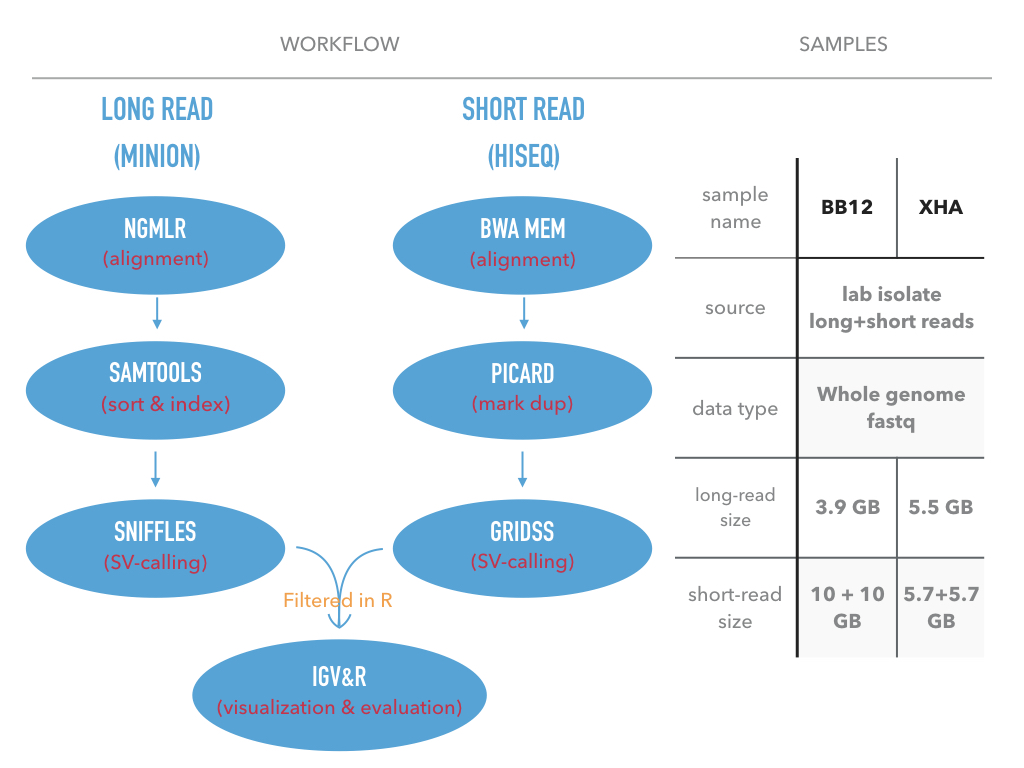
Structrual variant in Plasmodium falciparum samples

Litian Zhou

22/08/2018

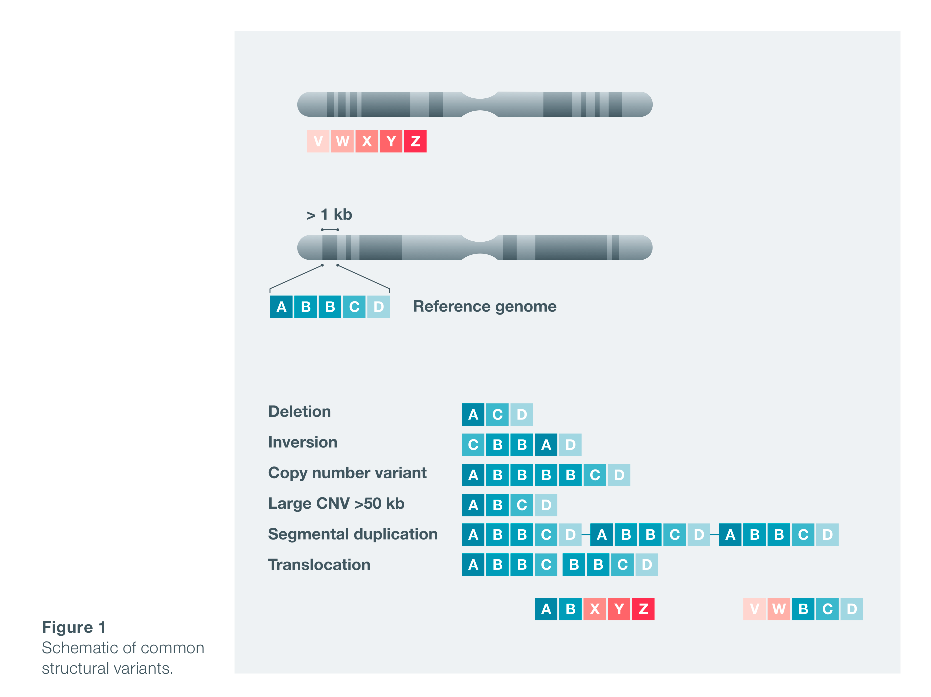
This is my work during my internship in Walter + Eliza Institute of Medical Research.  
In this document, I briefly introduce how to detect **structural variants (SVs)** from long-read data and validate those by short-reads.

