# Bioinformatics analysis of correlation between protein function and intrinsic disorder

Goran Vinterhalter\*1 Jovana J Kovačević<sup>1</sup> Gordana Pavlović Lažetić<sup>1</sup>

University of Belgrade, Faculty of Mathematics<sup>1</sup>

June 22, 2018





- Inspiration
- Method and data
  - Data
  - Dependency between protein length and predicted disorder
  - Functions dependency on (dis)order
- Oata preprocessing
  - Combining dataset and annotations
  - Mapping MF keywords to MF terms
- Results
  - Tabular comparison
  - GO subgraph
  - Level of disorder and p-value



- Inspiration
- Method and data
  - Data
  - Dependency between protein length and predicted disorder
  - Functions dependency on (dis)order
- 3 Data preprocessing
  - Combining dataset and annotations
  - Mapping MF keywords to MF terms
- 4 Results
  - Tabular comparison
  - GO subgraph
  - Level of disorder and p-value



Reference work: Xie. et. al 2006

Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions Hongbo Xie, Slobodan Vucetic, Lilia M. Iakoucheva, Christopher J. Oldfield, A. Keith Dunker, Vladimir N. Uversky, and Zoran Obradovic



Reference work: Xie. et. al 2006

Functional anthology of intrinsic disorder. 1.
Biological processes and functions of proteins with long disordered regions
Hongbo Xie, Slobodan Vucetic, Lilia M. Iakoucheva, Christopher J. Oldfield, A. Keith
Dunker, Vladimir N. Uversky, and Zoran Obradovic

• Swiss-Prot v48 from 2005. ( about 200 000 seq.)



#### Reference work: Xie. et. al 2006

Functional anthology of intrinsic disorder. 1.
Biological processes and functions of proteins with long disordered regions
Hongbo Xie, Slobodan Vucetic, Lilia M. Iakoucheva, Christopher J. Oldfield, A. Keith
Dunker, Vladimir N. Uversky, and Zoran Obradovic

- Swiss-Prot v48 from 2005. (about 200 000 seq.)
- UniprotKB Keywords

#### Reference work: Xie. et. al 2006

Functional anthology of intrinsic disorder. 1.
Biological processes and functions of proteins with long disordered regions
Hongbo Xie, Slobodan Vucetic, Lilia M. Iakoucheva, Christopher J. Oldfield, A. Keith
Dunker, Vladimir N. Uversky, and Zoran Obradovic

- Swiss-Prot v48 from 2005. ( about 200 000 seq.)
- UniprotKB Keywords
- PONDR VSL3 disorder predictor



- Inspiration
- Method and data
  - Data
  - Dependency between protein length and predicted disorder
  - Functions dependency on (dis)order
- Oata preprocessing
  - Combining dataset and annotations
  - Mapping MF keywords to MF terms
- 4 Results
  - Tabular comparison
  - GO subgraph
  - Level of disorder and p-value

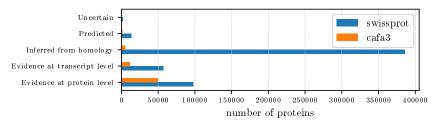


• CAFA3 dataset (training set from CAFA3 challenge (2016))

- CAFA3 dataset (training set from CAFA3 challenge (2016))
- GO terms

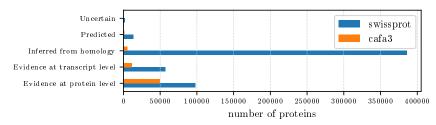
- CAFA3 dataset (training set from CAFA3 challenge (2016))
- GO terms
- CAFA3 dataset is a subset of Swiss-Prot

- CAFA3 dataset (training set from CAFA3 challenge (2016))
- GO terms
- CAFA3 dataset is a subset of Swiss-Prot



Protein existance

- CAFA3 dataset (training set from CAFA3 challenge (2016))
- GO terms
- CAFA3 dataset is a subset of Swiss-Prot



Protein existance

• We assume the CAFA3 dataset is not statistically redundant



# Predicting disorder

PONDR VSL2b disorder predictor



## Predicting disorder

PONDR VSL2b disorder predictor

#### Definition

Protein is labeled as **putatively disordered** if it contains a predicted IDPr at least 40 AA long, otherwise putatively ordered.



## Predicting disorder

PONDR VSL2b disorder predictor

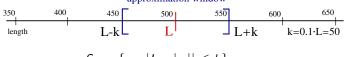
#### **Definition**

Protein is labeled as **putatively disordered** if it contains a predicted IDPr at least 40 AA long, otherwise putatively ordered.

# But!

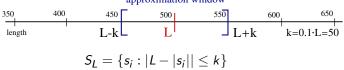
Likelihood of labeling a protein putatively disordered increses by it's length.

•  $S_L$ , set of proteins with lengths from interval [L-k, L+k], k=L/10approximation window



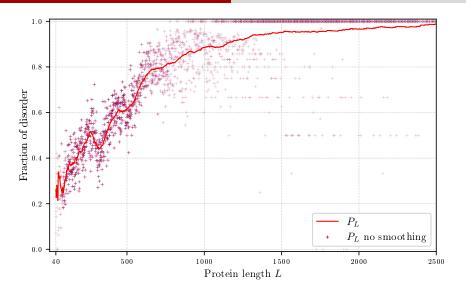
$$S_L = \{s_i : |L - |s_i|| \le k\}$$

•  $S_L$ , set of proteins with lengths from interval [L-k,L+k], k=L/10 approximation window



The probability a protein of length L is labeled putatively disordered:

$$P_L = rac{\sum_{s_i \in S_L} d(s_i)}{|S_L|}$$
  $d(s_i) = egin{cases} 1 & ext{if } s_i ext{ is putatively disordered} \ 0 & ext{otherwise} \end{cases}$ 



