Quantitative modelling predicts the impact of DNA methylation on RNA polymerase II traffic

Citation for published version:

Digital Object Identifier (DOI):
10.1073/pnas.1903549116

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Proceedings of the National Academy of Sciences

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Patterns of gene expression are primarily determined by proteins that locally enhance or repress transcription. While many transcription factors target a restricted number of genes, others appear to modulate transcription levels globally. An example is MeCP2, an abundant methylated-DNA binding protein that is mutated in the neurological disorder Rett Syndrome. Despite much research, the molecular mechanism by which MeCP2 regulates gene expression is not fully resolved. Here we integrate quantitative, multi-dimensional experimental analysis and mathematical modelling to indicate that MeCP2 is a novel type of global transcriptional regulator whose binding to DNA creates "slow sites" in gene bodies. We hypothesise that waves of slowed-down RNA polymerase II formed behind these sites travel backward and indirectly affect initiation, reminiscent of defect-induced shock waves in non-equilibrium physics transport models. This mechanism differs from conventional gene regulation mechanisms, which often involve direct modulation of transcription initiation. Our findings point to a genome-wide function of DNA methylation that may account for the reversibility of Rett syndrome in mice. Moreover, our combined theoretical and experimental approach provides a general method for understanding how global gene expression patterns are choreographed.

MeCP2 | Gene regulation | Mathematical modelling | DNA methylation

Introduction

Many eukaryotic chromatin-associated factors modulate transcription by binding to specific sites in gene promoters or enhancers (1, 2). Most transcription factors are thought to modulate the initiation rate of transcription by altering histone-DNA interactions (2, 3) or imposing promoter-proximal obstacles (4). However, transcription can also be affected by processes that occur in the bodies of genes. In particular, DNA methylation, which is widespread in gene bodies, appears to affect progression of RNA polymerase II (RNA Pol II) through densely methylated exons (5). The mechanism is unclear, but methyl-CpG binding proteins (6) may be involved. Since most gene bodies contain methylated CpGs, such proteins may have a global effect on transcription.

One putative global modulator is methyl-CpG binding protein 2 (MeCP2) (7, 8), which is highly expressed in neurons. MECP2 mutations, including loss-of-function or gene duplication, lead to severe neurological disorders (9, 10). MeCP2 does not behave as a conventional transcription factor with discrete targets, as its binding site occurs on average every ~100 base pairs. Evidence from in vitro systems (11, 12) and mouse models (13, 14) suggests that MeCP2 can mediate DNA methylation-dependent transcriptional inhibition. Transcriptional changes in mouse brain when MeCP2 is absent or over-expressed are relatively subtle but widespread (15-17), and the molecular mechanisms underlying these changes are unknown.

Here we set out to resolve the mechanism of MeCP2-dependent transcriptional regulation. Because MeCP2 binding sites occur in the vast majority of genes, we reasoned that most are likely to be influenced to some extent by its presence. To confront the technical and analytical challenges posed by modest changes in expression of large numbers of genes, we adopted a quantitative approach that combined deep, high quality datasets obtained from a uniform population of Lund Human Mesencephalic (LUHMES)-derived human dopaminergic neurons (18) with computational modelling. We created a spectrum of LUHMES cell lines expressing distinct levels of MeCP2. Using transposase-accessible chromatin sequencing (ATAC-seq) and chromatin immunoprecipitation (ChIP-seq) together with mathematical modelling, we detected a robust footprint of MeCP2 binding to mCG and mCA in vitro and determined the amount of MeCP2 bound to DNA. Quantification of mRNA abundance by RNA-seq revealed a relationship between changes in transcription and the density of mCG on gene bodies. To explain this observation, we proposed and tested several distinct mechanistic models. The only model consistent with our experimental results is one in which MeCP2 leads to slowing down of RNA polymerase II progression through a transcription unit. Importantly, mutant MeCP2 that is unable to bind the TBL1/TBLR1 subunits of the NCoR co-repressor complex fails to repress efficiently, suggesting that repression depends upon this interaction.

