FIRST INSIGHTS INTO THE GENETIC CONTROL OF ABIOTIC STRESS-RELATED METABOLITES IN SUNFLOWER USING GWAS

Marco Moroldo | ASTR team

Overview of the presentation

1) background and aim of the projet
2) experimental design and analysis workflow
3) results
4) next steps

(1) Background and aim of the project

(1) Background

- oxidative damage is common to many abiotic stresses, e.g. drought, cold and salinity
- antioxidant molecules play an important role in preventing oxidative damage
- most of the secondary metabolites induced by abiotic stress are antioxidants (Nakabayashi and Saito 2015)
- studying the genetics of secondary metabolites can contribute to the understanding of the tolerance to abiotic stress

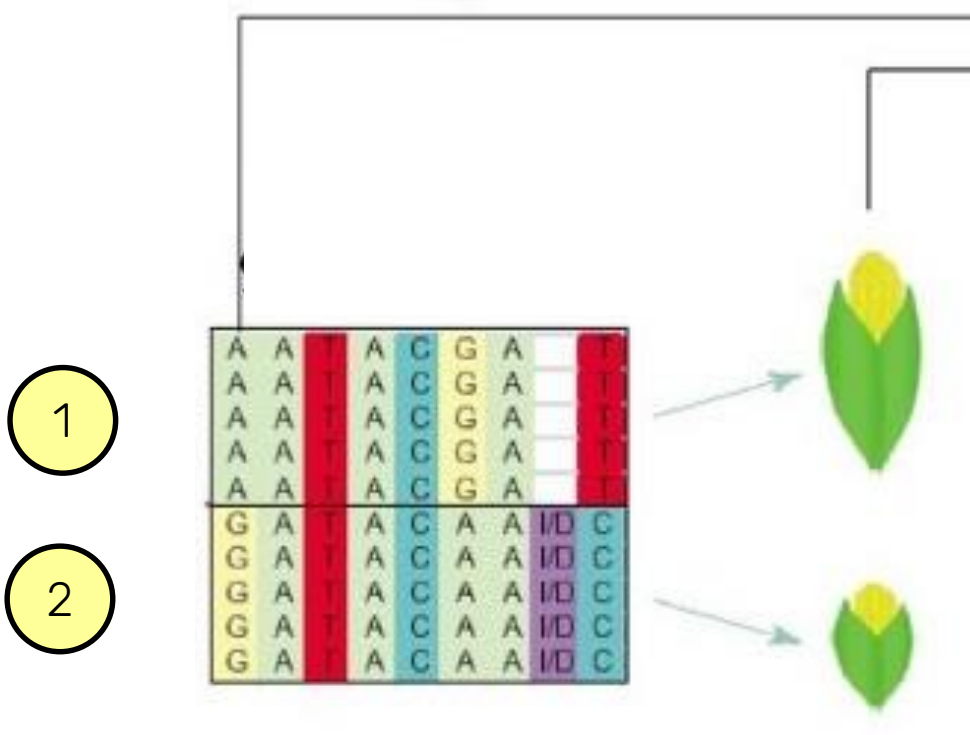
(1) Aim of the project

- to elucidate the bases of metabolic variation in response to abiotic stresses in sunflower through a genetic association approach (metabolic QTLs or mQTLs)
- to complement the information already available from transcriptomic and metabolomic studies

(1) GWAS

- GWAS = genome-wide association study
- association study = search for statistical associations among phenotypes and molecular variation in an appropriated population (for instance with high levels of diversity and recombination)
- typically based on SNPs

(1) GWAS

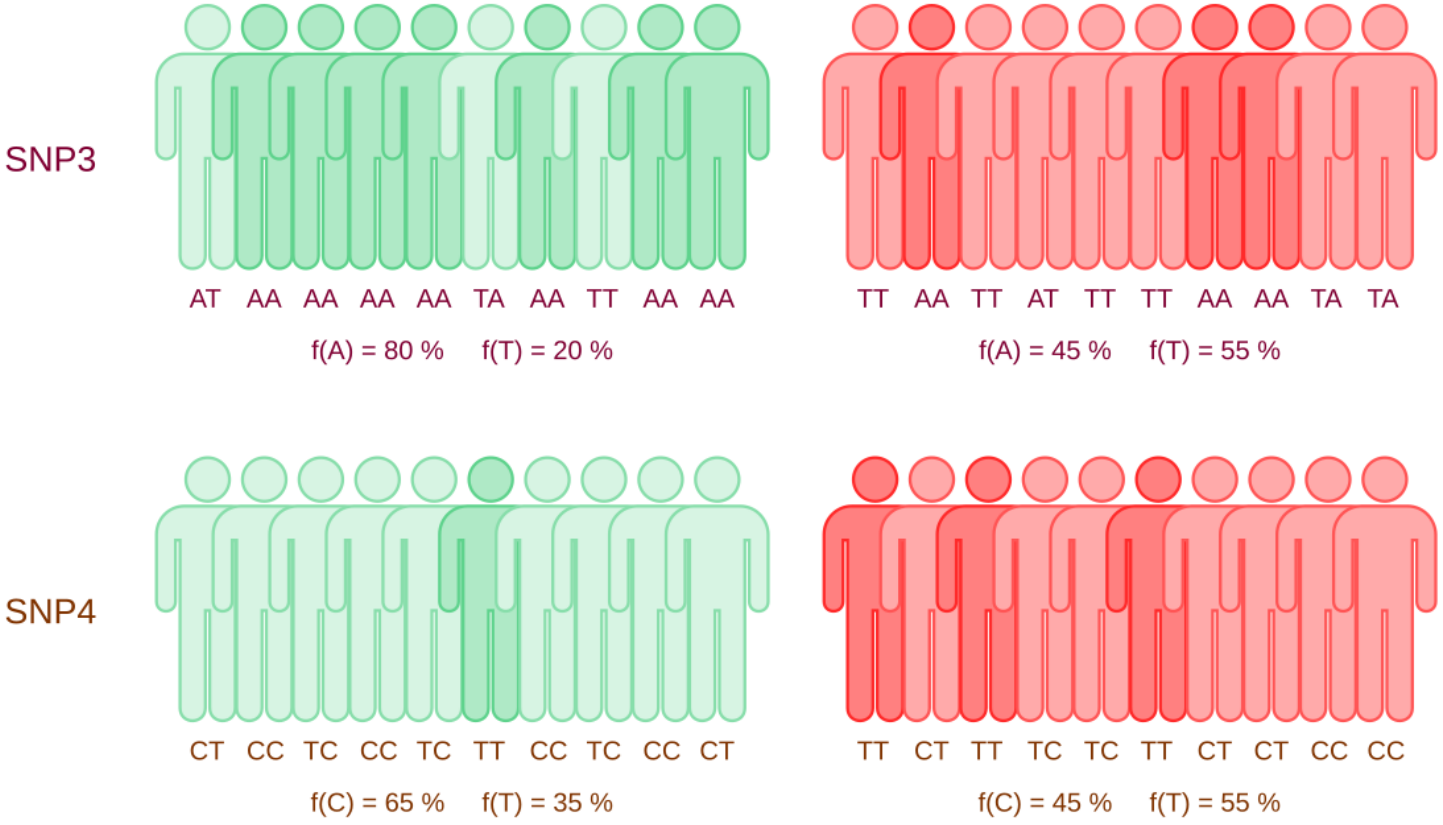


Phenotype	Genotype
1	1
1	1
1	1
1	1
1	1
0	0
0	0
0	0
0	0
0	0

(1) GWAS


SNP1
AG AA AA AA AA GA AA AA AA AA
f(A) = 18/20 x 100 = 90% f(G) = 2/20 x 100 = 10%
AA AA AA GA AA AA AA AA GA AA
f(A) = 18/20 x 100 = 90% f(G) = 2/20 x 100 = 10%
SNP2
GG GC GG GG CC CG GC CC GG CC
f(G) = 55 % f(C) = 45 %
CC GG GG CC GC GG CG GG CC CG
f(G) = 55 % f(C) = 45 %

(1) GWAS

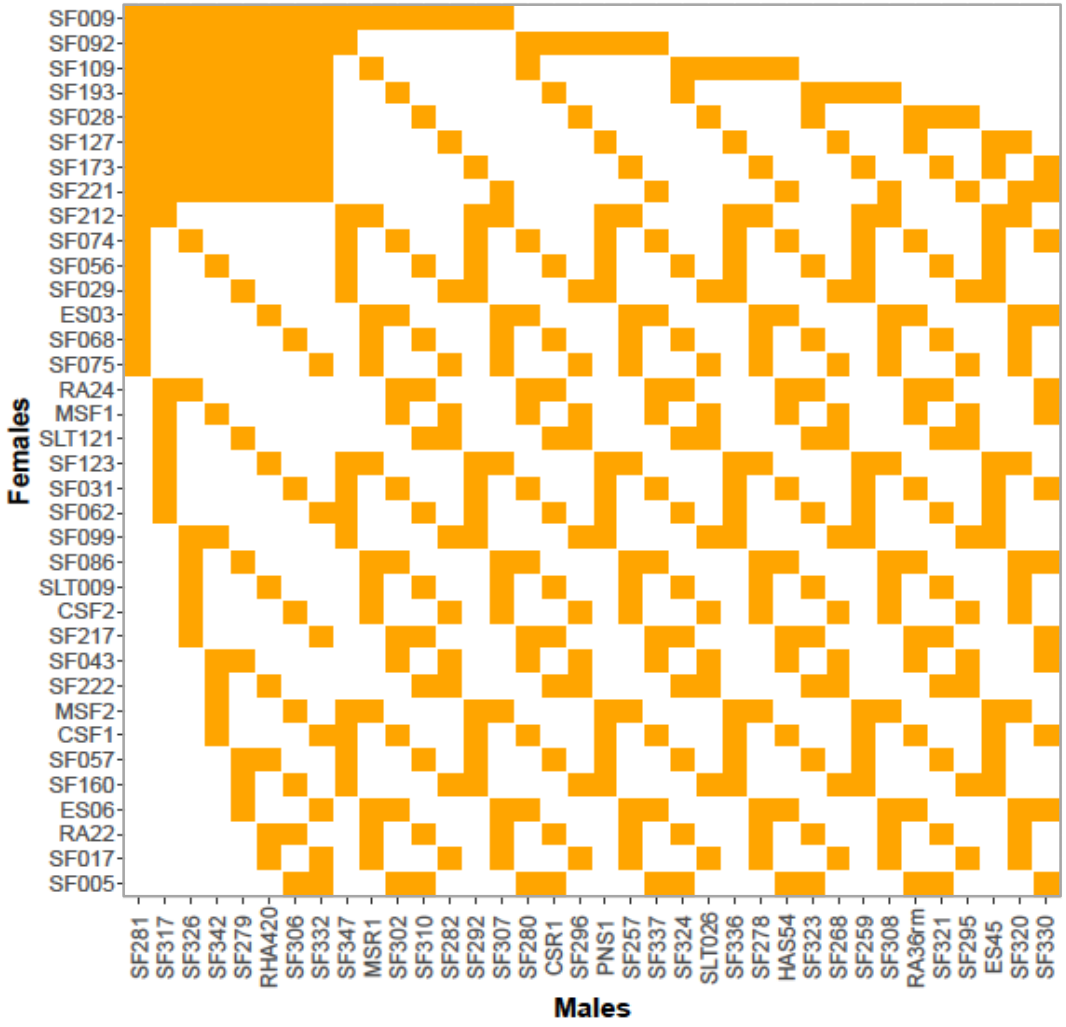

SNP3
AT AA AA AA AA TA AA TT AA AA
f(A) = 80 % f(T) = 20 %
TT AA TT AT TT TT AA AA TA TA
f(A) = 45 % f(T) = 55 %
SNP4
CT CC TC CC TC TT CC TC CC CT
f(C) = 65 % f(T) = 35 %
TT CT TT TC TC TT CT CT CC CC
f(C) = 45 % f(T) = 55 %

(2) Experimental setup

(2) Plant material

- 475 hybrid genotypes obtained by crossing 36 male genotypes and 36 female genotypes following an incomplete factorial design

(2) Plant material



(2) Plant material

- 475 hybrid genotypes obtained by crossing 36 male genotypes and 36 female genotypes following an incomplete factorial design
- good overall representation of the genetic diversity found in cultivated sunflower

