

Rare germline disorders implicate long non-coding RNAs disrupted by chromosomal structural rearrangements

Rebecca E. Andersen^{1,2,3}, Ibrahim F. Alkuraya^{4,5*}, Abna Ajeesh^{4*}, Tyler Sakamoto^{1,5}, Elijah L. Mena^{2,6}, Sami S. Amr^{2,7}, Hila Romi^{2,4}, Margaret A. Kenna^{2,8}, Caroline D. Robson^{2,8,9}, Ellen S. Wilch⁴, Katarena Nalbandian⁴, Raul Piña-Aguilar^{2,4}, Christopher A. Walsh^{1,2,3,10}, and Cynthia C. Morton^{2,3,4,7,11**}

¹Division of Genetics and Genomics and Manton Center for Orphan Diseases, Boston Children's Hospital, Boston, MA, USA; ²Harvard Medical School, Boston, MA, USA; ³Broad Institute of MIT and Harvard, Cambridge, MA, USA; ⁴Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, MA, USA; ⁵Harvard College, Cambridge, MA, USA; ⁶Division of Genetics, Department of Genetics, Brigham and Women's Hospital, Boston, MA, USA; ⁷Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA; ⁸Department of Otolaryngology, Boston Children's Hospital, Boston, MA, USA; ⁹Department of Radiology, Boston Children's Hospital, Boston, MA, USA; ¹⁰Howard Hughes Medical Institute, Boston, MA, USA; ¹¹University of Manchester, Manchester Center for Audiology and Deafness, UK

* indicates authors with equal contributions

** corresponding author: cmorton@bwh.harvard.edu

ORCID IDs:

Rebecca E. Andersen: 0000-0002-0702-3120

Ibrahim F. Alkuraya: 0000-0003-2489-2464

Abna Ajeesh: 0000-0002-9354-0190

Tyler Sakamoto: 0009-0001-5972-8288

Elijah L. Mena: 0000-0001-9980-9018

Sami S. Amr: 0000-0003-0927-6057

Hila Romi: 0000-0003-4071-4334

Margaret A. Kenna: 0000-0003-0851-9272

Caroline D. Robson: 0000-0001-5592-249X

Ellen S. Wilch: 0009-0008-5381-5437

Katarena Nalbandian: 0000-0002-2617-7387

Raul Piña-Aguilar: 0000-0002-2043-4282

Christopher A. Walsh: 0000-0002-0156-2238

Cynthia C. Morton: 0000-0003-2198-6756

Abstract

In recent years, there has been increased focus on exploring the role the non-protein-coding genome plays in Mendelian disorders. One class of particular interest is long non-coding RNAs (lncRNAs), which has recently been implicated in the regulation of diverse molecular processes. However, because lncRNAs do not encode protein, there is uncertainty regarding what constitutes a pathogenic lncRNA variant, and thus annotating such elements is challenging. The Developmental Genome Anatomy Project (DGAP) and similar projects recruit individuals with apparently balanced chromosomal abnormalities (BCAs) that disrupt or dysregulate genes in order to annotate the human genome. We hypothesized that rearrangements disrupting lncRNAs could be the underlying genetic etiology for the phenotypes of a subset of these individuals. Thus, we assessed 279 cases with BCAs and selected 191 cases with simple BCAs (breakpoints at only two genomic locations) for further analysis of lncRNA disruptions. From these, we identified 66 cases in which the chromosomal rearrangements directly disrupt lncRNAs. Strikingly, the lncRNAs *MEF2C-ASI* and *ENSG00000257522* are each disrupted in two unrelated cases. Furthermore, in 30 cases, no genes of any other class aside from lncRNAs are directly disrupted, consistent with the hypothesis that lncRNA disruptions could underly the phenotypes of these individuals. To showcase the power of this genomic approach for annotating lncRNAs, here we focus on clinical reports and genetic analysis of two individuals with BCAs and additionally highlight six individuals with likely developmental etiologies due to lncRNA disruptions.

Introduction:

Only ~2% of the human genome directly codes for proteins. Among the approximately 20,000 protein-coding genes, 4,000 have been implicated in Mendelian diseases (Avraham et al. 2022). The non-protein coding genome comprises a diverse array of elements including those that transcribe long non-coding RNA molecules (lncRNAs). lncRNAs are transcripts of at least 200 nucleotides in length but are not translated into proteins (Kopp and Mendell 2018). Current transcriptome annotations (Frankish et al. 2021) suggest that there are nearly 20,000 human lncRNAs, and the expression of many of these lncRNAs is highly regulated. However, the biological roles of most lncRNAs remain to be determined, which is made challenging by the fact that lncRNAs can have such a wide variety of different functions. Some lncRNAs regulate the transcription of nearby genes, whereas others regulate other biological processes, including splicing and translation (Taniue and Akimitsu 2021; Statello et al. 2021). One particularly intriguing class of lncRNAs is that of divergent lncRNAs, which have transcriptional start sites (TSSs) within 5kb of another gene and are transcribed in the opposite direction in a “head-to-head” configuration. Divergent lncRNAs have generally been associated with regulation of their neighboring gene, particularly when the neighbor is a transcription factor (Luo et al. 2016; Wang et al. 2020). More broadly, lncRNAs of several classes have been found to affect the expression of their neighboring genes (Statello et al. 2021), providing a mechanism for such lncRNAs to modulate key biological processes.

As a testament to the importance of lncRNAs in normal development, deletions of certain lncRNAs have led to lethal phenotypes in mice as well as abnormal development of the neocortex, lung, gastrointestinal tract, and heart (Sauvageau et al. 2013; Feyder and Goff 2016; Mattick et al. 2023). In addition, lncRNAs have been implicated in cancer and shown to affect cell division,

metabolism, and tumor-host interactions. Thus, lncRNAs are essential to maintaining proper cellular homeostasis (Statello et al. 2021). A prior observation of a variant in a lncRNA causing human disease in a Mendelian fashion is a 27 – 63 kb deletion of a locus that encompasses a lncRNA upstream of the engrailed-1 gene (*EN1*), which resulted in congenital limb abnormalities even though *EN1* itself was not disrupted (Allou et al. 2021). Most recently, a pre-print manuscript (Ganesh et al. 2024) reported three individuals with deletions in *CHASERR*, a lncRNA proximal to *CHD2*, a protein-coding gene that causes developmental and epileptic encephalopathy. Intriguingly, disruption of *CHASERR* leads to increased expression of *CHD2* in *cis*, leading to a distinct clinical presentation compared to individuals with *CHD2* haploinsufficiency.

Given the importance of lncRNAs to gaining a better understanding of developmental biology and improving clinical diagnoses, it is important to develop a better functional assessment of lncRNAs in the human genome. However, the methods of annotating protein-coding genes are not typically applicable to non-coding genes due to their fundamental differences (Mattick et al. 2023). For instance, single nucleotide variants (SNVs) can result in nonsense mutations that prematurely terminate proteins, and indels can cause translational frame shifts that alter the entire downstream amino acid sequence of a protein. However, because lncRNAs do not encode proteins, it is unclear what affect if any such mutations may have on lncRNAs. The study of lncRNAs remains to be elucidated with regard to definitive consensus over what constitutes a pathogenic lncRNA variant.

A prior landmark development in expanding a focus on DNA beyond protein-coding genes to the 3D genome was the discovery of topologically associated domains (TADs). TADs are megabase-sized genomic segments partitioning the genome into large regulatory units with frequent intra-domain chromatin interactions but relatively rare inter-domain interactions

(Lupiáñez et al. 2015). Conserved across different cell types and species, they are considered crucial for spatiotemporal gene expression patterns. Topological boundary regions (TBRs) block interactions between adjacent TADs, and TBR disruption by chromosomal structural rearrangements can result in rewiring of genomic regulators leading to abnormal clinical phenotypes. Rigorous interpretation of clinical phenotypes requires assessment of the boundaries of TADs following a chromosomal structural rearrangement because complex phenotypes may be dissected from rearrangements that reposition lncRNAs with respect to the relevant protein-coding region.

The Developmental Genome Anatomy Project (DGAP) (Higgins et al. 2008) and similar projects have historically explored balanced chromosomal rearrangements to establish possible relationships between genotypes and phenotypes through identifying nucleotide-level breakpoints via Sanger sequencing. Individuals with such rearrangements represent natural gene disruptions and dysregulations, and their chromosomal rearrangements can serve as ideal signposts for annotating the human genome. Unlike for protein-coding genes, it is hard to predict the pathogenicity of lncRNA variants because they are not translated and consequently frameshift and nonsense mutations may not disrupt their function. Here, we employ a foundational approach in human genetics using chromosomal rearrangements to interrogate potential phenotypic impacts of disrupted lncRNAs and their genomic repositioning resulting in dysregulation. Both disruption and dysregulation of lncRNAs therefore may increase the diagnostic yield of developmental disorders. We venture to make a call out to cytogeneticists to employ further the power of chromosomal rearrangements in yet another opportunity to contribute to annotating the genome, recognizing that there are many patients and families who still await diagnoses.

Methods

Human Subjects

Study ID numbers are a consecutive alphanumeric list that are not known outside of the research group. The Partners HealthCare System Internal Review Board (IRB) gave ethical approval for this work under protocol number 1999P003090.

Breakpoint Mapping

Genomic DNA from DGAP probands was sequenced to identify chromosomal breakpoints at nucleotide-level according to the previously published protocol (Talkowski et al. 2011; Hanscom and Talkowski 2014). Sanger sequencing results were aligned to the human genome using the UCSC Genome Browser BLAT tool (Kent 2002). Breakpoints were also compiled from previous publications (Talkowski et al. 2012b; Redin et al. 2017; Lowther et al. 2022). Breakpoint positions were converted from earlier genome builds to hg38 using the UCSC Genome Browser LiftOver tool (Kent et al. 2002).

