Title:

Validation and application of a novel integrated genetic screening method to a cohort of 1,112 men with idiopathic azoospermia or severe oligozoospermia

Authors:

Manon S. Oud¹, Liliana Ramos², Moira K. O'Bryan³, Robert I. McLachlan⁴, Özlem Okutman⁵, Stephane Viville⁵, Petra F. de Vries¹, Dominique F.C.M. Smeets¹, Dorien Lugtenberg¹, Jayne Y. Hehir-Kwa¹, Christian Gilissen¹, Maartje van de Vorst¹, Lisenka E.L.M. Vissers¹, Alexander Hoischen^{1,6}, Aukje M. Meijerink², Kathrin Fleischer², Joris A. Veltman^{1,7,8*}, Michiel J. Noordam^{8*}

Affiliations:

1. Department of Human Genetics, Donders Institute for Brain, Cognition and Behavior, Radboudumc, The Netherlands
2. Department of Gynaecology and Obstetrics, Radboudumc, The Netherlands
3. The Development and Stem Cells Program of Monash Biomedicine Discovery Institute and the Department of Anatomy and Developmental Biology, Monash University, Australia
4. Hudson Institute of Medical Research, Monash University, Australia
5. Laboratory of Genetic Diagnostics, UF3472-Genetics of Infertility, University Hospital Strasbourg, France
6. Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboudumc, the Netherlands
7. Institute of Genetic Medicine, Newcastle University, Newcastle, United Kingdom

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/humu.23312](#).

This article is protected by copyright. All rights reserved.

8. Department of Genetics and Cell Biology, Maastricht UMC+, The Netherlands

* Shared last authors

Correspondence

Correspondence should be addressed to J.A. Veltman (joris.veltman@newcastle.ac.uk)

Grant numbers:

This work was supported by a VICI grant from The Netherlands Organisation for Scientific Research (918-15-667 to JAV). MKOB and RML were supported by fellowships from the National Health and Medical Research Council of Australia (APP1058356 and APP1022327XXX).

Abstract:

Microdeletions of the Y chromosome (YCMs), Klinefelter syndrome (47,XXY) and *CFTR* mutations are known genetic causes of severe male infertility, but the majority of cases remain idiopathic. Here we describe a novel method using single molecule Molecular Inversion Probes (smMIPs), to screen infertile men for mutations and copy number variations (CNVs) affecting known disease genes. We designed a set of 4,525 smMIPs targeting the coding regions

of causal (n=6) and candidate (n=101) male infertility genes. After extensive validation, we screened 1,112 idiopathic infertile men with non-obstructive azoospermia or severe oligozoospermia. In addition to five chromosome YCMs and six other sex chromosomal anomalies, we identified 5 patients with rare recessive mutations in *CFTR* as well as a patient with a rare heterozygous frameshift mutation in *SYCP3* that may be of clinical relevance. This results in a genetic diagnosis in 11-17 patients (1-1.5%), a yield that may increase significantly when more genes are confidently linked to male infertility. In conclusion, we developed a flexible and scalable method to reliably detect genetic causes of male infertility. The assay consolidates the detection of different types of genetic variation while increasing the diagnostic yield and detection precision at the same or lower price compared to currently used methods.

Key words: Male Infertility, smMIPs, Targeted Sequencing, Diagnostics, *CFTR*, CBAVD, Azoospermia, Oligozoospermia, Reproduction

Introduction

Infertility defines the failure as a couple to conceive within one year of unprotected sexual intercourse (Zegers-Hochschild, et al., 2009). Approximately 10-15% of all couples are affected by infertility and about half of the cases can be assigned to a male factor (Neto, et al., 2016).

Male infertility has large etiological heterogeneity and in many cases a strong genetic component is suspected. Severely oligozoospermic (<5 million sperm cells per ml) or azoospermic (no sperm in the ejaculate) patients in whom the medical history and physical

examination does not reveal the infertility cause, are suspected to be affected by a monogenic cause of disease (Neto, et al., 2016). Identification of the underlying genetic anomaly is essential to elucidate the cause of spermatogenic failure, provide adequate genetic counseling to the couple and decide upon the optimal assisted reproductive technology approach to be used. The most common known genetic anomalies found in men with secretory azoospermia are Klinefelter syndrome (KS; 47,XXY) and recurrent micro-deletions on the Y-chromosome in Azoospermia Factor (AZF) regions, respectively explaining up to 8% and 5% of men with a sperm concentration 0.1 to 5 million/ml. These contributions have been reported to increase to 15 and 10%, respectively, in more severe forms of spermatogenic failure, leading to non-obstructive azoospermia (Krausz, 2011).

Despite many decades of research in the field of male infertility genetics, only a few genes have been identified to be causal for male infertility when mutated (Lima and Lopes, 2014), namely *CFTR* (Chillon, et al., 1995), *DDX3Y* (Foresta, et al., 2000), *SYCP3* (Miyamoto, et al., 2003), *TEX11* (Yang, et al., 2015; Yatsenko, et al., 2015), *AURKC* (Dieterich, et al., 2007; Dieterich, et al., 2009) and *DPY19L2* (Elinati, et al., 2012; Kosciński, et al., 2011). These genes play crucial roles in the development of the vas deferentia (*CFTR*) (Claustres, 2005), the pre-meiotic germ cell formation (*DDX3Y*) (Ramathal, et al., 2015), synaptonemal complex formation during meiosis (*SYCP3* and *TEX11*) (Dobson, et al., 1994; Yang, et al., 2008), cytokinesis and chromosomal segregation during meiosis (*AURKC*) (Yan, et al., 2005) and acrosome formation in maturing sperm cells (*DPY19L2*) (Pierre, et al., 2012). Hundreds of other genes have been implicated in the biology of male reproduction, due to their locations in chromosomal regions that have been recurrently deleted, their identification in single or familial cases of infertility, genetic association studies or in animal studies (de Boer, et al., 2015; Matzuk and Lamb, 2008). Except for *CFTR*, none of these genes is routinely

tested in patients with male infertility, and therefore there is little insight into the extent of their contribution to this disorder.

Next-generation sequencing (NGS) methods allows for the routine analysis of multiple genes, all protein-coding sequence (exome) or even the entire genome. In recent years these methods have accelerated not only genetic research but also genetic diagnosis of many disorders (van Dijk, et al., 2014). Target gene panel sequencing approaches, using for example single molecule Molecular Inversion Probes (smMIPs), provide an attractive and affordable method for sequencing in clinical settings (Hardenbol, et al., 2003; Neveling, et al., 2017; Nilsson, et al., 1994; O'Roak, et al., 2012). In addition, these approaches are of great value in research settings to study candidate disease genes for genetically heterogeneous disorders by rapidly screening these genes in large cohorts of patients (Stessman, et al., 2014). The potential of this highly flexible and affordable method for variant detection in daily practice is underlined by its low per-sample cost in combination with an easily manageable, compliant and scalable protocol that permits the inclusion and exclusion of genes as desired and requires only minimal input of patient DNA (Hiatt, et al., 2013).

Currently used genetic tests for male infertility in a diagnostic setting rely on different technologies to detect either chromosomal abnormalities or gene mutations, making nowadays commonly used diagnostic tests complex, costly and labor intensive. We set out to develop an integrated approach to screen infertile patients for common genetic causes of infertility such as KS and *AZF* deletions, as well as for pathogenic mutations in genes implicated in male infertility. To do so, we developed a smMIP assay targeting the coding regions of 6 male infertility genes, and an additional set of 101 candidate male infertility genes which includes genes on the X and the Y chromosome to allow for detection of sex

chromosome anomalies. We hypothesized that this approach will provide an excellent alternative to currently used genetic diagnostic approaches in male infertility. In addition, it could provide a labor and cost-efficient assay for large scale genetic studies in male infertility research. In this paper, we validated the assay on positive control samples that had been diagnosed using currently used methods. In addition, we studied the diagnostic value of this assay in a cohort of 1,112 idiopathic male infertile patients.

