Singmann *et al. Epigenetics & Chromatin (2015) 8:43*
DOI 10.1186/s13072-015-0035-3

Epigenetics & Chromatin

RESEARCH Open Access

Characterization of whole-genome autosomal differences of DNA methylation between men and women

Paula Singmann[1,2*], Doron Shem-Tov[3], Simone Wahl[1,2,4], Harald Grallert[1,2,4], Giovanni Fiorito[5,6], So-Youn Shin[7,8], Katharina Schramm[9,10], Petra Wolf[9,10], Sonja Kunze[1,2], Yael Baran[3], Simonetta Guarrera[5,6], Paolo Vineis[5,11], Vittorio Krogh[12], Salvatore Panico[13], Rosario Tumino[14], Anja Kretschmer[1,2], Christian Gieger[1,2], Annette Peters[2,4], Holger Prokisch[9,10], Caroline L. Relton[7,8], Giuseppe Matullo[5,6], Thomas Illig[1,15,16], Melanie Waldenberger[1,2] and Eran Halperin[3,17,18*]

Abstract

Background: Disease risk and incidence between males and females reveal differences, and sex is an important component of any investigation of the determinants of phenotypes or disease etiology. Further striking differences between men and women are known, for instance, at the metabolic level. The extent to which men and women vary at the level of the epigenome, however, is not well documented. DNA methylation is the best known epigenetic mechanism to date.

Results: In order to shed light on epigenetic differences, we compared autosomal DNA methylation levels between men and women in blood in a large prospective European cohort of 1799 subjects, and replicated our findings in three independent European cohorts. We identified and validated 1184 CpG sites to be differentially methylated between men and women and observed that these CpG sites were distributed across all autosomes. We showed that some of the differentially methylated loci also exhibit differential gene expression between men and women. Finally, we found that the differentially methylated loci are enriched among imprinted genes, and that their genomic location in the genome is concentrated in CpG island shores.

Conclusion: Our epigenome-wide association study indicates that differences between men and women are so substantial that they should be considered in design and analyses of future studies.

Keywords: DNA methylation, EWAS, Sex, Enrichment analysis, CpG, Imprinting, KORA, ALSPAC, EPICOR

Background

There is no doubt that men and women are different. Differences exist in risk and incidence of diseases between males and females, and sex is an important component of any investigation of the determinants of phenotype or disease etiology [1, 2]. At the molecular level, it has been shown that metabolic profiles of men and women differ substantially [3], whereas genomic differences on the SNP level are not reproducible [4]. However, it remains largely unknown to what extent men and women differ at the epigenomic level. Shedding light on the differences between men and women in terms of DNA methylation (DNAm) is particularly important in conducting epigenome-wide association studies (EWAS). Insights on extent and distribution of sex-dependent DNAm can potentially improve the identification of disease- or phenotype-related methylation signatures.

The relation between DNAm and sex in humans has been studied previously [5–8]. These studies, however, were limited in their scope and in their statistical power.

*Correspondence: paula.singmann@helmholtz-muenchen.de; heran@post.tau.ac.il
[2] Institute of Epidemiologie II, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Neuherberg, Germany
[18] International Computer Science Institute, Berkeley, CA, USA
Full list of author information is available at the end of the article

BioMed Central

© 2015 Singmann et al. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

In particular, several earlier studies considered 'global' measures of DNAm that captured repeated, noncoding regions in the genome (e.g., LINE-1 and Alu repeats [5, 9–12]). Others have focused on specific pathways or loci (e.g., cancer sites [13], specific genes [11, 14], low numbers of CpGs across the genome [15]), or restricted their analysis to certain chromosomes [16, 17].

In the current study, we aimed to systematically analyze whole blood DNAm differences between sexes across the genome in a large population-based cohort, and to validate our results in three additional cohorts. In addition, we sought to characterize the genomic loci that show substantial DNAm differences between men and women by means of different enrichment analyses and expression analysis.

Results

Discovery of sex-methylation associations

In our discovery sample from the population-based KORA F4 study, we sought to identify DNAm differences between 877 men and 922 women at 470,920 autosomal CpGs. Details on the cohort are given in Table 1. Association analysis of normalized methylation data (beta-values) from 391,885 CpGs revealed 11,010 sex-methylation associations (SMAs) ($p < 1.26E-07$) with some of them being highly significant ($p < 1E-300$). The sex-associated CpGs were spread across the autosomal genome (Fig. 1). These and all subsequent significance levels were corrected for multiple testing according to the Bonferroni method. Additional file 1 provides a flow-chart of our analyses.

Validation of sex-methylation associations in whole blood

We took the set of 11,010 SMAs forward to replication in three different cohorts coming from three different populations in three different geographic locations (Table 1, further details in the methods section). The replication results were meta-analyzed using a random-effects model, revealing 1184 CpGs ($p < 4.89E-06$). Effect sizes of revealed for these CpGs are in a range between −0.89 and 0.73 (given as theta values in Additional file 2). The histogram in Additional file 2 shows the frequency distribution of the effect sizes. Of note, the majority of the not replicated SMAs still showed consistent effect direction in the replication as in the discovery step ($p < 1E-200$, Binomial test), and SMAs were highly correlated between discovery sample and each of the three replication cohorts (correlation coefficients between KORA F4 and KORA F3, ALSPAC, EPIC, respectively, were 0.88, 0.65, 0.70), suggesting that the majority of the discovered SMAs are unlikely to be false positives (Fig. 2; Additional file 2).

Validation in buccal epithelial tissue

Since the identification of SMAs in whole blood may be biased by cell-type composition or other tissue-specific biases, we compared the SMAs found in KORA F4 to a public dataset on DNAm measured in buccal epithelial cells from 15-year-old adolescents ($n = 109$, 60 % females) with the Illumina Infinium HumanMethylation27 BeadChip [18]. Since the buccal cells study was performed on the HumanMethylation27 BeadChip, only a subset of the sites were measured in both studies and passed the KORA quality criteria (see "Methods"): a set of 22,773 CpGs remained. Out of these 22,773 CpGs, 912 were significantly associated with sex in KORA F4, and 96 CpGs were significantly associated with sex in the buccal cells study (both after Bonferroni with $n = 22{,}773$, resulting significance level $p = 2.20E-6$). Despite the low rate of associations observed in the buccal cells study, we note that out of the 96 sex-associated CpGs in the buccal cells study, 16 CpGs were also associated with sex in KORA F4 (Additional file 3) (p value <2.4E−07, hypergeometric test). Notably, most associations in the KORA F4 were not replicated in the buccal cells study, potentially due to the limited power, with the sample size being small ($n = 109$), and due to the differences between the populations, particularly the age distribution of the subjects.

Enrichment of SMAs in specific CpG island locations

We further tested whether there are specific regions in the genome that are enriched for SMAs by inspecting the tendency of SMAs to occur at specific locations with respect to CpG islands (CGIs). Since correlations between CpG sites can inflate the p value of an enrichment analysis, we randomly picked each site to participate in this analysis with probability 0.2. A set of 78,734 sites was therefore randomly chosen, out of which 231 sites are a subset of the 1184 CpGs identified in the meta-analysis. Table 2 shows a clear concentration of SMAs in CGI shores (north and south combined; Pearson Chi square test for overrepresentation, p value <1E−16), as opposed to a lower rate of SMAs in CGIs, shelves (north and south combined) and open sea CpG sites [19]. When excluding CGI shores from the test, there was no longer significance for any enrichment of SMAs among the remaining CGI locations (Pearson Chi square test, p value 0.75). These enrichment test results imply that the SMAs are overrepresented in the CGI shores.

