VELSERA

The Precision Engine Company

Velsera Clinical Genomics Knowledgebase

Rapidly, accurately identify clinically significant variants and produce an actionable report.

Introduction

As clinical NGS testing volumes grow in light of expanded medical knowledge and new targeted therapy approvals, commercial availability of large comprehensive gene panels, and favorable coverage decisions, the challenge of quality clinical interpretation and variant classification will also grow. To address this challenge, Velsera has developed the Clinical Genomics Workspace, a platform for streamlining complex genomics workflows.

The Velsera Clinical Genomics Knowledgebase is a key component of the Clinical Genomics Workspace. Built using an ideal combination of technology and human expertise, the Knowledgebase enables users to rapidly and accurately classify and interpret variants to produce an actionable report.

This technical note describes the Velsera Knowledgebase in the context of somatic cancer biomarker reporting.

Clinical Reporting Workflow

Following secondary analysis of data from a sequenced sample, which is also supported on Velsera's platform, the Clinical Genomics Workspace classifies variants using a rules engine and automatically creates a draft report that is populated with interpretation content from the Velsera Knowledgebase.

After a draft report is generated, variant scientists and medical professionals follow a software-aided workflow to review report content and its supporting clinical annotations, then perform final signout of the patient report.

Users can revise or add an addendum to the report at a later time after signing out a case. After a report is signed out, the Clinical Genomics Workspace can be configured to automatically route the report in a structured data format to the electronic medical record/electronic health record (EMR/EHR). Users can also route and/or export a PDF version of the report.

This workflow is depicted in Figure 1.

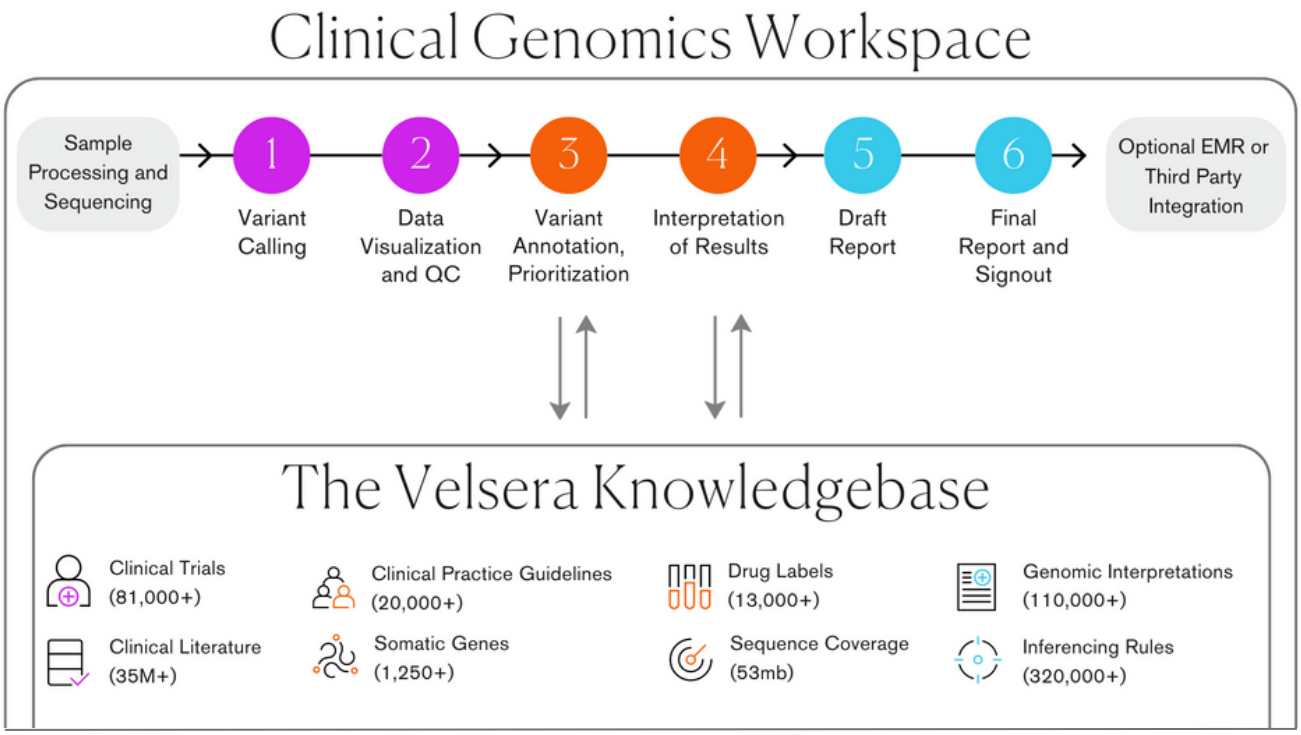

Figure 1:
The Velsera Clinical Genomics Knowledgebase is a key component of the Clinical Genomics Workspace and enables users to rapidly and accurately identify clinically significant variants to produce an actionable report.