Materials and methods

Literature analysis

We performed a literature search using Pubmed and OMIM for genes that either 1) are known to cause male infertility when mutated as shown by multiple cohort studies, or 2) are located in genomic regions that are deleted in infertile men (such as within the *AZF* regions), 3) genes that were identified in single/family cases and 4) genes that were candidate genes but failed to reach statistical significance due to small cohort sizes (Supp. Table S1). The search was performed using the search criteria ‘male infertility’, ‘infertile men’ or ‘spermatogenic failure’. We excluded genes in which only genetic risk factors have been found and genes in which variants have previously been reported to cause (severe) disorders unrelated to male infertility. In this way, we identified a set of six genes known to cause male infertility when mutated, and an additional set of 101 candidate male infertility genes. Ten of these genes are located on the X chromosome and 22 on the Y chromosome.

Patients

Validation cohort: we enrolled 26 anonymized samples with known genetic anomalies, including 21 patients who presented at the Radboudumc between 2005 and 2015 and were

tested in our routine diagnostics lab for (sex) chromosomal anomalies through karyotyping, *AZF* deletions through STS PCR analysis (Krausz, et al., 2014) or variants in *CFTR*, through mutation panel analysis (CFEU2v1 kit, Elucigene, Manchester, UK). Five patients were previously diagnosed with *DPY19L2* deletions or point mutations using breakpoint analysis and Sanger sequencing at Strasbourg University Hospital (Elinati, et al., 2012).

Application cohort: We enrolled a total of 1,119 patients with idiopathic male infertility that presented at Radboudumc (n=745) between July 2007 and July 2016 and at the Monash University (n=347) between May 1994 and November 2010. This study was approved by the local medical ethics committees (Radboudumc: CMO Regio Arnhem-Nijmegen (NL50495.091.14), Monash Surgical Day Hospital, Monash Medical Centre and Monash University) and written informed consent was obtained from all men. The men were diagnosed with either azoospermia (n=445), extreme oligozoospermia (<0.5 million sperm cells per ml; n=329), severe oligozoospermia (<5 million sperm cells per ml; n=337) or with mild to moderate oligozoospermia (<15 million sperm cells per ml; n=8). We excluded men with known causes of infertility, including a previous diagnosis of deletion of *AZF* (for all men tested after 1996), chromosomal anomalies, surgery on the vasa deferentia, orchitis, bilateral orchidectomy, varicoceles or prior radio- or chemotherapy. Men were also excluded if the diagnostic process identified retrograde ejaculation and/or proven Congenital Bilateral Absence of the Vas Deferens (CBAVD). For each man, a minimum of two semen analyses was performed according to WHO guidelines as part of standard patient care, and linked to each included man (WHO, 2010). All men were referred to one or more genetic screenings for chromosomal anomalies, *AZF* deletions and/or *CFTR* mutations to understand the underlying cause of infertility. We used residual genomic DNA (gDNA) extracted from a

venous blood sample that was obtained at the time of evaluation of the infertile men in our clinics.

smMIP design and sequencing

All smMIPs were designed using the MIPGEN pipeline (Boyle, et al., 2014). A total of 4,525 oligonucleotides were designed using the GRCh37/hg19 as the reference genome build to target the exomes and splice-site regions of our genes of interest. This sequence was supplemented with a 30 bp common linker arm containing a 5 bp random tag. This random single-molecule tag in the backbone of each MIP distinguished 1024 (4^5) unique molecules for each amplicon. These latter reduced potential PCR errors by removing PCR duplicates, and the resulting high-quality reads supported accurate quantification of mutation and copy-number ratios. Each smMIP covered a 112 bp genomic region with a maximum overlap of 40 bp with the adjacent smMIP. Combined, all smMIPs targeted approximately 412 kb of the human genome (Supp. Table S2).

The smMIP capture was performed according to the protocol previously described, with minor modifications (Hiatt, et al., 2013; O'Roak, et al., 2012). In short, our genomic regions of interest were captured in a reaction containing a molecular ratio between patient's gDNA and smMIP of 1:800. The smMIP capturing conditions were 10 minutes at 95°C which denatured the DNA, followed by an incubation period of 23 hours during which hybridization of the phosphorylated probes to the single stranded DNA strands occurred as well as the gap-filling and circulation of the probes via ligation. Next, all non-circular targets were digested via an exonuclease treatment. The circular targets were amplified with primers containing barcoded reverse primers via the following PCR conditions: 30 seconds at 98 °C followed by

17 cycles of 10 seconds at 98°C, 30 seconds at 60°C and 30 seconds at 72°C and 2 min at 72°C.

The primers used during this amplification, contained adapters compatible with Illumina sequencing (O'Roak, et al., 2012). After PCR, libraries were pooled each from 172 barcoded individual patient libraries. These combined libraries were then purified using Agencourt AMPure XP beads according to manufacturer's protocol (Beckman Coulter, Brea, CA). Each prepared library was diluted to a concentration of 1.7 pM and subsequently sequenced on a NextSeq500 platform (Illumina, San Diego, CA) according to the manufacturer's protocol (150 cycles High Output sequencing kit) resulting in 2x79 bp paired-end reads.

Rebalancing the smMIP pool

As described previously, sequence coverage can be highly variable in an initial MIP experiment (O'Roak, et al., 2012). The read depth variability was reduced by adding higher concentrations of the underperforming probes, whereas equimolar concentrations of unphosphorylated probes of overperforming smMIPs were added. The final concentrations of each smMIP are depicted in Supp. Table S2.

smMIP bioinformatics pipeline

Data was analyzed using an in-house smMIP-pipeline. First step of the pipeline was the bcl2fastq (v2.16) (<http://support.illumina.com/downloads.html>) conversion where raw sequence data was converted to two FASTQ files containing the forward and reverse reads from all samples. The 5 bp random tag was trimmed from the reads within the FASTQ files and kept in the key-identifiers of the reads for later use. Then the reads were aligned using BWA mem (v0.7.12) (Li and Durbin, 2009), directly followed by trimming of the arms and

overlap between the read-pairs. The latter occurred based on the smMIP design file, taking into account varying arm length with different insert sizes, as well as insertions and deletions compared to the reference. Finally, each read was written to a single BAM file per patient using the specified barcode.

For assessment of CNVs, reads from the same smMIP, e.g. containing the same random tag, were randomly discarded from the BAM file and the final coverages for each individual smMIP were written to a coverage report. The coverage report was then used for CNV calling per MIP target using CoNVaDing version 1.1.6 (Johansson, et al., 2016) using a minimum of 10 reference samples for comparison.

Single Nucleotide Variants (SNVs) were called using GATK (v3.4-46) (McKenna, et al., 2010) in two different ways. First the calling was performed on each individual BAM file using the HaplotypeCaller followed by the GATK joint genotyping step to retrieve a single VCF containing the SNVs of all samples. Secondly, the multi-sample calling from the UnifiedGenotyper was used to retrieve again a multi-sample VCF. The two VCFs were merged together and annotated using data from external databases like dbSNP (<https://www.ncbi.nlm.nih.gov/SNP/>) and OMIM (<https://www.omim.org/>) as well as in-house data such as allele-frequencies for these genomic loci in >15,000 individuals with disorders unrelated to male infertility as obtained by exome sequencing.

CNV analysis and validation

To detect recurrent chromosomal anomalies, such as sex-chromosomal aneuploidies and YCMs, the coverage of each smMIP region was normalized to the average coverage of all probes located on the autosomes in one patient. Then a fold-change in coverage over the

average coverage of the targeted region was calculated for each smMIP region, which was multiplied by the copy number of the region in the reference genome. Aneuploidy detection was obtained by calculating the average and standard deviation (SD) of the copy number estimate of all probes on each targeted chromosome. The predicted chromosome count, as well as the ratio for probes on the X and Y chromosome was used to predict common sex-chromosomal changes in infertile men (Supp. Table S3). A similar strategy was used to detect *AZF* duplications and deletions with the adjustment that copy number estimates were determined for each targeted gene, rather than the whole chromosome. The copy number of each gene was compared to reference values to detect (partial) *AZF* deletions (Noordam and Repping, 2006). *AZF* deletions were confirmed by STS PCR using sY142, sY1191, sY1197, sY1201, sY1206 and sY1291 as described earlier (Repping, et al., 2003).

