

# Precise Uncertain Significance Prediction Using Latent Space Matrix Factorization Models: Genomics Variant and Heterogeneous Clinical Data-driven Approaches - Supplementary File (Supplementary information, algorithms, figures and tables)

Abdollahi et al.

## 1 Supplementary Information

### 1.1 The feedforward neural network

In the feedforward neural network architecture, the output of a layer is passed on to the next layer and multiplied by different weight matrices. These results are summed and fed into a non-linear activation function:

$$\begin{aligned} a^{[1]} &= \sigma(W^{[1]}x + b^{[1]}) \\ a^{[2]} &= \sigma(W^{[2]}a^{[1]} + b^{[2]}) \\ a^{[3]} &= \sigma(W^{[3]}a^{[2]} + b^{[3]}) \\ y &= \text{sigmoid}(W^{[4]}a^{[3]} + b^{[4]}) \end{aligned} \quad (1)$$

where  $x \in \mathbb{R}^f$  is the input data vector with  $f$  features (genes), assuming that  $x = a^{[0]}$ ,  $y = a^{[4]}$ , and  $f = n^{[0]}$ ,  $W^{[i]} \in \mathbb{R}^{n^{[i-1]} \times n^{[i]}}$  is the weight matrix connecting  $n^{[i-1]}$  neurons of the  $(i-1)^{\text{th}}$  hidden layer to the  $n^{[i]}$  neurons of the  $i^{\text{th}}$  layer and  $b^{[i]}$  is the bias vector in the  $i^{\text{th}}$  layer for  $i = 1, 2, 3, 4$ .  $\sigma(\cdot)$  and  $\text{sigmoid}(\cdot)$  are the activation functions. The reason why we used the sigmoid functions for the output layer is that they give a number between zero and one. We interpreted these numbers as the probability of the types of CIPN occurrence. Since we have four output labels, we allocate four neurons in the output layer. Each sigmoid function declares the probability of a label activation. Experimentally, hyperbolic tangent  $\tanh(\cdot)$  returns values between  $-1$  and  $1$  so that the mean value of the activation functions in each hidden layer is roughly close to zero. Therefore, the learning for the next layer will be much easier. We subsequently defined the following loss function for each label:

$$\mathcal{L}(\hat{y}, y) = -[y \log \hat{y} + (1 - y) \log(1 - \hat{y})] \quad (2)$$

where  $y$  and  $\hat{y}$  are the observed and predicted value for each label. Equation (2) is also called the binary cross-entropy function. This loss function can be minimized by using different optimization algorithms such as root mean square prop (RMSprop) and adaptive moment estimation (AdaM).

### 1.2 Heterogeneous clinical data

All stage III CRC patients received standard surgical resection followed by adjuvant chemotherapy with the standard regimen of 12 cycles mFOLFOX6 (5-fluorouracil, leucovorin, and oxaliplatin). For endometrial or ovarian (Gynecological, GYN) cancer, 6 cycles of carboplatin plus paclitaxel were used as post-operative adjuvant chemotherapy. Clinical neuropathy evaluations were performed, including a neurological examination focusing on motor and sensory systems, nerve conduction studies (NCS) and quantitative sensory tests (QST) of thermal and vibratory sensation. In total, 34 variables were recorded for analysis. The motor NCS were assessed on median, ulnar, peroneal and tibial nerves including distal latency, amplitude of compound action potential, conduction velocity, F-wave latency and latency of H reflex. The sensory NCS were performed on median, ulnar and sural nerves including distal latency, amplitude of compound action potential and conduction velocity were obtained from the sensory NCS. The QST of thermal threshold, thermal pain thresholds, and vibratory sensation in the upper and lower limbs were evaluated. The follow-up period is every 3-6 months. Germline DNA with whole genome sequencing were tested at the time of enrollment. The NCKUH institutional review board approved this study (A-ER-103-395 and A-ER-104-153), and all the participants gave informed written consent. We measured the results of these examinations after the patients had gone through 12 months of chemotherapy. The boxplots of the values obtained from each examination are shown in Supplementary Figure S9.

### 1.3 Different ways for gene selection

For each neuropathy type (output label), we select  $n$  examinations which are highly-correlated with the output label. Then, for each selected examination obtained from the previous step, we select  $m$  genes which are highly-correlated with the selected examination. We used the following values for  $n$  and  $m$ :

$$\begin{aligned} n &= 2, 3, \dots, 10 \\ m &= 50, 100, 150, 200, 250, 300 \end{aligned} \quad (3)$$

We examined all conditions ( $9 \times 6 = 54$ ) to find the best couple  $(n, m)$ . The vertical lines in the supplementary Figs. S15 and S14 show the range of the results obtained from different 54 couples.

## 1.4 Mutational Signatures and Mutation Binary Vectors Extraction

In this part, we explain how to obtain mutational signatures and mutation binary vectors.

### 1.4.1 Mutational Signatures

There exists six classes of base substitution:

$$C > A, C > G, C > T, T > A, T > C, T > G \quad (4)$$

In addition, using the information of the 5' and 3' adjacent bases lead to 96 ( $= 4 \times 6 \times 4$ ) different mutation types. There are four possibilities for 3' and four possibilities for 5' adjacent bases. Firstly, We constructed Mutation matrix by extracting the number of occurrences of each mutation types for each patient. Mutation matrix,  $M$ , is a  $m \times s$  matrix with  $m = 96$  mutation types and  $s$  samples. Then, we tried to find the Mutation-Signature ( $P$ ) and Signature-Sample ( $E$ ) matrices by minimizing the following formula:

$$\min \|M - PE\|^2 \quad (5)$$

where,  $P$  and  $E$  are  $m \times k$  and  $k \times s$  matrices, respectively. Here,  $k$  is the number of mutational signatures. We performed this method by utilizing SigProfilerExtractor tool. We selected  $k = 30$  which gives us the minimum error.

### 1.4.2 Mutation Binary Vector

We calculated the number of occurrences of each mutations in a specific neuropathy type group. Then, according to the numbers, we ranked the mutations from high to low. Afterwards, we choose top 10,000 point mutation positions. Taking the positions, wild-type bases are extracted. We considered the wild-type bases as a reference sequence. To construct a mutation binary vector for each patient, we do the following procedure:

- Create an empty list with the size of the reference sequence length.
- If a point mutation occurs in the  $i^{th}$  position of the reference sequence, we assign 1 to the  $i^{th}$  position of the list; otherwise, 0.

As an example, suppose that ATGGCGTAA is the reference sequence and a patient's sequence is ACGGCGTTA. Then, the patient's mutation binary vector is 010000010.

## 1.5 Heatmap of the oxaliplatin and paclitaxel related genes

Several studies discuss the genome-wide association with oxaliplatin and paclitaxel (Argyriou et al., 2017; Reyes-Gibby et al., 2015). We extracted 9 oxaliplatin-related and 10 paclitaxel-related (totally 19) genes from these studies (Supplementary Table S5). In conjunction with these 19 genes, we also found 9 more genes which are clustered in the same group as paclitaxel-related genes and 4 genes which are clustered in the same group as oxaliplatin-related genes. Supplementary figure S19 illustrates a part of the score matrix heatmap which contains 32 genes consisting of 19 genes extracted from previous studies and 13 genes whose vectors are close to those 19 genes. The first 52 rows of the heatmap are the endometrial and ovarian cancer patients who received paclitaxel medication, whereas the bottom 105 rows represents colorectal cancer patients who received oxaliplatin medication. The columns of the heatmap signify the genes vectors in the score matrix which are grouped into two clusters:

- 1) paclitaxel- associated genes, and 2) oxaliplatin-associated genes.

The right side of Supplementary figure S19 represents genes whose pathogenic significance is higher in the paclitaxel-received patients than the oxaliplatin-received patients. As Supplementary Table S5 shows, EPHA4, EPHA8, XKR4, RRM1, ABCB1, CYP1B1, TUBB2A, RFX2, GSK3B, and FZD3 are genes which are significantly associated with paclitaxel, situated in the right-side cluster of the heatmap. For example, genetic variations in ABCB1 and CYP1B1 genes are involved in pharmacokinetic pathways of paclitaxel (Yamaguchi et al., 2006). On the other hand, the left-hand side of Supplementary figure S19 represents genes whose pathogenic significance is higher in the oxaliplatin-received patients than the paclitaxel-received patients. ERCC1, DPYD, GSTP1, FOXC1, POU2AF1, PELO, FARS2, ACYP2, and SCN4A are the oxaliplatin-related genes which are situated in the left-hand side cluster of the heatmap. For example, genetic variations in ERCC1, DPYD and GSTP1 genes are involved in pharmacokinetic pathways of chemotherapy with FOLFOX (Chang et al., 2009).

## 2 Supplementary Algorithms, Figures, and Tables

---

**Algorithm 1** Assign a score to each gene

---

```

1: function CoLASPGENESCORE(Gene g)
2:   gene_score  $\leftarrow -\infty$ , isPorLPexist  $\leftarrow \text{FALSE}$ 
3:   for each mutation m occurred in gene g do
4:     mutation_score  $\leftarrow \text{CoLASPMUTATIONSCORE}(m)
5:     if mutation_score  $\geq 2.5$  then
6:       if  $\neg \text{isPorLPexist}$  then
7:         gene_score  $\leftarrow \text{mutation_score}$ 
8:         isPorLPexist  $\leftarrow \text{TRUE}$ 
9:       else
10:        gene_score  $\leftarrow \text{gene_score} + \text{mutation_score}$ 
11:     else
12:       if  $(\neg \text{isPorLPexist}) \wedge (\text{mutation_score} > \text{gene_score})$  then
13:         gene_score  $\leftarrow \text{mutation_score}$ 
return gene_score$ 
```

---

**Algorithm 2** Assign a score to each mutation

---

```

1: function CoLASPMUTATIONSCORE(Mutation m)
2:    $\vec{PV}S, \vec{PS}, \vec{PM}, \vec{PP}, \vec{BP}, \vec{BS}, \vec{BA}, \text{InterVar\_type} \leftarrow \text{INTERVAR}(m)$ 
3:   Extract ClinVar_type, SIFT_type, and CADD_type from InterVar result
4:   ClinVar_value  $\leftarrow \text{GETCOLASPVALUE}(\text{CLINVAR}(m), "ClinVar", \text{ClinVar\_type})$ 
5:   SIFT_value  $\leftarrow \text{GETCOLASPVALUE}(\text{SIFT}(m), "SIFT", \text{SIFT\_type})$ 
6:   CADD_value  $\leftarrow \text{GETCOLASPVALUE}(\text{CADD}(m), "CADD", \text{CADD\_type})$ 
7:   CoLaSp_value  $\leftarrow \text{CALCULATECOLASPVALUE}(\vec{PV}S, \vec{PS}, \vec{PM}, \vec{PP}, \vec{BP}, \vec{BS}, \vec{BA}, \text{InterVar\_type})$ 
8:   if  $(\text{SIFT\_value} \neq -\infty) \wedge (\text{CADD\_value} \neq -\infty)$  then
9:     return  $\max\{\text{CoLaSp\_value}, \text{ClinVar\_value}, \text{SIFT\_value}, \text{CADD\_value}\}$ 
return  $\max\{\text{CoLaSp\_value}, \text{ClinVar\_value}\}$ 
```

---

**Algorithm 3** CoLaSp value is based on ACMG-AMP guidelines

---

```

1: function CALCULATECOLASPVALUE( $\vec{PV}S, \vec{PS}, \vec{PM}, \vec{PP}, \vec{BP}, \vec{BS}, \vec{BA}, \text{InterVar\_type}$ )
2:   if InterVar_type = "Uncertain Significance" then
3:     return 0
4:   pvs  $\leftarrow 4.165$ , ps  $\leftarrow 2.6$ , pm  $\leftarrow 0.835$ , pp  $\leftarrow 0.45$ 
5:   bp  $\leftarrow -0.7$ , bs  $\leftarrow -3.7$ , ba  $\leftarrow -5.5$ 
6:    $\vec{R} \leftarrow \text{CONCATENATE}(\text{pvs} \cdot \vec{PV}S, \text{ps} \cdot \vec{PS}, \text{pm} \cdot \vec{PM}, \text{pp} \cdot \vec{PP}, \text{bp} \cdot \vec{BP}, \text{bs} \cdot \vec{BS}, \text{ba} \cdot \vec{BA})$ 
7:   CoLaSp_value  $\leftarrow 0$ 
8:   for each value in  $\vec{R}$  do
9:     CoLaSp_value  $\leftarrow \text{CoLaSp\_value} + \text{value}$ 
return CoLaSp_value
```

