The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing

Ugrappa Nagalakshmi, *et al.*

Science **320**, 1344 (2008);
DOI: 10.1126/science.1158441

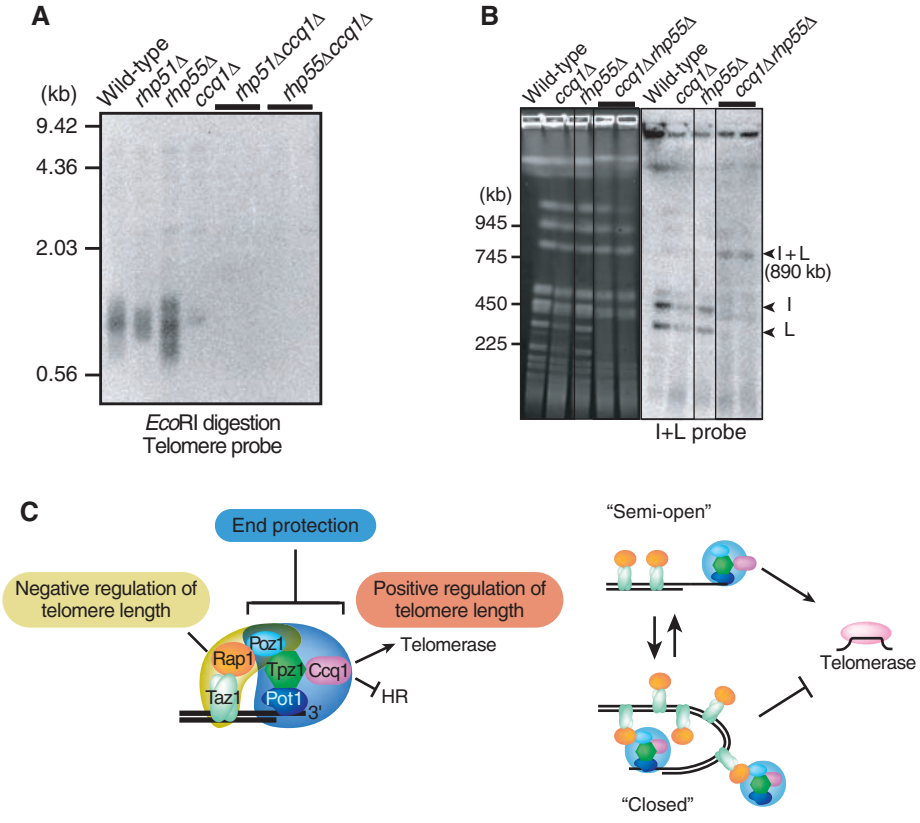

Fig. 4. Ccq1 inhibits HR at telomeres. (**A**) Telomeres are lost in *ccq1Δrhp51Δ* and *ccq1Δrhp55Δ*. (**B**) HR prevents chromosome circularization in *ccq1Δ*. Not I–digested chromosomal DNAs were fractionated by PFGE, stained with ethidium bromide (left), and subjected to Southern hybridization (right). (**C**) Model of Pot1-complex–mediated switching between closed and semi-open telomere configurations. The Pot1 complex exists at telomeres in two distinguishable modes (closed and semi-open configurations) that regulate telomerase in opposite manners. In both cases, Ccq1 and Poz1 protect telomeres from degradation, HR, and fusion.

plete loss of telomeric DNA and chromosome circularization (Fig. 4, A and B), indicating that terminal DNAs in *ccq1Δ* cells are maintained via HR. Thus, Ccq1 protects telomeres from HR. In *ccq1Δ* cells, the deregulated HR activity at telomeres contributes to the survival of cells without circular chromosome formation.

We propose that the Pot1 complex exists at telomeres in two distinguishable modes. When telomeres are relatively long, the Pot1 complex associates with ds telomeric DNA-Taz1-Rap1 via Poz1 because of a high concentration of local Taz1-Rap1 proteins, leading to inhibition of telomerase action (a closed configuration). When telomeres shorten, the Pot1 complex is dissociated from Taz1-Rap1, facilitating telomerase action aided by Ccq1 (a semi-open configuration). The structure and function of fission yeast telomeres revealed in this study highlight the conservation of major features of the shelterin complex across a wide evolutionary distance (fig. S12). We also predict that shelterin functions by recruiting unidentified effector molecules in higher eukaryotes, similar to the recruitment of Ccq1 in fission yeast.

References and Notes

1. D. E. Gottschling, V. A. Zakian, *Cell* **47**, 195 (1986).
2. C. M. Price, T. R. Cech, *Genes Dev.* **1**, 783 (1987).
3. M. P. Horvath, V. L. Schweiker, J. M. Bevilacqua, J. A. Ruggles, S. C. Schultz, *Cell* **95**, 963 (1998).
4. P. Baumann, T. R. Cech, *Science* **292**, 1171 (2001).
5. M. Lei, E. R. Podell, P. Baumann, T. R. Cech, *Nature* **426**, 198 (2003).
6. L. Wu *et al.*, *Cell* **126**, 49 (2006).
7. D. Hockemeyer, J. P. Daniels, H. Takai, T. de Lange, *Cell* **126**, 63 (2006).
8. F. Wang *et al.*, *Nature* **445**, 506 (2007).
9. H. Xin *et al.*, *Nature* **445**, 559 (2007).
10. T. de Lange, *Genes Dev.* **19**, 2100 (2005).
11. D. Loayza, T. de Lange, *Nature* **423**, 1013 (2003).
12. T. Sugiyama *et al.*, *Cell* **128**, 491 (2007).
13. M. R. Flory, A. R. Carson, E. G. Muller, R. Aebersold, *Mol. Cell* **16**, 619 (2004).
14. T. M. Nakamura, J. P. Cooper, T. R. Cech, *Science* **282**, 493 (1998).
15. T. Naito, A. Matsuura, F. Ishikawa, *Nat. Genet.* **20**, 203 (1998).
16. D. Hockemeyer *et al.*, *Nat. Struct. Mol. Biol.* **14**, 754 (2007).
17. J. P. Cooper, E. R. Nimmo, R. C. Allshire, T. R. Cech, *Nature* **385**, 744 (1997).
18. J. Kanoh, F. Ishikawa, *Curr. Biol.* **11**, 1624 (2001).
19. Y. Chikashige, Y. Hiraoka, *Curr. Biol.* **11**, 1618 (2001).
20. J. E. Croy, E. R. Podell, D. S. Wuttke, *J. Mol. Biol.* **361**, 80 (2006).
21. Materials and methods are available as supporting material on *Science* Online.
22. C. H. Haering, T. M. Nakamura, P. Baumann, T. R. Cech, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6367 (2000).

