

Computational Systems Biology
Karthik Raman
Department of Biotechnology
Indian Institute of Technology – Madras

Lecture - 71
Lab: Gene Deletions

(Refer Slide Time: 00:11)

Computational Systems Biology
Lab: Gene Deletions

- ▶ Gene Deletions using COBRA Toolbox
- ▶ Escher
- ▶ Single and Double Deletions

Karthik Raman
Department of Biotechnology, Bhupat & Jyoti Mehta School of Biosciences
Initiative for Biological Systems Engineering (IBSE)
Robert Bosch Centre for Data Science and Artificial Intelligence (RBC DSAI)
INDIAN INSTITUTE OF TECHNOLOGY MADRAS

The slide features a dark blue header with the title 'Computational Systems Biology Lab: Gene Deletions'. Below the header is a bulleted list of topics: 'Gene Deletions using COBRA Toolbox', 'Escher', and 'Single and Double Deletions'. The presenter's name, 'Karthik Raman', is centered below the list, followed by his affiliation: 'Department of Biotechnology, Bhupat & Jyoti Mehta School of Biosciences, Initiative for Biological Systems Engineering (IBSE), Robert Bosch Centre for Data Science and Artificial Intelligence (RBC DSAI), INDIAN INSTITUTE OF TECHNOLOGY MADRAS'. At the bottom, there are three circular logos: the IIT Madras logo, the IBSE logo, and the RBC DSAI logo.

In today's lab session, we will focus on gene deletions, how we perform gene deletions using the COBRA toolbox. I will also show you this tool called Escher which gives you a nice visual picture of what happens when you delete genes, how the rerouting of reactions happens and so on and we will also look at how we perform single or double or synthetic lethal deletions, identify synthetic lethals using the COBRA toolbox.

(Refer Slide Time: 00:33)

Gene Deletions

- Identify single lethals in the *E. coli* core model
 - [singleGeneDeletion](#)
- Identify double lethals in the *E. coli* core model
 - Using [doubleGeneDeletion](#)
- Find out a 4-5 gene deletion that is non-lethal
- Try this for multiple models
- Also check out [deleteModelGenes](#)

- What would be the minimal non-essential reaction set in a given model?

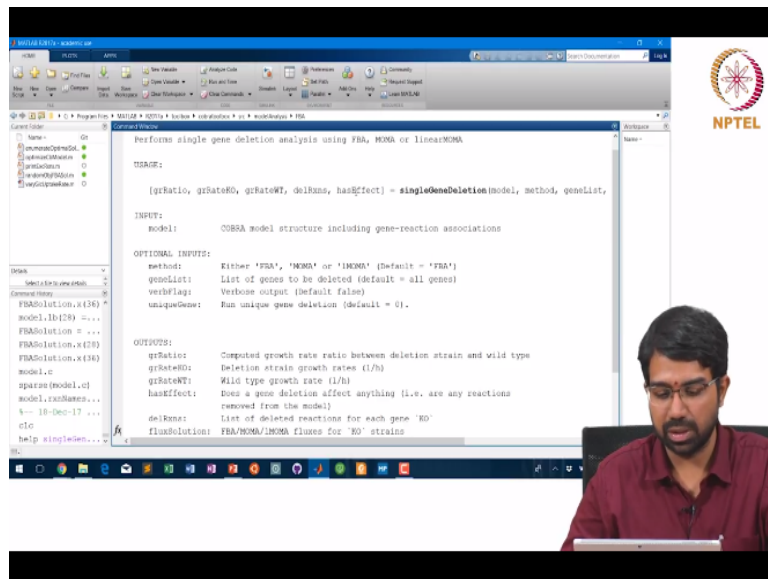
The slide has a white background with the title 'Gene Deletions' at the top left. A list of bullet points is centered on the slide. The first three items are about identifying single and double lethals in the E. coli core model, and finding a non-lethal 4-5 gene deletion. The fourth item is about trying this for multiple models. The fifth item is about checking out a specific tool. There is a gap in the list, followed by a final bullet point about the minimal non-essential reaction set. In the top right corner, there is the NPTEL logo. In the bottom right corner, there is a small video inset showing the presenter, Karthik Raman, sitting at a desk with a laptop.

So welcome back. Let us try out the following problems today in the lab. Let us identify single lethals in the E. coli core model using the single gene deletion function in COBRA toolbox. Let us identify double lethals in the E. coli core model using the double gene deletion and slightly more tricky problem is can we find out a 4-5 gene deletion so remove 4 or 5 genes from E. coli without killing it right.

And I will show you there is an interesting website that can be used to visualize these as well and try this for multiple models not just the E. coli core model but for the latest E. coli model or mycobacterium tuberculosis and so on. These can be downloaded from the BIGG database, bigg.ucsd.edu, so for deleting genes it is good to check out the function delete model genes right.

And lastly what would be the minimal non-essential reaction set in a given model, how would you compute that right? So sort of minimal genome right, so if you remove any reaction the organism is not going to survive so how would you come up with such a reaction set. So these are some of the questions I want you to tackle today.

