MeCP2 is involved in random mono-allelic expression for a subset of human autosomal genes

Marine Brousseau, Juliette Nectoux, Benjamin Saintpierre, Nicolas Lebrun, Nicolas Cagnard, Brigitte Izac, Emmanuelle Olivier, Franck Letourneur, Thierry Bienvenu

MeCP2 is involved in Random Mono-Allelic Expression for a Subset of Human Autosomal Genes

Marine Brousseau [a] .Juliette Nectoux[a] .Benjamin Saintpierre[b] .Nicolas Lebrun[c] .Nicolas Cagnard[d] .Brigitte Izac[b] .Emmanuelle Olivier[d] .Franck Letourneur [b] . Thierry Bienvenu[a,c]

a Assistance Publique - Hôpitaux de Paris, APHP. Centre Universitaire Paris, Hôpital Cochin, Laboratoire de Génétique et Biologie Moléculaires, Paris, France

b Plateforme Génomique, Institut Cochin, Paris, France

c Institut de Psychiatrie et de Neurosciences de Paris (IPNP), INSERM U1266, Team « Vulnérabilité aux troubles psychiatriques et addictifs », Université de Paris, Paris, France

d Plateforme Bioinformatique, Université Paris Descartes, Institut Imagine, Paris, France

Corresponding author : Thierry Bienvenu

Institut de Psychiatrie et de Neurosciences de Paris, 102 rue de la santé, 75014, Paris, France

Fax : 33 1 44 41 24 21 / 33 1 58 41 15 80 e-Mail: thierry.bienvenu@inserm.fr

Keywords :Rett syndrome ; MECP2 ; Mono-allelic expression

ABSTRACT

Widespread random monoallelic gene expression (RMAE) effects influence about 10% of human genes. However, the mechanisms by which RME of autosomal genes is established and those by which it is maintained both remain open questions. Because the choice of allelic expression is randomly performed cell-by-cell, the RMAE mechanism is not observable in non-clonal cell populations or in whole tissues. Several target genes of MeCP2, the gene involved in Rett syndrome (RTT), have been previously described as subject to RMAE, suggesting that MeCP2 may be involved in the establishment and / or maintenance of RME of autosomal genes. To improve our knowledge on this largely unknown phenomenon, and to study the role of MeCP2 in RMAE, we compared RMA gene expression profiles in clonal cell cultures expressing wild-type MeCP2 versus mutant MeCP2 from a RTT patient carrying a pathogenic non-sense variant. Our data clearly demonstrated that MeCP2 deficiency does not affect significantly allelic gene expression of X-linked genes, imprinted genes as well as the RMAE profile in the majority of genes. However, the functional deficiency in MeCP2 appeared to disrupt the mono-allelic or the bi-allelic expression of at least 49 genes allowing us to define a specific signature of *MECP2* mutated clones.

1.Introduction

Rett syndrome (RTT) is a rare neurodevelopmental disorder which affects approximately 1:15,000 live-born children and is predominantly observed in females [1]. RTT is linked to a mutation in the gene encoding methyl-CpG binding protein 2 (*MECP2*) which is located on the X-chromosome [2]. MeCP2 is an important regulator of neuronal function and development. It influences neuronal differentiation, maturation, morphology and synaptic plasticity. On the molecular level, MeCP2 acts as regulator of gene expression and, depending on the cofactor, can act as activator or inhibitor [1]. This dual role of MeCP2 in gene control agreed with the recent observation of the opposing effect of MeCP2 on DNMT3A as potential inhibitor and stimulator depending on the genomic context [3]. It also influences alternative splicing and microRNA expression. Moreover, several genome-wide expression profiling studies in RTT patients and Mecp2-null mice also showed that some imprinted genes were apparently deregulated, such as *IGF2* and *GNAS* in RTT lymphoblasts and *CDKN1C* in a differentiated MeCP2 blocked neuronal cell line. This suggested that MeCP2 is required to silence imprinted genes with functions that are essential during neurogenesis but unnecessary or detrimental in the mature brain [4]. Unlike imprinted genes, there is a class of autosomal genes that are expressed in a random and clonal fashion, from either the paternal or maternal allele as in X chromosome inactivation (XCI). This class traditionally included members of large gene families, such as olfactory receptors, antigen receptors and protocadherins, which are usually expressed in a single and highly specialized cell type, but also some genes which belong to a wide range of gene ontology categories, from cell adhesion to transcription factors and tend to be expressed in a lineage-specific manner. These genes can be expressed monoallelically from either the paternal or the maternal allele but also biallelically from their two alleles, which may contribute to generating some level of diversity within a cell population [4]. Similarly to XCI and genomic imprinting, epigenetic mechanisms may be at play to maintain these expression patterns. However, the mechanism by which random monoallelic expression of autosomal genes is established and those by which it is maintained both remain an open question [5,6]. Interestingly, several target genes of MeCP2 have been previously described as subject to random mono-allelic gene expression (RMAE), such as *BDNF, CADM1, SERPINE2, PCDHB1, STMN2* or *UCHL1* [7]. These results suggest that MeCP2 may be involved in the

establishment and / or maintenance of RME of autosomal genes and therefore that RTT may be in part related to the disruption of this RMAE mechanism. To improve our knowledge on this largely unknown phenomenon, and to study the role of MeCP2 in RMAE, we compared gene expression profiles in cell cultures expressing wild-type MeCP2 versus mutant MeCP2. Because the choice of allelic expression is randomly performed cell-by-cell, the RMAE mechanism is not observable in non-clonal cell populations or in whole tissues. To overcome this difficulty, we studied gene expression in clonal cell cultures from a RTT patient.