---

---

**Algorithm 4** To convert other interpreters scopes to CoLaSp scope

---

```
1: function GETCOLASPVALUE(interpreter_score, interpreter, type)
2:   if interpreter_score =  $-\infty$  then                                 $\triangleright$  When the score is unavailable
3:     return  $-\infty$ 
4:   if interpreter = "ClinVar" then
5:     if type = "Pathogenic" then
6:       return 5
7:     else if type = "Likely Pathogenic" then
8:       return 2.5
9:     else if type = "Likely Benign" then
10:      return -5
11:    else if type = "Benign" then
12:      return -15
13:    else
14:      return 0
15:   if type = "Pathogenic" then
16:     CoLaSp_lb  $\leftarrow$  5, CoLaSp_ub  $\leftarrow$  15
17:     SIFT_lb  $\leftarrow$  -0.025, SIFT_ub  $\leftarrow$  0
18:     CADD_lb  $\leftarrow$  20, CADD_ub  $\leftarrow$  40
19:   else if type = "Likely Pathogenic" then
20:     CoLaSp_lb  $\leftarrow$  2.5, CoLaSp_ub  $\leftarrow$  5
21:     SIFT_lb  $\leftarrow$  -0.05, SIFT_ub  $\leftarrow$  -0.025
22:     CADD_lb  $\leftarrow$  15, CADD_ub  $\leftarrow$  20
23:   else if type = "Benign" then
24:     CoLaSp_lb  $\leftarrow$  -15, CoLaSp_ub  $\leftarrow$  -5
25:     SIFT_lb  $\leftarrow$  -1.0, SIFT_ub  $\leftarrow$  -0.525
26:     CADD_lb  $\leftarrow$  0, CADD_ub  $\leftarrow$  10
27:   else
28:     CoLaSp_lb  $\leftarrow$  -5, CoLaSp_ub  $\leftarrow$  -1
29:     SIFT_lb  $\leftarrow$  -0.525, SIFT_ub  $\leftarrow$  -0.05
30:     CADD_lb  $\leftarrow$  10, CADD_ub  $\leftarrow$  15
31:   if interpreter = "CADD" then
32:     return  $\frac{(\text{interpreter\_score} - \text{CADD\_lb})(\text{CoLaSp\_ub} - \text{CoLaSp\_lb})}{(\text{CADD\_ub} - \text{CADD\_lb})} + \text{CoLaSp\_lb}$ 
33:   return  $\frac{(\text{interpreter\_score} - \text{SIFT\_lb})(\text{CoLaSp\_ub} - \text{CoLaSp\_lb})}{(\text{SIFT\_ub} - \text{SIFT\_lb})} + \text{CoLaSp\_lb}$ 
```

---

- A Pathogenic**
    - o 1 Very strong (**PVS1**) AND
      - o ≥1 Strong (**PS1 – PS4**) OR
      - o ≥2 Moderate (**PM1 – PM6**) OR
      - o 1 Moderate (**PM1 – PM6**) AND Supporting (**PP1 – PP5**) OR
      - o ≥2 Supporting (**PP1 – PP5**)
    - o ≥2 Strong (**PS1 – PS4**) OR
    - o 1 Strong (**PS1 – PS4**) AND
      - o ≥3 Moderate (**PM1 – PM6**) OR
      - o 2 Moderate (**PM1 – PM6**) AND ≥2 Supporting (**PP1 – PP5**) OR
      - o 1 Moderate (**PM1 – PM6**) AND ≥4 Supporting (**PP1 – PP5**)
  
  - C Benign**
    - o 1 Stand-alone (**BA1**) AND
    - o ≥2 Strong (**BS1 – BS4**)
  
  - B Likely pathogenic**
    - o 1 Very strong (**PVS1**) AND 1 Moderate (**PM1 – PM6**) OR
    - o 1 Strong (**PS1 – PS4**) AND 1-2 Moderate (**PM1 – PM6**) OR
    - o 1 Strong (**PS1 – PS4**) AND ≥2 Supporting (**PP1 – PP5**) OR
    - o ≥3 Moderate (**PM1 – PM6**) OR
    - o 2 Moderate (**PM1 – PM6**) AND ≥2 Supporting (**PP1 – PP5**) OR
    - o 1 Moderate (**PM1 – PM6**) AND ≥4 Supporting (**PP1 – PP5**)
  
  - D Likely benign**
    - o 1 Strong (**BS1 – BS4**) AND 1 Supporting (**BP1 – BP7**) OR
    - o ≥2 Supporting (**BP1 – BP7**)
  
  - E Uncertain Significance**
    - o Other criteria shown above are not met
    - o The other criteria for benign and pathogenic are contradictory

Figure 1: ACMG-AMP guidelines

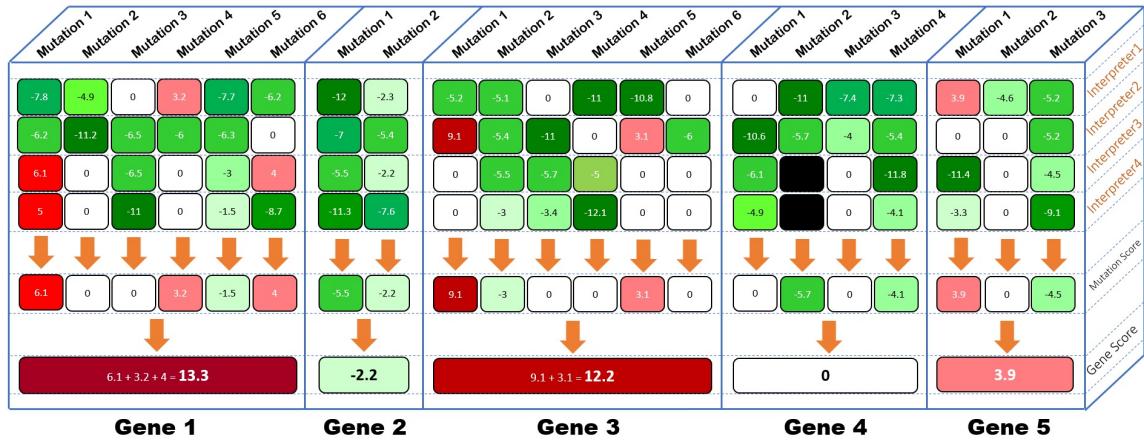


Figure 2: An example of the scoring system. Different number of mutations occur in each gene. Each mutation interpreters measure the degree of pathogenicity of each mutation in every patient. Red, green, white, and black blocks illustrate pathogenic/likely pathogenic, benign/likely benign, uncertain significance, and unavailable information, respectively.

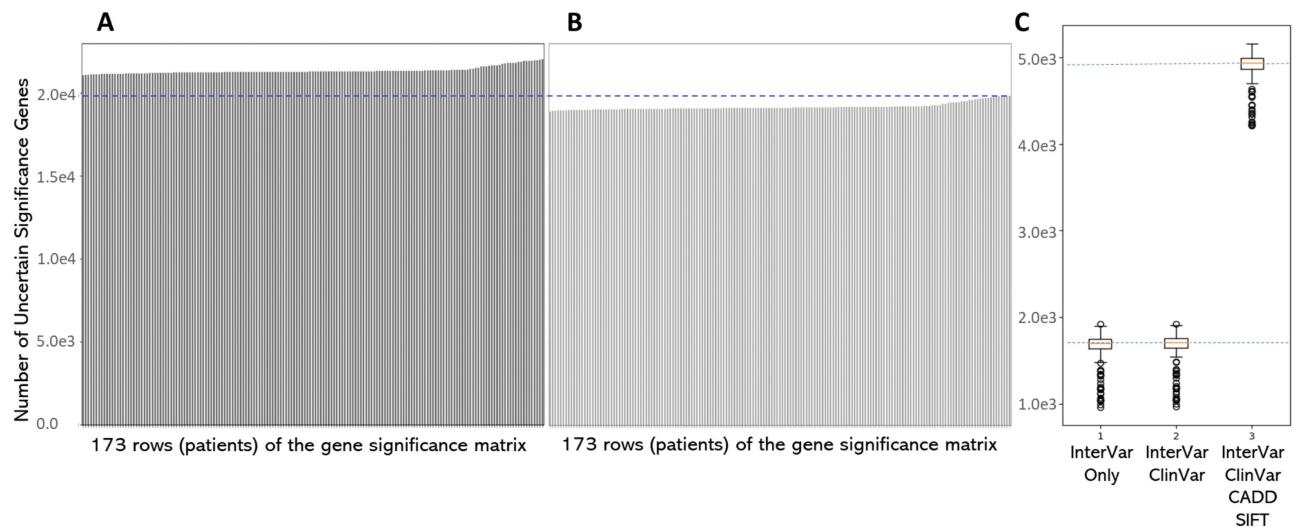


Figure 3: The number of genes in each patient which our method recognized them as the uncertain significance for 173 WGS patients when we use (A) InterVar only (B) InterVar, ClinVar, SIFT, and CADD as mutations interpreter(s). (C) The boxplots of the non-VUS mutations when we use (1) InterVar (2) InterVar and ClinVar (3) InterVar, ClinVar, SIFT, and CADD.

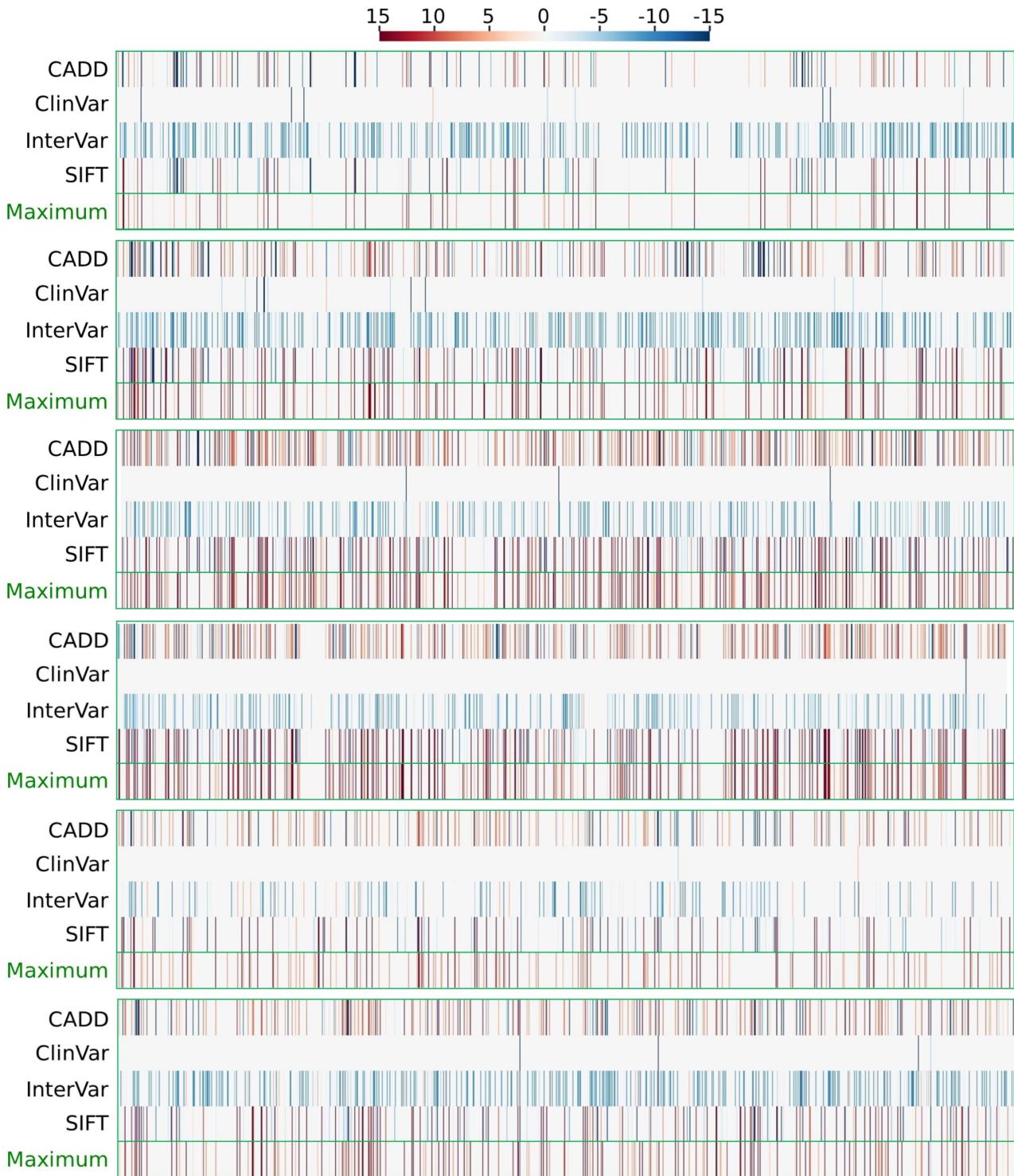


Figure 4: Fluctuations between the variant interpreters for individual variants in six randomly-selected TCGA samples. Blue, red, and white areas signify (likely) benign, (likely) pathogenic, and uncertain significance variants. SMFM and CoLaSp models accept the maximum pathogenicity score of each variant that is reported by different variant interpreters.

