AI-Powered Genomics

Unveiling Hidden Controls of Plant Gene Expression

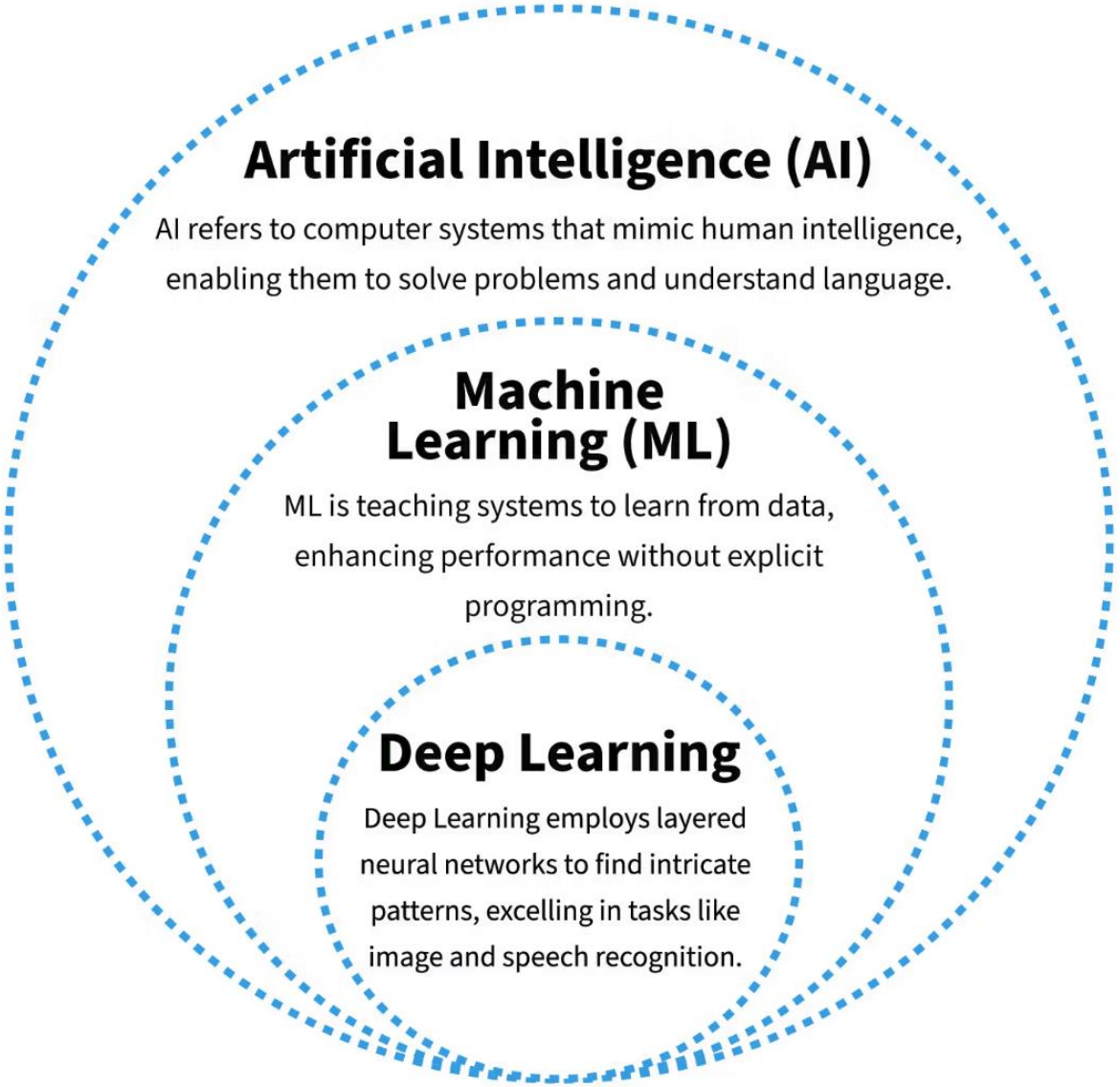
Jędrzej Jakub Szymański

@NAMlab www.szymanskilab.com

IPK LEIBNIZ INSTITUTE

CEPLAS Cluster of Excellence on Plant Sciences

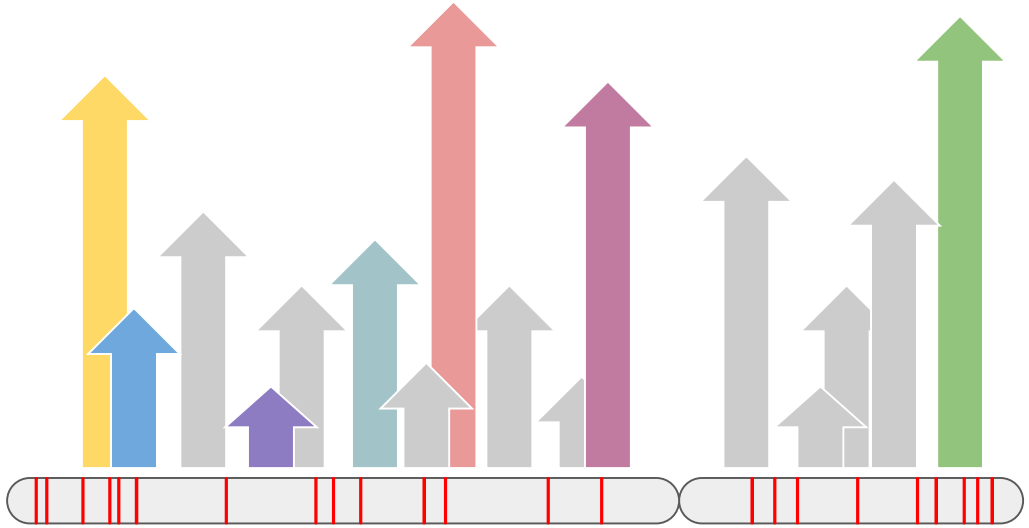
JÜLICH Forschungszentrum


Artificial Intelligence (AI)
AI refers to computer systems that mimic human intelligence, enabling them to solve problems and understand language.
Machine Learning (ML)
ML is teaching systems to learn from data, enhancing performance without explicit programming.
Deep Learning
Deep Learning employs layered neural networks to find intricate patterns, excelling in tasks like image and speech recognition.

Genotype to phenotype

Phenotype:

Genotype:



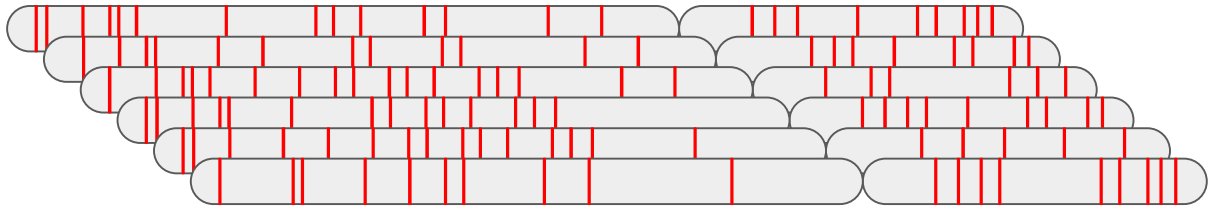
Genotype to phenotype

Phenotypes:

Photo: Uli Westphal
(IPK cultivar collections)
www.uliwestphal.de

?

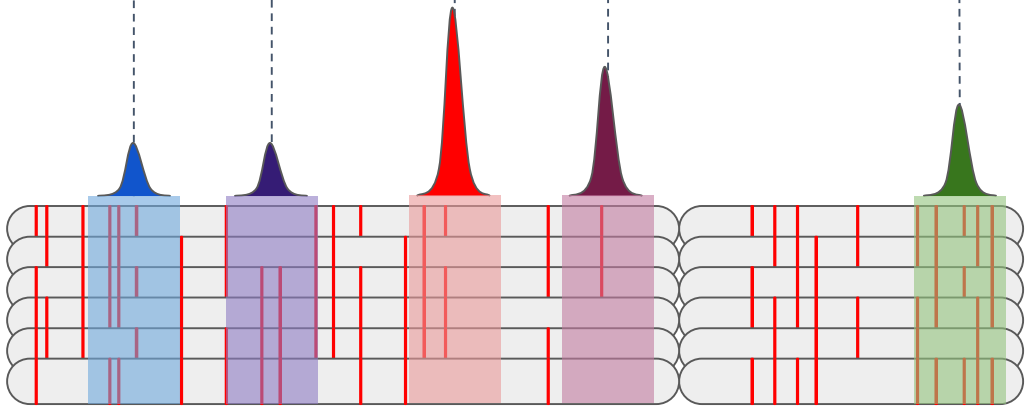
Genotypes:



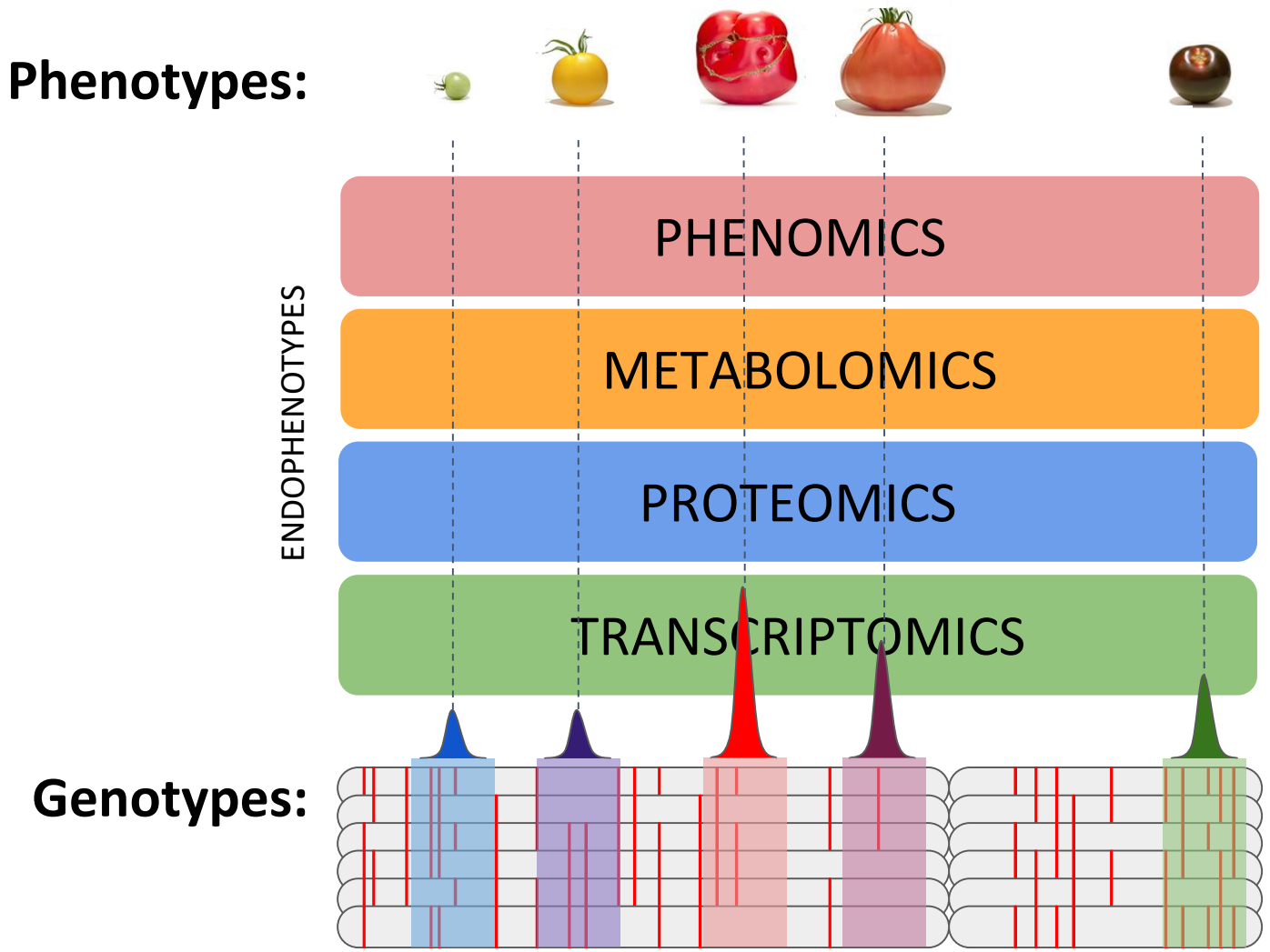
QTL / GWAS

Phenotypes:

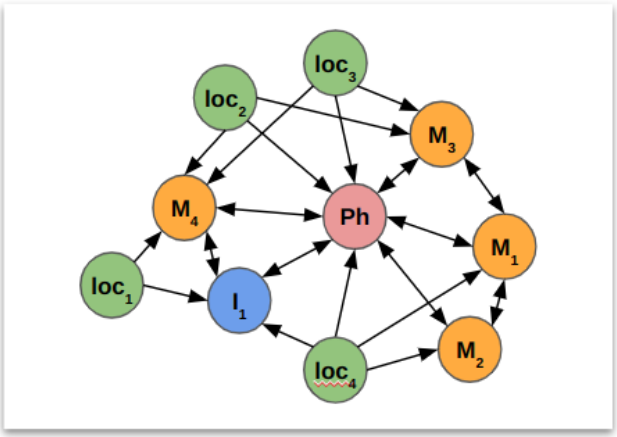
Genotypes:

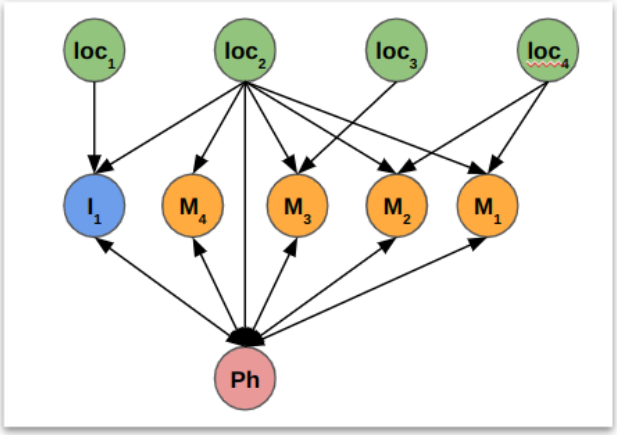


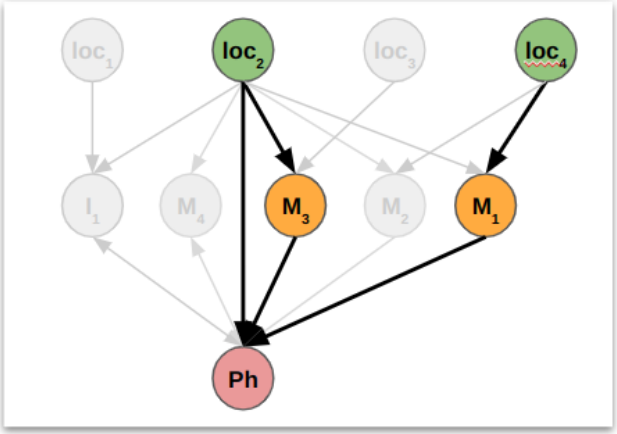
QTL / GWAS / Systems Genetics

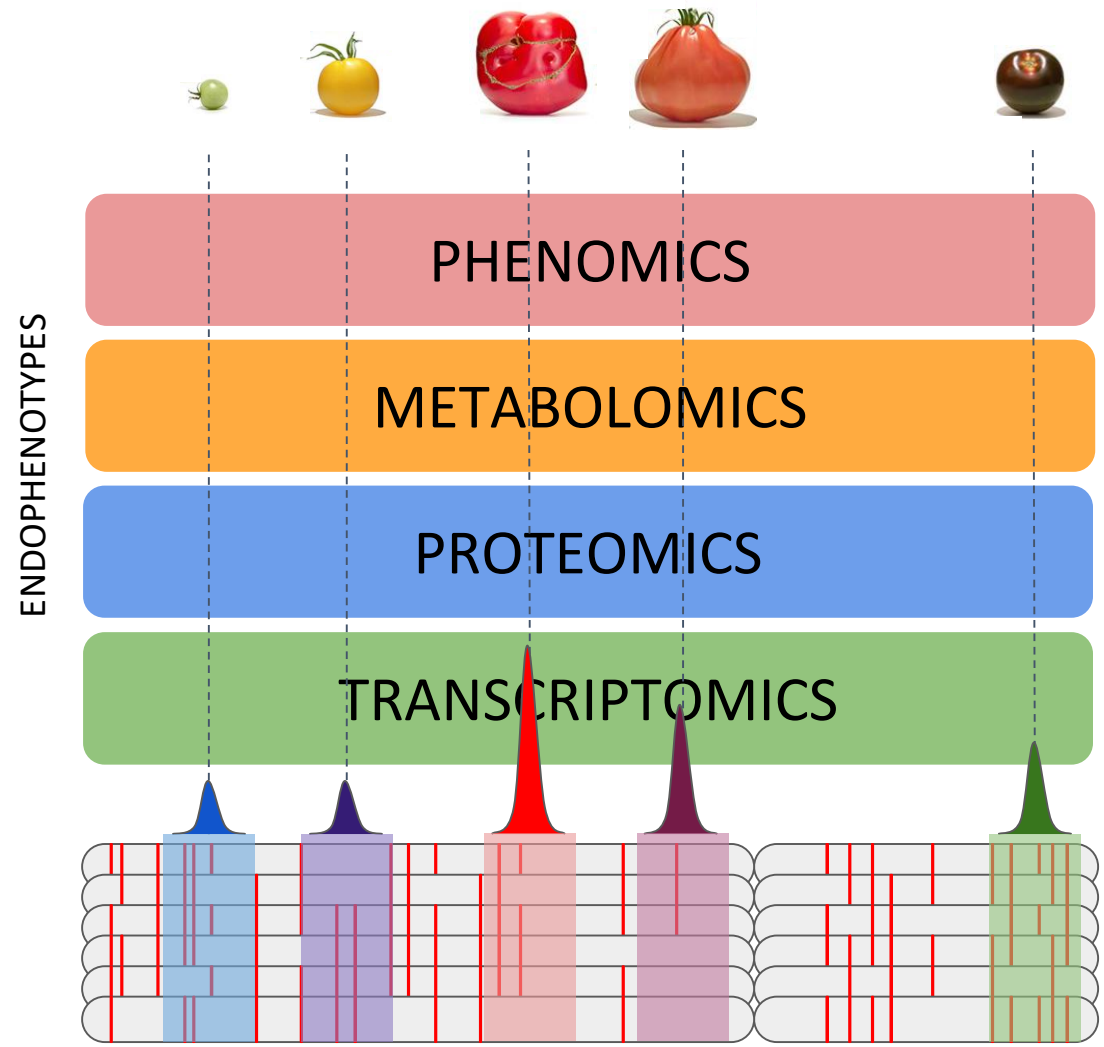


Genomic networking







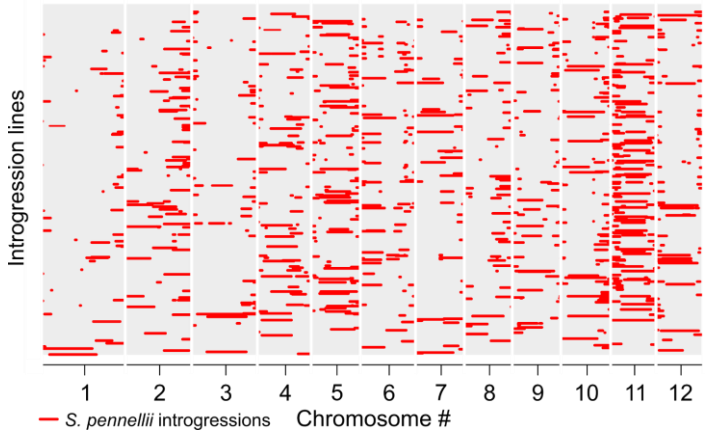


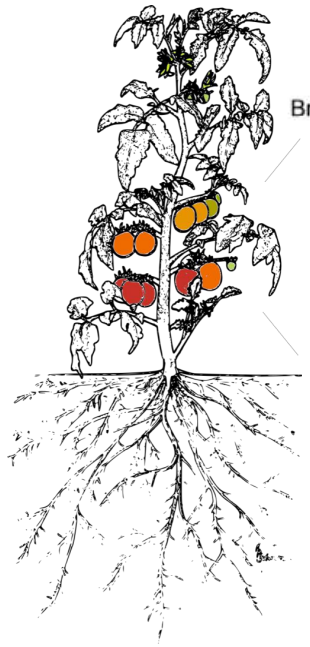
From genomic networking to molecular mechanisms

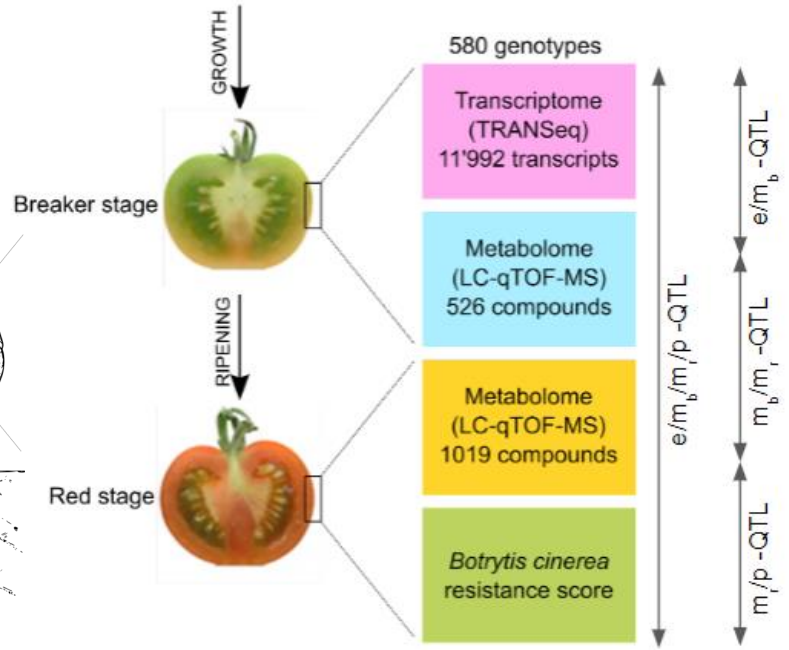
S. pennellii
#WildBeast #Resistant
#Toxic #StayAway

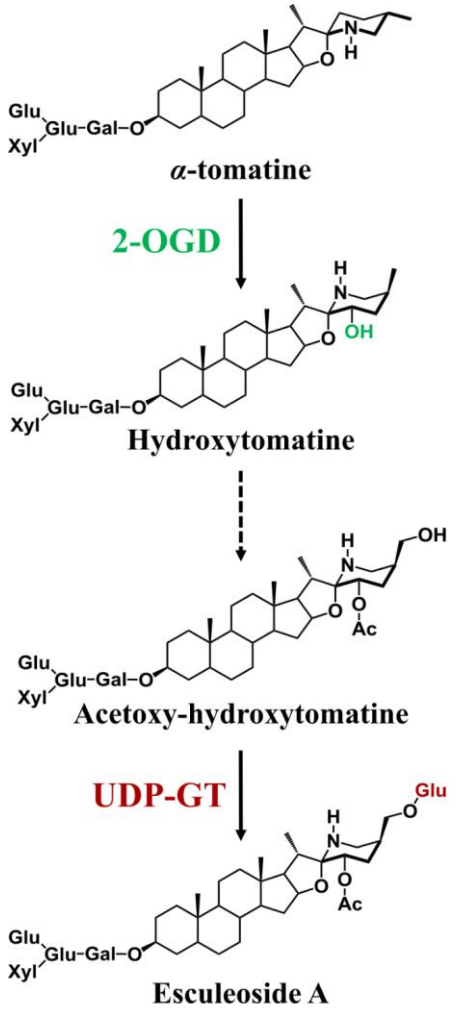
S. lycopersicum
#Domesticated #Sensitive
#Sweet #TakeMeHome

#Domesticated #Resistant
#Sweet #TakeMeHome



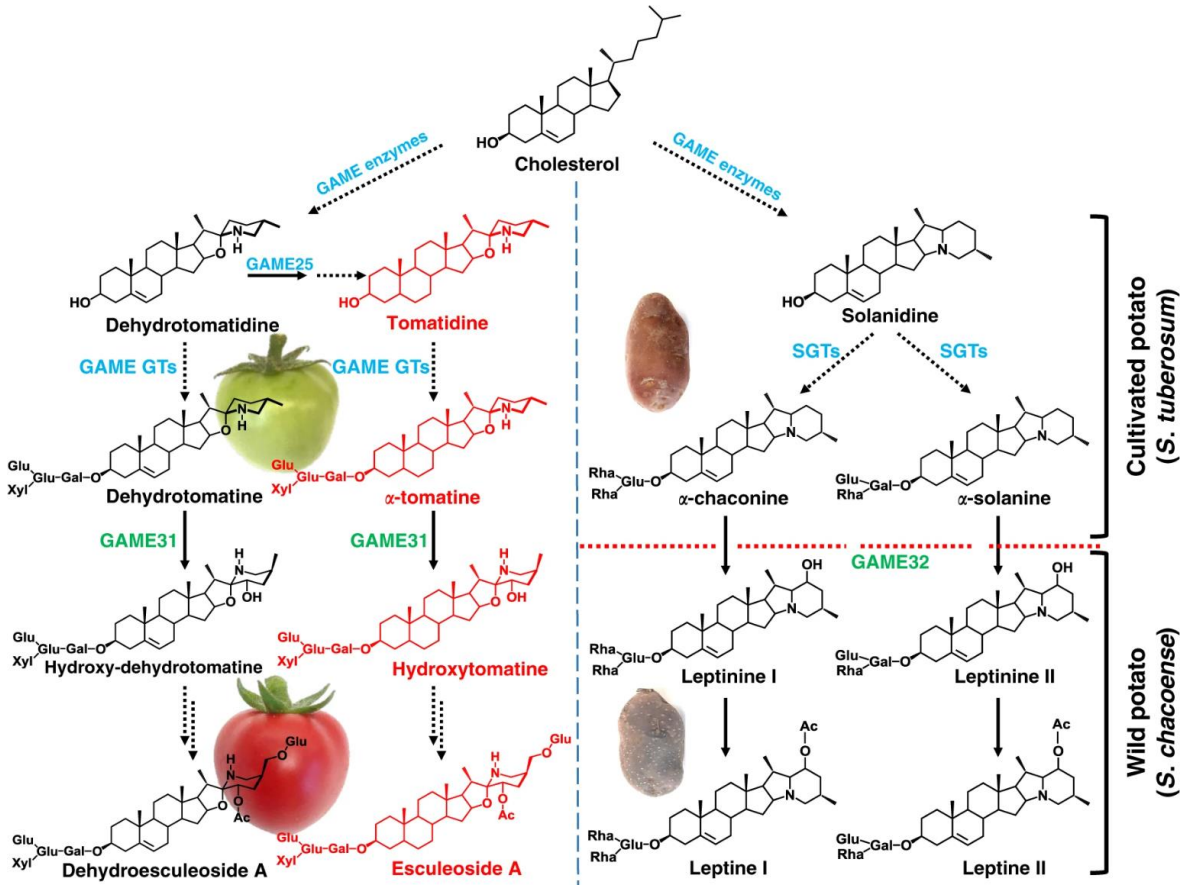


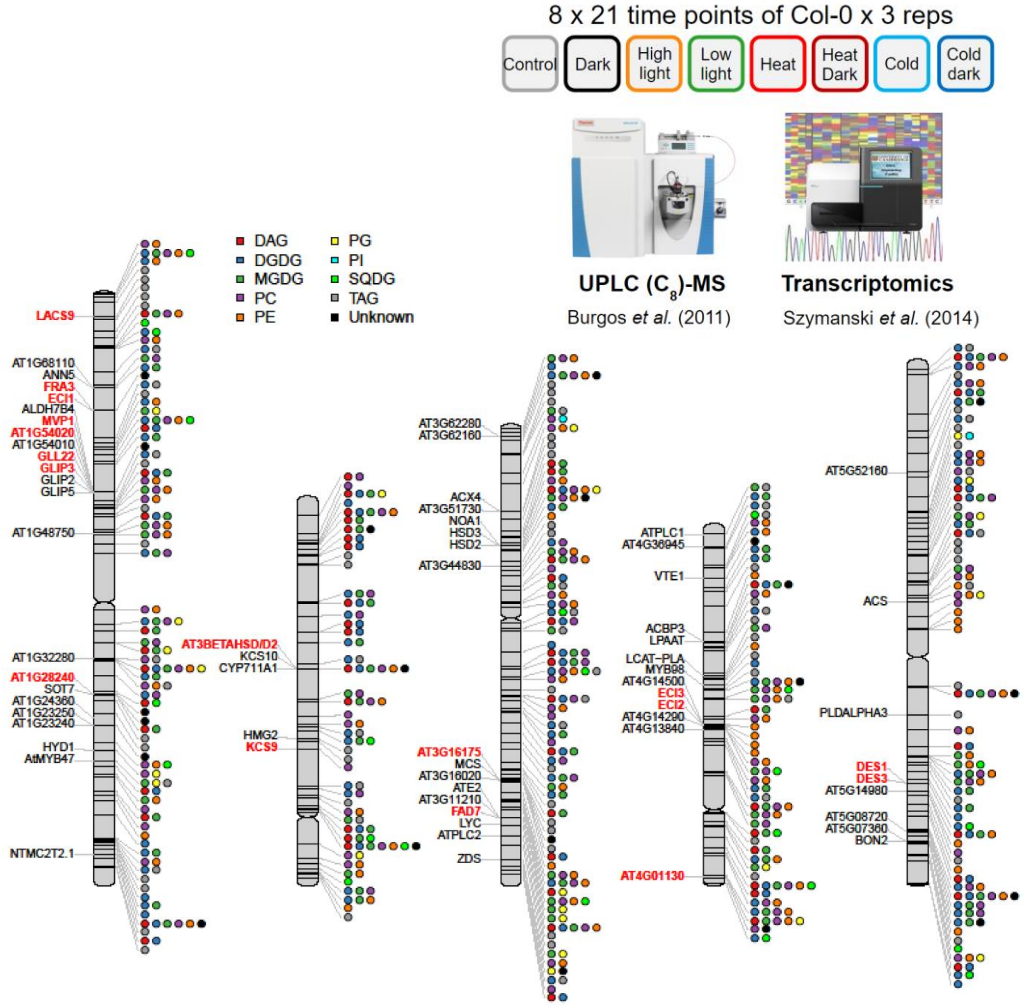




Metabolic mechanisms of *Botrytis cinerea* resistance in tomato fruits

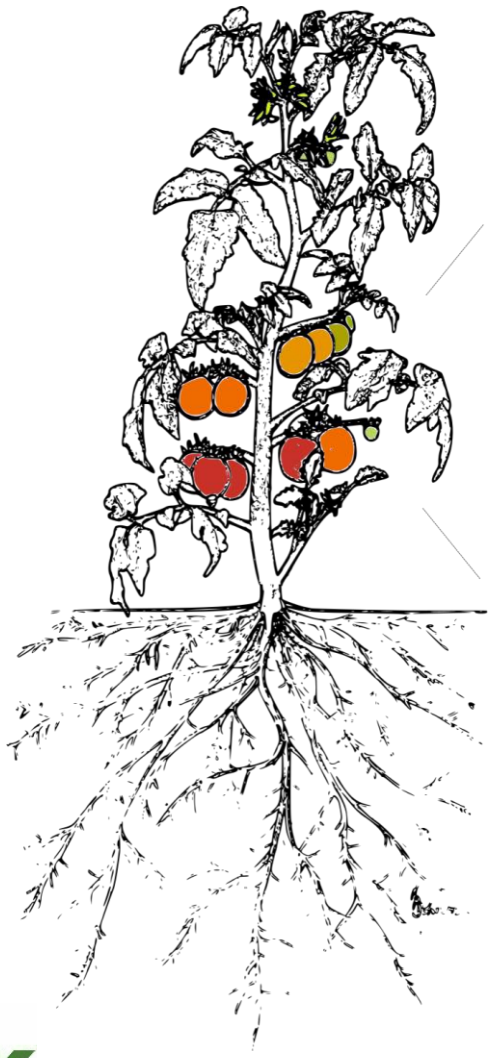
From genomic networking to molecular mechanisms

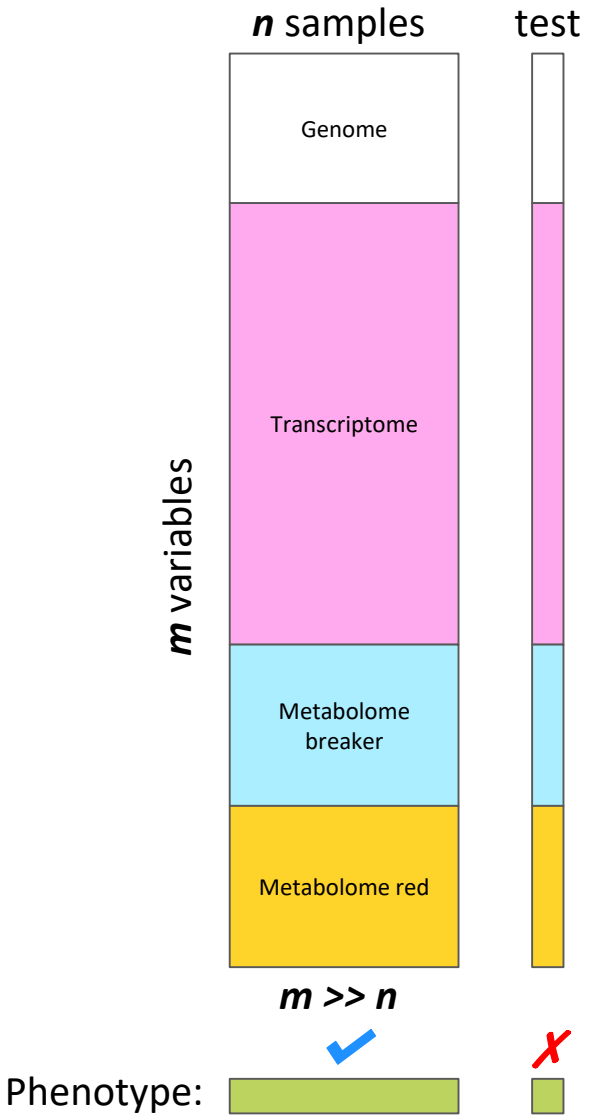


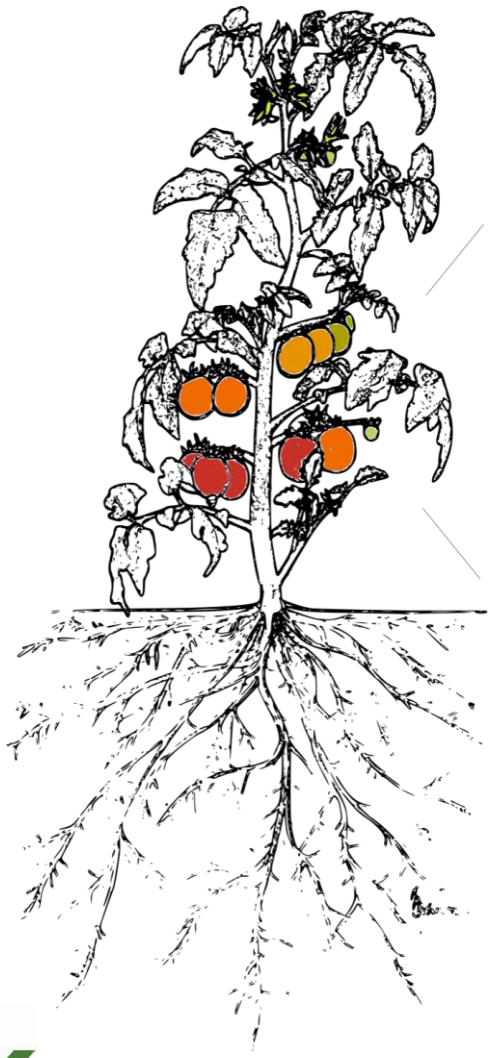


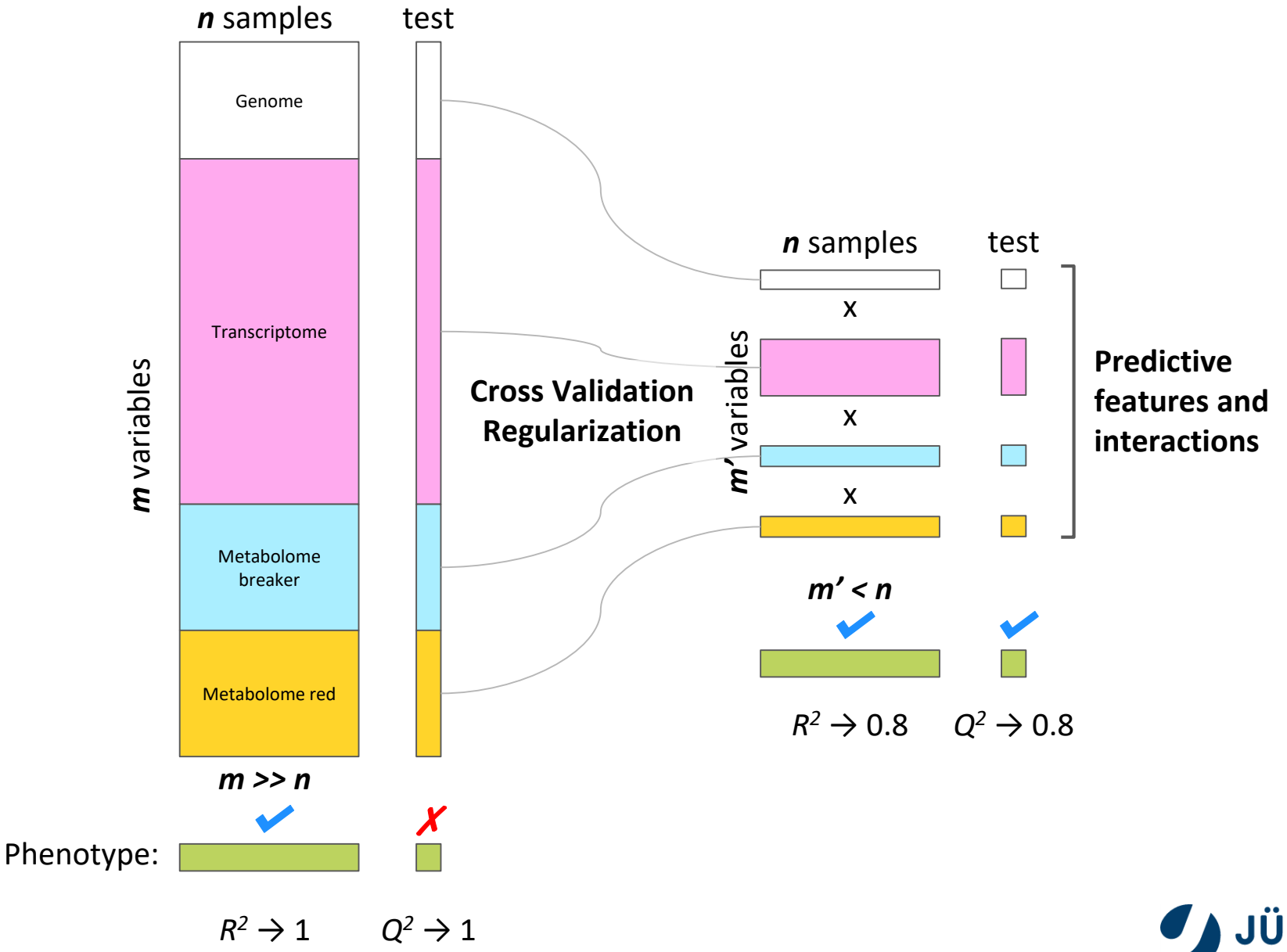
Metabolic mechanisms of herbivore defense in potato

Mechanisms of adaptation to environment in Arabidopsis

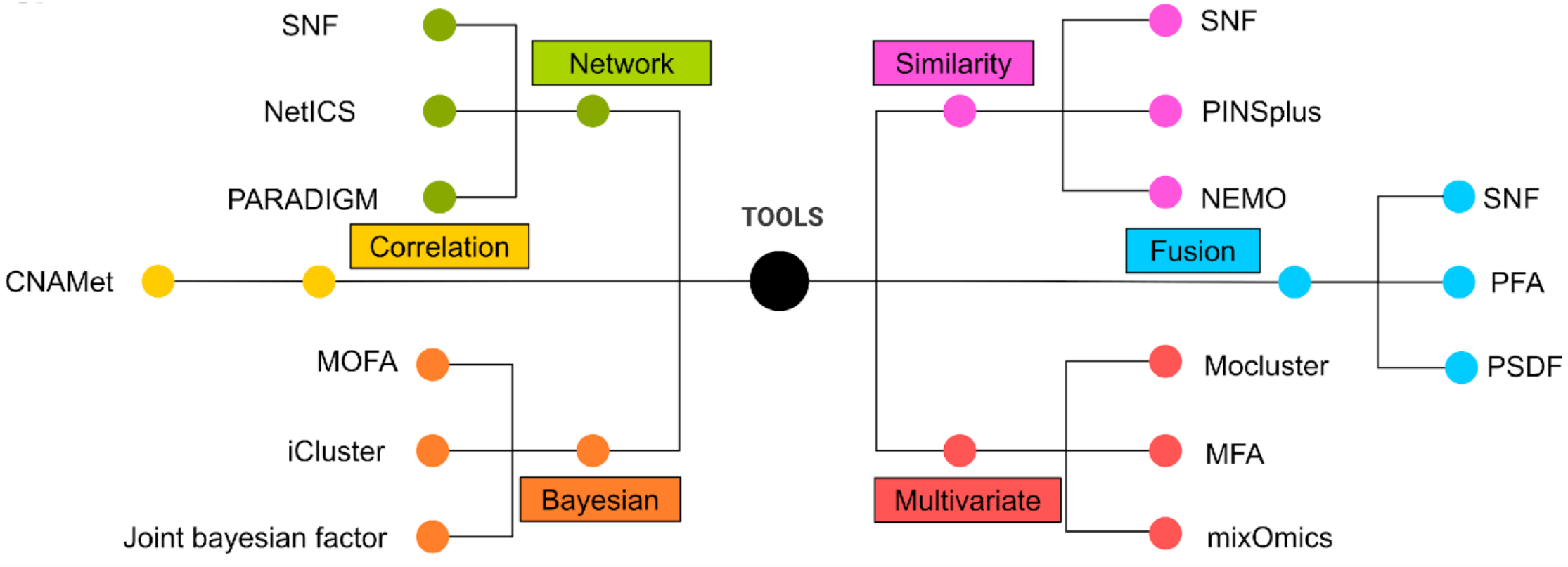



n samples
test
Genome
Transcriptome
Metabolome breaker
Metabolome red
m variables
m >> n
Phenotype:
R² → 1
Q² → 1

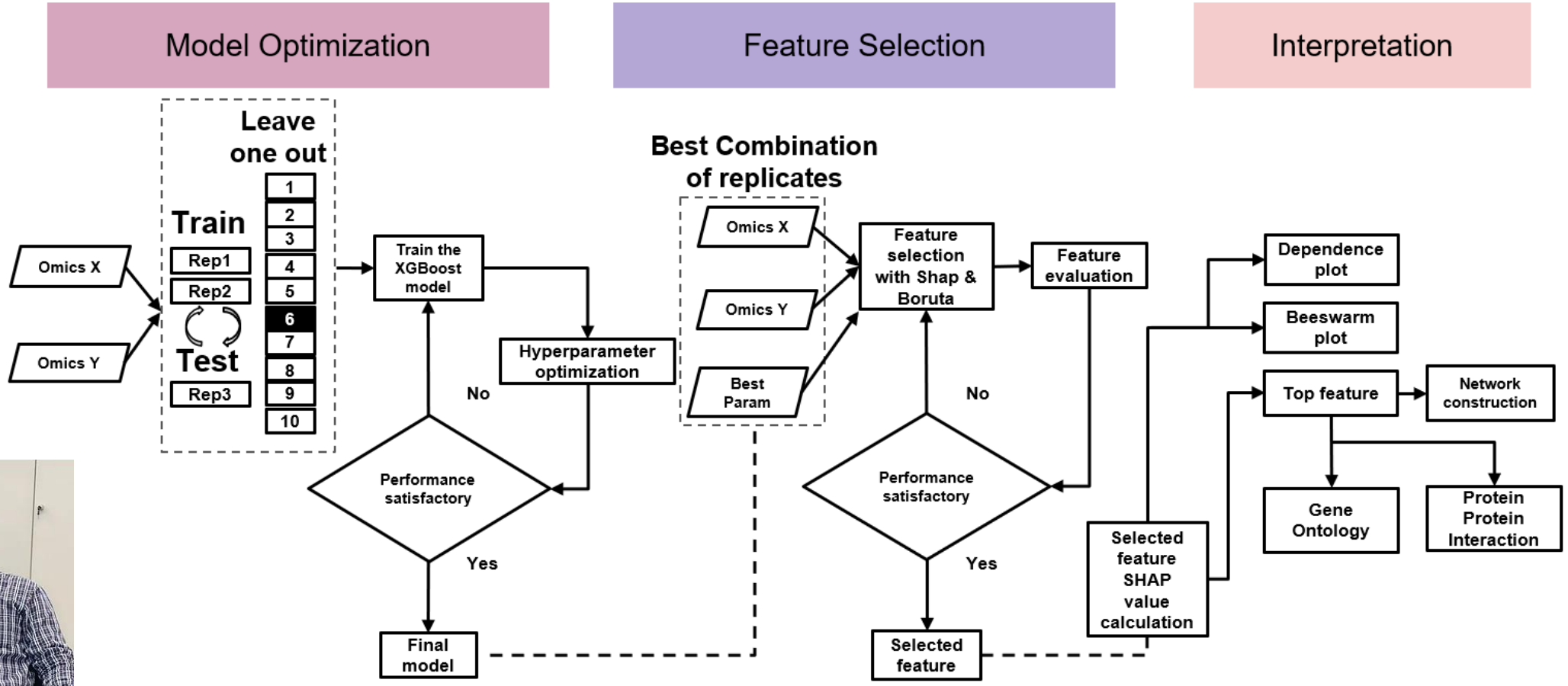



n samples
test
Genome
Transcriptome
Metabolome breaker
Metabolome red
m variables
m >> n
Phenotype:
R² → 1
Q² → 1
Cross Validation
Regularization
n samples
test
x
x
x
m' variables
m' < n
R² → 0.8
Q² → 0.8
Predictive features and interactions

What kind of machine learning / AI?



panomiX https://szymanskilab.shinyapps.io/panomiX/

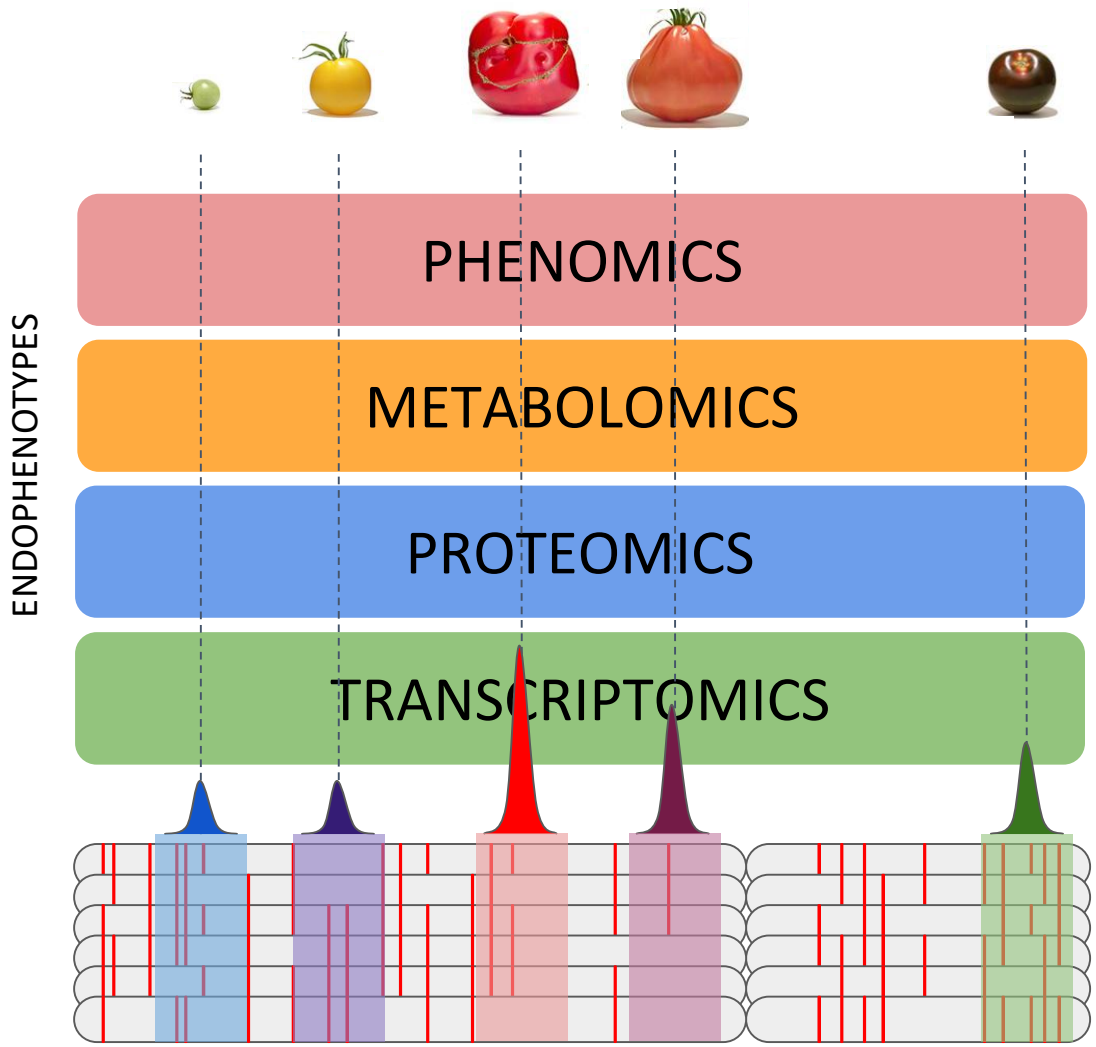


Ankur Sahu

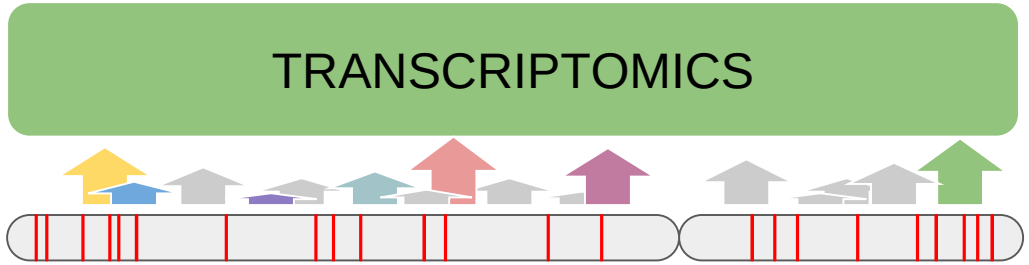
IPK LEIBNIZ INSTITUTE

JÜLICH Forschungszentrum

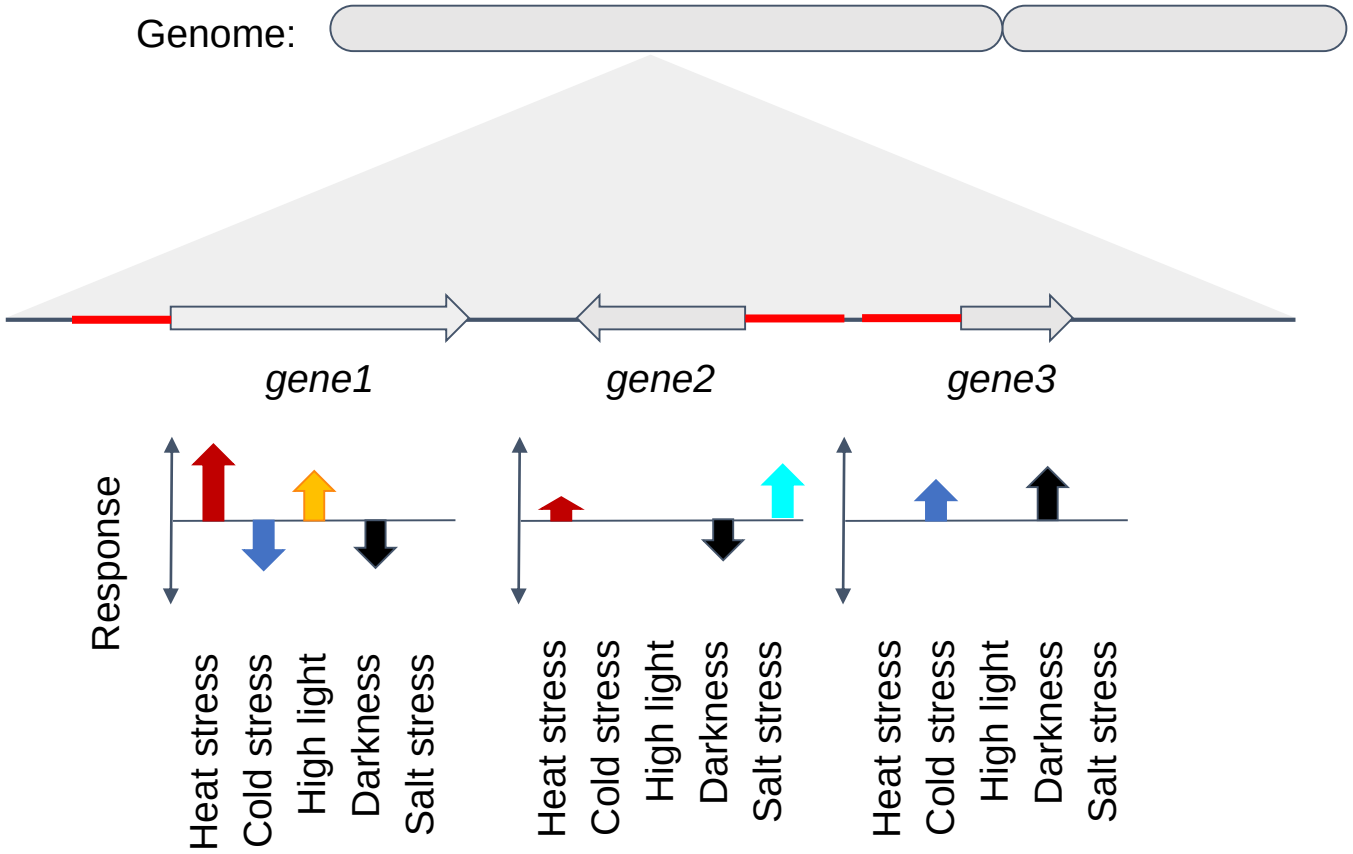
Systems genetics



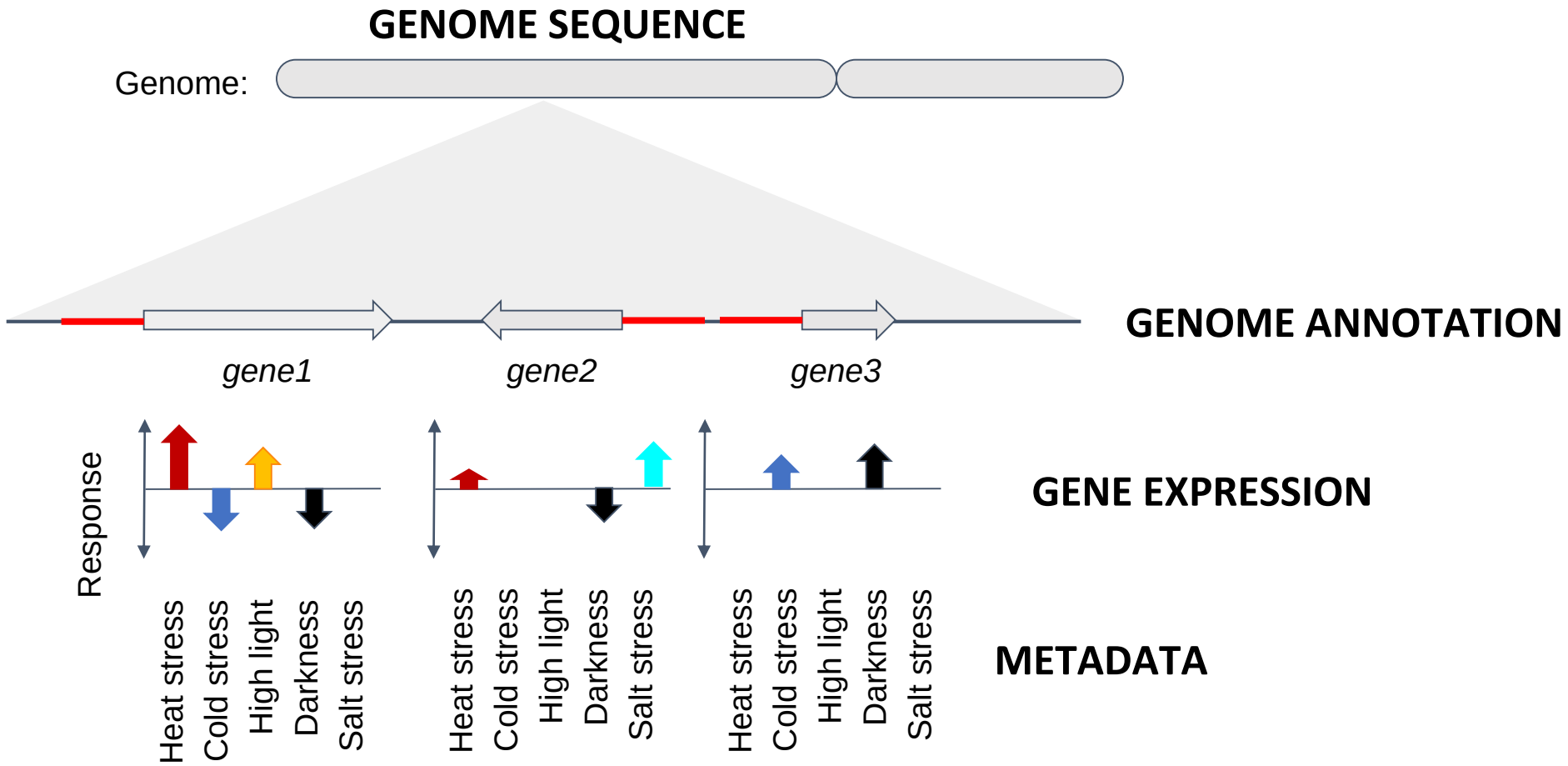
What about a simpler question?



