Foundations and profiles: splicing metaphors in genetic databases and biobanks

Ratto, Matt

Public Understand. Sci. **15** (2006) 31–53

Foundations and profiles: splicing metaphors in genetic databases and biobanks

Matt Ratto

In this paper we explore new developments in genomics, in particular the move from data sequencing efforts like the Human Genome Project (HGP), to newer forms of data-driven genomic work that focus explicitly on the complicated relationship between genes and environment. We compare the use of a key term within the HGP, the metaphor of "foundation," to the use of a different term, the metaphor of "profile," within the GenomEUtwin consortium, an exemplar of "post-genomic" projects. By doing so, we attempt to re-think the role of language and metaphor in scientific projects and explore new developments in post-genomic research. These developments include: first, the movement towards an explicit and programmatic acknowledgement of the complexity of gene and trait relationships, second, the use of bio-informatics techniques as exploratory tools of discovery rather than as part of a more straightforward "decoding" effort, third, the development of network infrastructures that link up and provide access to a vast array of different databases, and fourth, the alignment between various disciplines and interests within biology, clinical work, and public health initiatives. We use the metaphor of "splicing" to emphasize the heterogeneous work of scientists engaged in "weaving together" the diverse set of ideas, interests, and players necessary for the success of large-scale scientific projects.

1. Introduction

Many of the new "post-genomic" projects can be seen as a response to criticisms of the Human Genome Project, criticisms based on public disappointment as to its ultimate outcomes and dissatisfaction at the lack of medical and health benefits promised. One key difference appears to be an emphasis on the interactions of gene and environment, marked by new research projects and protocols aimed at exploring the complex relationship between genes, the environment, and physiological traits. In addition, the development of novel, networked databases replete with data consisting of digitized biological information from multiple sources can be understood as part of the "informational turn" in biology (Beaulieu, 2004; Lenoir, 1999; Wouters et al., 2002).

In this paper, we explore the shift in the biological sciences from genomics to "post-genomics" and the creation of new large-scale biological projects. One question that needs

 ISSN 0963-6625 DOI: 10.1177/0963662506056786

to be addressed is why these newer projects are being funded and developed, despite the criticisms noted above. We take the position that scientific knowledge and direct health benefits are not the only products of biological research. Instead, the construction of technological infrastructure, the development of training and the creation of expertise, as well as other types of value are part of what makes these projects successful. Equally, these projects provide the possibility for strategic alliances between scientists, policymakers, technology providers, public health experts, and many others. Understanding how these diverse players link up is thus of paramount importance for untangling the complex interests involved and the promises that are made. We argue that one of the tools by which these constellations negotiate perspectives, needs, and interests is through the ad-hoc use of conventionalized metaphoric expressions (Lakoff and Johnson, 1980) that provide an important resource for the managing of differences necessary for the successful construction and maintenance of large-scale scientific projects.

In this paper, we provide an overview of the development of two large-scale scientific projects in biology, the Human Genome Project (HGP) and the GenomEUtwin project. We detail some of the tensions around the HGP by examining scientific papers, project reports, science and technology studies (STS) scholarship, and journalism, in order to better characterize the maintenance of this project. We focus on the use of the term "foundation" by members of the HGP consortium to describe and evaluate the past, present, and future of the HGP. We then use our understanding of the tensions around the HGP to contextualize the development of GenomEUtwin, a project we see as being based on both the successes and failures of the HGP. Here, we rely on scientific papers, reports, and, importantly, interviews, to detail new directions in genomic research and better understand how these projects are constructed and maintained.

2. Metaphors in scientific communication

The call for the joint meeting of the Society for the Social Study of Science (4S) and the European Association for the Study of Science and Technology (EASST) in 2004 focused on the issue of public proofs and the problem of trust between scientists and publics. One element of this relationship has been called "public understanding of science," a series of interdisciplinary approaches which have, in the past, tended to emphasize scientific literacy and the problem of misapprehension of scientific ideas by publics.

A key aspect of the initial version of this model suggested that the problem of trusting scientists is due to an ignorant public—educate the public (i.e. improve widespread scientific literacy) and you create a more trusting and enthusiastic public relationship to science (Bodmer, 1985). As a recent article in the journal *Public Understanding of Science* commented: "This would, of course, be a rather happy outcome for the scientific research community" (Sturgis and Allum, 2004: 55). Such a comment also points to a dissatisfaction with such models.[1]

Although a deficit model of public understanding of science is an often critiqued theory (Irwin, 1996; Sturgis and Allum, 2004), it does have a way of creeping back in, particularly in criticisms of scientific language and "jargon." Such arguments say that the problem of public understanding is partly due to the reliance on expert speech genres within scientific domains (e.g. Leggett and Finlay, 2001). Publics can't understand or worse, misapprehend, scientific speech and come to faulty conclusions. Again, while a fair amount of recent work in STS has pointed to the more deliberate work of "expectation building" within scientific projects and its relation to grantsmanship and other overtly political projects, a linguistic

version of the deficit model often re-emerges. One linguistic form that often comes under fire is the metaphor (see below but also Hubbard and Wald, 1993; Lippman, 1992; Nelkin and Lindee, 1995).

For example, in a perspective article published in the 4 July 2003 issue of the journal *Science*, Arizona State University biologists and historians of science Matthew Chew and Manfred Laubichler examine what they see as a particularly rampant use of metaphors within ecological science. The article, entitled "Natural Enemies—Metaphor or Misconception?" addresses the use of war and conflict metaphors to describe the ecological ramifications of introduced species. The authors center their argument around the use of the term "natural enemies" and the way it carries with it "unscientific" connotations. In an interview published in their local university newsletter, they said: "Scientists claim that they are continually misunderstood, but we should examine how much they contribute to potential misunderstanding. Using this as an example, it seems disingenuous for them to complain" (Hathaway, 2003).

The article claims a number of problematic features of metaphors:

> Simplicity and intuitive appeal are . . . the main reasons why scientific language has never succeeded in "cleansing" itself from metaphorical "impurities" . . . Metaphors introduce a fundamental trade off between the generation of novel insights and the possibility of dangerous or even deadly misappropriation. (Chew and Laubichler, 2003: 52)

The quotations are demonstrative of two contradictory thoughts about metaphors, where the fear of misappropriation is balanced against the productive quality of novel thinking. In one way, then, we might equate these authors' feelings about metaphors to the fears around introduced species that they present in their article. In this light, scientific metaphors are productive within the tightly controlled boundaries of their natural environments—i.e. scientific work and discussion—but dangerous and destructive if they escape.

However, this is not the only view of scientific metaphors.[2] An alternative view focuses on the productive nature of metaphors. In the section below, we connect two slightly different positions that both analyze metaphors as productive. These are, namely, work on metaphors in scientific communication and scholarship on metaphors as part of social organization.

Three uses of metaphors have been most obvious within recent scholarship. First, metaphor analysis has been used to explore the work of scientists, typically by relating the development and transference of scientific theories and concepts, either within science (e.g. Black, 1993; Bono, 1990; Brown, 2003; Hesse, 1966; Leatherdale, 1974) or as used to communicate scientific concepts to a lay public (e.g. Bucchi, 1998; van Dijck, 1998). A second stream of work has focused on metaphors as carriers of knowledge or information at the level of communicative activity. Within this stream, metaphors have been used to explore the framing of public debates (e.g. Putnam et al., 1996; Schultze and Orlikowski, 2001; Stutman and Putname, 1994). Finally, metaphors have been analyzed as the ways users (whether as institutions, groups or individuals) are directed towards particular possibilities for action with technologies (e.g. Van Lente, 2000; Wyatt, 2000).

Thus, scholars have explored how metaphors provide ways to translate between various "frames" of experience (Black, 1993; e.g. Bono, 1990), make unfamiliar situations familiar (Schultze and Orlikowski, 2001), or construct a "common ground" within diverse communities (Maasen and Weingart, 2000). What unites all these perspectives is a focus on how metaphors work to maintain coherent understandings within disparate communities. Metaphors are robust enough to maintain certain characteristics, and flexible enough to allow for

several interpretations and various uses of the metaphors (Hellsten, 2002). In this sense, metaphors can be understood as "boundary objects" (Hellsten, 2002; Star and Griesemer, 1989) which serve to unite diverse interests and manage the work required to continue productive joint activity.

