The opportunities of epigenomic research using UK Biobank data

Elin Grundberg

Associate Professor, Center for Pediatric Genomic Medicine, Children's Mercy Kansas City, University of Missouri School of Medicine

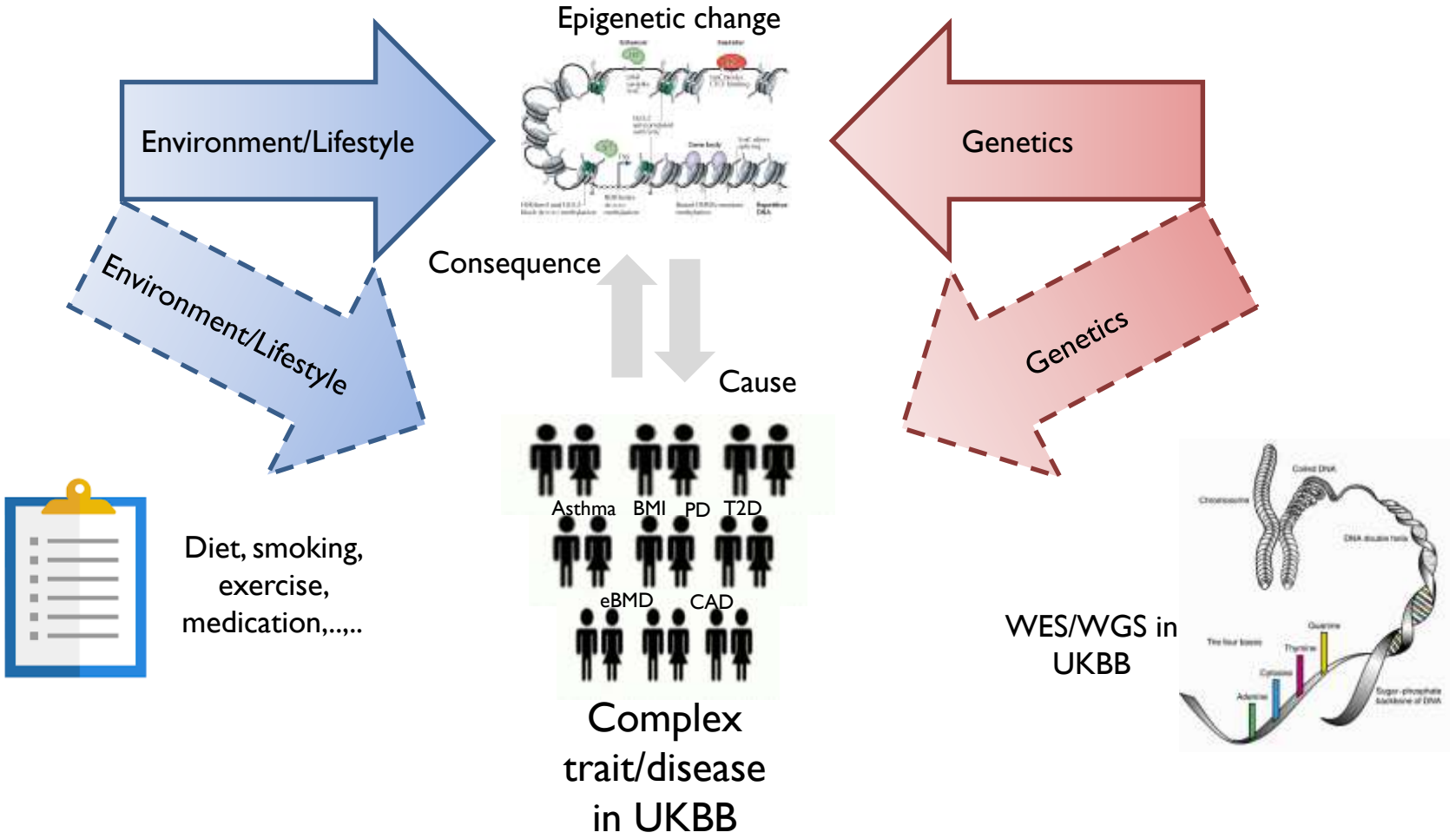
Adjunct Professor, McGill University

McGill

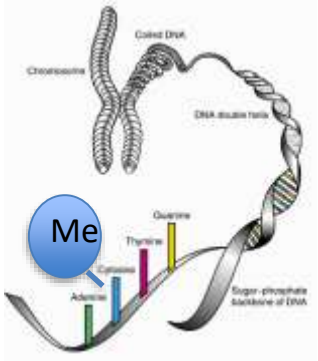
Children's Mercy KANSAS CITY

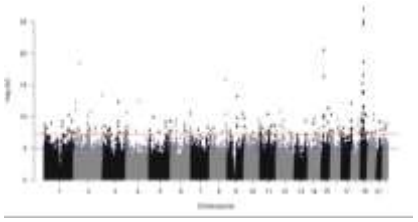
INTRODUCTION

Epigenetics connect environment and genetics to phenotypes and diseases



Epigenome-wide association studies





Study cohort

- Population-based
- Case/control

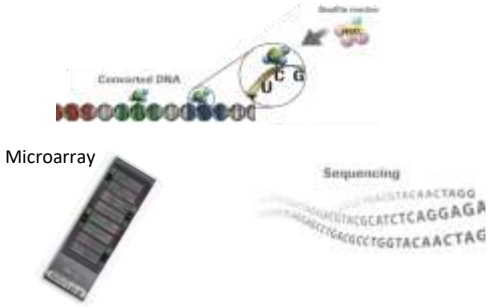
Epigenetic profiling

- Targeted
- Genome-wide

Epigenome variation associated with disease trait

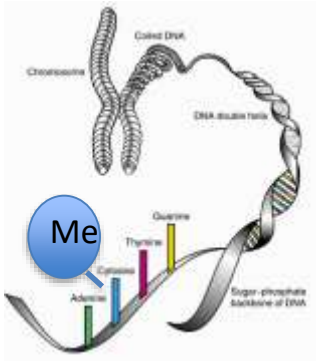
- Causal
- Reactive

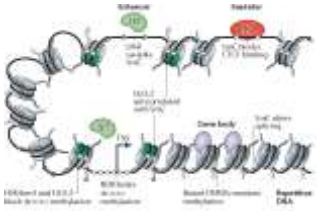
DNA methylation: commonly used epigenetic trait in large-scale disease studies



