

11-21-2014

Analysis for co-occurring sequence features identifies link between common synonymous variant and an early-terminated NPC1 isoform

Mercedeh Movassagh
George Washington University

Prakriti Mudvari
George Washington University

Maria Kokkinaki
George Washington University

Nathan J. Edwards
Georgetown University

Nady Golestaneh
Georgetown University

See next page for additional authors

Follow this and additional works at: http://hsrc.himmelfarb.gwu.edu/smhs_pharm_facpubs

 Part of the [Medical Pharmacology Commons](#), and the [Medical Physiology Commons](#)

Recommended Citation

Movassagh, M., Mudvari, P., Kokkinaki, M., Edwards, N.J., Golestaneh, N. et al. (2014). Analysis for co-occurring sequence features identifies link between common synonymous variant and an early-terminated NPC1 isoform. *Journal of Clinical Bioinformatics*, 4:14.

This Journal Article is brought to you for free and open access by the Pharmacology and Physiology at Health Sciences Research Commons. It has been accepted for inclusion in Pharmacology and Physiology Faculty Publications by an authorized administrator of Health Sciences Research Commons. For more information, please contact hsrc@gwu.edu.

Authors

Mercedeh Movassagh, Prakriti Mudvari, Maria Kokkinaki, Nathan J. Edwards, Nady Golestaneh, and Anelia Horvath

SHORT REPORT

Open Access

Analysis for co-occurring sequence features identifies link between common synonymous variant and an early-terminated *NPC1* isoform

Mercedeh Movassagh^{1,2}, Prakriti Mudvari^{1,2}, Maria Kokkinaki³, Nathan J Edwards^{4,5}, Nady Golestaneh^{3,4,5} and Anelia Horvath^{1,2*}

Abstract

Direct assessment of allelic phase for DNA and RNA features of diploid genomes has been challenging for Sanger sequencing, due to its allele-conflating base-calling signal. Massively parallel sequencing technologies are based on the generation of a continuous copy of a single strand sequence segments, thus preserving the allelic relation between the features of the original molecules. We have performed a transcriptome-wide search for co-occurrence of variant nucleotides and exon-intron boundaries positioned within the length of a single sequencing read. Analysis of 75 human transcriptomes from retinal pigment epithelia (RPE), glioblastoma, low-grade brain tumor, breast cancer and colon cancer, have identified an association between the synonymous variant rs1140458 and an early-terminated *NPC1* isoform lacking exons 19–25. Higher proportion of molecules bearing the variant nucleotide (versus the reference) incorporates the intron ($P < 0.0001$), which turns the last codon of exon 18 into a stop codon. The significance is highest in RPE cells ($P = 3.88 \times 10^{-12}$). *NPC1* protein is involved in the control of the cholesterol trafficking. *NPC1* mutations lead, in an autosomal recessive manner, to the neurological disorder Niemann-Pick syndrome type C (NP-C), and, ablation of *NPC1* causes age-progressive retinal degeneration in mice and drosophila. The vast majority of the NP-C causative variants consist of missense/nonsense substitutions, small indels, and, intronic splice variants. Rs1140458 is a common exonic synonymous substitution that has never been linked to alternative splicing or pathogenicity. Our analysis suggests that rs1140458 may affect the levels of the functional *NPC1* protein, and to contribute to some of the cholesterol-implicated cellular phenotype.

Introduction

Autosomal recessive mutations in the *NPC1* gene account for 95% of the cases with Niemann-Pick syndrome type C (NP-C, MIM #607623), a cholesterol storage and glycosphingolipid trafficking deficiency, with wide variability in the age of onset and in the number and severity of manifestations [1,2]. The patients manifest progressive neurological symptoms including cerebellar atrophy, loss of Purkinje neurons, and neurofibrillary tangles, which could eventually cause ataxia, loss of speech capabilities, psychiatric problems and learning difficulties [1,3]. Despite the clear autosomal mode of inheritance, in up to 30% of

the patients, only one deleterious *NPC1* allele has been identified [4-7]. In addition, ablation of *NPC1* expression is shown to lead to age-progressive retinal degeneration in mice and drosophila [8,9].