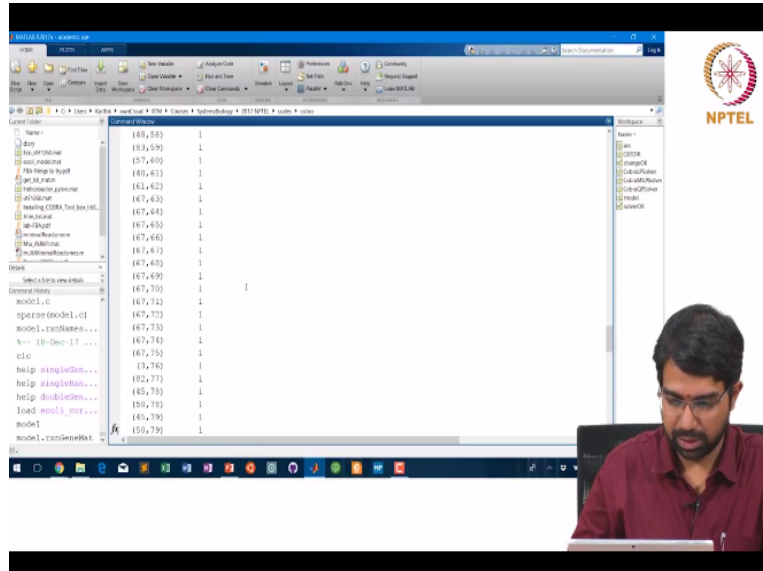
(Refer Slide Time: 01:46)



So this is a single gene deletion function, the outputs are growth ratio, growth rate on knockout so this is basically growth rate or knock out by growth rate of wild type right gr ratio and what are the reactions that are released and deleted and if there is any effect for the deletion hasEffect and you also have a help single reaction deletion which is again very similar.

You also have a double gene deletion right. This essentially gives you a sort of matrix. **“Professor - student conversation starts.”** Yes, the model has that information. **“Professor - student conversation ends.”**

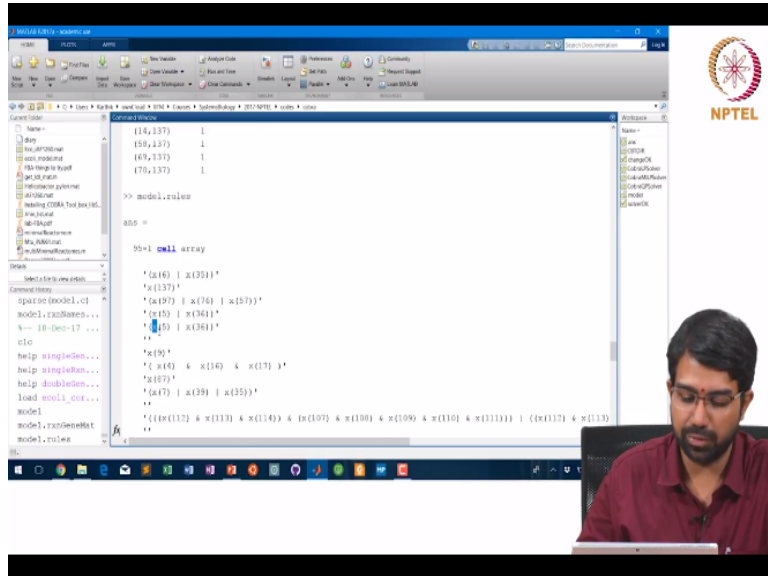
(Refer Slide Time: 02:56)



So if you look at this model.rxnGeneMat that is one place where this information is, I do not think it is there in rules. Yeah, rules also have this information. So you have several reactions that have no rules basically empty rules and this means that you know gene 31 and gene 32 or gene 20 and gene 21 are together necessary right, either these are necessary together or these are necessary together for the reaction to happen.

So the hasEffect is very useful, so when you delete 32, it will actually check if it has an effect right. At the gene level, it may have an effect when you remove a gene but may be it affects no reaction right. So if you remove x5, it is not going to have any effect assuming that it is not there individually in some other reaction elsewhere right.

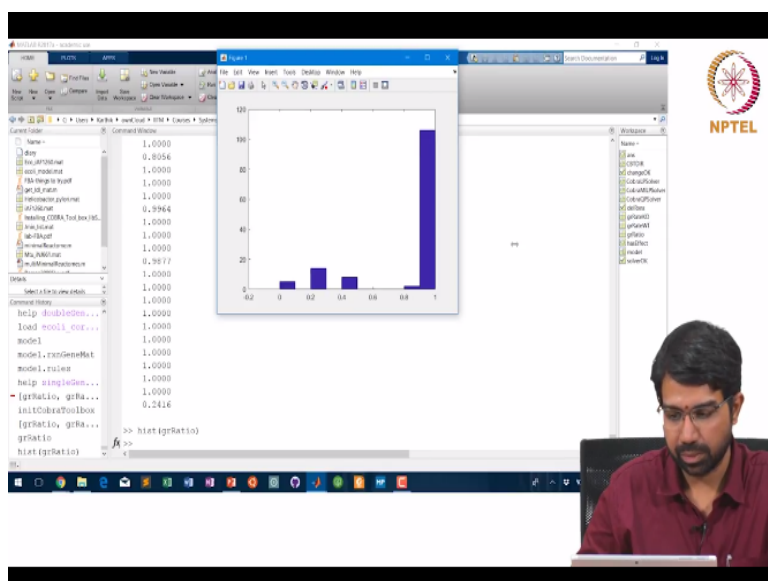
(Refer Slide Time: 03:54)



If no x5, there will be no effect because x36 will be able to take care of that reaction. Now `hasEffect` is one of the outputs, you look at `help single gene deletion`, the last output is `hasEffect`. **“Professor - student conversation starts.”** No, if the reactions are going to change, so look at the help so it says does the gene deletion affect anything that is our reactions removed from the model or not.

