12-1-2014

Retrotransposon-based profiling of mammalian epigenomes: DNA methylation of IAP LTRs in embryonic stem, somatic and cancer cells

Arundhati Bakshi
Louisiana State University

Joomyeong Kim
Louisiana State University

Recommended Citation

Bakshi, A., & Kim, J. (2014). Retrotransposon-based profiling of mammalian epigenomes: DNA methylation of IAP LTRs in embryonic stem, somatic and cancer cells. *Genomics, 104* (6), 538-544. https://doi.org/10.1016/j.ygeno.2014.09.009

Genomics 104 (2014) 538–544

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Retrotransposon-based profiling of mammalian epigenomes: DNA methylation of IAP LTRs in embryonic stem, somatic and cancer cells

Arundhati Bakshi, Joomyeong Kim *

Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA

ARTICLE INFO

Article history:
Received 13 June 2014
Accepted 21 September 2014
Available online 29 September 2014

Keywords:
IAP LTR
DNA methylation
Epigenomes
Epigenetic biomarkers
Retrotransposons
Endogenous retrovirus

ABSTRACT

In the current study, we have used HT-TREBS to individually analyze the DNA methylation pattern of 4799 IAP LTR retrotransposons in embryonic stem, somatic and Neuro2A cells. According to the results, half of the loci within this family show constant methylation patterns between the three cell types whereas the remaining half display variable levels of methylation. About half of the variably methylated IAP LTRs tend to be hypomethylated in ES cells, and nearly all in this group are hypomethylated in Neuro2A cells. The observed hypomethylation in both cell types occur in a non-uniform, locus-specific manner and to various degrees of severity, with some of them being easily detectible by COBRA. Overall, this study demonstrates the feasibility of HT-TREBS to study DNA methylation changes at retrotransposons in a locus-specific manner in multiple cell types and further suggests the potential utility of this technique in developing epigenetic biomarkers for tracking disease progression.

1. Introduction

Mammalian genomes have accumulated a large number of retrotransposons during evolution, making up about half of the genome in any given species [1,2]. These retrotransposons are usually repressed by two main mechanisms. In germ cells, DNA methylation is the major mechanism which represses the majority of retrotransposons [3,4]. Later, in fertilized eggs, the germ cell-driven DNA methylation is removed, and subsequently, retrotransposons are temporarily repressed by various histone modifications [5–9]. During the implantation stage, however, these transient histone marks are again replaced by DNA methylation, a more stable and permanent modification. Thus, histone modifications are usually responsible for the temporary repression of retrotransposons in transient stem cell populations whereas DNA methylation is responsible for the more stable and permanent repression in further committed and differentiated cell populations [4,10,11].

It is well known that improper epigenetic regulation of retrotransposons can result in unwanted consequences due to their transposition [12] and their ability to influence adjacent gene activity by acting as alternative promoters [13] or disrupting the endogenous exon structure. The latter has been well demonstrated by the two cases of mouse epialleles, where the ectopic expression of two independent IAP LTRs (Intracisternal A Particle Long Terminal Repeats) interfere with the transcription of two endogenous loci, A^{vy} (agouti viable-yellow) and $Axin^{fu}$ (Axin-fused). In the mouse, these ectopic expressions are responsible for visible phenotypic consequences, such as coat color variation for A^{vy} and tail kinkedness variations for $Axin^{fu}$ [14,15]. Similar hypomethylation on retrotransposons also is seen in cancer genomes in humans [16,17], which may result in functional consequences in cancer progression [18,19]. However, at the same time, hypomethylation of retrotransposons also provides a unique opportunity for mammalian genome evolution through allowing their co-option into the host genome as regulatory elements [20]. A recent study suggests that endogenous retroviruses (ERVs) in the human genome may have played an important role in the evolution of the central nervous system by affecting gene expression in key areas of the brain [21]. Overall, DNA methylation at retrotransposons is not only of interest toward understanding complex human disorders with an epigenetic underpinning, including cancers [22,23], but also for their role in shaping mammalian evolution.

Given the importance of epigenetic modifications on retrotransposons, much research is currently focused on understanding the extent to which they are repressed in normal as well as diseased cells and tissues. However, due to the paucity of methods allowing for deep sequencing coverage at repeat elements and flanking loci, it has thus far been difficult to perform a systematic analysis of the effect of DNA methylation at individual loci in a locus-specific manner. To address these issues, we have developed a new protocol termed HT-TREBS (High-Throughput Targeted Repeat Element Bisulfite Sequencing) [24], which is designed to survey the DNA methylation levels of a large

Abbreviations: COBRA, Combined Bisulfite Restriction Analysis; ERV, endogenous retrovirus; ES, embryonic stem cells; EST, expressed sequence tag; H3K9me3, histone 3 lysine 9 trimethylation; HT-TREBS, High-Throughput Targeted Repeat Elements Bisulfite Sequencing; IAP LTR, Intracisternal A Particle Long Terminal Repeat.

* Corresponding author. Fax: +1 225 578 2597.
E-mail address: jkim@lsu.edu (J. Kim).

http://dx.doi.org/10.1016/j.ygeno.2014.09.009
0888-7543/

number of interspersed repeat elements on an individual-locus basis. According to the results from a set of pilot experiments analyzing mouse IAP LTRs, the majority of this retrotransposon family is properly repressed in normal cells, but a very small number of IAP LTRs (about 5%) escape this repression mechanism [24,25]. In the current study, this protocol has been utilized to characterize the DNA methylation profile of this retrotransposon family in AB2.2 embryonic stem and Neuro2A cancer cells and subsequently compared with the somatic cell data derived from Ekram and Kim [24]. The results revealed that the DNA methylation levels of a large fraction of IAP LTRs, though not all, are dynamically fluctuating between different cell types, with some of them being easily detectible by quick methods such as COBRA. Overall, these analyses demonstrate the feasibility of using HT-TREBS to find potential epigenetic biomarkers of cell state in multiple different cell types.

