Approximate Bayesian bisulphite sequencing analysis

Owen J.L. Rackham[1,*], Sarah R. Langley[1,*], Thomas Oates[2,*], Eleni Vradi[3,*], Nathan Harmston[1], Prashant K. Srivastava[4], Jacques Behmoaras[5], Petros Dellaportas[6,§], Leonardo Bottolo[7,8,§] & Enrico Petretto[1,§]

We present a Bayesian smoothing approach to detect differentially methylated regions from whole-genome bisulfite sequencing (WGBS) data. The method exploits the Integrated Nested Laplace Approximation for fast and accurate model fitting, an alternative to computationally expensive sampling-based methods. We demonstrate the approach by extensive simulation studies and WGBS of macrophages in experimental glomerulonephritis, revealing differential *Ifitm3* promoter methylation in glomerulonephritis, supported by differential transcription factor binding and *Ifitm3* gene expression.

One of the most important epigenetic modifications directly affecting DNA is methylation, where a methyl group is added to a cytosine base in the DNA sequence creating 5-methylcytosine. Whilst it is still not fully understood how DNA methylation affects gene expression, it has been shown that depending on the location of the modification it can either have a positive or negative effect on the level of expression of genes[1]. The majority of functional methylation changes are found in methylation sites where cytosines are immediately followed by guanines, known as CpG dinucleotides. These are not positioned randomly across the genome but tend to appear in clusters called CpG islands (CpGI)[2]. It has been also shown that there are concordant methylation changes within CpGI and in the genomic regions immediately surrounding CpGI (also known as CpGI shores or CpGS). These "spatially correlated" DNA methylation patterns tend to be more strongly associated with gene expression changes than the methylation changes occurring in other parts of the genome[3].

High-throughput sequencing techniques, such as WGBS, now allow for genome-wide methylome data to be collected at single base-resolution[4]. However, the challenge remains on how to efficiently identify DNA methylation changes at the genome-wide level, and also account for the complex correlation structures present in the data. Beyond the univariate analysis of methylation changes at each CpG, recently the focus has shifted to identifying differentially methylated regions (DMRs), since CpG methylation status is highly dependent on the status of the surrounding CpG sites. To this aim, a number of tools have been proposed to detect DMRs from WGBS data. Typically, in the first level the number of methylated/unmethlyated reads is modeled (e.g., using binomial or negative-binomial distribution) and then, in a second level, a smoothing operator is applied, for instance a local-likelihood smoother (e.g., BSmooth[5]) or a lognormal-beta hierarchical model (e.g., DSS[6]). Other strategies have been developed (for example MethylKit[7], MethySig[8]) and reviewed in Robinson *et al*[9].

Here we propose a new approach, approximate Bayesian bisulphite sequencing analysis (or ABBA), designed to smooth automatically the underlying - not directly observable - methylation profiles and reliably identify DMRs. This approach takes advantage of the recently introduced Integrated Nested Laplace Approximation (INLA)[10], a technique that allows for a fast and accurate fitting of the parameters in terms of both convergence and computational time when compared to sampling-based methods such as Markov chain Monte Carlo[11] (MCMC) or Sequential Monte Carlo[12] (SMC). ABBA is a Bayesian structured generalized mixed additive model with a latent Gaussian field (i.e., the unobserved methylation profile), controlled by a few hyperparameters, and with a non-Gaussian response variable (i.e., the number of methylated/unmethlyated reads). ABBA provides the estimate of the posterior distribution of the smoothed unobserved methylation profile as well as the posterior methylation probability (PMP) at each CpG site. By contrasting the whole-genome PMPs between two groups, e.g. cases and controls, ABBA identifies DMRs at a specified False Discovery Rate (FDR) (**Fig. 1a**). Several intrinsic features of the data are incorporated into the model. For instance, the variability of the experimental replicates within each group is modeled through a random effect with a specific within-group variance. The spatial correlation of DNA methylation patterns is encoded in the latent Gaussian field equation, which reflects the neighborhood structure of the model. In particular, the *a priori* correlation between CpGs' methylation profile values depends on the distance between the CpGs and decreases as this distance increases. The degree of smoothing of the unobserved methylation profile is controlled by the variance of the marginal density of the latent Gaussian field. Rather than relying on a user-defined value for this parameter or fixing it by an automatic procedure (for instance through an empirical Bayes approach), ABBA assigns a non-informative prior to the variance of the marginal Gaussian density. This specification is key in our model since the data-adaptivity of the degree of smoothing permits the analysis of different scenarios rather than assuming fixed (genome-wide) values. All these features allow our model to adjust better to real-world scenarios, providing a purely data-driven way to describe the WGBS data without requiring any user-defined parameters. The

[1]Duke-NUS Graduate Medical School, Singapore. [2]Clinical Sciences Centre, Imperial College London, UK. [3]Department of Statistics, Athens University of Economics and Business, Athens, GR. [4]Division of Brain Sciences, Imperial College Faculty of Medicine, London, UK. [5]Centre for Complement and Inflammation Research, Imperial College London, UK. [6]Deparment of Statistics, University College London, UK. [7]Department of Medical Genetics, University of Cambridge, UK. [8]Department of Mathematics, Imperial College London, UK. *These authors contributed equally to this work. [§]Correspondence authors.

PMP allows the straightforward derivation of the posterior differential methylation probability (PDMP) at each CpG site between the two groups. This quantity is used in a Bayesian non-parametric FDR procedure to distinguish PDMPs that belong to the null distribution (no differential methylation) and the alternatives. As in Broët *et al.*[13], the classification of PDMPs into two groups permits the genome-wide calculation of the cut-off level to declare differentially methylated CpG sites at a fixed FDR level. Full details of ABBA model, INLA approximate inference and FDR procedure can be found in the **Online Methods**.

We benchmarked ABBA and compared it against recently proposed methods (DSS[6], MethylKit[7], MethySig[8], BSmooth[5] and the univariate Fisher's exact test (FET)). To ensure an objective comparison, we used WGBSSuite[14] to generate test datasets that are independent of the underlying statistical models of ABBA and of the other methods. Briefly, we generated test datasets to assess the performance of each method under several scenarios, which have been modeled using a combination of parameters related to data integrity and quality of the signal. The parameters considered are the following; the number of replicates within each group (r), the average read depth per CpG, background level of noise variance (s_0), the methylation probability difference between the two groups ($\Delta meth$) and the random perturbation parameter (δ) used here to simulate the errors due to incomplete bisulfite conversion (see **Online Methods** for details). For each simulated case we compared the DMRs called by each technique with the true simulated DMRs. To quantitatively assess the performance of ABBA with respect to competing methods, we evaluated false-positive and false-negative rates and generated receiver operator characteristic (ROC) curves, focusing on the partial area under the ROC curve (or pAUC) at a specificity of 0.75. The pAUC is considered to be more practically relevant than the area under the entire ROC curve[15] since in typical genomics studies only the features identified at very low false positive rates are selected for further biological validation. All results of the benchmark are detailed in **Supplementary Fig. 1-3** and summarized graphically in **Fig. 1b**, which shows, for a given combination of parameters, the best performing method based on its pAUC. Specifically, the colors in the "benchmark grid" indicate the best performing method in each of the 162 different simulated scenarios. Considering all simulated datasets, ABBA (black) showed to be the best method in 138 (85%) cases or to have a similar performance to another method (most often DSS) in 13 (8%) cases. Only in 11 (7%) simulated scenarios was ABBA not the best performing method, and in these cases DSS frequently showed the best performance (e.g., with random perturbation δ = 10% and $\Delta meth$ = 50% or 70%; right hand-side of **Fig. 1b**). Looking at the detailed ROC curves reported in **Supplementary Fig. 1-3**, the difference in performance between ABBA (and DSS) and all other methods is most apparent for simulated datasets with low coverage and relatively small methylation probability difference between the two groups. Overall the ABBA curve dominates the entire range of pAUC's sensitivity/specificity in 82% of the benchmarks. In particular, for all simulated scenarios with 10x coverage and $\Delta meth$ = 30% and irrespective of the values set for all other parameters, ABBA was the best performing method (with exception of one case where its performance was similar to that of BSmooth (cyan)), **Fig. 1b**. This suggests that in the simulated cases where it is most challenging to detect DMRs (i.e., low WGBS coverage, small differences in methylation, small number of replicates and in presence of incomplete bisulfite conversion errors) ABBA consistently outperformed all other methods. These data show the robustness of our procedure to recover the unobserved methylation profile and accurately identify DMRs when the number of observations is small and the reliability and/or sensitivity of the sequencing assay is low, or when the signal is contaminated by a high level of background noise.

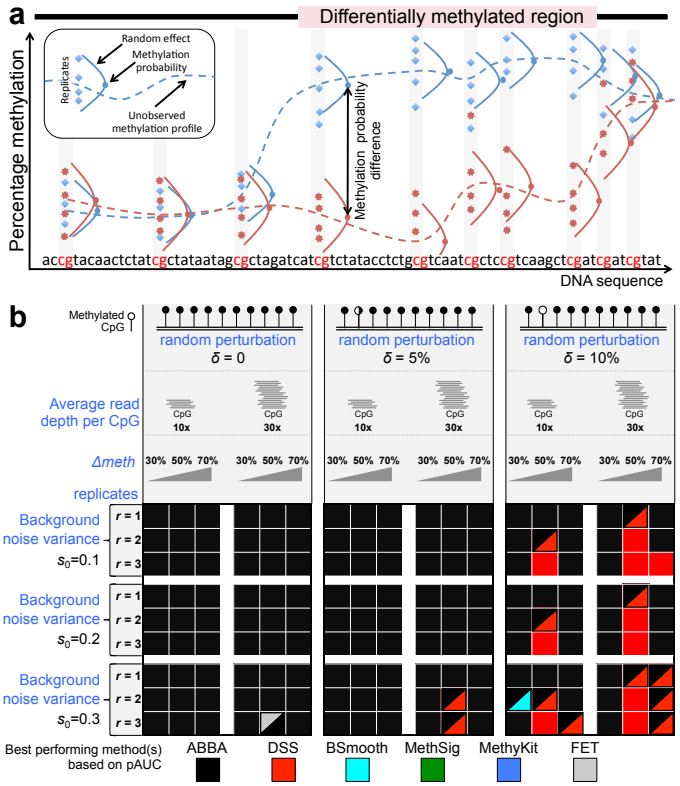


Figure 1. ABBA model and benchmarking results

*(**a**) ABBA estimates the unobserved methylation profiles from WGBS data. A random effect accounts for experimental replicates while the degree of smoothing is modeled by the latent Gaussian field that probabilistically connects consecutive methylation probabilities. Genome-wide methylation probability differences between groups identify DMRs at a fixed FDR level. (**b**) Global snapshot of the method's performance across all simulated datasets. A given combination of parameters is indicated by a square in the benchmark grid, and for each square we calculated the pAUC for each method and determined which method had the overall best pAUC (i.e., $pAUC_{method_1} > pAUC_{method_2}$). Colours in the benchmark grid indicate which method had the best performance. When pAUC*

of two methods are similar (±1%) we report the colours of both methods (e.g., black and red colours in the same square indicate similar performance of ABBA and DSS).

