Supplementary materials

Supplement to: Intraventricular fetus-in-fetu with extensive de novo gain in genetic copy number
Online Method

Genetic sequencing and data analysis

DNA extract and detect
Genomic DNA extracted from the tissue of the fetus-in-fetu and the peripheral blood of the host child and parents was fragmented to an average size of ~350bp and subjected to DNA library creation using established Illumina paired-end protocols. The Illumina Novaseq 6000 platform (Illumina Inc., San Diego, CA, USA) was utilized for genomic DNA sequencing in Novogene Bioinformatics Technology Co., Ltd (Beijing, China) to generate 150-bp paired-end reads with a minimum coverage of 10× for 99% of the genome (mean coverage of 30×).

Data analysis
After sequencing, basecall files conversion and demultiplexing were performed with bcl2fastq software (Illumina). The resulting fastq data were submitted to in-house quality control software for removing low quality reads, and then were aligned to the reference human genome (hs37d5) using the Burrows-Wheeler Aligner (bwa)\cite{1}, and duplicate reads were marked using sambamba tools\cite{2}.

SNP/INDEL calling
Single nucleotide variants (SNVs) and indels were called with samtools to generate gVCF\cite{3}. The raw calls of SNVs and INDELs were further filtered with the following inclusion thresholds: 1) read depth > 4; 2) Root-Mean-Square mapping quality of covering reads > 30; 3) the variant quality score > 20.

CNV calling
The copy number variants (CNVs) were detected with software ControlFREEC(v9.1)\cite{4}, using a 1-kb as threshold of duplication and deletion.

SV calling
The structural variants(SVs) were detected with software LUMPY(version v0.2.13)\cite{5}.

Annotation
Annotation was performed using ANNOVAR (2017June8) \cite{6}. Annotations included minor allele frequencies from public control data sets as well as deleteriousness and conservation scores enabling further filtering and assessment of the likely pathogenicity of variants.

Rare variants filtering
Filtering of rare variants was performed as follows: (1) variants with a MAF less than 0.01 in 1000 genomic data (1000g_all)\cite{7}, esp6500siv2_all\cite{8}, gnomAD data (gnomAD_ALL and gnomAD_EAS)\cite{9} and in house Novo-Zhonghua exome database
from Novogene; (2) Only SNVs occurring in exons or splice sites (splicing junction 10 bp) are further analyzed since we are interested in amino acid changes. (3) Then synonymous SNVs which are not relevant to the amino acid alteration predicted by dbscSNV are discarded; The small fragment non-frameshift (<10bp) indel in the repeat region defined by RepeatMasker are discarded. (4) Variations are screened according to scores of SIFT[10], Polyphen[11], MutationTaster[12] and CADD[13] softwares. The potentially deleterious variations are reserved if the score of more than half of these four softwares support harmfulness of variations[14]. Sites(>2bp) did not affect alternative splicing were removed.

4. Kinship analysis

Relationship between proband and parents was estimated using the pairwise identity-by-descent (IBD) calculation in PLINK[15]. The IBD sharing between the proband and parents in all trios is over 50%.

References

[15] Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based