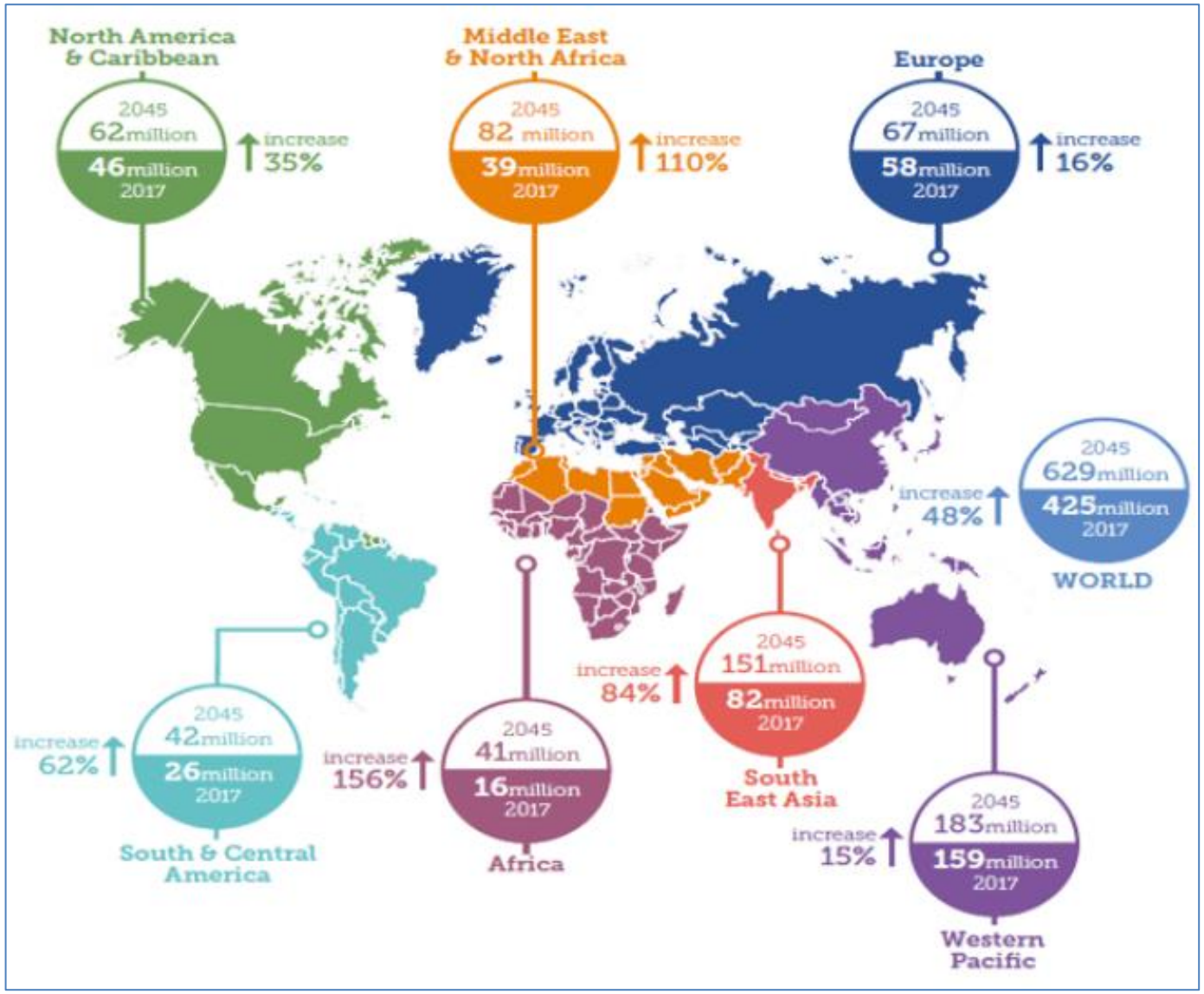
Advancing Precision Medicine for Diabetes and Cardiovascular disease

Prof John Chambers

Professor of Cardiovascular Epidemiology,
Lee Kong Chian School of Medicine, Nanyang Technological University

Chief Scientific Officer, PRECISE
PI, SG100K Study

Diabetes: the $1Trillion disease paradigm

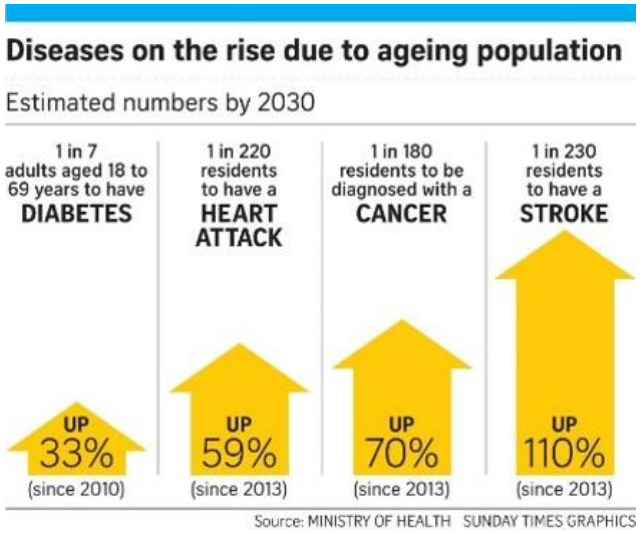


Importance

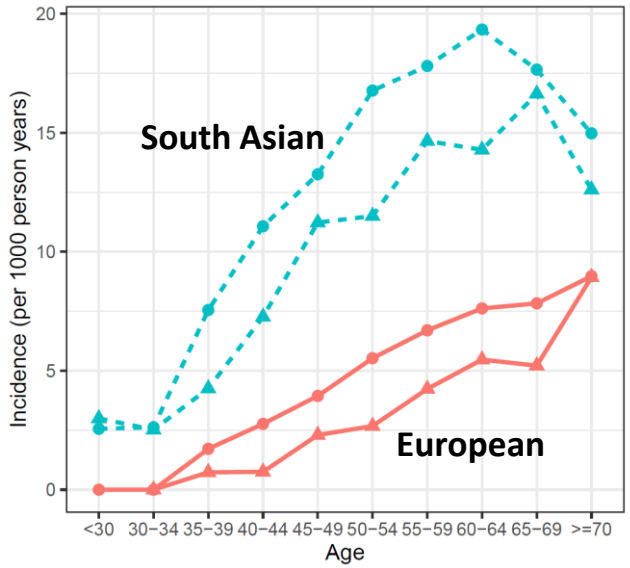
- Lifelong condition
- Affects 1/11 adults globally
- Multiple complications
- ~$988 billion 2021

