Title: *On the means of bio-production: bioinformatics and how to make knowledge in a high-throughput genomics laboratory.*

Note: This is a post-peer-review, pre-copyedit version of an article published in *Biosocieties*. The definitive publisher-authenticated version appears in volume 6, pp. 271-242 and is available online at: http://www.palgrave-journals.com/biosoc/journal/v6/n2/abs/biosoc201038a.html

Author: Hallam Stevens

<hstevens at ntu dot edu dot sg>

Phone: 617-495-3532

Address: Department of History of Science

Harvard University

Science Center 371

1 Oxford Street

Cambridge, MA 02138

USA.

Short Bio: The author received his PhD from the Department of History of Science at Harvard University in 2010 and now teaches classes on the history of biotechnology and the history of information technology at Harvard.

Abstract

Accounts of 'biocapital' abound in studies of the contemporary biosciences. However, these have tended to pay attention to the use and consumption of biological knowledge rather than the means and conditions of the production of data. This paper draws on an ethnographic account of a high-throughput genomics laboratory (the Eli and Edythe L. Broad Institute, Cambridge MA) to show how the means through which biological data is produced exerts a determinative effect on the kind of knowledge that is generated by the laboratory. High-speed, high-volume, high-efficiency production of data requires the high-throughput consumption of data by statistical and computational techniques. These techniques, in turn, generate general, broad-scale accounts of biological systems, rather than particular knowledge about individual genes or biological components. This cycle of production and consumption is described as 'bioinformatics' in order to indicate the centrality of computers and computing to the knowledge production process in contemporary biology.

Keywords: genomics, bioinformatics, production, consumption, biocapital, Broad Institute

Introduction

As science studies scholars have looked for new ways to characterize biological work in the twenty-first century, notions of 'biocapital' or 'biovalue' have become abundant.[1] Owing much to the work of Karl Marx and Michel Foucault, many accounts attempt to theorize the new relationships between biological life and economic forms that have emerged since about 1975. Edward Yoxen's (1981) notion of 'life as a productive force,' Marilyn Strathern's (1992) view of 'nature, enterprised-up,' Catherine Waldby's (2000) 'biovalue,' Sunder Rajan's (2006) 'biocapital,' and Nicolas Rose's (2007) 'bioeconomics' all seek to capture the essence of what is going on between biology, biotechnology, medicine, politics, and the economy. For the most part, these accounts focus on the consumption and marketing of biological objects – how living things have become a form of property or a commodity, how they have become involved in regimes of speculation and profit generation. These terms point to changes in the way knowledge about life is used, how it is consumed in the biomedical marketplace. But all this is dependent on changing modes of biological labor, changes in the means of *production* of biological data. Stefan Helmreich (2007) has argued that some notions of 'biocapital' trade off a vision or metaphor of organisms as laborers – as "natural factories or assembly lines," they are always and already generators of productivity, surplus, and profit. This points to the fact that these new forms of biology entail more than changes in how biology is used, but also changes in how biological knowledge is made. Just as capitalism involves the reciprocity of production and consumption, biocapitalism too emerges from an interplay between the production and consumption of biological data.

---

[1] For a review see Helmreich 2008.

The aim of this paper is to describe this cycle by scrutinizing biological knowledge in the making. The production-consumption dynamic will be described as 'bioinformatics.' As a term used by biologists, 'bioinformatics' describes a set of practices for storing, searching, organizing, and managing biological data (using information technologies) and making that data into biological knowledge. The usage of the term here is congruent with biologists' own but emphasizes that bioinformatics is not merely the application of computers to the same old biological problems. Rather, bioinformatics entails a fundamental shift in practices of knowledge production: computers have allowed the re-organization and re-valuation of biological work. At first glance, bioinformatics seems to entail a dematerialization of biological objects, "organizing property relations so that they circulate more freely through living bodies" (Mackenzie 2003: 330). Similarly, Eugene Thacker (2005) argues that bioinformatics is a 'virtualization' of biology that draws it closer to global regimes of capital and property.[2] However, bioinformatics is not merely a mobilization or liquefaction of biological stuff – it demands new means of producing data and making knowledge.

A vision of biology-as-production is perhaps by now unsurprising – after all, in 1986, before the Human Genome Project had even begun, Wally Gilbert declared that it was production, *not* science (Siniscalco 1987: 182). What *is* surprising about the descriptions offered in this paper is – first – that the kind of work that is now taking place has established entirely new modes of performing the science of biological work and – second – that these modes are transforming the kind (and not just the quantity) of knowledge produced in biology. What we are witnessing is the generation of new ways being 'productive' in biology and that this 'productivity' is tied not only to new kinds of work, but also to new kinds of knowledge in biology. Gilbert's distinction between science and production has gradually been erased. This hyper-productive, 'bioinformatic' biology – based on computers, automation, and high-throughput – creates a biology oriented towards large volumes of data, towards speed, and

[2] In this regard one might also point to the work of Bronwyn Parry who argues that "new modes of transaction" enable information to be "'rendered' in new, more transmissible forms" (Parry 2004: xviii).

towards accountability. Based on the tools of statistics and data management, this is a biology that requires informatics. This is not just individual biologists making sure to take "care of the data" (Fortun 2007), but large, institutionalized, managerial systems for the efficient production of data-as-product.

A number of historians have described the increasing industrialization of the scientific laboratory during the nineteenth and twentieth centuries. Scholars such as David Noble (1977) have pointed out the intricate interdependencies between science, technology and corporate capitalism in America. Studies have focused on how academic science has become entangled with industry either by relying on corporate money, through scientists sitting on the boards of companies, or acting as consultants, and through increasing dependence on industry for the production of standardized instruments and reagents (Gaudillière and Löwy 1998). In other work, the presence and influence of management practices in engineering has been documented in several contexts (Bugos 1993; Johnson 2002; Westwick 2007). Very meager attention has been given, however, to the direct influence of management practices on knowledge production within the 'fundamental' science laboratories. The influences described here are not mediated by instruments or materials or corporate contracts imported from outside the laboratory, but form the core organizational principles of the laboratory itself.

The present study is based on ethnographic fieldwork at Eli and Edythe L. Broad Institute in 2007 and 2008. Founded by Eric Lander in 1990, the Whitehead/MIT Center for Genome Research became one of the largest genome sequencing centers in the world, contributing about thirty percent of the sequence for the Human Genome Project. In 2002-2003, discussions between Lander, Harvard and MIT led to the creation of the Broad Institute, inaugurated by a $200 million gift from Eli and Edythe Broad (the Broad was formally launched in May 2004). One of the first orders of business was to find a home for the new Institute. The Broad secured ownership of a large building next door to the Whitehead Institute at 7 Cambridge Center (7CC) that had been designed for commercial use and

office space. Although sequencing was an integral part of the Broad's mission, it was quickly realized that there would be insufficient floor space to maintain the large number of heavy sequencing machines at the 7CC site. Consequently, the Broad leased (and later purchased) a second building (a few blocks away at 320 Charles Street) to house its sequencing operations. Hence the Broad is divided into two distinct sites: the impressive steel-and-glass structure at 7CC and the Sequencing Center at 320 Charles Street. The difference between these two buildings could not be more pronounced. Indeed, this paper uses the dual sites of the Broad to explore how bioinformatic knowledge is made by passing through and between these sites – the design and use of the respective buildings represent the 'production' and 'consumption' sides of biological knowledge-making. The Broad is an ideal site at which to study how computers and computation are transforming biology because its organization instantiates the evolving separation between 'wet' (lab bench) and 'dry' (computer) biology that is central to bioinformatics. After describing the 'production' practices at 320 Charles and the 'consumption' practices at 7CC in detail, this paper will show how bioinformatic knowledge emerges as a new kind of biological knowledge from the interplay between these two sets of practices.