Additional Genetic Analyses

To ensure that the phenotypes found in DGAP103 and DGAP353 could not be attributed to other variants aside from the chromosomal rearrangements, whole exome sequencing was performed for these cases by the Genomics Platform at the Broad Institute of MIT and Harvard (Cambridge, MA). Sequencing libraries were prepared from sample DNA (250 ng input) using the Twist Bioscience exome (~35Mb target) assay (San Francisco, CA), which were then sequenced (150 bp paired end) on the Illumina NovaSeq platform (Illumina, San Diego, CA) to generate a coverage of >85% of the target region at 20X read depth or greater. Sequencing data for each of the samples was processed through an internal pipeline using the BWA aligner for mapping to the human genome (GRCh38/hg38) and variant calling was performed using the Genome Analysis Toolkit

(GATK) HaplotypeCaller package. Variants were then annotated and assessed for pathogenicity using the Seqr software (Pais et al. 2022). Variants with >1% MAF were filtered out and variants in genes with an association with disease were prioritized for analysis. No candidate variants associated with either phenotype were found. For DGAP355, whole exome sequencing (GeneDX XomeDxXpress) was negative for relevant variants. For DGAP148, array comparative genomic hybridization was previously performed and found to be normal (Redin et al. 2017).

TAD Analysis and Visualization of Chromatin Interactions

TAD boundary positions previously identified by Dixon and colleagues (Dixon et al. 2012) were converted from hg18 to hg38 using the UCSC Genome Browser LiftOver tool (Kent et al. 2002). BEDTools was used to identify TADs that included DGAP breakpoints (Quinlan and Hall 2010). The USCS Genome Browser (Kent et al. 2002) was used to display these TAD regions. Along with the genes within these regions, we also displayed chromatin interactions identified through micro-C studies from H1-hESCs (Krietenstein et al. 2020).

Temporal Bone Computerized Tomography (CT)

Axial temporal bone CT without contrast for the mother of DGAP353 consisted of helical images with the following parameters: Discovery STE system (General Electric Healthcare, Waukesha, WI); 0.625 mm slice thickness, effective mAs of 17, mA of 158, rotation time of 2161 milliseconds, pitch of 0.5625 and kvp of 140. Images were viewed in a plane parallel to that of the horizontal semicircular canal. Coronal reformatted images were obtained in a plane perpendicular to the axial images at 0.74 mm thickness.

Axial temporal bone CT without contrast for DGAP353 consisted of helical images with the following parameters: Discovery STE system (General Electric Healthcare, Waukesha, WI); 0.625 mm slice thickness, effective mAs of 27, mA of 246, rotation time of 2161 milliseconds, pitch of

0.5625 and kvp of 100. Images were viewed in a plane parallel to that of the horizontal semicircular canal. Coronal reformatted images were obtained in a plane perpendicular to the axial images at 0.70 mm thickness.

Results

Identification of Human Subjects with Disrupted lncRNAs

We evaluated 279 cases of balanced chromosomal abnormalities and selected 191 cases with resolved breakpoints indicating a “simple” rearrangement (*i.e.*, breakpoints at only two genomic locations and no significant genomic imbalance) for further analysis (**Table S1**). Using the most recent Human Gencode Reference, Release 45, GRCh38.p14 (Frankish et al. 2021), we then identified 66 cases in which at least one breakpoint overlapped a lncRNA (**Table S2 and Table S3**). Overall, 79 unique lncRNAs were directly disrupted in these cases, and four lncRNAs including *MEF2C-AS1* and *ENSG00000257522* were each disrupted in two unrelated individuals. In 30 of the cases, no genes of any other class aside from lncRNAs were directly disrupted by the breakpoints. In this report, we primarily focus on two cases (DGAP353 and DGAP103) as examples of the potential value of assessing lncRNAs as diagnostic etiologies, and six additional cases are presented for further investigation (**Table 1**).

Clinical Report of DGAP353

The proband DGAP353 was diagnosed during gestation when her healthy mother (20-24 years old) underwent amniocentesis performed following an abnormal maternal serum screen for an elevated risk for trisomy 21. An apparently balanced translocation was detected in the female fetus between the long arms of chromosomes 14 and 17. Parental chromosome analyses revealed maternal inheritance and apparent structural identity to the maternal t(14;17) rearrangement.

DGAP353's G-banded karyotype is described as 46,XX,t(14;17)(q24.3;q23)mat and the mother's karyotype is 46,XX,t(14;17)(q24.3;q23). No clinical abnormalities were observed in the fetus and the pregnancy was continued. DGAP353 began developing signs of hearing loss between the ages of 10-14 years old, and her hearing loss was found to be primarily sensorineural with a conductive element. Around this time, surgery was performed to rectify the conductive abnormalities, but the sensorineural hearing loss remained. The mother of DGAP353 began wearing hearing aids around 35-44 years of age, after a gradual decline in hearing for an unspecified time period. Both DGAP353 and her mother were otherwise healthy, typical of nonsyndromic deafness of unknown genetic etiology. Computerized tomography (CT) imaging of the temporal bones of DGAP353 and her mother revealed abnormalities such as unusually small sinus timpani and narrowing of the round and oval windows (**Fig. S1**).

Breakpoint analysis of DGAP353

Both DGAP353 and her mother harbor a translocation between chromosomes 14 and 17 with a 7 base-pair (bp) insertion of DNA of non-templated origin at the breakpoint in the der(17) chromosome (**Fig. 1A**). Following suggested nomenclature (Ordulu et al. 2014), the next-generation cytogenetic nucleotide level research rearrangement is described in a single line as:
46,XX,t(14;17)(q24.3;q23)mat.seq[GRCh38] t(14;17)(14pter→14q23.3(+)(65,855,3 {58-60})::17q23.2(+)(61,393,84 {1-3})→17qter;17pter→17q23.2(+)(61,393,812)::TATATAC::14q23.3(+)(65,855,359)→14qter)mat.

TBX2-AS1 is a candidate lncRNA for an association with hearing loss

The DGAP353 breakpoints do not overlap any genes on chromosome 14 (**Fig. 1B**); however, this translocation results in the direct disruption of the lncRNA *TBX2-ASI* from chromosome 17 (**Fig. 1C**). The Gencode annotation also lists the lncRNA *ENSG00000267131* as a separate gene that is disrupted by these breakpoints, however this has been identified as an isoform of *TBX2-ASI* by LNCipedia (Volders et al. 2019). While little is known regarding the biological role of *TBX2-ASI*, particularly in the context of hearing, the orthologous mouse lncRNA (*2610027K06Rik*) has been detected in the cochlear and vestibular sensory epithelium of embryonic and postnatal mice (identified as “XLOC_007930”) (Ushakov et al. 2017). Using the Gene Expression Analysis Resource (gEAR) portal (Orvis et al. 2021), we further found that while *Tbx2-as1* is detected in supporting cell types (pillar and Deiters cells), it is predominantly expressed by sensory inner hair cells, as determined through cell-type-specific RNA-seq (Liu et al. 2018). Thus, the expression pattern of *Tbx2-as1* is consistent with the finding that the hearing loss demonstrated by DGAP353 and her mother is primarily sensorineural.

The lncRNA gene *TBX2-ASI* exists in a divergent configuration with the protein-coding gene *TBX2*. Divergent lncRNAs are a particularly interesting class because they have often been found to regulate their neighboring gene. In many cases, the divergent lncRNA affects expression of the neighbor in *cis* (Luo et al. 2016; Wang et al. 2020). Some divergent lncRNAs have also been found to modulate the downstream functions of the protein generated by the neighboring gene. For example, the divergent lncRNA *Six3OS* neighbors the gene *SIX3* and regulates the activity of the *SIX3* protein by functioning as a molecular scaffold (Rapicavoli et al. 2011). Similarly, *Paupar* is a lncRNA that is divergent to *Pax6* and physically interacts with the *PAX6* protein to affect how this transcription factor regulates target genes (Vance et al. 2014; Pavlaki et al. 2018). It has recently been proposed that *TBX2-ASI* may function in a similar manner to regulate *TBX2* target

genes in neuroblastoma cells (Modi et al. 2023). Thus, knowledge regarding the function of the protein-coding member of a divergent pair can provide insights into the potential biological role of the lncRNA partner.

Intriguingly, *TBX2* has previously been linked with hearing and inner ear development. In mice, *Tbx2* has been associated with otocyst patterning in inner ear morphogenesis, as mouse models in which *Tbx2* was conditionally knocked out exhibit cochlear hypoplasia (Kaiser et al. 2021). Previous studies have also shown that deletions encompassing *TBX2* and *TBX2-AS1* are found in individuals with hearing loss, albeit in conjunction with other deleted genes (Ballif et al. 2010; Nimmakayalu et al. 2011; Schönewolf-Greulich et al. 2011). In addition, a recent study has shown that *Tbx2* is required for inner hair cell and outer hair cell differentiation, demonstrating that it is a master regulator of hair cell fate (García-Añoveros et al. 2022). Therefore, we suggest that the translocation disrupting *TBX2-AS1* in DGAP353 and her mother may lead to altered expression or function of *TBX2*, ultimately resulting in the phenotype of hearing loss.

Clinical Report of DGAP103

The proband DGAP103 was referred to DGAP between 5-9 years of age with a complex overgrowth phenotype that we previously reported (Ligon and Moore et al., 2005). Between 0-4 years of age DGAP103 exhibited premature dentition, including potential supernumerary teeth identified by panoramic dental X-ray. Around this time, additional features including height, weight, and occipital-frontal circumference were all at or above the 95th percentile. Magnetic resonance imaging performed around this time due to macrocephaly revealed a two cm right cerebellar lesion ventral to cerebellar nuclei and stable by serial imaging as of 5-9 years of age. A bone marrow biopsy was performed at 5-9 years of age due to mild thrombocytopenia and

leukopenia (platelet counts as low as 60,000) with trilineal hematopoiesis and mildly reduced cellularity but no evidence of malignancy. A G-banded cytogenetic analysis was performed and reported as a constitutional 46,XY,inv(12)(p12.2q15), which was later revised through the DGAP study to 46,XY,inv(12)(p11.22q14.3)dn. Upon enrollment in DGAP at 5-9 years of age, DGAP103 had extreme overgrowth (height at the 50%ile for 15-year-old males, weight at the 50%ile for 14-year-old males, and head circumference at the 50%ile for adult males), facial dysmorphism, brachydactyly of hands and feet with shortened distal phalanges and redundant curved nails, and bilateral lower extremity nodules of adipocytes and fibrovascular tissue consistent with lipomas.

