Technical Note

PierianDx Knowledgebase

pieriandx

Introduction

As clinical NGS testing volumes grow in light of the FDA's recent approval of NGS-based *in vitro* diagnostics [2], the already time-consuming process of biological interpretation and variant classification will become even more challenging as our body of knowledge grows.

To address these challenges, PierianDx has developed Clinical Genomics Workspace. Driven by a comprehensive and well-annotated Knowledgebase that fosters sharing of data across medical sites, Clinical Genomics Workspace enables streamlined clinical genomics workflows for rapid and accurate interpretation and reporting of genetic variants.

This technical note delves into the engine that powers the rapid interpretation and reporting component of Clinical Genomics Workspace: The PierianDx Knowledgebase. The PierianDx Knowledgebase works with a powerful rules engine that includes an automated learning facet within Clinical Genomics Workspace to auto-classify variants and populate draft reports with relevant information that molecular pathologists and other medical staff can review and sign out.

Overview

The PierianDx Knowledgebase is a key component of Clinical Genomics Workspace, which provides informatics, genomic interpretation, and reporting workflow support for cancer and germline molecular diagnostic testing applications.

Developed at Washington University in St. Louis in one of the first laboratories to offer clinical next generation sequencing, Clinical Genomics Workspace is in use by over 50 health systems, academic medical centers, cancer centers, children's hospitals, and reference laboratories. It addresses all phases of a typical NGS testing workflow after a sample has been sequenced (see Figure 1).

Following secondary analysis of data from a sequenced sample, Clinical Genomics Workspace classifies variants using a rules engine and automatically creates draft reports that are populated with classification-specific content from the PierianDx Knowledgebase.

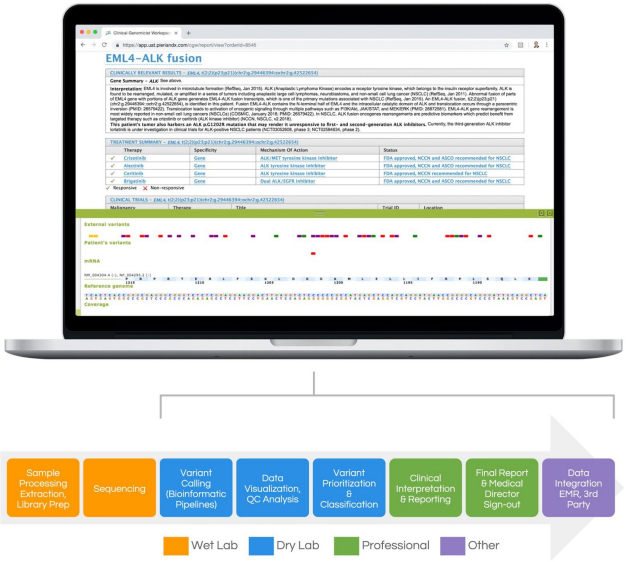

Figure 1: Clinical Genomics Workspace

After draft reports are generated, variant scientists and medical professionals follow a review workflow to review report content and supporting annotations and perform final sign-out of the clinical report for a patient.

Clinical Genomics Workspace also enables users to revise or add an addendum to the report at a later time after signing out a case. After the report is signed out, Clinical Genomics Workspace can be configured to automatically route the report to the electronic medical record/electronic health record

(EMR/EHR). Users can also export a PDF version of the report from Clinical Genomics Workspace.

Although the PierianDx Knowledgebase is an integral part of Clinical Genomics Workspace, the two can be decoupled, and the PierianDx Knowledgebase and rules engine can be accessed through the available application programming interface (API).

