**Sample QC ARIES data - Josine Min**

**Descriptives**

**ARIES samples**

In total there are 5469 samples for five timepoints (cord=1127; F7=1086; 15up=1073; antenatal; 1100; FOM=1083) in the LIMS belonging to the 1022 mother-child pairs. The 15up samples are a mixture of 277 samples of kids at TF3 clinic at 15 yrs and 796 TF4 clinic at age 17 yrs (Table 1). FOM consists of FOM (mothers at FOM clinic) and TF1-3 (mothers at child clinics TF1 to 3) (Table 1).

**Table 1 Timepoint versus sampletype for methylation data**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Timepoint** | **Timecode** | **Blood Spots** | **pbl** | **white cells** | **whole blood** |
| **cord** | **cord** | 241 | 0 | 886 | 0 |
| **F7** | **F7** | 0 | 0 | 68 | 1018 |
| **15up** | **F17** | 0 | 0 | 796 | 0 |
| **15up** | **TF3** | 0 | 0 | 277 | 0 |
| **antenatal** | **antenatal** | 0 | 0 | 503 | 597 |
| **FOM** | **TF1-3** | 0 | **184** | **5\*** | 0 |
| **FOM** | **FOM** | 0 | 0 | **894** | 0 |

**\*** Sampletype “DNA consent” is coded as white cells after correspondence with Wendy McArdle.

**Technical replicates**

There are 369 technical replicates (361 duplicates + 8 triplicates). Most of these samples are poor quality samples (low detection score) and were therefore repeated in the lab.

**Batches**

There are 134 batches of BCD plate (1-49 samples on each plate) and 518 batches of slides (1-12 samples on each slide). Samples from same plate ended on the same slide using a semi-randomisation procedure. Sample\_type2 is recoded from the samcode. There are differences in cell type composition for each sampletype. White cells are buffy coats (comprising lymphocytes, monocytes and granulocytes) and can be collected with heparin or EDTA tubes (coded in “additive”). PBLs are peripheral (mononuclear) blood lymphocytes and they comprise T cells, NK cells and B cells and collected with CPDA tubes. Blood spots are cord blood samples and are taken at birth. Blood spots are not taken from heel prick.

**Age**

Age are taken from the following variables: f7003c (F7), fh0011a (TF3), FJ003a (TF4), mz028a (antenatal), fm1a011 (FOM). This doesn’t include TF1-TF3 mums.

Ages for TF1-3 mums are taken from TF clinic files and compiled by Kate Northstone (file: ARIES\_age\_mums\_090315.xlsx). For TF1-3 mums, clinic and age of sampling is unknown. Kate Northstone: “There are a handful of duplicates”/ "Indicator vars have been created to show which clinic the sample MAY have been taken at (MAY because there was overlap between TF1 and TF2 - so we can't be sure which clinic she attended - match to actual clinic data to be sure)."

**Sample mismatches based on genotype QC**

4904/5469 ARIES samples have been successfully genotyped. Table 2 shows the exclusion criteria for genotype QC. 112/5469 samples failed genotype QC due to sample swaps, gender mismatches, high IBD or relatedness issues between mums and kids (“Other”). Table 3 shows samples that fail the combined mum-kid genotype QC. All samples in Table 3 should be removed.

**Table 2 Sample Counts for samples with methylation and genotype data and exclusion criteria in genotype QC (based on 5469 ARIES samples)**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Timepoint** | **450k\*** | **GWA QC Pass** | **no GWA** | **pop strat** | **MM** | **Sex MM** | **HZT** | **high IBD** | **Other\*\*** | **IBD10/CR** | **Miss** |
| cord | 1127 | 919 | 12 | 49 | **1** | **5** | 23 | **0** | **18** | 11 | 2 |
| F7 | 1086 | 887 | 11 | 47 | **1** | **5** | 22 | **0** | **15** | 11 | 4 |
| 15up | 1073 | 874 | 12 | 46 | **1** | **6** | 21 | **0** | **17** | 11 | 2 |
| antenatal | 1100 | 886 | 70 | 36 | **0** | **3** | 1 | **1** | **15** | 25 | 2 |
| FOM | 1083 | 870 | 69 | 31 | **0** | **3** | 1 | **1** | **19** | 23 | 2 |

**\*** Includes technical replicates

**\*\*** Passed mum and kid genotype QC but are not present in final genotype files i) Mother-offspring pairs with different aln and genetic relatedness of 0.5 were often excluded because their alns didn't map to their individual id in the original platemap file we used from the lab. ii) Mother-offspring pairs with same aln and genetic relatedness of zero. iii) Mother-offspring pairs with same aln and genetic relatedness of 1. MM=mismatch; HZT =heterozygosity; CR=cryptic relatedness; Miss=missingness

**Table 3 Sample counts for exclusion criteria in mum-kid genotype QC**

|  |  |
| --- | --- |
| **mum-kid genotypeQC\*** | **N** |
| mother-offspring pairs with a genetic relatedness of 1 | 6 |
| mother-offspring pairs with different aln and genetic relatedness of 0.5 and one other piece of conflict information | 56 |
| mother-offspring pairs with same aln and genetic relatedness of zero | 23 |

\* These samples are in Table 2 as “Other”.

**1A) Sample mismatches using GWA concordance (Figure 1)**

Sample mismatches are checked by comparing genotypes from 65 SNPprobes on the array to external genotypes for 4904 samples that passed genotype QC. I extracted the betas for the 65 snp probes and used kmeans with probabilities of 0.2, 0.5 and 0.8 to call genotypes. I checked the *SNP concordance* and 62/63 genotyped SNPs had a concordance above 95% with the SNPs called from the genotype arrays. For the 62 SNPs with >95% concordance, I calculated concordance across all samples and got a 5469 by 5469 matrix. I then checked the maximum concordance across unrelated individuals (ALN was not matching). The maximum concordance observed for unrelated individuals was smaller than 80%. For each of the 4904 samples with genotypes, I calculated *sample concordance*. We found 411 genotype mismatches with a concordance <80% and these samples are removed.

**1B) Sample mismatches using SNPprobes: concordance across family (Figure 1)**

To identify mismatches, I calculated a 5469 by 5469 concordance matrix. I performed multiple iterations to remove samples and checked the removal strategies against the GWA discordance list. I identified mismatches by:

1. mum samples have more than 90% concordance with a kids sample (22 samples)

2. checking of duplicates, 200 samples were removed as concordance was <80% between duplicates and less than 80% concordance with at least one other mums of kids sample.

3. 10 mums samples of which concordance was <80% between duplicates

4. 10 samples with low concordance (<80%) with other timepoints

5. 24 mums samples with low concordance (<80%) with other timepoints

**2) Sample mismatches using sex check (Figure 2)**

Sample mismatches are checked by a sex check. By comparing the median total intensity of the Y-chromosome-mapped probes to the median total intensity of the X-chromosome-mapped probes, where the total intensity is the sum of the methylated and unmethylated signals, it is possible to predict the gender of the sample by looking at the two distinct clusters of intensities. 161 samples showed clear sex mismatches whereas 30 samples were X-Y ratio outliers. Both categories are removed.

**2) ControlProbes (Figure 3 & Figure 4)**

The IlluminaHumanMethylation450k array was used comprising several types of probes.

Number of type I probes: 135476

Number of type II probes: 350036

Number of control probes: 850

Number of SNP type I probes: 25 Number of SNP type II probes: 40

**Control Type:**

bisulfate conversion I (N=12)

bisulfate conversion II (N=4)

hybridization (N=3),

extension

negative (N=614)

non-polymorphic (N=4)

norm A (N=32)

norm C (N=61)

norm G (N=32)

norm T (N=61)

specificity I (N=12)

specificity II (N=3)

staining (N=6)

target removal (N=2)

**Methylated vs Unmethylated (Figure 3)**

I checked the sample quality by checking the raw intensity of the 850 control probes. The QCplot in Figure 3 shows the median intensity methylated vs unmethylated signal for all control probes. It shows 68 outliers. These Methylated vs Unmethylated outliers are removed.

**ControlType intensities and Dyebias (Figure 4)**

Figure 4 shows plots for each controlType where average intensity per sample has been plotted and plots are ordered by BCD plate. Appendix shows the number of probes for each Control Type, dyebias is based on normalization probes.

Several issues are seen:

1) Nine samples are showing dyebias. Slides 8784225103 (4/7 samples), 8784225091 (4/6 samples) and 8963282048 (1/11) show dyebias issues. Thirteen samples from slides (8784225103 (7 samples), 8784225091 (6 samples) are completely removed from the analysis.

2) Staining probes have intensity of 0 (Green=131 samples; Red=73 samples; Both = 40 samples)

The results from 8667053024 & 8667053095 were reviewed by Illumina’s tech support because of the low values for the red and green staining controls. These are detected using control beads added to the bead pool by Illumina and again, the signal for these control beads are absent from some chips and their response was that the results are OK (email correspondence Wendy McArdle).

3) nonpolymorphic probes have intensity of 0 (Green: 65 samples)

There are issues with poor QC on the part of Illumina so some chips don't have all the controls present, an issue we see most often with one of the four non- polymorphic controls being absent  - Illumina's tech supports answer was that if all the other controls look OK then the absence of these can be ignored. For example, 6929793151 does have one of the non-polymorphic controls missing but the other three are present (email correspondence Wendy McArdle).

4) Outlying values for extension and specification probes. These values are based on a small number of probes and are ignored.

**Detection scores (Figure 5)**

*Samples*

Illumina reports a detection p-value, which represents the confidence that a given CpG site is distinguishable from the negative control probes. Standard protocol by Illumina recommends excluding probes that have a detection *P*-value greater than an *arbitrary* cutoff of 0.05. Detection pvalues were extracted from the original idat filesand I used p>0.01 as threshold for a poor measurement. I plotted the percentage of failed probes for each sample (Figure 5). There were 166 samples with a high proportion of undetected probes (proportion of probes with detection p-value > 0.01 is > 0.1).

**Number of Beads (Figure 6)**

*Samples*

For each sample and probe the number of beads can be extracted. I calculated the number of probes that failed the threshold of 3 beads. I plotted the percentage of failed probes for each sample (Figure 6). There were 2 samples with a high proportion of probes with low bead number (proportion of probes with bead number < 3 is > 0.1).

***Probes***

**1) Detection scores and nbeads (Figure 7 and 8)**

I calculated the fraction of failed samples for each probe based on det P and nbeads (Figure 7 and 8). We used a threshold of 0.1 to remove probes. There were 4176 probes with only background signal in a high proportion of samples (proportion of samples with detection p-value > 0.01 is > 0.1). There were 64 CpGs with low bead numbers in a high proportion of samples (proportion of samples with bead number < 3 is > 0.1). Six probes showed low bead numbers and detection scores. In total, 4246 probes are removed. I noticed that some regions behave poorly such as HLA region.

