

## Disruption of a Large Intergenic Noncoding RNA in Subjects with Neurodevelopmental Disabilities

Michael E. Talkowski,<sup>1,2,3</sup> Gilles Maussion,<sup>4</sup> Liam Crapper,<sup>4</sup> Jill A. Rosenfeld,<sup>5</sup> Ian Blumenthal,<sup>1</sup> Carrie Hanscom,<sup>1</sup> Colby Chiang,<sup>1</sup> Amelia Lindgren,<sup>6</sup> Shahrin Pereira,<sup>6</sup> Douglas Ruderfer,<sup>7</sup> Alpha B. Diallo,<sup>4</sup> Juan Pablo Lopez,<sup>4</sup> Gustavo Turecki,<sup>4</sup> Elizabeth S. Chen,<sup>4</sup> Carolina Giguek,<sup>4</sup> David J. Harris,<sup>8</sup> Va Lip,<sup>8</sup> Yu An,<sup>8,9</sup> Marta Biagioli,<sup>1,2</sup> Marcy E. MacDonald,<sup>1,2,3</sup> Michael Lin,<sup>3,10</sup> Stephen J. Haggarty,<sup>1,2</sup> Pamela Sklar,<sup>7</sup> Shaun Purcell,<sup>1,3,7</sup> Manolis Kellis,<sup>3,10</sup> Stuart Schwartz,<sup>11</sup> Lisa G. Shaffer,<sup>5</sup> Marvin R. Natowicz,<sup>12</sup> Yiping Shen,<sup>1,8,13</sup> Cynthia C. Morton,<sup>3,6</sup> James F. Gusella,<sup>1,2,3</sup> and Carl Ernst<sup>4,\*</sup>

Large intergenic noncoding (linc) RNAs represent a newly described class of ribonucleic acid whose importance in human disease remains undefined. We identified a severely developmentally delayed 16-year-old female with karyotype 46,XX,t(2;11)(p25.1;p15.1)dn in the absence of clinically significant copy number variants (CNVs). DNA capture followed by next-generation sequencing of the translocation breakpoints revealed disruption of a single noncoding gene on chromosome 2, *LINC00299*, whose RNA product is expressed in all tissues measured, but most abundantly in brain. Among a series of additional, unrelated subjects referred for clinical diagnostic testing who showed CNV affecting this locus, we identified four with exon-crossing deletions in association with neurodevelopmental abnormalities. No disruption of the *LINC00299* coding sequence was seen in almost 14,000 control subjects. Together, these subjects with disruption of *LINC00299* implicate this particular noncoding RNA in brain development and raise the possibility that, as a class, abnormalities of lincRNAs may play a significant role in human developmental disorders.

Large intergenic noncoding RNAs (lincRNAs), thought to number more than 4,000 in the human genome, have recently been identified as a new class of RNA molecule postulated to play important roles in gene regulation.<sup>1,2</sup> Those lincRNAs that have been characterized functionally exhibit diverse biological activities including involvement in X chromosome inactivation and regulation of gene expression in stem cells, cancer cells, and development.<sup>3–6</sup> It has been estimated that more than 30% of lincRNAs associate with chromatin-modifying complexes, such as PRC2 and co-REST, and subsequently target these complexes to specific genomic regions.<sup>1</sup> Evidence for lincRNA function comes from studies of *HOTAIR*, a lincRNA that is part of the Hox gene cluster, in breast cancer progression,<sup>7</sup> and *ANRIL*, a lincRNA involved in regulation of *CDKN2A* and *CDKN2B* and associated with atherosclerosis.<sup>8</sup> Still, despite their potentially broad functional impact in development, no lincRNAs have been implicated in human developmental abnormalities. Here, we identify a lincRNA specifically disrupted in independent subjects with developmental disabilities.

Arguably, the discovery of direct disruption of a single locus in the genome from a de novo balanced alteration is among the most powerful approaches in human genetics for isolating individual genomic loci directly associated with the phenotype. In this study, we describe a 16-year-old female (DGAP162) with developmental delay from the Developmental Genome Anatomy Project (DGAP)<sup>9</sup> whose clinical karyotype revealed a de novo balanced translocation: 46,XX,t(2;11)(p25.1;p15.1)dn. Array comparative genomic hybridization via 1 million probes (Agilent G3 microarray, Agilent, Santa Clara, CA) revealed no genomic deletions or duplications clinically interpreted as pathogenic. To rule out the possibility that a known pathogenic mutation involved in neurodevelopment was contributing to the phenotype of DGAP162, we sequenced genes *CDKLS*, *FOXG1*, *MECP2*, *SLC9A6*, and *TCF4* in the diagnostic clinical laboratory at Boston Children's Hospital, revealing only common polymorphisms in this case. No novel or known pathogenic mutations were discovered.

