Epigenetics

Jenny van Dongen
Vrije Universiteit (VU) Amsterdam
j.van.dongen@vu.nl

Boulder, Friday march 10, 2017

Epigenetics

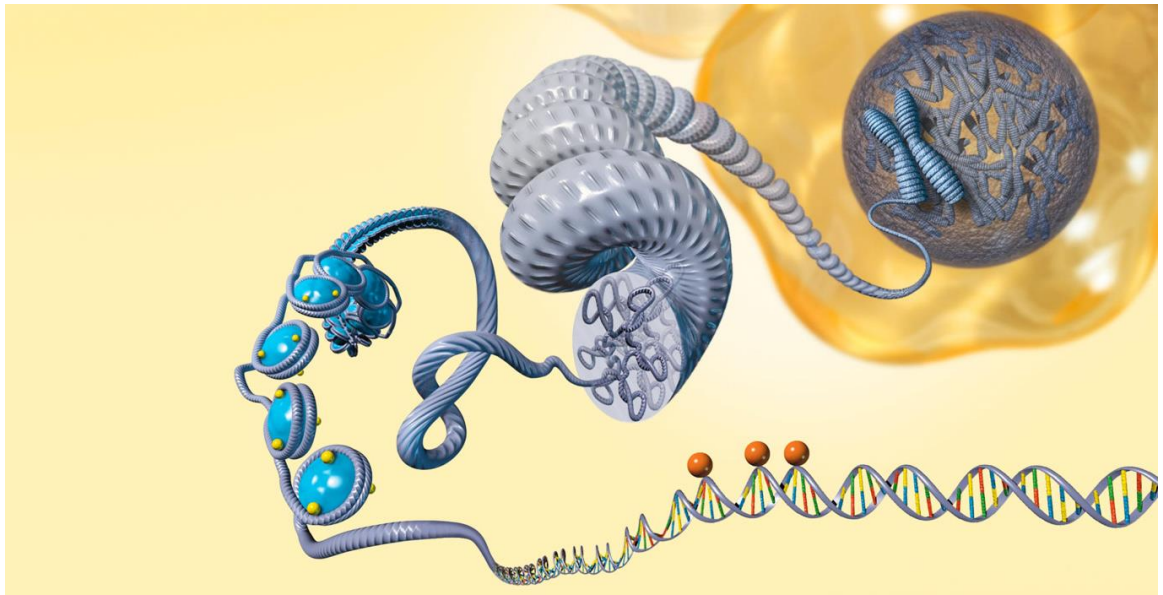
Epigenetics= The study of molecular mechanisms that influence the activity of gene expression and that are transmitted across cell division.
[*definition by Bird 2007 Nature*]

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

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

GGTGTACTCGTTGTCTAAGG
GCCTGGACTAGCTGGGACTT
A**G**GCTTCTGCCGGTT**C**CAAC
ACTCTGGTGCTGGAAGGCTG
GACTTGGGTGACTCAAGTTCC
CTCTTCCTACTGCAATGCAAG
AAAATAACAAAAGAAGTATGT
ATACCTTTAAGTATCTCAAAG
AGCTATCTCAGCTTCTGAATT
TCCTTCTAGGG**C**ACCTCTTCC
TGCGGTGTACTCGTTGTCTAA
GGGCCTGGACTAGCTGGGAC
TTAGGCTTCTGCCGGTTCCAA
CACTCTGGTGCTGGAAGGCT
G**G**ACTTGGGTGACTCAAGTT
CCCTCTTCCTACTGCAATGCA
AGAAAATACACAAAAGAAGTA
TGTATACCTTTAAGTATCTCAA
AGAGC**T**ATCTCAGCTTCT**G**AA
TTTCCTTCTAGGGCACCTC

