The battle inside our genome

Controlling transposable elements and the evolution of human gene regulatory networks

Haring, N.L.

Publication date
2020

Document Version
Other version

License
Other

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Discussion

Consequences of the Battle Within: Transposable Elements versus KRAB Zinc Finger Genes
Discussion
Consequences of the Battle Within: Transposable Elements versus KRAB Zinc Finger Genes

Introduction
Since the seminal work of Barbara McClintock in the ‘40s, an increasing amount of studies focused on understanding the role of TEs in the evolution of gene regulation. The last decade genome-wide profiling of the epigenetic landscape at TEs in combination with the characterization of TE-derived binding sites for many TFs in different species has shown that they provided species-specific cis-regulatory elements on a large scale (Wang et al. 2007; Kunarso et al. 2010; Xie et al. 2010; Lynch et al. 2011; Chuong et al. 2013; Sundaram et al. 2014; Notwell et al. 2015; Trizzino et al. 2017; Ito et al. 2017; Trizzino et al. 2018; Pontis et al. 2019). This was further substantiated by the functional assessment of individual TE insertions and their target genes (Chuong et al. 2016; Sundaram et al. 2017; Pontis et al. 2019).

The work presented in this thesis is dedicated to the identification of TE families co-opted as cis-regulatory elements in human neurons and to enhance our understanding of the impact of the battle between two KZNF genes and SVA elements on gene expression. Here we discuss the main findings described in the chapters and put forward hypotheses and future perspectives.

An evolutionary feedback loop to control TE activity
Today it is widely accepted that TEs are frequently embedded in gene regulatory pathways, instead of being continuously and robustly repressed. However, little is known about how their cis-regulatory activity is controlled. The KZNF gene family is specialized in repressing TEs through induction of local heterochromatin. The identification of the binding sites of over two hundred KZNF proteins has revealed the TE families they target (Jacobs et al. 2014; Najafabadi et al. 2015; Schmitges et al. 2016; Imbeault et al. 2017). However, how KZNF genes are regulated and how their expression relates to TE activity remained open questions. To address this we profiled the expression dynamics of primate specific KZNF genes and their target TEs in hESCs and cortical and mDA organoids in Chapter 1. We showed that both KZNF genes and TEs show a cell type specific expression pattern. It should be noted that we cannot rule out that some of these differences are the result of differences in library preparation methods. RNA seq data from hESCs and cortical organoids were derived from a ribosomal RNA depleted library, while the library from mDA organoids was poly-A selected. The influence of library preparation methods on relative transcript abundances may be limited however, because a recent study showed in a parallel comparison that the differences in transcript abundance were minimal (Kumar et al. 2017). This suggests that it is safe to assume that transcript abundances of KZNF genes and most TEs can be compared.
Our analyses revealed that groups of KZNF genes and TEs show a highly dynamic and cell type specific expression pattern, but KZNF genes frequently follow the same expression dynamics as their target TE. This is further supported by the fact that KZNF genes targeting the same TE family are co-expressed. Perhaps one would expect an inverse relationship between KZNF gene and their target TE, as higher levels of KZNF protein would lead to increased TE repression and thus reduced transcriptional activity. However, it has been reported before that KZNF gene expression is higher in cell types that also show higher TE activity (Imbeault et al. 2017). It may take time to effectively and collectively silence a TE subfamily explaining why KZNF expression is high at the same time the TE family is transcriptionally active.

