NUDIX5 overexpression in cancer patients correlates with a poor outcome in several cancer types, including breast cancer (fig. S8, A and B). The correlation with high expression of PARP1 (fig. S8C) may suggest that cancer cells depend on nuclear ATP generation. Hence, NUDIX5 is a promising target for PARP1 combinatorial drug therapy (*20*).

REFERENCES AND NOTES

1. G. J. Narlikar, M. L. Phelan, R. E. Kingston, *Mol. Cell* **8**, 1219–1230 (2001).
2. V. G. Allfrey, A. E. Mirsky, *Proc. Natl. Acad. Sci. U.S.A.* **43**, 589–598 (1957).
3. I. Betel, *Arch. Biochem. Biophys.* **134**, 271–274 (1969).
4. H. Imamura *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15651–15656 (2009).
5. J. Hageman, M. J. Vos, M. A. van Waarde, H. H. Kampinga, *J. Biol. Chem.* **282**, 34334–34345 (2007).
6. G. P. Vicent *et al.*, *PLOS Genet.* **5**, e1000567 (2009).
7. G. P. Vicent *et al.*, *Genes Dev.* **25**, 845–862 (2011).
8. C. Ballaré *et al.*, *Mol. Cell* **49**, 67–79 (2013).
9. F. Le Dily *et al.*, *Genes Dev.* **28**, 2151–2162 (2014).
10. R. H. G. Wright *et al.*, *Genes Dev.* **26**, 1972–1983 (2012).
11. T. Zhang *et al.*, *J. Biol. Chem.* **287**, 12405–12416 (2012).
12. A. G. McLennan, *Cell. Mol. Life Sci.* **63**, 123–143 (2006).
13. M. Zha, C. Zhong, Y. Peng, H. Hu, J. Ding, *J. Mol. Biol.* **364**, 1021–1033 (2006).
14. C. A. Sartorius, G. S. Takimoto, J. K. Richer, L. Tung, K. B. Horwitz, *J. Mol. Endocrinol.* **24**, 165–182 (2000).
15. H. Maruta, N. Matsumura, S. Tanuma, *Biochem. Biophys. Res. Commun.* **236**, 265–269 (1997).
16. S. L. Oei, M. Ziegler, *J. Biol. Chem.* **275**, 23234–23239 (2000).
17. H. Maruta *et al.*, *Biol. Pharm. Bull.* **30**, 447–450 (2007).
18. B. G. Ju *et al.*, *Science* **312**, 1798–1802 (2006).
19. M. Altmeyer *et al.*, *Nat. Commun.* **6**, 8088 (2015).
20. K. M. Frizzell, W. L. Kraus, *Breast Cancer Res.* **11**, 111 (2009).

ACKNOWLEDGMENTS

We thank H. Kampinga for the luciferase plasmids, H. Imamura for the Ateam plasmids, M. Jacobson for the PARG-GST plasmid, V. Schreiber for the PARP1-GST plasmid, S. Capdevila and J. M. Caballero for support with IVIS, and J. Valcarcel for advice with the manuscript and CRG core facilities. Supported by the Spanish MEC (CSD2006-00049; BMC 2003-02902 and 2010-15313), and Catalan government (AGAUR) grants BIO2014-57518-R (B.O.), SAF2011-30578 (O.Y.), and BFU2014-53787-P (M.J.M.). CRG and IRB Barcelona are recipients of the Severo Ochoa Award of Excellence from MINECO (Spain). Global data sets have been deposited in GEO with accession numbers GSE53855, GSE64136, GSE64161, and GSE64163.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6290/1221/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 to S8
Movies S1 and S2
References (*21*, *22*)

24 November 2015; accepted 9 May 2016
10.1126/science.aad9335

RNA TRANSCRIPTION

TT-seq maps the human transient transcriptome

Björn Schwalb,[1]* Margaux Michel,[1]* Benedikt Zacher,[2]* Katja Frühauf,[3] Carina Demel,[1] Achim Tresch,[4,5] Julien Gagneur,[2]†‡ Patrick Cramer[1,3]‡

Pervasive transcription of the genome produces both stable and transient RNAs. We developed transient transcriptome sequencing (TT-seq), a protocol that uniformly maps the entire range of RNA-producing units and estimates rates of RNA synthesis and degradation. Application of TT-seq to human K562 cells recovers stable messenger RNAs and long intergenic noncoding RNAs and additionally maps transient enhancer, antisense, and promoter-associated RNAs. TT-seq analysis shows that enhancer RNAs are short-lived and lack U1 motifs and secondary structure. TT-seq also maps transient RNA downstream of polyadenylation sites and uncovers sites of transcription termination; we found, on average, four transcription termination sites, distributed in a window with a median width of ~3300 base pairs. Termination sites coincide with a DNA motif associated with pausing of RNA polymerase before its release from the genome.