1. Robust and reproducible assays involving bisulfite conversion, automated solutions

2. Quantitative trait in bulk tissue (0-100%), single-base pair resolution, up to 30M measurable CpGs



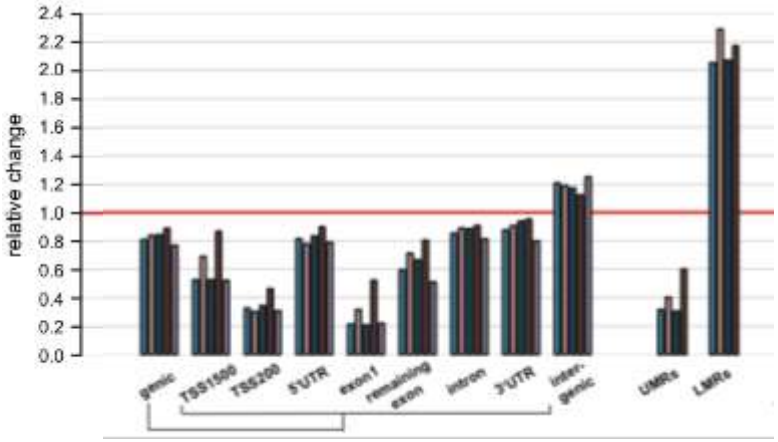


3. Stable cellular (DNA) phenotype, predictive of regulatory elements (e.g. enhancers vs. promoters)

VARIATION OF EPIGENOMES IN ASSAYS

DNA methylation variation is region-dependent

WGBS in TwinsUK (blood and adipose)

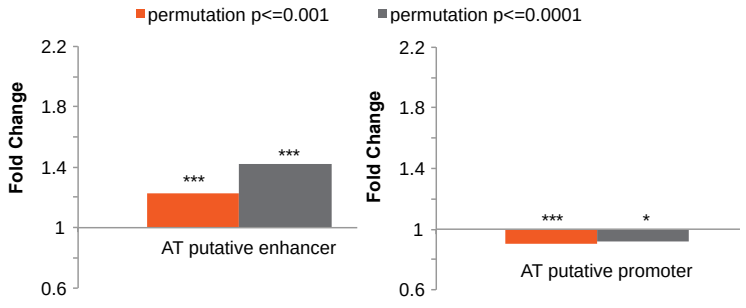


Promoter region mostly invariable and unmethylated

- Biased towards promoter regions on arrays (450K/EPIC)

Busche et al, Genome Biology 2015

Lipid EWAS (adipose)



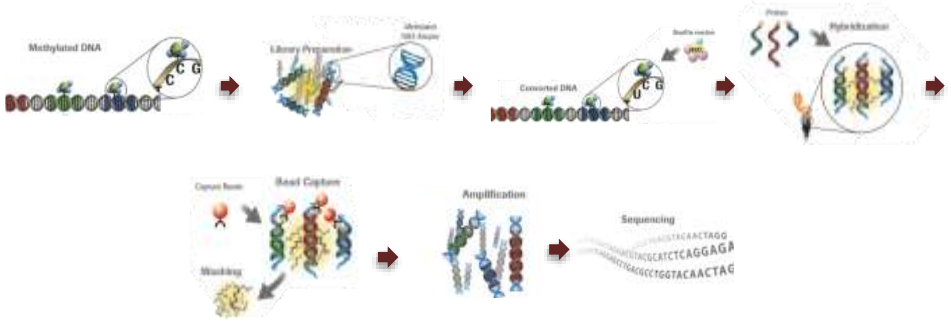
Disease-relevant CpGs are mapping to distal regulatory elements (enhancers)

- Underrepresented on arrays, only 50% of enhancers that are active in whole-blood covered (EPIC)