Alternatively, there may be other (transcription) factors establishing TE-driven transcription, which compete with KZNF proteins. Recently, it was proposed that TE-derived enhancers regulate the expression of the KZNF genes they are targeted by, ultimately leading to their own silencing (Pontis et al. 2019). This suggests the existence of an intriguing autoregulatory feedback loop between TEs and their partner KZNF gene(s). It should be noted that the analysis by Pontis et al (2019) connected the collective activity of many young TE families to the upregulation of many young KZNF genes, but that evidence of specific TE families regulating their cognate KZNF protein was thin. Notably, our results presented in Chapter 3 reveal that selective epigenetic activation of SVA elements induces the expression of a wide variety of KZNF genes in their proximity. This indicates that KZNF genes can show collective upregulation upon the activation of specific TE insertions that reside nearby. We do not expect, however, that KZNF genes are regulated this way under normal circumstances, because if this was the case KZNF gene clusters would show collective expression patterns in tissues based on their genomic location. Instead, the opposite was observed and KZNF genes with a similar expression pattern generally reside at different chromosomal regions (Huntley et al. 2006). Thus, it seems unlikely that active TE-derived enhancers in the neighborhood are the primary regulators of KZNF gene expression. Furthermore, the highly cell type specific expression of both KZNF genes and TEs suggests that the regulation is tightly controlled. Pontis et al (2019) showed with profiling of 3D DNA interactions (through Chromatin Interaction Analysis by paired-End Tag Sequencing, ChIA-PET seq) that H3K27ac positive TEs connect to the promoters of KZNF genes and that silencing of these elements with CRISPRi led to the downregulation of genes in the close vicinity. Re-analyzing these data would help elucidate if TEs specifically connect to their cognate KZNF and regulate their
expression levels. Whether similar 3D interactions also occur in somatic tissues remains to be explored. As different KZNF genes and TEs become active in brain tissue, exploring the 3D chromatin architecture at KZNF gene promoters in human neurons would elucidate what sequences are important for KZNF gene regulation and determine if TE-mediated KZNF gene regulation is a general mechanism through which KZNF genes are regulated.

On top of that, the extensive amount of TF ChIP data publicly available could provide insights into which factors function as (master) regulator of KZNF genes. Finally, it is possible that the KZNF gene family has an internal regulatory machinery consisting of KZNF genes regulating each other and themselves. ZNF143, for example, has been shown to regulate other KZNF genes in various cancer cell lines (Ngondo-Mbongo et al. 2013). Re-analysis of the available KZNF ChIP data to determine if and what KZNF proteins bind to the promoters of other KZNF genes could shed more light on this phenomenon.

**TE-derived enhancers in human neurons**

Some TE families have had a significant impact on gene regulatory networks essential for maintaining pluripotency during early development and orchestrating an innate immune response (Kunarso et al. 2010; Xie et al. 2010; Lynch et al. 2011; Chuong et al. 2016; Pontis et al. 2019). However, to what extent TE families may have been co-opted for gene regulatory purposes in other (adult) tissues remains largely unexplored. As the human brain is the most evolutionarily divergent organ, we hypothesized that TEs may have been a driving force in the innovation of gene regulatory programs that led to the evolution of novel structures and functions unique to the human brain. A few studies showed examples of co-option of cis-regulatory elements in the brain (Notwell et al. 2015; Pontis et al. 2019). However, a comprehensive characterization of TE-derived enhancers in the human brain is lacking. In Chapter 1 we performed genome-wide epigenetic and transcriptional analyses of hESCs and hESC-derived in vitro differentiated cortical and mDA organoids. First we showed that both neuronal tissues showed H3K27ac and expression profiles highly distinctive from their parental hESC line. The fact that cortical and mDA organoids share the enrichment of H3K27ac on a high number of TE families suggests that specific TE families are involved in the regulation of pan-neuronal gene expression. Indeed, we observe the presence of general neuronal TF motifs in some TEs that are H3K27ac enriched in both neuronal tissues. In addition, a similar amount of TE families show a neuronal subtype specific H3K27ac signature. We identified several primate/ape specific TE families that
showed H3K27ac levels above threshold on 12-35 insertions in cortical and/or mDA organoids. LTR5B and MER61F showed a mDA specific H3K27ac profile, while a group of LTR12E elements were increasingly adorned with H3K27ac in cortical organoids. The fact that the vast majority of these elements contained TF motifs of pan-neuronal and neuronal subtype specific TFs shows that they have the potential to function as cis-regulatory units in this cellular context. Together these findings lay ground to further explore their role as enhancers in neuronal gene regulation.

