THOMSON REUTERS

Outcome Evaluation of the National Cancer Institute's (NCI) Innovative
Molecular Analysis Technologies Program (IMAT)

August 2, 2013

Prepared by:

Thomson Reuters, Custom Analytics
Rockville, Maryland 20850

Report of IMAT Analyses

THOMSON REUTERS™

Executive Summary

The National Cancer Institute (NCI) Innovative Molecular Analysis Technologies Program (IMAT) was initiated in 1998 motivated by the premise that a focus on developing innovative cancer-relevant technologies will accelerate progress towards cancer treatment along the entirety of the cancer continuum from researchers and clinicians ultimately to patients. It has been recognized that to maintain program effectiveness, it is important to engage in on-going evaluation of the IMAT portfolio. To this end, this evaluation examines both IMAT applicants and awardees (Table 1).

Table 1: IMAT cohorts by evaluation task		Evaluation Task			
IMAT Portfolio	**Portfolio Description**	**1**	**2**	**4**	**6**
1	324 responsive applications (284 applicants) from RFAs issued with 3 receipt dates during 2012 (CA12-002,CA12-003,CA12-004,CA12-005)	✔	✔		
2	30 awards from RFAs issued with 3 receipt dates during 2009 (CA09-004,CA09-005,CA09-006,CA09-007,CA09-008)			✔	✔

This evaluation is built upon the idea that innovation may be understood and measured by its two separate components of novelty/inventiveness and utility/success[1] (see Figure 0.1).

Figure 0.1: The Evaluation Framework: measurable components of transformative innovation.

---
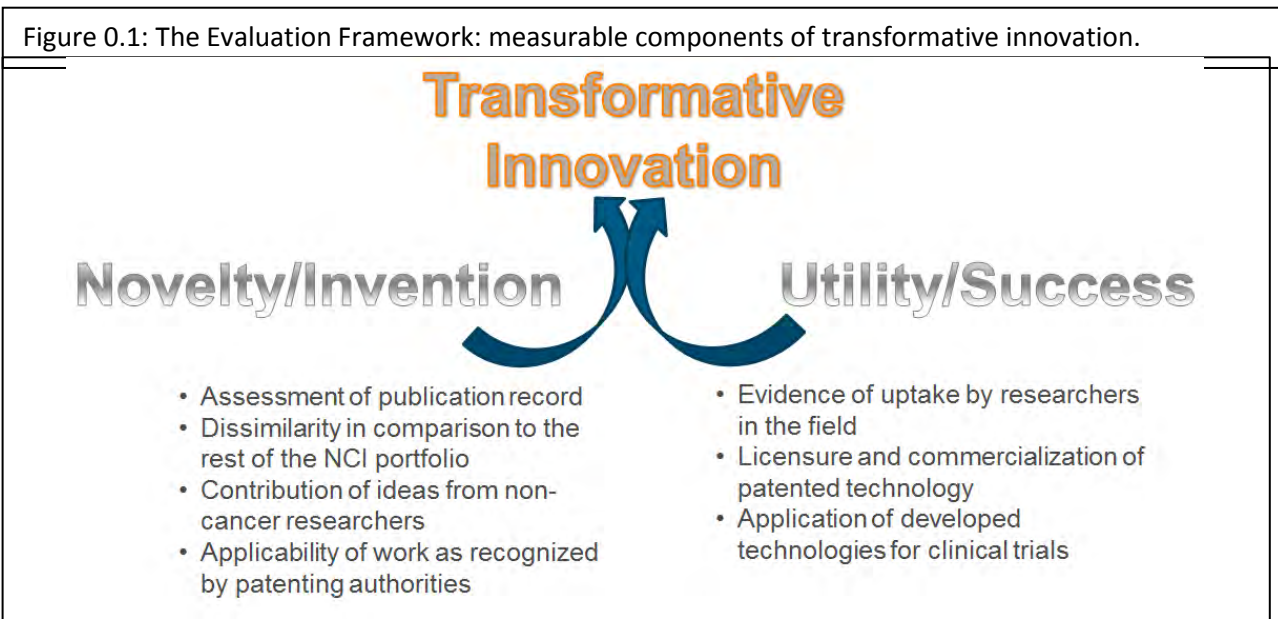
[1] Erwin and Krakauer. Insights into Innovation. 2004. Science Magazine. 304, 5674, 1117-1119.

The novelty/inventiveness of the proposed research from IMAT applications was assessed in three ways. The first approach (Task 1) used an automated method that compared text of the IMAT 2012 applications with text from a set of other similar NCI-referred program applications. The task 1 executive summary is listed below:

Task 1 Evaluation Question: Are submissions to and awards from the IMAT program significantly unique within the NCI portfolio?
Methodology Summary: We estimated a level of dissimilarity using a numerical measure of how different an IMAT application was from its most-similar counterpart in a large set of contemporaneous applications covering similar research topics (aka, "the companion cohort") and we compared the proportion of high-dissimilarity applications for IMAT to the high-dissimilarity proportion within the companion cohort itself.
Evaluation Result: The mean and median minimum calibrated dissimilarity (MCD) for IMAT relative to a companion cohort is significantly higher than that of the companion cohort relative to itself, excluding direct comparisons of an application to itself at the project level (1.13 times higher, 95% CI 1.09-1.17).

In the second approach (Task 2), the extent to which the IMAT program has been successful in attracting applicants with limited or no previous cancer relevant research -- with the assumption that these individuals may propose novel or inventive approaches to solving cancer-relevant problems -- was evaluated by determining the extent to which IMAT applicants engaged in cancer-relevant research prior to the current IMAT application.

Task 2 Evaluation Question: Is there evidence that the IMAT program attracted investigators to develop innovative technologies for cancer research, that may not have focused on cancer relevance otherwise?
Methodology Summary: The history of cancer-relevant research among the 2012 IMAT applications was examined from two perspectives:
- Whether the applicant was previously awarded a grant by NCI
- Whether and the degree to which their prior publications were relevant to cancer. This applied an analysis of the publication MeSH terms.
Evaluation Result: Less than 42% of the applications included a relatively higher indication of prior involvement in cancer-relevant research.

The third approach analyzed patent applications and awards (Task 4). The utility/success of technology products and knowledge generated from the IMAT program was assessed by examining commercial patent application and award activity (Task 4) and uptake of academic outputs by the scientific community at large (Task 6). The analysis of patent applications, awards and licensure was conducted to measure the inventiveness and utility of the research resulting from IMAT FY2010 awards to begin to examine if the IMAT program is accelerating patent and commercialization processes. A bibliometric analysis examined the utility of IMAT developed technologies for the broader research community and specifically compared the impact of the 2-year versus 3-year R21 awards from the IMAT 2010 award portfolio. See Table A0 in the Evaluation Summary Appendix for definitions of the data sources used to assess progress against the IMAT program goals.

Task 4 and 6 Evaluation Questions: Is there evidence that successfully developed technologies funded by IMAT were highly innovative? Have IMAT-supported technologies had a significant impact on the communities along the continuum of cancer research or clinical care?

Methodology Summary: We examined patent application, awards, licensure and commercialization resulting from IMAT funding as well as a variety of bibliometrics based on the articles produced from the FY2010 IMAT awards.

- Citation Based Bibliometrics
- Cancer relevance of citing publications
- Citing by Clinical Trials
- R21 Analysis: Compare 2-year and 3-year funded R21

Evaluation Result:

Investigators from the FY2010 IMAT awards demonstrate movement toward the commercialization process via patent applications, awards and licensure. An average of 4 publications has been generated per award and 16 of the 30 FY2010 IMAT awards have been cited by cancer-relevant research publications. The short-term assessment of outcomes suggested no significant difference in the median between the 2-year and 3-year level of R21 support groups, although there are visual differences in the interquartile ranges and in the first quartile for the number of publications and prestige ratio, with the 3-year level groups having a higher Q1 value.

The evaluation framework and processes established during this evaluation provide a foundation for an on-going evaluation of the IMAT portfolio. Overall, for most analyses in this evaluation, additional analyses are recommended including the incorporation of appropriate comparison groups.

Task 1.0: IMAT applications appear dissimilar from other technology related applications [2]

A unique quality of the IMAT program is their interest in high-risk applications if they demonstrate a strong potential to develop innovative technologies that will advance our understanding, diagnosis or treatment of cancer. The novelty/inventiveness of the IMAT program was estimated using a combination of subject matter expert review and an automated process based on measuring the dissimilarity of specific aims text from the IMAT 2012 application portfolio to specific aims text in a companion cohort of applications from other programs. The companion cohort was constructed to have temporal and topical overlap with IMAT, but also to be large enough to include cases of unanticipated similarity to IMAT so as reduce the chance of an overestimate.

The success of this methodology was contingent on four assumptions

- Dissimilarity can be estimated by a statistical comparison of IMAT text with a companion cohort text rather than by using heuristic rules applied to the IMAT text alone.
- Dissimilarity can be estimated using IMAT and an appropriately selected companion cohort if (a) the two cohorts have substantial topical overlap thereby reducing the influence of text differences

---

[2]Please note the careful use of terminology in Task 1. The deliverable for Task 1was a summary of a set of numbers calculated for each IMAT application. Those numbers are consistently referred to as Minimum Calibrated Dissimilarities (MCD) values or measurements, a term whose meaning will be explained in detail in this report. The goal of Task 1, however was to have the MCD values correlate as well as possible with a descriptive semantic feature of each application for which several terms could be used such as "technical innovation", "distinctiveness", "inventiveness", or "novelty". While such terms may be overloaded with specific meanings in other contexts, we use them here to indicate that an application proposes technologies that are fundamentally new and not just a refinement, extension, or correction of an established technology.

that are obviously not relevant to the IMAT topic areas of interest and (b) when differences exist and are correlated with innovation, IMAT is more likely to display technical innovation than the companion cohort.