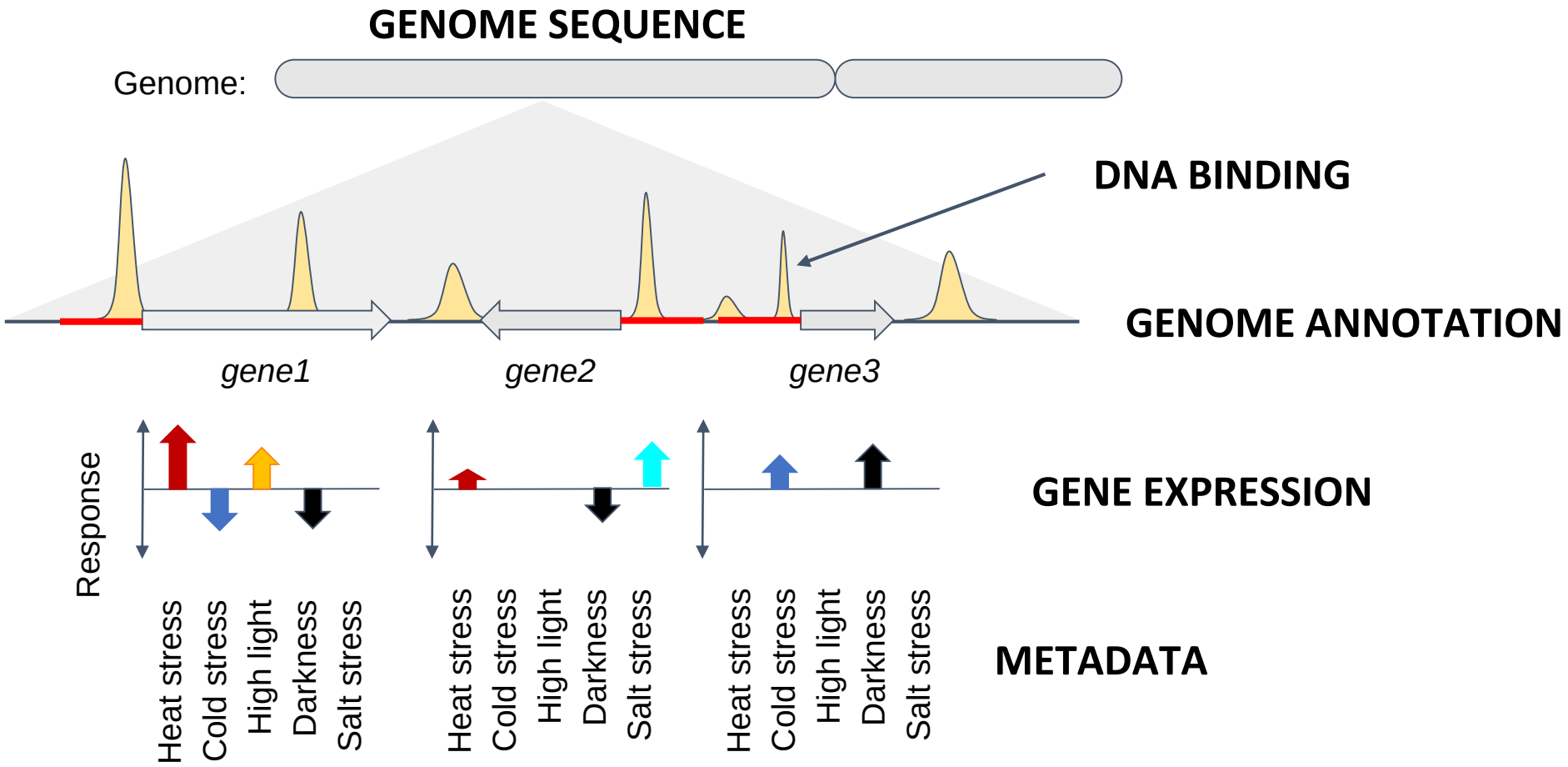
Gene expression regulation



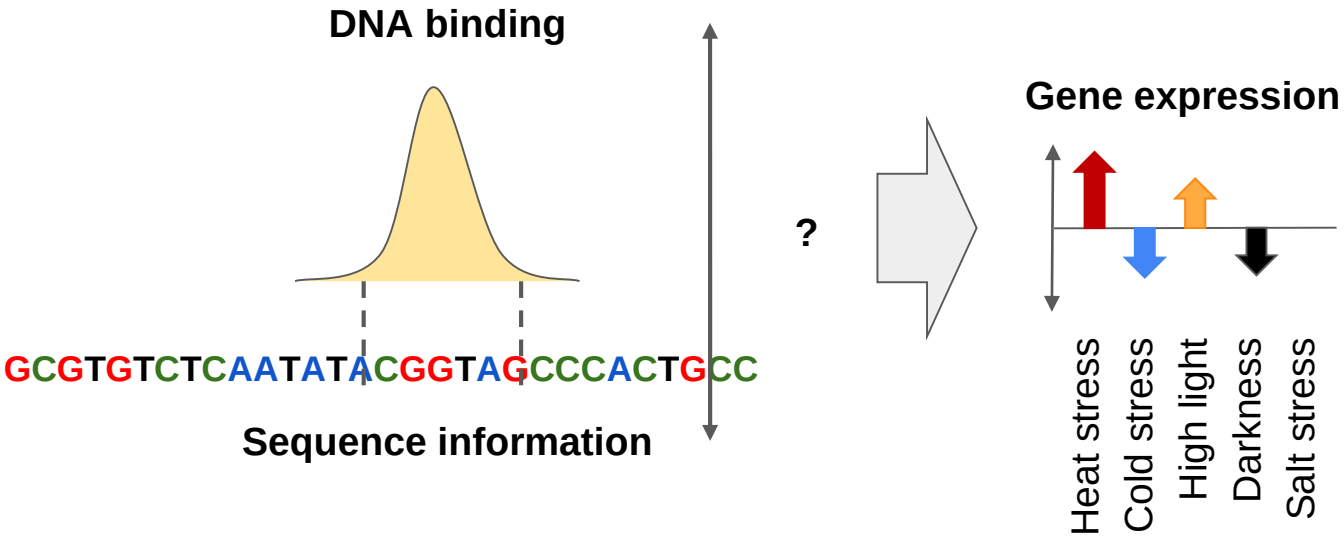
Gene expression regulation - data



Gene expression regulation - data

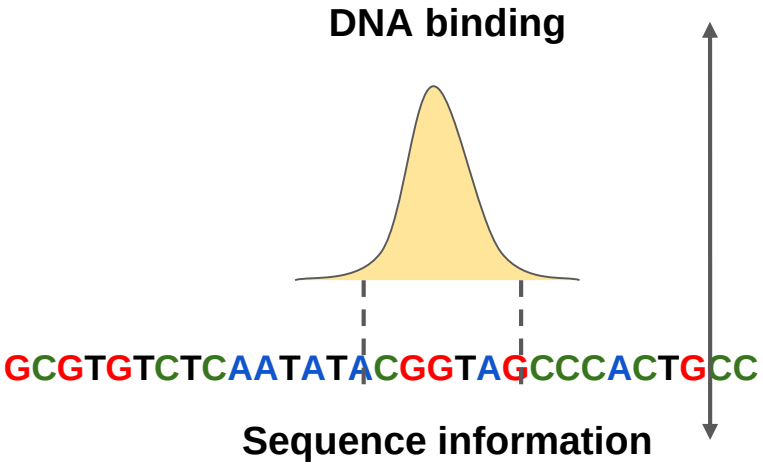


The task


DNA binding
GCGTGTCTCAATATACGGTAGCCCACTGCC
Sequence information
?
Gene expression
Heat stress
Cold stress
High light
Darkness
Salt stress

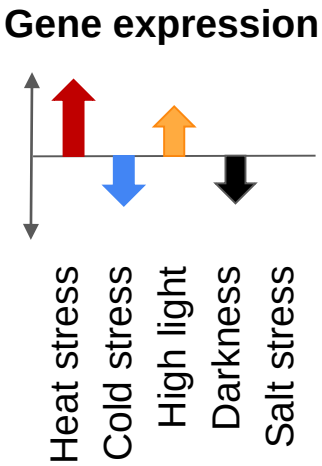
The task

> 500k DNA binding events



?

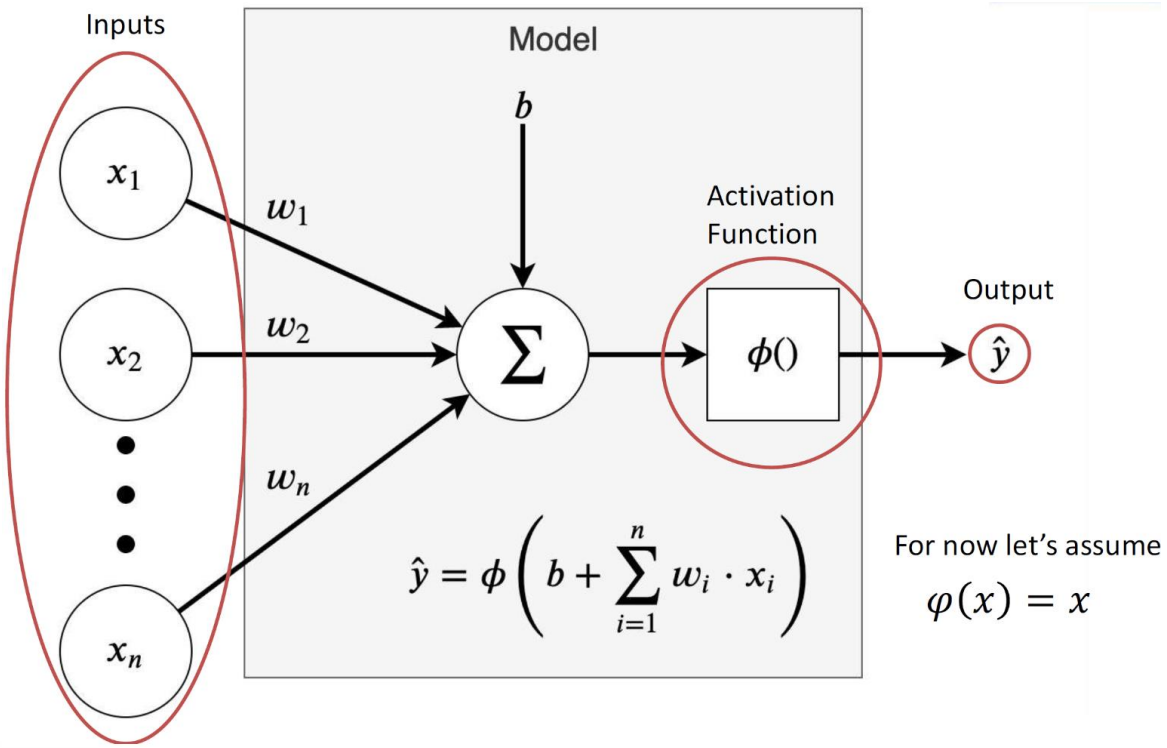
> 400 bln expression values



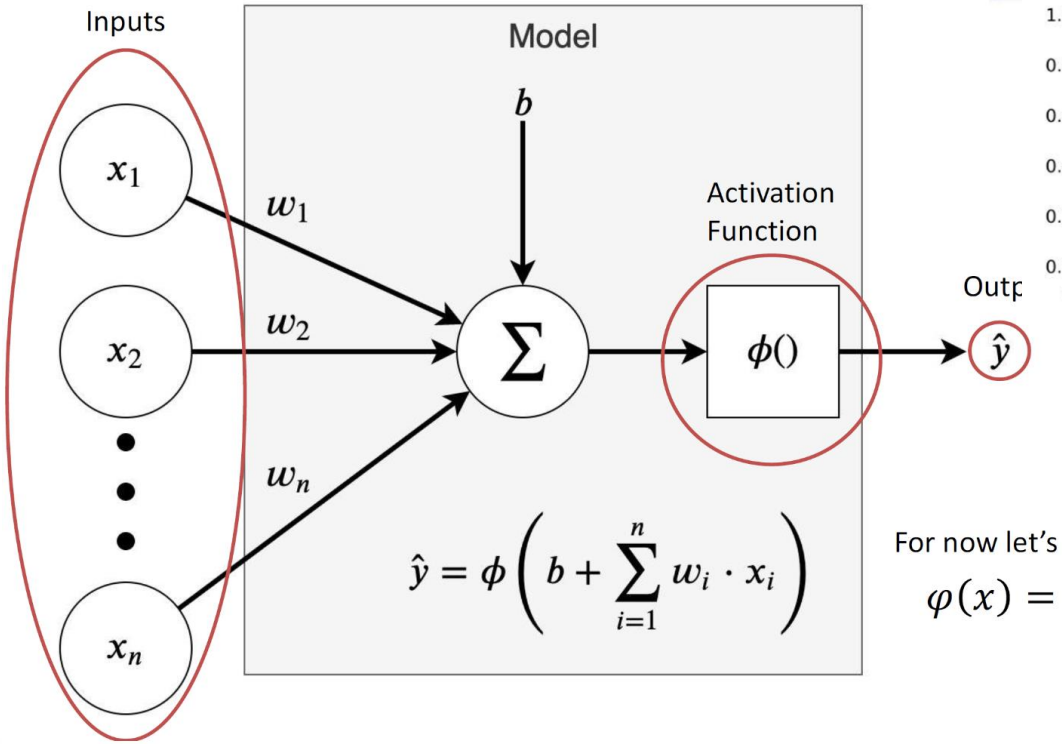
> 3 mln gene sequences

without accessions & cultivars

Artificial Neural Networks



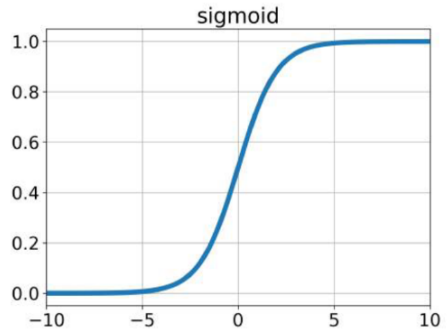
Artificial Neural Networks



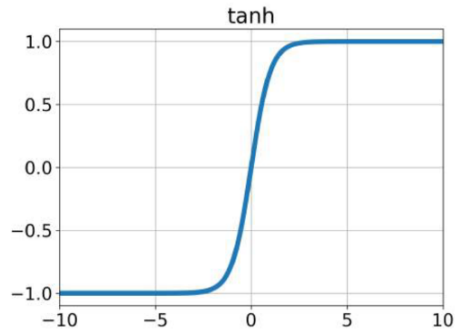
For now let's assume

$$\varphi(x) = x$$

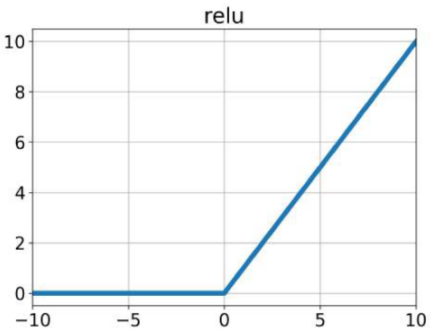
$$\varphi(x) = \frac{1}{1 + e^{-x}}$$



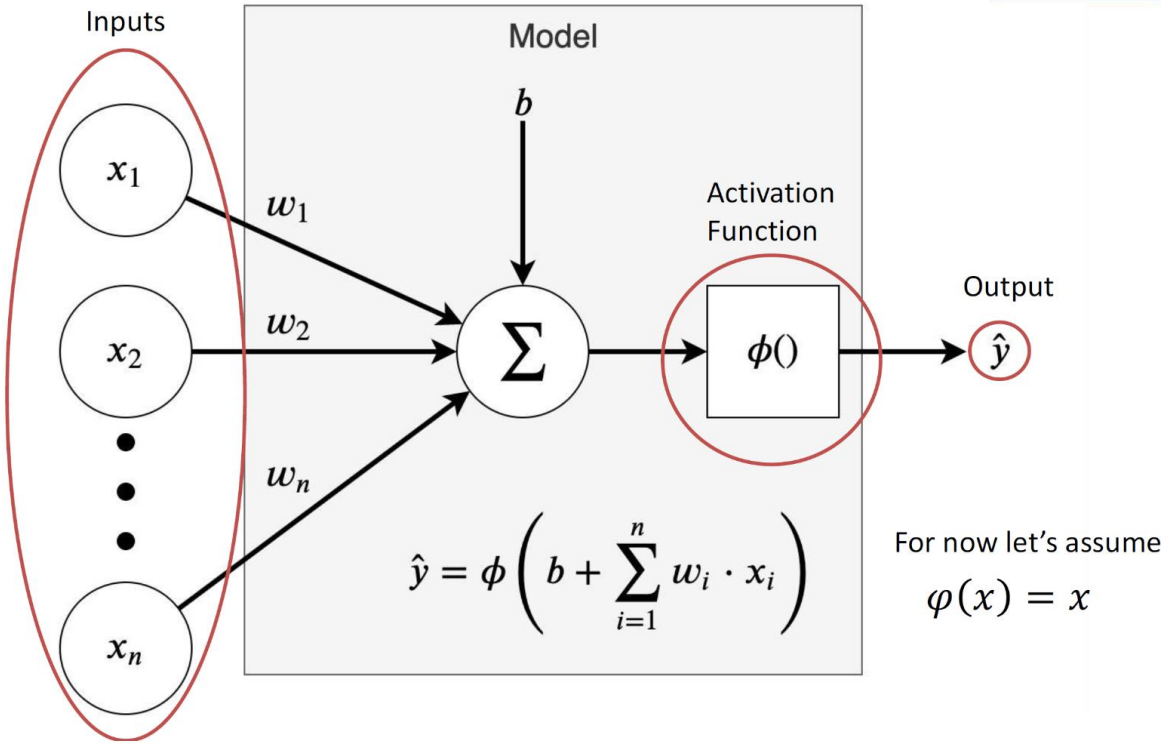
$$\varphi(x) = \frac{e^{2x} - 1}{e^{2x} + 1}$$

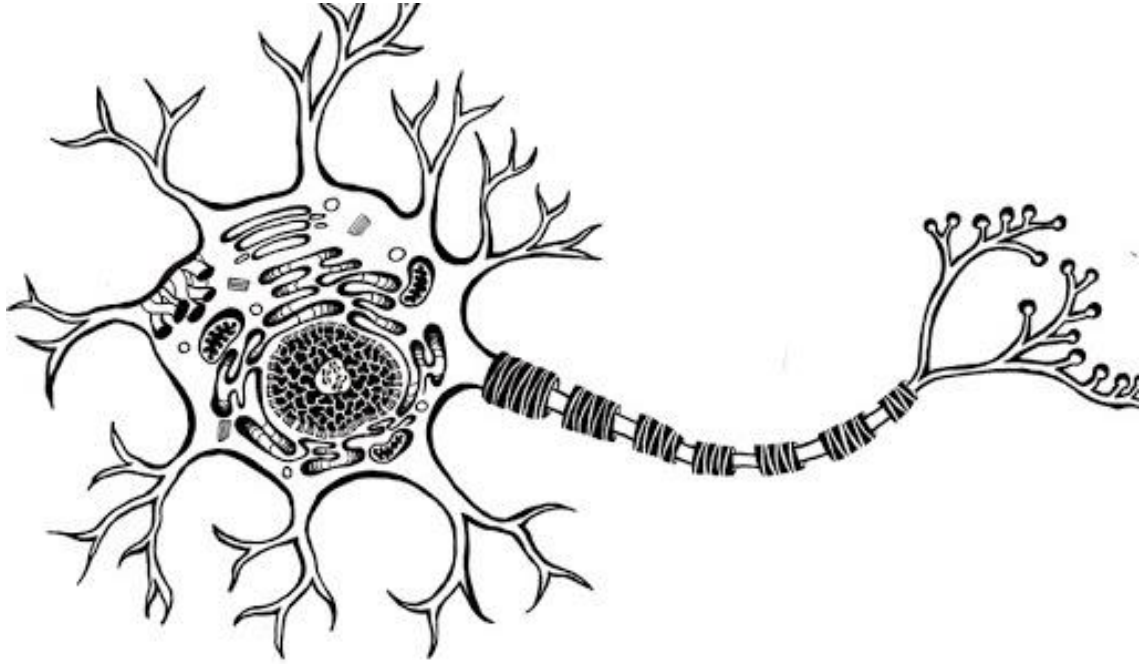


$$\varphi(x) = \max(0, x)$$

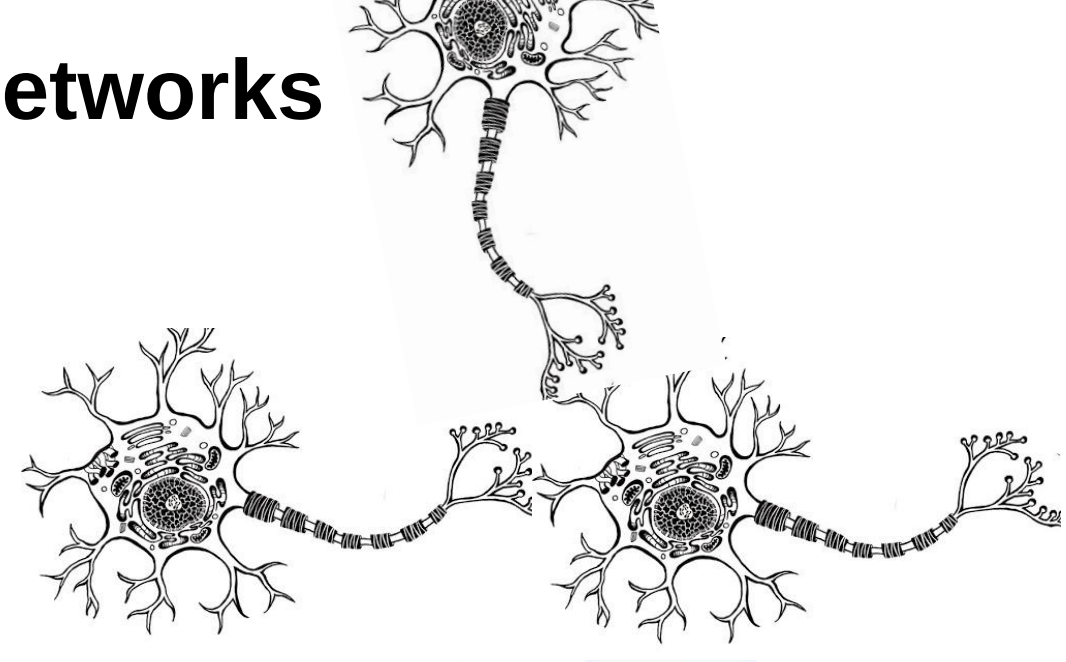


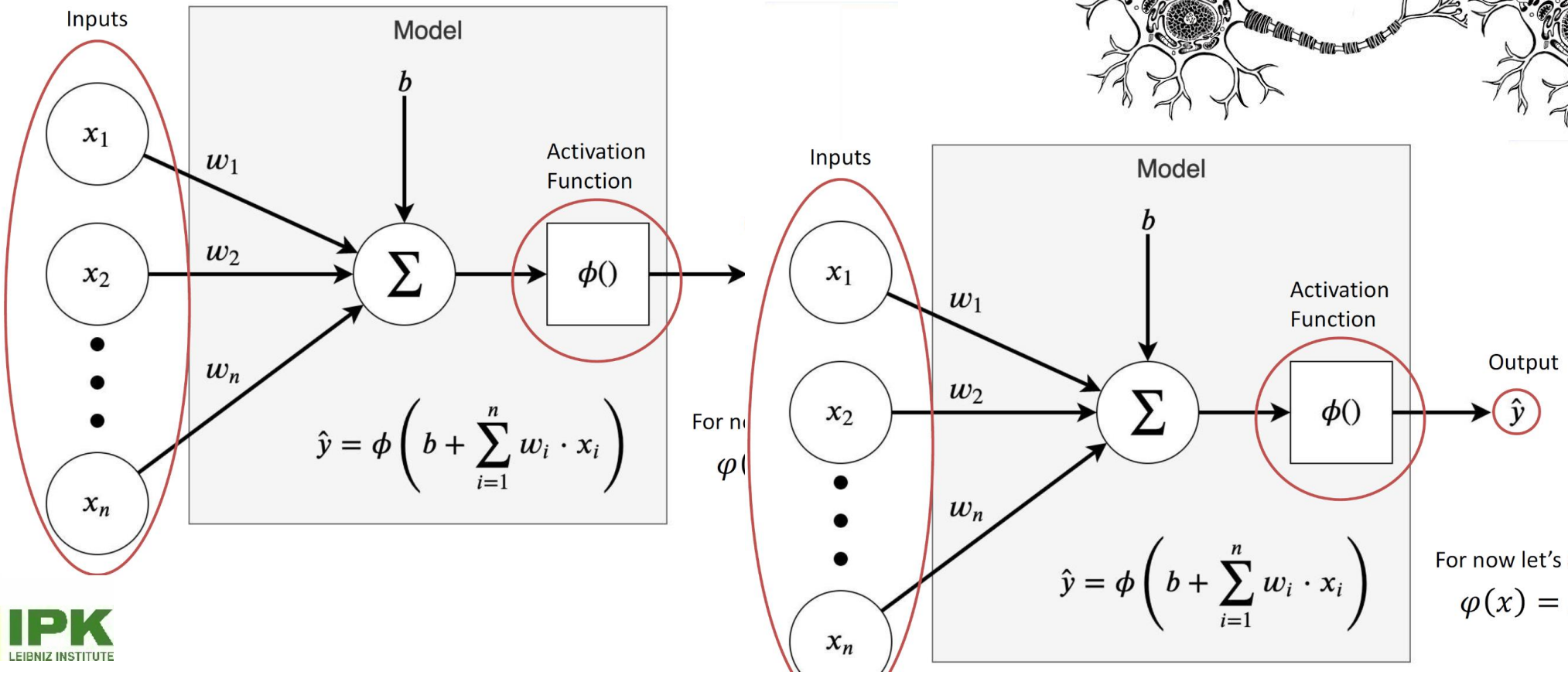
Artificial Neural Networks

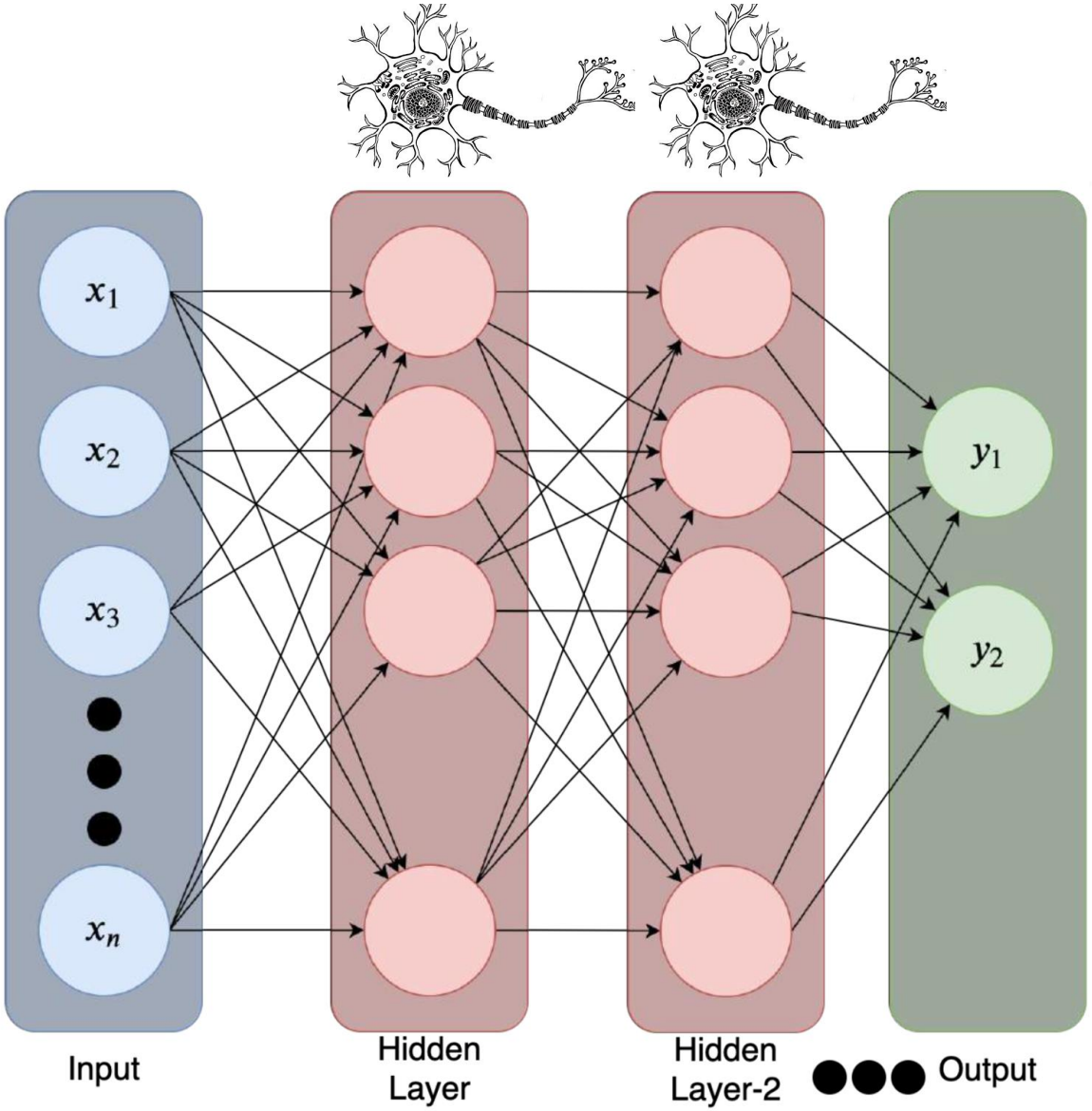




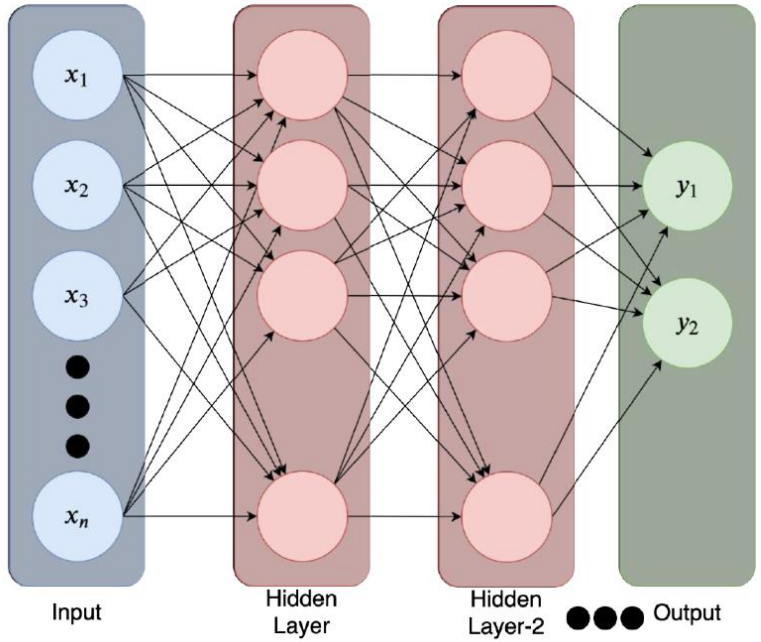
Artificial Neural Networks






x_1
x_2
x_3
x_n
y_1
y_2
Input
Hidden Layer
Hidden Layer-2
Output

Training the Artificial Neural Network model



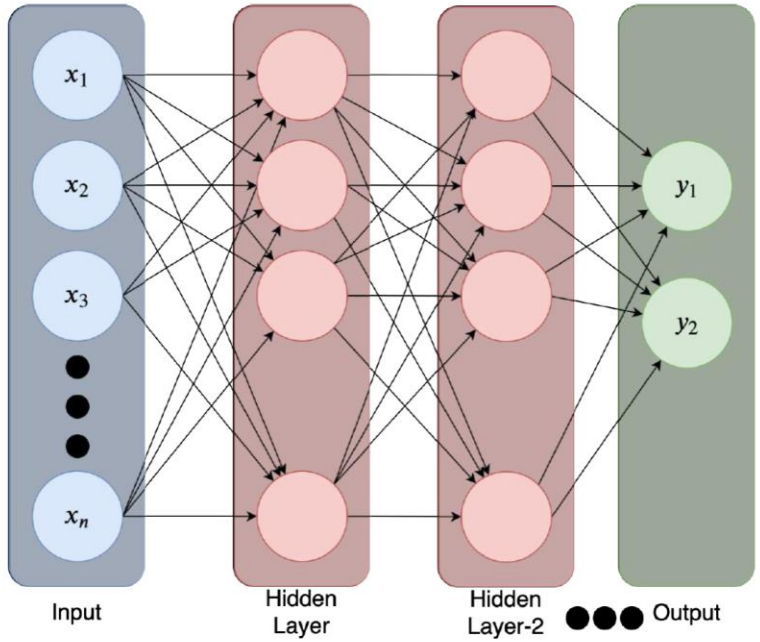
On every layer we multiply the output of the previous layer with the weights and apply activation function $\varphi(\cdot)$:

$$z^{(l)} = W^{(l)} X^T$$
$$a^{(l)} = \varphi(z^{(l)})$$

A DOG

Forward pass

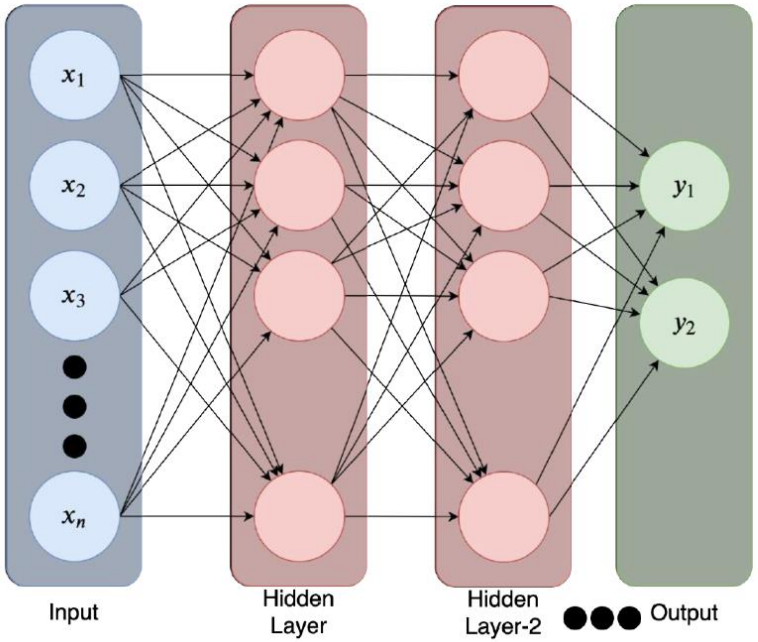
Training the Artificial Neural Network model



Weights are updated layer by layer to slightly improve the fit between expected and predicted result.

Backpropagation

Training the Artificial Neural Network model



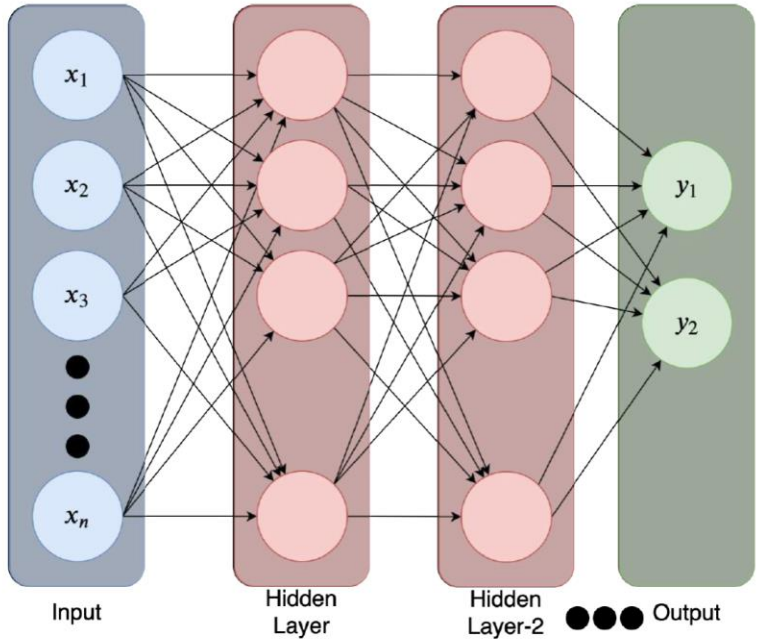
On every layer we multiply the output of the previous layer with the weights and apply activation function $\varphi(\cdot)$:

$$z^{(l)} = W^{(l)} X^T$$
$$a^{(l)} = \varphi(z^{(l)})$$

A CAT

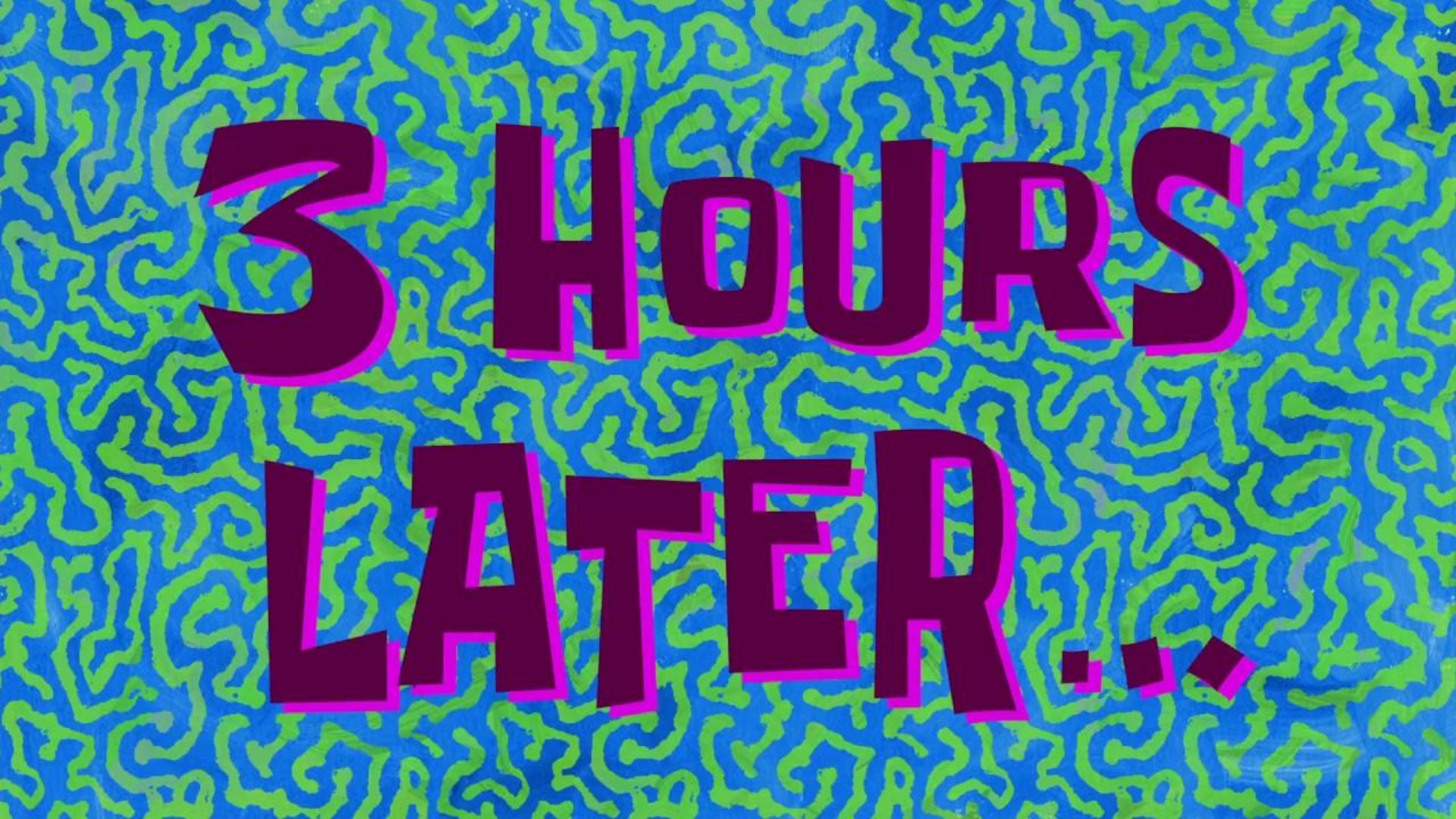
Forward pass

Training the Artificial Neural Network model

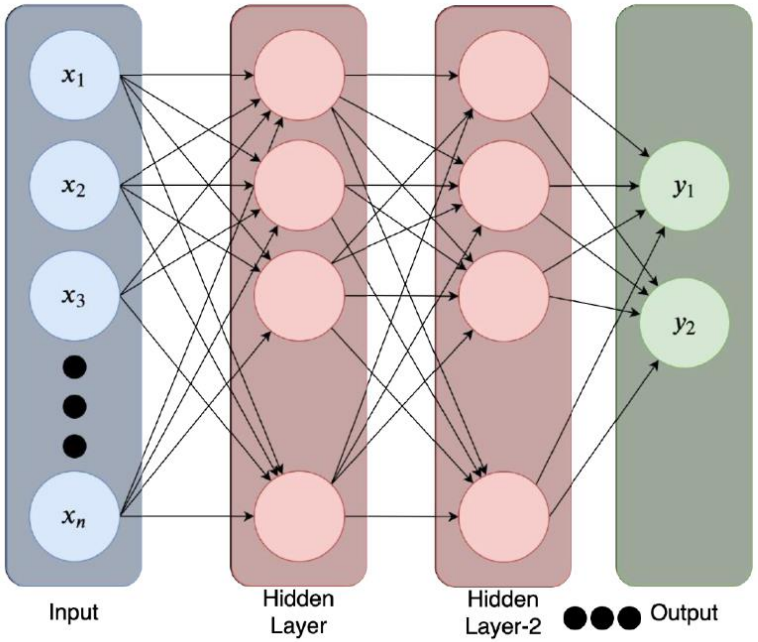


Weights are updated layer by layer to slightly improve the fit between expected and predicted result.

