INTEGRATING MOLECULAR AND CLINICAL DATA FOR REAL WORLD EVIDENCE RESEARCH IN PRECISION ONCOLOGY

Scott Phillips, B.S.[1], Xiaojun Xu Ph.D.[1], Rishaan Kenkre[2], Paulina Paul, M.S.[2], Wenhong Zhu PhD[3], Michael Hogarth M.D.[3,4], Olivier Harismendy, PhD [1,4]

1) Moores Cancer Center, 2) Department of Bioengineering, UC San Diego Jacobs Schools of Engineering, 3) Altman Clinical and Translational Research Institute, 4) Department of Biomedical Informatics. UC San Diego School of Medicine, La Jolla CA.

In precision oncology, Real World Evidence (RWE) research depends on clinicopathological data (stage, grade, diagnosis, outcome), which is not often available as structured data and hence requires manual curation of Electronic Health Record (EHR) data. Moreover, the practice of precision oncology increasingly relies on the results of DNA sequencing assays (NGS), which are available only as narrative text reports and not adequate for large-scale research use. Research initiatives such as the AACR GENIE consortium are aggregating data, relying on their members to share de-identified test reports accompanied with a minimal set of clinical information. Similarly, prospective clinical studies have collected molecular data and curated the clinical information to evaluate predictive biomarkers.

Here we present an alternate approach that minimizes curation in the assembly of molecular and clinical datasets and which was used to create the **Aggregated Registry for Molecular Oncology Research (ARMOR),** a large-scale registry supporting RWE research in precision oncology. We illustrate the validity and utility of ARMOR by addressing known and novel precision oncology questions

Clinical Phenotypes

6,098 tests (4,900 patients) ordered 2012-2020

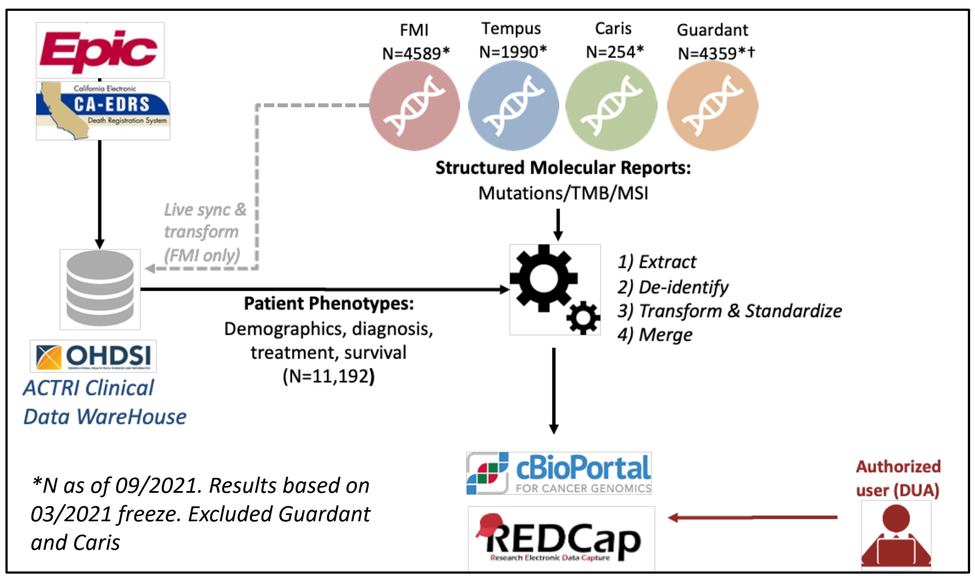
Demographics
- Sex
- Age (median 62)
- Race (66% White)
- Ethnicity (17% Hispanic)
- Insurance type (private vs non-private)

Cancer primary site standardization
- 86% mapped to Oncotree* organs (via ICD10). **Kundra R, et al. JCO Clin cancer informatics. 2021*

Survival (from 1st test date)
- 2097 deceased (median 8.7 months)
- 2797 alive (median 17 months)

Building ARMOR

A no-curation automated registry

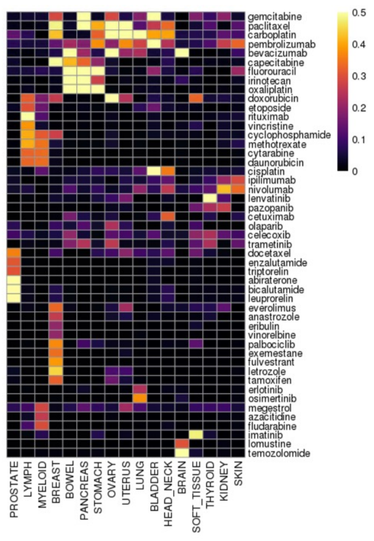


- Structured NGS report data are collected directly from the laboratories, extracted, and transformed to flat files following the cBioPortal extended formats.
- The associated meta-data including laboratory, assay-type and version, diagnosis, specimen site and primary organ at order, Tumor Mutational Burden, or Microsatellite instability are added to sample and patient tables.
- The clinical and demographic data is obtained from the UCSD Health Clinical Data Warehouse for Research
- The data is anonymized, and all dates converted to duration since first report and made available to approved UCSD Health investigators via a local cBioPortal server.

