RESEARCH

Canine hyper-sociability structural variants associated with altered three-dimensional chromatin state

Dhriti Tandon[1*], Enikő Kubinyi[2,3,4], Sára Sándor[3], Hannah Faughnan[1], Ádám Miklósi[2] and Bridgett M. vonHoldt[1*]

Abstract

Strong selection on complex traits can lead to skewed trait means and reduced trait variability in populations. An example of this phenomenon can be evidenced in allele frequency changes and skewed trait distributions driven by persistent human-directed selective pressures in domesticated species. Dog domestication is linked to several genomic variants; however, the functional impacts of these variants may not always be straightforward when found in non-coding regions of the genome. Four polymorphic transposable elements (TE) found within non-coding sites along a 5 Mb region on canine CFA6 have evolved due to directional selection associated with heightened human-directed hyper-sociability in domesticated dogs. We found that the polymorphic TE in intron 17 of the canine *GTF2I* gene, which was previously reported to be negatively correlated with canid human-directed hyper-sociability, is associated with altered chromatin looping and hence distinct *cis*-regulatory landscapes. We reported supporting evidence of an *E2F1*-DNA binding peak concordant with the altered loop and higher expression of *GTF2I* exon 18, indicative of alternative splicing. Globally, we discovered differences in pathways regulating the extra-cellular matrix with respect to TE copy number. Overall, we reported evidence suggesting an intriguing molecular convergence between the emergence of hypersocial behaviors in dogs and the same genes that, when hemizygous, produce human Williams Beuren Syndrome characterized by cranio-facial defects and heightened social behaviors. Our results additionally emphasize the often-overlooked potential role of chromatin architecture in social evolution.

Keywords Chromatin, Hyper-sociability, Domestication, Transposable elements

*Correspondence:
Dhriti Tandon
dtandon@princeton.edu
Bridgett M. vonHoldt
vonholdt@princeton.edu
[1] Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA
[2] Department of Ethology, ELTE Eötvös Loránd University, Budapest, Hungary
[3] MTA-ELTE Lendület "Momentum" Companion Animal Research Group, Budapest, Hungary
[4] ELTE NAP Canine Brain Research Group, Budapest, Hungary

Background

Social behavior is a quantitative complex trait defined through numerous biological, life history, developmental, and cognitive profiles. While candidate loci methods have identified genes encoding hormones, sensory receptors, or neurotransmitters that shape behaviors [1, 2], pleiotropy and polygenic infrastructure reduce the clarity of these primary structural findings [3]. The addition of multi-*omic* data to candidate loci findings has now uncovered large-effect variants that act as master regulatory loci shaping behavior, by rapidly innovating pleiotropic effects on transcription, chromatin conformations, and the underlying biological pathways [4–6]. More recently, genomic architecture has been implicated in the

Tandon *et al. BMC Genomics*   (2024) 25:767

Page 2 of 9

evolution of social behavior due to mechanisms of tightly linked supergene complexes [7, 8], copy number of structural variants (SVs) [9, 10] and, to a lesser extent, transposable elements (TEs) [11, 12]. TEs have the propensity to alter cis- regulation via coordinated genetic and epigenetic mechanisms. Such elements are targeted by the genome surveillance machinery for TE inactivation through various mechanisms, including DNA methylation and histone modifications [13, 14], with the epigenetically silenced TEs exerting cis-regulatory impacts on proximal (i.e., within 100 nucleotides) transcriptionally active sequences [15, 16]. TEs also harbor binding sites for transcription factors and regulatory proteins that impact higher order three-dimensional (3D) chromatin structure [17], making them strong candidates for drivers of phenotypic change and even social evolution.

Species domestication results from the rapid and significant phenotypic evolution in response to novel human-directed selective pressures, exemplified in the Farm-fox experiment where behavioral and morphologic evolution was documented [18]. Dogs were the first domesticated species [19], where strong artificial selection maintains strict impermeable boundaries between dog breeds and has removed a significant amount of intra-breed genomic and phenotypic variation, enhancing the efficiency of mapping and evolutionary studies [20, 21]. Among the several well-known morphological and behavioral changes that occurred during the domestication of gray wolves to dogs [22–24], human-directed sociability is one of the prominent phenotypes suspected to have been selected during domestication [25].

Dogs exhibit a magnified social interest towards humans, referred to as hyper-sociability [26, 27]. Through strong directed selection, this social phenotype has provided feedback to the underlying molecular infrastructure, evidenced by moderate estimates of heritability (h^2, social interactions towards humans = 0.23) with notable differences across breeds (h^2, willingness to make contact with humans: German shepherds = 0.38, Rottweilers = 0.03) [28]. Part of the supportive molecular infrastructure includes four previously mapped large-effect polymorphic retro-transposable elemental (TE) sequences, found on canine chromosome CFA6 with copy number of the derived allele positively correlated with human-directed hyper-sociability [27]. The derived alleles are located within or proximal to three genes: two TE insertions in intron 1 of Williams-Beuren Syndrome Chromosome Region 17 (*WBSCR17*), the lack of an insertion in intron 17 of the General Transcription Factor 2-I (*GTF2I*), and one insertion in intron 5 of *POM121* Transmembrane Nucleoporin (*POM121*) [27].

Here, we quantify the functional impacts of the TE sequence polymorphism, specifically the presence of a tRNA-based *S*hort-*I*nterspersed *N*ucleotide *E*lement in intron 17 of canine *GTF2I* [27] that encodes a multifunctional and ubiquitously expressed transcription factor 2-I (TFII-I) protein [6]. Within dogs, vonHoldt et al. (2017) [27] reported that the allele containing the TE insertion in gene *GTF2I* was the minor allele (*f* = 0.31 or 5/16 dogs). Concordant with the neural crest cell hypothesis [29], *GTF2I* is a plausible master regulator for domestication-driven morphological and behavioral phenotypes due to its involvement in neural tube closure and neural crest cell migrations [30]. Throughout embryogenesis, *GTF2I* is necessary for neuronal development [30] and is later central in regulating synaptic plasticity and biochemical pathways for anxiety and sociability [6, 31]. Perhaps more striking is that artificially constructed hemizygous knockouts of *GTF2I* are associated with heightened sociability in murine systems [32] and a large-scale naturally occurring hemizygous deletion on human chromosome 7, which includes *GTF2I*, causes Williams Beuren Syndrome (WS) [33], a neurodevelopmental disorder that is characterized by cranio-facial defects and extremely high levels of social behaviors [34, 35].

