Journal of Paediatrics and Child Health

doi:10.1111/jpc.12705

REVIEW ARTICLE

Phenotyping: Targeting genotype's rich cousin for diagnosis

Gareth Baynam,[1,2,3,4] Mark Walters,[5] Peter Claes,[6] Stefanie Kung,[2] Peter LeSouef,[2] Hugh Dawkins,[3,7,8] Matthew Bellgard,[9] Marta Girdea,[10,11] Michael Brudno,[10,11] Peter Robinson,[12] Andreas Zankl,[13,14,15] Tudor Groza,[16] David Gillett[5] and Jack Goldblatt[1,2]

[1]Genetic Services of Western Australia and [5]Cranio-Maxillo-Facial Unit, Princess Margaret Hospital for Children, Schools of [2]Paediatrics and Child Health and [8]Pathology and Laboratory Medicine, University of Western Australia, [3]Office of Population Health Genomics, Department of Health, Government of Western Australia, [4]Institute for Immunology and Infectious Diseases and [9]Centre for Comparative Genomics, Murdoch University and [7]Centre for Population Health Research, Curtin Health Innovation Research Institute, Curtin University of Technology, Perth, Western Australia, [14]Discipline of Genetic Medicine, University of Sydney and [13]Academic Department of Medical Genetics, Western Sydney Genetics Program, Sydney Children's Hospital Network, Sydney, New South Wales, [15]Centre for Clinical Research and [16]School of Information Technology and Electrical Engineering, University of Queensland, Brisbane, Queensland, Australia, [6]Medical Imaging Research Centre, Faculty of Engineering, KU Leuven, Leuven, Belgium, [10]Department of Computer Science, University of Toronto and [11]Centre for Computational Medicine, Hospital for Sick Children, Toronto, Ontario, Canada and [12]Institute of Medical Genetics and Human Genetics, Charité – Universitätsmedizin, Berlin, Germany

Abstract: There are many current and evolving tools to assist clinicians in their daily work of phenotyping. In medicine, the term 'phenotype' is usually taken to mean some deviation from normal morphology, physiology and behaviour. It is ascertained via history, examination and investigations, and a primary aim is diagnosis. Therefore, doctors are, by necessity, expert 'phenotypers'. There is an inherent and partially realised power in phenotypic information that when harnessed can improve patient care. Furthermore, phenotyping developments are increasingly important in an era of rapid advances in genomic technology. Fortunately, there is an expanding network of phenotyping tools that are poised for clinical translation. These tools will preferentially be implemented to mirror clinical workflows and to integrate with advances in genomic and information-sharing technologies. This will synergise with and augment the clinical acumen of medical practitioners. We outline key enablers of the ascertainment, integration and interrogation of clinical phenotype by using genetic diseases, particularly rare ones, as a theme. Successes from the test bed or rare diseases will support approaches to common disease.

Key words: deep phenotyping; general paediatrics; genetics; international child health; phenotype; precision medicine.

Phenotyping is a cornerstone of a doctor's daily work. In medicine, the term 'phenotype' is used to indicate manifestations resulting from deviation from normal morphology, physiology or behaviour.[1] It is ascertained via history, examination and investigations with the principal aim of making a diagnosis. Phenotypic information associated with the diagnosis is then used for selecting and monitoring response to treatment. Thus, doctors are by necessity expert 'phenotypers'. Fortunately, there is an expanding collection of phenotyping tools, such as next generation sequencing (NGS, Appendix S1), that are poised to both synergise with the clinical acumen of doctors and improve the utility of genetic testing technology.

Key Points

1. Clinical phenotypic information, including history, examination and investigations, is increasingly important for promoting clinical utility of advances in genomics technologies.
2. Rare diseases provide the ideal test bed for ascertainment and integration of clinical phenotype for diagnostics and treatment monitoring.
3. Approaches that support precise objective assessments and data sharing and that foster citizen science have the potential to promote step changes in diagnostics; particularly in rare diseases.

Correspondence: Dr Gareth Baynam, School of Paediatrics and Child Health, University of Western Australia, Perth, WA 6151, Australia. Fax: 61893401678; email: gareth.baynam@health.wa.gov.au

Conflict of interest: None declared.

