

Academia–Industry Collaboration: An Integral Element for Building “Omic” Resources

David E. Hill,^{1,10} Michael A. Brasch,² Anthony A. del Campo,³ Lynn Doucette-Stamm,⁴ James I. Garrels,⁵ Judith Glaven,⁶ James L. Hartley,⁷ James R. Hudson Jr.,⁸ Troy Moore,⁹ and Marc Vidal^{1,10}

¹Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA; ²Atto Bioscience, Rockville, Maryland 20850, USA; ³Office of Research and Technology Ventures, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA; ⁴Agencourt Biosciences Corporation, Beverly, Massachusetts 01915, USA; ⁵Garbrook Associates, Beverly, Massachusetts 01915, USA; ⁶Harvard Medical School, Boston, Massachusetts 02115, USA; ⁷SAIC/NCI, Frederick, Maryland 21702, USA; ⁸CityScapes, Huntsville, Alabama 35801, USA; ⁹Open Biosystems, Huntsville, Alabama 35806, USA

The availability of ~200 nearly completed genome sequences and >900 additional sequencing projects underway is changing the very fabric of biological research endeavors. With access to enormous amounts of sequencing data and rapidly expanding cloned gene collections, scientists have the opportunity to pursue research projects at any scale, from highly focused, one-gene-at-a-time studies to broader, more global genome and proteome-wide approaches. Although the former efforts are well within the standard purview of traditional research laboratories, global approaches necessitate a more complex collaborative environment involving multidisciplinary teams from academia, government, and industry. Such “large-scale science,” most recently demonstrated by the Human Genome Project, also demands open access to data and resources, regardless of where the primary data are generated, and a commitment to provide as complete a resource as is feasible. This special issue focuses on creating, improving, and using cloned “ORFeomes” and exemplifies successful partnerships between academia and industry. In this perspective we argue that long-term academia–industry collaborative relationships provide optimal solutions to the specific problems of discovery science.

From Blueprints to Finished Goods

The human genome sequence and that of various model organisms provide a necessary framework for a transition from molecular bi-

ology to systems biology. Although the human genome sequence is sometimes referred to as the “parts-list,” it is crucial to realize that genome sequence annotations, as they are available today, provide rough drafts of blueprints for the parts. The challenge to establish the precise number of parts, namely, the encoded proteins and RNAs, their actual structure, and their respective interactions, requires a dedicated effort to convert the blueprints into an accessible warehouse of available, well-characterized manufactured parts.

This issue of *Genome Research* highlights recent developments in the generation of various genome-wide resource collections that are expected to contribute to a more integrated understanding of biological processes (Ideker et al. 2001; Vidal 2001). As such, the efforts described herein (Dricot et al. 2004; Dupuy et al. 2004; Lamesch et al. 2004; Rual et al. 2004a,c), along with other established collections of cDNAs and ORFs (Hudson Jr. et al. 1997; Strausberg et al. 2002; Carninci et al. 2003; Reboul et al. 2003) constitute a foundation on which it will be possible to investigate and manipulate both specific genes and proteins and the global networks in which they participate. The creation of multiple types of genome resources, from large-insert genomic DNA libraries to specialized collections of individually cloned genes, cDNAs, and ORFs and their utilization across multiple disciplines as a way to understand biology from a systems approach is a direct consequence of the highly collaborative, interdisciplinary efforts such as those required for the Human Genome Project. However, to take full advantage of the growing collection of genome sequences and associated databases requires focused and committed efforts to create comprehensive resource collections that are not encumbered in any way (Collins et al. 2003a).

The public availability of ~200 genome sequences for humans, model organisms, and microbial species (Bernal et al. 2001) has provided tremendous impetus for creation of large-scale sets of cloned genes, expanding by orders of magnitude the numbers of genes and proteins readily accessible for further study (for review, see Rual et al. 2004b). Efficient utilization of these gene resources will require the deployment of robust and facile technologies for isolating full-length open reading frame (ORF) clones in order to carry out proteome-wide, protein-based studies and corresponding promoter regions for transcriptional regulation and localization studies. Analogous to the collaborative efforts at sequencing the human genome, “discovery science” will depend on collaborative arrangements involving both public and private partnerships (Committee on Large-Scale Science and Cancer Research 2003). A significant aspect of these collaborative enterprises is that the results and corresponding resources must be available to the public in an open-access setting, free of intellectual property (IP) entanglements.