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



* e.g. DNA methylation, histone modifications, microRNAs

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

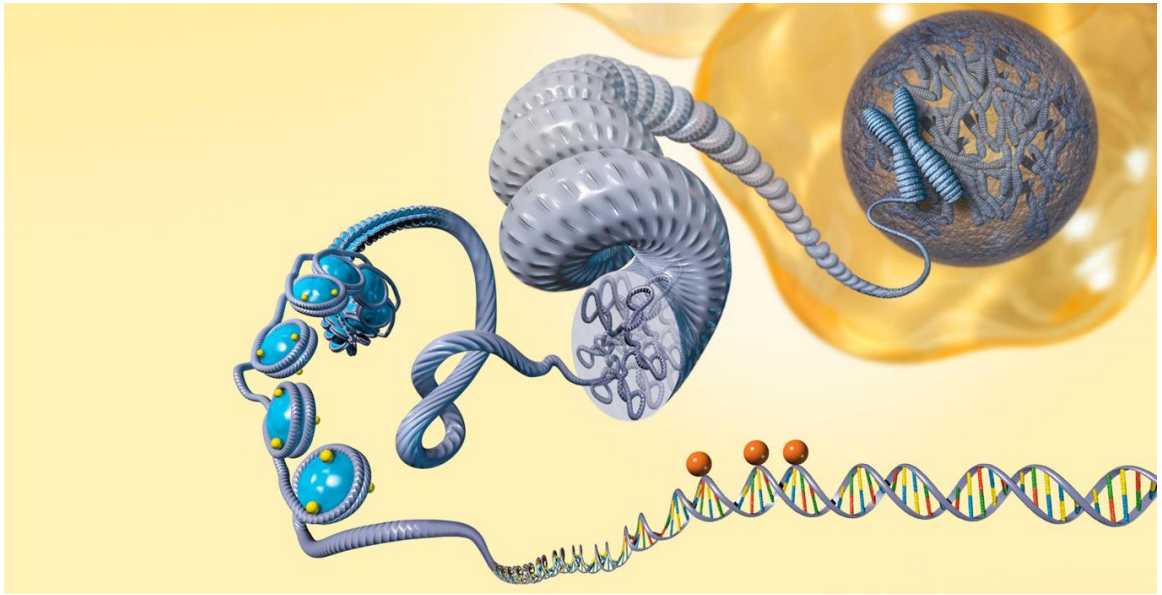
GGTGTACTCGTTGTCTAAGG
GCCTGGACTAGCTGGGACTT
A**G**GCTTCTGCCGGTT**C**CAAC
ACTCTGGTGCTGGAAGGCTG
GACTTGGGTGACTCAAGTTCC
CTCTTCCTACTGCAATGCAAG
AAAATAACAAAAGAAGTATGT
ATACCTTTAAGTATCTCAAAG
AGCTATCTCAGCTTCTGAATT
TCCTTCTAGGG**C**ACCTCTTCC
TGCGGTGTACTCGTTGTCTAA
GGGCCTGGACTAGCTGGGAC
TTAGGCTTCTGCCGGTTCCAA
CACTCTGGTGCTGGAAGGCT
G**G**ACTTGGGTGACTCAAGTT
CCCTCTTCCTACTGCAATGCA
AGAAAATACACAAAAGAAGTA
TGTATACCTTTAAGTATCTCAA
AGAGC**T**ATCTCAGCTTCT**G**AA
TTTCCTTCTAGGGCACCTC

