A Blueprint for Drug/Diagnostic Development: Facilitating Development and Use of Curated Genetic Databases

TABLE OF CONTENTS

Forum Goals

Expanding on the 2014 conference that explored regulatory considerations and strategies for FDA approval of an NGS platform, the objectives for this forum are to:

- Define strategies for developing and using curated genetic databases to establish clinical relevance for various intended uses
- Develop a method for evaluating the quality of such databases for different stakeholders, including (for example) researchers, regulators and payers
- Standardize the interpretation and reporting of genetic alterations to databases in order to facilitate interoperability

Summary

High-throughput genomic technologies, including next generation sequencing (NGS), allow for rapid assessment of many analytes, enabling diagnosis, and/or prognosis, of disease and helping to predict a patient's risk for developing certain conditions or response to therapies. There are many advantages to high throughput sequencing tests over those that detect a single analyte; however, with multiple tests and platforms under development, demonstrating adequate analytical and clinical performance for a set of variants with differing intended uses is challenging. There is a particular tension between the need to provide reasonable assurance of the analytical and clinical validity of variants and practical limitations for submitting such data for every possible variant the test could identify. Thus, consensus, evidence-based guidelines are needed at multiple levels, including: sample collection, preparation, and analysis; clinical reporting (including variant calling, interpretation, and report content), and data storage and protection.

With the increased use of genomic profiling in cancer therapy and the development of more powerful technologies, there is great interest in developing new regulatory pathways to facilitate the rapid approval of therapeutics and the accompanying diagnostics that enable precision medicine. Despite recent successes of drug-diagnostic combinations[1], there exists an opportunity to address some of the possible inefficiencies in regulatory review and coverage and reimbursement of these diagnostics. The FDA, aligned with the Precision Medicine Initiative, proposed[2] minimizing the burden of the traditional regulatory requirement for submitting clinical data by each sponsor through permitting the use of existing databases to support the associations between genetic variants and clinical outcomes (i.e., "clinical validity"). A precedent for such an approach was the 2013 FDA-clearance of the Illumina MiSeqDx system and accompanying assays (Cystic Fibrosis 139 Variant and Cystic Fibrosis Sequencing Assays) for diagnosing cystic fibrosis, which relied on a curated database of mutations associated with symptoms of cystic fibrosis (the Clinical and Functional TRanslation of CFTR database). Other efforts are underway to determine whether existing databases, such as ClinGen and ClinVar, could be leveraged to support the clinical significance of variants detected by NGS tests. A number of these databases are publicly accessible, and continually updated, with many additional resources available or in development. Identifying strategies to improve and connect existing private or public systems should facilitate the creation and sharing of critical knowledge and limit potential isolation of current genomic data collection efforts.

A working group, comprised of representatives from the academic, corporate, and government sectors with expertise in pharmaceutical and diagnostic development and regulatory affairs, explored strategies for using curated databases containing up-to-date clinical evidence and correlates of genomic alterations to support clinical validity of sequencing-based diagnostics. The working group focused on the proposals outlined below to

[1] DR Parkinson et al. *Evidence of Clinical Utility: An Unmet Need in Molecular Diagnostics for Patients with Cancer.* 2014. Clin Cancer Res; 20(6)

[2] Optimizing FDA's Regulatory Oversight of Next Generation Sequencing Diagnostic Tests—Preliminary Discussion Paper. 2015. http://www.fda.gov/downloads/MedicalDevices/NewsEvents/WorkshopsConferences/UCM427869.pdf

address the challenges associated with the integration of evidence from databases into NGS-based companion diagnostic (CDx) and complementary diagnostic development. Successful implementation of such an effort could facilitate making diagnostic tests available to patients more rapidly, thus increasing patient access to needed therapeutics by,

- Defining the clinical relevance of rare variants and building on existing knowledgebases
- Promoting sharing of data, standard mechanisms for querying shared data, and development of consistent, transparent data sources to support clinical validity
- Introducing efficiencies into biomarker development and validation
- Improving the design of novel biomarker-driven clinical trials and clinical decision-making using existing biomarkers

While the main focus of this document is addressing the regulatory challenges, the strategies that evolve may also be applicable for addressing coverage and reimbursement needs for payers. It will be important to consider the implications of these proposals on the payer community.

