GWAS with high-throughput machine vision phenomics uncovers genetic regulators of *in vitro* transformation and regeneration in *Populus trichocarpa*

Plant and Animal Genome 2023 Conference
Forest Trees Session

Michael F. Nagle
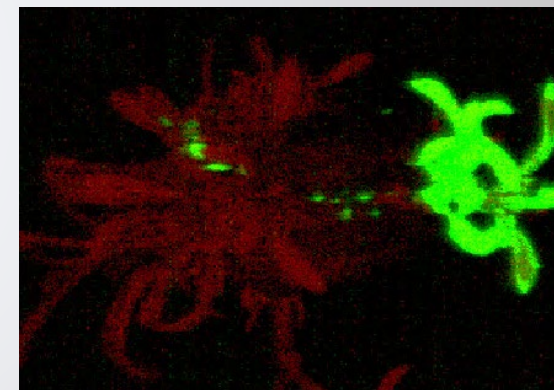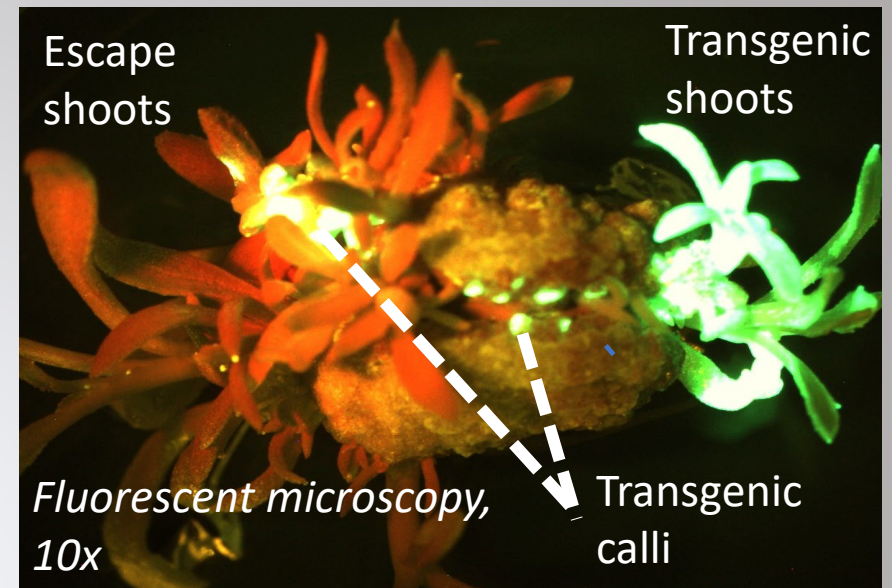Department of Forest Ecosystems & Society, Oregon State University, Corvallis OR

Presentation Overview

Objective:
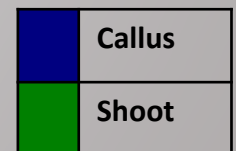Identify genetic regulators of plant transformation, a complex trait, with high statistical power

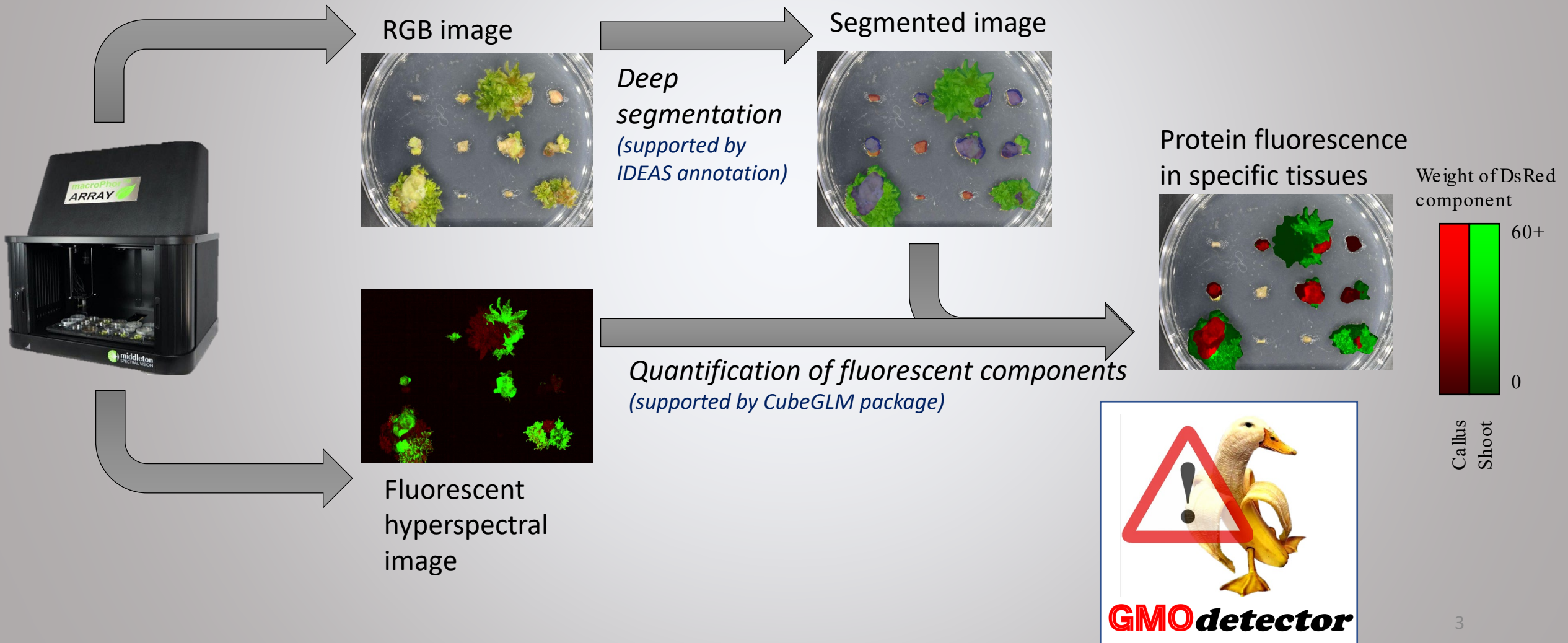I.   High-throughput phenomics

II. Multiple-workflow association mapping

Escape shoots

Transgenic shoots

Fluorescent microscopy, 10x

Transgenic calli

Hyperspectral imaging

Deep segmentation

	Callus
	Shoot

Strategy to integrate RGB, hyperspectral data to measure tissue-specific transformation + regeneration