Knowledgebase Sources and Curation

The Velsera Knowledgebase includes expertly curated content, real-world medical data, and public data sources, as well as ready-to-use interpretations synthesized from aggregated Knowledgebase information that users can include in their reports.

Expertly Curated Content

The Velsera Biocuration Team curates predictive, prognostic, and diagnostic biomarker associations from the following sources to support classification and interpretation of clinically important variants:

- FDA and EMA drug labels
- Clinical practice guidelines in oncology from NCCN, ASCO, and ESMO
- Published human studies research from PubMed, capturing emerging associations (or investigational therapies for emerging biomarkers)
- Clinical trials from clinicaltrials.gov and the European Clinical Trial Register for physicians and patients seeking treatment options under an actively enrolling clinical trial protocol

Real-World Medical Data

In addition to expertly curated associations, the Velsera Knowledgebase continuously adds shared genomic biomarker interpretations written and signed out by laboratories within the Velsera Clinical Sharing Network or by Velsera's own Interpretation Services Team.

This unique access to real-world knowledge in the Velsera Knowledgebase enables our users to learn from the classification and interpretation decisions of peer medical directors and rapidly issue an informative patient report with confidence.

Public Data Sources

The Velsera Knowledgebase also includes information loaded from the following public data sources:

- Human genome builds
- Gene-RNA-protein models
- Population frequency databases
- COSMIC, TCGA, ClinVar, dbNSFP
- PubMed literature search engine

The Clinical Genomics Workspace leverages these data for accurate annotation of genomic alterations, to automatically identify common polymorphisms, and to facilitate the review of variants of uncertain significance.

Systemized Curation

The Velsera team has established a quality process and assembled qualified team members to build a comprehensive and high-quality knowledgebase.

The Velsera Biocuration Team, which is composed of PhD- and MS-level scientists, uses a systematic approach to translate scientific prose from drug labels and practice guidelines into logical rules that the Knowledgebase can apply to properly annotate variants with their medical meaning. This process takes place weekly following validated SOPs that comply with the ISO 13485 medical device quality management standard and FDA guidance on the maintenance of genetic variant databases.

The newly curated content and logical rules created by the Velsera Biocuration Team undergo several rounds of review and formal verification and validation. The result is a formal Knowledgebase release accompanied by release documentation, including a quality verification report. Knowledgebase release quality is subject to regular review by a Velsera Change Control Board. Velsera also regularly consults with its external Knowledgebase Expert Panel of molecular pathologists for feedback and guidance on its biomarker curation strategies.

Knowledgebase Design

In all, the Velsera Knowledgebase encompasses over 320,000 logical rules across over 53 MB of sequence coverage (53 million genomic positions), covering 1,250 somatic cancer genes and supporting all genomic variant types: single nucleotide variants, insertions and deletions, copy number variants, fusions, large genomic structural variants, and splice isoform variants, as well as biomarkers including tumor mutational burden (TMB), microsatellite instability (MSI), and measures of homologous recombination deficiency (HRD).

The design of the Velsera Knowledgebase distinguishes it from the majority of commercially available genetic variant knowledgebases in that it uses a metadata model, rules engine, and robust logic to match content to variants and create ready-to-use interpretations.

Metadata Model

Due to the rapidly evolving practice of genomic medicine, the Velsera Knowledgebase has been architected to incorporate new data sets and technologies in a flexible manner. All sources incorporated—including public genetic databases, practice guidelines, drug labels, and shared interpretation content—vary in their update frequency and composition over time. Velsera uses configurable metadata to describe the model of these datasets, which enables the Velsera Knowledgebase and other solutions to rapidly adopt future changes in assay technologies and the evolving nature of other biomarkers that can be detected by NGS and non-NGS test modalities.