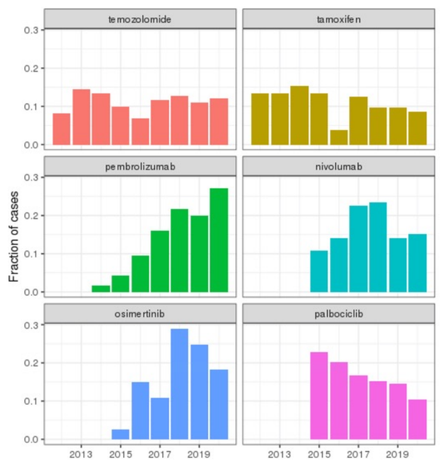
Daily structured reports. Bi-annual data freeze with automated ETL

Medical Treatment Data

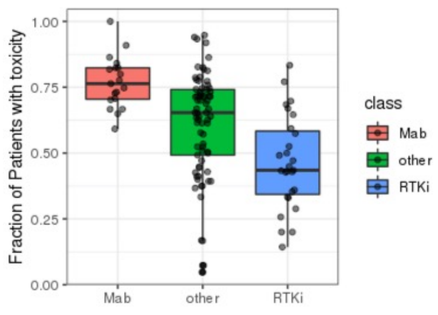
Antineoplastic and Endocrine treatment
- Identified in 76 % of cases
- 426 treatments (RxNorm) mapped to 163 agents (ATC 5th order)
- Treatment duration and drug holidays inferred from outlier intervals between repeated prescriptions



Treatment prevalence in each cancer type illustrates clinical guidelines. *The color indicate fraction of cancer type prescribed an agent.*
- Frequent combinations validate treatment regimen (e.g. FOLFOX in colorectal cancer)
- Some agents are ubiquitous (e.g. carboplatin or pembrolizumab)



Treatment evolution. Increase in prescription illustrates the approval of novel agents (e.g. Nivolumab in 2015), or novel line of therapy (e.g. Osimertinib as first line treatment in 2018)

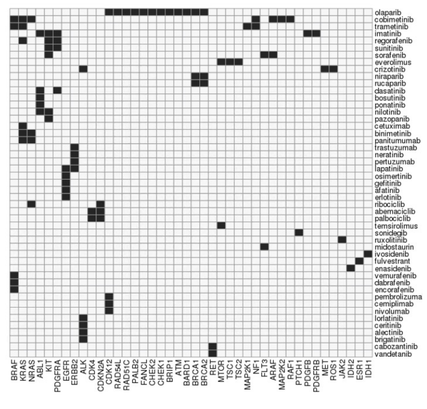


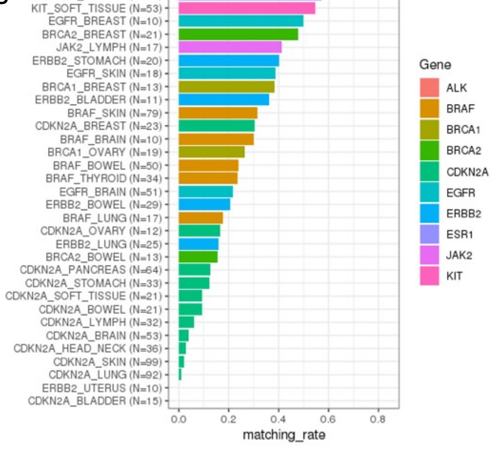
Inferred toxicity. The fraction of patients for which an agent dose or formulation is changed illustrates the increased safety of RTK inhibitors

Medical treatment data faithfully captures real world practice

Matching Targeted Therapies

2586 (55%) patients were prescribed Targeted Therapies





Targeted treatment matching reference. Solid cells represent a match between agent and targeted gene /pathway.
- From OncoKB: Chakravarty et al. JCO Precis Oncol. 2017