RGB image

Deep segmentation (supported by IDEAS annotation)

Segmented image

Protein fluorescence in specific tissues

Quantification of fluorescent components (supported by CubeGLM package)

Fluorescent hyperspectral image

Weight of DsRed component

60+

0

Callus

Shoot

GMOdetector

3

CubeGLM (Python) provides user-friendly Python classes and functions for hyperspectral analysis

- ***Challenge:*** Need to efficiently analyze 21,908 hyperspectral images (~1.5GB each, 32.8TB total)

- CubeGLM package key features
  - Quickly compute weights of fluorescent components per pixel
  - Allow visualization of results
  - Parallelization and efficiency

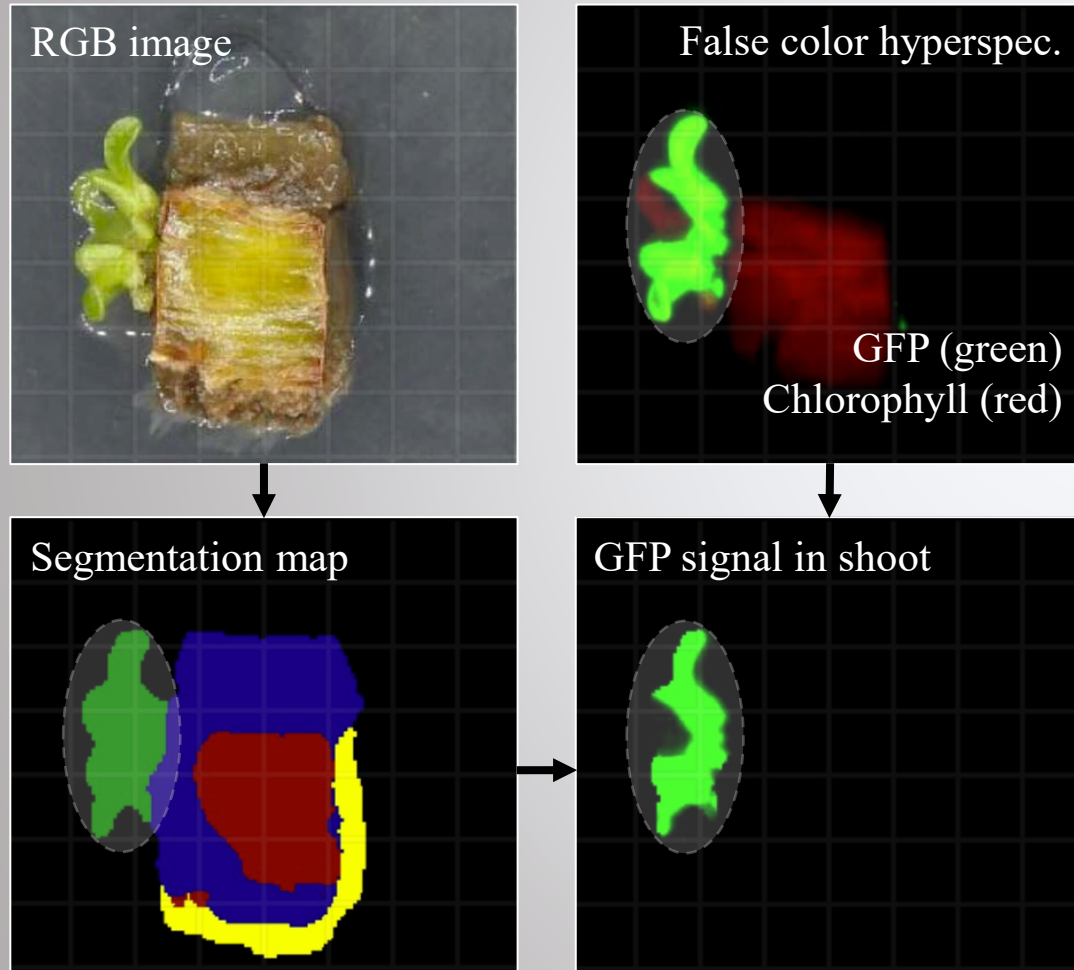- https://github.com/naglemi/gmodetector_py

FAC2_4_R25
SKWR-24-2_WOX11

Red: Chlorophyll

Green: GFP

Blue: Background / Noise

GFP shoot

GMOdetector final outputs:
Summary statistics of transformation and regeneration

RGB image

False color hyperspec.