Academia–Industry Collaborations: A Relationship Fostered by Governmental Action

Relationships among United States colleges and universities and commercial firms have existed since at least the 1860s, when the Morrill Act established the United States land-grant system of colleges, which fostered the transfer of new agricultural methods and technologies to farm operations (for review, see Hasselmo and McKinnell 2003). The Morrill Act also provided a mechanism for the federal government to fund investigator-initiated research projects (Committee on Large-Scale Science and Cancer Research 2003), a prelude to the subsequent develop-

¹⁰Corresponding authors.
E-MAIL marc_vidal@dfci.harvard.edu; FAX (617) 632-5739.
E-MAIL david_hill@dfci.harvard.edu; FAX (617) 632-5739.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.2771404>.

ment of extramural funding by the National Institutes of Health (NIH).

Throughout the 20th century, scientists have relied on commercial firms to provide critical reagents, materials, and technical know-how for their investigator-initiated efforts. For example, Fisher Scientific, founded in Pittsburgh in 1902 by Chester Garfield Fisher, was one of the first commercial sources of equipment and reagents for United States laboratories, initially as a reseller of quality instruments imported from Europe (<http://www.fisherscientific.com>). Various products from Fisher were used in government laboratories during the Manhattan Project to build the atomic bomb, one of the first of many "big science" projects undertaken by the United States government.

Although academic research has relied on industry for consumables and technology, much of the intellectual foundation and initial proofs-of-principle supporting a significant fraction of commercially available products generally derive from academic research endeavors. Obviously, both groups have developed seminal technologies (see below), and industry has provided the necessary means by which individual discoveries become value-added reagents, quickly and efficiently disseminated to the entire research community. The wealth of antibody-based commercial products available online from over 250 suppliers (for listing of companies with online antibody resources, see <http://www.antibodyresource.com/>) is directly attributable to the research efforts of Kohler and Milstein (1975) and is just one example of the commercialization process. It can be argued that this plethora of antibody products is available because there were no patents filed on the initial hybridoma technology by the inventors.

Commercialization Versus Public Access

Commercialization of the knowledge arising from academic and governmental research in the United States was inefficient at best prior to 1980. In that year, the United States Congress passed the Bayh-Dole and Stevenson-Wydler Acts, which changed the landscape of academia-industry relations. By these two acts, congress set forth a policy to expedite commercialization of products resulting from the federal government's investment in basic research (Blumenthal 2003; Committee on Large-Scale Science and Cancer Research 2003; Hasselmo and McKinnell 2003). A major consequence of the Bayh-Dole Act was a shift in Federal policy that allowed the IP rights resulting from the fruits of federally funded research to remain with the academic centers where the innovations were made. This decentralization of IP management,

combined with incentivized academia, became a motivating factor in the transfer of technology to the commercial sector (Committee on Large-Scale Science and Cancer Research 2003).

Prior to embarking on a full-scale effort to sequence the human genome, NIH was actively involved in filing patents on expressed sequence tags (ESTs); this activity elicited concern among many scientists (Olson 2002; Sulston and Ferry 2002), with public debate actually having an impact on the Human Genome Project (Sulston and Ferry 2002). Fortunately, there was a growing sentiment for public release of data that had evolved from the collaborative efforts to sequence model organisms, notably the *Caenorhabditis elegans* sequencing effort, coupled with a commitment by Merck to fund efforts leading to the public release of over 400,000 ESTs (Sulston and Ferry 2002). Concern over ownership of and restricted public access to the human genome resulted in the Human Genome Project providing daily release of DNA sequences to public databases once NIH abandoned its efforts to pursue patent protection on human genes (Olson 2002; Sulston and Ferry 2002). This very successful mechanism of public release of data has subsequently been adopted by the rat and mouse sequencing consortia (Waterston et al. 2002; Gibbs et al. 2004) and by the SNP Consortium to map human single nucleotide polymorphisms (SNPs; Holden 2002). The SNP Consortium is particularly noteworthy in that 13 companies provided funding to generate a SNP map in which all of the data were to be in the public domain with no IP entanglements. Clearly, the corporations funding the SNP project and Merck funding of EST sequencing concluded that the overall goals of the respective projects outweighed the potential monetary value accessible through patent protection of a small fraction of the total data set.