Some differences in method performance start to emerge in the presence of frequent errors occurring in the bisulfite conversion of DNA. Being the first step in a WGBS experiment, bisulfite conversion results in the selective deamination of cytosine to uracil of the DNA (leaving methylated cytosines unchanged), which can then be sequenced and ultimately provide a measure of DNA methylation at the single-base resolution. To mimic the error that can be introduced by failed bisulfite conversion, i.e., when an unmethylated cytosine fails to be deaminated therefore appearing as if it had been methylated, we introduced a random perturbation parameter (δ) to control when the methylation status of a small fraction of CpG sites (picked at random in one of the two populations) is inverted. With no random perturbation or 5% probability of switching methylation status (i.e., low or minimal errors in bisulfite conversion), ABBA was overall the best method or showed similar performance to DSS in two cases (**Fig. 1b**). However, when δ was as high as 10% (i.e., 1 in 10 CpGs is misclassified as unmethylated or vice versa), we observed that DSS performs as the best method in 10 out of 54 simulated scenarios, and in other 10 cases ABBA and DSS have comparable performance. This was more apparent when large probability differences between the two groups were simulated (*Δmeth* = 50% or 70%). In these cases, the drop in performance of ABBA in comparison with DSS might be due to the definition of a "significant DMR" between the two methods. In our simulations we used the default settings of DSS, which require only 80% of all CpGs within (at least) 100bp-long region to be significant in order to call a "significant DMR", whereas ABBA more conservatively requires all CpGs in a DMR to be significant at a specified FDR level. It is likely that in some instances this make DSS more robust to frequent errors introduced at random. However, we highlight that bisulfite conversion rates lower than 95% are considered suboptimal[16]; therefore the 10% random perturbation is not typically accepted in WGBS experiments. Under the more likely experimental scenario where δ is kept at 5% ABBA was the best performing method and on the whole ABBA's performance was the most robust across a variety of parameters tested.

To illustrate the practical utility of ABBA, we generated WGBS data in an established experimental model of crescentic glomerulonephritis (CRGN)[17], and assayed CpG methylation at single-nucleotide resolution in primary macrophages from four Wistar Kyoto (WKY) and four Lewis (LEW) isogenic rats (see **Online Methods**). Using an FDR cutoff of 5%, ABBA identified 1,004 DMRs genome-wide, with 1.07% falling within an annotated CpGI and 6.78% within an annotated CpGS (**Fig. 2a**). Of the 1,004 DMRs, 427 overlapped with annotated genes (**Supplementary Table 1**), and there was a significant enrichment for DMRs occurring within 1kb of the gene boundaries (P<0.001), within exons (P<0.05) and introns (P<0.05), **Fig. 2b**. The genes that are within 1kb of a DMR were enriched for pathways relevant to CRGN, including MAPK signalling[18], Phosphatidylinositol signalling[19] and Fc gamma R-mediated phagocytosis[20] (**Fig. 2c**). As DNA methylation can affect gene expression by interfering with transcription factor binding, we performed a transcription factor binding site (TFBS) analysis of the DMRs (**Fig. 2d**). This revealed enrichment for the ETS transcription factors family and a number of proteins that make the AP-1 TF complex (JUNB, FOS, JUN and JUND), which have been previously linked with CRGN[21,22]. To further investigate the potential effect of DMRs identified by ABBA, we carried out differential expression (DE) analysis in macrophages from WKY and LEW rats by RNA-seq. The list of DE genes (n=910, Benjamini–Hochberg (BH)-corrected P<0.05) was crosschecked with the genes impacted by DMRs, identifying 48 genes with both differential methylation and differential expression (**Supplementary Table 2**). Among these genes, *Ifitm3*, *Ydjc* and *Cd300Ig* showed differential methylation at their promoter region. Since the promoter is a key regulatory region where the effect of methylation is more clearly understood, we investigated these genes in detail and found the biggest change in mRNA expression was in interferon induced transmembrane protein 3 (*Ifitm3)*, with gene mRNA being almost undetected in unstimulated WKY macrophages (**Fig. 2e**). This observation is consistent with the differential methylation status of the promoter of *Ifitm3*, which was more methylated in WKY than in LEW rats (**Fig. 2f**). We have previously shown that JunD (AP-1) transcription factor is a major determinant of CRGN in WKY rats[21]. Therefore we scanned the DMR for canonical JunD binding site motifs, and identified three putative regions in the promoter region of *Ifitm3*. To further investigate whether the DNA methylation differences between WKY and LEW macrophages might underlie the differential gene regulation we re-analyzed ChIP-Seq data for JunD in WKY and in a congenic strain from LEW (see **Online Methods** for details). This analysis identified significant differences in JunD binding between WKY and LEW-congenic strain that overlapped with two of the four TFBS identified at the *Ifitm3* promoter (**Fig. 2g**). This is consistent with the hypothesis that the difference in DNA methylation is having a knock-on effect on the regulation of *Ifitm3*. Analysis of the same genomic region by other methods either failed to identify significant DMR (MethySig) or identified a large and unspecific genomic area as differentially methylated (DSS and BSmooth), **Supplementary Fig. 4**. Therefore, none of the competing methods pointed to methylation differences specific to the *Ifitm3* promoter where the TFBS and differential ChIP-seq signals were also identified. The combined evidence provided by ABBA and RNA-seq/ChIP-seq data therefore suggests that the effect of DNA methylation of the *Ifitm3* promoter in WKY rats may be restricting the binding of regulatory elements and, as a consequence, the gene is almost not expressed (<1TPM) in

unstimulated macrophages of WKY rats. DNA methylation alterations in IFN-related genes, including *Ifitm3*, have been previously observed and proposed to contribute to the pathogenesis of autoimmune diseases such as primary Sjögren's syndrome[23]. Our findings therefore are suggestive of a potential similar role for *Ifitm3* in rat glomerulonephritis.

Our extensive simulation studies and WGBS-analysis results demonstrate that ABBA is a powerful approach for the identification of DMRs from single-base resolution methylation data, without requiring user-defined parameters external to the data. ABBA outperforms other commonly used methods such as DSS[6], MethylKit[7], MethySig[8] and BSmooth[5], retaining a high degree of robustness of the results with respect to several factors affecting WGBS data integrity and quality, including sequencing coverage, number of replicates or different noise structures. This is particularly appealing in cases when considerable efforts have been expended toward generation of large-scale WGBS data from heterogeneous systems, e.g., the ENCODE project, and data quality can vary across experimental conditions and laboratories. As WGBS data generation is expected gains in scale, reliable and flexible methods will become increasingly important to ensure results reproducibility across studies.

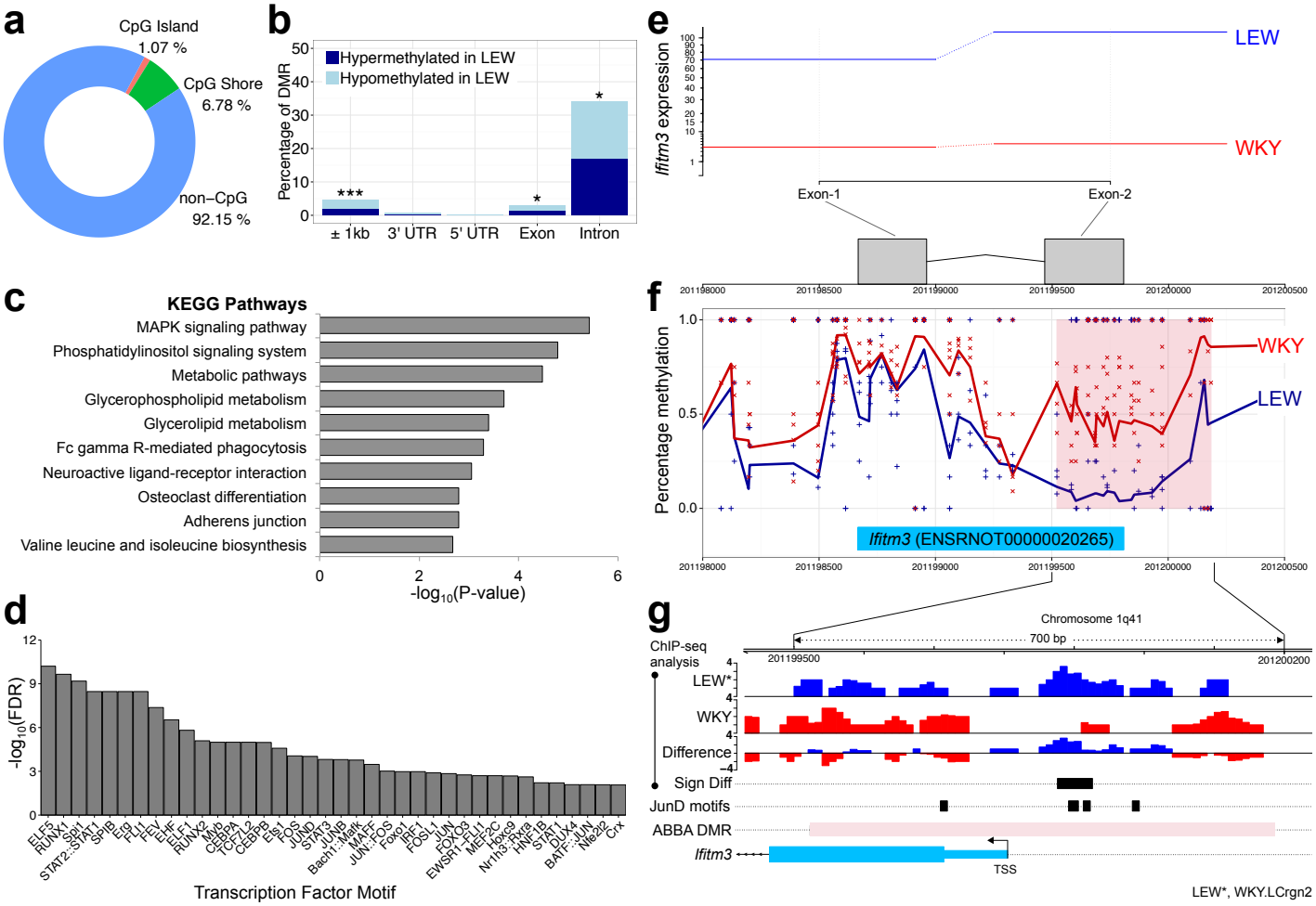


Figure 2. ABBA analysis of WGBS in rat macrophages

*(**a**) CpG-based annotation 1,004 DMR between WKY and LEW macrophages showing significantly higher proportions of CpGI and CpGS than those that would be expected by chance (P<0.009 for CpGI and P<0.001 for CpGS, respectively, obtained by 1,000 randomly sampled datasets of 1,004 CpG-matched regions). (**b**) Proportions of DMRs in different genomic features of overlapping genes. Feature annotation was retried from UCSC genome browser (RN4). (**c**) KEGG pathway enrichment for the genes overlapping with DMRs. Only significant pathways are reported (FDR<1%). (**d**) Enrichment for the TFBS within the DMRs was when compared to CG matched regions of the genome (FDR<0.05). (**e**) RNA-seq analysis in WKY and LEW macrophages shows Ifitm3 lack of expression in WKY rats. (**f**) Percentage methylation at each CpG in WKY (crosses) and LEW (plus) and smoothed methylation profiles by ABBA. The pink box highlights the significant DMR identified by ABBA (FDR<5%). (**g**) ChIP-seq analysis for JunD in LEW.LCrgn2 (LEW*) and WKY macrophages identified a region with differential binding of JunD (Sign Diff row, black box). This region overlapped with two (out of four) JunD binding sites motifs identified within the gene promoter (±500bp around the TSS). ABBA DMR, differentially methylated region identified by ABBA. TSS, transcription start site. *, P<0.05, ***, P<0.001.*

Author Contributions

L.B. and E.P. initiated, directed and supervised the project. P.D. and L.B. conceived the statistical model and the computational approach. P.D., E.V. and L.B. wrote the initial algorithm that was further developed by O.J.L.R. and L.B. to the presented approach. T.O. and E.P. generated WGBS data in the rat. S.R.L., O.J.L.R. and E.P. carried out analysis of WGBS and RNA-seq data in the rat and interpreted the results. N.H. and P.K.S. carried out ChIP-seq and TFBS analyses. J.B. provided RNA-seq and ChIP-seq data in the rat. O.J.L.R., L.B. and E.P. wrote the manuscript with input from all authors. All of the authors read and approved the final manuscript.