- The smallest text dissimilarity between an IMAT application and any companion cohort application is likely to be correlated with novelty/inventiveness.
- Text similarity scores from a standard tool can be calibrated to match manually reviewed "Gold-Standard" ratings of dissimilarity using a model.

The underline procedure to estimate dissimilarity involved seven steps[3]

1. A companion cohort was selected according to the assumptions in bullet two above. The specific aims text of the companion was utilized in the comparison (See appendix A1.1).
2. A gold-standard was established with a subject matter expert (NCI program officers) estimating the dissimilarities between a sample of 101 IMAT-companion cohort application pairs using a fixed scale between 0 and 1 where 1 was most dissimilar (See appendix A1.2).
3. A text similarity score for each IMAT-companion cohort pair was generated using automated text similarity methods. Each IMAT application was scored against each companion cohort application generating a set of scores for each IMAT application (See appendix A1.3. Also, see the supplementary materials for a sample showing the range of similarity/dissimilarity for a single IMAT application.).
4. A model was constructed from the gold-standard dissimilarity ratings and their corresponding auto-generated similarity scores, giving a best-fit formula for converting similarity scores with variable ranges to dissimilarity ratings in a fixed range (0-1) (See appendix A1.4).
5. **The minimum calibrated dissimilarity (MCD) for each IMAT application was then computed as the *smallest calibrated dissimilarity* of that IMAT application compared to any application in the companion cohort.** Higher values of MCD thus correspond to larger dissimilarities to even the most similar companion application (see Figure 1.1). Using the minimum dissimilarity rather than similarity allows our measurement to follow the more common "higher values are better" interpretation. (See appendix A1.6).
6. The accuracy of an MCD score to capture the *actual* dissimilarity of an IMAT application was assessed by a final quality review by NCI program scientists of MCD scores for a sample of 29 IMAT applications.
7. In addition to measuring the distribution of IMAT MCDs for the IMAT->companion cohort pair, we also examined 2 other pairs of cohorts, as an informal null hypothesis test. Since there is an assumed difference in the extent and character of the incentives in the IMAT program and programs that composed the companion cohort, we examined two cohort pairs in which this difference could be assumed to be absent, namely IMAT compared internally to itself and the companion cohort compared to itself. The differences found in the MCD distributions for these 3 cohort pairs are discussed below (See Figure 1.2). Also, several other pairs were formed using sub-cohorts of the companion (groups of projects categorized by research area as indicated by the review study section and, in the case of ZRG and ZCA sections, also by the specific review panel. Distributions for these additional 20 cohort pairs, as well as breakdowns of MCD for selected cohort pairs by selected project characteristics (e.g. funding) are included in the appendix (See appendices A.1.7 and A.1.8)

[3] See the Task 1 Appendix for further details of each procedural step.

Figure 1.1: Minimum Calibrated Dissimilarity (MCD) is computed as the minimum dissimilarity of a given IMAT application to the companion cohort as a whole. Dissimilarity is represented here as analogous to a physical distance, making the MCD, the "closest approach" of an IMAT application to the companion. The lengths of the lines from IMAT to the companion correspond to the calibrated dissimilarity values (longer line=more dissimilar). In this example, A has a higher estimated dissimilarity than B.
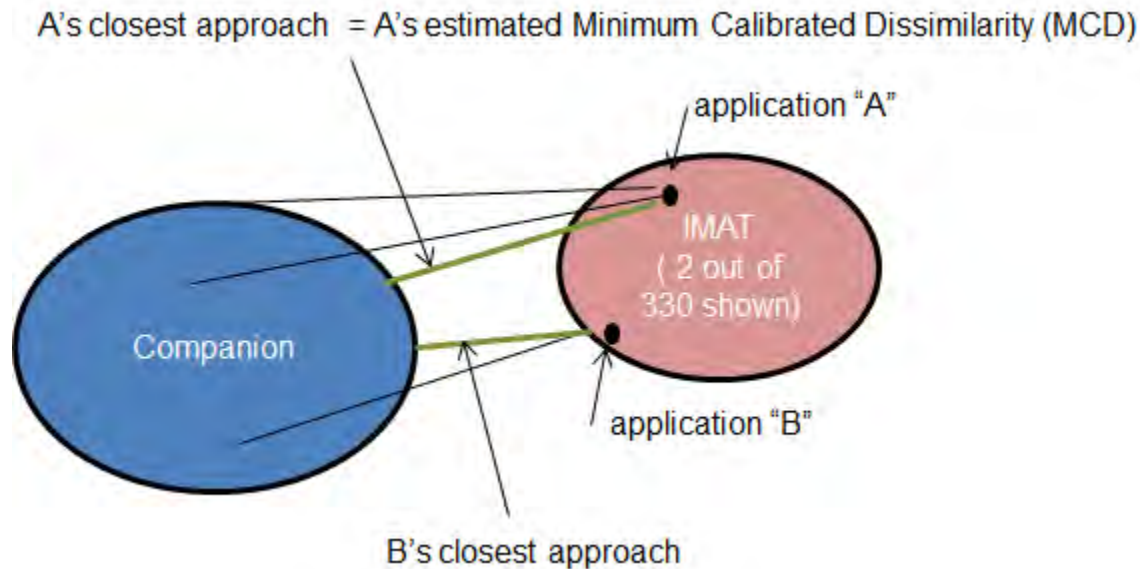
Boxplots of the distribution of MCD scores for the three cohort pairs, IMAT→companion, IMAT→IMAT and companion→companion are visualized in Figure 1.2[4]. The mean and median MCD for IMAT relative to the companion is significantly higher than that of the companion cohort relative to itself, excluding direct comparisons of an application to itself at the project level (1.13 times higher, 95% CI 1.09-1.17). The MCD values for IMAT measured relative to itself are significantly lower than that of IMAT measured relative to the companion, and there is a much higher variability in the MCD measurement for this cohort pair. This may be because IMAT is entirely focused on highly innovative technology development whereas the other study sections invite technology development as well as application of new technologies to pursue hypothesis-based research. It is reasonable that there are several innovative technology proposals competing in the IMAT solicitation with similar goals, compared to the variety of aims proposed to other solicitations.

---

[4]H. Wickham. ggplot2: elegant graphics for data analysis. Springer New York, 2009.

Figure 1.2: Distribution of the Minimum Calibrated Dissimilarity scores for three groups; IMAT→companion cohort, IMAT→IMAT and companion→companion. The scale is between 0 and 1 where 1 was most dissimilar.


Distribution of Minimum Calibrated Dissimilarity (MCD) measurements for selected pairs of NCI grant application cohorts
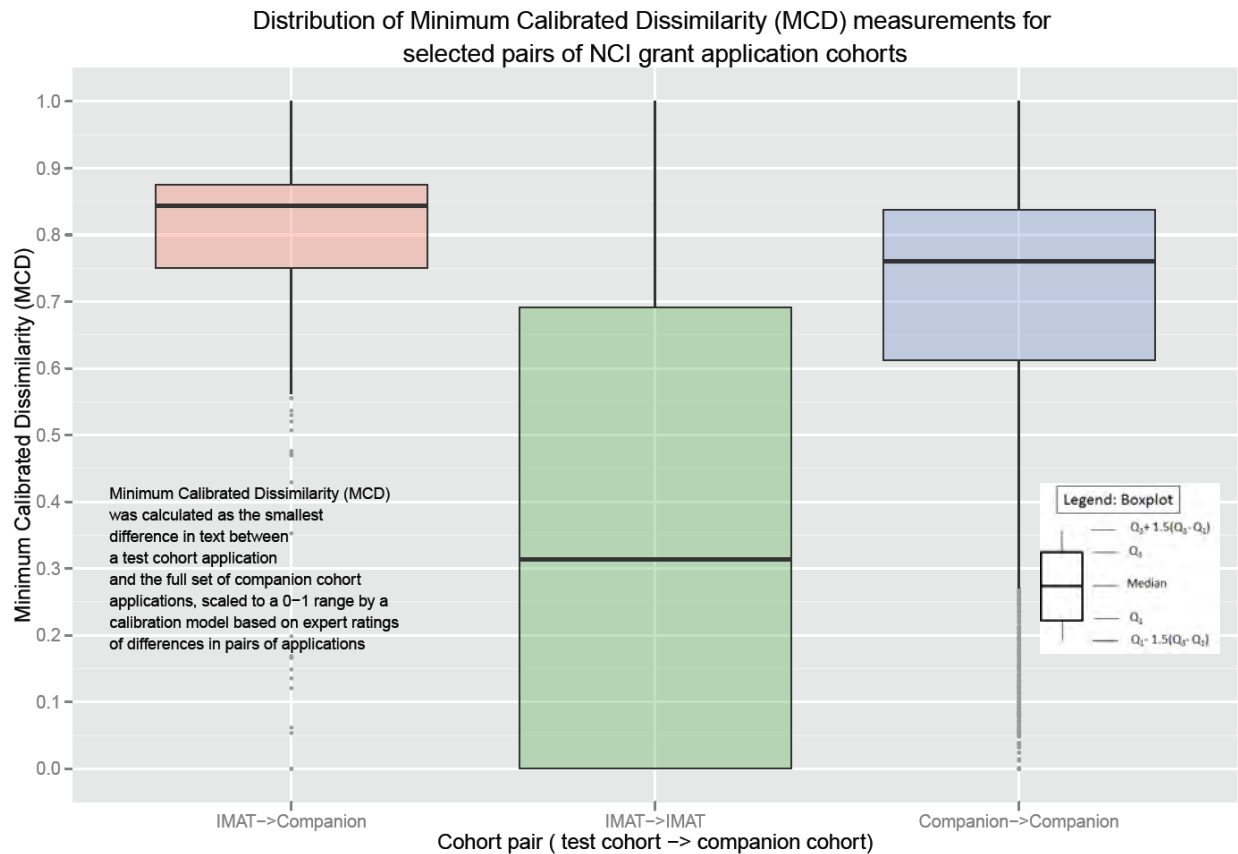
Note that the key distinction between the IMAT->companion measurement and the two "internal" measurements (IMAT→IMAT and companion→companion) is that, for IMAT->companion, a known differential existed between the level and character of the innovation incentives in their respective FOAs, and that this differential was absent in the two internal measurements.