2. Materials and methods

2.1. Cell culture, single cell cloning and nucleic acid extraction for fibroblasts

After informed consent obtained from the parents, a skin biopsy was carried out on a girl suffering from typical RTT. This patient presented the common recurrent p.Arg294* (c.880C>T) nonsense mutation within the *MECP2* gene. No X chromosome inactivation bias was detected in this studied patient. Clonal cultures were obtained from primary cultures of human fibroblasts by the limit dilution method. Briefly, fibroblasts from the patient were plated at equivalent one cell in 96-well plates in Dulbecco's modified Eagle's medium F-12 HAM (Sigma-Aldrich Chimie, Lyon, France) supplemented with 10% heat-inactivated fetal bovine serum and 50 units/ml penicillin and streptomycin. All cultures were routinely grown at 37°C, 5% CO2 in a humidified atmosphere. Wells indicating positive growth were systematically tested for clonality by two independent, complementary molecular approaches: allele-specific transcript analysis, and X chromosome inactivation study. Total RNA was extracted from fibroblast clone pellets with the Maxwell system (Promega, Madison, WI, USA) and the Maxwell 16 LEV simply RNA FABRIC AS 1280 kit (Promega, Madison, WI, USA) according to the manufacturer's recommendations. Prior to array hybridization, RNA quality was assessed with the NanoDrop spectrophotometry (NanoDrop Technologies, Wilmington, DE, USA) and the Agilent 2100 Bioanalyzer (Agilent Technologies, Massy, France). DNA was extracted using a standard extraction protocol to determine XCI status and to check their clonal status.

2.2. Allele-specific transcript analysis

After DNase I digestion (Roche Diagnostics, Meylan, France), reverse transcription of total RNA (1 µg/reaction) was carried out with the Superscript™ II RNase H-Reverse Transcriptase (Invitrogen, Cergy-Pontoise, France) using 500 ng of random hexamers. The genomic region containing the *MECP2* mutation was amplified by PCR and sequenced with an ABI 3130XL® (Applied Biosystems, Courtaboeuf, France) as previously described [7]. Cell clones expressing only the normal *MECP2* allele are called "wild-type clones", while cell clones expressing only the mutant *MECP2* allele are called "mutant clones".

2.3 X Chromosome Inactivation study

XCI status was investigated by sizing the polymorphic CAG trinucleotide repeat on the inactive androgen receptor allele. Digestion of four restriction sites located between the flanking primers and adjacent to the trinucleotide repeat were carried out with the methylation sensitive enzymes HpaII and CfoI. Amplification of the undigested allele including the trinucleotide repeat region was performed using the fluorescent-labeled, forward primer 5'-FAM-TGCGCGAAGTGATCCAGAAC-3', and the unlabeled, reverse primer 5'-CTTGCGGAGAACCATCCTCA-3'. Amplification products were sized with the ABI 3130XL® sequencer using GeneScan™ Fragment Analysis software (Applied Biosystems, Courtaboeuf, France). Undigested DNA samples of each original cell culture were amplified to size each allele.

2.4. RNA-sequencing

For quality check, total RNA was analyzed using Agilent Bioanalyzer 2100 (Agilent Technologies) and RNA 6000 Nano Kit (Agilent) to ensure appropriate RNA quality in terms of degradation. To construct the libraries, 1 µg of high quality total RNA sample (RNA Integrity Number (RIN)>8) was processed using TruSeq Stranded mRNA kit (Illumina, San Diego, United States) according to manufacturer instructions. Libraries were quantified by qPCR using the KAPA Library Quantification Kit for Illumina Libraries (KapaBiosystems, Wilmington, MA) and library profiles were assessed using the DNA High Sensitivity LabChip kit on an Agilent Bioanalyzer. Libraries were sequenced on an Illumina Nextseq 500 instrument using 75 base-lengths read V2 chemistry in a paired-end mode. Approximately 40 million reads were obtained per sample. Reads were extracted in FastQ format using CASAVA v1.8.2 or v1.8.3 (Illumina).

2.5. RNA-seq data analysis

Raw sequencing data were received in FASTQ format. Read mapping was performed using Tophat 2.0.6 and the human genome references assembly GRCh37 (http://feb2012.archive.ensembl.org/). The resulting SAM alignment files were processed using the HTSeq Python framework and the respective GTF gene annotation, obtained from the Ensembl database. Gene counts were further processed using the R programming language and normalized to Reads Per Kilobase of transcript per Million mapped reads (RPKM) values. In order to examine the variance and the relationship of global gene expression across the samples, different correlation values have been computed including Spearman's correlation of gene counts and Pearson's correlation of log2 RPKM values. The resulting correlation values were visualized using multi-dimensional scaling plots (MDS) and heatmaps. Subsequently, the Bioconductor packages DESeq and edgeR were used to identify differentially expressed genes (DEG). Both packages provide statistics for determination of differential expression in digital gene expression data using a model based on the negative binomial distribution. The non-normalized gene counts have been used here, since both packages include internal normalization procedures. The resulting p-values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate (FDR). Genes with an adjusted p-value < 0.05 found by both packages were assigned as differentially expressed. In the current study, we reported RNA-Seq to identify allele-biased gene expression. The applied filter allowed to retain sequences with a depth> 10X (in order to eliminate the "background noise" sequences corresponding to very little gene expression). The same filter was applied to study the distribution of RPKM values for all identified transcripts and for all RMAE genes in wild-type and mutant clones. Gene lists were examined using Enrichr [8,9] to investigate the transcription factor binding sites (ChIP Enrichment Analysis (ChEA)(http://amp.pharm.mssm.edu/lib/chea.jsp), the transcriptional regulatory interactions (TRRUST) (https://www.grnpedia.org/trrust/), the histone modifications (https://www.encodeproject.org/chip-seq/histone/),and biological pathways (https://www.wikipathways.org/index.php/WikiPathways) that were modulated. Clustering of wild-type and mutant cell clones was performed through a binary matrix based on SNPs frequency (1 if >80% or 0 if <80%) and a list of genes selected for their frequency modulations between the wild-type and mutated groups. Data have been sent to ncbi

database (https://www.ncbi.nlm.nih.gov/geo/subs/) and the GEO accession number is GSE140740.

