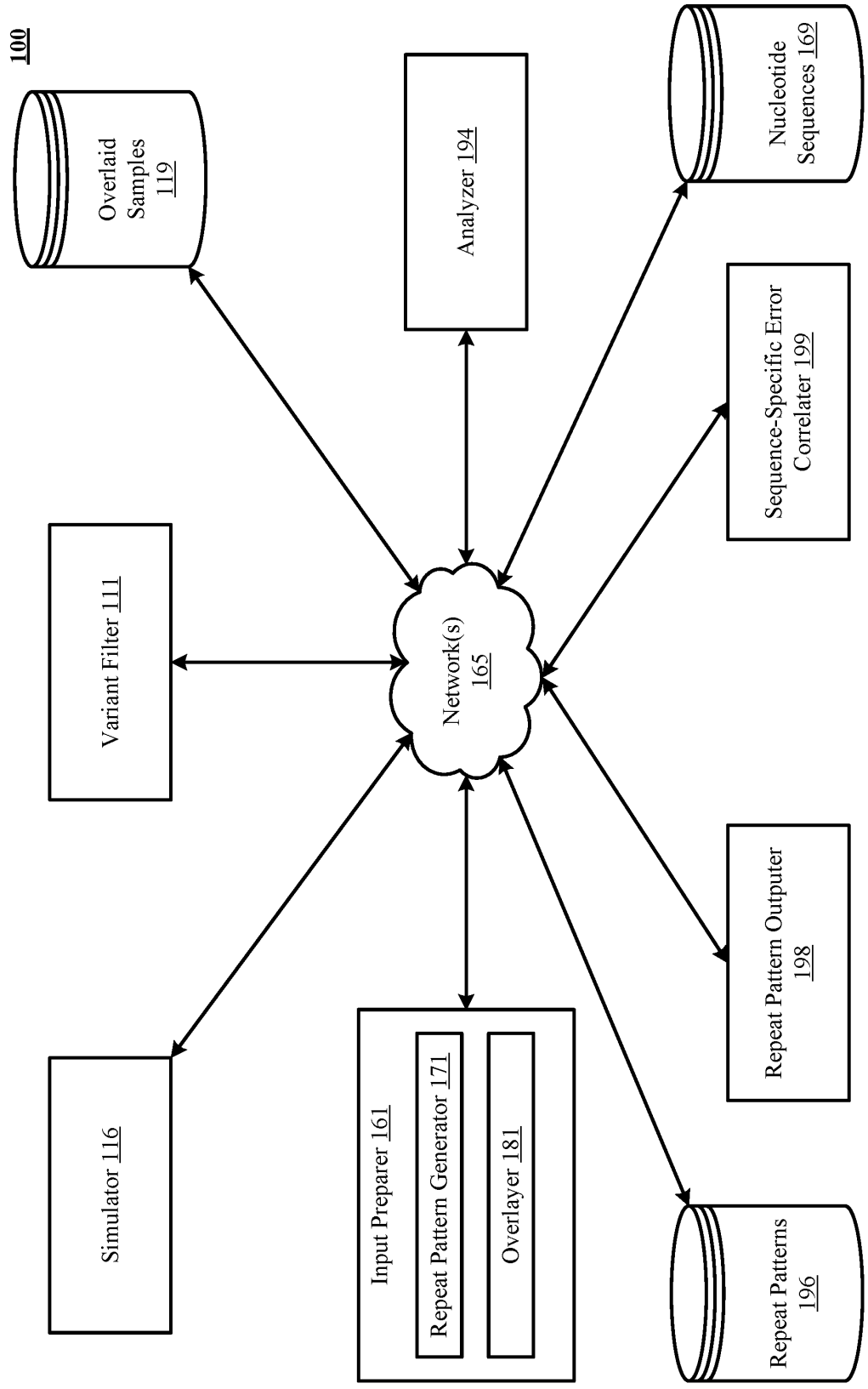

100
Overlaid Samples 119
Analyzer 194
Nucleotide Sequences 169
Variant Filter 111
Network(s) 165
Sequence-Specific Error Correlater 199
Simulator 116
Input Preparer 161
Repeat Pattern Generator 171
Overlayer 181
Repeat Pattern Outputer 198
Repeat Patterns 196

FIG. 1

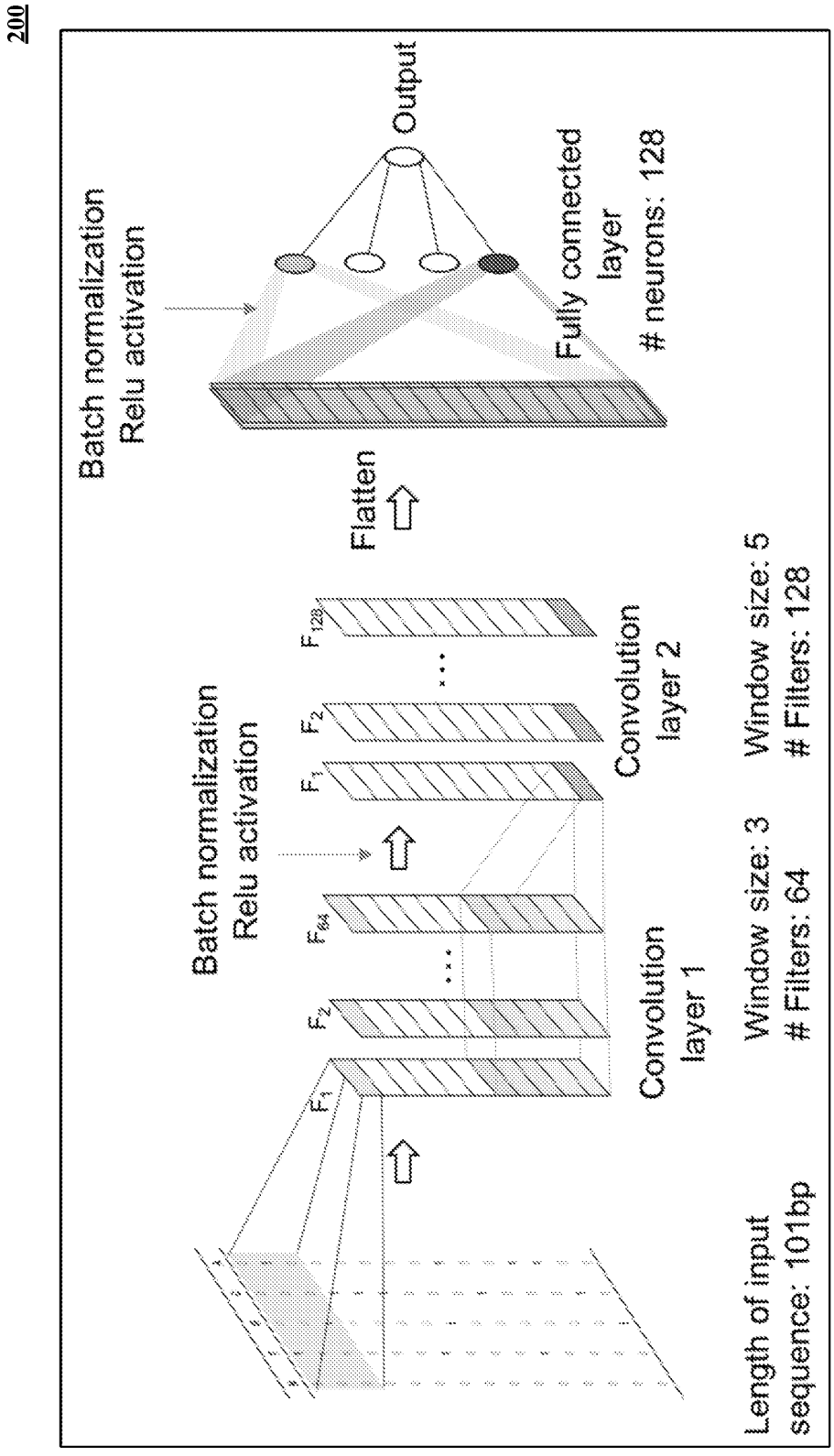

200
Length of input sequence: 101bp
Convolution layer 1
Window size: 3
Filters: 64
Batch normalization
Relu activation
Convolution layer 2
Window size: 5
Filters: 128
Flatten
Fully connected layer
neurons: 128
Output

FIG. 2

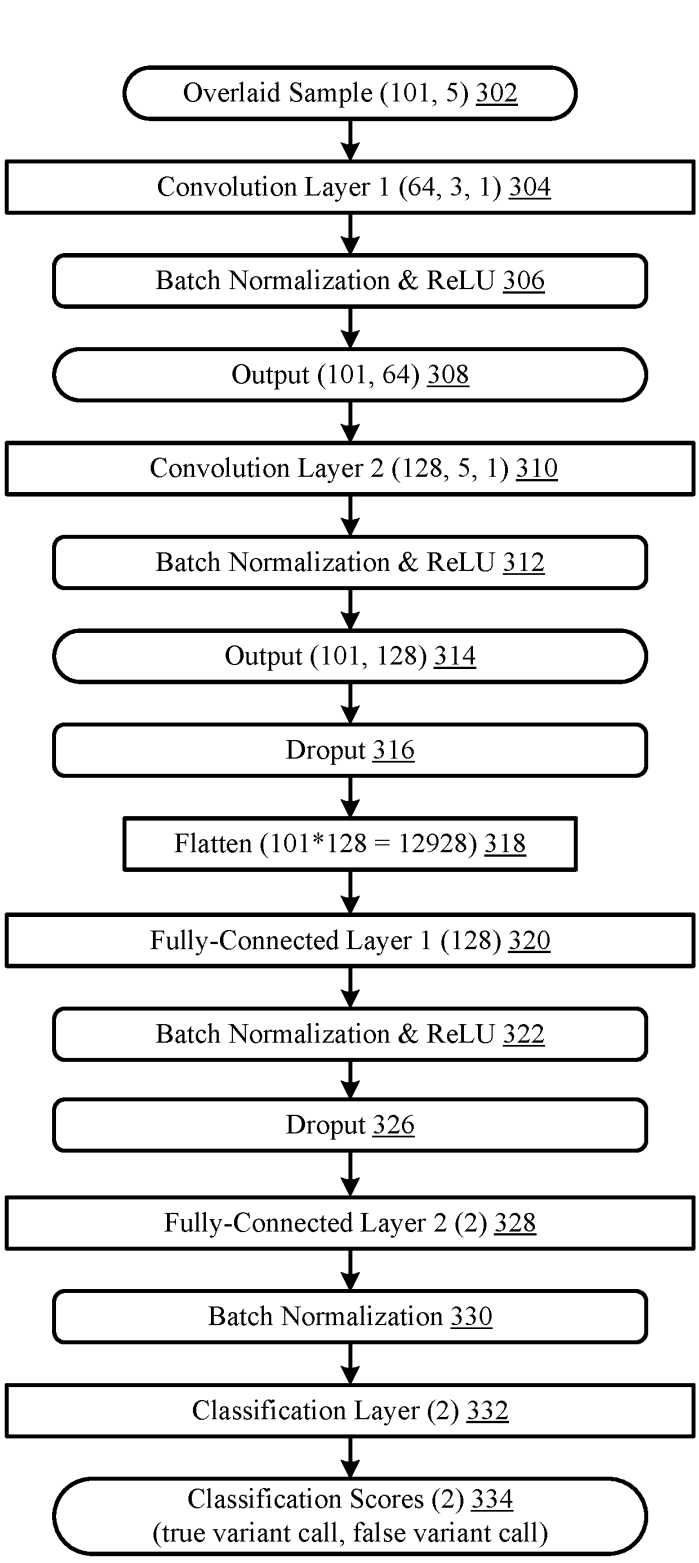

300
Overlaid Sample (101, 5) 302
Convolution Layer 1 (64, 3, 1) 304
Batch Normalization & ReLU 306
Output (101, 64) 308
Convolution Layer 2 (128, 5, 1) 310
Batch Normalization & ReLU 312
Output (101, 128) 314
Droput 316
Flatten (101*128 = 12928) 318
Fully-Connected Layer 1 (128) 320
Batch Normalization & ReLU 322
Droput 326
Fully-Connected Layer 2 (2) 328
Batch Normalization 330
Classification Layer (2) 332
Classification Scores (2) 334
(true variant call, false variant call)

