



Genome-Wide Identification of Human RNA Editing Sites by Parallel DNA Capturing and Sequencing

Jin Billy Li, *et al.*

Science **324**, 1210 (2009);

DOI: 10.1126/science.1170995

The following resources related to this article are available online at www.sciencemag.org (this information is current as of May 28, 2009):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/324/5931/1210>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/324/5931/1210/DC1>

This article **cites 28 articles**, 12 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/324/5931/1210#otherarticles>

This article appears in the following **subject collections**:

Molecular Biology

http://www.sciencemag.org/cgi/collection/molec_biol

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

12. T. Moore, K. M. Armstrong, *Nature* **421**, 370 (2003).
 13. L. B. Ekstrom, P. R. Roelfsema, J. T. Arsenault, G. Bonmassar, W. Vanduffel, *Science* **321**, 414 (2008).
 14. Materials and methods are available as supporting material on *Science* Online.
 15. P. Fries, T. Womelsdorf, R. Oostenvelde, R. Desimone, *J. Neurosci.* **28**, 4823 (2008).
 16. M. Zeitler, P. Fries, S. Gielen, *Neural Comput.* **18**, 2256 (2006).
 17. L. G. Nowak, J. Bullier, in *Cerebral Cortex*, K. S. Rockland, J. H. Kaas, A. Peters, Eds. (Plenum, New York, 1997), vol. 12, pp. 205–241.
 18. N. E. Huang *et al.*, *Proc. R. Soc. London Ser. A* **454**, 903 (1998).
 19. Y. Dan, M. M. Poo, *Neuron* **44**, 23 (2004).
 20. N. Kopell, G. B. Ermentrout, M. A. Whittington, R. D. Traub, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1867 (2000).
 21. A. von Stein, C. Chiang, P. Konig, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14748 (2000).
 22. A. Brovelli *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9849 (2004).
 23. B. Pesaran, M. J. Nelson, R. A. Andersen, *Nature* **453**, 406 (2008).
 24. P. R. Roelfsema, A. K. Engel, P. Konig, W. Singer, *Nature* **385**, 157 (1997).
 25. A. Sirota *et al.*, *Neuron* **60**, 683 (2008).
 26. T. J. Buschman, E. K. Miller, *Science* **315**, 1860 (2007).
 27. A. G. Siapas, E. V. Lubenov, M. A. Wilson, *Neuron* **46**, 141 (2005).
 28. T. Womelsdorf, P. Fries, *Curr. Opin. Neurobiol.* **17**, 154 (2007).
 29. We thank G. Pielli, D. Stock, and C. Alfes for help with the animal training and Z.-X. Liu for help with the

Granger causality analysis. We also thank N. Bichot, R. Landman, G. Mulliken, and A. Mitz for helpful discussions. Supported by grants EY017292 and EY017921 to R.D. S.J.G. was supported in part by grant MH64445 from NIH (USA).

Supporting Online Material

www.sciencemag.org/cgi/content/full/324/5931/1207/DC1
 Materials and Methods
 SOM Text
 Figs. S1 to S5
 Table S1
 References

26 January 2009; accepted 8 April 2009
 10.1126/science.1171402

Genome-Wide Identification of Human RNA Editing Sites by Parallel DNA Capturing and Sequencing

Jin Billy Li,^{1*} Erez Y. Levanon,^{1*} Jung-Ki Yoon,^{1†} John Aach,¹ Bin Xie,² Emily LeProust,³ Kun Zhang,^{1‡} Yuan Gao,^{2,4} George M. Church^{1§}

Adenosine-to-inosine (A-to-I) RNA editing leads to transcriptome diversity and is important for normal brain function. To date, only a handful of functional sites have been identified in mammals. We developed an unbiased assay to screen more than 36,000 computationally predicted nonrepetitive A-to-I sites using massively parallel target capture and DNA sequencing. A comprehensive set of several hundred human RNA editing sites was detected by comparing genomic DNA with RNAs from seven tissues of a single individual. Specificity of our profiling was supported by observations of enrichment with known features of targets of adenosine deaminases acting on RNA (ADAR) and validation by means of capillary sequencing. This efficient approach greatly expands the repertoire of RNA editing targets and can be applied to studies involving RNA editing–related human diseases.