**Normalization**

We used the R package meffil to normalize the data. Functional normalization is a between-array normalization method for the Illumina Infinium HumanMethylation450 platform and an extension to quantile normalisation. It removes unwanted *technical* variation by regressing out variability explained by the control probes present on the array. It is expected to be most useful in datasets with large differences between samples. For ALSPAC, it might be useful to remove the plate effects, chip row and slide (technical variation) but keep biological differences such as sample type and timepoint differences. The normalization procedure is applied to the Meth and Unmeth intensities separately, and to type I and type II signals separately. For the probes on the X and Y chromosomes, males and females are normalized separately using the gender information. For the Y chromosome, standard quantile normalization is used due to the small number of probes, which results in instability for functional normalization. As we observed dyebias and background signal based on the negative probes I used a background correction and a dyebias correction.

**Post normalization QC**

**Number of PCs to use for functional normalization (Figure 9).**

Figure 9A shows the variance explained for each PC after conducting a PCA on the control matrix. It shows that the recommended 2 PCs by Fortin *et al.* is not appropriate for our data. Figure 9B shows a validation curve where the data has been splitted in 10 subsets of which 9 datasets were used to predict the 10th subsets using the factor loadings of the 9 subsets. PC=10 seems to be a logical choice.

**Outlier check (Figure 10)**

To identify outliers in the data, I combined all mums and kids and conducted a PCA analysis on the 20,000 most variable autosomal probes on the normalized data. The data was normalized using 15 PCs. There are two clusters: cord and not cord. There were two clear outliers when I compared cord samples vs not cord samples and these are removed.

**PCA on control matrix (Figure 11)**

This PCA plot based on PCA of 840 control probes shows that there is no structure according to plate, slide, sample type or time point.

**PCA on normalized data (Figure 12)**

We used 15 PCs to regress out technical variation. To check the effect of normalization, PCA was done on the 20,000 most variable autosomal probes on the normalized data. PC1 discrimates cord samples from the other timepoints. Plots show clustering according to timepoint.

**Association between PC and batch variables (Figure 13)**

We performed Anova tests and t-tests to determine the association between PC1-10 and several batch variables (Slide, plate, sampletype, time code and timepoint) extracted from the control probes. Figure 13 shows that all these variables were associated with the control probe PCs.

**Association between PC and batch variables (Figure 14)**

We performed Anova tests and t-tests to determine the association between PC1-10 and several batch variables (Slide, plate, sampletype, time code and timepoint) on PCs extracted from the normalized data. Figure 14 shows that all these variables were associated with the PCs extracted from the betas. In PC3 and PC5, Slide 8963266138 is an outlier. This slide will be removed from the analysis. This chip contained a cell line control which also appeared to be an outlier.

**Clean data:**

**Sample QC**

All 615 outliers from Table 3 are removed. In total there are 4854 samples in the clean dataset.

**Table3: Number of outliers**

|  |  |
| --- | --- |
| **Issue** | **N** |
| Control probe (dye.bias) | 14 |
| Detection p-value | 166 |
| failed GWAS QC | 112 |
| genotype concordance (>90%) between mum-kid | 22 |
| genotype discordance (80%) across duplicates | 200 |
| genotype discordance (80%) across mums duplicates | 10 |
| genotype discordance (80%) across timepoints | 10 |
| genotype discordance (80%) across timepoints with mums | 24 |
| Genotype mismatch | 411 |
| Low bead numbers | 2 |
| Methylated vs Unmethylated | 68 |
| Sex mismatch | 161 |
| X-Y ratio outlier | 30 |
| Post-Normalization PCA | 2 |
| Post-Normalization t-test | 11 |

**Technical replicates**

In the clean data there are 71 replicates (between 9-25 for each timepoint). In the manifest file there is a column remove.duplicates which are 71 replicates that should be removed from the data if you don’t want to use replicates. We kept the sample with the highest number of detected probes.

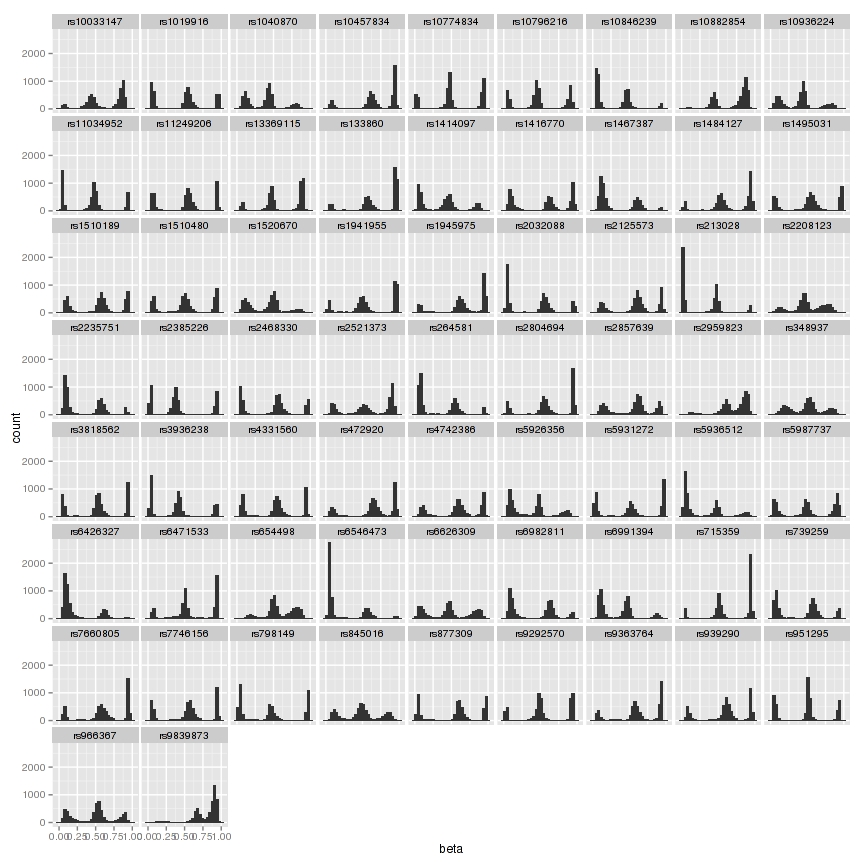
**Table4 Number of replicates (N=71) by time point**

|  |  |  |
| --- | --- | --- |
| **Timepoint** | **Timecode** | **Number of replicates** |
| Cord | Cord | 9 |
| F7 | F7 | 10 |
| TF3 | 15up | 1 |
| F17 | 15up | 10 |
| antenatal | antenatal | 25 |
| FOM | FOM | 15 |
| TF1-3 | FOM | 1 |

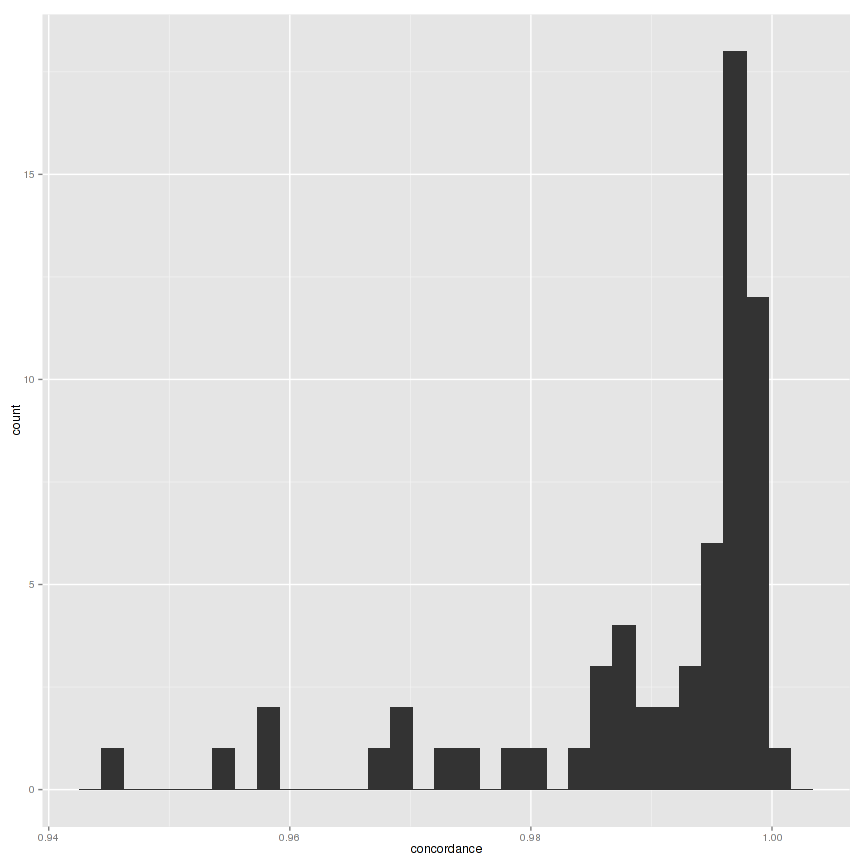
**Cell counts (Figure 16)**

Meffil has a function called meffil.cell.count.estimates. It normalizes your sample to a reference blood count dataset (blood gse35069) and then calculates cell counts using the Houseman et al. method.  We have normalized each sample individually to the cell type reference dataset rather than normalizing all samples in a dataset to the cell type reference in order to avoid having cell count estimates depend on the other samples being included in the normalization. Figure 16 shows the distribution of the blood cell types.

