MammaPrint and BluePrint Molecular Diagnostics Using Targeted RNA Next-Generation Sequencing Technology

Lorenza Mittempergher,* Leonie J.M.J. Delahaye,* Anke Witteveen,* Jacob B. Spangler,† Fariet Hassenmahomed,* Sammy Mee, Soufiane Mahmoudi,* Jiang Chen,† Simon Bao, Mireille H.J. Snel,* Sandra Leidelmeijer,* Naomi Besseling,* Anne B. Lucas,† Carlos Pabón-Peña,† Sabine C. Linn,† Christa Dreezen,* Diederik Wekhamp,* Bob Y. Chan,† René Bernard,*q

Laura J. van’t Veer,† and Annuska M. Glas*

Q1 Q2 From the Agenda NV,* Amsterdam, the Netherlands; the Agenda Inc.,† Irvine, California; the Diagnostics and Genomics Group,† Agilent Technologies, Santa Clara, California; the Divisions of Molecular Pathology and Medical Oncology* and Molecular Carcinogenesis,† Netherlands Cancer Institute, Amsterdam, the Netherlands; and the Laboratory Medicine,† University of California, San Francisco, San Francisco California

Accepted for publication April 16, 2019.
Address correspondence to Annuska M. Glas, Ph.D., Research and Development and Product Support, Agenda NV, Science Park 406, 1098 XH Amsterdam, the Netherlands. E-mail: annuska.glas@agenda.com.

Next-generation DNA sequencing is rapidly becoming an indispensable tool for genome-directed cancer diagnostics, but next-generation RNA sequencing (RNA-seq) is currently not standardly used in clinical diagnostics for expression assessment. However, multigene RNA diagnostic assays are used increasingly in the routine diagnosis of early-stage breast cancer. Two of the most widely used tests are currently available only as a central laboratory service, which limits their clinical use. We evaluated the use of RNA-seq as a decentralized method to perform such tests. The MammaPrint and BluePrint RNA-seq tests were found to be equivalent to the clinically validated microarray tests. The RNA-seq tests were highly reproducible when performed in different locations and were stable over time. The MammaPrint RNA-seq test was clinically validated. Our data demonstrate that RNA-seq can be used as a decentralized platform, yielding results substantially equivalent to results derived from the predicate diagnostic device. (J Mol Diagn 2019, 1:1–16; https://doi.org/10.1016/j.jmoldx.2019.04.007)

MammaPrint was one of the first gene-expression–based diagnostic tests introduced into the clinical management of breast cancer3–7 and the first in vitro diagnostic multivariate index assay using gene expression that obtained US Food and


Disclosures: Authors (except A.B.L., C.P.P., and S.C.L.) are employed by Agenda (R.B. and L.V.V. part-time). C.D. is a consultant for Agendia, the commercial entity that markets the 70-gene signature as MammaPrint and the 80-gene signature as BluePrint. L.V.V. and R.B. are shareholders of Agendia and are named inventors on the patent for the 70-gene signature used in this study. A.M.G. is named inventor on the patent for the 80-gene signature used in this study. A.B.L. and C.P.P. are employed by Agilent Technologies, the commercial entity that manufactures the microarrays and SureSelect next-generation sequencing probes and reagents for the MammaPrint and BluePrint assays. No writing assistance was used in the production of this article.

Copyright © 2019 American Society for Investigative Pathology and the Association for Molecular Pathology. Published by Elsevier Inc.

http://creativecommons.org/licenses/by-nc-nd/4.0

https://doi.org/10.1016/j.jmoldx.2019.04.007
Drug Administration (FDA) 510(k) clearances for both fresh-frozen (FF) and formalin-fixed, paraflin-embedded (FFPE) tissue. Extensive validation studies and, above all, the recent MINDACT clinical trial have demonstrated the clinical utility of MammaPrint at the highest level of clinical evidence (level 1A), making it a unique example of a microarray-based clinical diagnostic test that helps guide physicians in adjuvant treatment decisions for early-stage breast cancer patients. Specifically, MammaPrint determines the expression of 70 signature genes and stratifies early-stage breast cancer patients at low and high risk for developing distant metastases within 5 years after diagnosis. A second microarray test for breast cancer is BluePrint, which measures the expression of 80 signature genes to reveal breast cancer subtypes (ie, luminal, HER2, and basal types).

Currently, MammaPrint and BluePrint tests are performed at central laboratories in the Netherlands and United States using microarray technology. With a centralized setting, it is difficult to serve countries with legal restrictions on sending patient material outside the country; furthermore, reimbursement often requires local processing and forms a practical obstacle to clinical implementation of these tests. Therefore, decentralized diagnostic tests would offer to physicians an in-house solution, ideally without compromising the level of clinical validity. RNA-seq NGS technology is highly reproducible and currently easier to implement than microarray in preexisting infrastructure and experimental workflows because of the increased use of RNA-seq technology in many laboratories worldwide. In addition, RNA NGS shows good performance with RNA derived from FFPE tissues, even from small amounts of starting material. Herein, we evaluate the use of RNA NGS—based technology to perform clinical diagnostic testing, using the MammaPrint/BluePrint FFPE microarray-based assay as an example.

Several NGS technologies are available to date, including ion semiconductor sequencing, single-molecule real-time sequencing, nanopore sequencing and sequencing by synthesis. The sequencing-by-synthesis technology accounts for the largest share of the market, mostly because of the development of new and advanced NGS platforms and the increasing demand for the systems (including the HiSeq series and MiSeq) of Illumina (Hayward, CA). A targeted RNA-seq approach on the Illumina NGS technology that combines the RNA (cDNA) library preparation with the enrichment of our MammaPrint/BluePrint transcripts of interest using complementary capture probes was chosen. More important, such a targeted RNA NGS approach does not rely on poly-A tail enrichment or use ribosomal RNA depletion (eg, Ribo-Zero), and it can be reliably applied on degraded and cross-linked RNAs, such as those from FFPE tissues.