Also, for the other genes the total number of CNV calls on probes that indicate a heterozygous deletion or homozygous deletion in a single gene was calculated. For quality purposes, samples containing more than 500 CNVs were excluded, as were CNVs containing less than 2 smMIPs. Multiplex ligation dependent probe amplification (MLPA) was done to confirm *CFTR* deletions with the P091-D1 kit (MRC-Holland, Amsterdam, the Netherlands) according to standard protocol and *DPY19L2* breakpoint PCR was done to confirm *DPY19L2* deletions as described earlier (Elinati, et al., 2012).

SNV prioritization and validation

We selected variants in the coding and splice site regions of *AURKC*, *CFTR*, *DDX3Y*, *DPY19L2*, *SYCP3* and *TEX11*. The variants were present in >15% of at least 10 reads. We selected rare variants in the population, i.e. allele frequency <1% in local, dbSNP (Sherry, et al., 2001), ExAC (Lek, et al., 2016), GoNL (Consortium, 2014), Welllderly (Erikson, et al.,

2016) and Kaviar (Glusman, et al., 2011) databases. Then we prioritized all non-synonymous and/or canonical splice site variants which we will refer to as candidate variants. Candidate variants were validated using conventional Sanger Sequencing according to standard protocols. Validated variants were classified according to the guidelines of the Association for Clinical Genetic Science and the Dutch Society of Clinical Genetic Laboratory Specialists using five classes: clearly not pathogenic (class 1), unlikely to be pathogenic (class 2), unknown significance (VUS, class 3), likely to be pathogenic (class 4) and clearly pathogenic (class 5) (Wallis, et al., 2013). For classification of the *CFTR* variants, we used the most recent literature and the Clinical and Functional TRanslation of CFTR (CFTR2) database; available at <http://cftr2.org> (last accessed April 2nd 2017). Variants that were known as CF-causing, were classified as clearly pathogenic, variants that are known as non-CF causing were classified as clearly not pathogenic, unless recent literature provided additional evidence for a different classification for CBAVD.

Results

Molecular inversion probe sequencing descriptive statistics

We successfully designed a single molecule Molecular Inversion Probe (smMIP) targeted enrichment approach using a grand total of 4,525 equimolarly pooled smMIPs capturing the exons and splice-site regions of 6 infertility genes and 101 genes located on 21 different chromosomes.

After rebalancing our smMIP pool to optimize the capture efficiency, we sequenced 98.4% of the targeted regions covering a total region of approximately 412 kb with a mean (median) read depth before duplicate removal of 413.7x (192x) and after duplicate removal of 178.1x

(119x) unique reads per sample. Of all regions, 93.1%, i.e. 383 kb, were covered at least by at least 10 unique reads (Supp. Figure S1). Despite some remaining capture efficiency variation between and within genes, we captured all 107 targeted genes (Supp. Figure S2A), including six genes that are known causes of male infertility when mutated (Supp. Figure S2B-G). The percentage of coding exons and canonical splice site regions that was captured by independent probes at least 10 times in each patient was: *AURKC* 100%; *CFTR* 98.6% (TG/T stretch intron 8: 100%); *DDX3Y* 95.6%; *DPY19L2* 89.8%; *SYCP3* 96.7% and *TEX11* 97.2%.

To ensure optimal quality of point mutations and CNV calling, the average coverage of smMIPs was calculated in each sample. Only samples with a unique average coverage of >30 fold calculated over targets located on the autosomes, were analyzed. Of all 1,145 DNAs tested, 1,118 (97.6%) were successfully processed at first sequencing effort. Of the 27 (2.4%) samples that had unsatisfactory coverage, 20 were successfully sequenced upon repeat, together yielding a high quality sequencing result for 1,138 (99.4%) samples tested.

Validation of the method

First, we evaluated the efficacy of our assay to detect chromosomal abnormalities, point mutations and CNVs by screening 26 patients with known genetic aberrations who were diagnosed using conventional (cyto)genetic methods (Table 1). This analysis was done in a blinded fashion, and only after the result of the smMIP assay was interpreted, did we compare the results to those of the conventional methods.

A gain of sequence coverage on X chromosomal genes was detected in males with KS (Figure 1A), equal to patterns observed in female subjects (data not shown). Using this coverage information for X and Y chromosomal genes, we were also able to reliably detect

other sex chromosomal anomalies including XYY syndrome, *SRY*+ 46, XX male and isochromosome for the short arm of Y (Figure 1B-D). Similarly, a complete loss of coverage on genes in the respective AZF regions was detected in patients with complete *AZFc* and *AZFb+c* deletions (Figure 2A-C). These data together indicated that (partial) X and Y chromosomal gains and losses can be reliably detected.

We then assessed the strength of our assay for the detection of point mutations (SNVs) by sequencing the gDNA of five patients with known genetic mutations in the *CFTR* gene. All of these *CFTR* mutations were identified correctly by our smMIP assay. In addition, we sequenced the gDNA of five patients previously diagnosed with *DPY19L2* deletions and/or point mutations. In these five patients, we detected two homozygous *DPY19L2* deletions and three heterozygous *DPY19L2* deletions, consistent with previous diagnoses (Figure 3A-E, Table 1). SNV analysis was done on the patients with a heterozygous deletion. Our assay called a hemizygous c.2038A>T (p.Lys680*) mutation in one of the patients. However, a hemizygous c.1183del (p.Ser395Leufs*9) was missed by our calling algorithms in two other patients.

Overall we were able to detect 100% of all chromosomal abnormalities as well as all known copy number variations. The ability to detect point mutations varied between the infertility genes, ranging from 33% in case of *DPY19L2* to 100% in case of *CFTR*.

Cohort screening

We assembled a cohort of 1,119 males affected by a severe form of idiopathic male infertility. Of these patients, 445 were classified as azoospermic, 674 as oligozoospermic. All

patients were screened with our smMIP assay and in 1,112 successfully sequenced samples we assessed whole chromosome anomalies, AZF deletions and variation in six diagnostic infertility genes: *AURKC*, *CFTR*, *DPY19L2*, *DDX3Y*, *SYCP3* and *TEX11*.

First we generated molecular karyotypes of each patient based on coverage of the targeted regions in our smMIP assay. In this manner we identified five patients with abnormalities involving the sex chromosomes; two patients with a 47,XXY karyotype, two with a 47,XYY karyotype and one patient who was disomic for chromosome X, lacking the majority of chromosome Y markers, except for five genes on the p11.31 and p11.2 bands (Table 2).

Next, we analyzed the targeted genetic regions of chromosome Y and identified four patients with a classic *AZFc* deletion and one patient with a classic *AZFb+c* deletion (Table 2). Given this successful approach to screen for AZF deletions, we aimed to expose partial AZF deletions that lead to a smaller deletion not completely eradicating all genes within these regions. Indeed, by estimating the copy number of AZF genes based on the sequencing depth of targeted regions, we identified subjects with *gr/gr* (n=31), *b2/b3* (n=5) and *b1/b3* (n=1) deletions (Supp. Figure S3A-C). In addition, we identified patients with duplications of the *AZFb+c* (n=1) or *AZFc* (n=28) regions (Supp. Figure S3D and E).

Next, we performed CNV analysis for the six known infertility genes. One patient with a heterozygous deletion of exon 19 and 20 of *CFTR* and three patients with heterozygous deletions of *DPY19L2* were revealed (Figure 4).

CNV analysis was followed by SNV analysis. After excluding protein-altering variants with an allele frequency >1% in the general population a total of 157 non-synonymous or splice

site point mutations were called in the 6 known infertility genes: *AURKC* (n=7), *CFTR* (n=128), *DDX3Y* (n=2), *DPY19L2* (n=9), *SYCP3* (n=5) and *TEX11* (n=6) (Figure 5, Supp. Table S4). We supplemented these rare variants with point mutations in *CFTR* with higher allele frequencies as they are known to cause obstructive azoospermia. This selection yielded patients carrying the c.1521_1523del (p.Phe508del) (n=32) mutation or a 5T (either c.1210-11T>G, also known as TG(11)T(5) or c.1210-6_1210-7del, also known as TG(12)T(5)) variant in the T/TG tract of intron 8 (n=58) (Supp. Table S4).

