Project Name: Systematic Reviews on Selected Pharmacogenetic Tests for Cancer Treatment: *CYP2D6* for Tamoxifen in Breast Cancer, *KRAS* for anti-EGFR antibodies in Colorectal Cancer, and *BCR-ABL1* for Tyrosine Kinase Inhibitors in Chronic Myeloid Leukemia
Project ID: GENC0609

Table 1: Invited Peer Reviewer Comments

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
Peer Reviewer 1	CYP2D6	The authors do a nice job of collecting relevant studies which analyze the association between cyp2d6 status and outcomes on tamoxifen in the adjuvant breast cancer setting (and the one study in the metastatic setting). As they point out, these studies are difficult to compare and combine because they used varying definitions of "extensive," "intermediate" and "poor" metabolizing alleles. Moreover the analyses compare different combinations of these groups. This topic is further complicated by the use of SSRIs in many of these women which can change their metabolizer status. Many of the studies reported did not control for the use of SSRIs in their analyses.	We thank the reviewer for the kind comments. We agree that there is irreconcilable heterogeneity between studies and definitions used; this is the main reason we refrained from quantitative analyses.
Peer Reviewer 1	CYP2D6	While the authors do present the problems associated with variable definitions in these studies, their conclusions are not helpful. While the evidence for routine testing for metabolizer status may not yet be warranted, the question is far from fully vetted. Tamoxifen is one of the most widely-used drugs for treatment of breast cancer, and the question of efficacy modulation is still an important one. It is compelling that many of these studies across different ethnicities did have positive associations. And instead of dismissing these studies, a proposal	The TA summarizes the evidence on existing studies and reaches conclusions based on the available evidence. We agree that standardization of definitions, further studies and individual-patient data meta-analyses would benefit the CYP2D6 field. Regarding false positive associations (type I error) we have added the following comment: "First, most studies are relatively small and thus probably underpowered to detect what would be a

		to standardize definitions of metabolizer status could be added to the conclusion. The role of chance versus true associations should also be further discussed.	plausible effect size for modification of response to tamoxifen and susceptible to type I error (false positive findings)."
Peer Reviewer 1	CYP2D6	The authors mention that no formal interaction analysis was performed using women who were not treated with tamoxifen. The Schroth et al. publication did report the lack of association among women who did not receive tamoxifen. This finding should be added on page 23.	The study by Scroth did not perform (or at least did not report) and interaction test comparing the pharmacogenetic effect in tamoxifen treated and non-treated women. Results for the no tamoxifen arm are not presented, precluding us from conducting the interaction analysis. We have added a note that "Differences were not observed in the control group".
Peer Reviewer 2	General	In general, I found the reports to be clear, well organized, and appropriate in scope and methods.	We thank the Peer Reviewer for his kind comments.
Peer Reviewer 2	Summary	In the second paragraph of the background section in the summary and on page 2, paragraph 1 of the main text, reference is made to the Medicare beneficiary population. Consideration of PICO framework – (P) patient population; (I) intervention; (C) comparison: (O) outcome _ might be helpful – I, C and O components seem fine.	[no response needed]
Peer Reviewer 2	CYP2D6	Search was limited to MEDLINE until 24 August 2009. There appear to be a few more relevant articles since that time, for example, for CYP2D6, tamoxifen and breast cancer, which I found using the HuGENet Published Literature database, limited to pharmacogenomics, mammary neoplasms and CYP2D6. What was the rationale for not considering other databases such as EMBASE? I accept that these tend not to add much for genetic association studies compared with Medline/PubMed, but in the context of cancer therapy, it is possible that evaluations of treatment outcomes done by organizations such as EORTC, in which genotyping might have been performed secondarily, might be published in journals indexed in EMBASE and not in	We have updated the search strategies and data extraction to include data up to March, 2010 for the CYP2D6 and the KRAS systematic reviews. Regarding the EMBASE database, we agree that for biomarkers and genetic association studies EMBASE may not substantially increase the sensitivity of searches while reducing specificity.

		MEDLINE.	
Peer Reviewer 2	General	I was a bit unsure of the expected audience. Some of the information that is specific to cancer therapeutics might be unfamiliar to a group that was charged with more generic decision making about health technologies, while some that is specific to genetic testing might be unfamiliar to health professions handling cancer treatment decisions (e.g. on page 2 of main text, final paragraph, there is the clarification that an enzyme is a protein, but not of "chimeric oncogene".	We agree that the understanding of pharmacogenetic tests will require knowledge from diverse fields including clinical epidemiology, genetics and medicine. We have provided a "plain words" explanation of "chimeric oncogene".
Peer Reviewer 2	Summary	In the summary and the introduction, it is stated "The challenges in the integration of cancer pharmacogenetics and targeted therapies in clinical practice require proof of benefit to the healthcare system, incorporating patient preferences, improving provider education, and anticipating potential ethical and social implications." I agree, but it reads as if benefit to the patient is not a consideration! I also suggest referring to balance between benefits, harms and affordability, especially in contexts of ageing populations facing escalating health costs, and in some subsets at least, drift towards more interventions.	We have revised the sentence to read: "The challenges in the integration of cancer pharmacogenetics and targeted therapies in clinical practice require proof of benefit to the patients (a favorable balance of harms and benefits of testing), cost-effectiveness for the healthcare system, incorporating patient preferences, improving provider education, and anticipating potential ethical and social implications."
Peer Reviewer 2	Summary	Table S1 – it would be useful to have not only numbers of studies, but also indication of volume of evidence (i.e. number of patients overall and/or in smallest subgroup)	Given the extensive overlap between studies these estimates would not be representative of the true amount of evidence and may be .
Peer Reviewer 2	Summary	Page S-5 – second bullet point under "study design issues" _ Agree with point about "repurposing" already completed RCTs. What about use of case-only design embedded in such a repurposed RCT?	In epidemiological studies of disease causation the case-only design, which requires only diseased subjects, allows for estimation of multiplicative interactions between factors known to be independent in the study population. This design requires assumptions to be made, namely an independence assumption between the environmental factors and the genetic marker. When this assumption holds, interactions can be assessed based only on affected individuals. In most of the studies

			we reviewed all individuals were “cases”, i.e. treated. One “equivalent” of the case-only design in the “repurposed RCT” framework would be to only consider the “treated patients” and assess interactions with the treatment effect. As we extensively discuss in the Report, the majority of studies made this assumption and did not utilize any “controls”, i.e. individuals not receiving the treatment of choice.
Peer Reviewer 2	Summary	Page 2 – it is stated that tamoxifen metabolites are “biotransformed through a complicated metabolic pathway, in which CYP2D6 is a leading enzyme”. I did not find this clear. In what sense is CYP2D6 “leading”?	We agree that the term “leading” may be inappropriate for metabolic pathways. We have rephrased the sentence to read: “They are biotransformed through a complicated metabolic pathway, in which CYP2D6 is a key enzyme”.
Peer Reviewer 2	Summary	Sentence starting at bottom of page 2: “Therefore there may be pharmacogenetic associations of mutations of the BCR-ABL1 gene with potential impact on management decisions.(4)” This was unclear.	We have revised the sentence to read: “Based of these observations, the detection of mutations of the *BCR-ABL1* gene has been proposed as a pharmacogenetic test with potential impact on management decisions.”
Peer Reviewer 2	Summary	Page 3, key question 2: As “gender” appears to refer to biological aspects of treatment response, I think the appropriate term is “sex”	We have used “sex” as a more appropriate term.
Peer Reviewer 2	Summary	Page 3: example under key question 3 unclear	Seems ok to me. Are we allowed to change the KQs?
Peer Reviewer 2	CYP2D6	Page 9, para 2: are aromatase inhibitors ever given concurrently with tamoxifen?	To the best of our knowledge co-administration of tamoxifen and aromatase inhibitors is not an approved use of the drugs.
Peer Reviewer 2	CYP2D6	Para 3, last line – what is point about “(rare)” after mutations?	We have revised the sentence to read: “Tamoxifen resistance has been extensively investigated and a variety of biological mechanisms are considered as potentially mediating treatment resistance, including cross talk of the ER/PR-activated pathway and growth-factor signaling pathways, activation of alternative (non-ER-dependent) signaling pathways, loss of ER expression and ER mutations (a rare cause of resistance).”
Peer Reviewer 2	CYP2D6	Page 12: last para – I understand the reluctance to	Given the heterogeneity of outcomes reported,

