The MURDOCK Integrated Data Repository (MIDR): An integrative platform for biomarker research

Swati Chakraborty, Colette Blach, Matthew Gardner, Brian McCurt, Meredith Nahm, Jessica D. Tenenbaum

Duke Clinical Research Institute, Duke Center for Human Genetics, Duke Translational Medicine Institute

Abstract

Background: The first phase of the MURDOCK Study (Measurement to Understand the Reclassification of Disease) aimed to establish a biorepository that can help predict disease progression and response to treatment in distinct disease areas. A scalable architecture was developed for importing large datasets containing molecular information and demographical data, in addition to clinical data.

Methods: This platform is designed as a central database and is implemented with semi-automated data integration. It presents user-configured graphical queries and can integrate diverse datasets collected from multiple studies. Data elements from each of the four original studies were categorized into canonical data elements through use of metadata. Data elements from each of the four original studies were categorized into canonical data elements through use of metadata. Data elements from each of the four original studies were categorized into canonical data elements through use of metadata. Data elements from each of the four original studies were categorized into canonical data elements through use of metadata.

Results

Study code | Disease area | Principal investigator | Number of subjects | OA | WLM | HCV | Cathgen
--- | --- | --- | --- | --- | --- | --- | ---
1 | Cardiovascular disease | John G. McHutchison, MD | 159 | 89 | 64 | 99 | 86
2 | Obesity / weight loss | Virginia B. Kraus, MD | 17 | 12 | 11 | 10 | 8
3 | Hepatitis C | Annetje J. Mehlman, MD | 1326 | 84 | 4 | 8 | 86
4 | Other | Laura P. Svetkey, MD | 3294 | 59 | 9 | 5 | 9

Table 1. Individual Horizon 1 study information.

Figure 1. Overlap among data elements.

A shows the full clinical dataset from the original studies, the data contributed to the MIDR as a subset, and common data elements across studies as a subset. B Among the 4 studies, common data elements are identified as pooled data elements between any number of individual studies.

Figure 2. The MURDOCK Integrated Data Repository.

Figure 3. Data queries.

Methods

The MURDOCK Study (Measurement to Understand the Reclassification of Disease) aims to establish a biorepository that can help predict disease progression and response to treatment in distinct disease areas. A scalable architecture was developed for importing large datasets containing molecular information and demographical data, in addition to clinical data.

Results

Study code | Disease area | Principal investigator | Number of subjects | OA | WLM | HCV | Cathgen
--- | --- | --- | --- | --- | --- | --- | ---
1 | Cardiovascular disease | John G. McHutchison, MD | 159 | 89 | 64 | 99 | 86
2 | Obesity / weight loss | Virginia B. Kraus, MD | 17 | 12 | 11 | 10 | 8
3 | Hepatitis C | Annetje J. Mehlman, MD | 1326 | 84 | 4 | 8 | 86
4 | Other | Laura P. Svetkey, MD | 3294 | 59 | 9 | 5 | 9

Table 1. Individual Horizon 1 study information.

Figure 1. Overlap among data elements.

A shows the full clinical dataset from the original studies, the data contributed to the MIDR as a subset, and common data elements across studies as a subset. B Among the 4 studies, common data elements are identified as pooled data elements between any number of individual studies.

Figure 2. The MURDOCK Integrated Data Repository.

Figure 3. Data queries.