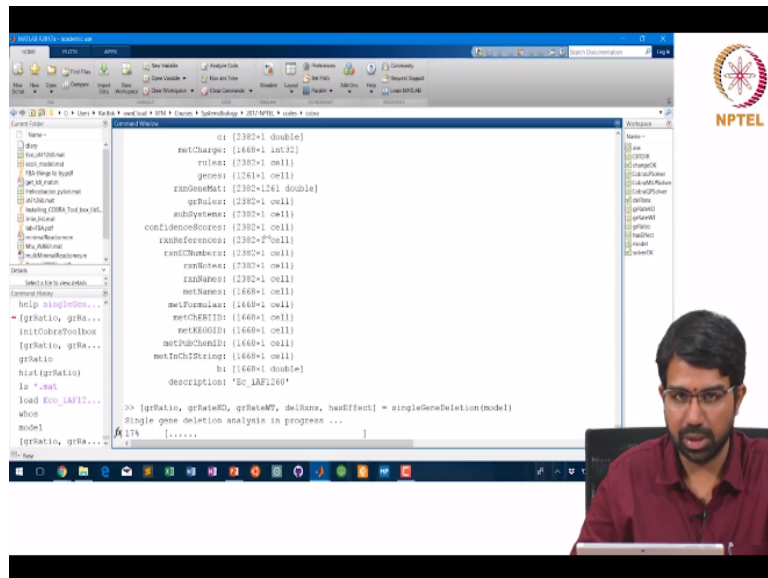
And you can use either FBA or MoMA and so on. It is, so it gives you the growth rate as well and the general assumption is something is lethal if it is $< 5\%$ of the wild type growth rate. So if you look at rate such as 99.64% of wild type or 0.8 of wild type, 0.42 and yeah for some there is no 0, for many you find it is 1.0, you find a few 0s as well.

(Refer Slide Time: 05:23)



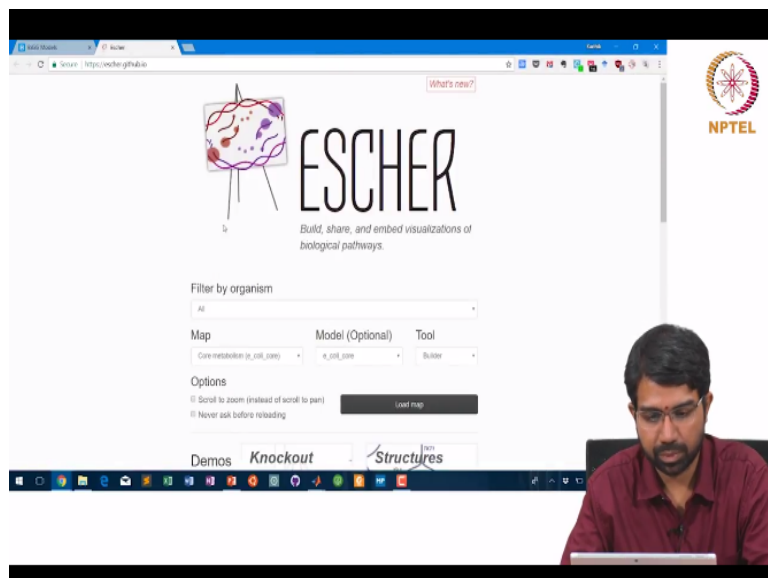
So this is the distribution, there are few lethals, few sick deletions, few very lightly sick deletions and many non-lethal deletions. So none of those are lethal so this is the E. coli model, so let us just do the same thing on the E. coli model. **“Professor - student conversation ends.”**

(Refer Slide Time: 06:03)



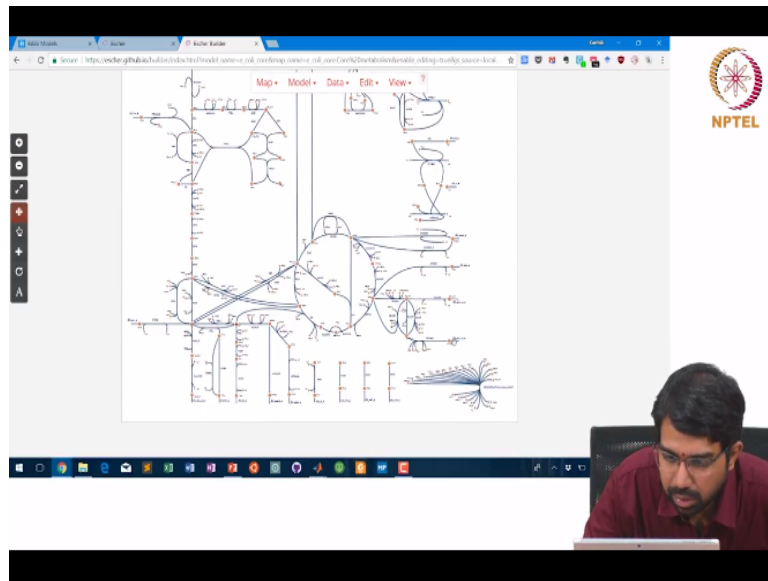
It is going to take a little longer to the single gene deletion because there are about 1260 genes in the E. coli model. We can download this model from the BIGG database. You will see that double gene deletion is like much, much lower the group force so you will have to use something like FastSL. While this is going on let me show you this nice web page.