Breakpoint analysis of DGAP103

We have now performed breakpoint analysis, including Sanger sequencing confirmation, which determined that DGAP103 harbors a pericentric inversion of chromosome 12 with a 9bp insertion of DNA of non-templated origin (**Fig. 2A**). Following suggested nomenclature (Ordulu et al. 2014), the next-generation cytogenetic nucleotide level rearrangement is described in a single line as:

46,XY,inv(12)(p12.2q15)dn.seq[GRCh38] inv(12)(pter→p11.22(+)(28,843,402)::TCTCAAAAA::q14.3(-)(65,851,667)→p11.22(-)(28,843,40{4})::q14.3(+)(65,851,66{9})→qter)dn.

Evaluation of genetic etiology of brachydactyly in DGAP103

The DGAP103 breakpoints at 12p11.22 do not overlap any known genes (**Fig. 2B**), but they occur within a TAD that contains *PTHLH*, a protein-coding gene. *PTHLH* is the only gene in a 5 Mb region around the breakpoint on the short arm of chromosome 12 with a pHaplo score above the threshold of 0.86 (*PTHLH* pHaplo = 0.92), indicating that it is predicted to exhibit

haploinsufficiency (Collins et al. 2022). While our initial study of DGAP103 left the cause of brachydactyly unaddressed (Ligon and Moore et al., 2005), it was subsequently reported that deletions and point mutations in *PTHLH* result in brachydactyly (Klopocki et al. 2010; Bae et al. 2018; Reyes et al. 2019). Moreover, *PTHLH* is described by ClinGen as having sufficient evidence for haploinsufficiency causing brachydactyly type E2 (Rehm et al. 2015). In DGAP103, *PTHLH* is not directly disrupted by the chromosomal inversion; however, previously published data from micro-C studies to identify genome-wide chromatin interactions (Krietenstein et al. 2020) demonstrate that the *PTHLH* locus interacts with a region in 12p11.22 that is centromeric to the breakpoints in DGAP103, which would separate this region from *PTHLH*. This region exhibits chromatin modifications associated with enhancer activity such as H3K4Me1 and H3K27Ac (**Fig. 2C**), as determined by chromatin immunoprecipitation analysis from the ENCODE consortium (ENCODE Project Consortium 2012). Thus, altered regulation of *PTHLH* is the most likely genetic etiology for brachydactyly.

Reevaluation of genetic etiology of overgrowth phenotypes in DGAP103 upon annotation of the lncRNA HMGA2-ASI

The DGAP103 breakpoints at 12q14.3 disrupt most of the isoforms of the protein-coding gene *HMGA2* (also known as *HMGI-C*) within the canonical third intron, as well as one isoform of its antisense lncRNA *HMGA2-ASI* (**Fig. 2D**). At the time of our initial study of DGAP103, *HMGA2-ASI* had not yet been identified, but *HMGA2* had been described as a member of the high-mobility group AT-hook (HMGA) family, which bind to DNA through three conserved AT-hook domains (Zhou and Chada 1998). *HMGA2* was already known to play crucial roles in the regulation of growth, with *Hmga2*-null mice exhibiting substantially reduced body size, while

heterozygous mice are more mildly affected (Zhou et al. 1995). More recently, microdeletions in 12q14 that include *HMGGA2* have been identified in individuals with short stature (Lynch et al. 2011; Alyaqoub et al. 2012), and disruptive variants in *HMGGA2* have been found to cause fetal growth restriction (Abi Habib et al. 2018), supporting the role of *HMGGA2* in human growth. Conversely, certain disruptions within *HMGGA2* had been linked to increased proliferation, and chromosomal rearrangements within the canonical third intron of *HMGGA2* had been described as the most frequent chromosomal aberration in human tumors (Kazmierczak et al. 1998). In some cases, chromosomal rearrangements had been identified in mesenchymal tumors including lipomas that result in the fusion of the *HMGGA2* DNA binding domains to other regulatory domains (Ashar et al. 1995; Schoenmakers et al. 1995). Therefore, we previously tested whether a fusion product between the first three exons of *HMGGA2* and *PTHLH* might exist in DGAP103, however no fusion product was detected (Ligon and Moore et al., 2005). It had also been demonstrated that overexpression of truncated *Hmg2a* causes gigantism and lipomatosis in transgenic mice (Battista et al. 1999). Thus, our initial study of DGAP103 proposed that truncation of *HMGGA2*, through an inversion with a breakpoint in canonical intron 3 that leaves canonical exons 1-3 intact, was the most likely genetic etiology for overgrowth and multiple lipomas (Ligon and Moore et al., 2005).

Since then, continued annotation of the human transcriptome has provided substantially greater detail regarding the region surrounding the 12q14.3 breakpoints in DGAP103. The most recent annotation, Human Gencode Reference, Release 45, GRCh38.p14 (Frankish et al. 2021), now includes 10 isoforms of *HMGGA2* and reveals that two short isoforms naturally terminate centromeric to the DGAP103 breakpoints (**Fig. 2D**). Unlike the longer isoforms of *HMGGA2*, these two short isoforms would remain intact in DGAP103, but they would be repositioned to the short arm of the derivative chromosome 12.

Additionally, updated transcriptome annotations have uncovered the lncRNA *HMGA2-ASI* (**Fig. 2D**). This lncRNA is antisense to *HMGA2* and is completely overlapped by the *HMGA2* gene but transcribed from the opposite strand. Such antisense lncRNAs can function to modulate the expression of their overlapping partner in *cis* (Statello et al. 2021). More broadly, lncRNAs of several different classes have been found to regulate their neighboring genes in *cis* (Ferrer and Dimitrova 2024). For instance, mouse models have shown that the lncRNA *Chaserr* represses the expression of its neighbor *Chd2* in an allele-specific manner, demonstrating that it functions strictly in *cis* (Rom et al. 2019). Interestingly, *Chd2* promotes the expression of *Chaserr*, and thus these neighbors participate in a regulatory feedback loop in which the *Chaserr* lncRNA serves as a sensor to tightly maintain appropriate levels of *Chd2*. A recent pre-print manuscript has further demonstrated that human *CHASERR* similarly regulates *CHD2* in *cis*; individuals with *de novo* deletions in the *CHASERR* locus exhibit increased *CHD2* expression from the neighboring allele, leading to severe developmental delay and facial dysmorphisms (Ganesh et al. 2024). Intriguingly, it has recently been found that another member of the *HMGA* family, *HMGA1*, is repressed by the nearby lncRNA *HMGA1-lnc* (Stewart et al. 2020). *HMGA2-ASI* could similarly play a critical role in regulating *HMGA2*.

New analyses have also led to a refined understanding of *HMGA2* and its effects on proliferation. It was previously thought that either the creation of an *HMGA2* fusion product or the truncation of *HMGA2* were required to cause overgrowth phenotypes, including mesenchymal tumors (Fedele et al. 1998; Battista et al. 1999). However, it has more recently been shown that the overexpression of either full length or truncated human *HMGA2* in differentiated mesenchymal cells is sufficient to cause mesenchymal tumors in transgenic mouse models, and it is now proposed that the overexpression of the three *HMGA2* DNA-binding domains is the key

requirement for these phenotypes (Zaidi et al. 2006). Importantly, the two short isoforms of *HMGA2* contain these DNA-binding domains, with identical sequence to that from the canonical full length *HMGA2* isoform. Thus, the overgrowth phenotypes of DGAP103 could be explained by altered regulation of *HMGA2* through the repositioning of its short isoforms to a new genomic location that lacks the antisense lncRNA *HMGA2-ASI*. Indeed, we previously detected increased expression of *HMGA2* in a lymphoblastoid cell line from DGAP103 (Ligon and Moore et al., 2005). Thus, we propose that the DGAP103 chromosomal rearrangement separating the short isoforms of *HMGA2* from the antisense lncRNA *HMGA2-ASI* may lead to increased expression of *HMGA2*, resulting in the phenotypes of overgrowth and multiple lipomas.

Recurrent disruptions of the lncRNA MEF2C-ASI in individuals with neurological phenotypes

We additionally identified two cases, DGAP191 and DGAP218, with chromosomal rearrangements that disrupt the lncRNA *MEF2C-ASI*. Following suggested nomenclature (Ordulu et al. 2014), the next-generation cytogenetic nucleotide level research rearrangements are described in a single line as:

DGAP191: 46,XY,t(5;7)(q14.3;q21.3)dn.seq[GRCh38] t(5;7)(5pter→5q14.3(+)(89,411,06{3-5})::7q21.3(+)(94,378,2{48-50})→7qter;7pter→7q21.3(+)(94,378,25{3-5})::5q14.3(+)(89,411,07{0-2})→5qter)dn.

DGAP218: 46,XX,inv(5)(p12q13.1)dn.seq[GRCh38] inv(5)(pter→p14.2(+)(24,272,19{3})::q14.3(-)(89,105,02{6})→p14.2(-)(24,272,189)::TATTTATATGACAAG::q14.3(+)(89,105,031)→qter)dn.

In both cases, the 5q14.3 breakpoints directly disrupt the lncRNA *MEF2C-ASI*. In DGAP191, the 7q21.3 breakpoints additionally overlap the lncRNA *ENSG00000285090*, but no

protein-coding genes are directly disrupted (**Fig. 3**). In DGAP218, *MEF2C-ASI* is the only gene of any class that is directly disrupted (**Fig. 4**).

We previously reported both of these individuals as part of a larger set of cases with breakpoints in 5q14.3 (Redin et al. 2017). This region is of particular interest due to 5q14.3 microdeletion syndrome, which is characterized by neurological phenotypes including intellectual disability and epilepsy (Zweier and Rauch 2012). This syndrome is now recognized to be driven by decreased *MEF2C* expression, either through direct disruption of *MEF2C* or due to distal mutations (Zweier and Rauch 2012). Indeed, when we previously described DGAP191 and DGAP218 (Redin et al. 2017), we noted that their phenotypes were similar to individuals with direct *MEF2C* disruptions. Furthermore, we determined that levels of *MEF2C* expression were reduced in lymphoblastoid cell lines from both DGAP191 and DGAP218 (Redin et al. 2017); however, no mention was made of *MEF2C-ASI*. Recent studies have further elucidated the functional effects of altering *MEF2C* or its topological organization (Mohajeri et al. 2022), but the potential role of *MEF2C-ASI* remains unclear.

