Genomics Annotation and Interpretation in Somatic Oncology with a Highly Structured Data Model (Poster# I026)

Jian Li, Lili Niu, Shuba Krishna, Fnu Kinshuk, Martin Jones, Carlos Hernandez, Michael Clark, Tara Love, Amin Gholami

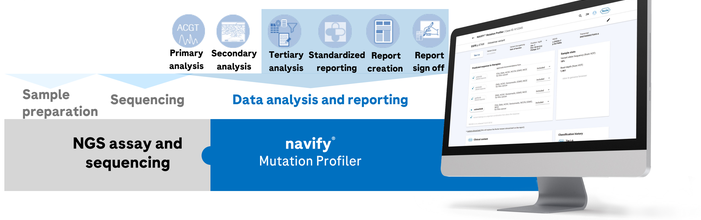
Roche Diagnostics Solutions, Santa Clara, CA

https://ter.li/lzb90a

Introduction

navify® Mutation Profiler is a Roche software product that provides an automated cloud-based solution to streamline the process of annotating, interpreting and reporting variants from next-generation sequencing (NGS) somatic oncology test results. The mutation profiler brings together multiple sources of information, including curated content, public databases, and laboratories internal data. The software supports a variety of NGS assays, and provides an user interface allowing users to reclassify biomarkers or customize final report. **navify** Mutation Profiler is for Research Use Only in the United States of America and other parts of the world. It is for *in vitro* diagnostic use in the European Union and other parts of the world.

Ephesus is an internally developed web application enabling expert curation for construction of a robust knowledgebase of NGS biomarkers in somatic cancer to support the mutation profiler. The Ephesus data model ensures adherence to best practices in evidence-based curation of genomic content. Variant classification follows the AMP/ASCO/CAP "Standards and Guidelines for Interpretation and Reporting of Sequence Variants in Cancer" [1] for somatic variant interpretation. Variant interpretations are derived from international regulatory approvals and professional practice guidelines (e.g. FDA, EMA, TGA, eVIQ) and recommendations (e.g. NCCN, ESMO).



Methods

The design of the data model for Ephesus starts from understanding key components for generating genomic testing reports in somatic cancer (Figure 1). The process of curation involved collecting evidence from drug approvals and guidelines, scientific literature, and recent meeting abstracts. These pieces of evidence are prioritized and summarized for each biomarker (at RNA or protein level) in the context of diseases based on actionability. Classifications for each biomarker and disease are assigned based on AMP/ASCO/CAP guidelines. Evidence items are collected and synthesized for genes, biomarkers, and biomarker profiles, which are structured as the main entities in the content database data model.

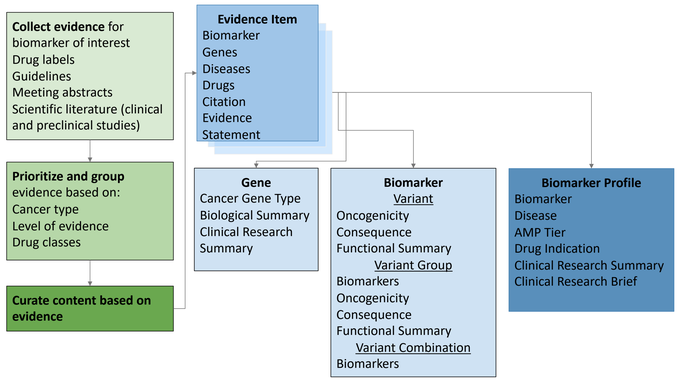


Figure 1. The data flow in Ephesus is based on the manual curation process

Methods (cont'd)

The most challenging aspect in architecting automated variant interpretation pipelines is often the integration of diverse sets of data [2]. Genomic profile and the biomedical evidence are highly heterogeneous. Therefore, how to efficiently decide that the genotype from a genomic profile matches the right collection of evidence and is classified appropriately in the context of the disease becomes the key resolution in such a pipeline. The ecosystem of our workflow adopts a hierarchical multi-facet data structure combined with the implementation of rule sets, to serve as the annotation and interpretation model.

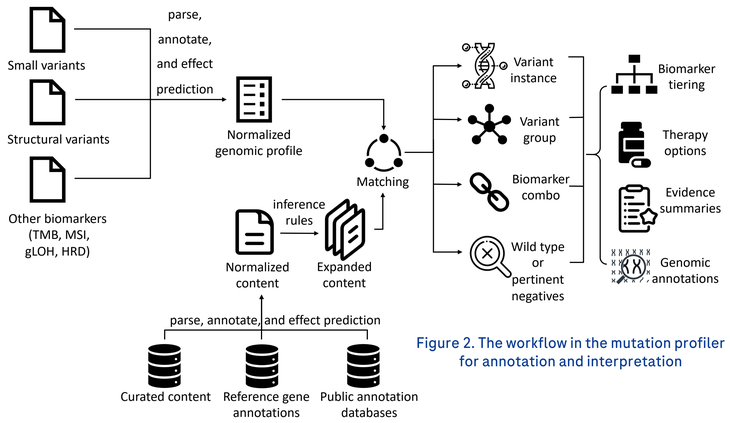


Figure 2. The workflow in the mutation profiler for annotation and interpretation

Results

The content drop for the mutation profiler from the Oct. 2024 snapshot contains 1,271 genes and 12,859 directly curated biomarker profiles, involving 34,058 small variants, 5,326 variant groups, 2,391 structural variants and other biomarkers such as TMB, MSI, etc., as well as biomarker combinations. After expansion with inference rules there are 7.9M+ profiles for 40+ major cancer types on top of their subtypes, covering NGS biomarker based drug approvals and professional guidelines in 14 countries and regions.