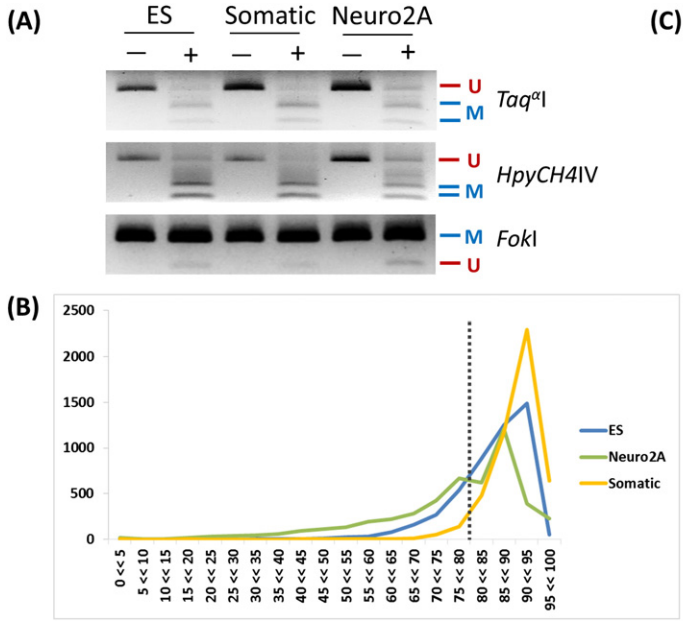
2. Results

2.1. HT-TREBS analyses of IAP LTRs in ES, somatic and Neuro2A cells

In the current study, we analyzed and compared the DNA methylation levels of the IAP LTR family using the following three cell types: embryonic stem (ES), somatic and Neuro2A cancer cells. Genomic DNA isolated from ES cells (AB2.2 cell, 129 origin, from Baylor College of Medicine) and a neuroblastoma cell line of strain A origin (Neuro2A) were treated according to an established protocol of HT-TREBS [24]. In brief, the genomic DNA was fractionated by sonication, ligated to an adaptor and modified with bisulfite conversion protocol [26]. It was then amplified with a PCR scheme designed to enrich IAP LTR-genomic regions. This was accomplished by designing one of the primers complementary to the 24-bp portion within the U3 region of the LTR that is conserved between five subtypes of IAP LTR (IAPLTR1, IAPLTR1a, IAPLTR2, IAPLTR2a, and IAPLTR2b), and is devoid of any CG dinucleotides. The region is located approximately 150 bp downstream of TGTTGGG (denoting the extreme 5′ end of the LTR), and includes the CAT box. The amplified library was finally size-fractionated with agarose gel electrophoresis. The prepared libraries were sequenced with a Next-Generation-Sequencing (NGS) protocol.

We have obtained 6.3 and 6.8 million raw sequence reads from ES and Neuro2A cells, respectively. Each set of bisulfite-converted sequence reads was mapped and processed to derive the methylation level of each IAP LTR locus as described below. In brief, individual raw sequence reads were mapped to a custom database containing the sequences of about 10,000 IAP LTR loci. In this database, each locus is represented with a 1030-bp sequence covering the 330-bp LTR plus two 350-bp flanking regions. A set of sequence reads mapped to a given locus was subsequently used for calculating its methylation level. We have successfully obtained the methylation values for 5637 and 5575 loci of IAP LTRs for ES and Neuro2A cells respectively. In both sets, 100 mapped raw sequences were used on average to derive methylation level of each IAP LTR locus, covering an average of 7 CpG sites per locus. Next, these two new data sets were compared with the somatic data set representing the average methylation values of IAP LTRs that were individually derived from the three organs (brain, liver and kidney) of a two-month-old C57BL/6N mouse (Fig. 1) as well as to the data sets derived from each of the organs representing the three tissue layers individually [24]. According to our analyses with the individual tissues (Supplemental Fig. 1), the liver and kidney showed overall more similarity in DNA methylation levels between each other than with the brain. However, on the whole, over 95% of all IAP LTRs were methylated at similar levels in all three organs, as observed by Ekram and Kim [24]. Therefore, in order to compare ES and Neuro2A cells to a representative DNA methylation value in normal somatic cells, all further analyses were performed with the average somatic data set. Finally, the three data sets (ES, somatic and Neuro2A) were compiled to derive 4799 IAP LTRs with a minimum of 15× coverage at each CpG site that were represented in all three cell types for further comparisons (Supplemental Data 1 and Fig. 1).

Initial tabulation of these data sets revealed the following immediate conclusions. The overall DNA methylation levels of IAP LTRs are variable among the three samples: the DNA methylation levels in somatic cells are the highest, followed by ES and Neuro2A cells, as was indicated by



(C)

% methylation	ES	Somatic	Neuro2A
0 >> 5	1	1	17
5 >> 10	0	0	7
10 >> 15	0	0	6
15 >> 20	1	1	19
20 >> 25	1	0	32
25 >> 30	0	0	37
30 >> 35	4	0	44
35 >> 40	7	0	60
40 >> 45	7	1	91
45 >> 50	8	1	110
50 >> 55	23	1	133
55 >> 60	33	4	193
60 >> 65	80	5	223
65 >> 70	158	11	281
70 >> 75	265	51	421
75 >> 80	537	140	668
80 >> 85	883	481	622
85 >> 90	1251	1167	1212
90 >> 95	1491	2297	393
95 >> 100	49	638	230

Fig. 1. HT-TREBS analyses of IAP LTRs in ES, somatic and Neuro2A cells. (A) COBRA analyses of the IAP LTRs in ES, somatic and Neuro2A cells. The bisulfite-treated DNA was amplified with a primer set targeting the entire family of this retrotransposon. Restriction enzymes recognizing methylated CpG (*Taq*$^{\alpha}$I and *HpyCH4*IV) as well as unmethylated CpG dinucleotides after bisulfite treatment and PCR (*Fok*I) indicate hypomethylation in Neuro2A cells compared to ES and somatic (brain) cells. Expected DNA fragment sizes based on CpG methylation status are marked by "U" (unmethylated) or "M" (methylated). The no-enzyme control for each restriction digestion is indicated by a minus sign. (B) The entire set of 4799 IAP LTRs in each cell type was first grouped into different bins based on their methylation levels, and subsequently presented as a line graph. A series of bins on X-axis represent a gradual increment of DNA methylation from 0 to 100% whereas the values on Y-axis represent the number of IAP LTRs in each bin. All IAP LTRs methylated under 80% (left of the dotted line) were considered hypomethylated. The actual tabulated numbers of IAP LTRs in each bin for the three cell types are also presented as a table (C).

