The Mappers' Torch Song

Jun Yu and Gane K.-S. Wong

The Human Genome Center, Department of Medicine, University of Washington, Seattle, Washington 98195 USA

Recently, after almost a decade of effort, two groups—one led by David Schlessinger at Washington University (Nagaraja et al. 1997), and another led by Eric Green at the National Human Genome Research Institute (Bouffard et al., this issue)—announced the completion of yeast artificial chromosome (YAC)-based sequence-tagged site (STS)-content physical maps for human chromosomes X and 7, respectively. Although we understand why most of the world is seeking alternatives to the many years of laborious work involved in building such maps, we feel that precisely because maps of such high quality are so rare, their completion should be celebrated. As these announcements signal the symbolic “passing of the torch” from mapping to sequencing, it is timely to reflect on what has been accomplished and on what lies ahead for the Human Genome Project (Watson 1990; Collins and Galas 1993).

A number of factors distinguish these two maps from others that have been published previously (Foote et al. 1992;

Ashworth et al. 1995; Bell et al. 1995; Chumakov et al. 1995; Collins et al. 1995; Crollius et al. 1996; Doggett et al. 1995; Hudson et al. 1995; Krauter et al. 1995; Quackenbush et al. 1995; Qin et al. 1996): First, the significant size of chromosomes X and 7 is estimated at 160 and 170 Mb, respectively; the next largest chromosome-specific map, for chromosome 11, is 130 Mb. Second, the high density of STS markers on these maps averages 80 kb in both cases, which is better than the 100 kb goal set by the Human Genome Project and comparable to the best of the chromosome-specific maps (e.g., chromosomes 11 and 22). Third, these chromosome maps contain a high fraction of uniquely ordered STSs, which is 90% on the chromosome 7 map and often not even specified on the other maps. Fourth, well-characterized and stable YAC clones (a few thousand in number for each map) are used in the contig construction; for example, hybrid cell-derived chromosome 7 YACs exhibited <15% rearrangements, compared with

50% for Centre d'Etude du Polymorphisme Humain (CEPH) mega-YACs along the same chromosome. And, fifth, there is complete integration of the STS maps with the existing radiation hybrid and genetic maps. Other STS mappers have achieved or, in some cases, exceeded some of these goals, but so far none except Nagaraja et al. (1997) and Bouffard et al. (this issue) has accomplished all of these goals.

It is the high quality of these two maps that makes them stand out and bring to issue the basic question of what it means to say that a map (and now, increasingly, a sequence) is “finished.” There is not a product in the world that cannot be made a little cheaper or finished a little sooner if the quality is allowed to slip. However, if we are to use a race track analogy, the race that we are engaged in is a relay, not an individual event; the success of the endeavor will be judged by the accomplishments of the entire team. Consider, then, the overall consequences of having a better map. The most significant parameter is

not the average distance between adjacent STSs but, rather, the average distance between uniquely ordered STSs (Olson and Green 1993; Cox et al. 1994). In the chromosome 7 map, these distances are 80 kb between adjacent STSs and 90 kb between uniquely ordered STSs. In contrast, although the map of Hudson et al. (1995) covers the entire human genome (providing its own unique advantages), the average distance between adjacent STSs is only 170 kb and the average distance between uniquely ordered STSs is estimated to be five times larger—or just under 1 Mb. STS-content maps are currently used to screen bacterial artificial chromosome (BAC) or P1-derived artificial chromosome (PAC) libraries with average insert sizes of ~140 kb (Shizuya et al. 1992; Ioannou et al. 1994; Kim et al. 1996). When the clone insert sizes are smaller than the distance between uniquely ordered STSs, gaps become the rule rather than the exception, and significant amounts of chromosome walking are usually needed to close the gaps. Even if the costs for generating the new STSs and rescreening the libraries can be subsumed into the much larger costs for sequencing, the true costs for mapping end up being hidden as sequencing costs. So what does it mean to “finish” a map if, in the end, it has to be rebuilt by the sequencers?