Results

Global changes in transcription correlate with MeCP2 expression level. We created progenitor cell lines capable of differentiation to a uniform population of human neurons (SI Appendix, Fig. S1A-C) that expressed seven widely different levels of MeCP2, including knock-out (KO), wild-type (WT) and 11-
fold over-expression (11x) (Fig. 1A,B and SI Appendix, Fig. S1D and Table S1). All lines differentiated into neurons with similar kinetics, expressed neuronal markers (SI Appendix, Fig. S1E), and had identical global levels of DNA methylation (~3.7% of all cytosines were methylated) (SI Appendix, Fig. S2A). Based on the known affinity of MeCP2 for methylated CG (mCG), we expected that the effect of MeCP2 on gene expression would depend on their mCG content. DNA methylation was therefore quantified for all genes in WT neurons using whole-genome bisulphite sequencing (TAB-seq) (SI Appendix, Fig. S2B,C). We calculated total methylation (total mCG, \( N_{mCG} \)) as the number of methylated CG dinucleotides, mCG density (\( n_{mCG} \)) as the number of mCGs per 100 bp, and mCG mean as the percentage of mCG in all CG dinucleotides. To determine the effects of MeCP2 on transcription, we performed RNA-seq on all seven cell lines. We included all expressed protein-coding genes (~17500 genes) in our analysis. Most genes responded to MeCP2 but changes were small, precluding definition of a subset of affected genes (SI Appendix, Fig. S3A). To enhance a possible relationship between expression changes and DNA methylation that otherwise might be obscured by other regulatory mechanisms and statistical noise, genes were binned according to methylation density, considering gene bodies and promoters separately.

The average change in expression versus appropriate controls (Log2FC) showed a strong relationship to mCG density (\( \rho_{mCG} \)) in gene bodies (Fig. 1C). The effect was the strongest for \( \rho_{mCG} = 0.8-4.0 \) mCG per 100bp which includes the vast majority of genes (SI Appendix, Fig. S3B). The apparent stimulation of expression at very low mCG densities in OE neurons is discussed in SI Appendix. Moreover, the maximum slope of the Log2FC versus \( \rho_{mCG} \) in gene bodies (Fig. 1C, black lines) was strikingly proportional to MeCP2 abundance. (E) Ratio between luciferase expressions from an unmethylated and gene-body methylated constructs, for three cases: no MeCP2, WT MeCP2, and an MBD mutant R111G that is unable to bind mCG. Points show individual replicates. In all panels, error bars represent +/- SEM.
MeCP2 binds predominantly methylated CG genome-wide. To map the binding of MeCP2 in human neurons, we performed MeCP2 ChIP-seq for KO, WT, OE 4x and OE 11x, and simultaneously developed a computer model that simulates the ChIP-seq procedure and MeCP2 binding in vivo (Fig. 2A). As expected, ChIP enrichment was proportional to the level of MeCP2 in each cell line (SI Appendix, Fig. S6a-C) and showed a strong peak centred at mCGs in MeCP2-positive lines (Fig. 2B) as well as a correlation between MeCP2 enrichment and mCG density (Fig. 2C). Conversely, enrichment was absent at non-methylated CGs (SI Appendix, Fig. S6E).

To derive an independent measure of absolute MeCP2 density on the DNA and to detect its molecular footprint with high resolution, we performed ATAC-seq (19) in which transposase Tn5 cuts exposed DNA to reveal DNA accessibility within chromatin (Fig. 2A). In agreement with the ChIP-seq data, ATAC-seq Tn5 insertion profiles (Fig. 2D) showed a graded depletion of insertion sites centered around mCG in WT, OE 4x and OE 11x neurons, whose amplitude was proportional to MeCP2 concentration (Fig. 2E) and therefore represents a "molecular footprint" of MeCP2 binding in vivo. The size and amplitude of the footprint agrees well with a computer model of ATAC-seq and MeCP2 binding (Fig. 2D, black lines) and previous in vitro data (20, 21), confirming that MeCP2 occupies 11 bp of DNA in living cells. No depletion of insertion sites was observed over unmethylated CGs (SI Appendix, Fig. S6F). The model revealed that only 6.5% of mCG sites are actually occupied by MeCP2 in OE 11x neurons, falling to less than 1% occupancy in WT (Fig. 2E), perhaps due in part to occlusion by nucleosomes. Excellent agreement between the models and ATAC-seq and ChIP-seq data allows us to predict MeCP2 occupancy from mCG density and MeCP2 level in each cell line (Fig. 2E and SI Appendix, Fig. S6D).