To assess the clinical relevance of all identified SNV and CNV variants, we classified all variants according to international guidelines on the evaluation of pathogenicity and clinical interpretation of sequence variants (Wallis, et al., 2013). Variant classification included the consultation of mutation databases HGMD® and ClinVar, and for *CFTR* variants the CFTR2 website. Of all 247 SNVs found in *AURKC*, *CFTR*, *DDX3Y*, *DPY19L2*, *SYCP3* and *TEX11*, 120 (48%) were classified as clearly not or unlikely to be pathogenic, 39 (16%) as variant of unknown significance (VUS) and 88 (36%) as likely or clearly pathogenic (Figure 5, Supp. Table S4).

Congenital Bilateral Absence of the Vas Deferens (CBAVD) is an autosomal recessive disorder caused by homozygous or compound heterozygous mutations in *CFTR* resulting in obstructive azoospermia (Cutting, 2015). Patients with azoospermia known to be caused by any type of obstruction, were excluded from our study. Ten patients harbored homozygous or compound heterozygous mutation in *CFTR*. Upon careful re-evaluation of the clinical patient information we could not fully exclude CBAVD in five of these ten patients (Table 2).

An earlier study showed that mutations in *SYCP3* cause autosomal dominant spermatogenic failure leading to azoospermia (Miyamoto, et al., 2003). We identified a likely pathogenic mutation in a patient with a severe oligozoospermia phenotype (Table 2).

Discussion

In this study we developed, validated and applied a novel smMIP assay for research and diagnostics in male infertility. Our findings demonstrate that targeted gene panel sequencing using the smMIP technology is a reliable method to screen for the different types of genomic abnormalities involved in male infertility, ranging from point mutations to chromosomal abnormalities. Our validation of the assay showed that common causes of male infertility such as *KS*, *AZFb+c* or *AZFc* deletions and *CFTR* point mutations were all diagnosed with similar or improved precision when compared to conventional methods used in our diagnostic workflows.

In order to demonstrate the diagnostic power of our assay, a large cohort of idiopathic male infertility cases was screened. In this cohort of 1,112 patients, we identified six patients with anomalies in the number of sex chromosomes and five patients with an *AZF* deletion. Of note, almost all of these 1,112 patients were pre-screened for chromosomal abnormalities and *AZF* deletions prior to inclusion in this study. Obviously, the diagnostics yield would have been significantly higher if this smMIP assay is applied to patients that have not yet undergone any genetic test. In addition, the diagnostic yield will increase when more genes are confidently linked to male infertility.

Evaluation of the genetic variation in six known male infertility genes uncovered a total of 247 point mutations and four rare CNVs in 229 patients. After combining the clinically relevant point mutations and CNVs in these genes and comparing this to the clinical phenotype observed in the patients involved, we made a likely diagnosis in one additional patient with a *SYCP3* mutation and identified five patients that should be referred to further genetic testing for segregation analysis of their *CFTR* point mutations and re-evaluation of CBAVD. In total, after careful and conservative diagnostic evaluation (excluding the patients with *CFTR* or *SYCP3* mutations), a definitive genetic diagnosis which explains the male infertility was made in 11 of the 1,112 patients.

It is noteworthy that other publications have found a higher diagnostic rate for some of the genes studied here (Miyamoto, et al., 2003; Yang, et al., 2015; Yatsenko, et al., 2015). While this may be explained in part by differences in patient inclusion criteria used, we want to stress the importance of using well-established diagnostic guidelines for evaluating the pathogenicity and clinical interpretation of sequence variants (Wallis, et al., 2013).

Our unbiased screening approach revealed some unexpected genotype-phenotype correlations and allowed us to make diagnoses in genes which were previously investigated. As a first example, a heterozygous frameshift mutation in exon 8 of *SYCP3* escaping nonsense-mediated RNA decay and leading to a truncated, but stable protein was previously described in two unrelated patients with azoospermia as a result of meiotic arrest (Miyamoto, et al., 2003). It had since then been assumed that such *SYCP3* mutations result in a very homogeneous phenotype, but no other patients with azoospermia and mutations in *SYCP3* could be identified in other studies (Gurkan, et al., 2013; Martinez, et al., 2007; Stouffs, et al., 2005). Here, we identified a

new heterozygous frameshift mutation in exon 7 of *SYCP3* that will likely result in nonsense-mediated RNA decay in a patient with oligozoospermia in combination with teratozoospermia. Although it is still possible that other, yet unknown, genetic factors may influence the phenotype of this patient, our finding suggests that the phenotypic spectrum resulting from *SYCP3* mutations may be broader than was initially presumed. Functional studies are needed to understand the role of this *SYCP3* mutation in infertility.

Other examples of unexpected findings were obtained for the *CFTR* gene. In most fertility clinics, *CFTR* mutation screening is only performed in patients with a presumed diagnosis of Congenital Bilateral/Unilateral Absence of the Vas Deferens (CBAVD/CUAVD). Although we excluded patients with a known diagnosis, we identified five patients with point mutations that may support a suspicion of CBAVD. All five patients presented with azoospermia. In two out of five males the vasa deferentia were difficult to palpate and in one patient the clinical information about CBAVD was discordant. In one patient only one of the vasa deferentia was identified suggesting Congenital Unilateral Absence of the Vas Deferens (CUAVD) and in one patient sperm cells were found by percutaneous epididymal sperm aspiration (PESA). Thus, in these five cases we cannot exclude a CBAVD and it is possible that these patients have obstructive azoospermia due to their *CFTR* mutations. Further segregation analysis of the *CFTR* mutations in the family is ongoing to confirm the molecular diagnosis.

Of clinical relevance are Patient624 and Patient1091 who were previously screened by a mutation panel of *CFTR* mutations frequent within populations of European origin, but in whom no mutations were reported. In each of the same patients, our assay detected two point mutations that are uncommon in the Dutch population. Although others have described targeted NGS methods for full *CFTR* gene analysis (Abou Tayoun, et al., 2013; Bonini, et al., 2015; Trujillano,

et al., 2015), most clinics still apply traditional mutation screening to detect only the most common mutations across local populations. Because of this, screening for *CFTR* mutations in CBAVD patients regularly leads to the identification of only one affected allele, because the second change is often a rare mutation or a genomic rearrangement affecting the *CFTR* gene (Giuliani, et al., 2010). Screening for such mutations depends on different detection strategies and is not always performed in these patients. Our findings demonstrate that the currently most used selection protocols for *CFTR* screening as well as commonly used screening methods miss diagnoses. It is of great importance to identify pathogenic mutations in the *CFTR* gene to predict the risk of a couple to deliver a child with Cystic Fibrosis. Our unbiased assay detects all point mutations throughout the coding region of *CFTR*, as well as CNVs affecting all or part of the *CFTR* gene.

One of the limitations of our targeted smMIP assay is the complication of variant detection in highly homologous regions of the genome. This is exemplified by the difficulty of detecting point mutations in *DPY19L2* as a result of strong homology to pseudogenes. This limitation is inherent to all NGS approaches using short read sequencing, including standard gene panel sequencing and Whole Exome Sequencing (WES) (Chan, et al., 2015; Eggers, et al., 2016). Although smMIPs may not be the most favorable method for sequencing of such regions, careful design of capturing probes in some cases can improve coverage (Nuttle, et al., 2013). As an example of this, our smMIP assay revealed a harmful c.2038A>T (p.Lys680*) *DPY19L2* mutation in a patient in whom WES did not reveal any *DPY19L2* mutations (unpublished data).

