

TECHNICAL NOTE

Open Access

Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc.

Jessica Roessler¹, Ole Ammerpohl², Jana Gutwein², Britta Hasemeier¹, Sumadi Lukman Anwar¹, Hans Kreipe¹ and Ulrich Lehmann^{1*}

Abstract

Background: The newly released 450k DNA methylation array from Illumina, Inc. offers the possibility to analyze more than 480,000 individual CpG sites in a user friendly standardized format. In this study the relationship between the β -values provided by the Illumina, Inc. array for each individual CpG dinucleotide and the quantitative methylation levels obtained by pyrosequencing were analyzed. In addition, the representation of microRNA genes and imprinted loci on the Illumina, Inc. array was assessed in detail. Genomic DNA from 4 human breast cancer cell lines (IPH-926, HCC1937, MDA-MB-134, PMC42) and 18 human breast cancer specimens as well as 4 normal mammary epithelial fractions was analyzed on 450k DNA methylation arrays. The β -values for 692 individual CpG sites from 62 different genes were cross-validated using conventional quantitative pyrosequencing.

Findings: The newly released 450k methylation array from Illumina, Inc. shows a high concordance with quantitative pyrosequencing if identical CpG sites are analyzed in cell lines (Spearman $r = 0.88$, $p < 0.0001$), which is somewhat reduced in primary tumor specimens (Spearman $r = 0.86$, $p < 0.0001$). 80.7% of the CpG sites show an absolute difference in methylation level of less than 15 percentage points. If different CpG sites in the same CpG islands are targeted the concordance is lower ($r = 0.83$ in cell lines and $r = 0.7$ in primary tumors). The number of CpG sites representing microRNA genes and imprinted loci is very heterogeneous (range: 1 – 70 CpG sites for microRNAs and 1 – 288 for imprinted loci).

Conclusions: The newly released 450k methylation array from Illumina, Inc. provides a genome-wide quantitative representation of DNA methylation aberrations in a convenient format. Overall, the congruence with pyrosequencing data is very good. However, for individual loci one should be careful to translate the β -values directly into percent methylation levels.

Background

The genome-wide assessment of DNA methylation patterns becomes more and more important in cancer research [1,2]. However, so far the comprehensive analysis of all potential CpG sites in the human genome, as demonstrated for the first time in 2009 by Lister et al. [3] is - in terms of costs per sample and required data processing resources - well beyond the options of most research groups. Therefore, various protocols have been developed to analyze a representative subset of the human genome [4,5]. The most

recent addition to this arsenal of methods is the 450k DNA methylation array from Illumina, Inc, which measures in parallel the methylation of approximately 480,000 CpG sites across the human genome. It was attempted to cover all known protein coding genes as well as a fairly large number of non-coding RNA genes and imprinted loci. Based on the well-established and widely used Infinium technology [6] this array platform promises an easy-to-use, standardized, and cost effective format for the analysis of a representative subset of the human methylome.

The array design, content, and technical performance are well described in two recent publications [7,8]. However, both publications are not independent from the manufacturer and so far only limited data are available

* Correspondence: Lehmann.Ulrich@MH-Hannover.de

¹Institute of Pathology, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1D-30625, Hannover, Germany

Full list of author information is available at the end of the article

about cross-validation of the 450k methylation array with other quantitative methods well established in the field of DNA methylation research. Dedeurwaerder et al. [9] compared the methylation level of altogether 15 CpG sites in 2 cell line samples (30 measurements) and 4 CpG sites in 6 primary samples (6 measurements) with quantitative pyrosequencing. An analysis of the representation of genes important in tumor biology (measured as number of CpG sites per gene) on the 450k array has not yet been published so far. Therefore, we compared the methylation level of altogether 692 individual CpG sites from 62 different genes in a series of cell lines and primary human tissue samples and analyzed in detail how microRNA genes and imprinted loci are represented on the array.

