Segmental duplications as a source of innovation in brain development

Lodewijk, G.A.

Publication date
2021

Document Version
Other version

License
Other

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.
Chapter 5
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

Gerrald A. Lodewijk¹, Matthijs de Geus¹, Rita L.F.P. Guimarães¹, Frank M.J. Jacobs*#¹

¹ University of Amsterdam, Swammerdam Institute for Life Sciences, Amsterdam, the Netherlands  
* Corresponding author (f.m.j.jacobs@uva.nl)  
# Lead author
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

Summary
KRAB zinc finger (KZNF) proteins are a class of transcriptional repressors important for epigenetic silencing of specific genomic regions. About 170 primate-specific KZNFs are present in the human genome. While KZNFs are primarily associated with repressing retrotransposon-derived DNA, evidence is emerging that they can be coopted for other gene regulatory processes. We focused on the 19p12 locus, which shows copy-number variations associated with neurodevelopmental disorders, implying it is required for human brain development. The two genes in this locus, ZNF675 and ZNF681, arose via duplication in primates. Genetic deletion of ZNF675 caused developmental defects in cortical organoids and our data suggest that the observed neurodevelopmental phenotype is mediated by ZNF675 binding to the promoter of the neurodevelopmental gene HES1. We show that ZNF675 interferes with HES1 auto-inhibition, a process essential for maintenance of neural progenitors. As a striking example of how some KZNFs have integrated into pre-existing gene expression networks, these findings suggest the emergence of ZNF675 has caused a change in balance of HES1 autoregulation. The association of ZNF675 CNV with human developmental disorders suggests that this new balance has become essential for normal human brain development.
**Introduction**

There are hundreds of KRAB zinc finger genes in our genome and while we are beginning to understand their communal roles as a gene family (Emerson and Thomas 2009; Nowick et al. 2010; Thomas and Schneider 2011; Jacobs et al. 2014; Schmitges et al. 2016; Ecco et al. 2017; Imbeault et al. 2017), the functions of individual KZNF genes and their importance for normal human development has remained largely elusive. Many of the KZNF genes in humans are primate-specific and based on the reported involvement of KZNFs in a variety of cellular processes, it is likely that some of them contributed to the evolution of gene expression networks and possibly the emergence of species-specific features. One of the primary roles of KZNFs is to silence retrotransposon insertions in our genome (Deininger et al. 2003; Cordaux and Batzer 2009; de Koning et al. 2011). KZNFs are DNA-binding proteins, that bind specific sequences via an array of C2H2 zinc-finger domains (Lee et al. 1989; Wolfe et al. 2000; Persikov et al. 2009; Najafabadi et al. 2017), and can recruit repressive co-factors via their N-terminal Krüppel associated box (KRAB) domain (Urrutia 2003; Zeng et al. 2008; Groner et al. 2010; Turelli et al. 2014). Even though the repressive function of KZNFs is initially important to restrict retrotransposon invasions, it has become clear that over time a decent number of KZNFs became co-opted as regulators of retrotransposon-mediated gene regulation (Jacobs et al. 2014; Ecco et al. 2016; Chen et al. 2019; Pontis et al. 2019) and regulators of genes directly through binding to specific elements in their promoters or other mechanisms (Nowick et al. 2011; Shin et al. 2011; Chauhan et al. 2013; Schmitges et al. 2016; Imbeault et al. 2017; Helleboid et al. 2019; Farmiloe et al. 2020). Disturbance of some KZNFs are associated with disease (Cassandri et al. 2017; Perdomo-Sabogal and Nowick 2019), suggesting certain KZNFs have become essential for normal human development. In this study we examine a primate-specific locus within the KZNF cluster on human chromosome 19p12, for which copy number variations (CNVs) are associated with neurodevelopmental disorders. Only two paralogous genes reside in this region: ZNF675 and ZNF681, emerged in our genome by gene duplication of a common ancestor KZNF gene specifically present in primates. In order to find a possible explanation for the neuronal phenotype associated with CNV of ZNF675 and ZNF681, we investigated their contributions to normal brain development using human cortical organoids. The findings reveal a role for ZNF675 in an essential
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

neurodevelopmental process, which may explain the neurodevelopmental phenotypes observed in 19p12 CNVs. Our results exemplify the potential of KZNFs to be co-opted as transcriptional regulators, which may have contributed important new gene-regulatory functions to the human genome throughout primate evolution. The incorporation of new primate-specific KZNFs into pre-existing gene expression networks highlights how the continuous battle between TEs and KZNFs leads to widespread fine tuning of gene expression dynamics in evolutionary adaptation.

Results

**ZNF675 and ZNF681 reside in a locus associated with neurodevelopmental disorders.**
Analysis of structural variation within the human genome has revealed several loci in our genome associated with human traits and diseases (Marques-Bonet et al. 2009; Girirajan et al. 2011). Here, we focus on copy-number variations (CNVs) in the 19p12 locus, which contains several primate-specific KZNF genes (Bellefroid et al. 1993; Eichler et al. 1998; Dehal et al. 2001; Looman et al. 2002; Hamilton et al. 2006; Tadepally et al. 2008) (Figure 1A). In this locus we identified a relatively small recurrent CNV, which encompasses only two genes: ZNF675 and ZNF681 (Figure 1B, Supplementary Figure 1). Data from the DECIPHER database (Firth et al. 2009) suggests structural variants involving these genes are in many cases linked to brain disorders, such as autism, intellectual disability, micro- and macrocephaly, ataxia and epilepsy (Figure 1C, Supplementary Table 1). In another large-scale analysis of genetic structural variation, duplications of this locus were significantly enriched in patients with developmental delay compared to controls. (Cooper et al. 2011; Coe et al. 2014). The fact that only 2 annotated genes reside in this locus and CNVs are recurrently associated with neurodevelopmental disorders, indicates that either or both ZNF675 and 56 ZNF681, have an important role in normal human neurodevelopment.

**Structural changes during ZNF675 evolution enabled it to bind a specific family of retrotransposons**
KZNFs represent a highly dynamic gene family in evolution. The repetitive character of DNA sequences in KZNF clusters cause frequent segmental
duplications and the formation of new KZNF genes. In addition, the repetitive character of the DNA sequences within each KZNF gene allow a rapid structural evolution of KZNF genes and diversification of their function. (Eichler et al. 1998; Tadepally et al. 2008; Emerson and Thomas 2009; Nowick et al. 2011; Najafabadi et al. 2017). These evolutionary changes can modify their DNA recognition properties and optimize the structure of the KZNF protein for silencing novel genomic target sites. To study the relationship between ZNF675 and ZNF681 in more detail, we reconstructed their evolutionary history. Based on sequence similarity between ZNF675 and ZNF681, these genes are closely related paralogues, suggesting ZNF675 and ZNF681 share a common parental gene (Figure 2A). Evidence for the presence of both ZNF675 and ZNF681 are found in...
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution.

Figure 2. Structural changes in ZNF675 co-occurred with an ERV invasion. A) Overview of ZNF675 and ZNF681 structures in primates. B-C) Analysis of ChIP-seq data showing type and number of transposons bound by ZNF675 and ZNF681. D) Luciferase assay using the pGL3-SV40-3x100bp-MSTA reporter containing the ZNF675 binding sequence derived from an MSTA-int element, n = 4 per condition, error bars denote standard deviation. P-values assessed by t-test followed by holm correction. E) Output from pairwise LiftOver of THE1/MST elements between human hg38 genome and other primate genomes. F) Variant data from gnomAD, showing a high frequency loss-of-function ZNF681 allele.
humans, great apes, old world monkeys and new world monkeys. In basal primate genomes, no homolog sequences of either ZNF675 or ZNF681 were found. This indicates that both the emergence of the ancestral ZNF675/ZNF681 gene and the segmental duplication that formed ZNF675 and ZNF681 happened in the last common ancestor of humans and newworld monkeys (Figure 2A). After the duplication event, ZNF675 and ZNF681 have undergone a number of structural changes that led to the diversification of their respective structural composition and DNA binding properties. ZNF675 lost 2 zinc-finger domains in the middle of the zinc-finger array. More recently, in the last common ancestor of apes and humans, the C terminus was altered by a stop codon, creating a slightly shorter protein. ZNF681 did not undergo major structural changes in the residues important for DNA sequence specificity, but has lost the overall binding capacity of several zinc-finger domains via alterations in codons coding for structurally required cysteine and histidine residues. In two large-scale KZNF binding analyses, ZNF675 was found to bind to the internal sequence of THE1 and MST types of retrotransposons, related to endogenous retroviruses (ERVs) (Najafabadi et al. 2015; Imbeault et al. 2017). Specifically, these are annotated as THE1A/B/C/D-int and MST/A/B/C/Dint, collectively named THE1/MST elements hereafter (Figure 2B). Human ZNF681 did not appear to bind any retrotransposons with high affinity in the human genome (Figure 2C). Binding of ZNF675 to THE1/MST elements was confirmed in a luciferase reporter assay, showing a strong ability of ZNF675 to repress a reporter plasmid containing a THE1/MST element (Figure 2D). As a control, ZNF681 did not display such an effect, suggesting the structural changes that diversified ZNF675 from ZNF681 and the ancestral sequence have been pivotal for its ability to recognize and repress THE1/MST elements. Indeed, the structural changes that ZNF675 underwent during evolution coincide with the timeframe during which THE1/MST elements invaded our ancestral genome. (Figure 2E, Supplementary Figure 2). Furthermore, population genetics data suggests that ZNF681 is dispensable for normal human development: We noticed a SNP in ZNF681 that causes a 2 bp deletion, resulting in a translational frameshift and early termination signal (rs61397759). According to gnomAD v2.1.1 data (Lek et al. 2016; Karczewski et al. 2019), this SNP has a 0.153 allele frequency, and a 0.025 homozygosity frequency (Figure 2F), meaning 2.5% of the human population have no functional ZNF681 in their genome at all. This variation further highlights the
instability of the genomic locus that ZNF675 and ZNF681 reside in. Taken together, our analysis indicates that ZNF675 is the only candidate for the developmental disorders associated with CNVs in this locus.

