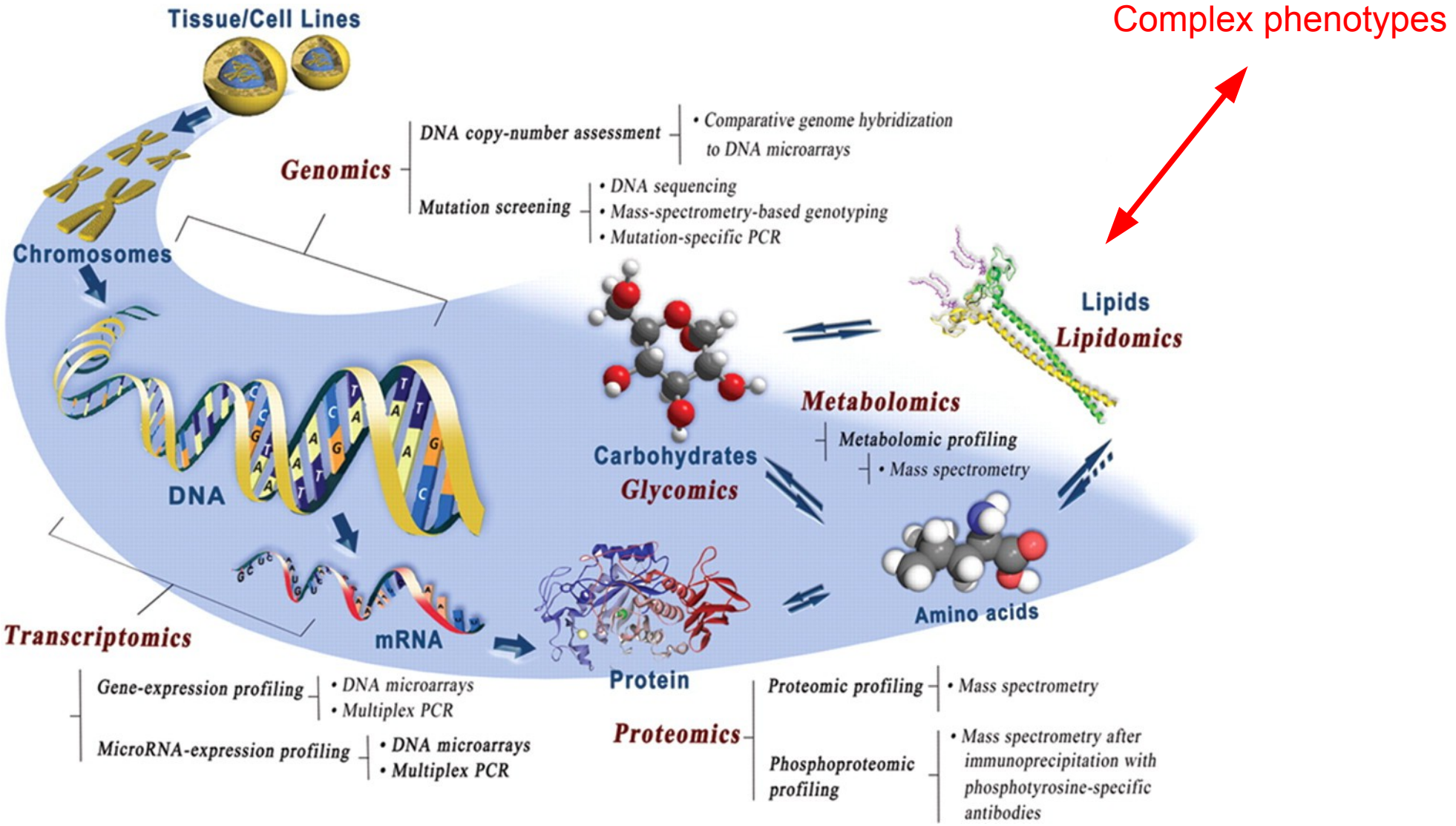
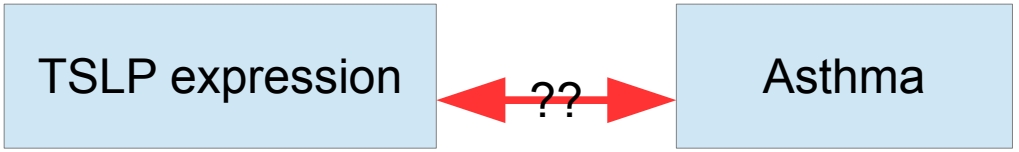
Rick Jansen VUMC Amsterdam

Functional genomics: eQTL analysis as a first step



Example: TSLP expression in Asthma



Levels of human TSLP RNA and protein are increased in the airways of patients with asthma.

Example: TSLP expression in Asthma

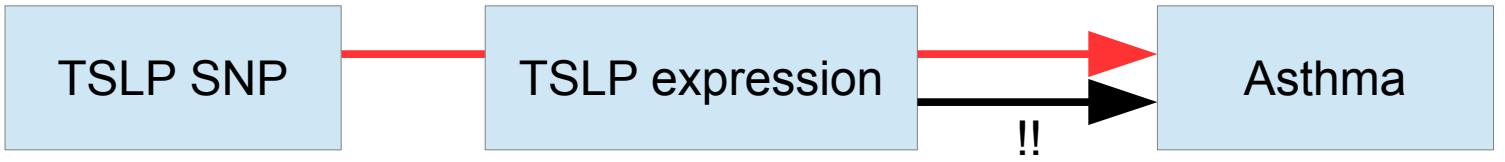
TSLP SNPs are risk loci for asthma (GWAS).

THE JOURNAL OF Allergy AND Clinical Immunology (2014)

Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype

Manuel A.R. Ferreira, PhD*, Melanie C. Matheson, PhD*, Clara S. Tang, PhD, Raquel Granell, PhD, Wei Ang,

Example: TSLP expression in Asthma



Reducing TSLP decreases asthma symptoms in double blind placebo controlled study.

The NEW ENGLAND JOURNAL of MEDICINE (2014)

ORIGINAL ARTICLE

Effects of an Anti-TSLP Antibody on Allergen-Induced Asthmatic Responses

Gail M. Gauvreau, Ph.D., Paul M. O'Byrne, M.B., Louis-Philippe Boulet, M.D., Ying Wang, Ph.D., Donald Cockcroft, M.D., Jeannette Bigler, Ph.D., J. Mark FitzGerald, M.D., Michael Boedigheimer, Ph.D., Beth E. Davis, Ph.D., Clapton Dias, Ph.D., Kevin S. Gorski, Ph.D., Lynn Smith, Ph.D., Edgar Bautista, B.S., Michael R. Comeau, B.S., Richard Leigh, M.B., Ch.B., Ph.D., and Jane R. Parnes, M.D.