FIG. 3

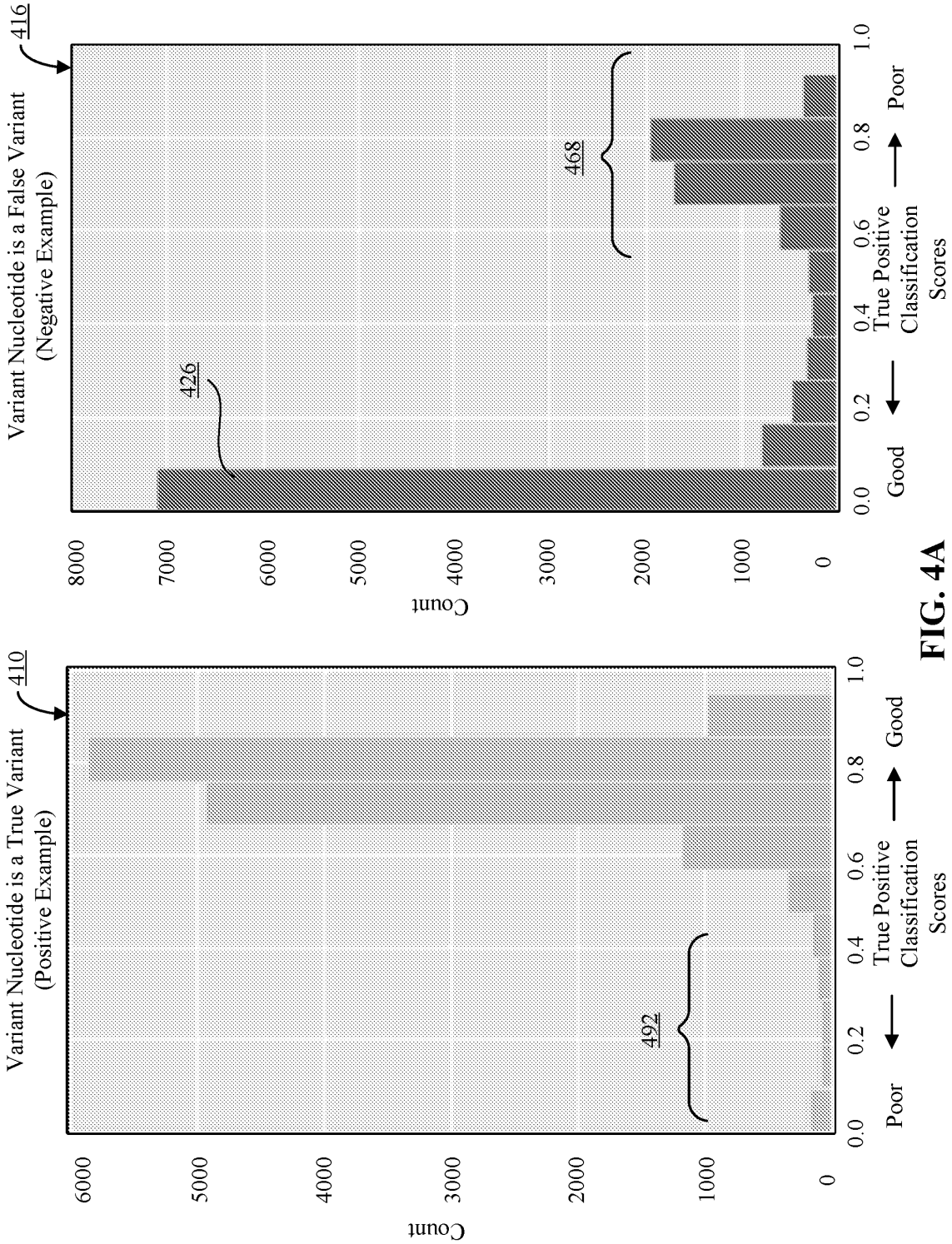

Variant Nucleotide is a False Variant
(Negative Example)
416
426
468
Count
8000
7000
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
Good
True Positive
Classification
Scores
Poor
Variant Nucleotide is a True Variant
(Positive Example)
410
492
Count
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
Poor
True Positive
Classification
Scores
Good

FIG. 4A

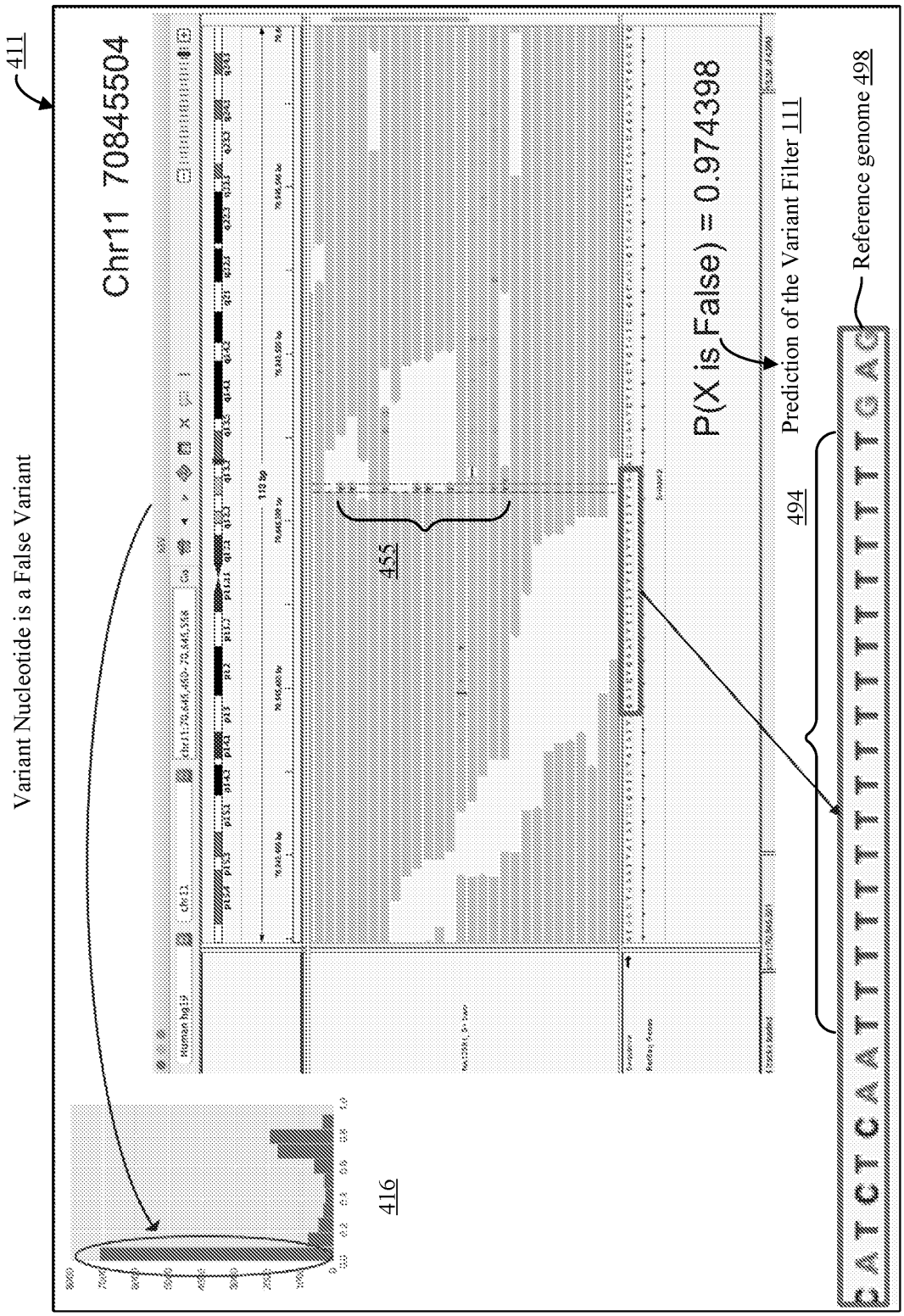

411
Variant Nucleotide is a False Variant
Chr11 70845504
455
P(X is False) = 0.974398
Prediction of the Variant Filter 111
494
Reference genome 498
CATCTCAATTTTTTTTTTTTTTTTGAG
416

