

# Chapter 11

## Mammalian Genome Plasticity: Expression Analysis of Transposable Elements

Brian B. Griffiths and Richard G. Hunter

### Abstract

Transposable elements (TEs) are mobile genetic elements, which constitute the single largest fraction of the mammalian genome. Though long assumed to be silent junk or parasitic, recent research has established that most of these elements are transcribed, often in a cell- and tissue-specific fashion and that this expression appears to be regulated in response to environmental influences. Therefore, it seems quite possible that these elements might prove to have a functional role in mammalian physiology and cell biology. Transposons have also been identified as pathogenic factors in both humans and animal models of diseases from cancer to neurodegeneration. These findings have stimulated further interest in transposon biology and created the need for further dissemination of the methods for analysis of TE expression, which is the goal of this chapter.

**Key words** Retrotransposon, RNA-sequencing, PCR, In situ hybridization, Chromatin immunoprecipitation

---

### 1 Introduction

Transposable elements (TEs) are genomic elements capable of moving themselves, transposing, around the genome. They do so in a variety of ways, which help define the various classes of transposable elements functionally and taxonomically. Transposons are divided into two major classes, retrotransposons, which use RNA intermediates produced by reverse transcriptase to copy and paste themselves around the genome, and DNA transposons, which transpose DNA directly without intermediates [1, 2]. Transposons comprise as much as half of most mammalian genomes [3–6], and their biology has been subject to extensive study over the years since their discovery by McClintock more than half a century ago [7–17]. However, due in part to labels like “junk” and “parasite” attached to these elements by eminent biologists such as Crick and Ohno [18, 19], as well as the more obvious functional significance of genes, this research effort has paled in comparison

to that spent on protein-coding genes. Further, it has distracted attention from what the functional role of these elements, in both health and disease, might be.

Starting with Kazazian's discovery of the role of transposons in the genetic pathology of Hemophilia [12], more research has focused on the capacity of transposition to produce pathology in humans and other species. More recently, it has become apparent that transposition can contribute to disease, as well as to genomic diversity in organs like the brain and immune system where it can have positive functional effects [20–22]. Further, transposition contributes substantially to individual genomic diversity and genome evolution, as genomic rearrangements caused by these elements accumulate at levels far higher than those visible by examination of coding sequences alone [23–27]. However, transposition is not the only feature worthy of attention with regard to these elements. Most transposons were long assumed to be transcriptionally silent. However, recent large-scale genomics efforts such as ENCODE and others, as well as next-generation sequencing efforts, have established that many of these elements are actively transcribed in a cell-type specific manner. Moreover, they appear to be involved not only in regulating their own expression, but the expression of nearby genes as well [28–30]. Transposon transcription can be regulated by environmental influences like stress in the mammalian brain [31–34]. Aberrant expression of transposons has been linked to a number of human diseases including neurodegeneration, autoimmune disorders, and cancer [2, 35, 36]. The fact that transposable elements are so actively transcribed, while somatic transposition rates are relatively low, has led to questions about what the role of transposon RNA might be in mammalian cells. To answer these questions, we must be able to analyze their expression, which is the aim of this chapter.

---

## 2 Materials

### 2.1 RNA Extraction

In order to analyze the expression of transposon RNA, it is necessary to extract the RNA from the tissue sample of interest. RNA extraction kits are widely available and should be chosen based on the optimal product for the sample and sample preparation you are using, e.g., for unfixed brain or adipose tissue a lipid extraction kit, such as the Qiagen RNeasy Lipid Tissue Kit (Qiagen) is ideal.

### 2.2 RT-PCR

RT-PCR reagents are widely available from a variety of manufacturers and should be selected based on familiarity and the target sequence to be examined. Generally, if detection of a specific transcript is the priority, then Taqman-based methods are to be preferred, whereas SYBR green and similar intercalating dyes are preferred when higher sensitivity is a priority.

