This is a repository copy of *RNAseq Supports the Molecular Genetic Diagnosis of Late-Onset ADA Deficiency*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/146756/

Version: Accepted Version

Article:

Watson, CM orcid.org/0000-0003-2371-1844, Stockdale, C, Berry, I et al. (5 more authors) (2019) RNAseq Supports the Molecular Genetic Diagnosis of Late-Onset ADA Deficiency. Journal of Clinical Immunology, 39 (3). pp. 270-273. ISSN 0271-9142

https://doi.org/10.1007/s10875-019-00625-4

© Springer Science+Business Media, LLC, part of Springer Nature 2019. This is an author produced version of a letter published in Journal of Clinical Immunology. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

RNAseq supports the molecular genetic diagnosis of late onset ADA deficiency

Christopher M. Watson, PhD[a,b], Claire Stockdale, MD[c], Ian Berry[a], Laura A. Crinnion[a,b], Ian M. Carr, PhD[b], Andrew Cant, MD[d], David T. Bonthron, FRCPath[a,b], Sinisa Savic MD, PhD[c,e,*]

[a] Yorkshire Regional Genetics Service, The Leeds Teaching Hospitals NHS Trust, St. James's University Hospital, Beckett Street, Leeds, LS9 7TF, United Kingdom

[b] Leeds Institute of Medical Research, University of Leeds, St. James's University Hospital, Beckett Street, Leeds, LS9 7TF, United Kingdom

[c] Department of Clinical Immunology and Allergy, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, United Kingdom

[d] Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne; and the Institute of Cellular Medicine, Newcastle University Medical School, Newcastle upon Tyne NE2 4HH, United Kingdom.

[e] National Institute for Health Research, Leeds Biomedical Research Centre and Leeds Institute of Rheumatic and Musculoskeletal Medicine (LIRMM), Wellcome Trust Brenner Building, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, United Kingdom

*Corresponding author:
Sinisa Savic
E-mail: s.savic@leeds.ac.uk
Tel: +44 (0)113 2065567
Fax: +44 (0)1132067250

To the Editor:

Adenosine deaminase (ADA) deficiency is an autosomal recessive disorder of purine metabolism that typically presents as a severe combined immunodeficiency in infancy (OMIM: 102700). Approximately 10-15% of patients have a delayed clinical onset (6-24 months) and a smaller proportion present later still (4 years to adulthood) with a milder phenotype and gradual immunological deterioration.[1]

Despite the expanding application of modern genetic techniques, the diagnosis of primary immunodeficiencies (PID) can be difficult due to the ever-increasing phenotypic heterogeneity of these disorders.[2] Here we report a case of late-onset ADA deficiency caused by two novel mutations. Both the clinical and genetic diagnoses were challenging and only achieved following the application of several molecular approaches. This case also illustrates the increasing importance of collaborative working between clinicians and laboratory diagnostic scientists.

A 7 year old girl was referred to the paediatric immunology clinic with a 12-month history of recurrent chest infections. She was previously reviewed by dermatologists for extensive molluscum contagiosum, which had been present for several years. Throughout early childhood the patient was generally well and received all primary immunisations without complications. Other past medical history included a recent diagnosis of hypothyroidism requiring thyroxine replacement. The patient was of non-consanguineous white British ancestry, with no family history suggestive of a PID.

Initial investigations showed normal levels of IgG, IgA and marginally elevated IgM. There was a lymphopenia affecting all subsets equally. T cell proliferation to phytohemagglutinin (PHA) was reduced but appeared normal to anti-CD3 (Supplementary Table E1). Specific antibody titers to tetanus, *Haemophilus influenzae* type b and pneumococcus, were within the protective range, suggesting previous adequate responses to vaccination. Interestingly, there were past records of the patient being lymphopenic [at least two CBC in the last 18-month showing lymphopenia of 0.2 and 0.3 10*9/l (reference range 1.5-4.5)].

