Epigenetics

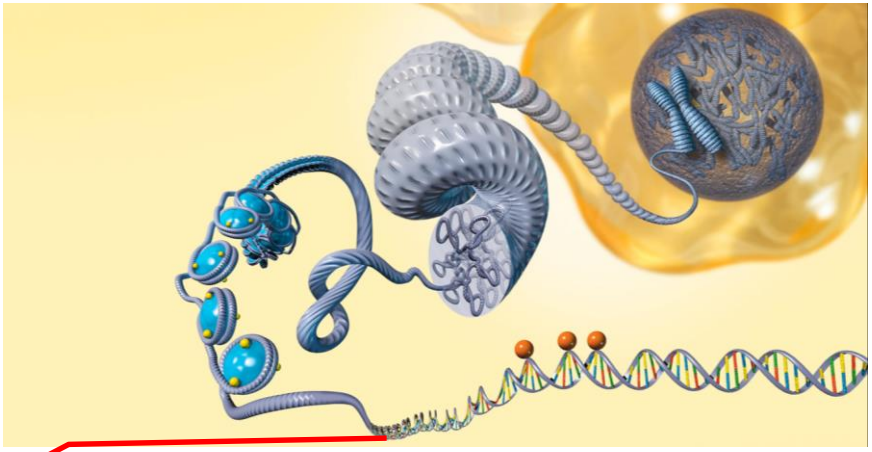
Jenny van Dongen, Michel Nivard
Vrije Universiteit (VU) Amsterdam

Boulder, Friday march 8, 2019

Genome: the DNA sequence
GWAS: *Which genetic variants are associated with a trait/disease?*

GGTGTACTCGTTGTCT
AAGGGCCTGGACTAG
CTGGGACTTA**G**GCTTC
TGCCGGTT**C**CAACACT
CTGGTGCTGGAAGGC
TGGACTTGGGTGACT
CAAGTTCCCTCTTCCT
ACTGCAATGCAAGAAA
ATAACAAAAGAAGTAT
GTATACCTTTAAGTAT
CTCAAAGAGCTATCTC
AGCTTCTGAATTTCCT
TCTAGGG**C**ACCTCTTC
CTGCGGTGTACTCGTT
GTCTAAGGGCCTGGA
CTAGCTGGGACTTAG
GCTTCTGCCGGTTCCA
ACACTCTGGTGCTGG
AAGGCTG**G**ACTTGGG
TGA

Epigenome: The collection of epigenetic marks* that regulate gene expression
Epigenome-wide association studies (EWAS)
Which epigenetic marks are associated with a trait/disease?



*multiple, inter-related layers of molecular information
e.g. DNA methylation, histone modifications, non-coding RNAs

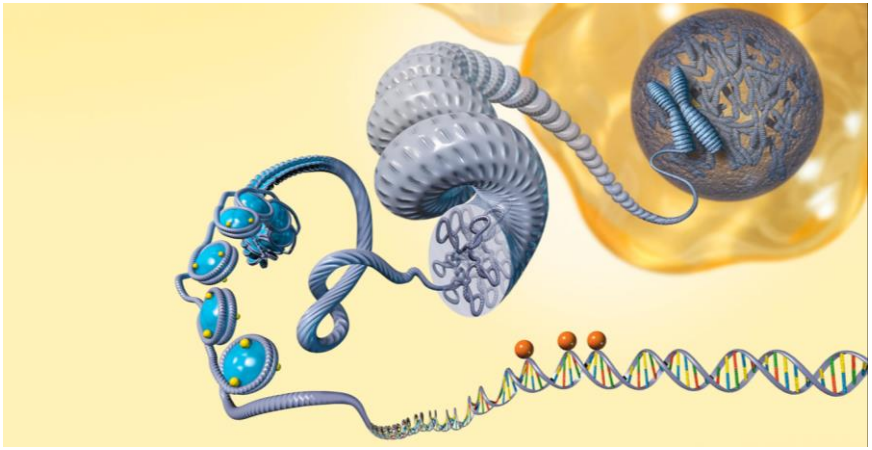
Genome: very stable throughout life (exception: *de novo* mutations)

GGTGTACTCGTTGTCT
AAGGGCCTGGACTAG
CTGGGACTTA**G**GCTTC
TGCCGGTT**C**CAACACT
CTGGTGCTGGAAGGC
TGGACTTGGGTGACT
CAAGTTCCCTCTTCCT
ACTGCAATGCAAGAAA
ATAACAAAAGAAGTAT
GTATACCTTTAAGTAT
CTCAAAGAGCTATCTC
AGCTTCTGAATTTCCT
TCTAGGG**C**ACCTCTTC
CTGCGGTGTACTCGTT
GTCTAAGGGCCTGGA
CTAGCTGGGACTTAG
GCTTCTGCCGGTTCCA
ACACTCTGGTGCTGG
AAGGCTG**G**ACTTGGG
TGA

Epigenome: dynamic
Programmed epigenetic changes (development and tissue differentiation)

Substantial changes in DNA methylation with ageing

Changes in response environment/exposures (e.g. cigarette smoke)



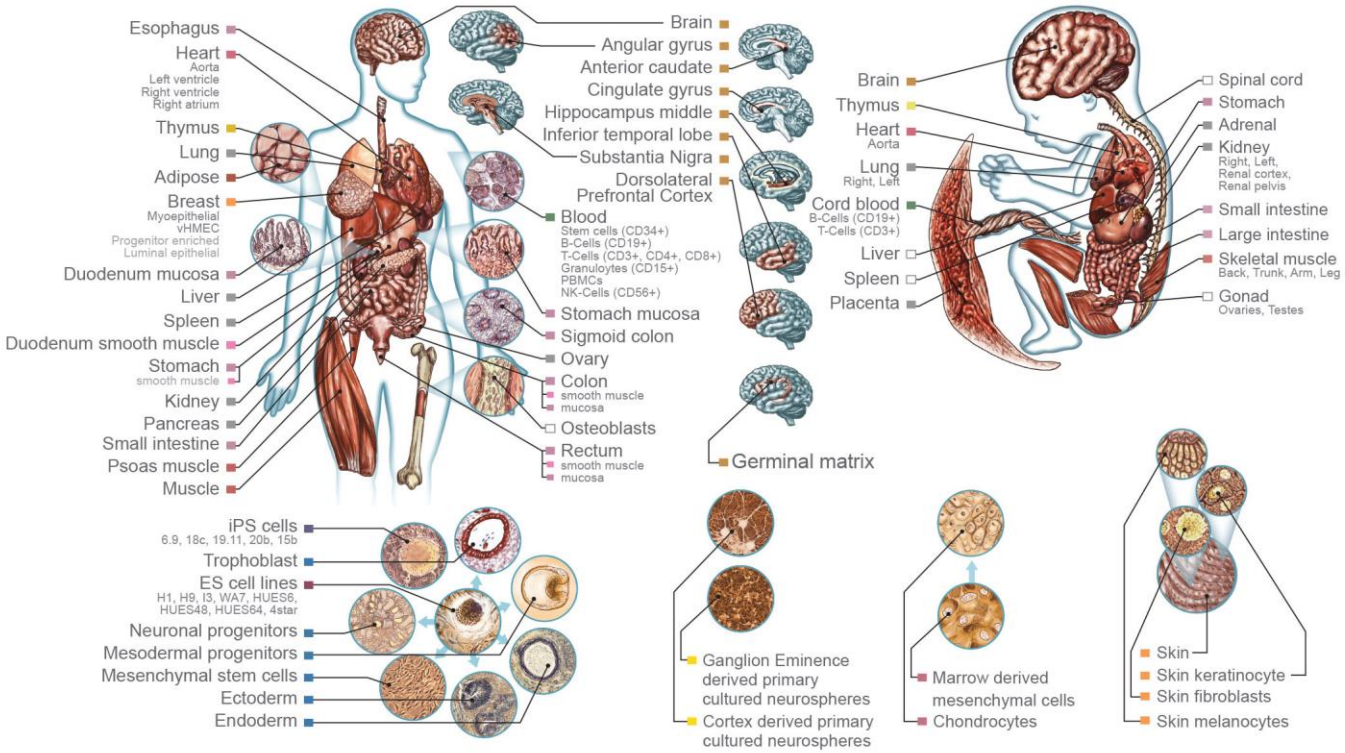
Epigenetics

Epigenetics= The study of molecular mechanisms that influence the activity of gene expression and that **are transmitted across cell division.**

[*definition by Bird 2007 Nature*]

- *epi-* (Greek: *επί-* over, above)
- epigenetics= “Above Genetics”

Each cell has its own *epigenome*



Kundaje, A, et al. "Integrative analysis of 111 reference human epigenomes." Nature 518.7539 (2015): 317-330.