(2) Plant material

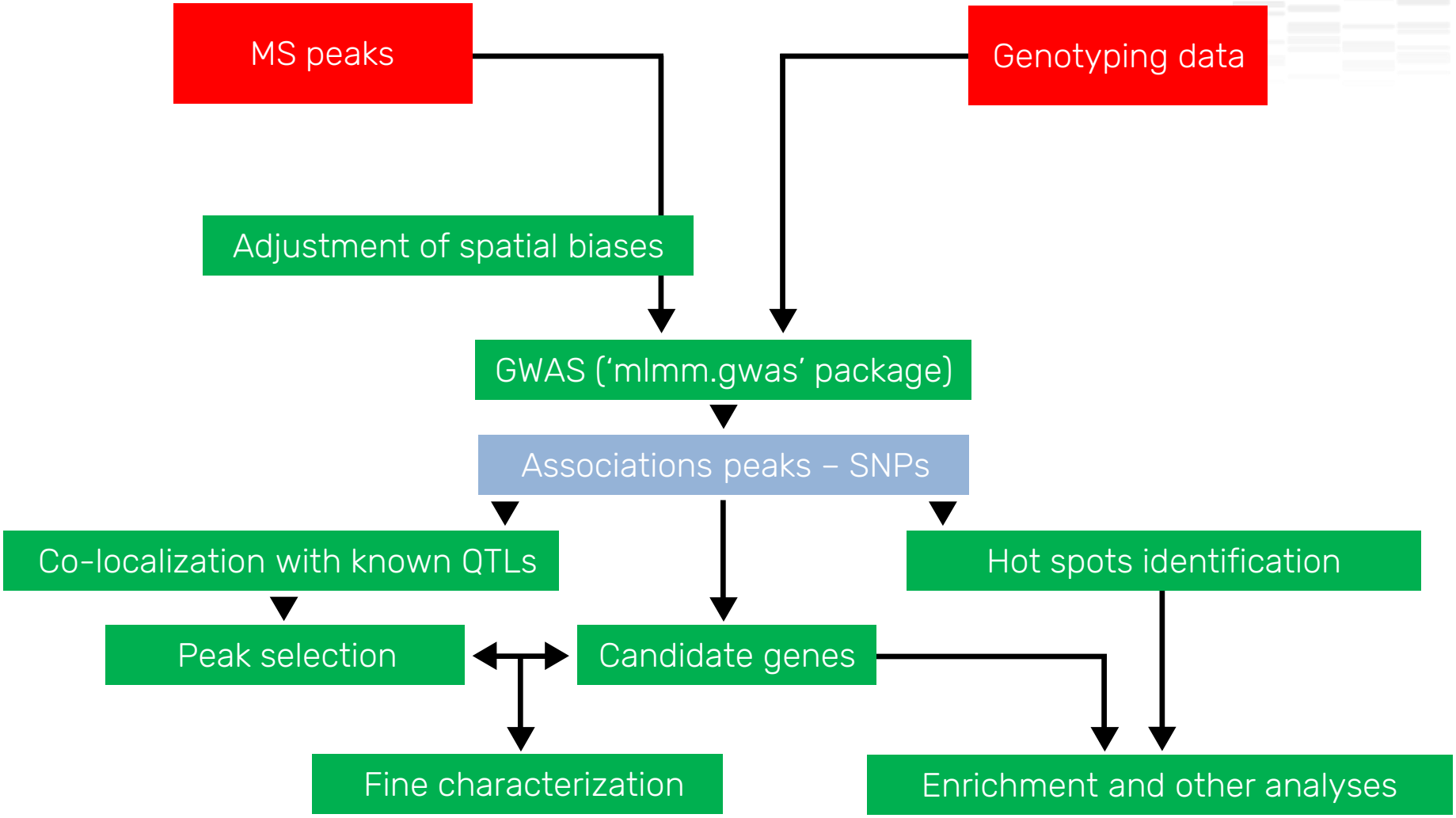
(2) Metabolome analysis

- untargeted analysis by LC-MS using an Orbitrap-MS (Thermo Fischer) after ethanol / water (80:20) extraction from *n*-4 topmost leaves
- protocol that isolates the semi-polar fraction of metabolome
- suited to target secondary metabolites
- single-step MS analysis: peaks can only be annotated *in silico*

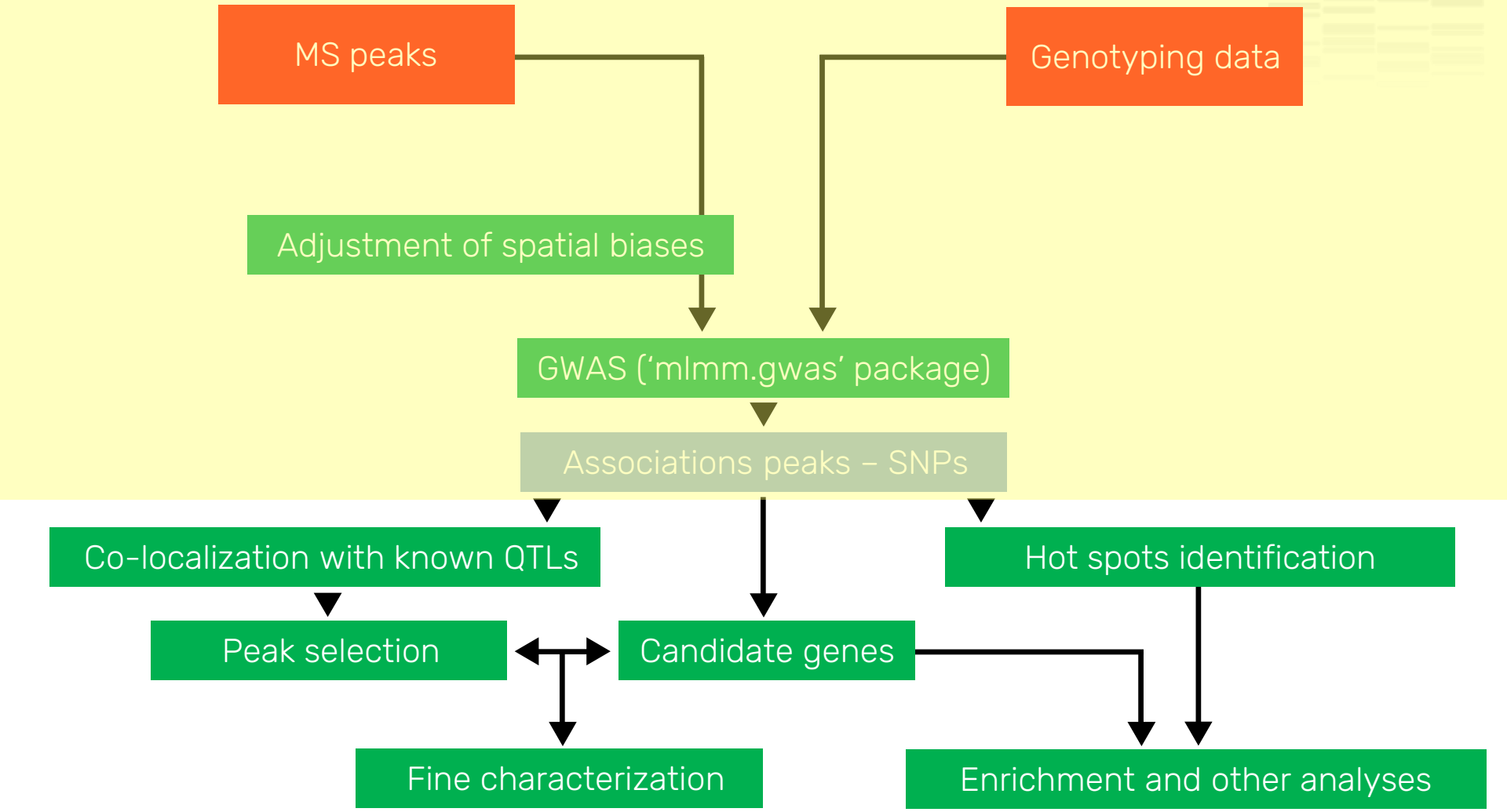
(2) Genotyping

- obtained by Illumina resequencing of the 72 parental lines (Badouin et al 2017)
- 14,127,553 SNPs initially detected
- filtering for MAF < 0.1
- filtering for SNPs in complete linkage disequilibrium: 1 SNP kept for each set of co-inherited SNPs
- 350,052 SNPs eventually retained (‘reference SNPs’)

