Prioritizing Germline and Somatic Variants for Precision Oncology

Scholarly Report submitted in partial fulfillment of the MD Degree at Harvard Medical School

Date: 6 March 2020

Student Name: Nicholas S. Moore, AB

Title: Prioritizing germline and somatic variants for precision oncology

Mentor:
Eliezer M. Van Allen, MD
Dept of Medical Oncology, Dana-Farber Cancer Institute

Collaborators:
Saud AlDubayan, MD
Dept of Medical Oncology, Dana-Farber Cancer Institute

Jennifer R. Brown, MD, PhD
Dept of Medical Oncology, Dana-Farber Cancer Institute

Catherine J. Wu, MD
Dept of Medical Oncology, Dana-Farber Cancer Institute

Gad Getz, PhD
Broad Institute

Publications:
Nassar AH*, Alaiwi SA*, AlDubayan SH*, **Moore NS**, Mouw KW, Kwiatkowski DJ, Choueiri TK, Curran C, Berchuck J, Harshman LC, Nuzzo PV, Chanza NM, Van Allen E, Esplin ED, Yang S, Callis T, Garber JE, Rana HQ, Sonpavde G. Prevalence of Pathogenic Germline Cancer Risk Variants in Urothelial Carcinoma. *Genet Med*. 2019 Dec 17.

Reardon B, **Moore NS**, Moore ND, Kofman E, Van Allen E. A molecular oncology almanac for integrative clinical interpretation of molecular profiles to guide precision cancer medicine. *Cancer Res*. 2019 July 1; 79(13 supplement):2470-2470.

Moore NS, AlDubayan SH, Taylor-Weiner A, Stilgenbauer S, Getz G., Wu CJ, Van Allen EM, Brown JR. Inherited DNA repair and cell cycle gene defects in chronic lymphocytic leukemia. *Journal of Clinical Oncology*. 2019 May 20; 37(15_suppl):1508.

Moore NS, Reardon B, Van Allen EM. Molecular Oncology Connector. Chrome Web Store. 2019 July 25.

Abstract

Molecular sequencing has become an integral part of diagnostic and treatment algorithms for many types of cancer, yet interpreting molecular sequencing in light of evolving knowledge and relevant data types remains a significant challenge. Interpretation and discovery is particularly challenging for germline sequencing routinely collected from normal tissue of cancer patients. Here I discuss enhancements to a paired precision oncology knowledgebase and interpretation algorithm, the Molecular Oncology Almanac, to add assertions relevant to medical and radiation oncology as well as improve accessibility of crowdsourcing through development of a public API and Chrome extension. To further characterize the contribution of pathogenic germline variants to the missing heritability of chronic lymphocytic leukemia, we performed a case-control study using WES data from 962 cases. Preliminary results demonstrate enrichment of pathogenic variants in the DNA damage repair gene, *CHEK2*. Significant enrichment in loss-of-function variants persists after removing known low-penetrance and founder variants, suggesting robustness to ancestry correction. Final analysis with enhanced QC, relatedness, and ancestry matching methods in expanded cohorts is pending. These results demonstrate potential for lowering the barrier to incorporate molecular sequencing data into precision oncology research and clinical practice for both medical and radiation oncology, in particular germline sequencing of normal tissue.

Contents

Table of Contents

Glossary

API: Application programming interface

CLL: Chronic lymphocytic leukemia

DDR: DNA-damage repair

DV: DeepVariant

LOF: Loss of function

LP: Low penetrance

MOA: Molecular Oncology Almanac

PARP: Poly(adenosine diphosphate-ribose)

RNA-seq: Whole-transcriptome sequencing

WES: Whole Exome Sequencing

WGS: Whole Genome Sequencing

Introduction

Background

Molecular sequencing has become an integral part of diagnostic and treatment algorithms for many types of cancer. Yet continued translation of cancer genomics into the clinic poses scientific, informatic, and practical challenges. As the number and types of known molecular alterations important for clinical decision-making have increased, interpretation of sequencing data has become increasingly complex. For cancer patients in whom germline sequencing data is collected alongside somatic tumor DNA, interpreting germline variants is particularly difficult owing to the challenge of classifying variants as pathogenic or benign, and an evolving understanding of which variants are relevant to cancer risk, prognosis, and treatment.

Computationally-enabled knowledgebases to aid in interpretation of cancer sequencing data have been developed in both industry and academic settings. Such knowledgebases have the potential to complement research and clinical interpretation efforts, for instance those of molecular tumor boards at multiple institutions (1), while also extending the reach of genomic interpretation to non-specialist settings. However, such databases vary in terms of academic rigor, thoroughness across the clinical-preclinical spectrum, curation process, and open access (Table 1). Furthermore, extant knowledgebases do not interface directly with software to interpret ‘omic readouts, foregoing an opportunity to use algorithms for rapid and accurate sequence interpretation. Finally, extant knowledgebases largely neglect molecular alterations relevant for clinical decision-making outside of medical oncology, for instance in radiation oncology, despite growing clinical and preclinical of molecular alterations that impact response and resistance to radiation (2).

Such shortcomings in extant knowledgebases for precision oncology are particularly concerning in light of emerging relevance for multiple types of sequencing data, which are poised to offer new insights into cancer care while further complicating interpretation efforts. Modalities such as whole genome sequencing (WGS), which to date have largely been used in a research context (3), introduce additional layers of nuance to interpretation of noncoding regions. While

the value of WGS and RNA-seq continues to be demonstrated in clinical studies, for instance by informing trial stratification and clinical decision-making around triple-negative breast cancer (4), guidelines for using these data continue to evolve (5).

More imminently, interpretation of germline sequencing data is growing in clinical importance for decision-making, not only in terms of cancer risk but also in terms of choice of therapy (6). Recent FDA approval for targeted agents specific to altered germline DNA, such as the poly(adenosine diphosphate-ribose) (PARP) inhibitor olaparib for pathogenic germline *BRCA1* or *BRCA2* mutations (7), signal that germline alterations may play an increasing role in choice of therapy. More broadly, genes involved in DNA damage repair pathways (DDR) (including *BRCA1/2)* are common drivers across a range of cancers and serve as attractive targets for therapies intended to cause synthetic lethality in cancer cells (8). Germline alterations in DDR genes may associate with cancer risk and severity across multiple cancer types, occurring for instance in 11.8% of men with metastatic prostate cancer as compared to 4.6% of men with localized prostate cancer (9). Specific DDR genes for which pathogenic germline alterations have been implicated in cancer include *CHEK2*, which associates with risk for testicular germ cell tumors (10), prostate cancer (11), colorectal cancer (12), and breast cancer (13); *ATM*, for which rare germline variants have been associated with chronic lymphocytic leukemia (CLL) (14) and colorectal cancer (15); and *PALB2*, which also associates with risk of colorectal cancer (15). Mechanistically, these genes are directly linked: *ATM* encodes a DNA damage-sensing protein (ATM) that phosphorylates the protein product of *CHEK2* (CHEK2), a kinase that in turn regulates over 20 downstream proteins (including BRCA1, BRCA2, p53, and Rb) involved in DNA double-strand break repair, regulation of the cell cycle, and apoptosis (16,17).

Furthermore, apart from their direct relevance to clinical care, studies of germline DNA have the potential to explain missing heritability in multiple types of cancer. While twin studies have reflected high heritability across cancer types (18), for many cancers the genetic source of this heritability remains unknown. In chronic lymphocytic leukemia (CLL) for instance, although first-degree relatives of CLL patients have a 7.5-fold increased risk and heritability is known to be

~60% (19), genome-wide association studies (GWAS) have only explained 15% of CLL risk across dozens of loci with small effects (20). Limited evidence has implicated individual genes in CLL risk: while studies of familial CLL have suggested roles for shelterin complex genes and possibly *ITGB2* (21–24), only *ATM* has been implicated in the sporadic setting (14,25,26). While large collections of paired tumor-normal CLL samples have been assembled for studies of somatic mutations (14,27,28), these have not been fully leveraged to study germline risk for CLL.

Compared to somatic sequencing data, interpreting germline alterations poses unique challenges owing to genetic diversity across populations and frequent ambiguity around classifying variants as pathogenic or benign. Furthermore, pathogenic germline mutations relevant to cancer risk, prognosis, and choice of therapy vary widely in terms of penetrance and frequency. Beyond the challenges of classifying variants and estimating their effects, computational tools to call variants from sequencing data also vary in terms of accuracy (29), which may have decreased the power of previous studies to identify variants.

Here I describe efforts to improve interpretation of sequencing data with the aim to advance precision medical and radiation oncology by 1) expanding the content and accessibility of a precision oncology knowledgebase and interpretation algorithm, the Molecular Oncology Almanac (MOA) and 2) identifying DDR genes enriched for germline pathogenic variants using a novel deep-learning based variant algorithm, DeepVariant (30), in a case-control study of CLL patients.

Student role

Working closely in the Van Allen lab with Brendan Reardon, who has been developing the MOA over the past several years, I updated the MOA precision oncology knowledgebase. I independently reviewed and added the first radiation oncology assertions to the knowledgebase. I also developed and published a web browser extension in the Chrome Web Store in order to enable submission of precision oncology content from the biomedical community, based on an API I added to the MOA website. I began work to enhance interpretation and reporting of germline genetic variants by MOA.