Although the TE found within *GTF2I*'s intron is strongly associated with the evolution of hyper-sociability in domestic dogs, the proximate mechanisms are currently undescribed. This is especially compounded by the fact that the high-effect TE is located within a non-coding region, 700 nucleotides from the nearest exon, and is relatively small (only 187 bp) as compared to most high-impact structural variants located in intronic and intergenic regions [27]. With past evidence for associative altered *cis-* regulation in canine blood [36], we hypothesize that the copy number of this intronic TE shifts the 3D chromatin state and the *cis-* regulatory landscape, which further impacts downstream transcriptional activities. We investigated chromatin architecture and gene expression differences associated with *GTF2I*'s intronic TE in dog brainstems, a tissue relevant to social behaviors as it contains neuronal projections for glutamatergic, serotonergic and dopaminergic neurons [37]. We found that the polymorphic TE located in intron 17 of *GTF2I* is associated with altered chromatin looping with its own intron 1 and putative alternative splicing of this gene. At a more global level, we find differences in pathways related to the extra-cellular matrix associated with the ancestral and derived forms of this gene.

Results

Ancestral TE insertion associated with higher expression of GTF2I exon 18

We collected data from brainstem tissue (pons) of six male dogs (12–16 years old) from three genomic dimensions: targeted chromatin conformation capture

sequencing (*Capture C*) at the polymorphic TE site, *E2F1* and *H3K27ac* chromatin immunoprecipitation sequencing (*ChIP-seq*), and RNA sequencing (*RNA-Seq*). Three of the six dogs were heterozygous for the ancestral allele, which contains the TE insertion, while the other three dogs were homozygous for the derived allele that lacked a TE insertion [27] (Table S1). To reduce the likelihood of false positives, we only considered 3D contacts with a concordant differential *ChIP* peak signal as biologically meaningful. We found that the TE sequence itself was enriched for binding motifs for the transcription factor *E2F1* and its co-factors (see Supplementary Text). In humans, *E2F1* has binding sites at genes within 1 Mb upstream and downstream (i.e. in *cis-* conformations) of *GTF2I* (see Supplementary Text). Samples that carried at least one copy of the ancestral allele exhibited unique 3D contacts, which we quantified from the contact frequency at each site relative to the surrounding background [38], with three regions on CFA6: intron 1 of *GTF2IRD2* at 5.65–5.67 Mb,

intron 1 of *GTF2I* at 5.82–5.84 Mb, and upstream of *GTF2I* spanning 5.85–5.87 Mb (log $q < $-10) (Fig. 1A; Table S2). When samples carried the derived allele lacking the TE, we discovered unique contacts with a novel CFA6 bin at 5.73–5.75 Mb (log $q < $-10) (Fig. 1B; Table S2). We conducted a negative binomial Wald Test in the R package *DiffBind*, to investigate differences in *ChIP* peak enrichment between samples as a function of mean signal quantified by normalized read counts in each peak bin [39]. Hence, *ChIP* peaks with significant differences, hereafter referred to as "differentially enriched peaks", are those whose signal strength differs more between samples with different genotypes than among samples with the same genotype [39]. We found significant differences in the *ChIP* peak signal for *E2F1* at CFA6:5,822,073–5,822,473 within intron 1 of *GTF2I* (log$_2$ Fold Change $_{[Derived/Ancestral]} = $-3.12; $p = 6.73 \times 10^{-4}$; $p_{adj} = 0.03$), which is located within the TE containing contact site at 5.82–5.84 Mb (Figs. 1A,C). We analyzed exon expression of the six tissues which were controlled
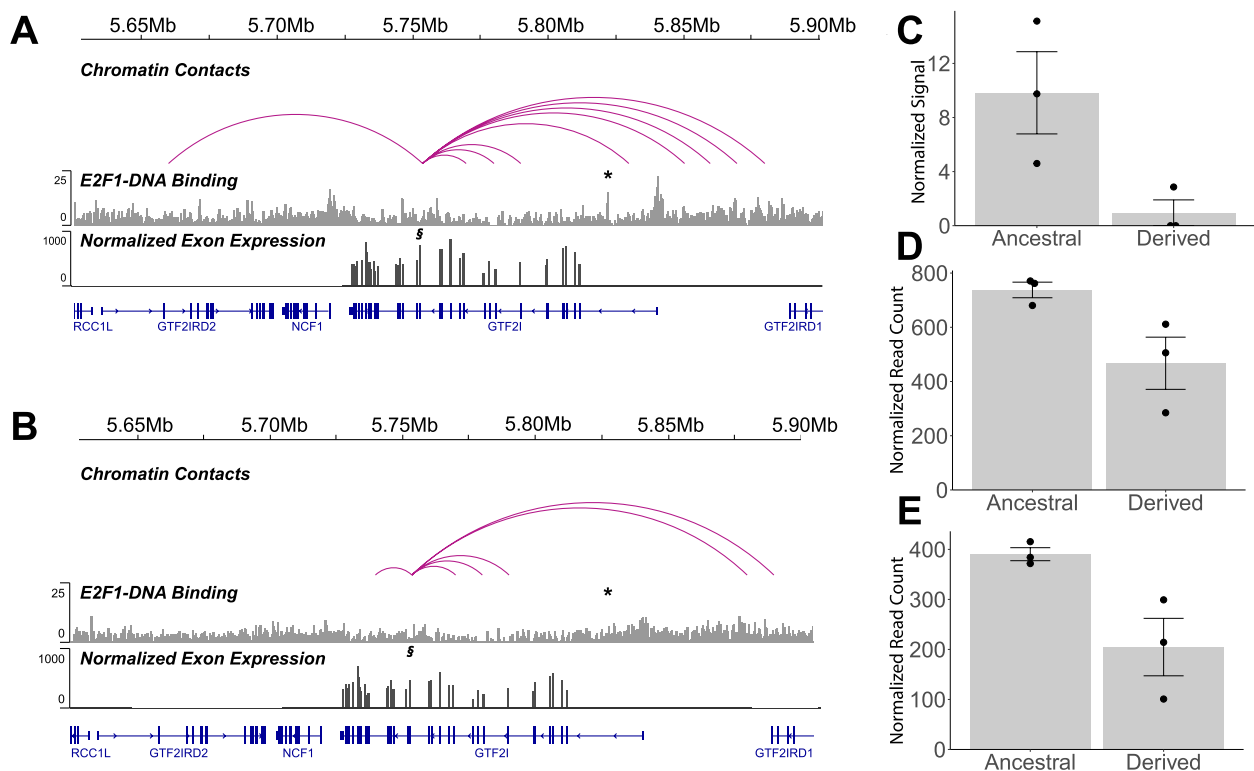