Enrichment of sex-methylation associations (SMAs) in imprinted genes

In order to investigate the distribution of SMAs across the genome further, we explored the possibility that DNAm varies in imprinted genes between sexes. Imprinted genes

Table 1 Population characteristics of discovery sample KORA F4, replication sample KORA F3, and replication studies ALSPAC and EPICOR

	Discovery			Replication								
	KORA F4			KORA F3			ALSPAC			EPICOR		
	All	Males	Females	All	Males	Females	All	Males	Females	All	Males	Females
N	1799	877	922	500	260	240	963	499	464	536	340	196
Age range (years)	32–81	35–79	32–81	34–83	34–77	35–83	14–19	14–19	15–19	35–71	35–64	35–71
Mean age	60 (8.89)	61 (8.92)	60 (8.85)	53 (9.64)	53 (9.80)	52 (9.48)	17.10 (1.00)	17.10 (1.10)	17.10 (1.00)	52.90 (7.38)	51.45 (6.95)	55.40 (7.45)
BMI (kg/m^2)	28.15 (4.80)	28.39 (4.27)	27.93 (5.24)	27.16 (4.54)	27.22 (3.97)	26.55 (5.07)	22.32 (3.84)	22.48 (4.08)	22.15 (3.55)	26.65 (3.87)	26.66 (3.07)	26.63 (4.96)
% Smokers	14.58	16.23	13.02	50.00	50.00	50.00	NA	NA	NA	32.83	35.29	28.57
Alcohol intake (g/days)	15.49 (20.43)	22.93 (24.17)	8.44 (12.53)	16.11 (19.59)	21.86 (21.79)	9.17 (14.10)	NA	NA	NA	17.99 (20.13)	24.08 (20.79)	7.43 (13.54)
Triglycerides (mmol/L)	1.52 (1.08)	1.72 (1.28)	1.34 (0.81)	1.88 (1.39)	2.18 (1.67)	1.56 (0.92)	0.84 (0.37)	0.83 (0.32)	0.84 (0.41)	1.72 (1.06)	1.81 (1.02)	1.57 (1.12)
Total cholesterol (mmol/L)	5.73 (1.02)	5.58 (1.00)	5.87 (1.01)	5.70 (0.99)	5.61 (1.03)	5.80 (0.93)	3.74 (0.68)	3.91 (0.69)	3.56 (0.63)	6.13 (1.19)	5.99 (1.17)	6.39 (1.19)
HDL (mmol/L)	1.46 (0.38)	1.32 (0.33)	1.59 (0.37)	1.51 (0.46)	1.33 (0.40)	1.69 (0.45)	1.25 (0.30)	1.32 (0.32)	1.18 (0.26)	1.48 (0.39)	1.38 (0.34)	1.65 (0.41)
LDL (mmol/L)	3.62 (0.91)	3.57 (0.88)	3.66 (0.93)	3.39 (0.86)	3.39 (0.84)	3.38 (0.87)	2.10 (0.60)	2.21 (0.63)	1.99 (0.55)	3.87 (1.05)	3.78 (1.02)	4.02 (1.09)
% Physical activity												
No	31.27	26.97	23.97	32.20	25.38	16.25	NA	NA	NA	21.27	15.88	30.61
Low	11.52	28.23	35.14	18.00	24.62	33.33	NA	NA	NA	43.84	41.18	48.47
Medium	31.78	12.23	10.85	28.80	18.85	17.08	NA	NA	NA	17.54	21.76	10.2
High	25.43	32.57	30.04	21.00	31.15	33.33	NA	NA	NA	17.35	21.18	10.72
% Diabetic	9.23	10.27	8.24	6.20	8.46	3.75	NA	NA	NA	1.68	1.18	2.55
% MI	3.562	5.59	1.63	1.60[a]	2.31[a]	0.83[a]	NA	NA	NA	52.98	53.23	52.55

Values given as mean (SD), otherwise indicated

[a] Myocardial infarction with hospital treatment (self-reported)

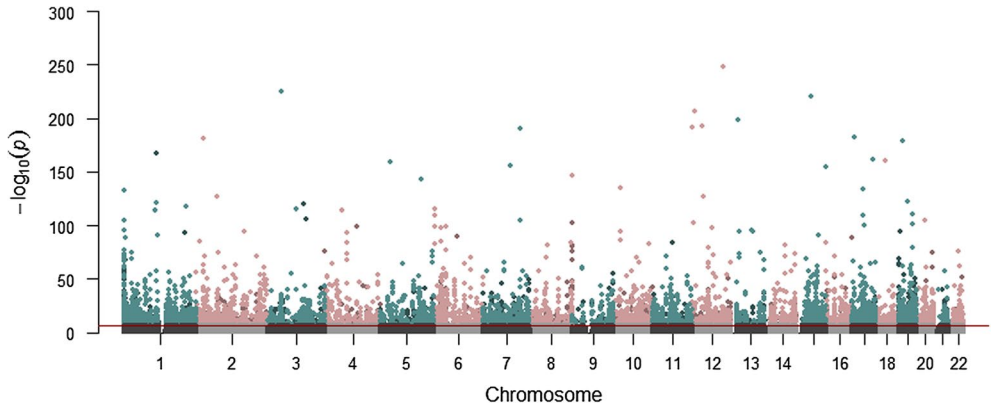

Fig. 1 Genome-wide sex differences in DNAm across the autosomes in the discovery study KORA F4 (Manhattan plot). Chromosomes (autosomes) are represented by alternating *colors* with the *lighter* color per chromosome representing hypermethylated CpGs and the *darker* color hypomethylated CpGs (male versus female). The *red line* represents the significance level of 1.26E−07

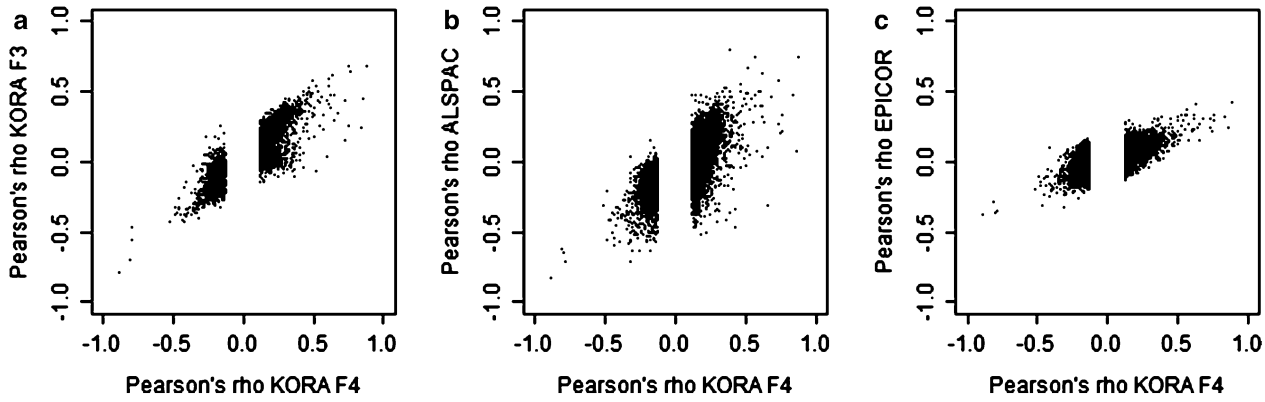

Fig. 2 Correlations between SMAs identified in KORA F4 and the respective associations in each replication study. Each *point* corresponds to one CpG site. SMAs in the discovery sample KORA F4 are plotted against SMAs in the three replication studies, with **a** KORA F4 against KORA F3, **b** KORA F4 against ALSPAC, **c** KORA F4 against EPICOR. Note that only the CpGs that were significant in KORA F4 and subsequently taken to replication are plotted, which results in the gap in the middle of the graphs

Table 2 Distribution and enrichment of sex-methylation associations (SMAs) among different CpG regions

CpG region[a]	Island	Shores[b]	Shelf	Open sea	Total
No. of CpG probes	26,159	18,888	7203	26,484	78,734
Significant SMAs	51	97	19	64	231

[a] Location with respect to the CGI: shore-0 to 2 kb from CGI, shelf-2 to 4 kb from CGI; north-upstream, south-downstream [19]

[b] North and south shores combined

have one silenced copy through DNAm in a parent-of-origin-specific manner; some examples of relations between autosomal DNAm patterns and the sex of the carrier of the imprinted gene are already known [20, 21]. We observed that SMAs tend to be of higher significance among CpGs at known imprinted genes compared to non-imprinted genes (Spearman correlation 0.025, $p = 4.0\text{E}{-04}$ considering discovery p values; 0.019, $p = 8.0\text{E}{-03}$ considering replication p values). Particularly, there were 3816 genes that passed the threshold for genome-wide significance in KORA F4 (where the best CpG near the gene had genome-wide significant p value), while out of those, 16 were imprinted genes (p value 0.008, hypergeometric test). Of these 16 genes, 10 also had significant SMAs in the meta-analysis; these were *GNAS*, *INPP5F*, *KCNQ1DN*, *GRB10*, *KCNQ1*, *DLGAP2*, *DLK1*, *PPP1R9A*, *MEG3*, and *PLAGL1*. This result suggests that imprinted genes tend to harbor sex-specific CpGs more often than non-imprinted genes.

