**The American Journal of Human Genetics, Volume *88***

## Supplemental Data

# *DPY19L2* Deletion as a Major Cause of Globozoospermia

**Isabelle Koscinski, Elias ElInati, Camille Fossard, Claire Redin, Jean Muller, Juan Velez de la Calle, Françoise Schmitt, Mariem Ben Khelifa, Pierre Ray, Zaid Kilani, Christopher L.R. Barratt, and Stéphane Viville**

### Supplemental Material and Methods

*Patients and controls*

The pedigree was identified at the Farah Hospital, Amman, Jordan. Other patients were identified at the University Hospital of Strasbourg or provided by collaborators Pr. C. Jimenez (Dijon), Dr. F. Carré-Pigeon (Reims), Pr. J.M. Grillot (Marseille), Dr. F. Brugnon (Clermont Ferrand), Dr. D. De Briel (Colmar), Dr. S. Declève-Paulhac (Limoges). This study was approved by the local Ethical Committee (Comité de protection de la personne, CPP) of Strasbourg University hospital. For each case analyzed informed written consent was obtained accordingly to the CPP recommendations.

*DNA preparation*

Genomic DNA was extracted either from peripheral blood leucocytes using QIAamp DNA Blood Midi Kit (QIAGEN, Germany) or from saliva using Oragene DNA Self-Collection Kit (DNAgenotech, Ottawa, Canada).

*SNP mapping*

A genomewide scan was performed on the 4 affected and the 3 healthy brothers of the family using the Affymetrix GeneChip®Mapping 10K array (Affymetrix, Santa Clara, CA). Sample processing and labeling were performed according to the manufacturer's instructions (Affymetrix Mapping 10K 2.0 Assay Manual, Version 1.0, 2004). The hybridization was performed using a GeneChip Hybridization oven 640, washed with the GeneChip Fluidics Station 450 and scanned with a GeneChip Scanner 3000. SNP allele calls were generated by the GeneChip DNA Analysis Software version 3.0.2 (GDAS). Regions of homozygosity were defined by the presence of more than 25 consecutive homozygous SNPs. The unique region of homozygosity by descent

segregating with the disease in the family was entered and screened in the UCSC Genome Browser (http://genome.ucsc.edu , version March 2006) for selection of candidate genes.

*Deletion analysis*

PCR products covering exon 2, 7, 9, 10, 13, 17 and 21 of *DPY19L2* and their exon-intron boundaries were amplified using genomic DNA from a control and one of the affected brothers and subsequently of the 24 patients with globozoospermia. A walk 5' and 3' was then carried out to identify the deletion, 3 amplicons were tested on the 5' and 2 on the 3' region of the gene. All primer sequences and PCR conditions are available in the supplementary data (Table 1). Because of the high conservation of the duplicated regions special care was taken to choose specific oligonucleotides. Even so, all amplicons were sequenced in order to control for the specificity of the PCR.

