Next Generation Sequencing in Clinical Practice: The Translation and Interpretation of Genomic Data in the Diagnostic Laboratory and the Clinic

Dr Chris Goldsworthy

Institute for Science Innovation and Society, University of Oxford

Abstract

As Next Generation Sequencing (NGS) becomes common practice in the diagnostic genetic setting it becomes important to examine the mediated and co-constructed impact NGS technologies have upon diagnostic laboratory and clinical practice. This is explored here in relation to how the mutual situated co-configuration of the technology and the users redefines understandings of certainty and uncertainty associated with genetic and genomic data. I show how a maintained pragmatic understanding that genomic data and technology is in a constant state of uncertainty, termed acceptable contingency, enables laboratory workers to be able to adapt the technology and the data produced to meet their situated needs. I suggest that this process of *reflexive standardisation* is only possible due to a deep technical understanding of the technology as well as an embeddedness within the space in which the technology and genomic data-bases are developed.

Introduction

Next Generation Sequencing (NGS) is increasingly making its way into diagnostic practice from the research setting. In doing so, NGS is designed to supplant existing sequencing technologies as the primary genetic and genomic diagnostic sequencing technology in the diagnostic laboratory. NGS serves as a broad category of modern sequencing technologies defined by their ability to sequence DNA and RNA at a much higher rate than and lower cost than their primary predecessor, Sanger sequencing. The NGS methodology employed at the diagnostic laboratory reported on here is Illumina Sequencing, this is the most popular commercial sequencing methodology due to its relatively low cost when compared to competitors, as well as the production of a high volume of data per run (around 15Gb of data per run using Illumina MiSeq (Illumina, 2017). Despite the reported benefits of the transfer to NGS from Sanger sequencing in the diagnostic genetics setting the translation of this technology from the research into the diagnostic setting is not a linear process. Instead it is considered a complex, dynamic process, defined by compromise and situated adaption to a mutually co-constituted (Shostak, 2005) ends.

Previous research has examines the clinical use of genetic testing (Latimer, 2013) and more recently Next Generation Sequencing (Timmermans, 2015), however this paper specifically examines accounts of how the increased quantity of genomic and genetic data is negotiated in the NHS laboratory and clinic. The interface and sharing of information between the laboratory and the clinic will also be examined, focussing upon how professionals share situated notions of confidence in data that are not consistent with defined practical standards. In doing so I will put into question the 'trust in numbers' hypothesis put forward by Theodore Porter (1995), instead asserting that big genomic data sets are viewed by clinicians and laboratory scientists as in a permanent state of acceptable contingency. By taking the perspective of those who contextualise, translate and interpret genomic data sets in practice in the clinic and the lab, I argue that clinicians and laboratory scientists are not 'data-dopes'. This terminology is in-keeping with MacKenzie and Spears (2012) definition 'model-dope' (p.7), which is itself a derivative of Garfinkel's (1967) 'cultural-dope' (p. 68). Both use the 'dope' as a rhetorical device designed to create an unreflexive other who naively reproduce the conditions of the culture in which they live or accept the outputs of a particular economic model without reflecting upon the model from which the output was derived. It is not clear that data-dopes exist in the clinical or laboratory setting, however Levin (2014) has observed an instance of what could be considered a data-dope in her study of a metabolomics research group. Levin gives the example of a biochemist who presented data to the metabolomics research group without reflecting upon the technology used to produce the data. The biochemist simply accepted the black-boxed explanation given by the developer of the technology, in doing so the biochemist failed to take into consideration the irreducible complexity of the metabolism held to be true by the metabolomics research group. I will show that laboratory and clinical professionals are aware of the effect increasingly large datasets have on their local practice and their profession as a whole, and that they do not naively accept big data outputs as wholly objective, or indeed useful. Through this I will show how clinicians and laboratory geneticists perform 'reflexive standardisation' (Timmermans, 2015) based on the perceived discrepancies between the technology and clinical utility. In examining these issues this article will consist of two sections. The first will discuss how developments in genomic science and the use of big data have shaped clinicians and laboratory scientists practice in the NHS, this section will draw upon scholarship showing how visualisations of data are active in the construction of the reality they are designed to represent (Levin, 2014; Lynch and Woolgar, 1990; Woolgar, 1991). The second section will discuss how these same clinicians and scientists mobilise and manipulate technologies as situated to their local practices and needs. I describe how clinical professionals are not passive recipients of NGS technologies, instead in line with Oudshoorn (2016), clinicians and scientists are positioned as actively engaged in re-configuring the technology and data output based upon

their situated needs. Taken together, this creates socio-technical network of local co-construction (Oudshoorn and Pinch, 2013) in which the technology shapes practice, however it does so within an existing system and thus is adapted to the existing practical and political infrastructure.

Methods

This article draws on data gathered as part of a research project examining the organisation of the space in which genetic testing for inherited cardiac conditions is undertaken within the UK health care system. During this time, many diagnostic services across the UK were in the process of transitioning aspects of their services from Sanger sequencing to Next Generation Sequencing. At this time (April 2014 – December 2015), many centres were actively negotiating how to deal with the impact of adopting this new technology on the organisation of the laboratory as well as the relations between clinicians, professionals in the genetics laboratory and the big genetic data produced by the NGS technology. The data were gathered from multiple sites across the UK (N=8) all of which specialised in the diagnosis and treatment of ICC's, either in the clinic, the laboratory or both. Across these sites, interviews were conducted with a variety of professional groups, including; clinical geneticists, laboratory geneticists, specialist cardiologists, genetics counsellors and specialist arrhythmia nurses. I also observed practices at one large NHS laboratory which serves as an international referral centre for cardiac genetic testing. In addition to this, I observed 8 cardiac genetics multi-disciplinary team meetings (MDT's) at a large tertiary hospital. These meetings were attended by 8-10 healthcare professionals who were actively involved in the management of patients, either from the discipline of cardiology or clinical genetics. The primary function of this MDT was to decide whether a patient should receive genetic testing or not as well as which test they should receive. These meetings also often discussed how to deal with genetic test results in terms of the impact these results would have on the diagnosis and treatment of the patient and their family. A multi-site approach was taken to gain insight into the way that different professional systems manage the use of genetic testing within their existing infrastructures.