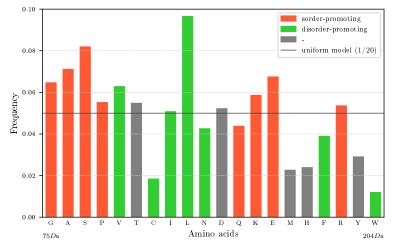
Dependency between protein length and fraction of putative disorder



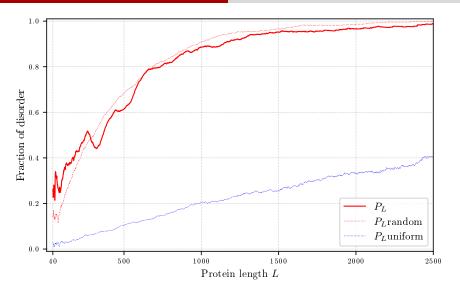
# Generating protein sequences

• random model ( *P<sub>L</sub>random* )

• uniform model ( *P<sub>L</sub>uniform* )



Amino acid fractions in CAFA3 dataset



Comparing P<sub>L</sub>, P<sub>L</sub>random and P<sub>L</sub>uniform on CAFA3 dataset



# Correlation between function and (dis)order

Let S<sub>j</sub> be a set of proteins annotated by function j.
 By disorder level (of function j) we assume the fractions (F<sub>j</sub>) of putatively disordered proteins inside set S<sub>j</sub>.

$$F_j = \frac{\sum_{s_i \in S_j} d(s_i)}{|S_j|}$$

# Correlation between function and (dis)order

Let S<sub>j</sub> be a set of proteins annotated by function j.
 By disorder level (of function j) we assume the fractions (F<sub>j</sub>) of putatively disordered proteins inside set S<sub>j</sub>.

$$F_j = \frac{\sum_{s_i \in S_j} d(s_i)}{|S_j|}$$

Null model

$$Y_j = rac{\sum_{s_i \in S_j} X_{|s_j|}}{|S_j|} \qquad \quad X_L : egin{pmatrix} 0 & 1 \ 1 - P_L & P_L \end{pmatrix}$$

# Correlation between function and (dis)order

Let S<sub>j</sub> be a set of proteins annotated by function j.
 By disorder level (of function j) we assume the fractions (F<sub>j</sub>) of putatively disordered proteins inside set S<sub>j</sub>.

$$F_j = \frac{\sum_{s_i \in S_j} d(s_i)}{|S_j|}$$

Null model

$$Y_j = rac{\sum_{s_i \in S_j} X_{|s_j|}}{|S_i|} \qquad \quad X_L : \begin{pmatrix} 0 & 1 \\ 1 - P_L & P_L \end{pmatrix}$$

- p > 0.95: function is **disorder related** (correlated with ...)
- p < 0.05: function is **order related**
- otherwise: nothing can be said about correlation (insufficient stat. sig.)

- Inspiration
- Method and data
  - Data
  - Dependency between protein length and predicted disorder
  - Functions dependency on (dis)order
- Oata preprocessing
  - Combining dataset and annotations
  - Mapping MF keywords to MF terms
- 4 Results
  - Tabular comparison
  - GO subgraph
  - Level of disorder and p-value



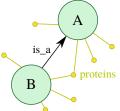
Merging Swiss-Prot and it's CAFA3 subset



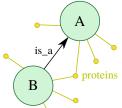
- Merging Swiss-Prot and it's CAFA3 subset
  - 66 530 valid proteins (out of 66 599 total)
  - Annotations (term & kw) we get from Swiss-Prot
  - Sequences we keep from CAFA3 dataset



- Merging Swiss-Prot and it's CAFA3 subset
  - 66 530 valid proteins (out of 66 599 total)
  - Annotations (term & kw) we get from Swiss-Prot
  - Sequences we keep from CAFA3 dataset
- Annotation grouping for MF terms

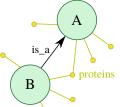


- Merging Swiss-Prot and it's CAFA3 subset
  - 66 530 valid proteins (out of 66 599 total)
  - Annotations (term & kw) we get from Swiss-Prot
  - Sequences we keep from CAFA3 dataset
- 2 Annotation grouping for MF terms



1781 MF terms that annotate min. 20 proteins. Without this step we wold only get 1146 MF terms.

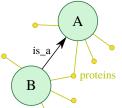
- Merging Swiss-Prot and it's CAFA3 subset
  - 66 530 valid proteins (out of 66 599 total)
  - Annotations (term & kw) we get from Swiss-Prot
  - Sequences we keep from CAFA3 dataset
- 2 Annotation grouping for MF terms



 $1781~\mathrm{MF}$  terms that annotate min. 20 proteins. Without this step we wold only get 1146 MF terms.

Mapping MF keywords to MF terms ???

- Merging Swiss-Prot and it's CAFA3 subset
  - 66 530 valid proteins (out of 66 599 total)
  - Annotations (term & kw) we get from Swiss-Prot
  - Sequences we keep from CAFA3 dataset
- 2 Annotation grouping for MF terms

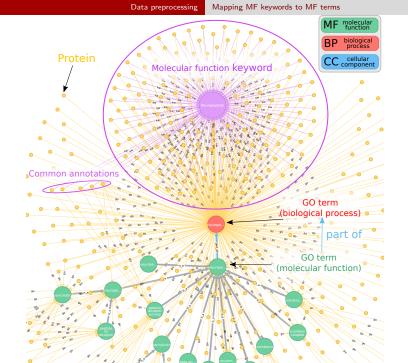


 $1781~\mathrm{MF}$  terms that annotate min. 20 proteins. Without this step we wold only get 1146 MF terms.