Adenosine-to-inosine (A-to-I) RNA editing converts a genomically encoded adenosine (A) into inosine (I), which in turn is read as guanosine (G), and increases transcriptomic diversity (1, 2). It is critical for normal brain function (3–7) and is linked to various disorders (8). To date, a total of 13 edited genes have been identified within nonrepetitive regions of the human genome (table S1). The limiting factor in the identification of RNA editing targets has been the number of locations that could be profiled by the sequencing of DNA and RNA samples. Even with recent developments in massively parallel DNA sequencing technologies (9), it still remains

expensive to sequence whole genomes and transcriptomes, both of which are required to identify RNA editing targets. Here, we report an efficient and unbiased genome-wide approach to identify RNA editing sites that uses tailored target capture followed by massively parallel DNA sequencing.

We first compiled a set of 59,437 genomic locations enriched with RNA editing sites, excluding repetitive regions such as *Alu* (fig. S1) (10). To reduce biases in detection, the key criteria for previous predictions of editing targets—conservation, coding potential, and RNA secondary structure (11–15)—were not taken into account. Over 90% of the previously identified editing targets are present in this data set (table S1). We designed padlock probes (16) for 36,208 sites that best satisfied our criteria for probe design (table S2) (10). Sites near splicing junctions required two different probes [targeting genomic DNA (gDNA) and cDNA], giving rise to a total of 41,046 probes designed for 36,208 sites (table S2).

To identify RNA editing sites, we used gDNA and cDNA from seven different tissues (cerebellum, frontal lobe, corpus callosum, diencephalon, small intestine, kidney, and adrenal), all derived from a single individual so as to rule out polymorphisms among populations. The pool of probes was hybridized to gDNA and cDNA in separate reactions (Fig. 1A and fig. S2). We sequenced the amplicons and identified sites where an A allele was observed in gDNA, whereas at least a fraction of G reads were present in the cDNA samples. A majority of sites were covered with multiple reads (Fig. 1B). Two independent

Table 1. Statistics of sequencing of samples used in this study.

Sample	Total reads	Mappable reads	Sites with ≥1 read	Fraction of sites with ≥1 read	RNA editing candidates*
gDNA (combined)	12,604,941	12,150,194	33,886	93.6%	N/A
Replicate 1	5,145,193	5,042,006	32,491	89.7%	N/A
Replicate 2	7,459,748	7,108,188	32,942	91.0%	N/A
cDNA					
Cerebellum	5,538,459	5,382,743	26,220	72.4%	126
Frontal lobe (combined)	14,065,388	13,360,868	28,382	78.4%	268
Replicate 1	6,950,660	6,563,630	26,617	73.5%	238
Replicate 2	7,114,728	6,797,238	26,628	73.5%	230
Corpus callosum	5,096,832	4,963,983	25,447	70.3%	180
Diencephalon	5,420,151	5,291,184	25,187	69.6%	172
Small intestine	6,516,258	6,172,901	26,845	74.1%	181
Kidney	6,354,025	5,984,709	26,299	72.6%	177
Adrenal	2,251,755	2,188,637	23,589	65.1%	121

*A site with evidence for RNA editing is required to have an editing level of ≥5% and a log-likelihood (LL) score of ≥2 (10).

¹Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA. ²Center for the Study of Biological Complexity, Virginia Commonwealth University, 1000 West Cary Street, Richmond, VA 23284, USA. ³Genomics Solution Unit, Agilent Technologies, 5301 Stevens Creek Boulevard, Santa Clara, CA 95051, USA. ⁴Department of Computer Science, Virginia Commonwealth University, 401 West Main Street, Richmond, VA 23284, USA.

*These authors contributed equally to this work.

†Present address: College of Medicine, Seoul National University, Seoul 110-799, Korea.

‡Present address: Department of Bioengineering, University of California, San Diego, CA 92093, USA.