Today, as a direct consequence of the Human Genome Project and the development of super high-throughput sequencing technologies, DNA sequencing has become a commodity in which academia and industrial laboratories can "outsource" their sequencing (Salisbury 2004). The trend to commodity-based reagents and services for sequencing has enhanced collaboration between academic and industry because the critical issue is assigning functions to a gene rather than simply sequencing a gene. Our own efforts at ORFeome cloning and interactome mapping have benefited by access to such IP-neutral relationships with corporations (Walhout et al. 2000a; Reboul et al. 2001, 2003; Lamesch et al. 2004; Li et al. 2004), albeit not without initial concerns over "ownership" issues that were eventually resolved by adhering to open access principles and practices (see below).

Large-Scale Resource Collections: Build-It-By-Collaboration

Public-private partnerships are considered critical elements for the future of basic and clinical research in the recently announced NIH "Roadmap" (<http://nihroadmap.nih.gov/>) and for successful outcomes involving "large-scale science" projects according to a recent report from the National Academy of Sciences (Committee on Large-Scale Science and Cancer Research 2003). However, building large-scale resource collections requires substantial and secured funding. Inevitably, these efforts entail extensive collaborations between public and private institutions, an issue that NIH has recently embraced through various initiatives such as the ENCODE Project (<http://www.genome.gov/10005107>). In academia, the traditional tenure-track, independent investigator-based environment is not ideally suited to large-scale, resource-building projects because the emphasis of such projects is perceived to be not based on hypothesis-driven inquiry (Committee on Large-Scale Science and Cancer Research 2003). Nevertheless, academia is insulated from the uncertainties of the business climate, which allows projects to be completed even when corporate partners are unable or unwilling to continue their involvement in the project due to a change in the partner's strategic direction. Industry, on the other hand, is well suited to carry out large-scale "production" style work and can keep overall costs low due to economies of scale. By incorporating industrial quality assurance and assessment practices and methods of production, suitably funded academic laboratories can embark on large-scale resource building projects and achieve a high degree of success. The major academic/institute sequencing centers that carried out a substantial portion of the full-scale sequencing of the human genome had adopted such production-style manufacturing processes (Hawkins et al. 1997; Huang 1999; Stojanovic et al. 2002; Collins et al. 2003b).

Completely Complete?

A key aspect of the Human Sequencing Consortium of public and private agencies is that the project has forged ahead with completing the sequencing effort and making the data freely available as they are generated. Although completion of the sequencing effort is essential for the building of comprehensive cDNA and ORFeome resources, it is arguably the hardest aspect of any sequencing project. Annotation and reannotation of partial and completed genomes have become the rate-limiting steps for building comprehensive resources of cloned ORFs and cDNAs as exemplified by the ongoing reannotation of genome sequences (Bernal et al. 2001; Gene Ontology Consortium 2001; Flybase Consor-

tium 2002, 2003; Garrels 2002; Crowe et al. 2003; Daraselia et al. 2003; Kellis et al. 2003, 2004; Meyers et al. 2003; Stein et al. 2003; Gibbs et al. 2004; Imanishi et al. 2004; Lamesch et al. 2004). As we use genome sequences to move further into the “discovery science” phase of resource building, efforts comparable to genome reannotation, that is, iterative versions of cDNA, ORFeome, and promoterome collections, will be required to achieve an acceptable level of completeness (Lamesch et al. 2004). Furthermore, comprehensive databases that compile multiple features for each gene and protein, are regularly updated, and are readily accessible to the entire scientific community will be essential for moving “discovery science” and systems biology forward (Costanzo et al. 2000; Bernal et al. 2001; Gene Ontology Consortium 2001; Matthews et al. 2001; Garrels 2002; Flybase Consortium 2003; Prince et al. 2004). Our own efforts in generating the *C. elegans* ORFeome and promoterome have been guided by the premise that “single-pass” high-throughput cloning has a success rate of ~65%, mainly due to limitations in gene annotation. To clone the remaining 35% requires integration of data from multiple venues, including genome reannotation and comparative genomic analyses. These subsequent efforts to build upon the initial successes are analogous to improved versions of computer software. As with multidisciplinary efforts to reannotate genomes, academia–industry collaborative endeavors to achieve completeness in ORFeome and promoterome resources will be essential.