the result of COBRA (COmbined Bisulfite Restriction Analysis [27]), showing a substantial level of DNA hypomethylation in Neuro2A cells as compared to the much greater levels of DNA methylation in both ES and somatic cells (Fig. 1A). Next, in order to quantify the level of DNA methylation difference, the set of 4799 loci was divided into two groups: hypermethylated (≥80% methylation level) and hypomethylated (<80% methylation level). IAP LTRs methylated above 80% showed no major differences among them, whereas those under 80% methylation level showed some unique patterns of hypomethylation, as discussed previously by Ekram and Kim [24]. The same was largely true for thresholds under 80% (70–75%), however the number of loci in the hypomethylated category was too few for meaningful comparisons with the hypermethylated category under these conditions. Therefore, the empirically selected 80% threshold, which is consistent with that established by Ekram and Kim [24], ensured an adequate number of loci in each group while staying close to the base of the curve shown in Fig.1B. According to the results, about 96% of IAP LTRs (4587 loci) showed at least 80% methylation level in somatic cells whereas only 77% (3689 loci) and 51% (2465 loci) were hypermethylated in ES and Neuro2A cells respectively (Fig. 1BC). In somatic cells, only 4% of the loci (212 IAP LTRs) belonged to the hypomethylated category, while 23% of IAP LTRs (1110 loci) and nearly half (2334 loci) showed less than 80% methylation levels in ES and Neuro2A cells respectively (Fig. 1BC). Overall, these initial analyses demonstrate the severe hypomethylation of IAP LTR loci in Neuro2A cells which is in stark contrast with the much higher levels of methylation seen in ES and somatic cells.

2.2. Comparative analyses of DNA methylation profiles in ES and Neuro2A vs. somatic cells

The DNA methylation pattern of each individual IAP LTR locus in ES and Neuro2A cells was analyzed in comparison to somatic cells using a two-dimensional dot plot display (Fig. 2). In a given plot, each dot represents one IAP LTR with two values: X- and Y-axis values representing DNA methylation levels derived from two comparing cell types. As shown in Fig. 2A, the majority of dots (IAP LTRs) are located within the 80–100% methylation range in both ES and somatic cells. Yet, a substantial fraction of IAP LTRs are spread horizontally over the methylation ranges less than 80% in ES cells but still greater than 80% in somatic cells. In terms of actual number, this group is estimated to contain around 932 IAP LTRs based on the initial tabulation (Fig. 1C). A small number of IAP LTRs also show even greater degrees of DNA methylation difference between ES and somatic cells based on their location in the upper left corner of the dot plot where the DNA methylation levels are less than 40% in ES but greater than 80% in somatic cells. Another set of IAP LTRs is located along the diagonal line connecting the lower left to upper right corner. These dots belong to the group of IAP LTRs showing hypomethylation regardless of cell type, whose methylation values, interestingly enough, are also known to be variable between different individuals [24,25]. Overall, this dot plot analysis highlights the observation that a large portion of IAP LTRs are hypomethylated in ES cells but not in somatic cells (Fig. 2A).

Comparing DNA methylation patterns of IAP LTRs in Neuro2A and somatic cells (Fig. 2B) indicates that a much greater number of IAP LTRs, around 2152 elements, show different levels of methylation between the two cell types. A majority of these loci (2137 elements), represented by a large number of dots spreading horizontally to the left from the upper right corner in the dot plot, show greater than 80% methylation in somatic cells but less than 80% methylation in Neuro2A cells. It is interesting to note that all the individual IAP LTRs of this group have different levels of DNA methylation change: some have much greater levels, such as the ones in the top left corner which are nearly completely methylated in somatic cells but completely unmethylated in Neuro2A cells, whereas others have more modest levels of change. This individuality of DNA methylation change is even more contrasting in the case of the other remaining half of the IAP LTRs, which still show similar levels of DNA methylation between somatic and Neuro2A cells. Overall, this analysis indicates that about half, but not all, of the IAP LTRs are hypomethylated in Neuro2A compared to somatic cells. Given all the different degrees and the various patterns of changes, the observed methylation difference between the two cell types is thought to be locus-specific rather than uniform, which may be reflecting the fact that not all IAP LTRs are equally sensitive to cell type in terms of their DNA methylation.

2.3. DNA methylation pattern-based grouping of IAP LTRs

The various different methylation levels of each individual IAP LTR in the three different cell types were further studied by Venn diagram analysis using Venny [28]. In this analysis, each IAP LTR has one of the two states in a given cell type, either High (≥80%) or Low (<80%) level of methylation based on the division established from Fig. 1BC. Since each IAP LTR is present in three different cell types, ES, somatic and Neuro2A, eight different combinations of DNA methylation states are possible for any given IAP LTR (Fig. 3). The first Venn diagram was