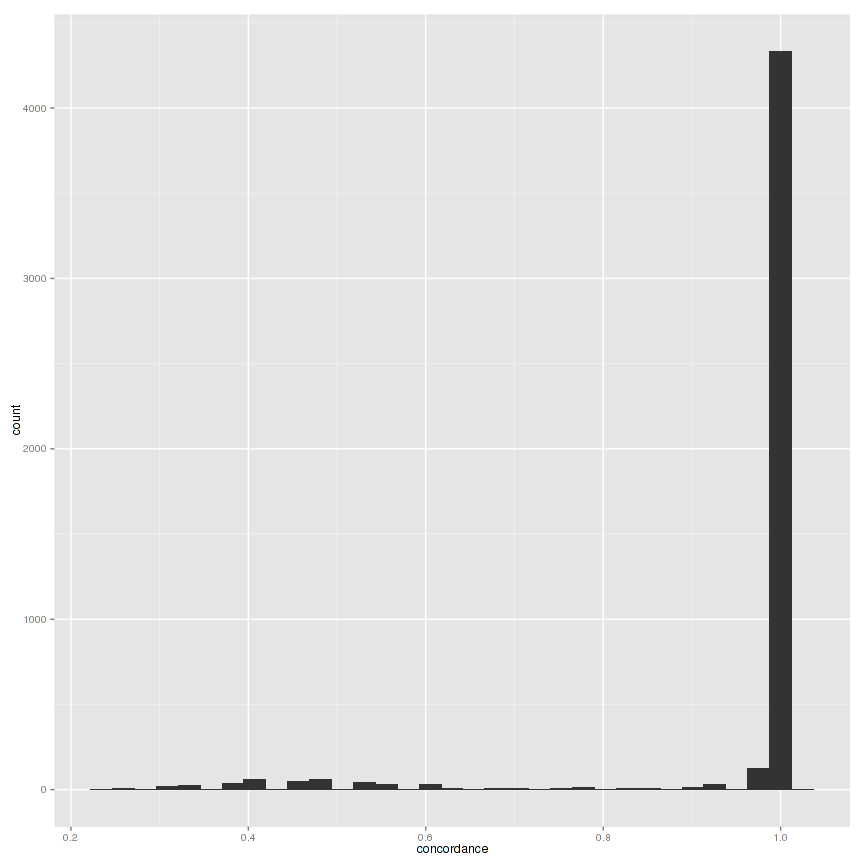
**Figures**

****

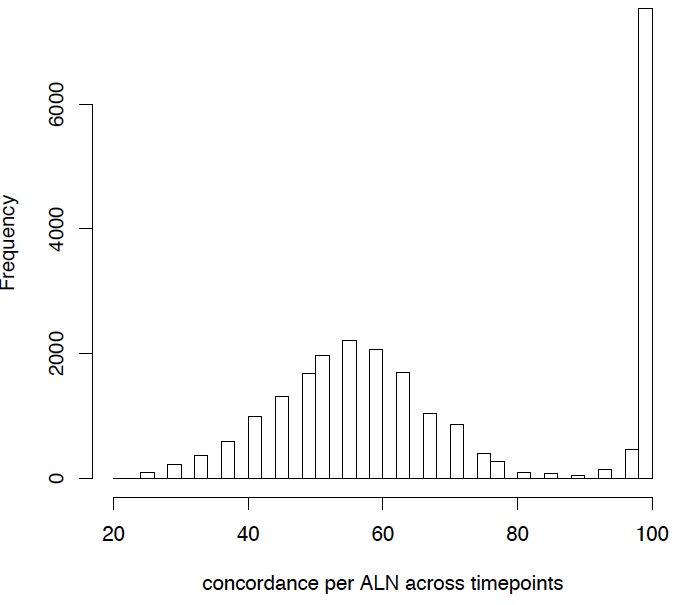
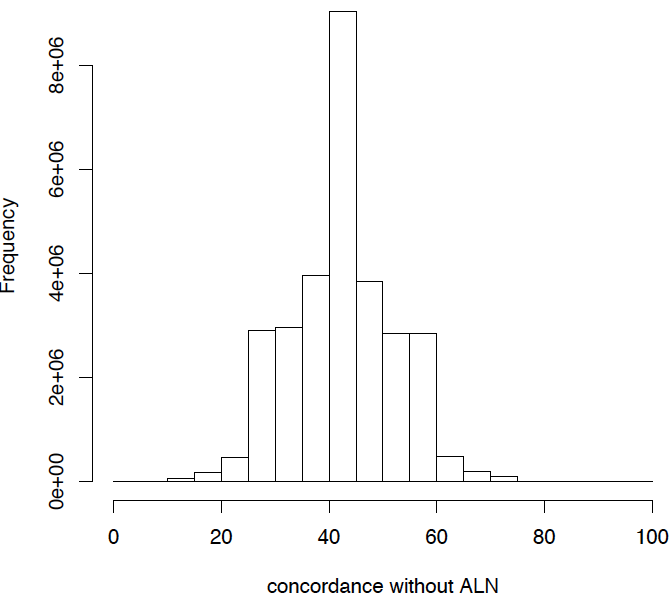
**A.**