The workflow and sample I use in this study: 

# Structrual variants in BB12

## What is structrual variants?

**Structural variation (SV)** refers to insertions, deletions, duplication, inversions,and translocations > 50 bp in length (Also includes large CNVs).



*paradigm from ONT*

## Table and snapshots:

The **structural variants (SV)** detection from **long-read** can give us more infomation than short-read. In this document, I use the SV detected from Plasmodium falciprum sample bb12. First, I overlap the SVs with nine popular genes. Second, I give some examples where long-read sequencing has advantages over short-read sequencing in terms of SV detection.

### Overlapping with genes

SVs in genes related with drug resistance and anti-folate resistance

## gene\_id gene\_name gene\_position  
## 1 PF3D7\_0304600 CSP Pf3D7\_03\_v3: 221,323 - 222,516(-)  
## 2 PF3D7\_0417200 DHFR Pf3D7\_04\_v3: 748,088 - 749,914(+)  
## 3 PF3D7\_0810800 DHPS Pf3D7\_08\_v3: 548,200 - 550,616(+)  
## 4 PF3D7\_0731500 EBA175 Pf3D7\_07\_v3:1,358,055 -1,362,929 (+)  
## 5 PF3D7\_1035300 GLURP Pf3D7\_10\_v3:1,399,195 - 1,402,896 (+)  
## 6 PF3D7\_1343700 K13 Pf3D7\_13\_v3: 1,724,817 -1,726,997 (-)  
## 7 PF3D7\_1343700 MDR1 Pf3D7\_05\_v3: 957,890 - 962,149(+)  
## 8 PF3D7\_1343700 MSP1 Pf3D7\_09\_v3: 1,201,812 - 1,206,974 (+)  
## 9 PF3D7\_1335900 TRAP Pf3D7\_13\_v3: 1,464,895 -1,466,619 (-)  
## gene\_function  
## 1 circumsporozoite protein  
## 2 Bifunctional dihydrofolate reductasethymidylate synthase  
## 3 Dihydropteroate synthetase  
## 4 erythrocyte bindingantigen-175  
## 5 glutamate-rich protein  
## 6 kelch protein K13  
## 7 multidrug resistanceprotein  
## 8 merozoite surface protein 1  
## 9 thrombospondin-related adhesion protein  
## reference sv\_type sv\_location sv\_length  
## 1 Weedall, G. D.,et al. (2007) INS within gene 32,56,76  
## 2 Nwakanma, D.C. et al. (2014) DEL upper stream 271  
## 3 Nwakanma, D.C. et al. (2014) DEL upper stream 52  
## 4 Baum, J et al.(2003) DEL + INS within gene 694 + 350  
## 5 Conway, D.J.(1997) DEL within gene 115+57+54  
## 6 Miotto, O. et al(2013) DEL down stream 33  
## 7 Nwakanma, D.C. et al. (2014) DUP cover all gene 95kbp  
## 8 Tetteh, K. K. A.et al. (2009) NOISE within gene 35kbp  
## 9 Weedall, G. D.,et al. (2007) DEL within gene 46 + 60  
## sample  
## 1 bb21 + xha  
## 2 xha  
## 3 xha  
## 4 xha  
## 5 bb12 + xha  
## 6 bb12 + xha  
## 7 bb12  
## 8 bb12  
## 9 bb12 + xha

#### Links to IGV visualization:

[CSP](pf_sv_pic/CSP_INS.png) [DHFR](pf_sv_pic/DHFR_DEL.png) [DHPS](pf_sv_pic/DHPS_DEL.png) [EBA175](pf_sv_pic/EBA175_DEL_INS.png) [GLURP](pf_sv_pic/GLURP_DEL.png) [K13](pf_sv_pic/K13_DEL.png) [MDR1](pf_sv_pic/MDR1_DUP.png) [MSP1](pf_sv_pic/MSP1_NOISE.png) [TRAP](pf_sv_pic/TRAP_DEL.png)

### Long-reads outperformance

chrom <- c(01,02,09,13,13)  
sv\_position <- c(463927,822520, 1241724, 754419, 872046)  
sv\_type <- c("DEL","DEL","DEL","DEL","DEL")  
sv\_length <- c(279, 482, "unprecise","unknown", "unprecise")  
sv\_info\_df <- data.frame(chrom=chrom, sv\_position=sv\_position, sv\_type=sv\_type, sv\_length=sv\_length)  
sv\_info\_df

## chrom sv\_position sv\_type sv\_length  
## 1 1 463927 DEL 279  
## 2 2 822520 DEL 482  
## 3 9 1241724 DEL unprecise  
## 4 13 754419 DEL unknown  
## 5 13 872046 DEL unprecise

#### Links to above SV visalization:

[1](pf_sv_pic/Pf3D7_01_v3-463,927_DEL.png) [2](pf_sv_pic/Pf3D7_02_v3-822,520_DEL.png) [3](pf_sv_pic/Pf3D7_09_v3-1,241,724_DEL.png) [4](pf_sv_pic/Pf3D7_13_v3-754,419_DEL.png) [5](pf_sv_pic/Pf3D7_13_v3-872,046-875,579_DEL_IGV.png)