Polony Multiplex Analysis of Gene Expression (PMAGE) in Mouse Hypertrophic Cardiomyopathy
Jae Bum Kim, *et al.*
Science **316**, 1481 (2007);
DOI: 10.1126/science.1137325

The following resources related to this article are available online at
www.sciencemag.org (this information is current as of June 8, 2007):

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/316/5830/1481

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/316/5830/1481/DC1

This article **cites 28 articles**, 12 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/316/5830/1481#otherarticles

This article appears in the following **subject collections**:
Medicine, Diseases
http://www.sciencemag.org/cgi/collection/medicine

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

Polony Multiplex Analysis of Gene Expression (PMAGE) in Mouse Hypertrophic Cardiomyopathy

Jae Bum Kim,[1,2]* Gregory J. Porreca,[2]* Lei Song,[2] Steven C. Greenway,[2] Joshua M. Gorham,[2] George M. Church,[2] Christine E. Seidman,[1,2,3]† J. G. Seidman[2]†

We describe a sensitive mRNA profiling technology, PMAGE (for "polony multiplex analysis of gene expression"), which detects messenger RNAs (mRNAs) as rare as one transcript per three cells. PMAGE incorporates an improved ligation-based method to sequence 14-nucleotide tags derived from individual mRNA molecules. One sequence tag from each mRNA molecule is amplified onto a separate 1-micrometer bead, denoted as a polymerase colony or polony, and about 5 million polonies are arrayed in a flow cell for parallel sequencing. Using PMAGE, we identified early transcriptional changes that preceded pathological manifestations of hypertrophic cardiomyopathy in mice carrying a disease-causing mutation. PMAGE provided a comprehensive profile of cardiac mRNAs, including low-abundance mRNAs encoding signaling molecules and transcription factors that are likely to participate in disease pathogenesis.

One of the central goals in cell biology is to define the complete repertoire of RNA transcripts that drive essential physiologic processes and to understand how transcription changes with pathology. Several powerful technologies are available to study large-scale transcriptional changes, including microarrays and serial analysis of gene expression (SAGE) (*1–3*). However, these technologies have limitations that preclude comprehensive interrogation of gene expression. For example, microarray-based approaches have a limited ability to detect low-abundance RNAs and they can produce misleading results owing to cross-hybridization (*4*). Because SAGE involves dideoxy sequencing of concatenated cDNA ditags, this approach provides the statistical rigor of digital quantification and addresses some of the limitations associated with hybridization-based platforms (*5*). In practice however, the cost of SAGE experiments limits sampling depth, which diminishes sensitivity for interrogation of rare transcripts.

To obtain comprehensive mRNA profiles that include rare and potentially important transcripts expressed at levels as low as <1 copy per cell, we developed polony multiplex analysis of gene expression (PMAGE) (*6*). This method permits accurate quantitative assessment of mRNA expression because individual cDNA molecules are directly subjected to sequencing without antecedent library amplification, concatenation, or subcloning. In PMAGE, individual cDNA template molecules are clonally amplified onto each polony bead in millions of parallel, compartmen-

talized droplets formed in a water-in-oil emulsion (*7*). Polony sequence-by-ligation (SBL) (*8*) is used to provide an accurate, inexpensive, and multiplexed platform for high-throughput DNA tag sequencing. SBL takes advantage of the high discriminatory power of DNA ligase to iteratively label each bead with a fluorophore encoding the identity of a base within the template. Microscopy-based detection of fluorescence ligation events permits the assessment of as many as 5 million cDNA molecules per run, and digital quantification of tag data accommodates rigorous statistical analysis over a broad dynamic range of mRNA expression.