Knowledgebase Sources

The PierianDx Knowledgebase includes a broad range of publicly available, proprietary, and user-created content. It includes:

- Curated assertions from:
    - FDA approved labels
    - Professional practice guidelines from National Comprehensive Cancer Network (NCCN) and the American Society of Clinical Oncology (ASCO)
    - Active, recruiting clinical trials from clinicaltrials.gov
- Shared medical content from tens of thousands of cases and millions of variants containing classifications and interpretations from the growing network of PierianDx customers running 150+ NGS panels
- Knowledge content derived from reviewing the published literature from over 130,000 peer-reviewed journal manuscripts
- Public data sources, such as The Cancer Genome Atlas (TCGA), COSMIC, and ClinVar

Built using a highly scalable and robust cross-platform database management system that supports Windows, Mac, and various Linux distributions, the PierianDx Knowledgebase uses a combination of software and human curation to determine additions or updates to content.

Genome Annotation Sources

Genome annotation sources that are loaded, versioned, provenance tracked, and maintained in the PierianDx Knowledgebase are normalized to the disease ontology SNOMED-CT where applicable. These include the following:

- Population frequency databases
    - gnomAD
    - ExAC
    - NHLBI Exome Sequencing Project
    - dbSNP
- TCGA
- ClinVar
- COSMIC
- dbNSFP
- SNOMED-CT

Note that while the structure of the data may be transformed to match that necessary for the PierianDx Knowledgebase, there is no manual curation done on these sources.

Systemized Curation

For those sources that are curated, the PierianDx Curation Team, which is comprised of PhD- and MS-level scientists, uses a systematic approach to review content on a weekly basis. This team screens the content, annotates it, and performs quality control checks over the following week. This process is cumbersome and time-consuming and involves turning scientific prose from different sources (manuscripts, FDA drug labels, etc.) into logical rules that the PierianDx Knowledgebase can call to properly annotate variants.

The newly curated content and logical rules created by the PierianDx Curation Team are then loaded into the PierianDx internal quality assurance environment, where they undergo functional testing. The result of this testing is a Knowledgebase Quality Control Report, which is reviewed and approved or rejected by the PierianDx Change Control Board.

If the new content are approved, they are released to the Test environment, to which PierianDx staff and customers have access, and to the Production environment shortly thereafter.

Medical Interpretations

In addition to the curated and genome annotation data sources, the PierianDx Knowledgebase also includes published medical interpretations. These interpretations originate from two sources:

- The PierianDx Interpretation Services Team
- PierianDx customers who publish interpretations on signed-out cases

The PierianDx Interpretation Services Team, which is an expert and clinically trained team of variant scientists led by the PierianDx board-certified Medical Director, routinely publishes new interpretations, and these are added to the PierianDx Knowledgebase. Likewise, when each PierianDx customer publishes an interpretation, it is automatically added to the PierianDx Knowledgebase.

Key Differentiators

The PierianDx Knowledgebase uses a highly dynamic metadata model and a rules engine to distinguish itself from other knowledgebases that are currently available.

Metadata Model

Due to the rapidly evolving practice of genomic medicine, the PierianDx Knowledgebase (and Clinical Genomics Workspace) has been architected to incorporate new data sets and technologies in a flexible manner. All data sets incorporated into the PierianDx Knowledgebase, including reference sequences, genomic annotations, external variant annotations, knowledgebases, and shared interpretation content vary in their update frequency and composition over time. PierianDx uses metadata to describe the model of these datasets, which enables the PierianDx Knowledgebase and other solutions to rapidly adopt future changes in assay technologies and the evolving nature of biomarkers such as tumor mutational burden (TMB) and microsatellite instability (MSI).

Rules Engine

Inferencing of knowledgebase content into clinical reports is a key requirement of any clinical reporting solution. PierianDx has a distinct approach to inference which differentiates it from other knowledgebases, and like the use of a metadata model, future-proofs the solution for changes in assay technologies and the evolving nature of biomarkers.

Other knowledgebases are limited to simple variant lookups, which means that if the variant being queried has not previously been curated, then the resulting report will not include any information on that variant. In contrast, the PierianDx Knowledgebase uses a comprehensive model that allows for the creation of complex rules on variations within or across genes--not only on variants using syntax but also having particular characteristics such as falling within particular exons, domains, fusing with known or novel partners, or requiring particular copy number

changes. This means that even if the variant is unknown or has never been seen before, then the PierianDx Knowledgebase will look across all other variations, apply logic inferred from an unlimited number of other gene-variant-disease combinations, and return rationalized content in the report.

