**CAW SV benchmark**

**Jesper Eisfeldt**

**Introduction**

Three callers were evaluated, Manta, Delly and Lumpy. They were evaluated using the CreateTranslocations pipeline, through the statistics in the Manta article, and statistics from the TIDDIT article

Manta contains a tumor normal mode, it outputs one tumor vcf, and one normal vcf. The tumor vcf contains no germline variants

Lumpy do not have a proper tumor normal mode, instead one file is outputted, where the tumor and normal sample is simply treated as two different samples. Thus, the Tumor sample will still contain germline variants, a filter would be needed to clear the germline variants from tumor variants.

Delly calls the variants as two different samples. However, there is an inbuilt DELLY script that is run afterwards to remove germline variants from the tumor file.

Manta supports threading. Lumpy and delly\* do not support threading, however, the normal and tumor sample could be run separately to achieve some kind of parallelization.

\*Delly supports multiple threads, however only one core may be used per sample.

**Results**

**The caller output**

Manta outputs two VCF files, one tumor, and one normal file

Delly outputs one joint file per variant type; these variants are deletions, inversions, translocations and duplications. These files are bcf files.

The tumor filter is run on each separate bcf file.

Lumpy generates one vcf file, where tumor and normal is treated as two different samples, each having a column in the format field of the vcf file

**Variant detection**

Table 1 presents which type of variant each caller may detect. Manta and Lumpy detect the same type of variants. Delly is unable to detect intrachromosomal translocations(aka insertions). This table is copied from the TIDDIT manuscript

|  |  |  |
| --- | --- | --- |
| **Currently used structural variant callers** | | |
| **Caller** | **WGS signal** | **Reported variant types** |
| Delly (Rausch, 2012) | split reads and discordant pairs | tandem duplications, deletions, inversions, translocations |
| Manta (Xiaoyu Chen, 2015) | split reads and discordant pairs | deletions, insertions, inversions, tandem duplications, translocations |
| Lumpy | Split reads and discordant pairs | deletions, insertions, inversions, tandem duplications, translocations |

**Supplementary figure 1)** A summary of some available structural variant callers. These callers use different methods and signals within the WGS data to detect different types and sizes of structural variants.

**Time consumption**

Note, some of these CPU times are copied from the TIDDIT article.

Each caller was run on a single core, the callers were tested on multiple datasets.

|  |  |  |
| --- | --- | --- |
| **CPU hour consumption on SV calling of NA12878** | | |
| **Delly** | **Manta** | **Lumpy** |
| 30 | **3** | 45 |

**Table 2. calling on NA12878 on the UPPMAX cluster, TIDDIT article,**

Manta is by far faster than the other tools, Lumpy is 50% slower than Delly

Table 3 is copied from the Manta article. Here Delly is compared to Manta, Delly seems inefficient at calling in tumor normal mode. Manta requires the double amount of CPU hours two perform paired calling.

|  |  |  |
| --- | --- | --- |
| **HCC1954 tumor-normal calling** | | |
|  | **DELLY** | **MANTA** |
| CPU hours | 100 | **6** |

**Table 3. calling on HCC1954, Manta article**

HCC1954, as well as a simulated dataset was analysed in tumor normal mode using Lumpy and Manta. The simulated data contains chromosome 1,2,3 simulated to a depth of 20x. Manta is faster than Lumpy.

|  |  |  |
| --- | --- | --- |
| **CPU hour comparison LUMPY versus Manta** | | |
| **Data set** | **Lumpy** | **Manta** |
| Simulated data | 0.75 | 0.25(17 min) |
| HCC1954 | > 48 | 6 |

**Table 4. Lumpy vs manta, CPU hours on uppmax, Lumpy ran out of core hours after 2 days of runtime.**

**Precision and sensitivity**

Delly versus Manta(manta article)

|  |  |  |  |
| --- | --- | --- | --- |
| Delly vs Manta on HCC1954 | | | |
|  | Caller | **Sensitivity** | **Precision** |
| **Inversions (n=100)** | Manta | **0.67** | **0.35** |
|  | Delly | 0.66 | 0.32 |
| **Translocations (n=87)** | Manta | **0.84** | **0.271** |
|  | Delly | 0.32 | 0.179 |
| **Duplications 10k+ (n=60)** | Manta | 0.53 | **0.29** |
|  | Delly | **0.55** | 0.26 |
| **Deletions 10k+ (n=56)** | Manta | **0.6** | **0.26** |
|  | Delly | **0.6** | **0.26** |
| **Deletions [1k,10k) (n=12)** | Manta | 0.4 | **0.23** |
|  | Delly | **0.5** | 0.15 |

**Table 5. Delly vs manta, precision and sensitivity**

Lumpy was cancelled after 2 days of running due to the time limit of the sbatch job (more than 2 days of run time was considered unpractical). Instead, lumpy and manta was compared on the previously described simulated data (table 6). Lumpy seem to perform better than Manta(table 6).

|  |  |  |  |
| --- | --- | --- | --- |
| **Lumpy and Manta on simulated data** | | | |
| Caller | Hits (normal/tumor) | Called variants | (Sensitivity/precision) |
| Manta | 206/4714 | 5564 | 0.9/0.88 |
| Lumpy | 228/4893 | 5213 | 0.94/0.98 |

**Table 6. Benchmark using the CreateTranslocations pipeline. The sample contains 250 normal (germline) variants and 5200 tumor variants.**

**Conclusion**

Manta is by far the fastest caller, it has high precision and specificity, as well as a proper normal tumor mode. However, it only supports Paired ends having innie orientation.

Delly is slow, and is worse than Manta in almost any SV category. Its complex to implement, does not output standardized translocation format, and outputs bcf instead of VCF. Moreover, Delly is unable to detect intrachromosomal translocations.

The merge bcf function of delly does not seem to work properly; it exits with a segmentation fault.

Lumpy is slow, and it does not really do any proper tumor-normal analysis, it simply runs the tumour/normal samples as separate individuals. A tumour normal filter would be needed to separate the tumor from normal variants. On simulated data, Lumpy has both higher precsion and sensitivity than Manta.

It seems running CAW using Manta is the best approach. I will stay in touch if TIDDIT gets a tumor normal mode.