

*Developing people
for health and
healthcare*

Developing
Clinical Bioinformatics
Training in the NHS
- a timeline for action



Report of the Clinical Bioinformatics
Task & Finish group – February 2015

Bioinformatics

Contents

Executive summary and key recommendations	3
Introduction	5
Criteria and delivery timeline	5
Training needs analysis	6
Short-term requirements: 2014/15 – 2017/18	7
<i>Figure 1</i> Staff groups requiring immediate training to support the 100,000 genomes project	8
<i>Figure 2</i> HEE Bioinformatics Roadmap: Longer-term training	10
Longer-term training needs: 2015/16 – 2018/19	11
Numbers and Delivery	12
Conclusions	13
References	14
Appendix 1: Remit and Membership of the Bioinformatics Task and Finish Group	15
Remit of the Clinical Bioinformatics Task and Finish Group:	15
Membership of the Group	15
Appendix 2: Roles of a clinical bioinformatician	16
Appendix 3: Clinical bioinformatics competency matrix	17
Appendix 4: Mapping of competency to existing HEE training in genomics or bioinformatics	19
Appendix 5: Challenges in the clinical application of genome sequencing	20
Appendix 6: Case studies in rare disease, cancer and infectious disease	21
Rare disease	21
Cancer (Case 1)	23
Cancer (Case 2)	23
Infectious disease	24

Executive summary and key recommendations

The Health Education England Genomics Education Programme (GEP) brought together a Bioinformatics Task and Finish Group in March 2014 (*Appendix 1*) to identify the bioinformatics training requirements related to the introduction of genomic technologies across the health and care system.

The key recommendations from the Group focus on the immediate need to support the work of the Department of Health's 100,000 genomes project and the longer term, but no less urgent, requirement to develop the workforce:

1. Bioinformatics training for staff using genomics and statistics in the NHS will be an evolving programme of workforce development and transformation at all levels.
2. To address immediate needs (Fig 1), continuing professional development (CPD) will be required for 3 distinct workforce groups:
 - The **specialist healthcare scientists** (NHS bioinformaticians and clinical scientists in genetics, molecular pathology and infectious disease) will require CPD training to handle and interpret genomic data. This training will be delivered through modules of a new MSc in Genomic Medicine, Modernising Scientific Careers (MSC) scientist training, Higher Specialist Scientific Training (HSST) and specialist courses provided by external organisations.
 - The **clinical staff** involved in recruiting patients to the 100,000 genomes programme will require Continuing Professional Development (CPD) training in order to understand the results of genome sequencing and counsel patients (and their relatives) accordingly. This training will include modules of the new MSc in Genomic Medicine and specialist courses provided by external organisations.
 - The **general workforce** will require training to provide awareness of genomic medicine and how it can improve patient care. This training will be provided by short online courses.
3. The following training interventions are needed to address the longer term goal (Fig 2) to establish a workforce fit for 'genomic medicine' and the implementation of other '-omics' technologies, the following training is recommended
 - **Doctorate and post-doctorate training programmes** in Genomic Medicine to develop future clinical academic leaders.
 - **Post-graduate MSc courses**, building on the work already underway through current healthcare science education and training, funded to ensure adequate outputs on workforce numbers. The size of the workforce will need to be properly estimated by HEE for their planning.
 - More **general training to raise awareness of bioinformatics** and its role in delivering genomic technologies for the current, non-genetic/genomic specialist workforce in areas such as rare diseases and cancer, first-contact professionals such as GPs and commissioners of services.

— **Increased teaching of genomics and bioinformatics within undergraduate curricula** for students of medicine, nursing, allied health professions and healthcare science.

- We recommend that HEE should allocate funding for all these areas or training and should work closely with other funders of training (eg MRC, NIHR and BBSRC) to ensure that these training needs are met in a timely fashion.
- We recommend that all training should take advantage of the 'Training Embassy' provided by Genomics England, to allow trainees to use real data gathered in a clinical setting. We envisage that a series of exemplars of genomic interpretation for different contexts will be provided, to enable practical training, suitable for genomic interpretation within the NHS.
- We recommend that there is a detailed analysis and longer term watching brief of current training provision, with a regular training needs analysis for genomic medicine and identification of expert groups that might deliver it. This would also identify synergies, encourage re-use of existing training material, and allow the programme to evolve as the field changes.

Introduction

The Government has set an ambitious and innovative objective of sequencing 100,000 whole genomes by the end of 2017. The Department of Health (DH) is responsible for the overall project but to lead on the capacity and infrastructure development, DH has established a company, Genomics England Limited, as the key delivery vehicle. The other key delivery partners are NHS England, Public Health England (PHE) and Health Education England (HEE).

Health Education England has established a Genomics Advisory Board with 5 Task and Finish Groups appointed in 2014. This paper summarises the outputs from meetings of the Task and Finish Group on Clinical Bioinformatics, which will inform the development of HEE bioinformatics education and training resources as part of their overall genomics programme. The group's remit was to identify the training requirements to support the 100,000 genomes project in the short-term and the adoption of genomic medicine in the longer term. The group noted that the provision of accredited training programmes contributes to the development of a competent workforce but additional training in the workplace with appropriate supervision and mentorship is essential to ensure diagnostic accuracy and patient safety. Further development of existing best practice guidelines (for both clinical and scientific pathways) and external quality assessment processes for genomic sequencing is also important for the delivery of high quality genomic medicine. The fast pace of development in genomics and other '-omics' technologies places a greater need for integration of researchers and NHS service providers than before.

The issues and recommendations identified by the Task and Finish group are focussed on:

- the immediate need to support the work of the Department of Health's 100,000 genomes project and Genomics England in achieving their goals
- the longer term, but no less urgent, requirement to develop the workforce, including:
 - Expanding the new cadre of clinical bioinformaticians working within the NHS (their role is outlined in Appendix 2)
 - Fostering a cohort of clinical academic leaders in bioinformatics and genomic medicine
 - Raising awareness and knowledge amongst specialist and general NHS staff

Criteria and delivery timeline

Bioinformatics training in the NHS will be an evolving programme of workforce development and transformation. The initial focus will be the development of existing NHS staff, to equip them with the new skills required (see Figure 1), together with the establishment of academic training programmes to develop research capacity and leadership in the NHS. The second objective is to set out a longer term roadmap that will see the emergence of a flexible workforce to lead and support the adoption of genomic medicine, working across science and healthcare (see Figure 2).