The translation of a diagnostic test from one technology platform to another one requires a series of experiments that assess the analytical and clinical performance of the newly developed test with respect to the standard technology. Therefore, different validation experiments were performed to assess the equivalence of the MammaPrint and BluePrint NGS test to the diagnostic microarray test and its reproducibility between different samplings of the same tumor block (surrogate for intratumor heterogeneity), over time (precision) and between different laboratories (interlaboratory agreement). Moreover, the clinical performance of the MammaPrint and BluePrint NGS test was assessed in a set of samples with known clinical outcome, which were previously used to validate the clinical utility of the MammaPrint FDA-cleared (Dx) diagnostic test.

We describe herein the development of targeted RNA NGS MammaPrint/BluePrint tests and their technical and clinical performance compared with the standard microarray-based MammaPrint and BluePrint diagnostic tests.

Materials and Methods

Sample Selection

FFPE samples selected for this study fulfilled the following criteria: i) They complied with the intended use of MammaPrint and BluePrint in vitro diagnostic tests: stage I, stage II, and operable stage III invasive breast cancer, tumor size ≤5.0 cm, and lymph node-negative and one to three lymph node-positive patients. ii) They had microarray MammaPrint results covering the entire MammaPrint index range. iii) They had at least five samples for each of the three BluePrint subtypes based on microarray analysis. In addition to these including criteria, FFPE samples selected for the clinical comparison had a minimum follow-up of 5 years from the date of diagnosis.

Targeted RNA-Seq Workflow

Overview

To develop the targeted RNA NGS MammaPrint/BluePrint test, a modified version of the Agilent SureSelectXT procedure, which combines a strand-specific RNA library with an RNA Enrichment System starting from FFPE tissue, was used. For the general laboratory procedure, please refer to the SureSelectXT RNA Direct Protocol (as per manufacturer’s instructions; Agilent Technologies, Santa Clara, CA); and for the Agendia MammaPrint/BluePrint NGS test procedure, please visit the Agendia website (Agenda NV, Amsterdam, the Netherlands). The SureSelect Target Enrichment workflow is a solution-based system using ultralong 120-mer biotinylated cRNA baits (so-called capture probes) to capture regions of interest, enriching them out of a NGS genomic fragment library. The Agilent SureSelectXT technology performs a post-capture indexing where samples are indexed (through a PCR enrichment step) after the capture is performed on each sample individually. After the target enrichment preparation, the cDNA libraries were sequenced using the MiSeq [research use only (RUO)]
and DX] Illumina sequencers. Subsequently, the expression of the MammaPrint and BluePrint signature genes was assessed using the sequence read count data and translated into the outcome test prediction. Supplemental Figure S1 represents the overall laboratory workflow performed in this study.

NGS Probe Design
First, the Agendia MammaPrint and BluePrint signature genes, previously reported,10,18 were translated into capture probes to be used in the Agilent SureSelect21 workflow. The target enrichment panel (ie, capture probes) was designed by first mapping the microarray oligonucleotide sequences to hg19 (ie, human assembly GRCh37), followed by target gene annotation. Genomic locations of probes mapping uniquely to intergenic or noncoding regions were confirmed against hg38 (ie, human assembly GRCh38) using BLAT19; 120-bp capture probes were then tiled across full transcripts based on RefSeq20, CCDS,21 Ensembl,22 GENCODE,23 and VEGA24 reference databases. The target regions used to design the SureSelect probes are reported in Supplemental Table S1. Each signature gene was targeted by multiple probes, from a minimum of 2 up to a maximum of 184 target probes.

The probe selection process resulted in an Agendia bait capture library covering 0.92 megabases of genomic regions. The Agendia bait capture library (alias MammaPrint BluePrint NGS Panel) targets a total of 178 unique genes.

RNA Target Enrichment
Prepared cDNA adapter—ligated library was hybridized to the Agendia bait capture library overnight. The hybridization reaction combined the prepared cDNA library with the SureSelect hybridization reagents and specific blocking agents. After hybridization was completed (17 to 24 hours), cDNA library/capture library hybrid was captured using streptavidin-coated beads that bind to the biotinylated cRNA baits. Next, the cDNA adapter—ligated and captured library was enriched by PCR amplification. The PCR amplification step adds index tags while it amplifies the captured library. Samples (cDNA adapter—ligated, captured, and indexed libraries) can then be pooled for multiplexed sequencing. Quality and quantity of the cDNA adapter—ligated and captured library were assessed using the Agilent 2100 Bioanalyzer instrument. To proceed with sequencing, at least 1 mmol/L of cDNA adapter—ligated and capture library (between 150 and 700 bp) was targeted.

Sequencing
The probe selection process resulted in a hybridization to equimolar concentration before sequencing. Sample pools of 1, 2, or 4 nmol/L starting concentration were denatured using sodium hydroxide and diluted with hybridization HT1 buffer, as per manufacturer’s instructions (Illumina). Samples were single-end sequenced on the Illumina MiSeq or MiSeq DX (RUO mode) instrument at the length of 150 bp using the MiSeq Reagent Kit V3 150 cycles). A single-end sequencing protocol was chosen because it provided sufficient read count and resolution to assess the expression of the genes in our target panel.

Sequencing Data Analysis
Sequence reads in the FASTQ format were generated and analyzed following a multistep procedure (Supplemental Figure S2). First, the quality of each FASTQ sequence was assessed using the FASTQC version 0.11.5 software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc; last accessed November 13, 2018). Next, the reads were trimmed with the Trimmomatic version 0.36 software,25 and trimmed sequence reads were mapped to the Ensembl GRCh37 human genome using the HISAT2 version 2.0.4 software.26 HISAT2 output files in a SAM (sequence alignment/map format) were compressed to binary sequence alignment/map files using SAMtools version 1.3.127 prior sequence read quantification. Afterward, mapped reads were counted using the feature Counts version 1.5.1 software.28 FASTQ files of the samples analyzed in this study have been submitted to the European Nucleotide Archive database (https://www.ebi.ac.uk/ena) under the study accession number PRJEB31802.