Findings

It is expected that the 450k DNA methylation array from Illumina, Inc. will be as widely used as its predecessor, the 27k methylation array, and will thereby influence the methylation field to a great extent. Therefore, careful and critical analysis of the strengths and weaknesses of this methodology as early as possible is of great importance.

Cross-validation of 450k methylation array and pyrosequencing

For the cross-validation of the 450k array we selected quantitative pyrosequencing which is a well-established and widely used method in the field of DNA methylation research. Several groups, including our own [10,11], could show that for many loci pyrosequencing provides a very good quantitative measure of the methylation level at individual CpG sites (see also Additional file 1). The comparison of β -values and pyrosequencing results was performed separately for cell lines and primary human tissue samples and included altogether 692 individual CpG dinucleotides (340 in cell lines and 352 in primary specimens). From the 352 individual CpG sites compared one-by-one in primary tumor specimens 80.3% show a difference in methylation level of less than 15 percentage points. If 10 percentage points are chosen as threshold, 60.5% of all data points are within this range of agreement. The corresponding numbers for the analyses in cell lines are very similar: 77.4% and 63.8%, respectively.

Figure 1 A) and B) show a very good concordance between both methods (cell lines: Spearman $r = 0.88$, primary tissue Spearman $r = 0.86$). However, the calculation of a correlation coefficient or a regression coefficient can be misleading, because the scatter plots are dominated by two agglomerations of data points near the origin of the scatter plot and near the 100% value. In between, the accordance between both methods is reduced. This can be seen more clearly in the corresponding Bland-Altman-Plots (Figure 1 C) and D)). In a Bland-Altman-Plot

the difference between the two methods under study is plotted against the mean of both methods for every individual pair of measurements. The interval of the mean of the difference \pm two times the standard deviation defines the 95% interval of the limits of agreement. In this kind of plot it is much more obvious that the data points become more dispersed in the range of 25 – 75% methylation (indicating less concordance). All data points outside the 95% limits of agreement (marked by the dotted lines) are lying in this middle region. If the CpG sites targeted by the pyrosequencing assay and the 450k array are not identical but from the same CpG islands, the agreement is - as expected - reduced (Spearman $r = 0.83$ for cell lines and $r = 0.71$ for primary specimens, see Additional file 2).

Common SNPs could be excluded as a major source of the discrepancies between pyrosequencing and 450k methylation array. All data points above or below the limits of agreement in Figure 1 C) and D) were checked for the presence of SNPs in the corresponding probes on the 450k array and also in the pyrosequencing assays. As a control 20 data points with very high concordance were randomly selected and checked for the potential presence of SNPs. No clear correlation between the level of agreement and the potential presence of SNPs could be identified.

If a threshold of 10 to 20% methylation is chosen for scoring CpG sites as “hypermethylated”, this reduced accordance in the middle range of methylation levels (25 – 75%) is not a problem at all. However, if cluster analyses are performed with the β -values for subsets of probes, which show intermediate methylation levels, results should be interpreted carefully.

If only cell line DNA samples are used for the evaluation and comparison of methods for the analysis of DNA methylation a certain bias is introduced because the vast majority of loci display a clear dichotomous distribution in cell lines: either no or very low level DNA methylation or very high, nearly complete methylation. Therefore, any scatter plot (and subsequent calculations of correlation or regression coefficients) illustrating the comparison of methods or samples is dominated by two populations of data points (i.e., close to 0 and close to 1 or 100%). This can easily be verified by visual inspection of Fig. 6 A) and B) in Sandoval et al. [8], Fig. 4 in Bibikova et al., [7], Fig. 4 A) in Dedeurwaerder et al. [9], and our own Figure 1.