Terminology and Considerations:

With the multiple terminologies in use for genetic alterations (also markers, or variants), we will use the terms interchangeably in this document, preferring the term most commonly applied in a given context.

Additionally, the following definitions will be used in this document:

- Actionable variant (or group of variants): Variants with supporting data that allow for a benefit-risk assessment of treatment choice, link patients to an FDA approved drug (on- or off-label) or an investigational drug enrolled in a registrational trial, or that are prognostic or predictive of outcome. Some markers may be specific (e.g., BRAF V600E), while others may represent a functional group (e.g., alterations in Exon 19 of the EGFR gene or loss-of-function mutations in a tumor suppressor gene such as TP53)
- Regulatory-grade database: Refers to a database that contains genetic information and related data (i.e., clinical) about a disease condition and is recognized by the FDA, or other health authorities, as able to provide evidence of clinical relevance of the variants detected for the intended use of such a test. Such a database would be reflected in the Intended Use statement and/or in the product label of a test
- Research-grade database: Refers to a database that contains genetic and clinical information about a disease condition from one or more research studies that may be supportive of clinical validity, but not recognized by the FDA as sufficient for regulatory approval

For the purposes of this paper, both targeted sequencing and whole genome or whole exome sequencing may be considered; while technical differences need to be addressed between the various sequencing technologies (such as the need for confirmation with orthogonal methods), the link between variant and phenotype should be independent of the tool used to identify the variant. Thus, this document aims to not be limited in scope by the current technologies used in clinical practice and also to be forward looking to those technologies in development.

While the effort to develop potential regulatory-grade databases could facilitate diagnostic development, there are significant challenges to overcome. First, the literature is not comprehensively or consistently curated. Second, even entries within the same database may not be equally valid (e.g., the clinical validity of variants may differ based on intended use). Additionally, the level of evidence supporting a link between a genetic variant and a clinical phenotype will fall into a continuum and investigational or preliminary findings may not be well described. To begin addressing some of the challenges, the working group makes the following proposals:

Proposals

Proposal 1: Define the minimum core data elements required for the interpretation of clinical significance of variants

Broad utility of genetic data within the clinical and research settings requires the establishment of standards of how information is collected and processed, capturing details on the functional consequences of variants and relevant clinical details. Such a set of standards should begin by considering the following aspects for inclusion:

1. Functional classification of variants:
    a. Hereditary cancer risk (germ-line) variants vs. somatic variants arising in tumors. Somatic variants are typically targets of cancer therapy, although some germ-line variants are likewise targetable (e.g., PARP inhibitors for *BRCA1/2*).
        i. Hereditary cancer risk variants are classified similarly to risk variants for other diseases: 1. Benign, 2. Likely benign, 3. Unknown, 4. Likely pathogenic, and 5. Pathogenic. Given the focus of this paper on variants relevant to the care of patients with established cancer and extensive ongoing efforts in hereditary disease risk field, cancer risk assessment is not further addressed here[3].
        ii. Somatic variants may be classified as cancer "driver" or "passenger" mutations or variants whose function is unknown. Somatic driver variants and germ-line cancer risk variants relevant to cancer care (e.g. *BRCA1/2*) should be prioritized for database inclusion and are the focus of these proposals
        iii. Driver mutation clinical implications may be impacted by additional factors which should be captured including:
            1. Clonal vs. sub-clonal status
            2. Level of copy number amplification
2. Grouping of variants into classes ("alteration groups") that may be interpreted equivalently on the basis of either clinical or pre-clinical data (e.g., all *EGFR* exon 19 deletions)
3. Association of variant groups with clinically relevant phenotypes and the indication for which this association would be pertinent, including
    a. Diagnosis
        i. Tumor type (i.e. lung, multiple)
        ii. Stage of cancer
        iii. Prior therapies, etc.
    b. Prognosis
    c. Therapy response prediction, including detail on evidence type (e.g., trial in biomarker selected population vs. all-comers) and type of response data
        i. Responsive (e.g., RECIST, PFS, OS)
        ii. Resistant
    d. Adverse events, safety signals, and patient reported outcomes
    e. Demographic group
        i. Age
        ii. Gender
        iii. Ethnicity
4. Levels of supporting evidence for the association (see proposal 2)
5. Data and technology source to ensure the validity and comparability of the data, including:

[3] Richards S. et al., 2015. **Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology**. *Genetics in Medicine* **17**, 405–423

a. Sequencing technology, sample type (blood, tissue, etc.), and key validation including analytical performance parameters (i.e., analytical performance characteristics with consideration of small or large deletions, insertions, rearrangements, and genomic context)that indicate accuracy, to ensure adequate coverage
b. Establishment of assay/platform-specific standards to ensure adequate analytical performance for various types of variants, such as small or large deletions, insertions, rearrangements, and genomic context

Ensuring the above parameters are captured when a variant is submitted should introduce transparency and consistency to the database. However, while some variants fall into clear categories of driver or passenger mutations, others will have conflicting classifications among researchers, clinicians and laboratories. This may be due in part to the specific disease setting, differences in bioinformatic analyses, and interpretation of the literature. Thus, it may be important to create a mechanism to adjudicate those variants, particularly those that may have clinical implications on regulatory and reimbursement decisions and ultimately treatment or prognosis.

One mechanism is to establish a panel of subject matter experts who would arbitrate inconsistent results and finalize the variant assignment based on the most current evidence. Such a panel would need to be structured, transparent in their methods, and composed of a variety of community experts both broad and deep in scope. Such efforts are already underway for germ-line variant classification, where expert working groups have been organized to apply a set of standards developed in the context of particular diseases. These panels would develop a "master" classification for an individual variant or group of variants, would facilitate review of the pertinent literature, and ensure the classification system is regularly updated. The standards would then need to be applied broadly among the databases used to support clinical performance. While this system would help minimize variability and provide verifiable content within a database, independent review of the raw data would still be needed. In fact, studies to evaluate the concordance among publicly available databases note significant levels of discrepancies in the classification of germ-line variants[4,5]. Further, with the standardization of content, it may be possible to query and generate automatic reports to disclose the level of inconsistency within a database (i.e., dbGaP), thus providing transparency to the degree of inconsistency within a database and among databases.

Proposal 1: Questions for discussion:

- Should additional data elements be considered?
- Are any proposed data elements not relevant and should be de-prioritized?
- Should "raw data" should be deposited into the database?
    - If so, through what mechanism?
- Is inconsistency, as long as it is transparent, acceptable in variant classification? To what degree should the community strive to resolve it?
- Can independent expert panel review help resolve database inconsistencies? If so, how would it be structured and governed? What other mechanisms can be put in place?
- Is it appropriate to include patient reported outcomes or other data in databases of these types?
- Could data be submitted from international sources? Does this pose any barriers?

Proposal 2: Establish a framework to evaluate the strength of evidence for genotype/phenotype associations

[4] Vail PJ, et al. 2014. Comparison of locus-specific databases for BRCA1 and BRCA2 variants reveals disparity in variant classification within and among databases. J Community Genet.
[5] Thompson BA et al. 2014. Application of a 5-tiered scheme for standardized classification of 2,300 unique mismatch repair gene variants in the insight locus-specific database. Nat Genet 46: 107-115.

Assessing levels of evidence for particular genotype/phenotype association will be essential to the appropriate application of any proposed database.