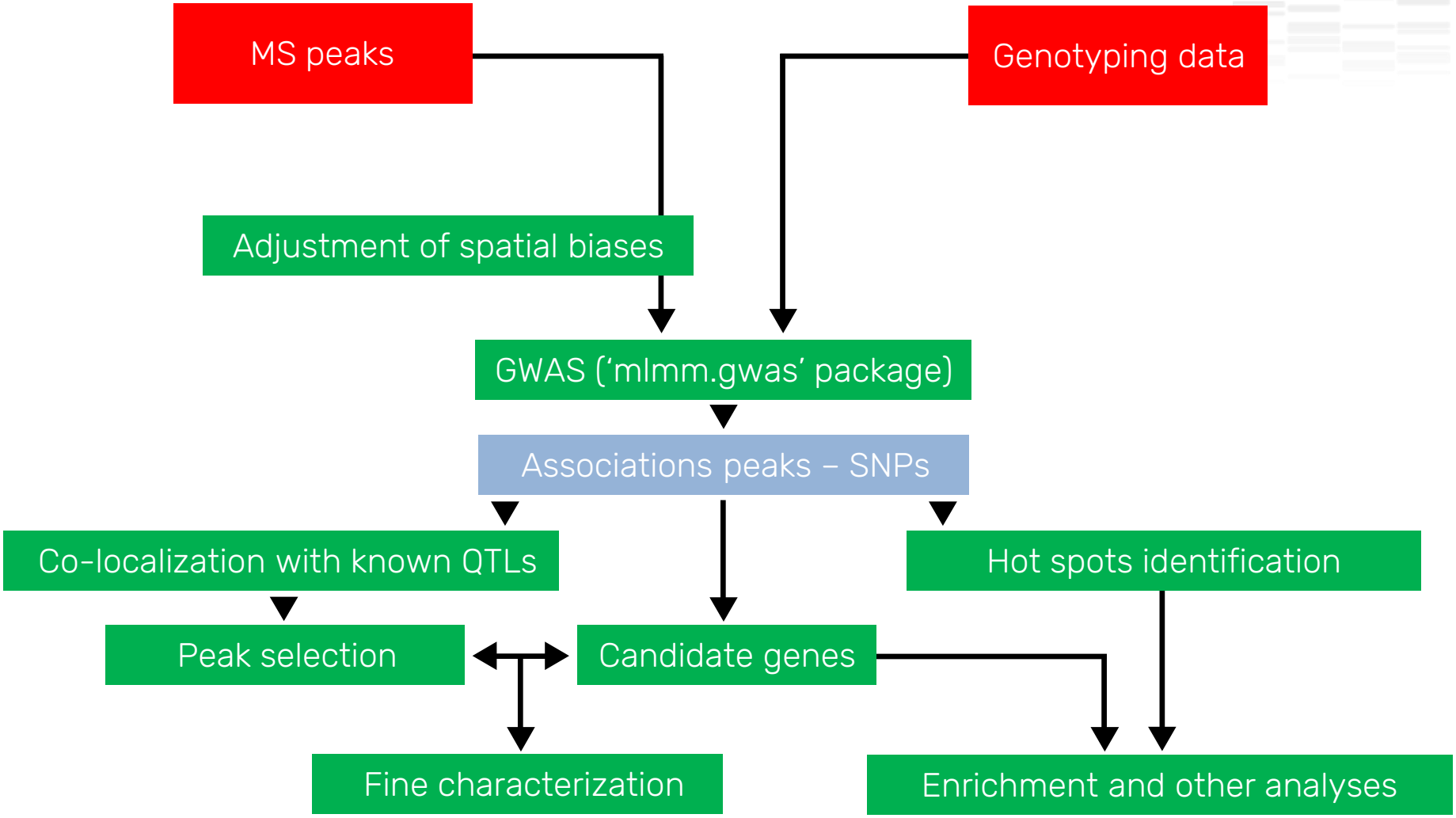
(2) Analysis workflow



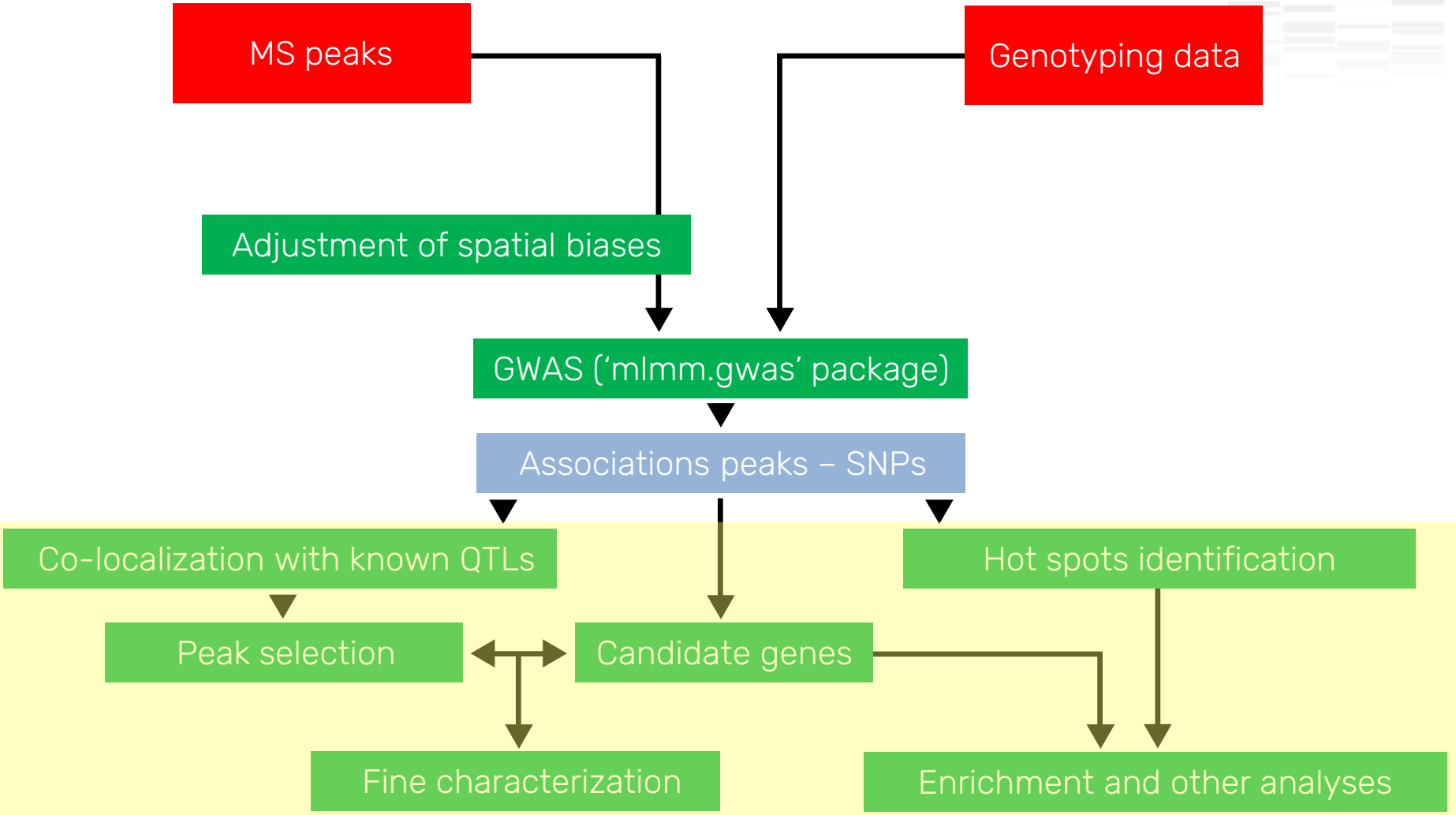
(2) Analysis workflow



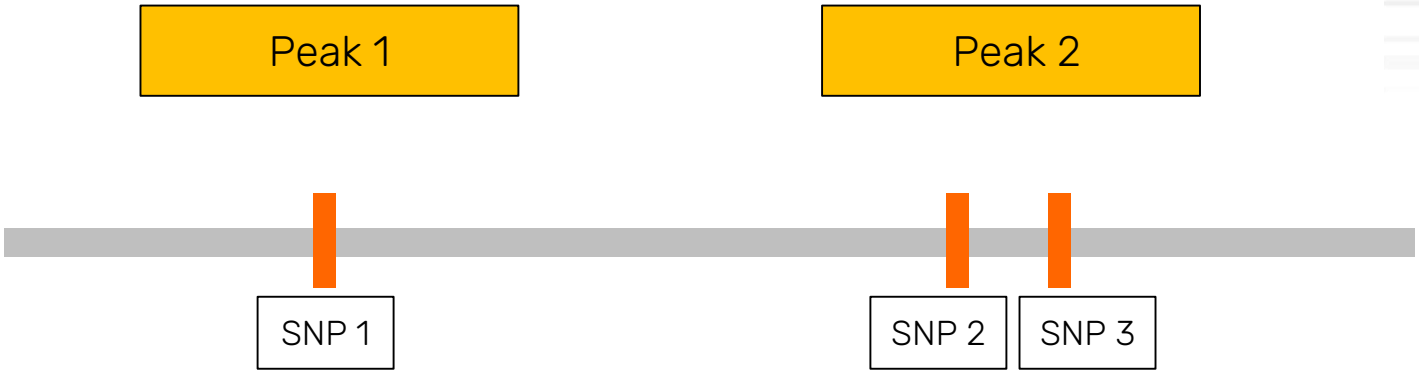
(2) Analysis workflow



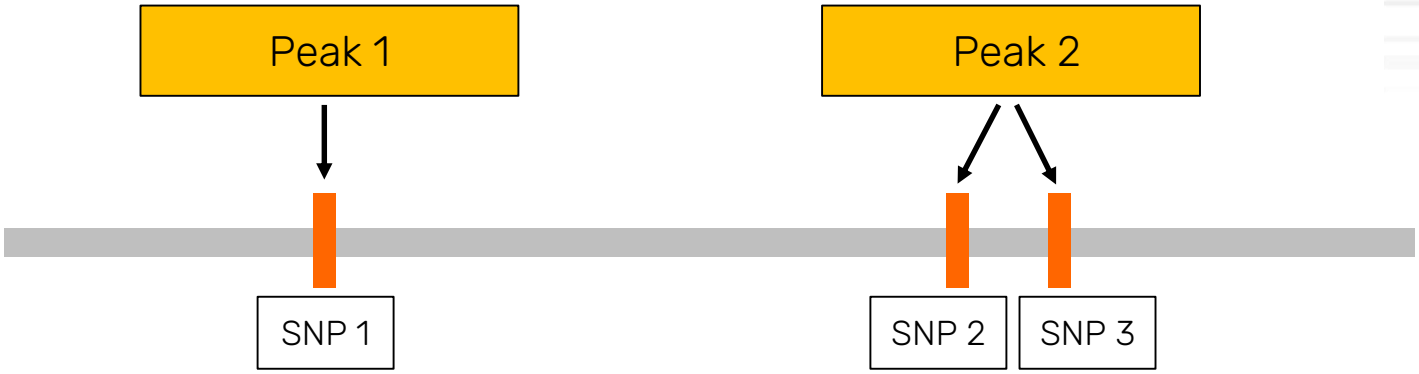
(2) Analysis workflow



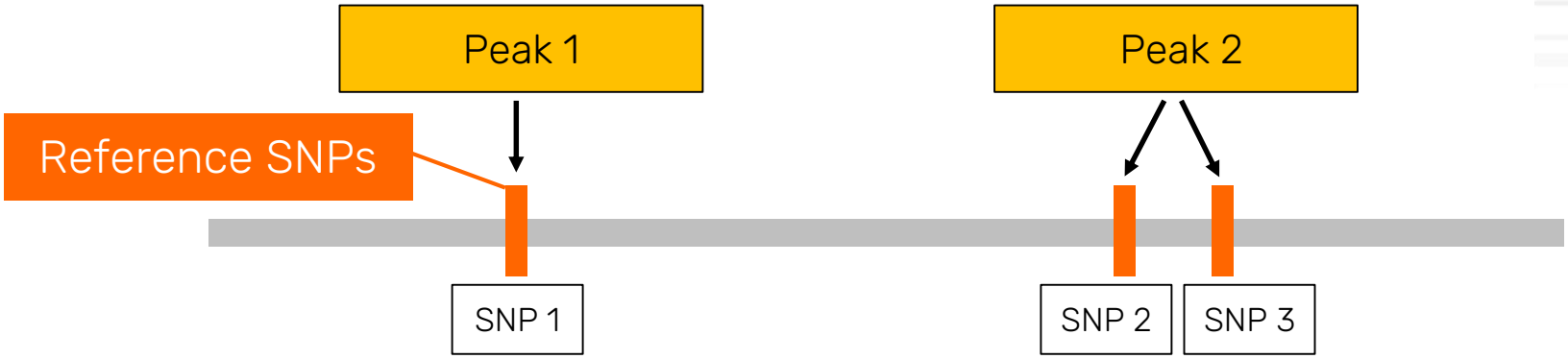
(2) GWAS



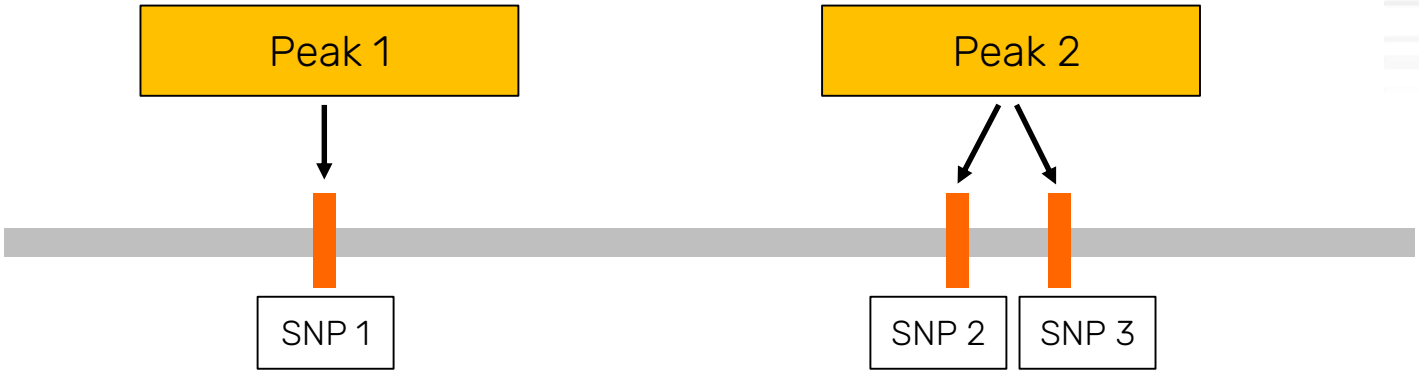
(2) GWAS



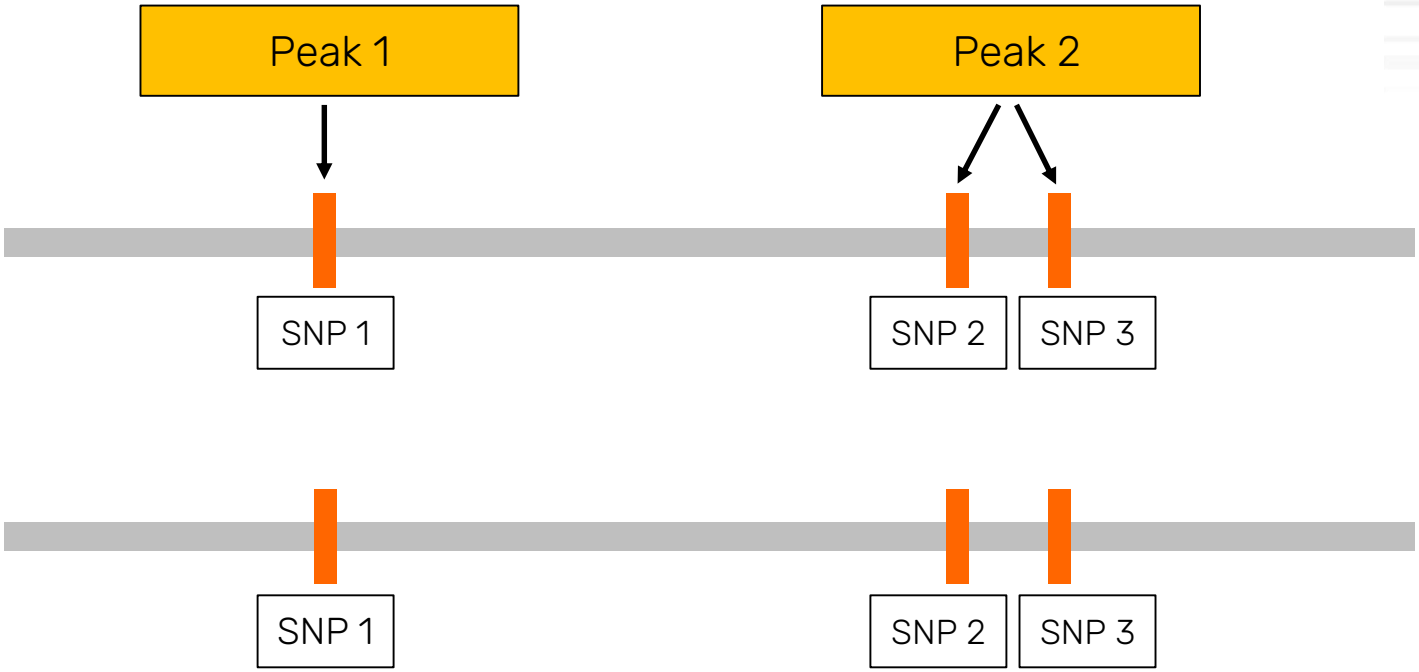