**Genetic deletion of ZNF675 results in altered neurogenesis in organoids**

Next, we investigated the function of ZNF675 in relation to the brain-developmental phenotypes in CNVs via gene knock-out. We generated isogenic human embryonic stem cell (hESC) lines deficient of ZNF675 by CRISPR-Cas9 mediated genetic deletion of exon 4, removing the complete zinc-finger array coding sequence (Figure 3A). To identify clones with successful removal of the target region, PCR and RT-qPCR genotyping was used (Supplementary Figure 3A-C). Next, under defined neuronal differentiation conditions (Field et al. 2019), cortical organoids were generated from hESC clones in two independent batches and isolated at day 28 of differentiation. The first batch consisted of hESC clones E5 +/+, F3 +/-, F5 +/- (Figure 3B) and a second independent batch consisted of clones G5 +/+, G10 +/- and B1 -/- (Figure 3C). To investigate the cortical organoid development in more detail, we compared the gene expression profile by RNA-sequencing of samples from wild-type, heterozygote, and knock-out organoids (Supplementary Table 2). The RNA-sequencing data confirmed successful knockout of ZNF675 (Supplementary Figure 3D). Cortical organoids generated from each cell line carrying the knockout, heterozygous or wild-type genotypes did not have major differences in size or growth (Supplementary Figure 3E). For differential gene expression analysis, only genes differentially expressed in both independent clone batches were selected (Figure 3D). A total of 472 genes showed expression changes in the same direction in both knockout datasets. Gene ontology analysis of upregulated and downregulated genes show that the ZNF675 knockout organoids show premature neuronal differentiation and aberrant self-renewal properties of neuronal stem cells (Figure 3E). Several key genes involved in neuronal stem cell maintenance such as HES1, SOX2 and early neurogenesis such as FEZF2, CUX2 and EMX2 were downregulated in ZNF675 knockout organoids (Figure 3F). In contrast, several genes involved in later stages of neuronal differentiation such as DCX, FOXP2 and ISL1 were upregulated in ZNF675 knockout organoids. These gene expression changes indicate that loss of...
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

Figure 3. Loss of ZNF675 leads to premature neuronal differentiation at the expense of neuronal stem cell maintenance. A) Location of CRISPR/Cas9 gRNAs for deletion of ZNF675 exon 4. B) Images of day 28 cortical organoids from 2 batches of ESC clones with indicated ZNF675 genotypes. C) Overlap of differentially expressed genes identified by RNA-seq profiling of 2 independent batches of day 28 organoids. D) GO-terms overrepresented in the selected 472 differentially expressed genes. E) Heatmaps showing expression patterns of upregulated and downregulated genes in RNA-seq data of all samples. G5 +/+: n = 2, G10 +/−: n = 1, B1 −/−: n = 2, E5 +/+: n = 2, F5 −/−: n = 2.

ZNF675 leads to premature neuronal differentiation at the expense of neuronal stem cell maintenance and early neurogenesis.

**ZNF675 regulates neuronal genes directly by binding to their promoters**

We next aimed to identify the mechanisms underlying the aberrant neuronal expression in the ZNF675 knockout cortical organoids. Since ZNF675 binds to and represses THE1/MST elements (**Figure 2D**), one possibility is that loss of ZNF675 may alter their epigenetic state and regulatory potential. Previous studies have shown the potential of LTR class retrotransposons to act as gene regulatory elements (Kunarso et al. 2010; Schmid and Bucher 2010; Gifford et al. 2013; Sundaram et al. 2014; Chuong et al. 2016; Ecco et al. 2016; Chen et al. 2019; Pontis et al. 2019). We tested a possible association between the differentially expressed genes identified in ZNF675 knockout cortical organoids and their proximity to THE1/MST elements (**Figure 4A**). The set of genes upregulated in ZNF675 knockout organoids showed a closer median distance to the nearest THE1/MST element compared to a random set of genes. This suggests a regulatory association between ZNF675-regulated neuronal genes and de-repressed THE1/MST elements, similar to what was shown for other classes of transposable elements (Jacobs et al. 2014; Ecco et al. 2016; Chen et al. 2019; Pontis et al. 2019). However, the overall effect of de-repressed THE1/MST elements on nearby gene expression seemed very modest, suggesting that other roles of ZNF675 may be important for driving normal neurodevelopment. It was recently established that some KZNFs directly bind to gene promoters, independent of the retrotransposons they have evolved to bind and repress (Schmitges et al. 2016; Imbeault et al. 2017; Farmiloe et al. 2020). We analyzed whether this was also the case for ZNF675. Based on two independent sources of
Figure 4. ZNF675 evolved to bind to endogenous gene promoters. A) Distance to the closest ZNF675 bound THE1/MST of differentially upregulated or downregulated genes from the ZNF675 knock-out RNAseq data, using a set of 485 random genes as control. P-values assessed by Kruskal-Wallis test, followed by Dunn’s test where applicable. B) Number of ZNF675 bound gene promoters identified from two independent ChIP-seq datasets. C) Chart showing the presence of repeats in ZNF675 bound gene promoters. D) Multiple sequence alignment of promoters bound by ZNF675, grouped by promoters containing no repeats and promoters with THE1/MST elements. E) Pileup of ChIP-seq reads showing ZNF675 binding the HES1 promoter. An MSTB-int element in the same region is shown for reference. Asterisks indicate zoom-in of HES1 promoter and the MSTB-int element. F) Luciferase assay using a pGL3-SV40-3x100bp-HES1 reporter containing the ZNF675 binding sequence in the HES1 promoter, n = 4 per condition, error bars denote standard deviation. P-values assessed by t-test followed by Holm correction.
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

KZNF binding data (Schmitges et al. 2016, Imbeault et al. 2017), we found that ZNF675 binds to 87 gene promoters (Figure 4B, Supplementary Table 3) and in 43 cases, the binding of ZNF675 to the gene promoter was not mediated by the presence of a THE1/MST element (Figure 4C). The DNA sequences bound by ZNF675 independent of THE1/MST elements share a clear similarity to the core ZNF675 binding site in the THE1/MST elements (Figure 5D). The importance of this 10 bp core motif for binding by ZNF675 is further strengthened by observations that show that the majority of THE1/MST elements are not bound by ZNF675 because they lack this specific core motif or show indels at the site where ZNF675 normally binds (Supplementary Figure 4A-C). This indicates that a small number of gene promoters have become bona fide ZNF675 targets after ZNF675 was optimized to recognize and repress THE1/MST elements. The close resemblance between the ZNF675 core binding motif in promoters and THE1/MST elements suggests that the promoter sequences carried a level of similarity to the invading retrotransposon class which caused them to be bound by ZNF675 as well. Within the 43 ZNF675-bound gene promoters, we noticed a number of important neurodevelopmental genes such as Microcephalin-1 (MCPH1), Sestrin 3 (SESN3) and Hes Family BHLH Transcription Factor 1 (HES1). Of particular interest was the binding of ZNF675 to the promoter of HES1, which is a key neurodevelopmental gene essential for regulating neural stem cell self-renewal (Figure 4E) (Nakamura et al. 2000; Imayoshi et al. 2010). Interestingly, HES1 is also differentially expressed in both batches of ZNF675 knock-out organoids (Figure 3F). In further support of a functional relation between ZNF675 and HES1 promoter regulation, we verified a regulatory potential of ZNF675 on the HES1 promoter in a reporter assay (Figure 4F, Supplementary Figure 5A). Even in the context of the endogenous HES1 promoter, ZNF675 showed the ability to reduce luciferase reporter gene expression (Supplementary Figure 5B). Combined, these results indicate that ZNF675 can affect HES1 gene expression in tissues where they are co-expressed. Given that tight control of HES1 expression is pivotal to the delicate balance between the maintenance of neural stem cells and neuronal differentiation, the regulatory influence of ZNF675 on the human HES1 promoter may be involved in the neurodevelopmental phenotypes associated with copy number variations in the ZNF675 locus.
ZNF675 interferes with HES1 autoregulatory feedback inhibition