As part of the 100,000 genomes project, Genomics England will provide a Training Embassy with access to genome sequence data gathered in a real clinical context. These data will be made available to all trainees, who are part of the HEE training programmes and can be used by trainers to develop some exemplars of genomic interpretation and its application in the clinic, thereby providing the hands-on practical training needed for NHS staff. We strongly recommend that training developed by the HEE takes advantage of this facility.

Training needs analysis

To analyse training needs, map these to currently available training and identify any gaps, we drafted a competency framework for bioinformatics as it relates to genomic medicine. We first defined the NHS roles involved in genomic medicine that require competency in bioinformatics and then divided bioinformatics, as it applies to genomic medicine, into defined areas of competency, for example 'Analyse genomics data using pre-existing software'. We then asked practitioners or experts in each role to define which competencies would be required by that role to fulfil the goals of the 100,000 genomes project. We also asked them to select the level at which each competency is required. Four options were provided: (1) no competency required; (2) awareness; (3) working knowledge; and (4) specialist knowledge.

The consensus matrix developed by this approach (See Appendix 3) allowed us to identify training needs, compare them with what is currently on offer, and formulate our recommendations. For the latter we selected three curricula that have specifically been developed with the aim of realising the potential of genomic medicine in the NHS:

4. HEE's Genomics Education Programme introductory courses
5. Modernising Scientific Careers (MSC) Scientist Training Programme MSc in Clinical Science – Clinical Bioinformatics
6. The new HEE Genomics Education Programme MSc Programme in Genomic Medicine and a new course on ethics and consent, currently in development.

Mapping the competencies to existing training provision or programmes under development (Appendix 4) showed how many of the specialist training requirements to support the 100,000 genomes project will be met through the HEE commissioned MSc in Genomic Medicine (and CPD accredited modules therein) or through the MSC clinical scientist training scheme in bioinformatics. There are also a number of specialist training courses available outside HEE. These can contribute to the development of bioinformatics skills but have not been specifically designed to meet the requirements for clinical staff and scientists delivering the 100,000 genomes project or for genomic medicine more generally. The mapping could be used to identify other appropriate course providers and/or develop new training.

Short-term requirements: 2014/15 – 2017/18

This section presents the short-term recommendations to address the immediate needs. These will be met by three distinct workforce groups:

- a. The first urgent need is for specialist healthcare scientists to assist clinical staff with the interpretation of genome sequencing data as part of the 100,000 genomes project. Existing NHS bioinformaticians and clinical genetic scientists provide a source of knowledge and expertise acquired through development and implementation of diagnostic targeted next generation sequencing and exome sequencing that can be exploited and further developed. The training modules will need to be tailored to meet the data formats delivered by Genomics England (or by Public Health England in the case of infectious diseases). Both NHS bioinformaticians and clinical scientists in genetics have skill sets that can respond to the immediate need but will need development through CPD. For the infectious disease component of the 100,000 genomes project, Public Health England will similarly need highly skilled individuals to further establish their current bioinformatics infrastructure and pipeline. In addition, scientific staff (including clinical scientists and healthcare science practitioners), medical and nursing staff will be able to enrol on the new HEE commissioned MSc programme in Genomic Medicine on a part-time, full-time or per module basis.
- b. The second group who will require urgent training are the clinical staff who will recruit patients with rare diseases or cancer for sequencing of their genome and will need to explain the relevant results. Existing personnel with specialist training, for example in genetics or oncology, will need additional training through CPD. A greater challenge is to provide training for clinical staff across all specialties to increase awareness and understanding of genomic technology, so that they can recognise those patients for whom genomic tests are appropriate, capture appropriate data required for analysis, explain the results to patients and manage their care accordingly. It is essential that these staff understand the challenges in the clinical application of genome sequencing and have appropriate training to minimise the risk of misdiagnoses (see Appendix 5). Additional training will also be required for clinicians managing the care of patients with infectious diseases.
- c. Thirdly the wider healthcare workforce, including commissioners and IT staff providing informatics support, needs to develop a level of awareness and understanding of how genomic data can be interpreted biologically and clinically and the critical role and requirements of bioinformatics for the adoption of genomic medicine.

The outcomes of this training should provide the following:

In rare diseases:

- NHS bioinformaticians and clinical scientists with enhanced skills enabling them to translate genome sequencing results from the 100,000 genomes project into fully interpretive clinical diagnostic reports.
- Clinical staff trained to recruit and consent patients for whole genome sequencing (WGS), capture standardised data sets, explain and take appropriate action according to the results described in the clinical diagnostic report.

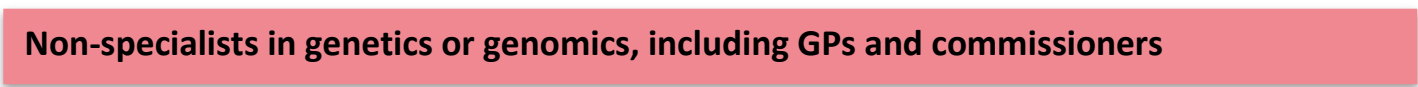
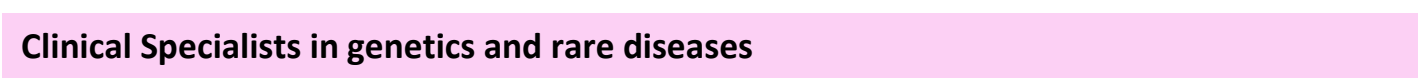
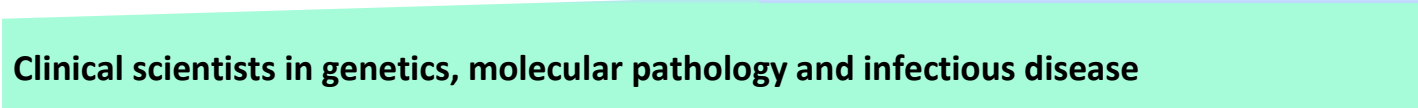
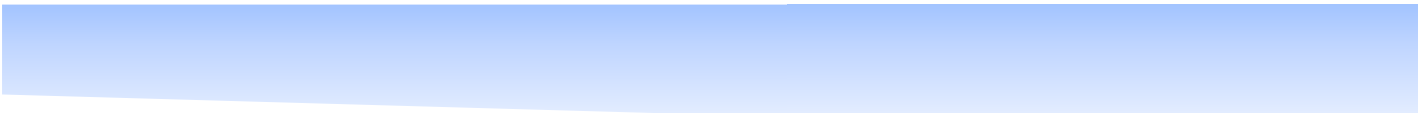
Figure 1

