IRC 2023; Sydney, `Omics Technology´ Session, 27.09.2023

# Advanced Virtuality and Augmented Reality Approaches in Seeds to Seeds

Thomas Altmann, Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK),

Discussion 26.06.2023

AVATARS

# Advanced Virtuality and Augmented Reality Approaches in Seeds to Seeds

Phase I (VI/2019 – III/2022) ➢ **Phase II (IV/2022 – IX/2024)**

Thomas Altmann[1], Amine Abbadi[2], Matthias Enders[2], Marius Kloft[3], Andrea Bräutigam[4], Olaf Sacher[5], Ljudmilla Borisjuk[1], Hardy Rolletschek[1], Jozephus Schippers[1], Hans-Peter Braun[6], Evgeny Gladilin[1], Mary-Ann Blätke[1], Matthias Lange[1]

**1** - Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, **2** - NPZ Innovation GmbH, Hohenlieth-Hof, **3** – Department of Computer Science, RPTU Kaiserslautern-Landau, **4** - Computational Biology, CeBiTEc, Bielefeld University, **5** - Breakpoint One GmbH, Berlin, **6** - Leibniz University Hannover

IPK LEIBNIZ INSTITUTE
NPZ INNOVATION
TECHNISCHE UNIVERSITÄT KAISERSLAUTERN
BREAK POINT ONE
Universität Bielefeld
Leibniz Universität Hannover
Federal Ministry of Education and Research

# Project Goal: Prediction and Explanation of Seed Traits

**AVATARS - harnessing the data explosion in life sciences towards enhanced gain of knowledge and facile utilization**

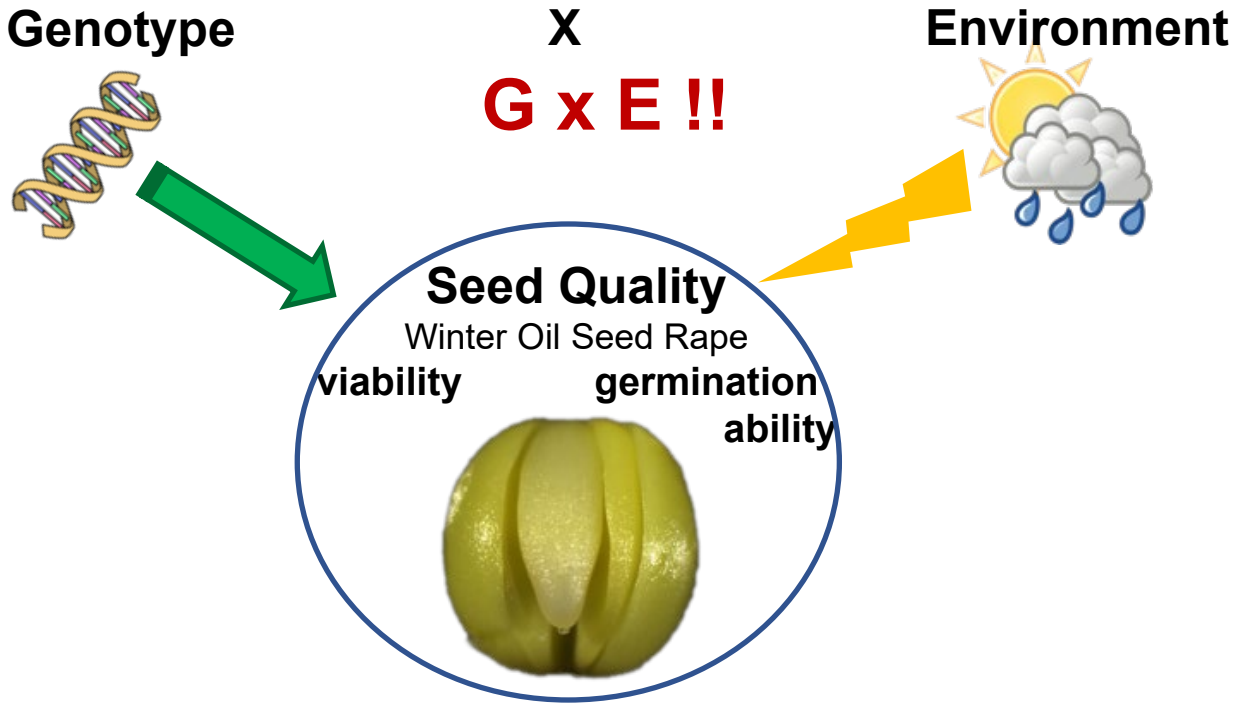
using
**Deep Learning** on high-throughput genotype, envirotype, and multilevel phenotype data

**and**

**mechanistic knowledge** gained from **Systems Analyses** using dynamic structural, compositional, and functional (multi-omics) data of seed formation.

**AVATARS – an integrative Data Science project**

# AVATARS – the central Question




## Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*

frontiers in Plant Science
ORIGINAL RESEARCH
published: 09 April 2015
doi: 10.3389/fpls.2015.00221

*Sarah V. Hatzig[1], Matthias Frisch[2], Frank Breuer[3], Nathalie Nesi[4], Sylvie Ducournau[5], Marie-Helene Wagner[5], Gunhild Leckband[6], Amine Abbadi[6] and Rod J. Snowdon[1]**

- **low - moderate correlation among 2 enviroments (SL2011 – SL2012):**
  GR72 (%) r = 0.44
- **very low narrow-sense heritability:**
  GR72 (%) $h^2$ (SL2011) = 0.07
  $h^2$ (SL2012) = 0.00

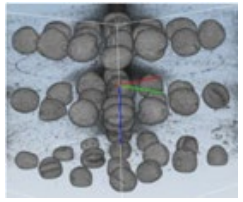
# AVATARS – the Interdisciplinary Concept

**Genotype** X **Environment**

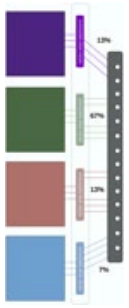
**G x E !!**

## Field Cultivation
**Plant & Evironmental Data**
**400/100 Lines**
**11 env.**
**genotype**
**15k SNPs**
**phenotype**
**weather data**

## Single Seed Data
**> 350.000 seeds**
**TD-NMR hyperspectral images**
**X-ray CT**
**germination**

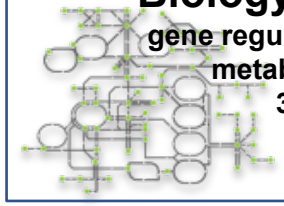


## Deep Learning
**neural networks using ultra high-throughput data**



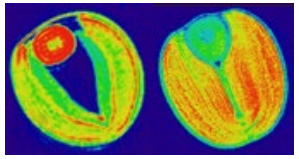
## Seed Quality
Winter Oil Seed Rape
**viability** **germination ability**

## Systems Biology
**gene regulatory network**
**metabolic networks**
**3D seed model**

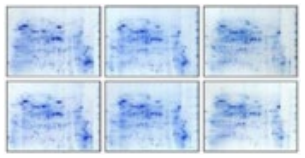


## Controlled Environment
**40/3 Lines**
**2 env. scenarios**

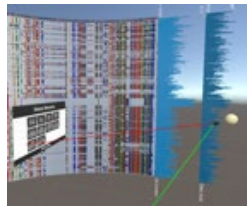
**Time-resolved Seed Architecture**
**MRI**
**Histology**



**Time- and Space-resolved Multi-Omics**



**DNA methylation**
**mRNA / sRNA**
**proteins**
**metabolites**
**phytohormones**

## VR visualization


**improved human access to Big Data and support of efficient exploration**
**and**
**Science Outreach**

**Plant Breeding 4.0**

**Seed Systems Biology**

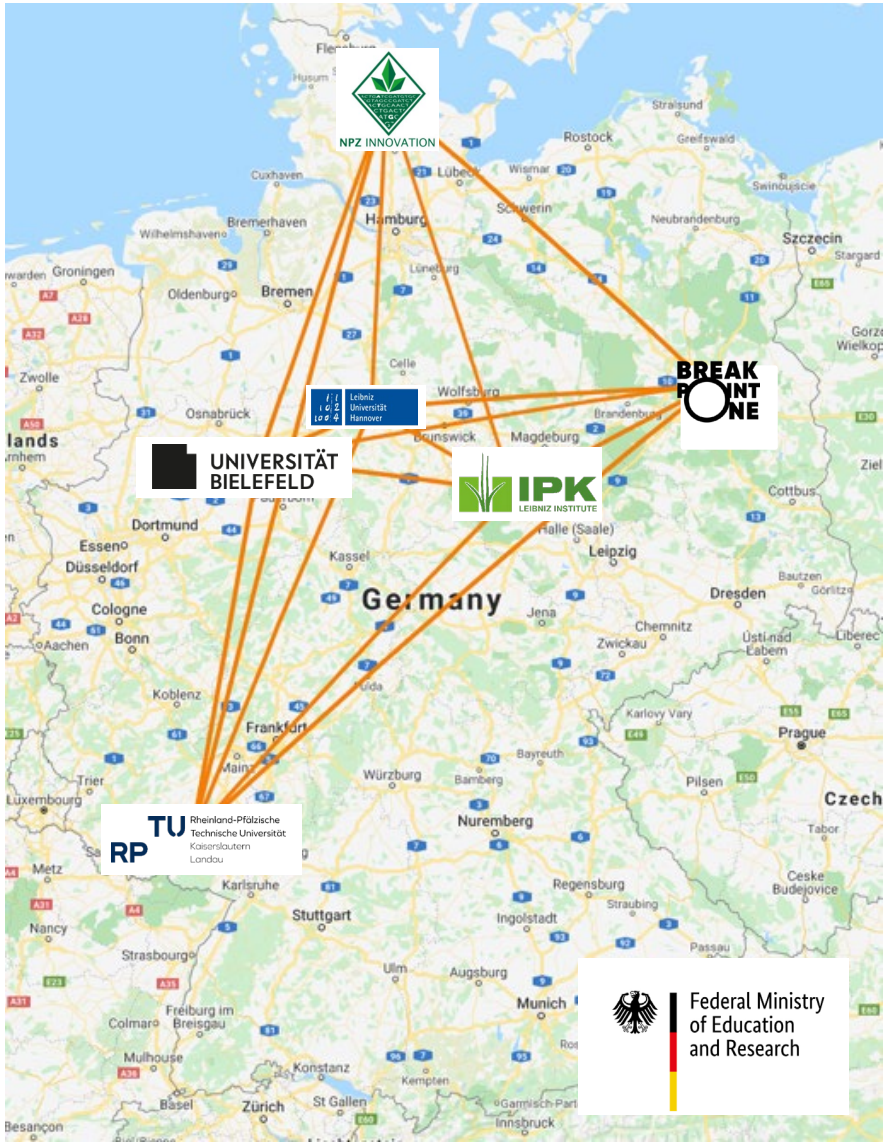
# AVATARS – the Partner Network

## IPK Gatersleben

- **Thomas Altmann** (project leader, IPK-HET), *growth & phenotyping in controlled environment, epigenetics*
- **Ljudmilla Borisjuk** (IPK-AAN), *MRI & histology, metabolome, respiration germination assay*
- **Jozefus Schippers** (IPK-SE), *RNA-seq & tissue sampling*
- **Evgeny Gladilin** (IPK-BA), *growing seed model & image analysis*
- **Mary-Ann Blätke** (IPK-IMM), *metabolic networks*
- **Matthias Lange** (IPK-BIT), *informatics infrastructure*

## Leibniz University Hannover

- **Hans-Peter Braun** (UHann), *shot-gun proteomics*




## NPZi

- **Amine Abbadi** (NPZi), *breeding nursery, genotyping, X-ray CT, NIRS & Hyperspec, germination*

## Breakpoint One

- **Olaf Sacher & Sebastian Tusk** (BPO), *virtual & augmented reality*

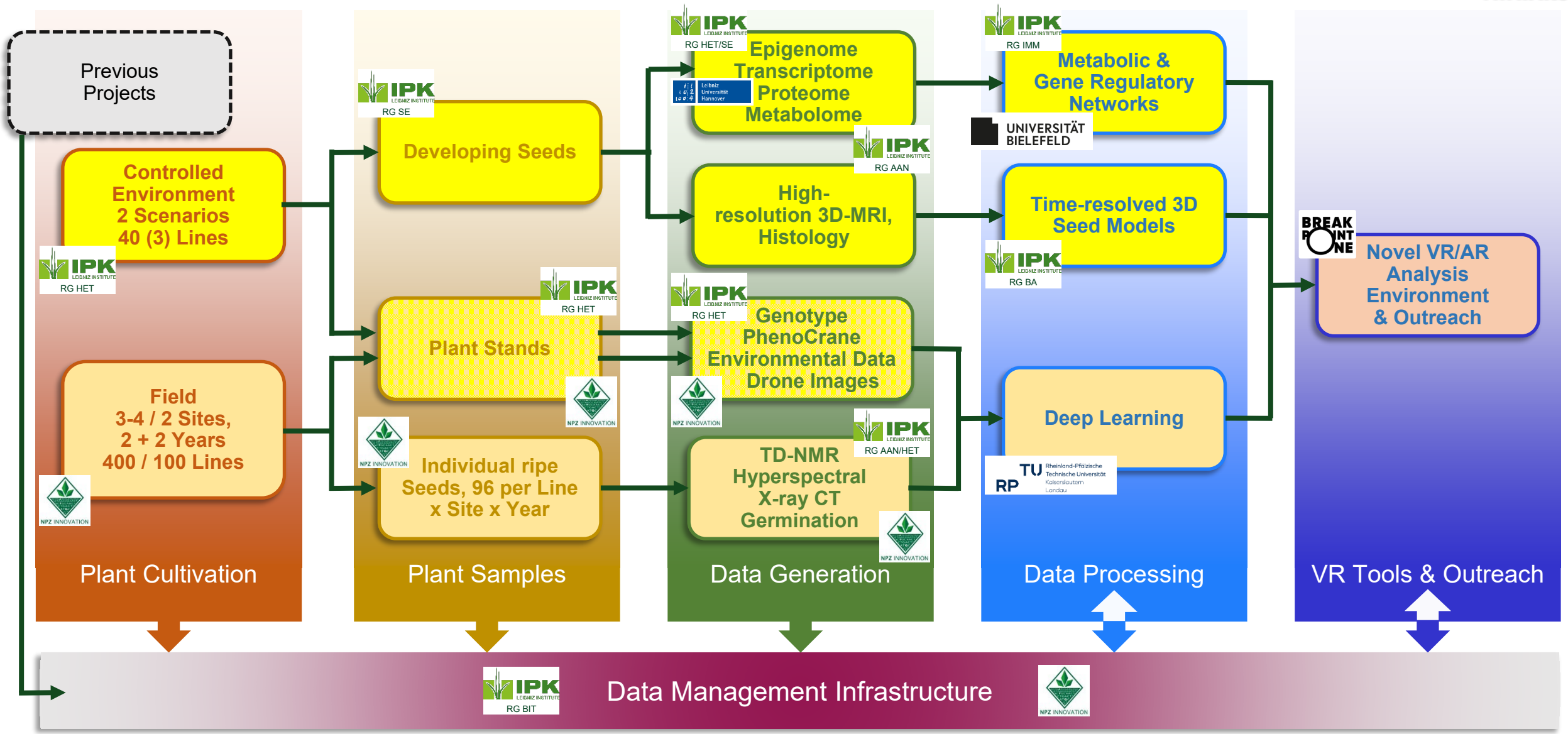
## University of Bielefeld

- **Andrea Bräutigam** (UBi), *regulatory networks*

## RPTU Kaiserslautern-Landau

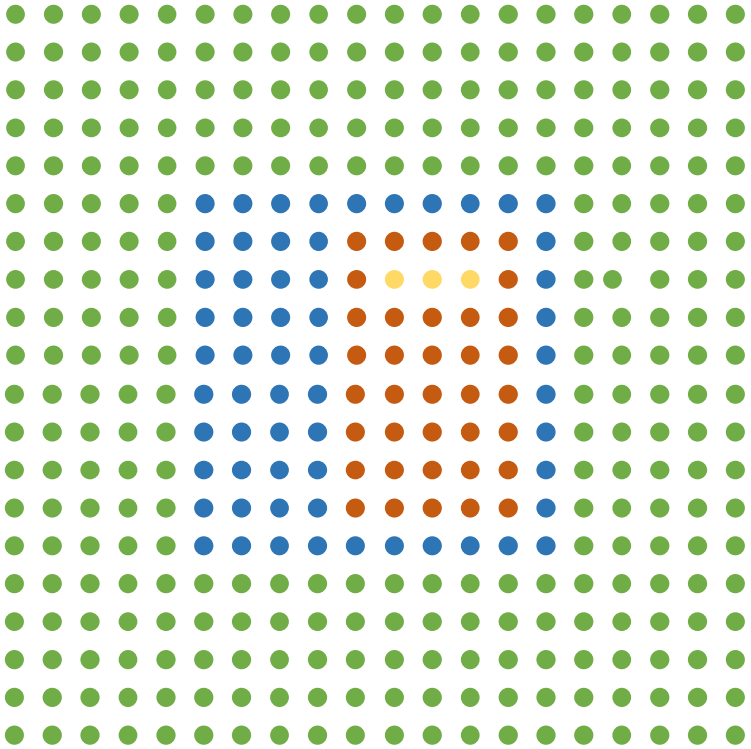
- **Marius Kloft** (RPTU), *deep learning*

# AVATARS – the Work Flows



# Population / Cultivation Scheme

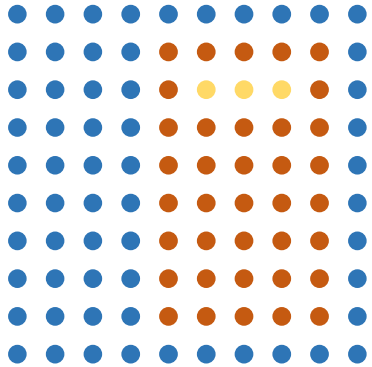
**Field Cultivation:**
**2019-20 & 2020-21**