Working closely with a team of computational biologists in the Van Allen lab, including Saud AlDubayan, Eric Kofman, Seunghun Han, Sabrina Camp, and Amaro Taylor-Weiner, I developed and tested cloud-based germline variant calling and analysis workflows based on DeepVariant. Using these workflows, alongside Saud AlDubayan, Eliezer Van Allen, and Jennifer Brown I designed a case-control study to detect DDR genes enriched for pathogenic germline variants in CLL patients. Jennifer Brown, Catherine Wu, and Gad Getz provided the data needed as well as domain knowledge. I curated the sample data and reviewed its quality with Aditi Gupta. I carried out the study with help from the individuals named in this paragraph.

Methods

Molecular Oncology Almanac

Knowledgebase enhancement

Molecular alterations predictive of sensitivity to therapy, resistance to therapy, or prognosis were collected from longitudinally reviewing the biomedical literature, FDA approvals, and NCCN guidelines; these were added to the alterations already documented in MOA. Clinical evidence in the MOA knowledgebase was harmonized with entries on OncoKB (31), with additional assertions gathered from the literature. A review of radiation oncology literature was performed for clinical and preclinical molecular features relevant to sensitivity or resistance to radiation. These were added to MOA in the same format as medical oncology assertions.

Chrome extension

The Flask web app, moalmanac.org, was updated with RESTful API endpoints for getting and posting alteration-action relationships to the MOA. Endpoints were designed within the OpenAPI 3.0 standard and documented in SwaggerHub. Using these endpoints, a Google Chrome browser extension for posting alteration-action relationships was developed. Every submitted assertion includes a citation, class of alteration, cancer type, alteration effect, and level of evidence; gene, alteration, and therapy may also be included where relevant. Testing of endpoints and the extension was performed by multiple users in the Van Allen Lab. Submitted assertions were reviewed by site administrators prior to final incorporation into the database.

Chronic Lymphocytic Leukemia

Cohorts

WES germline sequencing data from CLL cases were assembled from previous studies published on dbGaP (Table 2) as well as from collaborators. Samples were excluded if they had known tumor contamination, were sourced from peripheral blood with unknown tumor-in-normal contamination or >20% contamination per DeTiN, or mean coverage <10x per DepthofCoverage. Samples were combined to create at first 2 cohorts in the preliminary analysis (514 and 285 samples), followed by 3 independent cohorts in the final analysis as more

data became available (314, 275, and 339 samples) (Figure 8). Cases were not selected on the basis of any known clinical variables. For control cohorts, germline WES data for cancer-free samples was collected from multiple cohorts on dbGaP. All case and control samples were passed through identical workflows implemented on FireCloud/Terra: alignment to b37 reference genome (BWA 0.5.9) if not already, quality control, and calling with DeepVariant. Cohorts were merged using GATK Combine Variants. Cohort-level quality control was performed by examining distributions of InDel length and ratio of transitions to transversions. Related samples were excluded from cohorts using PC-Relate (32) and pruning one sample from every pair with relatedness >0.125. Continental ancestry was assigned to cases and control samples using principal-component analysis (PCA) and uniform manifold approximation and projection (UMAP) to classify samples using Random Forest with labeled continental ancestry for 1000 Genomes samples as reference. Initial analysis was performed with case and control cohorts known to share similar continental ancestry. In the final analysis, cases and controls will be matched according to continental ancestry, with a second PCA to select nearby controls for each case. All variants were annotated with Variant Effect Predictor (VEP).

Pathogenic Germline Variant Enrichment Analysis

For each case and control cohort, all variants called from a published set of DNA damage repair genes (15) were extracted. All variants known to have >1% frequency in the non-Finnish European population per gnomAD (33) were excluded. ClinVar annotations by VEP were used to filter down to variants listed as “pathogenic” or “likely pathogenic”; variants with conflicting ClinVar annotations were not included. Other variants were included if they were predicted to have a high pathogenic impact: frameshift, nonsense, splicing, or start loss variants. Variants were grouped by gene for gene burden testing on cases vs controls using two-sided Fisher’s exact test, with an false discovery rate significance threshold of $q<0.1$. All pathogenic variants in significantly enriched genes were manually reviewed in cases and controls using Integrated Genomics Viewer (IGV).

Results

Enhanced Molecular Oncology Almanac knowledgebase and improved accessibility for crowd-sourcing

Review of the biomedical literature and harmonization with OncoKB resulted in addition of multiple new assertions to the pre-existing MOA knowledgebase, bringing its total size to 543 action-alteration relationships across therapeutic modalities and levels of evidence, ranging from preclinical to FDA-approved (Fig. 6). Reviewing the radiation oncology literature resulted in addition of 18 new assertions, compared to 3 such assertions currently in CiVIC (34) and 0 in OncoKB (31). Initial work was done on the MOA interpretation algorithm to improve reporting on germline alterations and classify *BRCA* variants of unknown significance according to functional status determined in cell line models (35).

To open the MOA knowledgebase for use by the research community, in addition to the extant web portal www.moalmanac.org we added API endpoints enabling the public to query assertions by gene, therapy, predictive implication, source, or feature. Given that other crowd-sourced knowledgebases require a high degree of time-consuming detail in user submissions, we sought to lower the barrier for gathering assertions by allowing submission directly from the web browser. Based on the endpoints, to assist in rapid, expert crowd-sourcing of new precision oncology assertions from the biomedical literature we developed a web browser extension, the Molecular Oncology Almanac Connector, and published it in the Google Chrome web store (Fig. 7). To date the Connector has been downloaded 11 times and has been used to submit assertions to the database, which have been subsequently reviewed for accuracy and incorporated into the knowledgebase.

Pathogenic germline variants in *CHEK2* are associated with increased risk for CLL: preliminary analysis

Of 962 germline WES samples from CLL cases, 929 passed quality control requirements (Fig. 8). One sample was excluded due to aberrant variant calling statistics (Fig. 9). Samples were ultimately divided into 3 independent cohorts that showed reassuring QC metrics (Fig. 10). In a

preliminary analysis of a discovery cohort corresponding to 569 non-ICGC samples, pathogenic germline variants were found in many of the DDR genes studied (Fig. 2). In particular, *CHEK2* LOF variants were found in 15 (2.6%) of cases. Compared to a control cohort of 8614 WES samples where *CHEK2* LOF variants were found in 61 cases (0.7%), enrichment was significant after adjusting for multiple hypothesis correction (Fig. 3). To ensure that *CHEK2* enrichment was not driven by the common low-penetrance (LP) allele c.599T>C or the known founder mutation 1100delC, which might reflect spurious correlations driven by ancestry differences in cases and controls, analysis was repeated on other LOF variants. Although the founder variant 1100delC was independently enriched in cases vs controls (OR=6.5, 95%CI=2.1-18.2, p=0.001), significant enrichment in *CHEK2* LOF variants persisted after removal of 1100delC (OR=2.9, 95%CI=1.3-6.1, p=0.0070). The LP variant c.699T>C was also independently enriched. Estimated effect sizes for all tested DDR genes demonstrated enrichment to $p<0.005$ for both *CHEK2* LP and LOF variants, as well as *FANCE* and *TP53*, although only *CHEK2* variants were enriched to $q<0.1$ (Fig. 4). In a validation cohort of ICGC samples, *CHEK2* variants were also seen, however enrichment analysis was pending availability of an appropriate set of controls (Fig. 5). Continental ancestry classification of the control cohort to be used in the final analysis showed good representation of samples across continents, enabling appropriate case-control matching by continental ancestry (Fig. 11)

Discussion

Enhancing the MOA knowledgebase to include up-to-date assertions about medical and radiation oncology positions the MOA as a resource for physician scientists analyzing sequence data. Paired with the MOA interpretation algorithm and variant reports that any user can generate through the portal, the MOA has the potential to inform compassionate, exploratory decision-making for individual cases benefitting from panel sequencing and other modalities. In particular, patients with atypical presentations such as young age, or those enrolled in research studies, may benefit from rapid, automated interpretation of variants to inform their care and generate hypotheses for further research. The inclusion of initial alteration-action relationships relevant for radiation oncology opens the door for incorporating molecular alterations into decision-making for the field, a new and growing area of interest in field where "precision medicine" to date has largely meant anatomical and physical considerations (2). The establishment of API endpoints and a lightweight Chrome extension for submitting new evidence from the browser have the potential to further encourage growth in the knowledgebase, and its use as part of other resources.

Germline sequencing has an increasing role to play in precision cancer medicine, particularly in radiation oncology as the impact of therapeutic radiation on normal tissue is considered. The importance of germline alterations for cancer risk was illustrated in preliminary analysis of one type of malignancy, CLL. It appears that *CHEK2* is enriched for germline pathogenic variants in CLL patients, as is the case for multiple other tumor types. Enrichment for LOF germline variants persisted even after separating out the known LP variant and removing a known founder mutation that could confound results due to differences in case-control ancestry. This finding, once validated in an ongoing case-control study with rigorous ancestry matching, would contribute to explaining the missing heritability in CLL patients. Given the ongoing development of therapies targeting patients with germline variants in DDR genes, most notably *BRCA1/2* which like in the same molecular pathway as *CHEK2*, a DDR gene known to harbor causative variants may prove to be a viable therapeutic target across a number of cancer types. Notably,

previously reported enrichment for *ATM* mutations was not seen here (14), although that study analyzed rare missense variants regardless of predicted pathogenicity.

Limitations

Given the rapid rate of precision oncology research, enhancing the MOA knowledgebase to include all alteration-action relationships is a moving target that cannot be attained without large investment of resources or crowd-sourcing. The attempt at crowd-sourcing by creating the Connector is still nascent and has a small number of users. To reduce duplication of effort, harmonization with other large academic knowledgebases is a valid approach however mistakes may be carried forward in this way. Furthermore, differences in terminology across studies and knowledgebases complicate conversion of alteration-action relationships into machine-readable format, a necessity for interpretation algorithms to function. While we have attempted to consistently apply a unifying framework for collecting assertions, the constant emergence of new types of evidence relevant for precision cancer medicine means the framework must continuously evolve. Computational standards for reporting the results of genetic tests are still needed.