MeCP2 does not regulate transcription via condensation of chromatin or premature termination. To interpret these results mechanistically, we considered mathematical models based on a commonly accepted paradigm for gene expression (SI Appendix, Fig. S7A) (22). In the first class of models named Condensation models (Fig. 3A), MeCP2 affects the rate of transcription initiation via changes in chromatin structure. The possibility that MeCP2 affects the initiation rate by binding to promoters was rejected because it would imply a stronger correlation between gene expression and MeCP2 binding than for KO. Moreover, Queues of Pol II induced by MeCP2 can reach TSS (red dot) and block initiation if both the initiation rate (a) and the density of MeCP2 (p) are sufficiently high (left panel). (D) Schematic representation of Pol II (grey) density shock waves forming behind MeCP2 (blue). Black line is the local density of Pol II. (E) Log2FC (gene expression) versus mCG density in gene bodies obtained in computer simulations of the Dynamical obstacles model (black solid lines) fitted to the OE 11x/OE ctr RNA-seq dataset (red) agrees well with experimental data for OE 4x/OE ctr (orange) and KO/ctr (purple) datasets. Error bars represent ± SEM. (F) The maximum slope of Log2FC (gene expression) versus mCG density in gene bodies, predicted by the Dynamical obstacles model (black line). Points are experimental slopes from Fig. 1C.

Fig. 3. MeCP2 does not regulate transcription via condensation of chromatin or premature termination. (A) A cartoon of the Condensation model. Tangles represent regions of condensed chromatin that are inaccessible to RNA Pol II. (B) Chromatin accessibility (measured by ATAC-seq) at promoters rapidly decreases with increasing promoter methylation. In contrast, MeCP2 has a minor effect on accessibility (curves for 0x, 4x and 11x are slightly lower than for KO). (C) The Condensation model disagrees with Log2FC (OE 11x/KO) obtained from RNA-seq. (D) Schematic representation of the Detachment model. (E) Log2FC (gene expression) for KO/ctr (purple) versus the total number of mCGs per gene. Black lines represent predictions of Detachment model. Error bars represent ± SEM. (F) As (E) for OE 11x/ctr red.

Fig. 4. Mathematical modelling indicates that MeCP2 slows down transcriptional elongation. (A) Schematic representation of the Dynamical obstacles model. (B) Transcription rate predicted by the model, plotted as a function of the initiation rate , for different mean MeCP2 densities in gene bodies. (C) Space-time plots (kymographs) representing Pol II moving along the gene. Queues of Pol II induced by MeCP2 can reach TSS (red dot) and block initiation if both the initiation rate (a) and the density of MeCP2 (p) are sufficiently high (left panel). (D) Schematic representation of Pol II (grey) density shock waves forming behind MeCP2 (blue). Black line is the local density of Pol II. (E) Log2FC (gene expression) versus mCG density in gene bodies obtained in computer simulations of the Dynamical obstacles model (black solid lines) fitted to the OE 11x/OE ctr RNA-seq dataset (red) agrees well with experimental data for OE 4x/OE ctr (orange) and KO/ctr (purple) datasets. Error bars represent ± SEM. (F) The maximum slope of Log2FC (gene expression) versus mCG density in gene bodies, predicted by the Dynamical obstacles model (black line). Points are experimental slopes from Fig. 1C.
Finally, we considered a theoretical model: methyl-CpG binding domain (MBD) and NCoR-interaction domain (NID). The mutation R111G causes MeCP2 to lose the ability to bind specifically to mC. The mutation R306C prevents MeCP2 from binding the NCoR complex. (B) Level of MeCP2 (Western blot) in two overexpressed mutant cell lines (R111G and R306C) and the overexpression control cell line (OE ctrl). OE 11x is shown for comparison. Values are averaged over three biological replicates and normalised by the level of histone H3. (C) Log2FC (expression) of OE R111G/OE ctrl shows almost no dependence on mCG density in gene bodies (black). Grey line shows the maximum slope. (D) Log2FC (expression) of OE R306C/OE ctrl shows a small negative correlation with gene body mCG density (brown). Grey line shows the maximum slope. (E) Maximum slopes for all cell lines including OE R111G (black) and OE R306C (brown) from (C) and (D) versus MeCP2 level (Western blot). In all plots error bars represent ±SEM.