These perspectives form a strong grounding for our understanding of the usefulness of metaphors in scientific work. In order to better understand how large-scale scientific projects are developed and maintained, we emphasize the use of metaphoric expressions that provide connections across "contexts of knowledge" (Maasen and Weingart, 2000). Rather than focusing on metaphors at the conceptual or cultural level, as do many who have studied metaphor—i.e. root metaphors (Pepper, 1942), generative metaphors (Schoen, 1979), conceptual metaphors (Lakoff and Johnson, 1980)—we focus instead on more bounded metaphoric expressions (e.g. Zinken et al., 2005). Our goal has been to examine the use of specific key expressions in detail and analyze the variety of connections made within the discursive work of the projects we have studied. This work falls into a tradition of key word analysis (e.g. Callon et al., 1986; Maasen and Weingart, 2000). Equally, such analyses of key metaphoric terms in multiple contexts have also been mapped by Hellsten and Leydesdorff (2004). While their analysis focused on a quantitative "network" mapping of the proliferation and exchange of terms within and between politicians, journalists, and scientists, our focus is a more qualitative examination of how specific terms are used to further the goals of the projects we study. Our goal is to examine how these terms are used to connect together and communicate scientific insights, but also to legitimate organizational and technological decisions. In this sense, our focus in this project has been on words used in similar ways to what Miettenen (2002) calls "transdiscursive terms." Drawing upon the work of Foucault, Miettenen defined these as words that carry out important organizing functions for large-scale and diffuse organizations. Such functions include connecting social groups and institutions in shared discussions and acting as "epistemic organizers" (Elzinga, 2004: 114) by connecting past and present accounts. But whereas Miettinen explores how transdiscursive terms are intentionally generated by policy interests in order to unite diverse groups under a single "umbrella" concept, we focus on terms that are used in a more "ad-hoc" way to provide linkages between interests and ideas. These terms, which we call *splicing metaphors*, do have resonance across social and disciplinary contexts and provide references to divergent ideas and technologies. However, their task is less about "maintaining coherence across intersecting social worlds" (Star and Griesemer, 1989: 393), and more about providing a common point for connection. Rather than sitting at the heart of struggles over meaning, the diversity of meanings and associations connected to these terms are accepted and not necessarily debated. Thus, as opposed to "immutable mobiles" (Latour, 1987), whose power came from their ability to cross epistemic boundaries and maintain consistent definition, splicing metaphors, like boundary objects[3] more generally, are useful because they circulate, not because of stable meanings (Wouters et al., forthcoming).

3. Splicing metaphors

Splicing metaphors are terms used by the agents involved in the construction and maintenance of large-scale scientific projects. Rather than helping to create a stable consensus among disparate interests, the usefulness of these terms lies in their ability to help manage difference. The key element here is the issue of stability and the reflexiveness of the agents involved. While shared coherent meanings are necessary for the management of

large-scale scientific projects, we also believe that terms that provide mappings to a diversity of meanings are also important.

These terms are powerful organizers within scientific projects precisely because they do not rely on a shared consensus built on the fiction of linguistic stability or a singular relationship to an abstract level of conceptual or cultural meaning. We call these *splicing metaphors* in order to emphasize their deliberate use to "tie together" disparate players, ideas, and goals. Here, we use the metaphor of splicing to emphasize a number of important features. First, the term splicing originates from the same root word as the verb "split" and refers to a process of joining together that begins with "dividing or splitting the ends [of ropes or cables] into separate strands" (Webster's Revised Unabridged Dictionary, 1998 [1996]). These strands are then interwoven in such a way as to create a join between one or more pieces. The metaphor of splicing thus emphasizes the interfiliated nature of large-scale scientific projects, made up of various scientific, public, and private interests. The way "threads" from the different domains are woven together in a close, but not inseparable bond reminds us of the work that goes into creating splices, and also the need to maintain them. Without ongoing maintenance, the splices that help manage scientific projects can begin to unravel.

In addition, the term splicing has been used in the domain of genetics to refer to two different phenomena: first, as a description of the process whereby scientists connect together strands of DNA from two different organisms to create a wholly new third organism (commonly known as gene splicing); second, as a description of the process whereby introns are removed from mRNA precursors, and the remaining exons are combined together to form functional mRNA (commonly known as RNA splicing). This latter definition also serves as part of the description of a phenomenon known as "alternative splicing" that attempts to explain the multitudes of protein sequences that can be the result of a single DNA sequence.

We see the metaphor of splicing as productive for describing the use of specific key metaphoric terms within large-scale scientific projects. The most important purpose of splicing metaphors is to help legitimate scientific projects by "weaving together" ontological claims and specific scientific work. In addition, some splicing terms can be used to create a contiguous historical "strand"; by providing a temporal order for the project, these terms both characterize the past and predict future directions as necessary. Finally, the partial ambiguity of the metaphoric terms allows them to be used to link together various aspects of projects, providing the possibility for various "alternatives" (e.g. alternative meanings to be associated with the same term). Thus, splicing emphasizes the interwoven, temporal, and alternative possibilities that are indicative of the dynamic and heterogeneous nature of scientific and social work.

4. The Human Genome Project and "new biology"

The year 2003 marked 50 years since Watson and Crick first discovered the double-helix form of DNA. In addition, 2003 also marked the publicized completion of the human genome sequence, released by the Human Genome Project in April 2003.[4]

For many, these developments, and the new theories and techniques that surround them, are understood as having transformed the life sciences. The 2003 US National Academy of Sciences annual symposium was no exception. Entitled "The New Biology: Celebrating the Past, Imagining the Future," it provides a good overview of the changes that are seen as having taken place.

> Now the recognition of DNA as the fundamental determinant of life and the discovery of its molecular structure has certainly transformed biology from a largely descriptive and phenomenological science to one with an over-riding chemical and molecular perspective. Indeed, the biological principles established over one hundred years of research, can now be restated in molecular terms. (Berg, 2003)

Here, Berg, the Cahill Professor Emeritus in Biochemistry at the Stanford University Medical Center, describes the overall change as one from biology as a "largely descriptive" and experiential science to one based in theoretical and conceptual work. This is a theme that has been taken up by many advocates and critics of the "new biology," who associate with these developments changes in scientific practice and knowledge, as well as large-scale medical and public health ramifications. Key to these changes is the characterization of genes and DNA as the "fundamental determinant of life."

The main exemplar of the "new biology" is the Human Genome Project (HGP). Founded in 1991 as a consortium of various institutions in the United States, Europe, and Japan engaged in sequencing efforts, the HGP sought to "decode" a complete human genome sequence. This goal, sponsored by governments as well as private non-profit trusts, was often defended by its advocates as resulting in new visions of public health, new medical therapies "tailored" for individual genotypes, as well as other kinds of genetically based medical interventions. Upon the completion of a first draft of the human genome sequence, Bill Clinton stated that genomic science would ". . . revolutionize the diagnosis, prevention and treatment of most, if not all, human diseases" (Clinton et al., 2000). This was followed by a statement from Tony Blair that mirrored the language of revolution, ". . . what we are witnessing today—a revolution in medical science whose implications far surpass even the discovery of antibiotics . . ." (Clinton et al., 2000).

These quotations are indicative of the grand visions that developed around the HGP. In addition, scientists promised new medicines specifically tailored for an individual's genetic makeup, genetic engineering techniques that could extend and better life, as well as great advances in our knowledge of how biological organization and evolution take place.

Such grand claims were balanced by criticisms from vocal scientists such as Richard Lewontin. At the start of the HGP in 1990, Lewontin argued that the methodology and perspectives of the HGP were based in genetically determinist thinking (Lewontin, 2001). Equally, some anthropologists and philosophers have criticized the HGP as being anti-evolutionary and pre-Darwinian.[5]

Despite these early criticisms, public debate has taken longer to form. The ethical, legal, and social implications of the HGP and its associated technologies had often been addressed in public forums by journalists and pressure groups who raised fears about genetic discrimination, the misuse of public funds, and issues regarding intellectual property rights. However, the scientific merit of the HGP, i.e. whether or not sequencing efforts and gene discovery techniques would provide the knowledge described by their advocates, had remained largely outside the public purview.

This changed when in February of 2001, the international consortium responsible for the sequencing effort published a draft sequence containing an unforeseen insight—the human genome consisted of only approximately 30,000 genes, far fewer than the 150,000 predicted by some scientists, and too few to alone account for the complexity of inherited traits in humans or the differences between humans and other organisms. The gene was obviously not the only source of biological difference, a concept that caused some of the original claims of the HGP to fall into public dispute. Critics included Barry Commoner who, in a February 2002 feature article for *Harper's Magazine*, saw the HGP's outcome as

troubling some of the central theses of genetic science (Commoner, 2002). Although himself criticized for reductive thinking, Commoner's argument—and the public forum in which the issue was raised—demonstrate renewed attention to the scientific understandings associated with the HGP.

In addition, molecular biologists themselves have noted the "surprise" of the HGP outcome. Richard Strohman, an emeritus professor in the Department of Molecular and Cell Biology at UC Berkeley, wrote about this discovery:

> To me it suggests nothing less than a breakdown of the major paradigm guiding the entire HGP effort. That is, it was nothing less than the failure of genetic determinism . . . But after almost a century of life sciences dominated by this theory, and after ten years of the HGP dedicated to finding the genes for human diseases, their diagnosis and cure, and much more; . . . after all that, to now announce that the entire project was based on an incomplete theory, would have been much more than a shock . . . it would be a scandal. (Strohman, 2001)

And an article in the January 2002 issue of *Nature Genetics* states:

> The sequencing of the human genome has raised important questions about the nature of genomic complexity. It was widely anticipated that the human genome would contain a much larger number of genes (estimates based on expressed-sequence clustering ran as high as 150,000 genes) than *Drosophila* (14,000 genes) or *Caenorhabditis elegans* (19,000 genes). The report of only 32,000 human genes thus came as a surprise. (Modrek and Lee, 2002: 13)

This same article describes a process called alternative splicing, a theory that explains how one gene might be responsible for many different expressions of mRNA and thus of various proteins. Despite the process being well documented in several genes since the 1980s (articles from the journal *Cell* and the *Proceedings of the National Academy of Science* are cited in defense of this), the researchers state: "The study of alternative splicing has long been a valuable subfield within molecular biology, but has received comparatively little attention compared with major fields such as the discovery of new genes or transcriptional regulation" (Modrek and Lee, 2002: 13). Hidden within this brief comment is something of a criticism directed at past genomic research. The authors seem to be taking the small number of human genes revealed by the HGP to indicate that genomic research has not to date followed some of the possible productive paths available.