The reporter assay suggested a repressive effect of ZNF675 on the promoter of HES1, which is in line with the general role of KZNFs as transcriptional repressors. In seeming contradiction to the repressive role of ZNF675 in the reporter assay, genetic deletion of ZNF675 did not lead to an upregulation of HES1 expression in ZNF675 knock-out cortical organoids. Instead, HES1 expression was reduced. One scenario that could fit with these observations is that ZNF675 competes for an even stronger repressor on the HES1 promoter. To elucidate the role of ZNF675 on the endogenous HES1 promoter, we further examined the HES1 locus in order to find regulatory mechanisms that could explain these observations. HES1 is highly expressed during brain development and its expression dynamics play an important role in fate determination of neural stem cells. In these cells, it follows an intricate oscillatory expression pattern which is governed via negative autoregulatory feedback (Hirata et al. 2002; Kageyama et al. 2007; Ochi et al. 2020). Indeed, we observed a strong repressive effect of HES1 protein on the HES1 promoter in a reporter assay (Supplementary Figure 5C). When analyzing HES1 ChIP-sequencing data (The ENCODE Project Consortium 2012; Venkataraman et al. 2018), the peaks of HES1 and ZNF675 are partially overlapping with the centers of the peaks within only 60 bp of each other on the HES1 promoter (Figure 5A). This indicates there may be competition between ZNF675 and HES1 for binding to the HES1 promoter. We hypothesized that binding of ZNF675 to the HES1 promoter interferes with HES1 negative autoregulatory feedback (Figure 5B). To test this, we created a human dynamic reporter assay previously reported for mouse cells (Masamizu et al. 2006; Nakayama et al. 2008) to measure HES1 expression dynamics in the presence or absence of ZNF675. We constructed a luciferase reporter plasmid to assess HES1 expression dynamics, containing the human HES1 promoter and destabilization sequences for both luciferase mRNA and protein that mimic the rapid turnover of HES1. This plasmid was transfected in mouse 3T3 cells, which endogenously do not have primate-specific KZNFs. After synchronization of cells via 0.2% serum treatment, the reporter signal was monitored continuously in a LumiCycle assay. In the presence of ZNF675 the first oscillatory peak of the HES1 reporter was significantly elevated in comparison to presence of ZNF681 or the EV control. This indicates that ZNF675 affects the ability of endogenous HES1 protein to
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

autoregulate the HES1 promoter (Figure 5C, Figure S5D). Eventually, the HES1 oscillation was restored again, but there was a clear delay compared to conditions where EV and ZNF681 were co-expressed. In parallel experiments, we isolated samples at fixed time-points 0, 2 and 4 hours after synchronization. These samples were measured using the dual-luciferase assay, using the renilla signal for additional normalization. Here we confirmed the observed effect of ZNF675 on HES1 promoter regulation and found an increase in reporter activity after 2 hours in the ZNF675 condition (Figure 5D, Figure S5E), which was normalized after 4 hours. Intriguingly, our data indicate that ZNF675 interferes with the normal negative autoregulatory feedback of HES1 protein on the HES1 promoter, which has been described as essential for the balance between neuronal stem cell proliferation and neuronal differentiation. This model fits with the observations in ZNF675 knock-out cortical organoids, where HES1 expression is reduced in the absence of ZNF675 and premature neuronal differentiation at the cost of stem cell maintenance is observed. Together, these data support a role for ZNF675 during human brain development via HES1 autoregulation. Future work will have to determine whether impaired HES1 autoregulation is directly associated with neurodevelopmental phenotypes identified in 19p12 CNVs, including ZNF675 and ZNF681.

Discussion
In this study, we find evidence that ZNF675 evolved as a modulator of HES1 expression by binding directly to its promoter, amongst the promoters of other neurodevelopmental genes. The ZNF675 bound elements in the gene promoters show a clear similarity to the core ZNF675 binding motif in the THE1/MST elements which seems to be the main target of ZNF675 with many thousands of binding sites throughout the genome. The fact that the ZNF675 bound sequences within the gene promoters display a very high level of DNA sequence conservation throughout mammalian evolution, suggests that these genes came under ZNF675 control as accidental bystanders of the evolutionary arms race between ZNF675 and THE1/MST elements. Importantly, we find that these novel regulatory functions by ZNF675 seem to be required for normal brain development. CNVs of the ZNF675 locus, 19p12, are associated with neurodevelopmental disorders and deletion of ZNF675 in cortical organoids leads to premature neural differentiation.
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

Figure 5. ZNF675 interferes with HES1 autoregulatory feedback inhibition. A) Comparative ChIP-seq analysis of HES1, ZNF675 and ZNF681 in the HES1 locus. B) Potential effect of ZNF675 on HES1 expression regulation. C) Continuous luciferase measurements of the pGL3-promHES1-UbqLuc-HES1-3'UTR reporter plasmid in 3T3 cells, co-transfected with EV, ZNF675 or ZNF681. Line graphs indicate mean values, shaded regions the 95% confidence interval. n = 12 per condition in 4 independent experiments. D) Dual luciferase measurements at fixed time-points, co-transfected with EV, ZNF675 or ZNF681. n = 20 per condition in 5 independent experiments.

and downregulation of the important neurodevelopmental gene HES1. As a mechanistic explanation for these observations, we established that ZNF675 interferes with HES1 auto-regulatory functions and HES1 expression dynamics. This is likely to alter the tightly controlled dynamics of neurogenesis for which HES1 autoregulation and oscillatory expression are so crucial. Given the conserved role of HES1 oscillations in neural stem cells and neurogenesis, it is intriguing to see this mechanism is influenced by recent evolutionary changes. In the search for human-specific developmental features that emerged during primate evolution, it appears that introduction of ZNF675 positively acts on HES1 expression by competing with HES1 auto-inhibition, promoting neural stem cell self-renewal at the cost of neuronal differentiation. In line with our results, in recent work it was shown that shortening the period of Hes1 oscillation increased differentiation speed of neural stem cells in the developing mouse brain (Ochi et al. 2020). As another compelling example of how genomic evolution seems to play on the expression dynamics of oscillating genes, it was recently shown that some other oscillating genes also display differences in gene-expression dynamics between human and mouse. A longer oscillation period of HES7 in human compared to Hes7 in mouse was found to account for species-specific differences in somitogenesis (Matsuda et al. 2019; Diaz-Cuadros et al. 2020). The results presented here and in other studies highlight how evolutionary changes in gene expression dynamics, such as oscillations, can contribute to the emergence of species-specific features of developmental processes. It will be intriguing to see if HES1 expression dynamics in the developing brain are a hallmark of brain developmental differences between human and other species. In relation to ZNF675 CNVs, future work will have to establish if the effect of ZNF675 on gene promoters is directly underlying the neurodevelopmental disorder phenotypes.
Other ZNF675 gene promoter targets will also have to be considered, especially because ZNF675 binds to promoters of genes associated with important neuronal functions. Examples are MCPH1 and SESN3, which are related to brain size abnormalities and seizures respectively, phenotypes also seen in the DECIPHER CNV data of ZNF675. The findings in our study lead to some implications for the evolution of KZNFs in response to retrotransposons. In this process, the host genome needs to account and select for several parameters to optimize the function of new KZNFs. One parameter is the DNA binding motif, perhaps the most important, to specifically target a retrotransposon while minimizing binding to endogenous sequences. Retrotransposons often contain repetitive DNA sequences that closely mimic mammalian promoters, presumably in order for the retrotransposon to maximally utilize the host’s transcriptional machinery for retrotransposition. Therefore, it is not entirely unexpected that KZNFs that are optimized to target a retrotransposon, end up binding to endogenous promoter sequences as well (Bourque et al. 2008, Farmiloe et al., 2020). Another parameter is KZNF expression level, which in combination with binding affinity and avidity, may determine binding of the KZNF to on-target and off-target genome sequences. The expression level of a KZNF should be sufficient to target all individual retrotransposons genome-wide to effectively prevent on-going retrotransposition. However, too high KZNF expression may saturate binding sites in retrotransposons and lead to increased off-target binding to non-retrotransposon sequences that harbor similar DNA sequence motifs, like gene promoters. Copy-number variations of KZNF genes affect the availability of KZNF proteins and thus may alter this balance. In the case of ZNF675 promoter binding is observed under endogenous expression levels, further pointing towards a co-opted role in gene promoter regulation (Schmitges et al. 2016; Imbeault et al. 2017; Farmiloe et al. 2020). While direct interactions with gene promoters may not have been the original driver of KZNF adaptation, they may lead to novel regulatory functions of KZNFs in gene regulation. These novel functions of KZNFs may become subject to selective pressure after the original target of the KZNF gene, the retrotransposon it evolved to repress no longer poses a direct threat to the host genome. The acquired role of ZNF675 in modulating HES1 autoinhibitory oscillatory pattern and the association of ZNF675 CNVs with neurodevelopmental disorders provides a compelling example of the above scenario, showing how KZNF genes can force
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

new balances and dynamics of endogenous gene regulatory networks. As many other primate-specific KZNFs also show gene promoter binding, this presents the possibility of widespread KZNF-mediated adaptation of gene regulation in recent evolution. The KZNF gene family may act as a reservoir of genetic variation, enabling rapid adaptation of existing gene networks. Due to their repetitive composition, KZNF gene clusters are more vulnerable to CNVs, highlighting the balance between genome stability and evolutionary innovation in the human genome.