Big Genomic Data in the NHS

Genomic Data in the NHS

The benefits of Whole Genome Sequencing (WGS) for research purposes in the field of rare diseases have been made clear, the primary benefit is that it enables the identification of novel gene mutations associated with rare diseases (Genomics England, 2015b). This has been the

lynch pin of the rare disease mission of the 100,000 Genomes Project, positioned as 'Gene Discovery' (Genomics England, 2015a). Gene Discovery has been rationalised by Genomics England as a way of assisting in the interpretation of variants of uncertain significance (VUS), thus they present a reduction in uncertainty as a product of an increased quantity of data, committing to gather data from 50,000 people with rare diseases and their families. This prioritisation of quantity as a means of alleviating uncertainty has come to represent objectivity in modern times (Porter, 1995; Hacking, 1990). This subscribes to the common big data narrative in which 'Raw data' are a source of exploration and discovery, as an untapped, objective resource. In genomic science the larger the data set the more representative it is seen to be. An example of this is seen in the way researchers justify their findings or indeed criticise others. Elijah Behr et al. (2015) successfully used a big data approach to overpower a smaller research study (Hu, et al 2014) which claimed an association with mutations on the Sodium channel gene *SCN10A* and the rare cardiac channelopathy Brugada Syndrome. Behr and his colleagues main area of contention with this study was their small sample of 200 matched controls[1]. By conducting their own larger study which used the UK10K project database as matched controls, Behr et al. (2015) was able assert authority over the concerned claim, refuting Hu's research. The UK10K database is considered quantitatively superior to Hu's control in 2 ways; firstly it is a database of genome or exome sequences, whereas Hu's control was of the single gene *SCN10A* and secondly the database consists of over 7000 individuals' exome or genome sequence data to control against (The UK10K Consortium, 2015). The sheer weight of data was mobilised to increase the legitimacy of Behr's argument:

> 'Our data suggest that rare variation in *SCN10A*, particularly in *SCN5A* mutation negative cases, is unlikely to cause BrS (Brugada Syndrome). This contrasts markedly with a recent paper by Hu et al. which identified *SCN10A* mutations in 16.7% of 150 BrS probands... This difference in yield cannot be explained from a technical perspective as conventional Sanger sequencing was undertaken in both studies. Of note, Hu et al studied only 200 ethnically matched controls without finding any missense rare variants. This is unsurprising as ESP and UK10K data both show that there are plenty of rare variants in controls but larger numbers are required to detect them reliably... Thus

[1] Frequency/absence of mutations in the general population, through the use of matched controls, is 1 of 7 criteria for attributing significance to a variation in a gene, promoting it from the category of variant to mutation. The other criteria are:

- Reports in the literature;
- Co-segregation in (preferably) trios;
- Conservation (of the amino acid across species);
- Functional domains (does the function of the amino acid reflect the phenotype);
- Presence in unrelated individuals with the phenotype;
- And, Functional studies.

> our 'enriched' cohort and more stringent 'mutation' definition are more likely to be representative of the yield of novel rare SCN10A variants in BrS.'
>
> (Behr et al 2015, p. 16-17)

This ultimately resulted in mutations in *SCN10A* to be considered as of unknown significance by the research and clinical community in line with Behr's et al. (2015) findings (Conversation with cardiologist)[2].

While in genetic research papers, sample/dataset size is correlated with certainty[3], in the health service genomic data is used very differently, owing to a different functional priority. Clinicians contest that newer technologies and larger gene panels actually increase uncertainty, as opposed to the decrease in uncertainty promised by research:

> 'You have to be careful... the more genes you have the more uncertain you are.'
>
> (Clinical Geneticist 2)

In the clinical setting size does not equal certainty. It is important to note here that before complex NGS testing, where services were only testing at most a few genes per condition, clinicians were not more certain. However, clinicians were more certain of the validity of the gene mutations they did test for (Timmermans et al. 2016), during this time there were far more negative results, which does not preclude the possibility of a pathogenic genetic mutation. This is a distinction not in the scale of uncertainty but the type of uncertainty. NGS has heralded uncertainty associated with the validity of gene mutations found, where as pre-NGS testing was dominated by uncertainty associated with the ability of a test to capture gene mutations in the first place.

Negative genetic test results are welcomed in the clinical setting in particular circumstances, this is explained as the main difference between the use of genomics in research and the NHS. In the research setting the introduction of NGS technologies has had dramatic effects in terms of gene discovery for rare conditions as a Clinical Geneticist with a research interest described:

---

[2] This represents a shift in genomics research in which only those with adequate resources and connections can now produce valid scientific claims, creating class systems located around centres of excellence.

[3] This is not to say that in the researchers naively analyse large genomic data sets as objective fields of 'raw' data which can be unreflexively mined. But it is to say that the strength of a research paper is increasingly measured by the size of the sample or data set drawn from. As this research has not directly observed or conducted interviews in relation to the negotiation of big genomic data-bases within research projects, I cannot make any claims beyond the way in which this data is mobilised as a sign of authority.

> 'The biggest change for us... has been this adoption of NGS. From a research side of things one of the things we have been doing is trying to identify new genes that cause rare diseases... We have had a lot of experience at doing that and had a lot of success in identifying the genetic basis of a whole range of diseases.'
>
> (Clinical Geneticist 2)

Where as in the clinic there is:

> 'A huge shift in emphasis, in science they look for the most variance and this is exciting, where as in the clinical setting no variance is a good thing, it means the patient does not possess the pathogenic variant.'
>
> (Laboratory Geneticist)

This Geneticist is referring to the 'good-going' (MDT3) gene mutations, these are the gene mutations that are 'known' to be responsible for a large proportion of the conditions. If the patient is not found to have one of these gene mutations, the result is positive. This outlook is particularly useful when considering the testing of phenotype negative individuals. These may be family members who have experienced a sudden death (which is the most serious consequence of ICC's), or these may be family members of gene positive patients, in which case a negative genetic test can rule out a patient for further interventions as they 'do not possess the family gene'(Clinical Geneticist, 3). Thus genomic data, as is used in research, is not that useful in clinical practice for the majority of referrals, which are the patients with clearly visible phenotypes, who are found to have the 'good-going' mutations. These cases have become routinized, and it is more clinically useful and efficient to use traditional Sanger sequencing techniques focusing on a specific gene or genes. The clinic and the laboratory have a high level of certainty that these mutations cause the phenotype - these are the black and white cases. Where NGS does become useful is in the grey cases those with complex phenotypes who are not suspected to possess the most common gene mutations.

How Genomic Data has Shaped Genetics in the NHS

Contrary to the significant shifts in emphasis between Genomic research and the clinical genetics in the NHS, Big Genomic science has not only shaped the practices of NHS genetics services but also the way in which clinicians and scientists view their profession and future developments in the field. This extends debates instigated by Steve Woolgar (1990) in that genomic data sets and NGS technology attempt to denote its prescribed uses and users. The first way in which genomic data sets configure their users is by denoting who the users are, the

format of the data they access as well as how they use the data. For example, data outputs from the 100,000 Genomes Project will only be made available through Genomics England secure server and permitted users (members of healthcare or research organisations) have to agree to data access agreements (Genomics England, 2015a). 'Raw data' cannot be exported from this database so users are restricted to the format prescribed by Genomics England. This is equally the case for the reference genome which was produced at the end of the Human Genome Project. As such for a researcher or a clinician to compare their data to that held within this database they must format their data in the same way[4]. For the research community, legitimacy hinges on validation through large whole genome control samples as shown earlier. This has been extended to health service genetics, that are no longer solely reliant on reports in the literature as a way of validating uncertain findings:

> 'We look at population cohorts who have had whole genome or whole exome screening and we use them very much to try and determine whether a variant is pathogenic or not. So if you have found it at a high frequency you can have more confidence that it is not a pathogenic mutation and we use that information in our interpretations.'
>
> (Laboratory Geneticist).