The *C. elegans* ORFeome Project: A Microcosm of Genome-Wide Resource Building

The *C. elegans* ORFeome project could not have been undertaken without some form of collaboration, especially with respect to the actual cloning of ORFs and their subsequent structural analyses. In that regard, we were fortunate to have Research Genetics (ResGen), Life Technologies, Inc. (LTI), and Genome Therapeutics Corporation (GTC) as collaborators during the entire project. Each of the three collaborating institutions provided critical expertise to the overall project, and key individuals from each institution are co-authors on the various publications that resulted from this project.

We relied on ResGen for pairs of oligonucleotides, each primer nearly 50 nucleotides in length to accommodate the dual needs of being ORF-specific and containing the necessary elements for recombinational cloning (Hartley et al. 2000; Walhout et al. 2000b; Marsischky and LaBaer 2004), synthesized and delivered in 96-well format at a time when few groups were even contemplating high-throughput synthesis of thousands

of primers for PCR, and for help in overall organization of clone collections. Robotic-based liquid handling systems are still a rare entity in most academic laboratories, whereas companies such as ResGen had already developed processes for reliably handling large resource collections. Understanding those processes and learning from ResGen were important for our overall organization of the project.

Standard methods using restriction endonucleases for cloning ORFs are adequate at small scale but inefficient when attempting to clone entire ORFeomes (see Brasch et al. 2004; Marsischky and LaBaer 2004). Fortunately, we had been working with researchers at LTI on the commercialization of yeast two hybrid assay systems (Vidal et al. 1996). This established relationship subsequently evolved into a collaboration furthering the development of the Gateway recombinational cloning system. The collaborative efforts resulted in adapting Gateway for PCR-based cloning of *C. elegans* ORFs (Hartley et al. 2000; Walhout et al. 2000a) as the system was evolving from a basic corporate research and development project to a manufacturable product. Essentially, we were able to both access Gateway as a “beta-tester” (MacNeil 2004), albeit for a very large project involving 19,000 reactions, and expand the capabilities of the product. Although the specific components of Gateway can be produced successfully, nominally in small quantities, in any well-equipped molecular biology laboratory, a quality-controlled manufacturable process capable of producing large quantities of a critical reagent is a hallmark and strength of successful biotech companies. It made more sense to use the expertise at LTI than to try and cut corners by producing the critical reagents in-house.

The use of the Gateway technology did pose a potential problem, namely, the issue of clone ownership and the risk that there would be legal entanglements that would reduce or prevent open access to any gene cloned that way. For model organisms such as *C. elegans*, clone ownership was not an issue because it was very unlikely that a *C. elegans* gene might be the basis of a commercial product or human therapeutic agent, whereas for human genes, there was concern over such ownership issues. Invitrogen, which had acquired the rights to Gateway via acquisition of LTI in 2001, eventually responded to these concerns by publicly announcing a policy of open access to any gene cloned by using Gateway technology, most notably human genes obtained from the Mammalian Gene Collection (www.invitrogen.com/gateway/; http://home.businesswire.com/portal/site/google/index.jsp?ndmViewId=news_view&newsId=20040504006140&newsLang=en), thereby clearing the way for the development of Human ORFeome resource collections (Rual et

al. 2004c), analogous to what was done for *C. elegans* (Reboul et al. 2003). Such open access to physical resources complements the basic principles enunciated by Collins et al. (2003a) for access to other resource collections and databases.

Any cloning project is heavily dependent on DNA sequencing. Our ability to provide experimental verification of *C. elegans* predicted genes and to correct predicted exon–intron structures was based on obtaining high-quality sequences in a high-throughput manner. GTC, one of only two corporate entities that participated in sequencing efforts of the Human Genome Project (Lander et al. 2001), was able to provide sequencing services to the ORFeome project (and the interactome mapping project as well) at costs well below rates charged by core facilities in academia. In this regard, GTC was transitioning from conducting high-throughput genome sequencing to providing custom sequencing services on a smaller scale to multiple customers, both academic and industrial. Our need for sequencing individual ORF clones nicely coincided with their need to adapt to a changing market. Three separate and distinct collaborations all focused on the single goal of obtaining a comprehensive version of the *C. elegans* ORFeome.