Fig. 1 Differences in *cis-* regulation of *GTF2I* exon 18, between ancestral TE present and derived TE absent states of *GTF2I*. Visualization of the target region with chromosomal coordinates in Mb (top line) along canine chromosome CFA6 for *Capture C* contacts (top), average *E2F1 ChIP-Seq* Coverage (middle) with differential peak (*), and average normalized exon expression with exon 18 (§) values at *GTF2I* (bottom) for samples **A** heterozygous for the ancestral TE insertion in *GTF2I* and **B** homozygous for the derived allele lacking the TE insertion in *GTF2I*. For panels **C-E**, the ancestral state refers to the TE insertion in *GTF2I*, while the derived allele is the lack of the TE insertion. **C** Normalized *ChIP-Seq* signal for *E2F1* located at intron 1 of *GTF2I* (chr6:5,822,073–5,822,473 bp, $p_{adj} = 0.03$). **D** Expression of *GTF2I* exon 18 ($p_{adj} = 0.138$) as reported in the IGV track. **E** *GTF2I* exon-17—exon18 junction expression ($p_{adj} = 0.141$). Black circles show each data point, bar heights correspond to group means, and error bars correspond to standard errors

for age and sex, to account for any differences that are developmental-stage dependent expression patterns for *GTF2I* isoforms [40, 41] (see Supplementary Text). We used an R package *JunctionSeq* to conduct a likelihood ratio test to determine the effects of TE genotypic state on exon or splice junction expression [42]. We quantified exon expression using reads mapped within the exonic bins and splice junction expression using reads mapped across two exons [43], such that these reads visually mimic a discordant alignment similar to a deletion representing the spliced-out intron. We found higher usage of exon 18 (log2 Fold Change$_{[Derived/Ancestral]}$ = -0.528, $p = 0.001$, $p_{adj} = 0.138$) and evidence for a splice junction between exon 17 and 18 (log2 Fold Change$_{[Derived/Ancestral]}$ = -0.783, $p = 0.001$, $p_{adj} = 0.141$) as a function of the TE insertion state at *GTF2I* (Figs. 1D,E; Table S3). There were no changes in expression levels of *GTF2I* itself with respect to the TE insertion ($p > 0.1$). We found weak evidence for an additional loop between the polymorphic TE site and the gene *LIMK1* (Fig. S1; see Supplementary Text).

Differences in the enrichment of biological pathways related to extra-cellular matrix

Since *GTF2I* encodes a transcription factor regulating the expression of several genes [6], we evaluated the potential downstream impacts on global biological pathways as a function of TE genotypic state at *GTF2I*. First, we looked for differentially expressed genes across genome and found 10 genes differentially expressed as a function of the TE copy number at *GTF2I* (Fig. 2C; Table 1). Next, we constructed gene modules, which are sets of co-expressed genes that broadly represent biological pathways [44]. Since the construction of gene modules requires at least 12 samples [44], we collected *RNAseq* data from the brainstem tissue of 16 additional dogs across ages, sexes and breeds (Table S1) and conducted these analyses on all 22 samples with *RNAseq* data, while accounting for confounding variables (*Supplementary Information: Materials and Methods*). We generated signed gene networks and quantified module eigengene (ME) values, to represent the dimensionality reduced (i.e. the first principal component) gene expression values from all genes in the module, which are broadly associated with changes
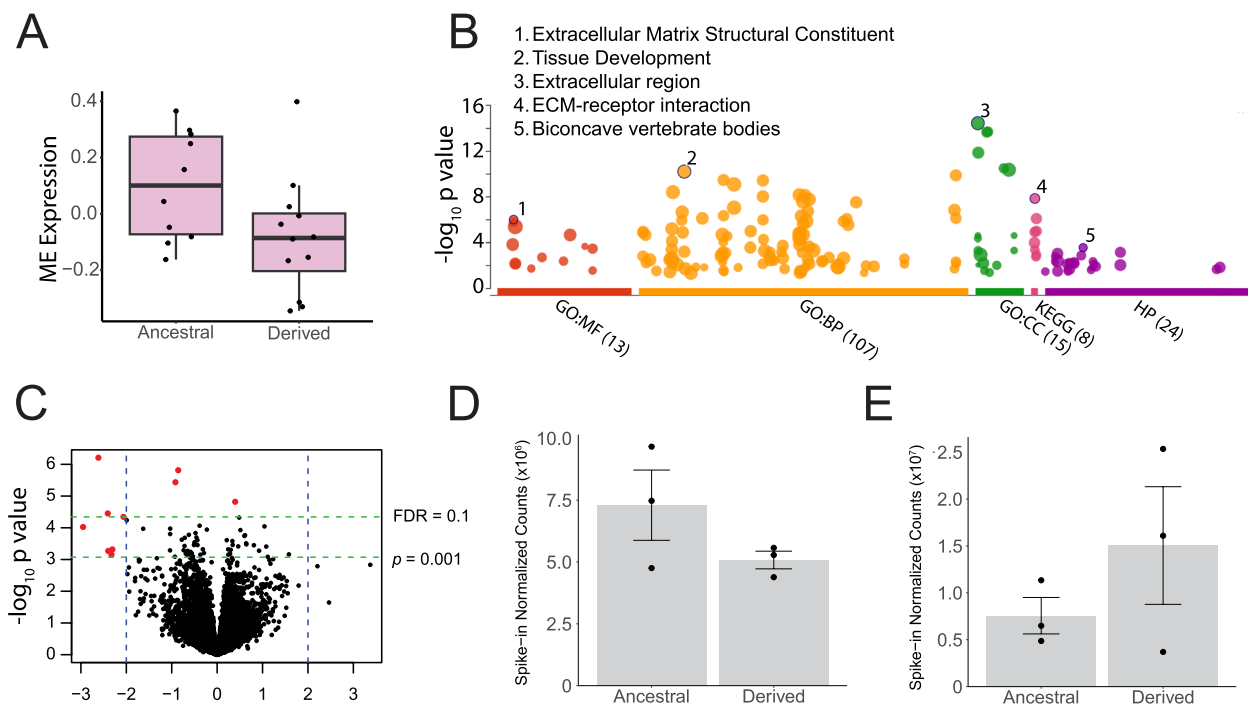