A diagnosis of combined immunodeficiency was suspected and molecular genetic investigations were performed on a panel of 37 PID-associated genes following clinical exome sequencing (see Supplementary Methods). Interrogation of these data identified the heterozygous missense variant, c.961G>A p.(Glu321Lys), in *ADA* exon 10 (NM_000022.3) (Supplementary Fig E1), which was absent from population control cohorts (including gnomAD[3] and in-house databases) and highly conserved across vertebrate species. This prompted biochemical analysis of ADA activity, which showed the presence of dATP (245 μmol/l), dADP (19 μmol/l) and dAMP (3 μmol/l) and an almost complete loss of ADA activity, pathognomonic of a diagnosis of ADA deficiency. Diagnostic interpretation, in accordance with ACMG criteria[4], supported classification of the c.961G>A variant as being "likely pathogenic". While this variant was determined to have been paternally inherited, complete molecular genetic characterization of the disorder could not be confirmed in the absence of a second pathogenic mutation. This highlights an increasingly common clinical scenario whereby a second deleterious allele, if present, cannot be identified using the "standard" genetic test.

Reported ADA expression profiles (https://gtexportal.org/) motivated our decision to perform globin-depleted whole transcriptome RNAseq on peripheral blood samples obtained from both the proband and her mother. We identified two coding region variants in each individual, for which genotypes were discordantly heterozygous in DNA and homozygous in RNA, suggesting monoallelic *ADA* expression (Fig 1). In the proband, it was the putative mutant allele (c.961G>A) that was expressed. These data justified our decision to perform whole genome sequencing (WGS) to identify the second pathogenic allele in this family. One read-pair, from the maternal WGS dataset, indicated the presence of a ~3.2 kb deletion (Supplementary Fig E2). Manual inspection of these data revealed reads that had been truncated by the alignment software at points within two SINE family repeats. To validate this variant, a breakpoint-spanning PCR amplicon was Sanger sequenced, confirming a clean deletion of 3,256 bp (with no inserted nucleotides) that included *ADA* exon 1 (NM_000022.3). Analysis of breakpoint-flanking sequence revealed a 26-bp stretch of 100% sequence identity, the likely cause of the non-allelic homologous recombination occurring at this locus (Fig 2A). A multiplex PCR was optimised to enable amplification of a normal-allele-specific fragment in conjunction with the deletion-specific allele. This confirmed that the deletion-containing allele was inherited by the proband from her mother (Fig 2B). The complete molecular characterisation of both pathogenic alleles provides the possibility to determine carrier status amongst the extended family and prenatal diagnosis in future pregnancies. The late-onset presentation we describe is likely caused by the 3.2-kb deletion (a presumed null mutation) being inherited in *trans* with the c.961G>A missense variant. Several investigators have previously characterized *ADA* missense variants using *E. coli* strain SØ3834, whose bacterial ADA gene is deleted.[7] These studies established a 4-group classification system defined by residual ADA activity of the mutant allele. Late-onset presentations were correlated with variants that retained a higher activity.

Retrospective analysis of the proband's WGS data did not reveal any read pairs spanning the deletion breakpoint. While comparative read-depth analysis demonstrated a reduction in read coverage across the deleted 3.2-kb interval (from a mean coverage of 13× to 5×), these data emphasise a known limitation in the use of short-read next-generation sequencing for the characterisation of structural variants at nucleotide resolution.[3] We further note that the small size of the *ADA* exon 1 coding region targeted by the enrichment reagent (33 bp) precluded reliable identification of the deletion by comparative read-depth analysis of the clinical exome dataset.

For genetic diagnostic laboratories, one effect of the ubiquitous adoption of next-generation sequencing has been a reduction in the variety of laboratory bench workflows. While this is leading to a standardised approach for DNA-based mutation detection, that predominantly relies on hybridisation-based target enrichment, the rate of detection of variants of uncertain clinical significance and the number of "partial diagnoses" (*i.e.* the identification of a single pathogenic mutation for an autosomal recessive disorder) continues to rise. The United Kingdom's ambition to sequence 5 million patient genomes over the next 5 years suggests that the burden of variant interpretation will worsen.[8]