Fig. 5. MeCP2 slows down transcription via a mechanism involving NCoR. (A) Location of two binding domains in MeCP2 that are relevant for the proposed mechanism: methyl-CpG binding domain (MBD) and NCoR-interaction domain (NID). The mutation R111G causes MeCP2 to lose the ability to bind specifically to mC. The mutation R306C prevents MeCP2 from binding the NCoR complex. (B) Level of MeCP2 (Western blot) in two overexpressed mutant cell lines (R111G and R306C) and the overexpression control cell line (OE ctrl). OE 11x is shown for comparison. Values are averaged over three biological replicates and normalised by the level of histone H3. (C) Log2FC (expression) of OE R111G/OE ctrl shows almost no dependence on mCG density in gene bodies (black). Grey line shows the maximum slope. (D) Log2FC (expression) of OE R306C/OE ctrl shows a small negative correlation with gene body mCG density (brown). Grey line shows the maximum slope. (E) Maximum slopes for all cell lines including OE R111G (black) and OE R306C (brown) from (C) and (D) versus MeCP2 level (Western blot). In all plots error bars represent ±SEM.

Finally, we considered a theoretical model: methyl-CpG binding domain (MBD) and NCoR-interaction domain (NID). The mutation R111G causes MeCP2 to lose the ability to bind specifically to mC. The mutation R306C prevents MeCP2 from binding the NCoR complex. (B) Level of MeCP2 (Western blot) in two overexpressed mutant cell lines (R111G and R306C) and the overexpression control cell line (OE ctrl). OE 11x is shown for comparison. Values are averaged over three biological replicates and normalised by the level of histone H3. (C) Log2FC (expression) of OE R111G/OE ctrl shows almost no dependence on mCG density in gene bodies (black). Grey line shows the maximum slope. (D) Log2FC (expression) of OE R306C/OE ctrl shows a small negative correlation with gene body mCG density (brown). Grey line shows the maximum slope. (E) Maximum slopes for all cell lines including OE R111G (black) and OE R306C (brown) from (C) and (D) versus MeCP2 level (Western blot). In all plots error bars represent ±SEM.

We next considered potential effects of MeCP2 on the elongation phase of transcription. The Detachment Model posits that MeCP2 causes transcription to prematurely abort (Fig. 3D). Since the probability of termination increases with each blocking site, under this model the Log2FC is a function of the total number of methylated CGs ($N_{mCG}$) in the gene: $\text{Log2FC} = \log(1 - N_{mCG} / N_{ac})$ where $N$ is MeCP2 concentration relative to WT, and the parameter $\gamma$ is proportional to the probability that Pol II aborts transcription when it encounters MeCP2 or an MeCP2-induced chromatin modification. The unknown parameter $\gamma$ can be obtained by fitting the model to the Log2FC (KO/WT) data (Fig. 3E). We found that the model failed to reproduce the Log2FC vs $N_{mCG}$ relationship for the OE 11x cell line (Fig. 3F). The model also fails to correctly predict the observed relationship between Log2FC and mCG density in gene bodies (SI Appendix, Fig. S7B,C). Therefore, it is unlikely that MeCP2 affects transcription via premature termination.