Accepted for publication 11 July 2014.

There is an inherent power in phenotypic information, and its potential has only been partly realised. In addition, phenotypic data are mandatory for the interpretation of genetic tests. For single-gene tests, such as testing of the gene for fibrillin-1 in suspected Marfan syndrome, phenotyping is crucial to refining the diagnosis to improve the chance that one has selected the correct gene and to improve the pre-test probability, that is, to develop an a priori understanding of the likelihood that a given identified variant in a gene is the cause of the observed phenotype (disorder). There are numerous examples of genetic variants being erroneously described as disease-causing pathogenic variants when they are either benign variants or variants of unknown significance. One study showed that about 10% of literature-annotated disease mutations in commonly used databases are wrong.[2] Errors in interpretation of the significance of a genetic variant can have marked consequences for clinical care when treatment decisions are based on these results. Increasing

the number of genes tested increases the probability of detecting variants of uncertain (or misattributed) significance. An average human genome has 3–4 million variants, and the average file size of stored information from whole-genome sequencing can be on the order of 100 or more GB. Deciphering which variants are harmless and which could cause disease requires extensive filtering. Phenotypic information is essential to this filtering. Technological advances and cost reductions for sequencing have outpaced developments in the ability to clinically interpret the sequencing data. Therefore, improvements in phenotyping approaches are required to support clinical utility; that is, 'next generation sequencing demands next generation phenotyping'.[3] Given limited health budgets, this will also be important for health system efficiencies.

In this paper, we outline aspects of current and emerging key enablers of phenotyping. We use genetic diseases, particularly rare ones, as a theme for this approach, as (i) there are a number of tools that are designed for this context; (ii) individual rare diseases provide a homogeneous window, 'the clarity in the extremity', that provides insights into common diseases; and (iii) the rarity of individual conditions has promoted an environment characterised by collaboration and community engagement. Rare diseases are a fecund test bed for the development of phenotypic tools that can be expanded to common diseases.

Below we outline factors that can improve the utility of medical phenotyping.

Accurate Description and Organisational Structure

Medical terminologies are embedded in clinical care. However, there are a number of challenges to their clinical utility. These include imprecise terms, the use of synonyms, and the lack of sufficient granularity (depth) of vocabularies to allow accurate description of individual phenotypic features (e.g. individual signs or symptoms or other disease manifestations).[4] Failure to address these challenges impairs exchange of information between clinicians and prohibits integrative analyses, for example with genetic test results.[5]

Ontologies are necessary tools for addressing these hurdles. An ontology provides a classification of the entities (terms) within a domain of knowledge (e.g. human (medical) phenotype) and specifies the semantic relationships between these entities. The Human Phenotype Ontology (HPO) contains more than 10 000 phenotypic terms. It was originally constructed from data from the Online Mendelian Inheritance in Man database, the most established source of information on genetic diseases. The HPO has been incorporated in to a number of international rare diseases initiatives (e.g. Care4Rare Canada and Australia, the US National Institutes of Health (NIH) Undiagnosed Diseases Program, the International Standards for Cytogenomic Arrays Consortium, and many institutes for human genetics). It is continually refined with a focus on clinical utility (see Table 1 for this and selected platforms and resources).

Synergistic developments include OrphaCodes,[6] a coding system for rare diseases; the Elements of Morphology,[7] standard descriptors for dysmorphology (abnormal form); and objective, deeply precise phenotyping measures such as 3D facial analysis.[8,9]

Efficient and Robust Access

As clinicians are time-poor, any new approach to the documentation of phenotype must, at a minimum, be time-neutral and provide added functionality. A number of platforms are under development. If designed and implemented appropriately, they have the capacity to streamline clinical workflow, standardise collection of patient phenotypes and promote time-efficient data entry and sharing. An example is PhenoTips, an open-source customisable platform for collecting and analysing the phenotypes of patients with genetic conditions. It has an intuitive user interface that can run on any device with a Web browser. To facilitate entry at the point of patient care, it is aligned to clinical workflow, and some tasks, such as plotting of growth charts, are automated. It incorporates the HPO to facilitate inbuilt computer-assisted diagnostics (decision support). Implementation sites include the Hospital for Sick Children, Toronto, Canada; the NIH Undiagnosed Diseases Program[10]; Charité Hospital, Berlin; the Western Australian Department of Health; and others. Another example is the Skeletome knowledgebase,[10] a platform for collecting phenotypic information on patients with rare skeletal disorders; it has been expanded to non-skeletal conditions. It selects phenotypic terms automatically from free text, such as clinical letters. This platform also allows sharing and discussion of cases with other users, for example to obtain an expert opinion from a colleague overseas.