Transcription of eukaryotic genomes produces protein-coding mRNAs and diverse noncoding RNAs (ncRNAs), including enhancer RNAs (eRNAs) (*1*, *2*). Most ncRNAs are rapidly degraded, difficult to detect, and thus far have not been mappable in their full range. Mapping of transient RNAs is required, however, for analysis of RNA sequence, function, and fate.

We developed transient transcriptome sequencing (TT-seq), a protocol that maps transcriptionally active regions and enables estimation of RNA synthesis and degradation rates. TT-seq is based on 4sU-seq, which involves a brief exposure of cells to the nucleoside analog 4-thiouridine (4sU) (Fig. 1A) (*3*). 4sU is incorporated into RNA during transcription, and the resulting 4sU-labeled RNAs are isolated and sequenced. 4sU-seq is more sensitive than RNA-seq in detecting transient RNAs. However, 4sU-seq fails to map human transcripts uniformly, because only a short 3′ region of nascent transcripts is labeled during a 5-min exposure to 4sU, and the long preexisting 5′ regions dominate the sequencing data. To remove this 5′ bias, TT-seq uses RNA fragmentation before isolation of labeled RNA fragments (Fig. 1A). Thus, TT-seq measures only newly transcribed RNA fragments and provides the number of polymerases transcribing a genomic position within 5 min.

When applied to human K562 cells, TT-seq samples newly transcribed regions uniformly, whereas 4sU-seq produces a 5′ bias (fig. S1A). The coverage of short-lived introns with respect to exons is estimated (*4*) to be 60% for TT-seq, whereas it is 23 and 8% for 4sU-seq and RNA-seq, respectively (figs. S1A and S2). TT-seq is highly reproducible (fig. S3) and enables complete mapping of transcribed regions, complementing the GRO-cap (*5*) and CAGE (*6*) protocols, which detect RNA 5′ ends (Fig. 1B). TT-seq monitors RNA synthesis, whereas protocols such as PRO-seq (*7*), NET-seq (*8*), and mNET-seq (*9*) detect RNAs attached to polymerase. Therefore, the latter protocols yield peak signals near the promoter where polymerase pauses (Fig. 1B), whereas TT-seq does not. For paused and active genes (*10*), TT-seq reveals higher rates of RNA synthesis near the promoter relative to other regions (fig. S1B).

Using TT-seq data and the segmentation algorithm GenoSTAN (*4*, *11*), we identified 21,874 genomic intervals of apparently uninterrupted transcription (transcriptional units, TUs) (Fig. 2 and fig. S4A). TT-seq is highly sensitive, recovering 65% of transcription start sites (TSSs) obtained by GRO-cap (overlapping annotations within ± 400 bp) (*5*). A total of 8543 TUs overlapped GENCODE annotations (*12*) in the sense direction of transcription (50% reciprocal overlap of annotated regions; fig. S4B). This analysis detected 7810 mRNAs, 302 long intergenic noncoding RNAs (lincRNAs), and 431 antisense RNAs (asRNA). The 2916 TUs that shared less than 50% of their length with GENCODE annotations were not classified. The remaining 10,415 TUs (48%) represented newly detected ncRNAs that we characterized further.

Transcripts arise from promoters but also from enhancers, which are regulatory elements with characteristic chromatin modifications (*13*, *14*). To detect chromatin regions comprising putative enhancers and promoters (chromatin states), we applied GenoSTAN (*11*) to ENCODE ChIP-seq (chromatin immunoprecipitation–sequencing) data (*15*) for the coactivator p300 and a series of histone modifications (H3K27me3, H3K36me3, H4K20me1, H3K4me1, H3K4me3, H3K9ac, and

[1]Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany. [2]Gene Center Munich, Ludwig-Maximilians-Universität München, Feodor-Lynen-Straße 25, 81377 Munich, Germany. [3]Department of Biosciences and Nutrition, Center for Innovative Medicine, and Science for Life Laboratory, Karolinska Institutet, Novum, Hälsovägen 7, 141 83 Huddinge, Sweden. [4]Department of Biology, University of Cologne, Zülpicher Straße 47, 50647 Cologne, Germany. [5]Max Planck Institute for Plant Breeding Research, Carl-von-Linné Weg 10, 50829 Cologne, Germany. *These authors contributed equally to this work. †Present address: Department of Informatics, Technische Universität München, Boltzmannstraße 3, 85748 Garching, Germany. **‡Corresponding author. Email: gagneur@in.tum.de (J.G.); patrick.cramer@mpibpc.mpg.de (P.C.)**

