Standards and legacies: Pragmatic constraints on a uniform gene nomenclature

Colin Michael Egenberger Halverson

Center for Biomedical Ethics and Society, Vanderbilt University, USA

Center for Bioethics, Indiana University, USA

Abstract

Over the past half-century, there have been concerted efforts to standardize how clinicians and medical researchers refer to genetic material. However, practical and historical impediments thwart this goal. In the current paper I argue that the ontological status of a genetic mutation cannot be cleanly separated from its pragmatic role in therapy. Attempts at standardization fail due to the non-standardized ends to which genetic information is employed, along with historical inertia and unregulated local innovation. These factors prevent rationalistic attempts to 'modernize' what is otherwise trumpeted as the most modern of the medical sciences.

Keywords

standardization, nomenclature, names and naming, classification, medical genetics

Correspondence:

Colin Michael Egenberger Halverson, Indiana University – Center for Bioethics, 410 W 10th St, Indianapolis, IN, 46202, USA

Email: chalver@iu.edu

This is the author's manuscript of the article published in final edited form as:

Halverson, C. M. E. (2019). Standards and legacies: Pragmatic constraints on a uniform gene nomenclature. Social Studies of Science, 49(3), 432–455. https://doi.org/10.1177/0306312719850335

Introduction

It was a beautiful summer afternoon in 2014, and I was indoors, observing a case conference in the Department of Medical Genetics at a Midwestern academic medical center. A young physician, whom I will call Dr. Novak, was presenting the case of a child with an undiagnosed rare disease. I will call the child Martin. Dr. Novak – who had not previously presented at one of these case conferences – appeared nervous as she stood in front of a table of other physicians, genetic counselors, nurses, and medical ethicists, who were gathered as a sort of expert panel in order to weigh in on her case. She later told me she felt a bit anxious talking in front of what she viewed as a very prestigious audience. The team of clinicians, on the other hand, was in good spirits, having spent much of the first five minutes of the conference joking about local politics before eventually transitioning to a more formal temper for Dr. Novak's presentation.

Discussions of patients' cases in the weekly conference were highly generic, progressing from the basic biographic characteristics of the patients, to their clinical history, the results of their genetic testing, and finally those results' potential interpretation and implications for healthcare. While she recounted the history of Martin's case, Dr. Novak clicked through PowerPoint slides projected on a screen at the front of the room. As often happens with rare disease patients, Martin had been seen in Medical Genetics since his birth. His parents struggled to understand what was causing his seemingly random inflammations and fevers, and he had already undergone several genetic tests in pursuit of a diagnosis. As Dr. Novak built the evidence for the case, keeping her final conclusions at a tantalizing remove, anticipation and intrigue grew palpable in the room. True to the genre, she paused after exhibiting the result of the test: a single mutation found in a gene called *CIAS1*.

With suspense at its peak, Dr. Novak finally turned to her interpretation of the case: Mutations in *CIAS1*, she explained to the room, have been linked to inflammatory syndromes; therefore, she and her clinical team had received the laboratory result with excitement. This mutation was indeed the cause of Martin's distress, they believed. However, upon further reflection, Dr. Novak was forced to conclude that this mutation could not be responsible for the patient's symptoms: 'It looked promising at first', she said, but the medical conditions associated with the mutation as presented in the scientific literature were simply not severe enough to explain the patient's acute health problems. That is why, she explained, she chose to bring Martin's case to this particular body of genetics experts. Perhaps they could suggest other paths for her to take in her continued pursuit of a diagnosis for her patient.

After some further discussion of Martin's phenotype and the health of his parents, a pathologist spoke up. 'Have you tested the *NLRP3* gene?' This gene, the pathologist emphasized, could be a prime suspect for mutations causing the specific symptoms expressed by the patient. No, Dr. Novak said as she enthusiastically wrote *NLRP3* on the yellow legal pad in front of her. What laboratory test – she asked, looking up from her notes – might she order that would sequence this particular gene?

The pathologist began to search a list of available tests on her laptop when she was interrupted by a genetic counselor who was also working on Martin's case. 'Wait, yes, we have already tested *NLRP3*. *NLRP3* is the gene formerly known as *CIAS1*. I'm not really sure why it got renamed, … but the official gene name is *NLRP3*.'

The senior pathologist balked for a moment before conceding. She was surprised that she had forgotten the old term. Dr. Novak, still the incumbent novice, apologized. She admitted that she had been unaware of the name change, and repeated that she was used to calling the gene

CIAS1, the name she had learned during her medical training. It would have been burdensome and negligent to order a repetition of the same test, she admitted, which at the time could have cost up to $3,000 and would have returned the same negative result.

In place of the *CIAS1*/*NLRP3* test, a medical geneticist who specialized in rheumatic diseases suggested another possible test. There was, however, noticeably less enthusiasm for this suggestion, as it also seemed only weakly to address the severity of the patient's distress. In the end, the attendees of the case conference determined that the best option for a diagnosis would be to bring the patient into the Department of Medical Genetics, where a specialist would conduct a physical examination and collect a detailed family history in order to better understand what the condition underlying Martin's debilitating symptoms could be.

Renaming *CIAS1*

The gene under discussion in the vignette above has a complicated history. It was first linked to inflammation around the turn of the millennium, as a part of a study on a so-called 'allergy to cold', a disorder that had been suspected since at least the late 1970s to have a genetic component (Wanderer, 1979). For this reason, scientists named the newly discovered gene *CIAS1*, an abbreviation of 'cold-induced autoinflammatory syndrome'. It is common for gene names to reflect a corresponding phenotype or clinical presentation. In fact, the Human Genome Organisation recommends that the name of newly discovered genes describe what they encode (Wain et al., 1999).

The ambiguity in this gene's name emerged in 2008 (Ting et al., 2008). However, a preference for the designation *NLRP3* eventually caused *CIAS1* to shrink from its prominent place in the literature.[1] This preference was based on a general trend away from identifying the

gene (solely) with this rare syndrome and toward identifying it with a broader range of associated disorders, some of which are not characterized by an 'allergy to cold' or even by inflammation at all (Nakanishi et al., 2017). A colleague of mine recently published an article calling for the renaming of another gene for the same reason, and he characterized this article as an attempt to improve the nomenclature's 'accuracy'. As did the *CIAS1* reformers, he saw the divergence of linguistic and conceptual objects to be iconic of a rankling 'irrational[ity]' in medical genetics. The renaming of the *CIAS1* gene was also meant to reflect the gene family – a gene family or superfamily is a set of genes that share similarities, generally in terms of their structure or function – of which it is now considered to be a part. Its new designation was decided through a consensus process, including the submission by email of votes from 'over 100 scientists', held under the auspices of the Human Genome Organisation (Ting et al., 2008: 285).

As is demonstrated in the opening vignette, however, old names – called 'legacy' names in medical genetics – have not ceased to be employed in informal conversation. Because of differences between Dr. Novak's and the pathologist's medical training and in their current engagements in independent speech communities, their norms and awareness of naming practices likewise differ. While this variation causes communicative friction between the two individuals, the fact of legacy nomenclature is well established in the metapragmatic awareness of clinicians in medical genetics. Many of my clinician interlocutors express overt frustration over the many different names for a single gene that circulate at any given time (see Halverson, in press). Major online databases list 'synonyms' of gene names, and in scientific publications, legacy and standard names may be used simultaneously or even interchangeably. For instance, Insalaco et al. (2014) refer to the gene from the vignette as 'CIAS1/NLRP3' in their article's title, but in the body of the article they write equally of 'the *NLRP3* gene' and 'the CIAS1 gene,'

seemingly varying freely between the standard and legacy alternants. Nonetheless, these names technically differ in how they group the gene and what attributes they ascribe to it. Moreover, 'many geneticists and molecular diagnosticians are probably unaware that [such colloquial names] are nonstandard' in the first place (Ogino et al., 2007: 1).

Nomenclature systems in medical genetics demonstrate that the ability of gene names to refer to a specific entity cannot be cleanly divorced from their ability to describe and classify. This multifunctionality, I contend, is a fundamental feature of naming in general. By considering the conflicted history of gene nomenclatures and their thwarted attempts at uniformization and standardization, I argue that gene names are tethered to the evolving conceptualization and utilization of the molecular structures to which they refer. As scientists' and clinicians' understanding of both individual genes and their relationship to the entire genome has changed, so too have their names and their referential potential. This has coupled with increases in the granularity with which genes are scrutinized, an issue to which I return below.