For example, rules are able to be written on deletions that are in frame in exon 19 of EGFR, or in frame deletions in exon 11 of KIT. Rules are also able to be written on one or more genes that are tied to a particular tumor type, such that the combination of variants that are found in a particular patient's case could have a different interpretation than for any one variant found alone.

Rapid, Accurate Interpretation

In addition to curated sources, such as NCCN/ASCO guidelines, FDA therapies, and clinical trials, the PierianDx Knowledgebase continually benefits from shared genomic interpretations that originate from the PierianDx Interpretation Services Team or PierianDx clinical customers.

Each interpretation is curated with an acute focus on its impact for clinical care, and each receives medical oversight through review and sign-out by either PierianDx's board-certified Medical Director or by a board-certified molecular pathologist or clinical geneticist at one of PierianDx's partner sites.

This unique aspect of the PierianDx KnowledgeBase enables comparison of classifications and interpretations across medical directors at more than one site and enables the use of real-world evidence to make knowledge assertions more rapidly and accurately.

Comprehensive

In all, the PierianDx Knowledgebase encompasses over 28,000 rules across over 6 MB of sequence coverage (6 million genomic positions) across 1,100 somatic cancer genes and supports all variant types: single nucleotide variants, insertions and deletions, copy number variants, fusions, and structural variants.

Up-to-Date

With real-time publishing of medical interpretations and regular updates to industry guidelines, FDA therapies and clinical trials, the PierianDx Knowledgebase contains the most up-to-date and clinically actionable information for next generation sequencing data.

Classification of Variants

The sources within the PierianDx Knowledgebase populate the Variant Details page within Clinical Genomics Workspace, and they also interact with the rules engine within Clinical Genomics Workspace to auto-classify variants and populate draft reports.

Classification Rule Sources

Some sources within the PierianDx Knowledgebase are prioritized above others when it comes to auto-classification. Clinical Genomics Workspace will prioritize sources in the PierianDx Knowledgebase in the following order.

1. Interpretations -- a customer's institution or published interpretations from other customer sites.
2. Practice guidelines/drug labels from NCCN, ASCO, and the FDA.
3. Databases--either population frequency databases or clinical variant databases such as ClinVar, COSMIC, etc.

Classification Schemes

Clinical Genomics Workspace has the ability to classify variants as:

- Polymorphisms
- Variants of unknown significance
- Variants that are clinically significant in a patient's disease/tumor type
- Variants that are clinically significant in other diseases/tumor type

The different sources within the PierianDx Knowledgebase drive different variant classifications.

Rules Engine

Clinical Genomics Workspace is able to classify variants and populate draft reports with relevant PierianDx Knowledgebase content using logic that is composed with a set of rules. The rules are expressions that when all conditions are met, lead to inference of certain content.

For example:

> IF (Condition A = True) AND (Condition B = True)
>
> THEN show Y content in Variant Details and classify variant as level X

When a rule's conditions are met with certain attributes or conditions--a predicate-- then the rule initiates and the annotated content appears in the draft report.

The predicate model is not a simple variant lookup, unlike other clinical genomics knowledgebases. Namely, it is capable of inferring on a variant based on the following:

- Human Genome Variation Society (HGVS) syntax (g-, c-, or p-syntax)
- Genomic, coding, or protein coordinates (e.g., codon ranges or exon ranges)
- Functional characteristics (e.g., frameshift, in frame, truncating)
- Matching known and novel partners for gene fusions (e.g., EML4-ALK vs MLL including coordinate-based limits on where breakpoints must fall)
- Limiting copy number variants based on a range of CNV gain or loss (e.g., ranges to support lower copy amplifications vs. high copy number gains)

This model also enables the inference where one or more variants are required to be present within or across genes (e.g., EGFR and KRAS mutations in lung cancer), and is completely configurable through the addition of new predicates (e.g., to support mRNA or RNA expression-based inferences or biomarker based inferences such as TMB or MSI that may be quantitative [mutations/MB] or enumerated [TMB High/TMB Low vs MSI High/MSI Low/MSI Stable]) via metadata changes to the model.