Acknowledgements
We thank Marco Hoekman and Swip Draijer for assistance in Lumicycle assays (SILS, University of Amsterdam); Lars von Oerthel for assistance in FACS operation (SILS, University of Amsterdam); Violeta Méndez Ovares for work on HES1 promoter cloning; The Galaxy Project for use of their servers to analyze RNA-seq and ChIP-seq data; Marco Hoekman, Swip Draijer, Marten Smidt, and members of the Jacobs lab for helpful discussions (SILS, University of Amsterdam). This work was supported by an ERC starting grant ERC-2016-StG-716035 (F.M.J.J.).

Author contributions

Data and Code Availability Statements
The cortical organoid RNA-seq data is available on GEO / SRA under accession numbers XX / YY. Previously published sequencing datasets used are indicated in materials and methods, as well as the data processing steps used for analysis.
Contact for reagent and resource sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Frank M.J. Jacobs (f.m.j.jacobs@uva.nl).

References


A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution


Chapter 5


A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution


The bHLH Gene Hes1 as a Repressor of the Neuronal Commitment of CNS Stem Cells. J. Neurosci. 20:283–293.

Chapter 5


A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution


Figure S1. DECIPHER structural variation in the 19p12 locus and extended region. Green highlight indicates ZNF675 and ZNF681 position.
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

**Figure S2.** Amount of elements from each THE1/MST subclasses identified in various primate genomes using pairwise LiftOver from the human genome.
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution.

**Figure S3. Overview of CRISPR/Cas9 strategy for ZNF675 knockout generation.** A) Location of gRNAs flanking exon 4, and genotyping PCR primers. B) Genotyping PCR 1 for a selection of hESC clones. C) RT-qPCR to measure relative ZNF675 and ZNF681 expression levels in hESC clones used for cortical organoid generation. n = 3 per condition, error bars indicate standard deviation. D) Visualization of RNA-sequencing data of day 28 cortical organoids in the ZNF675 locus. 1 replicate is shown for each genotype of hESC clone. E) Quantification of organoid size during differentiation.
Chapter 5

A

<table>
<thead>
<tr>
<th>Element subtype</th>
<th>MSTB-int</th>
<th>MSTC-int</th>
<th>MSTD-int</th>
<th>THE1A-int</th>
<th>THE1B-int</th>
<th>THE1C-int</th>
<th>THE1D-int</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of elements</td>
<td>3361</td>
<td>933</td>
<td>213</td>
<td>770</td>
<td>1505</td>
<td>4349</td>
<td>1852</td>
</tr>
<tr>
<td>Number of bound elements</td>
<td>423</td>
<td>102</td>
<td>10</td>
<td>13</td>
<td>303</td>
<td>888</td>
<td>365</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Indel</th>
<th>Indel (zoom)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSTB-int</td>
<td>MSTB-int</td>
</tr>
<tr>
<td>MSTC-int</td>
<td>MSTC-int</td>
</tr>
<tr>
<td>MSTD-int</td>
<td>MSTD-int</td>
</tr>
<tr>
<td>THE1A-int</td>
<td>THE1A-int</td>
</tr>
<tr>
<td>THE1B-int</td>
<td>THE1B-int</td>
</tr>
<tr>
<td>THE1C-int</td>
<td>THE1C-int</td>
</tr>
<tr>
<td>THE1D-int</td>
<td>THE1D-int</td>
</tr>
</tbody>
</table>

Position in consensus sequence per retrotransposon subtype

Relative position centered to ZNF675 binding motif
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution.

Figure S4. THE1/MST elements have acquired indels in the ZNF675 binding motif. A) Number of THE1/MST elements found in the human genome. B) Frequency of indels per position in the complete THE1/MST elements and in a zoomed-in region on the binding site of ZNF675. C) Frequency of SNVs per position in the complete THE1/MST elements and in a zoomed-in region on the binding site of ZNF675. Dashed lines in complete sequence plots indicate zoomed-in regions.
Figure S5. Additional reporter assays to quantify effects of ZNF675 and HES1 in HES1 promoter regulation. A) Dose-dependent ZNF675 effects on the pGL3-3x100bp-MSTA-SV40 and pGL33x100bp-HES1-SV40 reporters in mouse 46C ESCs, n = 3 per condition. Error bars denote standard deviation. B) Effects of ZNF675 and ZNF681 on the pGL3-promHES1 reporter in mouse 46C ESCs, n = 24 per condition in 6 independent experiments. Two clones, amplified from hESC gDNA, of the HES1 promoter were used. C) Effect of HES1 on the pGL3-promHES1-SV40 reporter in mouse 46C ESCs, in comparison to ZNF675 and ZNF681, n = 4 per condition. P-values assessed by t-test followed by holm correction. Error bars denote standard deviation. D) Setup 2, t = 56, continuous luciferase measurements of the pGL3-promHES1-UbqLuc-HES1-3'UTR reporter plasmid in 3T3 cells, cotransfected with EV, ZNF675 or ZNF681. Line graphs indicate mean values, shaded regions the 95% confidence interval. n = 12 per condition in 4 independent experiments. E) Setup 2, t = 56h dual luciferase measurements at fixed time-points, co-transfected with EV, ZNF675 or ZNF681. n = 16 per condition in 4 independent experiments.
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

**Table S1.** Details of individuals in DECIPHER database with CNVs in the 19p12 locus, including ZNF675 and ZNF681.

**Table S2.** DESeq2 analysis of day 28 organoid samples, including normalized read counts.

**Table S3.** List of gene promoters identified as ZNF675 bound from ChIP-seq data of Schmitges et al. 2016 and Imbeault et al 2017, divided in non-repeat associated and repeatassociated groups.

**Table S4.** THE1/MST element and gene positions analyzed for involvement in gene regulation based on hg38 coordinates.

**Materials and Methods**

**Cell Culture**

Mouse 46C embryonic stem cells (mESCs) (Ying et al. 2003) were obtained from Austin Smith. A gelatin-adapted line was used, that is supported by a 0.1% gelatin coating (Sigma G1890). The mESCs were grown in GMEM + L-glutamine (Thermofisher 21710025), supplemented with 10% HIFBS (Thermofisher 10500064), 100u/ml Pen/Strep (Thermofisher 15140122), 1x sodium pyruvate (Thermofisher 11360039), 1x non-essential amino acids (Thermofisher 11140035), 100uM beta-mercaptoethanol (Thermofisher 31350010) and 1000u/ml LIF (Merck ESG1106, added fresh before use). For routine passaging, medium was removed and cells washed once with PBS (Thermofisher 10010056). A sterile filtered 0.25% Trypsin (Thermofisher 15090046) + 0.5mM disodium-EDTA (Sigma E5134) solution in PBS was added, and incubated at 37°C for 2 minutes. 1/12 of the cell suspension was transferred to a new culture vessel of the same size, and medium replaced daily. H9 human embryonic stem cells (hESCs, WiCell WA09) were grown as colonies on a feeder layer of mouse embryonic fibroblasts (MEFs, Thermofisher A34958). MEFs were seeded on 0.1% gelatin coated plates 24 hours before hESCs were passaged hESCs were grown in DMEM/F12 + L-glutamine (Thermofisher 11320074), supplemented with 20% Knock-out serum replacement (Thermofisher 10828028), 100u/ml Pen/Strep, 1x non-essential amino acids and 100uM beta-mercaptoethanol. The medium was sterile filtered (Merck SCGPU05RE) and 8 ng/ml recombinant FGF2 was added fresh before use each time (Sigma F0291). For routine passaging, hESC colonies were manually cut into pieces with a 27gauge needle, and lifted from the plate using a cell lifter (VWR 734-1526). The cell lifter was cut using sterile scissors to decrease the size to 1-2mm width. Cell clumps were pipetted gently to a new plate already seeded with MEFs the day before.
Medium was replaced daily. 3T3 mouse embryonic fibroblast cells (Todaro and Green, 1963), were a gift from Marco Hoekman. The 3T3 cells were cultured in DMEM + GlutaMAX (Thermofisher 10566016), supplemented with 10% HIFBS (Thermofisher 10500064), 100u/mL Pen/Strep (Thermofisher 15140122). For routine passaging, medium was removed and cells washed once with PBS (Thermofisher 10010056). Sterile filtered 0.25% Trypsin (Thermofisher 15090046) + 0.5mM disodium-EDTA (Sigma E5134) solution in PBS was added, and incubated at 37°C for 2 minutes. Cells were split 1/3 and transferred to a new culture vessel of similar size.

