Introduction to Proteogenomics

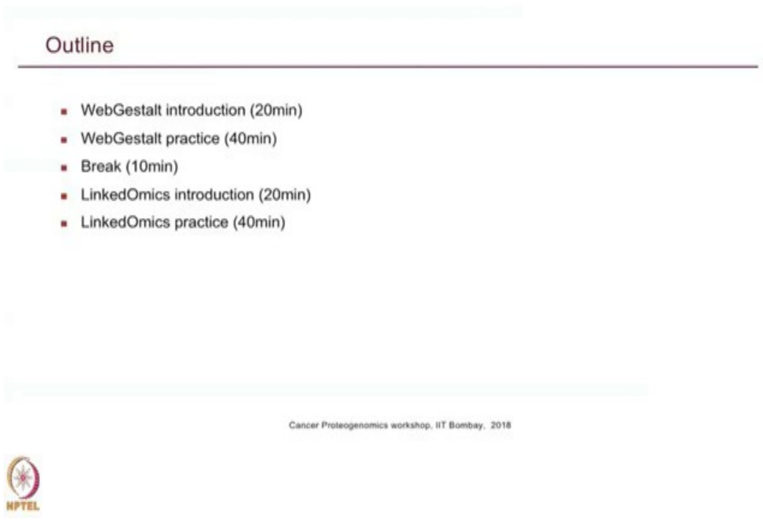
Dr. Sanjeeva Srivastava
Dr. Bing Zhang
Department of Biosciences and Bioengineering
Indian Institute of Technology, Bombay
Baylor College of Medicine

Lecture – 49
WebGestalt – I

Welcome, to MOOC course on Introduction to Proteogenomics. This is Dr. Bing Zhang's hands on session on WebGestalt. WebGestalt or Web based Gene Set Analysis Toolkit is a functional enrichment analysis web tool. Dr. Bing Zhang will briefly describe about the software and its three well established and complimentary methods for enrichment analysis including over representation analysis (ORA) gene set enrichment analysis (GSEA) and network topology based analysis (NTA). He will also discuss about different parameters of each method input gene list and run the analysis. So, let us welcome Dr. Bing Zhang for today's session.

I am going to introduce two tools that have been developed in my lab mostly for pathway and the network-based analysis and also for using these tools to analyze data that has been generated by TCGA and CPTAC cancer programs.

(Refer Slide Time: 01:42)



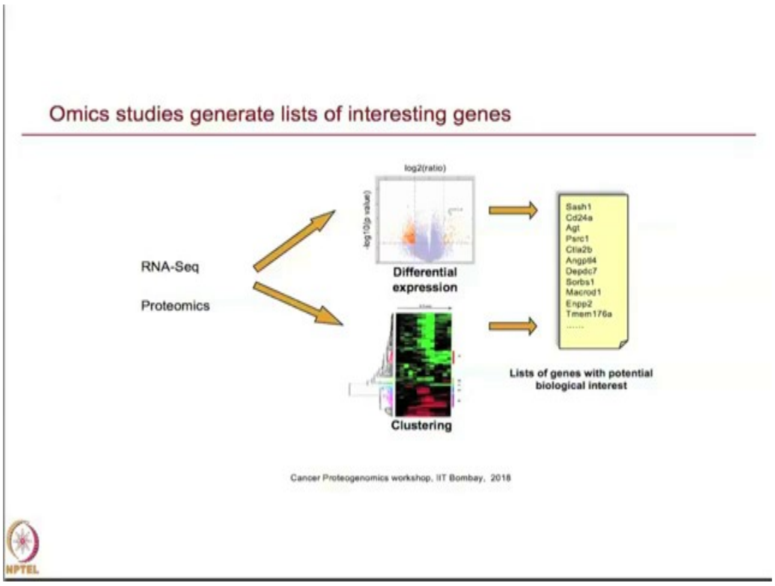
Outline

- WebGestalt introduction (20min)
- WebGestalt practice (40min)
- Break (10min)
- LinkedOmics introduction (20min)
- LinkedOmics practice (40min)

Cancer Proteogenomics workshop, IIT Bombay, 2018

So, one of the tool is called WebGestalt, the other one is called LinkedOmics. So, I think I have given a bit of structure for this hands on session. So, first I want to keep a very brief introduction to the WebGestalt tour and then we can have some practice and then we can have a little bit break if you want or we can go through the LinkedOmics introduction and hands on part, because I think the all the 9 method has been described both in some of the Karsten’s talk and also my talk yesterday. So, the introduction can go very quickly.

(Refer Slide Time: 02:20)



Omics studies generate lists of interesting genes

Sash1
Cd24a
Agt
Parc1
Ctla2b
Angptl4
Depdc7
Sorbs1
Macrod1
Enpp2
Tmem176a
....

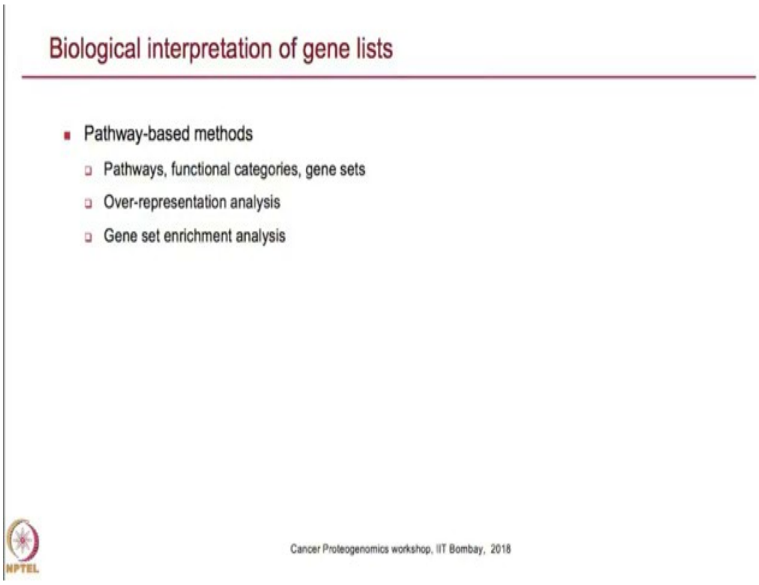
Lists of genes with potential biological interest

Cancer Proteogenomics workshop, IIT Bombay, 2018

So, for the WebGestalt so, this actually was my first bioinformatics project and it was through I created when I was a post doc. So, the motivation that I mean when you do RNA-Seq or proteomics and this high-throughput technologist actually when I started this I was doing microarray at that time and then you need to like differentiate expression analysis and clustering analysis is type of science right.

And eventually you end up with lists of genes data of possible interest on like a list of differentially expressed genes or through clustering analysis introduced by Dr. Mani during the previous lectures. You are going to get clusters of genes that have similar expression pattern, you want to explore those genes.

