

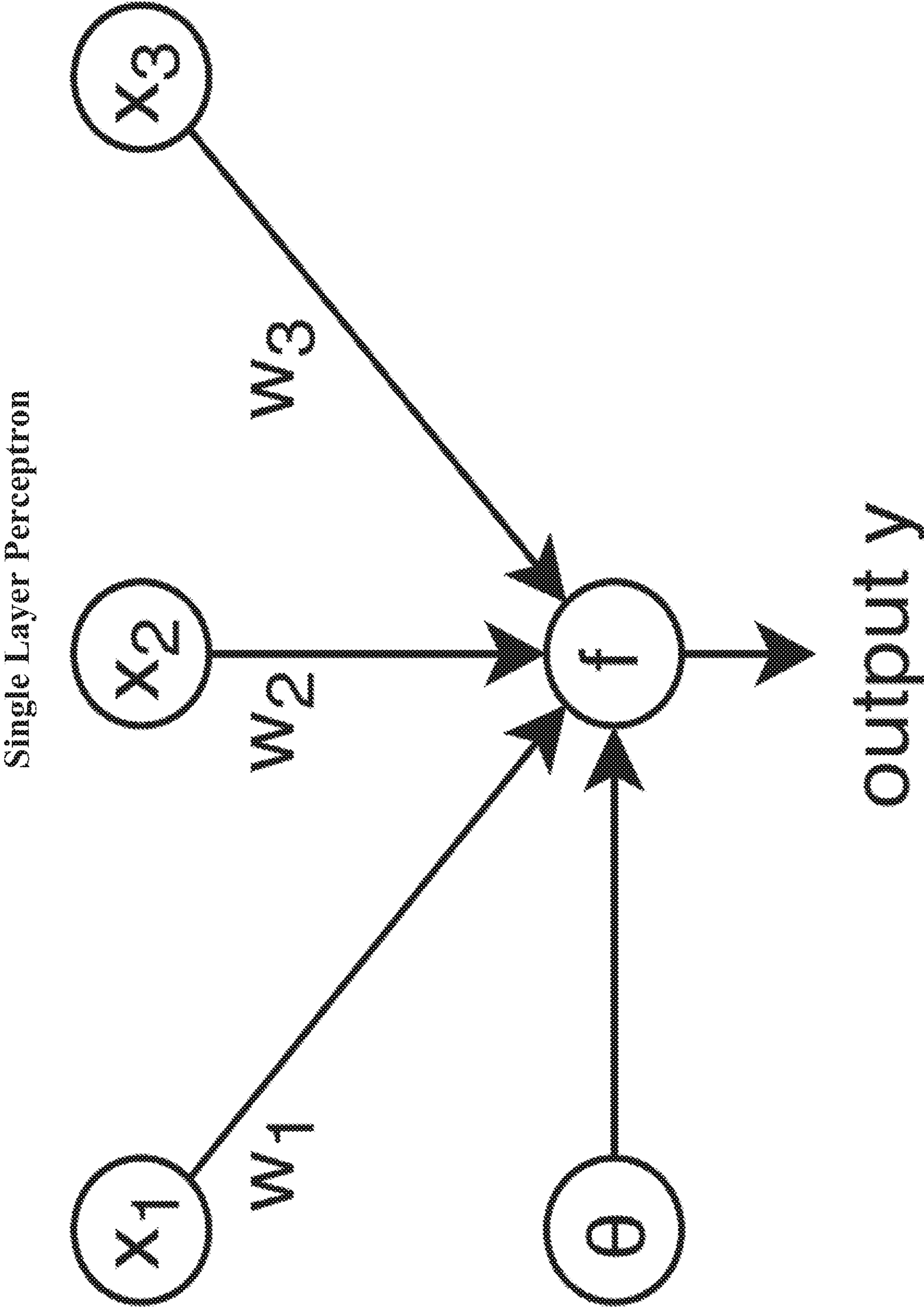


FIG. 1

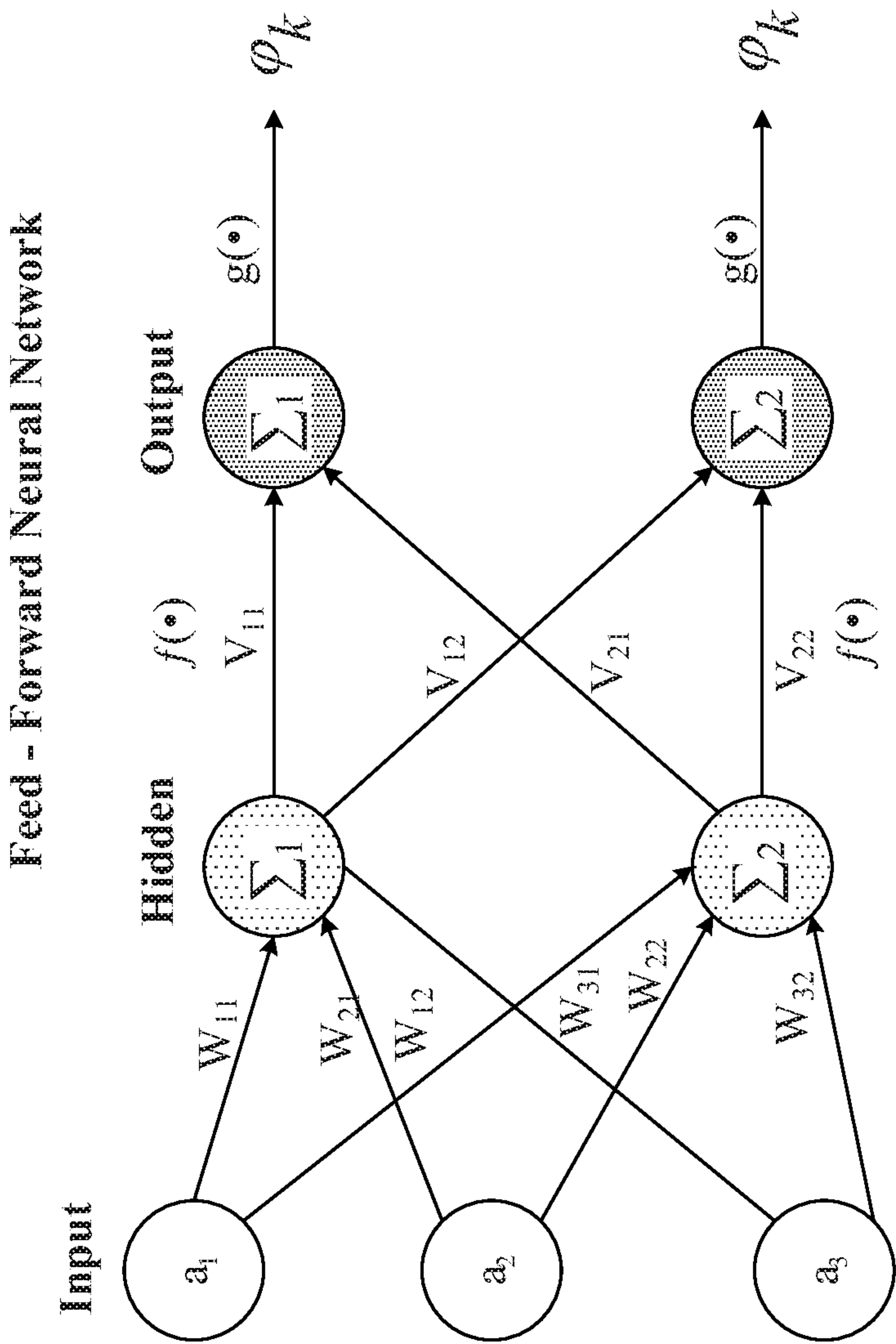


FIG. 2

Convolutional Neural Network

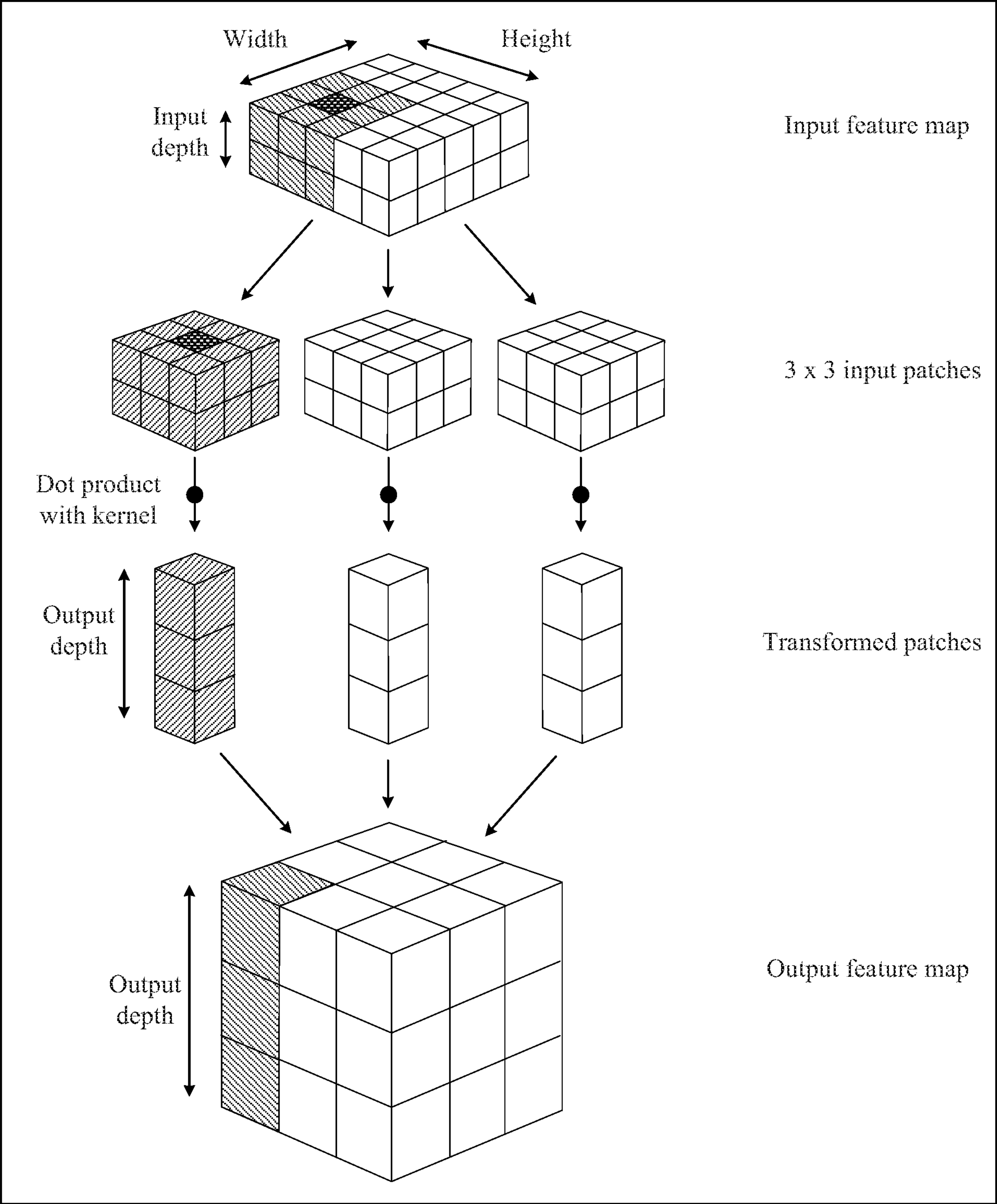


FIG. 3

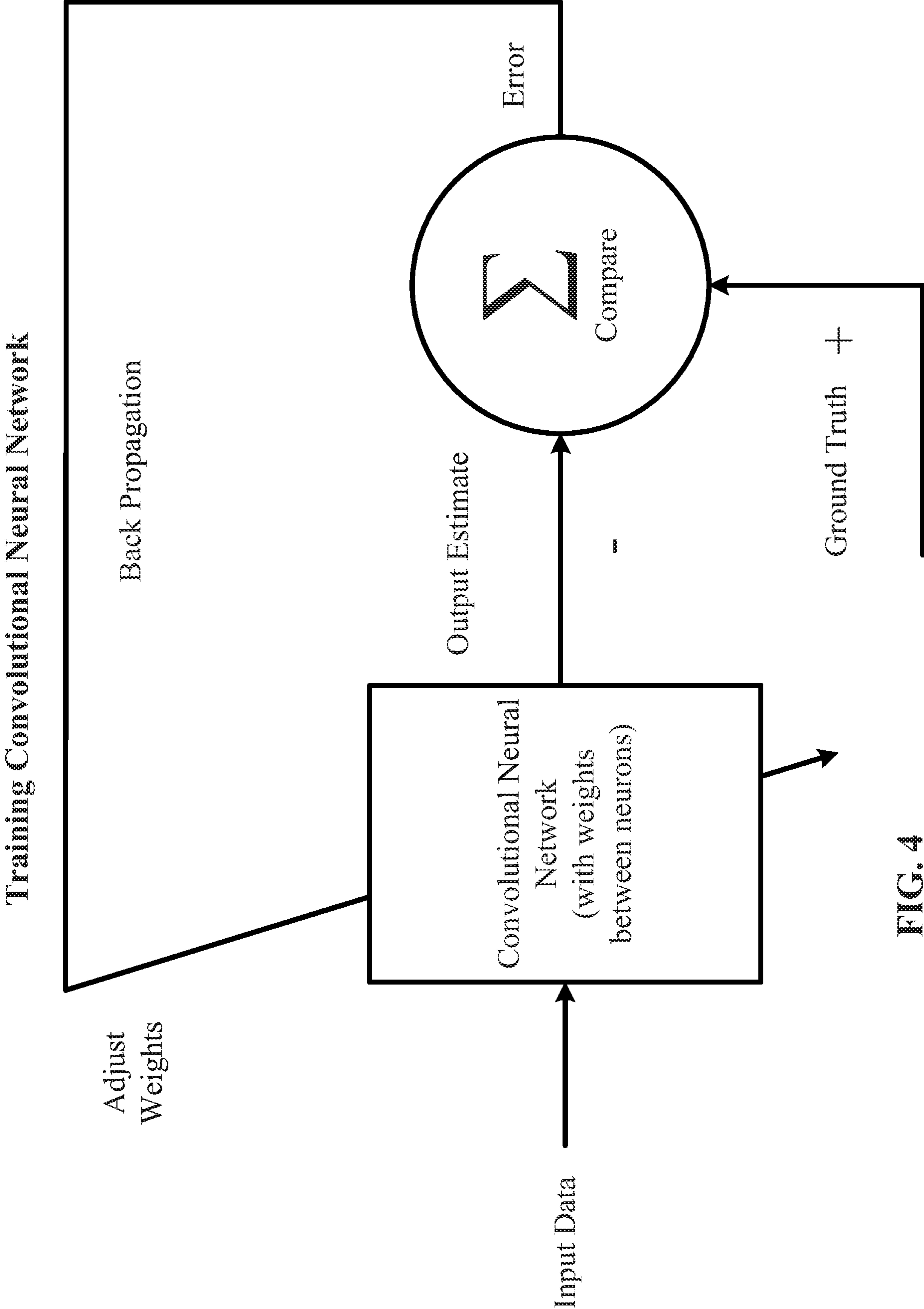
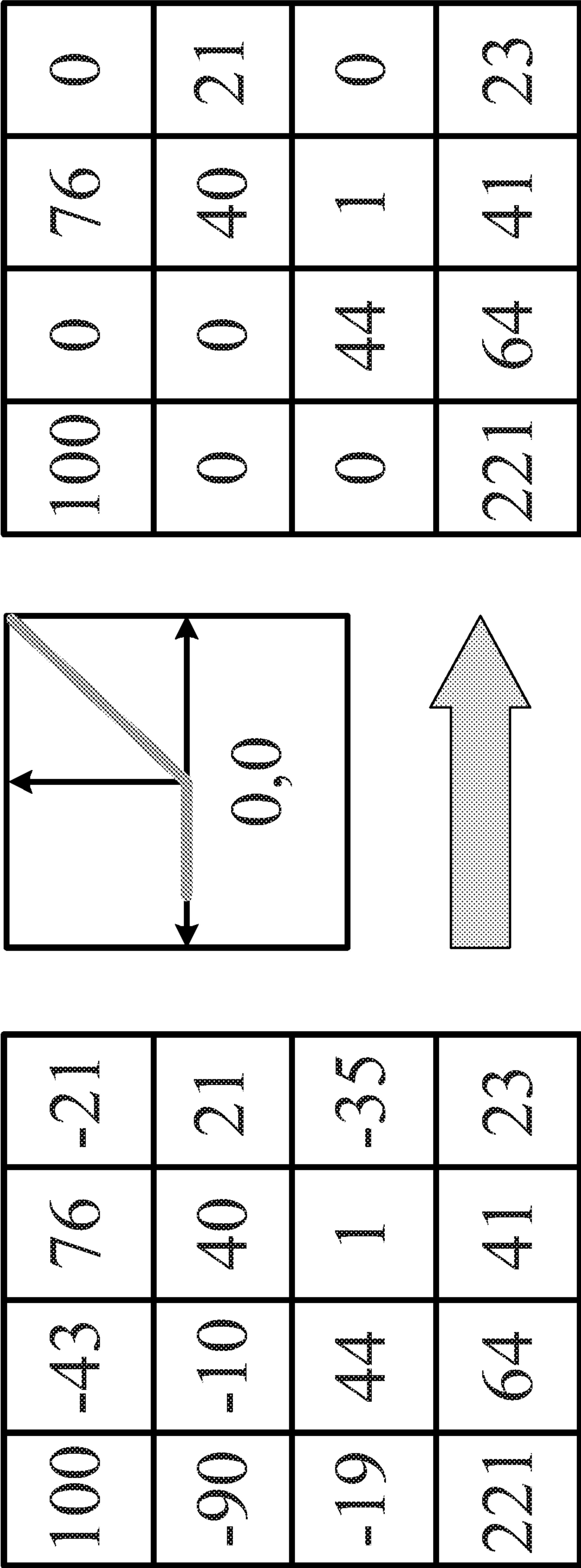


FIG. 4

Non-Linear Layers of Convolutional Neural Networks



ReLU Activation Function

FIG. 5

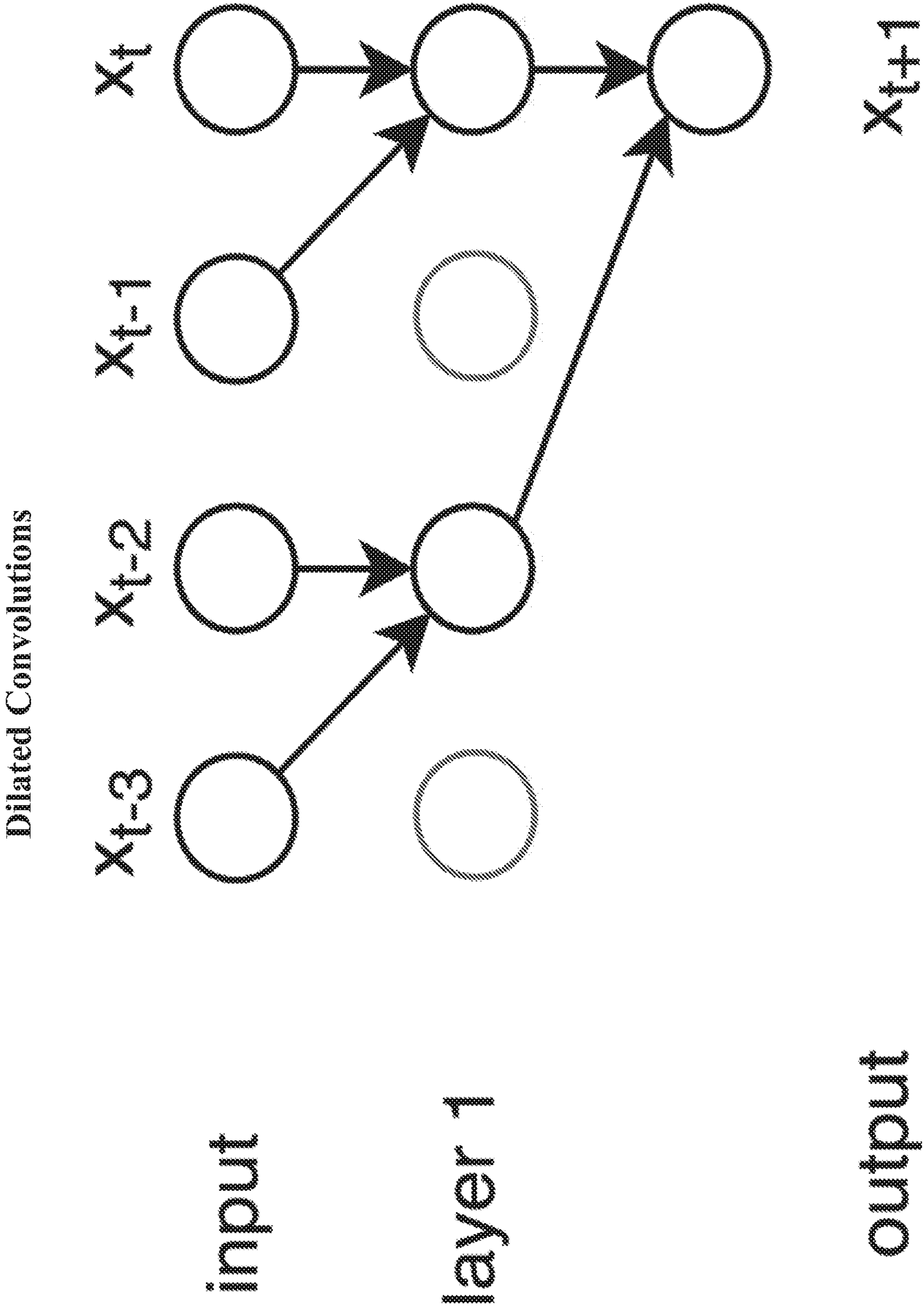


FIG. 6

Sub-Sampling Layers of Convolutional Neural Networks

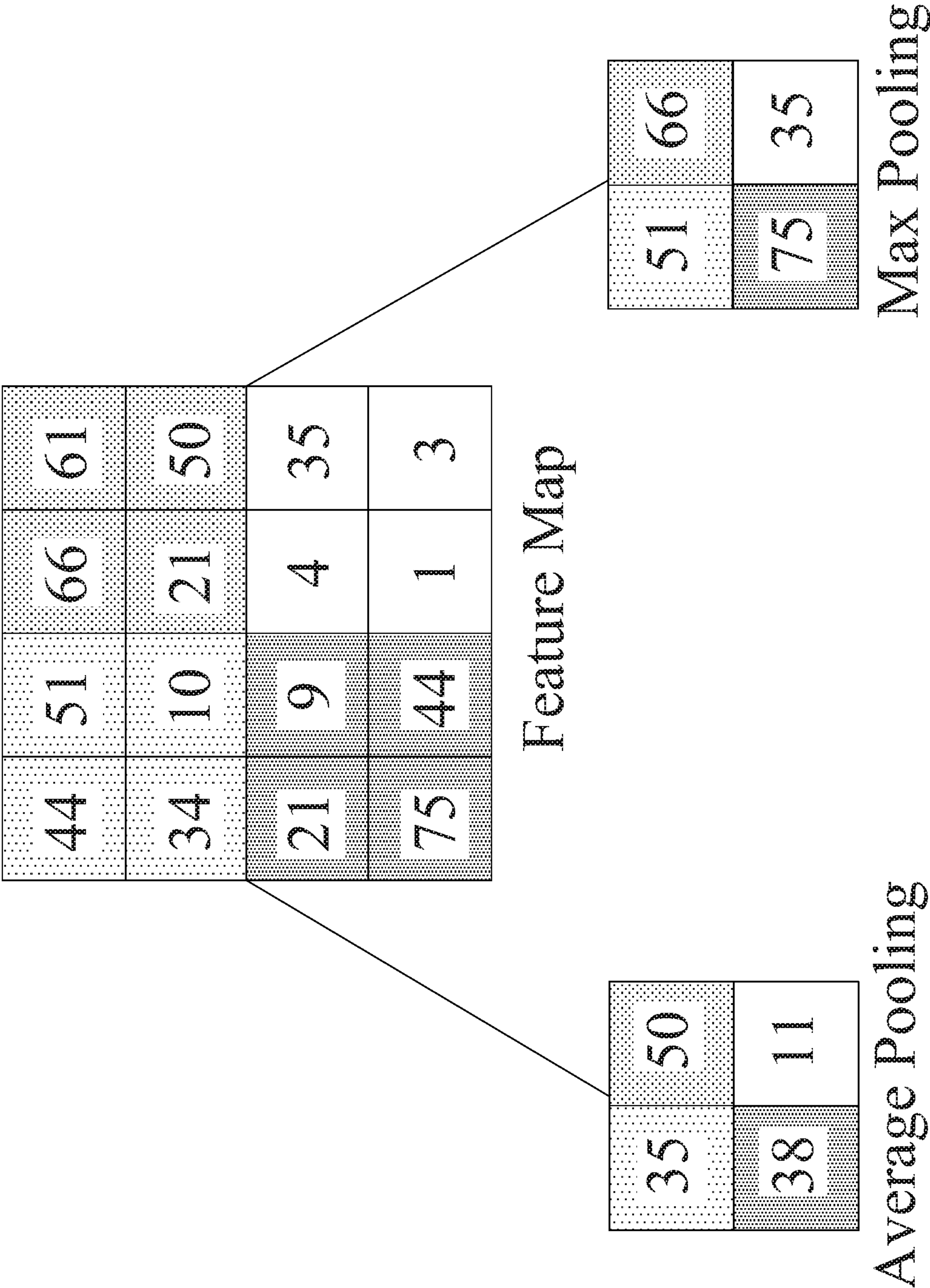


FIG. 7

Convolution Layers of Convolutional Neural Networks

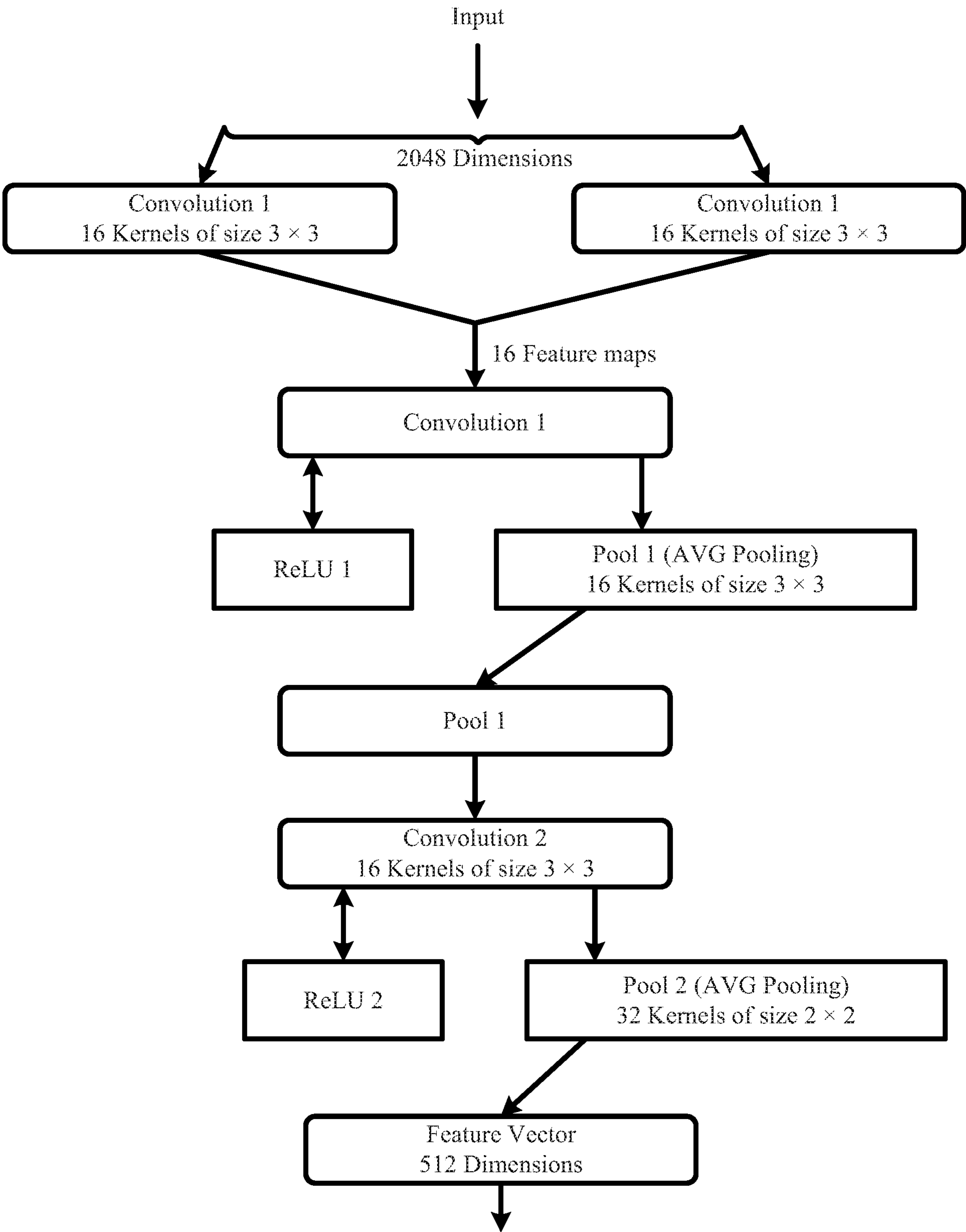


FIG. 8

Residual Connections used in Convolutional Neural Networks

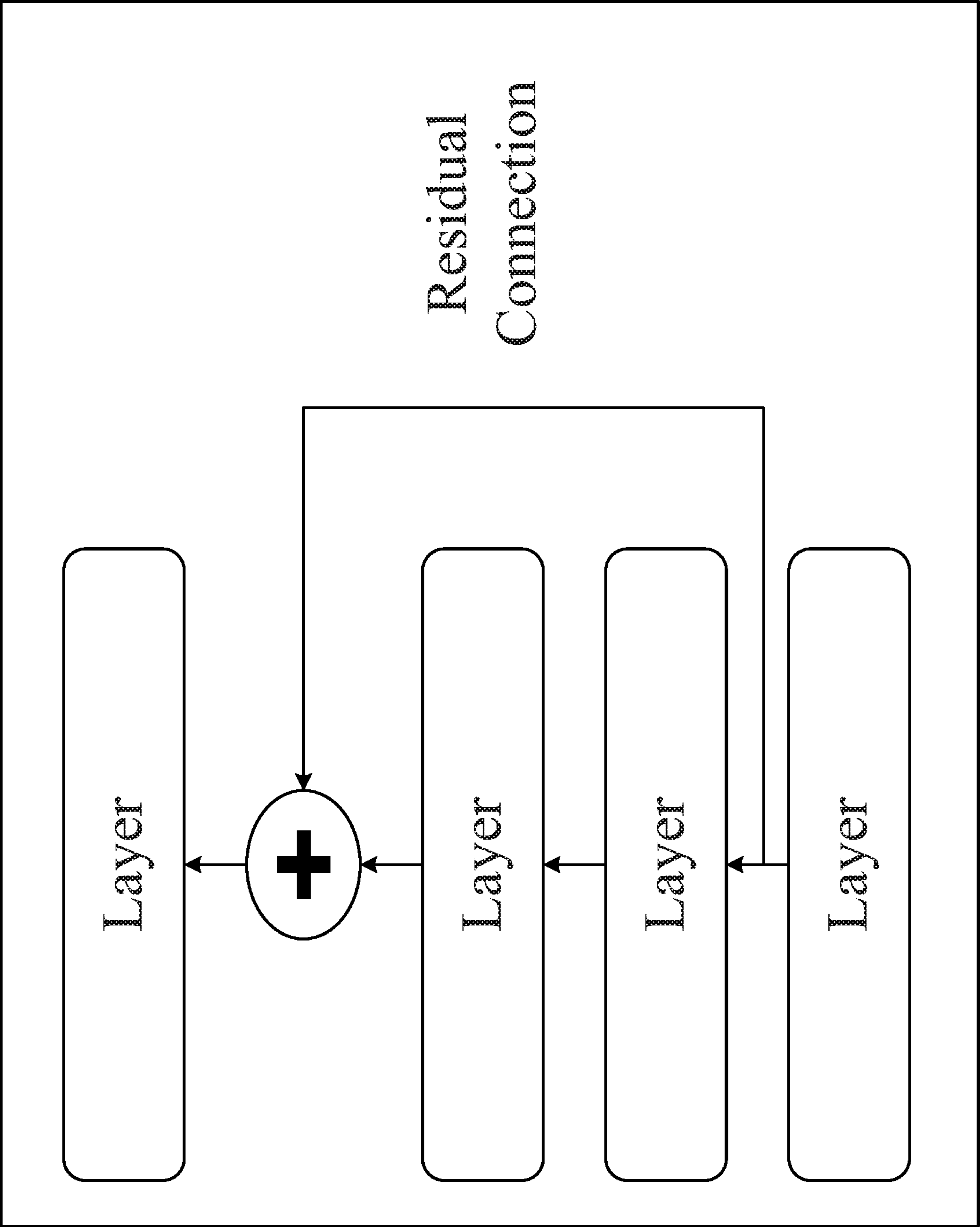
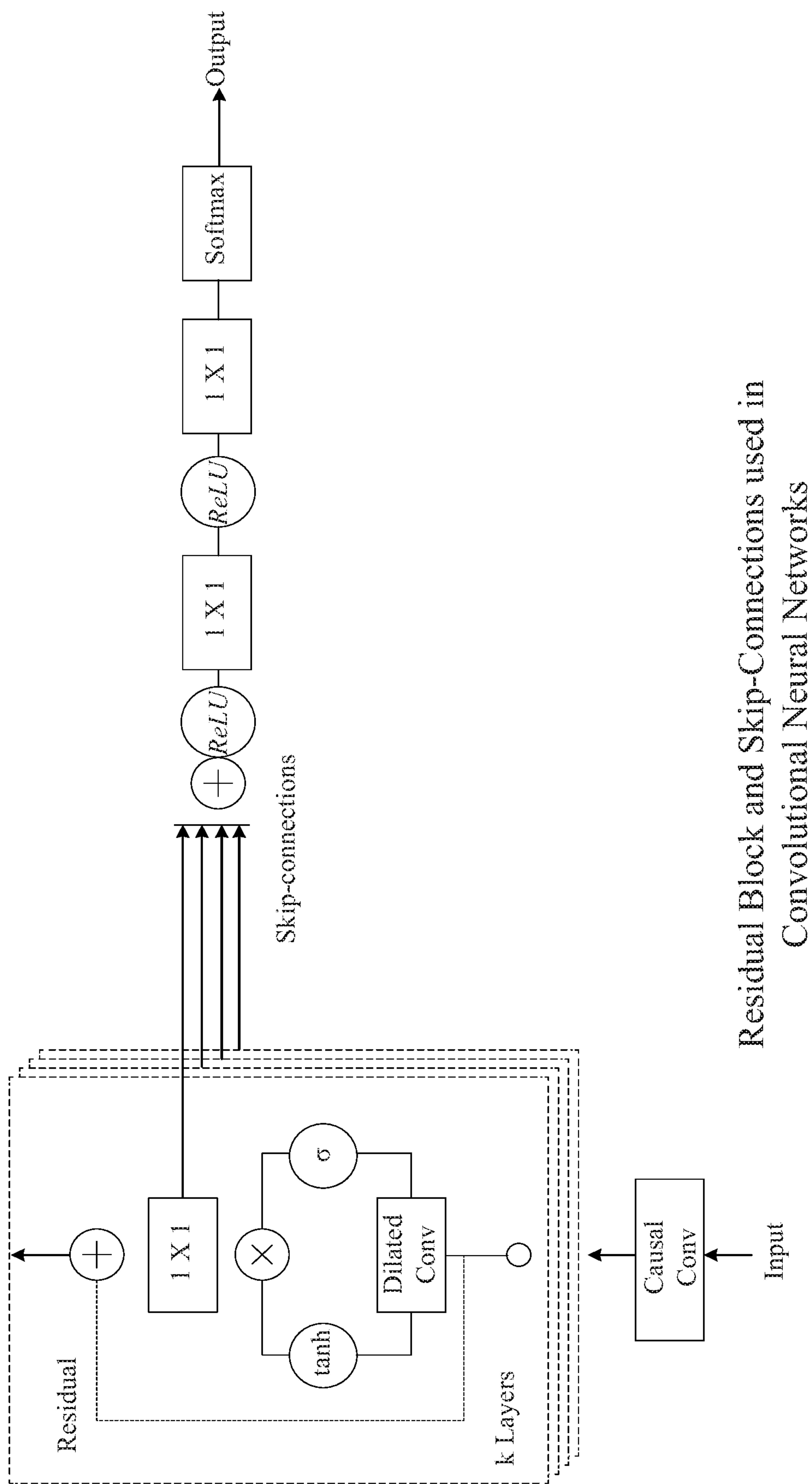


FIG. 9



Residual Block and Skip-Connections used in
Convolutional Neural Networks

FIG. 10

Stacked Dilated Convolutions

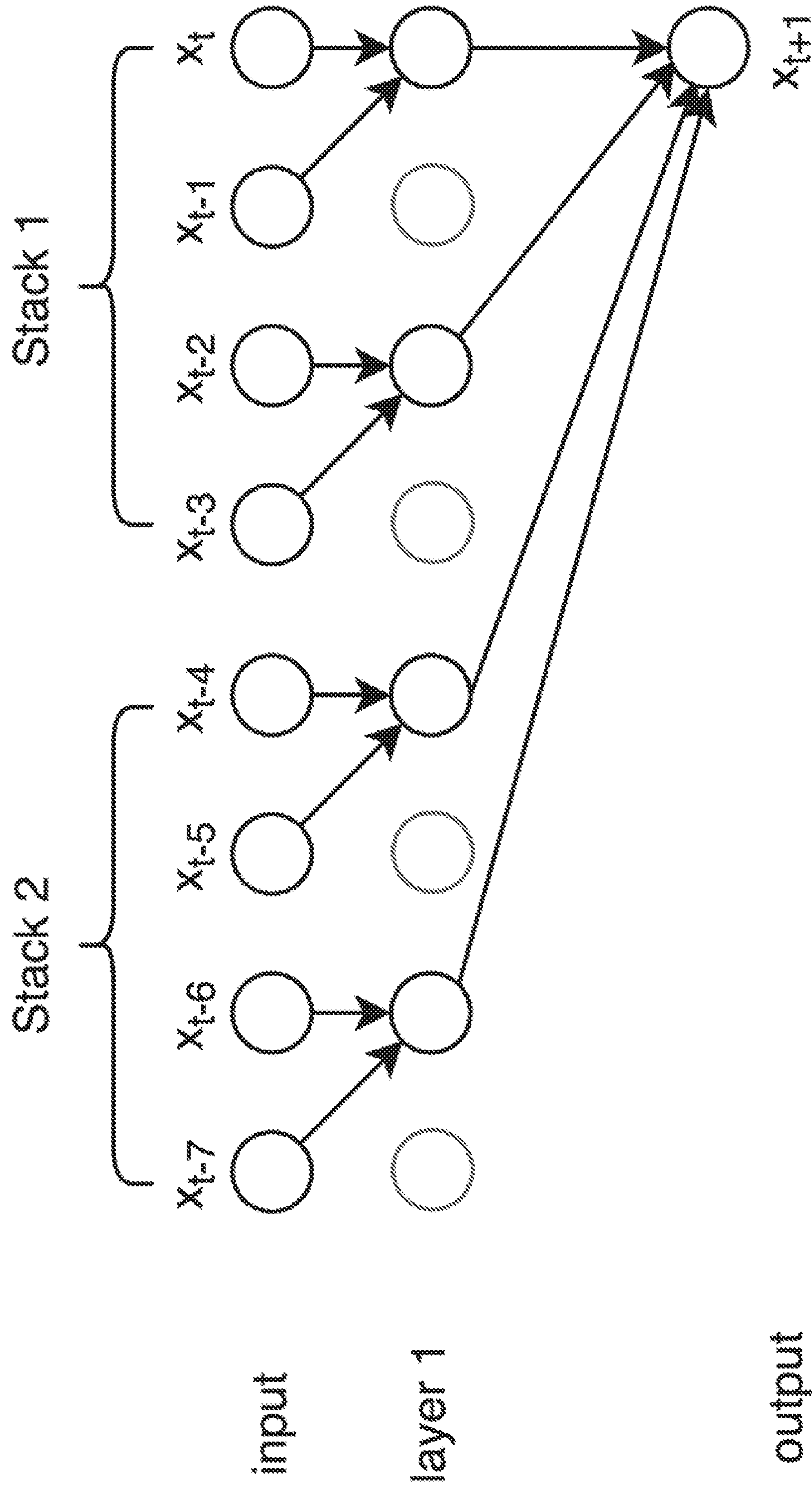


FIG. 11

Batch Normalization Forward Pass with Convolutional Neural Networks

$$\begin{aligned}\mu_B &= \frac{1}{n} \sum_{i=1}^n x_i^{(\ell-1)} \\ \sigma_B^2 &= \frac{1}{n} \sum_{i=1}^n (x_i^{(\ell-1)} - \mu_B)^2 \\ \hat{x}^{(\ell-1)} &= \frac{x^{(\ell-1)} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \\ x^{(\ell)} &= \gamma^{(\ell)} \hat{x}^{(\ell-1)} + \beta^{(\ell)}\end{aligned}$$