In examining these transformations, I build on previous work that has considered the multiple definitions of 'the Gene.' Using this capitalized orthography, I refer to the abstract concept of an epistemologically and practically uniform *type* of object, the emblem or epitome of the class of all individual genes. Some scholars have described interaction and conflict in conceptualizing 'the Gene' as due to clinical versus molecular perspectives (Fujimura, 1997; Kay, 2000; Keller, 2000), while others have seen it in terms of popular versus scientific perspectives (Hubbard and Wald, 1993; Nelkin and Lindee, 1995). Each of these scholars shows that 'the Gene' holds significantly different social value for different groups of people, and that these different conceptualizations frustrate and condition their exchange of ideas. I similarly problematize the dichotomy between the social and scientific character of 'the Gene'. However,

the present article's focus is specifically on the features of 'the Gene' that are selected as salient to various subfield's local practices of classification. An examination of these different subfields reveals a fundamental disunity in medical genetics, mirroring the disunities described in other scientific fields (e.g. Berg and Mol, 1998; Dupré, 1983; Galison and Stump, 1996; Knorr Cetina, 1999). However, rather than argue that these communities differ in what *counts as* a gene, I demonstrate that conflicts arise in terms of various genes' pragmatic value for the relevant parties. The systematization of genes takes place based on their clinical relevance, driving an articulation of medically agnostic molecular objects with local conventions for determining pathogenicity. Genes are 'truly biomedical entities' in Keating and Cambrosio's (2003: 331) formulation. In my analysis, the concept of 'the Gene' fails to be standardized because of disagreements over which of its features are salient to expert practice, and thus to its classification.

This article also works to contribute to the body of anthropological work on naming (e.g. Bodenhorn and vom Bruck, 2006) and standardization (discussed in more detail below) by investigating a set of scientific nomenclatures that has remained outside social science's investigative gaze. That said, parallel work exists describing the transformation of naming practices in other clinical domains, such as the effects on disease categorization of market forces (Moynihan, 2002; Payer, 1992), of popular and non-expert representations (Brown, 1995; Hogan, 2016; Kleinman, 1988), and of technological developments (Blaxter, 1978; Jutel, 2009; Keating and Cambrosio, 2000; Kendell, 1975; Navon, 2011; Rabeharisoa and Bourret, 2009). These scholars have shown that the linguistic categories of disease names are tethered to – and thus transform alongside – a variety of extra-linguistic forces. I draw on these literatures to argue that

the divergent and evolving social values of 'the Gene' for various biomedical subfields thwart linguistic attempts to standardize its nomenclature.

The current project arose from ethnographic questions that are central to my ongoing fieldwork in medical genetics clinics. However, because of my interest in nomenclature, this article has also come to draw additionally on an extensive engagement with primary sources produced by official and aspirant nomenclature committees. In order to explain the divergent social value of gene names, I turn to three separate case studies of actual attempts to standardize nomenclature systems. Each case demonstrates the effects of historical inertia and specialization on standardization in different ways. I conclude with a discussion of major themes from across each of these case studies and argue that a fundamental disunity in the social value of genes prevents their unification under the auspices of a general nomenclature of 'the Gene'.

Legacy nomenclatures

The gene described in the opening vignette, *CIAS1*, is hardly unique in its having been renamed. Many genes have similarly complicated histories and naming conventions. Schijvenaars and colleagues suggest that 'up to one third of human genes' have more than one name found in the published scientific literature (Schijvenaars et al., 2005: 149). Moreover, some of the same terms have referred to more than one gene. For instance, *'PAP'* has stood as a name for as many five unrelated human genes. These issues of renaming, homonymy, and synonymy cause consternations within medical genetics, a field that brings together 'bioclinical collectives' (Bourret, 2005; Rabeharisoa and Bourret, 2009) – researchers and clinicians from multiple professional generations, united in collaborative research and clinical care.

Active proponents of nomenclature standardization have been campaigning for uniformity for decades. 'One human genome – one language', declares one such manifesto (Shows et al., 1987: 12), imploring members of subfields to acknowledge a single, unified nomenclature. Heterogeneity in these systems has been derided as categorically irrational, with announcements of 'a genetic and molecular basis for a single human gene language without dialects' (Shows et al., 1987). Standardization is in part driven by an ideological insistence that names be singular and coherent because nature itself is singular and coherent, an insistence that 'signs at last correspond to causes, without obstacle, without evasion, without contradiction' (Barthes, 1972: 25). Researchers see standardization as a way to rationalize the nomenclature to reflect the science, to find the 'coherence relations' (Silverstein, 2006; see also Carruthers and Espeland, 1991) between its conceptual and linguistic objects.

Moreover, many researchers consider a general standard to be necessary for the successful implementation of genetic science into medical care. '[I]t is essential that information about mutations and variations in the human genome [is] communicated easily and unequivocally' (Ogino, 2007: 1). Without a general standard, many of its proponents worry, researchers in different communities will not be able to co-refer to unique genes, algorithms will not be able to comb large databases successfully, and clinicians will not be able reliably to derive classificatory information from test reports.

However, heterogeneity is not something that can be readily overcome. In what follows, I argue that the aspirations of medical genetics to a uniform nomenclature of 'the Gene' are frustrated by the disunity of clinical practice coupled with an historical inertia entailed in the process of training in medical genetics. I build on the extensive social scientific literature on standardization (e.g., Bowker and Star, 1999; Brunsson and Jacobsson, 2000; Epstein, 2007; Gal,

2006; Inoue, 2006; Silverstein, 1996; Star and Lampland, 2009; Timmermans and Epstein, 2010) and demonstrate that the implementation of standards can prove troubled – if not fundamentally impossible – under certain social circumstances.

Below, I discuss how medical subfields' unique practical needs affect naming systems, but first I turn to the issue of historical inertia. The problem of legacy nomenclatures arises for two primary reasons: The first is that multiple laboratories may investigate the same gene or hereditary disorder at the same time, naming their shared discoveries separately and more or less simultaneously. These names then circulate side by side in spoken language and scientific literature. The second reason is that researchers aggregate gene families and rename genetic material as they learn more about genes' relationship to the proteins they encode and the diseases to which they contribute, in order to reflect this new classificatory information.

Historical heterogeneity

There are currently at least six distinct nomenclature systems in general use in medical genetics and many more systems in use by specialists in subfields (Schoenbill et al., 2014). In this article I focus on one particular system with its own internal heterogeneity, a system of nomenclature called gene and allele *symbols*. This is the most common way scientists and clinicians name genes in the Department of Medical Genetics, and it includes designations such as *CIAS1* and other, more common names like *APOE*, *BRCA*, and *HOX*.[2] Gene names (called 'symbols' by researchers in the profession) are standardly written in italics in order to distinguish them from the proteins they produce, which – mutatis mutandis – often share the same abbreviations. For instance, the *EGFR* gene (mutations in which have been linked to a variety of cancers) encodes the protein EGFR.

As mentioned above, these 'symbols' have both a descriptive quality – representing the molecular products or the diseases associated with the gene – and a classificatory quality – grouping genetic material into families and superfamilies based on intragroup similarities in structure and function. In this way, symbols represent a classic system of naming (Bodenhorn and vom Bruck, 2006). However, neither of these qualities has been removed from the rapid conceptual transformations characteristic of medical genetics: Numerous genes have been shown to affect many different states of health and disease, and the state most salient to medical care is not always the first discovered – and thus neither the one for which a gene is named. In some instances, the original interpretation of the gene has even changed. For instance, the *SERPIN* gene family received its symbol because it was first thought to encode serine proteinase inhibitors. However, scientists now believe that many of the genes in the family do not act to inhibit gene expression at all, but rather serve in cell storage and transport. Furthermore, gene symbols reflect existing classificatory groups, but they also necessarily face pressure to change as the classification system changes around them. This is precisely what happened in the case of *CIAS1*/*NLRP3* described above.

These difficulties were already apparent to scientists early in the history of genetics research, and by the 1960s standardization had been recommended as a means to resolve them (Povey et al., 2001). In 1977, at the Fourth International Workshop on Human Gene Mapping in Winnipeg, specialists from around the globe sent in nominations by post, and a committee of geneticists was established to oversee the naming of all human genes.[3] The committee served as an institution of 'regulatory objectivity' (Cambrosio et al., 2006), constituting the objects of its oversight through the collective generation and proliferation of naming conventions. Through the consensus of the experts, they attempted to create an international standard out of the disunity of

existing local systems. Two years later, the committee published its first guidelines for unique and meaningful gene symbols (Shows et al., 1979). The committee has since circulated updates to these guidelines every few years.

The Human Genome Organisation's Gene Nomenclature Committee (HGNC) acts as one gatekeeper for ratifying official gene symbols. It oversees many subordinate nomenclature committees for individual gene families. In some cases, though, other organizations officiate the baptism of symbols within their specific subsystems. For example, the World Health Organization manages the naming of certain genes related to autoimmunity (the subject of Section 1.2), and the Centers for Disease Control and Prevention oversees the workgroup regulating the nomenclature of genes that affect drug metabolism (the subject of Section 2.2).

However, the true force of standardization comes from the gatekeeping and circulatory mechanisms of journals, online databases, and regulatory agencies like the FDA, which control both the funding and the publication of genetics research and thus the proliferation (and extinction) of specific linguistic systems. Many major publications require that authors use 'standard compliant' nomenclatures in their articles. These institutions are in a symbiotic relationship with the nomenclature committees in their joint aspirations to linguistic hegemony. The circulation of gene names lives and dies by their uptake in major journals. The Human Genome Organisation boasts that its standard is 'used in all the major secondary databases' and in 'most journals primarily concerned with human genetics' (Povey, 2002: 1). Other nomenclature committees petition journals and large, gatekeeping organizations to accept their recommendations and heed their particular concerns. For instance, researchers working on a naming system for one specific group of genes state at the end of their recommendations, '[w]e hope that this simplified and unified nomenclature will be useful to the [subfield] in general and

adopted by all our colleagues. The nomenclature described herein will be used in all publications by the authors of this manuscript' (Nebert et al., 1987: 10).