**400 Lines**

*x 3-4 Locations*

**Field Cultivation:**
**2021-22 & 2022-23**



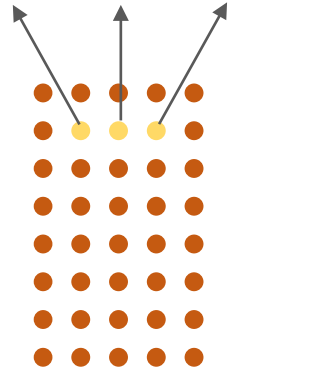
**100 Lines**

*x 2 Locations*

**Controlled Env. Cultivation**
**2020-21, 2021-22 & 2022-23**

IPK LEIBNIZ INSTITUTE RG HET

**AV1** **Express** **AV2**

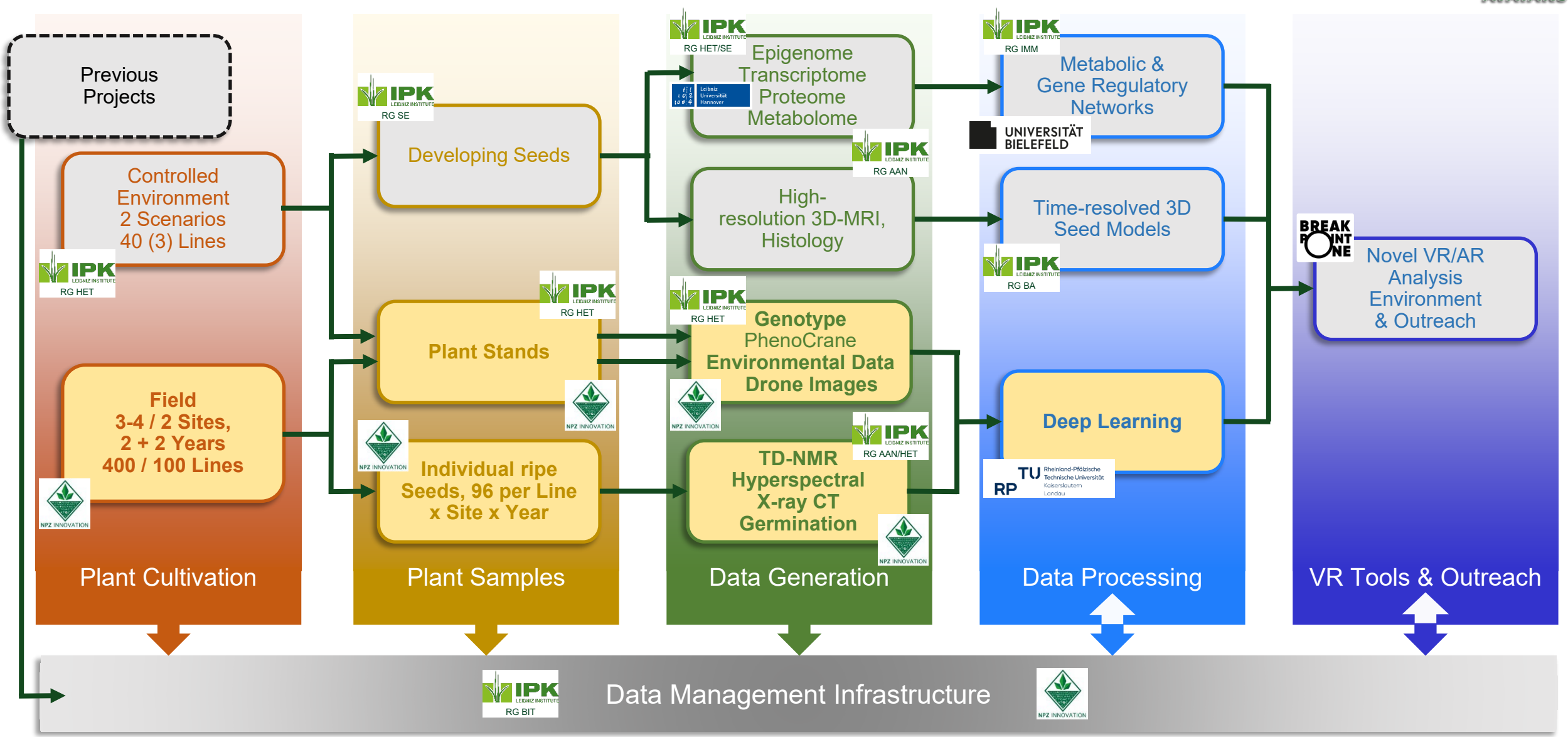


**40 Lines**

*x 2 Conditions (IPK-PS)*

- *benign*
- *detrimental*

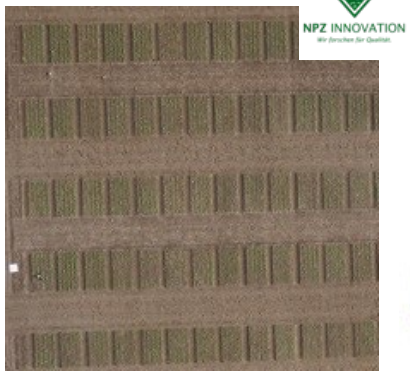
# AVATARS – the Work Flows



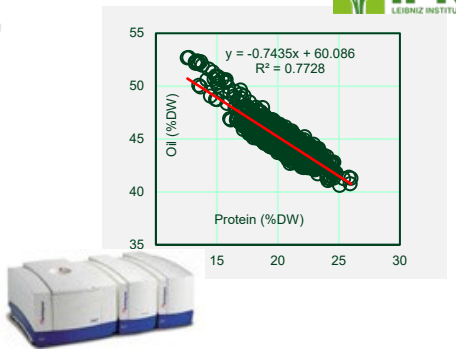
# Field Trials and uHTP Seed Data

## Deep Single-Seed Phenotyping

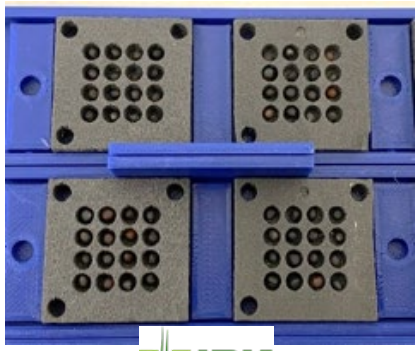
Field Trials



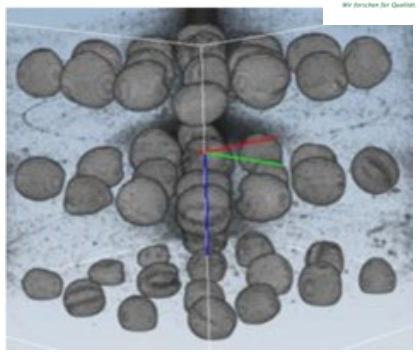
TD-NMR

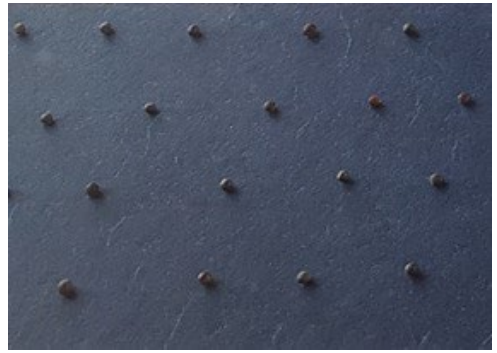



Single Object 2D & 3D tracing*
for all analysis techniques



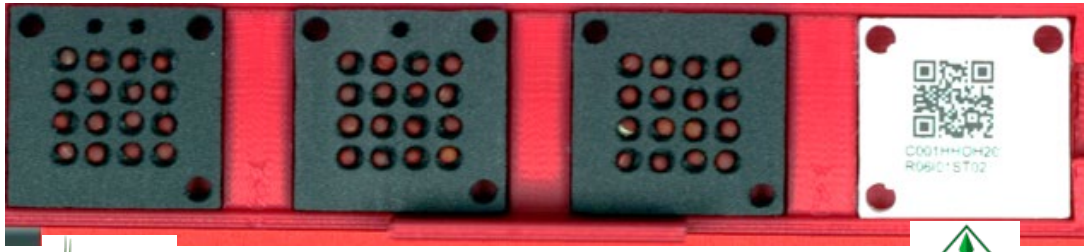
X-ray CT



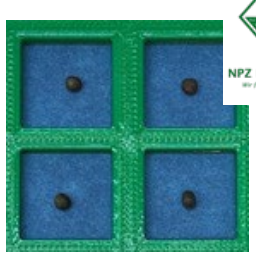


Germination Assay

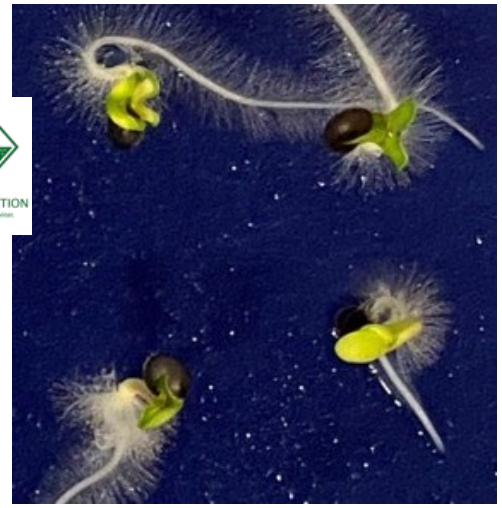
- UAV Images
- Phenotypic Scores
- Weather & Soil Data
- Genotypic Data
- NIRS
- TKM



Hyperspectral Imaging VNIR + SWIR



Accelerated Aging



In total, 364.896 seeds intended; > **260.000 done!**

*patent filed

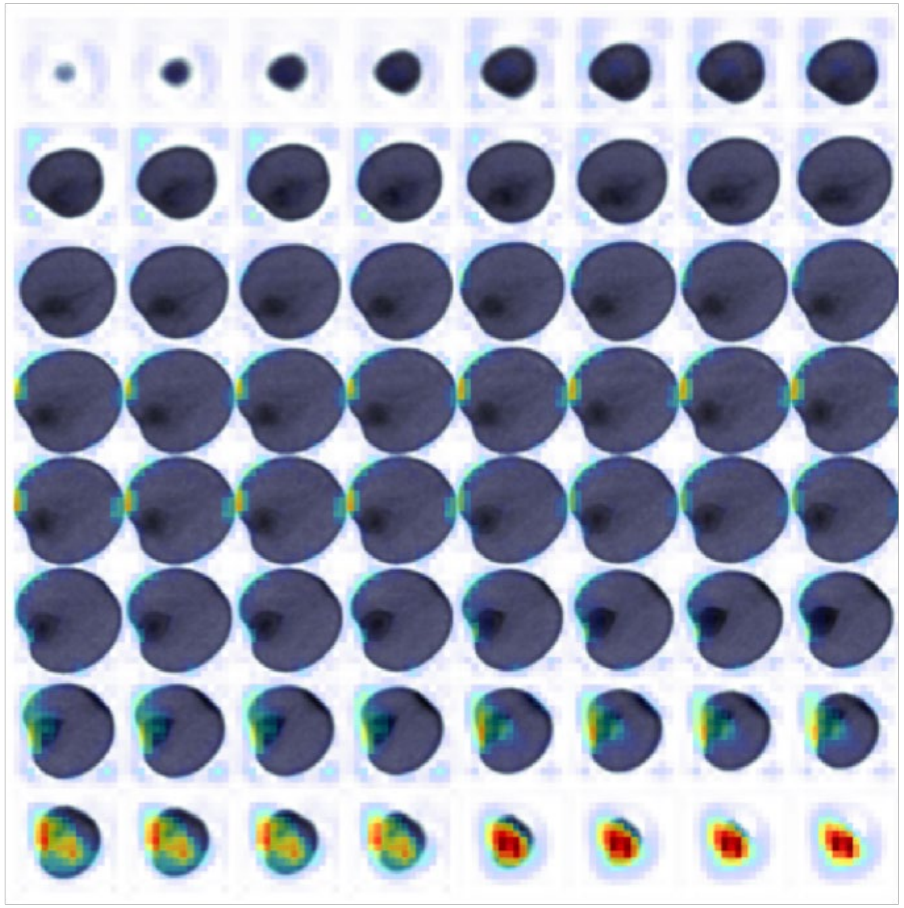
# Machine Learning using deep neural networks

AVATARS

TU RP Rheinland-Pfälzische Technische Universität Kaiserslautern Landau

## Explainable Anomaly Detection (AD)

**Heatmap for a CT scan of a non-germinating seed**



## Binary Classification Performance

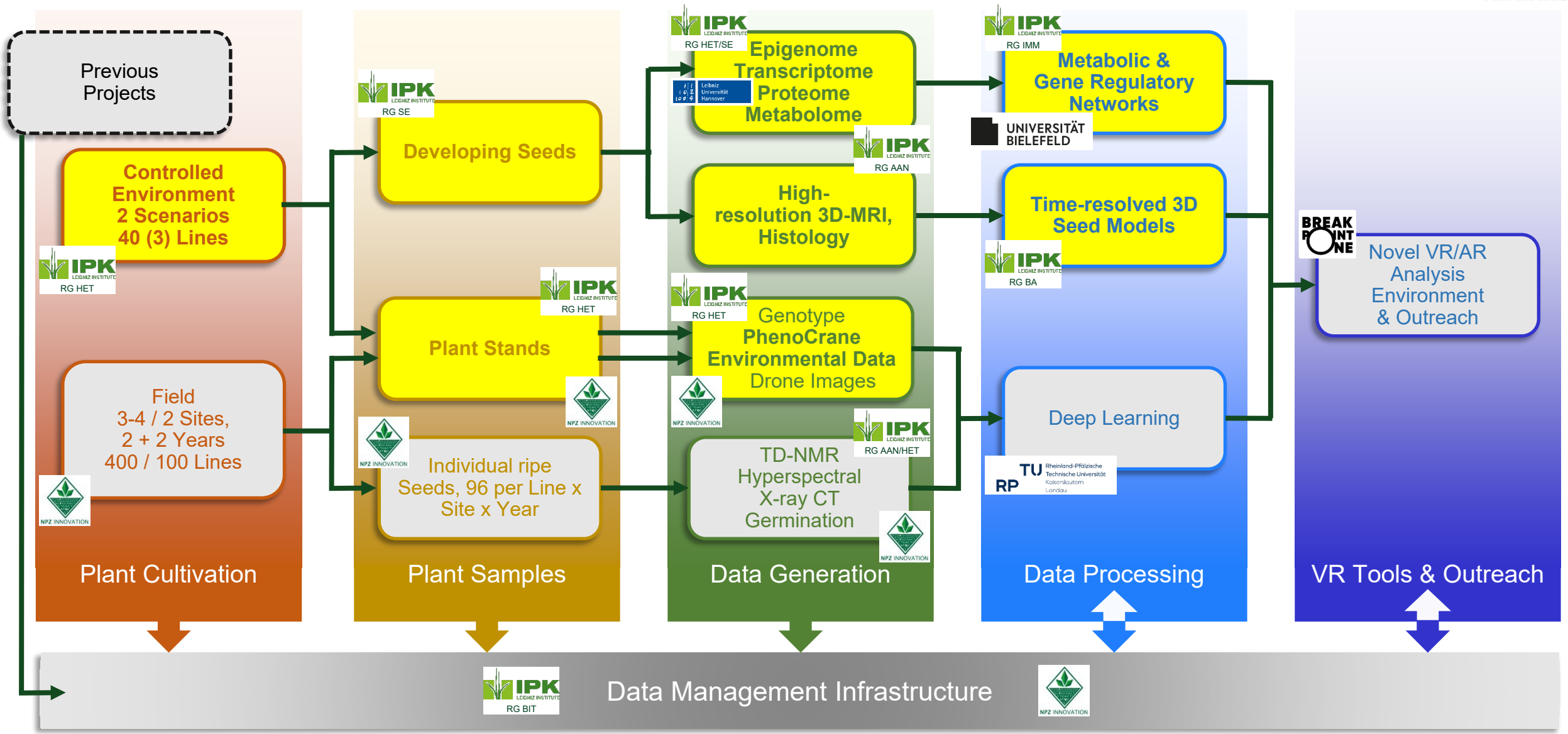
Table 1: Accuracy on hold-out validation data with models.

| Method | Modalities | Training Size | Accuracy |
|---|---|---|---|
| AD | CT | ≈95k | 58.2 |
| 3D CNN | CT | ≈95k | 58.5 |
| Tabular Nets | TD-NMR | ≈95k | 63.1 |
| Intra-data MM | CT | ≈95k | 63.6 |
| CoatNet | CT | ≈95k | 66.5 |
| Inter-data MM | TD-NMR + CT | ≈95k | 70.1 |
| CoatNet | CT | ≈200k | 71.6 |

Liznerski et al., *Proc. ICLR* (2021)

Liznerski et al., *Transact. M L Res.* (2022)

# AVATARS – the Work Flows



# CE cultivation, phenotyping and sampling for imaging and omics analyses

## PhenoSphere

➢ **Field-like controlled environments**

- Illumination (up to 1500 µmol/m²*s PAR )
- Air/Soil temperature
- Air humidity
- $CO_2$ level
- Wind simulation
- automated watering

**Regulation with fast changes of environmental conditions as they are typical for field conditions!**

**2 scenarios:**
**- benign (✓)**
**- detrimental (✓)**

➢ `PhenoCrane' for multi-imaging phenotyping

- RGB imaging + 3D laser scanner
- FluorCam (kin. CHL-fluorescence)
- hyperspectral imaging (VNIR: 350 – 900 nm)

➢ **Seed sampling**

# CE cultivation, phenotyping and sampling for imaging and omics analyses

## PhenoSphere

➢ **Field-like controlled environments**

- Illumination (up to 1500 µmol/m²*s PAR )
- Air/Soil temperature
- Air humidity
- $CO_2$ level
- Wind simulation
- automated watering

**Regulation with fast changes of environmental conditions as they are typical for field conditions!**

**2 scenarios:**
- benign (✓)
- detrimental (✓)

For detailed information see:

**Recordings of presentations and tour of the**

***IPPN-C$_E$PPG Workshop on Environment Simulation and Plant Phenotyping***

**https://meetings.ipk-gatersleben.de/ceppg2022/program/**

**M. Heuermann et al. (2023) Nat. Commun.,**
https://doi.org/10.1038/s41467-023-41332-4

➢ `PhenoCrane' for multi-imaging phenotyping

- RGB imaging + 3D laser scanner
- FluorCam (kin. CHL-fluorescence)
- hyperspectral imaging (VNIR: 350 – 900 nm)