5. The HGP as "foundation"

An article written by members of the US National Human Genome Research Institute and published soon after the 14 April 2003 announcement of the completion of the HGP, sets forth a "blueprint"[6] for the future of genomic research: "The completion of a high-quality, comprehensive sequence of the human genome, in the fiftieth anniversary year of the discovery of the double-helical structure of DNA, is a landmark event. The genomic era is now a reality" (Collins et al., 2003: 835). The authors depict genomic research as a building, with the HGP providing a foundation upon which future research will be built (Figure 1).

They go on to say: "The broadly available genome sequences of human and a select set of additional organisms represent **foundational**[7] information for biology and biomedicine" (Collins et al., 2003: 837). Peter Glasner (2002) has commented on this picture, seeing it as an example of a discursive twist by proponents of the HGP. Glasner notes that in order for

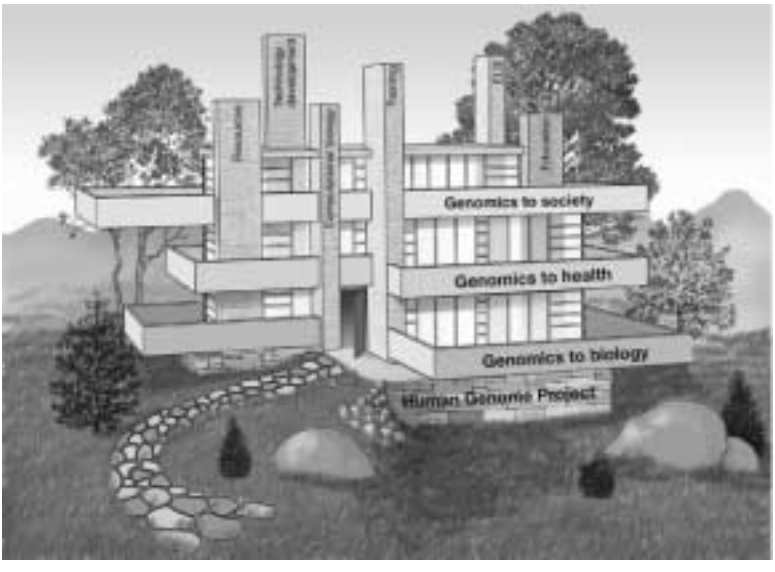

Figure 1. The future of genomics rests on the foundation of the Human Genome Project. *Source:* Collins, F., & Galas, D. (1993). A New Five-Year Plan for the United States Human Genome Project. Science, 262, 43-46

the HGP to become a "foundation" rather than the "Holy Grail" of biological science, the results of the HGP have had to be transformed into a "black box," accomplished "through the simple device of admitting that molecular biologists knew all along that the HGP would not provide access to the 'Holy Grail'" (Glasner, 2002: 269).

It is true that the metaphor of "foundation" is doing important work for the HGP and for new work being developed by previous members of the HGP consortium. But it should be noted that this is not the first time that the term has been used in the context of the HGP. A search through the project reports and press releases of the US part of the HGP, run by the National Human Genome Research Institute (NHGRI)[8] reveals that the term "foundation" is used in both press releases and project reports as far back as 1993 to describe the work of the HGP. The first usage occurs in the following quote:

> Thus, development of efficient technology for approaching detailed analysis of several megabase sections of the genome will provide a useful bridge between conventional genetics and genomics, as well as a **foundation** for innovation from which future methods for analysis of larger regions may arise. (NHGRI report 1993, originally published as Collins and Galas, 1993)

Similarly, the term is used in a 1996 press release:

> The technology development projects hope to provide not only the improvements needed to complete the gargantuan human DNA sequence by the year 2005, but also a **foundation** for 21st-century biomedical research technology, which will rely heavily on comparing and analyzing the sequence of entire genomes from several individuals or from different species. (NHGRI, 1996)

And in the 1998 report:

> Beyond that, research must be supported on new technologies that will make even

> higher throughput DNA sequencing efficient, accurate, and cost-effective, thus providing the **foundation** for other advanced genomic analysis tools. (NHGRI report 1998, originally published as Collins et al., 1998)

Such uses of the term would seem to defend Collins et al.'s depiction of the HGP as just the starting phase in a larger genomic project and deflect Glasner's criticism that this depiction is a new move by members of the HGP. However, Glasner is right in seeing this metaphor taking on new power at the end of the project, as is witnessed by the number of times the term is used in 2003: "The Human Genome Project has provided us with a wonderful **foundation**, but obviously having the human genomic sequence is not enough . . ." (NIH, 2003). And in a press release on 14 April 2003, announcing the completion of the human genome sequence and subtitled, "All Goals Achieved; New Vision for Genome Research Unveiled": "Biomedical researchers now have tremendous **foundation** on which to build the science and medicine of the 21st century" (NHGRI, 2003). And: "With this **foundation** of knowledge firmly in place, the medical advances promised from the project can now be significantly accelerated" (NHGRI, 2003). And in the triumphant article in *Nature*: "If we, like bold architects, can design and build this unprecedented and noble structure, resting on the firm bedrock **foundation** of the HGP, then the true promise of genomics research for benefiting humankind can be realized" (Collins et al., 2003: 847).

While many metaphors have been examined in the context of the HGP (e.g. Kay, 2000; Nerlich and Hellsten, 2004; Van Dijck, 2000), why focus on this one? We see the term "foundation" as acting as an important splicing metaphor for the HGP, due to the way it leverages ontological claims about genetics while remaining somewhat ambiguous, provides a temporal order to both the past and the future of the project, and works to connect up various public and private interests.

Some of the usefulness of "foundation" is specifically related to it being used in a metaphoric sense. Theories of metaphor have emphasized the way metaphors map meanings between two different domains, often referred to as the *source domain* and the *target domain* (Lakoff, 1993; Lakoff and Johnson, 1980). They do this by providing a variety (but not unlimited set) of characteristics, some of which are emphasized when the term is used in a metaphoric sense.

Such mappings are most obvious in the above picture where genomic science is represented as a building and the HGP (the target domain) presented as its foundation (source domain). Here, the physical world is brought to bear on the definition of the HGP—the cement and mortar of an actual foundation, with a selection of its physical characteristics such as its strength, are mapped onto the HGP. Metaphors, with their ability to map meanings, are a perfect candidate for acting as *splices*, not just between source and target domains (source–target), but also between source and multiple target domains (source–target–target).[9]

In other words, splicing metaphors do not just map meanings between a metaphoric term and that which it is intended to represent, i.e. the HGP and a physical foundation, but also help splice together the diverse set of elements that make up the HGP. The metaphor of foundation does this not by uniting all these elements under a single coherent shared sense of what the metaphor of foundation means, but by relying on the multitude of meanings associated with it.

"Foundation" is a successful characterization of the HGP for a number of reasons. First, "foundation" by explaining past genomic work and predicting future scientific developments provides a temporal order, placing the HGP at a specific point along a longer scientific trajectory of genomic work. In this way, foundation leverages the past by telling a story

about it and, at the same time, helps to set what Brown and Michael (2003) called the work of "scientific expectations." This use of the metaphor of foundation derives from the physical sense articulated above, but focuses on temporality. A foundation of a building comes, obviously, first. Thus, by calling the HGP a foundation for future work, rather than the work itself, Collins et al. make the future development of genomic science contingent on the development of the HGP.

Equally, the term foundation has a number of associated meanings. A glance at Roget's English Thesaurus (2005) reveals connections to terms that denote both *support*, such as ground, base, basis, and terra firma, but also terms denoting *cause*, such as origin, source, principle, element, and prime mover.[10] *Foundation*, in a metaphoric sense, partakes of these, as well as other meanings.

One of these is graphically illustrated in the picture from the vision paper of the NHGRI. Here, foundation is meant as *preparation*, as in provision, providence, anticipation, foresight, groundwork, cradle, stepping-stone, or scaffold. But this use belies a kind of artful ambiguity in this term, which could easily be perceived as meaning cause. A search through abstracts and titles from a corpus of molecular biology journals[11] reveals many ontologically similar metaphors at work. Examples include metaphors such as "genetic background," "genetic basis," and "genetic origin." Thus, the metaphor of foundation suggests the classic sense of genetics as the basis and cause of biological difference and, at the same time, provides a connection to a more sophisticated and less-ontologically bound sense of genetics as support for biological difference.

