



Random Shotgun Fire

Laurie Goodman

Genome Res. 1998 8: 567-568

Access the most recent version at doi:[10.1101/gr.8.6.567](https://doi.org/10.1101/gr.8.6.567)

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](#).



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

Cold Spring Harbor Laboratory Press

EDITORIAL

Random Shotgun Fire

Craig Venter's and Perkin-Elmer's May 9th announcement of a new joint venture to complete the sequence of the human genome in just 3 years set off a furor among the scientific community. The uproar, however, was unsurprising given that the earliest newspaper articles presented the plan as if it were a *fait accompli* and accused the publicly funded Human Genome Project of being a "waste" of money. The announcement, made just prior to the Genome Mapping, Sequencing, and Biology Meeting, held May 13–17 at Cold Spring Harbor Laboratory, was discussed, at least briefly, at the sequencing center director's meeting that preceded the CSHL meeting, in a very well-attended session during the CSHL meeting, and in numerous speculative debates over meals and beers among attending scientists.

The Facts

The publicly funded projects are generally sequencing the human genome by a strategy that might be called *map and then shotgun*. Physical maps of human chromosomes are constructed, typically first with YACs (yeast artificial chromosomes) and then with BACs (bacterial) or PACs (P1). A "minimal tiling path" of BACs/PACs that overlaps by the least amount is then selected and sequenced by a shotgun approach. All sequence data, both unfinished assembled sequence and completely finished clones, are released daily as agreed upon at the international sequence strategy meeting, held each of the last 3 years in Bermuda. The complete human sequence is anticipated in 2005, ending a 15-year project, with the last sequencing phase costing roughly \$1 billion.

The basic plan of the Venter/Perkin-Elmer venture for sequencing the human genome is to use a *whole-genome* shotgun sequencing approach. This means sequencing random inserts from completely unmapped clones from libraries containing fragmented DNA from the whole human genome. The new company will perform at least 10^8 sequence reads (a 10-fold redundancy—or what some might call a "deep shotgun"). They plan to do both forward and reverse reads from each PCR-prepared insert and will use libraries from different individuals, allowing them to identify polymorphisms during their analysis. Sequence assembly will be done incrementally using BAC-end sequence, ESTs, and STSs to link and position evolving sequence. The cornerstone of this initiative is a new 96-capillary sequencing instrument currently in development at Applied Biosystems, Inc. (ABI). About 230 instruments are expected to be in full operation at the new company by April 1999. The number of runs that can be accomplished in a day on one of these machines is projected to be about 8–12. The separation matrix is the same as that currently in use in the ABI 310 sequencer, and it should produce 500- to 600-base read

lengths. The sequencing machine will be accompanied by an automated workstation that does colony picking, extraction, PCR, and sequencing reactions. The company is committed to publicly releasing data on a quarterly basis on contigs greater than 2 kb, though the exact details for this release are still under discussion. The general business plan of the company includes the intention to patent between 100 and 300 *interesting* gene systems. Additionally, they plan to position themselves as a supplier of a sequence database and analysis tools. Finally, they intend to exploit the discovered single nucleotide polymorphisms (SNPs), probably through a new genotyping service. With their approach, they claim they can sequence the human genome in just 3 years at a cost of roughly \$200 million.

Opinion, Scientific or Otherwise

From an unbiased view, the two plans seem to be simply two different methods to obtain genomic information. So why all the excitement? The answer seems to be the apparent attack on the publicly funded projects that accompanied the announcement. Take, for example, the surprisingly unbalanced news article in *The New York Times* (May 11, 1998) that presented the Venter/Perkin-Elmer proposal as if the end point had already been achieved and that publicly funded project coordinators had conceded defeat. The article stated that "Recognizing the credibility of the new venture by Dr. Venter and Perkin-Elmer, N.I.H. officials [unnamed] are preparing to persuade Congress to continue funding the genome project but to switch the focus from getting the sequence to the enormous tasking of interpreting it." The article goes on to say that "Having forfeited the grand prize of the human genome sequence, they [Congress] should now be equally happy with the glory of paying for similar research in mice." Given that in press conferences, in later newspaper articles, and at the CSHL Genome Meeting, officials at the NIH, specifically Harold Varms and Francis Collins, made it clear that the news of the death of the U.S. publicly funded Human Genome Project was greatly exaggerated, *The New York Times* article appears to be sloppy news coverage at best and horribly biased at worst.