**Analysis of copy-number variation associated to disease**

The UCSC genome browser, human hg19, was used to analyze copy-number variation. The following tracks were accessed: DECIPHER: Chromosomal Imbalance and Phenotype in Humans (CNVs), last updated 2019-01-01 (Firth et al. 2009). Copy Number Variation Morbidity Map of Developmental Delay (Cooper et al. 2011; Coe et al. 2014). For DECIPHER data, the individuals identified from the CNV visualization on the genome browser were looked up in the DECIPHER database at https://decipher.sanger.ac.uk/, to retrieve information about possible additional CNVs for each case, and data regarding inheritance. Comparative sequence analysis for ZNF675 and ZNF681 evolutionary history

The human ZNF675 and ZNF681 cDNA and protein sequences were aligned to other species’ sequences using BLAT on the UCSC genome browser and NCBI BLAST. For comparative analysis, the following genomes were used:

- Human ZNF675, ZNF681: hg38
- Chimpanzee ZNF675, ZNF681: panTro5
- Gorilla ZNF675, ZNF681: gorGor4
- Orangutan ZNF675, ZNF681: ponAbe3
- Gibbon ZNF675, ZNF681: nomLeu3
- Rhesus macaque ZNF675, ZNF681: rheMac8
- Crab-eating macaque ZNF675, ZNF681: macFas5
- Marmoset ZNF675: calJac3
- Squirrel monkey: saiBol1
- White-faced Sapajou: cebCap1
- Ma’s night monkey: aotNan1
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

**ChIP-sequencing data analysis**

Raw fastq files were downloaded from the EBI SRA. Files were trimmed using Trimmomatic (Galaxy v0.36.5) using the following settings: SLIDINGWINDOW: window=4, qual=20 and MINLEN: 30. The Illuminaclip command was specified when necessary to remove adapters from reads. Reads were mapped to hg19 or hg38 reference genome using Bowtie2 (Galaxy v2.3.4.2) with the very sensitive end-to-end preset. BAM files were converted to Bigwig files for visualization on the UCSC genome browser with deepTools bamCoverage (Galaxy v3.0.2.0), bin size 1, without normalizing or scaling. The following datasets were used:

ZNF675: SRR5197211, SRR1370874  
ZNF681: SRR5197214  
HES1: SRR5444303, SRR5111087, SRR6212666

Association of KZNF ChIP-seq data with transposable elements

ChIP-seq datasets in bigwig format were analyzed using the UCSC genome browser Table browser tool (Kent et al. 2002). ZNF675 and ZNF681 data was converted to a BED file, selecting only regions were the peak signal was equal or higher than 20 mapped reads. This was intersected with the RepeatMasker table (update 2014-01-10) without simple repeats or satellite repeats. For the overview of bound elements, all THE1-int elements were summed and include THE1-int, THE1A-int, THE1B-int, THE1C-int and THE1D-int elements as annotated in RepeatMasker. For MST-int elements, these include MST-int, MST-A1-int, MST-B-int, MST-C-int and MST-D-int. Collectively these are named THE1/MST in this study. The promoter analysis was done in the same way, the cutoff for bigwig to BED conversion was set a signal of 10 or higher instead. Promoter regions were based on hg38 coordinates of GENCODE V27. For each transcript, the coordinates of the region 5kb upstream and 1kb downstream of the transcription start site were extracted, and used to intersect with the KZNF ChIP-seq data. Each hit was manually curated for artifacts to exclude false positives.

**Presence of THE1/MST elements in other primate genomes**

The coordinates for each subclass of THE1/MST element were extracted from the hg38 RepeatMasker table from the UCSC genome browser. The LiftOver tool was used to find the amount of homologous loci in other genomes, with the minimum amount of bases that must re-map set to 0.95. The analysis was done pairwise between the human genome and each primate genome available. The results were grouped per primate lineage and data
represents in a dot plot with 30 bins for the element number. The following genomes were included in the analysis:

**Great apes:** Chimpanzee (panTro6), Bonobo (panPan2), Gorilla (gorGor5), Orangutan (ponAbe3).

**Apes:** Gibbon (nomLeu3)

**Old-world monkeys:** Rhesus macaque (rheMac8), Crab-eating macaque (macFas5), Green monkey (chlSab2), Baboon (papAnu2), Golden snub-nosed monkey (rhiRox1), Proboscis monkey (nasLar1)

**New-world monkeys:** Marmoset (calJac3), Squirrel monkey (saiBol1)

**Basal primates:** Tarsier (tarSyr2), Bushbaby (otoGar3), Mouse lemur (micMur1)

**THE1/MST binding analysis for ZNF675**

To obtain the sequences of THE1*int and MST*int elements that are bound by ZNF675, a custom track was made on the UCSC genome browser using the Table Browser tool (Kent et al. 2002) from ChIP-seq data of ZNF675 (Schmitges et al. 2016; Imbeault et al. 2017). Subsequently, this data was intersected with the Repeats track filtered by repeat name set to match the THE1*int and MST*int elements. For the sequences of the THE1*int and MST*int elements that were not bound by ZNF675 the same steps were taken, but intersected instead with ZNF675 peaks lower than 20. The resulting sequences were exported as FASTA files. Next, the consensus sequence of each element was obtained from the UCSC genome browser on human transposons repeat consensus (build 05/14 assembly, repeats2), accessed at http://bit.ly/repbrowser (Fernandes et al. 2020). These were THE1A-int, THE1B-int, THE1C-int, THE1D-int, MST-A-int, MST-B-int, MST-C-int and MST-D-int. Using the LASTZ tool (Galaxy v1.3.2) the sequences of both bound and unbound elements were aligned against the corresponding consensus repeat sequence, using default parameters. Finally, the LASTZ output BAM files were used to calculate for each element the indel and substitutions per nucleotide, using pysamstats (v1.1.2) (https://github.com/alimanfoo/pysamstats).

**Design and cloning of CRISPR-Cas9 constructs**

The pX330 vector (Addgene 42230) (Cong et al. 2013) was used to insert sgRNA target sequences, and co-express Cas9 and gRNAs in cells. CRISPOR v4.0 [REF] was used to design gRNA sequences in the flanking regions of ZNF675 exon 4, and synthetic oligonucleotides were ordered with CACC and AAAC overhangs. The first 5’ nucleotide
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

was changed to guanines to improve sgRNA expression when necessary. The pX330 vector was digested with BbsI, and annealed oligonucleotides were inserted by ligation with Quick ligation kit (NEB M2200). Oligonucleotides used are listed below, overhangs for cloning are noncapitalized:

gRNA1_ZNF675_sense: caccGTACACTTAATTATGGCCTG
gRNA1_ZNF675_antisense: aaacCAGGCCATAATTAAGTGTAG
gRNA2_ZNF675_sense: caccGATATAAGTGACACATGAGG
gRNA2_ZNF675_antisense: aaacCCTCATGTGTCACTTATATG