➢ **Seed sampling**

# CE cultivation (PhenoSphere):

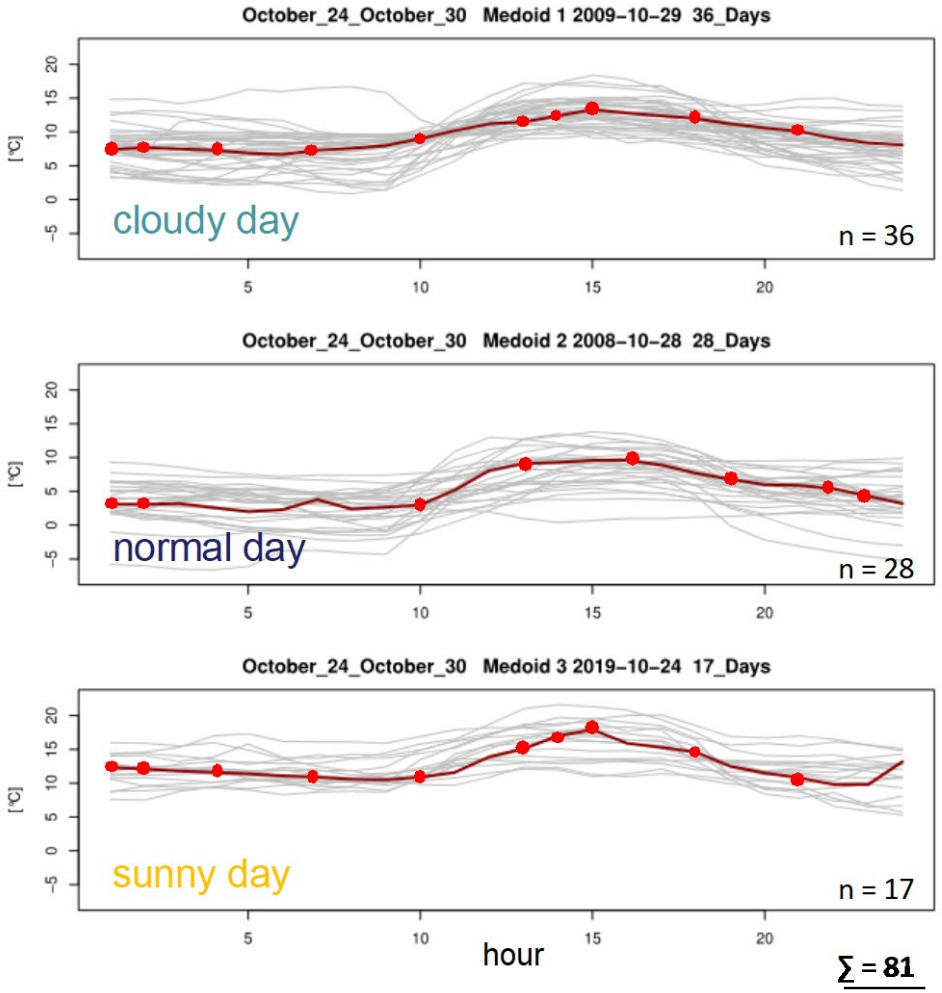
## benign weather regime

➢ **benign** = representative fluctuating conditions designed on a weekly basis

- based on **12 years of historic IPK weather data** (2008-2019)
- light changes in minutes / temperature & humidity in hour interv
  - **three light scenarios** programmed with fast changes of illumination in the range of minutes to simulate clouds

    "**cloudy**" (8 h of fluctuations, only LED)
    "**normal**" (5 h of fluctuations)
    "**sunny**" (2 h of fluctuations)

  - **daily temperature and humidity** ramped to approximate the diurnal curves
  - **weekly adjustment of sunrise and sunset** to account for the effect of day length
  - **monthly adjustment of light intensities** to simulate seasons



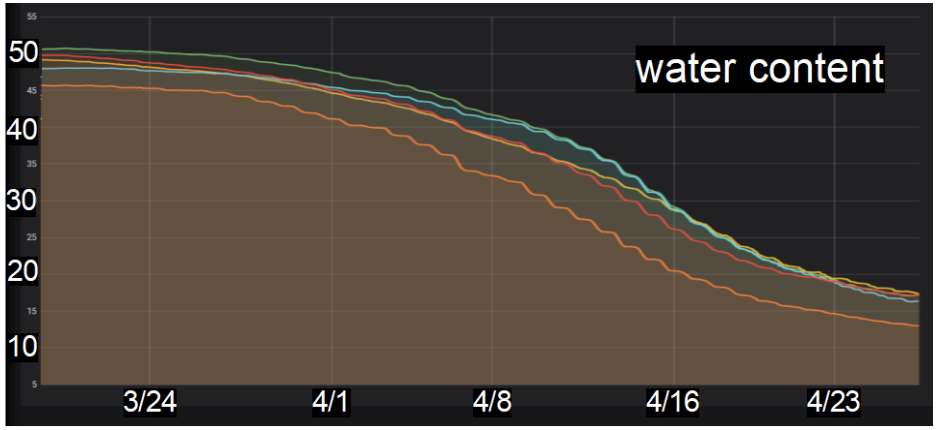
# CE cultivation (PhenoSphere):

## detrimental weather regime

- **detrimental** conditions are the same climate regime as for beneficial conditions, but with **two stress phases:**

**I. Spring/early summer drought during flowering** (BBCH 60-65) after 2nd and 3rd week of drought

- reduction in watering by mid of March
- three weeks without / reduced watering
- cloudy days → sunny days (start 14. April)
- monitoring of the effect on the plants




26. April 2022

5. Mai 2022

# CE cultivation (PhenoSphere):

## detrimental weather regime

- **detrimental** conditions are the same climate regime as for beneficial conditions, but with **two stress phases:**

**II. Simulation of heavy rain events during ripening (BBCH 86)**

- stress during ripening: BBCH 85-86 (seeds become dark)
- 2x three days of heavy watering in two weeks

  (heavy rain: 2x 10L at three consecutive days)

- replacement of sunny days → cloudy days
- increase relative air humidity to 80-90%
- additional spraying of pods with water 3-4x each day

# CE cultivation (PhenoSphere):

## detrimental weather regime

- **detrimental** conditions are the same climate regime as for beneficial conditions, but with **two stress phases:**

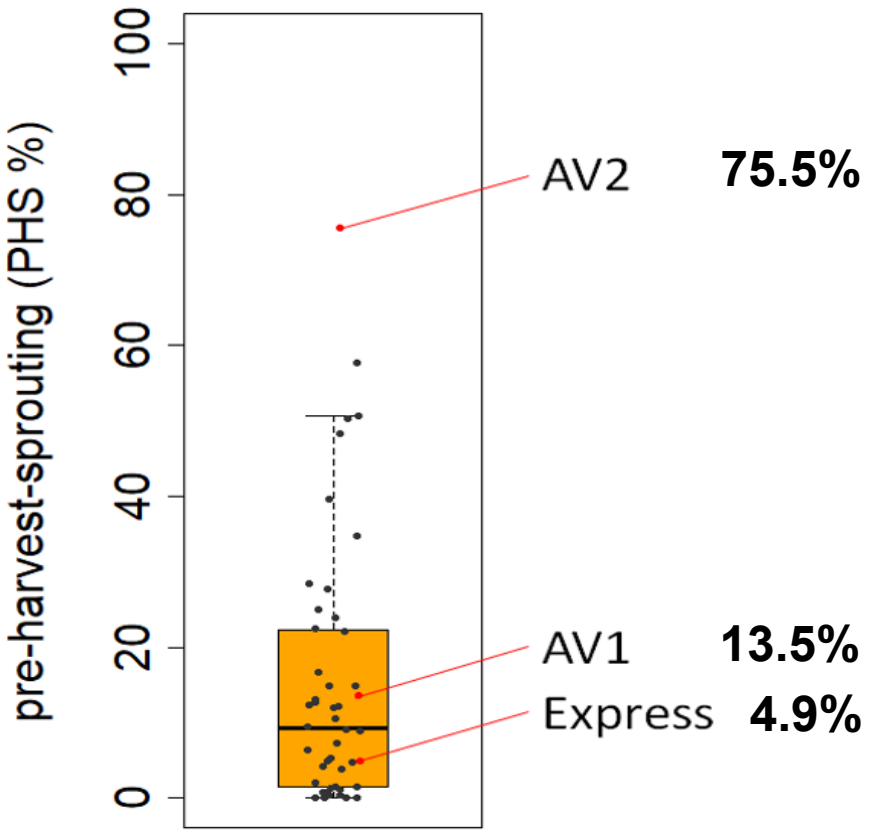
**II. Simulation of heavy rain events during ripening (BBCH 86)**

- stress during ripening: BBCH 85-86 (seeds become
- 2x three days of heavy watering in two weeks

  (heavy rain: 2x 10L at three consecutive days)

- replacement of sunny days → cloudy days
- increase relative air humidity to 80-90%
- additional spraying of pods with water 3-4x each day

after 2nd week of heavy rain simulation



# CE cultivation, phenotyping and sampling for imaging and omics analyses

## PhenoSphere

- **<u>Field-like controlled environments</u>**
  - Illumination (up to 1500 µmol/m²*s PAR )
  - Air/Soil temperature
  - Air humidity
  - $CO_2$ level
  - Wind simulation
  - automated watering

**Regulation with <u>fast changes of environmental conditions</u> as they are typical for field conditions!**

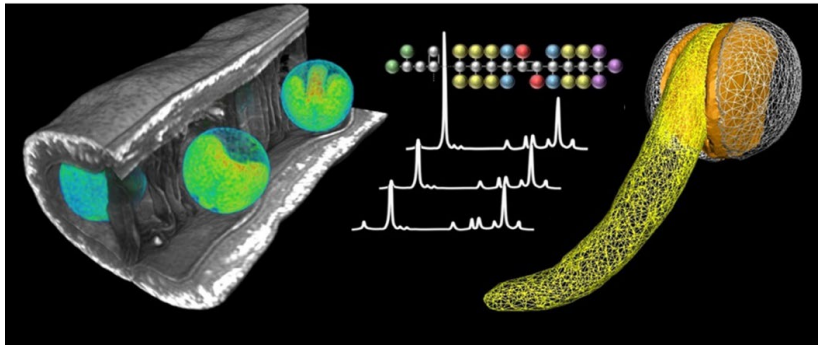
**<u>2 scenarios:</u>**
**<u>- benign (✓)</u>**
**<u>- detrimental (✓)</u>**

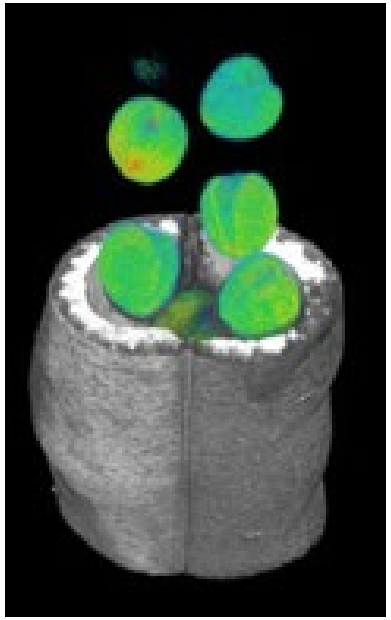
- <u>`PhenoCrane' for multi-imaging phenotyping</u>
  - RGB imaging + 3D laser scanner
  - FluorCam (kin. CHL-fluorescence)
  - hyperspectral imaging (VNIR: 350 – 900 nm)