FIG. 12

Batch Normalization – Inference with Convolutional Neural Networks

$$\hat{x}^{(\ell-1)} = \frac{x^{(\ell-1)} - \mu_{\mathcal{D}}}{\sqrt{\sigma_{\mathcal{D}}^2 + \epsilon}}$$

$$x_i^{(\ell)} = \gamma^{(\ell)} \hat{x}_i^{(\ell-1)} + \beta^{(\ell)}$$

FIG. 13

Batch Normalization Backward Pass with Convolutional Neural Networks

$$\begin{aligned}\nabla_{\gamma^{(\ell)}} \mathcal{L} &= \sum_{i=1}^n (\nabla_{x^{(\ell+1)}} \mathcal{L})_i \cdot \hat{x}_i^{(\ell)} \\ \nabla_{\beta^{(\ell)}} \mathcal{L} &= \sum_{i=1}^n (\nabla_{x^{(\ell+1)}} \mathcal{L})_i\end{aligned}$$

FIG. 14

Batch Normalization in Convolution Layers

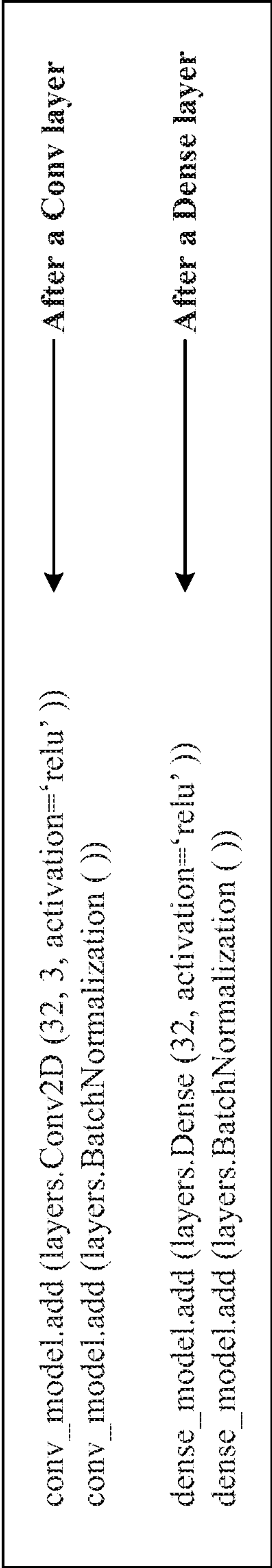


FIG. 15

1D Convolution used in Convolutional Neural Networks

Networks

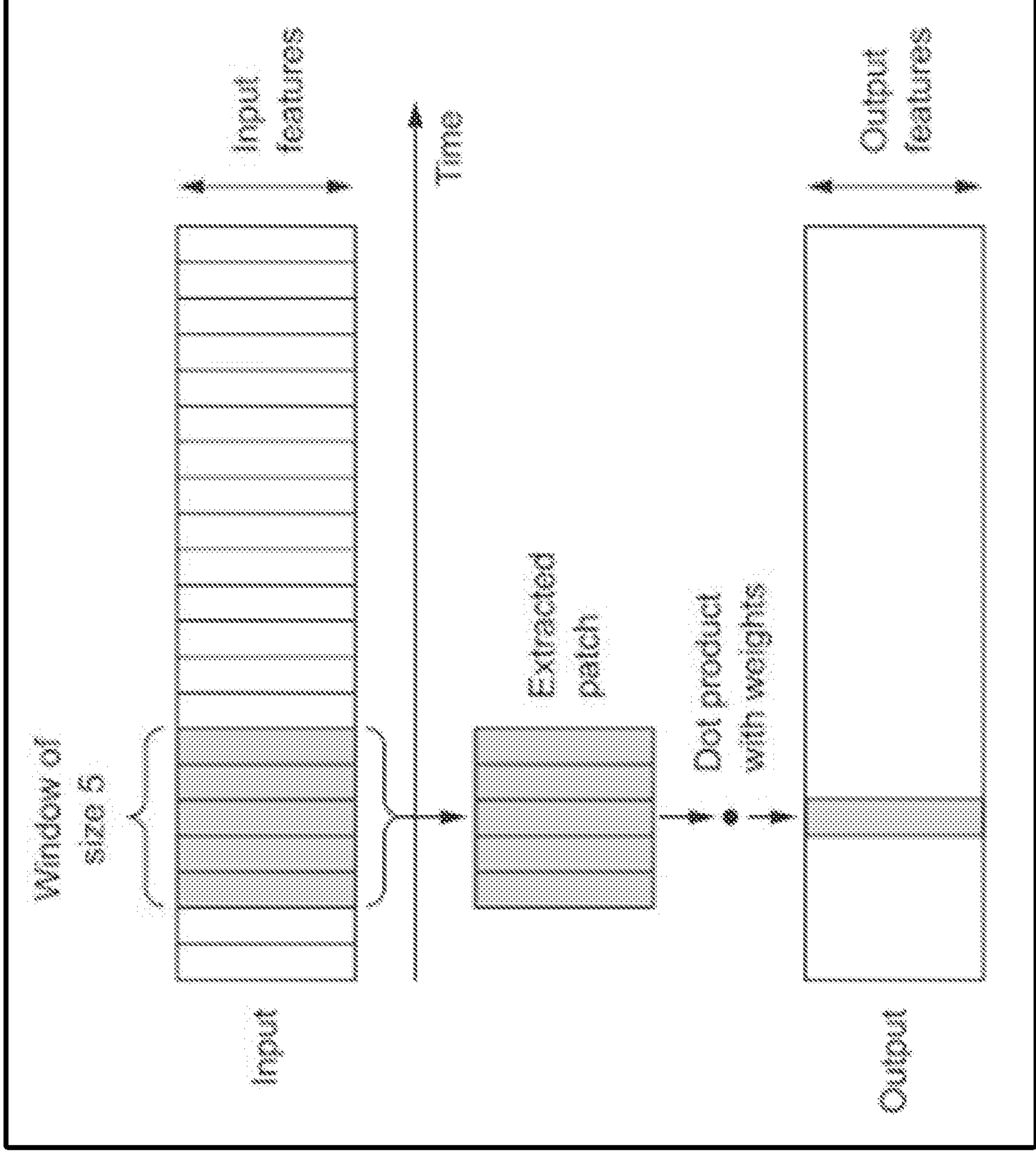


FIG. 16

Global Average Pooling (GAP) in Convolutional Neural Networks

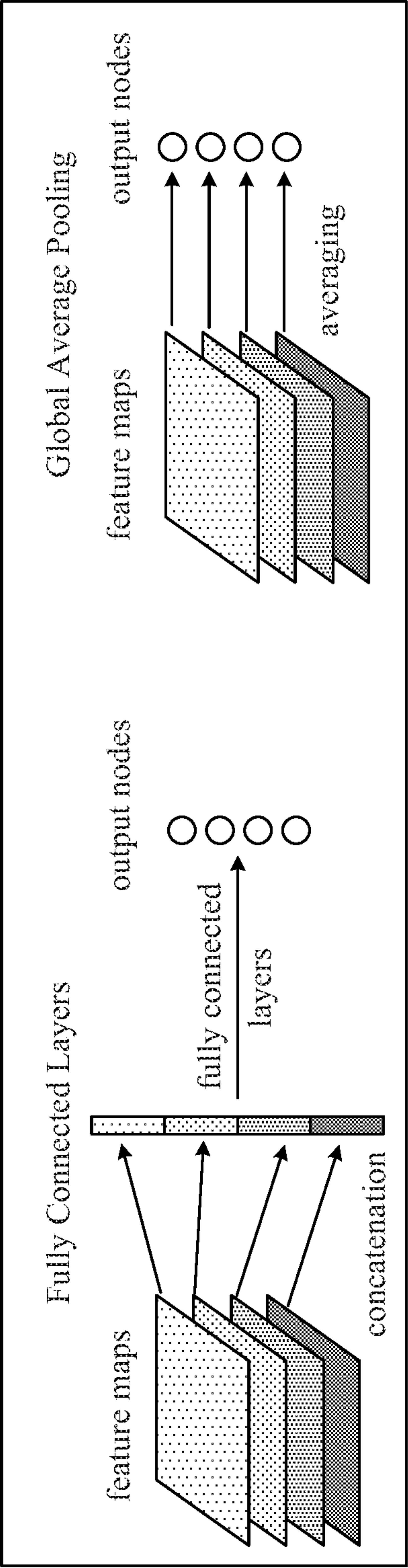
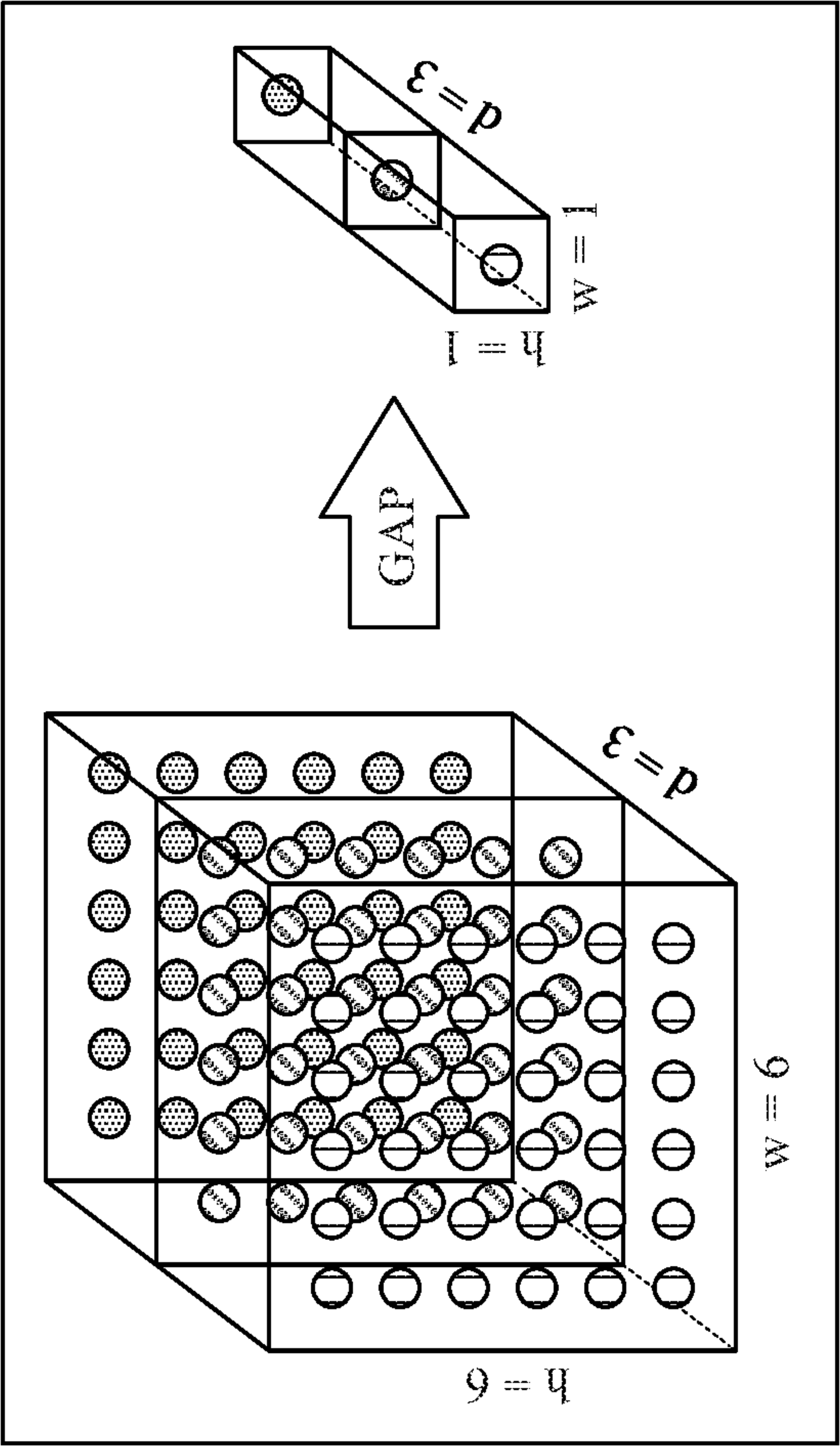


FIG. 17

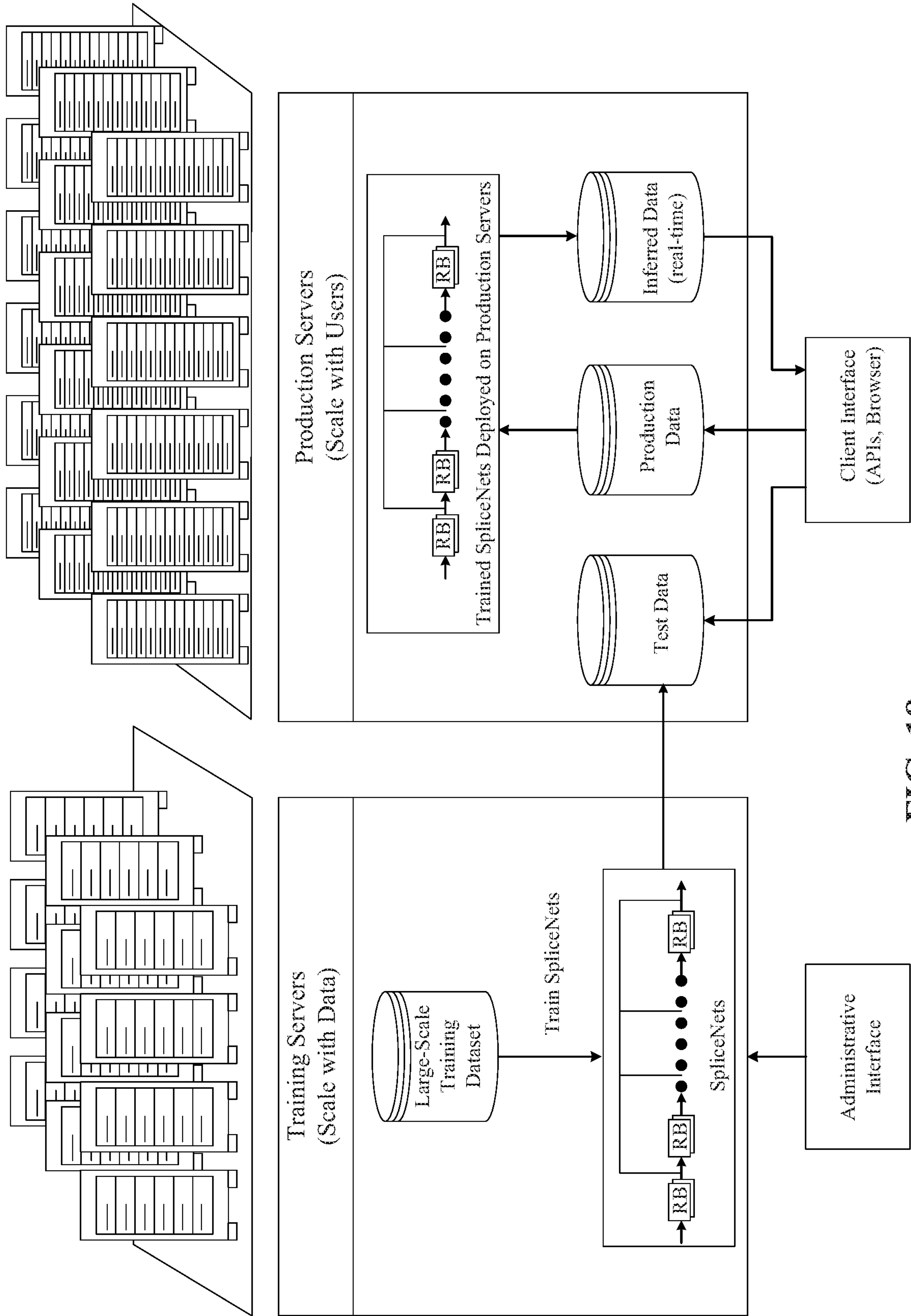


FIG. 18

SpliceNet

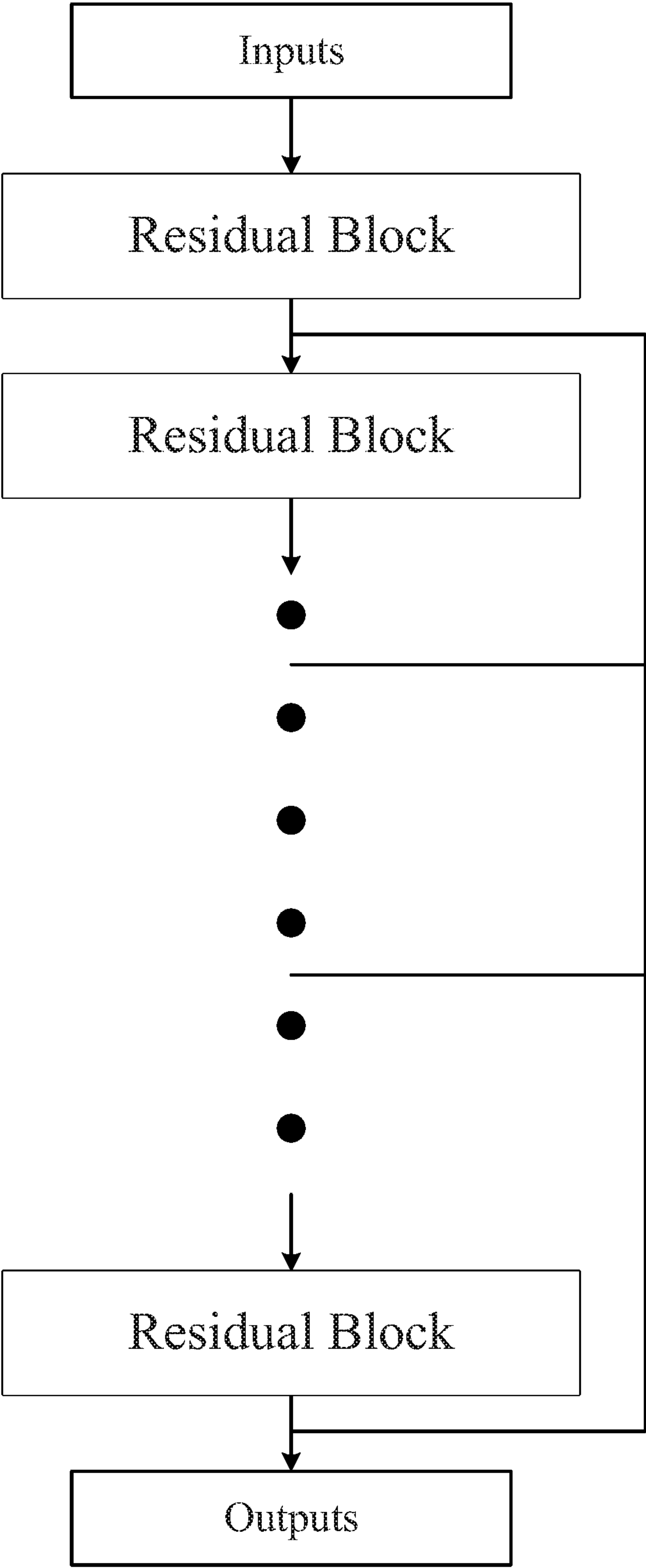
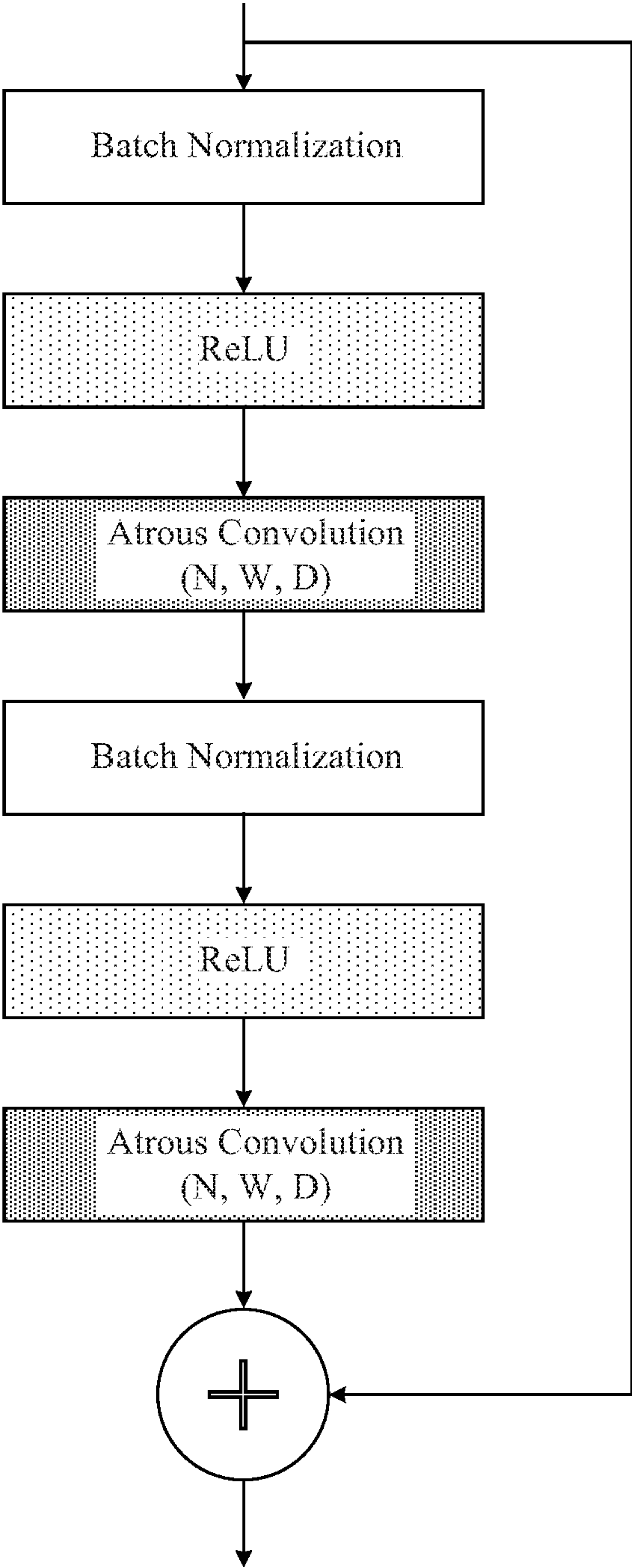


FIG. 19

Residual Block (RB) (N, W, D)



N: Number of Convolution Filters
W: Convolution Window Size
D: Atrous Convolution Rate

FIG. 20

SpliceNet80

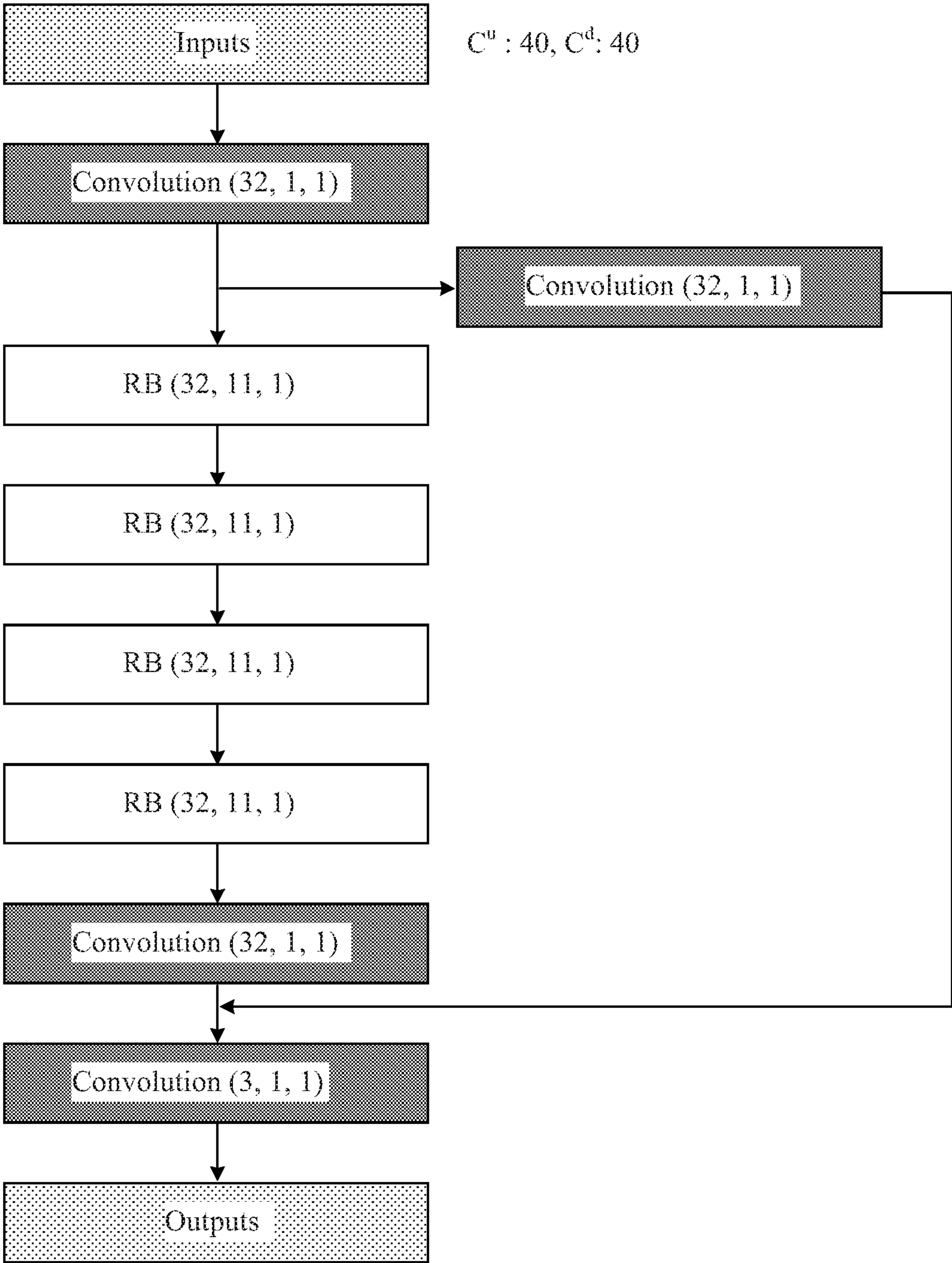


FIG. 21

SpliceNet400

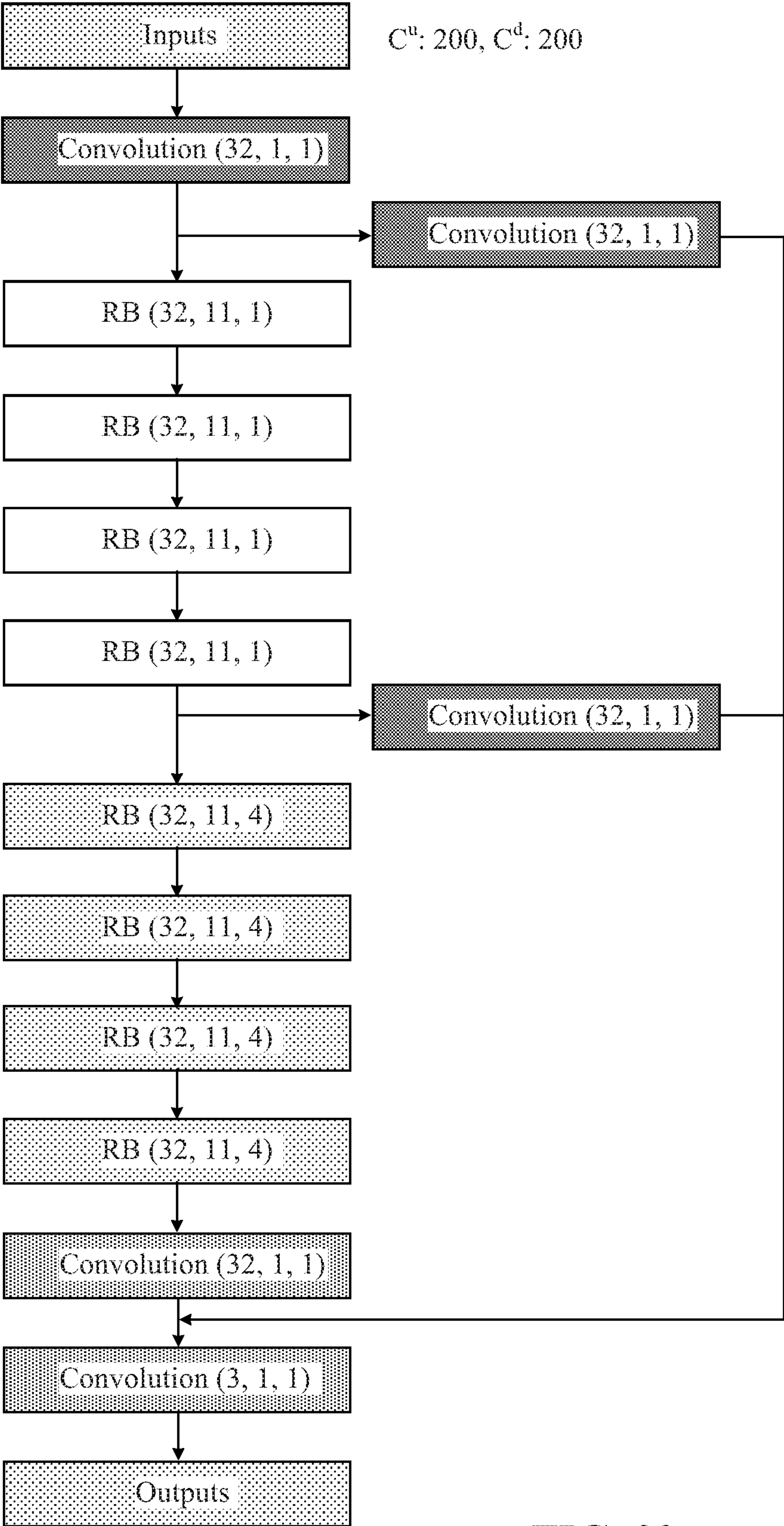


FIG. 22

SpliceNet2000

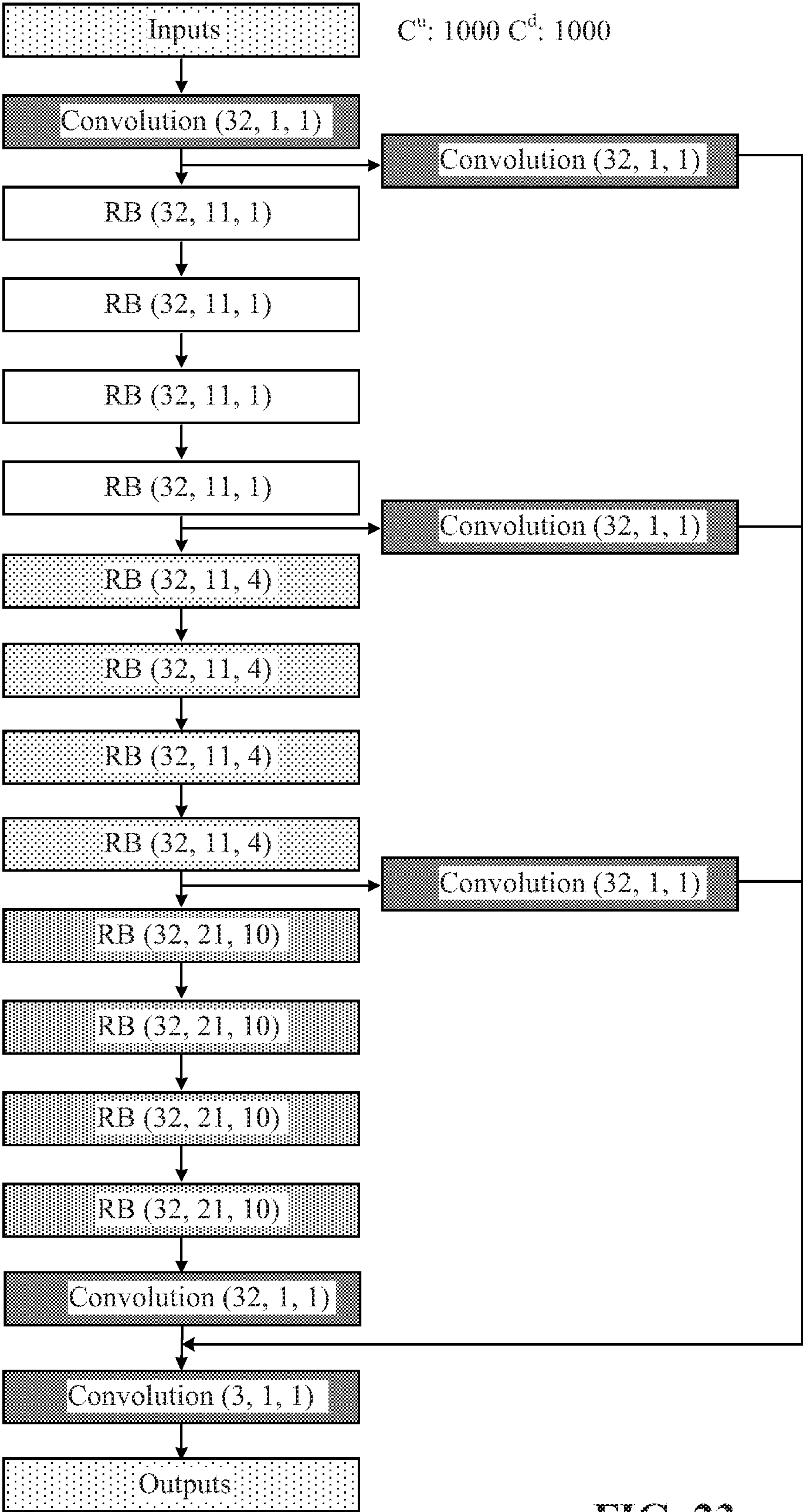


FIG. 23

SpliceNet10000

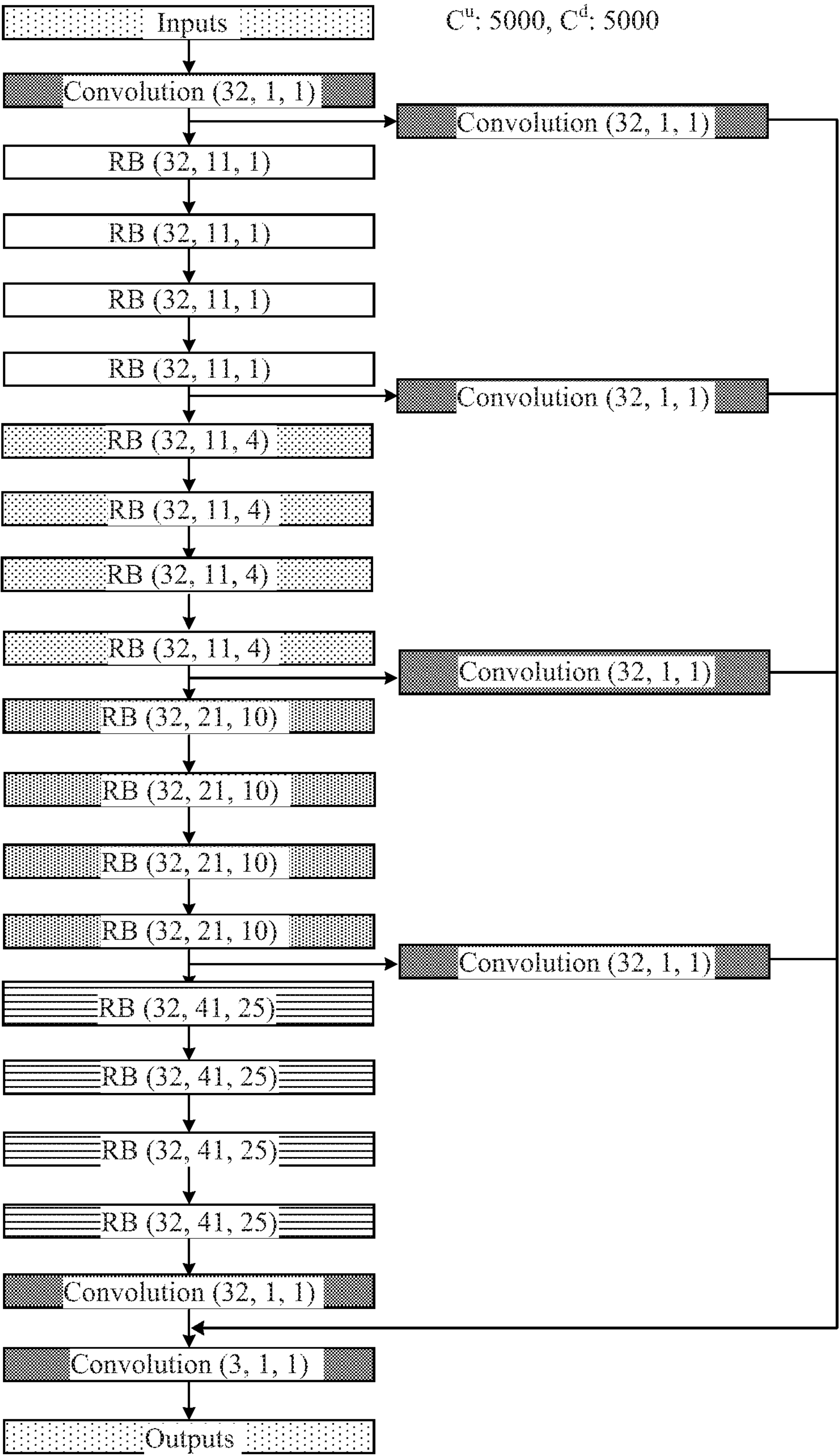
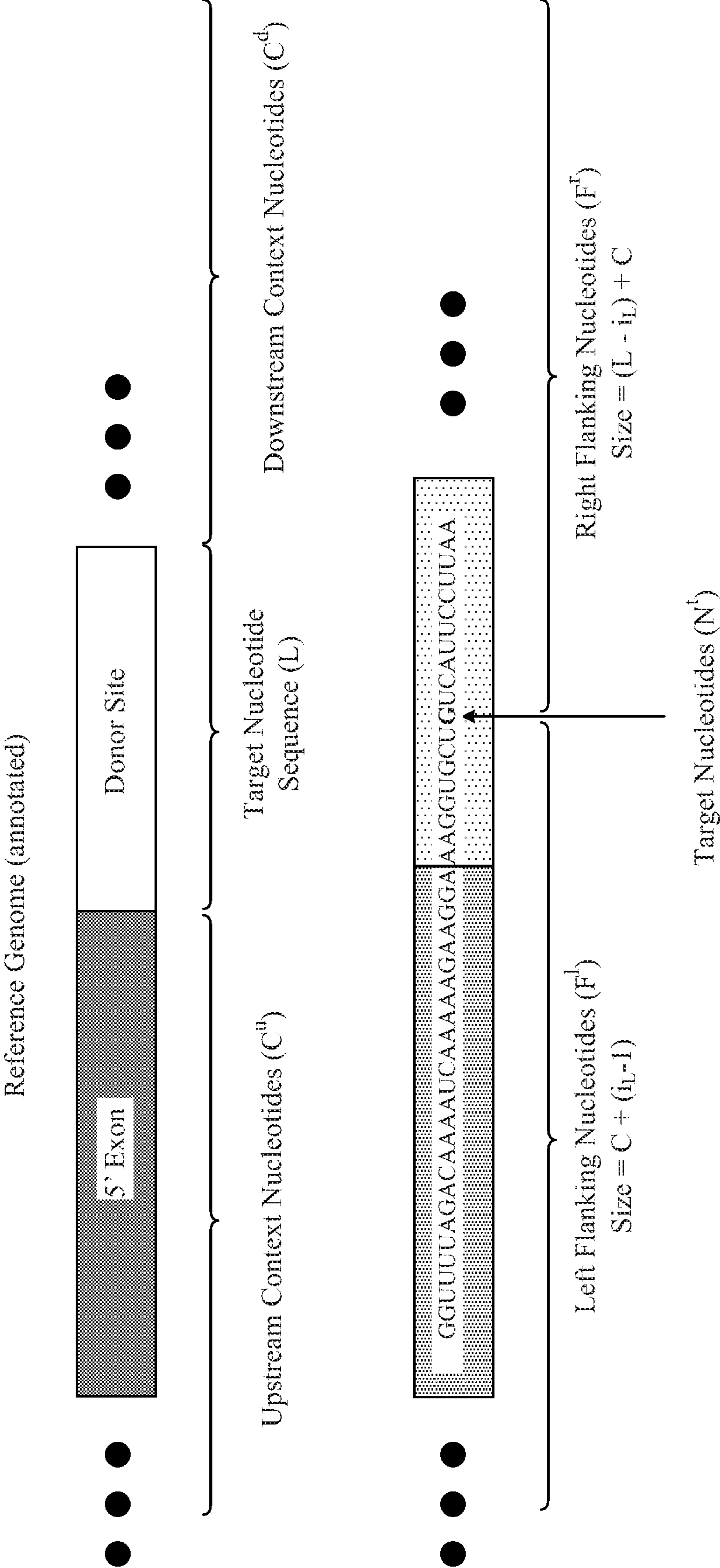
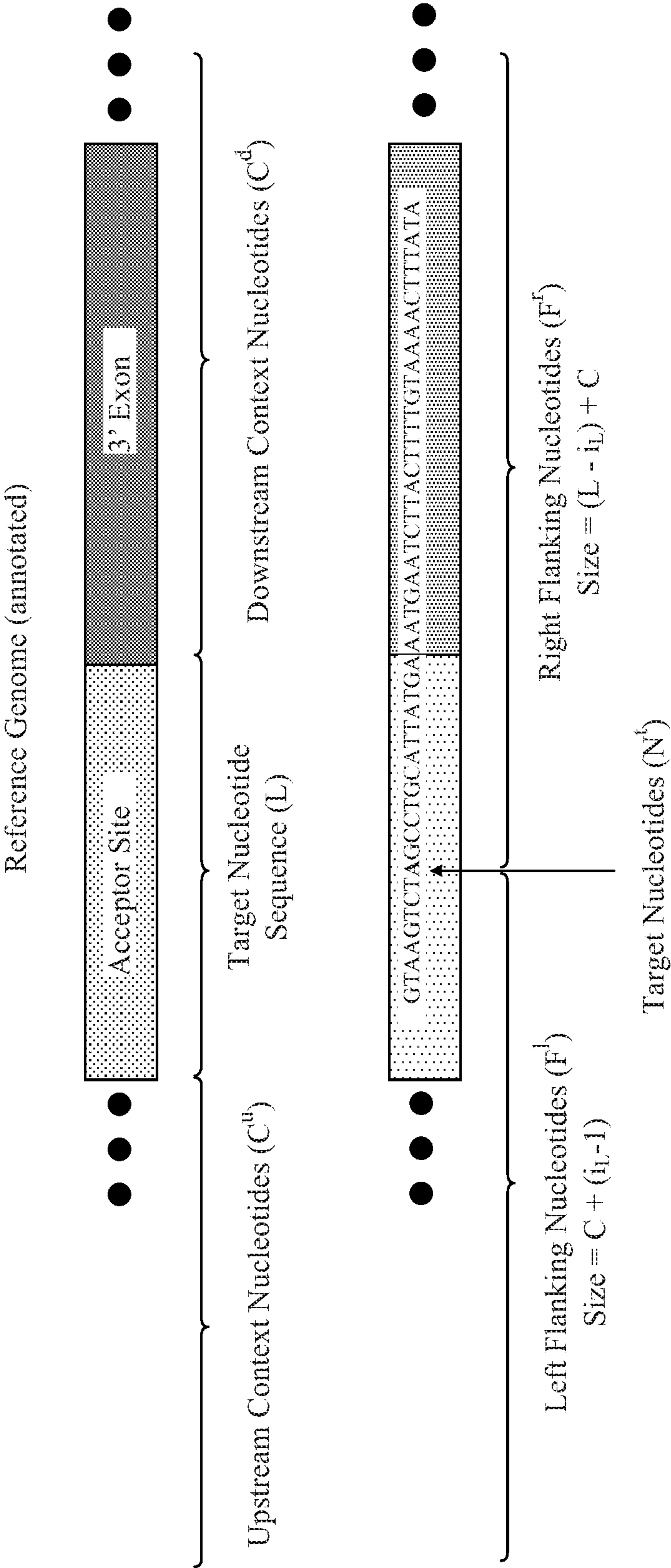