****

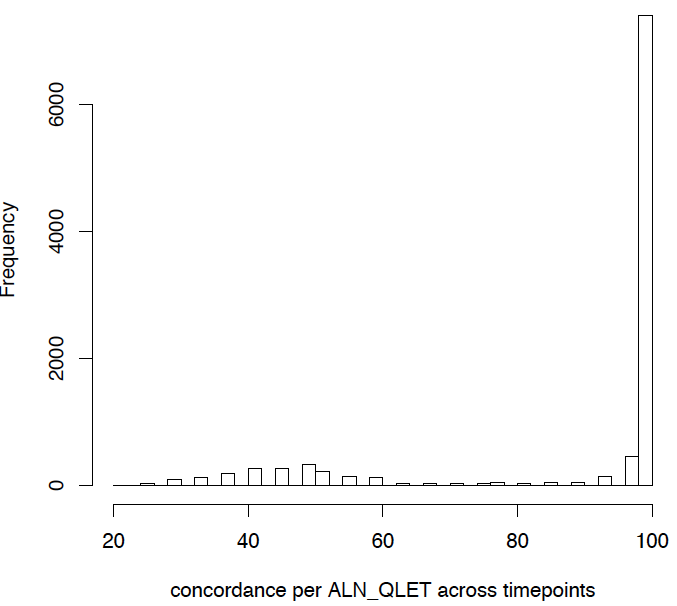
**B.**

****

**C**

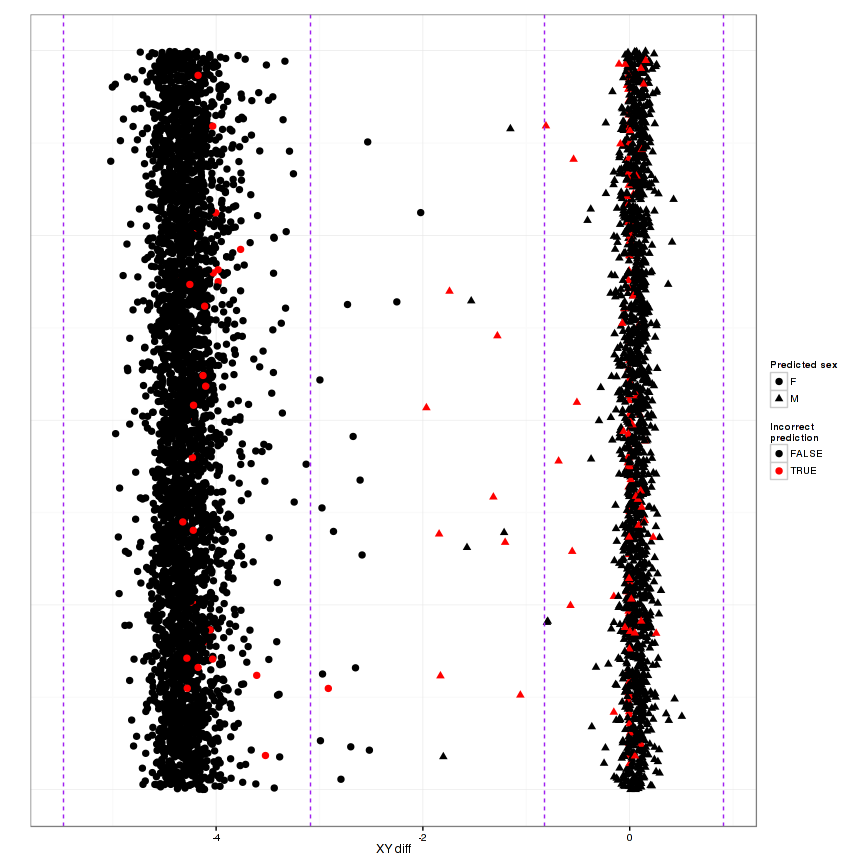
****

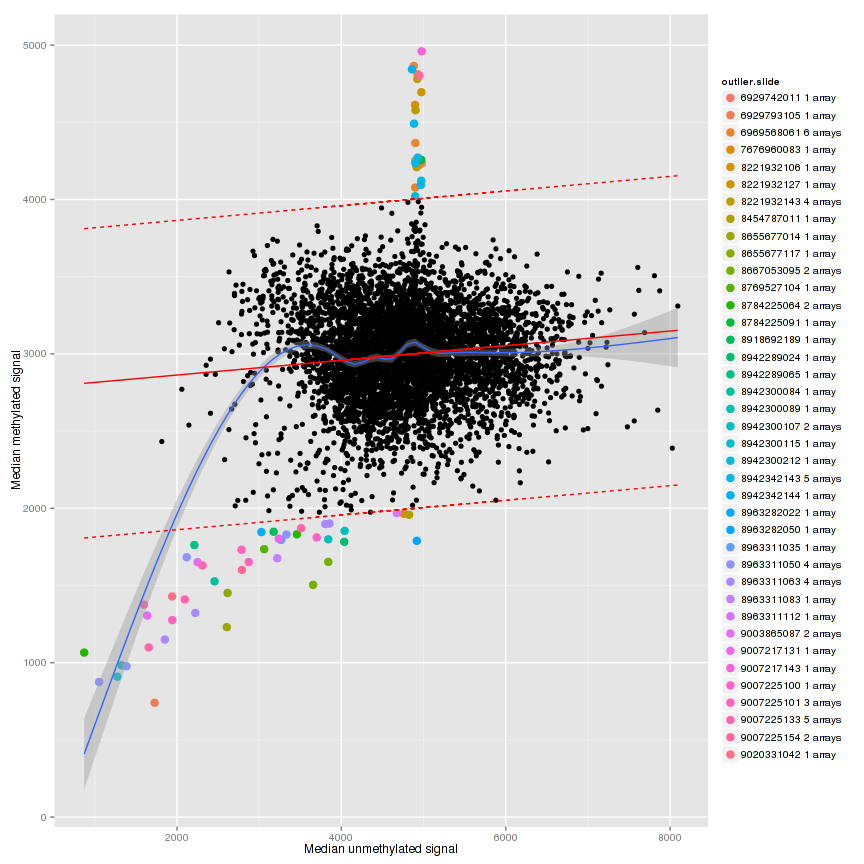
**D E**

****

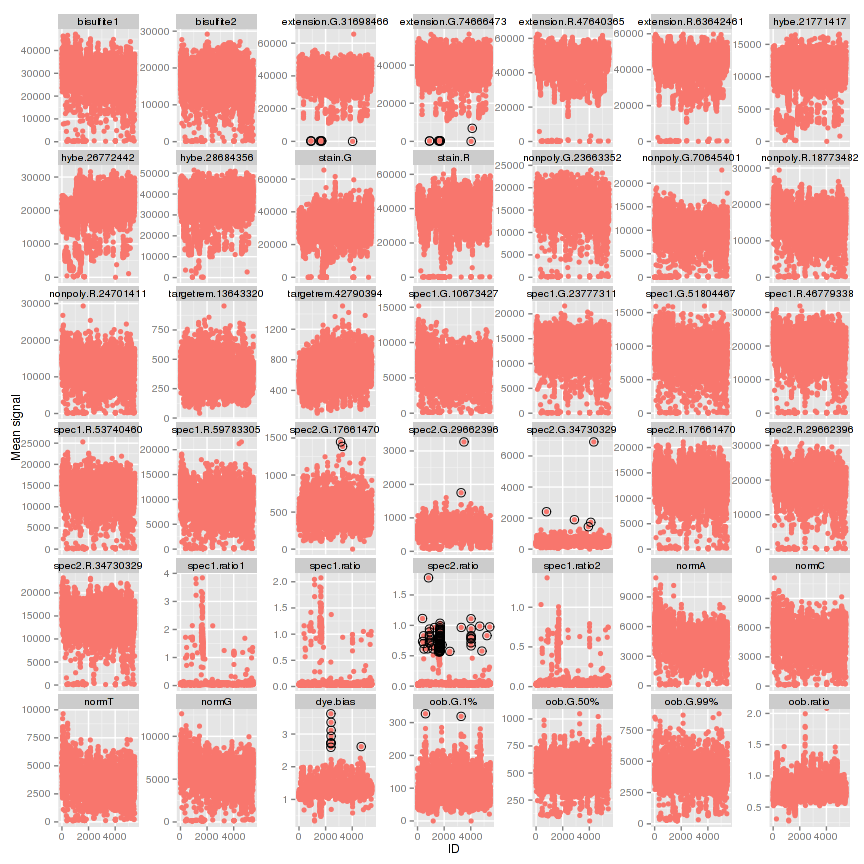
**F.**

**Figure 1:** A: Distributions of 65 SNPs measured on 450k. B. SNP Concordance between 65 SNPs on 450k versus SNPs measured on genotype arrays. One SNP (rs6471533) has a SNP concordance below 95% and is removed to compute sample concordance. C. Sample concordance based on GWA discordance. D Sample concordance for non-matching ALN sample pairs. E. Sample concordance for matching ALN sample pairs F. Sample concordance for matching ALN\_QLET sample pairs.

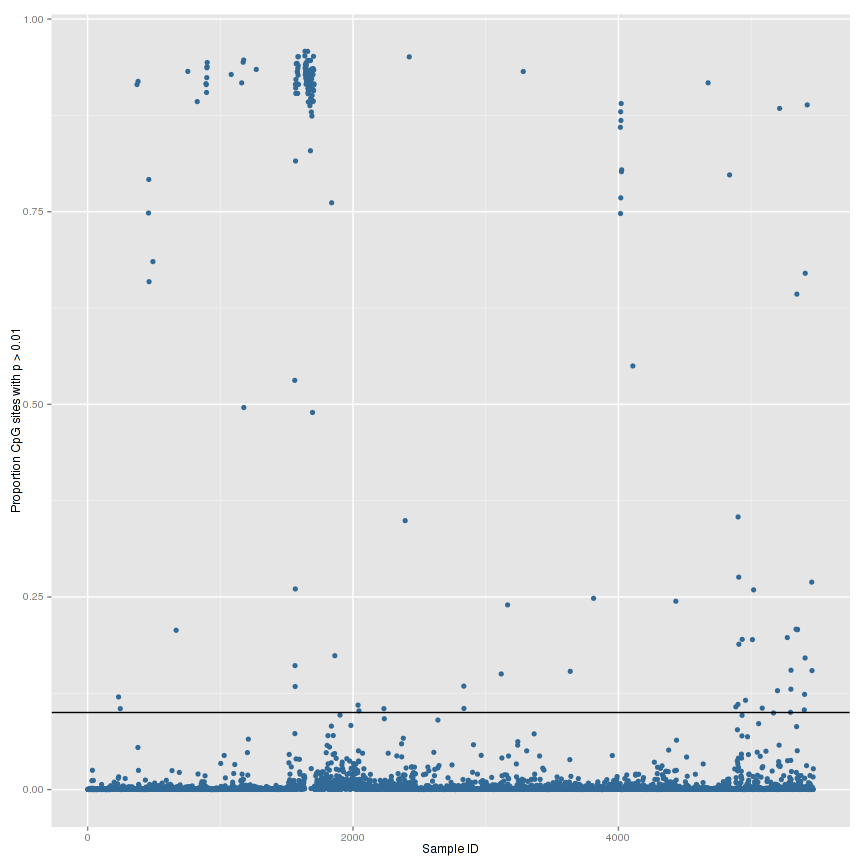
**Figure 2** Sex prediction by median total intensity of chrX and Y. Prediction has made using meffil using copy number information. There were 161 sex mismatches and 30 X-Y ratio outliers. Purple lines show 5SD outliers.



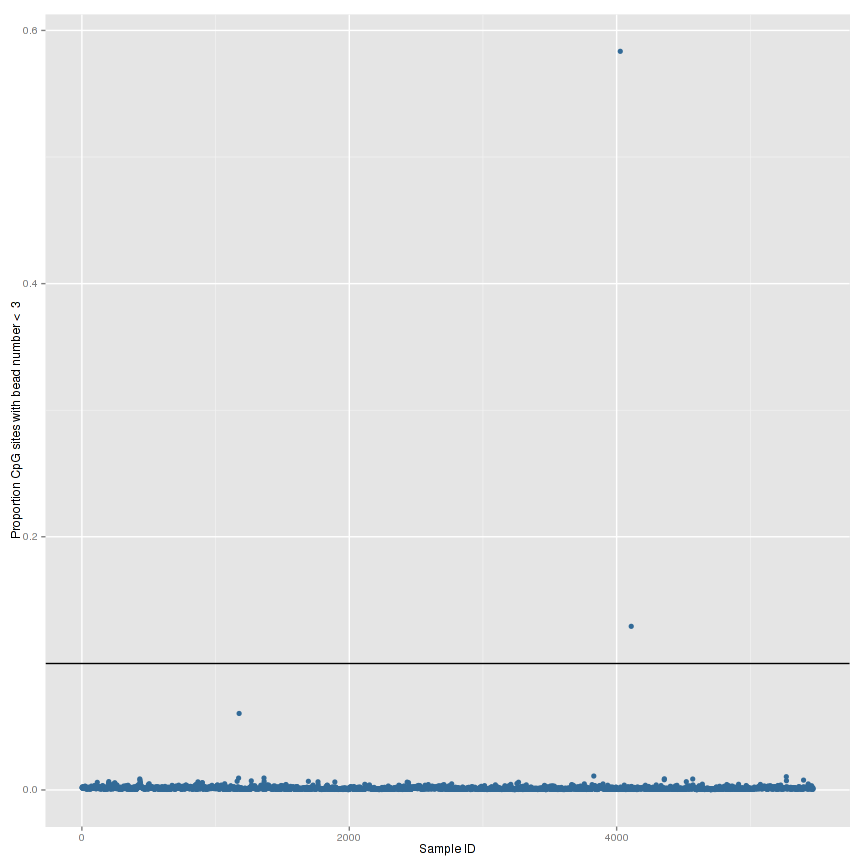
**Figure 3** QCplot of median intensity of the methylated signal against the median intensity of the unmethylated signal for all control probes. Colored samples are samples that are 5 SD from the mean and are removed.



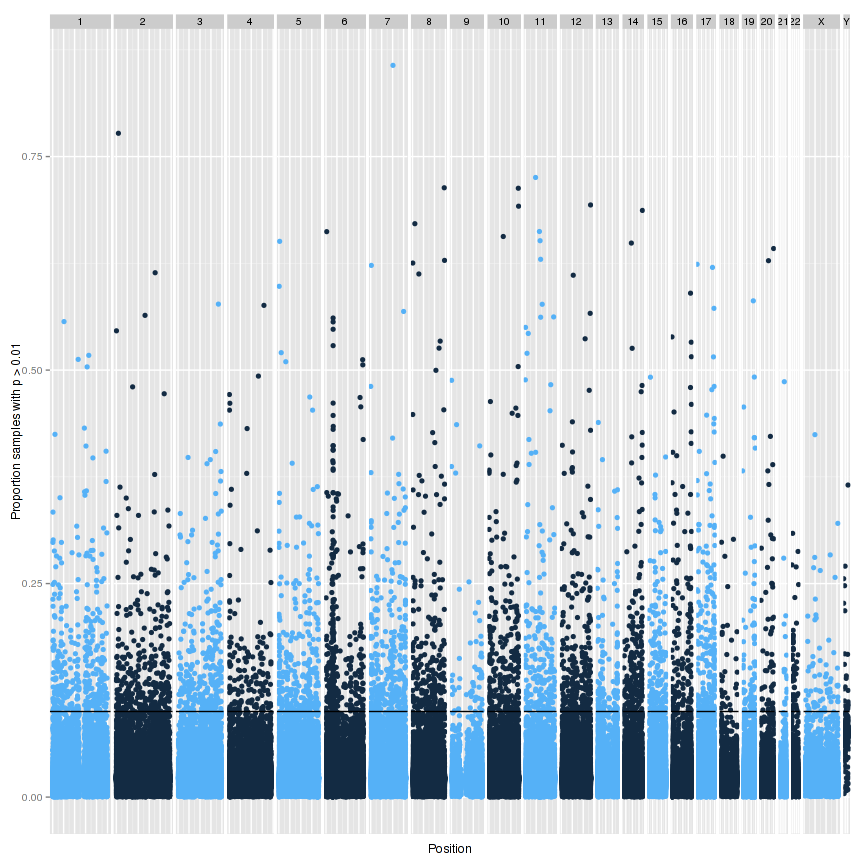
**Figure 4** Average intensities for individual control probe types ordered by BCD plate. Circles are samples that deviate 5SD of the mean. Bisulfite II are measured in red channel only and only for C-probes. Extension: Strong red signals for A and T and strong green signals for C and G. Hybe: Hybridization is monitored in the green channel only and measured at three different concentrations. The Target Removal Controls test the efficiency of stripping off DNA templates after the extension reaction and signal intensity is expected to be at background. The Non-Polymorphic Controls assess sample quality and the overall performance of the assay by querying non-polymorphic regions of the human genome. Note: Some samples have have intensities of zero for staining or nonpolymorphic probes. Some of these samples are checked by lIlumina and are fine.