At birth, DGAP162 was delivered normally, weighing 2.7 kg (<5<sup>th</sup> percentile), after a 37.5 week pregnancy

<sup>1</sup>Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; <sup>2</sup>Department of Neurology and Genetics, Harvard Medical School, Harvard University, Boston, MA 02114, USA; <sup>3</sup>Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02114, USA; <sup>4</sup>Department of Psychiatry, McGill University, Douglas Hospital Research Institute, Montreal, Quebec, ON H4H 1R3, Canada; <sup>5</sup>Signature Genomics, Inc., PerkinElmer, Spokane, WA 99207, USA; <sup>6</sup>Departments of Obstetrics, Gynecology and Reproductive Biology and of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; <sup>7</sup>Departments of Psychiatry and Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY 10029, USA; <sup>8</sup>Department of Laboratory Medicine, Boston Children's Hospital and Department of Pathology, Harvard Medical School, Boston, MA 02115, USA; <sup>9</sup>Institutes of Biomedical Sciences and MOE Key Laboratory of Contemporary Anthropology, Fudan University, Shanghai 200433, China; <sup>10</sup>Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02114, USA; <sup>11</sup>Genzyme Genetics, Inc., Labcorp, Research Triangle Park, NC 27709, USA; <sup>12</sup>Institutes of Pathology and Laboratory Medicine, Genomic Medicine, Neurology and Pediatrics, Cleveland Clinic, Cleveland, OH 44195, USA; <sup>13</sup>Shanghai Children's Medical Center, Shanghai Jiaotong University School of Medicine, Shanghai 200127, China

\*Correspondence: [carl.ernst@mcgill.ca](mailto:carl.ernst@mcgill.ca)

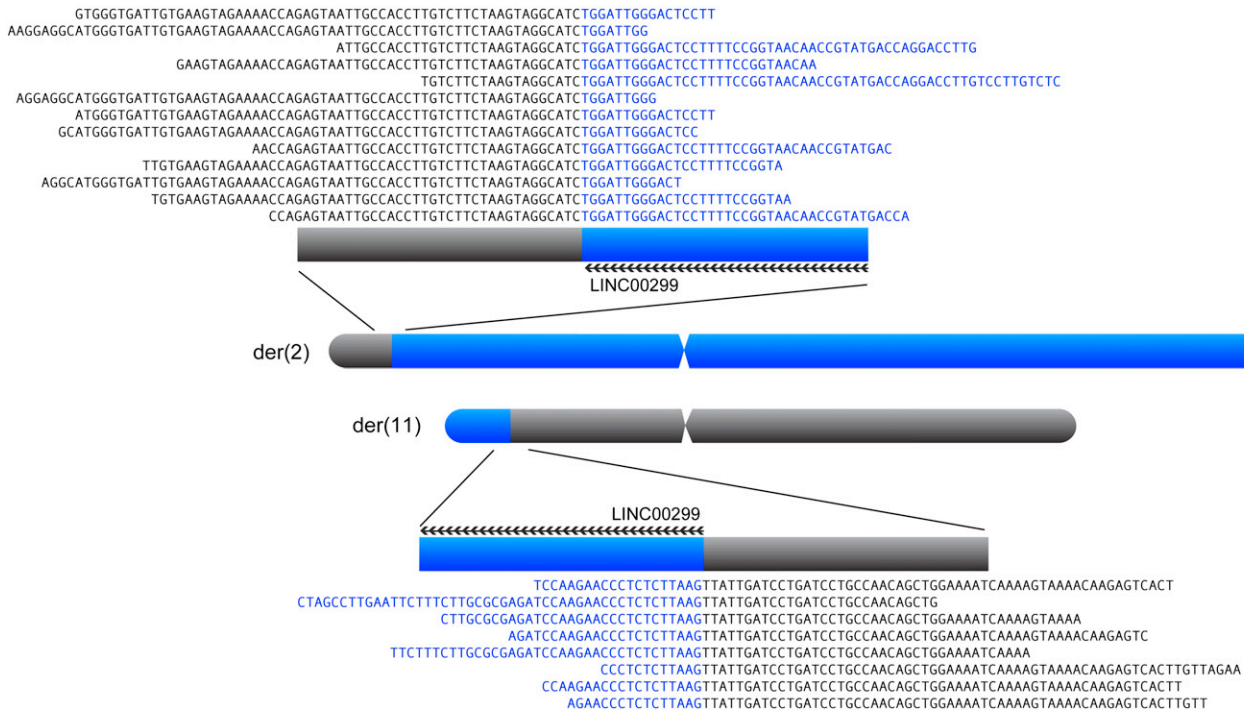
<http://dx.doi.org/10.1016/j.ajhg.2012.10.016>. ©2012 by The American Society of Human Genetics. All rights reserved.