Backpropagation


3 HOURS
LATER...

Training the Artificial Neural Network model

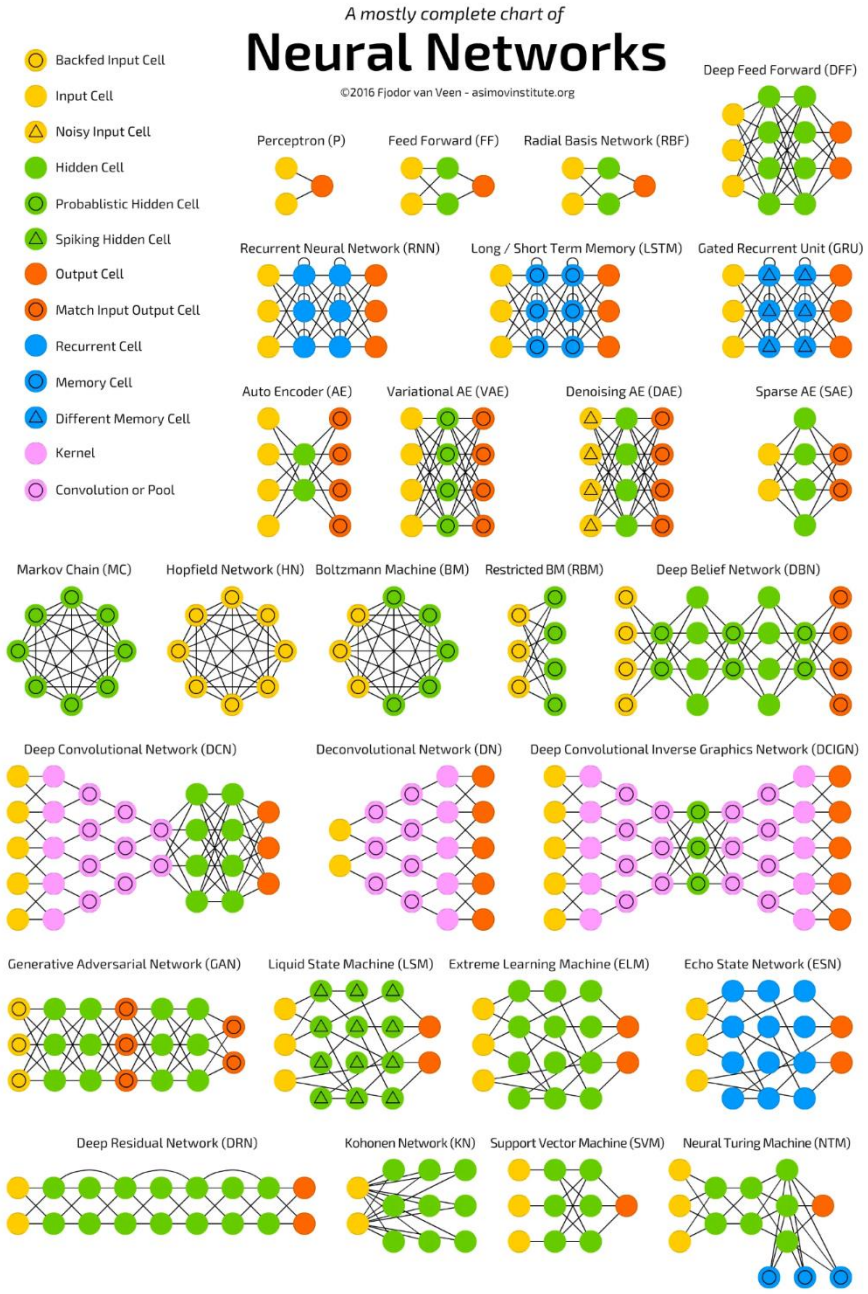


On every layer we multiply the output of the previous layer with the weights and apply activation function $\varphi(\cdot)$:

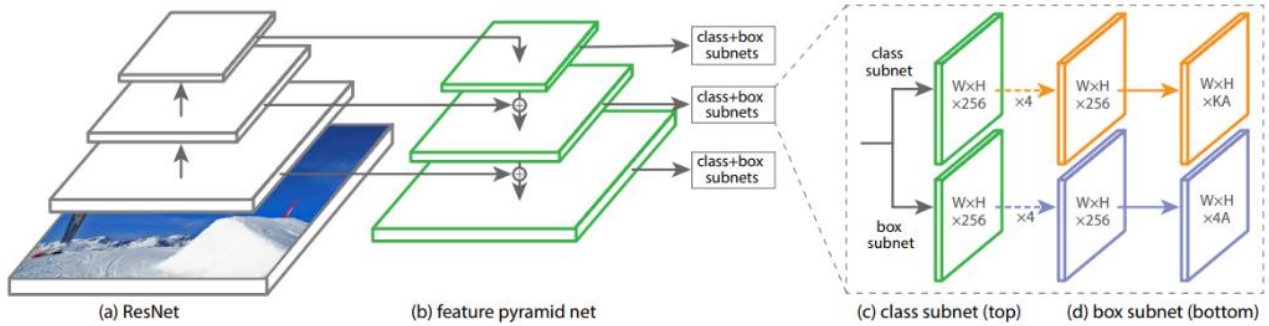
$$z^{(l)} = W^{(l)} X^T$$
$$a^{(l)} = \varphi(z^{(l)})$$

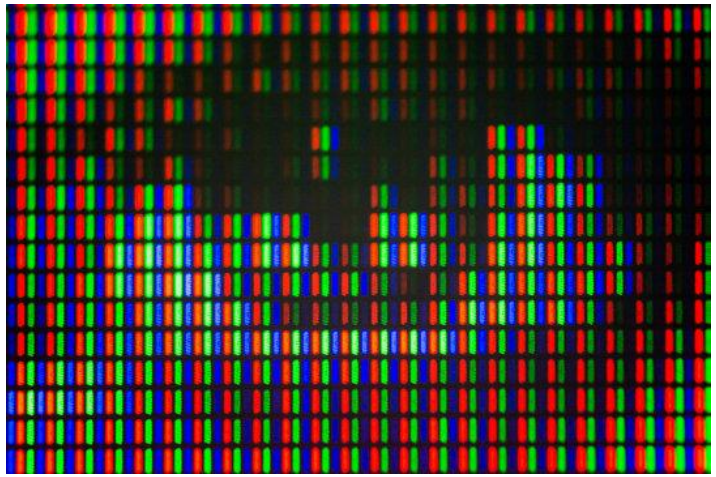
A CAT

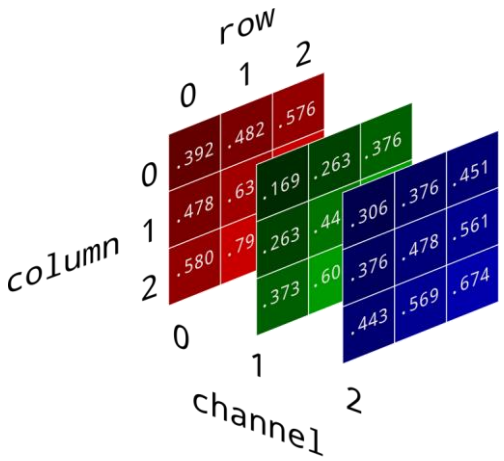
Forward pass



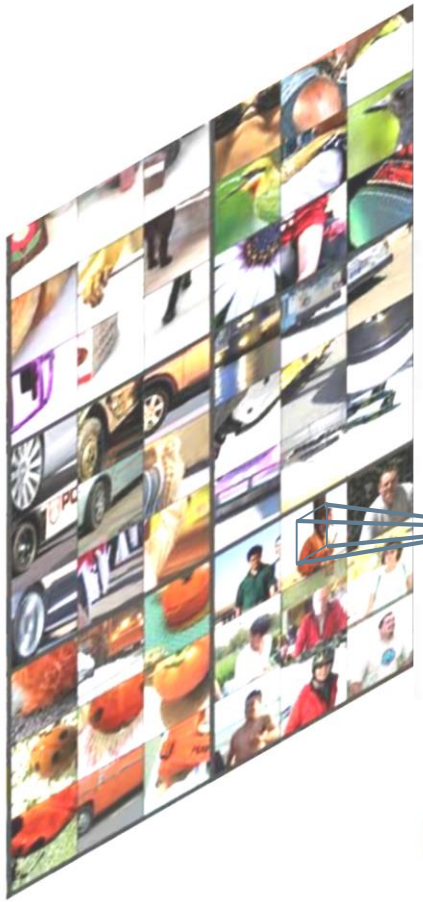
Custom Architectures - RetinaNet

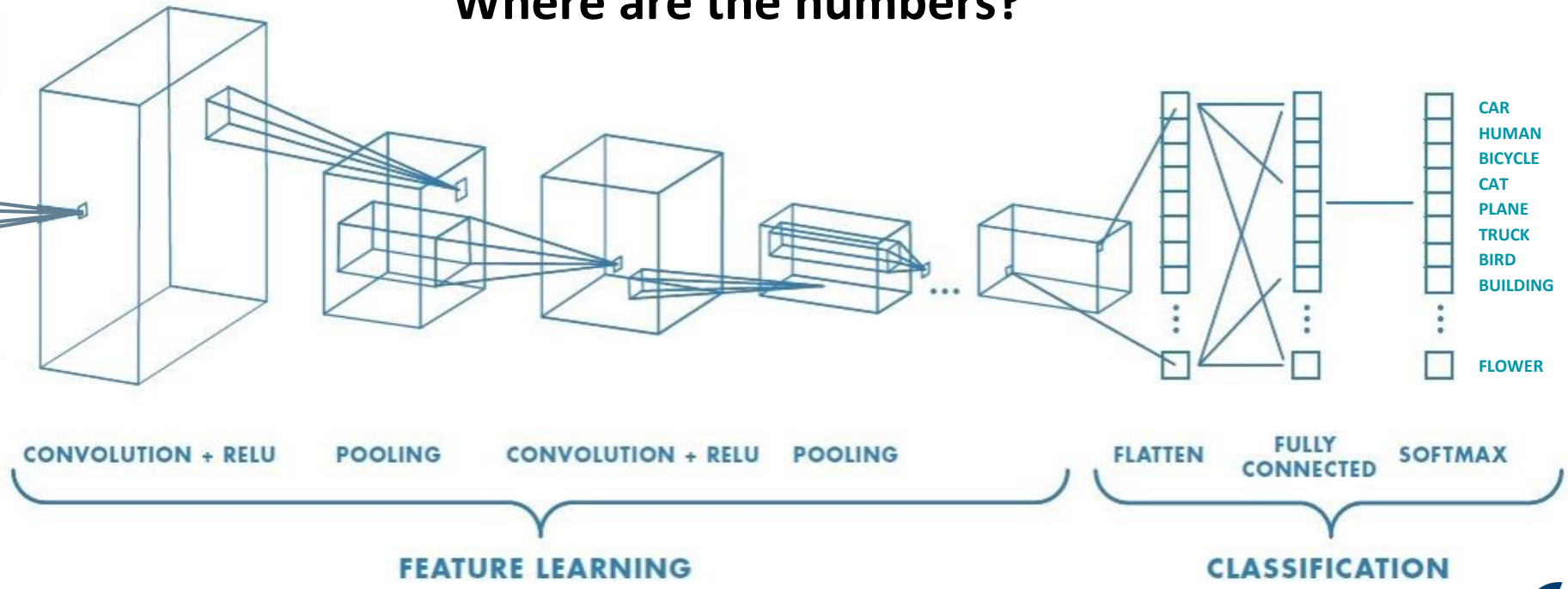


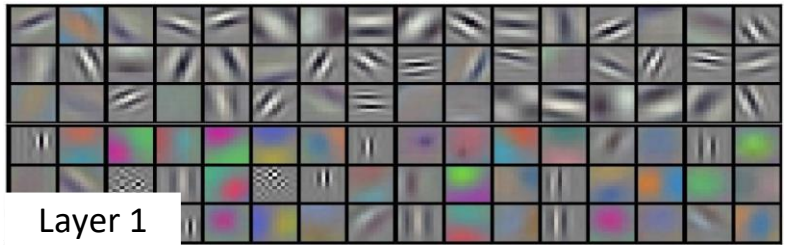


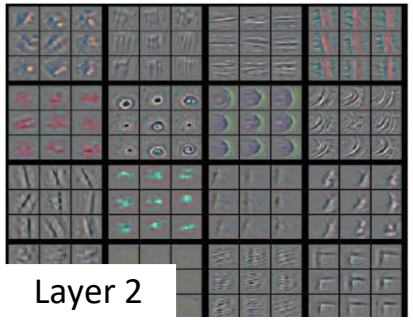


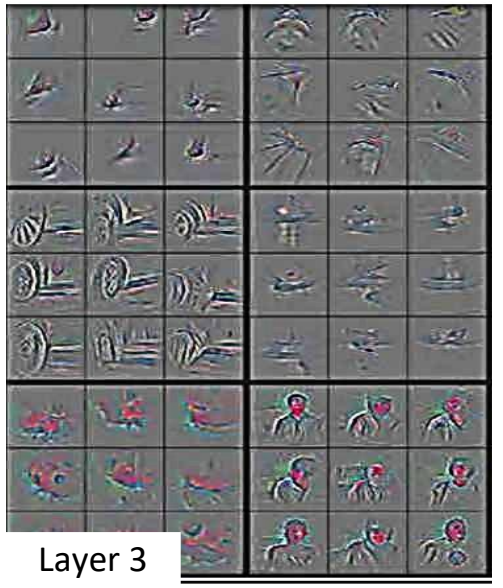
Where are the numbers?

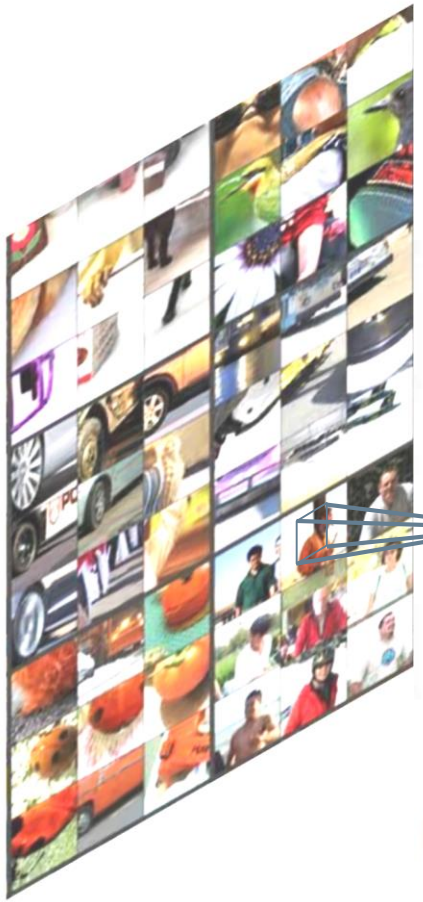


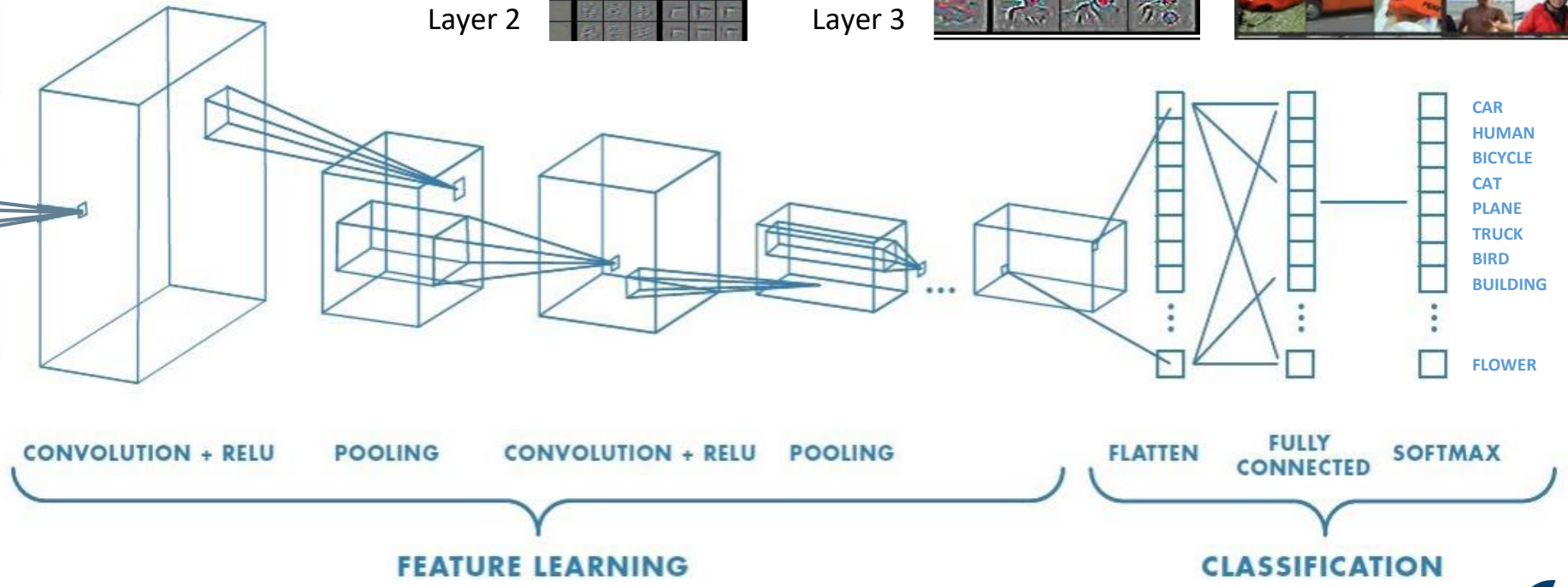



Layer 1

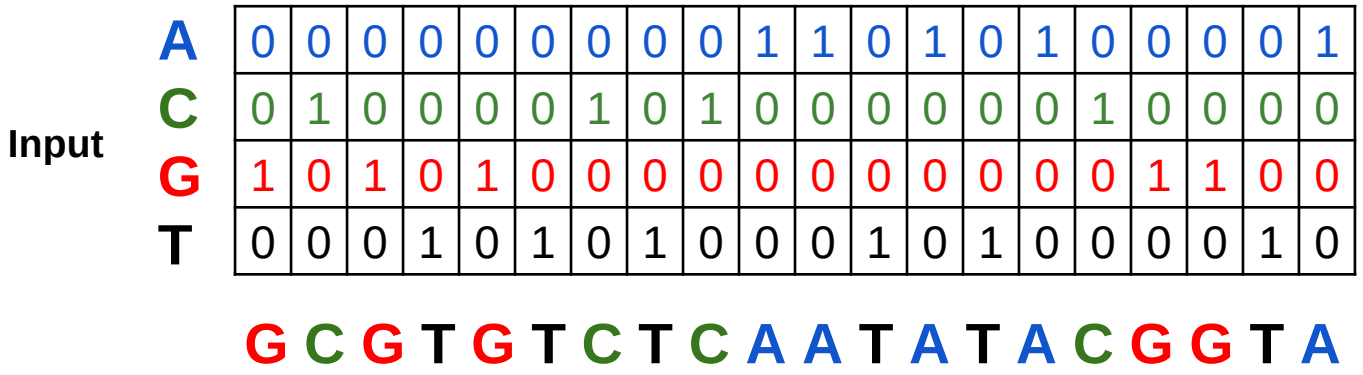

Layer 2


Layer 3




CAR
HUMAN
BICYCLE
CAT
PLANE
TRUCK
BIRD
BUILDING
FLOWER
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

Encode a DNA sequence as an image



Add more data?

TF binding?	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
Input **A**	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	1
C	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0
G	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
T	0	0	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	1	0

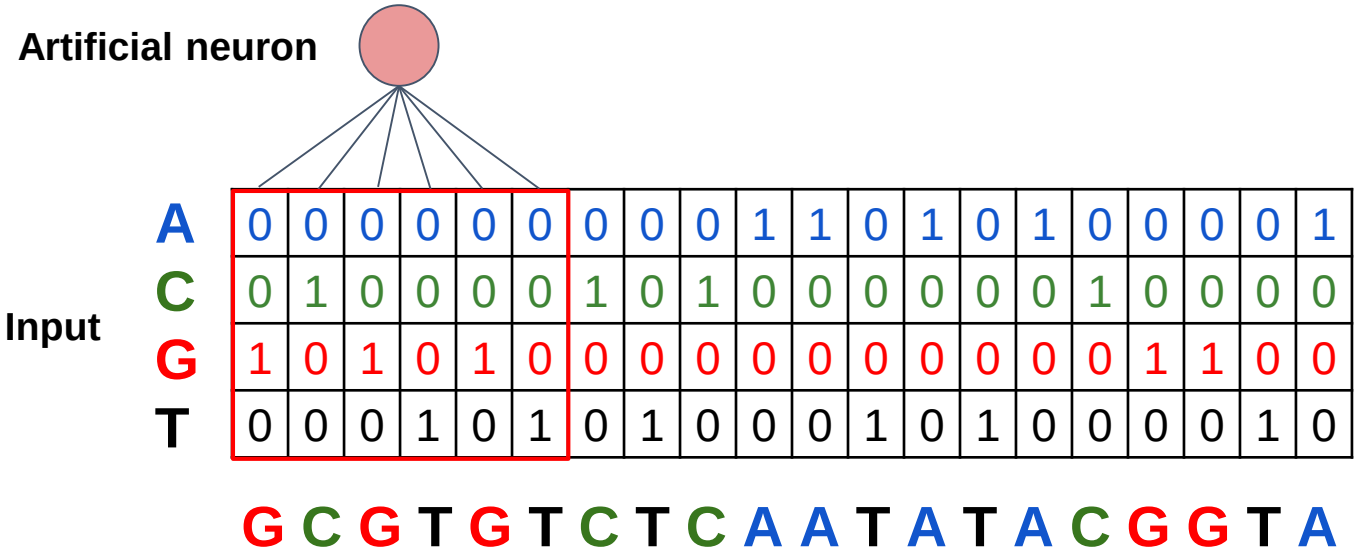
G C G T G T C T C A A T A T A C G G T A

Add more data?

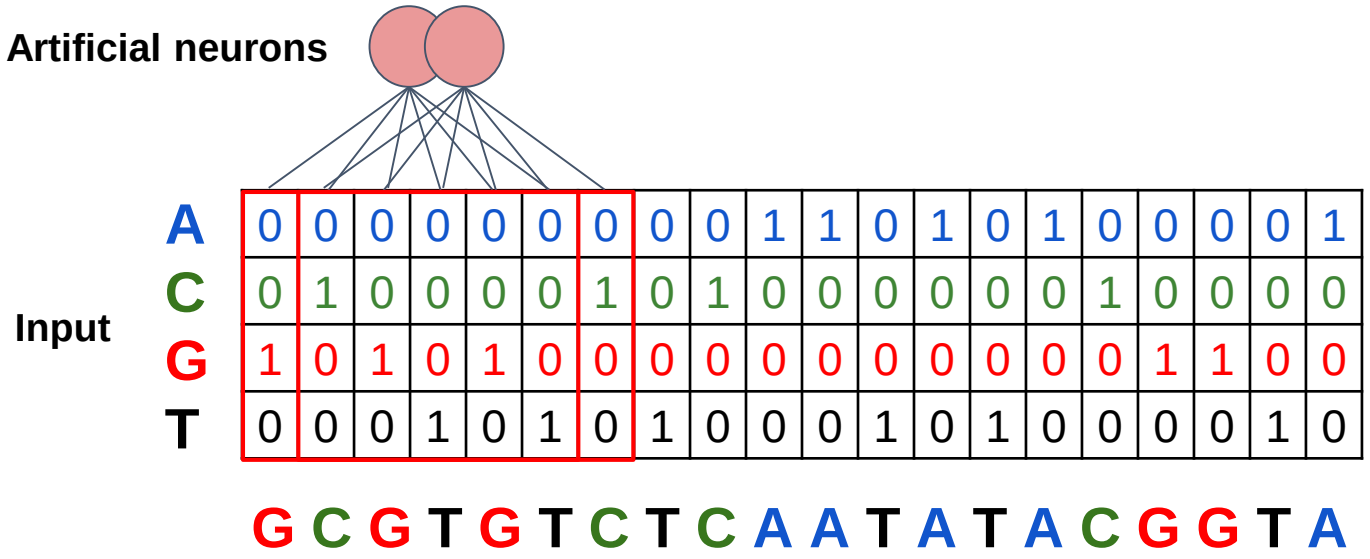
Methylation?	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0
TF binding?	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
Input **A**	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	1
C	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0
G	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
T	0	0	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	1	0

G C G T G T C T C A A T A T A C G G T A

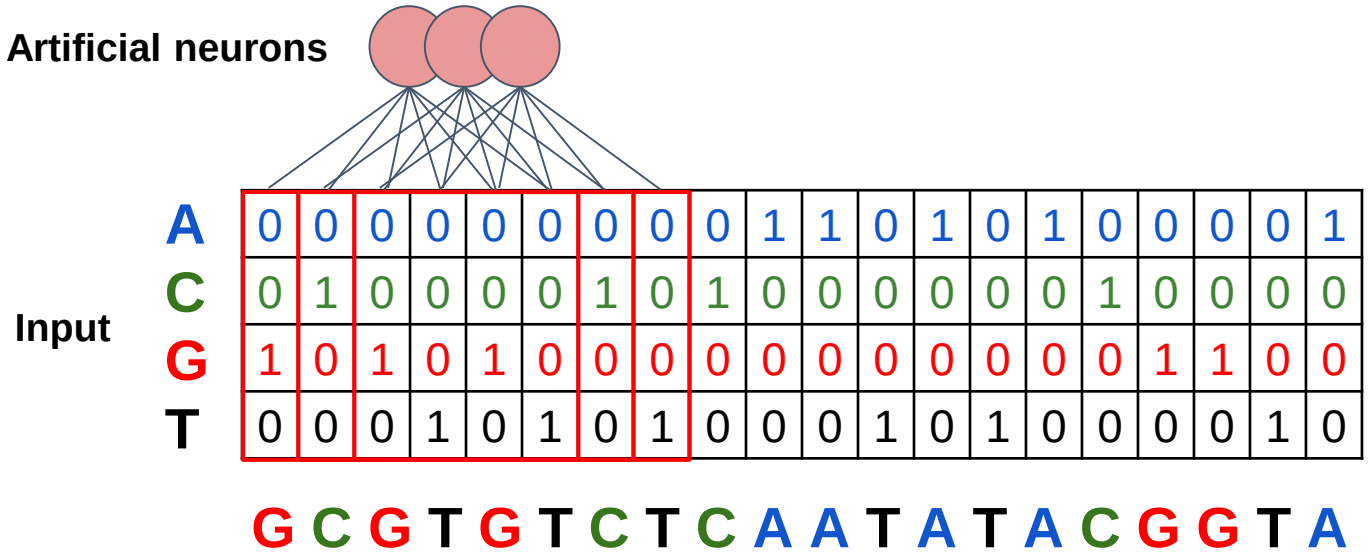
Build the artificial neural network



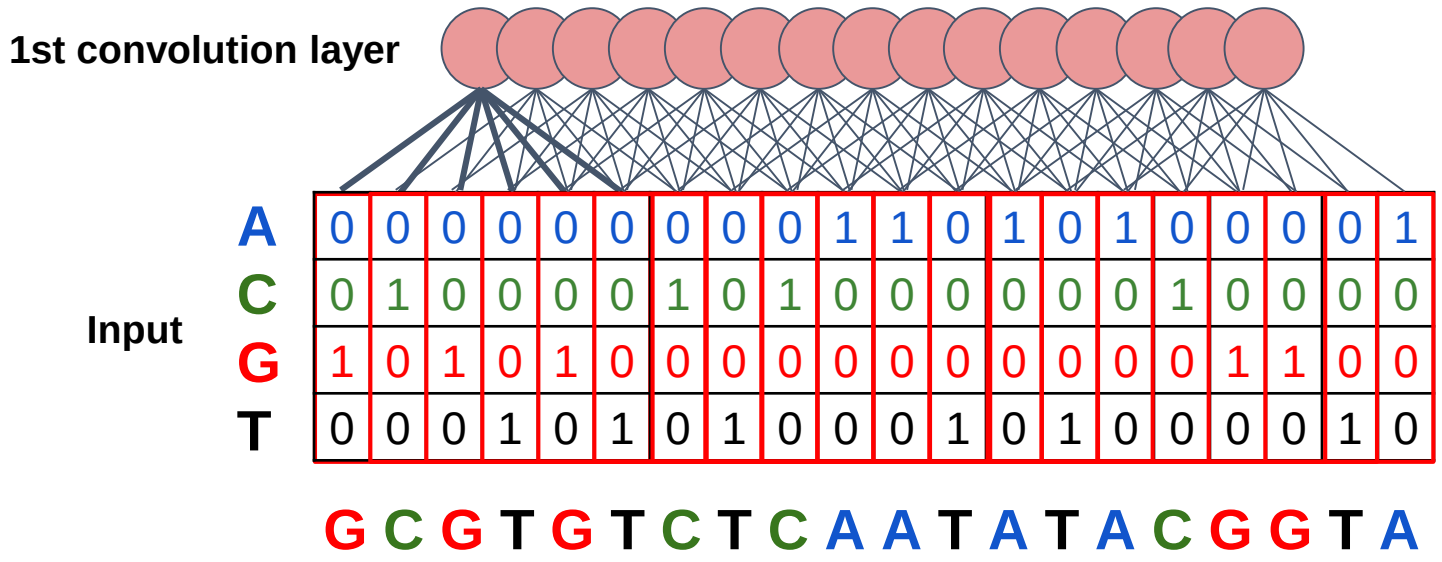
Build the artificial neural network



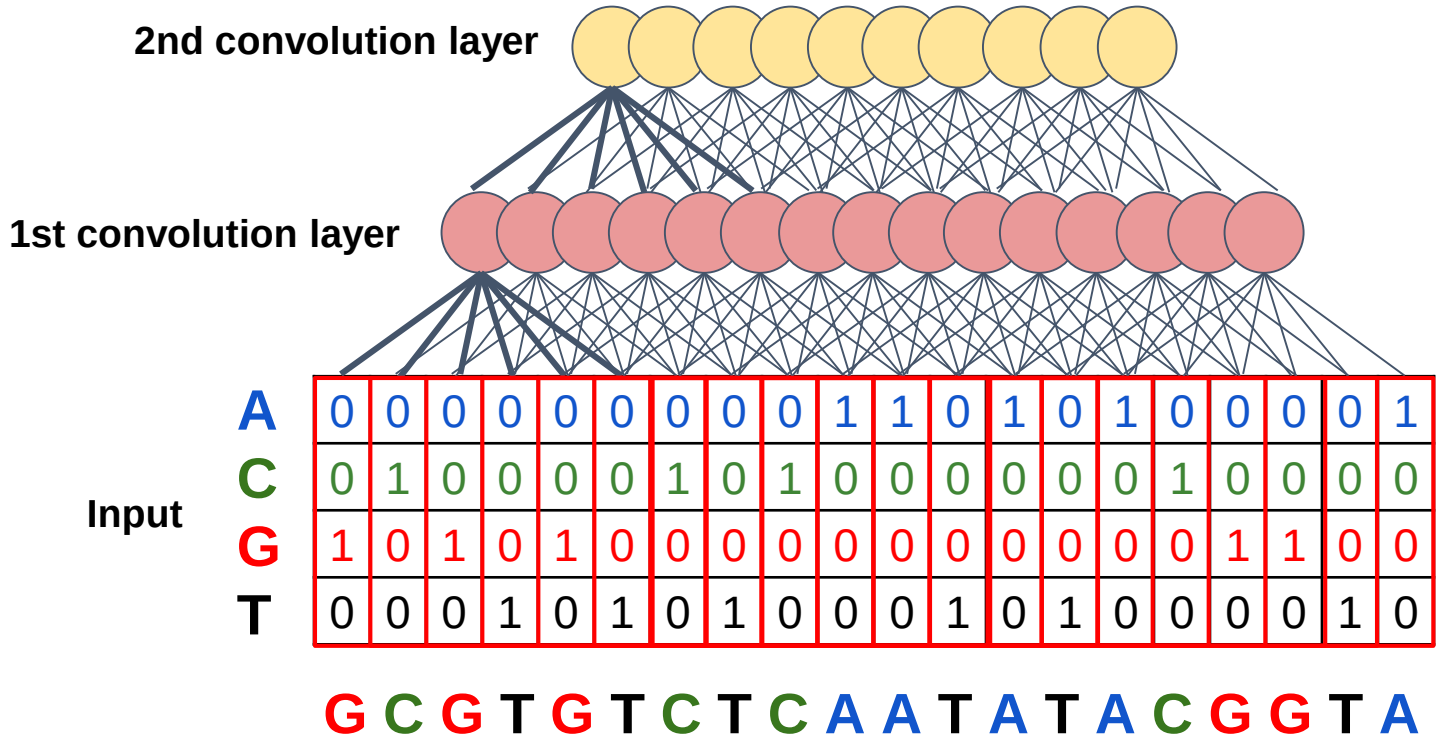
Build the artificial neural network

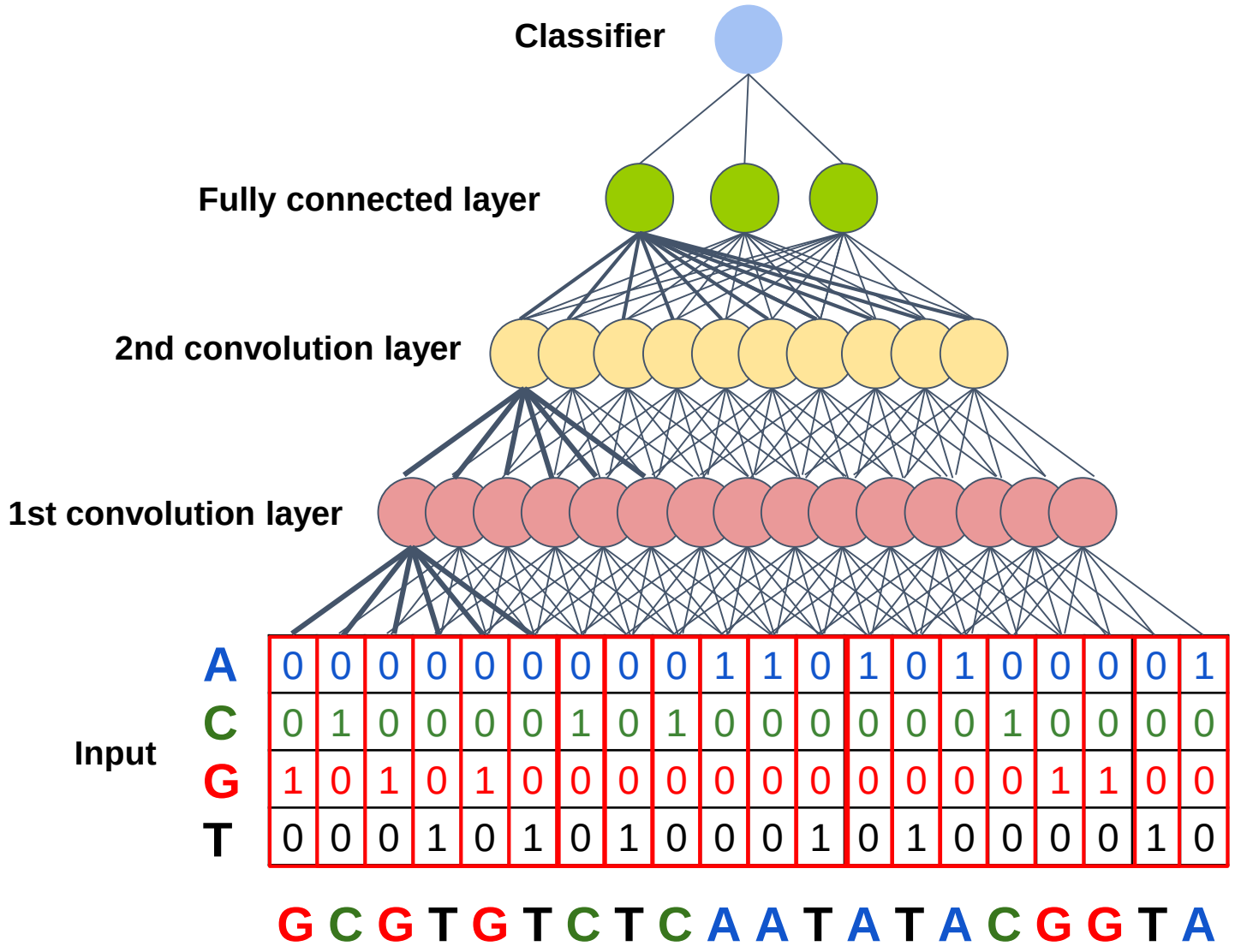


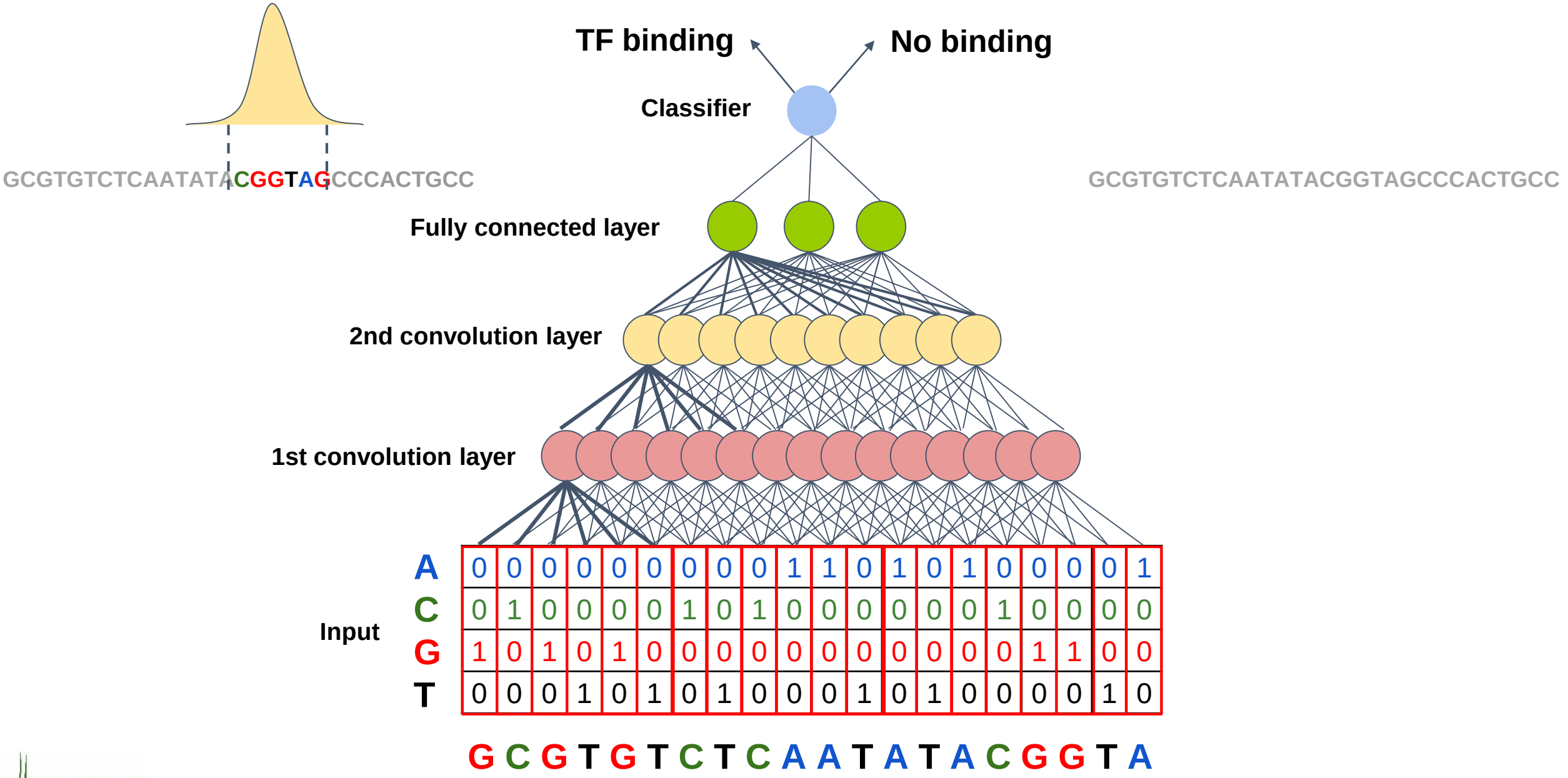
The first convolution layer

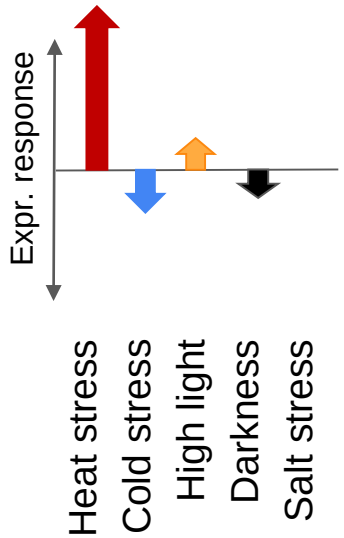


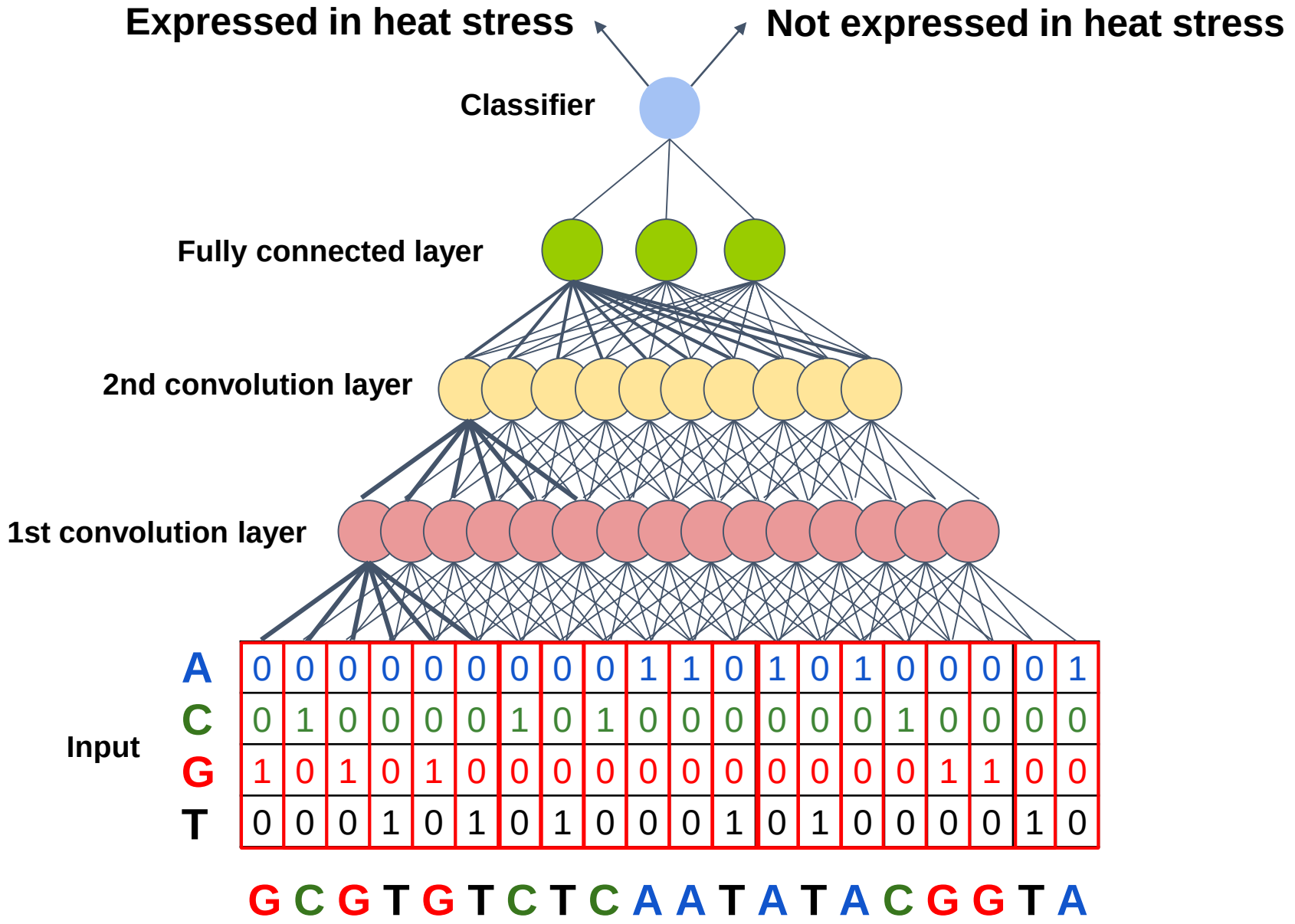
Add more convolution layers

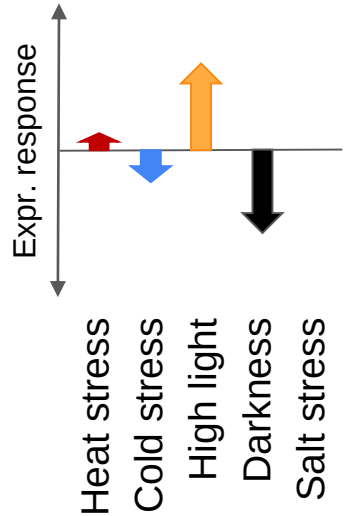


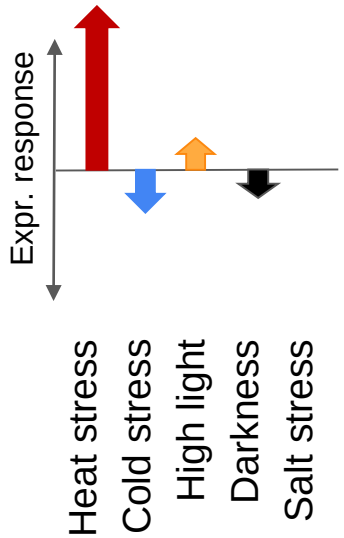

Classifier
Fully connected layer
2nd convolution layer
1st convolution layer
Input
A 0 0 0 0 0 0 0 0 0 1 1 0 1 0 1 0 0 0 0 1
C 0 1 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0
G 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0
T 0 0 0 1 0 1 0 1 0 0 0 1 0 1 0 0 0 0 1 0
G C G T G T C T C A A T A T A C G G T A

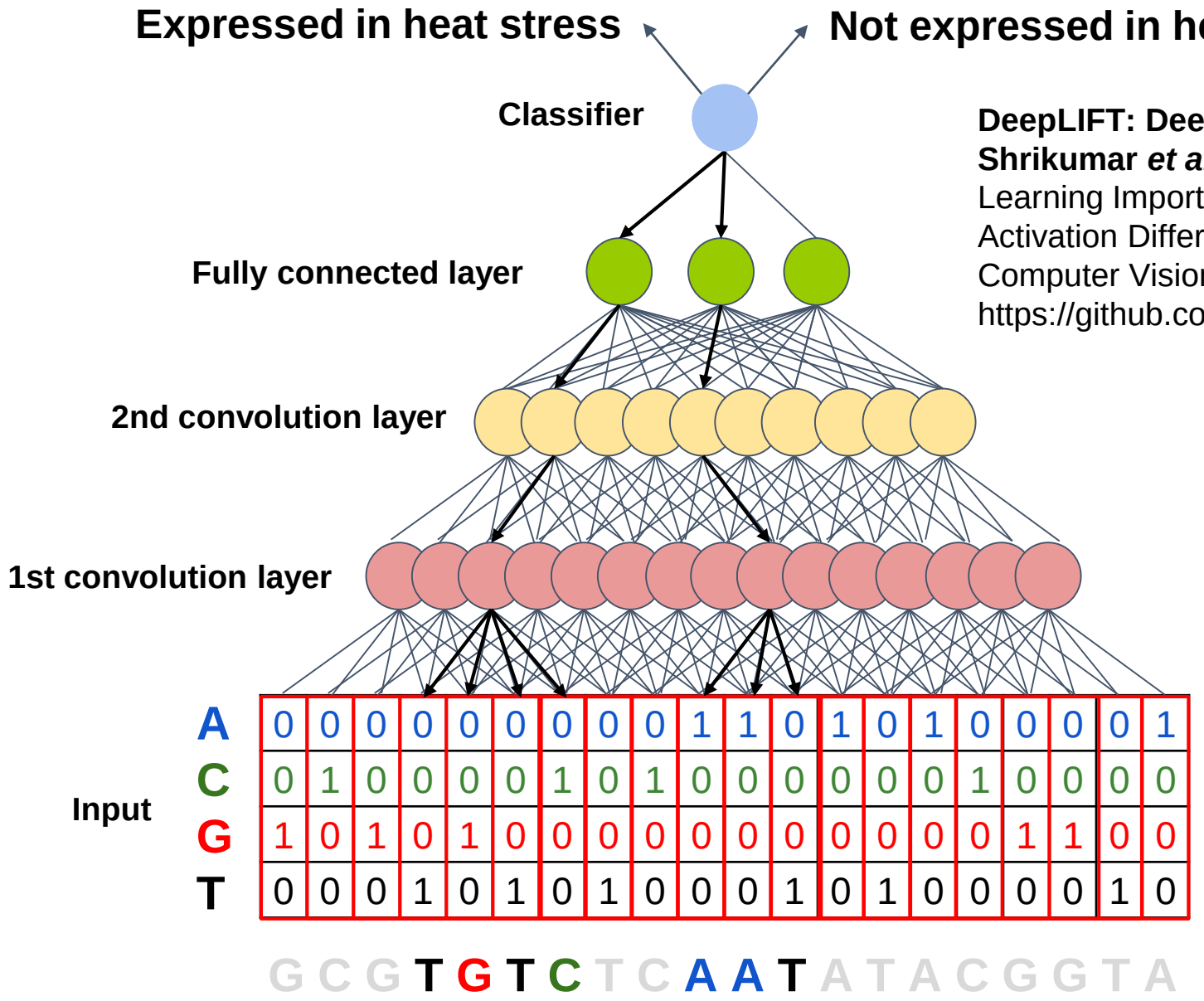

TF binding
No binding
Classifier
GCGTGTCTCAATATACGGTAGCCCACTGCC
GCGTGTCTCAATATACGGTAGCCCACTGCC
Fully connected layer
2nd convolution layer
1st convolution layer
Input
A 0 0 0 0 0 0 0 0 0 1 1 0 1 0 1 0 0 0 0 1
C 0 1 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0
G 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0
T 0 0 0 1 0 1 0 1 0 0 0 1 0 1 0 0 0 0 1 0
G C G T G T C T C A A T A T A C G G T A


Expr. response
Heat stress
Cold stress
High light
Darkness
Salt stress


Expressed in heat stress
Not expressed in heat stress
Classifier
Fully connected layer
2nd convolution layer
1st convolution layer
Input
A 0 0 0 0 0 0 0 0 0 1 1 0 1 0 1 0 0 0 0 1
C 0 1 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0
G 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0
T 0 0 0 1 0 1 0 1 0 0 0 1 0 1 0 0 0 0 1 0
G C G T G T C T C A A T A T A C G G T A

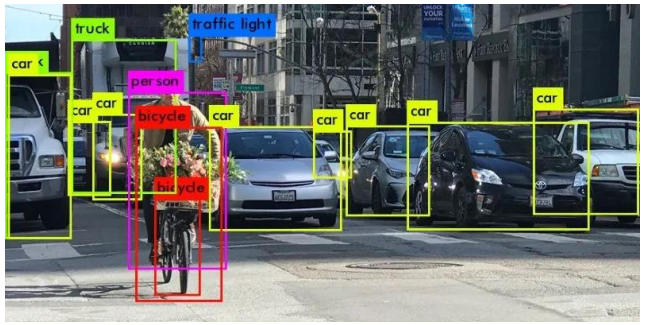

Expr. response
Heat stress
Cold stress
High light
Darkness
Salt stress





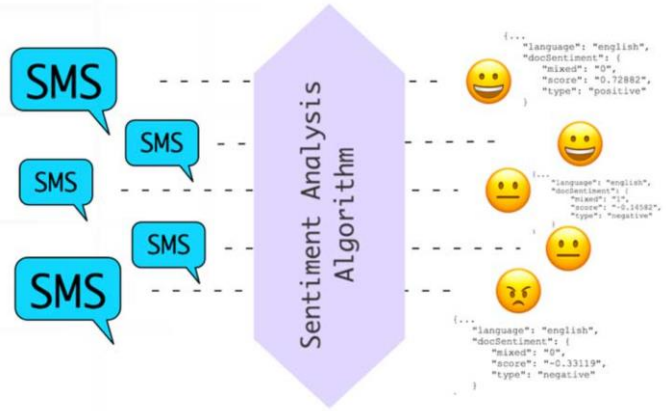
DeepLIFT: Deep Learning Important FeaTures
Shrikumar *et al.* 2019
Learning Important Features Through Propagating Activation Differences
Computer Vision and Pattern Recognition
https://github.com/kundajelab/deeplift

Implementation

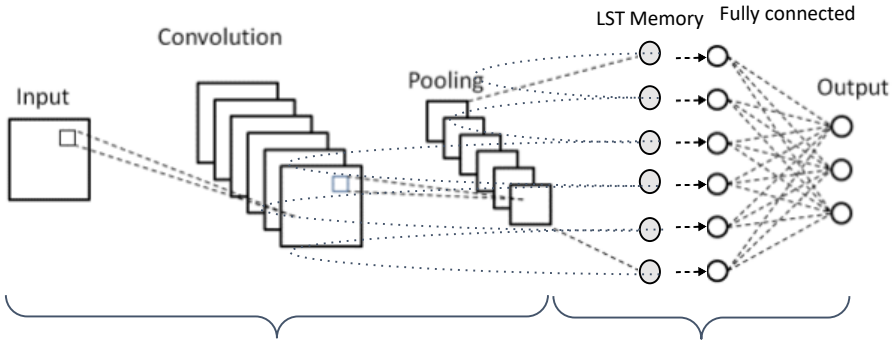


CNN-LSTM

Convolutional Neural Networks - Long Short Term Memory



gene sequences



expression values

Arabidopsis thaliana

Solanum lycopersicum

Sorghum bicolor

Zea mays

Implementation

Can we detect cis-regulatory sequence features predictive for gene expression?

Simon Zumkeller

Fritz Peleke

Arabidopsis thaliana

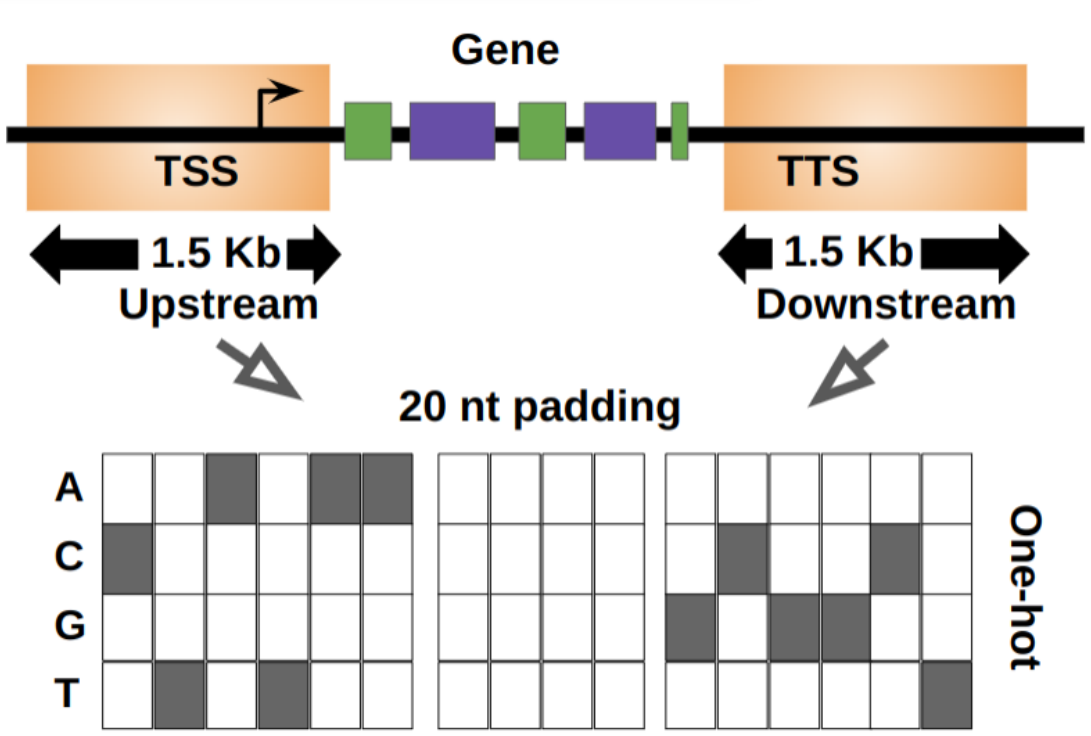
Solanum lycopersicum

Sorghum bicolor

Zea mays

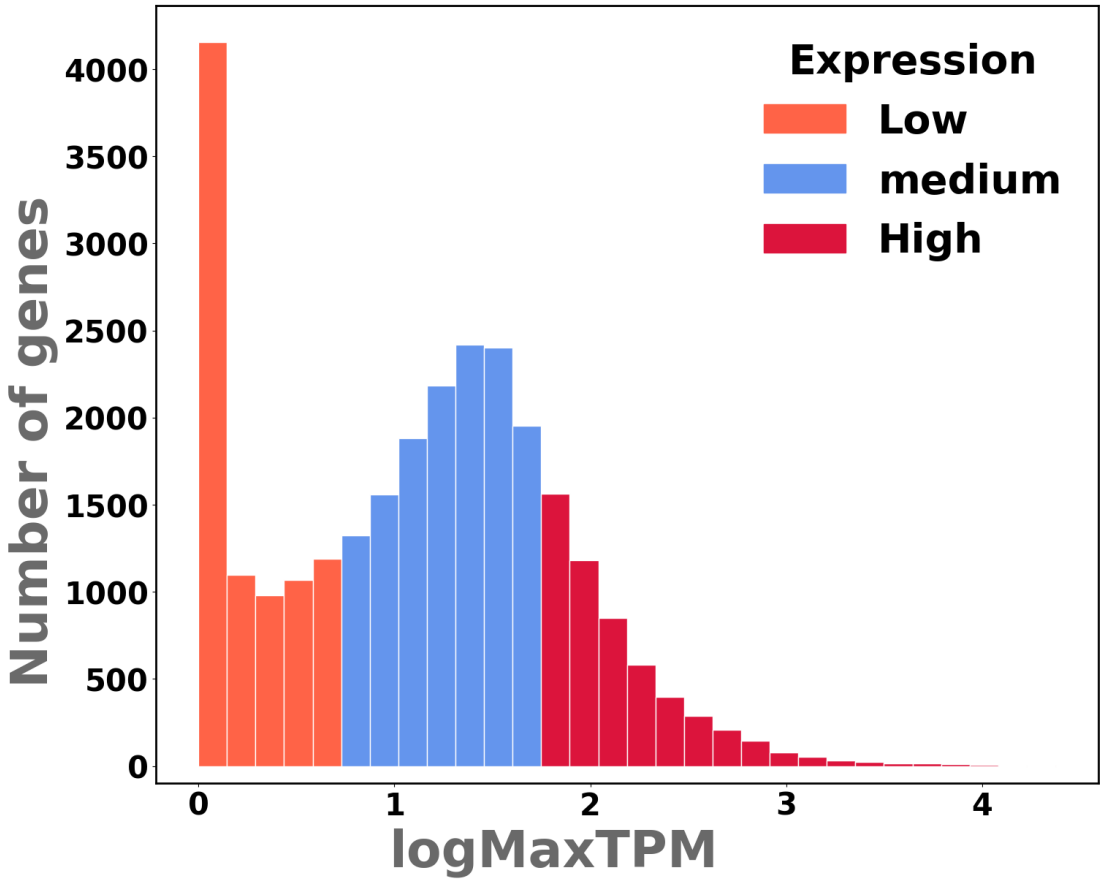
Input data

Can we detect cis-regulatory sequence features predictive for gene expression?