23. We thank M. Tamura and A. Hagiwara for technical assistance; P. Baumann, T. M. Nakamura, M. Yanagida, M. Shirakawa, and M. Katahira for discussions and materials; and A. Katayama and M. Sasaki for assistance. This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas (J.K.) and for Cancer Research (F.I.) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Supporting Online Material

www.sciencemag.org/cgi/content/full/320/5881/1341/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S3
References

4 January 2008; accepted 10 April 2008
10.1126/science.1154819

The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing

Ugrappa Nagalakshmi,[1]* Zhong Wang,[1]* Karl Waern,[1] Chong Shou,[2] Debasish Raha,[1] Mark Gerstein,[2,3] Michael Snyder[1,2,3]†

The identification of untranslated regions, introns, and coding regions within an organism remains challenging. We developed a quantitative sequencing-based method called RNA-Seq for mapping transcribed regions, in which complementary DNA fragments are subjected to high-throughput sequencing and mapped to the genome. We applied RNA-Seq to generate a high-resolution transcriptome map of the yeast genome and demonstrated that most (74.5%) of the nonrepetitive sequence of the yeast genome is transcribed. We confirmed many known and predicted introns and demonstrated that others are not actively used. Alternative initiation codons and upstream open reading frames also were identified for many yeast genes. We also found unexpected 3′-end heterogeneity and the presence of many overlapping genes. These results indicate that the yeast transcriptome is more complex than previously appreciated.

A major challenge in genomics is to identify all genes and exons and their boundaries. This information is crucial in order to understand the functional elements of the genome and determine when they are expressed and how they are regulated. Often, genes are identified

through the presence of large open reading frames (ORFs), sequence conservation, or cDNA probing of DNA tiling microarrays (*1*–*5*). These methods often fail to identify short exons, do not precisely reveal the boundaries of untranslated regions (UTRs), and/or have high false-positive rates.

In order to better map the transcribed regions of the yeast genome, we developed the RNA-Seq approach presented in Fig. 1. Briefly, polyadenylate [poly(A)] RNA was isolated from yeast cells grown in rich media (*6*) and used to generate double-stranded cDNA by reverse transcription with either random hexamers or oligo(dT) primers. The double-stranded cDNA was fragmented and subjected to high-throughput Illumina sequencing in which 35 base pairs (bp) of sequence were determined from one end of each fragment. Two technical and two biological replicates were performed for each hexamer and oligo(dT)-primed cDNA sample for a total of 15,787,335 and 14,125,182 reads, respectively. These sequence reads were analyzed with an algorithm that maps unique (that is, single-copy) sequences to the genome and allows for up to two mismatches (*6*). Of 29,912,517 total reads, 15,870,540 (56%) were mapped to unique genomic regions (fig. S1A and table S1).

We assessed our results by several criteria. First, none of the 29,912,517 sequence reads matched the deleted 3.5-kb regions of the genome (Fig. 1C), and very few, if any, matched the nontranscribed centromeres (fig S1B) (*6*); thus, our method is specific. Second, our replicates were in close agreement with one another, having a 0.99 Pearson correlation coefficient for technical replicates and a 0.93 to 0.95 coefficient for biological replicates (fig. S2). The data generated from hexamers and oligo(dT) primers also were closely correlated (0.97) and showed similar patterns of expression (fig. S2). Therefore, we merged all of these data sets, and the subsequent analyses were performed using the merged data.

RNA-Seq analyses revealed extensive expression of the whole yeast genome (Fig. 2A); 74.5% of the genome was expressed as RNA-Seq tags (Fig. 2B). We detected more reads from the 3′ ends than from the 5′ ends of annotated genes (fig. S3), presumably due to the enrichment of 3′ sequences during poly(A) purification as well as enhanced priming at the 3′ ends. Despite this bias, the deep sequencing allowed the detection of signals across the entire gene. Overall, 85% of the bases detected as expressed by RNA-Seq overlapped with those found with DNA-tiling microarrays (*7*).

We investigated the overall transcriptional activity and found that 4666 of the 5099 annotated ORFs (91.5%) in the Saccharomyces Genome Database (SGD) were expressed as tags above background (*6*). For this analysis, we removed 1178 ORFs whose 3′ ends lie within 100 bp of one another and whose transcripts might overlap. In addition, 327 ORFs that were not unique in their 3′ ends were not analyzed. We observed high expression for 20% of the genes; Gene Ontology (GO) analysis revealed that genes involved in biosynthetic pathways and ion transport were specifically enriched in the highly expressed category ($P < 2.3 \times 10^{-58}$; see table S2 for a complete list). Medium and low expression levels were observed for 39 and 33% of the genes, respectively. As expected, we did not detect the expression of genes involved in meiosis, mating, cell differentiation, sugar transport, or vitamin metabolism, the functions of which are not required during vegetative growth (*8*).

