Journal of Human Genetics (2011), 1–3

www.nature.com/jhg

SHORT COMMUNICATION

Exome sequencing can detect pathogenic mosaic mutations present at low allele frequencies

Alistair T Pagnamenta[1,6], Stefano Lise[1,6], Victoria Harrison[2], Helen Stewart[2], Sandeep Jayawant[3], Gerardine Quaghebeur[4], Alexander T Deng[1], Valerie Elizabeth Murphy[1], Elham Sadighi Akha[1], Andy Rimmer[1], Iain Mathieson[1], Samantha JL Knight[1], Usha Kini[2], Jenny C Taylor[1] and David A Keays[5]

The development of next generation sequencing (NGS) has radically transformed the scientific landscape, making it possible to sequence the exome of any given individual in a cost-effective way. The power of this approach has been demonstrated by a number of groups who have identified pathogenic mutations in small pedigrees that have been resistant to traditional genetic mapping. Recently it has become clear that exome sequencing has great potential with respect to sporadic disease and the identification of *de novo* mutations. This is highlighted by studies reporting whole-exome sequencing of patient–parental trios affected by learning disability, autism and schizophrenia. It is widely anticipated that the introduction of this technique into a clinical setting will revolutionise genetic diagnosis. However, the sensitivity of NGS exome sequencing is currently unclear. Here, we describe the exome sequencing of DNA samples from a patient with double cortex syndrome and her parents, resulting in the detection of a mosaic splicing mutation in *LIS1*. This variant was found at an allele frequency of just 18%, demonstrating that NGS methods have the capacity to identify pathogenic mosaic mutations present at a low level.
Journal of Human Genetics (2011) **0,** 000–000. doi:10.1038/jhg.2011.128

Keywords: next generation sequencing; mosaic; somatic; *PAFAH1B1*; neuronal migration

As part of an ongoing project to identify genes that cause neuronal migration disorders, we have been sequencing parental–patient trios where the affected individual presents with a sporadic disease state. Included in this study was a 7-year-old girl (BRC006) who presented with global developmental delay, impaired fine motor skills, and generalised tonic–clonic seizures. Her parents were non-consanguineous, and there were no complications in pregnancy. A magnetic resonance imaging (MRI) brain scan showed subcortical band heterotopia (double cortex) with a posterior to anterior gradient (Figure 1a). Molecular testing of *LIS1* and *DCX* was performed by an accredited clinical genetics laboratory who undertook Sanger sequencing of all coding exons with additional screening for copy-number variants (CNVs) using MLPA (P061-B1, MRC Holland, Amsterdam, the Nether lands). No mutations of clinical significance were found. In the absence of a specific genetic diagnosis, the parents were counselled as having a 5–10% recurrence risk, and a fetal MRI scan was suggested as a form of prenatal testing in future pregnancies. The patient was recruited into our study.

We performed genome-wide array-based comparative genomic hybridisation to identify any potential pathogenic CNVs. No losses or gains considered clinically relevant were identified. Next, we carried out whole-exome sequencing of DNAs extracted from patient and parental blood samples. Libraries were prepared using standard procedures, captured using the SureSelect Human all exon kit (Agilent Technologies, Santa Clara, CA, USA) and then sequenced using Illumina technology. The 51 bp paired-end reads were mapped using Stampy[1] and duplicate PCR reads were removed using Picard (http://www.picard.sourceforge.net/). We obtained a mean coverage 45.6× across the target region, with 85.3% of bases covered at 10× or greater (Table 1). Single-nucleotide variants were called using an in-house algorithm called Platypus. Platypus identifies single-nucleotide polymorphisms and short indels by locally realigning reads around candidate variants to find the most probable haplotype. Spurious calls from known error modes are filtered out, but unlike most other callers, there is no explicit requirement for a variant to be seen in a particular percentage of reads. This means that high quality variants can be called even if, as in this case, they appear in far fewer than 50% of the reads. After filtering out synonymous variants, and those variants observed at an allele frequency >5% (1000 genomes, June 2011 data release), 1241 variants remained. Comparison with parental exome data permitted exclusion of >99% of the rare functional variants seen in the proband. Similarly to Vissers *et al.*,[2] candidate

[1]NIHR Biomedical Research Centre, Wellcome Trust Centre for Human Genetics, Oxford, UK; [2]Department of Clinical Genetics, Oxford Radcliffe NHS Trust, Oxford, UK; [3]Department of Neurology, John Radcliffe Hospital, Oxford, UK; [4]Department of Neuroradiology, John Radcliffe Hospital, Oxford, UK and [5]Department of Neurogenetics, Research Institute of Molecular Pathology, Vienna, Austria
[6]These authors contributed equally to this work.
Correspondence: Dr DA Keays, Research Institute of Molecular Pathology, Dr-Bohr Gasse 7, Vienna 1030, Austria.
E-mail: keays@imp.ac.at
Received 14 September 2011; revised 17 October 2011; accepted 20 October 2011