### **2.3 Microarrays and Sequencing**

As these methods generally depend on the availability of a particular piece of capital equipment in a laboratory or core facility, the choice of platforms is usually predetermined. With regard to microarrays the difference between the most common platforms comes down to whether transposon transcripts are included on the arrays, and the extent to which these transposon transcripts overlap with one's research question. With regard to next-generation sequencers, the depth of sequencing possible with mid- to high-end sequencers such as the Illumina HiSeq 2500 or NextSeq 500 is better for the analysis of transposon RNA expression. As it improves the ability of alignment programs to distinguish highly similar and common transposon transcripts, paired-end sequencing, and greater sequence length, are preferable where possible.

---

## **3 Methods**

### **3.1 Choosing an Approach**

Which approach is best for analyzing TE expression depends very much on the research question asked. If the expression of a small number of known TEs is sufficient, then RT-PCR is the most cost effective and accessible approach for most laboratories. However, when the goals are more discovery-oriented or the question pertains to large-scale transcriptional or chromatin regulation of these elements, then next-generation sequencing approaches are more appropriate. Thus, if chromatin-TE interactions are at the root of our research program, we might start with a ChIP-sequencing experiment to identify which TEs are associated with a particular chromatin state or mark. In order to determine if this association influences the transcriptional activity of the elements identified, we would then proceed to use RNA-sequencing to build a global picture of the transcriptional impact of the particular Chromatin-TE interaction we are seeking to examine. Finally we might then use RT-PCR to examine a small subset of these TEs in more detailed mechanistic experiments, or in situ hybridization to assess their anatomical distribution.

### **3.2 Northern Blotting and In Situ Hybridization**

Prior to the development of the polymerase chain reaction (PCR), the northern blot was the dominant method for detecting specific RNAs extracted from cells or tissue. Northern blotting utilizes gel electrophoresis, much like Western blotting, to sort a sample of RNAs by size, these are then transferred to a blot and hybridized with either RNA or DNA antisense probes labeled to aid in visualization. RNA probes, due to their greater length, can increase specificity, while DNA probes are easier to work with. Choice of label is also important. Radioactive labeling with  $P^{32}$  provides maximum sensitivity, while colorimetric methods offer greater ease of use. Though less commonly used at present, due in part for the need for larger amounts of starting material, Northern blots have

the benefit of being able to identify transcript length, which is significant when examining TEs, as they are prone to the production of multiple transcripts, some of which are processed further into small RNAs by the RNAi machinery [37]. Of course the information can also be obtained from RNA-Seq experiments, but for research questions about a particular TE northern blotting remains a cost effective approach.

In situ hybridization (ISH) permits the localization of RNA transcripts to a high degree of cellular and anatomical resolution. In tissues like the brain, where anatomical and cellular specificity are significant factors, ISH may be the only approach available to analyze the spatial expression of TEs, especially given our limited understanding of the extent to which TE transcription results in protein expression. ISH, as its name suggests, involves hybridizing labeled RNA or DNA probes to tissue or cells in situ and visualizing them using either colorimetric agents or radioactive labeling. Radioactive labels are preferred when precise quantitation and sub-cellular localization are needed. They may also be the best means when transcript levels are very low, as signal can be increased with greater exposure time of the sample to either photographic film or emulsion (often weeks in length, though exposures of up to a year have been used in some cases, *see Note 1*).

Probe design for ISH and Northern blotting is similar. For RNA probes, a clone of the transcript of interest is cloned into an expression vector, which is then transcribed in competent cells, purified and labeled. Generally, the full-length probe will be used in Northernblots, while the probe is often fragmented for ISH to increase its ability to infiltrate tissue samples. DNA oligonucleotide probes can be built in a fashion similar to that utilized for PCR primer and probe design, with a target melting temperature ( $T_m$ ) between 60 and 65 °C being best for most protocols. As with all probe and primer design, running the probes and their complements through BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) is useful to reduce off-target hybridizations. Stringency is controlled with high salt and high temperature washes after the hybridization step [38].

### 3.3 RT-PCR

One method to test for the expression of TEs is to use Reverse-Transcription Polymerase Chain Reaction (RT-PCR; [39]). Although solutions exist for multiplexing RT-PCR [40], amplification biases and false-negatives that can occur mean best results are often achieved by looking at one target per well. This makes it time-intensive compared to other methods (though it is much faster than ISH), but is the present standard for showing expression changes. Many of the higher throughput assays are used to select candidates for RT-PCR confirmation, an approach used by Rowe and collaborators in showing that KRAB-associated protein 1 (KAP1) silences endogenous retroviruses in mouse embryonic stem cells [41], which used RNA-Seq and ChIP-Seq methods to find RT-PCR targets.

