

A12: Using Hilbert curves to visualize structural variations with Meander

Georgios A. Pavlopoulos*, Alejandro Sifrim, Jan Aerts

Faculty of Electrical Engineering - ESAT/SCD, Katholieke Universiteit Leuven,
Leuven, Belgium

Contact: georgios.pavlopoulos@esat.kuleuven.be

Background: In the past several years, the interest for browsing and analyzing structural variations in a genome increases exponentially. While sequencing techniques always improve and become cheaper over time, vast amounts of data get continuously produced. The analysis, the interpretation and the visualization of such an overload of information emerges and still remains a bottleneck.

Methods and Results: In this article we present Meander as java standalone application to visualize read-depth genome coverage. Meander uses a 2D plane of $512 \times 512 = 262.144$ pixels. DNA is computationally split into 262.144 buckets, each one holding the average coverage among the nucleotides that belong to the specific bucket. Each pixel on the Hilbert curve represents the read-depth coverage of the bucket. The coverage value is initially mapped to a color gray scale. Thus, the higher the coverage is, the darker the pixel appears and vice versa. In addition to the Hilbert representation, Meander also uses a linear representation of 512 pixels in length to show read-depth coverage at a lower resolution using bar charts. Each pixel/bar represents the average coverage of the nucleotide of each of the 512 buckets. The higher the bar height is, the higher the coverage. Meander is able to visualize data both linearly and as Hilbert curve at 5 different zoom-levels. For higher performance, files holding information about the read-depth coverage at different resolutions are pre-calculated. To directly compare the coverage between two genomes (sample vs reference) we use the log ratio. We use two different color schemes (red-green and blue-yellow) to visualize the ratio. Given a range of structural variations as they are calculated by external software applications, Meander is able to overlay such information by highlighting the areas on the Hilbert curve using rich color schemes.

Conclusions: Meander is tool to visualize read-depth coverage information and investigate inter-chromosome structural variations (currently deletions and duplications). Users are able to directly compare genomes among each other and overlay structural variations predicted by other software. It is a highly interactive standalone java application that utilizes a 2D Hilbert space to visualize the genome coverage. Currently it supports browsing at 5 different zoom and resolution levels. While Meander is currently limited to visualize intra-chromosomal variations we aim to support whole genome browsing using both read-depth and split read information. Meander will get further extended to support both balanced and unbalanced structural variations.

B35: Clinical Data Miner - an Electronic Data Capture software framework that improves interrater agreement

Arnaud Installé*, Dirk Timmerman, Thierry Van den Bosch, Bart De Moor

ESAT - SCD, Katholieke Universiteit Leuven, Belgium

Contact: arnaud.install@esat.kuleuven.be

Background: Clinical trials are used to increase medical knowledge. To that end, recorded variables must be of sufficient quality. For variables that are collected based on the interpretation of imaging modalities, this can pose a problem. Indeed, the meaning of variables can be interpreted differently by different clinicians potentially from different educational backgrounds. We implemented an Electronic Data Capture (EDC) software framework that allows to add pictograms as \hints\ to Case Report Forms (CRFs), enabling a more consistent interpretation of the variables, leading to better interrater agreement. To complement this, building on this software framework we implemented a module that can be used to conduct interrater agreement studies based on a set of imaging modalities.

Methods: The software framework was developed in Java, using software development best practices such as Test-Driven Development and Continuous Integration. Currently, a web-based user interface is available. However, its architecture is highly modular, enabling different user interface implementations, such as desktop or smartphone interfaces.

Results: The framework's high modularity allows not only changing the user interface, but also integration in hospital IT environments, or to define other types of studies than clinical trials or interrater agreement studies. The framework has been used to conduct two interrater agreement studies, which showed that showing pictograms in a CRF indeed does affect interrater agreement. It is also currently used to conduct the multi-centric studies conducted by the International Endometrial Tumor Analysis group.

Conclusions: The possibility to show pictograms as “hints” in CRFs, as well as the ease with which interrater agreement studies can be conducted in this software framework provides clinicians with all necessary tools to improve data quality in the evaluation of imaging modalities. Its high modularity allows its repurposing for a wide range of contexts, be they clinical or academic.