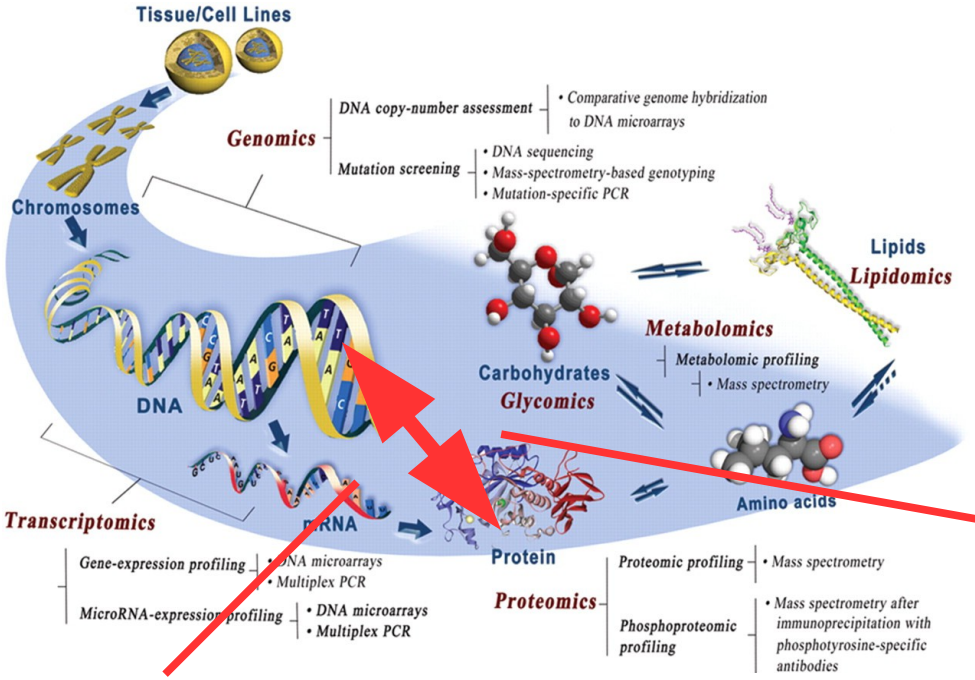
Example: TSLP expression in Asthma

TSLP SNPs are associated with TSLP expression in lung and blood tissues.

Example: TSLP expression in Asthma

It seems likely that the TSLP SNP influences asthma via the intermediating TSLP expression.

Functional genomics: protein QTLs



Variation and genetic control of protein abundance in humans

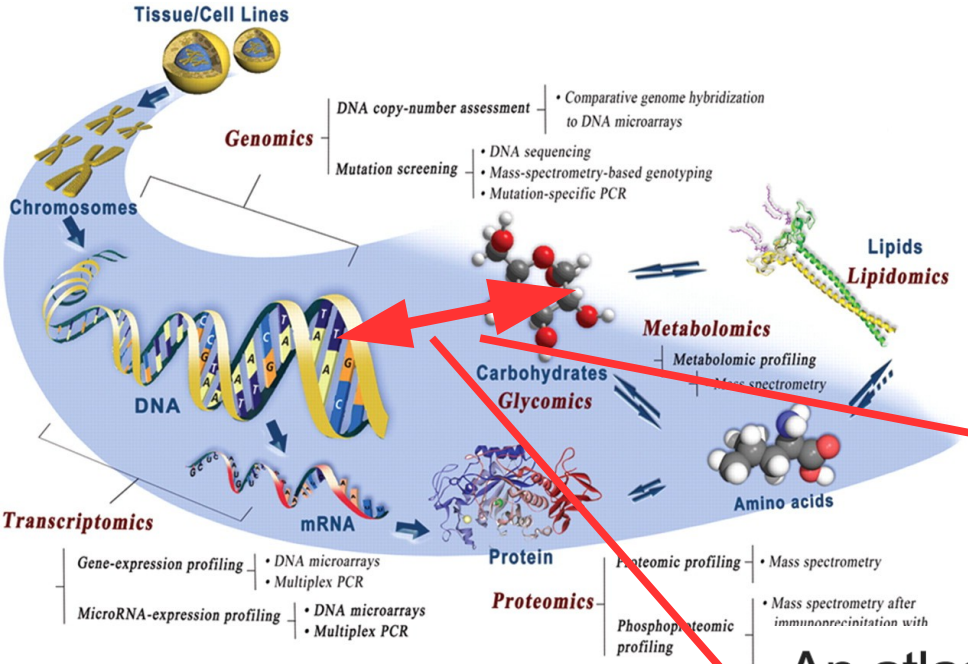
Linfeng Wu, Sophie I. Candille, Yoonha Choi, Dan Xie, Lihua Jiang, Jennifer Li-Pook-Than, Hua Tang & Michael Snyder

Affiliations | Contributions | Corresponding authors

Nature **499**, 79–82 (04 July 2013) | doi:10.1038/nature12223

5953 proteins in 95 subjects, 77 proteins were associated with DNA variant

Functional genomics: metabolite QTLs



An atlas of genetic influences on human blood metabolites

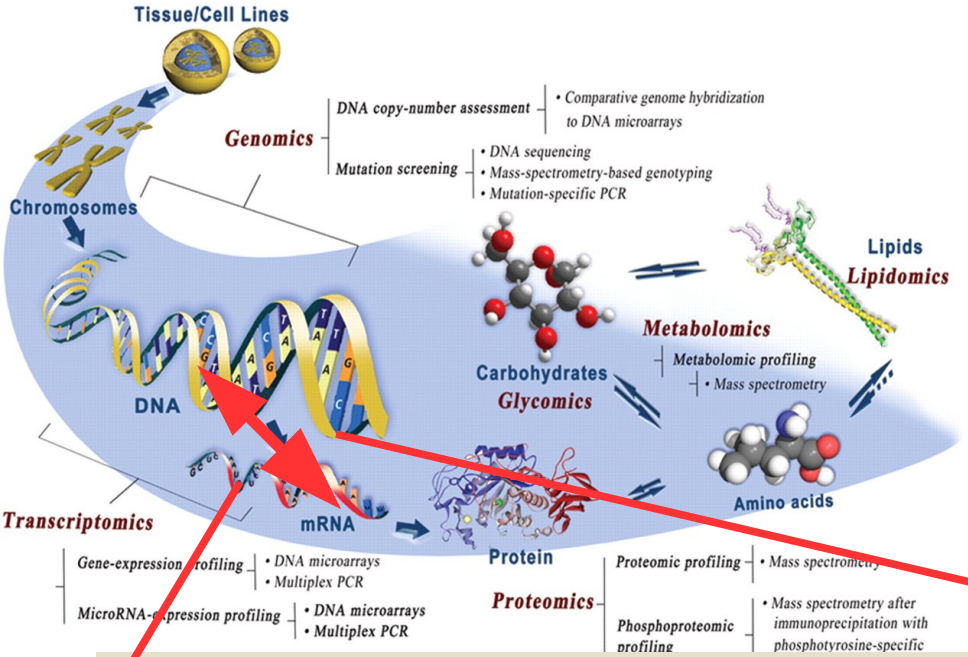
So-Youn Shin, Eric B Fauman, Ann-Kristin Petersen, Jan Krumsiek, Rita Santos, Jie Huang, Matthias Arnold, Idil Erte, Vincenzo Forgetta, Tsun-Po Yang, Klaudia Walter, Cristina Menni, Lu Chen, Louella Vasquez, Ana M Valdes, Craig L Hyde, Vicky Wang, Daniel Ziemek, Phoebe Roberts, Li Xi, Elin Grundberg, The Multiple Tissue Human Expression Resource (MuTHER) Consortium, Melanie Waldenberger, J Brent Richards, Robert P Mohney + *et al.*