Collaborations Come and Go, the Science Stays

All academia–industry collaborative ventures are at risk of being prematurely dissolved due to the vagaries of funding mechanisms and the business climate. This risk is heightened as the scale of the project increases, particularly when one partner is providing a unique component. In the worst-case scenario, the loss of a single partner can derail the entire enterprise. To guard against such an outcome in large-scale science projects requires one partner to serve as the “epicenter.” Because the business climate is generally less stable than academia, academic laboratories are the logical choice to be the focal point for such projects.

Ironically, all three of our corporate collaborators had “disappeared” prior to the publication of the *C. elegans* ORFeome (Reboul et al. 2003). The GenomeVision Sequencing Services of GTC was sold in 2003 to Agencourt Biosciences Corporation, with whom we now collaborate for our sequencing needs. During the creation of the *C. elegans* ORFeome, Invitrogen acquired our other industrial partners, ResGen and LTI. Because each partner was making significant contributions to the project, we were naturally concerned that the original arrangements might be abrogated in some fashion or that the project could be compromised or delayed. Fortunately, those fears were unfounded and we maintained our productive collaboration

with all of the groups, most likely because the project had internal champions supported by strong interpersonal relationships in which “good faith” promises were every bit as important as any legally binding contracts. Had we been unable to access the technologies provided by our partners, the project would have been terminated.

As the ORFeome project continued toward its version 1.1 completion and the interactome mapping project moved forward, our relationship with Invitrogen, now the owner of Gateway cloning technology developed by LTI, evolved from a customer-based one to a more collaborative one, which helped secure our access to Gateway. We also maintained working relationships with those individual collaborators who moved into other ventures after the Invitrogen acquisitions.

Large-Scale Science Requires “Disruptive Technologies” That Arise Anywhere and Impact All

In the past 30 years, there have been five major technologies that have exploded across the entire breadth of biological research, disrupting how molecular biology was formerly done. These are the generation of stable hybridomas leading to monoclonal antibody production in 1975 (Kohler and Milstein 1975), the ability to directly sequence DNA in 1977 (Sanger et al. 1977), the development of PCR in 1985 (Saiki et al. 1985, 1986), the use of immobilized arrays of nucleic acids for analyzing gene expression profiles across large numbers of genes (Schena et al. 1995, 1996), and the capability to specifically knockdown expression of any target protein through the use of small RNAs that catalyze the degradation of specific mRNAs (Fire et al. 1998; Montgomery et al. 1998; Timmons and Fire 1998; Timmons et al. 2001). Although hybridomas and DNA sequencing came out of publicly funded laboratories, the initial version of PCR was developed at Cetus and licensed to Roche. Today, there are a myriad of ways in which PCR, and related technologies, is used to selectively amplify target nucleic acids. Although many of these methods have been developed in academic and industrial laboratories, virtually all aspects of PCR are performed by using commercially available reagents and equipment with the exception of the source of template, and even then there are commercial sources of many of the standard cDNA libraries used for PCR. Microarray technology and RNAi were both developed in academic laboratories, and published methods allow one to carry out either technique by using standard reagents and minimal investment in equipment (Alizadeh et al. 1999; Cheung et al. 1999; Paddison et al. 2004). However, industry has responded in rapid fashion (Lipshutz et al. 1999) such that the commercially available systems for

conducting microarray studies have taken over the field while suppliers of commodity reagents have blanketed the research landscape with RNAi-based products (for a partial listing of companies offering products for RNAi, see <http://www.biocompare.com/nature/jump/1065/siRNA-Technology.html>).

The above examples demonstrate that academic laboratories have had a complex relationship with industry throughout the 30-year history of the biotech revolution. However, this complexity can be distilled to four distinct modes: (1) discoveries made in academic laboratories lead to the creation of new companies, new products, and new technologies through licensing efforts. These technology transfer activities are a direct consequence of the 1980 Bayh-Dole Act. (2) New technologies, products, and equipment developed in industry become key reagents/platforms/assays for academic projects. Such products can be accessed via collaboration, “beta testing,” or direct commercial purchase. (3) Industry provides contract services to academics. Both the service provider and academic customer may collaborate to improve the service product and avoid encumbrances. (4) Large-scale projects necessitate that academic laboratories and industry collaborate as full partners in which IP issues, project management, and staffing be well established before the project begins.