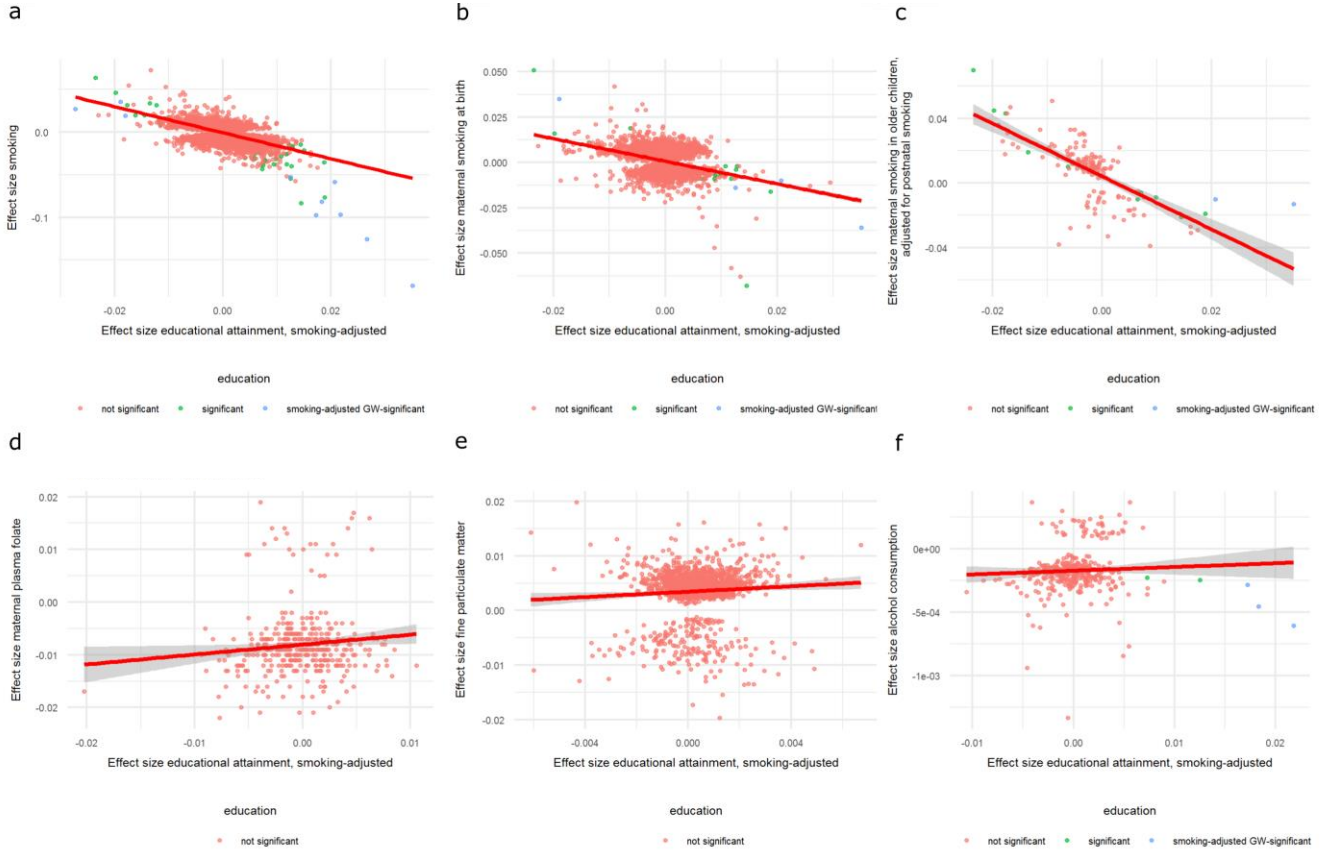
'Memory' functions of the epigenome

- Cellular identity
- Cellular response to environment

Blood EWAS of educational attainment shows epigenetic signatures of:

Smoking

Prenatal maternal smoking (at birth/in older children)



Dutch population (N=**4152, adults**)

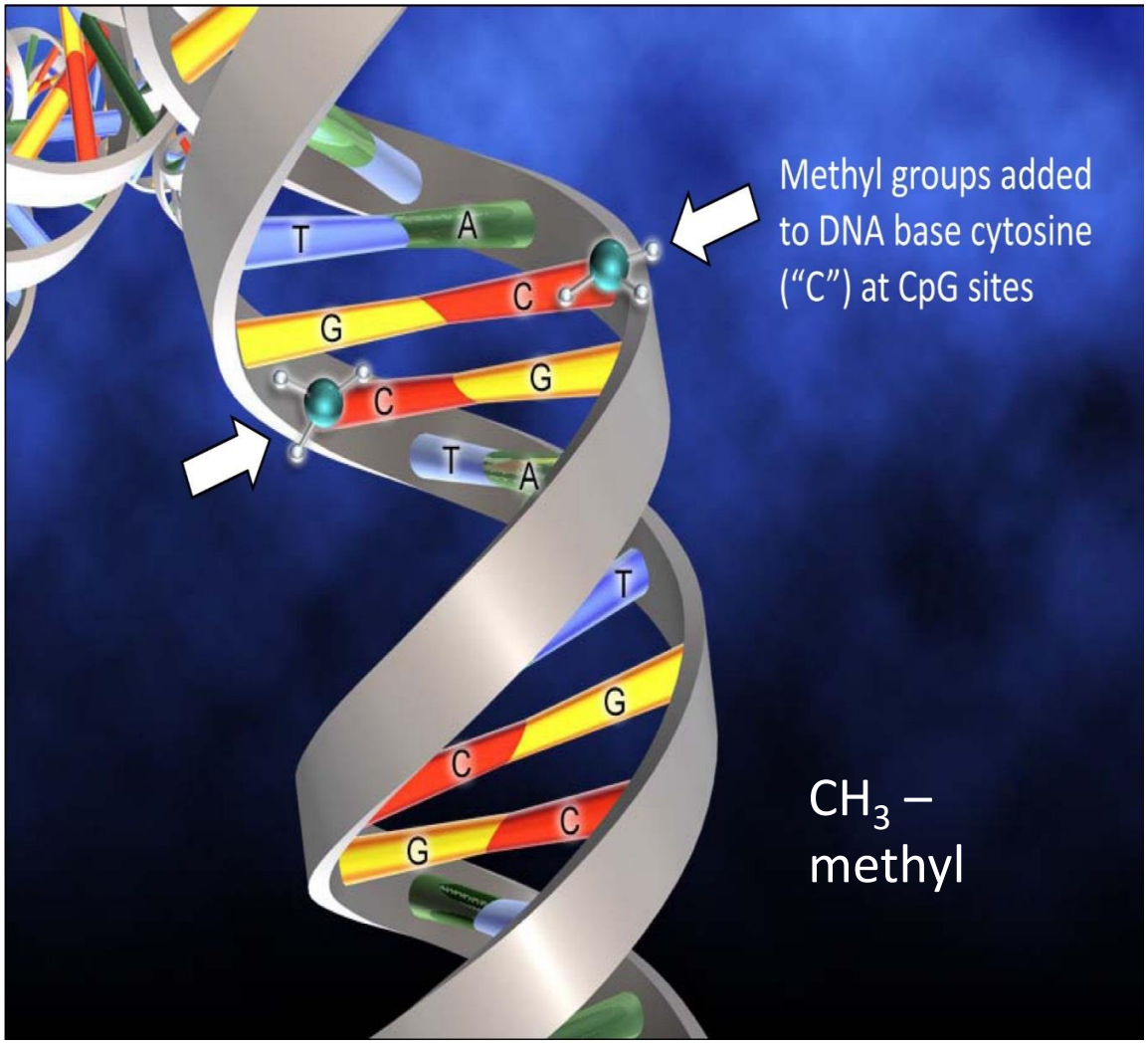
Prenatal maternal plasma folate

Air pollution (fine matter)

Alcohol consumption (n.s.)

van Dongen et al 2018 npj Science of Learning

DNA methylation

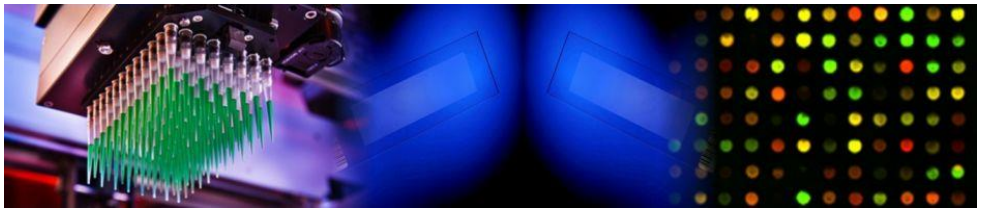


Most studied epigenetic mark in EWAS
→stable (covalent bond): no fresh tissue required

Relation with gene expression

- **at gene promoters**: usually associated with repression of transcription
- **at enhancers:** strongest relationship (often negative) with transcription
- **in gene bodies**: may regulate alternative splicing

Genome-wide methylation microarrays



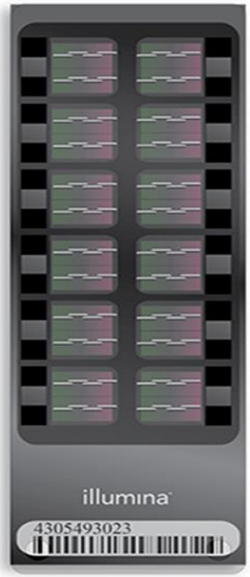
Illumina 450k array

- 485,000 methylation sites
- 99% of RefSeq genes
- average of 17 CpG sites per gene
- promoter, 5'UTR, first exon, gene body, and 3'UTR.
- No longer sold, but most commonly used currently in EWAS

PRACTICAL

Illumina EPIC array

- > 850,000 methylation sites
- Larger coverage of:
  - Enhancers (ENCODE, FANTOM)
  - ENCODE open chromatin
  - ENCODE transcription factor binding sites