Fig. 2 Downstream molecular impacts on global gene expression and regulation. **A** Box-and-whisker plot of module eigengene (ME) expression for the differentially expressed module. ME expression values have significantly different means ($t = 2.1348$, $df = 19.744$, $p = 0.0455$) between samples containing the ancestral and derived TE genotypes at GTF2I. Whiskers represent the data range excluding outliers. Horizontal edges correspond to 25th, 50th and 75th percentiles. **B** Gene ontology (GO) hits for genes in the differentially expressed module displaying their -log$_{10}p$ for each GO categories: Molecular Function (MF), Biological Processes (BP), Cellular Component (CC), Kyoto Encyclopedia of Genes and Genomes pathways (KEGG) and Human Phenotype (HP). Top GO term in each category is numbered and keyed. **C** Volcano plot of differentially expressed genes (red points; FDR < 0.1, or $p < 0.001$ and log$_2$ fold change >|2|), for the 22 samples. **D** Spiked-in normalized read counts for *E2F1* enrichment and; **E** *H3K27ac* marks. Bar heights represent mean values, error bars correspond to standard errors, and black circles depict replicate values

Table 1 Significantly differentially expressed genes with their respective false discovery rate (FDR < 0.1), log2 fold change ratios (log2FC > 2), and significance *p* values (*p* < 0.001)

Gene Symbol	Log2FC	*p* value	FDR
SLC47A1	-2.615	6.19×10^{-7}	0.010
SLC14A1	-0.857	1.53×10^{-6}	0.012
BNC2	-0.919	3.67×10^{-6}	0.019
PPIP5K1	0.397	1.51×10^{-5}	0.059
ERMAP	-2.064	4.53×10^{-5}	0.107
ADAMTS15	-2.409	3.53×10^{-5}	0.107
SLC6A20	-2.953	9.43×10^{-5}	0.123
WNT16	-2.399	5.38×10^{-4}	0.183
CCN3	-2.307	4.80×10^{-4}	0.183
SLC22A6	-2.327	6.99×10^{-4}	0.193

in biological pathways [44]. Since signed modules contain co-expressed genes with the same directionality of expression change (i.e., all with increased or decreased expression for a given condition) [44], these were preferred over the generation of unsigned modules to ease the interpretation of results and identify biological pathways with a general increased or decreased enrichment in the derived state of *GTF2I*. We found a single gene module with higher expression for samples carrying the ancestral *GTF2I* TE insertion (Pearson Correlation coeff $_{[Derived]}$ = -0.43; *p* = 0.047; p_{adj} = 0.63); however, this did not survive multiple comparison corrections (Figs. 3A, S2A-D; Table S4). Since this was the only gene module meeting the significance threshold of *p* < 0.05, we hereafter refer to this module as the "differentially expressed module". We then conducted a Gene Ontology Enrichment analysis on g:*Profiler* [45] using IDs of genes within the differentially expressed gene module. The top ranked gene ontology term is for the cellular component "extra-cellular region" (p_{adj} = 8.54×10^{-14}), and the top KEGG pathway [46] is for "Extra-Cellular Matrix (ECM) receptor interaction" (p_{adj} = 3.81×10^{-8}) (Fig. 2B). We further found *BNC2* among the top ranked transcription factors candidates that putatively regulates genes in the differentially expressed gene module (Mean Rank = 10.33; Percentage overlap *p* = 28.6%; *Enrichr* [47–49] FDR = 5.68×10^{-18}; *GTEx* [50] FDR = 2.51×10^{-5}; *ARC HS4* [51] FDR = 8.70×10^{-17}). *BNC2* was also found to be a differentially expressed gene (Tables 1, S5).

Given *GTF2I*'s multi-functional role in transcriptional and translational regulation [6, 52, 53], we additionally identified binding motifs at differentially enriched *E2F1* and *H3K27ac* peaks (as identified by *DiffBind analyses*) between samples with differential *GTF2I* genotypes. We conducted the motif enrichment analysis using the web-based tool, *CentriMo* [54], against the JASPAR2022

CORE vertebrates non-redundant v2 database [55]. We did not find enriched binding motifs among regions harboring differential histone acetylation marks. Differentially enriched *E2F1* peaks with higher signal in samples with the derived allele (log2FC > 2; FDR < 0.1) were significantly enriched for binding motifs of *ZNF740* (e = 6.6×10^{-6} to 9.1×10^{-3}; percent matching = 70.0%). However, differentially enriched *E2F1* peaks with higher signal in samples with the ancestral allele (log2FC > 2; FDR < 0.1) did not present any significantly enriched motifs. We found no changes in overall *H3K27ac* and *E2F1* enrichment (*p* > 0.1) (Figs. 2D, E).