Affiliations | Contributions | Corresponding authors

Nature Genetics **46**, 543–550 (2014) | doi:10.1038/ng.2982

7824 subjects, 400/529 metabolites associated with DNA variants

Functional genomics: expression QTLs (eQTLs)



ARTICLES

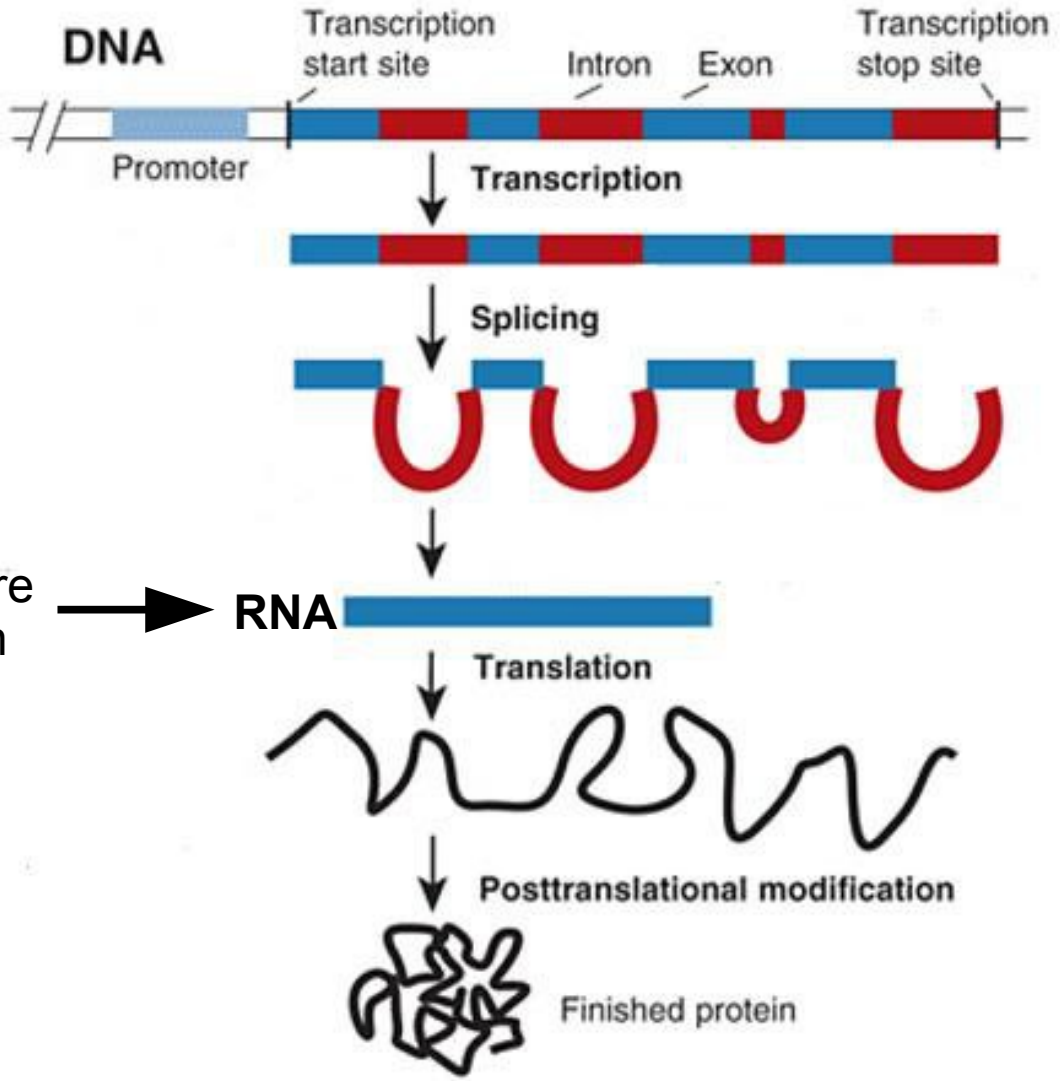
nature genetics (2014)

Heritability and genomics of gene expression in peripheral blood

Fred A Wright[1–3,13], Patrick F Sullivan[4,13], Andrew I Brooks[5], Fei Zou[6], Wei Sun[6], Kai Xia[6], Vered Madar[6], Rick Jansen[7], Wonil Chung[6], Yi-Hui Zhou[1,2], Abdel Abdellaoui[8], Sandra Batista[9], Casey Butler[9], Guanhua Chen[6], Ting-Huei Chen[6], David D'Ambrosio[10], Paul Gallins[4], Min Jin Ha[6], Jouke Jan Hottenga[8], Shunping Huang[9], Mathijs Kattenberg[8], Jaspreet Kochar[10], Christel M Middeldorp[8], Ani Qu[10], Andrey Shabalin[11], Jay Tischfield[5], Laura Todd[4], Jung-Ying Tzeng[1,2], Gerard van Grootheest[7], Jacqueline M Vink[8], Qi Wang[10], Wei Wang[12], Weibo Wang[9], Gonneke Willemsen[8], Johannes H Smit[7], Eco J de Geus[8], Zhaoyu Yin[6], Brenda W J H Penninx[7] & Dorret I Boomsma[8]

~5000 subjects, 52% of all gene expression associated with DNA variants in cis

Central Dogma of Molecular Biology



Here we measure gene expression →

genes: ~20.000

transcripts: ? > 50.000

proteins: ? >100.000

Gene expression is a quantitative measure:
the amount of transcribed RNA for each gene or exon.

Gene Expression in NTR (Netherlands twin registry) and NESDA (Netherlands study of anxiety and depression)

5362 peripheral blood samples:

-3370 samples from NTR (714 MZ twin pairs & 663 DZ twin pairs)

-1992 samples from NESDA (882 Major depressive disorder cases)

Affymetrix HGU219 micro array platform (3'-end design):

- 548.517 probes (25 bp), summarized in 44,241 probesets targeting 18,238 genes

Causes of gene expression variability

- Time and tissue
- Measurement noise (plate, well)
- Lab (blood draw)
- RNA quality (degradation)
- Tissue mixtures
- Demographic variables:
    - -sex (Jansen et al, *BMC Genomics* (2014)
    - -age (Van den Akker et al, *Aging Cell (2013)*)
    - -BMI (Van Dongen et al, International Journal of Obesity)
    - -smoking (Vink et al, *submitted*)

In 3563 unrelated whole blood gene expression samples, Bonferroni corrected:

covariate	# genes associated
plate	18,100
well	14,577
time between blood draw and Array measurement	1697
blood draw hour	1687
lab	1500
sex	2508
bmi	914
smoking	995
age	440

RNA: so close to DNA, and almost as complex as a 'complex trait' !!

eQTL analysis

Gene expression measures of:
-20.000 genes, 50.000 transcripts or 180.000 exons

When using 1000 Genomes imputation: >8 million SNPs

- Cis eQTL analysis: only compute associations between SNPs at < 1 MB distance of the gene
- Trans eQTL analysis: GWAS (180.000 times!)