Allum et al, Nat Commun 2014

Interrogation of functional methylome for disease association studies

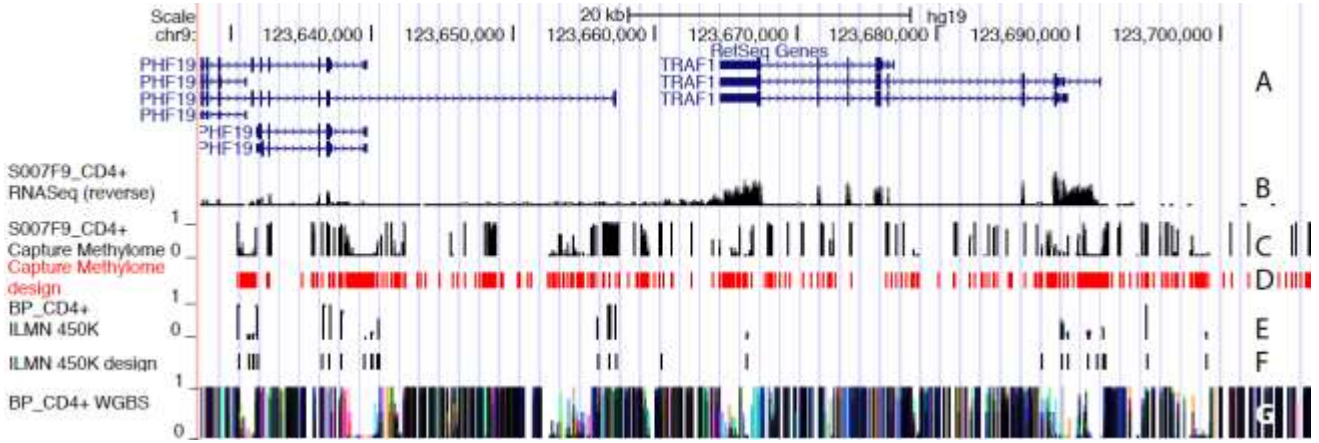
Implementation of MethylC-capture sequencing (MCC-Seq)



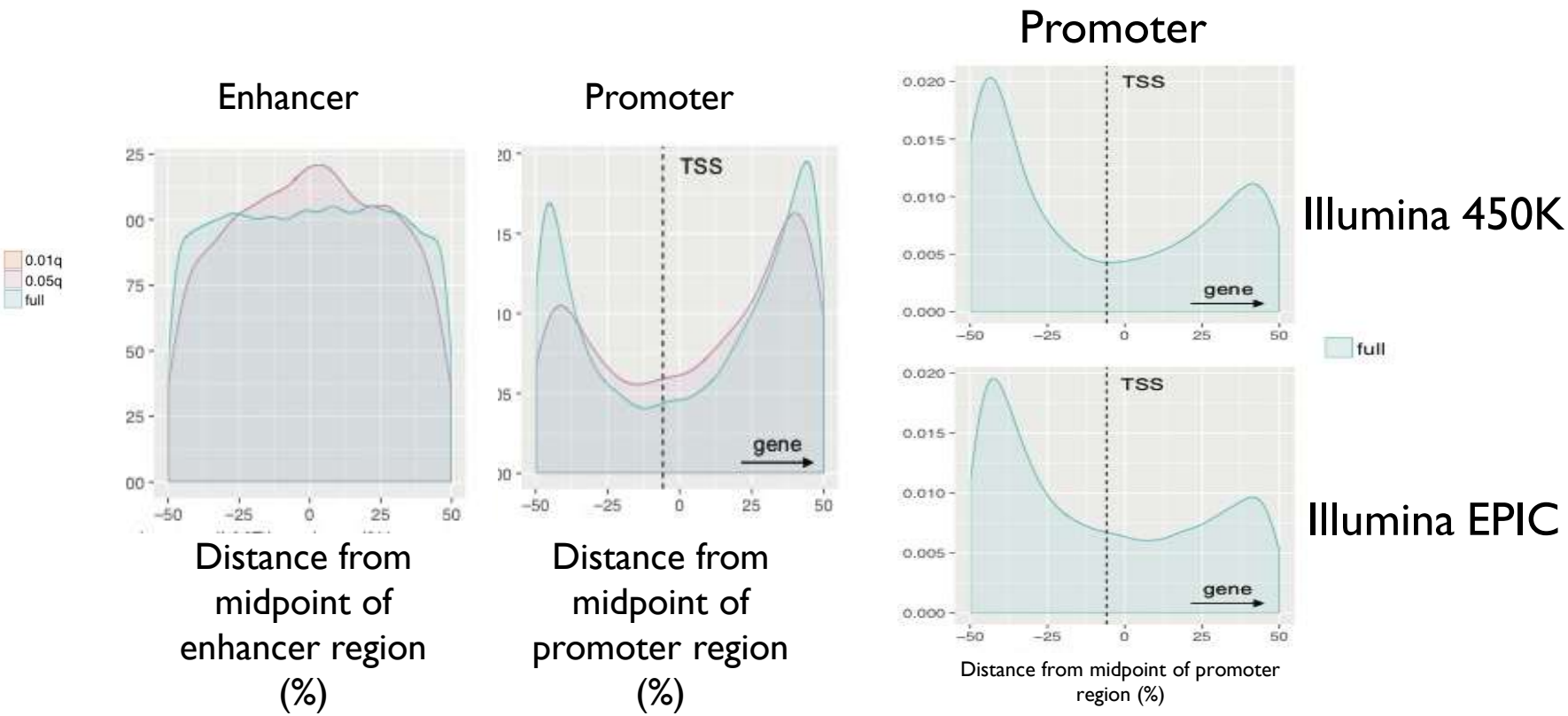
Selective enrichment of regulatory elements allows for multiplexing, increased coverage over regions of interest while reducing sequencing costs (Roche SeqCap Epi)