Mapping MF keywords to MF terms ???

	total	no map.	MF map.	BP map.	CC. map.
MF keywords	195	20	104	54	11





Jaccard index (Ji), 
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \in [0,1]$$



Jaccard index (Ji), 
$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \in [0, 1]$$



• Consider the whole Swiss-Prot (but really CAFA3 dataset is enough)

Jaccard index (Ji), 
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \in [0,1]$$



- Consider the whole Swiss-Prot (but really CAFA3 dataset is enough)
- No MF annotations grouping (smaller Ji, or wrong)

Jaccard index (Ji), 
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \in [0,1]$$



- Consider the whole Swiss-Prot (but really CAFA3 dataset is enough)
- No MF annotations grouping (smaller Ji, or wrong)
- expecting between 60 and 85% "correct" mapping

Jaccard index (Ji), 
$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \in [0, 1]$$



- Consider the whole Swiss-Prot (but really CAFA3 dataset is enough)
- No MF annotations grouping (smaller Ji, or wrong)
- expecting between 60 and 85% "correct" mapping
- We produce 64 mappings with Ji > 0.1

- Inspiration
- 2 Method and data
  - Data
  - Dependency between protein length and predicted disorder
  - Functions dependency on (dis)order
- 3 Data preprocessing
  - Combining dataset and annotations
  - Mapping MF keywords to MF terms
- Results
  - Tabular comparison
  - GO subgraph
  - Level of disorder and p-value



### Results

	reference	re	sults
	Xie2007 kw	MF kw	MF terms
total ( $\#prot \geq 20$ )	143	186	1781
p < 0.05 (order rel.)	37	53	699
p > 0.95 (disorder rel.)	51	44	616
insufficient stat. sig.	55	89	576



#### Results

	reference	re	sults
	Xie2007 kw	MF kw	MF terms
total (# $prot \ge 20$ )	143	186	1781
p < 0.05 (order rel.)	37	53	699
p > 0.95 (disorder rel.)	51	44	616
insufficient stat. sig.	55	89	576