Super fast software: Matrix eQTL (R package, Shabalin et al, Bioinformatics (2012))

For 2,201 genes and 57,333 SNPs in 840 samples:

Method	No covar.	10 covar.
Plink	9.4	583.3 days
Merlin	19.6	20.0 days
R/qtl	1.0	4.7 days
snpMatrix	3.2	5.1 days
eMap	17.8	N/A days
FastMap	10.3	N/A hours
Matrix eQTL, Matlab	11.8	11.8 **minutes**
Matrix eQTL, Rev R	14.6	14.6 **minutes**
Matrix eQTL, R+GOTO	19.4	19.4 **minutes**

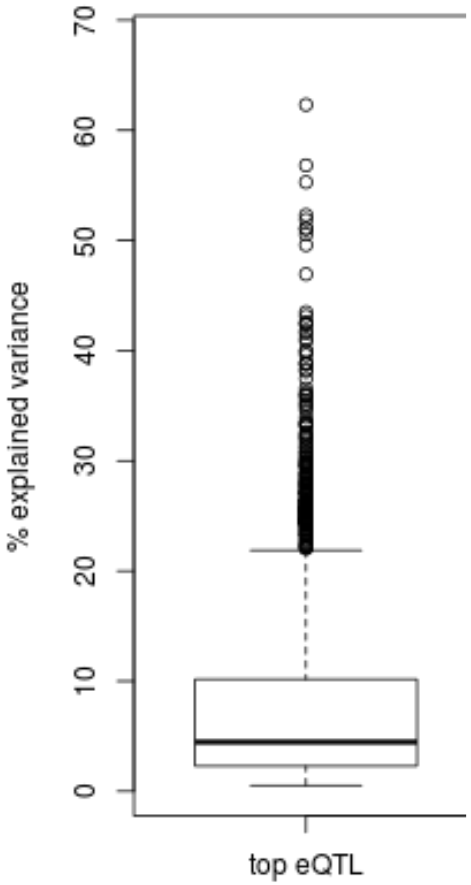
Trick: reduce testing to matrix multiplication,
Optimize matrix multiplication
(get the right R BLAS library)

We run ±40,000 1000 Genome GWAS's in 24 hour on a 10 core machine.

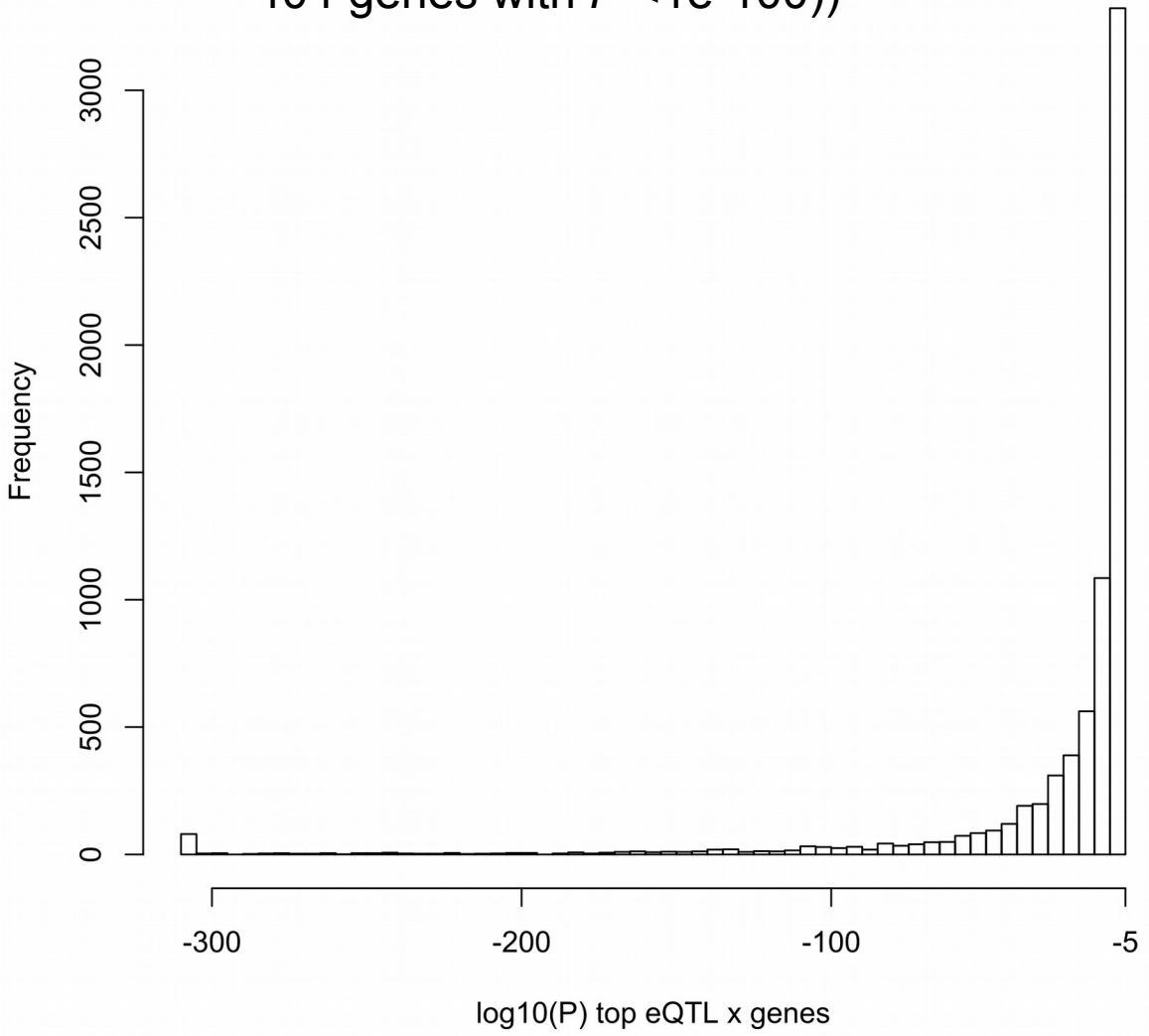
Cis eQTL results (N=5071, whole blood)

Cis regulation of 7,120 genes (**39%**)

