Questionnaire

Your name: [redacted]

Your email address: [redacted]

Your institution: Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK

Your position:

Undergraduate student	[ ]
PhD student	[ ]
Technician	[ ]
Postdoc	[ ]
Senior Postdoc	[X]
Faculty member	[ ]

Question 1: In a scale from 1 - 10, how important do you think it is to have quick access to the following type of information about each gene or protein?

a. Protein-protein interactions [1-10]:8

b. General function of the gene/protein [1-10]:8

c. Diseases a gene/protein is involved in [1-10]:9

d. Biochemical pathways a gene/protein is part of [1-10]:10

e. Cell types/tissues where your gene/protein is expressed [1-10]:8

f. Your protein's 3D structure (PDB) [1-10]:5

g. Popularity of the gene/protein in **social networks (Twitter, Facebook)** [1-10]:1

h. Knowing the **average impact factor** of the journals where a particular gene/protein is normally published [1-10]:1

i. The **relative scientific weight (e.g. by h-index)** of the scientists that work on your gene/protein [1-10]:2

j. How popular your gene/protein is in recently awarded grants (this is public information once a grant is awarded) [1-10]:2

k. What other genes/proteins are discussed in the context of your protein [1-10]:8

l. How your gene/protein is regulated at the transcriptional level [1-10]:10

m. How your gene/protein is regulated post-translationally (phosphorylation, ubiquitination) [1-10]:7

n. What is the **most popular type of experiment** other scientists typically do on your gene/protein [1-10]:4

o. What **biochemical kits** are available for doing these experiments [1-10]:4

p. Other (explain what type of information) [1-10]:

Question 2: What websites do you visit the most when analysing your list of genes/proteins?
What type of information do you expect to get from each of these websites?

Some examples of websites include:

- Ensembl (www.ensembl.org)
- NCBI's Entrez (https://www.ncbi.nlm.nih.gov/Class/MLACourse/Original8Hour/Entrez/)
- NCBI Databases (https://www.ncbi.nlm.nih.gov/search/)
- EuPathDB.org (Eukaryotic Pathogens Database)
- Galaxy (https://usegalaxy.org/)
- PubMed (https://www.ncbi.nlm.nih.gov/pubmed/)
- UniProt (http://www.uniprot.org/)
- KEGG Pathway Database (https://www.genome.jp/kegg/pathway.html)
- Any other resource you use routinely.

Website	**Type of information sought**	**Priority in your analytical pipeline**
Example: PubMed	Find out what's been	1

	published about my gene or protein	
Enrichr		
GeneCards		
MGI-mouse phenotypes		
OMIM		
(expand the table as needed)		

Question 3: How often do you perform these exploratory analyses on your genes or proteins:

[a] Daily [ ]

[b] Weekly [ ]

[c] Monthly [X]

[d] Several times a year [ ]

[e] Other (explain) [ ]

Question 4: If you could obtain the same type of information that you seek by doing these analyses **in 5 minutes only**, how often would you now perform these analyses?

[a] Daily [ ]

[b] Weekly [ ]

[c] Monthly [X]

[d] Several times a year [ ]

[e] Other (explain) [ ]

Question 5: When you get **a list of genes/proteins** from a proteomics or a differential expression experiment, what **steps and tools** do you follow for the analysis?

1. Run through Enrichr/GSEA

2. Check expression and genetic phenotypes at the MGI site

3. Check antibodies available at GeneCards

4.

Question 6: Now and related to the previous question, instead of telling me what steps you follow in your analysis, if I asked you **what type of information you would like to know** about each one of your genes or proteins, what would you be interested in knowing?

Here you can include some information you would love to have **but do not know** how to obtain it.

1. Contextual information for genes and proteins in biochemical pathways and TXN routes.

2. Reagents available for any follow-up experiments (GeneCards)
3. The most relevant field for the gene or protein.
4. Other diseases and contexts where the gene and protein has not been studied.

Question 7: If you have a long list of genes/proteins from a high-throughput experiment you ran in the lab, **what are the most useful factors in determining the next follow-up experiment?**

E.g. how easy the potential experiment is, how relevant the cell type, how much money this would cost, etc.

1. The transcription factors regulating the expression of my gene, and where they bind.

2. How easy is to do the experiments in terms of reagents available and how well they work.

3.

4.

3.

4.

5.

Question 8: What do you think is the **most competitive advantage against other competing laboratories** when trying to decide what the next follow-up experiment would be?

1. Getting the most information from the analysis of the genes and by doing a small set of experiments.

2.

3.

4.

5.

__END OF QUESTIONNAIRE__

Please return this questionnaire to d.mirandasaavedra@gmail.com