FIG. 4B

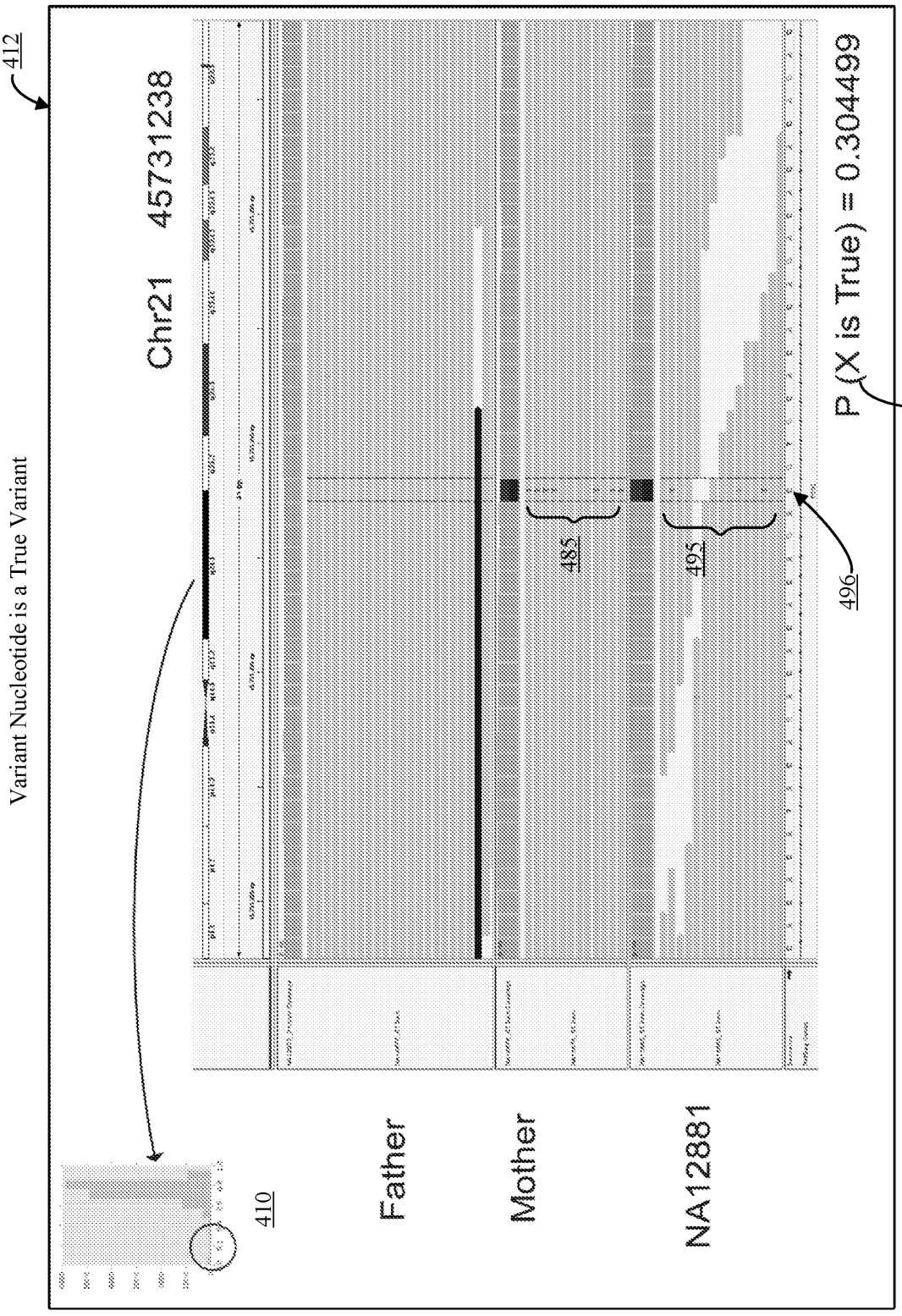

412
Variant Nucleotide is a True Variant
Chr21 45731238
P (X is True) = 0.304499
Prediction of the Variant Filter 111
410
Father
Mother
NA12881
485
495
496

FIG. 4C

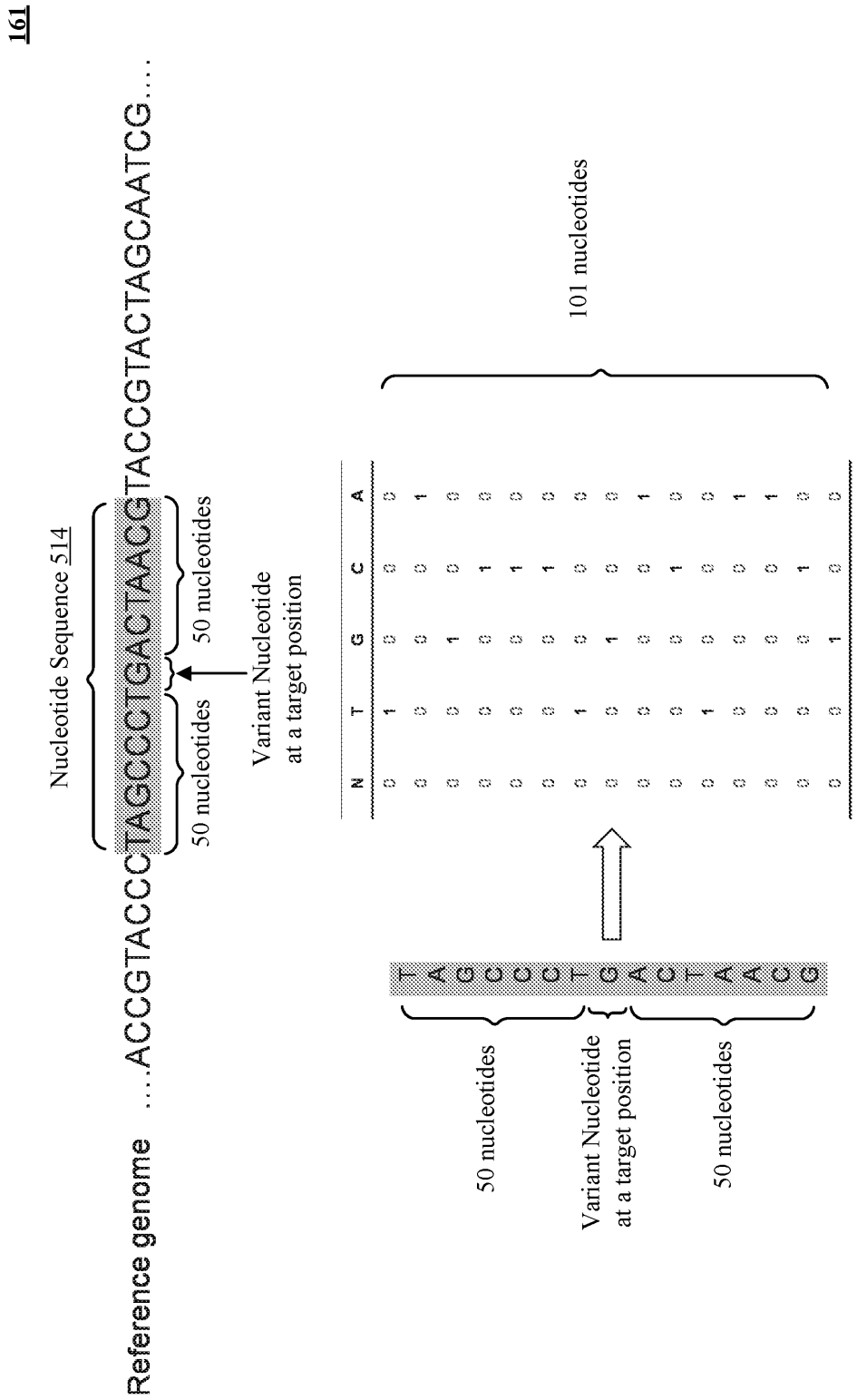

161
Nucleotide Sequence 514
Reference genome ....ACCGTACCCTAGCCCTGACTAACGTACCGTACTAGCAATCG....
50 nucleotides
50 nucleotides
Variant Nucleotide at a target position
101 nucleotides
N T G C A
50 nucleotides
Variant Nucleotide at a target position
50 nucleotides

FIG. 5

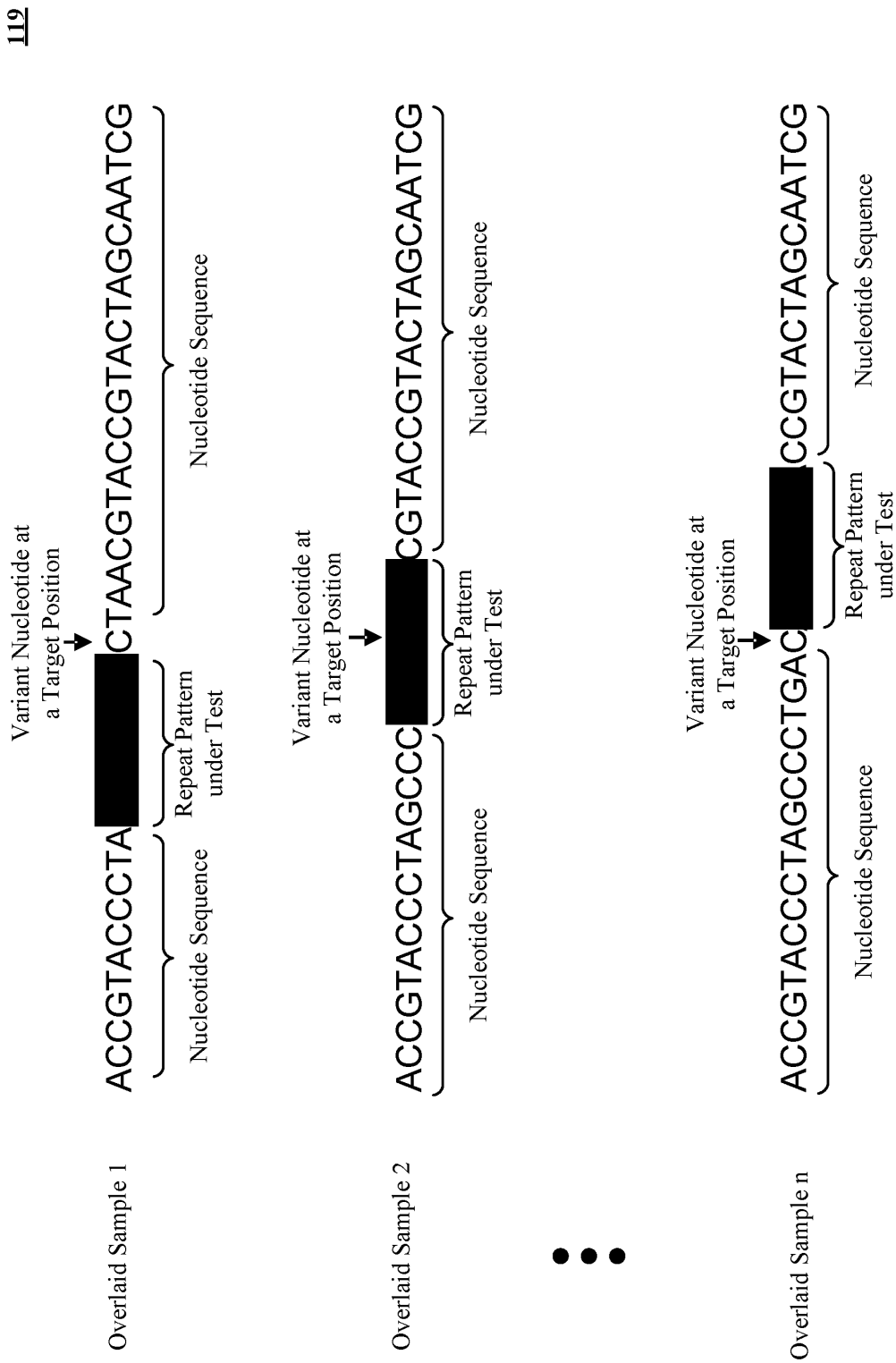

119
Variant Nucleotide at a Target Position
Overlaid Sample 1
ACCGTACCCTA
CTAACGTACCGTACTAGCAATCG
Nucleotide Sequence
Repeat Pattern under Test
Nucleotide Sequence
Variant Nucleotide at a Target Position
Overlaid Sample 2
ACCGTACCCTAGCCC
CGTACCGTACTAGCAATCG
Repeat Pattern under Test
Nucleotide Sequence
Nucleotide Sequence
Variant Nucleotide at a Target Position
Overlaid Sample n
ACCGTACCCTAGCCCTGAC
CCGTACTAGCAATCG
Repeat Pattern under Test
Nucleotide Sequence
Nucleotide Sequence