The first publication about the 450k array from Sandoval et al. [8] in collaboration with Illumina, Inc. gives a very good overview of the loci covered and the classes of CpG sites included in the 450k array design. However, it did not cross-validate the results directly with another method but only indirectly by comparing the 450k array with the well-established GoldenGate[®] array and the 27k array from Illumina, Inc. In the second

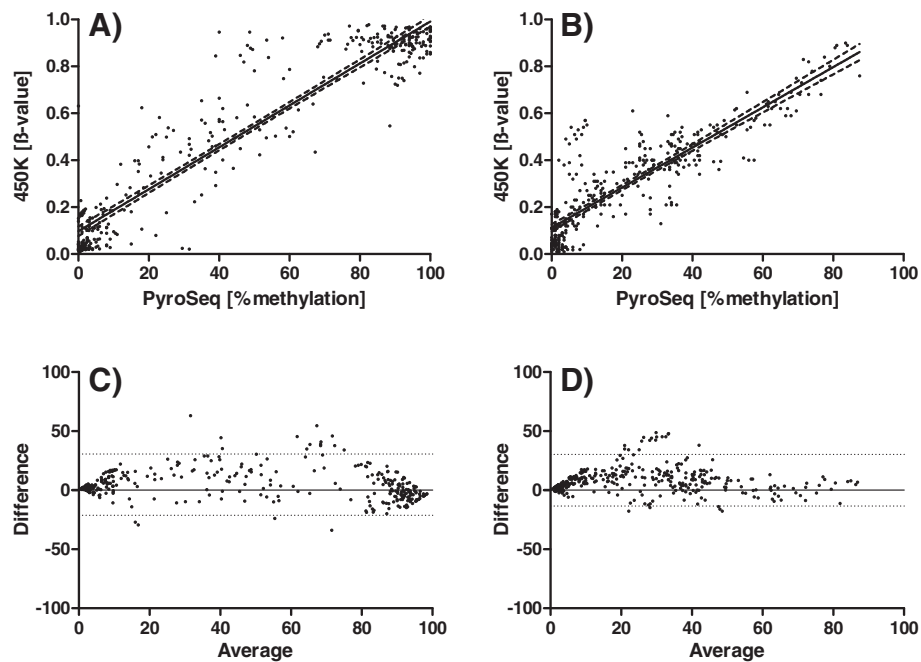


Figure 1 Cross-validation of 450k methylation array and quantitative pyrosequencing. Shown are scatter plots for the methylation values obtained by pyrosequencing (x-axis) and the β -values obtained from the 450k array for identical CpG sites. A) cell lines (Spearman $r=0.88$, linear regression $r^2=0.89$, slope: 0.89, 95% confidence interval: 0.86 – 0.93), B) primary human tumor tissues (Spearman $r=0.85$, linear regression $r^2=0.77$, slope: 0.86, 95% confidence interval: 0.81 – 0.91). Each dot in the figure represents the comparison of % methylation according to pyrosequencing versus the β -value for a single CpG site in one sample. C) and D) corresponding Bland-Altman plots. The difference between both methods for every individual measurement is plotted against the mean of both methods. The mean of the differences \pm two times the standard deviation denotes the 95% range for the limits of agreement (marked by the horizontal dotted lines). For the construction of these plots the β -values have been multiplied by the factor 100 in order to obtain data sets of the same size range. The coordinates of the individual CpG sites analyzed in this comparison are compiled in Additional file 3.

publication about the 450k array exclusively from Illumina, Inc. [7] a cross-validation with deep-sequencing for two samples is reported. Due to the high number of data points (189,000 and 167,000, respectively) the correlation seems to be extremely good ($r^2=0.96$). But a more careful inspection of Figure 5B from this publication shows that the data points form a “broad band”, which is 30 – 40 percentage points wide, meaning that many measurements differ by 30 – 40 percentage points. There are an unidentifiable number of measurements with no methylation in the sequencing analysis and up to 50% methylation level according to the array analysis. This is indicated by the black line at the bottom of Figure 5 B) parallel to the x-axis at 0.0 methylation value according to sequencing (y-axis) and ranging from a β -value of 0.0 to a β -value of approx. 0.5 at the x-axis. The same phenomenon can be seen at the upper limit of the scale at the top of Figure 5 B): A black line parallel to the x-axis at 1.0 methylation according to sequencing and ranging from a β -value of approx. 0.4 to a β -value of 1.0 indicating 100% methylation according to sequencing and 40 – 100% according to array analysis. This clearly shows in line with our results that whereas overall the

concordance is very high the derivation of methylation levels of individual genes or loci from β -values might be uncertain and requires independent validation.