§To whom correspondence should be addressed. E-mail: http://arep.med.harvard.edu/gmc/email.html

technical replicates were well correlated for both gDNA (Fig. 1C) and frontal lobe cDNA (Fig. 1D). In addition, the editing levels were highly correlated between the two replicates (Fig. 1E).

A total of 57.8 million reads were obtained, among which 55.5 million sequences were mapped to the target regions (Table 1) (10). To identify RNA editing sites, we searched for positions where

a homozygous A was seen in gDNA and more than 5% of reads were G in at least two of the seven cDNA samples with a log likelihood score of ≥ 2 (10). A total of 239 such sites (in 207 targets) with stringent thresholds were identified and referred to as class I (table S3), including 10 of all 13 known edited genes (tables S1 and S3).

To validate the class I set, we randomly selected 18 different sites, successfully amplified them with polymerase chain reaction, and sequenced them using the dideoxynucleotide (Sanger) method. We also tested gDNA and frontal lobe cDNA from two additional donors (a total of 12 samples per site). Fourteen of the 18 sites were clearly edited, with a majority in all three donors (Fig. 2A and fig. S3). One of the

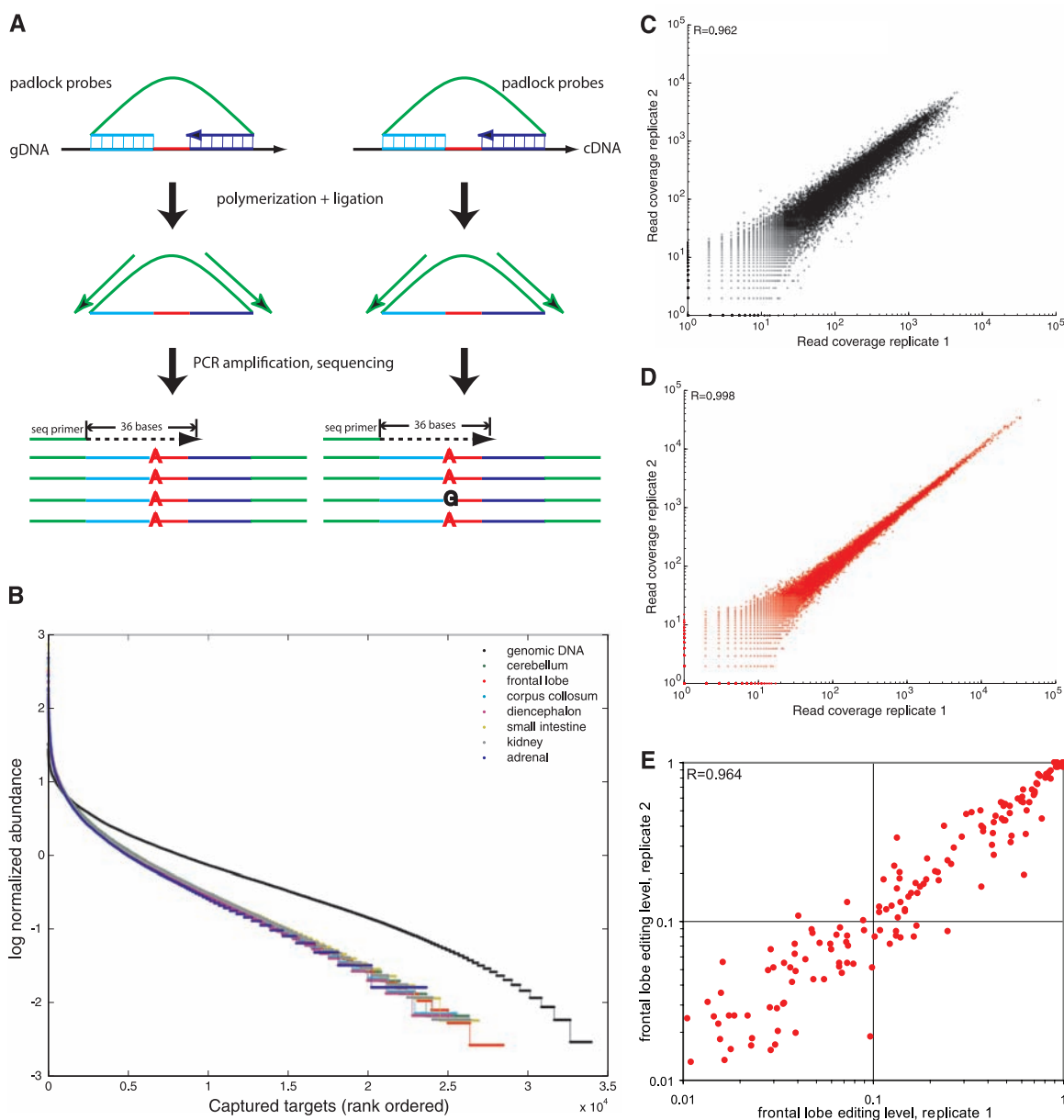
Table 2. Features of class I RNA editing sites.