Staff groups requiring immediate training to support the 100,000 genomes project

Year 1	2	3	4	5	6	
2014 /15	2015/16	2016/17	2017/18	2018/19	2019/2020	

Clinical bioinformaticians

100,000 genomes project



Note 1: The current distinction between the roles of clinical scientists in genetics, molecular pathology and bioinformatics within the NHS is likely to change as genomic medicine develops. The scope of practice of each of these professions may converge (perhaps to create clinical scientists in genomics) and training programmes will need to be flexible to provide staff with the appropriate skills mix.

Note 2: Genetics and genomics expertise must be extended across the clinical workforce in order for genomic medicine to become an integral part of patient pathways across all specialties caring for patients with rare disease, cancer and infectious disease.

- Multi-disciplinary team working of staff with iterative interaction of clinical/bioinformatics/scientific expertise to interpret genome sequencing results.
- Updated MSC clinical scientist and HSST training programmes in clinical bioinformatics and genetics that incorporate the above requirements for delivering the 100,000 genomes project.

In cancer:

- NHS bioinformaticians and clinical scientists with enhanced skills enabling them to translate genome sequencing results from Genomics England into clinical reports.
- Clinical staff trained to recruit and consent patients for WGS, capture standardised data sets, explain the results and integrate this information into their recommendations for treatment.
- Multi-disciplinary team working of staff with iterative interaction of clinical/bioinformatics/scientific expertise to interpret genome sequencing results.
- Updated MSC clinical scientist and HSST training programmes in clinical bioinformatics, molecular pathology and genetics that incorporate the above requirements for delivering the 100,000 genomes project.

In infectious diseases:

- PHE and NHS bioinformaticians and scientists with enhanced skills enabling them to translate PHE and Genomics England results into clinical diagnostic and public health reports.
- Clinical staff trained to explain the results described in the clinical diagnostic report.
- Multi-disciplinary team working of staff with iterative interaction of epidemiology/clinical/ bioinformatics/scientific expertise to interpret genome sequencing results.
- Updated MSC clinical scientist and HSST training programmes in clinical bioinformatics and infectious disease that incorporate the above requirements for delivering the 100,000 genomes project.
- Augmented RCPATH CPD training to include genomics approaches for infectious diseases.

Case study examples of the application of bioinformatics for improved clinical care in the areas of rare disease, cancer and infectious diseases are described in Appendix 6.

The implications of the 100,000 genomes project for the future of medical practice go well beyond the specialties of clinical genetics, oncology and infectious disease. Adoption of the principles of precision medicine (integrated, multi-disciplinary, multi-level approaches to the analysis of human samples and data to improve clinical care through increased precision in the understanding of mechanisms of disease, drug response and environmental exposure) by clinical and scientific staff is essential.

We recognise that the successful integration of genomic medicine into healthcare will require expert management and interpretation of the data. This will demand a combination of statistics and informatics skills, including software engineering and development, biological data and database management, user-interface design for databases and analysis tools, information governance in the NHS and knowledge of disease biology and clinical process.

Figure 2

HEE Bioinformatics Roadmap: Longer-term training to develop the workforce

Year 1	2	3	4	5	6	
2014 /15	2015/16	2016/17	2017/18	2018/19	2019/2020	

100,000 genomes project

Awareness for non-specialists in genetics or genomics, including GPs and commissioners

PhD and post-doctoral fellowships in clinical bioinformatics

MSc in Genomic Medicine

Increased genomics teaching for medical, nursing, allied health professional and healthcare science undergraduates

The ability to apply software engineering and data management skills within a robust governance framework that ensures patient safety is essential. Individual staff or teams that combine all these skills are rare and not routinely employed in the NHS. In the longer term, the aim should be to develop a cadre of such people. We also anticipate the growth of bioinformatics skills in industry for the development of software to analyse genome sequencing undertaken for the purposes of clinical diagnosis, prognosis and response to treatment. Finally, there is a need to improve the availability and quality of electronic healthcare records as genomic medicine will increasingly rely upon the integration of clinical data with genomic data. This will require the implementation of international data standards and mandatory quality-controlled data submission to allow sharing of anonymised datasets for the benefit of clinical care and translational research.

Longer-term training needs: 2015/16 – 2018/19

This section presents our recommendations for actions that need to be taken now to address the longer term needs.

Workforce development and transformation to embed genomics and bioinformatics expertise in the NHS is crucial to achieve the greatest patient benefit from genomic medicine. During this period the wider training programme for clinical bioinformatics will need to be established. Whilst the short term requirement is to respond to the needs of the 100,000 genomes project by building the current workforce through enhanced skills programmes, the development of the clinical bioinformatician profession in the longer term may be multi-layered, with career opportunities for specialisation involving both the clinical and data technology fields.

As well as clinical bioinformaticians supporting each step of the clinical genomic pathway, doctorate and post-doctorate training programmes are needed to develop a new cadre of professionals that will become global leaders in clinical bioinformatics, with the understanding, capability and confidence to work across both the clinical and academic research fields. We envisage an integrated approach whereby PhD students/post-doctoral fellows in research training programmes and NHS bioinformatics trainees will have the opportunity for cross-over training placements (including secondments to Genomics England) to facilitate innovation within clinical genomics. It will also be important to facilitate the development and integration of other technological innovations, for example proteomics, transcriptomics, metabolomics and epigenomics, so that the interpretation of such genome-scale datasets with respect to clinical intervention will become routine. The significant transformation required to deliver genomic medicine must be driven by clinical leadership. In order to be able to use ‘-omics’ in day-to-day clinical practice, clinicians need to be aware of both benefits and limitations of the technology, and all developments must be led by clinical need.

