***Chapter 3: organizing space***

*3.0 Introduction*

What do the spaces in which bioinformatic knowledge is produced look like? How are they arranged? How do people move around in them? What difference does this make to the knowledge that is produced? The dynamics of data exchange have driven a spatial reorganization of biological work. That is, data work demands that people and laboratories are arranged and organized in specific ways. I mean here how walls, hallways, offices, and benches are set out and how people move amongst them. These arrangements and movements are crucial in certifying and authorizing bioinformatic knowledge – the motion of data through space and between people renders it more or less valuable, more or less plausible. Because of this, spatial motion is also bound up with struggles between different kinds of work and the value of different kinds of knowledge in contemporary biology.

The volume and speed associated with high-throughput techniques have transformed biologists’ knowledge-making practices. But computers have also necessitated reorganization of workers and space that transform what counts as useful knowledge and work for biology. In particular, laboratory work itself is managed as data: samples, movements, and people are recorded in databases. This has led to the quantification and control of space and work. This entails not just speeding up, but rather making speed and efficiency *virtues* of biological work.

Biological data is made valuable – it can be shared, traded, exchanged, and even bought and sold under appropriate circumstances.[[1]](#footnote-1) It is a form of capital for performing biological work. Before the Human Genome Project began, Wally Gilbert quipped that genome sequencing would not be science, but production. But biology has *become* production – a large proportion of contemporary biology (particularly genomics) turns on the production of a product, namely data. The machines of this production are computers. The data-orientation of contemporary biology should also be understood as an orientation towards productivity. Data has generated new modes of *value* in biological work. Gilbert no doubt meant to denigrate work on the human genome as *mere* production – mindless, uncreative activity that stood in contrast to the work of proper science. But the pejorative connotation now rings hollow – successful production requires high technology, intricate management, difficult problem solving. It is these manufactures that will comprise the raw material in the manufacturing of new drugs and new medical treatments. Doing ‘good’ work in bioinformatics is now doing *productive* work – that is work that contributes to the rapid and effective creation and organization of data.

The link between space and the value of different kinds of biological work was already visible in the HGP. A significant part of the debate about the Project centered on whether to distribute resources for sequencing widely or to focus on a few large, centralized genome centers. Many feared that the HGP would compete with and ultimately reduce resources for the type of small-scale biology that had given rise to the most important discoveries.[[2]](#footnote-2) At the heart of concerns over the centralized approach was the suspicion that the HGP did not really amount to science. It was feared that a ‘technological’ focus would have a pronounced influence on the future direction of biological work, teaching young biologists just techniques, but not creativity.[[3]](#footnote-3) These debates were simultaneously a contest over both the organization of biology *in space* and the *value* of certain kinds of biological work. In particular, what was in question was the value of ‘technological’ work of using producing, storing, managing, and sharing data.

In the first part of this chapter, I describe a division in bioinformatics between ‘producers’ and ‘consumers’ of data. Production and consumption form a cycle in which both parts are crucial for the generation of biological knowledge, but data-producers are relegated to a lower status. I show how these divisions of labor are inscribed into the spaces of biological work and how biological knowledge is authorized through its movement between spaces. The last part of the chapter examines the production of data in more detail. It shows how production sequencing requires the kinds of efficient and accountable work that computers enable.

In describing Robert Boyle’s ‘House of Experiment,’ Steven Shapin has shown how the physical spaces (and the flow of people, materials, data within them) crucially structured the knowledge that was produced: “the siting of knowledge-making practices contributed toward a practical solution of epistemological problems.”[[4]](#footnote-4) The organization and policing of space, particularly the division into private and public, was central to the validation and authorization of knowledge claims. Despite the manifest differences between seventeenth century houses and modern laboratories, something similar can be said about bioinformatics: how labs are organized and how data flows between them contributes to the solution of the practical epistemological problems posed by computer scientists, biologists, and database managers having to work together. Biological knowledge, the forms of labor, and the spaces of bioinformatics reproduce one another. The kinds of epistemic changes described in Chapter Two require the simultaneous transformations in work and space that are detailed here.

*3.1 Divisions of labor: Producers and consumers*

Attention to the dynamics of contemporary biology shows that there are two distinct kinds of work being carried on. First, there are those working to transform samples into data; second, other individuals analyze this data and make into biological knowledge. This can be described as the production and consumption of data, and it is through a cycle of production and consumption that bioinformatic knowledge is generated. Both production and consumption require both biological and computational expertise. By marking consumption of data as high value work, while considering data production lower value, biologists are able to police the kinds of knowledge that are considered valuable and acceptable in bioinformatics. This can be understood as a form of continued resistance to (and discomfort with) computational and data driven approaches by biologists; data driven biology has required biologists to defend their skills and expertise within a landscape of new problem-solving approaches. By suppressing production as less ‘scientific’ or ‘biological,’ biologists are able to maintain control of biological work.