While there is still little known regarding the function of *MEF2C-ASI*, it has recently been found that *MEF2C-ASI* can positively regulate the expression of *MEF2C* in human cervical cancer cell lines (Guo et al. 2022). Interestingly, *MEF2C-ASI* is transcribed through multiple putative enhancers of *MEF2C* (D'haene et al. 2019), providing a potential mechanism for this lncRNA to regulate expression of its neighboring gene, as has previously been described for lncRNAs such as *Bendr* (Engreitz et al. 2016) and *Uph* (Anderson et al. 2016). Thus, for DGAP191 and DGAP218 we now propose that the disruption of *MEF2C-ASI* leads to decreased expression of *MEF2C*, resulting in neurological phenotypes.

The lncRNA ENSG00000257522 is recurrently disrupted in individuals with microcephaly

Our analysis further identified two cases, DGAP245 and NIJ1, with chromosomal rearrangements that disrupt the lncRNA *ENSG00000257522* (**Fig. 5 and Fig. 6**). These individuals exhibit shared phenotypes (**Table S1**) including microcephaly and defects of the corpus callosum. Following suggested nomenclature (Ordulu et al. 2014), the next-generation cytogenetic nucleotide level research rearrangements are described in a single line as:

DGAP245: 46,XY,t(3;14)(p23;q13)dn.seq[GRCh38] t(3;14)(3qter→3p22.2(-)(36,927,959)::CATTTGTTCAAATTTAGTTCAAATGA::14q12(+)(29,276,117)→14qter;14pter→14q12(+)(29,276,10{8-9})::3p22.2(-)(36,927,6{49-50})→3pter)dn.

NIJ1: 46,XX,t(8;14)(q21.2;q12)dn.seq[GRCh38] t(8;14)(8pter→8q21.12(+)(78,898,16{9})::14q12(+)(29,296,33{1})→14qter;14pter→14q12(+)(29,296,328)::AAAT::8q21.12(+)(78,898,172)→8qter)dn.

In both cases, the 14q12 breakpoints directly disrupt the lncRNA *ENSG00000257522* as well as the overlapping antisense lncRNA *ENSG00000258028*. In DGAP245, the 3p22.2 breakpoints additionally disrupt the protein-coding gene *TRANK1* (**Fig. 5B**), however this gene is not predicted to be haploinsufficient (pHaplo = 0.29) (Collins et al. 2022) and it has not been implicated in any human phenotypes by OMIM. In NIJ1, the 8q21.12 breakpoints disrupt the lncRNA *MITAI* (**Fig. 6B**). Given that the only shared disruptions between these cases are to the lncRNAs *ENSG00000257522* and *ENSG00000258028*, we focused on these for further analysis.

Using the GTEx database (Lonsdale et al. 2013), we found that *ENSG00000258028* is not readily detected in neural tissue, and thus it is unlikely to cause the patient phenotypes. In contrast, *ENSG00000257522* is primarily expressed in neural tissue (**Fig. 7A**), suggesting that it could play an important neurological role. Moreover, *ENSG00000257522* exists within the same TAD as the

protein-coding gene *FOXGI*, disruptions in which have been associated with a variant of Rett syndrome (MIM # 613454) (Ariani et al. 2008) as well as *FOXGI* syndrome (Kortüm et al. 2011). Core phenotypes of these syndromes include microcephaly and corpus callosum defects, implicating *FOXGI* dysregulation as the underlying genetic etiology in DGAP245 and NIJ1. Thus, we sought to identify potential regulatory elements that could be disrupted by the chromosomal rearrangements in these cases, and found three regions with prominent H3K4me1 chromatin modification (**Fig. 7B**), which is associated with enhancer activity (ENCODE Project Consortium 2012). Notably, one of these regions also exhibited H3K27Ac modification, which is also associated with enhancer activity (ENCODE Project Consortium 2012). Furthermore, these three regions each include a portion that has been demonstrated to drive reporter expression in neural tissue *in vivo* in transgenic mice (hs566, hs1539, and hs1168) (Visel et al. 2007), and thus these regions exert experimentally validated enhancer activity.

Strikingly, all three of these enhancers exist within the lncRNA *ENSG00000257522*. While the most distal enhancer is partially disrupted by the breakpoints in DGAP245, the other two enhancers remain in the appropriate position relative to *FOXGI*. In NIJ1, all three of the enhancers are proximal to the breakpoints and are not separated from *FOXGI*. Thus, these enhancers are not directly disrupted by the chromosomal rearrangements, and instead their activity could be impaired due to the disruption of the lncRNA in which they are embedded. Indeed, transcription of lncRNAs through enhancers is a well-documented mechanism through which lncRNAs can regulate gene expression (Statello et al. 2021). Thus, we propose that the lncRNA *ENSG00000257522* regulates the expression of *FOXGI* through its effects on the embedded enhancers.

Further supporting this, we also identified an individual with a complex *de novo* rearrangement that similarly disrupts *ENSG00000257522*. This individual, DGAP246, exhibits

consistent phenotypes including microcephaly (Redin et al. 2017). The complex rearrangement in DGAP246 consists of 14 pairs of breakpoints, including eight breakpoints in 14q12. Overall, this results in the direct disruption of the lncRNA *ENSG00000257522* while leaving the two most proximal enhancer elements in their correct position relative to *FOXG1*. Taken together, these three cases implicate the lncRNA *ENSG00000257522* in the regulation of *FOXG1*. Additionally, previous studies have reported several individuals with *FOXG1* syndrome that harbor disruptions in this region, including a translocation in “Patient 1” that directly disrupts *ENSG00000257522* (Mehrjouy et al. 2018). Thus, we propose that disruptions of this lncRNA can cause phenotypes including microcephaly and defects of the corpus callosum, consistent with *FOXG1* syndrome.

Potential regulation of KIRREL3 by its neighboring lncRNA ENSG00000255087

We previously described DGAP148 as an individual with a neurodevelopmental disorder including attention deficits and difficulty with spatial coordination (Talkowski et al. 2012b). We have recently received updated information from the referring clinical geneticist indicating that this individual is overall in good health but continues to treat attention-deficit/hyperactivity disorder (ADHD). She was not able to complete regular high school, however she is employed. While she does not live alone, she is autonomous for tasks of everyday living, including meals, laundry, exercise, and driving. She is also described as very sociable.

DGAP148 has a *de novo* translocation (**Fig. S2A**), and following suggested nomenclature (Ordulu et al. 2014), the next-generation cytogenetic nucleotide level research rearrangement is described in a single line as:

46,X,t(X;11)(p11.2;q23.3)dn.seq[GRCh38] t(X;11)(Xqter→Xp11.4(-)(39,882,592)::TCACTGT ACAG::11q24.2(+)(127,040,509)→11qter;11pter→11q24.2(+)(127,040,509)::CTC::Xp11.4(-)(39,882,591)→Xpter)dn.

The 11q24.2 breakpoints disrupt the lncRNA *ENSG00000255087* (**Fig. S2B**). No other genes are directly disrupted by this translocation (**Fig. S3A**).

At the time of our initial report of DGAP148 (Talkowski et al. 2012b), we were unaware of the lncRNA *ENSG00000255087*, which still lacks any PubMed publications. However, *ENSG00000255087* is approximately 20kb upstream of the protein-coding gene *KIRREL3*, which has been associated with neurodevelopmental phenotypes including attention deficits (Ciaccio et al. 2021; Querzani et al. 2023). We previously found that expression of *KIRREL3* was reduced in DGAP148 (Talkowski et al. 2012b), but the potential mechanism underlying this was unclear. Upon reanalyzing this case and determining that the lncRNA *ENSG00000255087* is directly disrupted, we used the GTEx database (Lonsdale et al. 2013) to assess the expression of *ENSG00000255087*. We find that *ENSG00000255087* is predominantly expressed in neural tissue (**Fig. S2C**), similar to the *KIRREL3* expression pattern (**Fig. S3B**) and consistent with a potential neurodevelopmental role. Considering that the translocation in DGAP148 directly disrupts *ENSG00000255087* and that DGAP148 exhibits decreased *KIRREL3* levels, we suggest that the lncRNA *ENSG00000255087* is a candidate for regulating expression of its neighboring gene *KIRREL3*. Furthermore, we propose that disruption of *ENSG00000255087* can thus lead to the neurodevelopmental phenotypes described for DGAP148.

The lncRNA SOX2-OT is implicated in an individual with epilepsy and autism spectrum disorder

DGAP355 is a nonverbal individual with global developmental delay, autism spectrum disorder (ASD), seizures, and epilepsy, whose mother has a history of multiple miscarriages. This individual has a *de novo* translocation between chromosomes 3 and 9 (**Fig. S4A**), and following suggested nomenclature (Ordulu et al. 2014), the next-generation cytogenetic nucleotide level research rearrangement is described here in a single line as solved by liWGS:

46,XX,t(3;9)(q26.3;q21.1)dn.seq[GRCh38] t(3;9)(3pter→3q26.33(+)(181,488,756)::9q21.13(+)(74,713,321)→9qter;9pter→9q21.13(+)(74,712,100)::3q26.33(+)(181,489,591)→3qter)dn.