Comparison of Infinium I and Infinium II assays

A potential problem of the 450k methylation array is the fact that it combines two different assays on a single array, namely the Infinium I and Infinium II technology (see Figure 1 in ref. [7] or Figure 2 in ref. [9] for a very good illustration of the principal of the two different assays). Dedeurwaerder et al. [9] describe in detail the effect of these two different assays on the β -value distribution (Fig. 1 in ref. [9]) observed now by many users in the DNA methylation community. Therefore, we analyzed all correlations between β -values and pyrosequencing results separately for Infinium I and II assays. Overall, the agreement between pyrosequencing and Infinium I and II, respectively, are very similar for both cell lines and primary patient samples (see Additional file 4). The above mentioned observation, that the concordance of both methods in the range of 25 – 75% methylation is reduced, affects both assay types to a similar extent. Our re-analysis of the data presented by Dedeurwaerder et al. [9] shows that the

peak-correction proposed by these authors does not improve the congruence (see Additional file 5). These authors performed also a much more limited comparison of pyrosequencing and the 450k methylation array, especially for primary human tissue samples: The results of only six measurements are presented (Fig 5 B in Dedeurwaerder et al. [9]), in comparison to 352 measurements in primary tissues samples evaluated by us.

The reduced concordance in primary patient samples might be due to the heterogeneous methylation patterns frequently encountered in primary samples compared with cell lines. As described in detail by Bibikova et al. [7] in the introduction section of their publication the basic assumption for the Infinium assay design is, that adjacent CpG sites display very similar methylation levels, thereby enabling the selection of closely spaced probes including potential methylation sites in their binding site. However, the two studies cited by Bibikova et al. as supporting this assumption [12,13] analyzed healthy normal tissue samples or healthy primary cell samples but not tumor tissue samples. Since many genes and loci show extensive heterogeneity in methylation patterns in tumor cells (e.g., *p16^{INK4A}* in HCC [11]), the methylation level of individual CpG sites might be assessed incorrectly.

Representation of microRNA genes and imprinted loci

Since we have a long standing interest in epigenetic regulation of microRNA genes [14] and imprinted loci [15] the representation of these two important classes of non-protein coding genes with pleiotropic regulatory functions was assessed in detail for several microRNA genes already under investigation in our group [11,14]. The number of CpG sites analyzed is in general comparable to the number of CpG sites included in pyrosequencing assays (i.e., 4 – 13 CpG

sites). However, due to the assay design the CpG sites covered are spread over a much larger region: 900-1500 bp, in comparison to 50 – 100 bp for an average pyrosequencing assay. Whether this is an advantage or disadvantage depends on the circumstances. The often very heterogeneous methylation patterns encountered especially in primary tumor specimens might be better recognized by the analysis of a continuous stretch of CpG sites as with pyrosequencing or conventional bisulfite sequencing. On the other hand, the spreading of CpG sites analyzed on the 450k array might give a better representation of the methylation status of a whole genomic locus.

A few microRNA loci are very well represented on the 450k array (e.g., *hsa-mir-1256* with 34 CpG sites and *hsa-mir-548H4* with 70 CpG sites) despite the fact that so far only very limited data about epigenetic regulation or the cellular function of these microRNAs are available (status: January 31st, 2012). Notwithstanding sparse published functional evidences the *miR-548* family is represented by altogether 252 CpG sites. However, this might be advantageous for future studies focusing on these microRNAs.