Any education and training programme must take account of the need to raise general awareness of bioinformatics and genomic medicine in the existing workforce and build genomics into undergraduate curricula for medical, nursing and healthcare science courses to ensure that future generations of NHS staff are aware of the importance of genomics at the earliest opportunity.

We envisage four components:

- Doctorate and post-Doctorate training programmes that will include analysis of data generated by the 100,000 genomes project.
- Delivery of post-graduate MSc courses, building on the work already underway through current healthcare science education and training, funded to ensure adequate outputs on workforce numbers. The size of the workforce will need to be properly estimated by HEE for their planning.
- More general training and awareness training on bioinformatics and its role in delivering genomic technologies, both for the current, non-genetic/genomic specialist workforce in areas such as rare diseases and cancer and for first-contact professionals, such as GPs.
- Increased teaching of genomics within undergraduate curricula for medical, nursing, allied health professionals and healthcare science students.

In 2016, the first cohort of trainees will complete the Modernising Scientific Careers (MSC) Scientist Training Programme in Clinical Bioinformatics and join the existing NHS bioinformaticians in the workplace. In 2019 the first cohort of PhD students will complete their four-year training. As genomics is embedded into clinical practice the focus is likely to shift towards implementation of transcriptomics, proteomics, metabolomics and epigenomics as the transformation of healthcare science advances.

Once the 100,000 genomes project is completed at the end of 2017 we anticipate that clinical bioinformatics skills will be sufficiently developed within the NHS for human genome sequencing to be commissioned as a diagnostic test for rare diseases and as a diagnostic/prognostic tool for cancer once clinical utility and cost-effectiveness have been established. The workforce transformation set out in this report will empower highly qualified healthcare professionals working across science and medicine to lead world-class research and healthcare throughout the NHS, supporting the innovation, health and wealth agendas.

Numbers and Delivery

How the proposed education and training programme is delivered will need to be carefully considered. In the medium term, it is likely that only a small number of specialist centres will have the capacity to deliver the quality of education and training needed. Programmes in such specialist centres will also provide a service delivery model that can bring together a critical mass of students that will ensure the benefits of shared learning. The creation and maintenance of standardised pipelines for genomic tests will be most readily achieved through networks of specialist centres for rare disease, cancer and infectious disease. Speculation on the total target for workforce numbers is also difficult at this stage and is something to keep under constant review as the 100,000 genomes project and clinical utility for genomic sequencing increases.

The training gap analysis summarised in Appendix 4 reveals overlap in some areas of current provision. This reflects the staff group-specific nature of some training programmes combined with the requirement for competency across a wider workforce. The apparent lack of any provision for some of the more specialised competencies is addressed in some areas through workplace-based training, for example by MSC clinical scientist and HSST training programmes. These issues will need to be addressed in planning future training.

The competency framework developed by this task force can form a basis on which to develop and maintain a responsive educational programme for clinical bioinformatics. By focusing on the immediate needs of the 100,000 genomes project, we have identified training gaps that need filling as a matter of urgency. However, bioinformatics as it relates to genomic medicine is evolving extremely rapidly, and curricula will need to keep pace – both for CPD and for higher education. As mentioned above, bioinformatics has clinical applications beyond genomic medicine, and this will need to be reflected in future training programmes. The framework will need to undergo regular updates, informed through consultation with a broader community of healthcare professionals and through other initiatives, for example ELIXIR-UK. Mechanisms for updating curricula, both to train the future workforce and to maintain state-of-the-art competency in the current workforce, will need to be linked to updates to the framework.

Conclusions

The 100,000 genomes project provides a unique opportunity to build the clinical bioinformatics profession in unison with sequence, annotation and variant interpretation service development. Whilst the prospect of more genomic technology in healthcare is becoming more accepted, the seismic shift in the patient/healthcare professional experience and its revolutionary and evolutionary nature should not be underestimated. Through our recommendations and building on the current service provision, we can develop the new skills required to deliver the full potential of genomic medicine for maximum clinical benefit.

References

Bell, C. J. et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci. Transl. Med* 2011; 3, 65ra4.

MacArthur, D. G. et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012;335;823–828.

MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014; 508(7497): 469-76.

Piton A, Redin C, Mandel J-L. XLID-Causing Mutations and Associated Genes Challenged in Light of Data From Large-Scale Human Exome Sequencing. *The American Journal of Human Genetics*. 2013;93(2): 368-83.

Wright CF, Middleton A, Burton H, Cunningham F, Humphries SE, Hurst J, et al. Policy challenges of clinical genome sequencing. *BMJ*. 2013; 347.

Appendix 1:

Remit and Membership of the Bioinformatics Task and Finish Group

Remit of the Clinical Bioinformatics Task and Finish Group:

To identify the bioinformatics training requirements to support the 100,000 genomes project in the short-term and the adoption of genomic medicine in the longer term.

Membership of the Group

(co-chair) Sian Ellard, Consultant Clinical Scientist and Professor of Human Molecular Genetics, Royal Devon & Exeter NHS Foundation Trust and University of Exeter Medical School

(co-chair) Janet Thornton, Director of EMBL-European Bioinformatics Institute

Tim Aitman, Professor of Clinical and Molecular Genetics, Academy of Medical Sciences

Chris Boustred, Bioinformatician, Great Ormond Street Hospital

Cath Brooksbank, Head of Training, EMBL-European Bioinformatics Institute

Mark Caulfield, Genomics England

Tom Clayton, Health Education England

Jonathan Colbourne, Chair of Environmental Genomics, University of Birmingham

Ann Dalton, Consultant Clinical Scientist, Sheffield Children's NHS Foundation Trust

Val Davison, Scientific Advisor, National School of Healthcare Science

Andrew Devereau, Genomics England Ltd

Helen Firth, Consultant Clinical Geneticist, Cambridge University Hospitals Trust