Data Sharing

Phenotyping remains a cornerstone for diagnosis, and in many cases, phenotype data are the most powerful predictors of natural course, therapeutic response and mortality.[11] Importantly, the relationship between genotype and phenotype, including disease states, can be circuitous and non-linear. It is modified by multiple internal biological processes (at the level of non-coding mRNA, proteins, etc.) and external (environmental) influences. Analysis at the phenotypic level allows for the assessment of the integrated endpoint of these various phenomena in a manner that is agnostic to the, often unknown, details of these processes. Identification of phenotypic patterns can be used for diagnosis and treatment monitoring, which can allow thin-slicing for clinical solutions.

An example of the power of phenotype sharing is DYSCERNE. This expert network of clinical dysmorphologists operated via a web service to diagnose rare dysmorphic disorders that have already been assessed by local and/or national specialists. A study of DYSCERNE found that remarkably, a consensus clinical diagnosis was achieved in 22.5% of cases.[12] This emphasises that coordinated collaborative assessment can support diagnosis, even in the most diagnostically intractable cases. This is corroborated by a study of the European Skeletal Dysplasia Network in which mutation detection rate increased from 36% to 81% when cases were reviewed by the expert panel prior to testing.[13] Ultimately, it may be the combination of human networks and computer-assisted analysis that brings most benefit.

Table 1 Tools, databases and platforms

Resource	Aim	Comments	Website
Human Phenotype Ontology	Standardised vocabulary of phenotypic abnormalities in human disease	A basis for clinical diagnostics in human genetics (e.g. Phenomizer), investigating the relationships between human phenotypic abnormalities and cellular and biochemical networks, and for providing a standardised vocabulary for clinical databases.	http://www.human-phenotype-ontology.org/ http://compbio.charite.de/phenomizer/
Elements of Morphology	Standardisation of terms used to describe human morphology	Developed as a consensus terminology by an expert group of clinical dysmorphologists; overlapping applications with the Human Phenotype Ontology and embedded in the former.	http://elementsofmorphology.nih.gov/
Online Mendelian Inheritance in Man	An online catalogue of human genes and genetic disorders	Curated summaries of phenotypic and genotypic information of human genes and genetic diseases. It includes clinical synopses of genetic diseases.	http://www.omim.org/ http://www.omim.org/search/advanced/clinicalSynopsis
Orphanet	A portal for rare diseases and orphan drugs	Inventory, classification and encyclopaedia of rare diseases. Assistance-to-diagnosis tool. Emergency and other guidelines. Inventory of orphan drugs. Directory of medical laboratories providing diagnostic tests. Directory of expert centres.	http://www.orpha.net/consor/cgi-bin/index.php http://www.orpha.net/national/AU-EN/index/homepage/
PhenoTips	Open source software for collecting and analysing the phenotypes of patients with genetic disorders	The user interface mirrors clinician workflows so as to facilitate the recording of observations made during the patient encounter. This easy-to-use front-end, compatible with any device that runs a Web browser, is coupled with a standardised database back-end where phenotypic information is represented using the Human Phenotype Ontology.	http://phenotips.org/
PhenomeCentral	A hub for secure data sharing within the rare-disorder community	A repository for secure data sharing targeted to clinicians and scientists working in the rare-disorder community. It encourages global scientific collaboration while respecting the privacy of patients profiled in this centralised database. Once users enter their patients' data, they are connected to other patient profiles within PhenomeCentral that share similar phenotypes and genotypes. It enables the discovery of multiple individuals affected by the same unnamed disorder, which ultimately may help with diagnosis.	https://phenomecentral.org/
Bio-LarK	Developing and aggregating platforms and tools to enable clinicians and scientists to represent, acquire and process biomedical data, using Semantic Web technologies	Progressively, it will comprise ontologies focused on representing and modelling disorder–phenotype–genotype relations, text mining, and concept recognition focused on disorders and phenotypes. It is creating diverse knowledge discovery and exploration tools to analyse and interpret data drawn from heterogeneous information sources, including clinical information like that found in clinical letters.	http://www.bio-lark.org/
Skeletome Knowledge Base	A community-driven knowledge curation platform within Bio-LarK	Provides information on all recognised bone dysplasias. Entries are continuously reviewed and updated by the global community of clinicians and researchers. It makes extensive use of ontologies to standardise the entered information and make it accessible to computational analysis (i.e. it can assist with diagnosis by, for instance, automatically identifying phenotypic terms in free text descriptions, that is, patient notes).	http://knowledge.skeletome.org/
Skeletome Archive	A platform within Bio-LarK	Allows sharing and discussion of entered cases with other users, for example to obtain an expert opinion from a colleague overseas.	http://skelarch.skeletome.org/
Phylo	A framework for harnessing everyday citizens for their pattern recognition ability	Not directly related to diagnosing conditions. Essentially provides an enormous free computing resource for a specific, otherwise computer-intensive task (i.e. multiple sequence alignment) and a way to inform developments of computer algorithms to improve their performance.	http://phylo.cs.mcgill.ca/