I. Production

It's eleven in the morning and I am at the Sequencing Center of the Broad Institute. The building itself – long, squat, and factory-like – is not what one might expect of a state of the art center in high-tech bioscience. Only a few blocks from the Broad's flashy flagship building opposite the Massachusetts Institute of Technology (7CC), the Sequencing Center remains hidden away in East Cambridge amidst older industrial buildings – its neighbors are a pipe-making company and a utility-company outstation (Figure 1).

I am here to interview Meredith, a manager in the Molecular Biology Production Group.[3] Her office sits on a floor above and overlooking the floor of the sequencing lab and as I stroll along the hallway I can look down on the workers busy at their lab benches. The first thing I notice in Meredith's office is that the bookshelves are almost empty except for about fifteen copies of a single book: *The machine that changed the world: the story of lean production* (1991). I ask the obvious question: why all the books? "It's required reading for my employees," she tells me, "every new person on my team gets a copy" (Meredith, personal communication). Perhaps surprisingly, this isn't a book about molecular biology, or about any natural science, but about assembly lines.

The tag line of *The machine that changed the world* is "How Japan's secret weapon in the global auto wars will revolutionize western industry." The book is based on the detailed study of the Japanese automobile industry by three of the directors of the International Motor Vehicle Program at MIT, James P. Womack, Daniel T. Jones, and Daniel Roos. 'Lean production' is the name they give to the techniques deployed in the Japanese car industry (developed largely by Eiji Toyoda and Taiichi Ohno), in order to manufacture high quality products at a low cost.

> The craft producer uses highly skilled workers and simple but flexible tools to make exactly what the consumer asks for – one item at a time.... The mass-producer uses narrowly skilled professional to design products made by unskilled or semiskilled workers tending expensive, single-purpose machines. These churn out standardized products in very high volume... The lean producer, by contrast, combines the advantages of craft and mass production, while avoiding the high cost of the former and the rigidity of the latter. Toward this end, lean producers employ teams of multiskilled workers at all levels of the organization and use highly flexible, increasingly automated machines to produce volumes of products in enormous variety (Womack, Jones, and Roos 1991: 12-13).

At the Broad Sequencing Center, this vision of flexible production is understood to be central to the Institute's goals of investigating human genetics and the causes of human disease. Indeed, in order to understand bioinformatics, it is necessary to understand the details of how its production is organized. Far from being secondary to the scientific activities of the Broad, this organization structures and authorizes genomics and the knowledge it makes.

In mass production, the expense of equipment and intolerance of disruption tends mass-

[3] Where I refer to individuals by first name only, their names of have been changed.

production towards over-supply workers, space, and raw materials; workers are bored and there is little variety in products. Lean production, on the other hand, has the potential to reduce human effort, reduce space, reduce inventory, reduce engineering time, and produce greater variety. One of the goals is also to "push responsibility far down the organizational ladder" making workers able to control their own work. Indeed lean production relies on "an extremely skilled and a highly motivated work force," in which employees become members of a 'community' that must make continuous use of its knowledge and experience. In the Toyota plants, for instance, Ohno placed a cord above every work station which workers could use to stop the entire assembly line if they encountered a problem they could not fix; this was in stark contrast to a mass-production line which could be stopped only by senior line managers in special circumstances (Womack, Jones, and Roos 1991: 57). Workers were encouraged to identify and rectify the cause of the problem. Lean production depends not only on teamwork but on proactive problem solving by every worker.

In keeping with the spirit of valuing workers and their work, lean production depends on the close coordination and cooperation between design engineering and production. At Honda, university-trained mechanical, electrical, and materials engineers spend their first three months of work assembling cars on the production line (Womack, Jones, and Roos 1991: 129). This not only makes the engineers aware of how different aspect of the business work, but also fosters teamwork and communication. It makes engineers acutely aware of the problems that their designs are likely to encounter on the assembly lines, and to anticipate or avoid those problems in their design work.

Lean production also depends on having small supply inventories (sometimes called the just-in-time system). This saves on storage space and builds close relationships with suppliers, allowing sharing of information about products. Having a small inventory means 'working without a safety net' – if there is any defect in a component from a supplier, production will be disrupted. This difficultly is mitigated by what Toyota calls the 'five whys' (presumably: "Why? Why? Why? Why? Why?): "Both

the supplier and the assembler are determined to trace every defective part to its ultimate cause and to ensure that a solution is devised that prevents this from ever happening again" (Womack, Jones, and Roos 1991: 152). Problems with components are solved rather than absorbed into the total cost or time of production.

The Broad Institute, and particularly it sequencing operations, had a commercial tenor from its beginning. Its founder and director, Eric Lander, had worked teaching managerial economics at the Harvard Business School before founding the Whitehead Institute/MIT Center for Genome Research in 1990. Robert Nicol, the director of the high-throughput genome sequencing platform, a chemical engineer by training, worked previously as a project manager for the Fluor Corporation, "the largest US-based, publicly traded engineering and construction firm." Nicol came to MIT in 1999 as a fellow in the Leaders in Manufacturing program to conduct research on manufacturing systems and processes, joining the Whitehead Institute/MIT Center for Genome Research in 2001 in order to implement industrial process design, control, and improvement techniques ("Robert Nicol," 2010).

During my fieldwork at the Broad Sequencing Center, the influence of Nicol's manufacturing and industrial process design ethos could be seen everywhere. Not only lean production, but Six Sigma (6σ) and a range of other manufacturing techniques had been put into practice in biology.[4] First, a large amount of time and effort had been invested in planning and streamlining processes and workflow for sequencing. Space and materials were carefully organized to economize human and sample movement around the labs. Beginning in 2003, the Broad recruited a series of MBA students from the Sloan School of Management (MIT's business school) to investigate the potential for improving the manufacturing capabilities of the lab. As Matthew Vokoun reported, "During the completion of the HGP in the 1990s and early 2000s, the purpose of the Broad Institute's sequencing operations was to

[4] Six Sigma is a business management strategy first implemented by Motorola that attempts to quantify and control variation in output by carefully monitoring and correcting product defects. The name reflects the aim to implement processes that produce products which defective only 0.00034% of the time; that is in which a defects are normally distributed, but occur only as rarely as events six standard deviations from the mean (6-sigma events). See Stamis 2004.

rapidly scale-up or 'industrialize' the genome sequencing process. This industrialization refers to its transition from being a highly skilled craft performed by a few very well-educated biologists to a large-scale, coordinated production process involving over one hundred technicians, engineers, managers, and scientists" (Vokoun 2005: 6). This was achieved by breaking the sequencing process down into small, repetitive steps which could be performed quickly and accurately.

This involved the careful management of both space and the people using that space. Some detailed examples of how space was redesigned and managed can be drawn from the studies undertaken by the Sloan students.

In 2003, Julia Chang (a student from the MIT Sloan Business School) was given the task of analyzing and improving the process of 'picking' *E. coli* colonies for sequencing (Chang 2004). Picking is an automated process through which *E. coli* colonies growing on agar plates – each containing a distinct fragment of DNA – are transferred to 384-well plates for sequencing. Colonies used to be picked by hand using a toothpick. At the Broad, a digital camera images the agar and a specialized software program analyzes the image to determine the position of colonies with desirable characteristics (size, roundness, not too close to other colonies). A computer arm fitted with specialized tips then transfers the suitable colonies to the wells. Chang's mandate was to identify sources of variation in this process and suggest steps to eliminate them. Chang worked with the team responsible for picking to devise a series of experiments to determine what variables most influenced the yield of colonies successfully transferred to wells. This data was analyzed using Chang's experience with operations management and process control theory. One of Chang's key findings was that significant variability in yield was caused by the density of colonies grown on the agar plate. This led to the development of a new process for plating the *E. coli* on the agar using microfluidic techniques, eliminating the inherent variability in the number of cells transferred to the plates with each dispense volume. As Chang noted in her conclusion, "While not widely available or referenced by those in the

organization, sufficient paper records contained the data required to build control charts of the picking process. The documented variability seemed typical of traditional industrial operations and suggested that operational methodologies would have some traction" (Chang 2004: 65). In other words, Chang collected data that had not been considered relevant or interesting to the Broad's technicians and mobilized it to formulate a new and more productive sequencing practices.