The *NPC1* gene is located on chromosome 18q11.2 and is composed of twenty-five exons. According to the Human Gene Database (<http://www.hgmd.cf.ac.uk/ac/index.php>, [7]) 345 pathogenic *NPC1* mutations have been described, consisting of missense and nonsense substitutions (69%), small insertions and deletions affecting the open reading frame (20%), intron-positioned splice variants (8%), and gross alterations (3%) (Additional file 1: Table S1). Rs1140458 is a common synonymous substitution (N931N, chr18:21119777 C > T, NM_000271.4) in *NPC1*, with allele frequency close to 0.5. It is located three nucleotides from the exon boundary inside of exon 18 and has not previously been associated with pathogenicity. Herein, we

* Correspondence: horvatha@gwu.edu

¹Department of Pharmacology and Physiology, The George Washington University, 2300 Eye Street NW, Suite 653, 20037 Washington, DC, USA

²McCormick Genomic and Proteomic Center, The George Washington University, 20037 Washington, DC, USA

Full list of author information is available at the end of the article

describe a link between the rs1140458 and a prematurely terminated *NPCI* isoform lacking exons 19–25, discovered through a novel approach for assessment of co-occurrence of RNA features.

Materials and methods

Human RNA-seq transcriptomes

We analyzed 75 transcriptomes derived from different human tissues and cell lines as follows: 10 in-house RPE primary cell lines from the eyes of organ donors, and 65 RNA-seq datasets downloaded from online sources: 7 glioblastoma tumor tissues and 8 low-grade brain tumor cell lines, 25 breast cancer tumor tissues and 6 breast cancer cell lines, 9 colon cancer tissues, and 10 colon cancer cell lines. The web-sources used were ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>) [10], and Cancer Genomic Hub (<https://cghub.ucsc.edu/>) [11]. The choice of samples is based on relevance to the cholesterol/glycosphingolipid metabolism, and availability of concordantly processed RNA-seq datasets from -different tissues for assessment of tissue-specificity of our observation.

RNA-seq analysis pipeline

All the RNA-seq datasets were aligned using Tophat 2 version 2.0.11 [12]. Variants calls were obtained using Mpileup utility of SAMTools (<http://samtools.sourceforge.net/mpileup.shtml>) [13]. Base Alignment Quality was used to score the variant calls. Consensus calling was one using bcftools. The variant calls were quality filtered as described previously [14], and annotated using SeattleSeq Annotation Tools version 8.01, dbSNP build 138 [15]. Parallel run of the 10 RPE samples through Varscan (<http://varscan.sourceforge.net/>) produced 100% identical calls on chr18:21119777 C > T in *NPCI*. Transcript abundance level of each sample was quantified using Cufflinks version 2.1.1 [16] as we have previously described [14].

Assessment for variants associated with intron-containing RNA-molecules

To identify potential splice-modulating variants, we have performed a transcriptome-wide search for nucleotides that selectively reside in unspliced molecules, using aligned RNA-seq datasets. To do this we applied SNplice - a Python-based computational approach developed in our lab (<https://code.google.com/p/snsplice/>, v.1.7.1) [17]. SNplice finds sequencing reads spanning SNP loci and nearby exon-intron boundaries, and classifies them based on the nucleotide at the SNP site and the presence or absence of the canonical exon-exon splice junction [17]. Following analysis of raw reads using Tophat, and Samtools, the resulting aligned reads (.bam format), junctions (.bed format), and variants (.vcf format) are read directly by SNplice. Four categories of reads are counted at a

given SNP–junction pair, and the statistical significance of non-independence of the SNP and splicing is assessed using Fisher-exact test: N_{VARee} indicates reads containing the variant base and exon-exon junction, N_{VARei} - variant base and exon-intron boundary, N_{REFee} - reference base and exon-exon junction, and N_{REFei} - reference base and exon-intron boundary. High N_{VARei} with low N_{REFei} indicate that the variant nucleotide is preferentially observed in intron-retaining molecules, and, may have splice-modulating potential. In addition, log odds-ratio (\log_2 , LOD) of intron-retaining reads among variant bearing reads versus intron-retaining reads among all reads is computed [17]:

$$\log_2 \frac{N_{VARei}/(N_{VARee} + N_{VARei})}{(N_{VARei} + N_{REFei})/(N_{VARee} + N_{VARei} + N_{REFee} + N_{REFei})}$$

The log odds-ratio increases when the intronic reads are enriched in the variant allele. A pooled ratio is used in the denominator since the number of reference intronic reads is often zero. For analyses across samples, SNplice will accept multiple reads, junctions, and variant files, taking the semantic union of the input reads, junctions, and variants.

