

# nature genetics

---

volume 25 no. 3 july 2000

---

## A focus on function

The completion of the human draft sequence inspires the question: what next? Large-scale, full-on sequence analysis is not going to fade away at any point in the near future, and there is still a great deal to be done on the human genome. But the other genomes to be sequenced, by whom and with what priority are a matter of debate. Another question relevant to resource management is how to balance the acquisition of additional sequence with exploiting the data already to hand. The recent appointment of Allan Bradley (presently of Baylor College of Medicine) as the next director of the Sanger Centre (Hinxton, UK) is an interesting one that, in several respects, speaks to these issues.

To begin with, it indicates that the Wellcome Trust—which helped to found and, for the most part, funds the Sanger Centre—is placing increasing emphasis on the elucidation of gene function. Having completed the genome of *Caenorhabditis elegans* in a joint effort with Washington University and contributed approximately one-third of the human draft, the Sanger Centre has focused intensely on the acquisition of high-throughput sequence and, along with the neighbouring European Bioinformatics Institute, its analysis. It has had a critical role in establishing a standardized set of procedures (known as the Bermuda agreement) that has facilitated coordinated international effort. These include release of sequence data into public databases within 24 hours of their acquisition, which has made it difficult to turn them into intellectual property. The subsequent development of Ensembl, a program that ‘functionally’ annotates sequence as it is generated (and before it is deposited into GenBank) provides a further disincentive to those who would seek to make similar annotations with a view to using them as an argument for patenting.

Much of this effort reflects a passion of John Sulston, the present director, who is primarily interested in “getting the data out there so that people can pick it up and use it,” rather than controlling its release which, he feels, would result in the Human Genome Project “clumping around specific projects”. Certainly, a crowning achievement of the project is that countries in three continents have found a way to expeditiously and successfully collaborate on a project of mammoth proportion. Whereas the extent to which immediate release has contributed to unification is difficult to assess, there seems little doubt that it has helped. Sulston retires from his position with a justifiable sense that the most urgent goals have been accomplished. He has no wish to prescribe strategy for making sense of sequence through functional genomics; with the bulk of human sequence there for the taking, he feels it is for others to orchestrate its dissection.

As a principal investigator of a grant to sequence the mouse genome, Bradley is committed to the rapid acquisition of finished mouse sequence and believes it critical to efforts to elucidate gene function—a conviction that may be fuelled by his part in developing methods and resources to genetically manipulate the mouse. These include a collection of embryonic stem cells for generating specified mouse knockouts, created under the auspices of Lexicon Genetics, a company he co-founded. Noting that these knockouts are “not very accessible,” he plans at least partial replication of this resource. Bradley’s mandate to recruit people to explore gene function in different systems is one that is consistent with this and other goals, which include more accurate phenotyping, the genesis of mutants (be they mouse or bacterial) and a robust growth of robotics (in the form of *in situ* hybridization as well as microarray) to query gene expression. He predicts that the future, so to speak, is physiology; that, when it comes to genetic mechanism, physiology will soon receive the same attention that development has enjoyed for some time—and that the integration of molecular ‘intelligence’, such as detailed expression data combined with annotation derived from *in silico* comparisons, will allow more accurate prediction of genes that contribute to physiological processes.

Increasing focus on gene function brings one closer to applied biology, which brings one closer to issues of intellectual property. Unlike the simple sequence of a nucleic acid, the utility of gene function and methods that exploit gene function is apparent. And as Bradley points out, attaching intellectual property to a method, say, for treating breast cancer may increase the odds that that method will be extrapolated to therapy, and is therefore in the public interest. Whereas Sulston’s point that academic biological research thrives in an open environment and one should therefore take care not to put up barriers is well taken, he too, acknowledges that applied biology is a “new area, and that [with respect to intellectual property] what is right for one moment is not necessarily right for another”. In broaching function, the Sanger Centre (and the Wellcome Trust) may find it necessary to review policy and philosophy—an activity with which it is well practiced, most recently through hosting a meeting to discuss application of the ‘Bermuda agreement’ to the release of coordinates of protein structures.

As for sequencing, Bradley is not alone in his desire to have mouse genome sequence of high quality (90% coverage) as soon as possible. The suggestion of parallel sequencing projects has raised the hackles of geneticists who think the power of the mouse to elucidate gene function, together with the perception that Celera Genomics will restrict access to mouse sequence (of which they have apparently sequenced 1.15 billion bases) to commercial bodies, makes rapid, thorough coverage of the mouse genome paramount. Whether Bradley’s concern will effect an increased number of ‘mouse’ reads at the Sanger Centre remains to be seen; there is speculation that a good number of its sequencers will—after the human sequence is shotgunned up to the point of hand-finishing—be churning out zebrafish sequence. Francis Collins (director of the National Human Genome Research Institute), who presented a case for the ‘parallel’ approach at the recent Cold Spring Harbor meeting on Genome Biology & Sequencing, recently announced that the rat genome is next on the list for US sequencers. The arguments for a more ‘parallel’ and less ‘serial’ approach is the power of comparative genomics to indicate biologically relevant motifs (an argument well familiar to those who seek to build a case for the pathologic effect of missense mutations) and, of course, the needs of the various model-organism communities.

With sequence capacity having increased to the extent that it has—the G5 centres alone sequence approximately 6 million reads per month—the present debate is a luxury that was hard to conceive of three years ago. The fact that resources are still perceived to be limiting (and not only with respect to sequence generation) bears testimony to the feeling that, as Karen Carpenter so presciently sang, we’ve only just begun.

