Of Mice, Men, and the Genome

Judith Margolin^{1,2}

¹Baylor College of Medicine, Houston, Texas 77030, USA; ²Texas Children's Cancer Center, Department of Hematology-Oncology, Houston, Texas 77030, USA

Master, I marvel how the fishes live in the sea. (W. Shakespeare, Pericles Act II, Sc. I, V. 28)

In this issue Carninci et al. (2000) and the RIKEN Genome Exploration Research Group introduce one of the “marvels” of nature that many have predicted but that we are just beginning to grasp in full scope. This is the first large-scale publication of a mammalian “transcriptosome,” or the genome as it is expressed, in which the sequenced clones are specifically enriched for full-length sequences. This massive effort represents the accumulated data from the RIKEN-MEI 4.02 release May 15, 2000 (<http://genome.rtc.riken.go.jp/>), including 914,452 3' end murine sequences representing 126,693 nonredundant clones found in >80 cDNA libraries generated from histologically and developmentally diverse tissues.

Perhaps even more important than the public identification of so many new mouse cDNAs is the clear explanation of the methods used to construct the Cap-Trapped, full-length, normalized, and subtracted libraries. These methods were responsible for a sequence redundancy level <2% and a rate of novel sequences discovery that ran as high as 20%–39% for the most heavily subtracted sublibraries. It is these two factors that made it economically feasible to consider sequencing such a large part of the transcriptosome. As pointed out by the authors, these methods are likely to be quite useful to the other ongoing or proposed cDNA discovery and sequencing efforts, including the recently announced Mammalian Gene Collection (MGC, <http://www.ncbi.nlm.nih.gov/>

MGC/index.html). Of the transcripts obtained by these methods, 88% appear to contain at least one probable ATG start site and evidence of full-coding length, and many demonstrate large amounts of both 5' and 3' untranslated sequence. These data will allow other researchers to move quickly from the small fragments of genes derived from more traditional methods or from homology criteria to obtaining a full-length actual cDNA clone.

A number of features distinguish this effort from both prior methods and other large-scale cDNA sequencing projects, which also have used normalization and subtraction techniques. (Bonaldo et al. 1996; Marra 1999) The first difference is the Cap-Trapper method itself (outlined in Fig. 1 of Carninci et al. 2000, in this issue), on which they have made and published serial improvements since 1996 (Carninci et al. 1996, 1998; Carninci and Hayashizaki 1999). This is not a method for the timid or for those inexperienced with the most technically demanding RNA work. Although the authors mention that they have been able to perform the method with RNA generated from microdissected tissues (usually no more than microgram quantities at best), generally speaking, it has required milligram quantities of high-quality RNA as starting material. The subtracted and normalized material used in the final cloning steps is frequently present in only nanogram or even subnanogram level (J. Margolin, unpubl.).

A second difference from prior efforts is the complete avoidance of the two methods that are the most likely causes of short and incomplete tran-

script insertion: There is no use of PCR and the cDNA is never amplified or otherwise manipulated in the plasmid state. Despite the improvements in PCR enzyme or enzyme mixture fidelity and processivity, the kinetics of the PCR reaction will inevitably favor short and easier to replicate templates. The PCR kinetic factor can be ameliorated but never eliminated with size fractionation schemas or by limiting cycle numbers. Inevitably, small amounts of contaminating shorter sequences or even sequences of the selected length that just happen to be easier to replicate or that were incompletely normalized will dominate the final product in geometric proportion to the amount of amplification (cycles) employed. The present methods avoid these pitfalls. Also, the Cap-Trapped, normalized, subtracted cDNA is cloned into a *cre-lox* containing a λ replacement vector where the ultimate circularization process does not favor shorter inserts. The importance of the *cre-lox* vector can not be understated, and it is hoped that Carninci et al. (2000, in this issue) will expeditiously publish both the vector and informatic details related to this project.

A third and final difference is how these authors set up the normalization, using the original mRNA as the driver, and their use of RoT calculations for the subtraction against pools of RNA expressed from previously sequenced clones. Other authors cited above have used similar normalization techniques using in vitro-expressed RNA or single-stranded DNA. Some of the most efficient prior subtraction techniques have had to rely on fragmentation of the cDNA (often employing a cutting re-

E-MAIL jmargolin@txccc.org **FAX (713) 825-4038**

Article and publication are at www.genome.org/cgi/doi/10.1101/gr.162800.

striction enzyme, i.e., the Diatchenko method) to obtain high efficiency with the subtraction (Diatchenko 1996). Carninci et al. (2000) have managed to avoid this and still obtain very efficient and specific subtractions. It is really a sum of the parts or, rather, a “sum-of-individual-great-parts,” including excellent informatics and quality-control measures (see Fig. 3 and Table 2 of the Carninci et al. 2000, in this issue, or the above-mentioned RIKEN Web site), that have led to the success of this effort.