As one example, take the work being done at the University of Washington Genome Center, where a significant fraction of the sequencing has been on chromosome 7. Having access to the map of Bouffard et al. (this issue) has proven quite fortunate. Use of the multiple complete digest (MCD) restriction fragment mapping technique (Wong et al. 1997) generated close to 2 Mb of contiguous sequence-ready cosmids. To guard against YAC rearrangements, every region was validated by two independently mapped YACs. Given the accuracy of the MCD maps, this meant that the sequenced materials were validated to an average resolution of 200 bp, far better than in the original STS map. A dozen YACs were subcloned into cosmids and MCD mapped to generate this 2-Mb region. No errors were found in the ordering of the YACs on the STS map, and only one of the YACs exhibited a rearrangement—one that was too small to have been detected by the STS map. These YACs were also MCD vali-

dated against more than a dozen BACs (J. Yu and G.K.-S. Wong, unpubl.). The point here is not to extol the virtues of YACs or to advocate any particular method for generating a high-resolution sequence-ready map from an STS-content map. Rather, it is to affirm the very high quality of the chromosome 7 map. Similarly positive experiences with the chromosome 7 map have been reported at the Genome Center at Washington University, using another fingerprint-based high-resolution mapping technique (R. Waterston, pers. comm.). It is also worth noting that a well-handled YAC library can be used as a sequencing substrate. This is particularly important, as having the flexibility to choose among a wide variety of cloning systems is essential for closure. For example, 20% of the nematode genome is not clonable in cosmids (Coulson et al. 1991), and work is currently under way to close these regions by shotgun sequencing directly from YACs (R. Waterston, pers. comm.).

As we gaze longingly at the still distant finish line, thoughts turn to the other chromosomes, where the STSs may not be as well ordered and where a high-quality large-insert clone library has yet to be mapped. The final phase of the Human Genome Project—the actual sequencing of those 3 billion base pairs—hinges on this point. We are not arguing against a continuous improvement in the STS maps as a part of any sequencing effort. BACs and PACs are emerging as the large-insert clones of choice, and these new libraries have to be placed on the maps. However, in this mad dash toward the finish line, we would caution against taking too cavalier an attitude toward data quality, particularly with regard to the issue of clone validation. STS mapping technology cannot detect clonal aberrations that are smaller than the distance between uniquely ordered STSs. Small insertions or deletions of just a few kilobases in size are possible. We would be remiss in our duties if we did not point out the following simple fact: Any problems that we choose to “sweep under the rug” will only come back to haunt the biologists who will be using our data—the people to whom the Human Genome Project chromosome mappers and sequencers will be “passing the torch.” It is because of the demonstrated high quality of the chromosomes X and 7 maps that we, the

immediate users of the maps, are grateful. Let us strive to ensure that the ultimate users of the sequence data being produced by the Human Genome Project will feel likewise in the years to come.