The higher degree of dissimilarity seen in IMAT->companion remains after controlling for mechanism and funding in a simple regression model with MCD as the outcome variable. See Task 1 Appendix A1.7 for more results from this model. Further analysis could 1) examine whether this effect remains significant after controlling for other variables such as institution, geo-location, prior PI performance, PI demographics, and other relevant variables and 2) explore options such as model adjustments, semantic rules, or thresholding/leveling to reduce the error in estimated innovation of individual IMAT applications.

As described in step 6 above, the final quality review of MCD scores for IMAT->companion measurements (see Figure 1.3) resulted in a median error of +0.2 (estimate-true), indicating an overestimation of the expert rating of technical innovation by the MCD score. Options to reduce this error exist as possible future research, which should also include an assessment of the reliability of the error measurement itself (i.e. to what extent is it a reflection of the difficulty, even for a subject matter expert, to rate technical innovation). The errors reduce the practical significance of the MCD difference

for IMAT->Companion vs. Companion->Companion since the median error is roughly twice the observed MCD difference.

Figure 1.3: Results of the final quality review. The error is the subject matter expert's rated MCD minus the calculated MCD from the automated system.

Errors vs MCD values for QA sample (n=29)
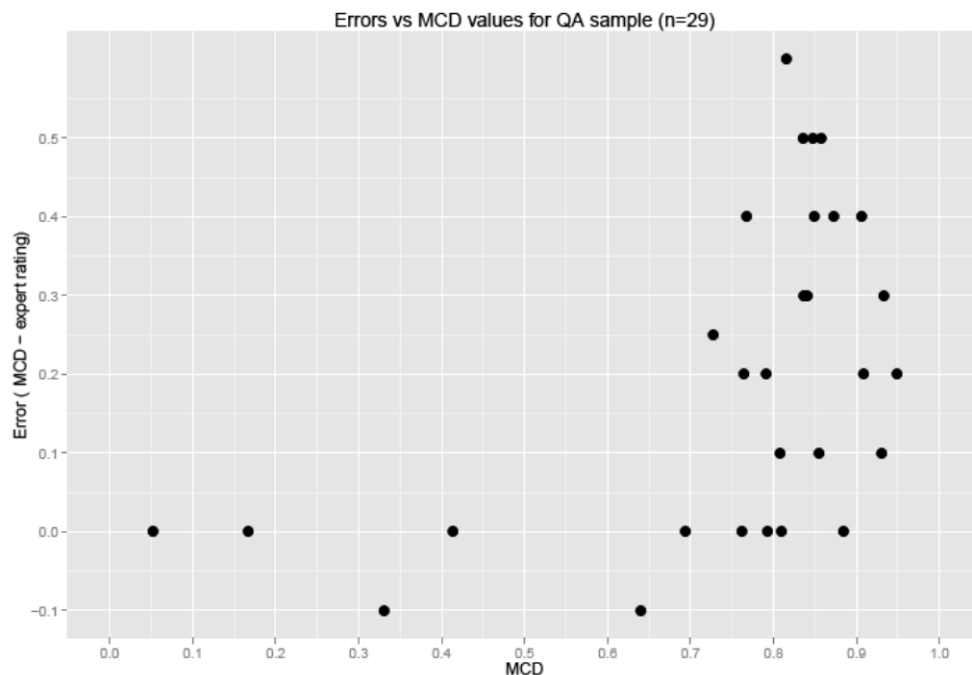
Task 2.0: The IMAT program attracts cross-disciplinary applications

To enable the cross-disciplinary application of innovative technologies to cancer research, the IMAT program is interested in attracting investigators who did not previously focus on cancer research to the IMAT program to support their development of cancer-relevant technologies. This evaluation assessed the IMAT 2012 application portfolio (see Table 1) from two perspectives: (1) whether the applicant was previously awarded a grant by the National Cancer Institute (NCI) and, (2) whether and the degree to which their prior publications were relevant to cancer. The 2012 application portfolio for the IMAT program included 308 different project proposals.[5]

All National Institute of Health (NIH) grants awarded to applicants[6] before 2012 were identified in the IMPAC II database, using a proprietary Thomson Reuters database with disambiguated investigators[7] from IMPAC II in combination with the 2012 IMAT grant application number and applicant email addresses provided by IMAT/NCI program staff.

---

[5] 432 total applications minus 116 non-responsive and 16 resubmitted applications.
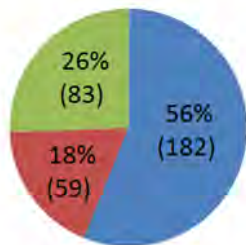[6] Prior publications for the contact principle investigator only were identified for each application.
[7] Using meta-data such as email address associated to each NIH application a single person who has been incorrectly assigned > 1 distinct person identification numbers in the IMPAC II database are unified in the proprietary database.

Publications prior to the IMAT application (between 2007-2011) for these same investigators were identified in *Web of Science* using an evidence-based algorithm. A total of 8,800 prior publications were identified for 308 applications (no prior publications were identified for 16 applications). A total of 5,051[8] of these were categorized into cancer-relevant and non-cancer-relevant publications by performing an analysis of MeSH terms indexed to publications assigned a Web of Science Journal Subject Category (JSC) of Oncology compared to publications in the other 265 Subject Categories (see appendix A2.3). These MeSH terms were examined for the period 2007 to present to identify terms that would be indicative of cancer research. These results were combined with manual review by an NCI subject matter expert resulting in a final set of 1,108 cancer-relevant MeSH terms. An adjusted precision and recall of 100% and 80%, respectively, was achieved validating the acceptability of the publication identification procedure (see A2.1 for description of terms and methodology). If a prior publication was indexed with a cancer-relevant MeSH term it was considered a cancer-relevant publication. See A2.2 for a list of cancer-relevant MeSH terms occurring on the 2012 IMAT applicant's prior publications.

Figure 2.1[9] differentiates the percentage of applications with zero, relatively lower and relatively higher prior cancer-relevant experience based on either having been previously awarded an NCI grant or having previously published cancer-relevant research. Fifty-six percent (N=182) of the 324 applications did not receive a previous grant from the NCI with 111 of these applicants not receiving any prior NIH grants and 71 having received prior NIH grants but no NCI grants. Twenty-one percent of the applications did not include a cancer-relevant publication history.

Figure 2.1: Prior Cancer Relevant Research.

Percentage of Applications with Prior NCI Grants

26% (83)
56% (182)
18% (59)

Percentage of Applications with Cancer-Relevant Prior Publications

29% (86)
21% (62)
50% (147)
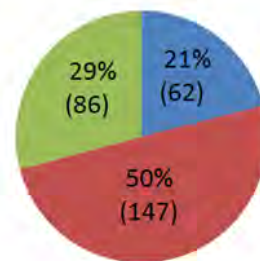
■ Zero prior cancer-relevance

■ Relatively lower prior cancer-relevance (1-50% cancer-relevant prior grant/publications)

■ Relatively higher cancer-relevance (51-100% cancer-relevant prior grant/publications)

---

[8] Only publications with specific document types (742 publications did not have a qualifying document type, see appendix A2.4) and indexed in Medline were included in the analysis (3,007 additional publications were not indexed in Medline, see appendix A2.3).

[9] The percentage of application with prior NCI grants is calculated out of the 324 applications. The percentage of applications with cancer-relevant prior publications is calculated out of the 295 applications that had prior publications indexed in Medline.

Table 2[10] combines the prior NCI grant and prior cancer-relevant publication data. Seventeen percent of the applications have no awarded prior NCI grants *and* no cancer-relevant publication history. Thirty-six percent of the applications have some cancer-relevant publication history but have not yet been funded by NCI (sum of orange highlighted percentages). Less than 42% of the applications included a relatively higher indication of prior involvement in cancer-relevant research by examination of prior grants or publications (sum of bold outlined percentages).