This kind of work has emigrated from the research community and is becoming increasingly important in the clinical diagnosis of ICC's, in which much of the research associated with variants are highly contestable. In many respects, genomic datasets define how clinicians and NHS laboratories assign validity to a mutation; the image of objectivity (Daston and Galison, 1992) has been engineered in genetic practice to be represented by validation through quantity of data outweighing the value of the corpus of scientific research in the area. Although clinical geneticists and cardiologists in the field of ICC's do not access genomic data sets directly, the ethic of validity based upon genomic population controls penetrates the value that they attribute to variants. A Cardiologist discussing this referred to the value he attributes to a generally accepted association between mutations on *SCN5A* and Brugada syndrome:

> '...for Brugada syndrome it [genetic testing] is totally unhelpful, the original study [(Chen et al. 1998)] was not even controlled properly... the guidelines say it might be helpful, but not in my experience.'
>
> (Cardiologist 2)

[4] The model for representing genome assemblies currently advocated by the Genome Reference Consortium which governs the reference genome is GRCh38. This format has been adopted across the world as to be compatible with the reference genome.

Such an insight was gained by this clinician through embedded experience with genomic data and genomic research which utilise much larger data sets. Prior to the prevalence of the use of genomic data in research the imperfections of Chen's et al. (1998) study were overlooked on the premise that an imperfect answer was better than no answer. Genomic data sets have enabled reflection upon the usefulness of previously commonly held beliefs. Thus, Genomic data sets have transformed the standards (Timmermans and Berg, 1997) of clinical genetics practice by redefining the standard of validity and certainty.

Data analysis and interpretation in the NHS laboratory is the most labour intensive task for the genetic scientist following the introduction of NGS, a by-product of the increased throughput and capacity offered by this technology is an exponential increase in the size of the data output. With gene panels of up to 72 genes in the field of cardiac genetics, the output can be considered big data in its own respect, in that the yield is far more data than an individual can analyse. This has dramatically changed the workflow of the NHS genetics laboratory. Technicians and genetic scientists specialized in particular conditions and followed patients through the lab before the introduction of NGS, however with increasing throughput and automation of key processes in the laboratory, workers increasingly specialize in one stage of the process, adopting post-fordist modes of production, similar to that reported in studies of large biological research laboratories (Stevens, 2011; Hilgartner, 2013; Mackenzie et al 2016). The impact of this on the genetic scientist is that they spend little to no time in the 'wet lab':

> 'We never come to the lab any more... I don't even know how to use the next gen machine. We just do the analysis, techs do the first line we do the rest... I liked the old ways it was like 'real science'. That's not saying that the job is any less interesting it is interesting in different ways now we can do so much more now.
>
> (Laboratory Observations)

The scale of data yielded from NGS necessitates this dramatic change in working practices in the laboratory setting to enable the management of the data.

Bioinformatics pipelines have been established to assist in this transition, filtering the data into a manageable quantity. Much of the bioinformatics software used in the clinical setting is outsourced to the technology provider. The software used at the laboratory I visited showed all the base pairs screened across the top of the screen, with all covered base pairs automatically highlighted by the program (there are often gaps in sequence data due to allelic dropout[5]). The

[5] Allelic dropout describes the process by which copies of alleles fail to be amplified by the PCR (Polymerase Chain Reaction), this results in missing data in the readout (Wang et al, 2012).

program then focussed the gaze of the genetic scientist to the 'known' variants in the sequence, these being variations from the control sequence data as well as 'known' pathogenic variants' associated with the phenotype. The interpretation by the genetic scientist is only undertaken on the remaining highlighted sections, her analytic gaze is focussed by the software to specific base pairs, as much of the process as possible is externally automated. Genetic Scientists are limited in this respect as the processes of data manipulation which lead to the output that they receive is 'black-boxed'.

One of the greatest successes of the translation of NGS technology to the clinical setting is the acceptance of the inherent uncertainty associated with many of the findings (Calvert, 2008). Uncertainty is nothing new for clinical cardiac genetics, ICC's are complex; there are issues of incomplete penetrance and there is a high proportion of mutations considered highly pathogenic in the general populations (See SCN5A). For one of the better understood cardiac genetic conditions, Long QT Syndrome, the 5 most common genes associated with this condition are only thought to explain 68% of cases (Splawski et al, 2000). However, the advent of NGS in the clinic altered the narrative of uncertainty, in that it strengthening ideas of temporality of uncertainty. When using NGS in the clinic, finding a variant does not always equate to attributing validity there is much more liminality, in which patients are between diagnostic categories. Much of this is based on the implicit understanding that knowledge relating to the genetic nature of ICC's is far from complete. A narrative of development from certainty to uncertainty has been presented using two analogies, with pre-NGS testing compared to 'picking the low hanging fruit', or 'catching the fish at the surface'[6]. This represents the notion that before NGS, geneticists were only able to identify the 'good-going' gene mutations. However as these mutations are found it becomes more difficult to explain the phenotypes of those without the good-going mutations. This often results in finding Variants of Uncertain Significance (VUS), the negotiation of which is a major problem when 'you throw your net wide looking at as many genes as possible' (Clinical Geneticist 1) as is done when using NGS:

> 'I always counsel about variants of uncertain significance, and we still have patients who come back and say; 'Well I don't understand you have found the gene change so why can't you just do the blood test'. Then you have to cover it again and say: 'We did discuss this possibility in which we would find a variant that we weren't certain about and that we wouldn't offer to people who weren't affected. We don't know enough about the

[6] These analogies were given by a representative from Genomics England on two occasions: in 2014 at the annual AICC meeting and in 2015 at the Cardiff International Cardiac Genetics Symposium. They were given as a way of presenting the rare disease gene identification agenda of the 100,000 genomes project, suggesting that the project could help identify the harder to reach mutations due to the use of whole genome sequencing.

> people with the condition who also have this spelling mistake and therefore we don't have enough proof that it's the cause.'
>
> (Genetic Counsellor 3)

Much of this kind of narrative rests on presumed limitations in the technology and assumes that finding quantitatively more patients with the mutation would increase certainty:

> '...as technology develops we can do more tests. So it's making the family understand that they are not necessarily missing something but that it is a limitation of the technology.'
>
> (Clinical Geneticist 1)

This is the first aspect of acceptable contingency, by this I mean the acceptance by clinicians and laboratory geneticists of the 'promissory narrative' (Stephens, 2013) provided by the genomic research community even though they understand the inherent uncertainties that the technology brings. Although genomic science does not constitute what Hedgecoe (2004. P.515) terms a 'promissory science', in that it is well established and has had huge implications and applications in the research and clinical setting. This extends Merton's (1942) notion of 'organised skepticism' in that I suggest no closure of this skeptisism, instead suggesting a pragmatic acceptance of the inconsistencies and uncertainties of a technology following critical scrutiny. This does not reduce the 'hope and hype' (Marris, 2005. p.1) narrative, promising a greater understanding of the nature of genotype-phenotype correlations through the exploration of ever-expanding genomic databases. This was not a difficult 'vision' (Martin, 2001) to sell to clinicians in this field as the majority of the specialist geneticists in cardiac genetics are also research active.