CLUSTALW 2.0.12 multiple sequence alignment

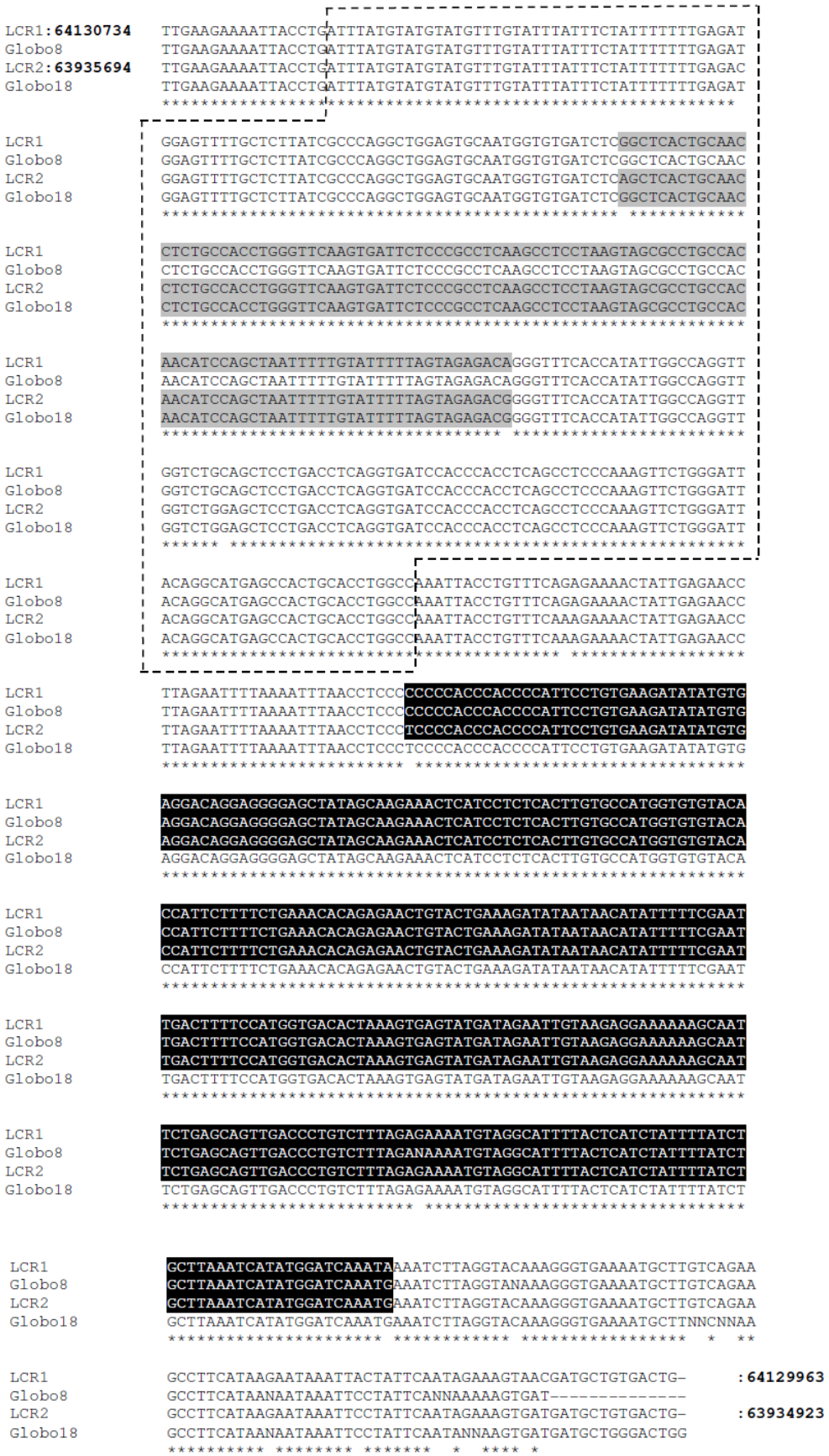

**Figure S1**. **Multiple Sequence Alignment of the Genomic Region Surrounding the Identified Breakpoints**

The alignment was performed using the clustalw 2.0 program[1] using as query sequences: the reference human sequence for both LCR1 (64 119 249-64 146 247 bp) and LCR2 (63 923 419-63 951 619 bp) and the sequences from 2 representative patients (Globo8 and Globo18) for each Breakpoint (BP) are presented. The exact genomic positions of both LCRs fragments shown are : LCR1-fragment: 64 129 963 - 64 130 734 bp; LCR2-fragment: 63 934 923 - 63 935 694 bp. Identical residues are marked with a star. Breakpoint 1 (BP1) is highlighted in gray and BP2 is highlighted in black. The block surrounded with dashed-lines indicates the AluSq2 repeat element.