2.6. Exome Analysis

Genomic DNA was extracted from fibroblast clones from the patient by a standard procedure using the Maxwell system (Promega, Madison, WI, USA) and the LEV DNA BLOOD AS 1290 kit (Promega, Madison, WI, USA). Whole exome capture, next-generation sequencing and data analysis were carried out by Cochin Institute and Paris Descartes Platform, respectively. Briefly, libraries were prepared from 3 µg genomic DNA extracted from whole blood using an optimized SureSelect Human Exome kit (Agilent) following the manufacturer's instructions. Captured, purified and clonally amplified libraries targeting the exome were then sequenced on a Nextseq 500 instrument (Illumina) according to the manufacturer's recommendations. Obtained sequence reads were aligned to the human genome (hg19) using BWA software. Downstream processing was carried out with the Genome analysis toolkit (GATK), SAMtools and Picard Tools (http://picard.sourceforge.net). Single-nucleotide variants and indels were subsequently called by the SAMtools suite (mpileup, bcftools, vcfutil). All calls with a read coverage ≤20× and a Phred-scaled SNP quality of ≤20 were filtered out. Variants located in introns were excluded. Substitution and variation calls were made with the SAMtools pipeline (mpileup). Variants were annotated with an in-house Paris Descartes bioinformatics platform pipeline based on the Ensembl database (release 67). In all cases, sequence variants in a gene are numbered starting from the first base of the ATG codon, numbering based on the reference sequence. For each variant, the numbers of reads for the reference and alternative alleles were counted and used for a binomial test to determine if the ratio of the two numbers significantly deviated from 0.5 reflecting the null hypothesis that both alleles were equally expressed. Only variants with allelic bias (up to 80/20%) between the two alleles were analyzed.

2.7 Sanger sequencing

DNA variants of interest were confirmed by RT-PCR amplification and direct sequencing. Exonic primer sequences and PCR conditions are available upon request to the

corresponding author. Sequencing reactions were carried out with the BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Courtaboeuf, France) and loaded on the ABI 3130XL genetic analyser (Applied Biosystems, Courtaboeuf, France).Results were visualized using the Sequencher software (Gene codes, Ann Arbor, United States).

3.Results

3.1.Isolation of wild-type and mutant-MECP2 fibroblast clones

As the choice of allelic gene expression is made on a cell-by-cell basis, RMAE is not observable in non-clonal cell populations or in whole tissues. Furthermore, it has been shown that the epigenetic choice of active allele is clonally stable and that monoallelic clonal cell lines give rise only to monoallelic subclones [5]. Taking into account these points, we isolated fibroblasts from a skin biopsy and cultured them to create clonal cell lines. We studied the clonality of the selected cells by two independent approaches: *MECP2* mRNA analysis by RT-PCR, and XCI status through methylation analysis. Only clones showing concordant results using the two approaches were selected for further studies. Three clonal cultures expressing exclusively the wild-type (WT) allele and three clonal cultures expressing exclusively the mutated (MT) allele were further analyzed.

3.2.Mono-allelic Expression of genes in Wild-type and MeCP2-deficient fibroblasts

Whole exome sequencing from the RTT patient's DNA allowed us to identify the heterozygous variants that are needed to identify allele-specific gene expression. After filtering our data, 2711 variants (single nucleotide variant, SNV) were identified at the heterozygous state in the RTT patient. In parallel, total RNA of each clone of fibroblasts (3 clonal WT cultures named C10, C13 and C15, and 3 clonal MT cultures named C3, C11 and C16) was used for reverse transcription and then subjected to massive RNA-sequencing. The resulting data were analyzed both qualitatively and quantitatively. By using the patient's exome data and the RNA-Seq results from both WT and MT clones, we defined two categories of genes according to the observed allelic bias: 1- RMAE-genes with at least one SNV> 90/10% (named "strict-RMAE") (Table S1), and 2- genes with "biased allele" expression with at least one SNV> 80/20% (named "biased-RMAE") (Table S2). All other expressed genes were considered not to be subject to mono-allelic expression in this cell type. In order

to evaluate the involvement of MeCP2 in RMAE, we defined three lists of RMAE genes: 1-genes with RMAE profile identified in both WT and MT clones, 2-genes with RMAE profile restricted to WT clones (named “WT-only”*i.e.* 1, 2 or 3 WT clones showing RMAE profile), and 3- genes with RMAE profile restricted to MT clones (named “mutant-only” 1, 2 or 3 MT clones showing RMAE profile).

3.3.Validation of Omic’s approaches

To check the quality of the selected clones and our data, we firstly checked that all the X-linked genes subjected to an X chromosome inactivation (and showing a heterozygous informative SNV) had a mono-allelic gene expression in all WT and MT clones. 39 SNVs located on 32 different transcripts corresponding to X-linked genes were identified. 34 of these 39 SNPs, located on 27 distinct transcripts (including *SLC25A5*, 2 SNVs; *XIAP*, 3 SNVs;*PLS3*, 2 SNVs;*MAP7D4*, 2 SNVs;*PLXNA3*, 2 SNVs;*FAM104B*, 2 SNVs), presented a total bias of inactivation and a mono-allele gene expression validating both experimental and data filtering strategies (Table 1). One of these SNVs corresponded to the pathogenic variant of the *MECP2* gene confirming again the validity of all the studied clones studied (WT, allele G, mutated, allele A) (Table 1). On the other hand, five SNVs corresponding to five X-linked genes were found to be bi-allelically expressed. One of these genes was known to escape to X chromosome inactivation (*RIBC1*) [10], and four (*DHRSX, ASMTL, ZBED1, CD99*) of these genes were localized in Xp22.33 in the pseudo-autosomal region (homolog to the Yp11.2 region) confirming again the validity of our experiment (Table 2).