FIG. 24



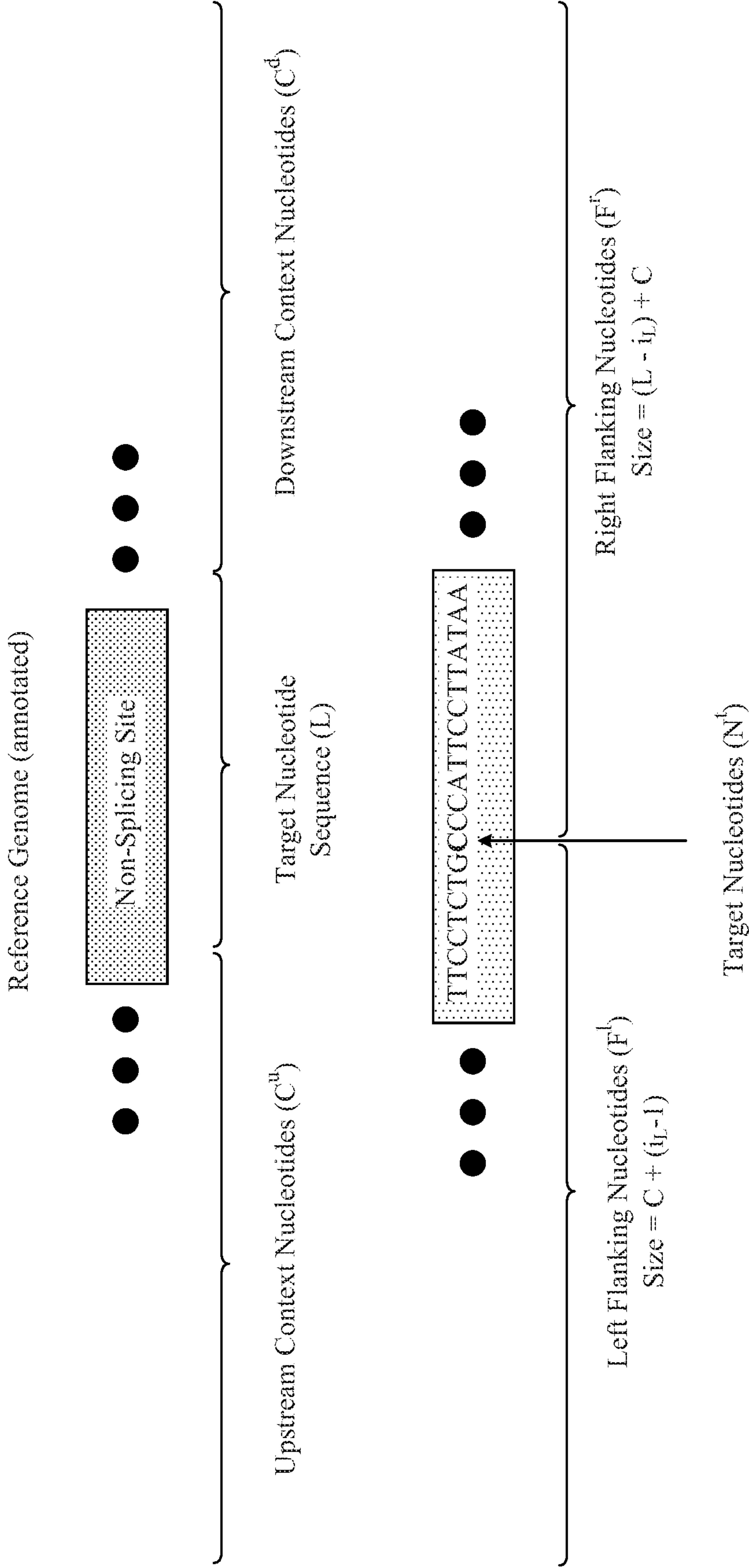
C^u : 40, 200, 1000, 5000
 C^d : 40, 200, 1000, 5000
 L : 5000

FIG. 25



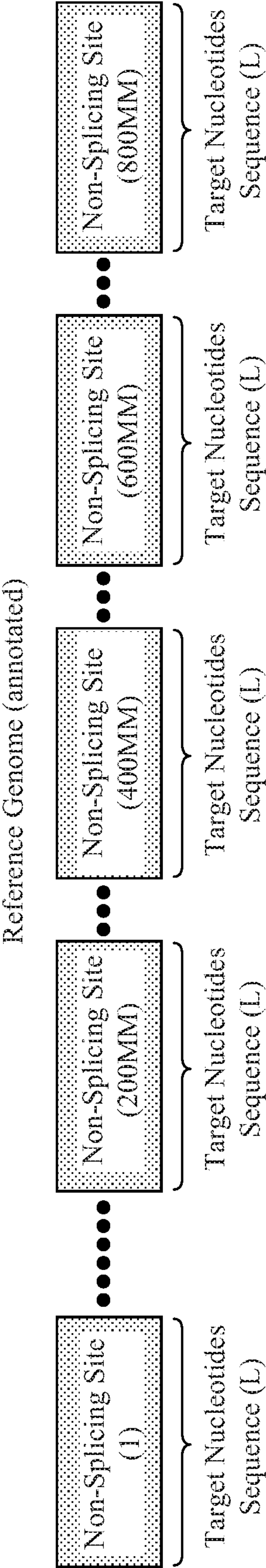
C^u : 40, 200, 1000, 5000
 C^d : 40, 200, 1000, 5000
 L : 5000

FIG. 26



C^u : 40, 200, 1000, 5000
 C^d : 40, 200, 1000, 5000
L: 5000

FIG. 27



L: 5000

FIG. 28

	A	C	G	T	T	G	C	A	C	A	C	A	C	G	?	G	T	A	T	A	T
A	1	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	1	0	1	0
C	0	1	0	0	0	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0
G	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0
T	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

One-hot Encoding

A C G T T T G C A C A C A C G ? G T A T A T

FIG. 29

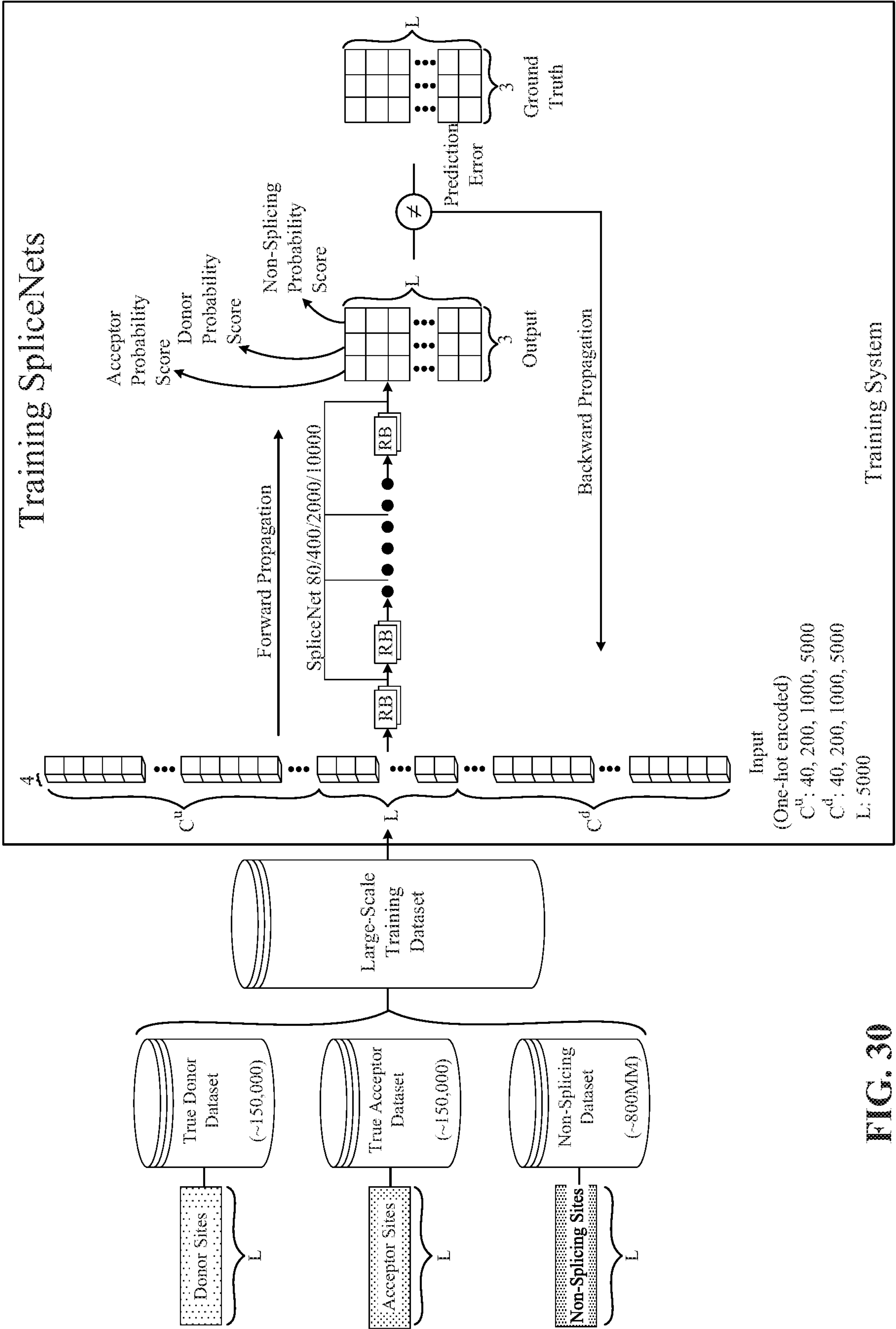


FIG. 30

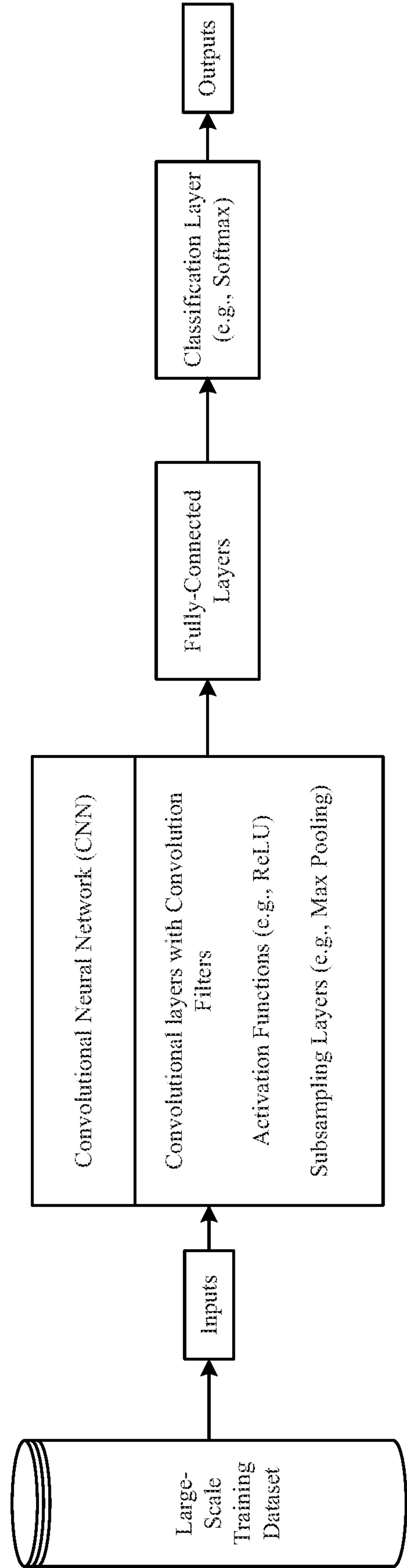


FIG. 31

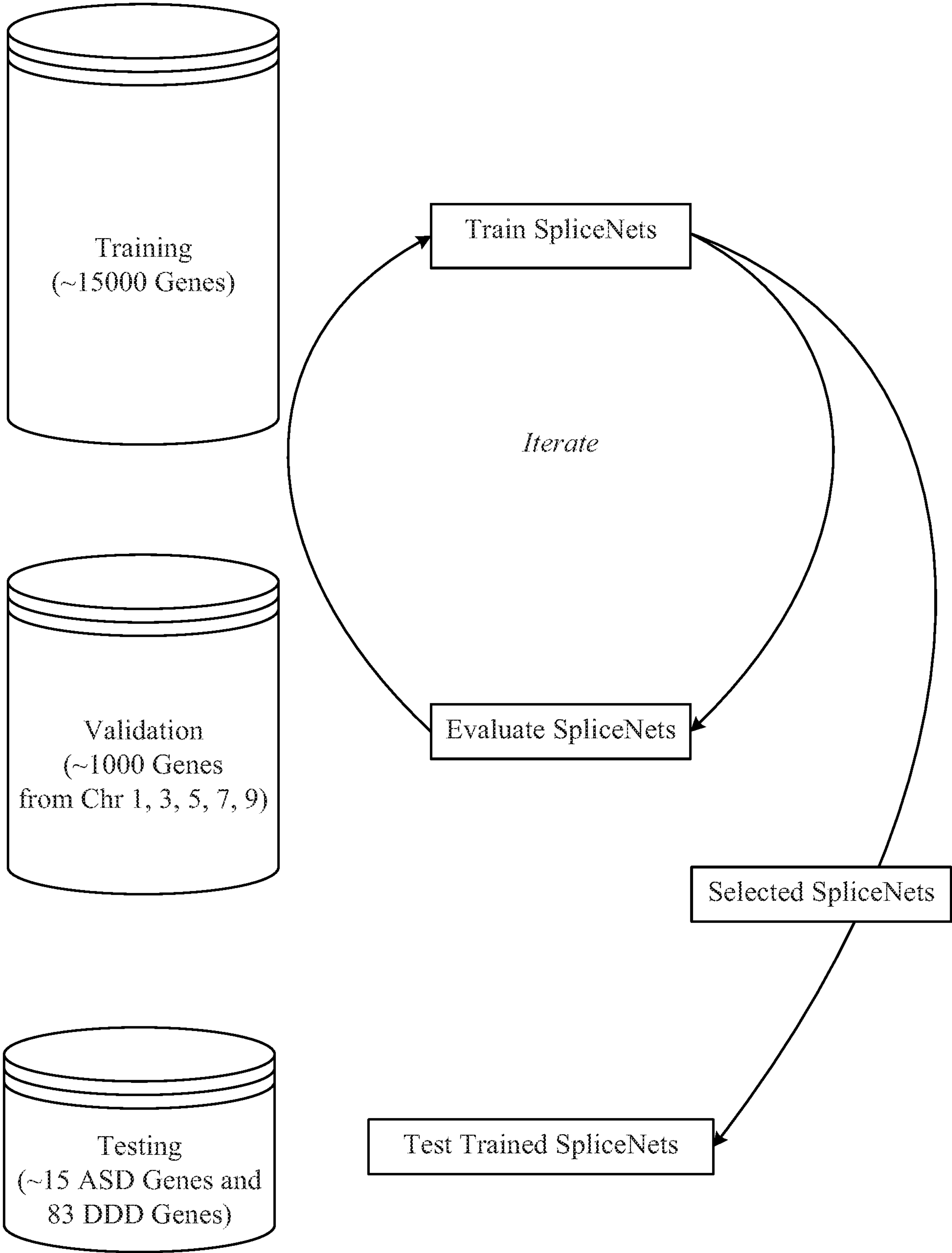
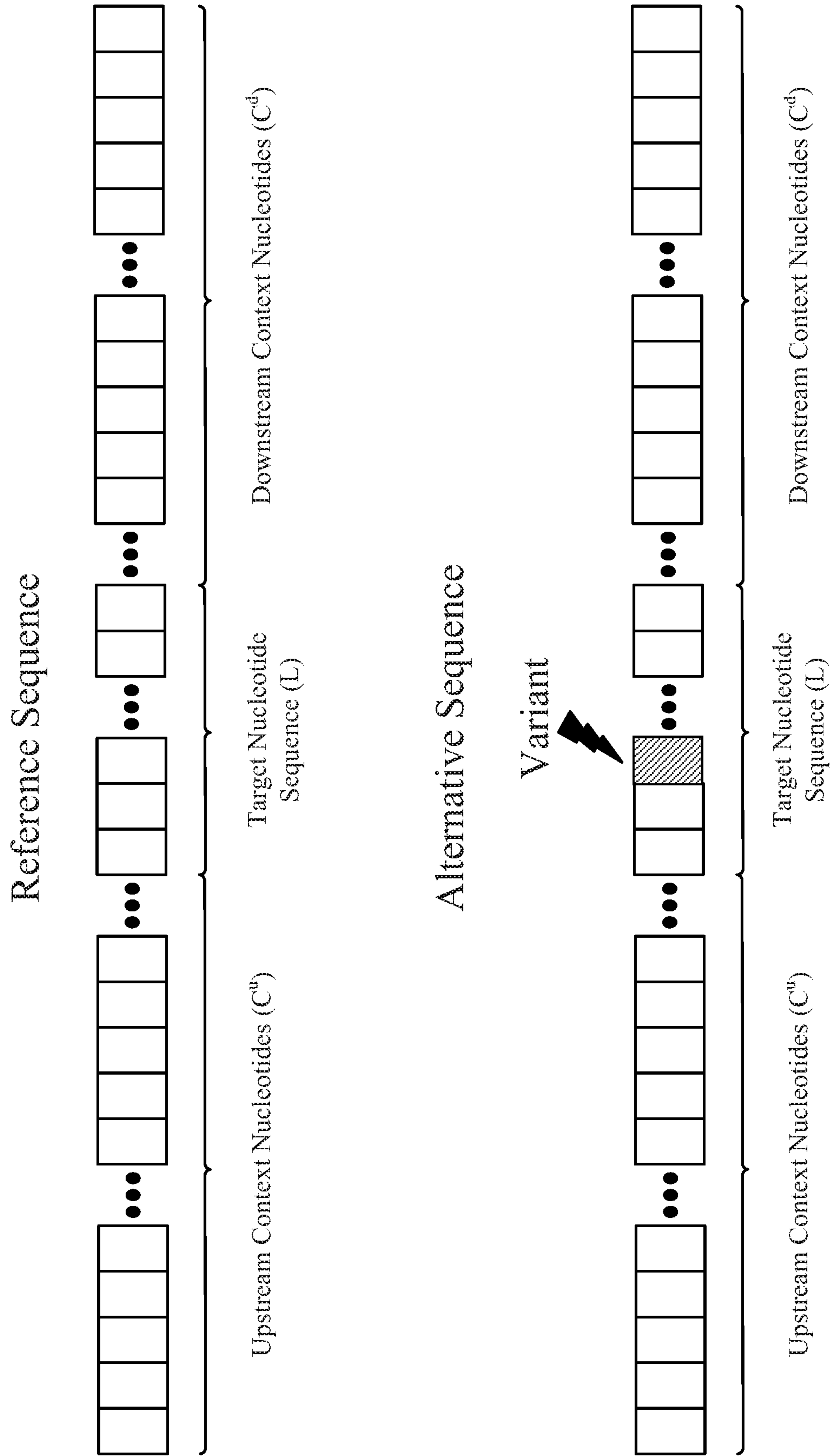


FIG. 32



C^u: 40, 200, 1000, 5000
C^d: 40, 200, 1000, 5000
L: 101

FIG. 33

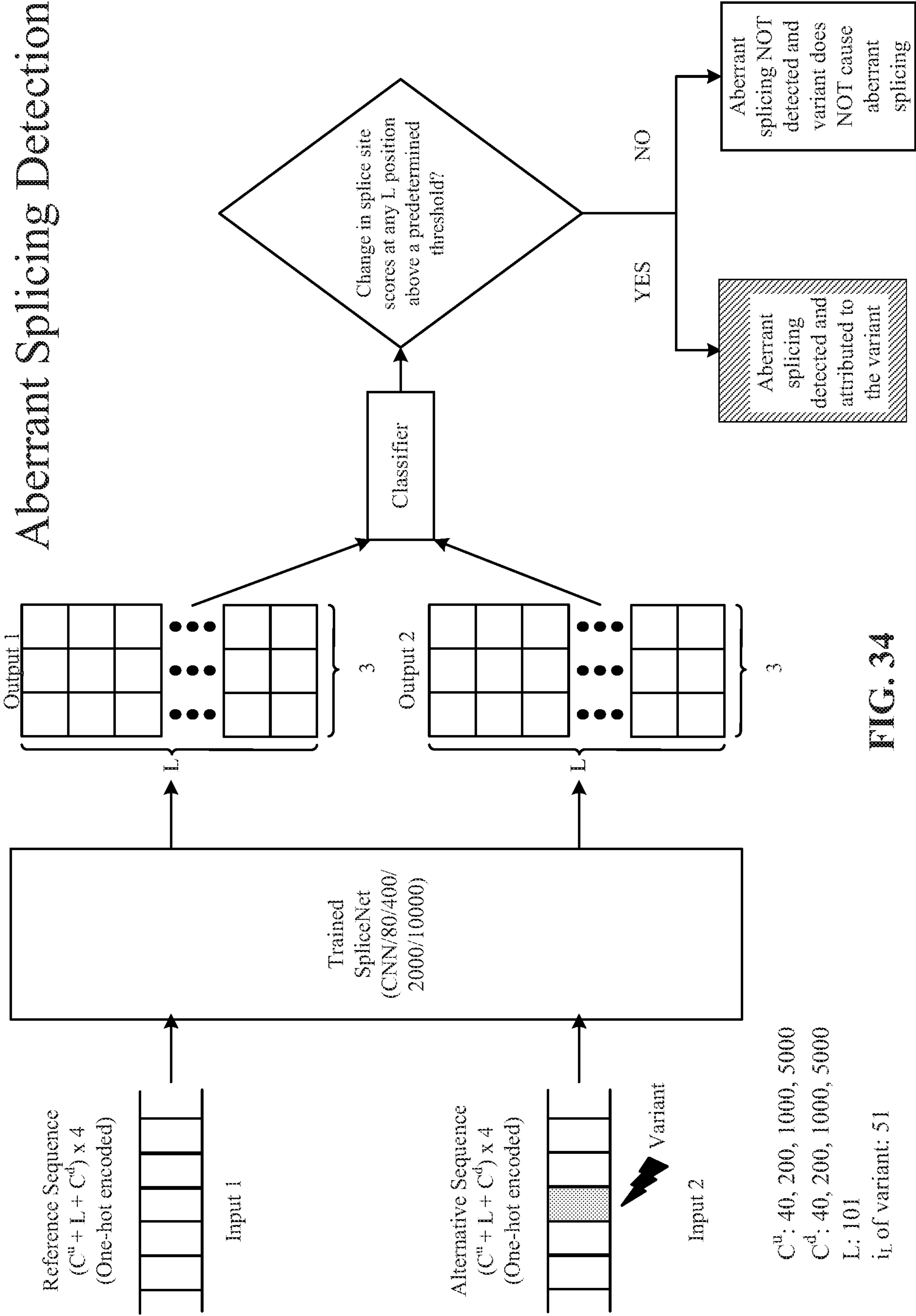


FIG. 34

SpliceNets 10000 Processing Pyramid for Splice Site Classification

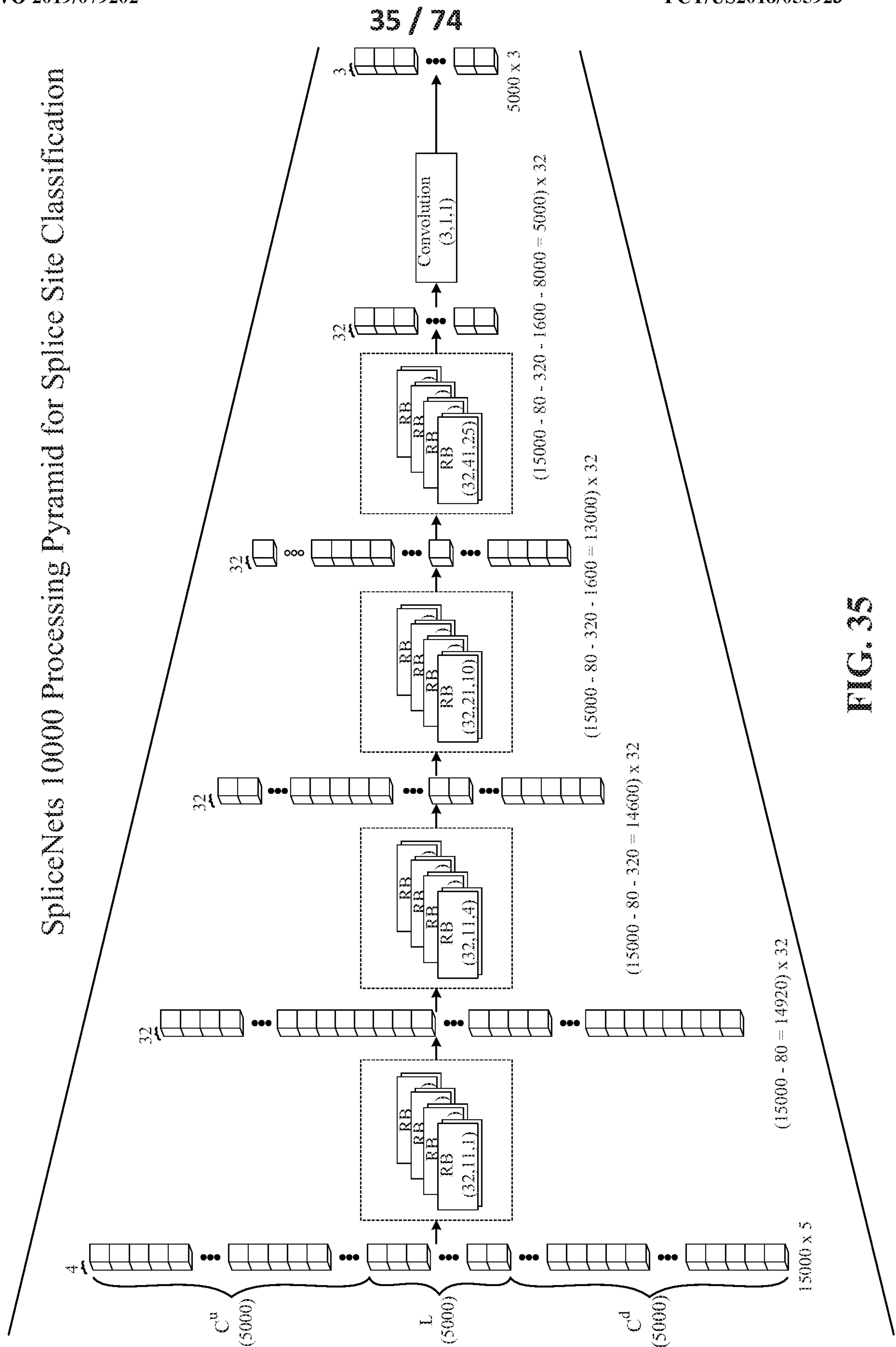


Fig. 35

SpliceNets 10000 Processing Pyramid for Aberrant Splicing Detection

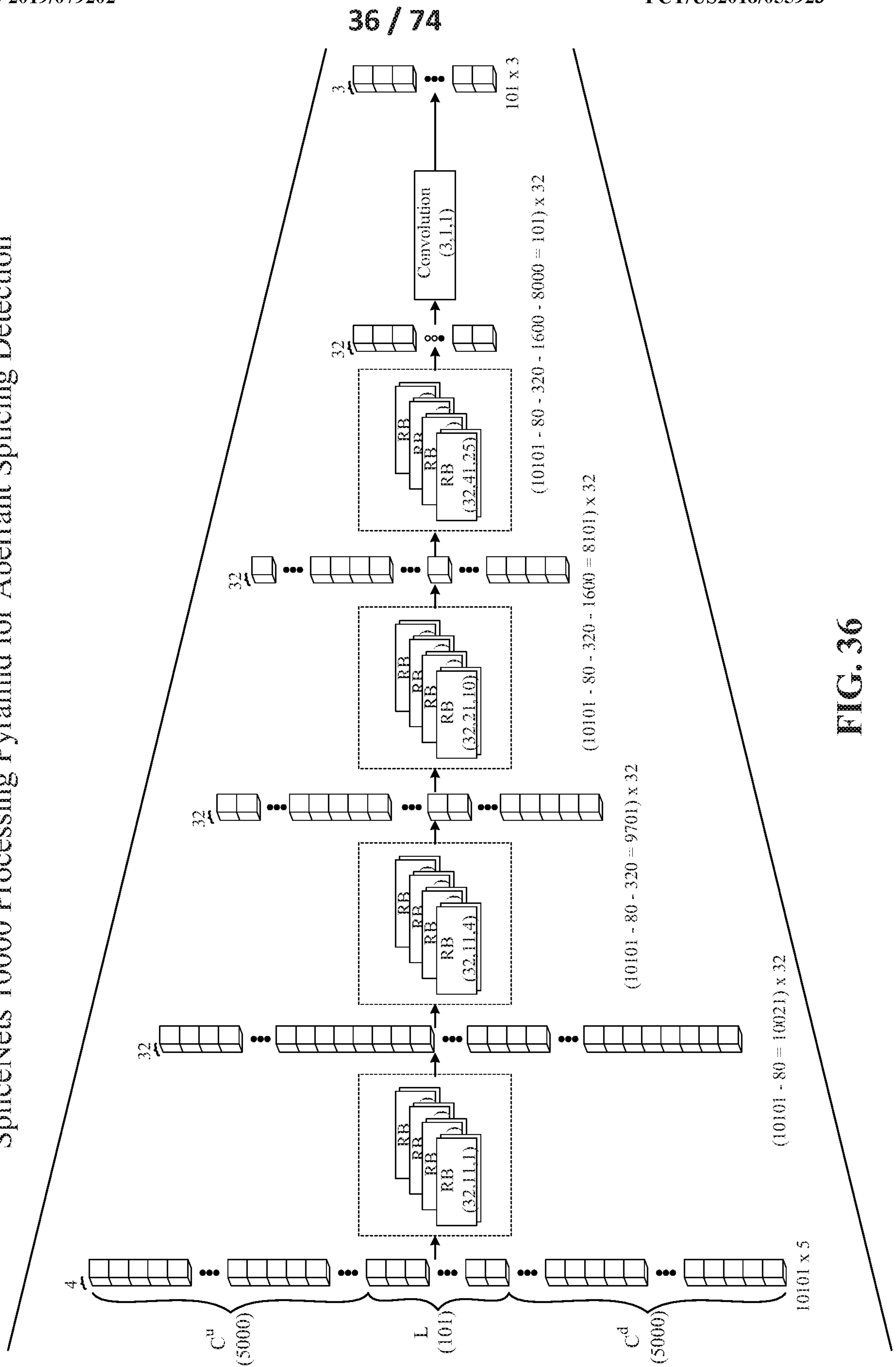


Fig. 3

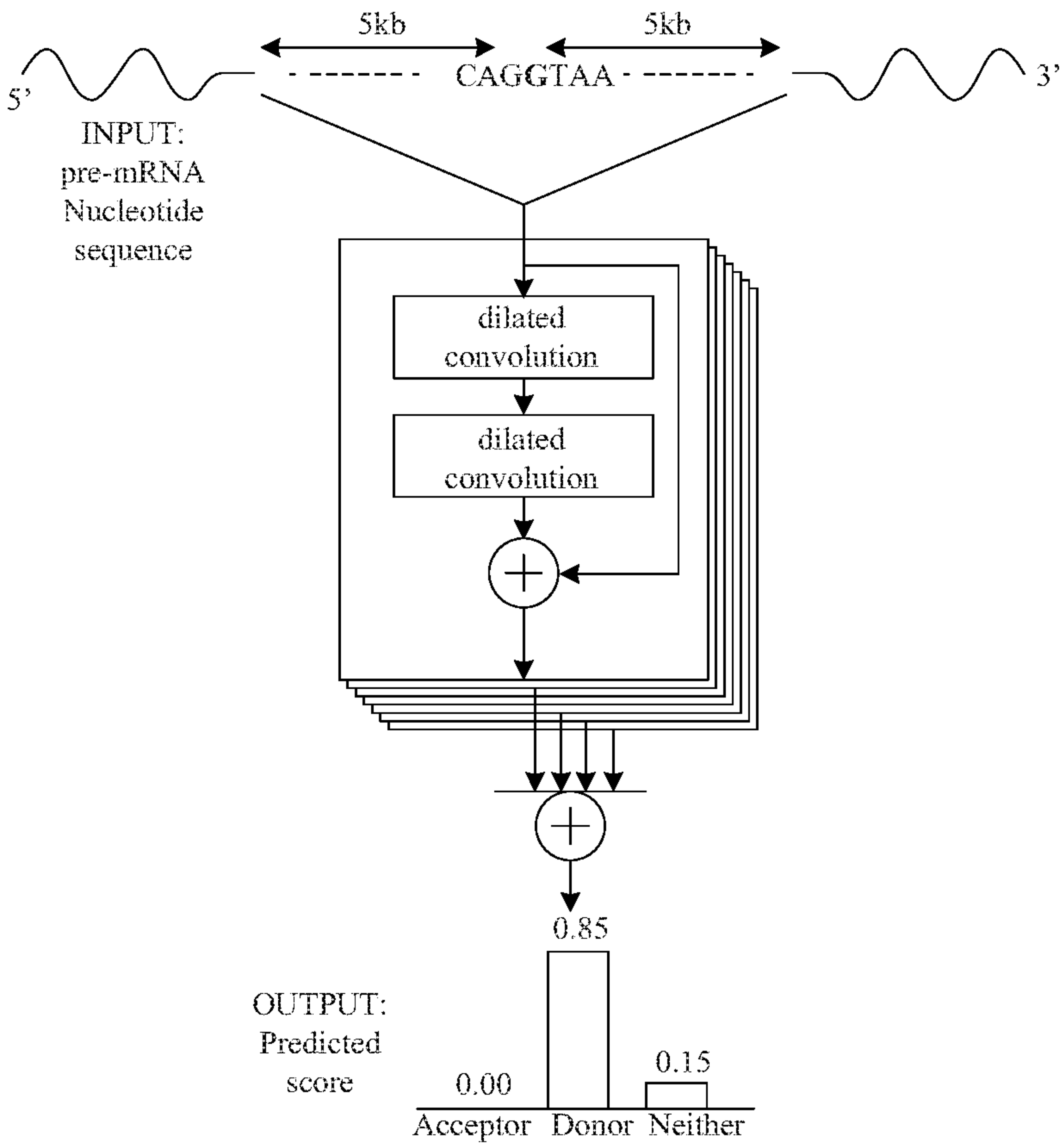


FIG. 37A

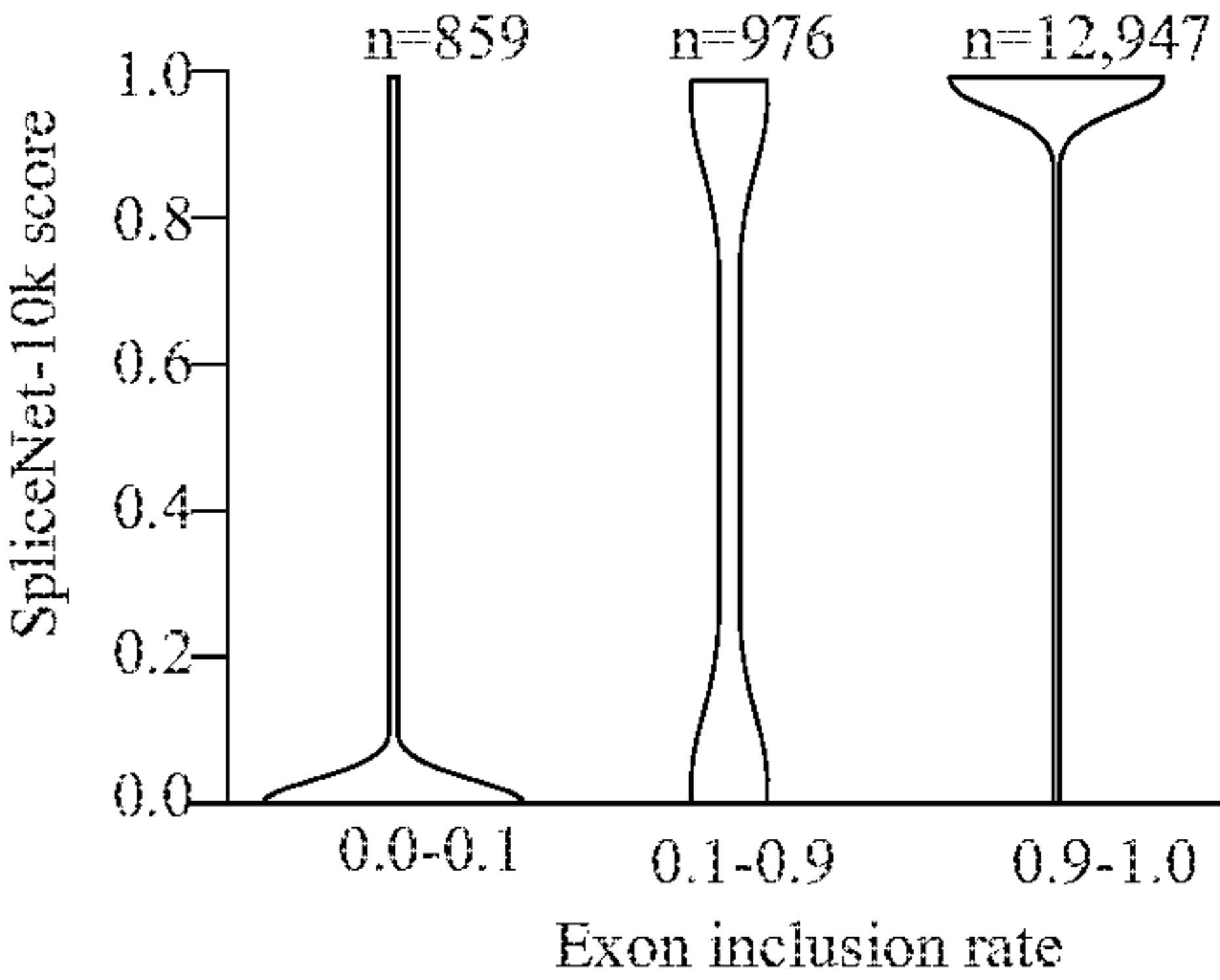


FIG. 37C

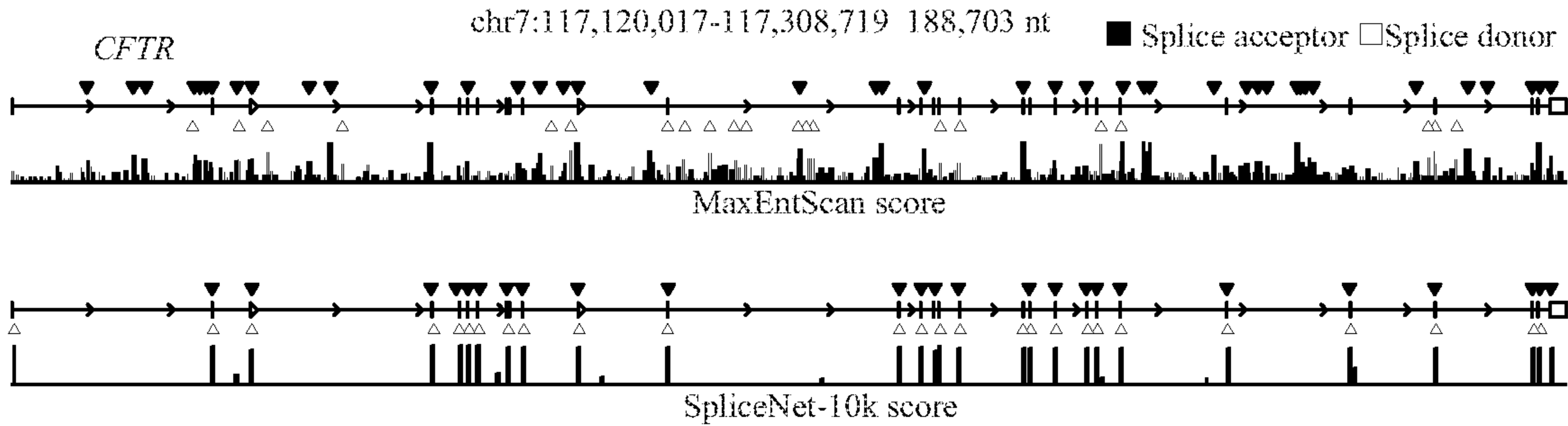


FIG. 37B

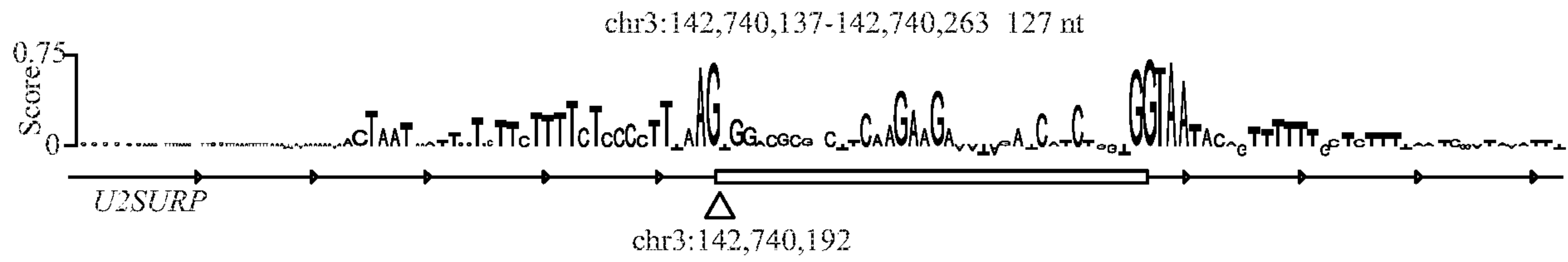


FIG. 37D

	Top- <i>k</i> accuracy	PR-AUC
SpliceNet-80nt	0.57	0.60
SpliceNet-400nt	0.90	0.95
SpliceNet-2k	0.93	0.97
SpliceNet-10k	0.95	0.98
GeneSplicer	0.30	0.23
MaxEntScan	0.22	0.15
NNSplice	0.22	0.15