notable for oligohydramnios and dilated fetal kidneys at 34 weeks. Her immediate postnatal period was remarkable for poor feeding but all scans were normal and she was discharged at 10 days. Her parents suspected some developmental issue at the age of 9 months because of her failure to sit on her own and she was referred for developmental delay at 12 months. At 22 months, DGAP162 was a friendly, happy, and very sociable girl with no dysmorphic features but who ate only soft food. Though Angelman syndrome was suspected, diagnostic tests were negative. Her workup also included serum amino acid levels, thyroid function, and blood count, all of which were normal. However, an MRI revealed slightly delayed myelination and partial agenesis of the corpus callosum. At 30 months, she was thriving with a weight of 11.4 kg (5<sup>th</sup>–10<sup>th</sup> percentile) and height of 84.6 cm (5<sup>th</sup> percentile), though her head circumference was noted at the 10<sup>th</sup> percentile and her developmental difficulties were worsening. A urine amino acid screen was unremarkable. By 3.5–4 years of age, she had only begun to walk and subsequently required a walker until age 7. She developed episodes of unresponsiveness/stiffness; an EEG was abnormal and she was started on, and continues to take, sodium valproate. She suffered from recurrent ear infections/conductive hearing loss, which necessitated adenoidectomy and insertion of grommets. At age 6.5 years, her weight and height were 20 kg (25<sup>th</sup> percentile) and 118 cm (50<sup>th</sup> percentile), respectively. She did not chew her food (instead she swallowed it whole) and often put hard objects in her mouth. Sensory issues were noted, along with prominent tip-toe walking, though there were no motor problems. At age 7, she was in a special school but showed no progress in learning, self-help skills, or development. She was sleeping 14 hr a day, and she showed reduced attention, though her hearing was normal. She also began to put her hands in her mouth and frequently engaged in self-stimulation. Her EEG remained abnormal in wake and sleep and was interpreted as resulting from a developmental abnormality in the frontotemporal area. By 12 years, she showed normal-set eyes, normal-set ears, no arched palate, and no abnormal hand creases. She had severe learning difficulties and many repetitive sensory behaviors, with a very brief attention span. She was very sociable and engaged well with family and caregivers. Her EEG was abnormal with continued changes in the bilateral frontotemporal region. At 16 years, DGAP162 is very thin with a high palate and is quite mobile despite an abnormal gait, but she requires complete care. She is double incontinent, cannot feed herself, and is nonverbal with only slight grunts and physical action for communication, with a laugh-like squeal when excited. She displays poor eye contact, seems “off in her own world,” and has particular routines and ritual behaviors, including hand-flapping, head-banging, stuffing her hands in her mouth, and masturbation. She has never been formally assessed for autism. DGAP162 has two younger, developmentally normal siblings. Partners HealthCare System institutional

review boards approved subject recruitment procedures and DNA screening and informed consent was obtained from the family.

We performed a DNA capture and sequencing experiment to determine the translocation breakpoints of DGAP162 by using methods (CapBP) that we have described previously.<sup>3</sup> In brief, standard insert paired-end genomic DNA libraries were prepared with NEBnext reagents (New England Biolabs, Ipswich, MA) and Illumina multiplex adapters (Illumina, San Diego, CA). Custom DNA capture arrays (Agilent, Santa Clara, CA) were designed based upon an approximate translocation region defined by FISH via mapped BAC clones. We targeted a 313,002 bp region on chromosome (chr) 11 and a 246,713 bp region on chr 2 over most unique and repeat-masked sequences (see Figure 4b in Talkowski et al.<sup>3</sup>). The captured DGAP162 DNA was sequenced within a pool of DNA from multiple subjects on a single lane of a flow cell (Illumina HiSeq2000). Reads were aligned and analyzed with BWA and samtools.<sup>10,11</sup> Chimeric read pairs were clustered with BamStat, which performs a single linkage clustering of anomalous pairs,<sup>3</sup> identifying the putative translocation cluster. Pile-ups and coverage were compared to expectations with a custom-designed mappability tool to verify the likelihood of good coverage after DNA capture<sup>3</sup> and split reads from translocation junction breakpoints were identified to base-pair resolution (Figure 1), then confirmed by Sanger sequencing.