We also analyzed the imprinted genes for which an informative SNV had been identified. We were able to identify in our gene list three imprinted genes. All imprinted genes were shown to have a total expected bias and mono-allelic gene expression in all clones for at least 3 SNVs per gene: *MAP2K3* (11 informative SNVs), *AK2* (3 informative SNVs), and *BCLAF1* (8 informative SNPs) (Table 3) [11-13].

3.4.Identification of autosomal genes subject to mono-allelic expression in human fibroblast clones

After validation of our data, we compared SNV results obtained from the patient’s exome and the sequences of the mRNAs obtained from each cell clone. From the "biased-RMAE"

list, we identified 188 informative SNVs presenting an inactivation bias in the absence of X-chromosome localization or a known parental fingerprinting mechanism and a significant expression (depth> 10, in at least one cell clone) in both WT and MT clones. These 188 SNPs corresponded to 83 transcripts encoded by autosomic genes. In other words, considering our samples, 83 genes were subjected to RMAE in human fibroblasts (Table 4). 37 were previously described as subject to random autosomal mono-allelic gene expression and 36 were new genes subjected to RMAE in humans (Table 4).

To check that the different genes identified as RMAE genes were not biased towards those associated with the lowest expression level, we plotted the distribution of RPKM values for all identified transcripts and for all RMAE transcripts in WT and mutant clones. Figure 1 showed that the profile of the distribution RPKM values appears to be highly similar for all genes and for RMAE genes, with a lower expression in the major peak of RMAE genes compared to all genes. The two peaks observed at the extreme right of the density plot of RMAE genes correspond to two outliers associated with a very important level expression (*FTH1* and *EEF1A1*). Moreover, Figures 2 C and D showed that gene expression appeared to be slightly lower for RMAE as well as for all genes in mutant clones, reinforcing the hypothesis that MECP2 mutant cells are associated with lower transcriptional level (Fig.2).

3.5.Identification of a specific RMAE signature of mutated MECP2 clones versus WT clones

Firstly, using the list of "biaised-RMAE" genes (393 genes), an unsupervised hierarchical clustering, obtained without definition of which cell clones were WT or MT, tend to distinguish two different groups of clones (Fig.3A). These results suggest that MeCP2 is involved in establishment and / or maintenance of RMAE of autosomal genes. Then, in order to investigate if Rett syndrome is related to the disruption of this RMAE mechanism we carried out a supervised hierarchical clustering and we clearly distinguished the group of MT clones from the group of WT clones (Fig.3B). Table S3 indicated the list of the 94 RMAE genes that clearly distinguished MT clones from WT clones (Table S3). These results identified a specific RMAE signature in mutated *MECP2* clones and suggested an impact of the MeCP2 deficiency on monoallelic expression profile for a subset of genes.

In parallel, we defined genes in which RMAE profile (both "strict RMAE" profile and "biased RMAE" profile) was restricted to WT clones ("WT-only" list) and genes in which RMAE was

restricted to MT clones ("mutant-only" list). The "WT-only" list corresponded to 61 SNVs localized on 57 genes (Table 5). 18 (32%) were known to present RMAE in humans (dbMAE database in humans and mice), and 18 (32%) showed RMAE in mice without known data in humans. For the 21 other genes, no information was available or the information was discordant. Our data suggested that the functional deficiency in MeCP2 may disrupt at least the mono-allelic expression of these 18 genes (Table 5). Similarly, 121 SNVs corresponding to 105 genes were identified as "strict or biased-RMAE" only in MT clones (Table 6). 39 (37%) of these genes were known as RMAE in humans and 22 (21%) in mice (without known data in humans). For the 44 other genes, no information was available or the information was discordant. Among the 44 genes identified as RMAE only in MT clones, 31 were known to present bi-allelic expression in humans in all data from the literature and databases (Table 6). Several of these genes encode proteins involved in intracellular organelle lumen (https://string-db.org/cgi/network.pl?taskId=KemsI1XK8kV3), such as *CDK2*, *NOTCH1*, *SMARCA4*, *CHD1L* and *REST.*

Based on these observations, and using the list of genes identified as "biased–RMAE" only in the MT or WT clones, we used EnrichR program to perform pathway analyses and investigate the transcription factor binding sites, the transcriptional regulatory interactions, and the histone modifications. Wikipathway analysis of these RMAE genes showed they were most enriched in the "Disorders of folate metabolism and transport" category (Table S4).Interestingly, Gene enrichment analysis by Enrichr indicated that the most enriched transcription factor sites and transcriptional interactions included GABP, GATA1, E2F6 and SIN3A (Table S4). Moreover, using encode histone modifications, we showed that these RMAE genes showed a particularly strong enrichment for histone H3 acetylation (Table S4). These observations are consistent with MeCP2 function, previous reports showing that DNA methylation and histone deacetylation can be linked by MeCP2 through the MeCP2/Sin3A/HDAC1/2 deacetylating complex [14] and that histone H3 alterations were identified in Mecp2 mice with truncated Mecp2 [15] and in autism brain with reduced MeCP2 expression [16].

3.6.Confirmation of RMAE profiles by Sanger sequencing of transcripts

We then selected a few candidate genes and analyzed by RT-PCR and Sanger sequencing the gene expression bias. New cell cultures were made from the WT and MT clones and mRNAs were again extracted from the different cell clones. RT-PCR were carried out, and sequenced. Firstly, we confirmed the mono-allelic expression of MeCP2 in all WT and MT clones. Then, we focused on two potential candidate genes, *HSPA2* previously identified to be regulated by MeCP2 in mice and *SPATA22* [17,18]. For the *HSPA2* gene, we observed and confirmed the bias and mono-allelic gene expression in all WT and MT clones (C allele in all clones), and for the *SPATA22* gene, we observed total bias and mono-allelic gene expression in all WT clones and no expression in all MT clones (Fig.S1).