Allele-specific Sanger sequencing

To illustrate the co-occurrence of variant nucleotide and exon-intron boundary, and the coherence with the alignment display through IGV, we designed and performed allele-specific reverse transcription PCR on cDNA derived from heterozygote samples from the same RPE primary cell lines, followed by Sanger sequencing. Three primers were designed: a common forward exonic primer positioned further inside the exon from rs1140458, and two reverse primers hybridizing in the downstream exon or intron, respectively (Additional file 2: Table S2, and Additional file 3: Figure S1A). The PCR products were gel-purified and subjected to bi-directional Sanger sequencing using the forward and reverse primers used for the amplification.

Results and discussion

The distribution of the categories of reads across the 75 samples is presented in Additional file 4: Table S3. In the nine heterozygote samples of the RPE subset, the variant nucleotide of rs1140458 was found more often on reads remaining unspliced at the exon 18 boundary, with seven of the samples' read counts meeting a p -value threshold of 5% (p -values between 0.04 and 0.009). When reads from all RPE samples, including the homozygote ones, were pooled, the p -value of the observed counts was 3.9×10^{-12} (See Additional file 4: Table S3). In the rs1140458 positive samples, between 8 and 21% of the sequencing reads harboring the variant nucleotide

retained the intron, while this percent was on average 1.7 for the reads carrying the reference nucleotide. Out of all intron-retaining molecules, overall 87.7% carried the variant nucleotide.

Next, we assessed the three other assemblies of non-neural cell lines and tissues: glioma/glioblastoma, breast cancer, and colon cancer (See Additional file 4: Table S3). In contrast to the neuroepithelial cells, when examined individually, fewer heterozygote samples presented with significantly higher number of intron-retaining reads harboring the variant nucleotide. For example, none of the heterozygote samples in the glioma/glioblastoma subset reached significance in the read distribution at $p < 0.05$. Nevertheless, a tendency was clearly seen, and the proportion of the intron-retaining reads with variant nucleotide was significant for the pooled reads in both cell lines ($p < 0.0001$), and tissues ($p = 0.03$). Similarly, few individual samples reached $p < 0.05$ for the read distribution in the breast and colon cancer subsets: one from colon cancer tissue, and two from the colon and breast cancer cell lines each. A total of 4 more heterozygote samples showed a tendency ($p < 0.1$). Again, when the pooled number of reads across all samples was analyzed, the difference in the distribution was significant ($p < 0.0001$) for all 4 datasets. The proportion of variant-bearing reads that retained the intron (out of all variant bearing reads) ranged between 3 and 8.3% as compared to between 0.2 and 1.6% for the intron-retaining reference reads. Of note, the overall proportion of the intron-retaining reads bearing the variant (out of all intron-retaining reads) in the non-neuroepithelial samples was the same as the one observed in the neuroepithelial subset – 87.6%. To assess if the observed strong association in the RPE cells is related to overall increased alternative splicing, we analyzed the variety and the abundance of transcripts across the studied samples. Higher expression fraction of complete length *NPC1* transcripts (over all *NPC1* transcripts) was estimated in the RPE samples, as compared to the cancer samples (0.88 vs 0.55, $p < 0.001$).

All the samples were examined at the rs1140458 locus using Integrative Genomics Viewer (IGV) [18], and the correspondence to the SNPllice assessment was visually confirmed (Figure 1 top). Furthermore, Cufflinks assembly (.gtf) of the local sequencing reads in samples bearing the variant showed the presence of an isoform retaining the intron 18 of the *NPC1* gene (Figure 1 bottom).

We performed allele-specific PCR followed by Sanger sequencing on two heterozygote samples from the RPE primary cell lines; resulting chromatograms are shown on Additional file 3: Figure S1. The chromatograms were specifically examined for relative signal (peak) of the variant versus the reference nucleotide in the spliced and unspliced amplicons. Consistent with the assessment, Sanger sequencing of the region flanking rs1140458

showed that the molecules bearing the variant nucleotide predominate in the PCR product amplifying the exon-intron boundary. In contrast, the amplicon of the canonically spliced exon-exon region showed an equal proportion of the variant and the reference allele (See Additional file 3: Figure S1).