(2) GWAS



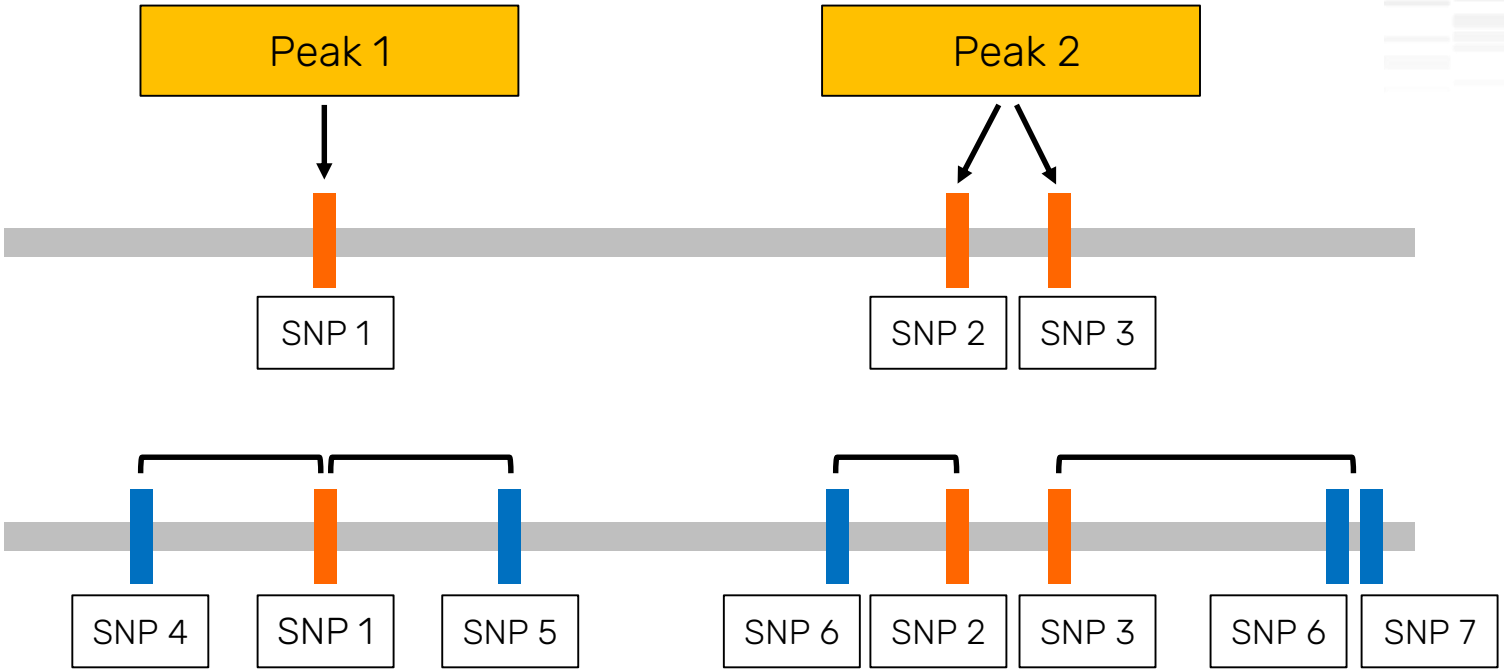
(2) GWAS



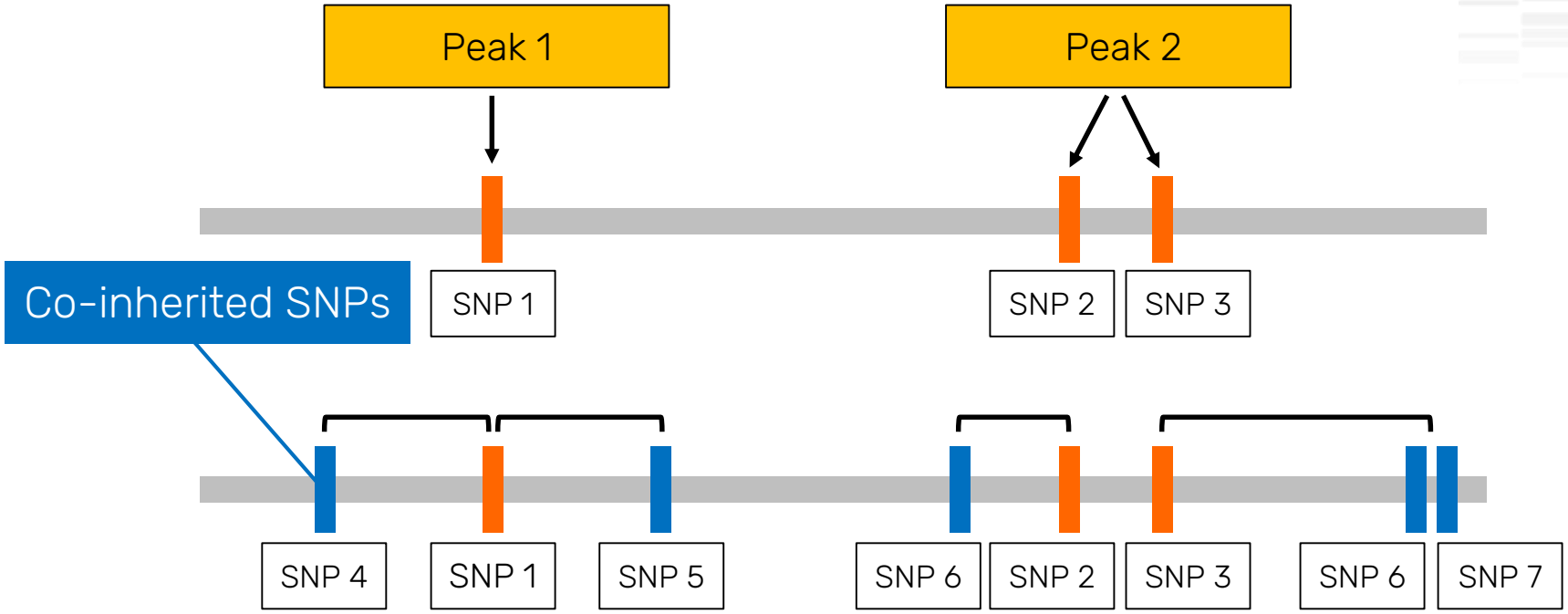
(2) GWAS



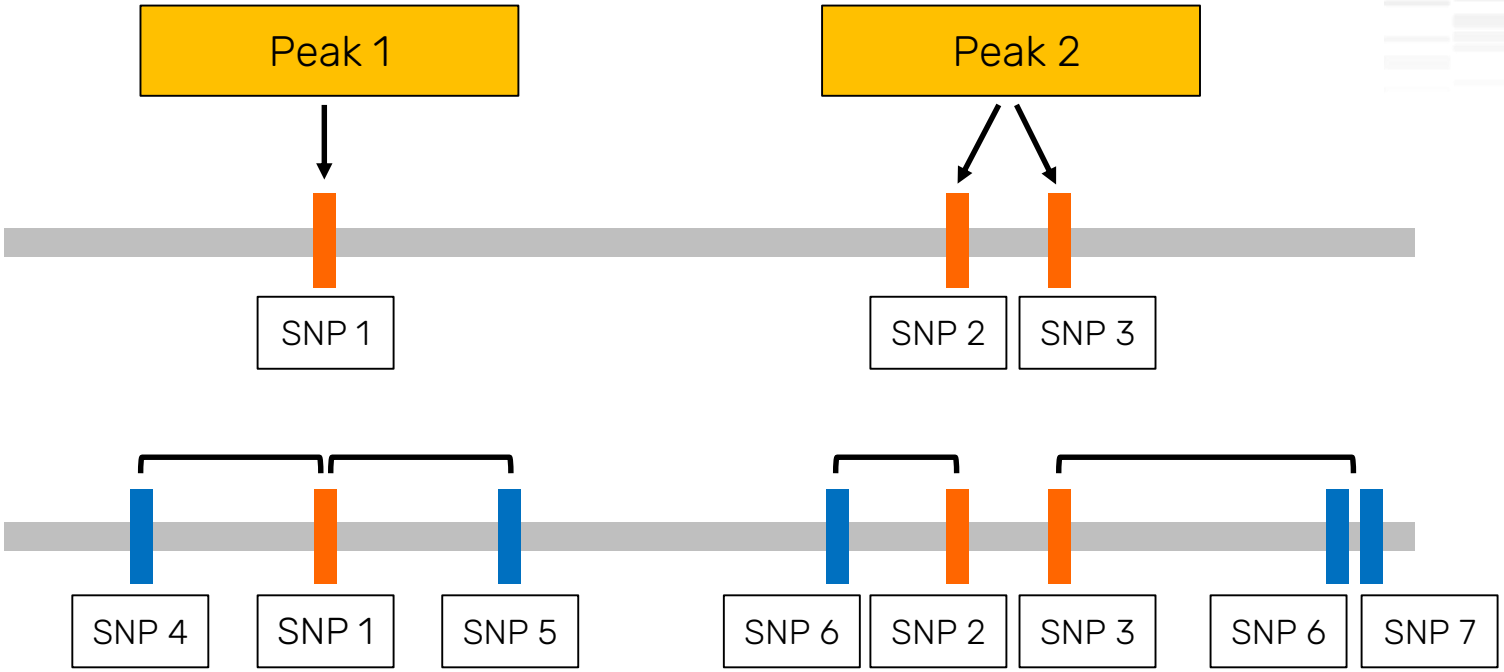
(2) GWAS


Peak 1
Peak 2
SNP 1
SNP 2
SNP 3
SNP 4
SNP 1
SNP 5
SNP 6
SNP 2
SNP 3
SNP 6
SNP 7

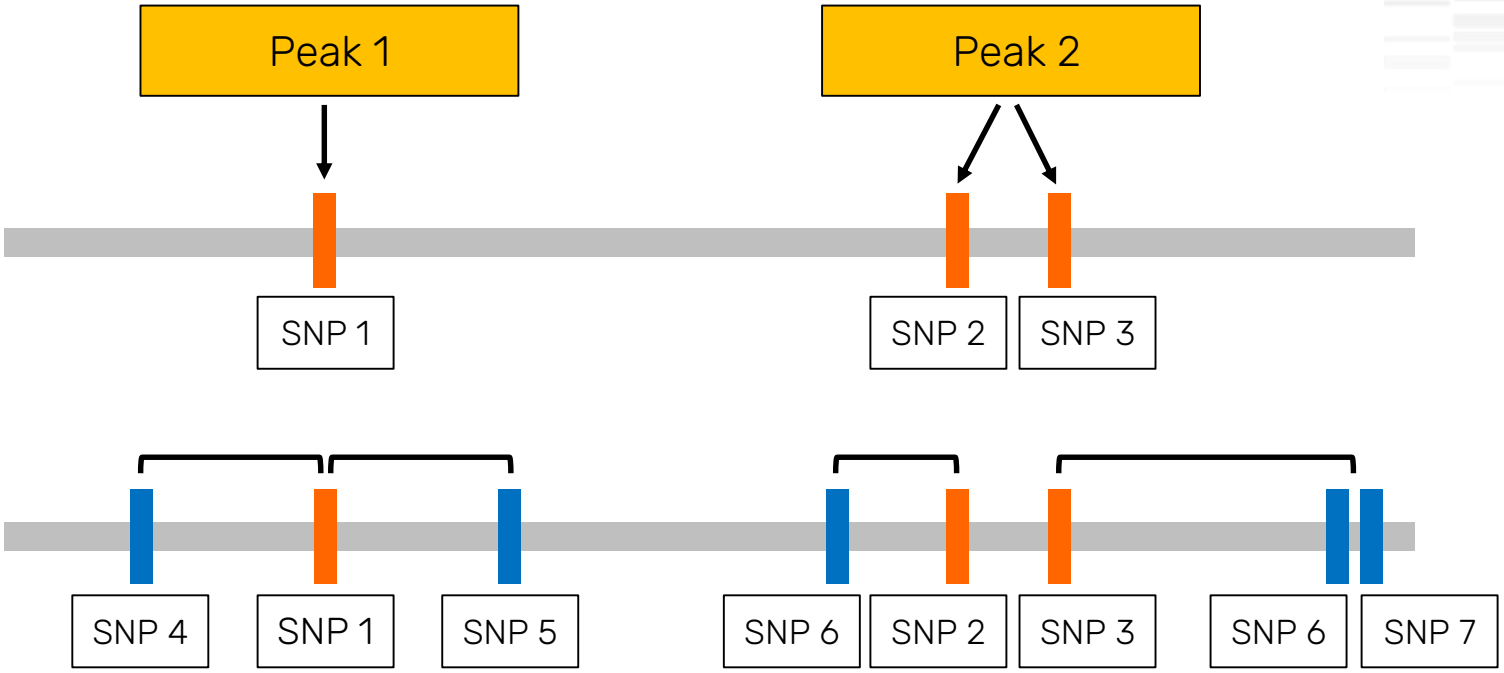
(2) GWAS


Peak 1
Peak 2
Co-inherited SNPs
SNP 1
SNP 2
SNP 3
SNP 4
SNP 1
SNP 5
SNP 6
SNP 2
SNP 3
SNP 6
SNP 7