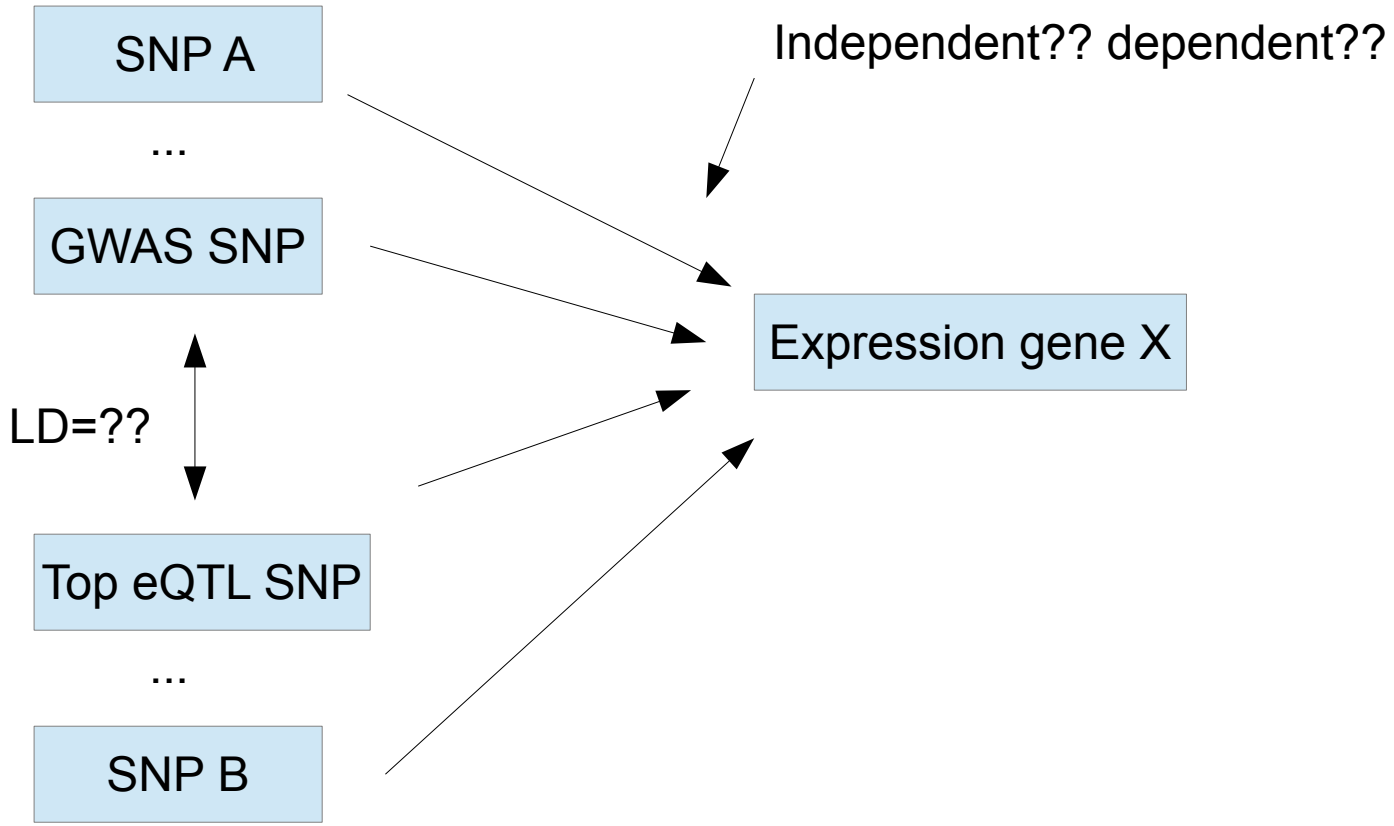
eQTLs explain on average 5%
of expression variability



FDR 0.05 : *P*< 1e-5
3798 genes with *P* <1e-10))
849 genes with *P* <1e-50))
404 genes with *P* <1e-100))

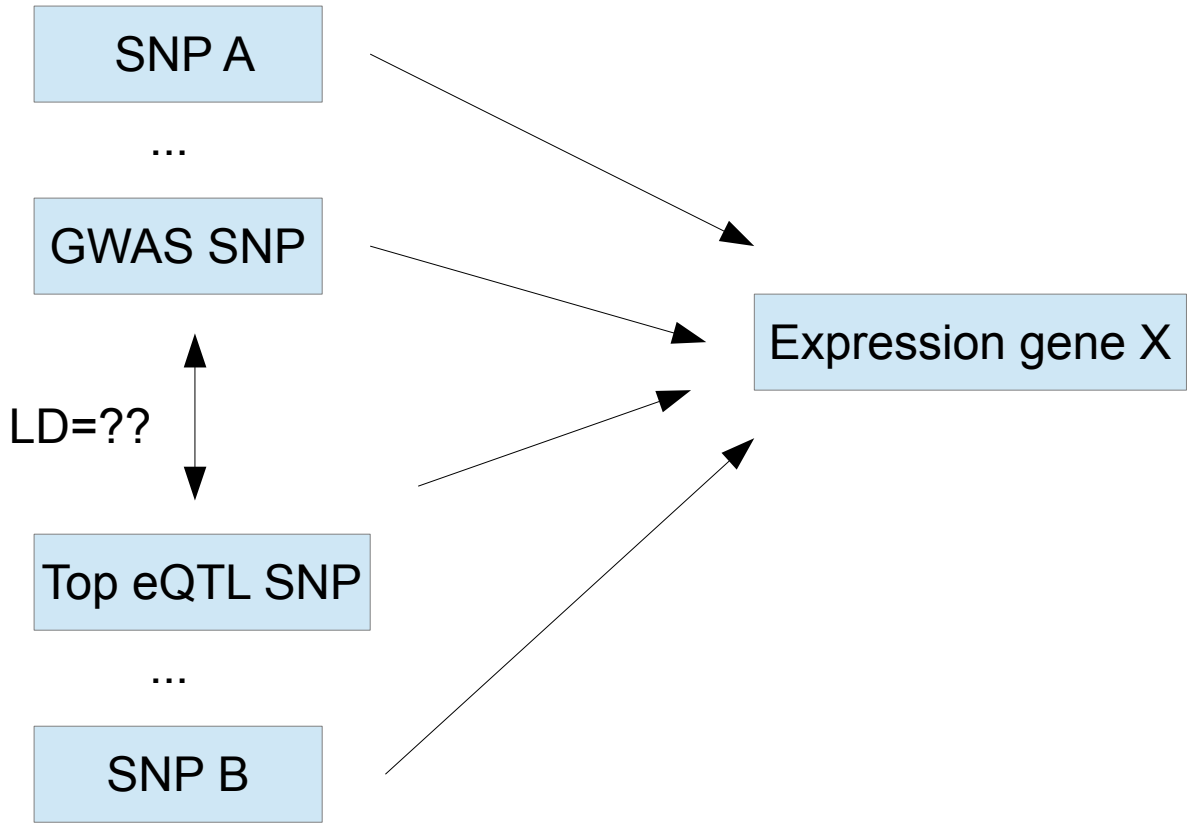


GWAS SNP look up in eQTL analysis



The number of eQTL SNPs per gene varies a lot (median=115, mean=237, SD=424, 410 genes with more than 1000 eQTLs).

GWAS SNP look up in eQTL analysis



Approach not using conditional analysis:
LD>0.8, GWAS SNP and top eQTL tag same locus: GWAS SNP is an eQTL
LD<0.2, GWAS SNP and top eQTL tag different locus: GWAS SNP is an eQTL
0.8>LD>0.2, GWAS SNP does not tag eQTL locus: GWAS SNP is not an eQTL

In case of conditional analysis: only conditioned on top eQTL and not on other possible independent SNPs.