CRISPR-Cas9 of H9 human embryonic stem cells

Cells were prepared using a modified published protocol (Hendriks et al. 2015). H9 hESCs colonies growing on MEFs were passaged to Matrigel (Corning 354277) coated dishes. For culture on matrigel, hESCs were grown in MEF-condition hESC medium. This was generated by adding hESC medium without FGF2 to a MEF feeder layer for 24 hours, then collecting and filtering the conditioned medium. This was used directly or stored at -20 and used within 3 months. FGF2 was added freshly before adding conditioned hESC medium to hESC cultures on matrigel. 24 hours before transfection, hESC cultures on matrigel were passaged using Accutase (Sigma A6964) to obtain a single cell culture and 2μM Thiazovivin (Sigma SML1045) was added to the medium to increase survival. On 1 60mm dish, 770000 cells were seeded. 1 hour before transfection, medium was replaced with hESC medium without pen/strep. Transfection mixes were made according to the Lipofectamine Stem protocol (Thermofisher STEM00003). For 1 60mm dish, 1250ng of pX330-sgRNA1_ZNF675, 1250ng pX330sgRNA2_ZNF675 and 500ng of pCAG-GFP were mixed in 75ul of OptiMEM. In a separate tube, 12.5ul of Lipofectamine STEM was mixed with 62.5ul of OptiMEM. The 2 solutions were mixed and incubated 20 minutes at room temperature and the transfection complex was added to the 60mm dish. 16 hours after transfection medium was replaced. 48 hours after transfection, hESCs were prepared for FACS. A single-cell solution was made using Accutase. PBS was added to dilute Accutase, and cells were centrifuged at 1000xg to pellet cells, following one more wash in PBS. Then, cells were resuspended in cold 1x PBS + 2μM Thiazovin. The cell suspension was applied to a 30μm cell strainer (Sysmex ZPS40464) column to remove remaining cell clumps, and collected in a Falcon tube (Corning 352058). During these steps, cells were kept on ice, and cell sorting and collection was done at 4°C. A FACSAlia III (BD Biosciences) was used for sorting, at 20psi and using a 100μm nozzle. Nontransfected cells were used to set the gates for GFP-negative cells. 60000-80000 transfected cells (GFP+, 4%-10% of total cells)
were collected in one well of a Matrigel coated 6-well plate, and cultured in hESC+MEF-conditioned hESC medium with 2μM Thiazovivin and 20mM HEPES pH7.4 for the first 24 hours. The next days, medium was replaced with MEF-conditioned hESC medium until ready to passage. Single cell clones were grown by seeding 800 cells in hESC medium with 2μM Thiazovivin on a 60mm dish containing a MEF feeder layer. After 24 hours, medium was replaced containing no Thiazovivin. Individual colonies were picked at about 1mm size, and transferred in MEF-conditioned hESC medium to a Matrigel coated 96wells plate. By pipetting up and down in the well, each colony was broken up into smaller pieces. When hESC colonies recovered, they were passaged again using Accutase, 1/10 to a new 96-well plate for further culturing, and 1/3 to a new 96-well plate for DNA isolation and genotyping. Selected clones according to genotyping were cultured on a MEF-feeder layer, by seeding 800 cells on a 60mm dish as described before, and cultured as colonies according to the regular protocol.

DNA isolation, RNA isolation of hESCs and genotyping PCRs
Genomic DNA was isolated directly from the 96-wells plate (Hendriks et al. 2015). Medium was removed each well, and washed twice with 1x PBS. 50μl of lysis buffer was added to each well (10mM Tris-HCl pH8.0, 10mM disodium-EDTA, 10mM NaCl, 0.5% w/v Sarcosyl, 40μg/ml proteinase K). The plate was sealed with parafilm and kept in a humid container, and placed in a 56°C incubator overnight. DNA was precipitated by adding 100μl of 75mM NaCl in 95% ethanol to each well, incubating for 2h at -20°C. DNA was pelleted by centrifugation for 5 minutes at 300xg. Supernatant was removed by pipetting, and the DNA pellet was washed twice by adding 200μl of 70% ethanol was added to each well, centrifuging 5 minutes at 300xg and removing the solution. The plate was air-dried for 15 minutes, and gDNA in each well was dissolved in a 50μl solution of 10mM Tris pH8.0, 1mM disodium-EDTA and 0.1mg/ml RNAse A. The plate was incubated at room temperature overnight, the first 2 hours on a shaking platform. For long-term storage, DNA samples were kept at -20°C. Genotyping PCR was done using the LongAmp Taq PCR kit (NEB M0323), according to the standard reaction protocol. 1ul of genomic DNA was used in PCR reactions with a total volume of 20ul. Primers used are listed below:

ZNF675 exon4 5’: ACCTGGCTGTAATTACTTCTG
ZNF675 exon4 3’: GGGATTACTAGAAATGTGTTC
ZNF675 del-internal: AGCTCAGAAATATCTAAGGC
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

For RT-qPCR analysis, selected hESC clones were passaged using Accutase. Cells from each clone were seeded in triplicate, at a density of 25000 cells per well on a Matrigel coated 24-well plate in MEF-conditioned hESC medium + 2μM Thiazovivin. After 24 hours, cells were cultured without Thiazovivin and after 48 hours cells were harvested in 300μl TRizol reagent per well (Thermofisher 15596018) and stored at -80°C. RNA was isolated using Direct-zol RNA microprep (Zymo Research R2062) with DNaseI treatment according to standard protocol. Relative expression was analyzed by RT-qPCR using QuantiTect SYBR green RTPCR kit (Qiagen 204243). 25ng of total RNA was used per 10μl reaction. Reactions were carried out on a LightCycler 480 II (Roche), using LightCycler 480 96-well plates (Roche 04729692001).

ZNF675 qFw: CAAGAAAAAGAGCCTTTGACTG
ZNF675 qRv: CACACTTTTACAGCCTTTTAACTG
ZNF681 qFw: GGACTAGAAAGAGACATAGG
ZNF681 qRv: CCTCCTTTTTGCACCTTTGCAC
ACTB qFw: TCCCTGGAGAAGAGCTACGA
ACTB qRv: GCACTGTGTTGGCGTACAG

Generation of cortical organoids.
H9 hESCs were grown as colonies on a MEF feeder layer. Before organoid formation, medium was replaced to hESC medium + 1x sodium pyruvate and without FGF2. Colonies sized 1.52mm on a 100mm dish were scooped using a cell lifter, which was cut using sterile scissors to decrease the size to 2-3mm width. The suspension was transferred to a 60mm ultra-low attachment dish (Sigma CLS3261) by gentle pipetting with a wide-bore tip. The dish was then placed overnight in the incubator for organoids to form. Then, 50% of the medium was replaced with differentiation medium consisting of DMEM/F12 + 2mM L-glutamine, 20% KOSR, 100u/ml Pen/Strep, 1x non-essential amino acids, 1x sodium pyruvate and 4 small molecule inhibitors: 1μM Dorsomorphin (Sigma P5499), 10μM SB431542 (Sigma S4317), 3μM IWR-1 (Sigma I0161) and 1μM Cyclopamine (Sigma C4116). For the initial 50% medium replacement, concentration of the 4 inhibitors was doubled to achieve the required concentration for each in the final volume. Thereafter medium was replaced every other day with differentiation medium until day 18. At this stage medium was replaced to Neurobasal medium (Thermofisher 21103049), supplemented with 2mM L-glutamine (Thermofisher 25030024), 100u/ml Pen/Strep and 1x N2 supplement (Thermofisher 17502048). Until day 24, 1μM Cyclopamine was added. At day 28, organoids were harvest for RNA isolation. Per replicate, 8-12 organoids were pooled
and dissolved in 1ml TRIzol reagent, stored at -80°C for later use. RNA was isolated from TRIzol using Direct-zol RNA microprep with DNaseI treatment.

RNA-sequencing data processing and analysis
RNA samples were prepared for sequencing using the NEBNext Ultra Library prep kit, and sequenced on an Illumina HiSeq 4000 platform with paired-end 150bp reads (GenomeScan). Data was analyzed using public Galaxy servers (Afgan et al. 2018). The output fastq files were trimmed using Trimmomatic (Galaxy v0.36.5) with parameters “ILLUMINACLIP: Truseq v3 paired-end, SLIDINGWINDOW: window=4, qual=20 and MINLEN: 30”. Reads passing QC were mapped onto the hg38 reference genome using HISAT2 (Galaxy v2.1.0) with paired-end settings: "strand information: reverse -RF, mate orientation: --fr" and using the GENCODE v27 .gtf file for known splice sites, and parameters “-k 20 and --max-seeds 20”. Raw read counts were generated with featurecounts (Galaxy v1.6.1.0) using GENCODE v27 annotation with settings “stranded -reverse, count fragments instead of reads, only allow fragments with both reads mapped, exclude chimeric reads, minimum mapping quality 12”. Read count normalization and paired wild-type versus knock-out statistics were done using DESeq2 (Galaxy v2.11.40.2). Coverage tracks in bigwig format were made using deepTools bamCoverage (Galaxy v3.0.2.0) with settings "bin size 1, do not normalize or scale, ignore missing data, minimum mapping quality 1", separating data into two files containing reads from either forward or reverse strand only. Bigwig files were visualized on the UCSC genome browser and scaled according to total read count per sample. The RNA-sequencing data confirmed successful knockout of ZNF675 exon 4 removal in clone F5, as no expression is observed in the targeted region. For clone B1, some expression was observed mapping on the positive strand, despite confirmed removal of exon 4 in each PCR approach. The origin of this expression is likely attributed to random integration of the excised ZNF675 sequence, but because transcription is derived from the non-sense strand, this transcript cannot produce functional ZNF675. Z-scores calculated across all samples using DESeq2 normalized read counts and visualized in heatmaps using Multiple Experiment Viewer (v4.9.0). For GO-term over-representation test, PantherDB (v13.1) was used (Mi et al., 2017). The background gene list consisted of all genes with a basemean expression of 50 or higher in the day 28 organoid RNA-seq data. There were 221 upregulated or 251 downregulated genes with a basemean expression of 50 or higher, which were tested against the background list for overrepresentation in the GO biological process complete annotation. Test type was Fisher’s Exact with FDR correction.
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

Distance between gene transcription start sites and THE1/MST elements
The transcription start site (TSS) of genes was extracted from the GENCODE V27 annotation. To select the control random genes, a list of all non-differentially expressed genes with basemean expression 50 or higher was made. 485 random genes were selected using the “sample()” function in RStudio (V1.1.463). THE1/MST elements were extracted from the RepeatMasker table in the UCSC genome browser (hg38), filtered with “repName” matching “THE1*-int AND MST*-int”. Elements with a length between 4000-1250 nucleotides were included in the analysis. The closest distance between TSS of all upregulated, all downregulated, or the random gene list, to THE1/MST elements was extracted using the tool “Fetch closest non-overlapping feature for every interval” (Galaxy V4.0.1) with parameter “either upstream or downstream”. For all positions analyzed, see Table S4.