(Refer Slide Time: 03:28)



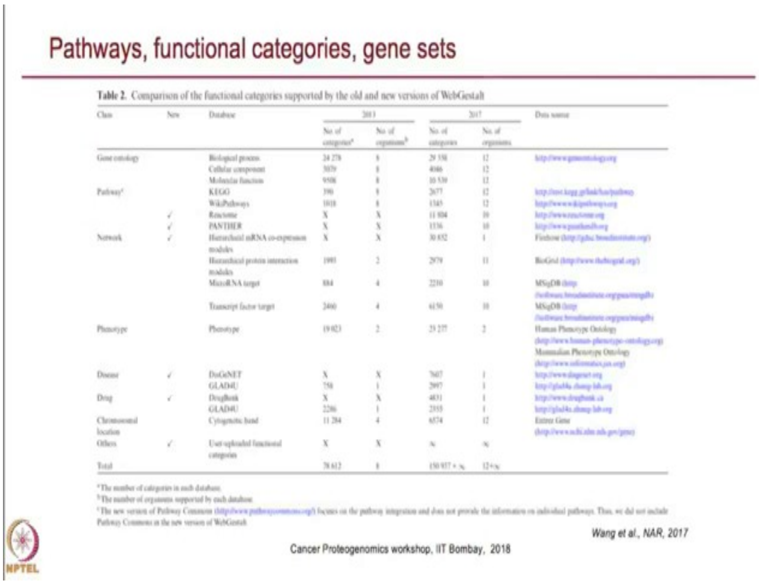
Biological interpretation of gene lists

- Pathway-based methods
  - Pathways, functional categories, gene sets
  - Over-representation analysis
  - Gene set enrichment analysis

Cancer Proteogenomics workshop, IIT Bombay, 2018

And, one way to do this is to leverage the information we already have about the pathways. So, and the typical path first in order to do this you have to have some pathway definition right and like the pathway databases or actually it is it can be also like loosely defined functional categories or just kind of any gene sets and then there are two types of analysis over representation analysis and gene set enrichment analysis. I think we already talked about this yesterday in Karsten's lectures.

(Refer Slide Time: 03:58)



Pathways, functional categories, gene sets

Table 2. Comparison of the functional categories supported by the old and new versions of WebGestalt

Class	New	Database	2013		2017		Data source
			No. of categories[a]	No. of organisms[b]	No. of categories	No. of organisms	
Gene ontology		Biological process	24 278	8	29 558	12	http://www.geneontology.org
		Cellular component	3079	8	4046	12	
		Molecular function	9508	8	10 539	12	
Pathway[c]		KEGG	390	8	2677	12	http://rest.kegg.jp/link/hsa/pathway
		WikiPathways	1018	8	1345	12	http://www.wikipathways.org
	✓	Reactome	X	X	11 804	10	http://www.reactome.org
	✓	PANTHER	X	X	1136	10	http://www.pantherdb.org
Network	✓	Hierarchical mRNA co-expression modules	X	X	30 852	1	Firehose (http://gdac.broadinstitute.org/)
		Hierarchical protein interaction modules	1993	2	2979	11	BioGrid (http://www.thebiogrid.org/)
		MicroRNA target	884	4	2210	10	MSigDB (http://software.broadinstitute.org/gsea/msigdb/)
		Transcript factor target	2460	4	6150	10	MSigDB (http://software.broadinstitute.org/gsea/msigdb/)
Phenotype		Phenotype	19 023	2	23 277	2	Human Phenotype Ontology (http://www.human-phenotype-ontology.org) Mammalian Phenotype Ontology (http://www.informatics.jax.org)
Disease	✓	DisGeNET	X	X	7607	1	http://www.disgenet.org
		GLAD4U	758	1	2997	1	http://glad4u.zhang-lab.org
Drug	✓	DrugBank	X	X	4831	1	http://www.drugbank.ca
		GLAD4U	2286	1	2355	1	http://glad4u.zhang-lab.org
Chromosomal location		Cytogenetic band	11 284	4	6574	12	Entrez Gene (http://www.ncbi.nlm.nih.gov/gene)
Others	✓	User-uploaded functional categories	X	X	∞	∞	
Total			78 612	8	150 937 + ∞	12+∞	

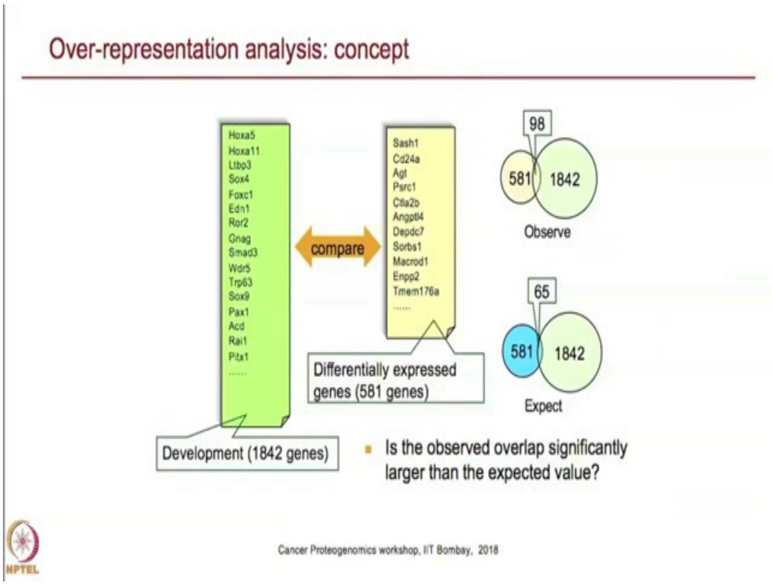
[a] The number of categories in each database.
[b] The number of organisms supported by each database.
[c] The new version of Pathway Commons (http://www.pathwaycommons.org/) focuses on the pathway integration and does not provide the information on individual pathways. Thus, we did not include Pathway Commons in the new version of WebGestalt.

Wang et al., NAR, 2017

Cancer Proteogenomics workshop, IIT Bombay, 2018

So, in WebGestalt, so, we have collected a lot of pathways, functional categories and other types of gene sets. So, basically it is from different databases or through some computational analysis I would not because it is too small I do not think you will be able to see it, but maybe you have the lecture note. So, you should be able to grid on your computer, but it can be separated into different categories. And, in total there is a very large number of gene sets that can be used for your analysis.

(Refer Slide Time: 04:39)



Over-representation analysis: concept

- Is the observed overlap significantly larger than the expected value?

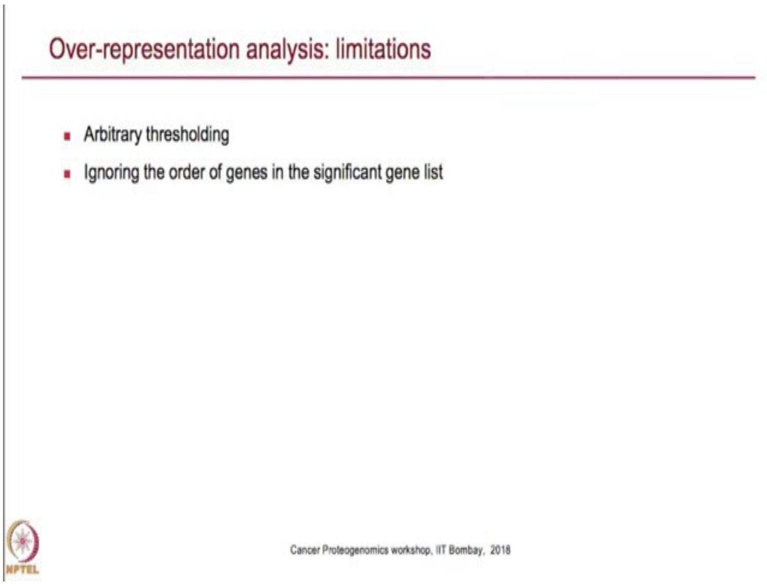
Cancer Proteogenomics workshop, IIT Bombay, 2018

For the over representation analysis just a quick reminder, you have a list of genes of interest let us say you do a RNA-seq or proteomics experiment and then you get a list of differentially expressed genes and then you want to compare with the gene set or pathway group to see whether there is any association between them. For example, we want to compare this with the different gene ontology and then you can do the overlap and then you can count the number of overlapping genes right.

