The sc-eQTLGen consortium framework: a federated pipeline for genotype- and phenotype-based association analyses

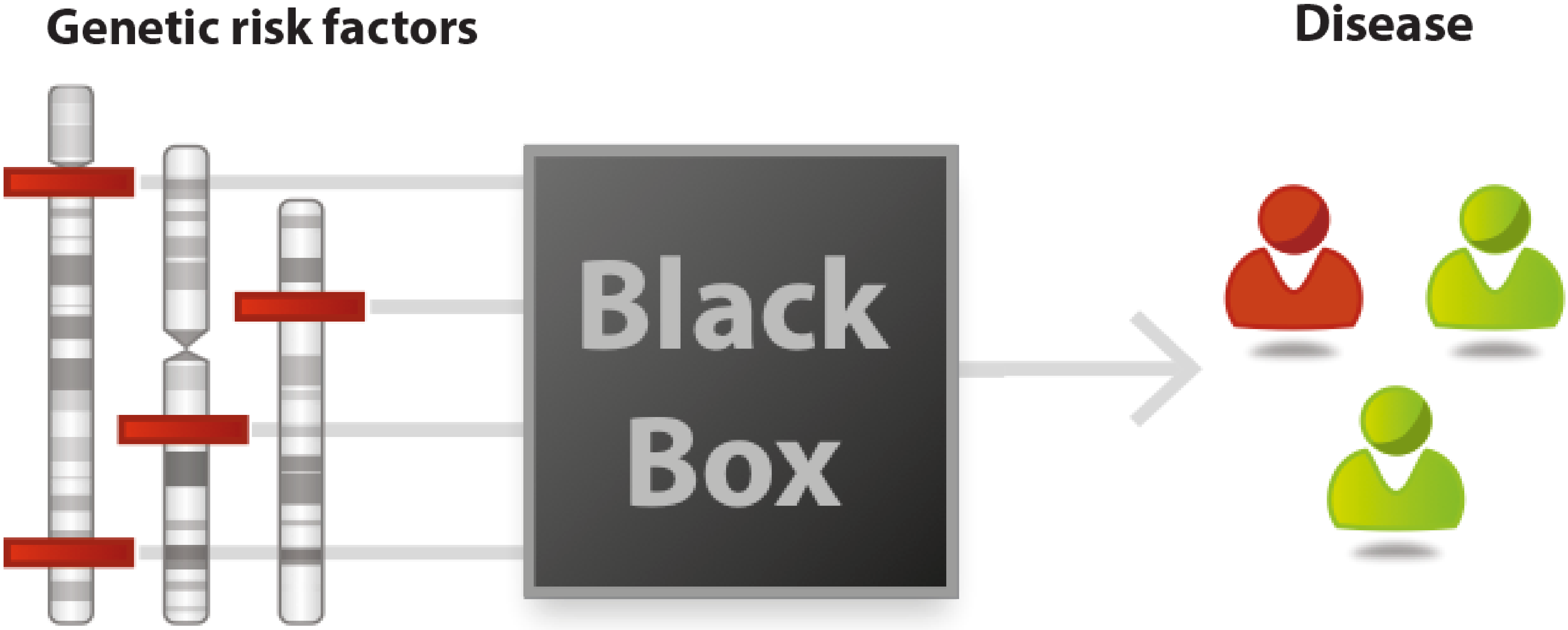
Monique van der Wijst

Department of Genetics

University Medical Center Groningen, The Netherlands

m.g.p.van.der.wijst@umcg.nl

From genotype to phenotype



Genetic risk factors

Black Box

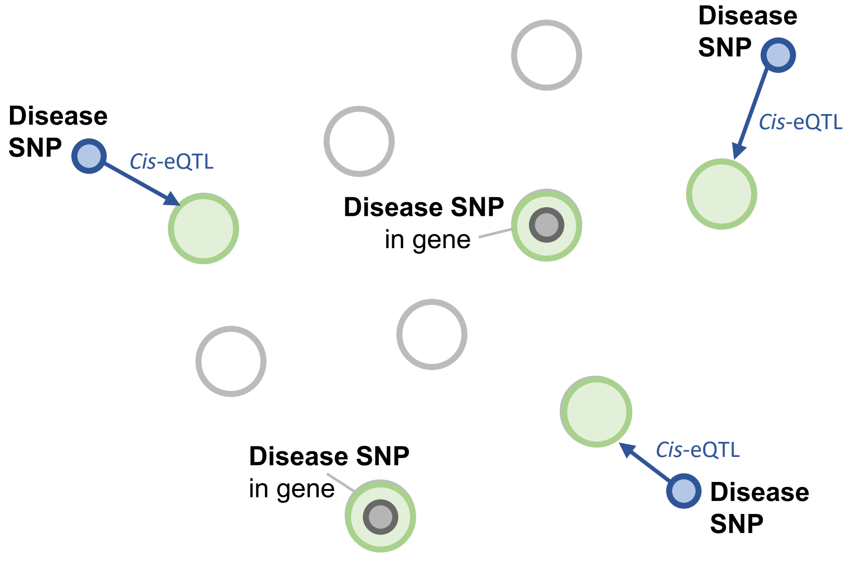
Disease

>275,000 known

>200 diseases

Interpretation of genetic risk variants