Allum et al, Nat Commun 2014

Blood Panel Design	Coverage
Regulatory Regions	Immune/circulating leukocyte-specific (DHS, ChIPseq, LMR/UMR WGBS) ~4.5M CpGs
Array content	482,421 CpGs (Illumina 450K)
GWAS SNPs	common SNPs and proxies (r^2>0.8) linked to autoimmune diseases ~30K CpGs



Disease epigenetic variants preferentially map *within* regulatory elements

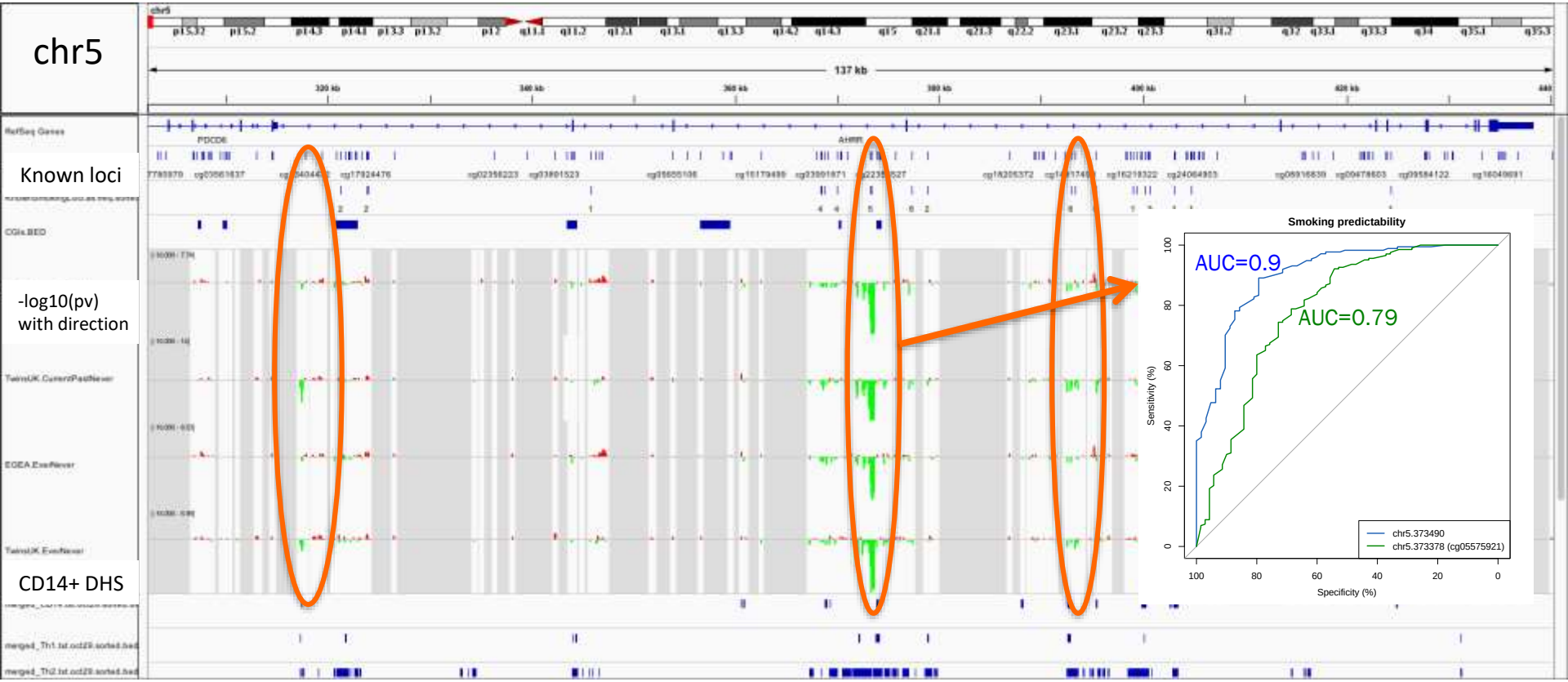


Allum et al, submitted 2018

VARIATION OF EPIGENOMES IN DISEASE STUDIES

Dense coverage allows fine-mapping of epigenetically regulated (disease) loci

Cigarette smoking impacts DNA methylation pattern in regulatory elements



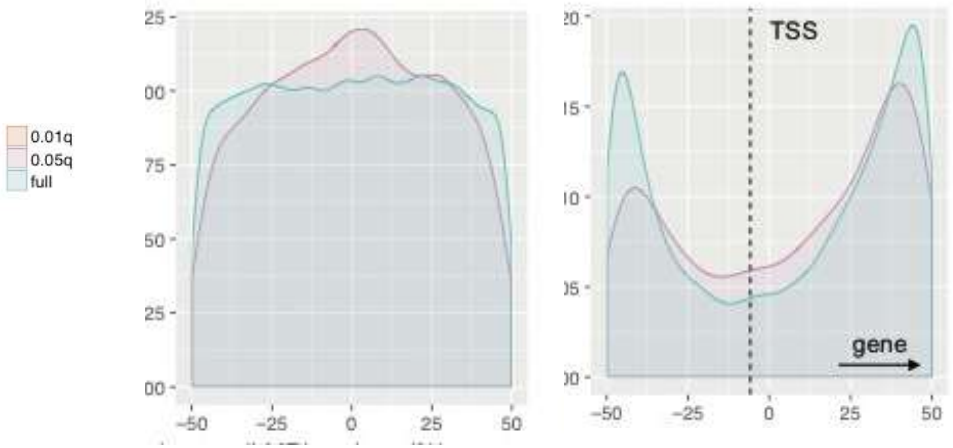
- Addition to previous reported loci, multiple regions are observed using MCC-Seq
- Fine-mapped CpG shows improved predictive value over known loci (AUC=0.9 vs. 0.79)

Shao et al, unbublished 2018

Disease epigenetic variants preferentially map *within* regulatory elements

Enhancer

Promoter



Distance from midpoint of enhancer region (%)

Distance from midpoint of promoter region (%)