If you say over any overlaps that indicates some of your genes are evolving this biological process, but the question is whether this is an enriched the representation of that category or not you do not know right and then what you can do is to randomly sample the same number of genes as you can from next 581 genes randomly sampled from the proteome you study for example and then you can do the overlap.

And the for example, here you observe the 98 gene overlapping here, but only 65 in a random experiment and then you can say I have more gene overlapping genes now here, but is that significant or not in order to do that you do the Fishers exact test or you can it is also called hyper geometric test and I think Karsten already talked about this yesterday. And, then you can get a p value that can help you to know whether it is significant or not and because when you use gene ontology or other databases to do this type of analysis, you are testing many different biological processes or pathways at the same time. So, you also need to do the multiple test adjustment in order to justify your observation.

(Refer Slide Time: 06:38)



Over-representation analysis: limitations

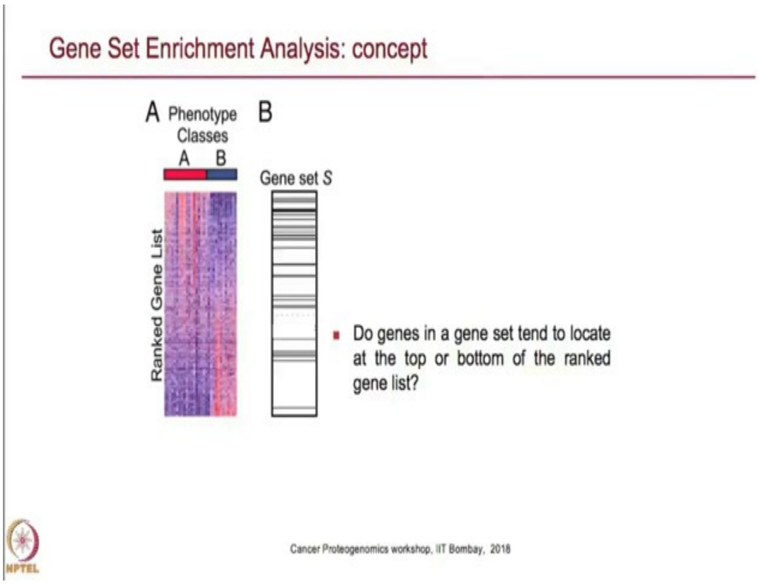
- Arbitrary thresholding
- Ignoring the order of genes in the significant gene list

Cancer Proteogenomics workshop, IIT Bombay, 2018

So, I want to mention the limitation of this approach. First I you have to define what is differentially expressed what is not differentially expressed; that means, you have to set up a cutoff for example, people usually use false discovery rate 0.01 let us say or 0.05 where the cutoff, but the question is like how about the gene right below the cutoff like 0.011, I mean that not useful at all probably not right, but setting this cutoff is very arbitrary.

And, certainly after you set the cutoff for example, for all the remaining genes with a p value or FDR NS 0.01 you consider them as a same right, but actually they are not so same, some of them may have a 10 fold change. But, others may only have a 2 fold change. They are not the same, but if you do this approach basically you consider all the significance genes are same. So, that is a major limitation.

(Refer Slide Time: 07:34)



Gene Set Enrichment Analysis: concept

- Do genes in a gene set tend to locate at the top or bottom of the ranked gene list?

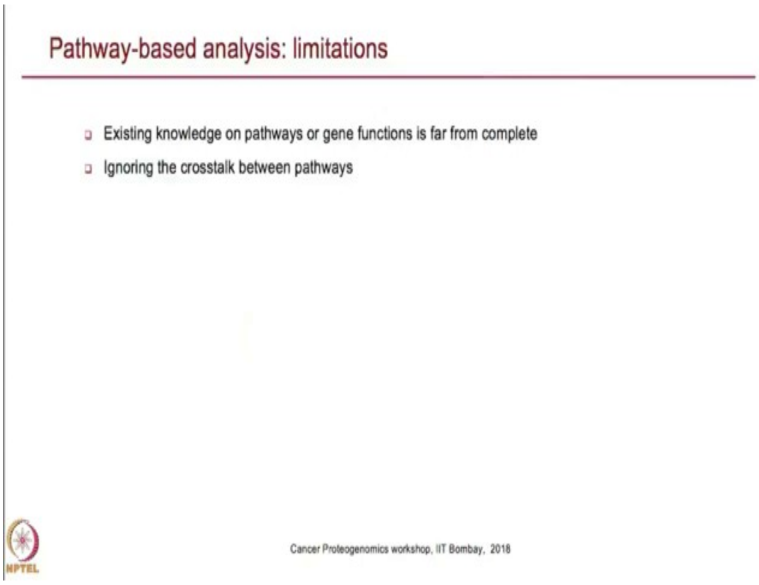
Cancer Proteogenomics workshop, IIT Bombay, 2018

And, the gene set enrichment analysis basically addresses this limitation a user rank needs in order to do this. So, basically instead of looking at the overlapping, it uses data from all the genes and basically you can rank all the genes from the most down regulated to the most up regulated. And, then for gene set of interest you can look at the location in the rank list. If there is no association between the rank list and the gene set and you would expect all the genes to be randomly distributed or evenly distributed across the rank list.

But, in this case for example, you see kind of the over representation of these genes at the top of rank list that is why you may think well something the might be an association between this gene set and this ranking and to do that we use a GSEA statistic or it is a statistic test derived from the Kolmogorov Smirnov Test I think Karsten has already talk about this today.

So, I would not go into the detail, but anyway from this and then you can generate the random simulation and then you get some random distribution and then you can get a p value. Again, if you do this for multiple genes sets like all the co-categories or the pathway databases, you need to do the multiple test adjustment.

(Refer Slide Time: 09:09)



Pathway-based analysis: limitations

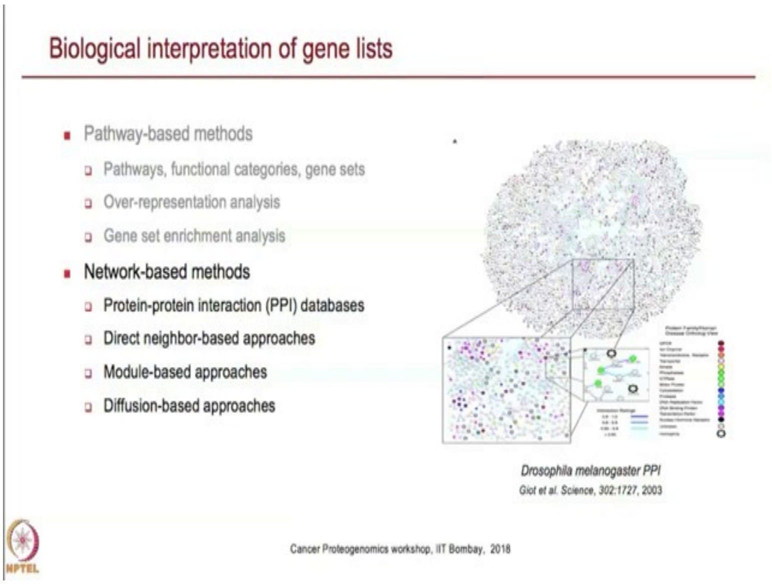
- Existing knowledge on pathways or gene functions is far from complete
- Ignoring the crosstalk between pathways

Cancer Proteogenomics workshop, IIT Bombay, 2018

This is better in a way than the over representation analysis, but still there are certain limitations like it relies on industry knowledge on the pathways and databases, but of course, I mean only knowledge and the pathway and the gene ontology I mean those annotations as the unlimited for a lot of genes we do not really know a lot about zero functions. And, also a treat one pathway as a separate identity it grows across talk between the pathways.

