Questionnaire

Your name: [redacted]

Your email address: [redacted]

Your institution: Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC)

Your position:

Undergraduate student	[ ]
PhD student	[x]
Technician	[ ]
Postdoc	[ ]
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]: 9

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

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]: 9

f. Your protein's 3D structure (PDB) [1-10]: 8

g. Popularity of the gene/protein in **social networks (Twitter, Facebook)** [1-10]: 5

h. Knowing the **average impact factor** of the journals where a particular gene/protein is normally published [1-10]: 7

i. The **relative scientific weight (e.g. by h-index)** of the scientists that work on your gene/protein [1-10]: 7

j. How popular your gene/protein is in recently awarded grants (this is public information once a grant is awarded) [1-10]: 8

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

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]: 10

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

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

p. Other (explain what type of information) [1-10]: the splicing variants of your gene, orthologs, parologs and homologs of your gene/protein, best antibodies you can use to study phosphorylation sites of your protein

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	1

	protein	
GeneCards	Detailed information on a gene and protein, and biochemical kits for it, including antibodies available	1
UniProt	A detailed and high quality annotation for my protein	1
NCBI's Entrez	The encyclopaedia for each gene or protein where I can also run some database analyses.	1
PDB	Three-dimensional structure	
KEGG Pathway Database	Signalling pathway analysis	1
(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 [ ]

[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 [x]

[c] Monthly [ ]

[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?

First of all, I would try to gather information about these proteins or genes in a public database like UniProt or Entrez. Once I have the protein sequence, I would run a BLAST analysis to identify homologs and highlight conserved regions and the phylogeny.

Also relevant is to find out the protein interactions (UniProt) and where the genes are expressed (tissues, cell types), as well as their involvement in diseases (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?

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

1. Which transcription factors and routes regulate the expression in my cell type of interest.

2. Any protein binding partners in the cytosol or the nucleus?

3. Are there studies of differential gene expression for this gene in my cell type of interest in multiple conditions?

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 involvement of the gene or protein in specific diseases, and how much is known about the disease.

2. how much intelligence I can collect about the gene or protein

3. how easy the experiments are

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. Being able to think of the right experiment, and do it quickly and be able to place the results in the right context.

2.

3.

4.

5.

__END OF QUESTIONNAIRE__