Bioinformatics III

Fifth Assignment

Thibault Schowing (2571837)
Wiebke Schmitt (2543675)

May 22, 2018

Exercise 5.1: asdf

(a) *asdf*

(b) *asdf*

(c) *asdf*

Exercise 5.2: aaas

(a) *Adding annotations to PPI-networks*

The listings for this exercise are at the end of the document.

(b) *Generating an overview*

For Chicken:

Table 1: Chicken network overview

Interactions in the network	300				
Proteins in the network	281	Protein without annotation	44	Percentage	15.6
Annotation per protein					
Smallest number	0	Average number	7.7	Biggest number	88
Protein per annotation					
Smallest number	1	Average number	1.55	Biggest number	27

For pig:

Table 2: Pig network overview

Interactions in the network	50				
Proteins in the network	51	Protein without annotation	13	Percentage	25.5
Annotation per protein					
Smallest number	0	Average number	5.5	Biggest number	40
Protein per annotation					
Smallest number	1	Average number	1.13	Biggest number	5

for Human:

Table 3: Human network overview

Interactions in the network	275472				
Proteins in the network	17087	Protein without annotation	2262	Percentage	13.2
Annotation per protein					
Smallest number	0	Average number	7.22	Biggest number	184
Protein per annotation					
Smallest number	1	Average number	10.6	Biggest number	1554

(c) *Examining the most/least common annotations*

Table 4: Function of the 5 most common GO identifiers of the human network.

GO id	**Quantity**	**Biological Process**
GO:0006351	1562	The cellular synthesis of RNA on a template of DNA.
GO:0045944	1029	Any process that activates or increases the frequency, rate or extent of transcription from an RNA polymerase II promoter.
GO:0007165	1010	Signal transduction
GO:0006357	960	Any process that modulates the frequency, rate or extent of transcription mediated by RNA polymerase II.
GO:0006355	765	Any process that modulates the frequency, rate or extent of cellular DNA-templated transcription

We can observe that these annotations concerns general process happening almost in every cell. This explains why they are the most common in opposition as the annotations in the table below, which concerns specific reaction or process concerning particular location on molecules.

Table 5: Function of the 5 least common GO identifiers of the human network

GO id	**Quantity**	**Biological Process**
GO:0000003	1	Reproduction
GO:0000011	1	Vacuole inheritance
GO:0000032	1	Cell wall mannoprotein biosynthetic process
GO:0000053	1	Argininosuccinate metabolic process
GO:0000097	1	Sulfur amino acid biosynthetic process

(d) *Investigating annotation enrichment*

Table 6: My caption

	Number	Percentage
p <0.05		
p >0.5		
p >0.95		

Table 7: Annotations with the five lowest pA and five highest pA

GO:ID	**pA**	**Nb Protein**	**Nb Interact. protein**	**Annotation**
GO:0009409	4.3907e-07	3	3	Response to cold
GO:0030154	1.7908e-05	7	4	Cell differentiation
GO:0007169	0.0002	3	2	Transmembrane receptor protein tyrosine kinase signaling pathway
GO:0000712	0.0002	3	2	Resolution of meiotic recombination intermediates
GO:0032570	0.0002	3	2	Response to progesterone
GO:0007049	1	10	0	Cell cycle
GO:0006096	1	9	0	Glycotic process
GO:0055114	1	9	0	Oxydation-reduction process
GO:0006457	1	9	0	Protein folding
GO:0006094	1	8	0	Gluconeogenesis

(e) *e) Investigating annotation combinations*