Feature	36,208 set	Class I set	P value*
Double-stranded RNA (dsRNA) structure†	16%	41%	2.7×10^{-21}
Downstream of base G	34%	8%	1.1×10^{-21}
Coding sequence	52%	23%	2.0×10^{-19}
Conserved region‡	42%	21%	1.5×10^{-11}
MicroRNA target sequences§	33%	20%	4.6×10^{-6}

*P values were calculated using Fisher's exact test. †Sequence centered on the site [4001 base pairs (bp) total] forms a dsRNA structure (10). ‡Sites in the "most conserved" track in the University of California Santa Cruz genome browser (<http://genome.ucsc.edu>). §Sequence centered on the site (13 bp total) contains 7 bp microRNA seeds (<http://microna.sanger.ac.uk>) (10).

Fig. 1. Screening for RNA editing sites using padlock capture and massively parallel DNA sequencing technologies.

(A) Schematic diagram of the padlock technology. The candidate RNA editing sites are specifically targeted by padlocks in both gDNA and cDNA samples from a single individual. Circles are formed when polymerase, deoxynucleotide triphosphate, and ligase are added, subsequently amplified, and sequenced with an Illumina genome analyzer (Illumina, San Diego, CA). (B) Uniformity of target abundance distributions in sequences obtained for all samples. Each graph shows the abundance of captured target sequences for each target over all samples, in which targets are given in ranked order. Abundance is represented by the \log_{10} of the target coverage normalized to the mean of the target coverage for the sample. The abundance of different sites is nonuniform because of capturing biases and expression-level variations. The gDNA and frontal lobe replicates were combined in this analysis. (C to E) Target capture is highly reproducible for technical replicates. (C) Correlation of coverage of sites for gDNA replicates (Pearson correlation, $r = 0.962$); (D) correlation of coverage of sites for frontal-lobe cDNA replicates ($r = 0.998$); and (E) correlation of RNA editing level in frontal-lobe replicates ($r = 0.964$). Editing level is the number of G reads divided by the sum number of A and G reads when the sum is ≥ 10 .



remaining sites, *ZNF7*, was edited at 1.1% level (2 of 187 individually sequenced clones). The false discovery rate of the set is thus up to 17% (3 of 18 sites).

RNA editing occurs when ADARs (adenosine deaminases acting on RNA) bind to an extended RNA duplex within target RNAs (17, 18). Indeed, the class I set is significantly enriched, as compared with the 36,208-candidates set, with sites that are located in RNA double-stranded regions (Table 2 and table S4) (10). Previous studies have indicated that ADARs have a sequence preference for strong G depletion in the

nucleotide 5' to the editing site (19). This observation is in agreement with our findings (Table 2 and fig. S4).

Of the 239 class I sites, 55 (23%) are located in coding regions, 38 of which change amino acids (table S3), including one that adds an additional 29 amino acids by changing a stop codon (UAG) to a tryptophan (UGG) (Fig. 2B). There is a clear bias against the coding regions (Table 2), where changes are less likely to be tolerated. Similarly, possible microRNA target sequences are significantly reduced in our set (Table 2).