Motivation

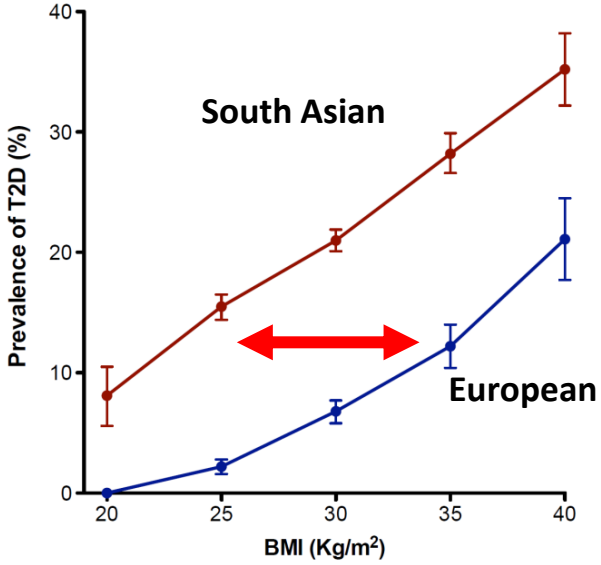
Importance of T2D in Singapore

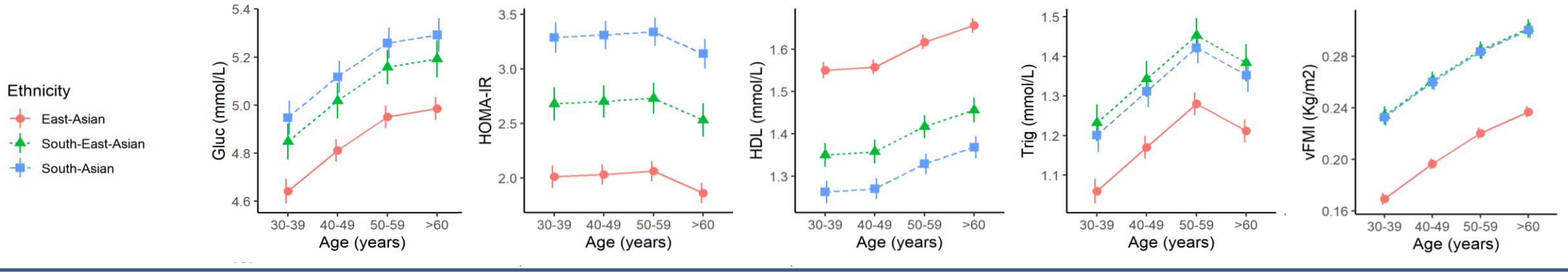


Unexplained T2D susceptibility



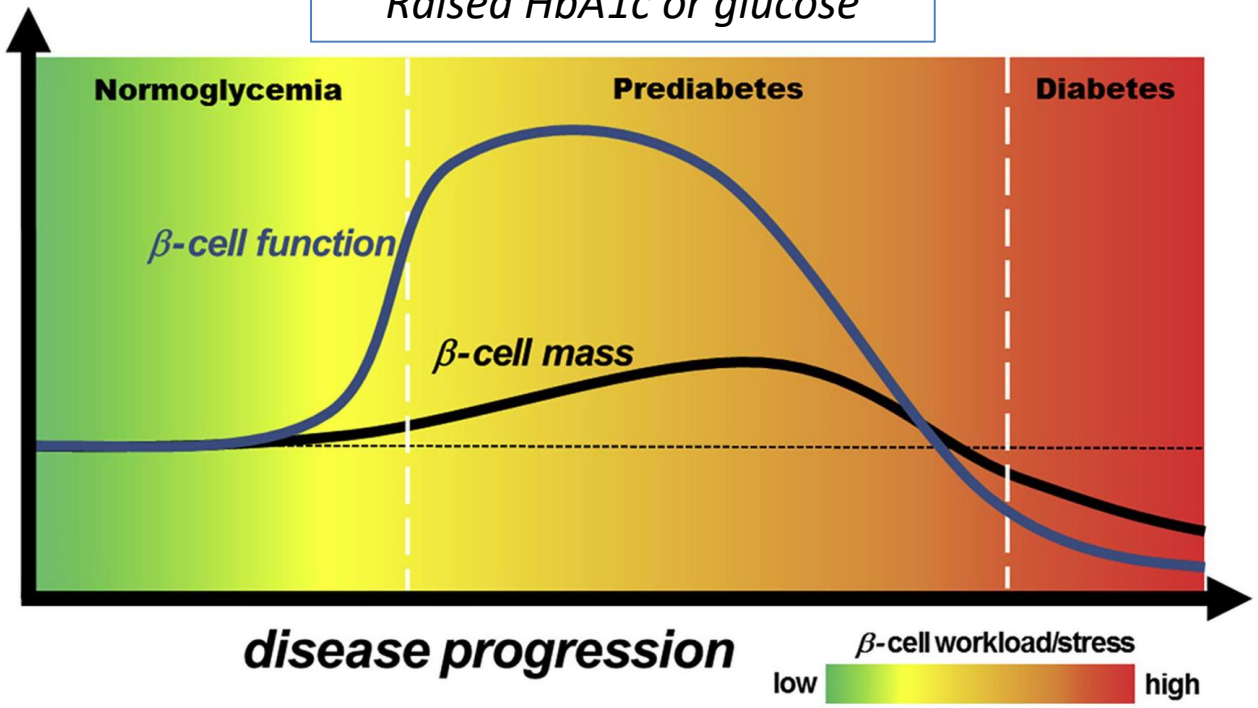
High risk not explained by adiposity





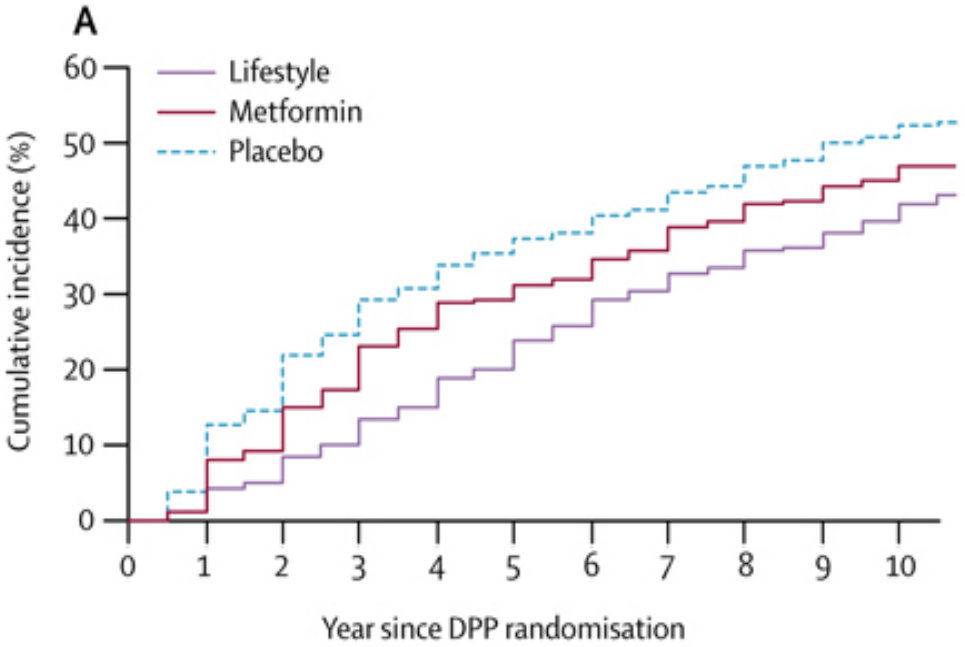
Progression to Diabetes

***Prediabetes**:*
Raised HbA1c or glucose



The problem with Prediabetes

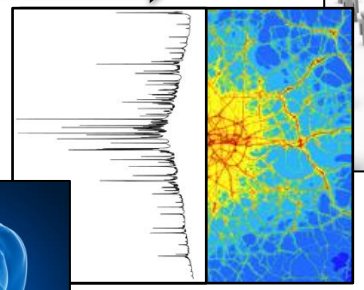
- Risk condition:
  - 2-fold increased risk of CVD
  - Microvascular disease (retinopathy, neuropathy, and nephropathy)
- High progression to T2D
  - 50% in Europeans over 10 years
  - 80% in Asians over 10 years
- Intervention modestly effective



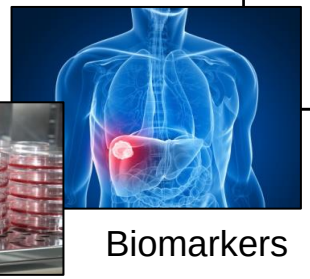
Population Science approach

Genes & Environment

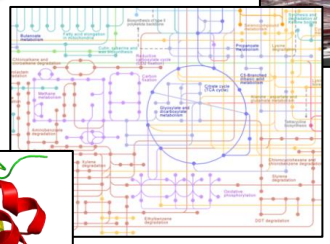
Population cohort



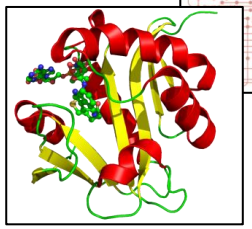
Molecular phenotyping



Biomarkers



Systems Biology



Pathways

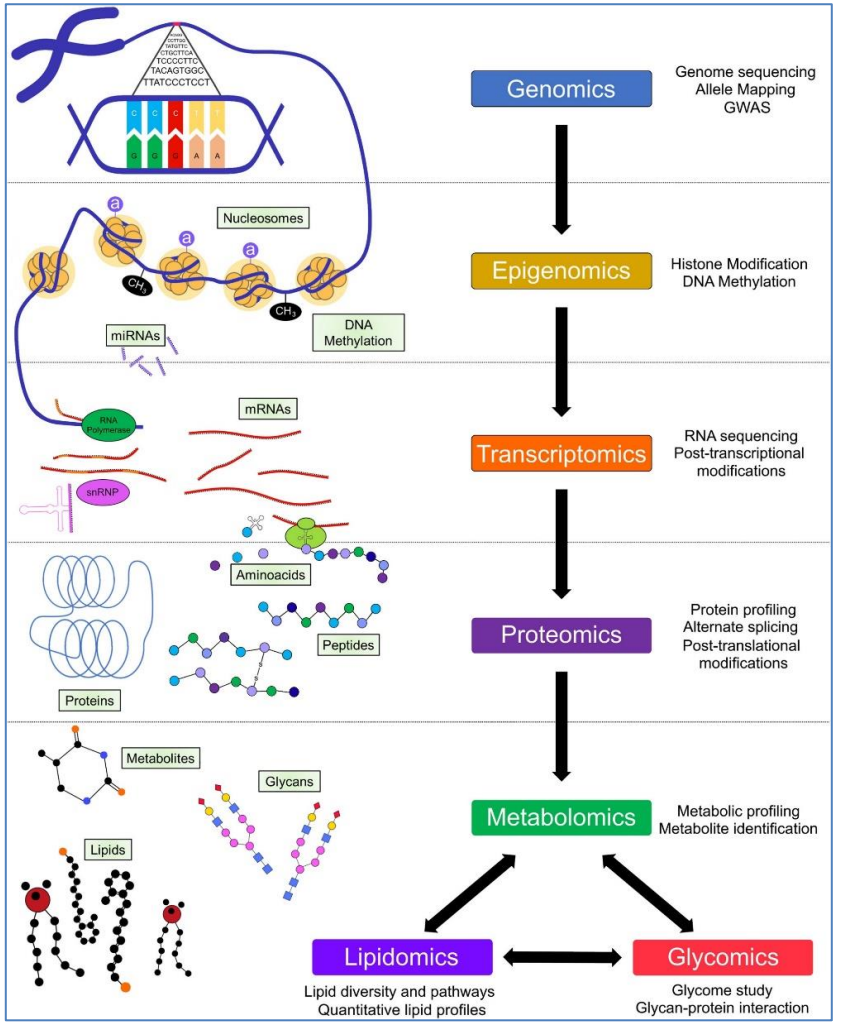
Molecular Insights

Linkage and Integration

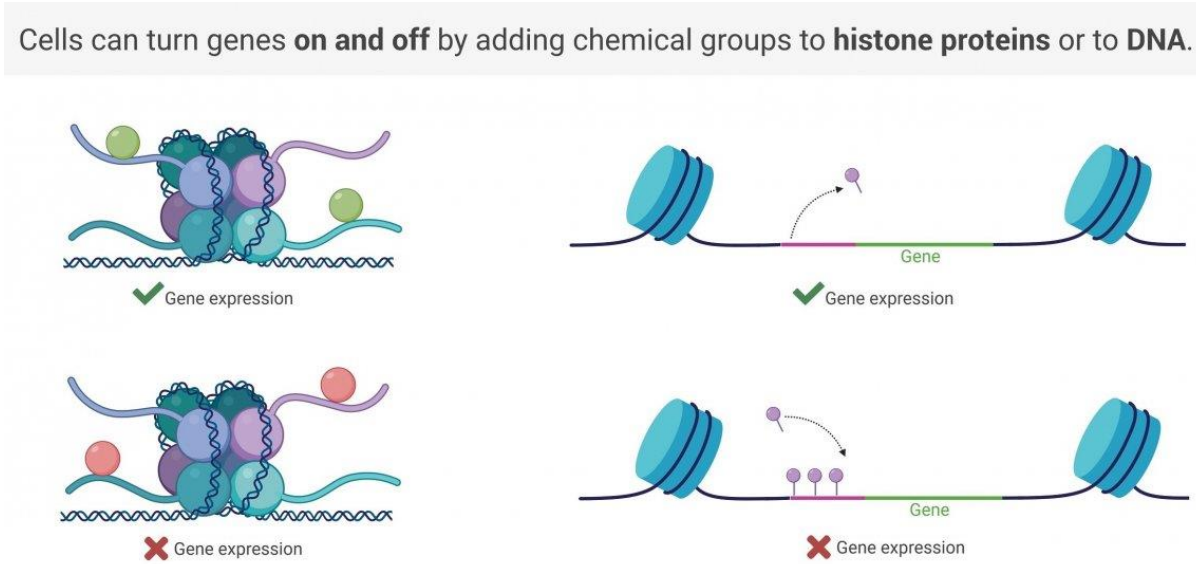
Precision medicine

Policy & Action

Molecular phenotyping

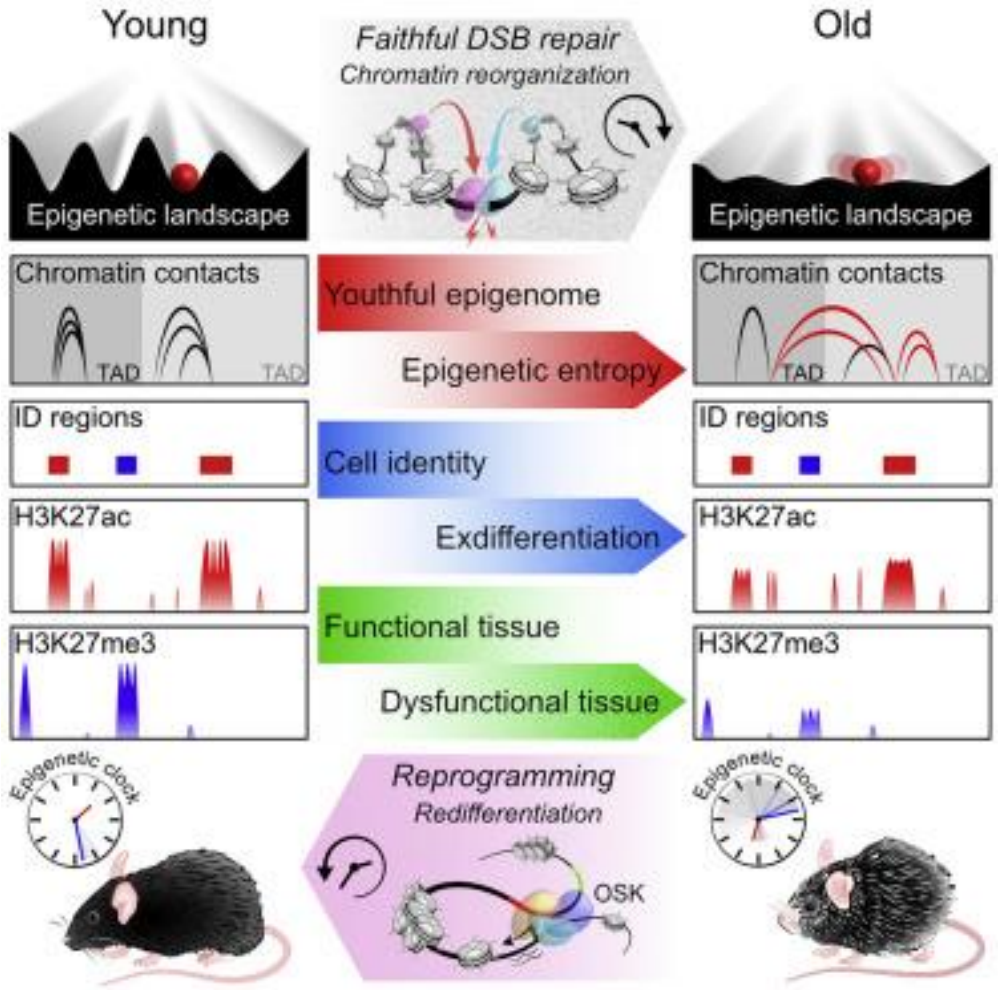


DNA methylation as a disease pathway



Why Methylation?