The 3q26.33 breakpoints occur within the lncRNA *SOX2-OT* (**Fig. S4B**), which is the only gene directly disrupted by this translocation (**Fig. S5A**). *SOX2-OT* consists of dozens of isoforms that together span a nearly 850kb genomic region. The protein-coding gene *SOX2* exists entirely within an intron of *SOX2-OT* and is transcribed in the same direction. *SOX2* is a transcription factor that serves as a crucial regulator of the potency and self-renewal capacity of several progenitor cell types (Arnold et al. 2011), and in particular *SOX2* is known to play important roles in neural progenitor cells (Graham et al. 2003). *SOX2-OT* exhibits a similar expression pattern to *SOX2*, with both genes primarily expressed in neural tissue (**Fig. S4C and Fig. S5B**). Recently, *SOX2-OT* has been found to affect *SOX2* expression in varying ways in different contexts (Shahryari et al. 2015; Knauss et al. 2018; Li et al. 2020; Yin et al. 2020). Thus, we propose that the translocation in DGAP355 that disrupts the lncRNA *SOX2-OT* may lead to dysregulated expression of *SOX2*, resulting in neurodevelopmental phenotypes including ASD and epilepsy.

Several lncRNAs are directly disrupted in DGAP cases

Additional DGAP cases in which lncRNAs are directly disrupted are listed in **Table S2**. These lncRNAs warrant further consideration, particularly for cases in which no other genes are

directly disrupted. Given the abundance of lncRNAs throughout the human genome, it is not rare for chromosomal rearrangements to disrupt lncRNAs, and yet this class of gene has remained largely overlooked. As updated human genome annotations continue to include new lncRNAs, it is increasingly likely to identify lncRNAs disrupted by chromosomal rearrangements. The cases described here emphasize the importance of carefully considering such disrupted lncRNAs when evaluating potential genetic etiologies underlying patient conditions.

Discussion

By virtue of their noncoding nature, it is difficult to assess the pathogenicity of lncRNA variants based on standards for protein-coding genes. As such, we propose a novel framework to implicate lncRNAs based on chromosomal rearrangements that disrupt the lncRNA function as we illustrate in two DGAP case examples, one involving deafness and the other a complex phenotype including overgrowth, lipomas and brachydactyly.

Deafness/hard-of-hearing (DHH) represents the most prevalent form of sensory deficit in humans. Approximately 5% of the global population are affected by the condition, and it is mostly of a genetic etiology in developed countries (Azaiez et al. 2018). Genetically determined DHH can be subdivided into Mendelian inheritance and complex inheritance, with the former being further classified into syndromic or nonsyndromic forms (Sheffield and Smith 2019). Over 100 genes are associated with nonsyndromic forms and more than 400 with syndromic forms (Alford et al. 2014). There has been an increased interest in investigating the role the non-coding genome plays in hearing loss. lncRNAs *Meg3*, *Rubie* and *Gm15083/lnc83* have been associated with proper functioning and development of the inner ear (Avraham et al. 2022). In addition, genomic duplications responsible for the DFNA58 form of deafness have been found to include certain

lncRNAs, although it is unreported as to whether a lncRNA(s) might be etiologic (Lezirovitz et al. 2020; Nascimento et al. 2022). Herein, we propose a divergent lncRNA disruption to cause human disease in a Mendelian fashion in a familial case of hearing loss.

We report the disruption of the lncRNA *TBX2-ASI* by a balanced chromosomal rearrangement that segregated with DHH from mother to daughter. Little is currently known about human *TBX2-ASI* other than that it is divergent to *TBX2*. It is not listed currently in OMIM, and available databases including gnomAD (Karczewski et al. 2020) and DECIPHER (Firth et al. 2009) cannot be used to determine the level of constraint in the human genome pool or the tolerance to haploinsufficiency for lncRNAs because these metrics are defined specifically for protein-coding genes. *TBX2*, however, has been linked to hearing and inner ear development, including through the identification of deletions encompassing *TBX2* and *TBX2-ASI* (among other genes) that were found in individuals with hearing loss (Ballif et al. 2010; Nimmakayalu et al. 2011; Schönewolf-Greulich et al. 2011).

In the DGAP353 proband presented herein, the breakpoint did not affect *TBX2* itself but interrupted *TBX2-ASI*. It has been shown that *TBX2* maps to the edge of a TAD and is linked to *TBX2-ASI* as a bi-directionally transcribed topological anchor point (tap)RNA (Amaral et al. 2018; Decaestecker et al. 2018). Expression levels of such lncRNAs have been found to be highly correlated with those of their nearest protein-coding genes, and this has also been observed between *TBX2* and *TBX2-ASI*, suggesting that *TBX2-ASI* and *TBX2* may be connected on a regulatory level (Wansleben et al. 2014; Decaestecker et al. 2018). Alternatively, *TBX2-ASI* could affect the function of the *TBX2* protein, as has been demonstrated for other divergent lncRNAs (Rapicavoli et al. 2011; Vance et al. 2014; Pavlaki et al. 2018). Thus, we propose the lncRNA *TBX2-ASI* as a candidate for an association with hearing loss.

Although there is a discrepancy between the ages of onset of the mother's and daughter's hearing loss, this may be attributed to anticipation or to confounding environmental exposures. There is also the possibility that the mother's hearing loss began at an earlier age than indicated because she reported a significant improvement in her hearing when habilitated with hearing aids, suggesting that her hearing began to decline at an earlier age. To assess fully whether *TBX2-ASI* disruption is the causal agent for their hearing loss, a mouse model where *TBX2-ASI* is knocked down while *TBX2* remains intact would be valuable. Additional cases of *TBX2-ASI* deleterious variants with hearing loss will be needed to confirm the proposed association.

We also present DGAP103, an individual with brachydactyly, overgrowth, and lipomas that we initially assessed nearly 20 years ago (Ligon and Moore et al., 2005). We have now performed additional analyses that have enabled us to refine the breakpoints in DGAP103 and to reinterpret the genetic etiology of the phenotypes. Since our initial assessment, it has been reported that haploinsufficiency of *PTHLH* can cause brachydactyly (Klopocki et al. 2010; Bae et al. 2018; Reyes et al. 2019). We have now identified a potential regulatory region that has previously been demonstrated to interact with the *PTHLH* genomic region through micro-C analyses (Krietenstein et al. 2020). This region is separated from *PTHLH* by the inversion in DGAP103, and thus we propose dysregulation of *PTHLH* as the genetic etiology of brachydactyly in DGAP103.

Furthermore, the inversion in DGAP103 results in the repositioning of two short isoforms of *HMGA2* away from the antisense lncRNA *HMGA2-ASI*. During our initial assessment of DGAP103, these short isoforms of *HMGA2* had not been described, and the lncRNA *HMGA2-ASI* had not yet been identified. However, we had found that the expression of *HMGA2* was increased in lymphoblastoid cells from DGAP103 (Ligon and Moore et al., 2005). This is consistent with the finding that overgrowth phenotypes including gigantism and lipomatosis can be caused by

overexpression of a short version of *Hmg2a* in transgenic mice (Battista et al. 1999). Given that a prominent mechanism of lncRNA function is through the regulation of neighboring genes (Statello et al. 2021), we now propose that *HMGA2-ASI* may repress *HMGA2* expression in *cis*; due to the inversion in DGAP103, such regulation has been lost, leading to increased levels of *HMGA2* resulting in the phenotypes of overgrowth and lipomas. This showcases the need to take into account the presence of lncRNAs to provide a more complete understanding of genetic etiologies. We present DGAP103 as an example to highlight the importance of reevaluating diagnoses in view of lncRNAs which were not previously annotated upon initial evaluation.

In conclusion, we have provided examples implicating lncRNAs in hearing loss (*TBX2-ASI*) and in a complex phenotype of overgrowth, lipomas and brachydactyly (*HMGA2-ASI*). We have also identified several additional lncRNAs that warrant further investigation, including *MEF2C-ASI*, *ENSG00000257522*, *ENSG00000255087*, and *SOX2-OT*. The potential connections between these lncRNAs and patient phenotypes were uncovered due to balanced chromosomal rearrangements in these loci; as such, we propose that such rearrangements are an untapped resource to functionally annotate lncRNAs. We propose that geneticists pay special attention to potential dysregulation of lncRNAs in patients where balanced chromosomal rearrangements do not disrupt protein-coding genes in a manner consistent with the observed phenotypes. With an increasing number of chromosomal rearrangements mapped due to inexpensive whole genome sequencing and optical genome mapping, additional lncRNAs that underly developmental diseases await characterization.

Acknowledgements

We thank the DGAP individuals, their families and their clinicians for participation in DGAP, and the Harvard Medical School-affiliated DGAP PIs and members of their laboratories. This study was supported by the National Institute of General Medical Sciences T32 GM007748 (awarded to C.C.M. and funding provided to R.E.A) and P01 GM061354 (awarded to C.C.M). C.C.M. is also supported by the NIHR Manchester Biomedical Research Centre. C.A.W. is an Investigator of the Howard Hughes Medical Institute and is supported by the National Institute of Neurological Disorders and Stroke 5R01NS035129 and by the Allen Discovery Center for Human Brain Evolution through the Paul G. Allen Frontiers Program. E.L.M. is supported by the National Institute on Aging K99 fellowship K99AG081456.

Figure and table legends

Figure 1. A) Chromosome diagrams depict the translocation between 14q23.3 and 17q23.2 in DGAP353. Above, TADs containing the breakpoints are shown, with the breakpoint positions indicated by vertical yellow bars. Protein-coding genes are shown in blue and non-coding genes in green, with a single isoform depicted per gene. TAD borders were defined in (Dixon et al. 2012). Triangular contact maps display micro-C data from (Krietenstein et al. 2020). B) Expanded view of the genomic region surrounding the 14q23.3 breakpoints in DGAP353. C) Expanded view of the genomic region surrounding the 17q23.2 breakpoints in DGAP353. The directly disrupted lncRNA *TBX2-ASI* is highlighted in red. *ENSG00000267131* has been identified as an isoform of *TBX2-ASI* by LNCipedia (Volders et al. 2019).