Collaborations Beget Collaborations

The various “omic” efforts described in this special issue further demonstrate the utility of collaborative efforts between academic laboratories and industry. The *C. elegans* ORFeome, interactome, and promoterome projects (Reboul et al. 2003; Dupuy et al. 2004; Lamesch et al. 2004; Li et al. 2004) were accomplished because of active involvement by our collaborative partners. Particularly in the case of the *C. elegans* ORFeome project, our corporate collaborations exemplified all four aspects above in that technology development, “beta testing,” contract sequencing, and project integration all played a role in the overall process. In addition, a successful collaboration actually generates more work for all parties. We continue to collaborate with nearly all of the individuals from industry who participated in the initial development of the *C. elegans* ORFeome (Reboul et al. 2001, 2003) despite the fact that many of them have changed companies. This continued interaction demonstrates that the formation of personal relationships among the partners is ultimately the critical factor to maintaining collaborations.

ACKNOWLEDGMENTS

We thank our colleagues and friends throughout academia and industry who have provided support, critical evaluations, tech-

nologies, ideas and/or contributed to the various “omic” projects. We thank D. Allinger and J. Albala for critical reading of the manuscript and acknowledge the efforts of G. Lucier and U. Caney in fostering open access of clone resources. This work was supported by grants from the National Cancer Institute and the National Human Genome Research Institute awarded to M.V.

REFERENCES

- Alizadeh, A., Eisen, M., Davis, R.E., Ma, C., Sabet, H., Tran, T., Powell, J.I., Yang, L., Marti, G.E., Moore, D.T., et al. 1999. The lymphochip: A specialized cDNA microarray for the genomic-scale analysis of gene expression in normal and malignant lymphocytes. *Cold Spring Harb. Symp. Quant. Biol.* **64**: 71–78.
- Bernal, A., Ear, U., and Kyripides, N. 2001. Genomes OnLine Database (GOLD): A monitor of genome projects world-wide. *Nucleic Acids Res.* **29**: 126–127.
- Blumenthal, D. 2003. Academic-industrial relationships in the life sciences. *N. Engl. J. Med.* **349**: 2452–2459.
- Brasch, M.A., Hartley, J.L., and Vidal, M. 2004. ORFeome cloning and systems biology: Standardized mass production of the parts from the parts-list. *Genome Res.* (this issue).
- Carninci, P., Waki, K., Shiraki, T., Konno, H., Shibata, K., Itoh, M., Aizawa, K., Arakawa, T., Ishii, Y., Sasaki, D., et al. 2003. Targeting a complex transcriptome: The construction of the mouse full-length cDNA encyclopedia. *Genome Res.* **13**: 1273–1289.
- Cheung, V.G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R., and Childs, G. 1999. Making and reading microarrays. *Nat. Genet.* **21**: 15–19.
- Collins, F.S., Green, E.D., Guttmacher, A.E., and Guyer, M.S. 2003a. A vision for the future of genomics research. *Nature* **422**: 835–847.
- Collins, F.S., Morgan, M., and Patrinos, A. 2003b. The Human Genome Project: Lessons from large-scale biology. *Science* **300**: 286–290.
- Committee on Large-Scale Science and Cancer Research. 2003. *Large-scale biomedical science: Exploring strategies for future research.* The National Academies Press, Washington, DC.
- Costanzo, M.C., Hogan, J.D., Cusick, M.E., Davis, B.P., Fancher, A.M., Hodges, P.E., Kondu, P., Lengieza, C., Lew-Smith, J.E., Lingner, C., et al. 2000. The yeast proteome database (YPD) and *Caenorhabditis elegans* proteome database (WormPD): Comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Res.* **28**: 73–76.
- Crowe, M.L., Serizet, C., Thareau, V., Aubourg, S., Rouze, P., Hilsen, P., Beynon, J., Weisbeek, P., van Hummelen, P., Reymond, P., et al. 2003. CATMA: A complete *Arabidopsis* GST database. *Nucleic Acids Res.* **31**: 156–158.
- Daraselina, N., Dernovoy, D., Tian, Y., Borodovsky, M., Tatusov, R., and Tatusova, T. 2003. Reannotation of *Shewanella oneidensis* genome. *Omic* **7**: 171–175.
- Dricot, A., Rual, J.-F., Lamesch, P., Bertin, N., Dupuy, D., Hao, T., Lambert, C., Hallez, R., Delroisse, J.-M., Vandenhaute, J., et al. 2004. Generation of the *Brucella melitensis* ORFeome version 1.1. *Genome Res.* (this issue).
- Dupuy, D., Li, Q., Deplancke, B., Boxem, M., Hao, T., Lamesch, P., Sequerra, R., Bosak, S., Doucette-Stamm, L., Hope, I.A., et al. 2004. A first version of the *Caenorhabditis elegans* promoterome. *Genome Res.* (this issue).
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. 1998. Potent and specific genetic interference by

