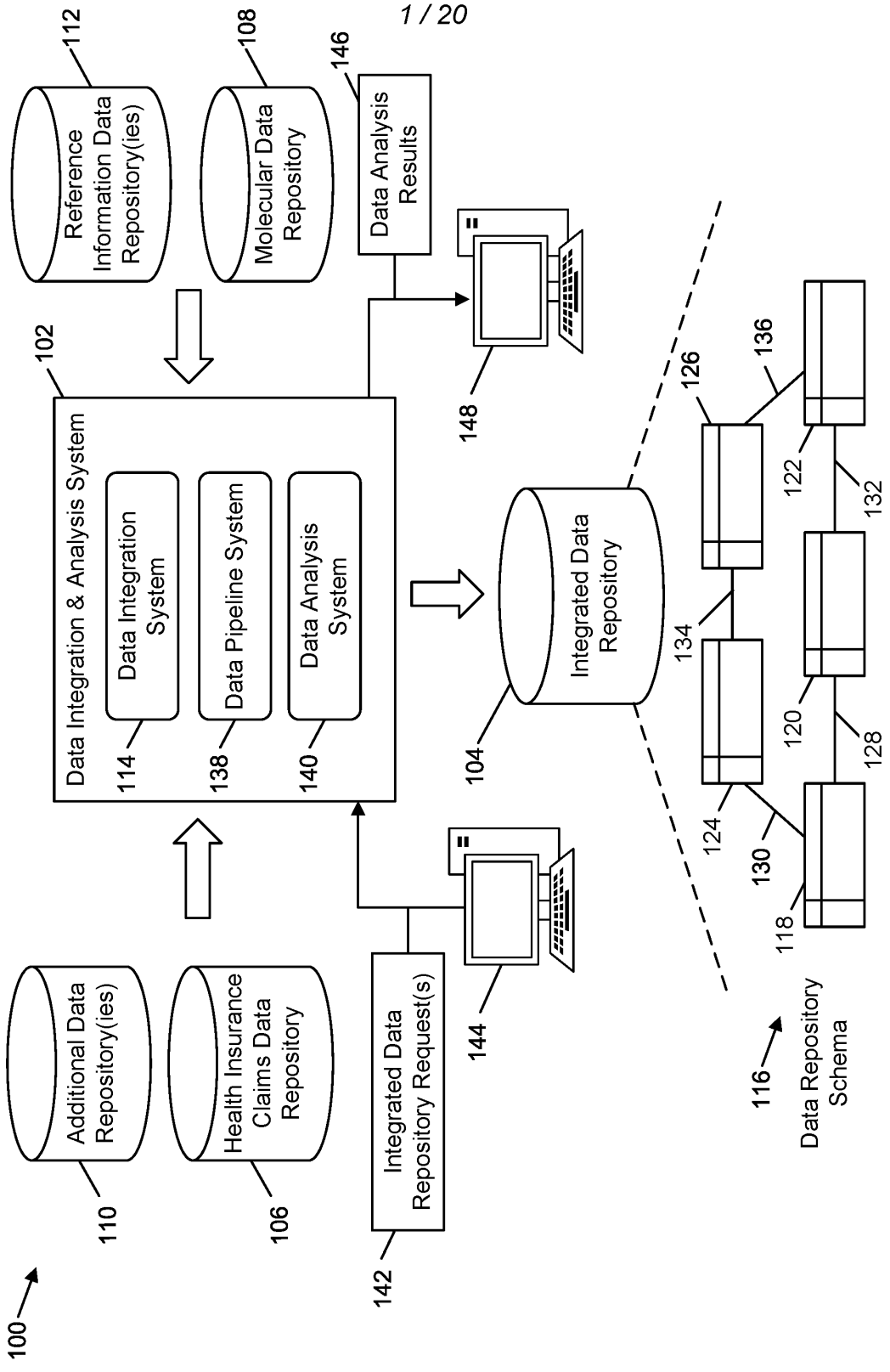

100
Additional Data Repository(ies)
110
Health Insurance Claims Data Repository
106
Integrated Data Repository Request(s)
142
144
Data Integration & Analysis System
102
Data Integration System
114
Data Pipeline System
138
Data Analysis System
140
Reference Information Data Repository(ies)
112
Molecular Data Repository
108
Data Analysis Results
146
148
Integrated Data Repository
104
116
Data Repository Schema
118
120
122
124
126
128
130
132
134
136

Figure 1

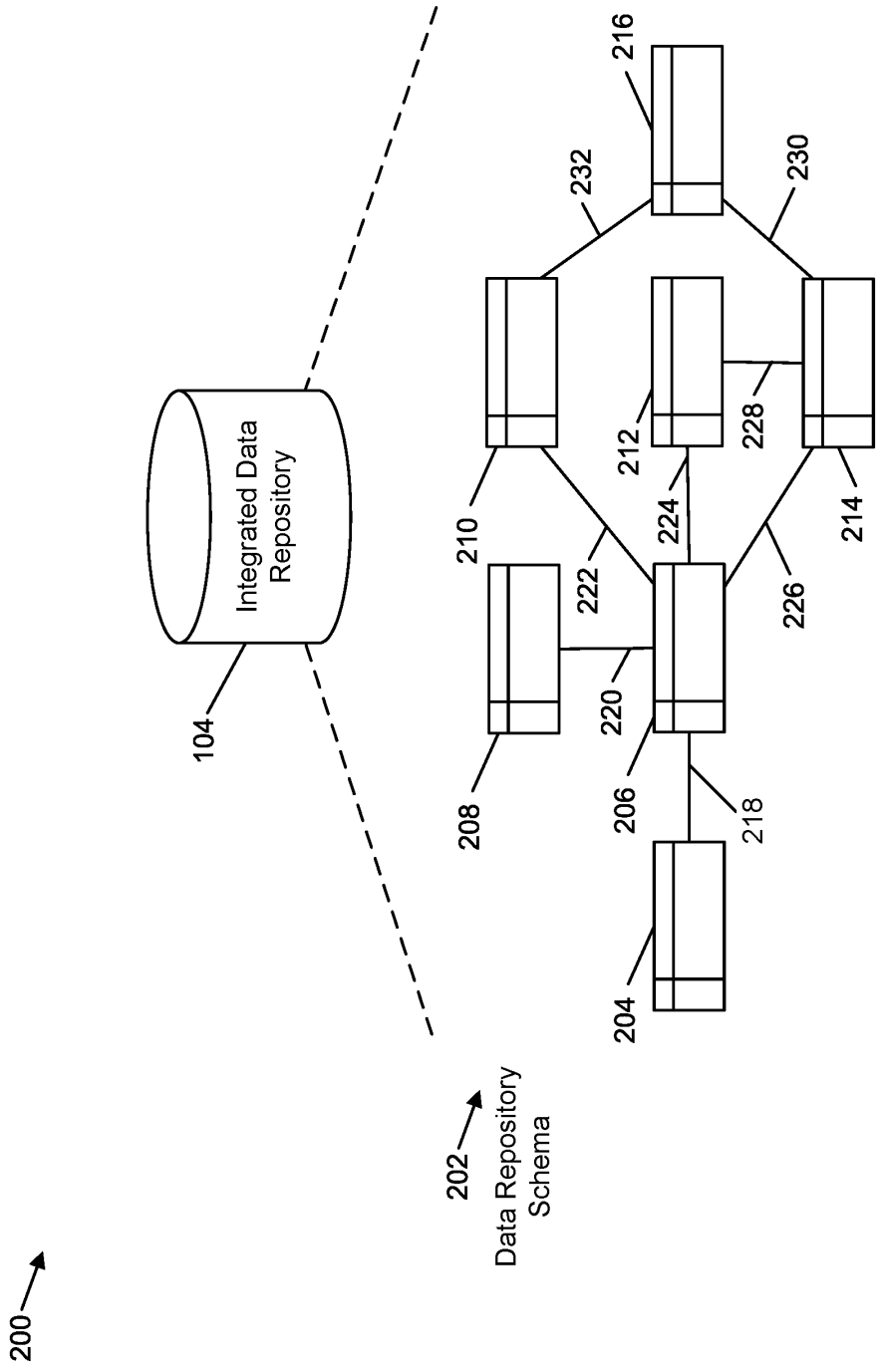

200
Integrated Data Repository
104
Data Repository Schema
202
204
206
208
210
212
214
216
218
220
222
224
226
228
230
232