The critical first step in PMAGE is the manufacture of tag libraries (Fig. 1A). Double-stranded cDNA is synthesized from total RNA and bound to oligo(dT) ferromagnetic beads. cDNAs are then cleaved with an anchor endonuclease, Nla III, leaving a 4–base pair (bp) 3′ overhang, which is ligated to an adapter containing an Acu I type IIs endonuclease recognition site. Digestion with Acu I cleaves 16 bases from the nonpalindromic recognition sequence and releases the adapter with its 10-bp cDNA sequence tag (plus a 4-base CATG anchor sequence) (fig. S1). The tag is subsequently ligated to a reverse adapter containing a 2-bp 3′ degenerate overhang. From the total RNA, 10 μg generates sufficient library template, in principle, to perform up to 90 experiments, each with >4 million polonies [supporting online material (SOM), note S1]. This method of in vitro library construction ensures a faithful representation of mRNA levels because each mRNA molecule is converted into one cDNA template and one polony.

Because PMAGE requires a highly accurate sequencing methodology, we improved upon polony SBL, which has a median raw sequencing accuracy of 99.7% for reads placed against a reference genome (*8*). We optimized the performance of polony SBL through five specific modifications [detailed in SOM, figs. S2 to S5,

and notes S1 to S4]. The polony beads were bound directly to glass rather than embedded in an acrylamide matrix (Fig. 1B and SOM, note S2), allowing an arraying density two or three times that of SBL. This modification allowed beads to coat the gel-less glass surface as a monolayer in a single focal plane (Fig. 1C), which improved microscopy-based imaging, data yield, and signal coherence (fig. S5).

To assess the utility of PMAGE for obtaining comprehensive transcriptional profiles, we studied mouse myocardial tissues. Cardiac tissue poses a challenge for detection of rare transcripts, because of the heterogeneity of resident cell types and the considerable abundance of mRNAs encoding sarcomere and mitochondrial proteins (fig. S6). We constructed two PMAGE libraries from cardiac left ventricles of 8-week-old mice; one library was derived from wild-type mice, the other from αMHC[403/+] littermates that carry an R403Q missense mutation (in which glutamine replaces arginine at codon 403) in one allele of the endogenous cardiac α-myosin heavy-chain gene. R403Q is orthologous to a human mutation that causes hypertrophic cardiomyopathy (HCM). αMHC[403/+] mice recapitulate human HCM (*9*) with development of cardiac pathology after 25 weeks of age. We studied young mutant mice (denoted as prehypertrophic), to identify early transcriptional changes involved in triggering hypertrophy, increased myocardial fibrosis, and other protean manifestations of HCM. For comparison, we constructed SAGE libraries (*10*) from the same RNAs.

We produced PMAGE libraries from normal and mutant tissue using different primer-bound beads, so that both libraries could be loaded onto the same array and sequenced simultaneously. Primers bound to each set of beads contained a unique oligonucleotide sequence, which allowed data from the two libraries to be distinguished by hybridization analysis (Figs. 1, D and E, and SOM, notes S1 and S3). The two PMAGE libraries were assembled into one array consisting of 53.8 million beads (that included polonies and beads without tags). About 5.5 million polony beads (~10%) met criteria for containing clonal and well-amplified cDNA tags. Sequence reads from ~4.7 million polonies (~86%) yielded 14-bp tags, which is consistent with the predicted cleavage behavior of Acu I (fig. S1).

Sequencing errors in abundant tags are a source for false-positive detection of rare transcripts. To minimize this error, we implemented an approach similar to the strategy used to evaluate ~3.5 million human SAGE tags in a meta-analysis of 84 published SAGE libraries (*11*). The sequence of each PMAGE tag was compared with all other tags in the library for potential error-derived substitutions. If criteria were met for sequencing error (SOM, note S4), the tag was excluded. This correction removed 278,955 tags from 4.7 million PMAGE tags. We also excluded single-copy PMAGE tags (59,302) from the combined libraries, on the assumption that

[1]Cardiovascular Division, Brigham and Women's Hospital, Boston, MA 02115, USA. [2]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [3]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: seidman@genetics.med.harvard.edu (J.G.S.); cseidman@genetics.med.harvard.edu (C.E.S.)