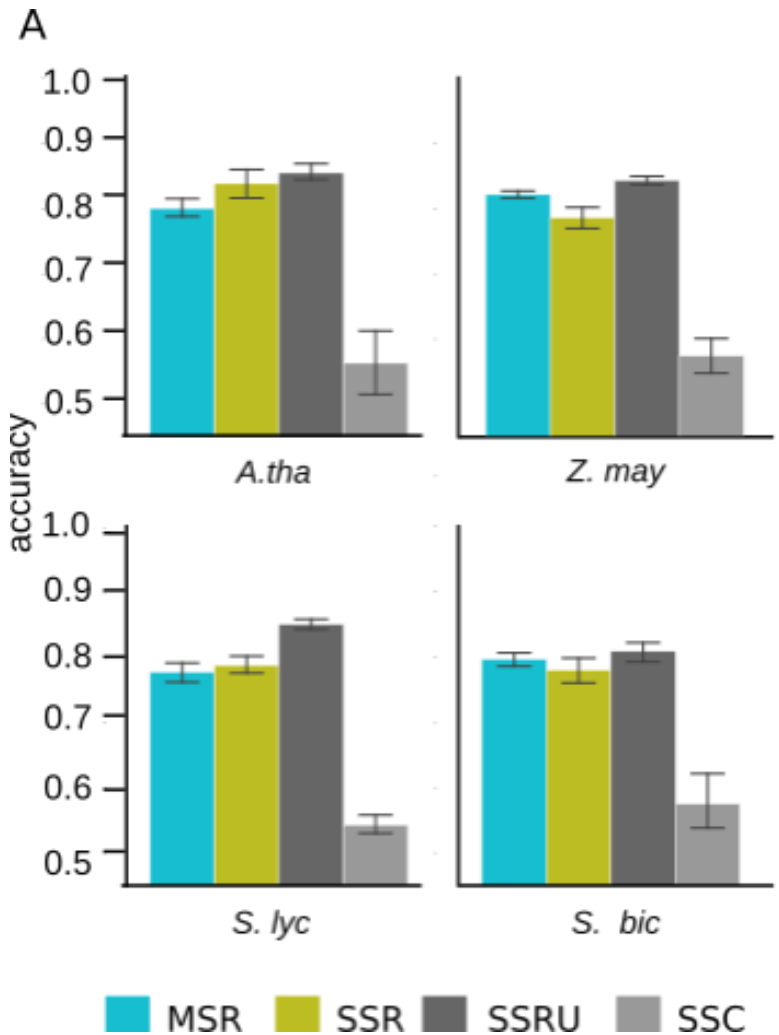


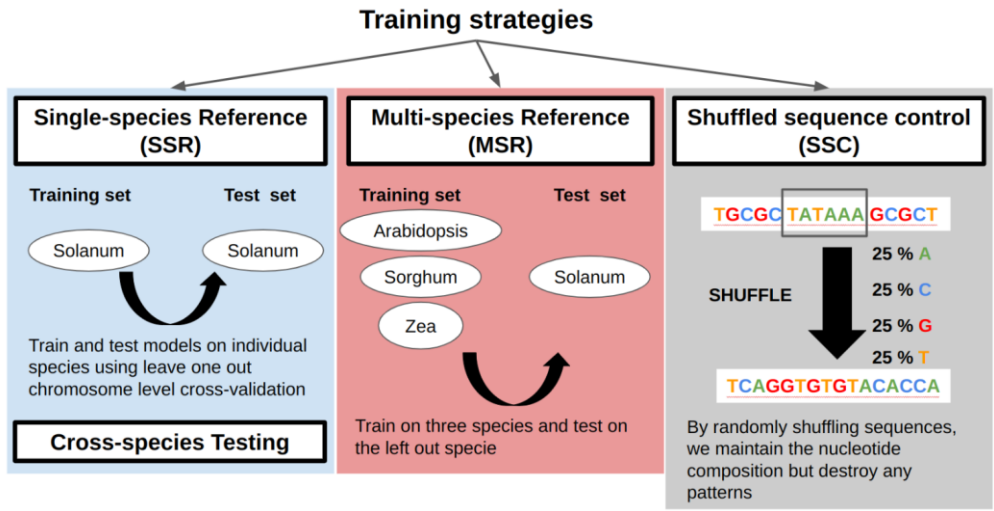
Input data

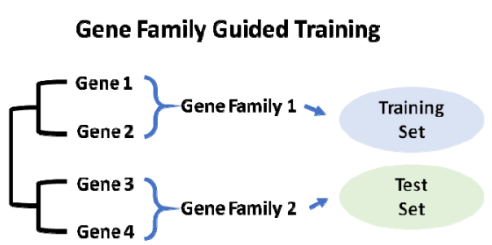
Can we detect cis-regulatory sequence features predictive for gene expression?

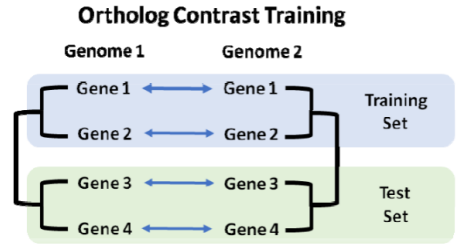


Prediction



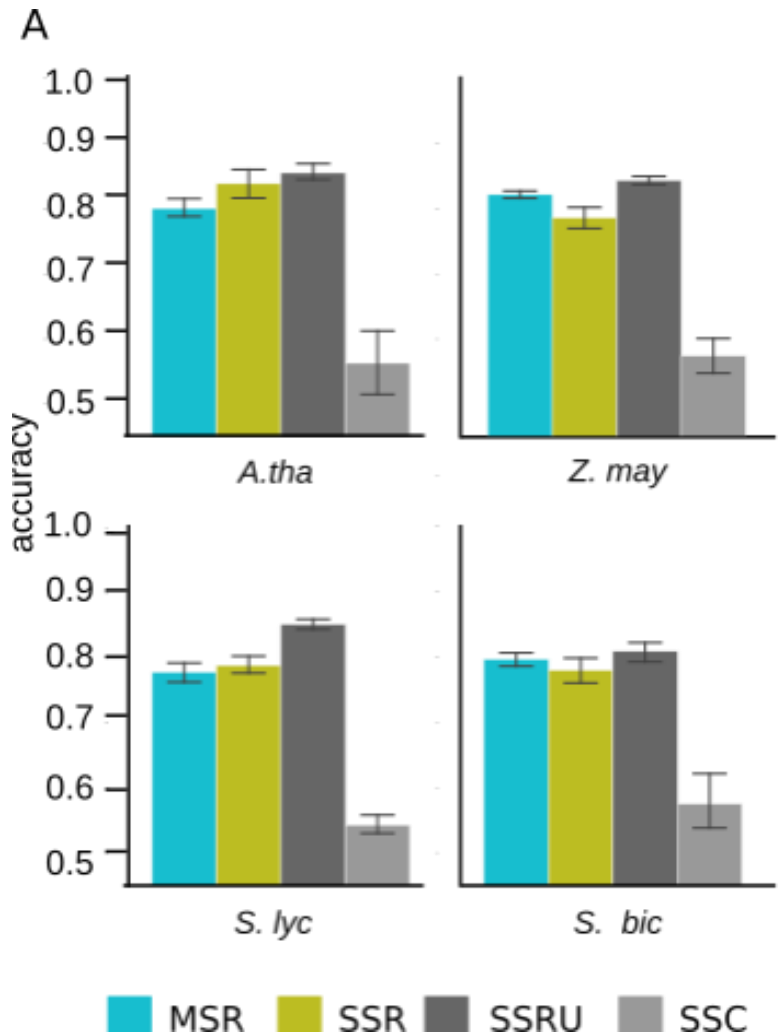


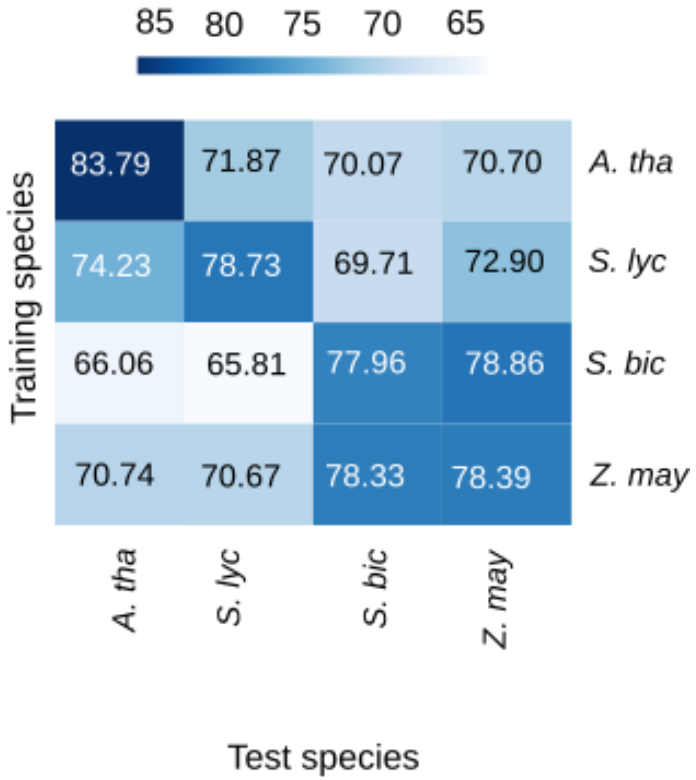




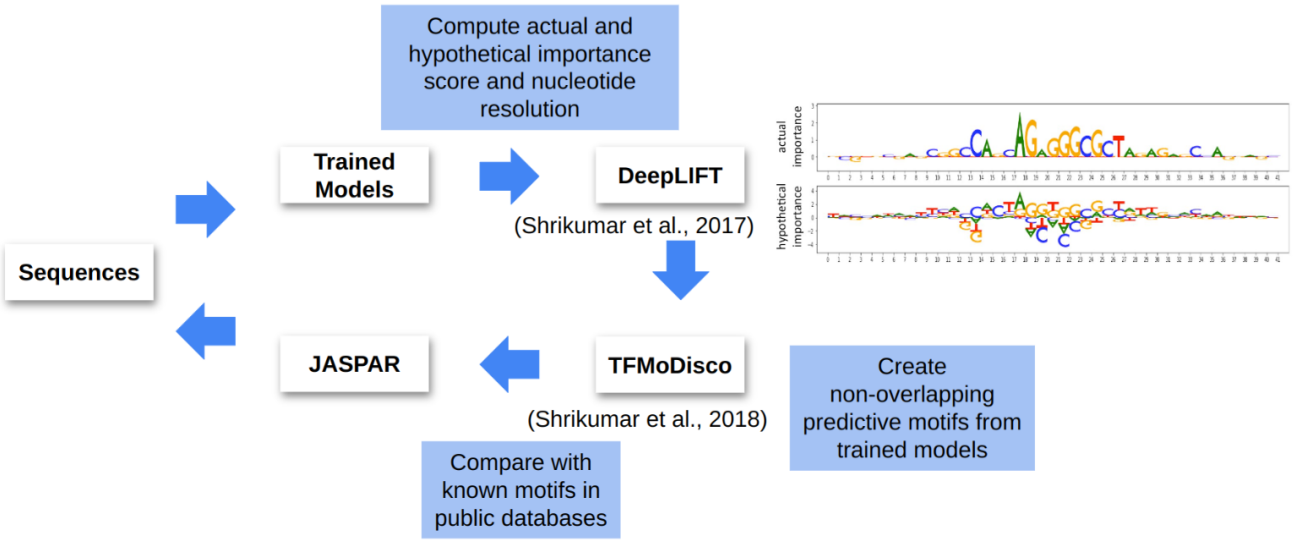
Washburn *et al.* 2019 *PNAS*

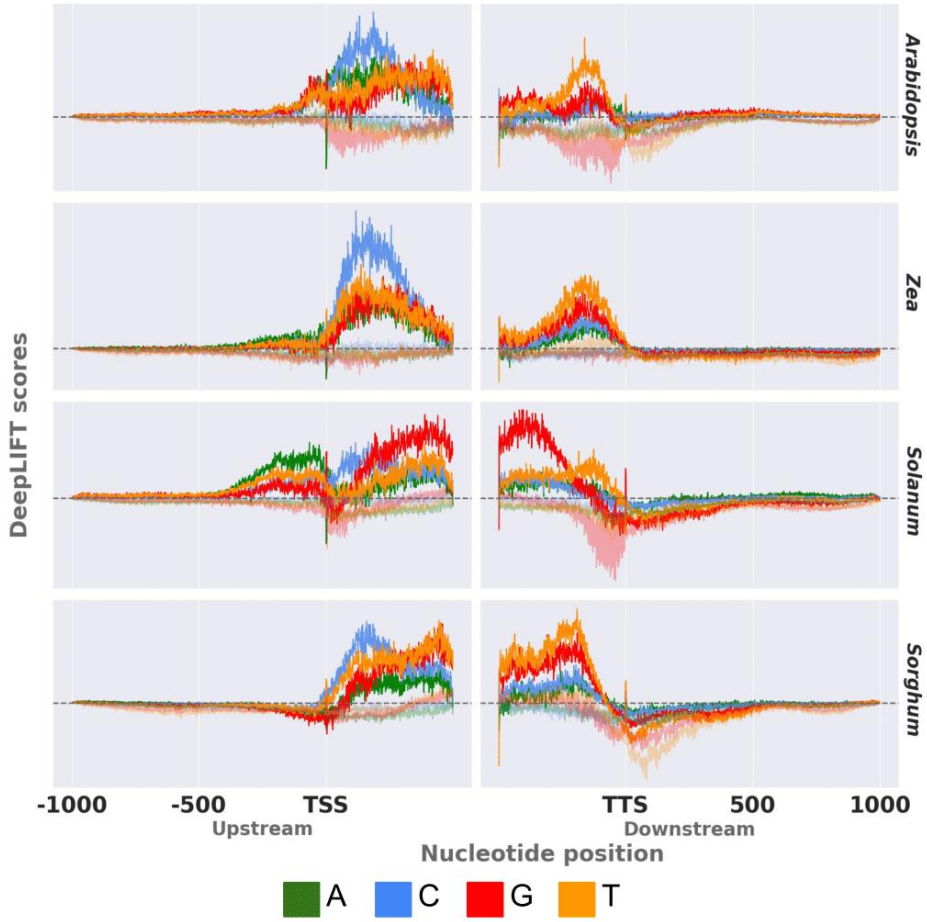
Prediction


A
1.0
0.9
0.8
0.7
0.6
0.5
accuracy
A.tha
Z. may
S. lyc
S. bic
MSR
SSR
SSRU
SSC

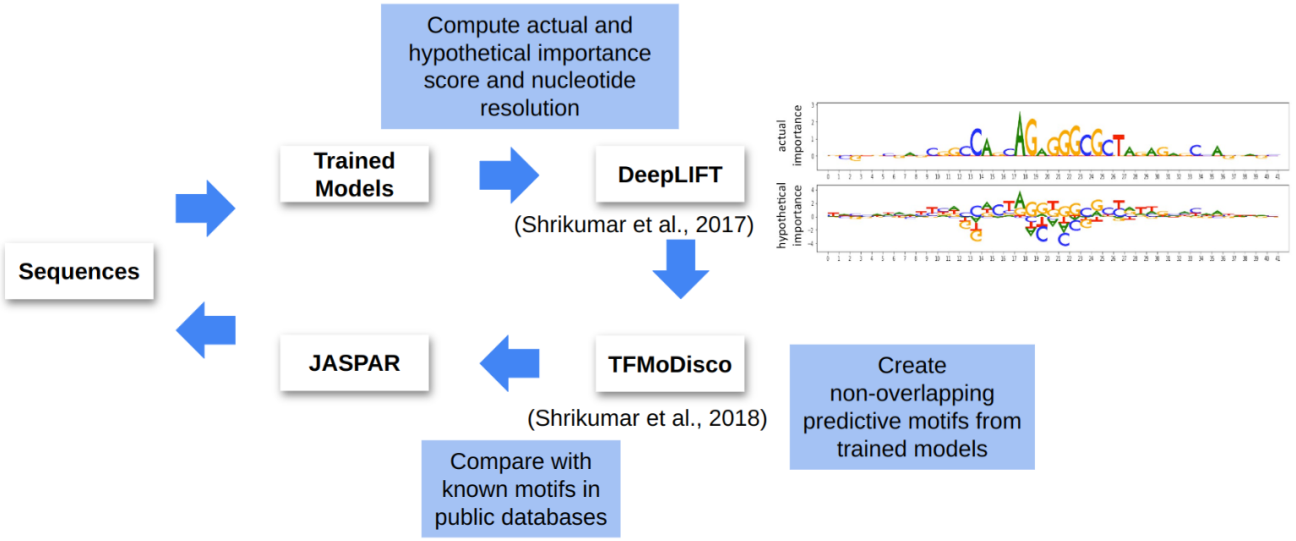

85 80 75 70 65
Training species
83.79 71.87 70.07 70.70 A. tha
74.23 78.73 69.71 72.90 S. lyc
66.06 65.81 77.96 78.86 S. bic
70.74 70.67 78.33 78.39 Z. may
A. tha
S. lyc
S. bic
Z. may
Test species

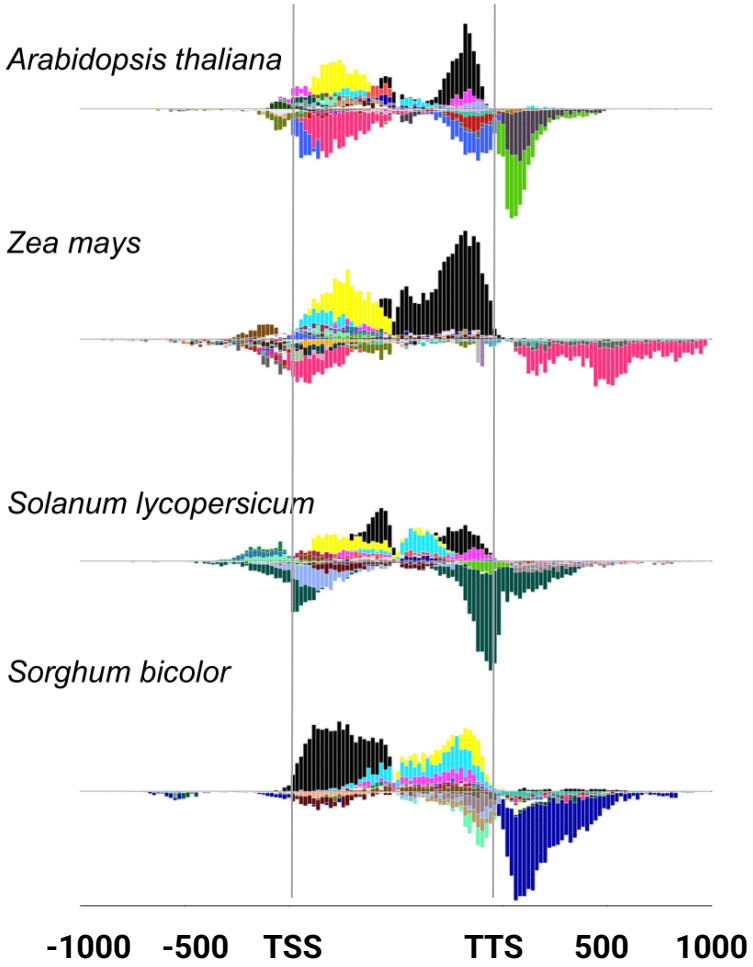
Feature selection





Feature selection





261 unique EPMs

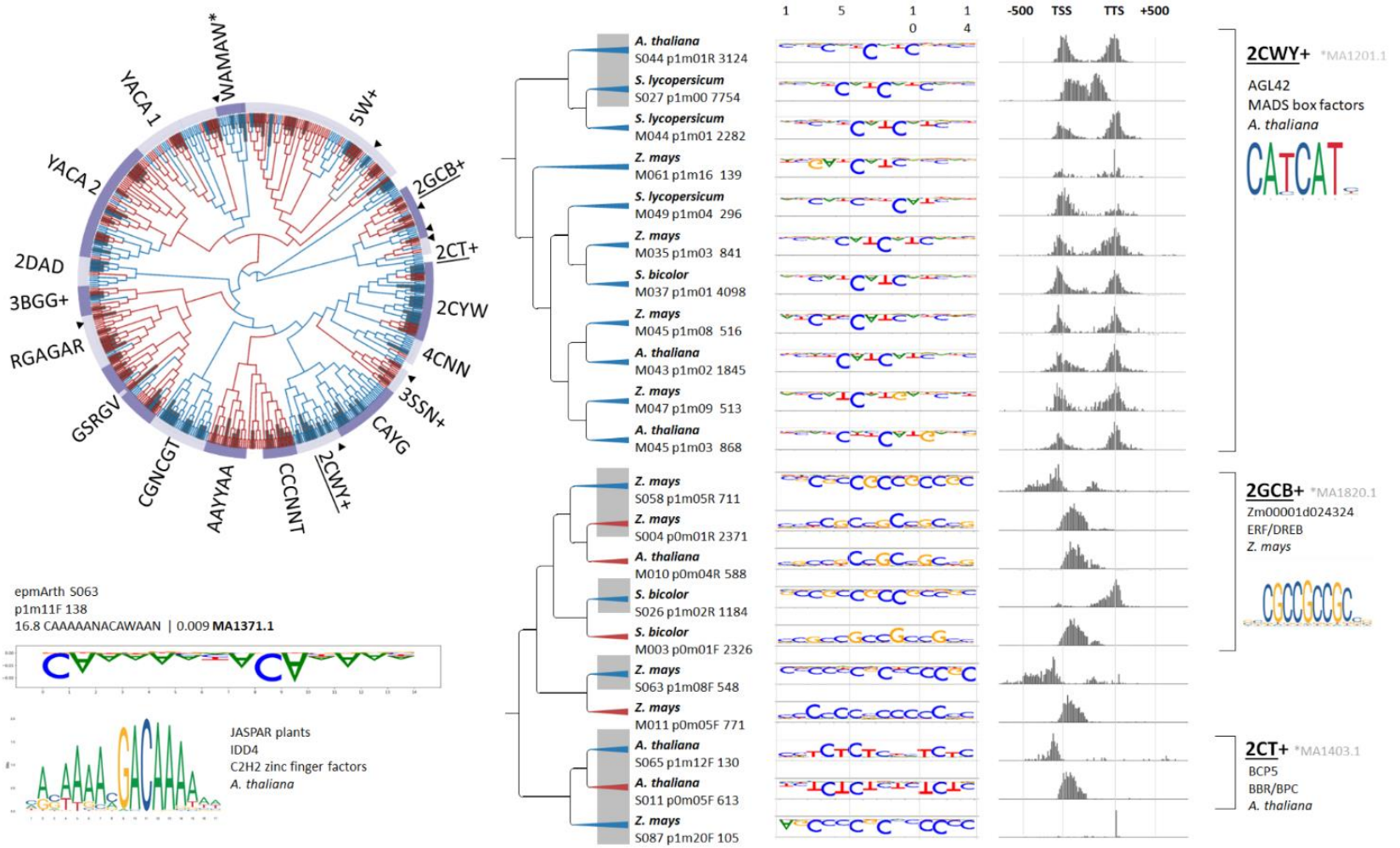
35 SSR / 28 MSR
Arabidopsis thaliana

29 SSR / 34 MSR
Sorghum bicolor

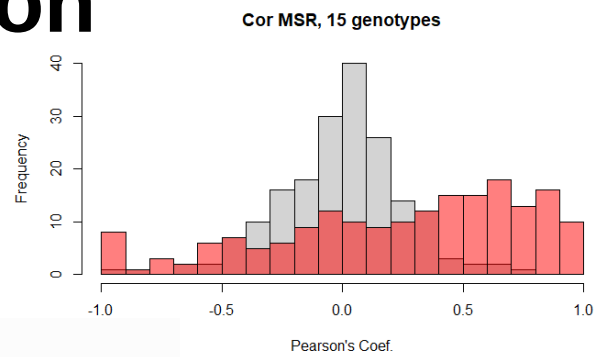
26 SSR / 30 MSR
Solanum lycopersicum

46 SSR / 33 MSR
Zea mays

Expression Predictive Motifs (EPM)

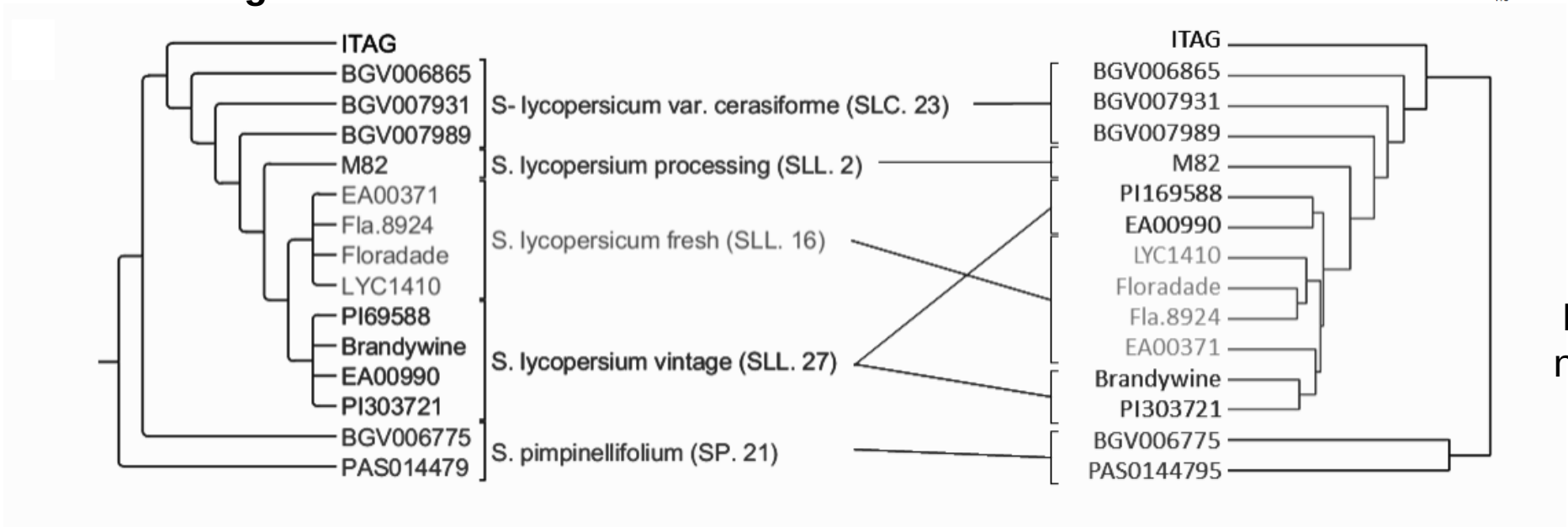


Annotate effects of genetic variation



Original data

Predicted data

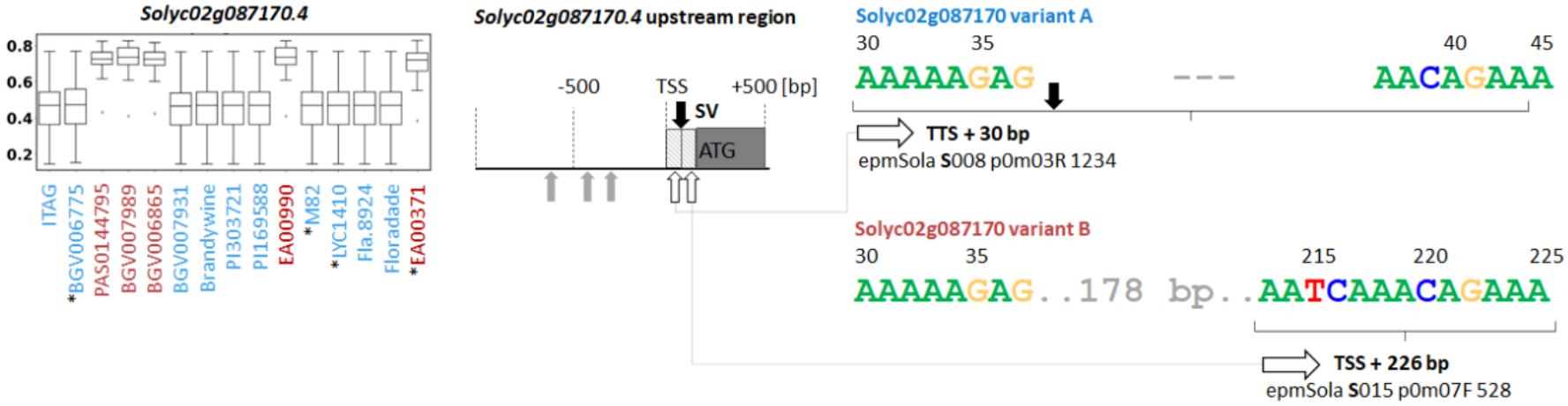


Prediction accuracies **>75%**

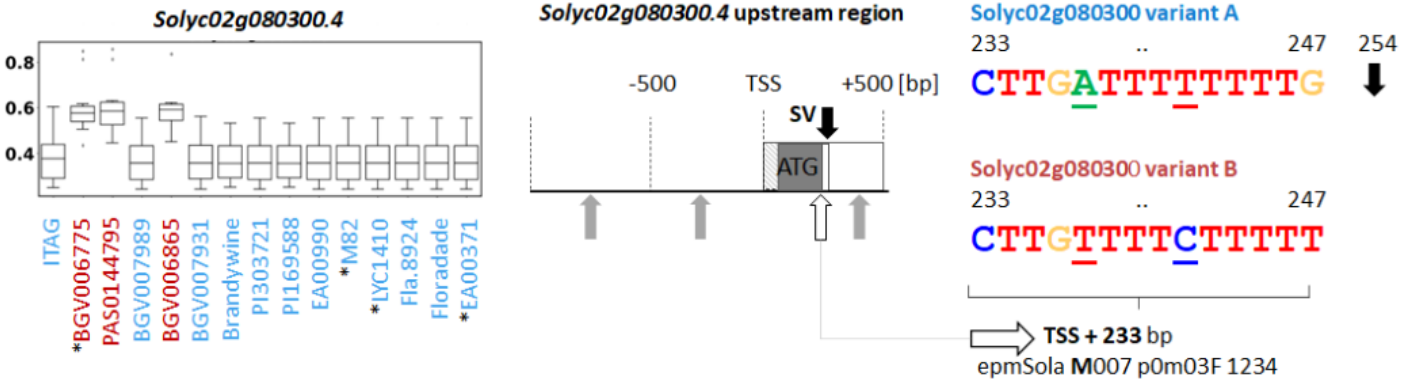
MSR model that never saw tomato

Alonge *et al* 2020, Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell*. **182**, 145–161.e23 (2020).

Annotate effects of genetic variation

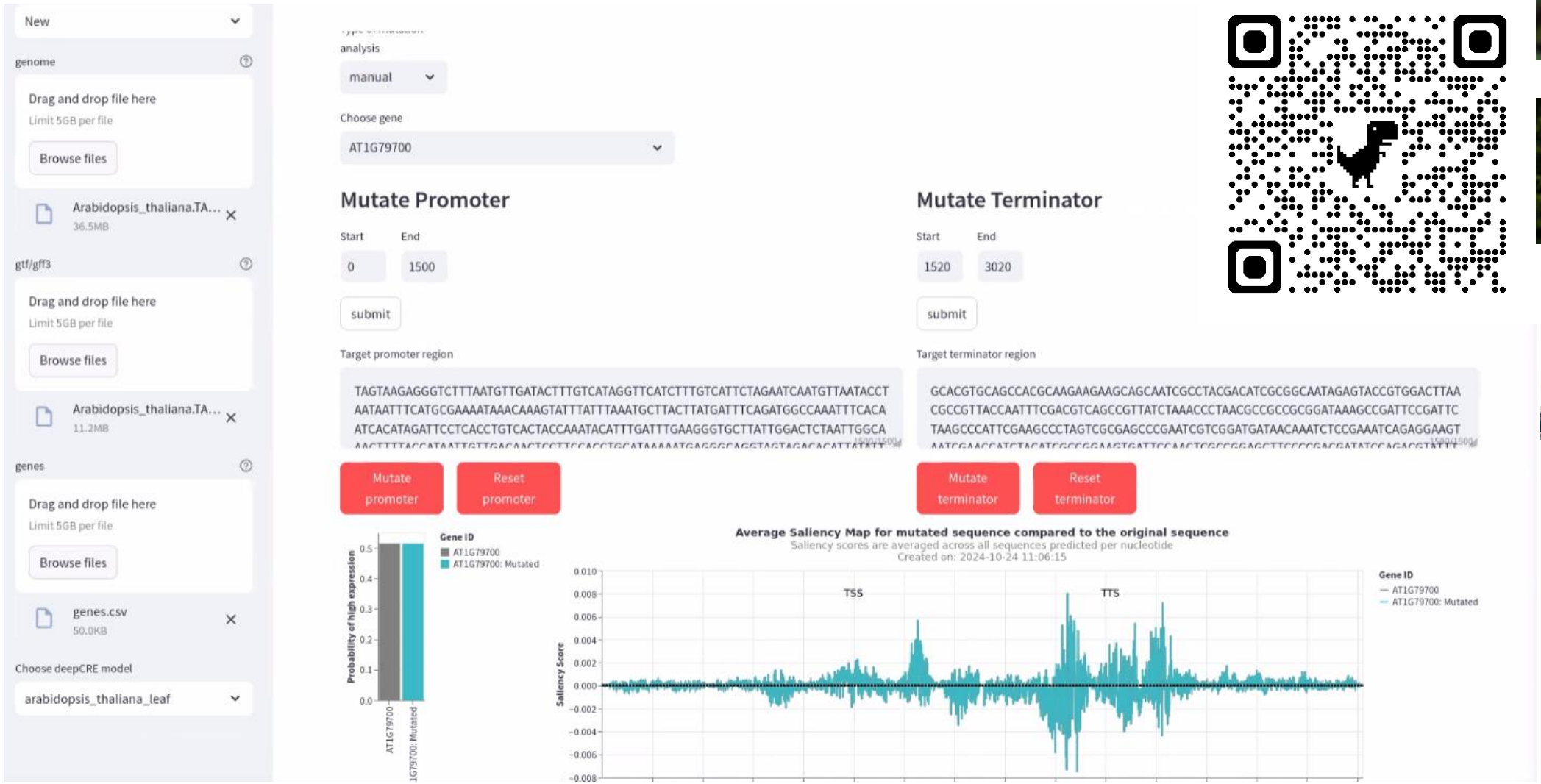


Introduction of functional EPM



Edit from neutral to functional EPM

DeepCRE - framework and toolbox: https://deepcre.ipk-gatersleben.de/



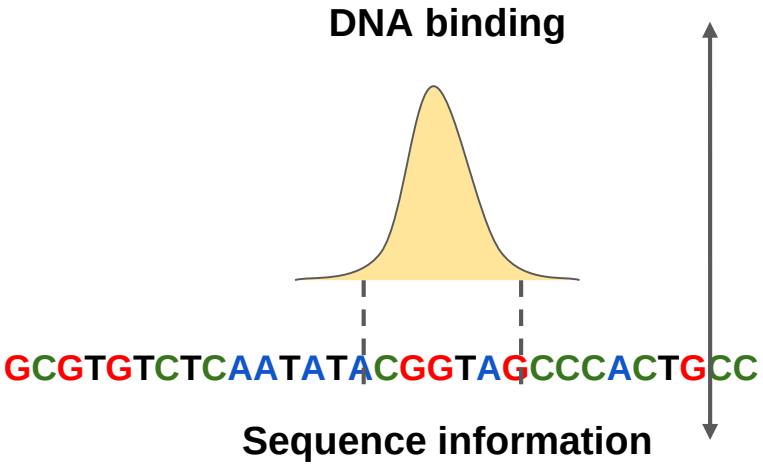
Merle Stein

Gernot Schmitz

Fritz Peleke

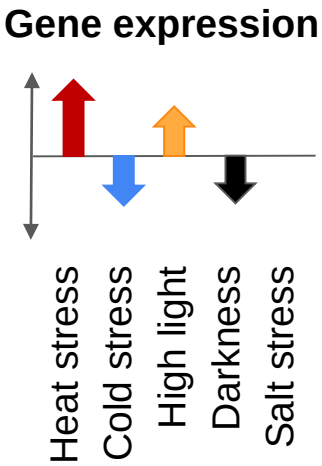
The task

> 500k DNA binding events



?

