AmpliconReconstructor: Integrated analysis of NGS and optical mapping resolves the complex structures of focal amplifications in cancer

Jens Luebeck[1,2], Ceyda Coruh[3], Siavash R. Dehkordi[2], Joshua T. Lange[4,5], Kristen M. Turner[5], Viraj Deshpande[2], Dave Pai[6], Utkrisht Rajkumar[2], Chao Zhang[1], Julie A. Law[3], Paul S. Mischel[5], Vineet Bafna[2]

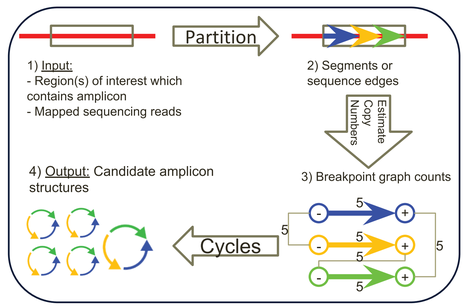
1) Bioinformatics & Systems Biology Graduate Program, UC San Diego, La Jolla, CA. 2) Department of Computer Science, UC San Diego, La Jolla, CA. 3) Salk Institute for Biological Studies, La Jolla, CA. 4) Biomedical Sciences Graduate Program, UC San Diego, La Jolla, CA. 5) Ludwig Institute for Cancer Research, UC San Diego, La Jolla, CA. 6) BioNano Genomics, San Diego, CA.

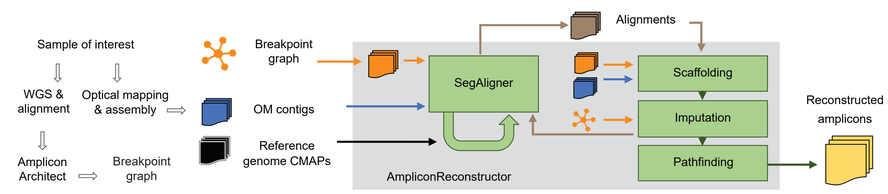
UC San Diego
JACOBS SCHOOL OF ENGINEERING
Computer Science and Engineering
UC San Diego
Bioinformatics & Systems Biology

Introduction

Copy number amplifications (CNA) are a hallmark of the cancer genome. The increase in copy number of oncogenes on focally amplified regions imparts positive selective pressure that mediates the rapid proliferation of cells. Presence of such focal amplifications has also been associated with genome instability and increased pathogenicity. Despite their importance, the mechanisms causing focal CNAs are incompletely understood. Proposed mechanisms include chromosomal translocation with duplication, chromothripsis and others. New reports suggest circular extrachromosomal DNA (ecDNA) exists in up to 40% of cancer types, and are an important driver for focal CNA. Thus, methods which reconstruct focal CNAs will enable a greater understanding of cancer biology.

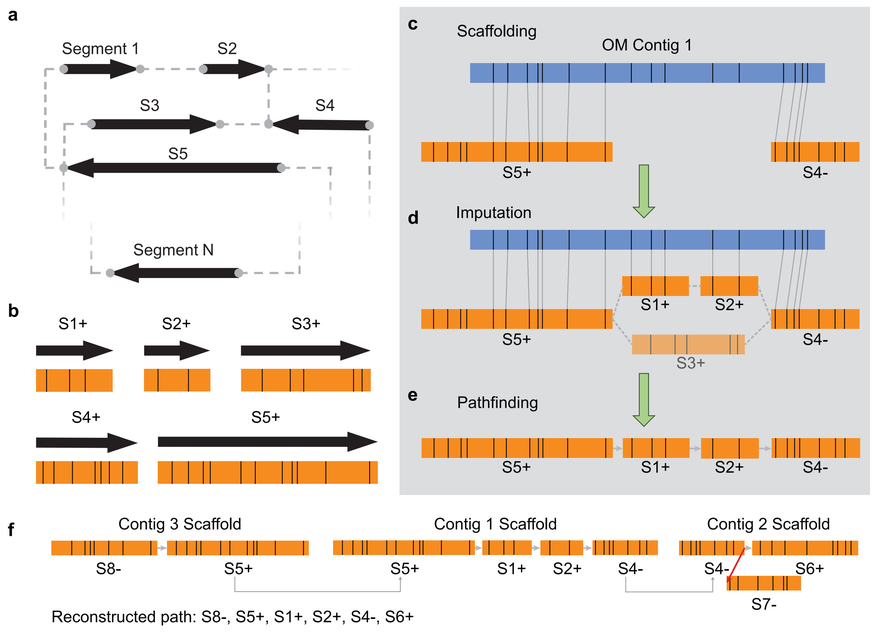
Breakpoint graph formation from NGS data with AmpliconArchitect





An earlier tool to analyze focal CNAs (Deshpande, et al. 2019), used next-generation sequencing (NGS) data to create a graph encoding rearrangement breakpoints, as a prelude to identifying the full structure. Paths and cycles extracted from breakpoint graphs provide the signatures of the rearrangement events, but are complex and rarely admit an unambiguous structure due to the complexity of focal CNAs. Here, we present a method, Amplicon Reconstructor (AR). AR employs a graph-based method to identify long paths and cycles in a breakpoint graph. We show reconstructions for a few paradigmatic examples of focal CNA, including a Philadelphia chromosome, circular ecDNA, and we provide the most comprehensive reconstruction of a breakage fusion bridge to date.