random errors were more likely to generate singleton tags. If we assume that all excluded tags reflected sequencing errors, the maximum calculated tag error rate for PMAGE was 7.1%, a value that compared favorably with reported tag error rates (6.8 to 9.4%) from dideoxy sequencing of SAGE libraries (2, 11). The final yield was 4,402,432 filtered PMAGE tags, with roughly equal yields for the normal and mutant libraries (Table 1). The full lists of differentially expressed PMAGE tags (table S1), UniGene clusters (table S2), and transcription factors (table S3) are provided as SOM (descriptions in note S4).

Technical replicates of a PMAGE library were highly reproducible ($R = 0.9986$; fig. S7), which compared favorably with previously published data on SAGE ($R^2 = 0.96$, $R \sim 0.98$) (12). Base composition in PMAGE data did not reveal noticeable base bias, and distributions did not deviate from expected (fig. S8). Comparison of tag counts in SAGE and PMAGE libraries constructed from the same RNA were highly correlated ($R = 0.8876$), which provided an independent validation of PMAGE data (Fig. 2A). However, data yields were markedly different between these two technologies (Table 1 and fig. S6). Using PMAGE, we sequenced more total tags by a factor of 31 than with SAGE, which included four times as many unique tags, and we identified twice as many genes as in SAGE. PMAGE identified expression of more than 10,000 genes that were undetected by SAGE.

Analyses of >2 million PMAGE tags per library provided sufficient sampling depth to capture rare transcripts (fig. S9) and to provide redundant coverage of the estimated 300,000 transcripts per cell (13). PMAGE data demonstrated a wide dynamic range in cardiac mRNA expression, with transcript counts ranging from 0.3 tags per cell [e.g., *Vgll2* (vestigial-like–2 homolog), (Table 2)] to ~29,000 tags per cell [e.g., cytochrome c oxidase III (fig. S6)]. Among 68,762 unique tags in the wild-type PMAGE library, 447 were expressed at >60 copies per cell and accounted for 65% of all mRNA molecules. Of the unique tags, 99% were expressed at low abundance (<60 copies per cell). Among 1337 transcription factors identified in the wild-type left ventricle PMAGE library (SOM, note S4), only 13 were present at >60 copies per cell; the remainder were expressed at low abundance.

