Precision medicine in diabetes: A non-invasive prenatal diagnostic test for the determination of fetal glucokinase mutations

Nouspikel, Thierry; Blouin, Jean-Louis; Puder, Jardena J; Koehler, Bettina; Schwitzgebel Luscher, Valérie

Supplementary methods

Panel design

98 SNPs were selected throughout a 0.7 Mbase genomic region surrounding the *GCK* gene, based on a rate of heterozygosity >0.3 in the general population (sources: gnomAD, 1000-genomes phase 3, and Illumina Omni microarray data, see supplementary figure S2 and supplementary xlsx file). Additionally, the panel included the *GCK*:p.Gly246Arg mutation. A QIASeq Targeted DNA custom panel featuring molecular barcodes (Qiagen) was designed to target 50 bp regions encompassing each SNP.

Samples and DNA extraction

Genomic DNA was extracted from venous blood or cord blood using the DNeasy blood and tissue kit and a Qiacube robot (Qiagen). Maternal venous blood was collected into Streck BCT tubes (Streck) between 12 weeks and 30 weeks amenorrhea. A 500 µl aliquot of blood was withdrawn to extract genomic DNA as above. Plasma was prepared by centrifugation for 10 min at 1'600 x g, 4°C, collected, centrifuged 10 min at 16'000 x g, 4°C and stored at -80°C. Circulating cell-free DNA was extracted from plasma with the QIAamp MinElute ccfDNA kit (Qiagen) according to manufacturer's instructions, quantified with a Qubit fluorimeter (Molecular Probes) and assayed for quality with Kapa hgDNA quantification and QC kit (Roche) as previously described [1].

Library construction

QIASeq targeted libraries were built from 40 ng genomic DNA, using the manufacturer's protocol for genomic DNA and from 7.2 ng, 15 ng and 10 ng ccfDNA for the 3 samples of family A, and 15 ng ccfDNA for family B, using the circulating DNA protocol. For family B, a second library was built from an additional 6.1 ng ccfDNA, and sequencing data were combined with data from the first library. Libraries were quantified by fluorimetry and size profiles were verified with a Fragment Analyzer (Agilent).

Sequencing

Samples were pooled and sequenced as 2 x 75 nucleotides with a NextSeq500 sequencer (Illumina, USA). For each sample, 4 to 5 million reads were obtained.

Analysis

Qiagen smCounter2 pipeline [2] was used to align and filter reads and, in parallel, to deduplicate barcodes and build consensus reads using fgbio subroutines (https://github.com/fulcrumgenomics/fgbio). Allele counts were extracted from the resulting

BAM files with bam-readcount (https://github.com/genome/bam-readcount). For family A, mean read depth was 14,000 for plasma sample 1 and 17,000 for plasma sample 2, whereas the plasma sample from the second pregnancy was sequenced 8 times with read depths varying from 1,600 to 113,000. For family 2, read depth was 7,000 for the first library and 15,500 for the second. Mean barcode depth after consensus building for family A, was 300 for sample 1, 520 for sample 2, and 330 to 720 for sample 3. For family B is was 234 for the first library and 124 for the second,

Parent and child genotypes were deduced from allelic counts and parental haplotypes were assembled (see supplementary xlsx file). In family A, the parents were homozygous for different alleles at 5 SNPs and these type 1 SNPs were used to calculate the fetal fraction, which was 8.9% at 12 weeks, 19.5% at 30 weeks and 14.4% for the second pregnancy at 22 weeks. For 18 SNPs the mother was heterozygous and the father homozygous and these type 4 SNPs were used for relative haplotype dosage (RHDO) analysis. Among these, 17 were alpha SNPs, i.e. the paternal allele corresponded to maternal haplotype 1 (high-risk, carrying the mutation) and one was a beta SNP, i.e. the paternal allele corresponded to maternal haplotype 2 (low-risk).

In family B, the parents were homozygous for different alleles at 31 type 1 SNPs, which were used to calculate fetal fraction. There were, however, only five type 4 SNPs, suitable for RHDO analysis, at which the mother was heterozygous and the father homozygous. Of these, there were 3 beta SNPs upstream of the mutation and 2 alpha SNPs downstream of it.

RHDO was performed in Excel, as described by Lo et al. [3]. Alpha and beta SNPs located upstream and downstream of the mutation were analyzed separately, to detect the unlikely occurrence of meiotic recombination within the region of interest. Once this possibility was excluded, a global likelihood ratio was calculated for all alpha SNPs and for all beta SNPs. For family A, the highest likelihood ratio was that of alpha SNPs and was 1.7×10^{9} for sample 1, 9.9×10^{22} for sample 2 and 1.4×10^{7} for the second pregnancy. For family B, the likelihood ratio was only 685 for the first library, i.e. inferior to the threshold of 1,200 proposed by Lo et al. This prompted us to analyze more DNA in a second library, and pooling the molecular counts for each SNP resulted in a satisfactory likelihood ratio of 2243.

A custom script in visual basic was run in Excel to randomly select subsets of SNPs for RHDO analysis and record the resulting global likelihood ratio. A maximum of 10'000 tests were performed for each number of SNPs and the fraction that failed to reach a likelihood ratio of 1200:1 was recorded.

1 Nikolaev S, Lemmens L, Koessler T, Blouin J-L, Nouspikel T. Circulating tumoral DNA: Preanalytical validation and quality control in a diagnostic laboratory. *Analytical biochemistry* 2018; **542**: 34–39.

2 Xu C, Gu X, Padmanabhan R *et al.* smCounter2: an accurate low-frequency variant caller for targeted sequencing data with unique molecular identifiers. *Bioinformatics (Oxford, England)* 2019; **35**: 1299–1309.

3 Lo YMD, Chan KCA, Sun H *et al.* Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Science translational medicine* 2010; **2**: 61ra91-61ra91.