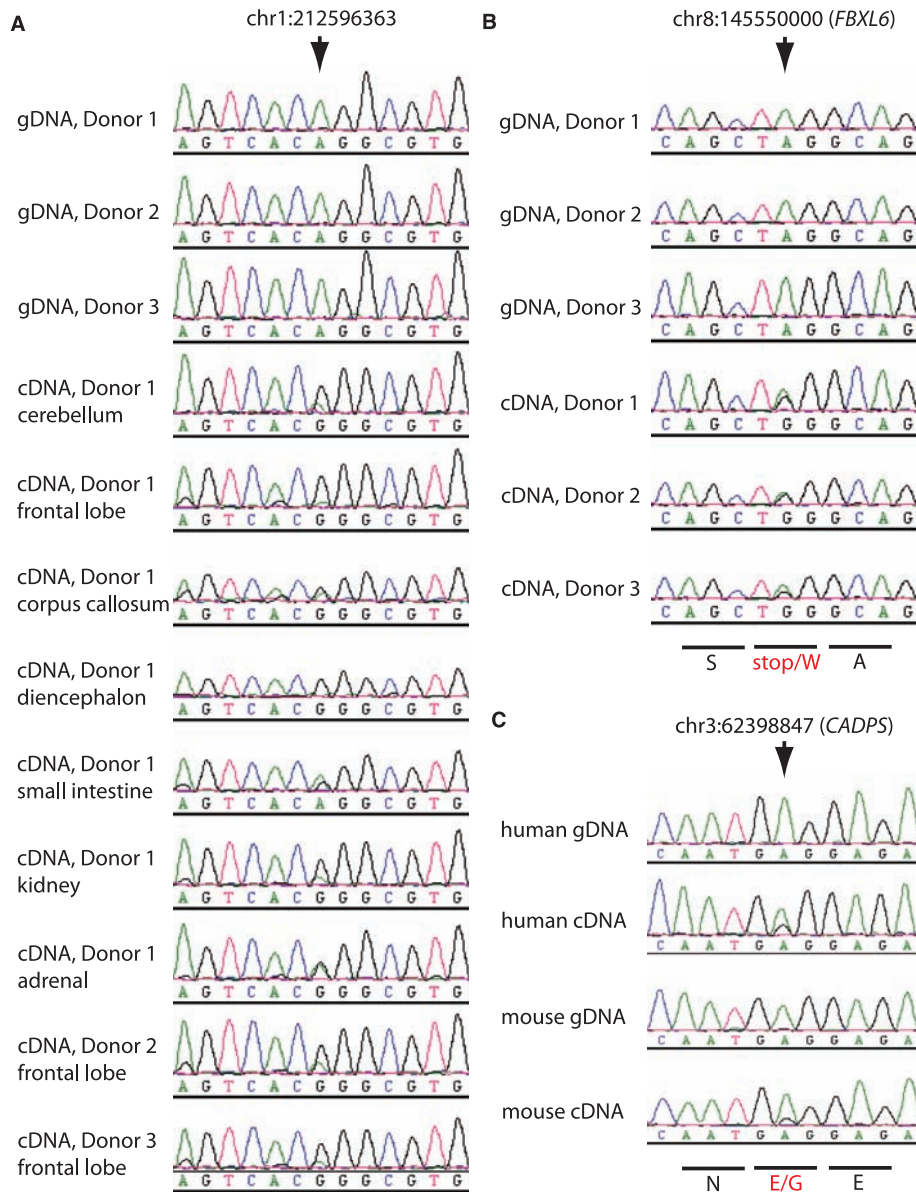


Fig. 2. Validation of RNA editing sites with conventional Sanger sequencing. **(A)** Sequencing chromatogram traces of an exemplary site, chr1:212596363, in gDNA and all seven tested cDNAs of the first donor and in gDNA and frontal lobe cDNA from two unrelated donors. Some nearby sites are also edited. A complete list of validated sites is in fig. S3. **(B)** At site chr8:145550000 [in *F-box and leucine-rich repeat protein 6* (*FBXL6*) gene], the genomic A in the stop codon (TAG) is highly edited, allowing the addition of 29 amino acids to the protein in all three donors. **(C)** The *CADPS* site, chr3:62398847, is edited in human (shown is the frontal-lobe cDNA), and the conserved site is edited in mouse as well (shown is the brain cDNA). The editing event leads to amino acid change from glutamic acid (GAG) to glycine (GGG).