MammaPrint and BluePrint Outcome Calculation
For each gene in every sample, raw sequence read count was normalized using a count per million normalization method. By correcting the total counts in each sample by the
total number of reads, samples sequenced in different pools can be compared with each other. Normalization was performed in R version 3.2.2. After the data were normalized, MammaPrint and BluePrint indexes were calculated using the same algorithm for current MammaPrint and BluePrint tests as previously described. In brief, the MammaPrint indexes are calculated by taking the expression of the 70 MammaPrint genes for each sample and correlating them to the MammaPrint Risk templates. If the MammaPrint index is >0, the sample is classified as low risk; when the index is ≤0, then the sample is classified as high risk. Similarly, the BluePrint indexes are calculated by taking the expression of the 80 BluePrint genes and comparing them with the three different templates, one for each of the subtypes (luminal, HER2, and basal types). The subtype with the highest (scaled) correlation is the reported subtype.

Data Analysis and Statistical Analysis

Data were analyzed using the MATLAB software version R2012a (MathWorks, Natick, MA) and RStudio software version 1.1.442, based on R software version 3.4.2. Data were visualized using MATLAB software, Multiple Experiment Viewer version 4.8.1, Microsoft PowerPoint (Microsoft Corp., Redmond, WA), and Excel 2016 MOS 16.0.9330.2124 (Microsoft Corp.). Unsupervised clustering analyses were performed with Multiple Experiment Viewer.

The following was done to compare gene expression levels obtained with NGS and microarray technologies. For each of the 70 MammaPrint and 80 BluePrint genes, expression levels were compared between 204 NGS samples and 204 matched microarray samples using Pearson correlation (NGS versus microarray paired). For each signature gene, a not paired comparison was also performed, where its NGS expression was correlated to the microarray or the NGS expression of any of the other signature genes and vice versa (NGS versus not paired microarray, NGS versus not paired NGS, microarray versus not paired NGS, and microarray versus not paired microarray). Pearson coefficients generated per gene from the not paired comparisons were averaged to obtain one not paired Pearson coefficient value per gene.

To compare the 70 MammaPrint gene expression levels in different MammaPrint risk groups, for each of the 70 MammaPrint genes, the average NGS and microarray median centered expression of the high- and the low-risk patient samples was calculated separately. NGS and microarray high- and low-risk aggregate expressions were compared using Pearson correlation. Similarly, for each of the 80 BluePrint genes, the average NGS and microarray median centered expression of the luminal-, HER2-, and basal-type patient samples was calculated separately. NGS and microarray luminal-, HER2-, and basal-type aggregate expression levels were compared using Pearson correlation.

The equivalence of MammaPrint and BluePrint indexes was determined by a Pearson correlation for assessment of the degree of linear correlation and a Passing-Bablok regression analysis to obtain the regression equation. Scatterplots were used to visually examine the existence of any constant bias in the difference of measurements between samples analyzed with targeted RNA-seq and microarray technologies. Acceptable limits were defined a priori, on the basis of microarray data generated at Agency between 2012 and 2016 using the Agenda’s Test Suite. MammaPrint and BluePrint test outcomes (MammaPrint: high/low risk; BluePrint: luminal, basal, and HER2 types), assessed using targeted RNA-sequencing technology, were compared with the standard microarray diagnostic outcomes using a contingency table. On the basis of the contingency table, concordance, negative percentage agreement (NPA), and positive percentage agreement (PPA) were determined. The NPA assesses whether the percentage of MammaPrint low-risk results obtained with microarray technology is concordant with the results obtained with the targeted RNA-seq technology. Similarly, PPA is the percentage of MammaPrint high-risk results obtained by one technique compared with that obtained by the other technique.

Concordance, NPA, and PPA should be ≥90%. The relative stability (reproducibility) is calculated by 100 minus the relative SD, which is the SD measured as percentile of the total MammaPrint or BluePrint NGS dynamic range assessed on a set of 326 samples sequenced in the Agenda laboratories. Clinicopathologic data and clinical outcome data were analyzed using the statistical package SPSS 22.0 for Windows (SPSS Inc., Chicago, IL). All validation experiments to assess technical equivalence, reproducibility, tumor heterogeneity, and clinical performance of the targeted RNA NGS MammaPrint/BluePrint test were designed according to guidelines of the FDA. Interlaboratory reproducibility was tested at the Agenda central laboratory (Amsterdam, the Netherlands; site 1), at University Hospital Leuven (Leuven, Belgium; site 2), and at Curie Institute (Paris, France; site 3).

Study Approval

Only data and not samples were collected for this study. All data and analyses used or performed in this study comply with the current ethical laws of the Netherlands. All patient sample data were anonymized in accordance with national ethical guidelines (Code for Proper Secondary Use of Human Tissues, Dutch Federation of Medical Scientific Societies), and study samples had institutional review board approvals.

Results

Targeted RNA Next-Generation Sequencing Performance

To represent the MammaPrint and BluePrint diagnostic signatures, a target gene panel consisting of a total of 178
genes (MammaPrint, BluePrint, and control genes), each represented by different 120-mer capture probes targeting their exome sequence (see Materials and Methods for details), was selected. To assess the specificity of the sequence capture strategy, a relevant parameter is the number of reads that map to the targeted sequence that, in this study, corresponds to 0.9 megabases in size.

To assess the analytical performance of the MammaPrint/BluePrint targeted RNA NGS test (referred to as NGS for brevity), a total of 327 FFPE breast cancer samples (Supplemental Table S2) were successfully sequenced in this study. On average, 21 samples were pooled per MiSeq run (minimum number of samples per pool = 9, maximum number of samples per pool = 32). Overall, an average of 9.45 × 10^5 reads per sample (minimum = 1.95 × 10^5, maximum = 4.03 × 10^6), of which 96.3% mapped to the human genome (hg19), were obtained. Of all reads, on average, 79.8% were on target (eg, uniquely assigned to the MammaPrint, BluePrint, and control genes). This indicates that the capture strategy can efficiently enrich the MammaPrint and BluePrint signature genes. Figure 1 depicts the relative mapping efficiency for the 327 samples by showing the relative percentage of reads mapped to the human genome (mapped read count) and the relative percentage of reads on target (on-target read count) versus the total read count set to 100%.