The comparison of PMAGE profiles from wild-type and prehypertrophic $\alpha MHC^{403/+}$ left ventricles (Fig. 2B) demonstrated a surprisingly
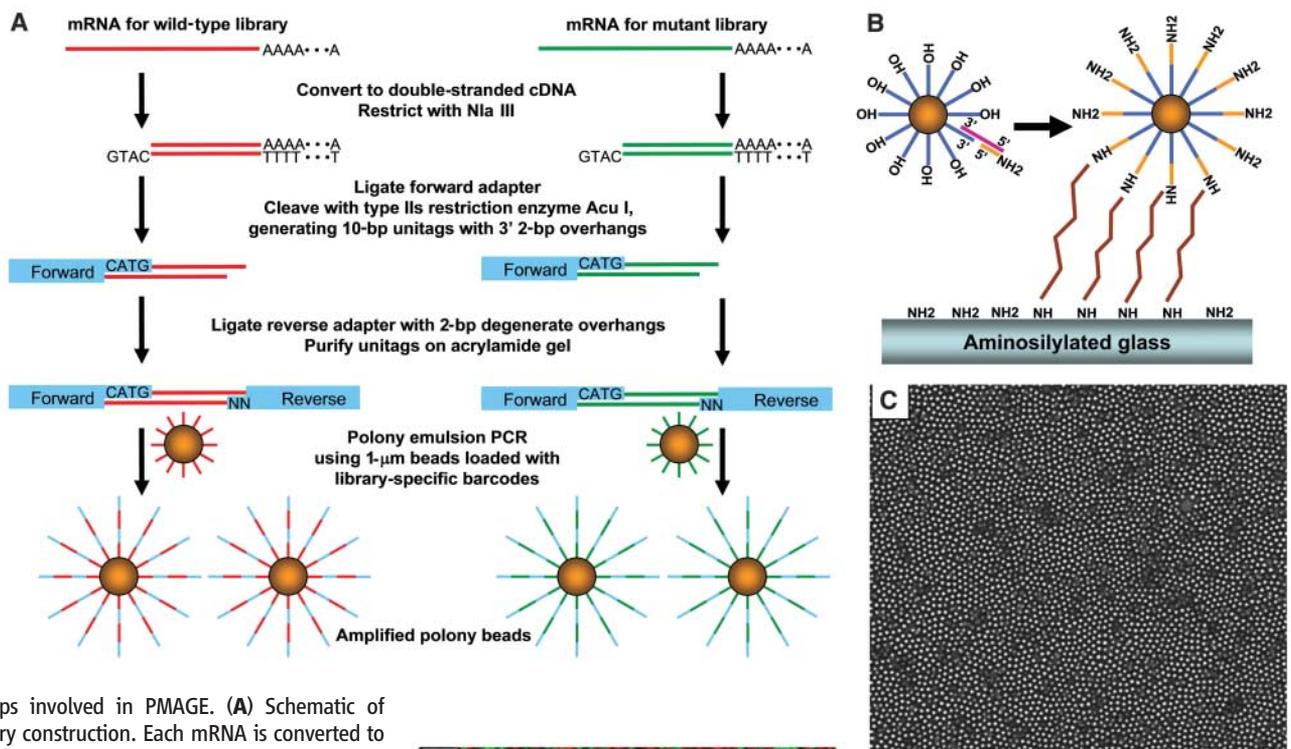
Fig. 1. Steps involved in PMAGE. (**A**) Schematic of PMAGE library construction. Each mRNA is converted to cDNA molecules (red orange and green bars), cleaved with anchor endonuclease, ligated to forward adapters, cleaved with tagging enzyme Acu I, ligated with reverse adapters, and amplified onto beads carrying primers with sequences that match the forward adapters. With emulsion PCR, each polony bead carries clonal molecules encoding one cDNA tag. (**B**) Polony bead capping and attachment to glass slide. Polony beads carry DNA templates (blue bars) that have terminal 3′ OH groups. 3′ Termini are capped by ligation, by using annealed bridging primers (purple bar), with oligonucleotides that contain 3′ amines (yellow orange bars). 3′ Amines on polony beads are cross-linked to aminosilylated glass with amino-ester bridges (jagged lines) (SOM, note S2), providing a gel-less milieu for sequence chemistry. (**C**) Bright-field image of the cardiac PMAGE library arrayed in a dense monolayer on a glass slide (each spherical object is a 1-μm polony bead). This gel-less platform enhances access to sequencing reagents and provides a single focal plane for microscopy-based imaging during polony sequencing. (**D**) Pseudofluorescence imaging of wild-type (red) and $\alpha MHC^{403/+}$ (green) polonies in a PMAGE sequencing array hybridized with library-specific fluorescence oligonucleotides (SOM, notes S2 and S3). (**E**) A fluorescence intensity plot of 10,000 polonies parsed into the wild-type (Cy3) and $\alpha MHC^{403/+}$ (Cy5) libraries that were simultaneously sequenced.

large number of differentially expressed mRNAs, given the absence of overt pathology in mutant hearts. There were 1486 unique PMAGE tags, corresponding to 706 genes, that were significantly up-regulated or down-regulated ($P < 0.01$) in the prehypertrophic ventricle (SOM, note S4 and tables S1 and S2). These widespread changes affected components in pathways involved in myocardial excitation-contraction coupling, Ca^{++} homeostasis, and energy metabolism (fig. S11). The vast majority of these differentially regulated transcripts escaped detection by SAGE, which identified only 53 genes with significantly ($P < 0.01$) altered expression.