#### Two types of results:

- Tabular comparison with keywords
- GO subgraph



Results	Tabular	com

									Results		abula	ar co	ompa	11130
									Order related MF ke	eywords	s (z-skor>	0.2)		
	Top 20 MF kw with strongest	correlatio	predicted	order (Xi	e et. al. 20	007)	1	#	name	n	avg_len	avg_dis	z	p
#	name	n prot	n families	avg_len	Z-score	P-value		0	Oxidoreductase	4126	472.25	0.28	-41.35	0
0	Oxidoreductase	14995	992	376.63	-29.54	0		1	Hydrolase	7564	614.81	0.51	-26.99	0
1	Transferase	26525	1606	445.17	-24.25	0	K 2	2	Lyase	1431	481.37	0.30	-23.62	0
2	Lyase	7262	347	377.92	-22.64	0	$\rightarrow$	3	Monooxygenase	555	503.36	0.20	-20.29	0
3	Hydrolase	20464	1995	430.68	-21.75	0	Y X	4	Transferase	8846	631.95	0.55	-19.72	0
	Isomerase	4487	220	383.98	-14.18	0	\ / <b>.</b>	5	Ligase	995	693.30	0.46	-18.05	0
	Glycosidase	1826	244	444.73	-13.98	0	$\sim 4$	6	Glycosyltransferase	1134	551.26	0.40	-17.04	0
5	Glycosyltransferase	2950	261	437.53	-12.51	0	- X	7	Glycosidase	697	570.50	0.37	-16.81	0
	Acyltransferase	2239	179	402.83	-10.85	0	k / 🌂	8	Isomerase	931	422.72	0.35	-13.60	0
	Methyltransferase	3524	224	349.60	-10.53	0		9	Protease	1863	674.42	0.54	-13.20	0
,	Kinase	7017	322	448.29	-10.22	0	1V/\7	10	Transducer	1703	482.28	0.41	-12.56	0
0	Ligase	8010	230	529.41	-10.06	0	YX XI	11	G-protein coupled receptor	1385	465.62	0.39	-12.45	0
1	Decarboxylase	1293	63	345.26	-9.66	0	K/ /¾	12	Acyltransferase	867	531.58	0.42	-11.28	0
2	Monooxygenase	1668	73	444.87	-9.26	0	V 🗸 🕶	13	Decarboxylase	195	488.21	0.25	-10.70	0
3	Metalloprotease	1100	109	553.73	-7.89	0	ιX	14	Aminotransferase	202	451.05	0.24	-10.23	0
4	Aminopeptidase	452	39	509.17	-7.55	0	$1 \times 1/4$	15	Aminopeptidase	130	668.72	0.37	-9.10	0
5	Dioxygenase	360	66	433.20	-7.32	0	1/X/	16	Serine protease	460	700.07	0.50	-8.87	0
6	Aminoacyl-tRNA synthetase	3402	37	571.83	-7.15	0	X ∕ 🍾	17	Metalloprotease	507	688.25	0.56	-8.43	0
7	Protease	4423	380	549.70	-7.1	0	YX ¥	18	Methyltransferase	874	611.22	0.47	-8.33	0
3	Aminotransferase	955	28	420.27	-6.02	0	V \	19	Carboxypeptidase	116	631.16	0.37	-8.25	0
							, /	20	Threonine protease	138	246.88	0.18	-7.50	0
							*	21	Dioxygenase	366	622.32	0.48	-7.39	0
								22	Serine esterase	141	423.09	0.28	-7.21	0
								22	Serine esterase Recentor	141 3424	423.09 647.92	0.28	-7.21 -7.09	0
								22 23 24	Receptor	141 3424 600	423.09 647.92 969.68	0.28 0.59 0.63	-7.21 -7.09 -6.18	0
								23		3424	647.92	0.59	-7.09	0
								23 24 25	Receptor Nucleotidyltransferase Peroxidase	3424 600 221	647.92 969.68 457.61	0.59 0.63 0.40	-7.09 -6.18 -5.50	0
								23 24 25 26	Receptor Nucleotidyltransferase Peroxidase Serine protease inhibitor	3424 600 221 182	647.92 969.68 457.61 497.66	0.59 0.63 0.40 0.38	-7.09 -6.18 -5.50 -4.92	0 0 0
								23 24 25 26 27	Receptor Nucleotidyltransferase Peroxidase Serine protease inhibitor Porin	3424 600 221 182 63	647.92 969.68 457.61 497.66 318.84	0.59 0.63 0.40 0.38 0.21	-7.09 -6.18 -5.50 -4.92 -4.55	0 0 0
								23 24 25 26 27 28	Receptor Nucleotidyltransferase Peroxidase Serine protease inhibitor Porin Dipeptidase	3424 600 221 182 63 22	647.92 969.68 457.61 497.66 318.84 522.27	0.59 0.63 0.40 0.38 0.21 0.23	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43	0 0 0 0
								23 24 25 26 27 28 29	Receptor Nucleotidyltransferase Peroxidase Serine protease inhibitor Porin Dipeptidase Integrin	3424 600 221 182 63 22 71	647.92 969.68 457.61 497.66 318.84 522.27 1038.25	0.59 0.63 0.40 0.38 0.21 0.23 0.70	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43 -4.29	0 0 0 0 0
								23 24 25 26 27 28 29 30	Receptor Nucleotidyltransferase Peroxidase Serine protease inhibitor Porin Dipeptidase Integrin Protease inhibitor	3424 600 221 182 63 22 71 284	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43 -4.29 -4.15	0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31	Receptor Nucleotidyltransferase Peroxidase Serine protease inhibitor Porin Dipeptidase Integrin Protease inhibitor RNA-directed RNA polymerase	3424 600 221 182 63 22 71 284	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43 -4.29 -4.15 -4.11	0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31	Receptor Nucleotidyltramsferase Peroxidase Serine protease inhibitor Porin Dipeptidase Integrin Protease inhibitor RNA-directed RNA polymerase Neurotoxin	3424 600 221 182 63 22 71 284 62 66	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43 -4.29 -4.15 -4.11 -3.80	0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33	Receptor Nucleotidyltramsferase Pervaidase Serine protease inhibitor Perin Dipeptidase Integrin Protease inhibitor RNA-directed RNA polymerase Neurotoxin Hemagglutinin	3424 600 221 182 63 22 71 284 62 66 23	969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17 0.13	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43 -4.29 -4.15 -4.11 -3.80 -3.37	0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34	Receptor Nucleotidyltramsferase Peroxidase Serine protease inhibitor Porin Dipeptidase Integrin Protease inhibitor RNA-directed RNA polymerase Neurotoxin Hemagglatinin Retinal protein	3424 600 221 182 63 22 71 284 62 66 23 47	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17 0.13 0.28	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43 -4.29 -4.15 -4.11 -3.80 -3.37 -3.37	0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35	Receptor Nucleotidyltransferase Persoidase Serine protease inhibitor Perin Dipeptidase Integrin Protease inhibitor RNA-directed RNA polymerase Neurotoxin Henagglatinin Retinal protein Myosin	3424 600 221 182 63 22 71 284 62 66 23 47	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06 1275.20	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17 0.13 0.28 0.72	-7.09 -6.18 -5.50 -4.92 -4.43 -4.29 -4.15 -4.11 -3.80 -3.37 -3.37 -3.29	0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35	Receptor Nucleotidyltransferase Peroxidase Serine protesse inhibitor Perin Dipeptidase Integrin Protesse inhibitor Protesse inhibitor Protesse inhibitor Protesse inhibitor Protesse inhibitor RNA-directed RNA-opyterases Nucurotosin Hemagglatinin Retinal protein Myosin Myosin	3424 600 221 182 63 22 71 284 62 66 23 47 185	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06 1275.20 540.68	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17 0.13 0.28 0.72 0.47	-7.09 -6.18 -5.50 -4.92 -4.43 -4.29 -4.15 -4.11 -3.80 -3.37 -3.37 -3.29 -3.02	0 0 0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35 36	Receptor Nucleotish Primiteriare Peroxidase Peroxidase Perioria Primiteriare Perioria Dipeptidase Integritoria Protectoria Integritoria Protectoria Integritoria Protectoria Integritoria Protectoria Integritoria Primiteria Protectoria Integritoria Primiteria Protectoria Primiteria Protectoria Integritoria Protectoria Integritoria Protectoria	3424 600 221 182 63 22 71 284 62 66 23 47 185 77	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06 1275.20 540.68 376.21	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17 0.13 0.28 0.72 0.47 0.31	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43 -4.29 -4.15 -4.11 -3.80 -3.37 -3.37 -3.29 -3.02 -2.59	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35 36 37	Receptor Nucleotidyltransferase Peroxidase Serine postease inhibitor Porin Dipoptidase Integrin Dipoptidase Integrin Protease inhibitor RNA-directed RNA opymerase Henagelutin Retatal postein Myosin Postoreceptor protein Prevyltransferase Elonggino factor	3424 600 221 182 63 22 71 284 62 66 23 47 185 77 39	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06 1275.20 540.68 376.21 467.68	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17 0.13 0.28 0.72 0.47 0.31	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43 -4.29 -4.15 -4.11 -3.80 -3.37 -3.37 -3.29 -3.02 -2.59 -2.59	