We assessed the reporting value of Ephesus/**navify** Mutation Profiler by querying somatic small variants represented across ~170K real cancer cases from the AACR GENIE® (v16.1) project[3].

AACR-Genie_v16.1 (169,686 cases)	CGI 2022-10-17	CIVIC 2024-10-01	ClinVar 2024-10-01	Roche 2024-10-14
#cases with any interpretation	47,378 (28%)	86,649 (51%)	155,490 (92%)	167,665 (**99%**)

Table 1. Comparison of number of cases in the GENIE cohort with any interpretation across the knowledgebases

AACR-Genie_v16.1 (1,105,422 biomarkers)	CGI 2022-10-17	CIVIC 2024-10-01	ClinVar 2024-10-01	Roche 2024-10-14
#biomarkers with any interpretation	264 (0.02%)	551 (0.05%)	160,695 (14.54%)	828,807 (**74.98%**)
#biomarker + disease	120	224	NA	1,327,551 (expanded)

Table 2. Comparison of number of biomarkers in the GENIE cohort with any interpretation across the knowledgebases

Results (cont'd)

Compared against 3 other major knowledgebases, the performance of our solution in terms of matching cancer cases and interpretations of potential clinical research significance exceeds all these major knowledgebases. In addition, the results show that the incorporation of variant groups, inference rules, and variant combinations dramatically increases interpretability.

Similarly, the structural variants(SVs) annotation data model and the corresponding pipelines in our solution are designed to capture the common patterns, fully resolve complex SV events, and improve the accuracy of SV interpretation. Here we use ALK fusions to showcase the implemented data model for SVs not only allows curating/matching such events at different levels of resolutions, but also allows covering novel fusion partners or breakpoints which essentially lead to the same functional results.

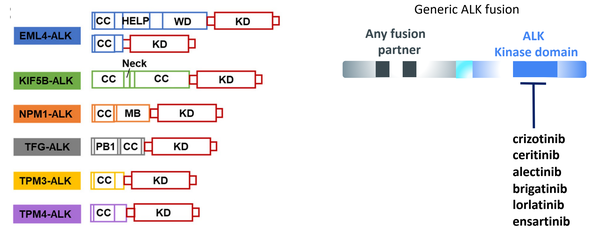


Figure 3. A: Domain schematic structural composition of frequently observed ALK fusion proteins [4]; B: a generic ALK fusion model adopted by the mutation profiler

COSMIC fusions (v100)	CIVIC 2024-10-01	Roche 2024-10-14
#Unique EML4 - ALK with interpretation	15 (44.12%)	34 (100%)
#Samples with EML4-ALK with interpretation	2600 (22.99%)	11309 (100%)
#Unique ALK fusions with interpretation	36 (37.89%)	95 (100%)
#Samples with ALK fusions with interpretation	4283 (25.42%)	16847 (100%)

Table 3. Number of unique breakpoint-pairs or unique samples for example fusions (EML4-ALK and all ALK 3' fusions) in COSMIC that can be correctly matched (Note: ClinVar and CGI are not included due to the lack of fusion data or clear breakpoint annotations)

Conclusion

The **navify** Mutation Profiler annotation and interpretation model allows robust, up-to-date and scalable processing of somatic case genomic profiles. At the same time, it saves significant time and resources for internal content curation, is extendable to new annotations or content data sources, and is applicable to NGS profiles agnostic to secondary pipelines or assay types.

References

1. Li MM, Datto M, Duncavage EJ, 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;19(1):4-23. doi:10.1016/j.jmoldx.2016.10.002
2. Cambrosio, A., Campbell, J., Vignola-Gagné, E., Keating, P., Jordan, B.R., Bourret, P. (2020). 'Overcoming the Bottleneck': Knowledge Architectures for Genomic Data Interpretation in Oncology. In Data Journeys in the Sciences. Springer International Publishing AG. 2020. p. 305-327 doi: 10.1007/978-3-030-37177-7_16
3. The AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine Through An International Consortium, Cancer Discov. 2017 Aug;7(8):818-83
4. Kong Y, Jiang C, Wei G, Sun K, Wang R, Qiu T. Small Molecule Inhibitors as Therapeutic Agents Targeting Oncogenic Fusion Proteins: Current Status and Clinical. Molecules. 2023;28(12):4672. doi:10.3390/molecules28124672

Conflicts of interest

Authors are employees and/or stock holders of Roche Diagnostics. This study was sponsored by F. Hoffmann-La Roche Ltd.

jian.li.jl7@roche.com
925-918-5392