Jonathan Green, Head of Bioinformatics Unit, Public Health England

Georgina Hall, Consultant Genetic Counsellor, Central Manchester NHS Foundation Trust

Patricia Oakley, Training Research Fellow, Kings College London

Colin Pavelin, Health Education England

Chris Ponting, Professor of Genomics, University of Oxford

Imran Rafi, Chair, Clinical Innovation and Research Centre, RCGP

Anna Schuh, Consultant Haematologist, Oxford Molecular Diagnostics Centre

Anneke Seller, Scientific Advisor, National School of Healthcare Science

Stuart Sutherland, National School of Healthcare Science

Ann-Marie Wright, Genomics Project Manager, National School of Healthcare Science

Caroline Wright, Senior Scientific Manager, Wellcome Trust Sanger Institute

Appendix 2:

Roles of a clinical bioinformatician

Work within multi-disciplinary teams to understand the needs of the service and communicate genomic data in the context of patient referrals

Provide support and advice to clinical and research services to develop appropriate analysis strategies

Evaluate, validate and understand the limitations of available tools and choose those most appropriate to meet the needs of the clinical service

Apply computer programming and statistical skills to the analysis of large data sets

Understand general principles of human genome sequence and variation

Analyse genomics data using the most appropriate pre-existing software tools and/or custom-built software, and relevant genomic/bioinformatics resources. Apply quality control and filtering strategies to produce accurate variant lists for further validation, clinical interpretation and reporting by genetic/genomic clinical scientists

Ensure compliance of bioinformatics pipelines, standard operating procedures and strategies with local quality management systems and in line with ISO 15189 standards for clinical laboratory accreditation

- Liaise with IT support services to ensure appropriate data storage and computing facilities and ensure compliance with data governance regulations
- Initiate and/or participate in formal and informal research programmes to provide new knowledge and insight into the applications of bioinformatics in genomic medicine
- Maintain up to date knowledge of bioinformatics developments and undertake CPD as appropriate
- Contribute to the development of professional standards of practice
- Raise awareness of the role of bioinformatics in genomic medicine to other healthcare professionals

Note: A briefing note produced by the Public Health Genomics (PHG) Foundation entitled “Defining the role of a bioinformatician” provides a general description of the different forms of bioinformatics and examines the roles of bioinformaticians in the delivery of genomic medicine (http://www.phgfoundation.org/briefing_notes/314/).

Appendix 3:

Clinical bioinformatics competency matrix

We have summarised a consensus view on the competency requirements for mid-career professionals in the roles defined by the columns. Roles are differentiated on the basis of the different levels of bioinformatics competency required, not on career stage or grade. Clearly this is a simplification, but one that we felt necessary to make our task tractable and to ensure that training to fill gaps at all career stages is developed as a result.

Where there was an even spread of opinion between three levels of competence we chose the middle level; where opinion was evenly divided between two levels of competence, we chose the higher level.

Notes

[1] Awareness required for raw data; specialist knowledge required for rare variants

[2] No knowledge required for raw data; awareness required for rare variants

[3] Specialist knowledge required for statistical/analytical epidemiologists

[4] Genetic technologists require working knowledge

[5] Genetic technologists require specialist knowledge

[6] Epidemiologists require working knowledge; statistical/analytical epidemiologists require specialist knowledge; data not collected for other roles.

Role	Clinical bioinformatician	Other bioinformatician	Specialist clinician with genomics/genomics expertise	Other specialist clinician	Other clinician	Clinical genetic Scientist	Other healthcare scientist	Specialist nurse/counsellor	Nurses and other allied health professionals	IT specialist	Data specialist
No. responses	11	6	5	5	6	8	6	6	5	7	7
Example	NHS diagnostic bioinformatician [1]	Academic bioinformatician, industry bioinformatician	Clinical geneticist or pathologist, haematologist, microbiologist with leadership responsibility in clinical lab	Cardiologist, neurologist, oncologist, paediatrician	General Practitioner	NHS diagnostic clinical scientist, microbiologist, statistical/analytical epidemiologist	Genetic technologist, Immunologist, epidemiologist	Genetic counsellor, Preimplantation genetic diagnosis nurse; clinical nurse specialist in surgery or oncology; Genetic Diabetes Nurse	Non-specialist nurse; physiotherapist	Systems administrator	Curator, data scientist
Competency	White computer programmes and algorithms that can analyse data	Specialist knowledge	Specialist knowledge	No knowledge required	No knowledge required	No knowledge required	Awareness	No knowledge required	No knowledge required	No knowledge required	Specialist knowledge
	Analyse genomics data using pre-existing software, including linking genotype to phenotype/microbial strain comparisons	Specialist knowledge	Specialist knowledge	Specialist knowledge[2]	Awareness	No knowledge required	Specialist knowledge	No knowledge required[3]	No knowledge required	Awareness	Working knowledge
	Employ good software development practice	Working knowledge	Specialist knowledge	No knowledge required	No knowledge required	No knowledge required	No knowledge required	No knowledge required	No knowledge required	Working knowledge	Specialist knowledge
	Apply computer science theory to computer system design	Working knowledge	Working knowledge	No knowledge required	No knowledge required	No knowledge required	No knowledge required	No knowledge required	No knowledge required	Specialist knowledge	Working knowledge
	Manage and organise genomics data and results	Specialist knowledge	Specialist knowledge	Awareness[2]	No knowledge required	No knowledge required	Working knowledge	Awareness	Awareness[3]	No knowledge required	Specialist knowledge
	Apply statistical research methods to genomics, medical, and population genetics	Working knowledge	Specialist knowledge	Working knowledge	No knowledge required	No knowledge required	Awareness[4]	Awareness	No knowledge required	No knowledge required	Awareness
	Use health informatics systems and understand their relevance to clinical genomics	Working knowledge	Awareness	Specialist knowledge	Awareness	Awareness	Awareness	Awareness	Working knowledge	No knowledge required	Working knowledge
	Principles of genetics, genomics and genome-sequencing technology	Specialist knowledge	Specialist knowledge	Specialist knowledge	Awareness	Awareness	Specialist knowledge	Working knowledge	Specialist knowledge	Awareness	No knowledge required
	Principles of genetic disease	Working knowledge	Working knowledge	Specialist knowledge	Working knowledge	Working knowledge	Specialist knowledge	Working knowledge	Specialist knowledge	Awareness	No knowledge required
	Principles of systems biology	Working knowledge	Working knowledge	Awareness	No knowledge required	Awareness	Awareness	Awareness	No knowledge required	Awareness	No knowledge required
	Principles of next-generation sequencing	Specialist knowledge	Specialist knowledge	Awareness	Awareness	Awareness	Specialist knowledge	Awareness[6]	Awareness	No knowledge required	No knowledge required
	Ethical, legal and social implications of clinical use of genomic data (including issues surrounding identification of patients, clinical benefits and risks, patient consent, incidental findings and ethical implications of unexpected clinically actionable findings)	Working knowledge	Working knowledge	Specialist knowledge	Working knowledge	Working knowledge	Specialist knowledge	Awareness	Specialist knowledge	Awareness	Awareness
	Interpret genetic variation in a clinical context, including understanding limitations of analysis, assessing quality and evidence for clinical interpretation	Specialist knowledge	Working knowledge	Specialist knowledge	Working knowledge	Awareness	Specialist knowledge	Awareness[5]	Specialist knowledge	No knowledge required	No knowledge required
	The role of various types of healthcare professional in genomic medicine	Working knowledge	Awareness	Specialist knowledge	Working knowledge	Working knowledge	Working knowledge	Awareness	Specialist knowledge	Awareness	Working knowledge
	The scientific discovery process and of the role of bioinformatics in it	Specialist knowledge	Specialist knowledge	Working knowledge	Awareness	No knowledge required	Working knowledge	Awareness	Awareness	No knowledge required	No knowledge required
The risks (and benefits) to patients and their families arising from the prediction of causal variants	Specialist knowledge	Awareness	Specialist knowledge	Working knowledge	Working knowledge	Specialist knowledge	Awareness	Specialist knowledge	Awareness	No knowledge required	
Integrate and jointly analyse genomic and other data						[7]	[7]				