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38	Receptor Niclostub/transferase Persvidase Persvidase Persvidase Series protoses inhibitor Deperdidase Integrin Integrin Integrin Integrin Persverses inhibitor Persverses inhibitor Persverses inhibitor Neurotextin Persverses inhibitor Persverses inhibitor Neurotextin Retusal protein Mysoni Persverses inhibitor Mysoni Persverses inhibitor Persverses inhibitor Persverses inhibitor Persverses inhibitor Endomorficase Engalion factor Endomorficase	3424 600 221 182 63 22 71 284 62 66 23 47 185 77 39 106	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06 1275.20 540.68 778.14	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17 0.13 0.28 0.72 0.47 0.31 0.48	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43 -4.29 -4.15 -4.11 -3.80 -3.37 -3.37 -3.29 -3.02 -2.59 -2.59 -2.59	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40	Receptor Nucleothyltramsferase Perroxidase Perroxidase Serine protease inhibitor Perin Depoptidase Integrin Protease inhibitor RNA-directed RNA polymerate Hemagglutinin Hemagglutinin Hemagglutinin Myotin Myotin Depolymeratese Depolymeratese Depolymeratese Enogation factor Endomaclease Toxin	3424 600 221 182 63 22 71 284 62 66 23 47 185 77 39 106 443 214	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06 1275.20 540.68 376.21 467.68 728.14	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17 0.13 0.28 0.72 0.47 0.31 0.48 0.60 0.39	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43 -4.29 -4.15 -4.11 -3.80 -3.37 -3.29 -3.02 -2.59 -2.59 -2.59 -2.56	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40	Receptor Nucleotidyltransferase Peroxidase Peroxidase Serine protease inhibitor Portin Diopoptidase Integrin Diopoptidase Integrin Protease inhibitor RNA directed RNA polymerase Neutroted RNA polymerase Neutroted RNA polymerase Neutroted RNA polymerase Postorocceptor protein Proceedings of the property protein Disconnection Endomoclose Toxin Bacteriolybic enzyme	3424 600 221 182 63 22 71 284 62 66 23 47 185 77 39 106 443 214	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 540.68 376.21 467.68 728.14 412.51	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17 0.13 0.28 0.72 0.47 0.31 0.48 0.60 0.39	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43 -4.29 -4.15 -4.11 -3.80 -3.37 -3.37 -3.29 -3.02 -2.59 -2.59 -2.59 -2.59 -2.56 -2.53	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42	Receptor Nucleothyltermaferase Pertoxidase Pertoxidase Serine protease inhibitor Porin Depeptidase Integrin Protease inhibitor RNA-directed RNA polymerase Neurotosi Hemagglutinin Hemagglutinin Hemagglutinin Protein Myotin Detoxico-protein Proteore-protein Proteore-protein Proteore-protein Protein-protein Detoxico-protein Protein-protein Protein-protein Retrial protein Baserir-dytic enzyme Baserir-dytic enzyme Nuclease	3424 600 221 182 63 22 71 284 62 66 23 47 185 77 39 106 443 214 35 703	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06 1275.20 540.68 376.21 467.68 728.14 412.51 434.66 668.54	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17 0.13 0.28 0.72 0.47 0.31 0.48 0.60 0.39 0.31	-7.09 -6.18 -5.50 -4.92 -4.55 -4.43 -4.29 -4.15 -4.11 -3.80 -3.37 -3.37 -3.29 -3.02 -2.59	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43	Receptor Nucleotidyltransferase Peroxidase Peroxidase Serine protease inhibitor Forin Dipopridase Integrin Dipopridase Integrin Protease inhibitor RNA directal RNA polymeras Neurotead RNA polymeras Neurotead RNA polymeras Pensugalatina Rensig Justina Rensig Justina Pensugalatina Pe	3424 600 221 182 63 22 71 284 62 66 23 47 185 77 39 106 443 214 35 703 60	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06 1275.20 540.68 728.14 412.51 447.68 728.14 412.51 434.66 668.54 74.92	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17 0.13 0.28 0.72 0.47 0.31 0.48 0.60 0.39 0.70 0.10	-7.09 -6.18 -5.50 -4.92 -4.43 -4.29 -4.15 -4.11 -3.80 -3.37 -3.37 -3.32 -2.59 -2.59 -2.59 -2.59 -2.59 -2.59 -2.53 -2.49 -2.47	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43	Receptor Nucleothyltumiferase Pertoxidase Pertoxidase Pertoxidase Serine protease inhibitor Portin Dipeptidase Integrin Protease inhibitor RNA-directed RNA polymerate Hemagglutinin Hemagglutinin Hemagglutinin Hemagglutinin Protein Myotin Photoexceptor protein Presyltramsferase Elongation factor Endomuclease Toxin Bacteriolytic enzyme Nuclease In channel impairing toxin Ratenal impairing toxin	3424 600 221 182 63 22 71 284 66 23 47 185 77 39 106 443 214 35 703 60 22	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06 1275.20 540.68 376.21 467.68 728.14 412.51 434.66 668.54 74.92 79.55	0.59 0.63 0.40 0.38 0.21 0.70 0.41 0.87 0.17 0.13 0.28 0.70 0.47 0.31 0.48 0.60 0.39 0.31 0.48 0.60 0.39 0.70	-7.09 -6.18 -5.50 -4.92 -4.45 -4.43 -4.29 -4.15 -4.11 -3.80 -3.37 -3.29 -3.02 -2.59 -2.59 -2.59 -2.56 -2.53 -2.47 -2.42	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45	Receptor Nucleotis/plannsferase Peroxidase Peroxidase Serine protease inhibitor Portin Dipeptidase Integrin Dipeptidase Integrin Protease inhibitor RNA-directed RNA polymerate Neurotost RNA polymerate Neurotost RNA polymerate Penninghalmin Retinal protein Mysous Mysous Department of the protein Department of the propertment of the pro	3424 600 221 182 63 22 71 284 66 23 47 185 77 39 106 443 214 214 5703 60 22 120	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06 1275.20 540.68 376.21 467.68 728.14 412.51 434.66 668.54 74.92 79.55 942.30	0.59 0.63 0.40 0.38 0.70 0.41 0.87 0.17 0.13 0.28 0.72 0.47 0.31 0.48 0.60 0.39 0.31 0.60 0.05 0.64	.7.09 -6.18 -5.50 -4.92 -4.15 -4.43 -4.29 -4.15 -3.37 -3.37 -3.37 -2.59 -2.59 -2.59 -2.54 -2.54 -2.54 -2.54 -2.54 -2.54 -2.54 -2.55 -2.54 -2.55 -2.54 -2.55	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45	Receptor Nucleothyltermaniferase Perroxidase Perroxidase Perroxidase Perroxidase Perin protesses inhibitor Integrin Dipeptidase Integrin Protesses inhibitor RNA-directal RNA polymerase Neurotorin Hemagglatinin Hemagglatinin Myosin Hemagglatinin Persyltramsferase Elongation factor Endomacleose Toxin Bacteriolytic enzyme Nuclease In channel impairing toxin Nuclease Ion channel impairing toxin Apapryl protease Hemoteasis impairing toxin	3424 600 221 182 63 22 71 284 62 66 23 47 77 39 106 443 214 35 703 60 22 120 55	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06 1275.20 540.68 376.21 467.68 728.14 412.51 434.66 668.54 74.92 79.55	0.59 0.63 0.40 0.38 0.21 0.23 0.70 0.41 0.87 0.17 0.13 0.28 0.72 0.47 0.31 0.49 0.49 0.39 0.39 0.05 0.60 0.15 0.60 0.15 0.60 0.15	-7.09 -6.18 -5.50 -4.49 -4.55 -4.43 -4.29 -4.11 -3.80 -3.37 -3.37 -3.39 -2.59 -2.59 -2.59 -2.59 -2.49 -2.47 -2.42 -2.49 -2.47 -2.42 -2.49 -2.49 -2.49 -2.49 -2.49 -2.59 -2	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
								23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45	Receptor Nucleotis/plannsferase Peroxidase Peroxidase Serine protease inhibitor Portin Dipeptidase Integrin Dipeptidase Integrin Protease inhibitor RNA-directed RNA polymerate Neurotost RNA polymerate Neurotost RNA polymerate Penninghalmin Retinal protein Mysous Mysous Department of the protein Department of the propertment of the pro	3424 600 221 182 63 22 71 284 66 23 47 185 77 39 106 443 214 214 5703 60 22 120	647.92 969.68 457.61 497.66 318.84 522.27 1038.25 446.29 2506.11 209.62 281.22 384.06 1275.20 540.68 376.21 467.68 728.14 412.51 434.66 668.54 74.92 79.55 942.30	0.59 0.63 0.40 0.38 0.70 0.41 0.87 0.17 0.13 0.28 0.72 0.47 0.31 0.48 0.60 0.39 0.31 0.60 0.05 0.64	.7.09 -6.18 -5.50 -4.92 -4.15 -4.43 -4.29 -4.15 -3.37 -3.37 -3.37 -2.59 -2.59 -2.59 -2.54 -2.54 -2.54 -2.54 -2.54 -2.54 -2.54 -2.55 -2.54 -2.55 -2.54 -2.55	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

