# Laboratory validation of the Fertilome® genetic test

Northrop LE, Bhardwaj N, DeGrazia J, Burnett B, Parfitt DE, Yurttas Beim P

Celmatix Inc., New York, NY, USA

## INTRODUCTION

Due to recent advances in next-generation sequencing (NGS), Celmatix Clinical Laboratories (CCL) has the ability to offer the Fertilome® test—the first multigene test for variants associated with risk for female reproductive conditions (RCs). CCL continually performs rigorous validation testing before introducing any new version of the Fertilome test[1]. The Fertilome assay has recently undergone another round of validation, which now allows for additional genomic content in hard-to-sequence regions and the ability to sequence saliva as a second sample option for our clinics and patients.

## MATERIALS & METHODS

### Single-nucleotide variant classification for the inclusion in the Fertilome panel

The Fertilome test is a New York State Department of Health (NYS-DOH)-approved laboratory-developed test (LDT) performed on genomic DNA (gDNA) sequenced on a targeted NGS panel. The testing panel was constructed by using natural language processing algorithms to identify 22,709 articles related to genetics and 11 RCs: diminished ovarian reserve, early menopause, premature ovarian insufficiency (POI), endometriosis, hyperandrogenism, idiopathic infertility, ovarian hyperstimulation syndrome, polycystic ovary syndrome (PCOS), recurrent implantation failure (RIF), recurrent pregnancy loss (RPL), and predictors of in vitro fertilization treatment response. Manual curation identified false positives and negatives, yielding 4,758 articles. Nomenclature was then harmonized across papers and mapped to 1,826 gene-condition combinations. We applied the principles of the Clinical Genome (ClinGen) Gene-Disease Clinical Validity Classification Framework prior to its expansion by Strande et al.[2] to rank the evidence linking the genes with the RCs. Articles pertaining to single-nucleotide variants (SNVs) within genes with strong evidence were further analyzed. Per PRISMA guidelines, 2 independent reviewers recorded ≥137 data points per study and met to resolve conflicts. Random effects model meta-analyses were run to assess associations on a variant level. Variants were excluded if they were analyzed in <2 published studies or in overlapping cohorts or if the risk allele could not be determined. A p-value of <0.05 was considered statistically significant, equivalent to a false discovery rate cutoff of approximately 0.10. Version 1.5 of the Fertilome genetic test panel comprises SNVs that withstood this rigorous statistical validation analysis. These SNVs fall within genes with strong evidence linking them to 1 or more of 6 RCs (endometriosis, PCOS, POI, RPL, idiopathic infertility, and RIF), and the Fertilome test report categorizes them according to effect size and allele frequency.

### Wet-lab bench validation of Fertilome genetic test

CCL's clinical laboratories (NJ #33D2127647 and NYS #33D2104293, PFI#9067) have been validated to run the automated NGS-developed assay and to achieve accurate results. The wet-bench workflow is based on laboratory products from industry leaders such as Illumina, Agilent, and Life Technologies. To enable deep sequencing, the Fertilome test is designed as a targeted panel encompassing only the flanking regions surrounding the SNVs. Specifically, a proprietary targeted gene panel, as annotated by our classification system above, was designed with Illumina TrueSeq Custom Amplicon Library enrichment and sequenced using a MiSeq sequencer (paired-end, 150 bp). Several quality control steps are incorporated throughout the wet-bench chemistry, from gDNA purification to loading the pooled library onto the MiSeq, to ensure high-quality data. The bioinformatics pipeline is built on the best practices in the field, using tools such as BWA, Samtools, Picard, GATK, and several scripts developed in house. The bioinformatics process implements quality control checks at several points to ensure that the data and the results are of the highest quality. Several criteria for rejection or acceptance of the data based on the quality have also been established: initial number of reads >2.5 million per specimen; per base sequence quality >Q30; min read length after quality trimming >30 bp; average coverage >80X; quality score of the variants >Q20; and depth of coverage at variant positions >30X. In addition, validated Sanger sequencing is used as a secondary method in cases of insufficient coverage for accurate variant detection (when depth <30X) or as confirmation for rarely identified variants (i.e., <1% allele frequency). All SNV data are evaluated by a board-certified molecular geneticist.