MeCP2 creates 'Dynamical obstacles' that impede transcriptional elongation. Finally, we considered a "Congestion model" whereby Pol II pauses when it encounters MeCP2 itself or an induced, transient structural modification of chromatin (Fig. 4A). The parameters are: the fraction $\gamma$ of mCGs bound by MeCP2, MeCP2 turnover (unbinding) rate $k_u$, and (specific to each gene) the length $l$ of the gene, the density $N_{mCG}$ of methylated CGs, and the initiation rate $\sigma$. Fig. 4B shows the transcription rate for OE 11x predicted by the model as a function of $\sigma$, for different mean MeCP2 densities ($N_{mCG}$). The assumed value of $k_u = 0.04$ s$^{-1}$ is compatible with the reported in vivo residence time of MeCP2 on chromatin (25-40s (23)). Inspired by non-equilibrium statistical mechanics approaches that have been utilised to model one-dimensional transport (24, 25), we expect a non-equilibrium phase transition from a low-density to a maximal-current (congested) phase as the initiation rate or the density of obstacles increase beyond a critical point. Indeed, all curves in Fig. 4B have a characteristic shape: a linear relationship $j \propto \sigma$ for small $\sigma$, followed by saturation at high initiation rates. Saturation occurs due to congestion as polymerases queue upstream of obstacles (Movies S1,2). However, even in the non-saturated regime of intermediate $\sigma$, excluded-volume interactions between polymerases that have been slowed down by an obstacle cause a density shock wave that propagates backwards (Fig. 4C). A small increase in the density of polymerases near the promoter decreases the rate of Pol II binding to the TSS. Thus, even though MeCP2 does not directly affect Pol II initiation, it does so indirectly by shock waves that form behind MeCP2-induced obstacles in gene bodies (Fig. 4D). To test the model against RNA-seq data, we estimated average initiation rates for genes with similar mCG densities by fitting the model to Log2FC data from one of the cell lines (OE 11x/OE ctrl; Fig. 4E left and SI Appendix, Fig. S8F). We then used the model to predict Log2FC for the remaining 6 cell lines. The model strikingly reproduces the data (Fig. 4E for OE 4x and KO) as well as the slopes of the Log2FC plots for all seven cell lines (Fig. 4F). A similar behaviour occurs in a modified model in which Pol II slows down (rather than completely stops) on permanent or long-lasting structural modifications of chromatin (SI Appendix, Fig. S8A-E, Movie S3). We conclude that both congestion models are compatible with the experimental data presented in Fig. 1C and D. The models also predict that Log2FC should decrease with increasing expression (measured as TPMs) in agreement with the data (SI Appendix, Fig. S8G).

MeCP2 binding to both DNA and NCoR are essential to slow down RNA Pol II. To address the question of whether MeCP2 impedes Pol II progression directly by steric interference or indirectly by altering chromatin structure (e.g., by histone deacetylation (26)), we overexpressed mutated forms of MeCP2 in the presence of WT MeCP2. The mutants were either unable to bind methylated DNA (R111G) (27) or unable to recruit the histone deacetylase complex NCoR (R306C) (14, 28) (Figs. 5A and SI Appendix, Fig. S9A). As expected, 7-fold overexpression of MeCP2-R111G caused no mCG-density dependent transcriptional changes (Figs. 5B,C and SI Appendix, Fig. S9B,C). The R306C mutant, on the other hand, was predicted to repress genes with similar mCG densities by fitting the model to Log2FC data from one of the cell lines (OE 11x/OE ctrl; Fig. 4E left and SI Appendix, Fig. S8F). We then used the model to predict Log2FC for the remaining 6 cell lines. The model strikingly reproduces the data (Fig. 4E for OE 4x and KO) as well as the slopes of the Log2FC plots for all seven cell lines (Fig. 4F). A similar behaviour occurs in a modified model in which Pol II slows down (rather than completely stops) on permanent or long-lasting structural modifications of chromatin (SI Appendix, Fig. S8A-E, Movie S3). We conclude that both congestion models are compatible with the experimental data presented in Fig. 1C and D. The models also predict that Log2FC should decrease with increasing expression (measured as TPMs) in agreement with the data (SI Appendix, Fig. S8G).

Concluding remarks

In summary, a close alliance between mathematical modelling and molecular biology has allowed us to discriminate molecular mechanisms underlying the relatively subtle global effects of MeCP2 on global gene expression. The proposed mechanism relies on MeCP2-NCoR interaction that slows down the progression of Pol II during transcription elongation. A candidate mediator of this effect is histone modification, in particular histone deacetylation, as cell transfection assays using methylated reporter gene data illustrate, and repression depends upon histone deacetylation activity (11, 12). According to this scenario, MeCP2...
recruitment of the histone deacetylase corepressor NCoR would
restrain transcription, perhaps by causing tighter binding of nucleo-
osomes to DNA (26). To explain the dramatic reversibility of Rett
syndrome in animal models (29) we propose that protec-
tion against the absence of MeCP2, DNA methylation patterns are unaffected, allowing
the re-expressed wildtype protein to bind within gene bodies and
commence normal modulation of transcriptional elongation. We
suggest that the Congestion model may apply to proteins other
than MeCP2. For example, other chromatin-binding factors that
bind short (and thus abundant) motifs, including other methyl-
binding proteins, may modulate gene expression by a similar
mechanism.

