

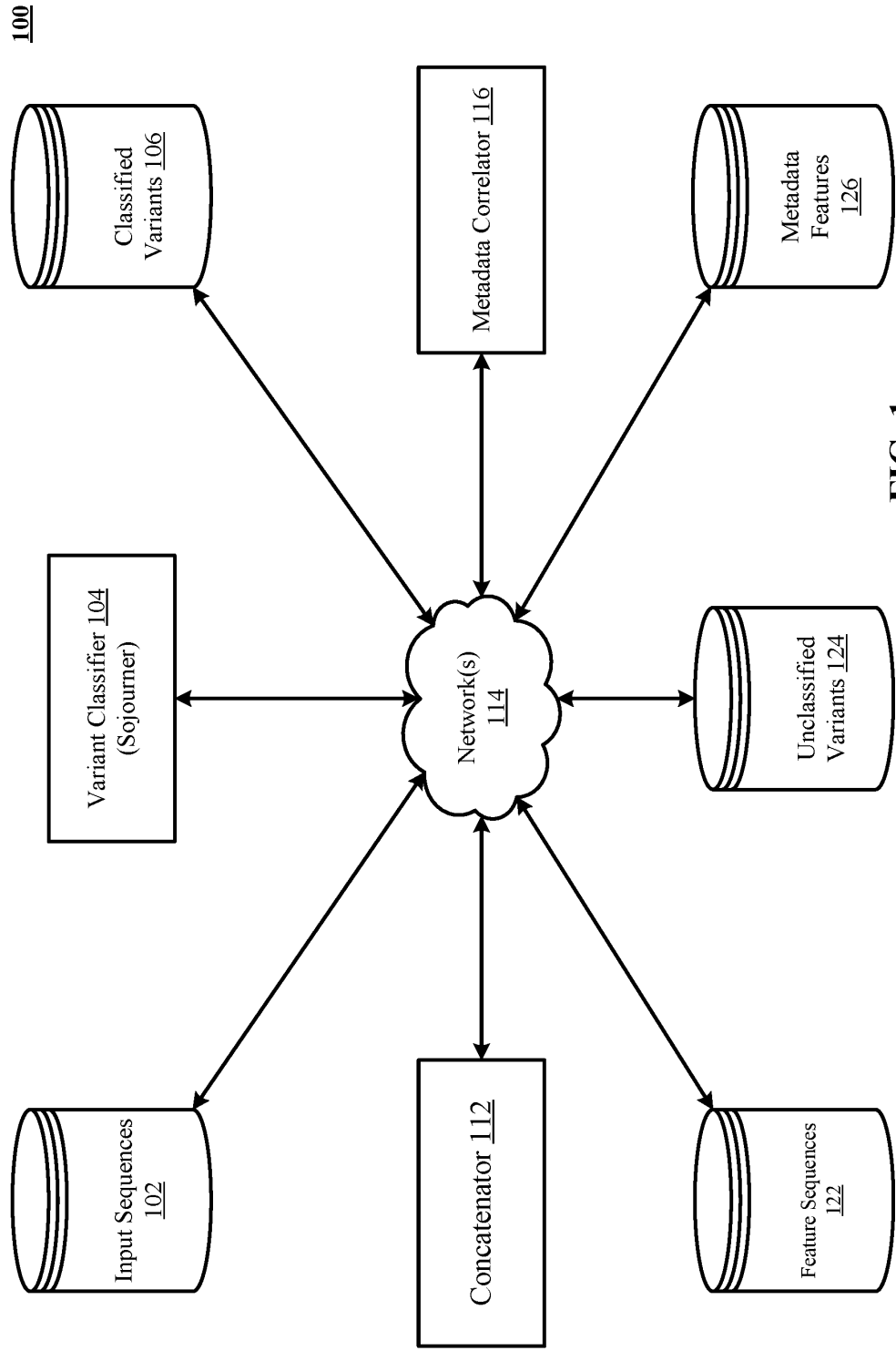


FIG. 1

200

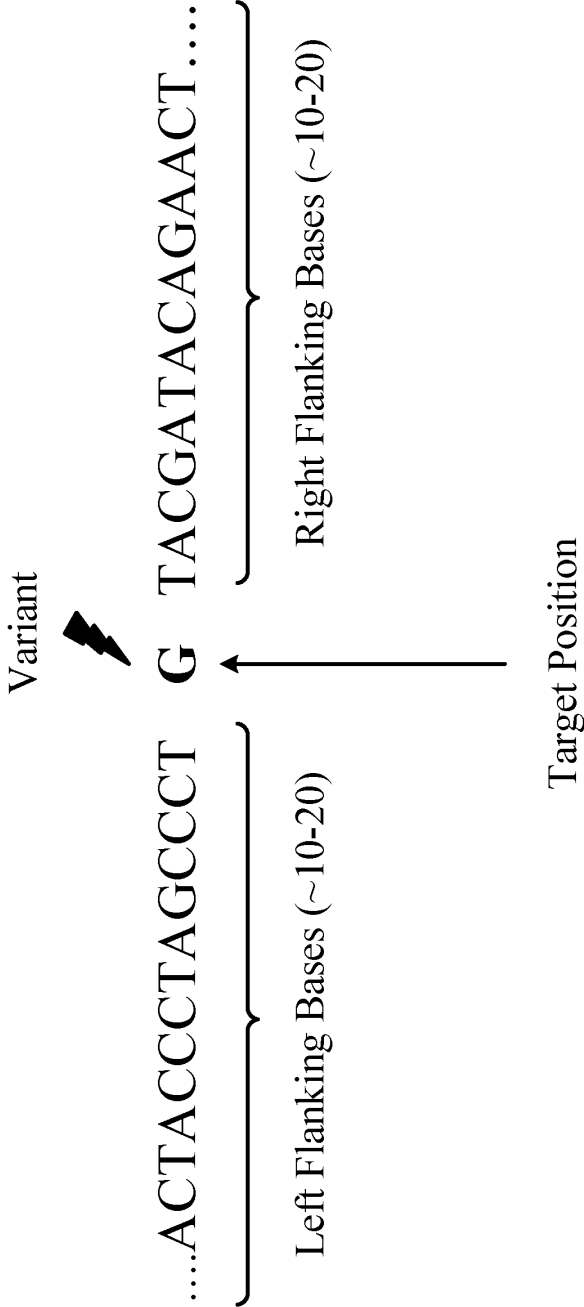


FIG. 2

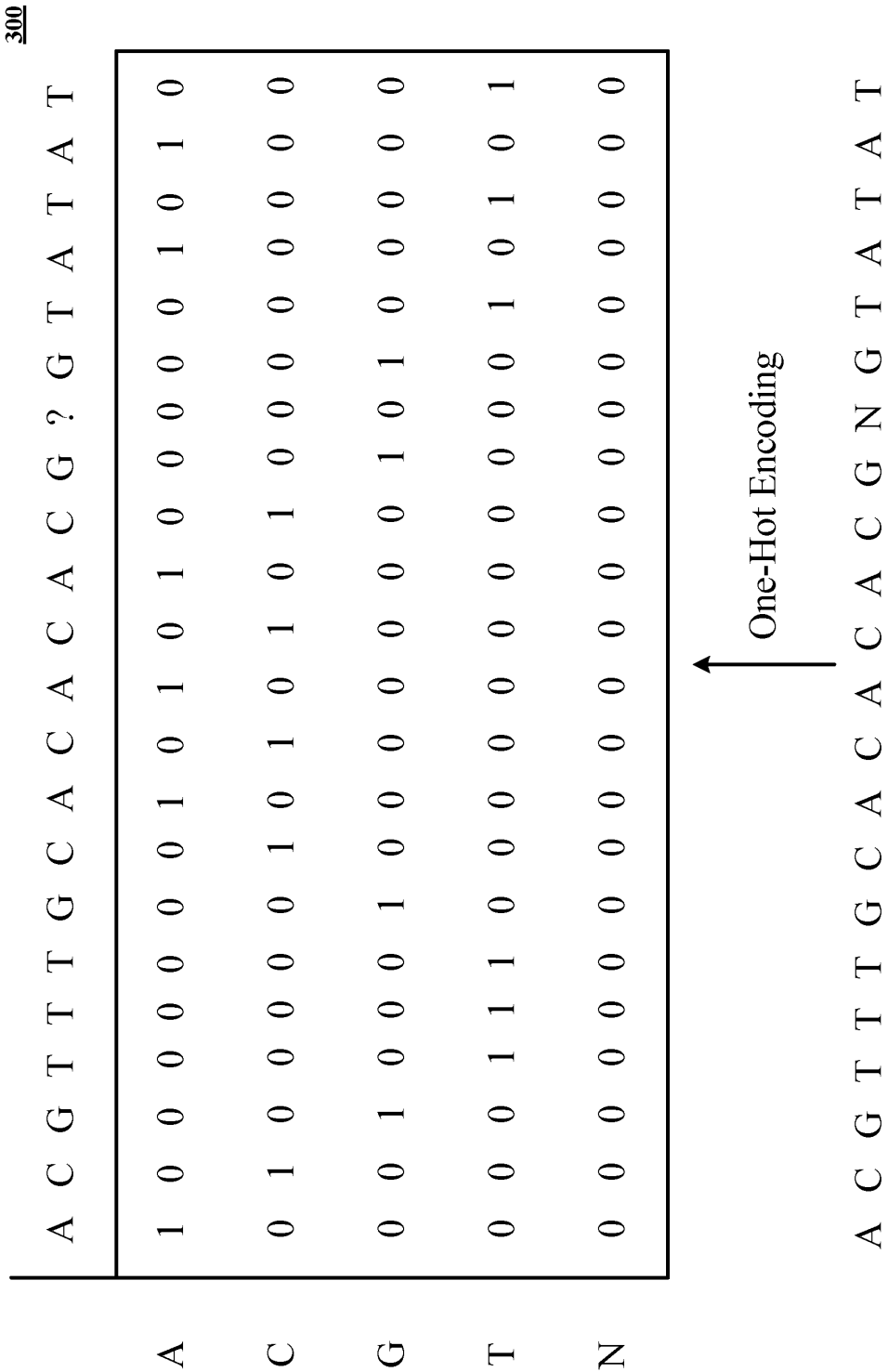


FIG. 3

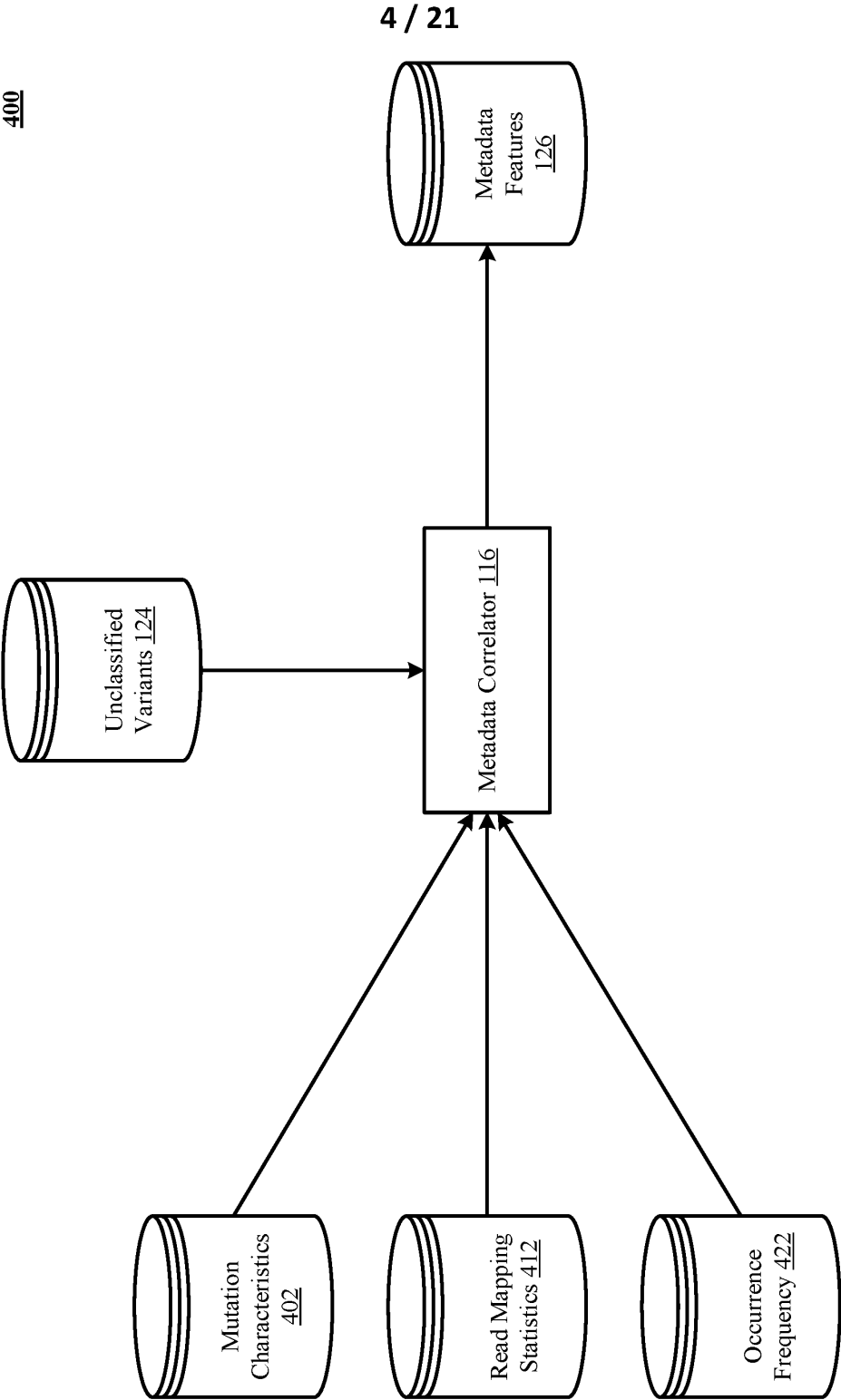


FIG. 4

500A

The features: context features

chr_all_ref_all_pos_all_alt_all_A alt_all_C alt_all_G alt_all_T alt_all_Other
VAF_all_dp_all_type_all_snv type_all_insertion type_all_deletion dbsnp_all
oneKg_all evs_all_exac_all nonsyn_all_true_nonsyn_all_false phylop_all
phylop_all_NA clinvarSig_all_drug_response clinvarSig_all_uncertain_significance
clinvarSig_all_likely_pathogenic clinvarSig_all_pathogenic clinvarSig_all_not
provided clinvarSig_all_nan clinvarSig_all_likely_benign clinvarSig_all_benign
clinvarSig_all_other polyPhen_all_benign polyPhen_all_possibly_damaging
polyPhen_all_nan polyPhen_all_probably_damaging polyPhen_all_unknown
sift_all_tolerated sift_all_deleterious - low confidence sift_all_nan
sift_all_deleterious sift_all_tolerated - low confidence CNT gnomadExome_allAf
gnomadExome_afrAf gnomadExome_asmrAf gnomadExome_easAf
gnomadExome_finAf gnomadExome_nfeAf gnomadExome_asjAf
gnomadExome_sasAfgnomadExome_othAf gnomad_allAf gnomad_afrAf
gnomad_asmrAf gnomad_easAf gnomad_finAf gnomad_nfeAf gnomad_asjAf
gnomad_sasAf gnomad_othAf ethno_P_AFR ethno_P_AMR ethno_P_EAS
ethno_P_EUR ethno_P_SAS