The following year, Vokoun, who had worked previously as a process development engineer in the Optical Systems Division at 3M, attempted to apply operations management techniques to Molecular Biology Production Group (MBPG) (Vokoun 2005). As the 'most upstream' part of the sequencing process, the MBPG was the least automated and most 'craft' dependent part of the lab. The aim of Vokoun's work was to transform the MBPG's "highly variable output" by implementing lean manufacturing, production forecasting, Six Sigma, and RFID (radio-frequency identification). Beginning in July 2004, Vokoun managed a five-month lean manufacturing implementation in MBPG with five goals: 1) eliminating all chances of mixing up DNA samples; 2) creating personal workstations with full sets of equipment and materials; 3) minimizing travel for samples and workers; 4) improving and standardizing materials flow; 5) cleaning up and organizing the MBPG area, recovering unused space (Vokoun 2005: 51). These changes were based on several principles of lean production, including 5S, pull production, and *kanban*. In order to understand the changes in lab practice that Vokoun implemented, it is necessary to explain these some detail. 5S, from the Japanese words *seiri, seiton*, *seiso*, *seiketsu,* and *shitsuke* (translated as sort, straighten, shine, standardize, and sustain) is a method for organizing workplaces and keeping them clean. Pull production also refers to a method of organizing workstations by simplifying material flow through the workspace. Workstations are a "sophisticated socio-technical system" in which there is "minimal wasted motion, which refers to any unnecessary time and effort required to assemble a product. Excessive twists or turns, uncomfortable reaches or pickups, and unnecessary walking are all components of wasted motion"

(Vokoun 2005: 49-50). *Kanban* is translated from Japanese as 'visible record' – it embodies the principle that the flow of materials must be carefully managed in order to limit inventory in the pipeline.

Vokoun used 5S, pull production, and *kanban* to recreate the modes of technical production within the MBPG. Working closely with the technicians, Vokoun gained hands-on experience with the ligation processes in order to identify problems: process travel maps were drawn, cycle times were measured, and equipment lists made. Figures 2 and 3 show hand-drawn maps of the movement of workers around the lab during the ligation step. Vokoun's redesigned workflow reduced the manual time involved in the process from 9.3 hours to 6.1 hours (Vokoun 2005: 55-56). Likewise this Figure 4 shows photographs of the ligation workstation before and after redesign according to the principles of 5S. The ligation team also created specialized 'kits' containing all the reagents needed for the preparation of one DNA library, avoiding multiple trips to the storerooms or freezers. As can be seen from these images, Vokoun's focus was on creating efficiencies by ordering space: moving equipment, economizing movement, creating visual cues, and making sure materials were where they could be best utilized. A similar redesign was undertaken for the 'transformation' and 'DNA preparation' steps, resulting in an overall redesign of the MBPG lab space. Vokoun concluded that the problems he encountered "had nothing to do with the actual molecular biology processes performed but rather were managed into the process by the policies, workflow designs, and organizational design of the MBPG" (Vokoun 2005: 105). What made the MBPG – and by extension the Broad as a whole – successful or unsuccessful was not the quality of the 'biology,' conventionally understood, but the attention to the details of the operation as a manufacturing and industrial process.[5]

The requirements of sequencing operations also demanded new ways of organizing people.

---

[5] Several other Sloan students also applied lean principles and other management techniques to aspects of the Broad: Scott Rosenberg (2003) analyzed computer finishing process and Kazunori Maruyama (2005) studied the electrophoretic sequencing process itself.

When I spoke with Will, who had worked on sequencing the human genome at the Broad, he told me that when they began to scale up to production sequencing the Broad "stopped hiring biologists and started hiring engineers and people with management experience" (Will, personal communication). Knowing something about biology was important, but it was even more important to know how to organize a team and how to manage projects. On the floor of the sequencing lab, PhDs in biology are few and far between. Many of the workers are young, often coming straight from undergraduate degrees in biology; there are also a disproportionate number of non-whites and immigrants.[6] The tasks to be performed are often repetitive, but depend on a high degree of skill and precision (for example, pipetting an identical, precise volume of solution over and over). This circumstance – depending on both the repetitiousness of mass-production and the high-skill of craft-production – lends itself to the deployment of lean production.

Critically, workers are given a large amount of responsibility for organizing and improving their own work practices. For instance, every three months workers in the MBPG are given a two week 'sabbatical' to reflect on their work and to come up with schemes for improving and streamlining the processes and workflows in which they are involved. Despite the repetitive nature of many tasks, managers realized that the success of projects ultimately dependent on the skill and the commitment of individuals. One study of computer 'finishers' for example, recognized the difference in interests between 'workers' and 'managers':

> The Center's senior management consisted primarily of academics and researchers, many of whom had pioneered modern gene sequencing. Typically PhDs, these managers held long-term career interests in the field of genomics. Though they ran a production facility, their ambitions also included publication, tenure, and senior roles in industry. This background contrasted sharply with that of the finishing personnel. Coordinators and finishers were typically young, in possession of a bachelor's degree, and at an early stage in their career. Some aspired to long-term careers in genomics or medicine. For others, finishing represented a temporary stopping point on the way to other careers (Rosenberg 2003: 75).

---

[6] Disproportionate with respect to the Broad Institute as a whole, and with respect to the profession of 'biologists' as a whole. The Broad is particularly proud of its large community of Tibetans, most of whom work at the sequencing center; this occasioned a visit by the Dalai Lama to the Broad Sequencing Center in 2003 during his visit to MIT. The pipette used by His Holiness is still mounted on the wall of the sequencing center together with his portrait.

Making finishing more efficient meant re-thinking incentives and re-organizing teams to bring management and workers goals into accord. The challenge was to maintain a sense of "pride and ownership" in work while fostering cooperation and teamwork. Scott Rosenberg, an analyst from MIT's Leaders in Manufacturing Program, proposed new metrics for measuring finisher performance that would foster employee growth, encourage teamwork, and reward innovation, as well as measuring individual performance (Rosenberg 2003: 82-83).

Moreover, Rosenberg proposed new ways to organize finishing teams in order to encourage collaboration and knowledge-sharing. The difference between the original 'skills-based' teaming, which assigned tasks to finishers on the basis of their experience, and the 'triage-based' teaming, which allowed junior finishers to try their hand at more difficult tasks, is illustrated in Figure 5 (Rosenberg 2003: 59-71). By allowing junior finishers to 'hand-off' their work to their more senior colleagues if they could not complete the task, triage promoted communication and knowledge-sharing amongst all finishers. When finishing had been a small-group activity, Rosenberg realized, "its self-image tended to reflect the dedication and individuality of its members," who often worked nights and weekends to complete tasks. But such individual 'heroics' were inappropriate and even counter-productive for a larger, production-line environment: "The organization was simply too large for its members to learn without the aid of better communication and collaboration" (Rosenberg 2003: 78). Triage teaming provided a way to increase productivity while still recognizing and exploiting the special skills of individuals.