In vitro differentiated mini brains as a tool to study the role of TEs in the human brain

Several lines of evidence, including our findings in Chapter 1, suggest a role for TEs in neuronal development and neurodegenerative disease. However, mechanistic experiments are necessary to prove they are causal to human (pathological) gene (dys)regulation. For a long time animal models, mainly mouse, have been the gold standard for neuroscience research and provided groundbreaking insights into the functioning of the central nervous system. However, many features, including the TE load in our genome, are unique to the human brain and can therefore not be studied in the mouse. Brain organoids have revolutionized the field and they are believed to provide answers to long standing questions regarding human brain (dys)functioning and evolution. Brain organoids outperform every other in vitro neuronal model due to their 3D structure, various regional identities, cellular diversity, advanced cytoarchitecture, and ability to form functional circuits (Reviewed in: Di Lullo and Kriegstein 2017; Andrews and Nowakowski 2019; Benito-kwiecinski and Lancaster 2019). Comparative analyses revealed that brain organoids resemble mid-gestational human fetal brains on the transcriptional and epigenetic level (Camp et al. 2015; Pasca et al. 2015; Amiri et al. 2018). Therefore, they have proven most valuable for studies with a neurodevelopmental focus. Brain organoids generated from patient-derived induced pluripotent stem cells (iPSCs) or genetically engineered PSCs have provided insights into disease pathology of several neurodevelopmental diseases (Lancaster et al. 2013; Mariani et al. 2015; Bershteyn et al. 2017; Birey et al. 2017; Amiri et al. 2018; Blair et al. 2018). Therefore, brain organoids are highly suitable to experimentally assess the impact of specific TE families or individual TE insertions on gene expression networks orchestrating neuronal development.

To model late-onset neurodegenerative diseases, with age as the most important risk factor, the immature identity of brain organoids is a major
obstacle. The fact that the conversion of somatic cells into iPSCs erases age-related phenotypes such as global loss of epigenetic repression (Miller et al. 2013), mitochondrial dysfunction (Suhr et al. 2010) and telomere shortening (Marion et al. 2009) further hampers the use of brain organoids to study neurodegenerative diseases. In order to assess the role of derepressed TEs in pathology underlying neurodegeneration it is particularly important that age-related hallmarks such as epigenetic drift are maintained. Miller et al (2013) showed that overexpression of progerin, a protein that causes accelerated aging in progeria patients, in iPSCs derived from Parkinson’s disease patients not only reestablished age-related markers, but also showed disease specific hallmarks when differentiated into mDA neurons. Similarly, when fibroblasts are directly differentiated into neurons, without an iPSC intermediate step, age-dependent transcriptomic signatures and nucleocytoplasmic compartmentalization were maintained (Mertens et al. 2015). Although these are promising strategies to study age-related neurodegenerative diseases, it remains to be established if they can be used in conjunction with the differentiation of 3D organoid structures to create a robust model encompassing all disease phenotypes.

However, organoids do provide an excellent platform to experimentally assess brain evolution by combining cross-species comparisons and (epi)genetic editing. Recently, a comprehensive single cell analysis of cerebral organoids from humans, chimpanzees and macaques revealed remarkable differences in neuronal development (Kanton et al. 2019). Not only did they show that human-specific chromatin accessibility correlated with human-specific gene expression changes, but also that the pace of neuronal development of the human cerebral organoids was lower compared to the other primates. On top of that, a similar comparison with bulk RNA seq data showed the expression of nearly 400 human-specific long non-coding RNAs (lncRNAs) (Field et al. 2019). Ectopic activation of these lncRNAs with CRISPRa in HEK293 cells showed major consequences for gene expression, suggesting they may have a gene regulatory role in the brain. These studies lay ground to further explore the role of TEs in the evolutionary innovation of gene regulatory networks. Species-specific neuronal gene expression could be linked to cis-regulatory properties of TEs. As the field of organoid differentiation is still young, it remains largely unclear to what extent brain organoids truly recapitulate the (developing) human brain. Despite rapid improvements of protocols, there are still many challenges and limitations that need to be overcome in order to make them robust models of neurodevelopment and disease. However, the high pace at which differentiation protocols and (epi)genetic tools evolve in combination with the widespread
efforts that are put in detailed analyses promise that in the near future brain organoids will reach their full potential as human brain model. In parallel this opens up avenues to experimentally assess the role of TEs in the human brain.