The potential of rs1140458 to modulate the splicing was further assessed through three independent splice modeling tools: SplicePort, Skippy and SpliceAid [19-21]. SplicePort estimated that the variant diminishes the donor strength in comparison to the reference (False Discovery Rate, FDR < 0.01). Skippy predicted loss of one exonic splice enhancer, and SpliceAid predicted the loss of the binding site for YB1 [22]. Overall, all three tools predicted a role for the variant in splice modulation.

Incorporation of the intronic sequence downstream of exon 18 leads to replacement of the third base of codon 932, which alters it from tyrosine-encoding to a stop codon, thus generating a prematurely terminated *NPC1* isoform lacking the seven downstream exons. Such an isoform has been previously described (uc010xba, UCSC genes). To our knowledge, this is the first study to link it to the presence of the variant nucleotide of the upstream located rs1140458.

The highest proportion of intron-retaining reads in an individual sample is 21% (RPE Sample #5, see Additional file 4: Table S3). Based on that, most of the samples homozygote for rs1140458 are expected to retain more than half of the active *NPC1*. If, however, rs1140458 is in compound heterozygosity with a strong deleterious *NPC1* allele, the level of the functional protein may decrease below 40%. Currently, it is unclear whether this level of NPC1 activity is sufficient for proper cellular functioning, or if it would lead to a cellular phenotype resulting from aberrant cholesterol trafficking [23]. Nevertheless, the study illustrates an important example of partial mutational effect – a phenomenon of particular interest for studying of incomplete penetrance of genetic changes.

Rs1140458 is positioned three nucleotides inside exon 18, thus residing on the border of a sequence traditionally annotated as “splice-site”, which consists of the two immediate nucleotides on each side of the exon-intron boundary. However, only a small percentage of exonic splice-site nucleotides have been experimentally verified to affect splicing. Our results suggest that the motif encompassing rs1140458 may be important for recognizing the donor splice sequence by the components of the splicing machinery.

The interesting observation in this study is the strong effect of the rs1140458 variant observed specifically in the retinal cells. This effect was present, but weaker in the compared cancer datasets, known to have increased general splice deregulation, as also shown by their increased