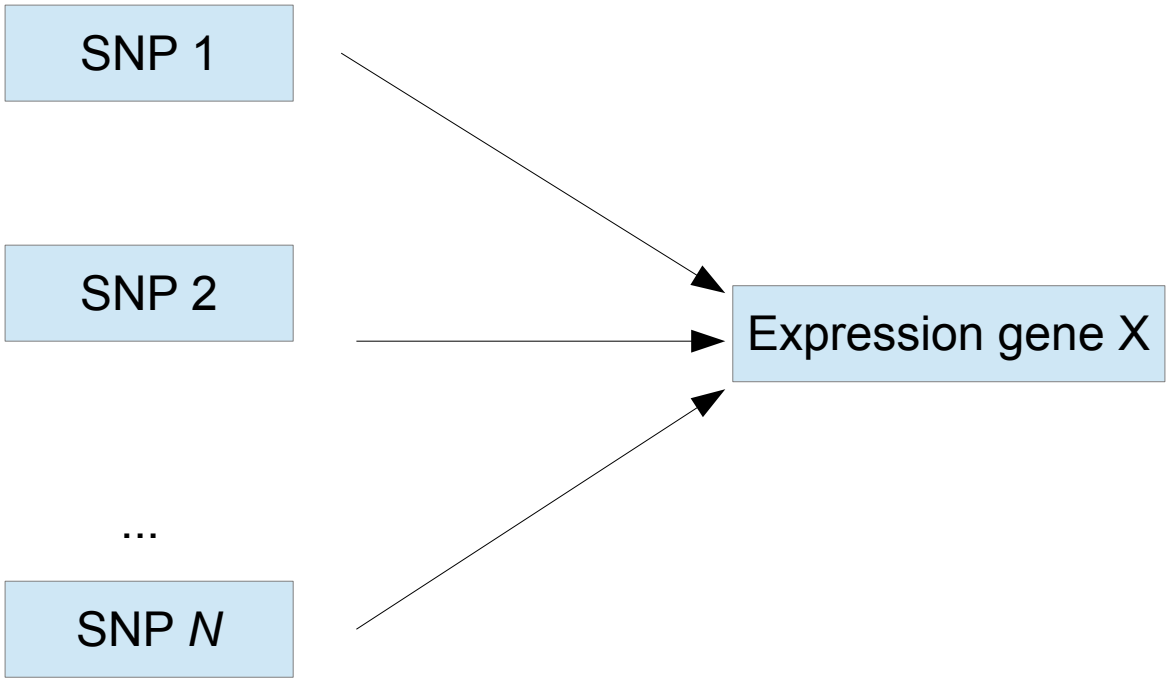
Example: 18 SNPs associated with CRP protein level in meta GWAS (>80.000 subjects)

15 'eQTLs':
1 GWAS SNP in LD>0.8 with top eQTL SNP (purple)
9 GWAS SNPs are in LD< 0.2 met top eQTL SNP (green)
After conditional analysis only 2 remain significant (blue)

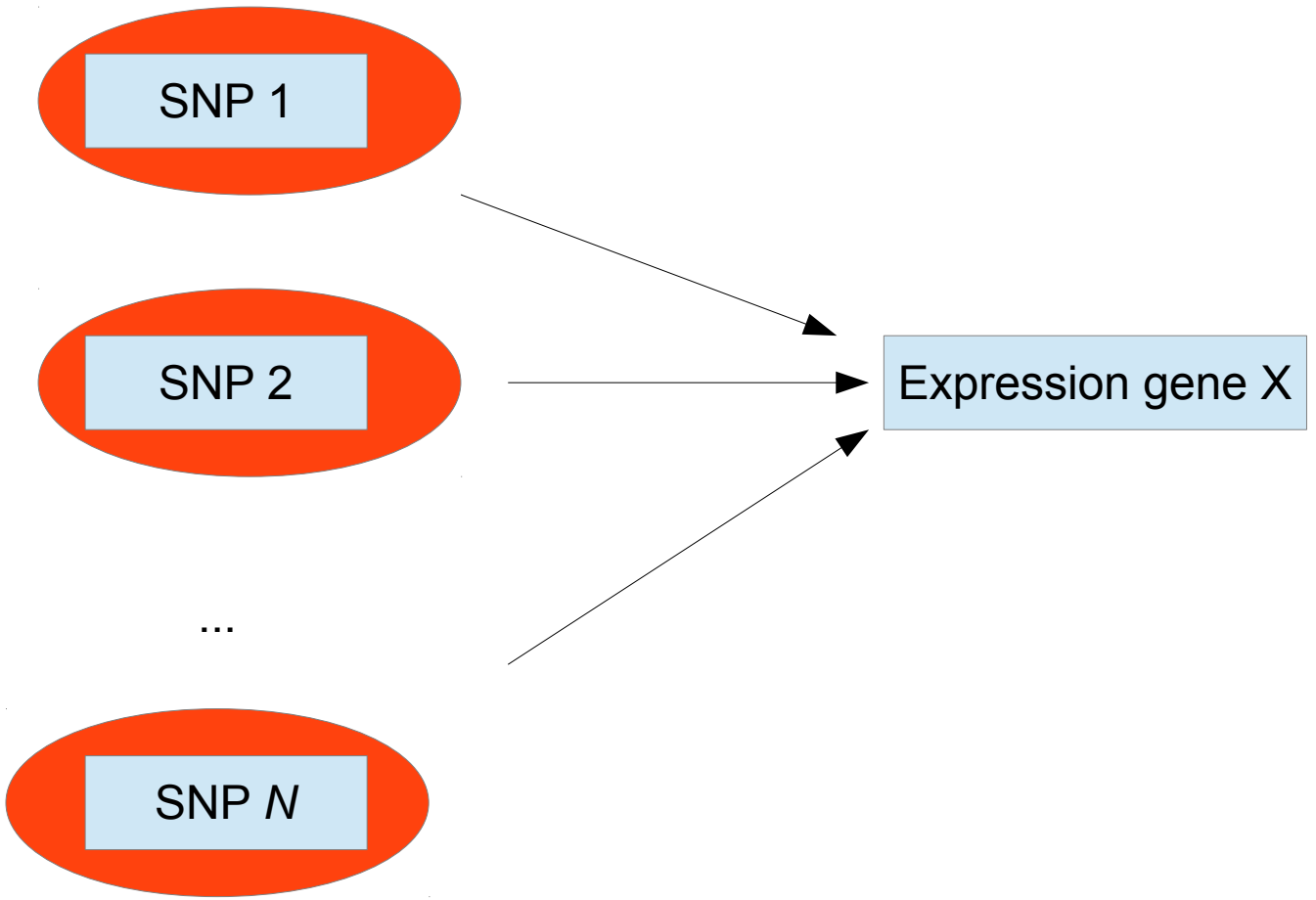
Gene expression gene	GWAS SNP	*P* GWAS SNP	Top eQTL SNP	*P* Top eQTL SNP	LD	*P* GWAS SNP Conditional on top eQTL SNP
IL6R	rs4129267	3.5e-06	rs55668699	8.5e-08	0.198	0.024
ADAR	rs4129267	4.2e-05	rs9426829	8e-13	0.055	0.015
CAMKK2	rs1183910	9.5e-11	rs3794207	2.8e-248	0.042	0.72
OASL	rs1183910	4.9e-07	rs2259690	5.2e-30	0.395	0.038
NLRP3	rs12239046	2.7e-14	rs10733112	2.9e-31	0.227	0.039
KIAA0754	rs12037222	6.9e-14	rs4660603	2.3e-24	0.575	0.85
MACF1	rs12037222	1.4e-11	rs645061	4.8e-20	0.064	7.7e-06
PABPC4	rs12037222	1.5e-10	rs72663521	2.4e-13	0.876	0.66
IL1RN	rs6734238	4.9e-06	rs315947	2.6e-28	0.11	0.34
SLC20A1	rs6734238	7.3e-05	rs11692751	7.7e-95	0.013	0.064
NRBP1	rs1260326	1.6e-06	rs1083864	8.4e-08	0.48	0.28
SNX17	rs1260326	2e-05	rs1728926	4e-13	0.403	0.74
IRF1	rs4705952	7.4e-10	rs10051145	1.7e-154	0.125	0.012
SLC22A4	rs4705952	4.7e-09	rs2631360	1.3e-61	0.087	0.27
FAM26F	rs6901250	1.3e-09	rs117361304	1.6e-258	0.008	3e-04

Low LD does not imply independent signals!!!