Acknowledgements

This research was funded by Engineering and Physical Sciences Research Council Grant EP/K030760/1 (L.B.), Royal Society IE110977 (L.B., P.D.), European Union (European Social Fund - ESF), Greek national funds through the Operational Program "Education and Lifelong Learning'' of the National Strategic Reference Framework (NSRF), project ARISTEIA (P.D.), Duke-NUS and Singapore Ministry of Health (O.J.L.R., E.P.), a Medical Research Council Chain-Florey fellowship (T.O), the Medical Research Council (MR/M004716/1 to J.B. and E.P.) and by Kidney Research UK (RP9/2013 to J.B.).

References

1. Gutierrez-Arcelus, M. *et al.* Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* **2,** e00523 (2013).
2. Mendizabal, I. & Yi, S. V. Whole-genome bisulfite sequencing maps from multiple human tissues reveal novel CpG islands associated with tissue-specific regulation. *Hum. Mol. Genet.* (2015).
3. Gutierrez-Arcelus, M. *et al.* Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. *PLoS Genet.* **11,** e1004958 (2015).
4. Harris, R. A. *et al.* Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat. Biotechnol.* **28,** 1097–105 (2010).
5. Hansen, K. D., Langmead, B. & Irizarry, R. A. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol.* **13,** R83 (2012).
6. Feng, H., Conneely, K. N. & Wu, H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res.* **42,** e69 (2014).
7. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13,** R87 (2012).
8. Park, Y., Figueroa, M. E., Rozek, L. S. & Sartor, M. A. MethylSig: a whole genome DNA methylation analysis pipeline. *Bioinformatics* **30,** 2414–22 (2014).
9. Robinson, M. D. *et al.* Statistical methods for detecting differentially methylated loci and regions. *Front. Genet.* **5,** 324 (2014).
10. Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B* **71,** 319–392 (2009).
11. Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. *Markov Chain Monte Carlo in Practice*. **39,** (1996).
12. Doucet, a, De Freitas, N. & Gordon, N. *Sequential Monte Carlo Methods in Practice*. *Springer New York* (2001).
13. Broët, P., Lewin, A., Richardson, S., Dalmasso, C. & Magdelenat, H. A mixture model-based strategy for selecting sets of genes in multiclass response microarray experiments. *Bioinformatics* **20,** 2562–71 (2004).
14. Rackham, O. J. L., Dellaportas, P., Petretto, E. & Bottolo, L. WGBSSuite: simulating whole-genome bisulphite sequencing data and benchmarking differential DNA methylation analysis tools. *Bioinformatics* **31,** 2371–3 (2015).
15. Ma, H., Bandos, A. I., Rockette, H. E. & Gur, D. On use of partial area under the ROC curve for evaluation of diagnostic performance. *Stat. Med.* **32,** 3449–58 (2013).
16. Genereux, D. P., Johnson, W. C., Burden, A. F., Stöger, R. & Laird, C. D. Errors in the bisulfite conversion of DNA: modulating inappropriate- and failed-conversion frequencies. *Nucleic Acids Res.* **36,** e150 (2008).
17. Aitman, T. J. *et al.* Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* **439,** 851–5 (2006).
18. Ryan, J. *et al.* Spleen tyrosine kinase promotes acute neutrophil-mediated glomerular injury via activation of JNK and p38 MAPK in rat nephrotoxic serum nephritis. *Lab. Invest.* **91,** 1727–38 (2011).
19. Wu, T. *et al.* Prevention of murine lupus nephritis by targeting multiple signaling axes and oxidative stress using a synthetic triterpenoid. *Arthritis Rheumatol. (Hoboken, N.J.)* **66,** 3129–39 (2014).
20. Page, T. H. *et al.* Role of novel rat-specific Fc receptor in macrophage activation associated with crescentic glomerulonephritis. *J. Biol. Chem.* **287,** 5710–9 (2012).
21. Behmoaras, J. *et al.* Jund is a determinant of macrophage activation and is associated with glomerulonephritis susceptibility. *Nat. Genet.* **40,** 553–9 (2008).
22. Raffetseder, U. *et al.* Mesangial cell expression of proto-oncogene Ets-1 during progression of mesangioproliferative glomerulonephritis. *Kidney Int.* **66,** 622–32 (2004).
23. Gottenberg, J.-E. *et al.* Activation of IFN pathways and plasmacytoid dendritic cell recruitment in target organs of primary Sjögren's syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 2770–5 (2006).

ONLINE METHODS

Latent Gaussian model. A latent Gaussian model (LGM) can be described by a three-stage hierarchical model

$$y_i|x_i,\boldsymbol{\theta} \sim \pi(y_i|x_i,\boldsymbol{\theta}), \quad (1)$$

$$\boldsymbol{x}|\boldsymbol{\theta} \sim N(\boldsymbol{\mu}(\boldsymbol{\theta}),\boldsymbol{Q}^{-1}(\boldsymbol{\theta})), \quad (2)$$

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}), \quad (3)$$

where y_i, $i=1,\cdots,n$, are the observed values, $\boldsymbol{x}$ is n-dimensional vector of latent variables and $\boldsymbol{\theta}$ is p-dimensional vector of model parameters. (1) is the *observations equation* and it describes the probabilistic model for each observation conditionally on the latent variable x_i and the model parameters $\boldsymbol{\theta}$, (2) is the *latent Gaussian field equation* with the latent variables distributed as a p-dimensional normal distribution, with mean vector $\boldsymbol{\mu}(\boldsymbol{\theta})$ and a sparse precision matrix $\boldsymbol{Q}(\boldsymbol{\theta})$. Both quantities can depend on the model parameters vector $\boldsymbol{\theta}$ whose distribution is described in the *parameter equation* (3). The Gaussian vector $\boldsymbol{x}$ exhibits a particular conditional dependence (or Markov) structure which is reflected in its precision matrix $\boldsymbol{Q}(\boldsymbol{\theta})$.

Given (1), (2) and (3), the joint posterior can be written as

$$\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\prod_{i=1}^{n}\pi(y_i|x_i,\boldsymbol{\theta})$$

$$\propto \pi(\boldsymbol{\theta})|\boldsymbol{Q}(\boldsymbol{\theta})|^{1/2}\exp\Big\{-\frac{1}{2}\big(\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\theta})\big)^T\boldsymbol{Q}(\boldsymbol{\theta})\big(\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\theta})\big) + \sum_{i=1}^{n}\log\big(\pi(y_i|x_i,\boldsymbol{\theta})\big)\Big\}.$$

As shown in[24] for Gaussian Markov random fields (GRMF), with LGM as a special case of GRMF, core linear algebra operations using a dense matrix $\boldsymbol{Q}(\boldsymbol{\theta})$ become computational expensive resulting in infeasible run times. Computational efficient inferential methods rely instead on the sparseness of the precision matrix $\boldsymbol{Q}(\boldsymbol{\theta})$ since the most common linear algebra operations can be performed much faster for sparse matrices than for dense ones. Further computational efficiency can be reached when the dimensional vector of model parameters is also small.

Integrated Nested Laplace Approximation. Integrated Nested Laplace Approximation (INLA) is a new approach to Bayesian statistical inference introduced by[25] and subsequently extended[10]. It provides a fast and accurate alternative to MCMC and other sampling-based methods such as Sequential Monte Carlo[12]. The main advantage of INLA over MCMC/SMC is that it is much faster to compute; it gives answers in minutes or seconds where MCMC requires hours or days.

INLA[10] builds upon the Laplace approximation which was proposed originally by [26] to posterior moments in a Bayesian set-up but can also be used to approximate marginal posterior densities. Below we describe the main approximation steps implemented in INLA. For LGM (1-3) the marginal posteriors of interest are

$$\pi(x_i|\boldsymbol{y}) = \int \pi(x_i|\boldsymbol{y},\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta},$$

$$\pi(\theta_j|\boldsymbol{y}) = \int \pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}_{\backslash j},$$

where $\boldsymbol{\theta}_{\backslash j}$ indicates the whole model parameters vector but the jth element. INLA seeks to approximate both quantities by

$$\tilde{\pi}(x_i|\boldsymbol{y}) = \int \tilde{\pi}(x_i|\boldsymbol{y},\boldsymbol{\theta})\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}, \quad (4)$$

$$\tilde{\pi}(\theta_j|\boldsymbol{y}) = \int \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}_{\backslash j}, \quad (5)$$

where (4) and (5) are obtained by approximate $\pi(x_i|\boldsymbol{y},\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$. Finally numerical integrations (finite sum) of the approximations to $\pi(x_i|\boldsymbol{y})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ are carried out on selected support points of $\boldsymbol{\theta}$ provided that the dimension of $\boldsymbol{\theta}$ is not too large

$$\tilde{\pi}(x_i|\boldsymbol{y}) = \sum_t \tilde{\pi}(x_i|\boldsymbol{y},\boldsymbol{\theta}^t)\tilde{\pi}(\boldsymbol{\theta}_t|\boldsymbol{y})\Delta^t,$$

$$\tilde{\pi}(\theta_j|\boldsymbol{y}) = \sum_t \tilde{\pi}(\boldsymbol{\theta}^t_{\backslash j}|\boldsymbol{y})\Delta^t_{\backslash j}$$

with Δ^t suitable weights. Crucially in (4) and (5), the approximation to the marginal posterior of $\boldsymbol{\theta}$ (up to the unknown constant of normalization $\pi(\boldsymbol{y})$)

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})}$$

is obtained by computing the Laplace approximation

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \left.\frac{\pi(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})}{\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})}\right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})} \quad (6)$$

with $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ the Gaussian approximation of the full conditional distribution

$$\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) \propto \exp\Big\{-\frac{1}{2}\big(\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\theta})\big)^T\boldsymbol{Q}(\boldsymbol{\theta})\big(\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\theta})\big) + \sum_{i=1}^{n}\log\big(\pi(y_i|x_i,\boldsymbol{\theta})\big)\Big\}$$

evaluated in the posterior mode $\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})$ that is obtained by a Newton-Raphson algorithm, for a given $\boldsymbol{\theta}$, i.e $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ is well approximated by a Gaussian distribution by matching the mode and the curvature at the mode. Although $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ might be non-Gaussian, the crude approximation $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is nonetheless very accurate

with no differences with the output produced by a long MCMC run[25].

To calculate (4), INLA introduces a further approximation to $\pi(x_i|\boldsymbol{y},\boldsymbol{\theta})$ (up to the unknown constant of normalization $\pi(\boldsymbol{y},\boldsymbol{\theta})$)

$$\pi(x_i|\boldsymbol{y},\boldsymbol{\theta}) \propto \frac{\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{x}_{\backslash i}|\boldsymbol{y},x_i,\boldsymbol{\theta})}$$

with the Laplace approximation

$$\tilde{\pi}_{LA}(x_i|\boldsymbol{y},\boldsymbol{\theta}) \propto \left.\frac{\pi(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})}{\tilde{\pi}_{GG}(\boldsymbol{x}_{\backslash i}|\boldsymbol{y},x_i,\boldsymbol{\theta})}\right|_{\boldsymbol{x}_{\backslash i}=\boldsymbol{x}^*_{\backslash i}(x_i,\boldsymbol{\theta})} \quad (7)$$

where $\tilde{\pi}_{GG}(\boldsymbol{x}_{\backslash i}|\boldsymbol{y},x_i,\boldsymbol{\theta})$ is the Gaussian approximation to the full conditional distribution $\pi(\boldsymbol{x}_{\backslash i}|\boldsymbol{y},x_i,\boldsymbol{\theta})$ and $\boldsymbol{x}^*_{\backslash i}(x_i,\boldsymbol{\theta})$ is its modal configuration. The computational expensive evaluation of the conditional posterior mode $\boldsymbol{x}^*_{\backslash i}(x_i,\boldsymbol{\theta})$ is also replaced by an approximated value to make the algorithm faster[10].