FIG. 5A

500B

The features: sequencing features

chr_all ref_all pos_all alt_all_A alt_all_C alt_all_G alt_all_T alt_all_Other
VAF_all dp_all type_all snv type_all insertion type_all deletion dbsnp_all
oneKg_all evs_all exac_all nonsyn_all true nonsyn_all_false phylop_all
phylop_all_NA clinvarSig_all drug response clinvarSig_all_uncertain significance
clinvarSig_all_likely pathogenic clinvarSig_all_pathogenic clinvarSig_all_not
provided clinvarSig_all_nan clinvarSig_all_likely benign clinvarSig_all_benign
clinvarSig_all_other polyPhen_all_benign polyPhen_all_possibly damaging
polyPhen_all_nan polyPhen_all_probably damaging polyPhen_all_unknown
sift_all_tolerated sift_all_deleterious - low confidence sift_all_nan
sift_all_deleterious sift_all_tolerated - low confidence CNT gnomadExome_allAf
gnomadExome_afrAf gnomadExome_asmrAf gnomadExome_easAf
gnomadExome_finAf gnomadExome_nfeAf gnomadExome_asjAf
gnomadExome_sasAfgnomadExome_othAf gnomad_allAf gnomad_afrAf
gnomad_asmrAf gnomad_easAf gnomad_finAf gnomad_nfeAf gnomad_asjAf
gnomad_sasAf gnomad_othAf ethno_P_AFR ethno_P_AMR ethno_P_EAS
ethno_P_EUR ethno_P_SAS

FIG. 5B

500C

The features: functional features

chr_all ref_all pos_all alt_all_A alt_all_C alt_all_G alt_all_T alt_all_Other
VAF_all dp_all type_all_snv type_all_insertion type_all_deletion dbsnp_all
oneKg_all evs_all exac_all nonsyn_all_true nonsyn_all_false phyloP_all
phyloP_all_NA clinvarSig_all_drug_response clinvarSig_all_uncertain significance
clinvarSig_all_likely pathogenic clinvarSig_all_pathogenic clinvarSig_all_not
provided clinvarSig_all_nan clinvarSig_all_likely benign clinvarSig_all_benign
clinvarSig_all_other polyPhen_all_benign polyPhen_all_possibly damaging
polyPhen_all_nan polyPhen_all_probably damaging polyPhen_all_unknown
sift_all_tolerated sift_all_deleterious_low confidence sift_all_nan
sift_all_deleterious sift_all_tolerated_low confidence CNT gnomadExome_allAf
gnomadExome_afrAf gnomadExome_asmrAf gnomadExome_easAf
gnomadExome_finAf gnomadExome_nfeAf gnomadExome_asjAf
gnomadExome_sasAfgnomadExome_othAf gnomad_allAf gnomad_afrAf
gnomad_asmrAf gnomad_easAf gnomad_finAf gnomad_nfeAf gnomad_asjAf
gnomad_sasAf gnomad_othAf ethno_P_AFR ethno_P_AMR ethno_P_EAS
ethno_P_EUR ethno_P_SAS

FIG. 5C

500D

The features: population features

chr_all_ref_all_pos_all alt_all_A alt_all_C alt_all_G alt_all_T alt_all_Other
VAF_all dp_all type_all snv type_all_insertion type_all_deletion dbsnp_all
onekg_all evs_all exac_all nonsyn_all_true nonsyn_all_false phylop_all
phylop_all_NA clinvarSig_all_drug_response clinvarSig_all_uncertain significance
clinvarSig_all_likely pathogenic clinvarSig_all_pathogenic clinvarSig_all_not
provided clinvarSig_all_nan clinvarSig_all_likely benign clinvarSig_all_benign
clinvarSig_all_other polyPhen_all_benign polyPhen_all_possibly_damaging
polyPhen_all_nan polyPhen_all_probably_damaging polyPhen_all_unknown
sift_all_tolerated sift_all_deleterious_low confidence sift_all_nan
sift_all_deleterious sift_all_tolerated_low confidence CNT gnomadExome_allAf
gnomadExome_afrAf gnomadExome_asrAf gnomadExome_easAf
gnomadExome_finAf gnomadExome_nfeAf gnomadExome_asjAf
gnomadExome_sasA gnomadExome_othAf gnomad_allAf gnomad_afrAf
gnomad_asrAf gnomad_easAf gnomad_finAf gnomad_nfeAf gnomad_asjAf
gnomad_sasAf gnomad_othAf ethno_P_AFR ethno_P_AMR ethno_P_EAS
ethno_P_EUR ethno_P_SAS