- **<u>Seed sampling</u>**

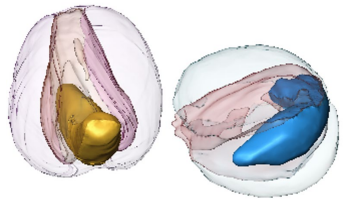
# Seed Imaging, …


**MRI** ( Ascent III/WB, NMR 9,4T)
High resolution imaging of living seeds

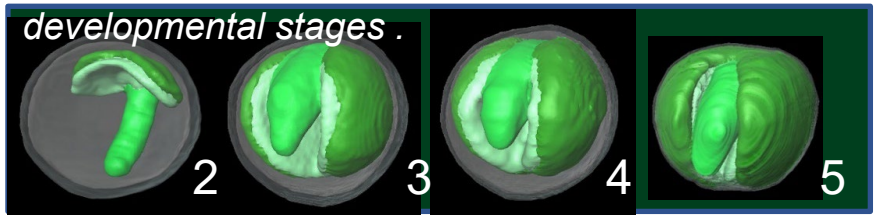


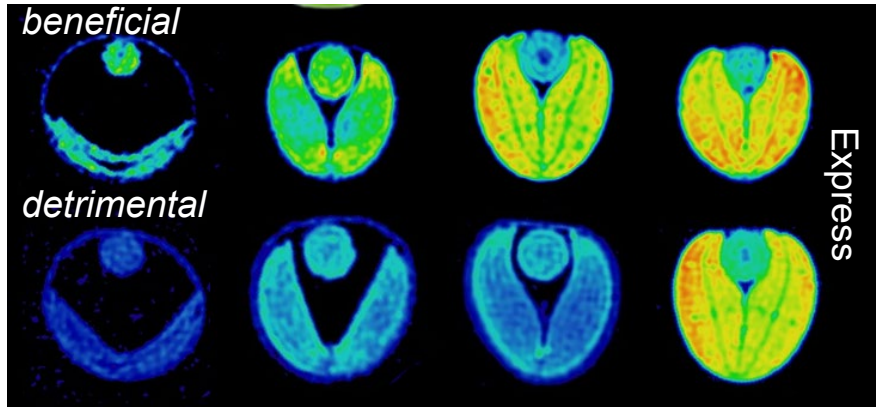


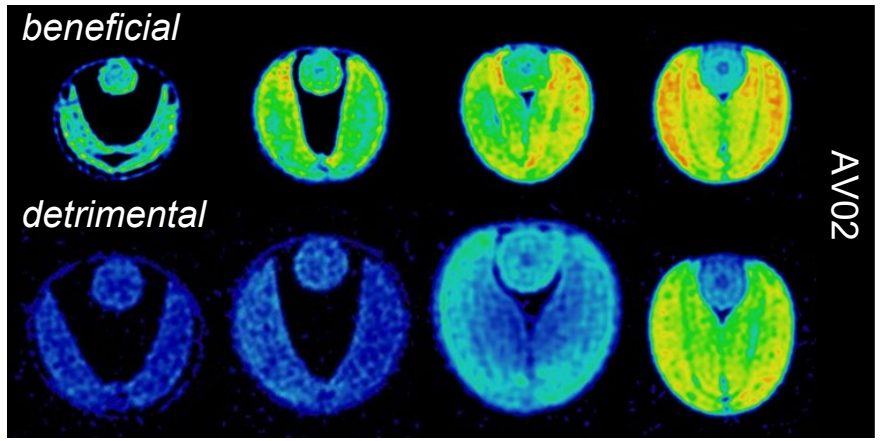
3D MRI data



3D labels

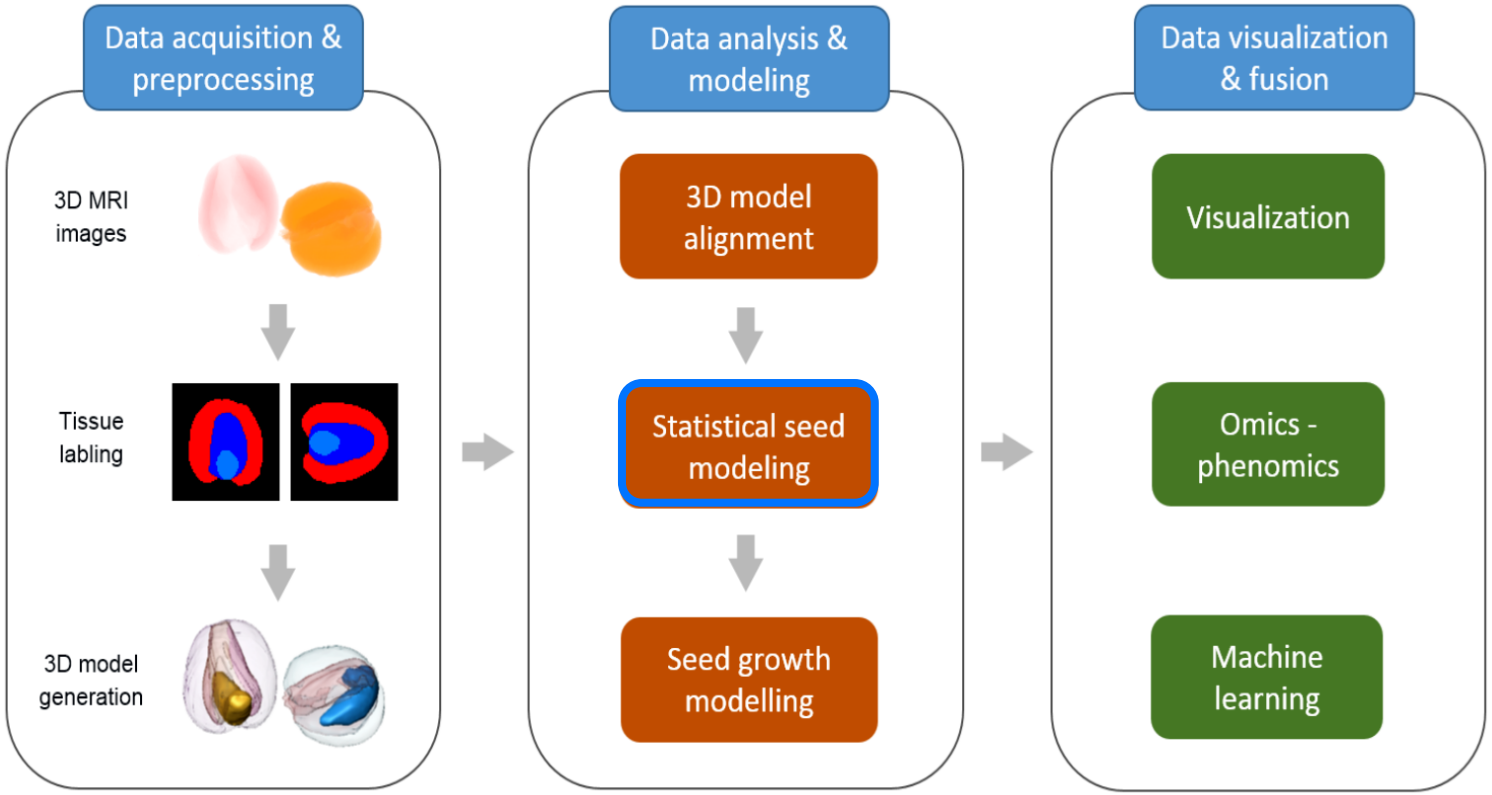


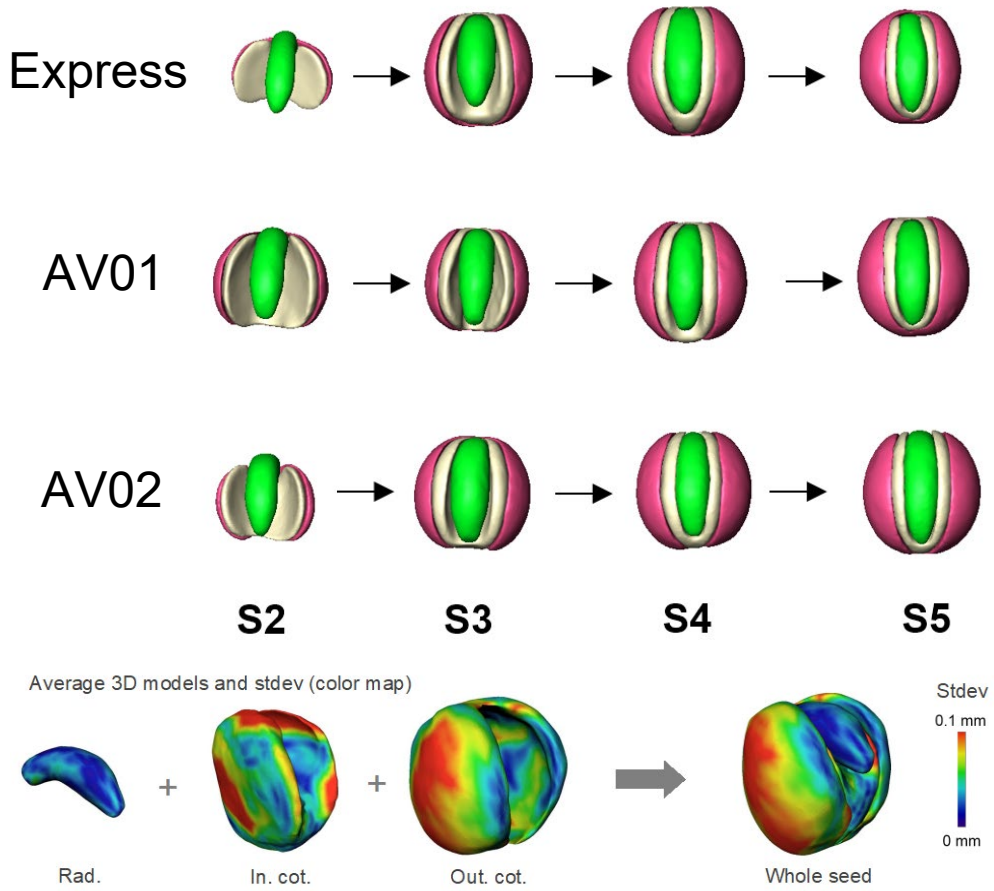



# …, Image Analysis, and Modeling

Image processing and analysis pipeline

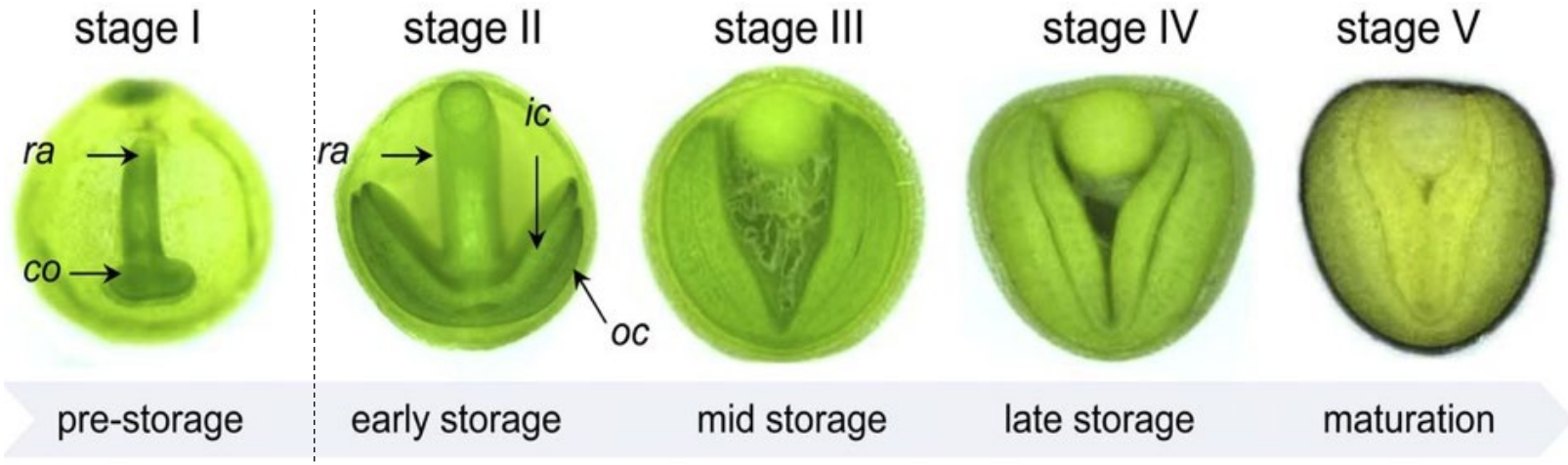



Average 3D models of *B. napus* seeds (3 lines, 4 stages; n = 5)



# AVATARS – Seed sampling and multi-omics analyses

**5 Developmental stages**

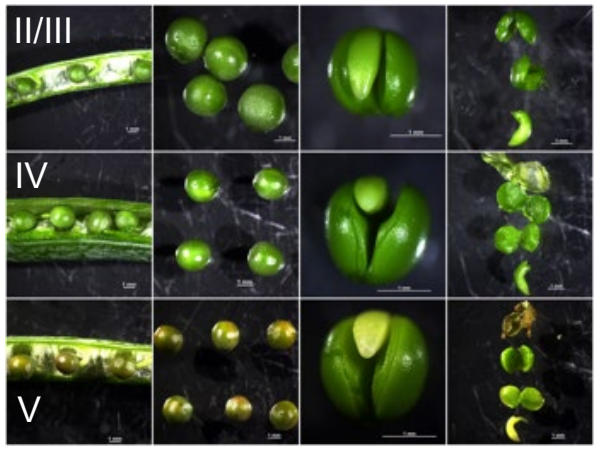



**3 Genotypes**

| | Samples |
|---|---|
| **Express** | 51 |
| 5 stages | |
| **AV01** | 27 |
| **AV02** | 27 |
| 3 stages (1 / 3 / 5) | |
| **Σ 105 samples** | |
| **Σ 18 samples** | |

**Whole seeds** **4 Seed organs:** inner, outer cotyledon (ic and oc), radicle (ra) and seed coat (sc)

| | I | II | III | IV | V |
|---|---|---|---|---|---|
| sRNA-Seq. | 3 G. | 1 G. x 4 O. | 3 G. x 4 O. | 1 G. x 4 O. | 3 G. x 4 O. |
| mRNA-Seq. | 3 G. | 1 G. x 4 O. | 3 G. x 4 O. | 1 G. x 4 O. | 3 G. x 4 O. |
| Proteome | 3 G. | 1 G. x 4 O. | 3 G. x 4 O. | 1 G. x 4 O. | 3 G. x 4 O. |
| Metabolome | 3 G. | 1 G. x 4 O. | 3 G. x 4 O. | 1 G. x 4 O. | 3 G. x 4 O. |
| Phytohormones | 3 G. | 1 G. x 4 O. | 3 G. x 4 O. | 1 G. x 4 O. | 3 G. x 4 O. |
| Methylome | 3 G. | - | - | - | 3 G. |



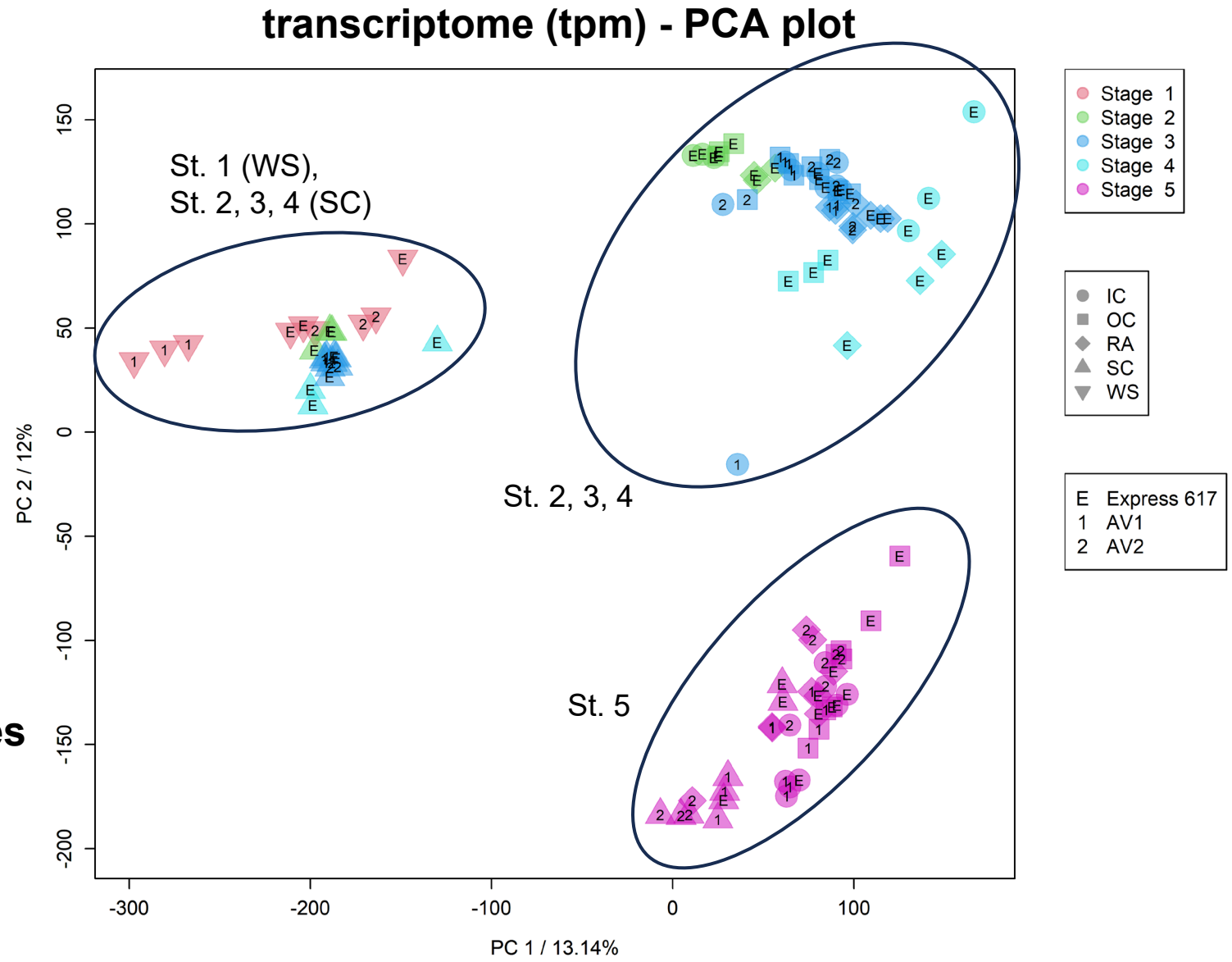

**Benign** and **Detrimental** **Conditions**: **105** + **105** = **210 samples** for **omics analyses** (+ methylome **18** + **18** = **36**)

Rolletschek et al., J. Pl. Physiol. (2021)

# AVATARS – Seed sampling and multi-omics analyses

**Benign Condition**:
**105 mRNA-Seq. Data Sets**

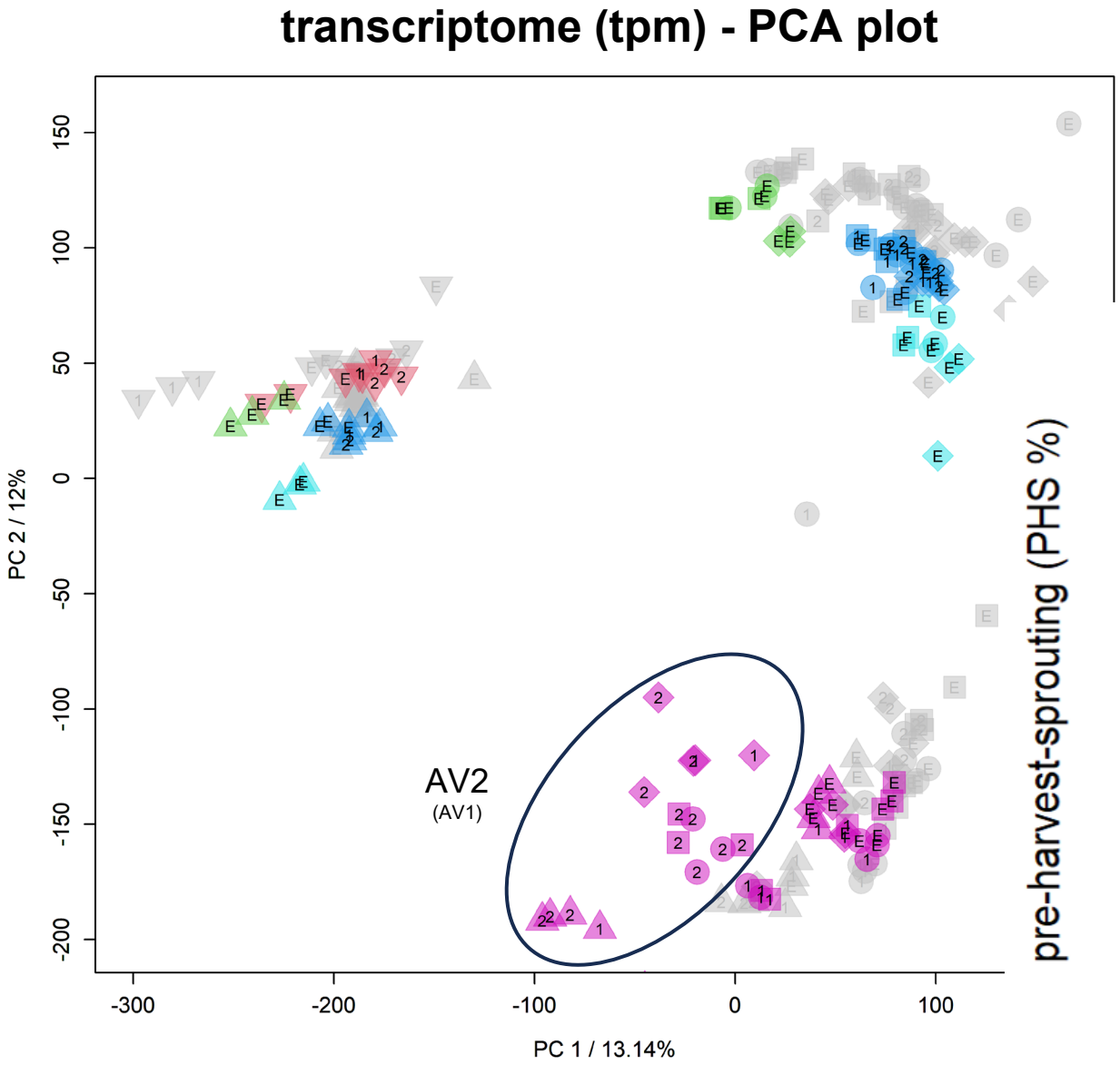
**Var.:**
**Stages > Organs > Genotypes**

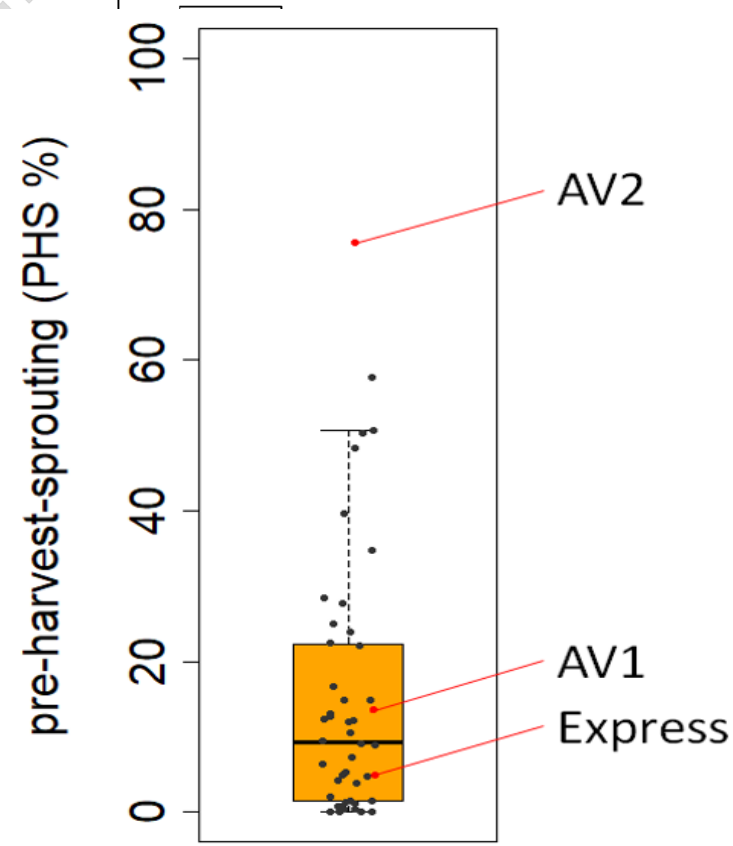


# AVATARS – Seed sampling and multi-omics analyses

Benign Condition:
105 mRNA-Seq.
Data Sets

Detrimental Condition:
105 mRNA-Seq.
Data Sets



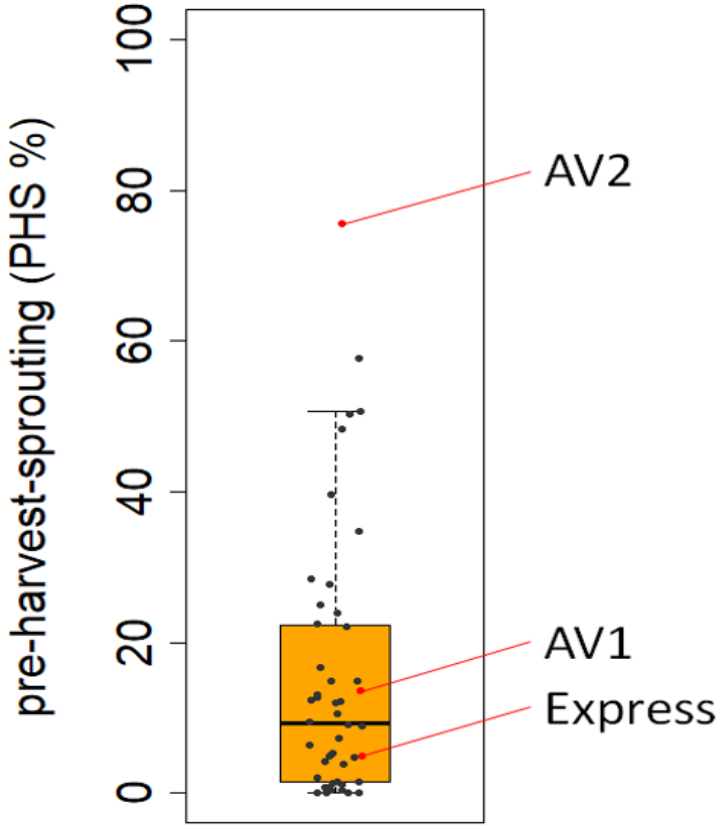
# AVATARS – Seed sampling and multi-omics analyses