Other investigators have used targeted amplification of gene-specific transcripts for RNA sequencing, in order to confirm molecular diagnoses in patients with either cystic fibrosis or primary ciliary dyskinesia.[9] In contrast, we have sought to establish a common laboratory workflow that can be applied concurrently to any peripherally-expressed transcripts. Furthermore, whole transcriptome sequencing is hypothesis-independent, removing assay bias associated with the positioning of primer pairs, therefore enabling a greater range of pathogenic mechanisms to be detected. Although the RNAseq data we report identified monoallelic expression, additional splicing aberrations including, for example, exon skipping, exon extension, exonic splice gain or intronic splice gain, could also have been detected.[10] Furthermore, our inference of monoallelic expression, which was informed by the identification of c.961G>A-containing transcripts, provided a locus-specific control for the sensitivity of this workflow for the detection of *ADA* expression. Had the patient been compound heterozygous for two structural variants, or variants that result in nonsense-mediated decay, the RNAseq assay may have been confounded by a lack of detectable transcription. While the current deficit of large RNAseq “control” datasets (which are likely to be influenced by patient demographics and need to be generated on a per-laboratory basis) prohibit robust assessments of quantitative read thresholds, it is possible that increased adoption and standardisation of RNA-based workflows will improve the tractability of more challenging variant combinations.

The patient has responded to pegylated ADA, resulting in normalisation of her total lymphocyte count, and is at home and well after haemopoietic stem cell transplantation. Our study demonstrates the importance of collaborative working between clinicians and diagnostic scientists and emphasises the increasing need to setup multidisciplinary teams to facilitate the reporting and interpretation of genetic investigations. Finally, we highlight to both referring clinicians and diagnostic scientists alike the utility of RNA as a complementary specimen type to aid molecular diagnosis. To enable its routine adoption and to upskill the molecular diagnostic workforce, we support the view that cases with an unresolved second pathogenic mutation should be prioritised for RNA-based analysis.

Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.

Funding sources

This work is supported by the National Institute for Health Research (NIHR) Leeds Biomedical Research Centre and a Medical Research Council grant awarded to Professor Bonthron (MR/M009084/1). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

REFERENCES

1. Flinn AM and Gennery AR. Adenosine deaminase deficiency: a review. Orphanet J Rare Dis 2018;13:65.

2. Rae W, Ward D, Mattocks C, Pengelly RJ, Eren E, Patel SV, et al. Clinical efficacy of a next-generation sequencing gene panel for primary immunodeficiency diagnostics. Clin Genet 2018;93:647-55.

3. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285-91.

4. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405-24.

5. Ameur A, Kloosterman WP, Hestand MS. Single-Molecule Sequencing: Towards Clinical Applications. Trends Biotechnol. 2018;S0167-7799:30204.

6. Teo SM, Pawitan Y, Ku CS, Chia KS, Salim A. Statistical challenges associated with detecting copy number variations with next-generation sequencing. Bioinformatics. 2012;28:2711-8.

7. Hershfield MS. Genotype is an important determinant of phenotype in adenosine deaminase deficiency. Curr Opin Immunol. 2013;15:571-7.

8. Department of Health and Social Care. Matt Hancock announces ambition to map 5 million genomes. https://www.gov.uk/government/news/matt-hancock-announces-ambition-to-map-5-million-genomes. Last accessed: 29th October 2018.

9. Ellingford JM, Beaman G, Webb K, O'Callaghan C, Hirst RA, on behalf of the 100,000 Genomes Project, Black GCM, Newman WG. Whole genome sequencing enables definitive diagnosis of Cystic Fibrosis and Primary Ciliary Dyskinesia. bioRxiv 2018;438838.

10. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci Transl Med 2017;9:386.

Figure legends

Fig 1: Globin-depleted whole transcriptome RNAseq data showing discordant DNA and RNA genotypes for both the proband **(A)** and her mother **(B)**. Nucleotide counts are displayed for corresponding genomic positions. Grey tracks represent per-base cumulative read coverage with *y*-axis ranges as indicated in square brackets. Mapped reads are coloured pink and blue for (+) and (−) strand alignments respectively. Turquoise lines in the RNAseq tracks indicate reads spanning exon-exon junctions. *ADA* is encoded on the antisense strand (blue track), with exons numbered according to transcript NM_000022.3. Genomic coordinates are reported using human genome build hg19. CES: Clinical exome sequencing. WGS: Whole genome sequencing.

Fig 2: (A) Sanger sequencing chromatogram confirming the deletion breakpoint (dashed vertical red line). Genomic coordinates are displayed for the 26-bp region of homology at the proximal (black) and distal (blue) ends of the chromosome 20 *ADA* deletion. **(B)** A diagnostic multiplex PCR assay showing amplification products for normal (611-bp) and deletion-containing (708-bp) alleles. The proband and her sibling are heterozygous for the deletion-containing allele, which was inherited from their mother. Genomic coordinates are reported using human genome build hg19.