In order to encourage commitment to the organization, the Broad provides a career path in biology for individuals without PhDs. In particular, it fosters ways to move from the lab floor into supervisory, management, and planning positions. Several individuals that I interviewed had progressed to their present roles in this way, starting out mixing chemicals on the lab floor and now responsible for large teams and the planning of workflows and sequencing processes. In 2008, Beth had

worked at the Broad for seven years. After working as a health inspector in a local town, Beth had worked for the Massachusetts State Laboratories while earning a masters degree in biology from the Harvard Extension School. Her first job at the Broad had been "making reagents and solutions," and at first she had "no idea what DNA sequencing was." After several years, Beth worked her way up into the technology development team. By the time that I spoke with her in early 2008, Beth had become a project manager in Quality Assurance team, in charge of logistics, supply chain, and quality control for many of the materials coming into the lab (Beth, personal communication). Likewise, Ben came to the Broad with a BA in biology, beginning his career mixing solutions in the materials lab. From here he moved to the "production floor" as part of the MBPG, and finally to the technology development group. In technology development Ben's role was to develop ways to scale up the processes for the new sequencing machines from the bench to mass-production (Ben, personal communication). These examples demonstrate how the Broad operates reward systems outside the traditional academic channels of publication and tenure. Individuals who can work in teams, who exhibit aptitude for logical thinking and planning, who can design processes that bring efficiencies to the data-production process, are promoted. Biological knowledge – especially of fundamental biological principles – is valuable, but it must be combined with an understanding of how a particular production process works and how it might be sped up by reorganizing materials or people. The success of the Broad Sequencing Center depends on a special kind of worker who is neither an automaton in the Fordist sense, nor a lab-bench scientist in the mode of a Pasteur or a Sanger. Instead, he or she (and both genders are well represented) is what might be called a 'lean biologist,' knowing only enough biology in order to perform their work efficiently. The lean biologist combines the individuality and creativity of the scientist with the work-ethic and team-orientation of the production line worker.

In addition to its careful organization of space, materials, and people, a final unique feature of the Broad was its orientation towards control. Keeping space, materials, and people in order means

constant oversight. Meredith told me about what she called Broad's sophisticated sense of 'operations': a few months before we spoke, certain sets of sequences had started to diminish in quality on the sequencing machines, producing shorter read lengths than average. At many labs such a problem would be (at worst) ignored or (at best) take months to resolve, leaving the sequencing machines running at sub-optimal capacity. At the Broad, however, the monitoring was careful and sophisticated enough that the problem could be quickly traced to a particular batch of reagent from a particular outside supplier; the supplier was notified of the defects in the product, quickly supplied a new batch, and the problem was resolved within a few days (Meredith, personal communication).

Such a feat could be achieved through the tracking and monitoring of everything within the sequencing center. From the moment samples enter the lab (and often before), they are given a two-dimensional barcode that links the contents of the sample to its record in the laboratory database. As the sample moves through the lab, the barcode is scanned at every step: each machine in the lab (for example the picking machines) is fitted with a barcode scanner so that the database can keep track of when each sample is run through the machine. Workers in the MBPG have scanners on their benchtops so that samples passing through their workspace are scanned in and out. Using the database, it would be possible to find the exact location of any given sample at any time; it would also be possible to find out which picking or sequencing machine it ran through, whose benchtops it passed over (and how long it spent there) and which batches of chemicals were used to treat it. All over the lab floor large signs remind workers "Never remove a barcode from anything!" The barcoding system is integral to the labs ability to control its operations and monitor its workflow.

Indeed, the barcoding system is just the front-end of a more through-going system of monitoring. This system goes by the names of Quality Control (QC) and Quality Assurance (QA). As a QA project manager, Beth was responsible for developing 'Bills of Materials' (BoMs), detailed lists of quantities of materials used for each step of the sequencing process – by comparing the BoMs with

sequence output, workers could be called to account for the quantities of materials they are using. For instance, a Bill of Materials might allow 100 milliliters of ethanol and 3 pairs of rubber gloves per megabase sequenced; significant deviations from this quickly attract the attention of the quality control teams who investigate the discrepancies. Others in quality control design tests to check the quality of both incoming reagents and outgoing products. Barcodes allow a certain degree of oversight – one could compare, for instance, sequence read lengths from a reagent from supplier A with a similar reagent from supplier B; or the quality scores of data coming from samples prepared by worker A compared with worker B. But often this was not enough – in order to improve processes, 'development' sub-teams in each sequencing team design specific tests to measure the effects of using more or less reagent, or a cheaper alternative, or a faster process. For instance: "could we be using less TAQ polymerase and getting the same quality output?" (Beth, personal communication). These processes allow the Broad to track workers on the lab floor, counting the number of pipette tips they discarded or the amount of a reagent they used in order to perform a particular sequencing step. If particular workers were found to use, for instance, fewer pipette tips for the same quality of product, their techniques could be adopted by a whole team. Little by little, the cost of the whole operation could be whittled down.

Meredith's lab maintained 'tracking sheets' for monitoring work from day-to-day. The tracking sheets record "which libraries we're making, who did it, when they started, how much they started with" (Meredith, personal communication). As well as handwritten notes on a worker's activities, the tracking sheet interfaces with the barcode system: in order to use a reagent the worker must peel off its barcode and attach it to the tracking sheet; at the end of the week the tracking sheets are scanned and the inventories of reagents updated. The electronic record of the tracking sheet is then linked to electronic files containing pictures of gels and results of QC tests. This database is maintained in SAP. Without such a sophisticated records, Meredith tells me, high-throughput would be impossible: the

database allows "fairies" (who resupply the lab) to make sure the lab never runs out of anything: "We don't stop for anything," Meredith reassures me.

> Before when I started here there was no standard tracking sheet. People would do your very common diary-type that molecular biologists do in the lab, page numbers... and they just say, 'this is what they did today, this is what they did today.' ... Which is great, except when you need to troubleshoot and figure out why this is so good or why this is so bad, you go back, and you need to go back many pages, and many times people didn't think that was a very important piece of information to keep... There is not much reliability in the data... When you do a standard tracking sheet, you know its there, and its always there. You also enforce, or at least you can see, that its been done the same way over and over again. This is a production environment and for us variability is hard to deal with, we want to have as little variability as possible and standard tracking sheets are very good for that (Meredith, personal communication).

The details with which such tracking is performed is illustrated in the kind of checklists that Vokoun proposed for the MBPG. Figure 6 shows how workers had to account for the numbers of tips, wipes, tubes, and caps at their workstation each day. Their managers then used a sheet to score their work on the basis of 'shiny clean' floors, 'unused pipettes, tools, [or] fixtures' cluttering their workspace, maintaining checklists and so on (Vokoun 2005: 69-70).

All this depends critically on machines. It is computers that maintain not only the detailed monitoring, but also the careful control over space and people. "Our life is spreadsheets," Meredith told me simply, "We love spreadsheets, we hate spreadsheets" (Meredith, personal communication). But Meredith also told me how some of their needs for managing data had far outgrown the ability of spreadsheets – by now it was really databases that ran the lab: SAP and the Broad's Laboratory Information Management Systems (LIMS) called SQUID. In one way at least the cephalopodic name is appropriate: SQUID's tentacles extend in all directions into all corners of the laboratory, sucking data back to a central repository. Any sample that passes through the lab leaves its trace in SQUID – the history of its movement is recorded in painstaking detail. Natalie, an associate director of the gene sequencing platform, described her work in coordinating and managing sequencing projects. Projects were initiated by generating records in SQUID, projects were monitored by watching their progress through SQUID on a daily and weekly bases, and projects ended when they were removed the database (Natalie, personal communication). At the Broad, the production of sequence becomes an information

management problem. The ability to manage leanly, to create spaces and people amenable to the principles of operational analysis means having the ability to measure, to quantify, and to track. The Broad's raw materials are samples – it deals with thousands; its products are bases of DNA sequence – it produces billions per year; in the middle, petabytes of data are generated. Measuring, quantifying, and tracking is only possible by computer. It is computers, in particular large and sophisticated databases, that have allowed the techniques of production management to be imported into molecular biology.