Enrichment for GO terms among the SMAs

With the aid of gene ontology (GO) terms, we analyzed the genes annotated to the sex-associated CpGs to identify common biological roles or processes in

which the products of these genes might be involved. GO term-enrichment test revealed three significantly enriched GO terms among the genes annotated to the 1184 meta-analysis hits compared with the 19,170 genes annotated to the CpGs that went into our analysis in KORA F4. These GO terms were 'negative regulation of glycolysis,' 'negative regulation of systemic arterial blood pressure,' and 'response to protozoan' (Table 3).

Enrichment of SMAs in sex hormone-related genes

Among the genes annotated to CpG sites with significant SMAs, we sought to explore the enrichment of genes involved in sex hormone biosynthesis, transport, receptors, genes of other hormones with sexual dimorphisms, as well as genes known to be involved in disorders of sexual development (identified by OMIM search [22]), excluding those located on X and Y chromosomes. A list of the genes considered for this analysis is given in Additional file 4. No significant enrichment among the SMAs was found for these genes (hypergeometric test, p value 0.41).

Association between SMAs and expression levels

To explore the biological consequences of the 1184 significant SMAs, we tested whether there were differences in mRNA expression levels of 16,904 genes associated with both DNAm at the validated CpGs and sex in KORA F4. We considered genes 1 Mb around each SMA and identified 2 CpG-expression associations in the discovery sample from KORA F4 ($p < 8.55E-08$). The effects of methylation and sex on the mRNA expression level were negative, i.e., expression appeared to be higher both with lower methylation values and female sex. These two associations encompassed two genes: the *cytokine-inducible SH2-containing protein* gene (*CISH*); and the *RAB23, member RAS oncogene family* gene (*RAB23*) (Table 4).

Discussion

We characterized sex differences in autosomal DNAm in a large population-based European sample and identified 11,010 sex-methylation associations (SMAs) after filtering for the highest associated CpG site per gene. Of these, 1184 were validated in a meta-analysis of additional three independent replication cohorts. The identified CpGs

Table 3 Enriched GO terms among the sex-associated CpGs from the meta-analysis

GO ID	GO term	*p* value[a]	Gene[b]	No. of CpG probes[c]
GO:0045820	Negative regulation of glycolysis	3.51E−07	HDAC4	393
			IER3	39
			FBP1	14
GO:0003085	Negative regulation of systemic arterial blood pressure	1.81E−06	ADRB1	8
			CALCA	40
			SOD2	7
			IER3	39
			BBS4	4
GO:0001562	Response to protozoan	2.69E−06	SLC11A1	20
			IL12B	12
			IER3	39

[a] *p* value for enrichment, Bonferroni-corrected significance level of 3.08E−08

[b] Genes annotated to the GO term restricted to those annotated to significant SMAs in KORA F4

[c] Number of CpG probes associated with each gene, according to Illumina's annotation file; enrichment tests were corrected for the probe numbers (see "Methods")

Table 4 Significant associations between SMAs and mRNA expression levels in KORA F4

Probe ID expression	Probe ID methylation	est.[b]	*p* value[c]	Methylation		Sex		Dist. CpG-gene (bp)[d]	Matched gene[a]
				est.[b]	*p* value	est.[c]	*p* value		
ILMN_1738207	cg01048561	0.16	9.85E−08	−0.03	0.03	−0.82	7.25E−08	99,832	CISH
ILMN_2346997	cg16616022	0.14	3.02E−07	−0.06	0.03	−1.70	8.16E−08	100,699	RAB23

[a] Gene annotation according to expression (not to CpG sites!)

[b] Beta estimate for the model expression $= b_0 + b_1 \times \text{sex} + b_2 \times \text{methylation} + e$

[c] Bonferroni-corrected significance level of 8.55E−08

[d] Genomic distance between CpG and the matched gene

were enriched in CGI shores, in imprinted genes, with three GO terms, but not in sex hormone-related genes.

Sex-specific methylation patterns

We identified marked differences in locus-specific DNA methylation between men and women throughout the autosomes with the largest proportion of effect sizes ranging between 0.5 and 1.5, but also a considerable proportion with substantial effect sizes (−0.89 to 0.73 in the meta-analysis). This is consistent with previous findings of sex-specific DNA methylation patterns. For instance, Liu et al. explored DNAm at 20,493 CpG sites in a small alcohol-abuse study ($n = 197$), identifying one gene on chromosome 4 to be significantly differentially methylated between men and women [9]. Furthermore, a recent study with main focus on the heritability and repeatability of DNAm in whole blood found 1687 CpGs associated with sex [23]. There is a significant overlap between that study and our results: 213 of the 1687 CpGs previously reported as being associated with sex are among the hits identified in the meta-analysis reported here (p value <1E−17, hypergeometric test). Inoshita et al. reported 292 discovery hits in peripheral blood of which we find 41 among our meta-analysis hits (25 %) and 228 among our discovery hits (2 %) [24]. As the publication does not provide a list of 98 replicated CpGs, a comparison between replicated results of their and our hits was not possible. To extend the comparison beyond blood, we compared our results with two studies with brain samples that were also analyzed with the 450-k chip. The intersection with discovery hits in the prefrontal cortex after correction for multiple testing from Xu et al. were only 2 and 3 % compared with our discovery and meta-analysis, respectively [25]. Of note, the study by Xu et al. included only 46 subjects. This fact might explain the little overlap. Spiers and co-workers published results on DNAm in fetal brain of which 14 and 18 % of their hits overlap with our discovery and meta-analysis hits, respectively [26].

Moreover, we looked at comparisons to DNAm changes associated with diabetes and with smoking as an example of environmental influence on DNAm. The incident diabetes-associated CpGs were all not significant in our data [27]. Among the 32 CpGs associated with smoking status, the majority was not significant in our results, i.e., 22 CpGs, whereas nine were significant in our KORA F4 results, two in KORA F3 and F4, one of her CpGs was excluded from our analysis [28]. Thus, none of the hits of the smoking study survived significance level in our replication or meta-analysis.

Since the four populations examined in our study were rather heterogeneous, particularly in terms of age and genetic background, the large number of replicated SMAs indicates that the majority of DNAm differences between the sexes are stable over time and independent of geographic origin.

SMAs are found to be enriched in CpG island shores and imprinted genes

Our results on the local agglomeration of CpGs revealed an enrichment of SMAs in CGI shores. It is thought that methylation of CpGs results in different functional consequences depending on their genetic location. For example, CpGs in the gene body are thought to be related to regulation of the gene's expression, whereas CpGs in a gene desert are thought to contribute to genomic stability [29]. With regard to CGI shores, methylation of these features seems to regulate gene expression, possibly by silencing effects [30]. However, according to Liu et al., a CpG location in genetic region is only a weak approximation to its functional consequence [31], which may explain the low number of gene expression differences linked to sex-specific DNAm in our study.