Figure 2

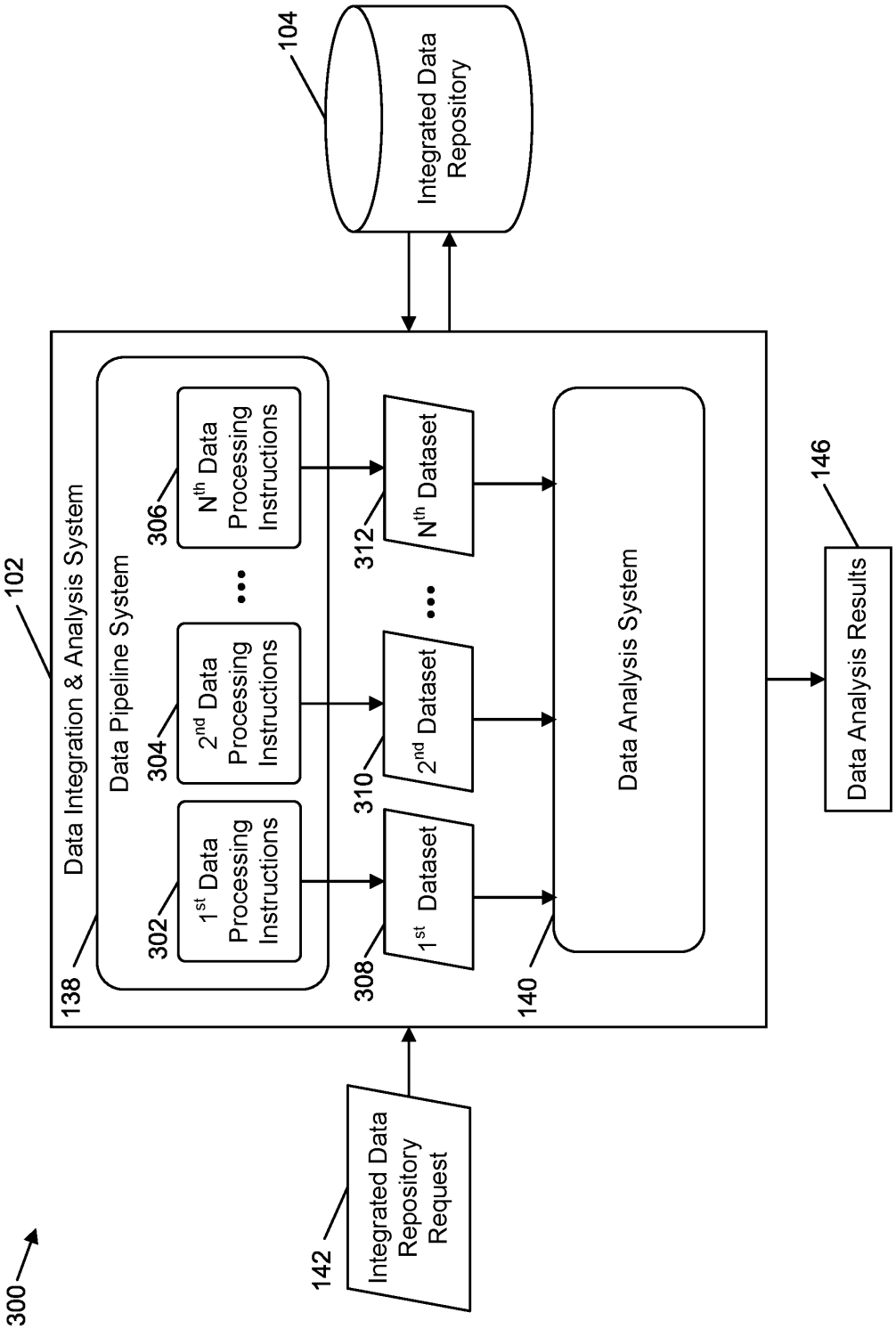

300
142
Integrated Data Repository Request
102
Data Integration & Analysis System
138
Data Pipeline System
302
1st Data Processing Instructions
304
2nd Data Processing Instructions
306
Nth Data Processing Instructions
308
1st Dataset
310
2nd Dataset
312
Nth Dataset
140
Data Analysis System
146
Data Analysis Results
104
Integrated Data Repository

Figure 3

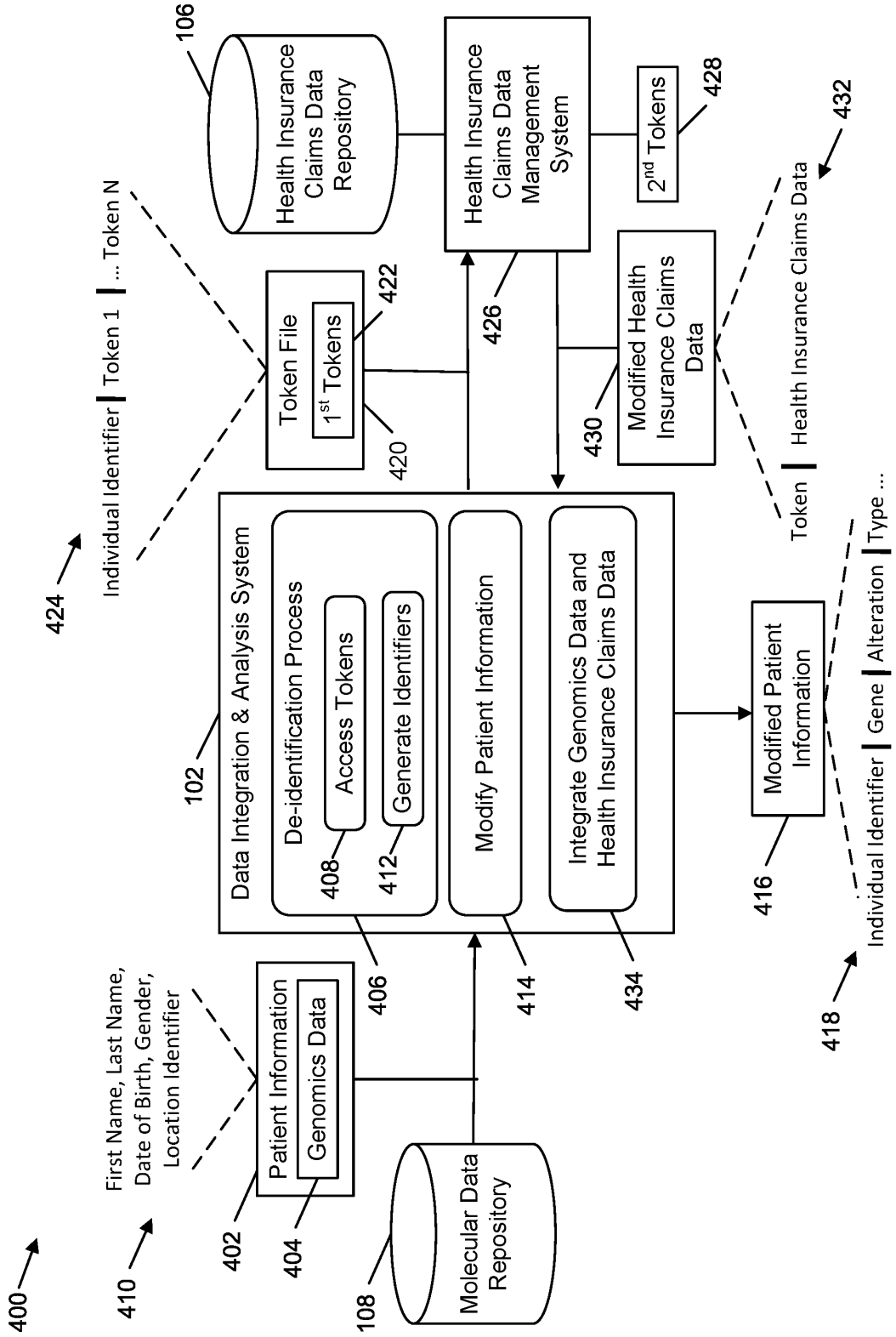
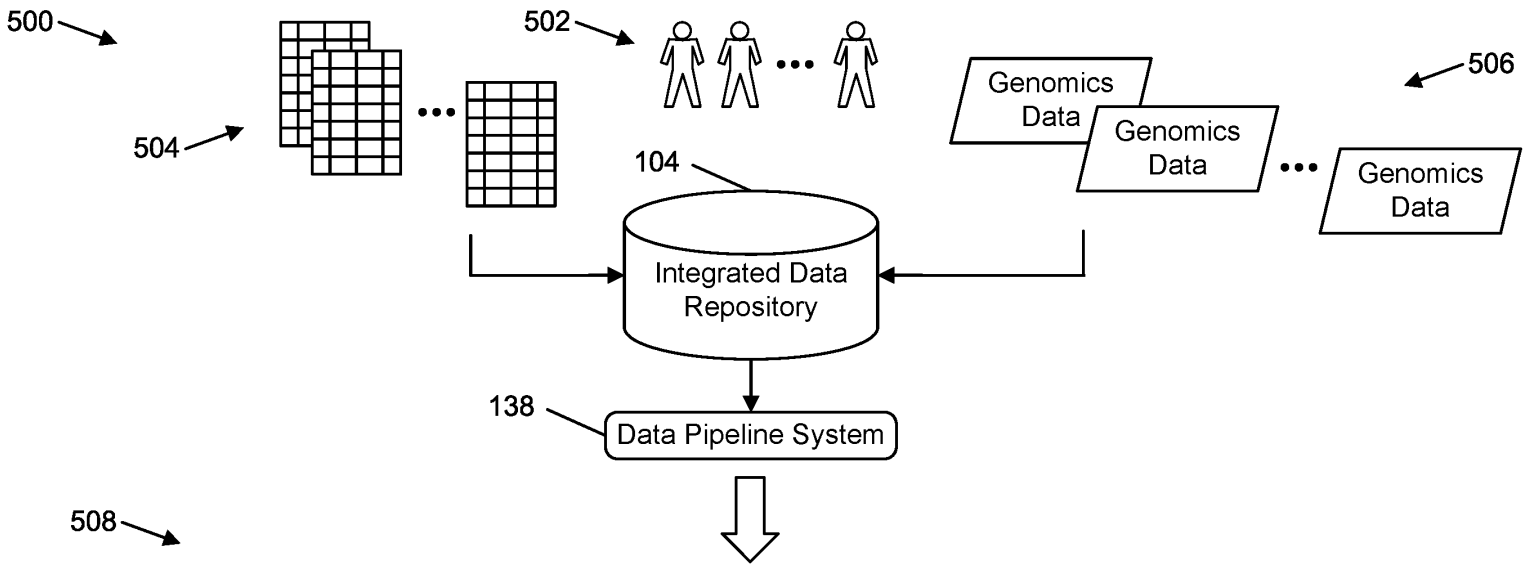

400
410
First Name, Last Name, Date of Birth, Gender, Location Identifier
402
Patient Information
404
Genomics Data
108
Molecular Data Repository
406
414
434
102
Data Integration & Analysis System
De-identification Process
408
Access Tokens
412
Generate Identifiers
Modify Patient Information
Integrate Genomics Data and Health Insurance Claims Data
424
Individual Identifier | Token 1 | ... Token N
Token File
1st Tokens
422
420
106
Health Insurance Claims Data Repository
426
Health Insurance Claims Data Management System
2nd Tokens
428
430
Modified Health Insurance Claims Data
432
Health Insurance Claims Data
Token
416
Modified Patient Information
418
Individual Identifier | Gene | Alteration | Type ...