In an interview with one biologist who had spent significant amounts of time in industry, spoke at length about how he understood the difference between different sorts of workers:

IT people are really like a dime a dozen... they are interchangeable parts for most large corporations; but a stats guy you just can't do that with... statistics is hard, computers are not... I experienced this first hand at the Whitehead... There would be these guys that would get hired in and they would get fired within a month because they were just programmers. And they would call themselves bioinformaticians. So we started to make a distinction at the Whitehead between bioinformaticians and computational biologists... If you went to the Whitehead or the Broad today and said: “I'm a computational biologist,” you’d have to prove yourself. And to prove yourself you’d be writing programs to do linear regressions models or peak finding for something, and if you’re a bioinformatician you’d be running the Oracle database. It’s like saying “I’m a computer scientist” versus a “I’m a computer engineer.” [At the Broad] we were having a problem where we would hire people who said they were bioinformaticians and we would ask them to do something statistical or mathematical and they would just fall on their faces... We would ask them to tell us what a Z-score means and they would be like: “uh, duh, I don't know.” How do you call yourself a bioinformatician then? “Well I know how to run a database and query GenBank and get some sequences out.” Yeah, everybody on the planet can do that! We would be like, “What do you mean, like type the sequence into the command line and get the file back and parse it with a parser you wrote ten years ago?” Yeah. Oh, that's a thing people do?... Millenium guys would be like “I'm a bioinformatician.” Okay, Millenium Guy, what did you do? “I ran all the sequencing pipelines to look at transcripts.” “Okay, cool, so did you use Hidden Markov Models or Bayesian analysis to do your clustering?” And they’d be like “Huh?” So we called bioinformaticians something different than industry was. And then the buckets became very clear: people who didn’t have math or statistics backgrounds those were the bioinformaticians, people who had math and statistics backgrounds, who were so important to the organization, those were your computational biologists. What happened … was that you started to lower the salary that you would pay to the bioinformaticians, because those people are commodities; any failed startup company, you could go and find anybody who could do this stuff. Anybody with a statistics background, those people you can’t find anywhere on a corner... so those people you would put a premium on... [Bioinformaticians] are just a tool...you’re a tool, you’re a shovel... go put together a database of this many sequences and make it run on the BLAST-farm while I go and do the tweaking of BLAST.

Although not all biologists saw the divisions quite this starkly, an observable difference certainly exists between the so-called ‘bioinformaticians’ and the so-called ‘computational biologists.’ Although they both spent the vast majority of their time in front of computers, the two groups usually occupied different work-spaces within the institutes and laboratories, ate lunch separately, and went to different parties.[[5]](#footnote-5)

As is suggested in the above quotation, however, the two groups are also performing different sorts of tasks. For the one group, those tasks were often described as ‘informatic’: building and maintaining databases, running BLAST queries on sequences, running data through a pipeline of ready-made software tools, or perhaps building and maintaining a website of biological data. ‘Computational’ tasks, on the other hand, included data analysis with sophisticated statistics, designing and improving algorithms for sequence analysis, and implementing sophisticated mathematical models for reducing biological data.[[6]](#footnote-6) It will be useful to provide one detailed example of each from my fieldwork:

**<space>**

Informatics: H is working at her computer. A biologist from the wet-lab enters her office and explains an experiment he doing on epitope tags[[7]](#footnote-7) in a particular species of bacteria. The biologist needs H to find out what epitopes have been characterized for this organism and provide him with a list. This is difficult task because it involves pulling information from published papers and potentially from databases. After several days of research scouring journals and the web, H puts together a list of online databases containing epitope tags. Being a skilled programmer, rather than retrieve the information by hand she writes a program to pull the information off the web from a variety of databases, put it into a standard format, and generate her own database with all the information the wet-lab biologist needs.

**<space>**

Computational: As part of the work for a paper he is writing, S, a graduate student, needs to know the relative levels at which a particular gene is expressed in different tissues of the body. S organizes a collaboration with a nearby lab who, after some months, provide him with the data he needs, drawn from experiments they have been doing on different cell lines. In the meantime, S develops a theoretical model of gene expression that he codes into an algorithm; it takes several days to run on the computer cluster to which he has access through his lab. When S receives the experimental data he writes a different program first to parse the data into a form he can work with, and then to perform a series of statistical checks: he knows he must be careful to control for differences between the experiments and differences in overall expression levels. S now has some results that he can compare to his own simulated model of gene expression. This comparison will form the basis of a paper published in a well-respected biology journal.

**<space>**

The attitude towards ‘bioinformatics’ expressed in the quotation above suggests that there is a strong hierarchy implied in the distinction between ‘bioinformatics’ and ‘computational biology.’ This attitude – that data production is an unimportant sideshow to the real activity of biology – was held by many (although not all) of my informants. Moreover, the attitude has been internalized by many practicing bioinformaticians, who see their roles as a “service” for biologists. During my fieldwork at the Broad sequencing center, the bioinformatics team understood their role as providing “informatic support” for the core work of the biologists in their labs. Although some resented the fact that biologists would often demand them to dig up data or reorganize a database at a moment’s notice, they saw it as an inevitable part of their day-to-day work in the laboratory. Like the stereotypical IT-support personnel in a large corporation, bioinformaticians saw themselves as under-appreciated, and playing an unacknowledged role in the success of the work being done. Some others became disillusioned with their ‘second class’ status: one of my interviewees had just quit his job at the Broad because he felt that his skills were undervalued. He had found a position at Google where he felt that his expertise in informatics and managing large data sets would be rewarded both financially and in terms of prestige.