**ZNF91 protects against the hidden gene regulatory potential of SVA elements**

Although the binding sites of many KZNF proteins are known, it remains to be established if they are required and essential for the repression of their target TE. The fact that many TE families are targeted by multiple KZNF proteins may illustrate the strategy our genome applies to establish their silencing. We set out to study the control of SVA elements, as they are an evolutionary young TE family that is still active in our genome and thus likely contributes to the innovation of gene regulatory networks. **Chapter 4** was aimed at identifying which KZNF proteins are most likely involved in SVA repression. We showed that ZNF91 and ZNF611 both bind to the majority of SVA elements, including elements from all subtypes. Notably, ZNF91 binding co-localizes with KAP1 at the border of the Alu-VNTR domains, whereas ZNF611 was restricted to the VNTR domain that shows lower KAP1 signal. It was previously established that ZNF91 evolved the ability to repress SVA elements in response to the invasion of SVA elements in the ancestor of human, chimp and gorilla (Jacobs et al. 2014). In **Chapter 4** we studied the evolutionary history of ZNF611 and revealed that, although to a lesser extent, this KZNF protein also showed great-ape specific structural changes that most likely enabled it to recognize SVA elements.

As ZNF91 and ZNF611 both strongly bind SVA elements, we aimed at understanding their role in the repression of SVA elements. We reasoned that ZNF91 and ZNF611 both may be essential for SVA repression. In **Chapter 3** we showed that genetic ablation of ZNF91 expression in hESCs leads to epigenetic and transcriptional activation of a fraction of SVA elements, whereas in **Chapter 4** transcriptional profiling of ZNF611 ko hESCs revealed that SVA element do not become transcriptionally active. Deletion of both ZNF91 and ZNF611 displayed no additive effect on SVA transcription compared to the ZNF91 single ko (**Chapter 4**). These results show that ZNF91 may be exclusively important for controlling SVA elements in hESCs, but the role of ZNF611 remains elusive. In **Chapter 3** we showed that when ZNF91 is deleted, approximately one third of SVA elements, mainly evolutionary young subgroups, gain epigenetic marks associated with active transcription and enhancers. Probably older subtypes are less likely activated due to accumulation of mutations of domains essential for activation and/or they are still repressed by KZNF proteins that selectively bind older subtypes (e.g. ZNF202 and ZNF257 bind SVA-A and ZNF141 binds
Consequences of the Battle Within: Transposable Elements versus KRAB Zinc Finger Genes

SVA-B+C). Furthermore, removal of repression revived the gene regulatory potential of SVA elements. We identified several mechanisms through which they control gene expression in close vicinity (Fig 1). Firstly, the upregulation of genes close to epigenetically activated SVA elements suggest that they function as cis-regulatory elements (Fig 1A). Several activated SVA elements that were located very close to a TSS extended the H3K27ac/H3K4me3 signal at the promoter and boosted gene expression (Fig 1B). However, further experiments are necessary to proof causality between SVA activation and nearby gene expression changes. A recent study used PCR and Capture-based chromosome conformation capture (3C) approaches to reveal that MER41 elements interact with adjacent gene promoters, most likely facilitated by STAT TFs (Raviram et al. 2018). This approach could be useful to compare the connectivity of active and repressed SVA elements with gene promoters in wt and ZNF91 ko hESCs. Our current data revealed several differentially expressed genes located close to an activated SVA, which could be used as bait. Noteworthy is that we cannot exclude the possibility that some of the effects on gene expression were caused by spreading of chromatin marks into regulatory regions of neighboring genes, as was shown by Rebollo et al. (2011) (Fig 1C). Even though this is likely to be

![Figure 1. Mechanisms of SVA-mediated gene regulation.](image)

A+B) SVA elements positive for H3K4me3 and H3K27ac function as cis-regulatory elements for genes located within kb distance (A, enhancers) or just down/upstream of the SVA (B, promoter). C) Passive gene regulation through changes in the local chromatin; SVA repression has a mild repressive effect on transcription in the area. D) Epigenetically active SVA elements can function as alternative TSS, producing chimeric SVA-Gene transcripts.
restricted to TEs close to a TSS, our current investigations were not able to assess this.