The first step in RT-PCR is to extract RNA from sample cells. Once the RNA is extracted, it must be reverse-transcribed into DNA. After the DNA copy is made, primers can be designed to target the TE of interest and the reaction can be quantified.

RNA extraction can be easily and reliably done with a variety of filter-tube kits for about \$5 per sample. In short, the process involves collecting a fresh or frozen sample, lysing the cells, filtering out the RNA in an affinity column, and washing all other cell products away. Both the quantity (concentration) and quality of extracted RNA can be measured using a spectrometer such as a NanoDrop. Then the extracted RNA can be frozen at  $-80^{\circ}\text{C}$  and stored until needed.

Reverse Transcription is a relatively straightforward process, with kits available from a variety of manufacturers. It is important to know the concentration of RNA template you are adding to each reaction, and vary the reagents accordingly. We generally prefer to perform the reverse transcription and PCR reactions separately when developing assays for new targets, as it makes the process easier to troubleshoot.

### **3.4 Designing PCR Primers**

Several tools now exist which allow the researcher to easily design and order PCR primers for testing DNA sequences of interest. With the NCBI's Primer-BLAST one can paste in a FASTA sequence or upload a file of multiple sequences [42]. Potential primer sequences are automatically analyzed with BLAST for specificity. Integrated DNA Technologies (IDT) offers a similar interface with their PrimerQuest software, but with the additional benefit of being able to purchase the primers directly from their website [43]. Both websites use the Primer3 software developed by the Whitehead Institute for Biomedical Research.

After entering a desired sequence to be amplified, several parameters can be adjusted according to the specific needs of the project. Primers can be ordered for general PCR, with an attached probe, or for use with intercalating dyes such as SYBR Green.

Because many TEs are small, it is important to keep amplicon length in mind. Designing a primer with padding on the 5' side of the desired TE target can increase the chance that it will bind to the specific target. This increases the specificity for the intended target, because the same core TE sequence may show up in multiple places in the genome, and may be expressed from more than one source, while flanking sequences are likely to be more unique. If overall levels of a particular class of TE are to be quantified, then primers can be designed to the core sequence of that class.

### **3.5 qRT-PCR**

Quantitative Reverse-Transcription PCR (qRT-PCR) uses changes in fluorescence intensity of special dyes in between cycles [44]. Taqman probes are specially designed primers that incorporate a fluorescent tag that becomes active when bound to DNA [45].

Intercalating dyes, such as SYBR Green, are only fluorescent when they are trapped in the grooves of double stranded DNA. For both of these methods, as the number of replicated targets increases, so does the intensity of the fluorescence. Specialized thermocyclers can record the level of intensity for each cycle. Because the replication is ostensibly exponential, a quantification curve can be generated with this data, and the number of cycles a target takes to reach a certain threshold ( $C_T$  score) directly relates to the amount of template that was present at the start. This value can be compared to other targets in the same sample, such as a positive control (usually  $\beta$ -Actin or GAPDH), or to other samples for the same target. While it is often assumed that each cycle of amplification represents a doubling of target sequence, no chemical reaction proceeds with 100 % efficiency, for this reason it is important to determine the efficiency of your thermocycler if exact quantitation is desired, or, if possible, use digital PCR.

### **3.6 Digital PCR**

In samples where the TE expression levels may be too low to be accurately quantified using traditional q-PCR methods, or where very precise quantitation is desired, digital PCR can help elucidate expression differences. The technique has been utilized very effectively to precisely quantitate the exact number of transpositions in the human cortex [46], and has many potential applications where exact assessment of sequence number is important. Digital PCR is so named because it utilizes a system of ones and zeros. A small, dilute aliquot of sample DNA is put into multiple wells per plate. Instead of using fluorescence intensity and exponential amplification to quantify total template present, as in RT-PCR, the well is counted as either containing the sample target (a one), or not (a zero). Because of the small amounts of DNA added to each well, not all wells will contain the target, and quantity can be deduced through fitting the results to a Poisson distribution [47].