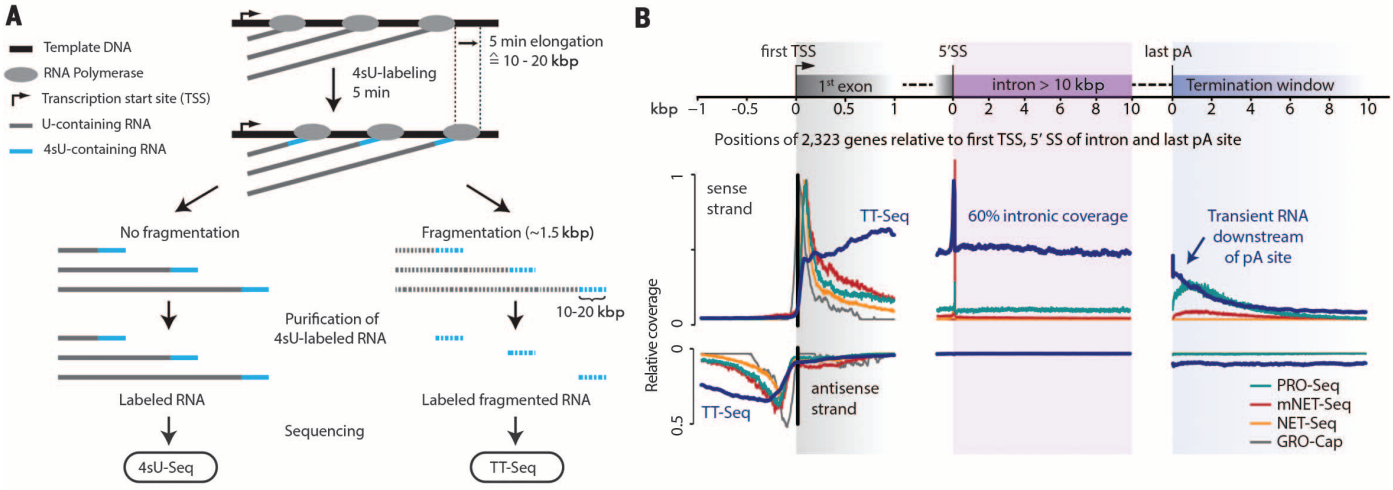

Fig. 1. TT-seq enables uniform mapping of the human transient transcriptome. (**A**) Workflow of 4sU-seq and TT-seq protocols. (**B**) Metagene coverage, comparing TT-seq with other transcriptomic methods. Shown is the average coverage for 2323 TUs lacking paused and active genes (*10*) around the first TSS (left), the 5′ splice site of intronic sequences > 10 kbp (first intron excluded), and the last pA site. Signals are relative to the maximum signal in the first kilobase pair from the first TSS.

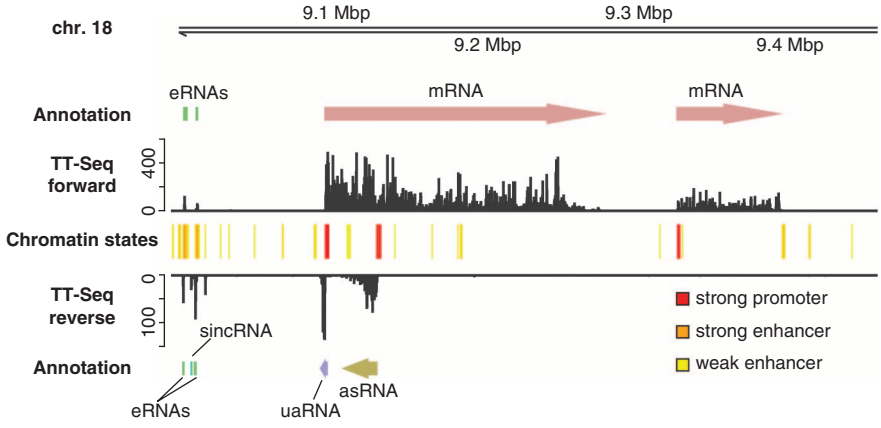

Fig. 2. Annotation of RNAs mapped by TT-seq. Example genome browser view showing RNAs from 5 of 7 transcript classes and 3 of 18 chromatin states [chromosome (chr.) 18, 9.00 to 9.54 million base pairs (Mbp)] (*4*). Arrows indicate the direction of transcription. Units on the *y* axes are read counts per 200-bp bin.