The representation of imprinted loci is on average much better. Up to 180 CpG sites cover a single differentially methylated region. The analysis of the following 11 loci revealed a very good representation of these imprinted loci (in terms of number and location of CpG sites assayed on the 450k array): *IGF2*, *IGF2R*, *SNRPN*, *CDKN1C*, *MEG3/DLK1*, *GNAS*, *PEG3*, *PLAGL1(ZAC)*, *PEG10*, *MEST(PEG1)*, *GRB10*. Figure 2 shows as an example how many CpG sites within the four imprint control regions of the *SNRPN* locus on chr. 15q11-13 are measured using the 450k array.

The concordance between the β -values, our pyrosequencing results, and the pyrosequencing values reported by Woodfine et al. [17] for four imprinted loci is shown in

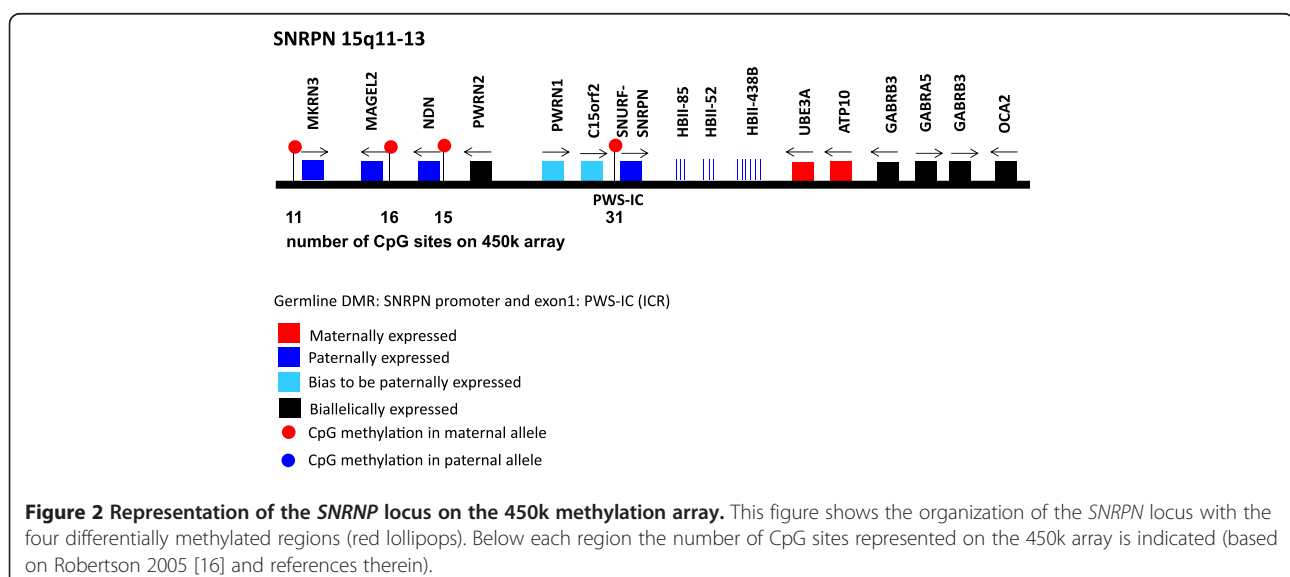


Figure 3. This graph represents altogether 36 measurements (9 different CpG sites in four samples).

Despite the above discussed reduced agreement between both methods in the middle range of methylation levels for some loci the results in Figure 3 demonstrate a very good concordance for four imprinted loci in normal primary mammary tissue samples between the pyrosequencing data from two independent laboratories and 450k DNA methylation array data.

A potential pitfall of the analysis of 450k array data is the annotation of probes which is not really consistent and straightforward in the small set of microRNA loci which we investigated in detail. A purely gene name-based search will miss many microRNA gene associated CpG sites due to inconsistent nomenclature. The assignment of a given CpG dinucleotide to a microRNA is still very difficult or impossible due to the lack of knowledge about microRNA gene promoters and the primary microRNA transcripts and due to the fact that many microRNAs are located within protein coding genes. Nevertheless, several inconsistencies in nomenclature could be eliminated in the next version of the probe identification sheet.