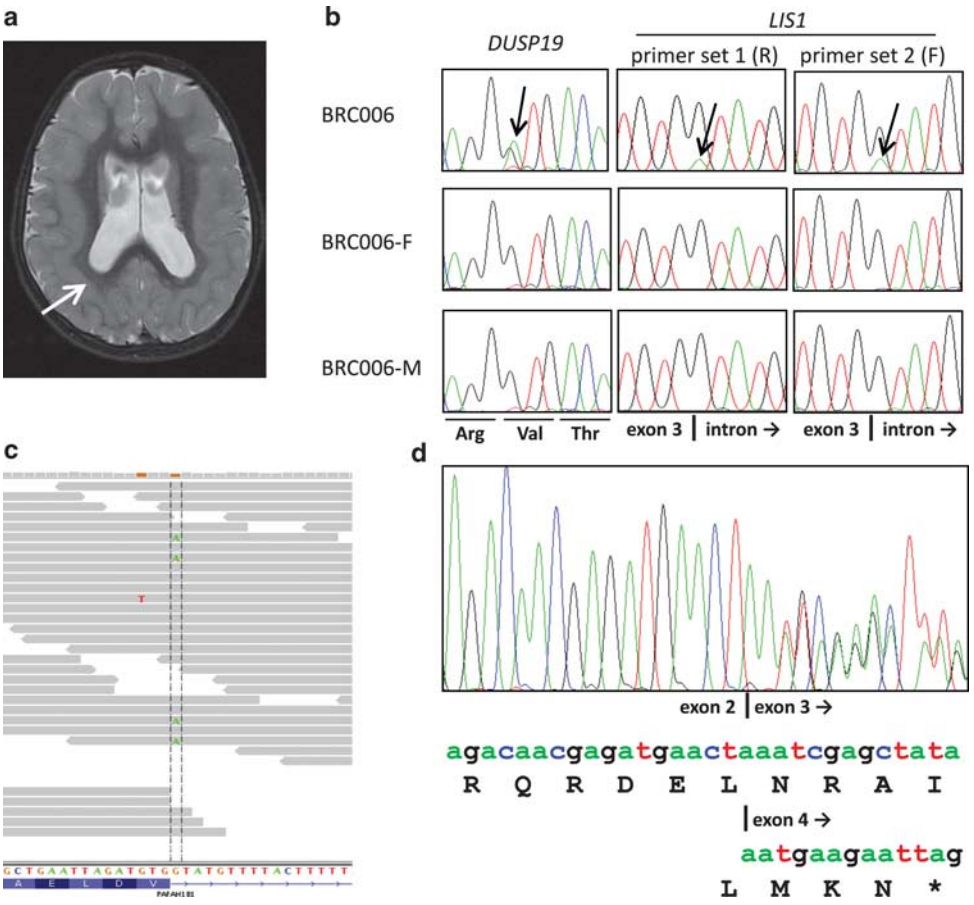

Figure 1 (**a**) Axial T2-weighted MRI image of the patient (BRC006). The white arrow shows the region of 'double cortex', predominant in the posterior hemispheres. (**b**) Sanger sequence analysis of the *DUSP19* variant (G>A) and the *LIS1* splicing mutation (c.117+1G>A) for the patient (BRC006), her father (BRC006-F) and her mother (BRC006-M). Black arrows show the *de novo* mutations in the patient. Quantification of the *LIS1* Sanger reads estimates the level of mosaicism to be 18%. (**c**) Illumina data showing 4 of 25 reads (16%) supporting the c.117+1G>A mutation. (**d**) Sanger sequencing of reverse transcriptase PCR products shows that the c.117+1G>A mutation leads to skipping of exon 3, and is predicted to result in three novel amino acids followed by a stop codon.

Table 1 Summary of exome data for BRC006

Data generated	*Mean coverage of target*	*Target bases covered 10× or more*	*Target bases covered 20× or more*	*Target bases covered 30× or more*	*Rare functional SNVs*	*Validated de novo mutations*
4.49 Gb	45.6×	85.3%	72.8%	60.8%	1241	2 (*DUSP19* and *PAFAH1B1*)

Abbreviation: SNVs, single-nucleotide variants.

variants were assumed to be inherited if present in at least one sequencing read from the proband's parents. This resulted in the identification of three putative *de novo* mutations in three different genes: *LIS1*, *DUSP19* and *EGFL7*.

