

BRIEF REPORT

Whole-transcriptome sequencing in blood provides a diagnosis of spinal muscular atrophy with progressive myoclonic epilepsy

Kristin D. Kernohan^{1,2*} | Laure Frésard^{3*} | Zachary Zappala^{3,4*} | Taila Hartley² | Kevin S. Smith³ | Justin Wagner² | Hongbin Xu⁵ | Arran McBride² | Pierre R. Bourque⁶ | Care4Rare Canada Consortium² | Steffany A. L. Bennett⁵ | David A. Dyment^{1,2} | Kym M. Boycott^{1,2} | Stephen B. Montgomery^{3,4†} | Jodi Warman Chardon^{1,2,6,7,8†}

¹Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada

²Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada

³Department of Pathology, Stanford University, Stanford, California

⁴Department of Genetics, Stanford University, Stanford, California

⁵Department of BMI, University of Ottawa, Ottawa, Ontario, Canada

⁶Division of Neurology, The Ottawa Hospital, Ottawa, Ontario, Canada

⁷Ottawa Hospital Research Institute, Ottawa, Ontario, Canada

⁸Centre for Neuromuscular Disease (CNMD), University of Ottawa, Ottawa, Ontario, Canada

Correspondence

Jodi Warman-Chardon and Kym M. Boycott, Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, ON K1H 8L1, Canada.

Email: jwarman@cheo.on.ca;

kboycott@cheo.on.ca

*Kristin D. Kernohan, Laure Frésard, and Zachary Zappala contributed equally to this study.

†Stephen B. Montgomery and Jodi Warman-Chardon contributed equally to this study as senior authors.

Contract grant sponsors: Genome Canada; Canadian Institutes of Health Research; Ontario Genomics Institute; Ontario Research Fund; Genome Quebec; Children's Hospital of Eastern Ontario Foundation (OGI-064); NIH (R01HG008150, R01MH101814, U01HG007436, T32HG000044, and U01HG009080); National Science Foundation GRFP (DGE-114747); Stanford Center for Computational, Evolutionary, and Human Genomics (CEHG).

Communicated by Garry R. Cutting

Abstract

At least 15% of the disease-causing mutations affect mRNA splicing. Many splicing mutations are missed in a clinical setting due to limitations of in silico prediction algorithms or their location in noncoding regions. Whole-transcriptome sequencing is a promising new tool to identify these mutations; however, it will be a challenge to obtain disease-relevant tissue for RNA. Here, we describe an individual with a sporadic atypical spinal muscular atrophy, in whom clinical DNA sequencing reported one pathogenic *ASAH1* mutation (c.458A>G;p.Tyr153Cys). Transcriptome sequencing on patient leukocytes identified a highly significant and atypical *ASAH1* isoform not explained by c.458A>G ($p < 10^{-16}$). Subsequent Sanger-sequencing identified the splice mutation responsible for the isoform (c.504A>C;p.Lys168Asn) and provided a molecular diagnosis of autosomal-recessive spinal muscular atrophy with progressive myoclonic epilepsy. Our findings demonstrate the utility of RNA sequencing from blood to identify splice-impacting disease mutations for nonhematological conditions, providing a diagnosis for these otherwise unsolved patients.

KEYWORDS

ASAH1, next-generation sequencing, spinal muscular atrophy with progressive myoclonic epilepsy (SMA-PME), transcriptome sequencing

Mutations that affect mRNA splicing are estimated to account for between 15% and 60% of genetic diseases (Krawczak, Reiss, & Cooper, 1992; Lopez-Bigas, Audit, Ouzounis, Parra, & Guigo, 2005). These variants can trigger a variety of effects, including exon skipping, intron retention, activation of cryptic splice sites, or creation of new splice

sites. Resultant defective transcripts may produce a partially or even nonfunctional protein, leading to disease.