The following table[6] describes a proposed grading system for the level of evidence supporting an association, ranging from the weakest association based upon pre-clinical evidence, at level 3B, increasing in strength to FDA approval (i.e. inclusion on drug label). Classification for a given variant is not intended to be static and will change over time as new evidence becomes available. Thus, to ensure robust classification of the variants, continued evaluation will be needed. This could be accomplished by ensuring that along with every new submission into a database, information is captured regarding: 1) the variant detected and phenotype reported according to the framework presented in proposal 1; 2) the level of concordance with the current classification; 3) descriptions of any differences or changes; and 4) the scientific rationale for these changes. Additionally, while intra- and inter-database discordances cannot be eliminated, the system could also facilitate the ranking of aggregate evidence with conflicting results.

Level		Supporting Evidence
Decreasing strength of evidence ↓	1A	Drug is FDA-approved for the patient's tumor type and indication and outcomes are associated with a specific biomarker (i.e., on-label)
	1B	An adequately powered, prospective study with biomarker selection/stratification or multi-study meta-analysis is available, demonstrating that a biomarker predicts tumor response (or resistance) to a drug or that the drug is clinically more (or less) effective in a biomarker-selected cohort in the patient's tumor type compared to an unselected cohort
	2A	Robust study demonstrating a biomarker is associated with tumor response or resistance to the drug in the patient's tumor type that may not meet statistical criteria for 1B.
	2B	Single or few unusual responder(s), or case studies, show a biomarker is associated with response or resistance to drug in the patient's tumor type, supported by scientific rationale
	3A	Clinical data is available that demonstrates that the biomarker predicts tumor response to drug in a different tumor type
	3B	Preclinical data (*in vitro* or *in vivo* models or functional genomics) demonstrates that a biomarker predicts cell-based response to drug treatment

*Adapted from Meric-Bernstam et al., 2015

Example 1: EGFR mutation in metastatic non-small cell lung cancer (mNSCLC)

a. An illustrative example includes kinase domain mutations in the EGFR gene in mNSCLC. The EGFR tyrosine kinase inhibitors (TKIs) afatinib, erlotinib, and gefitinib are FDA-approved for the treatment of patients with mNSCLC whose tumors harbor exon 19 deletion or the exon 21 L858R substitution mutation based on randomized controlled trials (RCTs) demonstrating improved progression-free survival (PFS), objective response rate (ORR) and favorable benefit-risk in this patient population compared to standard chemotherapy. Detection of exon 19 deletion or exon 21 L858R in mNSCLC could be considered **level 1A** evidence supporting biomarker patient selection for use of a drug
b. Other, less common EGFR kinase domain mutations are thought to be activating and confer sensitivity to EGFR TKIs such as afatinib and erlotinib. For example, exon 21 substitution mutation L861Q and exon 18 substitution mutation G719X are thought to be activating *in vitro* and are thought to confer sensitivity (though less so than L858R and exon 19 deletion) to EGFR TKIs. Patients with mNSCLC whose tumors harbor L861Q and G719X may be administered EGFR TKI in the community. While strictly speaking this may be considered as "off-label" use, pre-clinical and case report data suggest that

[6] Meric-Bernstam F. et al., 2015. A Decision Support Framework for Genomically Informed Investigational Cancer Therapy *JNCI J Natl Cancer Inst* 107(7)

patients with these mutations may be responsive to EGFR TKI, though a prospective RCT is not feasible. Detection of EGFR L861Q or G719X in mNSCLC could be considered **level 2B** evidence supporting the biomarker for patient selection for use of an EGFR TKI

c. In addition, there are activating mutations that are thought to confer resistance to EGFR TKI, such as exon 20 insertion mutations, or the T790M resistance mutation. Pre-clinical and clinical data suggest that first and second generation EGFR TKIs may not be active in patients whose tumors harbor exon 20 insertion or T790M. In the afatinib product label, the "forest plots" of a subgroup of patients with uncommon mutations (predominantly resistance mutations such as T790M and exon 20 insertions) show that PFS and OS tend to favor chemotherapy over afatinib. Detection of T790M and insertions in exon 20 in mNSCLC could be considered **level 2B** evidence arguing against the use of erlotinib and afatinib in these patients

