RESULTS

CLINICAL CHARACTERISTICS

Table 1. Clinical characteristics of patients in the genomic database. The distribution of features such as median age, smoking status, and gender are consistent with prior studies.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Total Patient Count</th>
<th>Stage IV</th>
<th>Stage III</th>
<th>Stage II</th>
<th>Stage I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>90,289 (63.4)</td>
<td>28,978</td>
<td>31,126</td>
<td>27,316</td>
<td>3,879</td>
</tr>
<tr>
<td>Sex</td>
<td>50,112 (35.5)</td>
<td>16,112</td>
<td>18,053</td>
<td>15,631</td>
<td>925</td>
</tr>
<tr>
<td>Race</td>
<td>17,088 (28.9)</td>
<td>4,635</td>
<td>5,903</td>
<td>4,541</td>
<td>209</td>
</tr>
<tr>
<td>Smoking History</td>
<td>71,893 (51.1)</td>
<td>23,381</td>
<td>25,392</td>
<td>22,024</td>
<td>1,115</td>
</tr>
<tr>
<td>History of Prior Cancer</td>
<td>60,899 (41.4)</td>
<td>17,597</td>
<td>19,870</td>
<td>16,473</td>
<td>752</td>
</tr>
<tr>
<td>History of Diagnosis</td>
<td>119,253 (83.5)</td>
<td>34,729</td>
<td>39,146</td>
<td>31,827</td>
<td>7,453</td>
</tr>
<tr>
<td>Unknown / Not Documented</td>
<td>40,998 (28.7)</td>
<td>11,091</td>
<td>13,361</td>
<td>11,140</td>
<td>5,406</td>
</tr>
<tr>
<td>Race</td>
<td>17,088 (28.9)</td>
<td>4,635</td>
<td>5,903</td>
<td>4,541</td>
<td>209</td>
</tr>
<tr>
<td>Black or African American</td>
<td>16,500 (23.9)</td>
<td>4,500</td>
<td>5,200</td>
<td>4,200</td>
<td>200</td>
</tr>
</tbody>
</table>

GENOMIC CHARACTERISTICS

Figure 3. Genetic characteristics of the NSCLC tumors in the genomic database are largely consistent with prior studies in large populations, including the TCGA. As expected, driver mutations (EGFR, ALK, ROS1, MET, BRAF, RET, or BRAF) and targetable driver mutations (GSK3A, AKT, ROS1, MET, BRCA, ERBB3, or ERBB2) were more associated with younger age, female gender, and non-smoking status.

Figure 4. Representation of integrating clinical and genomic features from the database.

RESULTS

CLINICAL AND GENOMIC PROGNOSTIC IMPLICATIONS

Figure 5. Kaplan-Meier overall survival analysis from initial diagnosis recapitulates known relationships with clinical and genomic features in NSCLC. Advanced stage, older age, a history of smoking, and presence of TP53 and KRAS mutations were associated with worse OS. The presence of a targetable driver was associated with a higher OS, with variation among the specific driver subtypes.

TESTING AND THERAPEUTIC RESPONSE PREDICTION

Figure 6. Using the CGDB to understand and predict response to therapy. The presence of an ALK or ROS1 fusion is associated with increased response to ALK/ROS1 inhibitor therapy (Nivolumab and Pembrolizumab). The inclusion of clinical features in the model improves the predictive accuracy of the model for patients receiving treatment in the database.

FUTURE DIRECTIONS

GROWTH OF THE GENOMIC DATABASE

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

FUTURE APPLICATIONS

• We have built a de-identified, HIPAA-compliant, real-world clinical-genomic database by linking longitudinal clinical data with high-resolution genomic information. The dataset consists of 25,000 cases, more than 20,000 total cases, and is both growing and updated on a quarterly basis.

• The clinical-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) genomic predictors 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., as seen in trials across controls).

CONCLUSIONS

Development and Validation of a Real-World Clinico-Genomic Database

Gaurav Singal1, Peter Grant Miller1, Vineta Agarwal1, Je He1, Anaka Gossay2, Gerald Lit3, Shannon Frank4, David Bourque5, Bryan Bosser6, Thomas Carton7, Ezra Bzybad8, Kathi Siddi-Radhik9, Ivan Ivanov10, Alex Parker1, Ansel Gura1, Gennadi Michael, Frighton1, Ann Jaskiew,1, Dana Feuchtbaum, Nathan Colemann, Amy Piak, Barterb11, Vincent A. Miller1

1Foundation Medicine, Inc., Cambridge, MA; 2Flatiron 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, registries, 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.

Methodology

• Oncology patients from community practices were identified for whom Flatiron EHR abstraction and Foundation Medicine tumor 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 public.

• Currently there are 21,394 non-small cell lung cancer (NSCLC) cases, which were used as a validation set for the database.

Results

• The CGDB covers 38 tumor types and continues to grow and receive updates on a quarterly basis.

• We have built a de-identified, HIPAA-compliant, real-world clinical-genomic database by linking longitudinal clinical data with high-resolution genomic information. The dataset consists of 25,000 cases, more than 20,000 total cases, and is both growing and updated on a quarterly basis.

• The clinical-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) genomic predictors 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., as seen in trials across controls).

Conclusions

Figure 3. Schematic of generation of CGDB (left) and cohort selection (right).

Figure 4. Representation of integrating clinical and genomic features from the database.

Figure 5. Kaplan-Meier overall survival analysis from initial diagnosis recapitulates known relationships with clinical and genomic features in NSCLC. Advanced stage, older age, a history of smoking, and presence of TP53 and KRAS mutations were associated with worse OS. The presence of a targetable driver was associated with a higher OS, with variation among the specific driver subtypes.

Figure 6. Using the CGDB to understand and predict response to therapy. The presence of an ALK or ROS1 fusion is associated with increased response to ALK/ROS1 inhibitor therapy (Nivolumab and Pembrolizumab). The inclusion of clinical features in the model improves the predictive accuracy of the model for patients receiving treatment in the database.