The clinical acceptance of the rhetoric of temporary uncertainty is now embedded in clinical practice, whereby clinical geneticists and genetic counsellors present findings as contingent upon scientific developments and more data. The primary mechanism by which this contingent uncertainty is performed, can be seen when a clinician is faced with a negative test result for a phenotype positive patient. In these cases, clinicians often wait for developments in the research setting to translate to the clinical setting and then re-tests the patient or family:

> 'We have a number of families where we have tested them right from the outset with our half a gene, and then the 3 gene, then the 4, the 13 and now the 16 gene screen. We still haven't found anything, it makes us think right 'have we missed something because the technology before hasn't allowed us to detect it and this technology has also missed

> the same thing'. We don't like having those families and they would be the first ones that we go back to and say: 'Oh by the way we have got a new test, fancy putting them on because we really want to find something'. We have got a family at the moment that has had a DCM [Dilated Cardiomyopathy] test, it's been one of our families for years, and we have just found a lamin A mutation and that completely explains the phenotype.... We offer the best that we can at the time and keep up to date with making changes.' (Laboratory Geneticist)

The emergent nature of our ability to interpret genetic findings also allows the possibility of 'Red Herrings' (Clinical Geneticist 1) in which new information sheds doubt upon the validity of a particular mutation and clinicians then have to re-categorise patients on this basis, certainty of findings is rarely presented by clinicians.

The narrative thus far has been of how the change to high-throughput genetic technologies such as NGS has shaped modern clinical genetic practice in the NHS. The big biological data revolution and the resultant bioinformatics emigration into the NHS laboratory has engineered the 'correct' (Levin, 2014) way to analyse genetic data as well as defining that which is valid for interpretation, and the form of the output produced by the Laboratory (Timmermans and Berg, 2003). This shift has also changed the material practices and tempo of the NHS genetics laboratory, with automated systems running 24 hours a day and technicians managing the technology. Perhaps the greatest impact the shift to NGS has made on cardiac genetic clinical practice has been the enculturation of the idea that genomic data is not only vast but emergent, flexible and dynamic (Rose, 2013), which has had the effect of realigning notions of certainty. Although of course VUS's were around long before NGS, they were rarer in the clinical setting. The introduction of NGS and the potential of WGS has heralded the possibility of a deluge of mutations, this creates a problem of quantitatively more uncertainty as well as the problem shifting from having to negotiate whether the VUS is significant to having to negotiate which VUS is significant and being able to say why.

Contextualisation, Transformation and Manipulation of Genomic Data

Situated data for situated practice

The previous section discussed the impact of developments in genomic technologies in the NHS clinical genetics setting. However presenting the relationship between the clinic and the technology in this way assumes that clinicians and laboratory scientists are uncritical consumers of the technology, it assumes they are 'data-dopes'. The potential of the consumer of

genetic technology to become a data-dope has been observed in previous research by Bourret et al. (2011), in which they reported on diagnostic tools used to identify genetic tumour signatures. These tools had been marketed as prognostic and predictive by their creators and utilised algorithms which provided results that the clinicians themselves could neither derive, confirm, nor validate independently, due to the lack of transparency as to the means by which the result is constructed. However, Bourret et al. (2011) reported strong oppositions to technologies which excluded clinical autonomy, to the extent that the FDA created a separate category to regulate such devices. This is not to reduce the impact of NGS in the NHS lab and the clinic, however the translation of research to the clinic cannot be represented as neutral or one sided where clinicians and scientists alike naively accept the data as objective. In fact, this process is a strongly negotiated one, underpinned by an understanding of genomic science and genomic datasets as mutable, dynamic and fallible and that clinicians and laboratory geneticists are able to mobilize the data beyond their pre-defined configurations. Clinicians and scientists are intrinsically aware that genomic data sets are both 'cooked' and 'noisy', as opposed to raw and clean. This section will discuss the effect of this in the NHS clinic and laboratory. This is part of acceptable contingency, as suggested earlier when clinical geneticists were shown to accept NGS in spite of the heightened uncertainty it yielded. This being for the most part owing to the acceptance of the 'promissory narrative' (Stephens, 2013) given by genomic science but also due to the perceived improvement in clinical utility that NGS yields. Utility in this respect is relative, based upon a definition from clinical chemistry, which describes a key aspect of utility as the extent to which the test (or technology) affects 'health outcomes relative to the current best alternative' (Bossuyt et al 2012, p. 1).

Drawing on Stefan Timmermans (2015) work on the negotiation of standards by clinicians in the clinical genetics setting, this section will discuss how clinicians and laboratory geneticists employ a process of 'reflexive standardization' (Timmermans, 2015. p.79). Examining how clinicians and laboratory geneticists ground the standards, in this case being the standard ways of negotiating genomic datasets and technologies within their situated practice, creating 'situated data'.

Situating the Data

This argument is supported by a strong pedigree of previous research, much of which centres around the idea that local clinical experience outweighs scientific consensus when making clinical decisions. Bosk was an early proponent, stating:

> 'in the case of discrepant opinions, arguments based on clinical expertise override those based on scientific evidence.'

(Bosk, 1979. p.85)

Latimer et al. (2006) and her colleagues similarly found this in relation to the value attributed to negative genetic testing showing that local clinical experience with the patient outweighs the findings of a genetic test and the presence of the genetic condition is not discounted. Hedgecoe (2008) went as far as to show how when considering the usefulness of APOE4 testing in Alzheimer's disease patients, clinicians would disregard genetic findings in favour of clinical findings and experience. Although this has been discussed in MDT's, it is generally avoided by clinicians, by focusing the technology only on what they see as clinically relevant prior to conducting the test. Although only certain clinicians who have a genetics lab within their NHS trust have the freedom to select which genes are analysed, ICC clinics across England and Wales have the autonomy to select which gene panel they use. This is firstly because there is a significant difference in the constitution of panels for individual and grouped conditions between laboratories, and secondly because each panel test within each centre is made up of different genes. This is most notable where a patient presents with a non-typical phenotype as was the case at an arrhythmia MDT:

'Cardiologist: I think we should definitely look for Danon

Clinical Geneticist: The thing is, LAMP2 is on the HCM (Hypertrophic Cardiomyopathy) panel. So we could look for others using this to cover more things like sarcomere and I don't think our funding stream would support just a DCM (Dilated Cardiomyopathy) panel.

Paediatric Cardiologist: So we are looking at the extended HCM panel. Well she did initially present with increased LV (Left-ventricular) mass so...'

