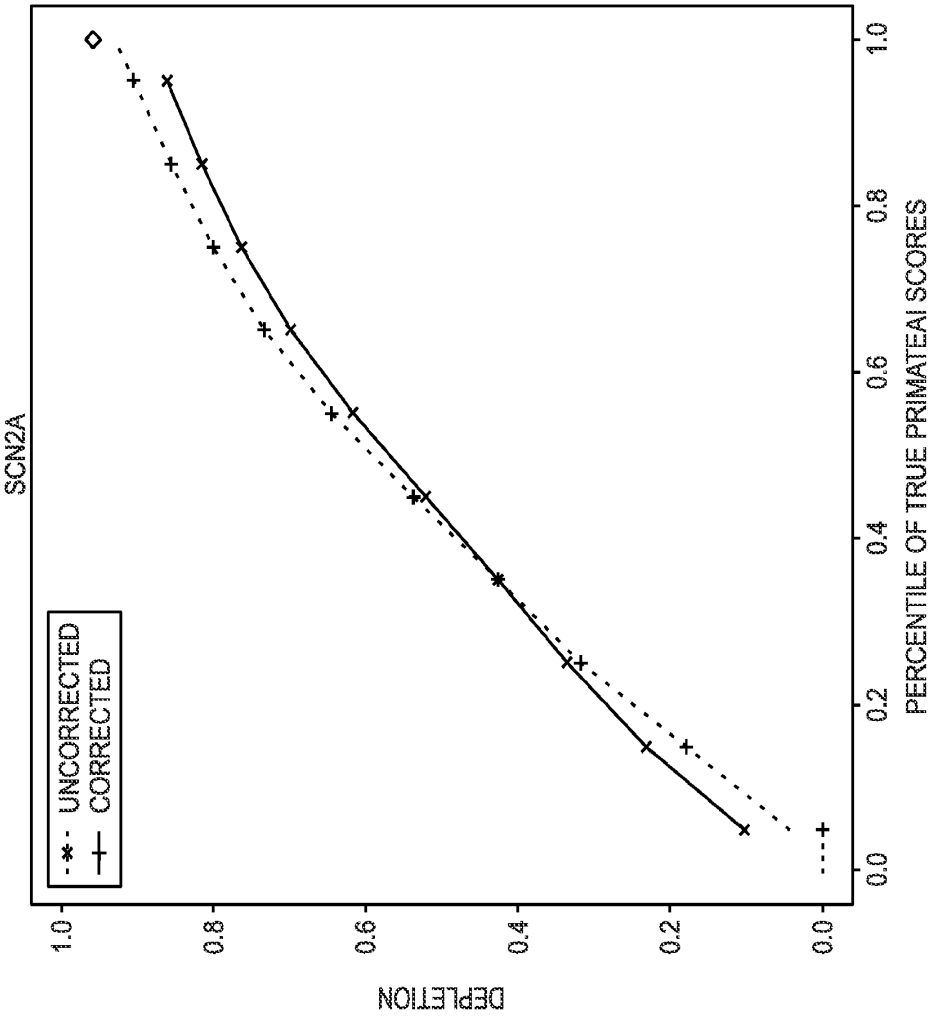
SCN2A
UNCORRECTED
CORRECTED
DEPLETION
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
PERCENTILE OF TRUE PRIMATEAI SCORES

FIG. 38

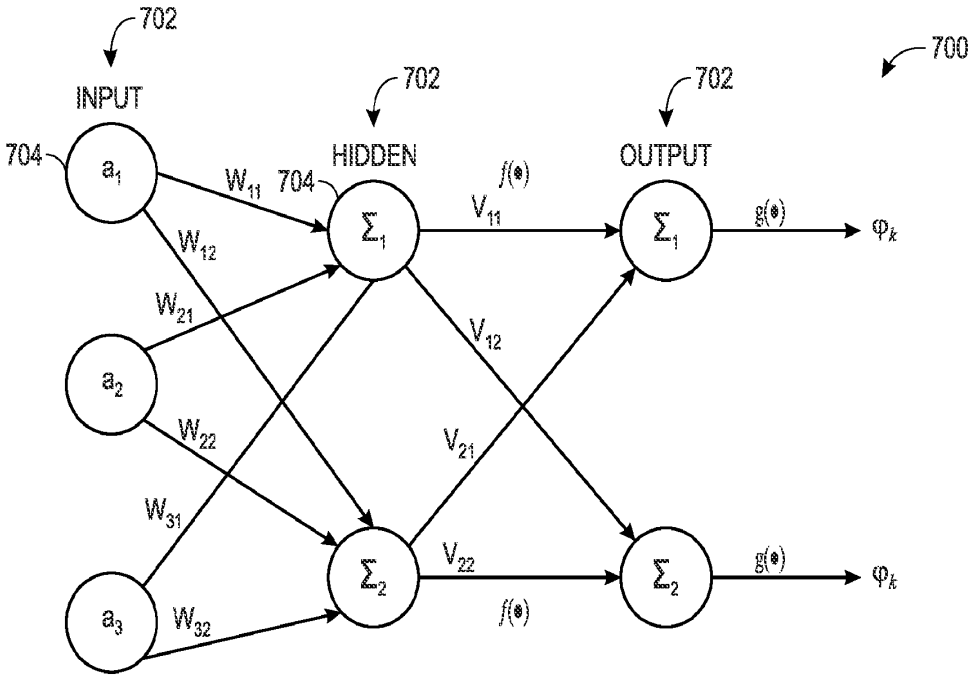
702
INPUT
704
a_1
a_2
a_3
W_{11}
W_{12}
W_{21}
W_{22}
W_{31}
W_{32}
702
HIDDEN
704
Σ_1
Σ_2
$f(\bullet)$
V_{11}
V_{12}
V_{21}
V_{22}
$f(\bullet)$
702
OUTPUT
Σ_1
Σ_2
$g(\bullet)$
$g(\bullet)$
φ_k
φ_k
700