GFP (green)
Chlorophyll (red)

Segmentation map

GFP signal in shoot

- Two types of statistical outputs of greatest interest
  - Transformation frequency
  - Transformed tissue size
- Tailored, trained for poplar transformation, but generalizable with further work
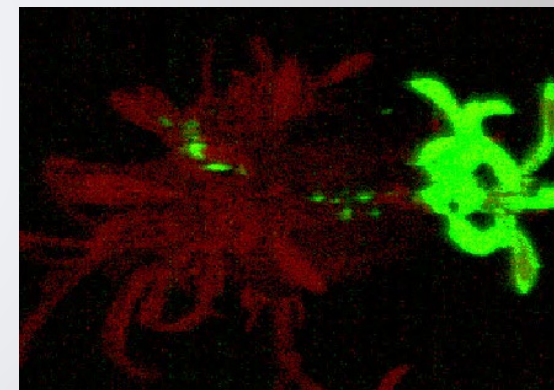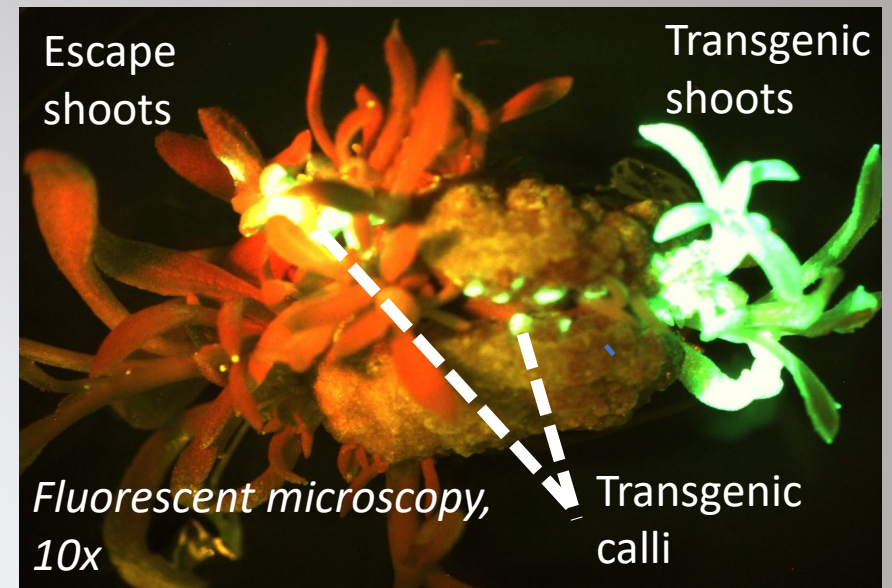 - https://github.com/naglemi/GMOnotebook

Presentation Overview

Objective:
Identify genetic regulators of plant transformation, a complex trait, with high statistical power
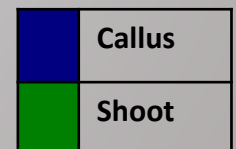
I. High-throughput phenomics

II. Multiple-workflow association mapping

Escape shoots

Transgenic shoots

Fluorescent microscopy, 10x

Transgenic calli

Hyperspectral imaging

Deep segmentation

|	
■	Callus
■	Shoot

Poplar GWAS resources an asset to statistical power

- 1323 wild genotypes collected over a diverse range in Pacific Northwest

- Rapid linkage disequilibrium (LD) decay ($R^2 < 0.2$ within 3kb)

- High SNP density
  - As many as ~30 million SNPs studied
  - ~13 million common SNPs

OAK RIDGE
National Laboratory

cbi

THE CENTER FOR
BIOENERGY INNOVATION

7
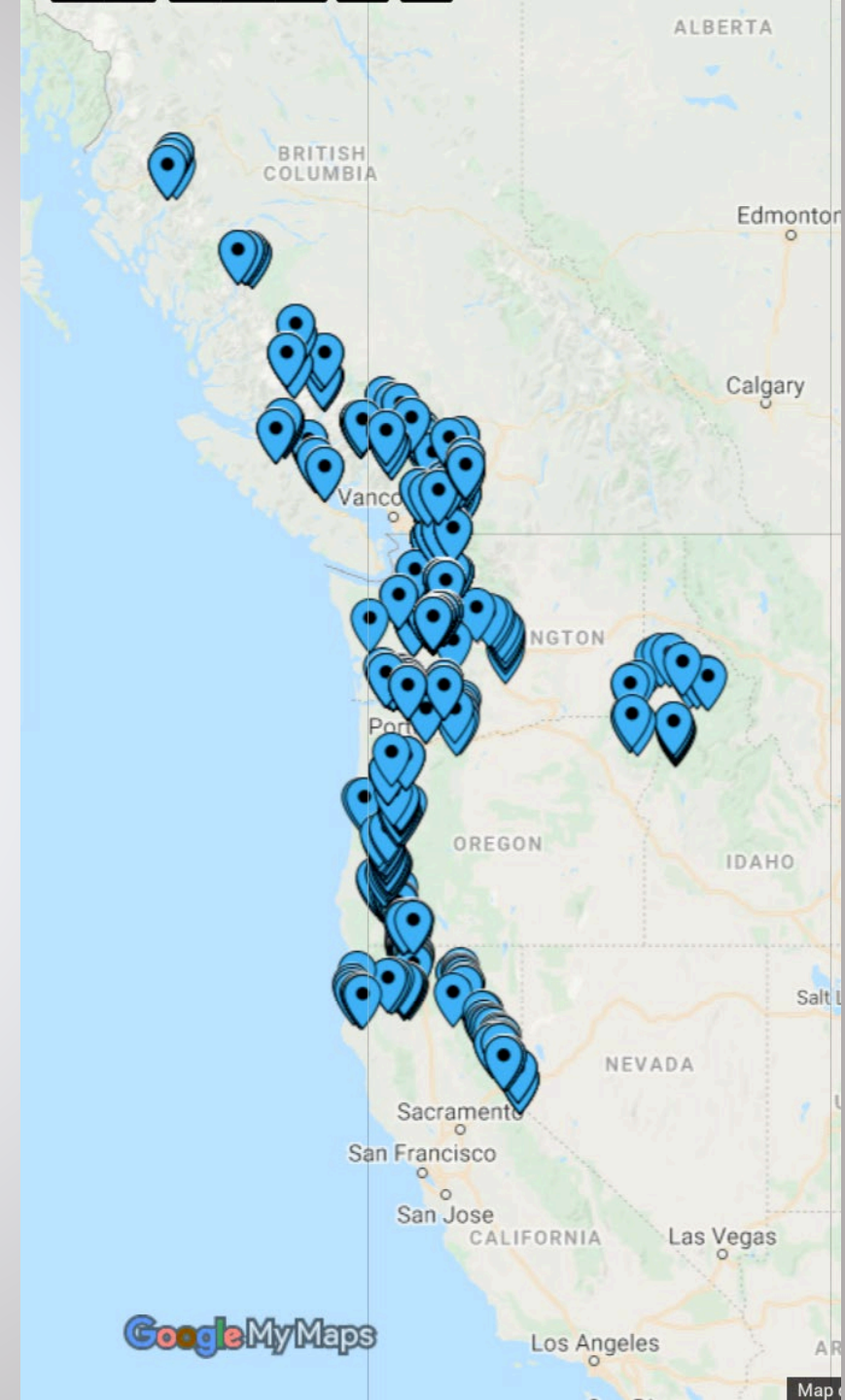
Non-normality, transformations obstacle to high power