We assessed 20 genes that PMAGE identified as both low abundance (54 to <0.3 per cell) and differentially expressed in prehypertrophic αMHC$^{403/+}$ hearts. Real-time reverse transcription polymerase chain reaction (RT-PCR) confirmed PMAGE results for 19 of 20 genes (fig. S10 and Table 2). Some of these genes [e.g., *Nfkbie* (nuclear factor kappa light polypeptide gene enhancer in B cells inhibitor epsilon *Nr1h3* (nuclear receptor subfamily 1, group h, member 3), and the gene for the retinoid-responsive nuclear hormone receptor LXRα (liver X receptor α)] have unknown roles in HCM pathogenesis, whereas other genes encode proteins that may stimulate the protean manifestations of HCM: myocyte enlargement, increased cardiac fibrosis, and abnormal calcium homeostasis (*14*). *Hod* (homeobox-only protein), *Hand2* (hand and neural crest derivatives 2), *Vgll2*, and *Egr3* (early growth response–3) are transcriptional regulators implicated in myocyte specification and growth during development (*15–19*). Differential expression of these transcripts in prehypertrophic αMHC$^{403/+}$ hearts may indi-

Table 1. Cardiac left ventricular PMAGE and SAGE libraries. Total tags are filtered tag counts. PMAGE libraries were filtered to exclude tags with potential sequence errors (SOM, note S4) and singleton tags from the combined dataset. SAGE libraries were filtered to exclude linker sequences. Unique tags assigned to the same UniGene cluster or gene symbol were combined into one vector.

	PMAGE		SAGE	
	Wild type	αMHC$^{403/+}$	Wild type	αMHC$^{403/+}$
Total tags	2,042,192	2,360,240	70,731	69,268
Unique tags	68,762	72,143	17,896	18,412
Unique genes	17,931	18,311	7,783	8,027

Table 2. Examples of low-abundance transcripts with changed expression in prehypertrophic αMHC$^{403/+}$ hearts. Low-abundance genes (<60 copies per cell) are listed in order of decreasing transcript abundance in the wild-type left ventricle. *Nppa*, an established hypertrophy-regulated gene that encodes atrial natriuretic peptide (*30*), is the only transcript among these 12 genes that was detected by SAGE as differentially expressed. Tag counts are followed by calculated transcript copies per cell (provided in parentheses). PMAGE tag counts were normalized to 2 million tags per library; actual SAGE tag counts are presented. Differential expression of each gene was confirmed by quantitative RT-PCR (fig. S10).

	PMAGE			SAGE		
Gene symbol	Wild type	αMHC$^{403/+}$	P value	Wild type	αMHC$^{403/+}$	P value
Hod	357 (54)	255 (38)	3.25E−05	6 (25)	1 (4)	n.s.
Hand2	342 (51)	282 (42)	1.52E−02	16 (68)	21 (91)	n.s.
Abcc9	302 (45)	103 (15)	4.88E−24	9 (38)	4 (17)	n.s.
Ctgf	233 (35)	476 (71)	4.17E−20	14 (59)	23 (100)	n.s.
Nppa	183 (27)	744 (112)	8.76E−81	15 (64)	66 (286)	2.91E−09
Sln	183 (27)	39 (6)	8.84E−24	2 (8)	0 (0)	n.s.
Postn	43 (6)	132 (20)	7.41E−12	1 (4)	3 (13)	n.s.
Tgfβ1	19 (3)	41 (6)	4.56E−03	0 (0)	0 (0)	n.s.
Nr1h3	2 (0.3)	12 (2)	7.47E−03	2 (8)	1 (4)	n.s.
Vgll2	2 (0.3)	12 (2)	7.47E−03	0 (0)	0 (0)	n.s.
Nfkbie	1 (0.2)	11 (2)	3.46E−03	1 (4)	0 (0)	n.s.
Egr3	0 (0)	7 (1)	7.88E−03	0 (0)	0 (0)	n.s.