It needs to be mentioned that overall gene expression changes were relatively mild. This may be caused by the low availability of TFs necessary to unleash the gene regulatory potential of SVA elements fully. A recent report demonstrated that primed hESCs require KLF4 overexpression to induce the cis-regulatory potential of SVA elements (Pontis et al. 2019).

Finally, in the absence of repression, some SVA elements function as alternative TSS and generate chimeric transcripts with the downstream located gene (Fig 1D). Of specific interest is the highly expressed fusion transcript composed of an SVA-F spliced into HORMAD1 as alternative exon 1. In healthy individuals HORMAD1 is selectively expressed in the testes. However, aberrant HORMAD1 expression has been observed in various types of cancer (Shahzad et al. 2013; Watkins et al. 2015; Nichols et al. 2018). As young TEs, including SVA elements, show loss of epigenetic silencing and transcriptional activity in cancer cells (Szpakowski et al. 2009; Barchitta et al. 2014; Attig et al. 2019) it is possible that aberrant HORMAD1 expression may be the consequence of SVA derepression. More analyses using RNA seq and histone ChIP data of these cancer cells could point out if SVA elements are indeed derepressed and drivers of chimeric transcripts such as SVA-HORMAD1 driven fusion transcripts, similar to what is observed for LTR elements (Yu et al. 2005; Macfarlan et al. 2012; Beyer et al. 2016; Krönung et al. 2016; Sokol et al. 2016; Brocks et al. 2017). If these transcripts can be linked to the origin of the disease, TEs offer an interesting therapeutic target. Taken together, we have shown that in hESCs ZNF91 plays an essential role in protecting against the disruptive gene regulatory potential of SVA elements.

Reestestablishment of SVA silencing in ZNF91 knockout cortical organoids
To study the hidden gene regulatory potential of SVA elements in the human brain we differentiated ZNF91 ko hESCs into cortical organoids and isolated RNA at day 14 (D14) and D42 (Fig 2A). Cortical organoids of ZNF91 wt and ko lines were very similar in general appearance, size and numbers (Fig 2B). Surprisingly, RNA sequencing revealed that SVA repression was almost fully restored again, as only very few elements were transcribed in ZNF91 ko cortical organoids (Fig 2C+D). On top of that, the number of differentially expressed genes was negligible (data not shown), with the exception of the SVA-HORMAD1 chimeric transcript that persisted throughout cortical diffe-
Figure 2. SVA elements repressed again in ZNF91 knockout cortical organoids.
A) Experimental design of cortical organoid differentiation from ZNF91 ko hESCs. B) Bright field images showing cortical organoids at D14 and D42. D) Principal component analysis of RNA seq data including only SVA-derived transcripts. E) Heatmap showing hierarchical clustering of SVA elements expressed in ZNF91 ko hESCs (z-score of normalized counts).
rentiation. Together this suggest that the repressive effect of ZNF91 and/or the transcriptional potential of SVA elements is highly cell type specific. This is in agreement with other studies that show limited derepression of ERVs upon KAP1 deletion in differentiated cells, opposed to ESCs (Rowe et al. 2010, 2013; Bojkowska et al. 2012; Tie et al. 2018). This supports a model in which more permanent epigenetic signatures, such as DNA methylation, secure TE repression in differentiated cells (Arand et al. 2012). Indeed, Tie et al (2018) suggested that SVA elements are less easily derepressed in somatic cells due to high levels of DNA methylation. Another plausible reason why SVA transcripts are not detected in ZNF91 ko cortical organoids is the expression of another KZNF gene that can repress SVA activity and is not expressed in hESCs. Noteworthy is that it was shown recently that a proportion of SVA elements gain H3K27ac marks in the fetal brain and iPSC derived human neurons (Pontis et al. 2019). Possibly the increase in H3K27ac and H3K4me3 on SVA elements as found in ZNF91 ko hESCs is maintained during cortical organoid development, but the histone ChIP analysis needed to establish this was not yet done at the time of this writing. To conclude, in cortical neurons ZNF91 is not essential for SVA repression or the (transcriptional) consequences of SVA derepression are negligible in the human neocortex.