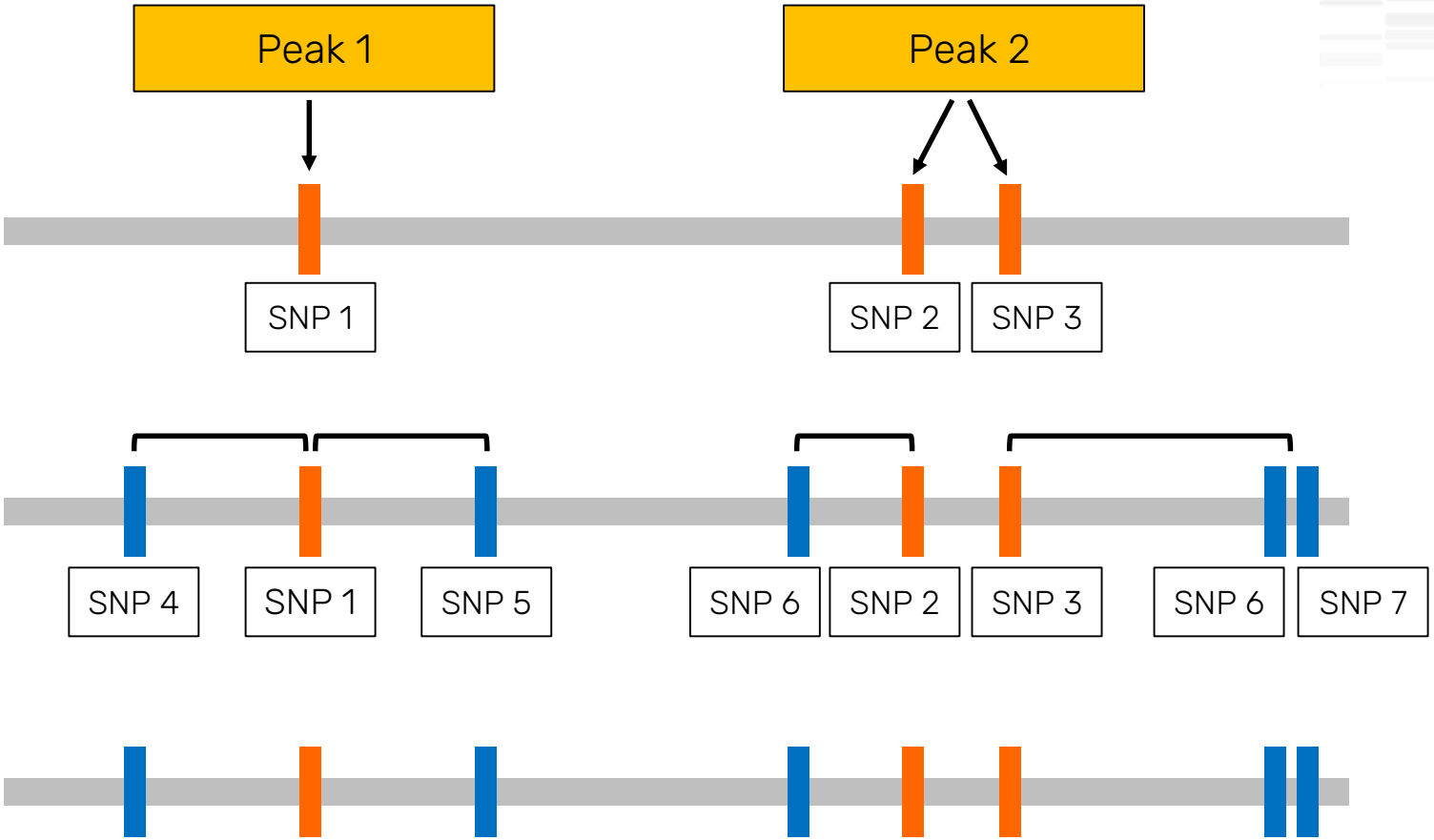
(2) GWAS


Peak 1
Peak 2
SNP 1
SNP 2
SNP 3
SNP 4
SNP 1
SNP 5
SNP 6
SNP 2
SNP 3
SNP 6
SNP 7

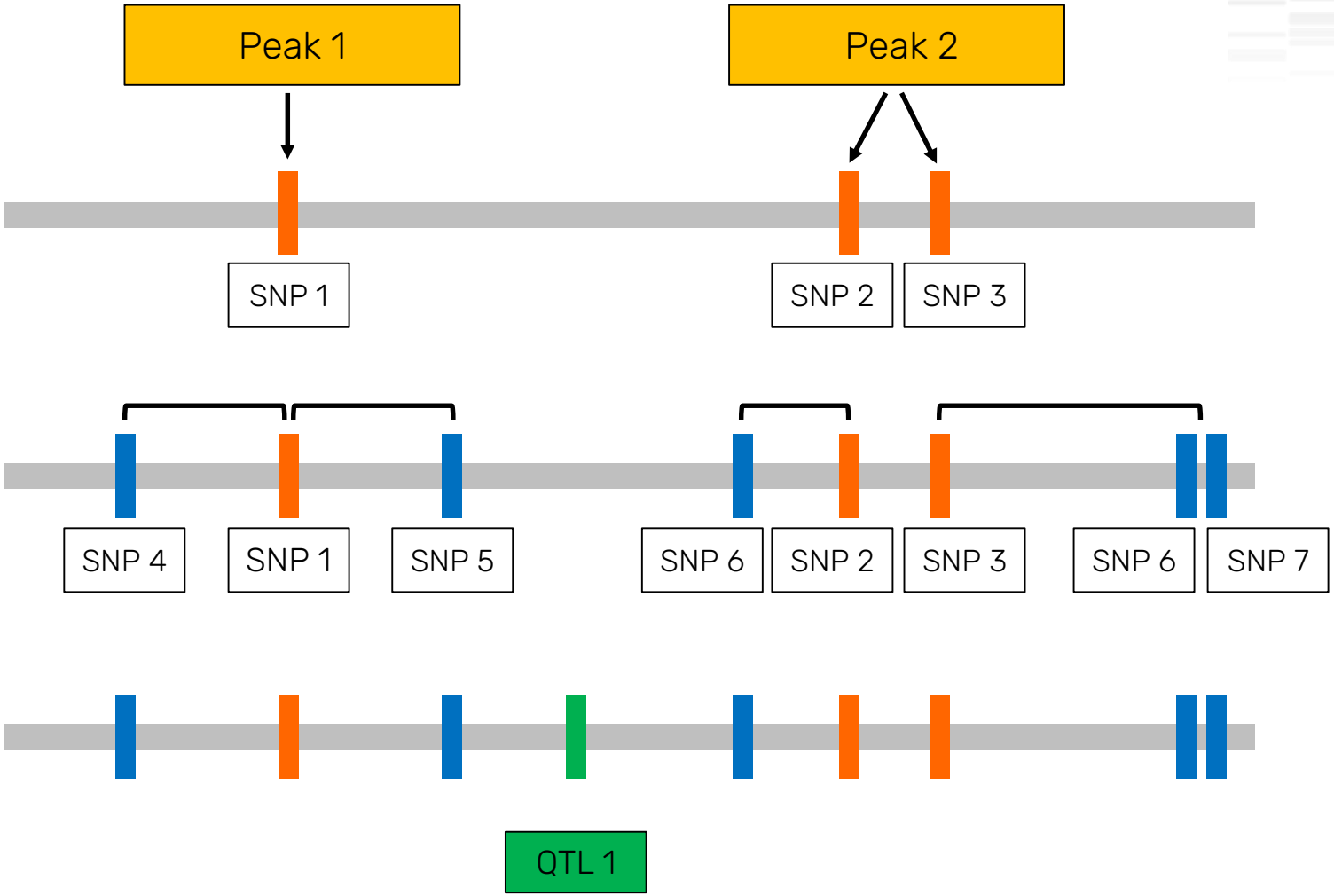
(2) Co-localization with QTLs



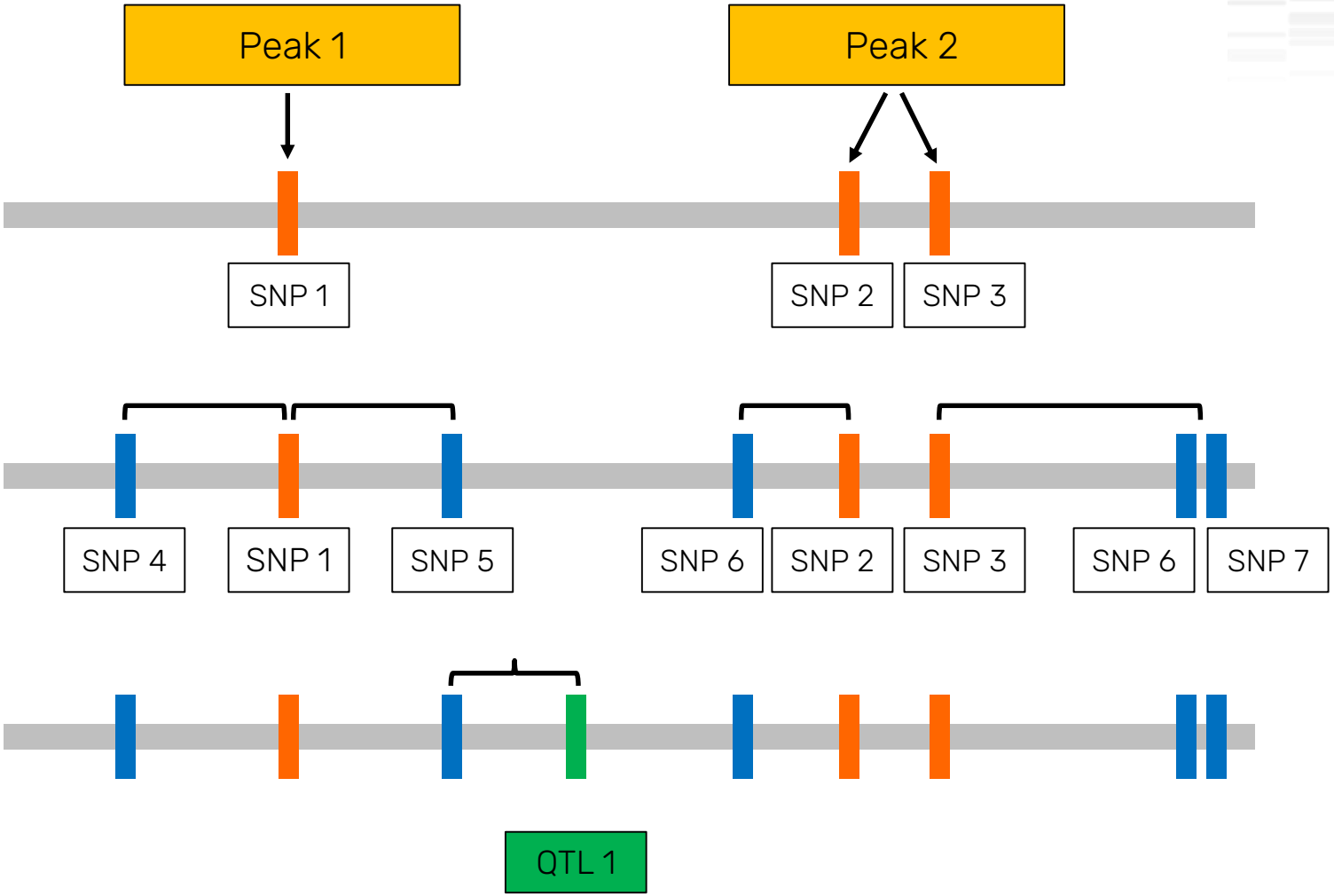
(2) Co-localization with QTLs



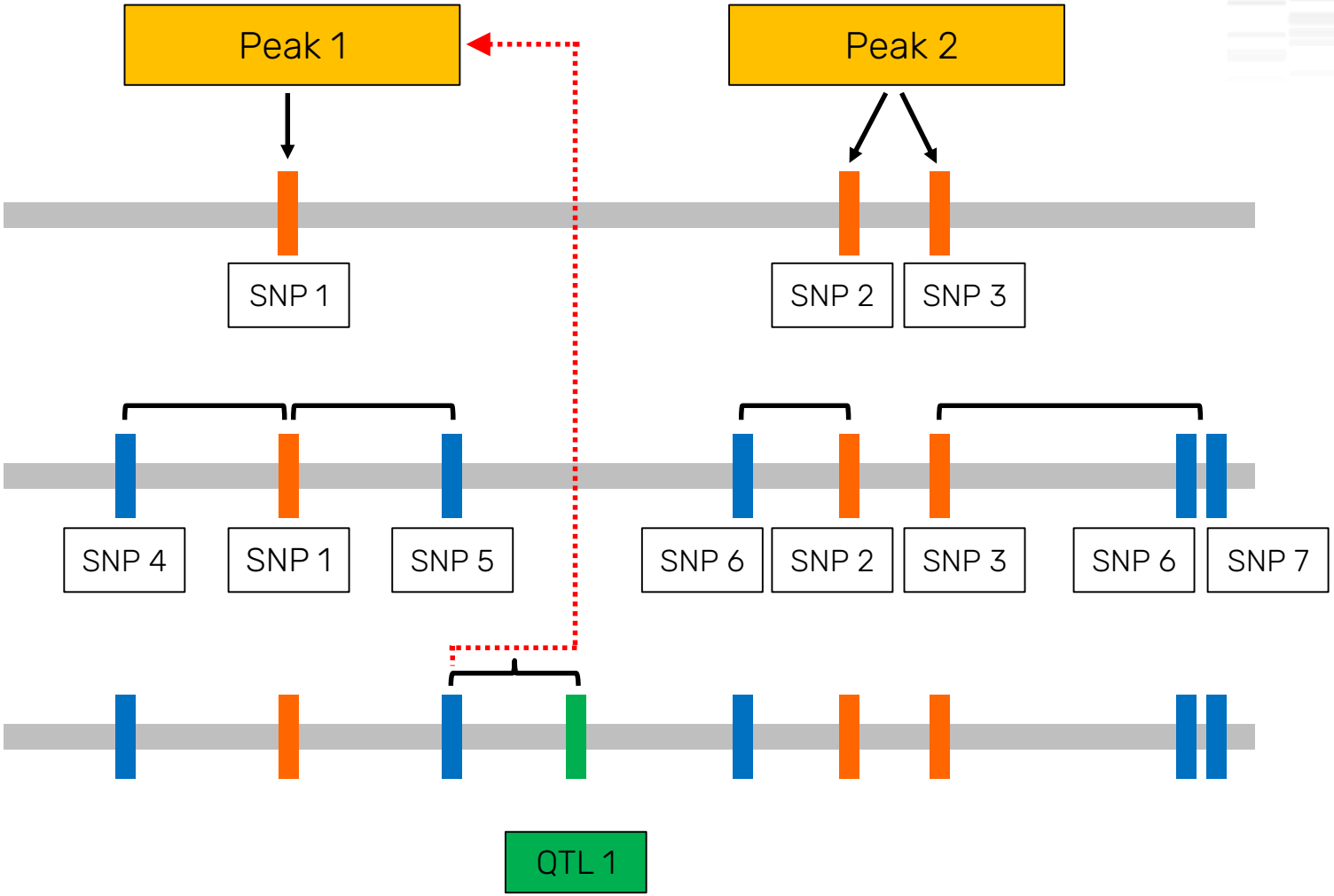
(2) Co-localization with QTLs



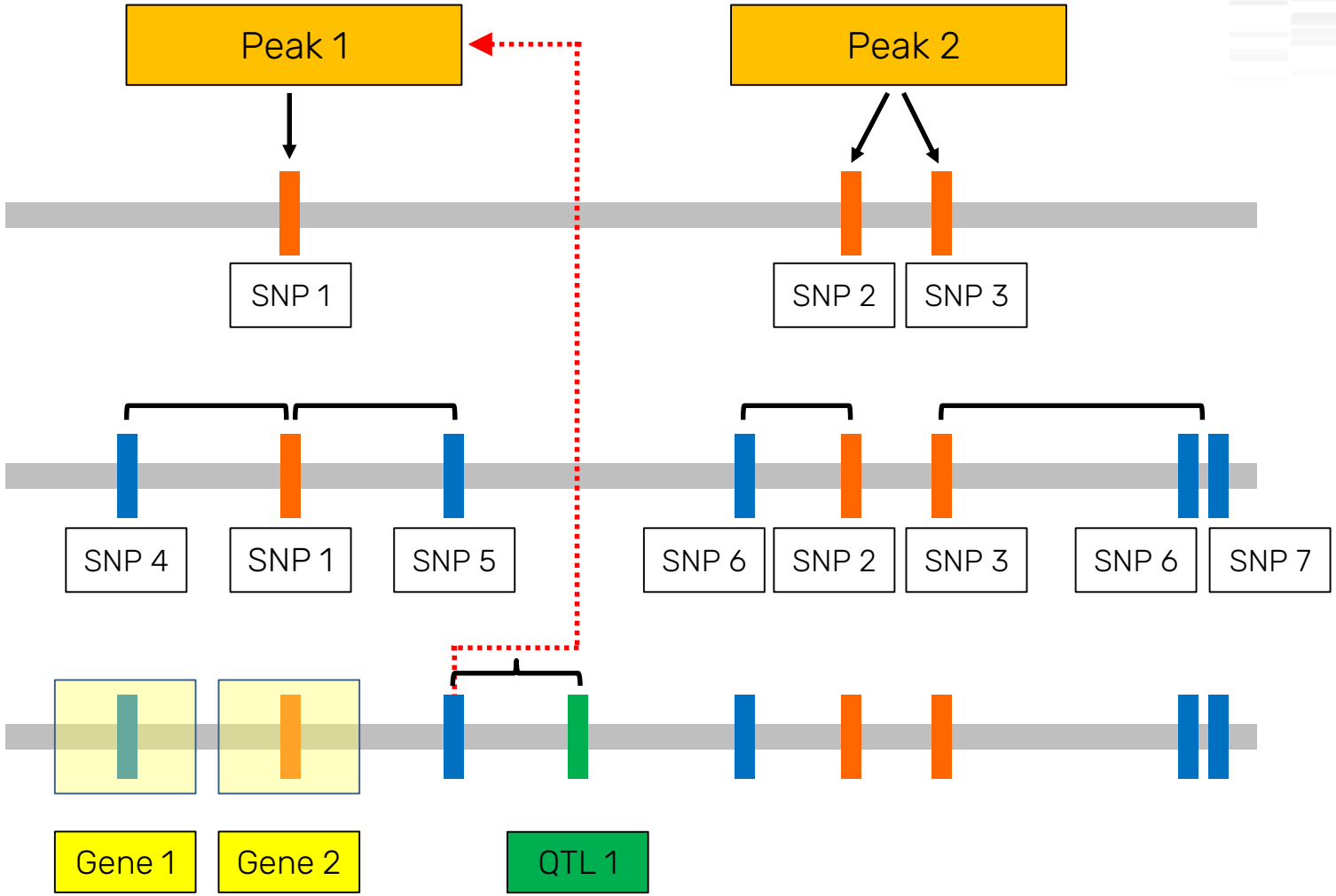
(2) Co-localization with QTLs



(2) Co-localization with QTLs



(2) Co-localization with QTLs



(3) Results

(3) GWAS

- 450 individuals retained for the analysis
- 2.557 MS peaks retained after filtering
- 955 MS peaks (37%) associated to 27,246 unique SNPs from GWAS
- 62,135 total associations among MS peaks and SNPs

(3) Co-localization with QTLs

- QTLs for drought, cold and nutrient stress and for productivity-related traits
- distance QTL – SNPs ≤ 50 kb
- 638 SNPs co-localize with QTLs
- 120 peaks co-localize with stress-related QTLs
- 25 peaks co-localize with productivity-related QTLs
- 137 total peaks co-localize with QTLs (8 peaks shared by the 2 groups)

(3) Peak annotation

- *in silico* only workflow:
    - analysis of raw chemical formulas (PubChem)
    - search on the 'KNApSAcK' database
    - search on the 'Natural Products' database
    - direct web search
- 30 putatively annotated peaks, among which 9 terpenes, 5 flavonoids, 4 polyacetylenes, 3 cinnamic acids and 2 phenolic acids
- 7 to 16 of these molecules belong to biochemical classes known as antioxidants

(3) Peak annotation

Peak	Molecule	Class
M201T581	4,5,9,10-dehydroisolongifolene isomer A	Sesquiterpenes
M203T379	Demethoxyencecalin	Chromenes
M287T548	Cyanidin	Flavonoids
M289T534	Pentahydroxychalcone	Flavonoids
M307T374	Brachystemidine A	Pyrroles
M349T519	Hexahydroxydimethylflavanone	Flavonoids
M131T321	2-Nonene-4,6,8-triyn-1-ol	Polyacetylenes
M294T138	deoxyfructosyl-leucine	Amino-acid derivatives
M145T650	1,9-Undecadiene-5,7-diyne	Polyacetylenes
M151T299	Coumaryl-alcohol isomer A	Cinnamic acids
M151T342	Coumaryl-alcohol isomer B	Cinnamic acids
M165T301	Coumaric acid	Cinnamic acids
M173T648	1,4-Tridecadiene-7,9-diyne	Polyacetylenes
M175T609	(2E,4E)-5-phenylpenta-2,4-dienoic acid	Styrenes
M201T687	4,5,9,10-dehydroisolongifolene isomer B	Sesquiterpenes
M202T649	4,5,9,10-dehydroisolongifolene isomer C	Sesquiterpenes
M217T336	(3S,9Z)-pentadeca-9,14-dien-4,6-diyn-3-ol	Polyacetylenes
M231T299	Dehydrocostus lactone isomer A	Sesquiterpenes
M231T362	Dehydrocostus lactone isomer B	Sesquiterpenes
M231T604	Dehydrocostus lactone isomer C	Sesquiterpenes
M243T272	Lumichrome	Alloxazines
M247T334	Annuolide A	Sesquiterpenes
M249T273	Annuolide E	Sesquiterpenes