At 320 Charles, 'productivity' means the lean management and operational analysis of space and personnel. Doing biology means producing high-throughput, high-quality sequence product. The elaborate organization described here is not secondary to or divorced from the core of bioinformatic knowledge. Rather, as we shall see, it plays a critical role in structuring and organizing what sorts of knowledge are made from biological data.

This analysis also suggests that bioproduction shares a great deal with other contemporary modes of capitalist production. In particular, the organization of the Broad could be described in terms of 'flexible specialization' (Harvey 1990) - ways of rapidly adapting workers and equipment to constant innovation. Described in this way, biocapital begins to look merely like one variety of a broader class of post-Fordist modes of production.

II. Consumption

What happens to a piece of sequence? If the sequence – as information – is a kind of commodity, how is it used or consumed in order to make biological knowledge? To witness this, we must visit the Broad's other spaces at 7CC. Here, the work is anything but a production-line. Unlike 320 Charles, the building at 7CC is the epitome of 21st century laboratory chic. Its eight stories of shimmering glass and

metal straddle almost the entire block (Figure 7). The vast double-story lobby houses a custom-designed information center assembled from hundreds of flat-screen televisions and clearly visible from the other side of the street; a small 'museum' of sequencing instruments, plush leather couches, and a serpentine glass staircase, more appropriate to a Californian mansion than a laboratory, fill the remainder of the space. The upper floors of the lab are accessible via a bank of elevators, which can only be activated by an RFID card. Although some of floors have been outfitted for specialized purposes (for instance, to house the large array of servers or the chemical screening robots), the basic pattern of many of the floors is identical. On the east side, and occupying about two-thirds of the floor space, are offices and conference rooms. Almost all of these have glass walls, allowing light to penetrate from the exterior windows into the central area of the building. On the west side of building, separated from the offices only by glass doors and walls, are the laboratory spaces proper.

On the one side, people sit at their computer terminals, on the other they stand at bench tops, pipetting or carefully adjusting various medium-sized instruments. From almost anywhere on either side, it is possible to see directly into the other, even though, for reasons of biological safety, doors must remain closed and air pressures must be maintained at different levels (Tom, personal communication). Standing in the open spaces around and between the offices, the overall impression is one of openness – ample light, neutral tones, glass, and the white spaces of the laboratory creates a space that feels both scientific and business-like, both a lab and a management consultant's office. Tom, one of the architects for the 7CC building told me how Eric Lander wanted 7CC to be a space in which many different types of people could be comfortable: undergraduate and graduate students, senior faculty scientists, software engineers, consultants, and visitors from the NIH and NSF. Lander wanted a "sock track" around each floor of the building – a 'racetrack corridor' which one could traverse with one's shoes off and still be able to see every aspect of the laboratory at work (of course no-one would be allowed in the wet lab spaces proper without shoes on). Such a pathway was designed

so that Lander and other lab leaders would be able to show visitors around the lab – to show off the biology that was being performed there to other scientists and to potential donors and funders. Originally the building at 7CC had been conceived as space for retail. However, Lander saw the open spaces and large amounts of glass as an opportunity to create a highly transparent laboratory space. "Eric Lander's instructions to us," Tom recalled, "were that he wanted [the lab] to be about transparency: clean, bright, and open. It was a philosophical desire. Since the purpose was to do genomic studies, and make results available, the transparency of discovery should make its way into the architecture" (Tom, personal communication).[7] As a place of both science and business, an amalgam between office and laboratory, 7CC reminds the observer of a physician's office – the cleanliness and order, the light and neutral tones, murmured conversations in the hallways. 7CC contrives to create an image of a highly collaborative space of Mertonian free inquiry, creativity, and open communication.

In order to characterize the kind of work that takes place at 7CC, it is useful to describe one example in detail. In 2005, the Broad Institute published a paper based on the draft of the chimpanzee (*Pan troglodytes*) genome (Chimpanzee Sequencing and Analysis Consortium 2005). The sequencing work alone – performed at both 320 Charles and at the Genome Sequencing Center at the Washington University School of Medicine – was not considered publishable material in and of itself. Indeed, the chimpanzee sequence had been uploaded to public databases (GenBank and EMBL-Bank) as it was completed and had been available for some time.[8] Rather, the 'chimpanzee genome paper' consisted largely of analysis and comparative work performed *on* the sequences themselves. In other words, it was the work of 7CC to turn biological data into publishable scientific knowledge.

---

[7] For more on the design of the Broad Institute, and especially its 'transparency' see Higginbotham 2006 and Silverberg. 2007.

[8] In other words, the sequence itself was not 'published.' This raises the dilemma that it would be possible for a third party to download and analyze the chimpanzee data and publish this analysis, thereby scooping the sequencing lab. However, a set of informal rules – agreed to at Fort Lauderdale in 2003 – allow the sequencing lab first rights to publish on the organism that they sequence. In theory, a journal would not accept a paper from another group.

In June 2007, I spoke with James, a lead author on the chimpanzee paper. Through this conversation (and many others like it) I developed a sense of the kinds of work entailed in making sequence data into knowledge. First, the data must be thoroughly characterized in terms of its quality and accuracy. In the chimpanzee, nucleotide-level accuracy was assessed not just by quality scores, but by aligning and comparing segments of the sequence to chimpanzee protein-coding regions sequenced by other methods (BACs rather than whole genome shotgun) using tools such as BLASTZ and BLAT. Structural similarity (the large-scale ordering of the various pieces of sequence) was also checked by comparison with sequences from other chimpanzees sequenced as part of the project. The fragmentation of draft sequence was also analyzed, again by comparison of shotgun-sequenced regions with BAC-sequenced regions. Such work entails downloading the sequences from various internal and external repositories and running it through software that performs sequence alignment and generates statistical information regarding the sequences.

Much of the software for performing such analysis is custom built. As such, a second aspect of this work involves the building of software tools. James told me how much of his work depended on "looking at the data himself." This usually meant writing short 'scripts' (programs) to perform some analysis and then compile it into a table or a chart (James, personal communication). For example, a simple script might locate the positions of all the genes on a chromosome and then generate a plot showing the density of genes across that chromosome. For James this was often about using visualizations to see patterns in the data, ideally a kind of "Google Earth for looking at the data." Programming tools allowed the vast amounts sequence data to be 'managed' – made visible and therefore comprehensible. Through building software, the raw sequence data was manipulated into biologically meaningful knowledge about the chimpanzee.

The third kind of work performed at 7CC is the statistical analysis of data. The distribution of genes, GC-content, single-nucleotide substitution rates, the frequency of indel events, the occurrence of

transposable elements, rates of evolutionary divergence, and other similar numbers are counted and calculated using more or less sophisticated statistical techniques. For instance, in the chimpanzee genome paper, the authors compute the nucleotide divergence rate (compared to humans, the chimpanzee's closest relative), in 1-megabase windows across the genome; a wide distribution of such rates suggests that there is regional variation in mutation rates across chromosomes (Chimpanzee Sequencing and Analysis Consortium 2005). Again, raw sequence data is reduced to a (publishable) biological fact about the mechanisms of evolutionary mutation. Statistical techniques are used to summarize and manage the vast amounts of sequence data, to render information into knowledge. Part of what is at issue here are new criteria for what counts as 'understanding' biological systems and mechanisms. Comparison of sequence does not directly explain the mechanisms responsible for differences in anatomy, physiology and pathology between the humans and chimpanzees. It may suggest specific genes or other sequences responsible for particular differences, while giving providing no direct evidence of the mechanisms or pathways through which such genes might act. Most often, sequence analysis provides overall patterns (for instance, patterns of mutation rates) that statistically summarize inter-species differences. Understanding the chimp genome means discovering these hidden patterns through computational analysis.

