

For questions related to these analyses, contact Barbara Ruggeri (barbara.ruggeri@kcl.ac.uk), Christine Macare (christine.macare@kcl.ac.uk), Tianye Jia (tianye.jia@kcl.ac.uk) or Sylvane Desrivières (sylvane.desrivieres@kcl.ac.uk)

This document summarises work performed in the labs of Sven Cichon (Stefan Herms) at Basel University, Georgy Bakalkin (Daniil Sarkisyan) and Tomas Axelsson at the University of Uppsala who generated the data (section 1, below), and of Yun Liu, Department of Biochemistry and Molecular Biology at Fudan University, who QC'ed the raw data (section 2).

1. DNA extraction and microarray hybridization

DNA was extracted (by the Centre National de Génotypage in Paris —Marc Lathrop's lab) from whole blood samples (~10mL) collected at the age of 14 and preserved in BD Vacutainer EDTA tubes (Becton, Dickinson and Company) using the Gentra Puregene Blood Kit (QIAGEN) according to the manufacturer's instructions.

Global blood DNA methylation levels were assessed by hybridizing DNA samples to the Infinium Human Methylation 450 Bead Chip, which interrogates over 485,000 methylation sites located throughout the genome (Illumina: http://www.illumina.com/products/methylation_450_beadchip_kits.html), following the manufacturer's protocol.

DNA samples were processed and hybridised in two waves:

Wave 1 included 708 samples randomly selected from the IMAGEN age 14 subjects and was processed by Tomas Axelsson (SNP& SEQ Technology platform, Uppsala University).

The following subject IDs were processed in Wave 1:

106871	11834339	23079648	35524006	48663051	60835696	72914941
215284	11944593	23194232	35587656	48727429	60919135	73236450
240546	12288410	23207511	35620929	48731895	61066708	73347940
297685	12582808	23639409	35661078	48846112	61507009	73628015
556983	12629290	23700572	35704260	48934345	61561596	73661031
602062	12646956	23746158	35734504	48935365	61581707	73668053
724314	12764460	23977691	35885865	48958188	61589239	73932354
737577	12809392	24041790	35968441	49311998	61630843	74104786
942170	12837283	24115048	36007181	49321994	61829162	74452500

DNA methylation analyses in IMAGEN samples

1023924	12967847	24686194	36043675	49403856	61869818	74536549
1088410	12975877	24715402	36214047	49433852	61943676	74601732
1283761	13025821	24782190	36232609	49483266	62182476	74731388
1330914	13361778	24786016	36503289	49850700	62632949	74757619
1338177	13368755	25000478	36525962	49899818	63031549	74934634
1357508	13377477	25073654	36529694	50775428	63301798	75126114
1380042	13424686	25103079	37509984	51006883	63400084	75142074
1441076	13679970	25139569	37740824	51133284	63489515	75233772
1647364	14141334	25193415	37805580	51293841	63770152	75465686
1664091	14185243	25229476	37836103	51342034	63873628	75478935
1717528	14540698	25254435	38429243	51387606	63951655	75578857
1938036	14837805	25378702	38436183	51436842	63985887	75698628
1939282	14900981	25453247	38521433	51671690	64210895	75760021
2296749	15398752	25821028	38666520	51829237	64564726	75809342
2371789	15411565	25843987	38685366	51999613	64672179	75993279
2758817	15439849	26045785	38690448	52086724	64800062	76254920
2904393	15539449	26047145	38852160	52375919	64881320	76310486
3023509	15734458	26193033	38965837	52419978	64893803	76390037
3328367	15765805	26200823	38979988	52774607	64913138	76544039
3610829	16003262	26400185	39215990	52886278	64924364	76631575
3712195	16010156	26545049	39678719	53000777	64970679	76762825
3726191	16069129	26722960	39711869	53120454	65032538	77008362
3970752	16140261	27553119	39826529	53431058	65084605	77380027
3991592	16352457	27553757	39923280	53515504	65086134	77511665
4004697	16408573	27718055	40021391	53554600	65259482	77545059
4032825	16448772	27981638	40302134	53703770	65270890	77689671