Allum et al, submitted 2018

Disease-CpG in regulatory elements - discovery

↓

Disease-CpG – replication

↓

TFBS analysis

Epigenetic variants identify disease-associated regulatory pathways

Rank 1: Liver receptor homolog-1 (LRH-1) or NR5A2, p=1.00E-4

Rank 2: Glucocorticoid receptor (GR), p=1.00E-3

Rank 3: Retinoic Acid Receptor Gamma (RARg), p=1.00E-3

Rank 4: Peroxisome proliferator-activated receptor gamma (PPARg), p=1.00E-3

Allum et al, submitted 2018

Epigenetic variants identify disease-associated regulatory pathways

Rank 3: Retinoic Acid Receptor Gamma (RARg), p=1.00E-3

Disease-CpG in regulatory elements - discovery

↓

Disease-CpG – replication

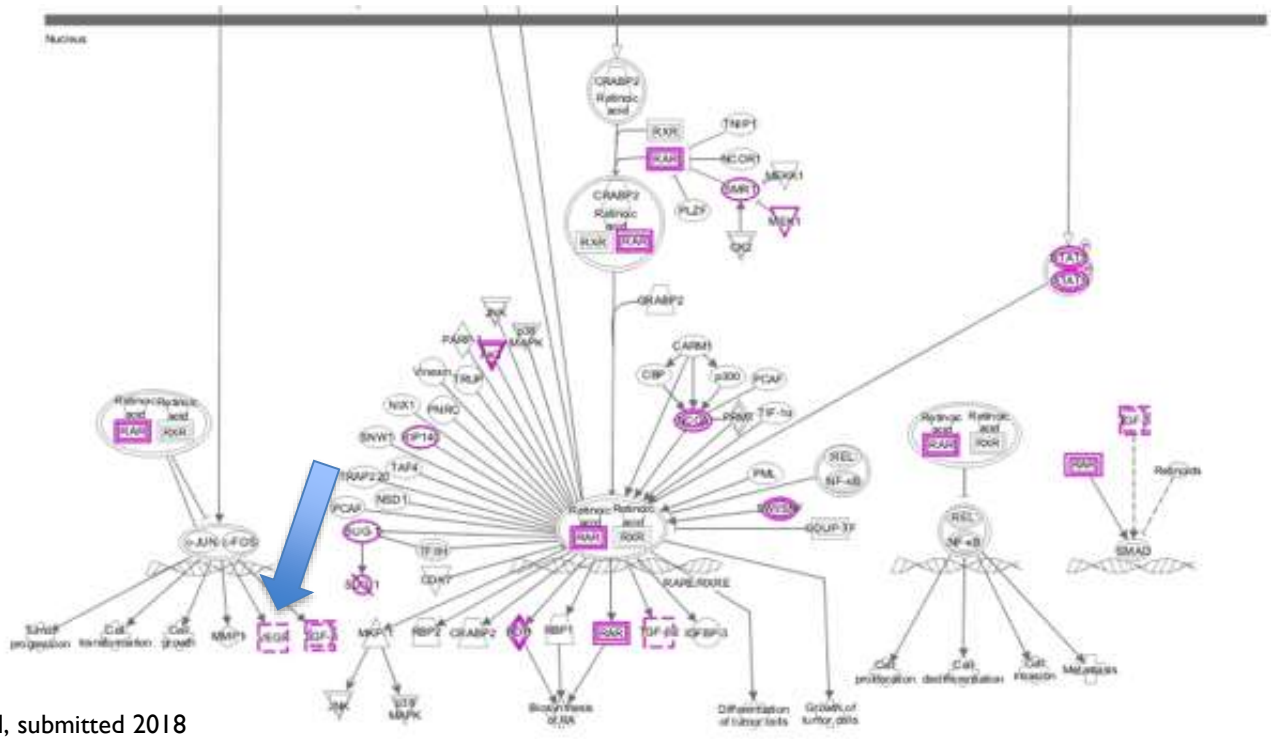
↓

Disease-CpG linked to Gx

↓