		perform quantitative mata-analysis. However, could this not have been done using the "simple" algorithm described in the last paragraph of the previous page, and then the effect of dropping each study in turn done as an influence analysis?	genotypes investigated, and results reported performing a meta-analysis of the set of CYP2D6-relevant studies would not be a valid approach.
Peer Reviewer 2	CYP2D6	Page 13, Fig 2: what does asterisk after "irrelevant" in bottom left box refer to?	There is a note at the bottom of the graph explaining the term "irrelevant". The note reads: * "Irrelevant" includes publications with no primary data, studies on healthy population, and studies on medications that inhibit CYP2D6.
Peer Reviewer 2	CYP2D6	Last paragraph - Avoid use of term "Caucasian" See Wikipedia on "Caucasian", and that it has generated a lot of debate. Also Bhopal R, Donaldson L. *Am J Pub Hlth* 1998; 88: 1303-7; Ma IW et al. *Journal of Clinical Epidemiology* 60 (2007) 572e578; Comstock RD et al. *Am J Epidemiol* 2004;159:611–619.	We have used the term "White" instead.
Peer Reviewer 2	CYP2D6	Page 14, line 1 – was there really no information about dosing in the RCTs?	We have qualified this statement to read: "Tamoxifen dosing was not reported in the majority of studies".
Peer Reviewer 2	CYP2D6	Last paragraph – I think it would be helpful to split off the study in the metastatic cancer setting – this was both the smallest study and had the shortest follow-up.	For the descriptive characteristics and genotyping methods and results, the data items we extracted from the metastatic setting study were the same as for the adjuvant setting studies. We agree with the Peer Reviewer that outcomes are different in the two settings and as such the outcomes tables are separate.
Peer Reviewer 2	CYP2D6	Page 15, Table 1 – Has overlap between the two Goetz studies been excluded? Could information be given TAM vs non_TAM for these?	There is substantial overlap between the two studies. For the CYP2D6 part of the report, given that no quantitative analysis was performed, we included all studies in the summary tables. We have made this explicit in the Methods and Results sections.
Peer Reviewer 2	CYP2D6	Page 16, col 2, rows 2-4, what is "RCS"?	There is a footnote at the bottom of the table spelling out these initials: Non-RCS = non-randomized comparative study.
Peer Reviewer 2	CYP2D6	Page 18 – further to comment on page 12, I accept that inferring metabolizer phenotype on basis of genotype complex, but would it not have been	Given the heterogeneity of outcomes reported, genotypes investigated, and results reported performing a meta-analysis of the set of CYP2D6-relevant studies

		possible also to do meta-analysis of ***4/***4 vs. wt/wt (5 studies), 10/10 vs wt/wt (3 studies) and **41/**41 vs wt/wt (also three studies), or are you arguing that this is pointless? (I accept that wt depends on what is tested for, but for 4/4, looks uniform apart from Schroth and Newman studies)	would not be a valid approach. We also suspect there is substantial potential for reporting bias. As such, any meta-analysis would be prone to identify spurious associations.
Peer Reviewer 2	CYP2D6	Page 19, Table 2: testing for departure from Hardy-Weinberg equilibrium mentioned without explanation. If variation in *CYP2D6* of etiologic importance in breast cancer, would equilibrium really be expected? (in case-control studies, often done in controls to indicate whether gross problem with genotyping, population stratification, other selection bias, but departure capable of many interpretations)	The review did not evaluate CYP2D6 variants as a causative factor for breast cancer. We evaluated these markers as predictors of response to treatment. It is true that if CYP2D6 polymorphisms were also associated with breast cancer development departure from HWE could occur. On the other hand, if the variants are not causally associated with breast cancer risk, then departure from HWE could indicate the problems highlighted from the Peer Reviewer or biased sampling from the study base.
Peer Reviewer 2	CYP2D6	Many footnotes incomplete in this table. Footnote a was unclear – what was the p value a test of?	Footnote was unclear and was removed.
Peer Reviewer 2	CYP2D6	Page 20, Kiyotani study – col 3 indicates available sample 67, genotyping success 100%, but total number in last column adds up to 58. Other totals seem to add up.	We have added the following footnote to explain the apparent discrepancy: "Other genotypes [*CYP2D6*
Peer Reviewer 2	CYP2D6	Footnote a – is this irrespective of genotype?	Cannot locate this comment
Peer Reviewer 2	CYP2D6	Pages 22-24: It would be helpful to comment on effects of adjustment in Tables 3-4, in view of comments on need for adjustment (Mendelian randomization principle) later.	We have added the following two comments in the relevant sections: "Many of the studies presented regression-adjusted estimates of the effect of CYP2D6 genotype on mortality risk, frequently for factors that could not confound the genotype-response association." "Many of the studies presented regression-adjusted estimates of the effect of CYP2D6 genotype on disease recurrence risk, frequently for factors that could not confound the genotype-response association."
Peer Reviewer 2	CYP2D6	Page 27, para 1, last line – "preventive setting" confusing.	This is statement verbatim from the American Society of Clinical Oncology and refers to the use of tamoxifen for

			the prevention of breast cancer occurrence in women at high risk of the disease.
Peer Reviewer 2	CYP2D6	Page 28 – is an issue the size of the study – if small, might there be departure from Mendelian randomization, just as can be inequaklities in co-variate diostributions between the arms of small RCTS?	"Mendelian randomization" is general principle governing segregation of alleles at miosis. Random fluctuations away from the proportions expected under Hardy Weinberg equilibrium, a related problem, is indeed a concern. Yet it can be argued than when such problems are due to small study sample sizes, adjustment for covariates will also be problematic due to sparse data.
Peer Reviewer 2	KRAS	In this section, numerous times spaces between words missing, and more typos than in previous sections. Needs thorough proof reading.	We have reviewed the section and attempted to correct all typographical errors.
Peer Reviewer 2	KRAS	Page 30, penultimate line – specify the tissue in which *KRAS* testing done – also applies to key question 1, page 31	Tables 10 and 11 report the tissue and type of material (fresh-frozen versus paraffin-embedded) used for DNA isolation. We have also collected information on whether metastatic or primary tumor foci were used for tissue collection and – from the few studies that reported relevant information – we have also collected whether primary and metastatic foci examination leads to the same mutational analysis results.
Peer Reviewer 2	KRAS	Page 34, line 3 – suggest changing "pre-treated" to "with metastatic disease who had previously been treated"	The suggested change was implemented.
Peer Reviewer 2	KRAS	Line 4 – "these" refers to what?	The sentence has been clarified. It now reads: "In studies conducted in the metastatic setting, the majority of patients had received prior treatment with at least one chemotherapy regimen; both the number and types of treatment regimens administered varied across studies"
Peer Reviewer 2	KRAS	Penultimate line before Table 7: specify drug dosing – may have changed by time the report is read	This sentence was rephrased to read: "Given that many of the patients in these studies were participants in larger, multicenter clinical trials, drug dosing in the studies included in this report can be expected to be similar to that employed in the prospective trials."

Peer Reviewer 2	KRAS	Page 36, Fig 5: nice, but text in ellipses virtually impossible to read	We have generated a new figure given the large number of studies captured by our search updata. The new figure is of much higher resolution.
Peer Reviewer 2	KRAS	Table 9: I like the way the table is ordered by study design. However, within the single arm studies, not sure of what order reflects. I thought it might be descending order of sample size, but last study in table larger than the previous six, so I'm flummoxed.	We have maintained the table structure based on design. All tables have been rearranged by year of publication. Studies with the same publication year are arranged by author name. In the rare case where year and author name are the same, we arranged studies by decreasing sample size.
Peer Reviewer 2	KRAS	Page 48, Table 10. After the three RCTs, I think a heading "single arm" and then Bengala reference missing.	We have corrected the Table. In addition, after obtaining author confirmation, we have re-classified the studies by Yen et al as first line studies. The studies did not report le specific line of treatment but Dr Wang, a corresponding author in one and author on both kindly provided this information. [Personal communication, Professor Jaw-Yuan Wang, MD, PhD Department of Surgery Kaohsiung Medical University and Hospital 100 Tzyou 1st Road Kaohsiung 807, Taiwan] Based on this re-classification, we have re-arranged the tables and repeated the meta-analyses.
Peer Reviewer 2	KRAS	More generally in Table 10, need (ref) after name and year, as in Table 9. This comment also applies to other tables in this section.	For all sections we have applied the policy of adding references after study authors only on the study characteristics Tables (Tables 1, 8 and 9, 20, 23 and 30-32). In all other tables the order of tables follows the order of the descriptive ones.
Peer Reviewer 2	KRAS	Page 54. Description of results of studies in first and second line studies mixed up, particularly for RCTs. Was difference in effect of treatment by presence of *KRAS* mutation in the same direction in the van Custem study as it was in the other?	The relevant quantitative analysis sections have now been updated to include the new studies identified by our search update. We have added the following comment: "The interaction test was non-significant in the study by

