Technology Assessment of Molecular Pathology Testing for the Estimation of Prognosis for Common Cancers
ID: CANG0212

Disposition of Comments

Draft review was available for public comment from November 20, 2013 to December 16, 2013.

Research Review Citation: Sreelatha Meleth, Ph.D., M.S., M.S., M.A., Katherine Reeder-Hayes, M.D., Mahima Ashok, Ph.D., M.S., Robert Clark, Ph.D., William Funkhouser, M.D., Ph.D., Roberta Wines, M.P.H., Christine Hill, M.P.A., Ellen Shanahan, M.A., Emily McClure, M.S.P.H., M.A., Katrina Burson, M.N., B.S.N., C.C.R.P., Manny Coker-Schwimmer, M.P.H., Nikhil Garge, M.S., Daniel E. Jonas, M.D., M.P.H, "Technology Assessment of Molecular Pathology Testing for the Estimation of Prognosis for Common Cancers." Technology Assessment. (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. CANG0212.) Rockville, MD: Agency for Healthcare Research and Quality. May 2014.

Comments to Research Review

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the Agency for Healthcare Research and Quality (AHRQ) Web site after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each invited peer reviewer comment is listed anonymously. Peer reviewers' names are acknowledged in the final report. Each public comment is listed with the name and affiliation of the commentator, if this information is provided. Public commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for the draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of AHRQ.

Table 1. Peer Reviewer Comments

Peer Reviewer	Section	Comment	Response
Peer Reviewer #1	Introduction	Page 18 Lines 8-10 While I understand that prediction of treatment response was not part of the review, however this is to some degree an artificial separation. In contrast to other prognostic factors, some of the genetic variants studied have some ability to predict response to treatment with specific agents. This has an impact on RFS, CSS and OS, as well as choice of alternative treatments. I think this distinction has to be made as evident and explicit as possible to mitigate the potential for misinterpretation. I'm concerned that the casual reader may confuse these two issues.	The mandate from CMS was to review the tests for the prognostic value of the test. We therefore omitted studies that looked at the predictive value of these tests. We have added a statement in both the executive summary and the beginning of the main report to clarify the intent of the report.
Peer Reviewer #2	Introduction	Refer to general comments about the clinical scenario. Otherwise, no comments.	Thank you. We have responded to each of the general comments.
Peer Reviewer #3	Introduction	Excellent	Thank you
Peer Reviewer #4	Introduction	Background and objectives were clearly stated. I refer to my attached comments.	Thank you
Peer Reviewer #1	Methods	The methods are appropriate and there are no significant problems. One issue related to clarity in defining the outcomes. Page 24 lines 3-4 The sentence states the "…outcome of interest in this review is recurrence of the cancer after the initial resection and treatment." If this is the outcome of interest, why are other outcomes (CSS, OS) included? I think it's valuable to include the other outcomes, but rationale for this should be provided	Thank you. The rationale behind including CSS and OS was that both of these outcomes would be affected by the recurrence of the cancer. Thus in the absence of data on the ability of the test to predict recurrence, its ability to predict CSS or OS was used as a surrogate. We have tried to clarify this in the statement on pg. 24.

Peer Reviewer	Section	Comment	Response
Peer Reviewer #2	Methods	Well done, no other comments	Thank you
Peer Reviewer #3	Methods	Excellent.	Thank you
Peer Reviewer #4	Methods	Overall, I think the methods were reasonable, but it is hard to cover so many different tests in a single review. Search strategies seem ok, eligibility appear appropriate for the questions asked, risk of bias and strength of evidence analysis look reasonable. Because prognostic evaluation will typically be in the context of therapy, the information gained is to some extent predictive, also different than the response to drugs that are directly targeted at a particular genomic change. Thus, the authors should mention the treatment setting in which prognosis has been evaluated.	We agree that it is challenging to cover so many tests in one review. The focus of the review per CMS was not to assess the impact of the tests on prediction of treatment efficacy. We thus did not include studies that were assessing tests for that purpose. We have clarified at the beginning of the executive summary (ES – Page 1 Para 1) and the main report (page 6 – 1st paragraph under Scope & Objectives of the Review) that the report is focusing on prognostic value of the tests, and that we do not evaluate studies of predicting response to treatment. We have also included information in our risk of bias assessment of studies based on whether treatment was controlled for or not.

Peer Reviewer	Section	Comment	Response
Peer Reviewer #1	Results	The results are presented with appropriate detail. Studies are clearly described with the exception of reference 40 (see comment below). Figures, tables and appendices are extensive and provide appropriate description. No obvious ommissions or inappropriate inclusions or exclusions, although see comment below regarding use of proficiency testing.	Thank you
Peer Reviewer #1	Results	There were a few issues in the Analytic Validity Section: Page 32 Key Question 2 Analytic Validity I'm surprised that no information was provided regarding proficiency testing programs. CAP/ACMG has proficiency testing programs in place for KRAS, BRAF and EGFR mutation testing as well as MSI testing although I don't think they are available for the other tests (with the possible exception of ALK). These data are publicly available and I have seen these used to address questions of analytic validity in other systematic reviews. Given the lack of published papers (which is not unusual in the genetic testing realm) this should be considered.	We have included CAP PT results for EGFR (pg 30 – EGFR Analytic Validity last para), KRAS (pg 30, KRAS Analytic Validity for Lung Ca; pg 31 Kras Analytic Validity for CRC, Last Para), BRAF (pg 31 Braf Analytic Validity for CRC), MSI (pg 32, MSI Analytic Validity for CRC), and Urovysion (pg 33- Analytic Validity of Urovysion), in the results section on analytic validity
Peer Reviewer #1	Results	Page 45 MammaPrint: Analytic Validity I think the conclusions of this paper [ref. 40] do not support analytic validity. I don't have access to the full paper, but based on the abstract and this paragraph, there are several errors and inaccuracies. First, not all BRCA1 mutant tumors are ER negative. In fact hormone receptor status has been shown to be an independent prognostic predictor in carriers of both BRCA1 and BRCA2 mutations. [Milne RL, Antoniou AC. Genetic modifiers of cancer risk for BRCA1 and BRCA2 mutation carriers. Ann Oncol. 2011 Jan;22 Suppl 1:i11-7]. Second the finding that Mammaprint distinguishes BRCA1 mutant tumors has not been replicated to my knowledge. The company does not assert this claim nor is this part of the FDA clearance. As such I don't think this article should be used to assert analytic validity.	Per the reviewer's comment, we deleted mention of identification of ER negative and BRCA1 mutant tumors by MammaPrint. We added the Milne et al 2011 article to our full-text literature review and excluded it from this report for wrong study design.

Peer Reviewer	Section	Comment	Response
Peer Reviewer #1	Results	Page 47 Oncotype DX Colon: Analytic Validity The statement is made, “The Oncotype DX test only indirectly determines the genetic mutations associated with CRC by measuring the expression levels of specific cancer-related genes.” This statement is not accurate. The intent of the Oncotype DX test is to measure gene expression, NOT to infer underlying gene mutation status. The differential gene expression pattern is the prognostic test. Analytic validity of the test would be to see if Oncotype DX produces the same expression pattern when applied to the same tumor multiple times. Since expression can vary from tumor to tumor and because different expression assays study different genes, it is difficult, if not impossible to have reference samples for proficiency testing.	We have corrected the statement.
Peer Reviewer #2	Results	See general comments about the clinical utility scenario, otherwise no comments--excellent.	Thank you
Peer Reviewer #3	Results	Excellent	Thank you

Peer Reviewer	Section	Comment	Response
Peer Reviewer #4	Results	See attached comments with respect to OncotypeDx breast. I think the authors found and presented or cited most of the relevant articles. Because prognostic evaluation will typically be in the context of therapy, the information gained is to some extent predictive, also different than the response to drugs that are directly targeted at a particular genomic change. Thus, the authors should mention the treatment setting in which prognosis has been evaluated.	We have clarified that the report is focusing on prognostic value of the tests (ES – 1 Ist Paragraph, Main Report page 6 – 1st paragraph under Scope & Objectives of the Review). We have also included information in our risk of bias assessment of studies about whether treatment was controlled for or not. However, per CMS the focus of the review was not to assess the impact of the tests on prediction of treatment efficacy. We thus did not include studies that were assessing tests for that purpose.

Peer Reviewer	Section	Comment	Response
Peer Reviewer #4	Results	In assessing analytic validity, the parameters evaluated by the authors include sensitivity and specificity, as well as positive and negative predictive values (e.g. p12). These latter 2 measures seem not appropriate to the analysis of analytic validity as they incorporate Bayesian concepts with respect to patient population, and in fact relate more to clinical validity. In addition, the authors' analysis is too limited, as they appear not to directly consider parameters normally used to evaluate analytic validity such as accuracy and precision, which are required by CLIA. In its recommendation regarding expression profiling in breast cancer, the EGAPP Working Group in addition to sensitivity and specificity mentions "reproducibility, robustness (e.g. resistance to small changes in preanalytic or analytic variables) and quality control."	Our calculations of NPV and PPV were based on the test's ability to obtain the true value, usually estimated with Sanger sequencing, and did not take population prevalence of mutations into account. Accuracy cannot be calculated if the true value is not known exactly. Precision is an intra- and inter- lab measurement, and was only done in one study for the mutations (Cronin et al 2007) and in one study for Mammaprint (Ach et al).

Peer Reviewer	Section	Comment	Response
Peer Reviewer #4	Results	Further, I think there are accurate and representative ways to assess analytic validity in the absence of published analyte-specific data. For example, there are extensive data on the performance characteristics of common techniques such as PCR and sequencing, Pyrosequencing, or real-time PCR for assays comparable/analogous to those in question. The lack of data in many ways is a reflection of the routine use and vast experience of these techniques for testing of common biomarkers. This lack of novelty renders publication unattractive and/or difficult without inclusion of new or novel assays, techniques, or analytes. For most, the analytic performance of genotyping and related assays that are routinely performed is likely to be very high. Performance on CAP proficiency testing challenges, which we have included in EGAPP Working Group recommendations in the past, e.g., KRAS and BRAF testing in colon cancer, are useful, albeit by no means definitive, data in this regard. Moreover, the CAP participant summary responses provide numeric data on the methods in actual use in clinical laboratories. This would better allow the authors to gear the discussion of analytic validity to the methods and tests laboratories use, and their relative frequencies.	We have included CAP PT data in the section on analytic validity. CAP PT results for EGFR (pg 30 – EGFR Analytic Validity last para), KRAS (pg 30, KRAS Analytic Validity for Lung Ca; pg 31 Kras Analytic Validity for CRC, Last Para), BRAF (pg 31 Braf Analytic Validity for CRC), MSI (pg 32, MSI Analytic Validity for CRC), and Urovysion (pg 33, Analytic validity of Urovysion), in the results section on analytic validity

Peer Reviewer	Section	Comment	Response
Peer Reviewer #4	Results	The authors also failed to address 2 substantial issues of concern implicated in somatic testing. The first deals with enrichment of the specimen typically through microdissection, in the context of the lower limit of detection of the assay. For Sanger sequencing, the lower level of detection is approximately 20% allele proportion (40% malignant cells for a heterozygous mutation), but mutations can be detected at lower allele proportions, or may occasionally may be missed at higher allele proportions. In some respects, this depends on optimization of the assay, but Phred- like quality scores of 50 or greater are readily achievable out of tissue and blood. Paraffin can be somewhat more challenging for Sanger sequencing assays, requiring PCR amplicons to be of reduced size because of DNA shearing (PCR is always performed prior to Sanger Sequencing), and the potential for artifacts. Pyrosequencing can achieve lower limits of detection as low as 5 – 10% allele proportion, and real - time PCR 2-5% or better. However, the point is that the percentage of tumor cells can be important for determining the accuracy of the result, and therefore impacts clinical validity. Second, tumor heterogeneity is a problem that has not been well explored, but seems particularly relevant for gene expression profiling assays (but could also be important for some somatic mutation detection). If you take 2 sections from the same tumor will both give you the same riskcore or classification? These data are lacking.	We agree that tumor heterogeneity is a problem but found no studies that addressed this. We have added text to reflect this gap.
Peer Reviewer #1	Discussion/Conclusion	The implications of the review are clearly stated. No important literature is omitted. There are a couple of issues with the limitations that could be addressed (these were also mentioned in the general comments).	Thank you

Peer Reviewer	Section	Comment	Response
Peer Reviewer #1	Discussion/Conclusion	Page 18 Limitations Another limitation of the study that needs to be addressed is that the review is supposed to focus on Medicare beneficiaries (most of whom are over the age of 65) [Page 20 lines 19-21 see next comment]. However, the eligible studies include all adults over the age of 18. I don't see that analysis was stratified by age. Prognostic factors may be different for younger patients compared to those over age 65.	We agree. We have added this to the limitations.
Peer Reviewer #1	Discussion/Conclusion	Page 20 lines 19-21 This is somewhat confusing. The assertion is that the population of interest is that of the Medicare beneficiary, however the question is examined from the perspective of adult patients. The authors note that most of the studies do not present information about the subjects' age other than some information about means/medians. Nowhere in the conclusion do the authors come back to this original question and state that the review cannot answer the KQs for Medicare beneficiaries specifically. This should also be presented as a limitation.	We agree. We have added this to the limitations.
Peer Reviewer #2	Discussion/Conclusion	See general comments about clinical utility scenario, otherwise no additional comments.	Thank you
Peer Reviewer #3	Discussion/Conclusion	Excellent	Thank you
Peer Reviewer #4	Discussion/Conclusion	I think the implications of the findings are adequately stated. However, in order to avoid potentially harmful confusion, I think the authors need to prominently state the limitations of the scope of the study, including that report focused on a use (prognosis) for which most of the tests are neither used nor recommended.	We have added more text to clarify this in the ES -1 Paragraph 1 and Main Report – Page 6 under Scope & Objectives.

Peer Reviewer	Section	Comment	Response
Peer Reviewer #4	Discussion/Conclusion	Although I do not disagree with the authors' conclusions with respect to prognosis and clinical utility for the 3 gene expression profiling tests, which are in the main not inconsistent with our recommendation of expression profiling in breast cancer, and our internal analysis of expression profiling in colon cancer, I believe that the authors should specifically mention evidence of the ability of the OncotypeDx breast test to predict response to non-targeted therapy. In particular, the authors should discuss the study by Paik et al. in the August 10, 2006 issue of the Journal of Clinical Oncology, cited as reference 110, page 183 that provides evidence of benefit from chemotherapy added to tamoxifen in patients with early stage ER+ breast cancer having a high OncotypeDx risk score vs. lack of benefit from chemotherapy in patients with low risk scores, as well as the study of Albain et al. in Lancet Oncology 2010;11:55-65, cited as reference 7 on page 148. Given the large volume use of OncotypeDx breast and support of ASCO, I think it is important to place these studies in proper context in light of the author's conclusions. On the EGAPP Working Group we found insufficient evidence for clinical utility of OncotypeDx breast, largely because of an inability to assess the harms to the apparently small numbers of patients with low risk scores who, if untreated, would otherwise develop distant metastasis. Finally, the OncotypeDx breast risk score is heavily dependent on ER, PR, and HER2 mRNA expression, along with a number of proliferative markers. Recent data suggest that comparable information could also be gained by immunohistochemistry (see e.g. NICE report from 2013). This alternative approach could be mentioned for completeness.	The focus of this report is the prognostic efficacy of the tests. We have revised the report to make it clearer that we do not address prediction of response to therapy (ES-1 para 1; Main Report pg 6 under Scope & Objectives).
Peer Reviewer #1	Clarity/Usability	The report is for the most part clear. Organization is fine. Given that the ultimate conclusion is that there is no evidence of clinical utility, it is difficult to know how this report will ultimately be used. A couple of issues related to clarity.	Thank you

Peer Reviewer	Section	Comment	Response
Peer Reviewer #1	Clarity/Usability	I don't think the title accurately reflects the review. The evidence review is not looking at just recurrent cancer. It is also focused on both prognosis and clinician decision making. I would propose a title something to the effect of: Technology Assessment of the Impact of Prognostic Genetic Testing on patient outcomes and clinician decision-making in adults with cancer.	We have changed the title.
Peer Reviewer #1	Clarity/Usability	Of particular concern is that, as mentioned above, several of these tests are routinely performed to determine choice of treatment, so even though there isn't evidence of clinical utility for prognosis, there is evidence of clinical utility for selection of treatment. I think that it needs to be made crystal clear, probably in the abstract and certainly in the executive summary that the report does not make any assertions about use of some tests for selection of therapy.	Question 4a addresses the impact on decision making (which includes impact on treatment decisions).
Peer Reviewer #2	Clarity/Usability	Very well done. Very useful and important review.	Thank you
Peer Reviewer #3	Clarity/Usability	Excellent	Thank you
Peer Reviewer #4	Clarity/Usability	Structure is ok.	Thank you

Peer Reviewer	Section	Comment	Response
Peer Reviewer #1	General	The report is clinically meaningful. The key questions are appropriate and are explicitly stated. A couple of comments about the target population that could be clarified. Page 18 Limitations Another limitation of the study that needs to be addressed is that the review is supposed to focus on Medicare beneficiaries (most of whom are over the age of 65) [Page 20 lines 19-21 see next comment]. However, the eligible studies include all adults over the age of 18. I don't see that analysis was stratified by age. Prognostic factors may be different for younger patients compared to those over age 65. Page 20 lines 19-21 This is somewhat confusing. The assertion is that the population of interest is that of the Medicare beneficiary, however the question is examined from the perspective of adult patients. The authors note that most of the studies do not present information about the subjects' age other than some information about means/medians. Nowhere in the conclusion do the authors come back to this original question and state that the review cannot answer the KQs for Medicare beneficiaries specifically. This should also be presented as a limitation.	Thank you We have modified text in the limitations section in the abstract, ES and the main report..
Peer Reviewer #1	General	Minor issues: Page 47 line 14 There is a formatting issue for Key Question 3 Clinical Validity. Should be bolded and start a new section. Page 47 line 31 Sentence beginning “Age of the women…” I think Age should be preceded by mean or median.	Thank you. Have corrected these.

Peer Reviewer	Section	Comment	Response
Peer Reviewer #2	General	This report is very well done and of great clinical importance. However, in my assessment, there is a critical missing element in the clinical scenario, analytic framework and key questions that impacts the overall report. The outcomes are addressed in terms of prognosis and survival, but the most common case for the clinical use of these tests is in predicting treatment response to specific agents. Specifically, clinicians use the results to NOT provide chemotherapeutic interventions to patients who have a low likelihood of response. The clinical benefit, then, is from avoiding the adverse effects associated with treatment with toxic drugs that will not improve survival. As such, one would not expect to necessarily see any significant impact on overall (OS) or cancer-specific survival (CSS), but in quality of life metrics, which do not appear to be mentioned in the review. Also, this use case represents an economic driver for the use of the tests, which also is not mentioned (though I have less concern about this issue). Knowing a bit about the evidence from other review efforts, I recognize that the evidence on the health benefits associated with withholding unuseful therapies is very sparse and the overall conclusions are not likely to change. But the use of KRAS mutation analysis in metastatic CRC to support decisions about cetuximab or panitumumab therapy is, according to other reviews, well-supported by evidence, leading me to have a concern with this one conclusion. I didn't have problems with the conclusions about KRAS/mCRC in terms of risk of recurrence, OS or CSS (though I found them interesting). Nor did not have trouble with any other conclusion. But Oncotype DX use is primarily being promoted to identify patients who should not be offered chemo due to differential risk of recurrence. This aspect of clinical utility is not captured in this review.	Thank you But, the outcomes addressed also include the impact on treatment decisions (KQ 4a). We did include quality of life metrics also, but the review is based on published results, and (although very important) studies generally have not reported quality of life outcomes. It is correct that this report does not address the use of the tests for the purpose of predicting response to treatment, and we mention that in several places in the report.

Peer Reviewer	Section	Comment	Response
Peer Reviewer #2	General	It appears to me that this clinical scenario was not necessarily the subject of the review, which again, I would not have a problem with, except that I feel strongly that this omission should be acknowledged and specifically outlined in the introduction and conclusions. With such a discussion, I think the review can stand as otherwise written, and is well-done.	Correct, it was not the subject of the review. We explain that in the report in several places, and have made edits to make that more clear.
Peer Reviewer #3	General	The study uses hazard ratios as its measure of prognostic accuracy. This idea that hazard ratios measure predictive accuracy is not really correct. A hazard ratio may be associated with an outcome; for example, a higher hazard ratio may be associated with a worse prognosis. But this association is not a measure of predictive accuracy. Accuracy is not a population distribution-related measurement; it is the association between each patient's predictor value for the outcome and each patient's true outcome. Thus, percent correct, sensitivity and specificity, and receiver operating characteristic are measures of predictive accuracy. I would suggest you change the idea that hazard ratios measure predictive accuracy to something like they are significantly associated with outcomes.	We agree that HRs do not measure predictive accuracy. The report was focused on assessing the prognostic value of tests. We have changed the text where we suggested that we were using HRs to measure predictive accuracy.
Peer Reviewer #4	General	The report has limited clinical value because it analyzed 11 tests for the clinical validity and utility of prognostic applications of the assays. However, most are not used in significant numbers as prognostic assays, but rather are primarily predictive (of targeted drug response) or diagnostic assays. Other for than the three expression profiling assays, and possibly microsatellite instability, there doesn't seem to have been much reason to undertake the analysis.	It is correct that this report does not address the use of the tests for the purpose of predicting response to treatment, and we mention that in several places in the report.

Peer Reviewer	Section	Comment	Response
Peer Reviewer #4	General	In the main, the report addressed 11 tests to assess the clinical utility of performing them for prognostic purposes. Eight of the tests are chiefly used for direct prediction of the response to targeted therapies, diagnosis, or detection. These tests are in widespread use and offered by many providers in the U.S. The other 3 tests are sole source, commercial tests that are sold with the expectation that the prognostic information they provide (in the case of the OncoTypeDx assays, a recurrence "risk score") will be used to inform decisions about the use of adjuvant chemotherapy in non-metastatic breast and colon cancer.	It is correct that this report does not address the use of the tests for the purpose of predicting response to treatment, and we mention that in several places in the report.
Peer Reviewer #4	General	Use of the OncoTypeDx breast test has received support from ASCO, and has considerable volumes nationally, although I think the test is still controversial. Our EGAPP Working Group recommendation was that there was insufficient evidence to recommend for or against, largely because of the lack of data assessing harms to women with low risk scores who do not receive adjuvant chemotherapy, but otherwise would have. I think these data still stand. There are 2 randomized controlled trials underway involving the test.	Thank you. We will add this point to our report.
Peer Reviewer #4	General	Mammaprint is a similar test, but with less data and somewhat different utilization. To my knowledge, it hasn't gained nearly the acceptance of OncoTyoeDx breast, although ironically it has been cleared by FDA whereas OncoType Breast has not.	The evidence regarding Mammaprint affecting treatment decisions is limited.
Peer Reviewer #4	General	The Oregon group did a systematic review on OncotypeDx colon within the last couple of years, and found little data to support its use. I don't think there have been significant changes to this. As far as I know it isn't widely used.	Thank you. There are a couple of publications that suggest that the test reduces intensity of treatment. These are included in the revised report.