Figure 4




| Patient ID | Insurance Transaction Identifier | ICD Code | Transaction Start Date | Transaction End Date | Lung Cancer Cohort | Confidence Level |
|---|---|---|---|---|---|---|
| ABC123 | 118UUW | C34 | 09-06-97 | 09-06-97 | Yes | High |
| | 762BBG | C34 | 12-11-97 | 12-11-97 | | |
| | 379NHJ | C162 | 06-15-98 | 06-17-98 | | |
| LMN001 | 992QAX | C34 | 04-22-95 | 04-23-95 | Yes | Low |
| | 659NHE | C64 | 10-19-95 | 10-19-95 | | |
| | 748VKI | C162 | 05-28-97 | 05-29-97 | | |
| P4O7Q3 | 014OSZ | C37 | 02-09-98 | 02-09-98 | No | Medium |
| | LLY821 | C162 | 02-09-98 | 02-09-98 | | |

Figure 5

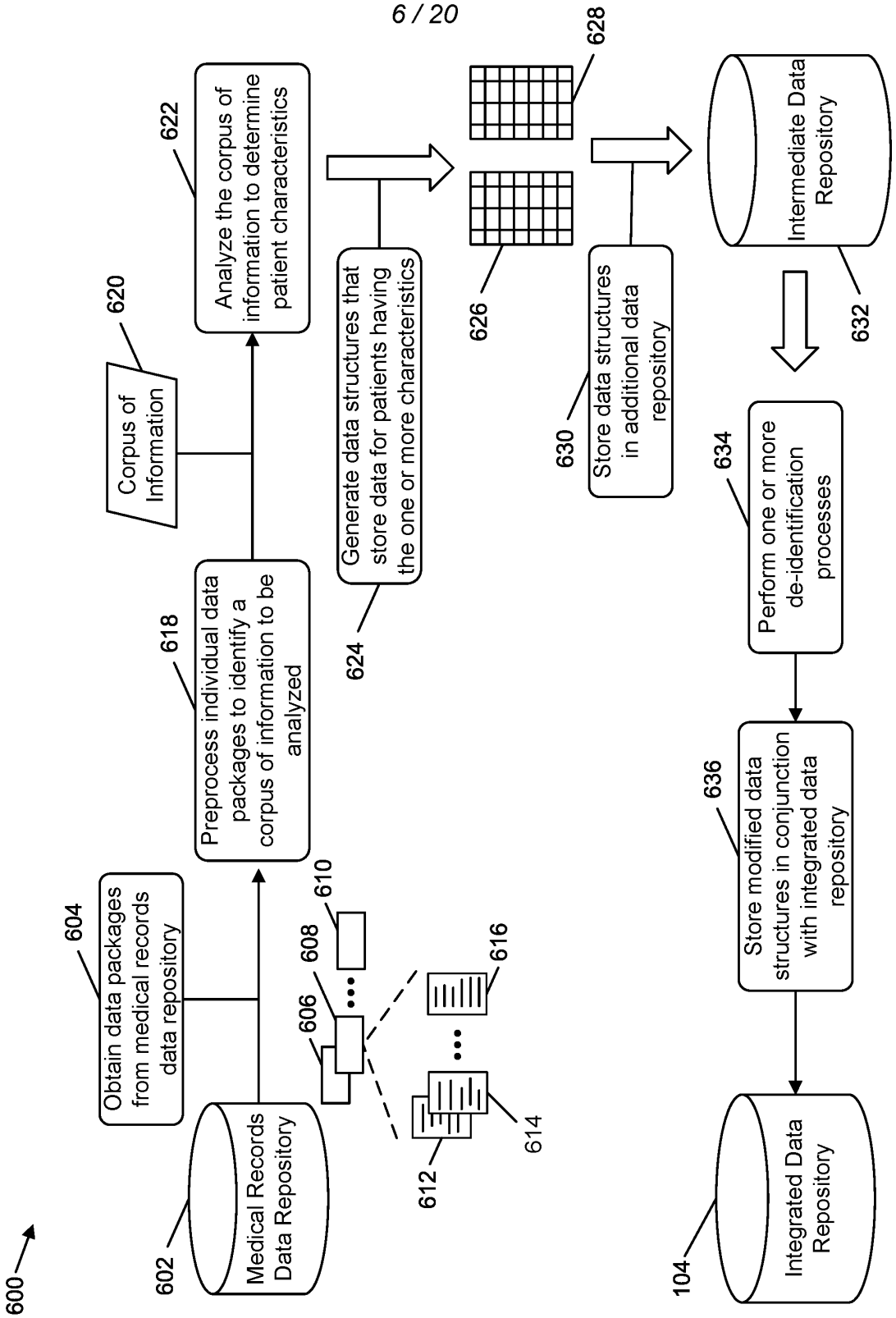

600
Medical Records Data Repository
602
Obtain data packages from medical records data repository
604
610
608
606
616
614
612
Preprocess individual data packages to identify a corpus of information to be analyzed
618
Corpus of Information
620
Analyze the corpus of information to determine patient characteristics
622
Generate data structures that store data for patients having the one or more characteristics
624
626
628
Store data structures in additional data repository
630
Intermediate Data Repository
632
Perform one or more de-identification processes
634
Store modified data structures in conjunction with integrated data repository
636
Integrated Data Repository
104

Figure 6

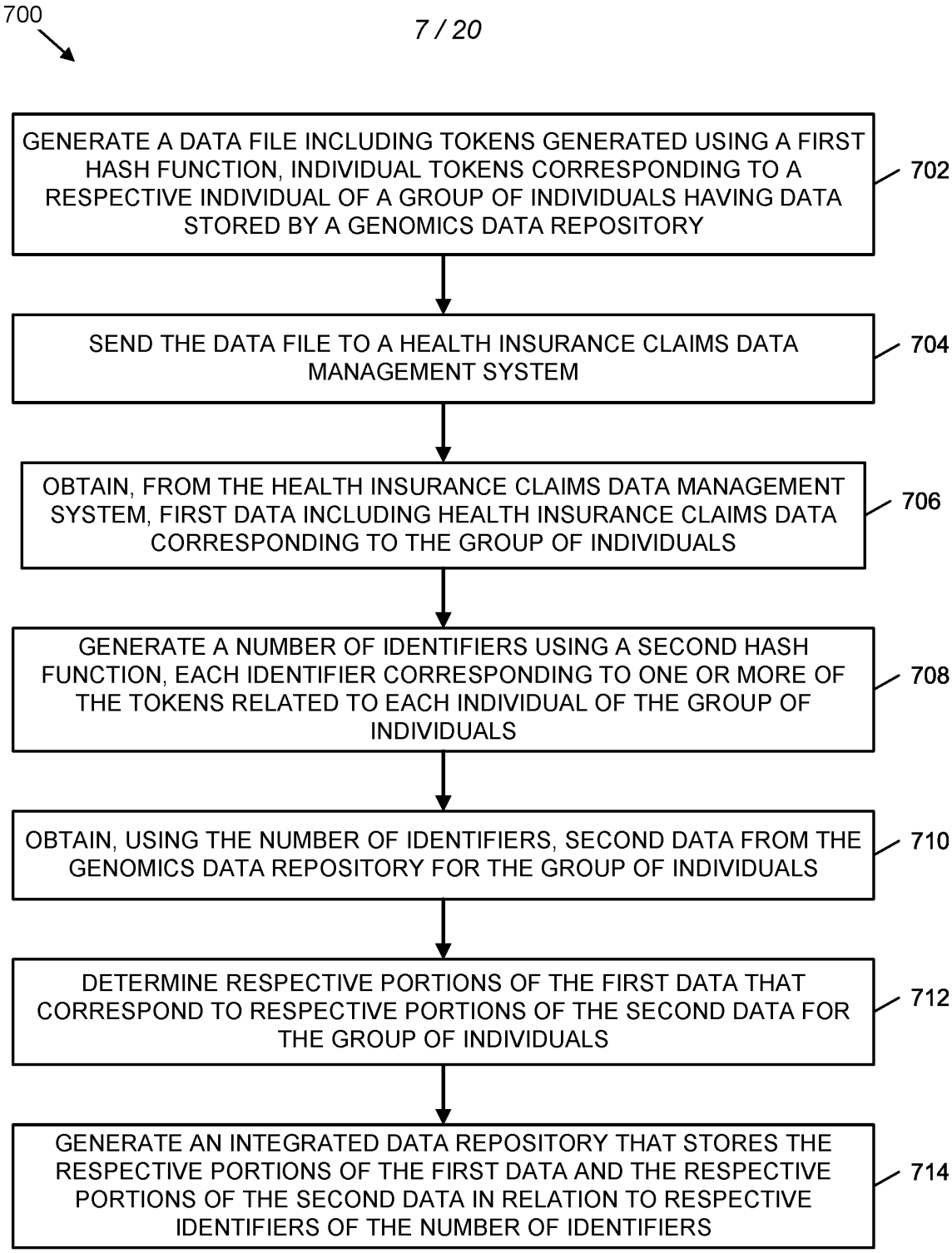

700
GENERATE A DATA FILE INCLUDING TOKENS GENERATED USING A FIRST HASH FUNCTION, INDIVIDUAL TOKENS CORRESPONDING TO A RESPECTIVE INDIVIDUAL OF A GROUP OF INDIVIDUALS HAVING DATA STORED BY A GENOMICS DATA REPOSITORY
702
SEND THE DATA FILE TO A HEALTH INSURANCE CLAIMS DATA MANAGEMENT SYSTEM
704
OBTAIN, FROM THE HEALTH INSURANCE CLAIMS DATA MANAGEMENT SYSTEM, FIRST DATA INCLUDING HEALTH INSURANCE CLAIMS DATA CORRESPONDING TO THE GROUP OF INDIVIDUALS
706
GENERATE A NUMBER OF IDENTIFIERS USING A SECOND HASH FUNCTION, EACH IDENTIFIER CORRESPONDING TO ONE OR MORE OF THE TOKENS RELATED TO EACH INDIVIDUAL OF THE GROUP OF INDIVIDUALS
708
OBTAIN, USING THE NUMBER OF IDENTIFIERS, SECOND DATA FROM THE GENOMICS DATA REPOSITORY FOR THE GROUP OF INDIVIDUALS
710
DETERMINE RESPECTIVE PORTIONS OF THE FIRST DATA THAT CORRESPOND TO RESPECTIVE PORTIONS OF THE SECOND DATA FOR THE GROUP OF INDIVIDUALS
712
GENERATE AN INTEGRATED DATA REPOSITORY THAT STORES THE RESPECTIVE PORTIONS OF THE FIRST DATA AND THE RESPECTIVE PORTIONS OF THE SECOND DATA IN RELATION TO RESPECTIVE IDENTIFIERS OF THE NUMBER OF IDENTIFIERS
714