Sequencing revealed a perfectly balanced translocation breakpoint on chr 2 and a loss of just three bases on chr 11 (hg19, Figure 1; der(2): 11pter–15,825,269::chr2 8,247,757–2qter, der(11): 2pter–8,247,756::chr11 15,825,273–11qter). There were no annotated genes, predicted genes, mRNAs, or expressed sequence tags within 500 kb on either side of the chr 11 breakpoint. However, on chr 2, the rearrangement breakpoint occurred within a large intron of *LINC00299*, directly disrupting this 316.9 kb gene that corresponds to cDNA clones BC043563 and AK127578; the former spans most of *LINC00299* and the latter provides its 5' portion and suggests alternative splicing. *LINC00299* (previously known as *FLJ45673*, *C2orf46*, and *NCRNA00299*), assigned the name “long intergenic nonprotein coding RNA 299” by the HUGO Gene Nomenclature Committee, produces a member of the recently identified class of lincRNAs, based upon its large size, lack of coding potential, and location outside of known genes. Indeed, we screened, in silico, all of the annotated exons for ATG sites preceded by a GCC(A/G)CC Kozak sequence and found no evidence for such a translational start site. Similarly, to confirm whether any DNA in the region spanned by *LINC00299* could give rise to an RNA with coding potential related to any known or suspected protein-coding RNA, we searched the entire genomic sequence with pfam, BlastX, and phyloCSF,<sup>13</sup> all with negative results. Thus, although RNA is both transcribed and processed from this locus, it appears that no known or potential RNA transcript encodes any protein



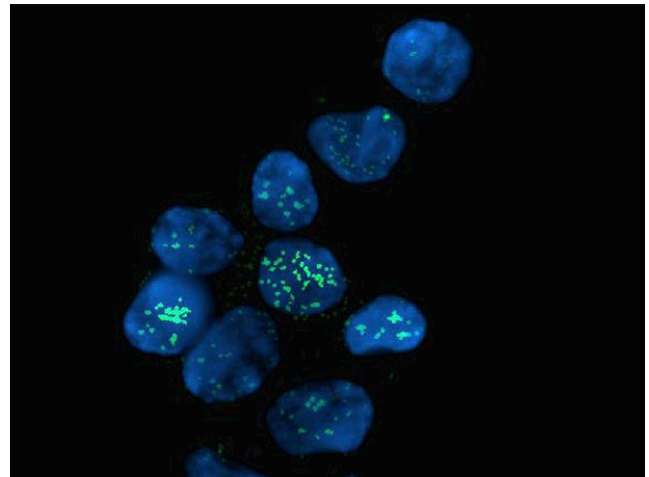
**Figure 1. Identification of Translocation Breakpoints from DGAP162 by CapBP Methodology**

A distal, short arm translocation involving chromosome 2 (blue) and chromosome 11 (gray) are shown, with split read sequences from fragments containing the breakpoint junctions provided. The next-gen cytogenetic karyotype is designated: 46,XX,der(2)(11pter-15,825,269::chr2 8,247,757-2qter),der(11)(2pter-8,247,756::chr11 15,825,273-11qter)dn (hg19). Additional details of sequence characteristics available in Chiang et al.<sup>12</sup>

that can be detected by homology to known or predicted proteins. Finally, to confirm expression and predicted splicing of this gene, we performed an RNA fluorescence in situ hybridization (FISH) experiment by using lymphocytes from a karyotypically normal human subject. FISH probes were made from cDNA and the probe targeted an ~500 bp region covering exons 5–7 (Figures 2 and 3A), suggesting that this gene is expressed and spliced in human lymphocytes.