But, we know and although you can consider that has a relatively independent the unit this they talk to each other. So, as I talked yesterday every protein is actually linked in the cell system.

(Refer Slide Time: 10:05)



Biological interpretation of gene lists

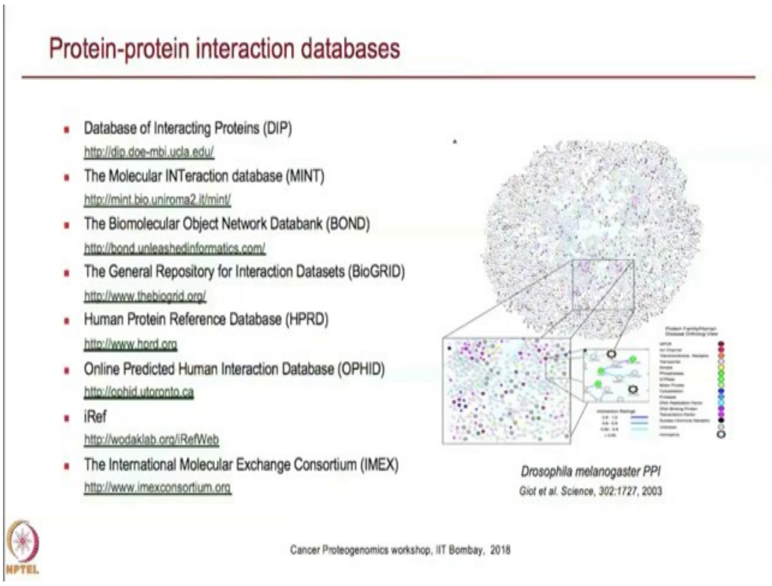
- Pathway-based methods
  - Pathways, functional categories, gene sets
  - Over-representation analysis
  - Gene set enrichment analysis
- Network-based methods
  - Protein-protein interaction (PPI) databases
  - Direct neighbor-based approaches
  - Module-based approaches
  - Diffusion-based approaches

Drosophila melanogaster PPI
Giot et al. Science, 302:1727, 2003

Cancer Proteogenomics workshop, IIT Bombay, 2018

So, that is why I the certain type of approach is to map your data to the network and then do some network based analysis like the we talked about a few different methods yesterday.

(Refer Slide Time: 10:19)



Protein-protein interaction databases

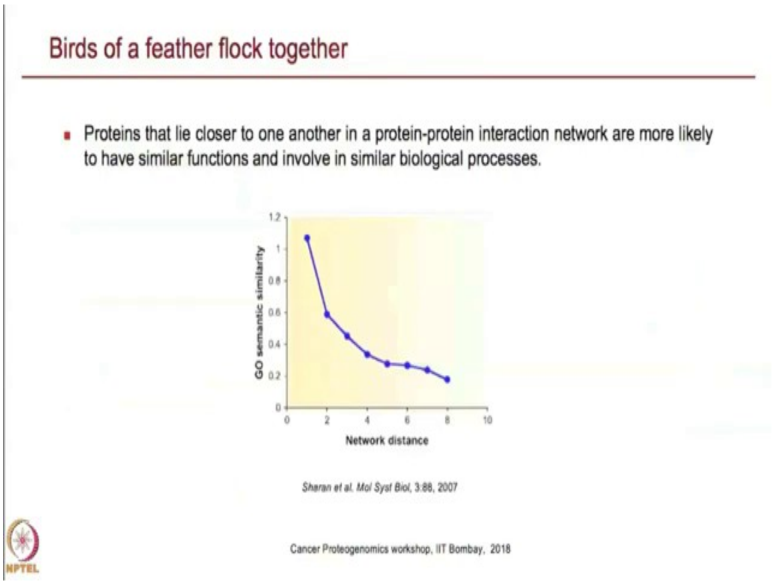
- Database of Interacting Proteins (DIP)
  http://dip.doe-mbi.ucla.edu/
- The Molecular INTeraction database (MINT)
  http://mint.bio.uniroma2.it/mint/
- The Biomolecular Object Network Databank (BOND)
  http://bond.unleashedinformatics.com/
- The General Repository for Interaction Datasets (BioGRID)
  http://www.thebiogrid.org/
- Human Protein Reference Database (HPRD)
  http://www.hprd.org
- Online Predicted Human Interaction Database (OPHID)
  http://ophid.utoronto.ca
- iRef
  http://wodaklab.org/iRefWeb
- The International Molecular Exchange Consortium (IMEX)
  http://www.imexconsortium.org

Drosophila melanogaster PPI
Giot et al. Science, 302:1727, 2003

Cancer Proteogenomics workshop, IIT Bombay, 2018

And, we also mentioned about this pathway interaction databases.

(Refer Slide Time: 10:22)



Birds of a feather flock together

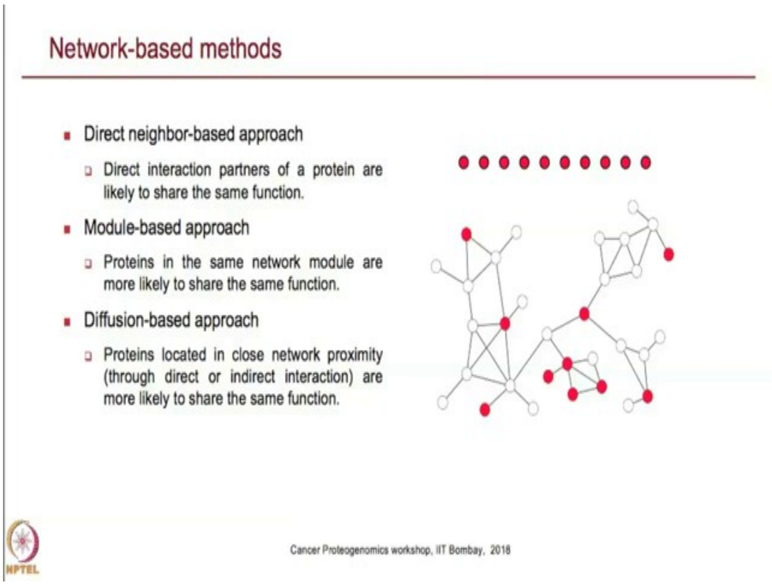
- Proteins that lie closer to one another in a protein-protein interaction network are more likely to have similar functions and involve in similar biological processes.

Sharan et al. Mol Syst Biol, 3:88, 2007

Cancer Proteogenomics workshop, IIT Bombay, 2018

And, the basis for using that this type of analysis is proteins lie close to each other in the network tend to have similar functions.