	catalytic ac
	oxidoreductas
	hydrolase a
	lyase acti
	monooxygenas
	transporter a
	transferase a
	transferase activ
	ligase act
	hydrolase activit
)	isomerase a
	transferase activ
	carboxy-lyase peptidase a
	peptidase a
ı	G-protein coupled
;	carboxylic ester l
,	serine-type peptic
	aminopeptidas
	transaminase
,	methyltransfera
)	carboxypeptida metallopeptida
	metallopeptidas
	ATP bine
,	dioxygenase
ı	peroxidase a
,	antioxidant a
,	threonine-type en
	nucleotidyltransf
:	porin acti
•	serine-type endop
)	phospholipase /
	phospholipase
!	prenyltransfera
	toxin acti
	dipeptidase a
į	metalloendopepti
	metalloendopeptis phospholipase A2
	metalloendopepti phospholipase A2 endonuclease
	metalloendopeptis phospholipase A2 endonuclease antigen bii
	metalloendopeptis phospholipase A2 endonuclease antigen bii
,	metalloendopepti phospholipase A2 endonuclease antigen bir translation elong nuclease ac
	metalloendopeptii phospholipase A2 endonuclease antigen bir translation elong muclease ac G-protein coupled
	metalloendopeptii phospholipase A2 endonuclease antigen bir translation elong muclease ac G-protein coupled
	metalloendopeptii phospholipase A2 endonuclease antigen bii translation elong nuclease as G-protein coupled chloride chunn RNA-directed 5:3
	metalloendopeptii phospholipase A2 endonuclease antigen bir translation elong muclease ac G-protein coupled