(MDT8)

The patient in this case was a young girl presenting with DCM a weakening and thinning of the heart muscle, however the clinicians suspected from her pedigree that she might have Danon disease, a rare disease presenting with either DCM or HCM. Guidelines dictate that when a patient presents with DCM they should receive targeted genetic testing (Ackerman et al, 2011), however targeted testing in the guidance does not cover LAMP2. Clinicians in this case asserted autonomy, not in their judgement over the validity of the test but of its ability to capture the nuances of the patient's phenotype, thus increasing their chances, from their experienced perspective, of finding a clinically useful mutation. This furthers discussions instigated by Latimer et al. (2006), who position the clinic as an active site of knowledge production. They show how genetic categorisation is achieved based upon the interaction between clinical

experience and genetic test results, they position genetic categorisation as a flexible accomplishment at the point at which biomedical collectives (Rabeharisoa and Bourret, 2009) converge. Similarly, Stivers and Timmermans (2016) show how genetic diagnostic uncertainty is actively negotiated in the genetics clinic between clinicians and patients. While these studies emphasise clinical autonomy in the negotiation of genetic test results, the above suggests that clinical negotiation begins at the point of selecting the which panel of genes to test.

In the laboratory, clinical needs and preferences are taken into account in the data filtering and analysis stage rather than at the point the test is undertaken. For every sample received for a cardiomyopathy panel or an arrhythmia panel test, the laboratory will run their entire panel and filter the findings so to only analyse the genes in the locations associated with the phenotype as given by the clinician:

> 'In the New Year we will have a new panel and essentially we will run it on every patient that we are requested a cardiac genetic test on. We run all 72 genes but we only analyse those dependant on the phenotype, we categorise them into different conditions, long qt, hypertrophic cardiomyopathy... So it might be for ARVC that we screen 6 of the genes out of the 72, the data is there for the 72 but we only look at 6. That speeds up the analysis, it means we have got a single pipeline for the testing but in terms of generating the result, it means that it's quicker and we are not looking at data that potentially isn't informative and that would delay and actually maybe even complicate things. If however new phenotypic information comes to light we could always come back and look at that data.'
>
> (Clinical Geneticist 2)

Much of this is cost related. It costs the lab the same amount to test for 6 genes as it does 72 by virtue of using the same technology and the same amount of reagent. The main variable in cost based on size of the panel is accumulated at the analysis stage. This is equally the case when clinical exome sequencing is undertaken:

> 'I see a number of patients with rare conditions where we have got a good idea of what the potential genes could be but there is no testing available for those in a routine diagnostic lab anywhere. The only way really to integrate those is by using an exome and we run the whole exome but we would only pick out certain genes that we were interested in looking at.'
>
> (Clinical Geneticist 2)

This filtering process is a process of practicality. In the NHS setting, genetic testing is not undertaken on an exploratory basis – rather, genetic testing for ICC's is commissioned on the basis that around 50% of test results will be gene positive (Cardiologist 3). However, this is also a sign of the culture of risk aversion in clinical genetics (Timmermans, 2015). Laboratories will not send out a report based on a mutation they are not sure about, in terms of association with the reported phenotype. It does not make sense for them to analyse a gene they would not report on. This is also a practical data management technique, as clinical laboratories simply do not have the time to analyse each base pair in a whole exome, or even each variant. Calling this whole exome sequencing brings up interesting questions about the relationship between the technology and the analyst. The whole exome has been sequenced so the process has been technically achieved, however the data output in its 'raw' form is simply stored away, so an analyst may never see it. Pragmatically this process has its advantages in that if no pathogenic variants are found in the genes analysed then the scope of the investigation can be broadened without having to go through the technical process of taking a blood sample, extracting the DNA and re-sequencing other parts of the patients exome. The process of storing sequence information is both economic and efficient, particularly where there is uncertainty associated with where the mutation is likely to be. It also serves to alleviate the ethical problem of incidental findings in that this information is only assigned meaning following human interpretation (for a discussion of the ethical issues of disclosing WES results in the clinical setting see Hallowell, et al, 2015).

Representational Uncertainty

Data is further filtered by the bioinformatician who works with the geneticists in the lab. By having a bioinformatician as part of the laboratory team, 'pipelines' and software can be developed to represent the particular needs of both the laboratory and the population they serve, ensuring that the geneticists get locally appropriate data. This is valued by the particular centre I visited due to their past involvement in WGS, in which they were made aware of the not inconsequential variation in the genomes they sequenced compared to the large data sets such as the Reference Human Genome:

> 'When you map the whole genome and we have done 30 here, you find 5-10 X excess [variation in the normal population] more than what you would expect. We have found many class 5 variations ['known' to be pathogenic] for long QT syndrome and Brugada Syndrome. This is a lot more than you would expect to find in the general population which makes you question earlier assertions.'
>
> (Clinical Geneticist 1)

This centre and many others had the perception that all genomes were more or less the same (over 99%). However, the validity of this claim is increasingly coming into question. The cause and resolution to this misunderstanding comes from genomic datasets. Because of the size of these datasets, deviations from the mean, become increasingly invisible – as the size of the dataset increases it is considered more representative of what a normal human genome should look like. However recent evidence suggests that even the genomes of monozygotic twins vary slightly, mostly through copy number variation (Bruder et al, 2008), and these changes can have significant consequences with reports of discordant monozygotic twins, where one suffers with a congenital heart defect. The result of this experience is the understanding that genomic datasets do not truly represent their local situation, thus an effort is made to situate the data themselves. This can be seen as a form of reflexive standardisation (Timmermans, 2015) as professionals in the clinic and the lab are not simply rejecting the assumptions made by large international genomic datasets but manipulate these assumptions to better fit their experience and practice.

Because clinical genetics services in the UK tend to get local referrals, the genetic variance in their clinical population is far more constrained than an international genome database and this has great implications for the utility of testing for certain genetic mutations. A good example of this come from genetic testing for Cystic Fibrosis in British Pakistani populations:

> 'We have a standard UK Cystic Fibrosis test but we have also designed one which is targeted for the Pakistani British population so it picks up all the mutations that arrive within that population, applying the British test is completely pointless, it doesn't have the same pick up rate.'
>
> (Clinical Geneticist, 2)

The most common Cystic Fibrosis mutation, *deltaF508,* is reported to be present in 74.1% of Cystic Fibrosis sufferers in the UK; however, it only represents 24.7% of British Pakistani Cystic Fibrosis sufferers (McCormick, et al 2002). The pickup rate of a Cystic Fibrosis panel would be very low for this minority population, which is a particularly large demographic for this geneticist. In this case, the genomic databases and population studies do not represent the local clinical population so the data is situated post-hoc to account for this. Although the situation is perhaps less striking in cardiac genetics it none the less persists and has implications for both the mutations that are looked for and value attributed to certain genes reported in the literature. A laboratory geneticist discussing her cohorts noted how this is particularly tricky:

> 'We get a lot of referrals from [location] and there is a specific mutation we have found in some [minority group] families, we don't know whether they are related as well. Now we tend to get asked for the [minority] mutation, if they send us a sample from [minority] families... Where we find that [mutation] we will contact the clinician and we'll say 'by the way we've found this', because they may not know that this is not common in the rest of our cohort.'
>
> (Laboratory Geneticist)

This centre is also an internationally renowned referral centre, receiving many referrals from New Zealand that causes a similar issue:

> 'We have got a couple of new Zealand families... who have got the same mutation that we've not found it in any of the UK families or any others in the world. Alongside that, there are other variants that you find in different populations as well. You usually end up classifying them as unlikely to be pathogenic but sometimes they are UV's (unknown variants). In Maori New Zealand people or black Africans, you would expect to find genetic variation or variants that you are not familiar with in Caucasians.... It does make it tricky to interpret because the cohorts of information that tend to be published are in Caucasian populations. We are not testing a lot of African people so it doesn't cause a big problem. But, if we were suddenly to have a collaboration with an African country or community we might need to think about what other data we would need to interpret those variants.'
>
> (Laboratory Geneticist)

Although this issue of diversity in datasets is reducing over time as more genomes are sequenced across the world, there remains an understanding at the clinical level that databases of genomic information are over populated by Caucasian samples.