**DEGs**

**Stage 5 Detrimental vs. Benign**

[log2FC] > 3 and p [FDR] < 0.01

| | | |
|---|---|---|
| **Express** | | |
| | **SC** | **204** |
| | **OC** | **289** |
| | **IC** | **131** |
| | **RA** | **189** |
| **AV1** | | |
| | **SC** | **543** |
| | **OC** | **2591** |
| | **IC** | **1274** |
| | **RA** | **1145** |
| **AV2** | | |
| | **SC** | **2198** |
| | **OC** | **4625** |
| | **IC** | **2575** |
| | **RA** | **4602** |



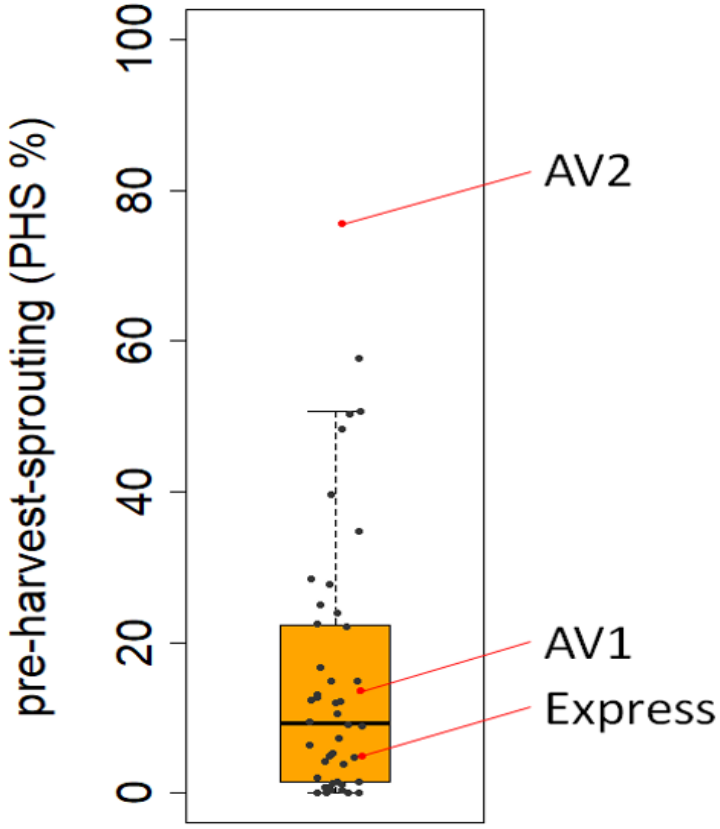
# AVATARS – Seed sampling and multi-omics analyses

**DEGs**

**Stage 5 Detrimental**

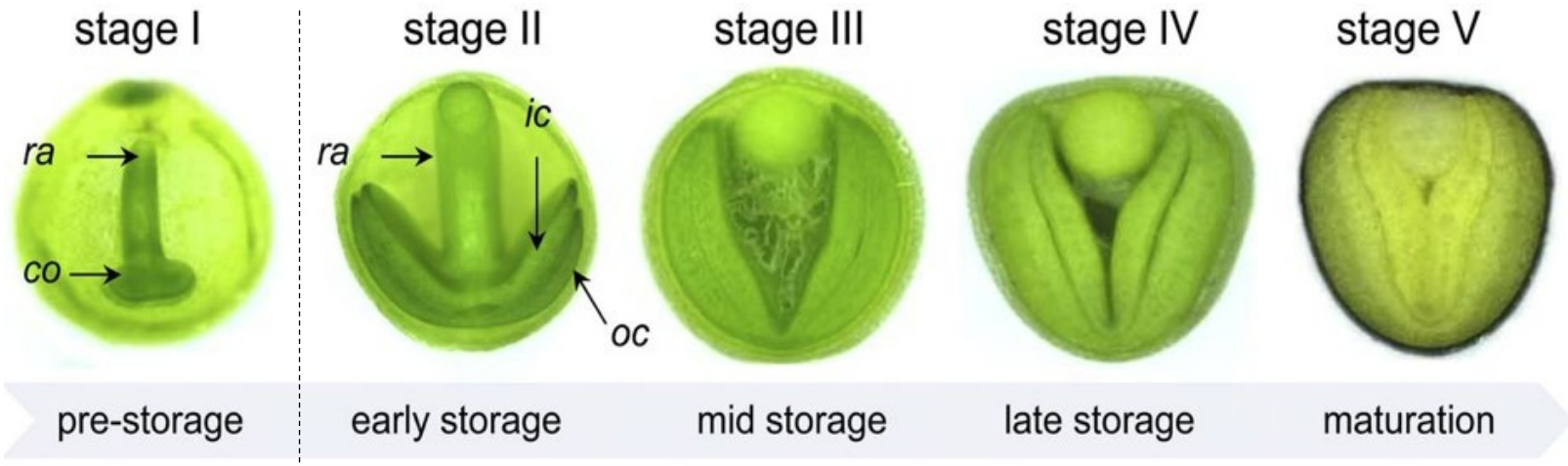
[log2FC] > 3 and p [FDR] < 0.01

| | | |
|---|---|---|
| **AV1 vs. Express** | | |
| | **SC** | **1519** |
| | **OC** | **2950** |
| | **IC** | **2049** |
| | **RA** | **1491** |
| **AV2 vs. AV1** | | |
| | **SC** | **2476** |
| | **OC** | **2591** |
| | **IC** | **2874** |
| | **RA** | **2665** |
| **AV2 vs. Express** | | |
| | **SC** | **4912** |
| | **OC** | **7176** |
| | **IC** | **5763** |
| | **RA** | **5823** |



# AVATARS – Seed sampling and multi-omics analyses

**5 Developmental stages**

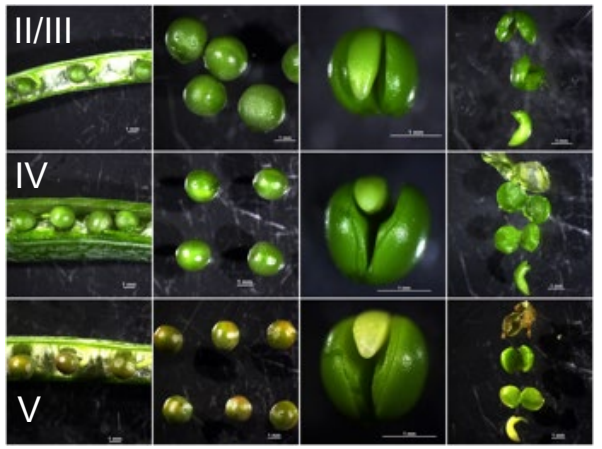



**3 Genotypes**

| | Samples |
|---|---|
| **Express** | 51 |
| 5 stages | |
| **AV01** | 27 |
| **AV02** | 27 |
| 3 stages (1 / 3 / 5) | |
| **Σ 105 samples** | |
| **Σ 18 samples** | |

**Whole seeds** **4 Seed organs:** inner, outer cotyledon (ic and oc), radicle (ra) and seed coat (sc)

| | I | II | III | IV | V |
|---|---|---|---|---|---|
| sRNA-Seq. | 3 G. | 1 G. x 4 O. | 3 G. x 4 O. | 1 G. x 4 O. | 3 G. x 4 O. |
| mRNA-Seq. | 3 G. | 1 G. x 4 O. | 3 G. x 4 O. | 1 G. x 4 O. | 3 G. x 4 O. |
| Proteome | 3 G. | 1 G. x 4 O. | 3 G. x 4 O. | 1 G. x 4 O. | 3 G. x 4 O. |
| Metabolome | 3 G. | 1 G. x 4 O. | 3 G. x 4 O. | 1 G. x 4 O. | 3 G. x 4 O. |
| Phytohormones | 3 G. | 1 G. x 4 O. | 3 G. x 4 O. | 1 G. x 4 O. | 3 G. x 4 O. |
| Methylome | 3 G. | - | - | - | 3 G. |



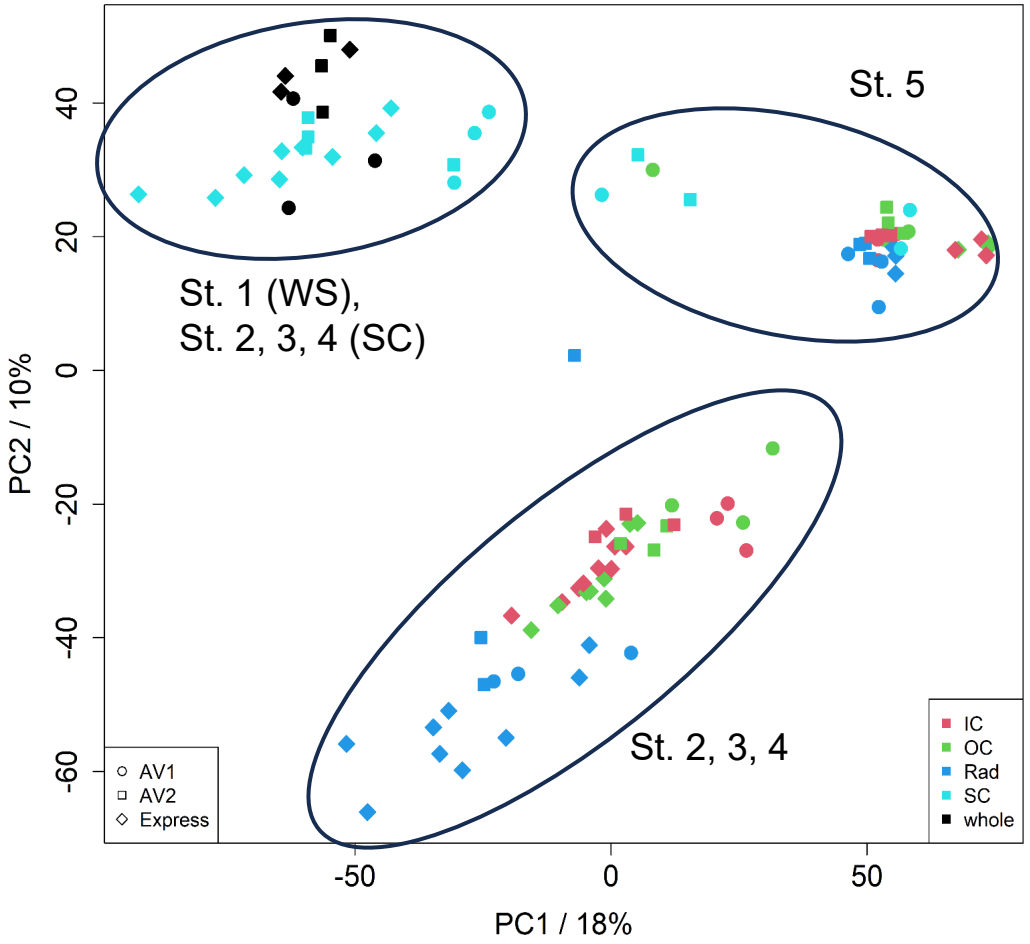

**Benign** and **Detrimental** **Conditions**: **105** + **105** = **210 samples** for **omics analyses** (+ methylome **18** + **18** = **36**)

Rolletschek et al., J. Pl. Physiol. (2021)

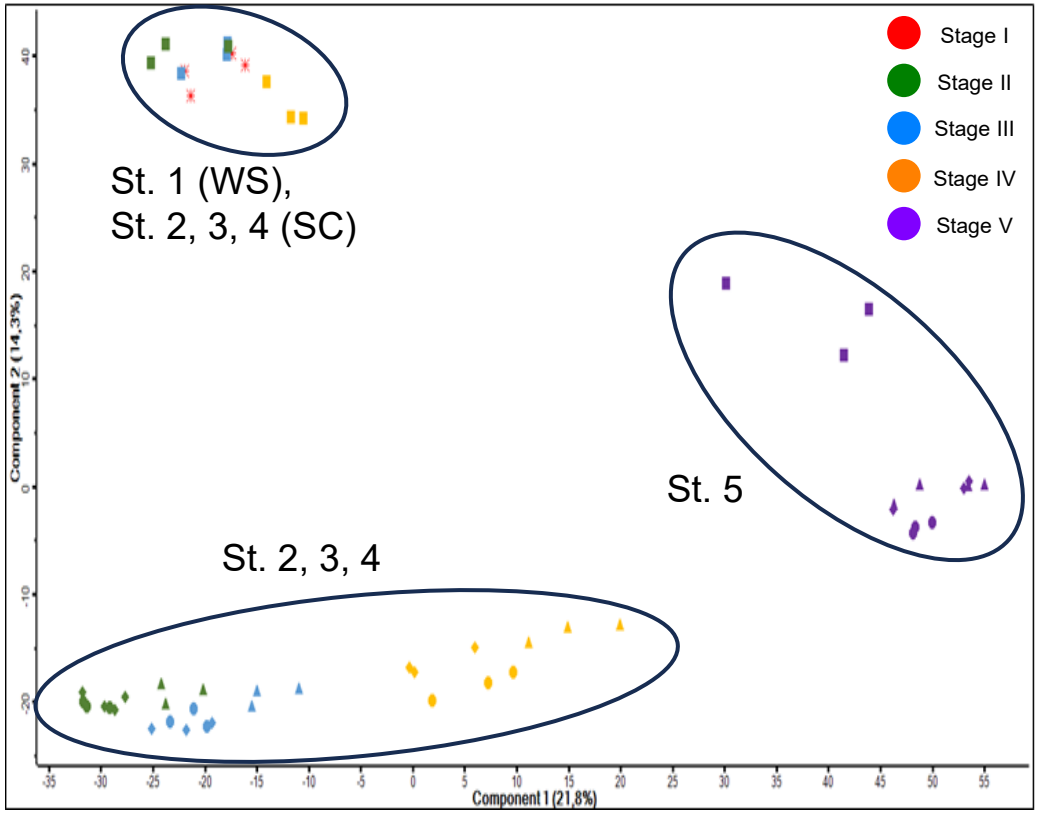
# AVATARS – Seed sampling and multi-omics analyses