The majority of yeast genes have been annotated primarily with ORFs and, to a lesser extent, with cDNA sequencing (*9*); thus, the 5′ and 3′ boundaries and UTRs of most yeast genes have not been precisely defined. To map the 5′ ends of genes with RNA-Seq, the 5′ ends of 1331 genes were first determined by generating and sequencing rapid amplification of cDNA ends (RACE) polymerase chain reaction (PCR) products (table S3). We then used 125 RACE ends to optimize parameters for determining 5′ ends by searching RNA-Seq data for a sharp signal reduction in the transcribed region; applying these parameters revealed the 5′ boundary regions for 4665 transcribed genes. Genes with very low levels of expression were excluded from the analysis. Comparison of these results with 1025 boundaries mapped with 5′ RACE showed that both methods identified 5′ boundaries within 50 bp of one another for 786 genes (77.9%) (Fig. 3A). Combining the 5′ RACE results with the RNA-Seq results defined the 5′ boundaries of 4835 yeast genes (Fig. 3B). The median length of 5′ UTRs was 50 bp with a range of 0 to 990 bp (Fig. 3A, top right). Two hundred forty-one genes contained a start codon (ATG) less than 10 bp from the 5′ end; we do not know if these ATGs represent true initiation codons.

We also globally mapped the 3′ boundaries of yeast genes by searching for a rapid transition in the RNA-Seq signal as well as by identifying end tags with poly(A) sequences containing a novel stretch of three or more consecutive As lying next to a genomic yeast sequence (*6*). Using these methods, we mapped the 3′ boundaries of 5212 transcribed genes and deduced the transcribed strand (Fig. 3, C and D, and table S4). The end tags allowed the precise assignment of 3′ boundaries even when transcripts were overlapping at their 3′ ends (Fig. 4). We found evidence that the transcription of a large number of yeast genes overlaps with transcription from the other strand. Of 4646 verified expressed ORFs, 275 transcribed pairs (11.8% of expressed genes) contained overlapping 3′ ends. Pervasive overlapping transcripts may be unique to *Saccharomyces cerevisiae* and other organisms lacking microRNA/small interfering RNA–processing components that might otherwise degrade double-stranded RNAs. Moreover, overlapping transcription at 3′ ends could be a form of gene regulation by which neighboring genes can potentially influence (positively or negatively) the expression of one another.

The median length of yeast 3′ UTRs is 104 bp, with a range of 0 to 1461 bp (Fig. 3A). Although most yeast genes have a single precise 3′ end (within 1 to 2 bp), many genes had heterogeneous 3′-end sequences. These occurred either in localized regions (usually 2 to 10 bp), suggesting some variability in 3′-end processing at the poly(A) signal, or, in 540 genes, in multiple peaks greater than 10 bp apart, suggesting that more than one poly(A) site is used (fig. S4).

In yeast, the first ATG at the 5′ end of an ORF is usually annotated as the start codon. However, in some databases the second ATG is annotated when the predicted amino-terminal protein–coding sequence is not conserved (*10*, *11*). RNA-Seq analysis revealed 35 genes with 5′ ends upstream of an ATG that is 5′ and in-frame to the annotated ATG initiation codon, suggesting that these proteins are potentially longer than predicted. We also found 29 genes whose 5′ end is located downstream of the annotated ATG, suggesting that these proteins are shorter than previously predicted; the 5′ ends of four of the latter genes were confirmed by RACE sequencing.

We also examined our data for the presence of introns; in yeast, introns are typically identified on the basis of sequence conservation, microarray analyses, or other studies. We detected 240 of 306 known introns with an algorithm that detects introns through the presence of discontinuous sequences whose boundaries contain GT and AG/AC and also detects sequence tags that span the intron boundary, which indicates that transcripts from these genes are spliced. For the 66 introns not validated with the algorithm, we examined 30 whose genes were expressed. In four cases, the intron sequences were clearly expressed at levels similar to those of the adjacent exons (Fig. 3E and fig. S5), and unspliced products, but not splice junction tags, were identified. Thus, transcripts from these genes are probably not spliced at appreciable levels in vegetative cells [see also (*12*)]. In two of these cases, the presence or lack of an intron affected the predicted protein sequence.

Recent analysis predicts that many eukaryotic 5′ UTRs may contain upstream ORFs (uORFs) (*13*). uORFs have been shown to regulate protein expression (*14*) and mRNA degradation (*15*) and have been annotated for 17 yeast genes. Our data predict uORFs upstream of the start codon for 321 (6%) of the yeast gene transcripts (Fig. 5 and table S6). GO analysis of these genes revealed that genes encoding DNA-binding proteins [$P < 0.0027$, false discovery rate (FDR)–adjusted] and anatomical structure and development (for example, sporulation; $P < 0.0045$, FDR-adjusted) were significantly enriched in uORFs (Fig. 5B). The presence of uORFs in DNA binding proteins suggests that these important genes are likely to be highly regulated.