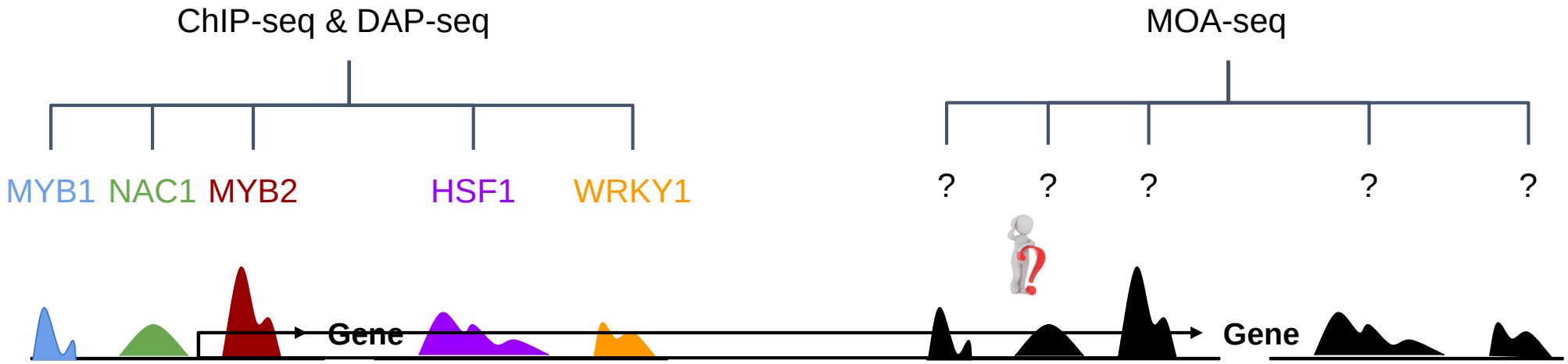
> 400 bln expression values



> 3 mln gene sequences

without accessions & cultivars

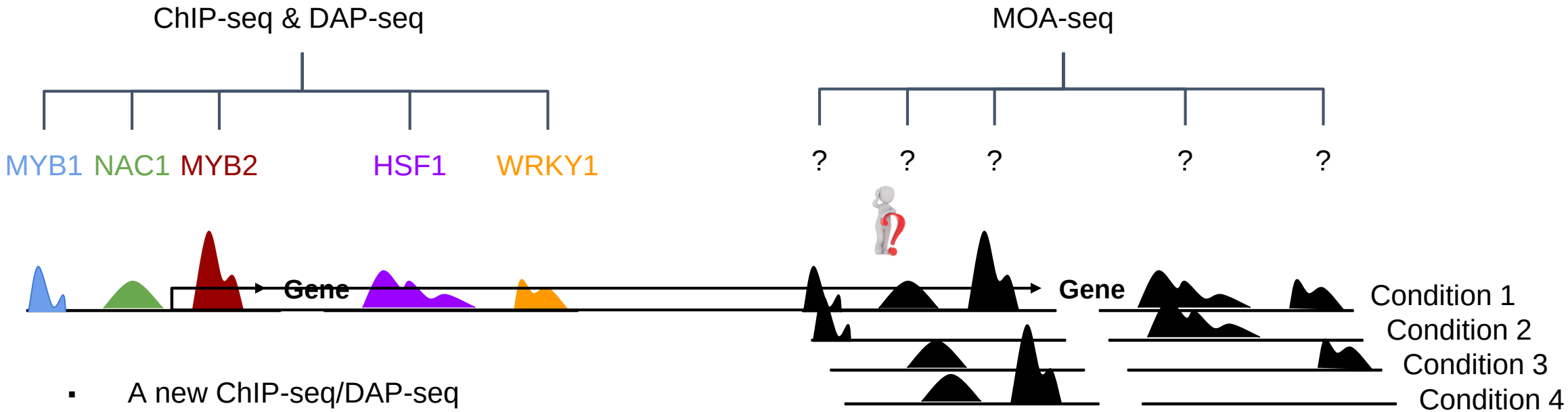
Beyond gene expression



- A new ChIP-seq/DAP-seq experiment for every new TF
- Costly

- An experiment for all TFs
- Not sure what the peaks are

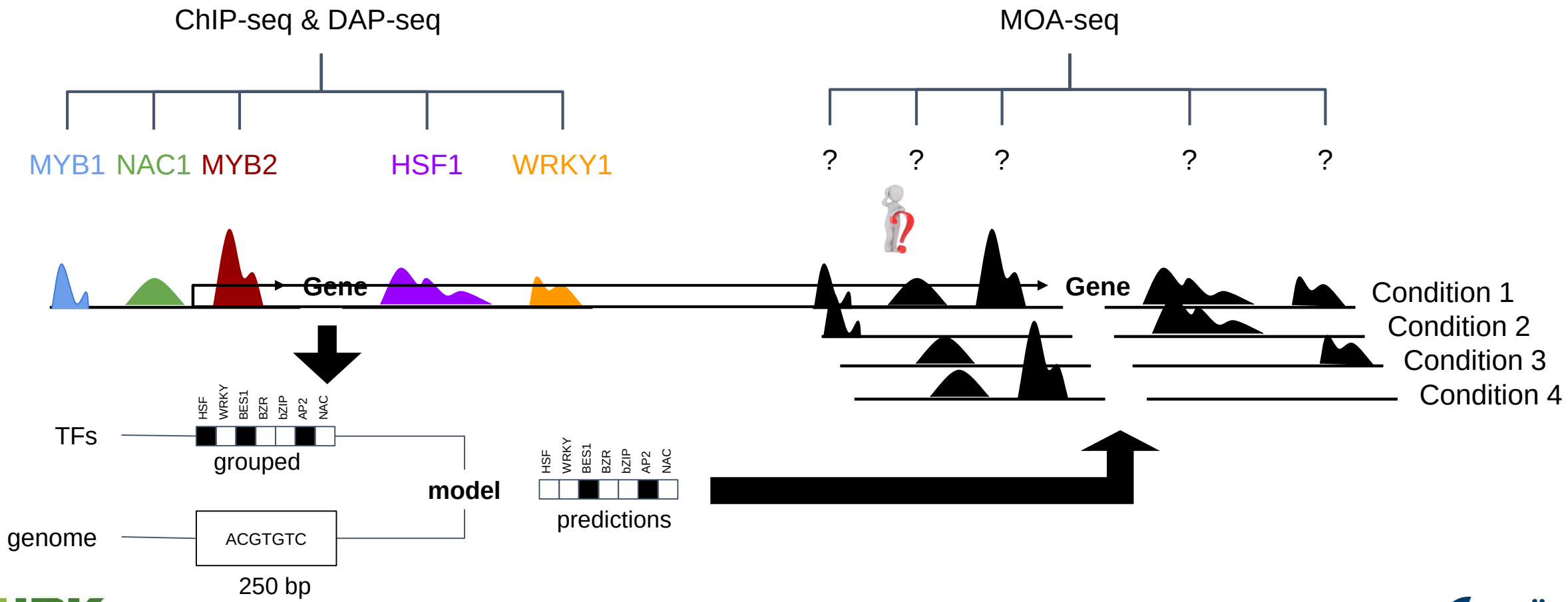
Beyond gene expression



- A new ChIP-seq/DAP-seq experiment for every new TF
- Costly

- We can do a lot of it

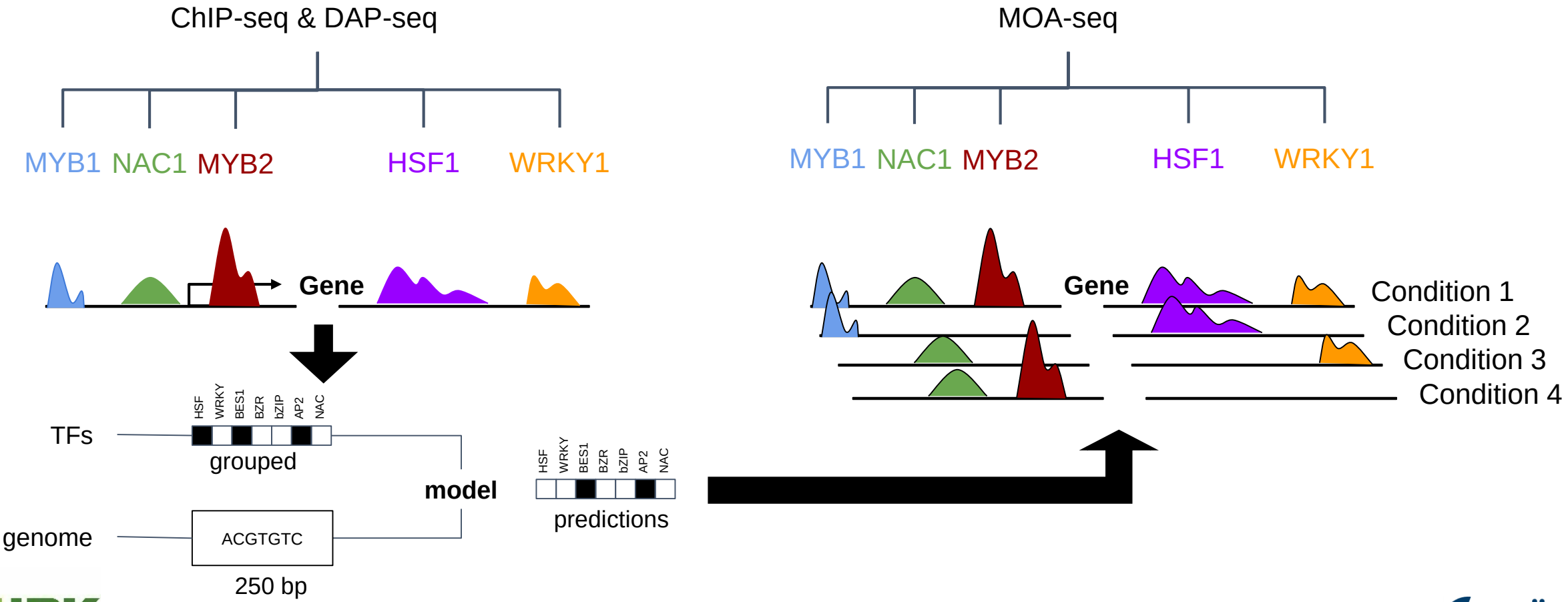
Beyond gene expression



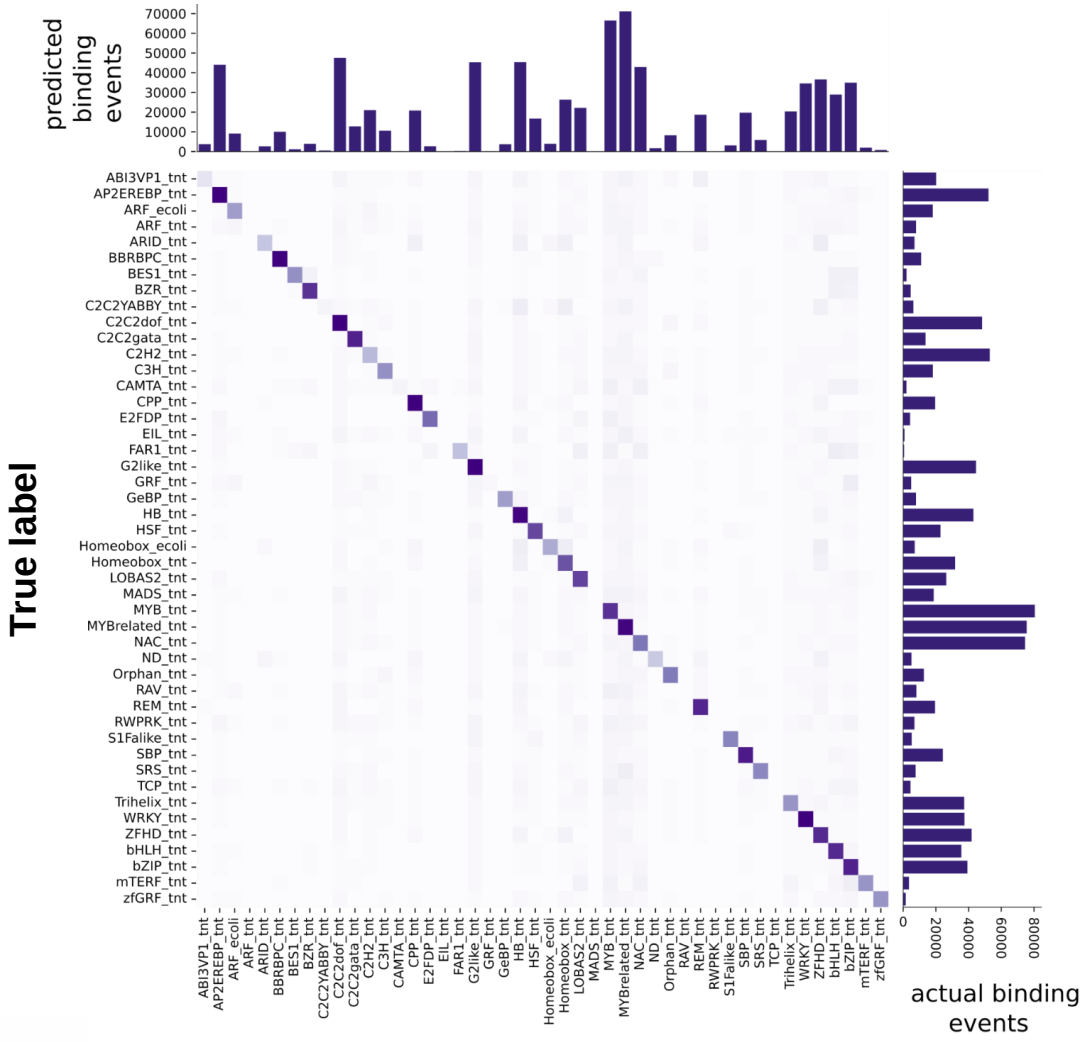
Beyond gene expression

Simon Zumkeller

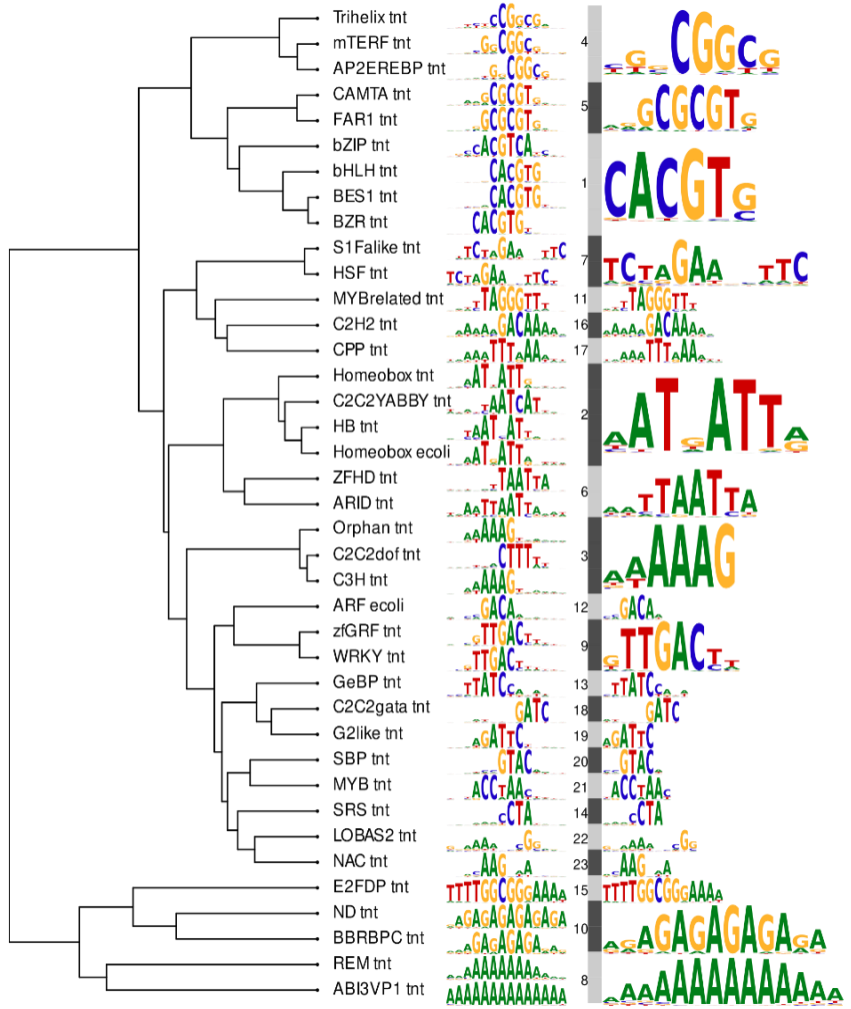
Fritz Peleke



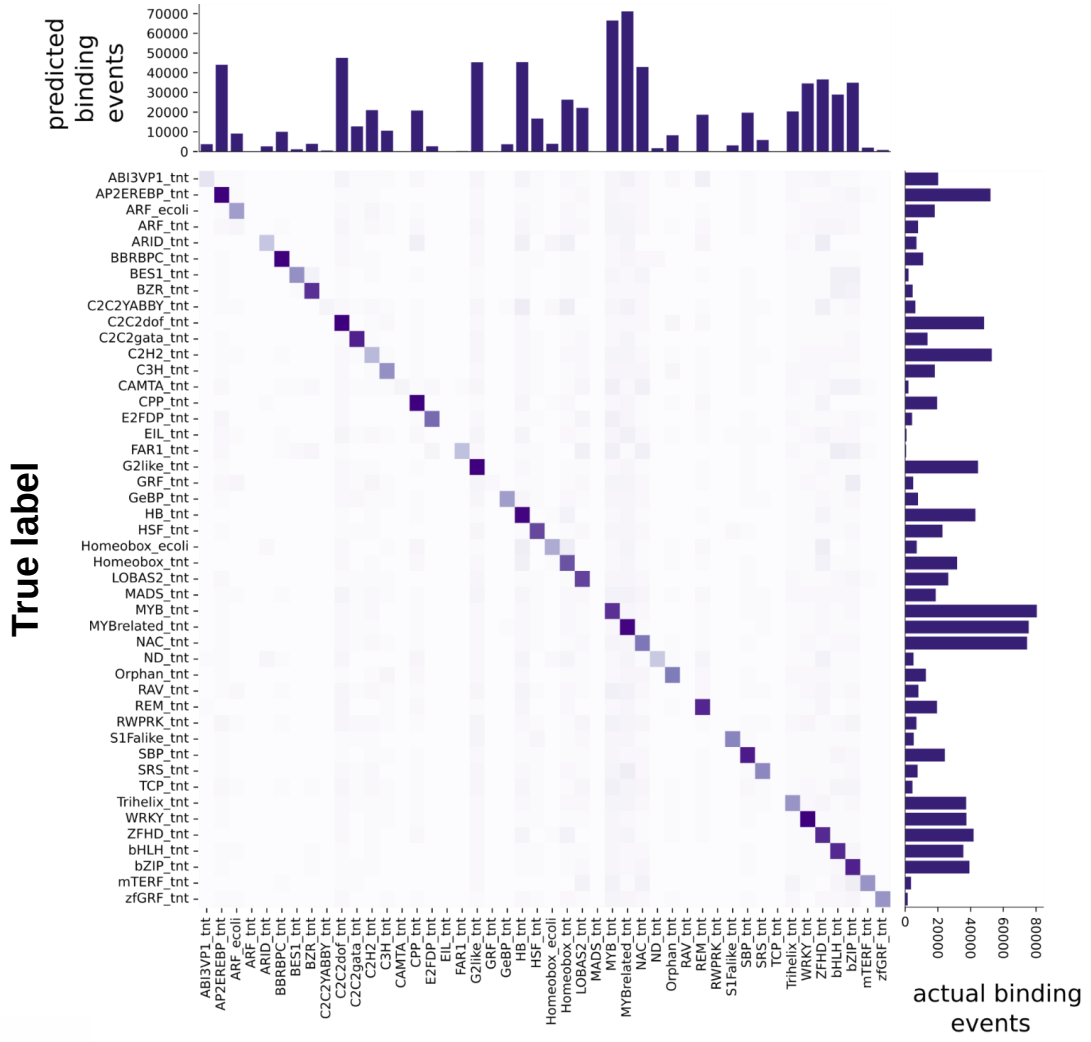
DNA binding predition



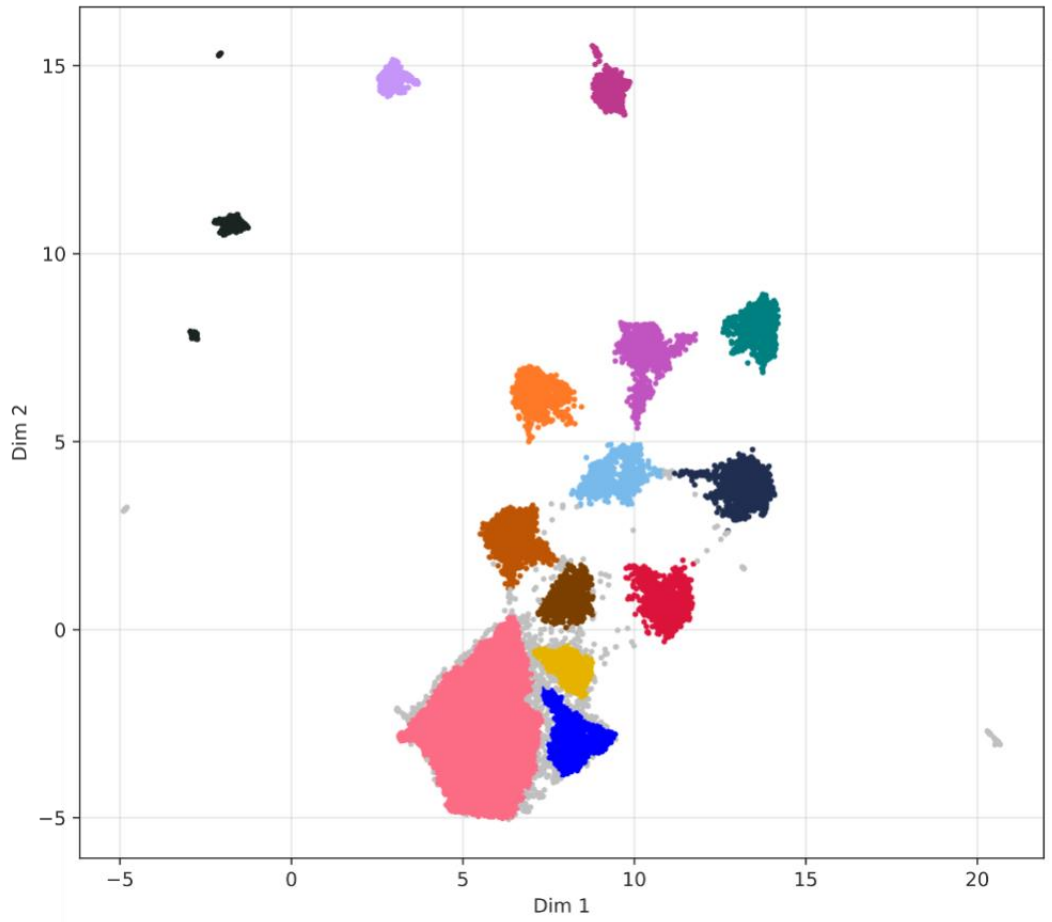
Interpretaion



DNA binding predition



Gene regulatory profiles



From genetic variation to phenotypic traits

1001 Genomes

AraGWAS Catalog 36878 phenotype associations

SNP / SV

↓

Δ TF binding

↓

Δ Gene regulation

↓

Δ Phenotype

Simon Zumkeller

Gernot Schmitz

From genetic variation to phenotypic traits

1001 Genomes

AraGWAS Catalog 36878 phenotype associations

SNP / SV

↓

Δ TF binding

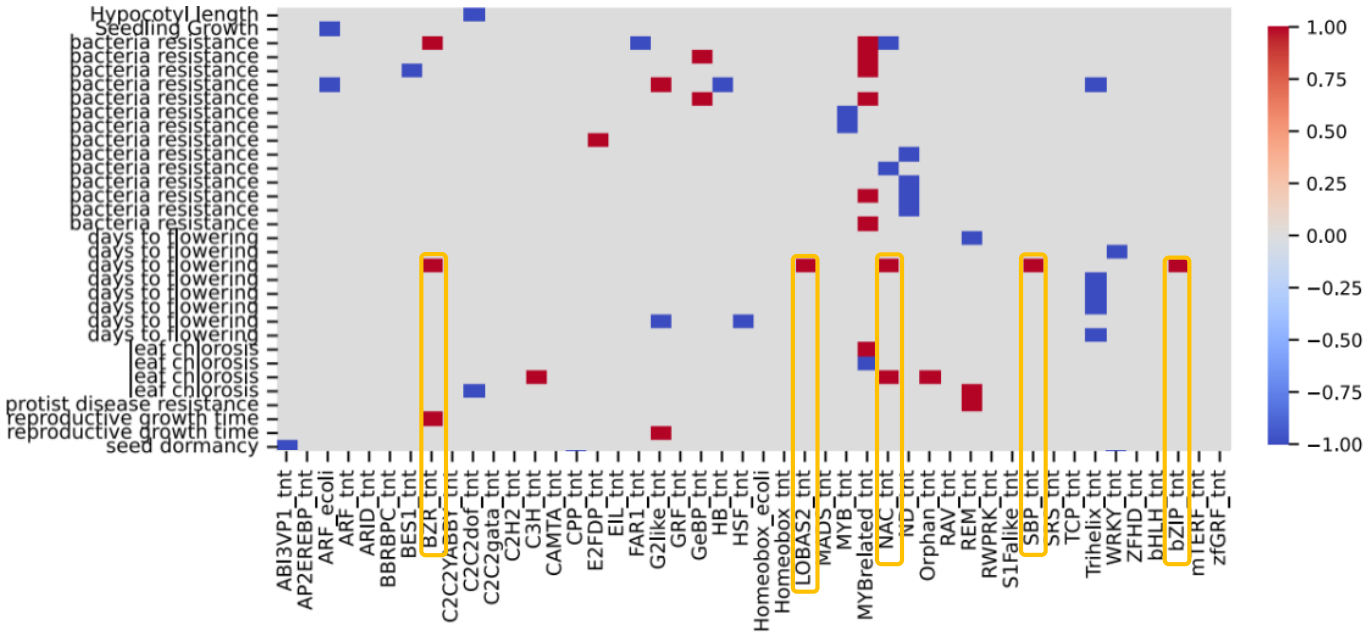
↓

Δ Gene regulation

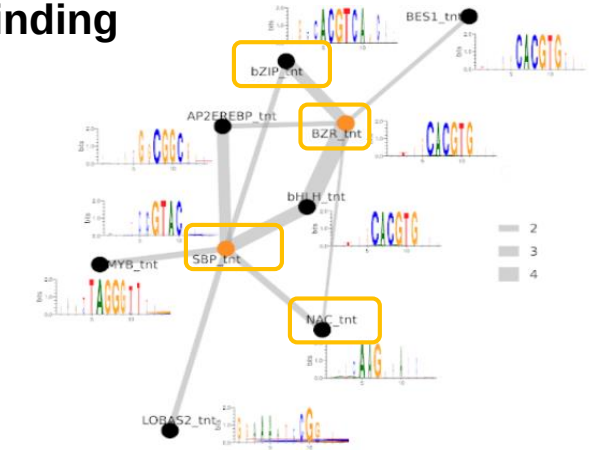
↓

Δ Phenotype

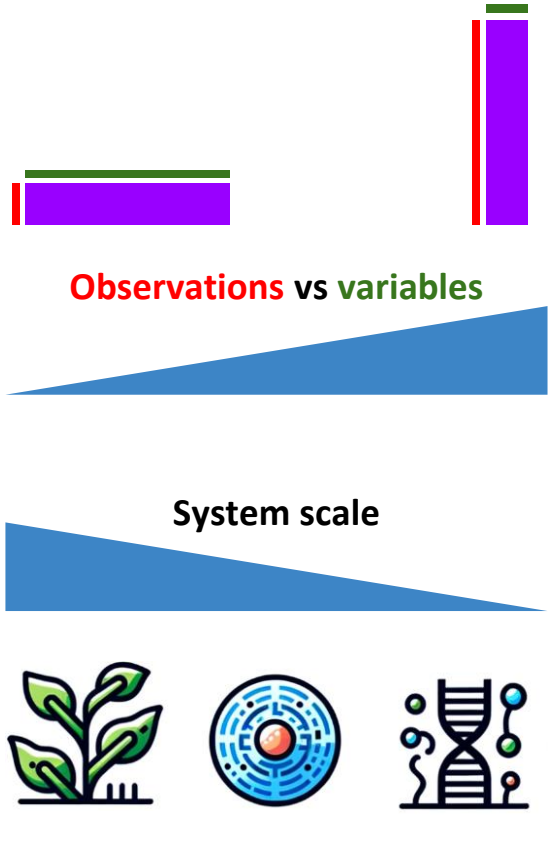
TF binding – phenotype associations



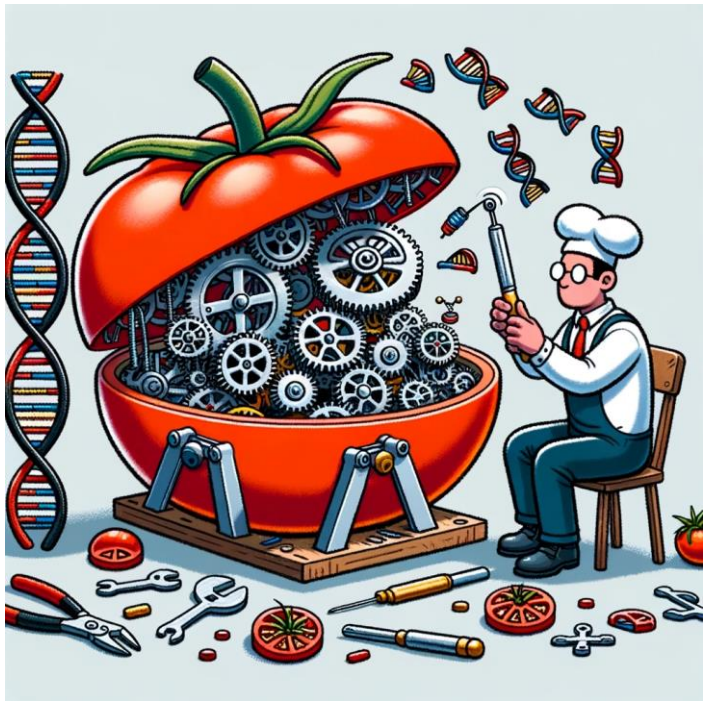
Combinatorial TF binding



AI for detective work



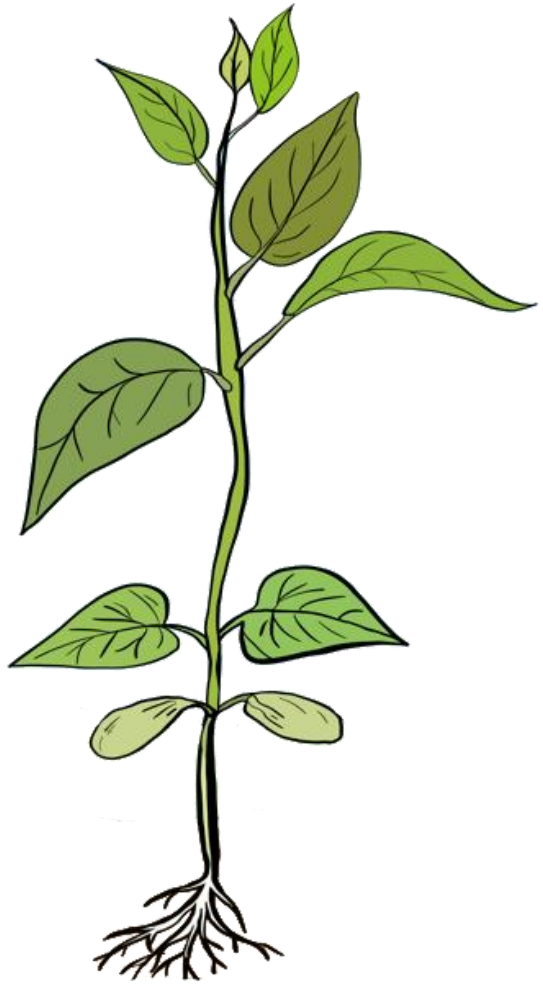
AI for reverse engineering



AI for computer games

The green game of survival

DALL-E nonsense



The green game of survival

DALL-E nonsense

The green game of survival

Balance resources

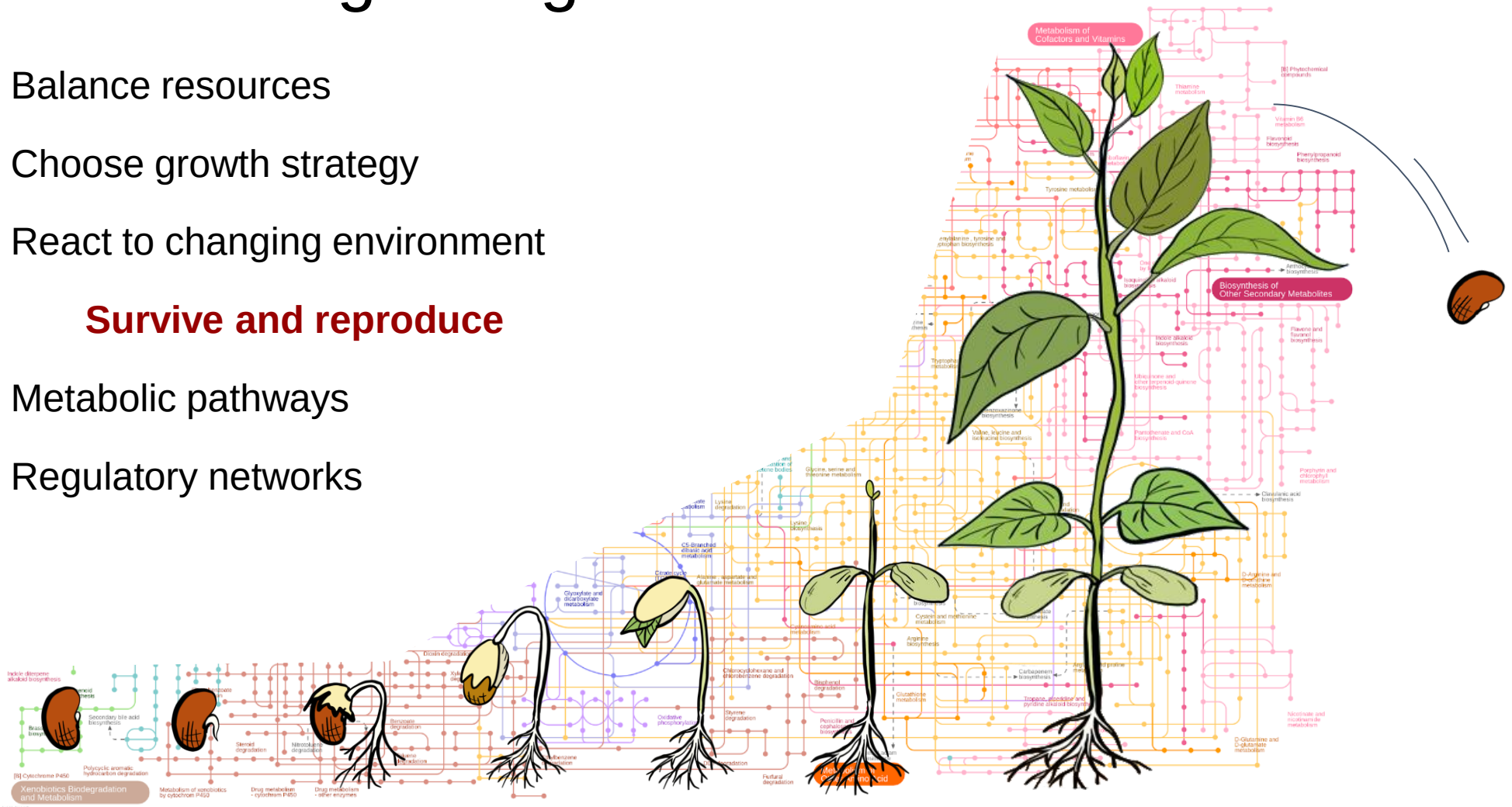
Choose growth strategy

React to changing environment

Survive and reproduce

Metabolic pathways

Regulatory networks



The green game of survival

So what if…

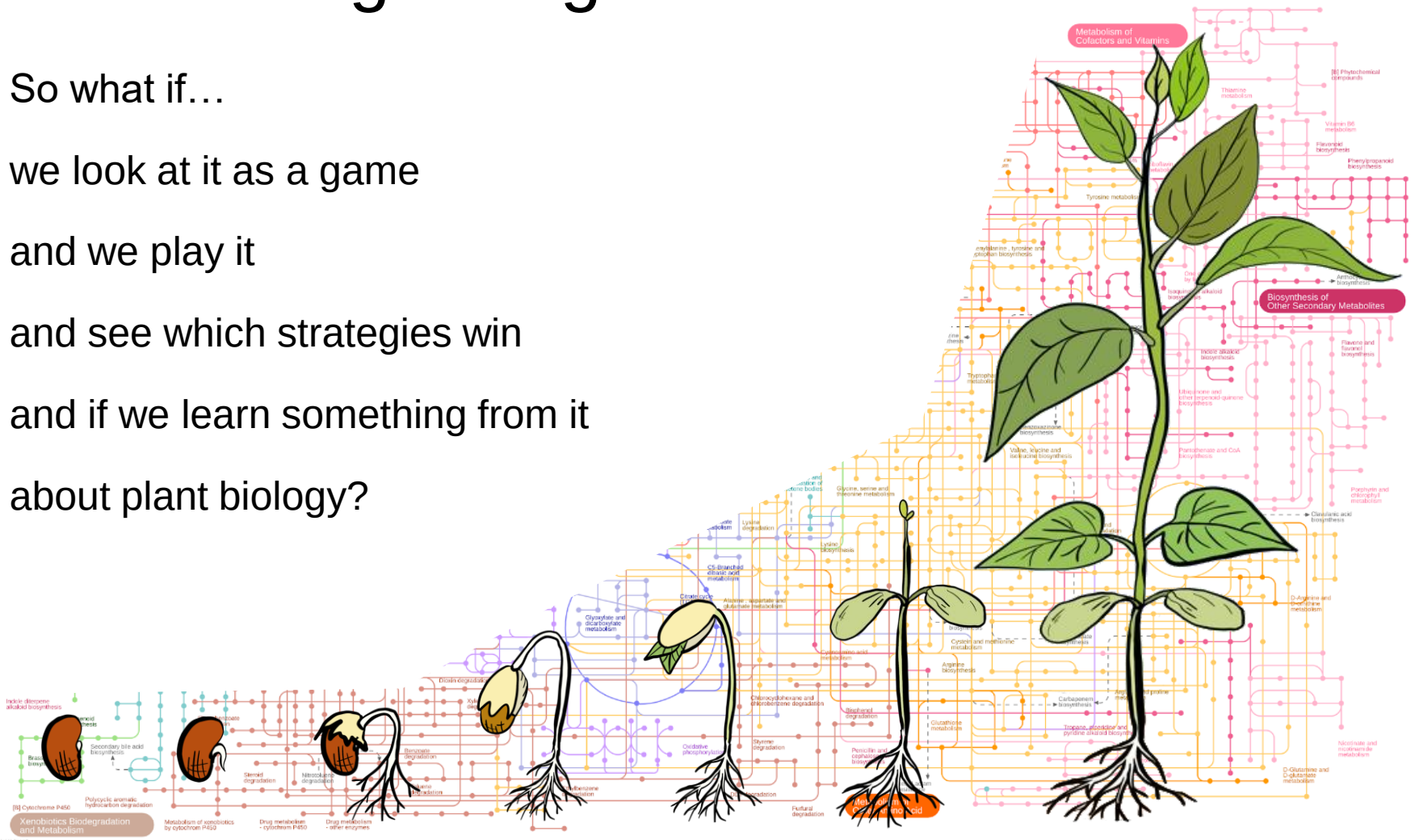
we look at it as a game

and we play it

and see which strategies win

and if we learn something from it

about plant biology?



The green game of survival

So what if…

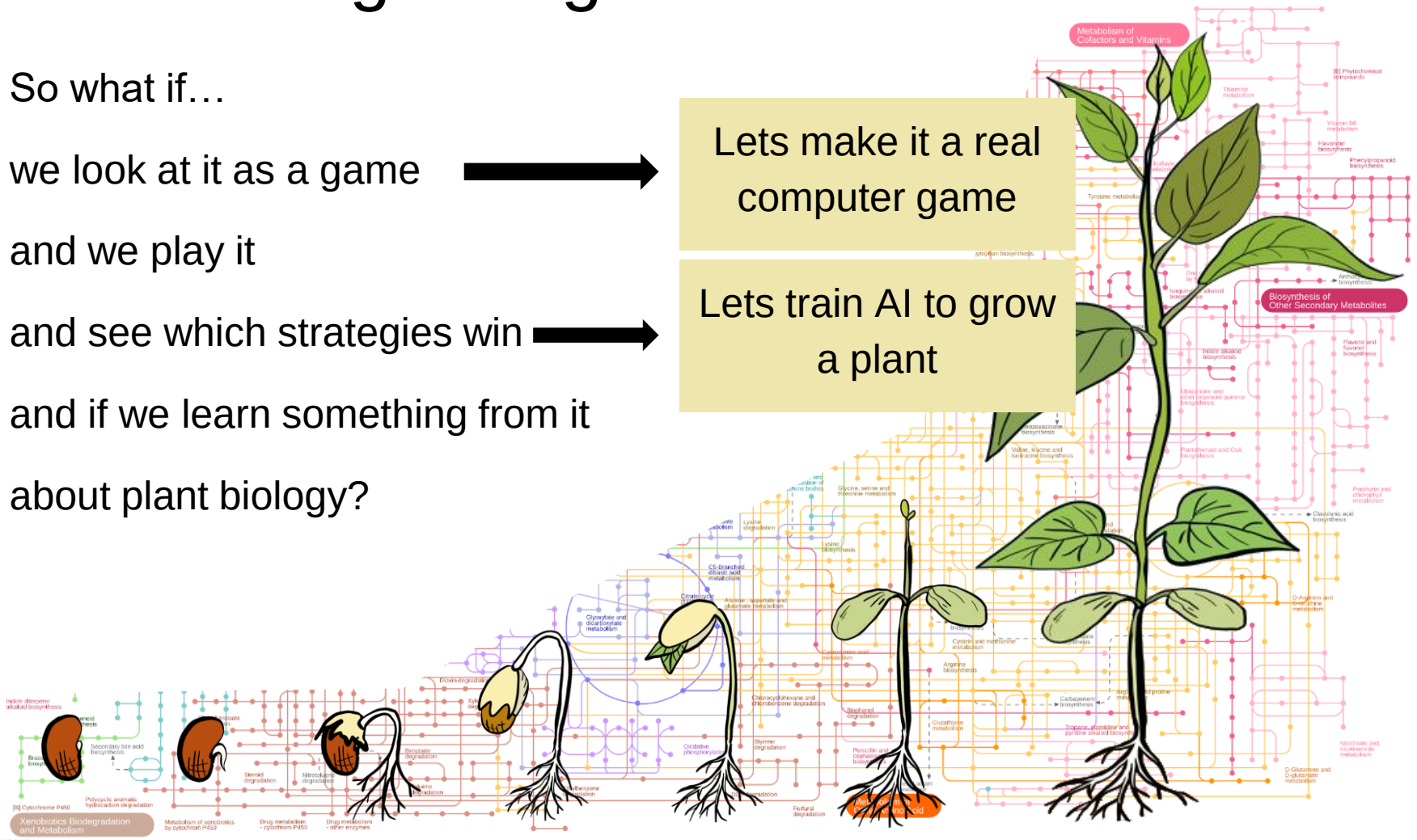
we look at it as a game → Lets make it a real computer game

and we play it

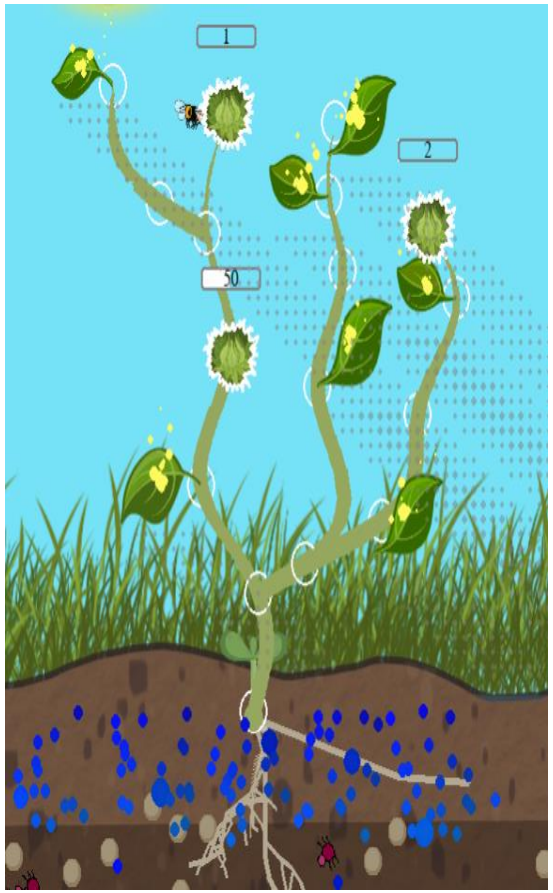
and see which strategies win → Lets train AI to grow a plant

and if we learn something from it

about plant biology?



A digital plant in a digital environment



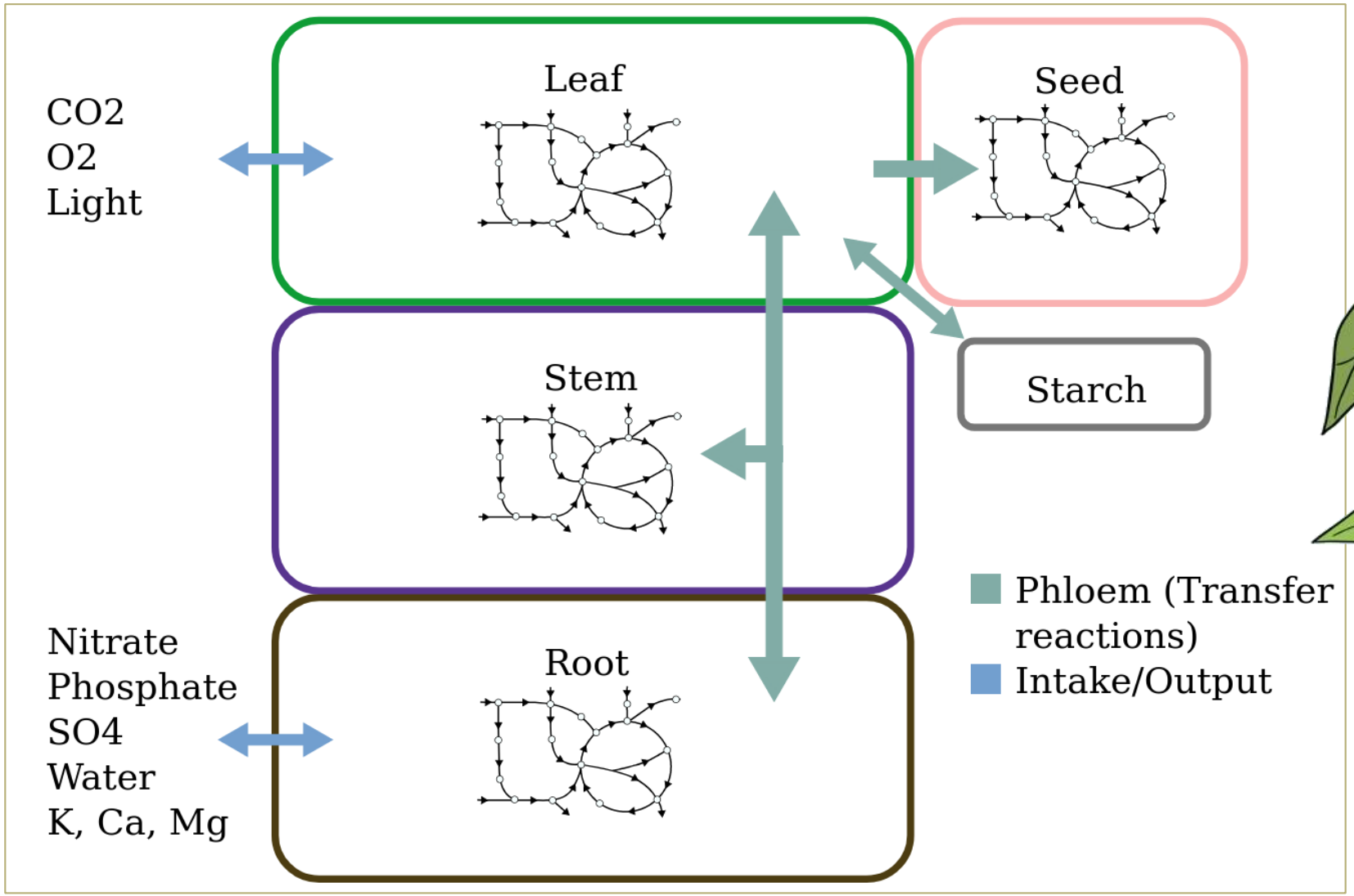
- Light intensity
- Light spectrum
- Temperature
- Humidity
- Water availability
- Nutrients
- .
- .
- CO_2
- O_2
- Time of the day
- Day/night cycle
- Season
- etc.
- .
- .
- .