Preliminary results identifying *CHEK2* as a CLL risk gene must be validated in multiple cohorts, now that the rigorous QC, relatedness, and ancestry-matching measures described here have been developed. Furthermore, only a subset of DDR genes was analyzed here; an exome-wide search for risk genes might produce more results, however such a search on this size of cohort may be underpowered to detect differences owing to the increased multiple-hypothesis correction that would be required.

Conclusions

Open access tools for posting and retrieving assertions, sources, genes, and specific alterations expand the scope of the Molecular Oncology Almanac for integration into external interpretation resources. These developments represent a step forward in accessible clinical interpretation of tumor molecular profiling for clinicians and researchers. Precision cancer medicine will require continued development of resources such as the MOA, powered by

expert review and crowd-sourcing methods, in order to improve the thoroughness, accuracy, consistency, and accessibility of molecular sequencing data for cancer patients.

Our preliminary results demonstrate enrichment for rare high impact germline variants in *CHEK2* among CLL patients of European descent. These findings suggest a role for DNA repair genes in germline susceptibility to CLL and are being validated with rigorous methods. Subsequent work will include evaluating rare missense variants to follow earlier findings of *ATM*.

Suggestions for Future Work

Continued expansion of MOA to include enhanced reporting of germline variants has been discussed and could help to clarify an area of variant interpretation that is still confusing to many in the field, given that tumor somatic variants are much more commonly detected and used for therapeutic decision-making. A patient cell-line matcher is under development by Reardon et al. and could offer an orthogonal method by which to expose therapeutic vulnerabilities in patient tumors. Across tumor types, many collections of paired tumor-germline WES have not been exploited to discovery important germline alterations with potential prognostic and predictive implications. Beyond CLL, similar work is underway within on many types of solid tumors. Continued studies of genetic alterations in cancer should leverage clinical data to identify molecular alterations associated with clinical phenotypes. (36)

Acknowledgements

I cannot thank Eliezer Van Allen enough for his mentorship and support. Working with him has opened my eyes to how a person can be a compassionate and outstanding mentor, role model, scientist, physician, supervisor, and human being all at once. His lab is a wonderful place and nothing I have been done would have been possible without many people who enabled me to have over a year and a half of learning and fun. All of the individuals named in the Student Role section above have been indispensable. Brendan Reardon is the perfect teammate and teacher: kind, patient, and incredibly knowledgeable about computer science, statistics, and biology. My growth in computational skills would have progressed much more slowly without his consistent presence, guidance, and positive attitude. I am grateful to Saud AlDubayan for sharing his passion for germline genetics, opening my eyes to the beauty of the subject and teaching me everything I know about it. In leading the Germline Team of Seunghun Han, Sabrina Camp, and Eric Kofman, Saud has made it possible for us to accomplish a lot of difficult work in a short amount of time with great attitude and perseverance. Thank you to the many collaborators on the CLL project for making the data collection and curation possible. Thank you to Scholars in Medicine for funding support.

References

1. Kumar-Sinha C, Chinnaiyan AM. Precision oncology in the age of integrative genomics. Nat Biotechnol. 2018;36(1):46–60.
2. Kamran SC, Mouw KW. Applying Precision Oncology Principles in Radiation Oncology. JCO Precis Oncol [Internet]. 2018;(2):1–23. Available from: http://ascopubs.org/doi/10.1200/PO.18.00034
3. Nakagawa H, Fujita M. Whole genome sequencing analysis for cancer genomics and precision medicine. Cancer Sci. 2018;109(3):513–22.
4. Staaf J, Glodzik D, Bosch A, Vallon-Christersson J, Reuterswärd C, Häkkinen J, et al. Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. Nat Med [Internet]. 2019;25(10):1526–33. Available from: http://dx.doi.org/10.1038/s41591-019-0582-4
5. Berner AM, Morrissey GJ, Murugaesu N. Clinical Analysis of Whole Genome Sequencing in Cancer Patients. Curr Genet Med Rep [Internet]. 2019;7(2):136–43. Available from: http://dx.doi.org/10.1007/s40142-019-00169-4
6. Thavaneswaran S, Rath E, Tucker K, Joshua AM, Hess D, Pinese M, et al. Therapeutic implications of germline genetic findings in cancer. Nature Reviews Clinical Oncology. Nature Publishing Group; 2019.
7. Moore K, Colombo N, Scambia G, Kim BG, Oaknin A, Friedlander M, et al. Maintenance olaparib in patients with newly diagnosed advanced ovarian cancer. N Engl J Med. 2018;379(26):2495–505.
8. Curtin NJ. DNA repair dysregulation from cancer driver to therapeutic target. Nat Rev Cancer. 2012;12(12):801–17.
9. Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, et al. Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. N Engl J Med [Internet]. 2016;375(5):443–53. Available from: http://www.nejm.org/doi/10.1056/NEJMoa1603144
10. Aldubayan SH, Pyle LC, Gamulin M, Kulis T, Moore ND, Taylor-Weiner A, et al. Association of Inherited Pathogenic Variants in Checkpoint Kinase 2 (CHEK2) with Susceptibility to Testicular Germ Cell Tumors. JAMA Oncol. 2019;5(4):514–22.
11. Hale V, Weischer M, Park JY. CHEK2∗ 1100delC Mutation and Risk of Prostate Cancer. Prostate Cancer. 2014;2014.
12. Xiang HP, Geng XP, Ge WW, Li H. Meta-analysis of CHEK2 1100delC variant and colorectal cancer susceptibility. Eur J Cancer. 2011;47(17):2546–51.
13. Nevanlinna H, Bartek J. The CHEK2 gene and inherited breast cancer susceptibility. Oncogene. 2006;25(43):5912–9.
14. Tiao G, Improgo MR, Kasar S, Poh W, Kamburov A, Landau D-A, et al. Rare germline variants in ATM are associated with chronic lymphocytic leukemia. Leukemia [Internet]. 2017 Oct 25 [cited 2018 Oct 17];31(10):2244–7. Available from: http://www.nature.com/doifinder/10.1038/leu.2017.201
15. AlDubayan SH, Giannakis M, Moore ND, Han GC, Reardon B, Hamada T, et al. Inherited DNA-Repair Defects in Colorectal Cancer. Am J Hum Genet [Internet]. 2018 Mar 1 [cited

2018 Oct 11];102(3):401–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29478780
16. Chen L, Gilkes DM, Pan Y, Lane WS, Chen J. ATM and Chk2-dependent phosphorylation of MDMX contribute to p53 activation after DNA damage. EMBO J. 2005;24(19):3411–22.
17. Bolcun-Filas E, Rinaldi VD, White ME, Schimenti JC. Reversal of Female Infertility by Chk2 Ablation Reveals the Oocyte DNA Damage Checkpoint Pathway. Science (80-) [Internet]. 2014 Jan 31;343(6170):533 LP – 536. Available from: http://science.sciencemag.org/content/343/6170/533.abstract
18. Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. JAMA [Internet]. 2016 Jan 5 [cited 2019 Feb 11];315(1):68. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2015.17703
19. Goldin LR, Pfeiffer RM, Li X, Hemminki K. Familial risk of lymphoproliferative tumors in families of patients with chronic lymphocytic leukemia: Results from the Swedish Family-Cancer Database. Blood [Internet]. 2004 Sep 15 [cited 2019 Feb 15];104(6):1850–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15161669
20. Law PJ, Berndt SI, Speedy HE, Camp NJ, Sava GP, Skibola CF, et al. Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. Nat Commun [Internet]. 2017 Apr 6 [cited 2019 May 28];8(1):14175. Available from: http://www.nature.com/articles/ncomms14175
21. Speedy HE, Kinnersley B, Chubb D, Broderick P, Law PJ, Litchfield K, et al. Germ line mutations in shelterin complex genes are associated with familial chronic lymphocytic leukemia. Blood [Internet]. 2016 Nov 10 [cited 2019 Feb 13];128(19):2319–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27528712
22. Tiao G, Improgo MR, Tausch E, Fernandes SM, Bahlo J, Robrecht S, et al. Analysis of ITGB2 rare germ line variants in chronic lymphocytic leukemia. Blood [Internet]. 2017 Nov;130(22):2443–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29051179
23. Goldin LR, McMaster ML, Rotunno M, Herman SEM, Jones K, Zhu B, et al. Whole exome sequencing in families with CLL detects a variant in Integrin $β$ 2 associated with disease susceptibility. Blood [Internet]. 2016 Nov;128(18):2261–3. Available from: http://www.bloodjournal.org/content/128/18/2261.long?sso-checked=true
24. Nicholas B. Blackburn, James R. Marthick, Annette Banks, Jac C. Charlesworth, Katherine A. Marsden, Ray M. Lowenthal JB and JLD. Evaluating a CLL susceptibility variant in ITGB2 in families with multiple subtypes of hematological malignancies. Blood. 2017;130(1):86–8.
25. Stankovic T, Weber P, Stewart G, Bedenham T, Murray J, Byrd PJ, et al. Inactivation of ataxia telangiectasia mutated gene in B-cell chronic lymphocytic leukaemia. Lancet [Internet]. 1999 Jan;353(9146):26–9. Available from: https://www.sciencedirect.com/science/article/pii/S0140673698101174?via%3Dihub
26. Beekman R, Campo E. Shelterins, a genetic crossroad in CLL. Blood [Internet]. 2016 Nov;128(19):2279–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28829750
27. Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, et al. Mutations driving CLL and their evolution in progression and relapse. Nature [Internet]. 2015 Oct 14 [cited 2019 Apr 9];526(7574):525–30. Available from:

http://www.nature.com/articles/nature15395
28. Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature [Internet]. 2015 Oct 22 [cited 2019 Feb 11];526(7574):519–24. Available from: http://www.nature.com/articles/nature14666
29. Kumaran M, Subramanian U, Devarajan B. Performance assessment of variant calling pipelines using human whole exome sequencing and simulated data. BMC Bioinformatics. 2019;20(1):1–11.
30. Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol [Internet]. 2018 Sep 24 [cited 2019 Feb 13];36(10):983. Available from: http://www.nature.com/doifinder/10.1038/nbt.4235
31. With P, Oncokb M, Food US, Comprehensive N, Network C, To R, et al. OncoKB : A Precision Oncology Knowledge Base. Am Soc Clin Oncol. 2018;
32. Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free Estimation of Recent Genetic Relatedness. Am J Hum Genet [Internet]. 2016;98(1):127–48. Available from: http://dx.doi.org/10.1016/j.ajhg.2015.11.022
33. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv [Internet]. 2019;531210. Available from: https://www.biorxiv.org/content/10.1101/531210v1
34. Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer [Internet]. Vol. 49, Nature Genetics. 2017 [cited 2018 Dec 7]. p. 170–4. Available from: http://www.nature.com/articles/ng.3774
35. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, et al. Accurate classification of BRCA1 variants with saturation genome editing. Nature [Internet]. 2018 Sep 12 [cited 2018 Sep 17];562(7726):217–22. Available from: http://www.nature.com/articles/s41586-018-0461-z
36. Wuhan T, Fudan T, Nature S. Cancer genomics gets new focus. Nature. 2020;578(Feb 6):7–8.

Tables and Figures

Table 1: Comparison of Molecular Oncology Almanac with Other Precision Cancer Medicine Resources

Service	Therapy types	Molecular feature types	Assertion types	Curation model	API	Browser extension	Open source code	Paired interpretation	Assertions	Tumor types	Citation
Molecular Oncology Almanac	Targeted, Immunotherapy, Chemotherapy, Hormone, Radiation, Combination	Somatic SNV&Indel, Somatic CN, Rearrangement, Germline SNV&Indel, Aneuploidy, Mutational Burden, Mutational Signature	Sensitivity, Resistance, Prognosis, Predisposition	Public + expert	Yes	Yes	Yes	Yes	723	56	
CIViC	Targeted, Immunotherapy, Chemotherapy, Hormone, Radiation, Combination	Somatic SNV&Indel, Somatic CN, Rearrangement	Sensitivity, Resistance, Prognosis, Predisposition	Public + expert	Yes	No	Yes	No	6473	274	[2]
OncoKB	Targeted, Immunotherapy, Chemotherapy, Combination	Somatic SNV&Indel, Somatic CN, Rearrangement	Sensitivity, Resistance	Expert	Yes	No	Yes	Yes	4932	45	[3]

Table 2: Previous CLL Study Cohorts

# CLL cases	Continental ancestry	Study type	Germline finding	Reference
646	European	Germline	Enrichment of rare variants in *ATM* (OR 1.6-1.8, q=0.00017)	Tiao et al. Leukemia 2017
538	European	Somatic	N/A	Landau et al., Nature 2015
506	European	Somatic	N/A	Puente et al., Nature 2015

Table 3: DNA damage repair and cell cycle genes included in gene burden tests

APC	FANCA	PALB2
ATM	FANCC	PMS2
ATR	FANCD2	POLD1
BAP1	FANCE	POLE
BARD1	FANCF	POLH
BLM	FANCG	PTEN
BMPR1A	FANCI	RAD51
BRCA1	FANCL	RAD51C
BRCA2	FANCM	RAD54L
BRIP1	GEN1	RECQL4
CHEK2	MLH1	RNF43
DDB2	MRE11A	SLX4
ERCC2	MSH2	SMAD4
ERCC3	MSH6	STK11
ERCC4	MUTYH	TP53
ERCC5	NBN	UBE2T
	NTHL1	WRN

Figure 1. Source of missing heritability in CLL has evaded extensive study. Few Mendelian genes of potential clinical importance are known. Our recent study identified rare missense variants in ATM as associated with CLL risk.

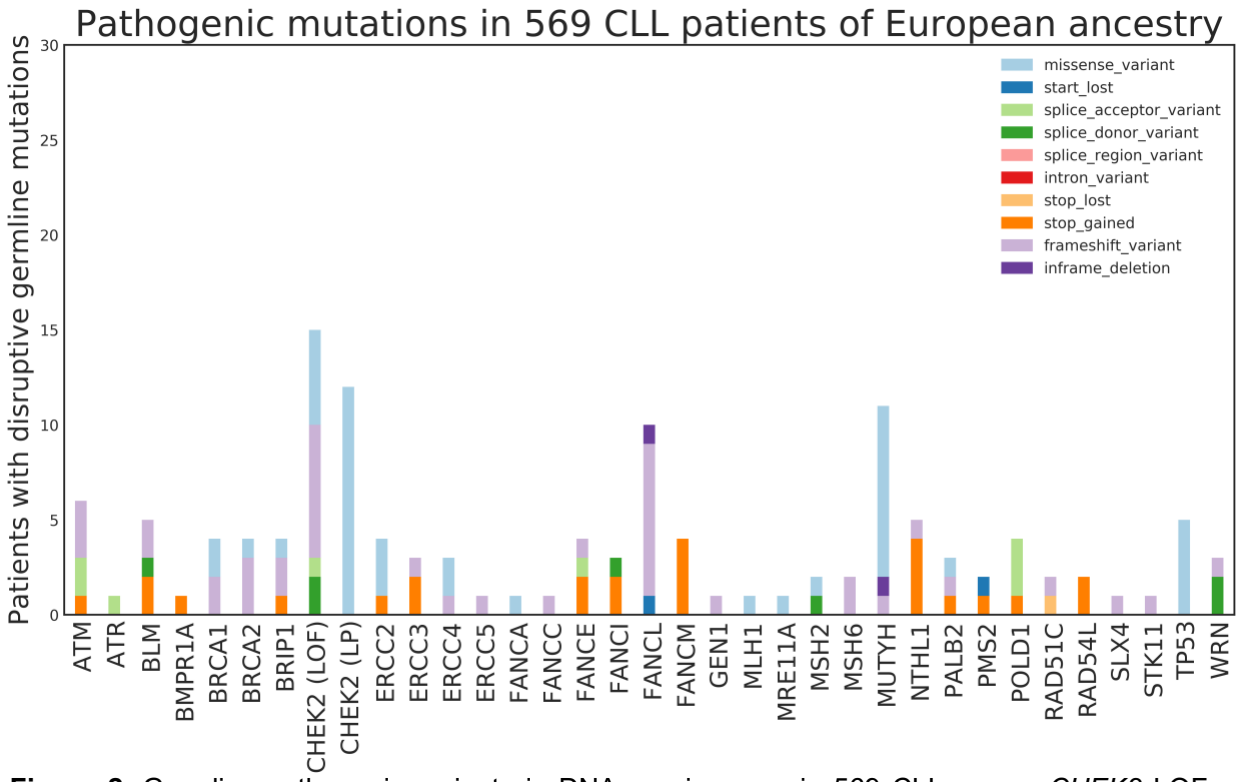

Figure 2. Germline pathogenic variants in DNA repair genes in 569 CLL cases. *CHEK2* LOF variants are found in 15 (2.6%) cases and 61 (0.7%) controls. *CHEK2* low penetrance (LP) variant c.599T>C is found in 12 (2.1%) cases and 24 (0.3%) controls.

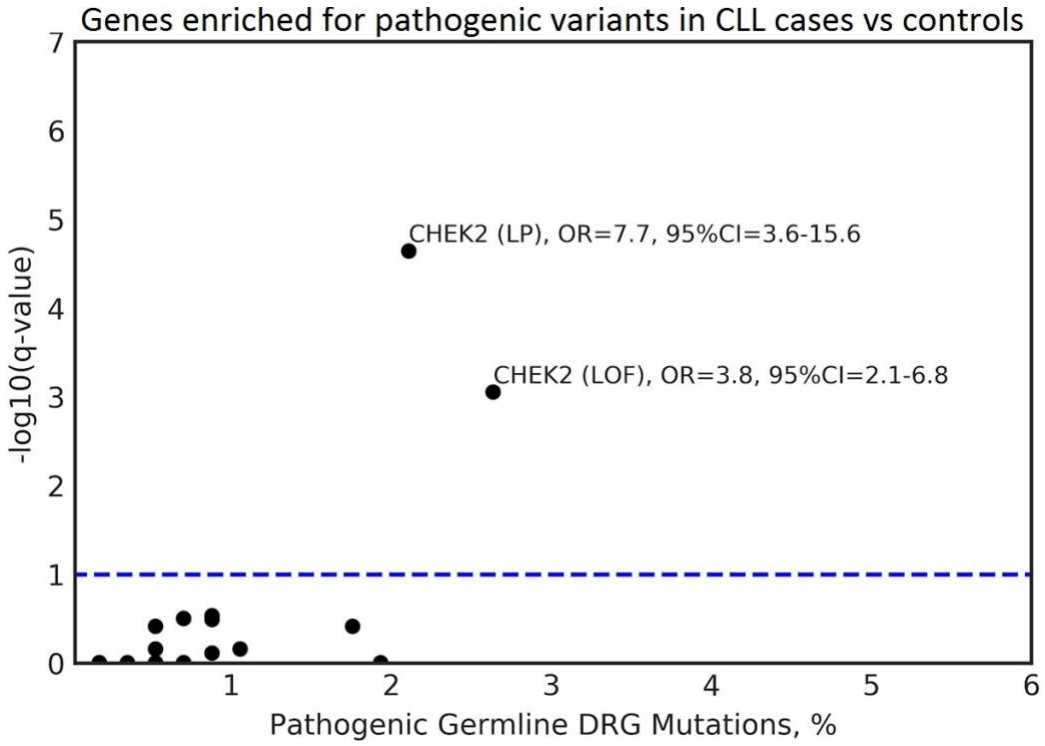

Figure 3. CLL cases are enriched for *CHEK2* pathogenic variants relative to ancestry-matched controls. Both LOF variants and the low-penetrance (LP) allele c.599T>C are enriched. Although the known founder *CHEK2* LOF variant 1100delC is independently enriched in cases vs controls (OR=6.5, 95%CI=2.1-18.2, p=0.001), significant enrichment in *CHEK2* LOF variants persists after removal of 1100delC (OR=2.9, 95%CI=1.3-6.1, p=0.0070). Blue line: q=0.10