M266T572	Heliannuol F	Sesquiterpenes
M273T341	8-Acetoxy-1,9,14-pentadecatriene-4,6-diyn-3-ol	Polyacetylenes
M273T635_2	Androst-5-en-4-one	Steroids
M299T342	3-phenyl-1-(2,3,4-trimethoxyphenyl)prop-2-en-1-one	Flavonoids
M346T682	Tambulin	Flavonoids
M372T338	4-(2-Amino-3-hydroxyphenyl)-4-oxobutanoic acid glucoside	Phenolic compounds
M515T343	Eugenol acetylrhamnosylglucoside	Phenolic compounds

(3) Peak annotation

Peak	Molecule	Class
M201T581	4,5,9,10-dehydroisolongifolene isomer A	Sesquiterpenes
M203T379	Demethoxyencecalin	Chromenes
M287T548	Cyanidin	Flavonoids
M289T534	Pentahydroxychalcone	Flavonoids
M307T374	Brachystemidine A	Pyrroles
M349T519	Hexahydroxydimethylflavanone	Flavonoids
M131T321	2-Nonene-4,6,8-triyn-1-ol	Polyacetylenes
M294T138	deoxyfructosyl-leucine	Amino-acid derivatives
M145T650	1,9-Undecadiene-5,7-diyne	Polyacetylenes
M151T299	Coumaryl-alcohol isomer A	Cinnamic acids
M151T342	Coumaryl-alcohol isomer B	Cinnamic acids
M165T301	Coumaric acid	Cinnamic acids
M173T648	1,4-Tridecadiene-7,9-diyne	Polyacetylenes
M175T609	(2E,4E)-5-phenylpenta-2,4-dienoic acid	Styrenes
M201T687	4,5,9,10-dehydroisolongifolene isomer B	Sesquiterpenes
M202T649	4,5,9,10-dehydroisolongifolene isomer C	Sesquiterpenes
M217T336	(3S,9Z)-pentadeca-9,14-dien-4,6-diyn-3-ol	Polyacetylenes
M231T299	Dehydrocostus lactone isomer A	Sesquiterpenes
M231T362	Dehydrocostus lactone isomer B	Sesquiterpenes
M231T604	Dehydrocostus lactone isomer C	Sesquiterpenes
M243T272	Lumichrome	Alloxazines
M247T334	Annuolide A	Sesquiterpenes
M249T273	Annuolide E	Sesquiterpenes
M266T572	Heliannuol F	Sesquiterpenes
M273T341	8-Acetoxy-1,9,14-pentadecatriene-4,6-diyn-3-ol	Polyacetylenes
M273T635_2	Androst-5-en-4-one	Steroids
M299T342	3-phenyl-1-(2,3,4-trimethoxyphenyl)prop-2-en-1-one	Flavonoids
M346T682	Tambulin	Flavonoids
M372T338	4-(2-Amino-3-hydroxyphenyl)-4-oxobutanoic acid glucoside	Phenolic compounds
M515T343	Eugenol acetylrhamnosylglucoside	Phenolic compounds

(3) Identification of candidate genes

Molecule	Class	Gene	Blast
Pentahydroxychalcone	flavonoids	Chr03g0130651	Cytochrome P450
Hexahydroxydimethyl-flavanone	flavonoids	Chr11g0515571	UDP-glycosyltransferase 73E1, likely family 1
		Chr11g0515581	UDP-glycosyltransferase 73E1, likely family 1
		Chr11g0515591	UDP-glycosyltransferase 73E1, likely family 1
Heliannuol F	terpenes	Chr09g0363371	UDP-glycosyltransferase 74G1, likely family 1
Dehydroisolongifolene	terpenes	Chr07g0314841	AP2/ERF transcription factor
		Chr07g0314871	AP2/ERF transcription factor

- genes considered as candidates only with SNPs directly landing in exons
- new tBLASTn analysis of all the sequences to obtain a more up-to-date annotation