This hierarchy is built into the structure of the Broad itself, which is divided into ‘programs’ (such as the Cancer Program, the Psychiatric Disease Program, and the Program in Medical and Population Genetics) and ‘platforms’ (such as Genome Sequencing Platform, the Imaging Platform, and the Biological Samples Platform). Broad personnel generally understand the programs to be performing the real research work of the Institute, while platforms provide support for imaging, cheminformatics, and sequencing. The ‘Computational Biology and Bioinformatics’ program recruits mostly individuals with strong mathematics and statistics backgrounds, while the lesser-status ‘bioinformatic’ work is distributed amongst the supporting platforms.[[8]](#footnote-8)

We should not be too hasty in following practitioners in dividing bioinformatics into ‘higher’ and ‘lower’ forms – both contribute to knowledge production in crucial ways. The attitude that collecting is “less valuable” to an organization than computational biology is contingent upon a particular notion of ‘value’ for biology. The private sector (represented by ‘Millennium Guy’) held a different view of what counted as valuable skill for biological work. For pharmaceutical companies like Millennium, value is generated by ‘high-throughput’: finding potential drug targets by testing thousands of proteins; finding a chemical with drug potential by screening hundreds of chemicals; or predicting potential toxic effects by checking thousands of possible ‘off-target effects.’ Such tasks do not necessarily require a detailed understanding of how a particular biological pathway or system works; since the space of potential compounds and potential targets is immense, speed and volume are crucial.

But this kind of bioinformatics is not only valuable within the private sector: these kinds of practices are also highly regarded *within* specific types of biological work. For instance, in the Broad Sequencing Platform, the aim is to produce as much sequence as quickly and cheaply as possible with the smallest number of errors. Workers who can design better databases, streamline sequencing pipelines, and find ways to process sequence data more quickly are highly sought after. Brendan, a programmer, told me how he designed a Laboratory Information Management (LIM) system – an internal lab database – that greatly improved the sequencing process. The SQUID-BetaLIMS system, which was still used by the Broad during my fieldwork, allowed the lab to track samples, diagnose errors, and track workflow in ways that were crucial to its performance as a state-of-the-art sequencing laboratory. Although Brendan had an undergraduate degree in biochemistry, he had worked as a computer programmer for twenty-five years – he assumed that his “domain knowledge” of biology had a very little to do with his being hired. Rather, it was his software and databases expertise that was exactly what was required for this kind of biological work.

The pronouncement that ‘bioinformatics’ occupies an under-laborer status within biology is designed to suppress the importance of certain new kinds of biological practice. Indeed, these practices are portrayed as ‘unbiological’ – they are, some biologists insist, just informatics or just engineering or just management and therefore distinct from the real business of doing biology. Such workers and the spaces in which they work should be removed, as far as is possible, from view.[[9]](#footnote-9) What is at stake is perceptions about what biological practice *should look like.* When Wally Gilbert claimed in the mid-1980s that sequencing the human genome would be not science but production, he was right. But it is a fact that has made many biologists profoundly uncomfortable; biology is, according to many, making hypotheses, doing experiments, and drawing conclusions about living systems (whether this be using computers or not). It does not building databases, managing information, and re-organizing laboratory workflow. Yet these latter practices are now as much a part of biological practice as the former. What we see in bioinformatics is biology trying to come to terms with these different forms of work and knowledge-making. As attributions of value, the terms ‘bioinformatics’ and ‘computational biology’ are weapons in a battle over what kinds of practices will count as legitimate biology in the years to come.

*3.2 Spaces of work*

In the rest of this chapter, I want to show how these divisions of labor are generated and maintained by the physical spaces in which these new forms of biology take place. The organization of space manages the tension between different forms of knowledge work. Data-driven biology requires not only specific divisions of labor, but also specific kinds of spaces and motions within them. The ‘higher’ (consumption) and ‘lower’ (production) forms of practice are accorded different sorts of spaces and involve different sorts of physical movements. The architecture and space within which the knowledge is produced contributes to its epistemic value. Specific movements and configurations of people, objects, and data act to verify and certify biological knowledge in various ways. This section will draw on fieldwork that took place during 2008 at the Broad Institute in Cambridge, MA.

**[figure 3.1 about here]**

The Broad Institute, endowed by $100 million gifts from Edith and Eli Broad, Harvard, and MIT, lies in the heart of Kendall square. A block from ‘Technology Square’ and standing directly across Main Street from the MIT Biology building and the lopsided ridges of Frank Gehry’s Stata Center, the building at Seven Cambridge Center is the epitome of 21st century laboratory chic. ‘7CC,’ as those who work there know it, is eight stories of shimmering glass and metal straddling almost the entire block (Figure 3.1). 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. 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. 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. For most visitors, scientific and otherwise, this is the Broad Institute. However, no DNA sequencing, the activity on which almost all other activities at the Broad depends, takes places at 7CC. Less than a ten minute walk away lies the building known as ‘320 Charles’ or the Broad Sequencing Center. Although only a few blocks distant, 320 Charles presents a very different scene. On the other side of Binney Street, behind AstraZeneca, BioGen Idec, and Helicos Biosciences, this part of East Cambridge reminds the observer of the abandoned and dilapidated mill towns of western Massachusetts. 320 Charles neighbors include an outstation for an electricity company and a pipe manufacturer. The building itself does shows almost no external signs of it the high-tech science within – only a small plaque near the entrance emblazoned with the Broad insignia would provide a clue to a passer-by. Also a block long, the 320 Charles is low and squat, barely two stories high. Its only external features are large ventilation shafts on the roof, and warehouse-style truck-loading docks at the rear (Figure 3.2). Indeed, the building used to be used by Fenway Park as a warehouse for Red Sox paraphernalia, and, for a time, the sequencers and the famous red and white jerseys must have shared the space before the Broad expanded and took it over in entirety.