Imprinting is a phenomenon of monoallelic gene silencing through DNAm in a parent-of-origin-specific manner [32]. DNAm differences in imprinted genes have been widely reported [33–37]; some studies showed sex-specific effects of these DNAm differences [20, 21]. Looking systematically at differences of DNAm between males and females at imprinted genes, we observed an enrichment of these 10 imprinted genes among significant SMAs in the meta-analysis: *GNAS*, *INPP5F*, *KCNQ1DN*, *GRB10*, *KCNQ1*, *DLGAP2*, *DLK1*, *PPP1R9A*, *MEG3*, and *PLAGL1*. Full names of these genes and those additionally found in the discovery study are listed in Additional file 5. Although examples of sex-specific DNAm differences in imprinted genes have been reported, these differences seem to be unknown for the imprinted genes reported here.

Three GO terms were enriched for SMAs

To achieve a more functional interpretation of the SMAs, enrichment was observed for three GO terms among the genes annotated to the CpGs in our meta-analysis: 'negative regulation of glycolysis,' 'negative regulation of systemic arterial blood pressure,' and 'response to protozoan.' Regarding the regulation of glycolysis, there are indications of differences between men and women in terms of glucose metabolism in the literature [38, 39]. Also, sex differences in the regulation of the arterial blood pressure are reported [40, 41]. These complementary strands of evidence suggest that sex-associated differences in DNAm may explain some of the underlying discordance in these, and possibly other traits and diseases.

The relationship of sex hormones and DNA methylation

An analysis for enrichment of genes related to sex hormone pathways (e.g., synthesis enzymes) revealed nothing of note in our data, suggesting that sex hormone-related autosomal genes show no sex-dependent DNAm differences. Sex hormones may, however, play an important role in determining DNAm levels; steroid hormones might mediate the transcriptional effects of co-regulatory complexes that associate with epigenetic modifiers [42]. Sex hormones might also exert regulatory effects on DNAm via downregulation of DNA methyltransferases expression in certain tissues [43]. Our observations of sex-dependent DNAm differences may reflect the evidence that steroidal hormones can influence DNAm, rather than vice versa.

Sex-associated methylation alters mRNA expression at some loci

Two genes, *CISH* and *RAB23*, showed significant DNAm-expression associations between men and women in the KORA F4 discovery sample. *CISH* and *RAB23* both exhibited lower levels of mRNA in women compared to men.

The function of RAB23 in humans is not completely unraveled yet, but it is known that this small GTPase acts as a negative regulator of the hedgehog pathway [44, 45]. Hedgehog signaling has a crucial role in response to injury, tissue stress, healing, and regeneration by helping to maintain and expand somatic stem cell populations [46]. Thus, hedgehog signaling imperatively needs to be tightly regulated. The relevance of sex-specific regulation of gene expression of this locus is unclear.

CISH is a member of the SOCS family and, together with SOCS1, 2, and 3, exhibits inhibitory functions as negative feedback to the JAK-STAT signaling pathway, triggered by cytokines like IL-2 and other stimuli [47–49]. This regulation is related to regulatory T-cell function and involved in immune polarization. Aberrations in the JAK-STAT pathway may be involved in hematopoietic disorders, autoimmune and inflammatory diseases [48], and affect susceptibility to infectious diseases [47, 49, 50].

Indeed, many infectious diseases appear to be more common in males that in females [51], but autoimmune diseases affect more females than males [52]. Thus, we hypothesize that DNAm might be a possible contributor to immunologic differences between the sexes.

Regarding the low number of genes expression of which was associated with differential DNAm in our study, one should have in mind that the consequences of changes in DNAm cover a much broader spectrum including effects in trans and genomic stability, for instance.

Strength and limitations

Important strengths of this study are the relatively large sample size and replication in several independent studies. In addition, validation was conducted in another tissue. Furthermore, we implemented a rather stringent and conservative threshold for statistical significance so that we likely underestimated the number of SMAs reported. Both the comprehensiveness in design along with several in-depth downstream analyses strengthens our study.

Some limitations arise from the methodology of the Illumina Infinium HumanMethlation450 BeadChip which covers only 1.7 % of known CpG sites across the genome, rather than more comprehensive sequence-based approaches. However, the 450-k chip is currently the most suitable method for high-throughput measurements for large epidemiologic studies in terms of sample throughput and expenses, and the chip serves a purpose in identifying differences in DNAm at the group level, despite its widely recognized technical limitations and idiosyncrasies.

Blood is an easily accessible source of DNA in epidemiologic studies, but tissue-specificity is an important issue when analyzing DNAm. Blood may function as surrogate tissue for systemic constitutional DNAm differences, but may not necessarily mirror specific changes in other tissues [29]. We assessed tissue specificity by comparing our results to a study in buccal cells. Despite the low rate of associations observed in the buccal data, out of the 96 sex-associated CpGs in the buccal data, 16 CpGs were also associated with sex in KORA F4. However, we must underline that potential sex-related differences in white blood cell proportions might not be captured by the applied estimation method and may thus confound our analysis to some extent.

Conclusion

We identified 1184 CpGs showing stable DNA methylation differences between men and women in four European cohorts. These sites were found to be enriched at CGI shores and at imprinted genes. Furthermore, we observed enrichment for three GO terms. From these results, we conclude that sex-dependent DNAm may be implicated in the observed sex discordance in various traits and diseases. Functional associations were demonstrated through mRNA expression analysis, which revealed two genes with significant sex- and DNAm-dependent expression differences, namely, *CISH* and *RAB23*.

Based on the substantial DNAm differences we found between men and women, we advocate that greater attention should be paid to sex differences in epigenetic

studies. Further work should be conducted to establish whether the extensive catalog of sex-associated DNAm differences observed are mediating the known sex discordance seen in many traits and diseases.

Methods

Study populations

Discovery

For the identification of SMAs, we used KORA F4 as the discovery sample and meta-analyzed the results from KORA F3, ALSPAC, and EPICOR.

KORA (*Cooperative Health Research in the Region of Augsburg*) is a population-based research platform with subsequent follow-up studies in the fields of epidemiology, health economics, and health care research [53]. Four surveys were conducted with participants living in the city of Augsburg and 16 surrounding towns and villages. The discovery dataset comprised 1799 individuals from the KORA F4 survey (all with genotyping data available) conducted during 2006–2008. Fasting blood samples were drawn into serum gel tubes in the morning. Further details are published elsewhere [54–56]. See "Array-based DNA methylation analysis" for the laboratory analysis details.

Replication

As one replication sample, we used 500 subjects of the *KORA F3* (50 % smokers, age comparable to KORA F4 sample, 50 % females), originally selected for the smoking study by Zeilinger et al. [28, 56, 57].

KORA F4 and F3 are found to be completely independent with no overlap of study subjects. Also, no indication of population stratification was seen in numerous publications on the KORA studies [58]. For all studies, we obtained written consent from participants and approval from the local ethical committees.

The *Avon Longitudinal Study of Parents and Children (ALSPAC)* is a longitudinal birth cohort in the Bristol area of the UK (details in [59, 60]). A subgroup of 963 ALSPAC children samples collected in adolescence was used for replication in this study. DNA methylation data were generated on this subgroup as part of the Accessible Resource for Integrated Epigenomic Studies (ARIES) [61]. Use of the data provided has previously been approved by the ALSPAC Ethics and Law committee. Please note that the study website contains details of all the data that are available through a fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/).

A third replication sample was from a nested case–control study from the Italian *EPICOR2* study [part of the European Prospective Investigation into Cancer and Nutrition (EPIC) Study]. The study sample includes 292 myocardial infarction (MI) cases and 292 matched controls that were healthy at recruitment, but diagnosed for nonfatal MI during the follow-up. Blood samples were collected at the time of enrollment. Further details are published in Fiorito et al. [62].

Array-based DNA methylation analysis

Genome-wide DNAm patterns of the 1814 KORA F4 samples (1 μg DNA from whole blood) were assessed using the Infinium HumanMethylation450 BeadChip Array (Illumina) as described elsewhere [28]. Also in the replication studies, DNAm was measured with the 450-k chip.

The percentage of methylation of a given cytosine is reported as a beta-value, which is a continuous variable between 0 and 1, corresponding to the ratio of the methylated signal over the sum of the methylated and unmethylated signals.

