A novel co-segregating *DCTN1* splice site variant in a family with Bipolar Disorder may hold the key to understanding the etiology

Supplementary materials

André Hallen[1] and Arthur J.L. Cooper[2]

[1]André Hallen, Ryde, Sydney, NSW 2112, Australia.

[2]Arthur J. L. Cooper, Department of Biochemistry and Molecular Biology, New York Medical College, Valhalla, New York 10595, USA.

Corresponding authors

André Hallen, andrehallen@yahoo.com.au. Arthur J.L. Cooper, arthur_cooper@nymc.edu.

Contributions

AH initiated this research project, was responsible for experimental design, and performed all analyses. AH and AJLC wrote and edited the manuscript.

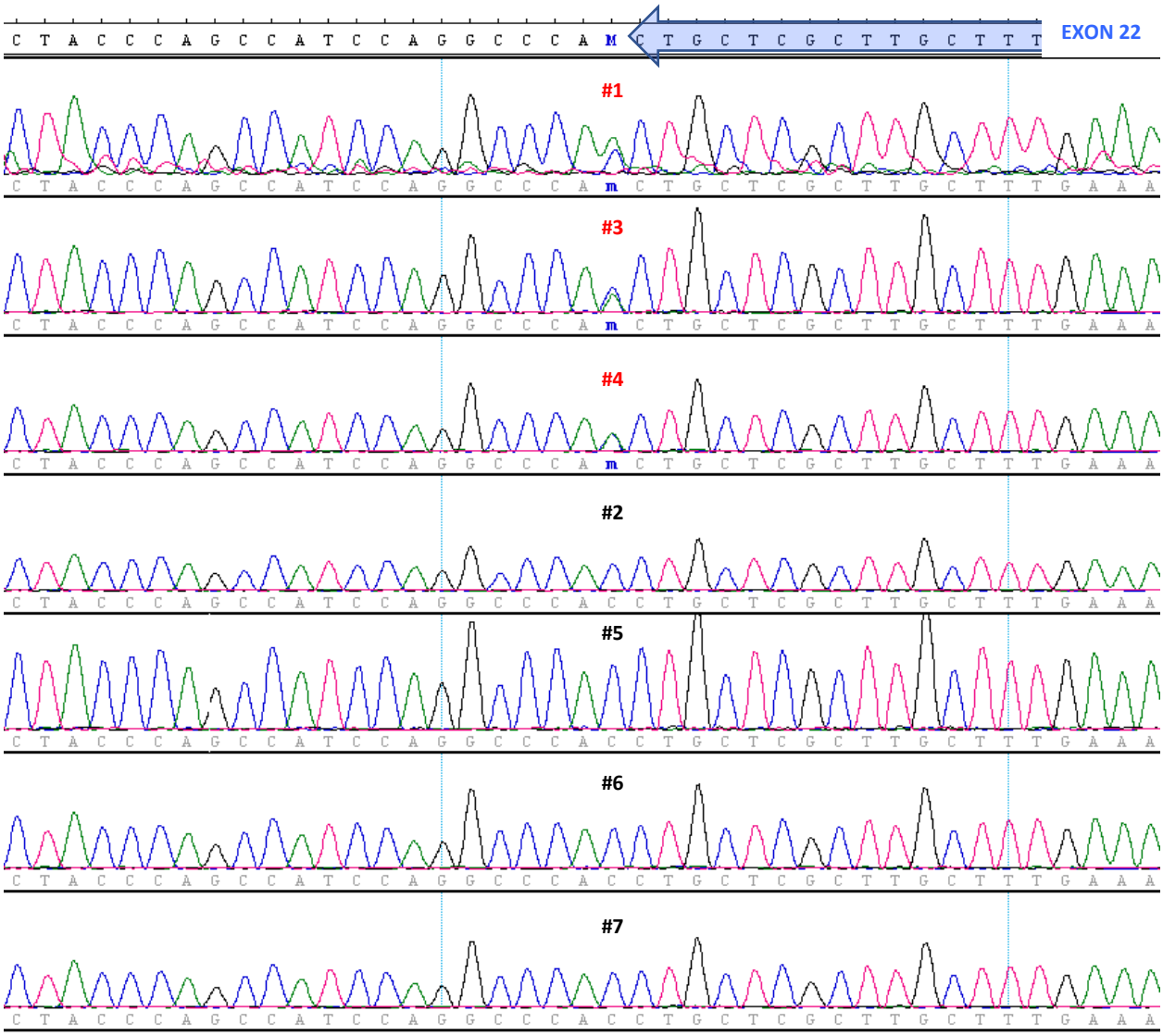

Figure S1. The *DCTN1* variant discovered using NGS was also verified using Sanger sequencing. The chromatogram shows heterozygous variant calls for the three affected family members (Fig. 1; #1,3,4), and a normal allele for related controls (Fig. 1; #2,5,6,7). The novel variant (C→A) is predicted to disrupt the invariant donor splice site (IVS22+1G>T [NM_004082.4: c.2628+1G>T]) of the *DCTN1* gene which is coded on the anti-sense strand of chromosome 2 (2p13.1). Chromatograms were aligned, and variants called using SeqMan Pro 15.1 software (DNASTAR Lasergene, USA). PCR primers used: Forward-primer 5'-TCATACTCCCCCTCCTGCAT-3' and Reverse-primer 5'-AATGAGGGGCTACTTGTGGC-3'. The forward PCR primer was used for Sanger sequencing.

Figure S2. (following page) p150^{Glued} protein (UniProt ID: Q14203, 1278AA). The colored bars under the amino acid sequence illustrate the different regions: ***Red***, the region predicted in this study to be missing in ct-p150^{Glued} due to a splice site variant; ***Green***, CAP_GLY and coiled-coil domains; ***Magenta***, microtubule binding domain; ***Purple,*** dynein IC binding region; ***Orange***, kinesin II binding region; ***Yellow***, ARPI binding region; ***Blue***, HAP1 binding region.