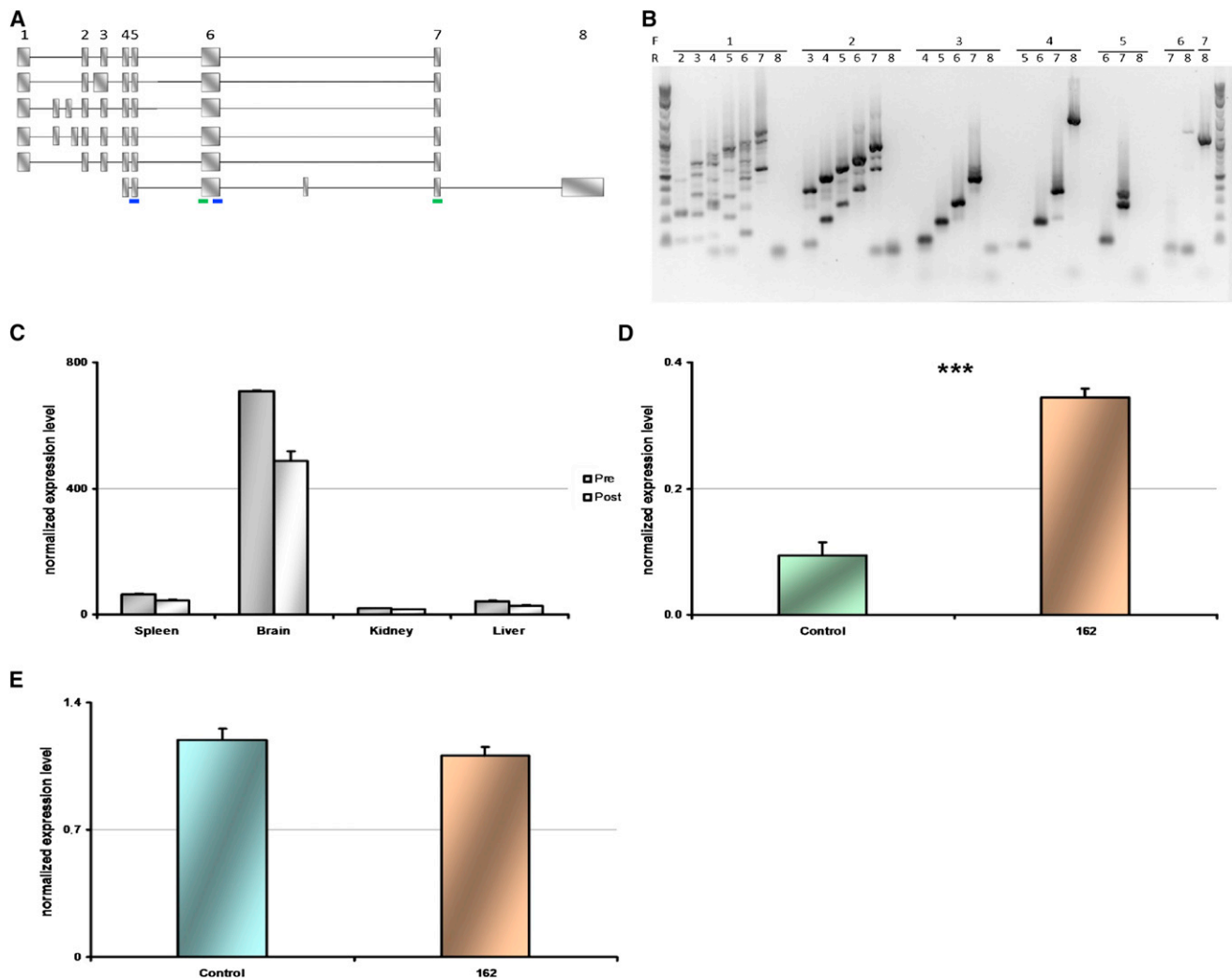
To determine the structure of *LINC00299* RNA product(s), we designed primer pairs to amplify across each exon of cDNA BC043563, numbered as predicted by RefSeq (Figure 3A). Taqman RT-PCR reactions were performed in a total volume of 20  $\mu$ l, on 384-well plates with an Applied Biosystems (Carlsbad, CA) 7900HT and a master mix commercialized by Quanta Biosciences. Serial dilutions provided amounts ranging from 0.04 ng to 10 ng of RNA. For each well, PCR mix included 10  $\mu$ l of 2 $\times$  Perfecta PCR 2 mix (Quanta Biosciences, Gaithersburg, MD), 1  $\mu$ l of primers/probe mix, 2  $\mu$ l of cDNA, H<sub>2</sub>O qsp 20  $\mu$ l. A comparison of amplification product sizes and sequences to those predicted by BC043563 revealed several alternatively spliced transcripts (Figures 3A and 3B). Three exons were identified between BC043563 exons 1 and 2 (at sites, hg19, chr2: 8,452,858–8,452,972, 8,442,650–8,442,949, and 8,442,193–8,442,559) and one was found between exons 6 and 7 (chr2: 8,383,695–8,383,770). We were consistently able to amplify from exon 4 to 8, but never

from upstream exons to exon 8, under a multitude of conditions, suggesting additional cryptic transcript complexity. All amplicons were cloned and Sanger



**Figure 2. *LINC00299* RNA Is Both Expressed and Spliced in Normal Human Lymphocytes**

An RNA FISH probe targeting the spliced product of *LINC00299* in EBV-transformed wild-type human lymphocytes. RNA FISH probe was ligated to an Alexa 488 dye, then hybridized to individual cells on a microscope slide. Individual cells are stained with the DNA dye 4',6-diamidino-2-phenylindole (DAPI). Green dots in the blue nucleus suggest individual molecules of *LINC00299* in each cell. Green stain, probe targeting exons 5–7 of *LINC00299*; blue stain, DAPI.