DNA methylation analyses in IMAGEN samples

4068298	16466990	28164869	40590177	53744289	65334806	77729708
4118866	16590495	28250707	40623634	54198873	65503534	77788846
4285802	16634839	28264513	40686493	54307255	65700991	77829727
4412810	16884888	28467356	41026861	54515697	65734961	77857132
4631192	17050113	28512342	41213271	54713646	65918990	77926173
4702312	17060890	28568776	41322973	54729776	66012310	77975320
4885471	17072628	28689545	41383458	54788610	66092324	78298644
5191496	17221560	28813248	41548748	54829951	66378811	78410552
5432841	17595567	29166355	41610677	55098445	66507352	78531436
5627426	17819722	29257136	42169862	55104004	66798817	78555266
5877888	17989042	29280369	42288666	55478711	66892780	78571667
5978534	18090882	29471033	42449851	55518941	66929613	78663263
6000160	18094307	29614424	42550252	55528763	67040028	78673570
6194918	18494678	29708729	42764536	55619168	67260838	78709917
6210895	18632052	30029083	43027131	55757334	67467964	79206781
6265257	18838739	30090381	43036636	55760095	67593241	79591035
6506393	18983258	30212082	43217738	55836894	67630794	79748525
6507225	19025504	30383377	43378066	55906043	67661866	79807538
6721723	19316057	30515864	43525440	55971614	67854391	79815065
6881937	19893323	30649663	43813881	55988721	67893172	79818517
6905767	20051604	30695561	43968174	56100212	67955800	79848243
7025997	20054476	30707637	44062132	56117209	68000499	79920058
7062184	20278160	30839590	44099260	56197206	68049177	79977505
7127936	20401625	30869248	44121224	56757425	68169738	80303743
7324890	20570622	30978066	44385922	56783265	68217535	80609714
7347207	20631530	31070544	44576096	56819618	68223643	80619139

DNA methylation analyses in IMAGEN samples

7420435	20667836	31561113	44647495	56864858	68649293	80732340
7478069	20796903	31733662	44753182	56896315	68824535	80931535
7705688	20830744	31825112	44836688	56896962	68849025	80996663
7787644	20841951	31834764	45526627	57133727	69026991	81007956
7824030	20843854	32062629	45710328	57135861	69156055	81019439
7839943	20879424	32110235	45747373	57172435	69295276	81257501
8309107	20992356	32581973	45793384	57348284	69537898	81584850
8633229	21131361	32642894	46161596	57401705	69548777	81745414
8685800	21245205	32782232	46457090	57743372	69643896	81805725
8962238	21424019	33045305	46765850	57872147	69832522	81924774
9252997	21496213	33775752	47028244	57930100	70065392	82149344
9474438	21659535	33860514	47110302	57951360	70086597	82195927
9715842	21682831	33925730	47119736	57968987	70124136	82621571
9740318	21729241	34252547	47174916	58060181	70528422	82765463
9803741	21834733	34255297	47405358	58140414	70910152	82922411
10058298	21934920	34268536	47621591	58577586	71090159	83358101
10111807	22105198	34343813	47713161	58656168	71332809	83490451
10400245	22202071	34406985	47804329	59057693	71765184	83522010
10482778	22245441	34487218	48054415	59325249	71766352	83648239
10643071	22302606	34754250	48190933	59750105	72165508	83762870
10646873	22456630	34829621	48284598	59973874	72303758	83785552
10813059	22531572	34862855	48310052	60150091	72525506	83827820
11019827	22698017	35004186	48386902	60184938	72545430	83862457
11104036	22825752	35104442	48473246	60186708	72579465	84037702
11335905	22858673	35197361	48541136	60281382	72628887	84104893
11481313	23063758	35247720	48616789	60301120	72869403	84121507

DNA methylation analyses in IMAGEN samples

11491576	35515967	60451857	99561404	99604669	99550415	84138979
99677574	99747799	99875982	99954902	98989002	99104307	99217838
84158838	86076571	88636908	92126350	94345446	97659232	95507335
84210744	86481815	88683330	92744330	94768500	97707135	95578455
84247324	86727453	88835784	93240591	94998851	98194253	95683953
84343095	86975524	88874143	93371309	95023810	98205180	95886795
84838602	86977385	88956613	93455351	95077976	98442422	95957207
84915055	86992161	88968610	93535053	95300616	98516748	96138957
85068141	87216041	89179801	93543288	93759994	98537667	96358046
85417248	87248884	89444665	93649142	94090604	98562537	96783194
85498115	87316942	90194741	91038192	94180305	98847514	96849050
85724167	87495487	90208234	91159380	94265316	98929068	97084811
85791572	87694854	90434244	91414961	88329552	87954468	97096149
85965300	87825984	90814109	91899826	88607616	88315066	97555742
90945437						

Wave 2 included 600 samples selected from the age 14 individuals and was enriched for participants scoring high on ADHD at that age. For this wave, N = 187 “ADHD” individuals, non yet included in the wave 1 methylation analyses, were selected based on ADHD score ≤ 6 in the parent-rated Strength and difficulties questionnaire [1], while the remaining N = 413 samples were randomly selected. Wave 2 samples were processed by Stefan Herms at the Life & Brain Centre, University of Bonn.