Materials and methods
Cell lines. The procedure for culture and differentiation of the LUHMES cell
line was previously described (18). To create two independent MeCP2 knock-
out lines, we used CRISPR-mediated gene disruption (30). To generate MeCP2
knock-downs, several shRNAs against MeCP2 were designed with a Chemene
Mission shRNA online software. Two shRNAs were chosen and cloned
into pLKO.1 vector including scrambled shRNA as a control and lentiviruses
were created (SI Appendix, Table S2). To increase the level of MeCP2 we
created lentiviruses expressing MeCP2 from two alternative promoters in the
pLKO.1 vector: Synapsin and cytomegalovirus (CMV). Calculation of standard
development, standard error of mean and t-tests for qPCR, Western blots,
methylation and total RNA quantification using HPLC were performed using
GraphPad Prism version 7.0.

Repression assay. CPG-free vector containing Firefly Luciferase with
CpGs was methylated by M.SssI methyltransferase in presence or absence
of SAM. Mouse embryonic fibroblasts were transfected using Lipofectamine
2000 with three plasmids containing: Firefly Luciferase, Renilla Luciferase
and MeCP2. Luciferase activity measurements were performed using Dual
Luciferase assay kit (Promega) according to manufacturer protocol.

Library preparation for Illumina sequencing. All libraries were se-
quenced as 75- or 100-nucleotide long paired-end reads on HiSeq 2000
and HiSeq 2500 Illumina platforms. Methylome of wildtype LUHMES-derived
neurons at day 9 was obtained by TAB-seq according to the published
protocol (31). RNA-seq library was performed according to manufacturer
protocol for ScriptSeq Complete Gold kit (Human/Mouse/Rat). Total RNA was
isolated from all generated cell lines (SI Appendix, Table S1) at day 9
differentiation using either the RNeasy Mini kit and the AllPrep DNA/RNA
Mini kit (Qiagen). ATAC-seq in four cell lines (KO, WT, OE 4x and OE 11x, SI
Appendix, Table S1) was performed as in (32).

MeCP2 ChIP-seq was performed using LUHMES-derived neurons at day
9 of differentiation with four levels of MeCP2. KO, WT, OE 4x and OE 11x (SI
Appendix, Table S1). Libraries were prepared using the NEBNext Ultra II DNA
Library Prep Kit for Illumina (NEB) according to manufacturer's instructions.

Data processing of raw reads from Illumina sequencing. All reads were
quality-controlled, trimmed to remove adapters (Trimmomatic) (33),
and duplicated reads, and mapped to the human hg19 reference genome.
Bismark (34) was used to extract cytosine methylation from TAB-seq. All raw
reads were deposited in GEO database (accession number GSE125660).

RNA-seq data analysis. We used a 4-let subset of protein-coding genes with
sufficient methylation coverage (BS-seq; Bismark (34) was used to extract cytosine methylation from TAB-seq. All raw reads were deposited in GEO database (accession number GSE125660).

The model predicts that
\[
\frac{N_{\text{meCP2}}(t)}{N_{\text{total}}(t)} = \left(1 - \frac{P(t)}{P_{\text{max}}(t)}\right)^{r(t)}
\]
where \(P(t)\) and \(P_{\text{max}}(t)\) are the insertion counts profiles for a given cell line and KO1, respectively, and \(N_{\text{total}}(t)\) normalizes the counts profiles such that their flanks have values close to one:

\[
N_{\text{total}}(t) = \frac{1}{2} \left(1 + \frac{P(t)}{P_{\text{max}}(t)}\right)
\]

Mathematical models of gene expression. The condensation model assumes
that the fraction \(f_j\) of cells in which gene \(j\) is actively transcribed depends on
promoter openness \(\alpha_j\) (measured by ATAC-seq) which in turn depends on
the level \(\text{MeCP2}\) and MeCP2 gene methylation \(\rho_j = f_j(\text{MeCP2})\). The
model predicts that \(\log_{2}C_{\text{MeCP2}}/C_{\text{rest}}\times\alpha_j\) the ratio of gene expression of cell
line X versus cell line Y should yield the same curve (plus a constant) as
the logarithm of the ratio of accessibility of X versus Y when plotted as a
function of \(\rho_j\). Data does not support this model (Fig. 3C). The detachment
model posits that the probability that RNA Pol II successfully terminates is
\(P = (1 - \lambda)^{\rho_j} = \rho_j^{\gamma}\), where \(\gamma\) is the number of "aborted sites" on the gene,
and \(\lambda\) is the proportional to the number of MeCP2 molecules on the gene, and \(\lambda\) is the aborption probability. We show that

\[
\log_{2}C_{\text{MeCP2}} = \text{const} - \gamma \times \rho_j(\text{MeCP2}) - 1
\]

where \(\alpha_j\) is an unknown parameter identical for all cell lines, and \(\text{MeCP2}\) are MeCP2 levels in cell lines X and Y. The model is rejected (Fig. 3F).