Peer Reviewer	Section	Comment	Response
Peer Reviewer #4	General	However, I believe that because of the risk of misinterpretation of this review the authors should state more prominently in the abstract, summary materials, key questions, and throughout the document that the study only addressed prognostic uses of these assays in non-metastatic cases, did not consider the uses of the tests for prediction of response to targeted therapies, and that prediction of response to targeted therapies, diagnosis, or detection are the primary or for most the sole recommended uses for these tests.	We have clarified.
Peer Reviewer #4	General	The 3 proprietary tests present somewhat different issues. These assays have applied statistical calculations to combinations mRNA levels of multiple genes in order to establish an association with disease prognosis. The intended of this prognostic information is to “stratify” patients into risk groups, in order to influence treatment decisions. Thus, in contrast to tests like KRAS or ALK, which have a direct predictive, and pathophysiologically-based relationship to an applied therapy in their standard use, these tests rely on general prognosis to influence the decision to treat or not treat with non-targeted, broader spectrum therapeutic agents (The analogy would be if the prognostic information derived from the 8 widely available tests were offered for a decision to use or not use broad, non-targeted chemotherapeutic regimens in a given setting, a situation as previously described that does not reflect the current standard of care.)	We agree with the comment and have added text to differentiate the two types of tests.
Peer Reviewer #4	General	Because prognostic evaluation will typically be in the context of therapy, the information gained is to some extent predictive, also different than the response to drugs that are directly targeted at a particular genomic change. Thus, the authors should mention the treatment setting in which prognosis has been evaluated.	We have clarified.

Peer Reviewer	Section	Comment	Response
Peer Reviewer #4	General	It was refreshing to see the term “molecular genetic pathology” used to describe the laboratory in which testing for many assays takes place, given the American Board of Pathology subspecialty certification that defines the field (it could also be extended to include ALK). With that in mind, and with the understanding that all the tests described in the report involve nucleic acid analytes, I believe the use of the term “genetic tests” to describe most of the assays in the report is potentially confusing and somewhat misleading from a medical standpoint, because almost all are unrelated to the practice of medical genetics. The described tests are molecular oncology/molecular pathology assays, which, in the context of this particular report, are used by oncologists and pathologists in the management of cancer cases, not geneticists diagnosing or counseling patients or family members about inherited diseases. Although a few of these tests are in practice used to diagnose Lynch Syndrome (microsatellite instability testing, BRAF, and MLH1 promoter methylation), which predisposes to heritable colon and other cancers, even in this context the tests described are typically part of the pathologic evaluation of the specimen, and in any event in this review are considered for their prognostic impact, not a predisposition to inherited disease. This distinction is significant from an evidence standpoint because although analytic evaluation will have similarities based on particular types of analytes or assay techniques, clinical utility of oncologic biomarkers is primary evaluated based on their medical uses. In contrast to tests for diseases usually thought of as falling within the field of genetics, most of which are performed to assist the diagnosis of rare germline disorders, oncologic biomarkers are normally used for therapeutic or other management decisions and should typically be approached utilizing a common framework, irrespective of the analyte or technique used to perform a given assay. The authors themselves illustrated the weakness and potential confusion in applying the term “genetic test” to molecular oncology assays by seemingly excluding the gene expression profiling assay for colon cancer from the category, stating,“The Oncotype DX test only indirectly determines the geneticmutationsassociatedwithCRCbymeasuringthe expressionlevelsofspecific cancer-related genes. Thus, this test has little direct analytic value as a genetic test for CRC.”	Based on this and other comments, we have revised the Title of the Report.

Table 2. Public Reviewer Comments

Public Reviewer & Affiliation	Section	Comment	Response
Bastiaan van der Bann Agendia, Inc.	Executive Summary	(Discussion) Please add remark on MammaPrint prospective outcome data "with the notable exception of the Oncotype DX assay in breast cancer, which does have a sizeable body of evidence to suggest an effect on treatment decisions, though not yet an effect on downstream outcomes." Please add to the discussion: MammaPrint is the only assay that has impact data with downstream outcome and showed that withholding chemotherapy in MammaPrint Low Risk patients did not compromise outcome.	We respectfully disagree from adding text suggested by Agendia. The updated search has provided more evidence re KQ4a and we have modified the report as appropriate in
Bastiaan van der Bann Agendia, Inc.	Executive Summary	(Conclusions) Based on our comments on Drukker et al 2013 please add And moderate prospective evidence that MammaPrint improves health outcome by safely forgoing chemotherapy based on the assay result.	The Drukker et al 2013 article was reviewed as part of our draft report and excluded for ineligible outcome on pg. B-21
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Introduction	Clinical biomarkers need to be evaluated within the context of the applicable disease state and the intent and purpose of the marker. As the use of clinical biomarkers increases, there is a need to clearly define appropriate analytical and clinical validation for results that are used for treatment decisions and in treatment guidelines.[1-3] Tests must be fit for purpose with evidence relevant to that specific purpose. Consistent results across multiple well-designed studies should provide evidence for analytic validity, clinical validity, and clinical utility. We believe that evidence in the literature and public domain clearly supports the conclusions that 1) the Onco*type* DX Breast Cancer Assay for invasive breast cancer is analytically and clinically valid, adds to established clinical and pathological prognostic features, and changes treatment decisions for newly diagnosed estrogen receptor (ER)-positive early stage invasive breast cancer and 2) the Onco*type* DX Colon Cancer Assay is analytically and clinically valid, adds to established clinical and pathological prognostic features, and changes treatment decisions for patients with early stage colon cancer. The key considerations for the validation and utilization of	We thank Genomic Health for providing these references. The 'level of evidence' guidelines provided in the paper by Simon et al (2007) are similar in many ways to the methods used by our Evidence Based Practice Center for Technology Assessments. We followed the Methods guidance for the Evidence-based Practice Centers, and included approaches used for EGAPP. These are well-established methods that national and international methodologists have

Public Reviewer & Affiliation	Section	Comment	Response
		biomarkers are now well-established in the peer-reviewed literature. Per Drs. Richard Simon (National Cancer Institute), Soon Paik (National Surgical Adjuvant Breast and Bowel Project) and Daniel Hayes (University of Michigan),[1] the three requirements for clinical acceptance of a tumor marker include: 1) "the specific setting and utility of the marker must be clear, 2) the magnitude in either outcomes or treatment effects between those patients who are "positive" for a marker must be sufficiently different from those who are "negative" for that marker that the clinician and/or patient would accept different treatment strategies for the two patients, and 3) the estimates of that magnitude must be reliable." This article (Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst. 2009;101(21):1446-52) will be cited several times throughout our reply; a full copy is attached (open access-http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2782246/). As reported through multiple peer reviewed publications, the Onco*type* DX Breast and Colon Cancer Assays meet/exceed these requirements and, as a result, the assays have helped inform treatment for over 400,000 patients.[4] However, the current technological assessment has an inadequate description of the "specific setting" in which either assay is used and thus poorly describes the utility, provides superficial reporting of isolated measures of each assay's "magnitude" of impact without demonstrating true understanding of the measures of impact, and makes several errors with respect to description of analytic validity that impugns the expertise available to critically evaluate relevant literature. Currently, the literature supports the rigorous assessment of biomarkers through the evaluation of analytic validity, clinical validity, and clinical utility. In order to ensure a common	developed. It is our assessment that the published evidence with low or medium risk of bias suggests that Oncotype Dx breast and colon are clinically valid tests. The published evidence on the clinical utility in terms of changing treatment decisions for breast cancer was updated based on our updated search. This essentially describes the ACCE model, which

Public Reviewer & Affiliation	Section	Comment	Response
		language through which to communicate our response to the AHRQ technological assessment, we will briefly define each and provide specific feedback as to how each was approached by the current technological assessment. Analytical validity is defined as the reproducibility, repeatability, and accuracy of an assay; in this case, the ability of the Onco*type* DX assay to accurately and reliably measure levels of mRNA. The analytical validity of a new test is usually determined through measurement comparisons with a "standard reference" or "gold standard." For novel gene expression tests, there is often no gold standard, and the assessment of analytical validity focuses on test variability, analytic precision, and reliability, as well as the reproducibility of patient classifications into clinically relevant risk groupings. Thus, a technological assessment of a gene expression assay requires a different approach than tests that detect mutations (as present or absent). There is no evidence that the basic difference between expression-based assays and mutation detection assays were understood or incorporated into the current technological assessment. We will provide references to the publications reporting the analytic validation of both the breast and colon assays and encourage AHRQ to seek appropriate expertise in evaluating the merits of the publications. Clinical validity is the degree to which a test accurately correlates to a clinical (as opposed to a laboratory) variable. The Onco*type* DX assays predict the risk of a clinical patient outcome (e.g. cancer recurrence). Because expression is not dichotomous, gene expression based assays offer a continuous measure that is correlated with outcomes using different statistical approaches. We agree with the AHRQ assessment that, in order to be most useful, the evaluation of a biomarker's clinical validation needs to include multivariate analysis that incorporates accepted clinical and pathological features commonly used to assess prognosis. The most appropriate methods for univariate and multivariate analysis of novel biomarkers and the most appropriate covariates to include is often a subject of much debate and requires	we used in this review. We have reviewed and included Cronin et al (2007) in the review for breast cancer. We have also included the Clark-Lagone paper in the review re colon cancer. We would also like to point out that all manufacturers / companies providing these tests were contacted by AHRQ with a request to provide information regarding the analytic validity of the tests. All publications we received were subjected to the review process

Public Reviewer & Affiliation	Section	Comment	Response
		disease expertise to understand the implications of the approach chosen and the included or excluded covariates. There is little evidence in the current AHRQ assessment that sufficient expertise was available and integrated into the evaluation of the different tests, including the Onco*type* DX Breast and Colon Cancer assays. We encourage the AHRQ to engage with disease and statistical experts who are able to review and interpret the literature accurately. Clinical utility studies are a way of measuring the net benefits of a test, incorporating aspects of analytic validity, clinical validity, and usefulness of the test in clinical practice. In this submission, clinical utility is demonstrated using: • studies showing the ability of the Onco*type* DX assay to prospectively change treatment management (decision impact studies), and • studies reporting quality of life changes directly as result of knowledge of the test score (e.g., reduced patient anxiety, decisional conflict and other quality of life measures) or indirectly though changes in the use of chemotherapy (and consequent changes in quality of life). While we acknowledge that an ideal clinical utility study would prospectively assess the impact of an assay on overall survival, disease-specific survival, or recurrence rate, the current technological assessments strict focus on prognosis (without allowing the evaluation of studies supporting the predictive strength of assays) and restriction to non-metastatic disease, places an unrealistic burden of evidence, as is underscored by the lack of any study for any of the 11 tests found to address KQ1. In fact, it is common and accepted practice that by understanding the disease context, clinically validating a biomarker with the specific clinical endpoints of interest (Overall Survival, Disease-Specific Survival, and Recurrence), and demonstrating that the biomarker can impact treatment choice (KQ4), there is often an obvious link to these clinical endpoints. In some cases,	outlined and included or excluded based on our review criteria.

Public Reviewer & Affiliation	Section	Comment	Response
		the trials required to demonstrate a direct impact on the key clinical endpoints are either not feasible due to the extremely large size of trial required or unethical due to the accepted standards in clinical practice. Before the availability of molecular assays to guide treatment decisions, oncologists had only general clinical and pathologic tumor features (e.g. age, stage, grade) as guidance for the choice of adjuvant chemotherapy treatment decisions. Often, the risk of recurrence is over- or underestimated when only clinicopathological factors are used in the risk assessment. Given the range of uncertainty these indices provided, past guidelines could only recommend that physicians discuss the potential benefits and risks of adjuvant chemotherapy, endocrine therapy, and/or biologic therapy in systemic disease. Multi-gene panel tests have been developed to meet the clinical need to 1) better and more accurately quantify recurrence risk and 2) predict benefit from therapies, specifically chemotherapy, for which the risk of recurrence must be weighed against the risk of overtreatment with no potential benefit. By identifying patients who are at a lower risk of distant recurrence and will be unlikely to benefit significantly from chemotherapy, a treatment plan that does not include chemotherapy reduces potential harms, including adverse events, while reducing the costs of chemotherapy, administration, and adverse event costs. The net therapeutic benefit ratio is raised when only patients who would clearly benefit from chemotherapy are directed towards chemotherapy. We found that further studies meet AHRQ's criteria and should be included within this review. Our feedback will be grouped to address the specific key question (KQ) by assay (Onco*type* DX Breast Cancer Assay and Onco*type* DX Colon Cancer Assay). In addition to the recommendations above, we further recommend: • <u>Analytic Validation:</u> Review and inclusion of additional relevant studies for both Onco*type* DX Breast and Colon Cancer assays, and inclusion of formalin fixed paraffin embedded (FFPE) analytical	We have reviewed and included Cronin et al (2007) in the review for

Public Reviewer & Affiliation	Section	Comment	Response
		validation studies for FFPE-based assays. These studies were inappropriately omitted in the initial review, as they directly address the analytical validity topic that was included in the initial statement of work. • Clinical Validation: Review and inclusion of additional relevant studies for both the Oncotype DX Breast and Colon Cancer Assays and consideration of predictive validation studies to support clinical evidence. Treatment Decision: Review and inclusion of additional relevant studies for both the Oncotype DX Breast and Colon Cancer assays.	breast cancer. We have also included the Clark-Lagone paper in the review re colon cancer. We would also like to point out that all manufacturers / companies providing these tests were contacted by AHRQ with a request to provide information regarding the analytic validity of the tests. All publications we received were subjected to the review process outlined and included or excluded based on our review criteria.
Bastiaan van der Baan Agendia, Inc.	Results	Please adjust the intended use of MammaPrint as a prognostic marker for distant metastasis in early stage breast cancer MammaPrint was not designed or intended to determine BRCA1 mutation status. Please amend the language on Analytical validity to include: MammaPrint® is an FDA cleared gene expression test for women with ER---positive or ER---negative, lymph node---negative or LN+0---3 nodes breast cancer. MP was developed by analyzing all 25,000 genes in the human genome in 78 early stage breast tumors of known outcome. The unbiased selection of the best prognosis reporter genes from the full complement of genes in the human genome yielded a gene signature representative of all the major pathways in cancer metastasis (van 't Veer, Nature 2002, Tian, 2010) MammaPrint (MP) was the first FDA cleared assay predicting clinical outcome for breast cancer patients. Relative precision is 98.7% with median standard deviation of 0.021. Reproducibility is 98.6% (median of the three control samples (0.023). Inter---	We have amended the text to clarify that Mammaprint was intended to predict risk of distant metastasis. We have included the van't Veer (2002). There were

Public Reviewer & Affiliation	Section	Comment	Response
		laboratory agreement of a series of 100 specimens tested independently in Agendia's two CLIA certified laboratories in The Netherlands and the USA was 100%. Positive predictive value (PPV) at 5 years was 0.22 (0.16---0.28) and the negative predictive value (NPV) at 5 years was 0.95 (0.91---0.99) and at 10 years PPV is 0.29 (0.22---0.35) and NPV is 0.90 (0.85---0.96). Diagnostic validation performed according to FDA and NCCLS guidelines. Precision of the assay over a 20--- day period equaled 98.7%. Repeatability equaled 98.8%. Proved microarray using more than 400 probes for normalization and printing the gene expression signature multiple times on the same array results in a robust, reproducible and reliable diagnostic assay. (Glas, BMC Genomics 2006)	two Tian papers (2008, 2012) which were reviewed and excluded. We did not find a Tian (2010) paper in PubMed. Suspect this is a typographical error re date of publication.
Bastiaan van der Baan Agendia, Inc.	Results	Pg. 69: Please add product for evaluation Missing ColoPrint, data included in this comment letter. [An a*ttachment to comments includes a table of study characteristics for 3 studies: Salazar et al. JCO 2011; Maak et al. Annals of Surgery 2013; Kopetz et al. ASCO GI 2013 (manuscript submitted). It also includes a table of results for relapse-free survival as reported in the three studies. See Attachment 1 at the end of this document.*]	The Salazar et al 2011 study was reviewed in our draft report and excluded for evaluating a test that is not eligible for our review. Per the reviewer's comment, we added the Kopetz et. al to our full-text literature review and excluded it from this report for wrong intervention/test.
Bastiaan van der Baan Agendia, Inc.	Results	Pg. 78: Please add data Discordance with clinical risk assessment was confirmed in the MINDACT trial. Recently the patient data on the MINDACT trial was presented and confirmed the 31% discordance between clinical risk assessment and MammaPrint. Out of 6694 patients, 2142 we discordant between clinical risk assessment and MammaPrint and randomized in treatment according to clinical risk assessment or MammaPrint Rutgers et al ESMO 2013 abs	Per the reviewer's comment, we added the Rutgers et al 2013 study and excluded it from this report because it does not report an eligible outcome.
Bastiaan van der Baan Agendia, Inc.	Results	Pg 79: Please add that a prospective randomized MINDACT trial showed 31% discordance with clinical risk assessment resulting in a 17% decrease in chemotherapy in ER+	The Drukker et al 2013 article was reviewed as part of our draft report and

Public Reviewer & Affiliation	Section	Comment	Response
		patients. The prospective observational RASTER trial showed that withholding chemotherapy in Low Risk patients did not compromise outcome. Rutgers et al ESMO 2013 abs Drukker et al See our earlier comments, Drukker et al 2013.	excluded because it does not report an eligible outcome. Per the reviewer's comment, we added the Rutgers et al 2013 study and excluded it from this report because it does not report an eligible outcome.
Bastiaan van der Baan Agendia, Inc.	Results	Pg 80: Please include missing data We found one randomized trial that showed a 31% discordance with clinical risk assessment and a 17% decrease in chemotherapy in ER+ LN--- patients(Rutgers et al). We found one impact study with outcome that provides moderate evidence that using MammaPrint to safely forgo chemotherapy results in a health benefit. See earlier comments on Drukker et al 2013 and Rutgers et al	The Drukker et al 2013 article was reviewed as part of our draft report and excluded because it does not report an eligible outcome. Per the reviewer's comment, we added the Rutgers et al 2013 study and excluded it from this report because it does not report an eligible outcome.
Bastiaan van der Baan Agendia, Inc.	Results	Pg 82: MammaPrint was not designed nor claims to predict local recurrence Please take the comment on local recurrence out Missing data on OS from vd Vijver en Buyse. There is evidence regarding the clinical validity of the MammaPrint signature in terms of oucome.	Buyse was reviewed and graded as high ROB and therefore excluded from our main analyses Vd Vijiver reported unadjusted values for OS. This report focuses on the the added prognostic value of the test over and above traditional prognostic factors. Therefore we included results only from studies that looked at the impact of the test after adjusting for other prognostic factors in

Public Reviewer & Affiliation	Section	Comment	Response
			a multivariable model.
Bastiaan van der Baan Agendia, Inc.	Results	Pg 85: Drukker et al was a well conducted observational trial, it does not warrant the statement unkown consistency and imprecision Please remove “we concluded that evidence was insufficient to determine the impact of MammaPrint on treatment decisions, primarily because of unknown consistency and imprecision.” Raster was a well conducted large observational trial with outcome.	The Drukker et al 2013 article which reported on the RASTER trial was reviewed as part of our draft report and excluded because it does not report an eligible outcome.
Bastiaan van der Baan Agendia, Inc.	Results	Pg 86: Please include Drukker 2013; Study with MammaPrint reveal pattern of less aggressive treatment without compromising outcome. Low risk patients which in large majority did forgo chemotherapy had a 97% DRFI at 5 years	The Drukker et al 2013 article was reviewed as part of our draft report and excluded because it does not report an eligible outcome.
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	KQ 1. Overarching Question: Is there direct evidence that the addition of the following genetic tests used alone or in combination with traditional prognostic factors changes physician decision making and improves outcomes for adult patients with CRC, breast, lung, or bladder cancer compared with the use of traditional factors to predict risk of recurrence (RR) for adults with these cancers? We agree with the finding of the technological assessment that no such study exists for the Onco*type* DX Breast Cancer assay. However, there are specific considerations due to disease context and the intended use of the assay that make such studies either impracticable, infeasible, or unethical. The only study design that would provide entirely independent and direct evidence that a prognostic test changes physician decisions *and* improves outcomes is a randomized trial in which patients are randomized to either receive or not receive the Onco*type* DX assay, with long term follow-up (five to ten years) to show that the arm that received the assay had significantly better clinical outcomes than the arm that did not. There are two core concerns with such a study design (usually termed “marker strategy design”) that result in such	We agree with the description of the type of study that would provide direct evidence for KQ1 and that such studies are unlikely to be found due to both logistic and ethical issues the reviewer raises. This why we have not recommended that future research should include RCTs. Nevertheless, it is important for such a review to search for such evidence and to clarify when it does or does not exist. Some of the rationale provided by this reviewer for not doing such studies are assumptions.