It is through recognising this *representational uncertainty* within genomic databases in terms of both the racial demography and the ability of the databases to accurately capture the genetic characteristics of their local populations that clinicians and scientists are able to reflexively attribute clinical value to some recognised variants over others. In the same vein this uncertainty offers clinicians the opportunity to attribute validity where scientific consensus would be at best uncertain, however officially this is far more difficult to achieve. Laboratory Geneticists are restricted in their reports in terms of the autonomy they have to attribute significance to variants that are not supported in the literature or are not recognised on OMIM (Online Mendelian Inheritance in Man) or HGMD (the Human Gene Mutation Data Base) (See,

Timmermans, 2015 for a discussion on these databases). As a result, the laboratory conveys their situated certainty in more subtle ways as to avoid any future comeuppance as a professional accountable group. This is only achievable through close collaboration with the clinic:

> 'We wouldn't necessarily report something because we think it might be doing something, everything is evidence based in terms of what our reports say. As much as we might think that a mutation... and I mean we sit there and think about a mutation because we are just not sure how to report it. There is one gene at the moment, Tropomyosin, we know just from our reading in terms of the function of the gene and how well conserved it is, we know that if we found a mutation in there it *is* going to be what has caused the phenotype, but there is not much reported out there. So in terms of evidence building we can't hang it on anything. You can't just say I think that's what's happening. The expectation of the clinician is that actually they may also think that it is pathogenic and they may also tell the patient that it's highly likely to be pathogenic but we haven't said that on our report and we can't say that.'
>
> (Laboratory Geneticist)

By not providing black and white answers to the clinician, which this geneticist claims they are no longer able to do, she is putting trust in the clinician to be able to interpret the genetic test report in relation to the patient's phenotype, attributing validity or discrediting the findings on this basis. The clinician-lab relationship is incredibly important for the laboratory geneticist, she needs to trust that the clinician has the ability to expertly interpret her report and situate it as part of the differential diagnosis of the patient and their family. As such, the laboratory will only accept referrals from within the community of cardiac genetics, including clinical geneticists as well as some specialist cardiologists:

> 'We tend to phone and check with a cardiologist because... they are very confident people and you know they, in their eyes they have requested the most appropriate test for their patient, but they might not necessarily understand all of the ins' and outs of it and trying to get that message across is sometimes quite difficult.'
>
> (Laboratory Geneticist)

The assumption that a certain level of expertise is required to negotiate genetic testing as opposed to other tests such as biochemistry permeates throughout the NHS genetics lab and clinic. The ability to interpret, manipulate and situate data is seen as a prerequisite entry requirement to the cardiac genetics professional community. This is precipitated by the

understanding that genomic datasets and the standards of practice in relation to these datasets do not reflect their local clinical needs and are more reflective of the needs of the research setting. However uncertainty of the value of genomic data is not solely based upon the inability of this data to fully encompass the characteristics of their local population, and by this I mean both demographically as well as in terms of the characteristics associated with gene mutations (i.e. phenotypical heterogeneity).

Mechanical Uncertainty

There is also an understanding that the technology itself is somewhat flawed in its ability to 'truly' represent the human genome. This second type of uncertainty can be referred to as *mechanical uncertainty* in that it is uncertainty based on a deep, technical understanding of the technology used to the extent that the inaccuracies of the technology are known. For the laboratory geneticist this uncertainty is garnered through a familiarity with the processes of sequencing genes and whole genomes, they are also aware of the technical challenges of using the technology and the ways in which one would overcome such issues in a local environment. This reflects MacKenzie's (1990) uncertainty trough where those close to the production of the knowledge or technology have a high level of uncertainty in relation to its ability to perform the task it is designed to do, and to produce an objective image of that which it is designed to capture. *Mechanical uncertainty* serves as a response to Daston and Galison's (1992) *Mechanical Objectivity* (p.82). WGS and NGS as a whole marked a move in the biological sciences to become more 'true to nature' (Daston and Galison, 1992, p. 85), in that it is presented as a technology, that with very little human intervention can read and present an objective image of the human genome. However just like the human eye, this technology skips over some sections or focuses too heavily on others resulting in a subjective image of the human genome. NHS Laboratory geneticists see these issues and correct them as part of their everyday practice, they are green fingered and can get the technology to cover what they want through work-arounds and modifications:

> 'For NGS technology, we know that it doesn't cover an entire gene or an entire region, there might be the odd base here and 2 bases there and 15 bases there that aren't covered and that is because of the way that the probes bind to the original DNA. For the genes that have the highest clinical utility, so where most of the mutations are found in people with a certain disease, we will Sanger sequence across those gaps to make sure we have got complete coverage of that gene or of that exon... We know what the coverage of our test is, we know it might not pick up huge duplications it might not pick up huge deletions, it might not be very sensitive in homo-polymer tracts, which is a tract

> of similar nucleotides all in a row, because the probes don't like binding there, or they bond too strong and don't dissociate when they should do. Over time and with experience you get to know more about the limitations of the test.'
>
> (Laboratory Geneticist)

In this process of local 'co-construction' (Clarke and Fujimura, 1992, p. 7), the laboratory ensures that the technology used is fit to perform the task they designate. This is gained through a deep knowledge of the technology and its capabilities to perform the necessary tasks. Moreover, knowledge of what the test covers and what it does not, impacts the reports that are given to clinicians, mechanical uncertainties are not ignored they are instead embedded into the practices of the lab and minimised as much as possible. However, this does effect the labs interpretation of 'objective' datasets and the validity attributed to particular research findings. Because the NHS laboratory uses the same sequencing technology as the research setting they know the mechanical uncertainties associated with it, this mediates the certainty attributed to research findings and of control sequence data. For example the laboratory will further investigate areas of genes associated with conditions, which might not be fully covered by NGS because, in their experience, and based on a technical understanding of genomics they 'know' functionally that a mutation in this area could be responsible for the phenotype, even where it is not reported in the literature. This is not to say that NGS is not 'the right tool for the job' (Clarke and Fujimura, 1992) but instead pragmatically asserts it is the best tool for the job at their disposal at this time (Bossuyt, et al, 2012). By understanding the uncertainties, clinicians and laboratory geneticists can manipulate the technologies and data to reduce the uncertainties experienced in their local practice.

