Nanopore Guided Regional Assembly

Eleni Adam*, Desh Ranjan*, Harold Riethman†
*Department of Computer Science, Old Dominion University
†School of Medical Diagnostic & Translational Sciences, Old Dominion University

INTRODUCTION

• The telomeres are the “caps” of the chromosomes and their vital role is to protect them.
• The integrity of our telomeres controls our health span: Telomere dysfunction caused by a various telomere rearrangements can be fatal for the cell.

• Telomeric and subtelomeric regions are hard to investigate because: (i) the current technology cannot provide their complete sequence, instead the DNA is given in multiple pieces and (ii) the current methods of assembling the pieces of these regions are not accurate enough due to the complicated structure of the subtelomeres (highly repetitive and variable).
• NPGREAT: An accurate assembly method for the region of the human subtelomeres.

METHOD

The NPGREAT is a hybrid method. It utilizes two of the latest available data (pieces): Linked-Reads and ultralong Nanopore reads.

The NPGREAT consists of five main steps:

i. The input selection: The pieces of both datasets corresponding to the subtelomere region in question are selected with the use of the adjacent single-copy region. Specifically for the Linked-Reads dataset, the REXTAL computational method is used to create the set of short-read contigs.
ii. The Orientation, Order and Enhanced Correction of the short pieces (color red) is obtained by using the long pieces (color blue) as scaffolds, upon which the short pieces are mapped to.

iii. In the Region Extraction, the segments of the multiple long pieces that can be used to connect the short pieces, are extracted.
iv. In the Gap Filling step, all possible segments are taken into account and one is selected to fill each gap.

v. In the Combination step, the corrected short pieces are combined with the connector segments. The output is the subtelomere region of the chromosome (purple color).

RESULTS

✓ High percent identity with the hg38 reference
✓ High quality assembly of the human subtelomeres.

Enhanced Correction step:
Potential misassemblies within the short pieces due to deletions in Tandem Repeat regions are corrected with the long pieces. The nanopore sequence is used to fill the missing portion.

Contact Email: eadam002@odu.edu