[1]Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520, USA. [2]Program in Computer Science and Computational Biology, Yale University, New Haven, CT 06520, USA. [3]Department of Molecular, Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: michael.snyder@yale.edu

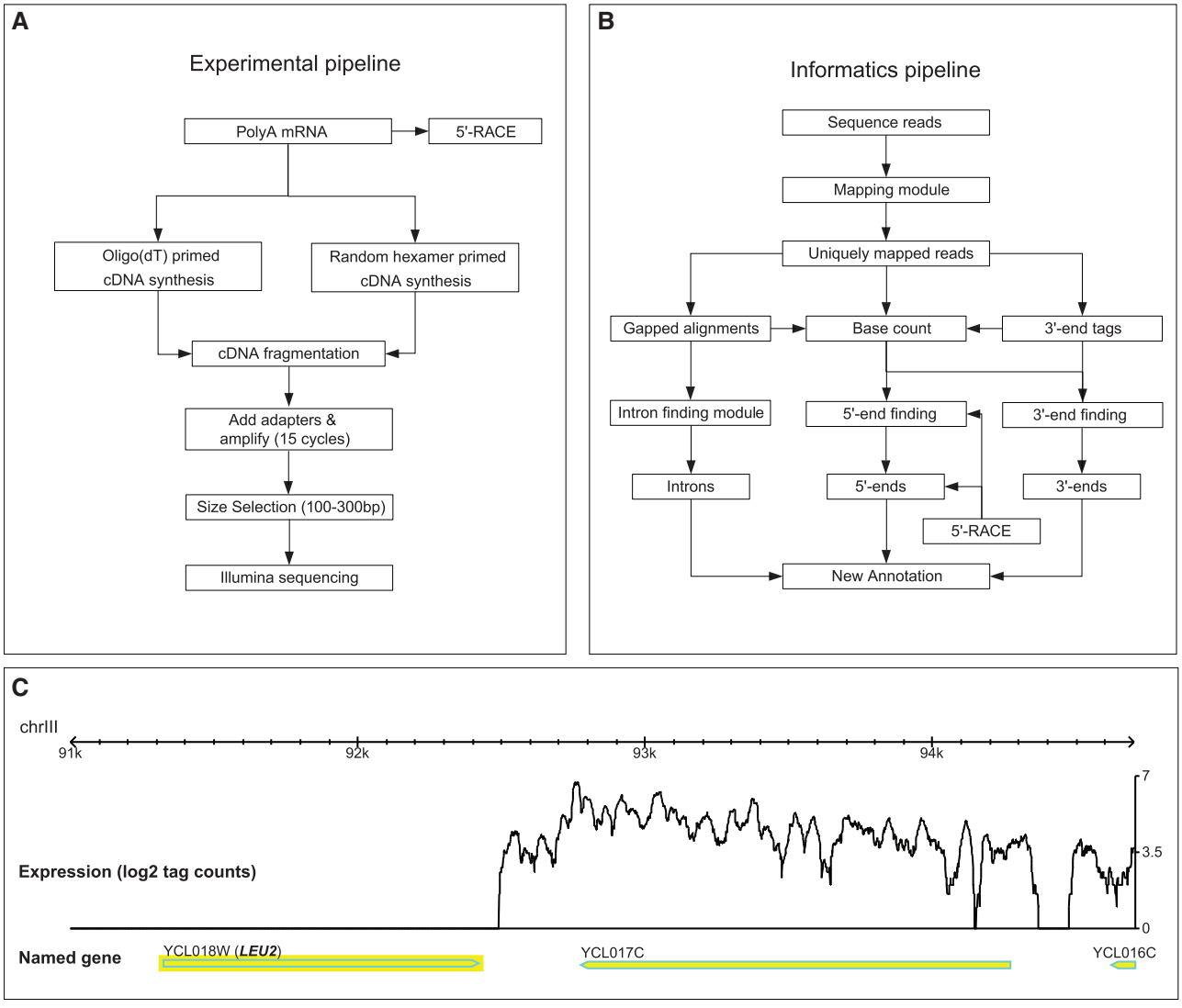

Fig. 1. (**A** and **B**) Flowcharts of the RNA-Seq method. (**C**) RNA-Seq signals are not evident at a deleted gene (*LEU2*) but are abundant at an expressed neighboring gene (*YCL017C*).

