Impact of sample pretesting in a high through-put genotyping facility

B Marosy, C Boehm, B Craig, J Romm, C Ongaco, M Ziliak, M Adams-Carr, Y Osimouk, I McMullen, J Zhang, M Jewell, K Hetrick, H Ling, EW Pugh, KF Doheny  Center for Inherited Disease Research (CIDR) and Genetic Resources Core Facility (GRCF) SNP Center, IGM, JHUSOM, Baltimore, MD

Introduction:
The Center for Inherited Disease Research (CIDR) was established at the Johns Hopkins University in 1996. CIDR provides high quality genotyping services and statistical genetics consultation to investigators working to discover genes that contribute to common disease. Providing high quality data is a continued priority in the current era of Genome Wide Association studies (GWAs) and publicly available data. DNA samples received by CIDR undergo pre-testing evaluation (if funding supports this step) prior to production genotyping. This pre-testing allows for samples and corresponding phenotypic data to be confirmed and included. Problems reported at the pre-testing phase were categorized based on workflow. The data presented here was collected on sample problems (LIMS) to collect information and data on samples at each phase of the production lab processing and data release process.

Table 1. Summary of pre-testing outcomes for GWAs and non-GWAs projects

<table>
<thead>
<tr>
<th>GWAs projects</th>
<th>non-GWAs projects</th>
<th>Total samples pretested</th>
<th>41,748</th>
<th>21,582</th>
<th>532</th>
<th>352</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total samples pretested</td>
<td>41,748</td>
<td>21,582</td>
<td>532</td>
<td>352</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total samples not fixed</td>
<td>1867</td>
<td>4%</td>
<td>10.7%</td>
<td>822</td>
<td>3.0%</td>
<td>8.6%</td>
</tr>
<tr>
<td>Problems fixed or dropped</td>
<td>1552</td>
<td>3.7%</td>
<td>10.4%</td>
<td>621</td>
<td>2.9%</td>
<td>6.1%</td>
</tr>
<tr>
<td>Problems not fixed or not able to be evaluated</td>
<td>124</td>
<td>0.3%</td>
<td>0.6%</td>
<td>191</td>
<td>0.9%</td>
<td>4.2%</td>
</tr>
<tr>
<td>Total new problems at project release</td>
<td>939</td>
<td>0.0%</td>
<td>0.8%</td>
<td>487</td>
<td>2.2%</td>
<td>9.3%</td>
</tr>
</tbody>
</table>

Discussion:
Pre-testing has proven to be helpful in providing high quality genotyping data by validating sample information, evaluating sample quality and minimizing costs. As large scale studies continue and new technologies emerge, CIDR is committed to continuing stringent quality control of data and development of strategies to lower costs and provide high quality data.

Pre-testing allows for:
- Confirmation of sample fitness – the sample is what it is claimed to be. Gender matches, few or no family inconsistencies present (in the case of family studies, where family information is available) or no two samples are the same (identifying unexpected duplicates).
- Elimination of samples that are truly poor performers and will most likely not match in production.
- Replacement of samples with better performance.
- Minimization of costs associated with producing low quality data.
- Identification of any Mendelian inconsistencies – confirm parent-offspring pedigrees information for family studies.

For GWAs projects:
- Most problems (46.7%) found in GWAs projects were related to Mendelian inconsistencies.
- Problems fixed or dropped: 86.7%
- Gender inconsistencies had the least amount of problems (20.3%) for GWAs projects.
- Mendelian inconsistencies had the highest rate of correction (92.1%) for GWAs projects.
- Problems not fixed or not able to be evaluated: 4.3%
- Total new problems at project release: 1.6%

For non-GWAs projects:
- Most problems (67.6%) found in non-GWAs projects were related to poor performance.
- Problems fixed or dropped: 76.8%
- Gender inconsistencies had the least amount of problems (20.3%) for non-GWAs projects.
- Mendelian inconsistencies had the highest rate of correction (92.1%) for non-GWAs projects.
- Problems not fixed or not able to be evaluated: 4.3%
- Total new problems at project release: 1.6%

Table 2. Summary of problem types and investigator responses

For GWAs projects:
- Problems fixed or dropped: 86.7%
- Gender inconsistencies had the least amount of problems (20.3%) for GWAs projects.
- Mendelian inconsistencies had the highest rate of correction (92.1%) for GWAs projects.
- Problems not fixed or not able to be evaluated: 4.3%
- Total new problems at project release: 1.6%

For non-GWAs projects:
- Problems fixed or dropped: 76.8%
- Gender inconsistencies had the least amount of problems (20.3%) for non-GWAs projects.
- Mendelian inconsistencies had the highest rate of correction (92.1%) for non-GWAs projects.
- Problems not fixed or not able to be evaluated: 4.3%
- Total new problems at project release: 1.6%

Table 3. Summary of pretesting outcome and investigator responses by DNA source

Table 4. Summary of sample performance by project

Table 5. Summary of cost savings

Discussion:
- Elimination of samples that are truly poor performers and will most likely not match in production.
- Minimization of costs associated with producing low quality data.
- Identification of any Mendelian inconsistencies – confirm parent-offspring pedigrees information for family studies.
- Replacement of samples with better performance.
- Minimization of costs associated with producing low quality data.
- Identification of any Mendelian inconsistencies – confirm parent-offspring pedigrees information for family studies.