FIG. 39

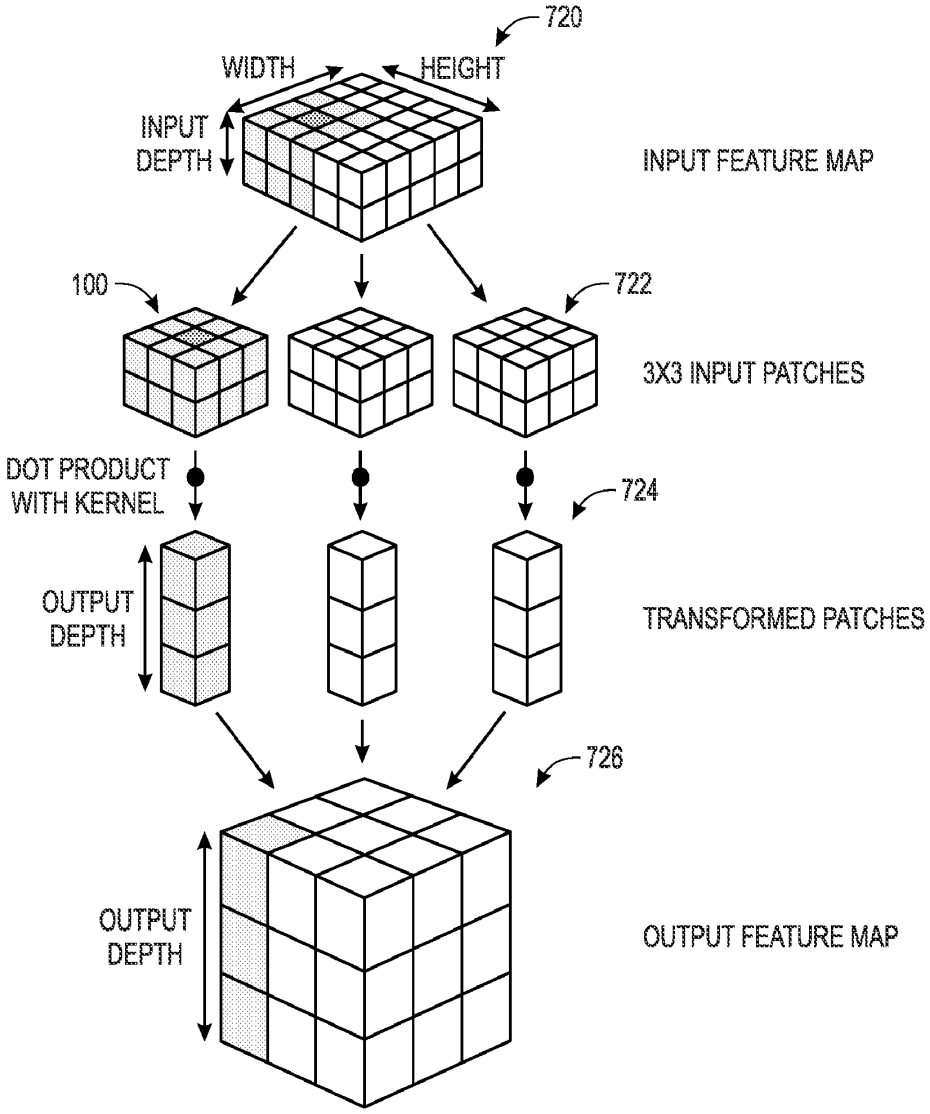
720
WIDTH
HEIGHT
INPUT DEPTH
INPUT FEATURE MAP
100
722
3X3 INPUT PATCHES
DOT PRODUCT WITH KERNEL
724
OUTPUT DEPTH
TRANSFORMED PATCHES
726
OUTPUT DEPTH
OUTPUT FEATURE MAP