Once a name is in circulation, though, it is hard to eradicate it from. As demonstrated in the opening vignette, a legacy name might never be fully excised from (even expert) discourse. As Blaxter notes regarding diagnostic categories, gene names exist and compete within 'a museum of past and present concepts of the nature of disease' (Blaxter, 1978: 10; see also Kendell, 1975; Rabeharisoa and Bourett, 2009).

Naming as description and classification

The linguistic terrain of medical genetics is simultaneously propelled by the centripetal force of standardization and the centrifugal force of specialization. Nomenclature committees for numerous genes and gene families have formed in order to standardize locally heterogeneous naming practices. Several superordinate committees have also come together to unify these standards across subfields and create a nomenclature for 'the Gene' as a generalized model. As noted above, the lack of naming standards is typically seen as irrational, as preventing the successful utilization of genetic information being translated into clinical care. However, the individual research teams exploring these different gene systems come with their own unique goals for the implementation of their discoveries. Moreover, different genes have radically unrelated implications for human health and medical care. Therefore, as gene names work simultaneously to describe and classify their referents, the question becomes: What information do researchers see as essential to convey with their various nomenclature systems? These differences may ultimately prove irreconcilable within the social system of medical genetics, and the failure of the numerous attempts at standardization has resulted not merely from historical

inertia but equally from the fundamental disunity of subfield conceptualizations of 'the Gene' as a clinically meaningful object (cf. Fujimura, 1997; Kay, 2000; Keller, 2000). What constitutes the essence of a given gene cannot be fully disarticulated from what constitutes its social value for the practice of medical genetics.

How, then, do researchers come to conclusions about what is essential about a gene? In the 1987 'Guidelines for human gene nomenclature', the authors caution that gene symbols 'should not attempt to indicate all known information about a gene' (Shows et al., 1987: 12). The amount of information produced in a genetic test can be enormous. Even the information about the sequence itself is too complex to distill it into a single symbol, let alone for that symbol additionally to convey how that genetic information corresponds to a particular disease state (Halverson, in press). As with all forms of classification, the question of naming as a descriptive act is thus necessarily a question of triage (Timmermans et al., 1998).

In certain instances, the only descriptive information that clinicians need from a gene symbol is the gene family itself. Hearing that a mutation has been found in either of the two *BRCA* genes (which are famously linked to breast cancer, among other things) leads clinicians down a very specific path: Certain medical interventions can be recommended regardless of which gene is affected, or how it is specifically affected. 'She has a *BRCA* mutation' is a common and meaningful statement within the Department of Oncology. Some cases require many more (and different) details. For example, knowing that a patient has a mutation in the *CYP* gene superfamily (discussed below) could lead clinicians in a radically divergent array of directions, depending on whether that mutation causes the patient to metabolize a given drug more quickly or more slowly. The utterance '#He has a *CYP* mutation' would be so vague as to be functionally meaningless and would fail Grice's maxims with regard to the quantity,

relevance and perspicuity of the information it provides (Grice, 1975). That is, it would fail to be a meaningful contribution to an ongoing conversation if the speaker had intended to be cooperative.

This tension drives a sort of Borgesian map–territory conundrum (Borges, 1998), where the world of representation threatens to collapse indistinguishably into the world being represented. Borges describes the map of an empire that – for the sake of capturing every minute detail – is the same size as the empire itself – and is, therefore, also completely useless as a form of representation. The primary way geneticists deal with this conundrum is to 'constrain a phenomenon within a particular set of dimensions' (Star and Lampland, 2009: 14) by classifying it within a standardized group. In order to prevent an 'N=∞' style reductio, in which every gene is treated as irreducibly unique, genes and their variations are organized into categories. The categories themselves are given names, each with sufficient descriptive granularity so as to prove meaningful within the context of medical genetics practice. That is, the gene's framework of meaningfulness within the social economy of this international project determines what features are selected for representation in standard nomenclatures and what other features are disregarded.

Gene names are therefore meant to reduce all the available data about a gene to only those most socially salient features. Some nomenclature systems explicitly reject monoglot standardization (Silverstein, 1996), meaning that they refuse to impose uniform mandates; instead, they promote local 'preferences' and methods for 'compliance'. They thereby allow for variation in name structure in order to accommodate the different clinical and scientific ends to which nomenclatures can be employed.

Certain genes and gene families have such radical particularity in terms of their structure and function that nomenclature committees of specialists have been formed to capture their

unique aspects in ways that escape the rationalizing univocality of general standards. As Star and Lampland note, a problem with general standards is that 'one person's well-fitting standard may be another's impossible nightmare' (Star and Lampland, 2009: 5). Drawing on the work of Epstein (2007, 2009), I label these local forms *niche standards*. Insular subfield committees create 'distinct standards, each appropriate to a subgroup' (Epstein, 2009: 50; see also Star and Lampland, 2009 on the nesting of standards) of genetic material. While the committees may ultimately turn to a larger body – such as the Human Genome Organisation – for validation and entry into broad and normative circulation, niche standards create their own systematics that are based on local values and are distinct from those of other nomenclatures.

HLA nomenclature: Vestigial legacies and the Borgesian map

For instance, consider the historical case of *HLA* nomenclature. *HLA* is a gene family important to human health because it encodes human leucocyte antigens (hence the gene family's symbol), which are molecules that help to regulate the immune system. The specific form of these genes varies greatly between individuals. In fact, they are 'amongst the most polymorphic in the human genome', with some 16,000 documented variations (Zheng et al., 2017). This amount of idiosyncratic variability is so great that they are called 'hyper-polymorphic' in the scientific literature. Being able to name the genes accurately is necessary for the successful matching of donors and recipients for cord blood and bone marrow transplantation (called histocompatibility). Thus, discriminating between individual forms of *HLA* holds significant social value for a major sector of modern biomedicine.

In 1964, two separate laboratories discovered the *HLA* gene family independently of each other. Each research team gave the family a name: *HU-1* and *LA*. It took three years and as many

international conferences to arrive at a compromise between these rival names, merging them into one standard symbol, as *HL-A* (Wain et al., 1999: 162). However, much work remained to be done. *HL-A* was merely the 'root symbol' for an entire family of genes – all of which are located within a specific region of chromosome 6, and all of which play similar roles in autoimmunity. Various laboratories had already begun assigning idiosyncratic names to individual genes within the family. Therefore, at the time of the Third Conference on Histocompatibility Testing, specialists elected a Nomenclature Committee that was tasked with overseeing the standardization of the names of these individual genes. On the committee were geneticists, hematologists, transplantation specialists, and immunologists from across Europe and North America.

In 1968, the Committee – under the auspices of the World Health Organization – published the first standardized nomenclature of the *HL-A* genes. The Committee had determined to select names for its standard based on their 'most common use, without regard to priorities or supposed genetic relationships' (WHO Nomenclature Committee 1968: 438). They did, however, standardize the form of these symbols. This produced names of the sort *HL-A3* in place of the old 'dialectal' variation between *To-10*, *LA3,* and *Lc-3*, among others.

The following year, further standardization took place: The niche *HLA* nomenclature system was incorporated into the general guidelines for all gene names, regardless of gene family (Shows et al., 1979: 108). The new *HLA* symbols were of the sort *HLAB7*. The system now more or less followed the major nomenclature conventions for 'the Gene'. This new, general standard removed the hyphen (though this convention later reemerged).[4] The root *HLA* was appended first with a letter, referring to the gene's individual locus within the chromosomal region, and then – after the asterisk – with a number, referring to the specific type of blood serum it encodes.

However, this standardization did not make the information that was encoded in the system uniform with that encoded by other subsystems. That is, the niche standard accounted for blood serum categories that the general standard did not.

In order to update the nomenclature to reflect changing interpretations of the *HLA* system and its relationship to human health, the Committee continued to meet every few years. Since that time, however, new concerns have arisen. For instance, as technology for sequencing genes has improved, it has become increasingly clear that genes vary not in discrete subfamilies but rather along a 'continuum' of diversity (Marsh et al., 2010: 294), what Fujimura and colleagues might call 'clines without classes' (Fujimura et al., 2014). This limits the descriptive meaningfulness of *HLA* symbols, as no symbol can capture enough specificity to discriminate between all the various permutations of nucleotide sequences within the gene family.

It had been announced in 1967 that '15 or more' categories of genetic variation might account for differences in certain types of autoimmune functions (Bach and Amos, 1967: 1506). Between 2005 and 2010 alone, though, more than 2500 new variations within the *HLA* system were given official symbols (Marsh et al., 2010: 294), and by 2015 over twenty distinct *HLA* genes had been described and named. Hence, the map–territory conundrum complicates the act of standardization: If variation is continuous, how can researchers manage to describe and classify it in a parsimonious manner?