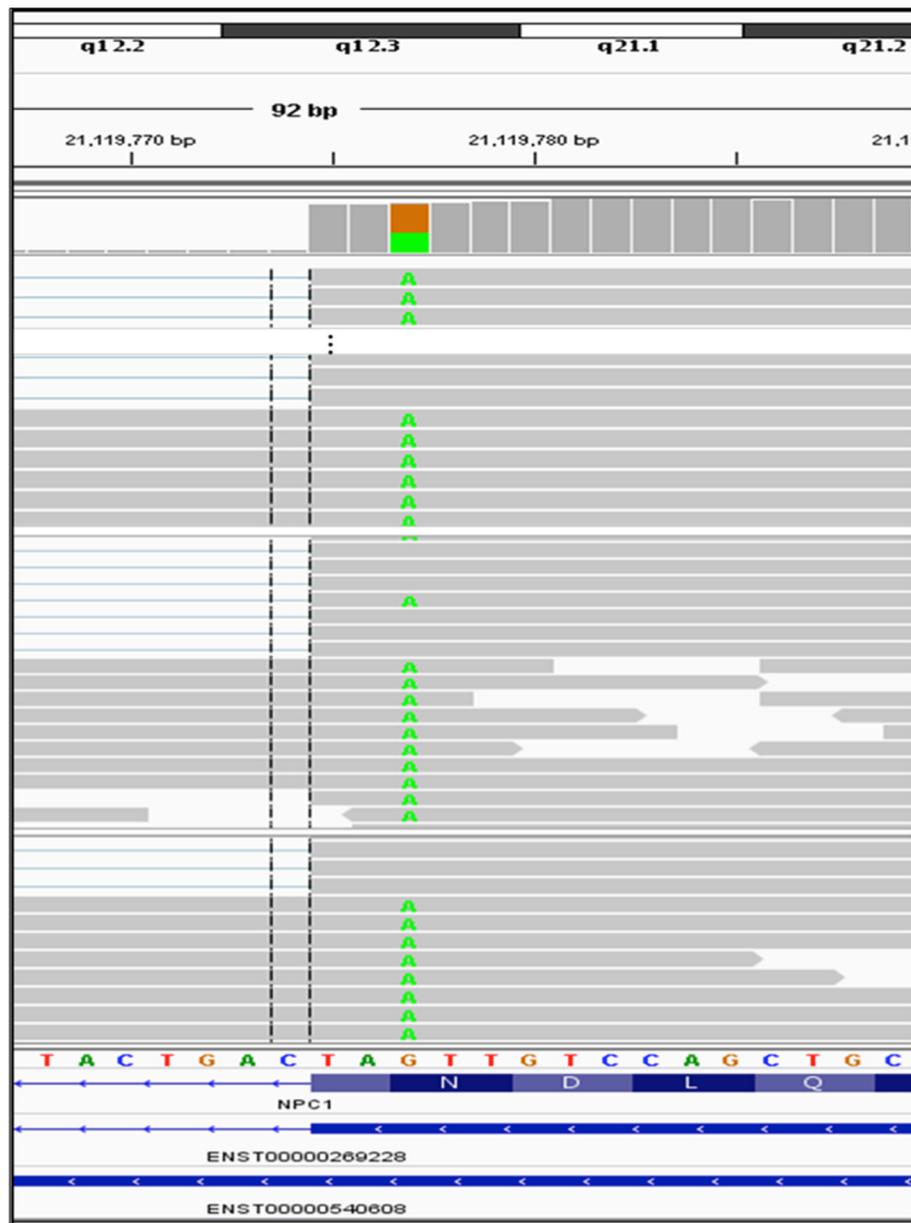


Figure 1 IGV visualization of the read alignment and assembly of rs1140458 region for three illustrative RPE samples. Top. Sequencing reads with variant (A, in green) and reference nucleotide at the SNP position in the proximity of an exon-intron boundary are shown. Variant-harboring reads often continue in the intron, indicating association with potential junction alteration. Bottom. Cufflinks assembly (gtf) showing the presence of an isoform retaining the intron 18 of the *NPC1* gene (ENST00000540608 represents spliced isoform, and ENST0000269228 represents intron-retaining isoform).

fraction of incomplete *NPC1* transcripts [24]. This may indicate cell-specific splicing control in the neuroepithelial cells, making this variant selectively essential for tissue-specific phenotype formation. Given the important role of the cholesterol for the retinal physiology [8,9,25], rs1140458 variant in *NPC1* is worth studying for potential implication in cholesterol-related retinal pathology.

Despite the clear tendency for the variant to reside in intron-retaining reads in non-neuronal tissue types, the

significance of the read distribution there was lower, mainly due to the overall lower proportion of intron-retaining reads (See Additional file 4: Table S3). This observation shows that even in cells where intron retention occurs less frequently, it is still selective for the variant-bearing molecules, which is also demonstrated by the similar proportion of variant intron-retaining reads (out of all intron-retaining reads) across the entire set of samples.

In summary, we demonstrate that substantial proportion of the RNA molecules bearing the variant nucleotide of rs1140458 in the *NPC1* gene remain unspliced at the nearby exon-intron boundary and incorporate an early stop codon; the effect being stronger in RPE cells. The implication of rs1140458 in NP-C phenotype variation, or in the cholesterol metabolism in retinal cells, is a subject for further investigation. However, our study illustrates an important mechanism of allele-specific splicing modulation, and reveals partial penetrance of the effect of a genetic variant in a cell-specific mode – a phenomenon of crucial importance for understanding complex genotype-phenotype relationships and multi-gene variance effects.

Additional files

Additional file 1: Table S1. Types of Reported *NPC1* Pathogenic Mutations (Human Gene Mutation Database, <http://www.hgmd.cf.ac.uk/ac/index.php>).

Additional file 2: Table S2. Primer sequences and amplicon sizes for AS-RT-PCR amplification of the exon-exon and exon-intron regions near rs1140458 in *NPC1*.