FIG. 40

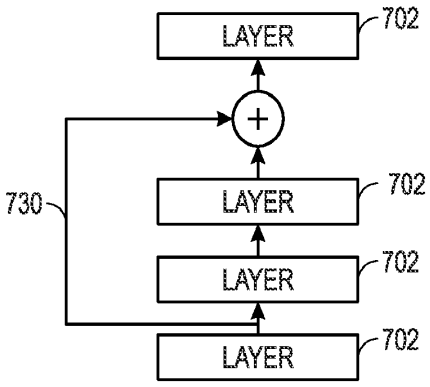
LAYER
702
730
LAYER
702
LAYER
702
LAYER
702

FIG. 41

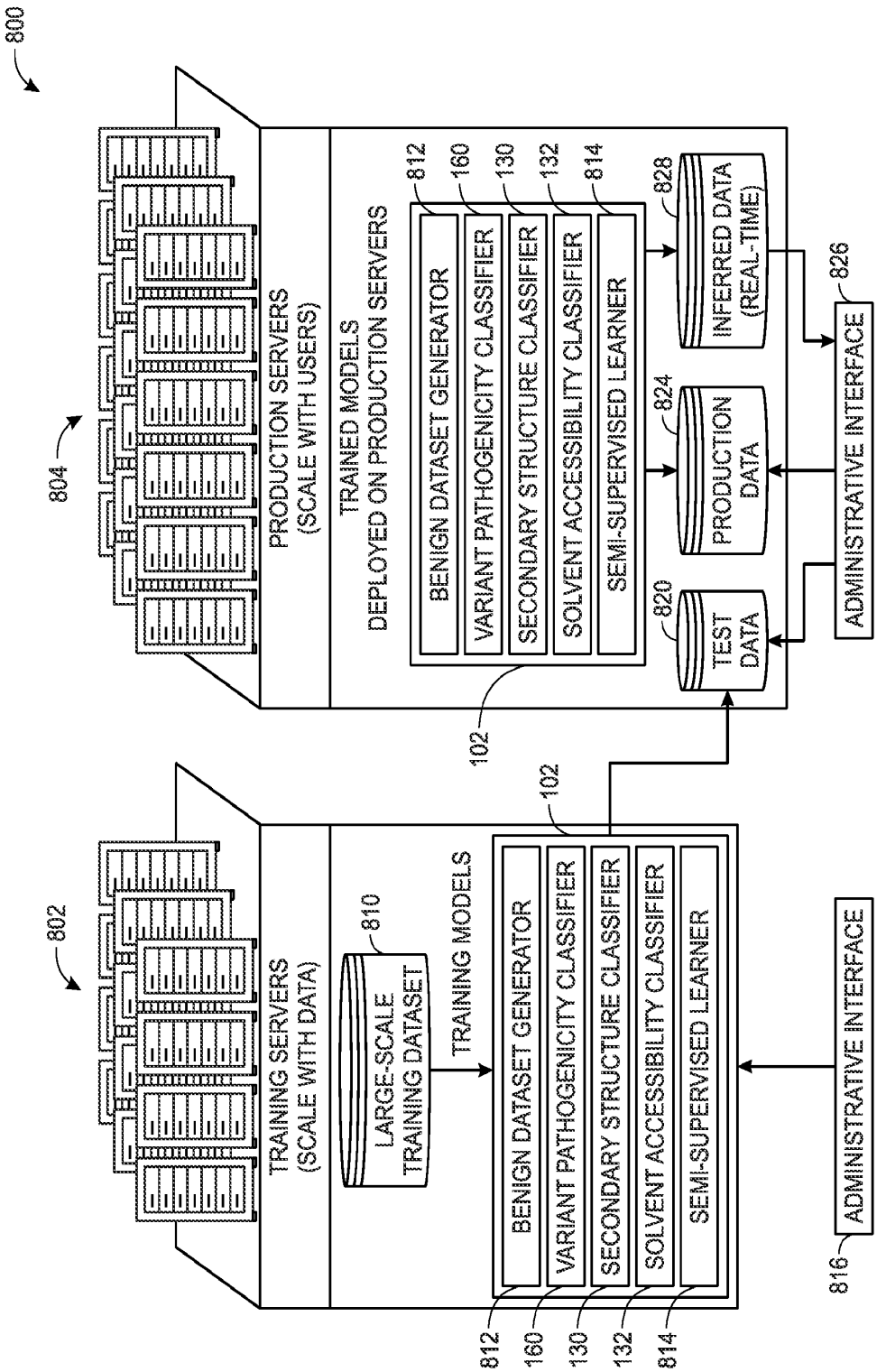
800
804
802
PRODUCTION SERVERS
(SCALE WITH USERS)
TRAINED MODELS
DEPLOYED ON PRODUCTION SERVERS
812
160
130
132
814
828
BENIGN DATASET GENERATOR
VARIANT PATHOGENICITY CLASSIFIER
SECONDARY STRUCTURE CLASSIFIER
SOLVENT ACCESSIBILITY CLASSIFIER
SEMI-SUPERVISED LEARNER
INFERRED DATA
(REAL-TIME)
826
824
PRODUCTION
DATA
820
TEST
DATA
ADMINISTRATIVE INTERFACE
102
TRAINING SERVERS
(SCALE WITH DATA)
810
LARGE-SCALE
TRAINING DATASET
TRAINING MODELS
102
BENIGN DATASET GENERATOR
VARIANT PATHOGENICITY CLASSIFIER
SECONDARY STRUCTURE CLASSIFIER
SOLVENT ACCESSIBILITY CLASSIFIER
SEMI-SUPERVISED LEARNER
ADMINISTRATIVE INTERFACE
816
812
160
130
132
814