Appendix 4: Mapping of competency to existing HEE training in genomics or bioinformatics

Key:

No training available but no requirement identified	No training provided by the courses that we mapped, but training available outside the NHS	No training course identified	Training available through other HEE routes
---	--	-------------------------------	---

Notes

1. STP genetics curriculum review is in progress
2. Could be included in Clinical Bioinformatics HSST

Competency	Level		
	Awareness	Working knowledge	Specialist knowledge
Write computer programmes and algorithms that can analyse data	Required by clinical genetic scientists ¹	MSc Clinical Bioinformatics	Required by clinical bioinformaticians ² , other bioinformaticians; data specialists
Analyse genomics data using pre-existing software, including linking genotype to phenotype/microbial strain comparisons	MSc Genomic Medicine	MSc Genomic Medicine, MSc Clinical Bioinformatics	MSc Clinical Bioinformatics
Employ good software development practice (software carpentry)	HEE Intro to Bioinformatics	Required by clinical bioinformaticians, IT specialists but available as part of specialist training or as standalone software carpentry workshops	MSc Clinical Bioinformatics
Apply computer science theory to computer system design	No training available but no requirement identified	MSc Clinical Bioinformatics	Required by IT specialists but may be provided through higher education in computer science
Manage and organise genomics data and results	HEE Intro to Bioinformatics, HEE intro to consent and ethics	MSc Clinical Bioinformatics	MSc Clinical Bioinformatics
Apply statistical research methods to genomics, medical, and population genetics	MSc Genomic Medicine	MSc Clinical Bioinformatics	Required by statistical/analytical epidemiologists
Use health informatics systems and understand their relevance to clinical genomics	HEE Intro to Bioinformatics, MSc Genomic Medicine	MSc Genomic Medicine	Required by specialist clinicians with genetics/genomics expertise; data specialists
Principles of genetics, genomics and genome-sequencing technology	HEE Intro to Bioinformatics, HEE intro to genomics, HEE intro to consent and ethics, MSc Genomic Medicine	MSc Clinical Bioinformatics, MSc Genomic Medicine	MSc Clinical Bioinformatics
Principles of genetic disease	HEE Intro to Bioinformatics, HEE intro to genomics, MSc Genomic Medicine	MSc Clinical Bioinformatics, MSc Genomic Medicine	STP Genetics, CCST Clinical Genetics, MSc Genetic Counselling
Principles of systems biology	HEE Intro to Bioinformatics	MSc Clinical Bioinformatics	No training available but no requirement identified
Principles of next-generation sequencing	HEE Intro to Bioinformatics, MSc Genomic Medicine	No training available but no requirement identified	MSc Clinical Bioinformatics
Ethical, legal and social implications of clinical use of genomic data (including issues surrounding identification of patients, clinical benefits and risks, patient consent, incidental findings and ethical implications of unexpected clinically actionable findings)	HEE Intro to Bioinformatics, MSc Genomic Medicine	HEE intro to consent and ethics, MSc Clinical Bioinformatics, MSc Genomic Medicine	STP Genetics, CCST Clinical Genetics, MSc Genetic Counselling
Interpret genetic variation in a clinical context, including understanding limitations of analysis, assessing quality and evidence for clinical interpretation	HEE Intro to Bioinformatics; HEE Intro to genomics, HEE intro to consent and ethics, MSc Genomic Medicine	MSc Clinical Bioinformatics, MSc Genomic Medicine	MSc Clinical Bioinformatics
The role of various types of healthcare professional in genomic medicine	HEE Intro to Bioinformatics, MSc Genomic Medicine	Required by clinical bioinformaticians; specialist clinicians; other clinicians; clinical genetic scientists; IT specialists	Required by specialist clinicians with genetics/genomics expertise; specialist nurses/counsellors
The scientific discovery process and of the role of bioinformatics in it	HEE Intro to Bioinformatics, MSc Genomic Medicine, MSc Clinical Bioinformatics	MSc Clinical Bioinformatics, MSc Genomic Medicine	Required by clinical bioinformaticians; other bioinformaticians
The risks (and benefits) to patients and their families arising from the prediction of causal variants	HEE Intro to genomics, MSc Genomic Medicine	HEE intro to consent and ethics, MSc Genomic Medicine, MSc Clinical Bioinformatics	MSc Clinical Bioinformatics
Integrate and jointly analyse genomic and other data	MSc Genomic Medicine	Required by epidemiologists	Required by statistical/analytical epidemiologists