- Transformations (one way to avoid violating assumption of residual normality) can overcorrect, reduce power

- Permutations (Monte Carlo methods) to compute empirical *p*-values
  - Avoid linear model assumptions and transformations
  - Great computational expense
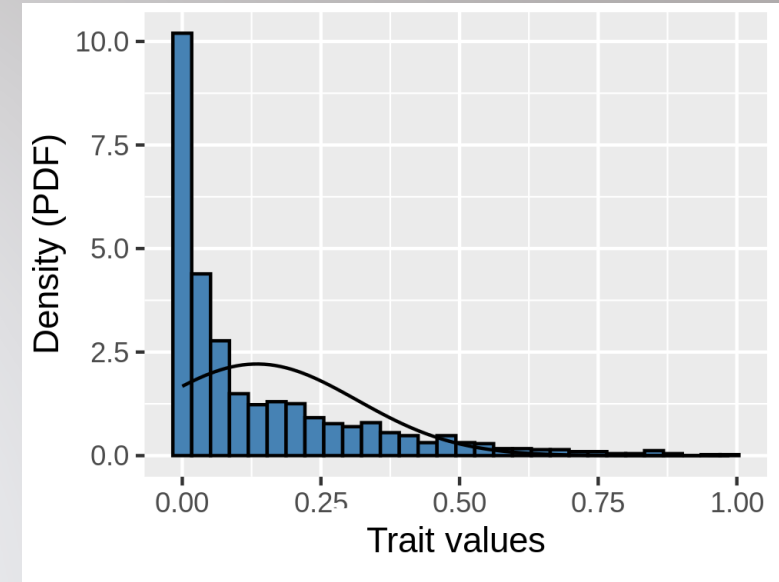
Box-Cox transformation
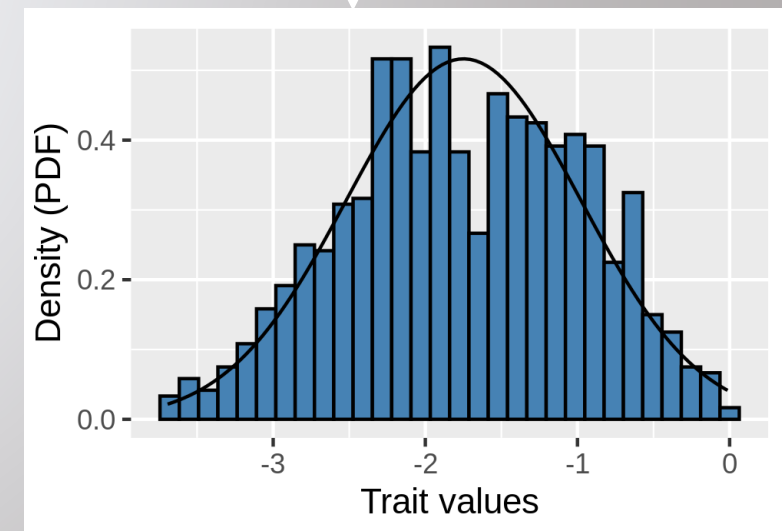
Two approaches to increase power by combining power from adjacent genetic markers

AJHG ARTICLE

Sequence Kernel Association Tests
for the Combined Effect of Rare and Common Variants

Iuliana Ionita-Laza,[1,6,*] Seunggeun Lee,[2,6] Vlad Makarov,[1] Joseph D. Buxbaum,[3,4,5] and Xihong Lin[2,*]

Detecting Weak Signals by
Combining Small P-Values in Genetic
Association Studies

Olga A. Vsevolozhskaya[1], Fengjiao Hu[2] and Dmitri V. Zaykin[1*]

frontiers
in Genetics

**SNP-set Kernel Association Test (SKAT) and
Multi-Threaded Monte Carlo SKAT (MTMC-SKAT)**

Augmented Rank Truncation (ART)

Resolution

- No clue on which marker(s) within "window" most critical

- Single-marker resolution from initial GEMMA, other single-SNP tools.

Assumptions

- Avoid assumptions by computing empirical *p*-values

- Relies on prior test with linear model assumptions

MTMCSKAT enables us to run SKAT with efficient resampling over many traits

- Three key feature additions to SKAT:

  1. **Adaptive resampling:**
     How many permutations needed for given p-val?

  2. **Adaptive multithreading:**
     How to parallelize most efficiently?