From GWAS SNP-to-gene link

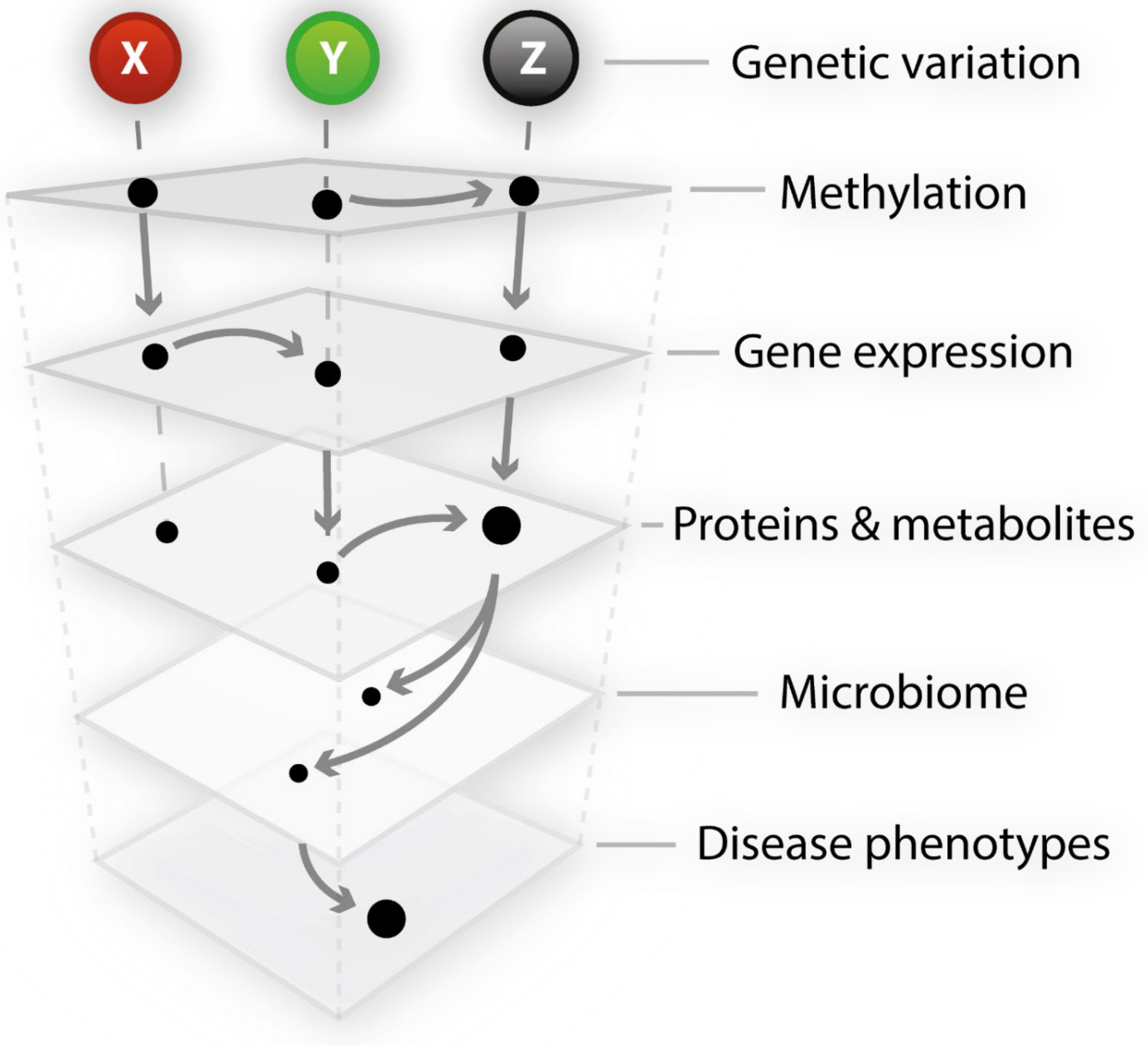


◉ = coding GWAS SNP

● = non-coding GWAS SNP

○ = Gene

🟢 = GWAS-associated gene

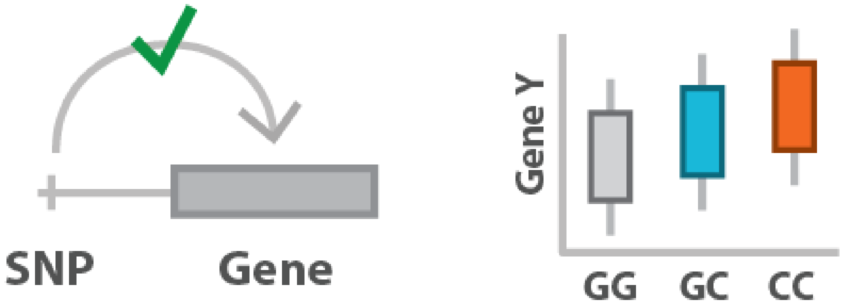


eQTLGen Consortium: a blood bulk pipeline and eQTL mapping resource (phase 1)

Phase 1

eQTLGen Consortium
eqtlgen.org

cis-eQTL:



(distance < 1Mb)

31,684 blood samples

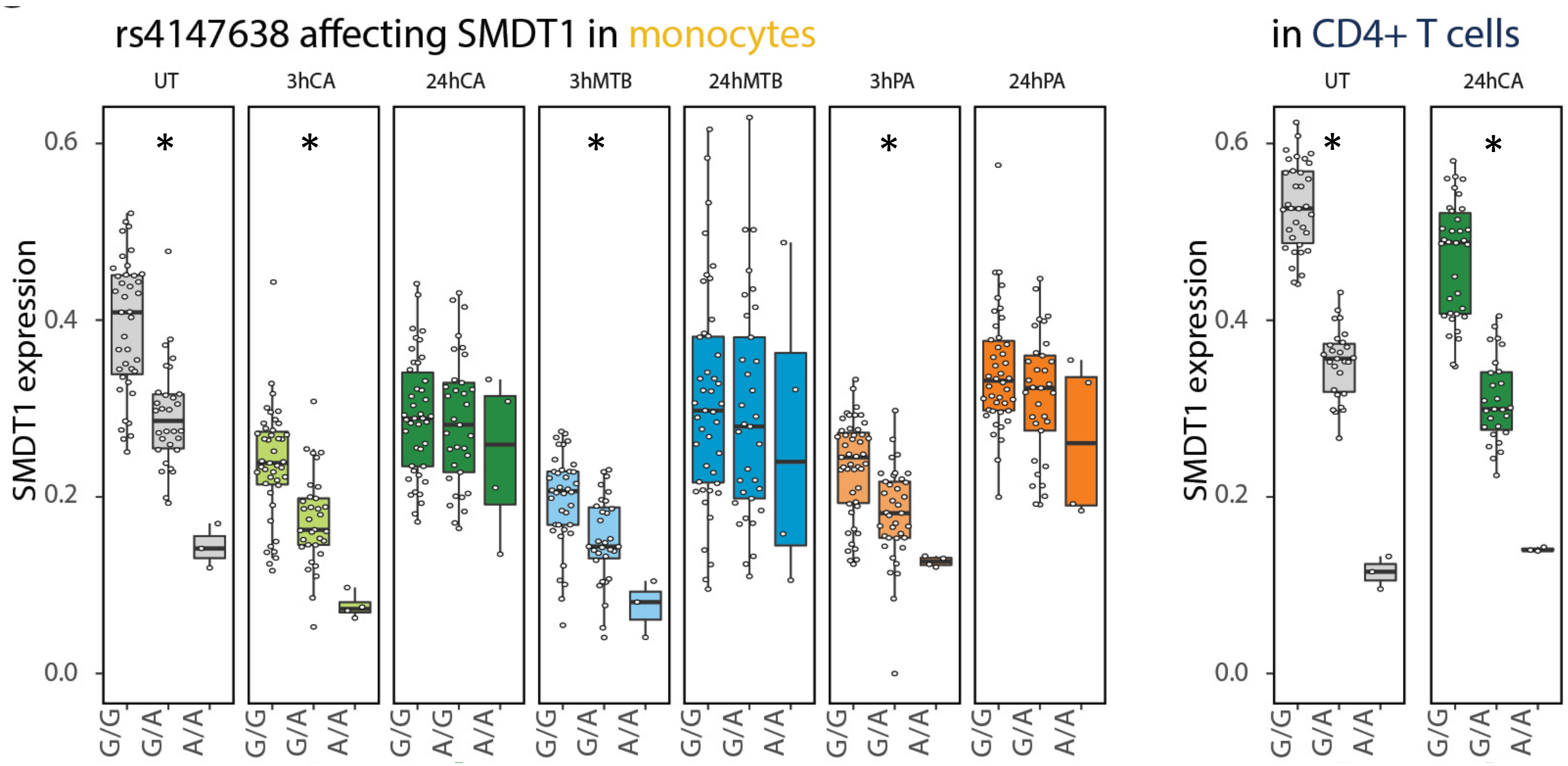
19,942 genes studied → 16,987 *cis*-eQTL genes = *cis*-eGenes