Within the context of the HGP, this balancing act is particularly important given the criticisms of "genetic determinism" that have been leveraged against the HGP. Foundation becomes a way of rearticulating powerful but deterministic genetic concepts and simultaneously incorporating ideas about genetic variation, as in the quote: "Develop the intellectual **foundations** for studies of sequence variation. The methods and concepts developed for the study of single-gene disorders are not sufficient for the study of complex, multigene traits" (Collins et al., 1998: 686).

Finally, as shown in the above quote reproduced here, the term foundation refers to not just the scientific ideas of the HGP, but also its larger mission of technological development and training: ". . . research must be supported on new technologies that will make even higher throughput DNA sequencing efficient, accurate, and cost-effective, thus providing the **foundation** for other advanced genomic analysis tools" (Collins et al., 1998: 685).

Thus foundation connects up a constellation of science and technology interest groups, including classical geneticists, molecular biologists, genetic epidemiologists, and sequencing technology manufacturers—as well as science policy advisors whose goal is the maintenance and development of a genomic science infrastructure. It does this by explaining why the HGP was funded in the first place, and positing what the value of this work might be for future research, and what the role of the HGP in genomic scientific and infrastructural development might be.

To briefly restate some of the points from above, the HGP has been accused of overvaluing the role of genomic sequences for the development of new medicine and health benefits and undervaluing the role of environmental and lifestyle factors in the progression of disease. These criticisms are based on two perceptions. First, that the HGP's construction of a single genomic "reference" sequence ignores normal genetic diversity (a perception that has caused some critics to call the HGP "anti-evolutionary" and "pre-Darwinian"). Second, that the HGP's emphasis on "pure" genetics enforces an understanding of genes as the "ultimate" and fundamental cause of biological difference. Above, we detailed the relation-

ship between these criticisms and the use of the metaphor of "foundation" in describing the scientific work and outcomes of the HGP. What then are some of the splicing metaphors at work in "post-genomic" work? First though, how are various people, scientific concepts, and technologies being connected in new project formations?

6. From genomics to "post-genomics"

The completion of the HGP was marked by increasing public debate as to the scientific merit of the project. The characterization (or re-characterization) of the HGP using the metaphor of "foundation" was one such response. Others have included a host of new projects, many of them focused on addressing and overcoming the claims of "genetic reductionism" associated with the HGP and genomics more generally.

Various scientists from within biology have begun to emphasize new strategies that explore the relationship between genes and biological traits. Scientific efforts include the creation of new genomic databases developed specifically to explore the phenomenon of alternative splicing,[12] as well as newly emphasized areas of exploration such as that of proteomics. In both these cases, the novelty of development typically lies in the creation of new bio-informatics tools focused explicitly on addressing the relationship between genes and expressed biological traits. An important example of this trend is the creation of new large-scale database projects, oriented towards linking genetic sequence information to environmental and lifestyle data. Key resources for such studies are existing population registers, including those devoted to tracing family lineages, as well as compilations of biological samples.

These projects are different from previous genetic databases. First, the digital data they are recording and archiving also include demographic, health care, and physiometric records—not just data from genotyping or protein sequencing—as well as the preservation of biological materials such as tissue and blood samples. Thus biobanks serve as a good example of four related strategies in current biology; first, the movement towards an explicit and programmatic acknowledgement of the complexity of gene and trait relationships, second, the use of bio-informatics techniques as exploratory tools of discovery rather than as part of a more straightforward "decoding" effort, third, the development of network infrastructures that link up and provide access to a vast array of different databases, and fourth, the alignment between various disciplines and interests within biology, clinical work, and public health initiatives.

This development is in part based on an increased understanding of the complex relationship between environment and genetics in the development of inherited traits and disease. Early genetic research saw the mapping of the human genome and the homological comparison of genotypes in order to isolate so-called "defective" genes as a valuable diagnostic tool. This research demonstrated its success by revealing the genetic causes of such conditions as Huntington's disease and certain forms of breast cancer. However, geneticists soon discovered that single gene conditions were comparably rare and that most diseases, such as those related to cardiovascular health, arthritis, and various cancers are caused by multiple gene interactions, themselves often "kicked-off" by environmental factors. Thus, while the HGP focused on creating a single "reference" genetic sequence, new "biobank" projects combine and compare various types of data, and use a variety of bio-informatics approaches.

One traditional way of mapping the relationship between gene function and structures is through the use of homologues—structurally characterized gene sequences whose functions are unknown are compared to sequences where the function is understood (typically from well-characterized model organisms). This method can work very well in increasing understanding of the pathways of genetic interaction that can lead to simple or monogenetic diseases—for example, how a mutation in a gene can result in the production of an abnormal protein such as in cystic fibrosis. However, in most common diseases, the relationship between genes and disease formation is much more complex. Multiple genes may interact to create a disease state, environmental factors may increase or decrease the likelihood of disease onset, and the intermediary steps between the existence of a genetic polymorphism and a disease might be not well understood.

As a response to the limitation of homologous searches and monogenetic approaches, new statistical methods are being developed and used in post-genomic research. These methods, including linkage analysis, linkage disequilibrium, and association analysis require large data sets and sophisticated statistical knowledge. An exemplar of these new kinds of genetic database projects is the GenomEUtwin project.

7. GenomEUtwin

GenomEUtwin was funded in 2002 by the European Union (EU) under a special grant structure aimed at facilitating the integration of various genome projects across Europe. The GenomEUtwin project is coordinated by Professor Leena Peltonen, located at the National Public Health Institute in Helsinki, Finland. Other partners are twin register studies located in various European countries. These twin registers, many of which date back to the early twentieth century, often contain blood and tissue samples from most twins born in the country as well as health records and follow-up interviews. In addition, each twin study group is responsible for a core specialization, such as database knowledge, statistics, or epidemiological expertise. However, rather than being in charge of all aspects of their specialization, the core groups act as advisors to all the other groups.

In order to verify the role of genetic variation in common traits such as stature or weight or in the cause of various diseases such as arthritis or heart disease, GenomEUtwin plans to compare the over 30,000 DNA samples already gathered within the six participating twin studies to two existing sources of trans-European population study data, MONICA and its continuation, the MORGAM project. These two projects have, between them, collected epidemiological and health care data relevant for cardiovascular diseases since the early 1980s and currently contain data from over 187,000 individuals, gathered at 12 different European centers. By bringing together twin cohort studies and traditional epidemiology (the MONICA and MORGAM projects), GenomEUtwin hopes to leverage the “unique features” of its European context.

With these goals in mind, GenomEUtwin plans to construct two linked databases. The first, called the Geno-Type database, or GTDB, is being developed as an open source software project.[13] This database will contain the minimal information necessary to do gene mapping studies, specifically data on markers, alleles, samples, and individuals. The second database, linked to the first by way of an anonymized identifier, will contain the records pulled from the various twin study databases. Rather than being a fully centralized database, the plan is to use what is called a “federated” or distributed database, to pull records “on the fly” from the twin registries.

8. New directions in genetic research

The GenomEUtwin project, like the HGP before it, can serve as an exemplar of new directions in genetic research. Its goals, listed on the project home page (GenomEUtwin, 2004b), emphasize its European context, the need to leverage this geographical location, and a desire to extend collaboration between European scientists:

> We aim to capitalize **special advantages of Europe** in population genetics by efficient collaboration of twin researchers, genetic epidemiologists, molecular geneticists and mathematicians. Our goal is to identify critical genetic and life-style risk factors for common diseases using **European strengths** in genetics, epidemiology and bio-computing. (GenomEUtwin, 2004b)

Further, the home page emphasizes its temporal location as well, explaining that *now* is the time for biological research: "The genome sequence, detailed information of genetic variations between individuals, high-throughput molecular technologies and novel statistical strategies **create new possibilities** to define genetic and life-style risk factors behind common health problems" (GenomEUtwin, 2004b).

Finally, the home page explains the steps that are necessary to create these new possibilities:

> **Studies of large population cohorts are needed** to transform the genetic information to detailed understanding of the predisposing factors in diseases affecting most human populations . . . This project will **apply and develop new molecular and statistical strategies** to analyze unique European twin and other population cohorts to define and characterize the genetic, environmental and life-style components in the background of health problems . . . (GenomEUtwin, 2004b)

These same issues were emphasized in our interviews with key GenomEUtwin participants—as well as some of the difficulties involved. In a joint interview with two members of the GenomEUtwin database core, Dr. Nancy Pederson, head of the Swedish Twin Register, and Dr. Jan-Eric Litton, Professor of Biomedical Computing Technology at the Karolinska Institute, each emphasized slightly different goals:

> Pederson: Well I think we have perhaps little bit different perspectives because Jan and I come from very different backgrounds and what-not. For me, being a director of a twin registry and a behavior geneticist and genetic epidemiologist, **the goals are to set up an infrastructure and be able to go ahead and start the basis for gene finding** for common complex disorders, but there are a lot of different aspects to the GenomEUtwin project in getting there . . .
>
> Litton: I'm the computer guy. **The goals are to connect Europe** . . .