Example 2: ROS1 rearrangement in mNSCLC

a. Crizotinib is approved for ALK-rearranged mNSCLC (~5% incidence) based on improved PFS, ORR, and favorable risk-benefit compared to conventional chemotherapy in RCTs. A recent robust single arm trial in patients with ROS1-rearranged NSCLC (~1% incidence) published in NEJM (Shaw et. al[7]) indicated that crizotinib in ROS1 fusion-positive mNSCLC has a large and durable ORR in this molecular subgroup, and current NCCN guidelines recommend crizotinib for ROS1 fusion-positive mNSCLC. Detection of ROS1 rearrangement in mNSCLC could be considered **level 2A** supporting use of crizotinib for these patients

Example 3: BRAF mutation

a. Dabrafenib and vemurafenib are approved for V600E metastatic melanoma (MM), and dabrafenib in combination with trametinib is approved for V600E and V600K MM. The basis for these approvals included improvements in ORR, PFS, and in certain circumstances overall survival (OS) as compared to conventional chemotherapy in RCTs. Detection of BRAF V600E or K in MM could be considered **level 1A** evidence supporting these biomarkers for patient selection for use of a drug or combination of drugs
b. At ASCO 2015, Planchard et al[8] reported high and potentially durable ORR with dabrafenib in combination with trametinib in patients with V600E positive mNSCLC based on a single arm trial. Detection of BRAF V600E in mNSCLC could be considered **level 2A** supporting use of dabrafenib with trametinib for these patients
c. Data from single arm trials along with anecdotal reports suggest that patients with V600E positive metastatic colorectal cancer (mCRC) have a low response to vemurafenib. Detection of BRAF V600E in mCRC could be considered **level 2A** evidence not to use vemurafenib in this setting

Example 4: *RAS* in mCRC

a. Prior to treatment of metastatic colorectal cancer patients with the anti-EGFR monoclonal antibodies cetuximab and panitumumab it must be determined if the patient's tumor harbors somatic mutations in exon 2 (codons 12 and 13), exon 3 (codons 59 and 61) and exon 4 (codons 117 and 146) of either *KRAS* or *NRAS* (*RAS*) since mutations in *RAS* have been shown to be predictive of resistance. A recent meta-analysis describes results from retrospective subset analyses of *RAS*-mutant or wild-type populations across nineRCTs conducted to investigate the role of *RAS* mutations on the clinical effects of anti-EGFR-directed monoclonal antibodies [Sorich reference]. Treatment with cetuximab or panitumumab was shown to provide prolonged survival over best supportive care and in combination with different chemotherapies over chemotherapies alone for mCRC patients having tumors that are *RAS* wild type.

---

[7] Shaw AT, et al. 2014. Crizotinib in ROS1-rearranged non-small-cell lung cancer. N Engl J Med. 371(21):1963-71

[8] Planchard D., et al. Interim results of a phase II study of the BRAF inhibitor (BRAFi) dabrafenib (D) in combination with the MEK inhibitor trametinib (T) in patients (pts) with *BRAF* V600E mutated (mut) metastatic non-small cell lung cancer (NSCLC). Presented at ASCO 2015. J Clin Oncol 33, 2015 (suppl; abstr 8006).

Average overall survival and progression free survival were significantly improved for the different anti-EGFR drugs regardless of line of therapy or therapy combination with hazard ratios of 0.87 (95% CI 0.77-0.99) and 0.62 (95% CI 0.50-0.76) respectively. Detection of *RAS* mutation in exon 2 (codons 12 and 13), exon 3 (codons 59 and 61) and exon 4 (codons 117 and 146) could be considered **level 1B** evidence arguing against use of cetuximab or panitumumab for these patients and level 1A if clinical utility is validated for a RAS companion diagnostic in conjunction with a RCT for one of these anti-EGFR agents.

Ongoing evidence assessment according to the framework proposed above will be required to ensure that the best available data and evidence are utilized at any given time to determine which variants are most relevant to patient care.