Epigenome: can be dynamic

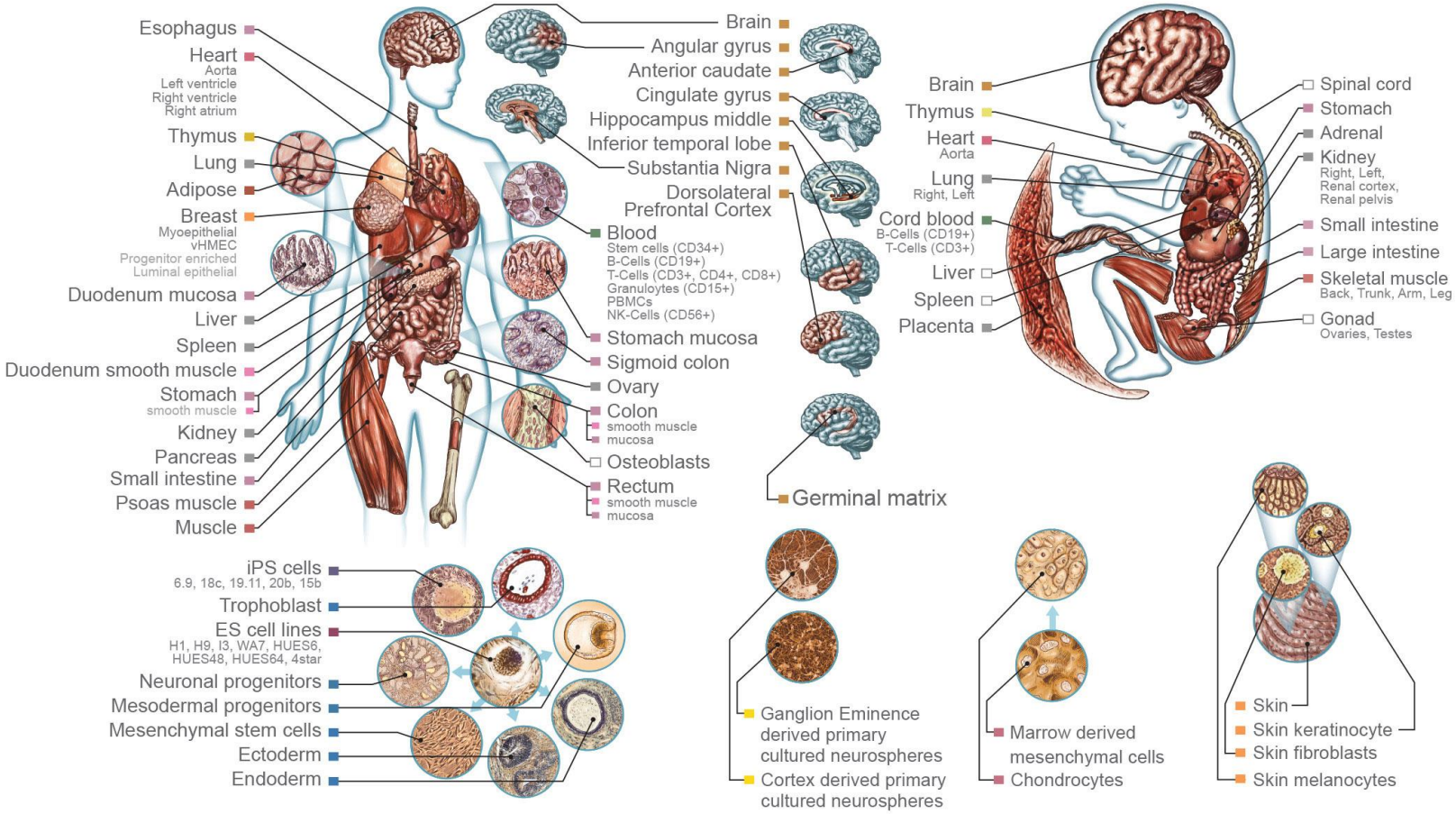
Programmed epigenetic changes (development and tissue differentiation)