Here, while Pederson emphasizes the need to set up a common infrastructure and start doing scientific work, Litton emphasizes the need to connect European researchers. Equally, they see the problems as slightly different:

> Litton: I mean the biggest problem is not technology. The biggest problem is I guess, **different ethics** from different countries, researchers to work together in a way they haven't done, ever I guess . . . so, that's the main problem.
>
> Pederson: Well the different countries have **different guidelines and laws** about access to data so that's one part of it and that is getting **people who come from monogenetic inheritance type of studies** with isolates and things like that, people who are used to

> looking at major gene effects to work with those of us who are in the field in a very uh, accustomed to working with quantitative traits and, uh . . . **complex gene interactions** and at the same time, those of us who have **no background in what a database is**, and how they need to be constructed to be communicated appropriately and what-not . . .

In these quotes, both Litton and Pederson emphasize the different ethics and legal frames within the various countries involved. However, Pederson also mentions the different scientific backgrounds of GenomEUtwin participants. She distinguishes between people working from traditional epidemiology and molecular biology "who come from monogenetic inheritance type of studies" and "those of us . . . accustomed to working with . . . complex gene interactions." She ends by including as a third category, the new technologies involved.

For GenomEUtwin to be successful a diverse set of elements must be connected. Scientists from molecular biology, traditional epidemiology, twin research, and bioinformatics must find common ground, despite sometimes conflicting scientific understandings and ways of working. Equally, these scientists are located in different national contexts, often with different ethics and laws governing appropriate scientific work. Finally, the above quotations underscore the importance of technology and infrastructure as mediating between various disciplines, this despite the fact that not all groups have the same level of technical expertise or equipment. For example, when asked to describe the biggest challenge to their research, the respondents replied:

> Pederson: . . . for the database core it's been the challenge that even two different genotyping laboratories, one of them had state of the art . . .
>
> Litton: and the other one had Excel . . .

Thus, for the GenomEUtwin project, the construction of a joint, distributed database *and* complex statistical analytic methods serve as both the focus of the work, and a place where collaboration can take place. This is an important point to emphasize. The database itself serves as a common object around which a complex set of interests can organize. This was particularly emphasized in the technical annex to the EU funding proposal:

> We aim to **network individual research groups with only partially overlapping interest** with the project by **creating a system that facilitates distribution of statistical analyses** of their data and even pooled data analyses over the Internet. (GenomEUtwin, 2001)

And:

> We aim to develop **a computer infrastructure** facilitating data mining of different existing databases. An important component of the system will be **effortless merge of genetic and environmental and life-event data**. We will to maximal extent harmonize the databases in participating centers and create a good structure for database management. (GenomEUtwin, 2001)

Another key development here is that while the author uses the metaphor of data mining, she or he extends this with notion of different databases and the concept of an "effortless merge" of various kinds of data.

Further, GenomEUtwin has the express goal of linking strategic interest (competitiveness) and scientific and health value:

> The expected outcome of this proposal will have a tremendous **strategic impact** on many areas of the European health care-associated industry. We will produce innova-

> tions that will have wide-ranging applications and have tremendous potential to improve **the competitiveness of Europe**. (GenomEUtwin, 2001)

Finally, it is stated that the GenomEUtwin project was created to address the complexity of gene–environment interactions and address dissatisfactions with current scientific perspectives on complex disease. In a special issue about GenomEUtwin in the journal *Twin Research*, some of the researchers state their opinion about using major gene effects to explain complex conditions like heart disease: "This may sound plausible enough, but when one considers that there might be several hundred polymorphisms involved in various systems and pathways which may affect atherosclerosis, **the prospect is daunting**" (Evans et al., 2003: 433). But as "daunting" as the prospect might be, the authors go on to criticize epidemiologists who look for gene effects within epidemiological data, but do not connect their findings to genetic sequence information. They consider this a practice that is "no longer acceptable" (Evans et al., 2003: 433) given the information and techniques available to researchers since the end of the HGP.

The goals of GenomEUtwin detailed above are revealing of a number of important trends in large-scale biological work. These include the recognition that biology projects can help create strategic resources for Europe in terms of technical infrastructures and training, the way computing projects like databases serve as a common scientific object around which diverse interests can unite, and an express acknowledgement of the need to develop new tools and technologies for addressing the complexity of genetics.

9. "Profile" as a splicing metaphor

The reasons for a shift from genomic to "post-genomic" research are described in an article by Leena Peltonen, head of the GenomEUtwin project:

> As **the tools of the genome project are now in hand** (Collins et al., 2003) **there is no excuse** for not minutely characterizing the **predisposing genetic profiles** underlying common traits. Based on this information we can then start to dissect the environmental and lifestyle components with high precision . . . (Peltonen, 2003: 354)

Here, Peltonen is both recognizing the contribution of the "foundation" created by the HGP with her use of Francis Collins' own language, and, at the same time, indicating that it is now not enough. The next step according to Peltonen is to use the tools and information generated by the HGP and to extend it by focusing on genetic variation as well as lifestyle and environmental data. Notably, Peltonen uses the term "genetic profiles" to underscore her emphasis.

We see "profile" as a particularly productive splicing metaphor at work within the GenomEUtwin project. Like foundation, it is not a new term; a search through the Biosis abstract database[14] reveals a steady growth in usage from the around 7,000 times it appeared in biology journal article titles and abstracts in 1993 to approximately 16,000 times in 2003. In addition, like foundation, "profile" describes a discrete concept with spatial, temporal, and physical characteristics that can be understood as important to the GenomEUtwin organization.

How is it used?

The metaphor of profile is used in a variety of documents associated with GenomEUtwin. These include funding proposals, working documents, as well as scientific reports. For

example, the technical annex for the GenomEUtwin project made liberal use of the term: "Using environmental and lifetime event data and genetic information we aim to use novel statistical analyses to produce genetic and life style **risk profiles** for these traits" (GenomEUtwin, 2001). This usage mirrors something of a shift in definition. Like foundation, profile is not a term that has been used uncontentiously within the domain of genetics. The concept of "genetic profiling" has been debated both in public policy (e.g. Brice, 2004; NHS, 2003) and in journalist accounts (e.g. GeneWatchUK, 2003; Lambert, 1998; Lawless, 2004). However, the debated definition of profiling in those contexts is often attached to the use of genetic indicators in forensics (i.e. in criminal cases) and the possibility of genetic discrimination based on genetic testing. The use of "profile" above to refer to a combination of genetic and environmental data is a slight but important move away from a previous pure genetic focus. While the previous quotation described the goal of GenomEUtwin as "characterizing the **predisposing genetic profiles** underlying common traits" this quotation focuses on "produce[ing] genetic and life style **risk profiles**." This double-use, then blurs the line between profile as a physical construct associated with an individual, and profile as the result of "**novel statistical analyses**." This connection to statistical analysis leverages the idea of a profile as "a formal summary or analysis of data, often in the form of a graph or table, representing distinctive features or characteristics" (The American Heritage® Stedman's Medical Dictionary, 2002 [2001, 1995]).

> These **profiles** provide the basis for the evaluation of gene-environment interaction behind these traits. Such **profiles** are of utmost importance for health care related industries as well as national health care and prevention programmes in all European countries. (GenomEUtwin, 2001)

This quote reveals that although the term profile does move towards a more complex theory of gene interaction, it does not entirely move away from the idea of genes as "foundational" for biological difference. Thus, while profile does not replace foundation, it does provide something of an ontologically less bound relationship between genes and environment. For example, the ideas of a profile as only an outline or contour or as "a human head represented sidewise, or in a side view; the side face or half face" (Webster's Revised Unabridged Dictionary, 1998 [1996]) are ways of recognizing that genes are only part of the overall picture of disease and biological trait development. Also, a direct link is made between the generation of profiles and industrial as well as public interest.

Just like foundation then, the metaphor of profile provides a way to splice together a series of ideas about genetics, the development of technological solutions, and corporate and public interest. The complex set of ideas, technologies, and various interests connected with the metaphor of profile is particularly obvious in the below quote, which also demonstrates the most powerful aspect of a splicing metaphor—its ability to map scientific concepts to organizational and social issues:

> The objective of the Project is to produce **genetic and life style risk profiles** at the population level. The eventual aim is to produce **risk profiling** that could be used in the future to advise an individual and/or her/his physician in making lifestyle and other choices that may influence health. **Risk profiles** are not open to abuse, for example, by insurance companies or employers. Information on such **profiles, including genetic, environmental and lifestyle profiles** will be invaluable in selecting study samples for phase II and III trials in drug development. Such **profiles** will pave the way for tailored drug choices for various patient groups, thus increasing the success of treatment, reducing complications and reducing drug development and treatment costs. (GenomEUtwin, 2004a)

This quotation makes obvious the diverse uses to which the splicing metaphor of profile is put. Most importantly, the depiction of profile as a predictive statistical outcome of risk at both the individual and the population level, is leveraged for both social ("not open to abuse") and economic ("reducing drug development and treatment costs") purposes.