(Refer Slide Time: 06:38)



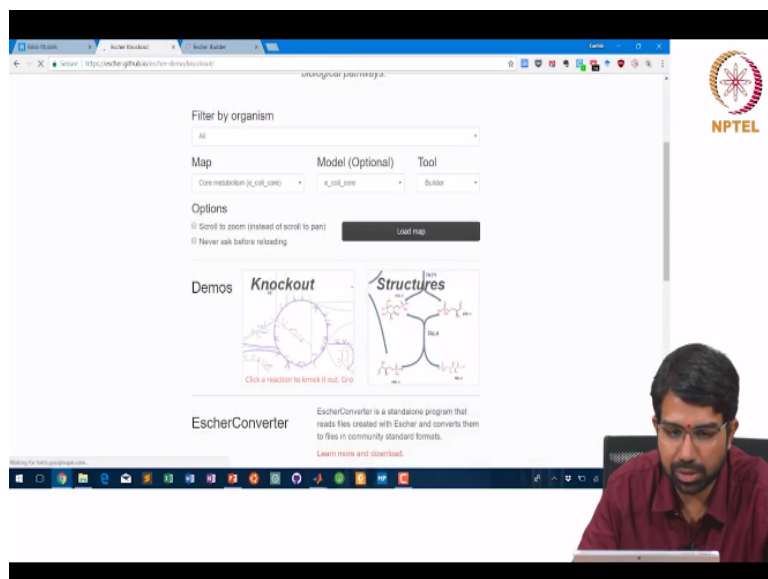
Right here it is called Escher.

(Refer Slide Time: 06:50)



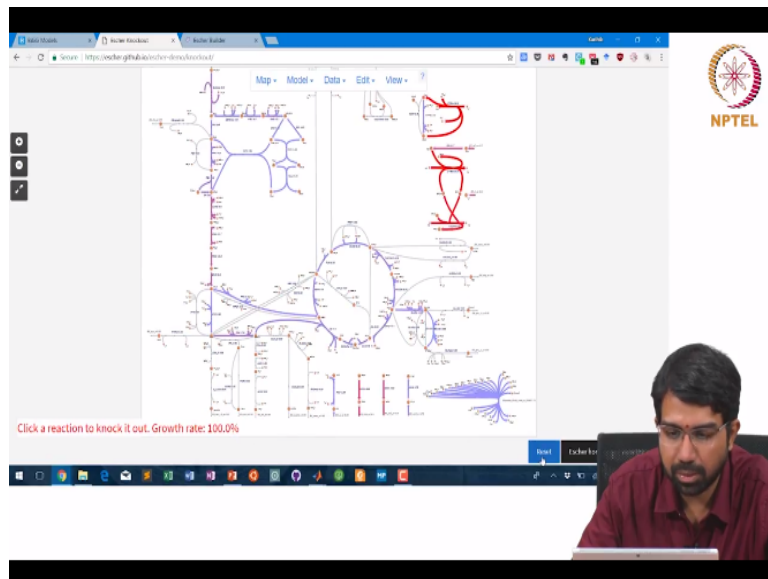
This is a visualization of E. coli and you can see what happens when you start deleting reactions. You delete a reaction here, there is a way to knock out the reactions let me check.

(Refer Slide Time: 07:41)



Then you have knock outs of reactions on Escher online, so you just a click a reaction to knock it out.

(Refer Slide Time: 07:50)

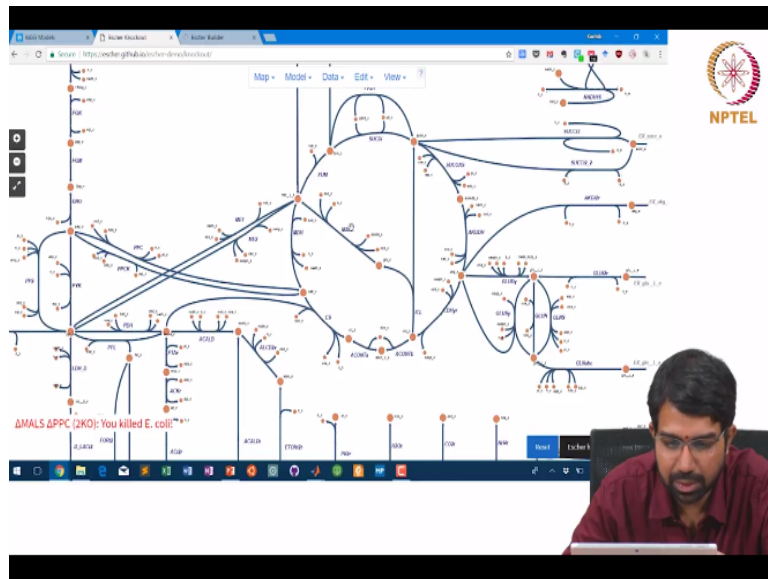


So inaugural from succinate dehydrogenase reaction or something. I have knocked out one more thing and growth rate is now 73%. Let us try to knock out something else, 3 knockouts it has gone to 28%, 4 knockouts it has gone to 26.5%, 5th knockout it killed E. coli. So this is like a quintuple synthetic lethal set right. So if you remove them separately, nothing happens. When you remove all 5 will not necessarily it could also have been that this is singly lethal.