When looking at the normalized read count of MammaPrint and BluePrint genes in the 327 samples, an average of 2.45 × 10^5 reads per MammaPrint gene (maximum = 3.23 × 10^5, minimum = 11) and an average of 4.93 × 10^5 reads per BluePrint gene (maximum = 2.87 × 10^6, minimum = 35) were observed (Supplemental Figure S3). These read count estimates indicated that all target genes were expressed in the sample set, and their expression was captured during the target enrichment step. High variation in the gene read counts is expected because of the biological variation in expression of the signature genes, depending on the MammaPrint risk group and BluePrint subtype of the sample.

**Gene Expression Comparative Analysis between NGS and Microarray**

To assess the correlation between gene expression measured with NGS and microarray, the MammaPrint 70-gene and BluePrint 80-gene expression levels were analyzed in a subset of 204 unique FFPE patient samples for which matched NGS and microarray data were available from the same RNA sample. NGS and microarray normalized expression was median centered per gene separately for each technique (ie, NGS and microarray). A similar strategy has been used previously in a different context, to adjust for platform bias.32–34

First, the Pearson correlation was calculated for each MammaPrint and BluePrint gene between its paired and not paired median centered expression across the 204 matched NGS and microarray samples. The distribution of Pearson correlation coefficients generated from the comparison between NGS and microarray paired gene expression profiles differed significantly from the distribution of the not paired profiles for both MammaPrint and BluePrint genes, indicating a good concordance between gene expression measured with the two techniques (Figure 2). Specifically, for NGS and microarray paired profiles, the average Pearson correlation was 0.582 ± 0.223 for the MammaPrint genes and 0.748 ± 0.188 for the BluePrint genes. For not paired profiles, the Pearson correlation coefficient significantly decreased (Wilcoxon signed-rank test P < 0.0001) for both MammaPrint and BluePrint genes (MammaPrint genes: 0.064 ± 0.089 for NGS versus not paired microarray, 0.064 ± 0.171 for microarray versus not paired NGS, 0.213 ± 0.194 for NGS versus not paired NGS, and 0.195 ± 0.120 for microarray versus not paired microarray; BluePrint genes: 0.104 ± 0.198 for NGS versus not paired microarray, 0.104 ± 0.125 for microarray versus not paired NGS, 0.102 ± 0.144 for NGS versus not paired NGS, and 0.246 ± 0.180 for microarray versus not paired microarray).

As expected, the average correlation between not paired samples of the same technique type is higher than the one measured between not paired samples of a different technique type (MammaPrint genes: 0.204 versus 0.064; BluePrint genes: 0.174 versus 0.104).

Next, the expression of the 70 MammaPrint genes and the 80 BluePrint genes measured with NGS, with the expression measured with microarray using heat maps, was visually compared. NGS expression was highly comparable with the microarray expression for both MammaPrint and BluePrint genes (Figure 3). For both MammaPrint and BluePrint genes, NGS expression intensity was greater than the
Materials and Methods

(see Materials and Methods).

Figure 2  Distribution of Pearson correlation coefficients generated from the comparison between next-generation sequencing (NGS) and microarray (MA) expression of the MammaPrint (A) and the BluePrint (B) genes in paired (NGS versus MA) and not paired (NGS versus not paired MA, NGS versus not paired NGS, MA versus not paired NGS, and MA versus not paired MA) gene expression profiles across 204 NGS and MA matched patient samples. The x axis reports the comparison type; the y axis reports the average Pearson correlation of each comparison type for each of the MammaPrint or BluePrint genes. Each dot represents a gene for which the average of the Pearson correlations for each comparison type is calculated. Normalized NGS and microarray expression data are median centered before correlation calculation. Color code: NGS versus MA paired, black; NGS versus not paired MA, white; NGS versus not paired NGS, gray; MA versus not paired NGS, light gray; MA versus not paired MA, light yellow. N = 70 (A); N = 80 (B).

MammaPrint NGS indexes showed high correlation with the matching microarray indexes (Pearson r = 0.978), as confirmed by the Passing-Bablok regression analysis.[F5] (Figure 4A and Supplemental Table S3). The Bland-Altman analysis showed no bias toward one platform, and the mean difference between the NGS and the microarray indexes was 0.1% of the observed dynamic range for the NGS test; this difference is similar to the technical variance observed using FF tissue on the microarray-to-microarray comparison.[F4] Concordance of categorical MammaPrint low- and high-risk classification was 97.6% with an NPA of 94.0% (95% CI, 81.3%–98.5%) and a PPA of 100%. A κ score of 0.82 (95% CI, 0.74–0.89) indicated almost perfect agreement.[F3] The samples that were discordant in MammaPrint classification results (N = 2) had indexes close to the classification threshold, with a difference between the NGS and the microarray indexes of 0.018 and to 0.137, respectively. These differences fall within 1 SD from the average index difference (0.083 ± 0.077) of the full sample set (N = 85).

Similar to MammaPrint, BluePrint NGS indexes for each subtype showed high correlation with the matching microarray indexes (luminal-type Pearson r = 0.979, HER2-type Pearson r = 0.788, basal-type Pearson r = 0.867), as confirmed by the Passing-Bablok regression analysis (Figure 5A and Supplemental Table S3). The correlation between NGS and microarray expression intensity, indicating a larger gene expression dynamic range for NGS than for microarray for MammaPrint and BluePrint genes. Of importance, the 70-gene heat map for both NGS and microarray resembles the heat map originally published using fresh-frozen RNA.[F3] Pearson correlation between the average NGS and microarray expression of the 70 genes in MammaPrint high-risk (N = 122) and low-risk (N = 82) patients was high (high-risk Pearson r = 0.870, low-risk Pearson r = 0.876). Similar to the 70 genes, strong correlations were observed between the average NGS and microarray expression of the 80 genes in BluePrint luminal-type (N = 133), HER2-type (N = 42), and basal-type (N = 29) patient samples (luminal-type Pearson r = 0.902, HER2-type Pearson r = 0.887, basal-type Pearson r = 0.869).