Above we pointed to the development of bio-informatics techniques as exploratory tools of discovery rather than as part of "decoding" efforts, as one of the novel aspects of "new genomics" projects. Here, a possibly useful comparison can be made between genetic profiling for genetic epidemiology and criminal profiling.

A key task of criminal profilers is to link up diverse cases in order to begin to determine what is called the "signature" of the criminal, "those behaviours or actions that fulfil a psychological or physical need in the offender" (Petherick, 2004). Questions related to "fitness" are often connected to the practice of profiling—figuring out how well a particular case or individual meets of the contours of the established profile. In addition, profiles are often sorted into larger categories or typologies for assessment and statistical analysis. This seems similar in many ways to the practices emphasized in the following quote:

> **Risk factor profiles** change throughout an individual's life span, necessitating knowledge of the status of these characteristics at multiple time points. Longitudinal data, when used in analyses of a twin cohort – and ultimately, combined with genotypic information – could provide valuable opportunities to learn how traits are affected by age-genotype and environment interactions over time. (Evans et al., 2003: 435)

Like the aim of criminal profiling, the main task of genetic and risk profiling is to take disparate forms of information and make them comparable so as to generate a single "picture" of a complex phenomenon. In criminal profiling, the goal is to merge forensic, contextual, information about the victim, and assumed behavioral and psychological characteristics, in order to construct a single view of an offender. For the GenomEUtwin project, profiling serves as a way of uniting a series of disparate data sources into a single and generalizable result and, at the same time, maintaining a connection between the individual case and more general, population-level phenomenon. Similar dynamics have been noted by Beaulieu (2004), in her work on digital brain databases.[15]

Profile thus serves as a good splicing metaphor of the GenomEUtwin project (and possibly other population biobanks) owing to the way it maps associations with statistics, entails horizontal rather than vertical approaches to gene finding, emphasizes longitudinal and environmental data, and de-emphasizes ontologically bound metaphors such as foundation, without throwing them out altogether. In addition, the results of this work—profiles that provide details about gene–environment–lifestyle interactions—are nicely objectified data sources that can be stored in databases, distributed (and possibly bought and sold), and compared and contrasted in order to reveal not just scientific insight, but also possible health benefits through the development of tailored drugs and public health regimes. Thus the *profile* becomes a way of both discursively and materially linking a series of important elements in the construction of the GenomEUtwin database project. One of the key aspects of this linking process is the way the insights ascribed to scientific work are connected to public (health policy, strategic infrastructure needs) as well as private (corporate) interests.

10. Comparison of foundation and profile

Although both terms are used as splicing metaphors within the respective projects, it is important to recognize the similarities and differences between them. The most immediate

difference is the focus of each term. While both implicitly leverage scientific claims in order to legitimate the past and/or present need for the organization, their explicit focus is different. While *foundation* is used in the context of the project to describe both the role of genetics and the role of the HGP, *profile* is used to describe desired outcomes and, more importantly, deliverable products of GenomEUtwin. *Foundation* leverages scientific claims in order to explain why the HGP was necessary, *profile* connects GenomEUtwin to the promissory science of, among other developments, tailored medicines, health benefits, and, interestingly, protections against genetic discrimination (see above). This difference in focus is connected to the current temporality of each project: whereas GenomEUtwin is still in development, the HGP is over—though this does not decrease its power as an exemplar.[16]

It is important to recognize the continuing power of the foundation term. GenomEUtwin does not attempt to replace this characterization. Instead, it leverages it by agreeing with the temporal argument made by describing the HGP as the foundation, and continues this work by constituting itself—and its product, profiles—as the necessary next step. In both cases, the physical and conceptual "shape" of the splicing metaphor provides resources for particular characterizations of the project or its result, and serves as a means by which the scientists involved in their development can weave together a necessary but diverse set of interests and ideas.

11. Conclusions: splicing metaphors and large-scale scientific work

In this paper we have attempted to detail some of the issues, tensions, and new developments in genomics, by focusing on two projects, the Human Genome Project and GenomEUtwin. We have argued a number of related points, the first being that large-scale biological projects have a number of goals in addition to the development of scientific knowledge and direct health benefits. These goals include the construction of technological infrastructures, the development of training, and the creation of networks of experts. Equally, these projects provide the possibility for strategic alliances between scientists, policymakers, technology providers, and public health experts. A key object for the development and maintenance of these constellations of interests is the technical objects themselves—the databases that exist as both one of the products of scientific projects and a tool around which the interests can organize. In this sense, databases serve as technical infrastructures, social arrangements, as well as policy objects.

We have used the concept of *splicing metaphor* to detail the constellations of science and technology interests, including classical geneticists, molecular biologists, genetic epidemiologists, sequencing technology manufacturers, and science policy advisors whose goal is the maintenance and development of a genomic science infrastructure. Here we have shown how *foundation* functions as a way to both link up various interests, and temporally position the HGP within a longer sequence of scientific progress. Equally, we have detailed how *profile* is used to extend more ontologically bound positions on genetics (criticized for being determinist), without throwing them away altogether. We have also linked the uses of profile within GenomEUtwin to what we see as the three other main strategies in new genomics projects; first, *profile* provides a description of the outcomes of new bioinformatics techniques used to explore genetic diversity and environmental influences; second, *profiles* serve as objects of exchange between and within the networks of experts using genomic databases; and third, the concept of genetic and lifestyle risk *profile* demonstrates the possibility of doing ethically secure scientific research (though this is never fully explained) with the desirability of creating tailored medicines and health therapies.

As we indicated in the section on metaphor theory (§2), the uses of metaphor within scientific communication and social organization are diverse. Our goal in this paper has been to extend these perspectives with a particular focus on the grounded analysis of terms that link together scientific concepts, ontological positions, multiple interests, and organizational needs. Our coining of the term "splicing metaphor" is an attempt to focus on the productive heterogeneity of language and, particularly, actors' own recognition and use of this fact.

We have attempted to remain somewhat neutral as to a normative evaluation on these projects, their promises, and their ultimate outcomes.[17] We see many aspects of these projects as positive, and demonstrative of a growing sensitivity towards social issues on the part of the scientists themselves. However, we also see the distributed networks of scientists, loosely organized by projects, trading in genetic information and located outside of any single national context, as an ethical and social problematic. While not denying the good will of the people involved (many of whom actively seek guidance on ethical and social issues), we would advocate for greater attention on these networks and the objects of their research.

To that end, we see terms like *foundation* and *profile* as productive tools for critical analysis. As Nerlich and Hellsten note, shifts in metaphoric expression are revealing of changes in the direction of large-scale genetic research (Hellsten, forthcoming; Nerlich and Hellsten, 2004). The shift from *foundation* to *profile* thus mirrors a similar transition from *sequencing* to *annotating* examined by Hellsten (forthcoming). Both metaphoric shifts are indicative of a movement away from genetic archiving and description and towards more sophisticated attempts to make sense of the results of past genomic research. We also see in the use of the metaphor of profile an often explicit recognition of the limitations of purely genetic techniques and the desire to direct attention towards the complicated relationship between genes, lifestyle, and the environment. It should also be noted that such analyses, while dependent on previous genomic projects like the HGP, require new constellations of experts, new types of technological infrastructures, and, in the case of GenomEUtwin, novel relationships between public and private interests in order to be successful.

Finally, terms that help to splice together these kinds of projects are also places where they can be unraveled. One of the key aspects of the splicing process noted above is the way the insights ascribed to scientific work are connected to public (health policy, strategic infrastructure needs) as well as private (corporate) interests. The distributed nature of large-scale biological projects and the increasingly sophisticated ways they link together multiple interests and diverse players make the analysis of such splices all the more important.

Acknowledgements

This work was done within the project "Metaphors of Databases as Early Indicators of Integration and Support in Scientific Research Networks" (NWO grant #458-02-005) at NERDI (Network Research and Digital Information). I would like to thank all my colleagues at NERDI for their comments and support, but especially Anne Beaulieu, Iina Hellsten, and Paul Wouters. Equally, I would also like to thank Sally Wyatt and Barbara Prainsack for their timely and helpful comments, as well as the members of the GenomEUtwin consortium for allowing access to their project.

Notes

1 For a more detailed analysis of "public deficit" models, the complicated construction of "publics" upon which they are based, and some of the reasons for knowledge gaps between scientific experts and laypeople, see Einsiedel (1998).

2 An interesting analysis would focus on exploring why various groups of scholars have tended to emphasize alternative views on metaphors. Such a disciplinary examination goes beyond the goals of this paper, but suffice it to say that we see a relationship between the explanatory aims of specific disciplines and their position on metaphor.

3 It is important to note that both "boundary objects" and "immutable mobiles" are typically understood as being socially productive (of organization or of knowledge), in part because they are material. While we focus on the materialized versions of the metaphors we study (e.g. in documents, pictures, and speeches) there is also something to be addressed in the non-material/conceptual aspect of these metaphors. A fruitful starting point for such an exploration might be the notion of the "imaginary" (Verran, 2001) and the "practice-bound imaginary" (Hyysalo, 2002, 2004).