Technical validation of the method has been reported elsewhere [63].

Data preprocessing and quality assurance

Raw methylation data were obtained from GenomeStudio software (Illumina, version 2011.1) methylation module (version 1.9.0) and preprocessed as proposed by Touleimat and Tost [64] with default option and an in-house updated list of CpGs with SNPs in the probe-binding regions, followed by beta-mixture quantile normalization (BMIQ) as a normalization step to correct for the InfI/InfII distribution shift of the beta-values ([65], using R-package wateRmelon, version 1.0.3 [66]). After the quality control, 391,885 CpGs for 1799 samples were eligible to enter the analysis. Details on data normalization can be found in Wahl et al. [67].

For the ALSPAC study, ARIES DNA methylation data version 1, released in October 2013, was used. During quality control, samples were excluded if they had detection p value ≥ 0.01 at more than 20 % probes, or unexpected mismatches with SNP genotype data or sex information. Finally, for normalization, the Touleimat and Tost method implemented in R-package *wateRmelon*, version 1.0.3 was applied with default parameters.

Similar procedures as in KORA were done in the EPICOR study, resulting in another 788 CpGs being excluded from the given KORA F4 hits.

Statistical analysis

We defined the sex-methylation association SMA at each CpG site to be the Pearson's correlation between sex and the methylation beta-value. Prior to association testing, methylation measurements were adjusted across the study sample by taking the residuals of the linear regression with the following covariates: Age, BMI, smoking

behavior, alcohol intake, triglycerides, total leucocytes, plate, HDL, LDL, physical activity, diabetes, myocardial infarction, and estimated white blood cell proportions. Specifically, the proportions of neutrophils, $CD4^+$ T, $CD8^+$ T, B cells, monocytes, and granulocytes were estimated according to Houseman et al. [68]. Experimental 96-well plates were represented by dummy variables. The Manhattan plot was drawn with an adapted version of the drawing function in the qqman package for R [69].

Allosomal cross-hybridization

Previous studies reported a subset of the 450k probes to be cross-reactive in silico [70, 71]. Cross-hybridization with the sex chromosomes leads to false discoveries as the measured methylation value is composed of a mixture of values from the target location and from the sex-chromosome, which obviously differs between sexes. Price et al. showed that probes with high homology to the sex chromosomes have a significant enrichment of sex-methylation associations [71].

We repeated the computational analysis of Price et al. in order to validate that the 450k probes have no correlation between allosomal homology and sex associations in our dataset. For each autosomal probe, we defined the cross-hybridization value to be equal to the number of matches between each probe and its best alignment to the allosomes. The cross-hybridization value equals the number of matching bases between each probe sequence and its best alignment to chromosomes X and Y. We observed that the highly significant SMAs tend to have high cross-hybridization values (see Additional file 6). We therefore removed all CpGs from the analysis with cross-hybridization values larger than a threshold of 40 (the point at which enrichment of SMAs starts). After the removal of these CpGs, the correlation between the SMA values and the cross-hybridization values was not statistically significant (p values were calculated empirically using 3000 permutations). This process resulted in the removal of 12,260 probes, of which 568 had significant sex associations (4.63 %). To be conservative, we additionally removed all nonspecific probes suggested by price; 391,885 CpGs were finally considered as safe probes.

Replication and meta-analysis

For replication of the significant CpGs from the discovery step, three studies were included: KORA F3, ALSPAC, and EPICOR. Association analysis was performed as described for the discovery sample. Covariates were included as follows: In KORA F3, all of the discovery covariates were included. In ALSPAC, age, BMI, total cholesterol, HDL-C, LDL-C, tissue contents estimated by the Houseman algorithm [68], and the bisulfite conversion batch (BCD plate) were included. In EPICOR, the discovery covariates were complemented with the center of recruitment (categorical: Torino, Varese, Napoli, Ragusa). Results from the three replication studies were meta-analyzed using the R function *metacor* (with default settings, i.e., Fisher's z-transformation of Pearson's rho values, random effects model). Significance level in the meta-analysis was Bonferroni-corrected for 10,222 tests corresponding to the number of CpGs analyzed in all replication cohorts (i.e., 4.89E−06).

Validation in monocellular tissue (buccal epithelium cells)

From the GEO database, we chose the largest publicly available dataset from a population-based study with even sex distribution and monocellular tissue as study sample [18]. In that study, the Illumina Infinium HumanMethylation27 BeadChip v1.2 was applied on buccal epithelium cells, and the data is publicly accessible at NCBI GEO database [72], Accession GSE25892.

The following preprocessing steps were already completed: Extraction of raw intensities using GenomeStudio v2010.1, background normalization by subtracting averaged negative probe intensity from signal A and B, quantile normalization, obtaining beta-values, and checking CpG site-wise call rate (CpGs removed if >50 % of data points had bad quality). According to our data preprocessing in KORA, we additionally set beta-values to missing where respective detection p values were 0.01 or below, excluded samples with call rates of ≤80 % (there were none). BMIQ was not necessary since it is designed to correct the InfI–InfII distribution shift, which is not an issue when 27k is used, where all CpGs are of InfI design.

We obtained a subset of 22,773 CpG sites, which were measured in both experiments, and passed the above quality-control steps. We tested for SMAs in the buccal dataset with a linear model as no other covariates were available. We used a conservative Bonferroni correction for 27k hypotheses throughout this analysis.

Test for enrichment of SMAs in CpG locations

CGI location information was taken from UCSC genome browser CGI specification which was available from Illumina's probe description file. Since test for enrichment assumes independency (both the hypergeometric test and the Spearman correlation), and since CpG sites are locally correlated, we preprocessed the data by analyzing a random subset of the sites which was obtained by picking each site with probability 0.2. Applying this filter resulted in 78,734 CpG sites, 231 of these had significant SMA after replication.

Enrichment of SMAs in imprinted genes

We tested a known set of imprinted genes for enrichment of significant SMA CpGs. For that purpose, we downloaded a list of 59 genes from 'Geneimprint' by Dec 2013 [73] and selected only those with status 'imprinted' in Homo sapiens [74, 75]. Of these genes, 48 had CpG sites that passed quality control. For each gene as to Illumina's Infinium 450k annotation, including these 48, we picked the most significant CpG sites in our analysis, resulting in 19,170 sites. We considered the vector of p values for these sites, and then considered another binary vector of length 19,170, where an entry is 1 if the corresponding gene is an imprinted gene. Spearman correlations were calculated for the pairs of vectors.

GO-enrichment analysis

An enrichment analysis for gene ontology (GO) terms was performed with all CpGs analyzed in KORA F4, resulting in 19,170 annotated genes (according to Illumina's annotation file). GO is an ontology of defined terms representing gene product properties [76]. According to Geeleher et al. [77], we applied the R-package *GOseq* [78], originally developed for expression analysis, to determine different probabilities of detecting a gene due to different CpG probe number per gene. Therefore, we defined those CpGs as differentially methylated which were replicated in the meta-analysis, looked at the annotated genes, and used the number of probes per gene as variable for weighting the results. Results were Bonferroni-corrected for 16,233 tests according to the number of genes that were processed by the *GOseq* function.

Enrichment in sex hormone-related genes

We created a list of genes involved in sex hormone biosynthesis, transport, receptors, genes of other hormones with sexual dimorphisms, as well as genes known to be involved in disorders of sexual disorders (identified by an OMIM search [22]), excluding those located on X and Y chromosomes (Additional file 4). We then compared those to our meta-analysis hits with a hypergeometric test.

mRNA expression analysis

Gene expression data were available for 731 KORA F4 subjects and were analyzed using the Illumina Human HT-12 v3 Expression BeadChip. The procedures are described elsewhere [79]. For all CpGs surpassing the threshold for statistical significance in the meta-analysis, mRNA expression levels were checked as expression of genes in 1 Mb distance to each CpG site as linear combination of sex effect and DNAm effect with the same cofactors as in the main analysis. The significance level of 8.55E−08 was calculated as 0.05 divided by the number of combinations between CpGs and genes, which was 585,031.