Matchmaking phenotypes can enable diagnosis. This is of particular relevance for rare diseases where there may be limited or no experience of that disorder within an institution or country. An increased diagnostic certainty can be achieved by matching features of one individual to one or more others previously documented. For instance, PhenomeCentral confidentially shares a limited amount of information about the features of an individual's rare condition to assist in making a diagnosis. This limited subset of information (without any personal identifiers) is shared with other doctors and scientists who are attempting to identify the causes of rare diseases. An automated feature-matching system identifies patients with the same or very similar features and informs the contributing doctors of matched individuals. This is performed without revealing additional, potentially identifiable information about the individuals or the doctors who contributed that information. The PhenomeCentral system can subsequently facilitate direct communication between individual doctors for further information sharing to assist in reaching a diagnosis.[14] The Skeletome platform offers similar features.

Development of Non-Invasive Precision Phenotyping Methods

In addition to subjective and expert-dependent clinical phenotyping, methods to objectively determine phenotypes with high precision are required. Such methods will optimally be non-irradiating, portable, scalable and relatively inexpensive. A picture is worth a thousand words; in the case of 3D facial analysis, it can be worth 10 000[15] or more data points. 3D facial analysis has been investigated for establishing facial signatures of rare diseases[8] as well as for monitoring therapies of rare disorders.[16] Other methods, including movement analysis (for instance, using motion-sensing capabilities of commercial gaming consoles) and voice analysis, are also promising. Such approaches may also provide the non-expert with tools that objectively capture standardised data and enable access to expert knowledge. Additionally, analogously to genetic reference sequences, which vary by population, phenotypic references (normal ranges) will vary by population, and it will be important to systematically document this diversity.[17]

The Marriage of Phenotype and Genotype

Rapid technological advances, including genome-wide assessments such as chromosomal microarray (CMA) and NGS technologies, are being applied to clinical need. Documenting and understanding the clinical phenotype has always been crucial to assessing whether an identified genetic variant is causative of the disorder under investigation; this is increasingly the case with technologies that provide genome-wide analyses. Accordingly, the ability to interpret CMA is improved when it is coupled with clinical phenotypic information,[18] and phenotype-centred analysis markedly improves interpretation of NGS data.[5] For immediate clinical utility, the simple act of providing more clinical information on a laboratory request form can increase diagnostic yield and reduce the likelihood of receiving a test result of uncertain significance. Thus, the 'marriage' of phenotype and genotype is better than either in isolation.

Also, as there are a number of existing and emerging tools for analysing phenotypic and genotypic information, it is imperative that software approaches be created to deliver, when required, combinations of these tools, and that they can be adjusted as new tools emerge.[19,20]

Future Avenues: A Role for Citizen Science?