### **3.7 Microarrays**

Microarrays are useful for looking at a large number of targets in a single assay. Most arrays utilize a hybridization probe: a single-stranded oligonucleotide that is anchored to a filter and denatured. The oligonucleotide is antisense to the sequence of interest. Standard microarrays can analyze the expression of tens of thousands of genes and expressed sequences. Some whole-genome microarrays include TEs, though most concentrate on annotated genes. Standard arrays have been used to discover changes in TE expression in mammalian brain [33, 34]. With off-the-shelf arrays, this method may be better suited for examining the expression of TEs that are situated in the introns and exons of genes, rather than TEs in the intergenic regions; however custom built arrays can be utilized to target TEs specifically. A TE-specific array recently been created by the Boeke group [48], and could be of great utility if it becomes commonly available.

Other microarrays look at a single target, but use binding frequency data for quantification purposes, similar to digital PCR. This method was used to determine that mice have approximately 3000 active L1 elements [49].

### 3.8 Next-Generation Sequencing

#### 3.8.1 RNA-Seq

Perhaps the best method for determining the expression of TEs is to sequence all expressed RNAs with RNA-Seq [50], align them to the source genome, and count the number of occurrences for all reads. The method is generally superior to microarray for a number of reasons [51]. As with PCR, the results of this method can be difficult to interpret with regard to small repeats. As many transposon sequences are found in multiple places in the genome and transposon transcripts can be variable in length, unique alignments can be difficult, thus the stringency with which multiple alignments are excluded from analysis may need to be relaxed depending on the research question. As with all next-generation sequencing methods, determination of ideal sequencing depth, either through preliminary experiments or by applying standard calculations and protocols, is important if a truly representative sampling is to be achieved [52]. Choice of RNA extraction method is also very important with regard to TEs. Standard RNA-Seq protocols typically utilize a poly-A selection to increase the representation of mRNA in the sample. Many TE-derived transcripts have poly-A tails, and so will be retained in these samples, however many transcripts are not polyadenylated and will be lost if this approach is used. For this reason either total RNA, or total RNA prepared so as to remove ribosomal RNA, with products such as RiboMinus™ (Life Technologies), will provide a more representative sample

RNA is extracted from a sample and fragmented. Fragmentation for next-generation sequencing is an important step, as differently sized fragments will lead to biases in sequencing. Several solutions exist for precise fragmentation, such as precision sonicators or enzymatic hydrolysis. After fragmentation, the RNA is reverse-transcribed into cDNA and purified for sequencing. After sequencing reads are aligned to a reference genome or transcriptome. With regard to TEs, *de novo* alignment or alignment to Repeatmasker may be needed in order to ensure that poorly annotated TE sequences are captured by the analysis [53]. Peak calling software is then utilized to determine the number and location of significant peaks of a particular quality (*see Note 2*). After alignment it is also possible to analyze transcript data for different transcript length and splicing utilizing programs such as TopHat [54].

#### 3.8.2 ChIP-Seq

Chromatin Immunoprecipitation (ChIP) is a method for examining DNA-Protein interactions [55]. As such, ChIP-Seq is useful to identify TEs which might be interacting with a particular element of chromatin or the transcriptional machinery in order to narrow down the number of TEs one might wish to subject to expression analysis.

Similar methods, such as methylated DNA Immunoprecipitation (MeDIP-Seq) can also be used depending on the nature of the research. The first step in ChIP is to cross-link the DNA to the bound proteins. The DNA is then precisely fragmented, usually into segments about 500 bases long. Metallic beads that contain an antibody with affinity for the protein of interest bind the protein while the rest of the DNA and cell contents are washed away. The DNA and protein are then unlinked with heat and proteinase K treatment, leaving behind only DNA that was bound by the protein *in vivo*. This DNA can then be analyzed via next-generation sequencing, or RT-PCR if a specific target is being examined. When developing ChIP assays, it is important to have positive and negative PCR controls to confirm antibody specificity (*see Note 3*).