FIG. 42

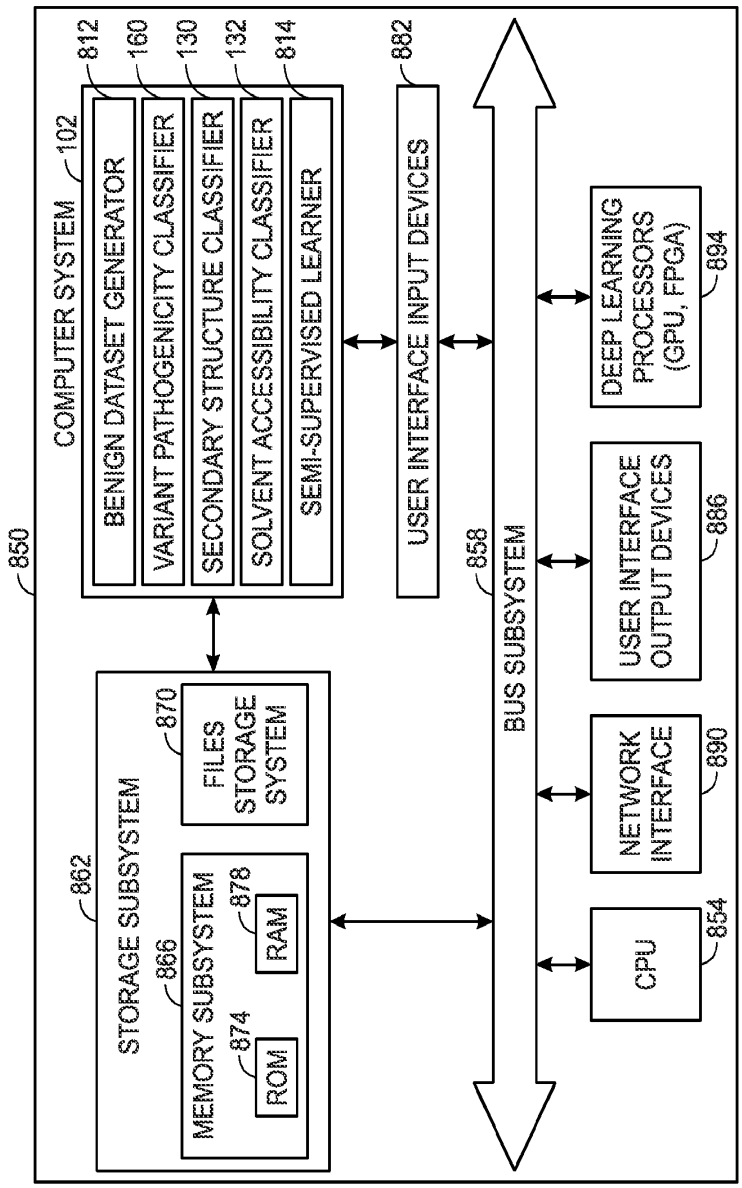
850
COMPUTER SYSTEM
102
812
BENIGN DATASET GENERATOR
160
VARIANT PATHOGENICITY CLASSIFIER
130
SECONDARY STRUCTURE CLASSIFIER
132
SOLVENT ACCESSIBILITY CLASSIFIER
814
SEMI-SUPERVISED LEARNER
882
USER INTERFACE INPUT DEVICES
858
BUS SUBSYSTEM
DEEP LEARNING PROCESSORS (GPU, FPGA)
894
USER INTERFACE OUTPUT DEVICES
886
NETWORK INTERFACE
890
CPU
854
STORAGE SUBSYSTEM
862
FILES STORAGE SYSTEM
870
MEMORY SUBSYSTEM
866
RAM
878
ROM
874

FIG. 43