Sequence conservation has been the main criterion in various attempts to identify new RNA editing sites. However, it has been shown that editing is enriched in the primate lineage, mainly because of widespread editing in *Alu* repetitive elements (20–24). In the class I set, the number of sites with flanking sequences conserved between human and mouse is significantly underrepresented (Table 2) (10). Of those sites that are highly conserved (fig. S5), we sequenced one located in the *CADPS* (Ca^{2+} -dependent secretion activator) gene in mouse gDNA and cDNA samples and observed an editing signal. This site is probably edited in all vertebrates based on A-to-G changes in supporting expressed sequence tags. Fourteen of the 50 editing sites located in conserved regions harbor a G in at least one of eight other vertebrate genomes (table S5), a phenomenon previously observed in flies (25). From an evolutionary perspective, RNA editing may thus play a role similar to genetic mutation in creating genetic diversity. In contrast to mutation, however, RNA editing provides a much wider spectrum of “genetic dosage”; our data demonstrate that the level ranges from very low to full editing (fig. S6).

In agreement with previous observations that targets of RNA editing are involved in nervous-system function (7, 11–15, 26–28), we found that the class I sites were enriched with functions such as synapse, cell trafficking, and membrane. Furthermore, many sites are located within genes that are implicated in human brain-related diseases (table S6). In addition to class I sites, many more sites are likely to be edited. When we relaxed our criteria to require only one tissue to be edited, we identified an additional set of 330 potential candidate editing sites as the class II set (table S7). We validated a selected candidate from this set, *GLII* (Glioma-associated oncogene homolog 1, at site chr12:56150891), which was highly edited in the frontal lobe of all three donors (fig. S3). An additional set of 141 sites was identified as class III when the editing level threshold was reduced to 2% (table S8), which suggests that many targets may be edited at very low levels. By sequencing 118 clones of the class III site chr11:74994333 in *MAP6* (microtubule-associated protein 6), we found 13 clones with a G at the editing site.

Although it is unclear if the extensive editing of primate *Alu* sites has any biological role, it may require an increased expression of ADAR proteins in humans, which in turn may lead to the editing of non-*Alu* RNAs. In support of this scenario, most of the nonrepetitive sites we identified do not seem to be conserved beyond the primate lineage and may play roles in primate-specific functions. Many of the identified editing sites are located in noncoding RNAs that have recently been linked to brain function (29).

The approach described herein can be readily extended to a wider variety of tissues in normal and diseased individuals in order to identify additional RNA editing sites and measure their

editing levels. The enlarged set of nonrepetitive RNA editing targets may help unravel rules of RNA editing in human diseases and behavior.