This is a valuable tool for examining the transcription factors or cofactors that may bind to TEs and affect their expression, as was done by Lynch and collaborators who assessed the large array of factors bind to the mammalian-specific TE MER20 during pregnancy [56]. As TEs are often found in heterochromatin, ChIP-Seq can be used to identify TEs which are associated with particular marks, some of which may be involved in regulating their expression, as the histone H3 lysine 9 trimethylation mark appears to be [31, 41].

The power of ChIP can be further increased by sequencing the DNA library after unlinking. This creates a read of the DNA sequence associated with the protein, which can then be analyzed for changes in binding. In one study, researchers looked at the binding of RNA polymerase II to TE sites to estimate expression levels [57], and ChIP-seq will likely be used with more frequency as the cost of DNA sequencing continues to decline.

Model-based Analysis of ChIP-Seq (MACS) is a popular tool for peak calling [58]. It uses a global and local average to find instances where more protein was bound to a specific region of DNA than is explainable by random chance. This tool is available as a command line tool for Linux systems, or with a graphical user interface through the Galaxy project [59–61].

### **3.9 Single-Cell Sequencing**

Because transposition can lead to a unique genotype for individual neurons, looking at population expression of TEs may not be sufficient for all hypotheses. Fluorescence Activated Cell Sorting (FACS) is a tool for selecting a common population of cells [62]. In investigation of the central nervous system, this is often used to sort glial cells from neurons in order to detect small expression changes that may be washed out by a heterogeneous population. In single-neuron sequencing done by Evrony et al. [46], it was used to select for using the neuron-specific antibody NeuN. The individual genome of 300 neurons was amplified and then individually sequenced in order to inspect the rate of L1 transposition.



While the rate of L1 transposition for individual neurons was determined to be very low (0.6–0.04 %), it was deemed sensitive enough to detect transposition in a single cell [46]. This method may be used in the future to explore other facets of TE expression in individual cells. Alternatively, laser capture microdissection (LCM) can be utilized to dissect single cells from tissue sections mounted on specialized microscope slides. For expression analysis, this approach may be superior to FACS as it does not require the lengthy dissociation and sorting steps, which will alter transcription globally. With LCM, expression can be fixed when the tissue is harvested either by fixation or freezing.

### **3.10 Identifying TE in Sequencing Data**

Because TEs number in the thousands, it is not practical to search sequencing data for possible TEs manually. RepeatMasker is software designed to mask out TEs for classic genetic studies, where TEs are seen as problematic [63]. After converting the sequencing file to FASTA format, the sequences can be run through RepeatMasker, which will then find all the TEs and generate a report, summarizing by TE class and family. Another file will list every individual TE found. This file can then be cross-referenced to sequencing data to use the statistics metrics provided by those analyses. TE transcripts can then be filtered by *P*-value, false discovery ratio and fold change in expression to identify targets for further analysis.

---

## **4 Conclusions**

Though transposable elements were discovered decades ago, they remain a significant and exciting frontier for a number of biological disciplines. We have attempted to provide an overview of some of the methods available for their analysis so that they might become a more accessible area of research. Nonetheless, given rapid advances, particularly in bioinformatics and next-generation sequencing technology, it is likely they will be quickly improved upon or even largely superseded, as has been the case with Northern blotting. Another area where change is likely is in the analysis of TE protein expression. To date only a few such proteins are known to be expressed in mammalian tissues, and some have proven difficult to detect even when circumstantial evidence for their presence is high, such as the LINE1 ORF2 protein [64]. Development of new antibodies or application of advanced mass spectrometry techniques are likely to be of assistance in analyzing the expression of TE-derived peptides in the immediately foreseeable future. Readers are encouraged to consult the recent literature in their area for refinements and additions to the framework presented here.