Figure 7

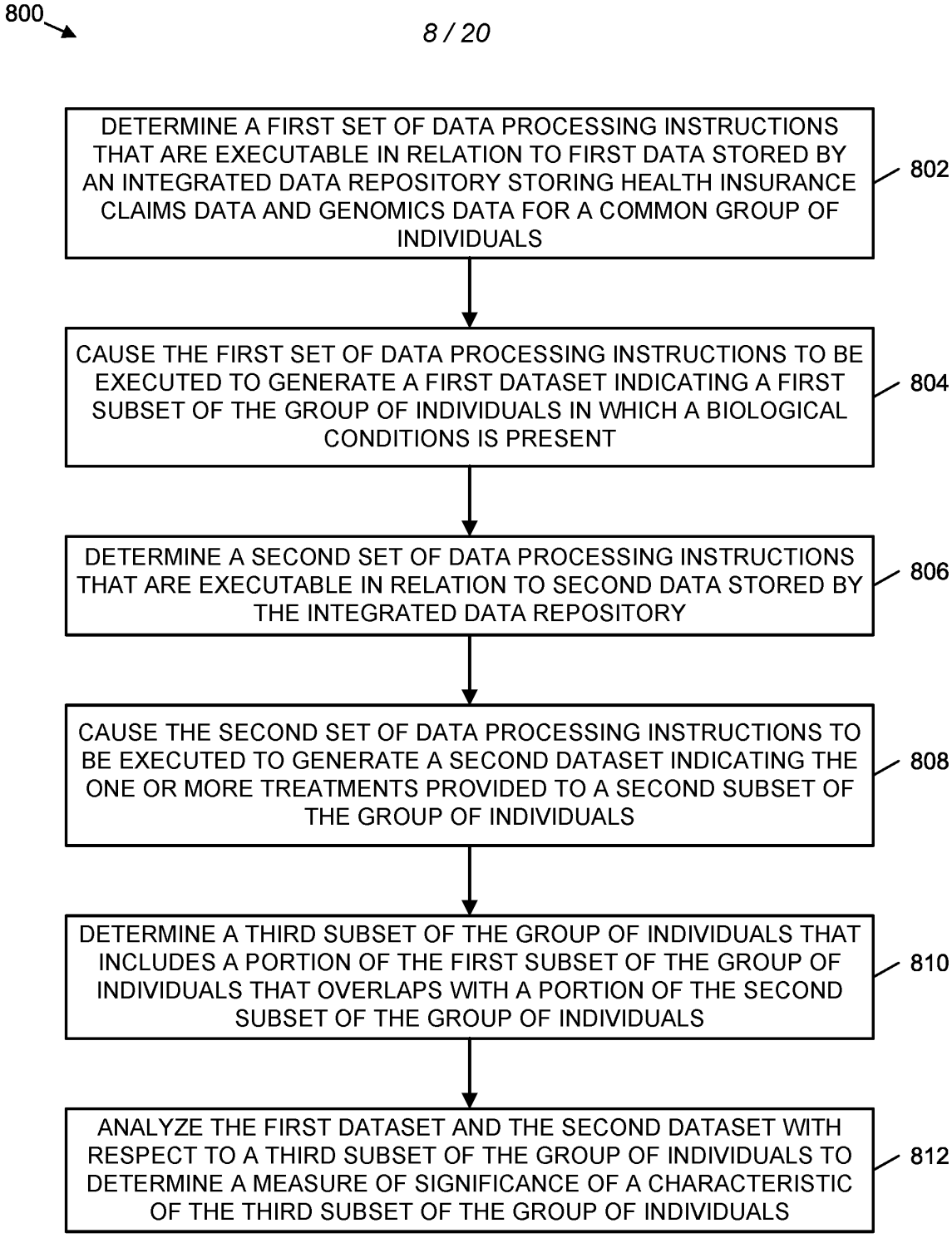

800
DETERMINE A FIRST SET OF DATA PROCESSING INSTRUCTIONS THAT ARE EXECUTABLE IN RELATION TO FIRST DATA STORED BY AN INTEGRATED DATA REPOSITORY STORING HEALTH INSURANCE CLAIMS DATA AND GENOMICS DATA FOR A COMMON GROUP OF INDIVIDUALS
802
CAUSE THE FIRST SET OF DATA PROCESSING INSTRUCTIONS TO BE EXECUTED TO GENERATE A FIRST DATASET INDICATING A FIRST SUBSET OF THE GROUP OF INDIVIDUALS IN WHICH A BIOLOGICAL CONDITIONS IS PRESENT
804
DETERMINE A SECOND SET OF DATA PROCESSING INSTRUCTIONS THAT ARE EXECUTABLE IN RELATION TO SECOND DATA STORED BY THE INTEGRATED DATA REPOSITORY
806
CAUSE THE SECOND SET OF DATA PROCESSING INSTRUCTIONS TO BE EXECUTED TO GENERATE A SECOND DATASET INDICATING THE ONE OR MORE TREATMENTS PROVIDED TO A SECOND SUBSET OF THE GROUP OF INDIVIDUALS
808
DETERMINE A THIRD SUBSET OF THE GROUP OF INDIVIDUALS THAT INCLUDES A PORTION OF THE FIRST SUBSET OF THE GROUP OF INDIVIDUALS THAT OVERLAPS WITH A PORTION OF THE SECOND SUBSET OF THE GROUP OF INDIVIDUALS
810
ANALYZE THE FIRST DATASET AND THE SECOND DATASET WITH RESPECT TO A THIRD SUBSET OF THE GROUP OF INDIVIDUALS TO DETERMINE A MEASURE OF SIGNIFICANCE OF A CHARACTERISTIC OF THE THIRD SUBSET OF THE GROUP OF INDIVIDUALS
812

Figure 8

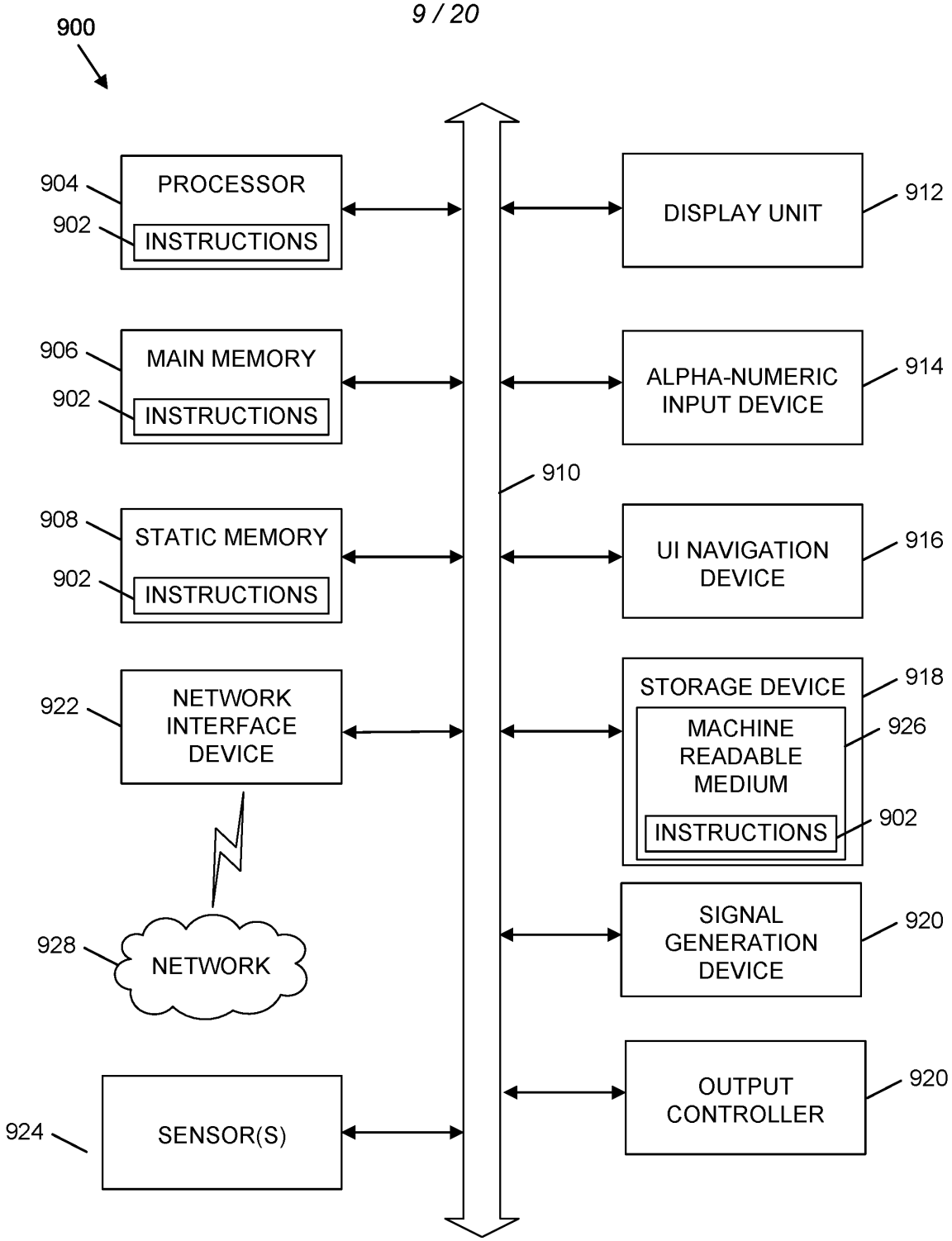

900
904
PROCESSOR
902
INSTRUCTIONS
906
MAIN MEMORY
902
INSTRUCTIONS
908
STATIC MEMORY
902
INSTRUCTIONS
922
NETWORK INTERFACE DEVICE
928
NETWORK
924
SENSOR(S)
910
DISPLAY UNIT
912
ALPHA-NUMERIC INPUT DEVICE
914
UI NAVIGATION DEVICE
916
STORAGE DEVICE
918
MACHINE READABLE MEDIUM
926
INSTRUCTIONS
902
SIGNAL GENERATION DEVICE
920
OUTPUT CONTROLLER
920