**Table S1. Results of the SNP Array for the Four Affected Brothers**

| Chromosome 12 | Genotype Call in Sibling | | | | |
|---|---|---|---|---|---|
| SNP | Globo1 | Globo2 | Globo3 | Globo4 | Globo18 |
| rs345945 | BB | BB | BB | BB | AA |
| rs505071 | BB | BB | BB | BB | BB |
| rs2839798 | BB | BB | BB | BB | AB |
| rs699603 | AA | AA | AA | AA | AA |
| rs722918 | AA | AA | AA | AA | BB |
| rs1146122 | AA | AA | AA | AA | BB |
| rs722526 | BB | BB | BB | BB | AA |
| rs1445442 | BB | BB | BB | BB | BB |
| rs1373877 | AA | AA | AA | AA | AA |
| rs4129000 | BB | BB | BB | BB | BB |
| rs974349 | AA | AA | AA | AA | AB |
| rs952562 | BB | BB | BB | BB | AB |
| rs2029692 | BB | BB | BB | BB | BB |
| rs1343807 | BB | BB | BB | BB | BB |
| rs1405467 | AA | AA | AA | AA | AA |
| rs1480065 | BB | BB | BB | BB | AB |
| rs2870793 | BB | BB | BB | BB | BB |
| rs3850590 | AA | AA | AA | AA | AA |
| rs1905444 | BB | BB | BB | BB | AB |
| rs722748 | AA | AA | AA | AA | AB |
| rs722749 | BB | BB | BB | BB | AB |
| rs2193047 | AA | AA | AA | AA | BB |
| rs2870951 | BB | BB | BB | BB | BB |
| rs2870950 | AA | AA | AA | AA | AA |
| rs973328 | BB | BB | BB | BB | AB |
| rs952218 | BB | BB | BB | BB | AB |
| rs1908660 | BB | BB | BB | BB | AB |
| rs1908682 | AA | AA | AA | AA | AB |
| rs2172989 | BB | BB | BB | BB | AB |

The thirty homozygous SNP shared by the four infertile brothers which cover an 6.4 Mb region on the chromosome 12 and the analysis of Globo18 patient showing a region of homozygosity covering the *DPY19L2* locus, showed in grey.