Appendix 5:

Challenges in the clinical application of genome sequencing

An appreciation of the wealth and novelty of variation in any individual genome and a recognition that we can currently only interpret <0.1% of variants is crucial to the safe application of genomics in healthcare (Wright, 2013). Variants which appear harmful are found not infrequently in the genomes of healthy individuals (MacArthur, 2012). Currently we do not understand what other genetic or environmental factors may protect an individual from the consequences of an apparently pathogenic variant. In classical genetics such variants are described as incompletely penetrant. At present, we do not have accurate measures of penetrance for the vast majority of variants. Our current knowledge is heavily skewed towards pathogenicity and disease caused by ascertainment biases in the data on which much of our existing knowledge is based. The situation is compounded by significant pollution of the current literature and of resources such as the Human Gene and Mutation Database (HGMD) with spurious assignments of pathogenicity (Ref. Bell, 2013, MacArthur, 2014 & Piton, 2014).

Expert clinical evaluation, combined where possible with statistical measures to evaluate the null hypothesis that even an apparently pathogenic variant represents a chance finding not causally related to the clinical features in the patient, is therefore appropriate. For all except the tiny minority of clinically actionable recurrent, well-established disease-causing mutations, variant interpretation for clinical use is likely to require expert multidisciplinary assessment by disease specialists, clinical geneticists and clinical molecular geneticists working in partnership (MacArthur, 2014). This will be especially true where there is an intention to cascade results to other family members and undertake predictive testing in apparently healthy individuals where there is a substantial risk of amplifying errors and propagating harm to patients and their relatives. These risks should be managed through standard governance processes.

Genome sequencing has huge potential to improve the accuracy and speed of diagnosis for many patients with rare diseases. However, this diagnostic power brings with it substantial risks for misdiagnosis. It is seductively easy to weave a plausible biological link between a gene mutation and a disease. Investigating the veracity of this link by segregation studies and independent diagnostic assays is not possible for many conditions, especially those that are highly genetically heterogeneous and those for which independent modes of diagnostic confirmation do not exist. Segregation studies are often undertaken by NHS Regional Genetics services, which have a duty of care to the family but may not be considered part of routine clinical practice elsewhere in the NHS where the focus of care is on the individual patient.

Training must address the challenges of interpreting variants identified by genome sequencing to ensure patient safety. NHS staff with expertise in the interpretation of genetic variants can play a key role in disseminating this knowledge, promoting training in genomic medicine and contributing to multidisciplinary team working.

Appendix 6:

Case studies in rare disease, cancer and infectious disease

Rare disease

Jamie was born with bilateral cataracts. Congenital cataracts most often have a genetic cause, but can be associated with environmental factors such as intrauterine infection. Whilst many patients have isolated cataracts, they can form part of a broader metabolic / developmental disorder. A next generation sequencing panel test for cataracts can now sequence 150 genes simultaneously.

Clinical presentation and testing

Jamie was referred to an ophthalmologist who discussed genetic testing with his parents. A blood sample was collected while he was under anaesthetic, aged 6 weeks, for his lensectomy.

Targeted genomic sequencing of 150 genes in which mutations can cause congenital cataracts was undertaken.

Bioinformatic analysis and interpretation

Bioinformatic analysis to map the sequence reads, call and annotate variants, then filter by frequency/variant type resulted in a list of 10 variants. Two of these variants were considered to be potentially causative of congenital cataracts and confirmed by Sanger sequencing.

The first variant was a previously unreported heterozygous 3 base pair deletion in *CRYGC*, a crystallin-encoding gene in which dominantly inherited mutations cause ~33% of congenital cataracts. The second variant was a heterozygous novel missense change in *COL4A1* in a region of the gene where a mutation was previously reported to cause cataracts and small-vessel brain disease. The glycine at position 758 is highly conserved across species and lies in a conserved triple helical domain where missense mutations have previously been reported in patients with porencephaly, a condition caused by brain infarction in utero or around the time of birth.

Examination of Jamie's medical records revealed that Jamie had recently suffered a cerebral infarction. An MR scan of the brain, performed because of a prior suspicion of a structural brain abnormality on the antenatal scan, was reviewed and demonstrated enlarged ventricles and a left middle cerebral artery infarct. In view of this clinical presentation, the crystallin variant was disregarded and the *COL4A1* variant was reported as consistent with the clinical presentation.

Delivering results

Jamie and his parents were seen in the joint genetic eye clinic with the consultant geneticist, ophthalmologist and genetic counsellor. Review of the MR scans in the joint neurology meeting had confirmed porencephaly subsequent to the brain infarction. Further eye examination showed anterior segment changes causing glaucoma. This clinical picture was consistent with HANAC syndrome, a very rare condition caused by mutations in *COL4A1*. The geneticist explained the *COL4A1* variant in Jamie as a likely cause of his cataracts, glaucoma

and MR brain scan changes. The genetic counsellor also talked through the incidental finding of the crystallin variant and the lack of current knowledge around the clinical significance.

COL4A1 mutations can have a variable effect and some families show 'non-penetrance' where individuals carry the gene mutation but develop no symptoms. The parents talked with the genetic counsellor about risks to future pregnancies and were keen to be tested for the *COL4A1* variant. Blood tests from the parents showed that neither carried the *COL4A1* variant. The genetic counsellor reassured the parents that the risk of HANAC syndrome in future pregnancies was low and there was no risk to the wider family. Family studies showing that the *COL4A1* variant had arisen de novo gave further evidence of its pathogenicity.

Clinical management

Renal cysts have been described in some children with HANAC syndrome. Kidney scans were arranged and Jamie will have ongoing follow up with the neurologist, ophthalmologist and paediatrician. He will remain at risk of haemorrhagic strokes later in life.

Summary

Targeted genomic sequencing and bioinformatic analysis revealed the diagnosis of HANAC syndrome in a 4 month old baby enabling appropriate clinical management and counselling for the family.

