Integrative analytics for precision medicine

Senevirathne S[1], Roddy A[1], Black J[1], Gilmore A[1] and McArt DG[1]

[1]Centre for Cancer Research and Cell Biology, Queens University Belfast, Ireland

Background

Modern 'omics' technologies have accelerated our abilities to profile the cancer landscape, in turn, creating a data lake of biological information. These new 'big data' generating mechanisms are still being treated with old analytical methods. Data being analysed in isolation and on an individual basis has hampered our ability to truly grasp new markers and therapies. Here, we feel, a more integrated and dynamic framework would allow new insights and key hypotheses to be investigated. In order to craft a spatial and temporal analytical platform we designed a dynamic and interactive suite that will place discovery back in the hands of the researcher.

Methods

The Cancer Integromics Research Application Framework (CIRAFm) was designed to emulate a user-friendly interface front-end tied to a robust architecture that fuses several cross-platform coding technologies. It was developed to appreciate a modularised framework where we can create an 'app-store' of key software to assist discovery, and sits astride of two NoSQL based database management systems, to improve efficiency and minimise data redundancy within the system. Module developments have begun to explore alignment-free applications on next generation sequencing (NGS) data allowing us to study spatial and temporal heterogeneity where key drivers can be analysed by assisting software deploying genetic algorithms to build correlative marker associations. Here, we can use the new embedded information to mine external platforms through a novel domain specific language (DSL).

Results

Unburdened by legacy pipelines and data, the framework utilises a 'plug and play' architecture orchestrated through the Angular framework with interactive and dynamic data visualisation techniques, such as D3.js, for assisted marker discovery. With such information embedded, we can use novel tools to explore new markers in large data that displays sample heterogeneity. Alignment-free phylogenetics of NGS data has revealed capabilities to display the true sequence evolution in the data. This refines new contrasts, creates markers of interest and builds connections to therapeutic compounds that could be targeted as a consequence via the DSL to our externally developed drug discovery software, QUADrATiC.

Conclusions

Integromics will be paramount in future stratified scientific endeavours where new insights are required to help define new treatment strategies. Flexible and robust models of architecture provide an effective and efficient framework for emerging research on our ever increasing search for precision medicine.

Conflict of Interest: the authors declare no conflicts of interest

Funding Acknowledgements: None