Substantial changes in DNA methylation with ageing

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

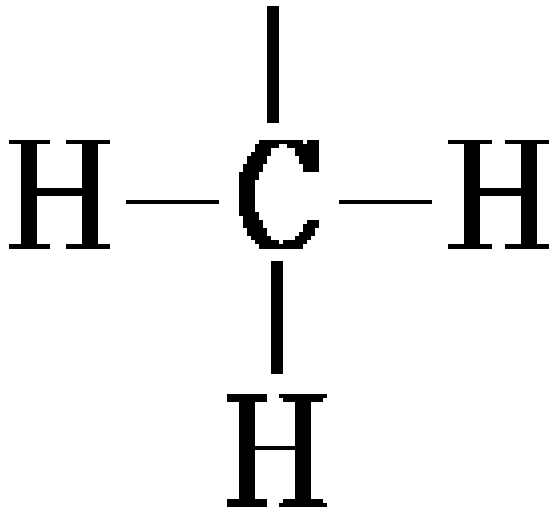


Each cell has its own *epigenome*



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

CH_3 – methyl



DNA methylation

- The covalent attachment of methyl-groups to DNA
- In host human cell types: only at cytosine bases next to guanine (*CpG sites*)
  - at promoters: usually represses gene expression
  - in gene bodies: may regulate alternative splicing
  - at enhancers - strongest relation to expression
- Most studied epigenetic mark in epigenome-wide association studies (stable molecular mark: no fresh tissue required)

Measuring genome-wide methylation

Illumina 450 array in brief

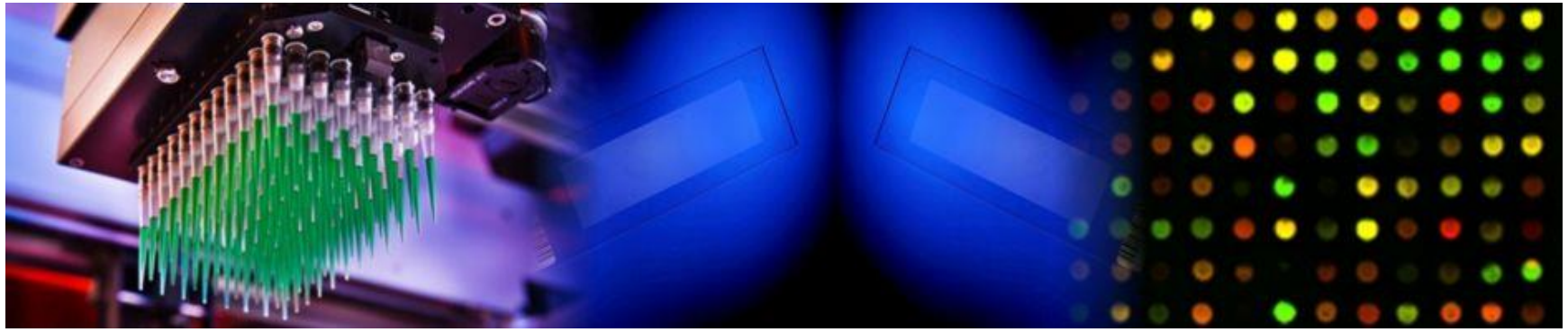
- 485,000 methylation sites (of the ~30 million CpGs)
- covers 99% of RefSeq genes
- average of 17 CpG sites per gene
- promoter, 5'UTR, first exon, gene body, and 3'UTR.

PRACTICAL

Illumina 850k array in brief

- > 850,000 methylation sites
- high coverage of enhancer regions
- ENCODE open chromatin
- ENCODE transcription factor binding sites