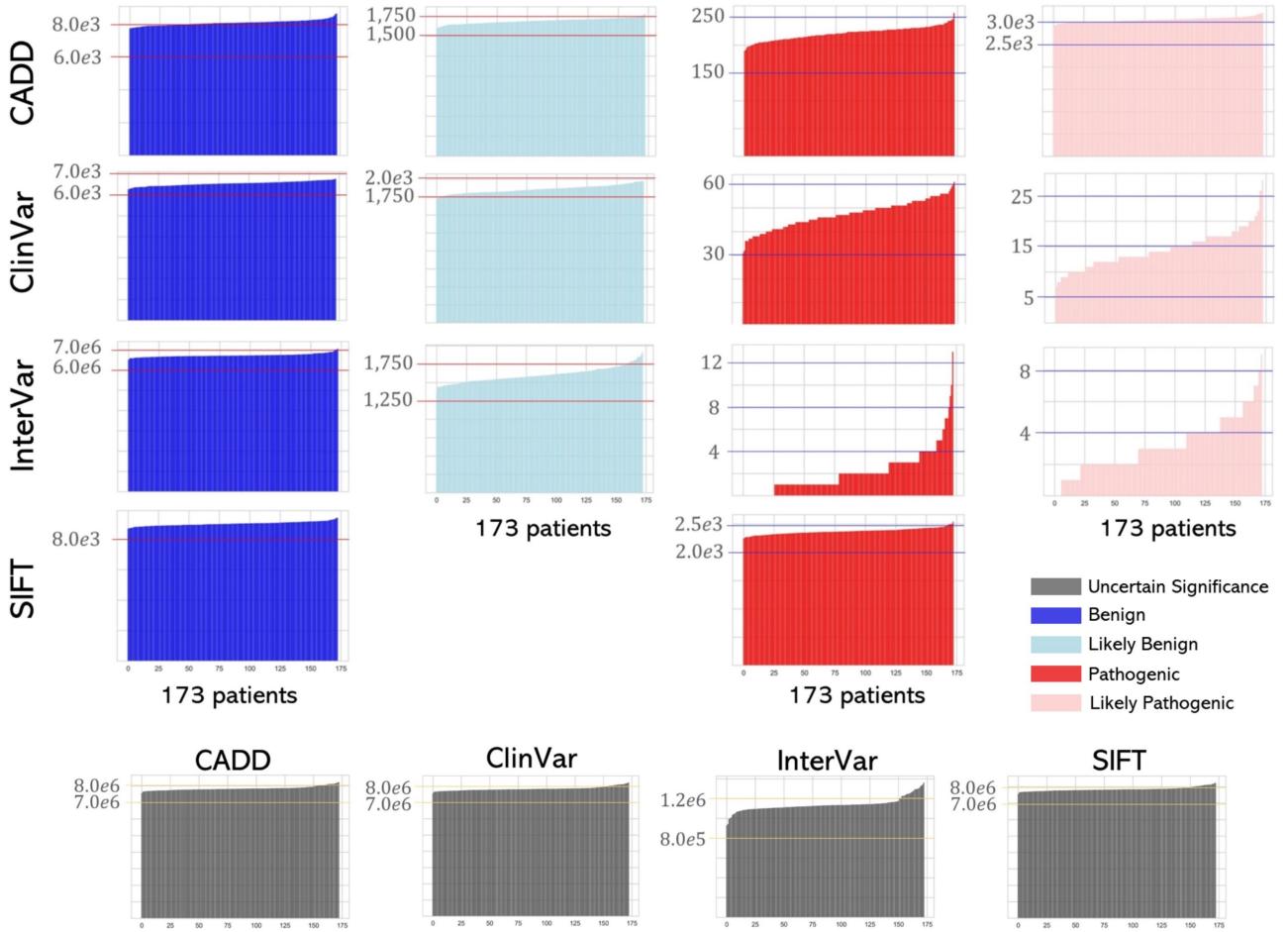


Figure 5: The number of mutations identified as Benign, Likely Benign, Variant of Uncertain Significance, Likely Pathogenic, Pathogenic for 173 patients WGS data using CADD, ClinVar, InterVar and SIFT.

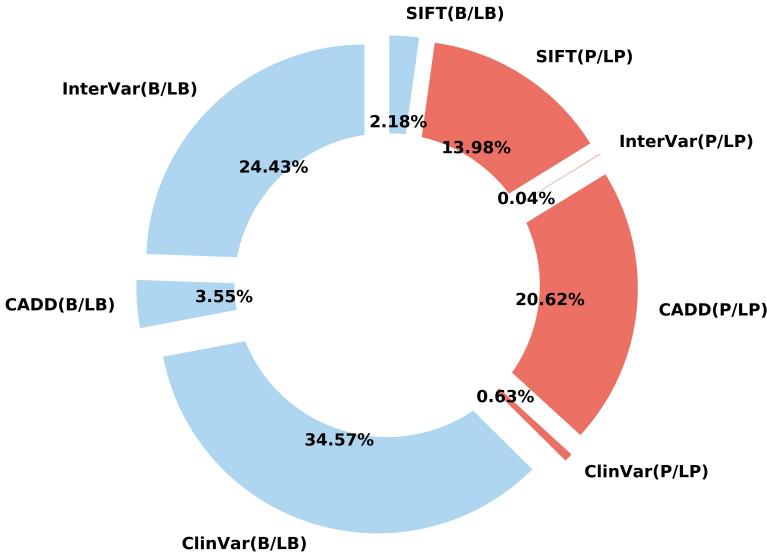


Figure 6: The contribution of each mutation interpreters in scoring the mutations. Four mutation interpreters assign different scores to a mutation. The proposed algorithm chooses the highest score (more pathogenic). This figure shows the interpreters' contribution to assign the highest score (among the four interpreters) to the mutations. Moreover, it shows the category (Pathogenic/Likely Pathogenic or Benign/Likely Benign) of the highest scores.

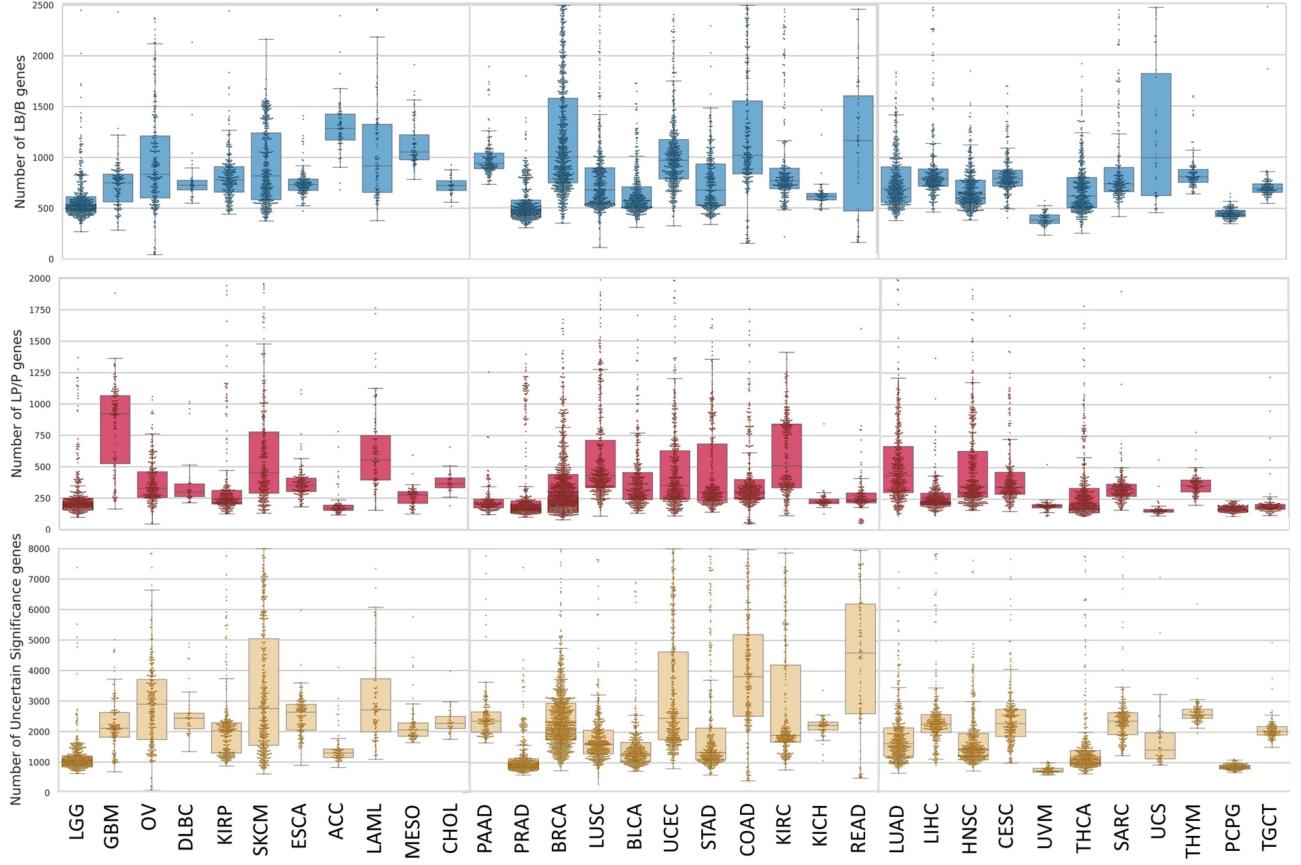


Figure 7: The distributions of the number of benign/likely benign (top), pathogenic/likely pathogenic (middle), and uncertain significance (bottom) genes detected by Supplementary algorithms S1-S4 for 33 different cancer types in TCGA cohort.

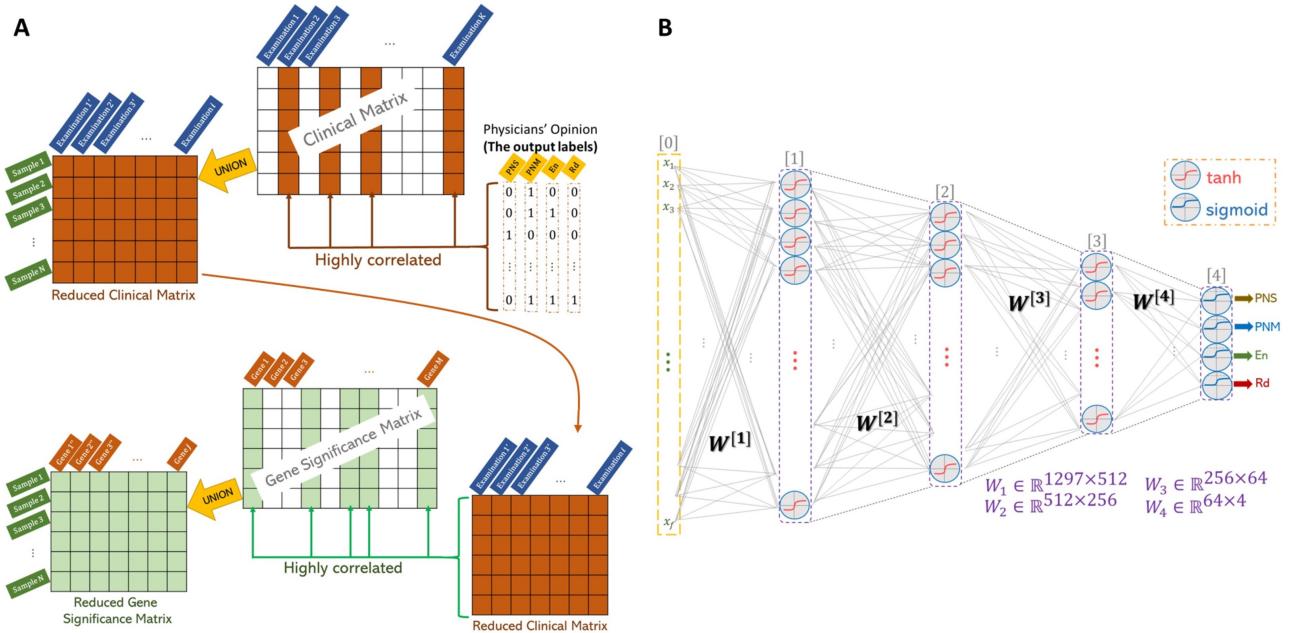


Figure 8: (A) The process of the gene selection. (B) The feedforward neural network (FFNN) architecture.