Same technology. Difference: Coverage (EPIC measures more methylation sites)
Note: only a small subset of the ~30 million CpG sites in the genome

- Often used in EWAS (suited for large numbers of samples)
- Cost-effective (much cheaper than bisulphite sequening)

Illumina 450k array data (peripheral blood)

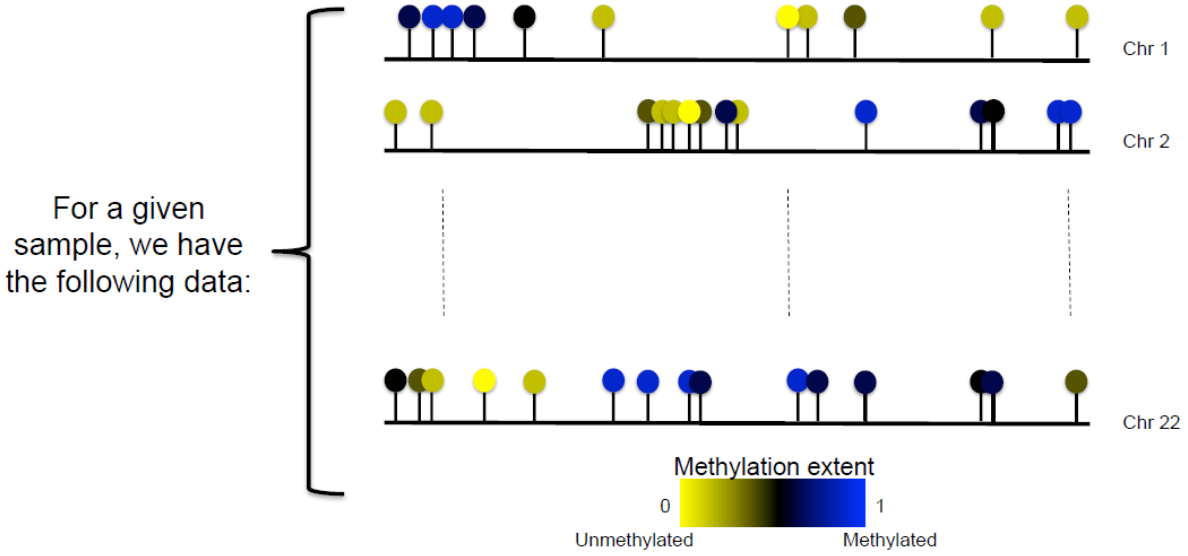
For > 450.000 sites:

- **methylation level**: proportion of methylated alleles.
- Continuous trait, range: 0-1

- DNA extracted from blood comes from billions of cells
- In some cells, the position may be methylated, while it is unmethylated in others

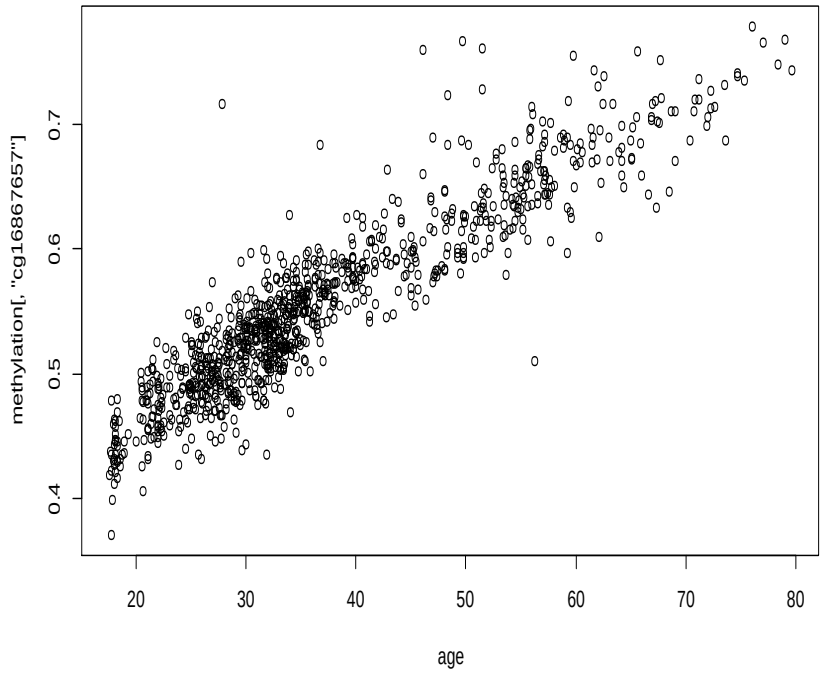
Methylation level at 1 location in 1 sample:

- 0= all DNA was unmethylated at this position
- 1= all DNA was methylated at this position



Epigenome-wide association study (EWAS)

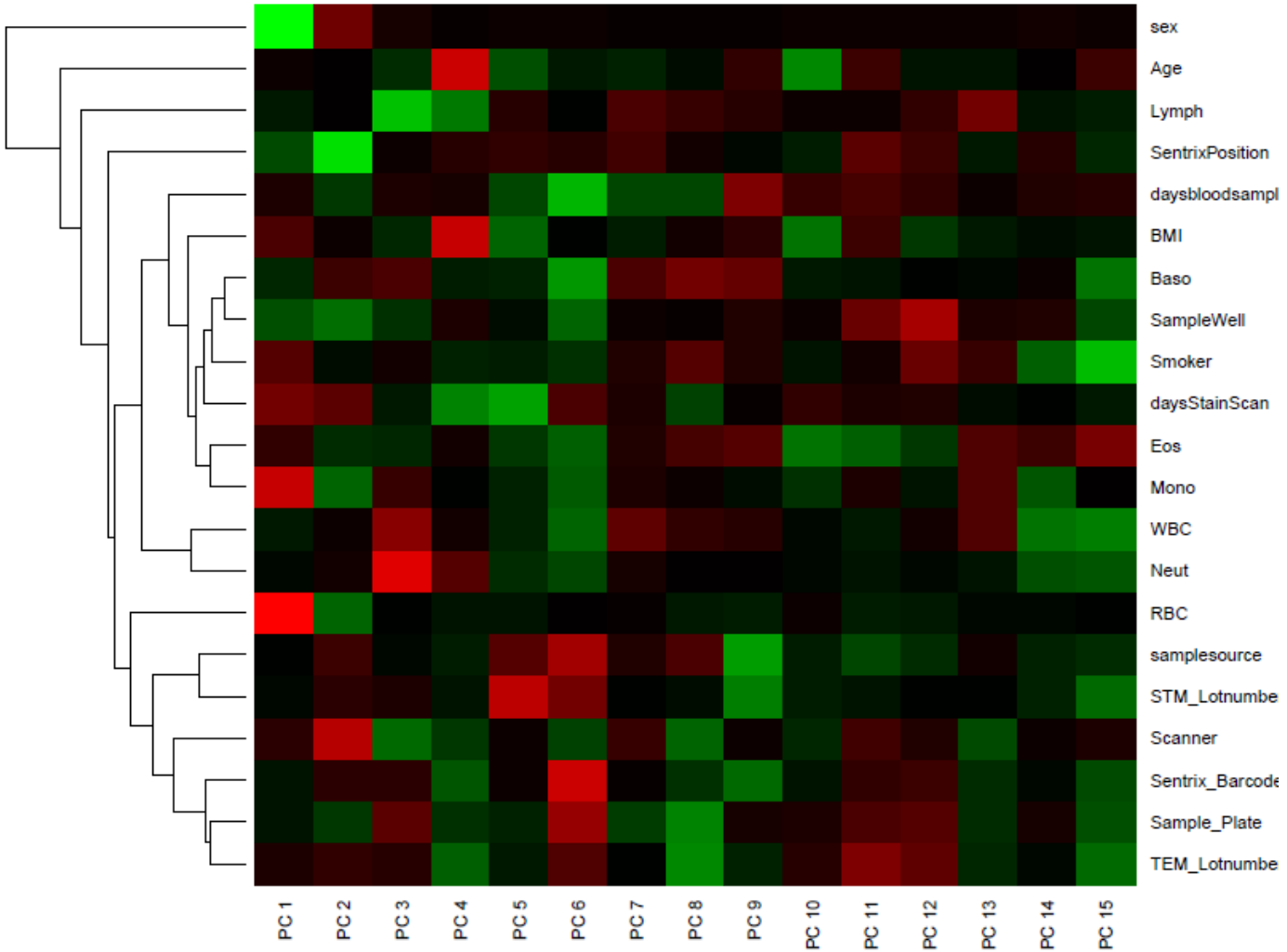
DNA methylation level at 1 CpG



Age

- Test if there is a significant relationship between DNA methylation level and trait of interest
- Repeat this test for thousands of methylation sites in the genome

PCA of a whole blood DNA methylation dataset (450k array) from > 3000 samples (NTR)
Main sources of biological variation: sex, white blood cell counts, age
Main sources of technical variation: Position on the chip (in particular row)



PCs 1 - 15

van Dongen et al Nat. Commun. 2016

How to deal with this in EWAS?