(Refer Slide Time: 10:36)



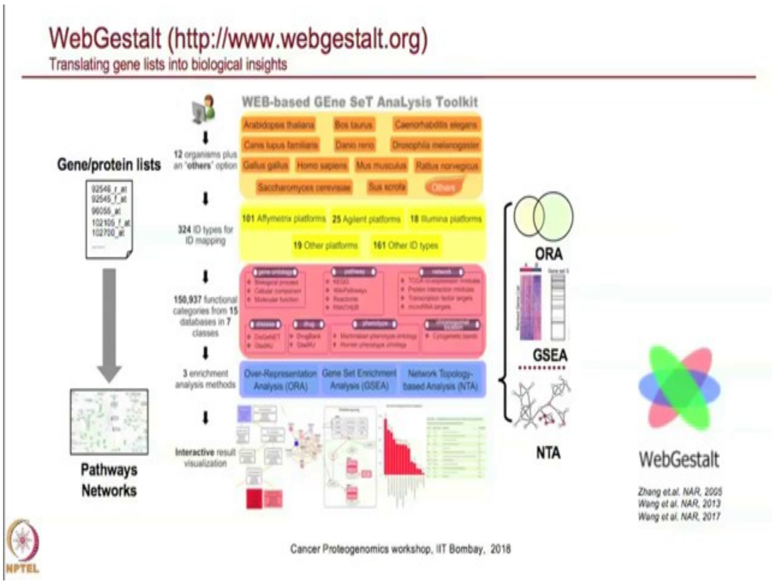
Network-based methods

- Direct neighbor-based approach
  - Direct interaction partners of a protein are likely to share the same function.
- Module-based approach
  - Proteins in the same network module are more likely to share the same function.
- Diffusion-based approach
  - Proteins located in close network proximity (through direct or indirect interaction) are more likely to share the same function.

Cancer Proteogenomics workshop, IIT Bombay, 2018

And so, basically we talked about a few different methods and the so, in the WebGestalt thought we are basically implemented to approach is the module based approach. Basically for each network we pre compute the modules and then we treat each module as a gene set and then you can do the enrichment analysis against those predefined modules or you can use a diffusion based approach and that will allow you to do gene prioritization.

(Refer Slide Time: 11:13)



WebGestalt (http://www.webgestalt.org)

Translating gene lists into biological insights

Cancer Proteogenomics workshop, IIT Bombay, 2018

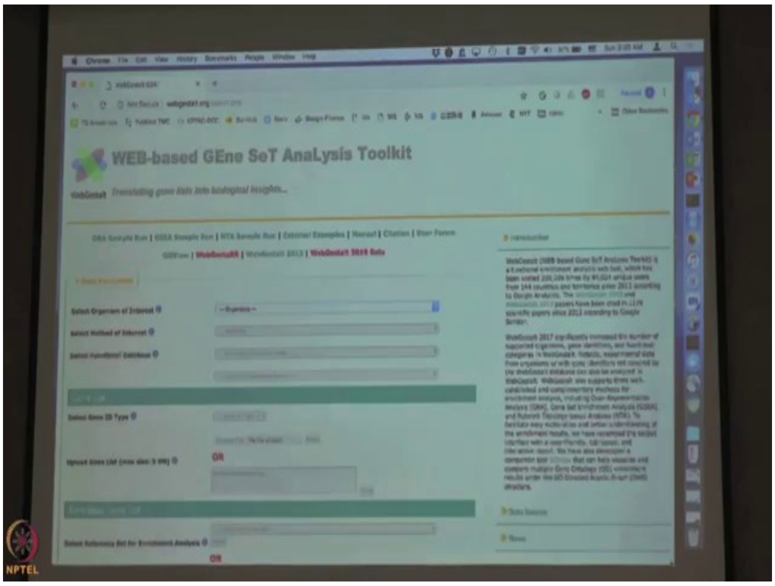
So, this is a kind of a overview of the WebGestalt. So, the goal is to translate on lists of genes of interest into some biological and pathway level understanding and the system can support tariff organisms. So, not only human or mouse, but other model organisms I can also be supportive. I talked to some of the students by in the audience and I know some of you do not actually work in human. Although all the examples today I am going to give being human, but if you are working *C. elegance* for example, you will be able to use this resource as well.

And, we support a lot of gene different types of gene IDs for example, whether you use uniprot or whether you use ensemble to do the proteomics database search you will be able to use this without worrying about the system do not recognize your ID. And, then there are a total of more than 100,000 gene sets different types of gene sets that can be used through your analysis I am coming from the gene ontology from different data pathway databases and they said from different types of protein-protein interaction networks or from the gene cancer derived the gene co expression networks and also neck phenotype associated gene sets drug related gene set etcetera.

And, we support all three types of analysis next over representation analysis GSEA and the network topology analysis which is I mean; that means, diffusion based analysis and then the output is very interactive and this was based on a figure part are in the 2017 paper, but today I am going to what I am going to demo is 2019 it is not 2019 yet, but they are preparing for that papers. So, it is actually this is my first demo of that new system and you guys will be the first group of people to know this new system. It is still the better version, but I think it is quite stable and you guys should be able to use it now.

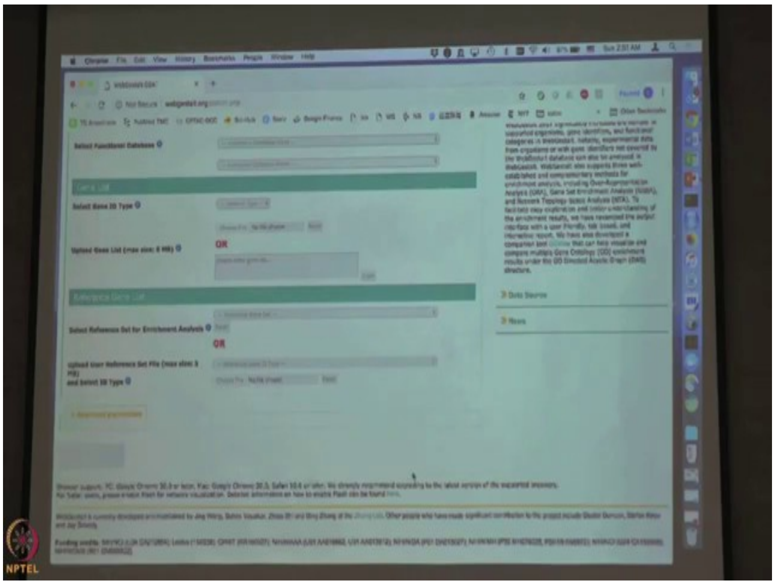
So, I put some hands on next step by step guide about the examples we are going to go through today. So, you can go through the workshop area and have everyone got that pdf file or word file about the steps to do the analysis.

(Refer Slide Time: 14:02)


WEB-based GEne SeT AnaLysis Toolkit
WebGestalt
Translating gene lists into biological insights...
Select Organism of Interest
Select Method of Interest
Select Functional Database
Select Gene ID Type
Upload Gene List (max size: 5 MB)
OR
Select Reference Set for Enrichment Analysis
OR
NPTEL

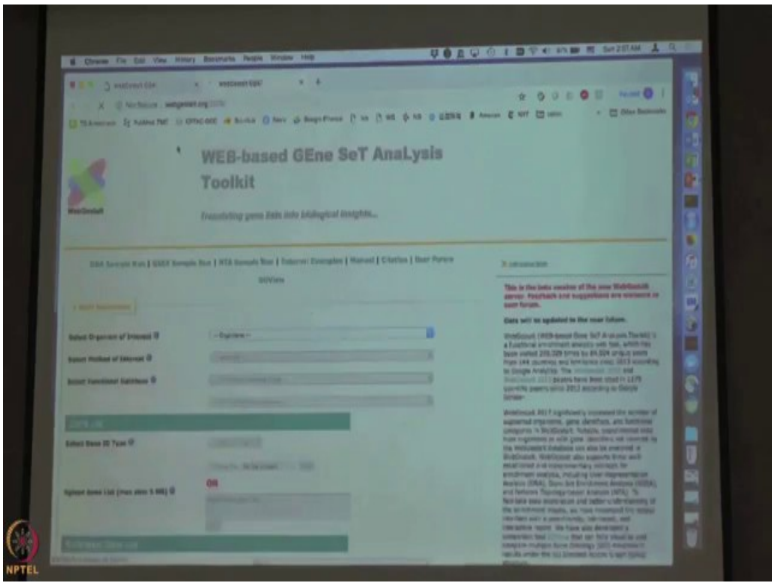
So, let us get started it is a web application. So, you do not need to install any anything or download anything you just go use your web browser to go to the right url and then you should be able to get started. I would recommend you to use Chrome as a in browser because most of our developers are I do not know what is causing the most of our developers are using the Chrome as primary web browser for the development. So, it is the best tested in that browser.