- Cost-effective options
- Often used in EWAS



Sequencing

- E.g. Bisulphite (BS) sequencing

DNA methylation data

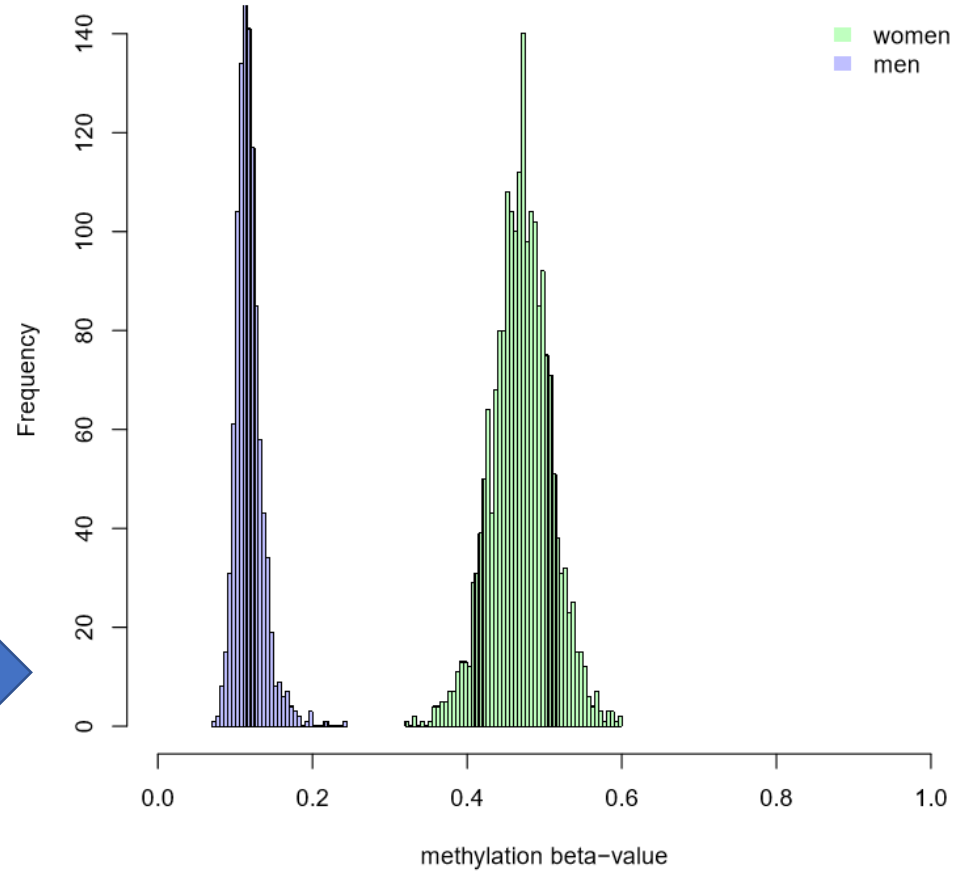
Illumina 450k array data on peripheral blood

- For > 450.000 sites:
- **methylation level**: proportion of methylated alleles.
- Continuous trait, range: 0-1
- This is because DNA extracted from blood comes from many different cells
- In some cells, the position may be methylated, while it is unmethylated in others

When looking at the methylation level at 1 location in 1 sample:

- 0= all DNA in the sample was not methylated at this position
- 1= all DNA in the sample was methylated at this position

- Distribution of the methylation level at 1 CpG site (chromosome X)
- In women, 1 X chromosome is methylated (X-chromosome inactivation)



Epigenome-wide association studies (EWAS)

Aim: To identify genomic regions whose epigenetic regulation:

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

Since 1-2 years: large-scale EWAS meta-analysis projects

-DNA from peripheral tissues (blood, buccal)