DNA sequencing using the smMIP platform has several advantages over other methods. First, the platform allows for conveniently adding or removing probes from the pool which permits

optimal capture efficiency and flexibility of the assay. This flexibility can also be used to update the pool with new probes at any time according to recent scientific insights on genes causative for male infertility, for example the new CBAVD gene *ADGRG2* (Patat, et al., 2016). Secondly, the smMIP assay work flow is straightforward and can be automated, making it an attractive method for diagnostic screening. The diagnostic application of the smMIPs has already been described for other genetic disorders (Hallam, et al., 2014; Neveling, et al., 2017). The smMIP platform does not only work on fresh DNA material, but we have obtained high quality data from gDNA samples dating back more than 20 years and gDNA derived from saliva samples. Also, other archived material such as formalin-fixed, paraffin-embedded (FFPE) tissue samples can be used for sequencing with this technique (Eijkelenboom, et al., 2016; Weren, et al., 2017).

The costs of our smMIP assay are relatively low compared to a combination of currently used genetic techniques used for male infertility diagnostics. Although smMIPs require a relatively high initial investment, the per-patient library preparation and sequencing costs are low. Currently the costs lie well below 1 dollar per gene per sample tested (O'Roak, et al., 2012). To date, genetic testing in male infertility is done in a targeted approach to avoid unnecessary investigation (Tournaye, et al., 2016). Currently karyotyping is offered to infertile men because chromosomal abnormalities have been reported in approximately 25% of the azoospermia cases and 13% of the severe oligozoospermia cases (Krausz, 2011; Neto, et al., 2016). Our method will miss balanced translocations and inversions accounting for approximately 0.9% of the infertile men (Neto, et al., 2016), and therefore should be combined with karyotyping, as is also recommended in a recent WHO guideline (Barratt, et al., 2017). Our method does, however, reliably detect aneuploidies and submicroscopic rearrangements affecting genes included in the assay. The ease and low costs may allow for

a change in strategy towards screening all males with severe oligozoospermia or azoospermia for common and rare genetic male infertility causes, irrespective of the clinical phenotype.

In conclusion, we describe a novel labor and cost-efficient method to screen for genetic causes of male infertility. We have shown that we can detect common, rare and novel mutations in six infertility genes, while also detecting CNVs and sex chromosomal anomalies. The smMIP assay has proved to be a powerful addition to current diagnostic screening techniques and both patients and doctors can benefit greatly from smMIP-based targeted sequencing approaches.

Acknowledgements

We are grateful for the participation of all patients in this study. The authors would like to thank Stefan Leliefeld, Marc Pieterse, Tuula Rinne, the Genome Technology Centre (Department of Human Genetics, Radboudumc, The Netherlands) and Duangporn Jamsai (Department of Anatomy and Developmental Biology, Monash University, Australia) for technical support.

Table 1: Validation cohort. Overview of all DNA samples enrolled in our validation cohort.

The samples were previously tested by our genetic diagnostics department using routine methods for different indications. All samples were tested in a blinded fashion and the original results were only disclosed after the smMIP results were in.

Table 2: Diagnostic yield in a cohort of 1,112 male infertility patients. Overview of all patients with a genetic anomaly that may explain an infertility phenotype. Classification and additional information about all point mutations can be found in Supp. Table S4. *CBAVD diagnosis discordant. ** Earlier *CFTR* testing did not reveal mutations. ***No information about presence of vas deferentia, patient successfully treated with Percutaneous Epididymal Sperm Aspiration (PESA).

Figure 1: Sex chromosomal anomalies in a validation cohort. The colored bars represent the average estimated copy number of each gene on chromosome X and Y which was calculated from the coverage data of each smMIP in one patient. The most left bar in each graph represents the estimated copy number of all targeted regions on the respective chromosome. The right bars represent the estimated copy number per gene. The whiskers depict the standard deviation of the estimations within the gene. The black lines indicate the expected copy number of each gene in a normal 46,XY male. A: Patient with Klinefelter syndrome (47,XXY) B: Patient with 47,XYY, C: Patient with 46,XX male syndrome positive for SRY. D: Patient with Isochromosome Yp.

Figure 2: AZF deletions in a validation cohort. The colored bars with whiskers represent respectively the average and SD estimated copy number of each gene on chromosome Y. The black lines indicate the expected copy number of each gene in a normal 46,XY male. The p and q arms of chromosome Y are depicted, as well as the AZF regions to indicate the locations of the genes. A: Classic AZFc deletion. B: Classic AZFb+c deletion. C: Atypical AZFc deletion deleting all PRY copies

Figure 3: *DPY19L2* deletions in a validation cohort. Plots of the calculated ratio of each smMIP (black dot) as calculated by CoNVaDing in the *DPY19L2* gene. The targeted exons are indicated by vertical blocks on the *DPY19L2* gene (horizontal black line). The normalized sequencing depth of the sample over the control gave ratio scores which are plotted for each smMIP. The ratios of a person without a *DPY19L2* deletion or duplication are approximately 1. Ratios above or below the blue and red lines indicate duplication and deletion events, respectively. In all five patients a deletion of *DPY19L2* was detected. A and E: Homozygous deletion. B-D: Heterozygous deletions. Residual coverage in some regions is seen even in homozygous deletions.

Figure 4: CNVs in known infertility genes in an idiopathic male infertility cohort. The ratio score for each smMIP was calculated by CoNVaDING from the expected read depth based on all samples in the cohort. Four heterozygous deletions were called. A: Heterozygous deletion of at least 4 kb affecting exon 19 and 20 in *CFTR*. B-D: heterozygous deletion of at least 108 kb affecting the full *DPY19L2* gene. The normal and deleted regions are indicated in respectively green and red.

Figure 5: Point mutations and CNVs in six known infertility genes in an idiopathic male infertility cohort. All variants were classified using the guidelines provided by Association for Clinical Genetic Science and the Dutch Society of Clinical Genetic Laboratory Specialists (Wallis, et al., 2013) and more details about the point mutations are available in Supp. Table S4. The color of the pinheads reflects the clinical classification from class 1 (green) to class 5 (red). The red dashed line represents all CNVs detected in the cohort. *AURKC* (A), *CFTR* (B), *DDX3Y* (C), *DPY19L2* (D), *SYCP3* (E), *TEX11* (F). Yellow: catalytic kinase domain. Brown: transmembrane domain. Pink: nucleotide binding domain. Dark green: Regulatory domain. Orange: Low complexity region (LCR). Purple: DEXDc domain. Light green: HELICc domain. Dark blue: Dpy-19 domain. Teal: coiled-coil domains. Red: meiosis-specific sporulation domain (SPO22). Protein domains from SMART (Letunic, et al., 2015).

Supp. Figure S1: Overall unique sequencing depth. Average cumulative read depth of unique smMIP molecules in all targeted regions of 1,112 patients and 26 control samples. The grey area depicts the standard deviation.

Supp. Figure S2: Sequencing depth in targeted genes. A: The median unique sequencing depth (horizontal black line) of each targeted gene is represented by Tukey box plots. The upper and lower hinges tally the 25th and 75th percentiles around the median. Outliers outside the interquartile range (whiskers) are represented as points. The plots are based on the unique read count per smMIP in 1138 patients. B-G: The unique sequencing depth on base-pair level of 6 known infertility genes: *AURKC*, *CFTR*, *DPY19L2*, *DDX3Y*, *SYCP3* and *TEX11*. The average depth (orange line) and standard deviation (grey area) were calculated from data of 1138 patients.

Supp. Figure S3: Partial AZF deletions and duplications. A: gr/gr deletion. B: b2/b3 deletion. C: b1/b3 deletion. D: AZFc duplication. E: AZFb+c duplication.

Supp. Table S1: All targeted genes.

Supp. Table S2: smMIP design.

Supp. Table S3: Requirements for the detection of sex chromosomal anomalies. The copy number estimate of each patient was compared to the ranges of expected ratios.

Supp. Table S4: All rare point mutations in the idiopathic male infertility cohort (n=1,112).