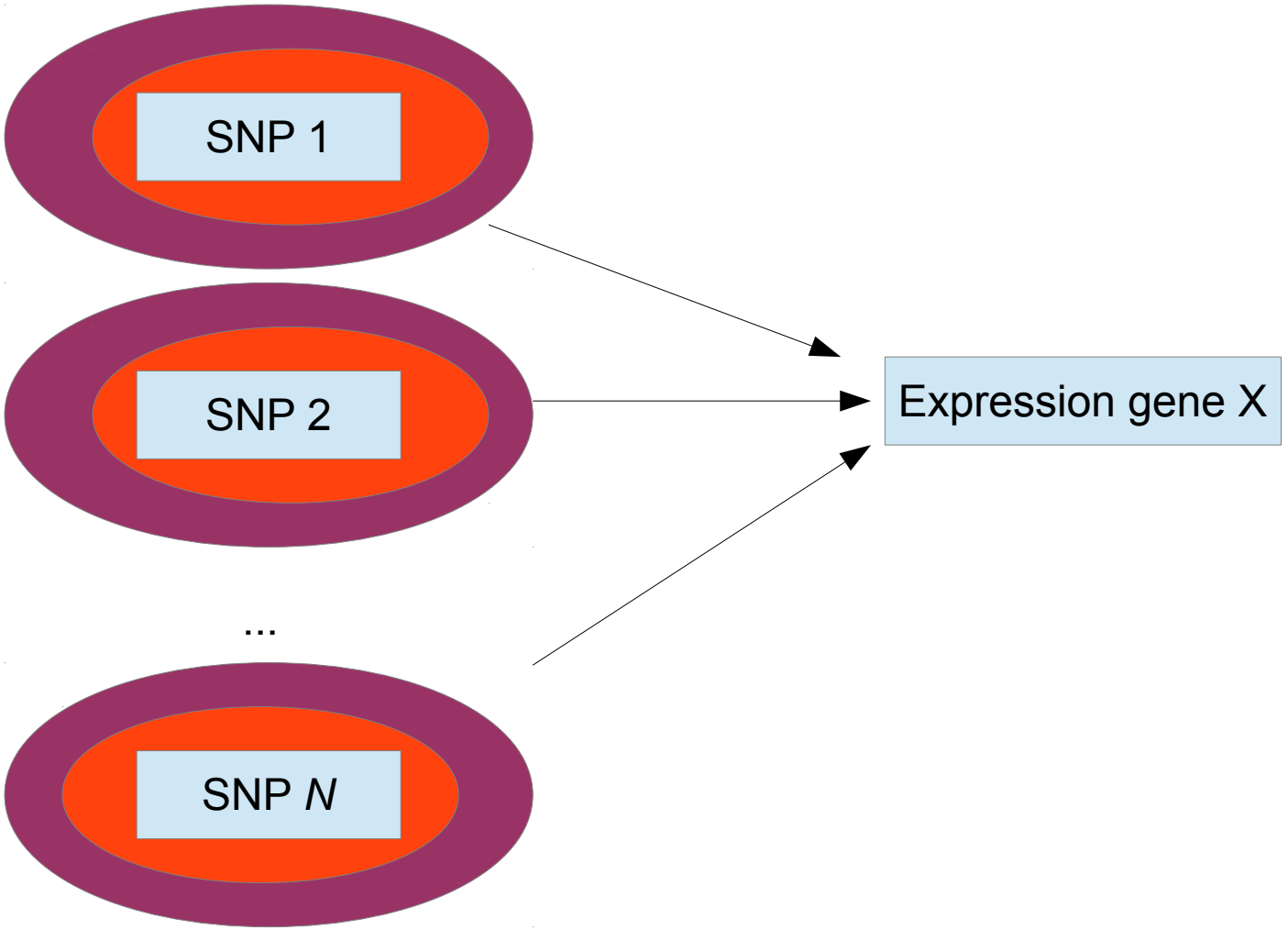
N independent SNPs associated with gene expression



N independent SNPs associated with gene expression
SNPs around the top SNPs(LD>0.8) are tagging the same locus

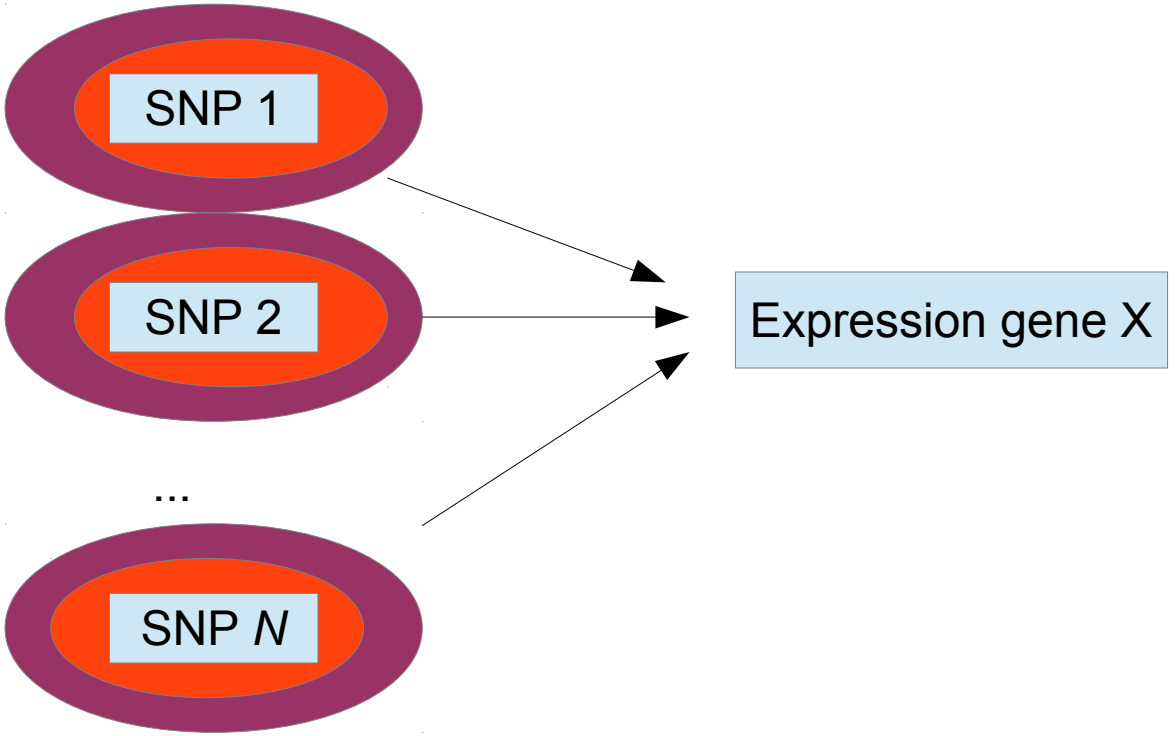


- *N* independent SNPs associated with gene expression
- SNPs around the top SNPs(LD>0.8) are tagging the same locus
- SNPs in low LD (<??) may still be associated with gene expression but are unlikely to tag the functional locus



GWAS SNP look up: compute LD between your GWAS SNP and SNP 1 - SNP *N*

N independent SNPs associated with gene expression
SNPs around the top SNPs (LD>0.8) are tagging the same locus
SNPs in low LD (<??) may still be associated with gene expression but are unlikely to tag the functional locus



How can we identity these 3 groups of SNPs?