Pathway analysis

RAR activation

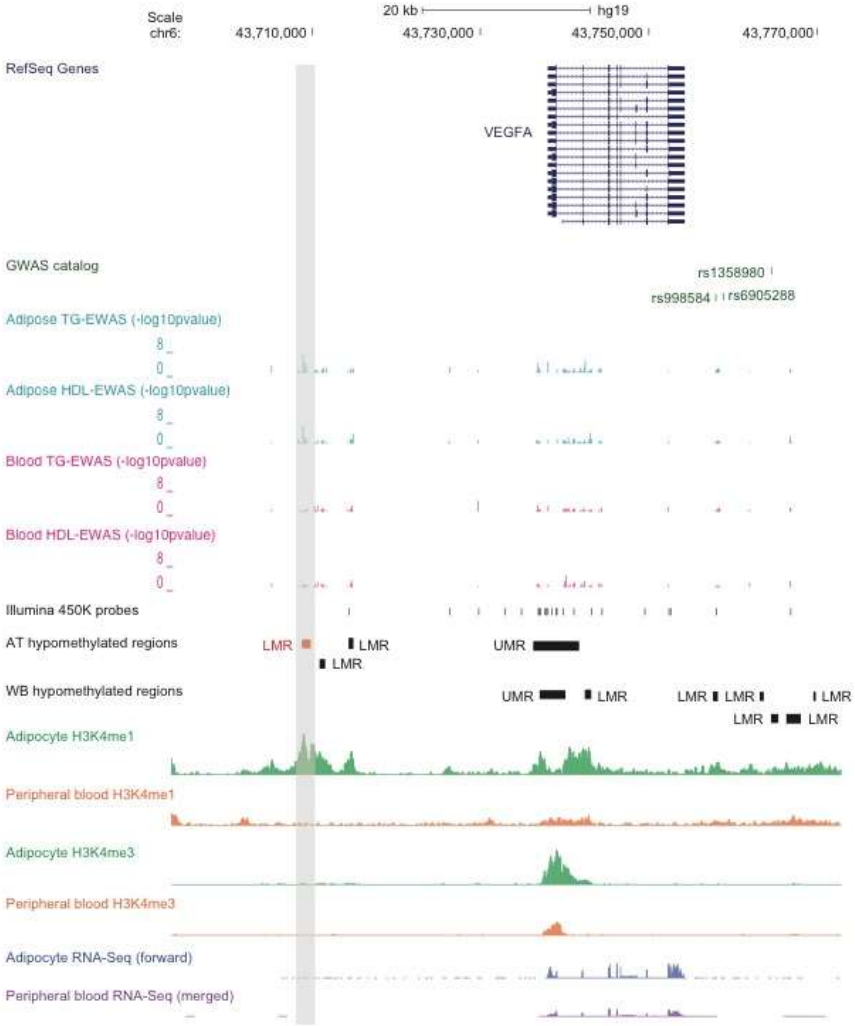


Allum et al, submitted 2018

Epigenetic variants identify non-genetic "additive" effect of disease loci

Vascular endothelial growth factor (VEGFA)

- Disease variant from large GWAS (CAD) in UKBB
- Disease-CpG (HDL, TG) mapping to distal regulatory element (enhancer), not covered by array (450k)
- Integration of genetic information show no association, potential combined (environmental) epigenetic and genetic risk



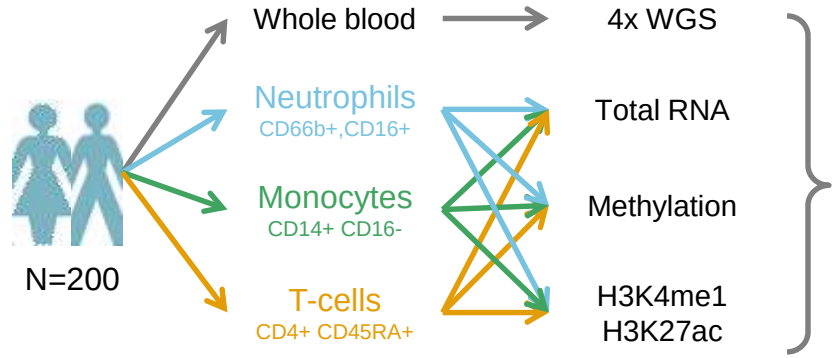
Allum et al, submitted 2018

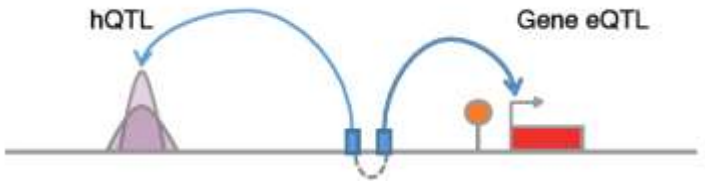
POPULATION EPIGENOMICS AND GWAS

Co-lozalization of molecular and disease traits

BLUEPRINT human diversity panel (Chen et al. Cell 2016)