By 2017, variation within the system came to be ascribed symbols such as *HLA-DRB1*30:20:12*, with a granularity of representation characteristic of a more baroque age of genetic science and technology. Although symbols are now allowed to have as many as eight digits following the gene root, the capacity to describe regular variation within certain genes is nonetheless 'fast approaching the maximum possible for the current naming convention' (Marsh

et al., 2010). Thus, there is a final limit set by present standards as to how much description can actually be embedded in a single symbol. In order to accommodate this limitation, the Nomenclature Committee has asserted that only empirically demonstrated variation should be given a name. As the researchers in charge of the online Immuno Polymorphism Database state, the system must not become 'a dataset of [all] theoretical combinations' of variation (Robinson et al., 2015: D423), but rather must limit itself to only those variants discovered within existing individuals.

We have now seen many of the concerns and features characteristic of this history of gene nomenclatures. As the science evolves, alternate systems emerge and decay. New science often pushes out old names, but vestigial 'legacies' can persist even – as in the case with *SERPIN* mentioned above – when their descriptive faculty contradicts current understandings of genetic meaningfulness. The case of *HLA* provides particular insight. While it began in a state of nomenclatural heterogeneity,[5] the push for standardization has managed to produce a set of names recognizable as gene symbols, which are uniform on their surface with symbols from other niche standards. There is historical inertia as well, though, in the form of symbols organized by a logic of coherence relations unique to the clinical meaningfulness of the specific molecules that the genes encode. We have also seen a centrifugal force of specialized description that has proliferated the digits after the *HLA* gene root in order to represent the unique forms of 'hyper-polymorphic' diversity specific to that gene family. All the while, this force is counteracted by the limits of the (albeit lax) general standards that the nomenclature is now required to fit. Thus, while *HLA* nomenclature has conceded to assimilate certain structural features of the general nomenclature system, it has nonetheless been constructed to represent the specific social value it holds for the medical subfields most intimately concerned with it.

Contemporary diversity

Certain gene families hold idiosyncratic – perhaps irreducibly idiosyncratic – social value for healthcare. As I have begun to suggest above, the utility of knowing any genetic information in the Department of Medical Genetics is judged based on its ability to inform a patient's care. Among research scientists, the interest is broader but is still generally limited by its relationship to a given disease state (see, e.g., Halverson, in press). That is, what constitutes the meaningful essence of a gene cannot be fully disengaged from what constitutes its pragmatic value for healthcare.

In the following section, I explain how the disunity of the scientific and clinical meaningfulness of genetic variation acts as a key driver of linguistic heterogeneity. This heterogeneity persists in the face of ongoing attempts to rationalize the numerous gene nomenclature systems under a single, general standard. The following two cases of gene family subsystems demonstrate the selection of different features of genetic sequence information in the constitution of nomenclatures as classificatory systems.

The case of KIR: Distinctive descriptions

Some systems resist otherwise generalizing attempts at uniform standards. The method for naming *KIR* genes is one such system. *KIR* genes encode killer-cell immunoglobulin-like receptors.[6] As with *HLA* genes, *KIR* genes are intimately involved in the body's self-defense and self-recognition.[7] The genes are important for the function of natural killer (NK) cells – immune cells that deliver toxins to kill virus-infected and tumor cells. Specifically, *KIR* genes encode receptor molecules that NK cells use to recognize the chemical signature of cells that belong to

the ‘self’ (and therefore should not be killed) versus those that are damaged or are ‘non-self’ (and thus need to be killed).

Again like *HLA* genes, *KIR* genes are highly idiosyncratic in their genetic sequences. This is evolutionarily significant, because it allows for distinctive chemical signals to emerge within individuals. These signals are specific to those individuals and thus allow their immune system to discriminate between a variety of substances, including benign substances that constitute parts of the individuals’ bodies themselves. In fact, their idiosyncrasy is so great that NK cells have ‘an unprecedented phenotypic and functional diversity within and between individuals’ (Rajalingam, 2012: 391).

Unlike *HLA*, however, the *KIR* gene family was much later to the game in terms of nomenclature standardization. The genes did not receive their own nomenclature committee until 2002, under the auspices of the superordinate Human Genome Organisation’s (HUGO) Genome Nomenclature Committee. At the time of its foundation, the *KIR* committee only needed to oversee standardization for 17 genes, but by 2014 more than 600 genes had been described, coding for more than 320 unique KIR protein sequences. Since then, the job of the committee has only continued to grow in scope and complexity.

In the committee’s initial report in 2002, it was decided that the *KIR* genes would receive symbols that reflected ‘the structures of the molecules they encode’ (Marsh, 2003: 80). For an illustration of these structures, see Figure 1. An example of *KIR* notation is *KIR2DL5A*.[8] The notation for the *KIR* gene family obligatorily encodes information about the protein it produces as well as information about the gene itself. The name *KIR2DL5A* notes the gene family with the root symbol *KIR*. This is followed by a numeral (either *2* or *3*) denoting the number of ‘immunoglobulin-like domains’ present in the KIR receptor. (This is followed by the

obligatory letter *D* for domain.) These 'domains' stand outside the NK cell itself and allow the killer cell to bind to other cells – both 'self' and 'non-self.' Knowing the number and form of these domains is significant because it determines what type of cells a particular receptor can and cannot detect.

[FIGURE 1 APPROXIMATELY HERE]

Next, the gene symbol contains either the letter *S*, *L*, or *P*. An *S* means that the receptor encoded by the gene has a short tail anchoring it in the cytoplasm of the NK cell; an *L* means it has a long tail; and a *P* means that the receptor is encoded by a 'pseudogene,' or an imperfect copy of an otherwise functional gene. Receptors with short tails form bonds with cells that they recognize as 'non-self,' thus triggering the NK cell to release toxins that destroy these foreign bodies. Receptors with long tails, on the other hand, form bonds with cells chemically marked as 'self,' which inhibit the 'natural killer' function of their hosts.

The *KIR* symbol concludes with a final digit. This represents the unique number of the individual gene encoding the receptor within the broader family of genes. As mentioned above, there were originally only 17 recognized *KIR* genes, but now there are hundreds reported in the literature. These genes are spread out across the various subfamilies, which are denoted by the preceding two segments of the gene symbol, and only the very last morphological segment of the symbol describes the unique, individual gene.

Thus, *KIR* symbols obligatorily encode the strikingly distinctive – and clinically salient – features of the molecules that the genes produce. The nomenclature describes its referents with enough granularity to demonstrate what makes individual genes 'functionally distinct' from one another (Rajalingam, 2012: 393). Unlike the *BRCA* genes, where the clinically salient information is conveyed by the symbol's root, much more information is relevant for researchers discussing *KIR* genes. While the *KIR* nomenclature committee operates under (and thus within)

the bounds of the general standards set down by the Human Genome Organisation, the characteristics selected for representation in the symbol directly reflect the gene's unique relevance to human health and disease. In this way, *KIR* nomenclature is a clear instance of niche standardization, in which the universalizing aspirations of naming 'the Gene' are constrained by the idiosyncrasies of the specific gene family itself.

The case of star alleles: Non-standard values

Another example of a gene system that has highly unique properties and whose variation has major medical significance is the superfamily of pharmacogenetic genes. They are so called because they affect a patient's response to different pharmaceuticals. Individual variation in these genes can determine whether a certain drug has a positive or negative effect on a patient, and whether that patient needs a lower or a higher dosage than is standardly prescribed. There are likely over 1000 human genes that condition drug response (Kalman et al., 2016), and even just the single gene *CYP2D6* is involved in the metabolism of more than seventy commonly prescribed drugs (Nebert and Jorge-Nebert, 2002). A particular variant of the *CYP2D6* gene (called **5*) causes adverse reactions to many of these drugs, whereas another variant (**1X13*) causes the body to metabolize them so quickly that the patient derives no therapeutic benefit from the standard prescription. Variation within this gene can have more than a 20-fold effect on what amount constitutes an appropriate drug dosage for an individual patient (Ingelman-Sundberg, 2002).

The efficacy of at least 150 FDA-approved drugs is conditioned by pharmacogenetic variation (Eichelbaum et al., 2006). Knowledge of a patient's genetic sequence in these genes can therefore provide very salient – even essential – information for the prescription of many

medications. Until this point, I have only discussed nomenclature at the level of the gene or gene family. However, in the case of pharmacogenetic variation, particularly meaningful is what is found below the level of the gene, in the idiosyncratic variation within the genes themselves. Genes can be quite large – with thousands, sometimes even millions of nucleotides constituting a single gene's sequence – and mutations can occur at many different loci within these sequences. Such difference between individual *tokens* of an abstract geno*type* is called allelic variation, and it is the focus of many attempts at standardizing pharmacogenetic nomenclatures.

[FIGURE 2 APPROXIMATELY HERE]

Efforts to produce standard naming systems at the scale of the gene can be traced back to the 1960s. However, the technological means of visualizing genes with the granularity needed to classify them based on their molecular sequence was not feasible at that time. It was not until decades later – in the early 1990s – that geneticists at Baylor College of Medicine began publishing well-attended suggestions for a standardized means of notating mutations within genes (e.g., Beaudet and Tsui, 1993). In fact, as the Human Genome Project neared completion, scientists began publicly expressing their surprise at the 'tremendous degree' of the allelic heterogeneity of individual genes, which was only then becoming visible through new developments in sequencing methods – but which was arriving on researchers' computers in the form of 'a staggering "information overload"' (Nebert, 2000: 279). By that point, there were already numerous independent research groups working on different genes, and many of these groups had developed their own niche methods for naming variation below the level of the gene. This meant that allelic variation had been named 'in a fairly random fashion', which researchers complained was leading to a 'chaotic situation' in the global scheme of pharmacogenetic

nomenclatures (Garte, 2001: 1305). Standardization was the obvious next step for many researchers who had been involved in similar efforts for other gene systems.