MAQSKRHVYS RTPSGSRMSA EASARPLRVG SRVEVIGKGH RGTVAYVGAT LFATGKWVGV ILDEAKGKND GTVQGRKYFT CDEGHGIFVR QSQIQVFEDG ADTTSPETPD SSASKVLKRE GTDTTAKTSK LRGLKPKKAP
140
CAP_GLY
MICROTUBULE BINDING SITE
TARKTTTRRP KPTRPASTGV AGASSSLGPS GSASAGELSS SEPSTPAQTP LAAPIIPTPV LTSPGAVPPL PSPSKEEEGL RAQVRDLEEK LETLRLKRAE DKAKLKELEK HKIQLEQVQE WKSKMQEQQA DLQRRLKEAR
280
MICROT... SITE
Coiled-coil region I
DYNEIN IC binding site
KEAKEALEAK ERYMEEMADT ADAIEMATLD KEMAEERAES LQQEVEALKE RVDELTTDLE ILKAEIEEKG SDGAASSYQL KQLEEQNARL KDALVRMRDL SSSEKQEHVK LQKLMEKKNQ ELEVVRQQRE RLQEELSQAE
420
Coiled-coil region I
DYNEIN IC binding site
STIDELKEQV DAALGAEEMV EMLTDRNLNL EEKVRELRET VGDLEAMNEM NDELQENARE TELELREQLD MAGARVREAQ KRVEAAQETV ADYQQTIKKY RQLTAHLQDV NRELTNQQEA SVERQQQPPP ETFDFKIKFA
560
Coiled-coil region I
DYNEIN IC binding site
ETKAHAKAIE MELRQMEVAQ ANRHMSLLTA FMPDSFLRPG GDHDCVLVLL LMPRLICKAE LIRKQAQEKF ELSENCSERP GLRGAAGEQL SFAAGLVYSL SLLQATLHRY EHALSQCSVD VYKKVGSLYP EMSAHERSLD
700
DYNEIN IC binding site
KINESIN II BINDING SITE
FLIELLHKDQ LDETVNVEPL TKAIKYYQHL YSIHLAEQPE DCTMQLADHI KFTQSALDCM SVEVGRLRAF LQGGQEATDI ALLLRDLETS CSDIRQFCKK IRRRMPGTDA PGIPAALAFG PQVSDTLLDC RKHLTWVVAV
840
DYNEIN IC binding site
KINESIN II BINDING SITE
LQEVAAAAAQ LIAPLAENEG LLVAALEELA FKASEQIYGT PSSSPYECLR QSCNILISTM NKLATAMQEG EYDAERPPSK PPPVELRAAA LRAEITDAEG LGLKLEDRET VIKELKKSLK IKGEELSEAN VRLSLLEKKL
980
Coiled-coil region II
MISSING TRUNCATED
DSAAKDADER IEKVQTRLEE TQALLRKKEK EFEETMDALQ ADIDQLEAEK AELKQRLNSQ SKRTIEGLRG PPPSGIATLV SGIAGEEQQR GAIPGQAPGS VPGPGLVKDS PLLLQQISAM RLHISQLQHE NSILKGAQMK
1120
Coiled-coil region II
MISSING TRUNCATED
ARPI BINDING SITE
HAPI BINDING SITE
ASLASLPPLH VAKLSHEGPG SELPAGALYR KTSQLLETLN QLSTHTHVVD ITRTSPAAKS PSAQLMEQVA QLKSLSDTVE KLKDEVLKET VSQRPGATVP TDFATFPSSA FLRAKEEQQD DTVYMGKVTF SCAAGFGQRH
1260
Coiled-coil region III
HAPI BINDING SITE
MISSING TRUNCATED
RLVLTQEQLH QLHSRLIS
MISSING TRUNCATED