**[figure 3.2 about here]**

Inside, the warehouse feeling is still evident. The interior, presumably fairly rapidly adapted to its new purpose, still seems somewhat uncomfortable in its new role. In contrast to the open, easily navigated spaces of 7CC, 320 Charles presents an almost impenetrable rabbit-warren to the newcomer. Windowless rooms, long corridors that seem to lead nowhere, unevenly partitioned spaces, unnecessary doorways, flights of stairs leading to odd mezzanine levels, are a result of the several rearrangements that the lab has undergone as it has had to be adapted to suit various needs. When I first took up my place in the lab, it was several weeks before I summoned the courage to enter by a door other than the one immediately adjacent to my office for fear of becoming embarrassingly lost in the interior. Rubber flooring, uniformly white walls, and metal staircases add to the provisional feel of the place. If 7CC is a consulting office, 320 Charles reminds one of nothing more than a manufacturing plant. Although there is office space, it is pushed towards the edges of the building. The largest, and most central, room in the building is the cavernous space housing the sequencing machines. Large ventilations pipes fitted across the length of the room maintained air quality and temperature at precisely monitored levels. The only noticeable motion comes from an electronic ticker-tape display on the back wall that scrolled the current average length of the sequencing runs in red and green letters.

The two main buildings of the Broad Institute represent different forms of biological practice and serve different scientific ends. It is useful here to draw on the sociology of Erving Goffman. Goffman’s analysis of ‘region behavior’ draws a distinction between the “front region,” in which a social performance is given, and a “back region,” where the performer can “drop his front.”[[10]](#footnote-10) It is in the back region that a good deal of hidden work is performed – work that is necessary for maintaining the performance taking place in the front region.[[11]](#footnote-11) Goffman notes that:

it is apparent that the backstage character of certain places is built into them in a material way, and that relative to adjacent areas these places are inescapably back regions... Employers complete the harmony for us by hiring persons with undesirable visual attributes for back region work, placing persons who ‘make a good impression’ in the front regions.[[12]](#footnote-12)

While I by no means wish to suggest that workers at 320 Charles have “undesirable physical attributes,” there are many ways in which the relationship between the two Broad buildings maps onto the division between Goffman’s front and back regions. First, the location of the two buildings makes clear their relative status. 7CC is a front region, designed for showing off biological work to scientific and non-scientific visitors. As a showpiece of Kendall Square, it’s grand, glassy lobby are designed to create and sell an image of biological work. 320 Charles, on the other hand, is sequestered in an industrial zone. The technical work of this back region sustains the performance of 7CC.

Second, 7CC and 320 Charles differ in terms of access. The former appears open and welcoming to the visitor – anyone can wander into the lobby to peruse the sequencing ‘museum’ or the televisual display. Although access to the upper floors is controlled, scientific visitors are afforded office space and free movement about the informal meeting spaces, kitchens, balcony garden, and library. 320 Charles, on the other hand, appears almost totally closed off – there are few exterior windows, access to any part of the building is by ID card only, doors are monitored by security cameras, and the interior space is closed-off and hard to navigate. Biological laboratory safety designations (BL1, BL2, BL2+) further divide the space into areas where only individuals with certain levels of training are permitted to enter. Third, human traffic between the two laboratories is controlled – few individuals find the need to regularly make the ten-minute walk from one building to the other. In this way, 320 Charles remains remarkably hidden and cut-off from its counterpart.[[13]](#footnote-13)

Robert Tullis, one of the principal 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,” Tullis 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.”[[14]](#footnote-14) 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.

The division between the front region of 7CC and the back region of 320 Charles inscribes a division between both practice and production. The latter building employs many individuals performing repetitive lab-bench work or managing databases. The output is large amounts of biological data. 7CC, on the other hand, houses mostly PhD scientists performing statistical and mathematical analyses, writing software, and performing experiments. Their work is mostly to produce scientific papers based on the analysis of large amounts of data. These are two distinct regimes of biological knowledge, each of which is represented in the physical spaces in which they are performed.

During the design of 7CC, it was realized that the Broad program had expanded such that the number of sequencing machines required would exceed the number that it would be possible to house on site. The solution was 320 Charles – the conversion of the warehouse would provide adequate square footage for the large number of sequencers. DNA sequencing is a highly repetitive activity: the factory-like layout of 320 Charles is appropriate to the performance of such labor. Indeed, the appearance and design of the building reassures the visitor that it is the sort of place suited to outputting large volumes of identically-replicated products (like the Red Sox jerseys that were originally stored here). Access is limited to those few who are needed to run and manage the day-and-night output of data. The central place given to the sequencing machines symbolizes the fact that 320 Charles would ideally run as an automaton. Careful attention is paid to the specific layout of lab-benches and essential equipment so that the lab can run, if not as a machine, then as Taylorist production-line in which every worker executes precise and repetitive motions. Figure 3.3 shows a work-flow diagram produced by a student at the MIT’s Sloan School of Business in order to streamline productivity. The ultimate measure of productivity is the error rate. This is carefully monitored by a team of quality control engineers. The appearance and the configuration of the space, and the motion within it, certifies the knowledge that is produced.