Figure 2. A) Chromosome diagrams depict the inversion between 12p11.22 and 12q14.3 in DGAP103. Above, TADs containing the breakpoints are shown, with the breakpoint positions

indicated by vertical yellow bars. Protein-coding genes are shown in blue and non-coding genes in green, with a single isoform depicted per gene. TAD borders were defined in (Dixon et al. 2012). Triangular contact maps display micro-C data from (Krietenstein et al. 2020). B) Expanded view of the genomic region surrounding the 12p11.22 breakpoints in DGAP103. C) The micro-C data for the TAD surrounding the 12p11.22 breakpoints in DGAP103 is expanded and annotated to highlight the interactions between the *PTHLH* locus and the potential regulatory region distal to the breakpoints. Below, the layered H3K4Me1 and H3K27Ac tracks show data from the Bernstein Lab at the Broad Institute. D) Expanded view of the genomic region surrounding the 12q14.3 breakpoints in DGAP103. The directly disrupted genes, protein-coding gene *HMG A2* and lncRNA *HMG A2-AS1*, are highlighted in red.

Figure 3. A) Chromosome diagrams depict the translocation between 5q14.3 and 7q21.3 in DGAP191. Above, TADs containing the breakpoints are shown, with the breakpoint positions indicated by vertical yellow bars. Protein-coding genes are shown in blue and non-coding genes in green, with a single isoform depicted per gene. TAD borders were defined in (Dixon et al. 2012). Triangular contact maps display micro-C data from (Krietenstein et al. 2020). B) Expanded view of the genomic region surrounding the 5q14.3 breakpoints in DGAP191. The directly disrupted lncRNA *MEF2C-AS1* is highlighted in red. C) Expanded view of the genomic region surrounding the 7q21.3 breakpoints in DGAP191. The directly disrupted lncRNA *ENSG00000285090* is highlighted in red.

Figure 4. A) Chromosome diagrams depict the inversion between 5p14.2 and 5q14.3 in DGAP218. Above, TADs containing the breakpoints are shown, with the breakpoint positions indicated by vertical yellow bars. Protein-coding genes are shown in blue and non-coding genes in green, with a single isoform depicted per gene. TAD borders were defined in (Dixon et al. 2012).

Triangular contact maps display micro-C data from (Krietenstein et al. 2020). B) Expanded view of the genomic region surrounding the 5p14.2 breakpoints in DGAP218. C) Expanded view of the genomic region surrounding the 5q14.3 breakpoints in DGAP218. The directly disrupted lncRNA *MEF2C-ASI* is highlighted in red.

Figure 5. A) Chromosome diagrams depict the translocation between 3p22.2 and 14q12 in DGAP245. Above, large regions containing the breakpoints are shown, with the breakpoint positions indicated by vertical yellow bars. The region shown surrounding 14q12 is a TAD, with its borders previously defined in (Dixon et al. 2012). No TAD was defined surrounding 3p22.2, so instead the region including 1Mb on either side of the breakpoints is displayed. Protein-coding genes are shown in blue and non-coding genes in green, with a single isoform depicted per gene. Triangular contact maps display micro-C data from (Krietenstein et al. 2020). B) Expanded view of the genomic region surrounding the 3p22.2 breakpoints in DGAP245. C) Expanded view of the genomic region surrounding the 14q12 breakpoints in DGAP245. The directly disrupted lncRNAs *ENSG00000258028* and *ENSG00000257522* are highlighted in red.

Figure 6. A) Chromosome diagrams depict the translocation between 8q21.13 and 14q12 in NIJ1. Above, TADs containing the breakpoints are shown, with the breakpoint positions indicated by vertical yellow bars. Protein-coding genes are shown in blue and non-coding genes in green, with a single isoform depicted per gene. TAD borders were defined in (Dixon et al. 2012). Triangular contact maps display micro-C data from (Krietenstein et al. 2020). B) Expanded view of the genomic region surrounding the 8q21.13 breakpoints in NIJ1. C) Expanded view of the genomic region surrounding the 14q12 breakpoints in NIJ1. The directly disrupted lncRNAs *ENSG00000258028* and *ENSG00000257522* are highlighted in red.

Figure 7. A) Expression of the lncRNA *ENSG00000257522* from the GTEx database (Lonsdale et al. 2013). B) Expression of the protein-coding gene *FOXG1* from the GTEx database (Lonsdale et al. 2013). C) Expanded view of the genomic region surrounding the 14q12 breakpoints in DGAP246, DGAP245, and NIJ1. Breakpoint positions are indicated by vertical yellow bars. The layered H3K4Me1 and H3K27Ac tracks show data from the Bernstein Lab at the Broad Institute. Genomic regions with experimentally validated enhancer activity (“VISTA enhancers”) are shown in red (Visel et al. 2007).

Table 1. Details regarding the breakpoints and the disrupted genes for the cases highlighted in this manuscript. Genomic coordinates refer to GRCh38/hg38.

Supplemental figure and table legends

Figure S1. A) Axial CT images of the right and left temporal bones of a 40-45 year old female, the mother of DGAP353. While the inferior basal turns appeared normal (not shown), the upper basal and middle turns of each cochlea appear flattened (wide arrow in image 1). The round windows appear mildly narrow. The right cochlear aperture (short line in image 3) measured 1.5 mm TR and the left measured 1.3 mm TR. There is variant anatomy of the internal auditory canals which appear mildly flared on axial images at the level of the porus acusticus, however normal in the coronal plane. The sinus tympani (posteromedial recess of the tympanic cavity) is unusually small bilaterally (short arrow in image 3). B) Reformatted coronal CT images of the right and left temporal bones of the mother of DGAP353. The right oval window is mildly narrow in height (long arrow in image 1). The round window is also narrowed (short arrow in image 2). Note the mildly small upper cochlear turns (arrowhead in image 3). The left oval window is also mildly narrowed and opacified. The inferior osseous margin of the tympanic segment of the facial nerve

canal is not clearly seen, raising concern for dehiscence at the level of the stenotic oval window (arrow in image 4). The subjacent round window appears normal in the coronal plane. C) Axial CT images of the temporal bones of a 10-14 year old female, DGAP353. Imaging reveals normal upper cochlear turns. The cochlear aperture measures 1.6 mm on each side. The sinus tympani is unusually small bilaterally (short arrows). Note the normal right stapedial crura (long arrow in image 1). The left stapedial crura are closely approximated and indistinct (long arrow in image 2). D) Reformatted coronal images of the temporal bones of DGAP353. These images reveal that the left oval window (arrow in image 1) and round window (arrow in image 2) are stenotic. The right sided oval and round windows were also slightly narrow (not shown).

Figure S2. A) Chromosome diagrams depict the translocation between Xp11.4 and 11q24.2 in DGAP148. Above, TADs containing the breakpoints are shown, with the breakpoint positions indicated by vertical yellow bars. Protein-coding genes are shown in blue and non-coding genes in green, with a single isoform depicted per gene. TAD borders were defined in (Dixon et al. 2012). Triangular contact maps display micro-C data from (Krietenstein et al. 2020). B) Expanded view of the genomic region surrounding the 11q24.2 breakpoints in DGAP148. The directly disrupted lncRNA *ENSG00000255087* is highlighted in red. C) Expression of the lncRNA *ENSG00000255087* from the GTEx database (Lonsdale et al. 2013).

Figure S3. A) Expanded view of the genomic region surrounding the Xp11.4 breakpoints in DGAP148. B) Expression of the protein-coding gene *KIRREL3* from the GTEx database (Lonsdale et al. 2013).

Figure S4. A) Chromosome diagrams depict the translocation between 3q26.33 and 9q21.13 in DGAP355. Above, TADs containing the breakpoints are shown, with the breakpoint positions indicated by vertical yellow bars. Protein-coding genes are shown in blue and non-coding genes

in green, with a single isoform depicted per gene. TAD borders were defined in (Dixon et al. 2012). Triangular contact maps display micro-C data from (Krietenstein et al. 2020). B) Expanded view of the genomic region surrounding the 3q26.33 breakpoints in DGAP355. The directly disrupted lncRNA *SOX2-OT* is highlighted in red. C) Expression of the lncRNA *SOX2-OT* from the GTEx database (Lonsdale et al. 2013).

Figure S5. A) Expanded view of the genomic region surrounding the 9q21.13 breakpoints in DGAP355. B) Expression of the protein-coding gene *SOX2* from the GTEx database (Lonsdale et al. 2013).

Table S1. Genetic and phenotypic details for all cases analyzed as part of this study. Genomic coordinates refer to GRCh38/hg38. Derivative A and Derivative B represent the chromosomal breakpoints listed in the order recommended by Orulu et al. 2014.

Table S2. Details regarding the breakpoints and the directly disrupted genes for all 66 cases in which we identified a disrupted lncRNA. The first tab lists cases in which only lncRNAs are directly disrupted. The second tab lists cases in which lncRNAs are directly disrupted along with other genes. Genomic coordinates refer to GRCh38/hg38. The disruption of the lncRNA *RMST* in DGAP032 was previously reported in (Stamou et al. 2020). The disruption of the lncRNA *LINC00299* in DGAP162 was previously reported in (Talkowski et al. 2012a).

Table S3. Additional details for the cases in which only lncRNAs were directly disrupted. The first tab lists the nearest protein-coding gene to each disrupted lncRNA. The second tab lists all genes of any class within 100kb of the breakpoints, excluding the lncRNAs that are directly disrupted (see Table S2 for directly disrupted lncRNAs). Genomic coordinates refer to GRCh38/hg38.