Figure 9

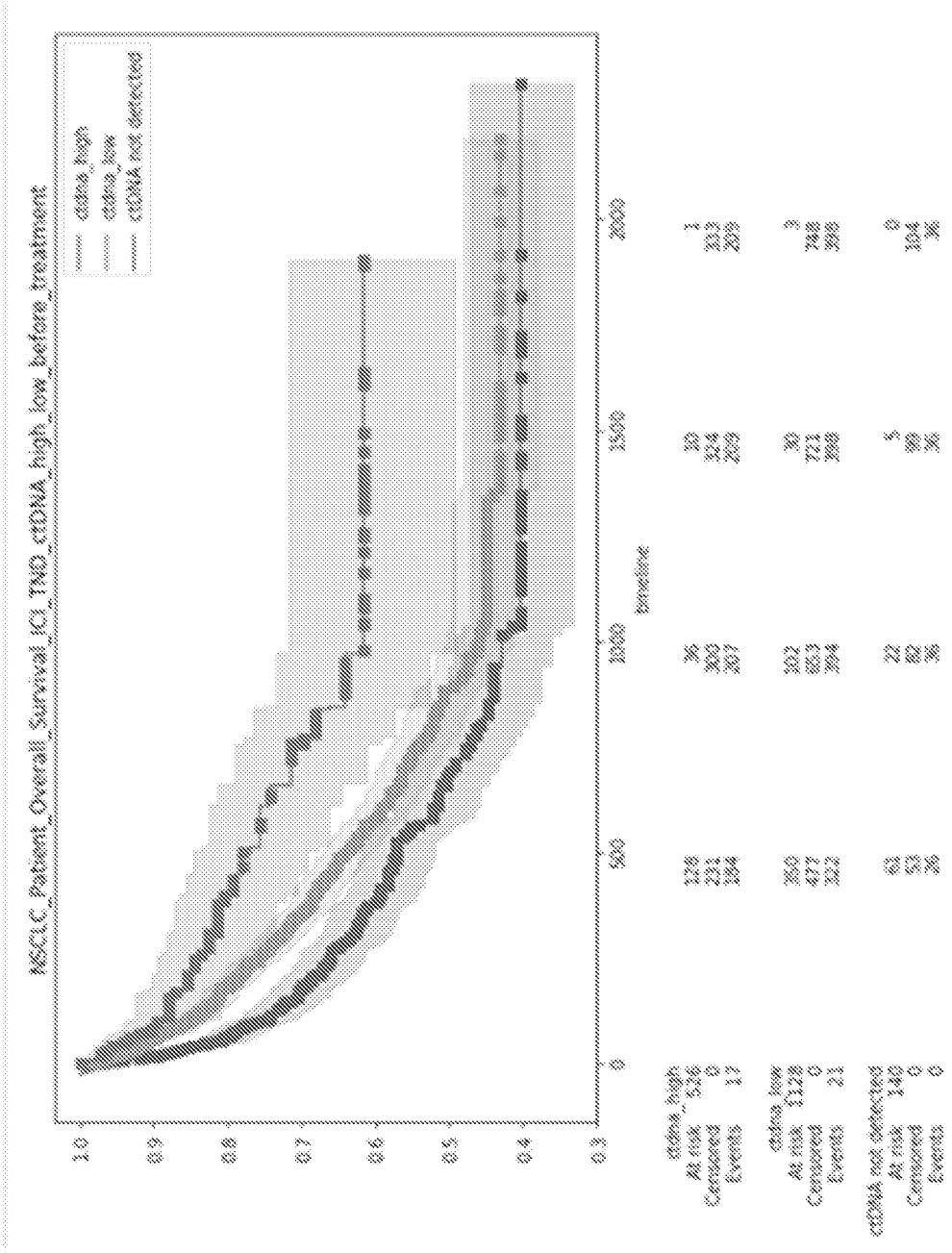

NSCLC_Patient_Overall_Survival_tO_TN0_ctDNA_high_low_before_treatment
ctdna_high
ctdna_low
ctDNA not detected
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
500
1000
1500
2000
timeline
ctdna_high
At risk 536 258 96 30 1
Censored 0 221 420 324 203
Events 12 134 207 128 20
ctdna_low
At risk 1128 560 132 30 3
Censored 0 67 603 721 248
Events 21 233 264 208 38
ctDNA not detected
At risk 140 62 22 5 0
Censored 0 36 68 88 104
Events 0 6 8 5 0

Figure 10

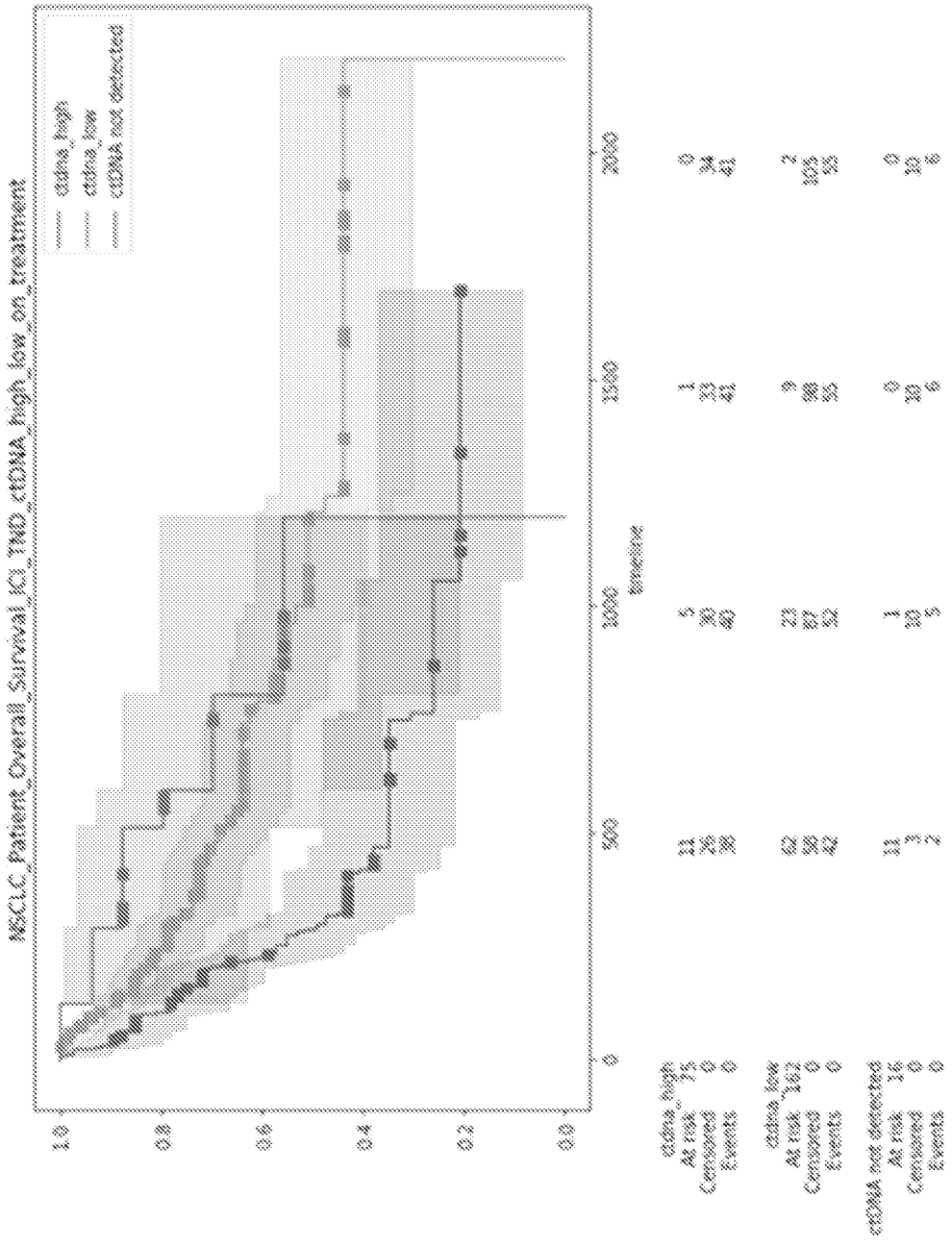

NSCLC_Patient_Overall_Survival_IO_TKD_ctDNA_high_low_on_treatment
ctdna_high
ctdna_low
ctDNA not detected
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
Timeline
ctdna_high
At risk 21 8 5 1 0
Censored 0 0 0 0 0
Events 0 8 13 18 21
ctdna_low
At risk 62 22 8 3 0
Censored 0 0 0 0 0
Events 0 40 54 59 62
ctDNA not detected
At risk 16 5 1 0 0
Censored 0 0 0 0 0
Events 0 11 15 16 16

Figure 11

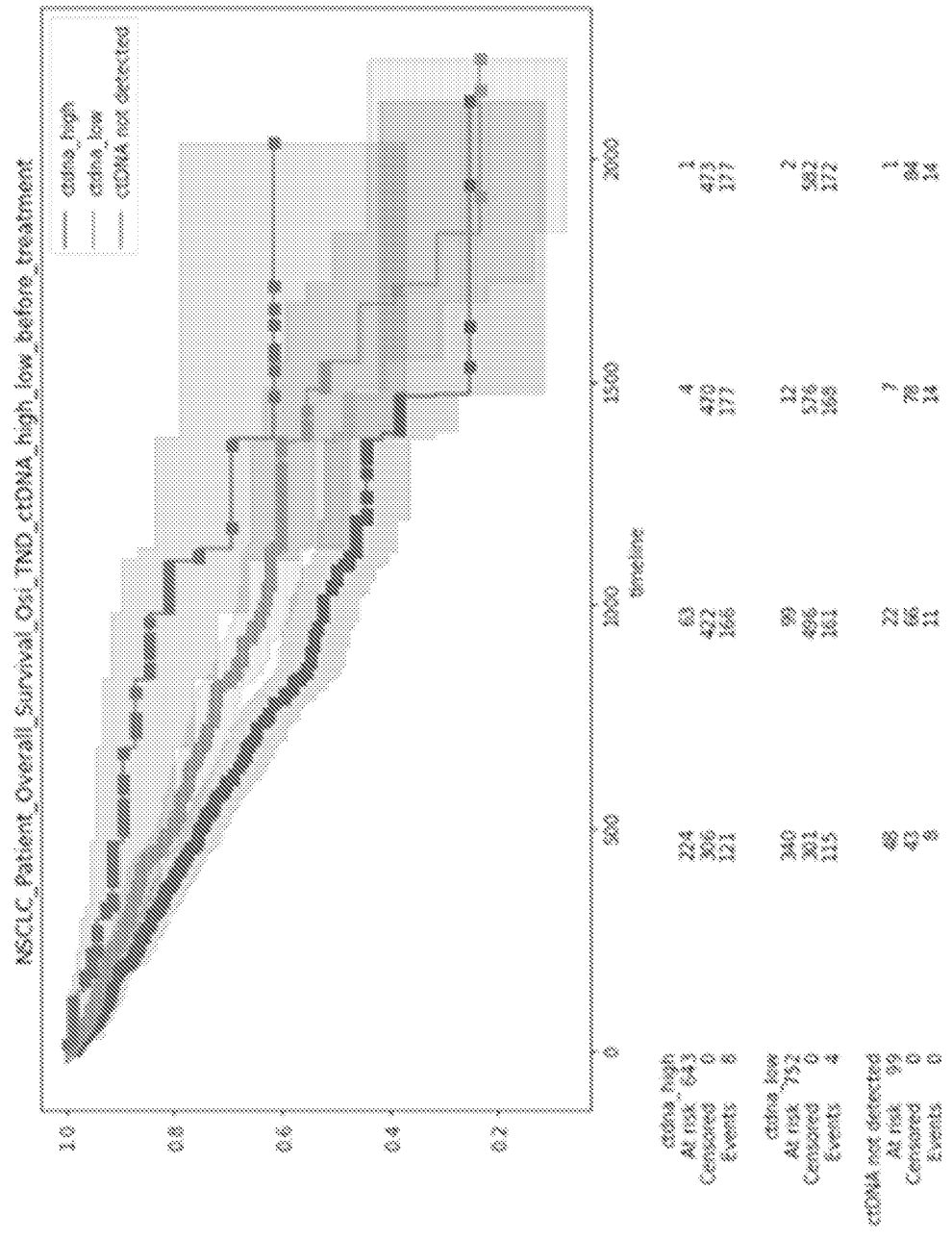

NSCLC_Patient_Overall_Survival_Osi_TKD_ctDNA_high_low_before_treatment
ctdna_high
ctdna_low
ctDNA not detected
1.0
0.8
0.6
0.4
0.2
0
500
1000
1500
2000
timeline
ctdna_high
At risk 643
Censored 0
Events 6
224
366
121
53
422
166
4
478
177
1
493
177
ctdna_low
At risk 762
Censored 0
Events 4
340
361
115
99
456
183
12
526
188
2
582
172
ctDNA not detected
At risk 88
Censored 0
Events 0
48
40
8
12
66
11
7
76
14
1
86
14