3.7.Identification of genes deregulated by the absence of functional MeCP2

Although this study was not designed to identify target genes of MeCP2 due to the small number of cell clones, we identified 117 genes for which mRNA expression was disrupted in MeCP2 deficient cells (adjusted p <0.05). Among these 117 genes, 62 genes were identified as over-expressed (Table S5), and 55 as under-expressed genes in the MT clones as compared to the WT clones from the same RTT patient (Table S6). For the great majority of the RMAE genes generated from our two category lists ("strict-RMAE" and "biased-RMAE"), we did not observe a significant change in the level of gene expression. We only observed gene expression deregulation for two identified RMAE genes *COL18A1* and *PEAR1*. *PEAR1*was identified as overexpressed in MT clones as compared to WT clones, whereas *COL18A1*was found to be under-expressed in MT clones as compared to WT clones (TablesS5 and S6).

4.Discussion

Autosomal monoallelic expression refers to mitotically stable, epigenetically controlled allele-specific expression of autosomal genes, with the initial non-predetermined choice of the transcriptional activity of the two alleles maintained in a given clonal cell lineage. Recently, RMAE was found to be established in a lineage-specific manner and stably maintained throughout terminal cell differentiation [19]. However, the mechanisms by which RME of autosomal genes is established and those by which it is maintained both remain open questions. Several studies suggested that histone methyltransferase EZH2, and possibly other components of the PRC2 complex could be involved in RMAE maintenance

[20]. Because PRC2 and MBD proteins such as MeCP2 could collaborate to different phases of the long-term silencing phenomenon [21,22], we suggested that MeCP2 could be involved in this mechanism. Moreover, several protocadherins such as PCDHB1 and PCDH7, known to be usually expressed in a single and highly specialized cell type, were found to be regulated by MeCP2 in human neuroblastoma cells and brain tissue [23].

It has been established that RME of autosomal genes is widespread in the human genome, affecting approximately 10 % of assessed genes [6]. In order to understand the extent of this phenomenon in human, we examined RMAE in human cells, using a genome scale-approach in clonal fibroblast lines from a RTT individual. Here, a small number of genes (~250) were found to be subject to RMAE in human fibroblasts. Among these RMAE genes, 83 were identified as RMAE in all WT and MT clones. 37 (44%) were previously identified as subject to mono-allelic expression in human and 23 (28%) in mice (without data in human)(in at least one cell type with moderate to high expression; https://mae.hms.harvard.edu/)[6]. For the remaining 23 genes, the data were not available in human (n=13) or not concordant (bi-allelic autosomal expression (BAE) for 10 genes). Taking into account that we observed a lower expression of RMAE genes in mutant and wild-type clones, we cannot exclude that this lower expression of RMAE genes may account for few differences in RMAE in MECP2 mutated clones. For RMAE genes identified by only one single clone, replicative studies will be needed to confirm these data. Our data identified at least 13 novel human genes subjected to RMAE, and 23 being already known to be RMAE in mice.

To understand the putative role of MeCP2 in this phenomenon, we identified 57 genes showing RMAE only in WT clones. 18 of these genes were previously known to present RMAE in humans, and 18 have been identified as RMAE gene in mice without any human data. For 5 other genes, no information was available. Information was discordant for 16 genes, previously found to present bi-allelic expression in humans. It is highly probable that a larger number of clones will identify the biallelic expression (BAE) profile for these 16 human genes. We also identified 105 genes showing RMAE only in MT clones. 39 were previously identified as RMAE genes in humans suggesting again that a larger number of WT clones will probably identify this RMAE profile.

Regardless of the type of list (more or less stringent), it is interesting to note that there are always more transcripts characterized by RMAE in the "mutant-only" list than in the "WT-only" list, with an approximate ratio of 1.5. These results suggest that, physiologically, the *MECP2* gene could be involved in the "relaxation" of RMAE. A mutation in the *MECP2* gene that would be involved in Rett syndrome might therefore lead to an increase in transcripts characterized by RMAE. This hypothesis was previously confirmed by studies showing that Mecp2 deficient cells show an absolute amount of mRNA transcripts about 15% lower than normal cells [24,25].

However, interestingly, RMAE was only observed in WT clones for 18 genes, as previously shown in the literature (dbMAE database in humans and mice). Our data suggest that the functional deficiency in MeCP2 may disrupt the mono-allelic expression of these 18 genes including several genes playing role in brain function (such as *AKAP6, NEDD9, SEPP1*) and the methylene tetrahydrofolate reductase (*MTHFR*) gene. Some polymorphisms of *MTHFR* may result in reduction of MTHFR activity as well as DNA methylation process, and may have significant impacts in various neurological and psychiatric disorders [26]. Moreover, among the 105 genes identified as RMAE genes in MT clones, 31 genes were always found to be biallelic in WT clones, but also in all data from the literature [6]. It is of interest to note that several of these genes encode proteins involved in synaptic plasticity (such as *CAPRIN2, NEURL4*), in brain function (*UBA6, NCAPD3, REST, NOTCH1, ATIC, VPS13C, NUP214*) or in epigenetic mechanisms and chromatin remodeling (*SMARCA4, SETD4*). Moreover, Notch1 involved in the axo-glial interaction, was found to be under the regulation of astrocytic Mecp2 [27]. All these data may suggest that MeCP2 may affect bi-allelic expression of several human genes involved in brain function.

5.Conclusion

Our data obtained in human clonal cells suggest that MeCP2 deficiency does not significantly affect X chromosome inactivation pattern, genomic imprinting as well as random autosomal allelic gene expression. However, MeCP2 deficit may affect specifically RMAE of a subset of autosomal genes without affecting their expression level defining a specific RMAE signature. Only two of our RMAE genes are considered as differentially expressed in Mecp2 deficient cells.