It is now well established that next-generation sequencing (NGS) of DNA, by either gene panel or whole-exome sequencing, is an effective method to identify disease-causing mutations (Sawyer et al., 2016;

Sun et al., 2015). However, these assays are usually limited to the coding sequence and minimal flanking DNA (± 10 bp), and rely on computational algorithms to predict the effects of mutations at the mRNA and protein level. Many splice-impact prediction programs have been developed (e.g., MaxEntScan [Yeo & Burge, 2004], NNSplice [Reese, Eeckman, Kulp, & Haussler, 1997], Human Splicing Finder [Desmet et al., 2009], GeneSplicer [Pertea, Lin, & Salzberg, 2001], GENSCAN [Burge & Karlin, 1997], VEP [McLaren et al., 2016], SPANR [Xiong et al., 2015], and SplicePredictor [Brendel, Xing, & Zhu, 2004]), and while each program has strengths and limitations, there is no standard approach to assess potential splicing impact. Furthermore, while evaluating the reference genome, these programs incorrectly predict between 2.6% and 76.1% of splicing events (Jian, Boerwinkle, & Liu, 2014). Therefore, it is highly likely that many splice-impacting mutations are not appreciated due to their location in noncoding regions or limitations of current splice-impact prediction.

A patient with a sporadic and atypical form of spinal muscular atrophy presented to the regional Neurogenetics Clinic for evaluation. His history included weakness with onset at the age of 3 years, followed by a slow progressive deterioration in muscle strength in childhood and adolescence. During his late teen years, he had pronounced progressive weakness in the proximal muscles with marked atrophy of the periscapular, pectoralis, and triceps muscles. He also developed a prominent facial tremor, myoclonic jerks, and sensorineural hearing loss. There was no sensory or autonomic dysfunction, and brain MRI was normal. EEG demonstrated frequent bursts of paroxysmal polyspikes and slow-wave activity. Nerve conduction studies were consistent with a diffuse motor neuropathy/neuronopathy. Left deltoid muscle biopsy primarily demonstrated neurogenic changes. Cognition was normal and family history unremarkable. DNA was sent for *SMN1* deletion testing, chromosomal microarray, and SMA gene panel testing (SMA panel at a CLIA-certified USA-based clinical laboratory that assesses 15 motor neuron-related disease genes). A single variant in *ASAH1* (NM_004315.4: c.458A>G; p.Tyr153Cys) was reported and classified as a variant of unknown significance, but was highly suspicious based on its rarity in control cohorts (seen in one of 70,372 alleles in ExAC Browser), and it is predicted to be deleterious to protein function (PolyPhen-2 [1.0], SIFT [0.0]). Subsequent *ASAH1* deletion and duplication testing to assess for a second mutation was negative.

Given the highly suspicious *ASAH1* variant and similarity between our patient phenotype and spinal muscular atrophy with progressive myoclonic epilepsy (SMA-PME) (OMIM 159950), a recessive condition caused by mutations in *ASAH1* (Zhou et al., 2012), we hypothesized the existence of a noncoding variant that was missed by the clinical panel sequencing. We chose to perform whole-transcriptome sequencing on a research basis. The patient was enrolled in the Care4Rare Canada research project, which was approved by the institutional research ethics board (Children's Hospital of Eastern Ontario); free and informed consent was obtained from the patient.

An important consideration in our approach to transcriptome sequencing was to utilize the most readily available and least invasive cell type for evaluation. Therefore, we chose to conduct RNA sequencing from blood. Importantly, we have a control dataset of 909 control samples with which to compare patient data (Battle et al.,

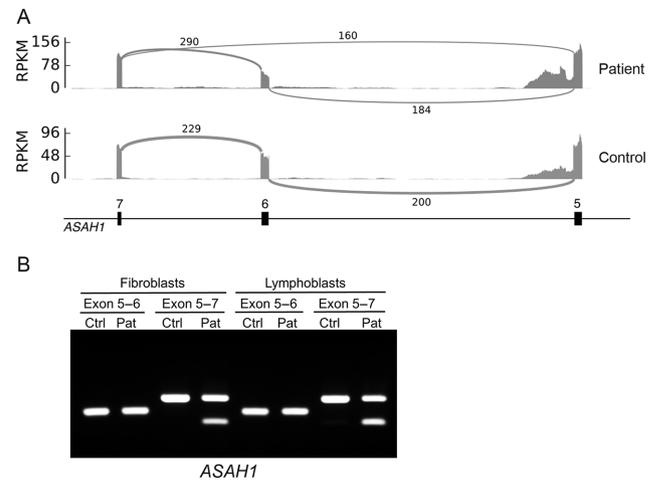


FIGURE 1 *ASAH1* differential splicing. **A:** Sashimi plot of *ASAH1* exons 5–7 in control and patient transcriptomes demonstrates the novel splice isoform observed in our patient lacking exon 6. Data are displayed as RPKM ($1e3 * [\text{depth}/\text{read_length}]/[\text{n_mapped_reads}/1e6]$), and the number of reads for each junction is noted on each junction arc. **B:** The novel isoform lacking exon 6 was confirmed by RT-PCR on RNA from patient-derived fibroblasts and lymphoblasts

