Enhancing the Ontological Representation of Cancer Cells Based on Analysis of AML Patient Data

Lucas M. Serra, BS1, William D. Duncan, PhD1, Paul K. Wallace, PhD2, Alexander D. Diehl, PhD1

1Department of Biomedical Informatics, University at Buffalo, Buffalo, NY; 2Roswell Park Comprehensive Cancer Center, Buffalo, NY

Introduction

Hematologic malignancies are a burdensome disease for over 500,000 U.S. residents. Flow cytometry is an established method in diagnosing and examining these diseases. We previously developed the Cancer Cell Ontology (CCL) as an extension of the Cell Ontology to represent types of hematologic malignancies including acute myeloid leukemia (AML) based on immunophenotype. Our current work aims to check the utility and coverage of CCL by analyzing de-identified flow cytometry data from past patients at Roswell Comprehensive Cancer Center (RPCCC) using CCL to classify cell populations in AML found for particular patients.

Methods

The Python programming language and FlowPy2 package were used to extract marker data from 1300 de-identified AML patients treated at RPCCC and were used to create AML cell populations for use in a modified FlowCL algorithm. The FlowCL algorithm essentially runs a SPARQL query and ranking algorithm to match cell populations and their markers with corresponding ontological representations in the CCL. We assessed the percentage of cell populations that matched and the degree to which they matched.

Results and Discussion

A sample extracted cell type can have an ontology class of ‘MPO+/CD13+/CD117+ myelomonocytic lineage leukemic cell’ which can have a child class that is CD33+. We were able to find partial matches for virtually all patient-derived AML cell populations (93%, 1209/1300). However, there were few complete matches (2%, 26/1300). We attribute these results to differences in the antibody panels used to characterize AML. Our ontology is based upon panels from the European Group for the Immunological Characterization of Leukemias (EGIL) which do not exactly align with panels used at RPCCC. We have used these results to develop over 100 additional cancer cell type definitions in CCL, most of which are either subclasses or siblings of EGIL-based classes. The lower half of Figure 1. shows how we intend to use the enhanced CCL as part of an automated system based on the ImmPort Galaxy service to perform automated classification of a larger set of patient data files.

Conclusion

Using patient-derived AML data, we have improved our past ontological descriptions of cancer cells and provide a strong foundation for a larger automated, analytic pipeline based upon pre-existing tools and the ImmPort Galaxy service in order to diagnosis hematologic malignancies.

References

2. FlowPy (http://flowpy.wikidot.com)