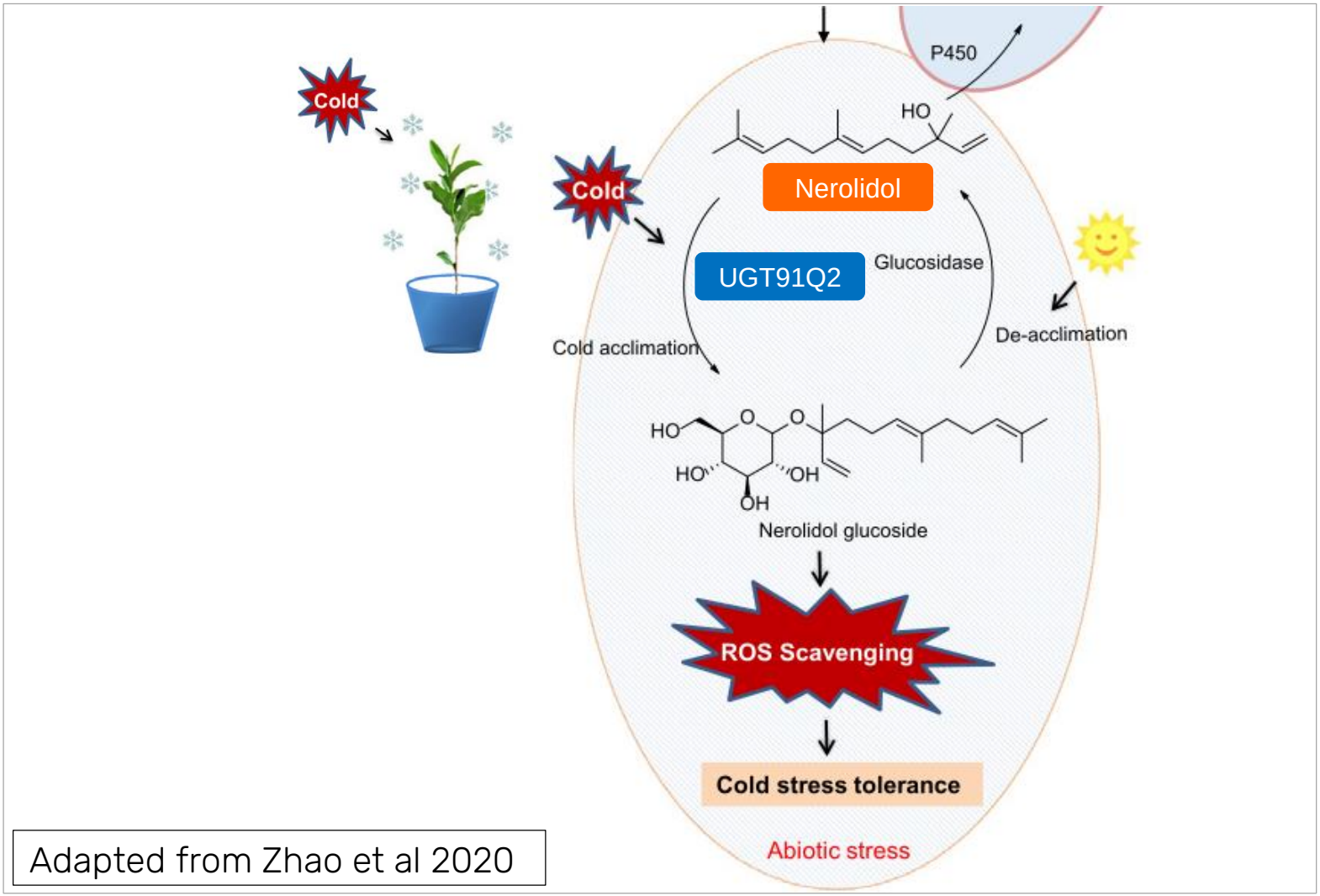
(3) UGTs

- uridine diphosphate (UDP) glycosyl-transferases (UGTs) transfer glycosyl residues to acceptor molecules
- ‘family 1’ UGTs is specific to secondary metabolites, e.g. terpenes, flavonoids, and cinnamic acids (Vogt and Jones 2000)
- glycosylation increases solubility and activity of these compounds
- the expression of UGTs is induced by several abiotic stresses

(3) Mechanism of UGTs

- UGTs enhance the plant antioxidant capacity by glycosylating secondary metabolites and provide, for instance, tolerance to:
  - cold stress in tea plant (nerolidol, Zhao et al 2020)
  - drought, cold and salinity stress in *Arabidopsis* (anthocyanin, Li et al 2017)

(3) Mechanism of UGTs



Adapted from Zhao et al 2020

(4) Next steps

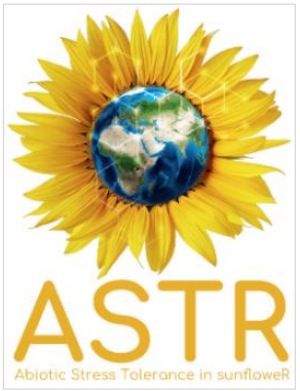
- further analysis on the candidate genes, e.g. search for missense mutations
- characterization of UGTs across different genotypes

Bibliography:

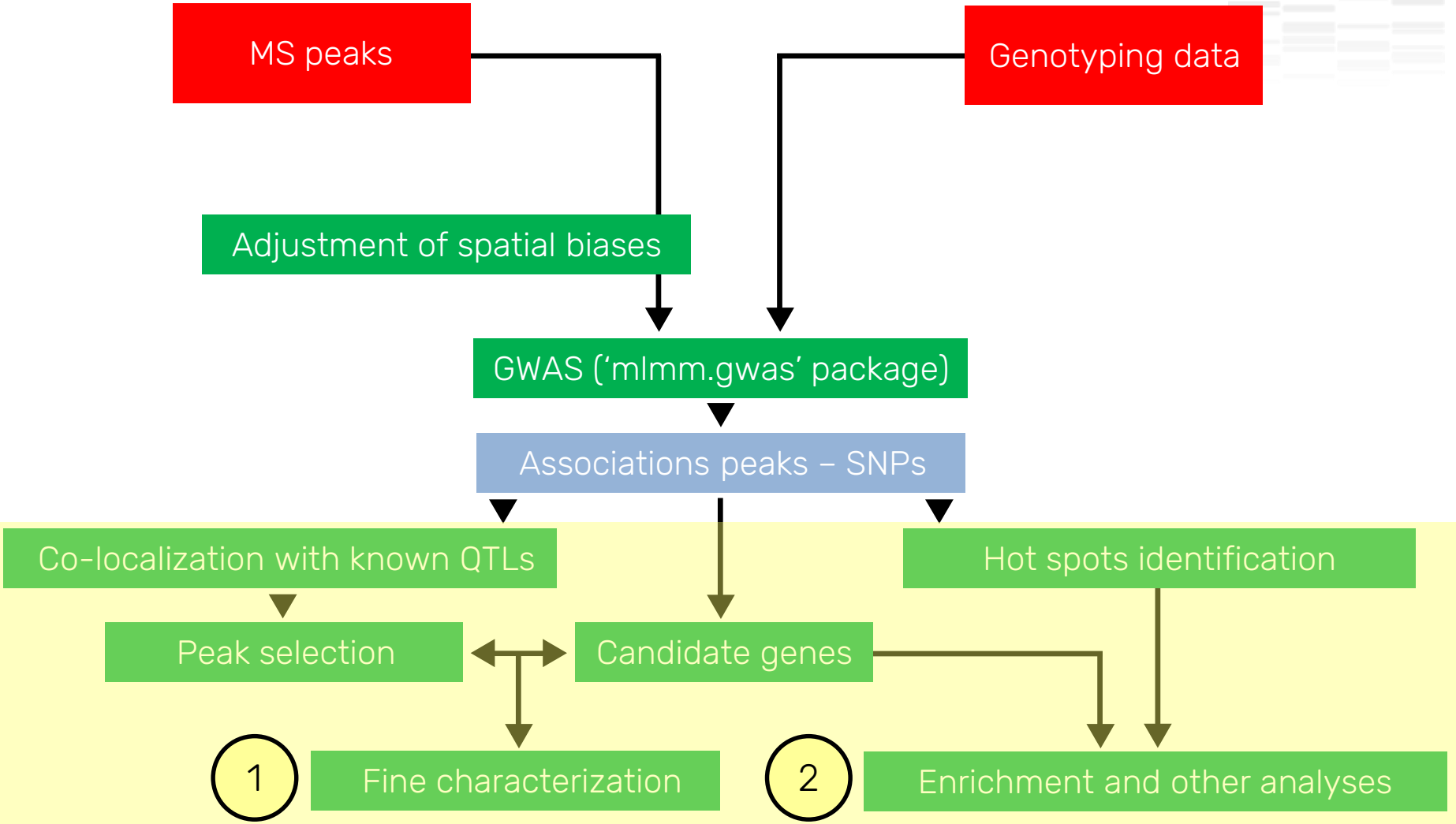
- Badouin et al 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*
- Li et al 2017. The Arabidopsis UDP-glycosyltransferases UGT79B2 and UGT79B3, contribute to cold, salt and drought stress tolerance via modulating anthocyanin accumulation. *The plant journal*
- Nakabayashi et al 2014. Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. *The plant journal*
- Nakabayashi and Saito 2015. Integrated metabolomics for abiotic stress responses in plants. *Current opinion on plant biology*
- Nützmann et al 2016. Plant metabolic clusters – from genetics to genomics. *New phytologist*
- Vogt and Jones 2000. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends in plant science*
- Zhao et al 2020. Sesquiterpene glucosylation mediated by glucosyltransferase UGT91Q2 is involved in the modulation of cold stress tolerance in tea plants. *New phytologist*

Many thanks to:

- Nicolas Langlade
- Annick Moing
- Stéphane Bernillon
- Nicolas Blanchet
- Harold Duruflé
- Vincent Segura
- David Pot
- Gabriela Bindea
- Elise Maigné
- Florie Gosseau
- Pierre Casadebaig



(2) Analysis workflow



(3) Unsupervised search for candidates

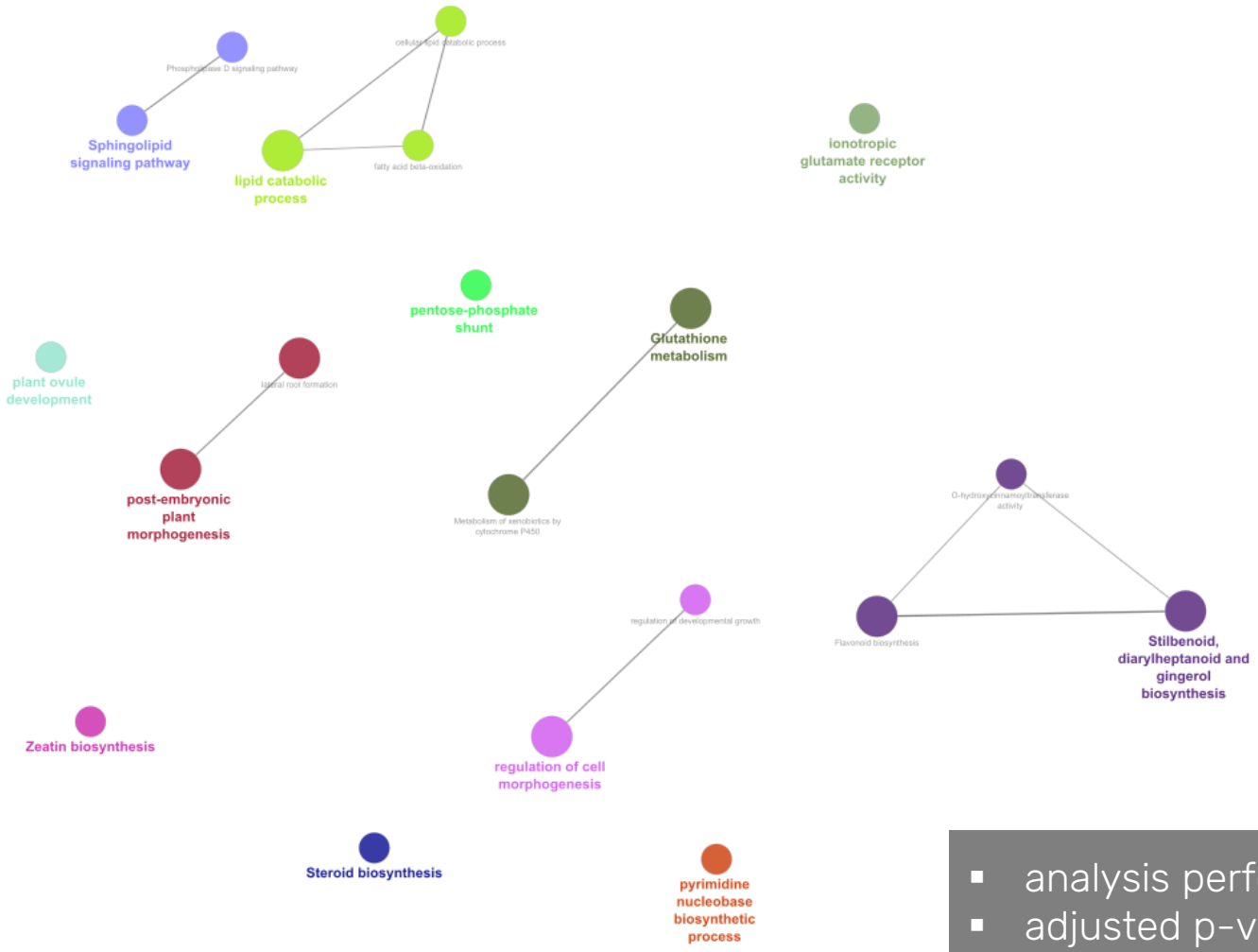
(3) THE ‘mlmm.gwas’ R PACKAGE

- mlmm = multi-locus mixed model
- linear mixed model: the kinship matrix is a random polygenic term
- multi-locus approach: no multiple testing correction needed
- several SNPs are identified, then filtered based on the extended Bayesian information criterion (eBIC)

(3) Unsupervised search for candidates

- based on all the SNPs associated to MS peaks, i.e. irrespective of their co-localization with previously known QTLs
- 1,768 SNPs in the exons of 533 genes
- enrichment analysis using ‘GO process’, ‘GO function’ and ‘KEGG’ pathways