			van Cutsem 2009 (p=0.44) but the direction of effects was consistent."
Peer Reviewer 2	KRAS	Pages 55-57: I appreciate need to report "null results" etc, but could there be a better way to signal this than entries in the tables where everything except author, year and study arm is "NR"?	We can remove all the NR, NR (but I would prefer not).
Peer Reviewer 2	KRAS	Pages 61-64, Table 15: I may be over-interpreting the limited information I see in the Table, but median survival times seem to be longer in the three studies in which panitumumab was used (Amado, Freeman, Muro) than the other studies. Is this worth a comment?	Although this observation is accurate, there are substantial differences in the populations included in the studies we analyzed (high potential for confounding). As such, comparisons of median survivals between studies may not be valid and we refrain from including specific comments in the report.
Peer Reviewer 2	KRAS	Page 73, Fig 7. Needs (ref) after author, year. The order of studies (year, then alphabetical first author) is different from the tables – I would have liked to see grouping by design and whether 1st line or salvage treatment. I see at different points Yen, 2008 and Yen, 2009, but I think there is just one reference – please check.	The studies in the forest plot are ordered by year of publication and then by author name. An explanatory footnote has been added. Results by specific subgroups are reported in the subgroup analysis table (Table 19). It is imposible to present all subgroup analyses in one figure. Following our update of the search strategy, there are now two studies by Yen et al., one published in 2009 and one in 2010. We have corrected the publication year in the figures and tables.
Peer Reviewer 2	KRAS	Page 78: Is there any possibility that publication bias is relevant?	It is hard to discern what the Peer Reviewer refers to. The consistency and magnitude of effects in the KRAS topic is reassuring regarding the threat of publication bias.
Peer Reviewer 2	BCR-ABL	Page 79, line 2 – this is annual number of newly diagnosed cases, not incidence	The sentence now reads: "Chronic myeloid or myelogenous leukemia (CML) is a relatively uncommon hematological malignancy with approximately 5,000 new cases diagnosed annually."
Peer Reviewer 2	BCR-ABL	Page 85, last 3 lines, Fig 12 and Fig 13 (p.90, with brief ref in text in last line of p. 87). I think it would make more sense to deal with these points when describing the second line TKI studies. For the 1st line and 3rd line studies, could be deal with briefly in text.	We adopted the Peer Reviewer's suggestion. Now the relevant sentence reads: "Most publications particularly in 2nd-line TKI treatments originated from MD Anderson Comprehensive Cancer Center."

Peer Reviewer 2	BCR-ABL	Page 87, para 1, line 3: does “noncomparative cohorts” mean that they were single arm studies – I would keep that nomenclature from section 2.	We replaced non-comparative cohorts with single-arm studies, as suggested.
Peer Reviewer 2	BCR-ABL	Page 95, para 2, line 2: “some studies” – as far as I understand from the tables, patients had received interferon in all but one of the studies.	We changed the sentence as follows: In most studies, the vast majority of patients had also received other therapies such as interferon (Table 23).
Peer Reviewer 2	BCR-ABL	Page 103, para 2, line 2: statement “17 to 71 percent for accelerated or blastic phase” is incorrect. This is range in Table 26 for miscellaneous phases. For the accelerated/blastic phase, range in Table 25 is 27-60%, very similar to chronic phase.	The Page 103 pointed out by the Peer Reviewer should be Page 102. We corrected the percentage numbers accordingly. Thank you.
Peer Reviewer 2	BCR-ABL	Last two lines: what studies do not conform to this pattern, and do their results look different?	This is our supposition and there is no supporting evidence presented in the report; therefore, we deleted the relevant two sentences (P102 the last two lines).
Peer Reviewer 2	BCR-ABL	Page 107, para 2: worth noting that the Jabbour study had lowest proportion of patients tested (Table 24)	We inserted a sentence based on the suggestion: “The study found no patients with the T315I mutation but assessed only 30% of the entire patient cohort for the presence of mutations .”
Peer Reviewer 2	BCR-ABL	Page 108, line 1: Figure 15	We corrected these converting errors. Thank you.
Peer Reviewer 2	BCR-ABL	Page 114, para 1, line 6: Figure 16	We corrected this accordingly. Thank you.
Peer Reviewer 2	BCR-ABL	Page 21, penultimate line: change “0 cells” to “cells with zero entries”	We made the suggested change.
Peer Reviewer 2	BCR-ABL	Page 126, Table 33: It would be helpful if the footnotes more clearly related to the columns they are explaining	Thank you for pointing out. We corrected as per suggestion.
Peer Reviewer 2	BCR-ABL	Page 127: a lot of abbreviations, not all of which are explained	We added following abbreviations: AP (=accelerated phase), BC (=blastic phase), CCyR (=complete cytogenetic response), CHR (=complete hematologic response), CP (=chronic phase), CR (=complete reponse), CyR (=cytogenetic response), MCyR (=major cytogenetic response), OS (=overall response), PFS

			(=progression free survival), and RR (=relative risk).
Peer Reviewer 2	BCR-ABL	Page 129, para 3: In general agree with conclusion that individual patient meta-analysis may be needed, but comment on last couple of lines about studies originating from limited number of referral centres raises question as whether more effort needed to assemble body of data from which greater generalizability will be possible	Again, there is no evidence to support our claim of lack of generalizability in the report; therefore, we deleted the relevant two sentences (P129 the last two lines of the paragraph 3).
Peer Reviewer 2	Cross-cutting	Page 132, 3rd bullet from end, factors	We have corrected the typographical error.
Peer Reviewer 2	Cross-cutting	Page 133: 2nd bullet _ insert “the reader” after “we remind”; further down “various biases” _ what are these?	We added “the reader”. We clarified regarding biases.
Peer Reviewer 2	Cross-cutting	Missing pages: I noted 66, 76, 92, 94, 128, 134	These are even pages left black when there is a new section starting in an odd page.
Peer Reviewer 3	General	Overall, this is an impressive and well-conducted evidence report – particularly in regard to the challenging analysis of the BCR-ABL mutations in CML.	We thank the Peer Reviewer for the kind comment.
Peer Reviewer 3	General	Key Questions: For future studies, could other ‘intermediate’ key questions be asked – e.g., related to potential benefits or harms?	For our review, the Key Questions were determined at the planning phase of the review.
Peer Reviewer 3	Summary	Lack of clarity in BCR-ABL summary: Summary (p. S-3): The first several sentences of the BCR-ABL section are extremely confusing and appear contradictory. Suggest re-writing, and defining TKI.	We rewrote the relevance sentence as follows: “ The presence of any BCR-ABL1 mutation (all mutations considered together) does not appear to predict differential response to tyrosine kinase inhibitor (TKI) treatments (defined as imatinib-, dasatinib-, and nilotinib-based regimens).”
Peer Reviewer 3	CYP2D6	Need to include recent, large studies on tamoxifen: Association Between CYP2D6 Polymorphisms and Outcomes Among Women With Early Stage Breast Cancer Treated With Tamoxifen. JAMA, October 7, 2009; 302: 1429 - 1436. Kiyotani K, Mushiroda T, Imamura CK, et al: Significant effect of polymorphisms in CYP2D6 and ABCC2 on clinical outcomes of adjuvant tamoxifen	We have updated the search strategies for all outcomes (including tamoxifen for breast cancer). The two studies suggested by the Peer Reviewer were among the new studies we identified.