Finally, much of the work at 7CC is directed towards understanding the evolution of genes and genomes. The most valuable knowledge from genomes often comes from closely comparing related genomes – the rationale for sequencing mice, dogs, and elephant genomes is based on the notion that it can aid in understanding the human genome (particularly, 'what makes us human'). James explained comparative genomics as a search for a 'parts list': if a particular part of the genome is very highly conserved (very similar) between different species then it is likely that that segment serves a particular and important function. Learning about genes and other functional elements in the human genome means comparing it to other more or less distantly related species. In the case of the chimpanzee, the

genome is expected to be so closely related to our own that the analysis proceeds differently. Instead of looking for similarities, comparison between human and chimpanzee aimed to discover the few sparse differences that distinguished the two. A large amount of effort was spent aligning the human and chimpanzee genome and identifying orthologs (highly conserved regions).[9] This analysis – using tools such as Clustal W – is computationally expensive, especially for generating alignments between more than two species. This kind of work is of crucial importance in posing and answering biological questions of large significance. In the Broad press release for the chimpanzee genome James made a strong case for the importance of this work: "By cross-referencing this catalog [of genetic changes] against clinical observations and other biological data we can begin to identify the specific changes that underlie the unique traits of the human species" (Broad Institute Communications 2005). This 'cross-referencing' transforms a set of nucleotide data into a crucial piece of biological knowledge about human uniqueness.

The kind of work at 7CC described here can be characterized as accountable, collaborative, data-driven, and oriented towards efficiency. Sequence data must first be verified and certified as reliable through computational and statistical processes. The software engineering and statistical analysis aspects of the work typically involve interdisciplinary, medium-sized teams: the chimpanzee genome paper listed sixty-seven individuals (including those trained in biology, computer science, mathematics, chemistry, statistics, physics, engineering, and other fields). Moreover, these practices are directed towards the management of large amounts of data and reducing it into readily digestible or understandable forms. 7CCs techniques aim to understand the genome in its totality – to draw conclusions about all genes, or all transposons, or all repeating elements in a genome. Software and statistics generate a high-throughput of knowledge, an efficient transformation of data into publications. James related how he thought that computers had changed the questions that were asked

[9] 13 454 1:1 human-chimpanzee orthologs were found. The Chimpanzee Sequencing and Analysis Consortium 2005.

in this sort of biological work – the emphasis falls on general problems, on big patterns and correlations (James, personal communication). The work of those at 7CC becomes to analyze and make sense of a vast output of data: comparing genomes, describing the whole human microbiome, analysis and classification of all forms of cancer, mapping the totality of human haplotypes, and so on.[10] Whereas pre-bioinformatic biologists usually dealt with a single organism, perhaps even a single gene – devoting their careers and the careers of the graduate students to characterizing it in detail – bioinformatic biology asks (and answers) bigger questions, questions that rely on knowledge of all human genes, or a comparison across twenty or more organisms. In short, they depend on having extraordinarily large volumes of reliable and accessible data. Indeed, the biologists at 7CC are in constant need of more sequence data. As one explained it to me:

> The strategy I'm trying to adopt is: let's take some of the classic tests and measures [for the adaptation and evolution of proteins] and recast them in ways where, if we had all the data in the world, then the result would be clearly interpretable. The *only* problem that we have is not having all the data in the world (Graham, personal communication).

Such work proceeds as if "not having all the data in the world" is a problem that will soon be solved for all practical purposes.

The consequence of an obsession with speed and quantity is that biologists are routinely answering different kinds of questions. In the Broad the specific nucleotide, the specific gene, the specific genome, even the specific disease, move into the background. It is the *general* problems that are at issue: how do all genes work? What are the overall rules or patterns for how genomes or diseases behave? Many informants remarked to me how they perceived this as a major break with other kinds of biology which deal with single proteins, single genes, and single species, spending years of work at the bench to determine the particularities and peculiarities of a locally defined object. Although it is true

---

[10] The chimpanzee paper is fairly typical of the kind of large-scale work that the Broad undertakes. Their website lists active areas of research as: "deciphering all the information encoded in the human genome; understanding human genetic variation and its role in disease; compiling a complete molecular description of human cancers..." As well as many genome projects, work has included attempts to completely characterize human genetic variation (HapMap, 1000 Genomes), work to completely characterize cancer and its genetics, and work to create an RNAi library which covers every known human gene.

that more traditional biologists are also interested in such general questions about life, their day-to-day focus usually remains specific; without genome-scale data (and bioinformatic techniques to digest it) answering such questions must rely on speculative extrapolation from a few examples. The big questions remain the same, but the answers that can be given are of a markedly different kind.

This genome-wide, brute force approach is certainly not without its opponents - some biologists argue that data collection should slow and that much more work should be done at the lab bench to help understand the flood of data already deposited in the databases. For instance, the Nobel Laureate Sydney Brenner argues:

> Right now, we're wandering through the databases, trying to get rid of all the weeds that mean nothing, and sometimes we can't recognize the weed… There are all these guys saying 'More is better; let's get huge amounts of data.' I say the least is better. You do the least and then you can calculate or compute whatever you want. Otherwise it's meaningless. (Quoted in Duncan 2004).

For Brenner, doing biology and answering biological questions means using the data to explore a particular system or mechanism in order to characterize it thoroughly. Bioinformatics is aiming at different kinds of answers: understanding means revealing the patterns and statistical trends in the data as a whole.[11] That the bioinformatic approach meets with skepticism and criticism, especially from older biologists, further suggests that it poses a challenge to traditional modes of doing and knowing.

The high volume of data generated at 320 Charles drives the work at 7CC: the regimes of productivity at that lab drive biological work not just on the sequencing floor, but also frame how biologists frame and answer questions. We can understand 320 Charles and 7CC as engaged in a dialectic of production and consumption: the data produced at 320 Charles is consumed at 7CC and converted into publications which result in scientific credit and (ultimately) more funding for the Broad as a whole. 7CC devises computational-statistical techniques to consume data ever more rapidly, leading to larger and more general conclusions about biological systems; this justifies an accelerating

[11] For example: a bioinformatic project to understand mRNA alternative splicing aims to discover the splicing 'code' -- the sequence patterns that influence or determine how the mRNA gets sliced. The approaches (based largely on sequence data) do not attempt to understand the splicing mechanisms (the spliceosome complex), but rather analyze large volume of alternative splicing sequence data to find common patterns.

output of data from 320 Charles, which in turn encourages 7CC to ask larger questions and devise even more powerful data-reduction techniques. 320 Charles and 7CC are bound together in an 'economic' cycle in which greater and greater volumes of data are traded. The kinds of biological knowledge that the Broad produces is dependent on the 'bioinformatic' regimes of productivity and data-management seen at 320 Charles – that is, high-throughput, high-speed, high-volume data analysis generates high-throughput, high-speed, high-volume knowledge. It is not merely that sequencing is production, as Gilbert predicted, but that bioinformatics – understood as data management – has fundamentally transformed what it means to do productive biological work and hence what sorts of questions biologists are asking and what knowledge they are making. 'Productivity' has penetrated to the core of understanding life.

III: Bioinformatics as a productive force

Bioinformatics is not merely the computerization or informatization of biology. Capitalist regimes of property, accumulation, and global circulation do not enter biology – as Thacker (2005) argues – through the 'coding' of life into virtual and computerized forms. Rather, bioinformatics entails a reorganization of biological practice that is changing the notion of valuable work in biology and that this in turn is impacting on the kind of knowledge being produced. The dialectic between production and consumption can be characterized in terms of three transformations in biological work.