The following subject IDs were processed in Wave 2:

22453	12409076	24118708	35635841	49374360	65007029	75436265	88106971
75717	12988699	24433939	35790894	49731299	65159486	75438006	88964325
458976	13097762	24436422	35868343	49743945	65320945	75460837	89047261
459848	13646768	24477110	35936114	50271212	65360091	76165026	89569994
540905	13697017	24748722	36032453	50323155	65567637	76279488	89690901
820294	14027707	24996321	36383037	51336668	65746428	76434221	89691577
829055	14045403	25033608	36630759	51445049	65833786	76466799	90253111

871047	14232183	25067315	37005099	51456409	65999822	76516256	90314183
1123104	14268884	25356336	37041127	51722290	66086086	76794351	90557023
1298929	14435937	25410823	37058553	51865546	66137615	76847922	90757912
1388096	14759699	25492584	37236832	51980538	66195730	76972990	90848850
1617607	14765341	25497997	37280746	52352041	66619767	77306439	90973178
1720362	14786947	25527440	37617376	52639093	66715400	77402830	91221714
1781386	14850118	25532814	37643558	52713040	66847819	77596543	91362411
1889268	14947520	25836330	38221741	52862216	67227272	77631822	91474448
1934375	15059321	26069147	38728463	53320582	67300471	77660233	91492921
2109942	15705369	26470206	38760566	53892023	67528603	78043249	91843900
2270310	16146243	26629318	38793721	53971707	67659985	78056005	92002804
2437251	16275727	26698359	38916315	54389991	67766736	78301795	92075592
2485525	16288795	27074970	38976393	54552397	68674859	78648508	92084654
2660085	16521332	27085653	39530825	54609183	68717425	78836365	92159995
2682961	16630568	27100173	39710910	54625022	68802565	78862327	92208408
2757585	16635570	27113512	39979193	54633971	69242825	79279134	92432902
2923651	16674081	27157405	40032718	54678674	69253270	79471229	92458933
2996147	16711878	27242238	40329097	54758073	69366941	80027030	92820066
3629479	16720517	27410033	40485442	54890823	69379423	80137080	92865408
3672597	17225963	27511167	40515445	55203542	69434928	80271792	92895731
4182298	17406332	27578624	40983027	55440249	69481559	80301458	92985697
4504055	17480176	27597482	41177354	55465932	69858614	80309952	93142747
4622874	17696979	27858889	41240671	55520374	69867258	80383784	93247658
4704480	17994216	27986310	41266477	55793162	69920617	80461839	93454722
4794212	18113022	28113492	41434646	55960242	70198928	80961475	94141243
5019076	18155653	28185341	41748802	55969914	70407053	81006258	94243275
5063402	18217001	28359931	41765496	56888495	70506641	81061941	94259613
5451728	18283421	28490711	41966278	57046701	70675464	81092960	94428008
5571072	18349150	28827850	42432941	57113773	70920329	81257882	94594928
5642417	18904077	29258014	42479340	57151570	71107361	81634564	94807913
6260065	18970311	29313096	42643344	57508486	71138289	81716527	94878465
6314205	19031295	29427900	42673088	57515374	71218732	81795463	94989507
6504813	19413655	29658673	42749560	57533759	71293261	81893696	95263539
6661511	19429616	29751086	42878935	57541508	71422292	81956801	95913035