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.
- N=15 907

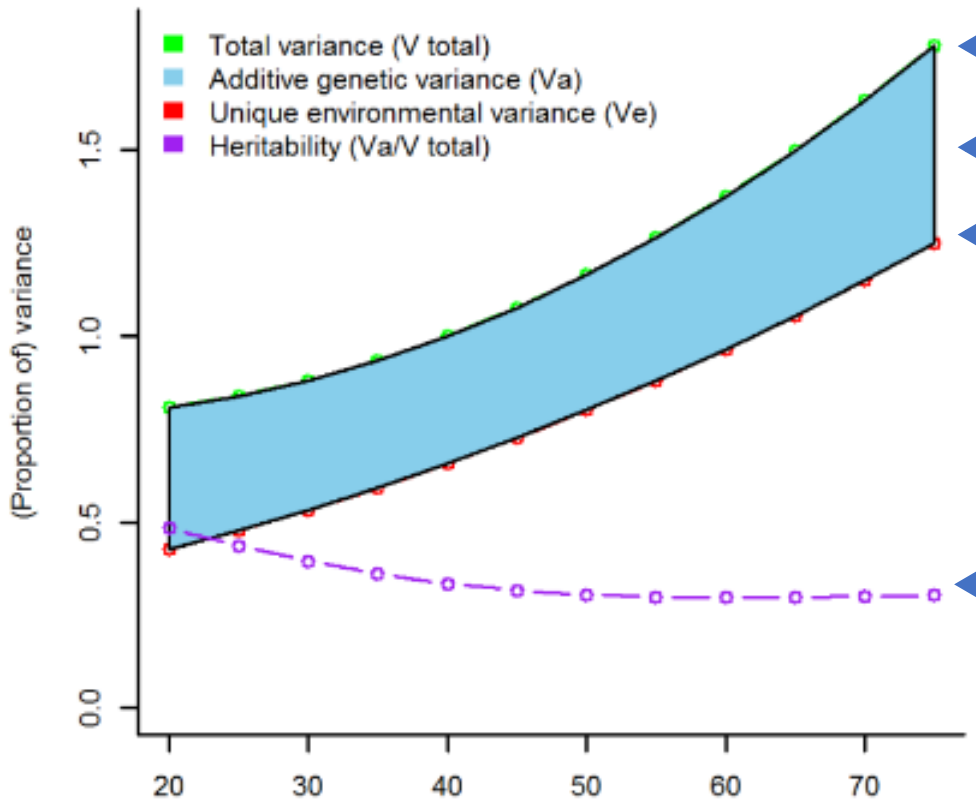
Variation in DNA methylation between people

- **Twin studies**: Average heritability of methylation levels in adults across 450.000 sites in the genome~19%
  (Illumina 450k array, peripheral blood, refs 1,2)

  - 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: 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?