Sanger sequencing validated the mutations in *LIS1* and *DUSP19*, but not in *EGFL7*. The *DUSP19* variant, a G>A transition in exon 1 (chr2:183,943,728, hg19/NCBI build 37), results in a V23M mutation (Figure 1b). The *DUSP19* gene encodes a protein phosphatase that is ubiquitously expressed and has no previous association with neurodevelopmental disease.[3] The second validated *de novo* variant identified was located in the donor site of intron 3 of *LIS1* (c.117+1G>A) and has been described previously in two patients with lissencephaly.[4] The mutation was only seen in 4/25 (16%) Illumina reads mapping to this region (Figure 1c). Assuming a binomial distribution for the number of reads with the mutant allele, the 95% confidence interval is 5–36%. This is consistent with the Sanger data, which gave a mean allele frequency of 18% (range 14–23%, Figure 1b, Supplementary Table S1). We therefore conclude that, that the patient is mosaic for c.117+1G>A. To determine the effects of this mutation, leukocyte RNA was extracted from fresh blood. Reverse transcriptase PCR from exons 1–5, followed by Sanger analysis showed that the c.117+1G>A mutation results in skipping of exon 3, introducing a premature stop codon (Figure 1d).

It is well established that both somatic and germline mutations in *LIS1* result in either lissencephaly or double cortex syndrome. LIS1 mediates neuronal migration by interacting with Ndel1 and dyenin,

facilitating nuclear translocation by modification of the microtubule cytoskeleton.[5,6] To date, relatively few somatic mosaic mutations have been detected in *LIS1*. Guerrini *et al*.[7] reported two patients with double cortex syndrome that harbour mosaic mutations in *LIS1* (R241P and R8X), and Boycott *et al*.[8] reported an affected child who had inherited a H389Y mutation from his mosaic mother. Our ability to detect the mosaicism in this case was because of the filtering strategy we employed when analysing our next generation sequencing (NGS) data. Although many studies include filters that exclude variants seen in $<$15–20% of reads,[2] the filters we used did not. Our results show that such filters, although excluding unwanted NGS artefacts, may not be beneficial as it compromises the sensitivity of NGS methods.

In summary, this study highlights the utility of whole-exome sequencing as a diagnostic tool specifically demonstrating that it has the sensitivity to detect mosaic mutations of clinical relevance when appropriate variant filtering is employed. As a consequence of this sensitivity, it is reasonable to contend that whole-exome sequencing can be applied to an expanded repertoire of biological problems such as the identification of somatic disease-causing mutations in focal cortical dysplasias.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was funded by the NIHR Biomedical Research Centre Oxford, with funding from the Department of Health's NIHR Biomedical Research Centres funding scheme and FWF Grant P21092. We thank the family members who participated in this study and the High-Throughput Genomics team at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust Grant 090532/Z/09/Z and MRC Hub Grant G0900747 91070). For more information about the Oxford Brain Abnormality Research Group, please see http:\\www.brainabnormalities.org.uk.

Ethics approval: This study was conducted within an approved ethical framework (08/MRE09/55).

1 Lunter, G., Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21,** 936–939 (2011).
2 Vissers, L. E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P. *et al*. A *de novo* paradigm for mental retardation. *Nat. Genet.* **42,** 1109–1112 (2010).
3 Patterson, K. I., Brummer, T., O'Brien, P. M. & Daly, R. J. Dual-specificity phosphatases: critical regulators with diverse cellular targets. *Biochem. J.* **418,** 475–489 (2009).
4 Saillour, Y., Carion, N., Quelin, C., Leger, P. L., Boddaert, N., Elie, C. *et al*. LIS1-related isolated lissencephaly: spectrum of mutations and relationships with malformation severity. *Arch. Neurol.* **66,** 1007–1015 (2009).
5 Tanaka, T., Serneo, F. F., Higgins, C., Gambello, M. J., Wynshaw-Boris, A. & Gleeson, J. G. Lis1 and doublecortin function with dynein to mediate coupling of the nucleus to the centrosome in neuronal migration. *J. Cell Biol.* **165,** 709–721 (2004).
6 Shu, T., Ayala, R., Nguyen, M. D., Xie, Z., Gleeson, J. G. & Tsai, L. H. Ndel1 operates in a common pathway with LIS1 and cytoplasmic dynein to regulate cortical neuronal positioning. *Neuron* **44,** 263–277 (2004).
7 Sicca, F., Kelemen, A., Genton, P., Das, S., Mei, D., Moro, F. *et al*. Mosaic mutations of the LIS1 gene cause subcortical band heterotopia. *Neurology* **61,** 1042–1046 (2003).
8 Mineyko, A., Doja, A., Hurteau, J., Dobyns, W. B., Das, S. & Boycott, K. M. A novel missense mutation in LIS1 in a child with subcortical band heterotopia and pachygyria inherited from his mildly affected mother with somatic mosaicism. *J. Child Neurol.* **25,** 738–741 (2010).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)