Luciferase reporter assays
Human ZNF675 and ZNF681 cDNAs were synthesized (Thermofisher GeneArt) codonoptimized for expression in mouse, and a kozak consensus sequence upstream of the start codon. generated. Human HES1 cDNA was retrieved from a previously constructed plasmid (Addgene #17624) (Yu et al. 2006) For expression in mammalian cells, the cDNAs were subcloned into the pCAGEN vector backbone (Addgene 11160) (Matsuda and Cepko, 2004). The reporter elements for MSTA-int and HES1 promoter were cloned in the pGL3SV40 vector (Promega E1761). These elements consists of a 3x100bp synthesized sequence, retrieved from a 100bp interval at the peak of a full-length MSTA element (pGL3-SV403x100bp-MSTA), or the HES1 promoter (pGL3-SV40-3x100bp-HES1) (Thermofisher GeneArt). Also, a HES1 promoter sequence of 854bp was synthesized and cloned into pGL3SV40 (pGL3-SV40-promHES1(854bp)).

100bp MSTA element sequence:
AGACCTTCAAGCGATCCCTCCCCCATCACAGCCTGGAGGTCTAGGAGGACAAAATGGTTTGTGGGCCAGCACAGGACAGCCTTGGGACACTGCTGCCTGCCAGCACAGGACAGCCTTGGGACACTGCTGCCTG

100bp HES1 element sequence:
CTAAACCAAAGCCCAGAGGGAGAGTAGAAGGGTTAATATCCCTTTGTAGGAGTTGTGTTGTTGTTTGTGGGCCAGCACAGGACAGCCTTGGGACACTGCTGCCTG

854bp HES1 promoter sequence:
GGTAGCCAGGCTCAGGAATGAAAGGGAAAGGGTGGTGAAGGAGGCCGGCAGTCTCGATCGATAAGTCCCTTCTCCACCCAGTTGGAAAAATATTCTCTTG
CGGAGGCCCCCTTTGTTGTTATTTTCAATTGGGAAAATACGGTTTCTCTCCCTCCACCCAGTTGGAAAAATATTCTCTTG
ATTGTCTTCCCTGGGAAAGCCCGGGGCTCAAGGCCAAAATCTAAAGGAGAAAGCTTCTTGTTGTTTGTGGGCCAGCACAGGACAGCCTTGGGACACTGCTGCCTG

GGTAGCCAGGCTCAGGAATGAAAGGGAAAGGGTGGTGAAGGAGGCCGGCAGTCTCGATCGATAAGTCCCTTCTCCACCCAGTTGGAAAAATATTCTCTTG
CGGAGGCCCCCTTTGTTGTTATTTTCAATTGGGAAAATACGGTTTCTCTCCCTCCACCCAGTTGGAAAAATATTCTCTTG
ATTGTCTTCCCTGGGAAAGCCCGGGGCTCAAGGCCAAAATCTAAAGGAGAAAGCTTCTTGTTGTTTGTGGGCCAGCACAGGACAGCCTTGGGACACTGCTGCCTG
Chapter 5

The reporter assays were done in 46C mESCs as these lack human-specific KZN F genes and transposon sequences. Cells were seeded on 0.1% gelatin coated 24-wells plates, at a density of 60000 cells per well. After 24 hours, cells were transfected with reporter mix as listed (Table 1). The amount of KZN F plasmid and pGL3-SV40 plasmid was scaled to molar ratio, according to the size of each plasmids, to ensure equal promoters for each condition. For control conditions, each KZN F was transfected with pGL3-SV40 (no insert) instead. Also, pCAGKZF was replaced with pCAG-EV control for each condition. pRL-TK (Promega E2241) was used for normalization. Transfections were done using Lipofectamine 3000 (Thermofisher L3000015). Medium was replaced 1 hour before transfection with 450ul complete medium without P/S. DNA-Lipid complexes were made by mixing 25μl OptiMEM (Thermofisher 31985047) containing 400ng DNA mix and 0.8μl P3000, with 25μl OptiMEM containing 0.8μl Lipofectamine 3000. After incubating for 20 minutes at room temperature, the mix was added to the well. Reactions were scaled up accordingly for multiple transfections and replicates. 24 hours after transfection cells were isolated for reporter assay. Each well was washed once with 1x PBS, and incubated with 100μl 1x passive lysis buffer (Promega E1941) for 15 minutes on a shaking platform. Plates were wrapped in parafilm and stored at -80°C. For measuring luciferase and renilla activity, the Dual-luciferase reporter assay system was used (Promega E1980). 20μl sample was pipetted to a 96-well optiplate (PerkinElmer 6005290). Samples were measured on a GloMax Navigator device (Promega GM2010), with the following settings: Injector 1, LARI II buffer (volume 50μl, speed 200μl/s). Wait 2s. Measure luminescence Luciferase (integration 10s, readings 1, interval 0.3s). Injector 2, Stop & Glo buffer (volume 50μl, speed 200μl/s). Wait 2s. Measure luminescence Renilla (integration 10s, readings 1, interval 0.3s).
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

<table>
<thead>
<tr>
<th></th>
<th>pCAG EV</th>
<th>pCAG ZNF675</th>
<th>pCAG ZNF681</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGL3-SV40 / +insert</td>
<td>50 / 53</td>
<td>50 / 53</td>
<td>50 / 53</td>
</tr>
<tr>
<td>pRL-TK</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>pCAG-EV</td>
<td>210</td>
<td>186</td>
<td>186</td>
</tr>
<tr>
<td>pCAG-KZNF</td>
<td>0</td>
<td>32</td>
<td>34</td>
</tr>
<tr>
<td>pBluescript</td>
<td>126</td>
<td>118</td>
<td>116</td>
</tr>
<tr>
<td>pCAG-GFP</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 1: Amount of plasmid DNA (ng) transfected per condition. pGL3-promoter 50 / 53 denote amount of plasmid used for EV / +insert conditions.

For the dose-dependent response of ZNF675 to the 3x100bp-MSTA and 3x100p-HES1 promoter reporter, the transfection mixes per reaction were scaled accordingly to maintain equal molar ratios (Table 2).

<table>
<thead>
<tr>
<th></th>
<th>pCAG EV</th>
<th>pCAG ZNF675 0.4ng</th>
<th>pCAG ZNF675 2.0ng</th>
<th>pCAG ZNF675 10ng</th>
<th>pCAG ZNF675 33ng</th>
<th>pCAG ZNF675 66ng</th>
<th>pCAG ZNF675 100ng</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGL3-SV40 / + insert</td>
<td>50 / 53</td>
<td>50 / 53</td>
<td>50 / 53</td>
<td>50 / 53</td>
<td>50 / 53</td>
<td>50 / 53</td>
<td>50 / 53</td>
</tr>
<tr>
<td>pTK-Renilla</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>pCAG-EV</td>
<td>210</td>
<td>210</td>
<td>208</td>
<td>203</td>
<td>186</td>
<td>161</td>
<td>136</td>
</tr>
<tr>
<td>pCAG-KZNF</td>
<td>0</td>
<td>0.4</td>
<td>2</td>
<td>10</td>
<td>33</td>
<td>66</td>
<td>100</td>
</tr>
<tr>
<td>pBluescript</td>
<td>126</td>
<td>126</td>
<td>126</td>
<td>123</td>
<td>117</td>
<td>109</td>
<td>100</td>
</tr>
<tr>
<td>pCAG-GFP</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 2: Amount of plasmid DNA (ng) transfected per condition. pGL3-promoter / + insert 50 / 53 denote amount of plasmid used for EV / +3x100bp element insert conditions.
HES1 luciferase reporter construct
In order to investigate the effects of ZNF675 on HES1 oscillations, a real-time luciferase assay was performed to track the activity of the HES1 promoter using the pGL3-promHES1-UbQuLucHES1-3'UTR vector. To generate this vector, first a 2835bp region of the human HES1 promoter was isolated by PCR using the following primers:

HES1_prom_F1_KpnI: TATAGGTACCTACCTCTATCCAGACATACG
HES1_prom_R1_NheI: GACTGCTAGCCAAACTACTGAGCAAGTAGC

Amplified region: UCSC genome browser hg38 chr3:194133446-194136281

This sequence was cloned into the pGL3-basic vector using KpnI and NheI restriction sites. Next, the sequences for the ubiquitin-G76V and the HES1 3'UTR were synthesized (Thermofisher GeneArt). Ubiquitin-G76V was N-terminally fused in-frame with the luciferase gene using the HindIII and NcoI restriction sites. The HES1 3'UTR was inserted using the XbaI restriction site downstream of the luciferase gene.