11 million SNPs (MAF ≥ 1%)

Challenge 1: eQTLs are cell-type- and context-dependent

Võsa and Claringbould *et al.*
Nature Genetics, 2021

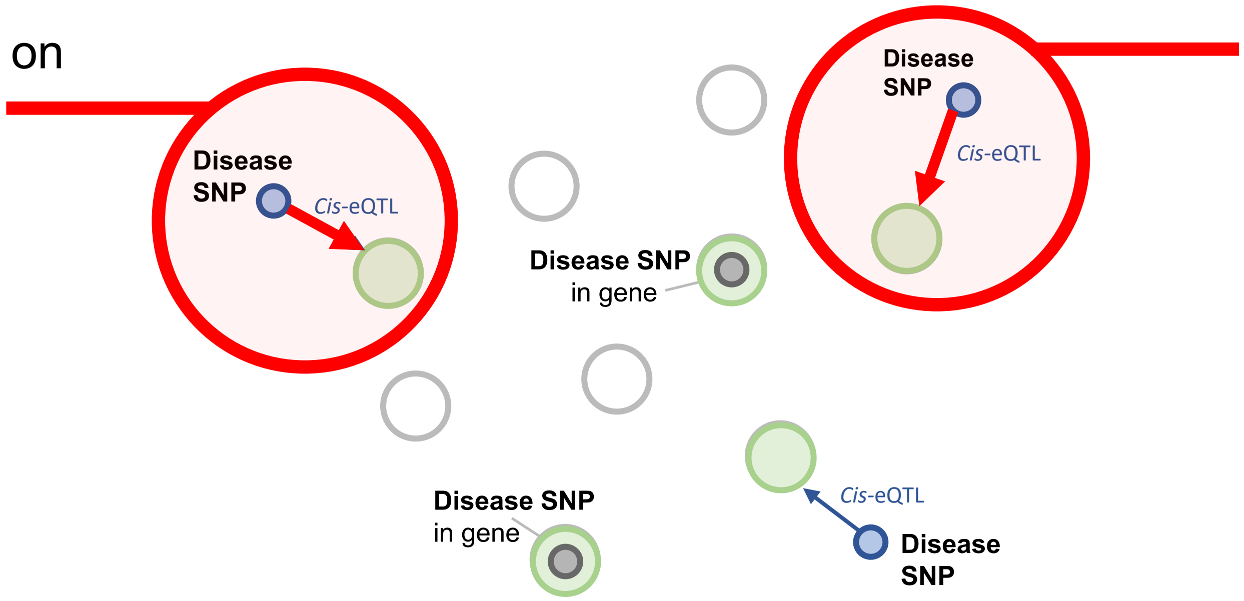
eQTLs are cell-type- and context-dependent



Oelen *et al.*
Nature Communications, 2022

Single-cell data provides the resolution needed to pinpoint the cell type and context at which eQTL effects take place

From GWAS SNP-to-gene link


Interaction dependent on **stimulation-context**

Interaction specific for a **cell type**

● = coding GWAS SNP

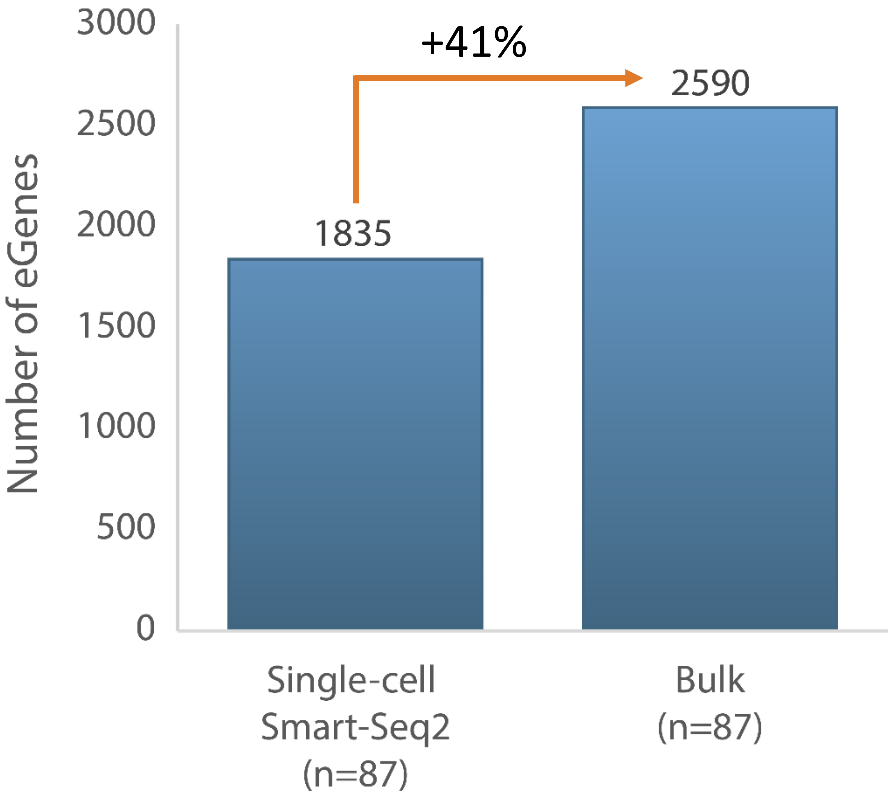
🔵 = non-coding GWAS SNP

○ = Gene

🟢 = GWAS-associated gene

Bulk-based datasets have larger eQTL discovery power than single-cell

Number of eGenes in matched single-cell and bulk iPSC datasets

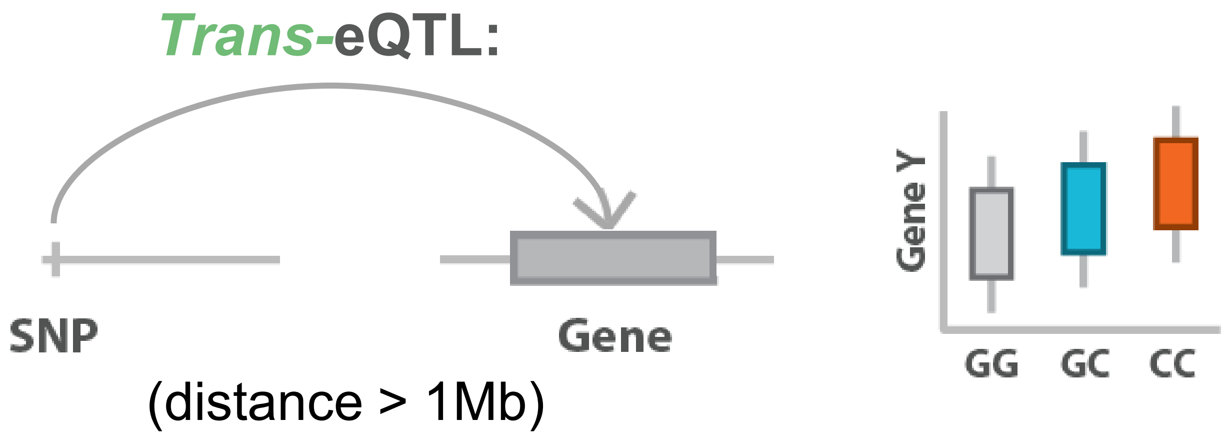


Cuomo *et al.*
Genome Biology, 2021

eQTLGen Consortium: a blood bulk pipeline and eQTL mapping resource (phase 2)