References

- Abou Tayoun AN, Tunkey CD, Pugh TJ, Ross T, Shah M, Lee CC, Harkins TT, Wells WA, Tafe LJ, Amos CI, Tsongalis GJ. 2013. A comprehensive assay for CFTR mutational analysis using next-generation sequencing. *Clin Chem* 59(10):1481-8.
- Barratt CLR, Björndahl L, De Jonge CJ, Lamb DJ, Osorio Martini F, McLachlan R, Oates RD, van der Poel S, St John B, Sigman M, Sokol R, Tournaye H. 2017. The diagnosis of male infertility: an analysis of the evidence to support the development of global WHO guidance—challenges and future research opportunities. *Hum Reprod Update* 1(21).
- Bonini J, Varilh J, Raynal C, Theze C, Beyne E, Audrezet MP, Ferec C, Bienvenu T, Girodon E, Tuffery-Giraud S, Des Georges M, Claustres M, Taulan-Cadars M. 2015. Small-scale high-throughput sequencing-based identification of new therapeutic tools in cystic fibrosis. *Genet Med* 17(10):796-806.
- Boyle EA, O'Roak BJ, Martin BK, Kumar A, Shendure J. 2014. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics* 30(18):2670-2.
- Chan LF, Campbell DC, Novoselova TV, Clark AJ, Metherell LA. 2015. Whole-Exome Sequencing in the Differential Diagnosis of Primary Adrenal Insufficiency in Children. *Front Endocrinol (Lausanne)* 6:113.
- Chillon M, Casals T, Mercier B, Bassas L, Lissens W, Silber S, Romey MC, Ruiz-Romero J, Verlingue C, Claustres M, et al. 1995. Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. *N Engl J Med* 332(22):1475-80.
- Claustres M. 2005. Molecular pathology of the CFTR locus in male infertility. *Reprod Biomed Online* 10(1):14-41.
- Consortium GotN. 2014. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 46(8):818-25.

- Cutting GR. 2015. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nat Rev Genet* 16(1):45-56.
- de Boer P, de Vries M, Ramos L. 2015. A mutation study of sperm head shape and motility in the mouse: lessons for the clinic. *Andrology* 3(2):174-202.
- Dieterich K, Soto Rifo R, Faure AK, Hennebicq S, Ben Amar B, Zahi M, Perrin J, Martinez D, Sele B, Jouk PS, Ohlmann T, Rousseaux S, Lunardi J, Ray PF. 2007. Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility. *Nat Genet* 39(5):661-5.
- Dieterich K, Zouari R, Harbuz R, Vialard F, Martinez D, Bellayou H, Prisant N, Zoghmar A, Guichaoua MR, Koscinski I, Kharouf M, Noruzinia M, Nadifi S, Sefiani A, Lornage J, Zahi M, Viville S, Sele B, Jouk PS, Jacob MC, Escalier D, Nikas Y, Hennebicq S, Lunardi J, Ray PF. 2009. The Aurora Kinase C c.144delC mutation causes meiosis I arrest in men and is frequent in the North African population. *Hum Mol Genet* 18(7):1301-9.
- Dobson MJ, Pearlman RE, Karaiskakis A, Spyropoulos B, Moens PB. 1994. Synaptonemal complex proteins: occurrence, epitope mapping and chromosome disjunction. *J Cell Sci* 107 (Pt 10):2749-60.
- Eggers S, Sadedin S, van den Bergen JA, Robevska G, Ohnesorg T, Hewitt J, Lambeth L, Bouty A, Knarston IM, Tan TY, Cameron F, Werther G, Hutson J, O'Connell M, Grover SR, Heloury Y, Zacharin M, Bergman P, Kimber C, Brown J, Webb N, Hunter MF, Srinivasan S, Titmuss A, Verge CF, Mowat D, Smith G, Smith J, Ewans L, Shalhoub C, Crock P, Cowell C, Leong GM, Ono M, Lafferty AR, Huynh T, Visser U, Choong CS, McKenzie F, Pachter N, Thompson EM, Couper J, Baxendale A, Gecz J, Wheeler BJ, Jefferies C, MacKenzie K, Hofman P, Carter P, King RI, Krausz C, van Ravenswaaij-Arts CM, Looijenga L, Drop S, Riedl S, Cools M, Dawson A, Juniarto AZ, Khadilkar V, Khadilkar A, Bhatia V, Dung VC, Atta I, Raza J, Thi Diem Chi N, Hao TK, Harley V, Koopman P, Warne G, Faradz S, Oshlack A, Ayers KL, Sinclair AH. 2016. Disorders of sex development: insights from targeted gene sequencing of a large international patient cohort. *Genome Biol* 17(1):243.
- Eijkelenboom A, Kamping E, Kastner-van Raaij A, Hendriks-Cornelissen S, Neveling K, Kuiper R, Hoischen A, Nelen M, Ligtenberg M, Tops B. 2016. Reliable Next-Generation Sequencing of Formalin-Fixed, Paraffin-Embedded Tissue Using Single Molecule Tags. *J Mol Diagn*.
- Elinati E, Kuentz P, Redin C, Jaber S, Vanden Meerschaut F, Makarian J, Koscinski I, Nasr-Esfahani MH, Demirel A, Gurgan T, Louanjli N, Iqbal N, Bisharah M, Pigeon FC, Gourabi H, De Briel D, Brugnion F, Gitlin SA, Grillo JM, Ghaedi K, Deemeh MR, Tanhaei S, Modarres P, Heindryckx B, Benkhalifa M, Nikiforaki D, Oehninger SC, De Sutter P, Muller J, Viville S. 2012. Globozoospermia is mainly due to DPY19L2 deletion via non-allelic homologous recombination involving two recombination hotspots. *Hum Mol Genet* 21(16):3695-702.
- Erikson GA, Bodian DL, Rueda M, Molparia B, Scott ER, Scott-Van Zeeland AA, Topol SE, Wineinger NE, Niederhuber JE, Topol EJ, Torkamani A. 2016. Whole-Genome Sequencing of a Healthy Aging Cohort. *Cell* 165(4):1002-11.
- Foresta C, Ferlin A, Moro E. 2000. Deletion and expression analysis of AZFa genes on the human Y chromosome revealed a major role for DBY in male infertility. *Hum Mol Genet* 9(8):1161-9.
- Giuliani R, Antonucci I, Torrente I, Grammatico P, Palka G, Stuppia L. 2010. Identification of the second CFTR mutation in patients with congenital bilateral absence of vas deferens undergoing ART protocols. *Asian J Androl* 12(6):819-26.
- Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC. 2011. Kaviar: an accessible system for testing SNV novelty. *Bioinformatics* 27(22):3216-7.
- Gurkan H, Aydin F, Kadioglu A, Palanduz S. 2013. Investigation of mutations in the synaptonemal complex protein 3 (SYCP3) gene among azoospermic infertile male patients in the Turkish population. *Andrologia* 45(2):92-100.

- Hallam S, Nelson H, Greger V, Perreault-Micale C, Davie J, Faulkner N, Neitzel D, Casey K, Umbarger MA, Chennagiri N, Kramer AC, Porreca GJ, Kennedy CJ. 2014. Validation for clinical use of, and initial clinical experience with, a novel approach to population-based carrier screening using high-throughput, next-generation DNA sequencing. *J Mol Diagn* 16(2):180-9.
- Hardenbol P, Baner J, Jain M, Nilsson M, Namsaraev EA, Karlin-Neumann GA, Fakhrai-Rad H, Ronaghi M, Willis TD, Landegren U, Davis RW. 2003. Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat Biotechnol* 21(6):673-8.
- Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. 2013. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res* 23(5):843-54.
- Johansson LE, van Dijk F, de Boer EN, van Dijk-Bos KK, Jongbloed JD, van der Hout AH, Westers H, Sinke RJ, Swertz MA, Sijmons RH, Sikkema-Raddatz B. 2016. CoNVaDING: Single Exon Variation Detection in Targeted NGS Data. *Hum Mutat* 37(5):457-64.
- Koscinski I, Elinati E, Fossard C, Redin C, Muller J, Velez de la Calle J, Schmitt F, Ben Khelifa M, Ray PF, Kilani Z, Barratt CL, Viville S. 2011. DPY19L2 deletion as a major cause of globozoospermia. *Am J Hum Genet* 88(3):344-50.
- Krausz C. 2011. Male infertility: pathogenesis and clinical diagnosis. *Best Pract Res Clin Endocrinol Metab* 25(2):271-85.
- Krausz C, Hoefsloot L, Simoni M, Tuttelmann F. 2014. EAA/EMQN best practice guidelines for molecular diagnosis of Y-chromosomal microdeletions: state-of-the-art 2013. *Andrology* 2(1):5-19.
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536(7616):285-91.
- Letunic I, Doerks T, Bork P. 2015. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res* 43(Database issue):D257-60.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-60.
- Lima AC, Lopes AM. 2014. Autosomal Mutations and Spermatogenic Failure. eLS: John Wiley & Sons, Ltd.
- Martinez J, Bonache S, Carvajal A, Bassas L, Larriba S. 2007. Mutations of SYCP3 are rare in infertile Spanish men with meiotic arrest. *Fertil Steril* 88(4):988-9.
- Matzuk MM, Lamb DJ. 2008. The biology of infertility: research advances and clinical challenges. *Nat Med* 14(11):1197-213.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297-303.
- Miyamoto T, Hasuike S, Yogev L, Maduro MR, Ishikawa M, Westphal H, Lamb DJ. 2003. Azoospermia in patients heterozygous for a mutation in SYCP3. *Lancet* 362(9397):1714-9.
- Neto FT, Bach PV, Najari BB, Li PS, Goldstein M. 2016. Genetics of Male Infertility. *Curr Urol Rep* 17(10):70.