> Table 2: Prior Cancer Relevant Research. This table combines the prior NCI grant and prior cancer-relevant publication data. See Figure 2.1 for the definitions of zero prior, relatively lower prior and relatively higher prior cancer-relevance.

		Prior Cancer-Relevant Publications		
		Zero Prior	Relatively Lower Prior	Relatively Higher Prior
Prior NCI Grants	Zero Prior	17%	26%	**10%**
	Relatively Lower Prior	2%	14%	**4%**
	Relatively Higher Prior	**2%**	**10%**	**15%**

These results indicate that the IMAT program has attracted applicants with limited or no significant attention toward cancer research prior to their IMAT application. Future analysis would benefit from a comparison with other NCI programs as well as accounting for other cancer-associated funding that did not come from NCI e.g. CDMRP, private investment, or an analysis of cancer-relevant NIH grants not funded by NCI).

Task 4.0: The IMAT program facilitates investigators through the patent and commercialization process

The main focus of the IMAT initiative is to support the development of cancer relevant innovative technologies such as cold PCR. Of the two components of innovation, novelty can be measured by an analysis of patent applications and awards, with *transformational* or translational technical innovation understood by measuring successful technology licensing. The IMAT FY2010 award portfolio was evaluated using the United States Patent and Trademark Office (USPTO) database, patent applications and awarded patents reported on the IMAT progress reports and a manual internet search for information on licensure.

---

[10] The percentages are out of the 295 applications with prior publications indexed in Medline (see appendix A2.3).

Table 3: Patent Analysis

Method to Identify Application/Award	Provisional Patent Application	Patent Application	Patent Award	Licensure
Acknowledgement of IMAT Grant Number in Patent Record	0	1	0	0
Match by Technology Short Name and Investigator Name	0	19	2	0
PI and NCI Program Report	3	45	2	6
Distinct Total	3	**37**	**4**	**6**

Table 3 displays the analysis results by the method used to identify the patent data. Sixteen out of the 30 IMAT FY2010 awards applied for patent applications, resulting in a total of 37 patent applications. All patent applications and awards were identified by matching the investigator name and a technology short name (provided by NCI/IMAT program staff, such as 'magnetic sifter', associated to the IMAT award). Two of the IMAT awards have succeeded in obtaining a patent award. These results indicate the work resulting from a large portion of the IMAT awards have some novelty. Since there are multiple applications per a single IMAT award there is an indication that either multiple dimensions of utility exist for a single developed technology or that investigators are developing multiple independent technologies from a single award. Future analysis would benefit from a comparison with other NIH programs that also have the specific initiative to facilitate investigators through the patent award and commercialization process.

Task 6.0: A short-term outcomes analysis indicates limited differences between the 2-year (traditional) and 3-year (non-traditional) IMAT R21 awards

The impact of the IMAT developed technologies on the broader research community contributes to our understanding of their utility, success and in turn their degree of technical innovation. The hypothesis behind this analysis is that differential funding support would result in differences in outputs that may affect the utility of those outputs resulting from IMAT awards.

Additional analysis examined the short-term differences in outcome between the two-year (traditional) and three-year (non-traditional) IMAT FY2010 R21 awards. Citation-based indicators, separated into 4 subtasks (see below), and the patent analysis in Task 4.0 were reviewed.

6.1 Standard and Normalized Citation-Based Bibliometrics: Total Citations, Impact Factor Quartile, One/Two-Year Citation Benchmark Quartile (see appendix A6.1).

6.2 Cancer-Relevant Citing Publications (see appendix A6.2).

6.3 Citations by Clinical Trials (see appendix A6.3).

6.4 Additional metrics for R21 Analysis: PrestigeCites/Prestige Ratio, Distinct Authors, Distinct Addresses, Return on Investment for Publications and Citations (see appendix A6.4).

Variations on some of these metrics such as the maximum, median or percentage in the top quartile were also incorporated into the analysis. Data was presented by publication, by IMAT FY2010 R21 award and by the two- and three-year groups.

Provided in Table 4 below is a summary of the bibliometric analysis for this task. At the time of this evaluation, 8 of the 30 projects[11] awarded in FY2010 were either in no-cost extension or still active. It must be noted that many research projects yield additional publications 2 years beyond the closeout, and an additional 6 months should be allowed for the accumulation of citation activity to assess potential impact. Therefore, the findings summarized below should be considered as preliminary, rather than as final finding for these projects.

Table 4: Summary of bibliometrics based on citations. See Task 6 Appendix for metrics definitions.

	2-yr R21 (15 projects)	3-yr R21 (10 projects)	R33 (5 projects)	Total (30 projects)
Total Publications[12]	53	43	12	108
Average Publications (Maximum)	4 (17)	4 (14)	2 (5)	4 (17)
Average Total Citations (Maximum)	28 (123)	40 (216)	9 (24)	29 (216)
Average Cancer-Relevant Citing Publications (Maximum)	4 (21)	3 (11)	1 (5)	3 (21)
Average Prestige Ratio (Maximum)	29 (69)	40 (77)	18 (50)	31 (77)
Median Impact Factor Quartile (Minimum)	1 (1)	1 (1)	2 (1)	1 (1)
Median Citation Benchmark Quartile (Minimum)	1 (1)	2 (1)	1 (1)	2 (1)

The short-term assessment of outcomes suggested no significant difference in the median (examined with the Mann-Whitney U test[13]) between the 2-year and 3-year level of R21 support groups, although

---

[11] 2 projects still active and 6 projects under no-cost extension.

[12] These publications are indexed in Web of Science with citation data available.

[13] The Mann-Whitney test is a nonparametric test (i.e. does not assume the data follow any particular distribution such as a normal distribution) and provides a more robust, conservative test for the statistical significance of an observed mean or median difference between two groups. The test uses a ranking of the intermixed values observed in two groups. Under the null hypothesis that the two groups do not have a location (e.g. median) difference, it computes the probability (p value) of the observed ranks. If this probability is below the significance threshold (a=0.05), then the null hypothesis is rejected and the group median difference is significant at the 0.05 level. The 2-year and 3-year groups do not show a significant difference by this test. However, the Mann-Whitney also assumes that the two groups have distributions of similar shape. Although the boxplots show obvious visual differences, plots of the estimated distributions are visually more similar. A sensitivity analysis using either alternative nonparametric tests or bootstrap methods would be needed to assess the [continued on next page] risk in applying the test under these conditions, but note that with this low sample size (N=25 / 15 2-year and 10 3-year) there is little power with any test to detect anything except a large effect. The large visual difference in the first quartile (Q1) value for the prestige ratio is caused by having 7 applications in the 2-year group with 0 prestige

there are visual differences in the interquartile ranges and in the first quartile for the number of publications and prestige ratio, with the 3-year level groups having a higher Q1 value. Figure 6 displays boxplots for some variables which suggested slight differences between the groups. Future analysis of these two groups is recommended, using outcomes data collected 2 years after the end date of the grant to provide a longer-range view of outcomes associated with the program.

Figure 6: R21 Analysis. Four citation based bibliometrics comparing R21 awardees funded for two-years or three-years.
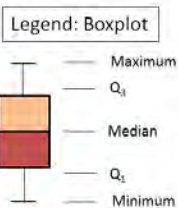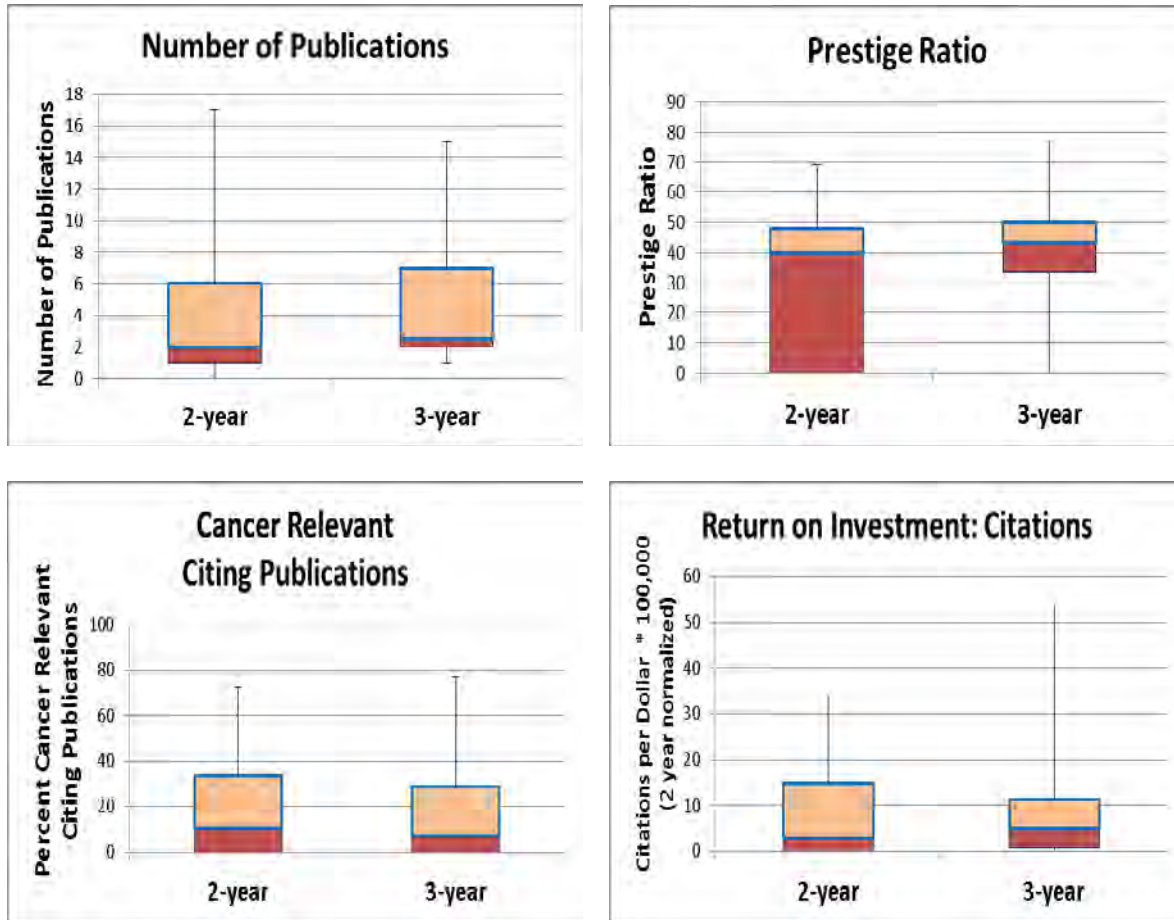
cites and 2 such applications in the 3-year group and by the nature of the ratio -- the lowest non-zero value is 33.33 = 1 prestige cite out of 3 total cites.
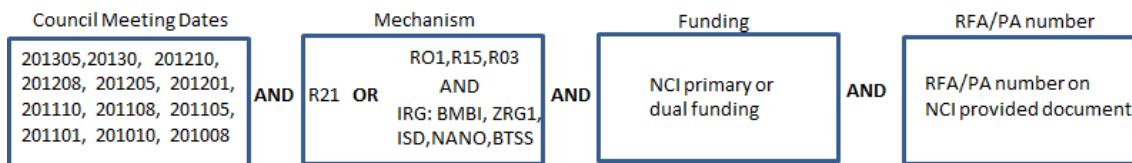
Appendix - Executive Summary