- Firstly: quality control and normalization to reduce technical variation

- Correction for known confounders/sources of variation
  -> Inclusion of biological covariates (e.g. age, sex, cell counts, smoking) & technical covariates in EWAS model

- Correction for hidden (unobserved) confounders
  - Inclusion of PCs as covariates in EWAS model
  - Batch correction tools (R packages)
  - OSCA (Zhang et al bioRxiv) – mixed-linear-model with all other distal probes fitted as random effect

- If you don't correct for confounders properly → can cause inflation of test statistics
- Bacon: R-package for estimating and adjusting for inflation of EWAS test statistics

OSCA: a tool for omic-data-based complex trait analysis

Futao Zhang[1], Wenhan Chen[1], Zhihong Zhu[1], Qian Zhang[1], Ian J. Deary[2], Naomi R. Wray[1,3], Peter M. Visscher[1,3], Allan F. McRae[1], Jian Yang[1,3,4,*]

METHOD | Open Access

Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution

Maarten van Iterson, Erik W. van Zwet, the BIOS Consortium and Bastiaan T. Heijmans

Genome Biology 2017 **18**:19
https://doi.org/10.1186/s13059-016-1131-9 | © The Author(s) 2017
Received: 23 June 2016 | **Accepted:** 12 December 2016 | **Published:** 27 January 2017

Epigenome-wide association studies (EWAS)

Goal

Identify genomic regions where DNA methylation:

- Differs between disease cases and controls (or correlates with a continuous trait)
- Differs between people with different lifestyles / environmental exposures

Motivation

1. To enhance **understanding of the biological mechanisms** that are involved in a disease/trait or that are modified b**y environmental exposures**
2. **To identify biomarkers** for a disease, a trait, or an environmental exposure

Last couple of years: large-scale EWAS meta-analysis projects

Usually: DNA from peripheral tissues (blood)

note: blood is not the primary tissue of interest for many traits/diseases, but is suitable for biomarkers

nature
International journal of science

Letter | Published: 21 December 2016

Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity

Simone Wahl, Alexander Drong [...] John C. Chambers

Nature **541**, 81–86 (05 January 2017) | Download Citation

Cold Spring Harbor Laboratory
bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

New Results

Comment on this paper

Epigenome-wide meta-analysis of blood DNA methylation and its association with subcortical volumes: findings from the ENIGMA Epigenetics Working Group

Tianye Jia, Congying Chu, Yun Liu, Jenny van Dongen, Nicola J Armstrong, Mark E Bastin, Tania Carrillo-Roa, Anouk den Braber, Mathew Harris, Rick Jansen, Jingyu Liu, Michelle Luciano, Anil P.S. Ori, Roberto Roiz Santianez, Barbara Ruggeri, Daniil Sarkisyan, Jean Shin, Kim Sungeun, Diana Tordesillas Gutierrez, Dennis van't Ent, David Ames, Eric Artiges, Georgy Bakalkin, Tobias Banaschewski, Arun L.W. Bokde, Henry Brodaty, Uli Bromberg, Rachel Brouwer, Christian Buchel, Erin Burke Quinlan, Wiepke Cahn,

nature COMMUNICATIONS

Article | OPEN | Published: 10 February 2016

Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns

Bonnie R. Joubert, Herman T. den Dekker [...] Stephanie J. London

Nature Communications **7**, Article number: 10577 (2016) | Download Citation

AJHG
Volume 98, Issue 4, 7 April 2016, Pages 680-696
Cell PRESS

Article

DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis

Bonnie R. Joubert [1, 58], Janine F. Felix [2, 3, 4, 58], Paul Yousefi [5, 58], Kelly M. Bakulski [6, 58], Allan C. Just [7, 58], Carrie Breton [8, 58], Sarah E. Reese [1, 58], Christina A. Markunas [1, 9, 58], Rebecca C. Richmond [10, 58], Cheng-Jian Xu [11, 12, 13, 58], Leanne K. Küpers [14, 58], Sam S. Oh [15, 58], Cathrine Hoyo [16, 58], Olena Gruzieva [17, 58], Cilla Söderhäll [18, 58], Lucas A. Salas [19, 20, 21, 58], Nour Baïz [22, 58], Hongmei Zhang [23, 58] ... Stephanie J. London [1, 59]

nature
International journal of science

Letter | Published: 21 December 2016

Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity

Simone Wahl, Alexander Drong [...] John C. Chambers

Nature **541**, 81–86 (05 January 2017) | Download Citation

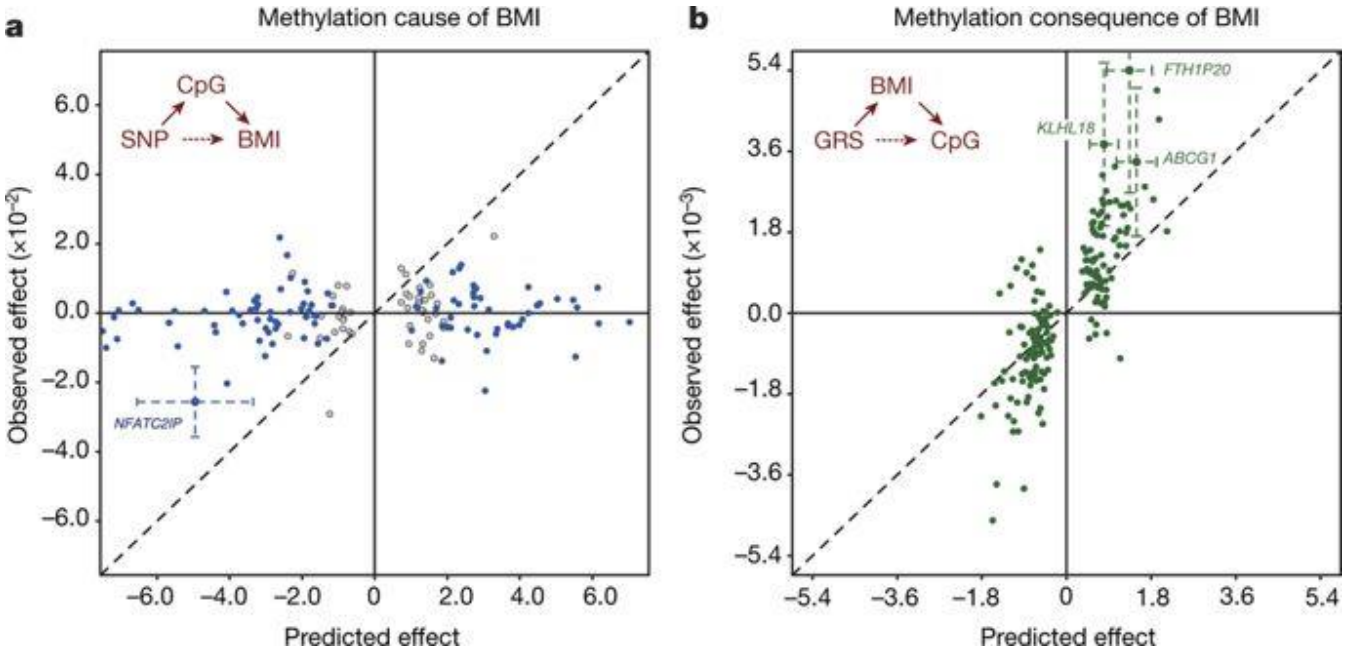
Dekkers *et al. Genome Biology* (2016) 17:138
DOI 10.1186/s13059-016-1000-6

Genome Biology

RESEARCH Open Access

Blood lipids influence DNA methylation in circulating cells

Koen F. Dekkers[1], Maarten van Iterson[1], Roderick C. Slieker[1], Matthijs H. Moed[1], Marc Jan Bonder[2], Michiel van Galen[3], Hailiang Mei[4], Daria V. Zhernakova[2], Leonard H. van den Berg[5], Joris Deelen[1], Jenny van Dongen[6], Diana van Heemst[7], Albert Hofman[8], Jouke J. Hottenga[6], Carla J. H. van der Kallen[9], Casper G. Schalkwijk[9], Coen D. A. Stehouwer[9], Ettje F. Tigchelaar[2], André G. Uitterlinden[10], Gonneke Willemsen[6], Alexandra Zhernakova[2], Lude Franke[2], Peter A. C. 't Hoen[3], Rick Jansen[11], Joyce van Meurs[10], Dorret I. Boomsma[6], Cornelia M. van Duijn[8], Marleen M. J. van Greevenbroek[9], Jan H. Veldink[5], Cisca Wijmenga[2], BIOS Consortium[12], Erik W. van Zwet[13], P. Eline Slagboom[1], J. Wouter Jukema[14] and Bastiaan T. Heijmans[1*]