DNA methylation analyses in IMAGEN samples

6975136	19432693	29980214	42979666	57598971	71566792	82047033	95958758
6997449	19543161	30151696	43040398	57728231	71572282	82205224	96196382
7047216	19683060	30374303	43102413	57847102	71640147	82447909	96245982
7167925	19912904	30385648	43189743	57944398	72006664	82880964	96498790
7580813	20013673	30527827	43347465	58066497	72155879	83037309	96528581
7719491	20035682	30742768	43784766	58616416	72291976	83057429	96892547
7779851	20320615	30954633	43926786	58634281	72299221	83095622	97067339
7843654	20355013	31622058	44165129	58846772	72358748	83194915	97386630
7965462	20451157	31630733	44775808	58908172	72747147	83237051	97456086
8089574	20523107	31655253	44982406	59169753	72882642	83252295	97759862
8592847	20537630	31851798	45049720	59343528	72947804	83539447	97798177
8996449	20776838	31966463	45243088	59355209	72966134	83656920	97970921
9219140	20785775	32067534	45376266	59791765	73113515	83670951	97982061
9329324	20838293	32116540	45835265	60380585	73216432	83780043	98241852
9354168	21053574	32506627	46320377	60587771	73231355	83830785	98301449
9551323	21073482	32617813	47102665	60774324	73420500	84123801	98600208
9616716	21114224	32899225	47117408	60941904	73639449	84167593	98606547
9736077	21218872	33183713	47157916	60942682	73858659	84791358	98720163
9808250	21415101	33889562	47188050	61050550	74172434	84828767	98737210
9963964	22053782	34196193	47381951	61298230	74179558	86071132	98958194
10121092	22465536	34241299	47501131	61504177	74247248	86116554	99002488
10394358	22482211	34277415	47507426	61538083	74304888	86129563	99297662
10867560	22888407	34279191	47690807	61588737	74327708	86259237	99514690
10917199	23032425	34458922	47864744	61781607	74384224	86458347	99529479
10974785	23426080	34508719	47938666	62050071	74586509	86642327	99559419
11310758	23580957	34532105	48135131	62210205	75074865	86673294	99580654
11662485	23896510	34617147	48797829	62252904	75198966	87249996	99616225
11922061	24002283	34621212	48828414	63043079	75273205	87280429	99888850
12029891	24055584	34646163	48908846	63366503	75278325	87298622	
12079621		34829301		64043761		87738590	
12089655		34924989		64233992		87817879	
12138673		35061633		64640797		87961112	
12203611		35402331		64814274		88060095	
12391077		35410225		64851530		88079695	

N.B: The raw (i.e., non-normalised) DNA methylation data is contained in the folders:
`/DNA_Methylation_Both_Waves/Raw_data/idat450k_wave1` and
`/DNA_Methylation_Both_Waves/Raw_data/idat450k_wave2`

2. Quality control of methylation data

After removal of samples with insufficient DNA (N= 14); duplicates (samples for which DNA was sent both in the first and in the second wave, N = 6) and mislabelled (during DNA extraction, N = 1), quality control and normalization was performed on DNA methylation from N = 1287 individuals.

2.1. Preprocessing / normalization

Preprocessing of raw microarray intensities into usable methylation measurements was performed on the combined sample (wave 1 + wave 2). Raw DNA methylation signals were computed into β values [$\beta = (\text{methylated signal})/(\text{methylated} + \text{unmethylated signals})$ at each locus]. These methylation ratios were transformed by logit transformation, as such transformation has been found to show heteroscedasticity [2]. Within- and between-sample normalization was performed using the stratified quantile normalization implemented in the Bioconductor minfi package [3].

(Note: the algorithm used for such a normalization method relies on the assumptions necessary for quantile normalization to be applicable and thus is not recommended for cases where global changes are expected, such as in cancer-normal comparisons).

2.2. Data Filtering

Detecting and removing low-quality samples that normalization cannot correct has been shown to improve downstream results. Thus, before proceeding to data analysis, quality control was performed using the minfi package [3].

Bad samples were removed by using the `removeBadSamples` argument in the `preprocessQuantile` function of the minfi package. Briefly, the median Meth and median Unmeth signal for each sample were calculated and the samples where their average falls below the default `badSampleCutoff` were removed (all samples from wave 1 and wave 2 passed this QC step). All probes, including those containing SNPs that might affect the measurement of DNA, were kept. Note, these SNP containing probes are recommended to be removed for DMP analyses, but can be kept for DMR analyses.

2.3. Quality assessment

2.3.1. PCA distribution of methylation signals

Figure 1 shows results of principal component analysis of methylation level for all probes on autosomes. The first 4 PCs were selected for further control analyses.