Public Reviewer & Affiliation	Section	Comment	Response
		studies being infeasible. The first problem is that the largest clinical impact of the Onco*type* DX Breast Cancer Assay is that it reduces the use of chemotherapy. However, a patient who is spared chemotherapy would not lower her risk of recurrence beyond that of a similar patient who receives chemotherapy that provided no clinical benefit. Hence, while use of the Onco*type* DX Breast Cancer Assay has been shown in many studies to reduce the use of adjuvant chemotherapy, it is unlikely that low risk patients who use the assay will have better clinical outcomes than the low risk patient who unnecessarily received chemotherapy. However, they would be exposed to chemotherapy toxicities, multiple office visits, infusions, growth factor agents, anti-emetics and for no additional long term survival benefit. The shortcomings of the "marker strategy design" for such assays have been characterized in the peer reviewed literature as requiring a "huge sample size" and yielding results that are "unlikely to be convincing."[1] An additional contemporary problem for the Onco*type* DX assays is that testing these hypotheses would require randomization of patients to an arm that does not allow them to receive testing that is widely incorporated into all oncology practice guidelines. This may raise ethical issues for physicians and institutions that would render this study infeasible for fully informed patients based on the large body of extant evidence. When these considerations are ignored, there is a risk that health technology assessments will produce findings which, on a larger level and as a tool for action, are not clinically applicable.	While the assumptions are based on decent observational evidence, they are not proven truths (e.g., the assumption that the risk of recurrence in the two randomized groups would not differ and the assumption that using less chemotherapy would not have any detrimental effect on recurrence). KQ1 was the overarching question that framed the Analytical framework. It is used to help find alternate evidence pathways that will help provide a satisfactory albeit indirect response to KQ1. The absence of studies that would be needed to directly answer that question does not necessarily reflect badly on the clinical utility of a test. The absence of studies that have prospectively looked at the impact of the test on patient outcomes would provide evidence for KQ4b which in turn will feedback to the overarching question. However, although there are several published studies on

Public Reviewer & Affiliation	Section	Comment	Response
			impact of Oncotype Dx on decision making, there were none that look at impact on health outcomes. The accumulation of the long term outcomes on patients does not have to be through RCTs.
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	KQ1 can and should be substantially refined to be specific to assays with different intended uses, so that it provides feasible avenues for obtaining the evidence that is necessary. Specifically KQ1 should be divided into two components: • Question KQ1a: *Is there evidence that the test provides actionable information beyond that provided by traditional measures (where actionable means that clinical outcomes are sufficiently different to clearly justify different treatment strategies)?* • Question KQ1b: *Is there direct evidence that the addition of the test used alone or in combination with traditional prognostic factors changes physician decision making?* Published studies addressing question KQ1a are the clinical validation studies for the Onco*type* DX Breast Cancer Assay, which are listed in the response to KQ3. These include: Breast Clinical Validation Studies (Prognostic and Predictive): • Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004; 351(27): 2817-26. *(prognostic validation)* • Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-	Thank you for the suggestion. The KQs were arrived at through a process of consensus that included AHRQ&CMS. We have not responded to article by article to this comment. All the of the articles referred to below were all included in our review and either included or excluded per our criteria

Public Reviewer & Affiliation	Section	Comment	Response
		negative, estrogen receptor-positive breast cancer. J Clin Oncol. 2006; 24(23): 3726-34. *(predictive validation)* • Habel LA, Shak S, Jacobs MK, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. Breast Cancer Res. 2006; 8(3): R25. *(prognostic validation)* • Albain KS, Barlow WE, Shak S, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. Lancet Oncol. 2010; 11(1): 55-65. *(prognostic and predictive validation)* • Toi M, Iwata H, Yamanaka T, et al. Clinical significance of the 21-gene signature (Oncotype DX) in hormone receptor-positive early stage primary breast cancer in the Japanese population. Cancer. 2010; 116(13):3112-8. *(prognostic validation)* • Dowsett M, Cuzick J, Wale C, et al. Prediction of risk of distant recurrence using the 21-gene Recurrence Score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a transATAC study. J Clin Oncol. 2010; 28(11): 1829-34. *(prognostic validation)* Published studies addressing question KQ1b are the treatment decision impact studies, which are listed in the response to KQ4A.	
Phillip Febbo, Chief Medical Officer	Results	Breast Decision Impact Studies: • Albanell J, Gonzalez A, Ruiz-Borrego M, et al.	Comment continued from above. Refers to changing

Public Reviewer & Affiliation	Section	Comment	Response
Genomic Health, Inc		Prospective transGEICAM study of the impact of the 21-gene Recurrence Score assay and traditional clinicopathological factors on adjuvant clinical decision making in women with estrogen receptor-positive (ER+) node-negative breast cancer. Ann Oncol. 2012; 23(3): 625-31. • Davidson JA, Cromwell I, Ellard S, et al. A prospective clinical utility and pharmacoeconomic study of the impact of the 21-gene Recurrence Score assay in oestrogen receptor positive node negative breast cancer. Eur J Cancer. 2013;S0959-8049(13):00211-6. • Lo SS, Mumby PB, Norton J, et al. Prospective multicenter study of the impact of the 21-gene recurrence score assay on medical oncologist and patient adjuvant breast cancer treatment selection. J Clin Oncol. 2010; 28 (10): 1671-6. • Asad J, Jacobson AF, Estabrook A, et al. Does Oncotype DX recurrence score affect the management of patients with early-stage breast cancer? Am J Surg. 2008; 196 (4): 527-9. • Lo SS, Mumby PB, Norton J, et al. Prospective multicenter study of the impact of the 21-gene recurrence score assay on medical oncologist and patient adjuvant breast cancer treatment selection. J Clin Oncol. 2010; 28 (10): 1671-6. • Klang SH, Hammerman A, Liebermann N, et al. Economic implications of 21-gene breast cancer risk assay from the perspective of an Israeli-managed health-care organization. Value Health. 2010; 13 (4): 381-7. • Oratz R, Paul D, Cohn AL, et al. Impact of a commercial reference laboratory test recurrence	KQ1 into two sub questions. Response is in the beginning of the comment.

Public Reviewer & Affiliation	Section	Comment	Response
		score on decision making in early-stage breast cancer. J Oncol Pract. 2007; 3 (4): 182-6. • Gligorov J, Pivot XB, Naman HL, et al. Prospective study of the impact of using the 21-gene recurrence score assay on clinical decsion making in women with estrogen receptor-positive, HER2-negative, early stage breast cancer in France. Poster presented at: American Society for Clinical Oncology Annual Meeting; June 2012; Chicago, IL. • Holt S, Bertelli G, Humphreys I, et al. A decision impact, decision conflict and economic assessment of routine Oncotype DX testing of 146 women with node-negative or pNImi, ER-positive breast cancer in the UK. Br J Cancer. 2013;108(11): 2250-8. • Eiermann W, Rezai M, Kümmel S, et al. The 21-gene recurrence score assay impacts adjuvant therapy recommendations for ER-positive, node-negative and node-positive early breast cancer resulting in a risk-adapted change in chemotherapy use. Ann Oncol. 2013; 24(3): 618-24. • Bargallo JER, Lara F, Shaw Dulin RJ, et al. A study of the impact of the 21-gene breast cancer assay on the use of adjuvant chemotherapy in women with breast cancer in a Mexican public hospital. Poster presented at: European Society for Medical Oncology Congress; September 2012; Vienna, Austria. • de Boer RH, Baker C, Speakman D, et al. The impact of a genomic assay (Oncotype DX) on adjuvant treatment recommendations in early breast cancer. Med J Aus. 2013;199: 205-8. • Yamauchi H, Nakagawa C, Takei H, et al. Prospective study of the effect of the 21-gene assay	

Public Reviewer & Affiliation	Section	Comment	Response
		on adjuvant clinical decision-making in Japanese women with estrogen receptor-positive, node-negative, and node-positive breast cancer. Clin Breast Cancer. 2013 Oct 26. • Oratz R, Kim B, Chao C, et al. Physician survey of the effect of the 21-gene recurrence score assay results on treatment recommendations for patients with lymph node-positive, estrogen receptor-positive breast cancer. J Oncol Pract. 2011; 7(2): 94-9. In combination, these studies provide the body of evidence necessary to address the intent of the overarching question, without the specific wording that would require an infeasible marker strategy study design.	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	KQ 2. Analytic Validity: Does existing evidence establish the technical accuracy and reliability of the tests for detecting the relevant genetic markers? The current technological assessment fails to capture commonly accepted measures of analytic validity for expression-based assays and did not include key publications that support the analytic validity of the Onco*type* DX Breast Cancer assay (Appendix: Table 1).[5-7] *[See Attachment 3 at the end of this document.]* Analytical validation involves characterization of the accuracy, precision, and reproducibility of the assay required for its intended use. The analytical validity of Onco*type* DX has been extensively studied and reported in peer-reviewed literature. The assay was developed using FFPE tissue samples. RT-PCR was chosen as the assay system primarily due to its quantitative accuracy and precision, reproducibility, and wide dynamic range. We published detailed analyses demonstrating that this approach was highly reliable. The data showed close concordance with analyses using freshly frozen tissue. The use of reference genes was investigated and established as an effective method of normalizing differences generated in the preparation of FFPE samples, and the methods and results	We have reviewed and included the Cronin (2007) paper on analytic validity in the report. Analytic validity section on Oncotype Dx has been modified. Based on our updated search and article -(Delahaye et al., 2013) which met our review criteria.

Public Reviewer & Affiliation	Section	Comment	Response
		were reported in great detail. The range of technical feasibility studies was extensive and was designed to determine and verify the following: 1) RNA yield and the quality of RNA after extraction from FFPE tissues, 2) gene expression differences and similarities between whole section and enriched tumor tissue sections, 3) gene expression heterogeneity within tumor tissues, 4) within block and between block gene expression heterogeneity, and 5) the selection of reference genes (important for normalization of pre-analytical factors). A number of subsequent studies (Appendix: Table 1) *[See attachment 3 at the end of this document.]*continued to expand on the published evidence for the sensitivity, specificity, limits of detection and quantitation, amplification efficiency, precision and reproducibility, as well as the success rate and other measures of validation required of a modern diagnostic assay. The methodology for the Onco*type* DX assays yields precise and highly reproducible results. As required by United States Federal and State laboratory regulations (Clinical Laboratory Improvement Amendments), proficiency testing of the Onco*type* DX assay (a blinded assessment of assay performance) on repeated testing of multiple characterized patient samples is regularly performed by our laboratory and has documented the consistency and reproducibility of the assay.[8-10]	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	KQ 3. Clinical Validity: Does existing evidence establish the prognostic accuracy of the tests for predicting recurrence? The Onco*type* DX Breast Cancer Assay result (the Recurrence Score® value) has been significantly correlated with distant recurrence, breast cancer-specific survival, disease-free-survival and overall survival in six major studies.[6,7,11-14] Additionally, the Recurrence Score result alone has been shown to provide significant additional information of the risk of distant recurrence beyond that provided by traditional clinicopathological predictors	The Albain paper was reviewed and excluded because it focused on the predictive value of the test; Toi (2010) was excluded because the study design did not meet our criteria. Paik (2006) was excluded because it did not have

Public Reviewer & Affiliation	Section	Comment	Response
		(Appendix: Tables 2a and 2b*). [See Attachment 3 at the end of this document]* The current AHRQ assessment does not include the Albain (Lancet Oncol. 2010), Toi (Cancer 2010), and Paik (J Clin Oncol. 2006) references from Table 2a. This is important because the Albain et al. study demonstrated that the Recurrence Score result was both prognostic of recurrence and/or death in patients treated with endocrine therapy alone, as well as predictive of the benefit from the addition of adjuvant chemotherapy in node-positive patients, and the Toi et al. study showed (using a design similar to the Paik et. al study of NSABP B-14) that the Recurrence Score result was prognostic in a population of Japanese women, demonstrating the relevance of the tumor biology assessed by the Onco*type* DX Breast Cancer Assay beyond the largely Caucasian populations in which the assay was originally validated. Although outside of the scope of the AHRQ summary, the Onco*type* DX assay is the only multigene assay currently available that has been validated for the prediction of chemotherapy treatment benefit in patients with ER-positive node negative and node positive early-stage breast cancer. While the current technological assessment focuses strictly on prognosis, it is impossible to understand the utility and value of the Onco*type* DX Assay without accounting for the ability of Onco*type* DX assay to predict chemotherapy benefit. Studies have consistently shown that patients with high Recurrence Score disease have a substantial benefit from chemotherapy whereas patients with low Recurrence Score disease experience minimal or no benefit from chemotherapy. For patients with intermediate Recurrence Score disease, the data indicate that there is little benefit of chemotherapy, but the confidence intervals could not exclude a clinically important benefit of chemotherapy. We believe that this evidence should be included in this review, as prediction of chemotherapy benefit is a significant component of the intended purpose of the Onco*type* DX Breast Cancer Assay to optimize the treatment decision regarding adjuvant	appropriate comparators. Also as mentioned previously Simon et al This essentially describes the ACCE model, which we used in combination with published practices for Evidence based Practice Centers in this review.

Public Reviewer & Affiliation	Section	Comment	Response
		chemotherapy (Paik et al. 2006, Albain et al. 2010). Additionally, there are five supportive trials in the neoadjuvant setting for the predictive ability of the Onco*type* coDX breast cancer assay (Appendix: Table 2).[15-20] [*See Attachment 3 at the end of this document]* In particular, the evidence from the Gianni (J Clin Oncol. 2005), Chang (Breast Cancer Res Treat. 2007), and Yardley (SABCS 2011) studies supports the predictive ability for neoadjuvant chemotherapy benefit using the Recurrence Score result for women with early stage breast cancer. The test identifies patients likely to benefit from chemotherapy who would not have been identified through standard clinical practice (with a resulting indirect impact on patient survival and cost of cancer recurrence). Importantly, the test also identifies many patients who are unlikely to derive meaningful benefit from chemotherapy, thus sparing them adverse effects and risks associated with chemotherapy. The Onco*type* DX Breast Cancer assay, therefore, provides insight into adjusting treatment plans for chemotherapy (based on whether chemotherapy is likely to be beneficial or not). The body of evidence shows that the Onco*type* DX invasive breast cancer assay meets tumor marker level IB evidence for clinical use.[1] The Onco*type* DX invasive breast cancer assay predicts the 10-year risk of distant recurrence and the likelihood of chemotherapy benefit in women with ER-positive, early stage invasive breast cancer. Six prospectively-designed analyses of archived samples from randomized clinical trials and confirmatory studies in >3,800 patients support a claim of level 1B evidence,[1] and an additional seven studies in 1,904 patients provide additional evidence to support these validation studies (Appendix: Table 3). *[See Attachment 3 at the end of this document.]* As a result of depth and breadth of these results across multiple studies, the Onco*type* DX test for invasive breast cancer is the only assay incorporated into all major oncology clinical guidelines (NCCN®, ASCO®, St Gallen, ESMO®, NICE)	

[1] Level 1B evidence described as 1) Study design: prospective using archived samples and 2) Validation Studies Available: one or more with consistent results. Simon et al. J Natl Cancer Inst. 2009.

Public Reviewer & Affiliation	Section	Comment	Response
		(Appendix: Table 4).[20-24] *[See Attachment 3 at the end of this document.]* The use of fully "prospective/retrospective" studies is considered a high level of evidence, as described by Simon et al. (J Natl Cancer Inst. 2009). All of the key Onco*type* DX assay validation studies (Appendix: Tables 2a and 2b) *[See Attachment 3 at the end of this document.]* have followed this approach. This approach involves the pre-specification of hypotheses and analytical study plans using archived tumor tissues collected from previously conducted large-scale clinical trials. In this revised standard of level of evidence criteria, the authors assign a Level I evidence to those tumor markers that demonstrate consistent results across multiple clinical validation studies conducted with prospective study designs using archived tumor samples. These "prospective retrospective" methods represent a disciplined and pragmatic approach for validating tumor markers. They should clearly be distinguished from the much more common observational retrospective studies that are less rigorous, examine many variables without pre-specification, and yield only hypothesis-generating data.	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	KQ 4. Clinical Utility: Does existing evidence support clinical utility of the genetic tests? • 4a. What is the evidence that the prognostic information provided by the genetic tests modifies physician decisions regarding use of adjuvant antineoplastic chemo- and/or radiotherapy, enhanced diagnostic testing for recurrence, and/or preventive surgery among adult patients with malignant tumors? • 4b. What is the evidence that modified decisions lead to improved outcomes, including patient-centered outcomes, overall survival, and disease-free survival, or change the likelihood of serious side effects of adjuvant therapy in adult patients with malignant tumors?	Based on the updated searches and review of articles submitted in this review process, we believe that the strength of evidence for the clinical utility of Oncotype Dx is moderate;

Public Reviewer & Affiliation	Section	Comment	Response
		The Onco*type* DX Breast Cancer Assay has been established worldwide as the standard of care in the evaluation of hormone receptor positive, HER2-negative invasive breast cancer. This is evidenced by the inclusion of the assay in guidelines and consensus statements from organizations such as NCCN, ASCO, ESMO, the St. Gallen International Breast Cancer Expert Panel, and the NICE review committee (Appendix: Table 4). *[See Attachment 3 at the end of this document.]* As of September 2013, Genomic Health's assays have been used to evaluate over 400,000 patients since its commercial release in 2004, helping to determine appropriate, necessary and, frequently, less-costly therapeutic options. The Onco*type* DX Breast Cancer Assay has been shown to have two very important and complementary properties on which clinical decisions can be based: 1) prognosis for 10-year risk of distant recurrence for patients treated with adjuvant endocrine therapy, and 2) predictive value for chemotherapy benefit for patients with ER-positive, HER2-negative invasive breast cancer. Regarding decision impact studies, many studies utilizing a consistent methodology including >2,200 patients from around the world show that use of the Onco*type* DX Breast Cancer Assay yields an approximate 30% change in treatment recommendations pre- vs post-assay (Appendix: Table 5).[25-37] *[See Attachment 3 at the end of this document.]* As noted correctly in the assessment, while both chemotherapy recommendations and utilization changes occur in both directions (lower Recurrence Score results guide away from and higher results guide towards chemotherapy), the majority of treatment changes after receipt of the Recurrence Score result are from a recommendation of chemohormonal therapy to hormonal therapy alone. We believe the analysis method yields an overly conservative rating of confidence. Taken as a whole, the large body of consistent evidence provides a high strength of evidence for KQ4a.	

Public Reviewer & Affiliation	Section	Comment	Response
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	In the current technological assessment, the following studies from Table 5 were omitted: • Lo SS, Mumby PB, Norton J, et al. Prospective multicenter study of the impact of the 21-gene recurrence score assay on medical oncologist and patient adjuvant breast cancer treatment selection. J Clin Oncol. 2010; 28 (10): 1671-6. • Asad J, Jacobson AF, Estabrook A, et al. Does Oncotype DX recurrence score affect the management of patients with early-stage breast cancer? Am J Surg. 2008; 196 (4): 527-9. • Klang SH, Hammerman A, Liebermann N, et al. Economic implications of 21-gene breast cancer risk assay from the perspective of an Israeli-managed health-care organization. Value Health. 2010; 13 (4): 381-7. • Oratz R, Paul D, Cohn AL, et al. Impact of a commercial reference laboratory test recurrence score on decision making in early-stage breast cancer. J Oncol Pract. 2007; 3 (4): 182-6. • Holt S, Bertelli G, Humphreys I, et al. A decision impact, decision conflict and economic assessment of routine Oncotype DX testing of 146 women with node-negative or pNImi, ER-positive breast cancer in the UK. Br J Cancer. 2013;108(11): 2250-8. • Bargallo JER, Lara F, Shaw Dulin RJ, et al. A study of the impact of the 21-gene breast cancer assay on the use of adjuvant chemotherapy in women with breast cancer in a Mexican public hospital. Poster presented at: European Society for Medical Oncology Congress; September 2012; Vienna, Austria. • Yamauchi H, Nakagawa C, Takei H, et al.	These studies were reviewed and excluded for reasons provided in appendices

Public Reviewer & Affiliation	Section	Comment	Response
		Prospective study of the effect of the 21-gene assay on adjuvant clinical decision-making in Japanese women with estrogen receptor-positive, node-negative, and node-positive breast cancer. Clin Breast Cancer. 2013 Oct 26. Oratz R, Kim B, Chao C, et al. Physician survey of the effect of the 21-gene recurrence score assay results on treatment recommendations for patients with lymph node-positive, estrogen receptor-positive breast cancer. J Oncol Pract. 2011; 7(2): 94-9.	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	A prospective US multi-center study (Lo J Clin Oncol 2010) examined how the results of the Onco*type* DX Breast Cancer Assay influenced adjuvant treatment selection both by medical oncologists and patients.[27] In this study, the Recurrence Score result increased the medical oncologist confidence in the treatment recommendation in 76% of cases, and patients reported greater satisfaction, lower conflict with decision-making, and decreased anxiety after learning their results. The study showed that 95% of patients were glad they received the Onco*type* DX Breast Cancer Assay as part of their care, and 83% of patients indicated that the result influenced their decision-making. In a large prospective German study (Eiermann Ann Oncol 2012),[33] use of the Onco*type* DX Breast Cancer Assay increased physicians' confidence in their treatment recommendations and moderately decreased patients' decisional conflict. After use of the assay, physicians classified their confidence as absolute or high in 81.9% of patients (compared to a baseline of 54.9%), showing that physicians' confidence increased in 45% of all cases ($p<0.001$), and there was a 6.2% improvement ($p=0.028$) in patients' decisional conflict. The health economic value of the Onco*type* DX Breast Cancer Assay has been well-established across multiple markets (Appendix: Table 6).[39-53] *[See Attachment 3 at the end of this document.]* Studies have shown a net savings of up to $2,000 US dollars per patient tested with the Onco*type*	Based on the updated searches and review of articles submitted in this review process, we believe that the strength of evidence for the clinical utility of Oncotype Dx is moderate;

Public Reviewer & Affiliation	Section	Comment	Response
		DX Breast Cancer Assay.[50] A reduction in chemotherapy use of 30% results in approximately $195,000 savings per 100 patients tested annually.[52] Results demonstrating that the assay is cost-effective/cost-saving are consistent around the world, regardless of country and local cost data. Patients with lower Recurrence Score results have minimal, if any, benefit from adjuvant chemotherapy, while patients with higher results have a distinct benefit from adjuvant chemotherapy. By identifying patients with low results who are at a lower risk of distant recurrence and who are predicted to be unlikely to benefit significantly from chemotherapy, opting for a treatment plan that does not include chemotherapy, direct (chemotherapy) and indirect (adverse event management and administration resources) costs can be avoided, thus reducing the financial burden of oncology treatment. Patients with a higher result, who would clearly benefit from chemotherapy, can thus be directed to chemotherapy to reduce the likelihood of cancer recurrence. Regarding 4B, we disagree that the chain of evidence is uninterpretable. The assessors have taken the position that there is no knowledge whether giving chemotherapy to patients with high risk of recurrence has any benefit, and that this is "unknown to physicians" making the decision. Benefit of adjuvant chemotherapy can only accrue to patients who would have recurred. We believe that the discussion of Section 4B should be re-written.	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	KQ 5. What are the harms associated with treatment decisions that are informed by the genetic tests? Because the Onco*type* DX Breast Cancer Assay is performed on specimens collected for diagnostic purposes, there are no immediate and direct harms from the assay. However, there are two possible harms from a diagnostic that is used to guide therapy (such as the Onco*type* DX assay). First, there is the potential harm associated with a test resulting in a women not receiving chemotherapy who may have benefitted from the therapy. In the context of use for the Onco*type* DX Breast assay, this risk is very small. It is estimated that approximately 4% of early stage, ER-positive,	Thank you. We agree with this assessment and have added relevant information in the report (pg 97 – Discussion).