Yeah, but you know for sure that these are all not single lethals right. **“Professor - student conversation starts.”** They must have done the similarity check there, I do not know what is the exact cut off they used. So this growth rate is 99.1, yeah so this might be a single lethal. So if you see I can keep deleting these reactions which are not going to make an effect. These are not reactions that are currently been used right but however if I even delete one of these it might already become lethal. **“Professor - student conversation ends.”**

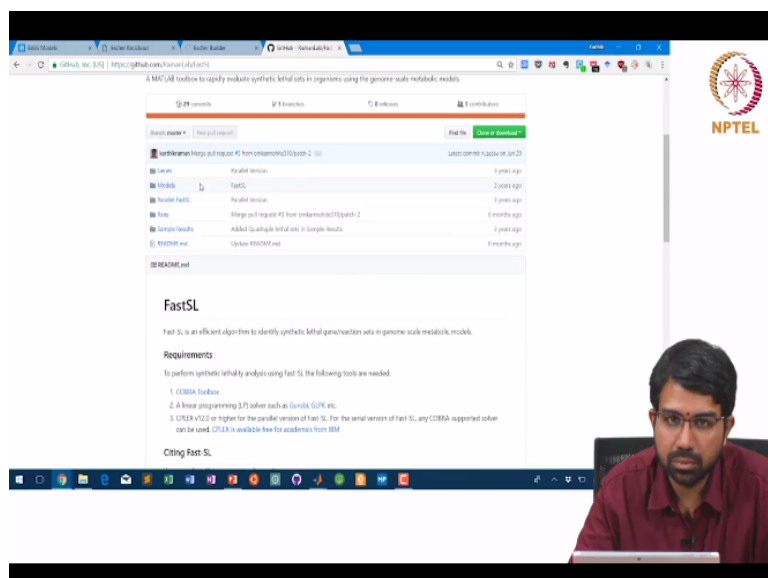
Yeah, so this was a nice example. So I kept deleting multiple genes nothing happened but then finally when I deleted one I think we have a synthetic lethal here, delta PPC and MALS looks to be a synthetic lethal.

(Refer Slide Time: 11:17)



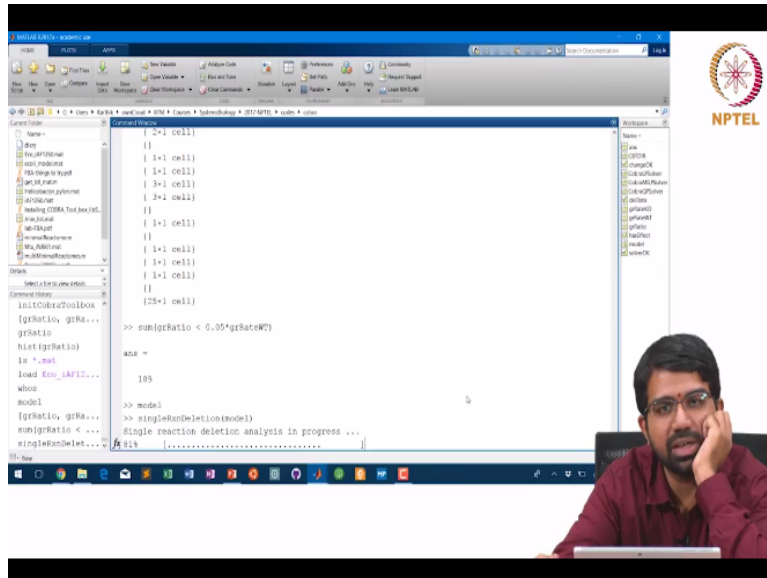
Yeah, this is a synthetic lethal, so PPC and MALS.

(Refer Slide Time: 11:55)

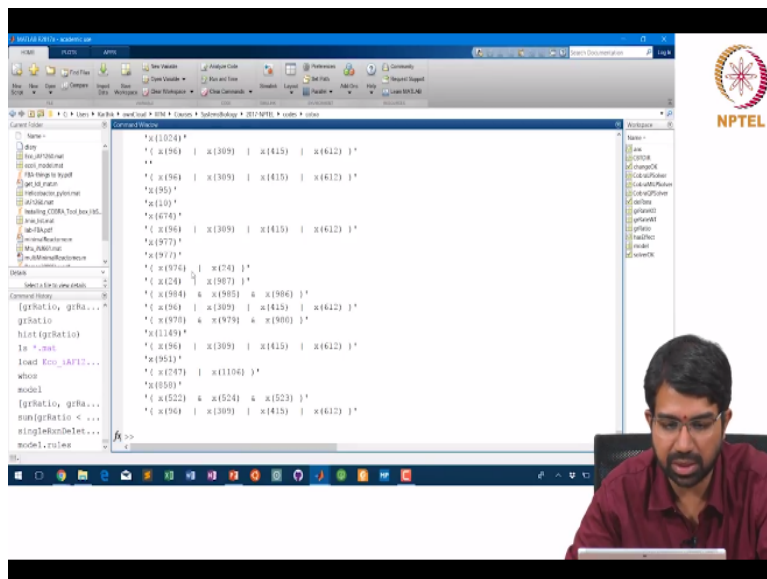


So if you download FastSL, you will also get these example matrices I mean MATLAB files which includes the stoichiometric matrix and the entire model actually. **“Professor - student conversation starts.”** That is not integrated with COBRA but the code is here, the code is (()) (12:28). So this is how you download a separate package and install it. **“Professor - student conversation ends.”** So how many single lethals in E. coli, how do you do it?