Ubiquitin-G76V sequence:
GACTAAAGCTTTGCCACACCAGACGATCCTCTGTTGAACACTGGACGGAGCATCAGACTGGGAAGTG
AACCACGGCAAGGCAACATCGAGAACTGAAAGCGGAACAGAGCATCAGACTGGGAAGTGG

HES1 3'UTR sequence:
CGAATTGGCGGAAGGCCGTCAAGGCCACGTGTCTTGTCCAGACATCAGACTGGGAAGTG
AACCACGGCAAGGCAACATCGAGAACTGAAAGCGGAACAGAGCATCAGACTGGGAAGTGG

For assessing effects of ZNF675 and ZNF681 on the 2835bp HES1 promoter sequence, a stable luciferase reporter containing only the HES1 promoter was tested first (pGL3promHES1). The transfection mixes per reaction were scaled accordingly to maintain equal molar ratios (Table 3).
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

<table>
<thead>
<tr>
<th></th>
<th>pCAG EV</th>
<th>pCAG ZNF675</th>
<th>pCAG ZNF681</th>
<th>pCAG HES1</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGL3-Basic/+ promHES1</td>
<td>50 / 79</td>
<td>50 / 79</td>
<td>50 / 79</td>
<td>50 / 79</td>
</tr>
<tr>
<td>pRL-TK</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>pCAG-EV</td>
<td>210</td>
<td>186</td>
<td>186</td>
<td>186</td>
</tr>
<tr>
<td>pCAG-KZN1 or pCAG-HES1</td>
<td>0</td>
<td>32</td>
<td>34</td>
<td>33</td>
</tr>
<tr>
<td>pBluescript</td>
<td>126</td>
<td>118</td>
<td>116</td>
<td>117</td>
</tr>
<tr>
<td>pCAG-GFP</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3: Amount of plasmid DNA (ng) transfected per condition. pGL3-Basic / + promHES1 50 / 79 denote amount of plasmid used for EV / + HES1 promoter insert conditions.

To also measure the effect of HES1 protein on the HES1 promoter, pCAG-HES1 was cotransfected with the pGL3-SV40-promHES1(854bp). As control, pCAG-ZNF675 or pCAGZNF681 were co-transfected too. The transfection mixes per reaction were scaled accordingly to maintain equal molar ratios (Table 4).

<table>
<thead>
<tr>
<th></th>
<th>pCAG EV</th>
<th>pCAG ZNF675</th>
<th>pCAG ZNF681</th>
<th>pCAG HES1</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGL3-SV40/+ promHES1(854)</td>
<td>50 / 59</td>
<td>50 / 59</td>
<td>50 / 59</td>
<td>50 / 59</td>
</tr>
<tr>
<td>pRL-TK</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>pCAG-EV</td>
<td>210</td>
<td>186</td>
<td>186</td>
<td>186</td>
</tr>
<tr>
<td>pCAG-KZN1 or pCAG-HES1</td>
<td>0</td>
<td>32</td>
<td>34</td>
<td>33</td>
</tr>
<tr>
<td>pBluescript</td>
<td>126</td>
<td>118</td>
<td>116</td>
<td>117</td>
</tr>
<tr>
<td>pCAG-GFP</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 4: Amount of plasmid DNA (ng) transfected per condition. pGL3-SV40 / + promHES1 50 / 59 denote amount of plasmid used for EV / + 854bp HES1 promoter insert conditions.
Real-time luciferase assay of HES1-promoter luciferase reporter
For the real-time luciferase measurements of the effect of ZNF675 on the 2835bp HES1 promoter, 3T3 cells were seeded on 9 cm² dishes 24 hours prior to transfection. Cells were transfected as described in the table below (Table 5, setup #1), using Lipofectamine 3000 (Thermofisher L3000015) following standard manufacturers procedure. 56 hours after transfection, cells were synchronised for 16 hours O/N in low serum medium (0.2% HiFBS) supplemented with 1 mM D-Luciferin (sodium salt) (Cayman Chemical 14682). At the start of the real-time luciferase measurements, medium was changed to normal DMEM without phenol red with 4.5 g/L D-Glucose, 4.0 mM L-Glutamine and 25mM HEPES (Thermofisher 21063029), supplemented with 10% HiFBS, 100 u/mL Pen/Strep, 25 ng/mL bFGF (Sigma F0291-25UG) and 1 mM D-Luciferin (sodium salt), after which the dishes were placed into the Lumicycle (Actimetrics). In an additional experiment (Table 5, setup #2) 3T3 cells were measured twice. Cells were first synchronized for 8 hours with low serum medium at 48 hours after transfection and subsequently measured in the Lumicycle. After this cells were synchronised a second time for 8 hours at 72 hours after transfection and measured in the Lumicycle.

<table>
<thead>
<tr>
<th></th>
<th>Setup #1</th>
<th>Setup #2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>ZNF675</td>
</tr>
<tr>
<td>pGL3-promHES1-UbqLuc-HES1-3'UTR</td>
<td>500 ng</td>
<td>500 ng</td>
</tr>
<tr>
<td>pCAG-GFP</td>
<td>20 ng</td>
<td>20 ng</td>
</tr>
<tr>
<td>pCAG-EV</td>
<td>840 ng</td>
<td>744 ng</td>
</tr>
<tr>
<td>pCAG-KZNF</td>
<td>0 ng</td>
<td>128 ng</td>
</tr>
<tr>
<td>pBluescript</td>
<td>660 ng</td>
<td>631 ng</td>
</tr>
</tbody>
</table>

Table 5: Co-transfections for the real-time luciferase assay.

Dual-luciferase assay of HES1-promoter luciferase reporter
Parallel to the two setups for the real-time luciferase assay, a dual-luciferase assay was also performed on specific timepoints after synchronization. Transfections were optimized for the Dual Luciferase assay (Table 6). Synchronization of the cells was initiated in the
A novel KRAB zinc finger gene has changed the balance of HES1 autoregulation during primate evolution

same fashion as in the two setups of the real-time luciferase experiment. Cells were isolated at the moment of synchronization and subsequently isolated every two hours according to standard manufacturer's procedure (Promega E1910). Luciferase activity was measured using the GloMax Navigator (Promega GM2010). Per reaction, 50 µL of Luciferase Assay Reagent II was injected to 20 µL substrate followed by a two seconds delay after which the firefly luciferase activity was measured for ten seconds. Subsequently 50 µL of Stop & Glo Reagent was injected followed by a two seconds delay after which the Renilla luciferase activity was measured for ten seconds.

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>ZNF675</th>
<th>ZNF681</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGL3-promHES1-UbqLuc-HES1-3’UTR</td>
<td>50 ng</td>
<td>50 ng</td>
<td>50 ng</td>
</tr>
<tr>
<td>pRL-TK</td>
<td>5 ng</td>
<td>5 ng</td>
<td>5 ng</td>
</tr>
<tr>
<td>pCAG-GFP</td>
<td>5 ng</td>
<td>5 ng</td>
<td>5 ng</td>
</tr>
<tr>
<td>pCAG-EV</td>
<td>210 ng</td>
<td>186 ng</td>
<td>186 ng</td>
</tr>
<tr>
<td>pCAG-KZNF</td>
<td>0 ng</td>
<td>32 ng</td>
<td>34 ng</td>
</tr>
<tr>
<td>pBluescript</td>
<td>126 ng</td>
<td>118 ng</td>
<td>116 ng</td>
</tr>
</tbody>
</table>

Table 6: Co-transfection for the Dual-luciferase experiments.

Statistics
For luciferase assays, two-tailed t-tests were done using Holm’s method for multiple testing correction. Replicate number and independent experiment number are listed in figure legends. Corrected p-value < 0.05 was considered significant. For RNA-sequencing data analysis, DESeq2 (Galaxy v2.11.40.2) was used. Per batch of organoid data, a paired analysis was done between wild-type and knockout datasets (n=2 each). A p-adj. value < 0.05 was considered significant. The resulting lists from each batch were overlapped with each other for final gene selection, only keeping genes that were significantly differentially expressed both in batches and in the same direction upregulated or downregulated.

Distance of transcription start site to the closest THE1/MST element was analyzed using Kruskal-Wallis test, followed by Dunn’s test for pairwise analysis. Corrected p-value < 0.05 was considered significant.