Keep your enemies close: Unleashed TEs activate an immune response against themselves

It has become increasingly clear that the KZNF gene cluster evolved in parallel and most probably in response to TE invasions (Thomas and Schneider 2011; Jacobs et al. 2014; Imbeault et al. 2017). However, how TE invasions are detected and what kind of short and long term responses are present in the host to stop them remains largely unclear. Our data sheds light on the potential mechanism through which aberrant TE activity can induce a defensive response by the host to counteract TE outbursts (Fig 3). Together with data from other studies we formulate a hypothetical model that may explain the genome defense against a TE outburst.

First of all it is important to know that chromosomal regions with a high KZNF gene density in both human and mouse are also filled with TE insertions (Wang et al. 2005; Kauzlaric et al. 2017). This may indicate that the DNA and chromatin around KZNF gene clusters are more permissive for TE integrations. A burst of novel TE insertions in KZNF dense genomic areas may provide several ways to repress the escaped TE. Some of the possible mechanisms are outlined below:
Collective upregulation of the KZNF gene family as acute response against a TE invasion

The collective upregulation of KZNF genes in response to widespread TE activation may be an acute defense response to limit a burst of novel insertions (Fig 3A). Possibly, elevated levels of KZNF protein allows for more off-target binding and potentially silencing of the novel TE. The more KZNF genes are upregulated, the more likely it is that one can (suboptimally) repress the new TE.

There is suggestive evidence for two mechanisms, potentially happening simultaneously, that play a role in the regulation of KZNF genes. Firstly, derepressed TEs can become cis-regulatory elements enhancing KZNF gene expression in the neighborhood. In support of this is the data that we have presented in Chapter 3+4 showing that KZNF genes located in proximity of epigenetically activated SVA elements are upregulated. In mouse cells, similar observations are reported upon global TE activation through KAP1 ablation (Kauzlaric et al. 2017). The fact that in naive hESCs many evolutionary young TEs connect to the promoters of KZNF genes further substantiate the hypothesis that KZNF genes are directly regulated a TEs (Pontis et al. 2019). Secondly, KZNF genes can be regulated as part of an TE-induced immune response. From cancer studies it is known that several DNA-demethylating agents owe their therapeutic effect, at least in part, to the induction of TE transcripts that in their turn trigger an interferon response (Chiappinelli et al. 2015; Roulois et al. 2015; Goel et al. 2017). For some KZNF genes it has been shown that they are upregulated by immune TFs in response to an exogenous viral infection and that they subsequently silence the virus (Nishitsuji et al. 2015; Li et al. 2018). However, evidence linking a TE-induced (auto-)immune response directly to the activation of KZNF genes is lacking so far.

Active TEs may provide recombination hotspots that fuel KZNF innovation in the long run

Heterochromatin formation at TEs not only protects against novel insertions and cis-regulation, but also against ectopic recombination causing chromosomal rearrangements (Peng and Karpen 2007; Vader et al. 2011; Kato et al. 2012; Robberecht et al. 2013; Vogt et al. 2014). On top of that, KZNF gene clusters are already sensitive to ectopic recombination as this allows for rapid adaptation to a TE outbreak (Lukic et al. 2014). Thus, active TEs within these clusters may facilitate and accelerate the generation of novel KZNF proteins that may have the ability to repress that very same TE (Fig 3B). As uncontrolled TE activity can have many deleterious effects that compromise the fitness of the host,
Genome defense against TE outburst