In summary INLA consists of three steps:

1. Compute the approximation to the marginal posterior of $\boldsymbol{\theta}$ (6) and, by-product, compute by finite sum the approximation to $\pi(\theta_j|\boldsymbol{y})$ (5);
2. Compute the approximation to $\pi(x_i|\boldsymbol{y},\boldsymbol{\theta})$ (7);
3. Combine (6) and (7) and, by finite sum, compute the approximation to $\pi(x_i|\boldsymbol{y})$ (4).

ABBA model. Based on LGM, ABBA can be described by a three-stage hierarchical model:

$$y_{igr}|\pi_{igr} \sim \text{Binomial}(n_{igr},\pi_{igr}), \quad (8)$$

$$\text{logit}(\pi_{igr})\,|\sigma_g^2 \sim \text{N}(\mu_{ig},\sigma_g^2) \quad (9)$$

$$\mu_{ig}|\rho_{ig}^2 \sim \text{N}(\mu_{i-1,g},\rho_{ig}^2) \quad (10)$$

$$\sigma_g^2 \sim \text{Gam}(0.1,0.1), \quad \rho_g^2 \sim \text{Gam}(0.1,0.1). \quad (11)$$

Eq. (8) is the *first part of the observations equation* where $i=1,\cdots,m$ denotes the CpG, $g=1,2$ the group (e.g., case and control group), and $r=1,\cdots,R$ the experimental replicate. y_{igr}, n_{igr} and π_{igr} are the observed number of methylated/unmethylated reads, the read depth and proportion of methylation for the ith CpG site, gth group and rth experimental replicate, respectively. Eq. (9) is the *second part of the observations equation* and it describes a random effect across the experimental replicates with a specific variance σ_g^2 for each group. In (9), $\text{logit}(z)$ indicates the logit transformation, $\text{logit}(z)=\log(1/(1-z))$. Together the observation equations (8) and (9) assume that the proportion of methylation are exchangeable within each group but are different between groups.

Eq. (10) is the *latent Gaussian field equation.* The spatial correlation between CpGs is modelled as a non-stationary random walk of order 1, RW(1): μ_{ig} follows a normal distribution with mean $\mu_{i-1,g}$ (defined in the $(i-1)$th CpG) and variance σ_{ig}^2. For parsimony reasons and for INLA implementation, we define $\rho_{ig}^2=\rho_g^2|p_i-p_{i-1}|$, where p_i and p_{i-1} are the chromosomal locations of two consecutive CpG sites. This automatically implies that the correlation between μ_{ig} and μ_{lg}, $i\neq l$, depends on the distance between the two CpG at sites i and l and, in particular, it decreases as this distance increases. This formulation implies also a sparse precision matrix for the LGF[27]. The model is completed by the prior specification of σ_g^2 and ρ_g^2 (11). Both quantities are distributed as a gamma density with mean 1 and variance 10 which are the default INLA values. Sensitivity analysis on the gamma density parameterization shows no departure from the results obtained using the default values (data not shown).

When a single replicate is available, since $\sigma_g^2=0$, Eq. (8) and (9) simplify to

$$y_{ig}|\pi_{ig} \sim \text{Binomial}(n_{ig},\pi_{ig}),$$
$$\text{logit}(\pi_{ig})=\mu_{ig}.$$

In summary, INLA inference of the ABBA model consists of three steps:

1. Compute the approximation to the marginal posteriors of σ_g^2, the variance of the random effect, and ρ_g^2, $g=1,2$ the smoothing parameters;
2. Compute the approximation to $\pi(\mu_{ig}|\boldsymbol{y})$, where $\boldsymbol{y}=(y_{igr})_{i=1,\cdots,n;g=1,2;r=1,\cdots,R}$;
3. Reconstruct the marginal posterior of the unobserved methylation profile $\pi(\pi_{ig}|\boldsymbol{y})$ using the inverse logit transformation of μ_{ig}, $z=\exp\{\text{logit}(z)\}/[1+\exp\{\text{logit}(z)\}]$.

Differential methylation and FDR calculation. Let's define $\pi(\pi_{ig}|\boldsymbol{y})$, $i=1,\cdots,m$, $g=1,2$, the posterior methylation probability (PMP) and $\pi(\pi_{i1}|\boldsymbol{y})-\pi(\pi_{i2}|\boldsymbol{y})$ the posterior differential methylation probability (PDMP). The posterior mean methylation probability $\text{E}(\pi_{ig}|\boldsymbol{y})$ is used to define the posterior mean differential methylation between the two groups, $d_i=\text{E}(\pi_{i1}|\boldsymbol{y})-\text{E}(\pi_{i2}|\boldsymbol{y})$. To distinguish between the null distribution (no differential methylation) and the alternatives, and invoking the central limit theorem for asymptotic considerations, we fit a normal mixture model with three components

$$d_i-c \sim \pi_-\text{N}(\theta_-,\xi_-^2)+\pi_0\text{N}(0,\xi_0^2)+\pi_+\text{N}(\theta_+,\xi_+^2), \quad (13)$$

where c is a constant (see below), $\pi_-, \pi_0, \pi_+ \in (0,1)$ with $\pi_- + \pi_0 + \pi_+ = 1$ are the mixing weights of the "negative" differentially methylated, no differentially methylated and "positive" differentially methylated with respect the control group, respectively, θ_-, θ_+ are the unknown centers of the differentially methylated groups and $\xi_-^2, \xi_0^2, \xi_+^2$ are the unknown variances. In (13) we also impose that $\pi_0 \geq \pi_- + \pi_+$ to reflect the hypothesis that the large majority of CpG sites are not differentially methylated.

Maximum likelihood estimates of (13) are obtained by the EM algorithm[28] taking particular care to avoid local maxima in the likelihood surface by running the EM algorithm many times initialized from different starting points. Using the EM algorithm the posterior probability the a CpG site belongs to each of the three component is

$$\mathrm{P}(z_i = "-") = \frac{\pi_- \mathrm{N}(d_i - c; \theta_-, \xi_-^2)}{C},$$

$$\mathrm{P}(z_i = "0") = \frac{\pi_0 \mathrm{N}(d_i - c; 0, \xi_0^2)}{C},$$

$$\mathrm{P}(z_i = "+") = \frac{\pi_+ \mathrm{N}(d_i - c; \theta_+, \xi_+^2)}{C}$$

with $C = \pi_- \mathrm{N}(d_i - c; \theta_-, \xi_-^2) + \pi_0 \mathrm{N}(d_i - c; 0, \xi_0^2) + \pi_+ \mathrm{N}(d_i - c; \theta_+, \xi_+^2)$.

The constant c used in (13) is the mode of central component and it allows us to center the null component in 0. It is obtained by fitting a kernel density to all d_i's that lie in the interquartile range and calculate the mode of the density numerically.

Similarly to Muller *et al*[29] and Broët *et al*[13], for a constant t, we define the estimated $\mathrm{FDR}(t)$ as

$$\widehat{\mathrm{FDR}}(t) = \sum_{i \in \mathcal{I}_-} \frac{\mathrm{P}(z_i = "-")}{n_-} + \sum_{i \in \mathcal{I}_+} \frac{\mathrm{P}(z_i = "+")}{n_+} \qquad (14)$$

where $\mathcal{I}_- = \{i: (d_i - c) \leq -t\}$, $\mathcal{I}_- = \{i: (d_i - c) \geq t\}$, $n_- = \#(\mathcal{I}_-)$ and $n_+ = \#(\mathcal{I}_+)$. Eq. (14) defined global FDR as the average local FDR which for posterior probabilities is defined as $1 - \mathrm{P}(z_i = "0")$. Finally the constant t is chosen such that $\widehat{\mathrm{FDR}}(t) \leq \mathrm{FDR}$ with the two cut-off values in the logit scale defined as $t_- = -t + c$ and $t_+ = t + c$.

In summary, FDR procedure consists of three steps:

1. Calculate the constant c using a non-parametric kernel density procedure;
2. Fit a normal mixture model with three components on the $d_i - c$ values; obtain the posterior probability that each $d_i - c$ belongs to each of the three components;
3. Calculate the constant t such that $\widehat{\mathrm{FDR}}(t) \leq \mathrm{FDR}$ for a desired level of FDR;
4. Obtained the cut-off levels t_- and t_+ to define "negative" and "positive" differentially methylated CpG sites (with respect the control group) controlling the FDR level.

WGBS data simulation. WGBS data have a number of intrinsic characteristics that can vary depending on the cell-types/tissue complexity being studied or on technical issues related to the sequencing. In order assess which method is the most robust for analyzing WGBS data it is important that changes in each of these characteristics are taken into account. Here we take advantage of our previously published WGBS-data simulator[14] that allows us to generate unbiased benchmarking datasets with several varying parameters. Wherever possible we will refer to the notation used in *Rackham et al.*[14] and detailed in the Supplementary Material. The parameters are the following:

1. The number of replicates - the parameter r was set to vary between $r = 1,2,3$ within each group;
2. Average read depth – at each CpG site for all replicates and groups, the number of reads n_{igr}, $i = 1, \cdots, m$ and $g = 1,2$, is simulated using a Poisson distribution with average read depth λ. The parameter λ was set to be either 10 or 30 reads on average per CpG site;
3. Background level of noise - the parameter s_0 controls the level of noise added the probability of methylation at each CpG site for all replicates and groups

$$p_{irg} = \mathrm{logit}^{-1}\big(\mathrm{logit}(p_{rg}) + \varepsilon_i\big),$$

where p_{rg} is the constant probability of methylation for all r, g and $\varepsilon_i \sim \mathrm{N}(0, s_0)$, $i = 1, \cdots, m$. This parameter was set to vary between 0.1 (small), 0.2 (medium) and 0.3 (large) to model different level of noise;
4. Methylation probability difference - the parameter $\Delta meth$ reported in [14] as "phase difference" was set to vary between 30%, 50% or 70% difference between the probabilities of methylation in each group;
5. We also considered an additional parameter δ (not available for modeling in WGBSSuite), which introduces a random perturbation to the probability of methylation. After selecting at random with a given probability δ a CpG site in the gth group for all replicates, we switch it's methylation status between the two groups, mimicking the error that is introduced when the bisulphite conversion is not 100% efficient such as failed conversion, i.e., an unmethylated cytosine fails to be deaminated, and thus appears as if it had been methylated. In our simulation study, the parameter δ has been varied from 0, 0.05 and 0.1

To perform the benchmarking we generate 5,000 CpGs for each combination of the above parameters. The resulted in a total of 162 benchmarking datasets (3 cases

for the number of replicates, 2 cases for the average read depth, 3 cases for the background level of noise, 3 cases for the methylation probability difference, 3 cases for the random perturbation) which are replicated 5 times to assess the Monte Carlo average performance for each combination of parameters.

WGBS data pre-processing for ABBA. To run ABBA efficiently at the genome-wide level we took advantage of cluster-computing environment that enables parallel computation, and to this aim we preprocessed the WGBS data as follows. After the raw WGBS data are aligned, we removed CpG sites where less than 50% of the samples contain reads. Next, we split the WGBS data into chunks such that the distance between the last CpG site in one chunk and the first CpG in the next chunk is greater than 3,000bp. It has been previously shown that the spatial correlation between CpG sites decreases dramatically after 400bp[30], so splitting the data in this way implies a particular conditional dependence structure in our data defined by a sparse block-diagonal precision matrix $\boldsymbol{Q}(\boldsymbol{\theta})$ where each block corresponds to a WGBS chunk. Chunks are then analyzed in parallel in a cluster-computing environment. We calculate the time required by ABBA to analyse chunks of different length (that span from 100 CpGs to 15,000 CpGs) on a single machine with 20 2.3GHz hyper-threaded cores and 32GB of RAM and found that the computational time (seconds) scales with the chunk length (N_{CpG}, number of CpG sites) following the power function: *time* (seconds) = $0.0045\ N_{\mathrm{CpG}}{}^{1.3985}$ (R^2 = 0.997). Scripts for the pre-processing step are embedded within ABBA at *abba.systems-genetics.net*.