It should also be mentioned that a major limitation of all bisulfite treatment based methodologies is the inability to differentiate between methylcytosine and hydroxymethylcytosine [18-20]. The latter one is playing a crucial role at least in developmental biology.

Conclusion

The newly released 450k methylation array from Illumina, Inc. provides a genome-wide quantitative representation of DNA methylation aberrations in a convenient format. For the majority of CpG sites the β -values represent a very good measure of the methylation status. However, for individual loci the direct transformation of β -values into methylation levels should be handled with care and validated by an independent method. The representation and annotation of functionally important loci could be improved in future versions of the array.

Methods

Tissue specimens and bisulfite modification of DNA

All primary human tissue samples were retrieved from the archive of the Institute of Pathology, Hanover Medical School (Germany) and analyzed anonymously following the guidelines of the local Ethics committee ("Ethik-Kommission der Medizinischen Hochschule Hannover", head: Prof. Dr. Tröger). Tumor cell content was determined to be greater than 70%. DNA was isolated by digestion with Proteinase K (Merck, Darmstadt, Germany) followed by phenol/chloroform extraction from a total of 22 specimens (Additional file 6). Genomic DNA (1 μ g) from tumor specimens was treated with sodium bisulfite using the EZ DNA Methylation™ kit (ZymoResearch, Irvine, CA, USA) following the protocol supplied by the manufacturer with the exception of eluting the treated DNA with distilled water instead of using the provided elution buffer.

Cell lines HCC1937, MDA-MB-134, and PMC42 were purchased from ATCC and cultivated following the provided protocols. The cell line IPH-926 was established in our institution and is described comprehensively elsewhere [21].

Methylation analysis using the 450k array

DNA methylation analysis using the Infinium HumanMethylation450k BeadChip (Illumina, Inc., San Diego, CA, USA) was performed according the manufactures' instruction. The HumanMethylation450 BeadChip was developed to assay more than 480,000 CpG sites selected CpG loci in parallel (Bibikova et al., 2011). DNA methylation data were processed using GenomeStudio software (ver. 2011.1; Illumina, Inc.) applying the default settings.

Methylation analysis using pyrosequencing

PCR products were generated in a 25 μ L reaction volume with 400 nmol/L of forward, 40 nmol/L reverse and 400 nmol/L universal biotinylated primers, 200 μ mol/L of each dNTP, 1.5 mmol/L or 2.5 mmol/L $MgCl_2$ (see Additional file 7 for all primer sequences and reaction conditions), 1x Platinum-Taq reaction buffer and 1.25 units PlatinumTaq™

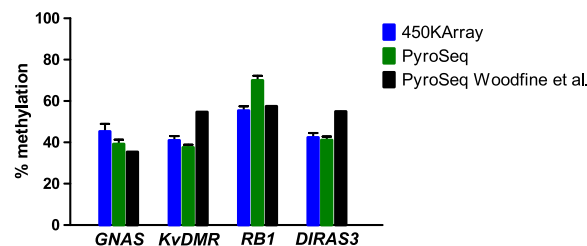


Figure 3 Comparison of methylation levels for four imprinted loci. The methylation level of *GNAS*, *KvDMR*, *RB1*, and *DIRAS3* was measured in four healthy normal mammary tissue samples using pyrosequencing and the 450k methylation array. Shown are the mean values, the error bars indicating the standard deviation. For *KvDMR* six CpG sites represented on the 450k array are measured by the pyrosequencing assay. For the remaining three loci only one CpG site was also measured by the pyrosequencing assay. In addition the data from Supplementary Table S2 from Woodfine et al. [17] are incorporated. These data are also obtained by pyrosequencing from 3 or 6 normal mammary tissue specimens. No standard deviations are reported for these data.

(Invitrogen, Karlsruhe, Germany). PCR conditions were 95°C for 5 minutes, followed by 45 cycles with denaturation at 95°C for 30 seconds, annealing at 55°C or 60°C for 45 seconds, and elongation at 72°C for 30 seconds finished with 1 cycle final elongation at 72°C for 5 minutes. The reverse primer is tagged by a sequence recognized by the universal primer. Therefore, a single (expansive) biotinylated primer can be used for all different gene-specific assays [22].