For GWAs projects:
- Most problems (46.7%) found in GWAs projects were related to Mendelian inconsistencies.
- Problems fixed or dropped: 86.7%
- Gender inconsistencies had the least amount of problems (20.3%) for GWAs projects.
- Mendelian inconsistencies had the highest rate of correction (92.1%) for GWAs projects.
- Problems not fixed or not able to be evaluated: 4.3%
- Total new problems at project release: 1.6%

For non-GWAs projects:
- Most problems (67.6%) found in non-GWAs projects were related to poor performance.
- Problems fixed or dropped: 76.8%
- Gender inconsistencies had the least amount of problems (20.3%) for non-GWAs projects.
- Mendelian inconsistencies had the highest rate of correction (92.1%) for non-GWAs projects.
- Problems not fixed or not able to be evaluated: 4.3%
- Total new problems at project release: 1.6%

Table 2. Summary of problem types and investigator responses

For GWAs projects:
- Problems fixed or dropped: 86.7%
- Gender inconsistencies had the least amount of problems (20.3%) for GWAs projects.
- Mendelian inconsistencies had the highest rate of correction (92.1%) for GWAs projects.
- Problems not fixed or not able to be evaluated: 4.3%
- Total new problems at project release: 1.6%

For non-GWAs projects:
- Most problems (67.6%) found in non-GWAs projects were related to poor performance.
- Problems fixed or dropped: 76.8%
- Gender inconsistencies had the least amount of problems (20.3%) for non-GWAs projects.
- Mendelian inconsistencies had the highest rate of correction (92.1%) for non-GWAs projects.
- Problems not fixed or not able to be evaluated: 4.3%
- Total new problems at project release: 1.6%

Discussion:
- Elimination of samples that are truly poor performers and will most likely not match in production.
- Minimization of costs associated with producing low quality data.
- Identification of any Mendelian inconsistencies – confirm parent-offspring pedigrees information for family studies.
- Replacement of samples with better performance.
- Minimization of costs associated with producing low quality data.
- Identification of any Mendelian inconsistencies – confirm parent-offspring pedigrees information for family studies.

For GWAs projects:
- Most problems (46.7%) found in GWAs projects were related to Mendelian inconsistencies.
- Problems fixed or dropped: 86.7%
- Gender inconsistencies had the least amount of problems (20.3%) for GWAs projects.
- Mendelian inconsistencies had the highest rate of correction (92.1%) for GWAs projects.
- Problems not fixed or not able to be evaluated: 4.3%
- Total new problems at project release: 1.6%

For non-GWAs projects:
- Most problems (67.6%) found in non-GWAs projects were related to poor performance.
- Problems fixed or dropped: 76.8%
- Gender inconsistencies had the least amount of problems (20.3%) for non-GWAs projects.
- Mendelian inconsistencies had the highest rate of correction (92.1%) for non-GWAs projects.
- Problems not fixed or not able to be evaluated: 4.3%
- Total new problems at project release: 1.6%

Discussion:
- Elimination of samples that are truly poor performers and will most likely not match in production.
- Minimization of costs associated with producing low quality data.
- Identification of any Mendelian inconsistencies – confirm parent-offspring pedigrees information for family studies.
- Replacement of samples with better performance.
- Minimization of costs associated with producing low quality data.
- Identification of any Mendelian inconsistencies – confirm parent-offspring pedigrees information for family studies.

For GWAs projects:
- Most problems (46.7%) found in GWAs projects were related to Mendelian inconsistencies.
- Problems fixed or dropped: 86.7%
- Gender inconsistencies had the least amount of problems (20.3%) for GWAs projects.
- Mendelian inconsistencies had the highest rate of correction (92.1%) for GWAs projects.
- Problems not fixed or not able to be evaluated: 4.3%
- Total new problems at project release: 1.6%

For non-GWAs projects:
- Most problems (67.6%) found in non-GWAs projects were related to poor performance.
- Problems fixed or dropped: 76.8%
- Gender inconsistencies had the least amount of problems (20.3%) for non-GWAs projects.
- Mendelian inconsistencies had the highest rate of correction (92.1%) for non-GWAs projects.
- Problems not fixed or not able to be evaluated: 4.3%
- Total new problems at project release: 1.6%