These unique characteristics as well as content expressed as rules make the PierianDx Knowledgebase much more comprehensive when compared to variant-based databases.

Knowledgebase Performance

To assess the performance of the current PierianDx Knowledgebase, three analyses were performed to demonstrate:

- Breadth of content

pieriandx

- Accuracy of variant classification
- Value beyond FDA labels and practice guidelines

Breadth of Content

First, a retrospective analysis was performed on cases where a gene in the Illumina TruSight Tumor 170 (TST170) panel was part of a panel that was or is being run by a current customer. This analysis allowed for the examination in how many cases variants in the gene may have been classified, interpreted, and reported.

This analysis revealed that approximately a third of genes had been reported in >10,000 patient cases, about half had been reported in >5,000 cases, about two thirds had been reported in >2,500 cases, and all genes had been reported in > 1,200 cases. Thus, clinically relevant variants in any gene in the TST170 panel with an overall incidence of 0.1% would have been observed and reported in at least 1.2-10 cases, indicating the PierianDx Knowledgebase already contains content for virtually every reportable variant.

Accuracy of Variant Classification

Second, the performance in assigning a classification was assessed by determining the classification of the variant after automated assignment by Clinical Genomics Workspace using the PierianDx Knowledgebase and comparing it to that assigned to the same variant after medical director sign-out. A total of 493 TST170 cases across three organizations and 731,496 variants were evaluated. Specificity, sensitivity, positive predictive value (PPV), and accuracy were determined.

Results showed that specificity, sensitivity, positive predictive value (PPV), and accuracy were 99.99%, 93.77%, 98.79%, and 99.95%, respectively. Overall, this analysis showed that only 397/731,496 variants (0.05%) required manual reclassification.

Additionally, results showed outstanding (specificity/PPV) to excellent (sensitivity) performance.

Value Beyond FDA Labels and Practice Guidelines

Third, it was sought to determine if there was value in the PierianDx Knowledgebase beyond that which could be inferred from FDA-approved labels and practice guidelines, i.e. “clinical management yield”. The same TST170 cases as in the previous analysis (493) were grouped as having at least one clinically actionable variant in the patient’s tumor type or another tumor type with the underlying evidence coming from an FDA-approved label, practice guideline, or other source (e.g. published literature). Results showed that while 68.15% of cases had at least one variant associated to an FDA-approved label or practice guideline, 26.98% of cases had a clinically relevant variant that could not be inferred using FDA or guideline content alone.

Overall, greater than 95% of cases had clinical management yield only if all sources were utilized, indicating that the PierianDx Knowledgebase has invaluable content to classify and interpret results from clinical NGS testing and that such reporting should not ignore content outside of FDA approved labels and guidelines.

Conclusion

Increasing test volumes and the resulting increase in genomic knowledge require tools that not only leverage this knowledge but that help efficiently and accurately assign biological significance to improve patient care.

The PierianDx Knowledgebase fosters the sharing of clinical content across medical directors at different sites and drives the creation of informed clinical

reports through use of its comprehensive rules engine for the efficient and accurate classification of genomic variants.

For more information on the PierianDx Knowledgebase, visit www.pieriandx.com or contact us at info@pieriandx.com to request a demonstration or consult.

Works Cited

[1] Kelly, B. (2018). *Clinical NGS Market Projected to Grow by 12% 0ver Next Five Years - BioInformatics Inc.*. [online] BioInformatics Inc. Available at: https://bioinfoinc.com/clinical-ngs-market-projected-grow-12-0ver-next-five-years/ [Accessed 30 Oct. 2018].