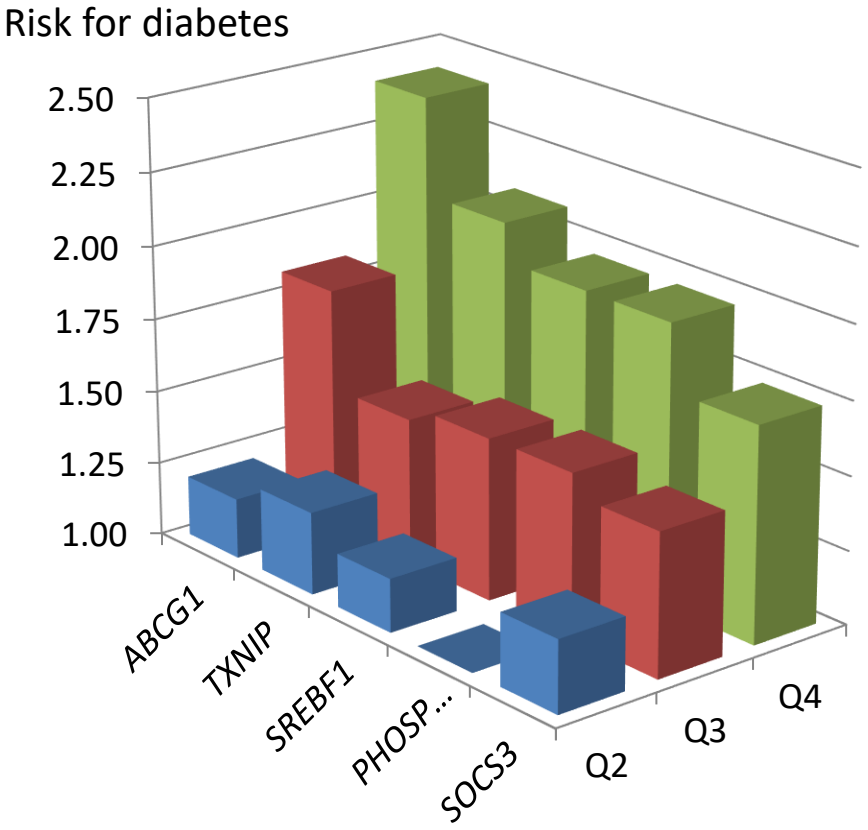
- Critical role in cellular programming and adaptation
- Proximal integration of gene, environment and behaviour
- Molecular heterogeneity: >30M CpG sites
- Implicated in chronic disease aetology



Yang, Sinclair and other Cell - Jan 2023

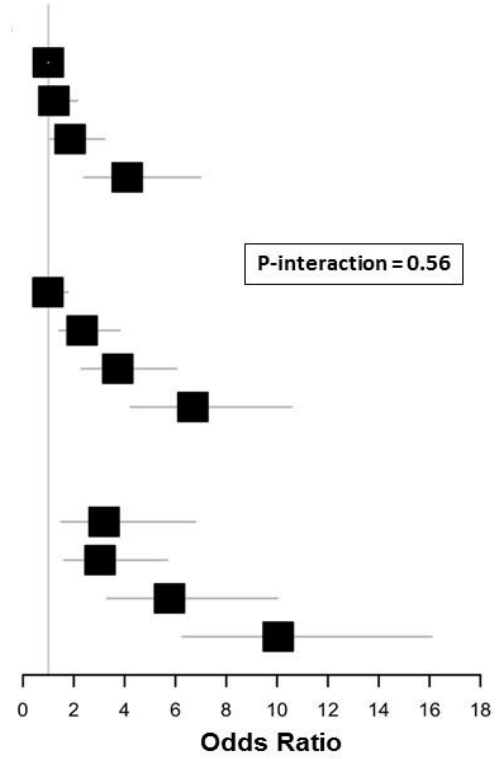
Prior work: DNA methylation signatures for T2D

Nested case-control study for T2D
N=2500, 450K Methylation array



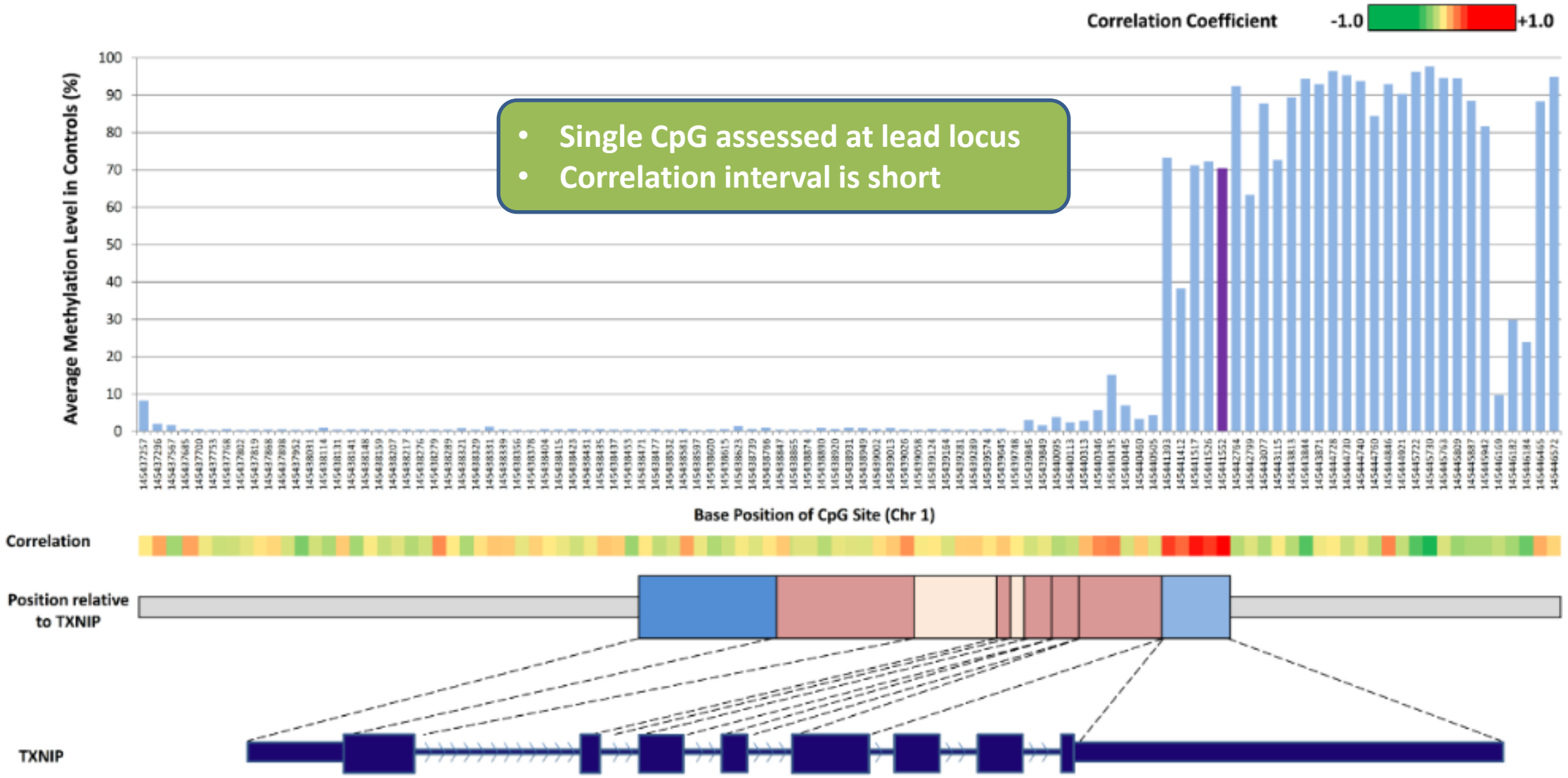
DNA methylation identifies metabolically unhealthy adiposity

	Controls/Cases	P	P-trend
Normal			
Q1	144/29	NA	3.85E-7
Q2	141/32	4.97E-1	
Q3	130/43	2.61E-2	
Q4	106/69	1.89E-7	
Overweight			
Q1	129/27	9.5E-1	5.66E-19
Q2	185/78	7.7E-4	
Q3	169/115	9.0E-8	
Q4	301/321	4.0E-16	
Obese			
Q1	27/17	2.5E-3	4.19E-7
Q2	50/28	5.2E-4	
Q3	59/61	5.1E-10	
Q4	149/255	7.9E-22	



Chambers et al (LDE2015, Nature 2017)

Insights from resequencing: Methylation arrays assess <3% of epigenome



Programme Goals

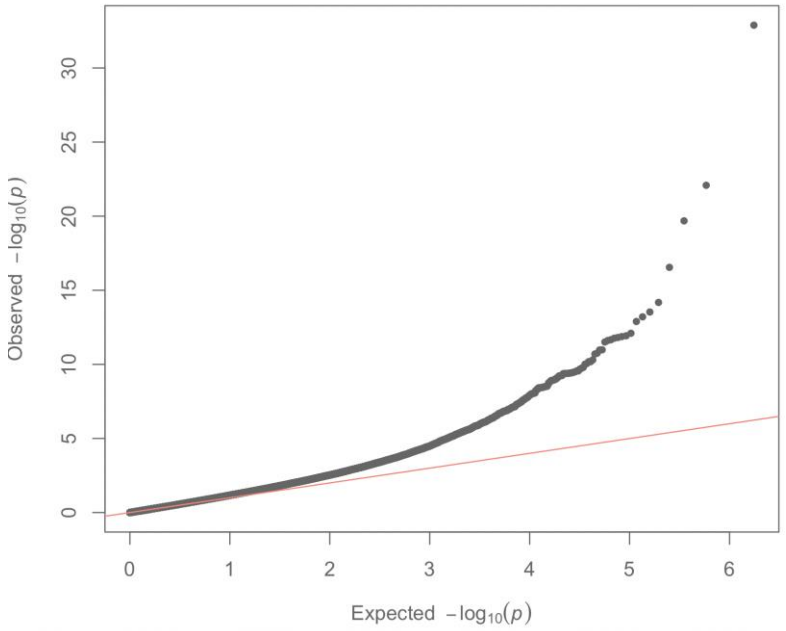
1. Extend discovery of methylation markers predicting T2D through
   a) Methylation array studies with larger sample size
   b) Methylation sequencing (whole methylome and targeted)
2. Functional genomic evaluation for molecular pathways
3. Determine predictive utility for future T2D