H3K27ac) and to deoxyribonuclease I hypersensitivity data (fig. S5A). Of the resulting strong enhancer state regions, 81% overlapped at least one TSS from GRO-cap (*5*) and 68% overlapped a polymerase II (Pol II) peak (*15*), compared with 52 and 37%, respectively, for ENCODE enhancer states (fig. S5, B to D).

The 10,415 nonannotated TUs were classified based on GenoSTAN-derived chromatin states and their positions relative to known GENCODE annotations (fig. S6A). TUs within 1 kilo–base pair (kbp) of a GENCODE mRNA TSS included 685 upstream antisense RNAs (uaRNAs) (*16*) and 778 convergent RNAs (conRNAs) (*8*). The 3115 TUs on the strand opposite an mRNA were classified as asRNAs when they were more than 1 kbp away from the GENCODE TSS. Remaining TUs were grouped according to their GenoSTAN chromatin state at their TSS. The 2580 TUs that originated from promoter state regions (*4*) were classified as short intergenic ncRNAs (sincRNAs) (fig. S6B). Most sincRNAs (67%) were located within 10 kbp of a GENCODE mRNA TSS. The remaining 3257 TUs originated from enhancer state regions (*4*) and were classified as eRNAs (*13*, *14*). The newly mapped ncRNAs are short (fig. S6C). On average, lincRNAs are five times as long as sincRNAs, and eRNAs have a median length of ~1000 nucleotides.

Kinetic modeling of TT-seq and RNA-seq data enabled us to estimate rates of RNA synthesis and degradation (Fig. 3 and fig. S7A) (*4*). We estimated rates of phosphodiester bond formation or breakage at each transcribed position and averaged these within TUs, thus obtaining estimates of relative transcription rates and RNA stabilities (*4*). We found that mRNAs and lincRNAs had the highest synthesis rates and longest half-lives. We determined a median mRNA half-life of ~50 min, compared with a previous estimate of ~139 min (*17*). Other transcript classes had low synthesis rates and short half-lives, explaining why short ncRNAs are difficult to detect. eRNAs had half-lives of a few minutes, consistent with prior data (*17*). Short RNA half-lives correlated with a lack of secondary structure (fig. S7B). The folding energy of eRNAs was comparable to the genomic background level (fig. S7C), and only 10% of their sequence was predicted to be structured, compared with 52% in mRNAs (fig. S7D).

We further found differences in transcription from promoters versus from enhancers (*2*). Enhancers showed lower occupancy of initiation factors TBP and TAF1 than mRNA promoters did (12- and 3.5-fold less, respectively; $P < 10^{-16}$, Fisher's exact test), whereas TFIIB and TFIIF had similar occupancies in enhancers and promoters. Occupancies were also similar for factors involved in polymerase pausing, such as NELF-E and the P-TEFb subunit cyclin T2 (fig. S8A). Synthesis of eRNAs terminated early (fig. S6C), probably because eRNAs are not enriched in U1 small nuclear ribonucleoprotein–binding sites (U1 signals; GGUAAG, GUGAGU, or GGUGAG) that can counteract early termination and lead to RNA stabilization (*18*–*20*). eRNAs contained U1 signals at the genomic background level (47%), whereas mRNAs were enriched (69%; $P < 10^{-16}$) (fig. S8B). In all transcript classes, longer RNAs were enriched with U1 signals in the first 1000 nucleotides (fig. S8C), suggesting that evolution of stable RNAs generally involves acquisition of U1 signals.

TT-seq also enabled us to uncover transcription termination sites (TTSs). TT-seq detected transient RNA downstream of the polyadenylation (pA) site (Figs. 1B and 4 and fig. S9). Such RNA is difficult to detect because RNA cleavage at the pA site leads to an unprotected 5′ end and RNA degradation by the XRN2 exonuclease (*21*, *22*). For a total of 6977 mRNA genes, we

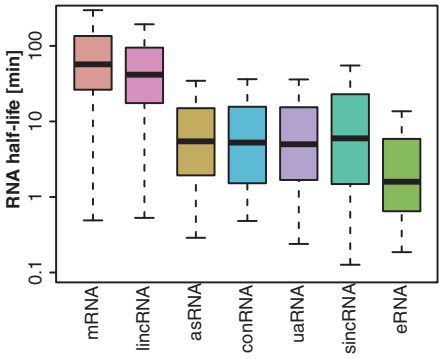

Fig. 3. Estimated RNA half-lives for different transcript classes. Black bars represent medians, boxes represent upper and lower quartiles, and whiskers represent 1.5 times the interquartile range (*4*).

derived, on average, four TTSs (figs. S10 and S11) (*4*). TTSs were located within a termination window that extended from the last pA site to an "ultimate TTS," where RNA coverage dropped to background levels (Fig. 4, A to C). The termination window had a median width of ~3300 bp and could be up to 10 kbp wide (fig. S10D), consistent with Pol II occupancy data (Fig. 4D) (*23*). For the 5113 TTSs with the strongest drop in TT-seq signal (*4*), Pol II peaks were obtained by PRO-seq (*7*), NET-seq (*8*), and mNET-seq (*9*) (fig. S12), indicating that Pol II pauses at the TTS.