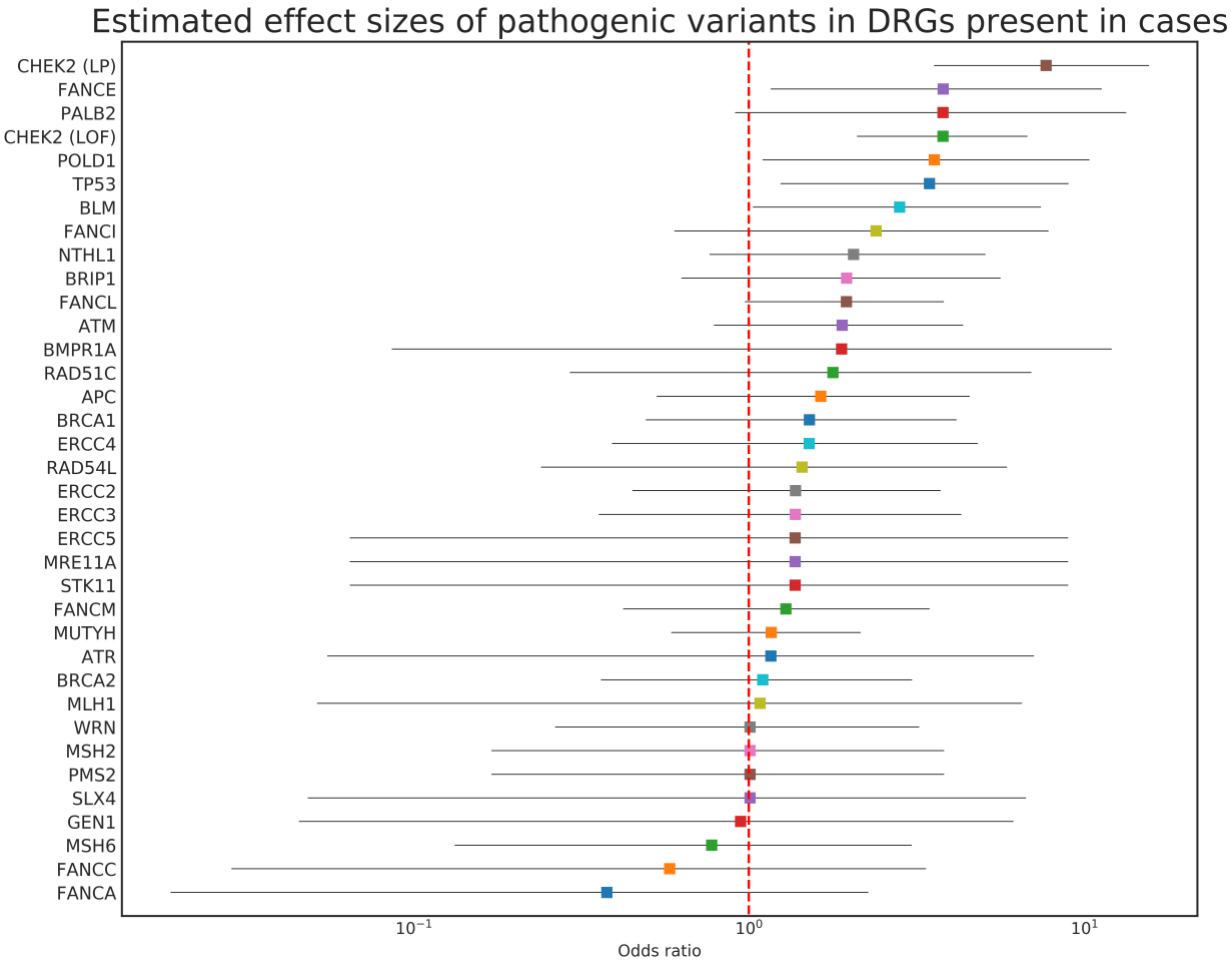

Figure 4. Estimated effect sizes and 95% confidence intervals of germline pathogenic variants in 569 cases . Both LOF variants and the low-penetrance (LP) allele c.599T>C are enriched. *FANCE* and *TP53* are enriched to p<0.05 although not to q<0.1.

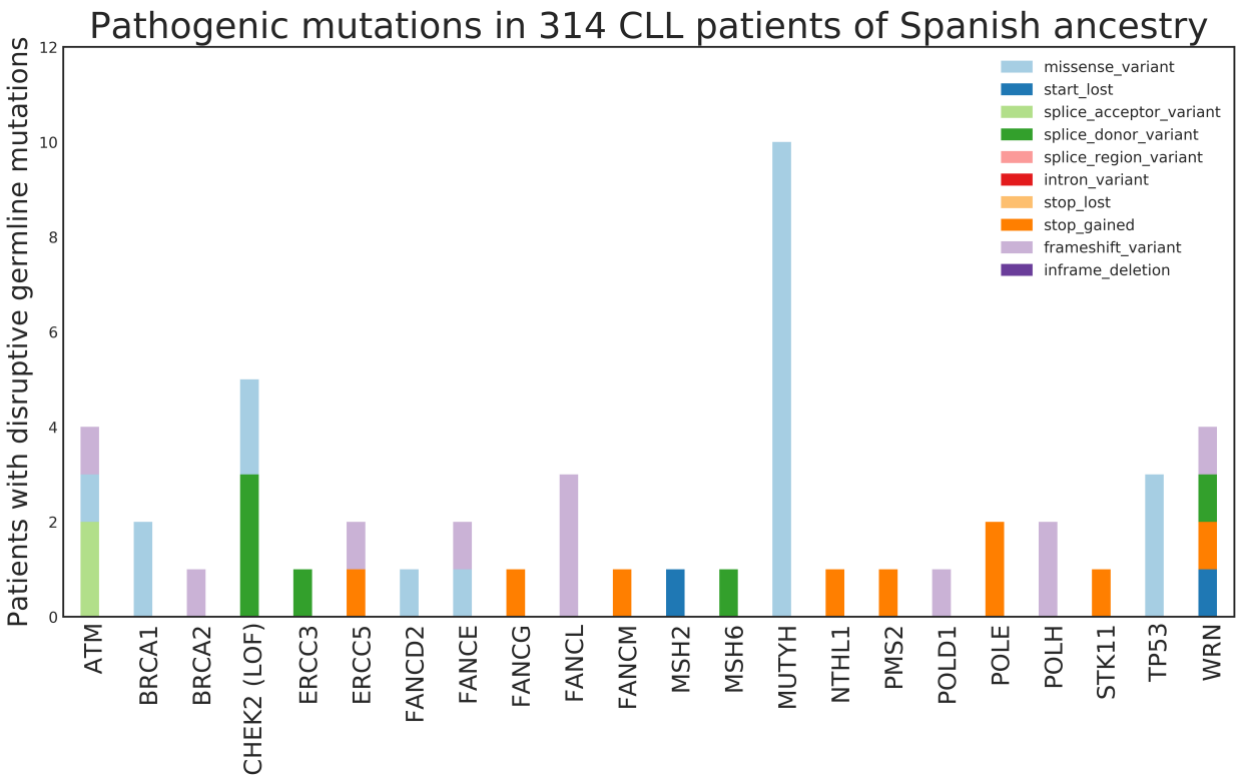

Figure 5. Germline pathogenic variants called in DRGs in CLL patients of Spanish ancestry. *CHEK2* LOF variants are found in 1.6% of cases. Enrichment analysis pending appropriate ancestry-matched controls.