4 The 2003 publication of a "complete" human genome sequence followed the "working draft" sequence that was released in 2001. The differences between these two sequences have to do with both the quality of the sequence, i.e. the error rate of assigned bases, as well as how contiguous it is. However, it has been noted that terms like "quality," "contiguous," and "complete" are defined differently by various sequencing efforts (*Nature* (15 February 2001) 409: 818–20).

5 For an historical overview of these criticisms, see Gannet (2003).

6 For more on the use of this term as a metaphor within the HGP, see Lippman (1992), Hubbard and Wald (1993), and Nelkin and Lindee (1995).

7 In the quotes that follow we will emphasize the terms under discussion by marking them in bold.

8 Four reports, written in 1991, 1993, 1998, and 2003, are available on the NHGRI web site at http://www.genome.gov/12010537.

9 This concept is explored in more detail in Fauconnier's (1994, 1997) work on conceptual blending.

10 This last definition is particularly interesting due to its connection to *primum mobile*, the term used by Aristotle to describe the "source of all motion." This connection between the concept of the DNA principle and Aristotelian physics was noted by Max Delbrück (Delbrück, 1971, cited in Kay, 2000: 38).

11 In order to create a corpus to compare to our research documents, we relied on a subject classification scheme developed by Loet Leydesdorff (2004). Using the journal titles associated with "molecular biology" we downloaded from PubMed all abstracts and titles included in these journals during 2003. Each set of abstracts and citations was compiled into a single text file which was then searched for all co-located word clusters associated with the term "genetic."

12 The Modrek and Lee (2002) article cited earlier provides a listing of various alternative splicing papers, projects, and databases. It is important to note that these authors themselves see purely bio-informatics approaches as suffering from a series of important problems (table 3 in the article), many of which can only be solved via experimental validations (i.e. "wet" laboratory work).

13 Details about the GTDB can be found at http://sourceforge.net/projects/gtdb/.

14 Biosis Previews is a subscription-based online searchable database of life science articles, conference proceedings, books, and patents. It constitutes a wider area of life science abstracts including, but not limited to molecular biology. We chose to use this database for our search in order to encompass the use of the term "profile" in epidemiology and genetic epidemiology in addition to molecular biology. It can be found at http://www.biosis.org/products/previews/.

15 She states,

> in a digitised informational setting such as that of the brain atlas, transformations of data are algorithmic and probabilistic. This means that the tension between the single case and the class of phenomena studied can be maintained, in a way that is not possible without significant computational power. (Beaulieu, 2004: 386)

16 Francis Collins, for instance, is currently head of the National Human Genome Research Institute. This institute was started to carry out the National Institute of Health's (NIH) role in the HGP. However, it has since become one of the 27 institutes that make up the NIH. As its Web site states, "The history of the HGP, the history of genomics, and the history of NHGRI, are inextricably intertwined" (http://www.genome.gov/10001763).

17 For an excellent overview of the social and ethical issues associated specifically with biobank research initiatives, see Rose (2001) and Einsiedel (2003).

References

Beaulieu, A. (2004) "From Brainbank to Database: the Informational Turn in the Study of the Brain," *Studies in History and Philosophy of Biological and Biomedical Sciences* 35(2): 367–90.

Berg, P. (2003) "The New Biology: Celebrating the Past, Imagining the Future," National Academy of Sciences

2003 Annual Meeting Symposium. URL: http://video.nationalacademies.org/ramgen/nas/042803_newbio_c.rm (accessed 16 December 2004).

Black, M. (1993) "More about Metaphor," in A. Ortony (ed.) *Metaphor and Thought*, 2nd edn, pp. 19–41. Cambridge: Cambridge University Press.

Bodmer, W. (1985) *The Public Understanding of Science*. London: The Royal Society.

Bono, J. (1990) "Science, Discourse, and Literature: The Role/Rule of Metaphor in Science," in S. Peterfreund (ed.) *Literature and Science: Theory and Practice*, pp. 59–89. Boston: Northeastern University Press.

Brice, P. (2004) "Barcode Babies: Prospects for Genetic Profiling," *Biolines* 1 June. URL: http://www.cambridgenetwork.co.uk/POOLED/ARTICLES/BF_NEWSART/VIEW.ASP?Q = BF_NEWSART_102586.

Brown, N. and Michael, M. (2003) "A Sociology of Expectations: Retrospecting Prospects and Prospecting Retrospects," *Technology Analysis and Strategic Management* 15(1): 3–18.

Brown, T.L. (2003) *Making Truth: Metaphor in Science*. Urbana: University of Illinois Press.

Bucchi, M. (1998) *Science and the Media: Alternative Routes in Scientific Communication*. London: Routledge.

Callon, M., Law, J. and Rip, A. (1986) *Mapping the Dynamics of Science and Technology*. London: Macmillan.

Chew, M.K. and Laubichler, M.D. (2003) "Natural Enemies—Metaphor or Misconception?," *Science* 301(5629): 52–3.

Clinton, B., Blair, T., Collins, F. and Venter, C. (2000) "Remarks by the President, Prime Minister Tony Blair (via satellite), Dr. Francis Collins, Director of the National Human Genome Research Institute, and Dr. Craig Ventner, President and Chief Scientific Officer, Celera Genomics Corporation, on the Completion of the First Survey of the Entire Human Genome Project," in Office of the Press Secretary, 26 June [online]. Office of Science and Technology Policy. URL: http://clinton4.nara.gov/WH/EOP/OSTP/html/00628_2.html.

Collins, F. and Galas, D. (1993) "A New Five-Year Plan for the United States Human Genome Project," *Science* 262: 43–6.

Collins, F.S., Green, E.D., Guttmache, A.E. and Guyer, M.S. (2003) "A Vision for the Future of Genomics Research," *Nature* 422: 835–47.

Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R. and Walters, L. (1998) "New Goals for the U.S. Human Genome Project: 1998–2003," *Science* 282: 682–9.

Commoner, B. (2002) "The Normal Genome in Twentieth-century Evolutionary Thought," *Harper's Magazine* February.

Delbrück, M. (1971) "Aristole-totle-totle," in E. Borek (ed.) *Of Microbes and Life*. New York: Columbia University Press.

Einsiedel, E.F. (1998) "Publics Problematized in the Public Understanding of Science," in M. Dierkes and C. van Grote (eds) *Between Understanding and Trust: the Public, Science and Technology*. Reading, MA: Harwood Academic Press.

Einsiedel, E.F. (2003) *Whose Genes, Whose Safe, How Safe? Publics' and Professionals' Views of Biobanks*. Ottawa: The Canadian Biotechnology Advisory Committee. URL: http://cbac-cccb.ca (accessed 16 December 2004).

Elzinga, A. (2004) "Metaphors, Models and Reification in Science and Technology Policy Discourse," *Science as Culture* 13(1): 105–21.

Evans, A., van Baal, G.C.M., McCarron, P., deLange, M., Soerensen, T.I.A., de Geus, E.J.C., Kyvik, K., Pedersen, N.L., Spector, T.D., Andrew, T., Patterson, C., Whitfield, J.B., Zhu, G., Martin, N.G., Kaprio, J. and Boomsma, D.I. (2003) "The Genetics of Coronary Heart Disease: The Contribution of Twin Studies," *Twin Research* 6(5): 432–41.

Fauconnier, G. (1994) *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge, MA: MIT Press.

Fauconnier, G. (1997) *Mappings in Thought and Language*. Cambridge: Cambridge University Press.

Gannet, L. (2003) "The Normal Genome in Twentieth-century Evolutionary Thought," *Studies in History and Philosophy of Biological and Biomedical Sciences* 34: 143–85.

GeneWatchUK (2003) "The Dangers of Genetic Testing Kits," *Guardian Unlimited* 19 January.

GenomEUtwin (2001) "Genome-wide Analyses of European Twin and Population Cohorts to Identify Genes Predisposing to Common Diseases," *QLRT-2001-01254*.

GenomEUtwin (2004a) *Ethical Issues for the GenomEUtwin Manual*. Unpublished.

GenomEUtwin (2004b) http://www.genomeutwin.org (accessed 17 August 2004).

Glasner, P. (2002) "Beyond the Genome: Reconstituting the New Genetics," *New Genetics and Society* 21(3): 267–77.

Hathaway, J. (2003) "Science Historians Ponder Naming 'Enemies' in Science Literature," *ASU News* 7 July.

Hellsten, I. (2002). *Politics of Metaphor: Biotechnology and Biodiversity in the Media*. Tampere: Acta Universitatis Tamperensis.

Hellsten, I. (forthcoming) "From Sequencing to Annotating: The Book of Life Enters the Post-genomic Era," *New Genetics and Society*.

Hellsten, I. and Leydesdorff, L. (2004) "Measuring the Meanings of Co-words in Contexts: Automated Analysis of 'Monarch Butterflies', 'Frankenfoods' and 'Stem Cells'," Paper presented at the Proceedings of the Sixth International Conference on Social Science Methodology: Logic and Methodology, Amsterdam, August.

Hesse, M.B. (1966) *Models and Analogies in Science*. Notre Dame, IN: University of Notre Dame Press.