Cancer (Case 1)

A 50 year old female with a 2 year history of lethargy and recurrent infections was admitted for pneumonia. On admission her full blood count (FBC) showed a monocytosis of $3.6 \times 10^9/l$, normal haemoglobin and platelet counts. An atypical infection was suspected, but never confirmed. The blood film was reported as leukoerythroblastic and a bone marrow examination was recommended. The bone marrow aspirate revealed some myeloid proliferation with a low percentage of blasts and no dysplasia. A bone marrow biopsy was also inconclusive. Bone marrow karyotyping and fluorescence *in situ* hybridisation (FISH) tests were normal, but a CT scan showed a 16 cm splenomegaly. Following discharge from hospital, the FBC remained abnormal and the patient developed weight loss and bone aches. The differential diagnosis of these findings included a reactive process, for example secondary to tuberculosis or a clonal disorder of the bone marrow.

A targeted NGS gene panel including genes most frequently mutated in myeloid diseases such as myelodysplastic syndrome and acute myeloid leukaemia was used to detect clonal markers in the bone marrow. Four pathogenic mutations were identified, all known to occur recurrently in myeloid disorders:

TET2 c.2145_2146delCT p.Ser716ThrfsTer6

ASXL1 c.1771_1772insA p.Tyr591Ter

SRSF2 c.284C>A p.Pro95His

SETBP1 c.2608G>A p.Gly870Ser

In addition to confirming the diagnosis of a malignant process, these mutations also provide prognostic information. The poor risk prediction from these mutations together with other signs of disease progression led to consideration of early bone marrow transplantation.

Cancer (Case 2)

A 67 year old female with chronic lymphocytic leukaemia presented with a 13cm x 12cm abdominal mass and anaemia. She received 6 treatments with FCR chemotherapy complicated by two hospital admissions for life-threatening neutropenic sepsis and pneumonia. A CT scan after treatment did not show any response. She had lost >10% of her body weight, was unable to walk, wheel chair-bound, with a life expectancy of 12 to 18 months with best supportive care. She was referred for consideration of a clinical trial for patients with *TP53* deletions or mutations, but in view of her low peripheral leukaemia count, conventional FISH/Sanger sequence analysis was not possible. The NGS gene panel confirmed presence of a pathogenic *TP53* mutation making her eligible for the clinical trial. She started on Ibrutinib and 18 months later is in remission and leading a normal life.

Infectious disease

Patient X is a recent immigrant from the Democratic Republic of Congo (DRC) who was 20 weeks pregnant and receiving antenatal care at a UK hospital. Routine serological testing of patient X indicated that she was antibody positive for HIV, but negative for HBV and HCV. She was referred to a Genitourinary Medicine (GUM) clinic where a serum sample was taken and used to establish a baseline HIV viral load and CD4 count.

Clinical presentation and testing

On interview, patient X said that she was aware of her HIV status but had received no HIV medication. She had a CD4 count of 346/ml and a viral load of 51,200 copies/ml. Given that she was pregnant, had a high viral load and low CD4 count, it was decided that patient X should begin highly active anti-retroviral therapy (HAART). It was anticipated that if successful, HAART would improve patient X's prognosis and prevent transmission of HIV to the unborn child. Prior to beginning HAART, the patient's HIV was sequenced using standard Sanger sequencing of the protease (PR) and reverse transcriptase (RT) coding regions and the whole genome was sequenced using NGS technology.

Bioinformatics analyses and interpretation

Sanger sequencing of PR and RT indicated that patient X was infected with a subtype C virus and that there was no evidence of drug resistance mutations present in the sequence.

NGS and bioinformatic analysis of the whole of the HIV virus indicated that the *pol* and *env* genes were of subtype C origin. However, there was a region within the *gag* gene that matched more closely to subtype A sequences. This indicated that X was infected with a recombinant A/C virus. Frequent detection of recombinant viruses is indicative of many African transmission events, lending weight to the likelihood that the infection occurred overseas. This finding did not rule out transmission within an immigrant community within the UK, but phylogenetic analysis of HIV sequences from within the UK did not suggest a UK-based transmission cluster.

NGS and bioinformatic analysis of the viral quasi-species within the PR- and integrase-coding regions did not show the presence of low frequency drug resistance mutations. Geno2pheno analysis of the quasi-species present within the V3 loop of the envelope protein did not indicate the presence of low-frequency CXCR4-tropic viruses. NGS analysis of RT resulted in identification of low frequency variants known to cause drug resistance to the NNRTI class of HIV inhibitors, including a Y181C mutation present in 6% of the data. The frequency of this mutation was below the limit of detection for standard Sanger sequencing. The Y181C mutation was unlikely to be a natural polymorphism and is known to revert to wild-type upon cessation of HIV treatment. The NGS analysis therefore suggested that X had received an NNRTI during the course of her infection.

Delivering results

Patient X was seen by the consultant within the GUM clinic and her medical history was discussed. X revealed that she had had another child whilst HIV positive and the consultant asked if she had been given single dose Nevirapine (an NNRTI) prior to labour to prevent

maternal transmission of HIV at birth. X was unsure, but conceded that this may have happened.

The consultant suggested that the CD4 count and viral load results indicated that patient X needed to begin HAART to protect herself and the unborn child. The consultant explained that data from recent clinical trials indicated that the presence of an NNRTI mutation (even at low frequency) meant that the standard front-line therapy of Truvada and Efavirin may be not be appropriate. The consultant explained that the Efavirin component of HAART would be replaced by a protease inhibitor because the coding region did not contain any resistance mutations.

The consultant stressed the need for X to adhere to the treatment regime being prescribed and that she should come back to the clinic if she experienced any side effects. If this were the case then the drugs could be varied or drugs targeting integrase or envelope utilised.

Clinical Management

The patient will be monitored for viral load, CD4 count and AIDS defining illnesses. NGS sequencing will be undertaken to monitor the development of drug resistance, where the viral load rebounds above 1000 copies/ml and sequence information will be used to inform the choice of drugs used within HAART.

Summary

Next Generation Sequencing and bioinformatic analysis provided information that improved the clinical management of an HIV infected patient.