Order related MF terms (ob	tained fro	om KW 2	GO mapp	ing)	
name	п	avg_len	avg_dis	z	Р
catalytic activity	28179	569.86	0.48	-52.25	0
oxidoreductase activity	4986	450.49	0.30	-39.62	0
hydrolase activity	11097	623.94	0.53	-27.76	0
lyase activity	1624	480.74	0.31	-24.47	0
monooxygenase activity	633	511.13	0.22	-21.40	0
transporter activity	5388	621.16	0.53	-19.15	0
transferase activity	10428	619.96	0.56	-18.80	0
ransferase activity, trans	1343	545.86	0.41	-18.49	0
ligase activity	1097	675.42	0.46	-16.68	0
ydrolase activity, acting	934	548.13	0.39	-16.32	0
isomerase activity	1046	425.93	0.35	-14.19	0
ransferase activity, trans	1235	540.97	0.43	-13.31	0
carboxy-lyase activity	292	494.47	0.26	-13.09	0
peptidase activity	2211	644.11	0.53	-13.05	0
protein coupled receptor	1473	465.64	0.40	-12.30	0
rboxylic ester hydrolase	628	482.21	0.37	-12.01	0
rine-type peptidase activity	708	614.43	0.46	-10.79	0
aminopeptidase activity	177	633.80	0.35	-10.74	0
transaminase activity	230	459.43	0.26	-10.46	0
methyltransferase activity	947	607.08	0.48	-8.77	0
carboxypeptidase activity	143	596.81	0.34	-8.54	0
metallopeptidase activity	632	663.90	0.56	-8.44	0
ATP binding	7585	801.90	0.69	-8.22	0
dioxygenase activity	419	604.98	0.48	-7.83	0
peroxidase activity	282	414.57	0.34	-7.11	0
antioxidant activity	499	344.42	0.36	-6.89	0
reonine-type endopeptidas	139	247.87	0.19	-6.79	0
ucleotidyltransferase acti	892	805.20	0.59	-6.07	0
porin activity	76	369.24	0.22	-5.22	0
rine-type endopeptidase i	218	485.65	0.40	-4.99	0
phospholipase A2 activity	93	377.39	0.30	-4.93	0
phospholipase D activity	37	778.27	0.43	-4.87	0
prenyltransferase activity	128	369.37	0.32	-4.70	0
toxin activity	285	349.98	0.32	-4.59	0
dipeptidase activity	41	557.73	0.34	-4.23	0
etalloendopeptidase activity	369	711.73	0.63	-4.22	0
nospholipase A2 activity (	70	365.39	0.30	-3.90	0
endonuclease activity	773	658.33	0.59	-3.85	0
antigen binding	277	345.87	0.38	-3.68	0
anslation elongation fact	116	464.12	0.46	-3.35	0
nuclease activity	1124	659.17	0.61	-3.20	0
protein coupled photorece	44	390.23	0.30	-2.99	0
chloride channel activity	195	622.49	0.62	-2.67	0.01
A-directed 5'-3' RNA poly	75	2341.31	0.87	-2.57	0.02
tRNA binding	247	566.35	0.57	-2.39	0.01
peptidase inhibitor activity	446	493.46	0.50	-2.38	0.01
photoreceptor activity	91	574.22	0.54	-2.30	0.01
ican endo-1,3-beta-D-gluc	28	522.18	0.43	-2.29	0.02
signal transducer activity	4149	619.06	0.62	-2.17	0.02
inoacyl-tRNA ligase activity	255	659.40	0.67	-1.93	0.04
steine-type endopeptidase	262	953.41	0.64	-4.65	.0.05

#### Results Tabular comparison

1

1

0.99

2.89

Disorder related MF keywords (z-skor > 0.2)