Software

Analyses were performed using Matlab and the Statistics Toolbox, Release 2013b, The MathWorks, Inc., Natick, Massachusetts, United States. Preprocessing and quality control steps as well as replications, meta-analysis, expression analysis, and GO and sex hormone-related gene-enrichment analyses were done using R, versions 2.15.3 3.0.1, and for the Manhattan plot version 3.1.3 [80].

Additional files

Additional file 1. Flow-chart of the studies and analyses. The figure provides an overview of the analyzed studies and the conducted analyses with corresponding numbers.

Additional file 2. Results from discovery (KORA F4) and replication analyses (KORA F3, ALSPAC, EPICOR). The table includes respective Pearson's rho and p-values for 391,885 CpGs in KORA, 11,010 CpGs in ALSPAC, and 10,222 CpGs in EPICOR as well as theta values and p-values for the meta-analysis. The file additionally contains a figure showing the frequency distribution of the effect sizes from the meta-analysis.

Additional file 3. Sex-associated CpGs in KORA F4 and the buccal cells study. The table includes CpGs significantly associated with sex in the KORA F4 and in the buccal cells study with chromosome number and position, CGI region, and annotated gene for each CpG, beta estimates and p-values per study.

Additional file 4. Sex-related genes considered for enrichment analysis. Given are the official gene symbol and the corresponding chromosomes. X and Y are marked in red since they were excluded from the enrichment analysis.

Additional file 5. Imprinted genes with significant SMAs enrichment.

Additional file 6. Enrichment of SMAs for probes with high Cross-Hybridization value (CHV). CHV is the number of matching bases to XY chromosomes. CHV values greater than 46 tend to be enriched for SMAs (green line). To be conservative, we removed all SMAs, represented as correlation coefficients on the y-axis, with CHV > 40 (red line).

Abbreviations

CGI: CpG island; DNAm: DNA methylation; GO: gene ontology; EWAS: epigenome-wide association study; SMA: sex methylation association.

Authors' contributions

PS designed this study, did statistical analyses and data interpretation in KORA, and wrote the manuscript. DST designed this study, did statistical analyses and data interpretation in KORA, and wrote the manuscript. SW designed this study, did the quality control for the 450k experiments in KORA, and wrote the manuscript. HG designed this study and supervised the research in general. GF did the quality control, statistical analyses, and data interpretation for EPICOR. SYS did statistical analyses and data interpretation for the ALSPAC study. KS performed the statistical analyses for the expression experiments. PW performed the statistical analyses for the expression experiments. SK performed 450-k experiments and did the quality control in KORA. YB did statistical analyses and data interpretation in KORA. SG performed experiments and quality control and interpreted the data for EPICOR. PV supervised research in EPICOR. VK supervised research in EPICOR. SP supervised research in EPICOR. RT supervised research in EPICOR. AK supervised the research in general. CG supervised the research in general. AP supervised the research in general. HP supervised research for the expression analyses in KORA. CLR supervised research in ALSPAC. GM interpreted the data for EPICOR and supervised research in EPICOR. TI designed this study and supervised the research in general. MW designed this study and supervised the research in general. EH supervised the research in general, did statistical analyses and data

interpretation in KORA, and wrote the manuscript. All authors contributed to the final manuscript. All authors read and approved the final manuscript.

Author details
[1] Research Unit of Molecular Epidemiology, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Neuherberg, Germany. [2] Institute of Epidemiologie II, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Neuherberg, Germany. [3] Blavatnik School of Computer Science, Tel-Aviv University, Tel-Aviv, Israel. [4] German Center for Diabetes Research (DZD e.V.), Neuherberg, Germany. [5] Human Genetics Foundation-Torino, Turin, Italy. [6] Department of Medical Sciences, University of Torino, Turin, Italy. [7] MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK. [8] Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK. [9] Institute of Human Genetics, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Neuherberg, Germany. [10] Institute of Human Genetics, Technical University Munich, Munich, Germany. [11] Epidemiology and Public Health, Imperial College London, London, UK. [12] Epidemiology and Prevention Unit, Fondazione IRCSS Istituto Nazionale Tumori, Milan, Italy. [13] Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy. [14] Cancer Registry and Histopathology Unit, "Civile-M.P. Arezzo" Hospital, ASP 7, Ragusa, Italy. [15] Hannover Unified Biobank, Hannover Medical School, Hannover, Germany. [16] Institute for Human Genetics, Hannover Medical School, Hannover, Germany. [17] Department of Molecular Microbiology and Biotechnology, George Wise Faculty of Life Science, Tel-Aviv University, Tel-Aviv, Israel. [18] International Computer Science Institute, Berkeley, CA, USA.

Acknowledgements
We are grateful to Prof. Dr. Jerzy Adamski (head of the Genome Analysis Center at the Institute of Experimental Genetics, Helmholtz Centre Munich and Editor-in-Chief J. Steroid Biochemistry and Molecular Biology) for helping with the list of genes of interest for the sex hormone-related gene enrichment. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provided core support for ALSPAC.

Funding
KORA KORA was initiated and financed by the Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany and supported by grants from the German Federal Ministry of Education and Research (BMBF), the Federal Ministry of Health (Berlin, Germany), the Ministry of Innovation, Science, Research and Technology of the state North Rhine-Westphalia (Düsseldorf, Germany), and the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ. We thank all members of field staffs who were involved in the planning and conduct of the MONICA/KORA Augsburg studies. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreements n° 261433 (BioSHaRe) to MW and AK and n° 603288 (sysVASC) to MW. CG was supported by the European Union's Seventh Framework Programme (FP7-Health-F5-2012) under grant agreement no. 305280 (MIMOmics) and by the Helmholtz-Russia Joint Research Group (HRJRG) 310. EH is a faculty fellow of the Edmond J. Safra Program at Tel Aviv University. EH and DST were supported in part by the Israel Science Foundation (Grant 1425/13), YB and EH by the United States-Israel Binational Science Foundation (Grant 2012304), DST and EH by the German-Israeli Foundation for Scientific Research and Development together with TI (Grant 1094-33.2/2010), YB and EH by the National Science Foundation (Grant III- 1217615), and DST and YB by fellowships from the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University.
EPICOR This work was supported by the Compagnia di San Paolo for the EPIC, EPICOR, and EPICOR2 projects (GM) and by the Human Genetics Foundation (HuGeF; GM, and PV). EPIC Italy is supported by grants from the Associazione Italiana per la Ricerca sul Cancro (AIRC, Milan) and by the European Union (EPIC-CVD 7th FP, Grant ref.: 279233).
ALSPAC This work was supported by the UK Medical Research Council Integrative Epidemiology Unit and the University of Bristol (MC_UU_12013_2). The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provided core support for ALSPAC. The Accessible Resource for Integrated Epigenomics Studies (ARIES) was funded by the UK Biotechnology and Biological Sciences Research Council (BB/I025751/1 and BB/I025263/1). SYS and CLR are supported by the MRC Integrative Epidemiology Unit. SYS is supported by a Post-Doctoral Research Fellowship from the Oak Foundation.

Competing interests
The authors declare that they have no competing interests.