Discussion

Linking Chromatin Architecture to Behavioral Evolution

Functional changes driven by intronic TEs have been well studied, although typically, such TEs are large in size and physically proximal to the nearest exon (i.e., within 100 nucleotides) or splice sites [56–58]. Our work confirms that the polymorphic 187 bp intronic TE alters *cis*-regulatory landscapes within CFA6:5.7–6.3 Mb, by impacting the chromatin structure itself. The TE promotes altered looping with intron 1 of *GTF2I*, which are further associated with differences in gene regulation. To our knowledge, the study provides the first evidence for a gene loop that is associated with social evolution as a consequence of animal domestication.

Our proposed model for the altered loop state is that the TE offers a binding site for an *E2F1* co-factor with subsequent recruitment of the *E2F1* transcription factor itself, which binds to an intron 1 site of *GTF2I*. This would facilitate looping between the TE insertion and chr6:5.82–5.84 Mb. The loop and its concordant *E2F1* peak are lost when *GTF2I* does not contain the TE in the derived state, which could be a consequence of binding site loss for the *E2F1* co-factor and the lack of *E2F1* in *GTF2I* intron 1. Given the proposed model, we would anticipate a peak in intron 17 of GTF2I. However, we could not survey DNA–protein binding activity at the TE itself due to its repetitive nature. We instead relied on chromatin conformation evidence that emerged from the polymorphic TE and a concordant *E2F1* peak at the putative contact site. In addition, our *in-silico* discovery also suggested that a co-factor of *E2F1*, such as *Sp1* (see Supplementary Text), rather than *E2F1* itself, could bind to the polymorphic TE. Hence, a lack of enrichment of this region could also be explained by immunoprecipitation targeting *E2F1* and compromised protein–protein interactions during the *ChIP* prep. Strikingly, we find no changes in expression levels of *GTF2I* itself; rather, we find an associated change in a single exon of *GTF2I* after controlling for confounding variables. While the differential exon usage analysis marginally misses the FDR < 10%

Tandon *et al. BMC Genomics*   (2024) 25:767

Page 6 of 9

threshold, multiple comparison corrections were performed with respect to all exons in the genome, including those of non-coding RNAs. Given this large sample size of exons, we still indicate that these results have the potential to be biologically meaningful and worthy of follow-ups for isoform discovery through long-read sequencing methods such as Iso-Seq. While we have not identified a causal mechanism linking the *E2F1* loop and splicing, *E2F1* is known to interact with splice-impacting co-factors such as the p100/TSN complex and regulates the splicing patterns of its target genes [59]. An alternative mechanism is that the altered looped state itself, rather than the molecular functions of *E2F1*, contributes to splicing that is facilitated by alternate promoter adoption whereby different promoters could express different isoforms of the same gene [60, 61].

Potential Convergence in Gene Regulatory Mechanisms Underlying Hypersocial Behavior

GTF2I exon 18 in humans encodes the R3 domain of the TFII-I protein [52, 62], which consists of a DNA binding leucine zipper domain followed by six loop-helix-loop repeat domains (R1-R6). The TFII-I loop-helix-loop domains are involved in protein–protein interactions [53, 62] and changes to the R3 domain could hence alter protein interactions with TFII-I. Although we did not directly investigate the changes in protein interactions with TFII-I to establish causality, our correlative evidence suggests the convergent outcome of reduced expression of *BNC2* and reduced expression of a gene module that includes *BNC2* target genes in samples that lack the TE insertion in *GTF2I*. Among the transcription factors that target *BNC2* include *USF1* and *MYC*, which have incidentally been identified as protein interactors of TFII-I [53, 62, 63]. Altered *GTF2I* exon 18 expression could drive changes in TFII-I proteomic interactions with *USF1* and *MYC*, thereby impacting molecular processes affecting canine hypersocial behavior. TFII-I also interacts with the histone deacetylase 3 protein (HDAC3) [64]. This protein facilities the removal of histone acetylation marks, thereby acting as a gene silencer in most cellular contexts [65]. While our findings suggest that samples with the ancestral allele of the TE insertion carry lower levels of global *H3K27ac*, this difference was not significant. This could be due to a variety of reasons that include an underpowered design for the specific analyses type, combinatorial confounding impacts of other histone deacetylases and nuanced proteomic impacts on TFII-I such that interactions with some proteins, and not others, are affected. Hence, future efforts to determine interactomes of TFII-I in samples with ancestral and derived forms of *GTF2I* could help identify downstream biological impacts of the TE locus. However, these assays

are currently technically challenging due to a limited availability of fresh tissues or cell lines from dogs.

In addition to higher sociability, patients diagnosed with Williams Beuren Syndrome also have cranio-facial abnormalities, explained by extra-cellular matrix anomalies [34]. Our results pertaining to gene module differences suggest that biological pathways related to the extra-cellular matrix show reduced expression. We also see functional changes related to the gene elastin (*ELN*), which is included in the differentially expressed gene module. Patients with WS can have varying lengths of deletions at the 7q11.23 locus, and 90% of patients with WS have a hemizygous deletion of *ELN* [66]. Extra-cellular matrix anomalies may not directly explain neurocognitive profiles and hence social behaviors. However, our results suggest that the derived *GTF2I* allele recapitulates some molecular characteristics of WS in domestic dogs: altered expression of extra-cellular matrix-related pathways and variants impacting *GTF2I* function.

