Questionnaire

Your name: [redacted]

Your email address: [redacted]

Your institution: Jawaharlal Nehru University

Your position: Assistant Professor

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]: 7

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

c. Diseases a gene/protein is involved in [1-10]: 8

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

e. Cell types/tissues where your gene/protein is expressed [1-10]: 6

f. Your protein's 3D structure (PDB) [1-10]: 7

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]: 3

i. The **relative scientific weight (e.g. by h-index)** of the scientists that work on your gene/protein [1-10]: 3

j. How popular your gene/protein is in recently awarded grants (this is public information once a grant is awarded) [1-10]: 4

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]: 10

n. What is the **most popular type of experiment** other scientists typically do on your gene/protein [1-10]: Expression profiling, coexpression, protein-protein interaction, identify downstream targets, mechanism of regulation, level of conservation across species, biochemical activity, subcellular localization, structural elucidation, functional characterization using transgenic approaches, etc.

o. What **biochemical kits** are available for doing these experiments [1-10]: Most of the commercial kits are available through major R&D companies like Sigma, Life technologies, Agilent, Qiagen, Thermofischer scientific, Ambion, Amersham, etc.

p. Other (explain what type of information) [1-10]: **Although I have answered questions a) to m) keeping in mind some of the genes our lab works on but answer to these really depend on the aim of your experiment and can vary from gene to gene. Considering very low number of genes characterized till date in plants, any information hinting on the gene function is useful and would be rated 10 by a researcher trying to understand its mechanism of function. However, personally I believe it is most beneficial for a plant biologist using high throughput omics technologies to be able to map your genes on relevant pathways.**

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
Plant genome databases like Phytozome, MSU rice database, TAIR etc.	Get information on structural attributes about the shortlisted genes	1
Gene Expression Omnibus	Gene Expression Patterns	2
Pathway mapping tools eg. KEGG, gramene, mapman etc.	Map on Pathways	3
Gene Ontology Analysis tools	Gene Ontology enrichment	4
Databases on gene networks e.g. RiceNet, String etc.	Get list of coexpressed/related genes and pathways	5
(expand the table as needed)		

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

[a] Daily [X]

[b] Weekly [ ]

[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. Gene Ontology Enrichment (Varies based on genome)

2. Mapping on Known Pathways (listed in table above)

3. Orthology Analysis (mostly Phytozome)

4. Experimental Validation

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. Mine literature to extract all possible information about the gene of interest.

2. For many genes, there are different names/accession IDs used in different studies that makes searching for relevant information difficult.

3. Be able to extract the data from high throughput experiments which were never submitted to any repository would be great but challenging. Most of the times, relevant information is buried in supplementary files and are hard to search unless you bump into relevant paper and read it thoroughly.

4. Ability to predict level of redundancy among genes with high accuracy to predict if there would be a phenotype expected on overexpression/silencing of relevant gene.

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. Being able to shortlist candidates for follow up experiments is real challenge for biologists. The other challenge is to predict level of redundancy among candidates. So, if one can make a network of genes where you virtually delete one gene and can see what happens to the rest of the pathway would be highly useful to predict and score the role of genes of interest. A scoring system to predict chances of a gene's involvement in a particular pathway would also be highly useful.

2.

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. Access and ability to mine high throughput omics data to predict best candidates for in-depth characterization.

2.

3.

4.

5.

__END OF QUESTIONNAIRE__

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