Benign Condition: 105 sRNA-Seq. and proteome data sets

**small RNAs - PCA plot**



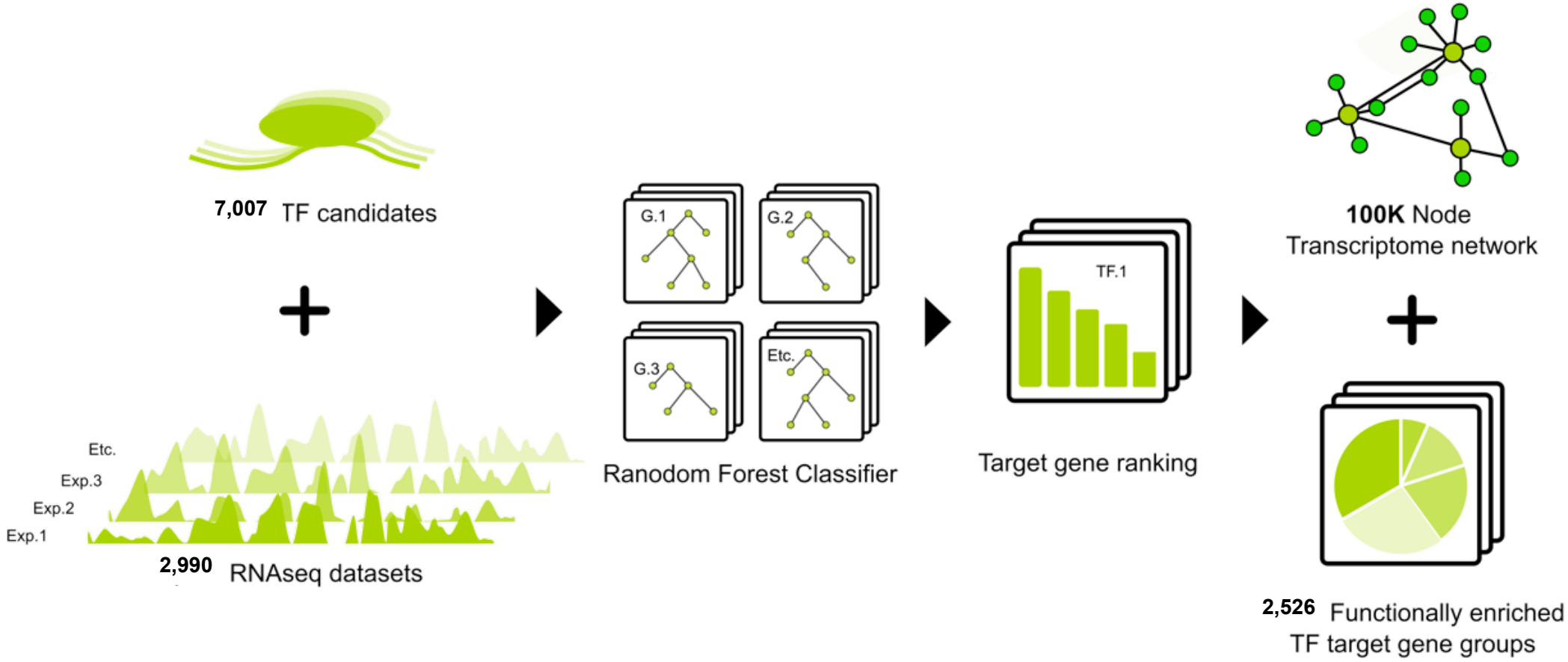

**'shotgun proteome' (Express) – 8,525 proteins**




Whole Seed
Seed Coat
Outer Cotyledon
Inner Cotyledon
Radicle

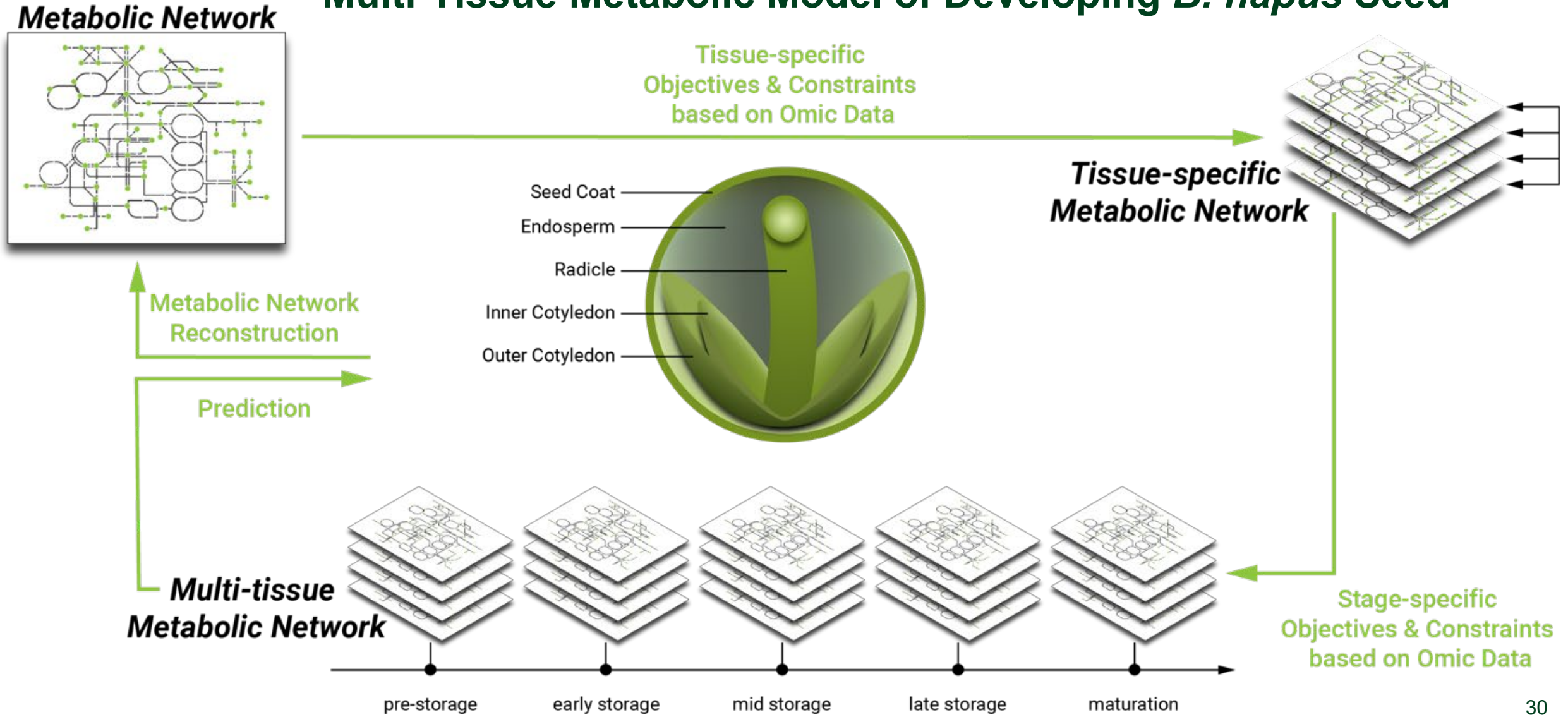
# Regulatory Network Analysis

## Machine learning based Gene Regulatory Network (GRN) construction

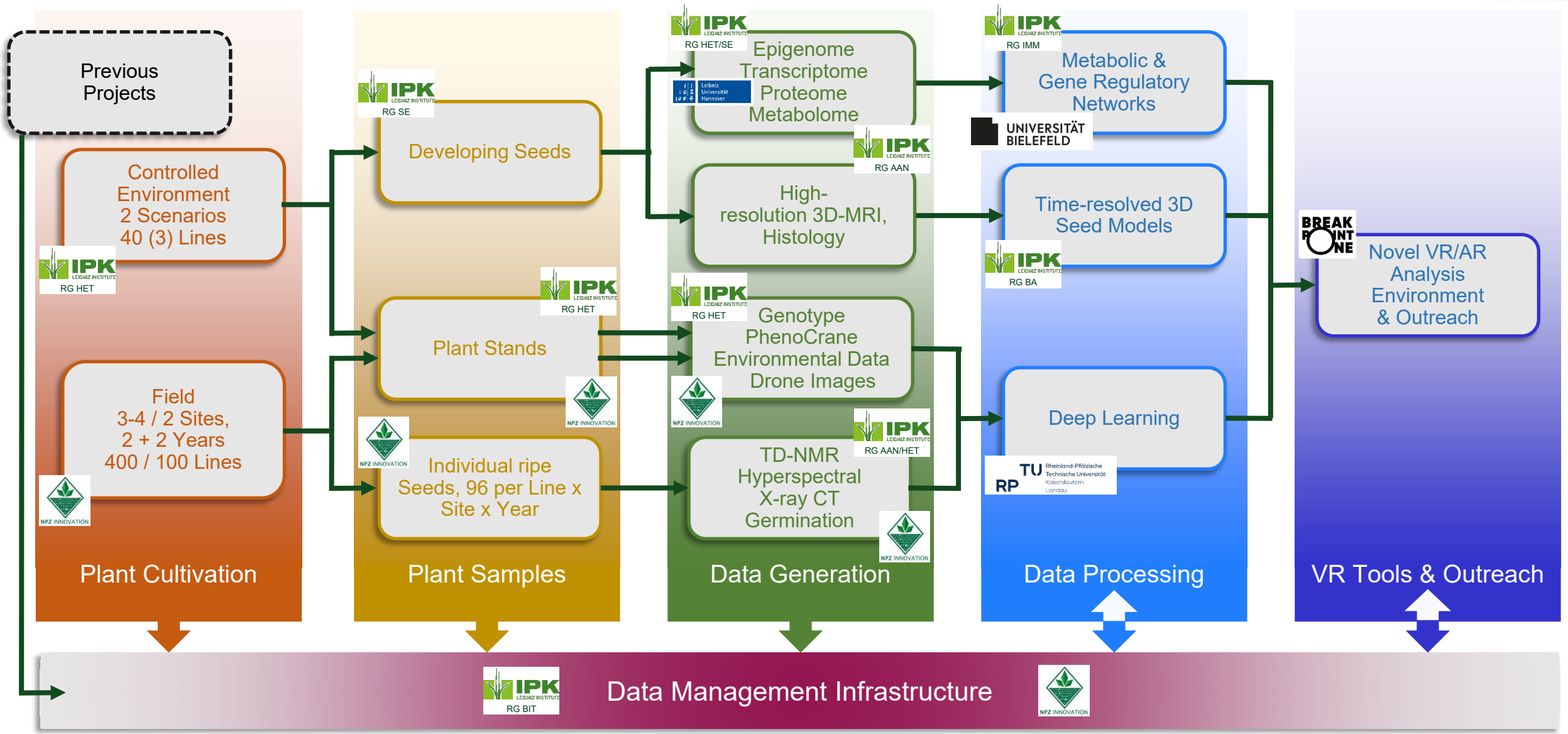


# Metabolic Modelling

## Multi-Tissue Metabolic Model of Developing *B. napus* Seed



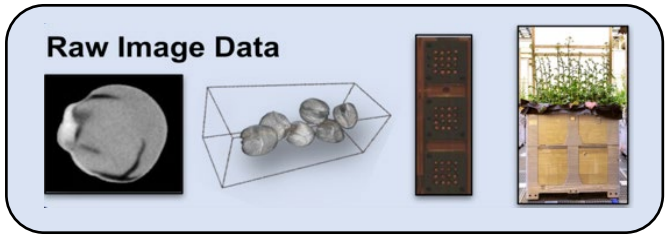
# AVATARS – the Work Flows

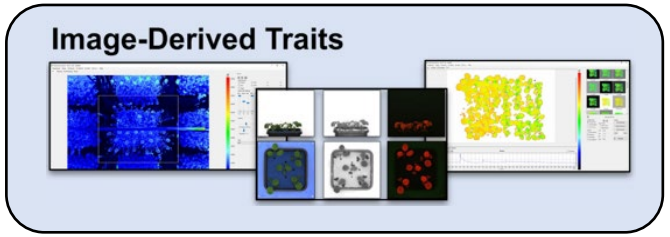


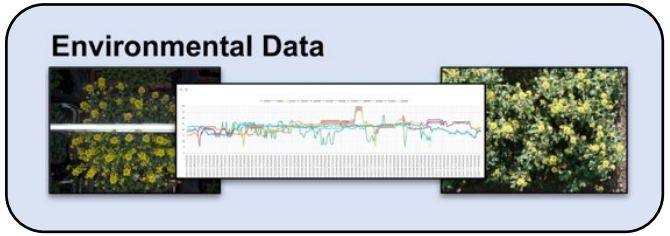
# Data Management Infrastructure

## Multiple Data Domains & Formats

Partner Data Acquisition Pipelines & Historic Data



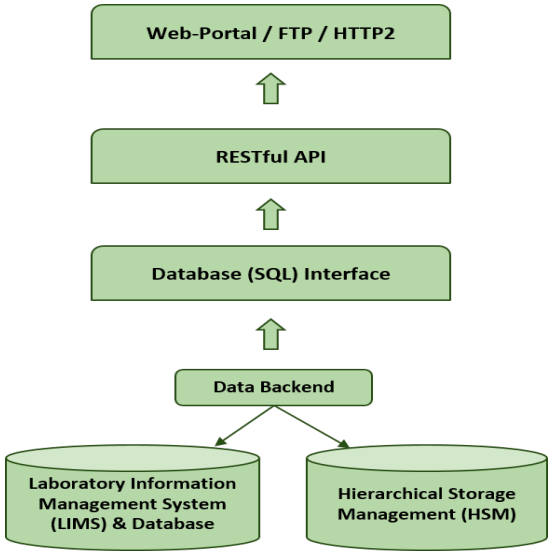




## AVATARS Integrated Data Hub & Gateway

**Digital Twins:** Connecting Digital Research Objects

| | MAT | DATA_TYPE | LOCATION_NAME | GENUS | SPECIES | SUBTAXA | NAME |
|---|---|---|---|---|---|---|---|
| 8 | _____L001HMOO20R01I01T02 | TD-NMR | MOO | Brassica | napus L. | napus | L001 |
| 9 | _____L001HMOO20R01I01T02SS001 | Single Seeds | MOO | | | | L001 |
| 10 | _____L001HMOO20R01I01T02SS016 | Single Seeds | MOO | | | | L001 |
| 11 | ____L001HMOO20R01N01 | NPZi Seedbags | NPZi | Brassica | napus L. | napus | L001 |
| 12 | __L001PHOH20R01 | Field Phenotyping | HOH | Brassica | napus L. | napus | L001 |
| 13 | ____L001HHOH20R01I01 | IPK Seedbags | IPK | Brassica | napus L. | napus | L001 |
| 14 | _____L001HHOH20R01I01T01 | TD-NMR | HOH | Brassica | napus L. | napus | L001 |
| 15 | _____L001HHOH20R01I01T01SS001 | Single Seeds | HOH | | | | L001 |
| 16 | _____L001HHOH20R01I01T01SS016 | Single Seeds | HOH | | | | L001 |

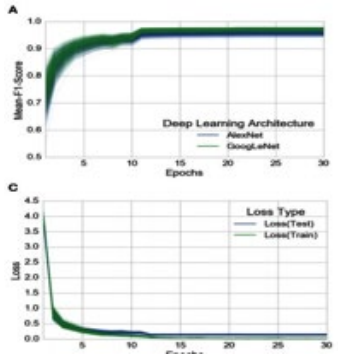
**Data Hub:** FAIR Storage & Access




## Data Visualization & Analysis

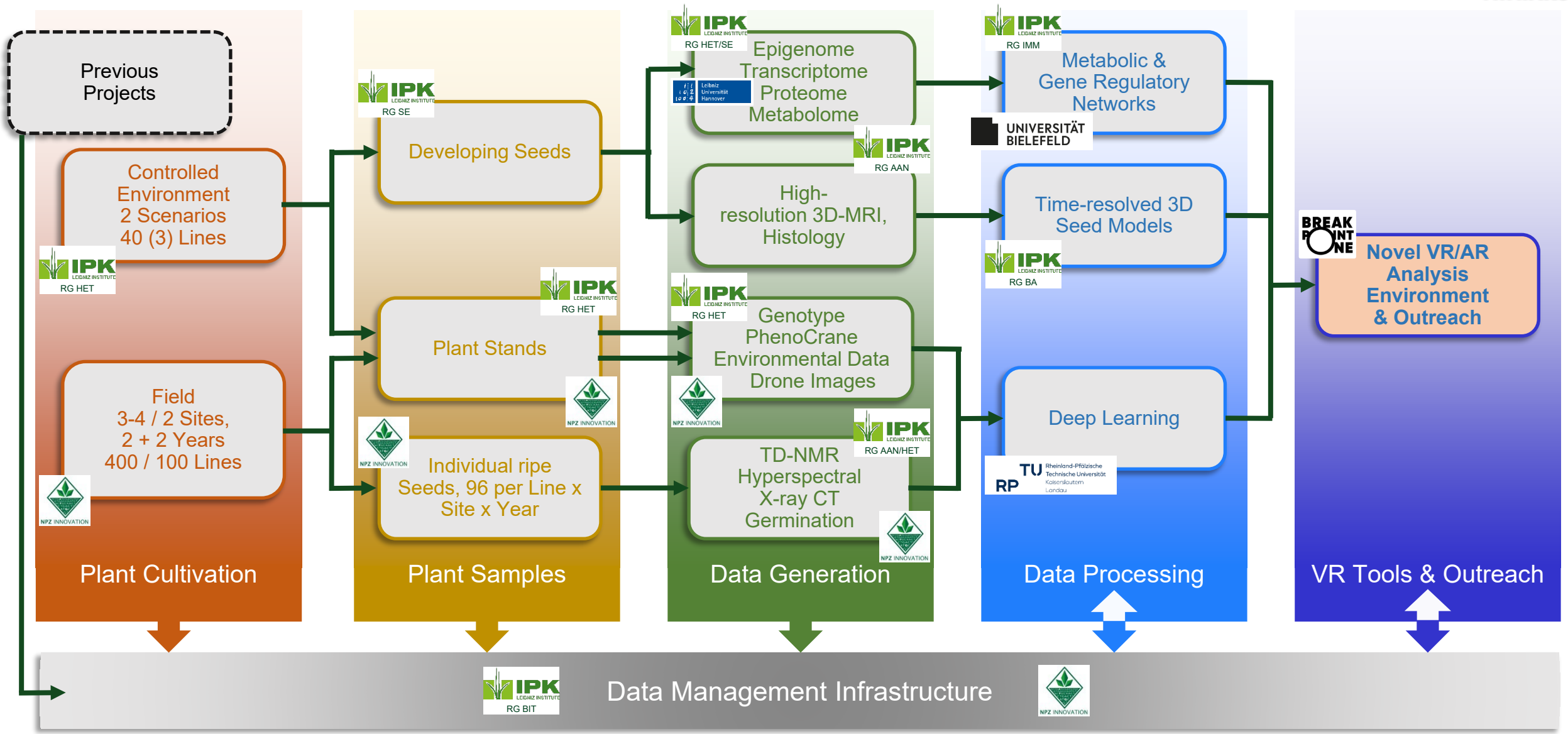
Explorative visualisation

Data analysis & model building




Arend et al., Plant J. (2022) & Arend et al., GigaScience 9 (2020) 1-11

# AVATARS – the Work Flows



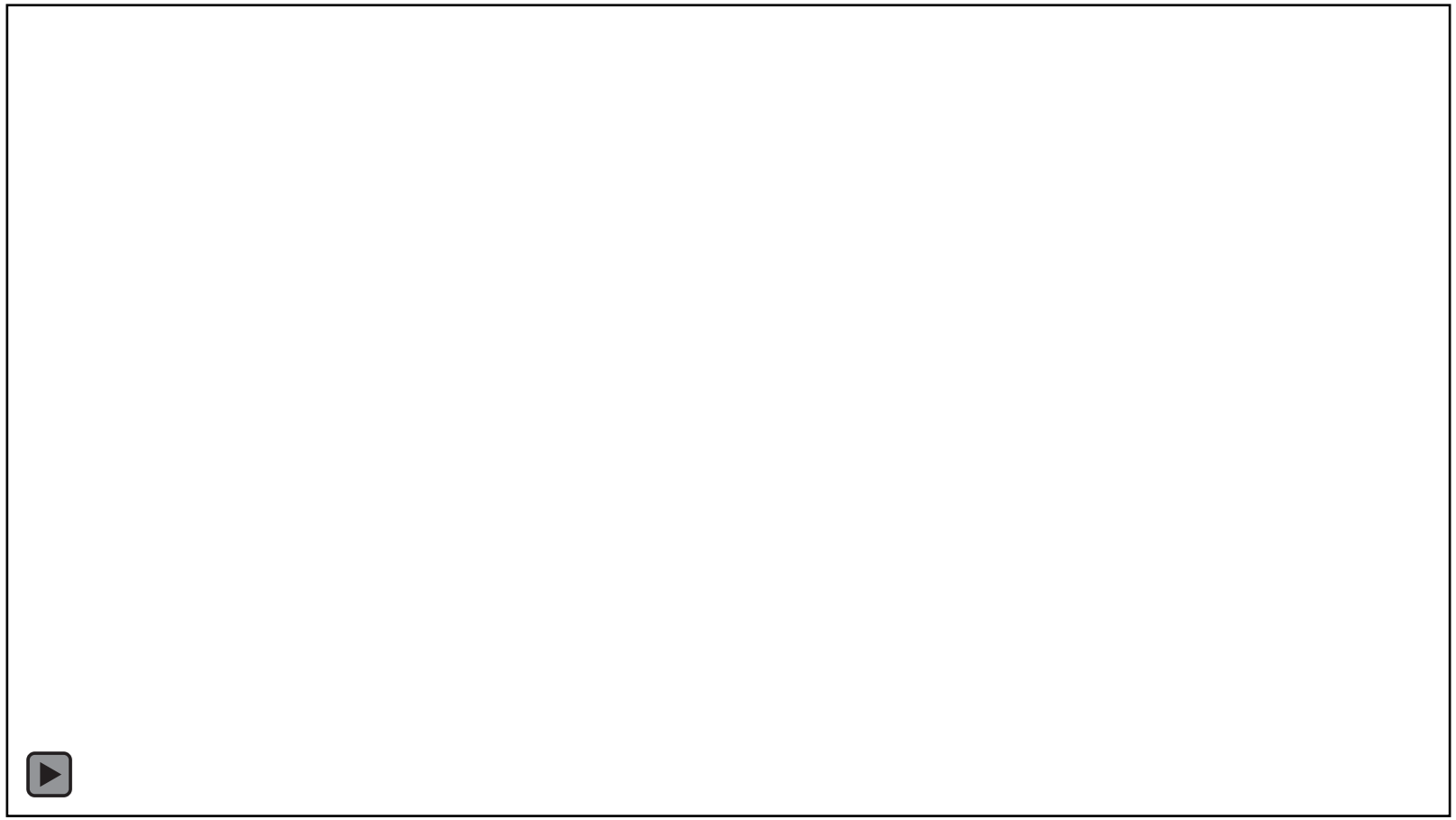
# VR for Science & Outreach

Research & Breeding

Education & Teaching

# Time-resolved 3D seed model – Virtual Seed

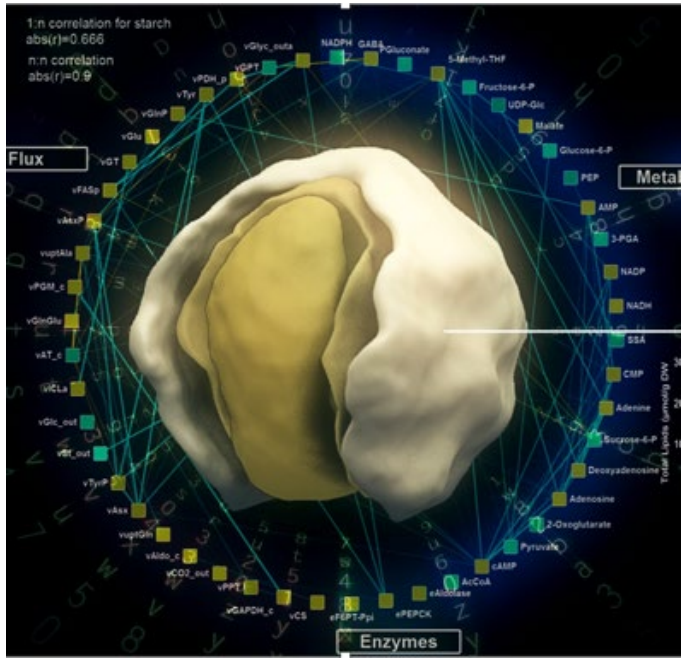
Center of a virtual and augmented reality environment for integrative and intuitive visualisation, evaluation, and exploration of complex data sets across multiple scales, domains and levels of processing