---

## 5 Notes

1. Though many molecular techniques are available in kit form, ISH often requires optimization when being established for the first time. For this reason a positive control is highly desirable. This can take the form of a section from a tissue known to be high in the target transcript, or a homogenate of the target tissue doped with synthetic target sequence. For radioactive ISH, standard micro scales should be used to calibrate radioactive signal to optical density on film. Controls for nonspecific binding are also important, typically a 100-fold excess of unlabeled probe added to the hybridization step, which will block hybridization of labeled probe and permit identification of background and nonspecific binding.
2. RNA-Seq is a very powerful technique, but this power and the volume of data it can produce contribute to a substantial risk of erroneous results. For this reason, genes whose expression changes are known should be used as controls within the data, and any significant change in expression should be confirmed with another method, such as RT-PCR in a separate confirmation experiment.
3. ChIP depends critically on three factors: proper fixation, sonication, and antibody specificity [55]. Fixation depends on the affinity of the protein of interest for DNA; histones can often be ChIPed without fixation (native ChIP) as their affinity for DNA is quite high, while many transcription factors or elements of the transcriptional machinery will require longer fixation times or more powerful fixation agents than the 1 % formaldehyde often used for histone marks. Sonication should reliably produce a smear of DNA fragments concentrated around 300–500 bp. Benchtop horn or bath sonicators are not reliable enough for this purpose, so more specialized sonicators, such as those produced by Diagenode or Covaris, are required. ChIP should never be performed with an antibody that has not been confirmed to work for ChIP. Many manufacturers do so, but it is worth confirming that the fraction of DNA IPed with a particular antibody is specific by using both positive and negative controls and comparing these for enrichment over an un-IPed input sample of DNA.

## References

1. Jurka J et al (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110(1-4):462–467
2. Levin HL, Moran JV (2011) Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* 12(9):615–627
3. Lander ES et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921
4. Waterston RH et al (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915):520–562

5. Gibbs RA et al (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428(6982):493–521
6. Nellaker C et al (2012) The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol* 13(6):R45
7. McClintock B (1951) Chromosome organization and genic expression. *Cold Spring Harb Symp Quant Biol* 16:13–47
8. Britten RJ, Kohne DE (1968) Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* 161(3841):529–540
9. Grimaldi G, Singer MF (1982) A monkey Alu sequence is flanked by 13-base pair direct repeats by an interrupted alpha-satellite DNA sequence. *Proc Natl Acad Sci U S A* 79(5):1497–1500
10. Boeke JD et al (1985) Ty elements transpose through an RNA intermediate. *Cell* 40(3):491–500
11. Daniels GR, Deininger PL (1985) Repeat sequence families derived from mammalian tRNA genes. *Nature* 317(6040):819–822
12. Kazazian HH Jr et al (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332(6160):164–166
13. Xiong YE, Eickbush TH (1988) Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. *Cell* 55(2):235–246
14. Britten RJ, Stout DB, Davidson EH (1989) The current source of human Alu retrotransposons is a conserved gene shared with Old World monkey. *Proc Natl Acad Sci U S A* 86(10):3718–3722
15. Dombroski BA et al (1991) Isolation of an active human transposable element. *Science* 254(5039):1805–1808
16. Batzer MA et al (1996) Genetic variation of recent Alu insertions in human populations. *J Mol Evol* 42(1):22–29
17. Moran JV, DeBerardinis RJ, Kazazian HH Jr (1999) Exon shuffling by L1 retrotransposition. *Science* 283(5407):1530–1534
18. Orgel LE, Crick FH (1980) Selfish DNA: the ultimate parasite. *Nature* 284(5757):604–607
19. Ohno S (1972) So much “junk” DNA in our genome. *Brookhaven Symp Biol* 23:366–370
20. Muotri AR et al (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435(7044):903–910
21. Singer T et al (2010) LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends Neurosci* 33(8):345–354
22. Kapitonov VV, Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol* 3(6):e181
23. Heard E et al (2010) Ten years of genetics and genomics: what have we achieved and where are we heading? *Nat Rev Genet* 11(10):723–733
24. Iskow RC et al (2010) Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141(7):1253–1261
25. Ewing AD, Kazazian HH Jr (2010) High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res* 20(9):1262–1270
26. Beck CR et al (2010) LINE-1 retrotransposition activity in human genomes. *Cell* 141(7):1159–1170
27. Huang CR et al (2010) Mobile interspersed repeats are major structural variants in the human genome. *Cell* 141(7):1171–1182
28. Faulkner GJ et al (2009) The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* 41(5):563–571
29. Djebali S et al (2012) Landscape of transcription in human cells. *Nature* 489(7414):101–108
30. Thurman RE et al (2012) The accessible chromatin landscape of the human genome. *Nature* 489(7414):75–82
31. Hunter RG et al (2012) Acute stress and hippocampal histone H3 lysine 9 trimethylation, a retrotransposon silencing response. *Proc Natl Acad Sci U S A* 109(43):17657–17662
32. Hunter RG, McEwen BS, Pfaff DW (2013) Environmental stress and transposon transcription in the mammalian brain. *Mob Genet Elements* 3(2):e24555
33. Ponomarev I et al (2010) Amygdala transcriptome and cellular mechanisms underlying stress-enhanced fear learning in a rat model of posttraumatic stress disorder. *Neuropsychopharmacology* 35(6):1402–1411
34. Ponomarev I et al (2012) Gene coexpression networks in human brain identify epigenetic modifications in alcohol dependence. *J Neurosci* 32(5):1884–1897
35. Reilly MT et al (2013) The role of transposable elements in health and diseases of the central nervous system. *J Neurosci* 33(45):17577–17586
36. Stetson DB (2012) Endogenous retroelements and autoimmune disease. *Curr Opin Immunol* 24(6):692–697