Technical Equivalence of MammaPrint and BluePrint Results for NGS and Microarray

Equivalence of NGS and microarray technologies was studied by performing both diagnostic tests on a set of 85 FFPE breast cancer samples for MammaPrint and on a set of 98 breast cancer samples for BluePrint (Supplemental Table S2). All samples fulfilled the selection criteria of the study (see Materials and Methods).
microarray HER2-type indexes was lower than for the luminal and basal-type indexes. This is because of the fact that most samples were not HER2 type and, therefore, the HER2 NGS and microarray indexes of non-HER2-type samples may differ more from each other. In addition, higher variability for the HER2 index is expected because the HER2 subtype is defined by a signature of four genes and expression changes occurring in each of the four HER2 signature genes can have a higher impact on the HER2 signature readout than expression changes occurring in each of the luminal and basal signature genes that are part of a more numerous set (58 luminal signature, 28 basal signature). Of importance, most genes that define the HER2 signature are underexpressed in non-HER2-type samples and, therefore, the measurement precision based on low expressed genes could be poorer. For all three subtypes, the Bland-Altman analysis showed no bias toward one platform, with a mean difference between the NGS and the microarray indexes smaller than the control technical variance observed for FFPE microarray. The mean difference between the NGS and the microarray indexes was 0.8%, 2.8%, and 2.5% of the observed NGS dynamic range for the luminal, HER2, and basal types, respectively (Supplemental Table S4). Overall concordance of categorical luminal-, HER2-, and basal-type classification was 100%.

Reproducibility of MammaPrint and BluePrint NGS Results Using Two Independent RNA Isolations from the Same Tumor Tissue Block

Cellular heterogeneity can occur within the same primary tumor (alias intratumor heterogeneity).36–38 Gene expression analysis results from different parts of the same tumor can, therefore, be variable. To evaluate the possible effect of intratumor heterogeneity, the reproducibility of the MammaPrint/BluePrint NGS test was also assessed between two different RNA isolations from the same FFPE tumor block.
Each sample block was sectioned twice: the first set of four slides for the first isolation and the second set of four slides for the second isolation. This analysis is an approximation of tumor heterogeneity; a true tumor heterogeneity study would analyze two different parts of the tumor. Moreover, this experiment aims to assess a more so-called biological variable (ie, cellular heterogeneity of the same primary tumor) than an analytical variable. Forty-three matched samples were collected for this analysis. A high correlation was observed in MammaPrint NGS indexes, as well as for BluePrint NGS indexes per subtype derived from these two independent isolations (MammaPrint Pearson r = 0.993, luminal-type Pearson r = 0.995, basal-type Pearson r = 0.993, HER2-type Pearson r = 0.990), as confirmed by the Passing-Bablok regression analysis. In this analysis, the intercepts for MammaPrint and BluePrint are close to zero and the slopes are close to one, indicating that there is no bias in results between the two isolations (Figure 4B, Figure 5B, and Supplemental Table S3). For both MammaPrint and BluePrint, the Bland-Altman analysis showed no bias toward one of the two isolations, with a mean difference between the matched NGS indexes smaller than the microarray control technical variance. The mean difference between the matched NGS indexes was 0.2%, 0.1%, 2.1%, and 0.4% of the observed NGS dynamic range for MammaPrint, BluePrint luminal type, BluePrint HER2 type, and BluePrint basal type, respectively (Supplemental Table S4). Concordance of categorical MammaPrint low- and high-risk classification between the two isolations was 97.7%, with an NPA of 100% and a PPA of 96.0% (95% CI, 80.5%–99.3%). Overall concordance of categorical BluePrint luminal-, HER2-, and basal-type classification was 100%.

**Figure 4** Scatterplot comparison analysis of MammaPrint indexes in different validation studies. Each dot represents a single breast cancer sample for which total RNA underwent next-generation sequencing (NGS) or microarray laboratory processing and analysis. A: Scatterplot showing equivalence of MammaPrint microarray [predicate device, US Food and Drug Administration (FDA) 510(K141142)] indexes (x axis) versus MammaPrint NGS (targeted RNA NGS test) indexes (y axis). B: Scatterplot showing equivalence of MammaPrint NGS indexes assessed from RNAs isolated from two independent isolations, isolation 1 (x axis) versus isolation 2 (y axis). C: Scatterplot showing equivalence of MammaPrint NGS indexes assessed using MiSeq instrument (x axis) versus MammaPrint NGS indexes assessed using MiSeq FDA-cleared (Dx) instrument (y axis). D: Scatterplot showing equivalence of MammaPrint NGS indexes derived from Qiagen RNeasy isolation method (x axis) versus MammaPrint NGS indexes derived from Qiagen AllPrep isolation method (y axis).
Figure 5  Scatterplot comparison analysis of BluePrint (BP) luminal, HER2, and basal indexes in different validation studies. Each dot represents a single breast cancer sample for which total RNA underwent next-generation sequencing (NGS) or microarray laboratory processing and analysis. A: Scatterplots showing equivalence of BluePrint microarray indexes (x axis) versus BluePrint NGS (targeted RNA NGS test) indexes (y axis) in the luminal, HER2, and basal subtypes. B: Scatterplots showing equivalence of BluePrint NGS indexes assessed from RNAs isolated from two independent isolations, isolation 1 (x axis) versus isolation 2 (y axis), in the luminal, HER2, and basal subtypes. C: Scatterplots showing equivalence of BluePrint NGS indexes assessed using MiSeq instrument (x axis) versus BluePrint NGS indexes assessed using MiSeq US Food and Drug Administration—cleared (Dx) instrument (y axis) in the luminal, HER2, and basal subtypes. D: Scatterplots showing equivalence of BluePrint NGS indexes derived from Qiagen RNeasy isolation method (x axis) versus BluePrint NGS indexes derived from Qiagen AllPrep isolation method (y axis) in the luminal, HER2, and basal subtypes.
Reproducibility of MammaPrint and BluePrint NGS Results Using MiSeq and MiSeq Dx Instruments