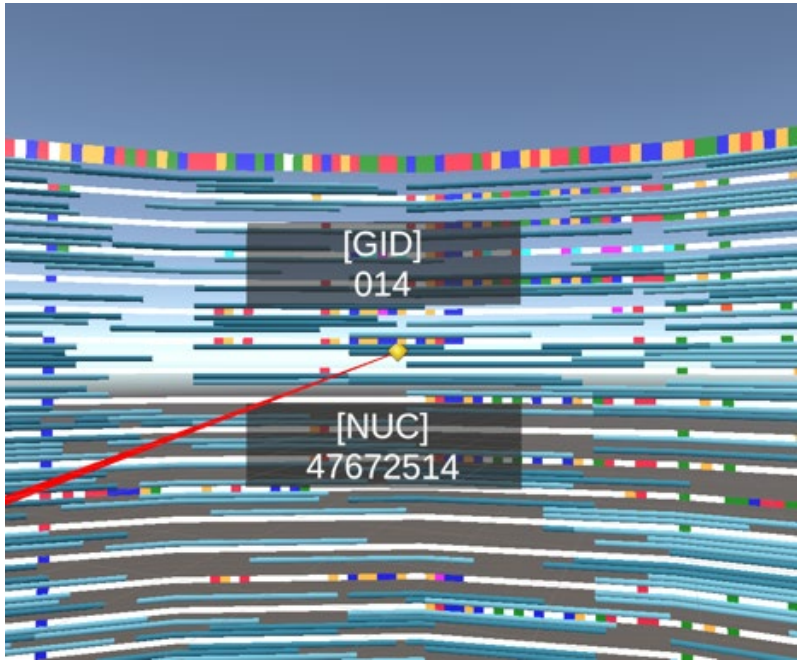
# Time-resolved 3D seed model – Virtual Seed

WINNER
BEST INDUSTRY VR
2021
NOW Awards

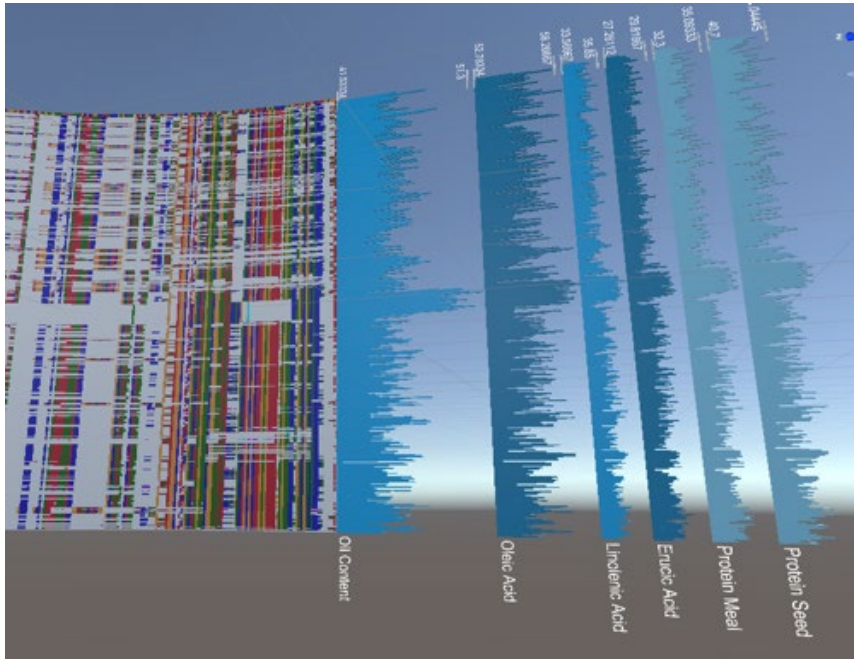
Center of a virtual and augmented reality environment for integrative and intuitive visualisation, evaluation, and exploration of complex data sets across multiple scales, domains and levels of processing




Choose stage and organ from dynamic 3D seed model




Display tissue-related transcripts aligned to genome tracks




Relate several phenotypic trait values to genome tracks

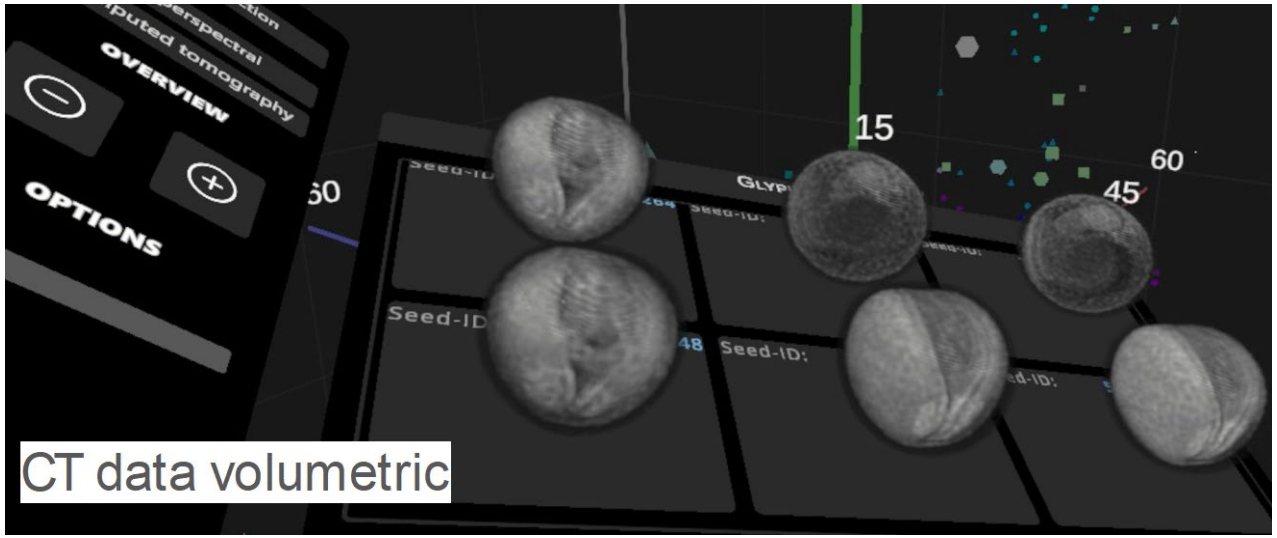
# Exploration of 2D/3D seed data in VR

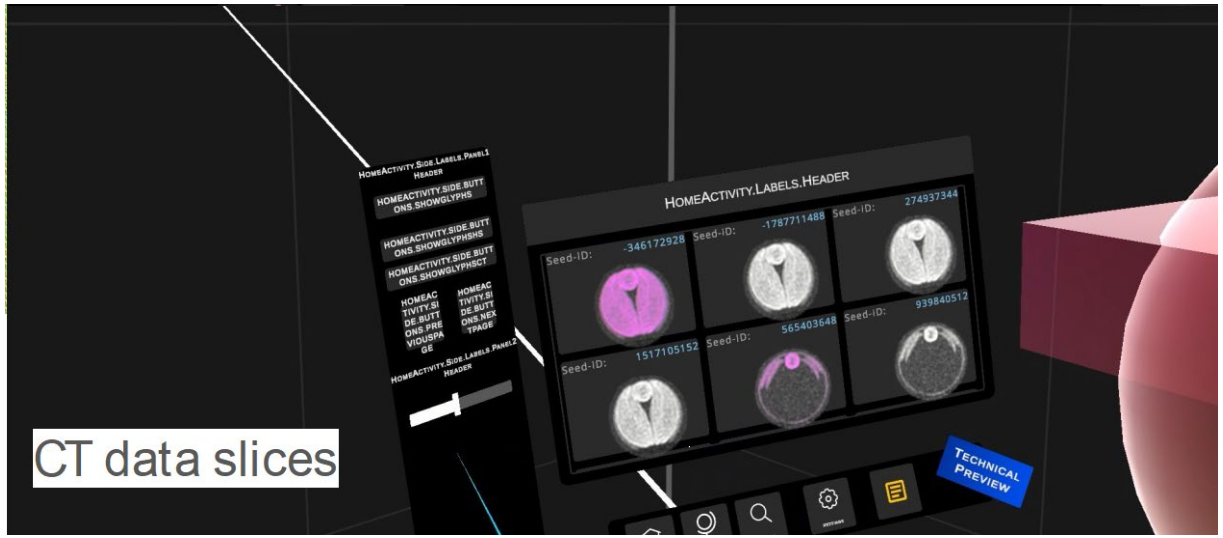
AVATARS

WINNER
BEST INDUSTRY VR
2021
NOW Awards

Center of a virtual and augmented reality environment for integrative and intuitive visualisation, evaluation, and exploration of complex data sets across multiple scales, domains and levels of processing

BREAK POINT ONE



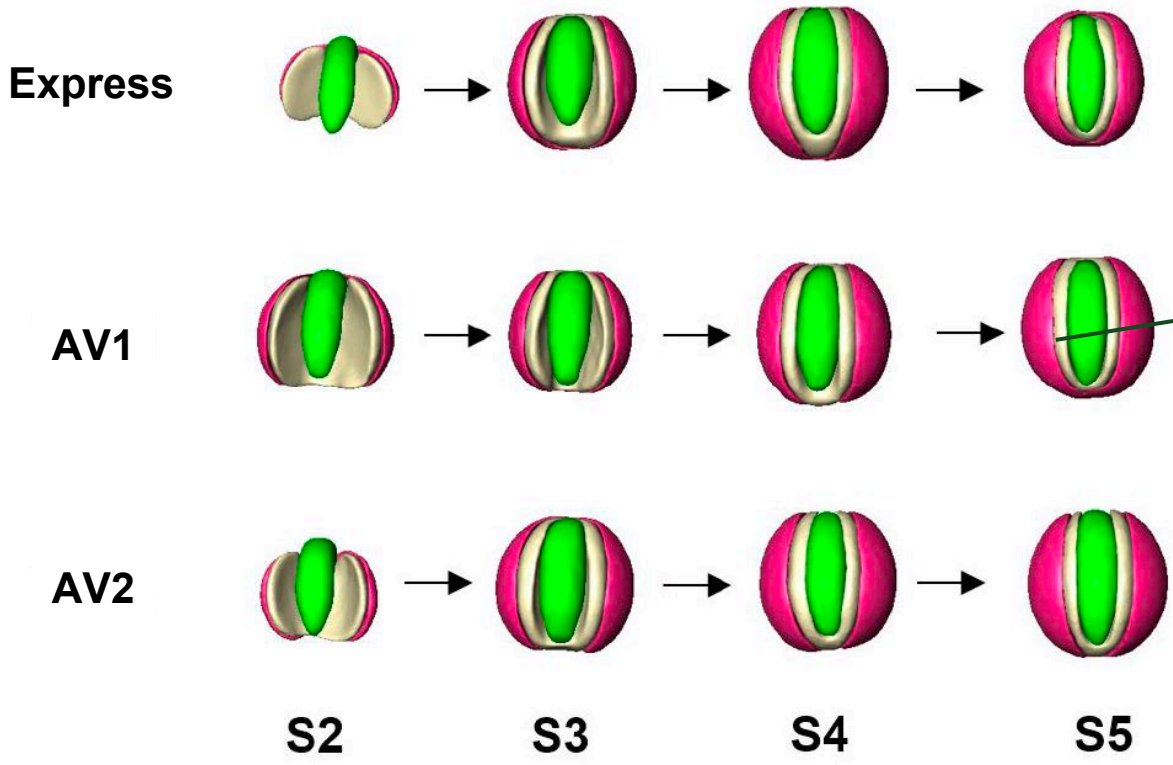

Visualization and exploration of x-ray CT images of seeds

# Exploration of 2D/3D seed data in VR

Center of a virtual and augmented reality environment for integrative and intuitive visualisation, evaluation, and exploration of complex data sets across multiple scales, domains and levels of processing

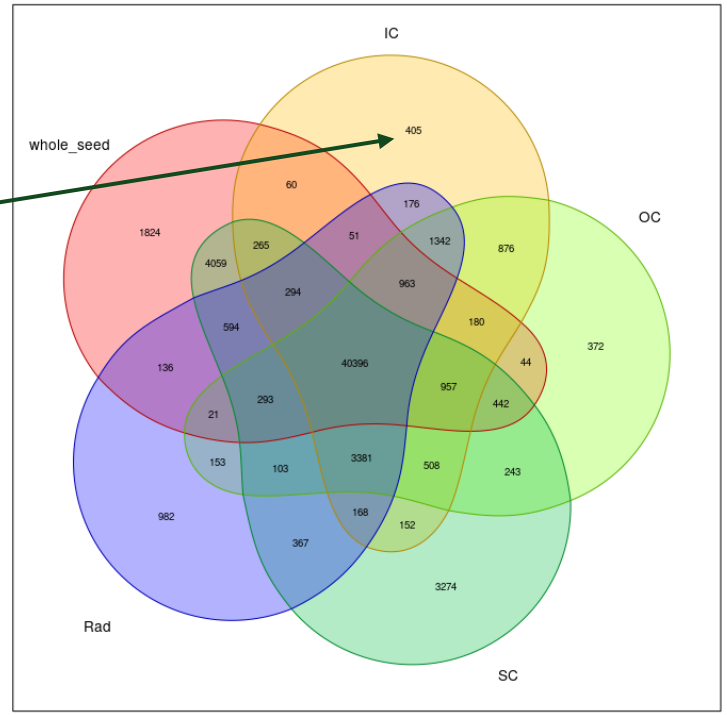
**3D structural organisation** (organs) and **changes**:
**Average seed models** across genotypes and stages

Organ-, stage, and genotype-resolved **multi-Omics data**




**organ-specific mRNA expression**

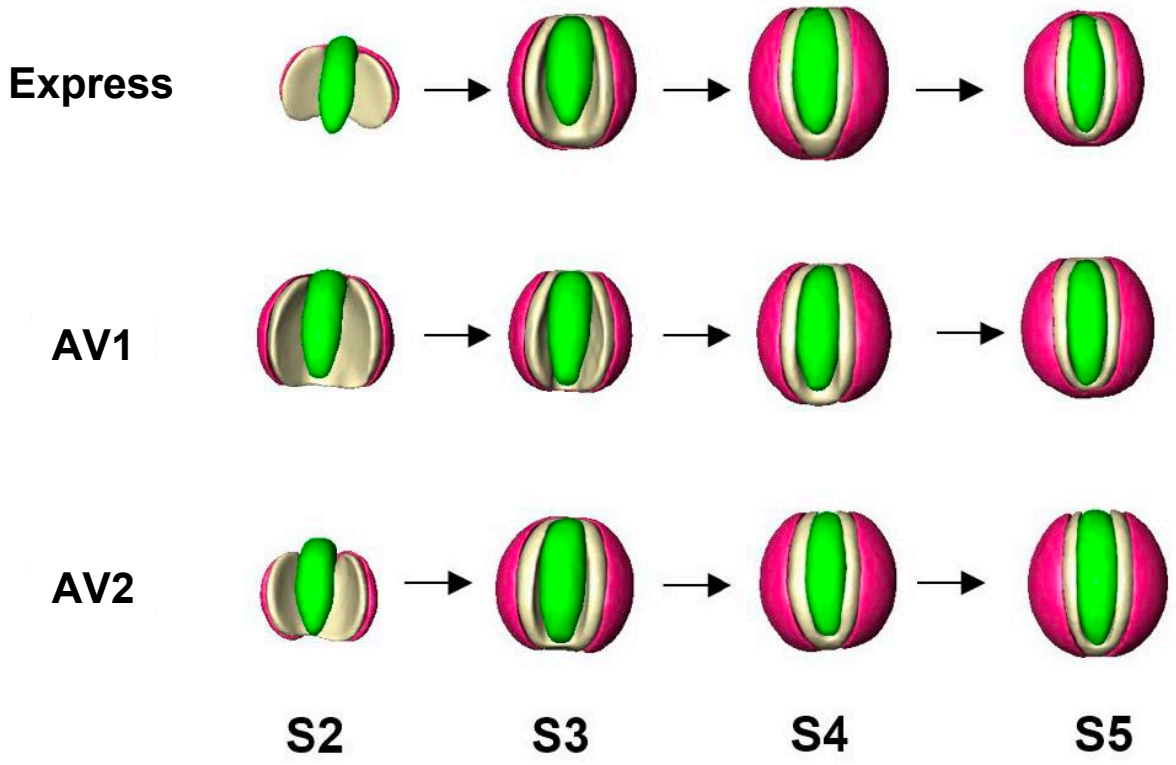
**expressed (>1 tpm in 2 of 3 replicates)**



# Exploration of 2D/3D seed data in VR

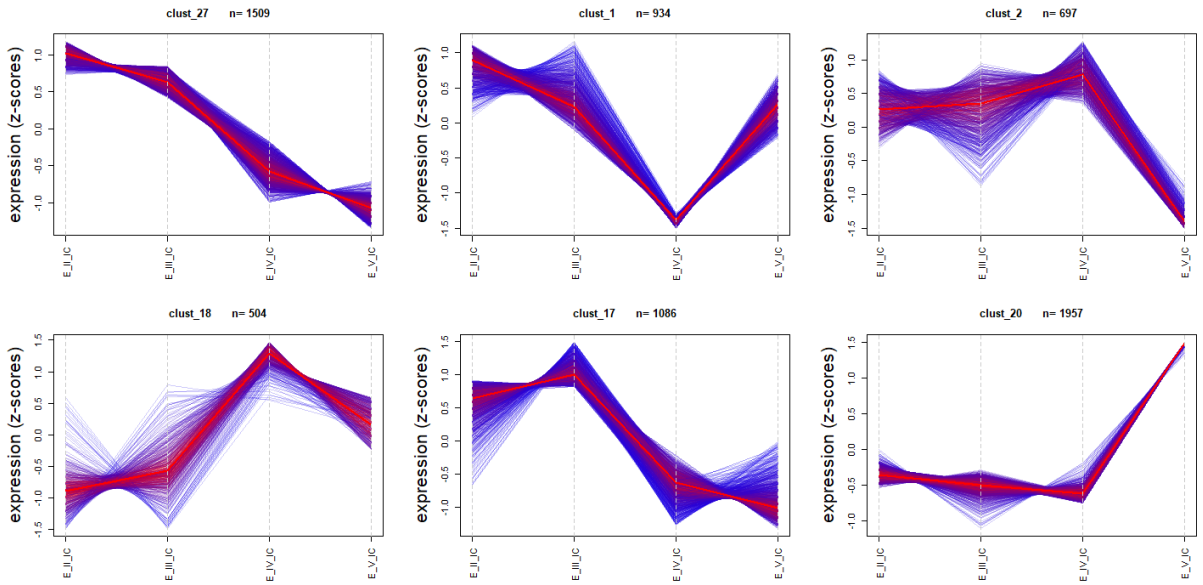
Center of a virtual and augmented reality environment for integrative and intuitive visualisation, evaluation, and exploration of complex data sets across multiple scales, domains and levels of processing