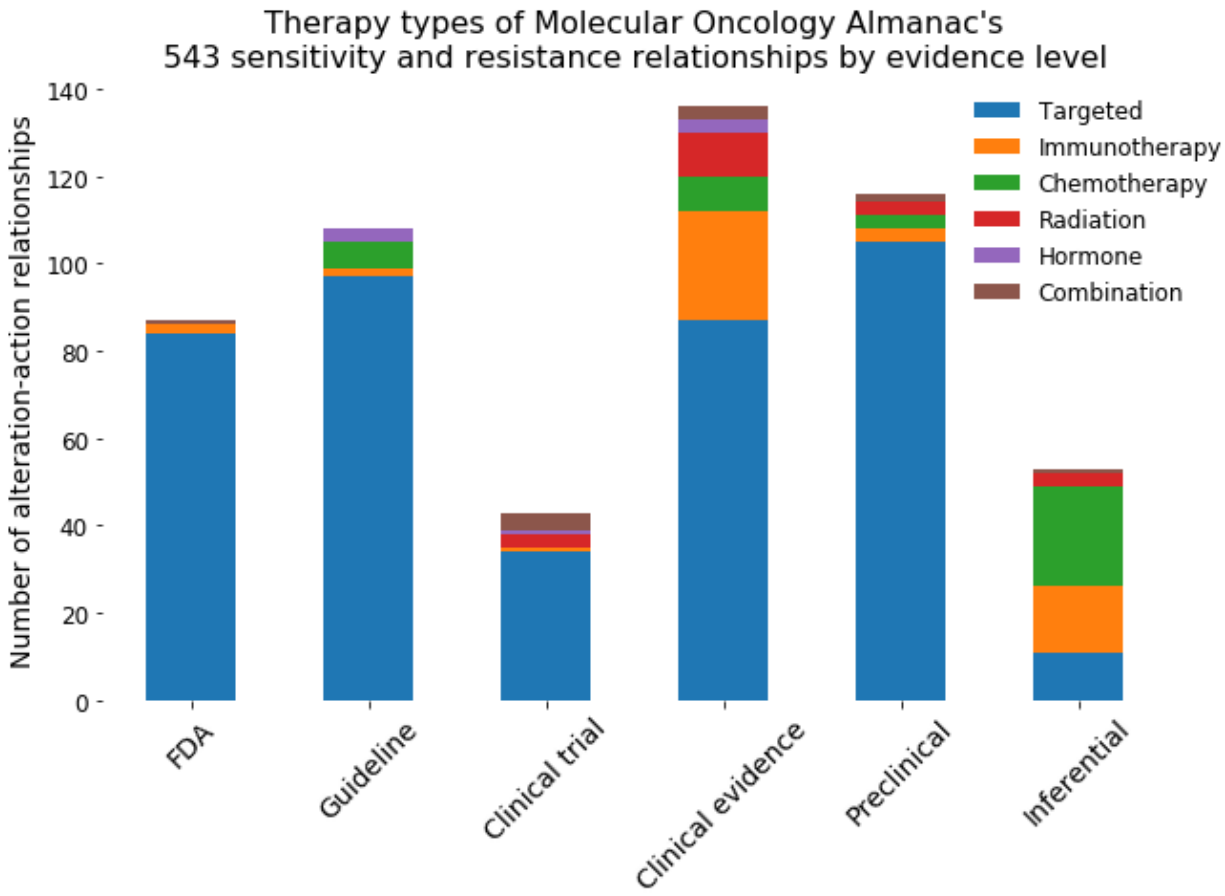

Figure 6. Molecular Oncology Almanac captures assertions across a wide range of therapy types.

f

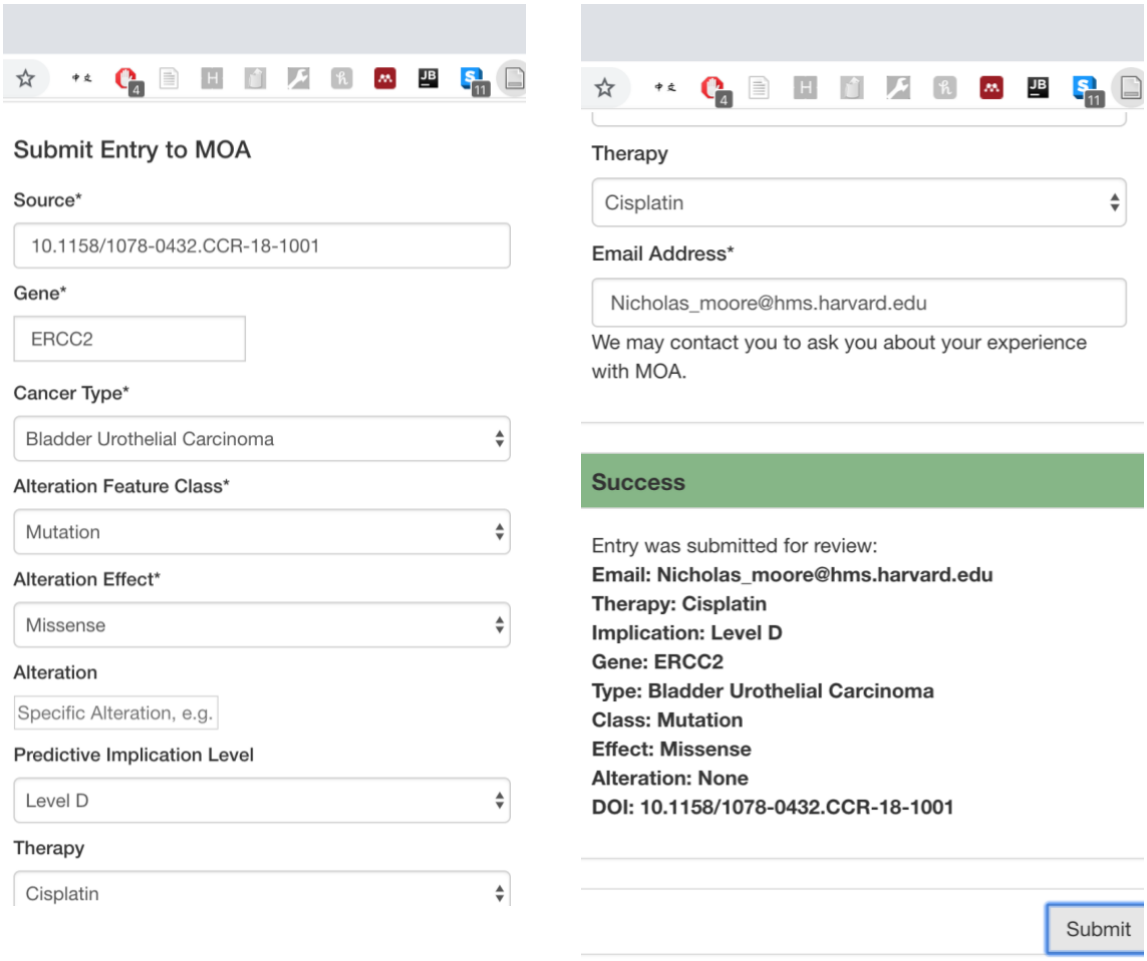

Figure 7. Molecular Oncology Almanac Connector browser extension available through the Chrome webstore enables rapid, lightweight addition of assertions to the knowledgebase.

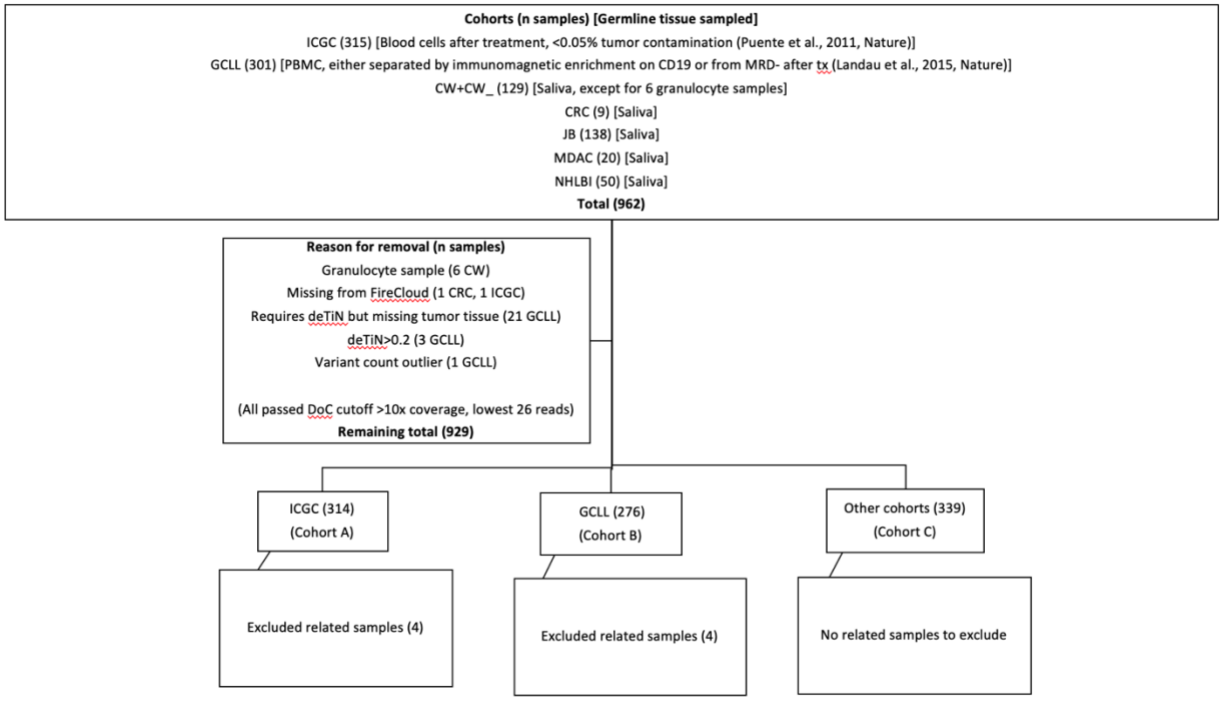

Figure 8: CLL case cohort procurement and QC results.