- double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**: 806–811.
- Flybase Consortium. 2002. The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res.* **30**: 106–108.
- . 2003. The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res.* **31**: 172–175.
- Garrels, J.I. 2002. Yeast genomic databases and the challenge of the post-genomic era. *Funct. Integr. Genomics* **2**: 212–237.
- Gene Ontology Consortium. 2001. Creating the gene ontology resource: Design and implementation. *Genome Res.* **11**: 1425–1433.
- Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**: 493–521.
- Hartley, J.L., Temple, G.F., and Brasch, M.A. 2000. DNA cloning using in vitro site-specific recombination. *Genome Res.* **10**: 1788–1795.
- Hasselmo, N. and McKinnell, H. 2003. *Working together, creating knowledge: The university-industry research collaborative initiative*, p. 95. Business-Higher Education Forum, Washington, DC.
- Hawkins, T.L., McKernan, K.J., Jacotot, L.B., MacKenzie, J.B., Richardson, P.M., and Lander, E.S. 1997. A magnetic attraction to high-throughput genomics. *Science* **276**: 1887–1889.
- Holden, A.L. 2002. The SNP consortium: Summary of a private consortium effort to develop an applied map of the human genome. *Biotechniques Suppl.* **22–24**, 26.
- Huang, G.M. 1999. High-throughput DNA sequencing: A genomic data manufacturing process. *DNA Seq.* **10**: 149–153.
- Hudson Jr., J.R., Dawson, E.P., Rushing, K.L., Jackson, C.H., Lockshon, D., Conover, D., Lanciault, C., Harris, J.R., Simmons, S.J., Rothstein, R., et al. 1997. The complete set of predicted genes from *Saccharomyces cerevisiae* in a readily usable form. *Genome Res.* **7**: 1169–1173.
- Ideker, T., Galitski, T., and Hood, L. 2001. A new approach to decoding life: Systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**: 343–372.
- Imanishi, T., Itoh, T., Suzuki, Y., O'Donovan, C., Fukuchi, S., Koyanagi, K.O., Barrero, R.A., Tamura, T., Yamaguchi-Kabata, Y., Tanino, M., et al. 2004. Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biol.* **2**: E162.
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**: 241–254.
- Kellis, M., Birren, B.W., and Lander, E.S. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617–624.
- Kohler, G. and Milstein, C. 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**: 495–497.
- Lamesch, P., Milstein, S., Hao, T., Rosenberg, J., Li, N., Sequerra, R., Bosak, S., Doucette-Stamm, L., Vandenhaute, J., Hill, D.E., et al. 2004. *C. elegans* ORFeome version 3.1: Increasing the coverage of ORFeome resources with improved gene predictions. *Genome Res.* (this issue).
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., et al. 2004. A map of the interactome network of the metazoan *C. elegans*. *Science* **303**: 540–543.
- Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* **21**: 20–24.
- MacNeil, J.S. 2004. Guide to a beta life. In *Genome technology*, pp. 24–27. D. Waters, pub., New York.
- Marsischky, G. and LaBaer, J. 2004. Many paths to many clones: A comparative look at high-throughput cloning methods. *Genome Res.* (this issue).
- Matthews, L.R., Vaglio, P., Reboul, J., Ge, H., Davis, B.P., Garrels, J., Vincent, S., and Vidal, M. 2001. Identification of potential interaction networks using sequence-based searches for conserved protein–protein interactions or “interologs.” *Genome Res.* **11**: 2120–2126.
- Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. 2003. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**: 809–834.
- Montgomery, M.K., Xu, S., and Fire, A. 1998. RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **95**: 15502–15507.
- Olson, M.V. 2002. The Human Genome Project: A player's perspective. *J. Mol. Biol.* **319**: 931–942.
- Paddison, P.J., Silva, J.M., Conklin, D.S., Schlabach, M., Li, M., Aruleba, S., Balija, V., O'Shaughnessy, A., Gnoj, L., Scobie, K., et al. 2004. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**: 427–431.
- Prince, J.T., Carlson, M.W., Wang, R., Lu, P., and Marcotte, E.M. 2004. The need for a public proteomics repository. *Nat. Biotechnol.* **22**: 471–472.