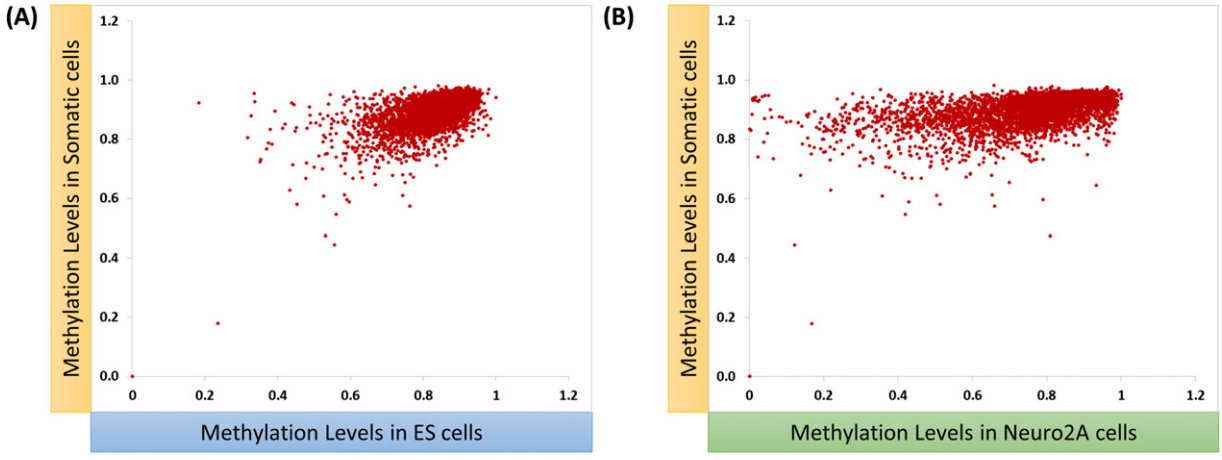
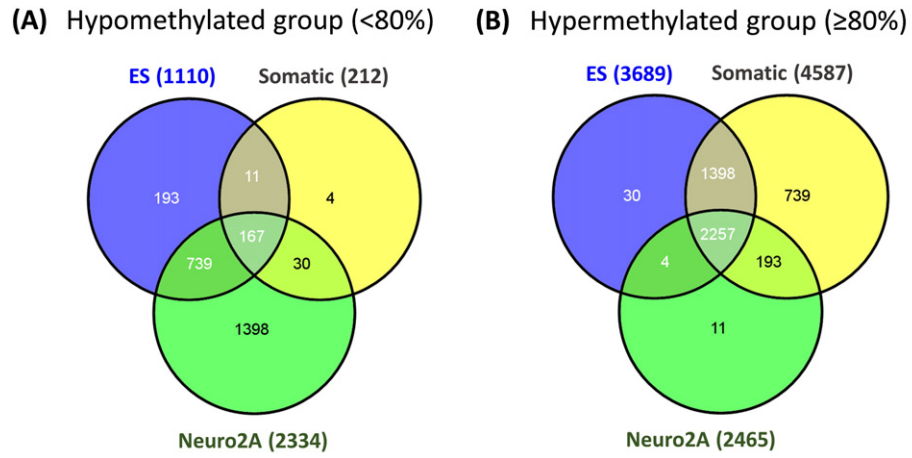


Fig. 2. DNA methylation profiles of IAP LTRs in ES and Neuro2A cells compared to somatic cells. The methylation levels of IAP LTRs were compared between ES and somatic cells (A) and between somatic and Neuro2A cells (B). For both scatter plot analyses, each given dot represents one IAP LTR, and the position of the dot (on the X and Y plane) indicates the methylation levels from the two cell types that are being compared. The dots spreading horizontally toward the left on the first plot indicates that a substantial fraction of IAP LTRs is hypomethylated in ES cells but not in somatic cells. Similarly, a large number of dots spreading horizontally to the left from the right corner of the second graph represent the hypomethylation of half of IAP LTRs in Neuro2A cells, but their methylation levels are still greater than 80% in somatic cells.



(C)

Group (No.)	ES	Somatic	Neuro2A	Group (No.)	ES	Somatic	Neuro2A
LLL (167)	Low	Low	Low	HLL (30)	High	Low	Low
LLH (11)	Low	Low	High	HLH (4)	High	Low	High
LHL (739)	Low	High	Low	HHL (1398)	High	High	Low
LHH (193)	Low	High	High	HHH (2257)	High	High	High

Fig. 3. DNA methylation pattern-based grouping of IAP LTRs. A given IAP LTR has one of two states, either Low (<80% methylation) or High (≥80% methylation) in each of the three cell types, resulting in eight total possible categories of DNA methylation patterns. The first Venn diagram was constructed through comparing the three sets of IAP LTRs showing less than 80% methylation level (Low) in the three cell types (A). The second Venn diagram was derived from comparing the three sets of IAP LTRs with their methylation levels being at least 80% (B). The resulting eight different categories of IAP LTRs are presented along with the actual numbers of loci in (C).

constructed through comparing three sets of IAP LTRs that show less than 80% methylation level in all three cell types (Fig. 3A). This, subsequently, derived seven groups of IAP LTRs, which are represented by the seven small sections of the first Venn diagram. The three non-overlapping sections indicate the three groups with one Low state in the three different cell types (LHH, HLH, HHL in Fig. 3C), whereas the three overlapping sections between two cell types (or circles) indicate the groups with two Low states (LLH, LHL, HLL). One remaining group is overlapped by three Low states (LLL), meaning that this group of IAP LTRs maintains Low methylation in all three cell types. One final group is missing in the first Venn diagram, but is found in the center section of the second Venn diagram that has been constructed with the three sets of IAP LTRs showing greater than 80% methylation level. The final group of IAP LTRs is characterized by greater than 80% methylation levels in all three different cell types (HHH) (Fig. 3B).

The actual numbers of IAP LTRs constituting each of the eight groups are summarized in Fig. 3C. According to this summary, there are three groups which are represented by very small numbers of IAP LTRs, which include LLH (11), HLL (30) and HLH (4). Of the remaining five groups, two of them, LLL (167) and HHH (2257), show constant DNA methylation and make up 3.5 and 47% of the entire IAP LTR family, respectively. The other three groups with fluctuating methylation levels altogether make up the remaining half: the HHL group (29%, 1398 members), the LHL group (15%, 739 members) and the LHH group (4%, 193 members).

These IAP LTRs within these Venn diagram groups were further analyzed in order to ascertain any biological significance behind the various patterns of methylation observed for these IAP LTRs. They were studied with regards to expressed sequence tags (ESTs), histone modifications (data not shown), gene association and genomic position preference along with the 20 kb flanking sequence in order to survey a representative portion of the mouse genome covered by the IAP LTRs (Supplemental Fig. 2). No specific epigenetic mark distinguished any one of the groups from others. IAP LTRs within all groups were associated with 0–2 genes, albeit only distally, with most IAP LTRs being positioned within 50–500 kb of transcription start sites (TSS) (Supplemental Fig. 2AB) as previously observed by Ekram and Kim [24]. There was no genomic location preference observed for any these groups of IAP LTRs with respect to their chromosomal location or distance to TSS (Supplemental Fig. 2BC) which distinguished them from each other, after accounting for the large variation in sample sizes. In summary, according to this series of Venn diagram analyses, the DNA methylation levels of half of the IAP LTRs are static and constant whereas the remaining half fluctuates between different cell types without any gene association or chromosomal position preference.