(Refer Slide Time: 13:05)



Now can you try out single reaction deletion? Is it much faster than what it is for me? **“Professor - student conversation starts.”** So reaction will remove reaction by reaction, gene will remove gene by gene and delete the corresponding reactions. Some genes if removed may not lead to the removal of any reaction. If gene 1 is not connected to any other reaction, if reaction 1 has a GPR which says like these are here right I think it is almost over. **(Refer Slide Time: 14:12)**



So model.rules yeah so this is 976 or 24, so if you remove 24 nothing will happen most likely. It would not have it okay, so this is an interesting point. So how does FBA work? It tries to maximize the growth rate through every reaction right. So it does not actually try to account if there are like multiple copies of a gene you can have a higher growth rate or any of those things right.

So if that happens see if you overexpress a gene, you can like basically you know double that value or something like that. So all this is regulatory information right which is not there in plane FBA. In future lecture, we will look at E-Flux which will essentially use this method. E-flux you know appropriately waits different changes the lb and ub for reactions depending upon the expression levels of the genes.

It helps you integrate say microarray data or any expression data set that you have with your flux model to make better predictions. So there are many methods to integrate, E-Flux is one of them. **“Professor - student conversation ends.”** I am just running this again and let us see what hasEffect looks like? So do you have an answer for the last question? What would be the minimal set of non-essential reactions in a given model?

“Professor - student conversation starts.” More or less yes you are in the right track. (()) (16:30) telling today in morning’s class. Good point, yes and no. The answer is no, why? Whether a reaction is known or not, so you might be removing a non-essential reaction. No, right so that again see your previous approach is little bit, Lavanya you have an answer, you did this already right.

That is a slow in blocked reactions but beyond that you can still remove some more reactions. That is just like that is read on it. It will just give you r1, r2, r3, r4, rn. Yeah, for reaction there is no question of doing hasEffect, so does reaction have it I was trying to remove it out of COBRA toolbox. It does have, what is the documentation say? It is just going to be all 1s. Look at the value of hasEffect, I think it is just going to be all 1s.

I did not try to edit it out of the toolbox, it is like we argued and was redundant but we may be kept it in for like backward compatibility. **“Professor - student conversation ends.”**

(Refer Slide Time: 18:45)

```
1 function [grRatio, grRateID, grRateAT, hasEffect, delRxn, FluxSolution] = singleRxnDeletion(model, method, rxnList, verbFlag)
2 % Performs single reaction deletion analysis using FBA, MOFA or LinearMOFA
3 %
4 % USAGE:
5 %
6 % [grRatio, grRateID, grRateAT, hasEffect, delRxn, FluxSolution] = singleRxnDeletion(model, method, rxnList, verbFlag)
7 %
8 % INPUT:
9 % model: COBRA model structure including reaction names
10 %
11 % OPTIONAL INPUTS:
12 % method: Either 'FBA', 'MOFA', or 'LINEAR' (Default = 'FBA')
13 % rxnList: List of reactions to be deleted (Default = all reactions)
14 % verbFlag: Verbose output (Default = false)
15 %
16 % OUTPUT:
```

So this is the code for single reaction deletion right and this is based on FastSL that we discussed a little earlier.

(Refer Slide Time: 19:02)

```
44 solWT = optimizeCvModel(model, 'max', 'one'); % by default uses the min manhattan distance norm FBA solution.
45 grRateAT = solWT.f;
46
47 % Identify reactions that do not carry a flux in solWT, none of these can be lethal
48 Jm = solWT.x==0; % reactions that carry a flux in the minimum norm solution
49 Jz = solWT.x==0; % reactions that do not carry a flux in the minimum norm solution
50
51 grRateID = ones(nDelRxn, 1)'*grRateAT;
52 hasEffect = true(nDelRxn, 1);
53 FluxSolution = repeat(solWT.x, 1, nDelRxn);
54 delRxn = columnVector(rxnList);
55 if (verbFlag)
56     fprintf('%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%', 'No', 'Per', 'Name', 'Growth rate', 'Del. GR');
57 end
58 showProgress(0, 'Single reaction deletion analysis in progress ...');
59 for i = 1:nDelRxn
60     showProgress(i/nDelRxn);
61     if ismember(rxnList(i), model.rxnList)
62         % If the reaction carries no flux in WT, deleting it cannot affect
63         % the flux solution. Assign WT solution without solving IP.
64         solID = solWT;
65         hasEffect(i) = false;
66     else
67         modelIDel = changeRxnBounds(model, rxnList(i), 0, 'U');
68         % [delRxn, method] = singleRxnDeletion(modelIDel, method, rxnList(i), verbFlag);
69     end
70 end
```

So here the key ideas here so if the reaction carries no flux in wild type, deleting it cannot affect the flux solution. See not all reactions will carry a flux in the wild type solution right. So corresponding to Biomass optimization, for any objective function you run FBA and you get a flux solution right a flux distribution. **“Professor - student conversation starts.”** Exactly, right so there are many equivalent flux distributions one this that is guaranteed for you is that the function value is the same right.