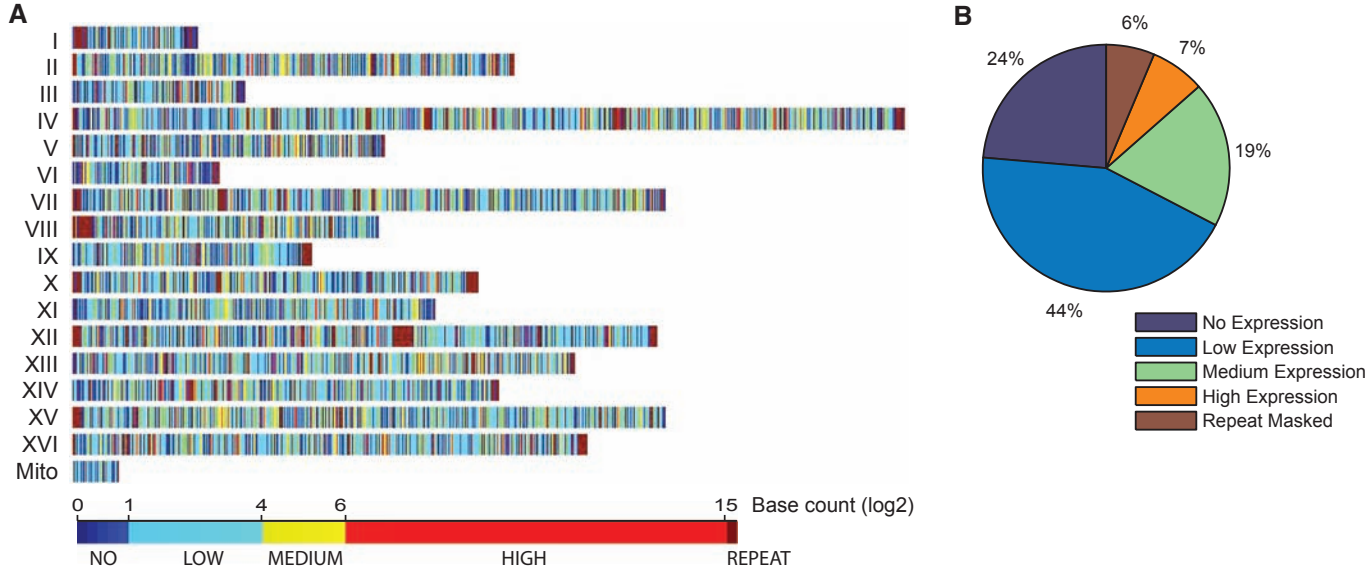

Fig. 2. (**A**) The genome distribution of transcribed regions in yeast. Colors represent different transcription levels for each base (log2 tag count). Numbers of chromosome regions are shown on the left. Mito, mitochondria. (**B**) A summary of the transcription level of the transcriptome.

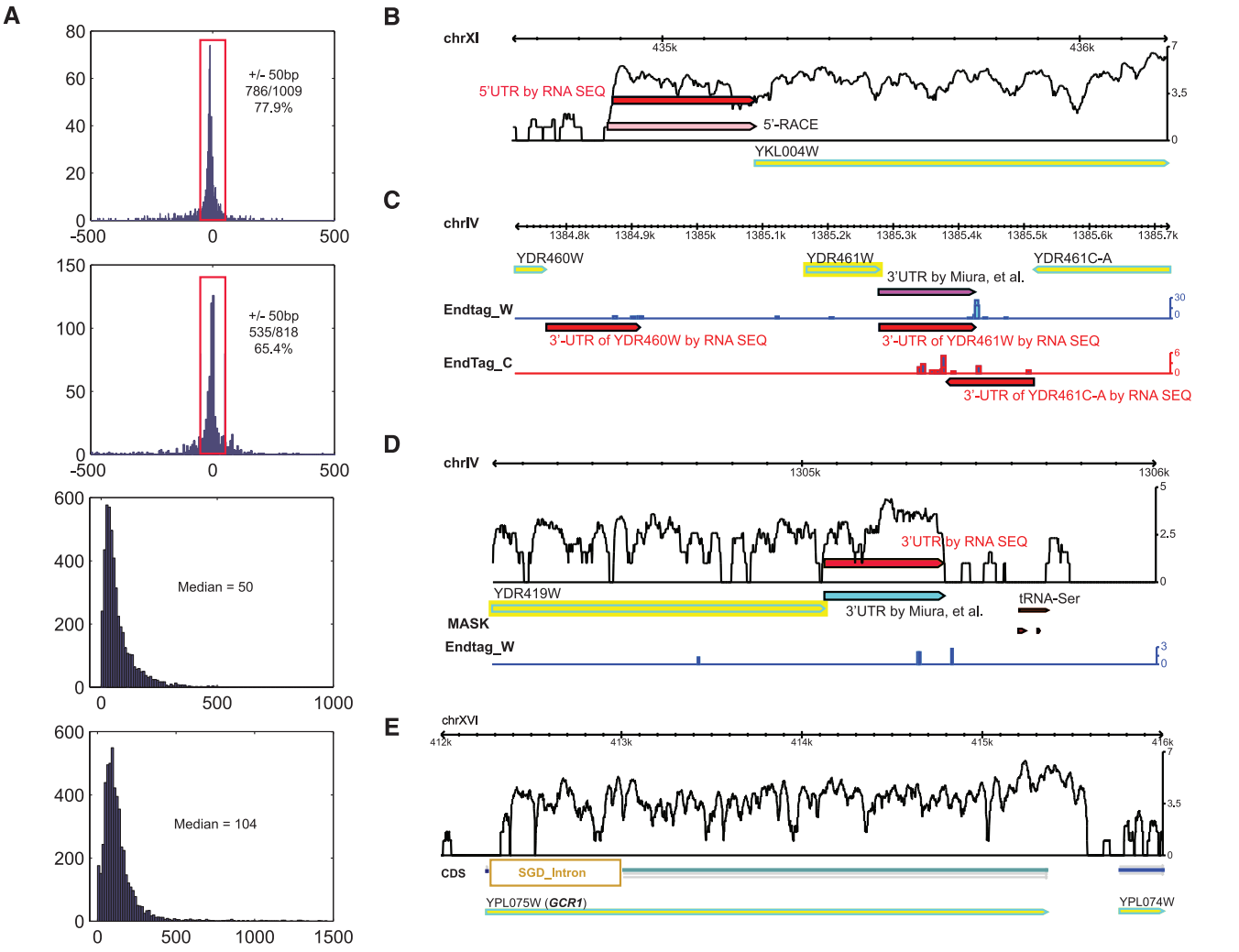

Fig. 3. (**A**) Size differences between mapped 5′-UTR data with RNA-Seq and RACE (top) and 3′-UTR data with RNA-Seq and cDNA sequencing (*9*) (second from top) next to the size distributions of the 5′ UTR (second from bottom) and 3′ UTR (bottom). (**B**) The 5′ UTR as determined by RNA-Seq and by 5′ RACE for gene *YKL004W*. In all figures, a colored box represents an ORF and an arrow indicates the transcription direction. chrXI, chromosone XI. (**C**) 3′ UTR determined by RNA-Seq on the basis of end tags for genes *YDR460W*, *YDR461W*, and *YDR461C-A*, or for *YDR461W* by cDNA sequencing (*9*). Endtag_W (blue) and Endtag_C (red) represent RNA-Seq reads containing poly(A) tails on either Watson or Crick strands, respectively, and the bars represent confidence scores of each 3′ end (*6*). (**D**) 3′ UTR determined by RNA-Seq based on a sharp expression decrease as compared with cDNA data (*9*). MASK, repeated mask sequence. (**E**) An SGD-annotated intron (box in brown color) in *GCR1* not supported by RNA-Seq. CDS, coding sequence.