FIG. 37E

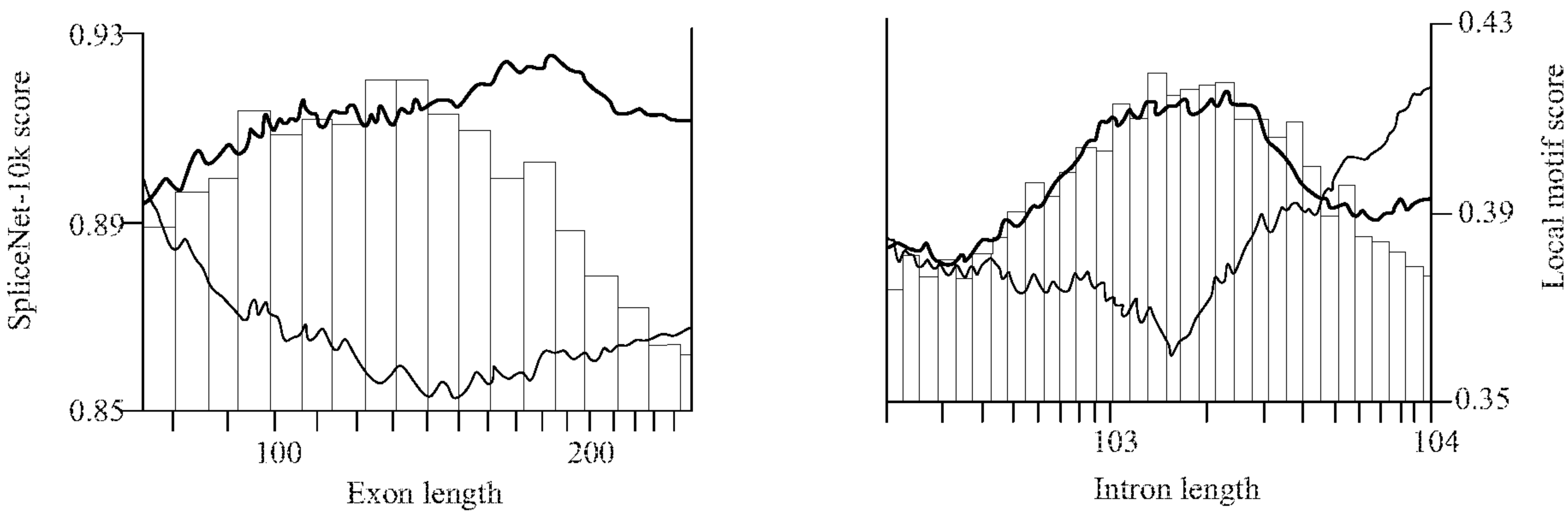


FIG. 37F

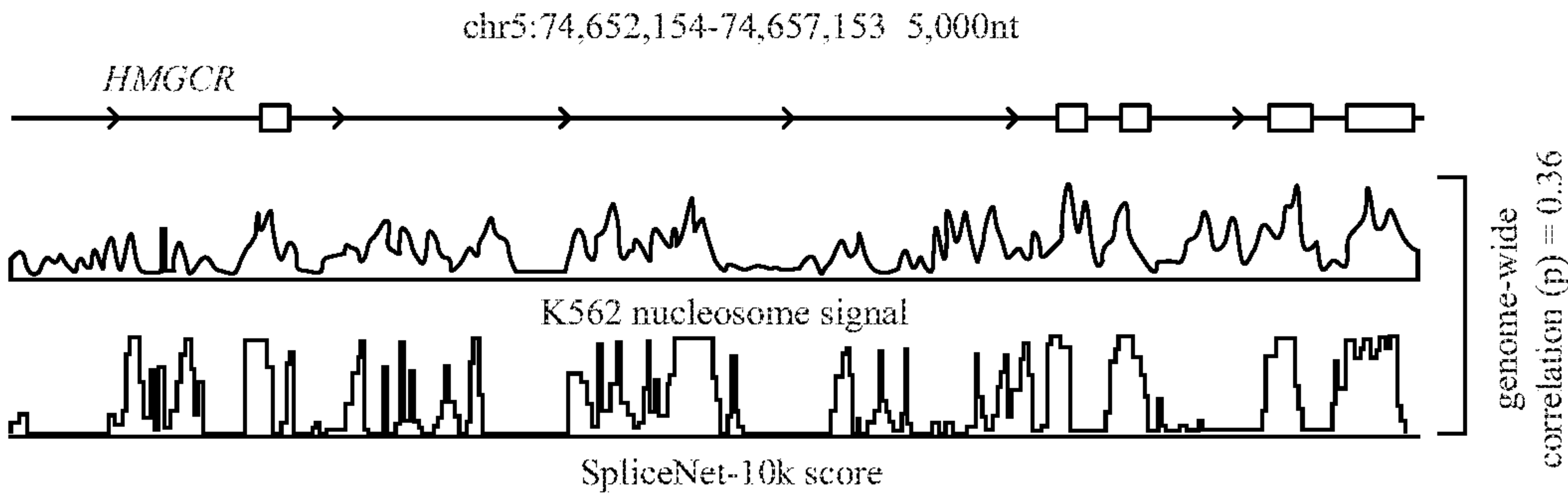


FIG. 37G

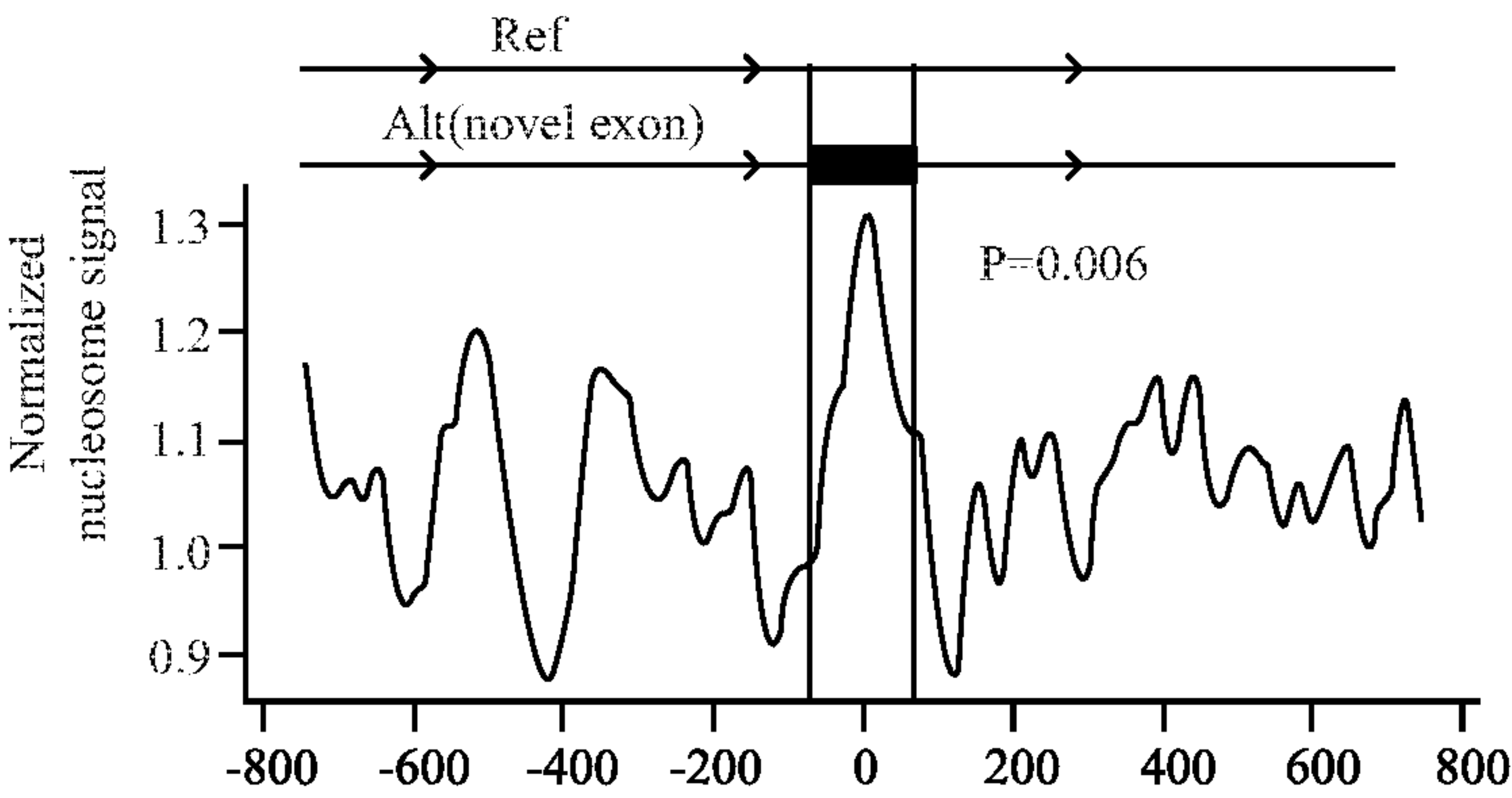


FIG. 37H

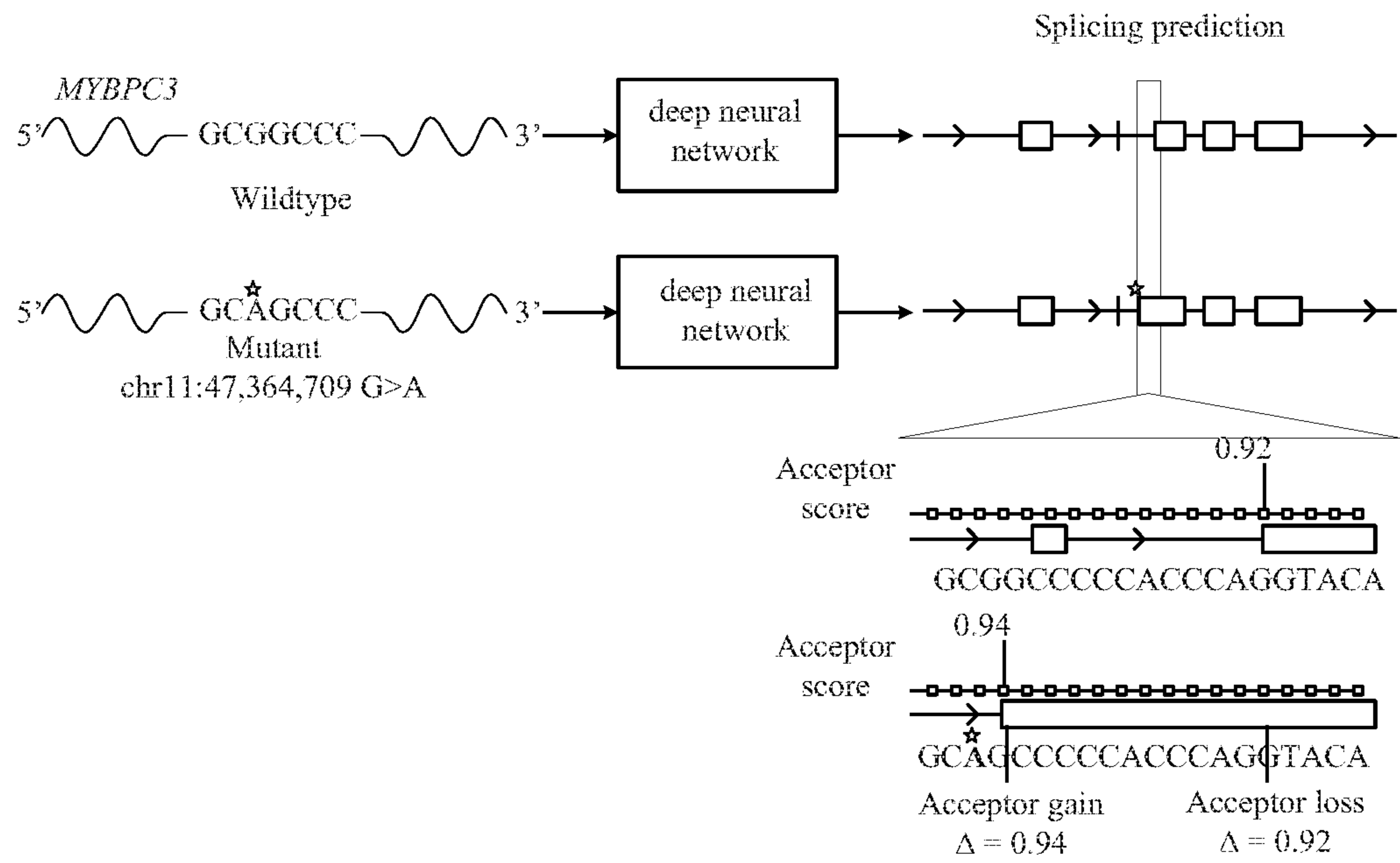


FIG. 38A

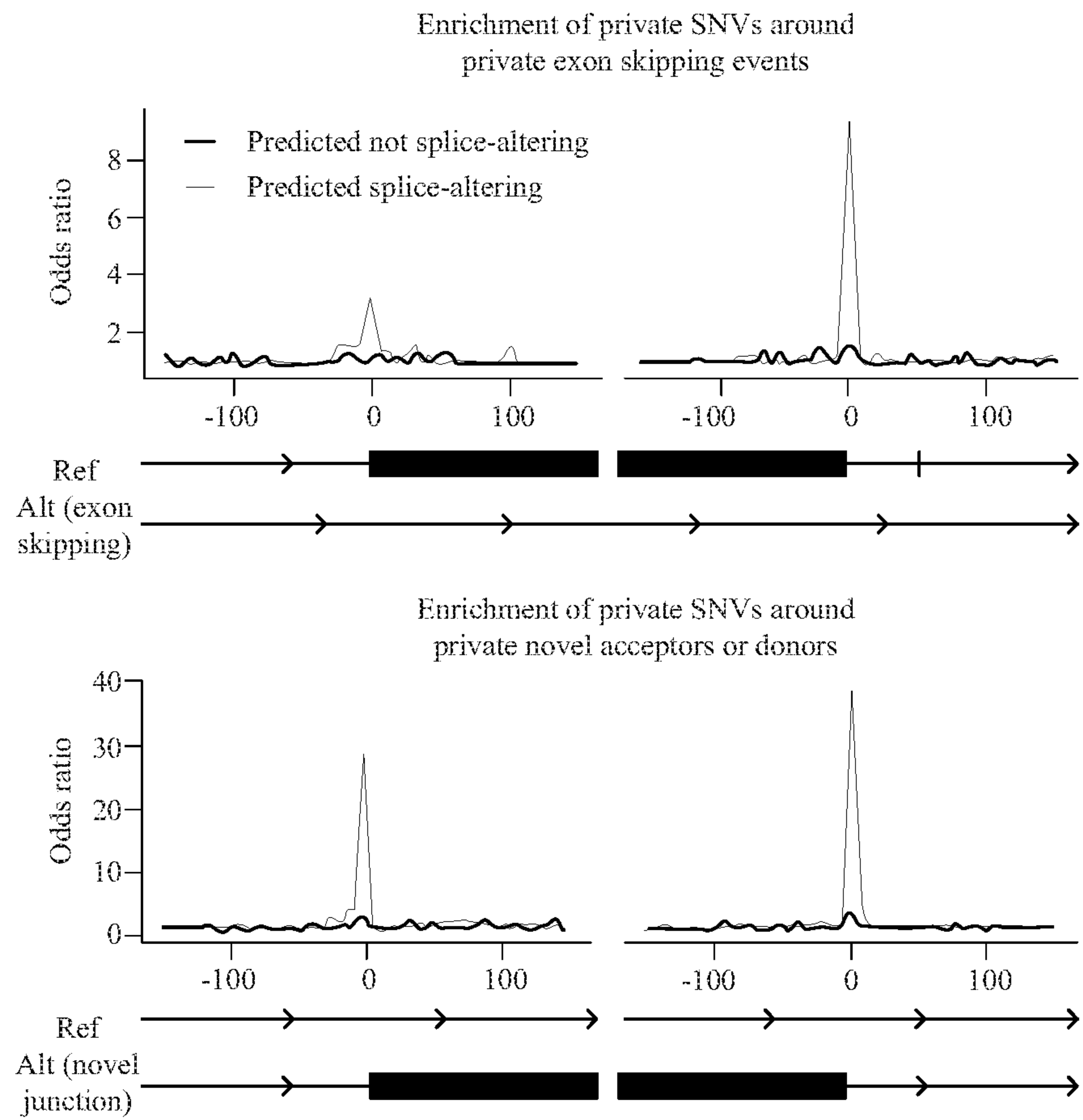
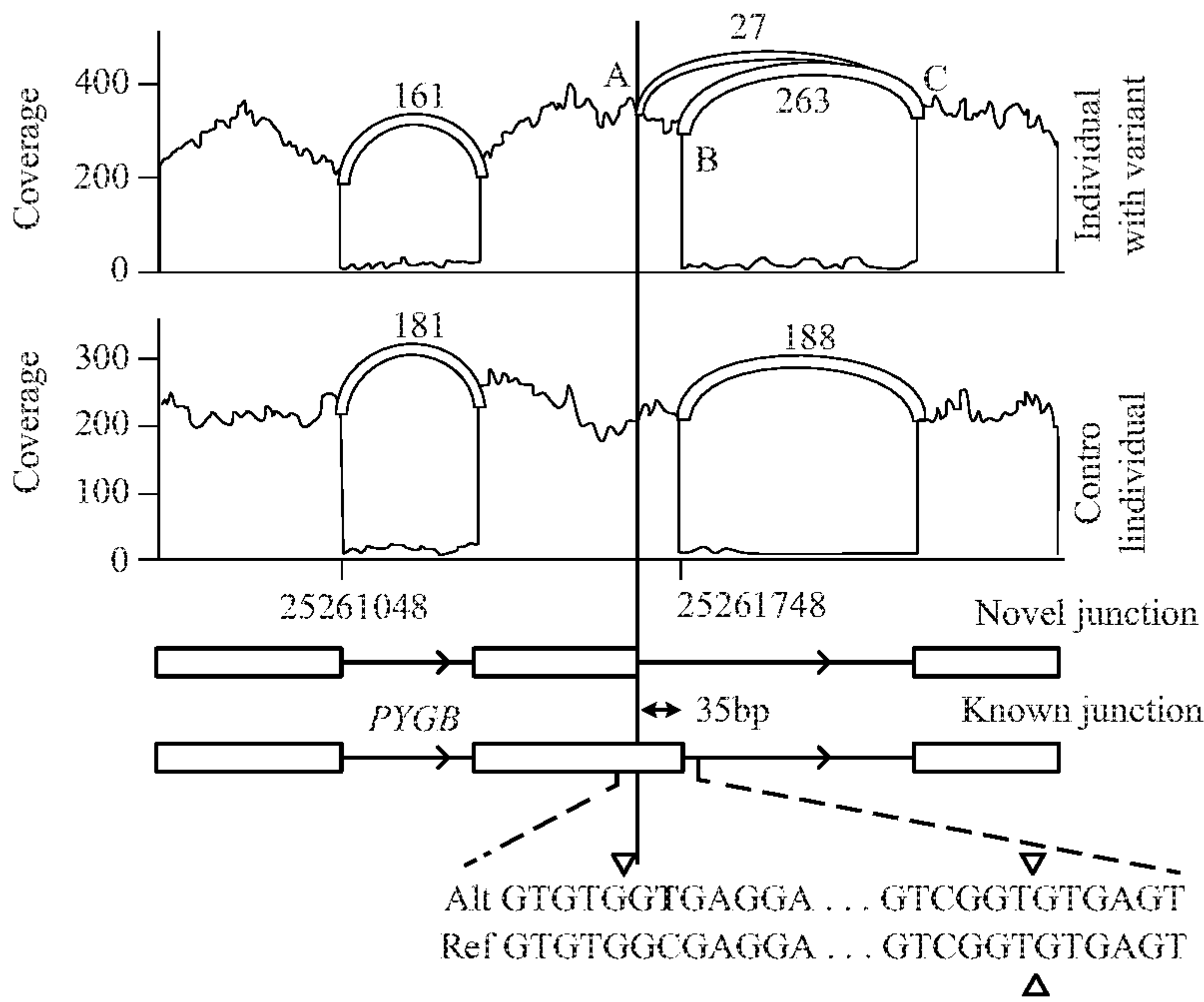


FIG. 38B

chr20:25261715 C>T Δ Score = 0.77



Relative usage of novel junction

$$\left(\frac{AC}{AC+BC}\right)_{mut} - \left(\frac{AC}{AC+BC}\right)_{ctrl} = \frac{27}{27+263} - 0 = 0.09$$