Conclusion

While often in an "arms-race" with the host and hence usually silenced, few TEs can be co-opted into regulatory sequences that promote specific transcriptional modules. Regulatory DNA sequences usually undergo purifying selection when the selective environment is stable [67]. However, new selective pressures that favor novel phenotypes will drive directional selection on regulatory loci. Such is the case for the canine intronic TE at *GTF2I*, which shows distinctive allele frequency differences between the ancestral gray wolf and dog genomes [27], with the wolf genome 'co-opting' the TE and the dog genome 'purging' it. While showing clear signatures of selection, functional impacts of TEs outside of coding regions are currently understudied across evolutionary scenarios, with most investigations limited to first-order genomic structure, overall TE distribution across the genome, and their putative relationship to candidate loci [12]. While more work is warranted to prove causality, we emphasize the relevance of high-effect non-coding variants and their role in regulating complex phenotypes in non-model systems by reporting their potential and indirect impacts on gene regulation through altered chromatin states. While our study does not prove this, we provide an evolutionary interpretation where the non-looped state of *GTF2I* could possibly provide a dog-specific fitness advantage from increased human-directed sociability. Molecular changes associated with an altered regulatory landscape can also introduce fitness costs [68] due to a loss in existing transcriptional machinery that could impact the extra-cellular matrix and activity of the multi-functional *GTF2I* gene, which could be validated by future efforts. We hope to provide a new framework

Tandon *et al. BMC Genomics*   (2024) 25:767

Page 7 of 9

for exploring the molecular infrastructure of social evolution by viewing the genome as more than a culmination of base-pairs and combining higher-order information on 3D genome architecture.

Limitations

Due to a lack of ability to do functional experiments and CRISPR-based gene editing in dogs, the polymorphic TE locus within *GTF2I* is only correlatively linked to hyper-social behaviors. In addition, the bulk of our findings that pertain to *cis*-regulation have been conducted in male dogs (12–16 years old). This represents technical difficulties with sample acquisition from dog brains as most euthanized younger dogs have serious health complications and hence may present results confounded by these complications. Upon limiting samples to those with no obvious brain-related conditions, we were unable to obtain a sample size powerful enough to reliably conduct this study on only early-life or female dogs. Therefore, we caution readers that some specificities of gene regulation may be age or sex specific. Owing to low cross-correlation scores of our E2F1 *ChIP* data, some E2F1-DNA binding sites may be missed. We also caution that results pertaining to differential gene module analysis did not survive FDR corrections, owing to the presence of a single gene module passing a significance threshold of $p < 0.05$. Nonetheless, we present substantial evidence supporting the propensity of this TE to impact *cis*-regulation through an altered chromatin state.

Methods

We obtained pons brainstem tissues preserved in *RNALater* (Thermo Fisher Scientific, Waltham, MA, USA), for 22 dogs collected by the Canine Brain and Tissue Bank (CBTB) at Eötvös Loránd University, Budapest, Hungary. Each sample was associated with metadata that included age, sex, and breed information (pure or mixed) (Table S1). Six of the 22 dog samples had an additional paired specimen available, which had been flash frozen at the time of collection. For these flash-frozen samples, we carried out *Capture C,* targeted at the polymorphic *GTF2I* TE site (CanFam 3.1; CFA6:5,753,797–5,753,983), and *ChIP-Seq* to quantify E2F1-DNA binding peaks and H3K27ac regions at chromatin loop bases. We collected RNA-seq data to further investigate differences in local gene regulation associated with the chromatin loops for the six samples. We conducted a higher-powered analysis of global differences in biological pathways among all 22 dog brainstem samples. Detailed methods, protocols and software used can be found in *Supplementary Information: Materials and Methods.*

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10614-6.

> Supplementary Material 1.

Declarations

Ethics approval and consent to participate
All samples were collected as per the ethical code established by Pest Vármegyei Kormányhivatal (Permit Number: PE/EA/00010–2/2023).

Consent for Publication
Not Applicable.

Competing Interests
The authors declare no competing interests.

References
1. Popova NK. From genes to aggressive behavior: the role of serotonergic system. BioEssays. 2006;28:495–503. https://doi.org/10.1002/bies.20412.
2. York RA. Assessing the Genetic Landscape of Animal Behavior. Genetics. 2018;209:223–32. https://doi.org/10.1534/genetics.118.300712.
3. Wittkopp PJ. Variable gene expression in eukaryotes: a network perspective. J Exp Biol. 2007;210:1567–75. https://doi.org/10.1242/jeb.002592.
4. Bendesky A, Kwon YM, Lassance JM, Lewarch CL, Yao S, Peterson BK, et al. The genetic basis of parental care evolution in monogamous mice. Nature. 2017;544:434–9. https://doi.org/10.1038/nature22074.
5. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, et al. Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell. 2017;171:557-572.e24. https://doi.org/10.1016/j.cell.2017.09.043.
6. Roy AL. Pathophysiology of TFII-I: Old Guard Wearing New Hats. Trends Mol Med. 2017;23:501–11. https://doi.org/10.1016/j.molmed.2017.04.002.
7. Küpper C, Stocks M, Risse JE, dos Remedios N, Farrell LL, McRae SB, et al. A supergene determines highly divergent male reproductive morphs in the ruff. Nat Genet. 2016;48:79–83. https://doi.org/10.1038/ng.3443.

Tandon *et al. BMC Genomics*   (2024) 25:767

Page 8 of 9

8. Lamichhaney S, Fan G, Widemo F, Gunnarsson U, Thalmann DS, Hoeppner MP, et al. Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). Nat Genet. 2016;4:84–8. https://doi.org/10.1038/ng.3430.

9. Donaldson ZR, Young LJ. The Relative Contribution of Proximal 5′ Flanking Sequence and Microsatellite Variation on Brain Vasopressin 1a Receptor (Avpr1a) Gene Expression and Behavior. Nachman MW, editor. PLoS Genet. 2013;9(8):e1003729. https://doi.org/10.1371/journal.pgen.1003729

10. Zhao Y, Long L, Wan J, Biliya S, Brady SC, Lee D, et al. A spontaneous complex structural variant in rcan-1 increases exploratory behavior and laboratory fitness of Caenorhabditis elegans. PLoS Genet. 2020;16(2): e1008606. https://doi.org/10.1371/journal.pgen.1008606.