Table S1. Variant counts for the *DCTN1* variant [NC_000002.11: g.74593585C>A (GRCh37)] in affected subjects (#1,3,4) and unaffected controls (#2,5-16).

Sample #	Diagnosis	*DCTN1* Count C	*DCTN1* Count A	*DCTN1* Sanger validation	NGS (Illumina)
1	**BDIII**	56	79	✓	30X WGS[a] 100X WES[d]
3	**BDI**	229	233	✓	300X WES[b] 100X WES[d]
4	**BDI**	309	310	✓	30X WGS[a] 300X WES[b] 100X WES[c] 100X WES[d]
2	RC	153	0	✓	30X WGS[a] 100X WES[d]
5	RC	433	0	✓	300X WES[b] 100X WES[d]
6	RC	125	0	✓	100X WES[c]
7	RC	146	0	✓	100X WES[c]
8	RC	120	0	ND	100X WES[d]
9	RC	114	0	ND	100X WES[d]
10	RC	58	0	ND	100X WES[d]
11	RC	108	0	ND	100X WES[d]
12	RC	102	0	ND	100X WES[d]
13	RC	107	0	ND	100X WES[d]
14	RC	87	0	ND	100X WES[d]
15	RC	158	0	ND	100X WES[d]
16	NRC	105	0	ND	100X WES[d]

[a] Illumina HiSeq 2000, 2×100 bp paired-end reads, Illumina, USA, (2011)

[b] Illumina HiSeq 2000, 2×100 bp paired-end reads, Illumina Nextera Rapid Capture Expanded Exome, Ramaciotti Sequencing Centre, Australia, (2014)

[c] Illumina HiSeq 4000, 2×100 paired-end reads, Agilent SureSelect All Exome V5, Macrogen, South Korea, (2016)

[d] Illumina HiSeq 4000, 2×100 paired-end reads, Agilent SureSelect All Exome V6, Macrogen, South Korea, (2017, 2018)

ND, not determined

Table S2. Genomic conservation data for the *DCTN1* splice variant*

Tool	**GERP++RS**[1]	**PhastCons100way_vertebrate**[2]	**PhyloP100way_verebrate**[2]
Score	5.08 (range 1-6.18)	1 (range 0-1)	5.75 (range -20-10)
Conservation	Highly conserved	Highly conserved	Highly conserved

* A multiple alignment of 100 vertebrate genomes in UCSC Genomes (genome.ucsc.edu) shows all vertebrate genomes sequenced thus far are completely conserved at this genomic coordinate except for that of the Medium Ground Finch (*Geospiza fortis*). This is most likely a bioinformatic error in the database.

Table S3. Bioinformatic predictions for the *DCTN1* splice variant

Tool	CADD[3]	DANN[4]	FATHMM-XF[5]	MutationTaster[6]
Score	25.6 [>20 = 1% most pathogenic]	0.9946 [range 0-1]	0.996 [range 0-1]	1.0 [range 0-1]
Prediction	deleterious	deleterious	deleterious	deleterious

1. Davydov, E.V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* **6**, e1001025 (2010).
2. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110-121 (2010).
3. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310-315 (2014).
4. Quang, D., Chen, Y. & Xie, X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **31**, 761-763 (2015).
5. Rogers, M.F. *et al.* FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* **34**, 511-513 (2018).
6. Schwarz, J.M., Cooper, D.N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361-362 (2014).