Table A0: Data Sources used for the IMAT Evaluation		Evaluation Task			
Data Source	**Data Source Description**	**1**	**2**	**4**	**6**
NCI Provided Document for 324 IMAT 2012 Applications	application number, mechanism, lead applicant full name, email address, application title and abstract, QVR query rules	✔	✔		
NCI Provided Document for Comparison Cohort	application number, mechanism, project start and end dates, application titles and abstracts, QVR query rules	✔			
NCI Provided Document for 30 IMAT 2009 Awards	award number, mechanism, lead awardee full name, email address, technology short name/key words, progress reports for patent analysis			✔	✔
ScienceWire	Thomson Reuters proprietary database that integrates Web of Science and Medline data	✔	✔	✔	✔
IMPAC II	National Institute of Health application and award database	✔	✔		
MeSH	National Institute of Health's Medical Subject Headings is a comprehensive controlled vocabulary for the purpose of indexing journal articles and books in the life sciences		✔		✔
United States Patent and Trademark Office (USPTO)	United State patent application and award database			✔	

Appendix - Task 1.0: Compare IMAT applications with those reviewed at selected concurrent NIH study section meetings

Additional details on methods and results are for Task 1.0 are included below.

A1.1   Below are listed query rules for identifying the companion cohort in the IMPAC II database. Any IMAT applications were removed from the final results.

Council Meeting Dates		Mechanism		Funding		RFA/PA number
201305,20130, 201210, 201208, 201205, 201201, 201110, 201108, 201105, 201101, 201010, 201008	AND	R21 OR	RO1,R15,R03 AND IRG: BMBI, ZRG1, ISD,NANO,BTSS	AND	NCI primary or dual funding	AND	RFA/PA number on NCI provided document

A1.2   To establish a Gold Standard the NCI IMAT program team was given a list of 101 IMAT-companion cohort pairs selected out of the set of 2 million pairs (from the original configuration in which only the IMAT->companion scores were collected), using 5 subsamples taken by a scrambled index defined against the record identifiers of the IMAT and companion applications. Each subsample was taken from one of 5 ranges of a naïve linear pre-calibrated dissimilarity (0-0.2, 0 .2-0.4, 0.4-0.6, 0.6-0.8, and 0.8-1.0). A 6th subsample was added drawing only from those measurements that appeared likely to be the most-similar to each IMAT (and thus be more likely to contribute directly to the MCD value). All pairs in the sample were provided in scrambled order and were not labeled with any indication of their similarity scores or pre-calibrated dissimilarities. Also, 2 labeled pairs were added with extreme, near maximal, values of similarity and dissimilarity across the entire set of 2.5 million pairs. The IMAT team determined a dissimilarity rating for each pair on a scale between 0 and 10 with 0 being the least dissimilar and 10 being the most dissimilar. This value was then divided by 10 to use in the calibration model.

A1.3   The text similarity score was generated using the Microsoft SQL Server Full Text Search FREETEXTTABLE function (http://msdn.microsoft.com/en-us/library/ms177652.aspx) which is based on the BM25 algorithm (Robertson, S. and Zaragoza H. The Probabilistic Relevance Framework: BM25 and Beyond. 2009. Foundations and Trend in Information Retrieval. 3, 4, 333-389).

A1.4   A model was constructed to predict the Gold Standard (see A1.2), fixed-scale (0-1) dissimilarities from the inherently variable-scale (and reversed scale) similarity scores in combination with the similarity score ranges (minimum to maximum) observed for each IMAT application. The model was fitted using the 101 Gold Standard dissimilarity ratings and their corresponding auto-generated similarity scores and ranges. The model estimated a "calibration 3D scalar field" designed to calculate a dissimilarity for any given score and range combination. Figure A1.1 displays a typical 1-dimensional calibration curve projected from the 3-dimensional best-fit model formula. For a given range (2D slice), the model provides a calibration curve (1D) between the similarity score and the dissimilarity (Figure A1.1, shows the calibration curve for a range from the mean minimum to the mean maximum). Several alternate models were fitted to the Gold Standard ratings. The model with the highest value for the adjusted R-squared measure of model fit was selected as the calibration model (adjusted R-squared = 0.83). This model was a General Additive Model (GAM[14]) based on the formula dissimilarity= f (score, min-score, max-score), where score is the value returned by FREETEXTTABLE for a given IMAT→companion pair of applications, the min/max-score values are the smallest and largest FREETEXTTABLE values observed for that IMAT across the entire companion cohort, and "f", is

---

[14]Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B) 73(1):3-36

some to-be-estimated function of those 3 input variables. The function f was configured as a product of 3 univariate functions each represented as a plate regression spline with a 3-dimensional basis.

Figure A1.1: Calibration curve to compute the dissimilarity of an IMAT and a companion cohort application, if the range of scores for the IMAT application is from the mean minimum to the mean maximum. The model that generates this curve was found to have a low error compared with the human subject matter expert ratings of document similarity (adjusted R-squared=0.83).

Calibration curve determined by model
of Gold Standard vs automated measurements
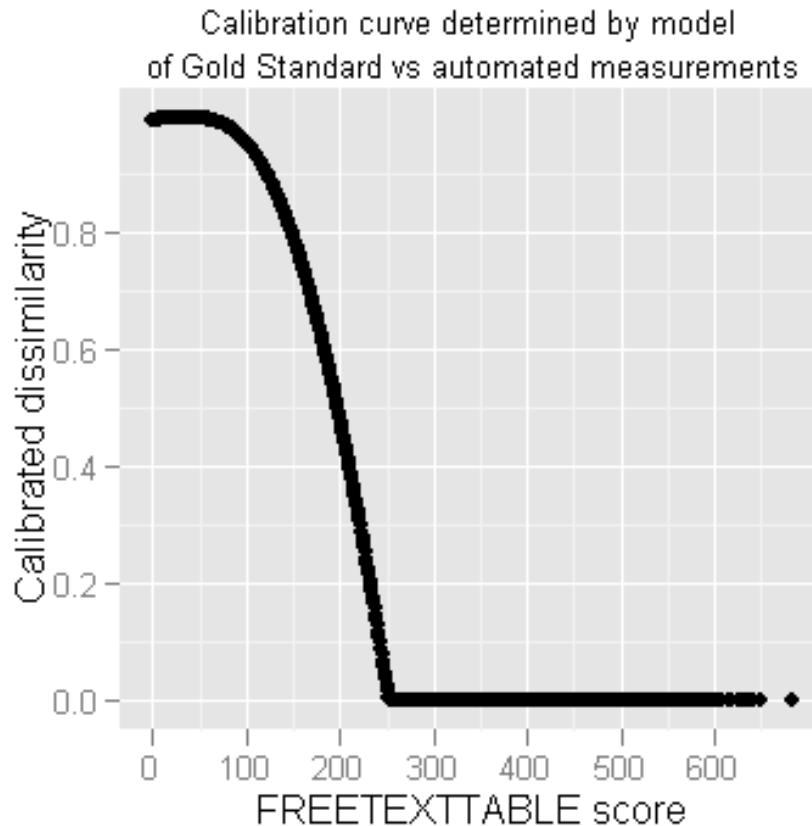
A1.5   The specific aims text was used for both IMAT and the companion cohort. However, for performance reasons only the first 4000 characters of the text of the left-side document in each pair was used. Many specific aims documents are longer than 4000 characters. The sensitivity of the MCD measurement to this text truncation was not investigated.