FIG. 6

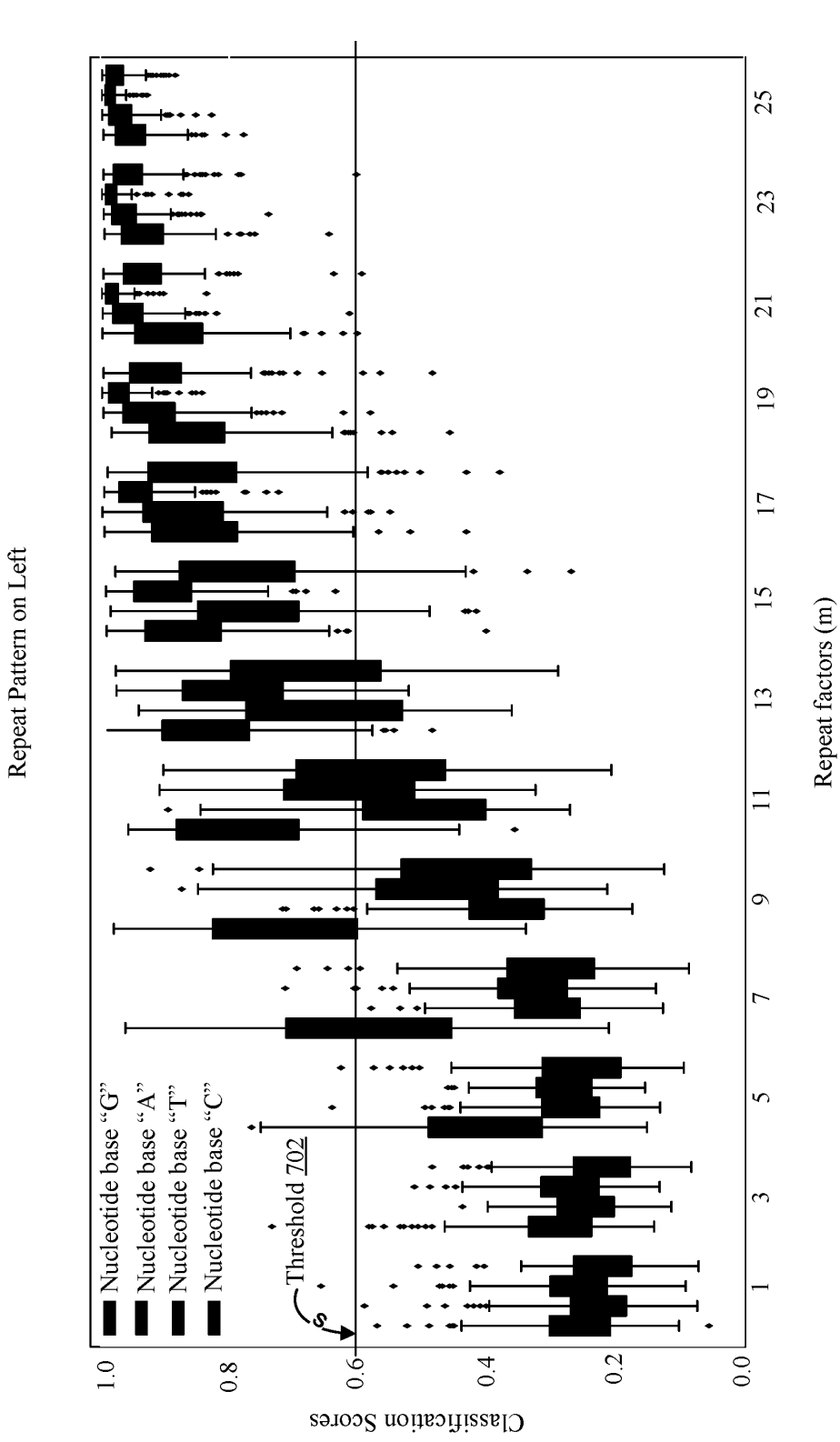

194
Repeat Pattern on Left
Nucleotide base "G"
Nucleotide base "A"
Nucleotide base "T"
Nucleotide base "C"
Threshold 702
s
Classification Scores
1.0
0.8
0.6
0.4
0.2
0.0
1
3
5
7
9
11
13
15
17
19
21
23
25
Repeat factors (m)

FIG. 7A

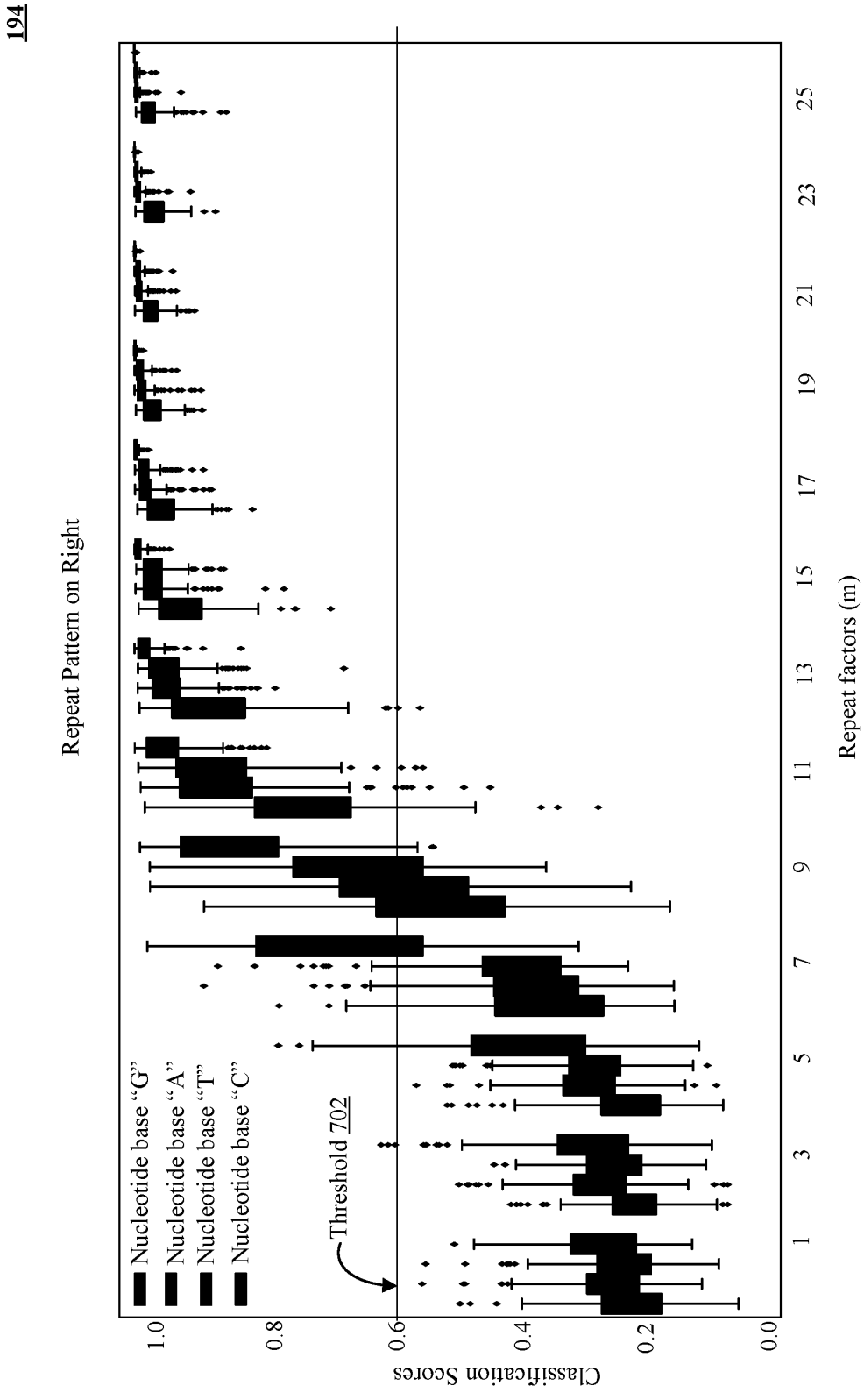

194
Repeat Pattern on Right
Nucleotide base "G"
Nucleotide base "A"
Nucleotide base "T"
Nucleotide base "C"
Threshold 702
Classification Scores
1.0
0.8
0.6
0.4
0.2
0.0
1
3
5
7
9
11
13
15
17
19
21
23
25
Repeat factors (m)