'Conditional eQTL mapping':
For each gene expression measure:
-rerun association conditional on SNP 1, identify SNP 2
-rerun association conditional on SNP 1 and SNP 2, identify SNP 3
..
-rerun association conditional on SNP 1, SNP 2 and SNP *N,* identify SNP *N*+1

Conditional eQTL mapping

- rerun association conditional on SNP 1, identify SNP 2
- rerun association conditional on SNP 1 and SNP 2, identify SNP 3
..
- rerun association conditional on SNP 1, SNP 2 and SNP *N,* identify SNP *N*+1

# independent eQTLs	≥1	≥2	≥3	≥4	≥5	≥6	≥7	≥8	≥9	≥10	≥11	≥12	13
# genes	7120	2485	830	323	154	75	40	21	13	7	3	3	3

30% of the genes with an eQTL have at least 2 independent eQTLs
12% of the genes with an eQTL have at least 3 independent eQTLs

Conditional eQTL mapping

# independent eQTLs	≥1	≥2	≥3	≥4	≥5	≥6	≥7	≥8	≥9	≥10	≥11	≥12	13
# genes	7120	2485	830	323	154	75	40	21	13	7	3	3	3

Linear model for HLA-C expression:

	Estimate	Std. Error	t value	Pr(>|t|)
SNP1	-0.10358	0.01545	-6.703	2.27e-11 ***
SNP2	1.07836	0.01597	67.532	< 2e-16 ***
SNP3	-0.12776	0.01571	-8.135	5.18e-16 ***
SNP4	0.15256	0.01322	11.539	< 2e-16 ***
SNP5	-0.18800	0.02023	-9.291	< 2e-16 ***
SNP6	0.45120	0.05504	8.198	3.09e-16 ***
SNP7	0.15734	0.02975	5.289	1.28e-07 ***
SNP8	-0.37238	0.05400	-6.896	6.04e-12 ***
SNP9	-0.18869	0.03403	-5.544	3.11e-08 ***
SNP10	-0.23124	0.04467	-5.176	2.35e-07 ***
SNP11	0.12676	0.02460	5.153	2.67e-07 ***
SNP12	-0.14983	0.02579	-5.811	6.62e-09 ***

R-squared: 0.7491

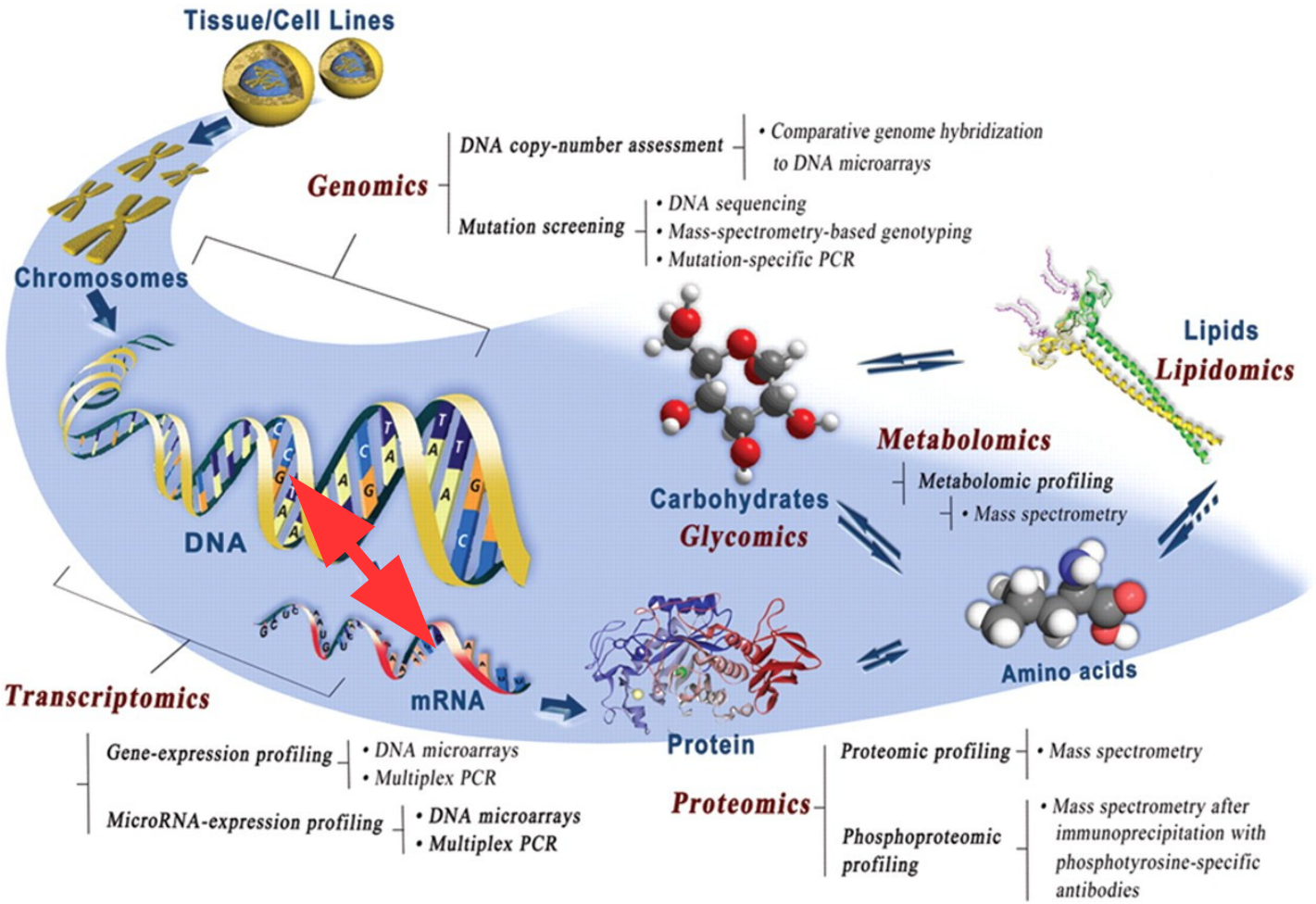
Conditional eQTL mapping

# independent eQTLs	≥1	≥2	≥3	≥4	≥5	≥6	≥7	≥8	≥9	≥10	≥11	≥12	13
# genes	7120	2485	830	323	154	75	40	21	13	7	3	3	3

From the 2485 SNPs identified in the analysis conditional on the top eQTLs, 17% were not identified in the unconditional analysis.

We will provide an eQTL catalog containing conditional eQTL results (up to conditioning on 12 SNPs)

Functional genomics: conditional eQTL analysis may be needed!



http://www.med.unc.edu/pgc/downloads (PGC website, NTR and NESDA eQTL results)
http://www.ebi.ac.uk/Tools/geuvadis-das/ (1000 Genomes RNA-seq in LCL DOWNLOADABLE!)
http://genenetwork.nl/bloodeqtlbrowser/ (meta eQTL analysis, whole blood)
http://snipa.helmholtz-muenchen.de/snipa/ (collection of catalogs, including eQTL and GWAS)
http://www.braineac.org/ (Brain eQTL database)
Conditional eQTL database (to be announced)