Public Reviewer & Affiliation	Section	Comment	Response
		HER2-negative, node-negative breast cancer will derive benefit from chemotherapy and the majority of women who get chemotherapy with this disease do not benefit.[54] In addition, there is strong data suggesting that the assay identifies women who specifically benefit from chemotherapy and, alternatively, those who are unlikely to benefit (KQ4a). Thus, there is a very low risk that use of the Onco*type* DX Breast Cancer Assay will cause harm by having a woman not treated with chemotherapy. The other potential harm is that a woman who would not have received treatment, and will not benefit, does receive treatment based upon the test result. Decision impact studies and clinical validation studies as described in KQ4 have demonstrated that women with low clinical risk do have tumors that will benefit from chemotherapy and there is a proportion, albeit small, who are more likely to receive treatment following use of the test. If these women do not benefit from chemotherapy, there is a risk that they suffer the side effects without potential benefit. These potential risks are balanced by the risks associated with treatment decisions for patients who do not receive the assay for clinical management. The Onco*type* DX Breast Cancer Assay Recurrence Score result has been validated for prediction of chemotherapy benefit and has been shown to reduce the use of adjuvant chemotherapy and subsequently reduce the risks associated with chemotherapy. According to one study of >12,000 women with newly diagnosed breast cancer, "chemotherapy-related serious adverse effects …may be more common than reported by large clinical trials and lead to more patient suffering and health care expenditures than previously estimated." The short-term risks of chemotherapy described in this study included fever or infection, neutropenia or thrombocytopenia, dehydration or electrolyte disorders, nausea, emesis, or diarrhea, anemia, deep venous thrombosis or pulmonary embolus, and/or malnutrition. The long-term risks of chemotherapy include cardiovascular toxicity, cognitive function, ovarian failure, and secondary malignancies.[55]	

Public Reviewer & Affiliation	Section	Comment	Response
		Use of the Onco*type* DX Breast Cancer Assay has consistently been shown to impact treatment decisions. This impact occurs both towards and away the use of chemotherapy, with a net reduction of chemotherapy recommendations and overall use. However, by identifying key patients with an anticipated chemotherapeutic benefit, physicians and patients can structure a management plan most appropriate for that individual patient.	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	KQ 1. Overarching Question: Is there direct evidence that the addition of the following genetic tests used alone or in combination with traditional prognostic factors changes physician decision making and improves outcomes for adult patients with CRC, breast, lung, or bladder cancer compared with the use of traditional factors to predict risk of recurrence (RR) for adults with these cancers? We agree with the finding of the technological assessment that no such study exists for the Onco*type* DX Colon Cancer assay. However, there are specific considerations due to disease context and the intended use of the assay that make such studies either impracticable, infeasible, or unethical. The only study design that would provide entirely independent and direct evidence that a prognostic test changes physician decisions *and* improves outcomes is a randomized trial in which patients are randomized to either receive or not receive the Onco*type* DX Colon Cancer Assay, with long term follow-up (five to ten years) to show that the arm that received the assay had significantly better clinical outcomes than the arm that did not. The concern with such a study design (usually termed "marker strategy design") results in such studies being infeasible. Testing these hypotheses would require randomization of patients to an arm that does not allow them to receive testing. This may raise ethical issues for physicians and institutions that would render this study infeasible. When these considerations are ignored, there is a risk that health technology assessments will produce findings which, on a larger level and as a tool for action, are not clinically applicable.	We agree with the description of the type of study that would provide direct evidence for KQ1 and that such studies are unlikely to be found due to both lostical and ethical issues the reviewer raises. This why we have not recommended that future research should include RCTs.Nevertheless, it is important for such a review to search for such evidence and to clarify when it does or does not exist. Some of the rationale provided by this reviewer for not doing such studies are assumptions. While the assumptions are based on decent observational evidence, they are not proven truths (e.g., the assumption that the risk of recurrence in the two randomized groups would not differ and the assumption that using less chemotherapy

Public Reviewer & Affiliation	Section	Comment	Response
			would not have any detrimental effect on recurrence). KQ1 was the overarching question that framed the Analytical framework. It is used to help find alternate evidence pathways that will help provide a satisfactory albeit indirect response to KQ1. The absence of studies that would be needed to directly answer that question does not necessarily reflect badly on the clinical utility of a test. The absence of studies that have prospectively looked at the impact of the test on patient outcomes would provide evidence for KQ4b which in turn will feedback to the overarching question. However, although there are several published studies on impact of Oncotype Dx on decision making, there were none that look at impact on health outcomes.
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	KQ1 can and should be substantially refined to be specific to assays with different intended uses, so that it provides feasible avenues for obtaining the evidence that is necessary. Specifically KQ1 should be divided into two components:	Thank you for the suggestion. The KQs were arrived at through a process of consensus that included AHRQ&CMS.

Public Reviewer & Affiliation	Section	Comment	Response
		• Question KQ1a: *Is there evidence that the test provides actionable information beyond that provided by traditional measures (where actionable means that clinical outcomes are sufficiently different to clearly justify different treatment strategies)?* • Question KQ1b: *Is there direct evidence that the addition of the test used alone or in combination with traditional prognostic factors changes physician decision making?* Published studies addressing question KQ1a are the clinical validation studies for the Onco*type* DX Colon Cancer Assay, which are listed in the response to KQ3. These include: Colon Clinical Validation Studies (Prognostic): • Gray RG, Quirke P, Handley K, et al. Validation study of a quantitative multigene reverse transcriptase–polymerase chain reaction assay for assessment of recurrence risk in patients with stage II colon cancer. J Clin Oncol. 2011;29(35):4611-9. • Venook AP, Niedzwiecki D, Lopatin M, et al. Biologic determinants of tumor recurrence in stage II colon cancer: validation study of the 12-gene recurrence score in cancer and leukemia group B (CALGB) 9581. J Clin Oncol. 2013; 31(14): 1775-81. • Yothers G, O'Connell M, Lee M, et al. Validation of the 12-gene colon cancer Recurrence Score in NSABP C-07 as a predictor of recurrence in stage II and III colon cancer patients treated with 5-FU/LV and 5-FU/LV+ oxaliplatin. J Clin Oncol. 2013 Nov 12. (electronic publication ahead of print). Published studies addressing question KQ1b are the treatment decision impact studies, which are listed in the	

Public Reviewer & Affiliation	Section	Comment	Response
		response to KQ4A.	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	Colon Decision Impact Studies: • Cartwright T, Chao C, Lee M, et al. Effect of the 12-gene colon cancer assay results on adjuvant treatment recommendations in patients with Stage II colon cancer. Curr Med Res Opin. 2013; [ePub ahead of print] • Srivastava G, Renfro LA, Behrens RJ, et al. Prospective evaluation of a 12-gene assay on treatment recommendations in patients with stage II colon cancer. Poster presented at: European Cancer Congress; September 2013; Amsterdam, Netherlands. • Brenner B, Lopatin M, Lee M, et al. Impact of the 12-gene colon cancer recurrence score assay on clinical decision-making for adjuvant therapy in stage II colon cancer patients in Israel. Poster presented at: European Cancer Congress; September 2013; Amsterdam, Netherlands. In combination, these studies provide the body of evidence necessary to address the intent of the overarching question, without the specific wording that would require an infeasible marker strategy study design.	This is a continuation of response above.
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	KQ2. Analytic Validity: Does existing evidence establish the technical accuracy and reliability of the tests for detecting the relevant genetic markers? As multi-analyte molecular diagnostic assays are becoming increasingly utilized in oncology, it is imperative that they be supported by published data on the associated analytical performance of the test. Prior to performing any clinical validation studies, the analytical performance of the Onco*type* DX Colon Cancer Assay, and its individual components, were validated in the context of meaningful pre-	We modified the relevant KQ2 sections. We directly address the reviewer's comments using our assessment of articles by Cronin and by Clark-Langone.

Public Reviewer & Affiliation	Section	Comment	Response
		defined acceptance criteria.[56] All potential sources of variability were included in the study, as to mimic the process in a commercial setting. All endpoints were successfully met, demonstrating a reliable, well-controlled process for reporting patient results. The development and analytic validation strategy for the Onco*type* DX Colon Cancer Assay was built upon the successful approach used to develop the Onco*type* DX Breast Cancer Assay.[5] This methodical and rigorous approach for developing and validating clinical assays to guide treatment decisions has been supported by leading authorities in statistics and clinical trial design. Prior to selecting genes for the Onco*type* DX Colon Cancer Assay, feasibility studies were conducted to optimize the Genomic Health platform for quantitative assessment of gene expression from FFPE colon tumor tissue. These studies in FFPE colon tumor tissue identified 1) the optimal method for reliably extracting RNA and measuring gene expression by quantitative RT-PCR technology, 2) the requirement for review of each case by a pathologist for manual microdissection to remove normal colon tissue adjacent to the tumor, and 3) reference genes for normalization of gene expression. The use of carefully selected reference genes to normalize gene expression in the context of sources of pre-analytical variability such as time of fixation and block age is a critical feature of this technology. The findings from these feasibility studies provided the technical foundation for subsequent studies.[57] Following definition of the gene list and algorithm for the assay, the process for conducting the assay, including all steps from receipt of the tumor specimen to generation of the assay result, was finalized and analytically validated prior to conducting the clinical validation study. Analytical validation of the assay ensured that the assay reports accurate, precise and reproducible results across different reagent lots, operators, patient samples, and days of the week. Amplification efficiency, linear range of expression, Limit of Quantitation (LOQ) and Limit of Detection (LOD) of individual	

Public Reviewer & Affiliation	Section	Comment	Response
		gene assays were assessed by performing a serial (15-point) dilution series of RNA, and processing through Reverse Transcription Quantitative Polymerase Chain Reaction (RTqPCR). Amplification efficiencies of the individual 12 gene assays were excellent and ranged from 96% to 107%; gene expression was found to be linear for all assays over at least an 11 $\log_2$ concentration range (2-6 to 25ng RNA input); the Limit of Detection for all assays was equivalent to signal with no sample, (40 CT) and the Limit of Quantitation (LOQ) for all assays was greater than would be expected for any patient sample at the standard RNA input. Precision and reproducibility were assessed by repeat examination of two different RNA samples through the Onco*type* DX Colon Cancer Assay RTqPCR process. The relative standard deviations (RSD) associated with each gene was very small, and well within the pre-defined acceptance criterion of 10%. The high precision of the individual gene assays translated into a similarly high level of precision for Recurrence Score (SD≤1.38). The differences in signal obtained between different robotic workstations across all 12 gene assays and the RNA samples were also extremely small (all ≤0.28 CT). In the current technical assessment, there is no meaningful discussion regarding the analytic validity of the Onco*type* DX Colon Cancer Assay that accounts for the specific analytes tested and appropriate measures of analytic validity. We believe the authors should work with investigators with appropriate technical expertise to evaluate the data supporting the analytic validity of each assay studied.	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	KQ 3. Clinical Validity: Does existing evidence establish the prognostic accuracy of the tests for predicting recurrence? Worldwide, nearly 1,200,000 new colorectal cancer cases occur annually, accounting for approximately 10% of all incident cancers.[58] In the United States, colorectal cancer is the fourth most prevalent cancer and is second only to lung cancer as a cause of cancer-related mortality.[59] Following potentially curative colon cancer resection the goal of adjuvant chemotherapy is to eradicate residual microscopic disease, thereby reducing the likelihood of	This is a long comment. The response is on the next row.

Public Reviewer & Affiliation	Section	Comment	Response
		recurrence. Fluoropyrimidine-based adjuvant chemotherapy regimens have been shown to benefit patients with stage II and III colon cancer following surgical resection. However, the absolute benefit from adjuvant chemotherapy is modest in stage II disease and may vary considerably within stage III disease.[60-61] As such, not all patients may benefit equally from chemotherapy, and it carries a significant risk of toxicities. Therefore, to assist with clinical decision-making, validated biomarkers are needed that accurately assess individual patient recurrence risk and discriminate absolute treatment benefit. The Onco*type* DX Colon Cancer Assay has been clinically validated as a predictor of recurrence risk in stage II colon cancer from prospectively-designed validation studies using archived tissue from the QUASAR, CALGB 9581, and NSABP C-07 trials.[62-64] The assay is based on an individual patient's colon tumor expression of 12 genes (seven cancer-related, five reference), which quantifies the likelihood of recurrence in early stage colon cancer following surgery. This test provides physicians with a precise genomic expression profile assessment for cancer-related genes within an individual tumor, extending beyond currently available clinical and pathological tools to quantify the risk of recurrence for each individual patient. The Recurrence Score result provides risk discrimination within stages and provides independent recurrence risk information beyond conventional factors. Use of this test allows clinicians and patients to make more informed decisions regarding adjuvant chemotherapy, which will help maximize the benefits of treatment while avoiding unnecessary risk.	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	Two studies were excluded from the assessment (Gray J Clin Oncol. 2011 and Yothers J Clin Oncol. 2013), and one was included (Venook J Clin Oncol. 2013). We will provide detail describing all three: • Gray RG, Quirke P, Handley K, et al. Validation study of a quantitative multigene reverse transcriptase–polymerase chain reaction assay for assessment of recurrence risk in patients with stage	Each of these studies was reviewed for this report. Gray et al. was excluded for ineligible outcome.

Public Reviewer & Affiliation	Section	Comment	Response
		II colon cancer. J Clin Oncol. 2011;29(35):4611-9. • Venook AP, Niedzwiecki D, Lopatin M, et al. Biologic determinants of tumor recurrence in stage II colon cancer: validation study of the 12-gene recurrence score in cancer and leukemia group B (CALGB) 9581. J Clin Oncol. 2013; 31(14): 1775-81. • Yothers G, O'Connell M, Lee M, et al. Validation of the 12-gene colon cancer Recurrence Score in NSABP C-07 as a predictor of recurrence in stage II and III colon cancer patients treated with 5-FU/LV and 5-FU/LV+ oxaliplatin. J Clin Oncol. 2013 Nov 12. (electronic publication ahead of print). The QUASAR study randomized patients to observation or adjuvant treatment with 5-fluorouracil and leucovorin (5-FU/LV) chemotherapy and demonstrated that 5-FU/LV benefit is significant but modest in stage II patients. Following development and analytical validation of the assay, clinical validation was conducted in a sample of 1,436 patients with resected stage II colon cancer from the QUASAR trial. For the primary analysis, a Cox proportional hazards regression model was fitted to the clinical endpoint of recurrence-free interval for the 711 patients who were randomized to surgery alone. The Recurrence Score results were significantly associated with risk of recurrence (p=0.004), disease-free survival (p=0.010), and overall survival (p=0.041). Kaplan-Meier estimates of recurrence risks at three years were 12%, 18%, and 22% for predefined low, intermediate, and high recurrence risk groups, respectively. In a multivariable analysis, the Recurrence Score result was a significant predictor of recurrence risk after controlling for the mismatch repair (MMR) status, T stage, tumor grade, number of nodes examined, and lymphovascular invasion. The Recurrence Score, MMR status, and T stage were the most significant independent predictors of recurrence risk following surgery. Therefore, the Recurrence Score result will have the greatest clinical utility in T3, MMR-Proficient patients, the majority of	Venook et al was included in the final report. Yothers et al did not address any of our KQs and therefore was excluded from the report.

Public Reviewer & Affiliation	Section	Comment	Response
		stage II colon cancers, where other conventional markers are not informative. CALGB 9581 was a randomized phase III clinical trial which reported no demonstrable effect of adjuvant edrecolomab (anti-EpCAM, Mab 17-1A) compared to observation in a lower risk population of patients with resected stage II colon cancer (patients with stage T4b and bowel obstruction or perforation were excluded). A prospectively-designed study using 690 patient specimens from CALGB 9581 was undertaken to confirm the findings from the QUASAR validation study. The Recurrence Score result was significantly associated with recurrence risk in univariate (p=0.013) and multivariable (p=0.004) analyses after controlling for MMR status, T stage, number of nodes examined, histologic grade, and lymphovascular invasion. NSABP C-07 was a randomized phase III clinical trial that evaluated the efficacy and safety of adding oxaliplatin to bolus 5-FU/LV chemotherapy following surgical resection in patients with stage II and III colon cancer. Prospective collection of tumor tissue from this landmark study provided the opportunity to study pathologic and molecular markers associated with outcomes in a large population of patients with stage II and III colon cancer treated with contemporary chemotherapy regimens. The prospectively-designed validation study evaluated the relationship between the continuous Recurrence Score result and recurrence risk in 892 stage II and III colon cancer patients randomized to 5-FU/LV (n=449) or 5-FU/LV plus oxaliplatin (n=443). A total of 264 (29.6%) patients had stage II disease. The study demonstrated that the continuous Recurrence Score result was a significant predictor of recurrence risk (p<0.001) in patients with stage II and stage III colon cancer treated with 5-FU/LV-based adjuvant chemotherapy after adjusting for stage and treatment. The continuous Recurrence Score result remained an independent predictor of recurrence risk (p<0.001) after controlling for the effects of N stage, treatment, MMR status, T stage, number of nodes examined, and tumor grade. With relative benefit of oxaliplatin being	

Public Reviewer & Affiliation	Section	Comment	Response
		similar across the range of Recurrence Score values, the absolute benefit of oxaliplatin increased with higher Recurrence Score results. While this study was reported at ASCO during the time interval included in the scope of the current technical assessment, the full manuscript was published after the end date of the technical assessment. Thus, while it should have been included based on the abstract and presentation, the full importance of the study would not have been appreciated with the data available to the assessors.	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	KQ 4. Clinical Utility: Does existing evidence support clinical utility of the genetic tests? • 4a. What is the evidence that the prognostic information provided by the genetic tests modifies physician decisions regarding use of adjuvant antineoplastic chemo- and/or radiotherapy, enhanced diagnostic testing for recurrence, and/or preventive surgery among adult patients with malignant tumors? • 4b. What is the evidence that modified decisions lead to improved outcomes, including patient-centered outcomes, overall survival, and disease-free survival, or change the likelihood of serious side effects of adjuvant therapy in adult patients with malignant tumors? Studies of the clinical application of the Onco*type* DX Colon Cancer Assay show that use of the assay results in meaningful changes in treatment recommendations.[65-67] The current technical assessment did not include any studies. The following studies should have been included and are discussed below: • Cartwright T, Chao C, Lee M, et al. Effect of the 12-gene colon cancer assay results on adjuvant	The studies have been reviewed and included in the updated report.

Public Reviewer & Affiliation	Section	Comment	Response
		treatment recommendations in patients with Stage II colon cancer. Curr Med Res Opin. 2013; [ePub ahead of print] • Srivastava G, Renfro LA, Behrens RJ, et al. Prospective evaluation of a 12-gene assay on treatment recommendations in patients with stage II colon cancer. Poster presented at: ASCO Gastrointestinal Symposium; January 2013; San Francisco, CA. Brenner B, Lopatin M, Lee M, et al. Impact of the 12-gene colon cancer recurrence score assay on clinical decision-making for adjuvant therapy in stage II colon cancer patients in Israel. Poster presented at: European Cancer Congress; September 2013; Amsterdam, Netherlands.	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	US medical oncologists who ordered the assay for three or more patients within the first two years of its commercial availability were asked to complete a survey regarding their most recent patient for whom the assay was ordered and report treatment recommendations before and after the assay (Cartwright Curr Med Res Opin 2013). Treatment recommendations changed in 29% of the patients suggesting that assay results impacted physicians' adjuvant treatment decisions for stage II colon cancer patients. Most changes in treatment recommendations resulted in decreases in treatment intensity. A prospective study of 187 patients with stage II colon cancer enrolled by 105 physicians across 17 academic and community practice sites within the Mayo Clinic Cancer Research Consortium was carried out to demonstrate the utility of the Onco*type* DX Colon Cancer Assay in real-world clinical practice (Srivastava ECC 2013). This study showed that the Recurrence Score result was associated with treatment changes in recommendations for 45% of T3 MMR-proficient stage II colon cancer patients compared to baseline assessments and that use of the Onco*type* DX Colon Cancer	This comment is a continuation. It describers the studies; as mentioned above, these studies have been reviewed and included in the report.

Public Reviewer & Affiliation	Section	Comment	Response
		Assay may lead to reductions in use of adjuvant chemotherapy for this subgroup. Similarly, in an Israeli population of patients with stage II colon cancer, use of the assay led to changes in treatment recommendation in 38% of cases, with most changes resulting in less treatment (Brenner ESMO 2013). Health economic analyses have concluded that the assay is projected to be cost-saving from a societal perspective. One modeling study showed that clinical use of the 12-gene assay to assess risk of recurrence in T3 stage II colon cancers with intact MMR may improve quality-adjusted life expectancy and be cost-saving from a societal perspective; there was an average increase of 0.035 QALY and average decrease of $2,971 per patient in direct medical costs.[68] A further study, using the results from a large treatment decision impact study, showed that use of the 12 gene assay increased quality-adjusted survival by 0.230 years due to avoidance of acute and long-term adverse events related to adjuvant chemotherapy. Further, use of the assay was shown to be cost-saving. On average, overall medical costs decreased $1,683 per patient, drugs and administration costs for adjuvant chemotherapy decreased by $3,978 per patient, and costs for the management of adverse events decreased by $3,168 per patient. These calculations were based off of the brand cost of oxaliplatin; savings are expected to persist even if the cost of oxaliplatin continues to decrease due to generic substitution. If the future cost of oxaliplatin is one-fourth of the current cost, use of the 12-gene assay would save on average $546 per patient.[69]	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	KQ 5. What are the harms associated with treatment decisions that are informed by the genetic tests? Because the Onco*type* DX Colon Cancer Assay is performed on specimens collected for diagnostic purposes, there are no immediate and direct harms from the assay. However, there are two possible harms from a diagnostic that is used to guide therapy (such as the Onco*type* DX Colon Cancer Assay). First, there is the potential harm associated with a test resulting in a patient not receiving chemotherapy who	Thank you for the suggestion. The KQs were arrived at through a process of consensus that included AHRQ&CMS.

Public Reviewer & Affiliation	Section	Comment	Response
		may have benefitted from the therapy. The benefit of chemotherapy within the early stage colon cohorts is uncertain and remains controversial, and use of chemotherapy varies. Thus, in the context of use for the Oncotype DX assay, this risk is very small. The other potential harm is that a woman who would not have received treatment, and will not benefit, does receive treatment based upon the test result. Decision impact studies and clinical validation studies, as described in KQ4, have demonstrated that patients with colon cancer are more likely not to receive therapy than to receive therapy if the test is used thus mitigating the potential for this risk. The harm associated with treatment decisions informed by the Oncotype DX Colon Cancer Assay is better answered by addressing the harm for patients who do not receive the assay for clinical management.. The use of chemotherapy is not without risks, and use of this test will allows clinicians and patients to make more appropriate decisions regarding adjuvant chemotherapy, which will help maximize the benefits of treatment while minimizing the risk. The primary risk associated with early stage colon cancer chemotherapy relates to neuropathy and gastrointestinal toxicities. Long-term, specific neurotoxicities remained significantly elevated for oxaliplatin-treated patients.[70-71] Use of the Oncotype DX Colon Cancer Assay has consistently been shown to impact treatment decisions. This impact occurs both towards and away the use of chemotherapy, with a net reduction of chemotherapy recommendations and overall use. By identifying key patients with higher risk of recurrence, physicians and patients can structure a management plan most appropriate for that individual patient.	
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	Results	[*The comments from Genomic Health include an appendix with the following tables*] *[See Attachment 3 at the end of this document.]* Table 1: Variability and Reproducibility, Oncotype DX Breast Cancer Assay Table 2: Oncotype DX Breast Cancer Assay Risk of Cancer	Thank you. We have reviewed the tables. All articles that did not come up in our searches or were not previously sent to us

Public Reviewer & Affiliation	Section	Comment	Response
		Recurrence (Disease-Free Survival, Recurrence-Free Survival, Distant Metastasis-Free Survival, Cancer-Specific Survival) Table 3: Clinical Evidence Studies of the Onco*type* DX Assay for Invasive Breast Cancer Table 4: Clinical Guidelines Describing Use of the Oncotype DX Assay for Invasive Breast Cancer Table 5: Decision Impact Trials Utilizing the Onco*type* DX Breast Cancer Assay Table 6: Health Economic Studies Utilizing the Onco*type* DX Breast Cancer Assay Table 7: Onco*type* DX Colon Cancer Assay Risk of Cancer Recurrence (Disease-Free Survival, Recurrence-Free Survival, Distant Metastasis-Free Survival, Cancer-Specific Survival) Table 8: Decision Impact Trials Utilizing the Onco*type* DX Colon Cancer Assay	by Genomic health were reviewed and included or excluded per our review criteria.
Molly Giammarco National Society of Genetic Counselors	Results	The National Society of Genetic Counselors (NSGC) appreciates the opportunity to comment on the Agency for Healthcare Quality & Research's (AHRQ) November 13, 2013 Technology Assessment of Genetic Testing for Risk of Recurrent Cancer draft. NSGC is the voice, authority, and advocate for over 2,800 genetic counselors, the largest group of clinical genetics care providers in the United States. NSGC recognizes genomic data sharing's potential to enhance research collaborations and improve our understanding of the contribution of variations in the human genome to health and disease states. NSGC is concerned that Medicare and third-party payers may use the term "Insufficient" that AHRQ uses to grade the strength of evidence as a reason to deny coverage for using microsatellite instability (MSI) and BRAF testing to detect Lynch syndrome. AHRQ's technical assessment did not consider these tests for the purpose of detecting Lynch syndrome. NSGC recommends that AHRQ add the following sentence to each of the sections within the Technology Assessment that address MSI and BRAF testing: "The MSI and BRAF tests can also be used to screen	This review does not examine the value of the tests for detecting Lynch Syndrome. We have referred your concerns re payment decisions to CMS. Per the reviewer's comment, we added Palomaki et al 2009 to the full-text literature review and excluded it in the final report because it does not include an eligible population.