****

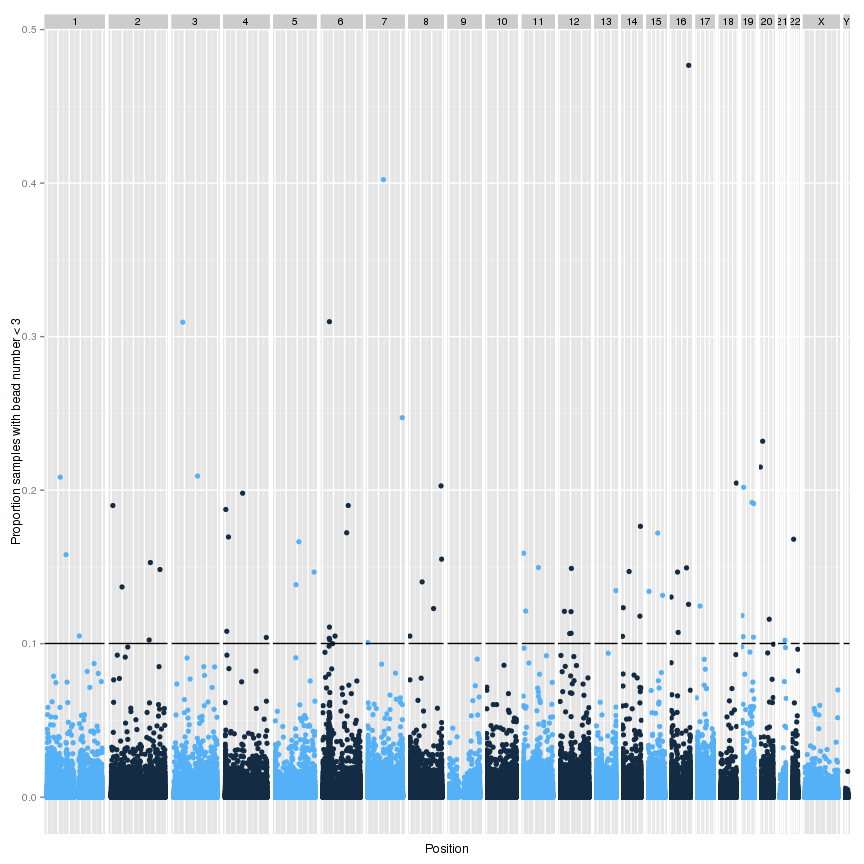
**Figure 5** Percentage of failed probes per sample based on detection pvalues. There were 166 samples with a high proportion of undetected probes (proportion of probes with detection p-value > 0.01 is > 0.1).

****

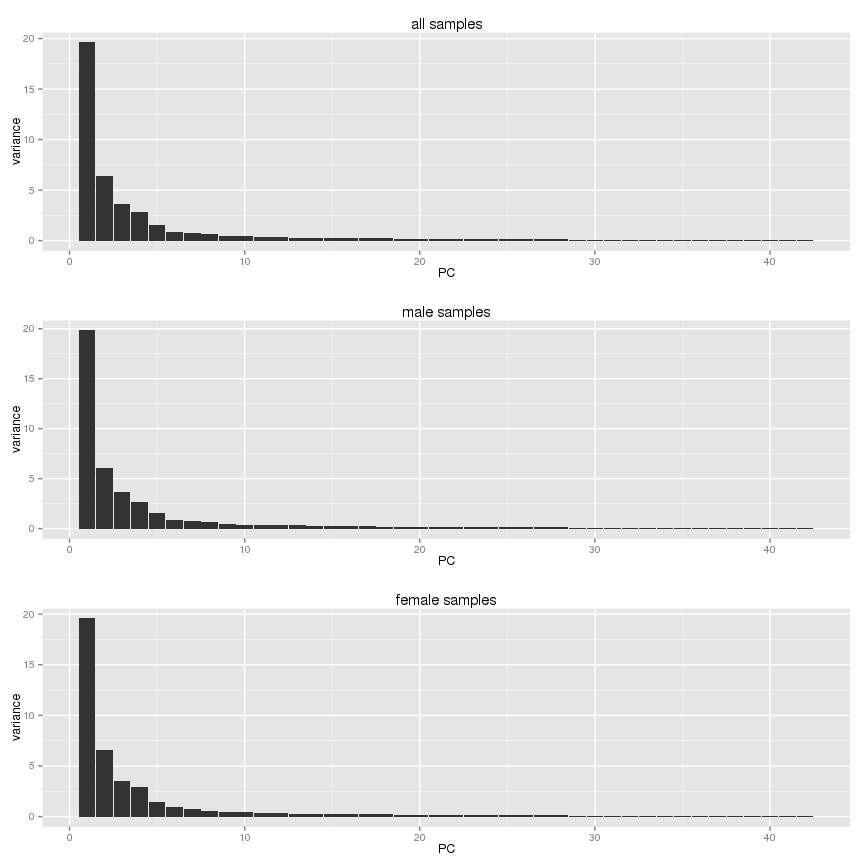
**Figure 6** Percentage of failed probes per sample based on bead number. There were 2 samples with a high proportion of probes with low bead number (proportion of probes with bead number < 3 is > 0.1).

****

**Figure 7.** Percentage of failed samples based on detection score per probe. For chrY, only males are included. There were 4182 probes with only background signal in a high proportion of samples (proportion of samples with detection p-value > 0.01 is > 0.1). Manhattan plot shows the proportion of samples.



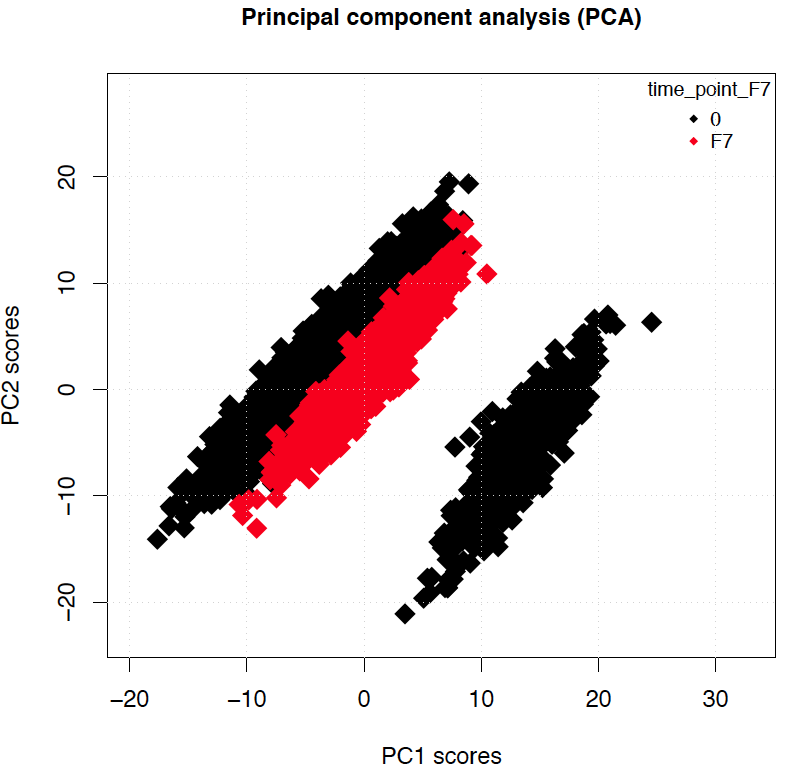
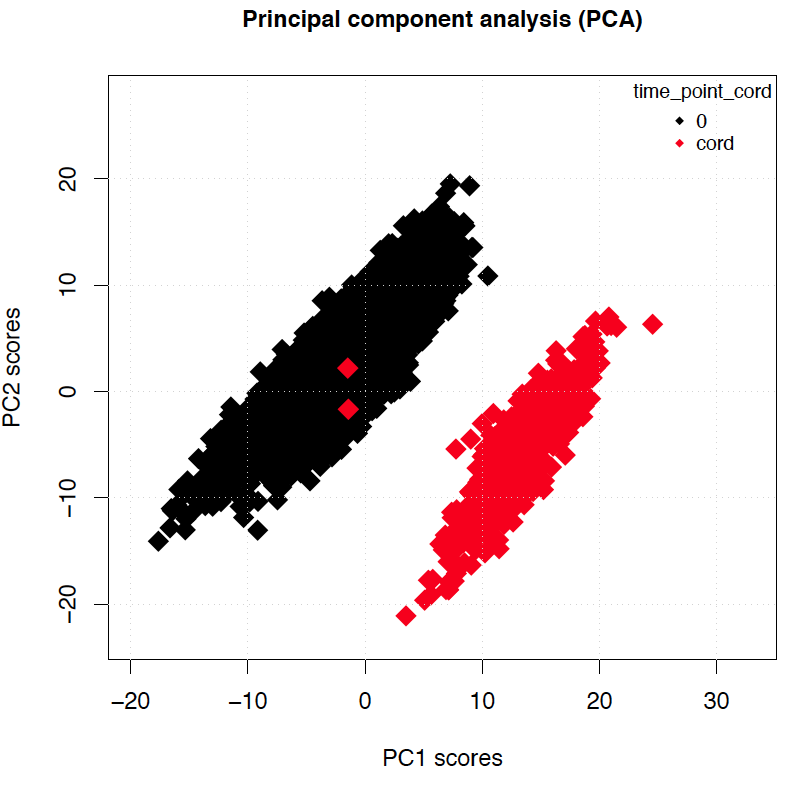
**Figure 8.** Percentage of failed samples based on bead number per probe. There were 70 CpGs with low bead numbers in a high proportion of samples (proportion of samples with bead number < 3 is > 0.1).