11. McKenzie SK, Kronauer DJC. The genomic architecture and molecular evolution of ant odorant receptors. Genome Res. 2018;28:1757–65. https://doi.org/10.1101/gr.237123.118.

12. Rubenstein DR, Ågren JA, Carbone L, Elde NC, Hoekstra HE, Kapheim KM, et al. Coevolution of Genome Architecture and Social Behavior. Trends Ecol Evol. 2019;34:844–85. https://doi.org/10.1016/j.tree.2019.04.011.

13. Hollister JD, Gaut BS. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. 2009;19:1419–28. https://doi.org/10.1101/gr.091678.109.

14. Almeida MV, Vernaz G, Putman AL, Miska EA. Taming transposable elements in vertebrates: from epigenetic silencing to domestication. Trends Genet. 2022;38:529–53. https://doi.org/10.1016/j.tig.2022.02.009.

15. Dolinoy DC, Huang D, Jirtle RL. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc Natl Acad Sci USA. 2007;104:13056–61. https://doi.org/10.1073/pnas.0703739104.

16. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13:484–92. https://doi.org/10.1038/nrg3230.

17. Sundaram V, Wysocka J. Transposable elements as a potent source of diverse cis -regulatory sequences in mammalian genomes. Phil Trans R Soc B. 2020;375:20190347. https://doi.org/10.1098/rstb.2019.0347.

18. Trut LN. Early Canid Domestication: The Farm-Fox Experiment: Foxes bred for tamability in a 40-year experiment exhibit remarkable transformations that suggest an interplay between behavioral genetics and development. Am. Sci. 1999;87:160–169. https://www.jstor.org/stable/27857815

19. Larson G, Karlsson EK, Perri A, Webster MT, Ho SYW, Peters J, et al. Rethinking dog domestication by integrating genetics, archeology, and biogeography. Proc Natl Acad Sci USA. 2012;109:8878–83.: https://doi.org/10.1073/pnas.1203005109

20. vonHoldt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, et al. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature. 2010;464:898–902. https://doi.org/10.1038/nature08837.

21. Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, et al. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. Nat Commun;2019,10:1489. https://doi.org/10.1038/s41467-019-09373-w

22. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, et al. A Single IGF1 Allele Is a Major Determinant of Small Size in Dogs. Science. 2007;316:112–5. https://doi.org/10.1126/science.1137045.

23. Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, Parker HG, et al. Coat Variation in the Domestic Dog Is Governed by Variants in Three Genes. Science. 2009;326:150–3. https://doi.org/10.1126/science.1177808.

24. Parker HG, VonHoldt BM, Quignon P, Margulies EH, Shao S, Mosher DS, et al. An Expressed Fgf4 Retrogene Is Associated with Breed-Defining Chondrodysplasia in Domestic Dogs. Science. 2009;325:995–8. https://doi.org/10.1126/science.1173275.

25. Topál J, Gergely G, Erdőhegyi Á, Csibra G, Miklósi Á. Differential Sensitivity to Human Communication in Dogs, Wolves, and Human Infants. Science. 2009;325:1269–72. https://doi.org/10.1126/science.1176960.

26. Kubinyi, E. Comparative Social Cognition: From wolf and dog to humans. Comp. Cogn. Behav. Rev. 2006;2. https://doi.org/10.3819/ccbr.2008.20002.

27. vonHoldt BM, Shuldiner E, Koch IJ, Kartzinel RY, Hogan A, Brubaker L, et al. Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. Sci Adv. 2017;3: e1700398. https://doi.org/10.1126/sciadv.1700398.

28. Persson ME, Roth LSV, Johnsson M, Wright D, Jensen P. Human-directed social behaviour in dogs shows significant heritability. Genes Brain Behav. 2015;14:337–44. https://doi.org/10.1111/gbb.12194.

29. Wilkins AS, Wrangham RW, Fitch WT. The "Domestication Syndrome" in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics. Genetics. 2014;197:795–808. https://doi.org/10.1534/genetics.114.165423.

30. Enkhmandakh B, Makeyev AV, Erdenechimeg L, Ruddle FH, Chimge NO, Tussie-Luna MI, et al. Essential functions of the Williams-Beuren syndrome-associated TFII-I genes in embryonic development. Proc Natl Acad Sci USA. 2009;106:181–6. https://doi.org/10.1073/pnas.0811531106.

31. Borralleras C, Sahun I, Pérez-Jurado LA, Campuzano V. Intracisternal Gtf2i Gene Therapy Ameliorates Deficits in Cognition and Synaptic Plasticity of a Mouse Model of Williams-Beuren Syndrome. Mol Ther. 2015;23:1691–9. https://doi.org/10.1038/mt.2015.

32. Martin LA, Iceberg E, Allaf G. Consistent hypersocial behavior in mice carrying a deletion of Gtf2i but no evidence of hyposocial behavior with Gtf2i duplication: Implications for Williams-Beuren syndrome and autism spectrum disorder. Brain Behav. 2018;8: e00895. https://doi.org/10.1002/brb3.895.

33. Merla G, Brunetti-Pierri N, Micale L, Fusco C. Copy number variants at Williams–Beuren syndrome 7q11.23 region. Hum Genet. 2010;128:3–26. https://doi.org/10.1007/s00439-010-0827-2

34. Pober BR. Williams-Beuren Syndrome. N Engl J Med. 2010;362:239–52. https://doi.org/10.1056/NEJMra0903074.

35. Järvinen A, Korenberg JR, Bellugi U. The social phenotype of Williams syndrome. Curr Opin Neurobiol. 2013;23:414–22. https://doi.org/10.1016/j.conb.2012.12.006.

36. vonHoldt BM, Ji SS, Aardema ML, Stahler DR, Udell MAR, Sinsheimer JS. Activity of Genes with Functions in Human Williams-Beuren Syndrome Is Impacted by Mobile Element Insertions in the Gray Wolf Genome. Genome Biol Evol. 2018;10:1546–53. https://doi.org/10.1093/gbe/evy112.