Citizen science includes 'the general public engagement in scientific research activities when citizens actively contribute to science either with their intellectual effort or surrounding knowledge or with their tools and resources'.[21] Drawing from experience in rare disease registries, data can be directly and reliably submitted by patients,[22,23] which can potentially save clinician time and increase the depth of information ascertained. Furthermore, citizen science approaches can be used to exploit the human computational power that is used every day in playing games. One example is Phylo,[24] a gamified approach for multiple sequence alignments (MSA). MSA is a computationally intensive task that is useful for investigating the causes of various genetic disorders. By turning this process into a visual pattern matching game, individuals with no formal scientific knowledge overcame a computational resource problem to enable medical research. Additionally, by analysing how these individuals reconciled these visual puzzles, algorithms could be refined to enhance computer processing. It is feasible that similar approaches that harness human pattern-matching and problem-solving capabilities could be used to match diagnoses to lists of signs or symptoms or to identify patterns of facial variation (facial dysmorphology) to assist with diagnosis of rare dysmorphic conditions.[8]

Conclusion

Systematic and precise phenotyping is part of the solution to current global challenges for genetics in medicine. However, and acknowledging significant efforts in this area, it might be perceived that it has not attracted the same level of interest and resources as genomic investigations. Ultimately, capturing precise phenotypes in an efficient pipeline will increase diagnostic yield in and of itself and enhance the utility of genomic investigations by refining analysis and decreasing downstream costs. It will also be crucial for documenting natural history, monitoring treatment and identifying homogeneous cohorts to improve the chance and reduce the cost of successful drug trials.[25] In many instances, clinical phenotypic data are already being obtained within the clinical workflow and can be ascertained and stored with less capital cost than some other types of data. While the potential of improved phenotyping is broad and will likely be realised across a range of time frames, arguably the most promising area for immediate translation to clinical utility is the diagnosis of rare diseases. Many rare diseases have distinctive manifestations that can help separate them out from the 'noise' of common variation. For instance, in highly defined (phenotyped) rare disease cohorts, the diagnostic yield of NGS approaches 40%[26]; with less specific phenotypes the rate of diagnostic success, and therefore cost efficiency, is far smaller. Additionally, by virtue of factors experienced by individuals

living with rare diseases and their families, including the 'diagnostic odyssey', they may be particularly motivated to participate in citizen science to provide novel solutions. Successes from the test bed of rare diseases could then be promulgated for common diseases.

Multiple Choice Questions

1. Which of the following statements about clinical phenotype is false?
   a. The term 'phenotype' is used to indicate some deviation from normal morphology, physiology or behaviour.
   b. Genotype invariably predicts phenotype.
   c. Objective assessment of phenotype can facilitate diagnostics.
   d. There are standard descriptors for dysmorphology.
   e. Ontologies are tools that are necessary to promote efficient use of phenotype.

a. True. This is a definition of clinical phenotype.

b. False. The path between genotype and phenotype is not linear, even in monogenic disorders. There are many interposing regulatory layers between gene sequence and phenotype. There are multiple factors that modify phenotype, including but not limited to genetic variants (e.g. at other genetic loci), epigenetic factors (changes in gene expression without change in DNA sequence, e.g. methylation) and environmental factors (e.g. cigarette smoking and obesity). These factors influence when, how and if a disorder manifests.

c. True. Reduced subjectivity can remove biases and promote phenotypic precision; this can be as simple as clinical measurements (e.g. head circumference).

d. True. The Elements of Morphology (http://elementsofmorphology.nih.gov/) were created by expert consensus to promote standardised dysmorphic descriptions for facial and other body regions.

e. True. An ontology provides a classification of the entities (terms) within a domain of knowledge (e.g. human (medical) phenotype) and specifies the semantic relationships between these entities. This provides a standard, controlled vocabulary for a scientific field that facilitates precise and unambiguous description within that field and that enables computer-assisted diagnostics.

2. Which of the following best describes an environment that impairs the clinical and computational use of phenotype?
   a. Imprecise terms.
   b. The use of synonyms.
   c. Lack of sufficient granularity of vocabularies to allow accurate description of individual phenotypic features.
   d. The use of bundled terms.
   e. All of the above.

e. Factors that reduce phenotypic ambiguity (as may occur with a, b, d) and facilitate depth (granularity) of phenotyping (e) promote diagnostic ability. For instance, 'fibrillation' has markedly different diagnostic and management implications if it is occurring in skeletal versus cardiac muscle.