Proposal #2: Questions for discussion:

- Are the levels of supporting evidence suggested appropriate? Are they consistent with the work that others in the field are undertaking?
- How should database consistency be captured and presented?
- Can retrospective studies be sufficient? Or are prospective confirmatory trials always required?
- What level of data should be reported to physicians and patients?
- What other level-of-evidence information should be retained and accessible in the database?
    - Data may become outdated with time, or determined to be unreliable. Is there any way to minimize the ability of this data to skew information?
- Should this type of categorization include prognostic as well as predictive markers?

Proposal 3: Determine the context of use of a database that encompasses information outlined in proposal #1 and supported with the levels of evidence captured in proposal #2

After the challenges associated with database creation are addressed, it is possible to specify how such data may be used. To begin addressing utility, the working group considered several applications, including 1) Research-grade (i.e., for use in scientific research and publications), 2) Regulatory-grade (i.e., for use in regulatory approvals), and 3) Reimbursement-grade (i.e., for use in the clinical setting).

Provided database characteristics outlined above, i.e. capturing the core data elements as well as the strength of evidence for variant/phenotype association, the working group proposed an approach, summarized in the figure below and adapted from the current clinical trial paradigm for developing companion diagnostics (Appendix 1), for determining the application of associations in the database. Briefly, associations with level 3 evidence, reflecting pre-clinical or clinical data from other tumor types, would be appropriate for further investigation and early clinical trials to determine more clearly the appropriateness of matching variant with therapy. There may be a possibility that associations with stronger evidence, level 2, could be used to support clinical validity, for example as part of a PMA review process for a CDx drug-diagnostic pair. Associations with level 1 evidence may possibly not require additional FDA review, only analytical validation of variant detection. Labeling for such a test would need to reflect the database used, the scientific rationale supporting the association, whether there was any discordance between the diagnostic test and database for the specific variant or grouping of variants and rationale for providing the result of such variant, etc.

Thus, elaborating on selected case studies above:

Example 3b: A diagnostic test that detects the BRAF V600E mutation in mNSCLC (level 2A), for use of the dabrafenib/trametinib combination therapy, may use a database capturing the therapeutic association between BRAF V600E and durable ORR in mNSCLC patients treated with dabrafenib in combination with trametinib to support clinical significance for FDA test approval.

Example 4a: Alternatively, a diagnostic detecting RAS-mutation in exon 2 (codons 12 and 13), exon 3 (codons 59 and 61) and exon 4 (codons 117 and 146), considered level 1B**,** may only require analytical validation of variant detection to guide against the use of cetuximab or panitumumab for these patients with mCRC tumors.

Capture minimum core data elements (proposal 1)				
↓				
Establish the strength of supporting evidence for variant – phenotype association: (proposal 2)				
Level 1		**Level 2**		**Level 3**
↓		↓		↓
Sufficient to meet definition of clinical validity for CDx and reimbursement purposes		FDA review of clinical validity required to assess adequacy for CDx claims and need/feasibility of additional studies		Further investigation via genotype-selected clinical trials

Proposal #3: Questions for discussion:

- Is this level-based framework appropriate to guide database use for regulatory approval?
- How could this framework translate to other contexts, e.g., payer community?
- If there are multiple databases that meet the requirements of a FDA-grade database, but contain discordant results, how will these be resolved? How would products that relied on these data be affected? How would the FDA resolve the discrepancy?
- What is meant by 'analytical data only'? (What is required to establish acceptable analytical performance?) When specific claims are made as to variants identified by a test, are accuracy studies required?
- How should data from databases be analyzed? E.g., Statistical analysis for hypothesis generation versus validation?

Proposal 4: Incentivize data sharing into publicly available research- and regulatory-grade databases

A significant fraction of data being generated from genetic testing is not publicly accessible. While there are multiple contributing factors, two key data sharing issues are:

1. clinical laboratories have limited resources to collate and upload their genetic findings into publicly accessible databases and
2. sponsors building internal databases are reluctant to make information publicly available due to competitive considerations

Given that accurate interpretation of genomic data is essential to patient health, developing incentives to facilitate all laboratories (small and large, public and private) to share data should ultimately help all players develop and perform more effective diagnostic tests. Following are suggestions to address these issues.