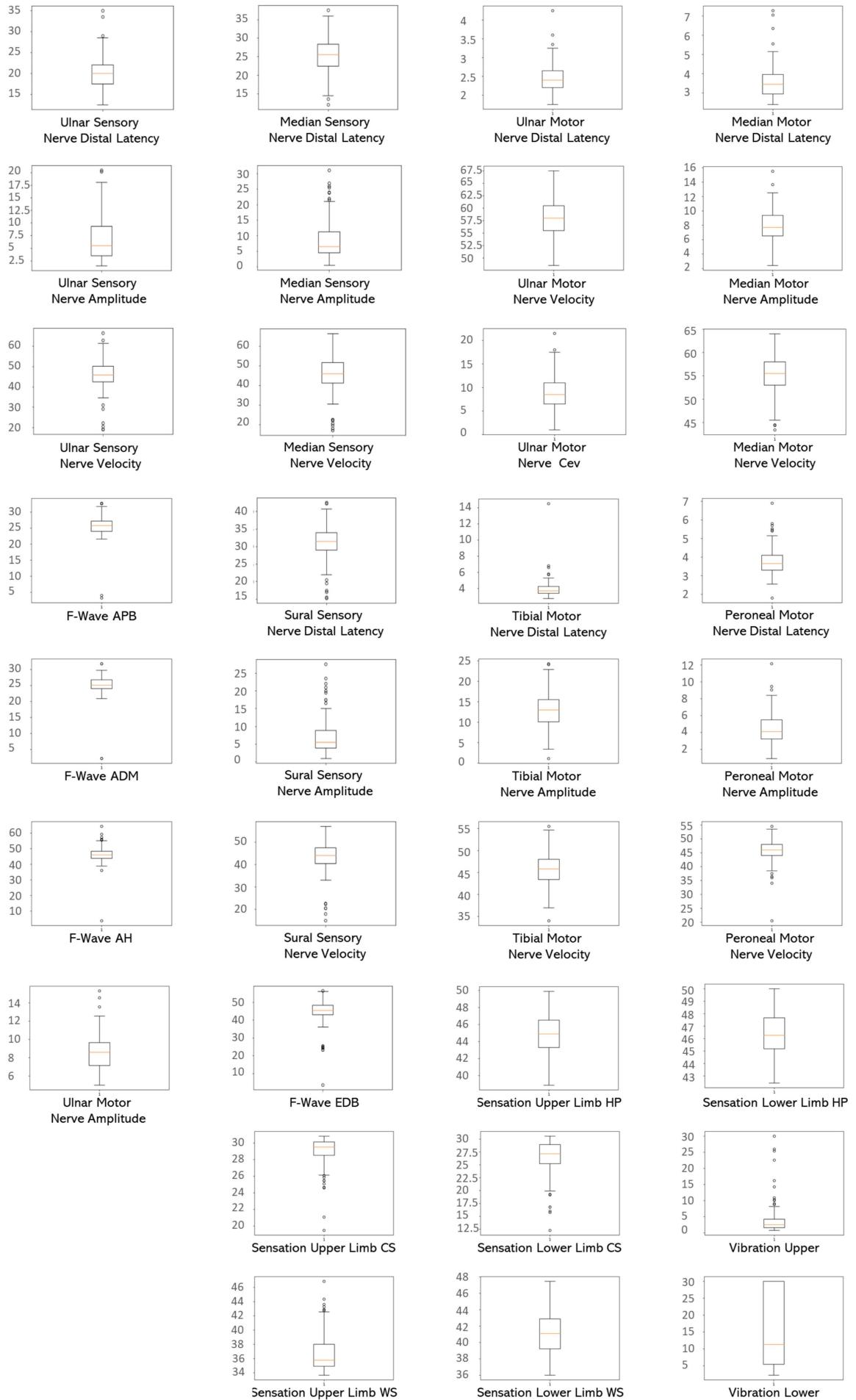


Figure 9: Boxplots of the values for 34 examinations for 173 patients.

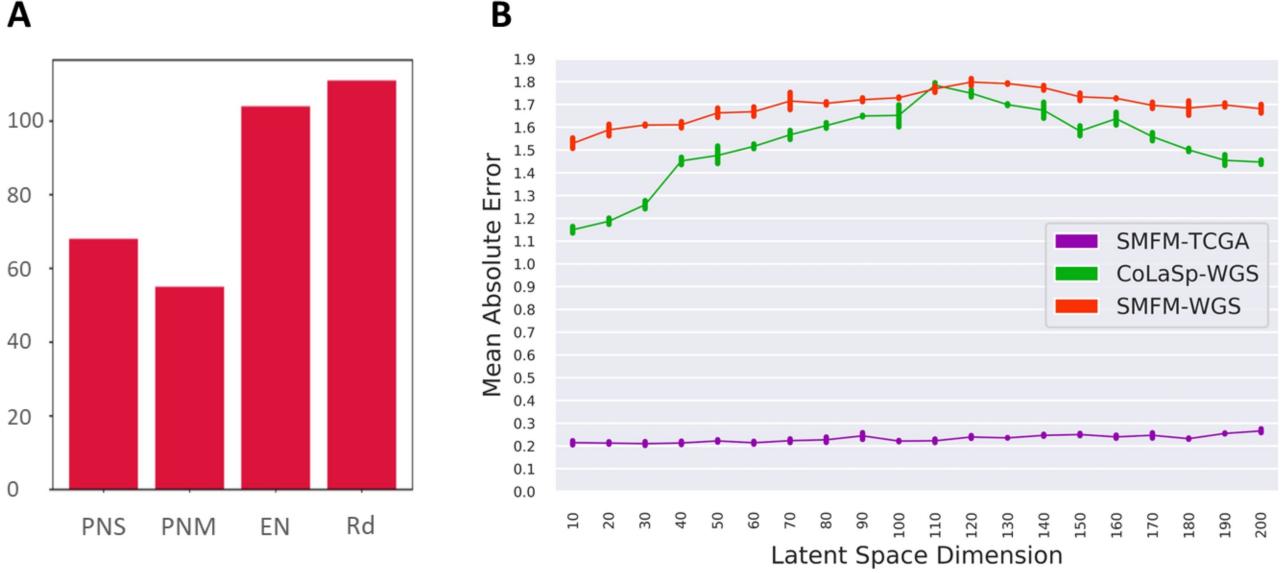


Figure 10: (A) The bar chart of the population for each output label. (B) Mean absolute error (MAE) comparison of CoLaSp and SMFM models over WGS and TCGA data.

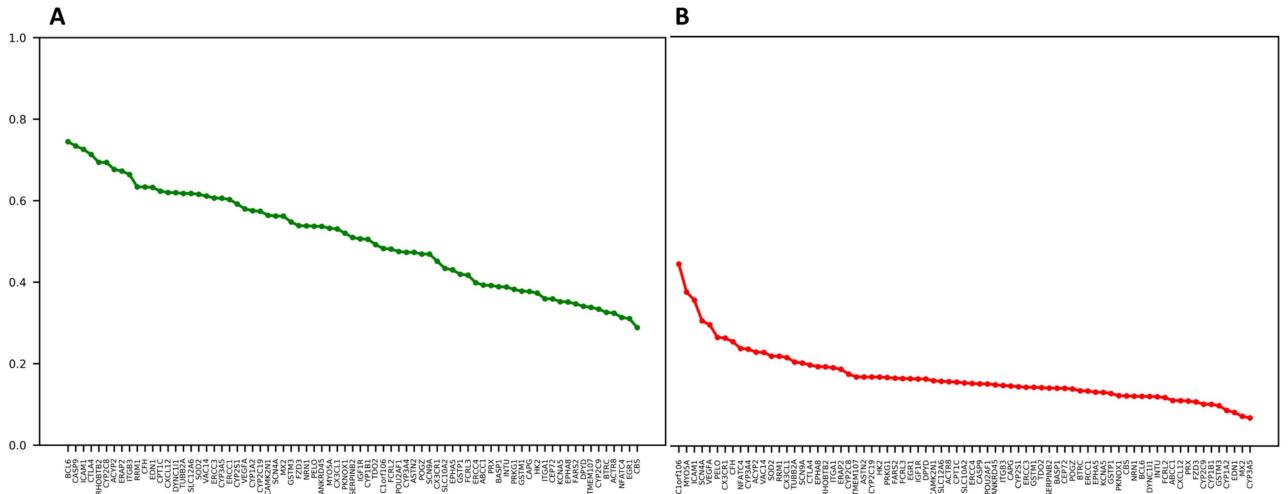


Figure 11: Pearson correlation values of the CIPN-related gene vectors and the output labels predicted by (A) CoLaSp (B) SMFM.

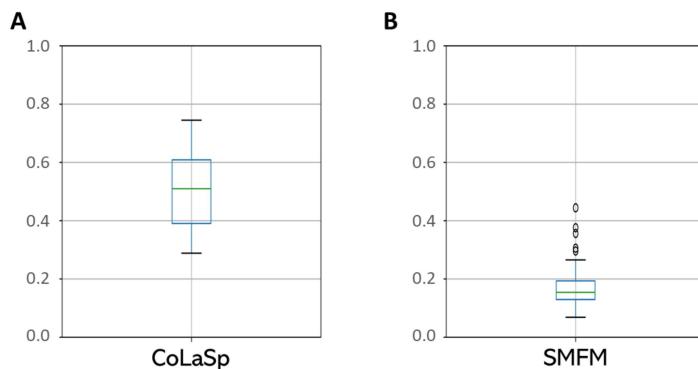
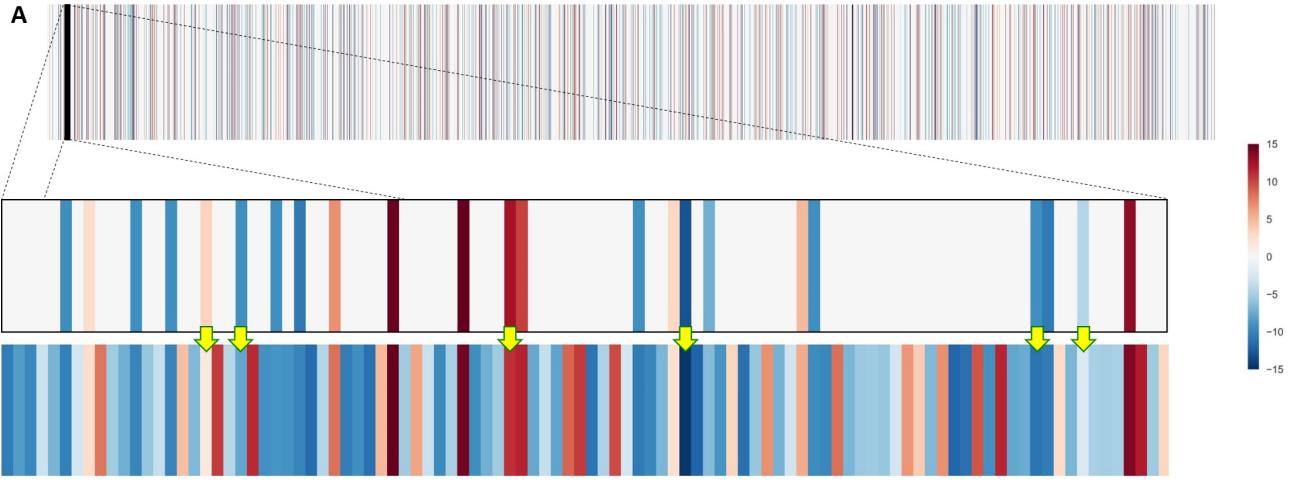
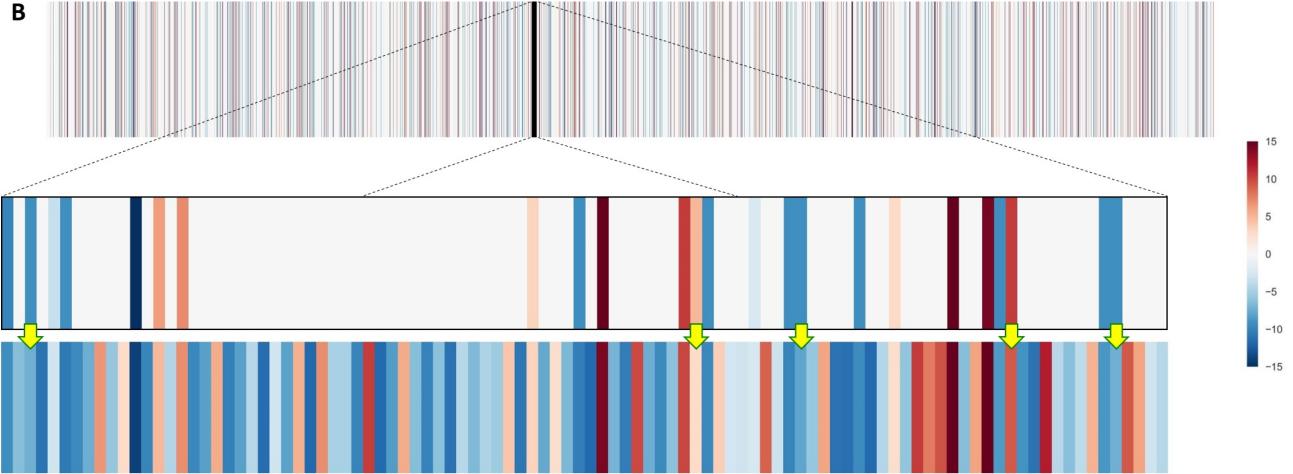