Circulation:
Cardiovascular Genetics

ORIGINAL ARTICLE

Epigenetic Signatures of Cigarette Smoking

Roby Joehanes, Allan C. Just, Riccardo E. Marioni, Luke C. Pilling, Lindsay M. Reynolds, Pooja R. Mandaviya, Weihua Guan, Tao Xu, Cathy E. Elks, Stella Aslibekyan, Hortensia Moreno-Macias, Jennifer A. Smith, Jennifer A. Brody, Radhika Dhingra, Paul Yousefi, James S. Pankow, Sonja Kunze, Sonia H. Shah, Allan F. McRae, Kurt Lohman, Jin Sha, Devin M. Absher, Luigi Ferrucci, Wei Zhao, Ellen W. Demerath, Jan Bressler, Megan L. Grove, Tianxiao Huan, Chunyu Liu, Michael M. Mendelson, Chen Yao, Douglas P. Kiel, Annette Peters, Rui Wang-Sattler, Peter M. Visscher, Naomi R. Wray, John M. Starr, Jingzhong Ding, Carlos J. Rodriguez, Nicholas J. Wareham, Marguerite R. Irvin, Degui Zhi, Myrto Barrdahl, Paolo Vineis, Srikant Ambatipudi, André G. Uitterlinden, Albert Hofman, Joel Schwartz, Elena Colicino, Lifang Hou, Pantel S. Vokonas, Dena G. Hernandez, Andrew B. Singleton, Stefania Bandinelli, Stephen T. Turner, Erin B. Ware, Alicia K. Smith, Torsten Klengel, Elisabeth B. Binder, Bruce M. Psaty, Kent D. Taylor, Sina A. Gharib, Brenton R. Swenson, Liming Liang, Dawn L. DeMeo, George T. O'Connor, Zdenko Herceg, Kerry J. Ressler, Karen N. Conneely, Nona Sotoodehnia, Sharon L. R. Kardia, David Melzer, Andrea A. Baccarelli, Joyce B. J. van Meurs, Isabelle Romieu, Donna K. Arnett, Ken K. Ong, Yongmei Liu, Melanie Waldenberger, Ian J. Deary, Myriam Fornage, Daniel Levy, Stephanie J. London

DOI https://doi.org/10.1161/CIRCGENETICS.116.001506
Circulation: Cardiovascular Genetics. 2016;9:436-447
Originally published September 20, 2016

- 2623 Bonferroni significant (current vs never smokers), 18 760 CpGs at false discovery rate <0.05.
- Tissue: Blood
- N=15 907

Blood-based epigenetic prediction of complex traits

ARTICLE

Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations

Sonia Shah,[1,2,14] Marc J. Bonder,[3,14] Riccardo E. Marioni,[1,4,5] Zhihong Zhu,[1] Allan F. McRae,[1,2] Alexandra Zhernakova,[3] Sarah E. Harris,[4,5] Dave Liewald,[4] Anjali K. Henders,[6] Michael M. Mendelson,[7,8,9] Chunyu Liu,[10] Roby Joehanes,[11] Liming Liang,[12] BIOS Consortium, Daniel Levy,[9] Nicholas G. Martin,[6] John M. Starr,[4,13] Cisca Wijmenga,[3] Naomi R. Wray,[1] Jian Yang,[1] Grant W. Montgomery,[6,14] Lude Franke,[3,14] Ian J. Deary,[4,13,14] and Peter M. Visscher[1,2,4,14,*]

Research | Open Access

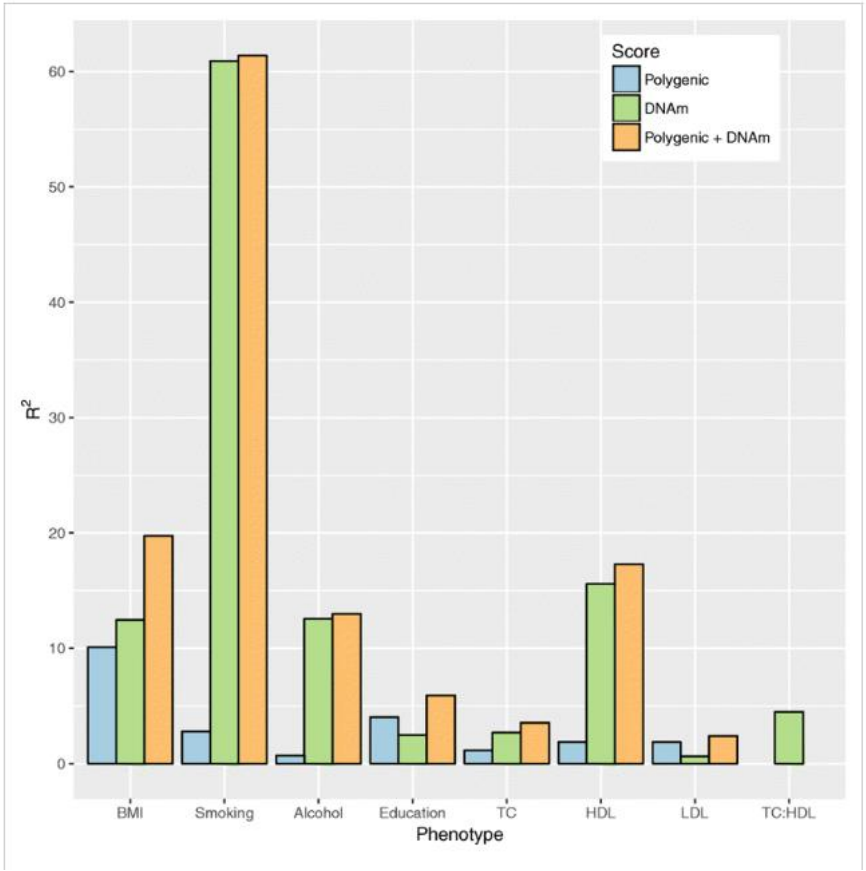
Epigenetic prediction of complex traits and death

Daniel L. McCartney †, Robert F. Hillary †, Anna J. Stevenson, Stuart J. Ritchie, Rosie M. Walker, Qian Zhang, Stewart W. Morris, Mairead L. Bermingham, Archie Campbell, Alison D. Murray, Heather C. Whalley, Catharine R. Gale, David J. Porteous, Chris S. Haley, Allan F. McRae, Naomi R. Wray, Peter M. Visscher, Andrew M. McIntosh, Kathryn L. Evans, Ian J. Deary and Riccardo E. Marioni

†Contributed equally

Genome Biology 2018 **19**:136
https://doi.org/10.1186/s13059-018-1514-1
Received: 3 April 2018 | **Accepted:** 22 August 2018 | **Published:** 27 September 2018

DNA methylation score - epigenetic equivalent of polygenic score



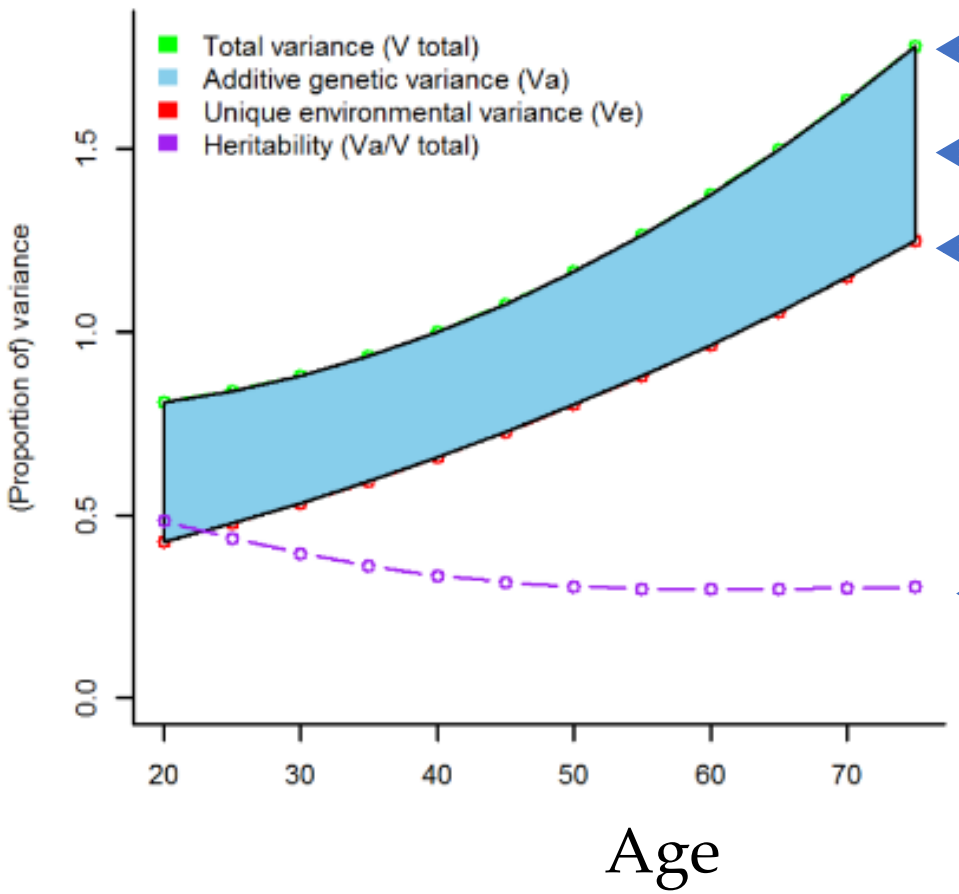
Causes of variation in DNA methylation

- Illumina **450k array**, whole blood (refs 1,2)
  - Average heritability = ~19%
  - Average SNP heritability= 7%

DNA-sequence contributes to its own regulation
Environment accounts for a large part of variation between people

1. McRae et al Genome Biology (2014).
2. van Dongen J. et al. Nature Communications (2016).

~10% of genome-wide methylation sites in whole blood (Illumina 450k array): variance components change with age



Total variance of DNA methylation (Vp)

Genetic variance (Va)

Environmental variance (Ve)

Heritability of DNA methylation $= \frac{Va}{Va + Ve}$

van Dongen J. et al. Nature Communications (2016).

Practical: Discordant monozygotic twins

Strong design to

- Identify epigenetic mechanisms involved in a disease / trait
- Identify epigenetic effects of environmental exposures

Monozygotic twins discordant for smoking

Okada, Haruko C., et al. "Facial changes caused by smoking: a comparison between smoking and nonsmoking identical twins." *Plastic and reconstructive surgery* 132.5 (2013): 1085-1092.