What emerged from the pursuit of standardization was a system my interlocutors in the Department of Medical Genetics call the 'star allele' system. The 'star' refers to the asterisk that the system uses to separate the gene symbol from the symbol representing allelic variation within the given gene. Star alleles, therefore, are names that allow scientists and clinicians to distinguish between medically meaningful variation within the broader system of pharmacogenetic genes. The use of the asterisk for this purpose was already in place in the first version of the international standard for gene nomenclatures (Shows et al., 1979), but when my interlocutors refer to 'star alleles' they are specifically talking about pharmacogenetic variants in genes associated with drug processing. That is, they are not talking about the difference between states of Tay Sachs disease (*HEXA*2*), gangliosidosis (*HEXA*3*), and statistically normal health (*HEXA*1*). None of these variants has anything to do with pharmacogenetics but they nonetheless all use the standard 'star' system for annotating allelic differences within a specific gene.

The historical heterogeneity of pharmacogenetic systems has certainly been a problem for attempts at standardization. Early researchers promoting the star allele system insisted that people 'speak the same language' and 'avoid "home-made" allelic designations that would only confuse the nomenclature system and the scientific literature' (Ingelman-Sundberg et al., 2001: 1307). They touted the 'universal acceptance and usage of one single system for all [pharmacogenetic] gene alleles' as their 'critically important goal' (Garte, 2001: 1305). However, these attempts have been frustrated further by the specificity of the various genes' idiosyncratic values for human health and disease. In fact, these values vary so greatly within the

pharmacogenetic system that the system has segmented into a number of relatively independent subsystems. Some of the different pharmacogenetic gene families have their own committees overseeing unique, niche naming conventions for their allelic variation. Some genes have no nomenclature committees at all, with allelic variation still being named without a standard for reference. Some subsystems are not actively updated to keep pace with developments in research and understanding, and others have numerous, competing nomenclatures that have not been unified under an agreed-upon standard (Kalman et al., 2016).

The most prominent of the pharmacogenetic gene families are referred to in HGNC standard notation as the *CYP* genes, an abbreviation of cytochrome P450, the specific drug-metabolizing enzyme that they encode. Like the *HLA* gene complex discussed above, the role of the *CYP* genes was discovered separately by three different laboratories (Robarge et al., 2007). In 1999, the complexity of the genes' relationship to clinical therapy had already been recognized, and the Cytochrome P450 (*CYP*) Allele Nomenclature Committee was convened to create a standard that would reflect the medical relevance of variation in this specific subset of genes.

We can take *CYP2D6*3* as an example of star allele nomenclature as developed by the Nomenclature Committee. *CYP2D6* is a particular gene in the larger *CYP* gene superfamily. The first numeral after the gene superfamily root (*CYP*) refers to a lower level of classification: *CYP* families 1 through 3 are involved in metabolizing foreign substances like drugs. Those with higher numeric designations are involved in metabolizing substances produced within the body. The following letter and number refer to the specific gene within that family. The asterisk – or 'star' of the star alleles – separates the symbol for the gene as an abstract concept from the actual allelic variant of the gene that is found within a particular individual. The variant represented is

the third variant of the gene to be reported to cause a difference in the function of the CYP protein that the gene encodes. There has historically been conflict over how to number these variants. One method has simply been to increase the number each time the description of a new allele is published. Another method has been to number alleles based on their physical position within the gene. A third method has to been to require their numbering to correspond to the 'most severe consequences' of the variant for a patient's health (Sim and Ingelman-Sundberg, 2010: 281). Under the star allele system, such waves of standardization attempts exist simultaneously as a sort of palimpsest, 'resulting in the gene names … being out of sequence' (Nelson, 2004: 2) when judged from different standards' unique vantages. The star allele system leaves this 'irrationality' unresolved, as the accumulative result of historical inertia. The 'myriads of interlocking conventions' (Cambrosio et al., 2006: 197) each get embedded within the single (arbitrary) standard.

Thus, in the *CYP* genes – and in pharmacogenetic genes more generally – we see the same historical, descriptive, and classificational forces at work *below* the scale of the gene as well. Historically heterogeneous and niche systems continue to operate within different pharmacogenetic gene families and persist as legacies within otherwise uniform nomenclatures. The star allele system has not managed to rationalize the numbering system of *CYP* variation, a fact that underscores the competitive (rather than self-evident) nature of classification in determining what information is most salient to the taxonomic task (see also Dupré, 1999). Even within this particular gene superfamily – with its highly specific relationship to medical genetics – researchers debate whether to include in the name descriptive information about the variant's relationship to an individual's health (phenotypic data about drug metabolism) or its structure within the gene (genotypic data about its position). The object of such naming thus walks the line

between more traditional clinical entities and what Navon (2011) calls ‘genomic designation’, oscillating between its more distal and more proximal products. Researchers continue to debate which features of the genes to select as relevant for the pragmatic act of classification. Standardization has so far failed to create a fully uniform and rationalized system of nomenclature because of the heterogeneous social value such names hold for their various users.

Discussion

A number of related pragmatic issues have constrained the repeated efforts to create general standards for naming genes and genetic variation. I have argued that these efforts are fundamentally thwarted by the inseparability of the meaningful essence of ‘the Gene’ from individual genes’ pragmatic value for healthcare and scientific discovery. ‘The Gene’ as an object of contemporary clinical practice exists as the articulation of a medically agnostic molecular structure and local value systems used to determine pathogenicity. In fact, it instantiates these two ‘material and discursive arrangements’ (Keating and Cambrosio, 2003: 332) as a thoroughly biomedical entity.

Standardization further suffers under the material fact of the boundaries between the relevant speech communities and bioclinical collectives within medical genetics itself. Archaic forms – so-called legacy nomenclatures – that do not meet current general standards are still with some regularity heard in case conferences or seen in print, because their siloed users are not always aware of changes in guidelines and recommendations. Thus, new and old criteria of classification exist side by side in a sort of palimpsest (cf. Rabeharisoa and Bourret, 2009). Since nomenclature systems strive to convey scientifically accurate information, they evolve alongside current beliefs about the structure and function of human genetic material (cf. Hogan, 2016). For

clinicians who have been removed from the lockstep of research, modifying their own linguistic practices appropriately can prove challenging, as was seen in the opening vignette.

In accommodating this disunity, many official nomenclature guidelines explicitly allow for or even promote the supplementary use of non-standard forms. Some legacy forms are so common that they appear in written materials as parentheticals behind the contemporary, standard-compliant name. This atavism is due to the assumption that for some readers, the former will be familiar while the latter will be unrecognizable. For instance, Berwouts and colleagues have suggested in the official journal of the Human Genome Variation Society (HGVS) that '[r]eports should include a description of identified sequence variants in both HGVS and traditional nomenclature … [which] ensures legibility and compatibility with the existing literature' (Berwouts et al., 2011: 3). In other instances, the standards themselves prevent the impulses behind them to rationalize the coherence relations within their own nomenclatures and instead elect archaic, non-systematized forms as official gene symbols. This was the case with the disorderly numbered *HLA* genes, which have been allowed to remain 'out of sequence' (Nelson 2004: 2), as vestiges within a contemporary niche standard.

Moreover, the success of standardization hinges on the uniformity of its object (Espeland and Stevens, 1998). If that uniformity does not already exist, it can be performatively constituted through standardization (e.g., Berg and Mol, 1998; Mol, 2002). However, as we have seen in the case of gene nomenclature, the process can be highly contested in this regard. Even in decidedly rationalistic systems such as those of medical genetics, the pragmatic disunity in the social value of genetic information leads to a disunity in the conceptualization of the abstract notion of 'the Gene' within medical genetics. While regulatory objectivity still functions within local niche nomenclature committees, it fails to establish general conventions at a broader scale.

Standardization ultimately fails to regiment the description of specific genes to parallel the spectrum of meaningfulness imputed to a universal model of 'the Gene'. It proves unable to 'eliminate the pragmatic features that constitute … [their] distinctive character' (Gal, 2006: 171) and social utility for their users. Nomenclatures do not simply carve nature at its joints, but neither do they allow a 'Borgesian map' to emerge. Only those aspects of genetic material that have a perceived relevance to particular medical genetic goals are preserved in the descriptive function of the gene symbol. *KIR* gene names encode information relevant to the molecular structures they produce, while *CYP* allele names differentiate between the clinical effects the variants have on drug metabolism.