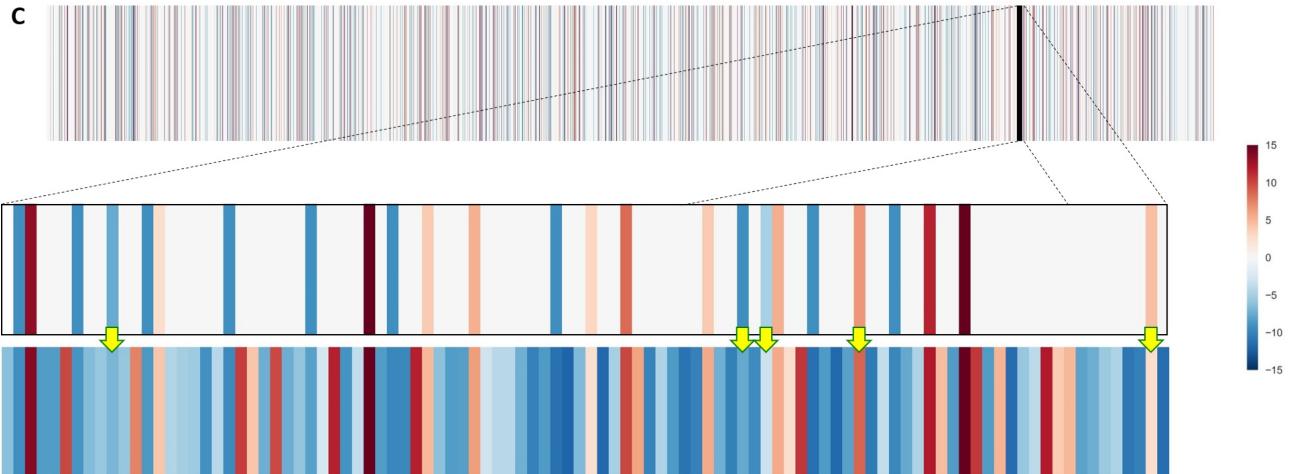
Figure 12: The correlation of 75 CIPN-related genes with the output labels boxplots in two cases: (A) CoLaSp and (B) SMFM



(a) 100 genes are selected randomly - Left part.



(b) 100 genes are selected randomly - Middle part.



(c) 100 genes are selected randomly - Right part.

Figure 13: A row (patient) extracted from the score matrix. The colors represent the pathogenicity of the genes in the patient. The values are between -15 (benign) and +15 (pathogenic). We divided this row into three parts and from each part we selected 100 genes randomly. The yellow arrows show the SMFM prediction results for the test set. The white cells (with value '0') represent uncertain significance and the rest of the cells belong to the training set. The top and bottom boxes in each subfigures represent the real and the predicted data, respectively.

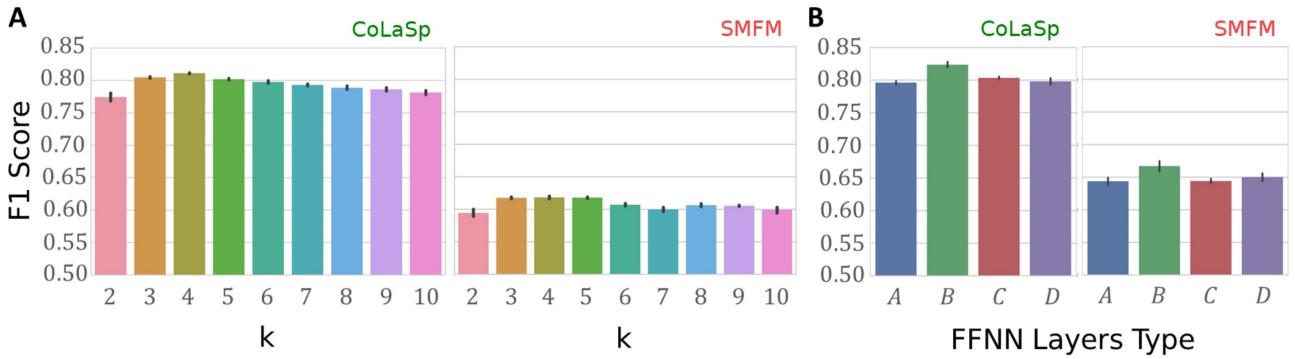


Figure 14: The bar charts compare the results of two cases: SMFM and CoLaSp. The bar charts show the *F1\_score* results of (A) the MLkNN classifier with different  $k$  values. Since MLkNN is a  $k$ -nearest-neighbor-based classifier,  $k = 2, \dots, 10$  represents the number of neighbors of each input instance. Each of the vertical lines in this figure illustrates the range of the results obtained from 54 different gene selection methods (Supplementary Information 1.4). The results are the mean of the 5-fold cross-validation. The figure shows that  $k = 4$  is the best parameter which gives the highest *F1\_score* result. (B) the feedforward neural network with different hyper-parameters. The x axis represents four different number of neurons in each layer of the feedforward neural network. "A" represents 1024, 256, and 32 neurons in first, second, and third layers, respectively. "B" represents 512-256-64. "C" represents 1024-512-64. "D" represents 512-64-32. Each of the vertical lines in this figure illustrates the range of the results obtained from 54 different gene selection methods (Supplementary Information 1.4). The results are the mean of the 5-fold cross-validation. The figure shows "B" is the best number of neurons in each layer of the feedforward neural network.

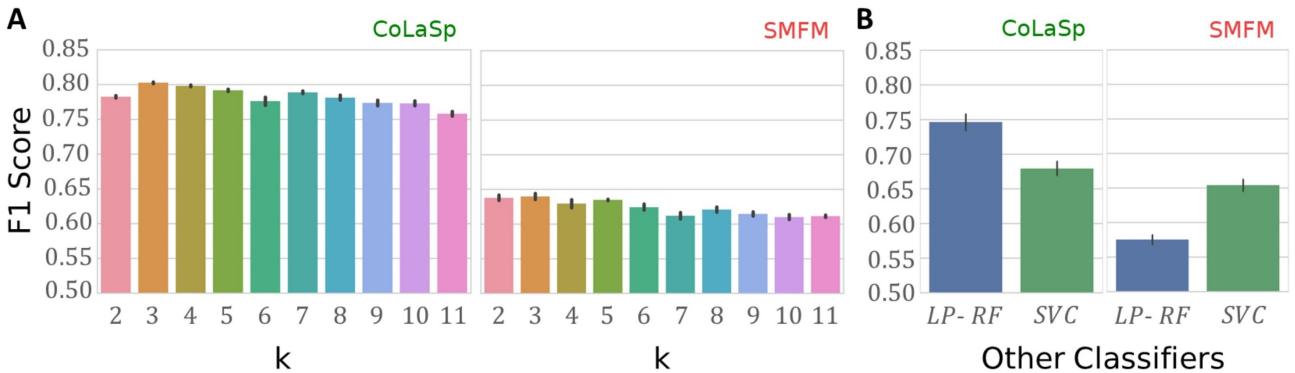


Figure 15: The bar charts compare the results of two cases: SMFM and CoLaSp. The bar charts show the *F1\_score* results of (A) the BRkNN classifier with different  $k$  values. Since BRkNN is a  $k$ -nearest-neighbor-based classifier,  $k = 2, \dots, 11$  represents the number of neighbors of each input instance. Each of the vertical lines in this figure illustrates the range of the results obtained from 54 different gene selection methods (Supplementary Information 1.4). The results are the mean of the 5-fold cross-validation. The figure shows that  $k = 3$  is the best parameter which gives the highest *F1\_score* result. (B) The bar charts compare the results of the SVC and LP-RF classifiers in two cases: SMFM and CoLaSp. Each of the vertical lines in this figure illustrates the range of the results obtained from 54 different gene selection methods (Supplementary Information 1.4). The results are the mean of the 5-fold cross-validation.

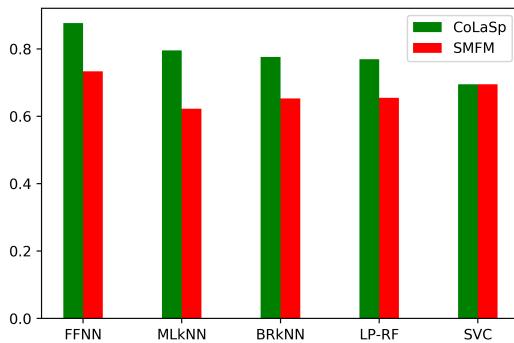


Figure 16: The AUC bar chart. Comparison of the five multi-label classifiers in two cases of using SMFM and CoLaSp.