(Refer Slide Time: 14:40)


Select Functional Database
Gene List
Select Gene ID Type
Upload Gene List (max size: 5 MB)
OR
Reference Gene List
Select Reference Set for Enrichment Analysis
OR
Upload User Reference Set File (max size: 5 MB)
and Select ID Type
NPTEL

But, if you use Safari or I think there is guidance at the bottom of this if you use Google Chrome, Safari or the latest version of the IE I think you should be fine, but Chrome is free. So, I think you anyone can download and use Chrome if possible. So, let us go to the WebGestalt website and you can either just typing url you have the 2019 one or you just Google WebGestalt and then go to the 2019 beta version.

(Refer Slide Time: 15:13)


WEB-based GEne SeT AnaLysis Toolkit

Translating gene lists into biological insights...

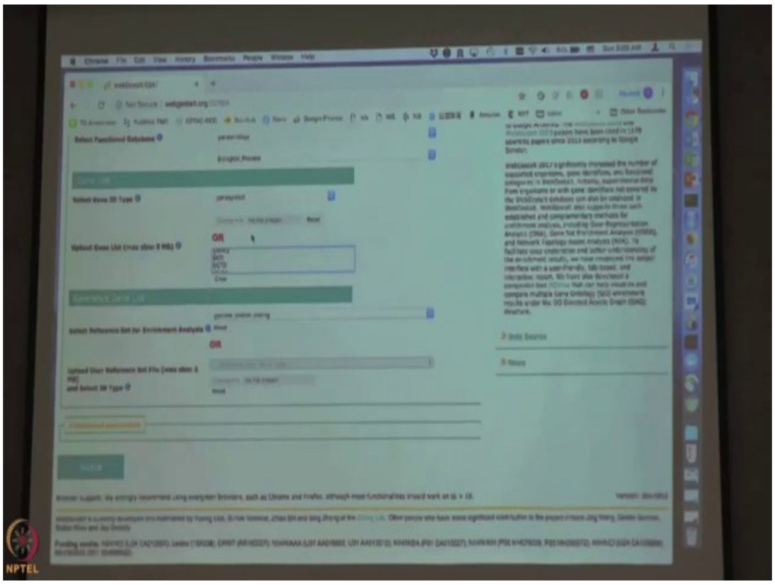
And, you should see this interface and the in the first example I will show you and how we can use WebGestalt to understand our needs of genes that are associated with colon cancer.

So, it is already pre-configured as ORA sample run. So, let us say you click on this ORA sample run. It will automatically filling all of the parameters and also upload the gene list.

So, this is a list of genes related to colon cancer it is we got this list from another tool developed in the lab called Glad For You, gene list is the automatically derived for you. So, that one so, basically you can type in anything you are interested in and then it will give you the list of genes that are related to the concept you are interested in.

It could be a disease, it could be a biological process or could be something you are interested in, but anyway this time we get 487 genes or that are related to colon cancer from that tool. So, we can go through the parameter setting to just try to understand what this means.

(Refer Slide Time: 16:36)



So, the first one is either it is select organism of interest because these are the genes from the human right. So, use select *Homo sapiens*, but if you have other data from other organisms in your future research you can select the right organism in the future studies. And, then select method of interest and work into the over representation analysis because we do not have a ranked news. This is the type of situation that you have a list of genes without any statistical metric or anything that you can use to rank them.

For example, you do clustering and then you get a list of genes just in that cluster and then there is not really a rank right then you cannot use GSEA and so, we choose ORA and we were just use gene ontology biological process for this basic analysis. But, here you can see in addition to gene ontology the function of the data base is related to pathways and network, modules, disease related the gene set, drug related gene sets, phenol type gene sets and genes in different chromosome locations or some community contributed the gene sets.

For example, some of you are interested in certain types of studies making infectious disease or other diseases and you may sometimes have a list of genes that you are interested in and you want to contribute to the community and other people to test their data against your gene set right. This is why we created this community contributed. These are from some labs they bound to share their gene set to let others to compare their new and studies against their gene sets.

So, let us see gene ontology and the biological process for today and then the gene ID type here what we have is gene symbol, but as you can see we support a lot of different gene IDs from all the different types of microarray IDs through the illumina array IDs and then for proteins we have the ensemble reps uniprot all these different types of IDs, but this one particular one we are using is gene symbol.

And, reference gene list that in this case we use a gene all the protein genes in the genome because the this was basically and it needs to come from the literature search and we do not really know what is space they used to we just consider every gene has been probably have some opportunities to be studied. But, in your own research in the future it is very important to consider what to use as a reference set for enrichment analysis.

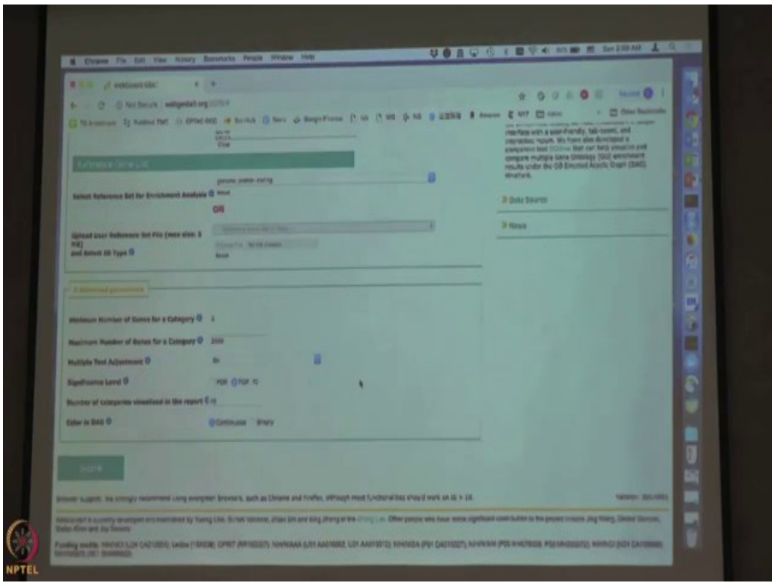
For example, if you do ionseq it might be safe to use all the genes because in ionseq is kind of umpires that you get access to you investigate all of the genes in the genome, but let us say if you use proteomics specifically I mean you can only get a part of the proteomes a lot of the proteins you did not identify. So, when you see you identify 100 differentially expressed proteins that is out of them maybe 10,000 proteins you quantified or even sometimes 8000 or 3000 right.