Fig. 2. PMAGE accurately detects mRNA abundance and differentially expressed transcripts. (**A**) Scatter plot (logarithmic scale) of tags in wild-type left ventricular SAGE (total tag count, 70,731) and PMAGE (normalized to a total count of 70,731) libraries derived from the same RNA. All tags with counts ≥1 copy were compared. $R = 0.8858$. (**B**) Scatter plot (logarithmic scale) of PMAGE tag abundances (normalized to total count of 2 million tags per library) from wild-type and αMHC$^{403/+}$ hearts (blue dots $P \geq 0.01$, red dots $P < 0.01$). $R = 0.9950$.
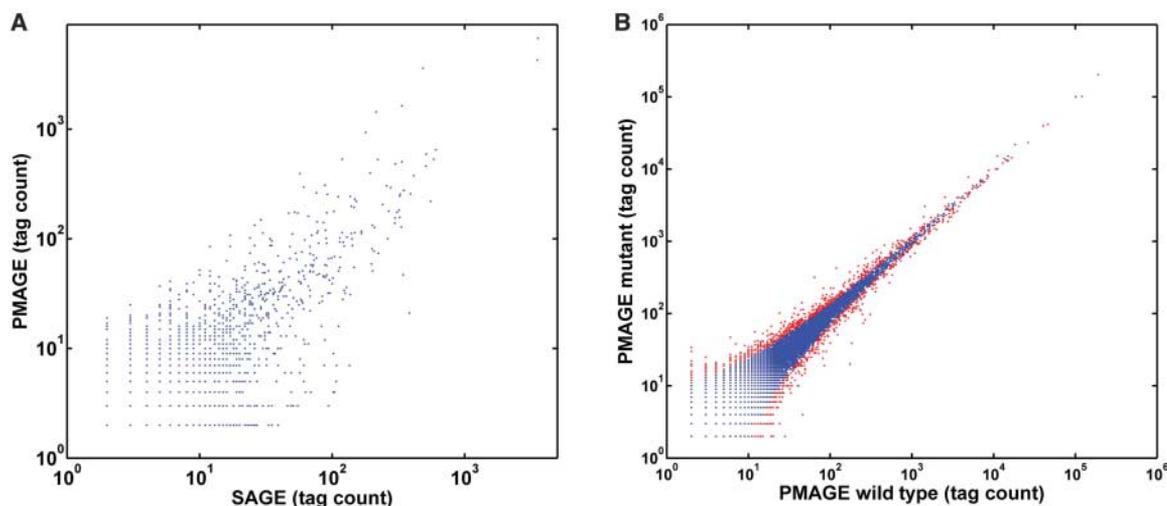
cate that these molecules also function in early growth responses to sarcomere dysfunction.

Myocardial fibrosis is characteristic of advanced HCM pathology and contributes to impaired cardiac relaxation, heart failure, arrhythmias, and sudden death (20, 21). The increased expression of *Tgfb1* (transforming growth factor–β1), *Ctgf* (connective tissue growth factor) and *Postn* (periostin), potent regulators of fibrosis and collagen deposition (22–24), in prehypertrophic ventricles indicates early activation of this pathway, which raises the possibility that fibrosis is not an advanced secondary phenomenon, but a primary contributor to myocardial dysfunction.