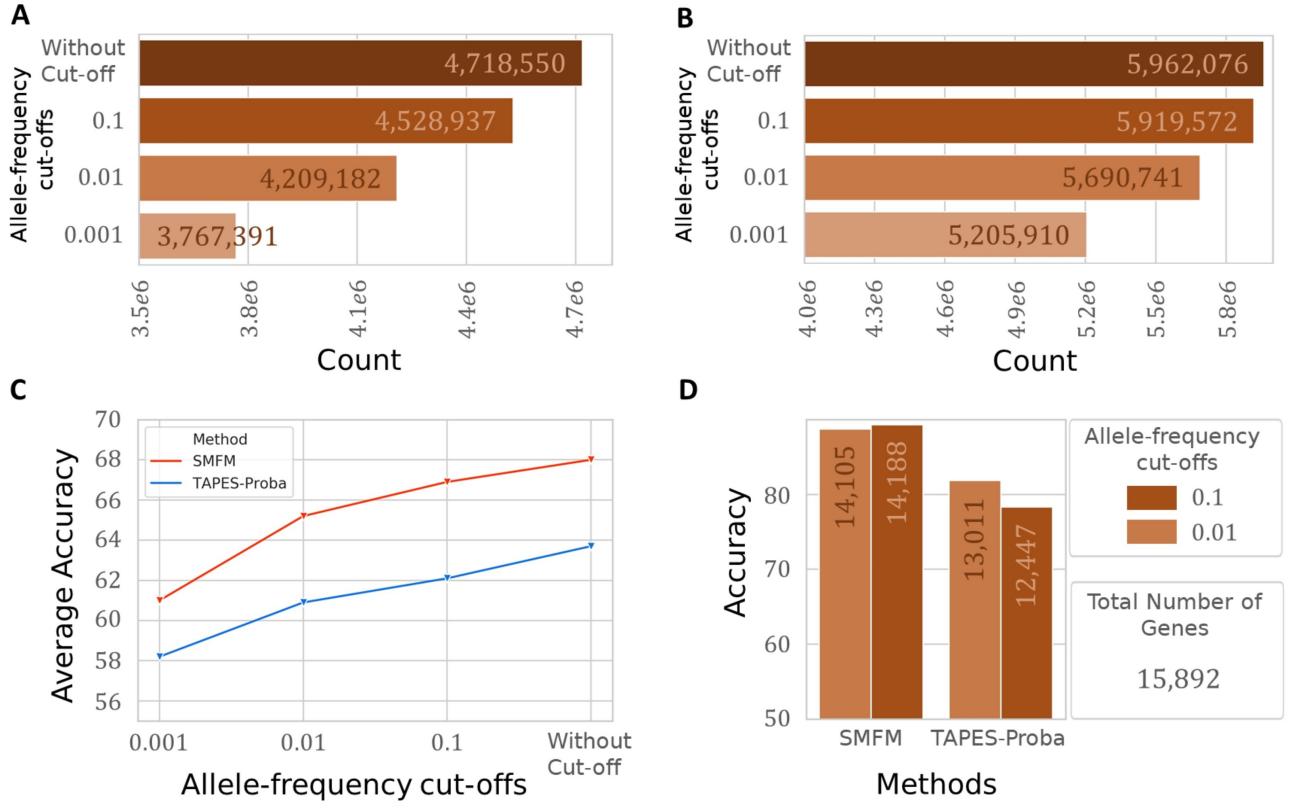


Figure 17: (A) Total number of **pathogenic genes** identified by our model in the TCGA samples with different allele frequency cut-offs. (B) Total number of **pathogenic variants** identified by our model in all the TCGA samples with different allele frequency thresholds. (C) Comparing the mean of all cancer type predictions' accuracy in the TCGA samples in two cases of using SMFM and TAPES-Proba as input features. (D) The percentage of identical calls between the average SMFM (and TAPES-Proba) score of the genes with RVIS scores (with different allele frequency cut-offs) and their comparison.

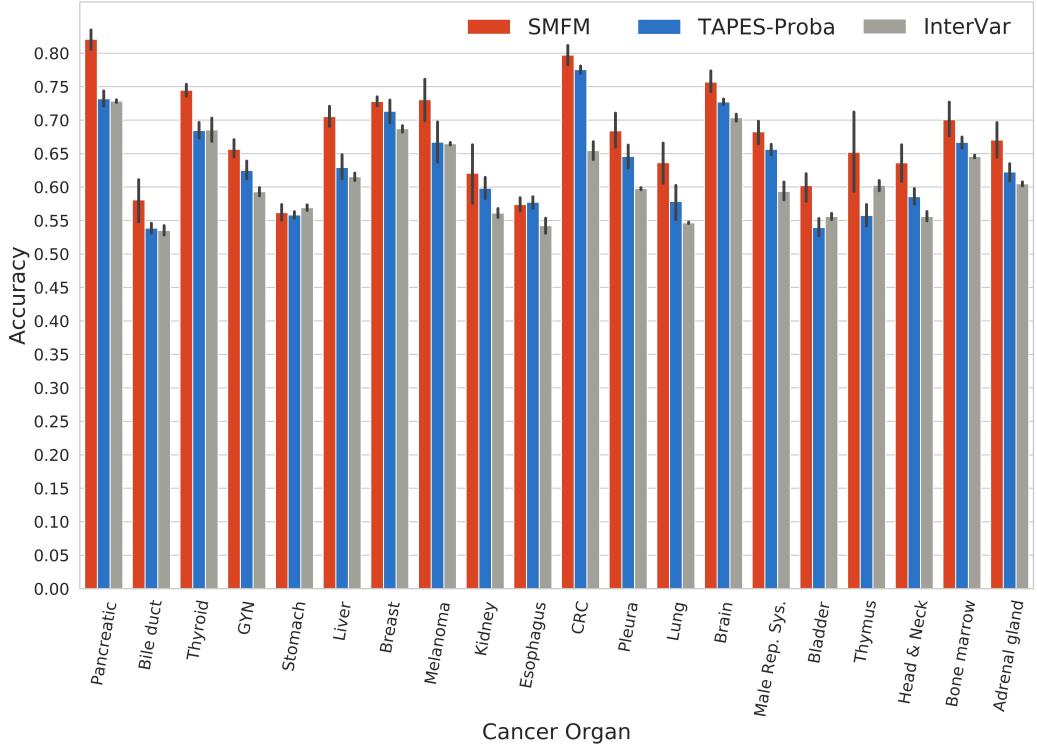


Figure 18: Cancer type prediction accuracy comparison in three cases of using SMFM, TAPES-Proba, and InterVar as input features of the feedforward neural network binary classifier.

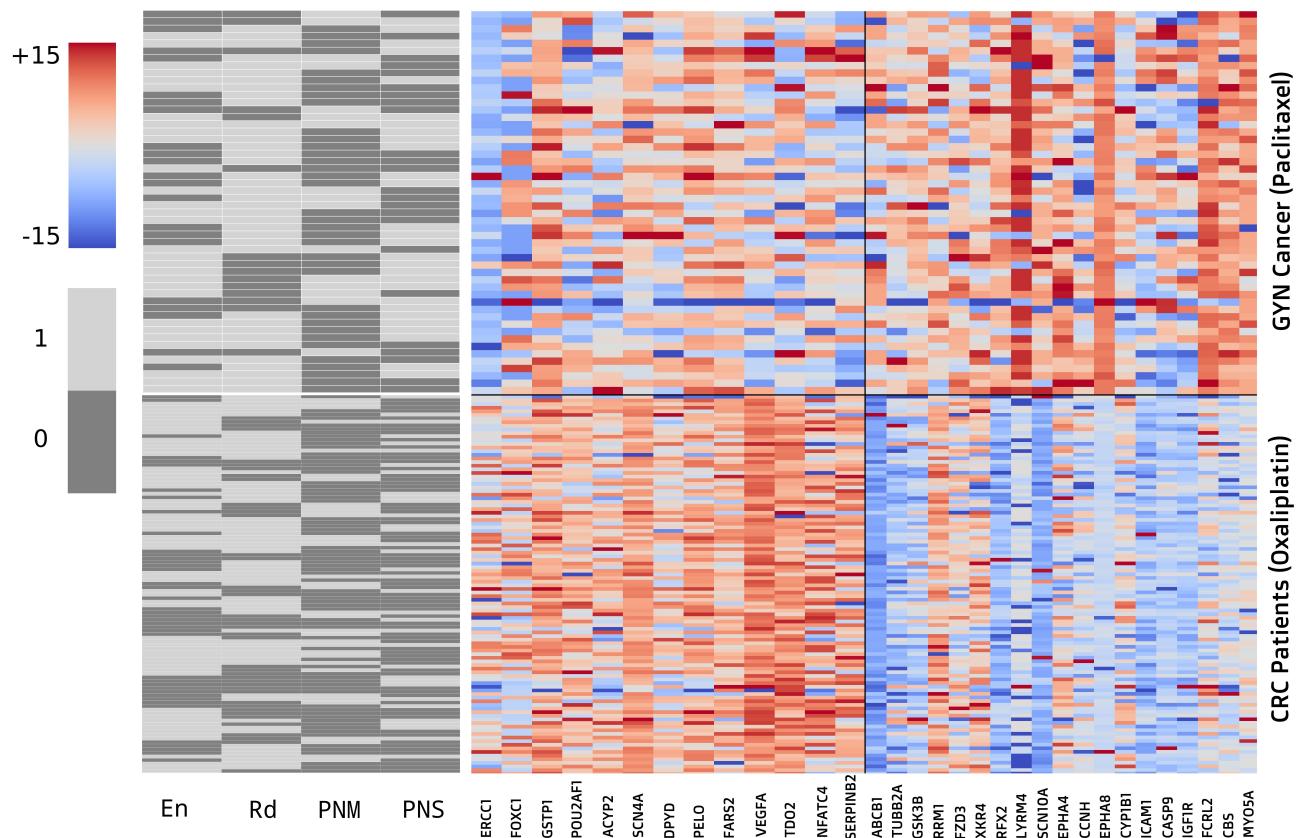


Figure 19: Heatmap of the genes associated with paclitaxel (right) and oxaliplatin (left) for colorectal cancer patients (bottom) and endometrial/ovarian cancer patients (top). Moreover, this figure shows the activated neuropathy types (Sensory Polyneuropathy (PNS), Motor Polyneuropathy (PNM), Entrapment Syndrome (En), and Radiculopathy (Rd)) in each patient.

Table 1: TCGA Whole Exome Sequencing dataset.

Number	Cancer Type	Population	Organ	Number	Cancer Type	Population	Organ
<b>1</b>	ACC	79	Adrenal Gland	<b>18</b>	LUSC	494	Lung
<b>2</b>	BLCA	406	Bladder	<b>19</b>	MESO	82	Pleura
<b>3</b>	BRCA	1038	Breast	<b>20</b>	OV	279	GYN
<b>4</b>	CESC	302	GYN	<b>21</b>	PAAD	175	Pancreatic
<b>5</b>	CHOL	36	Bile duct	<b>22</b>	PCPG	179	Adrenal Gland
<b>6</b>	COAD	427	CRC	<b>23</b>	PRAD	495	Male Rep. Sys.
<b>7</b>	DLBC	37	Bone marrow	<b>24</b>	READ	153	CRC
<b>8</b>	ESCA	163	Esophagus	<b>25</b>	SARC	253	Other
<b>9</b>	GBM	157	Brain	<b>26</b>	SKCM	468	Melanoma
<b>10</b>	HNSC	497	Head and Neck	<b>27</b>	STAD	379	Stomach
<b>11</b>	KICH	66	Kidney	<b>28</b>	TGCT	134	Male Rep. Sys.
<b>12</b>	KIRC	335	Kidney	<b>29</b>	THCA	491	Thyroid
<b>13</b>	KIRP	286	Kidney	<b>30</b>	THYM	118	Thymus
<b>14</b>	LAML	116	Bone marrow	<b>31</b>	UCEC	538	GYN
<b>15</b>	LGG	507	Brain	<b>32</b>	UCS	56	GYN
<b>16</b>	LIHC	369	Liver	<b>33</b>	UVM	80	Melanoma
<b>17</b>	LUAD	513	Lung				

Table 2: InterVar-ClinVar interpretations of 173 WGS patients.

InterVar		ClinVar				
		Pathogenic	Likely Pathogenic	Likely Benign	Benign	VUS
InterVar	Pathogenic	48	16	0	0	300
	Likely Pathogenic	150	19	1	0	376
	Likely Benign	397	120	21,581	11,712	243,789
	Benign	6,547	2,083	298,975	1,119,586	1,160,953,069
	VUS	1,029	245	999	724	196,736,964

Table 3: This table helps us to know when and which variant interpreters to use.

Interpreter	Uses ACMG-AMP?	Predicts Uncertain Significance?	Uses Clinical data?	Description
ClinVar	No	No	No	A manual corrected database that records pathogenicity of variants.
SIFT	No	No	No	A prediction tool that considers the sequence homology and the physical properties of amino acids.
CADD	No	No	No	A tool for prioritizing deleterious and disease causal variants.
InterVar	Yes	No	No	An automated process that classifies variants by the ACMG-AMP guidelines.
TAPES	Yes	No	No	Uses and transforms the ACMG-AMP guidelines into a continuous probabilities.
SMFM	Yes	Yes	No	A machine learning-based tool that utilizes four variant interpreters to predict the uncertain significance genes.
CoLaSp	Yes	Yes	Yes	A machine learning-based tool that utilizes four variant interpreters and relevant medical examination data to predict the uncertain significance genes.

Table 4: CIPN-related genes reported by different studies.