**A**

Acute response

- Collective KZNF upregulation
- Cis-regulation
- New insertion

Immune response

- IFNs
- ISGs

**B**

Long term response

- New TE repressor

structurral evolution

TE-derived Recombination hotspots

Legend:

- Active TE
- TE transcripts
- KZNF genes
- KZNF proteins
- Viral RNA sensors
- Immunity TF
- non-homologous recombination
ultimately individuals with the optimized/new KZNF protein will be selected for leading to fixation of this trait in the lineage.

Collectively, these findings suggest that the host is very well equipped to combat TE outbursts on multiple levels, but we are only beginning to understand the complex mechanisms that orchestrate these responses. ZNF91 ko hESCs provide a model to study the consequences of a derepressed and actively retrotransposing TE family, as if they were a novel TE family without suitable KZNF repressor. They could provide answers to several of the above formulated hypotheses. First of all the mechanism behind the defensive response against revived SVA elements could be elucidated. CRISPRi could be employed to selectively repress SVA elements in KZNF gene clusters and assess if KZNF genes are regulated by SVA-derived enhancers. In parallel, knockdown and overexpression of SVA transcripts could be used to assess the possibility that they indirectly cause the upregulation of KZNF genes, potentially through an immune response. Finally, culturing ZNF91 ko cells for an extensive amount of time would reveal the genomic consequences of unleashed SVA activity. Not only would it show the rate of retrotransposition, but also the frequency and locations of recombination events with specific interest in the KZNF gene clusters on chromosome 19.

**What is next: How to experimentally assess the functionality of candidate TE-derived enhancers in the brain?**

The identification of several TE families that show properties of cis-regulatory elements in *in vitro* differentiated brain organoids opens up possibilities for more experiments assessing their role in neuronal gene regulation (Fig 5).

**Figure 3. Hypothetical model showing genomic defense against TE invasions.**

A) The collective upregulation of KZNF genes in response to widespread TE activity possibly reflects an acute response to limit TE spread. We propose that two mechanisms may be controlling KZNF gene expression. First, active TEs located in KZNF rich genomic regions may function as enhancers for KZNF genes closely. In parallel, transcriptional activity of TEs is shown to induce an immune response and exogenous viral infections induce the expression of various KZNF genes through binding of immunity TFs such as STAT. However, evidence showing that endogenous TE activity induces a STAT-mediated upregulation of KZNF is lacking.

B) In the long run the KZNF gene cluster needs to adapt to repress the escaped TE. Active TEs may provide hotspots for ectopic recombination. As KZNF gene clusters are known to be sensitive to chromosomal rearrangements, this may be a mechanism facilitating innovation of KZNF genes ultimately leading to the formation of novel KZNF proteins that are able to repress the escaped TE family.

Dashed arrows indicate hypothetical mechanisms that need further experimental validation.
The most comprehensive way to confirm our identified candidate TE families would be to perform ChIP experiments for several key histone marks and neuronal TFs on human brain tissue (post mortem and in vitro differentiated) (**Fig 5A+B**). The amount of histone marks profiled for different brain regions is growing, mainly because of large collaborative efforts such as the ENCODE and Roadmap epigenomics projects. Many of these data sets are currently being re-analyzed to analyze the brain region specific epigenetic profile on TEs (Trizzino et al. 2018). Yet, most characterizations of TF binding sites available at the moment are done with (cancer) cell lines and are thus non-representative for the brain. Note that although profiling of binding sites of key neuronal TFs in human neurons are much needed to provide answers about the general regulatory landscape in the brain, they are highly laborious to realize, especially considering the heterogeneity of the human brain and the extensive list of TFs.