First, it is changing the scale at which biology is practiced. Small-scale bench work – where an investigator or a lab might examine a particular gene or protein for years – is increasingly rare. Instead, bioinformatics uses computers to analyze whole genomes, whole organisms. This is Big Biology where volume, speed, and scale are crucial. The work accorded value in the Broad Sequencing Center is not the highly individualistic, highly innovative work of the traditional bench scientist. What is valuable is

instead teamwork, and the careful organization and management of an interdisciplinary team. This has opened biology up to new sorts of workers with new sorts of skillsets. The culture of the Broad suggests a shift in what sorts of people are doing biological work and changes in the distribution of labor. Whereas previously biology was performed almost exclusively by PhD scientists (and their graduate students and post-docs), the biology at the Broad demands a workforce that is trained not only in biology, and whose skills might be transferable to a range of industries. While these others are busy with the laboratory/manufacturing work of sequencing, the PhD biologists are engaged in a quite distinct set of tasks, often physically and intellectually removed from the lab bench.

Second, work at the Broad is based on a new accounting of biological work; the lab is funded not only according to how many papers it publishes or how promising its research seems, but also on the basis of dollars per base. The National Human Genome Research Institute (NHGRI, from where a large proportion of the money comes) prefers the Broad because it can offer its product (sequence) at a cheaper rate than its competitors.[12] This 'accountability' is passed down through the hierarchy of the organization to the bench worker who must also be held to account for his or her own productions. This changes how work is rewarded and what is considered 'good' work. New forms of 'progress' have emerged which prioritize the accumulation of more and more sequencing data at an ever-decreasing cost. Progress is not only about novel ideas or novel facts or new understanding of organisms but can include discoveries that will increase productive output by decreasing variability.

Finally, this new kind of biological lab has become a space of surveillance to an extent previously unusual in the sciences. Through computers, people, objects, and spaces are constantly monitored; every pipette tip wasted or moment spent chatting to your colleague leaves a discernible informatic trace. Everything must be accounted for. Here, biology has become a sort of informatic Panopticon; Natalie told me that she liked her job because the 30 000 foot view of the Broad's work

---

[12] See the latest Request for Applications for NHGRI grants ("The Genome Sequencing Centers (U54)," 2006).

provided an appealing sense of control (Natalie, personal communication). Doing good and interesting work means keeping watch and being watched. What is valuable is the production of a high-quality and uniform product. This work is about attention to detail, repeatability, and accuracy. Doing biology has become manufacturing, has become commodity production. Moreover, this suggests new modes in which biological knowledge is authorized and authenticated. A large body of work in the history and sociology of science has elaborated on how the building of scientific knowledge depends on trust – in the early modern period this was trust between gentleman (Shapin and Schaffer 1985) and later trust guaranteed by mechanisms and institutions of expertise (eg. peer review). Bioinformatic knowledge is not verified or authenticated in these modes – rather it is trusted because it is a product. It is the carefully controlled and managerially monitored processes that produce sequence data that lead to its reliability. Just as we trust our Toyota cars or our Nokia cell phones to function appropriately and consistently, the base pairs produced by the Broad can be relied upon as mass-produced items.[13] This is a form of 'mechanical objectivity' (Daston and Galison 2007) refracted through labor and manufactures.

All this suggests that biological knowledge production – in genomics at least – has undergone a fundamental transformation. Certified and valuable knowledge is high quality, high quantity; it must be quality checked, scrutinized, and monitored throughout its production. It must be accountable, both in the sense that it be carefully costed, and its provenance (recorded in the database) can be rigorously checked. Once again it was the computer that allowed the concepts of lean management, six sigma, and so on, to be implemented in biology; the mass production of (sequence) data as a product required computers in order to make the sequencing process visible, manageable, and accountable. On the consumption side, ever more sophisticated statistical-computational tools are developed to consume ever larger quantities of data into valuable knowledge. At the Broad, the organization of biology in

[13] The author notes the irony here in the fact that in 2009 Toyota came under serious scrutiny for the failure of their processes to produce consistent and reliable products.

accord with 'business principles' seems to have reached its apotheosis. It is computers and databases that have made this transformation to a new form of knowledge-production possible.

This sort of work can be characterized by mass and speed, by asking questions involving large numbers sequences, genes, organisms, species, and so on. Computers have historically been tools for business – for speeding up, for making efficient, for managing and increasing productivity. Jon Agar has shown in detail how computational practices were linked to and evolved from efforts to rationalize government bureaucracy in the United Kingdom (Agar 2003). Martin Campbell-Kelly and William Aspray write that in the 1950s "the computer was reconstructed – mainly by computer manufacturers and business users – to be an electronic data-processing machine rather than a mathematical instrument" (Campbell-Kelly and Asprey 1996: 105). Many of these 'calculators' were used at first for accounting purposes, and later for administration and management. Computers were used for payroll calculations, sales statistics, and inventory control (Haigh 2001). The advent of the UNIVAC and the IBM 701 in the early 1950s made computers valuable to business as machines able to 'automate' and speed-up routine tasks. At General Electric (where one of the first UNIVACs was installed), the digital computer was used first to replace salaried clerks (and their overhead costs) and later for long-range planning, market forecasting, and revamping production processes (Ceruzzi 1998: 32-33). For IBM, and other companies attempting to compete with them, the computer had to be designed to the needs of business: alphanumeric processing, checking and redundancy mechanisms, 'buffers' for for high-speed transfers, magnetic tape storage, and variable length records. Such features later became commonplace in all digital computers. Even the von Neumann architecture – particularly its stored-program design – suited commercial needs, since it allowed for the storage and rapid sorting of data. As James Cortada has argued, "a quick look at how computers were used suggests that the history of the digital computer is every bit as much a business story as it is a tale of technological evolution" (Cortada 1996: 160).[14] In

[14] On the history of computers in business see also Edwards 2001.

the 1950 and 1960s, the computer developed as a tool for exercising close control over a corporation and making business more efficient. This preoccupation was reflected in their design.

In biology, computers were first deployed by mathematicians and physicists in the 1950s for the purposes of cracking the genetic code (Kay 2000). In the 1960s and 70s, some efforts were made to adapt computers as tools for biological work, especially for instrument control and as artificial intelligence for the laboratory (November 2006). These efforts were limited in scope and did not affect the mainstream of biological work. In the 1980s, computer re-entered biology as business machines – that is, as machines designed to speed up, to make efficient. For example, attempting to search biological databases by hand and eye was in principle possible, but in practice impossible: computer techniques brought speed, efficiency and accuracy to the process. As such, the principles and practices of bioinformatics were always and already 'industrial' in an important sense: they were attempts to streamline information flow and knowledge production in biology. This has two significant consequences. It is through computerization – through the principles and practices of bioinformatics – that genomics has become 'industrialized' and 'commercialized.' In many cases, industrialization began with and was made possible by automation and informatics; the regimes of accounting and business practices that dominate contemporary genomics find their origin in, and are made possible by computing practices. I want to suggest that computers carry with them certain modes of usage associated with their use as business machines – that is, they carry a 'politics' of productivity.

Michael Fortun has drawn attention to the cultures of 'acceleration' and 'speed' surrounding the Human Genome Project. He argues that much of the controversy surrounding the project was connected to its speed – many wondered why the Genome project was considered to be so urgent. The commitment to the HGP involved a commitment to prioritizing (and therefore speeding up) certain kinds of biological work in particular labs (Fortun 1999). Bioinformatics has become the space in which the conflict between these older and newer (slower and faster) forms of practice ultimately

played out. Already by the mid-1980s, "Informatics was becoming the most important speed nexus" (Fortun 1999: 30); computers were crucial for the HGP not just to achieve automation, but also to manage and maintain the speed of data production. The continued drive towards rapid and large-scale biology was created not by the HGP but by computers.

The importation of computers into the life sciences has brought with it not only changes in practice, but also a transformation of the values and knowledge regimes of biology. The descriptions of work at the Broad show how the 'business' modalities of the computer changed the kind of work that is performed and the kinds of questions that are asked and answered. These machines facilitate a cycle of data production and consumption which drives biological knowledge-making. Bioinformatics is often understood as the informatization of biology, an erasure of the boundaries between life *in vivo* and life *in silico.* But its most important consequences are not to be understood in terms of 'dematerialization' of life, but rather in the transformation of biology into a 'biocapitalist' system of production and consumption: life itself becomes a 'productive force' oriented towards the speeding up, making efficient, and scaling up of knowledge-making.