Rules Engine

Velsera has a distinct approach to matching Knowledgebase content with variants 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 been explicitly curated based on past description in the clinical genomics field, then the resulting report will not include any clinical information on that variant. In contrast, the Velsera Knowledgebase uses a comprehensive model that allows for the creation of complex rules on variants within or across genes—not only using syntax, but also based on particular variant characteristics such as falling within particular exons or domains, fusing with known or novel partners, or having particular copy number changes. This means that even if the specific patient variant has never been previously described, the Velsera Knowledgebase will look across all other variations, apply logic inferred from similar gene-variant-disease combinations, and return rationalized content in the report.

For example, the Velsera Biocuration Team can write rules qualitatively describing in-frame deletions in exon 19 of EGFR or exon 11 of KIT, or deleterious alterations in BRCA1 or BRCA2. Rules can also describe co-occurrence of multiple biomarkers, such that a combination of biomarkers found in a particular patient's case could have different clinical associations and a distinct interpretation than any single one of those variants found alone.

Powerful Logic for Content Matching

The Velsera Biocuration Team uses the flexible metadata model to describe a broad range of conditions for matching of content, including:

- Human Genome Variation Society (HGVS) syntax (HGVS p., c., or g. syntax)
- Genomic, coding, or protein coordinates (e.g., codon ranges, coding sequence ranges, or genomic sequence ranges)
- Functional characteristics (e.g., frameshift, in-frame, truncating, missense, splice site disruption)
- Variant type (e.g., SNV, MNV, deletion, insertion, and/or indel)
- Known and novel partners for gene fusions (e.g., EML4 fused to ALK, or MLL fused to any partner; can include conditions for directionality of the fusion and coordinate-based boundaries for breakpoint locations)
- Splice isoform variants detected from RNA sequencing
- Copy number variants based on the magnitude of CNV gain or loss (e.g., ranges to support lower copy amplifications vs. high copy number gains)

Examples of how the flexible data model is used to apply associations that are faithful to sources from which they are derived are provided in Table 1.

Implementing Evidence-Based Biomarker Curation Strategies

The Velsera Biocuration Team develops version-controlled strategies for the curation of complex biomarkers. These strategies describe how biomarkers referenced using common parlance in curated documents, including drug labels and clinical trials, will be converted into specific conditions in the Knowledgebase based on scientific and medical evidence and consensus in the field.

The Biocuration Team maintains a controlled library of biomarkers that represent the latest curation strategies. Each biomarker in the library comprises one or more rules that define the specific conditions under which detected alterations should be considered as belonging to that biomarker. Velsera Biocurators capturing genomic associations from source documents then leverage the biomarker library to ensure consistent representation of biomarkers across the Knowledgebase.

Table 1: How the Velsera Knowledgebase matches content based on biomarker conditions.

Biomarker Condition Scenario	"If" Conditions	"Then" Assertion	Example Patient Variant
HGVS Syntax HGVS p., c., or g. nomenclature in a gene is the most specific condition used in the Knowledgebase. BRAF p.V600E in cutaneous melanoma is a common example with multiple therapy associations. In the example at right, a coding DNA syntax in the TERT promoter mutation has diagnostic value.	**Tumor Type** CNS malignancy **Biomarker** TERT c.-124C>T	**Association** Diagnostic value in glioblastoma **Source** NCCN	**TERT** **c.-124C>T**
Codon Rules When diverse changes at an amino acid position or within a codon range are understood to share the same clinical meaning, specific types of alterations in that range are encoded in the Knowledgebase. In this example, the evidence indicates that any amino acid substitution at codon 132 in IDH1 is associated with benefit from ivosidenib in AML.	**Tumor Type** Acute myeloid leukemia **Biomarker** Missense IDH1 codon R132 mutations	**Association** May benefit from ivosidenib **Source** NCCN, ESMO	**IDH1** **p.R132C**
Exon/Domain Rules Insertions and/or deletions with specific characteristics in some protein functional domains can lead to targetable changes in protein function. In this example, diverse in-frame insertions across an exon of EGFR, with one demonstrated exception that is handled by the Knowledgebase, share sensitivity to amivantamab-vmjw in NSCLC.	**Tumor Type** NSCLC **Biomarker** In-frame EGFR exon 20 insertions, other than p.A763_Y764insFQEA	**Association** May benefit from amivantamab-vmjw **Source** FDA, EMA, NCCN	**EGFR** **p.A767_V769dup**
Protein Consequence Rules The Knowledgebase can describe changes across all or much of a gene with a specific protein consequence. In this example, truncating BRCA1 alterations—those that lead to a frameshift and/or early termination—prior to codon 1854 represent a subset of deleterious BRCA1 changes associated with an HRD phenotype and responsiveness to PARP inhibitors.	**Tumor Type** Ovarian cancer **Biomarker** Truncating BRCA1 alteration prior to codon 1854	**Association** May benefit from bevacizumab + olaparib **Source** FDA, EMA, NCCN, ASCO	**BRCA1** **p.K654fs*47**
Structural Variant Rules Copy number alterations, genomic rearrangements, and RNA detected fusions and splice variants are matched. Rearrangement/fusion conditions may or may not specify a mate gene or directionality of the event. In this example, an ETV6-RUNX1 fusion with ETV6 as the 5' partner confers favorable prognosis in B cell ALL.	**Tumor Type** B cell acute lymphoid leukemia **Biomarker** ETV6-RUNX1 fusion/ rearrangement	**Association** Favorable prognosis **Source** NCCN, ESMO	**ETV6-RUNX1** **fusion**
Co-occurring Rules When a combination of alterations confers a different medical meaning than when the alterations occur independently, a set of conditions are expressed together in the Knowledgebase. In this example, an inhibitor-resistant NTRK1 mutation in combination with an NTRK1 fusion negates the benefit of larotrectinib.	**Tumor Type** Any solid tumor **Biomarker** Co-occurring NTRK1 fusion and NTRK1 p.G595R	**Association** Unlikely to benefit from larotrectinib **Source** FDA, EMA	**TPM3-NTRK1** **fusion + NTRK1** **p.G595R**
Trial Exclusion Criteria Clinical trials may set enrollment conditions that require a biomarker of interest but exclude tumors that harbor other alterations. In this example, one of a set of KRAS mutations of interest serves as an inclusion criterion, while detection of any EGFR sensitizing mutation in the same tumor is an exclusion criterion.	**Tumor Type** NSCLC **Biomarker** KRAS mutation; EGFR sensitizing co-mutations excluded	**Association** Potential eligibility for Phase I/II trial NCT04263090 (clinicaltrials.gov)	**KRAS** **p.Q61H**