3. Which of the following statements about clinical phenotyping and diagnostics is false?
   a. Phenotype needs to be robustly and efficiently accessible.
   b. Advances in genetic sequencing technologies, such as next generation sequencing, will replace the need for phenotyping.
   c. Computers will not replace doctors for diagnosis.
   d. Ethnicity is an important factor.
   e. There is a potential role for citizen science.

a. True. As discussed above, precise phenotyping is important. Also, it is best ascertained in a manner that is, where possible, standardised and time-efficient. It is crucial that approaches to phenotyping mirror clinical workflow and do not unduly add to the time of clinical assessment, as this will incur an opportunity cost (i.e. morbidity and mortality).

B. False. Genotype is a very important but only partial influence on disease. There are many known non-genotypic factors that can influence clinical phenotype, for example environmental and epigenetic factors that are not assessed by sequencing. Also, there are types of mutation that are currently a challenge for sequencing technologies, such as triplet repeat disorders, including fragile X syndrome. There are numerous conditions for which the genetic basis remains to be identified or for which mutations have been identified only in a proportion of individuals. In these instances, diagnosis is made by clinical phenotype alone. For instance, in Noonan syndrome, even with mutation analysis of the multiple genes known to be associated with this condition, approximately one-quarter of individuals will not have an identifiable mutation. Additionally, the clinical phenotype of the proband and other family members is important, as it may be crucial for interpreting whether a given genetic variant is pathogenic (causative of the disease) or not. For instance, if a given genetic variant is found in the patient with the particular disease and is also present in an unaffected parent, then it is less likely to be the cause of the disease under investigation.

C. True. Computers can assist with diagnostics and may be particularly helpful for those who are non-expert in a particular clinical domain; however, they are unlikely, particularly in the intermediate term, to replace a nuanced assessment of the clinical gestalt.

D. True. The basis of the ability to detect abnormality (e.g. disease) is dependent on understanding the range of normality (e.g. reference ranges). In clinical genetic assessment, this is relevant at the level of both genetic sequence and phenotype. When an apparently novel genetic variant is found in a population with limited genetic reference information (e.g. in an Aboriginal Australian), it may potentially be erroneously considered to be causative of the disorder under investigation when it is in fact a relatively common variant in that population and not causative of that disorder. This also applies to phenotypic information; for example, synophrys (meeting of the eyebrows in the midline; http://elementsofmorphology.nih.gov/index.cgi?tid=5e417df50b2316f4) is a common, normal variant in some ethnic groups, although it may also, in combination with other findings, be indicative of one of a number of genetic syndromes, such as Cornelia de Lange syndrome.

E. True. Increasingly, individuals are becoming experts in their or their family member's condition, particularly in rare diseases; they have the lived experience of the manifestations (phenotype) of their or their family member's condition. If this information can be better captured from the individuals themselves, this may enhance diagnosis. Also, there are an increasing

number of crowdsourced pattern-matching initiatives that are being applied to provide scientific and medical solutions; this could potentially be harnessed for identifying clinical patterns (diagnoses and treatment response).