Impaired relaxation is the fundamental physiologic abnormality in HCM (25). Cardiac relaxation and contraction reflects Ca^{++} cycling between the sarcoplasmic reticulum and the sarcomere in cardiomyocytes. Ca^{++} uptake into the sarcoplasmic reticulum occurs via sarcoplasmic reticulum Ca^{++} transport adenosine triphosphatase (ATPase) (SERCA2a/Atp2a2), which is regulated by phospholamban (Pln) and sarcolipin (Sln) (26). Transcripts encoding each of these proteins were significantly decreased in prehypertrophic hearts (fig. S11), which may directly account for the early impairment in cardiac relaxation previously observed in this model (27). Downregulation of *Abcc9* [adenosine triphosphate (ATP)–binding cassette subfamily C member 9], which encodes SUR2, suggested another mechanism for Ca^{++} imbalance in prehypertrophic hearts. SUR2 is the ATPase-regulatory subunit of the inwardly rectifying cardiac K_{ATP} channel, which balances Ca^{++} homeostasis with energetic demands (28); *Abcc9*-null mice develop arrhythmias and myocardial calcium overload (29).

Notably, PMAGE also revealed significant ($P < 0.01$) differences in the expression of genes encoding 29 transcription factors between wild-type and prehypertrophic $\alpha MHC^{403/+}$ hearts (table S3). The biologic processes evoked by these molecules are likely to be considerable. By interrogating the temporal and spatial expression of these transcription factors, we can potentially dissect the networks activated in this cardiomyopathy, which, in turn, should help identify new molecular targets for therapeutic intervention.

In summary, PMAGE profiling provided reproducible, large-scale transcript identification, with sequence accuracy comparable to SAGE, and greater sensitivity for quantification of rare transcripts. We estimate that sampling ~2 million tags provides comprehensive assessment of most mRNAs (fig. S9); nevertheless, the current PMAGE platform has the capacity to read more than 4 million tags per experiment. Thus, PMAGE can be used for very deep sampling of one library or analyses of multiple libraries simultaneously by adapting polony beads that contain unique sequence identifiers. PMAGE offers several advantages over other currently available transcription profiling methods at a potentially lower cost (fig. S12). We anticipate that PMAGE studies will help further define mRNA regula-

tory networks that orchestrate critical cellular processes in healthy and diseased tissues.

References and Notes
1. G. A. Churchill, *Nat. Genet.* **32** (suppl.), 490 (2002).
2. V. E. Velculescu et al., *Cell* **88**, 243 (1997).
3. M. J. Holland, *J. Biol. Chem.* **277**, 14363 (2002).
4. S. Draghici, P. Khatri, A. C. Eklund, Z. Szallasi, *Trends Genet.* **22**, 101 (2006).
5. S. Audic, J. M. Claverie, *Genome Res.* **7**, 986 (1997).
6. Materials and methods are available as supporting material on *Science* Online.
7. D. Dressman, H. Yan, G. Traverso, K. W. Kinzler, B. Vogelstein, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8817 (2003).
8. J. Shendure et al., *Science* **309**, 1728 (2005).
9. A. A. Geisterfer-Lowrance et al., *Science* **272**, 731 (1996).
10. S. Blackshaw, J. B. Kim, B. St. Croix, K. Polyak, in *Current Protocols in Molecular Biology*, F. M. Ausubel, Ed. (Greene Publishing Associates and Wiley-Interscience, New York, 2002), pp. v (loose leaf).
11. V. E. Velculescu et al., *Nat. Genet.* **23**, 387 (1999).
12. S. Dinel et al., *Nucleic Acids Res.* **33**, e26 (2005).
13. N. D. Hastie, J. O. Bishop, *Cell* **9**, 761 (1976).
14. J. G. Seidman, C. Seidman, *Cell* **104**, 557 (2001).
15. F. Chen et al., *Cell* **110**, 713 (2002).
16. C. H. Shin et al., *Cell* **110**, 725 (2002).
17. D. Srivastava, P. Cserjesi, E. N. Olson, *Science* **270**, 1995 (1995).
18. T. Maeda, D. L. Chapman, A. F. Stewart, *J. Biol. Chem.* **277**, 48889 (2002).
19. Y. Albert et al., *J. Cell Biol.* **169**, 257 (2005).
20. A. M. Varnava, P. M. Elliott, N. Mahon, M. J. Davies, W. J. McKenna, *Am. J. Cardiol.* **88**, 275 (2001).
21. K. M. Harris et al., *Circulation* **114**, 216 (2006).
22. P. J. Lijnen, V. V. Petrov, R. H. Fagard, *Mol. Genet. Metab.* **71**, 418 (2000).
23. M. S. Ahmed et al., *J. Mol. Cell. Cardiol.* **36**, 393 (2004).
24. R. A. Norris et al., *J. Cell. Biochem.* **101**, 695 (2007).
25. B. J. Maron, *JAMA* **287**, 1308 (2002).
26. M. Asahi et al., *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9199 (2004).
27. D. Georgakopoulos et al., *Nat. Med.* **5**, 327 (1999).
28. M. Bienengraeber et al., *Nat. Genet.* **36**, 382 (2004).
29. L. V. Zingman et al., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13278 (2002).
30. R. T. Lee et al., *J. Clin. Invest.* **81**, 431 (1988).
31. We thank J. Shendure, S. Barr, S. DePalma, M. Maida, P. Teekakirikul, S. Blackshaw, Z. Arany, J. Loscalzo, and B. Vogelstein for advice and technical assistance. Funded by grants from NIH and Howard Hughes Medical Institute.

RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription

Philipp Kapranov,[1] Jill Cheng,[1] Sujit Dike,[1] David A. Nix,[1] Radharani Duttagupta,[1] Aarron T. Willingham,[1] Peter F. Stadler,[2] Jana Hertel,[2] Jörg Hackermüller,[3] Ivo L. Hofacker,[4] Ian Bell,[1] Evelyn Cheung,[1] Jorg Drenkow,[1] Erica Dumais,[1] Sandeep Patel,[1] Gregg Helt,[1] Madhavan Ganesh,[1] Srinka Ghosh,[1] Antonio Piccolboni,[1] Victor Sementchenko,[1] Hari Tammana,[1] Thomas R. Gingeras[1]*

Significant fractions of eukaryotic genomes give rise to RNA, much of which is unannotated and has reduced protein-coding potential. The genomic origins and the associations of human nuclear and cytosolic polyadenylated RNAs longer than 200 nucleotides (nt) and whole-cell RNAs less than 200 nt were investigated in this genome-wide study. Subcellular addresses for nucleotides present in detected RNAs were assigned, and their potential processing into short RNAs was investigated. Taken together, these observations suggest a novel role for some unannotated RNAs as primary transcripts for the production of short RNAs. Three potentially functional classes of RNAs have been identified, two of which are syntenically conserved and correlate with the expression state of protein-coding genes. These data support a highly interleaved organization of the human transcriptome.

A large fraction of the noncoding part of a eukaryotic genome is used to make RNA that is sufficiently stable in a cell to be detected by different technological approaches (1–4). The biological significance of this pervasive transcription is unclear and controversial. One possibility is that only very short regions of such unannotated RNA are biologically relevant (5). In-depth characterization of RNAs as to their subcellular compartmentalization, size, modifications, and genomic origins can potentially provide clues to their functions. This study reports two general observations derived from

the maps of nuclear and cytosolic polyadenylated [poly(A)$^+$] RNAs longer than 200 nucleotides (nt) (long RNAs, lRNAs) and whole-cell RNAs less than 200 nt (short RNAs, sRNAs) over the entire nonrepetitive portion of the human ge-

[1]Affymetrix Laboratory, Affymetrix, Inc., 3420 Central Expressway, Santa Clara, CA, 95051, USA. [2]University of Leipzig, Department of Computer Science, Leipzig, Germany. [3]Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany. [4]Institute for Theoretical Chemistry, University of Vienna, Austria.

*To whom correspondence should be addressed. E-mail: tom_gingeras@affymetrix.com