Moreover, presence of variant in these RMAE genes may affect the phenotype depending of this phenomenon. Several previous reports suggested that there is a clear potential for RMAE to contribute to phenotypic differences among individual organisms. This mechanism may lead to phenotypic differences in Mendelian and complex genetic diseases. Even in Rett syndrome, there is often phenotypic variability within families that can be reasonably ascribed to differences in genetic background or to gene-environment interaction [28]. Autosomal RME provides another potential explanation for phenotypic variability, in addition to the variability in phenotype observed in RTT patients due to X-chromosome inactivation [29]. This mechanism has been previously suggested to be implicated in skin, neuropsychiatric and developmental speech and language disorders [30-33]. Interestingly, allele-biased expression has been observed in more than 800 genes in differentiating neurons suggesting a plausible mechanism to explain incomplete penetrance and monozygotic twin discordance in neuropsychiatric disorders [31, 34] and in Rett syndrome [28].

Acknowledgements This study was supported by Institut National de la Recherche Médicale (INSERM) and by the Association Française du syndrome de Rett (AFSR). We thank the family for their enthusiastic participation.

Author's contributions TB and JN developed the project. TB, JN, MB, and BJ designed the study and analyzed and interpreted the data. MB, BI, NL, and FL performed the experiments. BS, NC, and EO performed and analyzed the bioinformatics data. All authors approved the final version of the manuscript.

Funding Compliance with Ethical standards All procedures performed in this study were in accordance with the ethical standards of our national research committee and with the 1964 Helsinki declaration and its later amendments. Informed written consent was obtained from the patient and parents included in the study.

Conflict of interest The authors declare that they have no conflict of interest.

References

[1] L.M. Lombardi, S.A. Baker, H.Y. Zoghbi, MECP2 disorders: from the clinic to mice and back, J. Clin. Invest. 125 (2015) 2914-2923.

[2] R.E. Amir, I.B. Van den Veyver, M. Wan, C.Q. Tran, U. Francke, H.Y. Zoghbi, Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2, Nat. Genet. 23 (1999) 185-188.

[3] A. Rajavelu, C. Lungu, M. Emperle, M. Dukatz, A. Bröhm, J. Broche, I. Hanelt, E. Parsa, S. Schiffers, R. Karnik, A. Meissner, T. Carell, P. Rathert, R.Z. Jurkowska, A. Jeltsch, Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2, Nucleic Acids Res. 46 (2018) 9044-9056.

[4] K.D. Kernohan, Y. Jiang, D.C. Tremblay, A.C. Bonvissuto, J.H. Eubanks, M.R. Mann, N.G. Bérubé, ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain,Dev. Cell 18 (2010) 191-202.

[5] A.A. Gimelbrant, A.W. Ensminger, P. Qi, J. Zucker, A. Chess, Monoallelic expression and asynchronous replication of p120 catenin in mouse and human cells, J. Biol.Chem. 280 (2005) 1354-1359.

[6] V. Savova, J. Patsenker, S. Vigneau, A.A. Gimelbrant, dbMAE: the database of autosomal monoallelic expression, Nucleic Acids Res. 44 (2016) D753-756.

[7] J. Nectoux, Y. Fichou, H. Rosas-Vargas, N. Cagnard, N. Bahi-Buisson, P. Nusbaum, F. Letourneur, J. Chelly, T. Bienvenu, Cell Cloning-Based Transcriptome Analysis in Rett Patients: Relevance to the Pathogenesis of Rett Syndrome of New Human MeCP2 Target Genes, J. Cell Mol. Med. 14 (2010) 1962–1974.

[8] E.Y.Chen, C.M. Tan, Y. Kou, Q. Duan, Z. Wang, G.V. Meirelles, N.R. Clark, A. Ma'ayan, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool, Bioinformatics 14 (2013) 128.

[9] M.V. Kuleshov, M.R. Jones, A.D. Rouillard, N.F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S.L. Jenkins, K.M. Jagodnik, A. Lachmann, M.G. McDermott, C.D. Monteiro, G.W. Gundersen, A. Ma'ayan, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, Nucleic Acids Res. 44 (2016) W90-7.

[10] G. Froyen, M. Corbett, J. Vandewalle, I. Jarvela,O. Lawrence, C. Meldrum, M. Bauters, K. Govaerts, L. Vandeleur, H. Van Esch, J. Chelly, D. Sanlaville, H. van Bokhoven, H.H. Ropers, F. Laumonnier, E. Ranieri, C.E. Schwartz, F. Abidi, P.S. Tarpey, P.A. Futreal, A. Whibley, F.L. Raymond, M.R. Stratton, J.P. Fryns,R. Scott, M. Peippo, M. Sipponen, M. Partington, D. Mowat, M. Field, A. Hackett, P. Marynen, G. Turner, J. Gécz, Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation, Am. J. Hum. Genet.82 (2008) 432-443.

[11] R.G. Tuskan RG, S. Tsang S, Z. Sun Z, J. Baer J, E. Rozenblum E, X. Wu X, D.J. Munroe DJ, K.M. Reilly, Real-time PCR analysis of candidate imprinted genes on mouse chromosome 11 shows balanced expression from the maternal and paternal chromosomes and strain-specific variation in expression levels, Epigenetics 3(2008) 43-50.

[12] F.A. Santoni, G. Stamoulis, M. Garieri, E. Falconnet, P. Ribaux, C. Borel, S.E. Antonarakis, Detection of Imprinted Genes by Single-Cell Allele-Specific Gene Expression, Am. J. Hum.Genet. 100 (2017) 444-453.

[13] T. Metsalu, T. Viltrop, A. Tiirats, B. Rajashekar, E. Reimann, S. Kõks, K. Rull, L. Milani, G. Acharya, P. Basnet, J. Vilo, R. Mägi, A. Metspalu, M. Peters, K. Haller-Kikkatalo, A. Salumets, Using RNA Sequencing for Identifying Gene Imprinting and Random Monoallelic Expression in Human Placenta, Epigenetics 9 (2014) 1397–1409.

[14] X. Nan, H.H. Ng, C.A. Johnson, C.D. Laherty,B.M. Turner, R.N. Eisenman, A. Bird, Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex, Nature 393 (1998) 386-389.