Phase 2

eQTLGen Consortium
eqtlgen.org


Trans-eQTL:

43,301 blood samples

19,942 genes →

11 million SNPs (MAF ≥ 1%)

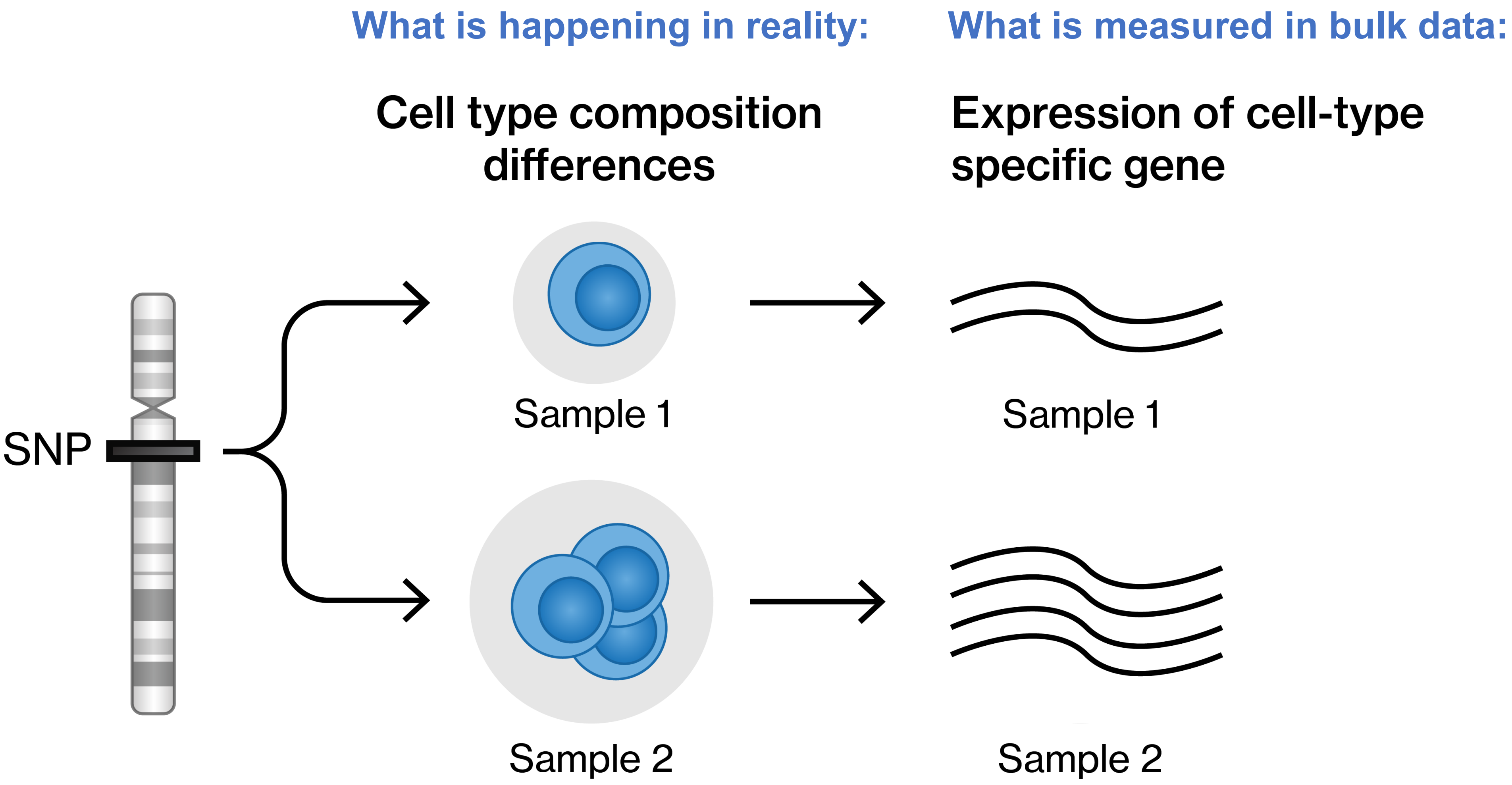
Genome-wide *trans*-eQTL mapping

58,956 *trans*-eQTLs (74.6% of the genes)

Challenge 2: difficult to distinguish true regulatory effects from cell type composition effects

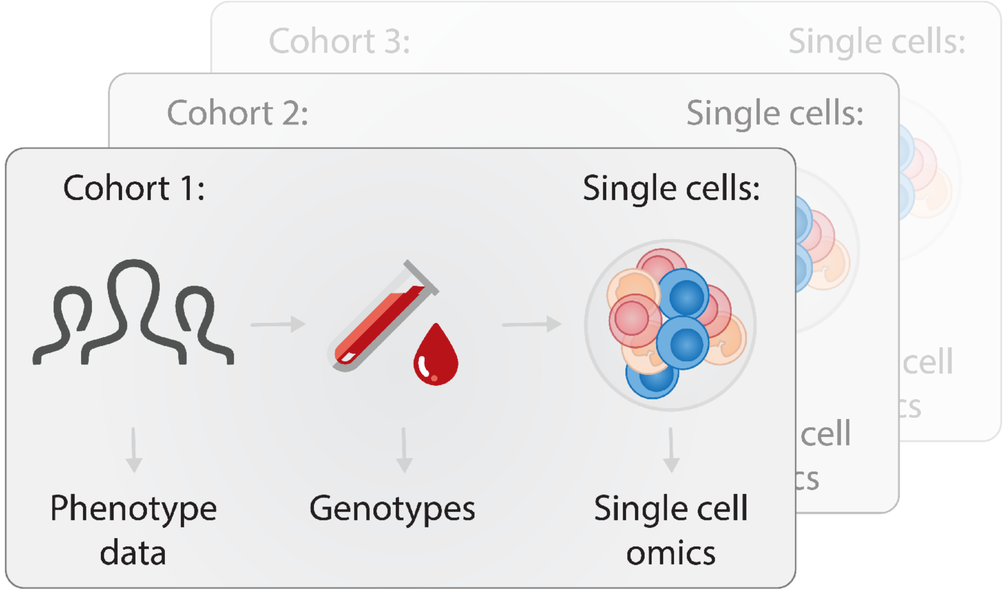
Urmo Võsa – University of Tartu
Robert Warmerdam – UMCG