The derived TTSs are strongly enriched for the sequence $(C/G)_{(2-6)}A$ (window ± 5 bp; $P < 10^{-16}$, Fisher's exact test; odds ratios, 2.98 and 1.63 for C_3A and G_3A, respectively; subscript numbers denote the number of nucleotide repeats) (Fig. 4E and figs. S13 and S14) (*4*). This sequence can contain up to six cytosines or guanines (Fig. 4F and fig. S10E). A G_3A element has also been found in a known termination signal (*24*). The $C_{(2-6)}A$ and $G_{(2-6)}A$ sequences are generally followed by a T-rich $[T_{(3-6)}]$ or an A-rich $[A_{(3-6)}]$ stretch, respectively ($P < 10^{-16}$; odds ratios, 2.39 and 1.24 for T_4 and A_4, respectively), that is located, on average, 15 bp downstream of the TTS (Fig. 4E). Such sequences were found at TTSs of all TUs (fig. S15) and can even be derived from published data (fig. S12B). In summary, the detected TTSs were highly enriched with the consensus motif $(C/G)_{(2-6)}AN_x(T/A)_{(3-6)}$, where N_x is a short stretch of nucleotides.

To test for the in vivo functionality of the derived TTS motif, we transfected expression plasmids into K562 cells that either lacked or contained four $C_3AN_8T_4$ or $C_7AN_8T_4$ motifs within 600 bp downstream of the pA site (fig. S16A and tables S1 to S4) (*4*). When the TTS motifs were present, significantly less RNA was detected downstream of the motifs, indicating termination of a fraction of polymerases (Wilcoxon test; fig. S16B). This experiment supports the functionality of the derived TTS motif in vivo. Termination depended on an upstream pA signal (fig. S16B), consistent with an occurrence of the motif in gene bodies, where they do not lead to transcription termination.