Public Reviewer & Affiliation	Section	Comment	Response
		colorectal cancer (CRC) patients to identify patients who are more likely to have Lynch syndrome (a hereditary cancer syndrome that causes CRC, endometrial, gastric, ovarian, and other cancers." NSGC appreciates this opportunity to provide comments on AHRQ's Technology Assessment of Genetic Testing for Risk of Recurrent Cancer draft. We look forward to collaborating with AHRQ to continue to properly facilitate healthcare advancements through research advancements and quality improvement. 1) Palomaki, GE, McClain MR, Melillo S, Hampel HL, Thibodeau SN. EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from Lynch syndrome. Genet Med. 2009 Jan; 11(1):42-65.	
Bastiaan van der Baan Agendia, Inc.	Discussion/Conclusion	MammaPrint: We would like to start by addressing the overarching question and your conclusion. Page 87. "We found no studies that directly addressed our overarching question (i.e., no studies directly assessed the impact of test use on downstream health outcomes to establish clinical utility)." In the continuing evolution of adjuvant therapy for early stage breast cancer, it is important to keep in mind that following NSABP B---20 in 1999, there was a prevailing thought in the US that all stages of breast cancer were shown to benefit from chemotherapy. As we started into the early 2000's, gene expression profiles facilitated clinical risk stratification for distant metastasis in early stage breast cancer that challenged the notion that all early stage breast cancer patients would benefit from chemotherapy. The development of the first two signatures: MammaPrint in 2002 and Oncotype DX in 2004 was significantly different both in terms of the selection of the genes that comprised the signatures but also with regard to the testing platforms, are PCR vs. Microarray and the choice of validation cohorts. The evolution of these two signatures along with physician adoption significantly changed over the last 10 years. MammaPrint, was evaluated for its impact and outcome very early in its course in the only published (RASTER Trial) prospective observational trial in real world, community based patients in 16 clinics in the Netherlands between 2004	We have responded to this comment earlier. The Drukker et al 2013 article was reviewed and excluded. Per the reviewer's comment, we added the Rutgers et al 2013 study and excluded it from this report because it does not report an eligible outcome. Similarly with the Drukker trial which estimates hypothetical treatment decisions based on current information. "One limitation of the comparison between the gene signature and AOL is that the actual treatment decisions in this study were based on the restrictive Dutch guidelines

Public Reviewer & Affiliation	Section	Comment	Response
		and 2006. The MammaPrint result was incorporated into the clinical decisions of 427 patients and compared to the performance of a validated clinical---pathological risk model of Adjuvant! Online. The published 5---year outcome of all 427 patients in that study (Drukker 2013) demonstrated a reduction of 32% in the number of high risk patients classified by the standard clinical risk model, and a 20% change in intended therapy by physicians. Most importantly, in the largest group of patients with a discordant risk classification, where the clinical model allocated patients to high risk and MammaPrint allocated patients to low risk, the physicians chose to follow the advice of MammaPrint and those patients had a 100% distant disease free survival at five years without ANY adjuvant systemic therapy. Moreover where the physicians chose to treat 81% of the high risk patients with chemotherapy,, those patients had a 5 year distant metastasis free survival of 92%, improved from an average of 71% without chemotherapy. MammaPrint thereby identifies not only those patients who can safely avoid chemotherapy without incurring avoidable harm, but also demonstrated its value in correctly identifying those high risk patients who would most benefit from chemotherapy. Both patient populations had improvement in net health outcomes. Recently the patient data on the MINDACT trial was presented and confirmed the 31% discordance (2142 out of 6694) between clinical risk assessment and MammaPrint (Rutgers et al 2013). Today, a significant (reported as 80%) majority of US based oncologists follow NCCN guidelines for determining chemotherapy. Based on 2013 NCCN guidelines most patients with primary tumors of >0.5 cm, hormone receptor positive breast cancer would be recommended to 'consider chemotherapy'. Oncologists choice for stratifying patients in the ER+, Her---2 negative, Lymph node negative population to chemotherapy or not, is to test by the gene expression profile of Oncotype DX which has, as far as we are aware, no published prospective outcome data or to test with MammaPrint with 5 year outcome data that is sufficient to withhold chemotherapy in the low risk group and give	of 2004 and doctor's and patients' preferences" Drukker et al (2013) also did not do a multivariate analysis to control for known risk factors so the article was excluded due to ineligible outcomes.

Public Reviewer & Affiliation	Section	Comment	Response
		chemotherapy to those patients in the high risk group. The RASTER trial is a well conducted and widely published nonrandomized observational trial, which, in the absence of any other prospective outcome data by any other commercial lab with an applicable gene expression profile, is at least at the level of moderate clinical utility and improvement of health outcome in a technical assessment.	
Bastiaan van der Baan Agendia, Inc.	Discussion/Conclusion	Please include Impact with outcome data on MammaPrint MammaPrint is the only assay with impact on treatment decision with 5 year outcome data showing excellent survival in Low Risk patients that did forego chemotherapy.	We thank you for the suggestion for additional text but respectfully disagree
Diane Allingham-Hawkins Hayes, Inc.	General	Title: The title is a misnomer. As designed, this is a TA for prognostic genetic tests for cancer, not of genetic tests for recurrence of cancer. Key Questions: The overarching key question is flawed in that it is not explicit enough regarding the fact that the only use of these tests that is being evaluated is their prognostic value. Other uses of these tests are not considered thus it is misleading to use such a general overarching question. Similarly, Key Question 4 is too general and should specify clinical utility from a prognostic perspective. Key Question 4a should specify prognostic information used to make decisions about whether or not to treat rather than which treatments to use. Key Question 5 should specify prognostic genetic tests rather than just genetic tests. Tests Included: It is not clear how tests were selected for inclusion. Many of the tests included have primary uses that are not prognostic. Only Mammaprint and Oncotype DX are primarily prognostic tests. The other tests evaluated are primarily diagnostic (MSI for colorectal cancer, UroVysion for bladder cancer) or pharmacogenetic (KRAS and EGFR in lung cancer, BRAF and KRAS in colorectal cancer) tests. Consequently, it is not surprising that there is little evidence	Thank you. We have tried to clarify the fact that the review looked only at the prognostic efficacy to the tests. And we have edited the title to make it more clear The tests were selected through discussions with the review team, AHRQ and CMS. One of the considerations was the assess the prognostic value of the tests that are currently used in clinical

Public Reviewer & Affiliation	Section	Comment	Response
		of prognostic value. Also, other prognostic genetic tests for cancer such as Mammostrat, BluePrint, DecisionDx and many other were not included. Recommendations: 1. Clarify title to specify prognostic genetic tests. 2. Clarify key questions to indicate that only prognostic indications of tests were considered. 3. Re-evaluate test inclusion criteria to ensure that only tests that are primarily prognostic in nature are included.	practice. We have modified the title. We have clarified that only prognostic uses were considered. The review is complete now and all reviewed tests will be included in the report.
Bastiaan van der Baan Agendia, Inc.	General	We thank the authors for the thorough review of the topic of genetic (and genomic) testing in cancer and for the technical report. We would like to add several comments to the document and provide the authors with more evidence in support of MammaPrint and ColoPrint. Currently ColoPrint is not mentioned in the document. ColoPrint is a commercially available diagnostic test for colorectal cancer patients that has been validated in 3 independent datasets. Please note that ColoPrint is also described in more details in the AHRQ technical brief “Gene Expression Profiling for Predicting Outcomes in Stage II Colon Cancer” from 2012.	Thank you. The ColoPrint diagnostic test was not part of the original scope requested by CMS; thus, is not included in this report.

Public Reviewer & Affiliation	Section	Comment	Response
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	General	We thank the AHRQ for the opportunity to review and submit feedback on the document "Technology Assessment of Genetic Testing for Risk of Recurrent Cancer." We appreciate that the AHRQ summary found good evidence supporting added prognostic accuracy beyond traditional prognostic measures for the Oncotype DX® Breast Cancer Assay, and moderate evidence that the breast assay leads to changes in treatment decisions. However, we believe the published evidence, assessed as a whole and in the appropriate clinical contexts, provides a much higher level of confidence than is currently reflected in the technological assessment. We believe that the methodological approaches taken and/or the execution of the methods resulted in the exclusion or omission of critical publications supporting both of our assays. As a result, the current technological assessment fails to accurately reflect the published literature. In this reply, we will present further data to support the analytic validity, clinical validity, and clinical utility of the Oncotype DX breast and colon assays.	Thank you for the comments. We responded to each of the specific issues that this comment introduces in other parts of this document. We stand by our assessments of the strength of evidence for the various tests and outcomes.
Phillip Febbo, Chief Medical Officer Genomic Health, Inc	References	[*The comments from Genomic Health include 71 references*] 1. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst. 2009;101(21):1446-52. 2. McShane LM, Cavenagh MM, Lively TG, et al. Criteria for the use of omics-based predictors in clinical trials. Nature. 2013;502(7471):317-20. 3. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. J Natl Cancer Inst. 2010;102(3):152-60. 4. Genomic Health, Inc; Data on file 5. Cronin M, Sangli C, Liu ML, et al. Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor-positive breast cancer. Clin Chem. 2007;53(6):1084-91. 6. Habel LA, Shak S, Jacobs MK, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. Breast Cancer Res. 2006; 8(3): R25. 7. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004; 351(27): 2817-26 8. College of American Pathologists. http://www.cap.org. Accessed December 12, 2013.	Per the reviewer's comment, we added the Simon et al 2009, McShane et al 2013, Freidlin et al 2010, Harris et al 2007, Senkus et al 2013 articles to our full-text literature review and excluded these articles from the final report for wrong study design. Per the reviewer's comment, we added the Cronin et al 2007, Yamachuch et al 2013, and Bargallo et al 2012, Brenner et al 2013 articles to our full-text literature

Public Reviewer & Affiliation	Section	Comment	Response
		9. Clinical Laboratory Improvement Amendments. Centers for Medicare & Medicaid Services. http://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA . Published October28, 2013. Accessed December 12, 2013. 10. Clinical Laboratory Evaluation Program. New York State Department of Health: Wadsworth Center. http://www.wadsworth.org/labcert/clep/clep.htm. Accessed December 12, 2013. 11. Dowsett M, Cuzick J, Wale C, et al. Prediction of risk of distant recurrence using the 21-gene Recurrence Score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a transATAC study. J Clin Oncol. 2010; 28(11): 1829-34. 12. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol. 2006; 24(23): 3726-34. 13. Albain KS, Barlow WE, Shak S, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. Lancet Oncol. 2010; 11(1): 55-65. 14. Toi M, Iwata H, Yamanaka T, et al. Clinical significance of the 21-gene signature (Oncotype DX) in hormone receptor-positive early stage primary breast cancer in the Japanese population. Cancer. 2010; 116(13):3112-8 15. Gianni L, Zambetti M, Clark K, et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. J Clin Oncol. 2005; 23 (29): 7265-77. 16. Chang J, Makris A, Gutierrez M, et al. Gene expression patterns in formalin-fixed, paraffin-embedded core biopsies predict docetaxel chemosensitivity in breast cancer patients. Breast Cancer Res Treat. 2008; 108 (2): 233-40. 17. Akashi-Tanaka S, Shimizu C, Ando M, et al. 21-Gene expression profile assay on core needle biopsies predicts responses to neoadjuvant endocrine therapy in breast cancer patients. Breast. 2009; 18 (3): 171-4. 18. Yardley DA, Peacock NW, Hendricks C, et al. Correlation of Oncotype DX Recurrence Scores with pathologic response following neoadjuvant ixabepilone and cyclophosphamide in patients with HER2-negative breast cancer. Poster presented at the San Antonio Breast Cancer Symposium; December 2011; San Antonio, TX. 19. National Comprehensive Cancer Network practice guidelines in	review and included these articles in the final report based on our inclusion criteria. The Dowsett et al 2010, Albanell et al 2012, Davidson et al 2013, Lo et al 2010, Klang et al 2010, Gilgorov et al 2012, Eiermann et al 2013, Clark-Langone et al 2010, Venook et al 2013, Cartwright et al 2013, Srivastava et al 2013 articles were reviewed in our draft report and included these articles in the final report based on our inclusion criteria. The Holt et al 2013 and de Boer et al 2013 articles were reviewed in our updated literature review and included these articles in the final report based on inclusion criteria. The Paik et al 2006, Asad et al 2008, Oratz et al 2007, Hornberger et al 2011, and Akashi-Tanaka et al 2009 articles were reviewed in our draft report and excluded for wrong or no comparator.

Public Reviewer & Affiliation	Section	Comment	Response
		oncology. Breast cancer. www.nccn.org, v3.2013. 20. Harris L, Fritsche H, Mennel R, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol. 2007; 25 (33): 5287-5312. 21. Goldhirsch A, Winer EP, Coates AS, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol. 2013;24:2206-2223. 22. Senkus E, Kyriakides S, Penault-Llorca F, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2013; {Epub ahead of print]. 23. National Institute for Health and Care Excellence. Gene expression profiling and expanded immunohistochemistry tests for guiding adjuvant chemotherapy decisions in early breast cancer management: MammaPrint, Oncotype DX, IHC4 and Mammostrat. NICE diagnostics guidance 10 www.nice.org.uk/dg10, 2013. 24. Albanell J, Gonzalez A, Ruiz-Borrego M, et al. Prospective transGEICAM study of the impact of the 21-gene Recurrence Score assay and traditional clinicopathological factors on adjuvant clinical decision making in women with estrogen receptor-positive (ER+) node-negative breast cancer. Ann Oncol. 2012; 23(3): 625-31. 25. Davidson JA, Cromwell I, Ellard S, et al. A prospective clinical utility and pharmacoeconomic study of the impact of the 21-gene Recurrence Score assay in oestrogen receptor positive node negative breast cancer. Eur J Cancer. 2013;S0959-8049(13):00211-6. 26. Lo SS, Mumby PB, Norton J, et al. Prospective multicenter study of the impact of the 21-gene recurrence score assay on medical oncologist and patient adjuvant breast cancer treatment selection. J Clin Oncol. 2010; 28 (10): 1671-6. 27. Asad J, Jacobson AF, Estabrook A, et al. Does Oncotype DX recurrence score affect the management of patients with early-stage breast cancer? Am J Surg. 2008; 196 (4): 527-9. 28. Klang SH, Hammerman A, Liebermann N, et al. Economic implications of 21-gene breast cancer risk assay from the perspective of an Israeli-managed health-care organization. Value Health. 2010; 13 (4): 381-7. 29. Lo SS, Mumby PB, Norton J, et al. Prospective multicenter study of the impact of the 21-gene recurrence score assay on medical oncologist and patient adjuvant breast cancer treatment selection. J Clin Oncol. 2010; 28 (10): 1671-6. 30. Klang SH, Hammerman A, Liebermann N, et al. Economic implications of 21-gene breast cancer risk assay from the perspective of an Israeli-managed health-care organization. Value Health. 2010; 13 (4): 381-7.	The Albain et al 2010, Gianni et al 2005, Vanderlaan et al 2011, O'Connell et al 2010 and Chang et al 2008, Hornberger et al 2012 articles were reviewed in our draft report and excluded for wrong intervention/test. If the tests was used to predict response to treatment it was the ' wrong' test for this review. The Toi et al 2010, Oratz et al 2011, Hannouf et al 2012, Blohmer et al 2012 articles were reviewed in our draft report and excluded for wrong study design. Per the reviewer's comment, we added the Yardley et al 2011, Lacey et al 2011, Holt et al 2011, Tsoi et al 2010, O'Leary et al 2010, Cosler et al 2009, Fisher et al 1997, Hassett et al 2006, Gray et al 2007, Yothers et al 2011, Yu et al 2013, Kuebler et al 2007, Kidwell et al 2012 articles to our full-text literature review and excluded these articles from the final report for

Public Reviewer & Affiliation	Section	Comment	Response
		31. Oratz R, Paul D, Cohn AL, et al. Impact of a commercial reference laboratory test recurrence score on decision making in early-stage breast cancer. J Oncol Pract. 2007; 3 (4): 182-6. 32. Gligorov J, Pivot XB, Naman HL, et al. Prospective study of the impact of using the 21-gene recurrence score assay on clinical decsion making in women with estrogen receptor-positive, HER2-negative, early stage breast cancer in France. Poster presented at: American Society for Clinical Oncology Annual Meeting; June 2012; Chicago, IL. 33. Holt S, Bertelli G, Humphreys I, et al. A decision impact, decision conflict and economic assessment of routine Oncotype DX testing of 146 women with node-negative or pNImi, ER-positive breast cancer in the UK. Br J Cancer. 2013;108(11): 2250-8. 34. Eiermann W, Rezai M, Kümmel S, et al. The 21-gene recurrence score assay impacts adjuvant therapy recommendations for ER-positive, node-negative and node-positive early breast cancer resulting in a risk-adapted change in chemotherapy use. Ann Oncol. 2013; 24(3): 618-24. 35. Bargallo JER, Lara F, Shaw Dulin RJ, et al. A study of the impact of the 21-gene breast cancer assay on the use of adjuvant chemotherapy in women with breast cancer in a Mexican public hospital. Poster presented at: European Society for Medical Oncology Congress; September 2012; Vienna, Austria. 36. de Boer RH, Baker C, Speakman D, et al. The impact of a genomic assay (Oncotype DX) on adjuvant treatment recommendations in early breast cancer. Med J Aus. 2013;199: 205-8. 37. Yamauchi H, Nakagawa C, Takei H, et al. Prospective study of the effect of the 21-gene assay on adjuvant clinical decision-making in Japanese women with estrogen receptor-positive, node-negative, and node-positive breast cancer. Clin Breast Cancer. 2013 Oct 26. 38. Oratz R, Kim B, Chao C, et al. Physician survey of the effect of the 21-gene recurrence score assay results on treatment recommendations for patients with lymph node-positive, estrogen receptor-positive breast cancer. J Oncol Pract. 2011; 7(2): 94-9. 39. Lacey L, Chien R, Hornberger J. Cost-utility of the 21-gene breast cancer assay (Oncotype DX®) in the Irish healthcare setting. Poster presented at San Antonio Breast Cancer Symposium; December 2011; San Antonio, TX. 40. Holt SDH, Bennett H, Bertelli G, et al. Cost-effectiveness evaluation of the Onco*type* DX® breast cancer assay in clinical practice in the UK. Poster presented at the San Antonio Breast Cancer Symposium, December 2011; San Antonio, TX. 41. Klang SH, Hammerman A, Liebermann N, et al. Economic implications	ineligible outcome. Some of the aforementioned studies made it to full-text review and were excluded in Appendix B. Per the reviewer's comment, we added the Goldhirsch et al 2013 article to our full-text literature review and excluded this article from the final report for wrong population Per the reviewer's comment, we added the National Institute for Health Care and Excellence 2013 article to our full-text literature review and excluded this article from the final report for systematic review. Per the reviewer's comment, we added the Madaras et al 2012 article to our full-text literature review and excluded this article from the final report for wrong or no comparator. Per the reviewer's comment, we added the de Lima et al 2013, Vataire 2012, Hornberger et al

Public Reviewer & Affiliation	Section	Comment	Response
		of 21-gene breast cancer risk assay from the perspective of an Israeli-managed health-care organization. *Value Health.* 2010;13(4): 381-7. 42. Tsoi DT, Inoue M, Kelly CM, et al. Cost-effectiveness analysis of recurrence score-guided treatment using a 21-gene assay in early breast cancer. Oncologist. 2010; 15(5): 457-65. 43. Hannouf MB, Xie B, Brackstone M, Zaric GS. Cost-effectiveness of a 21-gene recurrence score assay versus Canadian clinical practice in women with early-stage estrogen- or progesterone-receptor-positive, axillary lymph-node negative breast cancer. BMC Cancer. 2012;12: 447. 44. Madaras B, Rozsa P, Gerencser Z et al. The impact of chemotherapeutic regiments on the cost-utility analysis of Onctype DX® assay. Poster presented at: The European Breast Cancer Conference; March 2012; Vienna, Austria. 45. O'Leary B, Yoshizawa C, Foteff C, Chao C. Cost-effectiveness of the Oncotype DX assay in Australia: an exploratory analysis. Presented at International Society for Pharmoeconomics and Outcomes Research; May 2010; Atlanta, GA. 46. de Lima Lopes G, Chien R, Hornberger J. Cost-benefit of the 21-gene Recurrence Score® assay for patients in Singapore. Breast J. 2013;19(2): 220-1. 47. Vataire AL, Laas E, Aballéa S, et al. Cost-effectiveness of a chemotherapy predictive test. Bull Cancer. 2012;99(10): 907-14. 48. Blohmer JU, Rezai M, Kümmel S, et al. Using the 21-gene assay to guide adjuvant chemotherapy decision-making in early-stage breast cancer: a cost-effectiveness evaluation in the German setting. J Med Econ. 2012; 1–11. 49. Vanderlaan BF, Broder MS, Chang EY, et al. Cost-effectiveness of 21-gene assay in node-positive, early-stage breast cancer. Am J Manag Care 2011; 17 (7): 455-64. 50. Hornberger J, Lyman GH, and Chien R. Economic implications of 21-gene Recurrence Score assay: US multicenter experience. J Clin Oncol. 2010; 28(22): e382; author reply e383. 51. Cosler LE, Lyman GH. Economic analysis of gene expression profile data to guide adjuvant treatment in women with early-stage breast cancer. Cancer Invest. 2009; 27(10): 953-9. 52. Hornberger J, Chien R, Krebs K, et al. US insurance program's experience with a multigene assay for early-stage breast cancer. J Oncol Pract. 2011; 7(3 Suppl): e38s-45s. 53. Reed SD, Dinan MA, Schulman KA, et al. Cost-effectiveness of the 21-gene recurrence score assay in the context of multifactorial decision making to guide chemotherapy for early-stage breast cancer. Genet Med. 2013;15(3):203-11.	2010 articles to our full-text literature review and excluded these articles from the final report for wrong publication type. The Reed et al 2013 article was reviewed in our draft report and excluded for wrong publication type. Per the reviewer's comment, we added the Jemal et al 2010 article to our full-text literature review and excluded this article from the final report for wrong intervention/test. The Gray et al 2011 article was reviewed in our draft report and excluded for ineligible outcome on pg. B-28. The Yothers et al 2013 article was reviewed in our draft report and excluded for not applying to a key question.