Figure 12

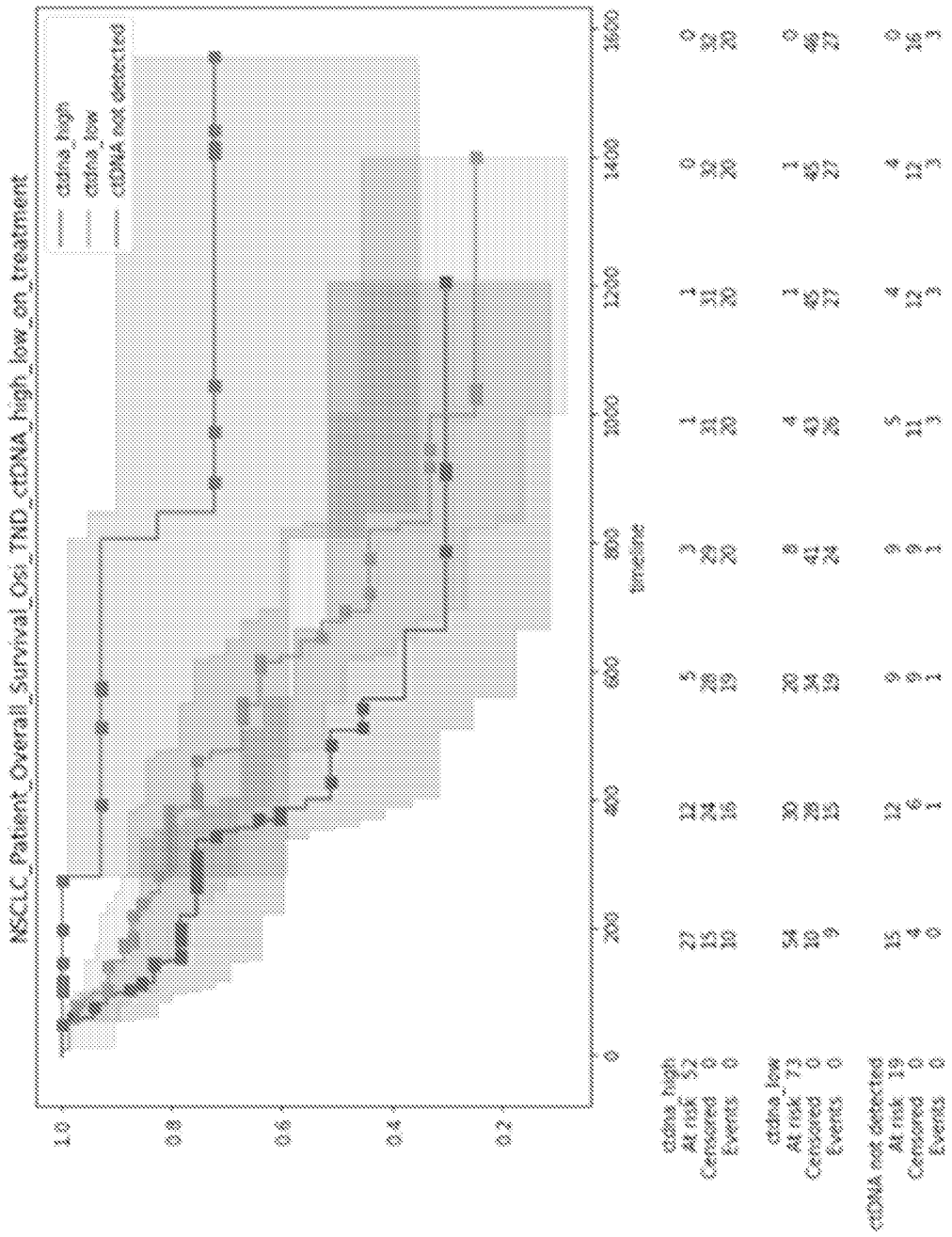

NSCLC_Patient_Overall_Survival_Osi_TND_ctDNA_high_low_on_treatment
ctDNA_high
ctDNA_low
ctDNA not detected
1.0
0.8
0.6
0.4
0.2
0
200
400
600
800
1000
1200
1400
1600
Months
ctDNA_high
At risk 52 22 18 7 2 1 0 0 0
Censored 0 0 0 0 0 0 0 0 0
Events 30 28 28 21 8 6 1 0 0
ctDNA_low
At risk 72 33 23 9 4 1 1 0 0
Censored 0 0 0 0 0 0 0 0 0
Events 39 32 22 9 4 1 0 0 0
ctDNA not detected
At risk 28 15 9 4 2 1 0 0 0
Censored 0 0 0 0 0 0 0 0 0
Events 13 6 5 3 2 0 1 0 0

Figure 13

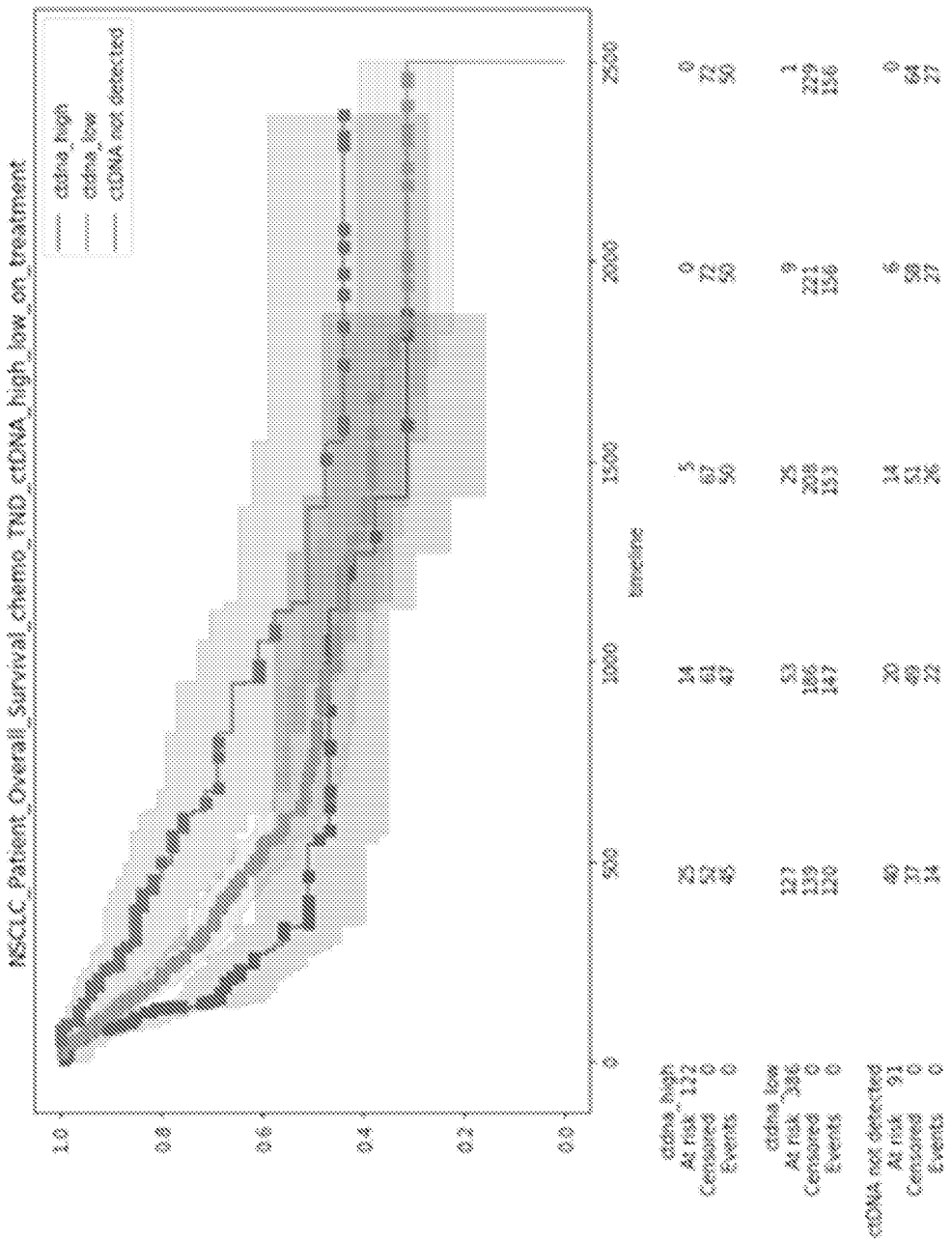

NSCLC_Patient_Overall_Survival_chemo_TND_ctDNA_high_low_on_treatment
ctDna_high
ctDna_low
ctDNA not detected
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
timeline
ctDna_high
At risk 212 0 0
Censored 25 34 0 0
Events 22 61 72 72
ctDna_low
At risk 286 0 0
Censored 137 50 1
Events 130 186 221 221
ctDNA not detected
At risk 91 0 0
Censored 49 28 6 6
Events 34 63 84 84