The foregoing demonstrates how data and practice mutually structure one another – both in terms of how technology and genomic data structures the way it is used in the clinical setting and how the users re-configure the technology and data to better fit their situated practices. The broader significance of these observations about the connection of data to practice is what they highlight about the concept of 'clinical usefulness'. The overarching argument of this paper is that genomic data only becomes clinically useful through a process of co-construction at a local level, through an interpretive employment of standards and a radical transformation of data.

Clinicians are acutely aware of the constitution of genomic databases, as a result of the regular enrolment of patients into clinical trials and WGS projects. As such, it is the clinician that construct who are counted, they are responsible for data acquisition. Although they are not involved in the technical process of counting, they are still aware that the data they use is not

'Raw', in that participants are pre-selected and thus not representative. Laboratory Geneticists are also acutely aware that the datasets are not 'clean'. Due to an understanding of the technical issues associated with DNA extraction and sequencing, the signal-to-noise ratio is far too high, as is shown by the way in which the laboratory post-hoc cleans the data output.

The clinical use of NGS has instigated a relationship between big genomic data and clinical practice. However, genomic datasets are not used, they are worked with and worked on in relation to the needs and practices of the patients in a process of *reflexive standardization* (Timmermans, 2015). To this end, I argue that an implicit trust in the objectivity (Porter, 1995; Daston and Galison, 1992) of genomic data and sequencing technology, as embodied within large genome research collectives (Cambrosio et al, 2006) is a flawed representation in the therapeutic context. It is by seeing the data and technology as in a state of acceptable contingency that it becomes clinically useful. By perceiving the imperfections (substantive and local) inherent in the technology, practitioners in the clinic and the laboratory are able to re-construct and manipulate the technology to a point of clinical usefulness. This expands debates concerning the resistance to or usefulness of a genetic technology (Hedgecoe, 2008) in the clinical setting to asking questions of how clinicians construct and re-construct technologies to fit their purpose.

With genomic data becoming more accessible and prominent in the clinical setting and beyond, it becomes increasingly important to discuss its use in practice beyond the research setting. This paper shows how these datasets are considered both 'noisy' and 'cooked' by its users in the clinical setting. Understanding genomic data in this way precludes the possibility of 'data-mining' in that the data genome sequencers produce are not naturally occurring nor can they be considered 'raw'. Populations are purposely selected for inclusion in databases and the data is manipulated and reconfigured. The data is externally configured and locally reconfigured creating situated data appropriate for informing clinical decisions. It is important here to emphasise the place of genetic testing in the cardiac genetic clinical setting:

> 'The whole purpose of testing a patient who has got long qt syndrome or hypertrophic cardiomyopathy… the result would benefit in targeted selection of other at risk family members, not to the person. Frankly speaking, you do not require genetic testing to make a diagnosis, you can make the same diagnosis of these conditions by ECG or [ajmaline]

> challenge[7]... the main advantage was that once you know the mutation ... in the person, then all first degree relatives... could be offered the genetic assessment.'
>
> (Clinical Geneticist 3)

In the vast majority of cases, genetic testing is used to confirm a diagnosis made based upon clinical presentations, or to cascade screen families of a patient with a clinical phenotype. Genetic testing is rarely predictive in this setting, but instead makes up part of the clinical picture, the weight and value of a genetic mutation is assessed by the clinician based upon his/her experience and expertise, there is no formula, validity is attributed on an individual basis. This is important to note when discussing the use of genomic data outside of the clinic in the insurance industry for example (Van Hoyweghen, 2007). Viewing genomic data as objective and representative risks a reduction in the complexities associated with using this data in practice. In saying this 'data-dopes' become a real risk, the possibility arises that certain groups may use the data without the pre-requisite expertise, without the skills to situate the data within their particular practice, which presents the very real risk of the inappropriate use of genomic data.

References

Ackerman, M. et al. 2011. HRS/EHRA Expert Consensus Statement on the State of Genetic Testing for the Channelopathies and Cardiomyopathies. *Europace* 13, p. 1077-1109.

Adams, JU. 2015. Genetics: Big Hope for Big Data. *Nature* 527 p. S108-S109.

Behr, E. et al. 2015. Role of Common and Rare Variants in *SCN10A*: Results from the Brugada Syndrome QRS Locus Gene Discovery Collaborative Study. *Cardiovascular Research* 106(3), p. 520-529.

Bosk, C. 2003[1979] *Forgive and Remember: Managing Medical Failure.* Chicago: University of Chicago Press.

Bossuyt, P., Reitsma, J.,Linnet, K. and Moons, K. 2012. Beyond Diagnostic Accuracy: The Clinical Utility of Diagnostic Tests. *Clinical Chemistry* 58(12), p. 1636-1643.

[7] Ajmaline challenge is an ECG Provocation test where the clinician injects ajmaline into a patient with suspected Brugada syndrome, the ajmaline can provoke the Brugada ECG which may lay dormant in the patient.

Bourret, P., Keating, P. and Cambrosio, A. 2011. Regulating Diagnosis in Post-Genomic Medicine: Re-Aligning Clinical Judgement? *Social Science and Medicine* 73, p. 816-824.

Calvert, J. 2008. The Commodification of Emergence: Systems Biology and Intellectual Property. *Biosocieties* 3, pp. 383-398.

Cambrosio, A., Keating, P., Schlich, T. and Weisz, G. 2006. Regulatory Objectivity and the Generation and Management of Evidence in Medicine. *Social Science and Medicine* 63, p. 189-199.

Chen, Q., et al.1998. Genetic basis and molecular mechanism for idiopathic ventricular fibrillation. *Nature*.392, p. 293–296.

Clarke, A. and Fujimura, J.H. 1992. *The Right Tools for The Job: At Work in Twentieth-Century Life Sciences*. New Jersey. Princeton University Press.

Daston, L. and Galison, P. 1992. The Image of Objectivity. *Representations* 0(1), p. 81-128.

Davies, G., Frow, E. and Leonelli, S. 2013. Bigger, Faster, Better? Rhetorics and Practices of Large-Scale Research in Contemporary Biosciences. *Biosocieties* 8(4), pp. 386-396.

Dinant, J.M., Lazaro, C., Poullet, Y., Lefever, N. and Rouvroy, A. 2008. Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data. Application of Convention 108 to the profiling mechanism: Some ideas for the future work of the consultative committee (T-PD).

Fukushima, 2015. Constructing Failure in Big Biology: The Socio-Technical Anatomy of Japan's Protein 3000 Project. *Social Studies of Science* [online first].

Garfinkel, H. 1967. *Studies in Ethnomethodology*. New-Jersey: Prentice-Hall.

Genomics England, 2015. Rare Diseases Eligibility Statements: 100,000 Genomes Project. Genomics England.

Genomics England. 2015. The 100,000 Genomes Project Protocol. Genomics England.