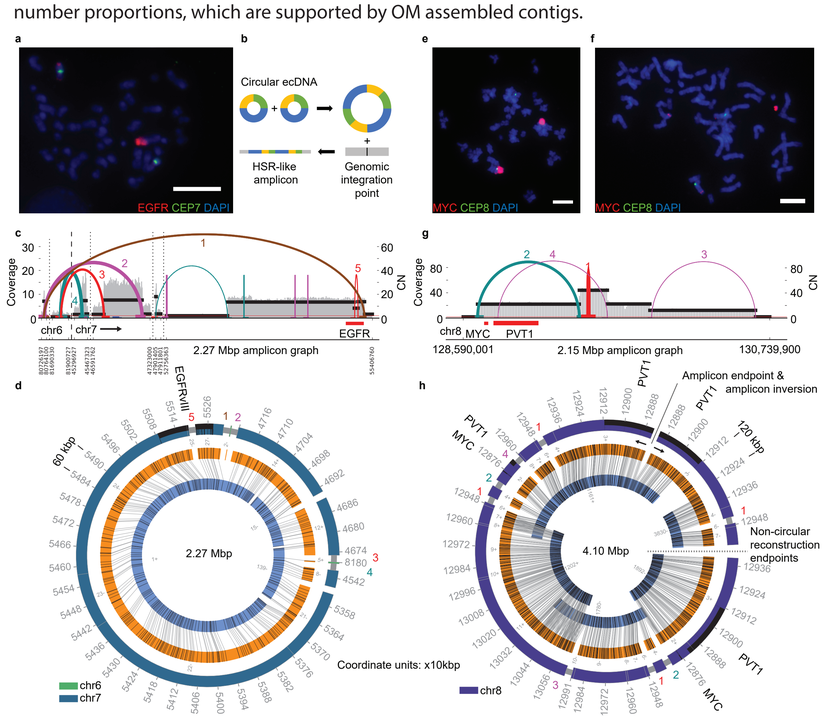
AmpliconReconstructor creates a graph of linked scaffolds from an NGS-derived breakpoint graph and long-range Bionano optical mapping (OM)



1. Turner KM, Deshpande V, Beyter D, et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. Nature. 2017;543(7643):122-125.
2. Deshpande V, Luebeck J, Nguyen ND, et al. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. Nat Commun. 2019;10(1):392.
3. Wu S, Turner KM, Nguyen N, et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. Nature. 2019;575(7784):699-703.
4. **Luebeck J, Coruh C, Dehkordi SR, et al. AmpliconReconstructor: Integrated analysis of NGS and optical mapping resolves the complex structures of focal amplifications in cancer. bioRxiv. 2020.**

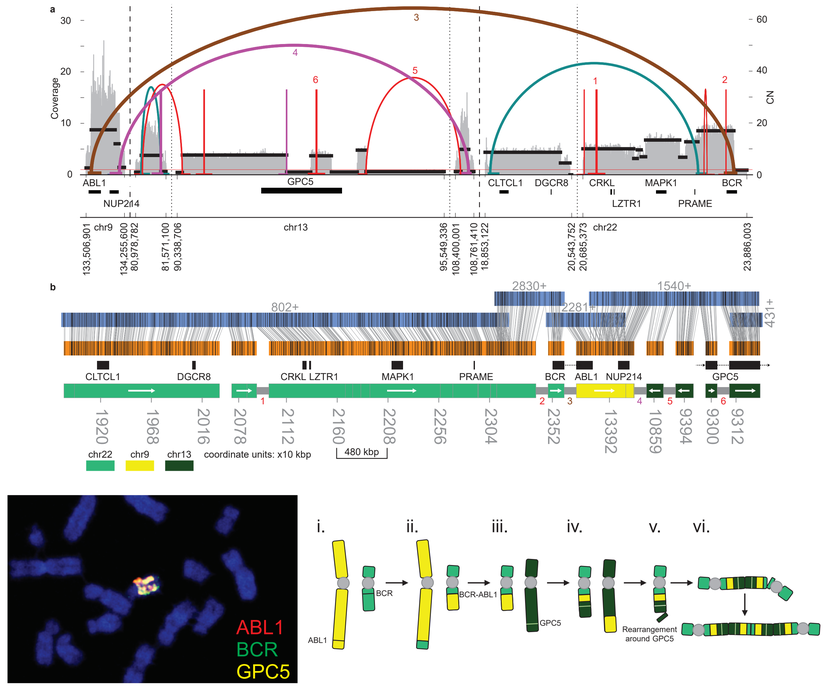
Reconstructing circular extrachromosomal DNA

We show two reconstruction examples below for circular ecDNA, amplicons from GBM39, a glioblastoma cell line bearing EGFR amplification, as well as for NCI-H460, a lung cancer cell line bearing MYC amplification. We show unambiguous reconstructions which explain the observed copy number proportions, which are supported by OM assembled contigs.