WGBS of rat macrophages. Bone-marrow derived macrophages (BMDM) were isolated from WKY and LEW rat strains. WGBS libraries were produced as follows: 6µg of genomic DNA was spiked with 10ng of unmethylated cl857 Sam7 lambda DNA (Promega) and sheared using a Covaris System S-series model S2. Sheared DNA was purified and then end-repaired in a 100µl reaction using NEBNext End Repair kit (New England Biolabs) incubated at 20C for 30 minutes. End-repaired DNA was next A-tailed using NEBNext dA-tailing reaction buffer and Klenow Fragment (also New England Biolabs) incubated at 37C for 30 minutes and then purified with the MinElute PCR purification kit (Qiagen) in a total final elution volume of 28µl. Illumina Early Access Methylation adapter oligos (Illumina) were then ligated to a total of 25µl of the A-tailed DNA sample using NEBNext Quick Ligation Reaction Buffer and Quick T4 DNA ligase (both New England Biolabs) in a reaction volume of 50µl. This mixture was incubated for 30 minutes at 20C prior to gel purification. Bisulphite conversion of 450ng of the purified DNA library was achieved using the Epitect Bisulfite kit (Qiagen) in a total volume of 140µl. Samples were incubated with the following program: 95C for 5 minutes, 60C for 25 minutes, 95C for 5 minutes, 60C for 85 minutes, 95C for 5 minutes, 60C for 175 minutes and then 3x repeat of 95C for 5 minutes and 60C for 180 minutes and held at 20C. Treated samples were then purified as per manufacturers instructions. Adapter bound DNA fragments were amplified by a 10-cycle PCR reaction and then purified using Agencourt AMPure XP beads (Beckman Coulter) before gel extraction and quantification using the Agilent Bioanalyzer 2100 Expert High Sensitivity DNA Assay. Then, libraries were quantified using quantitative PCR and then denatured into single stranded fragments. These fragments were then amplified by the Illumina cluster robot and transferred to the HiSeq 2000 for sequencing.

Despite ABBA being able to detect methylation changes at all genomic locations we focused only on those methylation changes that occur at CpG sites, and considered CpG sites where at least 4 out of the 8 samples contain reads (resulting in a total of 14,976,632 CpG sites genome-wide in BMDM from WKY and LEW rats). DMRs were called with ABBA (see above) using a 5 CpG minimum, a 33% or greater difference in methylation and a 5% FDR threshold. Genomic region annotations and Ensembl gene IDs for the rat reference genome 4 (rn4) were downloaded from the UCSC genome browser. Significant over-representations of genomic features (intron, exons, etc.) were determined empirically from 1,000 randomly sampled GC-matched regions per DMR. The genes overlapping with DMRs were further annotated and tested for enrichment in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using WebGestalt[31].

Identification of enriched TFBS motifs within the DMRs identified by ABBA was performed using HOMER [32]. HOMER was used to scan for motifs obtained from the JASPAR 2014 database[33]. Thresholds for motif identification using a p-value of 10^{-4}. Enrichments were calculated by comparing the motifs present in the DMRs against a large set of background sequences ($N = 10^6$) corrected for CpG content.

RNA-seq and ChIP-seq analysis of rat macrophages. RNA-seq data from BMDM in WKY and LEW strains were retrieved from[34] and reanalyzed in the context of WGBS analysis reported here. Briefly, total RNA was extracted from BMDM at day 5 of differentiation in three WKY rats and three LEW rats using Trizol (Invitrogen). 1 µg of total RNA was used to generate RNA-seq libraries using TruSeq RNA sample preparation kit (Illumina, UK). Libraries were run on a single lane per sample of the HiSeq 2000 platform (Illumina) to generate 100 bp paired-end reads.

An average of 72 M reads coverage per sample was achieved (minimum 38 M). RNA-seq reads were aligned to the rn4 reference genome using tophat2. The average number of mapped was 67M (minimum 36M) corresponding to an average mapping percentage of 93%. Sequencing and mapping were quality controlled using the fastQC software. Gene-level read counts were computed using HT-Seq-count with 'union' mode and genes with less than 10 aligned reads across all samples were discarded prior to analysis leading to 15,155 genes. Differential gene expression analysis between WKY and LEW BMDMs was performed using DESeq2[35] and significantly differentially expressed genes were reported at the 5% FDR level. The visualizations of the expression levels with gene structure were created with DEXSeq[36].

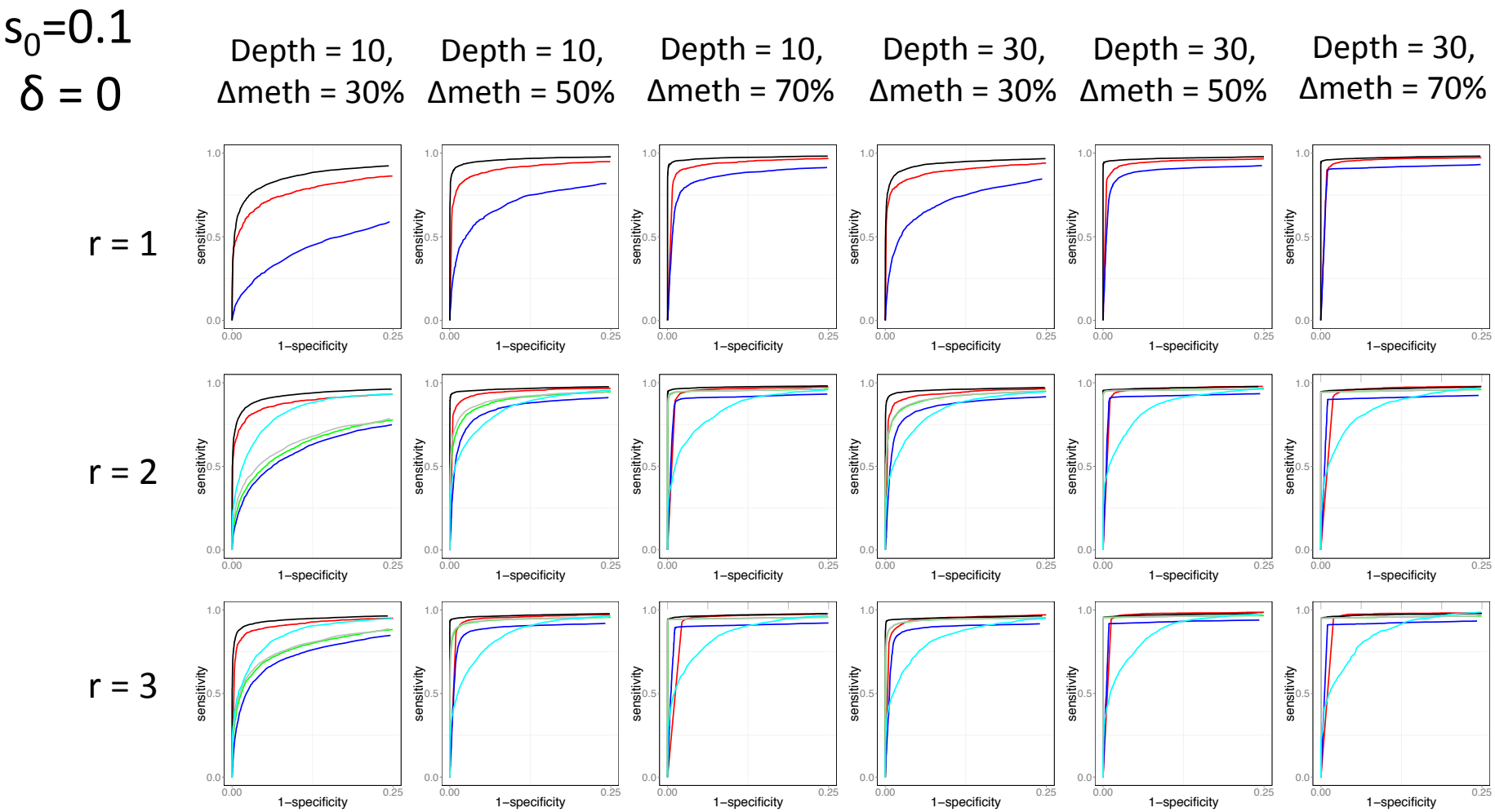
ChIP-seq data from BMDM isolated from the WKY and WKY.L*Crgn2* congenic strains (in which the LEW Crgn2 QTL was introgressed onto the WKY background) were retrieved from[37,38] and re-analyzed with respect to the *Ifitm3* locus. This congenic model (WKY.L*Crgn2*) has been extensively studied in previous studies where it has been shown that JunD expression levels are significantly higher in WKY when compared with the congenic[37] and that the canonical binding of AP-1 is significantly greater in WKY compared to WKY.L*Crgn2*[21]. Briefly, ChIP was performed with a JunD antibody (Santa Cruz sc74-X) and a negative IgG control (sc-2026). Single read library preparation and high throughput single read sequencing for 36 cycles was carried out on an Illumina Genome Analyser IIx and sequencing of the ChIP-Seq libraries was carried out on the high throughput Illumina Genome Analyzer II. Initial data processing was performed using Illumina Real Time Analysis (RTA) v1.6.32 software (equivalent to Illumina Consensus Assessment of Sequence and Variation, CASAVA 1.6) using default settings. Quality filtered reads were then realigned to the rn4 using the Burrows Wheeler Alignment tool v0.5.9 (BWA). Read ends were trimmed if Phred-scaled base quality scores dropped below 20. Significant ChIP-seq differences in the DMR of the *Ifitm3* gene promoter were called using a Fisher's exact test with a sliding window of 50bp.

Software availability and implementation. ABBA is implemented as a Perl/R program, which is available with instructions for download at *abba.systems-genetics.net.*