		therapy for breast cancer patients. J Clin Oncol 28:1287–1294, 2010	
Peer Reviewer 3	CYP2D6	Lack of pooling for tamoxifen studies: p. 12. It is surprising that no attempt was made at meta-analysis for tamoxifen. Despite 'irreconcilable' differences in definitions of metabolizer status, a random effects meta-analysis should be strongly considered. p. 28. If the adjusted values presented in several studies are misleading, why not pool unadjusted results? Furthermore, classifying a relative effect of 1.55 as 'modest' does nothing to confer the potential population impact. While the quality of the studies was clearly poor, the authors could conduct a random effects meta-analysis that would be more exploratory in nature. The results of such do not need to influence the overall conclusion of the report (and shouldn't given the limitations), but the lack of synthesis here is an omission. If the authors choose not to conduct a random effects meta-analysis, a more specific explanation for their rationale should be provided.	Random effects meta-analysis (under the commonly utilized models) assumes that there is a distribution of true effects and the analytical method attempts to identify the mean and the uncertainty around it. Based on the heterogeneity of outcomes reported, genotypes investigated, performing a meta-analysis of the unadjusted results would not overcome the limitations discussed above. In addition, very few studies reported anadjusted estimates or crude event rates to allow calculation of the unadjusted estimates. As such any pooled analysis would be subject to reporting bias. Since this report concerns pharmacogenetic tests, calculating the "population" effect would require use of the "average response rate" which is hard to calculate and cannot be representative of the patients from whom the relative effect was derived.
Peer Reviewer 3	KRAS	p. 78: Although the authors briefly discuss the weight of evidence for cetuximab vs. panitumumab, it seems that stratified pooled analyses are warranted, and mention in the summary of any differences in effect sizes or amount of evidence – particularly for RCTs.	We have conducted updated meta-analyses for the KRAS pharmacogenetic test and we have investigated the potential difference between panitumumab- and cetuximab-based studies when enough studies were available.
Peer Reviewer 3	General	p. 3: The statements "must be demonstrated" and "require proof of benefit" are somewhat extreme. 'should be' and 'evidence of' ?	The suggested changes have been implemented.
Peer Reviewer 3	Summary	Table S1: title should specify that numbers of studies is what is being listed.	Given the extreme overlap between studies (particularly for the KRAS topic) patient numbers would not be representative and may be misleading.
Peer Reviewer 3	CYP2D6	Fig. 3 is very interesting and helpful. Suggest adding bar color coding legend, rather than defining colors in figure caption.	This would not be feasible.

Peer Reviewer 3	KRAS	Fig. 5 is useful but legibility could be improved	We have updated the figure to include the new KRAS-relevant studies and have also increased the resolution to improve legibility.
Peer Reviewer 3	KRAS	Table 10. Delete 'could be color code'?	We have removed this artifact of the editing process.
Peer Reviewer 4	CYP2D6	The SACGT definition focuses on heritable variation, which would only seem relevant to CYP2D6 polymorphisms in the context of this report. The "lumping" of tests based on somatic mutations (which differentiate disease processes) from heritable mutations (which differentiate individuals) is unfortunate and confusing to the average practitioner and patient.	This is not accurate, the definition includes somatic mutations. (there is an "or" connecting several attributes of what falls under genetic testing, and somatic mutations are not explicitly excluded). We have provided further clarification regarding the definition of genetic tests. In the Methods of the TA, the introductions, and results of the specific parts and the Discussion of the methodological topics, we highlight the distinction between germ-line and somatic variation.
Peer Reviewer 4	CYP2D6	In regard to CYP2D6, the timing of the literature review was such that the most important paper on this topic (Schroth, JAMA, 2009) was not included. This well done retrospective study convincingly demonstrates a relationship between CYP2D6 genotype and rate of recurrence. In addition, a recent study by Kiyotani (J Clin Oncol, 2010) provides further evidence supporting the importance of CYP2D6 genotyping.	We have updated the search strategy and data extraction up to March 2010. The studies referenced by the Peer Reviewer were among the ones we identified in our update.
Peer Reviewer 4	CYP2D6	The lack of an association of genotype and mortality should be expected, given that tamoxifen's primary effect is to delay progression rather than cure micrometastatic disease. For the same reasons, there is no basis for looking at overall recurrence rate. Thus, a more appropriate meta-analysis would focus on those studies that focused on time to recurrence or recurrence-free survival.	We collected data both on mortality, overall survival, disease recurrence and time to recurrence. We did not perform meta-analysis for any of these outcomes because of the significant clinical heterogeneity.
Peer Reviewer 4	CYP2D6	The potential value of CYP2D6 genotyping is not well articulated. Randomized trials have demonstrated that letrozole is superior to tamoxifen for the overall population, but it is plausible that the two drugs have equivalent efficacy, except in CYP2D6 slow metabolizers. Thus, CYP2D6	Although the hypotheses brought forward by the Peer Reviewer are plausible the systematic review methodology is geared towards answering specific clinical questions based on available evidence. Although prediction of response to tamoxifen would be very important from a public health perspective, this does not

		genotyping could identify patients who would be spared the cost and toxicities of letrozole and other aromatase inhibitors.	validate CYP2D6 polymorphisms as a predictive (pharmacogenetic) test.
Peer Reviewer 4	CYP2D6	The potential risk of CYP2D6 genotyping is also not well articulated. If CYP2D6 slow metabolizers do not have an inferior result, then the administration of tamoxifen (in lieu of an aromatase inhibitor) to extensive metabolizers would be harmful.	Please see above.
Peer Reviewer 4	CYP2D6	Table 6 does not appropriately represent the issues regarding tamoxifen and CYP2D6 genotype. The underlying hypothesis is that all (or most) of the effects (both toxic and beneficial) of tamoxifen are a result of metabolites formed by CYP2D6. Thus, the assertion that the use of CYP2D6 inhibitors is not a relevant covariate is incorrect. If there is a relationship between CYP2D6 genotype and tamoxifen efficacy, that relationship would be masked if all patients were prescribed CYP2D6 inhibitors. Whether or not "confound" is the correct verb, the ideal analysis would exclude all patients who received CYP2D6 inhibitors, although that analysis would be confounded if there were more patients on CYP2D6 inhibitors who were extensive metabolizers (hypothetically due to a greater effect of tamoxifen in this population). For the same reason, adherence may vary among CYP2D6 genotypes, if the toxicity varies among CYP2D6 genotypes.	The use of the word "confound" is not a matter of terminology but one of substance. If a variable sis a potential confounder (and there are solid epidemiologic methods to identify potential confounders) then "conditioning" on the effect of the confounder (commonly achieved by performing multivariate regression including the confounder as a variable) would be necessary. On the other hand, if the variable of interest is not a confounder then simple inclusion in regression models will reduce precision and may also generate bias, thus adverse effects on study accuracy can occur when non-confounding factor is treated as such. Because the in the scenarios described by the Peer Reviewer use of CYP2D6-inhibiting medications cannot be a confounder adjusting for their use in regression (commonly performed in the studies we reviewed) is a serious flaw. Additionally, exclusion of individuals using CYP2D6 inhibiting medications for all analyses is problematic because it fails to account for all available information.
Peer Reviewer 4	CYP2D6	The most appropriate recommendation is the formation of a consortium to conduct an aggregated analysis of all data from the tamoxifen adjuvant trials. Ideally, all genotypes would be included, and there would be a quality control process to ensure that the genotypes are analytically valid. It would also be ideal to have uniform clinical data,	We agree that this would be a step forward for the CYP2D6 pharmacogenetics field. We have added the following comment in the discussion section: "Efforts to standardize the definitions of metabolizer groups based on genotype information would allow uniform reporting and facilitate patient-level synthesis of

		particularly time to recurrence. However, it will be difficult to have consistent information regarding concomitant CYP2D6 inhibitors and adherence for these retrospective studies.	results across studies."
Peer Reviewer 4	CYP2D6	The recommendation regarding "repurposing" of randomized clinical trials is a good one. In fact, one could consider recommending that federal funding agencies require collection of germline DNA on all federally funded clinical trials, unless the investigator provides an adequate scientific justification for not doing so (e.g., small sample size). (This is similar in concept to NIH requirements to study women, minorities, and children.) There may be an imbalance in any randomized trial in potential covariates. Thus, it is appropriate to consider adjusting for such covariates, but this should be prospectively specified in order to avoid the data dredging that is so prevalent in the pharmacogenomics literature. The emphasis on analyzing data from randomized trials is important, particularly when there is no a priori hypothesis based on other pharmacological or biological data. However, for CYP2D6, there is no basis for hypothesizing any relationship between CYP2D6 and prognosis (in the absence of treatment) and thus this concern is less relevant (albeit still valid). In contrast, one would anticipate that somatic mutations may have an effect on prognosis.	[no response needed]

[1] Peer Reviewers are not listed in alphabetical order.
[2] If listed, page number, line number, or section refers to the draft report.
[3] If listed, page number, line number, or section refers to the final report.