2.4. COBRA analyses of representative IAP LTR loci

We employed COBRA as an independent method of assessing DNA methylation at some of the loci representing the various Venn diagram groups. It provided us with a rapid view of the differences between cell types through testing the methylation level associated with 1–3 CpG sites associated with the retrotransposon. Bisulfite-treated brain DNA was chosen as a representative organ for somatic cells.

One representative example of COBRA analysis of the locus IAPLTR1a_Mm-ERVK-LTR chrX: 39388013-39388339, associated distally with the gene *Xiap*, is shown along with the corresponding HT-TREBS data set (Fig. 4A). The black triangles over the heatmaps indicate one of the two *HpyCH4*IV restriction sites in the amplified portion of the IAP LTR which was also sequenced by HT-TREBS. This particular CpG site was methylated in 133 reads out of a total of 191 reads in ES cells (70% methylation), 89 out of 137 in the brain (65% methylation) and 0 out of 187 reads (0% methylation) in Neuro2A cells. The stark difference in methylation states at this position between Neuro2A, and ES and somatic cells, is clearly demonstrated by the COBRA which shows an undigested DNA fragment for Neuro2A whereas complete digestion is observed for the ES and somatic cell samples.

COBRA results of four other loci from various different groups are shown in Fig. 4B. A full description of each of these loci, along with the gene they are associated with, may be found in Supplemental Data 2. It becomes apparent from these COBRA analyses that the difference in methylation level between Neuro2A and other cell types at the tested

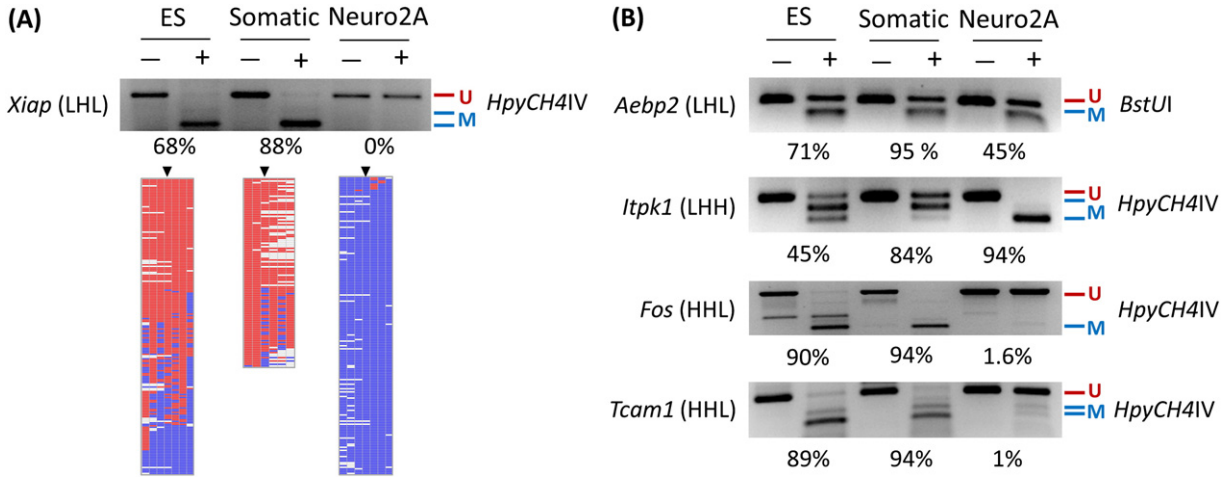


Fig. 4. COBRA analyses of representative IAP LTR loci. (A) Representative COBRA analysis shown along with heatmaps from HT-TREBS analysis. Red boxes indicate methylated CpG sites and blue boxes indicate unmethylated CpG sites in the heatmaps in the bottom panel, with a blank box indicating insufficient sequencing coverage. Black triangles indicate one of the restriction sites, which was also sequenced by HT-TREBS. (B) COBRA analysis from four randomly chosen loci representing the various Venn diagram groups. All loci in (A) and (B) are named after the gene they are (distally) associated with and the Venn diagram group they belong to. A full description of each locus may be found in Supplemental Data 2. The percent numbers shown under the images indicate the average methylation level of this locus in the indicated cell type. The expected fragments sizes following the restriction enzyme digestion as indicated are marked by "U" (unmethylated) or "M" (methylated) based on the CpG site methylation status. The no-enzyme control for each restriction digestion is indicated by a minus sign.

CpG sites are greater at some IAP LTRs compared to others. Two of the loci presented in this panel (IAP LTRs associated with *Fos* and *Tcam1*) show severe hypomethylation in Neuro2A cells, such that the undigested DNA fragments are easily detectible by COBRA. In contrast, the IAP LTR located within the second intron of *Itpk1* shows marked hypermethylation in Neuro2A, but much lower levels of methylation in ES and somatic cells. In general, these results are consistent with the detailed DNA methylation data obtained from these loci through HT-TREBS. However, not all loci can be tested as easily using COBRA due to technical limitations, such as low sensitivity and resolution capacity, as seen in the case of the IAP LTR associated with *Aebp2*. In conclusion, this set of analyses demonstrates that the DNA methylation difference at certain loci, though not all, could be verified using an independent method, COBRA.