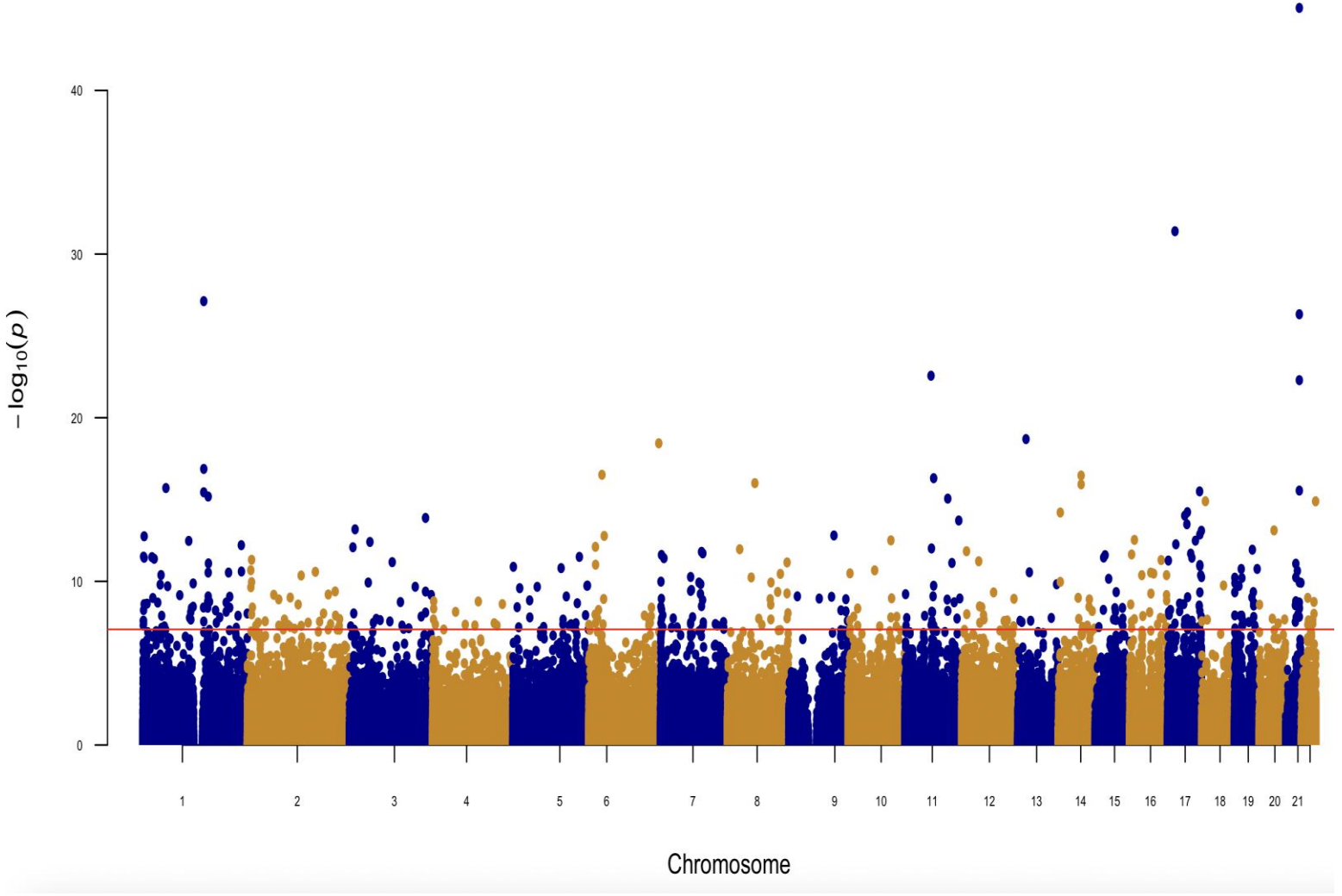
Study Samples

- Singapore Population Health Study (SPHS)/ Multi-Ethnic Cohort (MEC)
  - Cohort of ~50,000 adult Singaporeans and long-term residents from three major ethnic groups (Chinese, Malay and Indian)
  - EPIC array; 761 incident T2D cases, 743 controls

- LOLIPOP
  - Prospective population study of South Asian men and women
  - EPIC array
    - 1,000 incident T2D cases, 1,000 controls
  - 450K array (Chambers and Loh *et al* 2015)
    - 1,074 incident T2D cases, 1,590 controls

Model: T2D ~ Beta(QN) + Age + Gender + $WBC_{Houseman}$ + $PC1\text{-}30_{ctrl\text{-}probes}$ -> Meta-analysis





-> 420 EWAS signals; 314 discrete loci

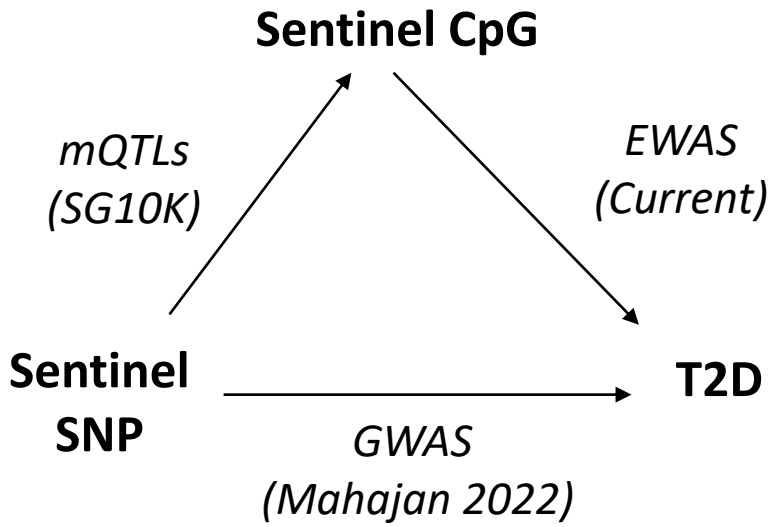
Leading associations

Marker	OR (95% CI)*	P	Chr	Position	Gene	Gene summary	Known?
cg06500161	1.18 [1.16, 1.21]	9.08E-46	chr21	43656587	ABCG1	Involved in macrophage cholesterol and phospholipids transport, and may regulate cellular lipid homeostasis	Y
cg11024682	1.15 [1.12, 1.17]	4.04E-32	chr17	17730094	SREBF1	Energy homeostasis: promotes glycolysis, lipogenesis, and adipogenesis	Y
cg19693031	1.07 [1.06, 1.09]	7.56E-28	chr1	145441552	TXNIP	Major regulator of glucose transport	Y
cg19758958	1.15 [1.12, 1.19]	2.72E-23	chr11	62319222	AHNAK	Regulates adipose tissue development via interaction with the SMAD1 protein and involved in metabolic homeostasis	N
cg19750657	1.08 [1.06, 1.10]	2.02E-19	chr13	38935967	UFM1	Increased expression linked to diabetes; Involved in the diabetic Iinflammatory response	Y

*per 1% change in methylation level

Are the methylation marks involved in disease aetiology?

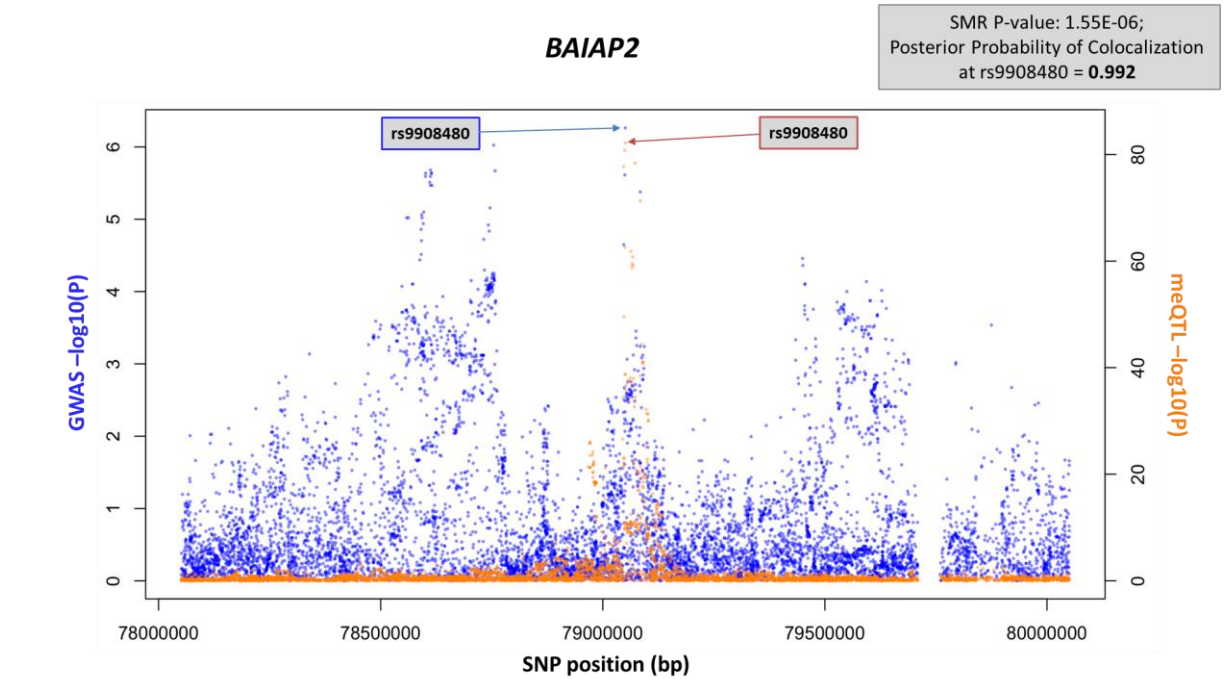
- 314: EWAS (meth-T2D) identifies Sentinel CpG
- SNP-CpG: Identifies Sentinel SNPs for the CpGs
- SNP-T2D: T2D GWAS result for the Sentinel SNPs *(Mahajan 2022)*