The international commitment to publicly support the Human Genome Project is equally strong. The Wellcome Trust in the UK immediately announced they were doubling their funding for human genome sequencing, adding another £110 million, and making it possible for the UK Human Genome Project to produce one-third of the human sequence. Discussion of a Wellcome Trust funding increase had been long under way and approved; however, the Trust felt it important in light of the Venter/Perkin-Elmer announcement that swift notification be made to make it clear that

GOODMAN

support for the Human Genome Project by public funding agencies remained extremely strong.

The Concerns

The main concern is that public perception (or misperception) will result in loss of public funding for the Human Genome Project, and this particular concern is why public airing of scientific information absolutely requires increased care to make clear to the public a project's true status, its overall aims, and its drawbacks. The final outcome of the Venter/Perkin-Elmer project is far from certain, and statements indicating that publicly funded projects are now a waste are irresponsible.

At this stage, no good scientist would say, especially given that there is no hard evidence whatsoever (not even a representative *model* has been tested to provide insight about problems or biases in a whole-genome shotgun for a repeat-rich genome), how successful or how high the quality—as an *assembled* sequence—the proposed Venter/Perkin-Elmer venture will be. The human sequence is unlike any other genome that has been done in a whole-genome shotgun fashion. The sheer number and length of repeat regions may make the assembly task next to impossible. Phil Green commented at the Genome Meeting that in a whole-genome shotgun, bridging repeat regions longer than 5–10 kb (e.g., LINEs and clusters of Alus) would often fail. Green estimated that following assembly of whole-genome shotgun data, the average contiguous sequence length would probably be only 40–50 kb. This should provide partial gene and SNP identification data, but little more. Many contigs would be only a few kb or less in size and would lack identifiable markers (BAC ends or ESTs), with the result that they would be entirely unpositioned orphans. Genes extending over several 100 kb or more (of which there are a great many in the A+T rich parts of the genome) would almost certainly not be obtained in intact form. Alu-rich regions—common in the G+C rich, gene-dense regions of the genome—are difficult to assemble correctly even at the single cosmid or BAC level, much less on a genome-wide scale. Thus genes in such regions would very likely be garbled or have significant pieces missing. Moreover, knowing the entire sequence from one end of the chromosome to the other is absolutely necessary for recognizing structural relationships for comparative evolutionary studies, and studies on chromosome dynamics will benefit more from a complete sequence than smaller, less accurately ordered fragments. As Aravinda Chakravarti commented at the Genome Meeting, the product that Craig Venter and Perkin-Elmer are aiming for “is not what we [the genome community] argued for; this is not what we bargained for,” articulating what the Human Genome Project is and what it is not—a project intended to provide high-quality, detailed information to aid all researchers in every area of biology, not just gene hunters.

Finally, leaving human genome sequencing entirely to private companies raises questions about pub-

lic availability of this information—an issue hotly debated in the newspapers, and rightly so, given its potential impact on the public's general welfare. Venter and Perkin-Elmer make it clear that they only plan to patent a small number of gene systems, but the effect of patenting any DNA sequence is uncertain, nor is it known what would ultimately hold them to patenting only a small number. Certainly the Wellcome Trust feels there is reason for concern; Michael Morgan announced at the Genome Meeting that the Trust is very carefully investigating gene patenting legalities and is preparing, if they believe it necessary, to contest any gene patents on the public's behalf.

Conclusions

At this stage, concerns about the Venter/Perkin-Elmer project—whether the project can be completed as stated, what the final product's form will be, and what impact privately owned human genome information would have—remain unresolved; thus, reduction in public funding would be extraordinarily premature. Finally, it should be noted that although there is clear concern over the effect on public funding and on the final form of the sequence to come from the Venter/Perkin-Elmer venture, it was equally clear at the CSHL Genome Meeting that numerous scientists are excited about the potential new information source and feel it will be a useful partner to the public projects. Likewise, they embraced the challenge of working toward providing the entire genome sequence, in complete order and orientation, in a timely manner and hopefully working in concert with Venter and Perkin-Elmer as well as with others. Privately and publicly funded projects have been working together so far, with both profiting from it. They should continue to do so.

One last point, to put things in perspective, especially given that many newspaper reporters wrote their stories as if Venter or the NIH or The Wellcome Trust—each on its own—was trying to win the *prize* of deciphering the human genome. This moment in the history of the Human Genome Project is unlikely to be remembered at all, except as one of many minor discussions on the best approach for attaining the final product for the human genome sequence. If one is looking for a model, the way the completed yeast genomic sequence is viewed undoubtedly reflects how the human genome sequence will ultimately be seen—as a tremendously successful international *group* effort and certainly not the work of any one individual or even country. It is ridiculous to act as if such might be the case and a waste of time behaving in such a fashion.

Laurie Goodman