**[figure 3.3 about here]**

The division between the front region of 7CC and the back region of 320 Charles largely matches the division between ‘higher’ and ‘lower’ bioinformatics described above. The hierarchy between biological practices based on mathematics, statistics, and bench work, and biological practices based on informatics and databases is represented in the physical spaces of the laboratories in which they are performed. The notion that the former practices constitute the ‘real’ work of biology and that production sequencing is mere technical support is reinforced by their front/back configuration. Selling biology to its funders, as Lander is well aware, means projecting an image (putting on a performance, if we are to follow Goffman’s terminology) of a biology that is not only open and transparent, but also a biology that presents immediate and tangible benefits for human health. In September 2008, the Broad Institute received an additional (and record-breaking) $400 million gift from Eli and Edythe Broad: “Of all our philanthropy,” Eli Broad announced, “the Broad Institute has been the investment that has yielded the greatest returns... We are convinced that the genomics and biomedical work being conducted here... will ultimately lead to the cure and even the prevention of diseases.”[[15]](#footnote-15) 7CC is highly conformable to the image of a place in which such medical advancements will be made; it's configuration not as a traditional laboratory but as a mixed lab-office space associates it with medical practice.

Thomas Gieryn has commented that “built places materialize identities for the people, organizations, and practices they house. Through their very existence, outward appearances, and internal arrangements of space, research buildings give meanings to science, scientists, disciplines, and universities.”[[16]](#footnote-16) This is an apt description of the two buildings comprising the Broad Institute. The segregation and spatial organization of different kinds of practices promotes particular vision of biological work. First, it renders some practices largely invisible, deeming them beyond the boundaries of ‘real’ biological practice. Second, the architecture denotes a vision of collaboration between biologists and computer scientists: by mixing their spaces into one another, the observer is given the impression that the two different sorts of practices are collaborating harmoniously. Following Gieryn once again, people are designed along with the walls and windows – the space gives people identities are particular kinds of workers and particular kinds of knowledge-producers. The spaces of 320 Charles identify those that work there as technicians, data-workers, and shop-floor managers; the design of 7CC suggests academic science, knowledge-work, and even medicine.

Lander’s notion of the “transparency of discovery” means a constant sharing and motion between wet and dry spaces. “Our research only works in the spirit of collaboration,” Alan Fein (a deputy director of the Broad) comments, “Everybody needs to be accessible, transparent, visible.”[[17]](#footnote-17) If everyone can see what everyone else is doing, they will be able to understand each other's work and collaborate most effectively, the designers hope. This is a vision of a new, integrated, hyper-productive biology. It is also a highly sell-able biology. But the division into higher and lower, front and back, obscures the full range of identities and practices that this knowledge relies upon; as we shall see in the next section, bioinformatics depends just as much on the productions of 320 Charles as it does on the work of 7CC.

For the Broad, the strong emphasis on medical applications requires a fluid and immediate interaction between biologists and computer scientists. For instance, a typical problem might be to determine the location of all the places in the human genome that are strongly linked to development of Crohn’s disease.[[18]](#footnote-18) Such work involves gathering samples from a large number of patients (usually at an affiliated hospital), processing the samples, performing the genotyping for each sample, then developing computational and statistical methods to analyze the entire set of genotypes, and finally examining the results of this analysis to see whether they provide any clues about the biological pathways through which the disease acts. Close and rapid communication between those collecting the samples (medical doctors), those preparing them (wet lab biologists) and those performing the analysis (computer scientists) speeds the process. 7CC presents a vision of integrated, medically-oriented, scientifically rigorous and productive biology. The work done here is perceived as valuable by those inside and outside the Broad because of its applicability to medicine. This applicability is achieved at least in part by creating a space that is designed and operated to produce this value.

*3.3* *The lab that changed the world*

But this is not the whole story. Let’s now examine the back spaces of the Broad in order to describe what goes on there in more detail. What we discover is that new modes of valuable work are emerging in this sort of activity. Here, principles of management and efficient manufacturing have been rigorously applied to the making of biological data. This is certainly production, but it is not *mere* production: it requires careful attention to the organization of space, people, and technology. Ultimately, these spaces are structured and monitored by computers: samples and people are represented as data too.

When I visited the Broad Sequencing Center at 320 Charles, one of the first people I spoke to was Meredith, the manager of the Molecular Biology Production Group. Her office sat on a floor above and overlooking the floor of the sequencing lab and as I traveled along the hallway I peered down on the workers busy at their lab benches. The first thing I noticed 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 asked the obvious question: why all the books? “It's required reading for my employees,” she told me, “every new person on my team gets a copy.” Perhaps surprisingly, this isn’t a book about molecular biology, or about any natural science, but about assembly lines.

The tagline 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’ (in contrast to Henry Ford’s ‘mass 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.[[19]](#footnote-19)

The expense of equipment and intolerance of disruption tends mass-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.[[20]](#footnote-20) Workers were encouraged to identify and rectify the cause of the problem. Lean production depends not only on teamwork but also 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.[[21]](#footnote-21) 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), which saves on storage space, and close relationships with suppliers, which allows 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.”[[22]](#footnote-22) 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.” Rob 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.[[23]](#footnote-23)

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.[[24]](#footnote-24) 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 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.”[[25]](#footnote-25) This was achieved by breaking the sequencing process down into small, repetitive steps that could be performed quickly and accurately.