[15] M. Shahbazian, J. Young, L. Yuva-Paylor, C. Spencer, B. Antalffy, J. Noebels,D. Armstrong, R. Paylor, H. Zoghbi, Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3, Neuron35 (2002) 243-254.

[16] K.N. Thatcher, J.M. LaSalle, Dynamic changes in Histone H3 lysine 9 acetylation localization patterns during neuronal maturation require MeCP2, Epigenetics 1 (2006) 24-31.

[17] D. Zhao, M. Lin, E. Pedrosa, H.M. Lachman, D. Zheng, Characteristics of allelic gene expression in human brain cells from single-cell RNA-seq data analysis, BMC Genomics 18 (2017) 860.

[18] F. Ehrhart, S.L. Coort,L. Eijssen, E. Cirillo, E.E. Smeets, N. Bahram Sangani, C.T. Evelo, L.M.G. Curfs, Integrated analysis of human transcriptome data for Rett syndrome finds a network of involved genes, World J Biol Psychiatry 25 (2019) 1-23.

[19] A. Nag, V. Savova, H.L. Fung, A. Miron, G.C. Yuan, K. Zhang, A.A. Gimelbrant, Chromatin signature of widespread monoallelic expression, Elife 2 (2013) e01256.

[20] A. Nag, S. Vigneau, V. Savova, L.M. Zwemer, A.A. Gimelbrant, Chromatin Signature Identifies Monoallelic Gene Expression Across Mammalian Cell Types, G3 (Bethesda) 5 (2015) 1713-1720.

[21] M.R. Matarazzo, M.L. De Bonis, M. Strazzullo, A. Cerase, M. Ferraro, P. Vastarelli, E. Ballestar, M. Esteller, S. Kudo, M. D'Esposito, Multiple binding of methyl-CpG and

polycomb proteins in long-term gene silencing events, J. Cell Physiol. 210 (2007) 711-719.

[22] J. Mann, D.C. Chu, A. Maxwell, F. Oakley, N.L. Zhu, H. Tsukamoto, D.A. Mann, MeCP2 controls an epigenetic pathway that promotes myofibroblast transdifferentiation and fibrosis, Gastroenterology 138 (2010) 705-714, 714.e1-4.

[23] K. Miyake, T.Hirasawa, M. Soutome, M. Itoh, Y. Goto, K. Endoh, K. Takahashi, S. Kudo, T. Nakagawa, S. Yokoi, T. Taira, J. Inazawa, T. Kubota, The protocadherins, PCDHB1 and PCDH7, are regulated by MeCP2 in neuronal cells and brain tissues: implication for pathogenesis of Rett syndrome BMC Neurosci. 12 (2011) 81.

[24] M. Yazdani, R. Deogracias, J. Guy,R.A. Poot, A. Bird, Y.A. Barde, Disease modeling using embryonic stem cells: MeCP2 regulates nuclear size and RNA synthesis in neurons, Stem Cells 30 (2012) 2128-2139.

[25] Y. Li, H. Wang, J. Muffat, A.W. Cheng, D.A. Orlando, J. Lovén, S.M. Kwok, D.A. Feldman, H.S. Bateup, Q. Gao, D. Hockemeyer, M. Mitalipova, C.A. Lewis, M.G. Vander Heiden, M. Sur, R.A. Young, R. Jaenisch, Global transcriptional and translational repression in human-embryonic-stem-cell-derived Rett syndrome neurons, Cell Stem Cell 13 (2013) 446-58.

[26] L. Wan, Y. Li, Z. Zhang, Z. Sun,Y. He, R. Li, Methylene tetrahydrofolate reductase and psychiatric diseases, Transl. Psychiatry 8 (2018) 242.

[27] B. Lipi, L. Jaldeep, P. Prakash, Role of astrocytic MeCP2 in regulation of CNS myelination by affecting oligodendrocyte and neuronal physiology and axo-glial interactions, Exp. Brain Res. 236 (2018) 3015-3027.

[28] K. Miyake, C. Yang, Y. Minakuchi, K. Ohori, M. Soutome, T. Hirasawa, Y. Kazuki, N. Adachi, S. Suzuki, M. Itoh, Y. Goto, T. Andoh, H. Kurosawa, M. Oshimura, M. Sasaki, A. Toyoda, T. Kubota, Comparison of Genomic and Epigenomic Expression in Monozygotic Twins Discordant for Rett Syndrome, PLoS One 8 (2013) e66729.

[29] J.I. Young, H.Y. Zoghbi, X-chromosome inactivation patterns are unbalanced and affect the phenotypic outcome in a mouse model of rett syndrome, Am. J. Hum. Genet. 74 (2004) 511-520.

[30] R. Happle, Monoallelic expression on autosomes may explain an unusual heritable form of pigmentary mosaicism: a historical case revisited. Clin. Exp. Dermatol. 34 (2009) 834-837.

[31] A.R. Jeffries, D.A. Collier, E. Vassos, S. Curran, C.M. Ogilvie, J. Price, Random or stochastic monoallelic expressed genes are enriched for neurodevelopmental disorder candidate genes, PLoS One 8 (2013) e85093.

[32] E. Ben-David, S. Shohat, S. Shifman, Allelic expression analysis in the brain suggests a role for heterogeneous insults affecting epigenetic processes in autism spectrum disorders, Hum. Mol. Genet. 23 (2014) 4111-4124.

[33] A.A. Adegbola AA, G.F. Cox GF, E.M. Bradshaw EM, D.A. Hafler DA, A. Gimelbrant A, A. Chess, Monoallelic expression of the human FOXP2 speech gene, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) 6848-6854.

[34] M. Lin, A. Hrabovsky, E. Pedrosa, T. Wang, D. Zheng, H.M. Lachman, Allele-biased expression in differentiating human neurons: implications for neuropsychiatric disorders, PLoS One 7 (2012) e44017.