**Figure 3. Structure of *LINC00299* Transcripts in Brain, Expression Level across Human Tissues, and Expression of *LINC00299* in Lymphoblastoid Cell Lines Comparing DGAP162 to Three Control Subjects**

(A) *LINC00299* alternative splice transcripts from RNA extracted from wild-type human prefrontal cortex (exons and introns not to scale). Numbers represent exons identified by RefSeq, and unreported exons are unnumbered. Each transcript represents a fully cloned RT-PCR fragment, confirmed by Sanger sequencing. Small bars below the final transcript refer to "pre" and "post" primer sets used for expression studies.

(B) Agarose electrophoresis gel showing amplification of *LINC00299* transcripts in brain prefrontal cortex via different primers. Forward primers (F, top line) and reverse primers (R, bottom line) are noted by numbers representing which exon the specific primer targeted. Exon numbers correspond to exon numbers noted in (A).

(C) Expression level of *LINC00299* transcripts in spleen, brain, kidney, and liver. "Pre" and "Post" refer to primers that bind 5' and 3', respectively, to the translocation breakpoint identified in DGAP162.

(D) Quantification of *LINC00299* transcripts in lymphoblastoid cell lines from controls and DGAP162 with a probe targeting exon 6 of *LINC00299* (pretranslocation), which amplified both the wild-type and mutant alleles of DGAP 162.

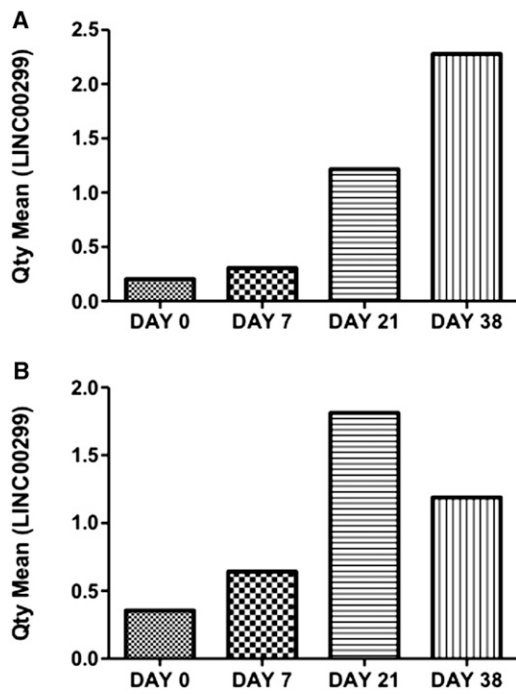
(E) Quantification of *LINC00299* transcripts in lymphoblastoid cells via primers specific to the wild-type allele (posttranslocation probe).

\*\*\*p < 0.001. All reactions were run in quadruplicate and *ACTB* was used as an internal control in all cases. All primer sequences can be found in Table S1. All error bars represent standard error of the mean.

sequenced. RNA sequencing data (not shown) suggests many more transcripts from this locus than identified here, highlighting the complexity of this lincRNA.

We next determined the expression pattern of *LINC00299* in selected human tissues including spleen, brain, kidney, and liver as well as in DGAP162 lymphoblastoid cell lines and three control lymphoblastoid cell lines. In DGAP162, the translocation occurs in an intron immediately upstream of exon 7; thus this translocation results

in a gene lacking the final two exons (exons 7 and 8, Figure 3A). One primer pair, referred to as "pre" (for "pretranslocation") is located in exon 6 of *LINC00299*. The second primer pair, specific to only the wild-type allele in the mutant subject (DGAP162), amplifies from exon 5 to exon 7 (Figure 3A shows the binding sites of these primer pairs on the transcript). In cross-tissue comparisons, both primer sets revealed high expression of *LINC00299* in brain compared to other regions (Figure 3C).



**Figure 4.** *LINC00299* Temporal Expression in Induced Pluripotent Stem Cell-Derived Neural Progenitor Cells

mRNA expression levels were quantified using qRT-PCR and normalized using two endogenous controls, *ACTB* and *GAPDH*.

(A) Quantity mean values represent expression levels of *LINC00299* in mRNA from exon 6.

(B) Quantity mean values represent expression levels of *LINC00299* mRNA from exon 8.

All error bars represent standard error of the mean.

There was a significant increase of *LINC00299* expression in the DGAP162 lymphoblastoid cell line (LCL) relative to controls via the pretranslocation primer set ( $p = 5.3 \times 10^{-4}$ ; Figure 3D), though the overall expression levels were quite low in LCL (see Figure S1 available online for qRT-PCR Ct value graphs). With the wild-type-specific primer set (i.e., the primer set that can amplify only the nonmutant chromosome from DGAP162), the level of expression from the single normal allele of DGAP162 was equivalent to the level of expression from two wild-type alleles in controls ( $p > 0.05$ ; Figure 3E). Quality control for RT-PCR experiments (no reverse transcriptase control experiments) can be found in Figure S1. These data suggest that an attempt at dosage compensation in DGAP162 cells in the presence of heterozygous inactivation results in up-regulation of transcription of both normal and translocated alleles, producing increases in the normal amount of wild-type RNA from the former and producing a transcript from the latter that includes exon 6 but not exons 7 and 8. The latter represents a truncated transcript comprising only sequences from the chr 2 side of the breakpoint, or, alternatively, a fusion transcript that incorporates sequences from the chr 11 side of the breakpoint.