In 2003, Julia Chang was given the task of analyzing and improving the process of ‘picking’ *E. coli* colonies for sequencing.[[26]](#footnote-26) 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.”[[27]](#footnote-27) In other words, Chang collected data that had not been considered relevant or interesting to the Broad’s technicians and mobilized it to formulate 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).[[28]](#footnote-28) 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.[[29]](#footnote-29)

These changes were based on several principles of lean production, including 5S, pull production, and *kanban*. 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.”[[30]](#footnote-30) *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 equipments lists made. Figure 3.4 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.[[31]](#footnote-31) Likewise figure 3.5 show 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 figures, 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.”[[32]](#footnote-32) 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.[[33]](#footnote-33)

**[figure 3.4 about here]**

**[figure 3.5 about here]**

The requirements of sequencing operations also demanded new ways of organizing people. 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.” 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-white and immigrants.[[34]](#footnote-34) 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.[[35]](#footnote-35)

Making finishing more efficient meant re-thinking incentive 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.[[36]](#footnote-36)

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 3.6[[37]](#footnote-37) 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.”[[38]](#footnote-38) Triage teaming provided a way to increase productivity while still recognizing and exploiting the special skills of individuals.

**[figure 3.6 about here]**

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 logistics, supply chain and quality control for many of the materials coming into the lab. 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 processes to scale up the processes for the new sequencing machines from the bench to the mass-production scales.

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 scientists 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.

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 designed 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?” 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, a whole team could adopt their techniques. 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.” 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’ ... 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.

The detail with which such tracking is performed is illustrated in the kind of checklists that Vokoun proposed for the MBPG. Figure 3.7 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.[[39]](#footnote-39)

**[figure 3.7 about here]**

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.” 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 basis, and projects ended when they were removed from the database. 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 is 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. Consonant with the spirit of lean management, the central role of the computer does not mean that workers have lost all agency in the sequencing process; indeed, individual workers have a high level for monitoring and improving their own work. However, it is the computer that draws all these individual actions together into a collective whole – it is the computerized monitoring and management that integrates a person’s work into that of the team. In other words, it is through the computer that individual labor becomes ‘productive.’

What are the consequences of this borrowing from business? What difference does it make that the Broad is organized more like Toyota than The Pasteur Institute? First, it has changed the notion of valuable work in biology. The work accorded value in the Broad Sequencing Center is not the highly individualistic, highly innovative work of traditional bench scientist. What is valuable is instead teamwork, attention to detail, precision, and efficiency. Second, work at the Broad is based on a new accounting of biological work; the lab is funded not according to how many papers it publishes or how promising its research seems, but 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.[[40]](#footnote-40) 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.

Third, what constitutes ‘progress’ in advancing biological knowledge has changed: progress can be understood as the accumulation of more and more sequencing data at an ever-decreasing cost. The immediate goals of day-to-day work are discoveries that will increase output by decreasing variability rather than make fundamental breakthroughs or shifting work in a qualitatively new direction. Fourth, the culture of the Broad Sequencing Center suggests a shift in what sorts of people are doing biological work and changes in the distribution of labor. Whereas before 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 not trained 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. Finally, this new kind of biological lab has become a space of surveillance to an extent previously unusual in the sciences. Through information, 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 provided an appealing sense of control. Doing good and interesting work means keeping watch and being watched.

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) rigorously checked. It was the computer – as a data management machine – 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. At the Broad, the organization of biology in accord with ‘business principles’ has deeply shaped practices of knowledge-making. Computers have created new ways of making authorized and valuable knowledge through careful accounting and management of data.

*3.4 Conclusions*

Wet lab biologists, computational biologists, and system administrators use and relate to the spaces around them differently. These different workers are acutely aware of these differences – the kind and status that is attributed to biological knowledge is dependent on the spaces from which it emerges. The highest status is accorded to knowledge produced in the highly visible front spaces and wet labs, while lower status accrues to knowledge associated with the sorts of technical production that takes place in the back spaces of the Broad.[[41]](#footnote-41) Moreover, the layout and design of the spaces recapitulate and reiterate the kinds of knowledge that are produced within them: back spaces promise reliable practices of replication and production while front spaces adhere to more stereotypical images of scientific practice. And the motion of people and data through and between spaces plays a crucial role in validating and certifying knowledge.

But the biological work of the Broad consists just as much of what is going on in the back spaces. A thick description of this work reveals the emergence of new forms of practice and with new forms of value in this work. This is a result of computers. They have transformed ‘production’ from a ‘bad’ to a ‘good.’ If bioinformatics can be characterized by mass and speed – by the asking of questions involving large numbers sequences, genes, organisms, species, and so on – the computer is the ideal tool for asking and answering such questions. They came into biology as machines designed to work with volume, to speed up, to make efficient.

The history of the computer suggests the reasons for this: from the 1950s, computational practices were linked to and evolved from efforts to rationalize government bureaucracy and commerce. For instance, the Treasury of the United Kingdom used punch-card-based computers for accounting, payroll, and the production of statistics.[[42]](#footnote-42) And it has been argued 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.”[[43]](#footnote-43) 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.[[44]](#footnote-44) 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.[[45]](#footnote-45) 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. 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.”[[46]](#footnote-46) In 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.

This history has conditioned the role that computers have come to play in biology. That is, as a tool for speed and efficiency. 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. Computers are tools of business, and demand and enforce the kinds of practices that have transformed biological work over the last two decades, re-orienting it towards speed, volume, productivity, accounting, and efficiency. However, this chapter has shown how it is not only the samples and sequences, but also the laboratory itself and its workers are managed and accounted as data. The computer brings changes that are at once social, technical, spatial, and epistemic. Bioinformatics entails reorganizations of space and people in tandem with reorganizations of practice and knowledge.