PCR products (5–20 µL) were added to a mix consisting of 3 µL Streptavidin Sepharose HP™ (Amersham Biosciences, Freiburg, Germany) and 47 µL binding buffer (Qiagen, Hilden, Germany) and mixed at 1200 rpm for 5 minutes at room temperature.

Using the Vacuum Prep Tool™ (Qiagen, Hilden, Germany), single-stranded PCR products were prepared following the manufacturer's instructions. The sepharose beads with the single stranded templates attached were released into a PSQ 96 Plate Low™ (Qiagen, Hilden, Germany) containing a mix of 12 µL annealing buffer (Qiagen, Hilden, Germany) and 500 nmmol/L of the corresponding sequencing primer (see Additional file 7). Pyrosequencing reactions were performed in a PyroMark MD System (Qiagen, Hilden, Germany) according to the manufacturer's instructions using the PyroGold SQA™ Reagent Kit (Qiagen, Hilden, Germany). CpG site quantification was performed using the methylation Software Pyro Q-CpG™.

Statistical analyses

All calculations were performed using GraphPad Prism5 software. p-values smaller than 0.5 were considered statistically significant.

For the comparison of the two methods Bland-Altman-Plots were generated [23]. In these plots the difference of two methods is plotted against the average of both methods. A comprehensive description of this type of data presentation by Altman and Bland ("Measurements in Medicine: the Analysis of Method Comparison Studies", *The Statistician* 32 (1983) 307 – 317) can be found freely available at: <https://person.hst.aau.dk/slc/Teaching/Papers/BlandAltman83.pdf> (availability checked: 17th February 2012). In order to obtain data sets of comparable range for the construction of the Bland-Altman-plots, the β -values were transformed using the following equation: β -value \times 100 = % methylation (Illumina).

Additional files

Additional file 1: Linearity of pyrosequencing for 4 different genes (*SFRP1*, *APC*, *DAPK*, *KvDMR*).

Additional file 2: Comparison of the concordance between pyrosequencing and the 450k array if the same CpG island is analyzed. For this purpose the mean methylation levels obtained by each method for a given CpG island were compared. Indicated are correlation coefficients according to Spearman (r) and linear regression coefficients (r^2).

Additional file 3: Coordinates for all individual CpG sites compared in Figure 1.

Additional file 4: Comparison of the concordance between pyrosequencing and the 450k array for Infinium I and II assays separately. Indicated are correlation coefficients according to Spearman (r) and linear regression coefficients (r^2).

Additional file 5: Scatter plot (A) and Bland-Altman-Plots without (B) and with (C) peak correction for the data from Dedeurwaerder et al. Table S2.

Additional file 6: Primary human specimens used in this study.

Additional file 7: Sequence of all primers used in this study.

Abbreviations

bp: Base pair.

Competing interest

The authors declare that they have no competing interests.

Acknowledgements

This research project was supported in part by grants from the Deutsche Forschungsgemeinschaft (SFB-TRR77, project B1) and Deutsche Krebshilfe (project 109435). The authors would also like to thank Elisa Schipper for excellent technical assistance.

Author details

¹Institute of Pathology, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1D-30625, Hannover, Germany. ²Institute of Human Genetics, University Hospital Schleswig Holstein, Kiel, Germany.

Authors' contributions

UL, OA and JR conceived the study; JR and BH prepared all samples and performed all pyrosequencing measurements; JG and OA performed the 450k array hybridization; JR, JG, OA, SLA and UL analyzed the 450k and the pyrosequencing data; HK selected and evaluated all cases; UL and JR wrote the manuscript with support from HK, SLA, OA, and JG. All authors read and approved the final manuscript.