  3. **Easy usage with high-performance clusters (HPCs)**

- https://github.com/naglemi/mtmcskat

MTMCSKAT development and deployment supported by NSF ACCESS HPC allocation

ACCESS | Advancing Innovation

Augmented Rank Truncation (ART) increases statistical power for GEMMA and GMMAT tests

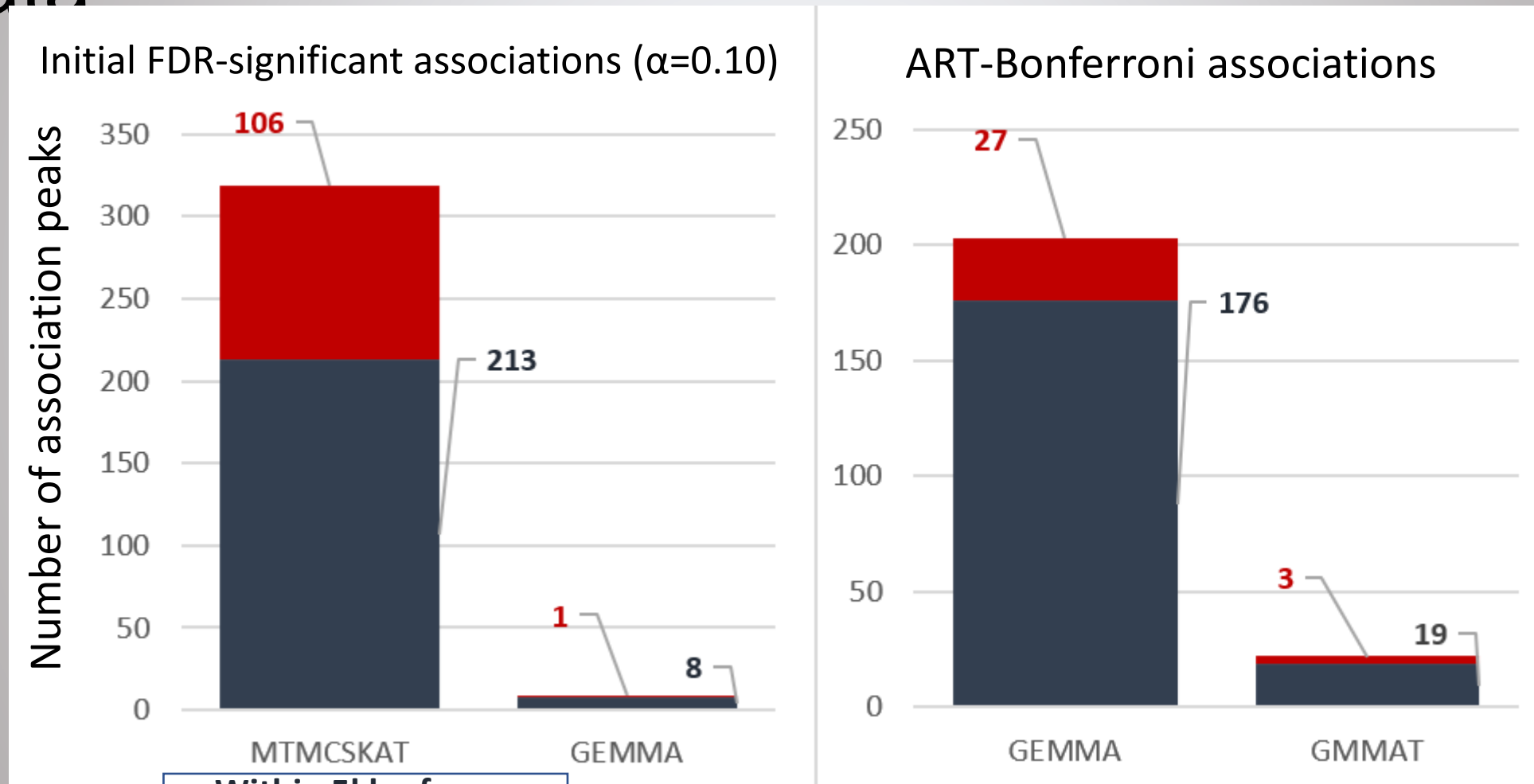Manhattan plot for GEMMA results for callus size trait
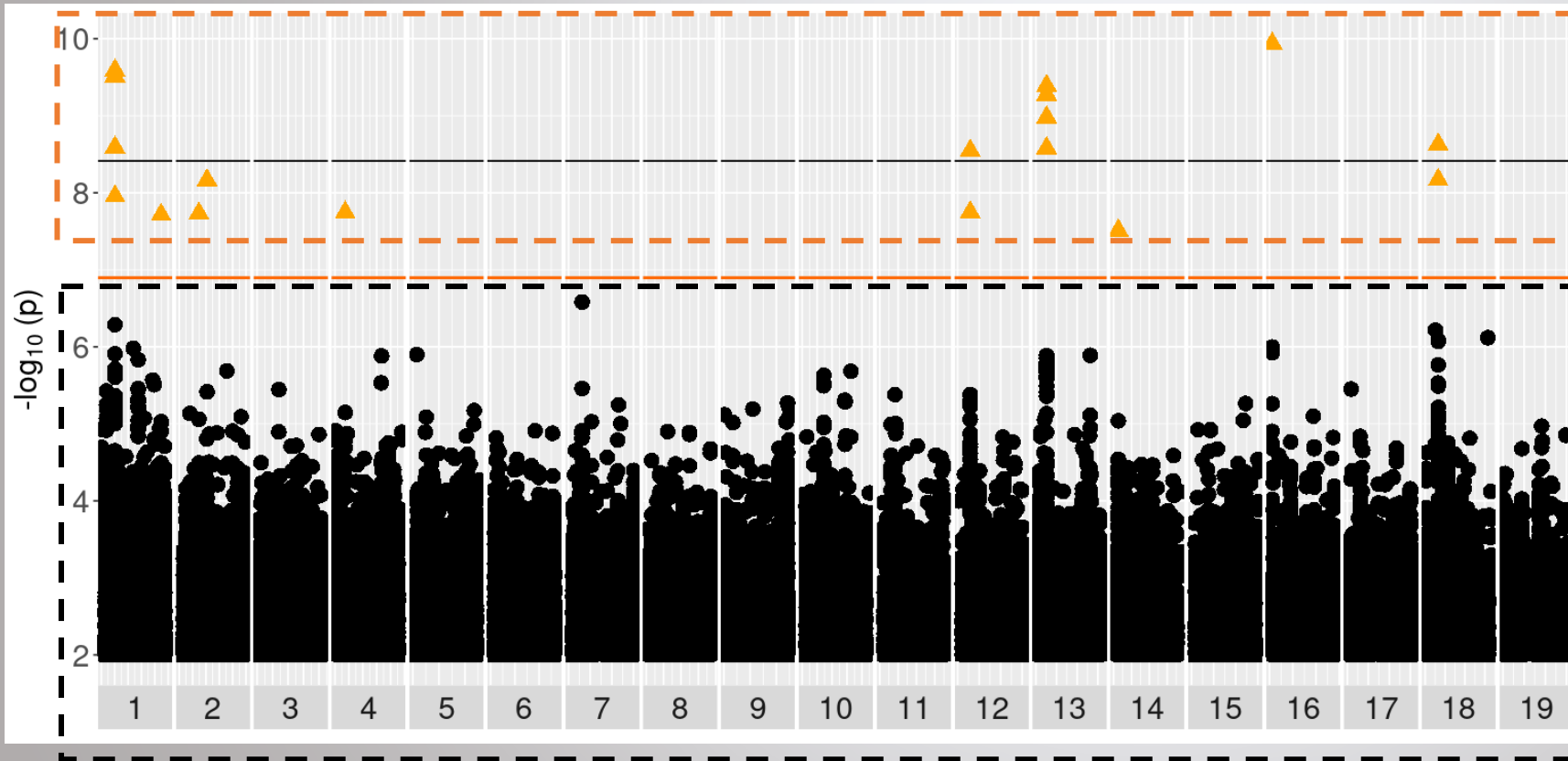
- With GEMMA-ART, windows integrate nearby SNPs.
- Window *p*-values often significant when individual SNPs not