These findings obviously rest heavily on observations of the Broad Institute. But similar patterns can be discerned elsewhere. The Sanger Institute (Hinxton, UK), the J. Craig Venter Institute (Rockville, MD), the Genome Institute at Washington University (St. Louis, MO) the Joint Genome Institute (Walnut Creek, CA), and the Beijing Genomics Institute (Shenzhen, China) all have large-scale production sequencing facilities.[[47]](#footnote-47) The aim of each is to produce high volume and high quality product at low cost.

Rather than a biology oriented around individual investigators and intra-disciplinary work, bioinformatics is a biology performed by large, multi-disciplinary teams, oriented towards efficiency rather then reproducibility, measured by accounting and QC rather than peer review, and ordered by information systems. Toyota’s concept of ‘lean production,’ as deployed at the Broad Institute, suggests a label for the ways in which the making of ‘biovalue’ depends on new forms of practice as well as new regimes of circulation. A ‘lean biology’ – a biology stripped back to its essentials, reduced to its elements (in particular, data and sequence), made efficient – is the kind of biology required to make bioinformatic knowledge. ‘Lean biology’ is a mode of practice that is designed to harness this productivity – to maximize the productivity by ignoring or stripping or abstracting away all functions extraneous to the generation of a product.[[48]](#footnote-48) In doing so, it creates new kinds of biological knowledge that are rooted in the practices of industrial management and in informatics. Lean biology – as the sets of practices that produce bioinformatic knowledge – is a way of not only describing this trend, but also suggesting how it is coupled to the commoditization of life in the age of biotechnology and biocapital.[[49]](#footnote-49)