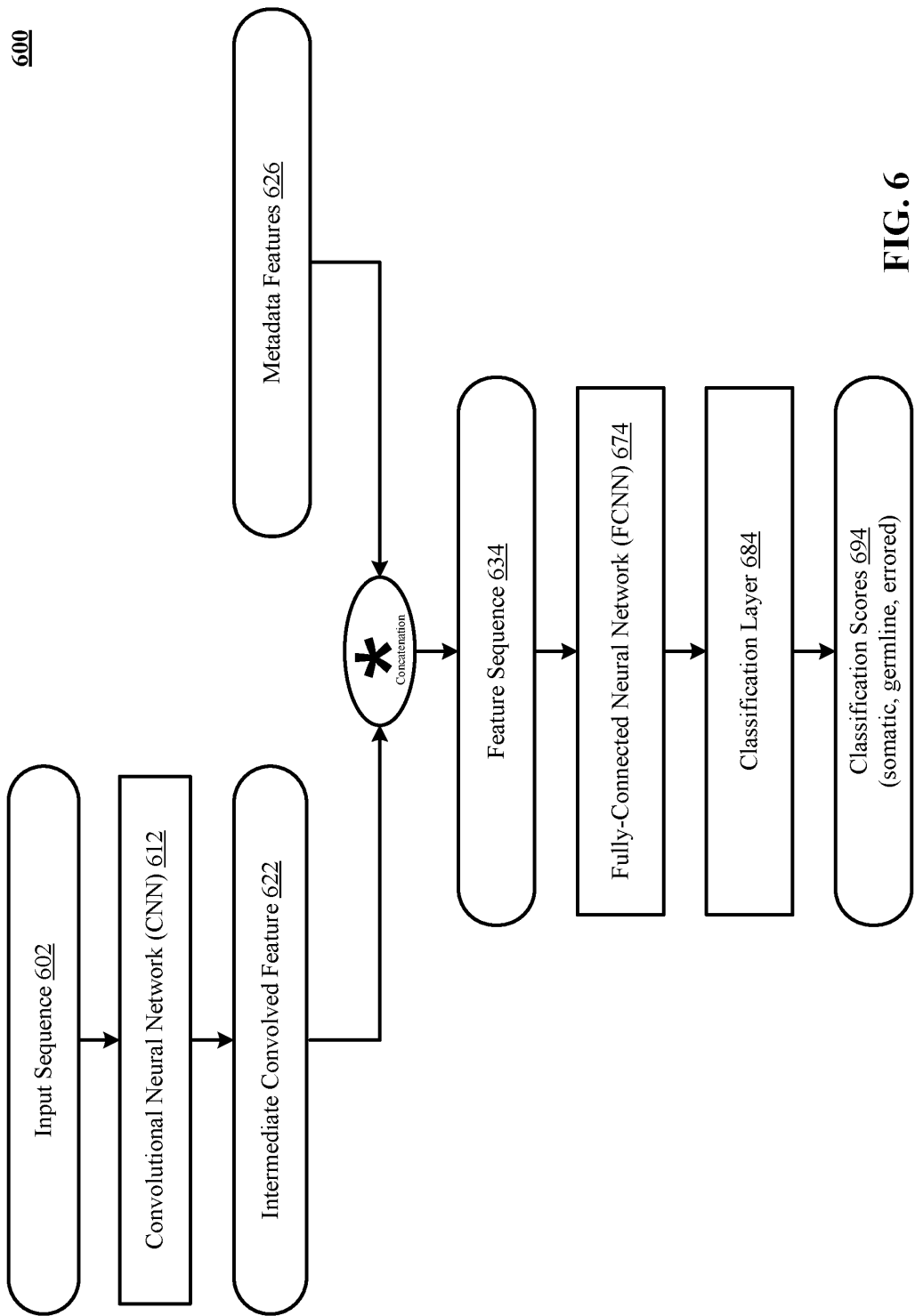
FIG. 5D

500E

The features: ethnicity features

chr_all_ref_all_pos_all_alt_all_A alt_all_C alt_all_G alt_all_T alt_all_Other
VAF_all_dp_all_type_all_snv type_all_insertion type_all_deletion dbsnp_all
oneKg_all evs_all_exac_all nonsyn_all_true_nonsyn_all_false phyloP_all
phyloP_all_NA clinvarSig_all_drug_response clinvarSig_all_uncertain_significance
clinvarSig_all_likely_pathogenic clinvarSig_all_pathogenic clinvarSig_all_not
provided clinvarSig_all_nan clinvarSig_all_likely_benign clinvarSig_all_benign
clinvarSig_all_other polyPhen_all_benign polyPhen_all_possibly_damaging
polyPhen_all_nan polyPhen_all_probably_damaging polyPhen_all_unknown
sift_all_tolerated sift_all_deleterious_low_confidence sift_all_nan
sift_all_deleterious sift_all_tolerated_low_confidence CNT gnomadExome_allAf
gnomadExome_afrAf gnomadExome_asrAf gnomadExome_easAf
gnomadExome_finAf gnomadExome_nfeAf gnomadExome_asjAf
gnomadExome_sasAfgnomadExome_othAf gnomad_allAf gnomad_afrAf
gnomad_asmrAf gnomad_easAf gnomad_finAf gnomad_nfeAf gnomad_asjAf
gnomad_sasAf gnomad_othAf ethno_P_AFR ethno_P_AMR ethno_P_EAS
ethno_P_EUR ethno_P_SAS