Variant Classification

The Clinical Genomics Workspace classifies variants according to established guidance, including AMP/ASCO/CAP joint recommendations and ComPerMed guidelines, and populates a draft report with relevant Velsera Knowledgebase content based on matching by the rules engine. Classification schemes like these serve as valuable systems for assessing potential clinical relevance and for prioritizing variants for reporting, particularly in the context of large comprehensive genomic profiling assays.

AMP/ASCO/CAP Tiering

The AMP/ASCO/CAP joint recommendations[1] place variant biomarkers in a tier based on clinical actionability. The Velsera Knowledgebase leverages curated, real-world, and public database evidence to automate assignment of AMP/ASCO/CAP tier, as described here:

- *Variants of Strong Clinical Significance, Tier IA:*
  - Drug label associations predictive of likely benefit (or lack of benefit) of a therapy, or practice guideline evidence of predictive, prognostic, or diagnostic associations for the patient's tumor type
  - Medical director interpretations from past cases for the patient's tumor type that assigned Tier IA
- *Variants of Strong Clinical Significance, Tier IB:*
  - Emerging biomarkers or established biomarkers with evidence of emerging predictive, prognostic, or diagnostic associations from well-powered published studies with expert consensus for the patient's tumor type
  - Medical director interpretations from past cases for the patient's tumor type that assigned Tier IB
- *Variants of Potential Clinical Significance, Tier IIC:*
  - Drug label and practice guideline associations for other tumor types
  - Medical director interpretations from past cases that assigned Tier IA for other tumor types or Tier IIC in the patient's tumor type
  - Emerging biomarkers or established biomarkers with evidence of emerging predictive, prognostic, or diagnostic associations from smaller published studies with some consensus for the patient's tumor type
  - Enrolling clinical trial matches for the patient's tumor type
- *Variants of Potential Clinical Significance, Tier IID:*
  - Medical director interpretations from past cases for the patient's tumor type that assigned Tier IID for the patient's tumor type
- *Variants of Unknown Clinical Significance, Tier III:*
  - Medical director interpretations from past cases that assigned Tier III
  - Variants with no clinical evidence that are not known polymorphisms or consensus benign
- *Benign or Likely Benign Variants, Tier IV*:
  - Known polymorphisms

ComPerMed Classification

The Commission for Personalized Medicine (CompPerMed) expert panel devised a system to harmonize the biological classification of somatic variants in cancer across various institutions[2]. This classification scheme primarily focuses on the biological contribution of each variant and its potential to promote tumorigenesis independent of tumor type.