<b>Gene Name</b>	<b>Correlation</b>	<b>Highly correlated</b>	<b>Highly correlated (examinations)</b>
BCL6	0.744651	PNM,PNS	MedSenNerVel, UlnSenNerVel, MedMotNerVel, UlnSenNerLat
CASP9	0.734028	PNM,PNS	Fwave-AH, TibMotNerVel, UlnMotNerVel
ICAM1	0.725844	PNS,En	MedMotNerLat, MedMotNerVel
CTLA4	0.713194	PNM,PNS	Fwave-AH, TibMotNerLat, UlnMotNerVel, TibMotNerVel
RHOBTB2	0.693988	PNS,En	MedSenNerAmp,SurSenNerVel
CYP2C8	0.69375	PNM,PNS	UlnMotNerVel
ACYP2	0.676479	PNS	SurSenNerVel
ERAP2	0.672513	PNS	MedSenNerAmp
ITGB3	0.663902	PNS	MedMotNerLat
RRM1	0.633799	PNM	MedMotNerVel, MedSenNerVel
CFH	0.633378	PNM	Fwave-AH, TibMotNerVel, MedMotNerVel, PerMotNerVel
EDN1	0.632546	PNS	UlnSenNerVel, UlnSenNerLat
CPT1C	0.623224	PNM	MedSenNerAmp, Fwave-AH
CXCL12	0.619722	PNM,PNS,En	TibMotNerVel, MedMotNerVel
DYNC1I1	0.61943	PNM	VibLower
TUBB2A	0.617653	PNS	SurSenNerVel
SLC12A6	0.617557	En	MedSenNerVel
SOD2	0.615553	En	MedMotNerLat
VAC14	0.611112	PNS	SurSenNerVel, UlnMotNerVel
ERCC3	0.606424	PNM	Fwave-AH, TibMotNerVel
CYP3A5	0.605989	PNM,PNS	Fwave-AH
ERCC1	0.602651	PNM,PNS	UlnSenNerVel, MedSenNerVel
CYP2S1	0.591629	PNM	Fwave-AH
VEGFA	0.579864	PNM,PNS	TibMotNerLat
CYP1A2	0.575128	PNM,PNS	Fwave-AH, TibMotNerVel
CYP2C19	0.573863	PNM	Fwave-AH
CAMK2N1	0.564034	PNM	Fwave-AH, UlnSenNerLat, MedMotNerLat
SCN4A	0.562204	PNS,Rd	UlnSenNerLat
MX2	0.561947	PNM,En	MedMotNerLat
GSTM3	0.547611	Rd	MedSenNerVel
FZD3	0.53864	PNM,PNS	VibLower, TibMotNerLat
NRN1	0.537943	PNM,PNS,En	PerMotNerVel
PELO	0.537228	PNM,PNS	TibMotNerVel, Fwave-AH
ANKRD45	0.53691	Rd	UlnSenNerLat, MedSenNerVel
MYO5A	0.532179	PNM	TibMotNerLat
CX3CL1	0.53051	PNM,PNS,En	VibLower
PKNOX1	0.520167	PNS	UlnSenNerVel
SERPINB2	0.50947	PNS	SurSenNerVel
IGF1R	0.506265	En,Rd	Fwave-AH
CYP1B1	0.504919	PNM,PNS	MedSenNerAmp, SurSenNerVel
TDO2	0.491916	PNM,Rd	MedMotNerLat
C1orf106	0.482388	PNS	SurSenNerVel
FCRL2	0.481091	PNM,PNS	SurSenNerVel
POU2AF1	0.475071	PNM	PerMotNerAmp
CYP3A4	0.473119	Rd	UlnSenNerLat
ASTN2	0.473082	Rd	UlnSenNerLat, MedSenNerVel
POGZ	0.468962	En	MedSenNerAmp
SCN9A	0.46895	PNM,PNS	MedSenNerAmp
CX3CR1	0.451255	PNM	TibMotNerLat, MedMotNerVel
SLC10A2	0.433804	En,Rd	VibLower
EPHA5	0.430081	PNM,PNS,En	VibLower
GSTP1	0.419169	PNS	MedSenNerAmp
FCRL3	0.417184	PNM	VibLower
ERCC4	0.398627	En	UlnSenNerLat, UlnMotNerCev
ABCC1	0.392721	PNM	PerMotNerAmp, PerMotNerVel
PRX	0.391839	PNM	MedSenNerAmp,MedSenNerVel, UlnSenNerVel

BASP1	0.388817	Rd	Fwave-AH, MedMotNerLat
INTU	0.387936	PNS	UlnSenNerVel
PRKG1	0.382342	En	UlnMotNerCev
GSTM1	0.377873	En	PerMotNerAmp, UlnMotNerCev
CAPG	0.377117	PNM,PNS	PerMotNerAmp, MedSenNerVel
HK2	0.3731	PNM,PNS	UlnMotNerVel, UlnSenNerLat
ITGA1	0.359346	En	MedSenNerAmp
CEP72	0.358846	Rd	UlnSenNerLat
KCNA5	0.35206	En	UlnMotNerVel
EPHA8	0.351318	PNS	TibMotNerVel
FARS2	0.346668	PNM	UlnSenNerLat
DPYD	0.340613	PNM	PerMotNerAmp, PerMotNerVel
TMEM107	0.338143	PNM,Rd	UlnSenNerVel
CYP2C9	0.333621	PNS	SurSenNerVel, MedMotNerLat
BTRC	0.32599	En	MedMotNerLat
ACTR8	0.323855	PNS,En	Fwave-AH, SurSenNerVel
NFATC4	0.313347	PNM,PNS	PerMotNerVel, SurSenNerVel
EGR1	0.310495	PNM,En	Fwave-AH
CBS	0.288383	Rd,PNS	PerMotNerAmp, UlnSenNerVel

Table 5: Oxaliplatin and Paclitaxel associated genes reported by different studies.

Oxaliplatin associated genes	Paclitaxel associated genes
ERCC1	EPHA4
DPYD	EPHA8
GSTP1	XKR4
FOXC1	RRM1
POU2AF1	ABCB1
PELO	CYP1B1
FARS2	TUBB2A
ACYP2	RFX2
SCN4A	FZD3
	GSK3B

Table 6: 45 Genes detected by CoLaSp MLC, which are highly-correlated with the output label (4 types of CIPN)

<b>1. UCHL5</b>
(A novel small molecule inhibitor of deubiquitylating enzyme USP14 and UCHL5 induces apoptosis in multiple myeloma and overcomes bortezomib resistance) Together, these data demonstrate that b-AP15 targets the deubiquitylating function of USP14 and UCHL5 in the cellular environment. <a href="#">Peripheral neuropathy</a> associated with bortezomib therapy is partly the result of blockade of neuronal cell survival protease HtrA2/Omi. In this study, we found no significant inhibition of HtrA2/Omi in response to b-AP15 treatment, whereas bortezomib inhibited HtrA2/Omi activity. These data highlight another distinction between bortezomib and b-AP15 and further indicate the selectivity of b-AP15.
<b>2. NR6A1</b>
Its expression pattern suggests that it may be involved in <a href="#">neurogenesis</a> and germ cell development
<b>3. MAGI2</b>
Seems to act as scaffold molecule at synaptic junctions by assembling <a href="#">neurotransmitter</a> receptors and cell adhesion proteins. May play a role in regulating activin-mediated signaling in <a href="#">neuronal cells</a> . Enhances the ability of PTEN to suppress AKT1 activation. Plays a role in nerve growth factor (NGF)-induced recruitment of RAPGEF2 to late endosomes and neurite outgrowth.
<b>4. NKX2-2</b>

The protein encoded by this gene contains a homeobox domain and may be involved in the morphogenesis of the [central nervous system](#).

#### **5. DPF1**

May have an important role in developing [neurons](#) by participating in regulation of cell survival, possibly as a neuro-specific transcription factor. Belongs to the neuron-specific chromatin remodeling complex (nBAF complex). During [neural](#) development a switch from a stem/progenitor to a post-mitotic chromatin remodeling mechanism occurs as neurons exit the cell cycle and become committed to their adult state. The transition from proliferating neural stem/progenitor cells to post-mitotic neurons requires a switch in subunit composition of the npBAF and nBAF complexes. As neural progenitors exit mitosis and differentiate into [neurons](#), npBAF complexes which contain ACTL6A/BAF53A and PHF10/BAF45A, are exchanged for homologous alternative ACTL6B/BAF53B and DPF1/BAF45B or DPF3/BAF45C subunits in neuron-specific complexes (nBAF). The npBAF complex is essential for the self-renewal/proliferative capacity of the multipotent neural stem cells. The nBAF complex along with CREST plays a role regulating the activity of genes essential for dendrite growth (By similarity)

#### **6. FKBP5**

The stress regulator FKBP51: a novel and promising druggable target for the treatment of persistent pain states across sexes. Deletion of FKBP51 reduced the mechanical hypersensitivity seen in joint inflammatory and [neuropathic](#) pain states in female and male mice. Furthermore, FKBP51 deletion also reduced the hypersensitivity seen in a translational model of chemotherapy-induced pain.

#### **7. RAPGEF2**

Involved in [neuron](#) migration and in the formation of the major forebrain fiber connections forming the corpus callosum, the anterior commissure and the hippocampal commissure during brain development. Involved in neuronal growth factor (NGF)-induced sustained activation of Rap1 at late endosomes and in brain-derived neurotrophic factor (BDNF)-induced axon outgrowth of hippocampal [neurons](#).

#### **8. PMCH**

MCH may act as a [neurotransmitter](#) or [neuromodulator](#) in a broad array of [neuronal](#) functions directed toward the regulation of goal-directed behavior, such as food intake, and general arousal. May also have a role in spermatocyte differentiation.

#### **9. ABHD12**

This gene encodes an enzyme that catalyzes the hydrolysis of 2-arachidonoyl glycerol (2-AG), the main endocannabinoid lipid transmitter that acts on cannabinoid receptors, CB1 and CB2. The endocannabinoid system is involved in a wide range of physiological processes, including [neurotransmission](#), mood, appetite, pain appreciation, addiction behavior, and inflammation. Mutations in this gene are associated with the neurodegenerative disease, PHARC (polyneuropathy, hearing loss, ataxia, retinitis pigmentosa, and cataract), resulting from an in-born error of endocannabinoid metabolism. Alternatively, spliced transcript variants encoding different isoforms have been noted for this gene.

#### **10. PCDHA2**

Potential calcium-dependent cell-adhesion protein. May be involved in the establishment and maintenance of specific [neuronal](#) connections in the brain.

#### **11. REEP1**

(Receptor Accessory Protein 1) is a Protein Coding gene. Diseases associated with REEP1 include Spastic Paraplegia 31, Autosomal Dominant and [Neuronopathy](#), Distal Hereditary Motor, Type Vb.

#### **12. NENF**

Acts as a [neurotrophic](#) factor in postnatal mature [neurons](#) enhancing neuronal survival. Promotes cell proliferation and [neurogenesis](#) in undifferentiated neural pro-genitor cells at the embryonic stage and inhibits differentiation of astrocyte. Its [neurotrophic](#) activity is exerted via MAPK1/ERK2, MAPK3/ERK1 and AKT1/AKT pathways. [Neurotrophic](#) activity is enhanced by binding to heme. Acts also as an anorexigenic [neurotrophic](#) factor that contributes to energy balance (By similarity). Plays a role in the human tumorigenesis

#### **13. STYXL1**

May play a role in the formation of neurites during [neuronal](#) development.

#### **14. TNFRSF1B**

Knockout studies in mice also suggest a role of this protein in protecting [neurons](#) from apoptosis by stimulating antioxidative pathways.

#### **15. NDNF**

Promotes [neuron](#) migration, growth and survival as well as neurite outgrowth.

#### **16. PCDHAC1**

Potential calcium-dependent cell-adhesion protein. May be involved in the establishment and maintenance of specific [neuronal](#) connections in the brain.

#### **17. CCR4**

The protein encoded by this gene belongs to the G-protein-coupled receptor family . It is a receptor for the CC chemokine - MIP-1, RANTES, TARC and MCP-1. Chemokines are a group of small polypeptide, structurally related molecules that regulate cell trafficking of various types of leukocytes. The chemokines also play fundamental roles in the development, homeostasis, and function of the immune system, and they have effects on cells of the [central nervous system](#) as well as on endothelial cells involved in angiogenesis or angiostasis GLIPR1 (GLI Pathogenesis Related 1) is a Protein Coding gene. Diseases associated with GLIPR1 include Glioma and [Nervous](#) System Cancer.