Fertilome was validated based on NGS guidelines from several regulatory agencies, including the NYS-DOH's Clinical Laboratory Evaluation Program guidelines for NGS LDT validation and those of the American College of Medical Genetics (ACMG), the College of American Pathologists (CAP), and the Nex-StoCT workgroup for Standardization of Clinical Testing[3-7]. Appropriately consented research specimens were collected from several hundred individuals across a broad age spectrum, with Institutional Review Board (IRB) approval. Individuals included women demonstrating different phenotypes associated with infertility as well as healthy women (e.g., egg donors). The specimens underwent massively parallel whole genome sequencing and variant identification at an independent reference laboratory (HudsonAlpha Institute for Biotechnology, Huntsville, AL). In addition, a positive control (NA12878) was sequenced in every run, as recommended by the National Institute of Standards and Technology, as a wet-bench chemistry control[8].

## RESULTS

### Study 1. Blinded validation

The blinded validation study investigated the reportable variants across 35 purified gDNA research specimens. These specimens contained a broad and representative spectrum of the reportable variants that the LDT is designed to detect. Specimens were evaluated in 3 separate rounds of sequencing and associated bioinformatics analysis.

Some specimens were run in triplicate for reproducibility and repeatability assessment (see Study 2), increasing the total number of samples to 44. From all 3 sequencing runs, a total of 2,024 variants (46 variants × 44 samples tested) were analyzed to determine accuracy of the test. Based on the results, the performance of the LDT was evaluated in terms of analytical specificity, analytical sensitivity, positive predictive value (PPV), negative predictive value (NPV), F-measure, and overall accuracy of the test[7].

To implement a blinded analysis, an external and independent honest broker (HB) was used and all steps were performed in a blinded fashion; that is, the identity and expected results of the specimens used in the validation exercise were not known to CCL staff performing the test and interpreting the results. The HB performed specimen selection and coded the specimens' identities. Blinded/coded blood samples were then subjected to the laboratory's standard operating procedure for this LDT, that is, DNA purification and quantification, library preparation, and sequencing for the interrogation of reportable SNVs. The results from the protocol were then returned to the HB, who compared reference and experimental datasets for the reportable SNVs detected and issued an unbiased report of concordance between the datasets to the laboratory director. Analysis was then performed by the laboratory director to determine assay performance (Table 1). Both accuracy and F-measure were found to be 99.9% (Table 1).

**TABLE 1.** Analytic Performance Metrics During Validation Of Fertilome V1.5

| Analytic Metric | Value |
|---|---|
| Sensitivity | 100.0% (1,046/1,046) |
| Specificity | 99.8% (976/978) |
| PPV | 99.8% (1,046/1,048) |
| NPV | 100.0% (976/976) |
| F-Measure | 99.9% (2,022/2,024) |
| Accuracy | 99.9% (2,022/2,024) |

Number of variants in each category: true positive (TP) = 1,046; true negative (TN) = 976; false negative (FN) = 0; and false positive (FP) = 2. Numbers in parentheses indicate the number of variants for calculating each metric.

### Study 2. Technical precision: reproducibility and repeatability

Repeatability of the assay (intra-assay concordance) was assessed by running triplicates of a different randomly selected sample in each of the 3 rounds. Also, 3 samples were included in each of the runs for assessing reproducibility (inter-assay concordance). For repeatability measurements, only 1 variant out of 138 (= 46 × 3) reported a different genotype in 1 of the 3 samples, placing the repeatability at 99.3% (the variant was identified in all 3 samples but had a different genotype in 1 sample). A reproducibility of 100.0% was registered (all 138 variants were concordant).

### Study 3. Cross-laboratory validation

The Fertilome test also underwent rigorous validation in the second laboratory located in NJ. The same specimens that underwent blinded validation in the NY laboratory were tested in the NJ laboratory, following the identical protocol. A total of 28 samples were tested, across 3 separate sequencing runs, including precision reproducibility testing. From all 3 sequencing runs, a total of 2,024 variants were analyzed to determine accuracy of the test. Both accuracy and F-measure were calculated to be 99.9%.

The validation data from the 2 laboratories were combined, and analytical validity was demonstrated in a total of 4,048 (= 2,024 × 2) reportable variants, with an accuracy of 99.9% when run on 2 different MiSeq instruments across different days by several personnel (Table 2). Previously sequenced samples were 99.99% concordant between the 2 laboratories (only 1 variant out of 2,024 was discordant) across all reportable SNVs in all of the targeted genes supporting reproducibility of the test.