Hubbard, R. and Wald, E. (1993) *Exploding the Gene Myth: How Genetic Information is Produced and Manipulated by Scientists, Physicians, Employers, Insurance Companies, Educators, and Law Enforcers*. Boston: Beacon Press.

Hyysalo, S. (2002) "Transforming the Object in Product Design," *Outlines* 4(1): 59–83.

Hyysalo, S. (2004) "Uses of Innovation: Wristcare in the Practices of Engineers and the Elderly," academic dissertation, University of Helsinki, Helsinki.

Irwin, A. and Wynne, B. (eds) (1996) *Misunderstanding Science? The Public Reconstruction of Science and Technology*. Cambridge: Cambridge University Press.

Kay, L.E. (2000) *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford, CA: Stanford University Press.

Lakoff, G. (1993) "The Contemporary Theory of Metaphor," in A. Ortony (ed.) *Metaphor and Thought*, pp. 202–51. Cambridge: Cambridge University Press.

Lakoff, G. and Johnson, M. (1980) *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Lambert, B. (1998) "Giuliani Backs DNA Testing of Newborns for Identification," *New York Times* 17 December.

Latour, B. (1987) *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Lawless, J. (2004) "Discoverer of DNA Fingerprinting Worries about Misuse of the Technology," Associated Press, New York, 8 September, p. 1. URL: http://www.ap.org (accessed 15 June 2005).

Leatherdale, W.H. (1974) *The Role of Analogy, Model, and Metaphor in Science*. Amsterdam/New York: North-Holland/American Elsevier.

Leggett, M. and Finlay, M. (2001) "Science, Story and Image: A New Approach to Crossing the Communication Barrier Posed by Scientific Jargon," *Public Understanding of Science* 10(2): 157–71.

Lenoir, T. (1999) "Shaping Biomedicine as an Information Science," in M.E. Bowden, T. Bellardo Hahn and R.V. Williams (eds) *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*, ASIS Monograph Series, pp. 27–45. Medford, NJ: Information Today, Inc.

Lewontin, R. (2001) *It Ain't Necessarily So: The Dream of the Human Genome and Other Illusions*. New York: New York Review of Books.

Leydesdorff, L. (2004) "Clusters and Maps of Science Journals: Based on Bi-connected Graphs in the Journal Citation Reports," *Journal of Documentation* 60(4): 371–427.

Lippman, A. (1992) "Led (Astray) by Genetic Maps: the Cartography of the Human Genome and Health Care," *Social Science and Medicine* 35(12): 1468–76.

Maasen, S. and Weingart, P. (2000) *Metaphors and the Dynamics of Knowledge*. London: Routledge.

Miettinen, R. (2002) *National Innovation System. Scientific Concept or Political Rhetoric*. Helsinki: Edita.

Modrek, B. and Lee, C. (2002) "A Genomic View of Alternative Splicing," *Nature Genetics* 30(January): 13–18.

Nelkin, D. and Lindee, S. (1995) *The DNA Mystique: The Gene as Cultural Icon*. New York: W.H. Freeman.

Nerlich, B. and Hellsten, I. (2004) "Genomics: Shifts in Metaphorical Landscape between 2001 and 2003," *New Genetics and Society* 23(3): 255–68.

NHGRI (1996) "Faster, Cheaper Methods to Read DNA Focus on Systems Integration, Miniaturization," URL: http://www.genome.gov/10000509 (accessed May 2005).

NHGRI (2003) "International Consortium Completes Human Genome Project," URL: http://www.genome.gov/11006929 (accessed May 2005).

NHS (2003) *Our Inheritance, Our Future—Realising the Potential of Genetics in the NHS*, Genetics White Paper. URL: http://www.dh.gov.uk/PolicyAndGuidance/HealthAndSocialCareTopics/Genetics/GeneticsGeneralInformation/ GeneticsGeneralArticle/fs/en?CONTENT_ID = 4016430&chk = RnGBgL.

NIH (2003) "Beyond Genes: Scientists Venture Deeper into the Human Genome," URL: http://www.genome.gov/11009066 (accessed May 2005).

Peltonen, L. (2003) "GenomEUtwin: A Strategy to Identify Genetic Influences on Health and Disease," *Twin Research* 6(5): 354–60.

Pepper, S.C. (1942) *World Hypotheses: A Study in Evidence*. Los Angeles, CA: University of California Press.

Petherick, W. (2004) "Criminal Profiling: How it Got Started and How it is Used," URL: http:// www.crimelibrary.com (accessed May 2005).

Putnam, L.L., Phillips, N. and Chapman, P. (1996) "Metaphors of Communication and Organization," in S.R. Clegg, C. Hardy, and W.R. Norad (eds) *Handbook of Organization Studies*, pp. 375–408. Thousand Oaks, CA: SAGE.

Roget's English Thesaurus (2005) URL: http://humanities.uchicago.edu/orgs/ARTFL/forms_unrest/ROGET.html (accessed May 2005).

Rose, H. (2001) *The Commodification of Bioinformation: The Icelandic Health Sector Database*. London: The Wellcome Trust.

Schoen, D.A. (1979) "Generative Metaphors: A Perspective on Problem Solving in Social Policy," in A. Ortony (ed.) *Metaphor and Thought*, pp. 254–83. New York: Cambridge University Press.

Schultze, U. and Orlikowski, W.J. (2001) "Metaphors of Virtuality: Shaping an Emergent Reality," *Information and Organization* 11(1): 45–77.

Star, S.L. and Griesemer, J.R. (1989) "Institutional Ecology, Translations and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39," *Social Studies of Science* 19(3): 387–420.

Strohman, R. (2001) "Human Genome Project in Crisis: Where is the Program for Life?," *California Monthly* 111(5) [online]. URL: http://www.alumni.berkeley.edu/Alumni/Cal_Monthly/April_2001/ (accessed June 2005).

Sturgis, P. and Allum, N. (2004) "Science in Society: Re-evaluating the Deficit Model of Public Understanding," *Public Understanding of Science* 13(1): 55–75.

Stutman, R.K. and Putname, L.L. (1994) "The Consequences of Language: a Metaphorical Look at the Legalization of Organizations," in S.B. Sitkin and R.J. Bies (eds) *The Legalistic Organization*, pp. 281–302. Thousand Oaks, CA: SAGE.

The American Heritage® Stedman's Medical Dictionary (2002 [2001, 1995]) URL: http://dictionary.reference.com/ (accessed May 2005).

Van Dijck, J. (1998) *Imagenation: Popular Images of Genetics*. New York: New York University Press.

Van Dijck, J. (2000) "The Language and Literature of Life: Popular Metaphors in Genome Research," in J. Marchessault and K. Sawchuk (eds) *Wild Science: Reading Feminism, Medicine and the Media*, pp. 66–79. New York: Routledge.

Van Lente, H. (2000) "Forceful Futures: From Promise to Requirement," in N. Brown, B. Rappert and A. Webster (eds) *Contested Futures: A Sociology of Prospective Techno-science*, pp. 43–64. Aldershot: Ashgate.

Verran, H. (2001) *Science and an African Logic*. Chicago, IL: University of Chicago Press.

Webster's Revised Unabridged Dictionary (1998 [1996]) MICRA, Inc. [online], URL: http://dictionary.reference.com/ (accessed May 2005).

Wouters, P., Beaulieu, A., Park, H.W. and Scharnhorst, A. (2002) "Knowledge Production in the New Digital Networks," 10 June. URL: http://www.niwi.knaw.nl/nerdi/ (accessed May 2005).

Wouters, P., Beaulieu, A., Scharnhorst, A., Hellsten, I., Fry, J., Ratto, M., et al. (forthcoming) "Network, Labor and Epistemic Culture: STS Meets the Challenge of the Internet," in E. Hackett, O. Amsterdamska, M. Lynch and J. Wajcman (eds) *New Handbook of Science, Technology, and Society*. Cambridge, MA: MIT Press.

Wyatt, S. (2000) "Talking about the Future: Metaphors of the Internet," in N. Brown, B. Rappert and A. Webster (eds) *Contested Futures, A Sociology of Prospective Techno-Science*, pp. 109–26. Aldershot: Ashgate.

Zinken, J., Hellsten, I. and Nerlich, B. (2005) "Discourse Metaphors," in R. Frank, R. Dirven, J. Zlatev and T. Ziemke (eds) *Body, Language and Mind. Volume 2: Sociocultural Situatedness* (forthcoming). Berlin and New York: Walter de Gruyter.

Author

Matt Ratto, Ph.D., is currently a Postdoctoral Research Fellow in the Virtual Knowledge Studio for the Humanities and Social Sciences (VKS), at the Royal Netherlands Academy of Arts and Sciences in Amsterdam. He has an academic background in communication and science studies. His research examines technologies as sites where material objects and discursive practices come together to influence the development of communities, disciplines, and other social groups. **Correspondence**: VKS-Royal Netherlands Academy of Arts and Science, Joan Muyskenweg 25, Postbus 95110, 1090 HC Amsterdam, The Netherlands, e-mail: matt.ratto@vks.knaw.nl