2014). Leukocytes were selected, transcriptome libraries were prepared (including Globin depletion and polyA capture), and the RNA was sequenced to a read depth of 63 million reads on an Illumina NextSeq500, using a combination of 150 and 75 bp paired-end technology. RNA reads were aligned to the genome, PCR duplicates were removed, and splice junctions were assessed (see Supp. Information for detailed RNA sequencing and analysis methods). We then tested for differential exon usage between the patient and all other individuals in order to identify aberrant splicing events unique to the patient (allowing a junction to be considered if it is observed in at least one sample with a minimum of 20 supporting reads). The most significant splicing difference between the patient and the 909 unaffected individuals was a skipping of the sixth exon in the *ASAH1* gene (Fig. 1A). This event was remarkable because it was not observed in any of the control samples, and is orders of magnitude more significant than any other differential splicing event in the patient (P value $< 1e-16$ after Benjamini–Hochberg correction). Haplotype analysis revealed that while there was a proportion of the normal allele from both haplotypes, 90% of the reads from one haplotype were of the alternative transcript. RT-PCR analysis on RNA extracted from fibroblast and lymphoblast cell lines derived from the patient was conducted to confirm this effect (Fig. 1B).

The initial *ASAH1* c.458A>G variant identified by clinical testing was not located near an exon 6 boundary and was therefore unlikely to have caused this splicing effect. We conducted Sanger sequencing in the Care4Rare research laboratory to assess for a second *ASAH1* variant. A c.504A>C, p.Lys168Asn (NM_004315.4) variant located at the 3' end of exon 6 (–2 bp from the splice junction) was identified. While informatics scores do not predict this variant to be damaging for protein function (PolyPhen-2 [0], SIFT [0.53]), this variant has been previously reported as a mutation in SMA-PME that effects splicing (Dyment et al., 2014). It is unclear why the second mutation was missed by clinical DNA testing and an explanation was not provided by the

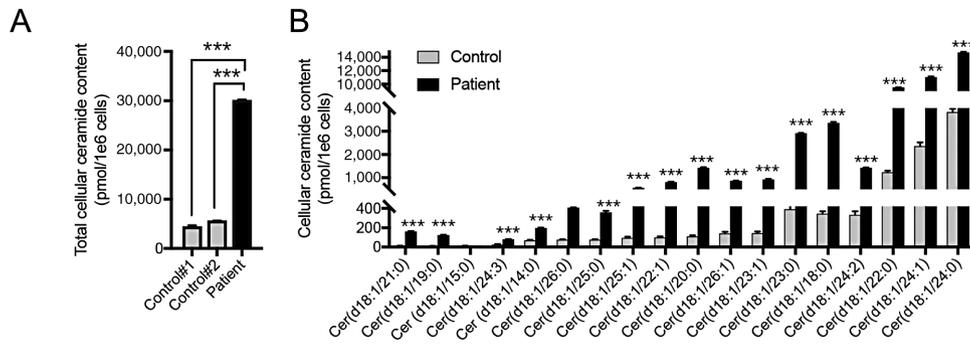


FIGURE 2 Defective ASAH1 function confirms the SMA-PME diagnosis. Total ceramide levels (A) and detailed ceramide profile (B) demonstrate a loss of ASAH1 function resulting in increased ceramides in patient-derived fibroblast cells compared with controls

clinical laboratory. We conclude that the differential splicing effect identified by transcriptome sequencing is likely a consequence of the c.504A>C *ASAH1* variant.