Genetically-driven cell type composition effects may present themselves as false *trans*-eQTLs



sc-eQTLGen consortium: single-cell eQTL meta-analysis to tackle these challenges

Single-cell eQTL
Consortium



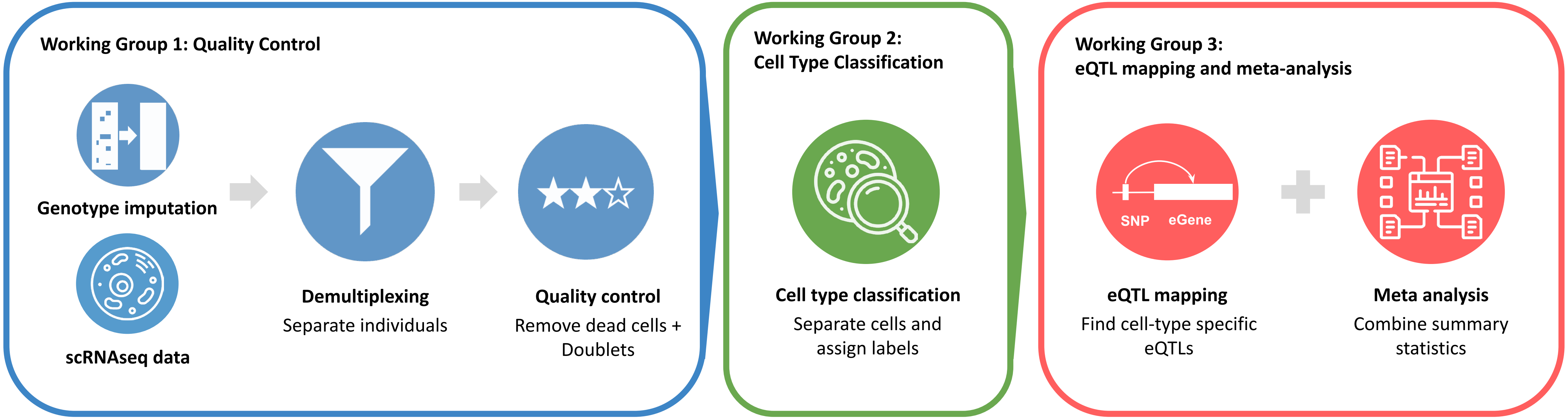
Phase 1:
14 Cohorts
2,032 Donors
6 Major PBMC cell types

Van der Wijst *et al.*
eLife, 2020

Overview sc-eQTLGen freeze 1 datasets

DS Name / Contributor	Individuals	Cells per Ind.	Technology
OneK1K	1,017	1,200	3’-10X
Burkina Malaria	178	500	3’-10X
SLE	170	4,935	3’-10X
Franke multiome	118	2,500	3’-10X
Cytoimmgen	117	1,000	3’-10X
UMCGv2	98	1,000	3’-10X
OASIS	91	8,374	5’-10X
CSF	73	1,200	3’-10X
UMCGv3	47	1,000	3’-10X
ARMS	45	318	SS2
Wijst	40	500	3’-10X
300BCG	38	1,298	3’-10X
Total	**2,032**	**3.7 million**	

Single-cell eQTLGen pipeline



Working Group 1: Quality Control

Genotype imputation

scRNAseq data

Demultiplexing
Separate individuals

Quality control
Remove dead cells + Doublets

Working Group 2:
Cell Type Classification

Cell type classification
Separate cells and assign labels

Working Group 3:
eQTL mapping and meta-analysis

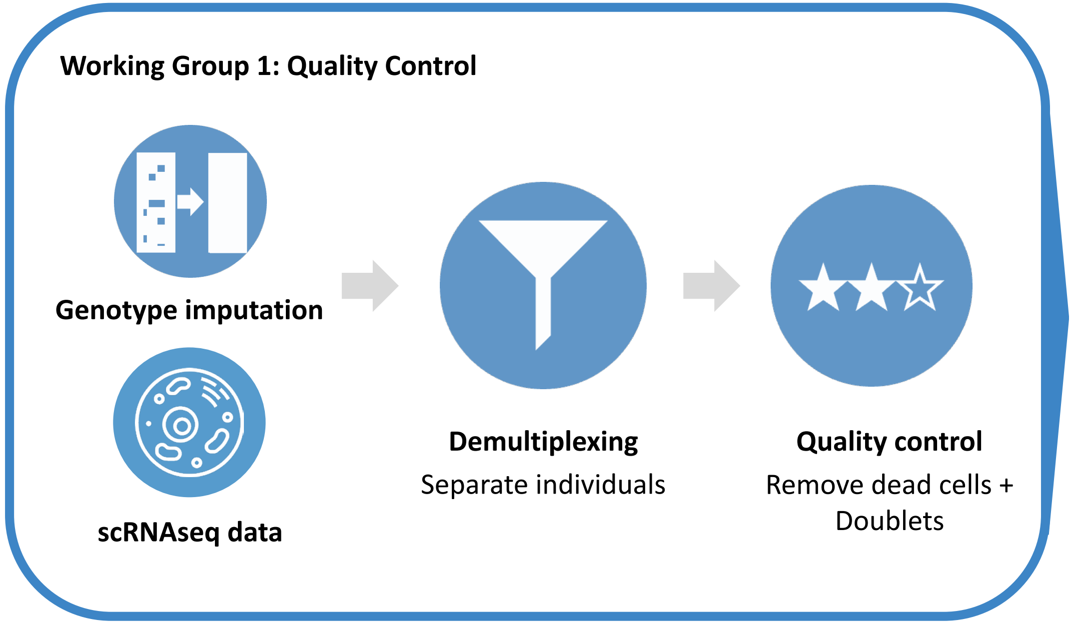
eQTL mapping
Find cell-type specific eQTLs

Meta analysis
Combine summary statistics

Ease of use:
Standardised harmonization
Pipelines built in snakemake
Software provided in singularity images