3. Discussion

In the current study, the DNA methylation patterns of the mouse retrotransposon family IAP LTR have been studied in three distinct cell types: embryonic stem (ES), somatic and a neuroblastoma cell line, Neuro2A (Fig. 1). Previous studies on DNA methylation of retrotransposons have focused on analyzing them as an entire group, without regard for locus-specific variations in methylation patterns. Therefore, thus far, it was unclear whether all loci behaved similarly in different cell types or whether some were more susceptible to changes in DNA methylation levels based on cell type and stage. Using HT-TREBS [24], we have now been able to analyze retrotransposons for the first time in a locus-specific manner, which has revealed its non-uniform nature of hypomethylation in ES and Neuro2A cells (Fig. 2). The results suggest that not all IAP LTRs are equally sensitive to cell types and stages in terms of their DNA methylation level. Certain IAP LTRs showed severe hypomethylation in ES and Neuro2A cells, while others were stably methylated at either high or low levels (Figs. 3 and 4). Interestingly, much greater levels of hypomethylation were observed in Neuro2A cells than in ES cells, which may be due to two reasons. First, it may be due to strain related differences between strain A, 129 and C57BL/6N or cell line related artifacts which are not relevant in vivo. Second, Neuro2A cell line is derived from a murine neuroblastoma and hence is likely to show some cancer related signatures. From this perspective, the severe DNA hypomethylation of Neuro2A cells is consistent with various other reports about genome-wide hypomethylation of retrotransposons in cancer cells [17,18,29–32]. According to the current study, however, it is apparent that not all retrotransposons are likely to behave in a similar fashion when a cell undergoes transformations such as in development or cancer. Only a subset of all loci are likely to undergo changes in DNA methylation in response to such stresses, with varying degrees of severity depending upon the locus.

The hypomethylation we observed in ES cells had also been previously predicted based on the propensity for temporary silencing marks in these transient cell populations, even on retrotransposons [5,6]. In fact, a recent whole-genome bisulfite sequencing analysis by Stadler et al. [33] showed that repeat elements comprised approximately 34% of all "low-methylation regions" (55–75% methylation) and 7% of all "unmethylated regions" (<55% methylation). The scope of this study, however, does not allow it to focus on the composition of the repeat elements analyzed within these under-methylated groups. Here we show with greater specificity and sequencing depth for IAP LTR retrotransposons that 25% of the loci tend to be hypomethylated (<80% methylation level) in these 129-derived ES cells. Such levels of hypomethylation in ES cells, however, is not a sign of increased transcriptional activity since only 0.23% of all ESTs in a blastocyst are associated with IAP elements [34]. Furthermore, *Dnmt1*-deficient ES cells are only found to overexpress IAP mRNA upon differentiation and not in their pluripotent state [11]. This apparent paradox is explained by recent studies which have shown that KAP1, SETDB1 and HP1 act in tandem to silence retrotransposons in a DNA methylation-independent manner through H3K9 trimethylation in ES cells [7–9]. These temporary, but strong, silencing histone marks sustain proper repression while still allowing a significant fraction of the retrotransposons to remain hypomethylated. A subset of these hypomethylated retrotransposons then have the opportunity to become co-opted into the host genome as regulatory elements. Instances of such adaptation of retrotransposons have been previously observed in ES cells [35,36] as well as other tissues [21,37] in both humans and mice. It is possible, then, that a small fraction of the 1110 hypomethylated IAP LTRs detected in this study may also function as regulatory elements for nearby genic regions, much as the stem cell-related promoters and enhancers originating from ERVs that were recently identified by Fort et al. [35].

In conclusion, through this study, we have been able to demonstrate the feasibility of using HT-TREBS as a Next-Generation-Sequencing based approach to reliably assess the DNA methylation levels of

retrotransposons in multiple different cell types. With greater specificity for targeted repeat elements and deep sequencing coverage, HT-TREBS may be used for identifying those retrotransposons which may be important for genome evolution or be associated with disease states in humans as well as mice. For instance, our observations in Neuro2A cells (Fig. 2), which are potentially cancer-like, allow us to predict that only a subset of retrotransposons are likely to suffer from hypomethylation in most cancers with some facing much greater levels of demethylation compared to other loci. Moreover, based on the heterogeneity of cancer, it is predicted that the loci facing hypomethylation will most likely be different based on the origin of the tumor and to different degrees depending on the locus. HT-TREBS can be employed on a large-scale basis to test which retrotransposons suffer hypomethylation and to what levels in different types of cancers. Since certain loci are likely to be severely hypomethylated, such as the ones demonstrated in Fig. 4, they can be easily and quickly detected by COBRA and used as cancer biomarkers. Furthermore, given that certain loci are hypomethylated at extreme levels in Neuro2A cells (0–5% methylation), it is possible that demethylation events at these IAP LTRs are additive over time. If one may indeed consider Neuro2A cells to be somewhat reflective of a neuroblastoma cell state, this could indicate that certain specific retrotransposable loci may start losing methylation very early during tumorigenesis events. Such loci, as biomarkers, could be useful for tracking cancer progression. The sensitivity of HT-TREBS also provides us with the opportunity to find the subset of retrotransposons which are potentially associated with tumor suppressors or oncogenes and can affect their expression patterns, since early demethylation at these loci could indicate their status as drivers of cancer progression. Overall, the various different patterns of DNA hypomethylation observed in ES and Neuro2A cells via HT-TREBS, in this study, introduce the possibility of identifying specific repeat elements which may serve as biomarkers for different cell states and be especially useful for tracking disease progression.