Received: 25 June 2015 Accepted: 29 September 2015
Published online: 19 October 2015

References
1. Kim AM, Tingen CM, Woodruff TK. Sex bias in trials and treatment must end. Nature. 2010;465(7299):688–9. doi:10.1038/465688a.
2. Tingen CM, Kim AM, Wu P-H, Woodruff TK. Sex and sensitivity: the continued need for sex-based biomedical research and implementation. Womens Health (Lond Engl). 2010;6(4):511–6. doi:10.2217/whe.10.45.
3. Mittelstrass K, Ried JS, Yu Z, Krumsiek J, Gieger C, Prehn C, et al. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. PLoS Genet. 2011;7(8):1002215. doi:10.1371/journal.pgen.1002215.
4. Boraska V, Jeroncic A, Colonna V, Southam L, Nyholt DR, Rayner NW, et al. Genome-wide meta-analysis of common variant differences between men and women. Hum Mol Genet. 2012;21(21):4805–15. doi:10.1093/hmg/dds304.
5. Zhu Z-Z, Hou L, Bollati V, Tarantini L, Marinelli B, Cantone L, et al. Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis. Int J Epidemiol. 2012;41(1):126–39. doi:10.1093/ije/dyq154.
6. Fuke C, Shimabukuro M, Petronis A, Sugimoto J, Oda T, Miura K, et al. Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. Ann Hum Genet. 2004;68(Pt 3):196–204. doi:10.1046/j.1529-8817.2004.00081.x.
7. Iwasaki M, Ono H, Kuchiba A, Kasuga Y, Yokoyama S, Onuma H, et al. Association of postmenopausal endogenous sex hormones with global methylation level of leukocyte DNA among Japanese women. BMC Cancer. 2012;12:323. doi:10.1186/1471-2407-12-323.
8. Liu J, Morgan M, Hutchison K, Calhoun VD. A study of the influence of sex on genome wide methylation. PLoS One. 2010;5(4):10028. doi:10.1371/journal.pone.0010028.
9. El-Maarri O, Becker T, Junen J, Manzoor SS, Diaz-Lacava A, Schwaab R, et al. Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males. Hum Genet. 2007;122(5):505–14. doi:10.1007/s00439-007-0430-3.
10. El-Maarri O, Walier M, Behne F, van Üüm J, Singer H, Diaz-Lacava A, et al. Methylation at global LINE-1 repeats in human blood are affected by gender but not by age or natural hormone cycles. PLoS One. 2011;6(1):16252. doi:10.1371/journal.pone.0016252.
11. Tapp HS, Commane DM, Bradburn DM, Arasaradnam R, Mathers JC, Johnson IT, et al. Nutritional factors and gender influence age-related DNA methylation in the human rectal mucosa. Aging Cell. 2013;12(1):148–55. doi:10.1111/acel.12030.
12. Zhang FF, Cardarelli R, Carroll J, Fulda KG, Kaur M, Gonzalez K, et al. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. Epigenetics. 2011;6(5):623–9.
13. Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Sommer IE, et al. The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. PLoS One. 2009;4(8):6767. doi:10.1371/journal.pone.0006767.
14. Sarter B, Long TI, Tsong WH, Koh W-P, Yu MC, Laird PW. Sex differential in methylation patterns of selected genes in Singapore Chinese. Hum Genet. 2005;117(4):402–3. doi:10.1007/s00439-005-1317-9.
15. McCarthy NS, Melton PE, Cadby G, Yazar S, Franchina M, Moses EK, et al. Meta-analysis of human methylation data for evidence

of sex-specific autosomal patterns. BMC Genom. 2014;15:981. doi:10.1186/1471-2164-15-981.
16. Bernardino J, Lombard M, Niveleau A, Dutrillaux B. Common methylation characteristics of sex chromosomes in somatic and germ cells from mouse, lemur and human. Chromosome Res. 2000;8(6):513–25.
17. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet. 2006;38(12):1378–85. doi:10.1038/ng1909.
18. Essex MJ, Boyce WT, Hertzman C, Lam LL, Armstrong JM, Neumann SMA, et al. Epigenetic vestiges of early developmental adversity: childhood stress exposure and DNA methylation in adolescence. Child Dev. 2013;84(1):58–75. doi:10.1111/j.1467-8624.2011.01641.x.
19. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98(4):288–95. doi:10.1016/j.ygeno.2011.07.007.
20. Barker DJP, Thornburg KL. Placental programming of chronic diseases, cancer and lifespan: a review. Placenta. 2013;34(10):841–5. doi:10.1016/j.placenta.2013.07.063.
21. Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, et al. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Hum Mol Genet. 2009;18(21):4046–53. doi:10.1093/hmg/ddp353.
22. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Maryland): Online mendelian inheritance in man, OMIM®. (2014). http://omim.org/. Accessed via NCBI in July 2014.
23. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. Genome Res. 2014;24(11):1725–33. doi:10.1101/gr.176933.114.
24. Inoshita M, Numata S, Tajima A, Kinoshita M, Umehara H, Yamamori H, et al. Sex differences of leukocytes DNA methylation adjusted for estimated cellular proportions. Biol Sex Differ. 2015;6:11. doi:10.1186/s13293-015-0029-7.
25. Xu H, Wang F, Liu Y, Yu Y, Gelernter J, Zhang H. Sex-biased methylome and transcriptome in human prefrontal cortex. Hum Mol Genet. 2014;23(5):1260–70. doi:10.1093/hmg/ddt516.
26. Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CCY, O'Donovan MC, et al. Methylomic trajectories across human fetal brain development. Genome Res. 2015;25(3):338–52. doi:10.1101/gr.180273.114.
27. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. Lancet Diabetes Endocrinol. 2015;3(7):526–34. doi:10.1016/S2213-8587(15)00127-8.
28. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One. 2013;8(5):63812. doi:10.1371/journal.pone.0063812.
29. Langevin SM, Kelsey KT. The fate is not always written in the genes: epigenomics in epidemiologic studies. Environ Mol Mutagen. 2013;54(7):533–41. doi:10.1002/em.21762.
30. Vanderkraats ND, Hiken JF, Decker KF, Edwards JR. Discovering high-resolution patterns of differential DNA methylation that correlate with gene expression changes. Nucleic Acids Res. 2013;41(14):6816–27. doi:10.1093/nar/gkt482.
31. Liu Y, Ding J, Reynolds LM, Lohman K, Register TC, De La Fuente A, et al. Methylomics of gene expression in human monocytes. Hum Mol Genet. 2013;22(24):5065–74. doi:10.1093/hmg/ddt356.
32. Plasschaert RN, Bartolomei MS. Genomic imprinting in development, growth, behavior and stem cells. Development. 2014;141(9):1805–13. doi:10.1242/dev.101428.
33. Talens RP, Jukema JW, Trompet S, Kremer D, Westendorp RGJ, Lumey LH, et al. Hypermethylation at loci sensitive to the prenatal environment is associated with increased incidence of myocardial infarction. Int J Epidemiol. 2012;41(1):106–15.
34. Murphy SK, Adigun A, Huang Z, Overcash F, Wang F, Jirtle RL, et al. Gender-specific methylation differences in relation to prenatal exposure to cigarette smoke. Gene. 2012;494(1):36–43. doi:10.1016/j.gene.2011.11.062.
35. Azzi S, Sas TCJ, Koudou Y, Le Bouc Y, Souberbielle J-C, Dargent-Molina P, et al. Degree of methylation of ZAC1 (PLAGL1) is associated with prenatal and post-natal growth in healthy infants of the EDEN mother child cohort. Epigenetics. 2014;9(3):338–45. doi:10.4161/epi.27387.
36. Vidal AC, Murphy SK, Murtha AP, Schildkraut JM, Soubry A, Huang Z, et al. Associations between antibiotic exposure during pregnancy, birth weight and aberrant methylation at imprinted genes among offspring. Int J Obes (Lond). 2013;37(7):907–13. doi:10.1038/ijo.2013.47.
37. El Hajj N, Pliushch G, Schneider E, Dittrich M, Müller T, Korenkov M, et al. Metabolic programming of MEST DNA methylation by intrauterine exposure to gestational diabetes mellitus. Diabetes. 2013;62(4):1320–8. doi:10.2337/db12-0289.
38. Fu MH, Maher AC, Hamadeh MJ, Ye C, Tarnopolsky MA. Exercise, sex, menstrual cycle phase, and 17beta-estradiol influence metabolism-related genes in human skeletal muscle. Physiol Genomics. 2009;40(1):34–47. doi:10.1152/physiolgenomics.00115.2009.
39. Boschmann M, Jordan J, Schmidt S, Adams F, Luft FC, Klaus S. Gender-specific response to interstitial angiotensin II in human white adipose tissue. Horm Metab Res. 2002;34(11–12):726–30. doi:10.1055/s-2002-38262.
40. Hart EC, Charkoudian N, Wallin BG, Curry TB, Eisenach J, Joyner MJ. Sex and ageing differences in resting arterial pressure regulation: the role of the beta-adrenergic receptors. J Physiol. 2011;589(Pt 21):5285–97. doi:10.1113/jphysiol.2011.212753.
41. Christou DD, Jones PP, Jordan J, Diedrich A, Robertson D, Seals DR. Women have lower tonic autonomic support of arterial blood pressure and less effective baroreflex buffering than men. Circulation. 2005;111(4):494–8. doi:10.1161/01.CIR.0000153864.24034.A6.
42. Kaminsky Z, Wang S-C, Petronis A. Complex disease, gender and epigenetics. Ann Med. 2006;38(8):530–44.
43. Zhang X, Ho S-M. Epigenetics meets endocrinology. J Mol Endocrinol. 2011;46(1):11–32.
44. Chi S, Xie G, Liu H, Chen K, Zhang X, Li C, et al. Rab23 negatively regulates gli1 transcriptional factor in a su(fu)-dependent manner. Cell Signal. 2012;24(6):1222–8. doi:10.1016/j.cellsig.2012.02.004.
45. Evans TM, Simpson F, Parton RG, Wicking C. Characterization of rab23, a negative regulator of sonic hedgehog signaling. Methods Enzymol. 2005;403:759–77. doi:10.1016/S0076-6879(05)03066-1.
46. Irvine DA, Copland M. Targeting hedgehog in hematologic malignancy. Blood. 2012;119(10):2196–204. doi:10.1182/blood-2011-10-383752.
47. Jacobsen M, Repsilber D, Kleinsteuber K, Gutschmidt A, Schommer-Leitner S, Black G, et al. Suppressor of cytokine signaling-3 is affected in t-cells from tuberculosis tb patients. Clin Microbiol Infect. 2011;17(9):1323–31. doi:10.1111/j.1469-0691.2010.03326.x.
48. Xu Z, Huang G, Gong W, Zhao Y, Zhou P, Zeng Y, et al. Activation of farnesoid X receptor increases the expression of cytokine inducible SH2-containing protein in HepG2 cells. J Interferon Cytokine Res. 2012;32(11):517–23. doi:10.1089/jir.2012.0008.
49. Khor CC, Vannberg FO, Chapman SJ, Guo H, Wong SH, Walley AJ, et al. CISH and susceptibility to infectious diseases. N Engl J Med. 2010;362(22):2092–101. doi:10.1056/NEJMoa0905606.
50. Periasamy S, Dhiman R, Barnes PF, Paidipally P, Tvinnereim A, Bandaru A, et al. Programmed death 1 and cytokine inducible SH2-containing protein dependent expansion of regulatory T cells upon stimulation with *Mycobacterium tuberculosis*. J Infect Dis. 2011;203(9):1256–63. doi:10.1093/infdis/jir011.
51. Guerra-Silveira F, Abad-Franch F. Sex bias in infectious disease epidemiology: patterns and processes. PLoS One. 2013;8(4):62390. doi:10.1371/journal.pone.0062390.
52. Pennell LM, Galligan CL, Fish EN. Sex affects immunity. J Autoimmun. 2012;38(2–3):282–91. doi:10.1016/j.jaut.2011.11.013.
53. Rathmann W, Strassburger K, Heier M, Holle R, Thorand B, Giani G, et al. Incidence of type 2 diabetes in the elderly German population and the effect of clinical and lifestyle risk factors: KORA S4/F4 cohort study. Diabet Med. 2009;26(12):1212–9. doi:10.1111/j.1464-5491.2009.02863.x.
54. Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, Prehn C, et al. A genome-wide perspective of genetic variation in human metabolism. Nat Genet. 2010;42(2):137–41. doi:10.1038/ng.507.