We also searched for transcription in intergenic regions (Fig. 5D) by identifying stretches of 150 bp or greater with significant expression above that of surrounding regions (Fig. 5F) (*6*). Of 487 expressed regions identified, 204 had not been observed by microarray analyses or cDNA studies (*7*, *9*). For 18 regions found by RNA-Seq but not other methods, we tested expression with real-time quantitative PCR (QPCR) and confirmed transcription in 16 cases (table S5).

Because RNA-Seq is a quantitative method, it can potentially be used to quantify RNA levels. We determined the median signal in a 30-bp window located immediately upstream of the 3′ ends of the annotated stop codon (table S4) after removing genes with overlapping 3′ ends in this region. The expression levels of 34 genes predicted to be expressed at a range of high, medium, and low levels were measured with QPCR. We found a strong correlation ($R = 0.98$) between the QPCR and RNA-Seq data (fig. S6A). As expected, discrepancies were the largest among the genes expressed at a low level and were better than those obtained with standard microarrays ($R = 0.72$) (fig. S6C) (*16*) or tiling DNA-microarray analysis ($R = 0.48$) (fig. S6B) (*7*). Moreover, the dynamic range of RNA-Seq is at least 8000-fold,

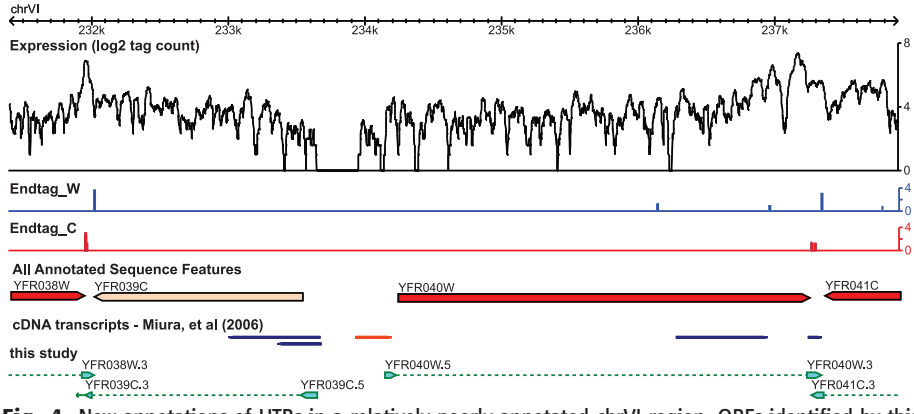

Fig. 4. New annotations of UTRs in a relatively poorly annotated chrVI region. ORFs identified by this study are denoted by dotted lines, and arrows denote the transcription direction. Green-shaded boxes are UTRs. cDNA transcripts in red are of high confidence and those in blue are of low confidence [defined in (*9*)].

compared with ~60-fold for DNA microarrays (see the scale of fig. S6, B and C).

RNA-Seq offers several advantages as compared with other technologies. First, it allows interrogation of all unique sequences of the genome, including those that are closely related, as long as unique bases exist that can be monitored. Second, because a large number of reads can readily be obtained, the method is very sensitive and offers a large dynamic range. Thus, RNA-Seq can detect and quantify levels of RNAs expressed at very low levels as compared with DNA microarrays; the sensitivity of the latter is probably reduced because of cross-hybridization effects. Third, RNA-Seq allows accurate determination of exon boundaries. Although we precisely mapped 3′ ends of many yeast genes using RNA-Seq, we did not attempt to determine the nucleotide reso-