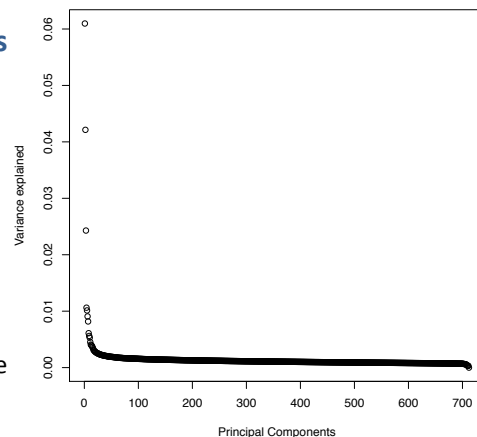


Figure 1. Screen plot indicating the percentage of variance explained with the components extracted.

2.3.2. Quality of overall methylation signals

Plotting hybridisation signals against the first PC (Figure 2) indicated that all samples showed good methylation signals, except the 8 methylation controls, that are either fully methylated or completely unmethylated.

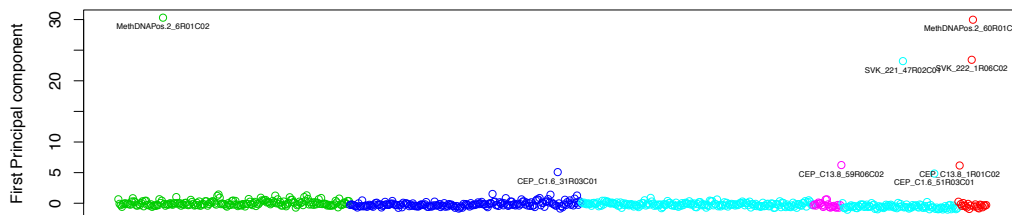


Figure 2. Overall methylation signals with control samples included. Colour indicates the date of hybridisation.

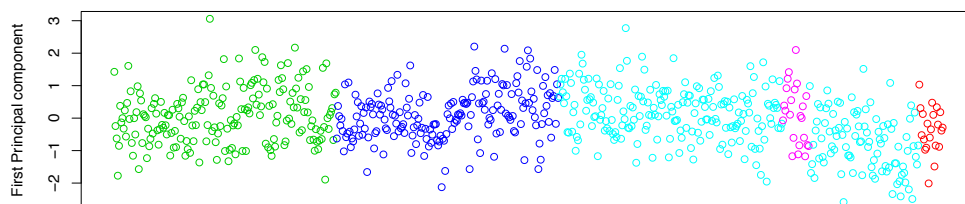


Figure 3. Methylation signal after removal of control samples. Colour indicates the date of hybridisation.

2.3.3. Predicting sex using methylation data

Sex information was estimated by using getSex function in the minfi package, which estimates samples sex based on the methylation values on the sex chromosomes. This estimated sample sex was verified with reported gender information. One subject showed gender data that did not match the reported gender information (Figure 4). This individual was removed from downstream analyses.

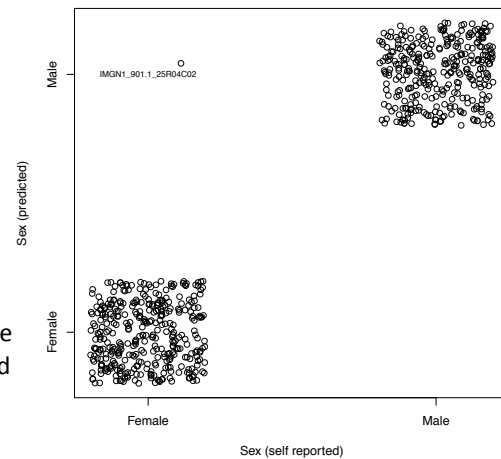


Figure 4: Prediction of gender using methylation data. One individual has wrong gender information and was removed from downstream analyses.

2.3.4. Batch effects

The signal recorded from the methylation data was assessed for confounding effects as outlined below.

- **Gender effects**

No effects of gender on the PCs for methylation levels on autosomes were detected (as shown in Figure 5a and b). The colour indicates gender.