Project Name: Systematic Reviews on Selected Pharmacogenetic Tests for Cancer Treatment: *CYP2D6* for Tamoxifen in Breast Cancer, *KRAS* for anti-EGFR antibodies in Colorectal Cancer, and *BCR-ABL1* for Tyrosine Kinase Inhibitors in Chronic Myeloid Leukemia
Project ID: GENC0609

Table 2: Public Review Comments

Peer Reviewer Name[1]	Peer Reviewer Affiliation[2]	Section[3]	Peer Reviewer Comments	Author Response[4]
Harry Burke	George Washington University	General	[The Peer Reviewer sent a draft of a manuscript submitted for publication for consideration]	Thank you for your contribution. We agree that a distinction between prognosis and prediction is important. The methods section of the TA describes the epidemiological, methodological and clinical characteristics of eligible studies. The submitted manuscript is not eligible for inclusion in the TA.
John Hermanek	Amgen	KRAS	[The comment referred to more recent information on panitumumab and KRAS, published after the last update of the draft TA.]	The TA has been updated to include publications on KRAS through March 2010. Also, according to the eligibility criteria only full publications are eligible.
Gregory D. Pawelski	None	General	[The Peer Reviewer presented an overall interpretation of translational research in personalized medicine and cancer]	Thank you for you contribution. The TA used a systematic review approach to address the posed key questions. The Methods section describes the approach, and the epidemiological, methodological and clinical characteristics of eligible studies, as well as the statistical analyses performed.
August J. Salvado	Novartis Pharmaceuticals	BCR-ABL1	[The Peer Reviewer makes clarifying comments also with respect to the MEDCAC. Two points that could be relevant to the report are brought up:] 1. […] It warrants mention that CML is the only hematological malignancy covered by the technology assessment, and as such it should be recognized that physicians utilize different	Thank you for your comments. We note that the MEDCAC had a broader set of questions that the TA. The key questions of the TA are listed in the methods section. With respect to the 2 clarification points: First, the TA states clearly that tumor load monitoring is outside its scope. Only mutation

			techniques to measure treatment response and clinical efficacy than are typically used to monitor solid tumors. PCR monitoring of BCR-ABL is the least invasive and most sensitive test currently available to hematologists to monitor treatment response and clinical outcomes in CML. 2. The second point of clarification focuses on the scope of the technology assessment in reviewing the utility of BCR-ABL testing to predict treatment response based upon polymorphisms of point mutations. We concur with the technology assessment's conclusion that there is little evidence at the current time supporting the predictive value of point mutations in the BCR-ABL1 gene to determine differences in clinical benefit of one TKI agent over another. However, one exception to this point that should be explicitly indicated in the technology assessment is that the T315I mutation predicts a decreased clinical response to all currently commercially available TKI therapies. In addition, when this mutation is present, physicians should prepare for an aggressive treatment regimen, potentially including stem cell therapy or experimental treatments, in the event that patients with T315I mutations fail currently available TKI therapies.	testing is in the scope of the TA and all our analyses pertain to mutations only. Second, the TA clearly singles out T315I and performs separate analyses for this mutation. However, the TA does not make recommendations for treatment or the management of patients with specific mutations.
Lawrence Solberg	American Society of Hematology (ASH)	General comments on BCR-ABL1 mutation testing	(This is a distillation, see pdf) The comments are the ASH answers to the key questions of the TA. 1. Does BCR-ABL1 mutation testing predict response to TKI therapy? Yes. There is enough evidence, both in vitro and in vivo that different mutations have different sensitivity to different inhibitors and that the in vitro sensitivity correlates well with the clinical response of patients […].In view of the recognized value of mutations analysis in predicting for response, this assay is	Thank you for your comments, which are a narrative (non-systematic) review of the positions of the ASH. The TA used a systematic review approach to address the posed key questions. The Methods section describes the approach, and the epidemiological, methodological and clinical characteristics of eligible studies, as well as the statistical analyses performed.

			recommended by a panel of international experts sponsored by the European LeukemiaNet in all patients after failure or suboptimal response to imatinib, and before changing therapy. 2. What patient- and disease-relevant factors affect the test results, their interpretation or their predictive response to therapy? The most important factor is the response to prior therapy […] 3. How does the gene testing impact the therapeutic choice? As mentioned earlier, patients with mutations of intermediate or low sensitivity to a certain tyrosine kinase inhibitor should be offered the alternative agent […]. 4. What are the benefits and harms or adverse effects for patients when managed with gene testing? 5. The obvious benefit is the possible selection of an agent that may offer the best probability of response. There are no known adverse events other than those implicated with the venipuncture. However, the test is usually obtained at the time other routine monitoring tests are obtained.	Studies that evaluated in-vitro sensitivity of BCR-Abl1 mutations to TKIs do not meet our inclusion criteria; thus, we did not assess these studies. Also, we did not assess the three studies reporting the association between in-vitro sensitivity of the mutations to TKIs and clinical outcomes (Hughes 2009, Muller 2009, and Jabbour 2009), which could be viewed as a way to lump some specific mutations together to perform subgroup analysis, because they were published after our literature search.
Mark Somerfield	American Society of Clinical Oncology	KRAS	First, on p. S-1, the authors list Key Question (KQ) 4 as "What are the benefits and harms or adverse effects for patients when managed with gene testing?" They note that no studies were identified that could be used to answer this question. Elsewhere (p. 3), however, the authors provide a broader explication of KQ4 that creates some confusion: "What are the benefits and harms or adverse effects for patients when managed with gene testing? Any cognitive, behavioral or other health effects of testing with the three tests of interest. These may be direct effects of the process of testing	We appreciate the comment. We agree that there is the potential for confusion, and we have rephrased the methods section to enhance clarity. Indeed KQ1 captures the downstream effects of testing, and KQ4 captures other (additional) benefits and harms that are related to testing and are beyond those covered in KQ1.

			(e.g., increased anxiety) or downstream effects stemming from treatment decisions informed by testing." Given this broader framework for KQ4, the statement that no studies were identified that could be used to answer this question is problematic. Thus, in the KRAS section the authors concluded that, "When treated with anti-EGFR antibodies, patients with KRAS mutations were less likely to experience treatment benefit, compared to patients whose tumors were wild-type for KRAS mutations, for all outcomes assessed. These results were confirmed in several RCT-based analyses of progression-free survival that demonstrated a significant treatment-by-KRAS mutation interaction in three out of the four cases where such analyses were reported." These results, which arguably reflect downstream effects stemming from treatment decisions informed by KRAS testing, would seem to address KQ4, although the authors do note that "Most of the latter [the downstream effects] would be captured by Key Question 1." Can the authors clarify this? Would it be simpler to restrict KQ4 to the benefits and harms of the testing itself, defined as "any cognitive, behavioral or other health effects of testing,? vs. clouding things by referring to the downstream effects element"	
Mark Somerfield	American Society of Clinical Oncology	CYP2D6	p. S-2, the authors note that "It is questionable whether pharmacogenetic testing of germline (heritable) variations in CYP2D6 can predict differential response to adjuvant tamoxifen in women with non-metastatic breast cancer. Further, evidence is severely limited for tamoxifen-treated women with metastatic disease. Our conclusions are in accordance with the 2009 American Society of	Thank you for providing this clarification. We rephrased the pertinent parts for clarity, both in the Executive summary and in the body of the TA adopting some of the Peer Reviewer's wording.