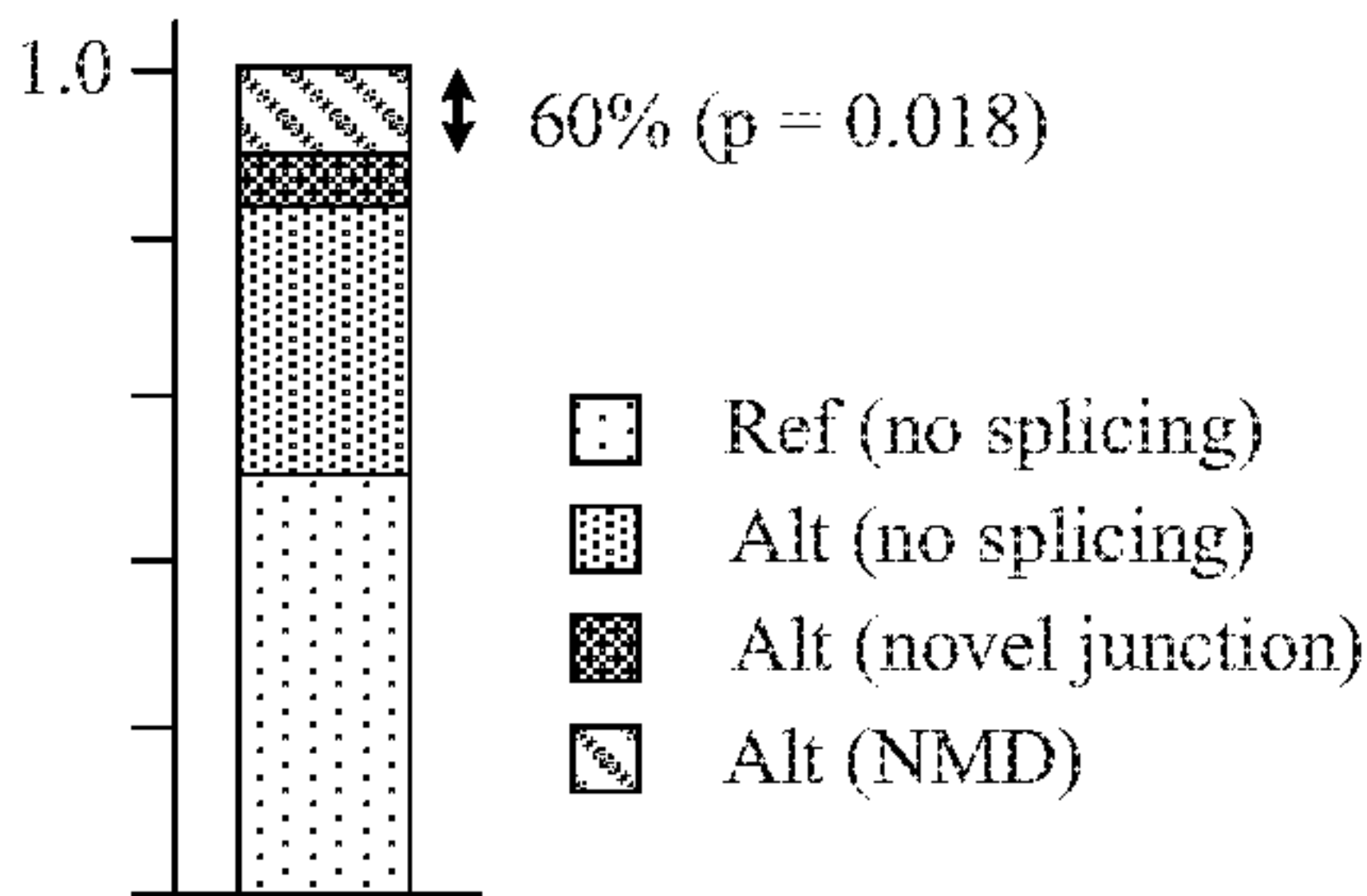


FIG. 38C

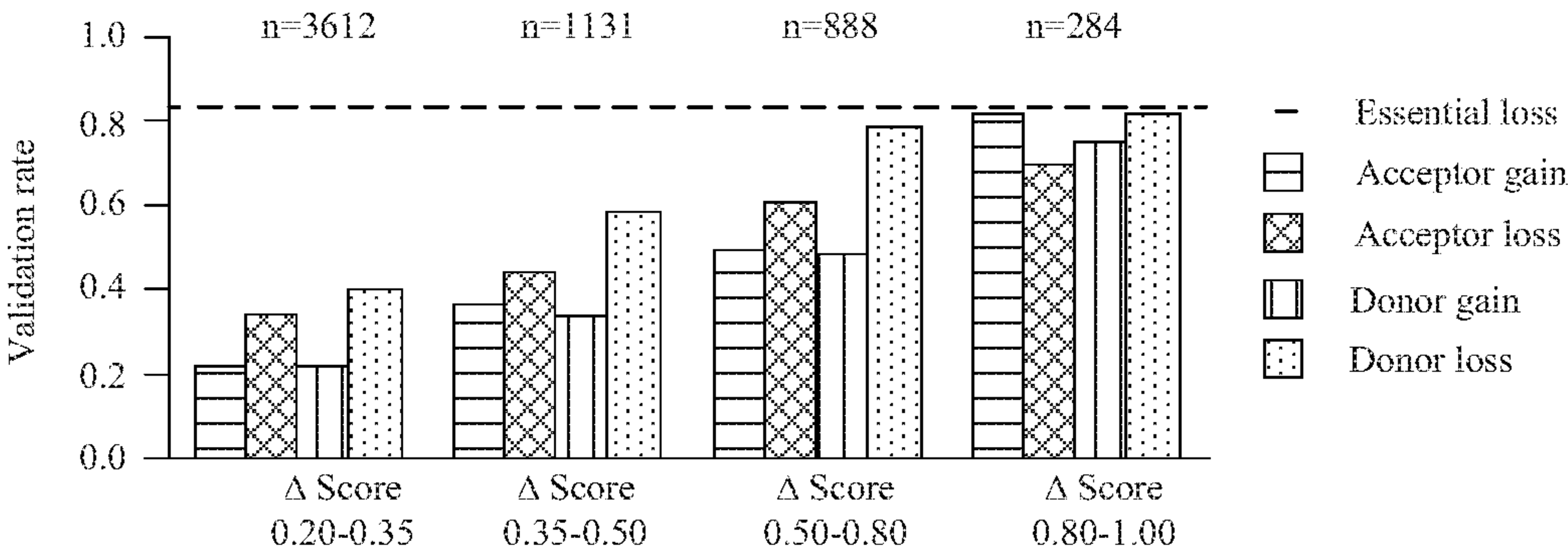


FIG. 38D

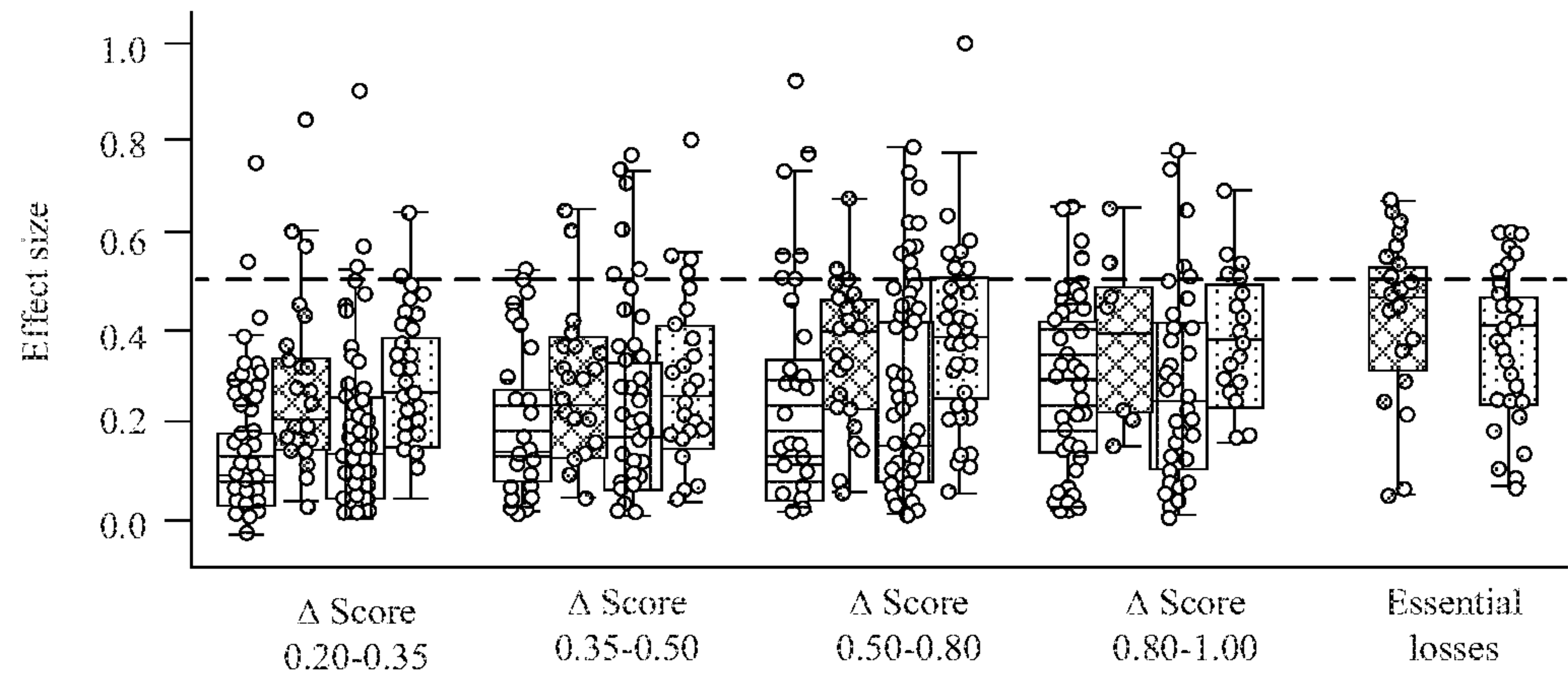


FIG. 38E

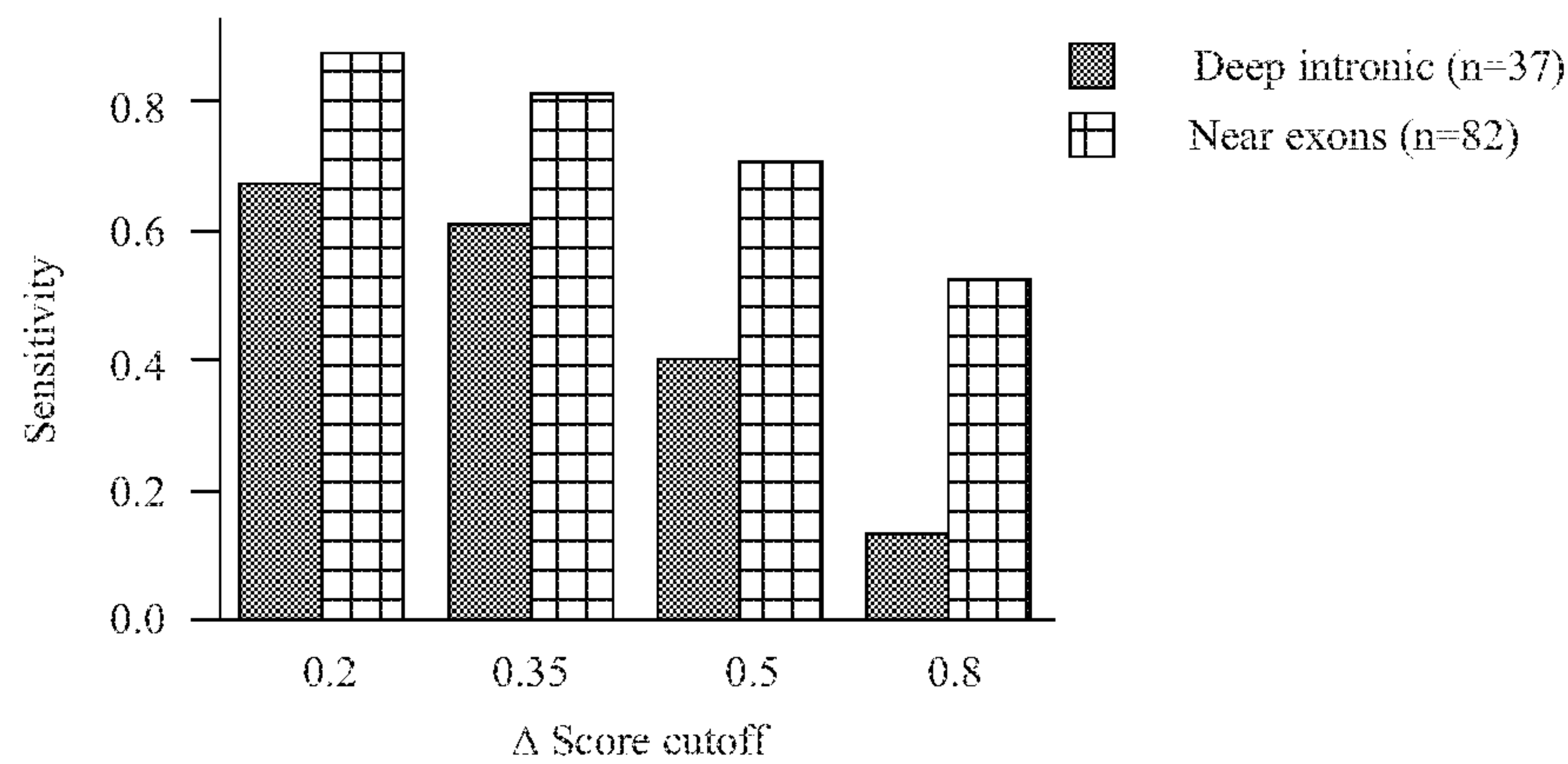


FIG. 38F

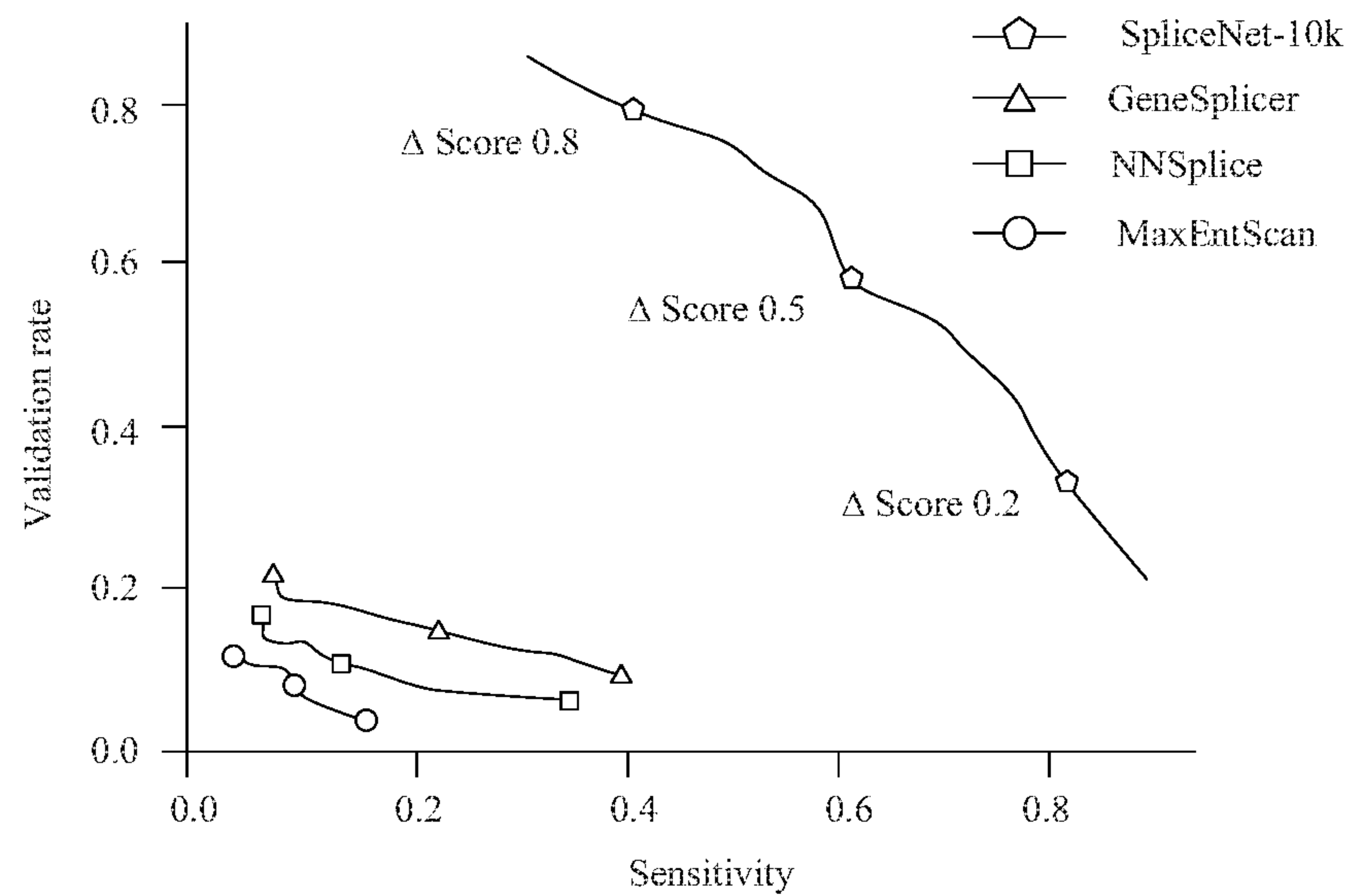


FIG. 38G

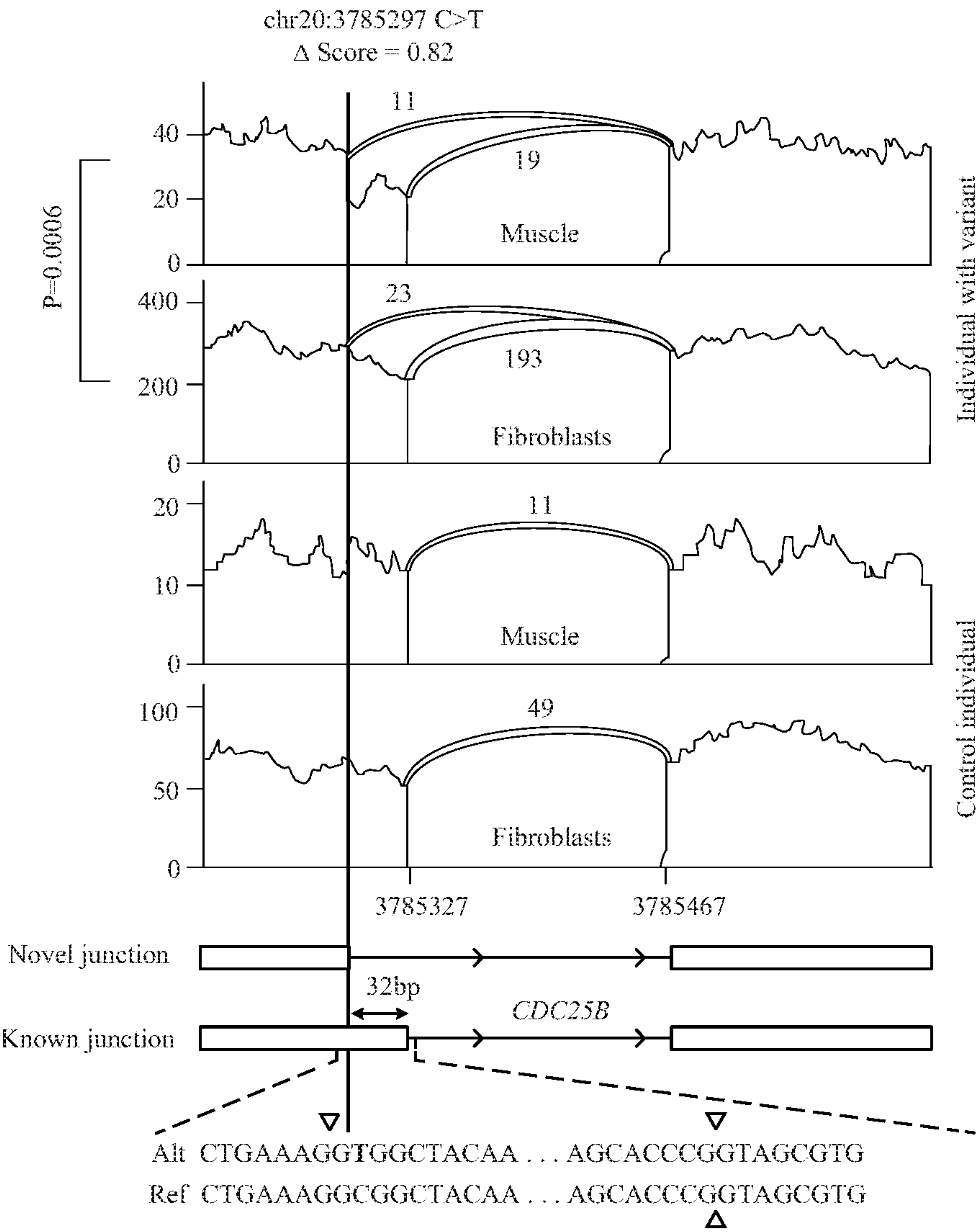


FIG. 39A

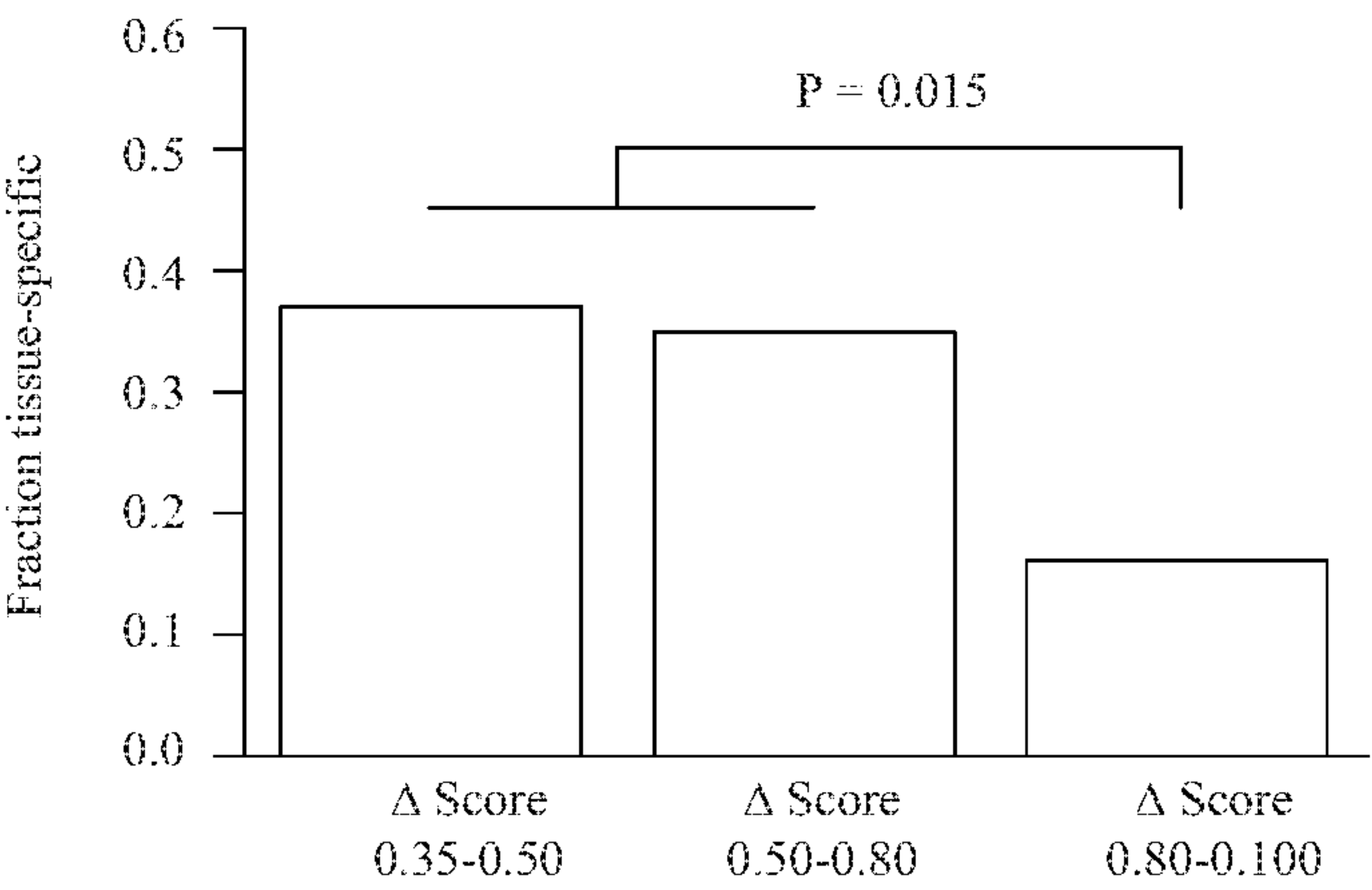


FIG. 39C

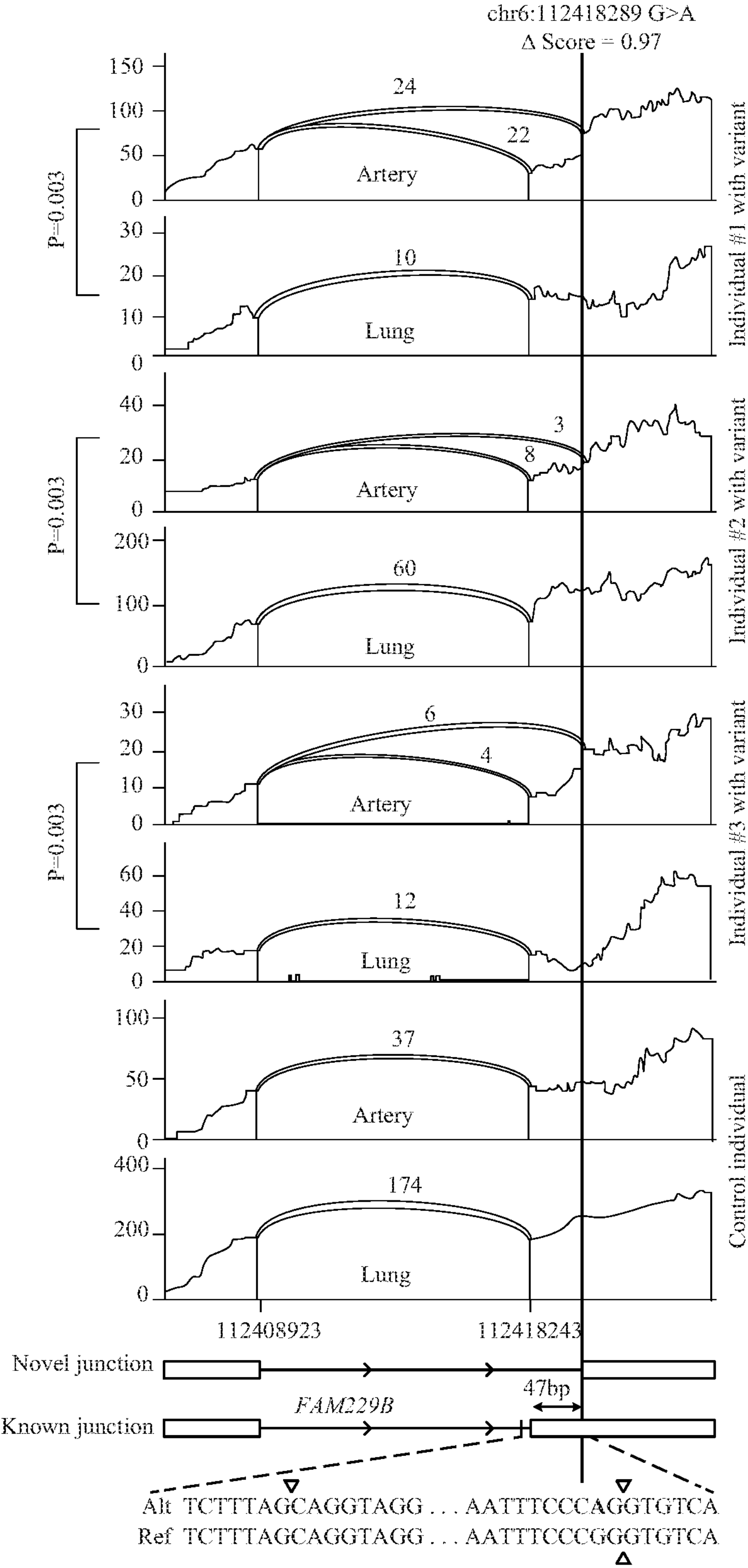


FIG. 39B

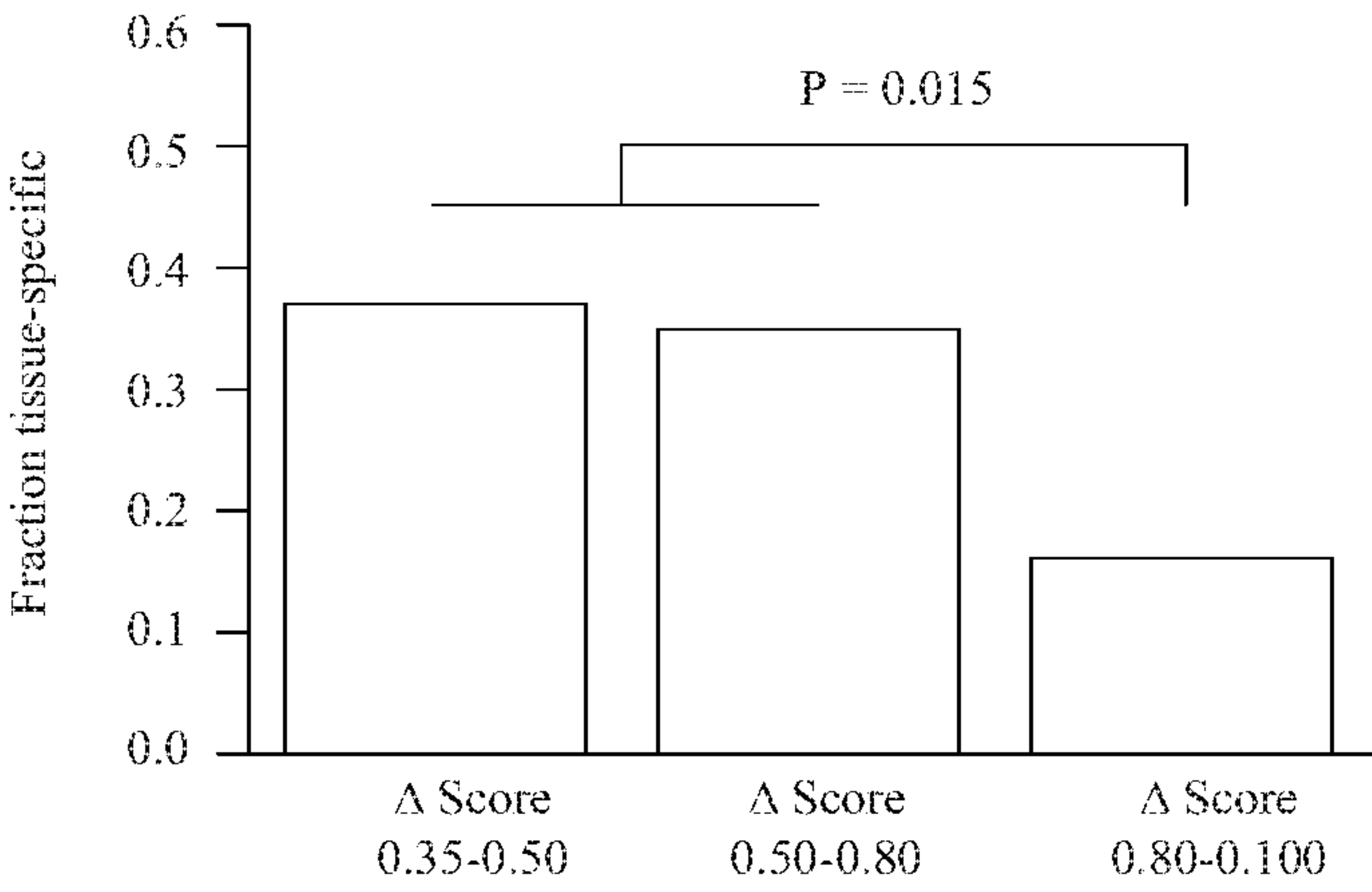


FIG. 39C

ExAC near exonic variants		
	Singleton	Common (AF ≥ 0.1%)
SNVs with Δ Score ≥ 0.8	10,369	212
SNVs with Δ Score < 0.1	1,687,004	158,177

Odds Ratio (OR) = 4.58 ($P < 10^{-127}$)
Fraction of deleterious variants = (1-1/OR)

FIG. 40A

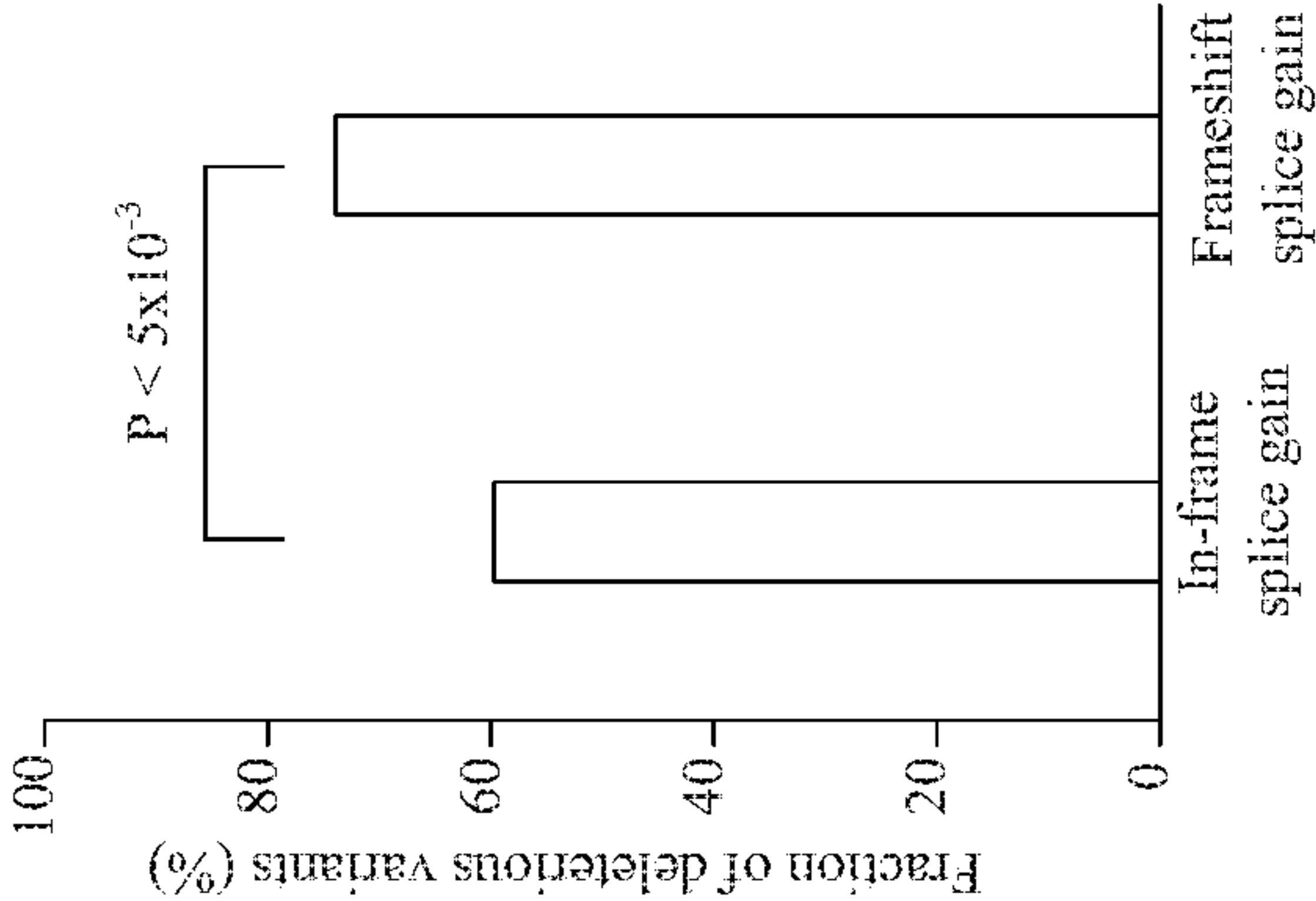
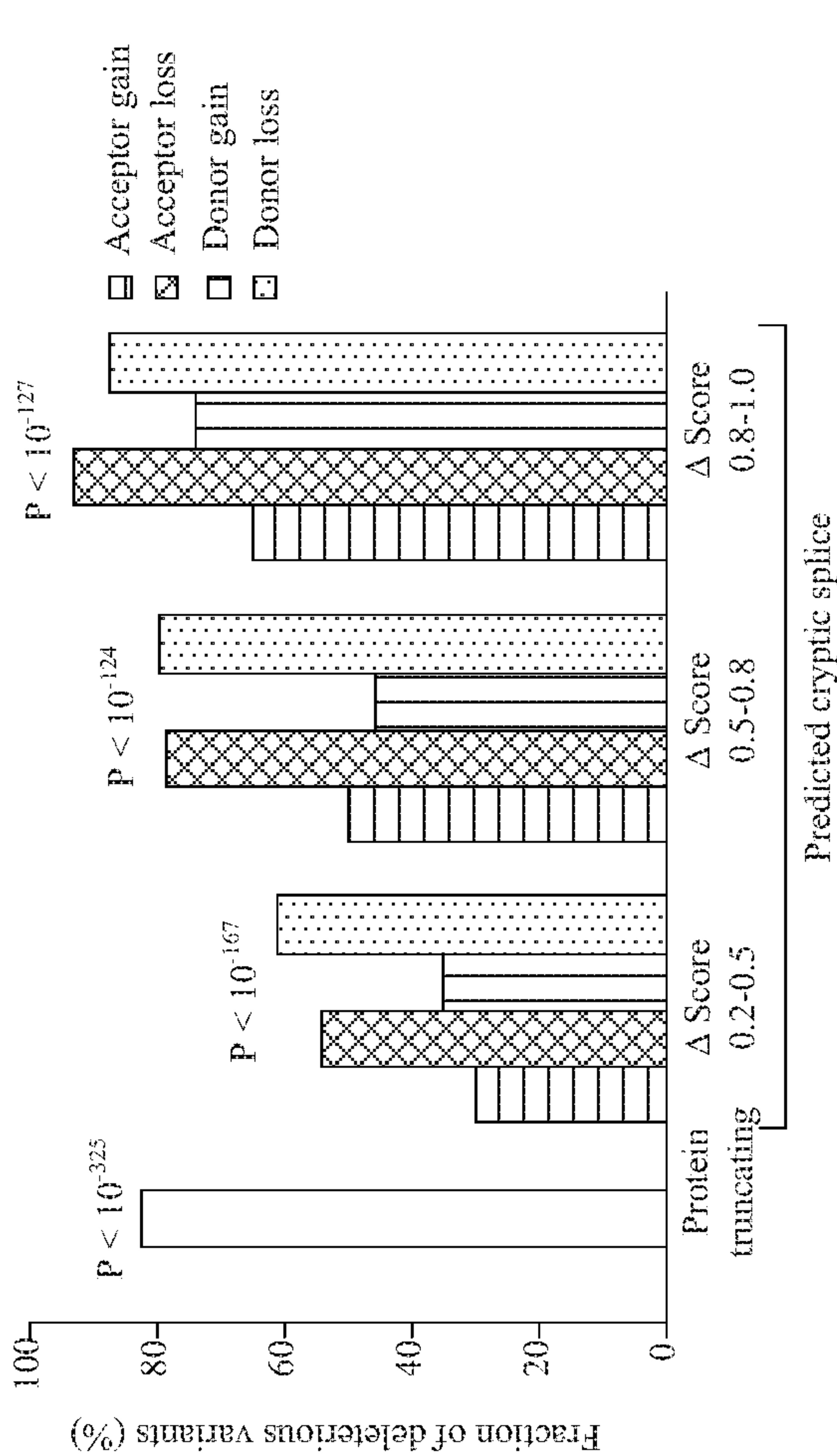


FIG. 40C

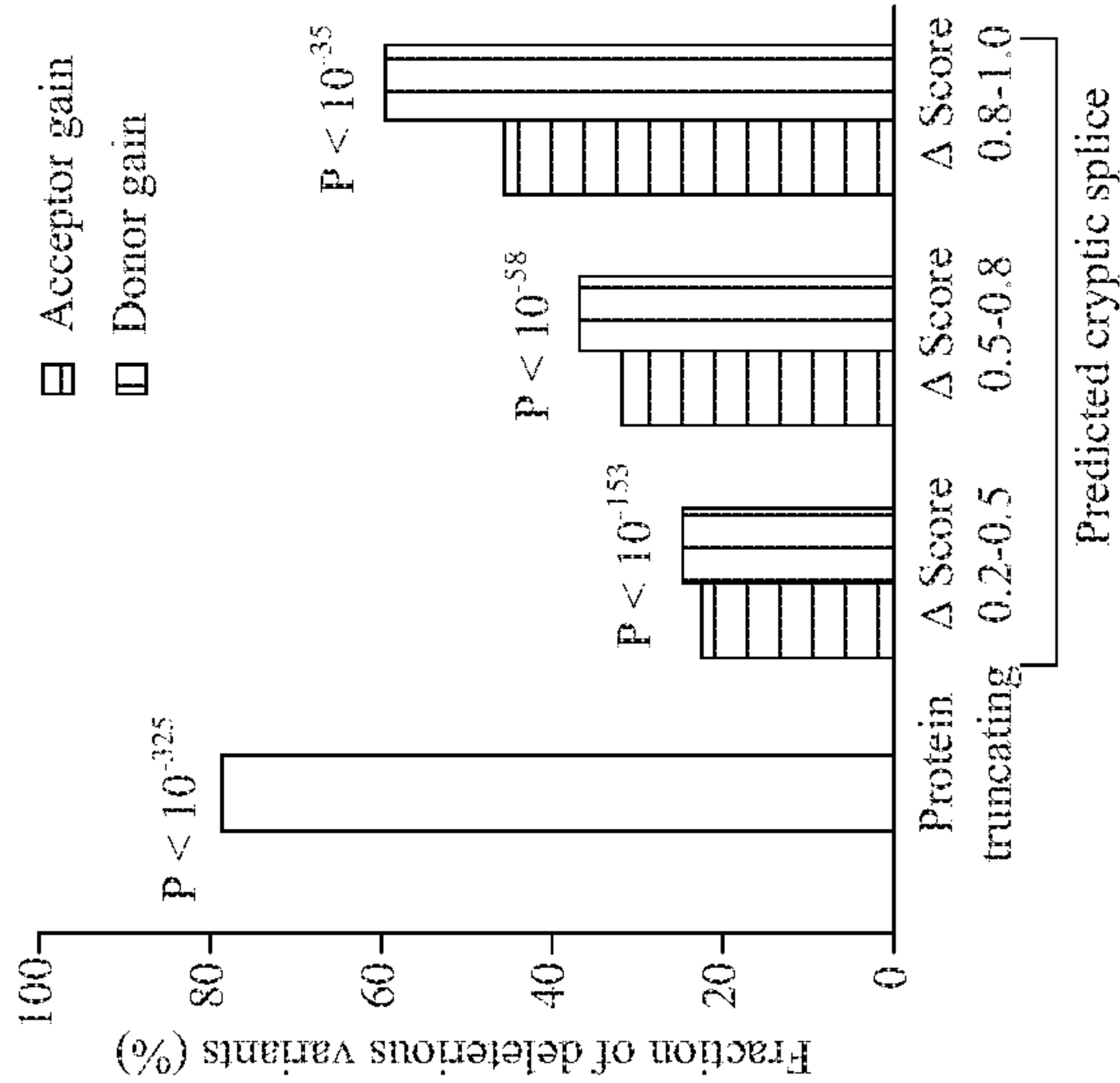


FIG. 40D

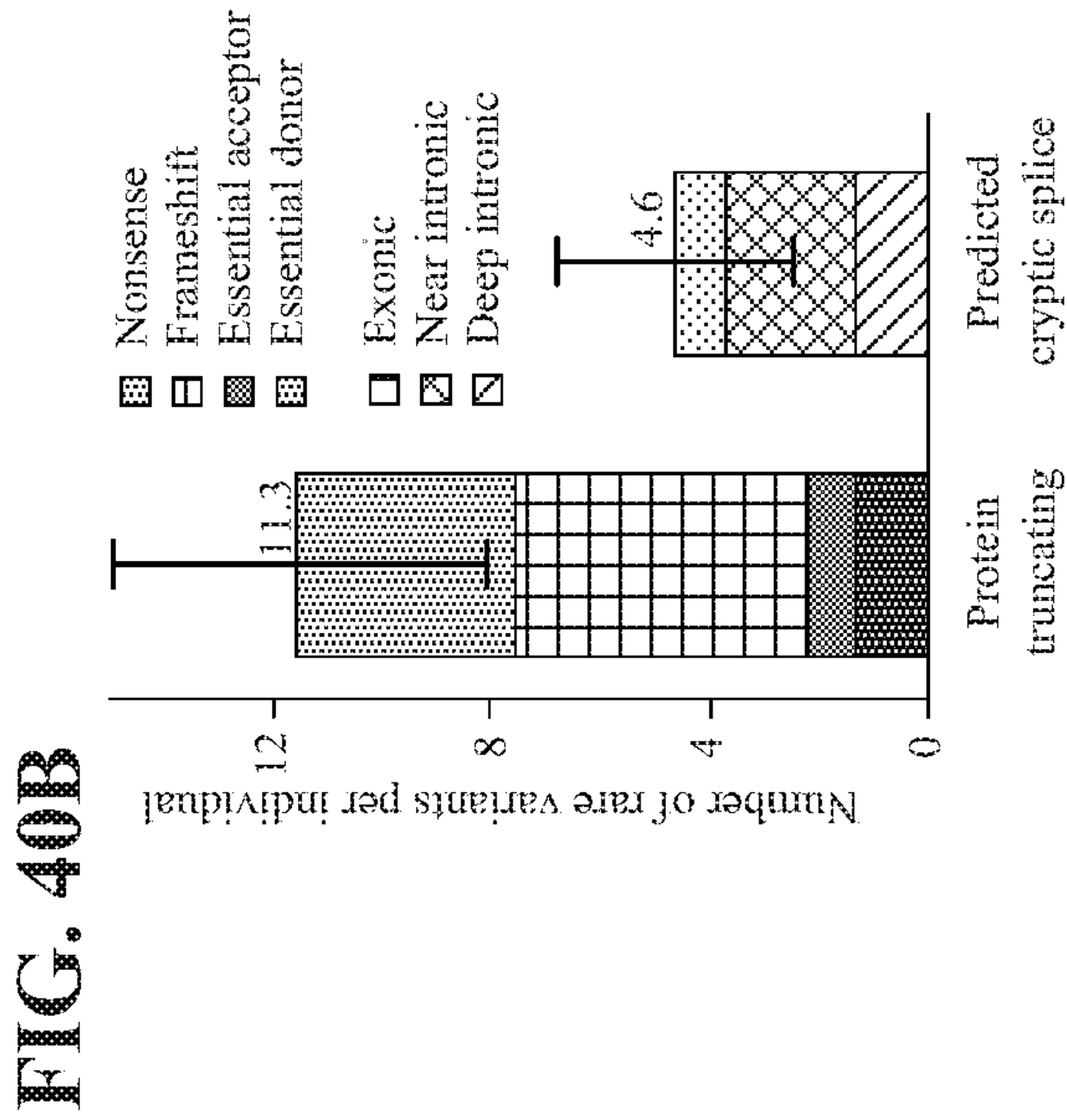


FIG. 40E

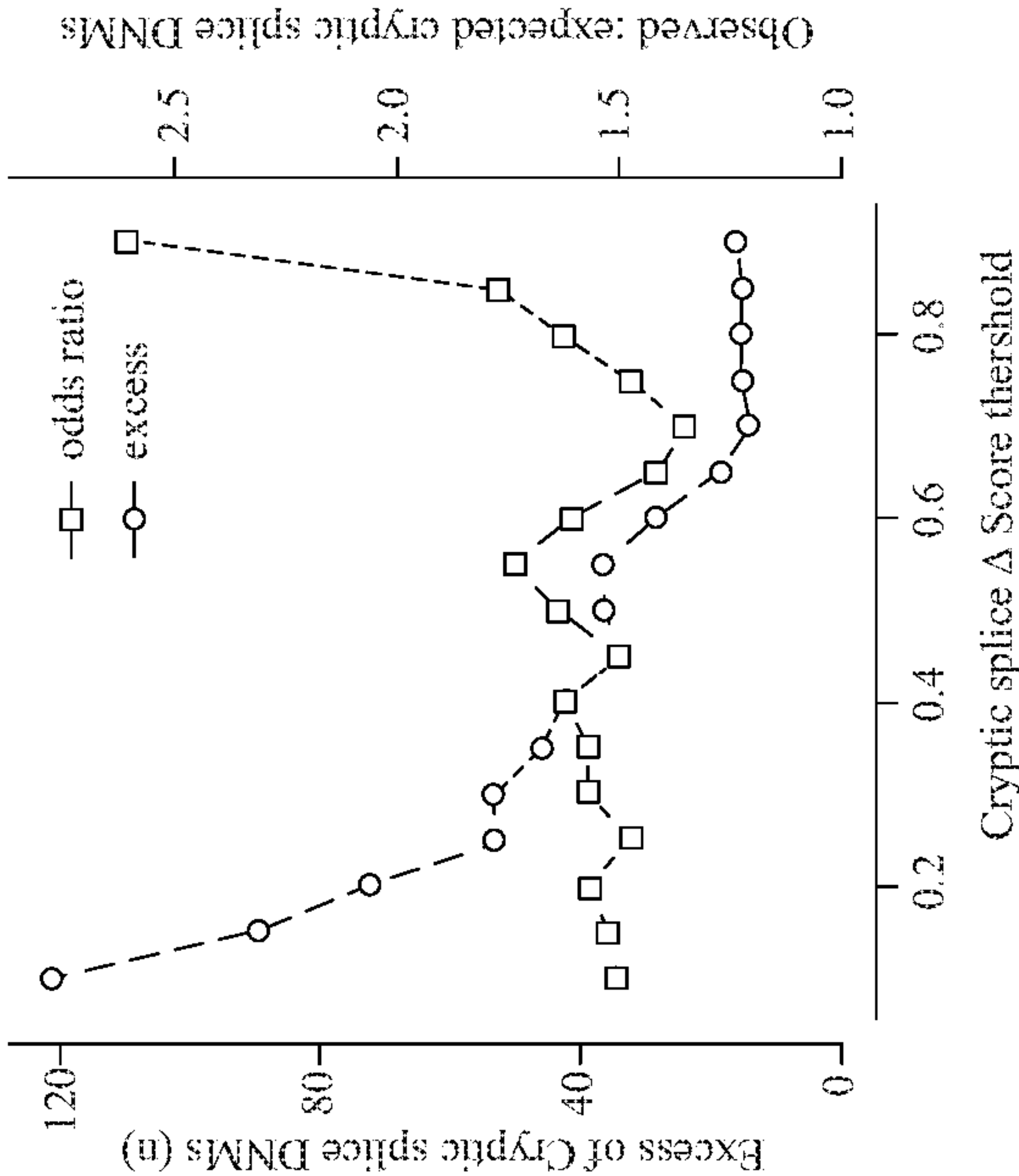


FIG. 41C

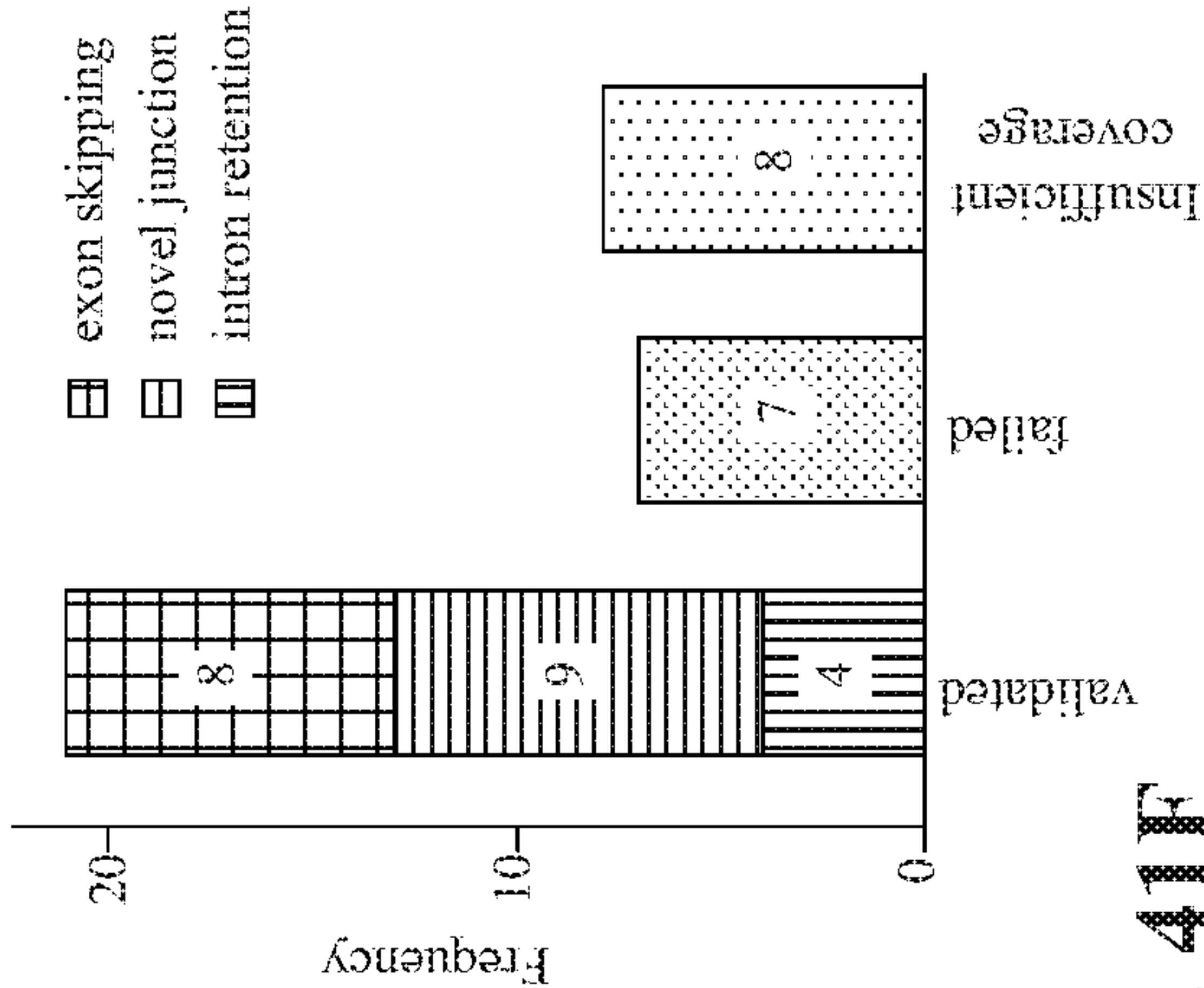


FIG. 41F

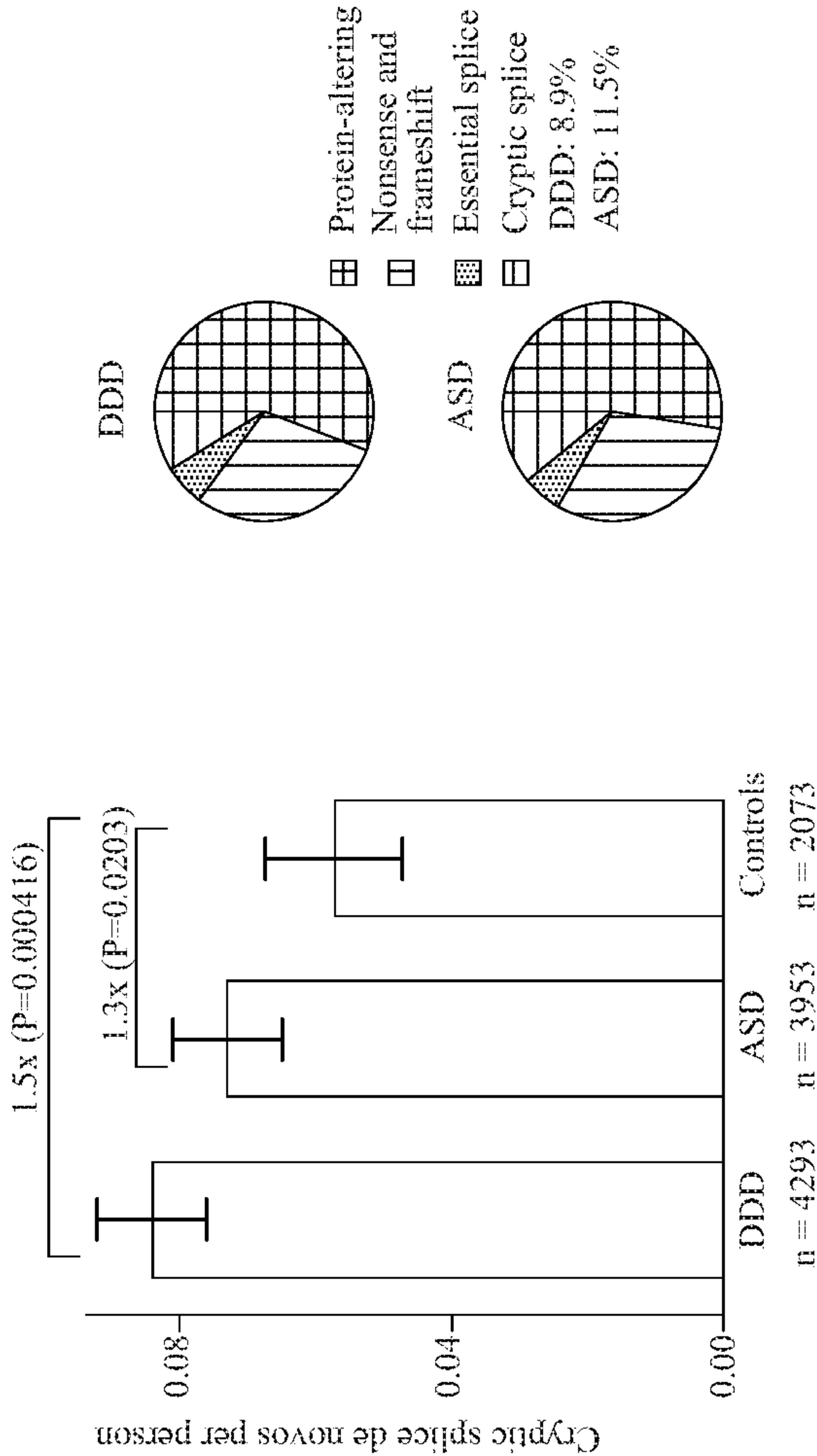


FIG. 41B

FIG. 41A

Symbol	PTV	Missense	Cryptic	P	Phenotypes
<i>CTNND1</i>	2	0	1	4.7 X 10 ⁻⁵	developmental delay
<i>HNRNPK</i>	2	1	1	9.7 X 10 ⁻⁶	developmental delay, scoliosis, orofacial cleft, vesicoureteral reflux
<i>PHF5A</i>	1	0	1	3.1 X 10 ⁻⁵	developmental delay
<i>SNAP25</i>	1	0	2	6.8 X 10 ⁻⁵	absence seizures, epicanthus
<i>TRIP12</i>	3	2	1	8.5 X 10 ⁻⁶	delayed speech
<i>KDM6B</i>	3	1	1	4.6 X 10 ⁻⁶	autism
<i>SLC6A8</i>	2	0	1	5.8 X 10 ⁻⁶	autism