References and Notes

- B. L. Bass, *Annu. Rev. Biochem.* **71**, 817 (2002).
- K. Nishikura, *Nat. Rev. Mol. Cell Biol.* **7**, 919 (2006).
- M. Higuchi *et al.*, *Nature* **406**, 78 (2000).
- R. Brusa *et al.*, *Science* **270**, 1677 (1995).
- M. Singh *et al.*, *J. Biol. Chem.* **282**, 22448 (2007).
- M. J. Palladino, L. P. Keegan, M. A. O'Connell, R. A. Reenan, *Cell* **102**, 437 (2000).
- H. Lomeli *et al.*, *Science* **266**, 1709 (1994).
- S. Maas, Y. Kawahara, K. M. Tamburro, K. Nishikura, *RNA Biol.* **3**, 1 (2006).
- J. Shendure, H. Ji, *Nat. Biotechnol.* **26**, 1135 (2008).
- Materials and methods are available as supporting material on Science Online.
- E. Y. Levanon *et al.*, *Nucleic Acids Res.* **33**, 1162 (2005).
- B. Hoopengardner, T. Bhalla, C. Staber, R. Reenan, *Science* **301**, 832 (2003).
- W. M. Gommans *et al.*, *RNA* **14**, 2074 (2008).
- D. R. Clutterbuck, A. Leroy, M. A. O'Connell, C. A. Semple, *Bioinformatics* **21**, 2590 (2005).
- J. Ohlson, J. S. Pedersen, D. Haussler, M. Ohman, *RNA* **13**, 698 (2007).
- G. J. Porreca *et al.*, *Nat. Methods* **4**, 931 (2007).
- A. G. Polson, B. L. Bass, *EMBO J.* **13**, 5701 (1994).
- K. Nishikura *et al.*, *EMBO J.* **10**, 3523 (1991).
- K. A. Lehmann, B. L. Bass, *Biochemistry* **39**, 12875 (2000).
- A. Athanasiadis, A. Rich, S. Maas, *PLoS Biol.* **2**, e391 (2004).
- M. Blow, P. A. Futreal, R. Wooster, M. R. Stratton, *Genome Res.* **14**, 2379 (2004).
- E. Eisenberg *et al.*, *Trends Genet.* **21**, 77 (2005).
- D. D. Kim *et al.*, *Genome Res.* **14**, 1719 (2004).
- E. Y. Levanon *et al.*, *Nat. Biotechnol.* **22**, 1001 (2004).
- N. Tian, X. Wu, Y. Zhang, Y. Jin, *RNA* **14**, 211 (2008).
- C. M. Burns *et al.*, *Nature* **387**, 303 (1997).
- M. Kohler, N. Burnashev, B. Sakmann, P. H. Seeburg, *Neuron* **10**, 491 (1993).
- B. Sommer, M. Kohler, R. Sprengel, P. H. Seeburg, *Cell* **67**, 11 (1991).
- T. R. Mercer *et al.*, *Neuroscientist* **14**, 434 (2008).
- We thank R. Emeson, M. P. Ball, and F. Isaacs for critical reading of the manuscript; P. Wang and Z. Liu (BioChain Institute) for helping collect human samples; M. Higuchi and P. Seeburg for providing *ADAR2*^{-/-} mouse brain cDNA; Harvard Biopolymers Facility for help with Illumina sequencing; and A. Ahlford, H. Ebling, and J. Santosuosso for assistance with Sanger sequencing. E.Y.L. was supported by the Machiah foundation. Funding came from National Human Genome Research Institute Centers of Excellence in Genomic Science grant to G.M.C. The Illumina sequencing data are deposited at the National Center for Biotechnology Information Short Read Archive under accession number SRA008181.

Supporting Online Material

www.sciencemag.org/cgi/content/full/324/5931/1210/DC1

Materials and Methods

Figs. S1 to S10

Tables S1 to S12

References

15 January 2009; accepted 1 April 2009

10.1126/science.1170995

Unstable Tandem Repeats in Promoters Confer Transcriptional Evolvability

Marcelo D. Vinces,^{1,2,3*} Matthieu Legendre,^{1,4*} Marina Caldara,¹ Masaki Hagihara,⁵ Kevin J. Verstrepen^{1,2,3†}

Relative to most regions of the genome, tandemly repeated DNA sequences display a greater propensity to mutate. A search for tandem repeats in the *Saccharomyces cerevisiae* genome revealed that the nucleosome-free region directly upstream of genes (the promoter region) is enriched in repeats. As many as 25% of all gene promoters contain tandem repeat sequences. Genes driven by these repeat-containing promoters show significantly higher rates of transcriptional divergence. Variations in repeat length result in changes in expression and local nucleosome positioning. Tandem repeats are variable elements in promoters that may facilitate evolutionary tuning of gene expression by affecting local chromatin structure.