Finally, given the variant of unknown significance classification of the c.458A>G change and the patient's atypical SMA-PME clinical presentation, we wanted to assess ASAH1 function to conclusively provide a SMA-PME diagnosis for this patient. *ASAH1* encodes a member of the acid ceramidase family, which catalyzes the degradation of ceramide into sphingosine and free fatty acids. Cellular ceramide content was quantified by high-performance liquid chromatography electrospray ionization tandem mass spectrometry in extracts from patient and control fibroblast cells. Eighteen distinct ceramide species with a d18:1 sphingosine backbone were detected; a significant increase was observed in 17 of the 18 ceramides (Fig. 2A and B), confirming the loss of ASAH1 function, and thus the pathogenicity of the *ASAH1* variants. We conclude that our patient has SMA-PME caused by biallelic *ASAH1* mutations, c.504A>C (p.Lys168Asn) and c.458A>G (p.Tyr153Cys).

Genes highly implicated in disease have an increased number of splice junctions and transcripts per gene, when compared with nondisease genes (Jian et al., 2014), speaking to the importance of splicing for human health. At present, approximately 10% of disease mutations cataloged in the Human Gene Mutation Database (HGMD) are annotated as splice impacting (Stenson et al., 2003). However, splice-impacting mutations may be missed by DNA NGS due to (1) failed detection of the DNA variant (due to location in a region not sequenced or in a GC-rich, AT-rich, or repetitive region difficult to capture and analyze), (2) failed reporting (due to sequence quality, population variant frequency, or human error), or (3) failed annotation of splice impact (due to limitations in prediction algorithms). Given these limitations, it is likely that the true proportion of splice-impacting mutations in rare disease is much higher than 10%. There is a clear need for improved methods to identify splice-impacting mutations.

Whole-transcriptome sequencing is a promising new tool to assess mRNA splicing. This may be particularly effective in recessive conditions when one pathogenic or likely pathogenic mutation is identified by DNA NGS and the second remains elusive. Additionally, the use of transcriptome sequencing could play a key role in interpreting variants of unknown significance by establishing any splicing effects in a gene of interest (splicing effects is an important ACMG pathogenicity classification criteria [Richards et al., 2015]). Whole-transcriptome sequenc-

ing is an ideal assay as this can be used for all disorders, as opposed to targeted approaches that must be individually designed and validated. The use of transcriptome sequencing in blood in the patient reported here with a neurological disease demonstrates the power of transcriptome sequencing to aid in the identification of splice-impacting DNA mutations. Interestingly, of the genes on the motor neuron panel, 71% are expressed in at least half of our control samples, and of the 78 genes in OMIM linked to myoclonic seizures, 40% are expressed in at least half of the control samples (a gene was considered expressed if at least 50% of the control cohort samples showed FPKM ≥ 1). This indicates that blood transcriptome sequencing would be a viable option for many of these disorders.

A recent study used transcriptome sequencing of muscle-derived RNA to identify a splice-impacting *DMD* mutation in a muscular dystrophy patient (Gonorazky et al., 2016). To our knowledge, our study is the first to identify a hereditary Mendelian disease mutation via transcriptome sequencing in blood. Importantly, the approach was successful in identifying a splice-impacting mutation for SMA-PME, a non-hematological condition. Relevant disease tissues are often difficult to obtain, whereas blood is a more readily available and less invasive sample. It is noted that some causative mutations may exist in transcripts that are not expressed in blood (e.g., some muscle or neuron-specific); conversely, many tissue-specific transcripts are expressed at low levels elsewhere, including blood. Given the limitations of sample acquisition from living patients, transcriptome sequencing from blood likely offers a feasible approach to identify splice-impacting DNA mutations for many hereditary Mendelian diseases.

In conclusion, we used whole-transcriptome sequencing from blood to identify a splice-impacting disease mutation and provide a diagnosis of SMA-PME in a patient where only a single variant of uncertain significance had been identified by the clinical laboratory. While DNA-based NGS applications are employed in many clinical diagnostic laboratories, RNA-based NGS applications have yet to be leveraged in the same fashion. Our results highlight the potential of combined multi-omic approaches, including transcriptome sequencing, to aid the diagnosis of rare disease patients.

ACKNOWLEDGMENTS

The authors would first and foremost like to thank the patient and their family for their active participation in this study. This work was

performed under the Care4Rare Canada Consortium (Enhanced Care for Rare Genetic Diseases in Canada).