A1.6   The calibration model obtained from the fit of the Gold Standard dissimilarity ratings to IMAT->companion similarity scores and ranges was extended to calculate the calibrated dissimilarities, and thus the MCD measurements, for the IMAT->IMAT and companion->companion cohort pairs.

A1.7   Additional results: breakdowns of MCD by funding and mechanism. Figures A1.2 and A1.3 below show how the dissimilarity distributions within each cohort pair vary by whether or not the

project was funded or not, and how the distributions vary by mechanism. In a linear regression model of MCD as a function of cohort pair, funding, and mechanism, only cohort pair differences discussed in the main body of the report under Task 1 are significant at the 0.05 level. The lower MCD for R03 is marginally significant at the 0.10 level, however there are only 3 R03 projects in the companion cohort, so this difference is of little practical significance. In a model restricted to only the 330 cases of the IMAT->IMAT comparison, the higher MCD for funded projects is significant, but the confidence interval is wide (31% higher, p=0.03, 95% CI 5% - 60%)

Figure A.1.2 MCD by funding. The left panel shows IMAT funded -> Companion and IMAT unfunded->Companion. The middle panel shows IMAT funded/unfunded -> all of IMAT. The right panel shows Companion funded/unfunded->Companion.

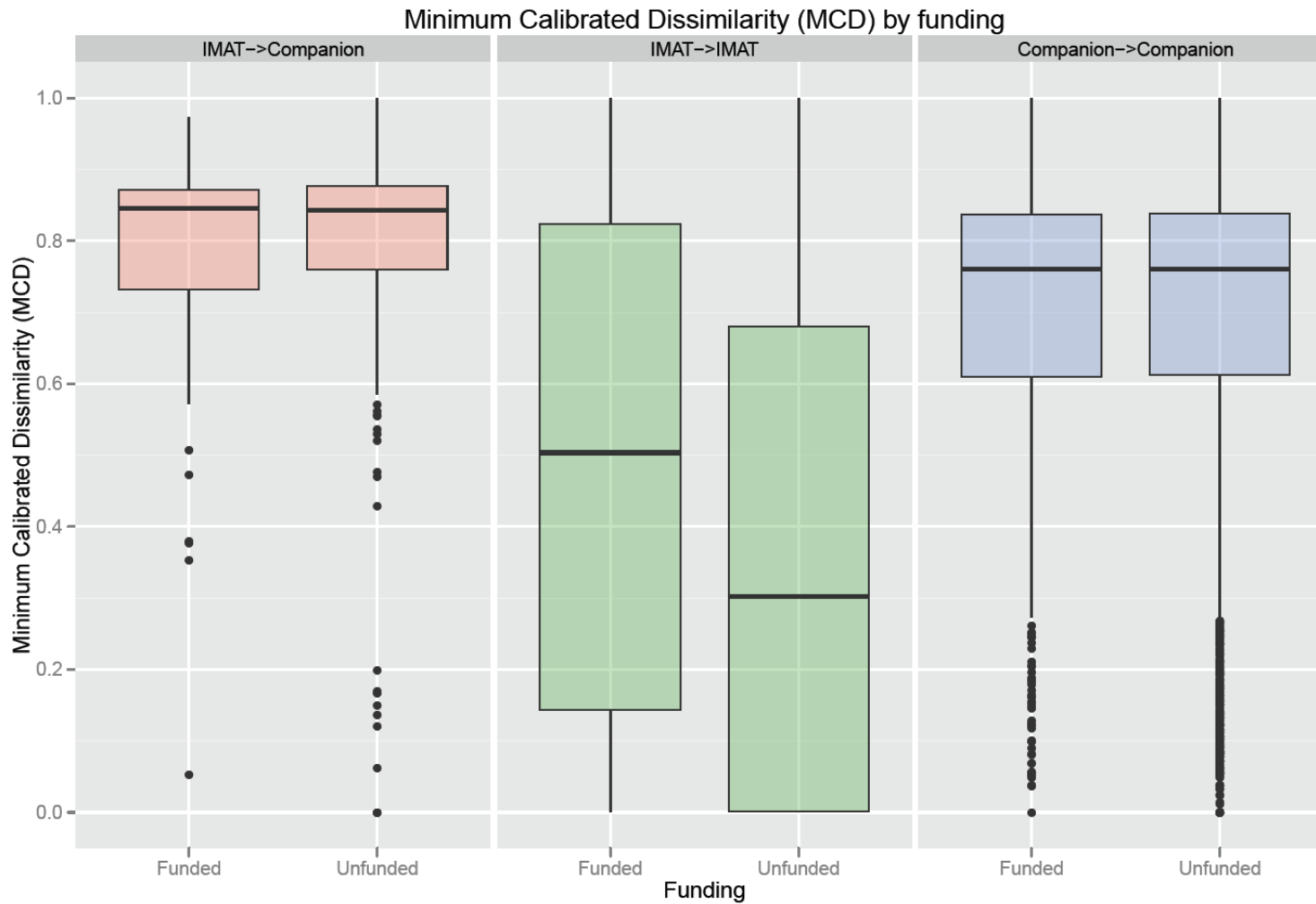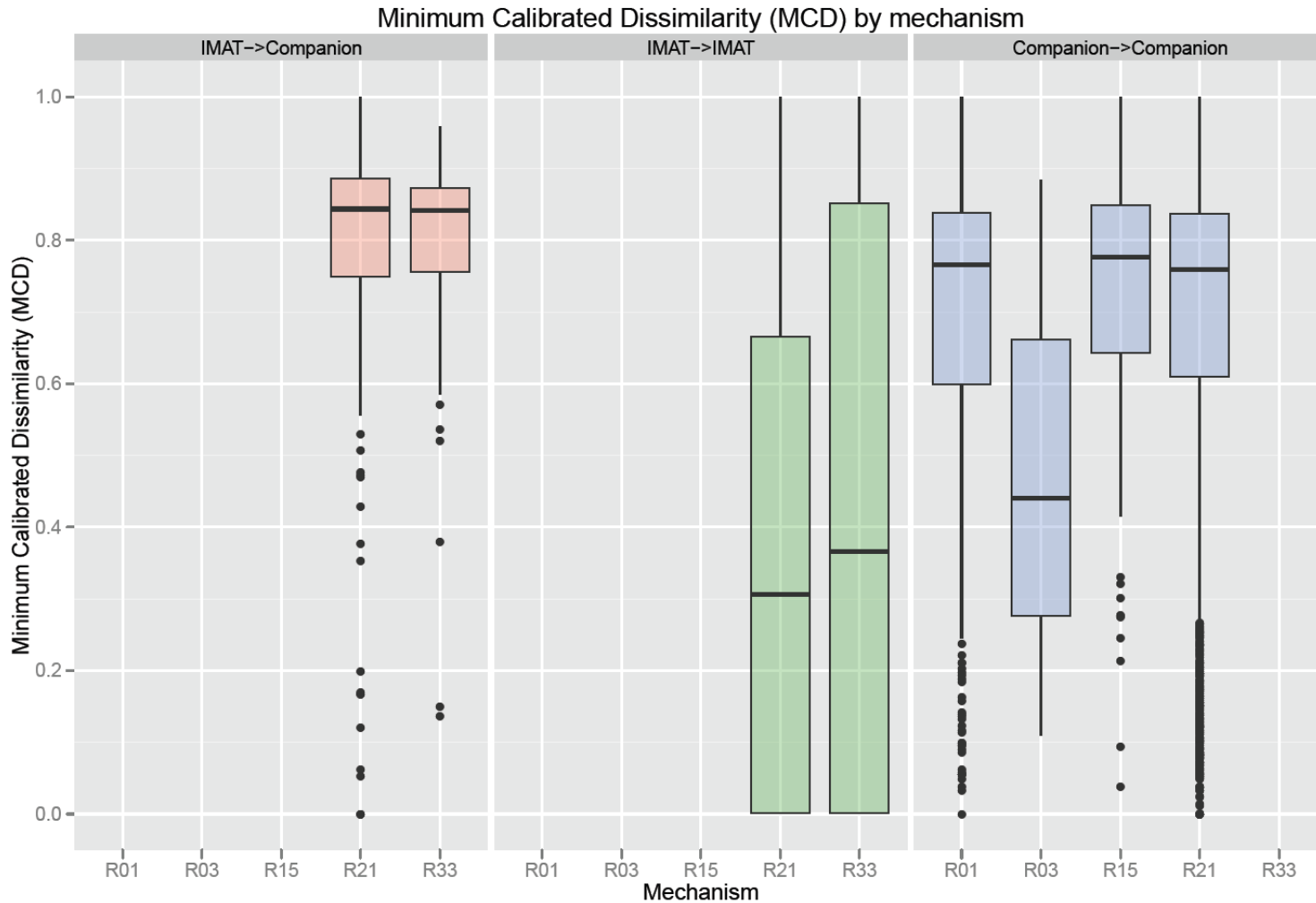Minimum Calibrated Dissimilarity (MCD) by funding

Figure A.1.3 MCD by mechanism. The project cohort on the left of the arrow in the panel labels at the top is broken down into each of the mechanisms shown on the x-axis. The MCD is measured against all of the cohort on the right of the arrow. For example, the left panel shows the 2 distributions of MCD for IMAT R21 -> all IMAT and IMAT R33 -> all IMAT.