Van der Wijst *et al.*
eLife, 2020

Workgroup 1: Preprocessing and quality control



Working Group 1: Quality Control

Genotype imputation

scRNAseq data

Demultiplexing
Separate individuals

Quality control
Remove dead cells + Doublets

Marta Melé

Martin Hemberg

Workgroup 1: Preprocessing and quality control

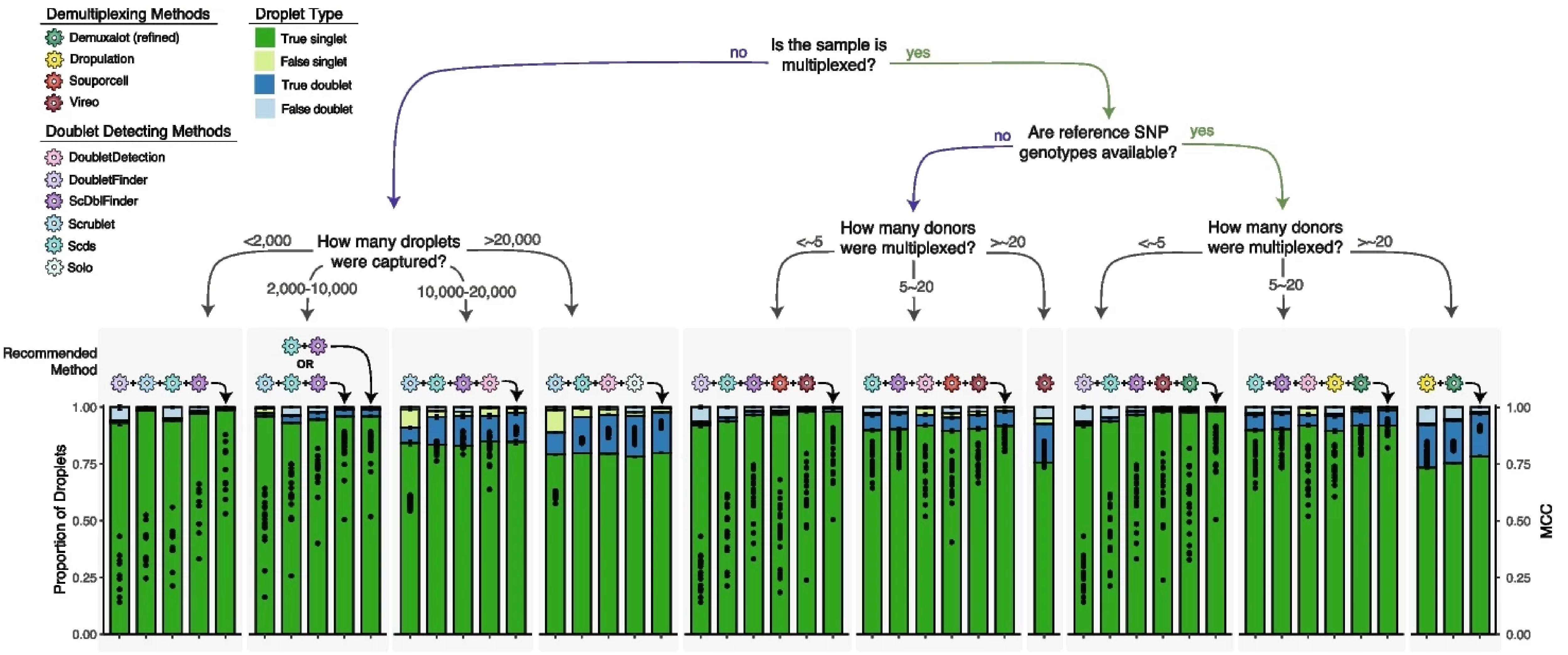
1. Impute SNP genotypes
   Reference: 1000 Genomes high coverage build hg38
   Minimac imputation software

