Development and Validation of a Real-World Clinico-Genomic Database
Gaurav Singal1, Peter Grant Miller1, Vinenta Agana2, Jie He3, Anaka Gossai4, Gerald Li5, Shannon Franklin5, David Bourque5, Bryan Boser5, Thomas Carson5, Ezra Bzyb5, Katlin Seid-Rathkopf5, Ivan Ivanov6, Alex Parker7, Amar Karla7, Gamlheit Michael Frankton8, Ann Jaskiew8, Dana Feuchtwang8, Nathan Coleman9, Amy Pickel Averbuch8, Vincent A. Miller1
1Foundation Medicine, Inc., Cambridge, MA; 2Flattion, Health, New York, NY.

BACKGROUND
• Genomic findings have diagnostic, prognostic, and predictive utility in clinical oncology.
• Population studies have been limited by reliance on trials, registers, or institutional chart reviews, which are costly and represent narrow populations.
• Integrating electronic health record (EHR) and genomic data collected as part of routine clinical practice may overcome these hurdles.

METHODLOGY
• Oncology patients from community practices were identified for whom Flatiron EHR abstraction and Foundation Medicine’s next generation sequencing (NGS) were performed.
• The information was linked in a HIPAA-compliant fashion through a third party to create the clinico-genomic database (CGDB), which is publicly available.
• Currently there are 2139 non-small cell lung cancer (NSCLC) cases, which were used as a validation set for the database.

RESULTS

CLINICAL CHARACTERISTICS
Table 1. Clinical characteristics of patients in the clinico-genomic database. The distribution of features such as median age, smoking history, and histology are consistent with prior studies.

Genomic characteristics of the NSCLC tumors in the clinico-genomic database are largely consistent with prior studies in large genomic database by Foundation Medicine, Inc., Cambridge, NY.

CRITICAL GENOMIC FEATURES

Table 2. Genomic characteristics of the NSCLC tumors in the clinico-genomic database are largely consistent with prior studies in large population-based databases including the TCGA. As expected, the clinico-genomic database is enriched for driver mutations (GSK3, AKT, ROS1, MET, BRAF, RET, or HER2) which were associated with younger age, female gender, and non-smoking status.

Table 3. Comparison of EHR and Flatiron Foundation Medicine, Inc., Cambridge, NY.

GROWTH OF THE CLINICO-GENOMIC DATABASE

Figure 7. Patient counts and growth of the clinico-genomic database, by disease (June 2017). The CGDB covers 38 tumor types and continues to grow and update on a quarterly basis.

FUTURE DIRECTIONS

Figure 6. Using the CGDB to understand and predict response to therapy. The presence of an EGFR or ALK driver mutation is associated with a higher OS, with variation among the specific driver subtypes.

TESTING AND THERAPEUTIC RESPONSE PREDICTION

Table 4. Comparison of EHR and Flatiron Foundation Medicine, Inc., Cambridge, NY.

FUTURE APPLICATIONS

• Unseen Needs
  • Populations for whom current treatments do not exist
  • Therapies for whom an appropriate population needs to be better defined

• Trial Design
  • Characterizing the natural history of a biomarker-defined population for trial design

• Integration of NGS testing into tumor biomarker and drug discovery

• Targeted Therapy
  • Refining genomic lesions for drug development
  • Better understanding of resistance to current therapies

• Immune Ovarian
  • Integrating tumor mutation burdens into our preclinical and predictive algorithms

• Detecting genomic subpopulations with differential sensitivity to checkpoint blockade

• Rational approaches to combining targeted therapy with checkpoint blockade

CONCLUSIONS

We have built a de-identified, HIPAA-compliant, real-world clinico-genomic database by linking longitudinal clinical data with high resolution genomic information. The database consists of 2139 NSCLC cases, more than 20,000 total cases, and is both growing and updating on a quarterly basis.

The clinico-genomic database shares similar genomic and clinical characteristics as NGS-tested population estimates, and recapitulates a broad array of expected findings regarding (X) genomic prognostic factors, (Y) clinical prognostic factors, and (Z) general observations for therapeutic response.

Future uses include novel biomarker discovery, better clinical trial design, comparative effectiveness of therapeutics, and better characterizing natural history of genomic subpopulations (e.g., to serve as in situ control arms).