Additional file 3: Figure S1. A. Schematic presentation of the AS-RT-PCR designed to detect co-allelic variant nucleotide and exon-intron boundary. For each allele-specific PCR, three primers were designed: a common forward exonic primer to amplify the SNV locus, and two reverse primers hybridizing in the downstream exon or intron, respectively. The example shows a variant nucleotide (A, in green) which, if splice modulating, is expected to over-dominate the reference allele in the PCR-amplicon containing the exon-intron junction. **B.** Sanger sequencing chromatograms of the allele specific amplification of the region flanking SNP rs1140458 near 3'5S of in the gene *NPC1*: the amplicon of the canonically spliced exon-exon region (top), shows an equal proportion of the variant and the reference allele. In contrast, in the intron-retaining molecules (second from the top) predominate the alleles with the variant nucleotide (indicated with an arrow). The results were consistent with the reverse primer (bottom two chromatograms).

Additional file 4: Table S3. Distribution of the four categories of sequencing reads (VAREi, VAREf, REFEi, REFEe) across the 75 transcriptomes in the region of rs1140458 by cell type, with P-values, and LOD scores. All the P-values lower than 0.2 are displayed (bold). "% VAR unspliced" shows the proportion of reads bearing the variant nucleotide and retaining the intron, out of all VAR harboring reads. "% REF unspliced" indicates the proportion of reads bearing the wild type nucleotide and retaining the intron, out of all wild type reads. "%VAR (of all unspliced)" indicates the proportion of variant bearing reads among all the intron-retaining reads. The blue rows in the table present homozygote wild type genotype (in regards to rs1140458), and the green rows represent homozygote variant genotype.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MM and PM performed the alignment, variation call and assembly of the transcriptome datasets. MM prepared the manuscript. NE wrote and optimized the computational approach and performed statistical significance computations for read counts. MK and NG provided RPE transcriptome datasets and performed the wet-lab studies. AH designed the study, the analytical strategy and the wet-lab confirmation, supervised the project, and finalized the paper. All authors read and approved the final manuscript.

Acknowledgments

This work is supported by Award Number UL1TR000075 from the NIH National Center for Advancing Translational Sciences and the Clinical and Translational Science Institute at Children's National Medical Center (CTSI-CN) to Anelia Horvath, and Georgetown University Dean's Pilot Award Number GX4002-753 to Nady Golestaneh.

Author details

¹Department of Pharmacology and Physiology, The George Washington University, 2300 Eye Street NW, Suite 653, 20037 Washington, DC, USA.

²McCormick Genomic and Proteomic Center, The George Washington University, 20037 Washington, DC, USA. ³Department of Ophthalmology, Georgetown University, School of Medicine, 20057 Washington, DC, USA.

⁴Department of Neurology, Georgetown University, School of Medicine, 20057 Washington, DC, USA. ⁵Department of Biochemistry and Molecular & Cellular Biology, Georgetown University, School of Medicine, 20057 Washington, DC, USA.