Public Reviewer & Affiliation	Section	Comment	Response
		54. Fisher B, Dignam J, Wolmark N, et al. Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. J Natl Cancer Inst. 1997;89(22):1673-82. 55. Hassett MJ, O'Malley AJ, Pakes JR, Newhouse JP, Earle CC. Frequency and cost of chemotherapy-related serious adverse effects in a population sample of women with breast cancer. J Natl Cancer Inst. 2006;98(16):1108. 56. Clark-Langone KM, Sangli C, Krishnakumar J, Watson D. Translating tumor biology into personalized treatment planning: analytical performance characteristics of the Oncotype DX® Colon Cancer Assay. BMC Cancer. 2010,10:691. 57. O'Connell MJ, Lavery I, Yothers G, et al. Relationship between tumor gene expression and recurrence in four independent studies of stage II/III colon cancer patients treated with surgery alone and surgery plus adjuvant fluorouracil plus leucovorin. J Clin Oncol. 2010;28(25):3904-7. 58. Jemal A, Center MM, DeSantis C, et al. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev. 2010; 19 (8): 1893-1907. 59. Cancer Facts and Figures. Atlanta, GA: American Cancer Society; 2013. 60. Gray R, Barnwell J, McConkey C, et al. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. Lancet. 2007; 370 (9604): 2020-9. 61. Yothers G, O'Connell MJ, Allegra CJ, et al. Oxaliplatin as adjuvant therapy for colon cancer: Updated results of NSABP C-07 trial, including survival and subset analyses. J Clin Oncol. 2011;29(28): 3768-74. 62. Gray RG, Quirke P, Handley K, et al. Validation study of a quantitative multigene reverse transcriptase–polymerase chain reaction assay for assessment of recurrence risk in patients with stage II colon cancer. J Clin Oncol. 2011;29(35):4611-9. 63. Venook AP, Niedzwiecki D, Lopatin M, et al. Biologic determinants of tumor recurrence in stage II colon cancer: validation study of the 12-gene recurrence score in cancer and leukemia group B (CALGB) 9581. J Clin Oncol. 2013; 31(14): 1775-81. 64. Yothers G, O'Connell M, Lee M, et al. Validation of the 12-gene colon cancer Recurrence Score in NSABP C-07 as a predictor of recurrence in stage II and III colon cancer patients treated with 5-FU/LV and 5-FU/LV+ oxaliplatin. J Clin Oncol. 2013 Nov 12. (electronic publication ahead of print). 65. Cartwright T, Chao C, Lee M, et al. Effect of the 12-gene colon cancer assay results on adjuvant treatment recommendations in patients with Stage II colon cancer. Curr Med Res Opin. 2013; [ePub ahead of print]	

Public Reviewer & Affiliation	Section	Comment	Response
		66. Srivastava G, Renfro LA, Behrens RJ, et al. Prospective evaluation of a 12-gene assay on treatment recommendations in patients with stage II colon cancer. Poster presented at: ASCO Gastrointestinal Symposium; January 2013; San Francisco, CA. 67. Brenner B, Lopatin M, Lee M, et al. Impact of the 12-gene colon cancer recurrence score assay on clinical decision-making for adjuvant therapy in stage II colon cancer patients in Israel. Poster presented at: European Cancer Congress; September 2013; Amsterdam, Netherlands. 68. Hornberger J, Lyman GH, Chien R, Meropol NJ. A multigene prognostic assay for selection of adjuvant chemotherapy in patients with T3, stage II colon cancer: impact on quality-adjusted life expectancy and costs. Value Health. 2012;15(8): 1014-21. 69. Yu T, Alberts SR, Behrens RJ, et al. Real-world comparative economics of a 12-gene assay for prognosis in stage II colon cancer. Presented at: American Society of Clinical Oncology; June 2013; Chicago, IL. 70. Kuebler JP, Colangelo L, O'Connell MJ, et al. Severe enteropathy among patients with stage II/III colon cancer treated on a randomized trial of bolus 5-fluorouracil/leucovorin plus or minus oxaliplatin: a prospective analysis. Cancer. 2007 Nov 1;110(9):1945-50. 71. Kidwell KM, Yothers G, Ganz PA, et al. Long-term neurotoxicity effects of oxaliplatin added to fluorouracil and leucovorin as adjuvant therapy for colon cancer: results from National Surgical Adjuvant Breast and Bowel Project trials C-07 and LTS-01. Cancer. 2012;118(22):5614-22.	
Susan Jewell, Ph.D. Abbott Molecular	General	Key Issues: -UroVysion™ is exclusively a diagnostic test, not a prognostic test. -Therefore, it has not been studied for prognostic value, nor would there be clinical impact nor health outcome studies on its “prognostic value.” - For this obvious reason, UroVysion should not be included in the final report, or at a minimum, extensive revisions must be added.	We have added text to the report to clarify this (pg 81).

Public Reviewer & Affiliation	Section	Comment	Response
Susan Jewell, Ph.D. Abbott Molecular	General	*Abbott Molecular is the inventor and manufacturer of the FDA-approved UroVysion Bladder Cancer Kit (UroVysion Kit).* We have reviewed the AHRQ document in detail, and we believe a substantial error has been made in the inclusion of UroVysion™ in the analysis. Simply stated, UroVysion is a diagnostic test, not a prognostic test. Therefore, it has not been studied for "prognostic" statistics (clinical validity) nor for its "prognostic" impact on management or survival. It is therefore a serious mistake to extrapolate from this viewpoint whether UroVysion has "value" as a genomic test. However, if UroVyison were included in the final report, it has excellent analytical accuracy, exceeds previous diagnostic tests, and has clinical impact as a diagnostic test. A full data presentation is attached. We highlight key summary points below:	We retained the findings but made it clear that Urovysion is not intended for use as a prognostic tool (pg 81).
Susan Jewell, Ph.D. Abbott Molecular	General	UroVysion was never approved or intended to be used for prognosis or predicting risk of recurrence. The two FDA-approved claims for UroVysion include: -Aid in initial diagnosis of bladder cancer in patients with hematuria in conjunction with standard diagnostic procedures. The AHRQ document confirms that UroVysion "has been shown to be sensitive in terms of diagnosing urothelial cancer." -Monitoring for tumor recurrence in patients previously diagnosed with bladder cancer. Here the focus is the detection or reappearance of cancer and not on prognosis or severity of cancer. -Use of the test for "prognosis" would be an off label use, and indeed, there is no such use in clinical custom.	We retained the findings but made it clear that Urovysion is not intended for use as a prognostic tool (pg 81).

Public Reviewer & Affiliation	Section	Comment	Response
Susan Jewell, Ph.D. Abbott Molecular	General	The reviewers state that their mandate was to review "prognostic tests" for common cancers in Medicare patients, i.e., bladder, lung, colon, breast. On page 5, the authors, Meleth, Reeder-Hayes *et al.,* state that particular tests were selected based on (1) literature searches, (2) clinical expert consultations, and (3) consultation with the funding agency. We believe that a look-back is appropriate and that there is no a priori clear evidence that UroVysion is a prognostic test. Even if a clinician suggested this, it is so out of keeping with the actual literature and usage that it was an inappropriate suggestion, possibly a misunderstanding or miscommunication.	We retained the findings but made it clear that Urovysion is not intended for use as a prognostic tool (pg 81).
Susan Jewell, Ph.D. Abbott Molecular	General	Three publications were reviewed in the AHRQ report, Kamat *et al*., Whitson *et al*., and Zellweger *et al.,* to "examine the added prognostic value of UroVysion for RR, CSS or OS in patients with bladder cancer". Abbott Molecular reviewed these publications with a focus on the product's intended use of monitoring for recurrence – that is, the DIAGNOSIS of recurrence. For example, in the publication, Kamat et al., we only approve the use of data from Figure 1 as being on label and aligned with the FDA-approved intended use of monitoring for recurrence. Here, the end point used was appropriately recurrence-free survival. Given our focus on purely monitoring and not prognosis, the end points of RR, CSS or OS are not relevant at all. This should have confirmed to the authors, Meleth, Reeder-Hayes *et al.,* that the suggestion, from whatever source, that UroVysion was a "prognostic" test was a miscommunication in the first place, due to the resulting complete absence of literature or guidelines suggesting it was prognostic.	We retained the findings but made it clear that Urovysion is not intended for use as a prognostic tool and the framework of what our report evaluated.

Public Reviewer & Affiliation	Section	Comment	Response
Susan Jewell, Ph.D. Abbott Molecular	General	Neither Abbott Molecular, nor, to our knowledge, any clinical guidelines anywhere in the world, have claimed that the utility of UroVysion lies in prognosis. In this narrow and unusual sense, it is correct that AHRQ found no supporting literature. Rather UroVysion's analytical and clinical validity and clinical utility lie in diagnosis and monitoring for recurrence, which is a form of diagnosis (the diagnosis of recurrent tumor).	We retained the findings but made it clear that Urovysion is not intended for use as a prognostic tool and the framework of what our report evaluated.

Public Reviewer & Affiliation	Section	Comment	Response
Susan Jewell, Ph.D. Abbott Molecular	General	We believe that the obvious conclusion is that UroVysion should be removed from the final report, since it is so clearly out of scope, as confirmed by the FDA label and AHRQ's own literature review. However, in the case that UroVysion is not removed, in the attachment *[see Attachment 2 at the end of this document]* we provide evidence around the two above mentioned actual intended uses of UroVysion for bladder cancer patients. A. Impact to Patient Management and Health Economics 2013 publication, (Gayed et al.) to demonstrate use of the UroVysion® Kit in cases of negative or equivocal cystoscopy and atypical cytology to decide whether to biopsy a patient significantly decreases bladder cancer costs. Additionally, it provides clinical utility for the UroVysion® Kit in patient management to avoid unnecessary biopsies. B. Evidence of Clinical Utility & Support from US/International Clinical Practice Guidelines a. Monitoring for Recurrence: Inclusion of FDA-Approved FISH in National Comprehensive Cancer Network (NCCN) Bladder Cancer Guidelines (v1. 2013) and Relevant Publications b. Diagnosis: Inclusion of FISH in 2013 International Consultation Urologic Diseases – European Association of Urology (ICUD – EAU) Recommendations and Relevant Publications C. Evidence of Clinical Utility The role of the UroVysion® Kit in diagnosis of bladder cancer, and comparison to cytology	We retained the findings but made it clear that Urovysion is not intended for use as a prognostic tool and the framework of what our report evaluated.

Public Reviewer & Affiliation	Section	Comment	Response
Susan Jewell, Ph.D. Abbott Molecular	General	*[A summary of evidence for the UROVYSION® KIT AMD-00001216 was provided as an attachment to comments submitted by the reviewer. See Attachment 2 at the end of this document.]*	Per the reviewer's comment, the Urovysion Kit was included in our full-text literature and excluded from the final report for ineligible outcome on pg. B-1.
James L. Madara American Medical Association	General	Dear Administrator Kronick: On behalf of the physician and medical student members of the American Medical Association (AMA), I appreciate the opportunity to provide comments on the Agency for Healthcare Research and Quality's (AHRQ) draft technology assessment entitled, "Technology Assessment of Genetic Testing for Risk of Recurrent Cancer." The AMA acknowledges the growing complexity of molecular tests designed to provide prognostic information for cancers, and like, AHRQ, notes that the evidence base for clinical use of the tests is still developing. However, we believe that the standard of "clinical utility" employed in the technology assessment is too narrow, and that the outcomes considered do not reflect the true nature and complexity of clinical decision-making and patient management that occurs for each patient diagnosed with cancer. This would compromise the quality of care patients receive. Further, we are very concerned that the clinical utility standard employed would substantially undermine the ability of physicians to utilize medically necessary and reasonable genetic tests that are supported by the weight of current and rapidly emerging clinical evidence and medical practice, including practice guidelines. We strongly object to the use of this poorly defined standard in this assessment and are very concerned with its broader application for purposes of making coverage determinations. Our concerns are addressed in more detail below, and recommendations for improving the technology assessment are included.	Your comments have been shared with CMS. We used standard definitions of clinical utility. We have evaluated the science based on the framework from the Evidence-based practice center and your comments regarding coverage decisions have been shared with CMS.
James L. Madara American Medical	General	Though the medically necessary and reasonable standard is associated with Medicare coverage, it is in fact the most appropriate standard when assessing clinical services and tests. The practice of medicine must be patient-centered, and	Your comments have been shared with CMS. We used standard

Public Reviewer & Affiliation	Section	Comment	Response
Association		physicians should have the professional discretion to utilize tests and services that are medically necessary and reasonable. Efforts to create more restrictive standards will undermine the ability of patients to receive care that improves health outcomes, is patient-centered, and in the long-term interest of the health care system. Furthermore, as health care delivery continues to evolve to a more outcome-focused approach, enabling physicians to determine which testing services are appropriate for their patient is critical. Going forward, we urge AHRQ to employ an assessment standard that is better designed to support patient-centered care, namely medically necessary and reasonable. Having stated the foregoing, to the extent that AHRQ elects to rely upon the more restrictive standard of clinical utility, for which limited consensus exists in defining, we provide the following comments:	definitions of clinical utility. We have evaluated the science based on the framework from the Evidence-based practice center and your comments regarding coverage decisions have been shared with CMS.
James L. Madara American Medical Association	General	Clinical utility, i.e., the ability of a test to improve outcomes when used in a clinical setting, can be assessed by a number of different factors. These include the test's ability to influence risk assessment and screening procedures, confirm or rule out a diagnosis, provide information on prognosis, guide therapeutic options, and inform genetic counseling. These factors are of utmost importance to physicians as they consider management options. The role of clinical utility as it relates to the patient also is exceptionally important; the ability of a test to end a diagnostic odyssey and/or to provide prognostic information that will help patients in life planning is extremely meaningful to patients. It is within the aforementioned context that we are alarmed at the narrow standard of clinical utility applied in the technology assessment. The technology assessment states that it examined whether evidence supported the test's ability to influence decision-making on the use of chemotherapy and/or radiotherapy, enhanced diagnostic testing for recurrence, and preventive surgery. It appears that there was no consideration of other clinically relevant factors, such as the test's ability to influence therapeutic decisions other than	The assessment reviewed all studies that affected decision making of any type. KQ4b was also designed to identify the impact of the tests on patient centered outcomes. The evidence on KQ4a was primarily related to Oncotype Dx for breast. All of the evidence available looks at the impact of the test on whether to provode chemohormonal therapy or just chemo or hormonal therapy. While we were expecting to find evidence on patient centered outcomes such as QOL,.

Public Reviewer & Affiliation	Section	Comment	Response
		those involving chemotherapy or radiation (such as risk-reducing medications or targeted therapeutics), enhanced screening procedures, adherence to the therapeutic regimen, satisfaction with the treatment choice, or incidence and severity of adverse events. These factors are extremely important in patient management decisions. We do not disagree that evidence assessing these clinical factors may be limited, but we are concerned that the technology assessment has limited itself to factors that represent a fraction of those physicians consider when developing management plans. We urge AHRQ to expand the clinical factors considered to more accurately reflect the complexity of decision-making that informs patient management decisions. It is becoming increasingly clear that cancers are molecularly distinct, even those originating in the same tissue. Concomitantly, clinical decision-making and patient management are unique to each patient and clinical situation. We believe that in an attempt to analyze available data on the tests under consideration, the authors neglected to consider the unique clinical contexts in which each of the tests is used. We urge AHRQ to utilize physicians who are experts in the molecular characterization of cancers to inform the Agency on which clinical utility factors and outcomes should be examined and to review the draft technology assessment. We are in a time of unprecedented growth in personalized medicine-based tools, making it difficult for many insurers to make coverage decisions informed by the most current evidence. Clinical trials soon to be completed could impact results of technology assessments, making them outdated soon after they are completed. Guidelines-making bodies such as the National Comprehensive Cancer Network and the American Society of Clinical Oncology are working intensively to update guidelines as new studies emerge. We fear that this technology assessment, which could be used to make coverage decisions contrary to those included in clinical guidelines, will introduce significant confusion into physician decision-making and insurer coverage decisions, and harm physicians' ability to deliver	the only published data that came close to patient centered outcomes in studies that used the tests were reports of decisional conflict.

Public Reviewer & Affiliation	Section	Comment	Response
		outcome-driven care. Additionally, non-coverage decisions could result in substantial access problems for patients whose physicians recommend the test but who cannot afford to pay for it themselves. This concern is heightened when taken in combination with our earlier point that the clinical utility factors considered in the technology assessment incompletely reflect the complexity of patient management for cancer. We urge AHRQ to evaluate the potential impact of this technology assessment on physician decisionmaking, insurance coverage decisions, and in turn, patient access to needed tests. Thank you for the opportunity to comment on the draft technology assessment. We would welcome the opportunity to meet with AHRQ along with other physician and allied professionals to discuss the general issues we have raised in this comment letter. The AMA believes that advances in personalized medicine hold significant promise for improving patient care, and looks forward to assisting in the development of policies that ensure appropriate clinical implementation.	

Public Reviewer & Affiliation	Section	Comment	Response
Stefan Sauerland Institute for Quality and Efficiency in Health Care (IQWiG), Germany	General	This most timely AHRQ report addresses an extremely important issue which contributes substantially to the care of patients with oncologic diagnoses. Furthermore, the topic relates to several methodological questions, the solutions of which will provide guidance on how to evaluate current and future biomarkers in the context of “targeted” cancer therapies. The AHRQ report aimed to assess the “clinical utility” of genetic tests, but no direct evidence was found that assessed the impact of diagnostic test performance on “downstream health outcomes.” It is mentioned in the discussion section that it was beyond the scope of the present review to examine whether the tests have clinical utility when used for predicting response to treatment. In our view, this preponderance of prognostic over predictive clinical utility is not helpful, as it narrows the research scope unnecessarily. Medical information (from either a diagnostic or a prognostic test) can only lead to a patient-relevant benefit if this information is used to guide therapeutic interventions. Biomarkers providing such information are called predictive (Sargent et al., J Clin Oncol 2005; 23: 2020-7), which should not be confused with the predictive value within the nomenclature of (diagnostic) accuracy. Prognostic information is not necessarily predictive in the former sense, while “vice versa” information without prognostic capacity may even be predictive. Prognostic information that is not predictive generally does not lead to lower mortality, lower morbidity or improved health-related quality-of-life. Therefore, prognostic accuracy is far less important than predictive information. The Methods Guide for Medical Test Reviews (AHRQ Publication No. 12-EHC017) describes that the best evidence on diagnostic tests consists of a randomized controlled trial (RCT) that compares patient management outcomes of the test to the outcomes of one or more alternative strategies. This type of RCT is commonly known as a “strategy design” or “test-plus-treatment RCT” (Tajik et al. Clin Cancer Res 2013; 19: 4578-88). However, Lijmer and Bossuyt have noted that”various randomized designs can be used to evaluate tests” (J Clin Epidemiol 2009; 62: 364-73).	The review responded to a specific request from AHRQ to examine the added prognostic value of these tests. We have clarified in many parts of the report that the review addresses prognostic value only – not predictive value. The articles by Sargent et al.2005, Tajik et al 2013, Lijmer and Bossuyt 2009, and Lord et al. 2009, were reviewed for background reading. The Paik et al 2006 article was reviewed in our draft report and excluded for wrong or no comparator. The Albain et a 2010 article was reviewed in our draft report and excluded for wrong intervention/test. The Tang et al 2011 article was reviewed in our draft report and excluded for wrong publication type.