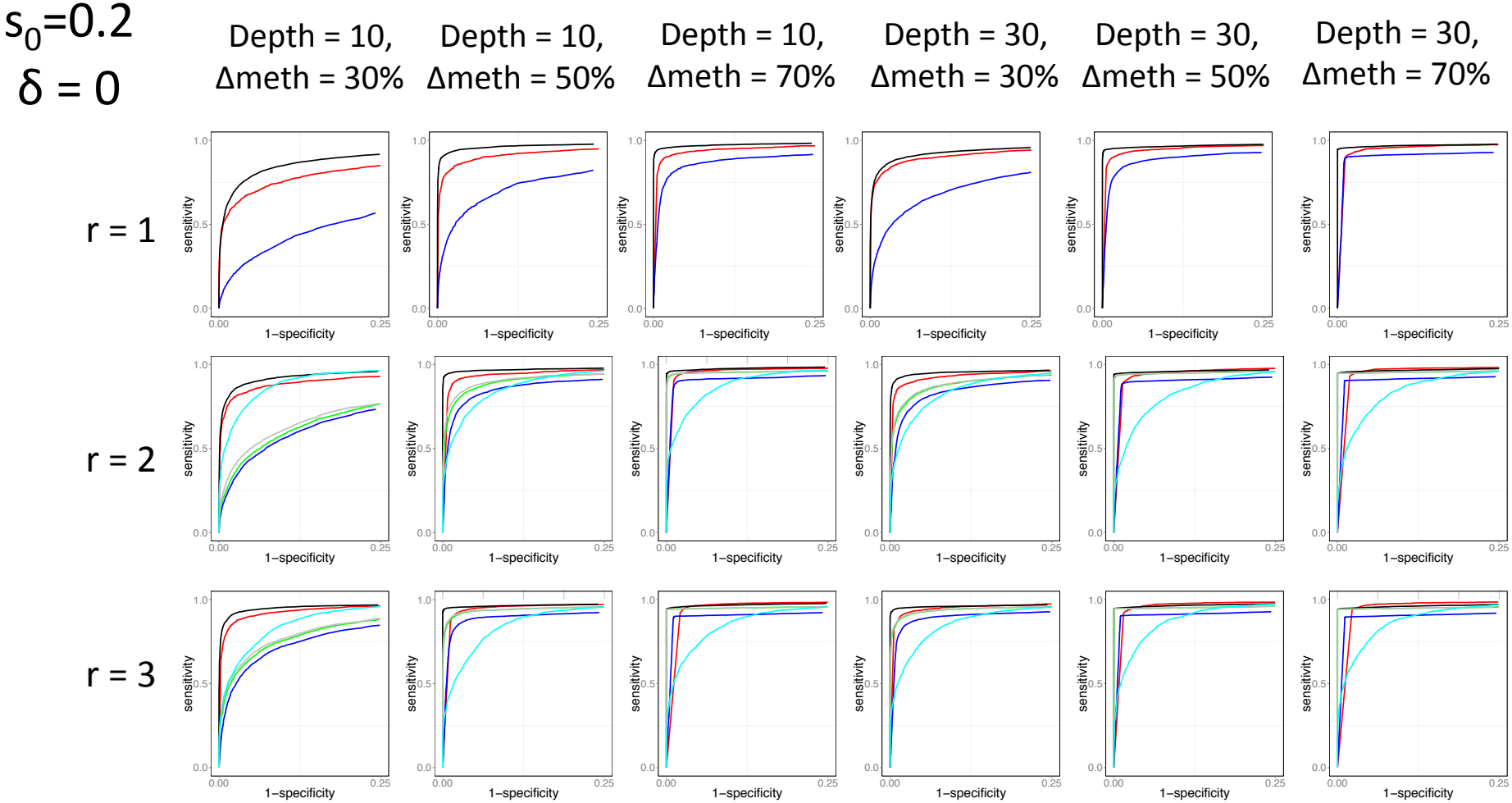
Additional References

24. Rue, H. & Held, L. *Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall/CRC press* (2005).
25. Rue, Hå. & Martino, S. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *J. Stat. Plan. Inference* **137,** 3177–3192 (2007).
26. Tierney, L. & Kadane, J. B. Accurate Approximations for Posterior Moments and Marginal Densities. *J. Am. Stat. Assoc.* (2012).
27. Lindgren, F., Rue, H. & Lindström, J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B* **73,** 423–498 (2011).
28. Dempster, A. P. A., Laird, N. M. N. & Rubin, D. D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **39,** 1–38 (1977).
29. Muller, P., Parmigiani, G. & Rice, K. FDR and Bayesian multiple comparisons rules. *Bayesian Stat. 8* 349–370 (2006).
30. Zhang, W., Spector, T. D., Deloukas, P., Bell, J. T. & Engelhardt, B. E. Predicting genome-wide DNA methylation using methylation marks, genomic position, and DNA regulatory elements. *Genome Biol.* **16,** 14 (2015).
31. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41,** W77–83 (2013).
32. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–89 (2010).
33. Mathelier, A. *et al.* JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **42,** D142–7 (2014).
34. Rotival, M. *et al.* Integrating phosphoproteome and transcriptome reveals new determinants of macrophage multinucleation. *Mol. Cell. Proteomics* **14,** 484–98 (2015).
35. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
36. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22,** 2008–17 (2012).
37. Hull, R. P. *et al.* Combined ChIP-Seq and transcriptome analysis identifies AP-1/JunD as a primary regulator of oxidative stress and IL-1β synthesis in macrophages. *BMC Genomics* **14,** 92 (2013).
38. Srivastava, P. K., Hull, R. P., Behmoaras, J., Petretto, E. & Aitman, T. J. JunD/AP1 regulatory network analysis during macrophage activation in a rat model of crescentic glomerulonephritis. *BMC Syst. Biol.* **7,** 93 (2013).

Supplementary Figure 1a



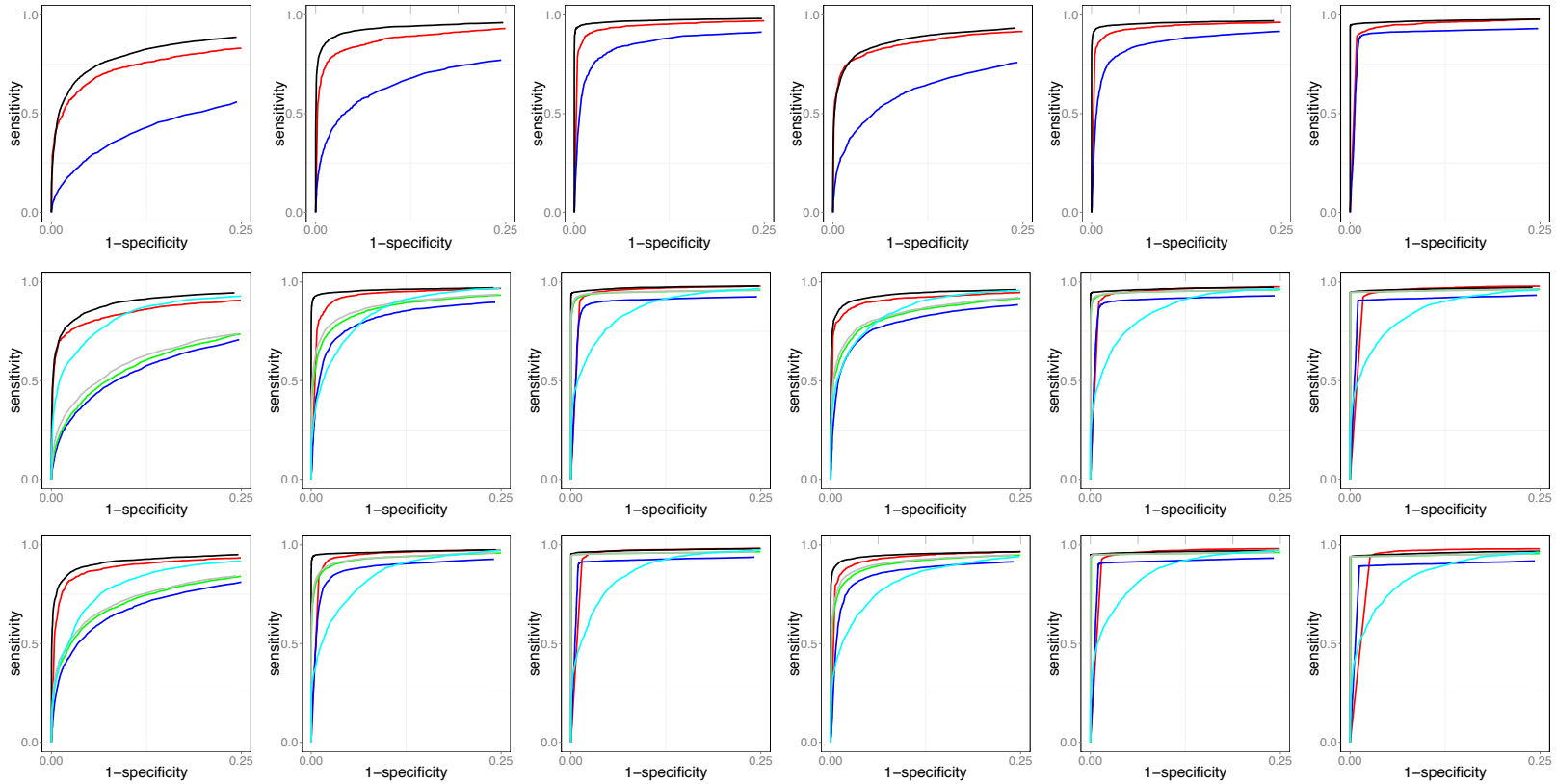
Supplementary Figure 1b



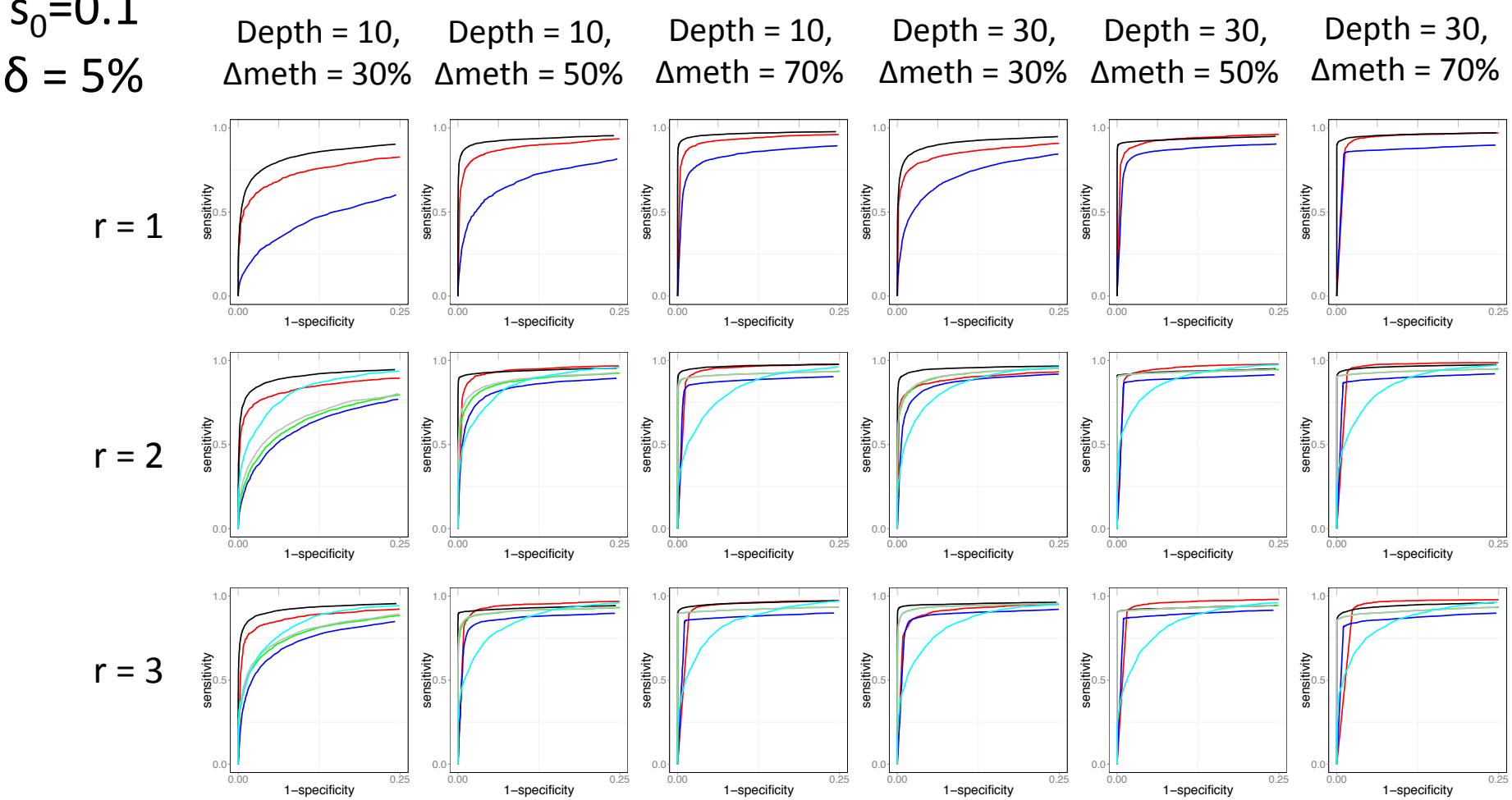
Supplementary Figure 1c

s_0=0.3
δ = 0

	Depth = 10, Δmeth = 30%	Depth = 10, Δmeth = 50%	Depth = 10, Δmeth = 70%	Depth = 30, Δmeth = 30%	Depth = 30, Δmeth = 50%	Depth = 30, Δmeth = 70%
r = 1						
r = 2						
r = 3						