Data Sharing Issue #1:

Many laboratories, particularly small labs, do not have the resources or the technical expertise required to gather and format data in a manner that is conducive to inclusion in clinical databases. Further, organizations

that are aggregating data may not have funding to reformat data from laboratories or create sophisticated pipelines that make data sharing easier. In order to address these issues, the working group suggests:

- Developing proposals for either establishing or upgrading databases to include the costs of creating easy-to-use pipelines to facilitate data transfer, or sufficient funds to allow for manual upload of data from smaller laboratories
- Creating mechanisms for industry stakeholders to either participate more broadly in the creation of public databases (without establishing ownership) or provide non-restricted funding to aid in their creation

Data Sharing Issue #2:

Competitive incentives often do not align with data sharing. For example, for FDA-cleared or -approved diagnostics, sponsors may be required to perform studies to establish the clinical validity of the marker the test detects. While this process sets a high standard for ensuring test efficacy and safety, it is also often lengthy and resource intensive for sponsors with limited value for sharing these data as it may lower the bar for market entry by subsequent product developers. For LDTs, the aggregated variant database is often proprietary, providing a critical competitive barrier needed to be successful in a highly competitive marketplace. In order to address these issues, the working group offers the following suggestions for discussion:

- Provide sponsors that contribute data to public databases access to priority review paths that provide opportunities for intensive guidance on efficient Dx development and accelerated review timelines for FDA-cleared or –approved products
- Similar to FDA approved pharmaceutical products, provide competitive protection (i.e., similar to patenting) that gives an organization that has made the investment to aggregate data a time-bound competitive advantage in the marketplace
- If data from sponsors are published in peer-reviewed literature the data must also be shared with at least one publicly available database (if available)

Finally, the development in recent years of FDA programs intended to facilitate and expedite the development and review of new drugs to address unmet medical need (e.g. fast track designation, breakthrough therapy designation, accelerated approval, and priority review), have greatly accelerated the approval of therapies for the treatment of serious or life-threatening conditions. In some disease settings such as oncology where targeted therapy development is highly dependent on the co-development of a CDx, the accelerated approval provisions have created unrealistic timeline requirements for the development of CDx. In a prior whitepaper published by this working group[9], it was proposed that an expedited review path be created to ensure that both drug and CDx can reach the market within an optimal timeframe. The use of aggregated publicly available databases to support the clinical relevance of variants in the CDx genes may enable cleared or approved products to reach the market expeditiously. Simultaneous inclusion of data submission requirements as a qualifying criteria for expedited review may help to establish a virtuous feedback loop and provide sponsors with the framework needed to share proprietary data sources.

Proposal #4: Questions for discussion:

- Would establishments of database standards alone incentivize data sharing? If so, would focusing on implementation strategies result in improved data sharing? I.e., development of bioinformatics support and data formatting, to facilitate submission and curation
- Would development of an alternate pathway to submit data for a diagnostic test that requires database submission and community/expert panel review to be a compelling path for regulatory approval?

[9] A Blueprint for Drug/Diagnostic Co-Development: Next-Generation Sequencing (NGS) in Oncology. 2014. http://focr.org/sites/default/files/FOCR-NGS-Report-115.pdf

- Is it possible to begin with promoting sharing into research-grade databases, using the established metrics, such that these could be elevated to regulatory-grade with improved evidence base? What would such a proposal look like?
- What legal mechanisms exist for expediting regulatory approval of CDx?
- What would competitive protection for diagnostics look like?
- Are these incentives sufficient to promote data sharing?
- What are the drawbacks of these suggestions?

Appendix 1: Using the current clinical trial paradigm, this figure illustrates the strength of evidence used to support CDx approval. Extrapolating from this process of approval, the working group suggests applying this model to broader utilization of genetic databases.