Gibbons, M., Limonges, C., Nowotny, H., Schwartzman, S., Scott, P, et al. 1994. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Society.* Thousand Oaks: SAGE.

Hacking, I. 1990. *The Taming of Chance*. Cambridge: Cambridge University Press.

Hallowell, N., Hall, A., Alberg, C., and Zimmern, R. 2015. Revealing the Results of Whole-Genome Sequencing and Whole-Exome Sequencing in Research and Clinical Investigations: Some Ethical Issues. *Journal of Medical Ethics* 41, p. 317-321.

Hedgecoe, A. 2004 *The Politics of Personalised Medicine: Pharmacogenetics in the Clinic.* Cambridge: Cambridge University Press.

Hedgecoe, A. 2006. It's Money that Matters: The Financial Context of Ethical Decision-Making in Modern Biomedicine. *Sociology of Health and Illness* 28(6), p. 768-784.

Hedgecoe, A. 2008. From Resistance to Usefulness: Sociology and the Clinical Use of Genetic Testing. *BioSocieties* 3 p, 183-194.

Hilgartner, S. 2013. Constituting Large-scale Biology: Building a Regime of Governance in the Early Years of the Human Genome Project. *Biosocieties* 8(4), pp. 397-416.

Hu, D., et al. 2014. Mutations in SCN10A are responsible for a large fraction of cases of Brugada syndrome. *Journal of the American College of Cardiology* 64(1), p. 66-79.

Illumina, 2017. Focused Power on the MiSeq System. *Illumina*[online]. Available at: https://www.illumina.com/systems/sequencing-platforms/miseq.html [Accessed: 01/06/17]

Kohli-Laven, N., Bourret, P., Keating, P., Cambrosio, A. 2011. Cancer Clinical Trials in the Era of Genomic Signatures: Biomedical Innovation, Clinical Utility, and Regulatory Scientific Hybrids. *Social Studies of Science* 41(4), p. 487-513.

Latimer, J. 2013. *The Gene, the Clinic and the Family: Diagnosing Dysmorphology, Reviving Medical Dominance.* Aldershot: Routledge.

Latimer. J., Featherstone. K., Atkinson. P., Clarke. A., Pilz. D.T., and Shaw. A. 2006. Rebirthing the Clinic: The Interaction of Clinical Judgement in the Production of Medical Science. *Science, Technology and Human Values,* 31(5), p. 599-630.

Levin, N. 2014. Multivariate Statistics and the Enactment of Metabolic Complexity. *Social Studies of Science* 44(4) p. 555-578.

Lynch, M. and Woolgar, S. 1990. *Representation in Scientific Practice*. Cambridge, Massachusetts: MIT Press.

MacKenzie, D. 1990. *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance.* Cambridge: MA. MIT Press.

McKenzie, D. and Spears, T. 2012. 'The Formula That Killed Wall Street'? The Gaussian Copola and the Material Cultures of Modelling. *Social Studies of Science.* 44(3) p. 393-417.

Marris, E. 2005. Unchecked By Government, Genetic Tests Sell Hope and Hype. *Nature Medicine* 11(6) p. 584.

Martin, P. 2001. Genetic Governance: The Risks, Oversight and Regulation of Genetic Databases in the UK. *New Genetics and Society* 20(2), p. 157-183.

McCormick, J., Green, M.W., Mehta, G., Culross, F. and Mehta, A. 2002. Demographics of the UK Cystic Fibrosis Population: Implications for Neonatal Screening. *European Journal of Human Genetics* 10, p. 583-590.

Merton, RK. 1942. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: Chicago University Press.

Oudshoorn, N and Pinch, T. Eds. 2013. *How Users Matter: The Co-Construction of Users and Technology*. Cambridge, Massachusetts: MIT Press.

Oudshoorn, N. 2016. The Vulnerability of Cyborgs: The Case of ICD Shocks. *Science, Technology and Human Values.* 41(5), pp. 767-792.

Porter, T. 1995. *Trust in Numbers*. Princeton University Press.

Rabeharisoa, V. and Bourret, P. 2009. Staging and Weighting Evidence in Biomedicine. *Social Studies of Science* 39(5), pp. 691-715.

Rose, N. 2013. The Human Sciences in a Biological Age. *Theory, Culture & Society* 30(1), p. 3–34.

Shostak, S. 2005. The Emergence of Toxicogenomics: A Case of Molecularization. *Social Studies of Science* 35(3), pp. 367-403.

Splawski, I., Shen, J., Timothy, K. W., Lehmann, M. H., Priori, S., Robinson, J. L., Moss, A. J., Schwartz, P. J., Towbin, J. A., Vincent, G. M., Keating, M. T. 2000. Spectrum of mutations in long-QT syndrome genes: KVLQT1, HERG, SCN5A, KCNE1, and KCNE2. *Circulation* 102, p. 1178-1185.

Stephens, N. 2013. Growing Meat in Laboratories: The Promise, Ontology, and Ethical Boundary-Work if Using Muscle Cells to Make Food. *Configurations* 21(2), p. 159-181.

Stevens, H. 2011. On the Means o Bio-production: Bioinformatics and how to make knowledge in the High-Throughput Genomics Laboratory. *Biosocieties* 6(2), pp. 217-242

Stevens, H. 2013. *Life Out of Sequence: A Data-Driven History of Bioinformatics.* Chicago: University of Chicago Press.

Stivers, T. and Timmermans, S. 2016. Negotiating the Uncertainty of Genomic Test Results. Social Psychology Quarterly 79(3), pp. 199-221.

Timmermans, S. and Berg, M. 1997. Standardization in Action: Achieving Local Universality Through Medical Protocols. *Social Studies of Science* 27, p. 273-305.

Timmermans, T. and Berg, M. 2003. *The Gold Standard: The Challenge of Evidence-Based Medicine and Standardization in Health Care*. Philadelphia: Temple University Press.

Timmermans, S. 2015. Trust in Standards: Transitioning Clinical Exome Sequencing from Bench to Bedside. *Social Studies in Science* 45(1), p. 77-99

Timmermans, S., Tietbohl, C. and Skaperdas, E. 2016. Narrating Uncertainty: Variants of Uncertain Significance (VUS) in Clinical Exome Sequencing. Biosocieties [online first].

The European Bioinformatics Institute. 2015. *Background*. EBI. Online. Available at: http://www.ebi.ac.uk/about/background [Accessed: 13/01/2016]

The UK10K Consortium. 2015. The UK10K Project Identifies Rare Variants in Health and Disease. *Nature* 526, p. 82-90.

Wang, C., Schroeder, K., Rosenberg, N. 2012. A Maximum-Likelihood Method to Correct for Allelic Dropout in Microsatellite Data with No Replicate Genotypes. *Genetics* 192(2), p. 651-669.

Woolgar, S. 1991. Configuring the User: The Case of Usability Trials. In: Law, J. (Ed) *The Sociology of Monsters: Essays on Power, Technology and domination*. London: Routledge. p. 57-102.