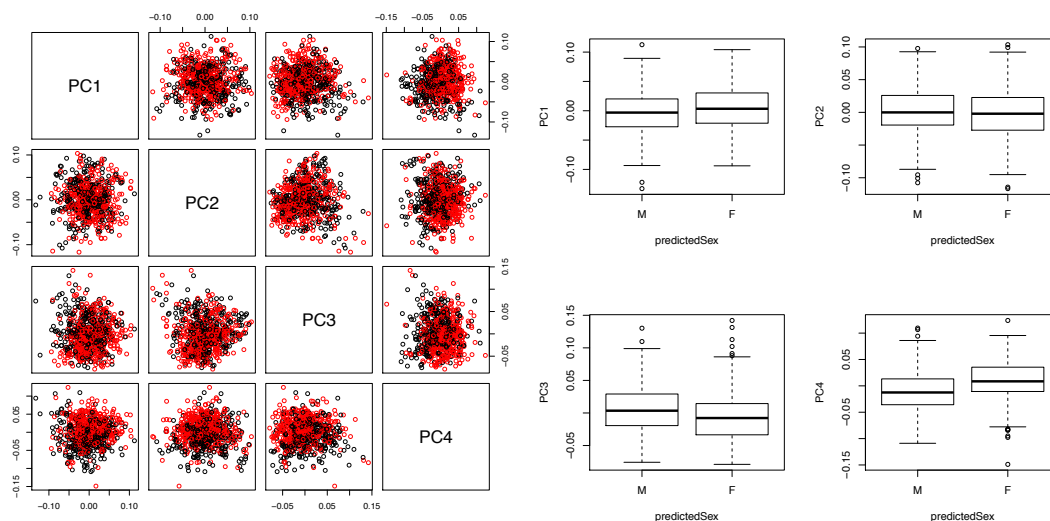


Figure 5. a) Scatterplots showing gender effects on PCs. Gender is colour coded. b) Main effects of gender on PCs of methylation levels.

- **Effects of location on the arrays**

No significant effects of location on the arrays on PCs of methylation levels were found (Figure 6).

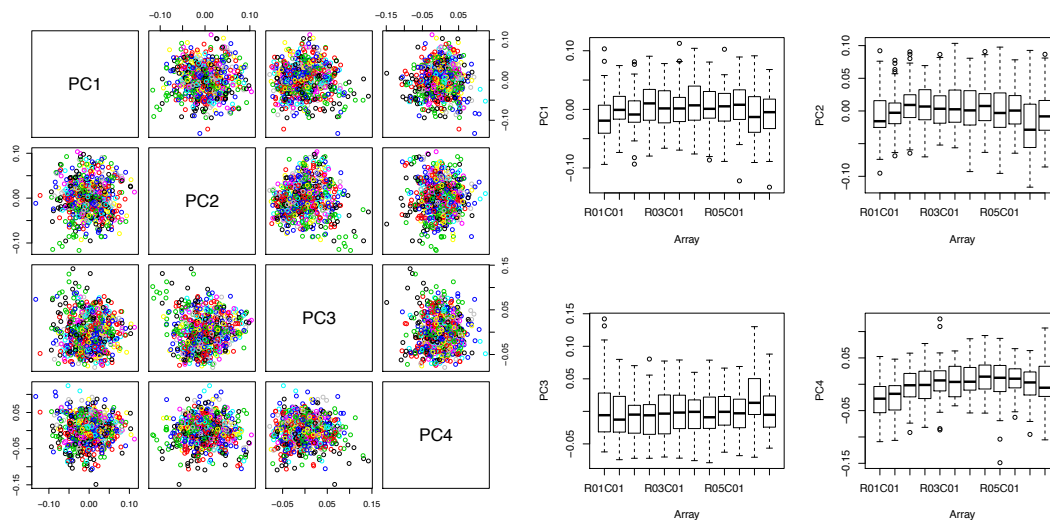


Figure 6. a) Scatterplots illustrating the associations between location on array and methylation level across the PCs. Colours indicates the array location. b) Location on array plotted on x-axes, level of methylation per PC plotted on y axes.

- **Effects of hybridisation date**

There were weak batch effects associated with hybridization date (PC3) (Figure 7).