References

1 Robinson PN. Deep phenotyping for precision medicine. *Hum. Mutat*. 2012; **33**: 777–80.
2 Bell CJ, Dinwiddie DL, Miller NA *et al*. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci. Transl. Med.* 2011; **3**: 65ra4.
3 Hennekam RC, Biesecker LG. Next-generation sequencing demands next-generation phenotyping. *Hum. Mutat.* 2012; **33**: 884–6.
4 Robinson PN, Mundlos S. The human phenotype ontology. *Clin. Genet.* 2010 Jun; **77**: 525–34.
5 Robinson P, Kohler S, Oellrich A *et al*. Improved exome prioritization of disease genes through cross species phenotype comparison. *Genome Res.* 2013; **2**: 340–8. PubMed PMID: 24162188.
6 Orphanet. *Orphanet: The Portal for Rare Diseases and Orphan Drugs*. 2013. Available from: http://www.orpha.net/consor/cgi-bin/index.php [accessed 5 June 2014].
7 Carey JC, Cohen MM Jr, Curry CJ, Devriendt K, Holmes LB, Verloes A. Elements of Morphology: standard terminology for the lips, mouth, and oral region. *Am. J. Med. Genet. A* 2009; **149A**: 77–92.
8 Baynam G, Walters M, Claes P *et al*. The facial evolution: looking backward and moving forward. *Hum. Mutat.* 2013; **34**: 14–22.
9 Hammond P, Suttie M. Large-scale objective phenotyping of 3D facial morphology. *Hum. Mutat.* 2012; **5**: 817–25. PubMed PMID: 22434506.
10 Girdea M, Dumitriu S, Fiume M *et al*. PhenoTips: patient phenotyping software for clinical and research use. *Hum. Mutat.* 2013; **34**: 1057–65.
11 Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. *Nat. Rev. Genet.* 2010; **11**: 855–66.
12 Douzgou S, Clayton-Smith J, Gardner S *et al*. Dysmorphology at a distance: results of a Web-based diagnostic service. *Eur. J. Hum. Genet.* 2013; **3**: 327–32. PubMed PMID: 23838594.
13 Zankl A, Jackson GC, Crettol LM *et al*. Preselection of cases through expert clinical and radiological review significantly increases mutation detection rate in multiple epiphyseal dysplasia. *Eur. J. Hum. Genet*. 2007; **15**: 150–4.
14 PhenomeCentral. *PhenomeCentral: A Hub for Secure Data Sharing Within the Rare Disorder Community*. 2013. Available from: https://phenomecentral.org/ [accessed 5 June 2014].
15 Kung S, Walters M, Claes P, Goldblatt J, Le Souef P, Baynam G. A dysmorphometric analysis to investigate facial phenotypic signatures as a foundation for non-invasive monitoring of lysosomal storage disorders. *JIMD Rep.* 2013; **8**: 31–9.
16 Baynam GS, Walters M, Dawkins H, Bellgard M, Halbert AR, Claes P. Objective monitoring of mTOR inhibitor therapy by three-dimensional facial analysis. *Twin Res. Hum. Genet*. 2013; **16**: 840–4.
17 Baynam GS. The need for genetic studies of Indigenous Australians. *Med. J. Aust*. 2012; **196**: 313.
18 Riggs ER, Jackson L, Miller DT, Van Vooren S. Phenotypic information in genomic variant databases enhances clinical care and research: the International Standards for Cytogenomic Arrays Consortium experience. *Hum. Mutat*. 2012; **33**: 787–96.
19 Hunter AA, Macgregor AB, Szabo TO, Wellington CA, Bellgard MI. Yabi: an online research environment for grid, high performance and cloud computing. *Source Code Biol. Med.* 2012; **7**: 1–10.
20 Bellgard MI, Macgregor A, Janon F *et al*. A modular approach to disease registry design: successful adoption of an internet-based rare disease registry. *Hum. Mutat*. 2012; **33**: E2356–66.
21 SOCIENTIZE Consortium. Citizen science for Europe: towards a better society of empowered citizens and enhanced research. Zaragoza, Spain: SOCIENTIZE Consortium, 2013.
22 Cart-Wheel. *Cart-Wheel: Center for Analysis of Rare Tumors*. 2013. Available from: https://www.cart-wheel.org/ [accessed 5 June 2014].
23 Reg4ALL. *Reg4ALL Beta*. 2013. Available from: https://www.reg4all.org/ [accessed 4 June 2014].
24 Kwak D, Kam A, Becerra D *et al*. Open-Phylo: a customizable crowd-computing platform for multiple sequence alignment. *Genome Biol.* 2013; **14**: R116.
25 Fishman MC. Power of rare diseases: found in translation. *Sci. Transl. Med.* 2013; **5**: 201ps11.
26 Sawyer SL, Schwartzentruber J, Beaulieu CL *et al*. Exome sequencing as a diagnostic tool for pediatric-onset ataxia. *Hum. Mutat*. 2013; **1**: 45–9. PubMed PMID: 24108619.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Glossary and related explanatory resources.