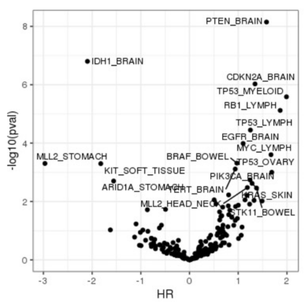
Cancer and Target Specific matching rate. Fraction of patients with mutation in a targetable gene and prescribed a matching targeted therapy.
- High matching rate in approved settings (ALK mutated lung cancer, JAK2 mutated heme malignancies).
- Low matching in speculative, untested settings (ERBB2 in Uterine cancer)

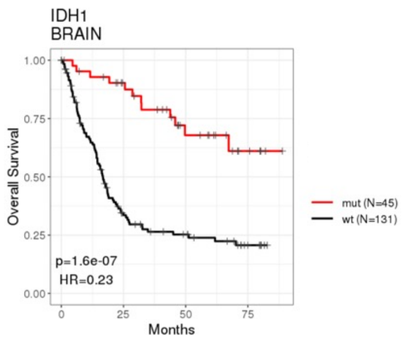
Correlations between mutational profiles and targeted therapy prescriptions captures real world practice

Prognostic Mutations

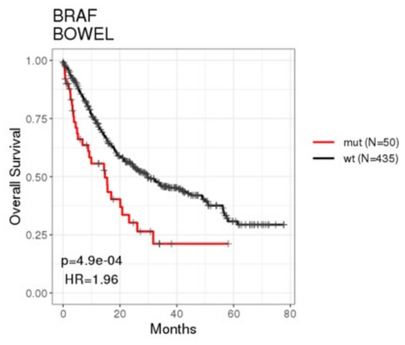
200 pairs of genes and cancer type had more than 10 patient per arm (WT or Mutated). Includes 50 genes and 18 cancer types

18 significant prognostic associations (q<0.05). 4 good prognosis (incl. TT), 14 poor (incl. 5 for BRAIN)





Example of a good prognosis interaction. IDH1 mutated Brain cancer patients have better survival. Consistent with Sanson et al. JCO (2009)

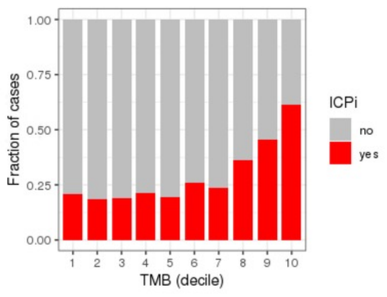


Example of a poor prognosis interaction. BRAF mutated Colorectal cancer patients have worse survival. Consistent with French et al. CCR (2008)

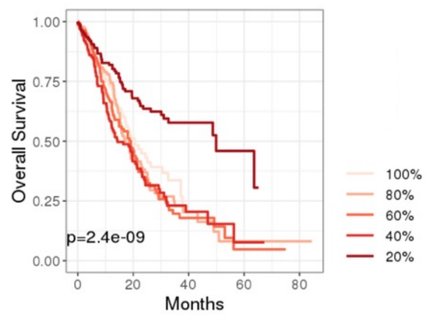
Prognostic mutational biomarkers identified or validated via retrospective design

Immunotherapy Response Prediction

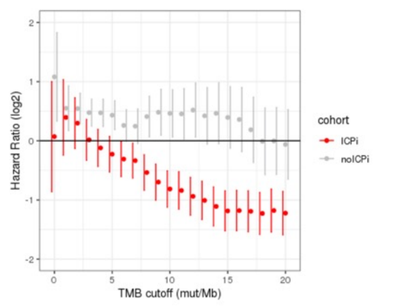
2,401 Patients with available Tumor Mutational Burden (TMB). 700 (29%) treated with immunotherapy (ICPi / checkpoint blockade)



Patient with High TMB are more likely to receive Immune checkpoint inhibitor



Patient with high TMB have better survival



TMB greater than 8 mutation/Mb is predictive of response to ICPi

Real World Data complements clinical trials and helps implement and optimize novel therapies

Limitations

Genetic labs use different gene panels. Hard to distinguishing false negative from un-mutated.

Data less complete than curated cancer registries. Missing
- Cancer stage
- Pathological diagnosis
- Initial diagnosis date

No Progression information / RECIST not available without chart review.

Treatment compliance and start/stop reasons not monitored.

Conclusions

ARMOR faithfully captures a snapshot of a molecular oncology clinic using automated and real time collection of both molecular and clinical information.

Despite inherent limitations, ARMOR provides a platform to study the practice of precision oncology at a high-level:
- rapidly validate research findings in an independent, real world patient population
- document patients' trajectories
- develop predictive models
- test novel therapeutic paradigms