The performative act of naming is famously tethered to institutions and institutional avatars that have been authorized to carry out such socially meaningful transformations (Althusser, 1971; Austin, 1975). 'To understand biomedicine is to understand its regulation', as Keating and Cambrosio (2003: 334) have put it. This, however, only further highlights the disunity within the process of translating scientific knowledge into clinical action. The right to 'baptize' a given gene name is a function of 'the sociolinguistic division of labor' (Silverstein, 2006; see also Putnam, 1975). The social legitimation of the Human Genome Organisation, for instance, only extends so far. General knowledge is not considered sufficient within the subfields of specialists, who constitute their own communities, with their own local gatekeepers authorized to officiate the entry of 'dialectal' words into niche nomenclatures. In this way, the history of gene nomenclatures is characteristic of broader trends in medical language. The clinic exists at the intersection of multiple speech communities, each with its own norms, and this causes even highly rationalistic scientific jargon to admit certain norms from other social groups (e.g., Hogan, 2016; Jutel, 2011).

It has not been my intention to suggest that researchers across bioclinical collectives vary in their ontological conceptions of what constitutes a gene at the molecular level. Disagreement is not over what comprises the essential or 'necessary' (Putnam, 1975) features of the object, but over the features salient to its sub-classification. In this way, my argument differs from those of many philosophers of science, who describe paradigmatic conflicts between multiple ontological conceptions of 'the Gene' (see Kitcher, 1982, and citations therein). I hope to have demonstrated that a gene in medical genetics is much more than simply its nucleotide sequence. The information gleaned from a genetic test couples with the collective knowledge available to link it with a disease state. This provides an overabundance of potential characteristics to be encoded in the gene's official symbol. It is the variability in the ascription of social value to particular characteristics and the erasure of others that constitute the disunity in the conceptualization of 'the Gene', which in turn prevents general standardization.

Coda

There are two major constraints on standardization in genetic nomenclature: the force of historical inertia, which promotes legacy nomenclatures; and specialization, which selects divergent aspects of 'the Gene' for representation in its official symbols. The former of these difficulties persists because of standardization's limited reach across speech communities: physician generations, disciplinary silos, laboratory and geographic insularity, etc. The latter of these difficulties persists, I have argued, because of a fundamental disunity in the conception and use of 'the Gene' as a socially meaningful object of medical and scientific intervention. The diversity of genes' products and functions leads to an irreducible plurality of potential classifications and descriptions, and thus of names. This map–territory conundrum forces

researchers to 'make assumptions about when and whether such variability is medically relevant' (Timmermans and Epstein, 2010: 37). In cases such as the *CYP* genes, the two forces interact, with historical waves of preference for particular methods of classification getting concatenated within a single, unsystematized standard.

As classificatory names, gene symbols attempt to capture enough salient descriptive information to be useful within the context of medical genetics. Yet they must simultaneously attempt to be general or generalizable enough to be able to represent every stretch of genetic variation that can be objectified as a socially recognizable scientific object.[9] In this way, we can draw informative parallels to what Gal (2006) describes as the 'contradiction' implicit in standardizing national languages: the pretention to being simultaneously *authentic* to an object's 'valuable, idiosyncratic properties' (p. 166) and *universal* (or 'general' as I have said here) in terms of the system's neutral and unconditional applicability. Again in somewhat parallel fashion to Gal's case, this 'contradiction' is suppressed through a play of scale. The ideological drive for standardization ends up creating numerous spin-off, niche standards that specialists develop for individual genes or gene families. This heterogeneity of nomenclatures was the primordial state of genetic science and persists in contemporary medical genetics, which at its core conceptualizes itself as a constantly emerging and evolving practice.

Nearly every update to the official, general standard for human gene nomenclatures includes a discussion about the future of genetics, scientific development and the constantly shifting nature of the objects to which its symbols are meant to refer. Even as these updates promote and lay down guidelines for a particular, uniform standard, they underscore the tenuousness of such a project. Just as '[a]ll pathology is subjective with regard to tomorrow' (Canguilhem 1978[1966]: 125), so too is its classification. In fact, in the most recent iteration of

the guidelines for pharmacogenetic nomenclature (Kalman et al., 2016), the authors propose the replacement of the contemporary 'star allele' system with a new, more universal system that applies to allelic variation in all gene families, overseen by a superordinate committee. Even here, Kalman and her coauthors note that the new system might be 'tedious or hard to understand' for researchers who have come of professional age with the star nomenclature system (Kalman et al., 2016: 18) and they therefore recommend the use of legacy nomenclature – 'familiar or alternative names' (p. 12) – alongside the new standard.

The new guidelines begin with a call to be prepared for change: ontological change and epistemological change, and, thus, linguistic change as well. For them, what makes the science behind gene naming scientific in the first place is that it is always changing. Today's standard always implicitly holds within itself the potential to become tomorrow's legacy.

Acknowledgments

Versions of this paper have benefited greatly from the close readings of Lisa Bastarache, Catherine Egenberger, Jenny Hua, Costas Nakassis, Anna Weichselbraun, and Lily Yeh, the discussion from attendees of the Semiotics Workshop at the University of Chicago, the anonymous reviewers for this journal, and Sergio Sismondo. I would also like to thank all of my clinician and scientist interlocutors for their generous consultations on this project. All errors, of course, remain my own.

Funding

This work was supported by the Wenner Gren Foundation (Gr. 8831) and the Hanna Holborn Gray Mellon Fellowship.

References

Althusser L (1971) Ideology and ideological state apparatuses: Notes towards an investigation. In: Brewster B (ed) *On Ideology*. New York: Verso, 85-126.

Austin J (1975[1962]) *How to Do Things with Words*. Cambridge: Harvard University Press.

Bach F and Amos D (1967) Hu-1: Major histocompatibility in man. *Science* 156(3781): 1506-1508.

Barthes R (1972) *Mythologies*. New York: Hill & Wang.

Beaudet A and Tsui L (1993) A suggested nomenclature for designating mutations. *Human Mutation* 2(4): 245-248.

Berg M and A Mol (1998) *Differences in Medicine: Unraveling Practices, Techniques, and Bodies*. Durham: Duke University Press.

Berwouts S, Morris M, Girodon E, Schwarz M, Stuhrmann M, and Dequeker E (2011) Mutation nomenclature in practice: Findings and recommendations from the Cystic Fibrosis External Quality Assessment Scheme. *Human Mutation* 32(11): 1197-1203.

Blaxter M (1978) Diagnosis as category and process: The case of alcoholism. *Social Science and Medicine* 12: 9-17.

Bodenhorn B and vom Bruck G (2006) 'Entangled histories': An introduction to the anthropology of names and naming. In: vom Bruck G and Bodenhorn B (eds) *The Anthropology of Names and Naming*. Cambridge: Cambridge University Press, 1-30.

Borges J (1998) On exactitude in science. In: Hurley A (ed) *Collected Fictions*. New York: Penguin, 325-326.

Bourret P (2005) BRCA patients and clinical collectives: New configurations of action in cancer genetics practices. *Social Studies of Science* 35(1): 41-68.

Bowker G and Star S (1999) *Sorting Things Out*. Cambridge: MIT Press.

Brown P (1995) Naming and framing: The social construction of diagnosis and illness. *Journal of Health and Social Behavior* 35(Extra issue): 34-52.

Brunsson N and Jacobsson B (eds) (2000) *A World of Standards*. Oxford: Oxford University Press.

Cambrosio A, Keating P, Schlich T and Weisz G (2006) Regulatory objectivity and the generation and management of evidence in medicine. *Social Science & Medicine* 63: 189-199.

Canguilhem G (1978[1966]) *On the Normal and the Pathological*. Boston: D. Reidel Publishing Company.

Carruthers B and Espeland W (1991) Accounting for rationality: Double-entry bookkeeping and the emergence of economic rationality. *American Journal of Sociology* 97(1): 31-69

Dupré J (1983) The disunity of science. *Mind* 92(367): 321-346.

Dupré J (1999) Are whales fish? In: Medin D and Atran S (eds) *Folkbiology*. Cambridge: MIT Press, 461-476.

Eichelbaum M, Ingelman-Sundberg M, and Evans W (2006) Pharmacogenomics and individualized drug therapy. *Annual Review of Medicine* 57: 119-137.

Epstein S (2007) *Inclusion: The Politics of Difference in Medical Research*. Chicago: University of Chicago Press.

Epstein S (2009) Beyond the standard human? In: Lampland M and Star S (eds) *Standards and Their Stories*. Ithaca: Cornell University Press, 35-53.

Espeland W and Stevens M (1998) Commensuration as a social process. *Annual Review of Sociology* 24: 313-343.

Fujimura J (1997) *Crafting Science: A Sociohistory of the Quest for the Genetics of Cancer*. Cambridge: Harvard University Press.

Fujimura J, Bolnick D, Rajagopalan R et al. (2014) Clines without classes: How to make sense of human variation. *Sociological Theory* 32(3): 208-227.

Gal S (2006) Contradictions of standard language in Europe: Implications for the study of practices and publics. *Social Anthropology* 14(2): 163-181.

Galison P and Stump D (1996) *The Disunity of Science*. Redwood City: Stanford University Press.

Garte S, Boffetta P, Caporaso N, and Vineis P (2001) Metabolic gene allele nomenclature. *Cancer Epidemiology, Biomarkers & Prevention* 10(12): 1305-1306.