55. Holle R, Happich M, Löwel H, Wichmann HE. MONICA/KORA Study Group: KORA—a research platform for population based health research. Gesundheitswesen. 2005;67(Suppl 1):19–25.
56. Wichmann H-E, Gieger C, Illig T, MONICA/KORA Study Group. KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen. 2005;67(Suppl 1):26–30.
57. Löwel H, Döring A, Schneider A, Heier M, Thorand B, Meisinger C, et al. The MONICA augsburg surveys—basis for prospective cohort studies. Gesundheitswesen. 2005;67(Suppl 1):13–8.
58. Steffens M, Lamina C, Illig T, Bettecken T, Vogler R, Entz P, et al. SNP-based analysis of genetic substructure in the German population. Hum Hered. 2006;62(1):20–9. doi:10.1159/000095850.
59. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol. 2013;42(1):111–27. doi:10.1093/ije/dys064.
60. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol. 2013;42(1):97–110. doi:10.1093/ije/dys066.
61. Relton CL, Gaunt T, McArdle W, Ho K, Duggirala A, Shihab H, et al. Data resource profile: accessible resource for integrated epigenomic studies (aries). Int J Epidemiol. 2015;. doi:10.1093/ije/dyv072.
62. Fiorito G, Guarrera S, Valle C, Ricceri F, Russo A, Grioni S, et al. B-vitamins intake, DNA-methylation of one carbon metabolism and homocysteine pathway genes and myocardial infarction risk: the EPICOR study. Nutr Metab Cardiovasc Dis. 2014;24(5):483–8. doi:10.1016/j.numecd.2013.10.026.
63. Petersen A-K, Zeilinger S, Kastenmüller G, Römisch-Margl W, Brugger M, Peters A, et al. Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. Hum Mol Genet. 2014;23(2):534–45. doi:10.1093/hmg/ddt430.
64. Touleimat N, Tost J. Complete pipeline for Infinium® HumanMethylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics. 2012;4(3):325–41. doi:10.2217/epi.12.21.
65. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450K DNA methylation data. Bioinformatics. 2013;29(2):189–96. doi:10.1093/bioinformatics/bts680.
66. Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genom. 2013;14:293. doi:10.1186/1471-2164-14-293.
67. Wahl S, Fenske N, Zeilinger S, Suhre K, Gieger C, Waldenberger M, et al. On the potential of models for location and scale for genome-wide DNA methylation data. BMC Bioinform. 2014;15:232. doi:10.1186/1471-2105-15-232.
68. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinform. 2012;13:86. doi:10.1186/1471-2105-13-86.
69. Turner SD. qqman: an R package for visualizing GWAS results using Q–Q and manhattan plots. bioRxiv. 2014. doi:10.1101/005165.
70. Chen Y-A, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8(2):203–9. doi:10.4161/epi.23470.
71. Price EM, Cotton AM, Lam LL, Farré P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics Chromatin. 2013;6(1):4. doi:10.1186/1756-8935-6-4.
72. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013;41(Database issue):991–995. doi:10.1093/nar/gks1193.
73. Jirtle RL. Geneimprint. http://www.geneimprint.com/-site/-genes-by-status.
74. Daelemans C, Ritchie ME, Smits G, Abu-Amero S, Sudbery IM, Forrest MS, et al. High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. BMC Genet. 2010;11:25. doi:10.1186/1471-2156-11-25.
75. Morison IM, Reeve AE. A catalogue of imprinted genes and parent-of-origin effects in humans and animals. Hum Mol Genet. 1998;7(10):1599–609.
76. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet. 2000;25(1):25–9. doi:10.1038/75556.
77. Geeleher P, Hartnett L, Egan LJ, Golden A, Raja Ali RA, Seoighe C. Gene-set analysis is severely biased when applied to genome-wide methylation data. Bioinformatics. 2013;29(15):1851–7. doi:10.1093/bioinformatics/btt311.
78. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2):14. doi:10.1186/gb-2010-11-2-r14.
79. Schurmann C, Heim K, Schillert A, Blankenberg S, Carstensen M, Dörr M, et al. Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. PLoS One. 2012;7(12):50938. doi:10.1371/journal.pone.0050938.
80. R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2011. R Foundation for Statistical Computing. ISBN 3-900051-07-0. http://www.R-project.org/.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central