Public Reviewer & Affiliation	Section	Comment	Response
		In addition, the AHRQ White Paper Series (Lord et al., Med Decis Making 2009; 29: E1-E12) contains many excellent examples of alternative RCT designs that can be used to assess the clinical utility of genetic tests. According to this article, RCTs that allow a comparison of treatment effects between patients with different test results will also provide optimal evidence. This is called an interaction between (diagnostic or prognostic) information and treatment effects. In the present report the publications of Paik et al. 2006 (J Clin Oncol 2006; 24: 3726-34) and Albain et al. 2010 (Lancet Oncol 2010; 11: 55-65) were unfortunately excluded, although they investigated such an interaction. Furthermore, only the prognostic capacity, but not the predictive ability (= clinical utility), was reported from Tang et al. 2011. In our opinion it would have been essential to consider the above ideas and present predictive rather than prognostic data. By not doing so, we think that very important information is missing in the report.	
Alec Stone, MA, MPA, Director of Health Policy	General	December 11, 2013 Agency for Healthcare Research and Quality (AHRQ) Technology Assessment Program (TAP) RE: Draft Technology Assessment of Genetic Testing for Risk of	Thank you

Public Reviewer & Affiliation	Section	Comment	Response
The Oncology Nursing Society		Recurrent Cancer The Oncology Nursing Society (ONS) is a professional organization of over 35,000 registered nurses and other healthcare providers dedicated to excellence in patient care, education, research, and administration in oncology nursing. ONS members are a diverse group of professionals who represent a variety of professional roles, practice settings, and subspecialty practice areas. In line with our mission and vision to lead the transformation of cancer care and promote excellence in oncology nursing and quality cancer care, we appreciate the opportunity to submit comments on AHRQ's draft Technology Assessment of Genetic Testing for Risk of Recurrent Cancer, which will play an important role in informing the development of Medicare coverage for cancer-related services. ONS appreciates AHRQ's work in preparing this comprehensive review of clinically relevant genetic tests that are increasingly utilized in practice. This is critical work to evaluate utility, reliability and validity across available evidence, as clinicians and payers tend to adopt new technology such as individual prognostic and predictive tests based on single or small groups of controlled studies leading to FDA approval. We especially appreciate the thoughtful construction and design of the analytic framework driven by the key questions, and consistently addressed throughout the document. The key questions are also pertinent and comprehensive, and it is valuable to identify the significant gaps in evidence that may not be apparent to the practicing clinician. Though the intention of highlighting the existence of these gaps may not be to discourage use of these tests in the near term, it may drive a more effective research agenda that prioritizes future work. The practice of oncology nursing is evidence-based and patient- and family-centered. Patients and their families, as well as other healthcare providers and team members, view oncology nurses as partners in cancer risk reduction, cancer treatment, skilled management of symptoms associated with cancer from diagnosis through the end of life, and professional and public education about cancer. To that end, ONS would like to extend itself as a partner to agency	

Public Reviewer & Affiliation	Section	Comment	Response
		as it continues work on this technology assessment or other cancer-related initiatives. We would be happy to discuss ways in which ONS may be of assistance in this endeavor, and would encourage you to contact Alec Stone, MA, MPA, Director of Health Policy, at astone@ons.org to coordinate a time to discuss. ***** Thank you for your consideration of our request, and we look forward to engaging in an ongoing dialogue to address issues of importance to cancer patients and oncology nurses. Sincerely, The Oncology Nursing Society	
Dr. Sophia M. Schild PSEMC	General	Thank yoiu for the comprehensive review of the varied genetic tests. It will be a guide for nursing and physicians as to the relaibilty fo the varied tests for prediction of needed cancer treatment. In light of the recent celebrity use of gentetic testing, this review will help us to educate future patients. Please keep us abreast of changes that will affect patients in the coming years. In addition, thank you for including nursing in your panel of reviewers.	Thank you
Bastiaan van der Baan Agendia, Inc.	Tables (Executive Summary)	(Table A) Please adjust that the evidence for MammaPrint for OS is also sufficient HR for OS is also significant. From vd Vijver et al (2002) estimated hazard ratio for distant metastases as a first event in the group with a poor-prognosis signature as compared with the group with a good-prognosis signature over the entire follow-up period was 5.1 (95 percent confidence interval, 2.9 to 9.0; P<0.001); the prognosis profile was associated with a significantly higher hazard ratio during the first five years of followup (hazard ratio, 8.8; 95 percent confidence interval, 3.8 to 20; P<0.001) than after five years (hazard ratio, 1.8; 95 percent confidence interval, 0.69 to 4.5; P=0.24). The hazard ratio for overall survival was 8.6 (95 percent confidence interval, 4 to 19; P<0.001). From Byuse et al 2006 The gene signature hazard ratio for overall survival was 2.79 (95% CI = 1.60 to 4.87) without adjustment and 2.69 (95% CI =1.53 to 4.73), 2.89	Buyse wasreviewed and

Public Reviewer & Affiliation	Section	Comment	Response
		(95% CI = 1.58 to 5.29), and 2.63 (95% CI =1.45 to 4.79) with the respective adjustments.	graded as high ROB and therefore excluded from our main analyses. Vd Vijier did not control for other prognostic factors in a multivariable model. Since the focus of the report was prognostic values in addition to traditional factors, unadjusted HRs were not used in the main analyses.
Bastiaan van der Baan Agendia, Inc.	Tables (Executive Summary)	(Table B, MammaPrint Breast OS Conclusions) Please change OS from insufficient to moderate See comments above on OS data. (Table B, MammaPrint Breast Decisions about Rx) Please change from Insufficient to Moderate Decisions about Rx: See comments on the Raster trial earlier in the document.	We respectfully decline the suggestion to change our assessment of the SOE.
Bastiaan van der Baan Agendia, Inc.	Table 2	Please change Frozen into Fresh frozen or FFPE Sample Requirements “frozen tissue” This information in incomplete. MammaPrint has also been validated for: RNAretail (fresh) preserved FFPE tissue Sapino et al MammaPrint molecular diagnostics on Formalin Fixed Paraffin Embedded tissue 2013 (accepted J Mol Diag) Bueno de Mesquita et al. *Lancet Oncol.* 2007; 8:1079---1087 http://www.accessdata.fda.gov/cdrh_docs/pdf7/K070675.pdf	We have done this
Bastiaan van der Baan Agendia, Inc.	Table 6	Please add missing data Data on analytical validity is incomplete. There is analytical validity data available in each of our 5 FDA clearances. http://www.accessdata.fda.gov/cdrh_docs/pdf6/K062694.pdf http://www.accessdata.fda.gov/cdrh_docs/pdf7/K070675.pdf	Per the reviewer’s comment, we added

Public Reviewer & Affiliation	Section	Comment	Response
		http://www.accessdata.fda.gov/cdrh_docs/pdf8/K080252.pdf http://www.accessdata.fda.gov/cdrh_docs/pdf8/K081092.pdf http://www.accessdata.fda.gov/cdrh_docs/pdf10/K101454.pdf A summary of the most important analytical and clinical validity data was recently published. Please include Delahaye et al *Personalized medicine* (2013) 10(8),801---811	Delahaye et al 2013 article to our full-text literature review and included the article because it met our inclusion criteria.
Bastiaan van der Baan Agendia, Inc.	Table 7	Please add missing studies Found 10 studies. There are 15 independent validation studies addressing the analytical validity, clinical validity and clinical utility. The following 5 studies are missing. We have not added the studies that address the predictive aspects of MammaPrint as your document aims to restrict itself to prognosis only. 1) isBS, Sgroi DC, Ryan PD *et al.* Analysis of the MammaPrint® breast cancer assay in a predominantly postmenopausal cohort. *Clin. Cancer Res.* 14(10), 2988–2993 (2008). 2) Ishitobi M, Goranova TE, Komoike Y *et al.* Clinical utility of the 70---gene MammaPrint® profile in a Japanese population. *Jpn J. Clin. Oncol.* 40(6), 508–512 (2010). 3) Glück et al (2013) Molecular subtyping of early---stage breast cancer identifies a group of patients who do not benefit from neoadjuvant chemotherapy. *Breast Cancer Res Treat*. 2013 Jun;139(3):759---67 4) Drukker CA, Bueno---de---Mesquita JM, Retèl VP *et al.* A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. *Int. J. Cancer* 133(4), 929–936 (2013). 5) Saghatchian M, Mook S, Pruneri G *et al.* Additional prognostic value of the 70---gene signature (MammaPrint®) among breast cancer patients with 4–9 positive lymph nodes. *Breast* 22(5), 682–690 (2012).	The Drukker et al 2013 article was reviewed as part of our draft report and excluded for ineligible outcome on pg. B-21. The Wittner et al 2008 article was reviewed as part of our draft report and excluded for ineligible outcome on pg. B-100. The Ishitobi et al 2010 article was reviewed as part of our draft report and excluded for wrong study design. Per the reviewer's comment, the Glück et al 2013 article was added to our full-text literature review and excluded for ineligible outcome on pg. B-27. Per the reviewer's comment, the Saghatchian 2012 article was added to our full-text literature review. Based on our inclusion criteria, we

Public Reviewer & Affiliation	Section	Comment	Response
			included this article in our final report.
Bastiaan van der Baan Agendia, Inc.	Table 9	Please add data There are three papers with CSS data missing: Vd Vijver et al 2002 Buyse et 2006 Saghatchian 2012	Per the reviewer's comment, the Saghatchian 2012 article was added to our full-text literature review. Based on our inclusion criteria, we included this article in our final report. Buyse and vd Vijier were reviewed and graded as high ROB and therefore excluded from our main analyses.
Bastiaan van der Baan Agendia, Inc.	Table 10	Please add data There are two papers with data on Overall Survival missing: Vd Vijver et al 2002 Buyse et al 2006 See comment 1 [*this is comment #1: Please adjust that the evidence for MammaPrint for OS is also sufficient HR for OS is also significant. From vd Vijver et al (2002) estimated hazard ratio for distant metastases as a first event in the group with a poor-prognosis signature as compared with the group with a good-prognosis signature over the entire follow-up period was 5.1 (95 percent confidence interval, 2.9 to 9.0; P<0.001); the prognosis profile was associated with a significantly higher hazard ratio during the first five years of followup (hazard ratio, 8.8; 95 percent confidence interval, 3.8 to 20; P<0.001) than after five years (hazard ratio, 1.8; 95 percent confidence interval, 0.69 to 4.5; P=0.24). The hazard ratio for overall survival was 8.6 (95 percent confidence interval, 4 to 19; P<0.001). From Byuse et al 2006 The gene signature hazard ratio for overall survival was 2.79 (95% CI = 1.60 to 4.87) without adjustment and 2.69 (95% CI =1.53 to 4.73), 2.89 (95% CI = 1.58 to 5.29), and 2.63 (95% CI =1.45 to 4.79) with the respective adjustments.*]	We respectfully decline your suggestion to change our assessment of the SOE Buyse was reviewed and graded as high ROB and therefore excluded from our main analyses. Vd Vijver was not included in the analyses for OD because it did not analyse OS after adjusting for traditional prognostic factors in a multivariavle model.

Public Reviewer & Affiliation	Section	Comment	Response
Bastiaan van der Baan Agendia, Inc.	Table 41	Please include missing data Analytical validity data is incomplete. FDA clearance data is publicly available for Analytical Validity Summary of the most important data was recently published http://www.accessdata.fda.gov/cdrh_docs/pdf6/K062694.pdf http://www.accessdata.fda.gov/cdrh_docs/pdf7/K070675.pdf http://www.accessdata.fda.gov/cdrh_docs/pdf8/K080252.pdf http://www.accessdata.fda.gov/cdrh_docs/pdf8/K081092.pdf http://www.accessdata.fda.gov/cdrh_docs/pdf10/K101454.pdf A summary of the most important clinical validity data from the FDA filings was recently published. Delahaye et al *Personalized medicine* (2013) 10(8),801---811	Same comment and refs as above. Per the reviewer's comment, we added Delahaye et al 2013 article to our full-text literature review and included the article because it met our inclusion criteria.
Bastiaan van der Baan Agendia, Inc.	Table 42	Please include missing data See earlier comments on missing data on OS vd VIjver et al 2002 Buyse et al 2006	Buyse and vd Vijier were reviewed and graded as high ROB and therefore excluded from our main analyses.
Bastiaan van der Baan Agendia, Inc.	Table 43	Please change strength of evidence to Moderate See earlier comments on Drukker et al en Rutgers et al	The Drukker et al 2013 article was reviewed as part of our draft report and excluded for ineligible outcome on pg. B-21. Per the reviewer's comment, we added the Rutgers et al 2013 study and excluded it from this report for ineligible outcome on pg. B-77.

Attachment 1. ColoPrint information from Agendia comments

Colorectal Cancer: ColoPrint

Characteristics of Included Studies Assessing ColoPrint

Three cohort studies examined the added prognostic value of ColoPrint for Relapse free Survival (RFS) in patients with CRC (Table xxx according to Table 33).

Table xxxx (according to Table 33). Characteristics of included studies: ColoPrint for colorectal cancer

Characteristics of included studies	N Length of follow-up	Country	Disease Stage Other characteristics	Study Groups	Race Ethnicity	Overall Age % Female
Salazar et al. JCO 2011	N=206 54 months	Spain	Stage II/ III Colorectal cancer	Recurrent and non-recurrent patients	White/ Hispanic	69 years 36%
Maak et al. Annals of Surgery 2013	N=135 101 months	Germany	Stage II Colon cancer	Recurrent and non-recurrent patients	White	65 years 43%
Kopetz et al. ASCO GI 2013 (manuscript submitted)	N=190 64 months	US	Stage II/ III Colorectal cancer	Recurrent and non-recurrent patients	unknown	62 years 39%

ColoPrint: Risk of Relapse

ColoPrint is a 18-gene expression signature that was developed based on unbiased gene selection, searching the whole genome for genes that have the highest correlation to a tumor relapse event (Saalzar et al. JCO 2011). The signature was translated into a

diagnostic test, was validated in three independent studies, and has been shown to be technically reproducible and robust (Maak et al, Annals of Surgery 2013).

Relapse Free Survival (RFS) was defined as survival until first event of local, regional or distant recurrence or death of cancer. In all studies, ColoPrint was the only significant prognostic factor in multivariate analysis when analyzed in stage II patients only and compared to known risk factors (MMR, age, gender, T-stage, numbers of lymph nodes assessed, grade, lymphatic/ angio-invasion) and clinical risk assessment as described in the NCCN guidelines (NCCN Guideline Version 2013.3 Colon Cancer).

Table xxxx (designed according to Table 34). Colorectal cancer: ColoPrint relapse-free survival

Study	Adjusted HR (95% CI), p-value	Variables Used in Multivariate Model
Salazar et al. JCO 2011 N=206	Stage II+III HR 2.51 (1.33-4.73) p=0.005 Stage II HR 3.34 (1.24-9) p=0.017	ColoPrint, MMR (MSI-status), age, gender, T-stage, numbers of lymph nodes assessed, grade, therapy lymphatic, perineural, venous invasion
Maak et al. Annals of Surgery 2013; N=135	Stage II HR 4.12 (1.3 –13.01) p= 0.009	ColoPrint, MMR (MSI-status), age, gender, T-stage, numbers of lymph nodes assessed, grade, lymphatic/ angio-invasion
Kopetz et al. ASCO GI 2013 N= 190	Stage II and III ColoPrint HR 2.55 (1.17-5.52) p=0.018	ColoPrint, MMR (MSI-status), age, gender, stage, numbers of lymph nodes assessed, grade, Lymphvascular/ Perineural invasion

Attachment 2. Abbott Molecular UroVysion Kit

Abbott
A Promise for Life

UROVYSION® KIT

AMD-00001216

SUMMARY OF EVIDENCE

UroVysion® Kit is a FDA-approved, patented assay with proprietary combination of probes – Chromosome Enumeration Probes (CEPs) 3, 7, 17 and Locus Specific Identifier (LSI) 9p21. This document includes current evidence / publication references for the UroVysion® Kit that were recently published. The evidence addresses:

A. **Impact to Patient Management and Health Economics**
   2013 publication to demonstrate use of UroVysion® Kit in cases of negative or equivocal cystoscopy and atypical cytology to decide whether to biopsy a patient significantly decreases bladder cancer costs. Additionally, it provides clinical utility for UroVysion® Kit in patient management to avoid unnecessary biopsies.

B. **Evidence of Clinical Utility & Support from US/International Clinical Practice Guidelines**
   a. Monitoring for Recurrence:
      Inclusion of FDA-Approved FISH in National Comprehensive Cancer Network (NCCN) Bladder Cancer Guidelines (v1. 2013) and Relevant Publications

   b. Diagnosis:
      Inclusion of FISH in 2013 International Consultation Urologic Diseases – European Association of Urology (ICUD – EAU) Recommendations and Relevant Publications

C. **Evidence of Clinical Utility**
   The role of the UroVysion® Kit in diagnosis of bladder cancer, and comparison to cytology

A. Impact to Patient Management and Health Economics
Use of UroVysion® Kit in cases of negative or equivocal cystoscopy and atypical cytology to decide whether to biopsy a patient results in significant decrease in bladder cancer costs. Additionally, it provides clinical utility for UroVysion® Kit in patient management to avoid unnecessary biopsies.

Bladder cancer is the 5th most expensive cancer in the US. Almost $ 3.4B is spent per year, of which the direct treatment related costs are about $ 2.9B. Much of this spend is related to surveillance costs. However, cystoscopy could miss around 10-30% of cancers and cytology has low sensitivity and often is atypical/indeterminate (1).

Gayed et al. (2013) combined data from two large prospective trials, **Lotan et al. (2008)** and **Schlomer et al., (2010)** leading to 263 patients (1, 2, 3). Upon analysis, they concluded that the use of the UroVysion® Kit assay "should be considered as a significantly more cost effective method in determining the role of biopsy in patients with atypical cytology and negative or equivocal cystoscopies" (1):

1. **FISH in Patients with Equivocal Cystoscopy and Atypical Cytology**
   Operating Room-Based Biopsies
   Cost Impact - Using the base case analysis, the paper found that biopsying everyone leads to a cost of $3267 / patient vs. doing so with patients with FISH results leads to $1527 / patient.

   Clinical Impact - When the clinical decision was based on FISH, there was a 68% decrease in biopsies.

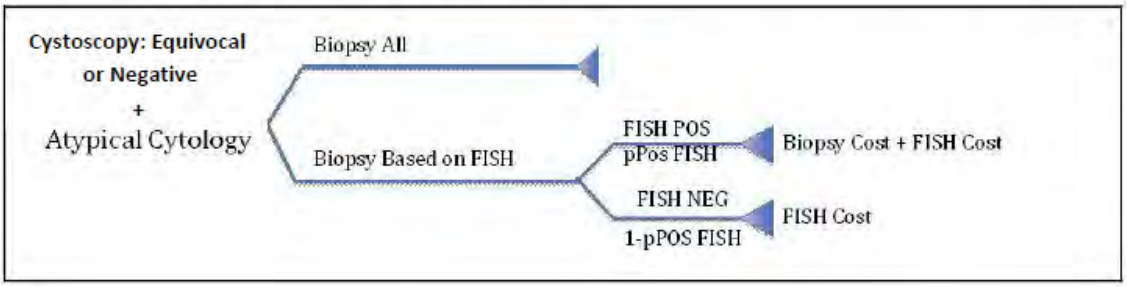
   Office-Based Biopsies
   Cost Impact - The paper found that biopsying everyone leads to a cost of $836 / patient vs. doing so with patients with FISH results leads to $741 / patient.

2. **FISH in Patients with Negative Cystoscopy and Atypical Cytology**
   Operating Room-Based Biopsies
   Cost Impact - Using the base case analysis, the paper found that biopsying everyone leads to a 3x cost of $3267 / patient vs. doing so with patients with FISH results leads to $1026 / patient.

   Clinical Impact - When the clinical decision was based on FISH, there was 83% decrease in biopsies.

   Office-Based Biopsies
   Cost Impact - The paper found that biopsying everyone leads to a cost of $836 / patient vs. doing so with patients with FISH results leads to $620 / patient.

MODEL USED TO ASSESS BIOPSY COSTS FOR EQUIVOCAL / NEGATIVE CYSTOSCOPY



B. Evidence of Clinical Utility and Support from US Clinical Practice Guidelines
a. Monitoring for Recurrence:
Inclusion of FDA-approved FISH in 2013 NCCN Guidelines and Relevant Publications

1. NCCN Bladder Cancer Guidelines

FDA-approved FISH was included in the National Comprehensive Cancer Network's (NCCN's) guidelines for bladder for monitoring for recurrence:

NCCN Bladder Cancer Guidelines v1. 2013
"Consideration may be given to FDA-approved urinary biomarker testing by fluorescence in situ hybridization (FISH) or nuclear matrix protein 22 (NMP22) for monitoring for recurrence." (2b) (4)

NCCN Bladder Cancer Guidelines v2. 2012
"Because the clinical benefit of ploidy, vascularity, p53 status, other urinary tumor markers (e.g., NMP-22, BTA, M344), and chromosomal alterations by FISH is uncertain, they are not used to guide treatment decisions outside of the experimental protocol setting." (5)

This update is a significant move in recognizing the role that FISH could play in guiding treatment decisions considering the prior guidelines referred to FISH being used in experimental protocol setting. Currently the grading of the new 2013 guidelines is 2b. This is similar to that breast cancer assay, Oncotype Dx, by most payors.

2. Kamat et al., 2012 – Prospective study demonstrating that FISH (UroVysion Kit) can identify patients at risk for tumor recurrence and progression during BCG immunotherapy (6)

The update to NCCN guidelines referred to the **Kamat et al. (2012)** publication from the MD Anderson Cancer Center published in 2012. This publication mentions that cytology and cystoscopy rely on detection of actual tumor recurrence and are poor predictors of therapy failure. The UroVysion® Kit is DNA based and is not affected by BCG. In comparison to Cystoscopy / Histology for detection of bladder cancer in patients on BCG therapy within 3 months completion from initiation of therapy, the UroVysion® Kit shows a Negative Predictive Value of 94.1% and a clinical sensitivity of 92.3%. Kamat et al. found positive FISH results identified patients who were 3 to 5 times more likely to develop tumor recurrence as compared to patients with negative FISH results. The authors concluded results of FISH assays correlated with the risk of tumor recurrence. The earlier a FISH result converted to positive from a negative baseline, the higher the risk of recurrence and progression (a positive FISH result at 6 weeks indicated a 50% overall risk of recurrence and a 30% overall risk of disease progression). The authors' conclusions from this study were *"patients can be counselled with even greater accuracy based on individual history of FISH results"*. Finally, in patients who do not respond to BCG therapy, radical cystectomy can improve bladder cancer patient survival by 20% in patients when performed within 24 months after diagnosis

3. In addition, multiple studies evaluated the use of FISH (UroVysion® Kit) in monitoring the response to intravesical therapies in patients with high-risk superficial bladder tumor (HRSBT). (57;58) Each study (described below) concluded that a positive FISH test at the end of intravesical therapy was predictive of eventual relapse, with one study also showing higher chance of progression of disease:
   a. **Kipp et al. (2005)** studied US patients prospectively and concluded that patients with a positive FISH at the end of treatment were at high risk for progression to muscle invasive bladder cancer (7)
   b. **Mengual et al. (2007)** studied Spanish patients prospectively and concluded FISH appeared to be useful for the surveillance of patients with HRSBT following BCG therapy. HRSBT patients could be monitored more carefully and treated more aggressively to prevent tumor relapse, progression and metastasis (8)
   c. **Whitson et al. (2009)** studied US patients retrospectively and concluded that in patients with high-risk superficial bladder tumors undergoing intravesical therapy, a positive UroVysion test after treatment is highly predictive of recurrence, even in a multivariate mode. (9)

B. Evidence of Clinical Utility and Support from International Clinical Practice Guidelines
b. Diagnosis and Screening:
Inclusion of FISH in 2013 ICUD - EAU Recommendations and Relevant Publications

1. International Consultation of Urologic Diseases (ICUD) Guidelines

The 2nd International Consultation on Bladder Cancer on the screening, diagnosis and the role of molecular markers in bladder cancer was held at the 2011 European Association of Urology Congress. This consultation developed recommendations using evidence-based strategy involving large retrospective and prospective studies. (10)

"The role of urinary markers – in particular FISH (fluorescence in situ hybridization) – appears to be most useful in the setting of a negative cystoscopy and atypical cytology (Grade C)"

Kamat A, et al. ICUD-EAU International Consultation on Bladder Cancer 2012: Screening, Diagnosis and Molecular Markers. 2013. European Urology. 63: 4-15