			Clinical Oncology (ASCO) practice guideline update." The ASCO guideline that is referenced here is an update of a guideline on breast cancer risk reduction. The ASCO guideline update did consider CYP2D6 and concluded that, "Given the limited evidence, CYP2D6 testing is currently not recommended in the preventive setting." However, it is not accurate in our view to state that the conclusion of the AHRQ systematic review are "in accordance with" the ASCO guideline; the ASCO guideline concerns the very different setting of risk reduction or chemoprevention, and the AHRQ report focuses on the adjuvant and metastatic settings. Later, in the body of the report (p. 27), the authors do note indirectly that the ASCO guideline concerned the risk reduction or prevention setting: ?Our conclusions are in accordance with the relevant 2009 ASCO practice guideline update, which states ?Given the limited evidence, CYP2D6 testing is currently not recommended in the preventive setting.?? However, as many may read only the executive summary of the technology assessment, this point should be made in that summary as well. Ideally, the authors would change the phrasing from ?in accordance with? to ?[our conclusions] are analogous to the conclusion of the ASCO guideline that addressed the role of CYP2D6 in the breast cancer risk reduction setting.?	
Anonymous Reviewer 1	NA	General	Utilization of pharmacogenetic information to prescribe medications on the basis of the needs of the individual patients can save money and lives. Knowing that Tamoxifen will not work for you, if you were a cancer patient, is something you would like to know before taking the medication for years just to learn that your cancer is back. Money and time are wasted when a patient takes a drug that has no effect	Thank you for your comments. No reply needed.

			because of variants in a patient's genetic make-up! Pharmacogenetics is a tool that physicians can use in many areas of medicine to better and more safely help patients. More knowledge is always better for any disease state.	
Anonymous Reviewer 2	Friends of Cancer Research	General	[The letter contains several general thoughts and three more specific mentions, listed below:] 1. We believe that it is premature to accurately determine impact of CYP2D6 polymorphisms on response to tamoxifen based solely this literature review Currently, multiple trials are underway or have been completed that will prospectively examine the role of CYP2D6 genotype on the clinical effect of tamoxifen treatment in both the metastatic and adjuvant setting.1 In addition, a study conducted by Schroth et. al., published after the timeframe of the technology assessment, retrospectively examined 1,325 patients treated with adjuvant tamoxifen for early stage breast cancer. The study concluded that, "Among women with breast cancer treated with tamoxifen, there was an association between CYP2D6 variation and clinical outcomes, such that the presence of two functional CYP2D6 alleles was associated with better clinical outcomes and the presence of nonfunctional or reduced-function alleles with worse outcomes." While this study is not a randomized controlled trial, of which it may be several years until clinical results will demonstrate prospectively what the actual impact of altered drug metabolism is, the conclusions of this study are indicative of the complexity of the topic and weakness of the currently available evidence. Therefore, it would be difficult to conclude that pharmacogenetic testing for variations in CYP2D6 should not be used by physicians and patients with breast cancer to determine course of	Thank you for your comments. The TA used a systematic review approach to address the posed key questions. The Methods section describes the approach, and the epidemiological, methodological and clinical characteristics of eligible studies, as well as the statistical analyses performed. 1. CYP2D6: The TA summarized the published evidence and does not make recommendations for practice. The final version of the TA, based on an updated search and data extraction, includes the Schroth paper (among others), and reaches the same conclusions. 2. The Methods section describes exactly what is meant by each key question, which is quite specific. As described in the TA there were no eligible studies that quantified the impact of testing on therapeutic decisions in the form of how many times the physicians changed treatment, etc. This is qualitatively different from the comments made by the Peer Reviewer. 3. The TA clearly states that only BCR-ABL1 mutations are in the scope of the pertinent systematic review. As per our reply to other Peer Reviewers, transcript levels are not examined in the TA. This was predetermined in out review protocol and clearly stated in the methods section

			treatment. 2. Perhaps one of the most notable and recent changes in the practice oncology is the utilization of KRAS testing to aid therapeutic decision making for colorectal cancer patients. The conclusions described in the technology assessment are in accordance with the body of clinical evidence that has prospectively demonstrated, in several cases, the predictive value of testing for KRAS mutational status prior to delivering the anti-EGFR therapies, cetuximab and panitumumab, to colorectal cancer patients. However, the technology assessment suggests that the data are not being used in therapeutic decision making (S-5: "In all three examples, we found no evidence on whether testing impacts therapeutic decisions…"). […] 3. In the case of evaluating how variations in BCR-ABL1 impact response to three drugs used for treatment of chronic myeloid leukemia (CML), the technology assessment concludes that "The presence of any BCR-ABL1 mutation does not appear to predict differential response to treatment…with any of the three tyrosine kinase inhibitor (TKI)-based regimens." While this is an accurate summary of existing published evidence, there is a rapidly developing body of new clinical evidence that will provide additional information on this complex topic, including over a hundred studies examining the use of imatinib, dasatinib or nilotinib in chronic myeloid leukemia.4 However, to date BCR-ABL pharmacogenetic testing has not been shown to be a predictive factor to guide first-line treatment decisions. But the value of molecular evaluation in this case should not be dismissed due to the existence of molecular methods, such as RQ-PCR transcript testing, that	of the TA.

			can be used to monitor treatment response	
Anonymous Reviewer 3	Association for Molecular Pathology	ABCR-ABL1 – General	[comment put into numbered paragraphs:] 1. We very much appreciate the authors' statistical expertise of this technology assessment report; however, we note a number of shortcomings that compromise the relevance of the report's conclusions. We believe that many of these shortcomings could have been avoided had there been prior input from clinicians and molecular pathologists intimately familiar with the performance and clinical utilization of these tests This lack of clinical input is immediately evident in the definition of genetic test adopted by the study: "The analysis of human DNA, RNA, chromosomes, proteins, and certain metabolites in order to detect heritable disease-related genotypes, mutations, phenotypes, or karyotypes for clinical purposes. Such purposes include predicting risk of disease, identifying carriers, establishing prenatal and clinical diagnosis or prognosis. Prenatal, newborn, and carrier screening, as well as testing in high-risk families, are included. Tests for metabolites are covered only when they are undertaken with high probability that an excess or deficiency of the metabolite indicates the presence of heritable mutations in single genes." Two of the three tests evaluated in the study do not fulfill this definition, highlighting a superficial understanding of the biology underlying these tests and how they are used clinically. Acknowledging that pharmacogenomic tests can be a special type of genetic tests, it is noteworthy that the authors fail to appreciate that of the three	Thank you for your comment. Please, see point to point replies: 1. We now clarify in the methods the nebulous text (which was a verbatim quotation). Please also see our replies to the peer-Peer Reviewers. The definition includes somatic mutations. (there is an "or" connecting several attributes of what falls under genetic testing, and somatic mutations are not explicitly excluded). In any case we make clear in the report. As described in the Methods of the TA, the introductions, and results of the specific parts and the Discussion of the methodological topics, we are cognizant of the distinction between germline and somatic variations. Both the draft and the final report provide a detailed discussion of methodological issues unique to heritable vs somatic genetic variation. We selected dasatinib as a second-generation TKI after discussion with AHRQ and CMS. 2. The Methods section presents detailed criteria for study eligibility. Also, note that contrary to narrative reviews, a systematic review uses structured and rigorous methodologies to provide answers to *specific* key questions . 3. As described in the methods section, analytic validity was not in the scope of the

			tests evaluated in the report, only the CYP2D6 qualifies as being heritable. Alterations in the KRAS gene and in the BCR-ABL translocation are not heritable, but are tumor specific, intrinsic to the neoplastic process. This distinction is not moot. Polymorphisms that influence drug metabolism can be identified in healthy individuals and can have bearing on dosing or drug selection of numerous therapeutic agents. Tumor specific genetic changes, in contrast, have significance beyond simple choice of drug, influencing disease recognition, disease prognosis, tumor aggressiveness, and potential response to multiple and combinational chemotherapeutic agents. Therefore, the value of a genetic test in specific malignancy is more than for the selection of one specific chemotherapeutic agent. These genetic changes need to be considered in the clinical context of the specific tumor for each patient. The clinical decision to treat or not treat with a specific agent takes into account all of these factors and is not made on the basis of a single test result. 2. The naivete' of the concept of "one bioanalyte - one drug" becomes apparent in consideration of the drug dasatinib, one of the drugs used in the setting of Bcr-Abl+ leukemias resistant or intolerant to prior therapy. There is evidence that this drug also has activity against Src family kinases as well as Flt3 and c-Kit. (Corey, et al, Clin Cancer Res 16:1149-58, 2010). 3. Additionally, pre-analytic issues are critical to the performance of each assay, and must be given consideration. For example, the choice of method used for the detection of a mutation will	TA. Please, also refer to the replies above. 4. Replies to quotations by Dr Jones in points 5-8 below: 5. Thank you for your comments. The Methods section presents an explanation of the analyses performed in the TA. Briefly, the TA did not perform a meta-analysis (did not "lump") studies exactly because of clinical heterogeneity (dissimilarity). We agree that the incidence of mutations is heterogeneous, which is already described in the Results section of the TA. However, we disagree with the comment that the incidence of the T315I mutation depends on the phase of disease. This is not necessarily the case in our results (see Figure 13). We agree that response to therapy depends upon the phase of disease and did not "lump" different disease phases together. Further, the general methods section has a brief explanation of graphical ("qualitative") analyses of published studies for readers who are not familiar with evidence-based methods or clinical research methodology. 6. It is unclear to us what the comment means. The key questions were refined by AHRQ, CMS and Tufts EPC, influenced by the methodological framework of the EGAPP initiative. The methods section includes a description of eligible studies. In particular, no study performed interaction analyses. The TA includes a description of the rational for focusing on interaction analyses to identify modifiers of the pharmacogenetic effect. 7. The methods section explains what is