So out of these equivalent flux distributions which of them you know so can you use that to find out which reactions are essential and not essential? So there are some reactions which

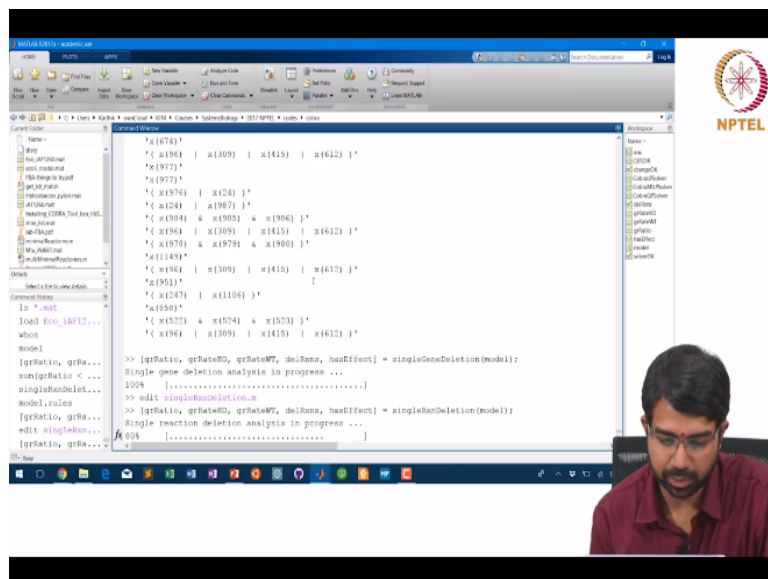
always stay 0 right like your blocked reaction, so you can be done a flux variability analysis and if both the lower and upper bounds are I mean the maximum and minimum values are 0, you can discard those reactions but beyond that even out of a pair of reactions like we were seeing here, let me just reset this.

You see that only MALS is not currently carrying a flux right but it could very well be carrying a flux. If I delete this right the growth rate has not changed for practical purposes right and this is carrying a flux instead right. So now if I go and touch this it gets you know the knocked out and you have a synthetic lethal right but basically which of these do you pick initially? It is almost arbitrary it depends upon the solver, both these are equivalent states in which the cell can exist in reality right.

So to come back let us look at this part of the code now. So if the reaction carries no flux in the wild type right in a FBA solution in any given FBA solution you can remove it without affecting the flux solution and you can basically assign the wild type solution to the knock out solution without even computing or solving the LP and that is where we say that hasEffect is false.

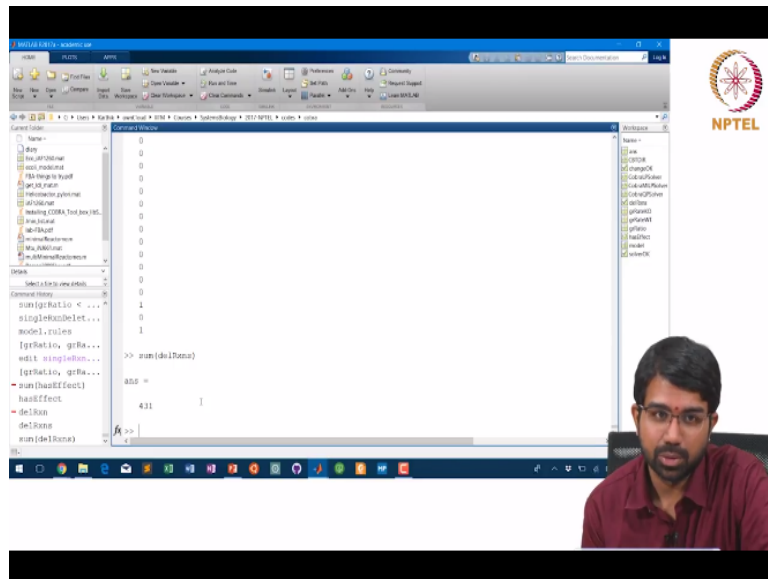
There is no effect actually on deleting that reaction. So if you run this, you will now see this is much faster. **“Professor - student conversation ends.”**

(Refer Slide Time: 22:04)



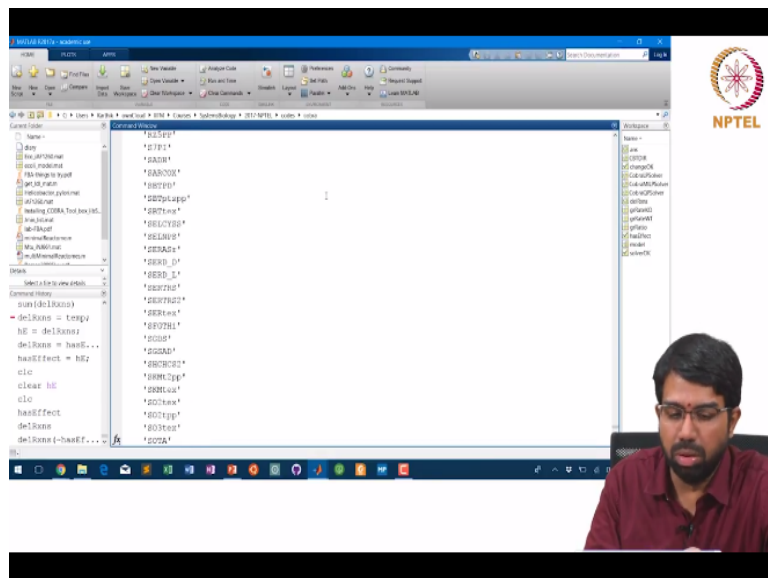
Yeah, you can see this is way faster than the deletion that I was running previously, single reaction deletion right. So now oops I think it has a reaction delRxn are swapped.

(Refer Slide Time: 23:10)



I thought I fix this anyway, so there are 431 reactions that cannot be knocked out so if I start with a subset of this reaction set so let us just **“Professor - student conversation starts.”** No, no, no this was like variable mismatch. This is actually hasEffect. The code has an error, oh god I have to put a pull request for this. No, not all that is exactly precisely what I was trying to tell you all along right.