FIG. 5E



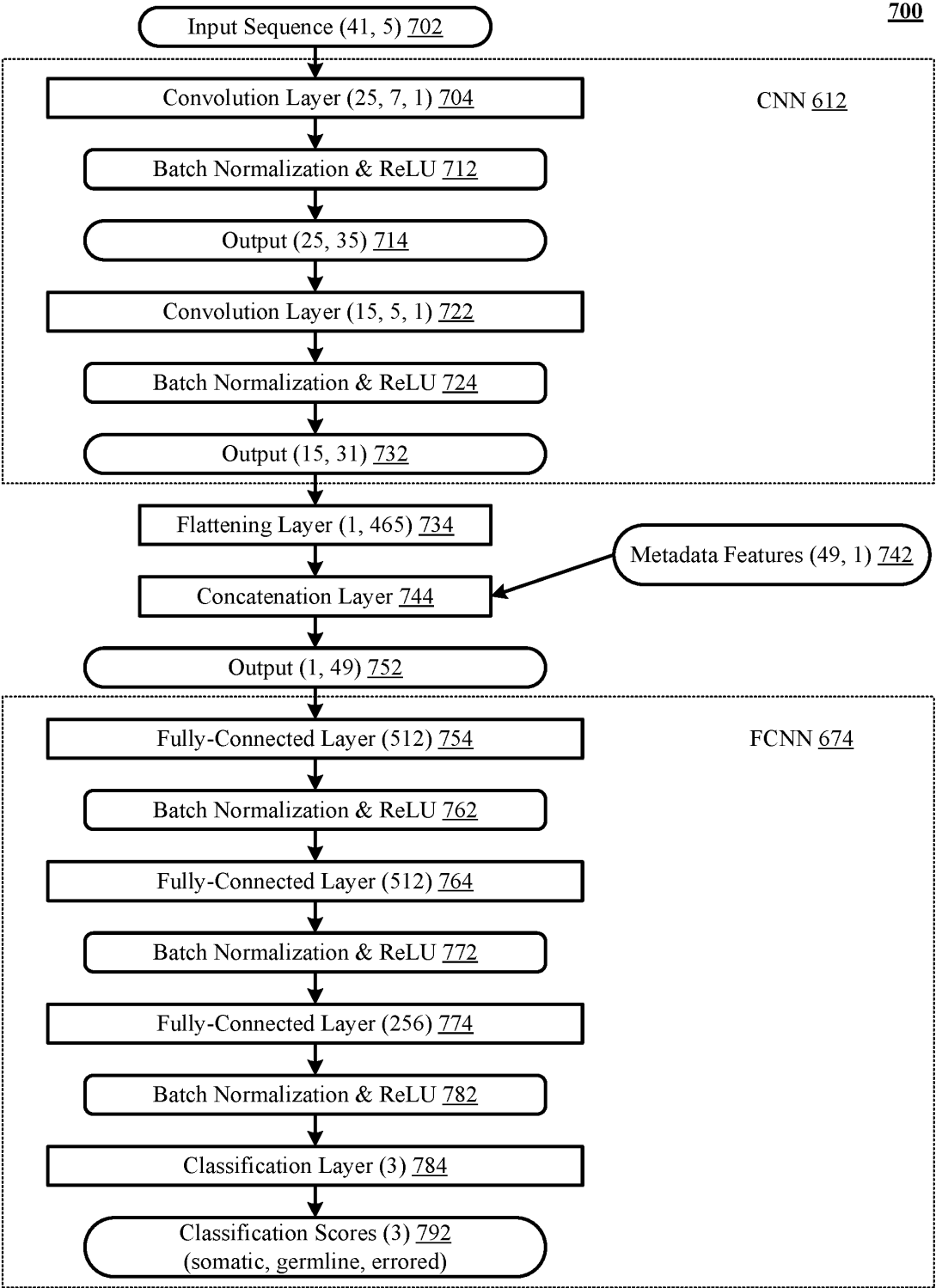


FIG. 7

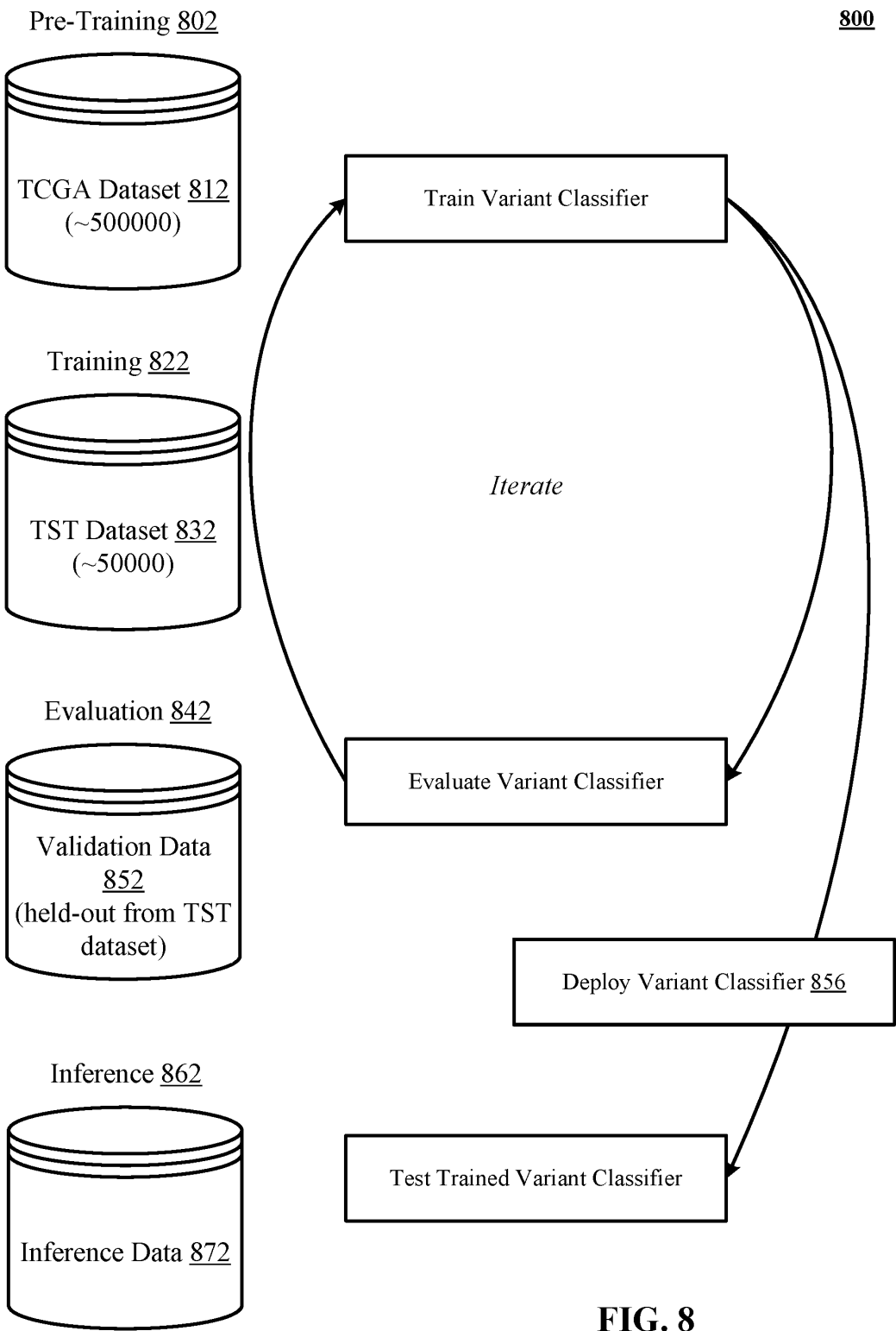


FIG. 8

Comparison Results – Sensitivity/Specificity

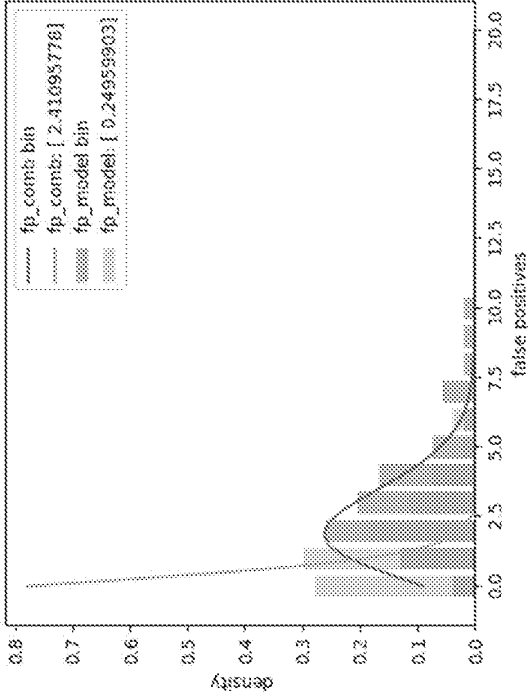
Performance on Test Set (Exons)

	Sensitivity	Specificity
Sojourner - Best F1	0.891509	0.991798
Sojourner - Default	0.786626	0.996810
Non-Deep Neural Network Classifier - Combined Filter (db threshold 1)	0.636424	0.994201
Non-Deep Neural Network Classifier - Combined Filter (db threshold 2)	0.736895	0.993484
Non-Deep Neural Network Classifier - Combined Filter (db threshold 3)	0.777889	0.992996
Non-Deep Neural Network Classifier - Combined Filter (db threshold 5)	0.831710	0.992475

FIG. 9

1000

False Positive Analysis (Exons)



- Using all 54 test samples
- Orange: Sojourner
- Blue: Non-Deep Neural Network Classifier
 - db cutoff 1 (most stringent)
 - Lines are best Poisson fit, with Poisson parameter in the legend
- U-test two-sided p-value: 5.3206e-06

FIG. 10

1100

**Mean Absolute Adjusted TMB Difference and Confidence Interval
(TMB<30) (Exons)**

- **Soujourner**
 - Confidence interval 1.00229046316 - 2.15874919358
 - Mean 1.5291033807
 - Std 0.297960960183
- **Non-Deep Neural Network Classifier**
 - Confidence interval 1.47142489617 - 3.33821858223
 - Mean 2.33678889251
 - Std 0.484091917501