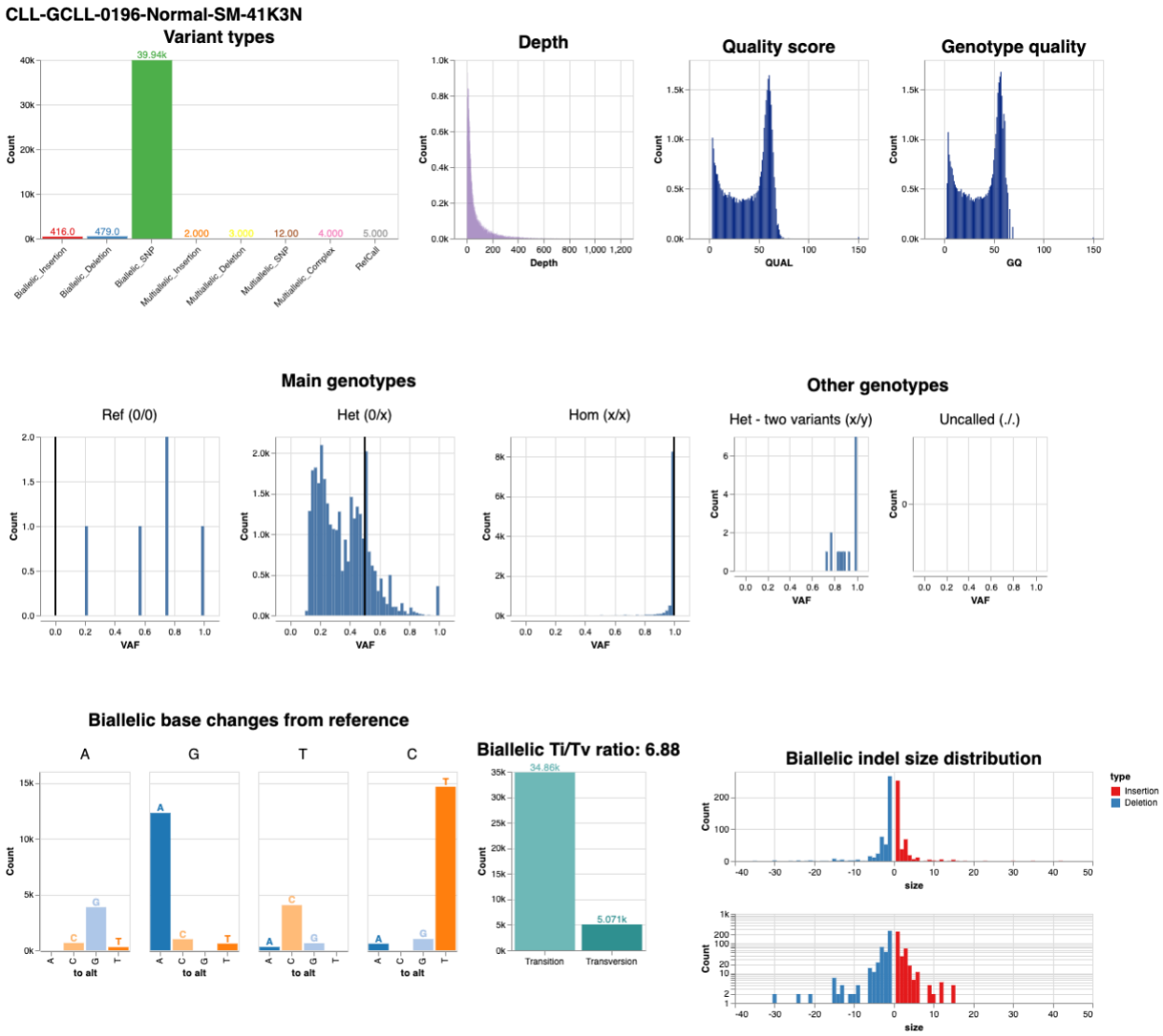

Figure 9: Quality control metrics for single sample rejected due to being an outlier with 41k variants.

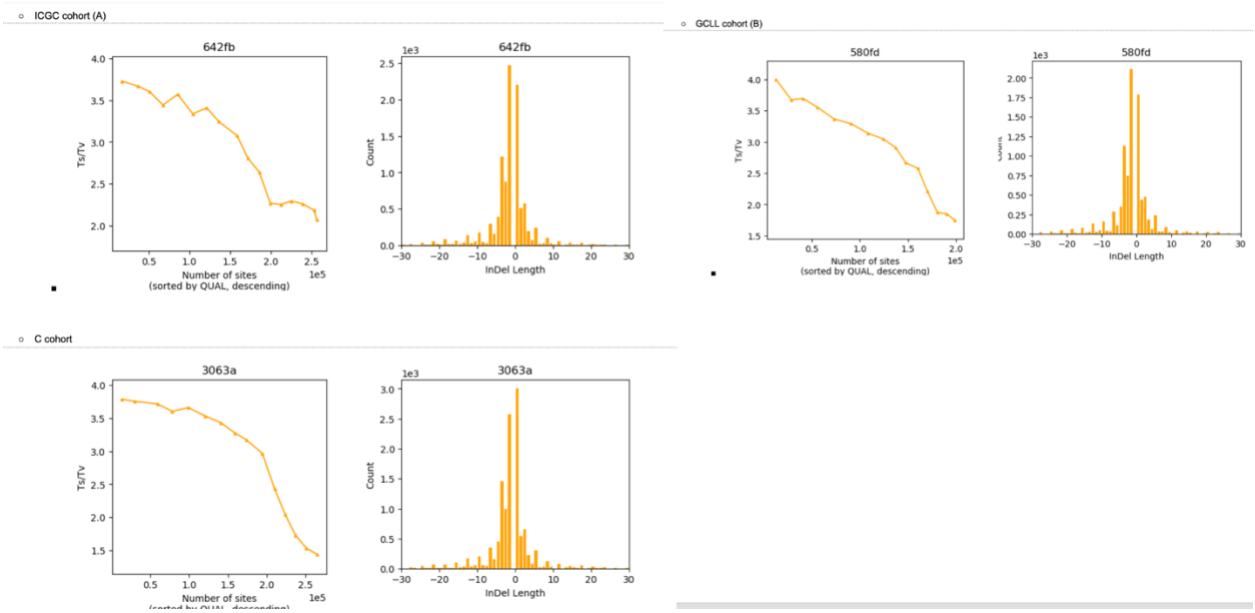

Figure 10: Quality control metrics at cohort level for 3 CLL case cohorts.

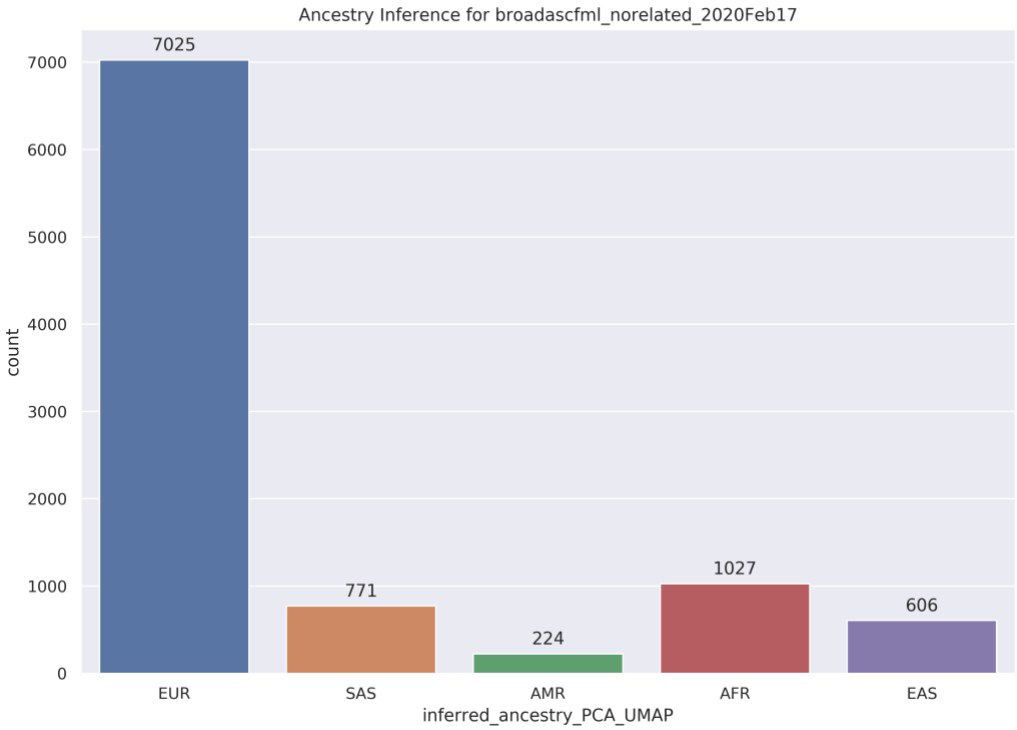

Figure 11: Continental ancestry groupings for control cohort used in preliminary analysis. EUR: European, SAS: South Asian, AMR: Admixed American, AFR: African, EAS: East Asian.