2. Demultiplex and remove doublets
   Demuxafy pipeline

3. Calculate QC metrics for filtering threshold selection
   Number of UMI and mitochondrial RNA percentage

Neavin *et al.* Genome Biology, 2024

Workgroup 1: Preprocessing and quality control



Neavin *et al.* Genome Biology, 2024

Workgroup 2: Cell type classification

Working Group 2:
Cell Type Classification

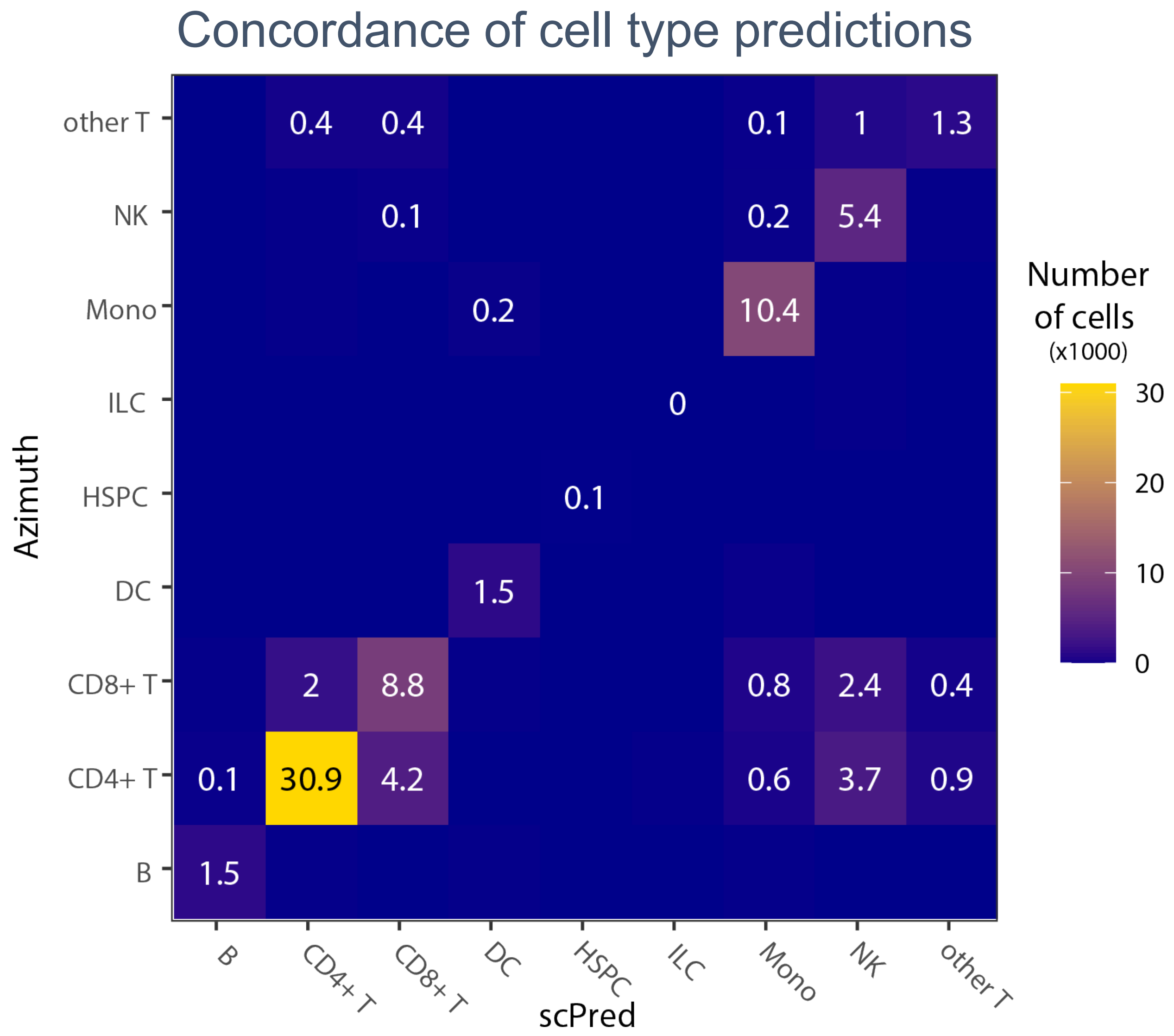
Cell type classification

Separate cells and
assign labels


Joseph Powell


Ahmed Mahfouz

Workgroup 2: Cell type classification



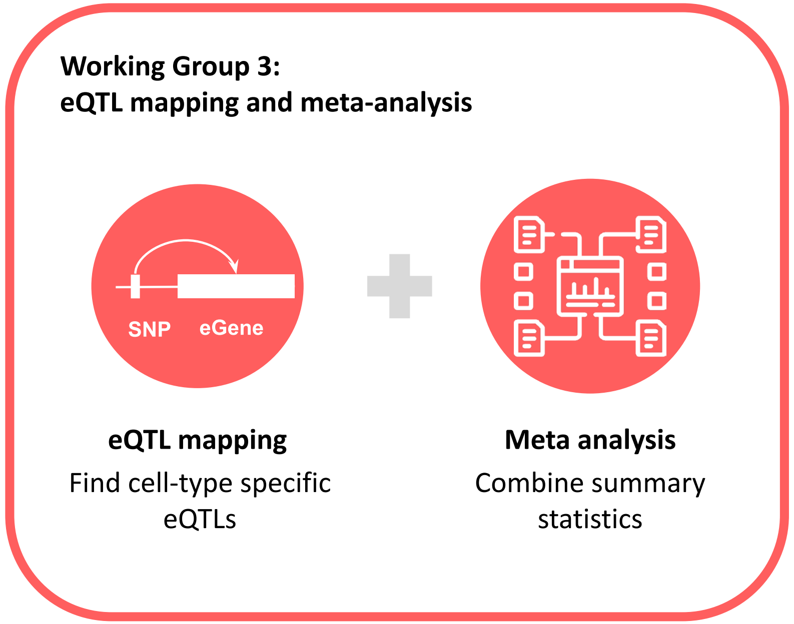
Azimuth and Hierarchical scPred approaches

Combining cell type prediction removes low quality cells

~80% of the cells remain

Hao *et al.* Cell, 2021
Alquicira-Hernandez *et al.* Genome Biology, 2019
Michielsen *et al.* Nature Communications, 2021

Workgroup 3: eQTL mapping and downstream analyses



**Working Group 3:
eQTL mapping and meta-analysis**

eQTL mapping
Find cell-type specific eQTLs

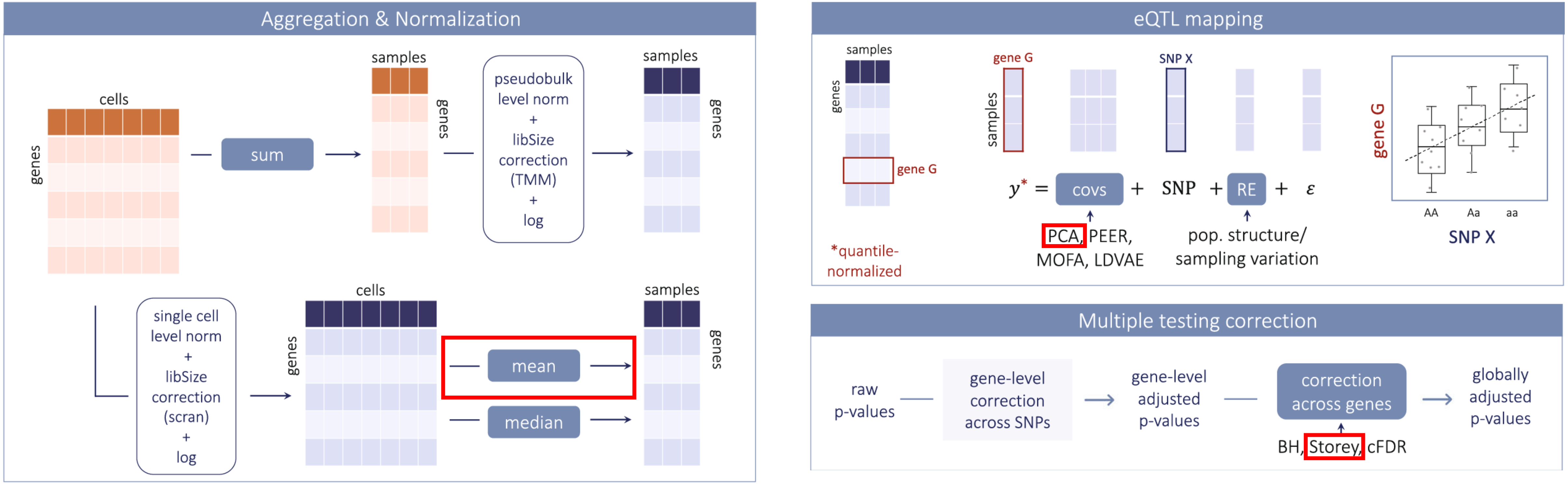
Meta analysis
Combine summary statistics

Privacy-sensitive data: ‘Bring the algorithm to the data’