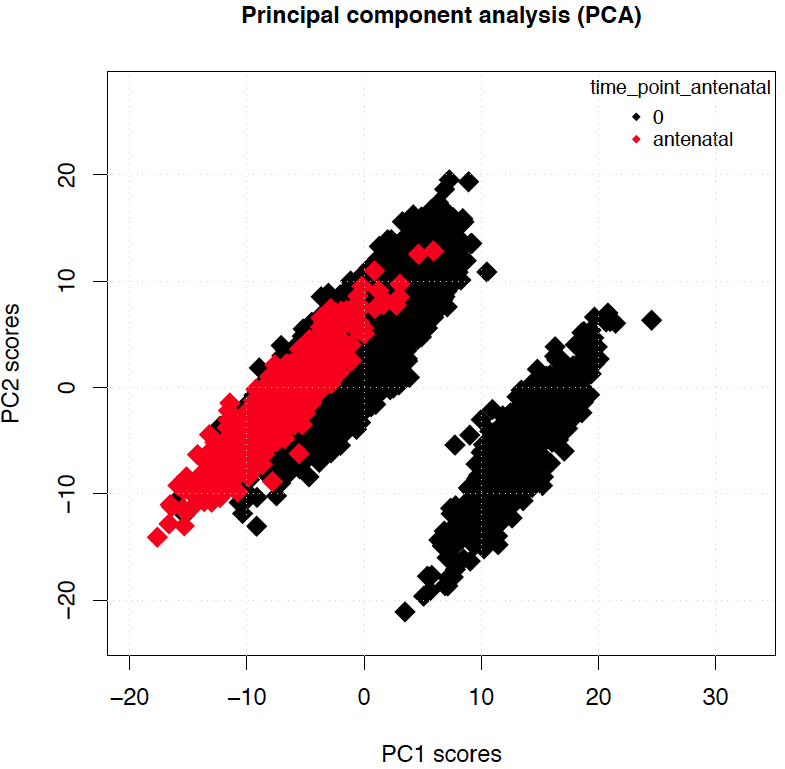
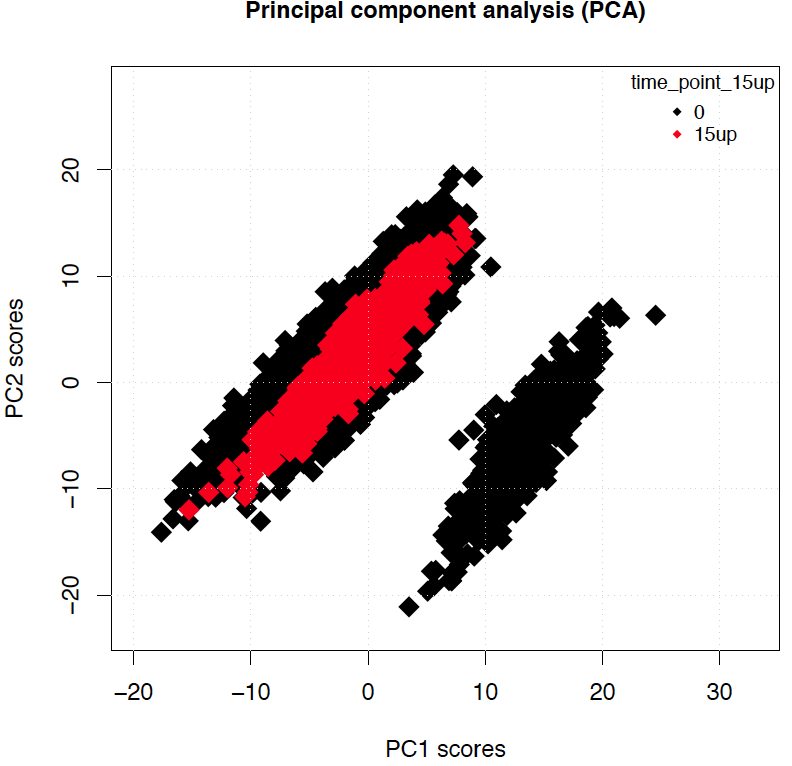
****

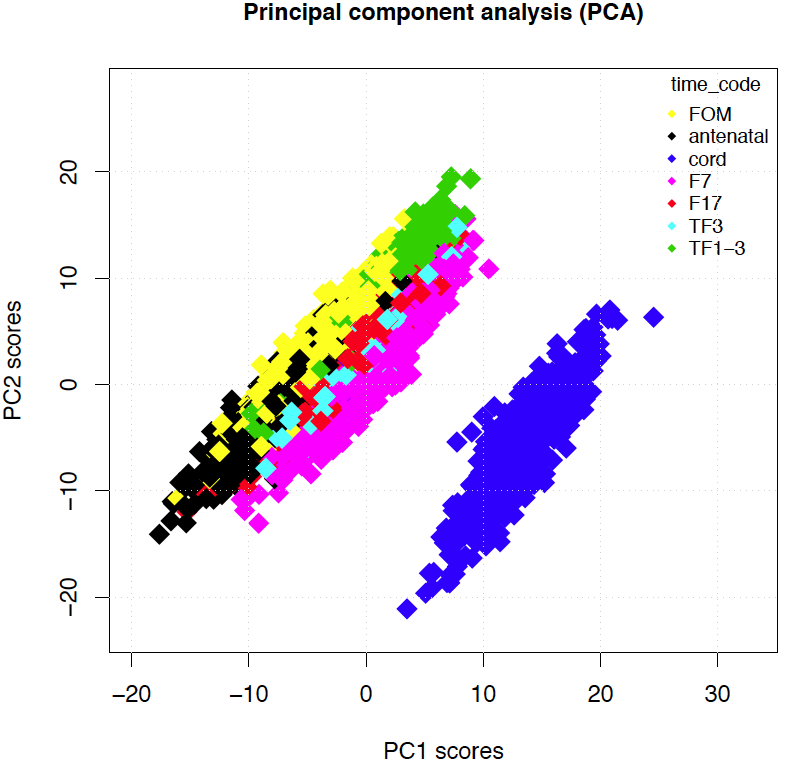
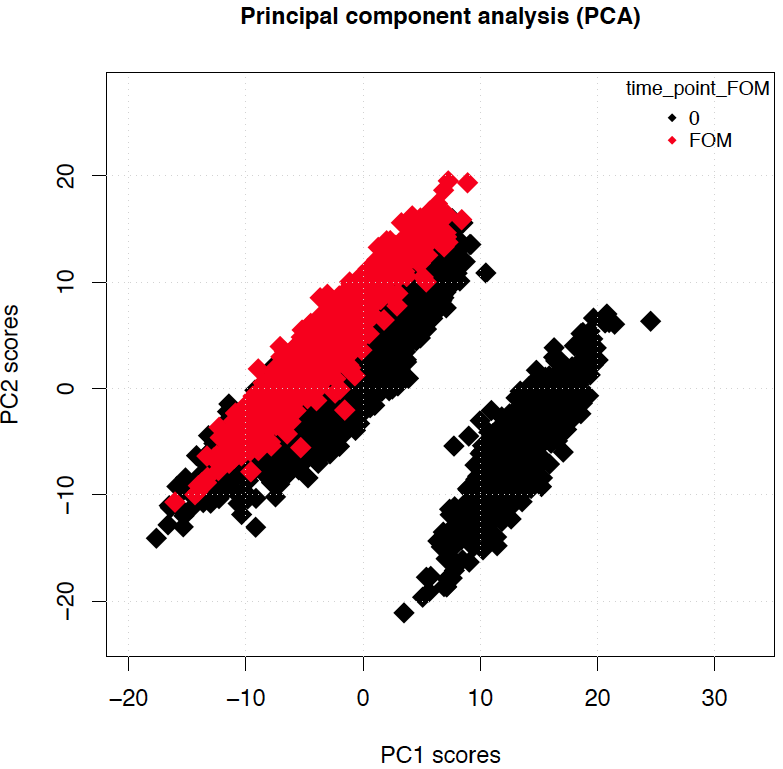
**A**

****

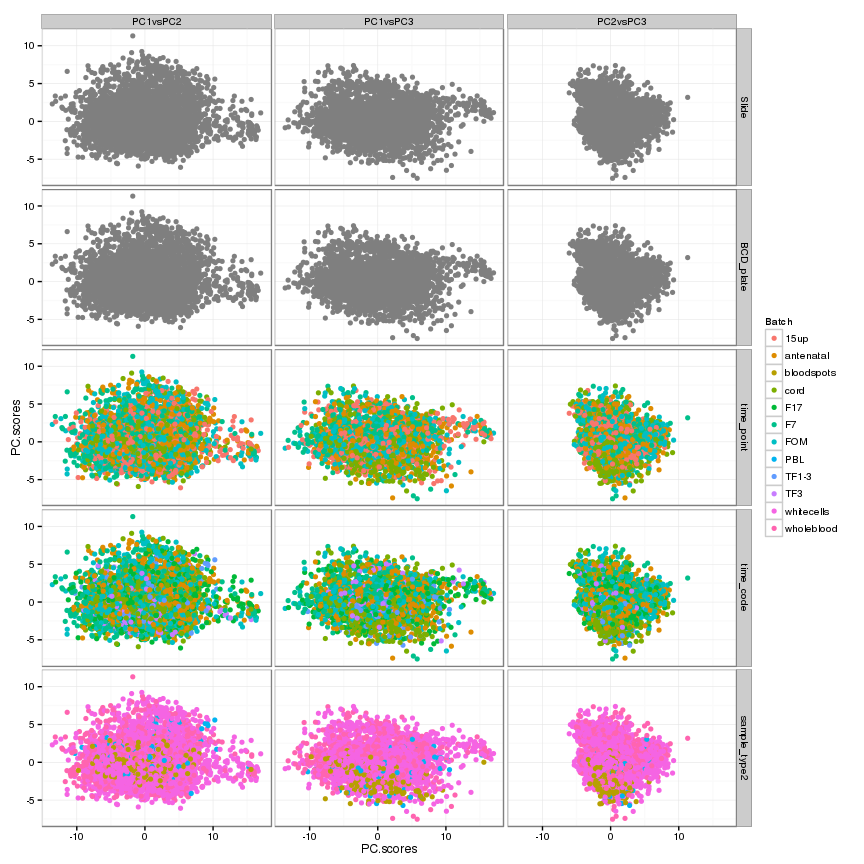
**Figure 9** A.Variance explained by PCs on control matrix. The plot shows clearly that you should use more than 2 PCs (recommended by minfi). Choosing the optimal number of PCs in your normalization. PC=10 seems to be a good cut-off.