Appendix: Related peer-reviewed publications

American Medical Informatics Association 2019 Annual Symposium

Enhancing Accessibility of a Molecular Oncology Almanac for Precision Cancer Medicine

Nicholas S. Moore, BA[1,2,3], Brendan Reardon, BS[2,3], Eric Kofman, BS[2,3], Nathanael Moore, BS[4], Eliezer Van Allen, MD[2,3]
[1]Harvard Medical School, Boston, MA; [2]Dana-Farber Cancer Institute, Boston, MA; [3]Broad Institute of MIT and Harvard, Cambridge, MA; [4]Indiana University School of Medicine, Indianapolis, IN

Problem

Cancer care is increasingly informed by molecular alterations that predict therapeutic sensitivity, therapeutic resistance, and prognosis. Clinical and preclinical studies continue to identify somatic and germline alterations that impact clinical decision-making, increasing the complexity of interpreting tumor molecular profiles for patient care and research. Increasing access to point-of-care molecular profiling holds great potential for patients but will only translate to improved outcomes if clinicians and researchers are able to interpret patient data in the context of a rapidly changing and broadening research landscape.

Towards solving this problem, we have built and continue to extend the Molecular Oncology Almanac, an open-access combined genomic interpretation algorithm, knowledge system, and web portal for informing treatment decisions through rapid assessment of tumor actionability. Analysis encompasses not only single nucleotide variants, insertions/deletions, and copy number alterations, which are commonly used at the point of care, but also anticipates growing application of fusions, mutational burden, mutational signatures, microsatellite instability, and aneuploidy. A major obstacle to maintaining an up-to-date knowledgebase of actionable alterations lies in the difficulty of incorporating relevant assertions from the literature as soon as they are available. To improve the accessibility of these assertions for the wider oncology research community, we adapted the Molecular Oncology Almanac by implementing open API endpoints and a browser extension for the purpose of rapid assertion acquisition.

Aims

1. Increase accessibility of the Molecular Oncology Almanac precision medicine knowledgebase through open-access API endpoints.
2. Implement a user-friendly interface for submitting actionable alterations to the knowledgebase directly from journal websites.

Methods

RESTful API endpoints for getting and posting alteration-action relationships to the Molecular Oncology Almanac were developed within the existing Flask web framework. Endpoints were designed within the OpenAPI 3.0 standard and documented in Swagger. Using these endpoints, a Google Chrome browser extension for posting alteration-action relationships was developed. Every submitted assertion includes a citation, class of alteration, cancer type, alteration effect, and level of evidence; gene, alteration, and therapy may also be included where relevant. Testing of endpoints and the extension was performed by multiple users in the Van Allen Lab. Submitted assertions are reviewed by site administrators prior to final incorporation into the database.

Results

The Connector browser extension was submitted to the Google Chrome Web Store, with final release pending approval. Multiple test users have submitted assertions through the pre-release extension. The Connector streamlines the process of submitting actionable alterations while reviewing the literature, reducing the effort required for knowledge base maintenance.

Conclusion

Open access tools for posting and retrieving assertions, sources, genes, and specific alterations expand the scope of the Molecular Oncology Almanac for integration into external interpretation resources. These developments represent a step forward in providing accessible clinical interpretation of tumor molecular profiling to clinicians and researchers. The Connector will be presented to cancer research audiences aiming for wide adoption in the community.

1679

Journal of Clinical Oncology®
An American Society of Clinical Oncology Journal

CANCER PREVENTION, HEREDITARY GENETICS, AND EPIDEMIOLOGY

Inherited DNA repair and cell cycle gene defects in chronic lymphocytic leukemia.

Nicholas S. Moore, Saud H. Aldubayan, Amaro Taylor-Weiner, Stephan Stilgenbauer, Gad Getz, Catherine J. Wu, Eliezer Mendel Van Allen, Jennifer R. Brown

Dana-Farber Cancer Institute, Boston, MA; The Broad Institute of MIT and Harvard, Cambridge, MA; Broad Institute, Cambridge, MA; Department of Internal Medicine III, University of Ulm, Ulm, Germany; Broad Institute of MIT and Harvard, Cambridge, MA

Abstract

1508

Background: Chronic lymphocytic leukemia (CLL) is among the most heritable cancers, with 60% of disease risk genetically determined. However, most of the genetic heritability of CLL remains unexplained. Previously, we identified *ATM* as the first CLL risk gene. Here, we leverage a deep-learning-based germline variant calling algorithm to explore germline mutational enrichment in DNA repair and cell cycle genes in CLL. **Methods:** A two-stage case-control analysis was conducted using gene-based mutational enrichment analysis of 50 established cancer predisposition DNA repair and cell cycle genes. In the discovery phase, a total of 285 Spanish patients and 5,608 ancestry-matched controls were evaluated. In the validation stage, an independent cohort of 514 European patients and 27,173 ancestry-matched controls were analyzed. An FDR correction was applied to both datasets and genes with a q-value < 0.2 in both cohorts were considered significant. **Results:** Our joint analysis of 799 CLL patients from 2 genetically distinct cohorts and 32,781 ancestry-matched cancer-free controls identified *ATM* and *CHEK2* as significantly enriched in both CLL datasets. First, our analysis recaptured the previously reported finding of *ATM* variant enrichment in CLL patients. Carriers of pathogenic *ATM* mutations in our cohorts (n = 9 patients, discovery: 1.05%, validation: 1.17%) were 2.8–3.7 times more likely to develop CLL compared to cancer-free individuals (discovery: OR = 2.8, 95%CI = 0.7–9.0, q-value = 0.181; validation: OR = 3.7, 95%CI = 1.6–8.3, q-value = 0.0454). In addition, our analysis identified 21 CLL patients carrying pathogenic *CHEK2* alterations (discovery: 1.40%, validation: 3.31%), making CLL patients 4.4-8.0 times more likely to carry such alterations compared to controls (discovery: OR = 8.0, 95%CI = 2.3–27.0, q-value = 0.026; validation: OR = 4.4, 95%CI = 2.5–7.3, q-value < 0.001). **Conclusions:** Our analysis of genetically distinct CLL cohorts, using a high-sensitivity variant calling algorithm, supports CHEK2 as a potentially novel CLL predisposition gene that may explain a portion of the missing monogenic heritability of CLL. In addition, this study highlights the DNA repair and cell cycle regulation pathways as potential drivers of CLL susceptibility.

ARTICLE CITATION

DOI: 10.1200/JCO.2019.37.15_suppl.1508
Journal of Clinical Oncology 37, no. 15_suppl (May 20, 2019) 1508-1508.

Published online May 26, 2019.

Cancer Research

Bioinformatics, Convergence Science, and Systems Biology

Abstract 2470: A molecular oncology almanac for integrative clinical interpretation of molecular profiles to guide precision cancer medicine

Brendan Reardon, Nicholas Moore, Nathanael Moore, Eric Kofman, and Eliezer Van Allen

DOI: 10.1158/1538-7445.AM2019-2470 Published July 2019 Check for updates

Proceedings: AACR Annual Meeting 2019; March 29-April 3, 2019; Atlanta, GA

Abstract

Background: Tumor molecular profiling is increasingly used to detect first-order genomic alterations associated with therapeutic actions (e.g. BRAF V600E & RAF/MEK inhibition). Simultaneously, more complex molecular features are being discovered and applied to clinical scenarios (e.g. mutational signatures, somatic-germline interactions). As patients receive expanded profiling, such as clinical whole-exome and RNA sequencing, novel algorithms are needed to integrate interpretation of multiple data modalities. Furthermore, the clinical-preclinical gap continues to widen as data from high-throughput screens of cancer cell lines are generated without accessibility at the point of care. Here, we introduce a paired interpretation algorithm and knowledge system for cancer genomic data, the Molecular Oncology Almanac, to inform treatment decisions through rapid assessment of tumor actionability.

Methods: We implemented a cloud-based interpretation algorithm that annotates and evaluates variants from WES and RNA-seq (SNVs from WES and RNA-seq, InDels, CNAs, and fusions) and infers additional features such as mutational burden, mutational signatures, MSI, somatic-germline interactions, and aneuploidy. Predictive implication levels were assigned to reflect confidence in the database's catalogued relationships to therapeutic response and prognosis for each molecular feature. We also developed a patient-preclinical matchmaker function to expand the theoretical therapeutic modalities for any given patient. Towards timeliness of updates and knowledge system accessibility, we developed API endpoints, a browser extension for suggesting citations, and workflows in the FireCloud framework.

Results: A total of 260 patients with metastatic castration-resistant prostate cancer (n=150) and metastatic melanoma (n=110) were evaluated with 569 alteration-action relationships catalogued in the Molecular Oncology Almanac. Overall 80% of patients had at least one alteration suggesting therapeutic sensitivity based on FDA approval, clinical trials, or studies in humans; which increased to 95.8% by also considering preclinical and inferential associations. Per patient, the matchmaker function on average

highlighted 1.56 additional therapies that would not have otherwise been nominated. At least one feature associated with resistance or prognosis was observed in 85% and 90% of patients, respectively.

Conclusion: Clinical actionability of sequence data was increased by including integrative molecular profiling of DNA and RNA, global molecular features, and preclinical alteration-action relationships. Increased accessibility of clinical interpretation through our cloud-based web portals and API endpoints may aid in sample contextualization.

Source code and a web portal for this project are available at moalmanac.org.

Citation Format: Brendan Reardon, Nicholas Moore, Nathanael Moore, Eric Kofman, Eliezer Van Allen. A molecular oncology almanac for integrative clinical interpretation of molecular profiles to guide precision cancer medicine [abstract]. In: Proceedings of the American Association for Cancer Research Annual Meeting 2019; 2019 Mar 29-Apr 3; Atlanta, GA. Philadelphia (PA): AACR; Cancer Res 2019;79(13 Suppl):Abstract nr 2470.

Previous

Back to top

July 2019
Volume 79, Issue 13 Supplement
Table of Contents

Search this issue

Sign up for alerts

Request Permissions
Article Alerts
Email Article
Citation Tools

Share
Tweet
Like 0