Monique van der Wijst

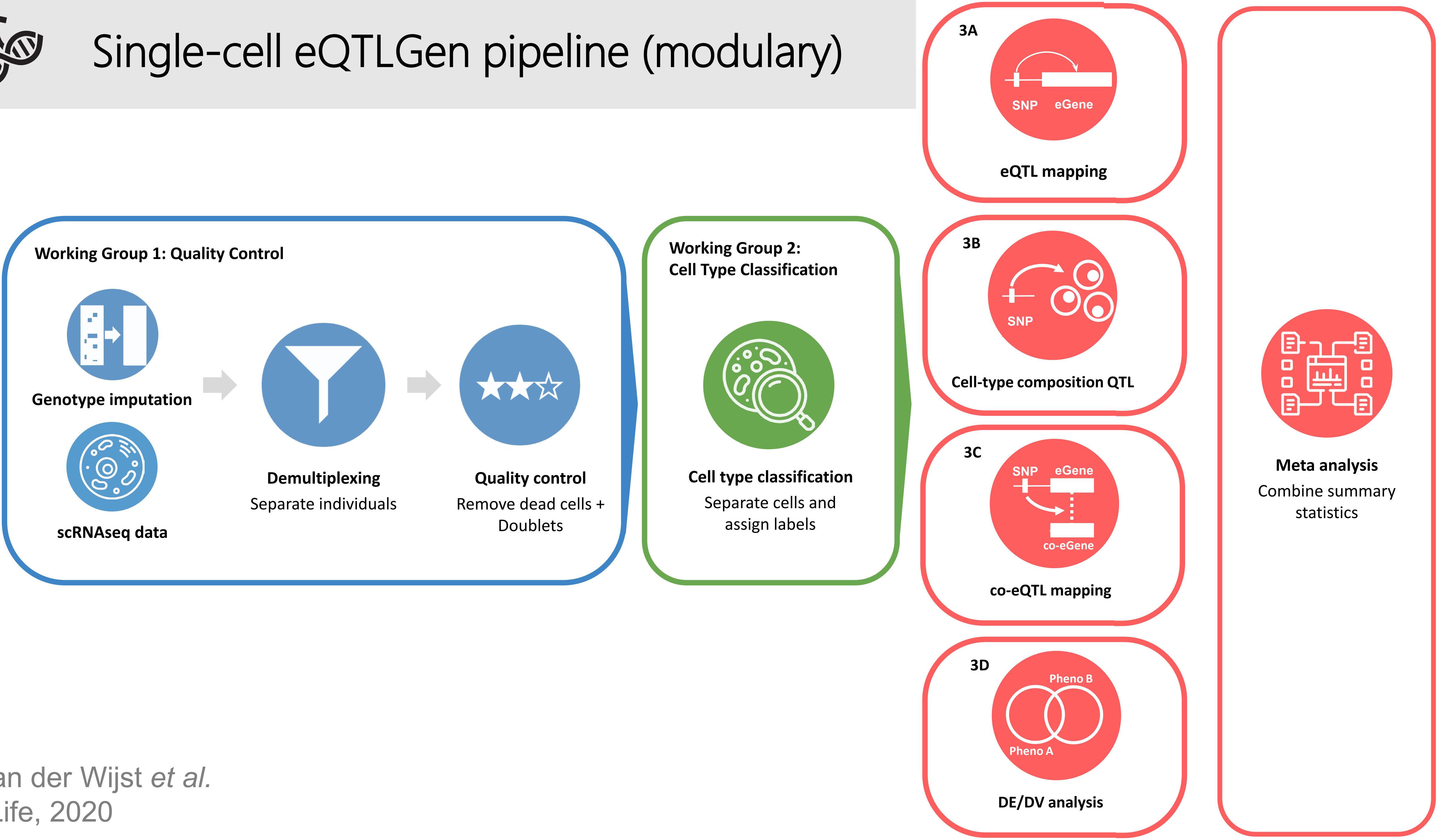
Marc Jan Bonder

Workgroup 3: eQTL mapping and downstream analyses



Cuomo *et al.* (Genome Biology, 2021)

Single-cell eQTLGen pipeline (modulary)



Working Group 1: Quality Control

Genotype imputation

scRNAseq data

Demultiplexing
Separate individuals

Quality control
Remove dead cells +
Doublets

Working Group 2:
Cell Type Classification

Cell type classification
Separate cells and
assign labels

3A
eQTL mapping

3B
Cell-type composition QTL

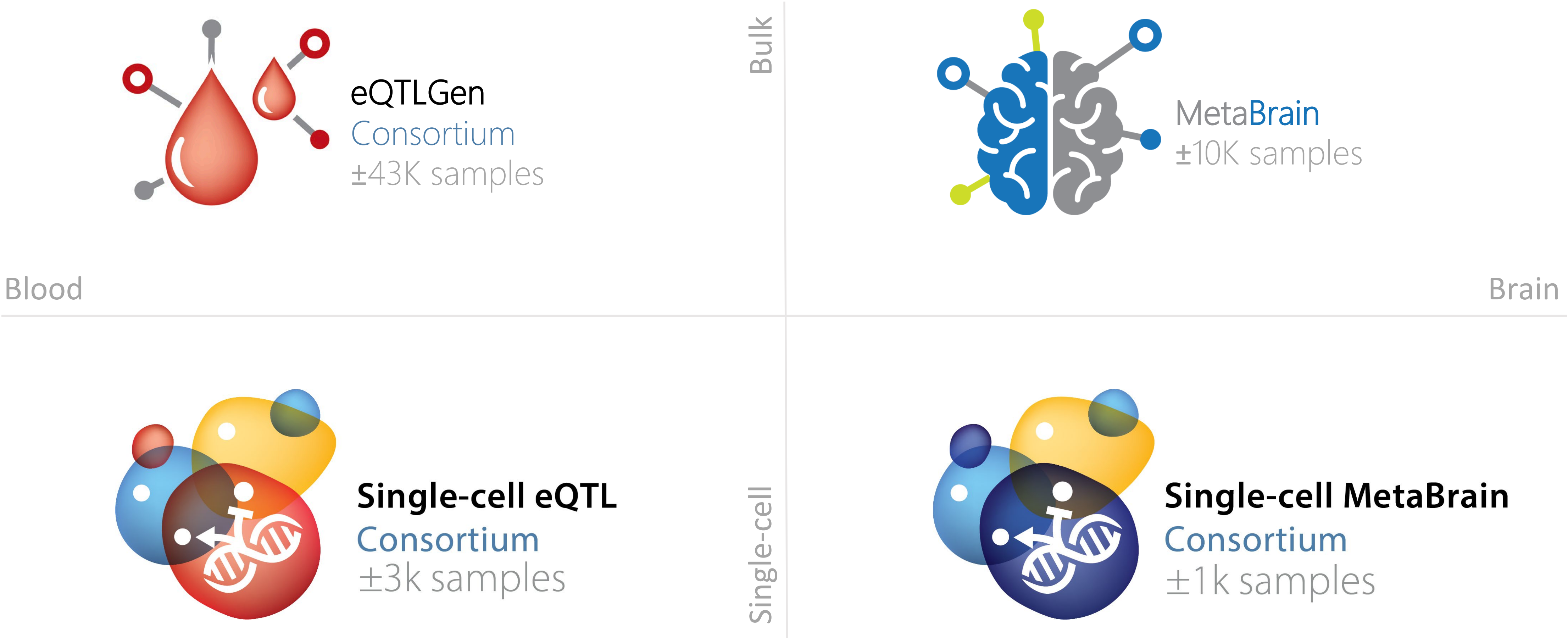
3C
co-eQTL mapping

3D
DE/DV analysis

Meta analysis
Combine summary
statistics

Van der Wijst *et al.*
eLife, 2020

Transferability of the framework to other tissues



eQTLGen
Consortium
±43K samples

MetaBrain
±10K samples

Single-cell eQTL
Consortium
±3k samples

Single-cell MetaBrain
Consortium
±1k samples

Conclusions

- Genotype-to-phenotype associations can be interpreted using QTL studies.
- Combining single-cell and bulk-based QTL mapping provides the best of both worlds: maximum eQTL discovery power and the highest cellular and context-specific resolution.
- To do this in the best possible way, consortia in which data is combined after preprocessing it in a harmonized manner are essential.

Outlook for sc-eQTLGen

- Expand to higher resolution cell types
- Go beyond only unstimulated blood cells (as in current data freeze 1): integrate all samples in one meta-analysis (healthy and disease, unstimulated and stimulated, etc) using a topic-/module-based QTL mapping approach [Popp *et al*. Cell Genomics, 2024].
- Assess the impact of donor phenotypes on eQTL effect
- Identify the upstream regulators of SNPs through co-expression QTL mapping and facilitate their interpretation using single-cell multiomics data

Will continue to add more datasets and increase sample size in following phases, consider joining!

eqtlgen.org/single-cell.html

Acknowledgements

Roy Oelen
Maryna Korshevniuk
Harm-Jan Westra
Lude Franke

Dan Kaptijn
Robert Warmerdam
Marc-Jan Bonder

Urmo Võsa

Corrina Losert
Matthias Heinig

Single-cell eQTLGen Consortium

eQTLGen Consortium

lifelines

NWO Netherlands Organisation for Scientific Research

We are looking for PhD students and post docs!

Contact: m.g.p.van.der.wijst@umcg.nl