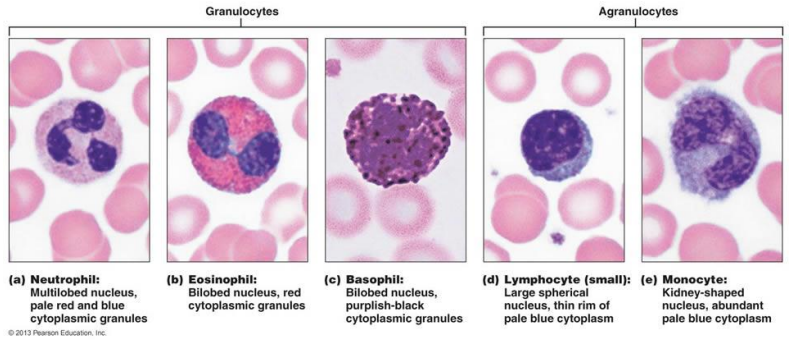
Methylation data: Illlumina 450k array

- Methylation level residuals (adjusted for several covariates including white blood cell counts)
- Based on a real methylation 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/2017/friday/* .
```

Open the R-script in R studio

White blood cell counts

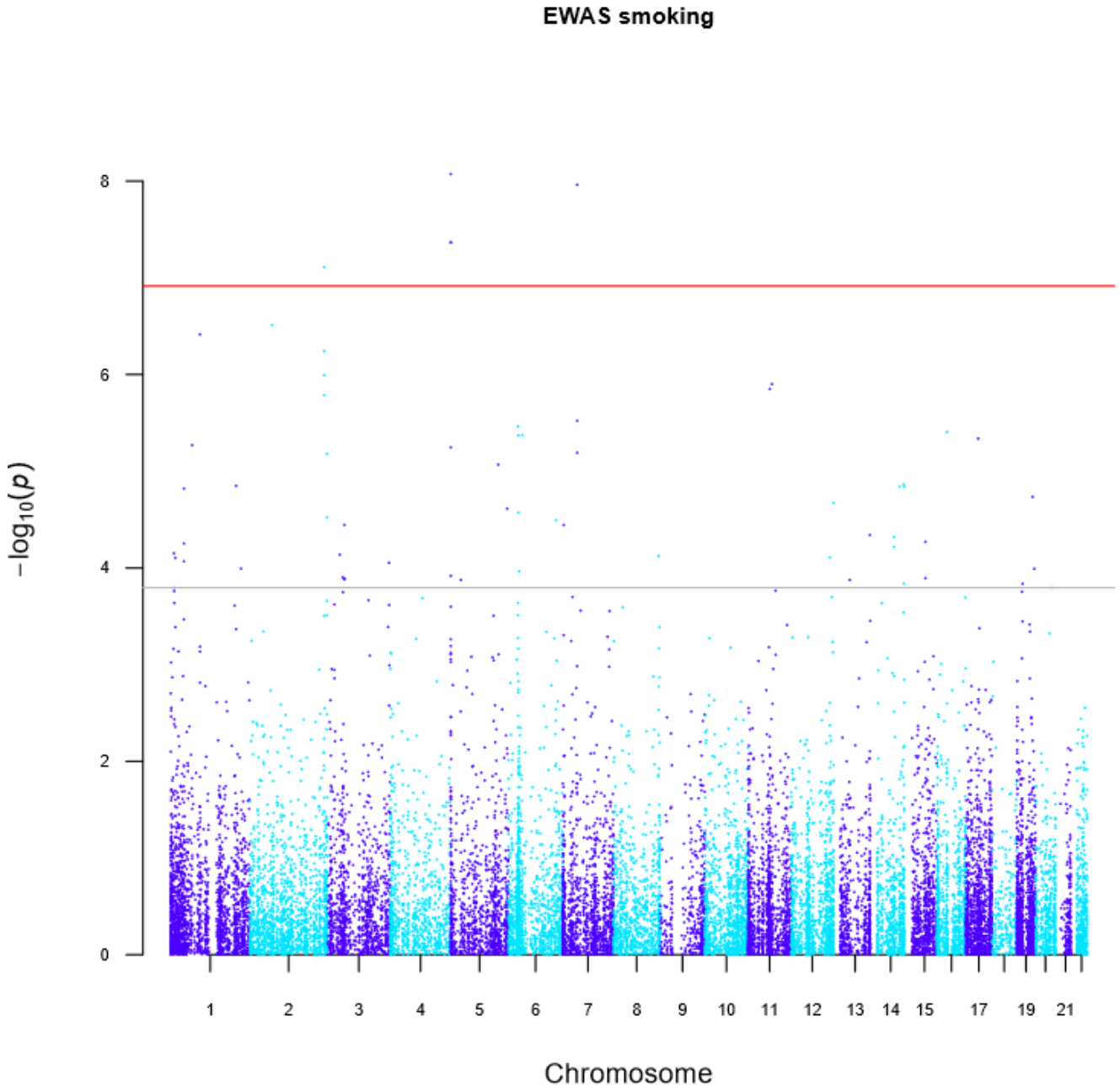
```
> results
           MeanDifference_current_minus_never     pvalue 95confint_L 95confint_H
Neut_Perc                              1.1300 0.53826841  -2.5510909  4.81109087
Lymph_Perc                            -0.2600 0.87437086  -3.5643963  3.04439631
Mono_Perc                             -0.7075 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!