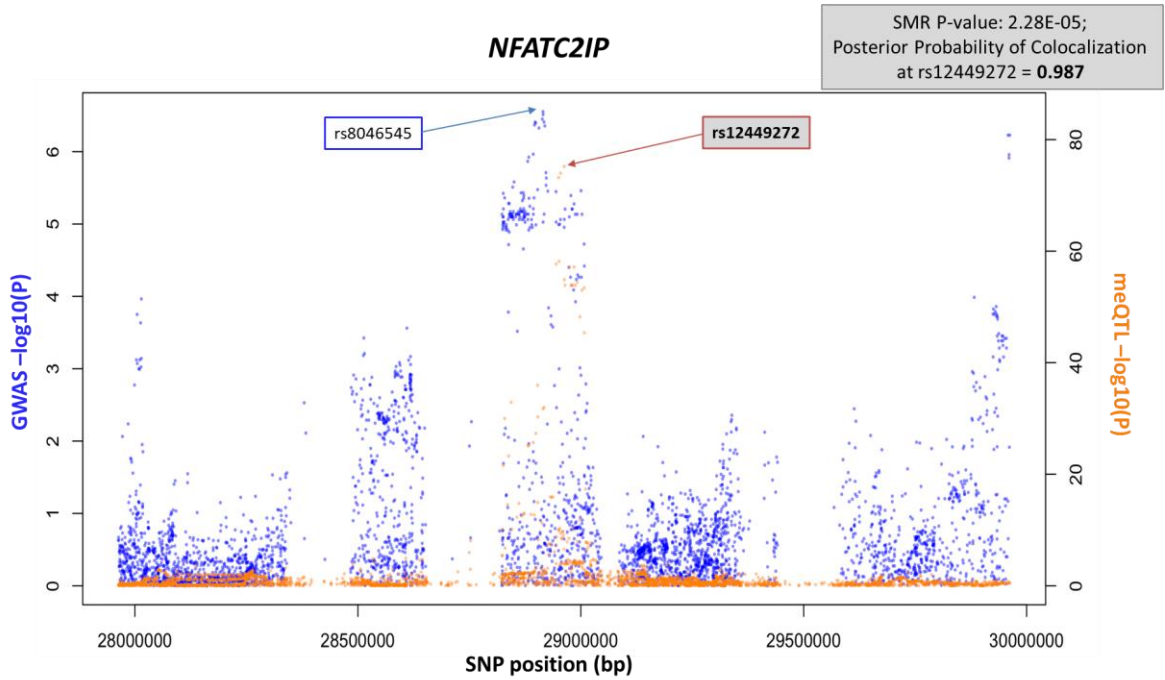


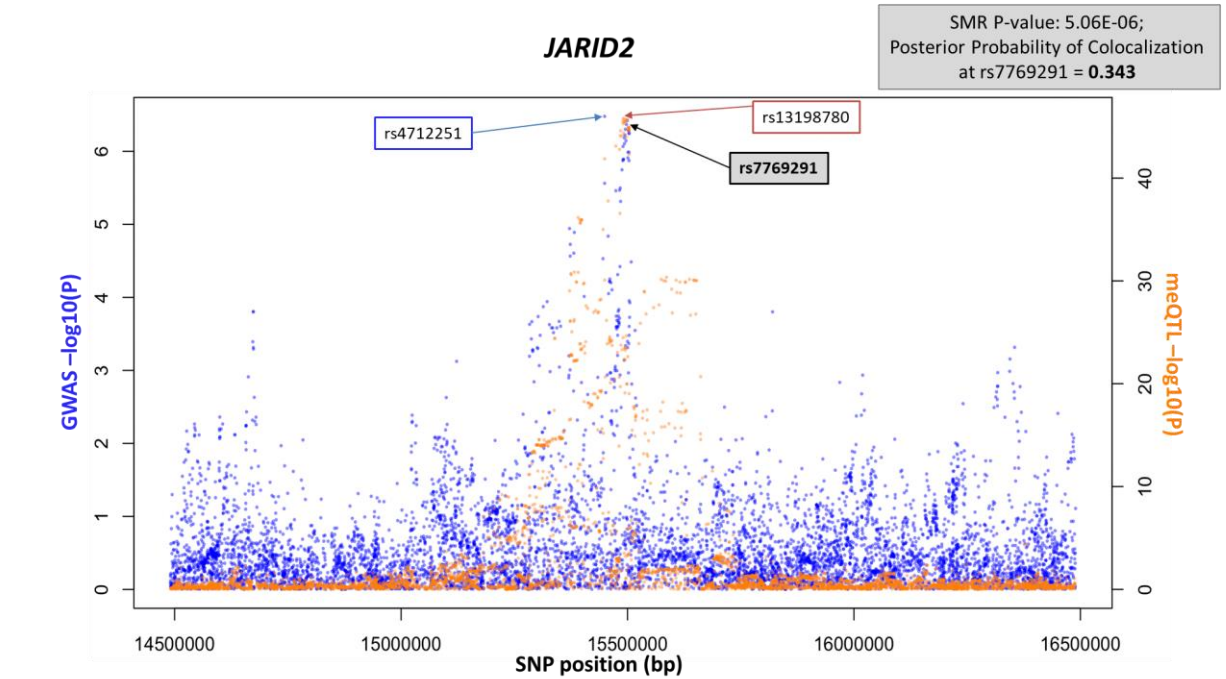
Colocalization Analysis

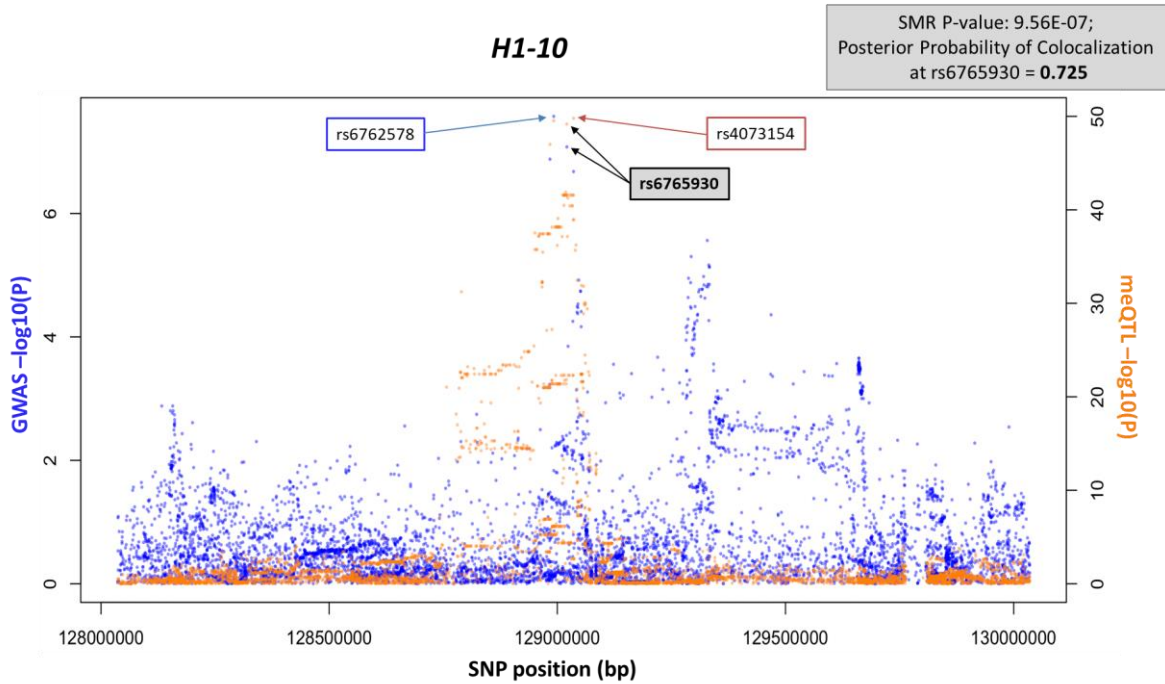
H_0: neither trait has a genetic association in the region
H_1: only trait 1 has a genetic association in the region
H_2: only trait 2 has a genetic association in the region
H_3: both traits are associated, but with different causal variants
H_4: both traits are associated and share a single causal variant

CpG	Nearest Gene	EWAS P	PP.H0.abf	PP.H1.abf	PP.H2.abf	PP.H3.abf	PP.H4.abf
cg26663590	*NFATC2IP*	4.194E-11	4.69E-75	1.01E-73	1.52E-02	3.27E-01	6.58E-01
cg06710464	*BAIAP2*	4.233E-11	1.68E-69	8.46E-69	1.14E-03	4.72E-03	9.94E-01
cg26405097	*JARID2*	1.063E-08	4.94E-37	7.66E-36	1.35E-03	1.99E-02	9.79E-01
cg14468090	*H1-10*	7.959E-08	2.49E-57	7.99E-56	2.27E-04	6.29E-03	9.93E-01


BAIAP2
SMR P-value: 1.55E-06;
Posterior Probability of Colocalization
at rs9908480 = 0.992
rs9908480
rs9908480
GWAS −log10(P)
meQTL −log10(p)
SNP position (bp)
78000000
78500000
79000000
79500000
80000000


NFATC2IP
SMR P-value: 2.28E-05;
Posterior Probability of Colocalization
at rs12449272 = 0.987
rs8046545
rs12449272
GWAS −log10(P)
meQTL −log10(P)
SNP position (bp)
28000000
28500000
29000000
29500000
30000000


JARID2
SMR P-value: 5.06E-06;
Posterior Probability of Colocalization
at rs7769291 = 0.343
rs4712251
rs13198780
rs7769291
GWAS −log10(P)
meQTL −log10(p)
SNP position (bp)
14500000
15000000
15500000
16000000
16500000


H1-10
SMR P-value: 9.56E-07;
Posterior Probability of Colocalization
at rs6765930 = 0.725
rs6762578
rs4073154
rs6765930
GWAS −log10(P)
meQTL −log10(P)
SNP position (bp)
128000000
128500000
129000000
129500000
130000000

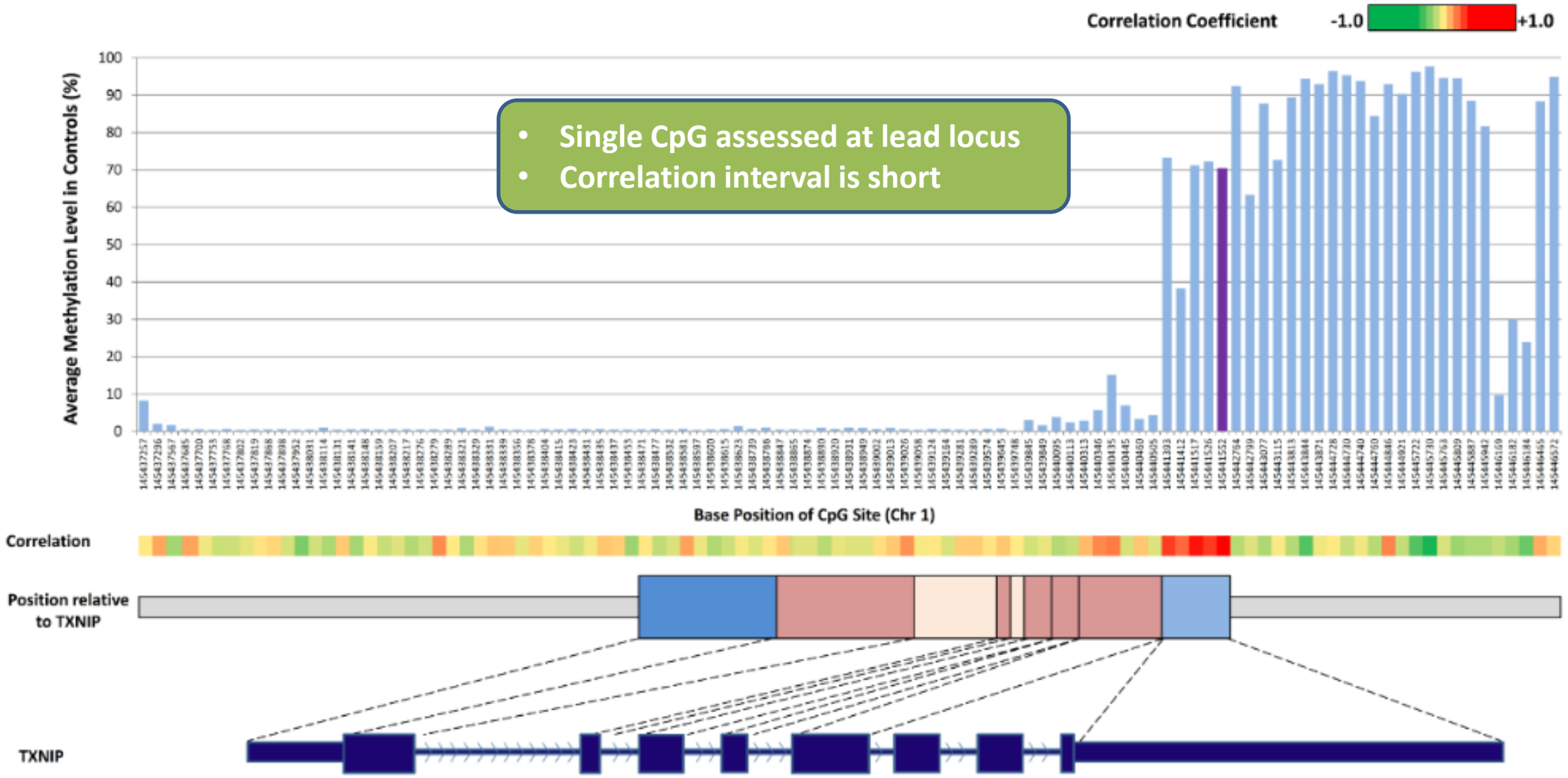
Plausible candidates

- *JARID2*. Involved in development of endocrine pancreas. Conditional ablation in mice impairs endocrine cell differentiation, and activation of the insulin-producing β-cell differentiation program.
- *BAIAP2* encodes the Insulin Receptor Substrate p53 (IRSp53), a target for insulin activated tyrosine kinase in the CNS. May be involved in myogenic differentiation.