2. Two prospective studies have demonstrated that atypical and equivocal cytology can be categorized by the UroVysion® Kit and help clinicians to *"help avoid unnecessary evaluation in patients with atypical cytology and equivocal or negative cystoscopy, while identifying those who would need further evaluation."* It is also a *"worthwhile approach in patients with previous CIS, at high risk for the development of CIS, or previous unequivocal cytology suggestive of CIS, especially during or shortly after instillation therapy."*(8, 9)
   a. The prospective study by **Lotan et al. (2008)** (2) evaluated the clinical usefulness of the UroVysion® Kit for diagnosing patients with atypical cytology who were at risk for bladder cancer. In this study, 50 patients (of which 37 had hematuria) who underwent cystoscopy and cytology with atypical or suspicious cytology underwent a reflex FISH test. This prospective evaluation of a reflex FISH test in patients with atypical cytology showed that UroVysion® Kit was unnecessary in patients with obvious tumors on cystoscopy but beneficial in patients with equivocal or negative cystoscopy. In this study, the UroVysion® kit identified all high grade cancers (2)
   b. Similarly, a US prospective study by **Schlomer et al. (2010)** (3) performed a reflex FISH test on every patient who underwent cystoscopy and cytology with atypical cytology for the diagnosis of bladder cancer (of which 63% had hematuria). A comprehensive review was then performed to evaluate clinical and pathological data on each patient. In this study, the UroVysion® Kit identified all urothelial carcinoma tumors in patients with equivocal or negative cystoscopy (3)

3. **Yoder *et al* (2007)** (11) followed up 250 patients to understand the course of these cases and the time to bladder tumor recurrence. The data presented in this 2007 study showed that performing reflex FISH analysis on routine urine cytologic specimens allowed early detection of recurrent urothelial carcinoma in patients with a negative cystoscopic and cytologic diagnosis. In this study, authors also estimated that approximately 26% of patients with negative cystoscopic results will have a positive concurrent FISH result. According to the data presented, recurrent urothelial carcinoma will develop in approximately 50% to 80% of this population within 29 months of a positive FISH result. Yoder *et al.* also estimated that in 34% to 62% and 40% to 67% of the population, recurrent urothelial carcinoma developed within 6 and 10 months, respectively, of the positive FISH result. Yoder *et al.* concluded that the UroVysion® Kit was "an excellent adjunct to ThinPrep-based urine cytology, with the capacity to detect recurrent urothelial carcinoma before cystoscopically visible lesions can be identified and to resolve equivocal cytologic findings". (11)

4. A 2006 literature review by **Jones *et al.*** (12) evaluated the role of FISH in bladder cancer surveillance. Jones *et al.* found FISH outperformed conventional cytology across all stages and grades in all published reports, and FISH detected abnormalities before the development of lesions visible by cystoscopy. One of the publications found in this literature review demonstrated that the reflex FISH concept adds significant sensitivity to cytology, detecting cancer in 85% of false-negative cytology cases. (12)

C. Evidence of Clinical Utility
The role of the UroVysion Kit in diagnosis of bladder cancer, and comparison to cytology

Historically cytology has been less sensitive in all stages and grades of disease of bladder cancer vs. UroVysion® Kit FISH. (13, 14, 15) Additionally, cytology can be reported out as atypical or equivocal in which the clinician has no definitive answer. As previously mentioned, the UroVysion® Kit compares performance of UroVysion vs. Cystoscopy / Histology for detecting bladder cancer recurrence by tumor stage and grade, and UroVysion demonstrates 75% and 73.5% clinical sensitivity across all stages and grades respectively. The UroVysion® Kit shows greatest clinical sensitivity (100%) among tumors T2 and Tis, when compared to cystoscopy and cytology. (16)

UroVysion® Kit (FISH) vs. Cytology in the Detection of Urothelial Carcinoma

	Halling *et al.* (17)		Sarosdy *et al.* (16)		Junker *et al.* (18)		Laudadio *et al.* (19)	
	Cytology	FISH	Cytology	FISH	Cytology	FISH	Cytology	FISH
Sensitivity								
Stage								
pTa	17/36 (47%)	24/37 (65%)	8/32 (25%)	21/32 (66%)	N/A (15%)	N/A (36.1%)	N/A	N/A
pTis	14/18 (78%)	17/17 (100%)	2/6 (33%)	7/7 (100%)	N/A	N/A	N/A	N/A
pT1	9/15 (60%) [1]	18/19 (95%) 1	4/6 (67%)	5/6 (83%)	N/A (25.7%)	N/A (65.2%)	N/A	N/A
pT2	N/A	N/A	1/3 (33%)	3/3 (100%)	N/A (66.7) [2]	N/A (100) 2	N/A	N/A
	Halling *et al.* (17)		**Sarosdy *et al.* (16)**		**Junker *et al.* (18)**		**Laudadio *et al.* (19)**	
	Cytology	**FISH**	**Cytology**	**FISH**	**Cytology**	**FISH**	**Cytology**	**FISH**
Grade								
1	3/11 (27%)	4/11 (36%)	4/22 (18%)	12/22 (55%)	N/A (14%)	N/A (37%)	8/25 (32%) [3]	14/25 (56%) 3
2	13/24 (54%)	19/25 (76%)	4/9 (44%)	7/9 (78%)	N/A (40%)	N/A (65.4%)	N/A	N/A
3	24/34 (71%)	36/37 (97%)	7/17 (41%)	17/18 (94%)	N/A (50)	N/A (91.7)	7/17 (41%) [4]	18/19 (95%) [4]
Specificity								
	85%	76%	N/A	66%	90.5%	82.6%	93%	65%

[1] T1-T4
[2] T2-3
[3] Low
[4] High

References

1. Gayed BA, Seideman C, Lotan Y. Cost Effectiveness of Fluorescence in Situ Hybridization in Patients with Atypical Cytology for the Detection of Urothelial Carcinoma. J Urol 2013; Acceptance Date: 29 March 2013.
2. Lotan Y, Bensalah K, Ruddell T, Shariat SF, Sagalowsky A, Ashfaq R. Prospective evaluation of the clinical usefulness of reflex fluorescence in situ hybridization assay in patients with atypical cytology for the detection of urothelial carcinoma of the bladder. J Urol 2008; 179(6):2164-2169.
3. Schlomer BJ, Ho R, Sagalowsky A, Ashfaq R, Lotan Y. Prospective validation of the clinical usefulness of reflex fluorescence in situ hybridization assay in patients with atypical cytology for the detection of urothelial carcinoma of the bladder. J Urol 2010; 183(1):62-67.
4. NCCN Bladder Cancer Guidelines v1. 2013
5. NCCN Bladder Cancer Guidelines v2. 2012
6. Kamat AM, Dickstein RJ, Messetti F, Anderson R, Pretzsch SM, Gonzalez GN et al. Use of fluorescence in situ hybridization to predict response to bacillus Calmette-Guerin therapy for bladder cancer: results of a prospective trial. J Urol 2012; 187(3):862-867.
7. Kipp BR, Karnes RJ, Brankley SM, Harwood AR, Pankratz VS, Sebo TJ et al. Monitoring intravesical therapy for superficial bladder cancer using fluorescence in situ hybridization. J Urol 2005;(173):401.
8. Mengual L, Marin-Aguilera M, Ribal MJ, Burset M, Villavicencio H, Oliver A et al. Clinical utility of fluorescent in situ hybridization for the surveillance of bladder cancer patients treated with bacillus Calmette-Guerin therapy. Eur Urol 2007; 52(3):752-759.
9. Whitson J, Berry A, Carroll P, Konety B. A multicolour fluorescence in situ hybridization test predicts recurrence in patients with high-risk superficial bladder tumors undergoing intravesical therapy. BJU Int 2009;(104):336.
10. Kamat A, et al. ICUD-EAU International Consultation on Bladder Cancer 2012: Screening, Diagnosis and Molecular Markers. 2013. European Urology. 63: 4-15.
11. Yoder BJ, Skacel M, Hedgepeth R, Babineau D, Ulchaker JC, Liou LS et al. Reflex UroVysion testing of bladder cancer surveillance patients with equivocal or negative urine cytology: a prospective study with focus on the natural history of anticipatory positive findings. Am J Clin Pathol 2007;(127):295-301.
12. Jones JS. DNA-based molecular cytology for bladder cancer surveillance. Urology 2006; 67(3 Suppl.):35-45, discussion: 45-7.
13. Abbott Molecular. UroVysion Bladder Cancer Kit, Package Insert. 30-608385/R4. 2012.
14. Bubendorf L, Grilli B, Sauter G, Mihatsch MJ, Gasser TC, Dalquen P. Multiprobe FISH for enhanced detection of bladder cancer in voided urine specimens and bladder washings. Am J Clin Pathol 2001;(116):79.
15. Yoder BJ, Skacel M, Hedgepeth R, Babineau D, Ulchaker JC, Liou LS et al. Reflex UroVysion testing of bladder cancer surveillance patients with equivocal or negative urine cytology: a prospective study with focus on the natural history of anticipatory positive findings. Am J Clin Pathol 2007;(127):295-301.
16. Sarosdy MF, Schellhammer P, Bokinsky G, Kahn P, Chao R, Yore L et al. Clinical evaluation of a multi-target fluorescent in situ hybridization assay for detection of bladder cancer. J Urol 2002; 168(1950).
17. Halling KC, King W, Sokolova IA, Meyer RG, Burkhardt HM, Halling A et al. A comparison of cytology and fluorescence in situ hybridization for the detection of urothelial carcinoma. J Urol 2000;(164):1768.
18. Junker K, Fritsch T, Hartmann A, Schulze W, Schubert J. Multicolor fluorescence in situ hybridization (M-FISH) on cells from urine for the detection of bladder cancer. Cytogenet Genome Res 2006; 114(3-4):279-283.
19. Laudadio J, Keane TE, Reeves HM, Savage SJ, Hoda RS, Lage J et al. Fluorescence in situ hybridization for detecting transitional cell carcinoma: implications for clinical practice. BJU Int 2005; 96(9):1280-1285.

Attachment 3. Appendix of Tables from Genomic Health, Inc.

Appendix

Table 1: Variability and Reproducibility, Oncotype DX Breast Cancer Assay

Study, year	Aims and Methods	Results	Conclusions
Cronin, 2007	To assess individual gene and RS reproducibility Repeated measurements of 2 aliquots of a single RNA across multiple days, operators, RT-PCR plates, 7900HT instruments, and liquid-handling robots Mixed-effect ANOVA was used to estimate components of variance	Reproducibility, CT measurements SD for the individual genes: Total SD range: 0.06 to 0.15 CT units Between day SD range: 0 to 0.055 CT units Between plate SD range: 0 to 0.090 CT units Within plate SD range: 0.057 to 0.147 CT units At a CT of 30 a maximum SD of 0.15 translates into a CV of 0.5% The largest differences between operators, liquid handling robots, and 7900HT instruments < 0.5 CT Reproducibility, CT measurements SD for the RS: Total SD was 0.792 RS units Between day SD was 0 RS units Between plate SD was 0 RS units Within plate SD was 0.792 RS unit	Authors reported that the following procedures were performed to assure the reproducibility of the assay: A standard RNA control sample is assayed at least once per batch of patients (46 samples) PCR controls are run in every assay plate RT-PCR failures are excluded from analysis Expression values are assigned when at least 2 of 3 assay wells provide acceptable RT-PCR results All 21 genes must have an expression value assigned for an RS to be calculated and reported
Habel, 2006	To assess RS reproducibility Pearson's correlation and ANOVA to assess within-patients correlation and variability: 60 blocks that did not undergo macro-dissection from 20 patients (2 to 5 blocks per patient); 49 core biopsies or tumor resection blocks	RS (as a continuous value) SD and Pearson's correlation observed in two unpublished studies: Overall between blocks SD was 3.0 RS units For 16 of the 20 patients, the between blocks SD was < 2.5 RS units Pearson's correlation = 0.86 Similar results from the second study	
Paik, 2004	To assess individual genes and RS reproducibility Reproducibility within and between blocks was assessed by performing the assay in five serial sections from six blocks in two patients	Reproducibility evaluation: 16 Cancer genes SD ranged from 0.07 to 0.21 CT units; Within-block RS SD = 0.72 RS unit (95% CI = 0.55 to 1.04); Total within-patient SD (including between and within-block SD) = 2.2 RS units; Similar variability in the RS was observed in reanalysis of clinical trial samples on separate days with different reagent lots (data not shown).	

Abbreviations: RS = Recurrence Score result; RNA=ribonucleic acid; RT-PCR = reverse transcriptase polymerase chain reaction; ANOVA=analysis of variance; CT = cycle threshold; SD = standard deviation

Table 2a: Oncotype DX Breast Cancer Assay Risk of Cancer Recurrence (Disease-Free Survival, Recurrence-Free Survival, Distant Metastasis-Free Survival)

Author, Year, Design, Number of Patients	Estimates (95% CI), p-values	Variables Used in Multivariate Model
Paik, 2004 Cohort 668	Distant Recurrence HR: 3.21 (2.23 go 4.61) per 50 RS units, p<0.001	RS, age, clinical tumor size
Dowsett, 2010 Cohort 1,231	Distant Recurrence N0 patients HR: 5.25 (2.84 to 9.73) per 50 RS units, p<0.001 N+ patients HR: 3.47 (1.64 to 7.38) per 50 RS units, p<0.001	 RS, tumor size (>2 vs. ≤2 cm), central grade (well, moderate, poor), age (<65 vs. ≥65) and treatment (tamoxifen vs. anastrozole) RS, tumor size (>2 vs. ≤2 cm), central grade (well, moderate, poor), age (<65 vs. ≥65), treatment (tamoxifen vs. anastrozole) and positive nodes (≥4 vs. 1-3)
Paik, 2006 Cohort 651	Distant Recurrence RS by treatment (chemohormonal vs. hormonal only) interaction p=0.038 HR for treatment 1.31 (0.46 to 3.78), 0.61 (0.24 to 1.59), 0.26 (0.13 to 0.53) for RS<18, RS 18-30, RS≥31, resp.	
Albain, 2010 Cohort 367	Disease-free survival HR 2.64 (1.33 to 5.27) per 50 RS units, p=0.006 RS by treatment (chemohormonal vs. hormonal only) interaction in the first 5 years (p=0·029); no additional prediction beyond 5 years (p=0·58), although cumulative benefit remained at 10 years. HR for treatment 1.02 (0.54 to 1.93), 0.72 (0.39 to 1.31), 0.59 (0.35 to 1.01) for RS<18, RS 18-30, RS≥31, resp.	RS, number of positive nodes
Toi, Cohort 2010	Distant Recurrence HR 6.03 (2.17 to 16.7) per 50 RS units, p<0.001	RS, age, clinical tumor size

Table 2b: Oncotype DX Breast Cancer Assay Cancer-Specific Survival

Author, Year, Design, Number of Patients	Estimates (95% CI), p-values	Variables Used in Multivariate Model
Habel, 2006 Case-control 220 cases, 570 controls	Tamoxifen treated OR: 5.3 (1.6 to 17.2) per RS 50 units, p=0.003 Tamoxifen untreated OR 2.4 (1.1 to 5.2) per RS 50 units, p=0.025	Matched for age, race, calendar year of diagnosis, and tamoxifen treatment (yes/no). Adjusted for tumor size (continuous) and grade (well, moderate, poor).

Table 3: Clinical Evidence Studies of the Oncotype DX Assay for Invasive Breast Cancer

Study	Prognosis	Prediction	Validation/Confirmatory vs. Supportive	Clinical Outcome Assessed
Paik 2004 (NSABP B-14)	X		Validation/Confirmatory	10-year distant recurrence rate (6.8% , 14.3% , 30.5 % for low, intermediate, high Recurrence Score risk groups)
Gianni 2005		X	Supportive	Pathologic complete response to neoadjuvant chemotherapy (high Recurrence Score result was associated with higher likelihood of pCR; p=0.005)
Paik 2006 (NSABP B-20)		X	Validation/Confirmatory	10-year distant recurrence rate (risk reduction from the addition of chemotherapy in the high risk group; no demonstrable relative risk reduction in the low risk group (0.26 vs 1.31))
Habel 2006 (Kaiser)	X		Validation/Confirmatory	10-year risk of breast cancer death (2.8% , 10.7%, 15.5% for low, intermediate, and high risk groups)
Chang 2007 (MD Anderson)		X	Supportive	Clinical complete response to neoadjuvant chemotherapy (> 1.7-fold increase for patients with high vs low score)
Goldstein 2008 (ECOG 2197)	X		Supportive	11.5-year disease free recurrence (score was a significant predictor of recurrence including node negative and node positive disease; p< 0.0001)
Akashi-Tanaka 2009		X	Supportive	5-year recurrence free survival (100%, 84%, and 73% for low, intermediate, and high risk groups)
Albain 2010 (SWOG 8814)	X	X	Validation/Confirmatory	10-year disease free survival (60% vs 43% for low vs high risk groups); 10-year breast cancer specific survival (low 92% Tam vs 87% CAF-Tam and high 54% Tam, vs 73% CAF-Tam; test for interaction between score and treatment p=0.021)
Dowsett 2010 (TransATAC)	X		Validation/Confirmatory	9-year distant recurrence (4%, 12%, and 25% in low, intermediate, and high node-negative risk groups, 17%, 28%, and 49% in node-positive risk groups)
Toi 2010	X		Validation/Confirmatory	10-year distant recurrence risk (3.3%, 0%, 24.8% for low, intermediate, high risk node-negative groups; low vs high p<.001)
Ueno 2013		X	Supportive	Clinical response to neoadjuvant hormonal therapy (59%, 59%, 20% for low, intermediate, and high risk groups)
Yardley 2011		X	Supportive	Pathologic complete response to neoadjuvant chemotherapy (0%, 0%, 26% for low, intermediate, and high risk groups; Mantel-Haenzel chi-square p=0.002)
Mamounas 2012 (NSABP B-28)	X		Supportive	10-year distant recurrence free interval (75.8%, 57.0%, and 48.0% for low, intermediate, and high risk groups; p<0.001)

Table 4: Clinical Guidelines Describing Use of the Oncotype DX Assay for Invasive Breast Cancer

National Comprehensive Cancer Network® (NCCN®)	Consider use in > 0.5 cm, hormone receptor positive, HER2-negative disease (pT1, pT2, or pT3; pN0 and pN1mi (≤ 2 mm axillary node metastasis)
American Society of Clinical Oncology® (ASCO®)	The Oncotype DX assay may be used in newly diagnosed patients with node-negative, estrogen-receptor positive breast cancer: • to predict risk of recurrence in patients treated with tamoxifen • to identify patients who are predicted to obtain the most therapeutic benefit from adjuvant tamoxifen and may not require adjuvant chemotherapy Patients with high Recurrence Score results appear to achieve relatively more benefit from adjuvant chemotherapy (specifically CMF) than from tamoxifen.
St Gallen Consensus	The Oncotype DX assay has been shown to predict chemotherapy benefit among patients with hormone receptor positive disease
European Society for Medical Oncology (ESMO®)	The Oncotype DX Recurrence Score result may be used to gain additional prognostic and/or predictive information to complement pathology assessment and to predict response to adjuvant chemotherapy, in particular in patients with ER-positive early breast cancer.
NICE	Includes the Oncotype DX Breast Cancer Assay as an option to guide chemotherapy treatment decisions for qualified early-stage breast cancer patients

American Society of Clinical Oncology (ASCO), National Comprehensive Cancer Network (NCCN), and European Society for Medical Oncology (ESMO) are trademarks of ASCO, NCCN, and ESMO respectively. ASCO, NCCN, ESMO, NICE, and St Gallen guidelines do not endorse any product or therapy.

Table 5: Decision Impact Trials Utilizing the Oncotype DX Breast Cancer Assay

Study	Country	Type	Patients	Nodal status	Change in Treatment Recommendation (%)
Albanell	Spain	Prospective	107	N0	31.8
Davidson	Canada	Prospective	150	N0	30
Lo	US	Prospective	89	N0	31.5
Asad	US	Retrospective	85	N0	44
Klang	Israel	Retrospective	313	N0	39.9
Oratz	US	Retrospective	68	N0	33.8
Gligorov	France	Prospective	96	N0/N1mi	36
Holt	England	Prospective	142	N0/N1mi	26.8
Eiermann	Germany	Prospective	366	N0/N1	33.1
Bargallo	Mexico	Prospective	96	N0/N1	32
De Boer	Australia	Prospective	151	N0/N1	23.8
Yamauchi	Japan	Prospective	90	N0/N1	37.8
Oratz	US	Retrospective	160	N1	51

**N0- node-negative; N1- node-positive*

Table 6: Health Economic Studies Utilizing the Oncotype DX Breast Cancer Assay

Country	Country Threshold (Willingness to pay for 1 QALY($))	Reported Findings (ICER in cost per QALY gained)	Impact
Ireland	EUR 20,000	EUR 9,462	Cost Effective
UK	GBP 20,000	GBP 6,232	Cost Effective
Israel	USD 35,000	USD 10,700	Cost Effective
Canada	CAD 75,000	CAD 63,421	Cost Effective
Canada	CAD 75,000	> CAD 29,000	Cost Effective
Japan	USD 50,000	USD 3,848	Cost Effective
Canada	CAD 75,000	CAD 9,591	Cost Effective
Hungary	EUR 12,600-25,300	EUR 9,730	Cost Effective
Australia	AUS 18,000	AUS 9,986	Cost Effective
Singapore	Improved outcomes (QALYs), reduced costs		Cost Saving
France			Cost Saving
Germany			Cost Saving
USA			Cost Saving
USA			Cost Saving
USA			Cost Saving

Table 7a: Oncotype DX Colon Cancer Assay Risk of Cancer Recurrence (Disease-Free Survival, Recurrence-Free Survival, Distant Metastasis-Free Survival)

Author, Year, Design, Number of Patients	Estimates (95% CI), p-values	Variables Used in Multivariate Model
Gray, 2011 Cohort 1436	Recurrence HR: 1.43 (1.11 to 1.83) per interquartile range of 18.2 units [equivalent to HR of 1.63 (1.15 to 2.29) per 25 units], p=0.006	RS, T stage, MMR status, number of nodes examined, tumor location, tumor grade, lymphovascular invasion, and age
Venook, 2013 Cohort 690	Recurrence HR 1.68 (1.18 to 2.38) per 25 units, p=0.004	RS, T-stage, MMR status, number of nodes examined, grade, and lymphovascular invasion
Yothers, 2013 Cohort 892	Recurrence HR 1.57 (1.19 to 2.08) per 25 units, p=0.001	RS, N stage (II, IIIA/B, IIIC), T stage, MMR status, number of nodes examined, grade, and treatment (FU vs FU+Ox)

Table 7b: Oncotype DX Colon Cancer Assay Overall Survival

Author, Year, Design, Number of Patients	Estimates (95% CI), p-values	Variables Used in Multivariate Model
Gray, 2011 Cohort 1436	HR: 1.23 (1.01 to 1.51) per interquartile range of 18.2 units [equivalent to HR of 1.33 (1.01 to 1.76) per 25 units], p=0.041	RS alone (unadjusted)
Yothers, 2013 Cohort 892	HR: 1.89 (1.46 to 2.44) per 25 units, p=0.001	RS alone (unadjusted)

Table 8: Decision Impact Trials Utilizing the Oncotype DX Colon Cancer Assay

Study	Country	Type	Patients	Change in Treatment Recommendation (%)
Cartwright	USA	Retrospective	92	29.3
Srivastava	USA	Prospective	141	44.7
Kuchel	Israel	Prospective	269	37.9