1. **Assembly based detection of SV**
   1. Obtain a previously made ONT based de novo assembly: yeast\_pilon\_assembly.fasta.gz
   2. Align it using

*nucmer -maxmatch -l 100 -c 500 Saccharomyces\_cerevisiae.R64-1-1.dna.toplevel.fa yeast\_pilon\_assembly.fasta*

* 1. Compress delta.out :

*gzip delta.out*

* 1. Download the delta.out.gz
  2. Go to <http://assemblytics.com/> and upload the delta.out.gz
  3. What plots and summary statistics do you see?
  4. Download the results and upload it back to your working instance.
  5. Run: unzip user\_data
     1. Go into the folder and take the “my\_favorite\_organism.Assemblytics\_structural\_variants.bed” file.
     2. Convert this with SURVIVOR:

*SURVIVOR convertAssemblytics my\_favorite\_organism.Assemblytics\_structural\_variants.bed 0 assemblytics.vcf*

1. **Short read mapping based detection of SV**
   1. Obtain the previously generated mapped reads: illumina\_mapped.sort.bam
   2. Index the reference and mapped reads using:

*samtools index illumina\_mapped.sort.bam*

*samtools faidx Saccharomyces\_cerevisiae.R64-1-1.dna.toplevel.fa*

* 1. Now we are going to infer SVs using Delly:

*delly call –o illumina\_mapped.delly –g Saccharomyces\_cerevisiae.R64-1-1.dna.toplevel.fa illumina\_mapped.sort.bam*

* 1. Conver the output to VCF file:

*bcftools view illumina\_mapped.delly > illumina\_mapped.delly.vcf*

* 1. Obtain insight into what variants we identified.

SURVIVOR stats illumina\_mapped.delly.vcf -1 -1 -1 Illumina\_summary

* 1. Look at the individual output files.
     1. Illumina\_summary
     2. Illumina\_summary\_CHR

1. **Long read mapping based detection of SV**
   1. Obtain the Nanopore data set that was previously mapped. Ont\_mapped.sort.bam
   2. Index the mapped reads:

samtools index Ont\_mapped.sort.bam

* 1. Now we want to identify SVs using Sniffles:

sniffles -m Ont\_mapped.sort.bam -v Ont\_mapped.sniffles.vcf

* 1. Obtain insight into what variants we identified.

SURVIVOR stats Ont\_mapped.sniffles.vcf -1 -1 -1 ONT\_summary

* 1. Look at the individual output files.
     1. ONT\_summary
     2. ONT\_summary\_CHR

1. **Comparison of the SV calls obtained**
   1. As the last step today we want to compare the calls that we generated. Make sure you use the right paths and files! First we generate a file including the file names and paths.

*ls \*vcf > files*

* 1. Next merge the vcf files and generate a new file using:

*SURVIVOR merge files 1000 1 0 0 0 comparison.vcf*

* 1. Open the file using: less -S *comparison.vcf*
  2. We want to identify the overlap:

*perl -ne 'print "$1\n" if /SUPP\_VEC=([^,;]+)/' comparison.vcf | sort | uniq -c*

* + 1. What does that command do?
  1. Why are there so many singeltons in the short read vs. long read caller?
     1. *grep 'SUPP\_VEC=010;' comparison.vcf | perl -ne 'print "$1\n" if /SVTYPE=([^,;]+)/' | sort | uniq -c*
     2. *grep 'SUPP\_VEC=001;' comparison.vcf | perl -ne 'print "$1\n" if /SVTYPE=([^,;]+)/' | sort | uniq -c*
     3. What do you see?