Practical: effects of smoking on the methylome

- Perform an EWAS of smoking in monozygotic twin pairs discordant for smoking status (current smoker & never smoker)
- Analysis: paired t-test in R
- **AIM**: To identify DNA methylation differences between smokers and non-smokers

→ Which genomic locations are differentially methylated in smokers?

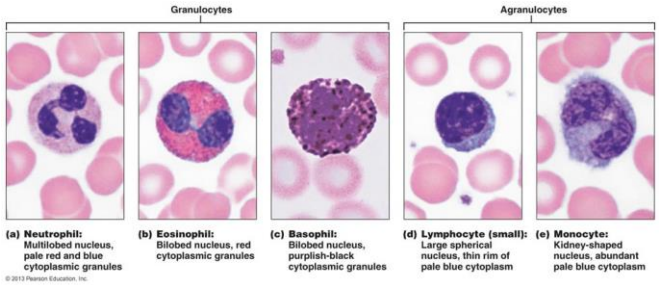
Methylation data: Illumina 450k array

- Methylation level residuals (adjusted for several covariates including white blood cell counts)
- Real dataset (subset of probes and twins) from the Netherlands Twin Register
- Quality control of the dataset is described in van Dongen et al 2016 Nature Communications.

Phenotype data

White blood cell counts: percentage of monocytes, neutrophils, basophils, lymphocytes, and eosinophils

→ Are there differences in white blood cell counts between the smoking and the non-smoking twin?



mkdir EWAS

cd EWAS

cp -r /faculty/jenny/2019/friday/* .

Open the R-script in R studio

Run line by line

Read the comments and follow the instructions in the comments (in the R-script)

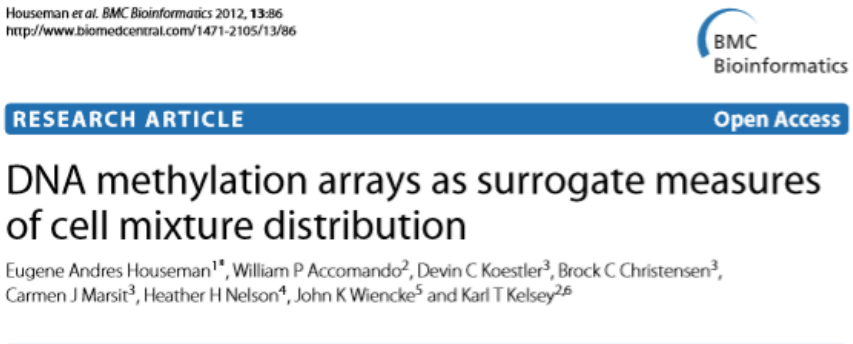
White blood cell counts

```
> results
           MeanDifference_current_minus_never     pvalue 95confint_L 95confint_H
Neut_Perc                              1.1300 0.53826841  -2.5510909  4.81109087
Lymph_Perc                            -0.2600 0.87437086  -3.5643963  3.04439631
Mono_Perc                             -0.7075 0.02997538  -1.3427447 -0.07225533
Eos_Perc                              -0.1300 0.79641609  -1.1422838  0.88228376
Baso_Perc                             -0.0275 0.86788932  -0.3597032  0.30470319
> 
```

- Twins who smoke tend to have lower levels of monocytes
- Different white blood cell types each have distinct methylation patterns
- Therefore, when comparing DNA methylation (in whole blood) between smokers and non-smokers, it is important to correct for white blood cell counts!

Cell count prediction tools

- R packages exist for predicting cellular proportions based on your DNA methylation data
- Use these if you don't have measured cell count data!



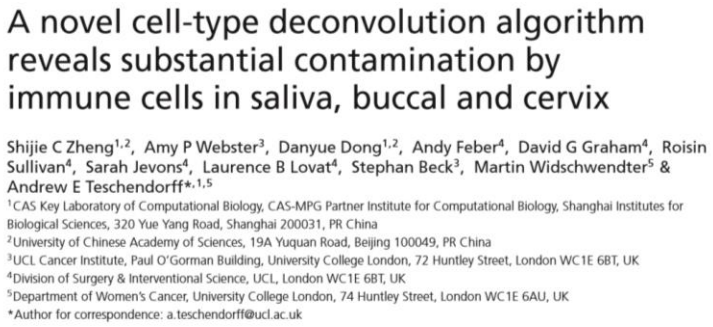
Houseman *et al. BMC Bioinformatics* 2012, **13**:86
http://www.biomedcentral.com/1471-2105/13/86

BMC Bioinformatics

RESEARCH ARTICLE Open Access

DNA methylation arrays as surrogate measures of cell mixture distribution

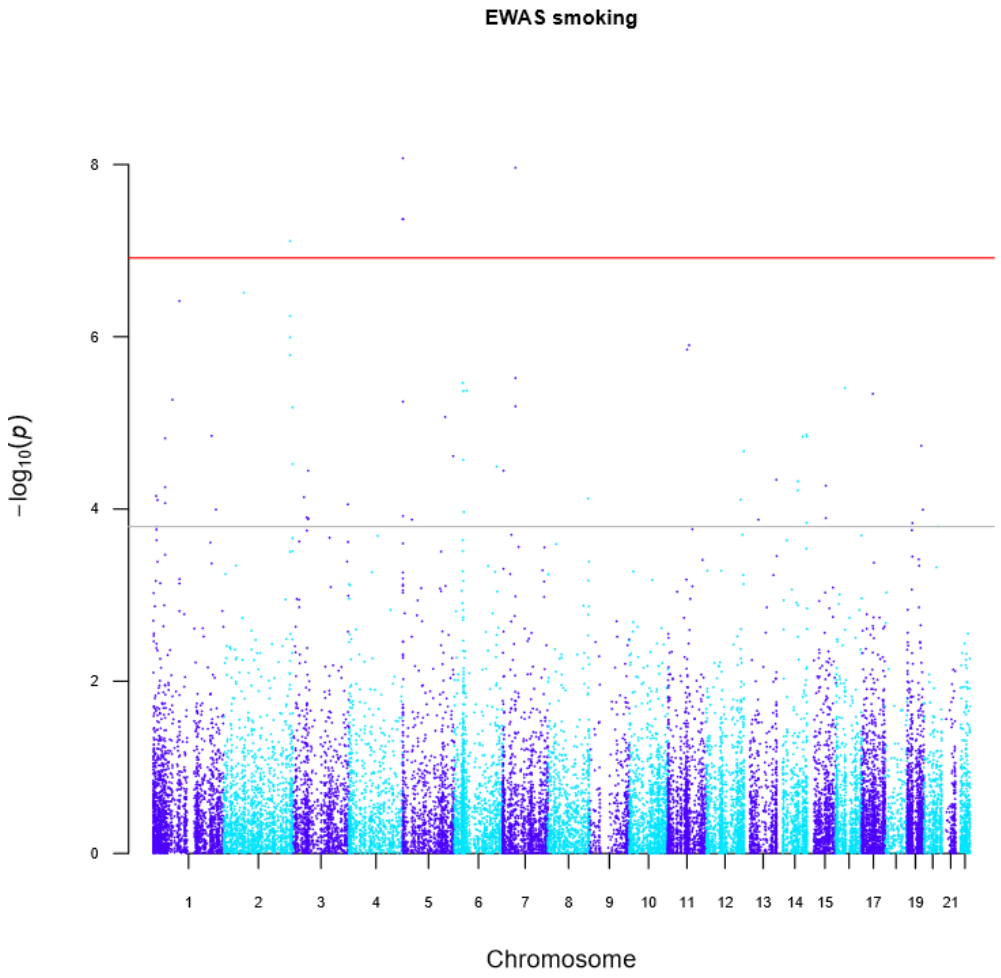
Eugene Andres Houseman[1*], William P Accomando[2], Devin C Koestler[3], Brock C Christensen[3], Carmen J Marsit[3], Heather H Nelson[4], John K Wiencke[5] and Karl T Kelsey[2,6]



A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix

Shijie C Zheng[1,2], Amy P Webster[3], Danyue Dong[1,2], Andy Feber[4], David G Graham[4], Roisin Sullivan[4], Sarah Jevons[4], Laurence B Lovat[4], Stephan Beck[3], Martin Widschwendter[5] & Andrew E Teschendorff*[,1,5]

[1]CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, 320 Yue Yang Road, Shanghai 200031, PR China
[2]University of Chinese Academy of Sciences, 19A Yuquan Road, Beijing 100049, PR China
[3]UCL Cancer Institute, Paul O'Gorman Building, University College London, 72 Huntley Street, London WC1E 6BT, UK
[4]Division of Surgery & Interventional Science, UCL, London WC1E 6BT, UK
[5]Department of Women's Cancer, University College London, 74 Huntley Street, London WC1E 6AU, UK
*Author for correspondence: a.teschendorff@ucl.ac.uk