REFERENCES

- Ashworth, L.K., M.A. Batzer, B. Brandriff, E. Branscomb, P. de Jong, E. Garcia, J.A. Garnes, L.A. Gordon, J.E. Lamerdin, G. Lennon et al. 1995. *Nature Genet.* 11: 422–427.
- Bell, C.J., M.L. Budarf, B.W. Nieuwenhuijsen, B.L. Barnoski, K.H. Buetow, K. Campbell, A. Colbert, J. Collins, P.R. Desjardins, D. DeZaawn et al. 1995. *Hum. Mol. Genet.* 4: 59–69.
- Bouffard, G.G., J.R. Idol, V.V. Braden, L.M. Iyer, A.F. Cunningham, L.A. Weintraub, J.W. Touchman, R.M. Mohr-Tidwell, D.C. Pelusson, R.S. Fulton et al. 1997. *Genome Res.* (this issue).
- Chumakov, I.M., P. Rigault, I. LeGall, C. Bellann'e-Chantelot, A. Billault, S. Guillou, G. Guasconi, E. Poullier, I. Gros et al. 1995. *Nature* 377: 175S–179S.
- Collins, F. and D. Galas. 1993. *Science* 262: 43–46.
- Collins, J.E., C.G. Cole, L.J. Smink, C.L. Garrett, M.A. Leversha, C.A. Soderlund, G.L. Maslen, L.A. Everett, K.M. Rice, A.J. Coffey et al. 1995. *Nature* 377: 367S–379S.
- Cox, D.R., E.D. Green, E.S. Lander, D. Cohen, and R.M. Myers. 1994. *Science* 265: 2031–2032.
- Coulson, A., Y. Kozono, B. Lutterbach, R. Shownkeen, J. Sulston, and R. Waterston. 1991. *BioEssays* 13: 413–417.
- Crollius, H.R., M.T. Ross, A. Grigoriev, C.J. Knight, E. Holloway, J. Misfud, K. Li, M. Playford, S.G. Gregory, S.J. Humphray et al. 1996. *Genome Res.* 6: 943–955.
- Doggett, N.A., L.A. Goodwin, J.G. Tesmer, L.J. Meincke, D.C. Bruce, L.M. Clark, M.R. Altherr, A.A. Ford, H.C. Chi, B.L. Marrone et al. 1995. *Nature* 377: 335S–366S.
- Foot, S., D. Vollrath, A. Hilton, and D.C. Page. 1992. *Science* 258: 60–66.
- Hudson, T.J., L.D. Stein, D.S. Gerety, J. Ma, A.B. Castle, J. Silva, D.K. Slonim, R. Baptista, L. Kruglyak, S.H. Xu et al. 1995. *Science* 270: 1945–1954.
- Ioannou, P.A., C.T. Amemiya, J. Garnes, P.M.

Insight/Outlook

Kroisel, H. Shizuya, C. Chen, M.A. Batzer, P.J. de Jong. 1994. *Nature Genet.* 6: 84–89.

Kim, U.J., H. Shizuya, H.L. Kang, S.S. Choi, C.L. Garret, L.J. Smink, B.W. Birren, J.R. Korenberg, I. Dunham, M.I. Simon et al. 1996. *Proc. Natl. Acad. Sci.* 93: 6297–6301.

Krauter, K., K. Montgomery, S.-J. Yoon, J. LeBlanc-Straceski, B. Renault, I. Marondel, V. Herdman, L. Cupelli, A. Banks, J. Lieman et al. 1995. *Nature* 377: 321S–333S.

Nagaraja, R., S. MacMillan, J. Kere, C. Jones, S. Griffin, M. Schmatz, J. Terrell, M. Shomaker, C. Jermak, C. Hott et al. 1997. *Genome Res.* 7: 210–222.

Olson, M.V. and P. Green. 1993. *Cold Spring Harbor Symp. Quant. Biol.* 58: 349–355.

Qin, S., N.J. Nowak, J. Zhang, S.N.J. Sait, P.G. Mayers, M.J. Higgins, Y. Cheng, L. Li, D.J. Munroe, D.S. Gerhard et al. 1996. *Proc. Natl. Acad. Sci.* 93: 8034–8038.

Quackenbush, J., C. Davies, J.M. Bailis, J.V. Khristich, K. Diggle, Y. Marchunk, J. Tobin, S.P. Clark, A. Rodkins, S. Marceno et al. 1995. *Genomics* 29: 512–525.

Shizuya, H., B. Birren, U.J. Kim, V. Mancino, and T. Splepak. 1992. *Proc. Natl. Acad. Sci.* 89: 8794–8797.

Watson, J.D. 1990. *Science* 248: 44–49.

Wong, G.K.-S., J. Yu, E.D. Thayer, and M.V. Olson. 1997. *Proc. Natl. Acad. Sci.* 94: 5525–5230.

The Mappers' Torch Song

Jun Yu and Gane K.-S. Wong

Genome Res. 1997 7: 666-668

Access the most recent version at doi:[10.1101/gr.7.7.666](https://doi.org/10.1101/gr.7.7.666)

References

This article cites 21 articles, 12 of which can be accessed free at:
<http://genome.cshlp.org/content/7/7/666.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](#).

Affordable, Accurate
Sequencing.

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