FIG. 7B

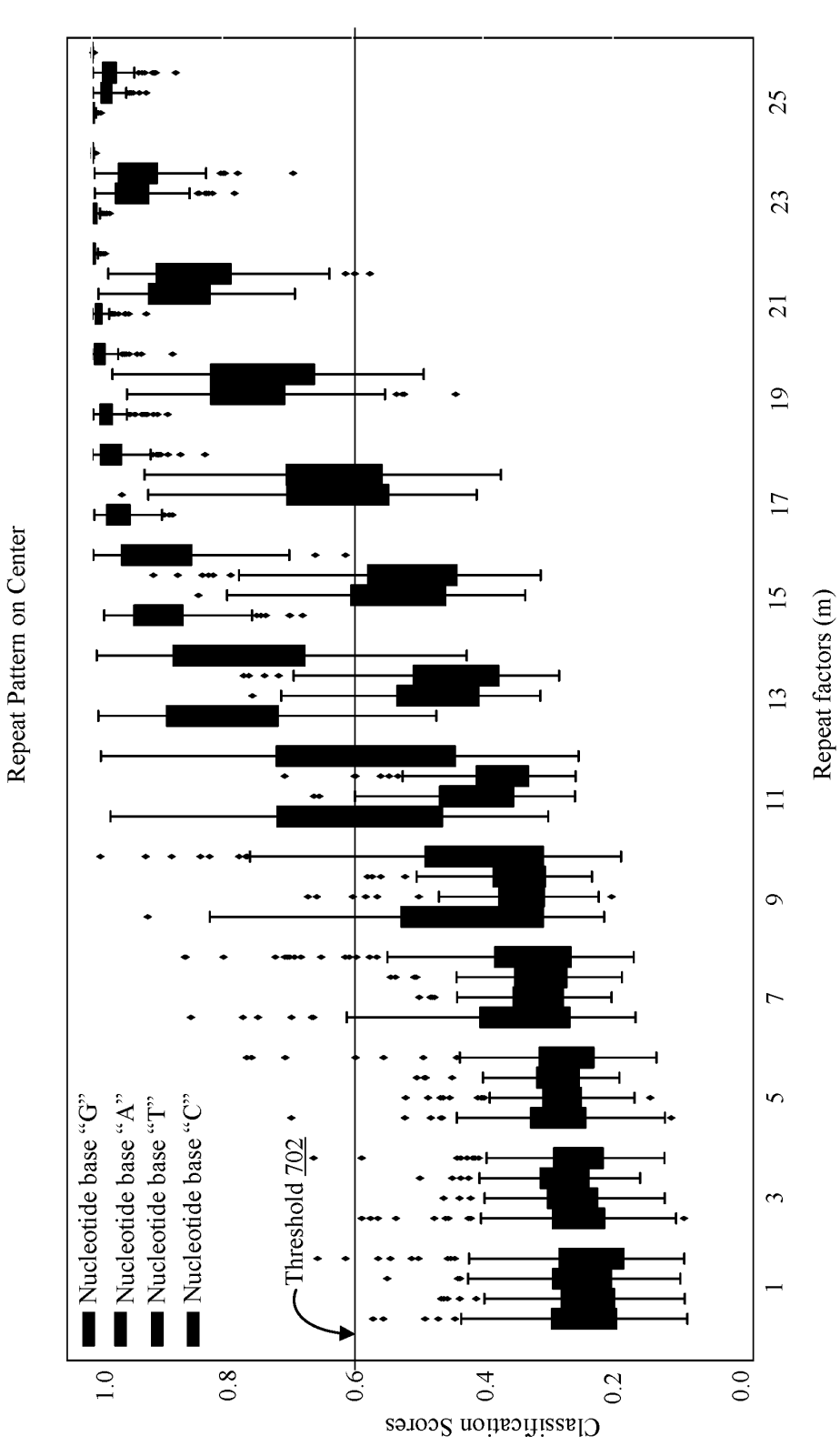

194
Repeat Pattern on Center
Nucleotide base "G"
Nucleotide base "A"
Nucleotide base "T"
Nucleotide base "C"
Threshold 702
Classification Scores
1.0
0.8
0.6
0.4
0.2
0.0
1
3
5
7
9
11
13
15
17
19
21
23
25
Repeat factors (m)

FIG. 7C

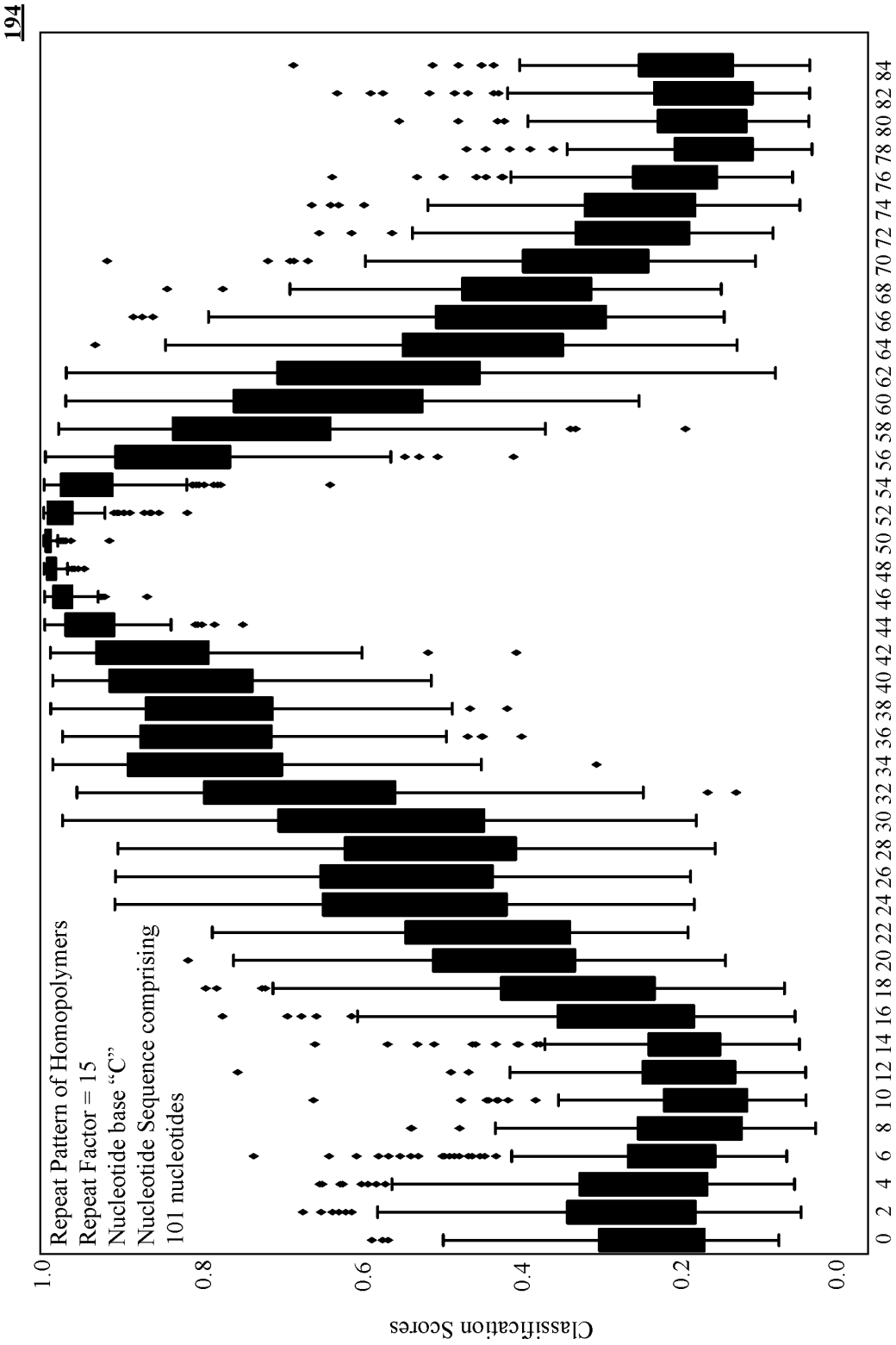

194
Repeat Pattern of Homopolymers
Repeat Factor = 15
Nucleotide base "C"
Nucleotide Sequence comprising
101 nucleotides
Classification Scores
1.0
0.8
0.6
0.4
0.2
0.0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80 82 84
Offset position

FIG. 8A

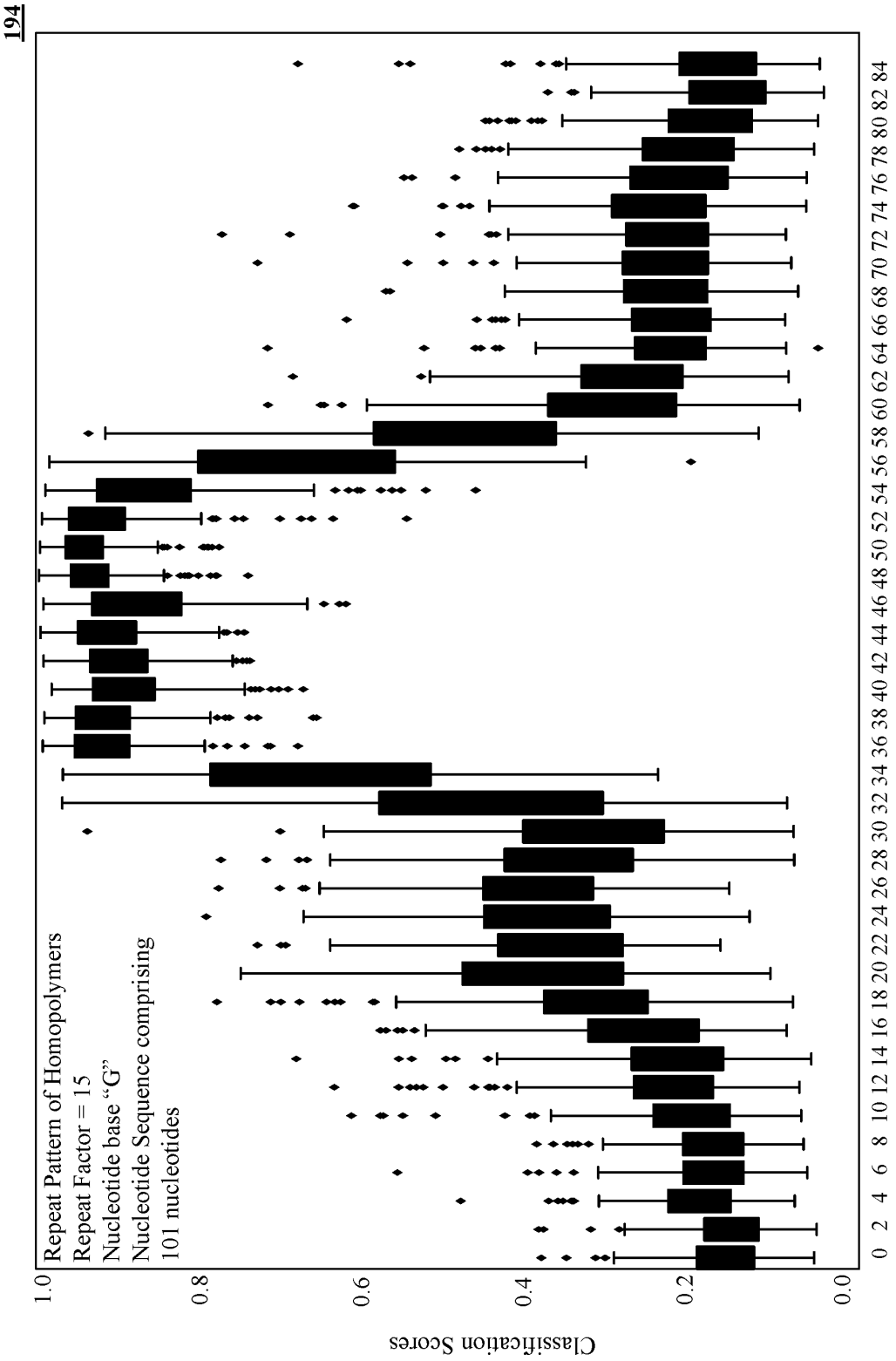

194
Repeat Pattern of Homopolymers
Repeat Factor = 15
Nucleotide base "G"
Nucleotide Sequence comprising
101 nucleotides
Classification Scores
1.0
0.8
0.6
0.4
0.2
0.0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80 82 84
Offset position