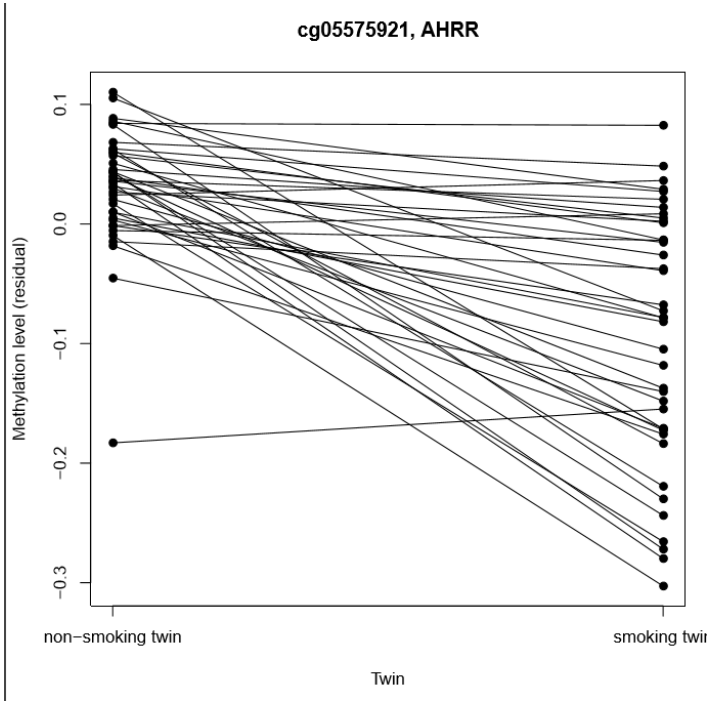


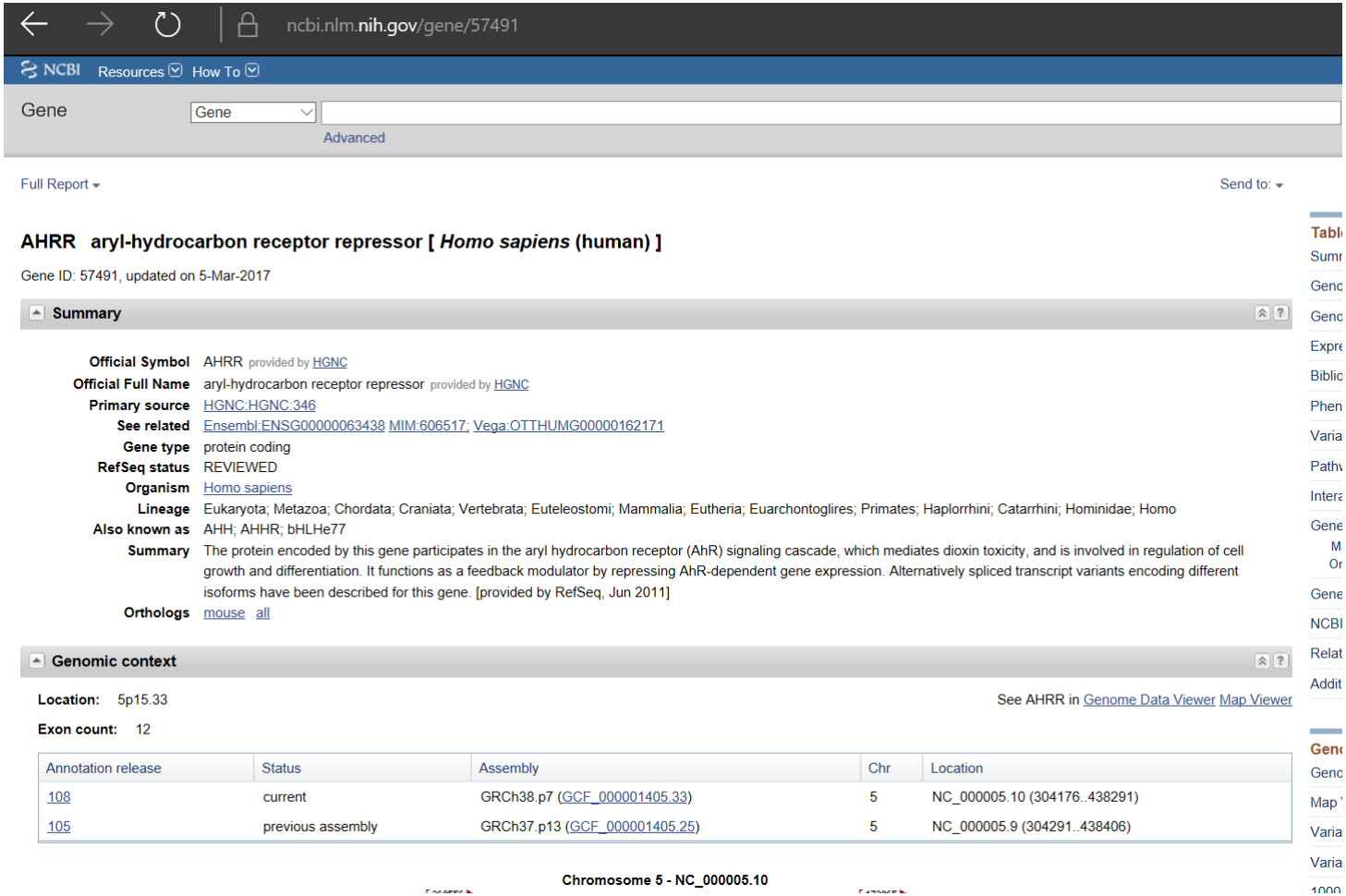
Bonferroni threshold (1×10^{-7})

FDR 5%

- In which genes are the top hits located?
- What is the function of the gene associated with the CpG with the lowest p-value (look up online, e.g. NCBI, OMIM)

	IlmnID	CHR	MAPINFO	UCSC_RefGene_Name	UCSC_RefGene_Group	cgid	MeanDifference_current_minus_never	pvalue
cg05575921	cg05575921	5	373378	AHRR	Body	cg05575921	-0.12446768	8.457223e-09
cg19089201	cg19089201	7	45002287	MYO1G	3'UTR	cg19089201	0.03772695	1.089340e-08
cg21161138	cg21161138	5	399360	AHRR	Body	cg21161138	-0.04359654	4.291994e-08
cg23067299	cg23067299	5	323907	AHRR	Body	cg23067299	0.02122614	4.315941e-08
cg21566642	cg21566642	2	233284661			cg21566642	-0.08473326	7.736723e-08




ncbi.nlm.nih.gov/gene/57491
NCBI Resources How To
Gene
Gene
Advanced
Full Report
Send to:
AHRR aryl-hydrocarbon receptor repressor [Homo sapiens (human)]
Gene ID: 57491, updated on 5-Mar-2017
Summary
Official Symbol AHRR provided by HGNC
Official Full Name aryl-hydrocarbon receptor repressor provided by HGNC
Primary source HGNC:HGNC:346
See related Ensembl:ENSG00000063438 MIM:606517; Vega:OTTHUMG00000162171
Gene type protein coding
RefSeq status REVIEWED
Organism Homo sapiens
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as AHH; AHHR; bHLHe77
Summary The protein encoded by this gene participates in the aryl hydrocarbon receptor (AhR) signaling cascade, which mediates dioxin toxicity, and is involved in regulation of cell growth and differentiation. It functions as a feedback modulator by repressing AhR-dependent gene expression. Alternatively spliced transcript variants encoding different isoforms have been described for this gene. [provided by RefSeq, Jun 2011]
Orthologs mouse all
Genomic context
Location: 5p15.33
See AHRR in Genome Data Viewer Map Viewer
Exon count: 12
Annotation release Status Assembly Chr Location
108 current GRCh38.p7 (GCF_000001405.33) 5 NC_000005.10 (304176..438291)
105 previous assembly GRCh37.p13 (GCF_000001405.25) 5 NC_000005.9 (304291..438406)
Chromosome 5 - NC_000005.10

Circulation: Cardiovascular Genetics

Hello, Gu

MY ALERTS | SIGN IN

HOME | ABOUT THIS JOURNAL | ALL ISSUES | SUBJECTS | BROWSE FEATURES | RESOURCES | AHA JOURNALS

ORIGINAL ARTICLE

Epigenetic Signatures of Cigarette Smoking

Roby Joehanes, Allan C. Just, Riccardo E. Marioni, Luke C. Pilling, Lindsay M. Reynolds, Pooja R. Mandaviya, Weihua Guan, Tao Xu, Cathy E. Elks, Stella Aslibekyan, Hortensia Moreno-Macias, Jennifer A. Smith, Jennifer A. Brody, Radhika Dhingra, Paul Yousefi, James S. Pankow, Sonja Kunze, Sonia H. Shah, Allan F. McRae, Kurt Lohman, Jin Sha, Devin M. Absher, Luigi Ferrucci, Wei Zhao, Ellen W. Demerath, Jan Bressler, Megan L. Grove, Tianxiao Huan, Chunyu Liu, Michael M. Mendelson, Chen Yao, Douglas P. Kiel, Annette Peters, Rui Wang-Sattler, Peter M. Visscher, Naomi R. Wray, John M. Starr, Jingzhong Ding, Carlos J. Rodriguez, Nicholas J. Wareham, Marguerite R. Irvin, Degui Zhi, Myrto Barrdahl, Paolo Vineis, Srikant Ambatipudi, André G. Uitterlinden, Albert Hofman, Joel Schwartz, Elena Colicino, Lifang Hou, Pantel S. Vokonas, Dena G. Hernandez, Andrew B. Singleton, Stefania Bandinelli, Stephen T. Turner, Erin B. Ware, Alicia K. Smith, Torsten Klengel, Elisabeth B. Binder, Bruce M. Psaty, Kent D. Taylor, Sina A. Gharib, Brenton R. Swenson, Liming Liang, Dawn L. DeMeo, George T. O'Connor, Zdenko Herceg, Kerry J. Ressler, Karen N. Conneely, Nona Sotoodehnia, Sharon L. R. Kardia, David Melzer, Andrea A. Baccarelli, Joyce B. J. van Meurs, Isabelle Romieu, Donna K. Arnett, Ken K. Ong, Yongmei Liu, Melanie Waldenberger, Ian J. Deary, Myriam Fornage, Daniel Levy, Stephanie J. London

DOI https://doi.org/10.1161/CIRCGENETICS.116.001506
Circulation: Cardiovascular Genetics. 2016;9:436-447
Originally published September 20, 2016

- 2623 Bonferroni significant (current vs never smokers), 18 760 CpGs at false discovery rate <0.05.
- N=15 907

"Genes annotated to these CpGs were enriched for associations with several smoking-related traits in genome-wide studies including pulmonary function, cancers, inflammatory diseases, and heart disease."