Figure 14

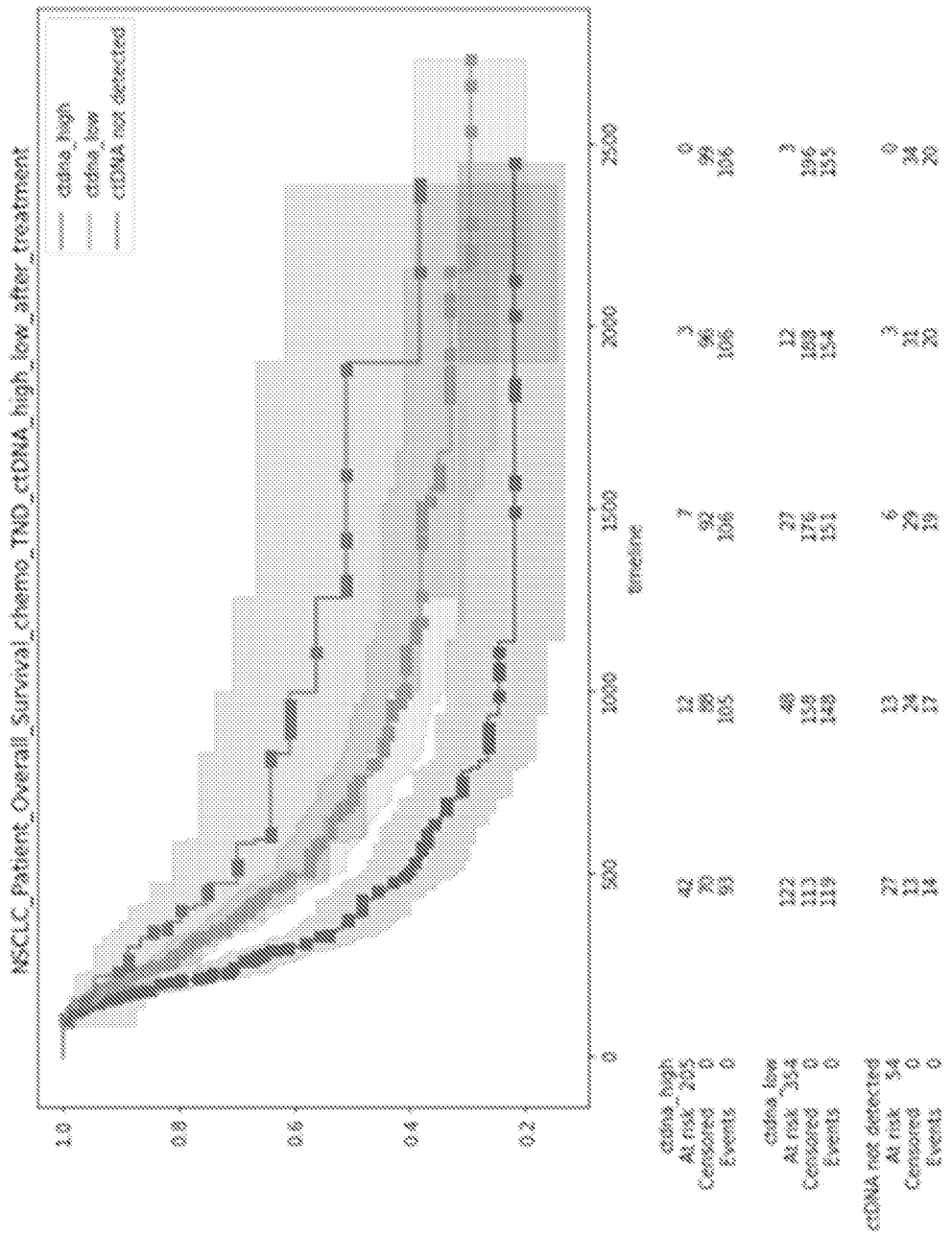

NSCLC_Patient_Overall_Survival_chemo_TND_ctDNA_high_low_after_treatment
ctdna_high
ctdna_low
ctDNA not detected
1.0
0.8
0.6
0.4
0.2
0
500
1000
1500
2000
2500
timeline
ctdna_high
At risk 42 28 13 7 3 0
Censored 0 0 0 0 0 0
Events 0 0 0 0 0 0
ctdna_low
At risk 122 80 48 27 12 0
Censored 0 0 0 0 0 0
Events 0 0 0 0 0 0
ctDNA not detected
At risk 22 13 3 2 0 0
Censored 0 0 0 0 0 0
Events 0 0 0 0 0 0

Figure 15

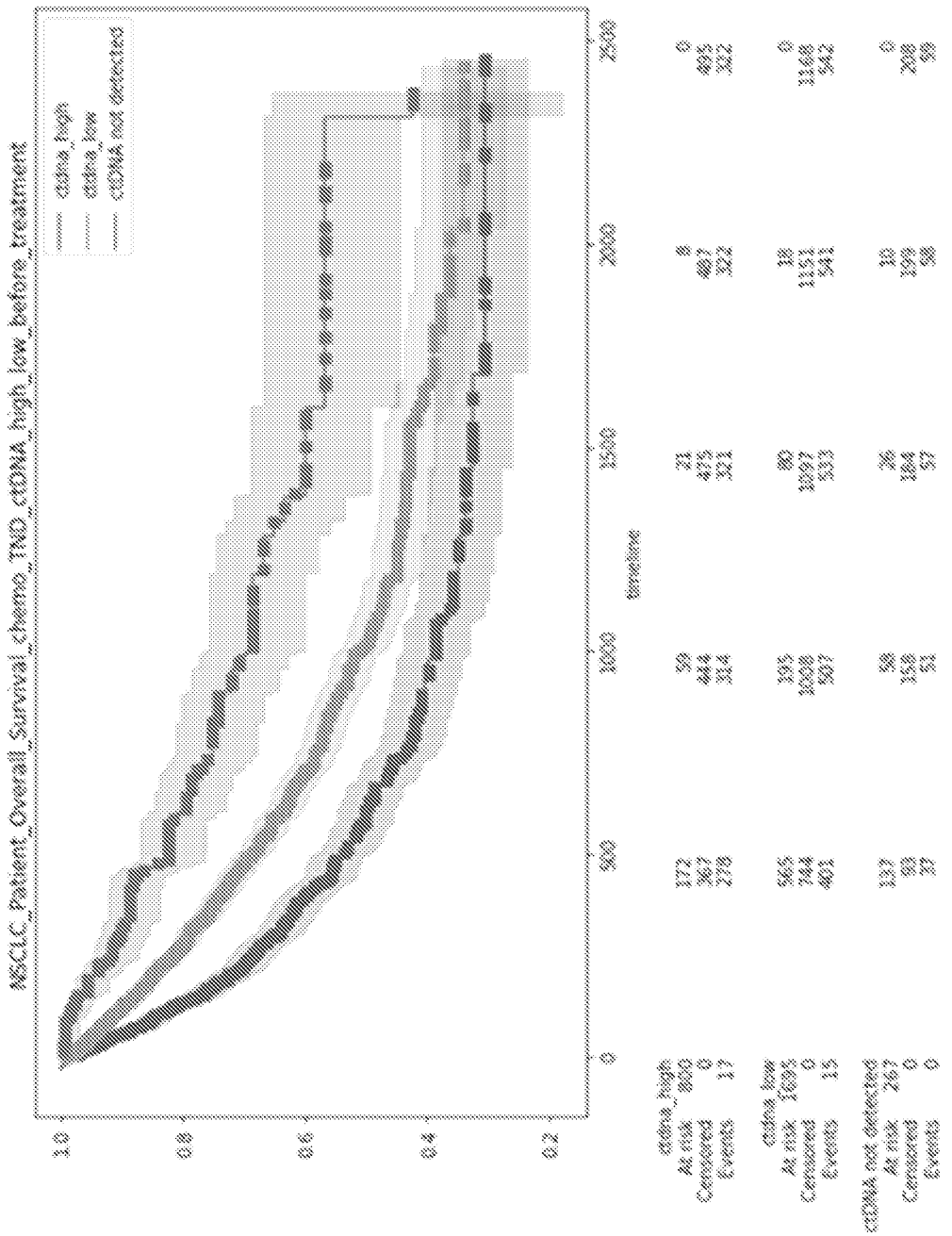

NSCLC_Patient_Overall_Survival_chemo_TWO_ctDNA_high_low_before_treatment
ctDNA_high
ctDNA_low
ctDNA not detected
1.0
0.8
0.6
0.4
0.2
0
500
1000
1500
2000
2500
timeline
ctDNA_high
At risk 800 172 59 21 8 0
Censored 0 367 444 475 497 495
Events 17 278 314 321 322 322
ctDNA_low
At risk 1091 565 195 80 18 6
Censored 0 744 1008 1092 1151 1148
Events 15 491 587 633 641 642
ctDNA not detected
At risk 267 137 58 36 13 0
Censored 0 88 159 184 196 208
Events 0 37 53 67 68 68