**TABLE 2.** Analytic Performance Metrics Across 2 Validation Studies

| Accuracy Measurement | Value |
|---|---|
| Accuracy | 99.9% (4,045/4,048) |
| Sensitivity | 100.0% (2,092/2,092) |
| Specificity | 99.8% (1,953/1,956) |
| PPV | 99.9% (2,092/2,095) |
| NPV | 100.0% (1,954/1,954) |
| Repeatability | 99.8% (2,023/2,024) |
| Reproducibility | 100.0% (2,024/2,024) |

Number of variants in each category across 2 validation studies run in 2 separate laboratories: TP = 2092, TN = 1953, FN = 0, and FP = 3. Numbers in parentheses indicate the number of variants for calculating each metric.

### Study 4. Validation with saliva as the reference material

Appropriately consented research specimens were collected from 23 individuals across a broad age spectrum with IRB approval. From each individual, 2 types of samples were collected in a blinded fashion: blood and saliva. These matched specimens were used to validate the accuracy of using saliva as the reference material (RM). In this validation study, a blinding scheme as in Study 1 was used and an HB was involved (see above for details).

From the set of specimens, a total of 5 unique sample pairs were subjected to the optimized NGS methods for concordance testing. Since they were matched samples of both blood and saliva, the total number of samples was 10. The final variant call files were compared between blood and saliva for concordant genotypes between the matched samples.

One of the 5 saliva samples was run in triplicate to test reproducibility and precision of the assay using saliva as the RM. There were a total of 12 samples (5 blood samples and 7 saliva samples).

The final variant calls generated were provided to the HB, who determined concordance between blood and saliva samples on the basis of the variants called and the genotypes identified for them. The performance metrics were identical to those used for Study 1. Accuracy and F-measure of 100.0% were obtained for this validation study using saliva as RM (Table 3).

TABLE 3. Analytical Performance Metrics For Saliva Validation

| Analytic Metric | Value |
|---|---|
| Sensitivity | 100.0% (280/280) |
| Specificity | 100.0% (308/308) |
| PPV | 100.0% (280/280) |
| NPV | 100.0% (308/308) |
| F-Measure | 100.0% (588/588) |
| Accuracy | 100.0% (588/588) |

Number of variants in each category: TP = 280, TN = 308, FN = 0, and FP = 0. Numbers in parentheses indicate the number of variants for calculating each metric.

To determine intra-assay reproducibility, a different specimen was run in triplicate in the validation run. The variant calls made for each of the triplicates were compared to quantify intra-assay reproducibility. Identical calls were obtained for the variant genotypes across all 3 replicates, placing the intra-assay reproducibility at 100.0%.

## CONCLUSION

CCL validated the Fertilome test following the guidelines established by NYS-DOH and CAP. For validation of blood as RM, both F-measure and accuracy were 99.9%. For saliva validation, all matching blood and saliva samples recorded 100.0% concordance in terms of variants identified and genotype reported. CCL is continually validating the test and confirming precision with bi-annual proficiency testing and technologist competency runs.

## REFERENCES


1. Validation results for previous versions of the Fertilome test are on file at CCL and available upon request.
2. Strande, N.T., et al., *Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource.* Am J Hum Genet, 2017. **100**(6): p. 895-906.
3. *Guidelines for Validation Submissions of Next Generation Sequencing (NGS) assays under the NYS Testing Category of Genetic Testing – Molecular.* 2015.
4. Rehm, H.L., et al., *ACMG clinical laboratory standards for next-generation sequencing.* Genet Med, 2013. **15**(9): p. 733-47.
5. Aziz, N., et al., *College of American Pathologists' laboratory standards for next-generation sequencing clinical tests.* Arch Pathol Lab Med, 2015. **139**(4): p. 481-93.
6. Gargis, A.S., et al., *Assuring the quality of next-generation sequencing in clinical laboratory practice.* Nat Biotechnol, 2012. **30**(11): p. 1033-6.
7. Halling, K.C., I. Schrijver, and D.L. Persons, *Test verification and validation for molecular diagnostic assays.* Arch Pathol Lab Med, 2012. **136**(1): p. 11-3.
8. Zook, J.M., et al., *Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls.* Nat Biotechnol, 2014. **32**(3): p. 246-51.