No significant associations with GEMMA by itself
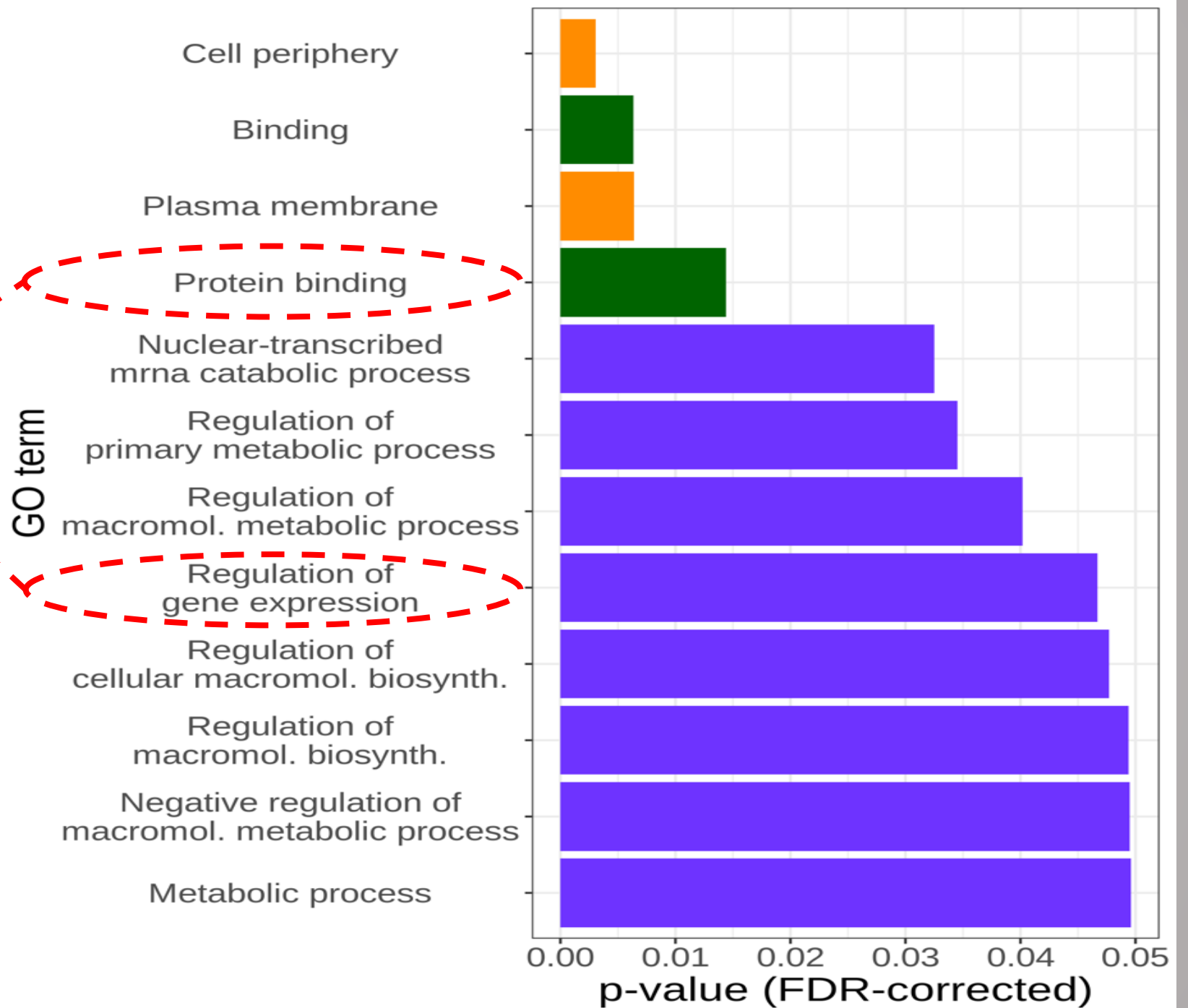
Method
- ● GEMMA
- ▲ GEMMA-ART

