Questionnaire

Your name: [redacted]

Your email address: [redacted]

Your institution: Dartmouth College

Your position:

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]:

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

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

d. Biochemical pathways a gene/protein is part of [1-10]: 2

e. Cell types/tissues where your gene/protein is expressed [1-10]: 4

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

g. Popularity of the gene/protein in **social networks (Twitter, Facebook)** [1-10]: 8

h. Knowing the **average impact factor** of the journals where a particular gene/protein is normally published [1-10]: 8

i. The **relative scientific weight (e.g. by h-index)** of the scientists that work on your gene/protein [1-10]: 9

j. How popular your gene/protein is in recently awarded grants (this is public information once a grant is awarded) [1-10]: 6

k. What other genes/proteins are discussed in the context of your protein [1-10]: 3

l. How your gene/protein is regulated at the transcriptional level [1-10]: 4

m. How your gene/protein is regulated post-translationally (phosphorylation, ubiquitination) [1-10]: 3

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

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

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

NCBI Databases	Gene info	1
UniProt	Protein info	1
https://www.genecards.org/		1
The Cancer Genome Atlas Program - National Cancer Institute	Cancer cases, types of cancer, genes associated, transcriptome profiling, mutation types,	2
http://www.cbioportal.org/		2
https://cancer.sanger.ac.uk/cosmic/signatures		2
https://www.cancerrxgene.org/	Drug/compound search cancer treatment	3

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

[a] Daily [ ]

[b] Weekly [x]

[c] Monthly [ ]

[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 [x]

[b] Weekly [ ]

[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?

1. I only use excel, prioritize based on largest fold change

2. **Reproducibility/repetition** of similar fold change between experiments

3. Number of hits, i.e. number of peptides detected in a proteomics screen

4. Significance and % confidence

5.

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. Fastest and **best** way to determine % of confidence of gene hits
2. Information on gene/ type of disease association (if known)
3. Rapid and easy gene analysis information (mutations, SNPs, etc) already available
4. Signaling pathways the protein/gene is involved in
5. Hits grouped based on common signaling pathways?
6. Hits grouped based on functional/enzymatic roles?
7. Homologs/orthologs of hits grouped based on signaling pathways

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. Interest in further analyzing the “big hits”. Is there an urgent need to investigate a gene or set of genes further? Assessment of the impact of further investigation on future publications and funding applications

2. Availability of reagents

3. Amount of time experimental approaches will take

4. Amount and level of competition on same genes

5. Cost of experiments

6. Availability of on-site collaborations with specialists that could accelerate the process of the investigation and potentially minimize costs

7.

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. Funding

2. Number and type of specialists working in the lab (post-docs, researchers)

3. Facilities available on-site (platforms)

4. Collaborations

5.

__END OF QUESTIONNAIRE__