We consider two mechanisms by which MeCP2 could affect elongation.
To implement the slow site models we use the totally asymmetric simple
exclusion process (TASEP) with open boundaries (24). A gene is represented
as a chain of 1s and 0s. Each site (equivalent to 60bp of the DNA) is either
occupied by a particle (RNA Pol II) or is empty. Particles enter the chain at
site \(i = 1\) with rate \(a\) (transcription initiation rate), move along the chain
and exit at site \(i = L\). For a fast inclusion site, we expect the chain to
be "fast" and not include sites that represent mCG affected by the interaction with MeCP2, whereas fast sites are all other sites (methylated or not). Particles jump with rate \(\nu = 1\) sec\(^{-1}\) (equivalent to Pol II speed 60bp/s) on fast sites and \(\nu = 0.05\) sec\(^{-1}\) on slow sites. Slow sites are randomly and uniformly distributed with density \(\beta_s = 0.09\text{MCG}\) where the probability that an mCG is occupied by MeCP2. To relate this model to the mRNA-seq differential expression data we calculate

\[\log_{2}C_{mRNA}\]

Footline Author
We thank Tanja Waldmann for introducing us to the LUMHES cell line, Beatrice Alexander-Hovden for technical support, David Kelly for microscopy assistance, Martin Waterfall for help with FACS and Jim Sefridge for help in preparing samples for HPLC. We thank Sabine Lagger and John Connolly for critical assessment of the manuscript. The work has made use of resources provided by the Edinburgh Compute and Data Facility (ECDF; www.ecdf.ed.ac.uk) and was supported by a Wellcome Trust Programme Grant and Investigator Award to AB. AB is a member of the Simons Initiative for the Developing Brain. JCW was supported by a grant from the Rett Syndrome Research Trust. RS was supported by a Wellcome Trust 4 year PhD studentship. BW was supported by a Personal Research Fellowship from the Royal Society of Edinburgh. Author contributions J.C.W and R.S. designed experiments, made and characterised cell lines and interpreted data; J.C.W performed the NGS experiments; S.W and K.C. performed NGS data preparation; B.S. performed HPLC analysis, M.Y. performed initial reaction for TAB-seq; P.G. and B.W. developed the “Congestion model”. B.W. conceived the mathematical modelling, analysed NGS data and interpreted data. A.B. coordinated the study and, together with J.C.W., R.S., and B.W., wrote and edited manuscript.

Acknowledgments

We thank Tanja Waldmann for introducing us to the LUMHES cell line, Beatrice Alexander-Hovden for technical support, David Kelly for microscopy assistance, Martin Waterfall for help with FACS and Jim Sefridge for help in preparing samples for HPLC. We thank Sabine Lagger and John Connolly for critical assessment of the manuscript. The work has made use of resources provided by the Edinburgh Compute and Data Facility (ECDF; www.ecdf.ed.ac.uk) and was supported by a Wellcome Trust Programme Grant and Investigator Award to AB. AB is a member of the Simons Initiative for the Developing Brain. JCW was supported by a grant from the Rett Syndrome Research Trust. RS was supported by a Wellcome Trust 4 year PhD studentship. BW was supported by a Personal Research Fellowship from the Royal Society of Edinburgh. Author contributions J.C.W and R.S. designed experiments, made and characterised cell lines and interpreted data; J.C.W performed the NGS experiments; S.W and K.C. performed NGS data preparation; B.S. performed HPLC analysis, M.Y. performed initial reaction for TAB-seq; P.G. and B.W. developed the “Congestion model”. B.W. conceived the mathematical modelling, analysed NGS data and interpreted data. A.B. coordinated the study and, together with J.C.W., R.S., and B.W., wrote and edited manuscript.