4. Materials and methods

4.1. HT-TREBS analyses of DNA isolated from ES and Neuro2A cells

For the current study, the HT-TREBS protocol developed and employed by Ekram and Kim to characterize somatic tissues (brain, liver and kidney) from a two-month-old C57BL/6N mouse [24] was applied to ES (AB2.2 from Baylor College of Medicine) and Neuro2A cells. For each sample, 1 μg of the purified genomic DNA was fragmented with sonication (Bioruptor NGS, Diagenode) to generate a pool of DNA fragments with the peak size being around 700 bp in length. The fragmented DNA was end-repaired using the NEBNext® End Repair Module (New England BioLabs), and ligated to custom-made Ion Torrent 'A' adaptors in which all the cytosines have been methylated (Integrated DNA Technologies). The adaptor-ligated DNA fragments were further size-selected to remove any excess adaptors and DNA fragments smaller than 300 bp in length using the Agencourt AMPure XP beads (Beckman Coulter). The adaptor-ligated DNA library was modified using the bisulfite conversion reaction according to the manufacturer's protocol (EZ DNA Methylation™ kit, Zymo Research). The bisulfite-converted library was used as template for a round of PCR (Maxime PCR Premix Kit, Intron Biotech) using the following two primers: the forward primer (5′-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3′) designed to bind to the 5′ end of the 'A' adaptor region and the reverse primer (5′-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAG TCGGTGAT^CTCCCTAATTAACTACAACCCATC-3′) designed to bind to the 24-bp region that is well conserved among the IAP LTR subtypes (IAP LTR1, 1a, 2, 2a, and 2b). The sequence in the 5′-side of the reverse primer marked by ^ corresponds to the 'P1' adaptor, which is part of the amplification strategy used for the Ion Torrent NGS scheme (Ion Torrent, Life Technologies). The PCR product was finally size-selected for a range of 250–300 bp in length using agarose gel electrophoresis. Each of the two PCR products was then individually sequenced in the Ion Personal Genome Machine (PGM) Sequencer using Ion 318 Chips (Ion Torrent, Life Technologies). The sequence reads generated from the two Ion PGM runs were individually mapped using the aligner Bowtie2 [38] to a curated reference genome made up of bisulfite-converted IAP LTR sequences. The mapped reads were filtered through several custom Perl scripts to extract only the sequences covering the IAP LTR and flanking unique regions. The filtered reads from each sample were separately analyzed using the BiQAnalyzerHT tool [39]. The detailed information regarding Perl scripts and bioinformatic pipelines are available upon request. ES and Neuro2A datasets have been added to the NCBI's Gene Expression Omnibus [40] data repository and can be viewed under the accession number GSE60007 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60007).

4.2. COBRA analyses

Approximately 500 ng of purified genomic DNA from AB2.2 ES cell, Neuro2A and C57BL/6N brain was treated using the EZ DNA Methylation™ kit according to the manufacturer's protocol (Zymo Research). This bisulfite-treated DNA was then used for the COBRA (COmbined Bisulfite Restriction Assay) analyses. Specifically, 1 μL (~20 ng) of the converted DNA was used for methylation-unbiased PCR (Maxime PCR Premix Kit, Intron Biotech) using bisulfite primers which lacked any CpG dinucleotides or cytosines (all cytosines were converted to thymines). Next, amplified DNA was digested using appropriate restriction enzymes which recognized at least one CpG site as part of their recognition sequence (New England BioLabs). All primers, PCR conditions and amplified region coordinates (according to the mm9 version of the mouse genome) may be found in Supplemental Data 2, along with the restriction enzymes used for COBRA.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2014.09.009.

Competing Interests

None declared.

Acknowledgments

We would like to thank Drs. Scott Herke and Mark Batzer for their help on NGS sequencing, Drs. Muhammad B. Ekram and Dong-Ha Oh for their help on bioinformatics, and Isabel Lorenzo at Baylor College of Medicine for providing ES cells. We also thank Dr. Hana Kim, Dr. Hongzhi He, Dr. Su Man Lee, An Ye, B. Pinithi Perera, Wesley Frey and Corey Bretz for their careful reading and discussion of the manuscript. This research was supported by the National Institute of Health (J.K. R01-GM066225 and R01-GM097074).

References

[1] A.T. Chinwalla, L.L. Cook, K.D. Delehaunty, G.A. Fewell, L.A. Fulton, R.S. Fulton, T.A. Graves, L.W. Hillier, E.R. Mardis, J.D. McPherson, et al., Initial sequencing and comparative analysis of the mouse genome, Nature 420 (2002) 520–562.
[2] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.
[3] J.A. Yoder, C.P. Walsh, T.H. Bestor, Cytosine methylation and the ecology of intragenomic parasites, Trends Genet. 13 (1997) 335–340.
[4] H. Sasaki, Y. Matsui, Epigenetic events in mammalian germ-cell development: reprogramming and beyond, Nat. Rev. Genet. 9 (2008) 129–140.
[5] J.H.A. Martens, R.J. O'Sullivan, U. Braunschweig, S. Opravil, M. Radolf, P. Steinlein, T. Jenuwein, The profile of repeat-associated histone lysine methylation states in the mouse epigenome, EMBO J. 24 (2005) 800–812.
[6] T.S. Mikkelsen, M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T.K. Kim, R.P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E.S. Lander, B.E. Bernstein, Genome-wide maps of chromatin state in pluripotent and lineage-committed cells, Nature 448 (2007) 553–560.