Whole-plant metabolic model

Stefano La Cruz
Uni Köln

Nadine Töpfer
Uni Köln

töpfer lab

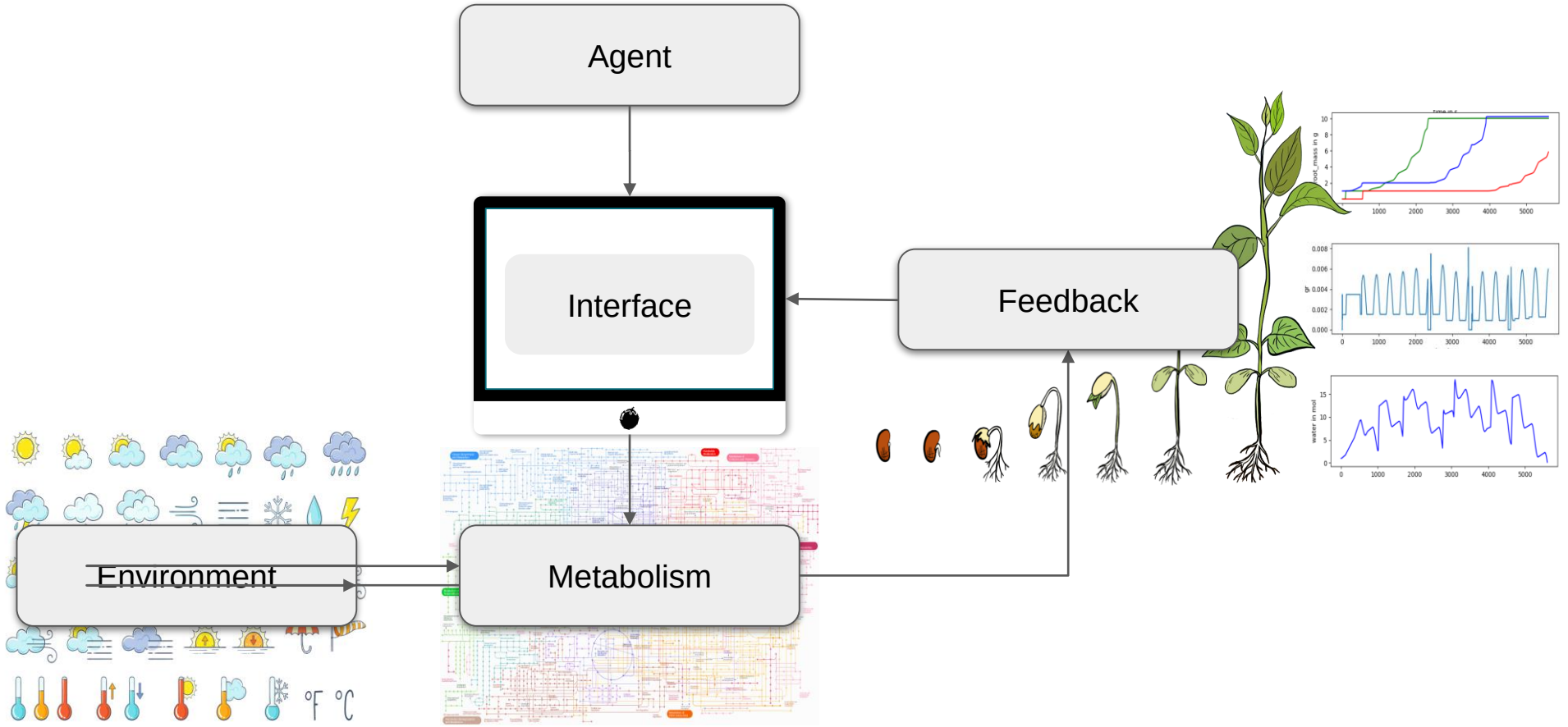


IPK LEIBNIZ INSTITUTE

JÜLICH Forschungszentrum

The game loop

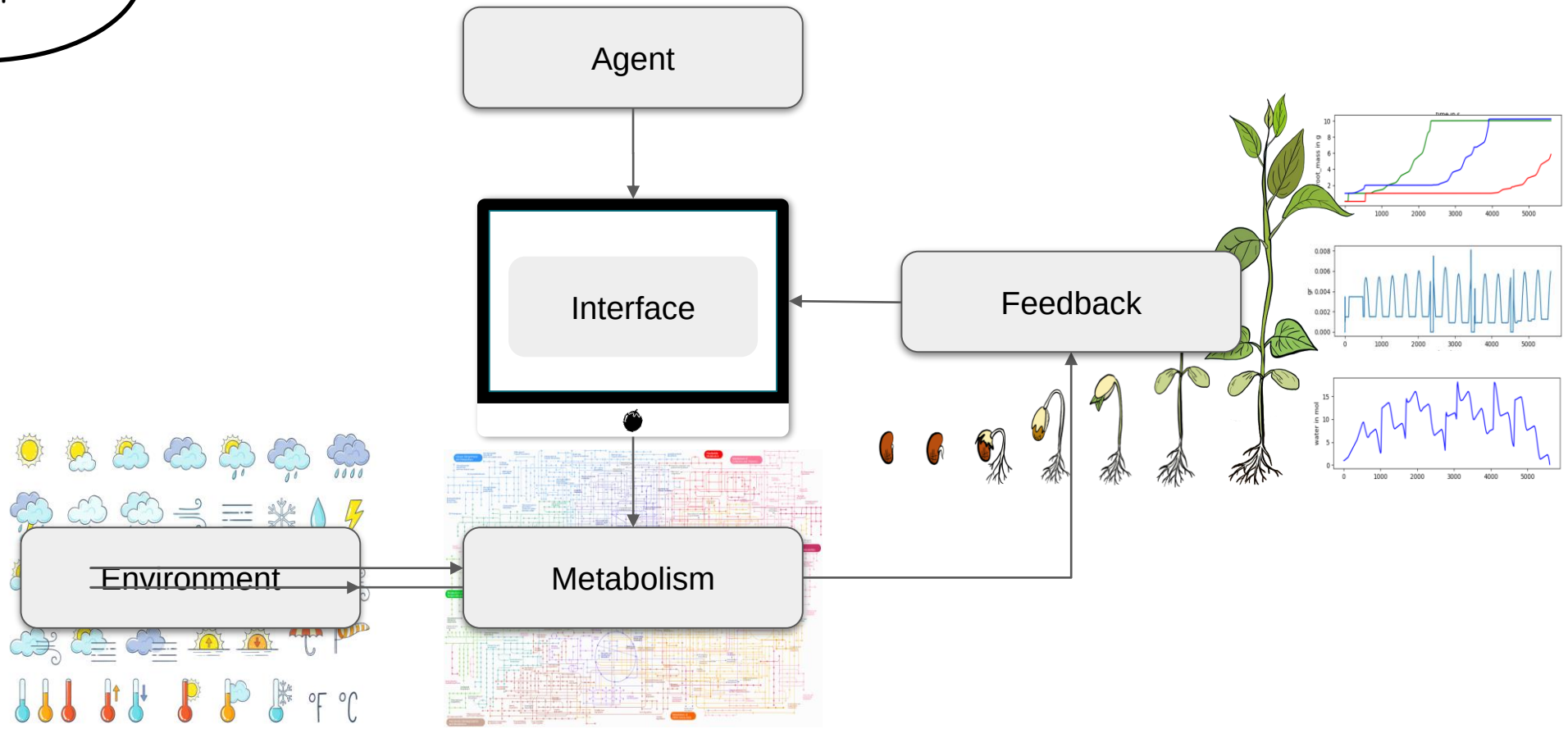
Daniel Koch



Koch *et al.* in preparation

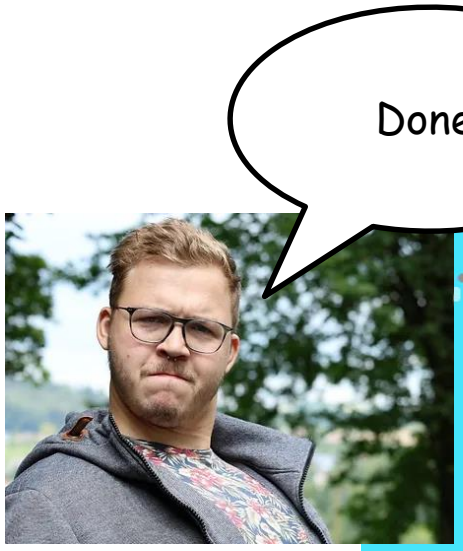
The game loop

Daniel Koch

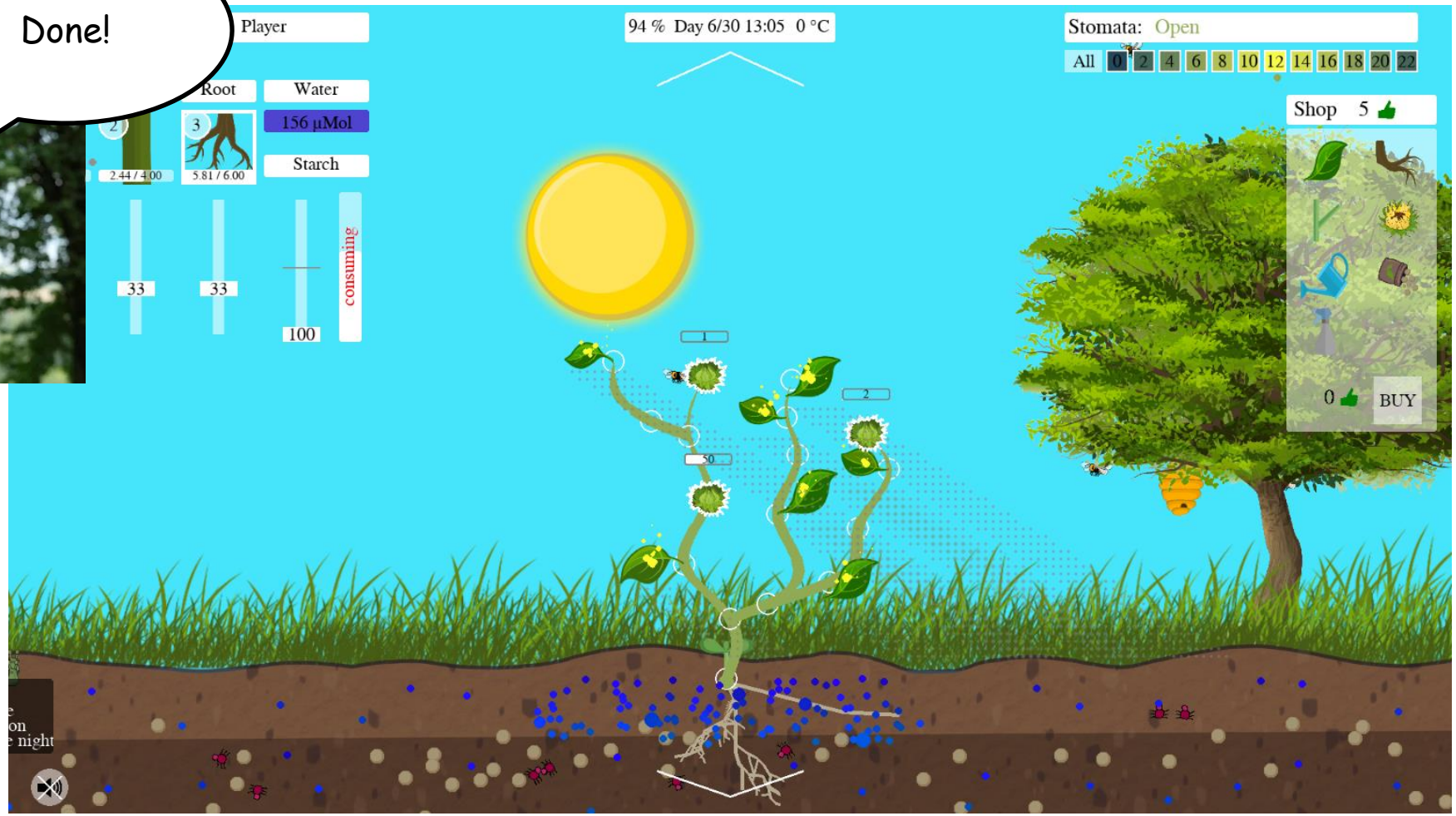


Koch *et al.* in preparation

Interface



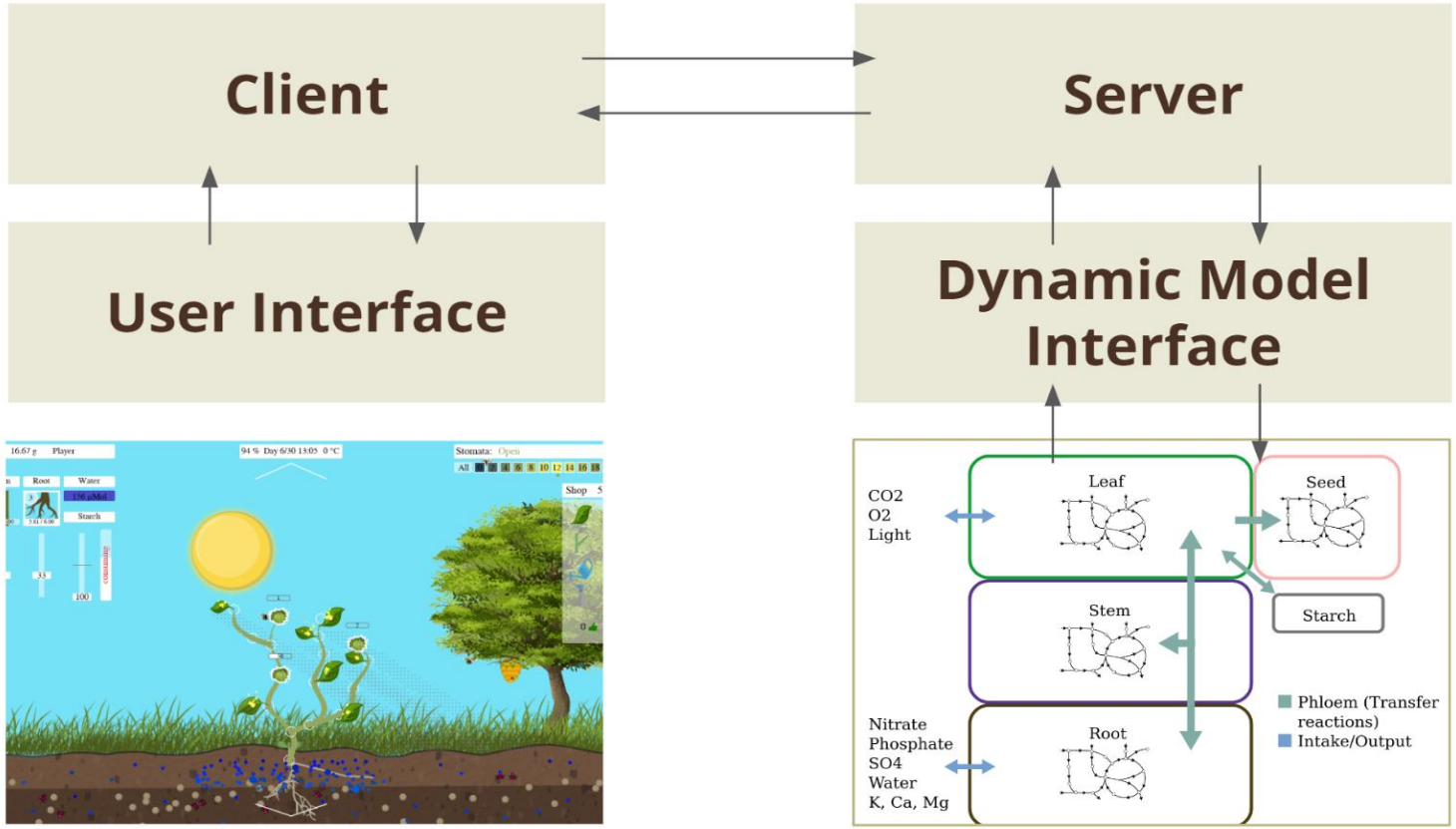
Daniel Koch



https://danielkoch.itch.io/planted

Implementation

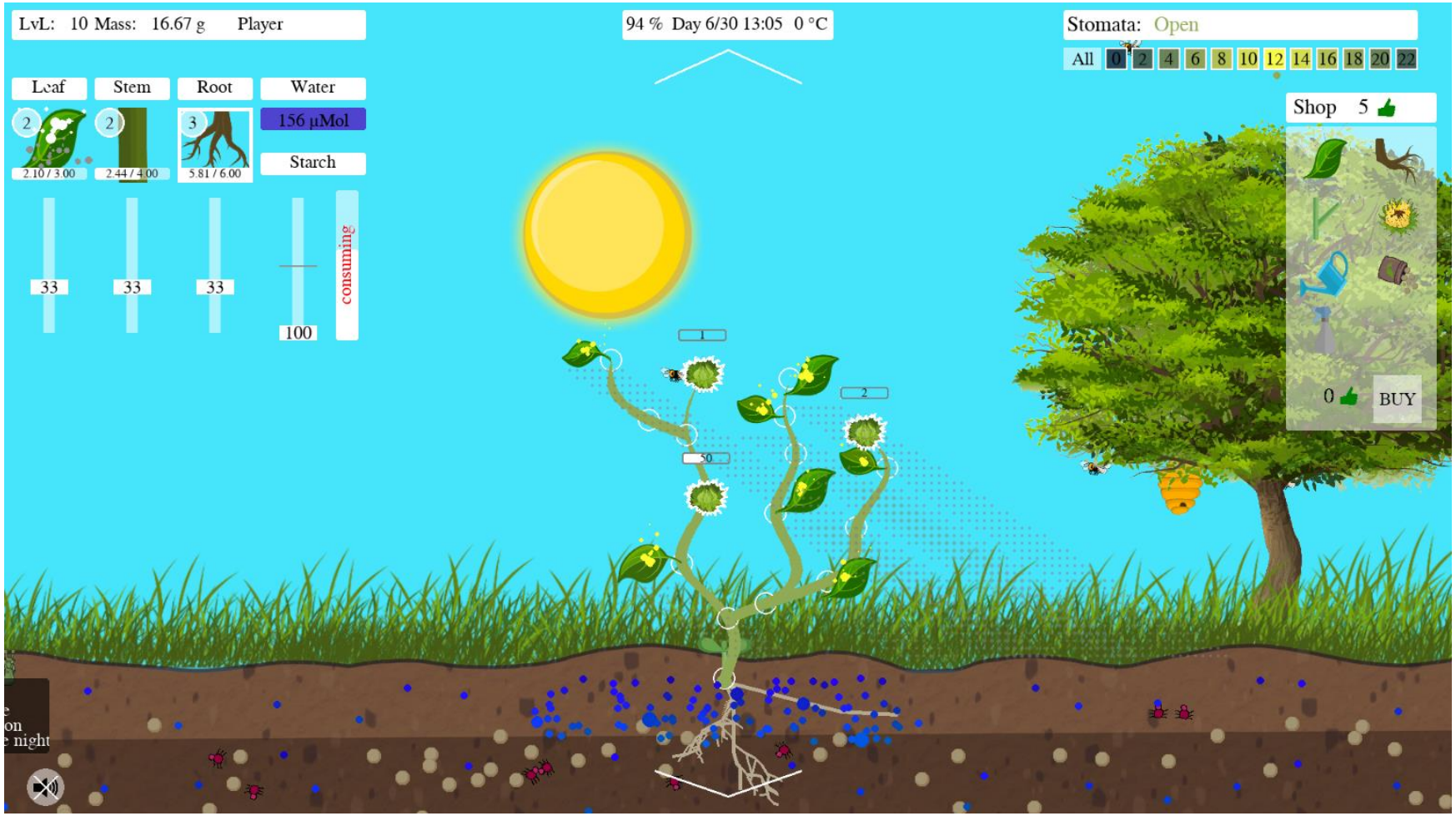
Daniel Koch



Jan-Niklas Weder
Uni Köln

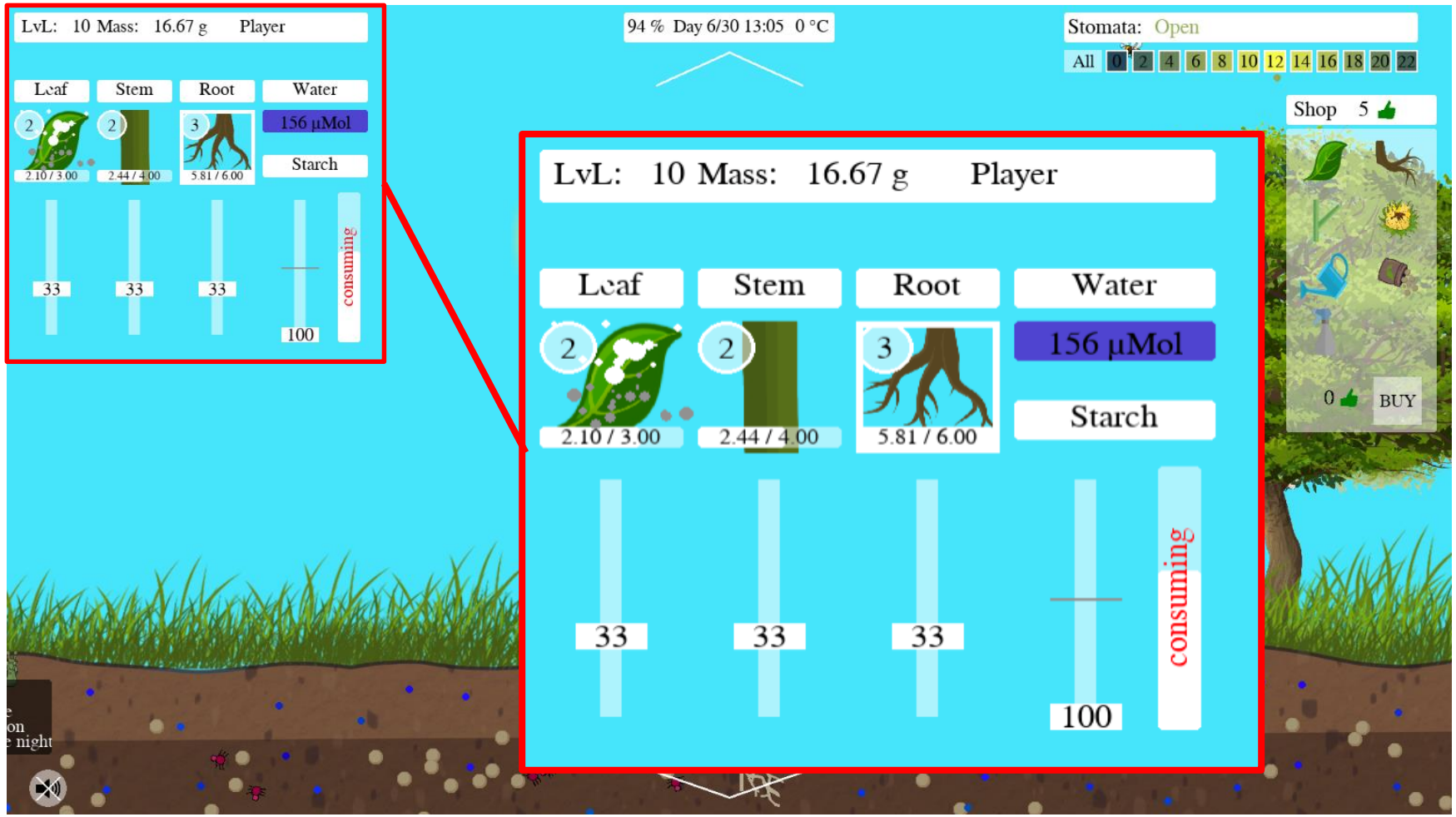
Dennis Psaroudakis
IPK Gatersleben

Interface



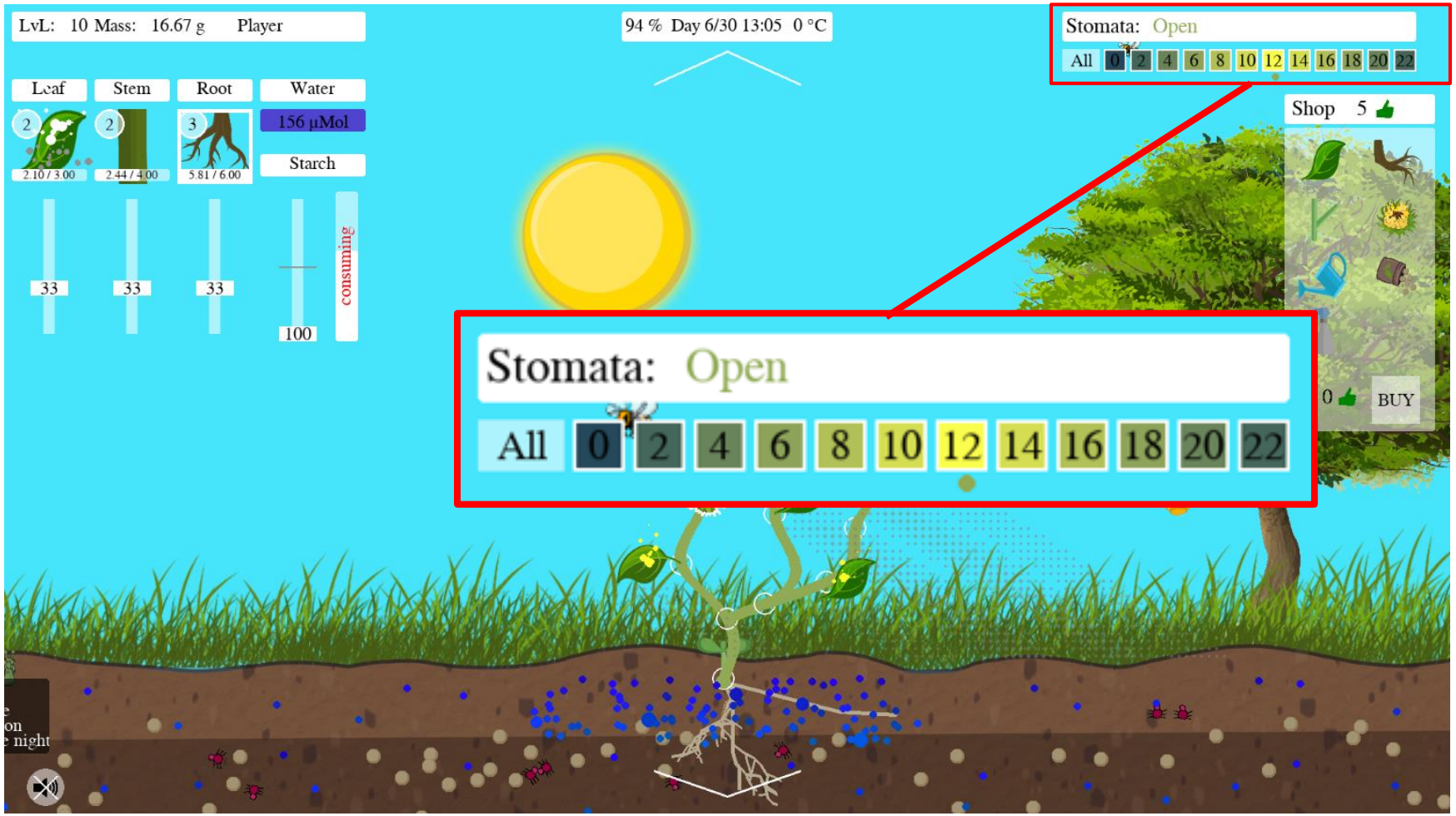
https://danielkoch.itch.io/planted

Interface



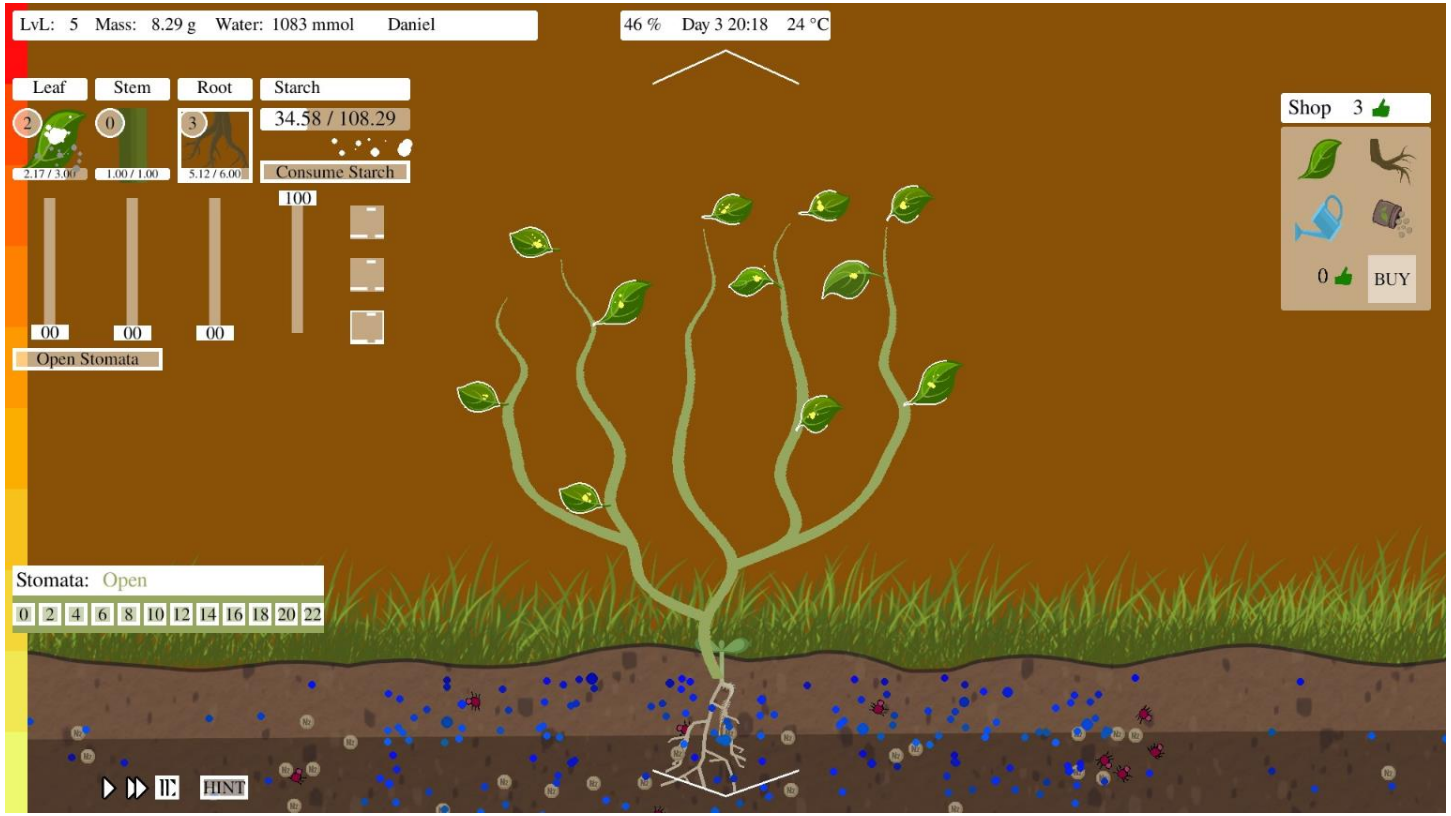
https://danielkoch.itch.io/planted

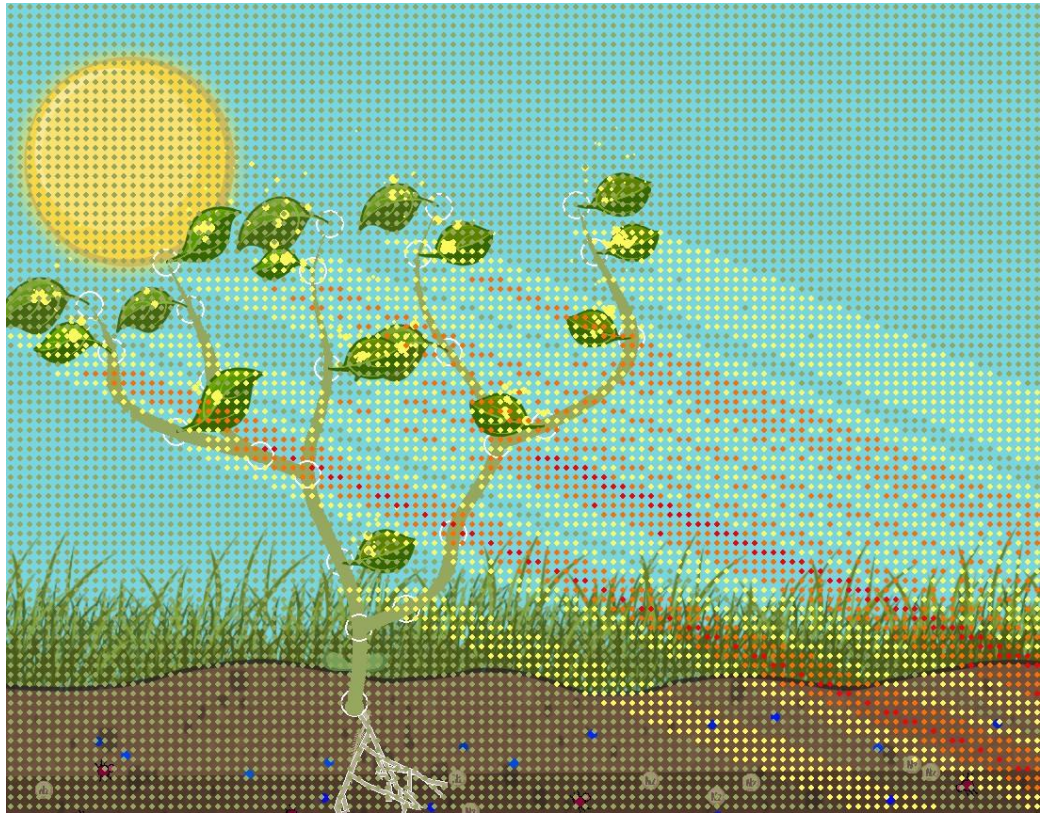
Interface



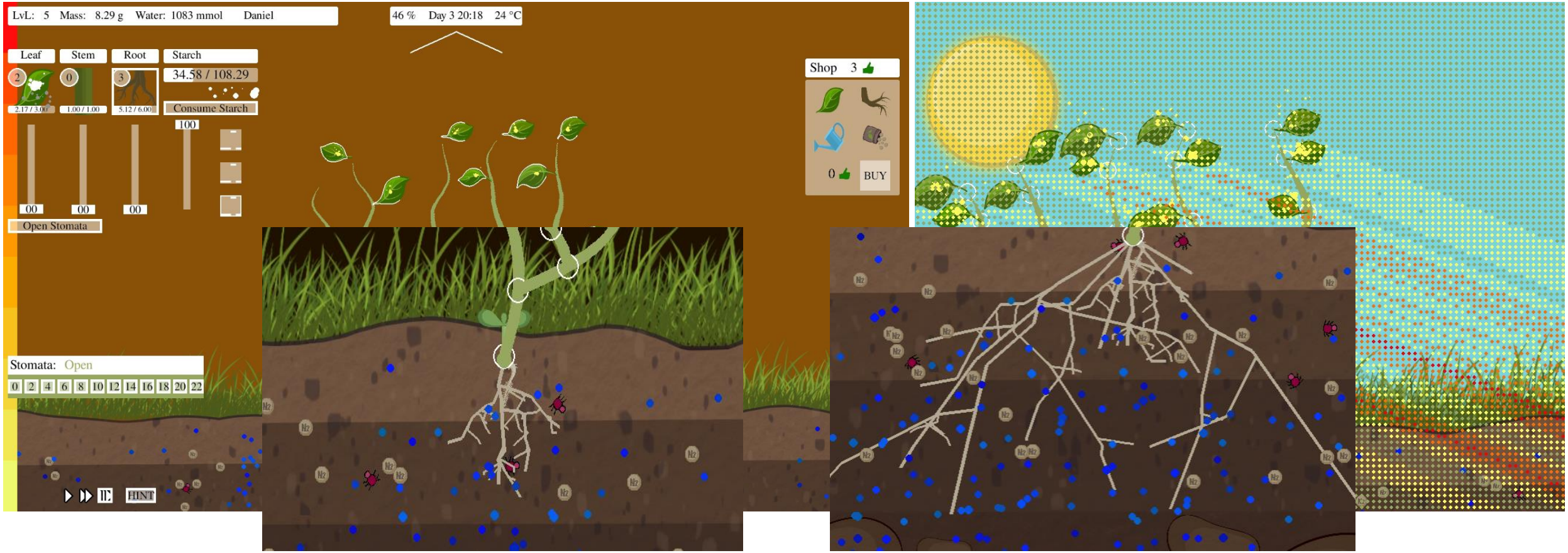
https://danielkoch.itch.io/planted

Interface

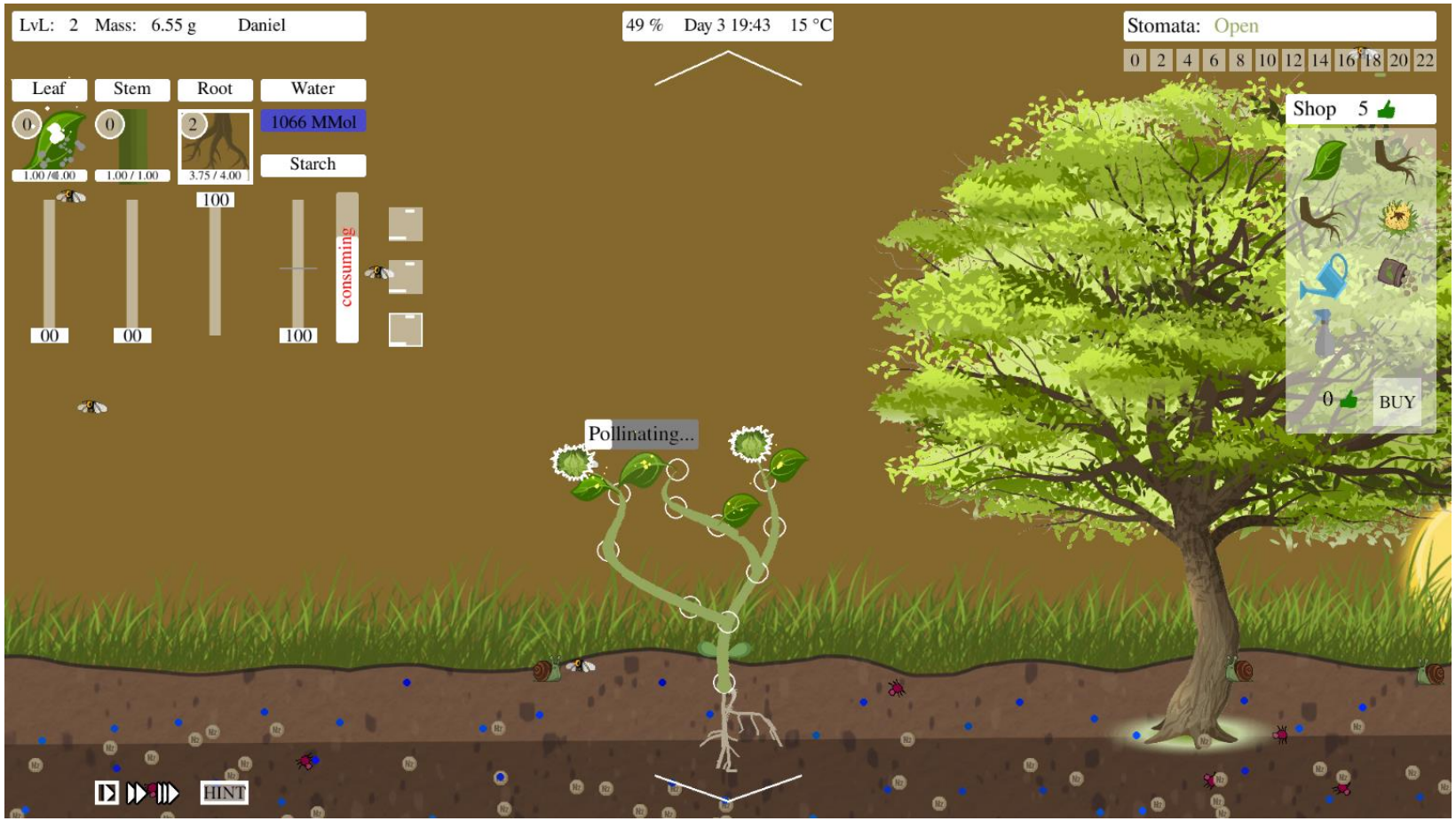

LvL: 5 Mass: 8.29 g Water: 1083 mmol Daniel
46 % Day 3 20:18 24 °C
Leaf
Stem
Root
Starch
34.58 / 108.29
Consume Starch
Shop 3
BUY
Open Stomata
Stomata: Open
0 2 4 6 8 10 12 14 16 18 20 22
HINT



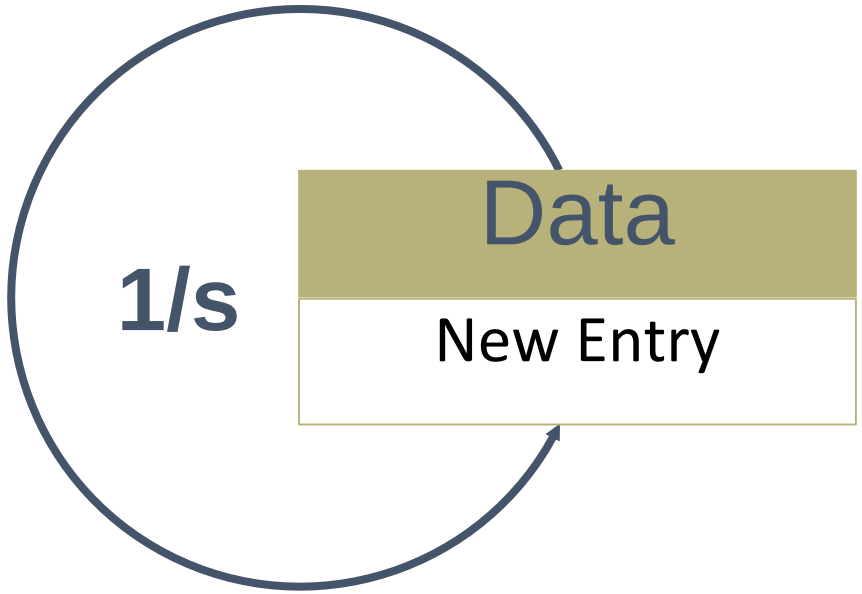
Interface

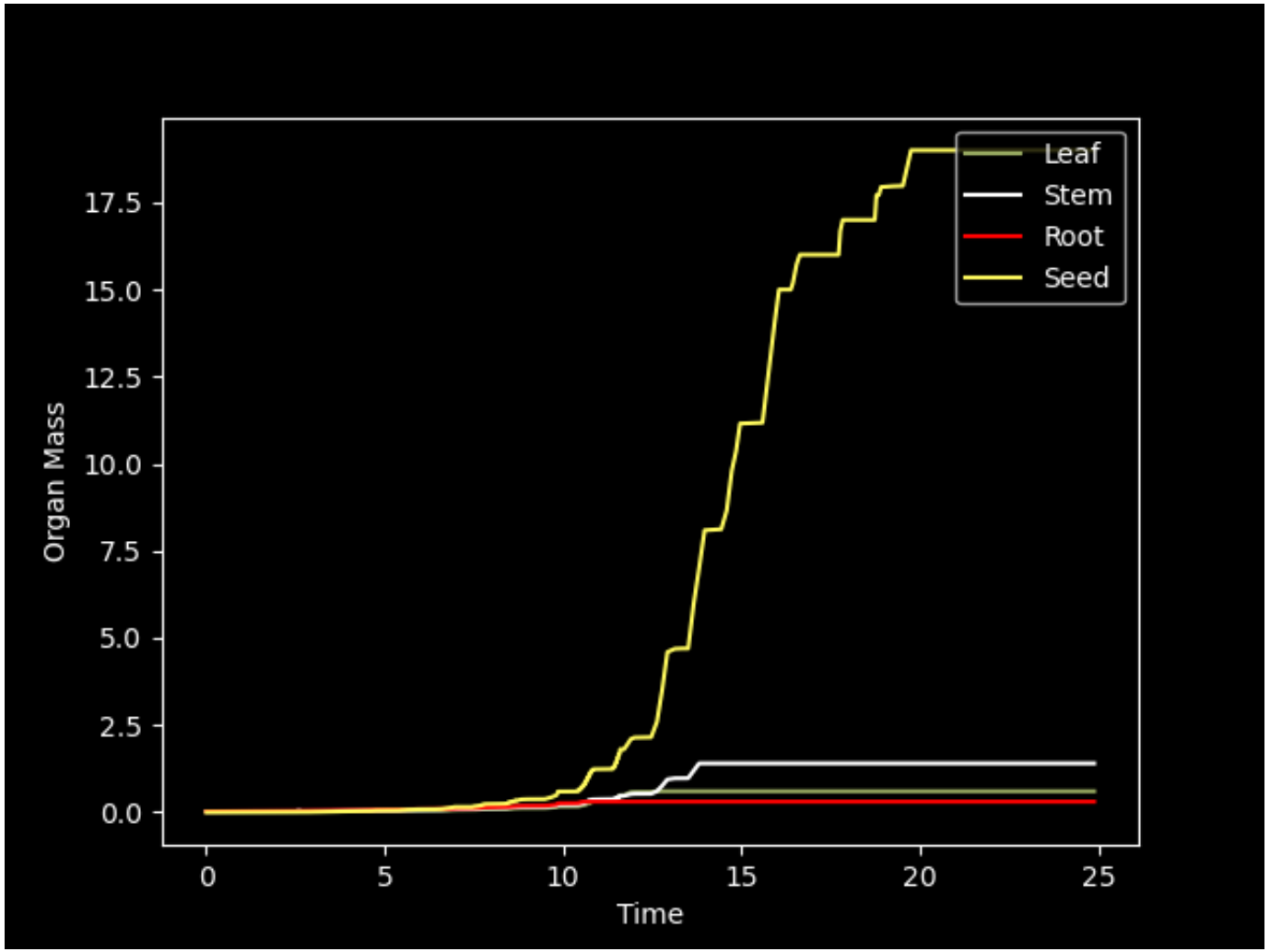

LvL: 5 Mass: 8.29 g Water: 1083 mmol Daniel
46 % Day 3 20:18 24 °C
Leaf
Stem
Root
Starch
34.58 / 108.29
Consume Starch
Open Stomata
Shop 3
BUY
Stomata: Open
0 2 4 6 8 10 12 14 16 18 20 22
HINT

Interface

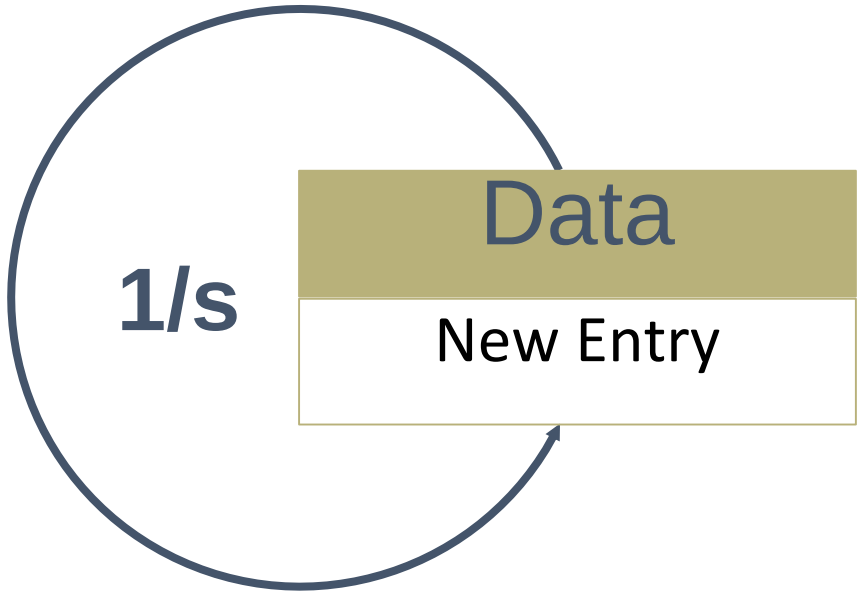

LvL: 2 Mass: 6.55 g Daniel
49 % Day 3 19:43 15 °C
Stomata: Open
0 2 4 6 8 10 12 14 16 18 20 22
Leaf
Stem
Root
Water
1066 MMol
Starch
1.00 / 4.00
1.00 / 1.00
3.75 / 4.00
100
consuming
00
00
100
Shop 5
0
BUY
Pollinating...
HINT

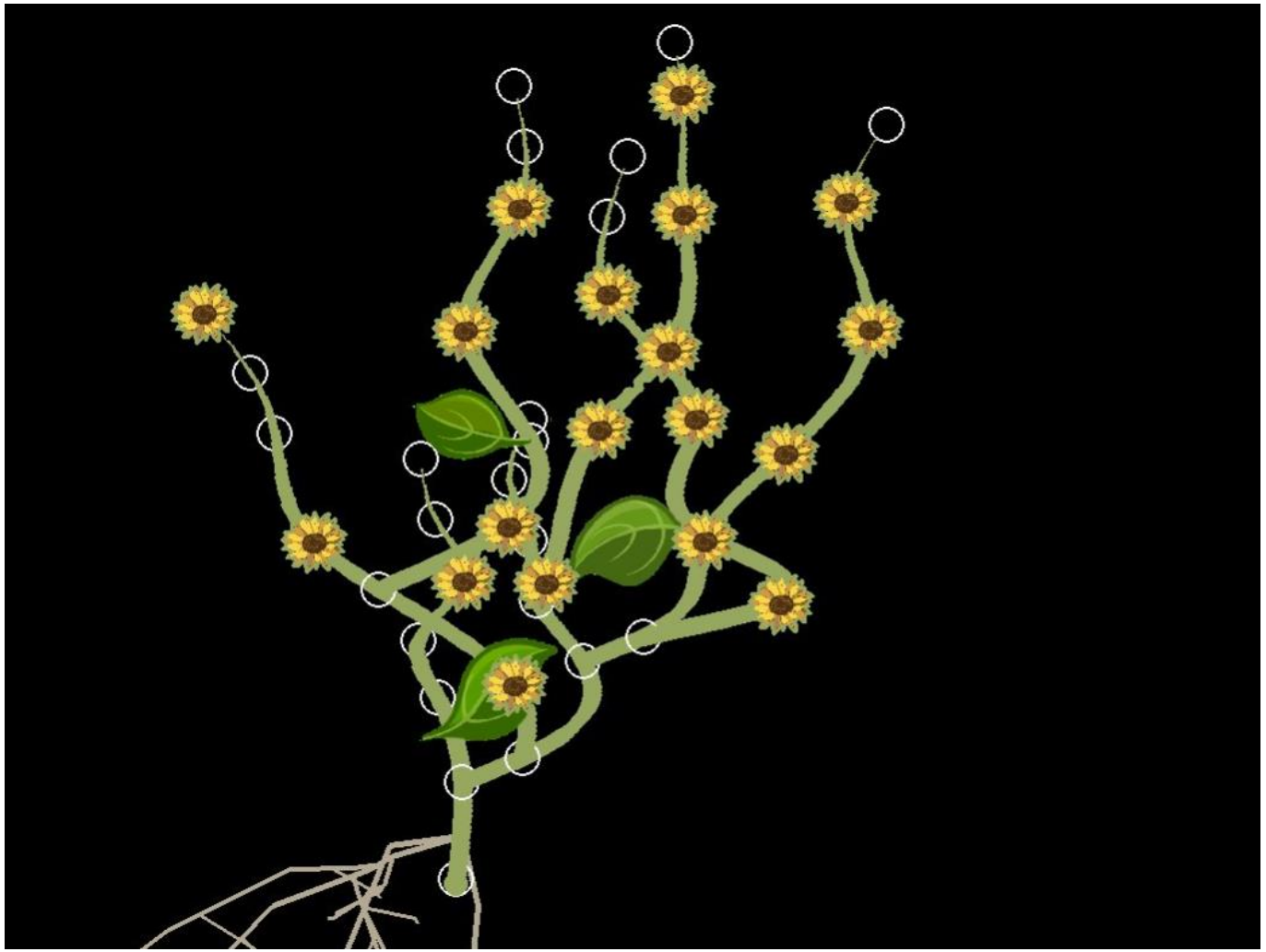
Data collection

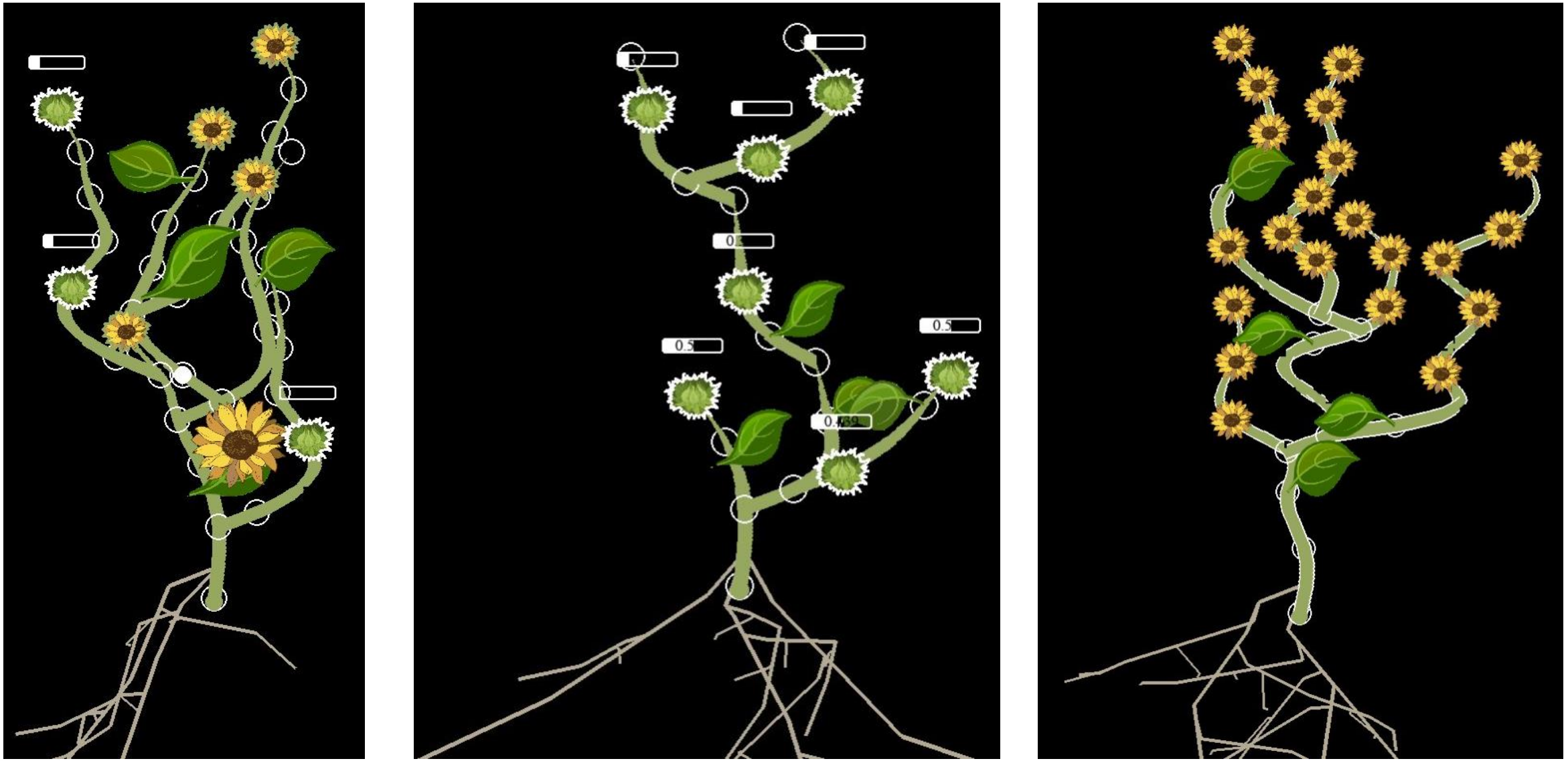




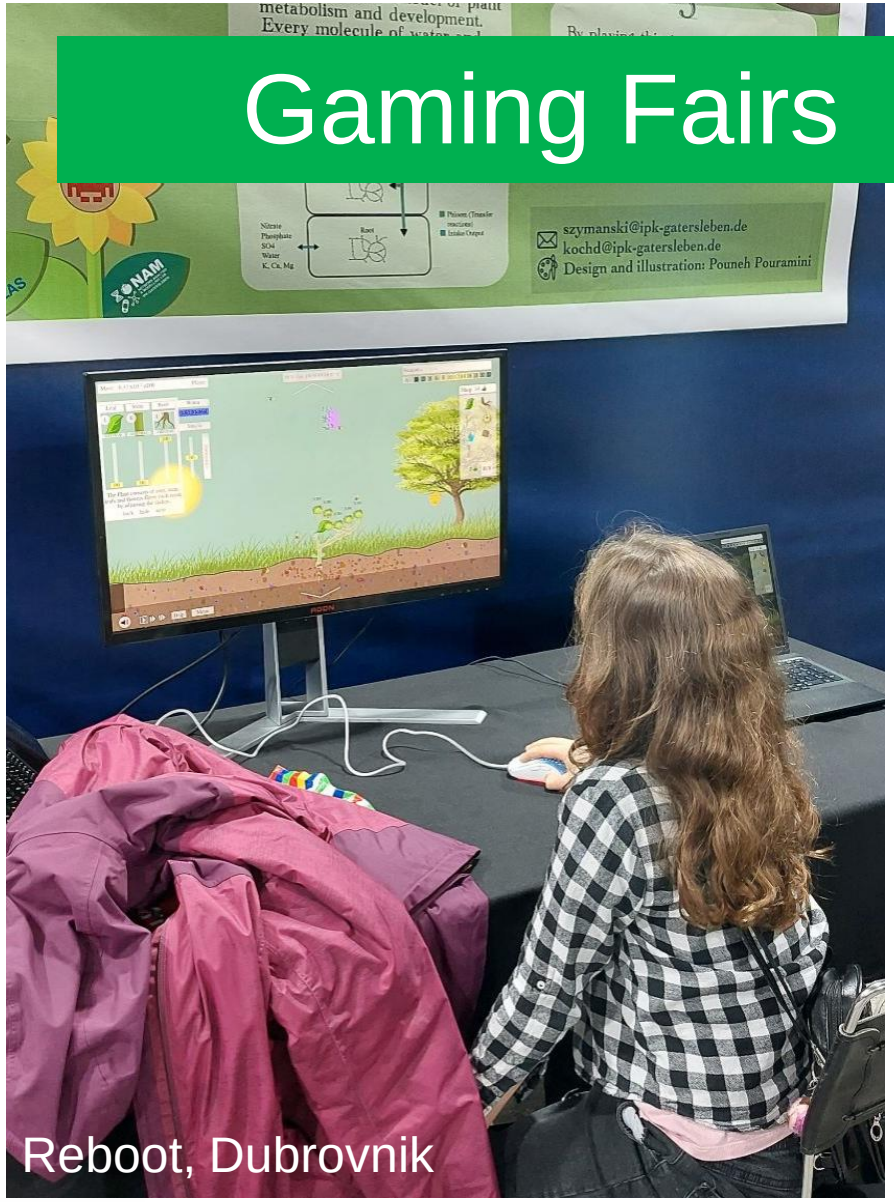
Data collection



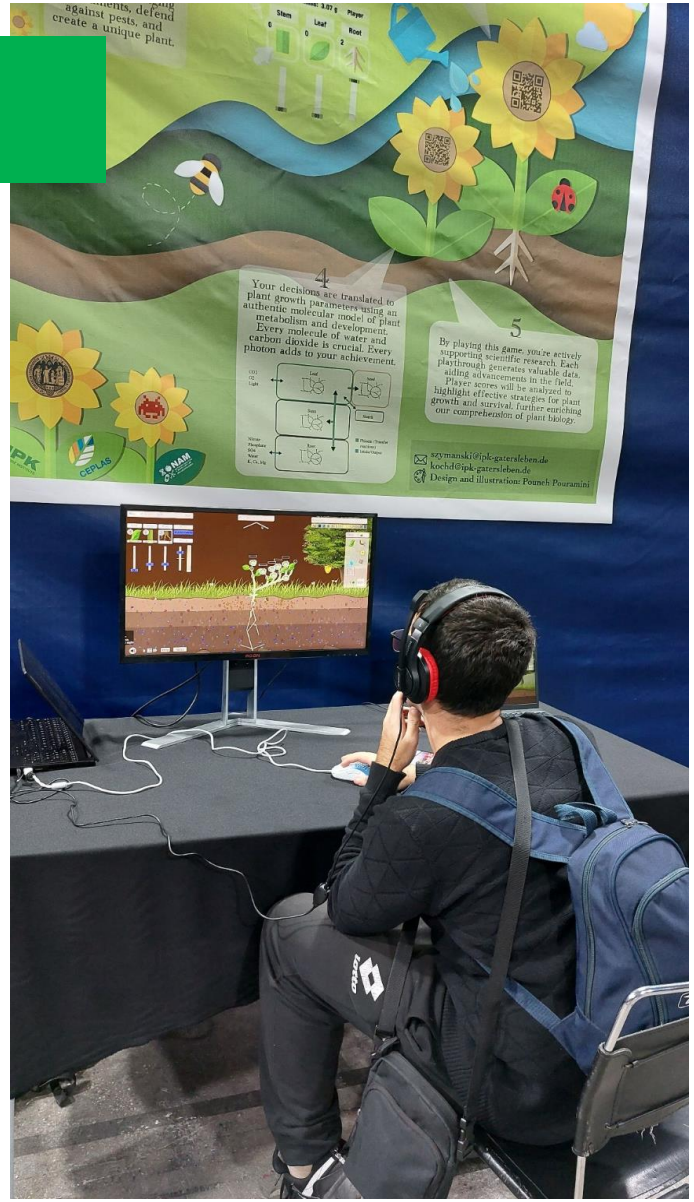



0
0.5
0.5

Gaming Fairs



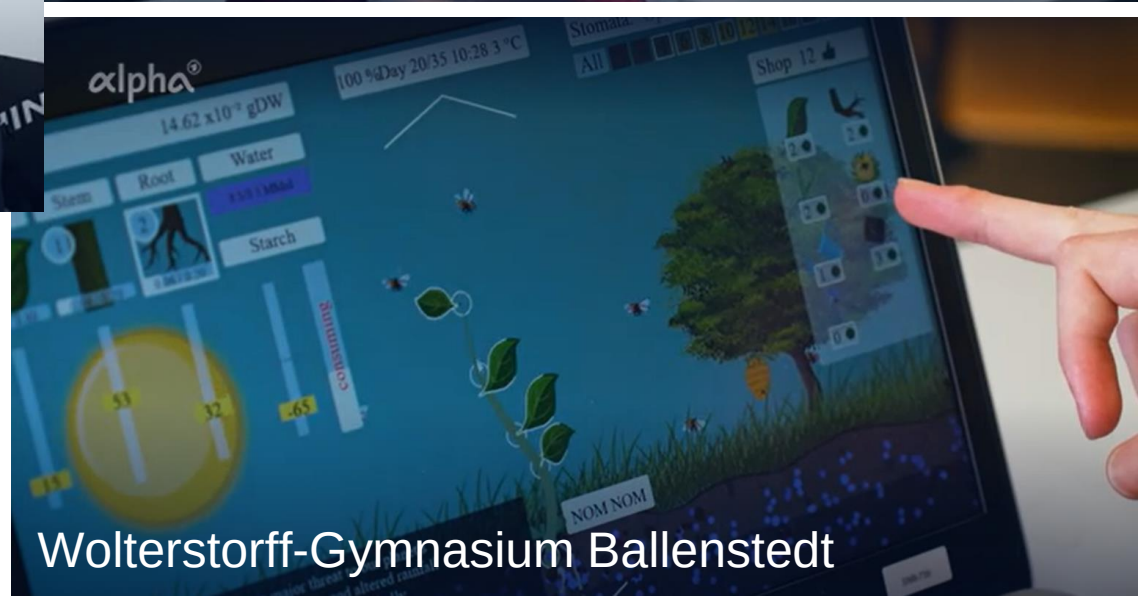
Reboot, Dubrovnik



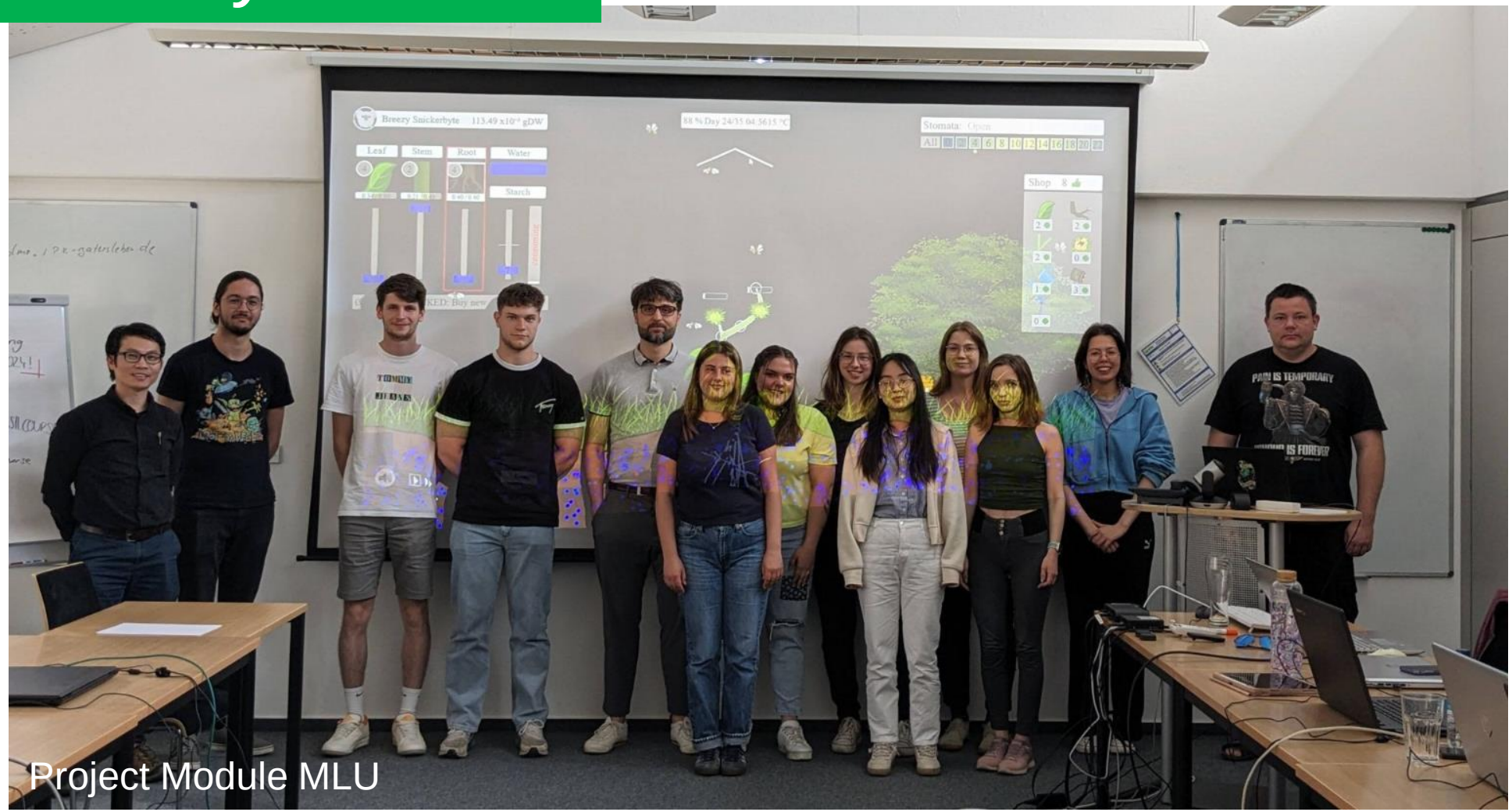
Schools


alpha
Die Teams müssen die Pflanze **gut versorgen,**
damit sie **viele Samen produziert.**


alpha


alpha
Wolterstorff-Gymnasium Ballenstedt

University courses



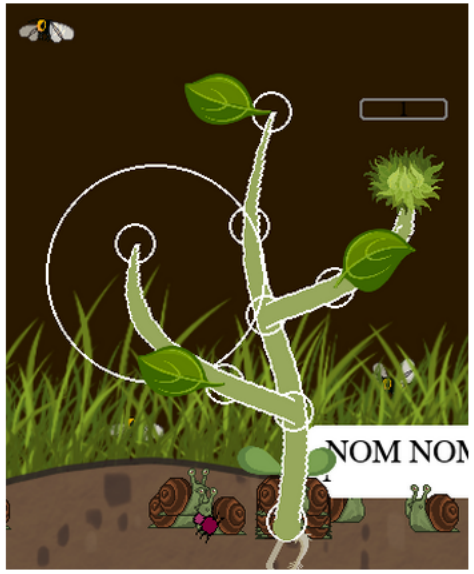
Project Module MLU

https://danielkoch.itch.io/planted

PlantEd

PlantEd beta

PlantEd is an engaging plant growth simulation game where players nurture a seed or seedling into maturity. Strategically allocate biomass to develop leaves, branches, roots, and flowers, aiming to maximize seed mass within a 35-day timeframe. Navigate dynamic weather conditions, manage resources efficiently, and interact with fun elements like snails, bees, and mites. Challenge yourself to optimize growth strategies and achieve the highest scores. Immerse into the fascinating world of plant biology with PlantEd!