The scheme classifies variants into one of the following categories: Pathogenic, Likely Pathogenic, Variant of Uncertain Significance (VUS), Likely Benign, and Benign. These are assigned automatically by the Clinical Genomics Workspace based on the following criteria:

1. Based on gnomAD frequency, high frequency/common variants are assigned Likely Benign or Benign
2. Consensus pathogenic variants maintained by the ComPerMed expert panel are assigned Pathogenic
3. Clear loss-of-function mutations (e.g., frameshift, truncation, or canonical splice donor or acceptor variants) in tumor suppressor genes are assigned Likely Pathogenic; clear loss-of-function mutations in oncogenes are assigned VUS
4. Non loss-of-function mutations are evaluated via a scoring system that considers the number of COSMIC entries for a variant, genomic database information such as ClinVar, and in silico prediction tools, with classification assigned accordingly

5. Several exceptions are applied to the above logic, as defined by the ComPerMed panel, with specific conditions determined based on biological and clinical evidence captured in our biomarker library or from other sources, including for TP53, BRCA1, BRCA2, MET, CALR, NPM1, and CEBPA

Knowledgebase Performance

To assess the performance of the Velsera Knowledgebase, analyses were conducted to demonstrate accuracy of 1) assignment of clinical relevance and 2) identification of clinical associations relative to orthogonal opinions.

Accuracy of Assignment of Clinical Relevance

As per AMP/ASCO/CAP joint recommendations, clinically significant variants are classified as Tier I or Tier II and variants that lack clinical significance are classified as Tier III or IV. We assessed the performance of the Velsera Knowledgebase for the assignment of clinical relevance to genomic variants by comparing its automated calls to the consensus classification assigned by medical directors in the Clinical Genomics Workspace. Consensus classification was determined from signed out reports by assessing agreement on medical director assigned tier for a variant-disease combination within and across laboratories for a specified timeframe.

Over 8,500 cases across 31 organizations and almost 5.8 million variant-disease combinations were evaluated to assess specificity, sensitivity, and accuracy. Out of 52,410 variant-disease combinations that qualified as having a consensus classification, the Velsera Knowledgebase agreed with laboratory professional assignment of clinical relevance in signed out clinical cases for 52,379 alterations (99.9% concordance). Thirty-one variant-disease combinations were discrepant. These were justified as correctly assigned by the Knowledgebase based on time effects of emerging/evolving biomarkers, on medical director reliance on low-level or outdated PubMed evidence, or on medical director overclassification of variants of high minor allele frequency in the population.

Furthermore, our findings show that the Velsera Knowledgebase accurately classified over 100,000 benign or likely benign variants from ClinVar (99.9% concordance) unless there was evidence to classify them as clinically significant. Only 60 ClinVar variants were categorized differently due to weak evidence from a single submitter. Some of these variants were still deemed clinically significant in multiple studies or practice guidelines.

Sensitivity and Accuracy of Reporting Clinical Associations

The Velsera Knowledgebase performed favorably in comparison to an independent knowledgebase, Clinical Interpretation of Variants in Cancer (CIViC), in a comparison that leveraged the Clinical Genomics Workspace as a real-world variant database. Velsera ascribed clinical relevance to 78% of CIViC's predictive, prognostic, or diagnostic assertions. The remaining 22% were either based on older evidence items without sufficient consensus, lacked support from recent clinical research publications, or were represented in CIViC as single biomarker associations while the KB association was with the co-occurrence of two biomarkers.

The Velsera Knowledgebase also exhibits high accuracy in clinical trial matching. Our independent team of specialists queried clinical trials on clinicaltrials.gov and the European Clinical Trials Register (EUCTR) for over 50 disease-biomarker combinations. They created a list of expected trial matches, and we compared it to the matches found by the Velsera Knowledgebase. We found a 99.97% accuracy rate for clinicaltrials.gov and 99.98% for EUCTR.

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 Velsera 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 Velsera Knowledgebase, visit www.velsera.com or contact us at info@Pieriandx.com to request a demonstration or consult.

Works Cited

1. Li MM et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists J Mol Diagn 2017 Jan;19(1):4-23
2. Froyen G et al. Standardization of Somatic Variant Classifications in Solid and Haematological Tumours by a Two-Level Approach of Biological and Clinical Classes: An Initiative of the Belgian ComPerMed Expert Panel. Cancers (Basel) 2019 Dec;11(12):2030

Schrafft's City Center,
529 Main St, Suite 6610,
Charlestown, MA 02129