So, in the context of rapidly changing genomic techniques, where other forms of sequencing approaches (i.e., SAGE) and microarrays allow researchers glimpses of the transcriptosome for far less effort and cost, what is the importance of the techniques and clone set described in the Carninci (2000, in this issue) article? I believe the answer to this question is that in order to extract the most information from the other methods, we need an accurate picture of what is possible. Furthermore, the techniques described in this article are likely to become applicable beyond the context of large gene discovery programs. What are now the methods of large, expensive gene discovery efforts may become just another research tool as methods for accurately amplifying diverse mRNA pools improve and the price of sequencing and computing continues to fall. It is not hard to imagine researchers performing individual experiments in which the changes in transcriptosome function are traced with quick subtractions that are performed with small aliquots of available subtraction pools at

specific RoT levels. Conceivably, differences in individual cases of specific diseases or response to therapies could be traced this way.

A frequent criticism of large-scale cDNA sequencing efforts is that they are expensive, non-hypothesis-driven fishing expeditions. These issues have been hotly debated on the pages of both this journal and many others (Goodman 1999). I would posit that it is not only the individual researchers who have an immediate need for a particular clone who have a vital interest in these arguments. In conjunction with the sequencing of the human genome, this information is critical to understanding the actual biology.

Thus, the choice of the fish quotation I used at the beginning of this commentary was, of course, deliberate. I would not be true to the Bard or to the readers of this essay if I did not give you the full quotation and the context in which it was stated:

Fisherman 3: . . . Master, I marvel how the fishes live in the sea.

Fisherman 1: Why, as men do a-land, the great ones eat up the little ones. I can compare our rich misers to nothing so fitly as to a whale: a' plays and tumbles, driving the poor fry before him, and at last devour them all at a mouthful. Such whales have I heard on a' th' land, who never leave gaping till they swallow'd the whole parish, church, steeple, bells, and all.

(*Pericles*, Act II Sc. I)

Shakespeare was making a political as well as an entertaining comment.

Within the context of our own times, biotechnology fish from minnows to whales are scrambling to identify and patent sequence. It behooves the public-domain sequencing groups to move swiftly with these improved methods to make sure that the information and the clones are freely available to all.

REFERENCES

- Bonaldo, M., Lennon, G., and Soares, M. 1996. *Genome Res.* **6**: 792–806.
- Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K., Sasaki N., Izawa, M., et al. 1996. *Genomics* **37**: 327–336.
- Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. 1998. *Proc. Natl. Acad. Sci.* **95**: 520–524.
- Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. 2000. *Genome Res.* **10**.
- Carninci, P. and Hayashizaki, Y. 1999. *Methods Enzymol.* **303**: 19–44.
- Diatchenko, L., Lau, Y., Campbell, A., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E., et al. 1996. *Proc. Natl. Acad. Sci.* **93**: 6025–6030.
- Goodman, L. 1999. *Genome Res.* **9**: 673–674.
- Marra, M., Hillier, L., Kucaba, T., Allen, M., Barstead, R., Beck, C., Blistain, A., Bonaldo, M., Bowers, Y., Bowles, L., et al. 1999. *Nature Genet.* **21**: 191–194.
- Shakespeare, W. *Pericles, Prince of Tyre*. In *The Riverside Shakespeare*, pp 1490–174. Houghton Mifflin Co., Boston.
- Soares, M., Bonaldo, M., Jelene, P., Su, L., Lawton, L., Efstratiadis, A. 1994. Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci.* **91**: 9228–9232.

Of Mice, Men, and the Genome

Judith Margolin

Genome Res. 2000 10: 1431-1432

Access the most recent version at doi:[10.1101/gr.162800](https://doi.org/10.1101/gr.162800)

References

This article cites 8 articles, 3 of which can be accessed free at:
<http://genome.cshlp.org/content/10/10/1431.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Focused on your science.

USC
SCIENTIFIC
SIDE BY SCIENCE

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>