Received: 2 August 2014 Accepted: 9 October 2014

Published: 21 November 2014

References

1. Vanier MT: **Niemann-Pick disease type C.** *Orphanet J Rare Dis* 2010, **5**:16.
2. Blom TS, Linder MD, Snow K, Pihko H, Hess MW, Jokitalo E, Veckman V, Syvänen AC, Ikonen E: **Defective endocytic trafficking of NPC1 and NPC2 underlying infantile Niemann-Pick type C disease.** *Hum Mol Genet* 2002, **12**:257–272.
3. Patterson MC, Hendriksz CJ, Walterfang M, Sedel F, Vanier MT, Wijburg F: **Recommendations for the diagnosis and management of Niemann-Pick disease type C: an update.** *Mol Genet Metab* 2012, **106**:330–344.
4. Bauer P, Knoblich R, Bauer C, Finckh U, Hufen A, Kropp J, Braun S, Kustermann-Kuhn B, Schmidt D, Harzer K, Rolfs A: **NPC1: complete genomic sequence, mutation analysis, and characterization of haplotypes.** *Hum Mutat* 2002, **19**:30–38.
5. Fernandez-Valero EM, Ballart A, Iturriaga C, Lluçh M, Macias J, Vanier MT, Pineda M, Coll MJ: **Identification of 25 new mutations in 40 unrelated Spanish Niemann-Pick type C patients: genotype-phenotype correlations.** *Clin Genet* 2005, **68**:245–254.
6. Fancello T, Dardis A, Rosano C, Tarugi P, Tappino B, Zampieri S, Pinotti E, Corsolini F, Fecarotta S, D'Amico A, Di Rocco M, Uziel G, Calandra S, Bambi B, Filocamo M: **Molecular analysis of NPC1 and NPC2 gene in 34 Niemann-Pick C Italian patients: identification and structural modeling of novel mutations.** *Neurogenetics* 2009, **10**:229–239.
7. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeyasinghe S, Krawczak M, Cooper DN: **Human Gene Mutation Database (HGMD): 2003 update.** *Hum Mutat* 2003, **21**:577–581.
8. Claudepierre T, Paques M, Simonutti M, Buard I, Sahel J, Maue RA, Picaud S, Priefer FW: **Lack of Niemann-Pick type C1 induces age-related degeneration in the mouse retina.** *Mol Cell Neurosci* 2010, **43**:164–176.
9. Phillips SE, Woodruff EA, Liang P, Patten M, Broadie K: **Neuronal loss of Drosophila NPC1a causes cholesterol aggregation and age-progressive neurodegeneration.** *J Neurosci* 2008, **28**:6569–6582.
10. Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Ison J, Keays M, Kurbatova N, Malone J, Mani R, Mupo A, Pedro Pereira R, Pilicheva E, Rung J, Sharma A, Tang YA, Ternent T, Tikhonov A, Welter D, Williams E, Brazma A, Parkinson H, Sarkans U: **ArrayExpress update—trends in database growth and links to data analysis tools.** *Nucleic Acids Res* 2013, **41**:D987–D990.
11. Cancer Genome Atlas Research Network: **Comprehensive genomic characterization defines human glioblastoma genes and core pathways.** *Nature* 2008, **455**:1061–1068.
12. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics* 2009, **25**:1105–1111.
13. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup: **The sequence alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**:2078–2079.
14. Horvath A, Pakala SB, Mudvari P, Reddy SD, Ohshiro K, Casimiro S, Pires R, Fuqua SA, Toi M, Costa L, Nair SS, Sukumar S, Kumar R: **Novel insights into breast cancer genetic variance through RNA sequencing.** *Sci Rep* 2013, **3**:2256.

15. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J: **A general framework for estimating the relative pathogenicity of human genetic variants.** *Nat Genet* 2014, **46**:310–315.
16. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nat Protoc* 2012, **7**:562–578.
17. Mudvari P, Movassagh M, Kowsari K, Seyfi A, Kokkinaki M, Edwards NJ, Golestaneh N, Horvath A: **SNPllice: splice-modulating SNPs from RNA-sequencing data.** *Bioinformatics* 2014, in press.
18. Thorvaldsdottir H, Robinson JT, Mesirov JP: **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.** *Brief Bioinform* 2013, **14**:178–192.
19. Dogan RI, Getoor L, Wilbur WJ, Mount SM: **SplicePort-an interactive splice-site analysis tool.** *Nucleic Acids Res* 2007, **35**:W285–W291.
20. Piva F, Giulietti M, Burini AB, Principato G: **SpliceAid 2: a database of human splicing factors expression data and RNA target motifs.** *Hum Mutat* 2012, **33**:81–85.
21. Woolfe A, Mullikin JC, Elnitski L: **Genomic features defining exonic variants that modulate splicing.** *Genome Biol* 2010, **11**:R20.
22. Ray D, Kazan H, Chan ET, Peña Castillo L, Chaudhry S, Talukder S, Blencowe BJ, Morris Q, Hughes TR: **Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins.** *Nat Biotechnol* 2009, **27**:667–670.
23. Wongsiriroj N, Jiang H, Piantadosi R, Yang KJ, Kluwe J, Schwabe RF, Ginsberg H, Goldberg IJ, Blaner WS: **Genetic dissection of retinoid esterification and accumulation in the liver and adipose tissue.** *J Lipid Res* 2014, **55**:104–114.
24. Chen J, Weiss WA: **Alternative splicing in cancer: implications for biology and therapy.** *Oncogene* 2014, Epub ahead of print.
25. Pikuleva IA, Curcio CA: **Cholesterol in the retina: the best is yet to come.** *Prog Retin Eye Res* 2014, **41**:64–79.

doi:10.1186/2043-9113-4-14

Cite this article as: Movassagh *et al.*: Analysis for co-occurring sequence features identifies link between common synonymous variant and an early-terminated NPC1 isoform. *Journal of Clinical Bioinformatics* 2014 **4**:14.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

