Questionnaire

Your name: [redacted]

Your email address: [redacted]

Your institution: The Francis Crick Institute

Your position: Senior Research Scientist

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

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]: 9

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

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

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

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

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

g. Popularity of the gene/protein in **social networks (Twitter, Facebook)** [1-10]: 7

h. Knowing the **average impact factor** of the journals where a particular gene/protein is normally published [1-10]: 9

i. The **relative scientific weight (e.g. by h-index)** of the scientists that work on your gene/protein [1-10]: 8

j. How popular your gene/protein is in recently awarded grants (this is public information once a grant is awarded) [1-10]: 9

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]: 8

m. How your gene/protein is regulated post-translationally (phosphorylation, ubiquitination) [1-10]: 8

n. What is the **most popular type of experiment** other scientists typically do on your gene/protein [1-10]: 9

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

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 published about my gene or protein	1

Ensembl (www.ensembl.org)	Genetic features (i.e.: gene structure, exons, alternative splicing, transcription and translational regulation)	Depending on the project the priority could go from 5 to 10.
UniProt (http://www.uniprot.org/)	Basic information about the protein of interest. I'm particularly interested about the proteins descripted function	10
PubMed	Find out what's been published about my gene or protein	10
KEGG Pathway Database (https://www.genome.jp/kegg/pathway.html)	In general, I want to know which signaling pathway is relevant as I work in signalling, however I don't find KEGG as the most suitable platform to find this out.	9
IMMGEN (http://www.immgen.org/)	Expression level profiles on the immune system	10
NCBI's Entrez (https://www.ncbi.nlm.nih.gov/Class/MLACourse/Original8Hour/Entrez/)	Don't really use this until I have read the literature	4

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

[a] Daily [ ]

[b] Weekly [ ]

[c] Monthly [ ]

[d] Several times a year [X]

[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	[ ]
[d] Several times a year	[X]
[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. I personally don't do the proteomics analysis. I ask the bioinformaticians to check on relevant pathways affected

2. When I have candidate genes/proteins selected, I search for a general description of the gene product. I.e: what is known about this protein in general terms? In general, I google the protein, and trust information coming from Generacard, Uniprot, NCBI Gene, Ensembl. I sometimes read Wikipedia as well but check on sources for reliability of the information.

3. tissue/cell type expression. Is it relevant to my system? As an immunologist I check Immgen.org for expression patterns. But I also use get info from NCBI.

3. signalling pathway. Google + KEGG

3. interactome/ binding partners. *https://bio.tools › ihop ;;* ***iHOP - omicX***

5. Role in disease. OMIM

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?

1. the same as in question 5
2. reagents available and protocols already tried for testing protein function
3.
4.
5.

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. relevance to experimental system

2. availability of reagents

3. cost

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. how easy and fast are the experiments for the follow-up

2. Novelty. I.e.: would it be easy to predict the target even without a screen?

3.

4.

5.

__END OF QUESTIONNAIRE__