1027 861.88

386 816.98 0.90 2.46 2.25 2.19

39 1202.79

725.99 24 712.75

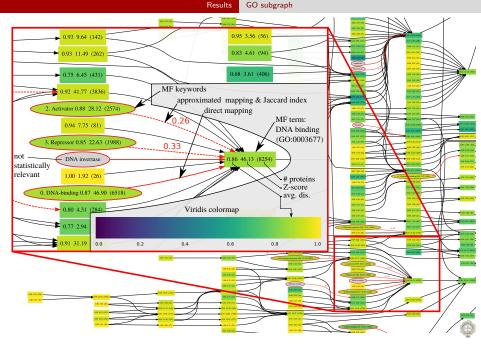
Voltage-gated channel

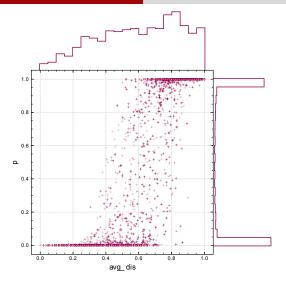
Tumor antigen

									#	name	n	avg_len	avg_dis	z	Ī
									0	DNA-binding	6518	546.53	0.87	46.90	
									T	Developmental protein	3897	655.21	0.86	31.10	Ī
									2	Activator	2574	600.51	0.88	28.12	Ī
Т	op 20 MF kw with strongest co	rrelatio p	redicted di	sorder (X	ie et. al.	2007)	1	//	3	Repressor	1988	589.29	0.85	22.63	Ī
#	name	n prot	n families	avg_len	Z-score	P-value	1 /	/ //	4	RNA-binding	2728	575.76	0.76	16.62	I
0	Ribonucleoprotein	12236	412	150.55	22.13	1	$\sim$	///	5	Chromatin regulator	1038	847.06	0.90	13.91	Ī
1	Ribosomal protein	11692	330	140.58	20.63	1	$V^{-}$	7//	6	Ribonucleoprotein	1886	272.29	0.60	13.39	İ
2	Developmental protein	3260	721	477.93	19.28	1	$r \sim$	///	7	Serine/threonine-protein ki	1782	802.24	0.84	11.56	Ī
3	Hormone	1187	161	141.13	15.58	1	lΛ	X.	8	Chaperone	937	430.43	0.71	10.02	
4	Growth factor	785	84	255.70	11.16	1	14/	// >	9	Ribosomal protein	1408	186.38	0.53	9.34	Ī
5	Cytokine	899	110	213.28	10.21	- 1	1 /X	7	10	Growth factor	398	299.63	0.70	8.98	Ī
6	Neuropeptide	268	209	95.08	9.65	1	V //	\	11	Protein kinase inhibitor	49	337.20	0.96	8.34	
7	Activator	3086	573	428.47	9.04	1	Y\//	/	12	Calmodulin-binding	520	1229.00	0.90	7.57	I
8	GAP protein	47	2	232.96	7.42	- 1	1 <i>X</i>	*	13	Hormone	338	221.13	0.59	7.24	Ī
9	Antigen	1113	455	437.48	6.99	- 1	I//\		14	Cyclin	133	422.71	0.87	7.18	Ī
10	Repressor	2309	449	374.46	6.92	1	// /	· /	15	Signal transduction inhibitor	115	408.43	0.84	6.76	I
11	Chromatin regulator	334	100	801.24	6.70	1	Y	\/	16	Guanine-nucleotide releasin	319	1144.39	0.96	6.40	I
12	Pyrogen	37	2	262.59	6.44	1	1	Хя	17	GTPase activation	424	867.35	0.88	6.28	Ī
13	Vasoactive	125	39	160.39	5.56	1	l .	М	18	Growth factor binding	50	593.98	1	6.09	I
14	Amphibian defense peptide	123	148	50.64	5.44	1	\//	/ •	19	Neuropeptide	105	234.96	0.68	5.99	
15	GTPase activation	311	70	831.03	5.36	1	KV		20	Potassium channel	191	621.52	0.85	5.10	I
16	Endorphin	42	4	226.68	5.35	- 1	1 M		21	Calcium channel	193	1397.77	0.93	5.09	Ī
17	Opioid peptide	24	4	216.96	5.14	- 1	lΛ,	↳	22	Protein phosphatase inhibitor	64	352.86	0.81	5.07	Ī
18	Protein phosphatase inhibitor	47	8	366.51	5.07	1	17	_	23	Tyrosine-protein kinase	376	863.00	0.89	5.04	Ī
19	Cyclin	182	25	430.58	4.88	1	γV	, \	24	Mitogen	137	286.03	0.68	4.80	Ī
								\ *	25	Vasoactive	46	267.00	0.76	4.29	
								\	26	Heparin-binding	221	650.97	0.73	3.88	I
								-\	27	Muscle protein	193	920.01	0.73	3.81	I
								\	28	Actin-binding	837	974.92	0.77	3.80	Ĭ
								•	29	Amphibian defense peptide	49	85.80	0.53	3.34	ĺ
									30	Helicase	739	1086.05	0.87	3.29	Í
									31	Prion	22	497.05	0.91	3.06	ĺ

	nome		-		_	_
# 0		n 8254	avg_len	avg_dis	z	P
_	DNA binding		574.89	0.86	46.13	·
1	sequence-specific DNA binding	3836	536.95	0.92	41.77	1
3	chromatin binding	1538	713.64	0.93	23.54	1
4	RNA binding	5797 4109	565.80	0.73	19.79	1
•	receptor binding		529.68	0.68		
5	protein serine/threonine ki	1942	784.87	0.84	12.88	1
6	structural molecule activity	2884	558.88	0.64	10.84	1
7	structural constituent of r	1350	182.95	0.53	9.25	1
8	calmodulin binding	710	1079.79	0.88	9.22	1
9	growth factor activity	458	312.38	0.70	8.99	1
10	protein kinase inhibitor ac	237	401.60	0.82	8.94	1
11	hormone activity	452	218.09	0.61	8.84	1
12	cell adhesion molecule binding	1090	805.74	0.80	8.60	- 1
13	GTPase activator activity	633	820.60	0.87	8.15	-1
14	actin binding	1228	1000.47	0.80	7.73	1
15	cyclin-dependent protein se	142	357.65	0.77	6.58	1
16	protein tyrosine kinase act	504	933.02	0.86	6.26	-1
17	guanyl-nucleotide exchange	575	990.55	0.86	6.22	-1
18	protein phosphatase inhibit	89	425.83	0.83	6.19	1
19	heparin binding	403	652.76	0.76	6.01	-1
20	neuropeptide hormone activity	117	158.64	0.65	5.88	-1
21	microtubule motor activity	225	1394.78	0.97	5.65	- 1
22	enzyme inhibitor activity	1059	462.66	0.63	5.60	-1
23	potassium channel activity	339	612.39	0.79	4.88	1
24	damaged DNA binding	284	680.80	0.80	4.31	-1
25	growth factor binding	351	737.85	0.79	4.20	1
26	chloride channel inhibitor	25	827.64	0.96	3.96	1
27	calcium channel activity	327	1205.13	0.86	3.54	1
28	morphogen activity	23	331.57	0.83	3.20	1
29	helicase activity	766	1068.02	0.86	3.15	1
30	motor activity	447	1275.57	0.86	3.10	1
31	voltage-gated ion channel a	501	767.73	0.76	3.06	1
32	rRNA binding	541	266.55	0.52	3.00	1
33	calcium ion binding	1878	859.00	0.70	2.80	1
34	RNA-directed DNA polymerase	75	1427.97	0.97	2.72	
35	kinase activity	3593	726.02	0.71	2.70	1
36	ATP-dependent helicase acti	455	926.44	0.84	2.48	1
37	exonuclease activity	371	749.41	0.75	2.40	0.9
38	translation initiation fact	268	482.02	0.65	1.99	0.9
39	acetylcholine-gated cation	60	504.17	0.75	1.99	0.9
40	metalloendopeptidase inhibi	40	323.35	0.65	1.86	0.9
41	cysteine-type endopeptidase	153	424.63	0.58	1.48	0.9







Level of disorder (avg\_dis) and p-value (1781 MF terms)



# Thank you