Disease variants alter transcription factor levels and methylation of their binding sites

Marc Jan Bonder, René Luijk, Daria V Zhernakova, Matthijs Moed, Patrick Deelen, Martijn Vermaat, Maarten van Iterson, Freerk van Dijk, Michiel van Galen, Jan Bot, Roderick C Slieker, P Mila Jhamai, Michael Verbiest, H Eka D Suchiman, Marijn Verkerk, Ruud van der Breggen, Jeroen van Rooij, Nico Lakenberg, Wibowo Arindrarto, Szymon M Kielbasa, Iris Jonkers, Peter van 't Hof, Irene Nooren, Marian Beekman, Joris Deelen + *et al.*

Affiliations | Contributions | Corresponding authors

Nature Genetics **49**, 131–138 (2017) | doi:10.1038/ng.3721

Received 02 December 2015 | Accepted 18 October 2016 | Published online 05 December 2016

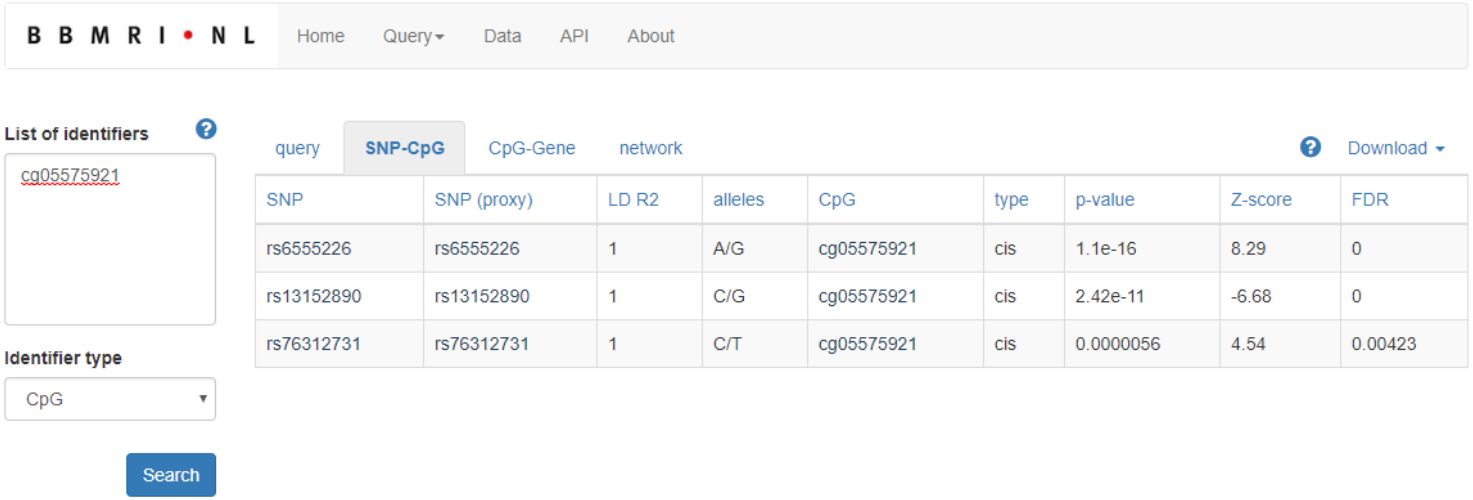
N= 3,841 individuals

- Is the methylation level at our smoking-associated CpGs associated with **gene expression** in cis?

- Is the methylation level at our smoking-associated CpGs influenced by SNPs (**methylation QTLs**)?

http://bbmri.researchlumc.nl/atlas/#query

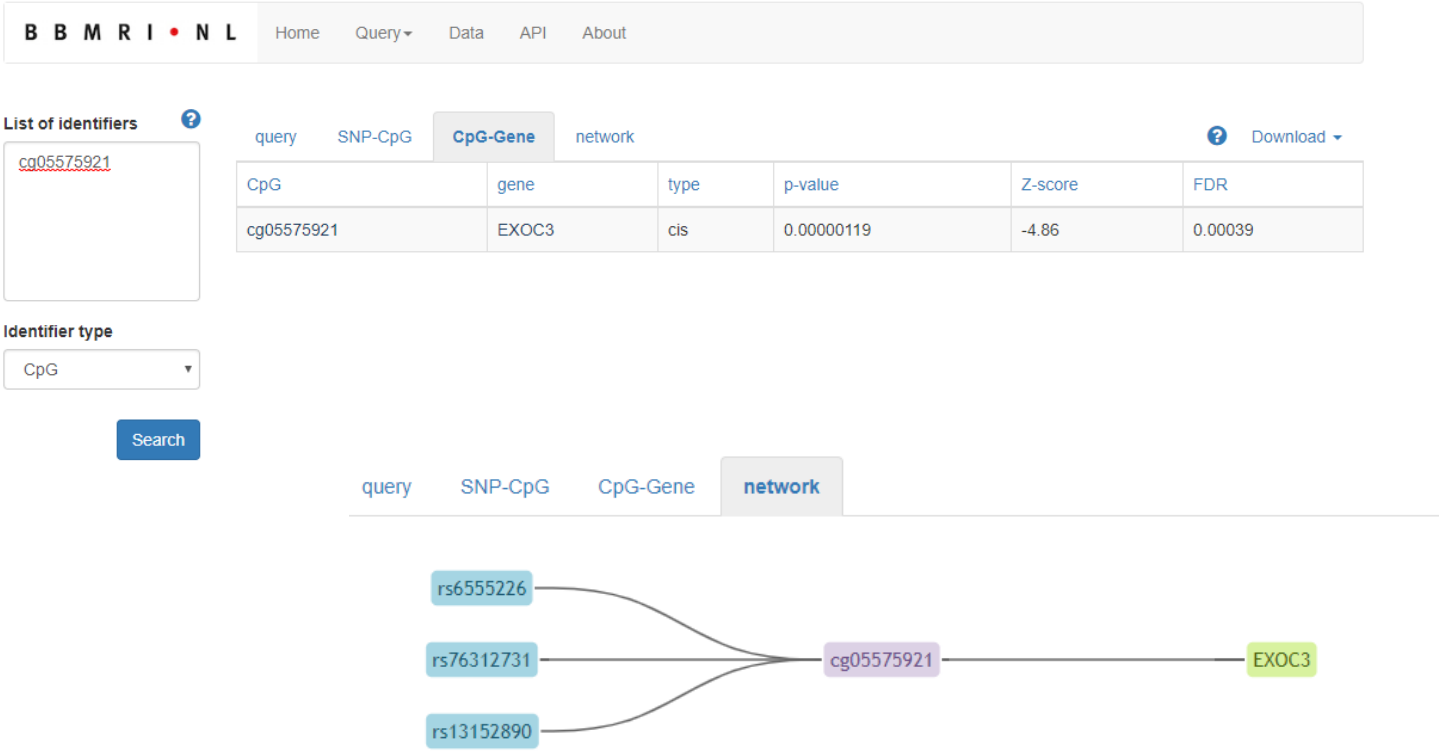
Query - top CpG (cg05575921)



SNP	SNP (proxy)	LD R2	alleles	CpG	type	p-value	Z-score	FDR
rs6555226	rs6555226	1	A/G	cg05575921	cis	1.1e-16	8.29	0
rs13152890	rs13152890	1	C/G	cg05575921	cis	2.42e-11	-6.68	0
rs76312731	rs76312731	1	C/T	cg05575921	cis	0.0000056	4.54	0.00423

note: no significant trans mQTLs for this CpG

Query - top CpG (cg05575921)



GoDMC (Genetics of DNA methylation) consortium

- http://www.godmc.org.uk/
- In progress
- 32,851 individuals from 37 population-based and disease datasets
- Blood DNA methylation

Concluding remarks

- DNA methylation –> potential molecular intermediate of environmental exposures and genetic variants
- Tissue-specific
- Promising biomarker of environmental exposures
- Options to examine causality
  - Additional data from former smokers allows to examine **reversibility**
  - Reversible genes (Vink et al 2015): gene expression/methylation goes back to the level of non-smokers in individuals who quit smoking

  → This is in line with a causal effect of smoking on DNA methylation/gene expression
  - Mendelian randomization (Dekkers et al 2016)