- Neveling K, Mensenkamp AR, Derks R, Kwint M, Ouchene H, Steehouwer M, van Lier B, Bosgoed E, Rikken A, Tychon M, Zafeiropoulou D, Castelein S, Hehir-Kwa J, Tjwan Thung D, Hofste T, Lelieveld SH, Bertens SM, Adan IB, Eijkelenboom A, Tops BB, Yntema H, Stokowy T, Knappskog PM, Hoberg-Vetti H, Steen VM, Boyle E, Martin B, Ligtenberg MJ, Shendure J, Nelen MR, Hoischen A. 2017. BRCA Testing by Single-Molecule Molecular Inversion Probes. *Clin Chem* 63(2):503-512.
- Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U. 1994. Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* 265(5181):2085-8.
- Noordam MJ, Repping S. 2006. The human Y chromosome: a masculine chromosome. *Curr Opin Genet Dev* 16(3):225-32.
- Nuttall X, Huddleston J, O'Roak BJ, Antonacci F, Fichera M, Romano C, Shendure J, Eichler EE. 2013. Rapid and accurate large-scale genotyping of duplicated genes and discovery of interlocus gene conversions. *Nat Methods* 10(9):903-9.
- O'Roak BJ, Vives L, Fu W, Egerton JD, Stanaway IB, Phelps IG, Carvill G, Kumar A, Lee C, Ankenman K, Munson J, Hiatt JB, Turner EH, Levy R, O'Day DR, Krumm N, Coe BP, Martin BK, Borenstein E, Nickerson DA, Mefford HC, Doherty D, Akey JM, Bernier R, Eichler EE, Shendure J. 2012. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338(6114):1619-22.
- Patat O, Pagin A, Siegfried A, Mitchell V, Chassaing N, Faguer S, Monteil L, Gaston V, Bujan L, Courtade-Saidi M, Marcelli F, Lalau G, Rigot JM, Mieusset R, Bieth E. 2016. Truncating Mutations in the Adhesion G Protein-Coupled Receptor G2 Gene ADGRG2 Cause an X-Linked Congenital Bilateral Absence of Vas Deferens. *Am J Hum Genet* 99(2):437-42.
- Pierre V, Martinez G, Coutton C, Delaroche J, Yassine S, Novella C, Pernet-Gallay K, Hennebicq S, Ray PF, Arnoult C. 2012. Absence of Dpy19l2, a new inner nuclear membrane protein, causes globozoospermia in mice by preventing the anchoring of the acrosome to the nucleus. *Development* 139(16):2955-65.
- Ramathal C, Angulo B, Sukhwani M, Cui J, Durruthy-Durruthy J, Fang F, Schanes P, Turek PJ, Orwig KE, Reijo Pera R. 2015. DDX3Y gene rescue of a Y chromosome AZFa deletion restores germ cell formation and transcriptional programs. *Sci Rep* 5:15041.
- Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, de Vries JW, Oates RD, Silber S, van der Veen F, Page DC, Rozen S. 2003. Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat Genet* 35(3):247-51.
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29(1):308-11.
- Stessman HA, Bernier R, Eichler EE. 2014. A genotype-first approach to defining the subtypes of a complex disease. *Cell* 156(5):872-7.
- Stouffs K, Lissens W, Tournaye H, Van Steirteghem A, Liebaers I. 2005. SYCP3 mutations are uncommon in patients with azoospermia. *Fertil Steril* 84(4):1019-20.
- Tournaye H, Krausz C, Oates RD. 2016. Concepts in diagnosis and therapy for male reproductive impairment. *Lancet Diabetes Endocrinol*.
- Trujillano D, Weiss ME, Koster J, Papachristos EB, Werber M, Kandaswamy KK, Marais A, Eichler S, Creed J, Baysal E, Jaber IY, Mehaney DA, Farra C, Rolfs A. 2015. Validation of a semiconductor next-generation sequencing assay for the clinical genetic screening of CFTR. *Mol Genet Genomic Med* 3(5):396-403.
- van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. *Trends Genet* 30(9):418-26.
- Wallis Y, Payne S, McAnulty C, Bodmer D, Sistermans E, Robertson K, Moore D, Abbs S, Deans Z, Devereau A. 2013. Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics.

- Weren RD, Mensenkamp AR, Simons M, Eijkelenboom A, Sie AS, Ouchene H, van Asseldonk M, Gomez-Garcia EB, Blok MJ, de Hullu JA, Nelen MR, Hoischen A, Bulten J, Tops BB, Hoogerbrugge N, Ligtenberg MJ. 2017. Novel BRCA1 and BRCA2 Tumor Test as Basis for Treatment Decisions and Referral for Genetic Counselling of Patients with Ovarian Carcinomas. *Hum Mutat* 38(2):226-235.
- World Health Organisation 2010. WHO laboratory manual for the examination and processing of human semen. Geneva: World Health Organization.
- Yan X, Cao L, Li Q, Wu Y, Zhang H, Saiyin H, Liu X, Zhang X, Shi Q, Yu L. 2005. Aurora C is directly associated with Survivin and required for cytokinesis. *Genes Cells* 10(6):617-26.
- Yang F, Gell K, van der Heijden GW, Eckardt S, Leu NA, Page DC, Benavente R, Her C, Hoog C, McLaughlin KJ, Wang PJ. 2008. Meiotic failure in male mice lacking an X-linked factor. *Genes Dev* 22(5):682-91.
- Yang F, Silber S, Leu NA, Oates RD, Marszalek JD, Skaletsky H, Brown LG, Rozen S, Page DC, Wang PJ. 2015. TEX11 is mutated in infertile men with azoospermia and regulates genome-wide recombination rates in mouse. *EMBO Mol Med* 7(9):1198-210.
- Yatsenko AN, Georgiadis AP, Ropke A, Berman AJ, Jaffe T, Olszewska M, Westernstroer B, Sanfilippo J, Kurpisz M, Rajkovic A, Yatsenko SA, Kliesch S, Schlatt S, Tuttelmann F. 2015. X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. *N Engl J Med* 372(22):2097-107.
- Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, Sullivan E, Vanderpoel S. 2009. International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology, 2009. *Fertil Steril* 92(5):1520-4.

Table 1: Validation cohort. Overview of all DNA samples enrolled in our validation cohort.

The samples were previously tested by our genetic diagnostics department using routine methods for different indications. All samples were tested in a blinded fashion and the original results were only disclosed after the smMIP results were in.