Candidate genes: *cis* eQTM analyses (Methylation to Gene expression relationships)

- RNAseq N=1,200 SG samples:
  - NEBNext® Ultra™ II Directional RNA Library Prep
  - GLOBINClear for depletion of globin gene RNA and Ribosomal RNA (rRNA).
  - NovaSeq6000, paired-end run (2 x 150bp)
  - 30M aligned reads per library (~9Gb of data)
- Test 314 Sentinel CpGs for association with expression of nearest gene
- 65 genes as QTMs at $P<0.05$ after Bonferroni correction

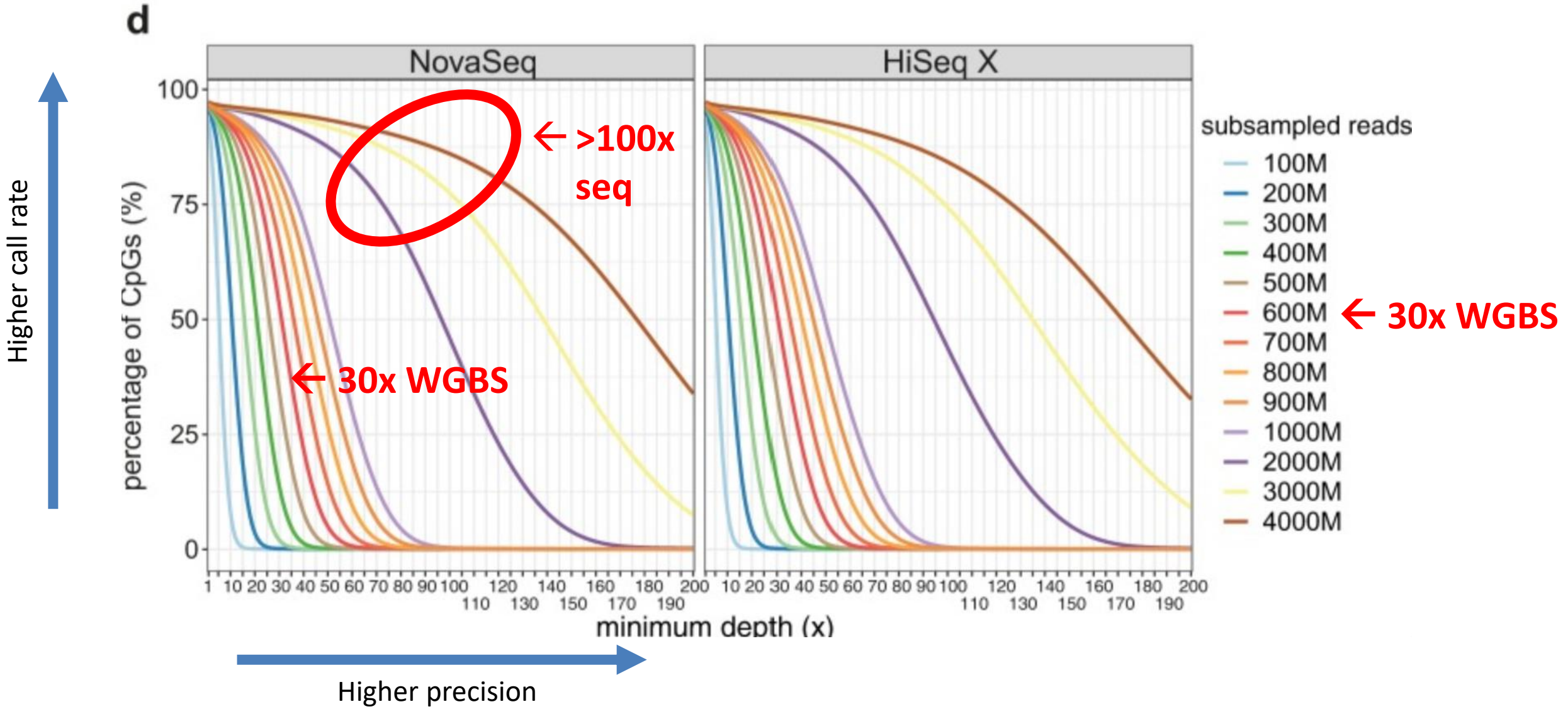
The need for resequencing: Methylation arrays assess <3% of epigenome



Whole Methylome sequencing for T2D

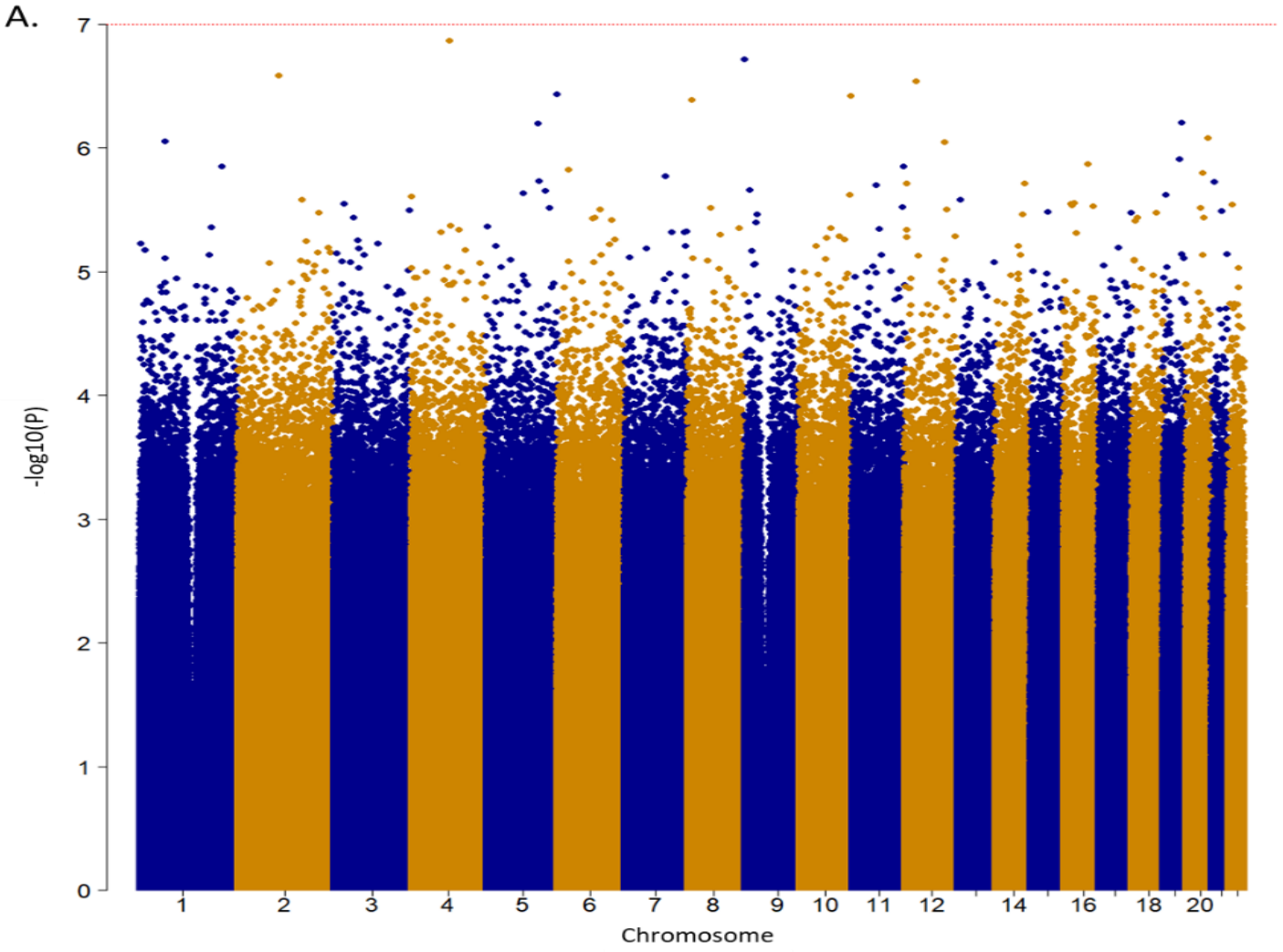
- 264 incident T2D cases and 286 controls
- Bisulfite conversion: EZ DNA Methylation Gold Kit (Zymo)
- Library construction: Accel-NGS® Methyl-Seq Kit (Swift Biosciences)
- Sequencing: Illumina® NovaSeq 6000 platform
- 2x150bp paired-end; target 30x coverage
- Minimum depth 5x

The problem with methylome sequencing



30x Whole methylome association results

→ No additional loci at $P<10^{-7}$



MAE for replicates

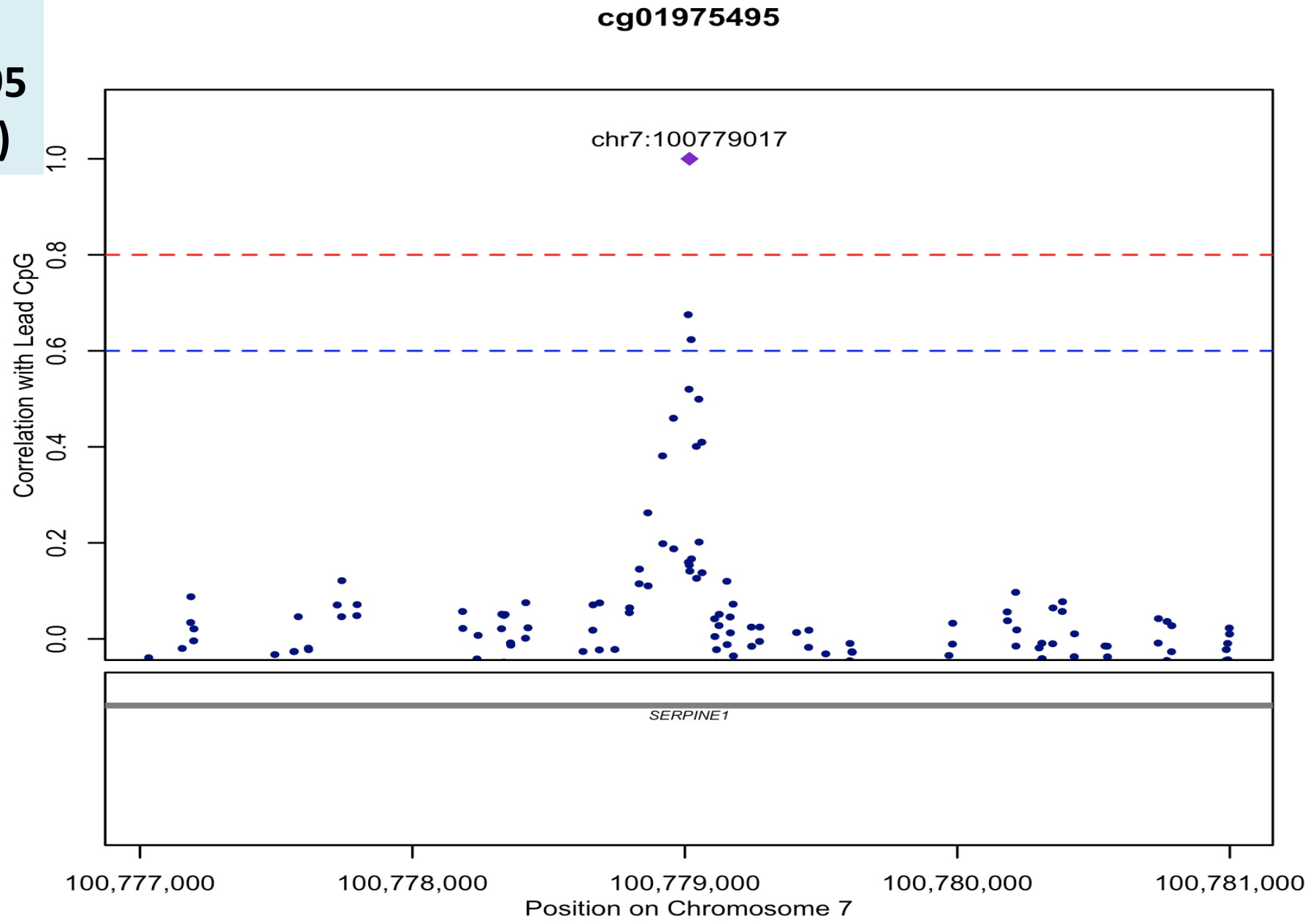
- Array: 1.4%
- 30x WGBS: 7.5%

$S^2_{total} = S^2_{biological} + S^2_{technical}$

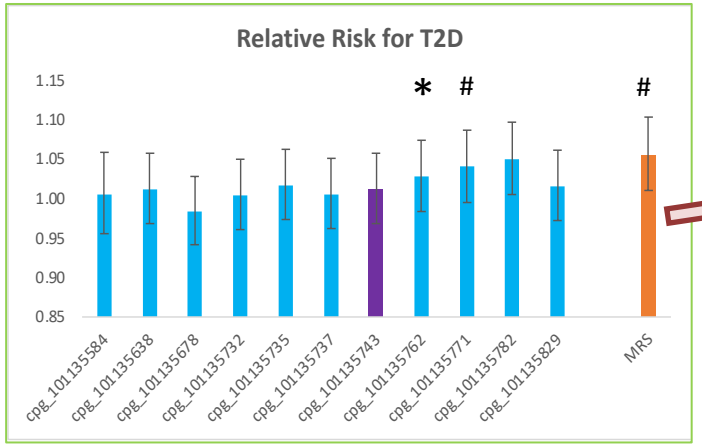
30x WGBS:

$S^2_{technical} >> S^2_{biological}$

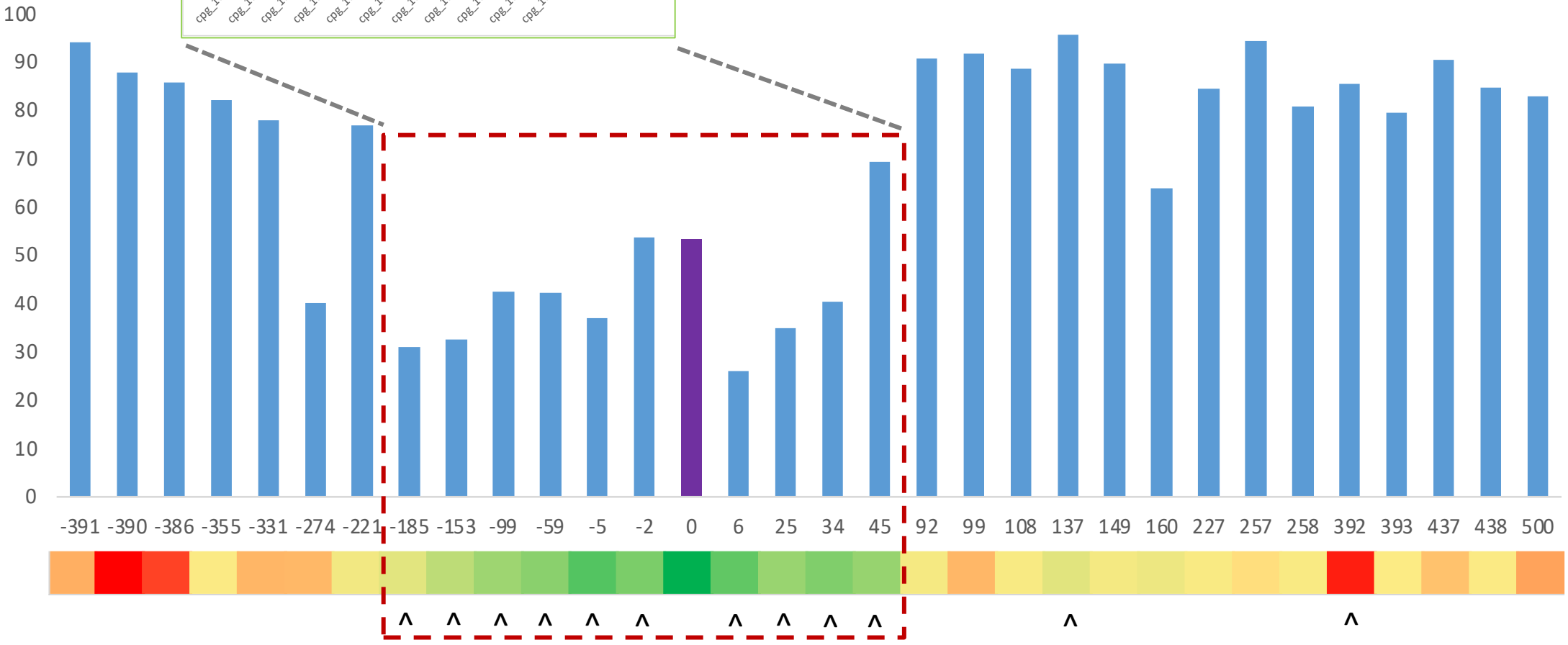
Example: cg01975495 (*SERPINE1*)



Example: cg01975495 (*SERPINE1*)



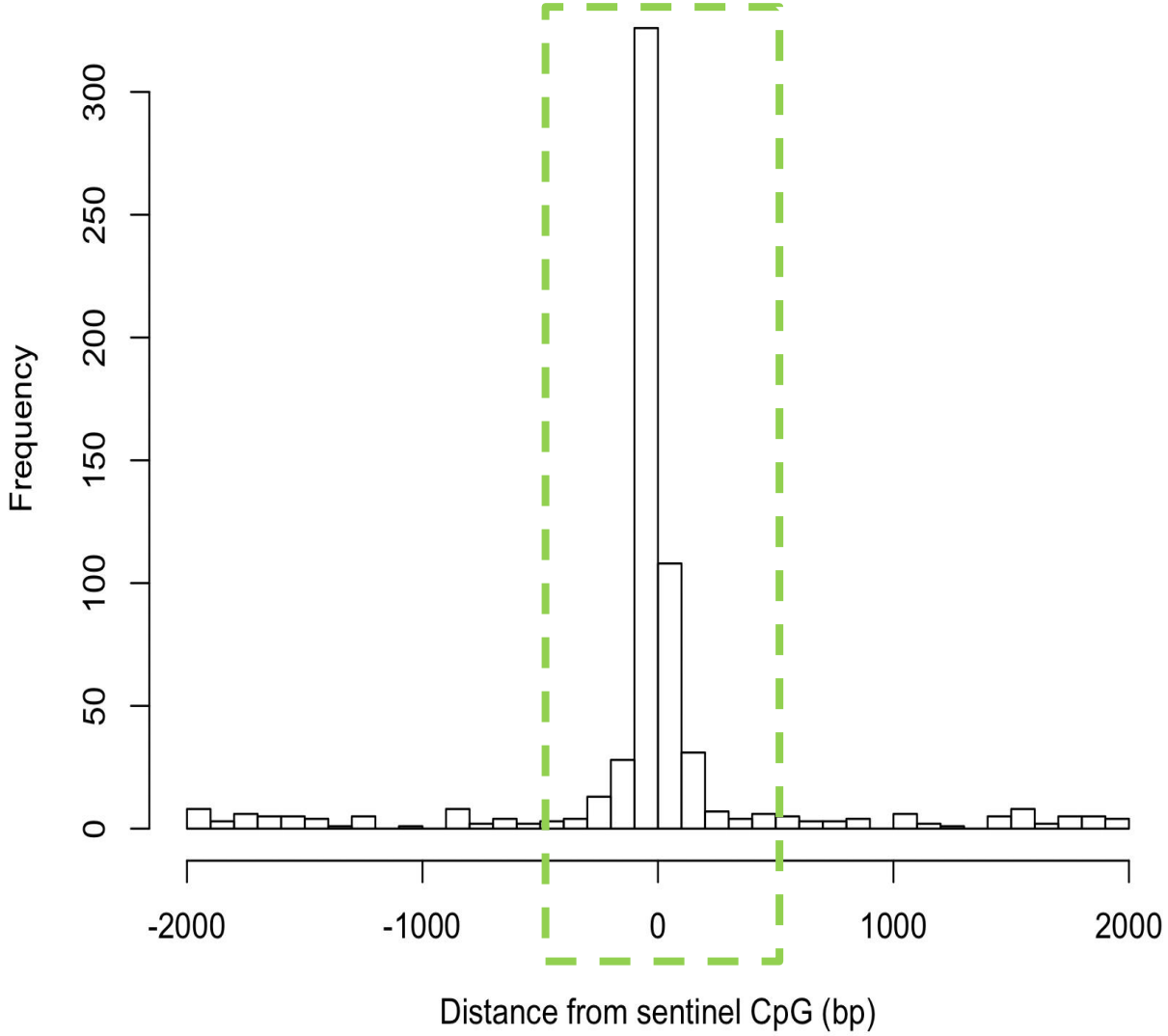
MRS comprises of cpg_101135771 and cpg_101135782 (marked with * and #; P<0.1 and <0.05 respectively for association with incident T2D in WGBS (n=500)



Correlation Coefficient -1 1

^: P<0.05 for correlation with sentinel CpG (marked in purple)

Distribution of distance of most outlying CpG with |r|>0.2 from sentinel CpG

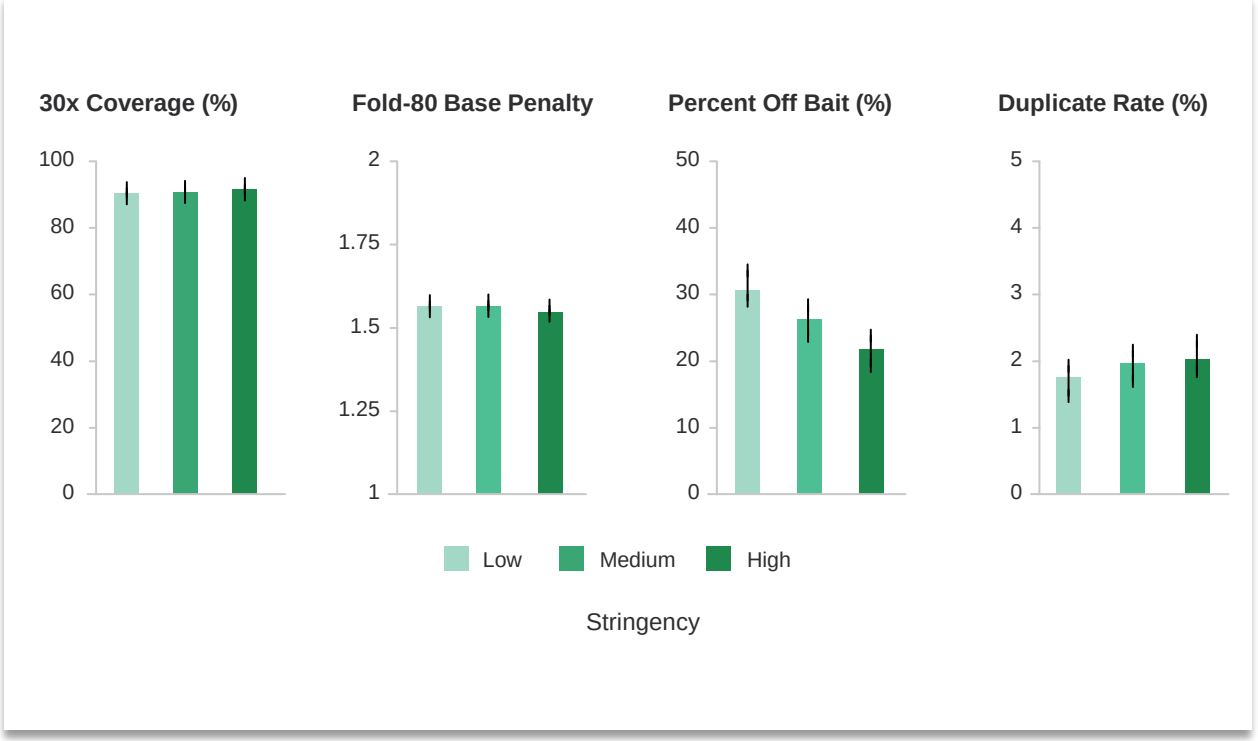


- Inspected correlation based on WGBS (within +/- 2kb)
- Most of the correlation appears to be within +/-500bp from sentinel CpG

Twist Methylation Sequencing

New Solutions for Methylation Sequencing:

- Exceptional Performance
- Unparalleled Uniformity
- Sensitivity
- Rapid Panel Iteration
- Design Flexibility
- End-to-end Workflow with Gentle Enzymatic Conversion