- Reboul, J., Vaglio, P., Tzellas, N., Thierry-Mieg, N., Moore, T., Jackson, C., Shin-i, T., Kohara, Y., Thierry-Mieg, D., Thierry-Mieg, J., et al. 2001. Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in *C. elegans*. *Nat. Genet.* **27**: 332–336.
- Reboul, J., Vaglio, P., Rual, J.F., Lamesch, P., Martinez, M., Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R., et al. 2003. *C. elegans* ORFeome version 1.1: Experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat. Genet.* **34**: 35–41.
- Rual, J.-F., Ceron, J., Koreth, J., Hao, T., Nicot, A.-S., Hirozane-Kishikawa, T., Vandenhaute, J., Orkin, S.H., Hill, D.E., van den Heuvel, S., et al. 2004a. Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome Res.* (this issue).
- Rual, J.F., Hill, D.E., and Vidal, M. 2004b. ORFeome projects: Gateway between genomics and omics. *Curr. Opin. Chem. Biol.* **8**: 20–25.
- Rual, J.F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P.-O., et al. 2004c. Human ORFeome version 1.1: A platform for reverse proteomics. *Genome Res.* (this issue).
- Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. 1985. Enzymatic amplification of β -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350–1354.
- Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B., and Erlich, H.A. 1986. Analysis of enzymatically amplified β -globin and HLA-DQ α DNA with allele-specific oligonucleotide probes. *Nature* **324**: 163–166.
- Salisbury, M.W. 2004. Say goodbye to sequencing. In *Genome Technology*, pp. 28–36. D. Waters, pub., New York.
- Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**: 5463–5467.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467–470.
- Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W. 1996. Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci.* **93**: 10614–10619.
- Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al. 2003. The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. *PLoS Biol.* **1**: E45.
- Stojanovic, N., Chang, J.L., Lehoczy, J., Zody, M.C., and Dewar, K. 2002. Identification of mixups among DNA sequencing plates. *Bioinformatics* **18**: 1418–1426.
- Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., et al. 2002. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl. Acad. Sci.* **99**: 16899–16903.
- Sulston, J. and Ferry, G. 2002. *The common thread: A story of science, politics, ethics, and the human genome*. The Joseph Henry Press, Washington, DC.
- Timmons, L. and Fire, A. 1998. Specific interference by ingested dsRNA. *Nature* **395**: 854.
- Timmons, L., Court, D.L., and Fire, A. 2001. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**: 103–112.
- Vidal, M. 2001. A biological atlas of functional maps. *Cell* **104**: 333–339.
- Vidal, M., Braun, P., Chen, E., Boeke, J.D., and Harlow, E. 1996. Genetic characterization of a mammalian protein–protein interaction domain by using a yeast reverse two-hybrid system. *Proc. Natl. Acad. Sci.* **93**: 10321–10326.
- Walhout, A.J., Sordella, R., Lu, X., Hartley, J.L., Temple, G.F., Brasch, M.A., Thierry-Mieg, N., and Vidal, M. 2000a. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**: 116–122.
- Walhout, A.J., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorton, M.A., van den Heuvel, S., and Vidal, M. 2000b. GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* **328**: 575–592.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.

WEB SITE REFERENCES

- <http://www.genome.gov/10005107>; National Human Genome Research Institute: The ENCODE Project: Encyclopedia of DNA Elements.
- <http://www.fisherscientific.com/>; Fisher Scientific International.
- <http://www.antibodyresource.com/>; The Antibody Resource Page.
- <http://nihroadmap.nih.gov/>; NIH Roadmap: Accelerating Medical Discovery to Improve Health.
- <http://www.biocompare.com/nature/jump/1065/siRNA-Technology.html>; Nature products siRNA technology.
- www.invitrogen.com/gateway; Gateway technology.
- http://home.businesswire.com/portal/site/google/index.jsp?ndmViewId=news_view&newsId=20040504006140&newsLang=en; Gateway technology.