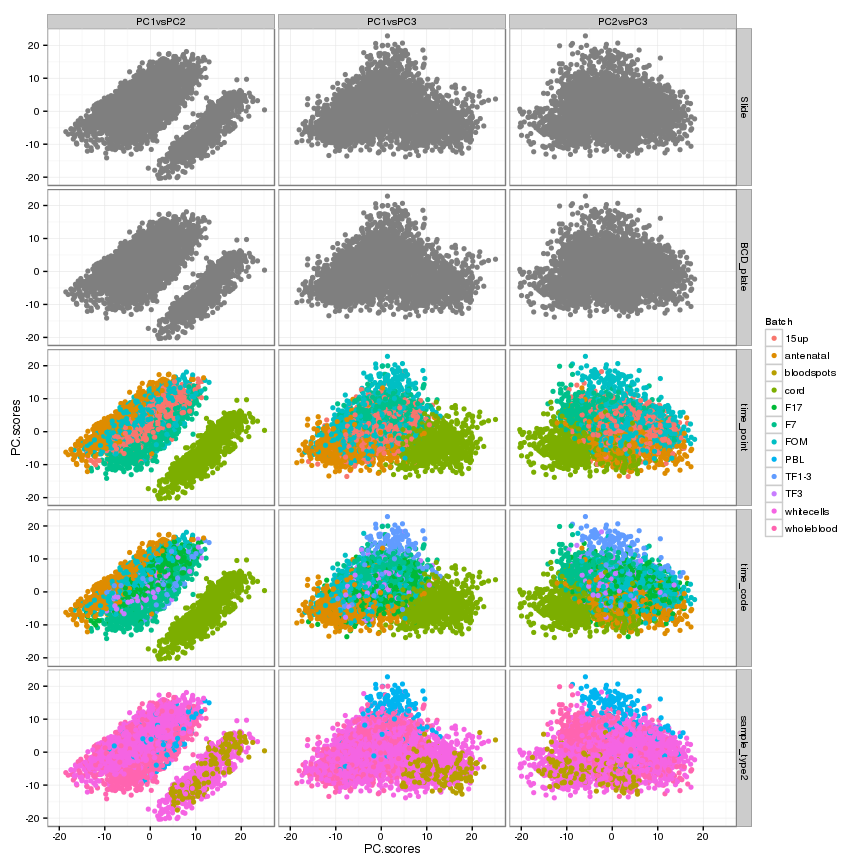




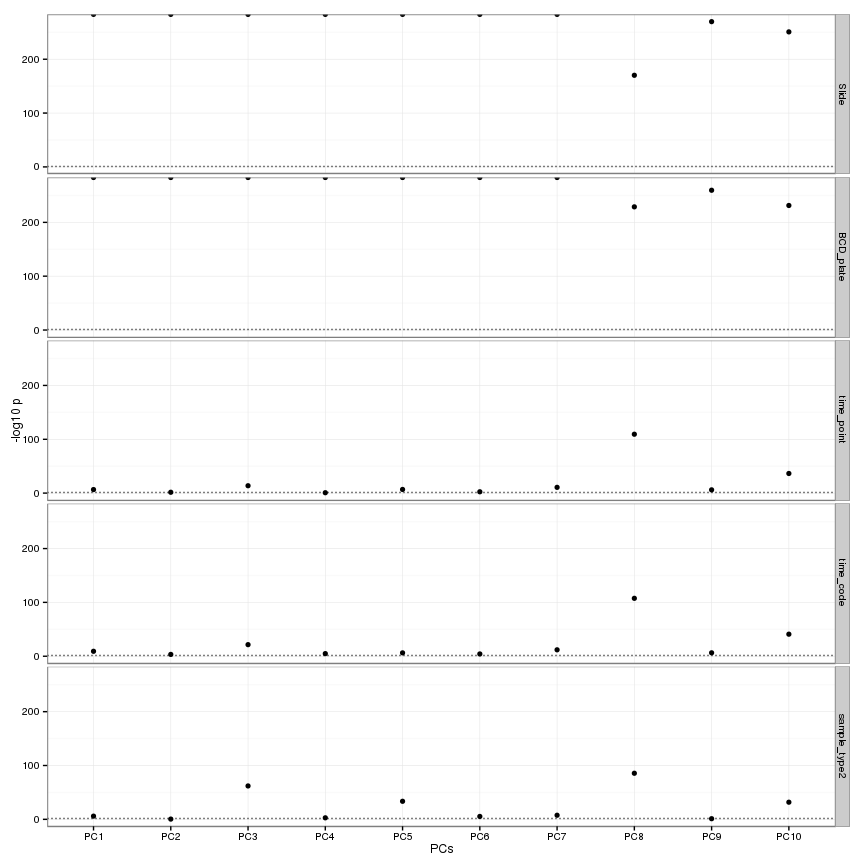
**Figure 10 PCA analysis on normalized data. Samples are colored by timepoint.** Two outliers were found.

****

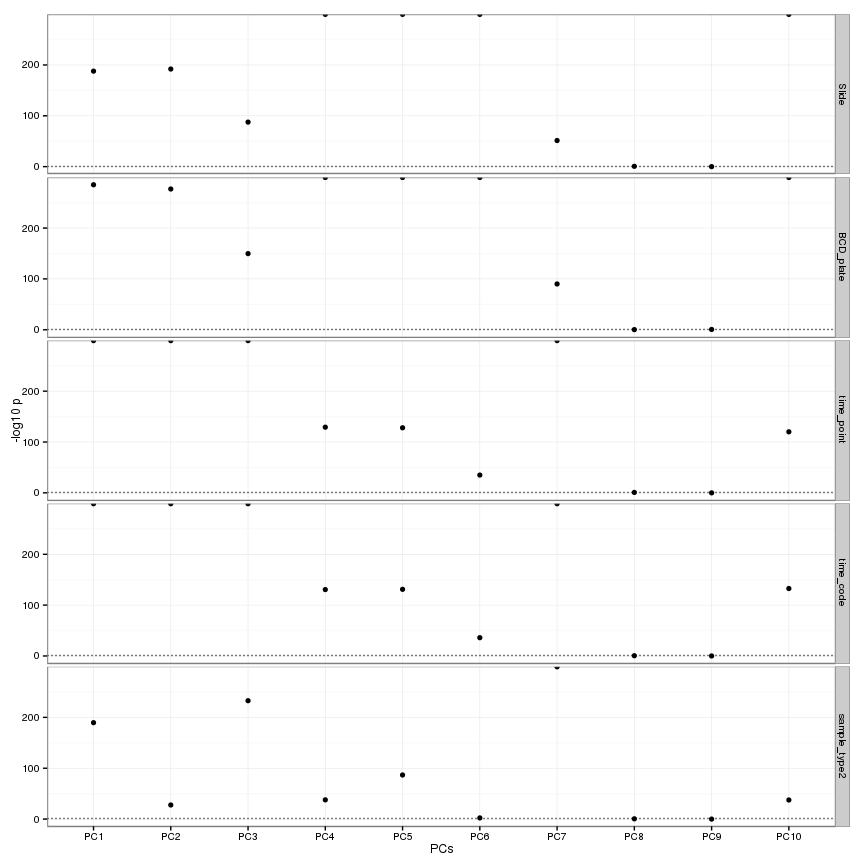
**Figure 11 PCA plot of controlmatrix: PC1-3**

****

**Figure 12 PCA plot of normalized betas: PC1-3**

****

**Figure 13:** Anova test pvalues for PC~batch effect in the control matrix. The PCs calculated from the control probes were associated with batch variables. Horizontal dotted line represents the \(-log\_{10}p = 0.05\).



**Figure 14** Anova test pvalues for PC~batch effect in the normalized data. The most variable normalized probes were extracted and decomposed into principal components, which were each regressed against each batch variable. If the normalization has performed well then there will be no associations between normalized probe PCs and batch variables. Horizontal dotted line represents the \(-log\_{10}p = 0.05\).



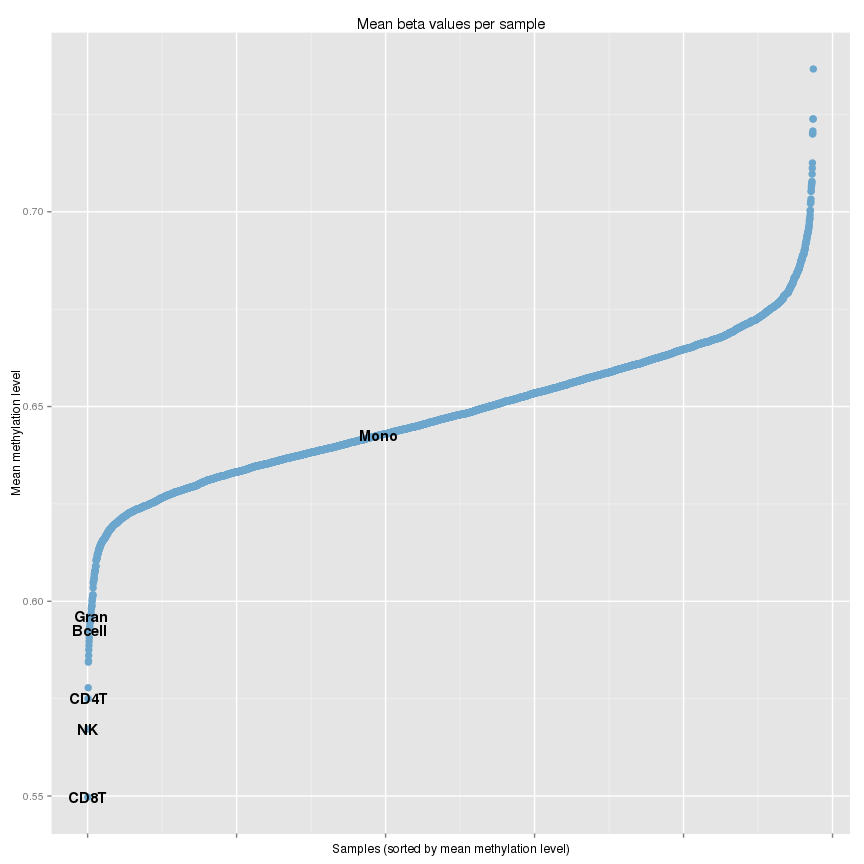


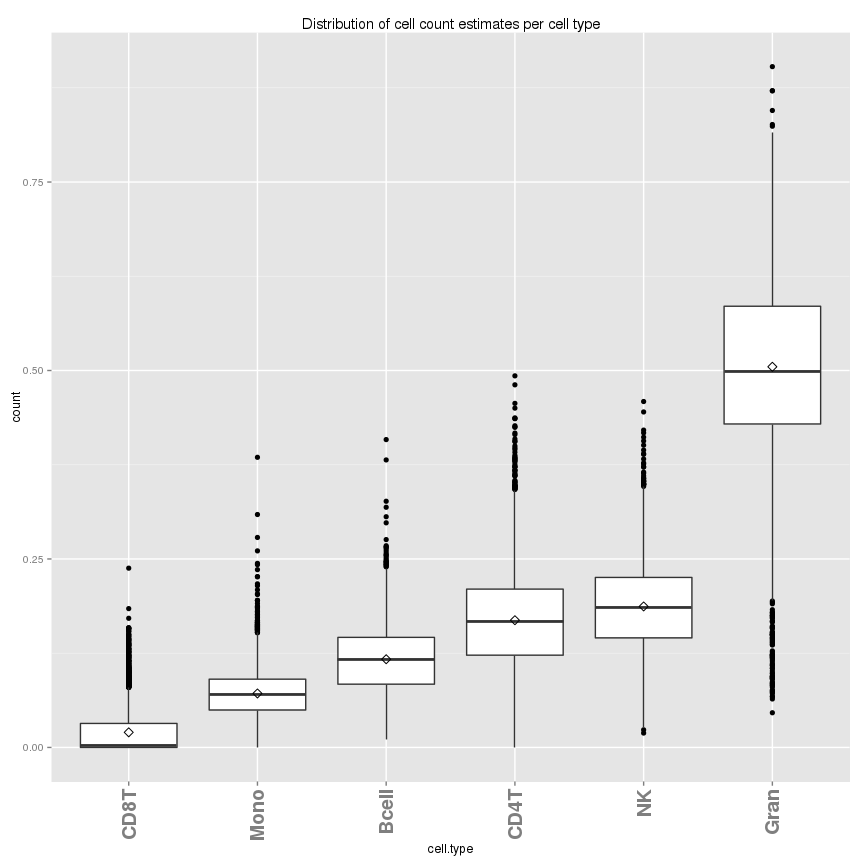






**Figure 15**: Coefficient plot. Effect sizes and their 5% confidence intervals of the association between PC1-10 and batch variable using a t-test. PC3 and PC5 show that slide 8963266138 is a clear outlier.