(3) Enrichment analysis



- analysis performed with ClueGO
- adjusted p-value for groups = 0.05

(3) Enrichment analysis

Term	Ontology source	BH-adjusted *p*-value
Post-embryonic plant morphogenesis	GO	0.0042
Glutathione metabolism	KEGG	0.0014
Sphingolipid signaling pathway	KEGG	0.0242
Flavonoid biosynthesis	KEGG	0.0016
Lipid catabolic process	GO	0.0012

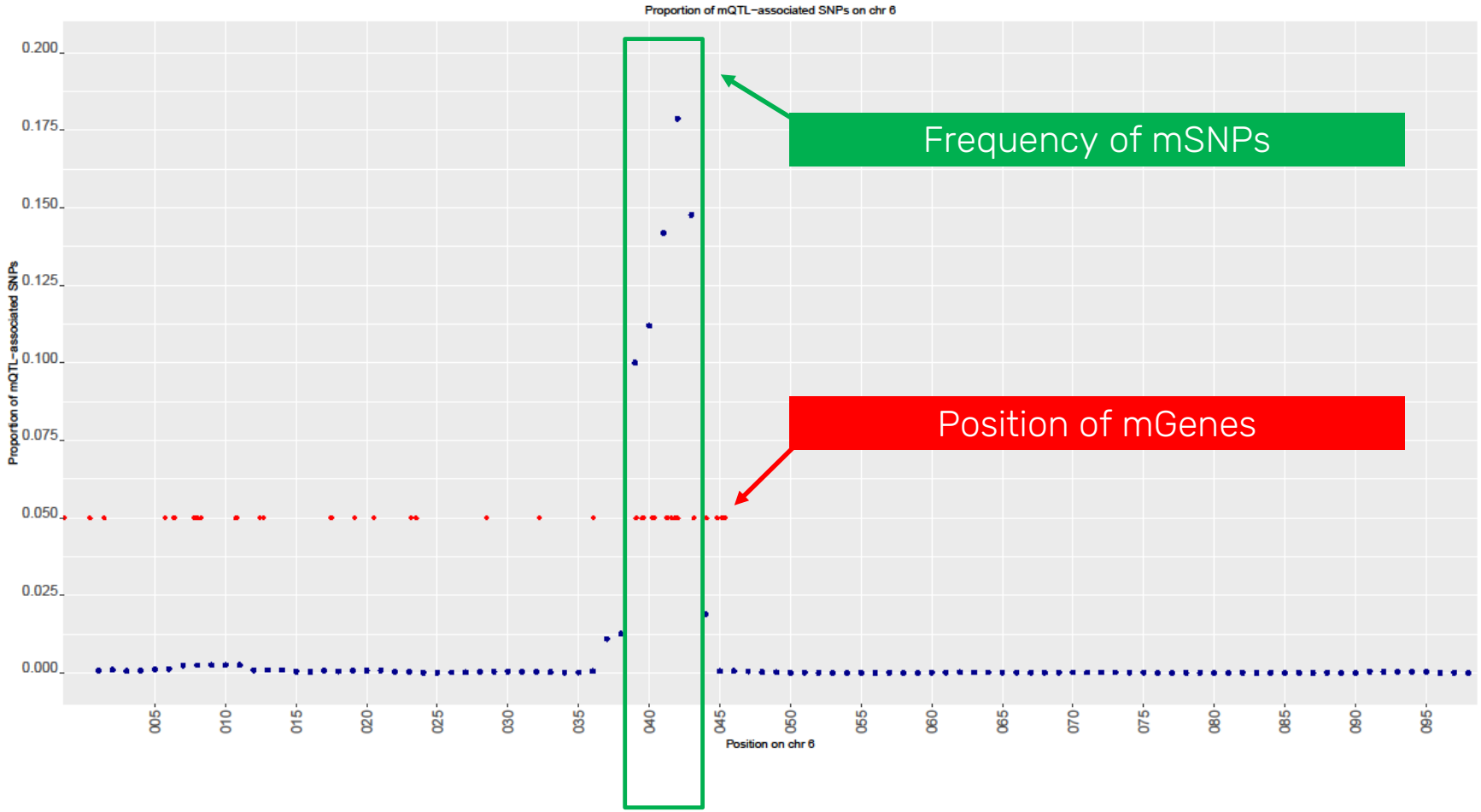
- *p*-values are not especially robust, but can be considered acceptable in this context
- no new tBLASTn analysis to to obtain a more up-to-date annotation

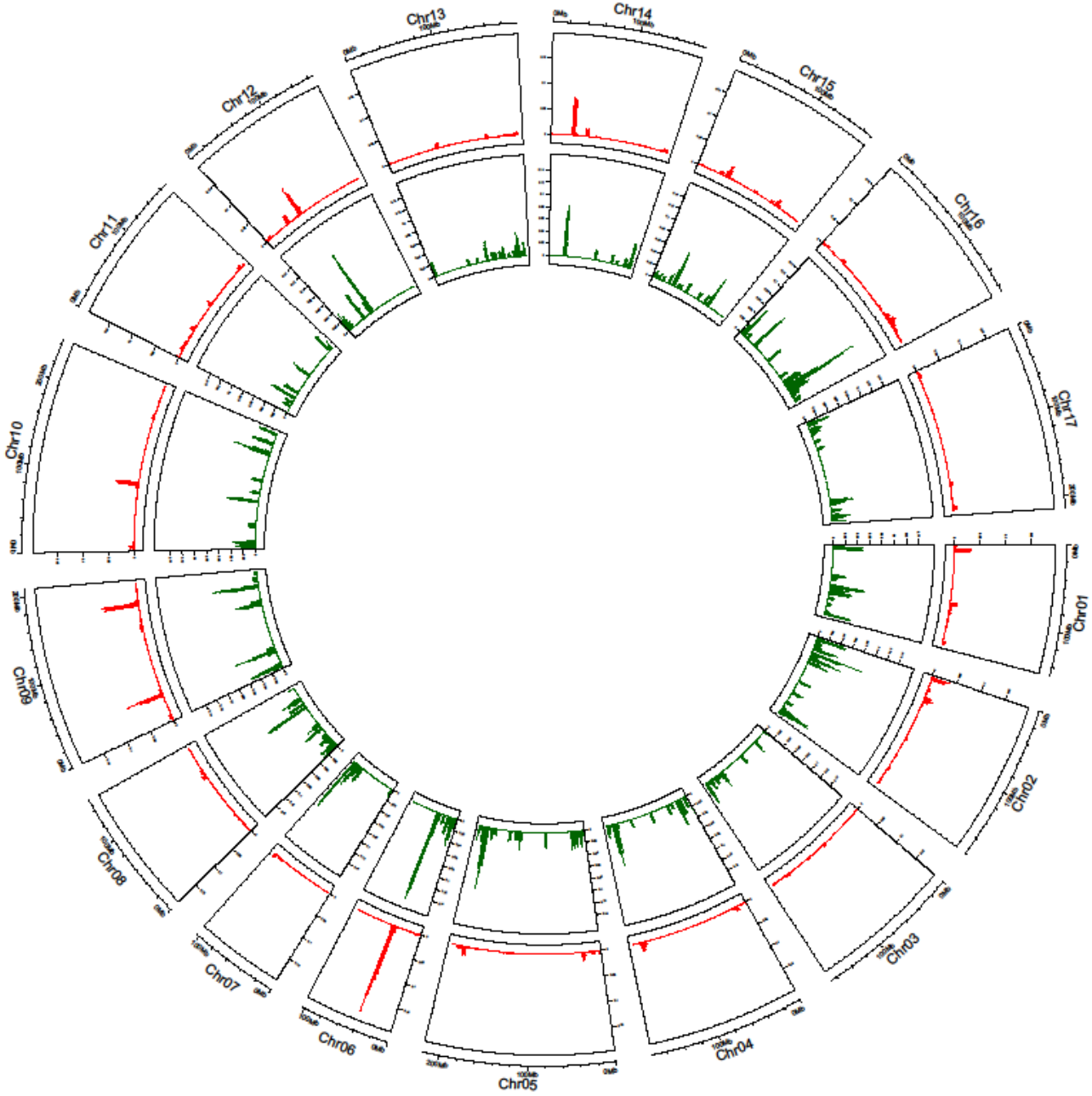
(3) Identification of hot spots

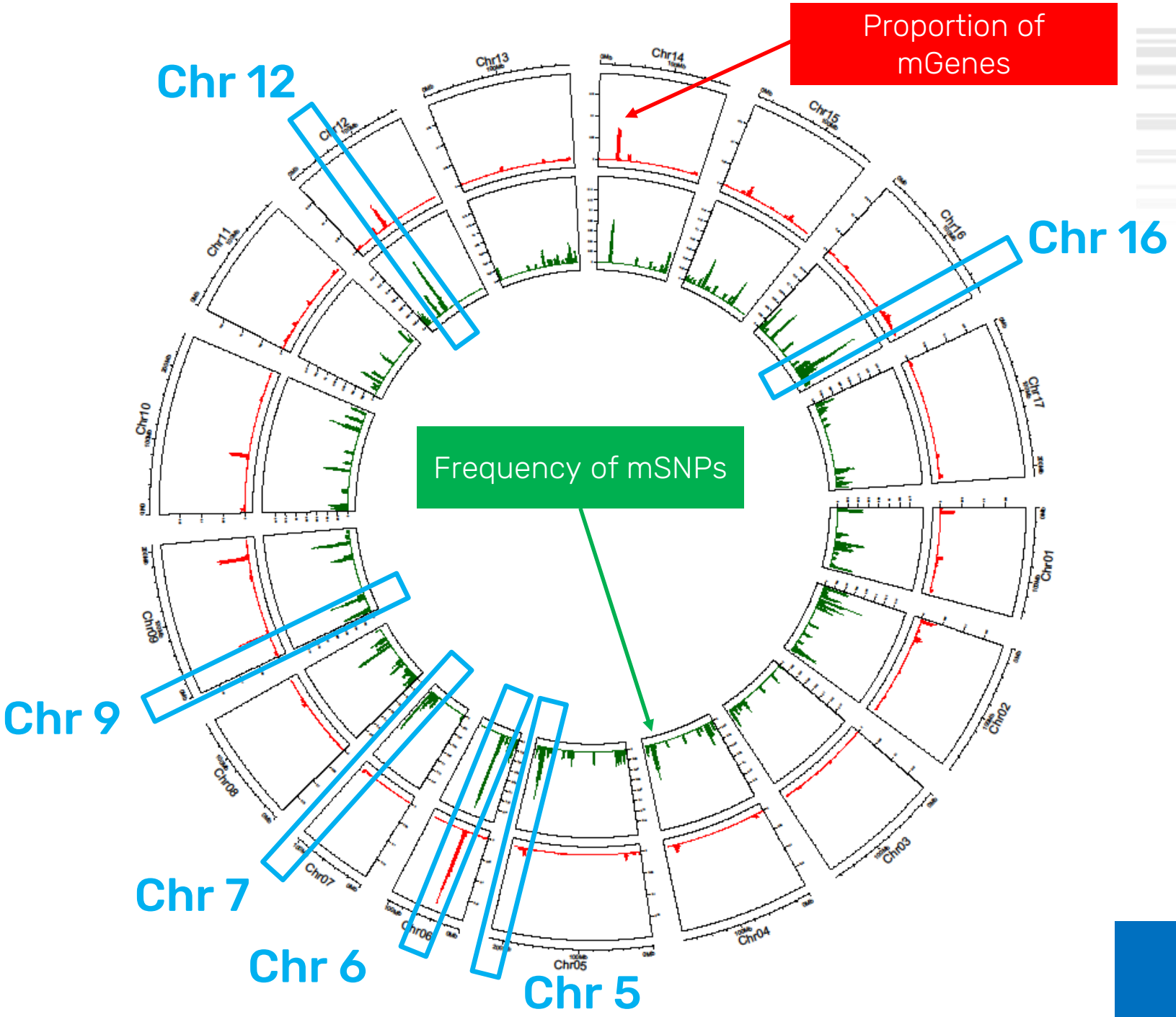
(3) Identification of hot spots

- at least 20 clusters of metabolite-related genes are reported in species such as rice, tomato, *Arabidopsis* and *Lotus*
- mainly related to the biosynthesis of terpenes, cyanogenic glucosides and alkaloids
- typically consisting of 3 – 10 genes (Nützmann et al 2016)

(3) Identification of hot spots




Chr13
Chr14
Chr15
Chr16
Chr17
Chr01
Chr02
Chr03
Chr04
Chr05
Chr06
Chr07
Chr08
Chr09
Chr10
Chr11
Chr12


Proportion of mGenes
Frequency of mSNPs
Chr 12
Chr 16
Chr 9
Chr 7
Chr 6
Chr 5
Chr01
Chr02
Chr03
Chr04
Chr05
Chr06
Chr07
Chr08
Chr09
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17

(3) Identification of hot spots

- the 6 identified hot spots span 24 Mbs, i.e. 0.7 % of the genome
- they harbour a relevant percentage of mSNPs, i.e. 6,574 out of 27,246 (24 %)
- they also harbour a relevant percentage of mGenes: 92 out of 533 (17 %)

(4) Next steps