FIG. 8B

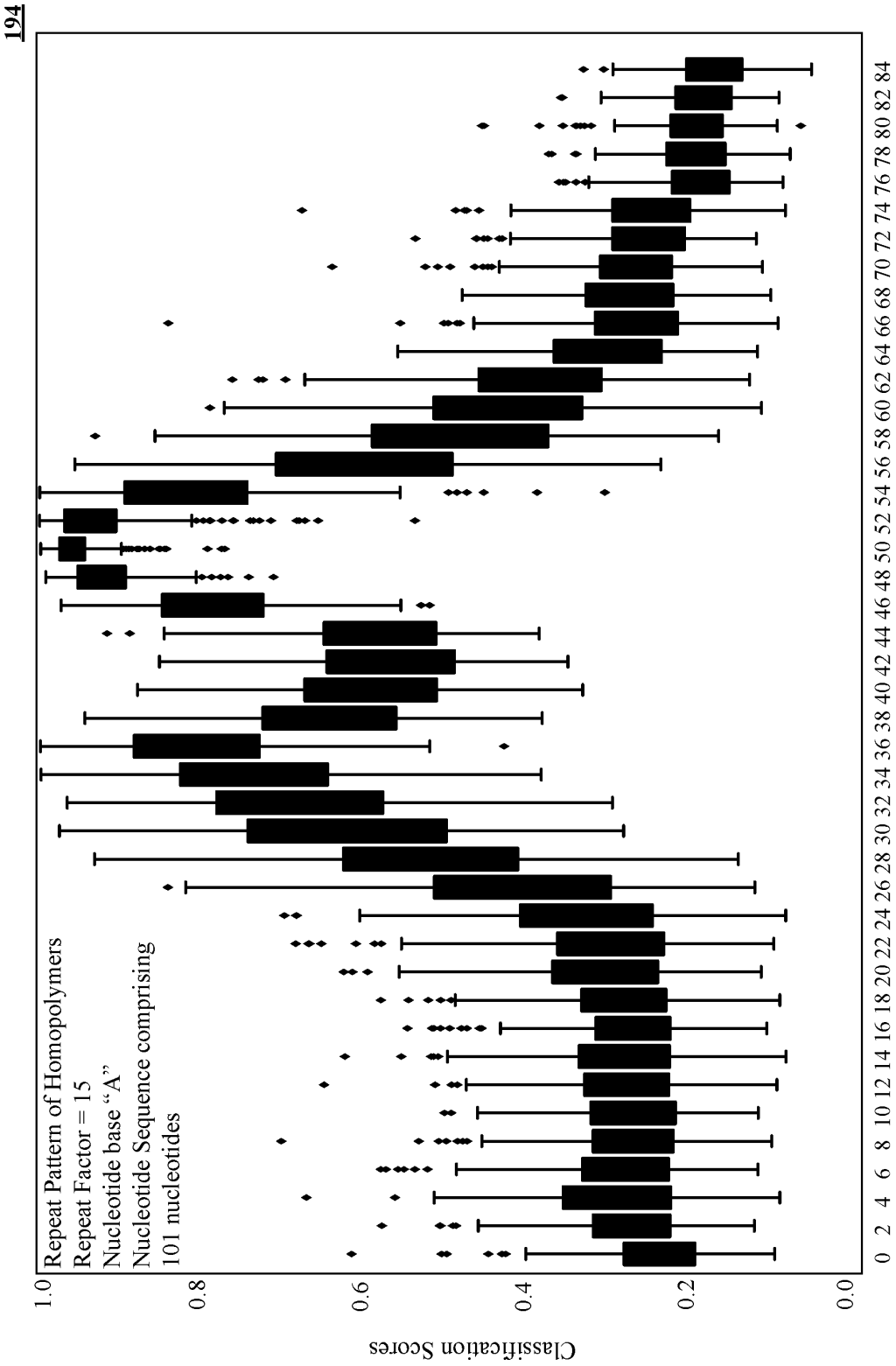

194
Repeat Pattern of Homopolymers
Repeat Factor = 15
Nucleotide base "A"
Nucleotide Sequence comprising
101 nucleotides
Classification Scores
1.0
0.8
0.6
0.4
0.2
0.0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80 82 84
Offset position

FIG. 8C

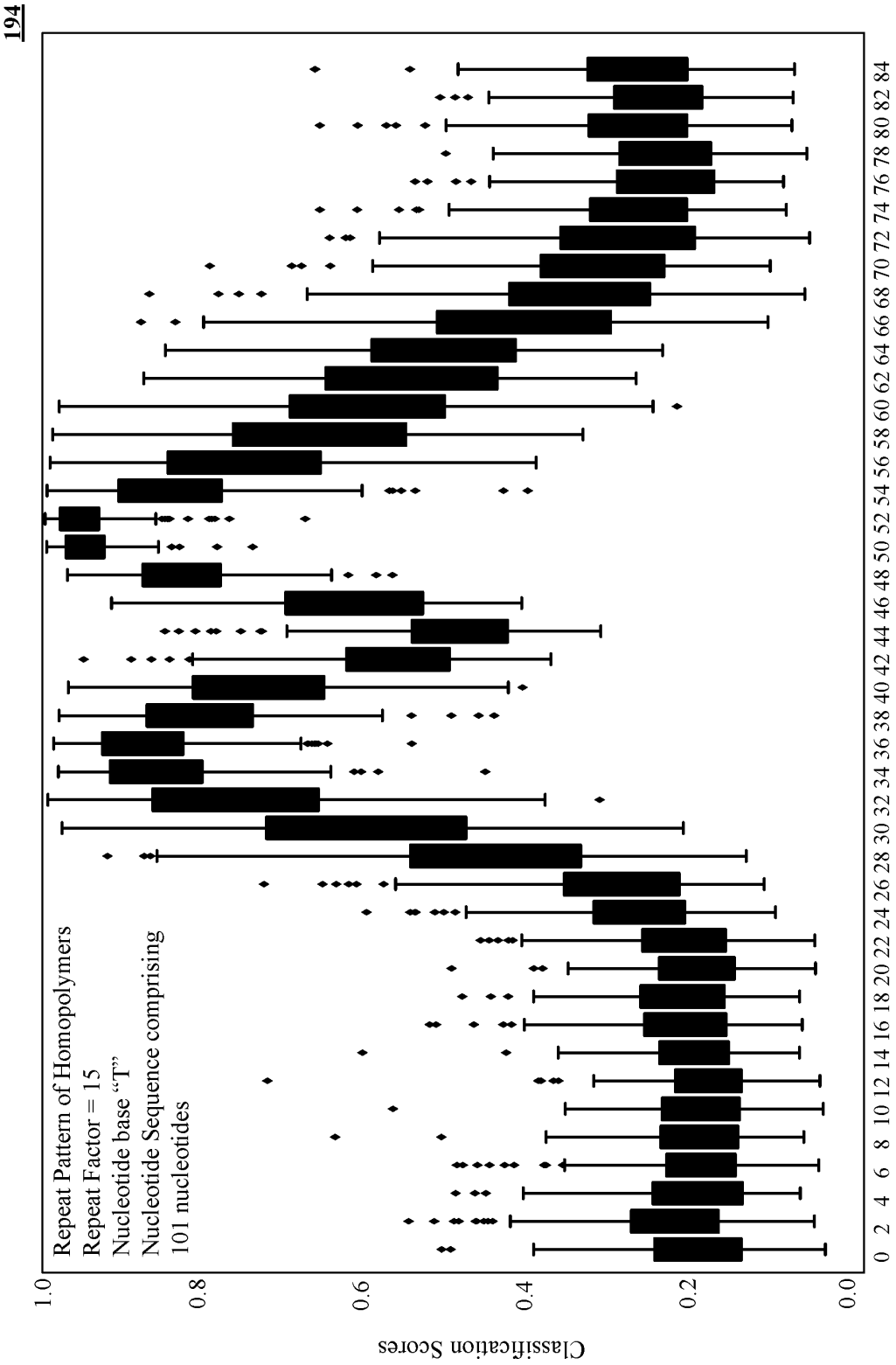

194
Repeat Pattern of Homopolymers
Repeat Factor = 15
Nucleotide base "T"
Nucleotide Sequence comprising
101 nucleotides
Classification Scores
1.0
0.8
0.6
0.4
0.2
0.0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80 82 84
Offset position