**A**

**B**

**Figure 16 A:** Cell counts were estimated using the blood gse35069 cell type methylation profile references. Plot compares methylation levels of CpG sites used to estimate cell counts for each sample and reference methylation profile. Methylation levels of samples should generally overlap with reference methylation levels otherwise estimation will have simply selected the cell type reference with the nearest mean methylation level. **B.** Boxplot shows the distributions of estimated cell counts for each reference cell type across all samples.

**Appendix** Control probes on the 450k array.

|  |  |  |  |
| --- | --- | --- | --- |
| **ControlType** | **No of probes (Red/Green)** | **Comment** | **Description** |
| bisulfite1 | 4/0 | average of red channel only | test efficiency of bisulphite conversion by query of C/T polymorphism |
| bisulfite2 | 3/3 | average of 6 probes (probes C1-C6); U-probes ignored | test efficiency of bisulphite conversion by query of C/T polymorphism |
| extGrn1 | 0/1 | C-probe | test efficiency of extension of A, T, C and G nucleotides from a hairpin probe; Strong positive signals are expected for Extension (C) and Extension (G) data points, while signals for Extension (A) and Extension (T) are expected to be at background levels. |
| extGrn2 | 0/1 | G-probe | test efficiency of extension of A, T, C and G nucleotides from a hairpin probe; Extension Green Strong positive signals are expected for Extension (C) and Extension (G) data points, while signals for Extension (A) and Extension (T) are expected to be at background levels. |
| extRed1 | 1/0 | A-probe | test efficiency of extension of A, T, C and G nucleotides from a hairpin probe; Extension Red Strong positive signals are expected for Extension (A) and Extension (T) data points, while signals for Extension (C) and Extension (G) are expected to be at background levels. |
| extRed2 | 1/0 | T-probe | test efficiency of extension of A, T, C and G nucleotides from a hairpin probe; Extension Red Strong positive signals are expected for Extension (A) and Extension (T) data points, while signals for Extension (C) and Extension (G) are expected to be at background levels. |
| hybe1 | 0/1 | High concentration | Hybridization should be monitored only in the green channel; measured at three different concentrations |
| hybe2 | 0/1 | Medium concentration | Hybridization should be monitored only in the green channel; measured at three different concentrations |
| hybe3 | 0/1 | Low concentration | Hybridization should be monitored only in the green channel; measured at three different concentrations |
| stain.green | 0/1 | 1 Biotin probe; blue biotin probe ignored | Staining Red Strong positive signals are expected for the dinitrophenyl (DNP) (High while background is expected for the DNP (Bgnd), Biotin (High), and Biotin (Bgnd) signals) |
| stain.red | 1/0 | 1 DNP probe; purple DNP probe ignored | Staining Green Strong positive signals are expected for the Biotin (High) data points,while background is expected for the DNP (High), DNP (Bgnd), and Biotin (Bgnd) data points. |
| nonpolyGrn1 | 0/1 | NP (A) | The Non-Polymorphic Controls assess sample quality and the overall performance of the assay by querying non-polymorphic regions of the human genome. Strong positive signals are expected for the green NP (C) and blue NP (G) data points, while the red NP (A) and purple NP (T) signals are expected to be at background levels. |
| nonpolyGrn2 | 0/1 | NP (T) | The Non-Polymorphic Controls assess sample quality and the overall performance of the assay by querying non-polymorphic regions of the human genome. Strong positive signals are expected for the green NP (C) and blue NP (G) data points, while the red NP (A) and purple NP (T) signals are expected to be at background levels. |
| nonpolyRed1 | 1/0 | NP (C) | The Non-Polymorphic Controls assess sample quality and the overall performance of the assay by querying non-polymorphic regions of the human genome. Strong positive signals are expected for the red NP (A) and purple NP (T) data points, while the green NP (C) and blue NP (G) signals are expected to be at background levels.. |
| nonpolyRed2 | 1/0 | NP (G) | The Non-Polymorphic Controls assess sample quality and the overall performance of the assay by querying non-polymorphic regions of the human genome. Strong positive signals are expected for the red NP (A) and purple NP (T) data points, while the green NP (C) and blue NP (G) signals are expected to be at background levels.. |
| targetrem1 | 0/1 |  | The Target Removal Controls test the efficiency of stripping off DNA templates after the extension reaction. Target Removal Red Target removal controls are monitored in the green channel, thus, signal intensity in the red channel is expected to be at background |
| targetrem2 | 0/1 |  | The Target Removal Controls test the efficiency of stripping off DNA templates after the extension reaction. Target Removal Red Target removal controls are monitored in the green channel, thus, signal intensity in the red channel is expected to be at background levels. |
| spec1Grn1 | 0/1 | GT Mismatch 1 (PM) | check for non-specific detection of methylation signal over unmethylated background. Specificity controls are designed against non-polymorphic T sites (G/T mismatch) |
| spec1Grn2 | 0/1 | GT Mismatch 2 (PM) | check for non-specific detection of methylation signal over unmethylated background. Specificity controls are designed against non-polymorphic T sites (G/T mismatch) |
| spec1Grn3 | 0/1 | GT Mismatch 3 (PM) | check for non-specific detection of methylation signal over unmethylated background. Specificity controls are designed against non-polymorphic T sites (G/T mismatch) |
| spec1Red1 | 1/0 | GT Mismatch 4 (PM) | check for non-specific detection of methylation signal over unmethylated background. Specificity controls are designed against non-polymorphic T sites (G/T mismatch) |
| spec1Red2 | 1/0 | GT Mismatch 5 (PM) | check for non-specific detection of methylation signal over unmethylated background. Specificity controls are designed against non-polymorphic T sites (G/T mismatch) |
| spec1Red3 | 1/0 | GT Mismatch 6 (PM) | check for non-specific detection of methylation signal over unmethylated background. Specificity controls are designed against non-polymorphic T sites (G/T mismatch) |
| spec2Grn1 | 0/1 |  | check for non-specific detection of methylation signal over unmethylated background. Specificity controls are designed against non-polymorphic T sites (G/T mismatch) |
| spec2Grn2 | 0/1 |  | check for non-specific detection of methylation signal over unmethylated background. Specificity controls are designed against non-polymorphic T sites (G/T mismatch) |
| spec2Grn3 | 0/1 |  | check for non-specific detection of methylation signal over unmethylated background. Specificity controls are designed against non-polymorphic T sites (G/T mismatch) |
| spec2Red1 | 1/0 |  | check for non-specific detection of methylation signal over unmethylated background. Specificity controls are designed against non-polymorphic T sites (G/T mismatch) |
| spec2Red2 | 1/0 |  | check for non-specific detection of methylation signal over unmethylated background. Specificity controls are designed against non-polymorphic T sites (G/T mismatch) |
| spec2Red3 | 1/0 |  | check for non-specific detection of methylation signal over unmethylated background. Specificity controls are designed against non-polymorphic T sites (G/T mismatch) |
| spec1.ratio1 | 3/3 | avg(Red)/avg(Green) for GT Mismatch 1,2,3 |  |
| spec1.ratio | 3/3 | (spec1.ratio1 + spec1.ratio2) / 2 |  |
| spec2.ratio | 3/3 | avg(Green)/avg(Red) for 3 probes |  |
| spec1.ratio2 | 3/3 | avg(Green)/avg(Red) for GT Mismatch 4,5,6 |  |
| normA | 32/0 |  | Strong positive signals are expected for the red NP (A) and purple NP (T) data points, while the green NP (C) and blue NP (G) signals are expected to be at background levels. |
| normC | 0/61 |  | Strong positive signals are expected for the green NP (C) and blue NP (G) data points, while the red NP (A) and purple NP (T) signals are expected to be at background levels. |
| normT | 61/0 |  | Strong positive signals are expected for the red NP (A) and purple NP (T) data points, while the green NP (C) and blue NP (G) signals are expected to be at background levels. |
| normG | 0/32 |  | Strong positive signals are expected for the green NP (C) and blue NP (G) data points, while the red NP (A) and purple NP (T) signals are expected to be at background levels. |
| dyebias | 93/93 | (avg(normC) +avg( normG))/(avg(normA) +avg(normT)) |  |
| oob1 | 0/1 | Green: 178374; Red: 92578 (colQuant, prob=1%) | Retrieves the so-called “out-of-band” (OOB) probes. These are the measurements of Type I probes in the “wrong” color channel. Return value is a list with two matrices, named ‘Red’ and ‘Grn’ |
| oob50 | 0/1 | Green: 178374; Red: 92578 (colQuant, prob=50%) | Retrieves the so-called “out-of-band” (OOB) probes. These are the measurements of Type I probes in the “wrong” color channel. Return value is a list with two matrices, named ‘Red’ and ‘Grn’ |
| oob99 | 0/1 | Green: 178374; Red: 92578 (colQuant, prob=99%) | Retrieves the so-called “out-of-band” (OOB) probes. These are the measurements of Type I probes in the “wrong” color channel. Return value is a list with two matrices, named ‘Red’ and ‘Grn’ |
| oob.ratio | 3/3 |  |  |