1. One way to appreciate the special value of sequence data is to examine the Bermuda Rules. In 1996 and 1997, a set of meetings amongst the genomics community established rules for how and when data was to be shared and who was entitled to use it. These rules recognized the unique epistemic status of data (as something less than knowledge) and attempted to prevent the direct exchange of data for financial value. [↑](#footnote-ref-1)
2. Davis, “Sequencing the human genome,” 121. Davis compared the HGP to Nixon’s ill-fated ‘war on cancer’ of the 1970s. [↑](#footnote-ref-2)
3. Davis, “Sequencing the human genome,” 21. [↑](#footnote-ref-3)
4. Shapin, “House of experiment.” [↑](#footnote-ref-4)
5. At the start of my fieldwork it was usually my practice to ask informants and interviewees who else of their friends, colleagues and professional acquaintances I should talk to. However, the separation of the two groups made it such that, having begun talking to computational biologists it was difficult to gain an introduction to a bioinformatician. [↑](#footnote-ref-5)
6. One bioinformatics blog, written by a graduate student, half-jokingly divided bioinformatics into six different ‘career paths’: linux virtuoso (“the LV performs all their research at the command line: vi edited bash scripts chained together using shell pipes”), early adopter (“always working on the latest area of research, system biology synthetic biology, personal genomics”), old school (“blinkered to change in tools and technology, the Old School is doing their analysis in Fortran on a Windows 95 Pentium II”), data miner (“their everyday tools are mixed effect regression, hidden Markov models, and the fearsome neural gas algorithm”), perfect coder (“produces code like poetry and, after a five second glance, even your dog knows what the script does”), wet lab bioinformatician (“while others have their heads in the clouds thinking about theories and algorithms, the WB is getting his hands dirty with real data as it is being produced”). In analogy with a fantasy role playing game (called World of Bioinformatics Quest), these careers have different attributes for coding, statistics, presentation/writing, research focus, and collaboration. “Getting the right attributes is therefore critical, and playing to your strength will result in more Papers(TM) and Grants(TM).” Barton, “World of Bioinformatics Quest.” This provides a detailed, if amusing, description of the various sorts of practices in bioinformatics and how biologists perceive the differences between them. [↑](#footnote-ref-6)
7. An epitope tag is a region of a folded protein that is recognized by an antibody. By tagging a protein with an epitope, allow it to be recognized by antibodies, which allow the construction of a simple test for the presence of a protein. [↑](#footnote-ref-7)
8. Broad Institute, “Who is Broad?” [↑](#footnote-ref-8)
9. This invites comparison with Shapin’s arguments about seventeenth century laboratory spaces in which certain technical practices were hidden from view (Shapin, “Invisible technician”). [↑](#footnote-ref-9)
10. Goffman, *Presentation of self*, chapter 3. [↑](#footnote-ref-10)
11. Goffman gives the example of a watch-repair shop in which the watch is taken into a back room to be worked on: “it is presented to [the customer] in good working order, an order that incidentally conceals the amount and kind of work that had to be done, the number of mistakes that were first made before getting it fixed, and other details the client would have to know before being able to judge the reasonableness of the fee that is asked of him.” Goffman, *Presentation of self*, 117. [↑](#footnote-ref-11)
12. Goffman, *Presentation of self*, 125. He adds that the back region is often reserved for ‘technical’ standards while the front for ‘expressive’ ones (126). [↑](#footnote-ref-12)
13. For several months while I was working in the MIT biology department, and regularly interviewing scientists across the road at 7CC, I remained completely unaware of the existence of 320 Charles. [↑](#footnote-ref-13)
14. For more on the design of the Broad Institute, and especially its ‘transparency’ see Higginbotham, “Collaborative venture,” Higginbotham, “Biomedical facility,” Silverberg, “The glass lab.” [↑](#footnote-ref-14)
15. Broad Institute, “Philanthropists.” [↑](#footnote-ref-15)
16. Gieryn, “Two faces,” 424. [↑](#footnote-ref-16)
17. Silverberg, “The glass lab.” [↑](#footnote-ref-17)
18. Crohn’s disease a genetic disorder affecting the intestinal system. On the Broad’s work on Crohn’s disease see: Rioux et al., “Genome-wide association study.” [↑](#footnote-ref-18)
19. Womack et al., *Machine*, 12-13. [↑](#footnote-ref-19)
20. Womack et al., *Machine*, 57. [↑](#footnote-ref-20)
21. Womack et al., *Machine*, 129. [↑](#footnote-ref-21)
22. Womack et al., *Machine*, 152. [↑](#footnote-ref-22)
23. Broad Institute, “Rob Nicol.” [↑](#footnote-ref-23)
24. 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 Stamatis, *Six sigma fundamentals*. [↑](#footnote-ref-24)
25. Vokoun, “Operations capability improvement,” 6. [↑](#footnote-ref-25)
26. Chang, “Control and optimization.” [↑](#footnote-ref-26)
27. Chang, “Control and optimization,” 65. [↑](#footnote-ref-27)
28. Vokoun, “Operations capability improvement.” [↑](#footnote-ref-28)
29. Vokoun, “Operations capability improvement,” 51. [↑](#footnote-ref-29)
30. Vokoun, “Operations capability improvement,” 49-50. [↑](#footnote-ref-30)
31. Vokoun, “Operations capability improvement,” 55-56. [↑](#footnote-ref-31)
32. Vokoun, “Operations capability improvement,” 105. [↑](#footnote-ref-32)
33. Several other Sloan students also applied lean principles and other management techniques to aspects of the Broad: Scott Rosenberg analyzed computer finishing process (Rosenberg, “Managing a data analysis production line”); Kazunori Maruyama studied the electrophoretic sequencing process itself (Maruyama, “Genome sequencing technology”). [↑](#footnote-ref-33)
34. 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. [↑](#footnote-ref-34)
35. Rosenberg, “Managing a data analysis production line,” 75. [↑](#footnote-ref-35)
36. Rosenberg, “Managing a data analysis production line,” 82-83. [↑](#footnote-ref-36)
37. Rosenberg, “Managing a data analysis production line,” 59-71. [↑](#footnote-ref-37)
38. Rosenberg, “Managing a data analysis production line,” 78. [↑](#footnote-ref-38)
39. Vokoun, “Operations capability improvement,” 69-70. [↑](#footnote-ref-39)
40. National Human Genome Research Institute, “Genome Sequencing Centers (U54).” [↑](#footnote-ref-40)
41. All economies benefit some and disadvantage others: they are sets of power relations. The production and consumption of bioinformatic knowledge accumulates prestige to academically trained biologists, while restricting credit for data producers. This has the important consequence that bioinformatic work continues to be described and understood as fundamentally ‘biological’ – that is, it causes us to underestimate and undervalue the contribution of managers, physicists, engineers, technicians, mathematicians, and computer scientists to this new kind of biology. This has significance for those who might point to the need for greater value and emphasis to be placed on new skills of data management and data curation in biology: the argument here suggests that there is reason why such pursuits remain undervalued. Namely, the status of biological knowledge as biological knowledge partly depends on the hierarchy between producers and consumers. [↑](#footnote-ref-41)
42. Agar, *Government machine*. On the Treasury see chapter 8. [↑](#footnote-ref-42)
43. Campbell-Kelly and Aspray, *Computer*, 105. [↑](#footnote-ref-43)
44. Haigh, “Chromium-plated tabulator.” [↑](#footnote-ref-44)
45. Cerruzi, *History*, 32-33. [↑](#footnote-ref-45)
46. Cortada, *Information technology*, 160. On the history of computers in business see also Edwards, “Making history.” [↑](#footnote-ref-46)
47. The Beijing Genomics Institute is an especially interesting example since it, unlike Broad, places its sequencing activities front and center. Rather than suggesting a fundamentally different model, though, the BGI may be revealing of the global dynamics of science: China and the BGI itself may be said to constitute a ‘back space’ that produces materials to be consumed in the ‘front spaces’ of the US and European biomedical institutions. The world’s largest manufacturer is also emerging as the major manufacturer of sequence. [↑](#footnote-ref-47)
48. This kind of view is also prevalent in synthetic biology where attempts have been made to strip organisms down to the bare essentials necessary for life (the so-called ‘minimum genome project’). See Glass et al., “Essential genes.” [↑](#footnote-ref-48)
49. Edward Yoxen’s notion of ‘life as a productive force’ (Yoxen, “Life as a productive force”), Marilyn Strathern’s view of ‘nature, enterprised-up’ (Strathern, *After* nature), Catherine Waldby’s ‘biovalue’ (Waldby, *Visible human project*) Sunder Rajan’s ‘biocapital’ (Rajan, *Biocapital*) and Nicolas Rose’s ‘bioeconomics’ (Rose, *Politics of life itself*) all seek to capture the essence of what is going on between biology, biotechnology, medicine, politics, and the economy. For more on the relationship between bio-capital and the production of sequence see Stevens, “On the means of bio-production.” [↑](#footnote-ref-49)