More information

Download

Download PlantEd_V1.0.0.zip 221 MB

Download PlantEd_V1.0.0.tar.xz 183 MB

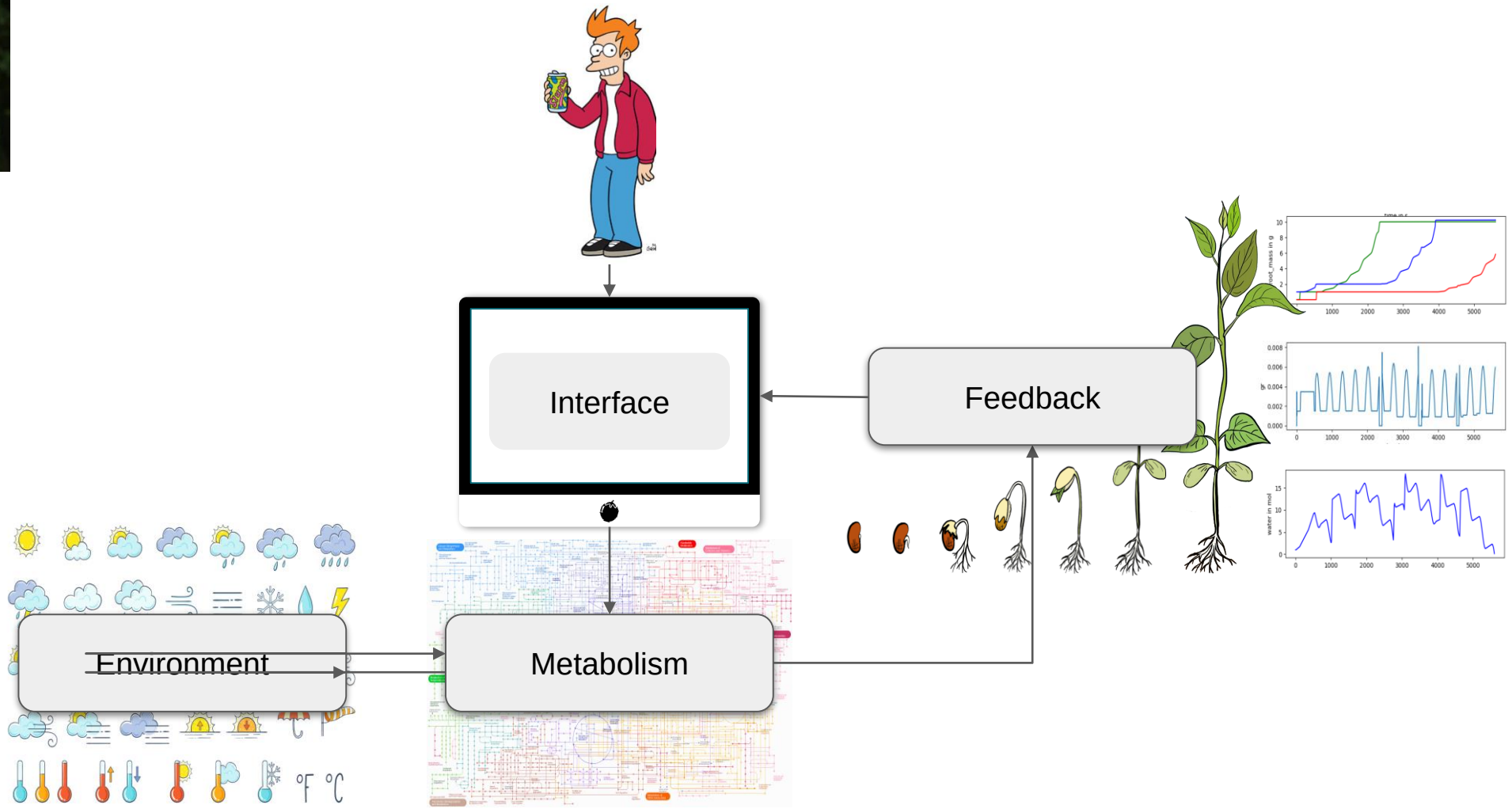
Install instructions

Tipps and tricks:

1. **Early Focus on Leaves:**
   - Begin by investing in leaves to maximize photon intake, as they are crucial for the plant's energy production through photosynthesis.
   - Higher photon intake will contribute to increased energy availability for other growth processes.
2. **Strategic Branch Development:**
   - Once a solid leaf foundation is established, invest in branches strategically. Branches enable more leaves and branches to grow, amplifying the overall photon intake.
   - Consider balancing branch and leaf growth to optimize energy production.
3. **Root Expansion for Resources:**

Human agent

Daniel Koch



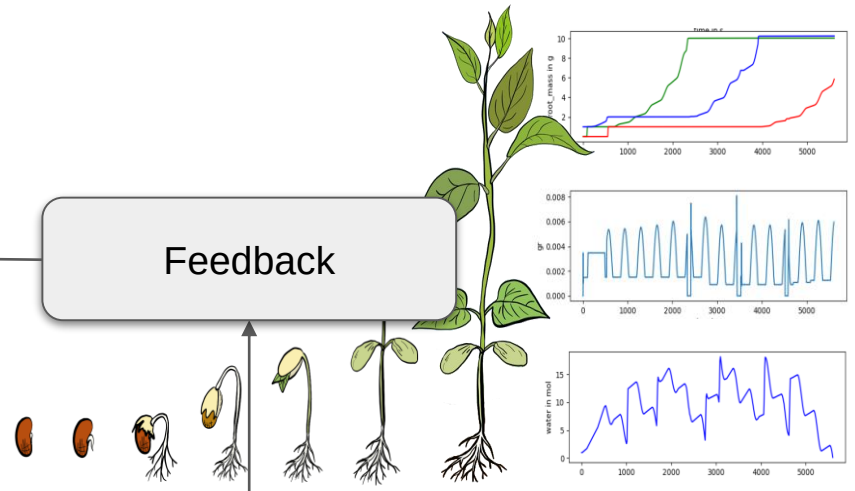
Koch *et al.* in preparation

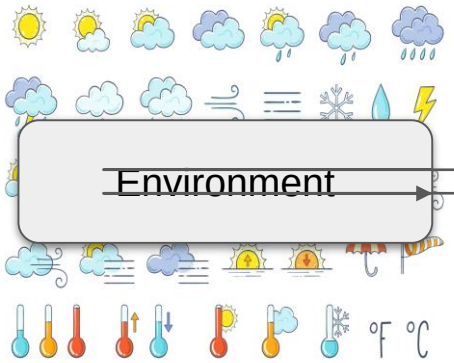
Daniel Koch

AI agent

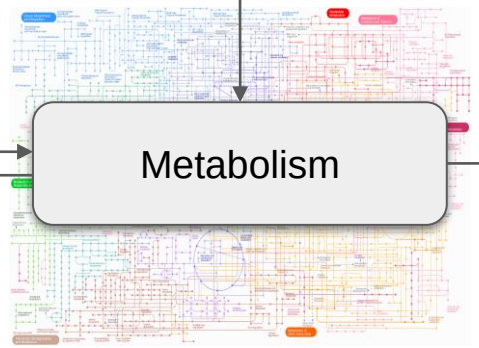
Interface

Feedback





Environment



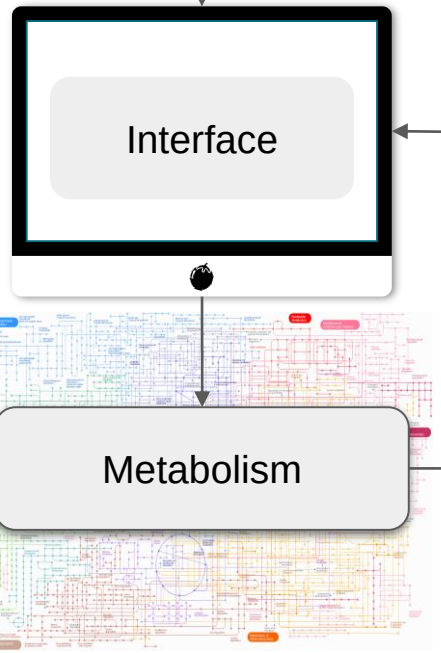
Metabolism

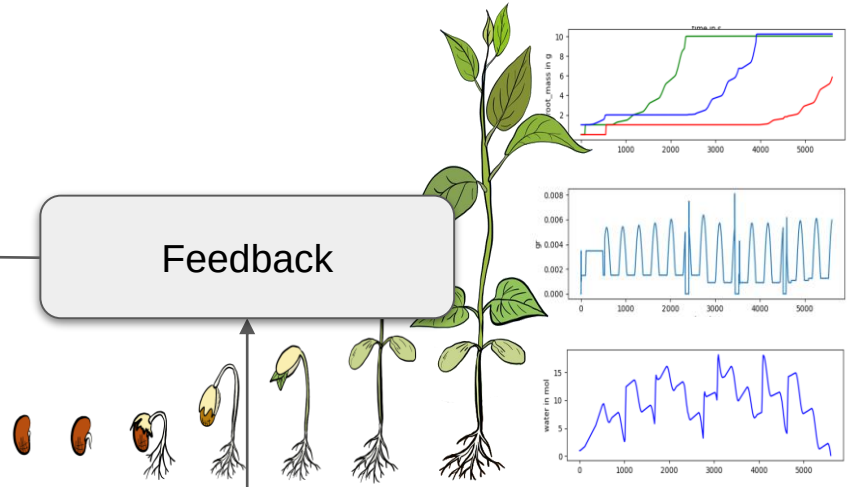
Koch *et al.* in preparation

AI agent

Daniel Koch

Dennis Psaroudakis



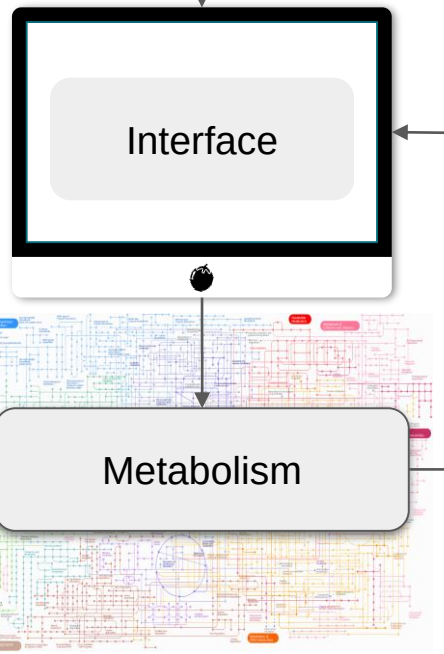


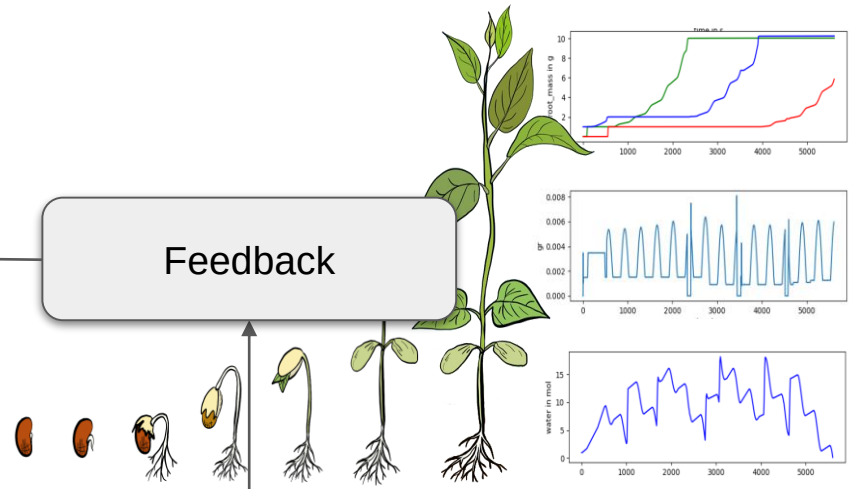
Koch *et al.* in preparation

AI agent

Daniel Koch

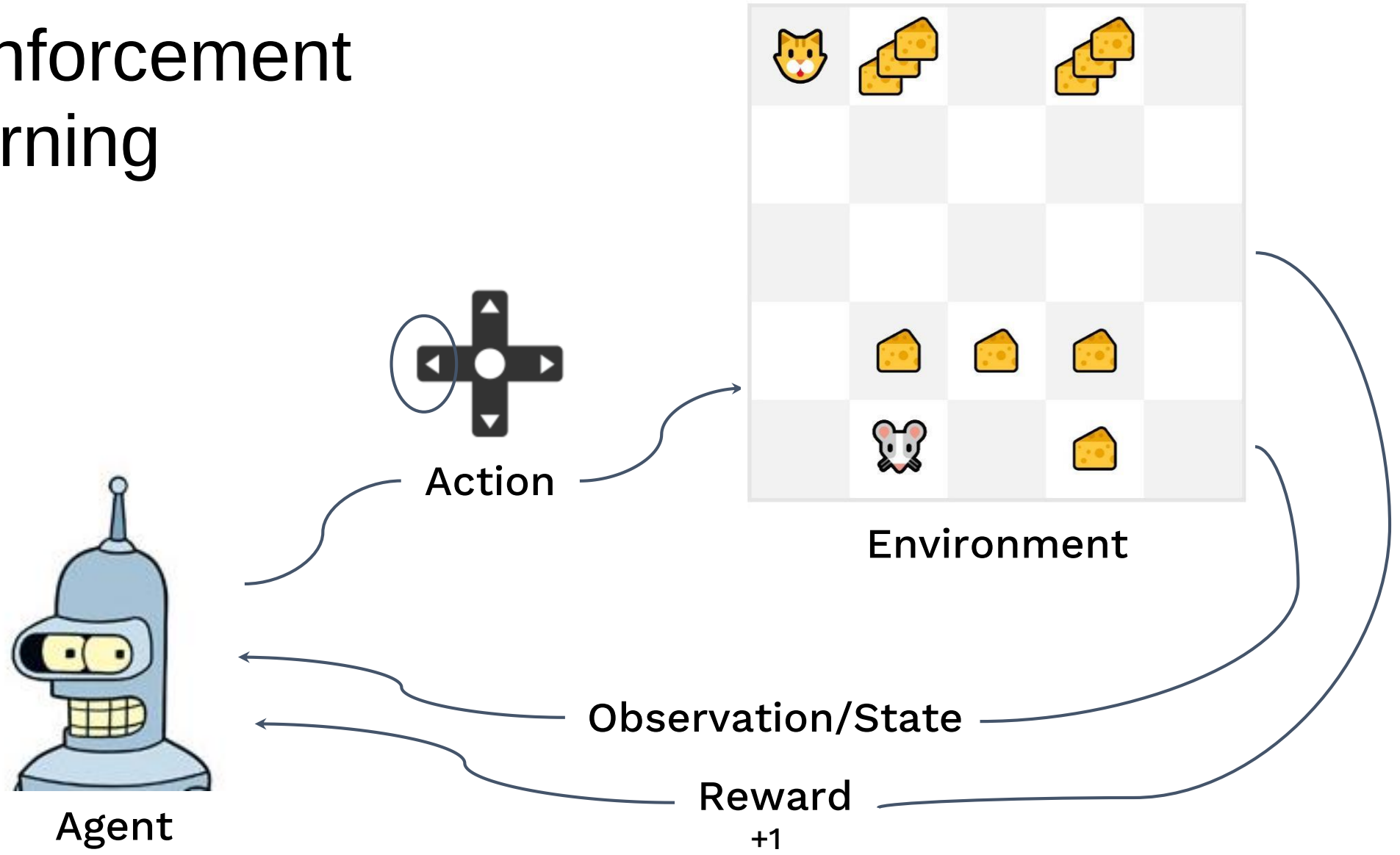
Dennis Psaroudakis





Koch *et al.* in preparation

Reinforcement Learning

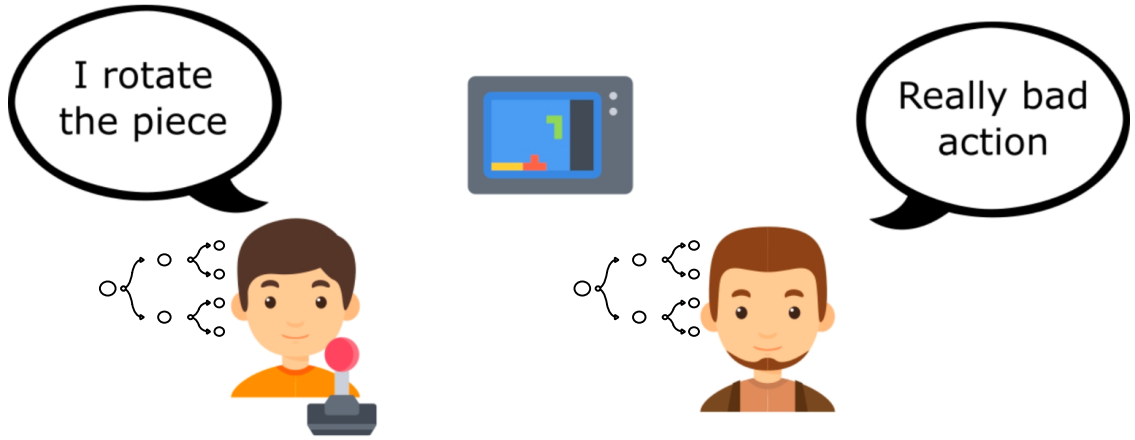


Our Agent

Action Space

- Grow Leaves
- Grow Stem
- Grow Roots
- Produce Starch
- Open Stomata
- Close Stomata
- Buy new Leaf
- Buy new Branch
- Buy new Root
- Buy new Flower

Algorithm: Proximal Policy Optimization (PPO)



Both see observation

learns from Critic

learns from rewards

Agent plays multiple different environments in parallel

PPO Algorithm: Schulman et al. 2017. Implementation: Stable Baselines 3, Raffin et al. 2021

Level 1

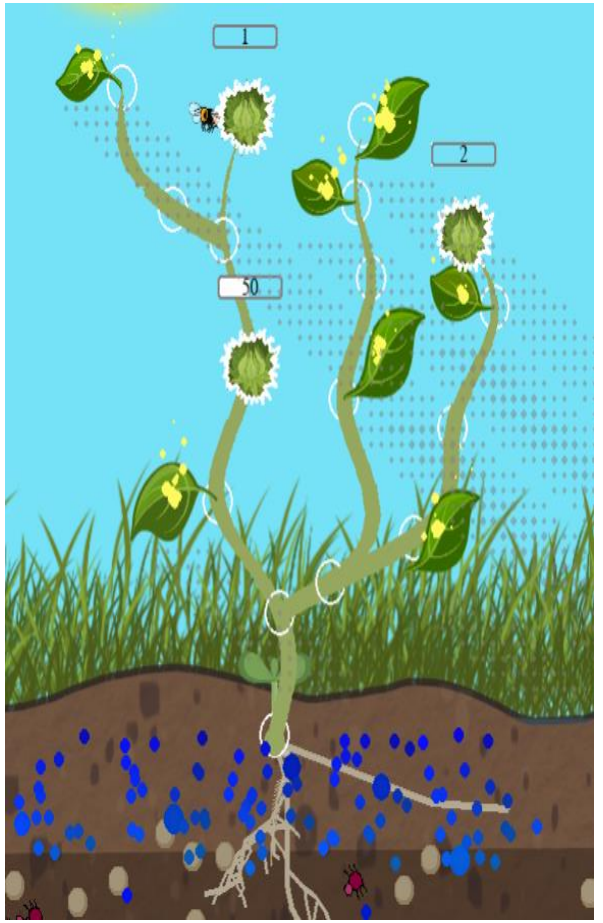
Just grow!

Reward Function

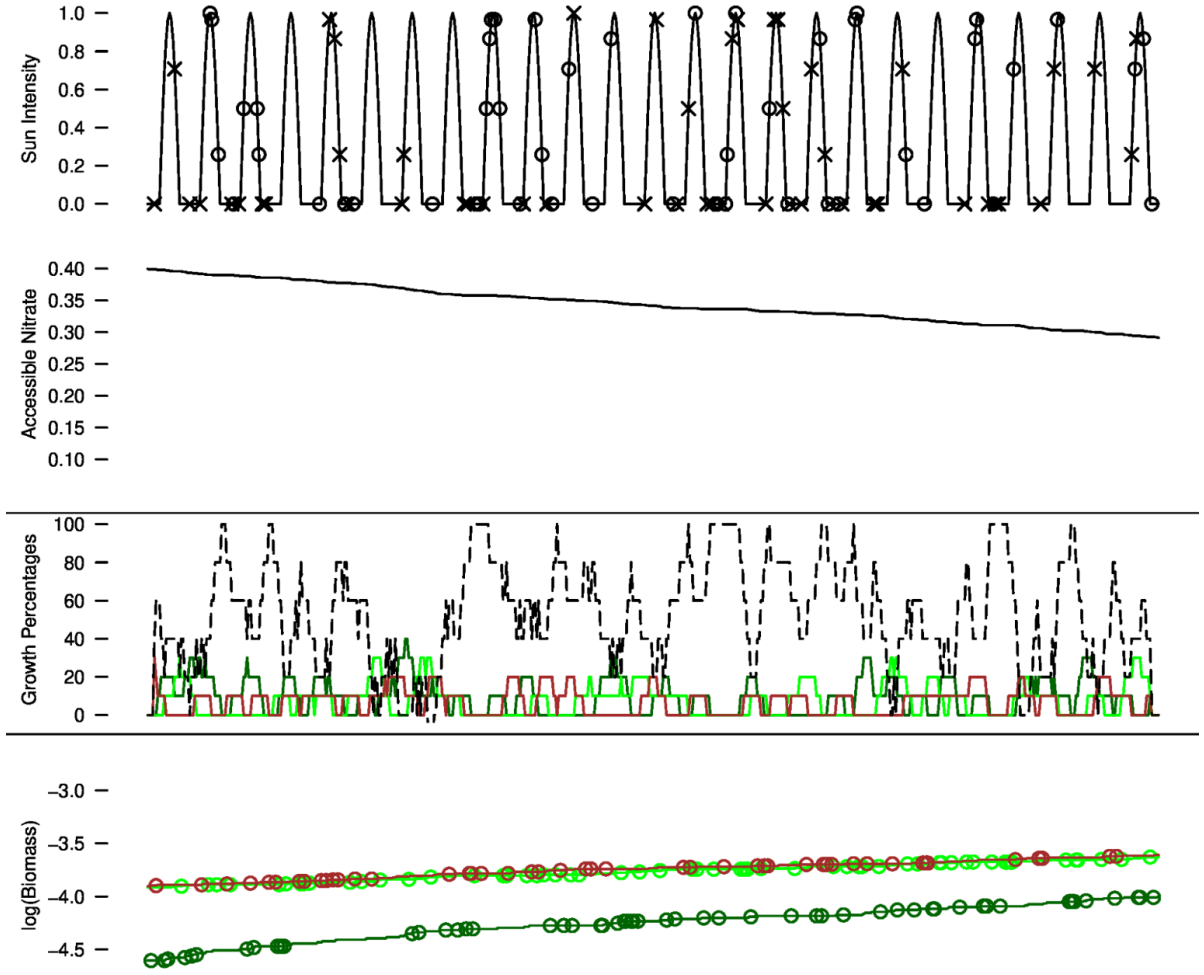
Difference in total biomass (g) since last step

Action Space

- Grow Leaves
- Grow Stem
- Grow Roots
- Produce Starch
- Open Stomata
- Close Stomata
- Buy new Leaf
- Buy new Branch
- Buy new Root
- Buy new Flower



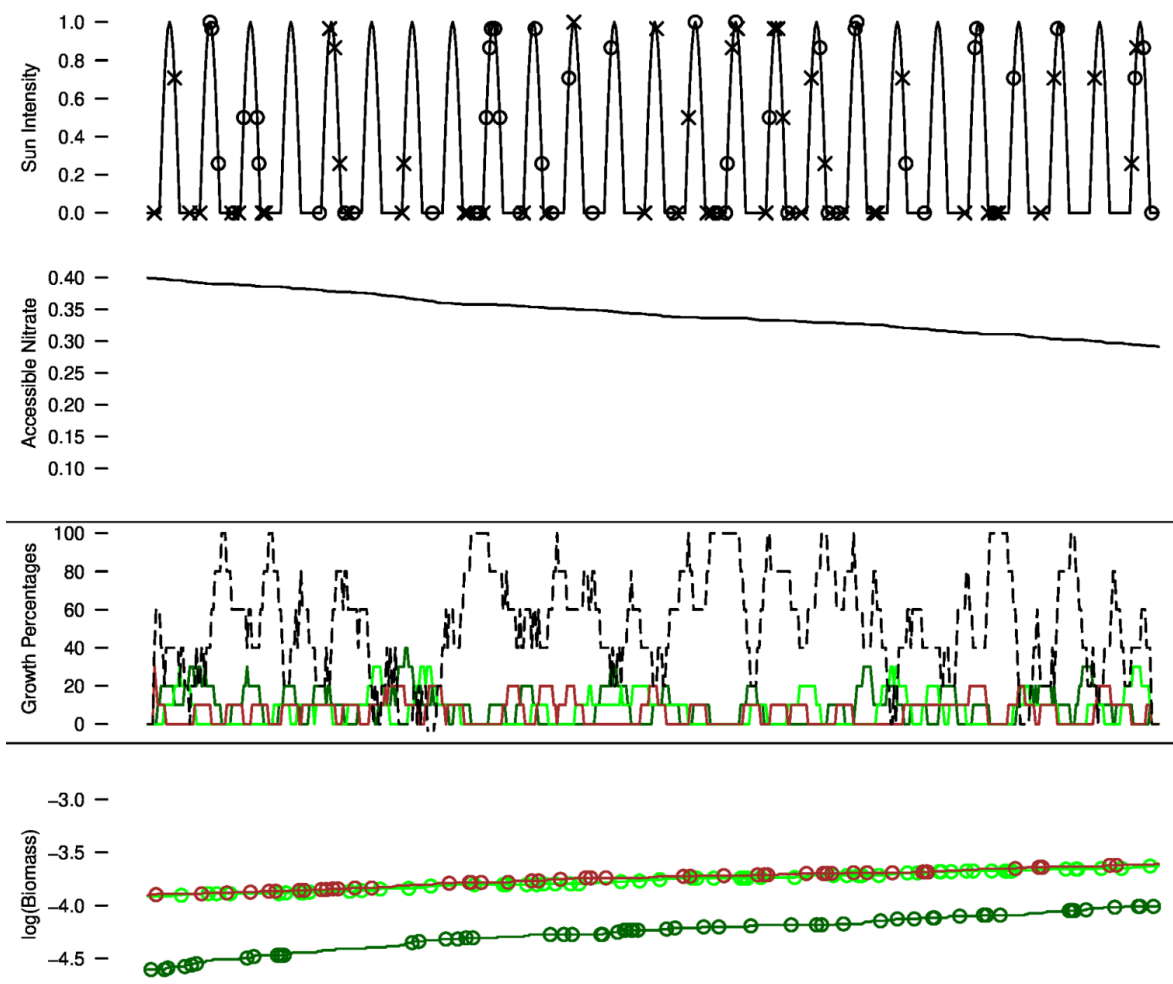
Low Nitrate



Episode 1

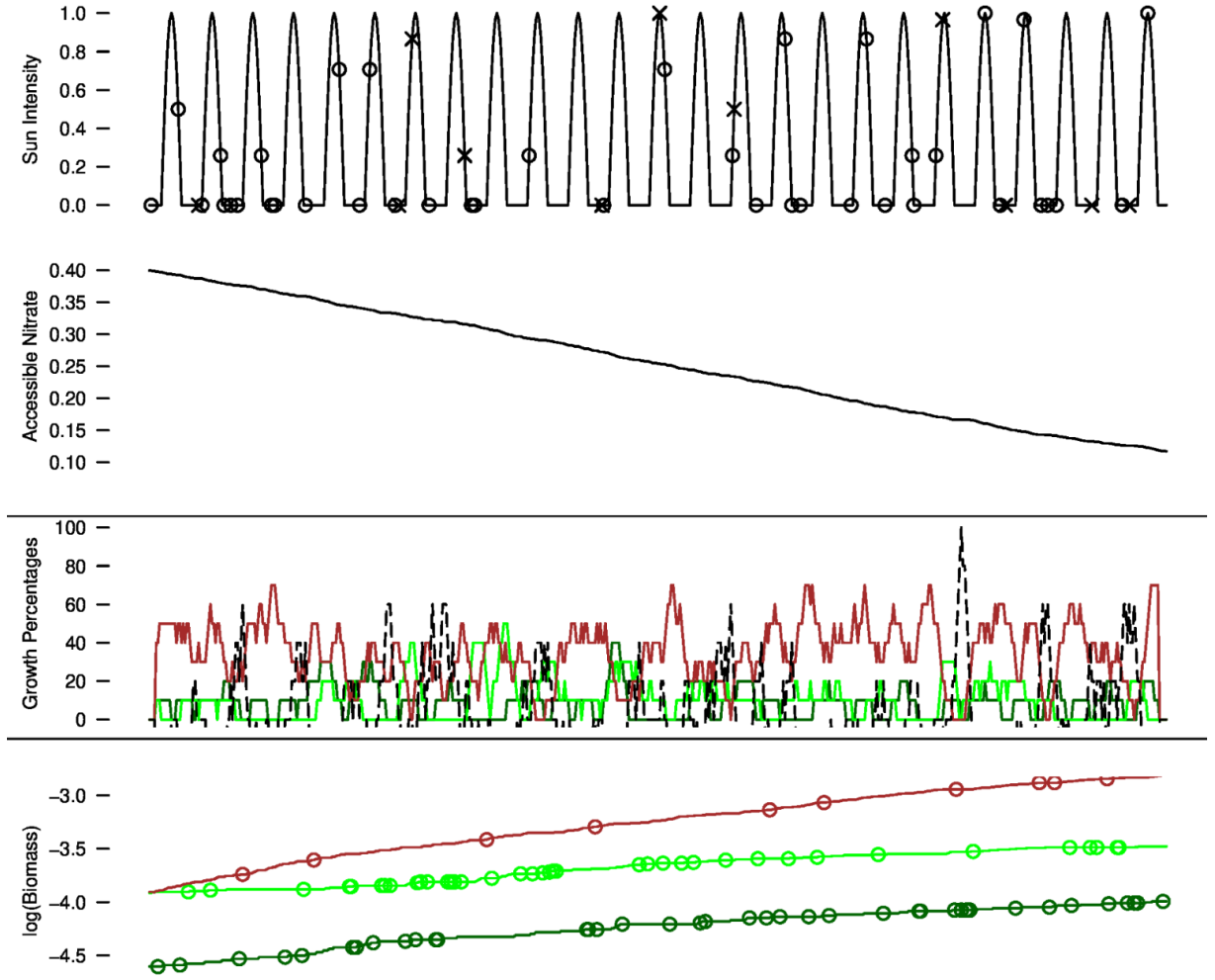
root
stem
leaf

Low Nitrate



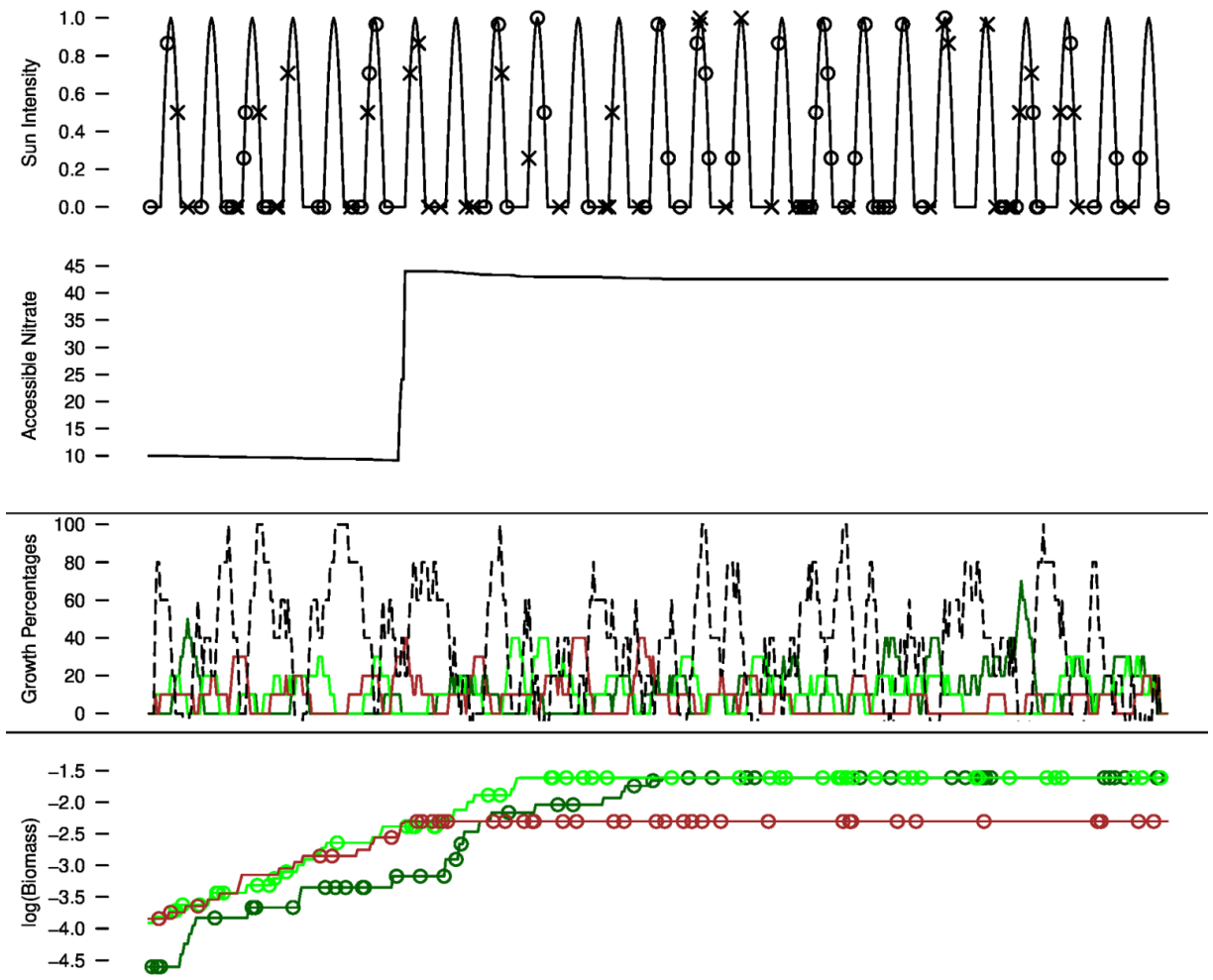
Episode 1

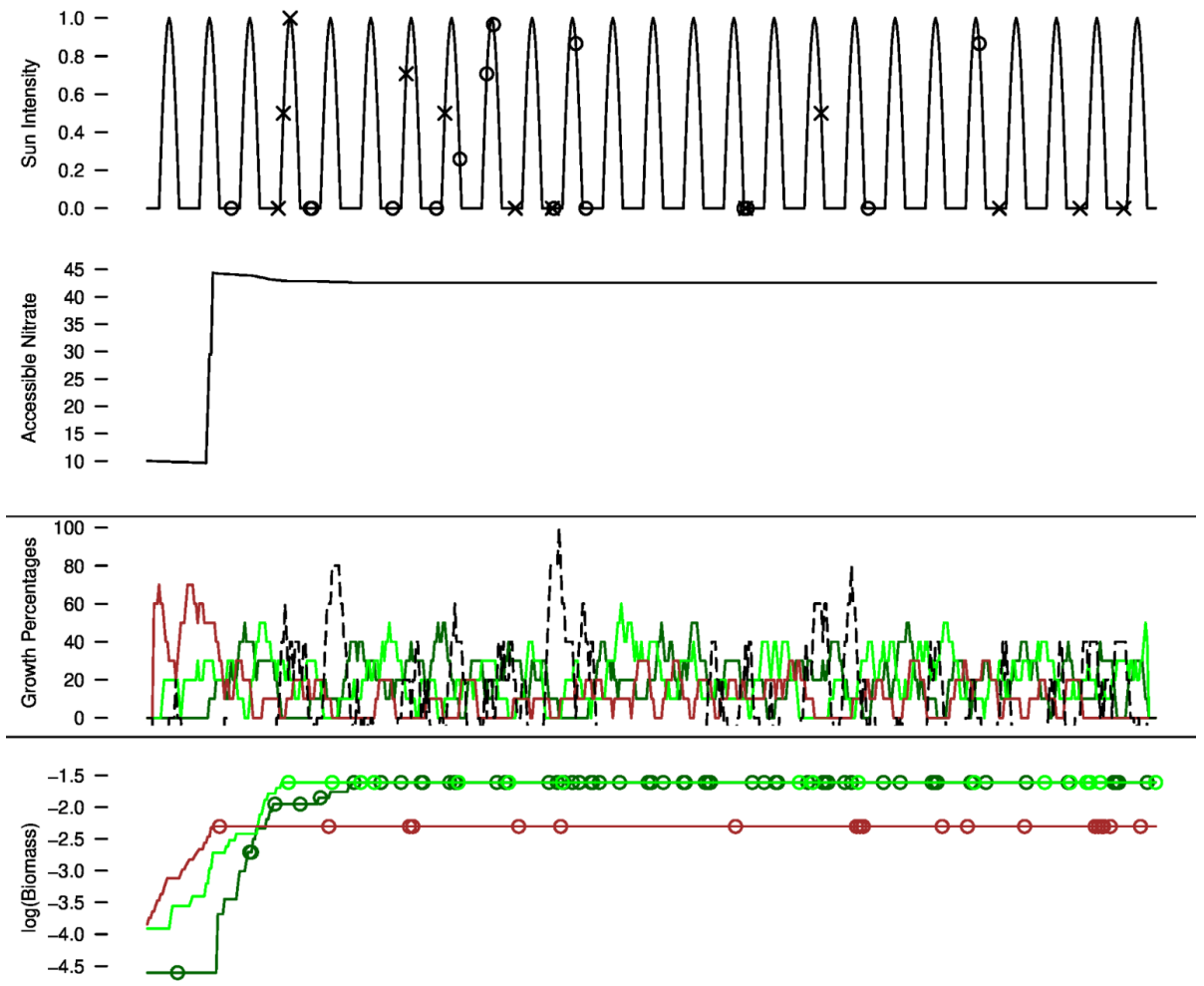
root
stem
leaf



Episode 9

High Nitrate





Episode 1

root
stem
leaf

Episode 15

Let’s go to Seed!

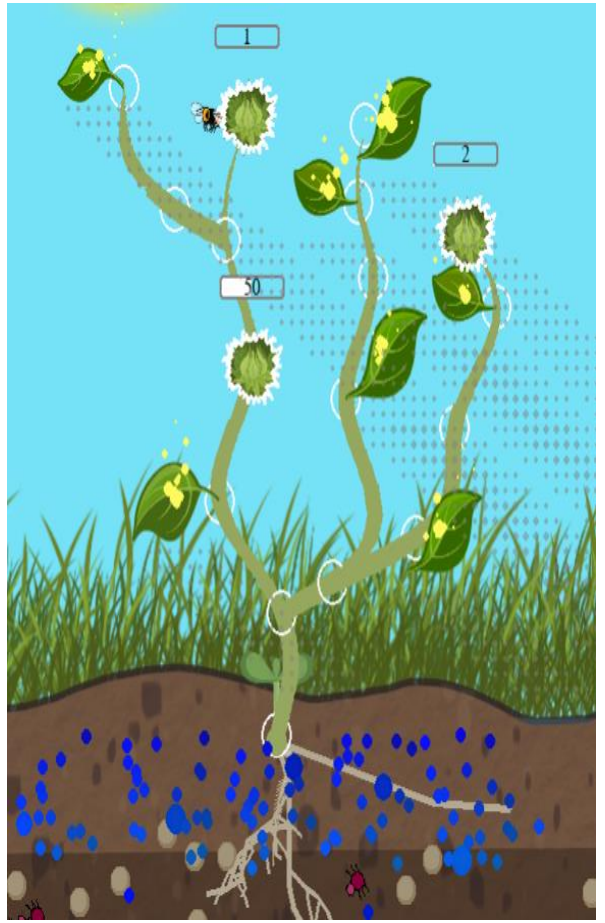
Level 2

Double Reward for seed mass!

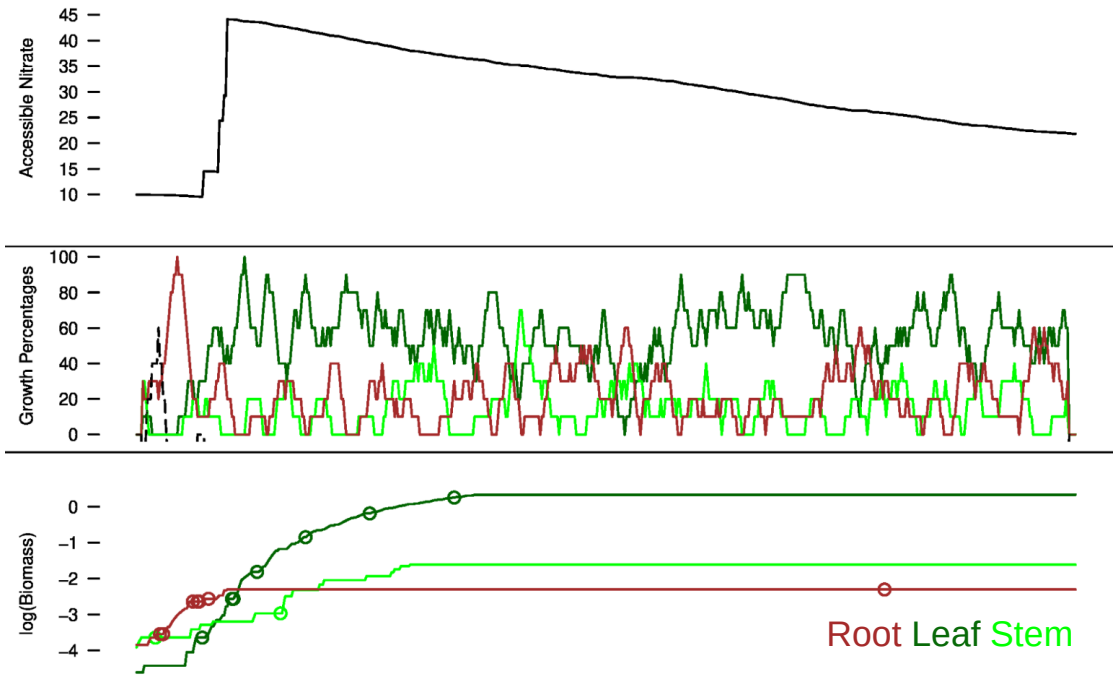
Penalties (in relation to current biomass)

Action Space

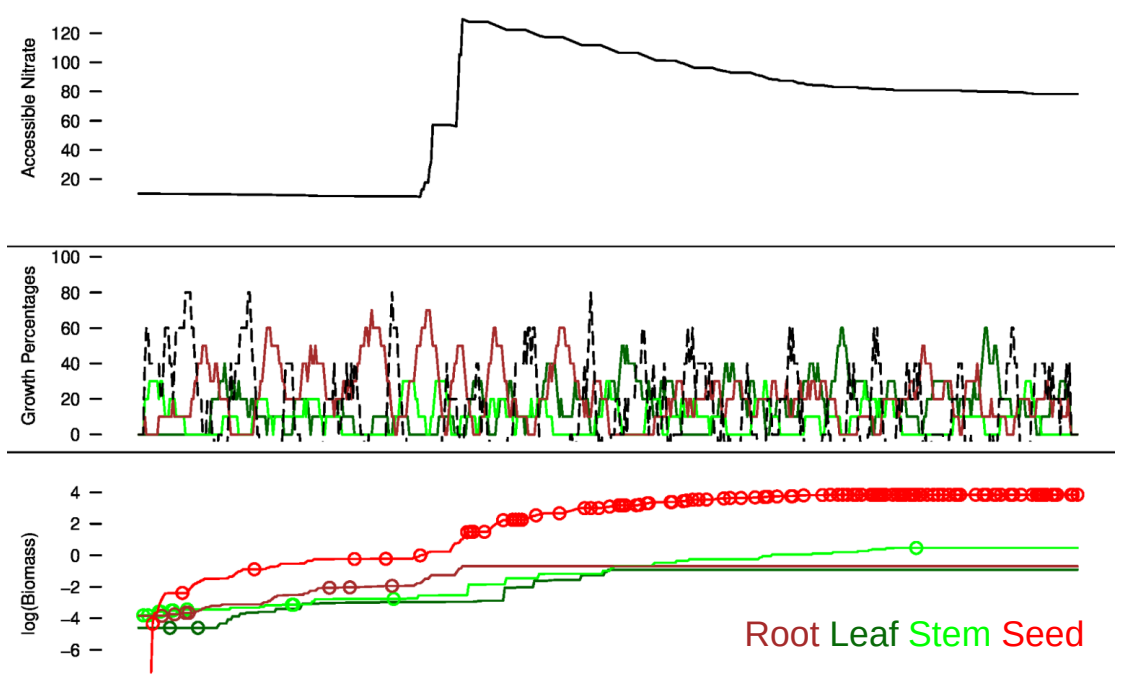
- Grow Leaves
- Grow Stem
- Grow Roots
- Produce Starch
- Open Stomata
- Close Stomata
- Buy new Leaf
- Buy new Branch
- Buy new Root
- Buy new Flower



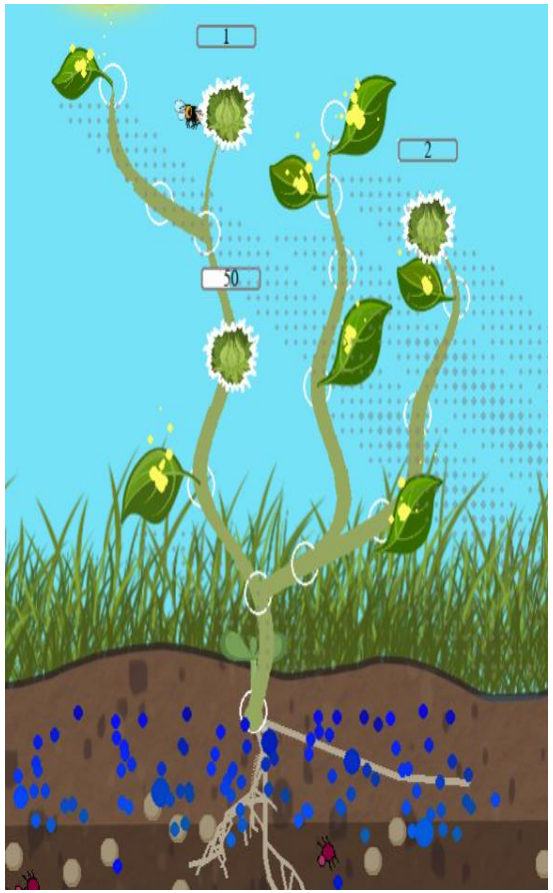
Spring (High Nitrate) max biomass



Spring (High Nitrate) max seed production



A digital twin

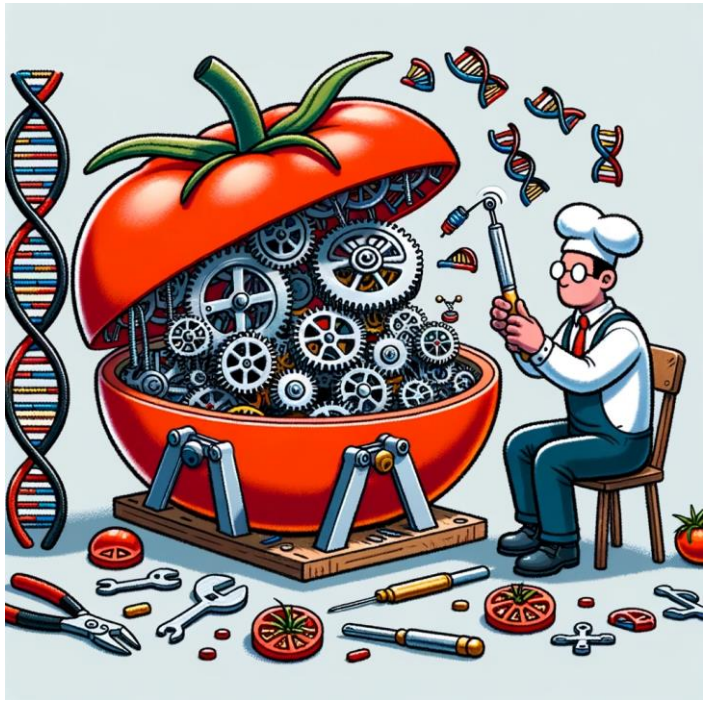


- Light intensity
- Light spectrum
- Temperature
- Humidity
- Water availability
- Nutrients
- .
- .
- CO_2
- O_2
- Time of the day
- Day/night cycle
- Season
- etc.
- .
- .
- .

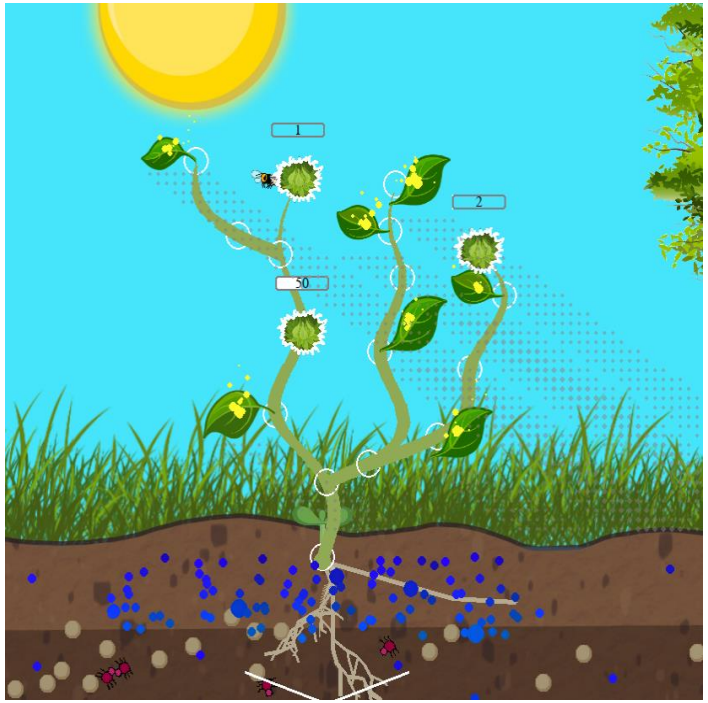
Summary

AI for detective work

AI for reverse engineering



AI for computer games



Thanks to:

Simon Zumkeller

Fritz Peleke

Dennis Psaroudakis

Ankur Sahu

Merle Stein

Gernot Schmitz

Daniel Koch

Kalyan Pininti

Sana Sangeen

Adeboye Adejoro

Max Opitz

Uni Cologne
Nadine Töpfer

Uni Cologne
Julietter de Meaux

MPIPZ @Cologne
Thomas Hartwig and **Julia Engelhorn**

IPK Gatersleben
Jozefus Schippers

FZ Jülich
HHU
Björn Usadel

..and many many others!

Thank you for your attention!

@NAMlab

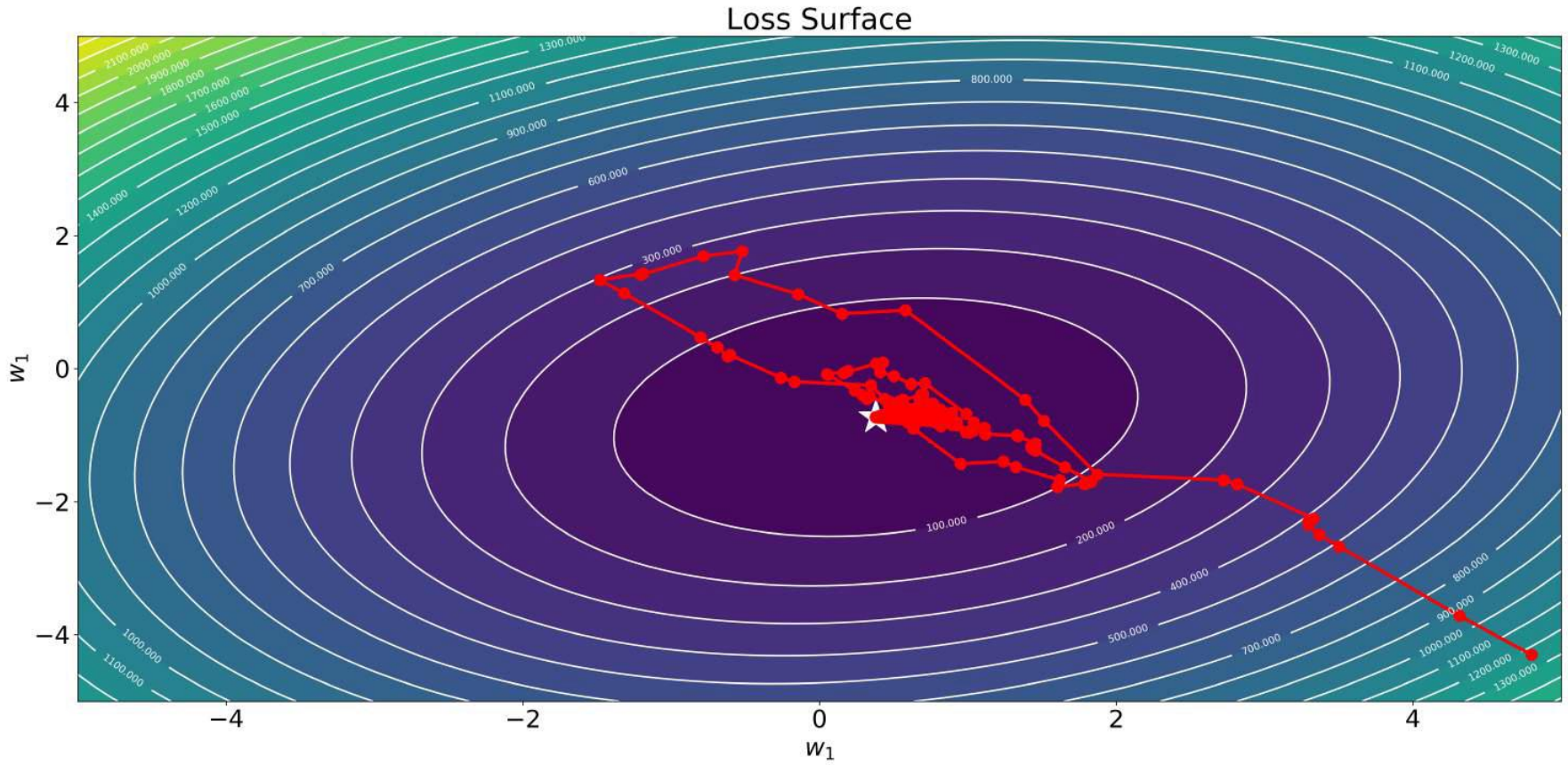
www.szymanskilab.com

szymanski@ipk-gatersleben.de

WE ARE HIRING!

postdoc & PhD positions in machine learning and gene regulation

Gradient descent – tracking the learning progress



AI

Dennis Psaroudakis
IPK Gatersleben
FZ Jülich

Growth
Plant Architecture
Morphogenesis
Stomata
Pathway regulation
Defense
etc

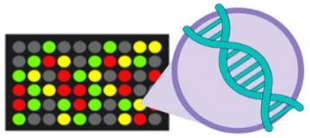
Light
CO2
Water
O2
NO3
etc...