**Table S2. List of Genes in the Homozygous Region Shared between the Infertile Patients**

| Ensembl Gene ID | Description |
|---|---|
| ENSG00000251857 | 5S ribosomal RNA [Source:RFAM;Acc:RF00001] |
| ENSG00000240075 | |
| ENSG00000252883 | Small nucleolar RNA SNORD112 [Source:RFAM;Acc:RF01169] |
| ENSG00000238475 | Small nucleolar RNA U13 [Source:RFAM;Acc:RF01210] |
| ENSG00000238440 | Small nucleolar RNA U13 [Source:RFAM;Acc:RF01210] |
| ENSG00000223294 | Small nucleolar RNA SNORD83 [Source:RFAM;Acc:RF00137] |
| ENSG00000238592 | Small nucleolar RNA U13 [Source:RFAM;Acc:RF01210] |
| ENSG00000238528 | Small nucleolar RNA U13 [Source:RFAM;Acc:RF01210] |
| ENSG00000206650 | Small nucleolar RNA SNORA70 [Source:RFAM;Acc:RF00156] |
| ENSG00000200814 | U6 spliceosomal RNA [Source:RFAM;Acc:RF00026] |
| ENSG00000202034 | U6 spliceosomal RNA [Source:RFAM;Acc:RF00026] |
| ENSG00000200296 | U1 spliceosomal RNA [Source:RFAM;Acc:RF00003] |
| ENSG00000212298 | U6 spliceosomal RNA [Source:RFAM;Acc:RF00026] |
| ENSG00000251788 | U5 spliceosomal RNA [Source:RFAM;Acc:RF00020] |
| ENSG00000221564 | U6atac minor spliceosomal RNA [Source:RFAM;Acc:RF00619] |
| ENSG00000207099 | U6 spliceosomal RNA [Source:RFAM;Acc:RF00026] |
| ENSG00000252770 | U7 small nuclear RNA [Source:RFAM;Acc:RF00066] |
| ENSG00000241412 | |
| ENSG00000245867 | |
| ENSG00000246324 | |
| ENSG00000250314 | |
| ENSG00000250748 | |
| ENSG00000248947 | |
| ENSG00000247008 | |
| ENSG00000245302 | |
| ENSG00000251695 | |
| ENSG00000247363 | |
| ENSG00000111530 | cullin-associated and neddylation-dissociated 1 [Source:HGNC Symbol;Acc:30688] |
| ENSG00000127334 | dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2 [Source:HGNC Symbol;Acc:3093] |
| ENSG00000111537 | interferon, gamma [Source:HGNC Symbol;Acc:5438] |
| ENSG00000111536 | interleukin 26 [Source:HGNC Symbol;Acc:17119] |
| ENSG00000127318 | interleukin 22 [Source:HGNC Symbol;Acc:14900] |
| ENSG00000111554 | Mdm1 nuclear protein homolog (mouse) [Source:HGNC Symbol;Acc:29917] |
| ENSG00000127314 | RAP1B, member of RAS oncogene family [Source:HGNC Symbol;Acc:9857] |
| ENSG00000111581 | nucleoporin 107kDa [Source:HGNC Symbol;Acc:29914] |
| ENSG00000175782 | solute carrier family 35, member E3 [Source:HGNC Symbol;Acc:20864] |
| ENSG00000135679 | Mdm2 p53 binding protein homolog (mouse) [Source:HGNC Symbol;Acc:6973] |
| ENSG00000226959 | cDNA, FLJ92387 [Source:UniProtKB/TrEMBL;Acc:B2R594] |
| ENSG00000135678 | carboxypeptidase M [Source:HGNC Symbol;Acc:2311] |
| ENSG00000111605 | microRNA 1279 [Source:HGNC Symbol;Acc:35357] |
| ENSG00000090382 | lysozyme [Source:HGNC Symbol;Acc:6740] |