Currently, Illumina offers an RUO and an Dx version of the MiSeq instrument. The MiSeq Dx can be also used in a RUO mode to allow users to run customized sequencing protocols. Many diagnostic laboratories around the world are starting to use a MiSeq Dx instrument; therefore, results generated from both instruments (MiSeq and MiSeq Dx in RUO mode) were compared using the same setting, to assess their degree of reproducibility. The experiments were performed in two centralized Agenda laboratories, one in Irvine, CA, using the MiSeq Dx instrument and one in Amsterdam, the Netherlands, using the MiSeq instrument. This analysis also included possible effects of site-to-site variation in the performance of the NGS test. Fifty FFPE samples were selected and processed in both locations, starting from previously isolated RNA. Two samples failed the internal quality control (QC) assessment and were excluded from the analysis; therefore, 48 matched samples were left for further analysis. Figures 4C and 5C show scatterplots of MammaPrint and BluePrint indexes generated at the two sites with MiSeq and MiSeq Dx. A high correlation was observed in MammaPrint NGS indexes, as well as for the BluePrint NGS indexes for each subtype derived from the two-site NGS processing and sequencing (MammaPrint Pearson $r = 0.991$, luminal-type Pearson $r = 0.998$, HER2-type Pearson $r = 0.992$, basal-type Pearson $r = 0.995$), as confirmed by the Passing-Bablok regression analysis (Supplemental Table S3). For both MammaPrint and BluePrint, the Bland-Altman analysis showed no bias toward one of the two isolations, with a mean difference between the matched NGS indexes smaller than the control technical variance. The mean difference between the matched NGS indexes was 1.0%, 1.1%, 1.6%, and 1.7% of the observed NGS dynamic range for MammaPrint, BluePrint luminal type, BluePrint HER2 type, and BluePrint basal type, respectively (Supplemental Table S4).

Concordance of categorical MammaPrint low- and high-risk classification between the two sites was 95.8% with an NPA of 89.5% (95% CI, 68.6%—97.1%) and a PPA of 100%. Two discordant cases were observed; however, the absolute differences between the NGS indexes obtained with the two instruments were smaller than the control technical variance. Overall concordance of categorical BluePrint luminal-, HER2-, and basal-type classification was 97.9%. The use of two different sequencing instruments in two different locations had no significant influence on the outcome of the test.

Reproducibility of MammaPrint and BluePrint NGS Results Over Time

Reproducibility over time is a key indicator of precision for a diagnostic test. Therefore, three control patient samples (control 1, control 2, control 3) were repeatedly processed and sequenced over a 6-month period, starting from total RNA onwards. Controls 1 and 2 were MammaPrint low risk, and control 3 was MammaPrint high risk. For control 1, there were 26 NGS measurements collected, of which 25 passed the QC assessment; for control 2, there were 23 NGS measurements collected, of which 17 passed the QC assessment; for control 3, there were 16 NGS measurements collected, of which 14 passed the QC assessment.

In this long-term validation experiment, nearly all potential sources of variation are included by default (ie, different operators, reagent batches, and environmental factors). For both MammaPrint and BluePrint categorical classifications, the concordance was 100% between all the measurements of each control sample. The reproducibility was measured by the relative SD of the samples. MammaPrint showed stable results over time (Figure 6) for all three control patient samples, with relative SDs of 4.1% (control 1, 16 measurements), 3.9% (control 2, 14 measurements), and 2.2% (control 3, 16 measurements). The MammaPrint NGS test showed a reproducibility (ie, relative stability) of 98.5%, 98.0%, and 98.2% in the three control samples, respectively (Supplemental Table S4). This reproducibility is slightly higher than using the microarray-based test.39 Similar to MammaPrint, BluePrint showed stable results over time, with an average relative SD over the three control samples of 1.5% (luminal type), 2.6% (HER2 type), and 1.5% (basal type) (Supplemental Table S4). On the basis of these three control samples, the reproducibility of the BluePrint NGS test was 98.5% for the luminal type, 98.4% for the basal type, and 97.4% for the HER2 type (Supplemental Table S4). None of the three samples analyzed over time was classified as HER2 subtype and, therefore, the HER2 subtype index reproducibility was not measured. This had no effect on the BluePrint outcomes.

Because there was a higher failure rate for control sample 2, which was isolated with a different RNA isolation method than the other two control samples (ie, Qiagen All Prep kit versus Qiagen RNeasy kit), it was further studied. Twenty-eight matched FFPE samples were selected, and each block was sectioned twice; and the first set of four slides was isolated with the RNeasy method, whereas the second set of four slides was isolated with the All Prep isolation methods. Of the 56 samples analyzed (28 pairs), four samples failed for both isolation methods, two for the RNeasy method, and six for the All Prep method. This result would support the idea that All Prep gives a higher failure rate than the RNeasy method. However, larger sample numbers are needed to confirm and elucidate this finding. Additional validation of this hypothesis is still ongoing. Of importance, when looking at the results obtained for 16 RNeasy-AllPrep pairs successfully sequenced, high correlation was observed between the two isolation methods for both MammaPrint NGS indexes and BluePrint NGS indexes for each subtype (MammaPrint Pearson $r = 0.992$, luminal-type Pearson $r = 0.997$, HER2-type Pearson $r = 0.947$, basal-type Pearson $r = 0.996$) (Figure 4D, Figure 5D, and Supplemental Table S3). The
mean difference between the matched NGS indexes was 1.5%, 0.7%, 1.8%, and 0.04% of the observed NGS dynamic range for MammaPrint, BluePrint luminal type, BluePrint HER2 type, and BluePrint basal type, respectively (Supplemental Table S4). Concordance of categorical MammaPrint low- and high-risk classification and BluePrint subtype classification between the two isolation methods was 100%.

Reproducibility of MammaPrint and BluePrint NGS Results between Different Laboratories

The main objective of developing a MammaPrint/BluePrint NGS test is the implementation of a decentralized diagnostic setting. Therefore, reproducibility was assessed between different laboratories by processing the same set of 15 FFPE RNA samples in three independent laboratories: Agendia central laboratory (site 1) and two decentralized academic laboratories (sites 2 and 3) (see Materials and Methods). The three sites used a MiSeq instrument with the same setting. One sample analyzed at site 2 did not pass the QC assessment and, therefore, 14 samples were used for the analysis.