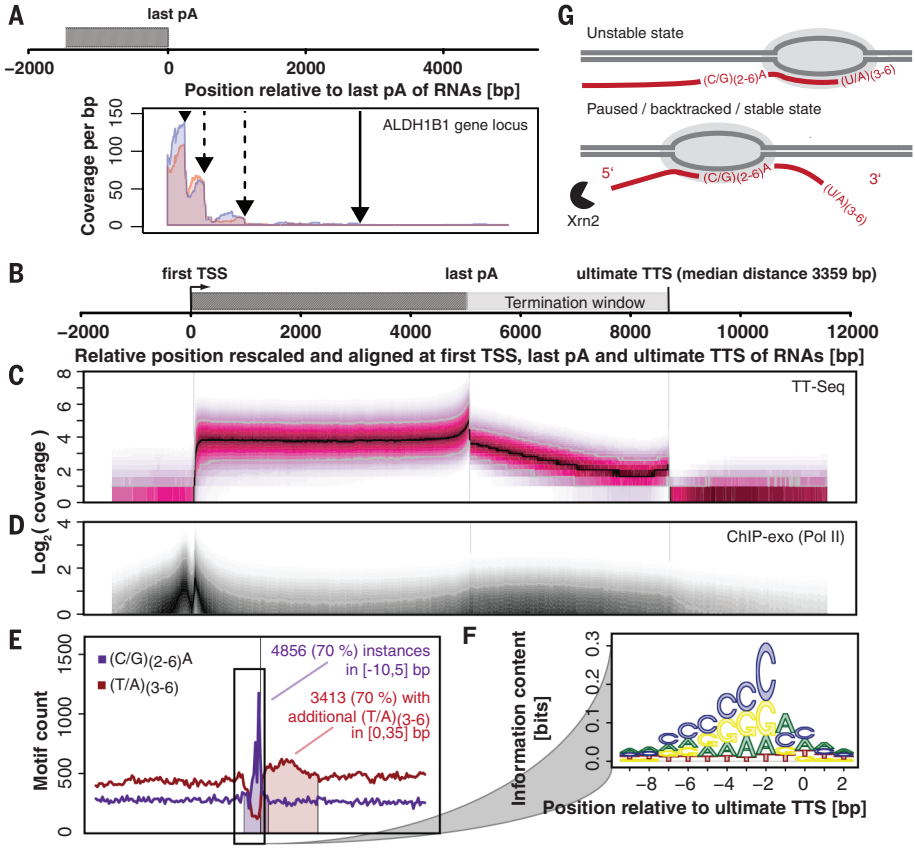

Fig. 4. Transcription termination sites. (**A**) TT-seq coverage for two replicates (red and blue) downstream of the pA site in the *ALDH1B1* gene locus. Arrows indicate TTSs obtained from segmentation (solid arrow, ultimate TTS). The last annotated pA site (per GENCODE) was aligned at zero. (**B**) Generic gene architecture. The first TSS was aligned at zero, and the last pA site was set at a rescaled distance of 5000 bp from the TSS (the real median distance is 24,079 bp for 6977 investigated genes) (*4*). The ultimate TTS is depicted at a median distance of 3359 bp from the last pA site (rescaled). (**C**) TT-seq coverage over quantiles 0.05 to 0.95 (pink area; black line, median), rescaled and aligned as in (B). (**D**) Pol II occupancy determined by the ChIP-exonuclease method (*23*), visualized as in (C) (black line, median; white line, upper quartile). (**E**) $(C/G)_{(2-6)}A$ and $(T/A)_{(3-6)}$ sequence count within ±100 bp of ultimate TTSs. (**F**) PWM (position weight matrix) logo representation of nucleotides at positions −9 to +2 around the ultimate TTS (position 0). (**G**) Predicted polymerase states at the T-rich stretch downstream of the TTS (top) and after backtracking to the TTS (bottom) (gray, DNA; red, RNA).

Transcription over the $(C/G)_{(2-6)}AN_x(T/A)_{(3-6)}$ sequence is predicted to destabilize the polymerase complex (*24*, *25*) because the melting temperature (*4*) of the DNA-RNA hybrid is low in the T/A-rich region (fig. S17). This may trigger backtracking and trap polymerase at the TTS (Fig. 4G). At the TTS, the hybrid is C/G-rich and stable, and RNA may be cleaved from its 3′ end to yield a terminal A residue. Polymerase can then be released from DNA by XRN2 (*21*, *22*). TT-seq has afforded insights into the determinants of human genome transcription and provides a complementary tool for transcriptome analysis.

REFERENCES AND NOTES

1. T. H. Jensen, A. Jacquier, D. Libri, *Mol. Cell* **52**, 473–484 (2013).
2. R. Andersson *et al.*, *Nature* **507**, 455–461 (2014).
3. M. D. Cleary, C. D. Meiering, E. Jan, R. Guymon, J. C. Boothroyd, *Nat. Biotechnol.* **23**, 232–237 (2005).
4. Materials and methods are available as supplementary materials on *Science* Online.
5. L. J. Core *et al.*, *Nat. Genet.* **46**, 1311–1320 (2014).
6. R. Kodzius *et al.*, *Nat. Methods* **3**, 211–222 (2006).
7. H. Kwak, N. J. Fuda, L. J. Core, J. T. Lis, *Science* **339**, 950–953 (2013).
8. A. Mayer *et al.*, *Cell* **161**, 541–554 (2015).
9. T. Nojima *et al.*, *Cell* **161**, 526–540 (2015).
10. L. J. Core, J. J. Waterfall, J. T. Lis, *Science* **322**, 1845–1848 (2008).
11. B. Zacher *et al.*, http://biorxiv.org/content/early/2016/03/06/041020 (2016).
12. J. Harrow *et al.*, *Genome Res.* **22**, 1760–1774 (2012).
13. S. Djebali *et al.*, *Nature* **489**, 101–108 (2012).
14. T. K. Kim *et al.*, *Nature* **465**, 182–187 (2010).
15. ENCODE Project Consortium, *Nature* **489**, 57–74 (2012).
16. R. A. Flynn, A. E. Almada, J. R. Zamudio, P. A. Sharp, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10460–10465 (2011).
17. M. Rabani *et al.*, *Cell* **159**, 1698–1710 (2014).
18. D. Kaida *et al.*, *Nature* **468**, 664–668 (2010).
19. M. G. Berg *et al.*, *Cell* **150**, 53–64 (2012).
20. A. E. Almada, X. Wu, A. J. Kriz, C. B. Burge, P. A. Sharp, *Nature* **499**, 360–363 (2013).
21. S. West, N. Gromak, N. J. Proudfoot, *Nature* **432**, 522–525 (2004).