Dataset





Disease SNP

McGill EPIGENOMICS MAPPING CENTRE

wellcome sanger institute

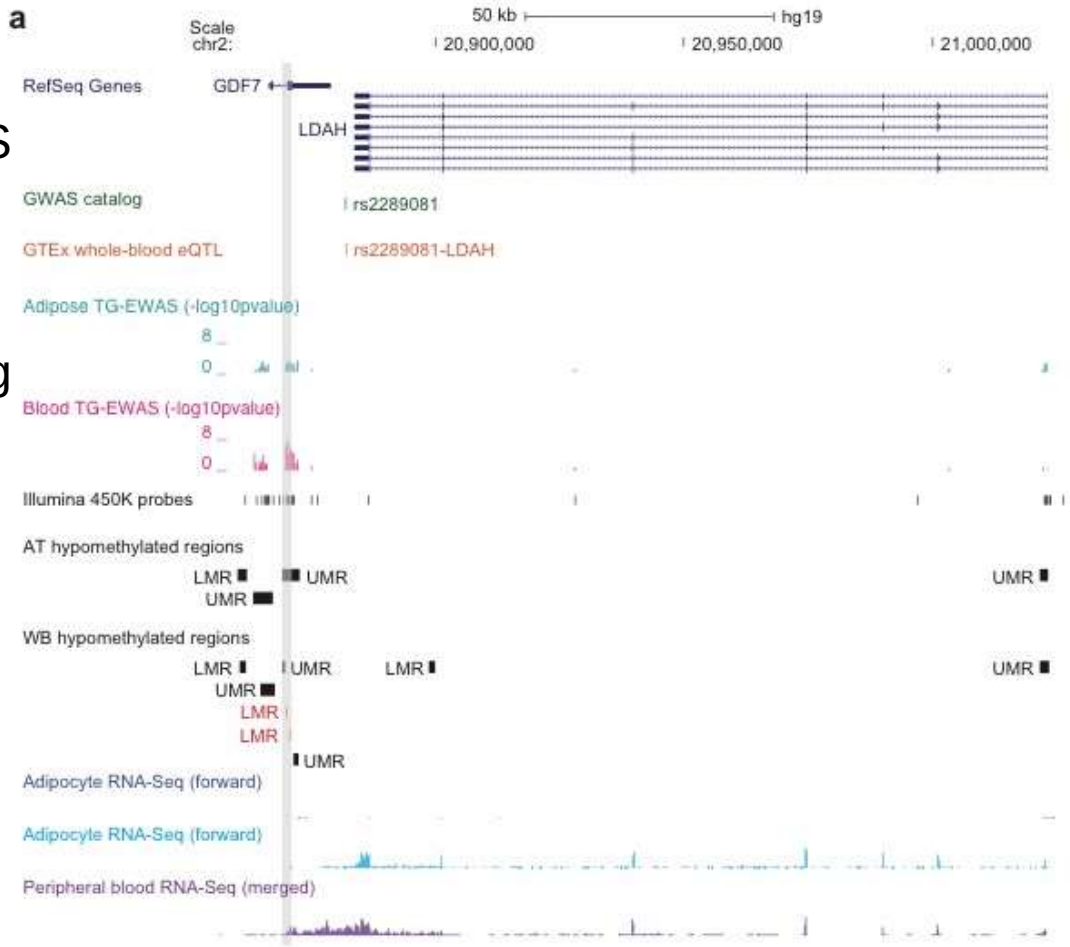
- Gene expression accounts for only a minority of disease links
- Epigenetics allow further insight of biological validity to mapped traits beyond traditional eQTL studies

QTL	N overlap ($r^2 \geq 0.8$)	Disease colocalized (%)	FE relative to eQTL
Gene	168	91 (54%)	1
PSI	86	47 (55%)	1.01
H3K27ac	189	117 (62%)	1.14
H3K4me1	190	102 (54%)	0.99
Meth	439	300 (68%)	1.26
All	1072	657 (61%)	-