For both decentralized sites, high correlation was observed between their MammaPrint and BluePrint NGS indexes and the Agendia MammaPrint and BluePrint NGS indexes (site 1 versus site 2: MammaPrint Pearson $r = 0.998$, luminal-type Pearson $r = 0.995$, HER2-type Pearson $r = 0.990$, basal-type Pearson $r = 0.994$; Agendia site 1 versus site 3: MammaPrint Pearson $r = 0.997$, luminal-type Pearson $r = 0.996$, HER2-type Pearson $r = 0.986$, basal-type Pearson $r = 0.991$), as confirmed by the Passing-Bablok regression analysis (Figure 7 and Supplemental Table S3). For both MammaPrint (Figure 7A) and BluePrint (Figure 7, B and C), the scatterplot and the Bland-Altman analysis showed no bias toward one of the sites (site 1 versus site 2, site 1 versus site 3), with a mean difference between the matched NGS indexes smaller than the control technical variance (site 1 versus site 2: mean difference in NGS-microarray indexes for MammaPrint was 0.6%, for
BluePrint luminal type was 0.8%, for HER2 type was 4.8%, and for basal type was 0.5% of the observed NGS dynamic range; site 1 versus site 3: mean difference in NGS-microarray indexes for MammaPrint was 2.3%, for BluePrint luminal type was 0.7%, for HER2 type was 1.0%, and for basal type was 1.1% of the observed NGS dynamic range (Supplemental Table S4). Concordance of 100% for categorical MammaPrint low- and high-risk classification and BluePrint subtype classification was observed between the three laboratories.

**Clinical Performance Comparison**

As described above, the NGS-based test proved to be technically equivalent to the microarray diagnostic test, showing high correlation between the NGS- and the microarray-based indexes. Nevertheless, highly correlated continuous values may result in a lower concordance when these continuous variables are presented as binary categories, as is the case for MammaPrint low- and high-risk classification groups. In

![Figure 7](jmd.amjpathol.org) Scatterplot comparison analysis of MammaPrint and BluePrint (BP) next-generation sequencing (NGS) indexes (luminal, HER2, and basal types) assessed at the Agenda (Amsterdam, the Netherlands) central laboratory and at decentralized laboratories at sites 2 and 3. Each dot represents a single breast cancer sample for which total RNA underwent NGS or microarray laboratory processing and analysis. **A:** Scatterplot showing equivalence of MammaPrint NGS indexes assessed at Agenda (x axis) versus MammaPrint NGS indexes assessed at site 2 (y axis; left panel) and site 3 (right panel). **B:** Scatterplot showing equivalence of BluePrint NGS indexes assessed at Agenda (x axis) versus MammaPrint NGS indexes assessed at site 2 (y axis). **C:** Scatterplot showing equivalence of BluePrint NGS indexes assessed at Agenda (x axis) versus MammaPrint NGS indexes assessed at site 3 (y axis).
particular, samples with a result close to a classification threshold with a small difference between two measurements may still result in a switch in binary outcome.39 Therefore, the clinical performance of the NGS-based test was assessed in a data set with known clinical outcome. A large series of breast tumors were used from the microarray-prognostic-in-breast-cancer (RASTER) study, for which 5-year outcome data are available and were used to compare different versions of the MammaPrint microarray diagnostic test (ie, MammaPrint FF versus FFPE).5,39 All patients were aged 18 to 61 years and had a histologically confirmed invasive adenocarcinoma of the breast (cT1-3N0M0). For 341 samples of the original series of 427, total RNA was available. Because of the age of the samples (on average, 12 years after surgery), 316 were successfully processed with the NGS test (Supplemental Table S2). Clinicopathologic characteristics of the 316 patients analyzed with NGS and grouped by MammaPrint risk outcome are reported in Supplemental Table S5. The 316 patient subset and the original 427 patient set show similar clinicopathologic characteristics, as confirmed by the P value after testing for difference (Pearson \( \chi^2 \) statistic for categorical variables and \( U \)-test for continuous variables, \( P > 0.05 \)) (Supplemental Table S6). Survival analysis of the 316 patient samples for 5-year distant recurrence-free interval was performed for MammaPrint FF microarray [predicate device, FDA 510(K062694/K070675)] as well as for MammaPrint NGS. Figure 8 shows that the Kaplan-Meier curves are similar between the microarray and NGS results and that MammaPrint NGS high- and low-risk groups significantly differ in 5-year survival, with a log-rank \( P \) value smaller than the one observed for MammaPrint microarray (NGS log-rank \( P = 0.002 \) versus microarray log-rank \( P = 0.038 \)). Moreover, 5-year survival percentages of distant recurrence-free interval for both the low- and high-risk patient groups were comparable between microarray and NGS, as reported in Table 1.

Concordance of categorical MammaPrint low- and high-risk classification was 94.0% with an NPA of 96.3% (95% CI, 92.2%–98.3%) and a PPA of 91.4% (95% CI, 85.9%–94.9%). In summary, we show herein that diagnostic testing results using targeted NGS and microarray are comparable and lead to similar clinical performance of both tests in a cohort of patients with known outcome.

**Discussion**

Gene expression–based diagnostic tools are increasingly used to assess risk of recurrence in breast cancer.6,9,40,41 The two most commonly used tests, MammaPrint,6,8,9 and Oncotype DX,42,43 are currently only available to breast...
Mittempergher et al
cancer patients through central laboratory services. This limits the availability of these tests, as some countries have legal restrictions on the export of human tissue. To overcome this barrier to clinical use, an RNA-seq kit version of MammaPrint was developed and its reproducibility, stability, and clinical validity were tested.

In recent years, RNA NGS technology evolved rapidly, and it is increasingly chosen by genomics laboratories because of its low background, high sensitivity, and reproducibility. Efforts toward the establishment of the best practices for RNA NGS as a clinical test (reproducibility, accuracy, and precision) are ongoing.

In light of the potential of the RNA NGS technology for clinical applications, this study was initiated with the objective of translating the current MammaPrint and Blueprint FFPE microarray test to an RNA NGS–based test. A targeted RNA-seq approach was chosen because of the limited size of the gene target panel (approximately 200 genes, which cover 0.9 megabases) and because a target sequencing approach delivers high sequencing depth and high coverage on target. On average, a percentage of reads on target >79% was observed, which indicates a high-capture performance, especially for targeted sequencing of FFPE-derived RNA.