FIG. 41D

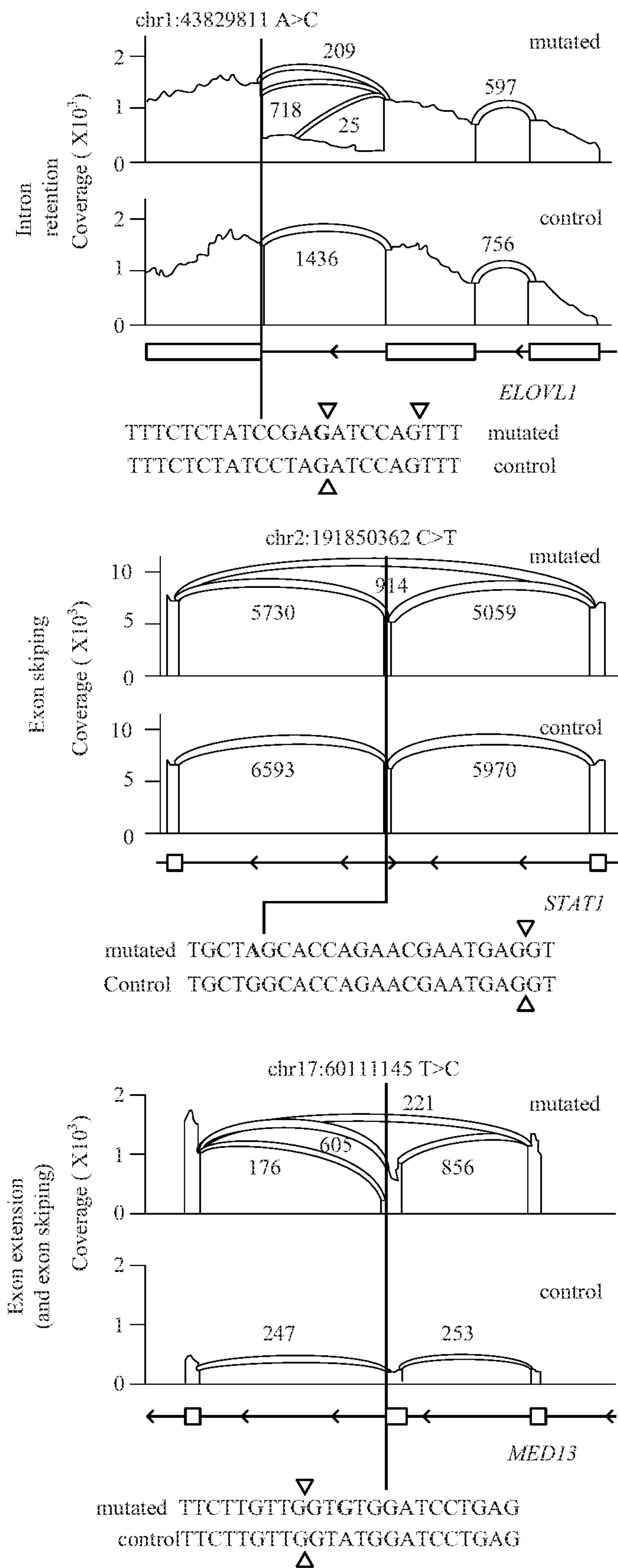


FIG. 41E

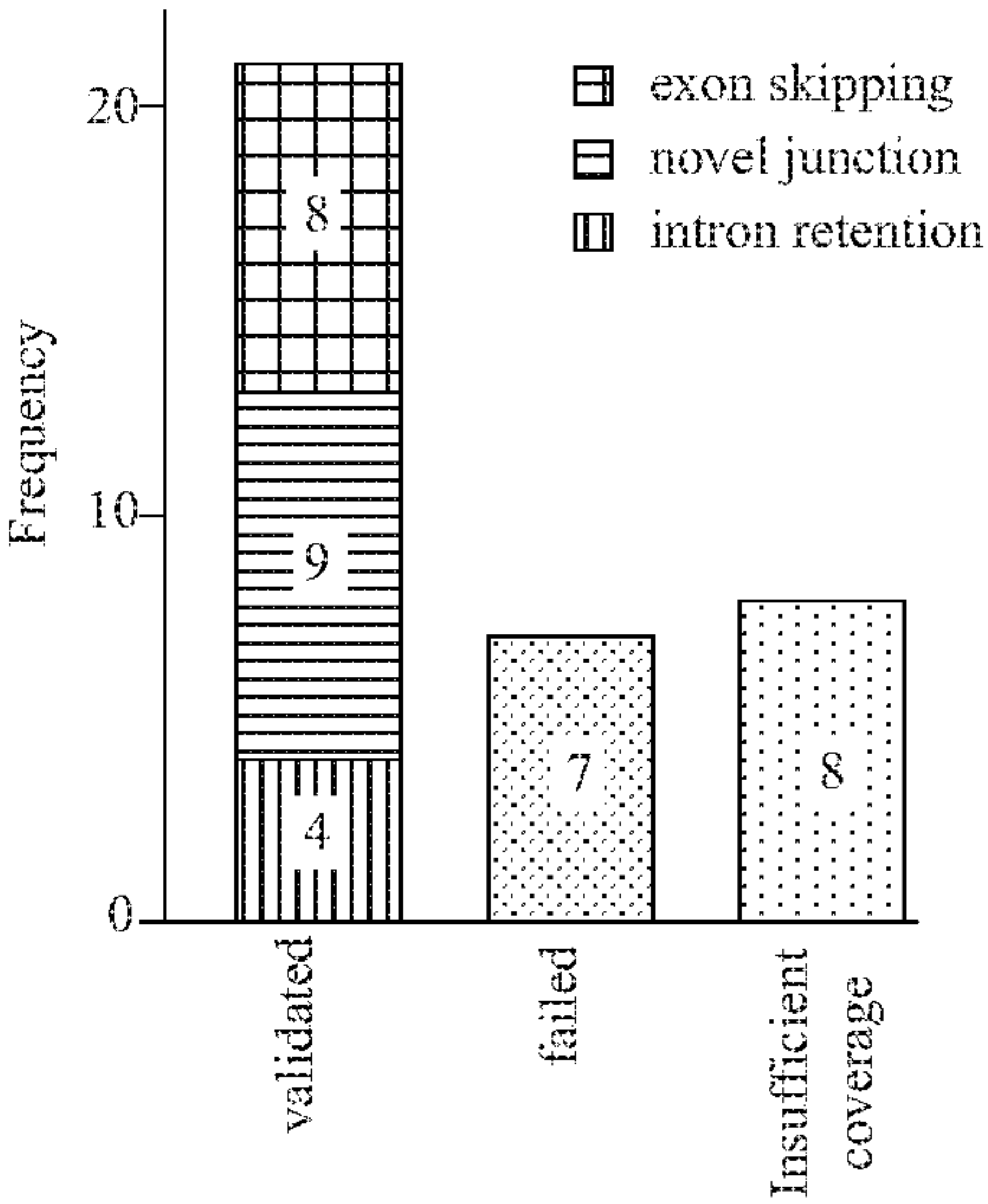


FIG. 41F

	Top- <i>k</i> accuracy		Precision-Recall AUC	
	Acceptor	Donor	Acceptor	Donor
SpliceNet-80nt	0.5129	0.5110	0.5195	0.5273
SpliceNet-400nt	0.6848	0.7221	0.7320	0.7982
SpliceNet-2k	0.8090	0.8405	0.8742	0.9047
SpliceNet-10k	0.8223	0.8586	0.8896	0.9107
GeneSplicer	0.3238	0.3381	0.2234	0.2982
MaxEntScan	0.2932	0.3687	0.2233	0.2937
NNSplice	0.2655	0.3687	0.1702	0.3003

FIG. 42A

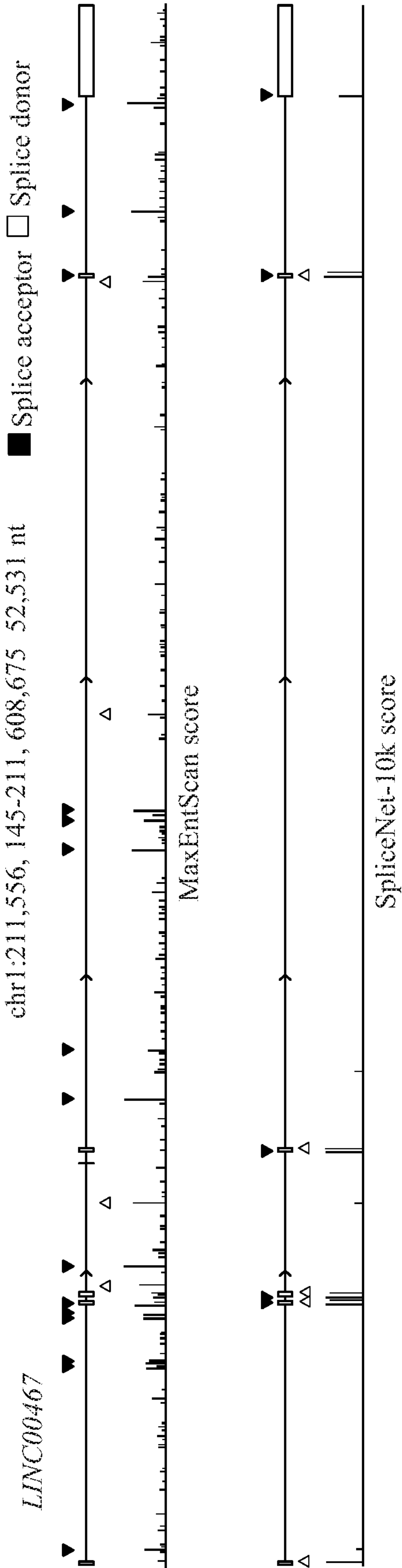


FIG. 42B

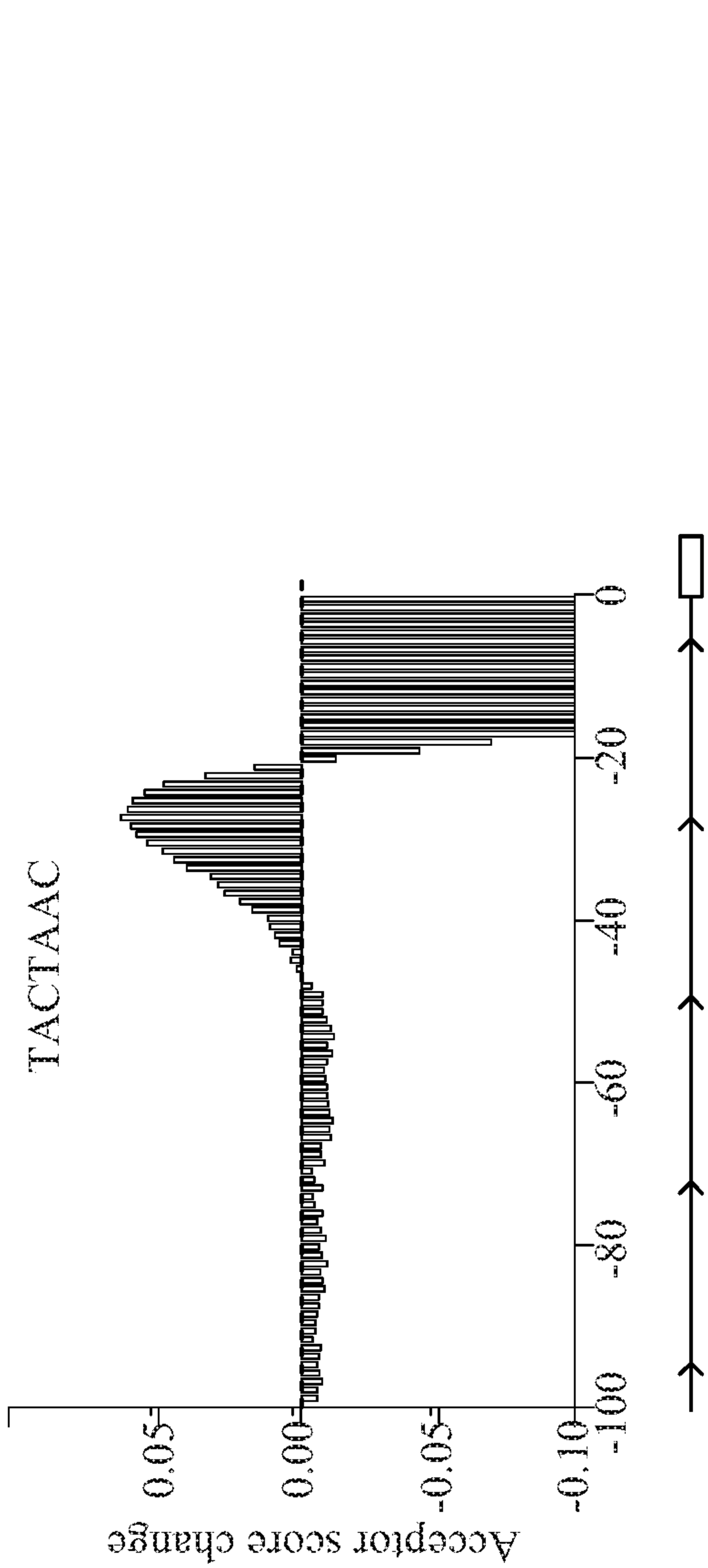


FIG. 43A

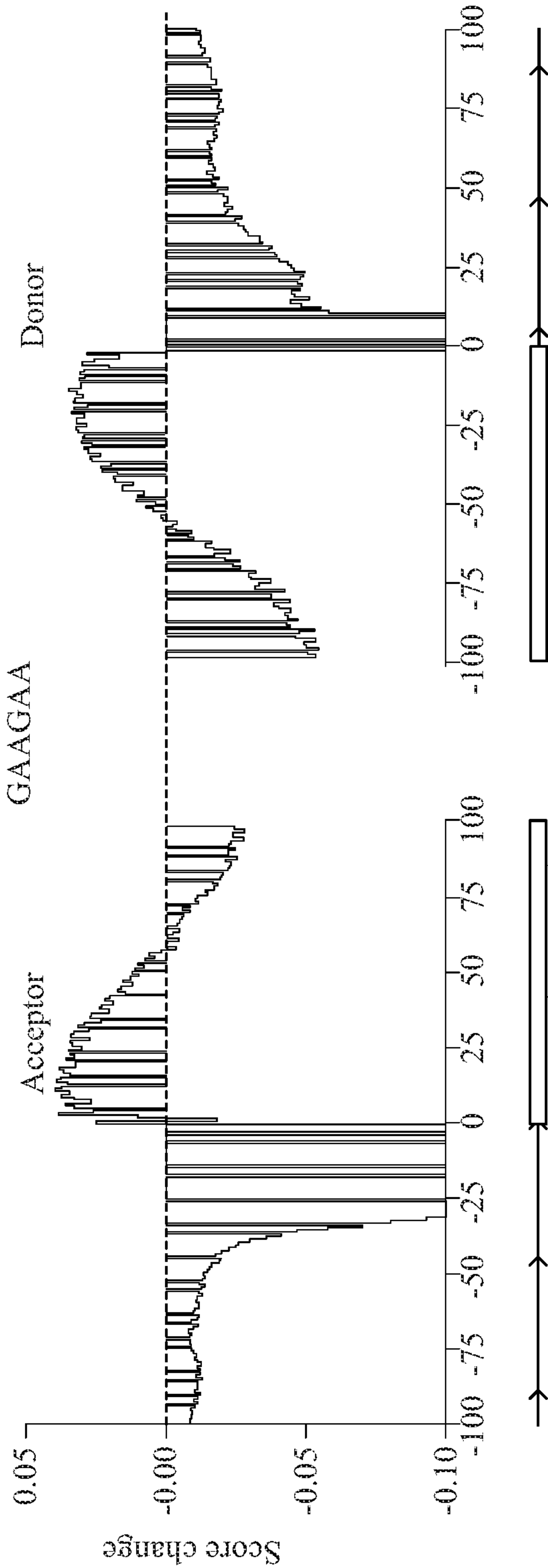


FIG. 43B

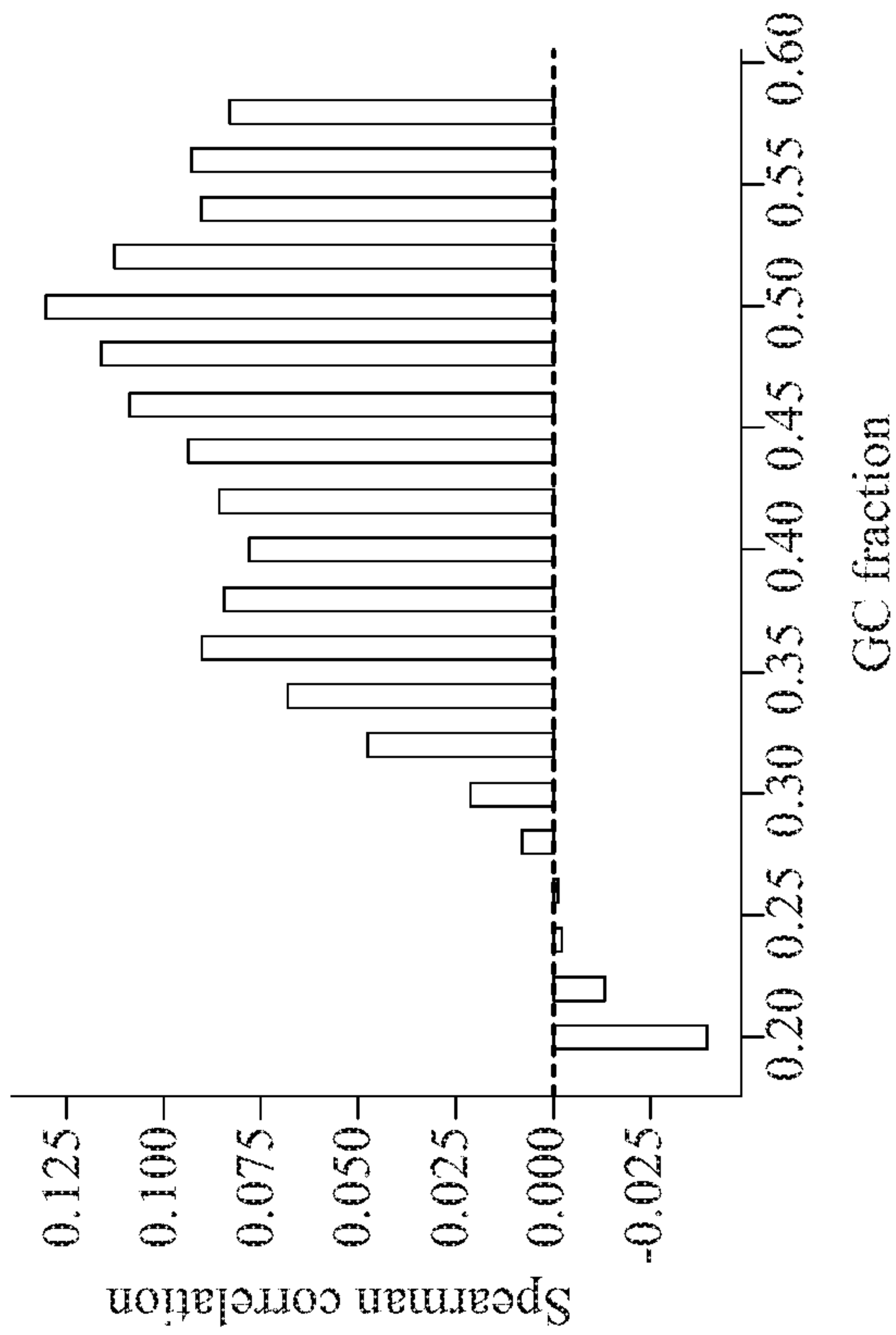


FIG. 44A

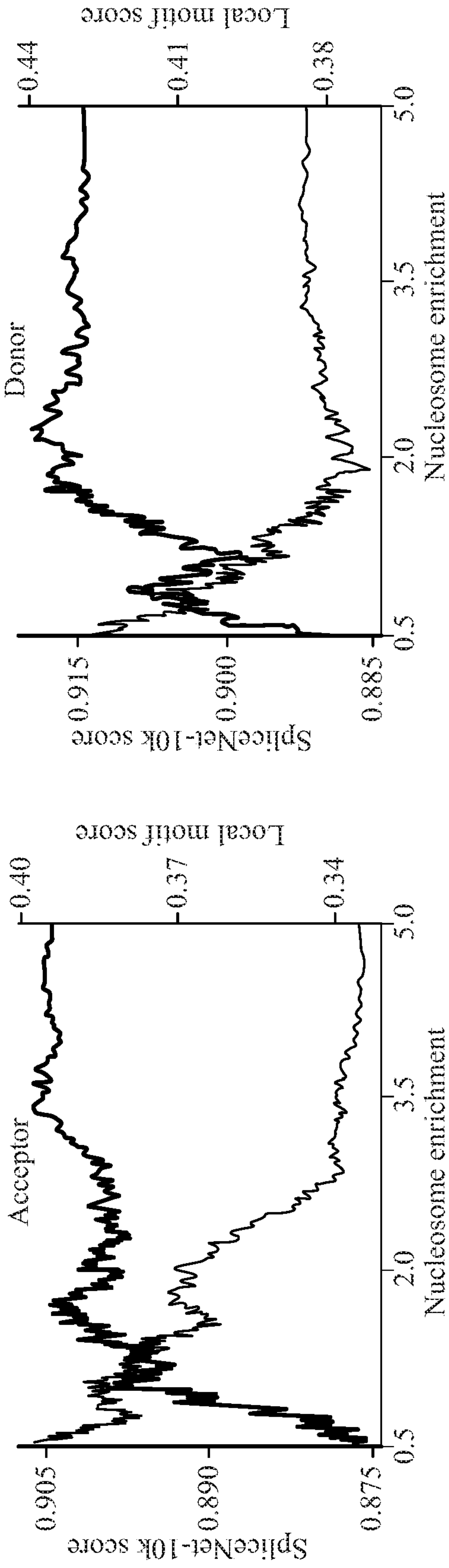
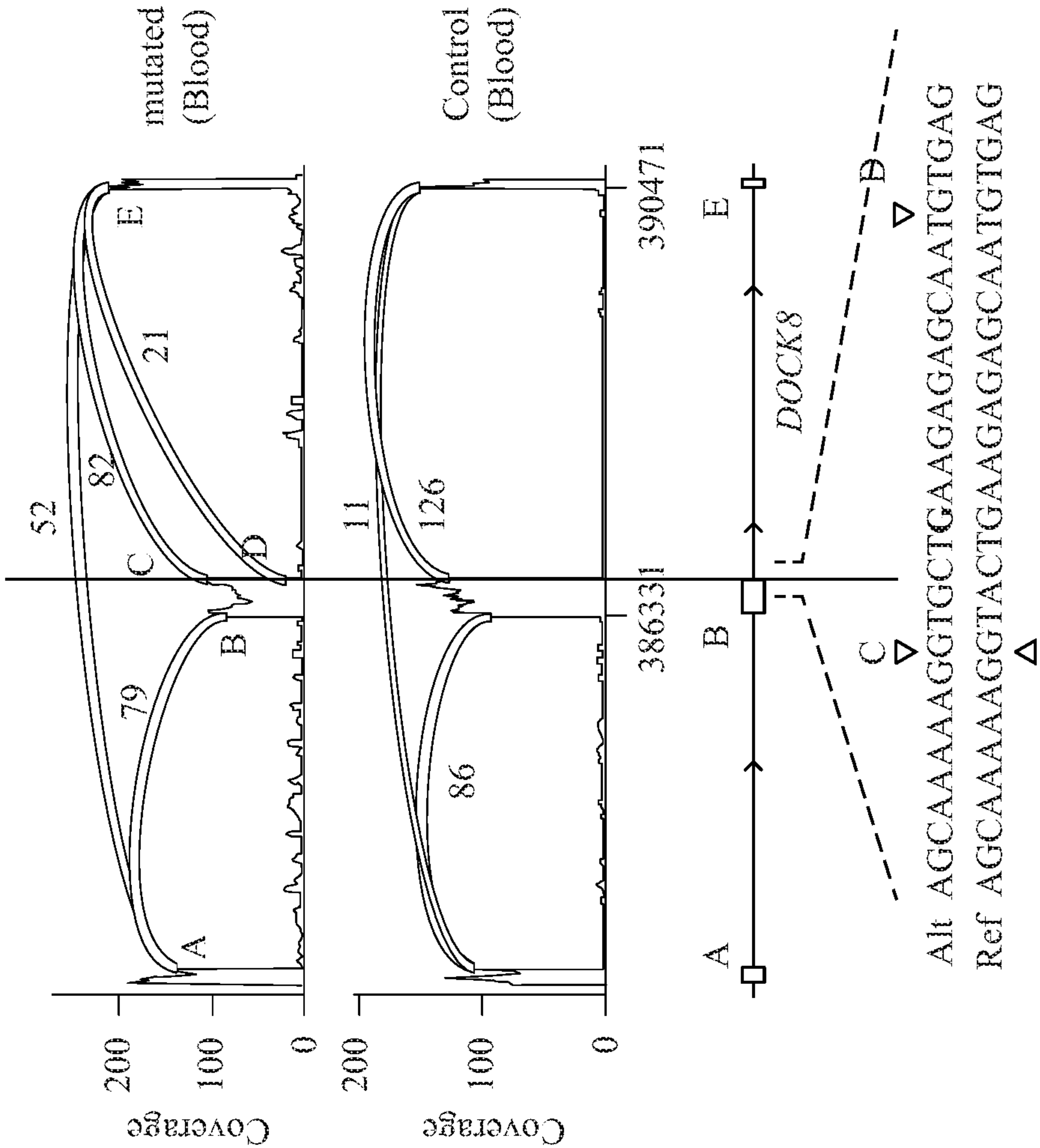


FIG. 44B

chr9:386429 A>GΔ Score = 0.65



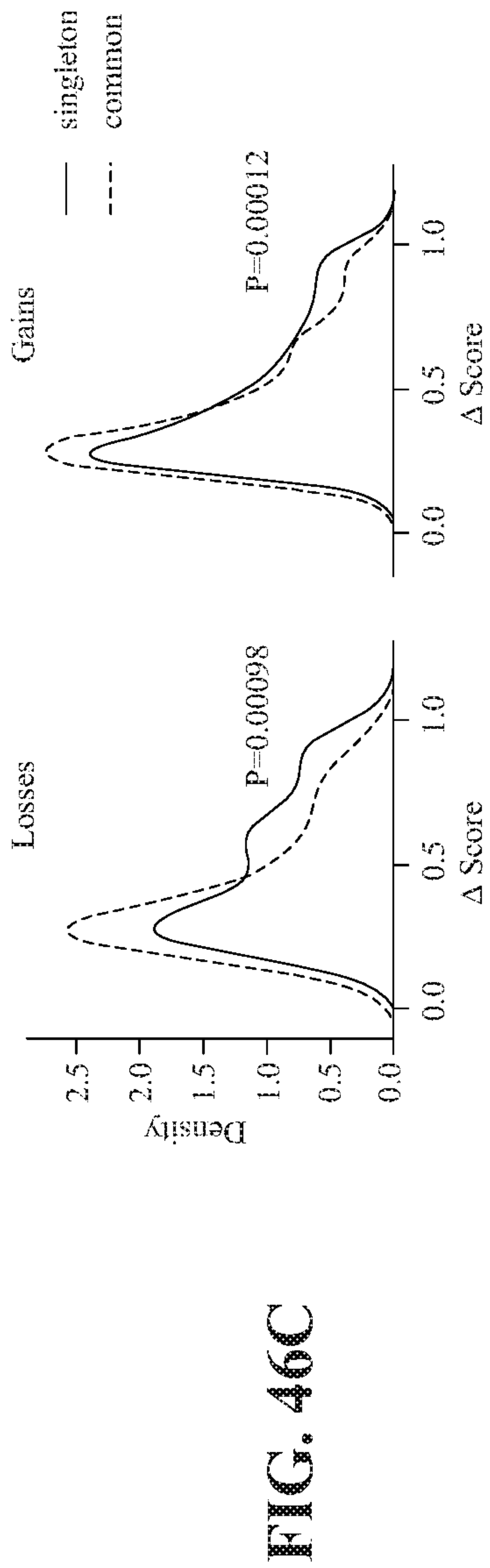
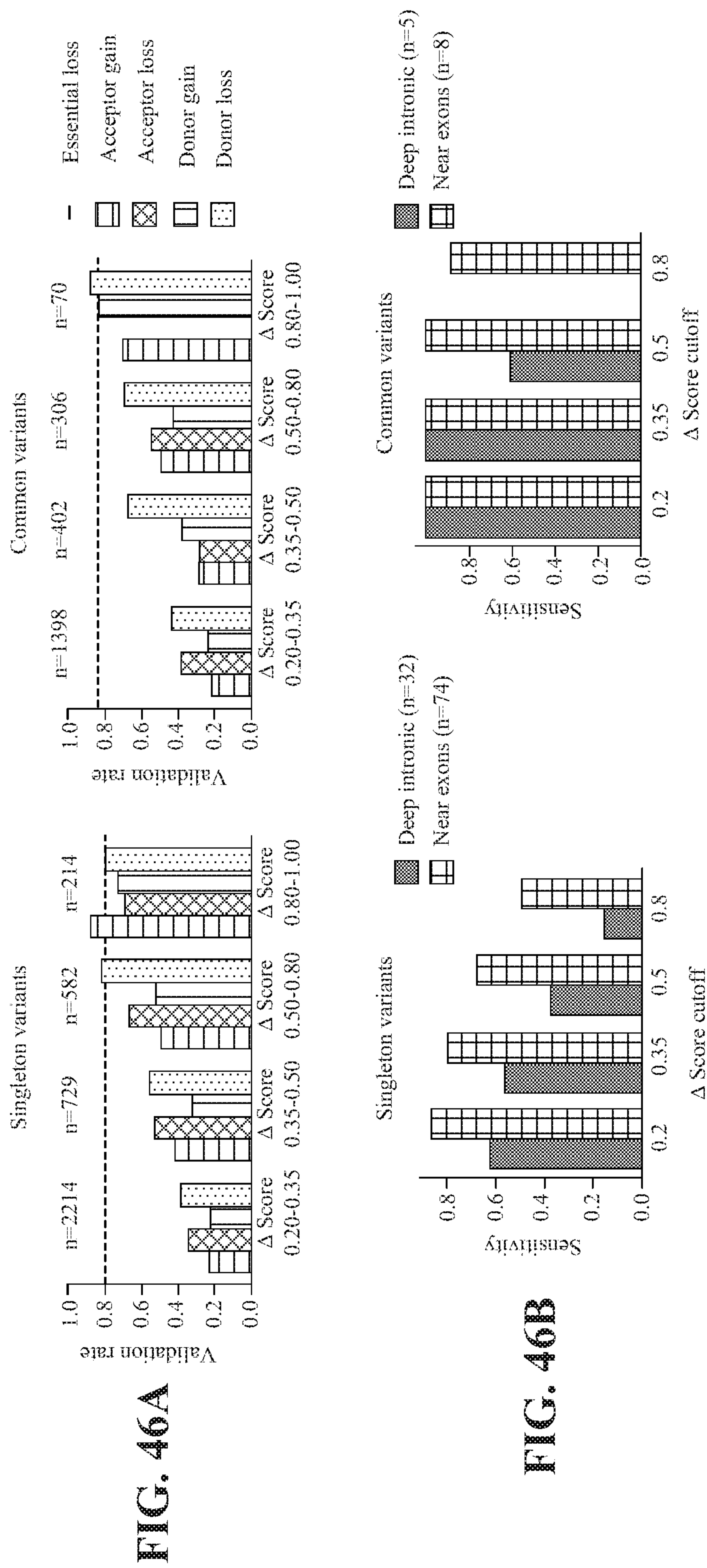
Relative increase in exon skipping:

$$\left(\frac{AE}{AE+(AB+CE)/2} \right)_{\text{mut}} - \left(\frac{AE}{AE+(AB+CE)/2} \right)_{\text{ctrl}} = \frac{52}{52+(79+82)/2} - \frac{11}{11+(86+126)/2} = 0.30$$

Relative loss of annotated junction:

$$\left(\frac{CE}{AE + CE + DE} \right)_{\text{ctrl}} - \left(\frac{CE}{AE + CE + DE} \right)_{\text{mut}} = \frac{126}{11 + 126} - \frac{82}{52+82+21} = 0.39$$