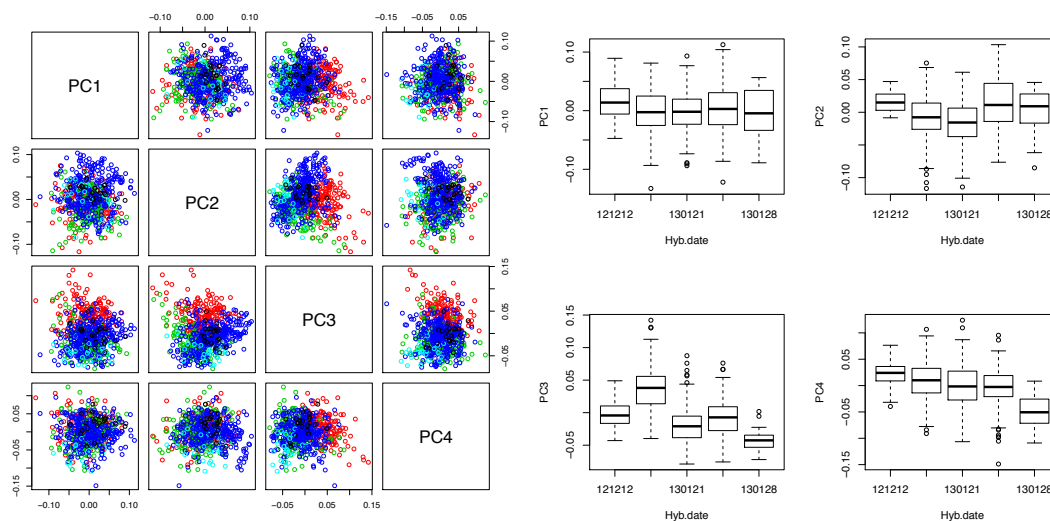


Figure 7. a) Scatterplots illustrating the associations between date of hybridisation and methylation level across the PCs. The colour indicates date of hybridisation. b) Main effects of hybridisation date on methylation levels across PCs.

- **Effects of site**

There are weak batch effect associated ImagineCentreID (PC3) (Figure 8).

Note: ImagineCentreID is strongly associated with hybridization date

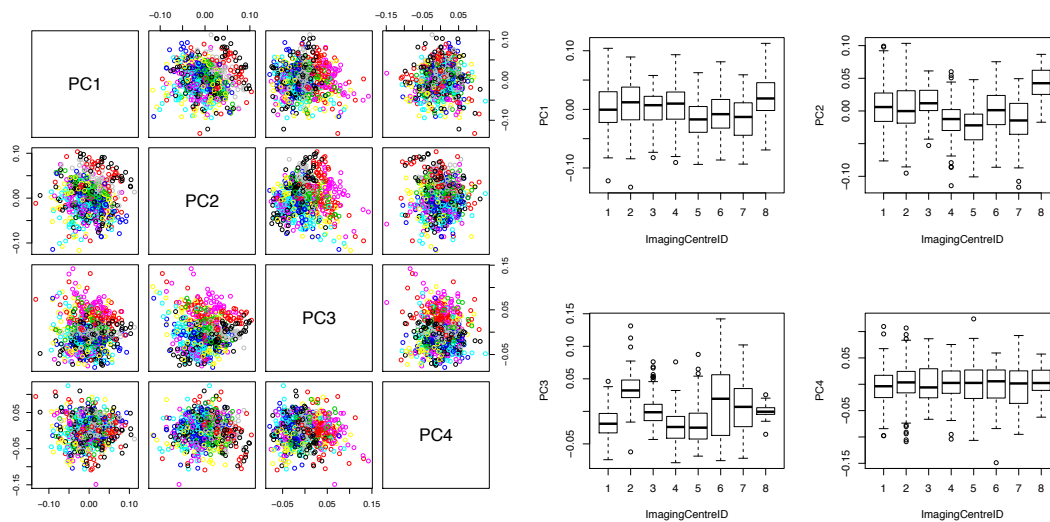


Figure 8. a) Scatterplots illustrating the associations between methylation levels and site. The colour encodes the sites. b) Main effects of site on methylation levels across the PCs.

- **Effects of cell types proportions**

Strong association between methylation data (PC1 and PC2) and estimated cell proportion, as expected (Figure 9). For each individual, cell counts for the 6 major cell types present in blood (i.e., CD8⁺ cytotoxic T cells (CD8T), CD4⁺ helper T cells (CD4T), natural killer cells (NK), B cells, monocytes (Mono) and granulocytes (Gran)) were estimated using the estimateCellCounts function in the minfi package. This function obtains sample-specific estimates of cell proportions based on reference information on cell-specific methylation signatures [4]. As shown in Figure 9a, the proportion of cell types varies, therefore we have to include the cell type proportion into the analyses, especially for granulocytes, as these are associated with methylation levels across the first and second PC, see Figure 9b.