37. Dumesic PA, Madhani HD (2014) Recognizing the enemy within: licensing RNA-guided genome defense. *Trends Biochem Sci* 39(1): 25–34
38. Dagerlind A et al (1992) Sensitive mRNA detection using unfixed tissue: combined radioactive and non-radioactive in situ hybridization histochemistry. *Histochemistry* 98(1): 39–49
39. Bustin SA (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 25(2):169–193
40. Henegariu O et al (1997) Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques* 23(3):504–511
41. Rowe HM et al (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* 463(7278):237–240
42. Ye J et al (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform* 13(1):134
43. IDT. *PrimerQuest® program*. 2012. Accessed on 1 July, 2014, from: <http://www.idtdna.com/Scitools>
44. Heid CA et al (1996) Real time quantitative PCR. *Genome Res* 6(10):986–994
45. Medhurst AD et al (2000) The use of TaqMan RT-PCR assays for semiquantitative analysis of gene expression in CNS tissues and disease models. *J Neurosci Methods* 98(1):9–20
46. Evrony GD et al (2012) Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* 151(3):483–496
47. Pohl G, Shih Ie M (2004) Principle and applications of digital PCR. *Expert Rev Mol Diagn* 4(1):41–47
48. Gnanakkan VP et al (2013) TE-array – a high throughput tool to study transposon transcription. *BMC Genomics* 14:869
49. DeBerardinis RJ et al (1998) Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat Genet* 20(3): 288–290
50. Mortazavi A et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621–628
51. Marioni JC et al (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18(9): 1509–1517
52. Li H et al (2008) Determination of tag density required for digital transcriptome analysis: application to an androgen-sensitive prostate cancer model. *Proc Natl Acad Sci U S A* 105(51):20179–20184
53. Chaparro C, Sabot F (2012) Methods and software in NGS for TE analysis. In: Bigot Y (ed) *Mobile genetic elements*. Springer, New York, NY
54. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111
55. Milne TA, Zhao K, Hess JL (2009) Chromatin immunoprecipitation (ChIP) for analysis of histone modifications and chromatin-associated proteins. *Methods Mol Biol* 538: 409–423
56. Lynch VJ et al (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* 43(11):1154–1159
57. Sienski G, Donertas D, Brennecke J (2012) Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151(5): 964–980
58. Zhang Y et al (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137
59. Blankenberg D et al. (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol*. Chapter 19: Unit 19 10 1–21
60. Giardine B et al (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* 15(10):1451–1455
61. Goecks J, Nekrutenko A, Taylor J (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11(8):R86
62. Bonner WA et al (1972) Fluorescence activated cell sorting. *Rev Sci Instrum* 43(3):404–409
63. Smit A, Hubley R, Green P. RepeatMasker Open-3.0. 1996–2010. Accessed on 1 July, 2014, from: <http://www.repeatmasker.org>
64. Dai L et al (2014) Expression and detection of LINE-1 ORF-encoded proteins. *Mob Genet Elements* 4:e29319