Grice H (1975) Logic and conversation. In: Cole P and Morgan J (eds) *Studies in Syntax and Semantics III: Speech Acts*. New York: Academic Press, 183-198.

Halverson C (in press) Evidence and expertise in genetic nomenclatures. *Journal of Linguistic Anthropology*.

Hogan A (2016) Medical eponyms: Patient advocates, professional interests and the persistence of honorary naming. *Social History of Medicine* 29(3): 534-556.

Hubbard R and Wald E (1993) *Exploding the Gene Myth*. Boston: Beacon Press.

Ingelman-Sundberg M, Oscarson M et al. (2001) Human cytochrome P-450 (*CYP*) genes: A web page for the nomenclature of alleles. *Cancer Epidemiology, Biomarkers & Prevention* 10(12): 1307-1308.

Ingelman-Sundberg M (2002) Polymorphism of cytochrome P450 and xenobiotic toxicity. *Toxicology* 27(181-182): 447-452.

Insalaco A, Prencipe G, Buonomo P et al. (2014) A novel mutation in the CIAS1/NLRP3 gene associated with an unexpected phenotype of cryopyrin-associated periodic syndromes. *Clinical and Experimental Rheumatology* 32: 123-125.

Inoue M (2006) Standardization. In: Brown K (ed) *Encyclopedia of Language & Linguistics*. Oxford: Elsevier, 121-127.

Jha S and Ting P (2009) Inflamosome-Associated Nucleotide-Binding Domain, Leucine-Rich Repeat Proteins and Inflammatory Diseases. *Journal of Immunology* 183(12): 7623-7629.

Jutel A (2009) Sociology of diagnosis: A preliminary review. *Sociology of Health and Illness* 31(2): 278-299.

Kalman L, Agúndez J, Appell ML et al. (2016) Pharmacogenetic allele nomenclature: International workgroup recommendations for test result reporting. *Clinical Pharmacology & Therapeutics* 99(2): 172-185.

Kay L (2000) *Who Wrote the Book of Life?* Redwood City: Stanford University Press.

Keating P and Cambrosio A (2000) Real compared to what? Diagnosing leukemias and lymphomas. In: Lock M, Young A and Cambrosio A (eds) *Living and Working with the New Medical Technologies: Intersections of Inquiry*. Cambridge: Cambridge University Press, 103-134.

Keating P and Cambrosio A (2003) *Biomedical Platforms: Realigning the Normal and the Pathological in Late-Twentieth-Century Medicine.* Cambridge: MIT Press.

Keller EF (2000) *The Century of the Gene*. Cambridge: Harvard University Press.

Kendell R (1975) The concept of disease and its implications for psychiatry. *British Journal of Psychiatry* 127(4): 305-315.

Kitcher P (1982) Genes. *British Journal of the Philosophy of Science* 33(4): 337-359.

Kleinman A (1988) *The Illness Narratives: Suffering, Healing, and the Human Condition*. New York: Basic Books.

Knorr Cetina K (1999) *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge: Harvard University Press.

Marsh SGE, Parham P, Dupont B et al. (2003) Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. *Tissue Antigens* 62(1): 79-86.

Marsh SGE, Albert ED, Bodmer WF et al. (2010) Nomenclature for factors of the *HLA* system, 2010. *Tissue Antigens* 75(4): 291-455.

Mol A (2002) *The Body Multiple*. Durham: Duke University Press.

Moynihan R, Heath I and Henry D (2002) Selling sickness: The pharmaceutical industry and disease mongering. *British Medical Journal* 324(7342): 886-891.

Nakanishi H, Kawashimi Y, Kurima K et al. (2017) *NLRP3* mutation and cochlear autoinflammation cause syndromic and nonsyndromic hearing loss DFNA34 responsive to anakinra therapy. *PNAS* 114(37): E7766-E7775.

Navon D (2011) Genomic designation: How genetics can delineate new, phenotypically diffuse medical categories. *Social Studies of Science* 41(2): 203-226.

Nebert D, Adesnik M, Coon M et al. (1987) The P450 gene superfamily: Recommended nomenclature. *DNA* 6(1): 1-11.

Nebert D (2000) Suggestions for the nomenclature of human alleles: Relevance to ecogenetics, pharmacogenetics and molecular epidemiology. *Pharmacogenetics* 10(4): 279-290.

Nebert D and Jorge-Nebert L (2002) Pharmacogenetics and pharmacogenomics. In: Rimoin D, Connor J, Pyeritz R and Korf B (eds) *Emery & Rimoin's Principles and Practice of Medical Genetics*. London: Churchill Livingstone, 590-631.

Nelkin D and Lindee S (1995) *The DNA Mystique: The Gene as a Cultural Icon*. Ann Arbor: University of Michigan Press.

Nelson D, Zeldin D, Hoffman S et al. (2004) Comparison of cytochrome P450 (*CYP*) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics* 14(1): 1-18.

Ogino S, Gulley M, den Dunnen J and Wilson R (2007) Standard mutation nomenclature in molecular diagnostics: Practical and educational challenges. *Journal of Molecular Diagnostics* 9(1): 1-6.

Payer L (1992) *Disease-Mongers: How Doctors, Drug Companies, and Insurers Are Making You Feel Sick.* New York: John Wiley and Sons, Inc.

Povey S, Lovering R, Bruford E et al. (2001) The HUGO Gene Nomenclature Committee (HGNC). *Human Genetics* 109(6): 678-680.

Povey S (2002) Smelling of roses? *Genome Biology* 3(6): 1-2.

Putnam H (1975) The meaning of 'meaning'. In: *Philosophical Papers*. New York: Cambridge University Press, 215–271.

Rabeharisoa V and Bourret B (2009) Staging and weighting evidence in biomedicine: Comparing clinical practices in cancer genetics and psychiatric genetics. *Social Studies of Science* 39(5): 691–715.

Rajalingam R (2012) Overview of the killer cell immunoglobulin-like receptor system. In: Christiansen T, Talt B (eds) *Immunogenetics: Methods and Applications in Clinical Practice*. New York: Humana Press, 391-414.

Robarge J, Li L, Desta Z et al. (2007) The star-allele nomenclature: Retooling for translational genomics. *Clinical Pharmacology & Therapeutics* 82(3): 244-248.

Robinson J, Halliwell JA, Hayhurst JD et al. (2015) The IPD and IMGT/HLA database: Allele variant databases. *Nucleic Acids Research* 43(D1): D423-D431.

Schijvenaars B, Mons B, Weeber M et al. (2005) Thesaurus-based disambiguation of gene symbols. *BMC Bioinformatics* 6(1): 149.

Schoenbill K, Fost N, Tachinardi U, Mendonca E (2014) Genetic data and electronic health records: A discussion of ethical, logistical and technological considerations. *Journal of the American Medical Informatics Association* 21(1): 171-180.

Shows T, Alper CA, Bootsma D et al. (1979) International System for Human Gene Nomenclature (1979). *Cytogenetics and Cell Genetics* 25: 96-116.

Shows T, McAlpine PJ, Boucheix C et al. (1987). Guidelines for human gene nomenclature. An international system for human gene nomenclature (ISGN, 1987). *Cytogenetics and Cell Genetics* 46(1-4): 11-28.

Silverstein M (1996) Monoglot 'standard' in America: Standardization and metaphors of linguistic hegemony. In: Brenneis D (ed) *The Matrix of Language: Contemporary Linguistic Anthropology*. Boulder: Westview, 284-306.

Silverstein M (2006) Old wine, new ethnographic lexicography. *Annual Review of Anthropology* 35: 481-496.

Sim S and Ingelman-Sundberg M (2010) The Human Cytochrome P450 (*CYP*) Allele Nomenclature website: A peer-reviewed database of CYP variants and their associated effects. *Human Genomics* 4(4): 278-281.

Star S and Lampland M (2009) Reckoning with standards. In: Lampland M and Star S (eds) *Standards and Their Stories*. Ithaca: Cornell University Press, 3-24.

Tamames J and Valencia A (2006) The success (or not) of HUGO nomenclature. *Genome Biology* 7(5): 402.

Timmermans S, Bowker G and Star SL (1998) The architecture of difference: Visibility, controllability, and comparability in building a nursing intervention classification. In: M Berg and A Mol (eds) *Differences in Medicine: Unraveling Practices, Techniques and Bodies*. Durham: Duke University Press, 202-225.

Timmermans S and Epstein S (2010) A world of standards but not a standard world: Toward a sociology of standards and standardization. *Annual Review of Sociology* 36: 69-89.

Ting J, Lovering R, Alnemri E et al. (2008) The NLR gene family: A standard nomenclature. *Immunity* 28(3): 285-287.

Wain H, White J and Povey S (1999) The changing challenges of nomenclature. *Cytogenetics and Cell Genetics* 86(2): 162-164.

Wain H, Bruford E, Lovering R et al. (2002) Guidelines for human gene nomenclature. *Genomics* 79(4): 464-470.

Wanderer A (1979) An 'allergy' to cold. *Hospital Practice* 14(6): 136-137.

WHO Nomenclature Committee (1968) Nomenclature for factors of the *HL-A* system. *Bulletin of the World Health Organization* 39: 483-486.