[7] T. Matsui, D. Leung, H. Miyashita, I.A. Maksakova, H. Miyachi, H. Kimura, M. Tachibana, M.C. Lorincz, Y. Shinkai, Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET, Nature 464 (2010) 927–931.
[8] H.M. Rowe, J. Jakobsson, D. Mesnard, J. Rougemont, S. Reynard, T. Aktas, P.V. Maillard, H. Layard-Liesching, S. Verp, J. Marquis, F. Spitz, D.B. Constam, D. Trono, KAP1 controls endogenous retroviruses in embryonic stem cells, Nature 463 (2010) 237–240.
[9] I.A. Maksakova, P.J. Thompson, P. Goyal, S.J. Jones, P.B. Singh, M.M. Karimi, M.C. Lorincz, Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells, Epigenetics & chromatin 6 (2013) 15.
[10] C.P. Walsh, J.R. Chaillet, T.H. Bestor, Transcription of IAP endogenous retroviruses is constrained by cytosine methylation, Nat. Genet. 20 (1998) 116–117.
[11] L.K. Hutnick, X. Huang, T.C. Loo, Z. Ma, G. Fan, Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism, J. Biol. Chem. 285 (2010) 21082–21091.
[12] D.C. Hancks, H.H. Kazazian Jr., Active human retrotransposons: variation and disease, Curr. Opin. Genet. Dev. 22 (2012) 191–203.
[13] G.J. Faulkner, Y. Kimura, C.O. Daub, S. Wani, C. Plessy, K.M. Irvine, K. Schroder, N. Cloonan, A.L. Steptoe, T. Lassmann, K. Waki, N. Hornig, T. Arakawa, H. Takahashi, J. Kawai, A.R. Forrest, H. Suzuki, Y. Hayashizaki, D.A. Hume, V. Orlando, S.M. Grimmond, P. Carninci, The regulated retrotransposon transcriptome of mammalian cells, Nat. Genet. 41 (2009) 563–571.
[14] N.C. Whitelaw, E. Whitelaw, How lifetimes shape epigenotype within and across generations, Hum. Mol. Genet. 15 (2) (2006) R131–137.
[15] R.L. Jirtle, M.K. Skinner, Environmental epigenomics and disease susceptibility, Nature reviews, Genetics 8 (2007) 253–262.
[16] G. Howard, R. Eiges, F. Gaudet, R. Jaenisch, A. Eden, Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice, Oncogene 27 (2008) 404–408.
[17] Y. Baba, C. Huttenhower, K. Nosho, N. Tanaka, K. Shima, A. Hazra, E.S. Schernhammer, D.J. Hunter, E.L. Giovannucci, C.S. Fuchs, S. Ogino, Epigenomic diversity of colorectal cancer indicated by LINE-1 methylation in a database of 869 tumors, Mol. Cancer 9 (2010) 125.
[18] N. Rodic, K.H. Burns, Long interspersed element-1 (LINE-1): passenger or driver in human neoplasms? PLoS Genet. 9 (2013) e1003402.
[19] E.M. Wolff, H.M. Byun, H.F. Han, S. Sharma, P.W. Nichols, K.D. Siegmund, A.S. Yang, P.A. Jones, G. Liang, Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer, PLoS Genet. 6 (2010) e1000917.
[20] R. Rebollo, M.T. Romanish, D.L. Mager, Transposable elements: an abundant and natural source of regulatory sequences for host genes, Annu. Rev. Genet. 46 (2012) 21–42.
[21] M. Suntsova, E.V. Gogvadze, S. Salozhin, N. Gaifullin, F. Eroshkin, S.E. Dmitriev, N. Martynova, K. Kulikov, G. Malakhova, G. Tukhbatova, A.P. Bolshakov, D. Ghilarov, A. Garazha, A. Aliper, C.R. Cantor, Y. Solokhin, S. Roumiantsev, P. Balaban, A. Zhavoronkov, A. Buzdin, Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene PRODH, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 19472–19477.
[22] V.K. Rakyan, T.A. Down, D.J. Balding, S. Beck, Epigenome-wide association studies for common human diseases, Nature reviews, Genetics 12 (2011) 529–541.
[23] P.A. Jones, S.B. Baylin, The epigenomics of cancer, Cell 128 (2007) 683–692.
[24] M.B. Ekram, J. Kim, High-Throughput Targeted Repeat Element Bisulfite Sequencing (HT-TREBS): genome-wide DNA methylation analysis of IAP LTR retrotransposon, PLoS ONE 9 (2014) e101683.
[25] M.B. Ekram, K. Kang, H. Kim, J. Kim, Retrotransposons as a major source of epigenetic variations in the mammalian genome, Epigenetics 7 (2012) 370–382.
[26] S.J. Clark, J. Harrison, C.L. Paul, M. Frommer, High sensitivity mapping of methylated cytosines, Nucleic Acids Res. 22 (1994) 2990–2997.
[27] Z. Xiong, P.W. Laird, COBRA: a sensitive and quantitative DNA methylation assay, Nucleic Acids Res. 25 (1997) 2532–2534.
[28] J.C. Oliveros, VENNY. An interactive tool for comparing lists with Venn Diagrams, 2007.
[29] K. Brennan, J.M. Flanagan, Is there a link between genome-wide hypomethylation in blood and cancer risk? Cancer Prev. Res. 5 (2012) 1345–1357.
[30] I. Katoh, S.I. Kurata, Association of endogenous retroviruses and long terminal repeats with human disorders, Front. Oncol. 3 (2013) 234.
[31] S.N. Akers, K. Moysich, W. Zhang, G. Collamat Lai, A. Miller, S. Lele, K. Odunsi, A.R. Karpf, LINE1 and Alu repetitive element DNA methylation in tumors and white blood cells from epithelial ovarian cancer patients, Gynecol. Oncol. 132 (2014) 462–467.
[32] K.D. Hansen, W. Timp, H.C. Bravo, S. Sabunciyan, B. Langmead, O.G. McDonald, B. Wen, H. Wu, Y. Liu, D. Diep, E. Briem, K. Zhang, R.A. Irizarry, A.P. Feinberg, Increased methylation variation in epigenetic domains across cancer types, Nat. Genet. 43 (2011) 768–775.
[33] M.B. Stadler, R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Scholer, E. van Nimwegen, C. Wirbelauer, E.J. Oakeley, D. Gaidatzis, V.K. Tiwari, D. Schubeler, DNA-binding factors shape the mouse methylome at distal regulatory regions, Nature 480 (2011) 490–495.
[34] A.E. Peaston, A.V. Evsikov, J.H. Graber, W.N. de Vries, A.E. Holbrook, D. Solter, B.B. Knowles, Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos, Dev. Cell 7 (2004) 597–606.
[35] A. Fort, K. Hashimoto, D. Yamada, M. Salimullah, C.A. Keya, A. Saxena, A. Bonetti, I. Voineagu, N. Bertin, A. Kratz, Y. Noro, C.H. Wong, M. de Hoon, R. Andersson, A. Sandelin, H. Suzuki, C.L. Wei, H. Koseki, F. Consortium, Y. Hasegawa, A.R. Forrest, P. Carninci, Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance, Nat. Genet. 46 (2014) 558–566.
[36] G. Kunarso, N.Y. Chia, J. Jeyakani, C. Hwang, X. Lu, Y.S. Chan, H.H. Ng, G. Bourque, Transposable elements have rewired the core regulatory network of human embryonic stem cells, Nat. Genet. 42 (2010) 631–634.
[37] C.J. Cohen, W.M. Lock, D.L. Mager, Endogenous retroviral LTRs as promoters for human genes: a critical assessment, Gene 448 (2009) 105–114.
[38] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359.
[39] P. Lutsik, L. Feuerbach, J. Arand, T. Lengauer, J. Walter, C. Bock, BiQ Analyzer HT: locus-specific analysis of DNA methylation by high-throughput bisulfite sequencing, Nucleic Acids Res. 39 (2011) W551–W556.
[40] R. Edgar, M. Domrachev, A.E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, Nucleic Acids Res. 30 (2002) 207–210.