#### **18. NPAS4**

Transcription factor expressed in [neurons](#) of the brain that regulates the excitatory-inhibitory balance within neural circuits and is required for contextual memory in the hippocampus (By similarity). Plays a key role in the structural and functional plasticity of neurons (By similarity). Acts as an early-response transcription factor in both excitatory and inhibitory neurons, where it induces distinct but overlapping sets of late-response genes in these two types of neurons, allowing the synapses that form on inhibitory and excitatory [neurons](#) to be modified by neuronal activity in a manner specific to their function within a circuit, thereby facilitating appropriate circuit responses to sensory experience (By similarity). In excitatory neurons, activates transcription of BDNF, which in turn controls the number of GABA-releasing synapses that form on excitatory neurons, thereby promoting an increased number of inhibitory synapses on excitatory neurons (By similarity). In inhibitory neurons, regulates a distinct set of target genes that serve to increase excitatory input onto somatostatin [neurons](#), probably resulting in enhanced feedback inhibition within cortical circuits (By similarity). The excitatory and inhibitory balance in neurons affects a number of processes, such as short-term and long-term memory, acquisition of experience, fear memory, response to stress and social behavior (By similarity). Acts as a regulator of dendritic spine development in olfactory bulb granule cells in a sensory-experience-dependent manner by regulating expression of MDM2 (By similarity). Efficient DNA binding requires dimerization with another bHLH protein, such as ARNT, ARNT2 or BMAL1.

#### **19. GRIA1**

Glutamate receptors are the predominant excitatory neurotransmitter receptors in the mammalian brain and are activated in a variety of normal [neurophysiologic](#) processes. These receptors are heteromeric protein complexes with multiple subunits, each possessing transmembrane regions, and all arranged to form a ligand-gated ion channel. The classification of glutamate receptors is based on their activation by different pharmacologic agonists. This gene belongs to a family of alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA) receptors. Alternatively, spliced transcript variants encoding different isoforms have been found for this gene.

#### **20. PCDHGA12**

Potential calcium-dependent cell-adhesion protein. May be involved in the establishment and maintenance of specific [neuronal](#) connections in the brain.

#### **21. PRDM10**

The protein encoded by this gene is a transcription factor that contains C2H2-type zinc-fingers. It also contains a positive regulatory domain, which has been found in several other zinc-finger transcription factors including those involved in B cell differentiation and tumor suppression. Studies of the mouse counterpart suggest that this protein may be involved in the development of the [central nerve system](#) (CNS), as well as in the pathogenesis of neuronal storage disease. Multiple alternatively spliced transcript variants encoding distinct isoforms have been observed.

#### **22. IL3**

The protein encoded by this gene is a potent growth promoting cytokine. This cytokine is capable of supporting the proliferation of a broad range of hematopoietic cell types. It is involved in a variety of cell activities such as cell growth, differentiation and apoptosis. This cytokine has been shown to also possess neurotrophic activity, and it may be associated with [neurologic](#) disorders.

#### **23. EWSR1**

Mutations in this gene are known to cause Ewing sarcoma as well as [neuroectodermal](#) and various other tumors.

#### **24. KCNJ2**

The protein encoded by this gene is an integral membrane protein and inward-rectifier type potassium channel. The encoded protein, which has a greater tendency to allow potassium to flow into a cell rather than out of a cell, probably participates in establishing action potential waveform and excitability of [neuronal](#) and muscle tissues.

#### **25. ACTL7B**

(Actin Like 7B) is a Protein Coding gene. Diseases associated with ACTL7B include Dysautonomia and Hypertensive [Nephropathy](#). Gene Ontology (GO) annotations related to this gene include structural constituent of cytoskeleton.

#### **26. LARGE1**

The protein encoded by this gene is the glycosidase that adds the final xylose and glucuronic acid to alpha-dystroglycan and thereby allows alpha-dystroglycan to bind ligands including laminin 211 and neurexin. Mutations in this gene cause several forms of congenital muscular dystrophy characterized by cognitive disability and abnormal glycosylation of alpha-dystroglycan. Alternative splicing of this gene results in multiple transcript variants that encode the same protein.

### **27. NDE1**

Required for centrosome duplication and formation and function of the mitotic spindle. Essential for the development of the cerebral cortex. May regulate the production of [neurons](#) by controlling the orientation of the mitotic spindle during division of cortical neuronal progenitors of the proliferative ventricular zone of the brain. Orientation of the division plane perpendicular to the layers of the cortex gives rise to two proliferative neuronal progenitors whereas parallel orientation of the division plane yields one proliferative neuronal progenitor and a post-mitotic neuron. A premature shift towards a [neuronal](#) fate within the progenitor population may result in an overall reduction in the final number of neurons and an increase in the number of neurons in the deeper layers of the cortex.

### **28. VWC2L**

May play a role in [neurogenesis](#). May play a role in bone differentiation and matrix mineralization.

### **29. AIFM1**

Mutations in this gene cause combined oxidative phosphorylation deficiency 6 (COXPD6), a severe mitochondrial encephalomyopathy, as well as Cowchock syndrome, also known as X-linked recessive Charcot-Marie-Tooth disease-4 (CMTX-4), a disorder resulting in [neuropathy](#), and axonal and motor-sensory defects with deafness and cognitive disability.

### **30. ROBO1**

Bilateral symmetric [nervous](#) systems have special midline structures that establish a partition between the two mirror image halves. Some axons project toward and across the midline in response to long-range chemoattractants emanating from the midline. The product of this gene is a member of the immunoglobulin gene superfamily and encodes an integral membrane protein that functions in axon guidance and [neuronal](#) precursor cell migration. This receptor is activated by SLIT-family proteins, resulting in a repulsive effect on glioma cell guidance in the developing brain.

### **31. RBFOX3**

This gene produces the neuronal nuclei (NeuN) antigen that has been widely used as a marker for post-mitotic neurons. This gene has its highest expression in the central nervous system and plays a prominent role in neural tissue development and regulation of adult brain function. Mutations in this gene have been associated with numerous [neurological](#) disorders.

### **32. NAA50**

Diseases associated with NAA50 include [Neuropathy](#), Hereditary Sensory And Autonomic, Type Ia and [Neuropathy](#), Hereditary Sensory And Autonomic, Type Ib.

### **33. OR1E1**

Olfactory receptors interact with odorant molecules in the nose, to initiate a [neuronal](#) response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome. The nomenclature assigned to the olfactory receptor genes and proteins for this organism is independent of other organisms.

### **34. GDAP1**

This gene encodes a member of the ganglioside-induced differentiation-associated protein family, which may play a role in a signal transduction pathway during neuronal development. Mutations in this gene have been associated with various forms of Charcot-Marie-Tooth Disease and [neuropathy](#).

### **35. LHX8**

The protein encoded by this gene is a member of the LIM homeobox family of proteins, which are involved in patterning and differentiation of various tissue types. These proteins contain two tandemly repeated cysteine-rich double-zinc finger motifs known as LIM domains, in addition to a DNA-binding homeodomain. This family member is a transcription factor that plays a role in tooth morphogenesis. It is also involved in oogenesis and in [neuronal](#) differentiation. This gene is a candidate gene for cleft palate, and it is also associated with odontoma formation. Alternative splicing of this gene results in multiple transcript variants.

### **36. OR9A4**

Olfactory receptors interact with odorant molecules in the nose, to initiate a [neuronal](#) response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome. The nomenclature assigned to the olfactory receptor genes and proteins for this organism is independent of other organisms.

#### **37. PCDHGA1**

Potential calcium-dependent cell-adhesion protein. May be involved in the establishment and maintenance of specific [neuronal](#) connections in the brain.

#### **38. KIF21B**

Plus-end directed microtubule-dependent motor protein which displays processive activity. Is involved in regulation of microtubule dynamics, synapse function and [neuronal](#) morphology, including dendritic tree branching and spine formation. Plays a role in learning and memory. Involved in delivery of gamma-aminobutyric acid (GABA(A)) receptor to cell surface.

#### **39. PCSK1N**

The protein encoded by this gene functions as an inhibitor of prohormone convertase 1, which regulates the proteolytic cleavage of [neuroendocrine](#) peptide precursors. The proprotein is further processed into multiple short peptides. A polymorphism within this gene may be associated with obesity.

#### **40. PRR12**

This gene encodes a proline-rich protein that contains two A-T hook DNA binding domains. A chromosomal translocation and gene fusion between this gene and zinc finger, MIZ-type containing 1 (Gene ID: 57178) may underlie intellectual disability and neuropsychiatric problems in a human patient. Enriched expression of this gene in embryonic mouse brain suggests that this gene may play a role in [nervous system](#) development.

#### **41. ZFYVE27**

Key regulator of RAB11-dependent vesicular trafficking during neurite extension through polarized membrane transport. Promotes axonal elongation and contributes to the establishment of neuronal cell polarity (By similarity). Involved in nerve growth factor-induced neurite formation in VAPA-dependent manner. Contributes to both the formation and stabilization of the tubular ER network. Involved in ER morphogenesis by regulating the sheet-to-tubule balance and possibly the density of tubule interconnections. Acts as an adapter protein and facilitates the interaction of KIF5A with VAPA, VAPB, SURF4, RAB11A, RAB11B and RTN3 and the ZFYVE27-KIF5A complex contributes to the transport of these proteins in [neurons](#). Can induce formation of neurite-like membrane protrusions in non-neuronal cells in a KIF5A/B-dependent manner.

#### **42. FKBP8**

The protein encoded by this gene is a member of the immunophilin protein family, which play a role in immunoregulation and basic cellular processes involving protein folding and trafficking. Unlike the other members of the family, this encoded protein does not seem to have PPIase/rotamase activity. It may have a role in [neurons](#) associated with memory function.

#### **43. RTN4RL2**

Cell surface receptor that plays a functionally redundant role in the inhibition of neurite outgrowth mediated by MAG (By similarity). Plays a functionally redundant role in postnatal brain development. Contributes to normal axon migration across the brain midline and normal formation of the corpus callosum. Does not seem to play a significant role in regulating axon regeneration in the adult [central nervous system](#). Protects motoneurons against apoptosis; protection against apoptosis is probably mediated by MAG (By similarity). Like other family members, plays a role in restricting the number dendritic spines and the number of synapses that are formed during brain development (PubMed:22325200). Signaling mediates activation of Rho and downstream reorganization of the actin cytoskeleton.

#### **44. PRICKLE2**

Diseases associated with PRICKLE2 include Progressive Myoclonic Epilepsy Type 5 and Sensory Ataxic [Neuropathy](#), Dysarthria, And Ophthalmoparesis. Among its related pathways are Wnt Signaling Pathway and Pluripotency. An important paralog of this gene is PRICKLE3.

#### **45. VIP**

Vasoactive intestinal peptide (VIP) receptors are a group of Gs -protein-coupled receptors that are currently divided into two subtypes; VPAC1 and VPAC2. They are widely distributed throughout the [central nervous system](#).

## References

- Chang, P. M.-H., Tzeng, C.-H., Chen, P.-M., Lin, J.-K., Lin, T.-C., Chen, W.-S., Jiang, J.-K., Wang, H.-S., and Wang, W.-S. (2009). ERCC1 codon 118 C- $\rightarrow$ T polymorphism associated with ERCC1 expression and outcome of FOLFOX-4 treatment in Asian patients with metastatic colorectal carcinoma. *Cancer Science*, 100(2):278–283.
- Yamaguchi, H., Hishinuma, T., Endo, N., Tsukamoto, H., Kishikawa, Y., Sato, M., Murai, Y., Hiratsuka, M., Ito, K., Okamura, C., Yaegashi, N., Suzuki, N., Tomioka, Y., and Goto, J. (2006). Genetic variation in ABCB1 influences paclitaxel pharmacokinetics in Japanese patients with ovarian cancer. *International Journal of Gynecologic Cancer*, 16(3):979–985.
- Reyes-Gibby, C., Wang, J., Yeung, S., and Shete, S. (2015). Informative gene network for chemotherapy-induced peripheral neuropathy. *BioData Mining*, 8:24.
- Argyriou, A. A., Bruna, J., Genazzani, A. A., and Cavaletti, G. (2017). Chemotherapy-induced peripheral neurotoxicity: management informed by pharmacogenetics. *Nature Reviews Neurology*, 13(8):492–504.
- Spyromitros, E., Tsoumakas, G., and Vlahavas, I. P. (2008). An empirical study of lazy multilabel classification algorithms. In *Proc. 5th Hellenic Conf. Artif. Intell., Syros, Greece*, pages 401–406.
- Luaces, O., Díez, J., Barranquero, J., del Coz, J. J., and Bahamonde, A. (2012). Binary relevance efficacy for multilabel classification. *Progress in AI*, 1(4):303–313.
- Boutell, M. R., Luo, J., Shen, X., and Brown, C. M. (2004). Learning multi-label scene classification. *Pattern Recognit.*, 37(9):1757–1771.
- Zhang, M. and Zhou, Z. (2007). ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognit.*, 40(7):2038–2048.
- Chang, C. and Lin, C. (2011). LIBSVM: A library for support vector machines. *ACM TIST*, 2(3):27:1–27:27.