Legends to the figures and tables :

Table 1. Genes located on the X chromosome and subjected to X chromosome inactivation. Each column represents an individual clone. WT indicated wild-type clones and MT indicated mutant clones. The percentage of mono-allelic gene expression is indicated for each gene per clone, and showed mono-allelic gene expression in all analyzed wild-typeand mutated *MECP2*clones.

Table 2. Genes located on the X chromosome in the pseudoautosomal region or known to escape to X chromosome inactivation. Each column represents an individual clone. WT indicated wild-type clones and MT indicated mutant clones. The percentage of mono-allelic gene expression is indicated for each gene per clone, and showed bi-allelic gene expression in all analyzed wild-type and mutated *MECP2*clones.

Table 3. Autosomal genes previously known to be subject to the parental imprinting. Each column represents an individual clone. WT indicated wild-type clones and MT indicated mutant clones. The percentage of mono-allelic gene expression is indicated for each gene per clone, showing a mono-allelic gene expression in all analyzed wild-type and mutated *MECP2*clones.

Table 4. Autosomal genes identified to be subject to mono-allelic gene expression in wild-type and mutated *MECP2* clones. Each column represents an individual clone. WT indicated wild-type clones and MT indicated mutant clones. Symbol 1 means allelic biais in gene expression. Concordance of the RMAE status between this study and previous works and the DbMAE database is indicated (https://mae.hms.harvard.edu/)(MAE, mono-allelic expression; BAE biallelic expression; UTD, undetermined)(MAE human in light grey; UTD human in dark gray) [4].

Table 5. Autosomal Genes identified to be subject to mono-allelic gene expression and showing restricted RMAE to WT clones. Each column represents an individual clone. WT indicated wild-type clones and MT indicated mutant clones. Symbol 1 means allelic biais in gene expression. Concordance of the RMAE status between this study and previous work

and the DbMAE database is indicated (https://mae.hms.harvard.edu/)(MAE, mono-allelic expression; BAE biallelic expression; UTD, undetermined)(MAE human in light grey; UTD human in dark gray) [17].

Table 6. Autosomal Genes identified to be subject to monoallelic gene expression and showing restricted RMAE to MT *MECP2* clones. Each column represents an individual clone. WT indicated wild-type clones and mut indicated mutant clones. Symbol 1 means allelic biais in gene expression. No symbol means no allelic biais. Concordance of the RMAE status between this study and previous work and the DbMAE database (https://mae.hms.harvard.edu/)(MAE, mono-allelic expression; BAE biallelic expression; UTD, undetermined)(MAE human in light grey; UTD human in dark gray) [17].

Table S1. List of autosomal genes identified (by at least one SNV >90/10%) as random monoallelically expressed in human cell clones. Each column represents an individual clone. WT indicated wild-type clones and MT indicated mutant clones. Symbol 1 indicates allelic biais in gene expression.

Table S2. List of autosomal genes identified (by at least one SNV >80/20%) as random monoallelically expressed in human cell clones. Each column represents an individual clone. WT indicated wild-type clones and MT indicated mutant clones. Symbol 1 means allelic biais in gene expression.

Table S3: List of RMAE genes selected for their frequency modulations between the wild-type and mutated groups, and used for clustering of wild-type and mutant cell clones.

Table S4. Gene list enrichment analysis using ChEA_2016, Trrust_transcription factor_2019, Encode_ChEA consensus Factor transcriptions_2015, Encode_Histone modifications_2015, and Wikipathways_2019 datasets in EnrichR. Only significant data (adjusted $p<0.05$) are indicated.

Table S5. Genes identified as overexpressed (adjusted $p<0.05$) in mutated versus wild-type clones. Each column represents an individual clone. WT indicated wild-type clones and MT indicated mutant clones. The normalized read count value from DESeq2 package was indicated for each gene as well as the mean value (mean, in light grey).

Table S6. Genes identified as underexpressed (adjusted $p<0.05$) in mutated versus wild clones.Each column represents an individual clone. WT indicated wild-type clones and MT indicated mutant clones. The normalized read count value from DESeq2 package was indicated for each gene as well as the mean value (mean, in light grey).

Fig.1. Distributions of expression levels (log 2 RPKM values) of all genes (dark grey) and RMAE genes (light grey) are shown for mutant (left side) and wild-type (right side) clones. (A) using the list of autosomal genes identified (by at least one SNV >90/10%) as random monoallelically expressed in human cell clones (Table S1); (B) using the list of autosomal

genes identified (by at least one SNV >80/20%) as random monoallelically expressed in human cell clones (Table S2). Because these data did not follow a normal law, a non-parametric test (Wilcoxon-Mann-Whitney test) was performed.

Fig.2. Distributions of gene expression levels (log 2 RPKM values) in mutant (dark grey) and wild-type (light grey) clones are shown for all genes(left side) and RMAE genes (right side). (A) using the list of autosomal genes identified (by at least one SNV >90/10%) as random monoallelically expressed in human cell clones (Table S1); (B) using the list of autosomal genes identified (by at least one SNV >80/20%) as random monoallelically expressed in human cell clones (Table S2). Because these data did not follow a normal law, a non-parametric test (Wilcoxon-Mann-Whitney test) was performed.

Fig. 3.A. Unsupervised hierarchical clustering of fibroblast clones according to their random monoallelic gene expression pattern using the "biased-RMAE" list (393 SNV_genes). WT clones are wild-type clones and MUT are mutant clones. B. Supervised hierarchical clustering of fibroblast clones according to their random monoallelic expression pattern for genes in the "biased-RMAE" list (94 SNV_genes).

Fig. S1. Sanger sequencing of genomic DNA and cDNA clones (A): * position of the SNV in the *MECP2* gene. This mutation is responsible for Rett syndrome in the RTT patient, (B): * position of the SNVin the *HSPA2* gene, (C): * position of the SNV in the *SPATA22* gene.