22. M. Kim *et al.*, *Nature* **432**, 517–522 (2004).
23. B. J. Venters, B. F. Pugh, *Nature* **502**, 53–58 (2013).
24. R. Ashfield *et al.*, *EMBO J.* **13**, 5656–5667 (1994).
25. M. L. Kireeva, N. Komissarova, D. S. Waugh, M. Kashlev, *J. Biol. Chem.* **275**, 6530–6536 (2000).

ACKNOWLEDGMENTS

The sequencing data and the annotation file have been deposited in the Gene Expression Omnibus database under accession code GSE75792. We thank H. Blum, S. Krebs, and A. Graf (Laboratory for Functional Genome Analysis, Gene Center, Ludwig-Maximilians-Universität München) for help with sequencing and J. A. Feuillet, J. Soeding, A. Sawicka, and C. Bernecky for help and discussions. K.F. was supported by the Center for Innovative Medicine (CIMED) at Karolinska Institutet and by the Science for Life Laboratory (SciLifeLab) in Stockholm. C.D. was supported by a Deutsche Forschungsgemeinschaft (DFG) fellowship at the Graduate School of Quantitative Biosciences Munich. A.T. was supported by a German Federal Ministry of Education and Research (BMBF) e:Bio grant and by DFG grant SFB 680. J.G. was supported by the Bavarian Research Center for Molecular Biosystems and the Bundesministerium für Bildung und Forschung, Junior verbund in der Systemmedizin "mitOmics" (grant FKZ 01ZX1405A). P.C. was funded by the Advanced Grant TRANSIT of the European Research Council, the DFG, the Volkswagen Foundation, CIMED, and SciLifeLab.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6290/1225/suppl/DC1
Materials and Methods
Figs. S1 to S17
Tables S1 to S4
References (*26*–*39*)

2 December 2015; accepted 6 May 2016
10.1126/science.aad9841

NONHUMAN GENETICS

Genomic and archaeological evidence suggests a dual origin of domestic dogs

Laurent A. F. Frantz,[1]*† Victoria E. Mullin,[2]† Maud Pionnier-Capitan,[3,4] Ophélie Lebrasseur,[1] Morgane Ollivier,[3] Angela Perri,[5] Anna Linderholm,[1,6] Valeria Mattiangeli,[2] Matthew D. Teasdale,[2] Evangelos A. Dimopoulos,[1,7] Anne Tresset,[4] Marilyne Duffraisse,[3] Finbar McCormick,[8] László Bartosiewicz,[9] Erika Gál,[10] Éva A. Nyerges,[10] Mikhail V. Sablin,[11] Stéphanie Bréhard,[4] Marjan Mashkour,[4] Adrian Bălăşescu,[12] Benjamin Gillet,[3] Sandrine Hughes,[3] Olivier Chassaing,[3] Christophe Hitte,[13] Jean-Denis Vigne,[4] Keith Dobney,[14,15] Catherine Hänni,[3] Daniel G. Bradley,[2]* Greger Larson[1]*

The geographic and temporal origins of dogs remain controversial. We generated genetic sequences from 59 ancient dogs and a complete (28x) genome of a late Neolithic dog (dated to ~4800 calendar years before the present) from Ireland. Our analyses revealed a deep split separating modern East Asian and Western Eurasian dogs. Surprisingly, the date of this divergence (~14,000 to 6400 years ago) occurs commensurate with, or several millennia after, the first appearance of dogs in Europe and East Asia. Additional analyses of ancient and modern mitochondrial DNA revealed a sharp discontinuity in haplotype frequencies in Europe. Combined, these results suggest that dogs may have been domesticated independently in Eastern and Western Eurasia from distinct wolf populations. East Eurasian dogs were then possibly transported to Europe with people, where they partially replaced European Paleolithic dogs.

Dogs were the first domestic animal and the only animal domesticated before the advent of settled agriculture (*1*). Despite their importance in human history, no consensus has emerged regarding their geographic and temporal origins, or whether dogs were domesticated just once or independently on more than one occasion. Although several claims have been made for an initial appearance of dogs in the early upper Paleolithic [~30,000 years ago; e.g. (*2*)], the first remains confidently assigned to dogs appear in Europe ~15,000 years ago and in Far East Asia over 12,500 years ago (*1*, *3*). Although archaeologists remain open to the idea that there was more than one geographic origin for dogs [e.g. (*4*, *5*)], most genetic studies have concluded that dogs were probably domesticated just once (*6*), disagreeing on whether this occurred in Europe (*7*), Central Asia (*8*), or East Asia (*9*).