References

- Abi Habib W, Brioude F, Edouard T, et al (2018) Genetic disruption of the oncogenic HMGA2-PLAG1-IGF2 pathway causes fetal growth restriction. *Genet Med* 20:250–258. <https://doi.org/10.1038/gim.2017.105>
- Alford RL, Arnos KS, Fox M, et al (2014) American College of Medical Genetics and Genomics guideline for the clinical evaluation and etiologic diagnosis of hearing loss. *Genet Med* 16:347–55. <https://doi.org/10.1038/gim.2014.2>
- Allou L, Balzano S, Magg A, et al (2021) Non-coding deletions identify Maenli lncRNA as a limb-specific En1 regulator. *Nature* 592:93–98. <https://doi.org/10.1038/s41586-021-03208-9>
- Alyaqoub F, Pyatt RE, Bailes A, et al (2012) 12q14 microdeletion associated with HMGA2 gene disruption and growth restriction. *Am J Med Genet A* 158A:2925–30. <https://doi.org/10.1002/ajmg.a.35610>
- Amaral PP, Leonardi T, Han N, et al (2018) Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. *Genome Biol* 19:32. <https://doi.org/10.1186/s13059-018-1405-5>
- Anderson KM, Anderson DM, McAnally JR, et al (2016) Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature* 539:433–436. <https://doi.org/10.1038/nature20128>
- Ariani F, Hayek G, Rondinella D, et al (2008) FOXP1 is responsible for the congenital variant of Rett syndrome. *Am J Hum Genet* 83:89–93. <https://doi.org/10.1016/j.ajhg.2008.05.015>
- Arnold K, Sarkar A, Yram MA, et al (2011) Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice. *Cell Stem Cell* 9:317–29. <https://doi.org/10.1016/j.stem.2011.09.001>
- Ashar HR, Fejzo MS, Tkachenko A, et al (1995) Disruption of the architectural factor HMGI-C: DNA-binding AT hook motifs fused in lipomas to distinct transcriptional regulatory domains. *Cell* 82:57–65. [https://doi.org/10.1016/0092-8674\(95\)90052-7](https://doi.org/10.1016/0092-8674(95)90052-7)
- Avraham KB, Khalaily L, Noy Y, et al (2022) The noncoding genome and hearing loss. *Hum Genet* 141:323–333. <https://doi.org/10.1007/s00439-021-02359-z>
- Azaiez H, Booth KT, Ephraim SS, et al (2018) Genomic Landscape and Mutational Signatures of Deafness-Associated Genes. *Am J Hum Genet* 103:484–497. <https://doi.org/10.1016/j.ajhg.2018.08.006>
- Bae J, Choi HS, Park SY, et al (2018) Novel Mutation in PTHLH Related to Brachydactyly Type E2 Initially Confused with Unclassical Pseudopseudohypoparathyroidism. *Endocrinol Metab (Seoul)* 33:252–259. <https://doi.org/10.3803/EnM.2018.33.2.252>
- Ballif BC, Theisen A, Rosenfeld JA, et al (2010) Identification of a recurrent microdeletion at 17q23.1q23.2 flanked by segmental duplications associated with heart defects and limb abnormalities. *Am J Hum Genet* 86:454–61. <https://doi.org/10.1016/j.ajhg.2010.01.038>
- Battista S, Fidanza V, Fedele M, et al (1999) The expression of a truncated HMGI-C gene induces gigantism associated with lipomatosis. *Cancer Res* 59:4793–7
- Ciaccio C, Leonardi E, Polli R, et al (2021) A Missense De Novo Variant in the CASK-interactor KIRREL3 Gene Leading to Neurodevelopmental Disorder with Mild Cerebellar Hypoplasia. *Neuropediatrics* 52:484–488. <https://doi.org/10.1055/s-0041-1725964>
- Collins RL, Glessner JT, Porcu E, et al (2022) A cross-disorder dosage sensitivity map of the human genome. *Cell* 185:3041–3055.e25. <https://doi.org/10.1016/j.cell.2022.06.036>

- Decaestecker B, Denecker G, Van Neste C, et al (2018) TBX2 is a neuroblastoma core regulatory circuitry component enhancing MYCN/FOXM1 reactivation of DREAM targets. *Nat Commun* 9:4866. <https://doi.org/10.1038/s41467-018-06699-9>
- D'haene E, Bar-Yaacov R, Bariah I, et al (2019) A neuronal enhancer network upstream of MEF2C is compromised in patients with Rett-like characteristics. *Hum Mol Genet* 28:818–827. <https://doi.org/10.1093/hmg/ddy393>
- Dixon JR, Selvaraj S, Yue F, et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–80. <https://doi.org/10.1038/nature11082>
- ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74. <https://doi.org/10.1038/nature11247>
- Engreitz JM, Haines JE, Perez EM, et al (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539:452–455
- Fedele M, Berlingieri MT, Scala S, et al (1998) Truncated and chimeric HMGI-C genes induce neoplastic transformation of NIH3T3 murine fibroblasts. *Oncogene* 17:413–8. <https://doi.org/10.1038/sj.onc.1201952>
- Ferrer J, Dimitrova N (2024) Transcription regulation by long non-coding RNAs: mechanisms and disease relevance. *Nat Rev Mol Cell Biol*. <https://doi.org/10.1038/s41580-023-00694-9>
- Feyder M, Goff LA (2016) Investigating long noncoding RNAs using animal models. *J Clin Invest* 126:2783–91. <https://doi.org/10.1172/JCI84422>
- Firth H V, Richards SM, Bevan AP, et al (2009) DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* 84:524–33. <https://doi.org/10.1016/j.ajhg.2009.03.010>
- Frankish A, Diekhans M, Jungreis I, et al (2021) GENCODE 2021. *Nucleic Acids Res* 49:D916–D923. <https://doi.org/10.1093/nar/gkaa1087>
- Ganesh VS, Riquin K, Chatron N, et al (2024) Novel syndromic neurodevelopmental disorder caused by de novo deletion of CHASERR, a long noncoding RNA. *medRxiv*. <https://doi.org/10.1101/2024.01.31.24301497>
- García-Añoveros J, Clancy JC, Foo CZ, et al (2022) Tbx2 is a master regulator of inner versus outer hair cell differentiation. *Nature* 605:298–303. <https://doi.org/10.1038/s41586-022-04668-3>
- Graham V, Khudyakov J, Ellis P, Pevny L (2003) SOX2 functions to maintain neural progenitor identity. *Neuron* 39:749–65. [https://doi.org/10.1016/s0896-6273\(03\)00497-5](https://doi.org/10.1016/s0896-6273(03)00497-5)
- Guo Q, Zhang L, Zhao L, et al (2022) MEF2C-AS1 regulates its nearby gene MEF2C to mediate cervical cancer cell malignant phenotypes in vitro. *Biochem Biophys Res Commun* 632:48–54. <https://doi.org/10.1016/j.bbrc.2022.09.091>
- Hanscom C, Talkowski M (2014) Design of large-insert jumping libraries for structural variant detection using Illumina sequencing. *Curr Protoc Hum Genet* 80:7.22.1-7.22.9. <https://doi.org/10.1002/0471142905.hg0722s80>
- Higgins AW, Alkuraya FS, Bosco AF, et al (2008) Characterization of apparently balanced chromosomal rearrangements from the developmental genome anatomy project. *Am J Hum Genet* 82:712–22. <https://doi.org/10.1016/j.ajhg.2008.01.011>
- Kaiser M, Wojahn I, Rudat C, et al (2021) Regulation of otocyst patterning by Tbx2 and Tbx3 is required for inner ear morphogenesis in the mouse. *Development* 148:. <https://doi.org/10.1242/dev.195651>

- Karczewski KJ, Francioli LC, Tiao G, et al (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581:434–443. <https://doi.org/10.1038/s41586-020-2308-7>
- Kazmierczak B, Bullerdiek J, Pham KH, et al (1998) Intron 3 of HMGIC is the most frequent target of chromosomal aberrations in human tumors and has been conserved basically for at least 30 million years. *Cancer Genet Cytogenet* 103:175–7. [https://doi.org/10.1016/s0165-4608\(97\)00348-8](https://doi.org/10.1016/s0165-4608(97)00348-8)
- Kent WJ (2002) BLAT--the BLAST-like alignment tool. *Genome Res* 12:656–64. <https://doi.org/10.1101/gr.229202>
- Kent WJ, Sugnet CW, Furey TS, et al (2002) The human genome browser at UCSC. *Genome Res* 12:996–1006. <https://doi.org/10.1101/gr.229102>
- Klopocki E, Hennig BP, Dathe K, et al (2010) Deletion and point mutations of PTHLH cause brachydactyly type E. *Am J Hum Genet* 86:434–9. <https://doi.org/10.1016/j.ajhg.2010.01.023>
- Knauss JL, Miao N, Kim S-N, et al (2018) Long noncoding RNA Sox2ot and transcription factor YY1 co-regulate the differentiation of cortical neural progenitors by repressing Sox2. *Cell Death Dis* 9:799. <https://doi.org/10.1038/s41419-018-0840-2>
- Kopp F, Mendell JT (2018) Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* 172:393–407. <https://doi.org/10.1016/j.cell.2018.01.011>
- Kortüm F, Das S, Flindt M, et al (2011) The core FOXG1 syndrome phenotype consists of postnatal microcephaly, severe mental retardation, absent language, dyskinesia, and corpus callosum hypogenesis. *J Med Genet* 48:396–406. <https://doi.org/10.1136/jmg.2010.087528>
- Krietenstein N, Abraham S, Venev S V, et al (2020) Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell* 78:554-565.e7. <https://doi.org/10.1016/j.molcel.2020.03.003>
- Lezirovitz K, Vieira-Silva GA, Batissoco AC, et al (2020) A rare genomic duplication in 2p14 underlies autosomal dominant hearing loss DFNA58. *Hum Mol Genet* 29:1520–1536. <https://doi.org/10.1093/hmg/ddaa075>
- Li P-Y, Wang P, Gao S-G, Dong D-Y (2020) Long Noncoding RNA SOX2-OT: Regulations, Functions, and Roles on Mental Illnesses, Cancers, and Diabetic Complications. *Biomed Res Int* 2020:2901589. <https://doi.org/10.1155/2020/2901589>
- Ligon AH, Moore SDP, Parisi MA, et al (2005) Constitutional rearrangement of the architectural factor HMGA2: a novel human phenotype including overgrowth and lipomas. *Am J Hum Genet* 76:340–8. <https://doi.org/10.1086/427565>
- Liu H, Chen L, Giffen KP, et al (2018) Cell-Specific Transcriptome Analysis Shows That Adult Pillar and Deiters' Cells Express Genes Encoding Machinery for Specializations of Cochlear Hair Cells. *Front Mol Neurosci* 11:356. <https://doi.org/10.3389/fnmol.2018.00356>
- Lonsdale J, Thomas J, Salvatore M, et al (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45:580–585. <https://doi.org/10.1038/ng.2653>
- Lowther C, Mehrjouy MM, Collins RL, et al (2022) Balanced chromosomal rearrangements offer insights into coding and noncoding genomic features associated with developmental disorders. *medRxiv* 2022.02.15.22270795. <https://doi.org/10.1101/2022.02.15.22270795>
- Luo S, Lu JY, Liu L, et al (2016) Divergent lncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells. *Cell Stem Cell* 18:637–652. <https://doi.org/10.1016/j.stem.2016.01.024>