Received: 9 February 2012 Accepted: 30 April 2012

Published: 30 April 2012

References

1. Laird PW: Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* 2010, **11**(3):191–203.
2. Satterlee JS, Schubeler D, Ng HH: Tackling the epigenome: challenges and opportunities for collaboration. *Nat Biotechnol* 2010, **28**(10):1039–1044.
3. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al: Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009, **462**(7271):315–322.
4. Zuo T, Tycko B, Liu TM, Lin HJ, Huang TH: Methods in DNA methylation profiling. *Epigenomics* 2009, **1**(2):331–345.
5. Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Gnirke A, Stunnenberg HG, Meissner A: Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol* 2010, **28**(10):1106–1114.
6. Steemers FJ, Gunderson KL: Whole genome genotyping technologies on the BeadArray platform. *Biotechnol J* 2007, **2**(1):41–49.
7. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, et al: High density DNA methylation array with single CpG site resolution. *Genomics* 2011, **98**(4):288–295.
8. Sandoval J, Heyn HA, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M: Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 2011, **6**(6):692–702.
9. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F: Evaluation of the Infinium Methylation 450K technology. *Epigenomics* 2011, **3**(6):771–784.
10. Brakensiek K, Wingen LU, Langer F, Kreipe H, Lehmann U: Quantitative high-resolution CpG island mapping with Pyrosequencing reveals disease-specific methylation patterns of the CDKN2B gene in

- myelodysplastic syndrome and myeloid leukemia. *Clin Chem* 2007, **53**(1):17–23.
11. Potapova A, Albat C, Hasemeier B, Haeussler K, Lamprecht S, Suerbaum S, Kreipe H, Lehmann U: **Systematic cross-validation of 454 sequencing and pyrosequencing for the exact quantification of DNA methylation patterns with single CpG resolution.** *BMC Biotechnol* 2011, **11**:6.
 12. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, et al: **DNA methylation profiling of human chromosomes 6, 20 and 22.** *Nat Genet* 2006, **38**(12):1378–1385.
 13. Shoemaker R, Deng J, Wang W, Zhang K: **Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome.** *Genome Res* 2010, **20**(7):883–889.
 14. Lehmann U, Hasemeier B, Christgen M, Muller M, Romermann D, Langer F, Kreipe H: **Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer.** *J Pathol* 2008, **214**(1):17–24.
 15. Tessema M, Langer F, Bock O, Seltsam A, Metzsig K, Hasemeier B, Kreipe H, Lehmann U: **Down-regulation of the IGF-2/H19 locus during normal and malignant hematopoiesis is independent of the imprinting pattern.** *Int J Oncol* 2005, **26**(2):499–507.
 16. Robertson KD: **DNA methylation and human disease.** *Nat Rev Genet* 2005, **6**(8):597–610.
 17. Woodfine K, Huddleston JE, Murrell A: **Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue.** *Epigenetics Chromatin* 2011, **4**(1):1.
 18. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A: **The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing.** *PLoS One* 2010, **5**(1):e8888.
 19. Jin SG, Kadam S, Pfeifer GP: **Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine.** *Nucleic Acids Res* 2010, **38**(11):e125.
 20. Nestor C, Ruzov A, Meehan R, Dunican D: **Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA.** *Biotechniques* 2010, **48**(4):317–319.
 21. Christgen M, Bruchhardt H, Hadamitzky C, Rudolph C, Steinemann D, Gadzicki D, Hasemeier B, Romermann D, Focken T, Krech T, et al: **Comprehensive genetic and functional characterization of IPH-926: a novel CDH1-null tumour cell line from human lobular breast cancer.** *J Pathol* 2009, **217**(5):620–632.
 22. Colella S, Shen L, Baggerly KA, Issa JP, Krahe R: **Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites.** *Biotechniques* 2003, **35**(1):146–150.
 23. Bland JM, Altman DG: **Statistical methods for assessing agreement between two methods of clinical measurement.** *Lancet* 1986, **1**(8476):307–310.

doi:10.1186/1756-0500-5-210

Cite this article as: Roessler et al: Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc.. *BMC Research Notes* 2012, **5**:210.

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