FIG. 11

1200

Mean absolute adjusted TMB difference and confidence interval
(TMB<30) (Exons)

- Comparison
- Soujourner has lower TMB delta: 9269/10000 bootstraps

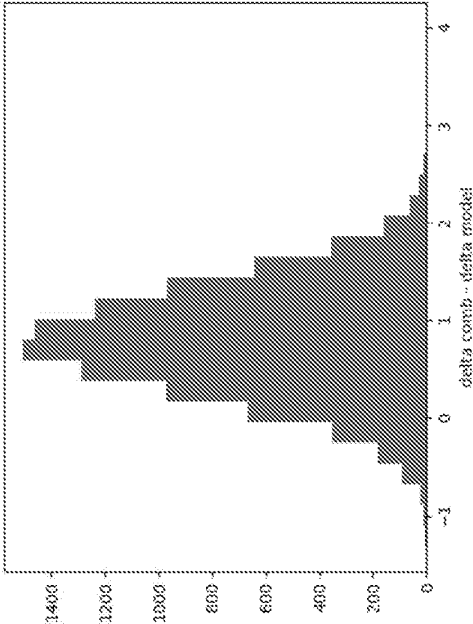


FIG. 12

Comparison Results – Sensitivity/Specificity

1300

Performance on Test Set (CDS)

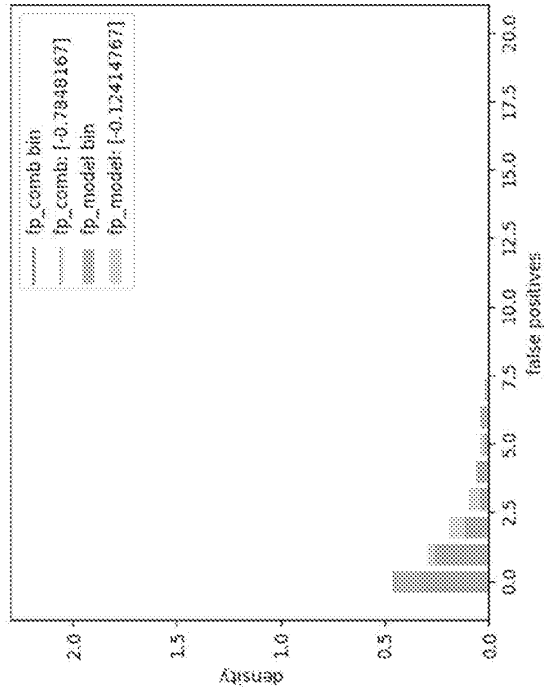
	Sensitivity	Specificity
Sojourner – Ratio 97	0.754998	0.998112
Sojourner - Default	0.813798	0.997111
Non-Deep Neural Network Classifier - Combined Filter (db threshold 1)	0.645123	0.997380
Non-Deep Neural Network Classifier - Combined Filter (db threshold 2)	0.751664	0.996610
Non-Deep Neural Network Classifier - Combined Filter (db threshold 3)	0.792792	0.996109
Non-Deep Neural Network Classifier - Combined Filter (db threshold 5)	0.852722	0.995609

17 / 21

FIG. 13

1400

False Positive Analysis (CDS)



- Using all 54 test samples
- Orange: Sojourner
- Blue: Non-Deep Neural Network Classifier
 - db cutoff 1 (most stringent)
 - Lines are best Poisson fit, with Poisson parameter in the legend
- U-test two-sided p-value: 0.2507

FIG. 14

1500

Mean Absolute Adjusted TMB Difference and Confidence Interval
(TMB<30) (CDS)

- Soujourner
 - Confidence interval 1.09335716413 - 2.78749505783
 - Mean 1.83634036739
 - Std 0.459655305695
- Non-Deep Neural Network Classifier
 - Confidence interval 0.988798777219 - 2.40507762924
 - Mean 1.66277809651
 - Std 0.357688784948

FIG. 15

1600

**Mean absolute adjusted TMB difference and confidence interval
(TMB<30) (CDS)**

- Comparison
- Soujourner has lower TMB delta: 3905/10000 bootstraps

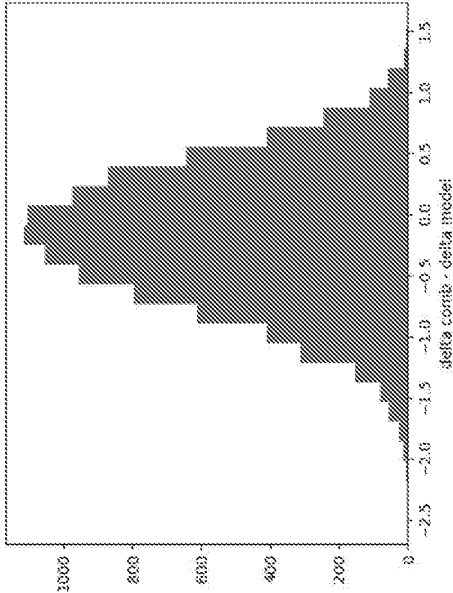


FIG. 16