37. Franklin TB, Silva BA, Perova Z, Marrone L, Masferrer ME, Zhan Y, et al. Prefrontal cortical control of a brainstem social behavior circuit. Nat Neurosci. 2017;20:260–70. https://doi.org/10.1038/nn.4470.

38. Cairns J, Freire-Pritchett P, Wingett SW, Várnai C, Dimond A, Plagnol V, et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. Genome Biol. 2016;17:127. https://doi.org/10.1186/s13059-016-0992-2.

39. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012;481:389–93. https://doi.org/10.1038/nature10730.

40. Makeyev AV, Bayarsaihan D. Alternative splicing and promoter use in TFII-I genes. Gene. 2009;433:16–25. https://doi.org/10.1016/j.gene.2008.11.027.

41. The UniProt Consortium, Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, et al. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51:D523–31.

42. Hartley SW, Mullikin JC. Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. Nucleic Acids Res. 2016;gkw501. https://doi.org/10.1093/nar/gkw501

43. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21. https://doi.org/10.1093/bioinformatics/bts635.

44. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9:559. https://doi.org/10.1186/1471-2105-9-559.

45. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 2019;47:W191–8. https://doi.org/10.1093/nar/gkz369.

46. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:27–30. https://doi.org/10.1093/nar/28.1.27.

47. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinform. 2013;128. https://doi.org/10.1186/1471-2105-14-128

48. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene

Tandon *et al. BMC Genomics*   (2024) 25:767

Page 9 of 9

set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44:W90-97. https://doi.org/10.1093/nar/gkw377.

49. Xie Z, Bailey A, Kuleshov MV, Clarke DJB., Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, Jeon M, Ma'ayan A. Gene set knowledge discovery with Enrichr. Curr Protoc. 2021;e90. https://doi.org/10.1002/cpz1.90

50. Consortium G. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369:1318–30. https://doi.org/10.1126/science.aaz1776.

51. Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, Silverstein MC, Ma'ayan A. Massive mining of publicly available RNA-seq data from human and mouse. Nat Commun. 2018;9. https://doi.org/10.1038/s41467-018-03751-6

52. Roy AL. Biochemistry and biology of the inducible multifunctional transcription factor TFII-I: 10years later. Gene. 2012;492:32–41. https://doi.org/10.1016/j.gene.2011.10.030.

53. Roy AL, Meisterernst M, Pognonec P, Roeder RG. Cooperative interaction of an initiator-binding transcription initiation factor and the helix–loop–helix activator USF. Nature. 1991;354:245–8. https://doi.org/10.1038/354245a0.

54. Bailey TL, Machanick P. Inferring direct DNA binding from *ChIP*-seq. Nucleic Acids Res. 2012;40:e128–e128. https://doi.org/10.1093/nar/gks433.

55. Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles Nucleic Acids Res. 2004;D91–4. https://doi.org/10.1093/nar/gkh012

56. Zhang Y, Romanish MT, Mager DL. Distributions of Transposable Elements Reveal Hazardous Zones in Mammalian Introns. PLoS Comput Biol. 2011;7: e1002046.

57. Payer LM, Steranka JP, Ardeljan D, Walker J, Fitzgerald KC, Calabresi PA, et al. Alu insertion variants alter mRNA splicing. Nucleic Acids Res. 2019;47:421–31. https://doi.org/10.1093/nar/gky1086.

58. Khan AR, Enjalbert J, Marsollier AC, Rousselet A, Goldringer I, Vitte C. Vernalization treatment induces site-specific DNA hypermethylation at the VERNALIZATION-A1 (VRN-A1) locus in hexaploid winter wheat. BMC Plant Biol. 2013;13:209. https://doi.org/10.1186/1471-2229-13-209.

59. Roworth AP, Carr SM, Liu G, Barczak W, Miller RL, Munro S, et al. Arginine methylation expands the regulatory mechanisms and extends the genomic landscape under E2F control. Sci Adv. 2019;5(6):eaaw4640. https://doi.org/10.1126/sciadv.aaw4640

60. Xin D, Hu L, Kong X. Alternative Promoters Influence Alternative Splicing at the Genomic Level. PLoS ONE. 2008;3: e2377. https://doi.org/10.1371/journal.pone.0002377.

61. Doshi J, Willis K, Madurga A, Stelzer C, Benenson Y. Multiple Alternative Promoters and Alternative Splicing Enable Universal Transcription-Based Logic Computation in Mammalian Cells. Cell Rep. 2020;33: 108437. https://doi.org/10.1016/j.celrep.2020.108437.

62. Doi-Katayama Y, Hayashi F, Inoue M, Yabuki T, Aoki M, Seki E, et al. Solution structure of the general transcription factor 2I domain in mouse TFII-I protein. Protein Sci. 2007;16:1788–92. https://doi.org/10.1110/ps.072792007.

63. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. 2020;48:D882–9. https://doi.org/10.1093/nar/gkz1062.

64. Wen YD, Cress WD, Roy AL, Seto E. Histone Deacetylase 3 Binds to and Regulates the Multifunctional Transcription Factor TFII-I. J Biol Chem. 2003;278:1841–7. https://doi.org/10.1074/jbc.M206528200.

65. Hui Ng H, Bird A. Histone deacetylases: silencers for hire. Trends Biochem Sci. 2000;25:121–6. https://doi.org/10.1016/s0968-0004(00)01551-6.

66. Nickerson E, Greenberg F, Keating MT, McCaskill C, Shaffer LG. Deletions of the elastin gene at 7q11.23 occur in approximately 90% of patients with Williams syndrome. Am J Hum Genet. 1995;56:1156–61.

67. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017;18:71–86. https://doi.org/10.1038/nrg.2016.139.

68. Lang GI, Murray AW, Botstein D. The cost of gene expression underlies a fitness trade-off in yeast. Proc Natl Acad Sci USA. 2009;106:5755–60. https://doi.org/10.1073/pnas.090162010.