Across methods and significance thresholds, MTMC-SKAT and ART most powerful for our data

Gene ontology overrepresentation

Key roles of transcription factors, protein-protein interactions suggest complex regulatory network

Importance of wound/defense signaling in regeneration:
Many gene candidates co-regulated by PI4Kβ (salicylic acid signaling) and auxin/cytokinin signaling regulators

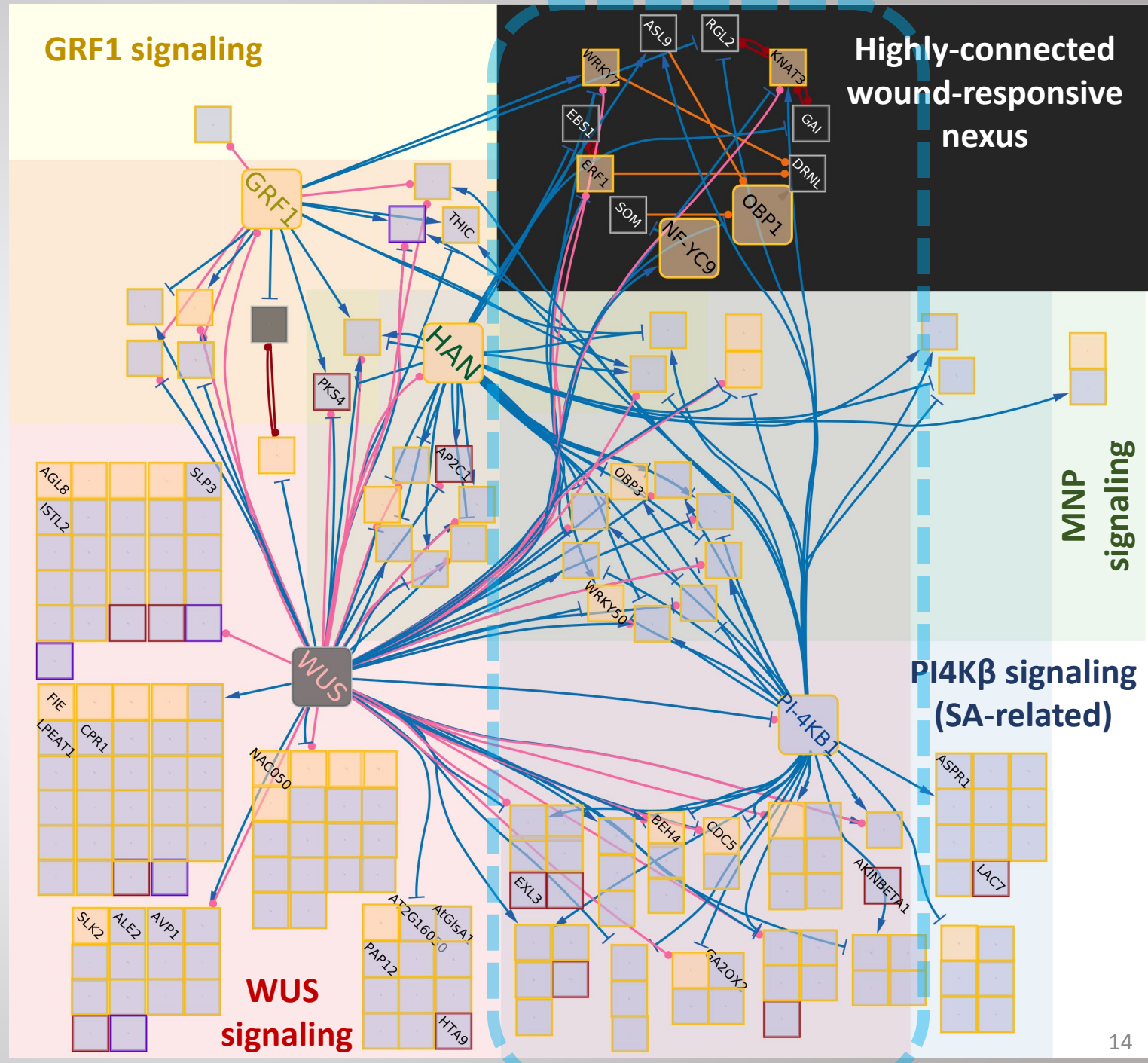Category of gene
- Gene expression regulator candidate gene
- Other candidate gene
- Not candidate, but key interactor

Relationships between genes
- DEG (downreg.)
- DEG (upreg.)
- Yeast Two-Hybrid
- Yeast One-Hybrid
- ChIP-seq

Ten published high-throughput datasets from Arabidopsis leveraged to build network

GRF1 signaling

Highly-connected wound-responsive nexus

MNP signaling

PI4Kβ signaling (SA-related)

WUS signaling

14

Key gene candidates regulate hormone signal within and across phytohormone pathways