(Refer Slide Time: 24:45)



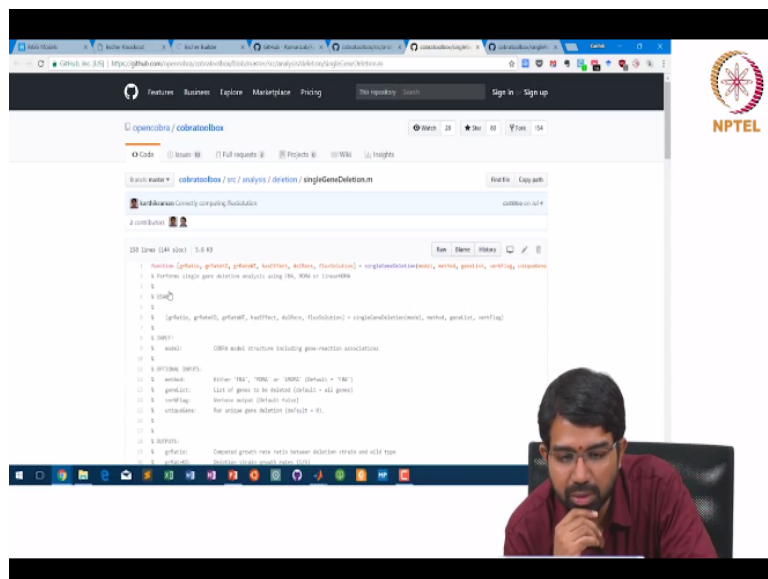
So some reactions okay so let us just see okay wait. So all these reactions do not have an effect, when you delete them nothing happens. So that reaction is not carrying a flux, it is not currently important for the cell. So there are two, three possibilities. When you delete a reaction and nothing happens, there are three possibilities. What are the three possibilities? We discussed them before at least two of them we clearly discussed them before.

Yeah but that is like just saying you know $A=B$, $B=A$ right. So I want you to tell what is the reason that that happens, it could be blocked reactions, it is not like a major name. It could be a conditionally blocked reaction or it could be a gap or an unconditionally blocked reaction. So these are two possibilities. The third possibility is that it is not lethal but that is also a possibility, it could be not lethal I do not know if that is some usually exclusive cases not lethal.

The fourth possibility is that it could be part of a higher lethal set. When you remove a reaction out of a triple lethal set, nothing is going to happen. If you remove two reactions out of a triple lethal set nothing is going to happen. Only when you remove all three, the organism is going to die. Who is referring to? HasEffect basically says removing your reaction changes has an effect on the growth rate or not.

No, not lethally well whether it even changes growth rate or whether it even changes the reaction flux, whether it is even important, whether it is unblocked so it is actually a slightly debatable thing. HasEffect for gene is the much easier thing to understand. The code is not wrong. I had given the arguments in the wrong order here. **“Professor - student conversation ends.”** The gene deletion and reaction deletion orders are different, let us just see.

(Refer Slide Time: 27:50)

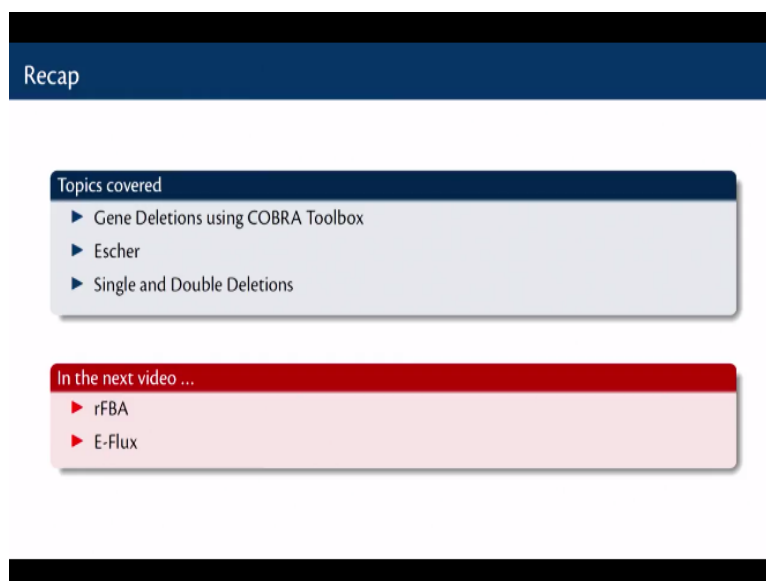


This is growth ratio, growth rate KO, growth rate wild type, hasEffect, del reactions. This is growth ratio, growth rate KO, growth rate wild type, hasEffect, del reactions. It is similar

maybe okay maybe the version I had was slightly different. This is del reactions hasEffect that is where the problem is okay. So where are we, how much of this is done? Let me try all of this, this is easy, seriously easy stuff.

You have to learn out your delete model genes. What will be the naïve way to do this? Find the minimal non-essential reactions. The naïve way to do is to remove a reaction. If it does not kill the organism, keep iterate. Then, remove another reaction, if it kills the organism, revert that deletion repeat.

(Refer Slide Time: 29:36)



So in today's lab, we had an overview of how we perform gene deletions using the COBRA toolbox. I also showed you this interesting tool called Escher which gives you a visual picture of what happens when you delete genes and we can also see when the organism stops growing and so on and we also looked at single and double gene deletions.

In the next video, we will see how to integrate regulatory information into flux balance models and get better predictions. We will look at two techniques, one called regulated flux balance analysis or rFBA and the other called E-Flux which gives a nice way to constraint the reactions based on transcription data or transcriptomic data.