So, a lot of proteins you did not have the opportunity to get the statistics and then you should limit your search space or the background reference space to those proteins you can actually analyze. This is very very important I review a lot of papers and sometimes people just use the for example, in a lot of proteomics studies. They use computers like all every other genes in the proteome as a reference that is not correct because that will would not like me you can easily identify.

For example, ribosome or those type of abundant proteins as enriched, but it is just because those proteins has better chance to be detected not because they are differentially expressed just to be careful in selecting your reference gene list. Upload the gene list, here this one was from I because we clicked on the sample run it is automatically feared from the I mean as I said from the 417 genes from another tool, but in the future let us see if you have a 100 gene symbols you just copy and paste are here or you can go here and you can save this in a text the file and then choose the file and then you can upload the file anywhere.

There are two options either you can just copy and paste or you can save that as a file and upload as the text the file. But, make sure when you upload that the file your file extension has to be .txt and then you do not want to have any special characters in your file name.

(Refer Slide Time: 22:04)



Select Reference Set for Enrichment Analysis

OR

Upload User Reference Set File (max size: 5 MB) and Select ID Type

Minimum Number of Genes for a Category 5

Maximum Number of Genes for a Category 2000

Multiple Test Adjustment BH

Significance Level ○ FDR ● TOP 10

Number of categories visualized in the report 10

Color in DAG ● Continuous ○ Binary

And, the for the advance the parameter settings the minimum number of genes in the category we do not want to look at the categories that have only 1 or 2 genes because that is not very interesting right. But, we also do not want to look at the gene ontology terms based maybe more than 2000 genes because those are too broad to be meaningful and then for the multiple test adjustment we choose Benjamin Hochberg correction.

So, this is one of the most popular or I recommend just use this method for multiple tests adjustment it is nest conservatives and the Bonferroni for example. And, then for is that a significant level there are two options here and you can use FDR cutoff for example, 0.05 or

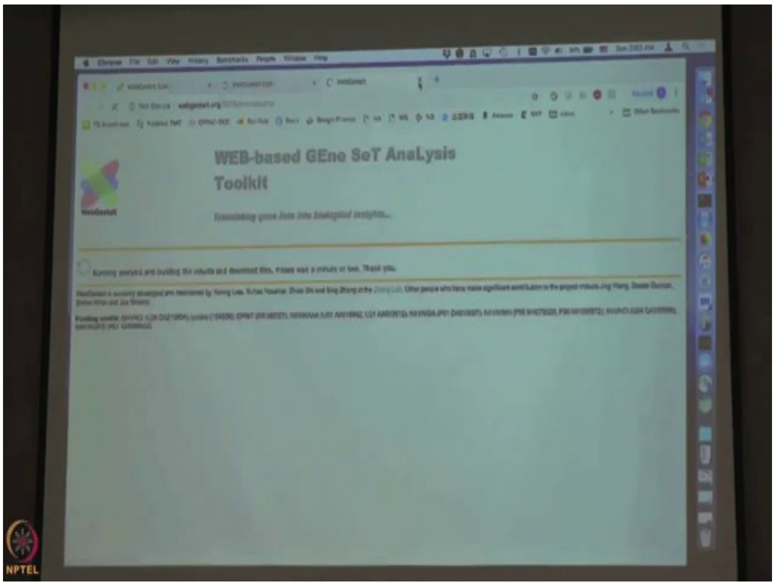
0.01 or but when you do the first round of analysis I you and I have choose a top 10 or top and you can change the number, but.

So, this way you can always get some readout back like you know what are the most enrich the term look like and then because sometimes if you pick for example, 0.01 and the number of the terms come back as significant and, then you end up with empty readout set that is not interesting right. So, and sometimes if you your difference is really huge you end up with thousands of in richer terms and then you cannot comprehend all those either.

So, to the top 10 either the first step can give you a sense or how strong the signal is your dataset and the number of category is realized in the report. So, I think we generate results for all the significant gene sets, but for example, in this case we are going to realize the results in a depth if the number of gene sets is too big then for example, if you have one thousand significant sets and then the realization will be too crowded and you would not be able to see any sense.

So, that is why we need to set a cut off here and then you want to color the terms in using continuous data meaning, using the significance level in the deck. You can also choose binary, but then you know the continuous is a better option.

(Refer Slide Time: 24:34)



WEB-based GEne SeT AnaLysis Toolkit

Translating gene lists into biological insights...

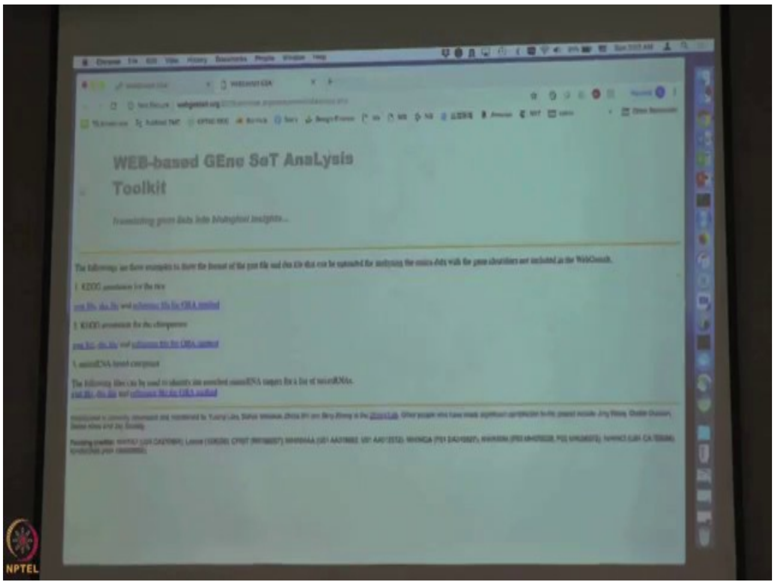
Running analysis and building the results and download files. Please wait a minute or two. Thank you.

[illegible]

Funding credits: [illegible]

Well, you can download the I will I actually downloaded the result on my I can use that as well.

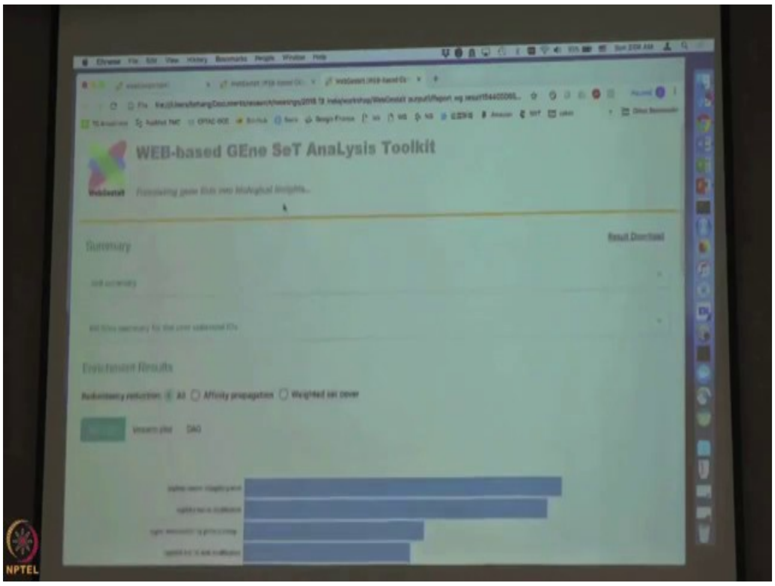
(Refer Slide Time: 24:41)



WEB-based GEne SeT AnaLysis Toolkit

But, let me try yesterday I found sometimes I submit the first time that did not work and submit the second time.