The big advantage of our neuronal in vitro differentiation protocols is that high potential TE-enhancer candidates can be experimentally assessed through (epi) genetic editing of the parental hESC line and study their regulatory impact in the context of the human genome. CRISPR/Cas9 can be used to delete specific TE insertions in hESCs and study their impact on neuronal gene expression in brain organoids (**Fig 5C**). Alternatively, CRISPRi/a offers opportunities to collectively or individually repress/activate TE families or specific insertions (**Fig 5D**). This would be an excellent way to assess if TEs ensure the collective regulation of genes in close vicinity. To study their role in neurons, stable integration of dCAS9-gRNAs into the parental hESC line could be established, which subsequently can be used as a source for neuronal tissues. Alternatively, as described in Chapter 2, hESC-derived neurons can be plated and transfected at the progenitor stage.

As mentioned earlier, profiling of 3D chromosomal interactions with chromosome conformation capture (3C) based techniques, such as hiC, could be a complementary approach to assess the target genes of specific TE-derived enhancers (**Fig 5E**). However, the usability of this technique might be restricted to long-range interactions due to the limited resolution of current hiC technology. An alternative approach, called Interaction Analysis by Paired-End Tag (ChIA-PET) Sequencing, combines ChIP with 3C to show long range chromatin interactions at TF binding sites. This provides an attractive strategy to elucidate the target genes of TEs bound by a specific TF.
Figure 5. Experimental strategies to elucidate the role of specific TE families in neuronal gene regulation.

A+B) Genome wide approaches to identify TE families that show epigenetic characteristics of cis-regulatory elements and are bound by lineage-specific TFs. Analysis of different regions of the post mortem brain and in vitro differentiated brain structures (similar to analyses done in Chapter 1) will reveal candidate TE-derived enhancers. C-E) Display strategies to experimentally assess the cis-regulatory capacities of specific TE families or insertions and their target genes. C) Deletion of specific TE insertions in hESCs, which can subsequently be differentiated in brain organoids will reveal if they play an essential role in neuronal gene regulation. D) Selective activation or repression of specific TE insertions or entire TE families with CRISPR/dCas9 fused to a repressive or activating domain will show their regulatory capacities. The dCas9-gRNA construct can be stable integrated in a hESC line, which can subsequently be differentiated into brain organoids, or transfected in neuronal progenitors. E) Chromosome conformation capture (3C) based techniques, including Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET), provide ways to profile the interactions between TEs and their target genes. Antibodies against specific TFs that bind TE families can be used for the ChIA-PET strategy. Alternatively, Specific TEs can be used as bait to identify their target sequences with 3C.
Thus, if done systematically for a number of TE insertions and TE families, above described approaches will not only lead to the identification of individual TE-derived neuronal enhancers and their target genes but also the co-regulation of gene groups involved in generating specific structures or functions. Taken together, the identification of several TE subfamilies that display a region-specific H3K27ac profile in the cortical and mDA organoids opens up new possibilities to mechanistically assess their role in neuronal gene regulation. Our in vitro differentiation protocols, described in Chapter 1, provide an excellent platform to do this.

**Final remarks**

The findings presented in this thesis support the notion that TEs are more than selfish DNA sequences that need to be silenced at all times. Our analyses reveal that KZNF proteins are the guardians of the genome that regulate TE activity in a time and place dependent manner. We show that sometimes specific TE families are co-opted as cis-regulatory elements, while KZNF-mediated repression mostly protects against, potentially harmful, cis and trans regulatory capacities of TEs. We also found that widespread activation of the SVA family elicits a genome defense response, potentially through co-option of their gene regulatory capacity, suggesting that aberrant TE activity induces a response directed to their own silencing. Together, these findings contribute to our understanding of the evolutionary battle between TEs and the KZNF gene cluster and the impact it has on gene regulation and potentially fitness of the host. The next challenge we are faced with is to connect TE-mediated innovation of gene regulatory networks to phenotypic changes. Showing which new gene groups collectively contribute to novel structures and/or functions will show the extent at which TEs are drivers of evolution. On top of that, it will open up possibilities to explore their role in disease.

The internal battle between the host genome and TEs has posed conflicts and benefits for millions of years, but they shaped our genome and inevitably contributed to the evolution of us humans as a species.
References