Zheng J, Xie J, Sun Y and Jian Y (2017) Characterization of the novel *HLA-DRB1*11:01:20* allele. *HLA Immune Response Genetics* 89(4): 258-259.

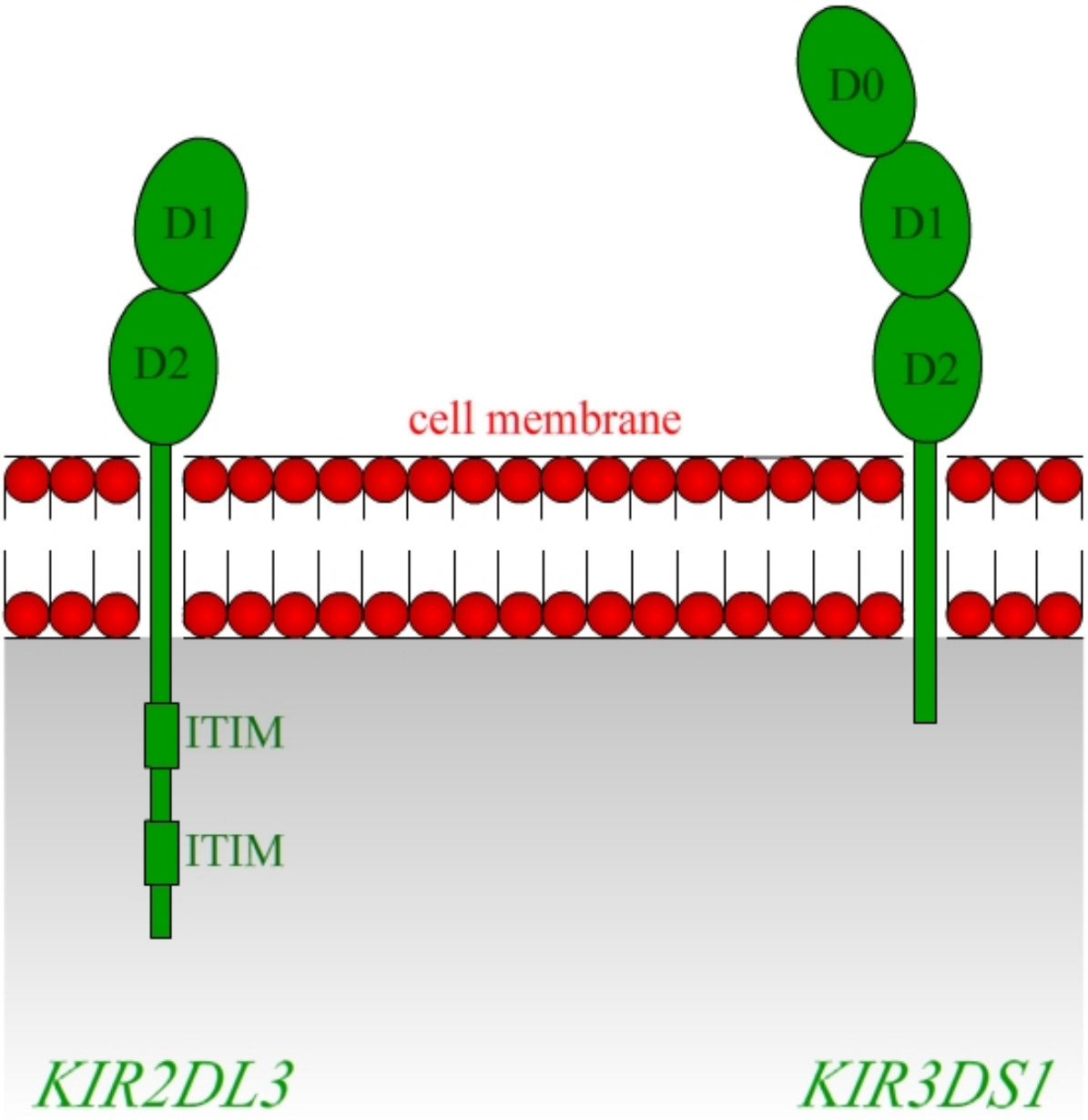

Figure 1. Two KIR receptors.

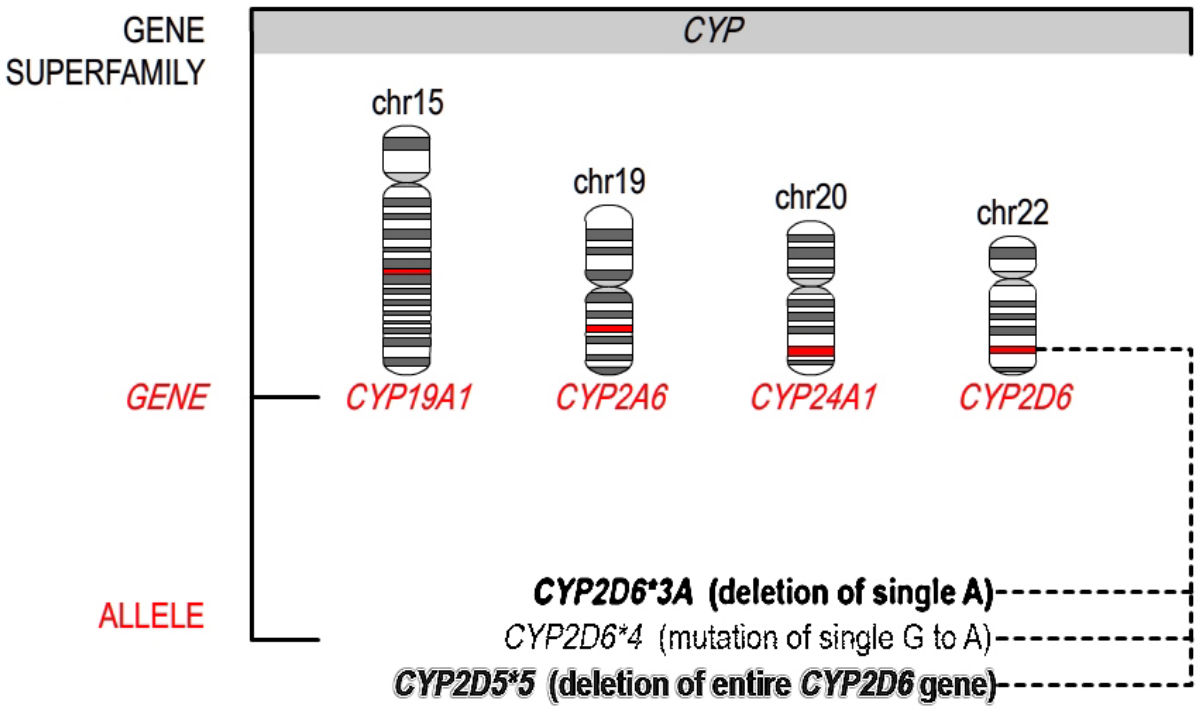

Figure 2. Four genes within the *CYP* superfamily, and three allelic variants of one of those genes.

Notes

[1] *NLRP3* is a name more general than *CIAS1* and stands for NLR (a family of genes, all of which having a nucleotide-binding domain that produces protein segments called leucine-rich repeats) and pyrin-like protein (the downstream product of the gene). It is not the purpose of this article to explain what any of this means, but an interested reader may turn to Jha and Ting (2009) for further information.

[2] *APOE* (which encodes apolipoprotein E) is linked most prominently to Alzheimer disease, *BRCA* to breast cancer, and *HOX* genes (a type of homeobox gene) are involved in embryo development.

[3] The relationship between human and animal gene nomenclatures is interesting in its own right – and with serious implications for medical questions of disease modeling and ontological questions of homology – but it is too complex and too far afield to discuss here.

[4] *HLA* gene symbols are one of only three subsystems that are allowed to use punctuation in their gene symbols (for reasons of historical inertia, see Wain et al., 2002).

[5] The *HLA* system is not unique in this regard, though its history is particularly well documented. Another example is the chemokine lymphotactin gene (related to inflammation and immunological response), which was reported nearly simultaneously in 1995 by independent laboratories on three different continents. The gene quickly amassed multiple different local

symbols – *ATAC*, *LTN*, *LPTN*, and *SCM1* – before the HGNC could agree on a (completely unrelated) official name (*XCL1*) many years later (Tamames and Valencia, 2006).

[6] The 'I' in this name originally stood for 'inhibitory' rather than 'immunoglobulin-like'.

[7] It is worth noting that the *HLA* and *KIR* systems do not merely have biochemical traits in common. Some researchers have been key personnel in the efforts to standardize the nomenclature systems of both groups as well. The geneticist Steven GE Marsh, for instance, has been first author on a number of publications pertaining to both immunological gene families.

[8] Like many symbols, once an author has established what gene family is being discussed – by initially using the full gene symbol – the root can be elided, with only the gene-specific portion of the symbol being used. For example, *KIR3DP1* may be abbreviated as *3DP1*, as in 'The KIR2DP1 and 3DP1 … pseudogenes' (Rajalingam, 2012: 400).

[9] As with the case of *HLA* genes, genetic variation may ultimately be seen as a 'continuum' (Marsh et al., 2010: 294), but it is nonetheless made discrete through naming.