FIG. 45



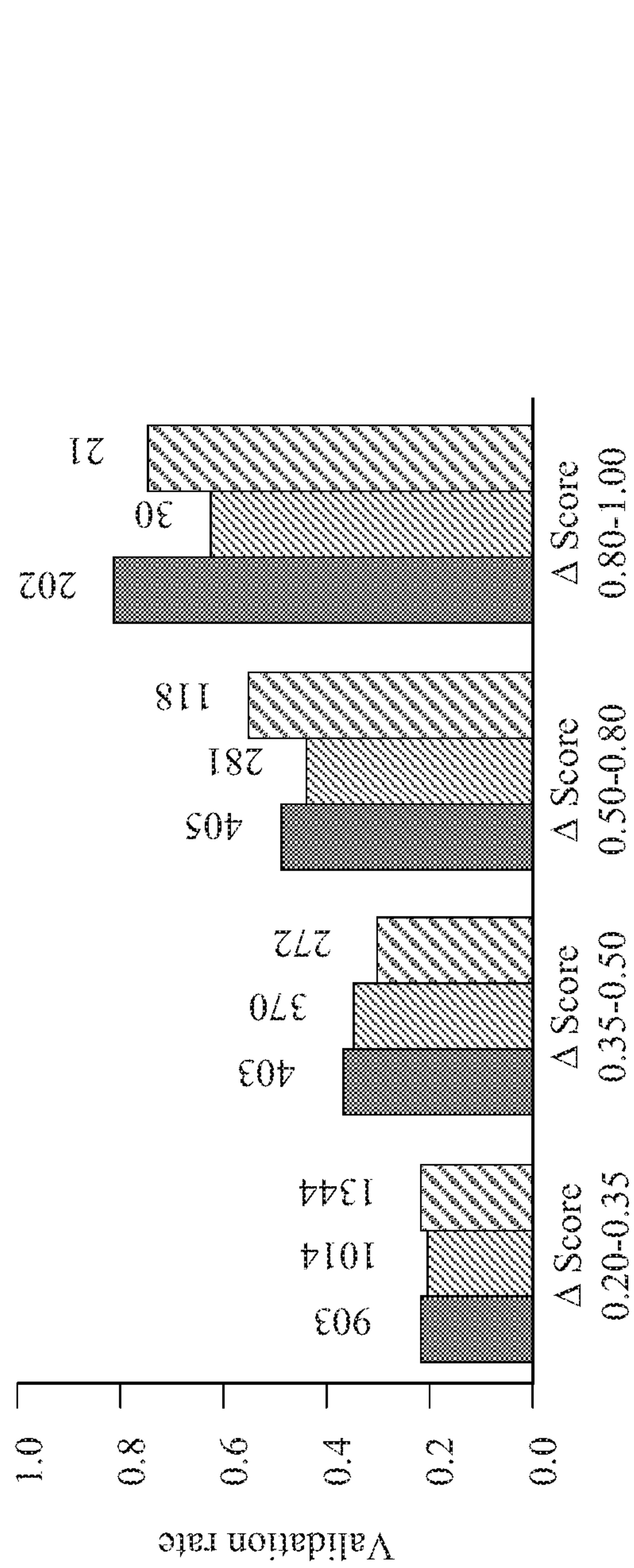


FIG. 47A

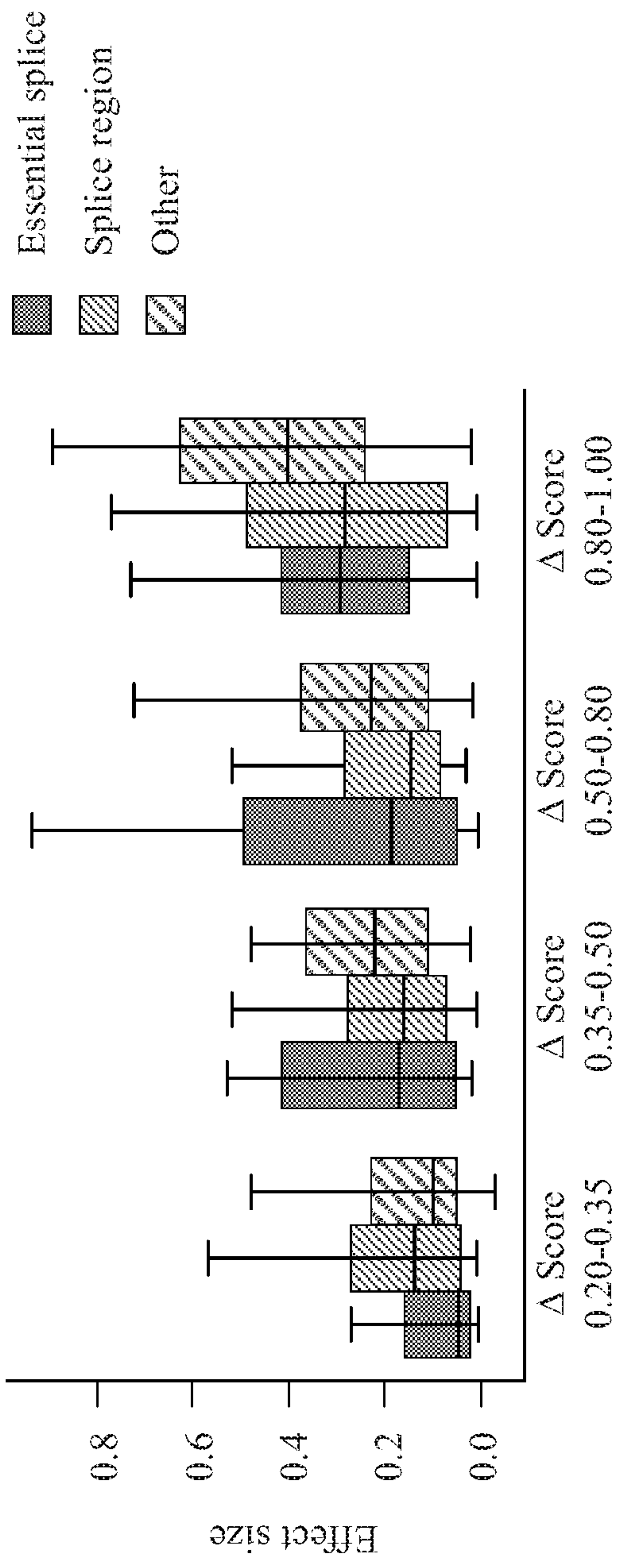


FIG. 47B

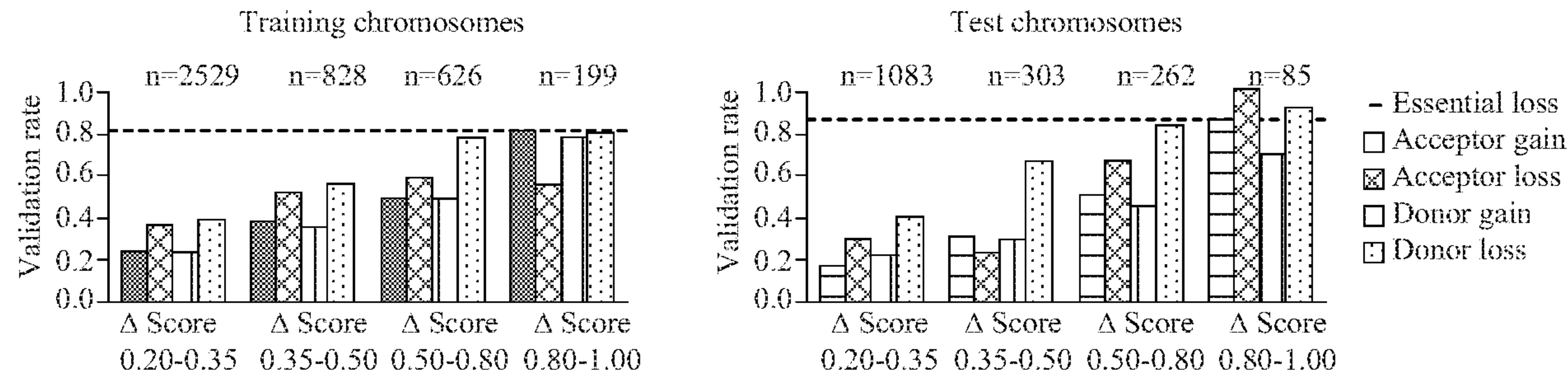


FIG. 48A

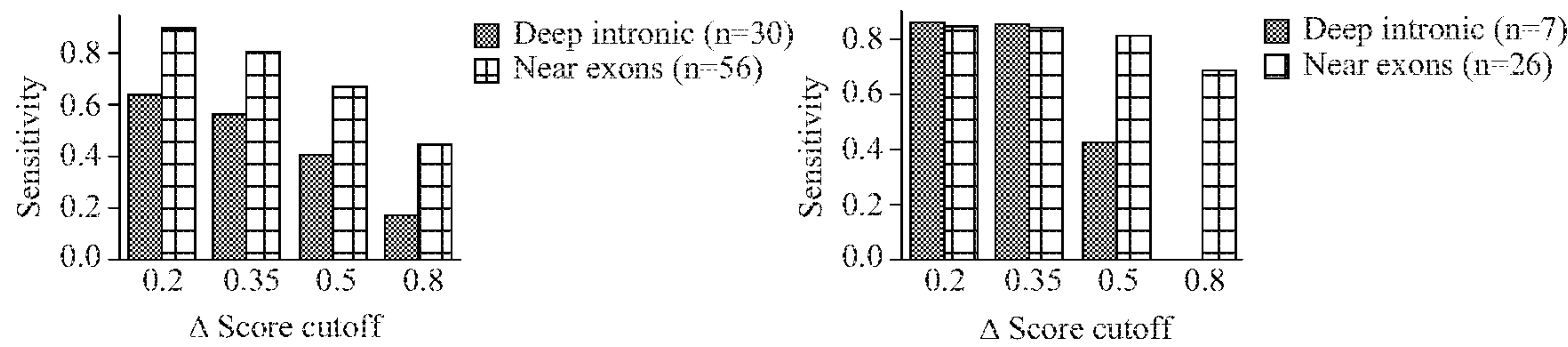


FIG. 48B

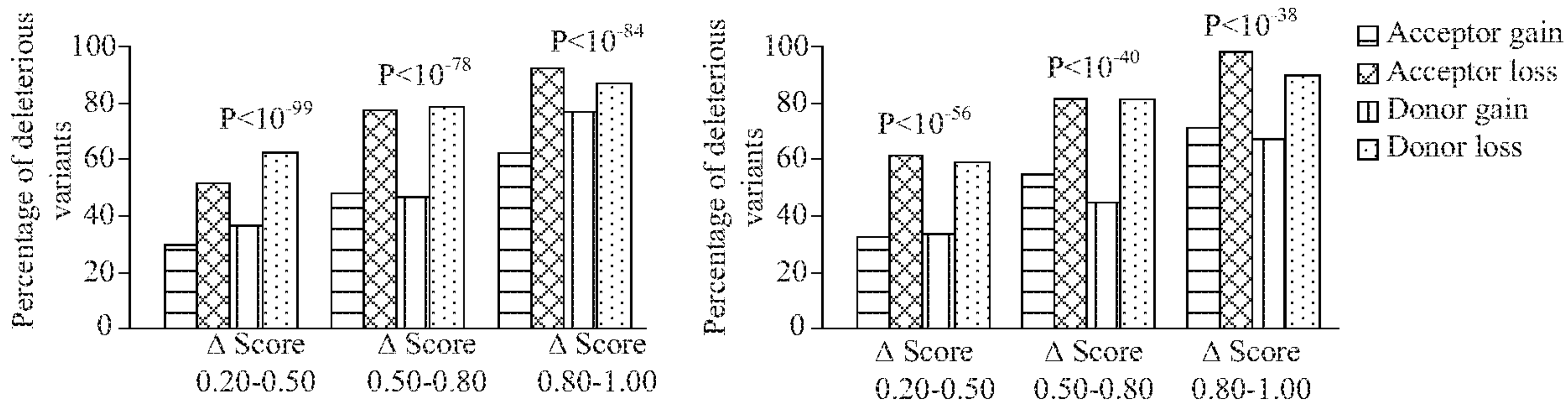


FIG. 48C

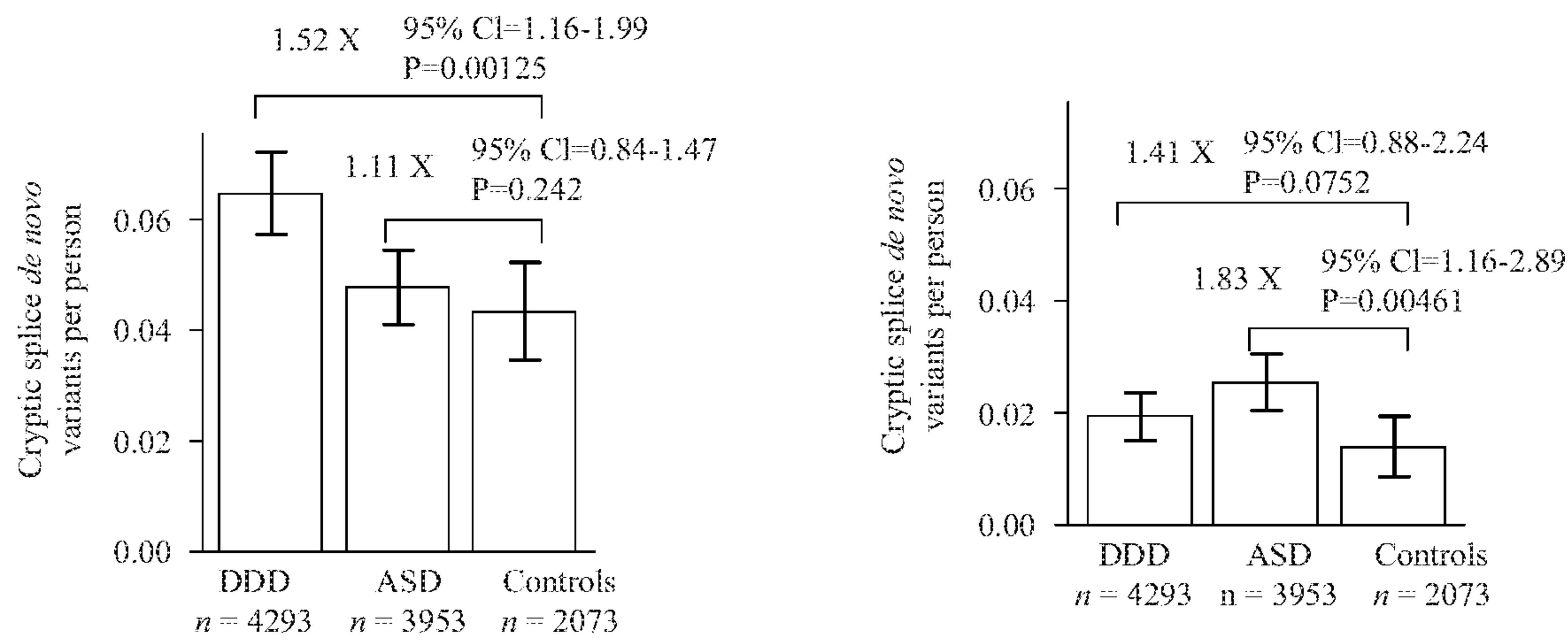


FIG. 48D

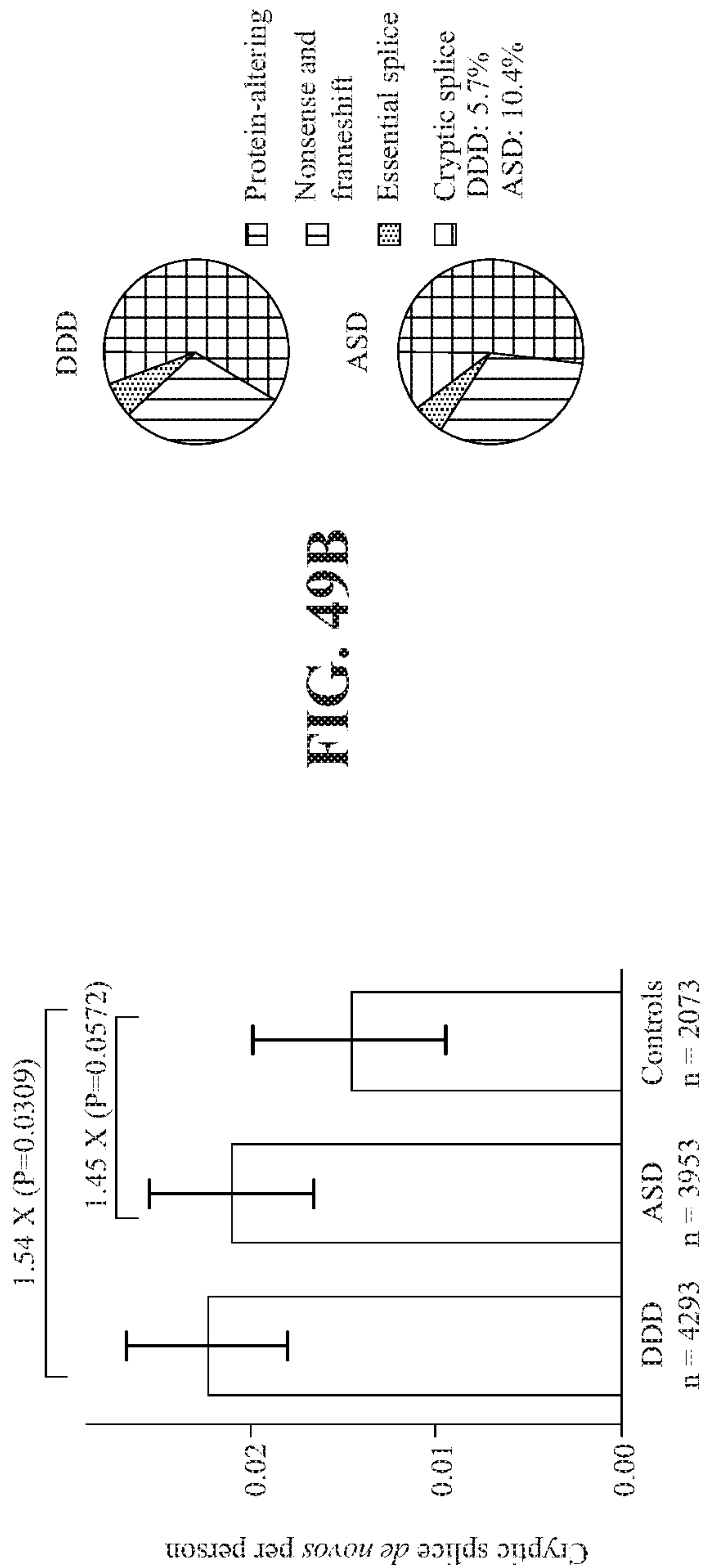
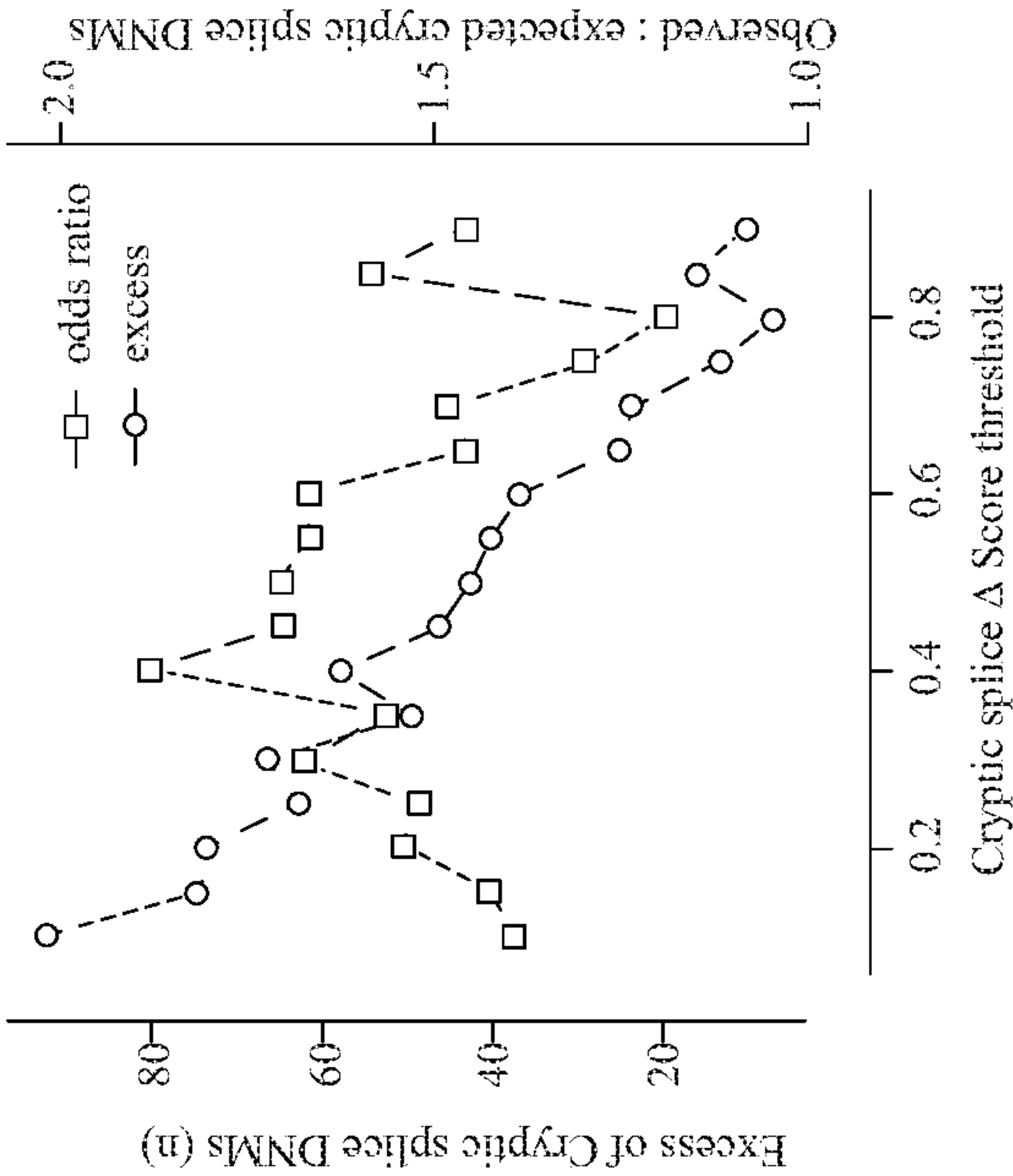
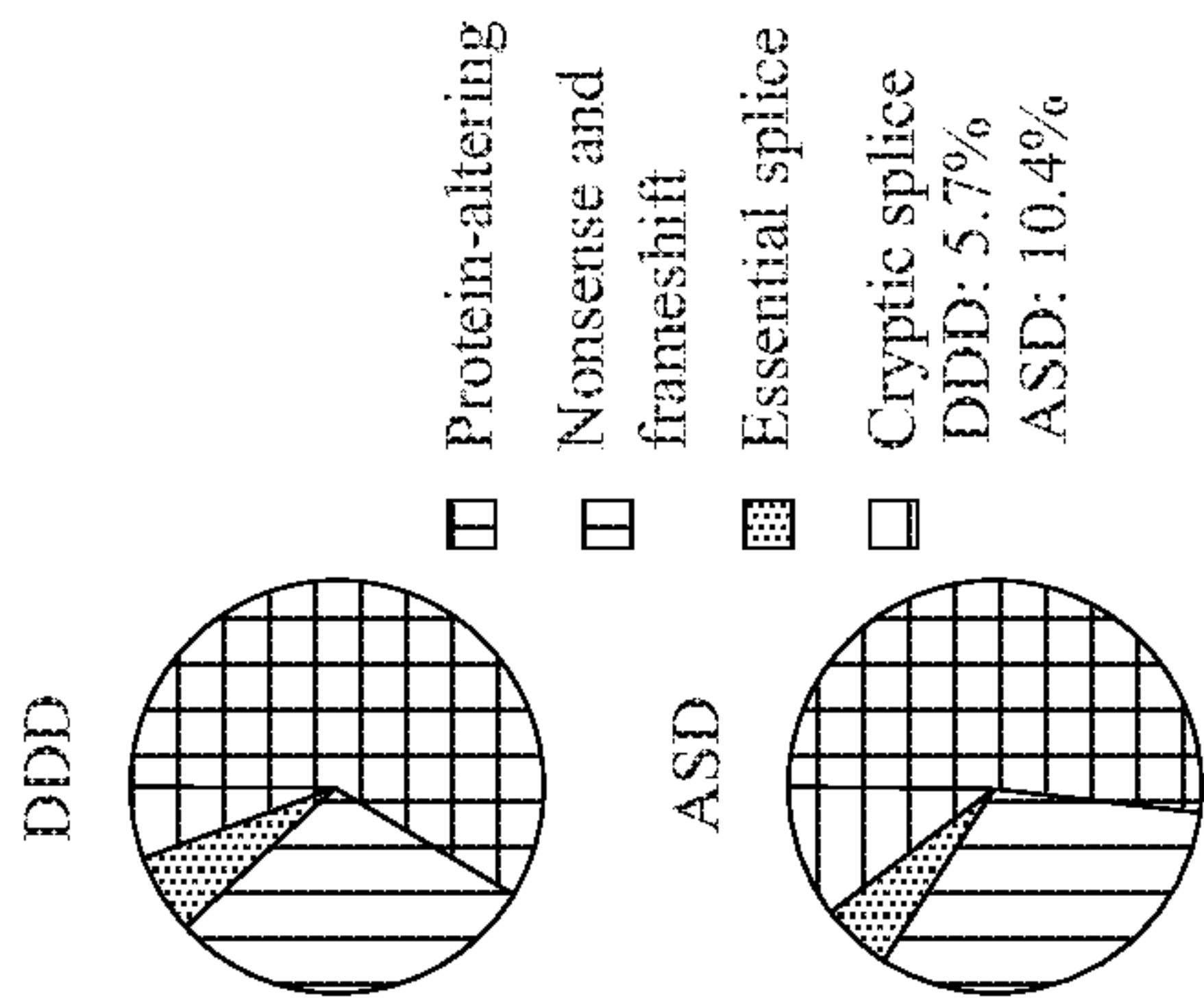