FIG. 8D

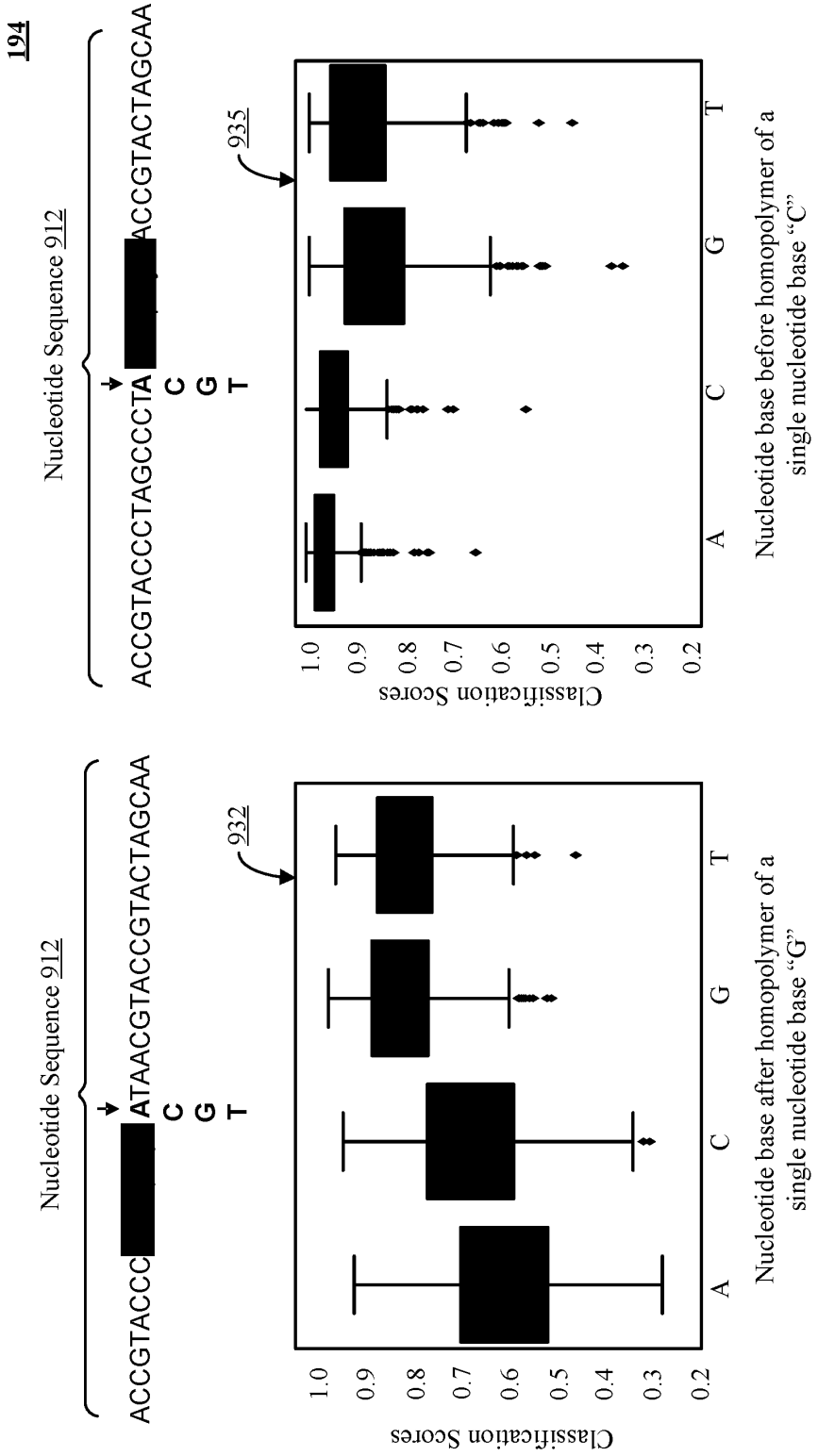

194
Nucleotide Sequence 912
ACCGTACCC
ATAACGTACCGTACTAGCAA
C
G
T
932
Classification Scores
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
A
C
G
T
Nucleotide base after homopolymer of a single nucleotide base "G"
Nucleotide Sequence 912
ACCGTACCCTAGCCCTA
ACCGTACTAGCAA
C
G
T
935
Classification Scores
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
A
C
G
T
Nucleotide base before homopolymer of a single nucleotide base "C"