REFERENCES

- Battle, A., Mostafavi, S., Zhu, X., Potash, J. B., Weissman, M. M., McCormick, C., ... Koller, D. (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Research*, *24*, 14–24.
- Brendel, V., Xing, L., & Zhu, W. (2004). Gene structure prediction from consensus spliced alignment of multiple ESTs matching the same genomic locus. *Bioinformatics*, *20*, 1157–1169.
- Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology*, *268*, 78–94.
- Desmet, F. O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., & Beroud, C. (2009). Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*, *37*, e67.
- Dyment, D. A., Sell, E., Vanstone, M. R., Smith, A. C., Garandeanu, D., Garcia, V., ... Boycott, K. M. (2014). Evidence for clinical, genetic and biochemical variability in spinal muscular atrophy with progressive myoclonic epilepsy. *Clinical Genetics*, *86*, 558–563.
- Gonorazky, H., Liang, M., Cummings, B., Lek, M., Micallef, J., Hawkins, C., ... Dowling, J. J. (2016). RNAseq analysis for the diagnosis of muscular dystrophy. *Annals of Clinical and Translational Neurology*, *3*, 55–60.
- Jian, X., Boerwinkle, E., & Liu, X. (2014). In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Research*, *42*, 13534–13544.
- Krawczak, M., Reiss, J., & Cooper, D. N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences. *Human Genetics*, *90*, 41–54.
- Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G., & Guigo, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Letters*, *579*, 1900–1903.
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., ... Cunningham, F. (2016). The ensembl variant effect predictor. *Genome Biology*, *17*, 122.
- Pertea, M., Lin, X., & Salzberg, S. L. (2001). GeneSplicer: A new computational method for splice site prediction. *Nucleic Acids Research*, *29*, 1185–1190.
- Reese, M. G., Eeckman, F. H., Kulp, D., & Haussler, D. (1997). Improved splice site detection in Genie. *Journal of Computational Biology*, *4*, 311–323.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... ACMG Laboratory Quality Assurance Committee. (2015). 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. *Genetics in Medicine*, *17*, 405–424.
- Sawyer, S. L., Hartley, T., Dyment, D. A., Beaulieu, C. L., Schwartzentruber, J., Smith, A., ... Boycott, K. M. (2016). Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: Time to address gaps in care. *Clinical Genetics*, *89*, 275–284.
- Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shiel, J. A., Thomas, N. S., & Cooper, D. N. (2003). Human Gene Mutation Database (HGMD): 2003 update. *Human Mutation*, *21*, 577–581.
- Sun, Y., Ruivenkamp, C. A., Hoffer, M. J., Vrijenhoek, T., Kriek, M., van Asperen, C. J., & Santen, G. W. (2015). Next-generation diagnostics: Gene panel, exome, or whole genome? *Human Mutation*, *36*, 648–655.
- Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K., & Frey, B. J. (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science*, *347*, 1254806.
- Yeo, G., & Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology*, *11*, 377–394.
- Zhou, J., Tawk, M., Tiziano, F. D., Veillet, J., Bayes, M., Nolent, F., ... Melki, J. (2012). Spinal muscular atrophy associated with progressive myoclonic epilepsy is caused by mutations in *ASAH1*. *American Journal of Human Genetics*, *91*, 5–14.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Kernohan KD, Frésard L, Zappala Z, et al. Whole-transcriptome sequencing in blood provides a diagnosis of spinal muscular atrophy with progressive myoclonic epilepsy. *Human Mutation*. 2017;38:611–614. <https://doi.org/10.1002/humu.23211>

SUPPORTING INFORMATION

Transcriptome Analysis

RNA from this patient was processed alongside 52 other samples (including other rare disease patients as well as healthy controls). For each sample, 0.5 μ g of total RNA was processed and analyzed on an Agilent Bioanalyzer 2100 for quality. Globin mRNA was removed from all samples with GLOBINclear prior to cDNA library construction. cDNA libraries were constructed following the Illumina TrueSeq Stranded mRNA Sample Prep Kit protocol and dual indexed. The size and concentration of each cDNA library was determined using the Bioanalyzer for proper dilution and balanced sequencing across samples. Pooled samples were run in three distinct sequencing runs on an Illumina NextSeq 500 (High Output cartridge): the first generated 151 bp paired end reads and the second and third runs generated 76 bp paired end reads. Overall we generated an average of ~55 million reads per sample across all three runs - the patient in question was sequenced during the second run and had ~63 million reads. Before further processing, we also included an additional set of reads from 909 samples from the Depression, Genes and Networks cohort (Battle et al., 2014) to serve as external healthy controls.