Acknowledgments:

Thanks to Peter Galison, Stefan Helmreich, the STS Circle at the Kennedy School of Government, and the Modern Science Working Group in the Department of History of Science at Harvard for reading early drafts of this paper. This work was supported in part by the Wenner-Gren Foundation for Anthropological Research (#7745) and a Doctoral Dissertation Research Improvement Grant from the National Science Foundation (#0724699).

Reference list

Agar, J. (2003). *The government machine: a revolutionary history of the computer.* Cambridge, MA: MIT Press.

Bugos, G. (1993). Manufacturing certainty: testing and program management for the F-4 Phantom II. *Social Studies of Science,* 23, 265-300.

Broad Institute Communications (2005). Comparison of human and chimpanzee genomes reveals striking similarities and differences. Press Release, August 31. Retrived May 10, 2010 from, http://broadinstitute.org/news/263

Campbell-Kelly, M. & Aspray, W. (1996). *Computer: a history of the information machine.* New York: Basic Books.

Cerruzi, P.E. (1998). *A history of modern computing.* Cambridge, MA: MIT Press.

Chang, J.L. (2004). Control and optimization of E. coli picking process for DNA sequencing. Masters Thesis submitted to Sloan School of Management and Department of Chemical Engineering, MIT.

The Chimpanzee Sequencing and Analysis Consortium. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature,* 437, 69-87.

Cortada, J.W. (1996). *Information technology as business history: issues in the history and management of computers.* Westport, CT: Greenwood Press.

Daston, L. and Galison, P. (2007). *Objectivity.* Boston: Zone Books.

Duncan, D.E. (2004). "Discover dialogue: Sydney Brenner" *Discover* (April). Retrieved October 11, 2010 from http://discovermagazine.com/2004/apr/discover-dialogue/

Edwards, P.N. (2001). Making history: new directions in computer historiography. *IEEE Annals of the history of computing* 23(1), 85-87.

Fortun, M. (1999). Projecting speed genomics. In: M. Fortun and E. Mendelsohn (Eds.), *The Practices of Human Genetics*, 25-48. Boston: Kluwer Academic.

Fortun, M. (2007). "The care of the data" Retrieved October 11, 2010, from

http://www.fortuns.org/?page_id=72

Gaudillière, J-P. & Löwy, I. Eds. (1998). *The invisible industrialist: manufactures and the production of scientific knowledge.* New York: St. Martin's Press.

The Genome Sequencing Centers (U54) (2006). Retrieved August 28, 2009, from http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-06-001.html

Haigh, T. (2001). The chromium-plated tabulator: institutionalizing an electronic revolution, 1954-1958. *IEEE Annals of the history of computing* 23(October-December), 75-104.

Harvey, D. (1990). *The condition of postmodernity: an enquiry into the origins of cultural change.* Malden, MA: Blackwell.

Helmreich, S. (2007). Blue-green capital, biotechnological circulation and an oceanic imaginary: a critique of biopolitical economy. *Biosocieties* 2(3), 287-302.

Helmreich, S. (2008). Species of biocapital. *Science as culture* 17(4), 463-478.

Higginbotham, J. (2006). Collaborative venture produces versatile open laboratory. Retrieved October 3, 2008, from http://www.rdmag.com/

Johnson, S. B. (2002). *The secret of Apollo: systems management in American and European space programs*. Baltimore: Johns Hopkins University Press.

Kay, L. E. (2000). *Who wrote the book of life? A history of the genetic code.* Stanford: Stanford University Press.

Mackenzie, A. (2003). "Bringing sequences to life: how bioinformatics corporealizes sequence data" *New genetics and society* 22(3), 315-332.

Maruyama, K. (2005). Genome sequencing technology: Improvement of electrophoretic sequenceing process and analysis of the sequencing tool industry. Masters Thesis, Sloan School of Management and Department of Chemical Engineering, MIT.

Noble, D. (1977). *America by design: Science, technology and the rise of corporate capitalism.* New York: Alfred A. Knopf.

November, J. (2006). "Digitizing life: the introduction of computers into biology and medicine" PhD Thesis, Department of History, Princeton University.

Parry, B. (2004). *Trading the genome: investigating the commodification of bio-information.* New York: Columbia University Press.

Rajan, K. S. (2006). *Biocapital: the constitution of post-genomic life*. Durham: Duke University Press.

Robert Nicol. (2010) Retrieved 20 August, 2009, from http://www.broadinstitute.org/about/bios/bio-nicol.html

Rose, N. (2007). *The politics of life itself: biomedicine, power, and subjectivity in the twenty-first century.* Princeton: Princeton University Press.

Rosenberg, S.A. (2003). Managing a data analysis production line: an example from the Whitehead/MIT Center for Genomic Research. Masters Thesis, Sloan School of Management and Department of Electrical Engineering / Computer Science, MIT.

Shapin, S. and Schaffer, S. (1985). *Leviathan and the air pump: Hobbes, Boyle, and the experimental life.* Princeton: Princeton University Press.

Silverberg, M. (2007). The glass lab. Retrieved October 3, 2008, from http://www.metropolismag.com/cda/story.php?artid=2517

Siniscalco, M. (1987). On the strategies and priorities for sequencing the human genome: a personal view. *Trends in genetics* 3(7), 182-184.

Stamatis, D.H. (2004). *Six sigma fundamentals: a complete guide to the systems, methods, and tools.* Productivity Press.

Strathern, M. (1992). *After nature: English kinship in the late twentieth century.* Cambridge: Cambridge University Press.

Thacker, E. (2005). The *global genome: biotechnology, politics, and culture.* Cambridge, MA: MIT Press.

Vokoun, M. R.(2005). Operations Capability Improvement of a Molecular Biology Laboratory in a High Throughput Genome Sequencing Center. Masters Thesis submitted to Sloan School of Management and Department of Chemical Engineering, MIT.

Waldby, C. (2000). *The visible human project: informatic bodies and post-human medicine*. London: Routledge.

Westwick, P.J. (2007). Reengineering engineers: management philosophies at the Jet Propulsion Laboratory in the 1990s. *Technology and culture* 48.1, 67-91.

Womack, J.P., Jones, D.T. & Roos, D. (1991). *The machine that changed the world: the story of lean production*. New York: Harper Perennial.

Yoxen, E. (1981). Life as a productive force: capitalizing upon research in molecular biology. In. L. Levidow and R. Young (Eds). *Science, technology and labour processes,* 66-122. London: Blackrose Press, pp. 66-122.

Figure captions:

Figure 1: Exterior of Broad Institute at 320 Charles Street (Sequencing Center). Source: Image by author.

Figure 2: Hand drawn map of worker movements in ligation step before lean-production re-design. Source: Vokoun 2005. By permission of Massachuetts Institute of Technology.

Figure 3: Hand drawn map of worker movements in ligation step after lean-production re-design. Source: Vokoun 2005. By permission of Massachuetts Institute of Technology.

Figure 4: Ligation workstation before and after redesign according to the principles of 5S. Source: Vokoun 2005. By permission of Massachuetts Institute of Technology.

Figure 5: 'Skills-based' (top) versus 'triage-based' (bottom) teams in sequence finishing. Source: Rosenberg 2003. By permission of Massachuetts Institute of Technology.

Figure 6: Molecular biology production group worker checklist. Source: Vokoun 2005. By permission of Massachuetts Institute of Technology.

Figure 7: Exterior of Broad Institute of Harvard and MIT at 7 Cambridge Center, Main Street, Kendall Square. Source: Image by author.