To understand whether *LINC00299* might have a role in neurodevelopment, we used previously described human induced pluripotent stem (iPS) cell-derived neural progen-

itor cells that were developed from a healthy control subject.<sup>14</sup> We differentiated these neural progenitor cells to functional, electrically active neurons over a 38 day period, sampling RNA at distinct time points and measuring *LINC00299* expression. We are able to record stimulus-evoked action potentials from these cells 18 days postdifferentiation (initiated by the removal of FGF2 and EGF), suggesting that over a 38 day time period many neurodevelopmental changes are occurring (data not shown). We measured expression of pre- and post-translocation exons from *LINC00299* in these wild-type cells and found an increase in expression level over time (Figures 4A and 4B).

We tested whether disruption of *LINC00299* also occurred in normal individuals, which would argue against it being a risk factor for abnormal neurodevelopment. We surveyed 13,991 adult controls from a series of genome-wide association studies, as we have previously described (see Talkowski et al.<sup>15</sup>). Among these controls, no individuals harbored a structural variant of any kind that disrupted any exonic region of this locus.

To corroborate the association of *LINC00299* disruption with abnormal neurodevelopment, we also sought to identify other developmentally abnormal subjects where the locus was disrupted, and to compare their phenotypes where available. We analyzed diagnostic array data from Signature Genomics, LabCorp, DECIPHER, and ISCA (International Standards for Cytogenomic Arrays), selecting for subjects that had CNVs that overlapped *LINC00299*, a single CNV <10 Mb, and no other pathogenic CNVs in the genome. We identified four subjects with disruptions at *LINC00299* (Table 1). Notably, all subjects with disruptions over *LINC00299* had developmental delay. Together, subjects with nonrecurrent deletions in this region allow us to narrow a critical region to chr2: 7,554,804–8,945,097, which is less than 1.4 Mb. This region contains only *ID2*, *MBOAT2*, *KIDINS220* 5' from *LINC00299*, and *BC104747* 3' from *LINC00299*. Support for the specific involvement of *LINC00299* in phenotypic expression comes from the translocation subject DGAP162 with a breakpoint directly in *LINC00299* and subject L1 with a 60 kb deletion in the final two exons of *LINC00299*.

In one case (L1), we were able to recontact the physician for additional clinical information. This subject had an approximately 60 kb deletion that removed the final two exons of *LINC00299* (chr2: 8,140,756–8,200,476), without altering neighboring genes. The individual is a 43-year-old female with a history of speech delay, mild-moderate intellectual disability, and bouts of confusion and abnormal behavior and is currently receiving treatment for bipolar disorder. She reportedly also has a seizure disorder with negative EEG assessment and has a disorder of articulation. She is said to be nondysmorphic, with weight 56.3 kg and height 164.5 cm. No brain imaging studies have been performed, and no cell line nor tissue were available for expression studies.

**Table 1. Subjects with Disruption Only at the *LINC00299* Locus**

Sample	Chromosome Region	Event	Length	Phenotype
L1	chr2: 8,058,207–8,117,927	CN loss	60 Kb	DD, seizures, bipolar
S1	chr2: 5,148,670–8,945,097	CN loss	3.8 Mb	DD
S2	chr2: 7,554,804–16,844,265	CN loss	9.3 Mb	DD
ISCA nssv577684	chr2: 6,588,755–16,161,372	CN loss	9.6 Mb	global DD

DD, developmental delay; CN, copy number.