Illumina 450k array annotation file

UCSC_RefGene_Group: Location of CpG site within gene

- TSS200 = within 200 bp of transcriptional start site
- TSS1500= within 1500 bp of transcriptional start site
- 1stExon
- 3'UTR
- Body= gene body

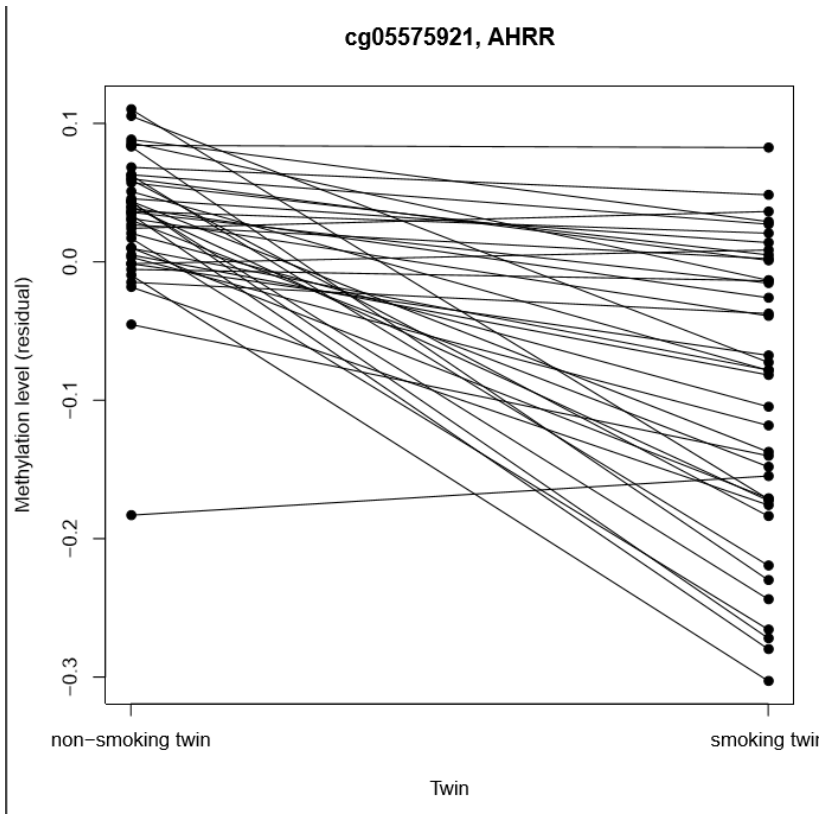


Bonferroni threshold (1×10^{-7})

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



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

Tabl
Summ
Genc
Genc
Expre
Biblic
Phen
Varia
Pathv
Intera
Gene
M
Or
Gene
NCBI
Relat
Addit
Genc
Genc
Map
Varia
Varia
1000

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.
- 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."

- 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**)?

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

http://genenetwork.nl/biosqtlbrowser/

Query our top CpG (cg05575921)

Trans-meQTLs

P-value	SNP	SNP Chr.	SNP Chr. position	CpG	CpG Chr.	CpG Chr. position	SNP Alleles	Assesed Allele	Z-score	Gene name	FDR

No records found.

Cis-meQTLs

P-value	SNP	SNP Chr.	SNP Chr. Position	CpG	CpG Chr.	CpG Chr. position	SNP Alleles	Assesed Allele	Z-score	Gene name	FDR
1.10E-16	rs6555226	5	389589	cg05575921	5	373402	A/G	G	8.29	EXOC3	0.00
2.42E-11	rs131528905		489598	cg05575921	5	373402	C/G	G	-6.68	EXOC3	0.00
5.60E-06	rs763127315		401734	cg05575921	5	373402	C/T	T	4.54	EXOC3	0.00

Cis-eQTMs

P-value	CpG	CpG Chr.	CpG Chr. Position	Probe	Probe Chr.	Probe Chr. position	Z-score	Gene name	FDR
1.19E-06	cg05575921	5	373378	ENSG00001801045		443273	-4.86	EXOC3	0.00

Concluding remarks

- DNA methylation –> potential molecular intermediate of environmental exposures and genetic variants
- Causality: see Vink et al 2015
  - additional data from former smokers → allows to examine reversibility of smoking effects
  - Reversible genes: 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