FIG. 9

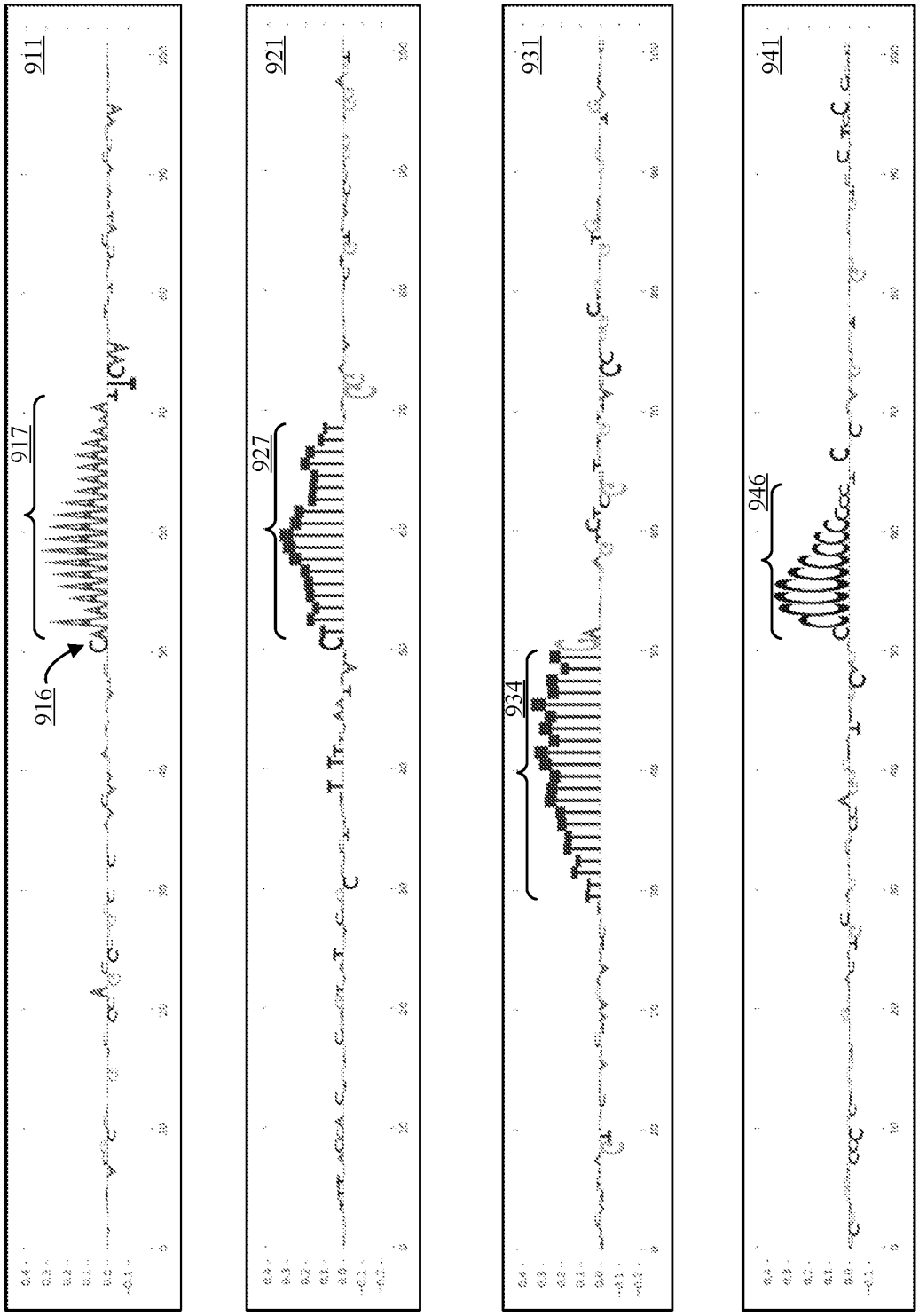


FIG. 10A

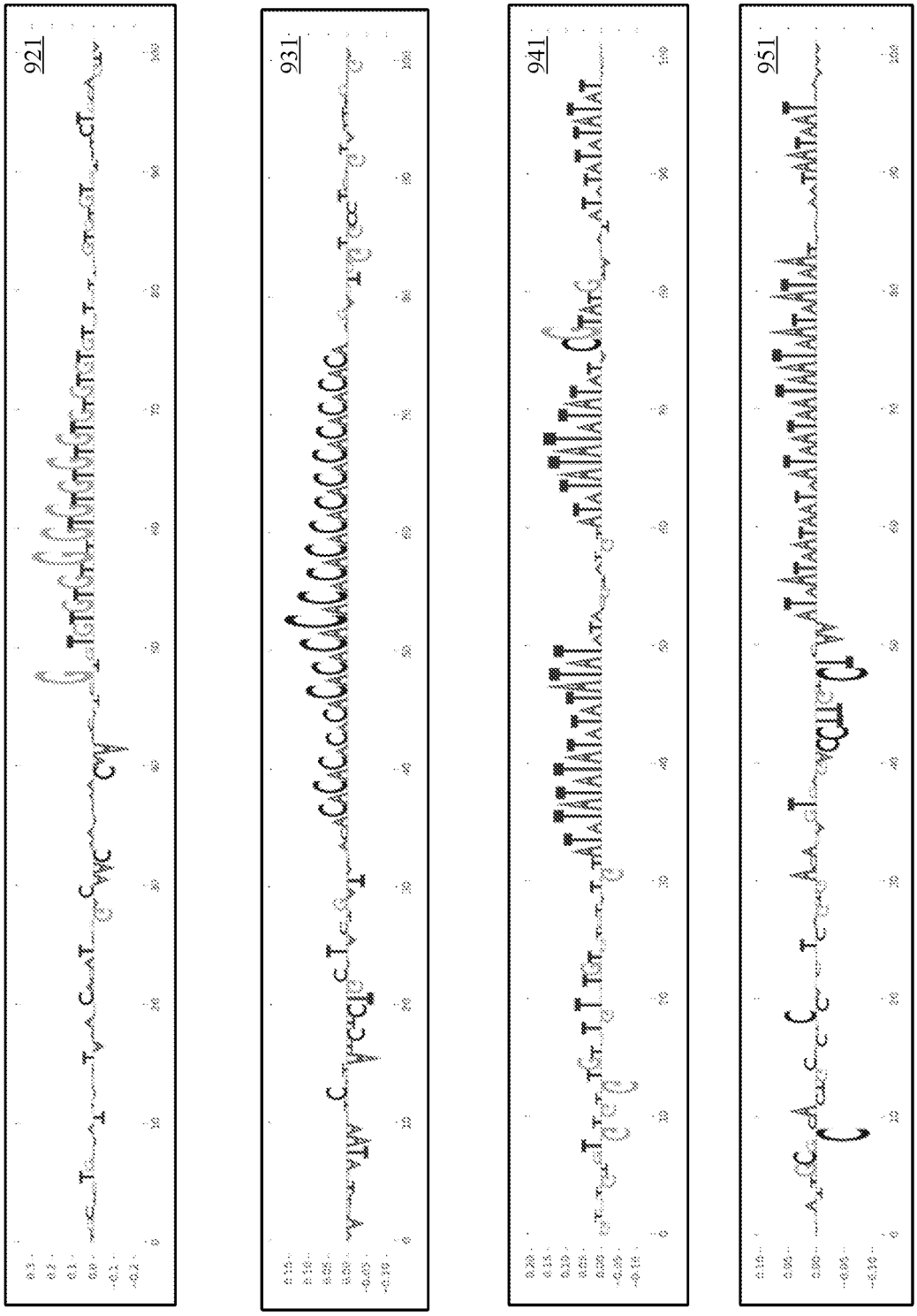


FIG. 10B

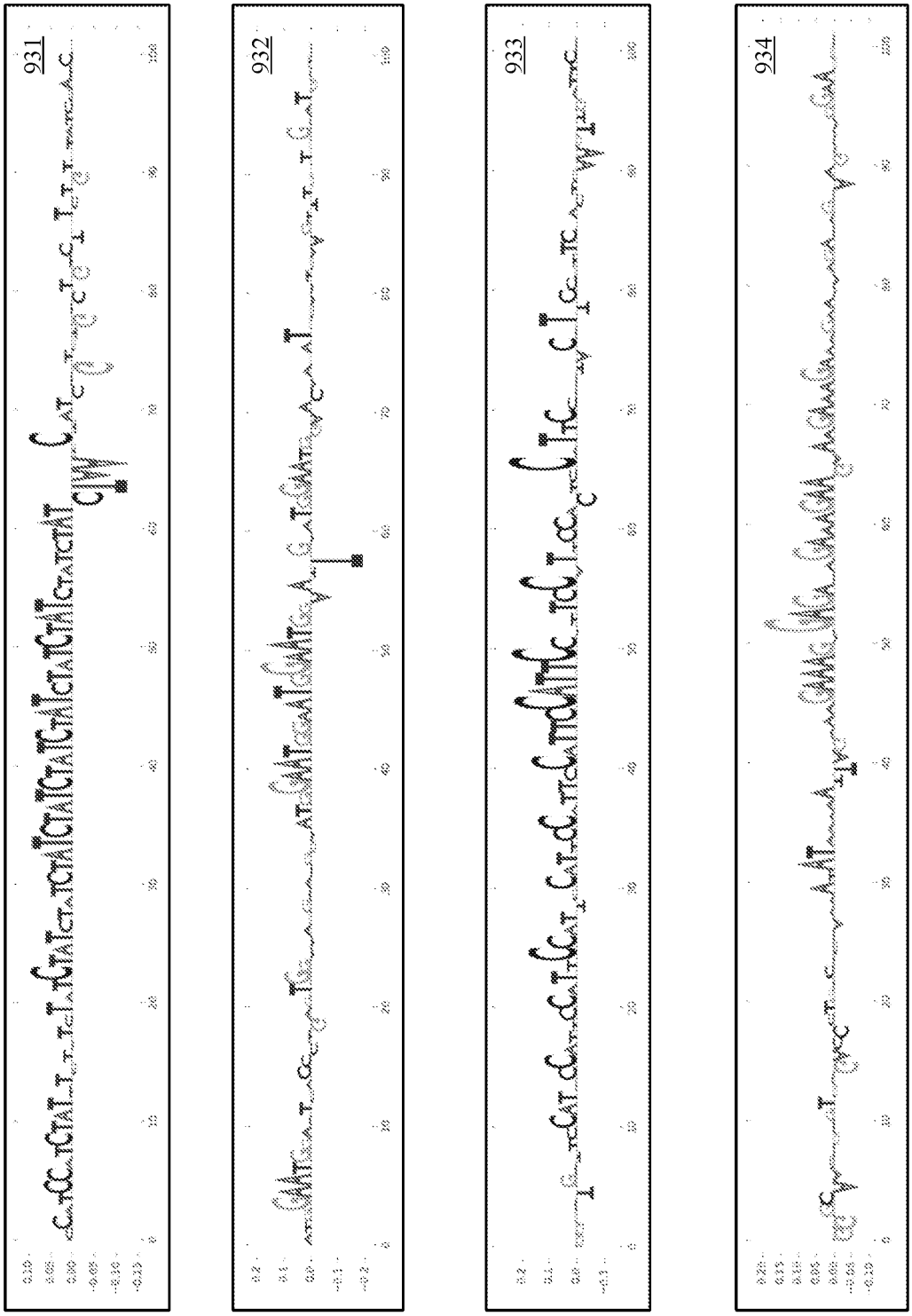


FIG. 10C

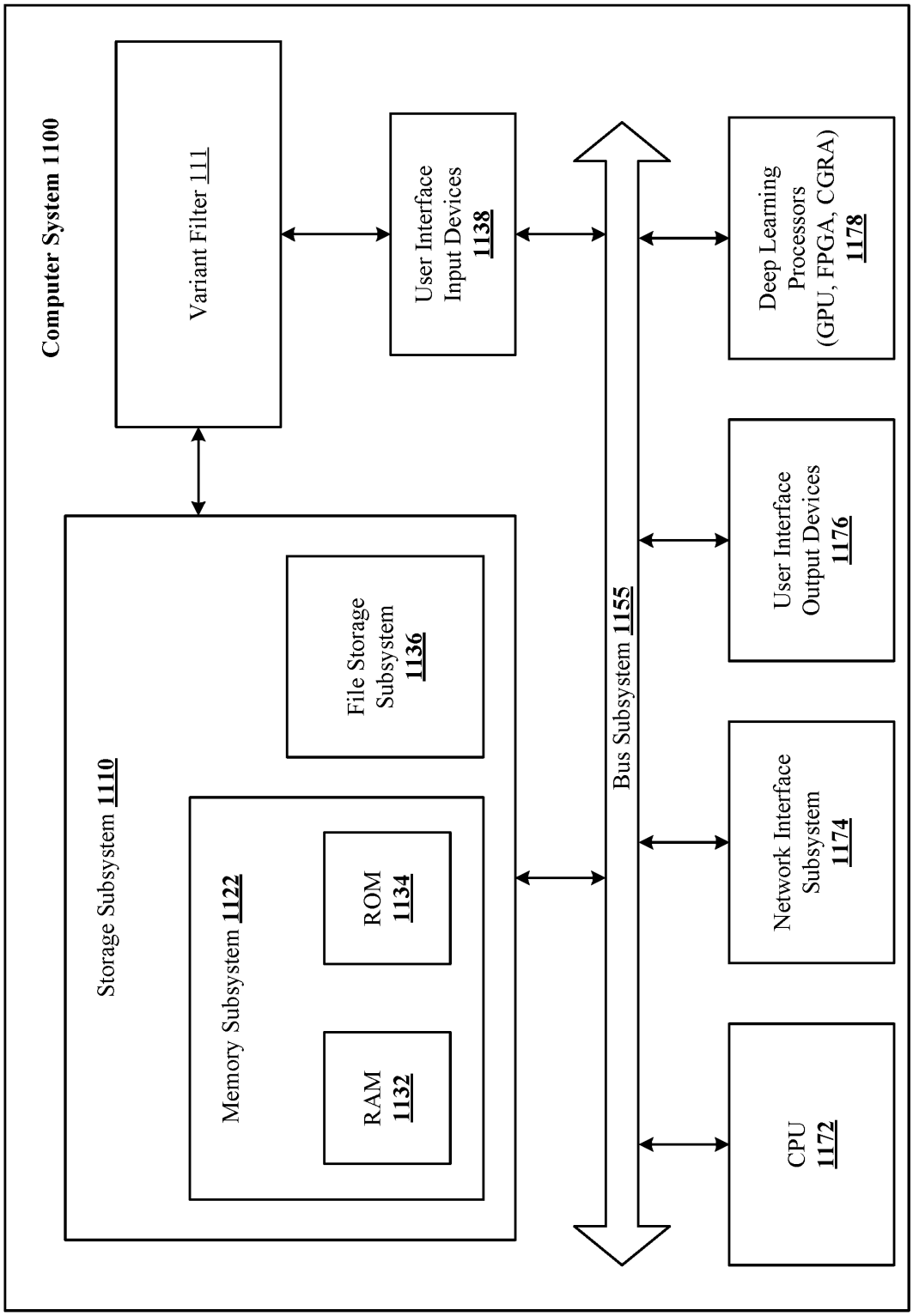

Computer System 1100
Variant Filter 111
User Interface Input Devices 1138
Deep Learning Processors (GPU, FPGA, CGRA) 1178
Storage Subsystem 1110
File Storage Subsystem 1136
Memory Subsystem 1122
ROM 1134
RAM 1132
Bus Subsystem 1155
User Interface Output Devices 1176
Network Interface Subsystem 1174
CPU 1172

FIG. 11

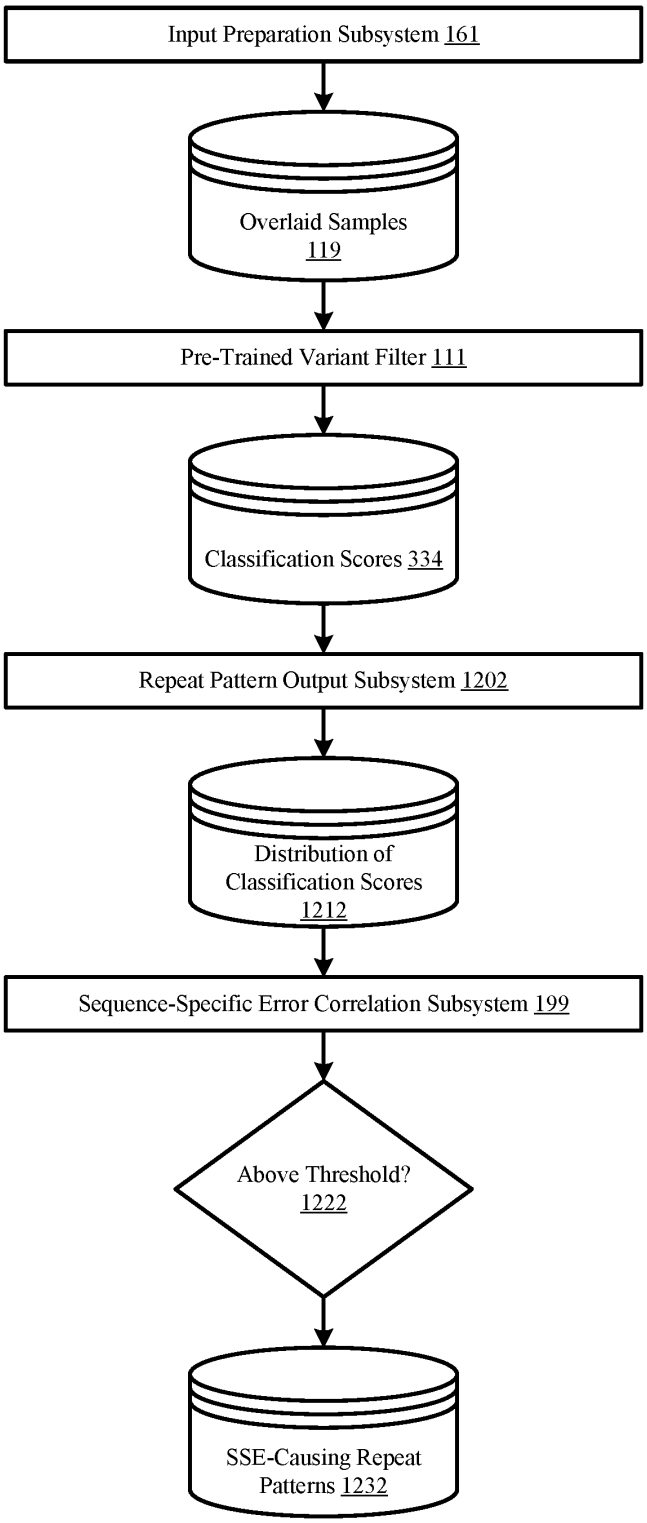

Input Preparation Subsystem 161
Overlaid Samples 119
Pre-Trained Variant Filter 111
Classification Scores 334
Repeat Pattern Output Subsystem 1202
Distribution of Classification Scores 1212
Sequence-Specific Error Correlation Subsystem 199
Above Threshold? 1222
SSE-Causing Repeat Patterns 1232

FIG. 12