**3D structural organisation** (organs) and **changes**:
**Average seed models** across genotypes and stages




Organ-, stage, and genotype-resolved **multi-Omics data**

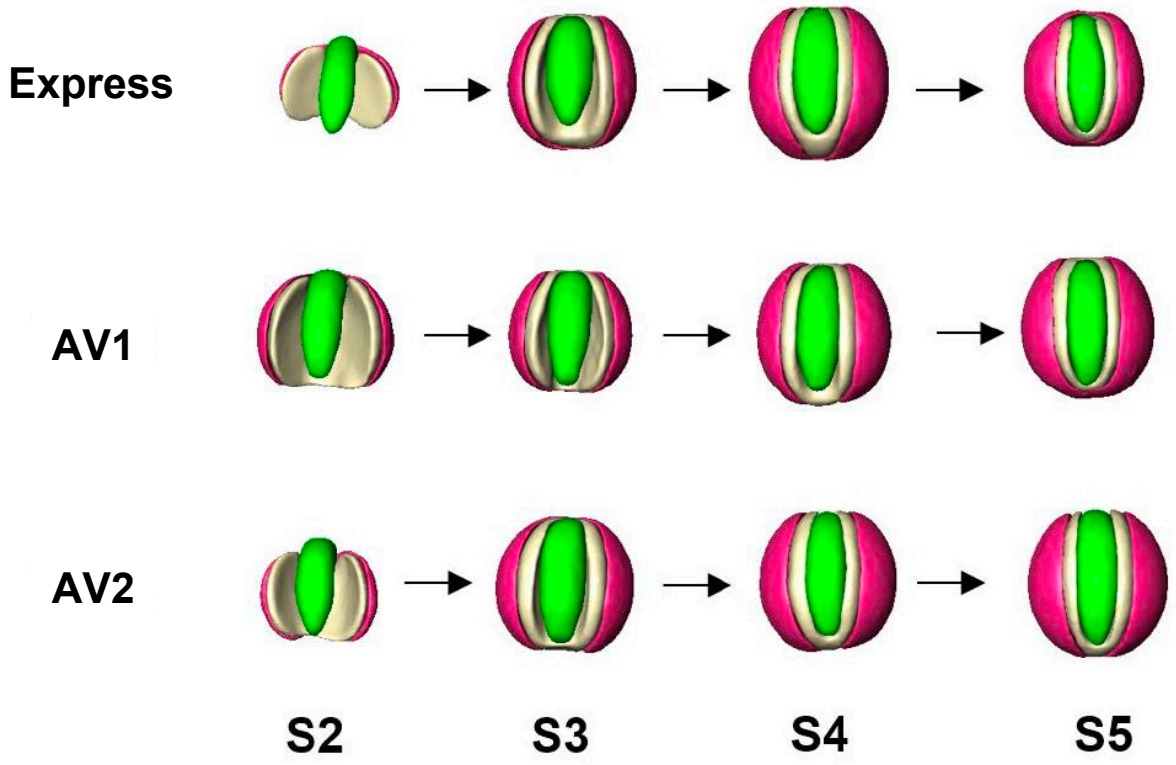
**fuzzy c-means clustering**
**(Express | inner cotyledons)**



# Exploration of 2D/3D seed data in VR

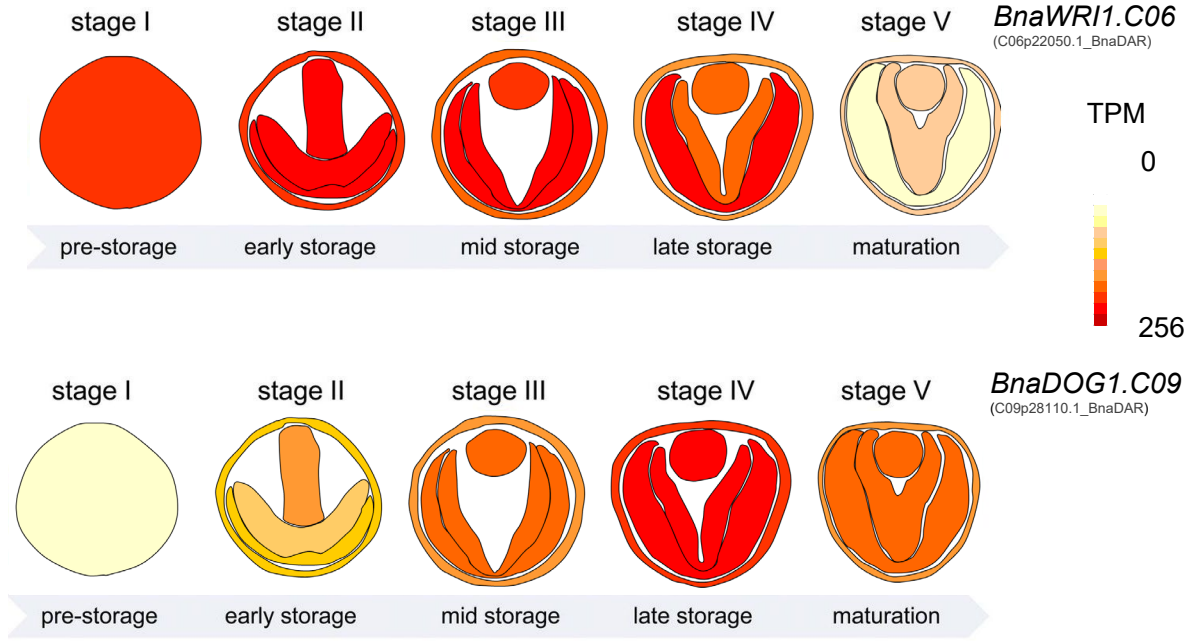
Center of a virtual and augmented reality environment for integrative and intuitive visualisation, evaluation, and exploration of complex data sets across multiple scales, domains and levels of processing

**3D structural organisation** (organs) and **changes**:
**Average seed models** across genotypes and stages




Organ-, stage, and genotype-resolved **multi-Omics data**

**False color displays of organ-/stage-specific gene expression values**



# VR for Science & Outreach

Research & Breeding

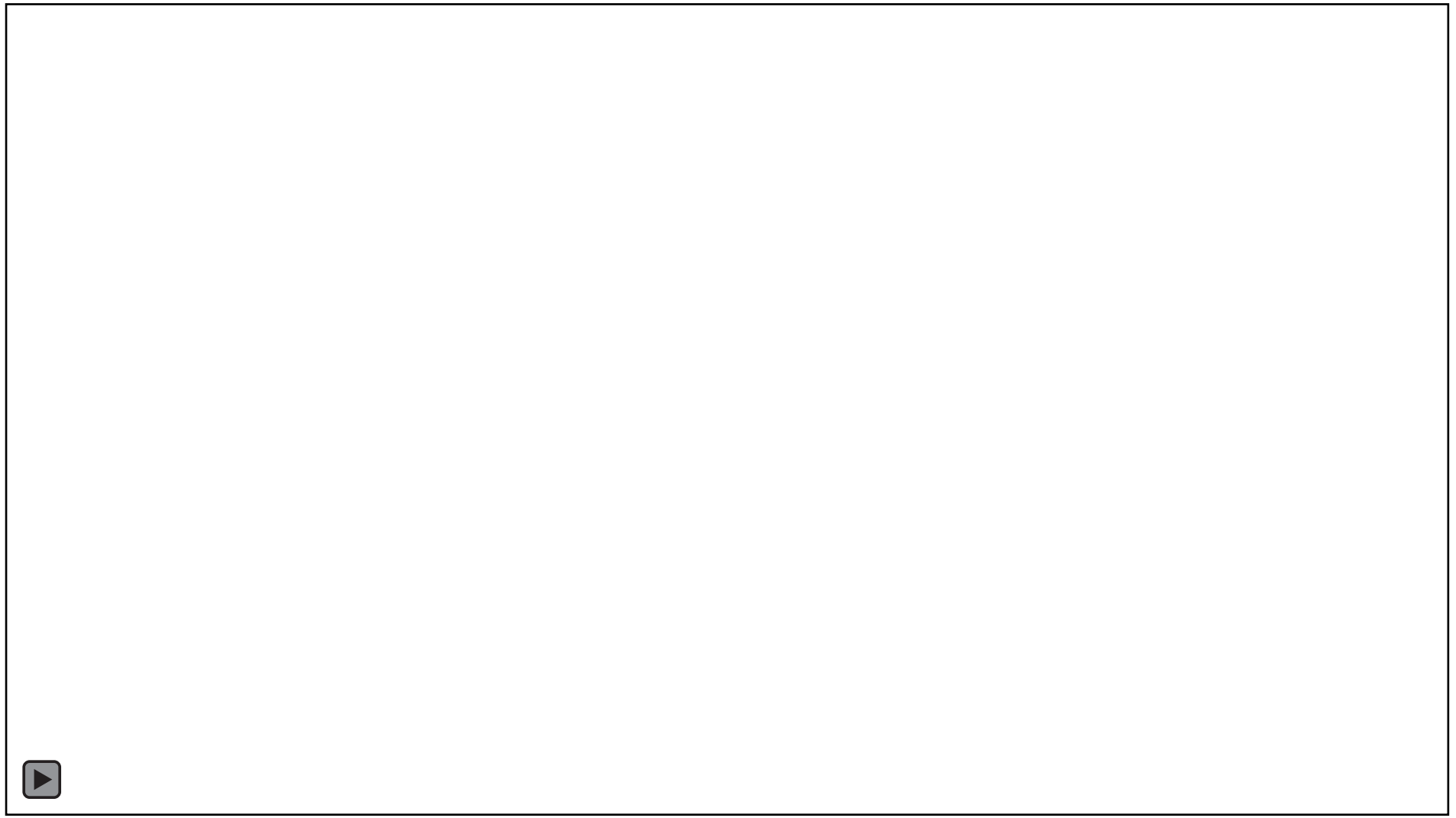
Education & Teaching

# VR Experience Plant Journey


The "VR Plant Journey" is a 15 minutes long room scale virtual reality edutainment experience

Federal Minister for Education and Research Anja Karliczek: *"This VR Plant Journey should be experienced in every school."* IdeenExpo 2019, Hannover

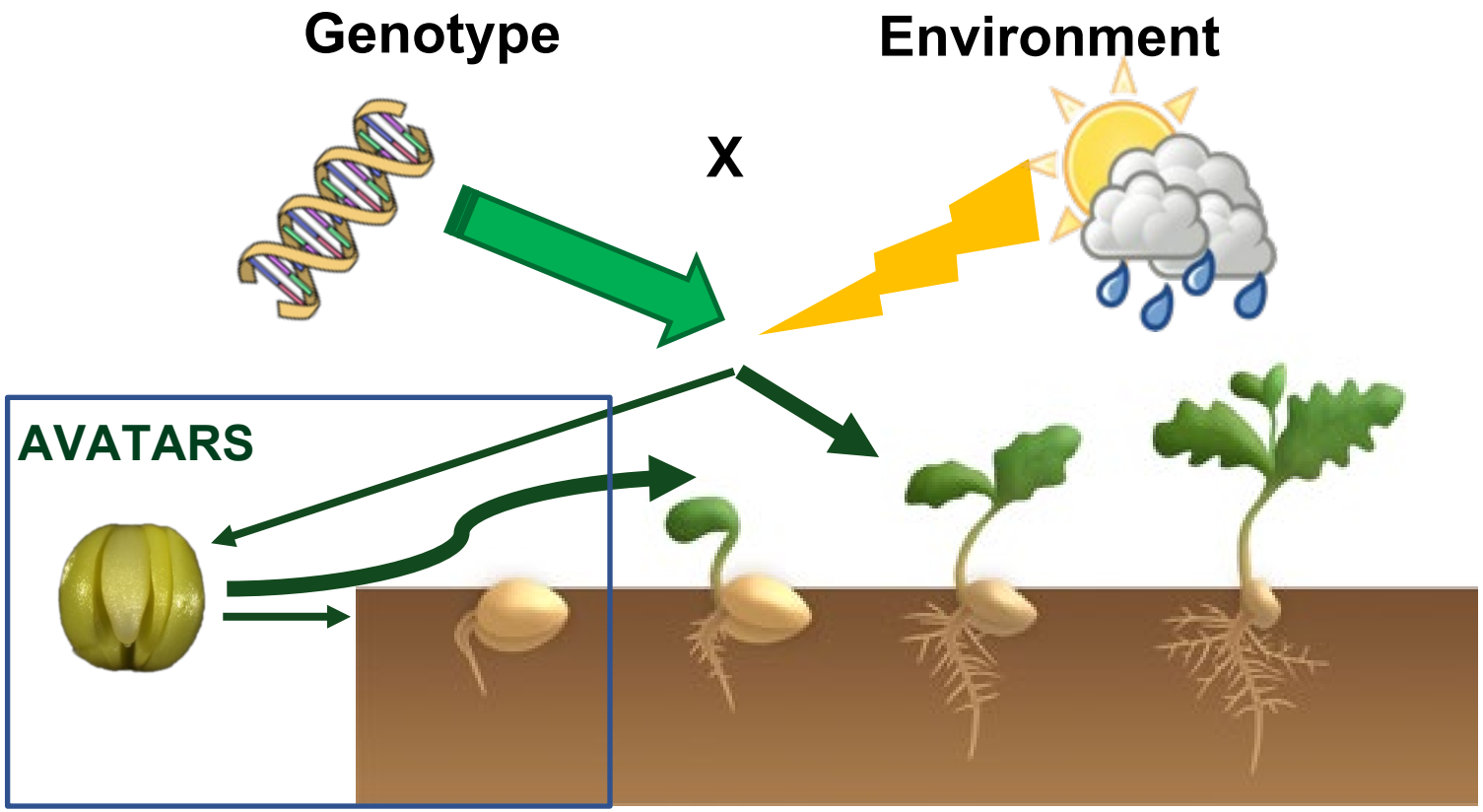
# VR Experience Plant Journey



# AVATARS – outlook

## `AVATARS II – the next level´

an **integrated data science** project on **seed germination, seedling growth,** and **crop establishment** addressing **environment resilience** to migitate consequences of **climate change**



# Acknowledgements


**Dominic Knoch**
**Lothar Altschmied**
**Marc Heuermann**
**Markus Kuhlmann**
**Rhonda C. Meyer**
IPK, RG HET

**Henning Tschiersch**
**Astrid Junker***
**Kerstin Neumann**
IPK, RG ADP/APP

**Michael Henke***
**Narendra Narisetti**
**Evgeny Gladilin**
IPK, RG BA

**+ many technicians and gardeners!!**

**Steffen Wagner**
**Stefan Ortleb**
**Alexander Hilo**
**Peter Keil**
**Hardy Rolletschek**
**Ljudmilla Borisjuk**
IPK, RG AAN

**Johannes Thiel**
**Jozefus Schippers**
IPK, RG SE

**Paride Rizzo**
**John C. D´Auria**
IPK, RG MD

**Mary Ann Blätke**
IPK, RG IMM

**Daniel Arend**
**Elena Rey Mazon**
**Matthias Lange**
**Uwe Scholz**
IPK, RG BIT

**Jens Freitag**
IPK, GS

**Matthias Enders**
**Jennifer Zur**
**Simon Görtz**
**Amine Abbadi**
NPZi

**Nils Rugen**
**Hans-Peter Braun**
UHann

**Bart Verwaaijen**
**Anja Meierhenrich**
**Andrea Bräutigam**
(IPK, RG NAM) UBi

**Philipp Liznerski**
**Marius Kloft**
TUKl

**Christine Ripken**
**Gergö Kovács**
**Sebastian Tusk**
**Olaf Sacher**
BPO

WINNER
BEST INDUSTRY VR
2021
NOW Awards

Federal Ministry of Education and Research

PLANT 2030

# Thank you!

www.avatars-project.de

@AvatarsProj

80 ANNIVERSARY IPK LEIBNIZ INSTITUTE

UNIVERSITÄT BIELEFELD

NPZ INNOVATION

Leibniz Universität Hannover

RP TU Rheinland-Pfälzische Technische Universität Kaiserslautern Landau

BREAK POINT ONE

# AVATARS – an integrative Data Science project

Genotype X Environment

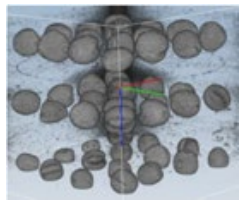
G x E !!

## Field Cultivation

Plant & Evironmental Data
400/100 Lines
11 env.
genotype
15k SNPs
phenotype
weather data

## Single Seed Data

> 350.000 seeds
TD-NMR
hyperspectral images
X-ray CT
germination



## Deep Learning

neural networks using ultra high-throughput data

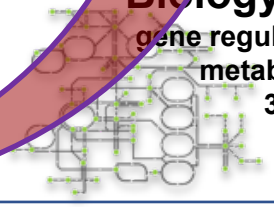
## Seed Quality

Winter Oil Seed Rape

viability germination ability

## Systems Biology

gene regulatory network
metabolic networks
3D seed model

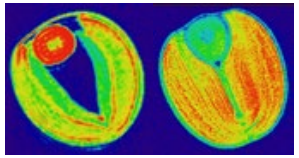


## Controlled Environment

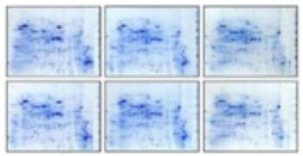
40/3 Lines
2 env. scenarios

Time-resolved Seed Architecture

MRI
Histology

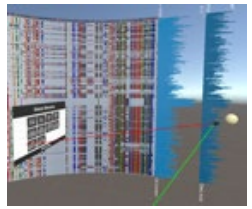


Time- and Space-resolved Multi-Omics



DNA methylation
mRNA / sRNA
proteins
metabolites
phytohormones

## VR visualization



improved human access to Big Data and support of efficient exploration and Science Outreach

**Plant Breeding 4.0** ← VR visualization → **Seed Systems Biology**