Metabolic network

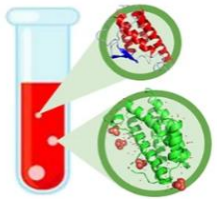
CO2
O2
Biomass

Dynamic growth, environment, interaction

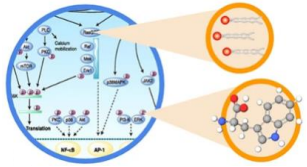
What kind of machine learning / AI?



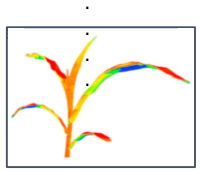
Transcriptomics
~ 100k transcripts



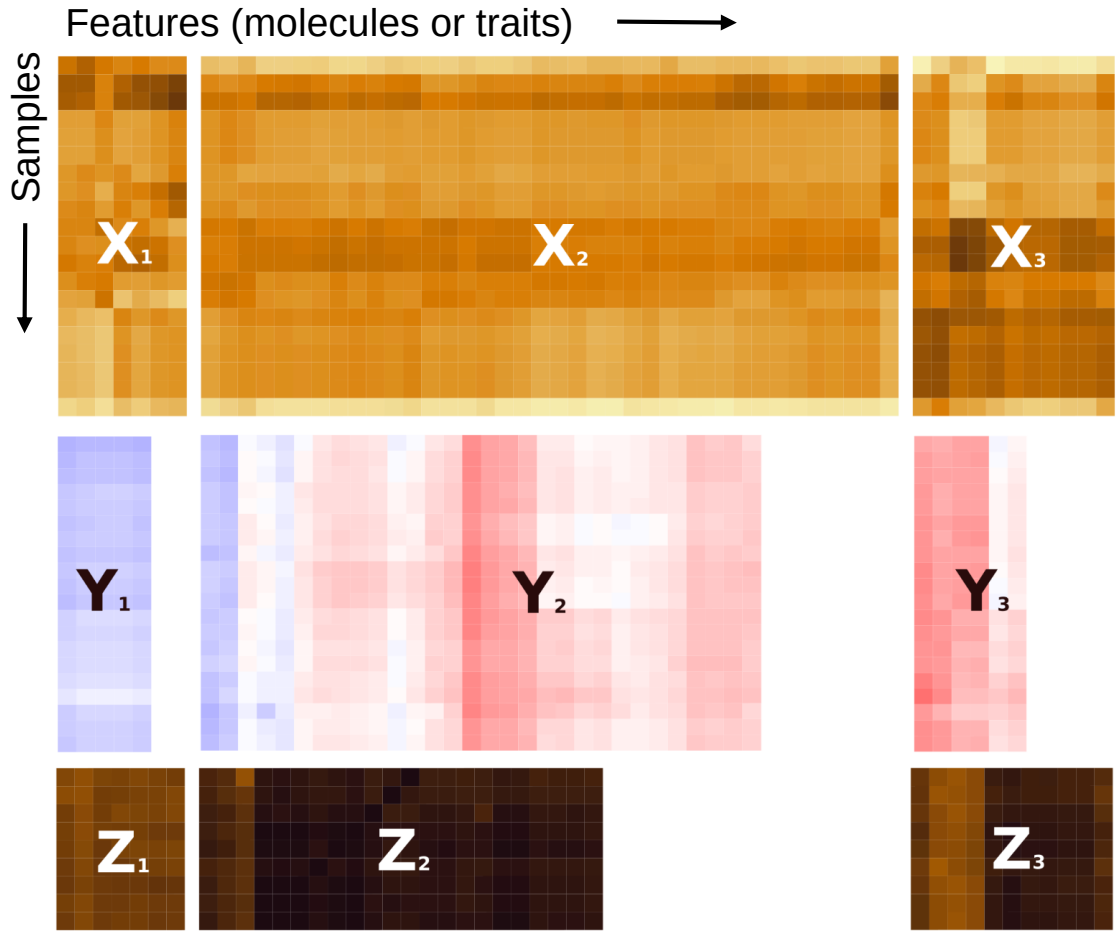
Proteomics
~ 10k proteins



Metabolomics
~ 3k metabolites



Phenomics
~ 300 phenotypes



AI

Dennis Psaroudakis
IPK Gatersleben
FZ Jülich

Growth
Plant Architecture
Morphogenesis
Stomata
Pathway regulation
Defense
etc

Light
CO2
Water
O2
NO3
etc...

Metabolic network

CO2
O2
Biomass

Dynamic growth, environment, interaction

AI

Done

Dennis Psaroudakis
IPK Gatersleben
FZ Jülich

Growth
Plant Architecture
Morphogenesis
Stomata
Pathway regulation
Defense
etc

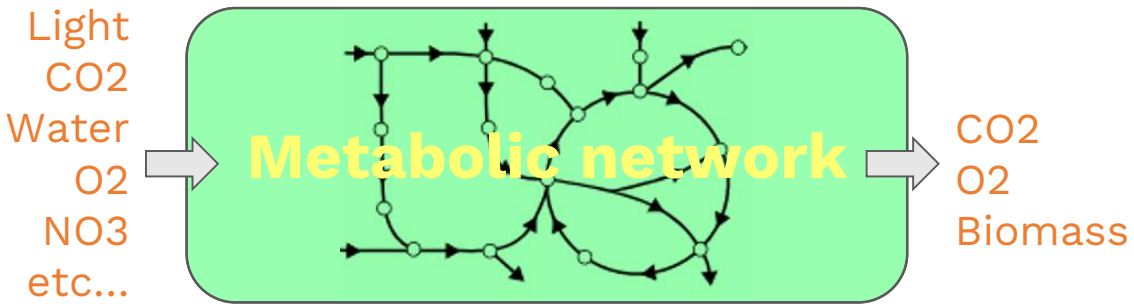
Light
CO2
Water
O2
NO3
etc...

Metabolic network

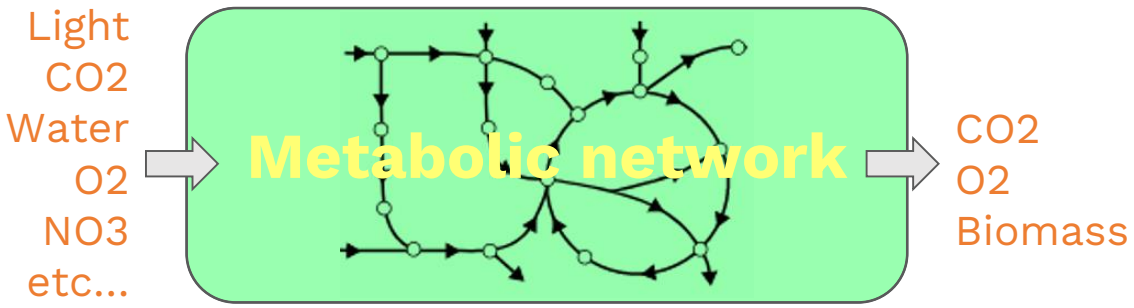
CO2
O2
Biomass

Dynamic growth, environment, interaction

The challenge

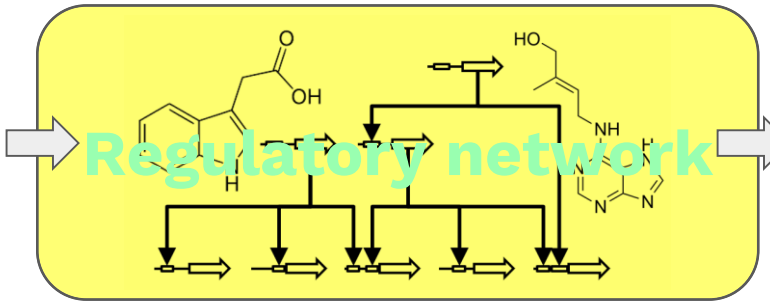


The challenge



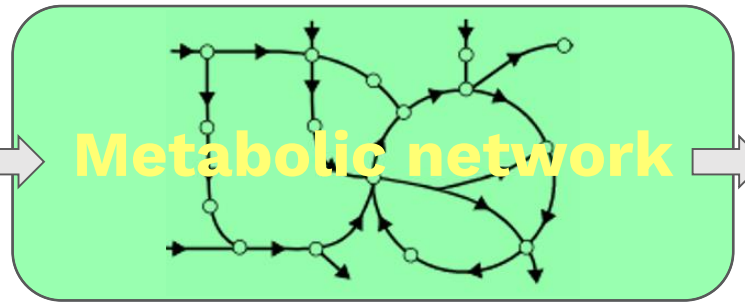
The challenge

TFs
Phytohormones
Membranes
Phytochromes
Small RNAs
etc



Growth
Plant Architecture
Morphogenesis
Stomata
Pathway regulation
Defense
etc

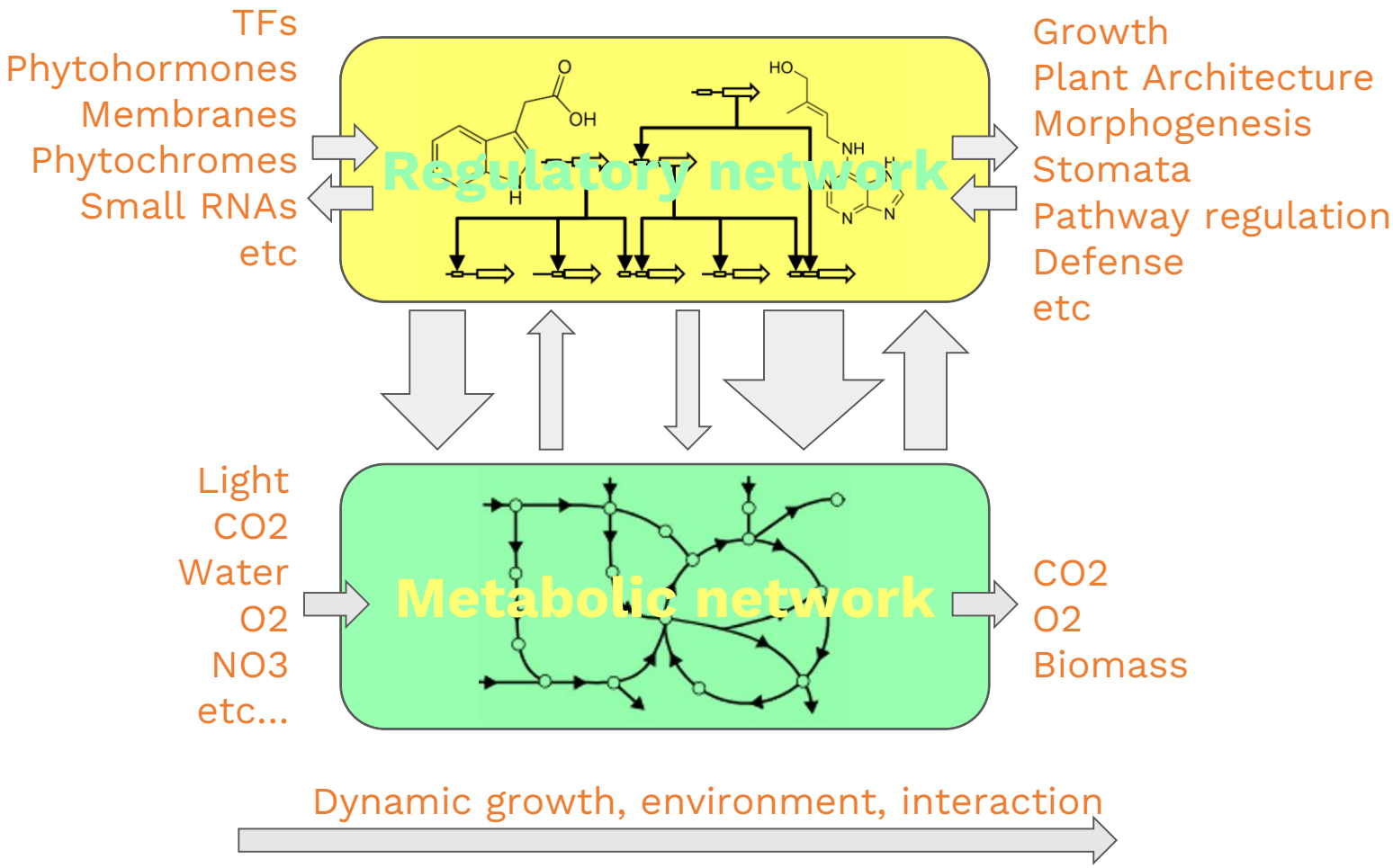
Light
CO2
Water
O2
NO3
etc...



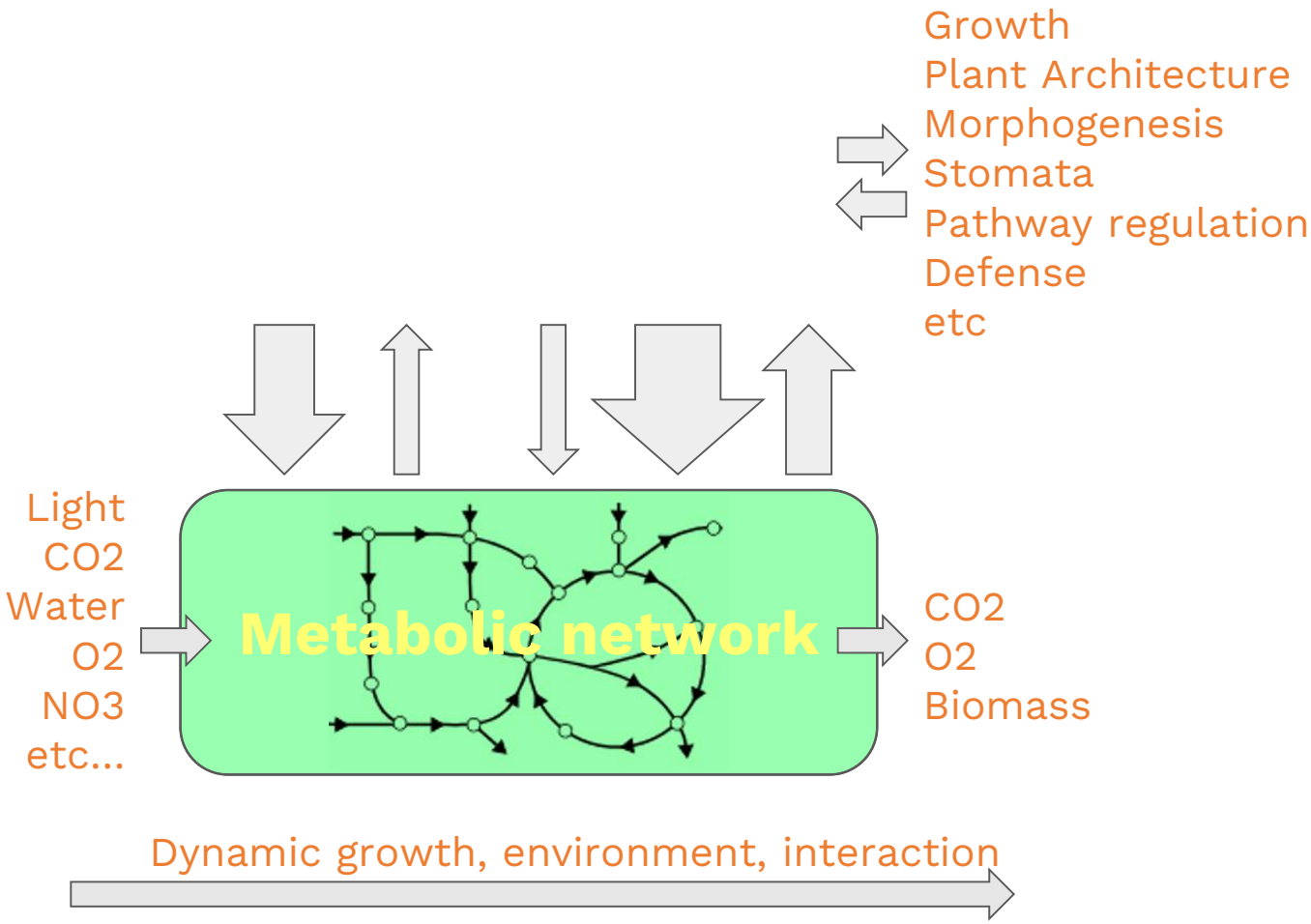
CO2
O2
Biomass

Dynamic growth, environment, interaction

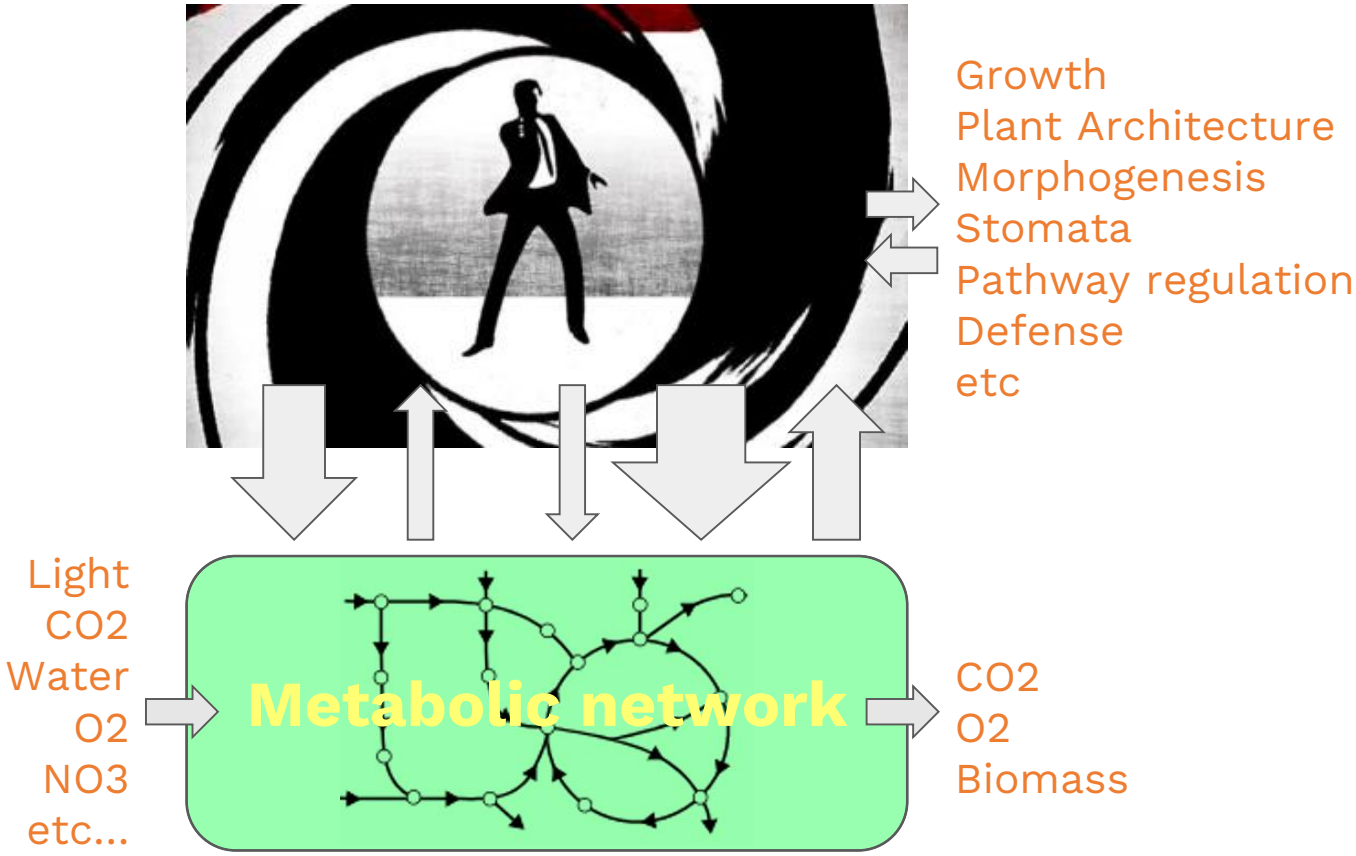
The challenge



The challenge



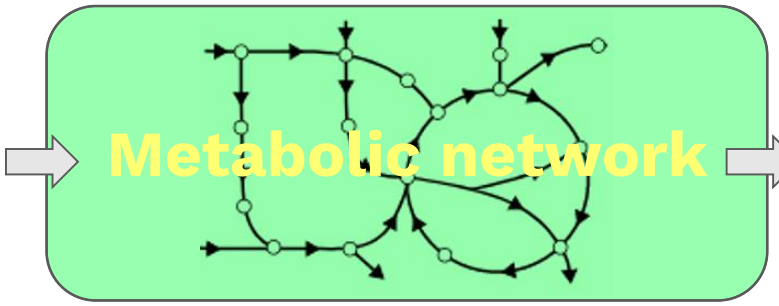
Agent


Growth
Plant Architecture
Morphogenesis
Stomata
Pathway regulation
Defense
etc
Light
CO2
Water
O2
NO3
etc...
Metabolic network
CO2
O2
Biomass
Dynamic growth, environment, interaction

Agent

Growth
Plant Architecture
Morphogenesis
Stomata
Pathway regulation
Defense
etc

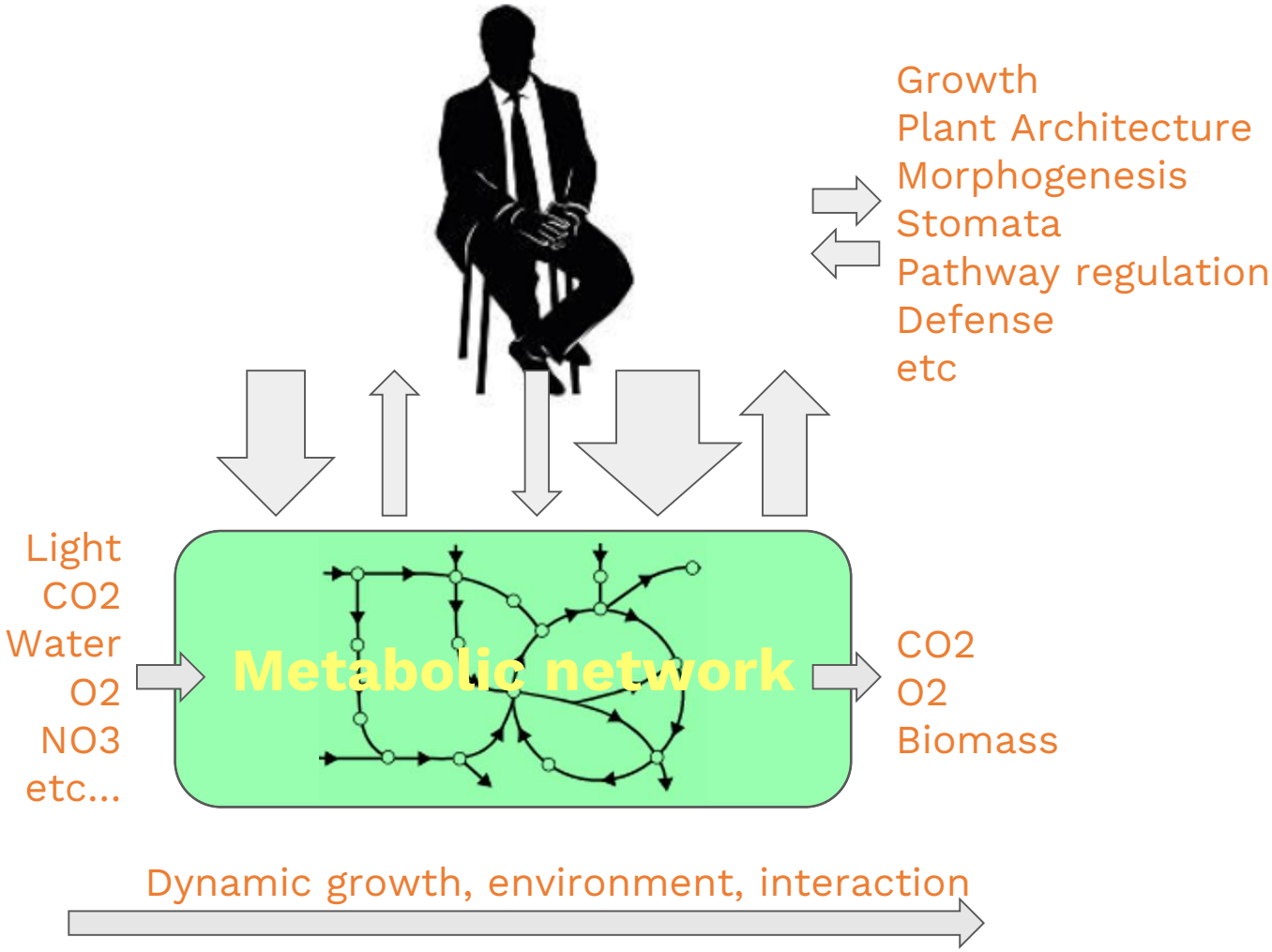
Light
CO2
Water
O2
NO3
etc...

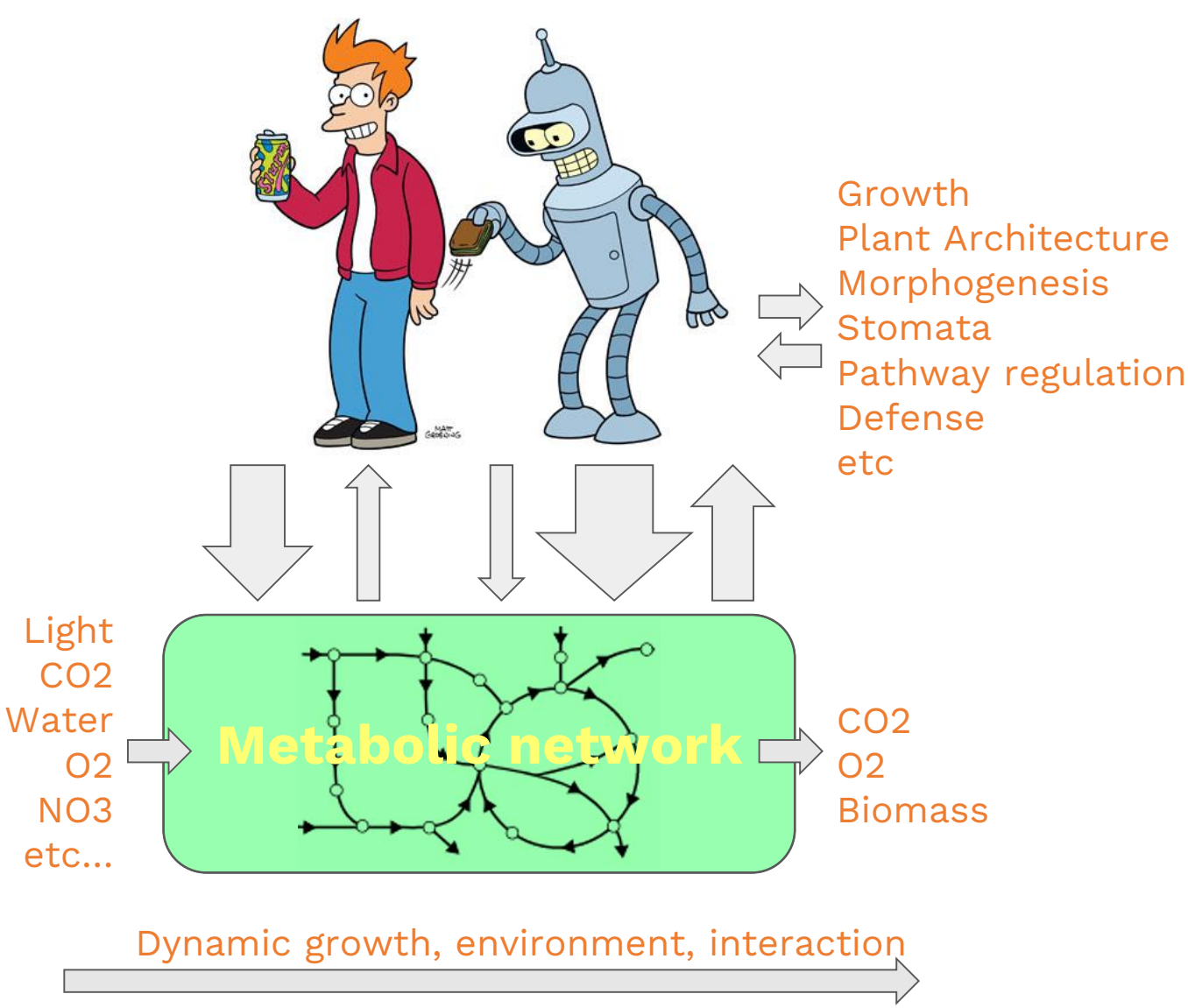


CO2
O2
Biomass

Dynamic growth, environment, interaction

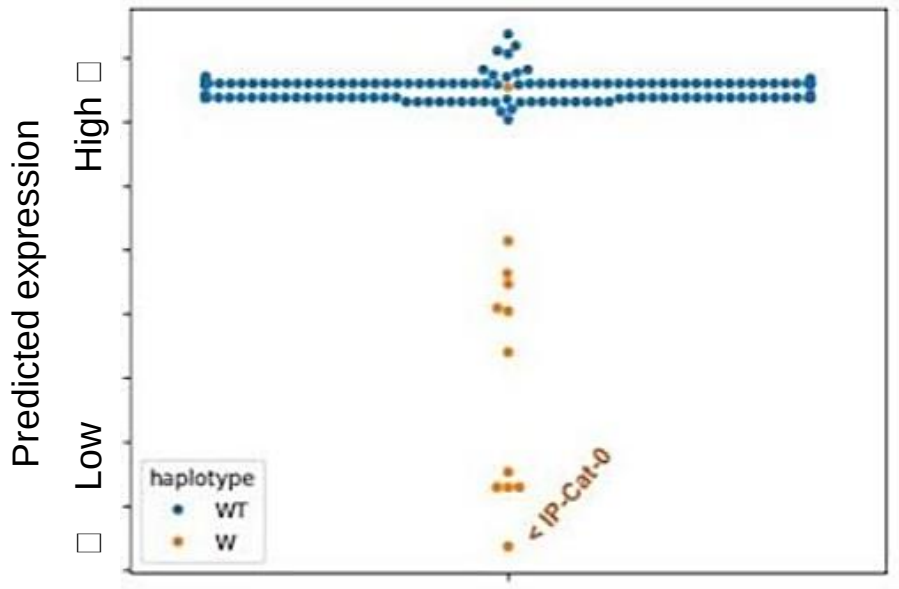
Agent


Growth
Plant Architecture
Morphogenesis
Stomata
Pathway regulation
Defense
etc
Light
CO2
Water
O2
NO3
etc...
Metabolic network
CO2
O2
Biomass
Dynamic growth, environment, interaction


Human
AI
Growth
Plant Architecture
Morphogenesis
Stomata
Pathway regulation
Defense
etc
Light
CO2
Water
O2
NO3
etc...
Metabolic network
CO2
O2
Biomass
Dynamic growth, environment, interaction

Regulatory gene variants

SNP effects on expression of key regulatory genes



Example gene RAP2.2 in *A. thaliana*

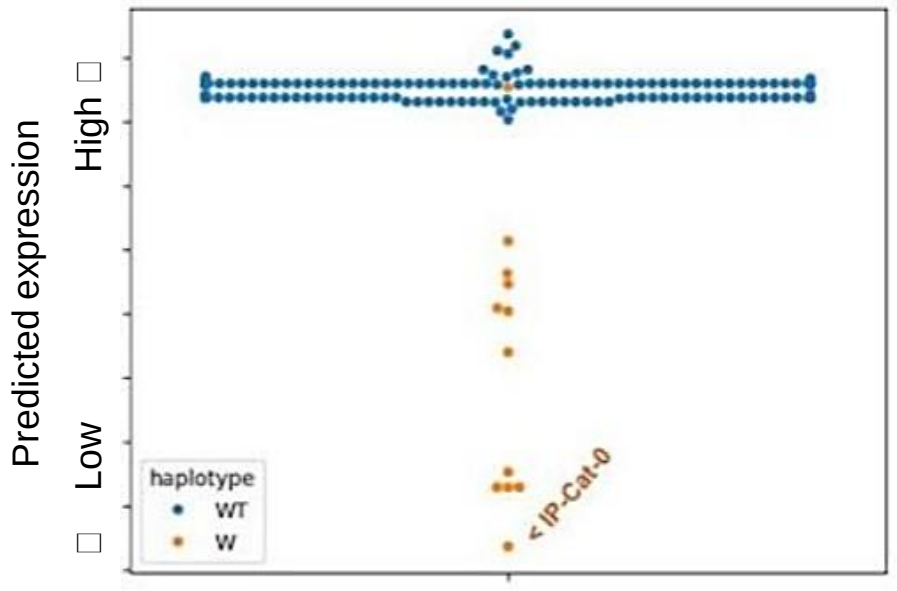
Related to SUB1A in rice

Responsible for survival during hypoxia / flooding

Gernot Schmitz
(PhD Student)

Regulatory gene variants

SNP effects on expression of key regulatory genes



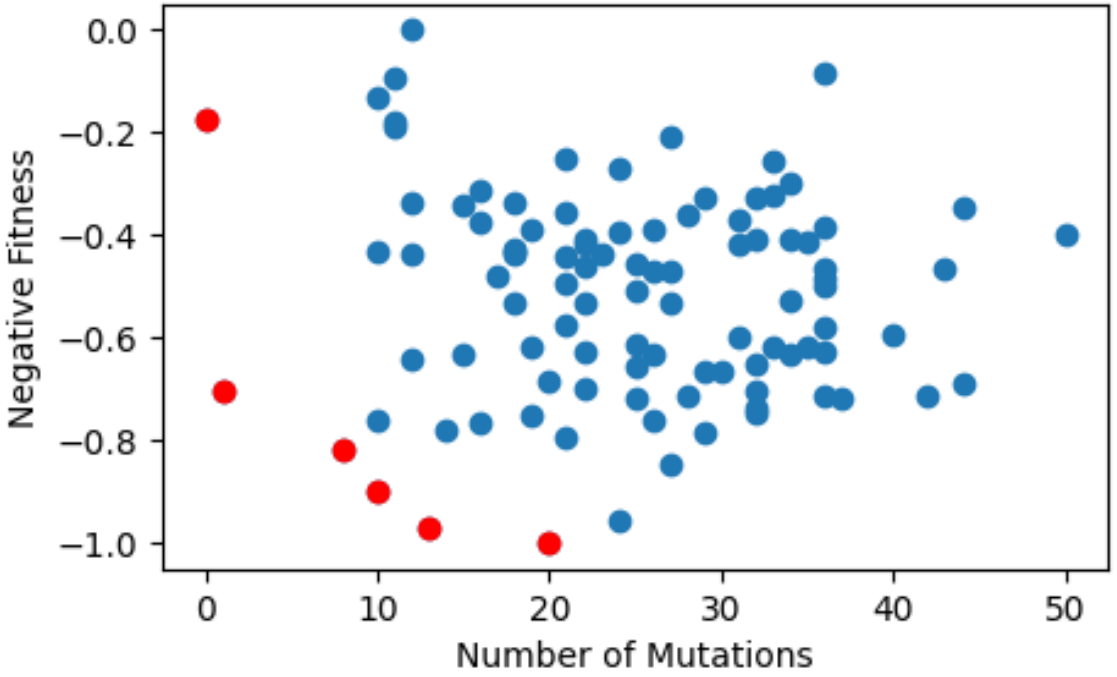
Example gene RAP2.2 in *A. thaliana*

Related to SUB1A in rice

Responsible for survival during hypoxia / flooding

Promoter optimization

Smart gene editing for max effect with minimum mutations

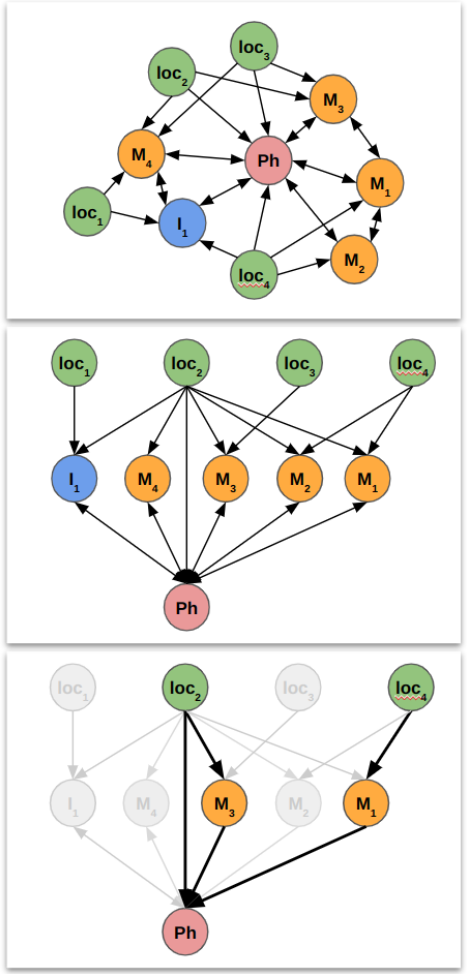


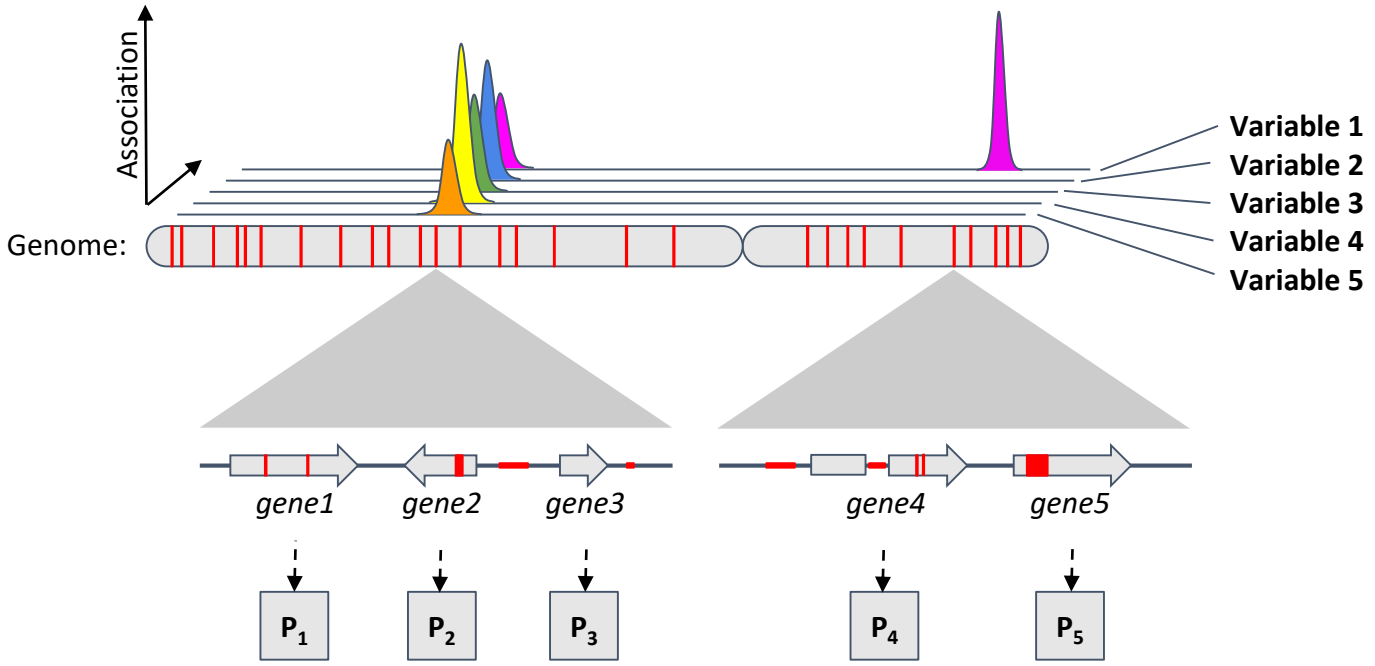
Gernot Schmitz
(PhD Student)

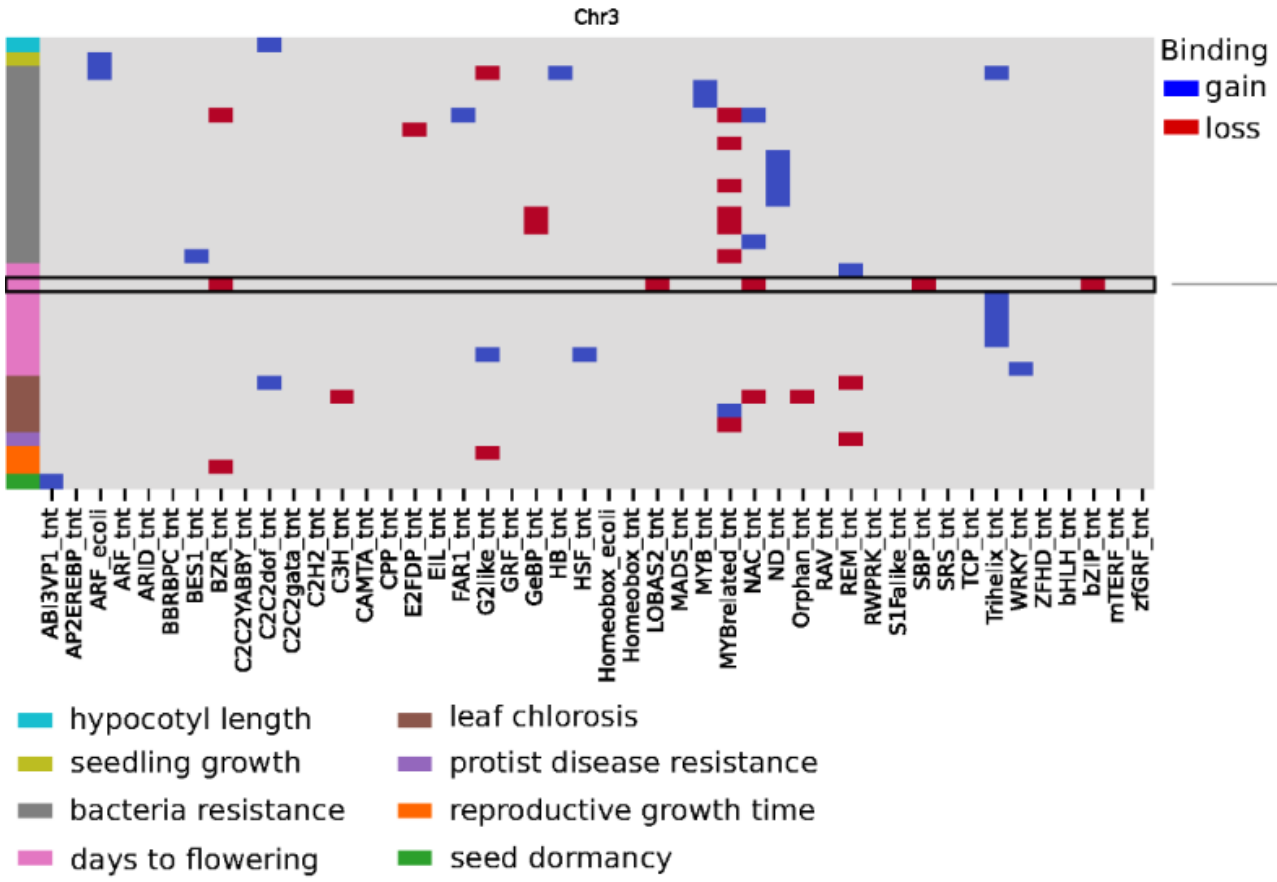
Genomic networking

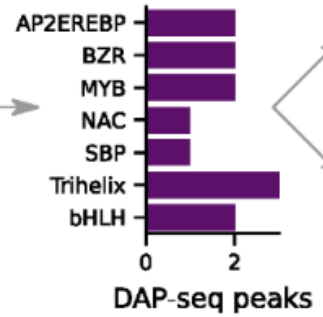
Genomic networking

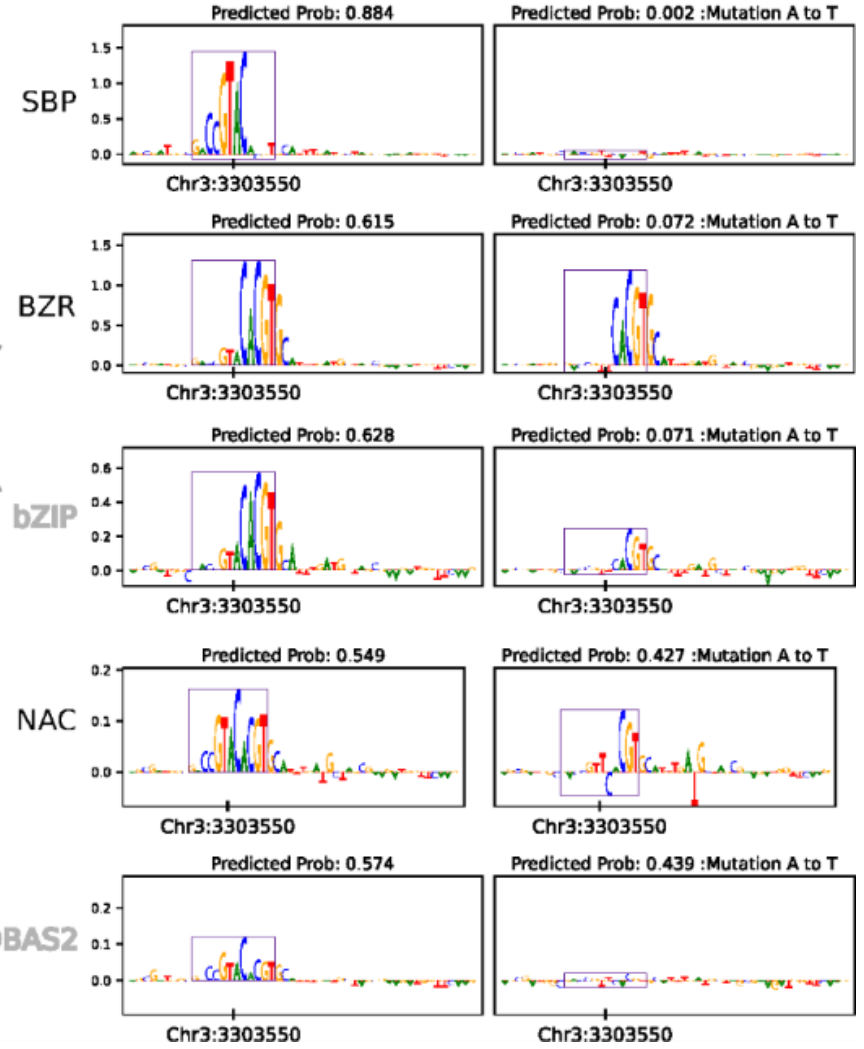
multivariate GWAS, path analysis, mixed graphical models






Chr3
Binding
gain
loss
ABI3VP1_tnt
AP2EREBP_tnt
ARF_ecoli
ARF_tnt
ARID_tnt
BBRBPC_tnt
BES1_tnt
BZR_tnt
C2C2YABBY_tnt
C2C2dof_tnt
C2C2gata_tnt
C2H2_tnt
C3H_tnt
CAMTA_tnt
CPP_tnt
E2FDP_tnt
EIL_tnt
FAR1_tnt
G2like_tnt
GRF_tnt
GeBP_tnt
HB_tnt
HSF_tnt
Homeobox_ecoli
Homeobox_tnt
LOBAS2_tnt
MADS_tnt
MYB_tnt
MYBrelated_tnt
NAC_tnt
ND_tnt
Orphan_tnt
RAV_tnt
REM_tnt
RWPRK_tnt
S1Falike_tnt
SBP_tnt
SRS_tnt
TCP_tnt
Trihelix_tnt
WRKY_tnt
ZFHD_tnt
bHLH_tnt
bZIP_tnt
mTERF_tnt
zfGRF_tnt
hypocotyl length
seedling growth
bacteria resistance
days to flowering
leaf chlorosis
protist disease resistance
reproductive growth time
seed dormancy


AP2EREBP
BZR
MYB
NAC
SBP
Trihelix
bHLH
0
2
DAP-seq peaks


SBP
Predicted Prob: 0.884
Predicted Prob: 0.002 :Mutation A to T
BZR
Predicted Prob: 0.615
Predicted Prob: 0.072 :Mutation A to T
bZIP
Predicted Prob: 0.628
Predicted Prob: 0.071 :Mutation A to T
NAC
Predicted Prob: 0.549
Predicted Prob: 0.427 :Mutation A to T
LOBAS2
Predicted Prob: 0.574
Predicted Prob: 0.439 :Mutation A to T
Chr3:3303550

1001 Genomes

AraGWAS Catalog 36878 phenotype associations

SNP / SV

↓

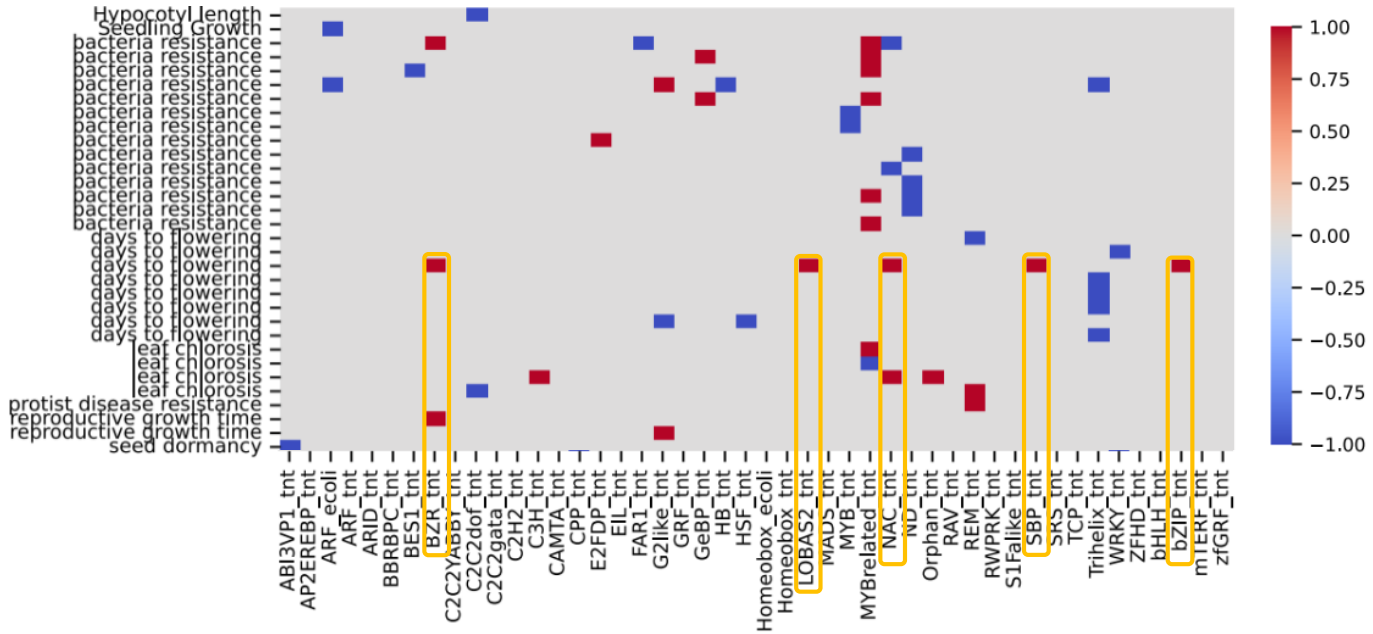
Δ TF binding

↓

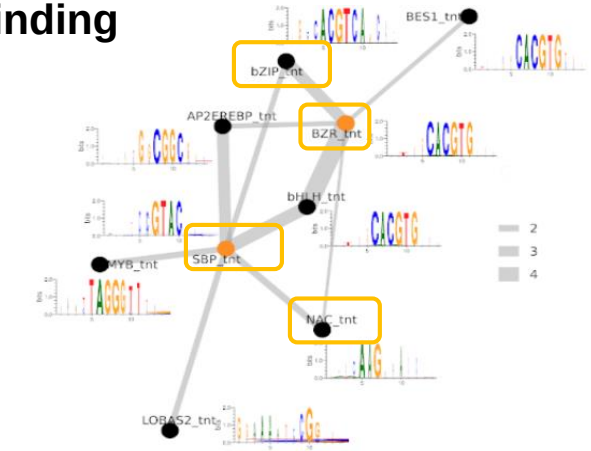
Δ Gene regulation

↓

Δ Phenotype



Combinatorial TF binding



131