The two subjects (DGAP162 and L1) with their similarly defined genomic lesions altering only *LINC00299* and the absence of such disruptions in normal individuals provide evidence for a role for this lincRNA gene in normal human development. Apart from the genetic lesions disrupting *LINC00299*, neither subject had a clinically interpreted dosage imbalance elsewhere in the genome, within the limits of the array analysis. The variability in the neurological and developmental phenotypes observed between the two subjects lacking the final two exons of this gene may reflect the actions of genetic, environmental, or stochastic modifiers. We acknowledge the differences in severity of phenotype between DGAP162 and L1. These issues, like delineation of the biological function of *LINC00299*, will require understanding its various isoforms and the timing and importance of their expression during development of the brain and other tissues. Because of the rarity of genomic lesions disrupting transcripts of *LINC00299*, this study does not have sufficient power to derive a statistically significant association between disruptions in this gene and pathology; these data show a suggestive clinical association and evidence for complex splicing patterns of *LINC00299* and a potential role in neurodevelopment. Subjects with disruptions in *LINC00299* have developmental disabilities of varying severity, though the most prominent effects appear to be in development and function of the brain. Both subjects for whom clinical data were available were also described as very thin, possibly suggesting some form of metabolic disorder. The model of dysregulation that we propose to explain the deficit is not haploinsufficiency, but rather either a gain-of-function or a dominant-negative action of a truncated form of *LINC00299* where targets of the transcript are bound by a nonfunctional transcript. Extensive additional analyses are warranted to understand this mechanism and its potential specificity to lincRNA function. Overall, the implication of a lincRNA as an important component of proper neurodevelopment provides an entrée into the biology of such lincRNAs.

### Supplemental Data

Supplemental Data include one figure demonstrating reverse-transcriptase-negative RT-PCR data compared to reverse-transcriptase-positive RT-PCR data for pre- and posttranslocation primers and one table with all primers used in the study and can

be found with this article online at <http://dx.doi.org/10.1016/j.ajhg.2012.10.016>.

### Acknowledgments

We thank the subjects and their families for participating in this study. This work was funded by NIH GM061354 (J.F.G. and C.C.M.), HD065286 (J.F.G.), K99MH095867 (M.E.T.), and R33MH087896 (S.J.H.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. M.E.T. and J.F.G. are supported by the Simons Foundation for Autism Research and the Nancy Lurie Marks Family Foundation. C.E. is funded by a Canada Institute of Health Research Canada Research Chair in Psychiatric Genetics.

Received: April 9, 2012

Revised: October 9, 2012

Accepted: October 19, 2012

Published online: December 6, 2012

### Web Resources

The URLs for data presented herein are as follows:

BamStat, <http://talkowski.mgh.harvard.edu/>  
 BLAST and BLASTX, <http://blast.ncbi.nlm.nih.gov/>  
 dbSNP, <http://www.ncbi.nlm.nih.gov/projects/SNP/>  
 DECIPHER, <https://decipher.sanger.ac.uk/>  
 DGAP database, <http://www.bwhpathology.org/dgap/>  
 DGV, <http://projects.tcag.ca/variation/>  
 ISCA, <https://www.iscaconsortium.org/index.php/search>  
 Online Mendelian Inheritance in Man (OMIM), <http://www.omim.org/>  
 Pfam, <http://pfam.sanger.ac.uk/>

### References

1. Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* 106, 11667–11672.
2. Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223–227.

3. Talkowski, M.E., Ernst, C., Heilbut, A., Chiang, C., Hanscom, C., Lindgren, A., Kirby, A., Liu, S., Muddukrishna, B., Ohsumi, T.K., 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**, 469–481.
4. Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124.
5. Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., et al. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419.
6. Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S., et al. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* **42**, 1113–1117.
7. Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076.
8. Burd, C.E., Jeck, W.R., Liu, Y., Sanoff, H.K., Wang, Z., and Sharpless, N.E. (2010). Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet.* **6**, e1001233.
9. Higgins, A.W., Alkuraya, F.S., Bosco, A.F., Brown, K.K., Bruns, G.A., Donovan, D.J., Eisenman, R., Fan, Y., Farra, C.G., Ferguson, H.L., et al. (2008). Characterization of apparently balanced chromosomal rearrangements from the developmental genome anatomy project. *Am. J. Hum. Genet.* **82**, 712–722.
10. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.
11. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
12. Chiang, C., Jacobsen, J.C., Ernst, C., Hanscom, C., Heilbut, A., Blumenthal, I., Mills, R.E., Kirby, A., Lindgren, A.M., Rudiger, S.R., et al. (2012). Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. *Nat. Genet.* **44**, 390–397, S1.
13. Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M.F., Kellis, M., Lindblad-Toh, K., and Lander, E.S. (2007). Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl. Acad. Sci. USA* **104**, 19428–19433.
14. Sheridan, S.D., Theriault, K.M., Reis, S.A., Zhou, F., Madison, J.M., Daheron, L., Loring, J.F., and Haggarty, S.J. (2011). Epigenetic characterization of the FMR1 gene and aberrant neurodevelopment in human induced pluripotent stem cell models of fragile X syndrome. *PLoS ONE* **6**, e26203.
15. Talkowski, M.E., Rosenfeld, J.A., Blumenthal, I., Pillalamarri, V., Chiang, C., Heilbut, A., Ernst, C., Hanscom, C., Rossin, E., Lindgren, A.M., et al. (2012). Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* **149**, 525–537.