Recent paleogenetic studies have transformed our understanding of early human evolution [e.g. (*10*, *11*)]. We applied a similar approach to reconstruct the evolutionary history of dogs. We generated 59 ancient mitochondrial DNA (mtDNA) sequences from European dogs (from 14,000 to 3000 years ago) as well as a high-coverage nuclear genome (28x) of an ancient dog dated to ~4800 calendar years before the present (*12*) from the Neolithic passage grave complex of Newgrange (*Sí an Bhrú*) in Ireland. We combined our ancient sample with 80 modern publically available full genome sequences and 605 modern dogs (including village dogs and 48 breeds) genotyped on the CanineHD 170,000 (170 K) single-nucleotide polymorphism (SNP) array (*12*).

We first assessed characteristics of the Newgrange dog by typing SNPs associated with specific phenotypic traits and by inferring its level of inbreeding, compared to other breed and village dogs (*12*). Our results suggest that the degree of artificial selection and controlled breeding during the Neolithic was similar to that observed in modern free-living dogs. In addition, the Newgrange dog did not possess variants associated with modern breed-defining traits, including hair length or coat color. And although this dog was likely able to digest starch less efficiently than modern dogs, it was able to do so more efficiently than wolves (*12*).

A phylogenetic analysis based on 170 K SNPs revealed a deep split separating the modern Sarloos breed from other dogs (Fig. 1A). This breed, created in the 1930s in the Netherlands, involved breeding German Shepherds with captive wolves (*13*), thus explaining the breed's topological placement. The second deepest split [evident on the basis of both the 170 K SNP panel (Fig. 1A) and genome-wide SNPs (fig. S4)] separates modern East Asian and Western Eurasian (Europe and the Middle East) dogs. Moreover, the Newgrange dog clusters tightly with Western Eurasian dogs. We used principal components analysis (PCA), D statistics, and the program TreeMix (*12*) to further test this pattern. Each of these analyses unequivocally placed the Newgrange dog with modern European dogs (figs. S5 to S7). These findings demonstrate that the node separating the East Asian and Western Eurasian clades is older than the Newgrange individual, which was directly radiocarbon dated to ~4800 years ago.

Other nodes leading to multiple dog populations and breeds [including the basal breeds (*1*)

[1]The Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK. [2]Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland. [3]CNRS/ENS de Lyon, IGFL, UMR 5242 and French National Platform of Paleogenetics, PALGENE, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France/Université Grenoble Alpes, Laboratoire d'Ecologie Alpine (LECA), F-38000 Grenoble, France. [4]CNRS/Muséum National d'Histoire Naturelle/Sorbonne Universités, Archéozoologie, Archéobotanique: Sociétés, Pratiques et Environnement (UMR 7209), CP56, 55 rue Buffon, F-75005 Paris, France. [5]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [6]Department of Anthropology, Texas A&M University, College Station, TX 77843-4352, USA. [7]School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece. [8]School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, University Road, Belfast, Northern Ireland, UK. [9]Osteoarchaeological Research Laboratory, University of Stockholm, Stockholm, Sweden. [10]Archaeological Institute, Research Centre for the Humanities, Hungarian Academy of Sciences, Budapest, Hungary. [11]Zoological Institute, Russian Academy of Sciences, Universitetskaya Nab. 1, 199034 Saint-Petersburg, Russia [12]The National Museum of Romanian History, 12 Calea Victoriei, 030026 Bucharest, Romania. [13]Institut de Génétique et Développement de Rennes, CNRS-UMR6290, Université de Rennes 1, Rennes, France. [14]Department of Archaeology, School of Geosciences, University of Aberdeen, St. Mary's, Elphinstone Road, AB24 3UF, UK. [15]Department of Archaeology, Classics and Egyptology, University of Liverpool, 12-14 Abercromby Square, Liverpool L69 7WZ, UK. ***Corresponding author. Email: laurent.frantz@arch.ox.ac.uk (L.A.F.F.); greger.larson@arch.ox.ac.uk (G.L.); dbradley@tcd.ie (D.G.B.)** †These authors contributed equally to this work.

Editor's Summary

TT-Seq maps a transient transcriptome

RNA expression is related to protein abundance and cellular function. However, the amounts of RNA generated at any one time-point have been difficult to determine. Schwalb *et al.* developed a method, transient transcriptome sequencing (TT-Seq), to collect and sequence all RNA segments synthesized over 5 minutes. Because 5 minutes is not long enough to fully degrade even the most transient RNA, this method can detect the synthesis of most RNA without bias. Applying this method to human K562 cells, TT-Seq detected thousands of noncoding transcripts, providing a snapshot of RNA synthesis rates and RNA half-lives, and full-length maps of short-lived RNAs such as enhancers and short intergenic noncoding RNAs.

Science, this issue p. 1225