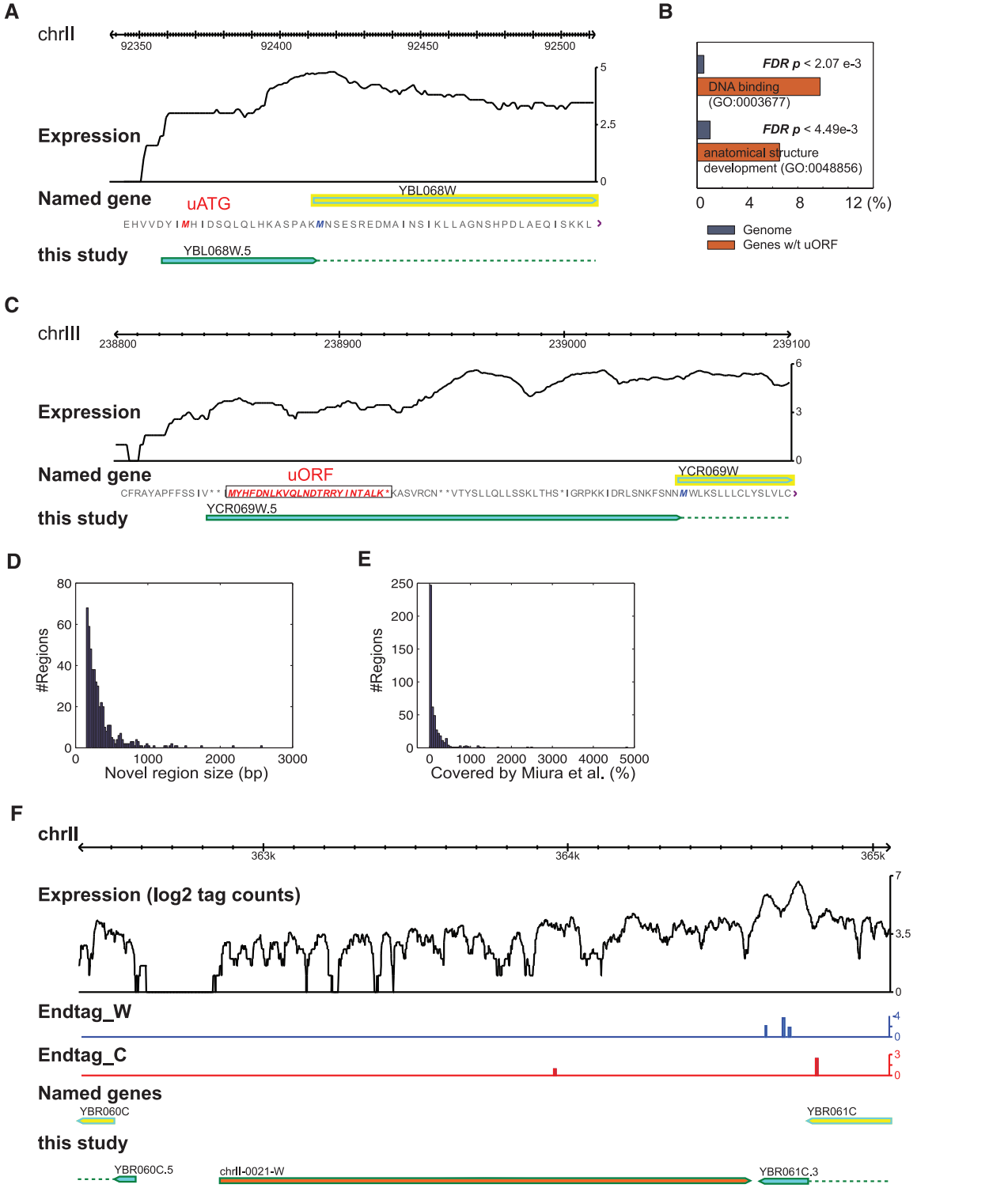

Fig. 5. (**A**) RNA-Seq reveals putative upstream start codons (uATG, in red) relative to existing annotations (blue ATG). (**B**) Significantly more upstream uORFs were found relative to annotated ORFs for certain biological functions. *P* values are FDR–adjusted. (**C**) An example of an uORF (boxed and in red). (**D**) Size distribution of previously undescribed transcribed regions. (**E**) Transcribed regions previously covered by cDNA sequencing (*9*) in percentages. (**F**) An example of a previously undescribed transcribed region containing a poly(A) signal (shaded in red).

lution of 5′ boundaries because yeast 5′ ends are often heterogeneous (*9*, *17*) and we performed an amplification step. Instead, an approximate location was deduced by a sharp transition in signal over a small interval. Nonetheless, overall, RNA-Seq provides a useful map of exon boundaries. Our RNA-Seq method allowed us to map the transcriptional landscape of the yeast genome and define UTRs and previously unknown transcribed regions. In the future, application of this method should help to precisely determine the transcriptional landscape of other genomes.

References and Notes

1. M. Snyder, M. Gerstein, *Science* **300**, 258 (2003).
2. M. B. Gerstein *et al.*, *Genome Res.* **17**, 669 (2007).
3. M. D. Adams *et al.*, *Nature* **377**, 3 (1995).
4. P. Kapranov *et al.*, *Science* **296**, 916 (2002).
5. P. Bertone *et al.*, *Science* **306**, 2242 (2004).
6. Materials and methods are available as supporting material on *Science* Online.
7. F. Perocchi, Z. Xu, S. Clauder-Munster, L. M. Steinmetz, *Nucleic Acids Res.* **35**, e128 (2007).
8. L. David *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5320 (2006).
9. F. Miura *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17846 (2006).
10. M. Kellis, N. Patterson, M. Endrizzi, B. Birren, E. S. Lander, *Nature* **423**, 241 (2003).
11. P. Cliften *et al.*, *Science* **301**, 71 (2003).
12. K. Juneau, C. Palm, M. Miranda, R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1522 (2007).
13. F. Mignone, C. Gissi, S. Liuni, G. Pesole, *Genome Biol.* **3**, REVIEWS0004 (2002).
14. A. G. Hinnebusch, *Annu. Rev. Microbiol.* **59**, 407 (2005).
15. M. J. Ruiz-Echevarria, S. W. Peltz, *Cell* **101**, 741 (2000).
16. F. C. Holstege *et al.*, *Cell* **95**, 717 (1998).
17. C. F. Albright, R. W. Robbins, *J. Biol. Chem.* **265**, 7042 (1990).

18. We thank S. P. Dinesh-Kumar for comments. Funded by two grants from NIH and one from the Connecticut Stem Cell Fund (P6SCE01). The Gene Expression Omnibus accession number for sequences is GSE11209.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1158441/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S6
References

31 March 2008; accepted 22 April 2008
Published online 1 May 2008;
10.1126/science.1158441
Include this information when citing this paper.