Overall, a high concordance was observed between results obtained using NGS and microarray technologies (all >97%). When NGS and microarray result from the same FFPE RNA were compared, a discordance rate of 2.4% was observed, which can only be ascribed to the assay-related variation. After combining assay-related variation and intratumor heterogeneity from our study (2.3%), a discordance rate of approximately 5% was obtained, which matches the discordance rate observed for the MammaPrint-fresh microarray test between two biopsy isolations of the same tumor. Interestingly, if it is considered that the concordance between the MammaPrint FFPE and FF test is 91.5% (95% CI, 87%–95%), it can be concluded indirectly that assay-related variation (NGS versus microarray techniques) is lower than the variation observed between the MammaPrint FF and MammaPrint FFPE microarray tests, when excluding the contribution of intratumor heterogeneity (2.4% FFPE NGS versus FFPE microarray; 3.5% FFPE microarray versus FF predicate microarray). This study suggests that sample preservation method (FF or FFPE) is a bigger contributor to variation than the technological platform (microarray or NGS). This result is in line with the study of Zhang et al, in which they showed that RNA-seq and microarray-based models perform similarly in clinical end point prediction for neuroblastoma patients. In their study, prediction accuracies were most strongly influenced by the nature of the clinical end point, whereas the technological platforms (RNA-seq versus microarray) did not significantly affect performances of the model.

Next, the reproducibility of the NGS test was assessed between different central laboratories using different MiSeq instruments. The objective was to measure the reproducibility of the NGS test, excluding the RNA isolation procedure; and therefore, for this experiment, the same total RNA was processed at the different Agenda sites. Again, high correlation was observed in a set of 48 matched RNA samples, between the NGS and microarray MammaPrint and BluePrint indexes (Pearson r coefficients all >0.99) generated at the two central laboratories, as well as high concordance between the categorical MammaPrint and BluePrint classifications (>95%).

In the technical validation studies described so far, a few samples that had a discordant MammaPrint and/or BluePrint NGS classification with respect to the microarray classification were observed. However, all these discordant cases showed an absolute index difference smaller than the microarray technical variance and, therefore, these discordant events could be accepted as part of the expected technical variation of the NGS test.

To assess more extensively the precision of the NGS test, three patient control samples (two low-risk and one high-risk MammaPrint) were repeatedly measured over a 6-month period. Reproducibility of the MammaPrint and BluePrint NGS indexes was >98% (minimum = 97.0%, maximum = 99.1%). The data showed that the MammaPrint and BluePrint NGS test is stable over time and that the same result is generated regardless of the operator, reagent lot, day, and other factors switching in the laboratory process.

To assess the reproducibility of the NGS test between decentralized laboratories (ie, not central laboratories), two independent academic European cancer centers (Curie Institute and University Hospital Leuven) processed the same set of RNA samples (n = 15) using the MammaPrint and BluePrint NGS test. Considering that the main reason of developing the MammaPrint/BluePrint test is to enable a decentralized use of diagnostic testing, assessing the reproducibility of the NGS test between independent laboratories is important to assess the feasibility of such a decentralized setting in an academic setting. A 100% concordance was observed between the MammaPrint and BluePrint outcome results obtained at the three sites (Agenda, University Hospital Leuven, Curie Institute), with high correlation (Pearson r > 0.98) between the MammaPrint and BluePrint NGS indexes. This result highlights once again the robustness of the NGS test, and it paves the way for a larger validation study in which samples are processed locally with the NGS kit.

In this study, its clinical performance was also assessed using a set of 316 breast cancer samples with known clinical outcome. Samples that switched between the different platforms may affect the clinical performance of the test. Therefore, 316 samples with known clinical outcome were analyzed and compared with MammaPrint results generated using the microarray platform (FF predicate device, FDA 510(K062694/K070675), same as used for the MINDACT series). The survival curves of MammaPrint high- and low-risk patients, based on the NGS...
result, were equivalent to those based on the microarray result. Both NGS and microarray have a significant difference in 5-year distant recurrence-free interval (log-rank \( P = 0.002 \)). The NGS kit showed a survival of 97% in the MammaPrint low-risk group in mostly untreated patients. When the MammaPrint risk prediction results obtained with the NGS test were compared with the one based on the reference microarray test, overall high concordance was obtained, with PPA and NPA values both >90%. This result indicates that in nearly all of the cases, patient treatment recommendations based on the NGS result are the same as those based on the microarray result; and the same patient group will be identified for safely forgoing chemotherapy (low-risk patients) or for undergoing necessary chemotherapy treatment (high-risk patients).

Discordant results were also observed, as could be expected from the technical equivalence study; however, this is part of the expected technical variation of the NGS test compared with the microarray test.

In summary, results presented herein show that MammaPrint and BluePrint indexes generated with the targeted RNA NGS test are not only technically equivalent to those generated with the microarray FFPE predicate device, but also highly reproducible between centralized laboratories, stable over time, and clinically meaningful. A larger study is required to confirm reproducibility between decentralized laboratories, but preliminary data generated in collaboration with two European academic cancer centers confirm high reproducibility between results from Agenda and external laboratories. The successful translation of the MammaPrint microarray test to an NGS-based test highlights the robustness of the biology behind the MammaPrint test, which underwent several translations over time6,8,30,39 and still proves to be a reliable tool for personalized medical care.

Acknowledgments

We thank Petra Sintubin, Liesbet Vliegen, and Laurence Slemrouk (University Hospital Leuven) and Virginie Raynal, Bohec Mylene, Cécile Reyes, and Audrey Rapinat (Curie Institute) for performing the next-generation sequencing (NGS) laboratory experiments for the ‘Reproducibility of MammaPrint and BluePrint NGS Results between Different Laboratories’ study.

Supplemental Data

Supplemental material for this article can be found at http://doi.org/10.1016/j.jmoldx.2019.04.007.

References


Mittempergher et al


1875. Food and Drug Administration: Statistical guidance on reporting results from studies evaluating diagnostic tests Available at: https://www.fda.gov/RegulatoryInformation/Guidances/ucm071148.htm; 2007


1892. 2012