Minimum Calibrated Dissimilarity (MCD) by mechanism

A1.8   The first graph, Figure A1.4 below, shows how the distribution of minimum calibrated dissimilarities varies between the entire IMAT cohort and 4 research area subcohorts of the companion cohort, all measured against the full companion cohort. Subcohorts were selected based on study section assignment for each application. The categorization of study sections to either Biotech (predominantly technology development focused review), Bioeng (predominantly bioengineering focused review, with significant technology involved), or Bioinf (predominantly bioinformatics and statistical methods focused review) groups were executed by NCI program officers familiar with these standing study sections. The median MCD for IMAT as a whole is greater than that of any of the companion subcohorts. Projects reviewed under the study sections and special emphasis panels focused on biotechnology, bioengineering, and bioinformatics appear to have a slightly higher median MCD than projects reviewed by other sections/panels. This is analogous to the higher MCD seen for IMAT vs. the entire companion than for the companion-to-companion measurement. In both cases a known enhanced level of focus on technological innovation in the review are reflected in higher MCD measures.

A1.9   The second graph, Figure A1.5 below, shows all 20 combinations generated by splitting the companion cohort into the 4 subcohorts and then measuring minimum dissimilarities for all of IMAT and the 4 companion subcohorts against those very same 4 subcohorts. The dominant feature is that all comparisons to the Other (this category accounts for all study sections that did not fit into any of the three groups described above) subcohort have the lowest MCD values, most likely due to simply having a larger set of applications, so the chance of the companion-wide minimum being in this subcohort is higher. Apparently, just the larger size itself takes precedence over an assumed tendency for higher topical overlap and similar review focus to produce the "closest approach" application. Another noticeable feature is that the pattern of distributions for IMAT is most similar to the pattern for the companion biotechnology subcohort, which an indication of similarity between IMAT and that subcohort, which is confirmed by the "tech" graph for IMAT having the next-lowest MCD after the graph against Other. However, both of these features mostly seem to provide confirmation of the consistency of the MCD calculations and do not provide any new insights into the IMAT measurement. Note that these conclusions are based on a visual comparison of median lines and IQR sizes in the boxplots and have not been cross-checked with a statistical test.

Figure A.1.4 Companion -> Companion MCD measurements, by research area subcohort

Minimum Calibrated Dissimilarity (MCD) for the IMAT project cohort
and the companion cohort research area subcohorts,
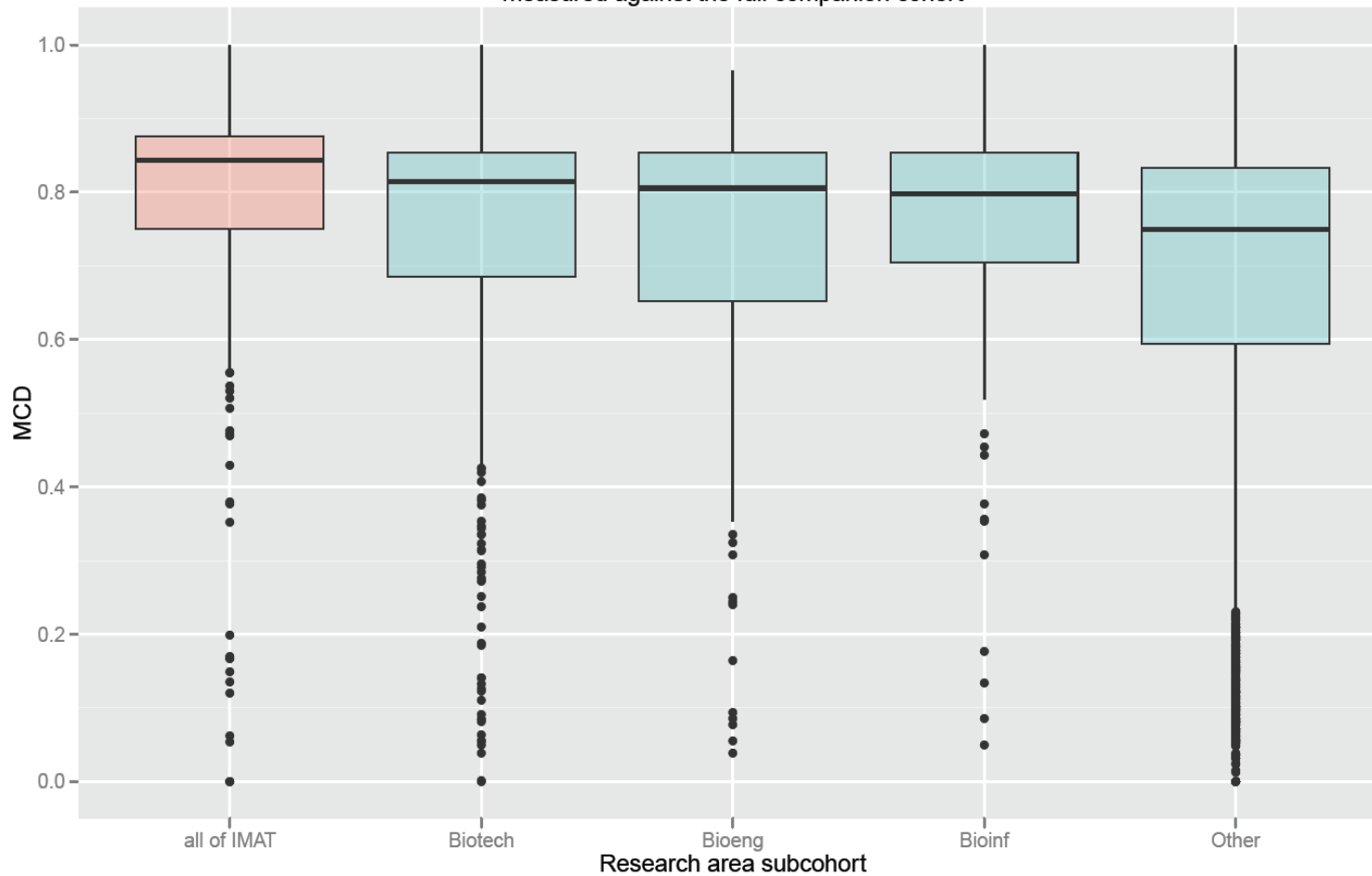measured against the full companion cohort

Figure A.1.5 Restricted MCD for IMAT and the 4 companion subcohorts compared with the companion subcohorts

Restricted Minimum Calibrated Dissimilarity (rMCD) measured for a left−>right pair of subcohorts, with the companion cohort partitioned into research area subcohorts. The left subcohort whose rMCD is measured is shown in the panel label at the top. The right subcohort against which the minimum is taken is shown on the x axis.
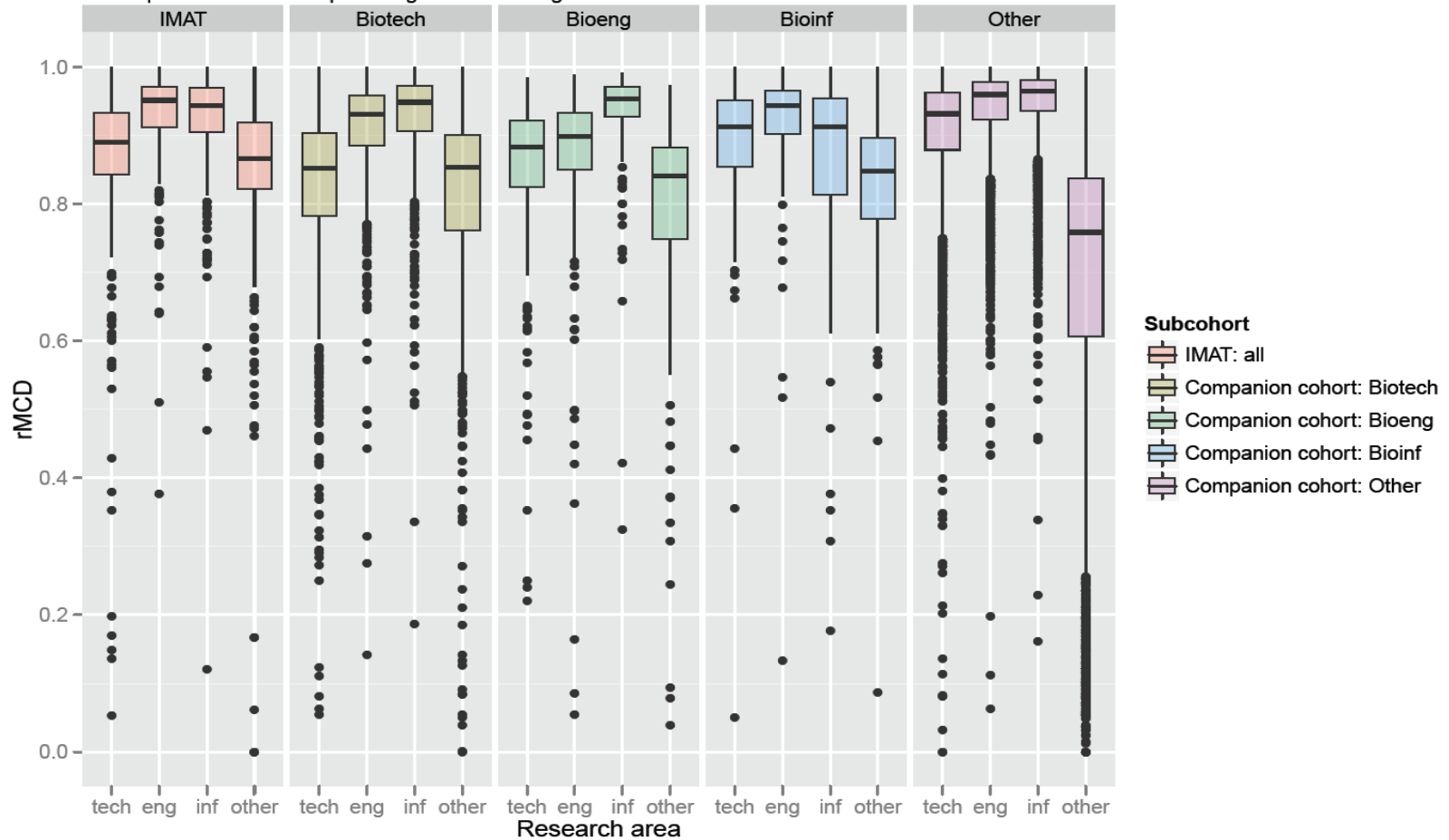
Appendix - Task 2.0: Screen each responsive applicant for evidence of past NCI support or publication of prior cancer research