FIG. 49B



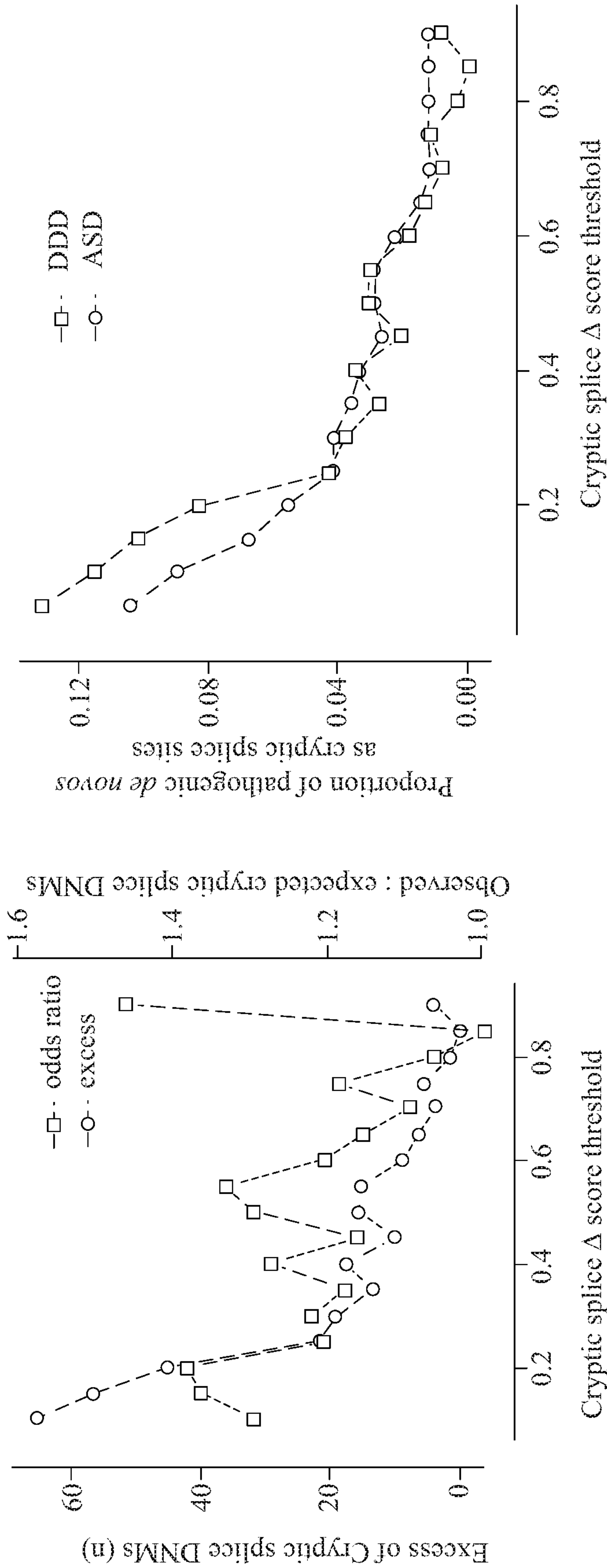


FIG. 50A

FIG. 50B



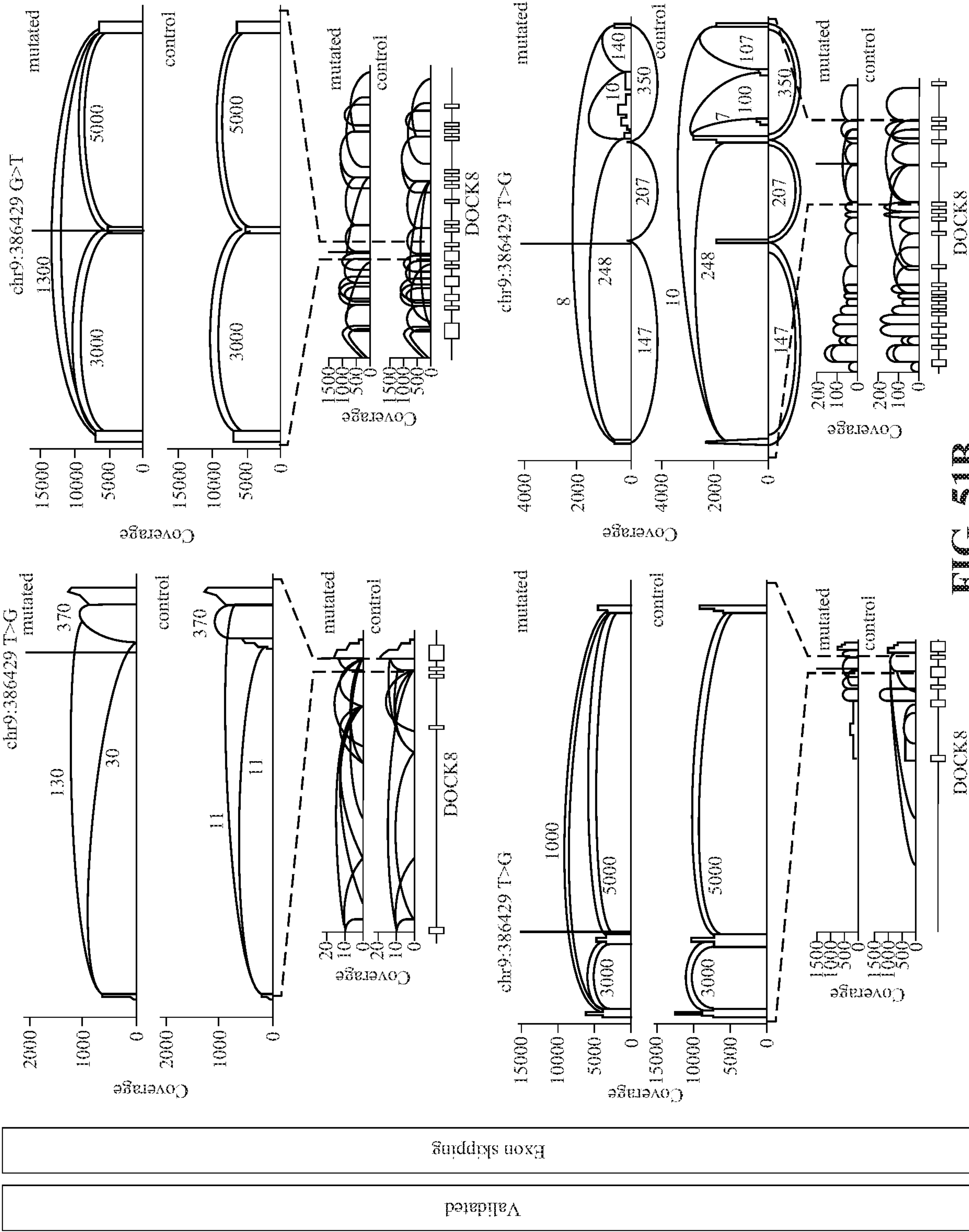


FIG. 51B

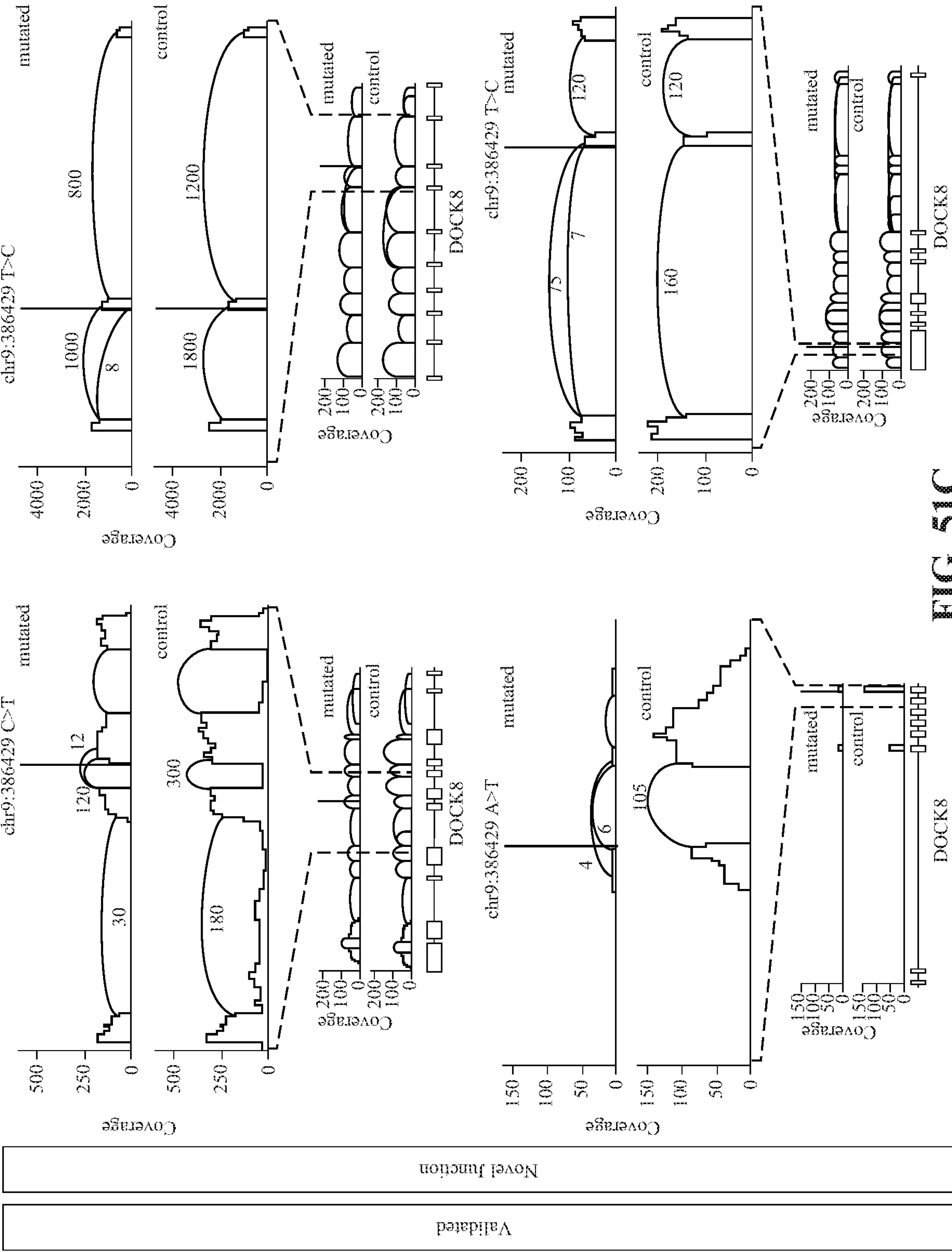


FIG. 51C

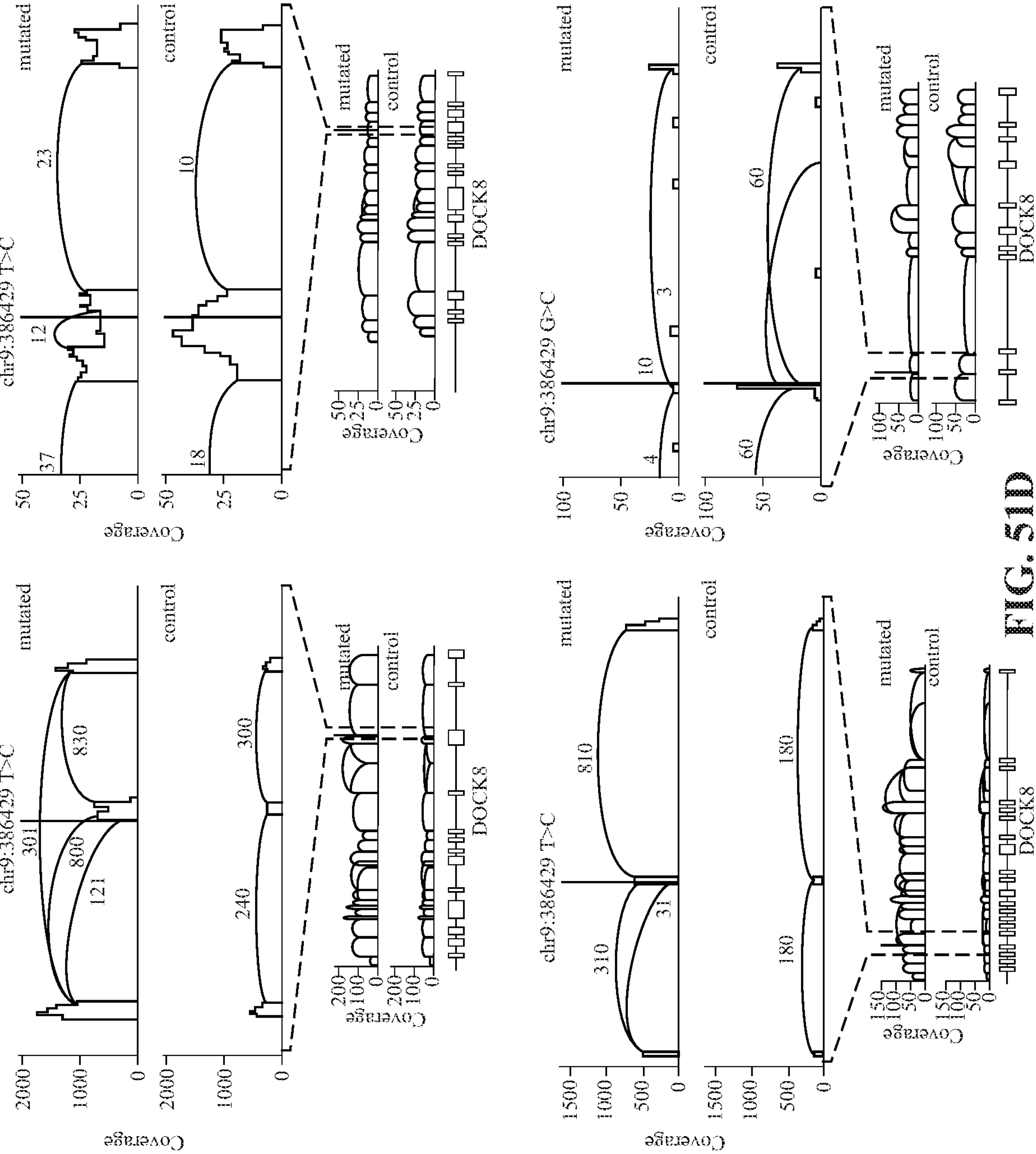


FIG. 51D

	Validated	Novel Junction
--	-----------	----------------

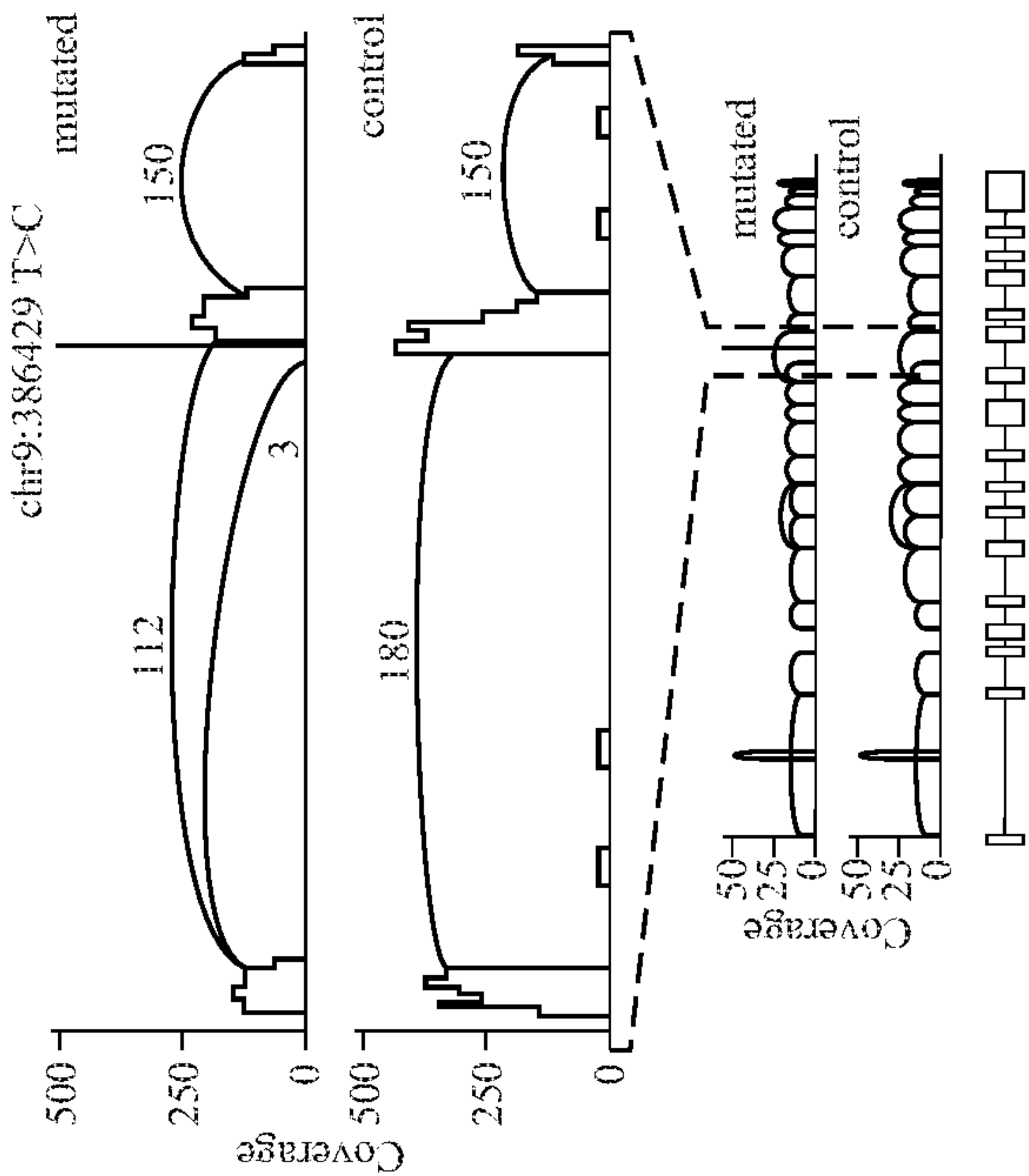


FIG. 51E

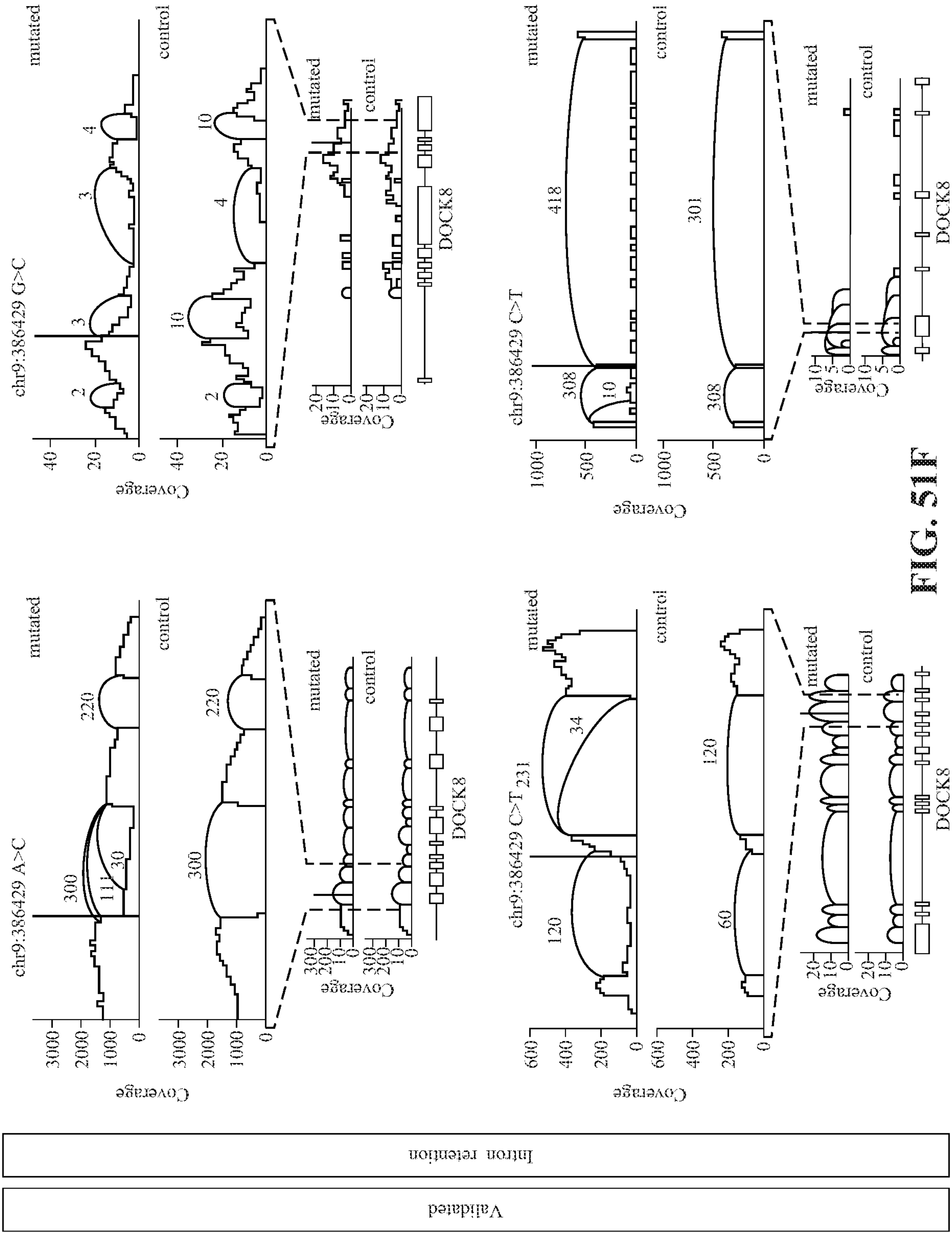


FIG. 51F

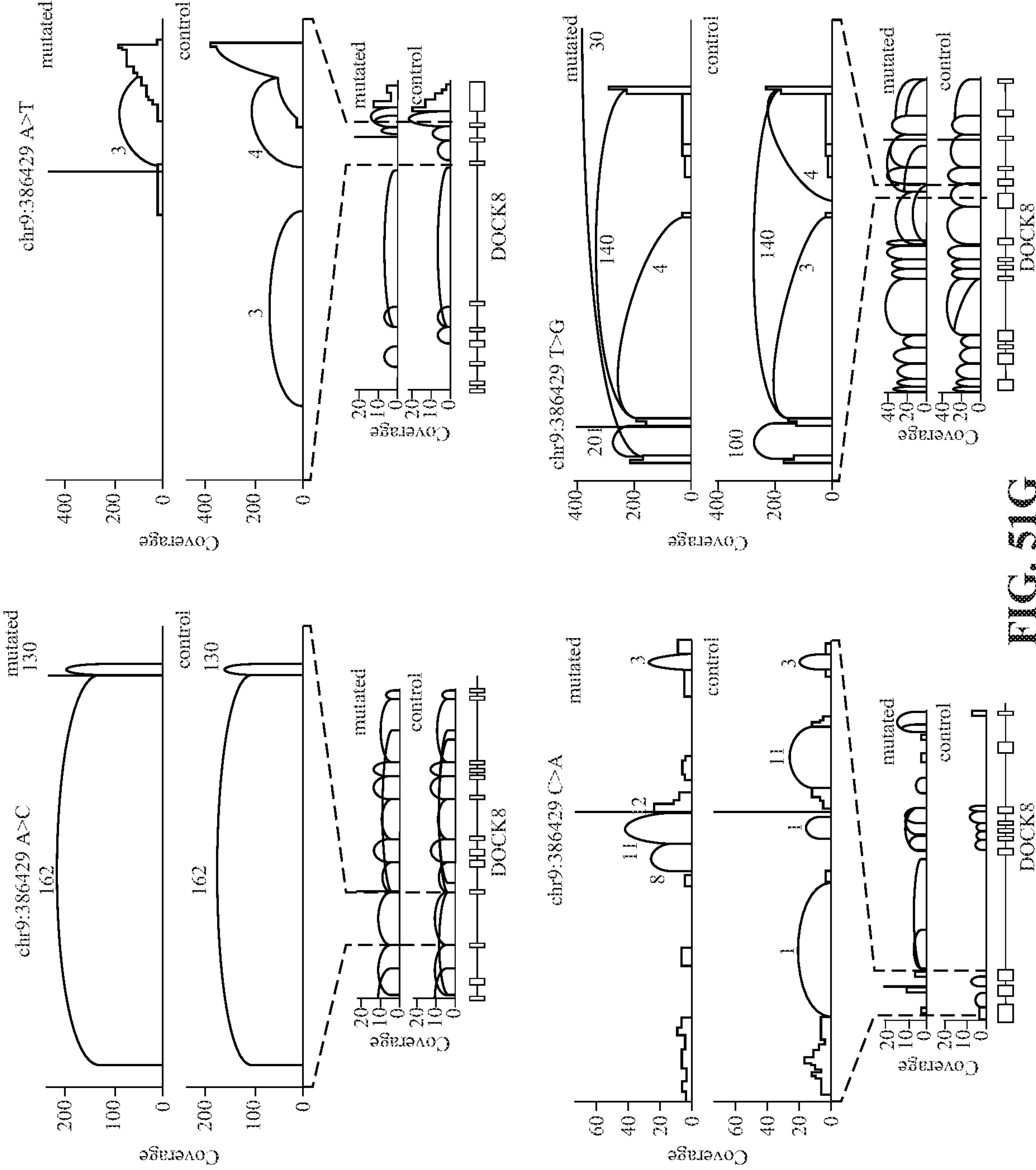


FIG. 51G

Failed

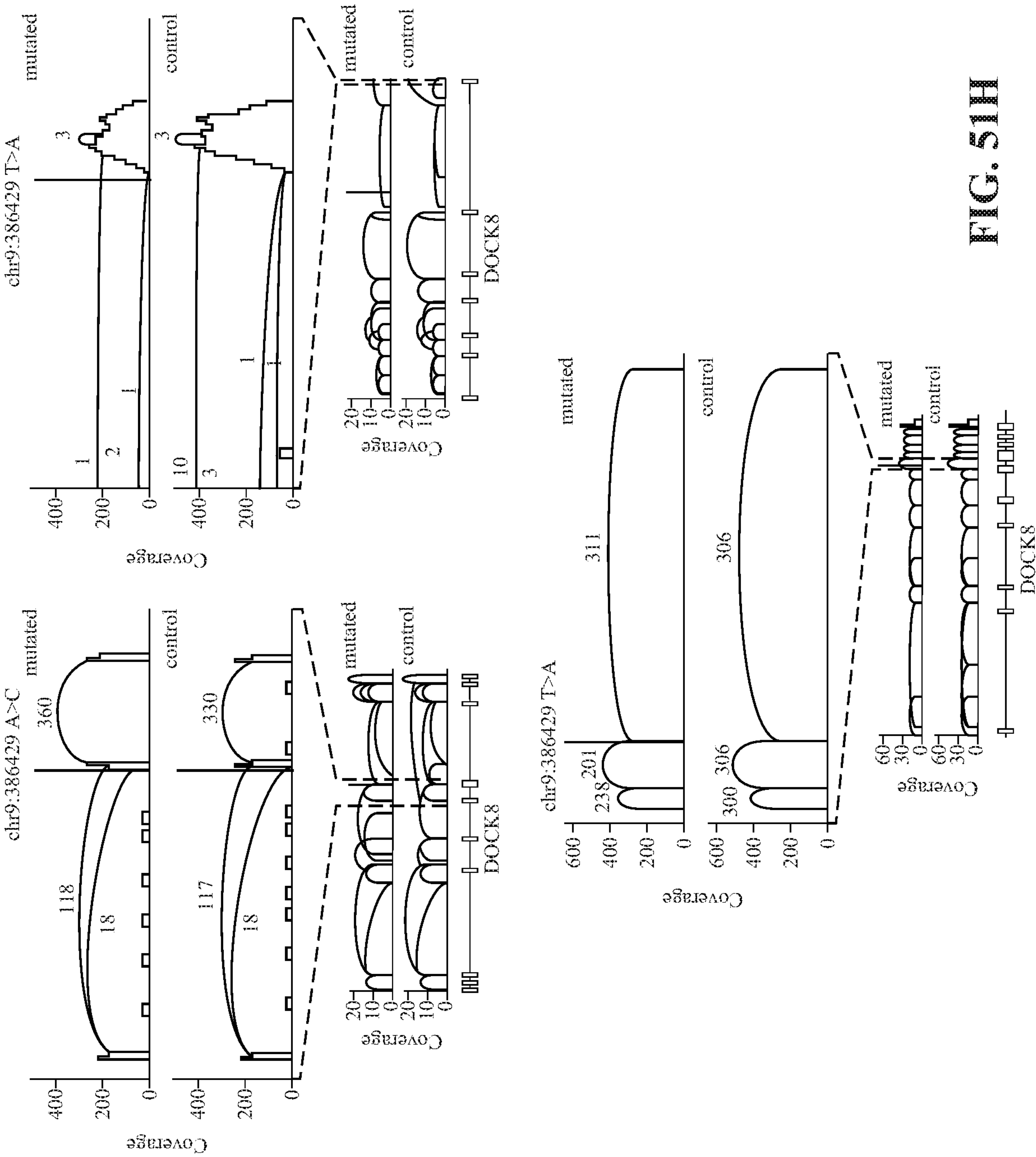


FIG. 51H

Failed

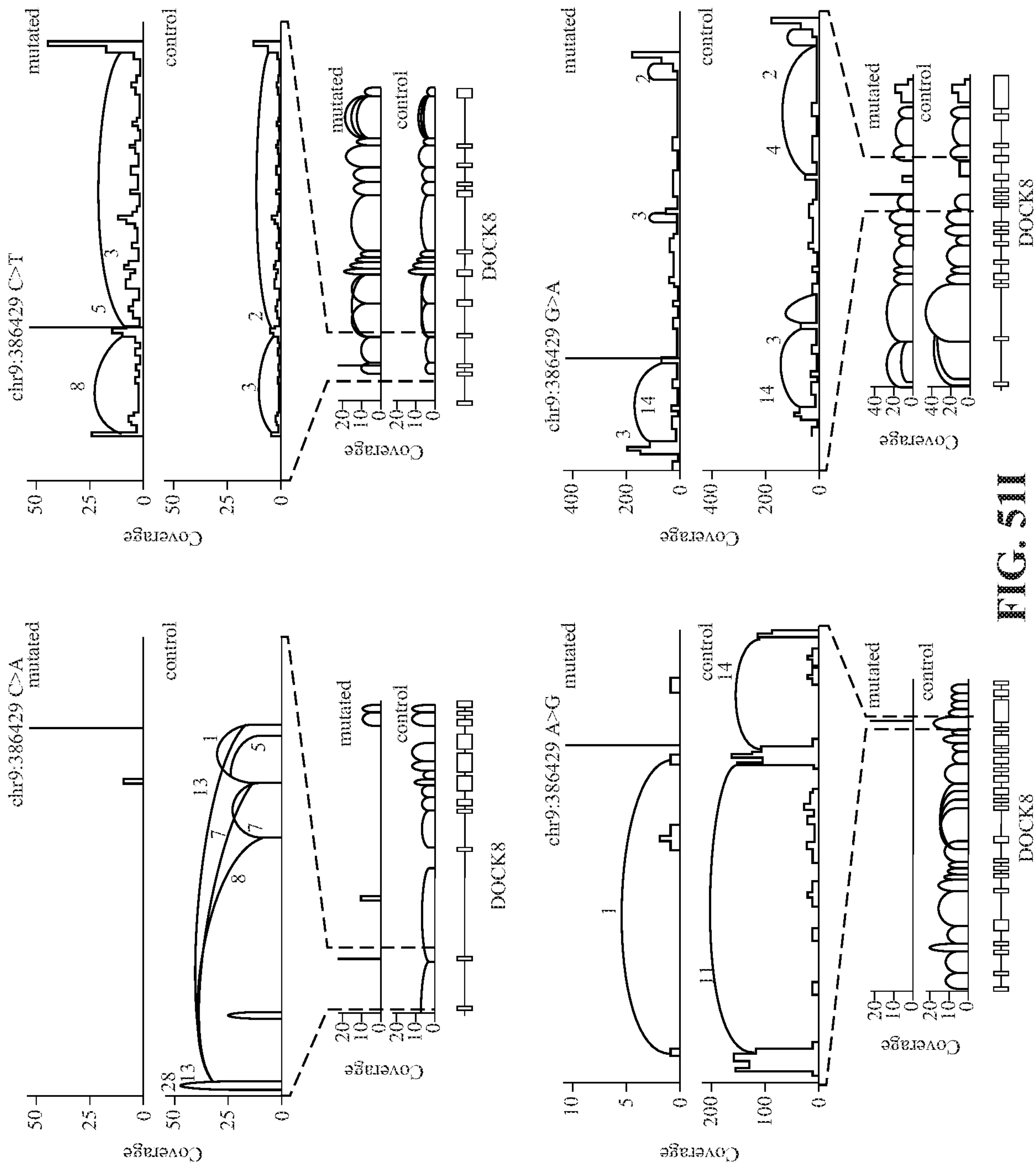


FIG. 511

Uncertain

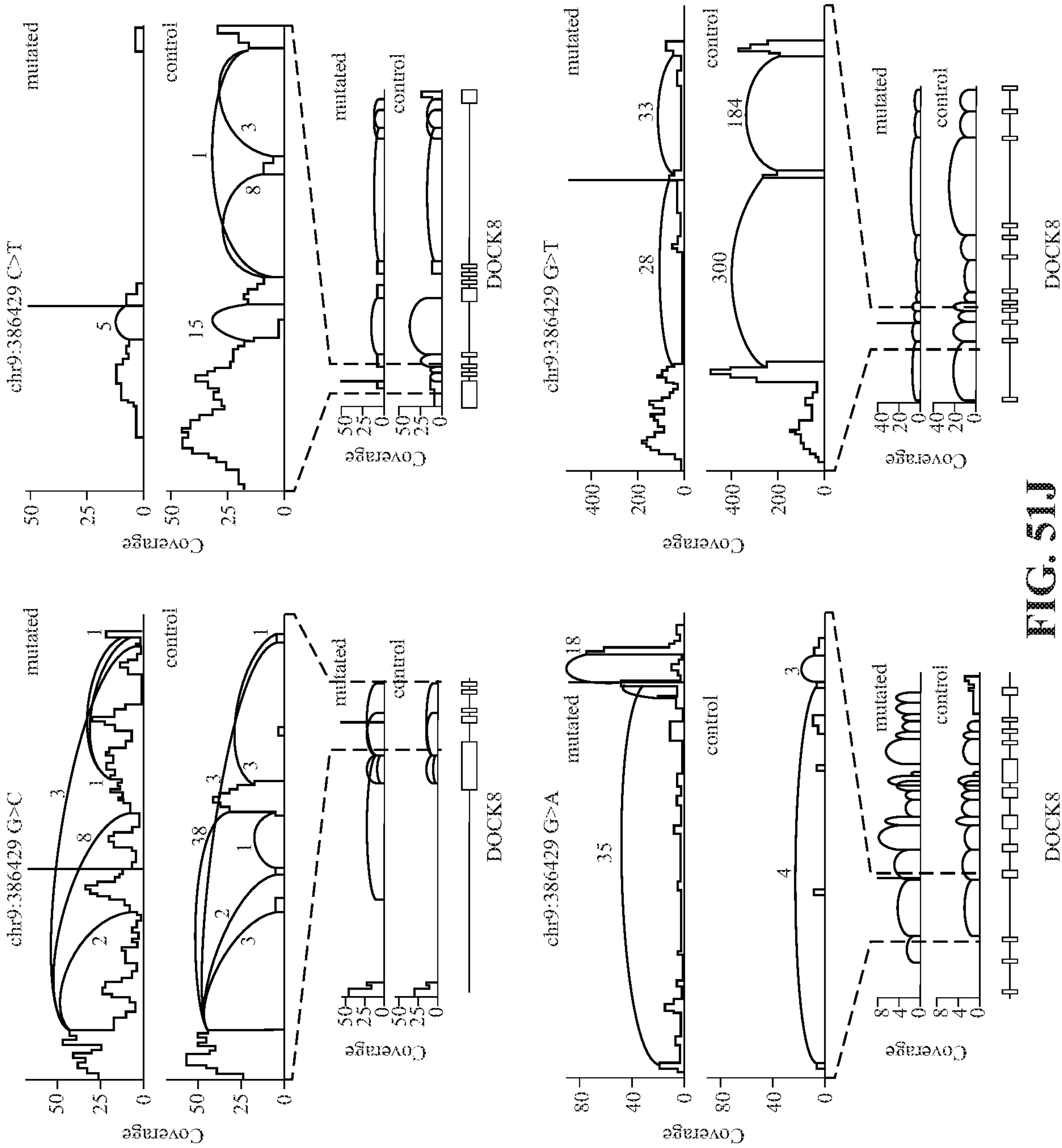


FIG. 51J

Uncertain

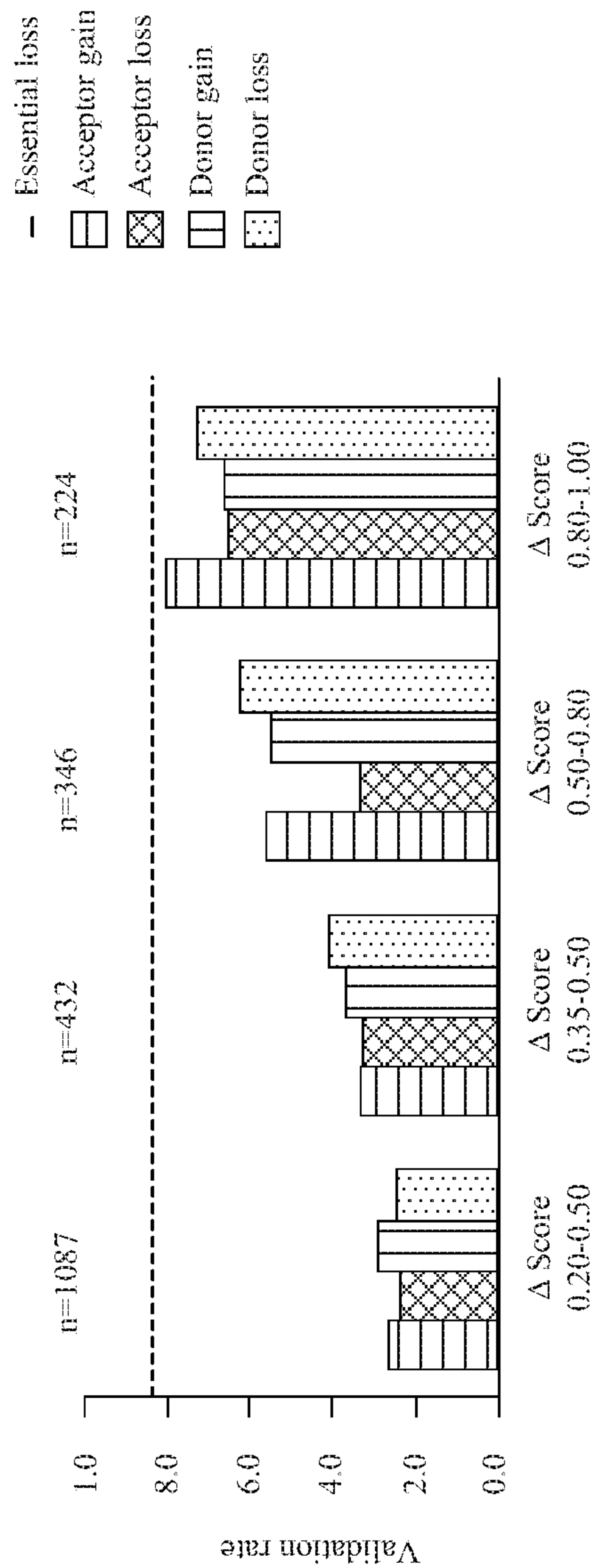


FIG. 52A

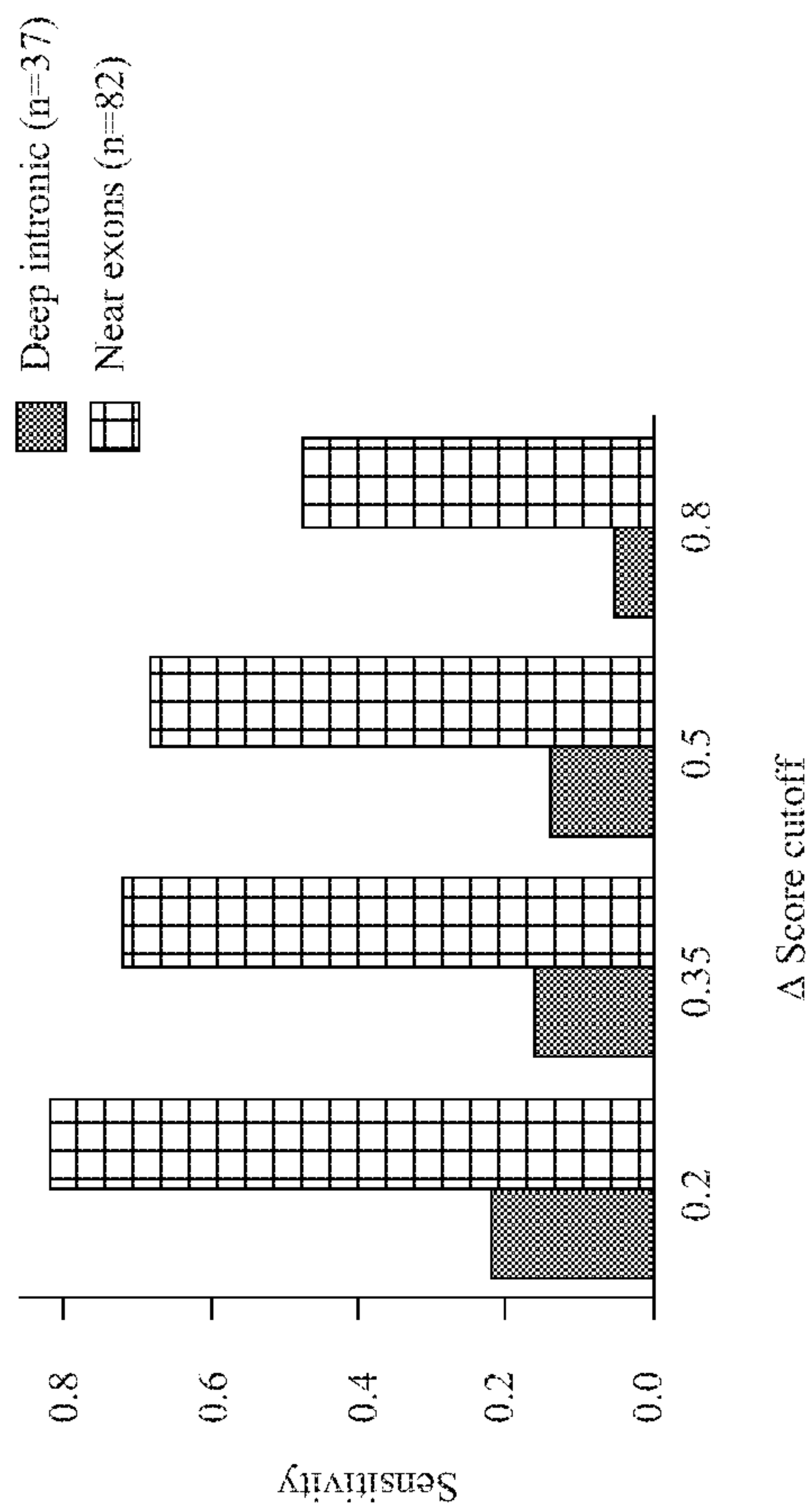


FIG. 52B

	Top- <i>k</i> accuracy		Precision-Recall AUC	
	Acceptor	Donor	Acceptor	Donor
Model #1	0.9304	0.9353	0.9630	0.9682
Model #2	0.9297	0.9354	0.9649	0.9695
Model #3	0.9288	0.9373	0.9650	0.9702
Model #4	0.9222	0.9307	0.9616	0.9659
Model #5	0.9267	0.9297	0.9640	0.9673

	Top- <i>k</i> accuracy		Precision-Recall AUC	
	Acceptor	Donor	Acceptor	Donor
1 model	0.9304	0.9353	0.9630	0.9682
2 models	0.9376	0.9428	0.9694	0.9742
3 models	0.9409	0.9472	0.9715	0.9758
4 models	0.9424	0.9474	0.9727	0.9765
5 models	0.9429	0.9479	0.9737	0.9771

FIG. 53A

FIG. 53B

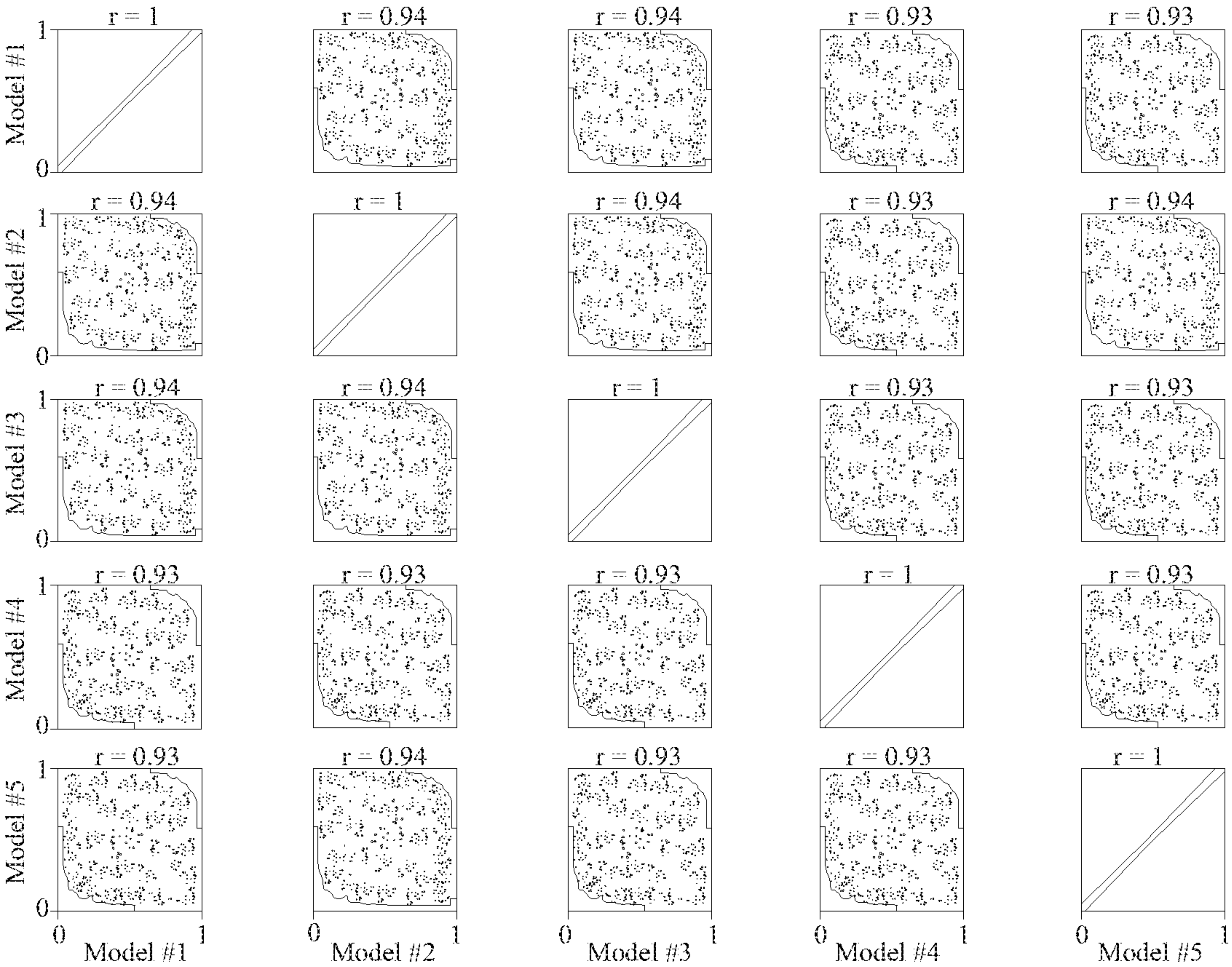


FIG. 53C

A				
SpliceNet-10k				
Exon density	Top-k accuracy		Precision-Recall AUC	
	Acceptor	Donor	Acceptor	Donor
1 exon	0.9165	0.9292	0.9592	0.9659
2 exons	0.9307	0.9357	0.9645	0.9674
3 exons	0.9519	0.9512	0.9816	0.9822
4 exons	0.9556	0.9540	0.9833	0.9845
≥5 exons	0.9579	0.9665	0.9826	0.9872
B				
MaxEntScan				
Exon density	Top-k accuracy		Precision-Recall AUC	
	Acceptor	Donor	Acceptor	Donor
1 exon	0.1834	0.2487	0.1140	0.1660
2 exons	0.2180	0.3124	0.1512	0.2375
3 exons	0.2616	0.3348	0.1964	0.2773
4 exons	0.2950	0.3820	0.2307	0.3347
≥5 exons	0.3583	0.4652	0.3103	0.4314

FIGs. 54A-B

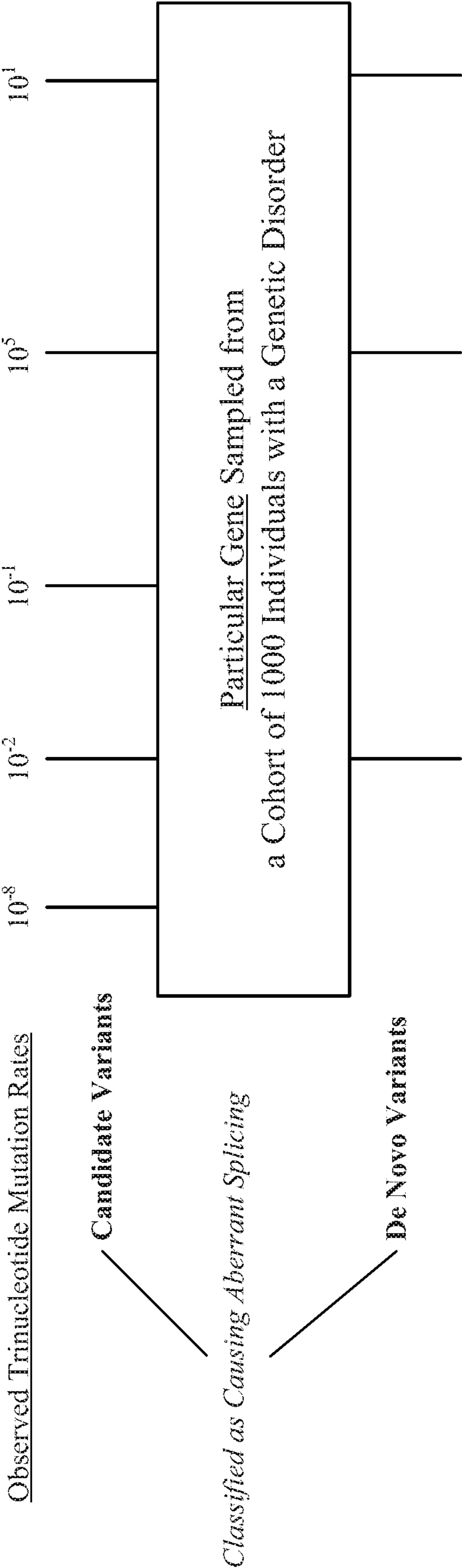
Tissues	Variant individuals	Control individuals
Skin - Sun Exposed (Lower leg)	T6MO	OXRK
Muscle - Skeletal, Transformed fibroblasts	WHSB	WK11
Artery - Tibial, Lung	XOT4, S341, POMQ	WFON
Whole Blood	TMMY	T8EM

FIG. 55

Gains					Losses				
SpliceNet-10k	GeneSplicer	NNSplice	MaxEntScan	SpliceNet-10k	GeneSplicer	NNSplice	MaxEntScan	SpliceNet-10k	GeneSplicer
0.1	103.5145	0.9024	0.901	0.1	1.5183	0.1063	0		
0.2	105.6838	0.9288	0.9171	0.2	2.6295	0.2315	0.0088		
0.3	107.1061	0.9542	0.9419	0.3	3.4707	0.3554	0.0312		
0.4	108.0655	0.967	0.9582	0.4	4.0856	0.4625	0.0846		
0.5	108.9011	0.9739	0.9694	0.5	4.6085	0.5349	0.1003		
0.6	109.6138	0.9781	0.9759	0.6	5.2084	0.6345	0.1624		
0.7	110.388	0.9817	0.9826	0.7	6.4456	0.7573	0.2838		
0.8	111.2775	0.9839	0.9875	0.8	8.1053	0.8202	0.4271		
0.9	112.6826	0.9858	0.9915	0.9	10.2586	0.8529	0.641		

FIG. 56

Per-Gene Enrichment Analysis



$$\sum (\text{observed trinucleotide mutation rates}) = 10^{-5}$$

Transmission (Chromosome) Count

Baseline Number of Mutations = $2 \times 1000 \times 10^{-5} = 0.2$

Cohort Size

De Novo Variant Count = 3

FIG. 57

Genome-Wide Enrichment Analysis

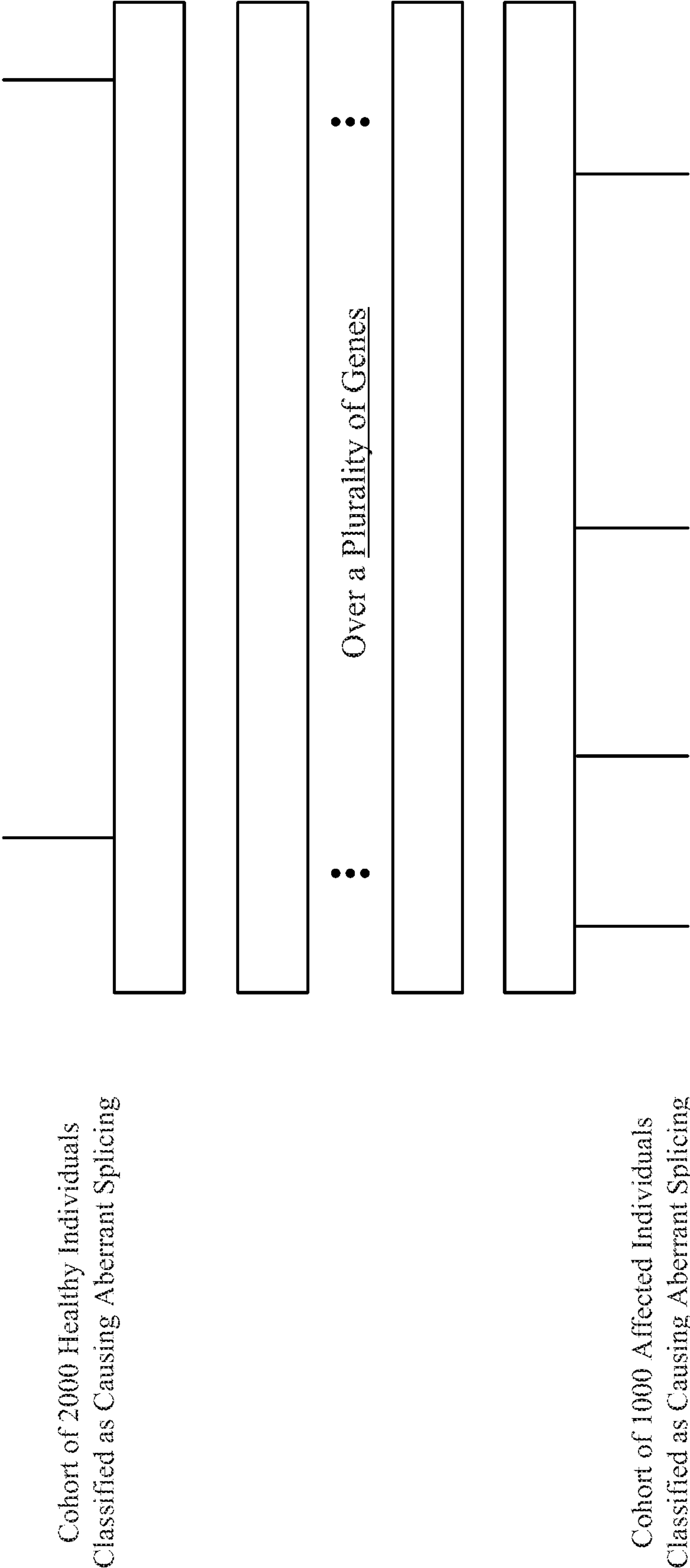


FIG. 58

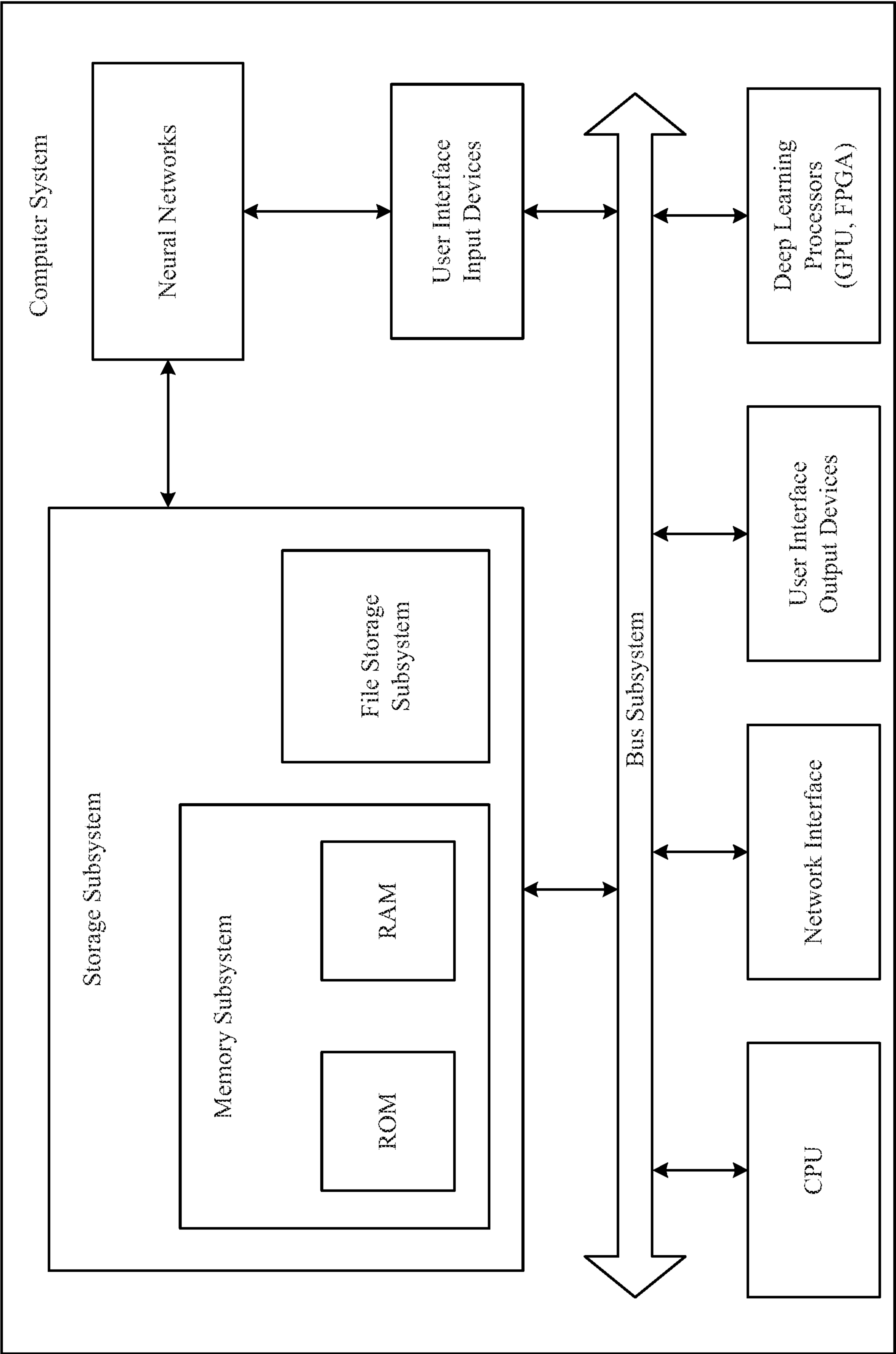


FIG. 59