We aligned reads for each sample to hg19 with STAR (version 2.4.0j). We filtered these alignments to include only uniquely mapped reads and removed PCR duplicates using Picard tools (<https://github.com/broadinstitute/picard>). Using Leafcutter (Li et al., 2016), we extracted all observed splice junctions from the filtered alignments and clustered them to identify introns (requiring each intron to have at least 30 supporting reads or at least 0.1% of the total reads in the cluster). We then used custom scripts to assess differential exon usage between the patient and all other individuals in order to identify aberrant splicing events unique to the patient (allowing

an intron to be considered if it is observed in at least one sample with at least 20 supporting reads).

Ceramide Assay

Patient cells and two lines of healthy human fibroblasts cultured in DMEM/F12 (ThermoFisher) supplemented with 10% fetal bovine serum (Sigma Aldrich), 1% penicillin/streptomycin, and 2 mM L-glutamine (complete media) were extracted using the modified method of Bligh and Dyer as we have previously described (Bligh and Dyer, 1959; Ryan, et al., 2009; Whitehead, et al., 2007) with 235 pmol of Cer(d18:1/16:0-d31) (Avanti Polar Lipids) added at time of extraction as internal standard. Cells were washed with PBS and counted prior to lipid extraction. Dried lipid extracts were re-dissolved in ethanol, stored in amber glass vials, flushed with nitrogen gas and kept at -80°C until use. Intracellular ceramide content was quantified by high performance liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS) using an Agilent Infinity II LC and a triple quadrupole-linear ion trap mass spectrometer QTRAP 5500 (AB SCIEX). LC separation was achieved on a 100 mm x 250 µm (i.d.) capillary column packed with ReproSil-Pur 120 C8 (particle size of 3 µm and pore size of 120 Å, Dr. A. Maisch, Ammerbruch, Germany) with 8 µl sample injections by an autosampler maintained at 4°C at a flow rate of 10 µl/min. was used. The binary solvent gradient consisted of water with 0.1% formic acid and 10 mM ammonium acetate (solvent A) and ACN/acetonitrile/IPA/isopropanol (5:2; v/v) with 0.1% formic acid and 10 mM ammonium acetate (Solvent B). Gradient elution began started at 30% B, reached 100% B in 5 min, was maintained at 100% B for 30 min, before returning to 30% B within 1 min and then maintained for 10 min to re-equilibrate the column. Data acquisition was performed in the positive ion mode using multiple reaction monitoring

mode (MRM) for transitions corresponding to all possible biological ceramides with a d18:1 sphingosine backbone monitoring the MS/MS ion transition from $[M+H]^+ \rightarrow 264.3$ (corresponding to the didehydrated 18:1 sphingosine backbone). Instrument control and data acquisition were performed with Analyst software (v. 1.6.2, AB SCIEX). MultiQuant software (v. 3.0.2, AB SCIEX) was used for processing of quantitative MRM data. Ceramide abundance was expressed as pmol equivalents of Cer(d18:1/16:0-d31) per 1E6 cells. Total ceramide content was analyzed statistically using a one-way analysis of variance with post-hoc Holm-Sidak tests. Changes in the abundance of individual ceramide species was determined statistically by multiple t-tests with false discovery rate (FDR) set at 0.01%. We used the two-stage linear step-up FDR procedure of Benjamini, Krieger and Yekutieli (Benjamini, et al., 2006) without assuming consistent standard deviation. Prism (GraphPad) was used for all statistical analyses.

References

Battle, A, Mostafavi, S, Zhu, X, Potash, JB, Weissman, MM, McCormick, C, Haudenschild, CD, Beckman, KB, Shi, J, Mei, R, Urban, AE, Montgomery, SB, Levinson, DF, Koller, D (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.*, 24, 1:14-24.

Li, YI, Knowles, DA, Pritchard, JK (2016) LeafCutter: Annotation-free quantification of RNA splicing. *bioRxiv.*, 044107; doi: <http://dx.doi.org/10.1101/044107>

Benjamini, Y, Krieger, AM, Yekutieli, D (2006) Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*. 93(3), 491-507