The Transcription/Migration Interface in Heart Precursors of *Ciona intestinalis*

Lionel Christiaen,[1]* Brad Davidson,[1]† Takeshi Kawashima,[2] Weston Powell,[1] Hector Nolla,[3] Karen Vranizan,[4] Michael Levine[1]*

Gene regulatory networks direct the progressive determination of cell fate during embryogenesis, but how they control cell behavior during morphogenesis remains largely elusive. Cell sorting, microarrays, and targeted molecular manipulations were used to analyze cardiac cell migration in the ascidian *Ciona intestinalis*. The heart network regulates genes involved in most cellular activities required for migration, including adhesion, cell polarity, and membrane protrusions. We demonstrated that fibroblast growth factor signaling and the forkhead transcription factor FoxF directly upregulate the small guanosine triphosphatase RhoDF, which synergizes with Cdc42 to contribute to the protrusive activity of migrating cells. Moreover, RhoDF induces membrane protrusions independently of other cellular activities required for migration. We propose that transcription regulation of specific effector genes determines the coordinated deployment of discrete cellular modules underlying migration.

There has been considerable progress in elucidating the gene regulatory networks controlling cell fate specification during animal development (*1*–*3*). In parallel, traditional in vitro assays coupled with more-recent proteome analyses have characterized the protein interaction networks controlling dynamic cellular processes, such as actin-based membrane protrusions (*4*) and adhesion (*5*). Comparatively little is known about how transient regulatory states interface with the dynamic cellular processes underlying morphogenesis. We investigated this problem using the migrating heart precursors of the ascidian *Ciona intestinalis* as a relatively simple model. The ascidian heart develops from the B7.5 pair of blastomeres that specifically express the basic helix-loop-helix transcription factor Mesp in response to the T-box factor Tbx6 (*6*, *7*). Subsequently, a fibroblast growth factor (FGF) signal activates the Ets1/2 transcription factor, a presumed Mesp target, and induces heart specification and cell migration of the anteriormost B7.5 granddaughter cells (Fig. 1A) (*8*). As a consequence, the heart precursors, called trunk ventral cells (TVCs), migrate into the trunk, whereas their sibling cells form anterior tail muscles (ATMs) (Fig. 1B). FGF signaling upregulates the forkhead box transcription factor FoxF in the TVCs. Interfering with FoxF function inhibits cell migration, but not heart muscle differentiation, showing that TVC migration is predominantly controlled by FGF signaling and the FoxF transcription factor (*9*).

To determine how FGF and FoxF control TVC migration, we developed a method for lineage-specific transcription profiling using fluorescence-activated cell sorting (FACS) and microarray analysis. The cis-regulatory DNAs from *Mesp* and the myogenic differentiation *MyoD* were used to express green and yellow fluorescent proteins in the B7.5-lineage and surrounding mesodermal cells, respectively [Fig. 1B; Mesp, green fluorescent protein (GFP) transgenes; MyoD, yellow fluorescent protein (YFP) transgenes]. B7.5-lineage cells were sorted based on their GFP fluorescence, after targeted manipulations of Mesp, FoxF, and FGF signaling using the *Mesp* cis-regulatory DNA (Fig. 1, A to C) (*7*–*9*). Targeted expression of a dominant-negative FGF receptor (dnFGFR) converts the entire B7.5 lineage into ATMs (*8*) and is therefore expected to inhibit the expression of TVC-specific genes (Fig. 1A and fig. S3). In addition, modified versions of Mesp and FoxF (Mesp:VP16 and FoxF:WRPW, respectively; both are fusion proteins) block TVC migration but not cardiomyocyte differentiation (*7*, *9*), thereby providing an opportunity to identify migration-specific genes (Fig. 1, A and E).

Microarray assays captured differential expression of known TVC- and tail muscle–specific marker genes (Fig. 1D and table S1). Moreover, at least 130 genes were found to be downregulated in all three conditions that inhibit TVC migration as compared with wild-type samples (Fig. 1E, fig. S4, and table S2). Transcriptional profiling of late gastrula stage B7.5 cells (LG sample) and whole tailbud embryos (whole sample) indicated that these 130 genes are upregulated in wild-type TVCs at the onset of migration (fig. S4). In situ hybridization assays validated the experimental design: 51 of 56 randomly selected candidate genes were expressed in migrating TVCs (Fig. 1, F and G, and fig. S3). Candidate migration genes include a broad spectrum of functional classes [for example, the RhoDF small Ras-homolog guanosine triphosphatase (Rho GTPase) and wunen-like phospholipid phosphohydrolase] (Fig. 1, F and G, and table S2). This diversity supports the view that many facets of cell migration are controlled transcriptionally (*10*, *11*).

Specific gene families and biological processes have been implicated in directed cell migration, such as polarity, cell-matrix adhesion, and actin dynamics regulators (figs. S7 to S9). We compared the expression levels of individual genes in wild-type and dnFGFR samples, which permitted the identification of genes specifically upregulated in either TVCs or ATMs (Fig. 1A and figs. S5 and S9). Cell type–specific genes were found to func-

[1]Department of Molecular and Cell Biology, Division of Genetics, Genomics and Development, Center for Integrative Genomics, University of California, Berkeley, CA 94720, USA. [2]Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA. [3]Cancer Research Laboratory, University of California, Berkeley, CA 94720, USA. [4]Functional Genomics Laboratory, University of California, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: lionelchristiaen@berkeley.edu (L.C.); mlevine@berkeley.edu (M.L.)

†Present address: Department of Molecular and Cellular Biology, Molecular Cardiovascular Research Program, University of Arizona, Tucson, AZ 85724, USA.