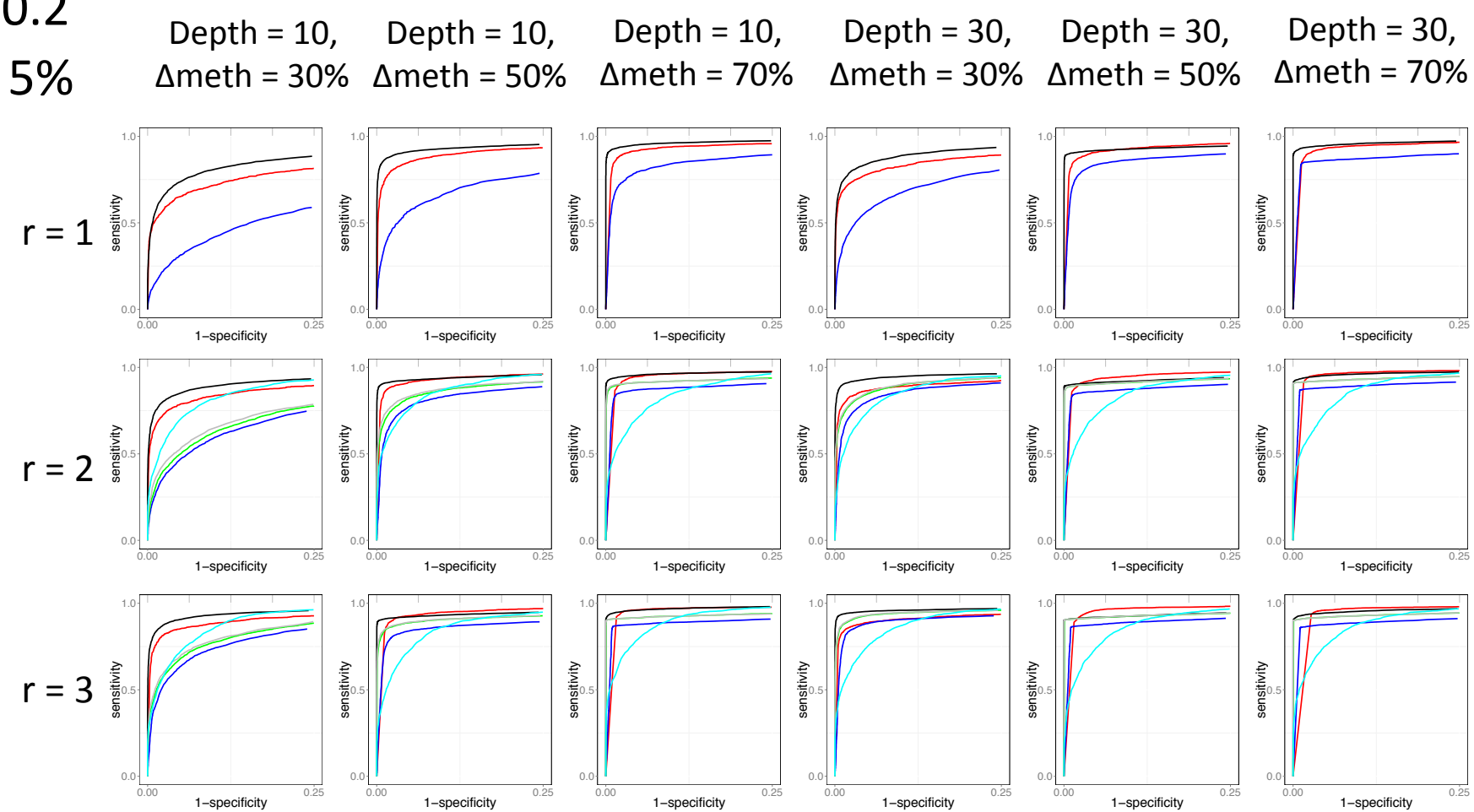


Supplementary Figure 2a

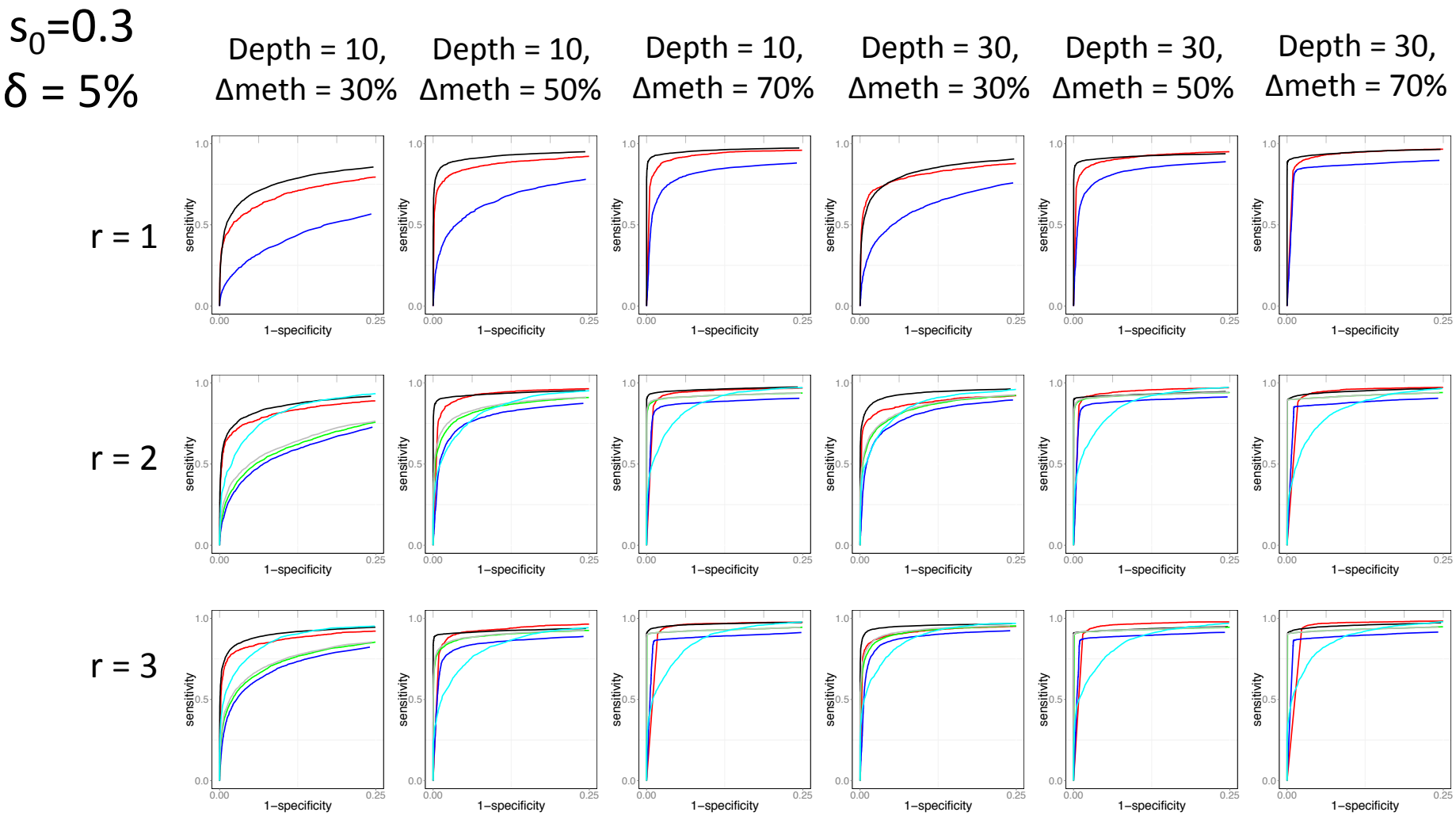


Supplementary Figure 2b

s_0=0.2
δ = 5%



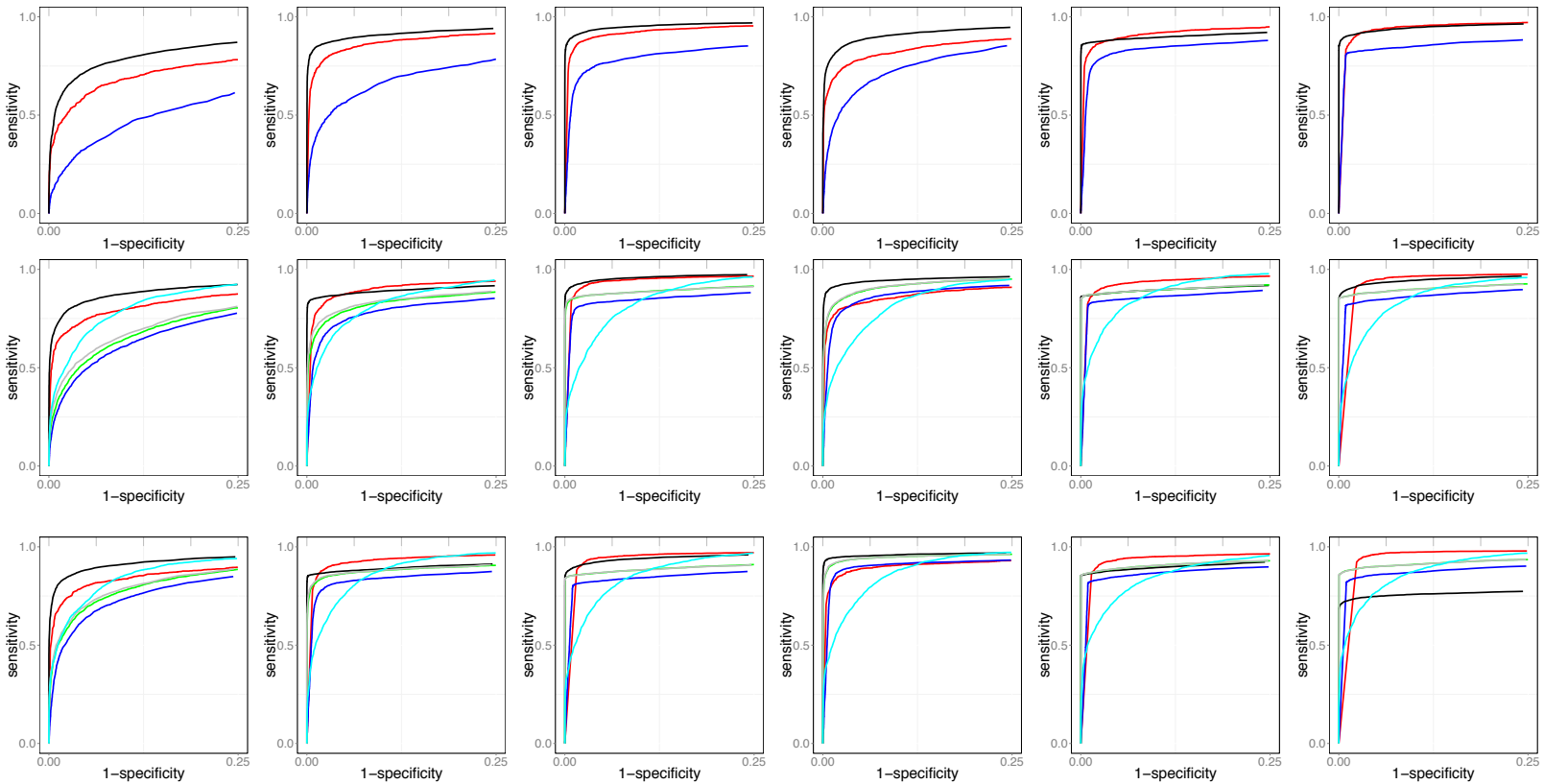
Supplementary Figure 2c



Supplementary Figure 3a

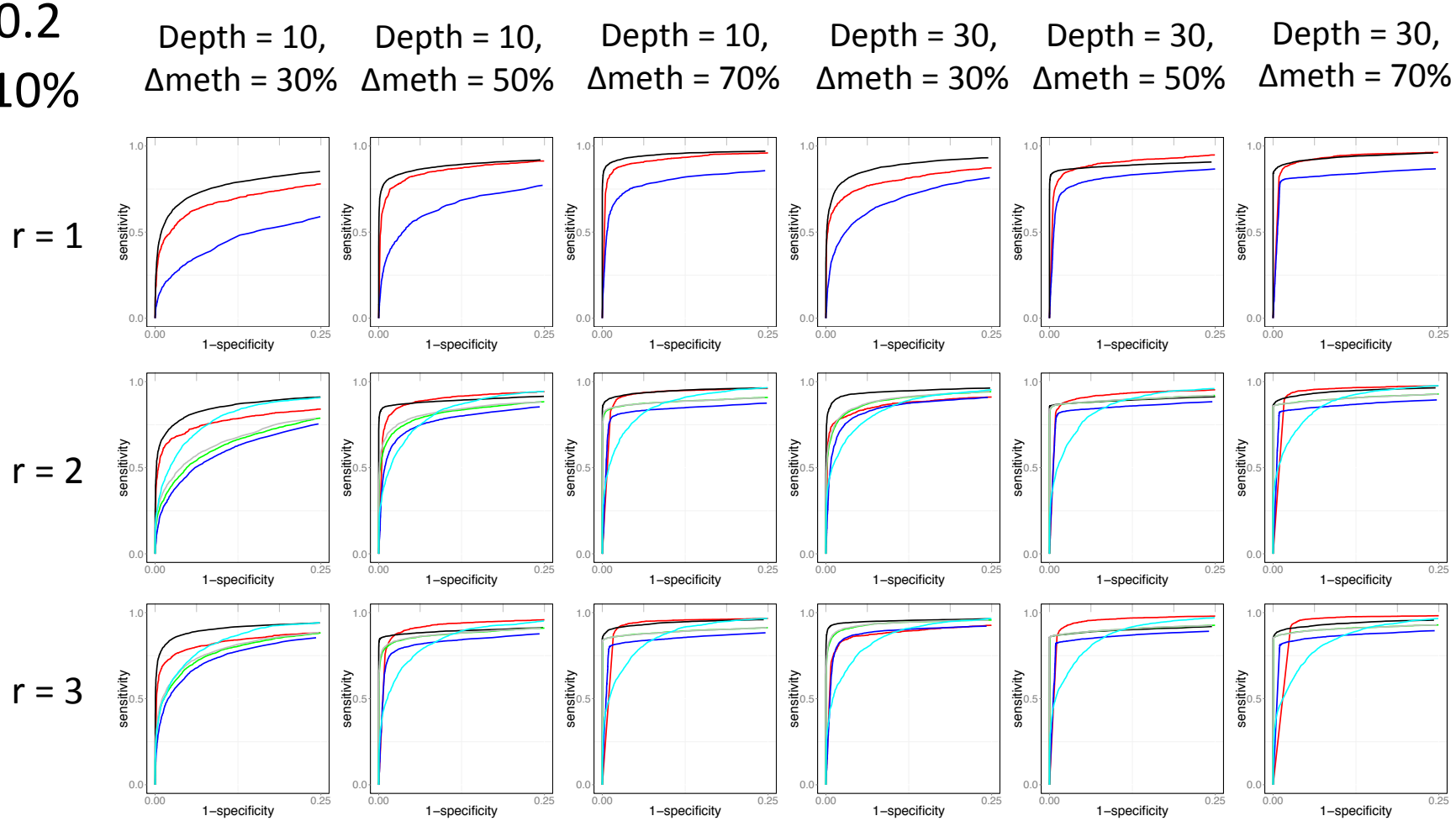
s_0=0.1
δ = 10%

	Depth = 10, Δmeth = 30%	Depth = 10, Δmeth = 50%	Depth = 10, Δmeth = 70%	Depth = 30, Δmeth = 30%	Depth = 30, Δmeth = 50%	Depth = 30, Δmeth = 70%
r = 1						
r = 2						
r = 3						



Supplementary Figure 3b

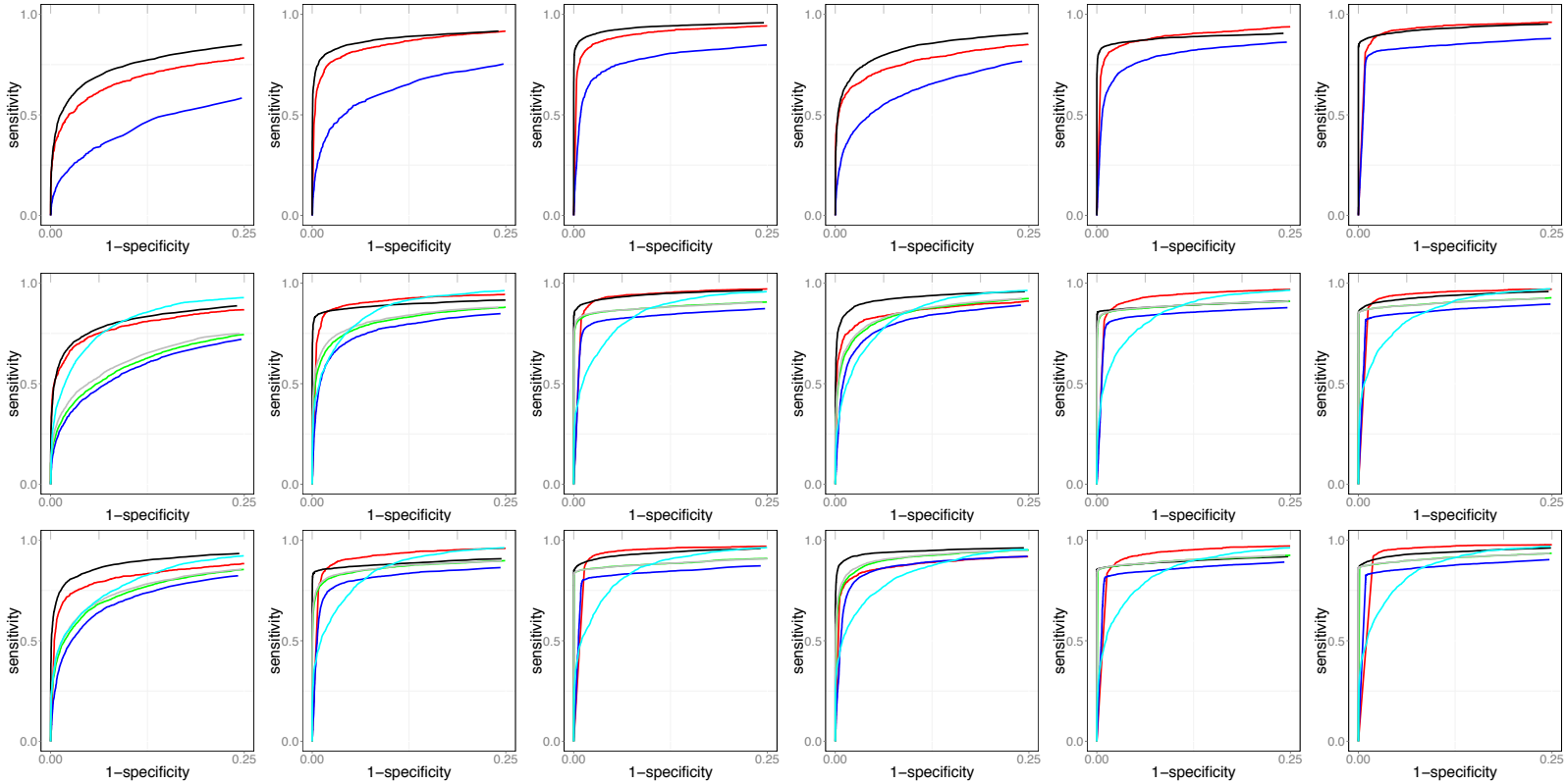
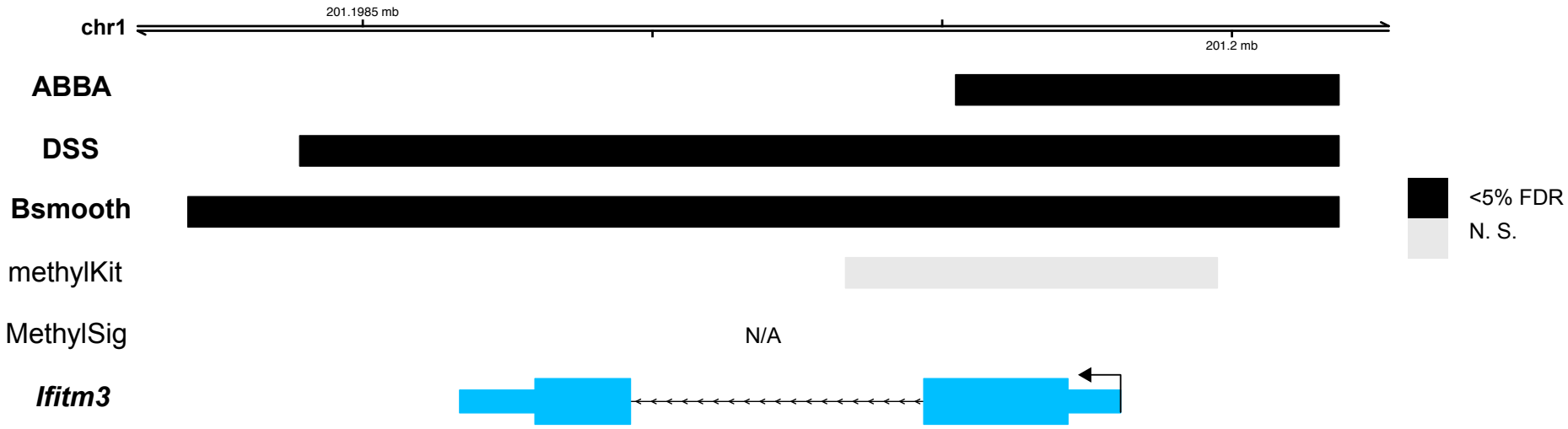
s_0=0.2
δ = 10%



Supplementary Figure 3c

s_0=0.3
δ = 10%

	Depth = 10, Δmeth = 30%	Depth = 10, Δmeth = 50%	Depth = 10, Δmeth = 70%	Depth = 30, Δmeth = 30%	Depth = 30, Δmeth = 50%	Depth = 30, Δmeth = 70%
r = 1						
r = 2						
r = 3						



Supplementary Figure 4. Schematic representation of the DMR identified by each method at the *Ifitm3* locus. WGBS of bone marrow derived macrophages from WKY and LEW rats were independently analyzed using ABBA, DSS, BSmooth, MethylKit and MthylSig. For each method we report the location of the DMR identified at 5% FDR level. Only ABBA, DSS and BSmooth identified differential methylation at the *Ifitm3* locus. MthylKit identified a signal for differential methylation which was not genome-wide significant. Amongst the methods that detected significant differential methylation at this locus, only ABBA identified differential methylation specifically overlapping with *Ifitm3* promoter where the TFBS and differential ChIP-seq signals were also identified (see Figure 2g). In contrast, DSS and BSmooth reported a large and rather unspecific region with differential methylation between WKY and LEW macrophages.