Sample	Indication	Method Diagnostics	Result Diagnostics	Result smMIPs	Result
Validation sample1	Chromosomal anomaly	Karyotyping	47,XXY (Klinefelter)	Duplication of X	Consistent
Validation sample2	Chromosomal anomaly	Karyotyping	47,XXY (Klinefelter)	Duplication of X	Consistent
Validation sample3	Chromosomal anomaly	Karyotyping	47,XXY (Klinefelter)	Duplication of X	Consistent
Validation sample4	Chromosomal anomaly	Karyotyping	47,XXY (Klinefelter)	Duplication of X	Consistent
Validation sample5	Chromosomal anomaly	Karyotyping	47,XYY	Duplication of Y	Consistent
Validation sample6	Chromosomal anomaly	Karyotyping	46,XX SRY+ male	46,XX male (SRY, RPSY1, ZFY, TGIF2LY still present)	Improved precision
Validation sample7	Chromosomal anomaly	Karyotyping	46,X,i(Yp)	46,X,i(Yp)	Consistent
Validation sample8	AZF	STS PCR	AZFc	AZFc	Consistent
Validation sample9	AZF	STS PCR	AZFc	AZFc	Consistent
Validation sample10	AZF	STS PCR	AZFc	AZFc	Consistent
Validation sample11	AZF	STS PCR	AZFc	AZFc	Consistent
Validation sample12	AZF	STS PCR	AZFc	AZFc	Consistent
Validation sample13	AZF	STS PCR	AZFc	AZFc	Consistent
Validation sample14	AZF	STS PCR	P5/distal-P1 (AZFb+AZFc)	P5/distal-P1 (AZFb+AZFc)	Consistent
Validation sample15	AZF	STS PCR	P5/distal-P1 (AZFb+AZFc)	P5/distal-P1 (AZFb+AZFc)	Consistent
Validation sample16	AZF	STS PCR	AZFc	Atypical AZFc	Improved precision
Validation sample17	CFTR	Sanger	Homozygous	Homozygous	Consistent
Validation sample18	CFTR	Sanger	c.1521_1523del (p.Phe508del) Heterozygous	c.1521_1523del (p.Phe508del) Heterozygous	Consistent
Validation sample19	CFTR	Sanger	c.1521_1523del (p.Phe508del) c.3705T>G (p.Ser1235Arg) Heterozygous	c.1521_1523del (p.Phe508del) c.3705T>G (p.Ser1235Arg) Heterozygous	Consistent
Validation sample20	CFTR	Sanger	c.1624G>T (p.Gly542*) T5-TG12 (c.1210-6_1210-7del) Heterozygous	c.1624G>T (p.Gly542*) T5-TG12 (c.1210-6_1210-7del) Heterozygous	Consistent
Validation sample21	CFTR	Sanger	c.948del (p.Phe316Leufs*12) c.350G>A (p.Arg117His) Homozygous	c.948del (p.Phe316Leufs*12) c.350G>A (p.Arg117His) Homozygous	Consistent
Validation sample22	DPY19L2	PCR	c.579+1G>A Homozygous deletion	c.579+1G>A Homozygous deletion	Consistent
Validation sample23	DPY19L2	PCR/Sanger	Heterozygous DPY19L2 deletion c.1183delT (p.Ser395Leufs*9)	Heterozygous DPY19L2 deletion	Indel missed
Validation sample24	DPY19L2	PCR/Sanger	Heterozygous DPY19L2 deletion c.1183delT (p.Ser395Leufs*9)	Heterozygous DPY19L2 deletion	Indel missed
Validation sample25	DPY19L2	PCR/Sanger	Heterozygous DPY19L2 deletion c.2038A>T (p.Lys680*)	Heterozygous DPY19L2 deletion c.2038A>T (p.Lys680*)	Consistent
Validation sample26	DPY19L2	PCR	Homozygous deletion	Homozygous deletion	Consistent

Patient	Genetic diagnosis smMIPs	Phenotype	Clinical information	Sperm concentration (10 ⁶ per ml)	Diagnostic significance
Patient956	46, XX (Y p11.31 and p11.2 positive)	Azoospermia	-	0	Diagnosis
Patient699	47, XXY (Klinefelter)	Azoospermia	-	0	Diagnosis
Patient1047	47, XXY (Klinefelter)	Azoospermia	-	0	Diagnosis
Patient1078	47, XXY (Klinefelter)	Azoospermia	-	0	Diagnosis
Patient41	47, XYY	Severe oligoteratozoospermia	-	4-11	Risk factor
Patient263	47, XYY	Oligozoospermia	-	4-11	Risk factor
Patient938	AZFb+c deletion	Azoospermia	-	0	Diagnosis
Patient11	AZFc deletion	Severe oligoasthenoteratozoospermia	-	0.5-5	Diagnosis
Patient122	AZFc deletion	Severe oligozoospermia	-	0.5-5	Diagnosis
Patient161	AZFc deletion	Severe oligozoospermia	-	0.5-5	Diagnosis
Patient993	AZFc deletion	Azoospermia	-	0	Diagnosis
Patient616	CFTR p.Arg117Cys (het) p.Arg1158* (het)	Azoospermia	CBAVD not excluded*	0	Further investigation required
Patient624	CFTR p.Asp110His (het) p.Ser945Leu (het)	Azoospermia	CBAVD not excluded**	0	Further investigation required
Patient638	CFTR p.Phe508delPhe (het) p.Arg117His (het)	Azoospermia	No CBAVD suspected	0	Diagnosis not possible
Patient684	CFTR p.Phe508delPhe (het) p.Gly463Ser (het)	Azoospermia	CUAVD	0	Further investigation required
Patient800	CFTR p.Phe508delPhe (het) p.Cys76Trp (het)	Azoospermia	No CBAVD suspected	0	Diagnosis not possible
Patient884	CFTR	Azoospermia	No CBAVD suspected	0	Diagnosis not possible

Patient970	p.Phe508delPhe (het)	Azoospermia	No CBAVD suspected	0	Diagnosis not possible
	p.Arg117His (het)				
	CFTR				
Patient1035	p.Arg117His (het)	Azoospermia	PESA successful ***	0	Further investigation required
	5T-12TG (het)				
	CFTR				
Patient1086	p.Phe508delPhe (het)	Azoospermia	No CBAVD suspected	0	Diagnosis not possible
	p.Arg117His (het)				
	CFTR				
Patient1091	p.Arg117His (hom)	Azoospermia	CBAVD not excluded **	0	Further investigation required
	CFTR				
	p.Arg248Thr (het)				
Patient178	p.Phe508Cys	Severe oligoteratozoospermia	-	0.5-5	Further investigation required
	SYCP3				
	p.Ile175Asnfs*8 (het)				

Table 2: Diagnostic yield in a cohort of 1,112 male infertility patients. Overview of all patients with a genetic anomaly that may explain an infertility phenotype. Classification and additional information about all point mutations can be found in Supp. Table S4. *CBAVD diagnosis discordant. ** Earlier *CFTR* testing did not reveal mutations. ***No information about presence of vas deferentia, patient successfully treated with Percutaneous Epididymal Sperm Aspiration (PESA).

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Oud, MS;Ramos, L;O'Bryan, MK;McLachlan, RI;Okutman, O;Viville, S;de Vries, PF;Smeets, DFCM;Lugtenberg, D;Hehir-Kwa, JY;Gilissen, C;van de Vorst, M;Vissers, LELM;Hoischen, A;Meijerink, AM;Fleischer, K;Veltman, JA;Noordam, MJ

Title:

Validation and application of a novel integrated genetic screening method to a cohort of 1,112 men with idiopathic azoospermia or severe oligozoospermia

Date:

2017-11-01

Citation:

Oud, M. S., Ramos, L., O'Bryan, M. K., McLachlan, R. I., Okutman, O., Viville, S., de Vries, P. F., Smeets, D. F. C. M., Lugtenberg, D., Hehir-Kwa, J. Y., Gilissen, C., van de Vorst, M., Vissers, L. E. L. M., Hoischen, A., Meijerink, A. M., Fleischer, K., Veltman, J. A. & Noordam, M. J. (2017). Validation and application of a novel integrated genetic screening method to a cohort of 1,112 men with idiopathic azoospermia or severe oligozoospermia. HUMAN MUTATION, 38 (11), pp.1592-1605. <https://doi.org/10.1002/humu.23312>.

Persistent Link:

<http://hdl.handle.net/11343/293427>