			have a major impact the sensitivity of the assay, with limits of detection ranging from 1 cell (or less) in one million for a PCR approach targeting the mutant allele, to the requirement that greater than 15-20% of cells contain the mutation for most sequencing methods. The selection of pure tumor cells prior to sample processing can further exaggerate apparent variations in analytic sensitivity, so that a study utilizing relatively insensitive conventional sequencing, along with selection for tumor cells, will likely vastly underestimate the true occurrence of the mutation in a case series being studied. These "false negative" results will lead to an inaccurate assessment of the clinical correlation or clinical utility. 4. In January 2009, AMP published laboratory practice guidelines for detecting and reporting BCR-ABL drug resistance mutations in CML and ALL. Those guidelines effectively discussed the state of knowledge regarding BCR-ABL mutation testing not only in considering analytical factors, but also in the clinical contexts for which such testing has import (Jones, et al, J Molec Diag 11:4 ? 11, 2009). We asked Dr. Dan Jones, one of the authors of that report, to comment (in qutoes below) on the technology assessment's conclusions regarding BCR-ABL mutation testing: 5. KQ1: "The commentary in Key Question 1 is fair. However, the literature on CML and mutations is pretty vast right now and some studies have been omitted. Therefore, some qualifications on the conclusions reached in that Key Question is recommended. The Authors need to emphasize that there are big differences in the incidence of	sought after in key question 4. Briefly, eligible studies are those that report the number of times that the treatment decision changed before compared to after testing. Regarding the first sentence of Section 3.4 ("In our systematic review of the literature, presence of any BCR-ABL1 mutation does not appear to predict differential response to treatment in CML patients treated with imatinib-, dasatinib-, or nilotinib-based regimens."): This is not pertinent to KQ3. Further, we feel that the statement is accurate statement, as it refers to having "any mutation". This is based on the results described in KQ1. Regarding the Jones paper: This was published after the last search. Further, it is not in the scope of KQ3, since this paper does not describe the data on treatment decision impact of the test (i.e., how frequently treatments were changed after performing the test).In addition, T315I mutation testing to predict treatment outcomes of homoharringtonine alone or TKI-MK-457 is beyond the scope of this technology assessment review. 8. The Methods section of the TA describes the epidemiological, methodological and clinical characteristics of eligible studies, as well as what is sought after in KQ4. None of the eligible studies provided clinical data for KQ4. No studies provided specific data on clinical benefit or harm

			mutations (particularly T315I) and the therapy responses depending on the phase of disease (chronic, accelerated and blast phase) and lumping all together as "CML" is probably not useful for interpretation of test results." 6. KQ2: "There is some gathering data on levels of drug metabolizing genes on responses to TKIs but agree that this question is not really relevant to interpretation of BCR-ABL testing." 7. KQ3: "I would encourage the Authors to include the reference Jabbour E, Jones D, Kantarjian HM, O'Brien S, Tam C, Koller C, Burger JA, Borthakur G, Wierda WG, Cortes J. Long-term outcome of patients with chronic myeloid leukemia treated with second-generation tyrosine kinase inhibitors after imatinib failure is predicted by the in vitro sensitivity of BCR-ABL kinase domain mutations. Blood. 2009 Sep 3;114 (10):2037-43 which does show (retrospectively) that if mutations are matched to the Kd for in vitro inhibition of second (or third) TKIs that there are differences in outcome in chronic phase CML. This would contradict the general statement in the first line of 3.4 Discussion." Given the already extensive data on in vitro responses to particular TKIs, a prospective study is unlikely to be done in CML to randomize treatment choice based on mutation result. However, the European LeukemiaNet guidelines (Baccarani M, Cortes J, Pane F, Niederwieser D, Saglio G, Apperley J et al. Chronic Myeloid Leukemia: An Update of Concepts and Management Recommendations of European LeukemiaNet. J Clin Oncol 2009 November) are an attempt to codify current clinical practice on	defined in details in the general method section. 9. Thank you for the suggestions.As detailed in the Methods, the Results and Discussion sections, we did not perform a meta-analysis in our systematic review. A meta-analysis was performed only in the KRAS section of the TA. No further reply necessary.

			how detection of T315I impacts choice of therapy. The homoharringtonine clinical trial (published in abstract form, Khoury HJ, Michallet M, Facon T, Guilhot F, Jones D, Hochaus A, Benichou A-C, Schwartz R, Cortes J. Safety and efficacy study of subcutaneous homoharringtonine (SC HHT) in imatinib-resistant chronic myeloid leukemia (CML) with the T315I BCR-ABL kinase domain mutation ? initial report of a Phase II trial. Blood 110(11):318a, 2007.) and to some extent the TKI-MK-457 clinical trial (Blood, 15 January 2007, Vol. 109, No. 2, pp. 500-502) use presence of the T315I mutation as enrollment criteria, based on the selective responses of those particular agents against that mutation." 8. Key question 4 "Different 2nd and 3rd-generation TKIs (and non-KI therapies) have different toxicity profiles so the use of BCR-ABL mutation data to influence choice of a particular TKI will have benefits and harms to patients." 9. In summary, we believe that if the authors had access to appropriately qualified clinical and technical input, the value of their study would have been markedly enhanced. Certainly, surveying the literature at a single point in time for a rapidly growing field suffers the danger of being irrelevant by the time the results are analyzed. This deficit would be very apparent to anyone with true clinical experience. As it is, the conclusions can only be regarded as having limited marginal value. We offer the Guidelines published by AMP in 2009 as an example of a rational, coherent approach to assessment of test efficacy and utility that recognizes that such an assessment must be a dynamic, clinically	

			relevant process. We strongly urge that future meta analyses of published reports include appropriate scientific and clinical expertise to better design the inquiries and better assess the outcomes for reasonableness. We further urge that any such technical assessments be presented in the appropriate clinical contexts. We believe that introducing these elements will significantly enhance the validity and utility of future studies.	
Anonymous Reviewer 4	College of American Pathologists	BCR-ABL1 General	[the comments are presented as a numbered list to facilitate point to point replies] 1. The College commends the authors at the Tufts Evidence-Based Practice Center for a well written and insightful report. It has a consistent format throughout and explains the statistical analyses clearly. However, the College is concerned about the performance of meta-analyses and other literature reviews divorced from an understanding of the clinical use of the tests which can result in the wrong questions being asked or incorrect framing of the questions. The College believes the questions should be more nuanced than stated in the report. Question 1 asks "Does the genetic test result predict response to therapy?" Testing and treatment are complicated, and reducing the issue to a simple one bioanalyte-one drug (i.e. companion diagnostic) issue is naive. The article by Jones et al. provides a very cogent review of the status of BCR-ABL mutation monitoring which does address the complex clinical contexts in which that question of imatinib resistance arises and provides a useful model for this type of review. Tests may answer questions important to the treatment of the patient but not directly	Thank you for your comments. Please find point to point replies below: 1. The General Methods Section presents the rational behing the Key Questions asked and the epidemiological, methodological and clinical characteristics of eligible studies. We make the following clarifications: first, the eligibility criteria of studies per key question are listed in the Mehods Section (general and topic-specific). Further, the Methods section clearly defines what is sought after in each question of the systematic review (population, intervention, comparator, outcome, design). Please refer to the methods section for a clear description of eligible research. The key questions of the report were refined by the Tufts EPC in discussions with AHRQ and CMS. Further, the TA **did not perform meta-analyses** in BCR-ABL1. This is described in the methods, the results and the discussion. Also, the TA used predefined and explicit definitions of what research designs/clinical settings would be