Quantitatively Tune On-target Performance:

Stringency Control for High Quality Pre-optimized Designs

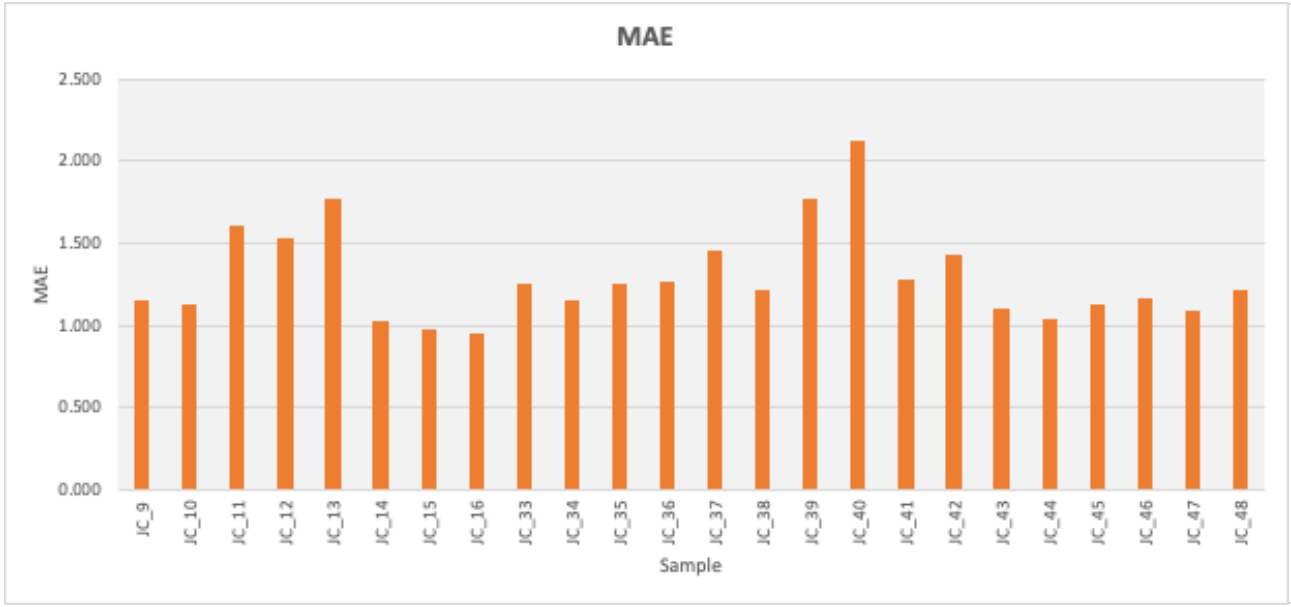
Pilot Study

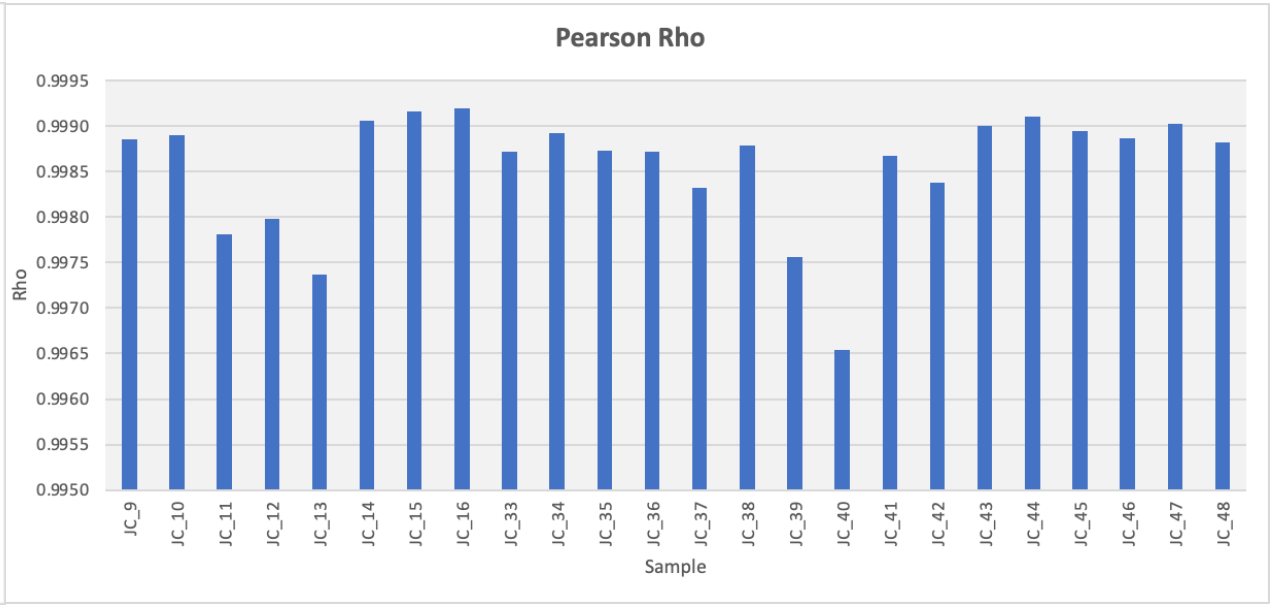
- TWIST Methylome Panel (~5M markers)
- 16 samples (500ng DNA in total)
- Samples split equally
- Library prep / sequencing in duplicate (SG and US)
- Aimed to achieve ~500x raw sequencing

Concordance between duplicates

>100x coverage at 77% of sites

	rho	MAE	#Common	#Unique
Mean	**0.9986**	**1.29**	**6,367,108**	**157,562**
S.D.	0.0007	0.28	95,816	84,273
Min	0.9965	0.95	6,195,538	92,080
q1	0.9983	1.11	6,277,992	105,863
q2	0.9988	1.21	6,362,474	125,642
q3	0.9990	1.44	6,418,289	192,993
Max	0.9992	2.12	6,601,365	457,757

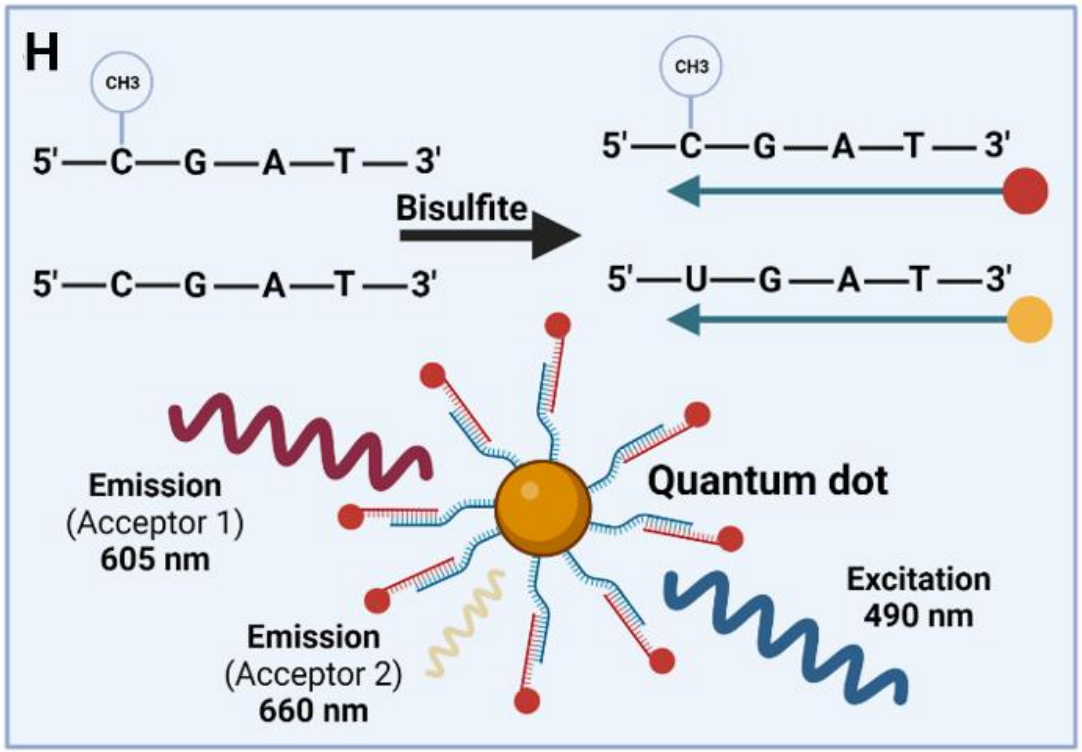


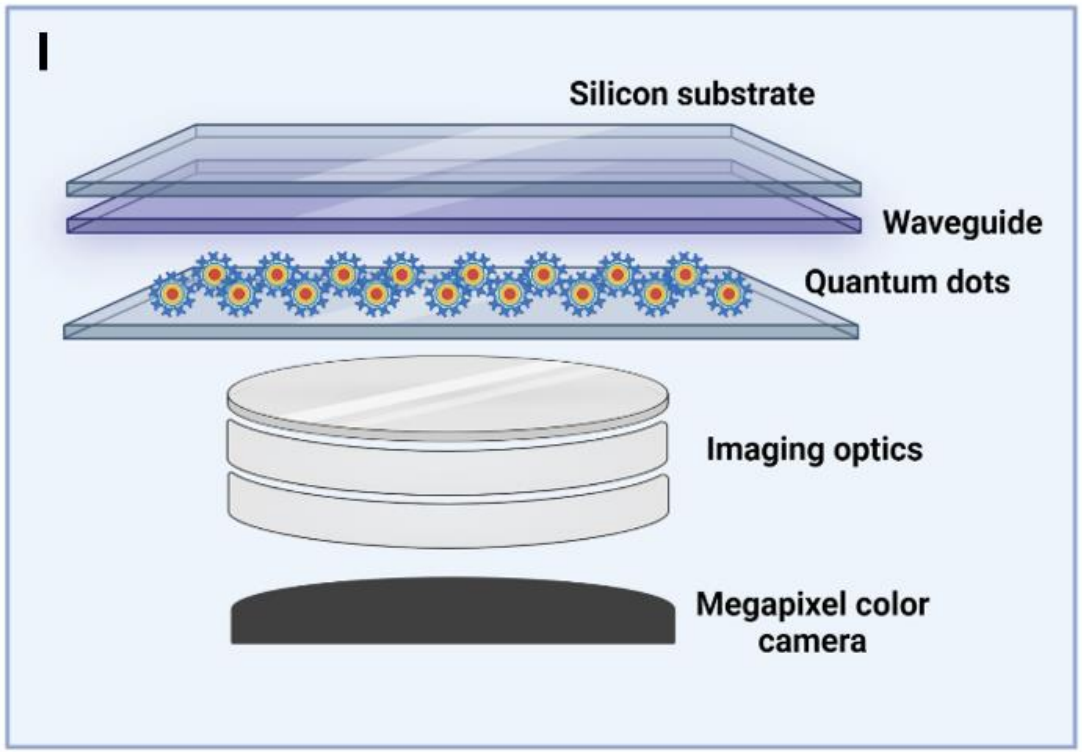


Strategy for TWIST targeted methylation sequencing

- 2,000 samples, 500x sequencing
- 500MB sequence target selected as:
  - Minimum +/-500bp from sentinel CpG for all loci
  - Inspected +/-2kb for correlated markers (2kb prior literature), and extend interval up/downstream to include CpGs with $|r|>0.2$
  - Where the outermost CpG is close to (within 100bp) 2kb boundary of initial evaluation of correlation, further extend the region of evaluation to +/-5kb.
- Results …. Watch this space

Translation to clinic





Wrap-up

- Diabetes is a critical public health problem, with a need for new predictive biomarkers and preventative treatments.
- Methylation array studies identify >400 methylation markers for future diabetes
  - Biologically compelling
  - Independent predictive value >> genetic risk
- Fine mapping studies designed based on methylome sequencing and underway using targeted methylation panels