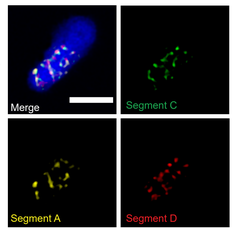
AR reveals the complex context of a BCR-ABL1 fusion

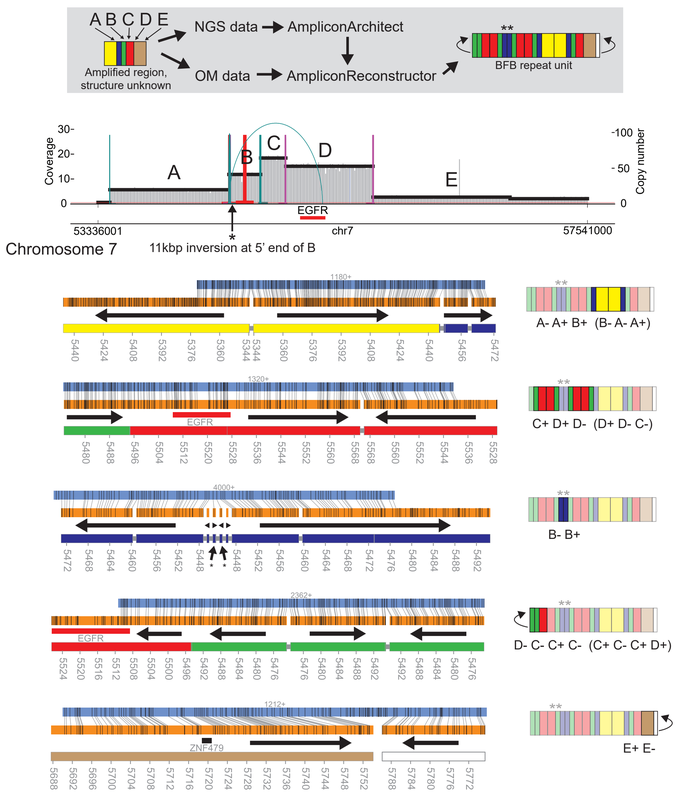
The K562 leukemia cell line contains a complex focal amplification bearing a BCR-ABL1 fusion. FISH verifies the presence of the BCR-ABL1 fusion, which is characteristic of the Philadelphia chromosome, a translocation rearrangement between the p-arms of chromosomes 9 and 22 which creates a highly oncogenic fusion protein. AR in concert with the breakpoint graph and OM scaffolding unravels a more complete picture of the complex focal amplification. AR's reconstruction joins segments from three chromosomes, representing the first time a Philadelphia chromosome has been assembled from sequencing-based and long range scaffolding techniques.



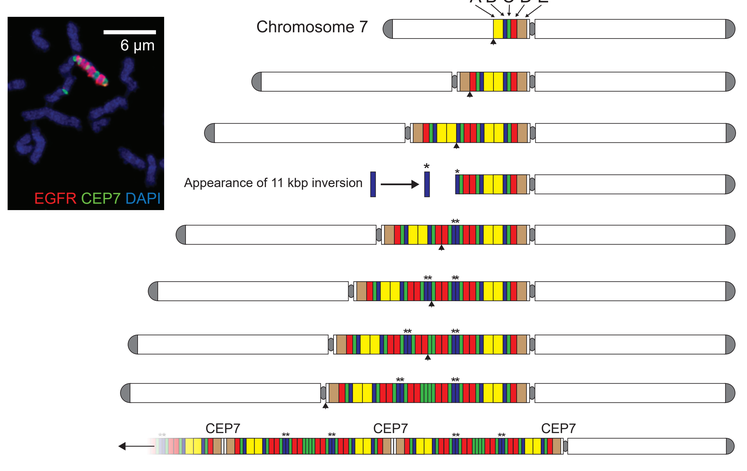
Unraveling breakage fusion bridge architecture

In the HCC827 lung cancer cell line, we demonstrate one of the most comprehensive reconstructions of a breakage fusion bridge (BFB) chromosome. First described by Barabara McClintock in the 1930's, the BFB phenomenon is characterized by telomere loss followed by chromosome end-joining and chromosomal breakage. BFBs result in an uneven chromatid pair forming palindromic subunits which may accumulate in high copy number. In HCC827, we demonstrate that the EGFR amplification is borne on a BFB structure through reconstruction of a BFB repeat unit.





Proposed mechanism of HCC827 BFB formation



Source code available on

GitHub

@Jens_Luebeck

AmpliconArchitect (AA)
https://github.com/virajbdeshpande/AmpliconArchitect

AmpliconReconstructor (AR)
https://github.com/jluebeck/AmpliconReconstructor

CycleViz
https://github.com/jluebeck/CycleViz