Figure 16

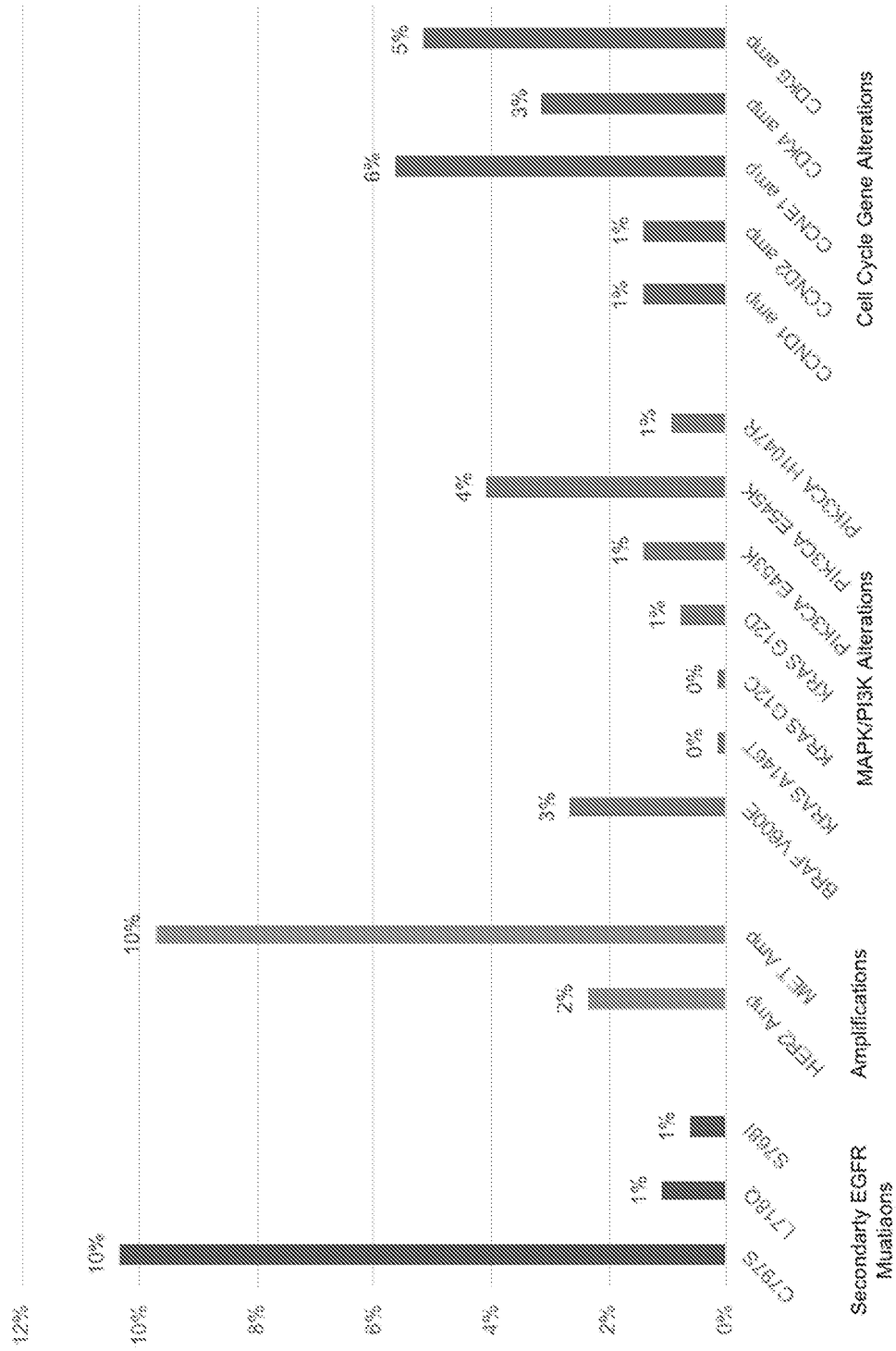

12%
10%
8%
6%
4%
2%
0%
C797S
L718Q
S768I
EGFR amp
MET amp
HER2 amp
BRAF V600E
KRAS A146T
KRAS G12C
KRAS G12D
PIK3CA E545K
PIK3CA E542K
PIK3CA H1047R
CCND1 amp
CCND2 amp
CCNE1 amp
CDK4 amp
CDK6 amp
10%
1%
1%
2%
10%
3%
0%
0%
1%
1%
4%
1%
1%
1%
6%
3%
5%
Secondary EGFR Mutaions
Amplifications
MAPK/PI3K Alterations
Cell Cycle Gene Alterations

Figure 17

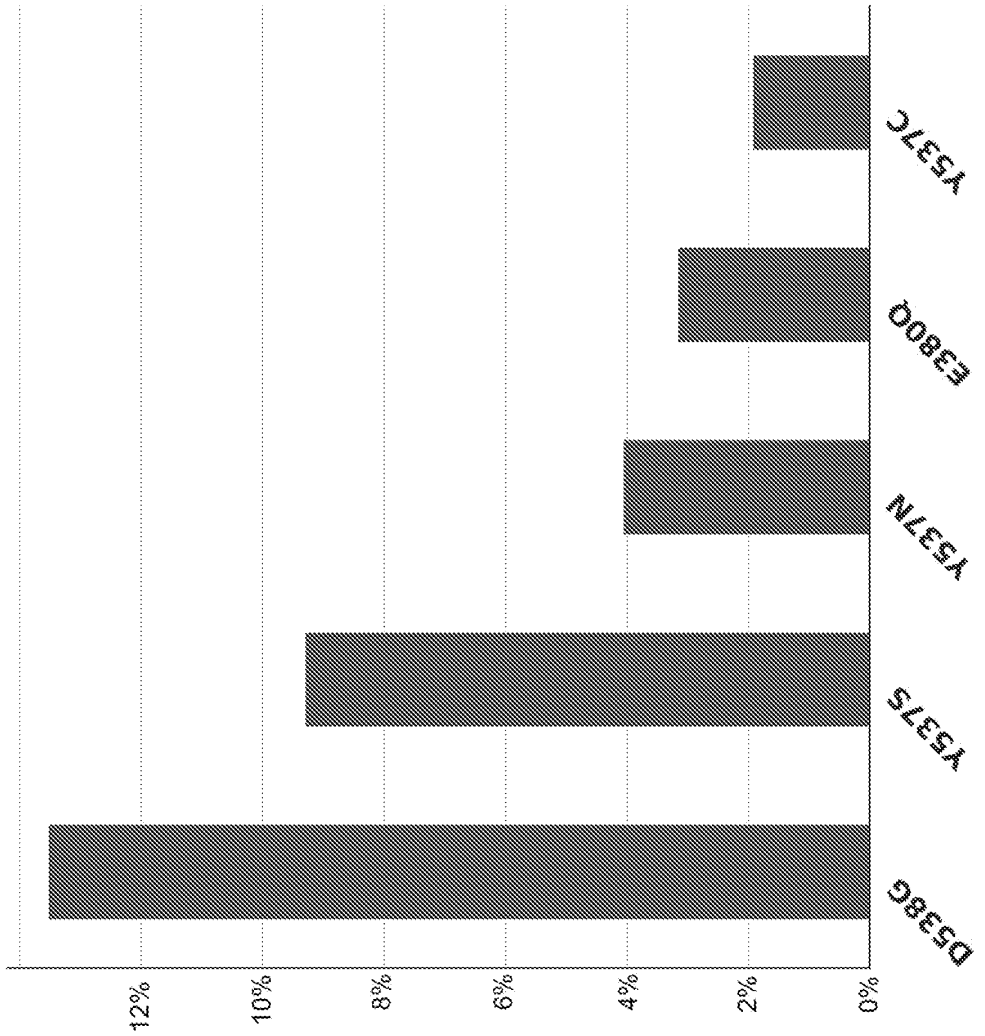



Figure 18

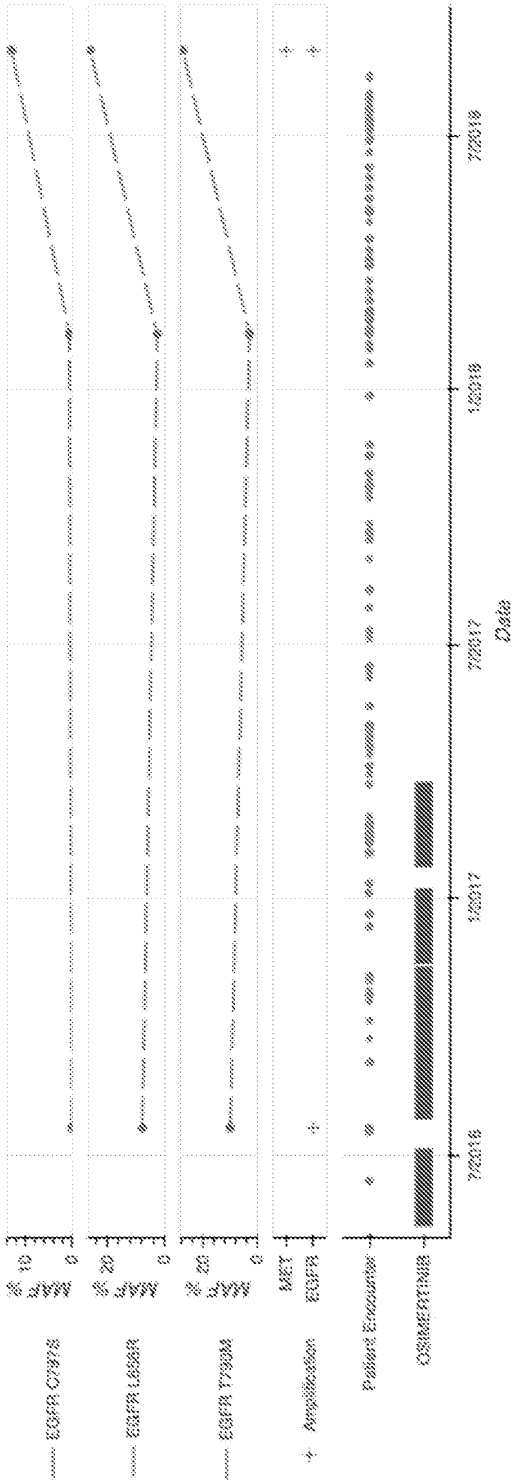

EGFR C797S
EGFR L858R
EGFR T790M
Amplification
MET
EGFR
Patient Encounter
OSIMERTINIB
7/2018
1/2017
7/2017
1/2018
7/2018
Date

Figure 19

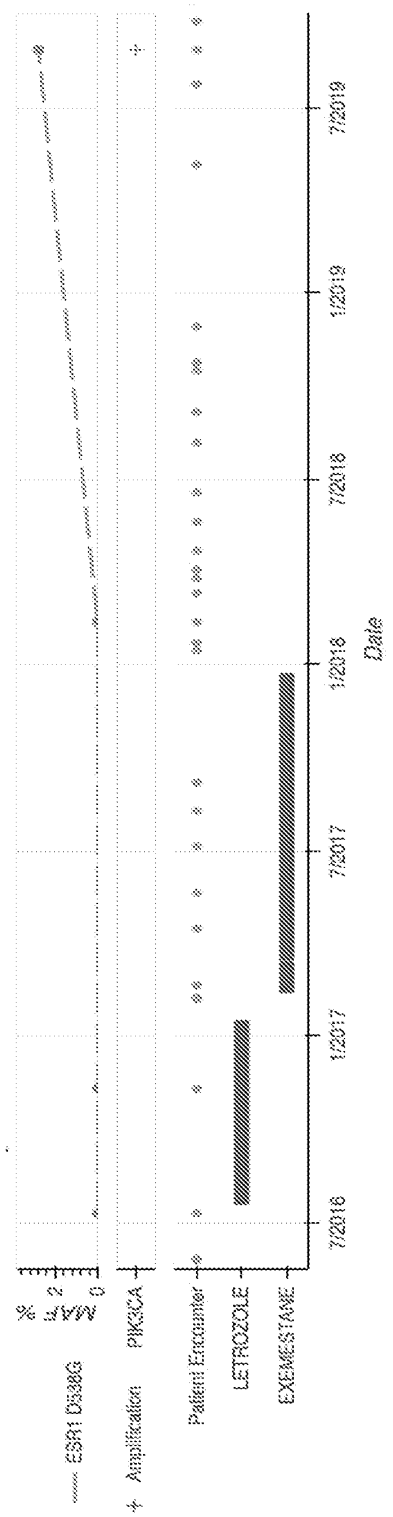

%
4
2
0
ESR1 D538G
Amplification
PIK3CA
Patient Encounter
LETROZOLE
EXEMESTANE
7/2016
1/2017
7/2017
1/2018
7/2018
1/2019
7/2019
Date

Figure 20