- Lupiáñez DG, Kraft K, Heinrich V, et al (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161:1012–1025. <https://doi.org/10.1016/j.cell.2015.04.004>
- Lynch SA, Foulds N, Thureson A-C, et al (2011) The 12q14 microdeletion syndrome: six new cases confirming the role of HMGA2 in growth. *Eur J Hum Genet* 19:534–9. <https://doi.org/10.1038/ejhg.2010.215>
- Mattick JS, Amaral PP, Carninci P, et al (2023) Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol* 24:430–447. <https://doi.org/10.1038/s41580-022-00566-8>
- Mehrjouy MM, Fonseca ACS, Ehmke N, et al (2018) Regulatory variants of FOXP1 in the context of its topological domain organisation. *Eur J Hum Genet* 26:186–196. <https://doi.org/10.1038/s41431-017-0011-4>
- Modi A, Lopez G, Conkrite KL, et al (2023) Integrative Genomic Analyses Identify LncRNA Regulatory Networks across Pediatric Leukemias and Solid Tumors. *Cancer Res* 83:3462–3477. <https://doi.org/10.1158/0008-5472.CAN-22-3186>
- Mohajeri K, Yadav R, D’haene E, et al (2022) Transcriptional and functional consequences of alterations to MEF2C and its topological organization in neuronal models. *Am J Hum Genet* 109:2049–2067. <https://doi.org/10.1016/j.ajhg.2022.09.015>
- Nascimento LR do, Vieira-Silva GA, Kitajima JFW, et al (2022) New Insights into the Identity of the DFNA58 Gene. *Genes (Basel)* 13:. <https://doi.org/10.3390/genes13122274>
- Nimmakayalu M, Major H, Sheffield V, et al (2011) Microdeletion of 17q22q23.2 encompassing TBX2 and TBX4 in a patient with congenital microcephaly, thyroid duct cyst, sensorineural hearing loss, and pulmonary hypertension. *Am J Med Genet A* 155A:418–23. <https://doi.org/10.1002/ajmg.a.33827>
- Ordulu Z, Wong KE, Currall BB, et al (2014) Describing sequencing results of structural chromosome rearrangements with a suggested next-generation cytogenetic nomenclature. *Am J Hum Genet* 94:695–709. <https://doi.org/10.1016/j.ajhg.2014.03.020>
- Orvis J, Gottfried B, Kancherla J, et al (2021) gEAR: Gene Expression Analysis Resource portal for community-driven, multi-omic data exploration. *Nat Methods* 18:843–844. <https://doi.org/10.1038/s41592-021-01200-9>
- Pais LS, Snow H, Weisburd B, et al (2022) seqr: A web-based analysis and collaboration tool for rare disease genomics. *Hum Mutat* 43:698–707. <https://doi.org/10.1002/humu.24366>
- Pavlaki I, Alammari F, Sun B, et al (2018) The long non-coding RNA *Paupar* promotes KAP1-dependent chromatin changes and regulates olfactory bulb neurogenesis. *EMBO J* 37:e98219. <https://doi.org/10.15252/embj.201798219>
- Querzani A, Sirchia F, Rustioni G, et al (2023) KIRREL3-related disorders: a case report confirming the radiological features and expanding the clinical spectrum to a less severe phenotype. *Ital J Pediatr* 49:99. <https://doi.org/10.1186/s13052-023-01488-7>
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–2. <https://doi.org/10.1093/bioinformatics/btq033>
- Rapicavoli NA, Poth EM, Zhu H, Blackshaw S (2011) The long noncoding RNA Six3OS acts in trans to regulate retinal development by modulating Six3 activity. *Neural Dev* 6:32. <https://doi.org/10.1186/1749-8104-6-32>
- Redin C, Brand H, Collins RL, et al (2017) The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat Genet* 49:36–45. <https://doi.org/10.1038/ng.3720>

- Rehm HL, Berg JS, Brooks LD, et al (2015) ClinGen--the Clinical Genome Resource. *N Engl J Med* 372:2235–42. <https://doi.org/10.1056/NEJMSr1406261>
- Reyes M, Bravenboer B, Jüppner H (2019) A Heterozygous Splice-Site Mutation in PTHLH Causes Autosomal Dominant Shortening of Metacarpals and Metatarsals. *J Bone Miner Res* 34:482–489. <https://doi.org/10.1002/jbmr.3628>
- Rom A, Melamed L, Gil N, et al (2019) Regulation of CHD2 expression by the Chaserr long noncoding RNA gene is essential for viability. *Nat Commun* 10:. <https://doi.org/10.1038/s41467-019-13075-8>
- Sauvageau M, Goff LA, Lodato S, et al (2013) Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* 2:e01749. <https://doi.org/10.7554/eLife.01749>
- Schoenmakers EF, Wanschura S, Mols R, et al (1995) Recurrent rearrangements in the high mobility group protein gene, HMGI-C, in benign mesenchymal tumours. *Nat Genet* 10:436–44. <https://doi.org/10.1038/ng0895-436>
- Schönewolf-Greulich B, Ronan A, Ravn K, et al (2011) Two new cases with microdeletion of 17q23.2 suggest presence of a candidate gene for sensorineural hearing loss within this region. *Am J Med Genet A* 155A:2964–9. <https://doi.org/10.1002/ajmg.a.34302>
- Shahryari A, Jazi MS, Samaei NM, Mowla SJ (2015) Long non-coding RNA SOX2OT: expression signature, splicing patterns, and emerging roles in pluripotency and tumorigenesis. *Front Genet* 6:196. <https://doi.org/10.3389/fgene.2015.00196>
- Sheffield AM, Smith RJH (2019) The Epidemiology of Deafness. *Cold Spring Harb Perspect Med* 9:. <https://doi.org/10.1101/cshperspect.a033258>
- Stamou M, Ng S-Y, Brand H, et al (2020) A Balanced Translocation in Kallmann Syndrome Implicates a Long Noncoding RNA, RMST, as a GnRH Neuronal Regulator. *J Clin Endocrinol Metab* 105:e231-44. <https://doi.org/10.1210/clinem/dgz011>
- Statello L, Guo C-J, Chen L-L, Huarte M (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* 22:96–118. <https://doi.org/10.1038/s41580-020-00315-9>
- Stewart GL, Sage AP, Enfield KSS, et al (2020) Deregulation of a Cis-Acting lncRNA in Non-small Cell Lung Cancer May Control HMGA1 Expression. *Front Genet* 11:615378. <https://doi.org/10.3389/fgene.2020.615378>
- Talkowski ME, Ernst C, Heilbut A, et al (2011) Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am J Hum Genet* 88:469–81. <https://doi.org/10.1016/j.ajhg.2011.03.013>
- Talkowski ME, Maussion G, Crapper L, et al (2012a) Disruption of a large intergenic noncoding RNA in subjects with neurodevelopmental disabilities. *The American Journal of Human Genetics* 91:1128–1134. <https://doi.org/10.1016/j.ajhg.2012.10.016>
- Talkowski ME, Rosenfeld JA, Blumenthal I, et al (2012b) Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* 149:525–37. <https://doi.org/10.1016/j.cell.2012.03.028>
- Taniue K, Akimitsu N (2021) The Functions and Unique Features of LncRNAs in Cancer Development and Tumorigenesis. *Int J Mol Sci* 22:. <https://doi.org/10.3390/ijms22020632>
- Ushakov K, Koffler-Brill T, Rom A, et al (2017) Genome-wide identification and expression profiling of long non-coding RNAs in auditory and vestibular systems. *Sci Rep* 7:8637. <https://doi.org/10.1038/s41598-017-08320-3>

- Vance KW, Sansom SN, Lee S, et al (2014) The long non-coding RNA Paupar regulates the expression of both local and distal genes. *EMBO J* 33:296–311. <https://doi.org/10.1002/embj.201386225>
- Visel A, Minovitsky S, Dubchak I, Pennacchio LA (2007) VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* 35:D88-92. <https://doi.org/10.1093/nar/gkl822>
- Volders P-J, Anckaert J, Verheggen K, et al (2019) LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res* 47:D135–D139. <https://doi.org/10.1093/nar/gky1031>
- Wang Y, Chen S, Li W, et al (2020) Associating divergent lncRNAs with target genes by integrating genome sequence, gene expression and chromatin accessibility data. *NAR Genom Bioinform* 2:lqaa019. <https://doi.org/10.1093/nargab/lqaa019>
- Wansleben S, Peres J, Hare S, et al (2014) T-box transcription factors in cancer biology. *Biochim Biophys Acta* 1846:380–91. <https://doi.org/10.1016/j.bbcan.2014.08.004>
- Yin J, Shen Y, Si Y, et al (2020) Knockdown of long non-coding RNA SOX2OT downregulates SOX2 to improve hippocampal neurogenesis and cognitive function in a mouse model of sepsis-associated encephalopathy. *J Neuroinflammation* 17:320. <https://doi.org/10.1186/s12974-020-01970-7>
- Zaidi MR, Okada Y, Chada KK (2006) Misexpression of full-length HMGA2 induces benign mesenchymal tumors in mice. *Cancer Res* 66:7453–9. <https://doi.org/10.1158/0008-5472.CAN-06-0931>
- Zhou X, Benson KF, Ashar HR, Chada K (1995) Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. *Nature* 376:771–4. <https://doi.org/10.1038/376771a0>
- Zhou X, Chada K (1998) HMGI family proteins: architectural transcription factors in mammalian development and cancer. *Keio J Med* 47:73–7. <https://doi.org/10.2302/kjm.47.73>
- Zweier M, Rauch A (2012) The MEF2C-Related and 5q14.3q15 Microdeletion Syndrome. *Mol Syndromol* 2:164–170. <https://doi.org/10.1159/000337496>