Functional annotation of GWAS loci in blood tissue

GDF7-LDAH

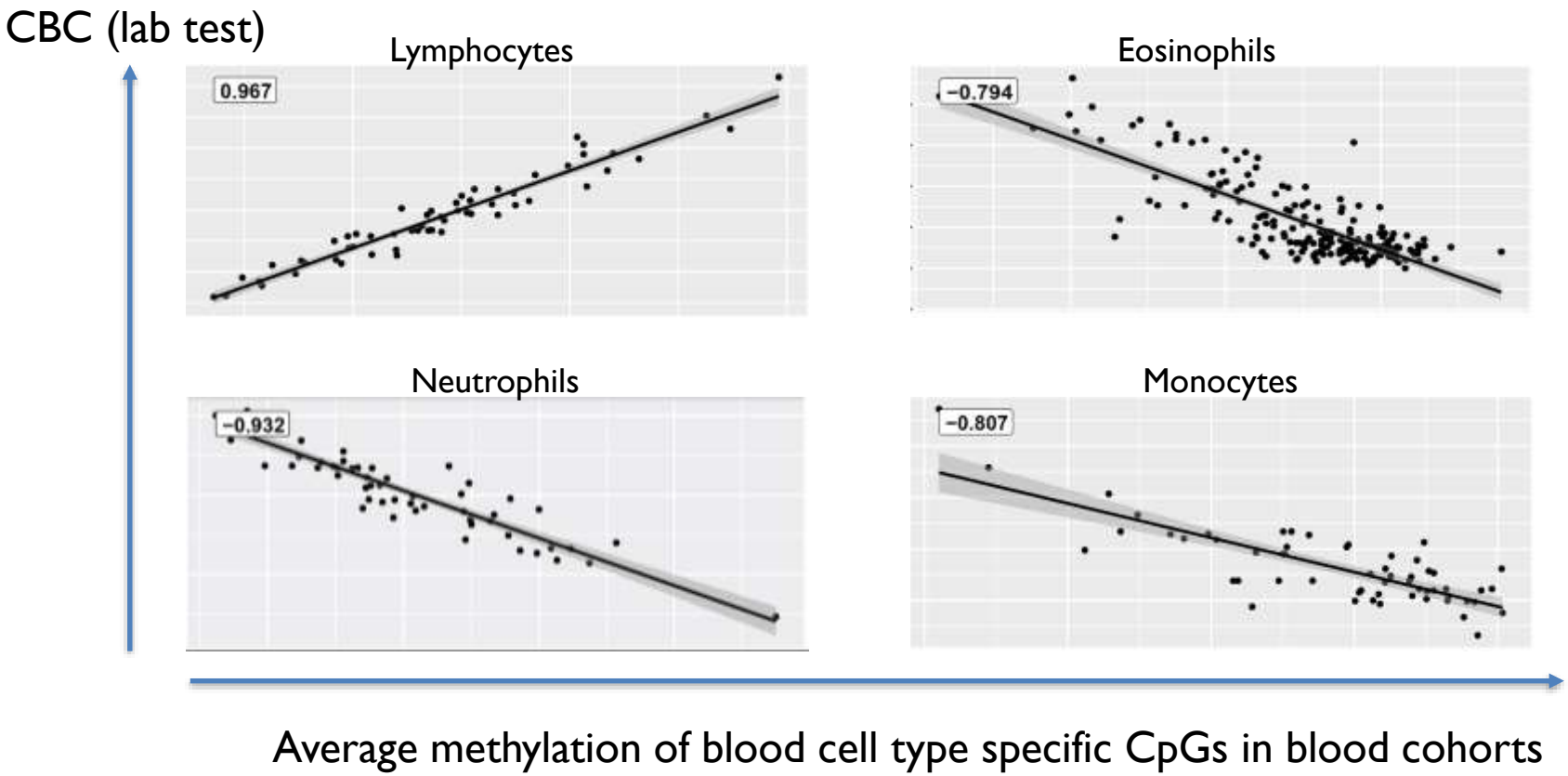
- Disease variant from large GWAS (BP) in UKBB
- Disease-CpG (HDL, TG) mapping to distal regulatory element (enhancer) in blood tissue only
- Integration of genetic information show colocalization (metQTL) of GWAS SNP
- GTEx resource confirms target gene (LDAH)



TISSUE HETEROGENEITY

Tissue architecture (purity) needs to be accounted for

Supervised prediction of cell counts in 1000 MCC-Seq blood samples based on CpGs that are specifically hypomethylated or hypermethylated in a cell-type specific manner (identified by wgbs of 19 purified cell types)



Cheung et al, unpublished 2018

Summary: why epigenetics?

- Discover regulatory pathways underlying disease/trait
  - Caused by environment, lifestyle or genetics
  - High resolution NGS-based approaches yield additional unique insight into distinct regulatory programs
- Annotate putative functional consequences of genetic variants associated with disease/trait
  - Dense coverage of methylation variation in regulatory elements captures majority of regulatory disturbances caused by non-coding disease SNPs
- Allow building combined epigenetic and genetic risk scores for predictions of later life outcomes
  - Dependent on large well-phenotyped (longitudinal) cohorts

Acknowledgment

McGill
Tomi Pastinen
Mark Lathrop
Guillaume Bourque
Fiona Allum
Elodie Boulier
Xiaojian Shao
Marie-Michelle Simon
Jinchu Vijay
Albena Pramatarova
Tony Kwan
Bing Ge
Warren Cheung
Alfredo Staffa

Roche NimbleGen
Dan Burgess
Todd Richmond
Jennifer Wendt

U Laval
Marie-Claude Vohl
Andre Tchernof
Frédéric Guénard
Marie-Frederique Gauthier
Mélissa Pelletier

UKBB
Rory Collins
Paul Matthews
Jimmy Bell
Jemma Hopewell

TwinsUK/MuTHER
Tim Spector
Panos Deloukas
Mark McCarthy
Manolis Dermitzakis
Asa Hedman
Kerrin Small

Children's Mercy
Tom Curran (CSO)
Tomi Pastinen
Emily Farrow
Laurel Willig
Nhu Bui
Rebecca Biswell
Maggie Gibson
Erica DelAquila
Warren Cheung
Neil Miller
Brad Beldon
Suzy Herd
Steve Leeder
Jason Fraser
Carrie Vylidahl
Donna Pacicca

compute canada | calcul canada

CIHR IRSC
Canadian Institutes of Health Research | Instituts de recherche en santé du Canada

Fonds de recherche Santé Québec

INNOVATION.CA
CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION

Canada Research Chairs

Children's Mercy KANSAS CITY