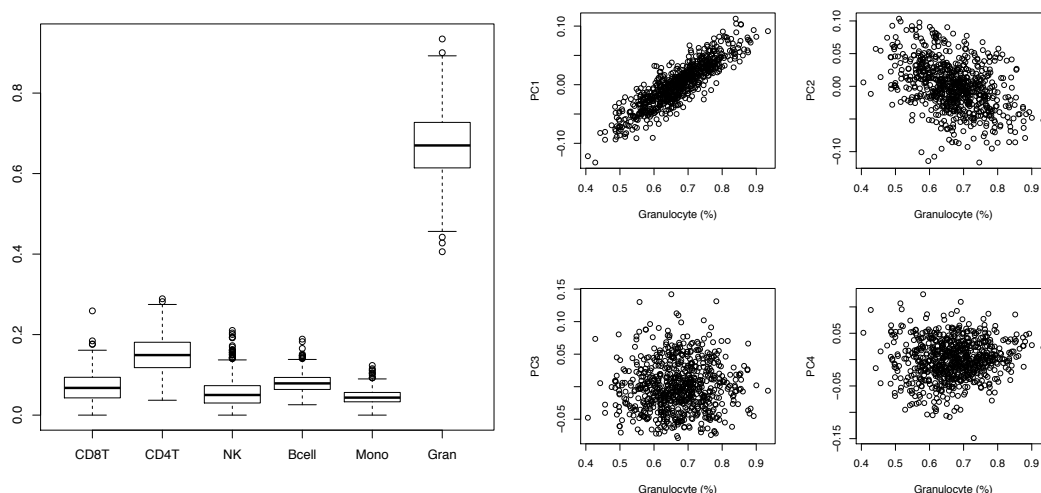


Figure 9. a) Proportions of cell types and their variability. b) Scatterplots illustrating associations between granulocytes and methylation level across PCs.

- **Effects of different waves**

Strong effect of wave on global methylation profile: as shown in Figure 10, samples from the two different waves have different global methylation measurements. Thus, we have to include the experimental wave information in the downstream analyses.

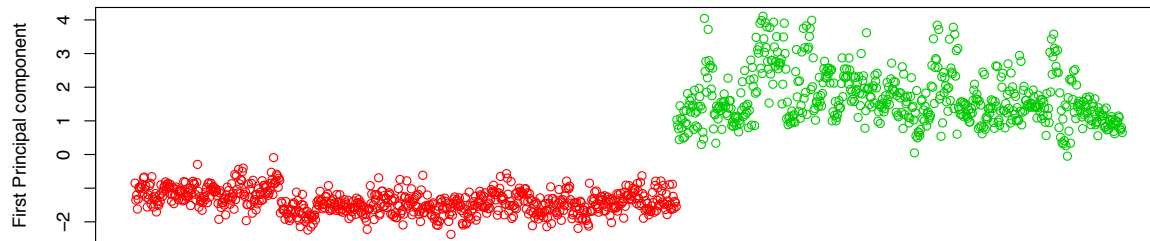


Figure 10. Methylation measurements for samples from different waves. Red: samples from wave 1; Green: samples from wave 2.

2.3.5. Other information important for secondary analyses

As mentioned above, the methylation data was preprocessed together (wave 1 + wave 2) using the stratified quantile normalization method implemented in the Bioconductor minfi package (see section 2.1.). To identify differentially methylated regions (DMRs) associated with a phenotype, we used the bump hunting method [5], implemented by the `bumphunter` function in the minfi package, with adjustment for the following confounders: gender, experimental waves and the first two principle components of estimated differential cell counts. Regions that had a family-wide error rate (FWER) less than 0.1 with 1000 resamples and contained at least 2 probes were identified as phenotype-associated DMRs.

N.B:

1. **Data:** Normalised & QCed DNA methylation data is contained in the folder: **DNA_Methylation_Both_Waves/Normalised_data/Beta.rds**
2. **Covariates:** Covariates to include in analyses are in the folder: **DNA_Methylation_Both_Waves/ Covariates.rds**

References

1. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry*. 1997;38:581-586.
2. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC bioinformatics*. 2010;11:587.
3. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30:1363-1369.
4. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH,
5. Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell
6. mixture distribution. *BMC Bioinformatics*. 2012 May 8;13:86.
7. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *International journal of epidemiology*. 2012;41:200-209.