Mechanistic connections between gene candidate Arabidopsis homologs
from *in vitro, in planta* experiments across 78 studies
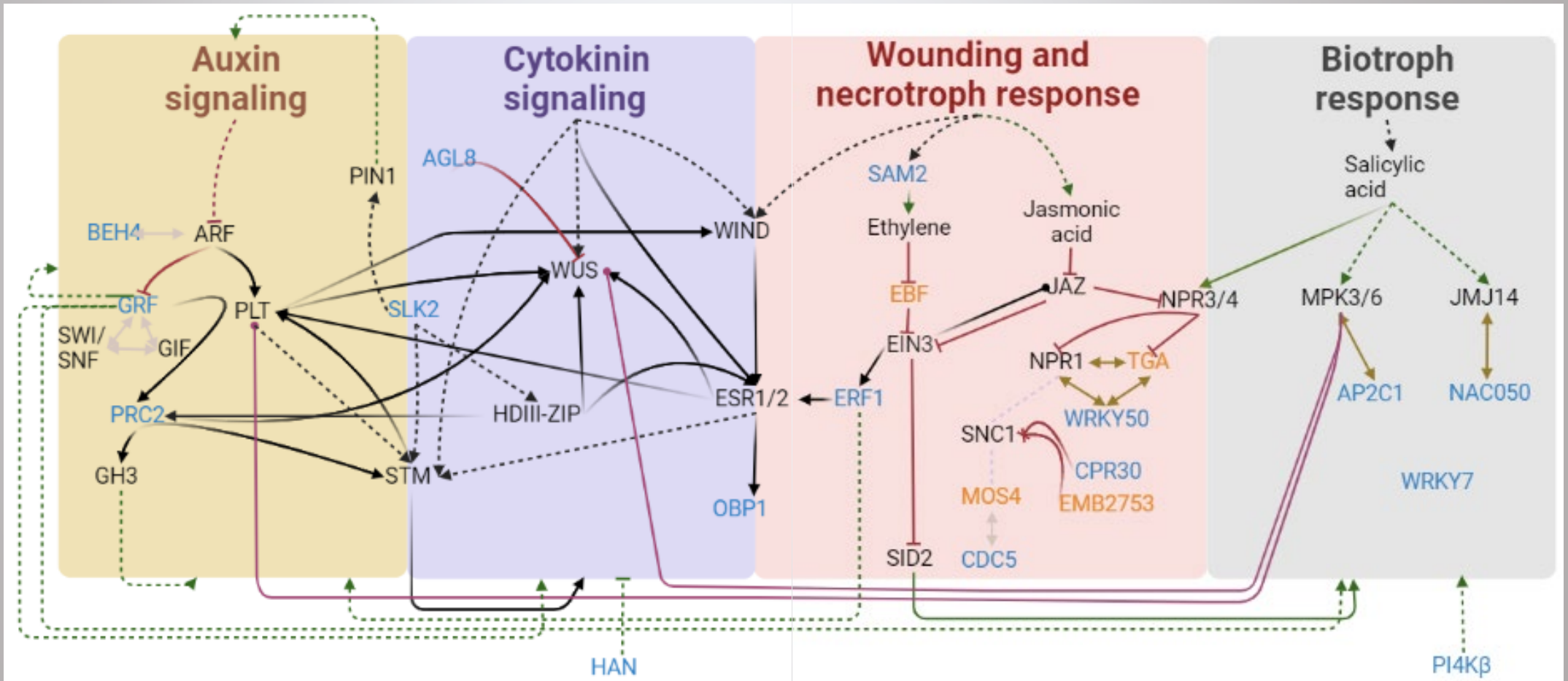
Summary

- GWAS relies on statistical power, which is improved with high-throughput phenomics and combined variant testing.

- Found 409 QTLs, investigated functions and relationships with gene ontology and network analysis

- Regeneration/transformation in poplar depends not only on auxin/cytokinin pathways, but wound/defense pathways associated with salicylic acid, jasmonic acid and ethylene

Acknowledgements

Cathleen Ma
Transformation & Greenhouse Experiments

Kate Peremyslova
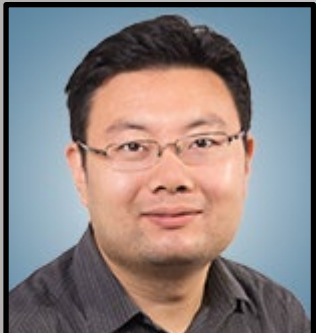GWAS, Transformation Experiments

Julie Kucinski,
GWAS, in vitro experiments
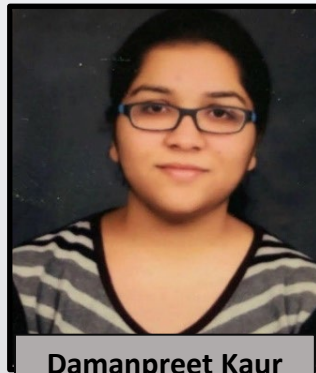
Steve Strauss
PI, Professor

Amanda Goddard
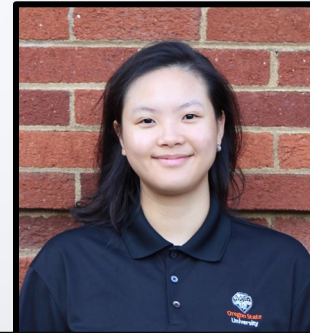Program & Field (clone bank) Manager

Fuxin Li
Co-PI, Professor, Machine Vision

Jialin Yuan
PhD Student, Machine Vision

Damanpreet Kaur
Master's Student, Machine Vision

Jia Yi Li
Undergraduate Student, Machine Vision

Yuan Jiang
Co-PI, Professor, Statistics