(Refer Slide Time: 24:50)



WEB-based GEne SeT AnaLysis Toolkit

Result Download

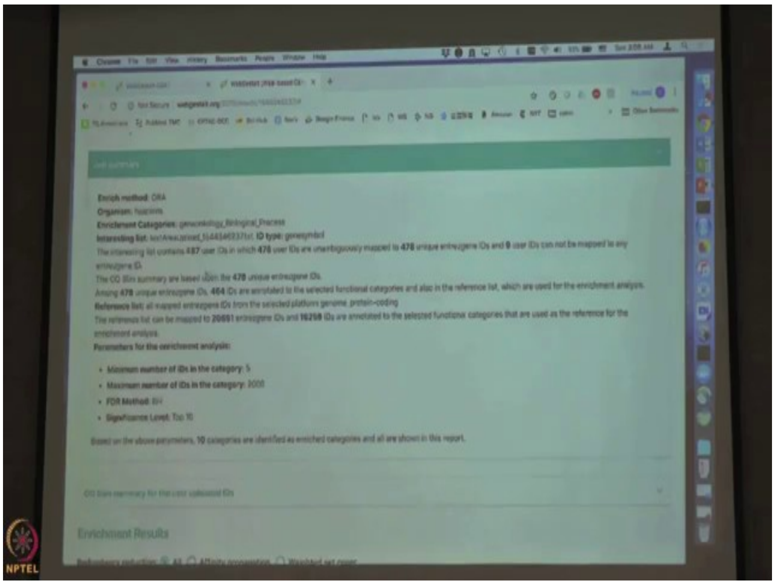
Summary

Enrichment Results

Redundancy reduction: ● All ○ Affinity propagation ○ Weighted set cover

So, I can show you the result I got here, but I also got the result. So, many let us use this. So, to both of you get your results back like this. So, in this report at the very top I mean some of you might have used the old version of the WebGestalt before, but I think this new version is much improved under user interface the results are much simpler, much easier to understand and easy to browse. So, at the very top you see the job summary.

(Refer Slide Time: 25:24)



Job summary

Enrich method: ORA
Organism: hsapiens
Enrichment Categories: geneontology_Biological_Process
Interesting list: [illegible] **ID type:** genesymbol
The interesting list contains **487** user IDs in which **478** user IDs are unambiguously mapped to **478** unique entrezgene IDs and **9** user IDs can not be mapped to any entrezgene ID.
The GO Slim summary are based upon the **478** unique entrezgene IDs.
Among **478** unique entrezgene IDs, **464** IDs are annotated to the selected functional categories and also in the reference list, which are used for the enrichment analysis.
Reference list: all mapped entrezgene IDs from the selected platform genome_protein-coding
The reference list can be mapped to **20691** entrezgene IDs and **16258** IDs are annotated to the selected functional categories that are used as the reference for the enrichment analysis.
Parameters for the enrichment analysis:

- **Minimum number of IDs in the category:** 5
- **Maximum number of IDs in the category:** 2000
- **FDR Method:** BH
- **Significance Level:** Top 10

Based on the above parameters, **10** categories are identified as enriched categories and all are shown in this report.

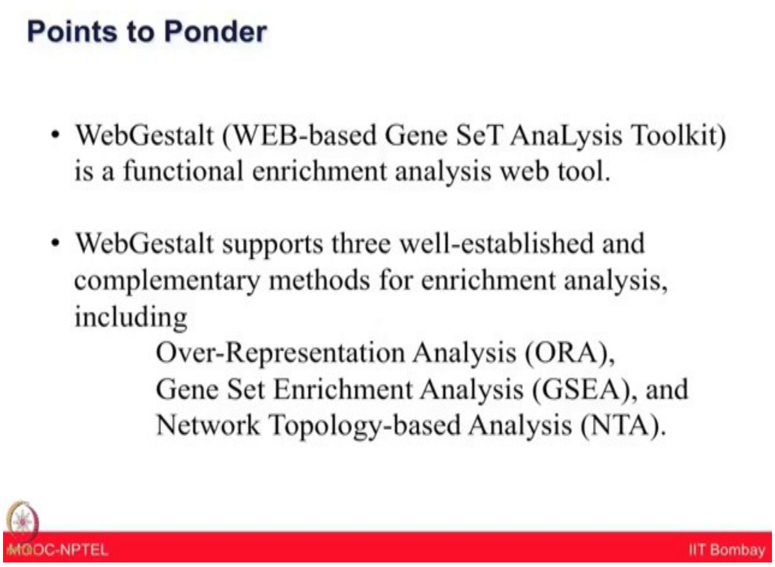
GO Slim summary for the user uploaded IDs

Enrichment Results

Redundancy reduction: ● All ○ Affinity propagation ○ Weighted set cover

If you expand this basically will remind you about all the parameters you used or I mean what you did for this analysis you can expand and the resist, but you can also close it.

(Refer Slide Time: 25:38)

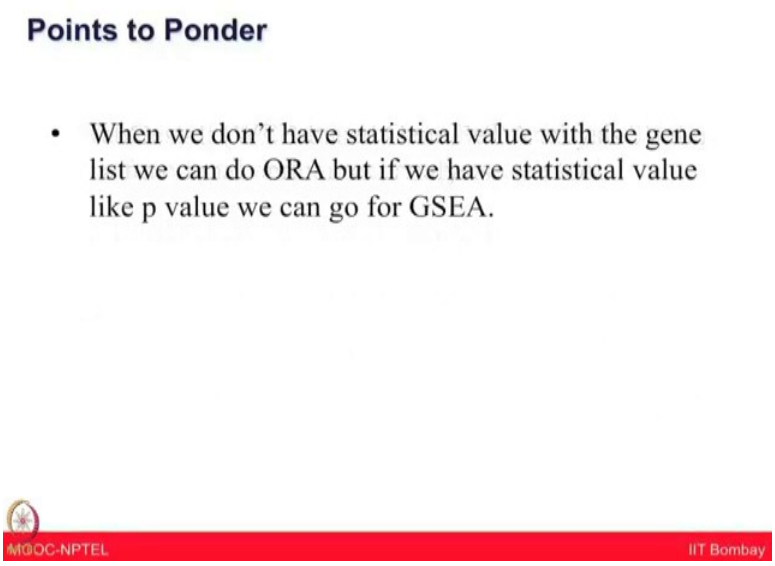


Points to Ponder

- WebGestalt (WEB-based Gene SeT AnaLysis Toolkit) is a functional enrichment analysis web tool.
- WebGestalt supports three well-established and complementary methods for enrichment analysis, including
  Over-Representation Analysis (ORA),
  Gene Set Enrichment Analysis (GSEA), and
  Network Topology-based Analysis (NTA).

MOOC-NPTEL IIT Bombay

(Refer Slide Time: 25:53)



Points to Ponder

- When we don’t have statistical value with the gene list we can do ORA but if we have statistical value like p value we can go for GSEA.

MOOC-NPTEL IIT Bombay

I hope today’s session was useful, where you got a brief idea about WebGestalt. Dr. Bing Zhang discussed about different type of input ID and how the three methods are different from each other. In conclusions when we do not have a statistical value with the gene list we can do ORA, but if we have a statistical values like p value we can go for GSEA lastly he also described briefly about the job summary, how it looks like and what are the information we can obtain from it.

In the next hands on session, we will learn more about the result visualization, protein-protein interaction modules, pathway based method and network based methods.

Thank you.