| Ensembl Gene ID | Description |
|---|---|
| ENSG00000127337 | YEATS domain containing 4 [Source:HGNC Symbol;Acc:24859] |
| ENSG00000166225 | fibroblast growth factor receptor substrate 2 [Source:HGNC Symbol;Acc:16971] |
| ENSG00000166226 | chaperonin containing TCP1, subunit 2 (beta) [Source:HGNC Symbol;Acc:1615] |
| ENSG00000198812 | leucine rich repeat containing 10 [Source:HGNC Symbol;Acc:20264] |
| ENSG00000242922 | |
| ENSG00000214304 | keratin 8 pseudogene 19 [Source:HGNC Symbol;Acc:33371] |
| ENSG00000213363 | |
| ENSG00000241941 | |
| ENSG00000213352 | |
| ENSG00000139239 | |
| ENSG00000250517 | |
| ENSG00000240027 | |
| ENSG00000242087 | |
| ENSG00000243024 | ribosomal protein S11 [Source:HGNC Symbol;Acc:10384] |
| ENSG00000243020 | |
| ENSG00000215208 | |
| ENSG00000213344 | |
| ENSG00000241749 | ribosomal protein SA pseudogene 52 [Source:HGNC Symbol;Acc:35752] |
| ENSG00000213343 | |
| ENSG00000228144 | |
| ENSG00000225422 | RNA binding motif, single stranded interacting protein 1 pseudogene 1 [Source:HGNC Symbol;Acc:9908] |
| ENSG00000244432 | |
| ENSG00000236946 | |
| ENSG00000240087 | |
| ENSG00000241825 | |
| ENSG00000241765 | |
| ENSG00000198673 | family with sequence similarity 19 (chemokine (C-C motif)-like), member A2 [Source:HGNC Symbol;Acc:21589] |
| ENSG00000135655 | ubiquitin specific peptidase 15 [Source:HGNC Symbol;Acc:12613] |
| ENSG00000061987 | MON2 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:29177] |
| ENSG00000221949 | Putative uncharacterized protein C12orf61 [Source:UniProtKB/Swiss-Prot;Acc:Q8N7H1] |
| ENSG00000111110 | protein phosphatase, Mg2+/Mn2+ dependent, 1H [Source:HGNC Symbol;Acc:18583] |
| ENSG00000166148 | arginine vasopressin receptor 1A [Source:HGNC Symbol;Acc:895] |
| ENSG00000177990 | Protein dpy-19 homolog 2 (Dpy-19-like protein 2) [Source:UniProtKB/Swiss-Prot;Acc:Q6NUT2] |
| ENSG00000118600 | transmembrane protein 5 [Source:HGNC Symbol;Acc:13530] |
| ENSG00000196935 | SLIT-ROBO Rho GTPase activating protein 1 [Source:HGNC Symbol;Acc:17382] |
| ENSG00000174206 | UPF0536 protein C12orf66 [Source:UniProtKB/Swiss-Prot;Acc:Q96MD2] |
| ENSG00000185306 | Uncharacterized protein C12orf56 [Source:UniProtKB/Swiss-Prot;Acc:Q8IXR9] |
| ENSG00000184575 | exportin, tRNA (nuclear export receptor for tRNAs) [Source:HGNC Symbol;Acc:12826] |
| ENSG00000183735 | TANK-binding kinase 1 [Source:HGNC Symbol;Acc:11584] |
| ENSG00000153179 | Ras association (RalGDS/AF-6) domain family member 3 [Source:HGNC Symbol;Acc:14271] |
| ENSG00000135677 | glucosamine (N-acetyl)-6-sulfatase [Source:HGNC Symbol;Acc:4422] |
| ENSG00000111490 | TBC1 domain family, member 30 [Source:HGNC Symbol;Acc:29164] |
| ENSG00000156076 | WNT inhibitory factor 1 [Source:HGNC Symbol;Acc:18081] |

| Ensembl Gene ID | Description |
|---|---|
| ENSG00000174106 | LEM domain containing 3 [Source:HGNC Symbol;Acc:28887] |
| ENSG00000174099 | methionine sulfoxide reductase B3 [Source:HGNC Symbol;Acc:27375] |
| ENSG00000149948 | high mobility group AT-hook 2 [Source:HGNC Symbol;Acc:5009] |
| ENSG00000197301 | Putative uncharacterized protein ENSP00000348547 [Source:UniProtKB/TrEMBL;Acc:B7WPI5] |
| ENSG00000139233 | LLP homolog, long-term synaptic facilitation (Aplysia) [Source:HGNC Symbol;Acc:28229] |
| ENSG00000155957 | transmembrane BAX inhibitor motif containing 4 [Source:HGNC Symbol;Acc:24257] |
| ENSG00000090376 | interleukin-1 receptor-associated kinase 3 [Source:HGNC Symbol;Acc:17020] |
| ENSG00000127311 | helicase (DNA) B [Source:HGNC Symbol;Acc:17196] |
| ENSG00000155974 | glutamate receptor interacting protein 1 [Source:HGNC Symbol;Acc:18708] |
| ENSG00000199179 | hsa-let-7i [Source:miRBase;Acc:MI0000434] |
| ENSG00000252801 | |
| ENSG00000207546 | hsa-mir-548c [Source:miRBase;Acc:MI0003630] |
| ENSG00000211577 | |
| ENSG00000252547 | |
| ENSG00000221422 | |
| ENSG00000252660 | Y RNA [Source:RFAM;Acc:RF00019] |
| ENSG00000222744 | 7SK RNA [Source:RFAM;Acc:RF00100] |

The first column reports the Ensembl Gene number of the 101 genes contained in the homozygous region. From this, forty genes have been selected according to their testis expression and their putative role in the gametogenesis. These genes are presented in the second column.