			related to response to therapy including the following: a. Is mutation testing performed simply to select a drug, or is it also informative about likely disease aggressiveness and potential response to any therapy? b. Are there differences in degree of resistance to the various drugs? c. After therapy has failed (for whatever reason), is testing used differently than upon initiation of therapy? 2. Evidence-based medicine requires published evidence, and for diagnostic tests there may be sufficient data to implement an assay using intermediate outcomes rather than waiting for studies of long term outcome or survival. Diagnostics are critical to patient care but are also just one piece of medical decision-making processes, with many other clinicopathologic and socioeconomic variables contributing to patient outcomes. We are troubled that the investigators were unable to find any studies that answered Key Questions 2, 3, and 4. For example, the whole point of BCR-ABL mutation testing is to understand the basis for treatment failure in a very complex clinical context; yet they report no patient or disease-related factors were found that affect the test results. Also, it does not make sense that the authors found no information on Key Question 3: How does the gene testing impact the therapeutic choice? Clearly K-RAS testing impacts therapeutic choice. 3. On Key Question 4: What are the benefits and harms or adverse effects for patients when	analyzed. This information is presented in the Methods section of the BCR-ABL topic. With respect to 1a-1c: The posed questions span too broad a range of decisional contexts. The TA has a clearly defined decisional context, as detailed in the Methods Section of the BCR-ABL1 section. 2. The General Methods section and the specific methods sections of the TA presents the inclusion and exclusion criteria employed as well as what we considered as relevant evidence regarding each Key Question. We believe that the TA is accurate in stating that none of the identified studies fulfilled the eligibility criteria for KQ2-4. 3. The Methods section clarifies what is meant under KQ4. Note that it includes direct data on benefits and harms not summarized under KQ1. 4. The methodological discussion of cross-cutting issues provides a detailed explanation behind the rational of requiring interaction tests between treated and untreaded groups. We kindly note that for the CYP2D6 and the KRAS topic randomized controlled trials were available and included in the review. The interpretation of our statement is out of context. We will rephrase to enhance clarity. 5. Indeed the TA refers to the presence of any mutation. We revisited the phrasing to enhance clarity: “The presence of any BCR-ABL1 mutation (all mutations

			managed with gene testing? It is not clear how one defines "benefits or harms" in the context of genetic testing ? i.e. long term versus short term or surrogate outcomes, or a prospective study in which patients are managed with or without the genetic test. According to this report, none of the included studies reported benefits of testing. Similarly, no study reported harms or adverse effects of testing. We believe a greater understanding of the clinical contexts in which this testing is used might allow an appropriate search for the benefits and harms and adverse effects of inappropriate therapy for CML for example. The effectiveness of various treatment regimens has been well documented in multiple clinical studies and therefore unnecessary to include in studies of this testing. In studies addressing treatment failure, of which BCR-ABL1 mutations are only one cause, the evaluation of outcomes was not a primary goal, so it is not surprising that these issues were not specifically addressed. This is a deficiency of the very narrow approach taken. 4. In the conclusion the authors note that most studies analyze only treated patients, "effectively assuming that effects in untreated patients are zero." The assumption of the authors is not necessarily correct; the reason the studies do not have untreated patient arms is that we are treating human beings, not mice. To not offer some form of treatment in cancer therapy has potential ethical implications. Although an untreated arm may be necessary for good data collection, clinician cannot do this when treating patients.	considered together) does not appear to predict differential response to tyrosine kinase inhibitor (TKI) treatments (defined as imatinib-, dasatinib-, and nilotinib-based regimens)." 6. Thanks for pointing out the discrepancy with the direct quotation. We have rephrased for clarity. 7. This section has been updated in the final report to include 3 additional papers including the suggested one. The conclusions do not change. 8. As described in the methods section, analytic validity was not in the scope of the TA. We also note that in the current version of the TA the studies are not penalized for issues arising from less than perfect analytic validity (we do not account for measurement error). 9. Thank you for the pointer on gene names. In the Methods section we state that the TA refers to gene symbols in italics and protein symbols in regular font. We have verified that the convention is followed throughout. 10. We have defined what we mean as (relatively) rare. We do not change the verbiage. In our review, the frequency of T315I is not necessarily the highest of all identified mutations across the included studies (see out Figure 13). 11. Please refer to the discussion section of BCR-ABL1. Thank you for the suggestions. We have revised the initial part of the summary to clarify the points made. 12. This is not what the TA results were nor

			Specific comments on the Summary and Introduction: 5. In Section 3 of the Summary (page S-3), we suggest the removal of the second sentence ?The presence of any BCR-ABL1 mutation does not appear to predict differential response to treatment? and begin the summary with the third sentence ?There is consistent evidence that presence of the [most common] T315I mutation can predict TKI treatment failure?..? The second sentence as currently written implies that no BCR-ABL mutations are associated with treatment response, whereas it should read that if you query for the presence of ANY mutation, treatment response cannot be predicted, likely due to the fact that different mutations confer differential resistance to TKIs. The fourth and fifth sentences summarize the effect of the presence of ANY BCR-ABL mutation more clearly. 6. In the introduction the authors choose to use a definition of genetic tests from National Human Genome Research Institute (NHGRI) which includes only heritable mutations; then proceed to review studies of two tests that do not meet the definition provided on page one of the report. We do not have an issue with the definition provided, but suggest that the report be internally consistent. **Specific comments on Section 1:** 7. We suggest that the omission of a recently published and relevant study by Schroth et al. in JAMA , due to the timing of the review process, highlights the difficulty in taking a snapshot in	what the TA concludes. The Methods Section describes what was studied, how, and in which clinical context. We have already proposed that an appropriate next step would be a collaborative international patient registry or patient-level meta-analyses using standardized definitions of clinical context, disease stage, follow-up, and outcome assessment to identify particular mutations beyond T315I to predict treatment outcomes.

			time approach to review of a rapidly changing field. The paper by Schroth et al. was published in October 2009. **Specific comments on Section 2:** 8. Other factors not reviewed also impact the quality of the studies reviewed. The College notes that the analytic validity of assays should be mentioned as a factor impacting the quality of KRAS studies. Pre-analytic specimen preparation is also critical for somatic mutation tests such as KRAS (e.g. adequate tumor cell percentage). Though the test performance characteristics such as precision (reproducibility), analytic sensitivity, etc are important for evaluating the utility of these test but we understand that this topic is beyond the scope of this review. Quality is often assumed because CAP accreditation and proficiency programs provide successful oversight for consistent laboratory performance. 9. On page 29, this document should describe HUGO Gene Nomenclature Committee-approved official gene symbols along with alias in parenthesis. For example: Replace c-erbB-1 with ERBB1, and c-erbB-2 with ERBB2, etc. (Correct symbols and colloquial names for every gene and its encoded protein can be found at www.genenames.org.) Gene symbols are in italics (KRAS) while protein symbols are not (KRAS). **Specific comments on Section 3:** 10. We suggest a change to the verbiage throughout the document referring to the BCR-ABL1 T315I mutation as ?relatively rare,? as T315I is one of	

			the most commonly identified BCR-ABL mutations. (see Jones, et al. J Mol Diagn 2009;11:4-11). 11. On the top of page 130 Section 3-4, we suggest the following addition, ?A limitation of this systematic review is that all BCR-ABL1 mutations were lumped together, potentially restricting our ability to identify all but the strongest associations between a given mutation and drug resistance. Further complicating the process was the variability in study design, including varying indications for testing and varying spectra of analytes tested. Specific mutations beyond T315I that predict TKI resistance cannot be ruled out. Generally, it was found that patients who developed mutations during treatment experienced higher imatinib resistance compared with those with no mutations detected during the follow-up.? 12. It is striking to know that all the mutation tests performed so far have not shown to be predictive to patient outcomes except T315I. The authors may suggest research designs for future studies to answer key questions and items important for future data mining such as the detection sensitivity of an assay for future analysis.	

[1] Names are alphabetized by last name. Those who did not disclose name are labeled "Anonymous Peer Reviewer 1," "Anonymous Peer Reviewer 2," etc.

[2] Affiliation is labeled "NA" for those who did not disclose affiliation.

[3] If listed, page number, line number, or section refers to the draft report.

[4] If listed, page number, line number, or section refers to the final report.