Definitions, examples of cancer relevant MeSH terms and caveats related to Task 2.0 are listed below.

A2.1   Precision Recall Analysis
Unadjusted Precision/Recall Analysis
Precision measures the percentage of publications in common between the applicant's CV listing of publications and their publications identified in ScienceWire out of the total number of publications identified in ScienceWire.

Adjusted Precision/Recall Analysis
The adjusted precision further qualifies those publications identified in ScienceWire but not listed on an applicant's CV by a manual examination of the publication meta-data such as associated email addresses, co-authors and institutes. Recall measures the percentage of publications in common between the CV and the ScienceWire set out of the total number of publications listed on the CV. The adjusted recall examines if the publications listed on an applicant's CV are indexed in ScienceWire. The unadjusted precision and recall values were 74% and 73%, respectively.

A2.2   List of the top 120 cancer-relevant MeSH terms occurring on the 2012 IMAT applicant's prior publications (2007-2011).

1	Breast Neoplasms	31	Proto-Oncogene Proteins c-akt	61	Wilms Tumor	91	Lymphatic Metastasis
2	Prostatic Neoplasms	32	Neoplasm	62	ras Proteins	92	Gallbladder Neoplasms
3	Pancreatic Neoplasms	33	Burkitt Lymphoma	63	Tumor Suppressor Proteins	93	Epstein-Barr Virus Infections
4	DNA Methylation	34	Glioblastoma	64	Non-Small-Cell Lung	94	Peritoneal Neoplasms
5	Ovarian Neoplasms	35	STAT3 Transcription Factor	65	Protein-Tyrosine Kinases	95	Rhabdomyosarcoma
6	Colorectal Neoplasms	36	TNF-Related Apoptosis-Inducing Ligand	66	Head and Neck Neoplasms	96	Adenoma
7	Neoplasms	37	Glioma	67	Receptor Protein-Tyrosine Kinases	97	Kininogens
8	Melanoma	38	Esophageal Neoplasms	68	Antineoplastic	98	Thymidylate Synthase
9	Lung Neoplasms	39	Neoplasm Proteins	69	Deoxycytidine	99	Antimetabolites

10	Tumor Markers	40	Protein Kinase Inhibitors	70	Proto-Oncogene Proteins c-met	100	Mesothelioma
11	Epidermal Growth Factor	41	Neoplastic Cells	71	Astrocytoma	101	Gliosarcoma
12	Carcinoma	42	Mammary Neoplasms	72	Mitogen-Activated Protein Kinase Kinases	102	Tumor Suppressor
13	Tumor Necrosis Factor-alpha	43	Bone Neoplasms	73	Oropharyngeal Neoplasms	103	Hodgkin Disease
14	Acetylation	44	Medulloblastoma	74	Radioimmunotherapy	104	Matrix Metalloproteinase 2
15	Antineoplastic Agents	45	Proto-Oncogene Proteins c-myc	75	Telomerase	105	Neurilemmoma
16	Brain Neoplasms	46	erbB-2	76	Cyclin D1	106	Adenomatous Polyposis Coli Protein
17	Lymphoma	47	Cerebellar Neoplasms	77	Insulinoma	107	Iron Regulatory Protein 1
18	Urinary Bladder Neoplasms	48	Intestinal Neoplasms	78	CpG Islands	108	Osteosarcoma
19	Colonic Neoplasms	49	Sentinel Lymph Node Biopsy	79	Neoplasm Invasiveness	109	Mouth Neoplasms
20	Liver Neoplasms	50	Neoplasm Metastasis	80	Tumor Necrosis Factors	110	Tumor Virus Infections
21	Leukemia	51	Xenograft Model Antitumor Assays	81	Proto-Oncogene Proteins	111	Proto-Oncogene Protein c-fli-1
22	Neoplastic	52	Precursor Cell Lymphoblastic Leukemia-	82	Estrogen	112	Molecular Targeted Therapy
23	Receptor	53	Anticarcinogenic Agents	83	Quinazolines	113	Oligodendroglioma
24	Uterine Cervical Neoplasms	54	Proto-Oncogene Proteins c-bcl-2	84	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	114	Integrin alpha3
25	Multiple Myeloma	55	Sarcoma	85	Cisplatin	115	Vascular Endothelial Growth Factor Receptor-2
26	Kidney Neoplasms	56	Blast Crisis	86	Neuroblastoma	116	Tongue Neoplasms

27	Skin Neoplasms	57	Thyroid Neoplasms	87	Epithelial-Mesenchymal Transition	117	Nuclear Receptor Coactivators
28	Antineoplastic Combined Chemotherapy	58	Stomach Neoplasms	88	Fas-Associated Death Domain Protein	118	Carcinoma in Situ
29	Tumor Suppressor Protein p53	59	Mucin-1	89	Osteopontin	119	IGF Type 1
30	Adenocarcinoma	60	BRCA1 Protein	90	Cell Hypoxia	120	Choriocarcinoma

A2.3   An analysis dependent on MeSH terms is limited to those publications that are indexed in MEDLINE and assigned a MeSH term. 36% of the prior publications associated to the 2012 IMAT applicants did not have MeSH terms. An analysis based on text extraction of the titles and abstracts of the publications could alleviate the exclusions problem, but was not pursued due to time and cost restrictions.

A2.4   Document Types:
- Included in analysis: Article (journal article), Article (conference proceedings), Letter, Meeting Abstract.
- Excluded from analysis: Biographical item, Comment, Congresses, Correction, Editorial, English Abstract, Historical Article, Interview, News Items, Reprint, Review, Software review, Video-audio media.

Appendix - Task 6.0: Perform bibliometric analyses to characterize quality of publications

Definitions of bibliometric indicators are calculated and reviewed below for Task 6.0

A6.1   Standard and Normalized Citation-Based Bibliometrics
- Total Citations: All citations accumulated up to current date (07/29/13) with ~2 month lag in database.
- Impact Factor Quartile: All journals in a Journal Subject Category (262 categories) are ranked and assigned a quartile according to their Journal Impact Factor within each journal subject category. Quartile 1 is the quartile with the higher journal impact factors.
- One-Year and Two-Year Citation Benchmark Quartile: The one-year (two-year) citation counts of all articles published within 6 months, in the same journal, and the same document type as the IMAT publication being analyzed are ranked and assigned a quartile within the appropriate journal subject category. The quartile position of the IMAT publication one-year (two-year) citation count is then determined where quartile 1 contains those with the higher citation counts.

A6.2   Cancer-Relevant Citing Publications
- The cancer relevance of publications citing IMAT publications was determined using cancer-relevant MeSH terms as described in Task 2. Please see analysis caveat in A2.3.

A6.3   Citations by Clinical Trials

- Identified clinical trials that either cite a publication resulting from IMAT funding or include an IMAT short technology name/keyword. Potential results were manually reviewed. No relevant clinical trials were identified for the IMAT grants that were examined.

A6.4   Additional Metrics for R21 Analysis
- Prestige Cites/Prestige Ratio: The Impact Factor ranking of a citing publication's journal for each of its assigned Journal Subject Categories is obtained as a percentile. If the minimum percentile is <=10%, which is to say that the citation is coming from a journal in the top 10% of journals in that subject category, then the citing publication is considered prestigious. The ratio of the number of prestige cites to the total citations times 100 is the Prestige Ratio.
- Distinct Authors.
- Distinct Addresses.
- Return on Investment, Publications: Ratio of the number of publications to 2-years of the award cost multiplied by 100,000.
- Return on Investment, Citations: Ratio of the number of citations to 2-years of the award cost multiplied by 100,000.