The genomes of most organisms are not uniformly prone to change because they contain hotspots for mutating events. An abundant class of sequences that mutate at higher frequencies than the surrounding genome is composed of tandem repeats (TRs, also known as satellite DNA), DNA sequences repeated adjacent to one another in a head-to-tail manner (1). Errors during replication make TRs unstable, generating changes in the number of repeat units that are 100 to 10,000 times more frequent than

point mutations (2). Variable TRs are often dismissed as nonfunctional “junk” DNA. However, some TRs located within coding regions (exons) have demonstrable functional roles. For example, TR copy numbers in genes such as *FLO1* in *Saccharomyces cerevisiae* generate plasticity in adherence to substrates (3). In canines, variable repeats located in *Alx-4* and *Runx-2* confer variability to skeletal morphology, which may have facilitated the diversification of domestic dogs bred by humans (4). Thus, repeats located in coding regions may increase the evolvability of proteins.

There is also evidence that repeats influence expression of certain genes (5–7). To investigate the involvement of TRs in gene expression variation, we first mapped and classified all repeats in the S288C yeast genome (8) (data set S1). TRs are enriched in yeast promoters (table S1). Of the ~5700 promoters in the genome, 25% (1455) contain at least one TR. Many TRs in promoters consist of short, A/T-rich sequences (table S2, fig. S1, and data set S2). Comparison of orthologous regions in genomes of different *S. cerevisiae* strains showed that many of the TRs are variable (data set S1). For example, 24.1% of

orthologous TR loci in promoters differ in the number of repeat units between the two fully sequenced strains, S288C and RM11 (8). To confirm this, we sequenced 33 randomly chosen promoter repeats in seven *S. cerevisiae* genomes (Fig. 1A, figs. S2 and S3, and data set S3). Twenty-five of the 33 TRs differed in repeat units in at least one of the seven strains. The repeat variation frequency is 40-fold higher than the frequency of insertions and deletions (indels) and of point mutations in the surrounding non-repetitive sequence ($P < 10^{-15}$) (figs. S2 and S3).

To determine whether promoter TR variation affects gene expression, we compared repeat variability to expression divergence (ED), which represents how fast the transcriptional activity of each gene evolves (9–11). Promoters containing TRs showed significantly ($P < 1.75 \times 10^{-4}$) higher amounts of ED than did promoters lacking TRs when comparing yeast species (*S. cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. kudriavzevii*) (Fig. 1, B to D, and fig. S4A) and *S. cerevisiae* strains (S288C and RM11) (Fig. 1, E to G, and fig. S4, B and C). This difference was independent of factors known to affect transcriptional divergence, for example, the presence of TATA boxes (fig. S5). Only promoters containing variable numbers of repeat units between strains or species showed the elevated ED (Fig. 1, D and G). Furthermore, when variable TRs were binned into variable and highly variable (10% most variable) groups, highly variable repeats displayed even higher ED. Hence, ED correlates not merely with TRs in promoters but more specifically with repeat number variation.

To directly test whether changes in promoter TRs affect transcriptional activity, we varied the TR repeat number in the promoters of yeast genes *YHB1*, *MET3*, and *SDT1* (Fig. 2 and fig. S6A). For each construct, expression increased as the length of the TR increased from zero, until a certain size was reached, after which expression dropped off. To determine whether natural variation between strains corresponded to similar changes in gene expression, we cloned promoters of several strains

¹FAS Center for Systems Biology, Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA. ²Laboratory for Systems Biology, Flanders Institute for Biotechnology (VIB), Katholieke Universiteit Leuven (K.U. Leuven), B-3001 Heverlee, Belgium. ³Genetics and Genomics Group, Centre of Microbial and Plant Genetics (CMPG), K.U. Leuven, Gaston Geenslaan 1, B-3001 Leuven (Heverlee), Belgium. ⁴Structural and Genomic Information Laboratory, CNRS-UPR 2589, IFR-88, Université de la Méditerranée Parc Scientifique de Luminy, Avenue de Luminy, FR-13288 Marseille, France. ⁵The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, 567-0047, Japan.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: Kevin.Verstrepen@biw.vib-kuleuven.be