Paweł KUŚ*[1], Roman JAKSIK[1], Marek KIMMEL[2]

# Chapter 12. TUMOR SUBCLONAL RECONSTRUCTION PIPELINES – COMPARISON OF RESULTS

## 12.1. Introduction

Tumors arise from mutated single cells in the multistep evolutionary process, driven by two phenomenons: 1) the occurrence of new mutations, which can drive the tumor progression, and 2) the process of selection, which eliminates weak populations of cells and promotes the growth of the strongest ones. The resultant population is usually highly heterogeneous - even if the resulting evolutionary process has a branching, rather than a linear structure, old populations might not be entirely replaced by the new ones. This polyclonal nature of cancer carries a serious clinical impact: disease relapses and evolution of therapy-resistance [1], clonal cooperation [2] and, in general, poor prognosis. The knowledge of tumor subclonal structure might be the key to the development of effective, personalized cancer therapies and the improvement of patients' survival.

However, despite the recent advances in Next Generation Sequencing (NGS), the recognition of clonal structure of the tumor is still not a trivial task. Since the sequencing of single cells is still an expensive technology, struggling with the issue of accurate and uniform amplification of the whole genome, bulk sequencing remains the main method of tumor genotyping. Information provided by bulk sequencing is however averaged over the millions of cells and the distinction of subclonal populations, especially if they are small and share similar cellular prevalences, is therefore hard or even impossible.
Over the last decade, numerous algorithms aiming to resolve tumor heterogeneity from the bulk sequencing data have been created. Most of them define subclones by clustering together the detected single-nucleotide variants (SNVs) with similar allelic fre-

* Corresponding author: pawel.kus@polsl.pl, Akademicka 2A, 44-100 Gliwice, PL.
[1] Department of Systems Biology and Engineering, Silesian University of Technology.
[2] Department of Statistics, Rice University, Houston, USA.

quencies, usually correcting them according to the detected copy number variation (CNV) of the mutation loci. Next, they estimate the cellular frequencies of detected subclones, which might be followed by the inference of phylogenetic relationships between them. Unfortunately, until now no golden standard of tumor clonality analysis has been established [13]. Lack of independent and wide assessment of the existing approaches makes the methodology selection a hard process. In this work, we apply and compare the results of few pipelines, consisting of two different methods of copy-number changes detection and two most common algorithms for the reconstruction of tumor subclonal structure.

### 12.1.1. Overview of the chosen algorithms

Tests of the selected algorithms were performed on two datasets, for which multiple samples per patient were available. The first one included 4 female patients with HER2 positive breast cancer (BRCA), for which we performed Whole Exome Sequencing (WES) of two tumor samples (primary tumor and lymph node metastasis) along with the one normal sample. The second dataset, consisting of control-diagnosis-relapse triplets of samples from 11 patients with acute myeloid leukemia (AML) was retrieved from European Genome-phenome Archive (EGA) [4]. From both datasets we extracted 2 BRCA and 8 AML patients for which the number of detected high-quality non synonymous/intron/intergenic mutations did not exceed 3,000 to decrease the required computing time.

Two popular algorithms for the clonal reconstruction were tested: PhyloWGS [5] and PyClone [6], both supporting simultaneous analysis of multiple samples per patient and using SNVs as well as CNVs associate mutations into clones. In addition, PhyloWGS performs the evolutionary analysis of detected clones, generating and evaluating thousands of possible phylogenetic trees, which is not done by PyClone and might be performed using additional software like ClonEvol [7]. Two more algorithms were considered but were finally discarded for different reasons: SciClone [8]- which only clusters the mutations on the copy-number (CN) neutral regions, whereas CNVs were found common in our samples and CliP [9] - due to the lack of the multi-sample mode. Two methods of CNVs detection were used: FACETS [10] and TitanCNA [11]. They were chosen since both support the latest version of the reference genome (hg38), providing integer allele-specific copy numbers and easy integration with the

PhyloWGS, through the available CNV parser. We did not apply the popular GATK workflow for somatic CNV detection as it does not calculate integer CN.SNV and Indel mutations were called using Mutect2 [16].

Application of the two selected algorithms of clonal reconstruction to the mutation calls from two methods of CNV detection and one of SNV/Indel detection resulted in 4 pipelines being tested in this work (Fig. 12.1). In addition, as the NGS-based tumor purity estimates may significantly differ from the pathologist's assessments [3], to investigate the impact of purity estimations on the reconstruction results, all 4 pipelines were run twice: 1) with the purity estimates delivered by CNV callers and using default 100% purity only if the estimate was missing or below 0.1, and 2) assuming the perfect purity of all samples.

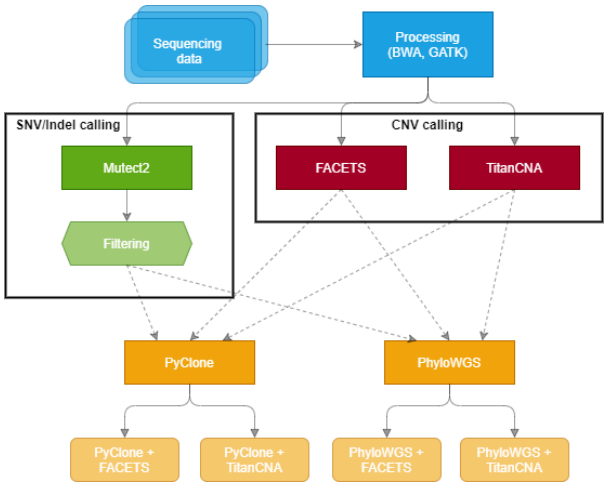


Fig. 12.1. Graph representing the 4 pipelines run in the study
Rys. 12.1. Graficzne przedstawienie 4 kombinacji metod zastosowanych w pracy

## 12.2. Results

### 12.2.1. CNV callers – comparison of FACETS and Titan calls

Both FACETS and Titan were run to call changes in total and minor allele copy numbers (Fig. 12.2). FACETS returned a single solution for each sample, Titan generated four different solutions across a set of given ploidies (2, 3) and numbers of clusters (1, 2) and successfully chose the optimal solution in all samples except those from the patient AMLRO-9, which except for some small regions was CN-neutral among non-sex

chromosomes according to FACETS. The computational time required by both methods was significantly shorter for FACETS than for Titan (data not shown).

For one patient (AMLRO-6) FACETS and Titan consistently showed different main ploidy (Titan: 3, FACETS: 2) and for 4 patients in both datasets (AMLRO-2 and 3, G30 and G31) Titan showed significantly different results for paired tumor samples, whereas all FACETS results for these patients were concordant.

Both algorithms were found prone to the detection of numerous, short CN calls, equally distributed along the genome, which after manual inspection should be classified as noise. The occurrence of these artifacts was more common in the FACETS results and among samples with limited or no true large-scale CN changes.

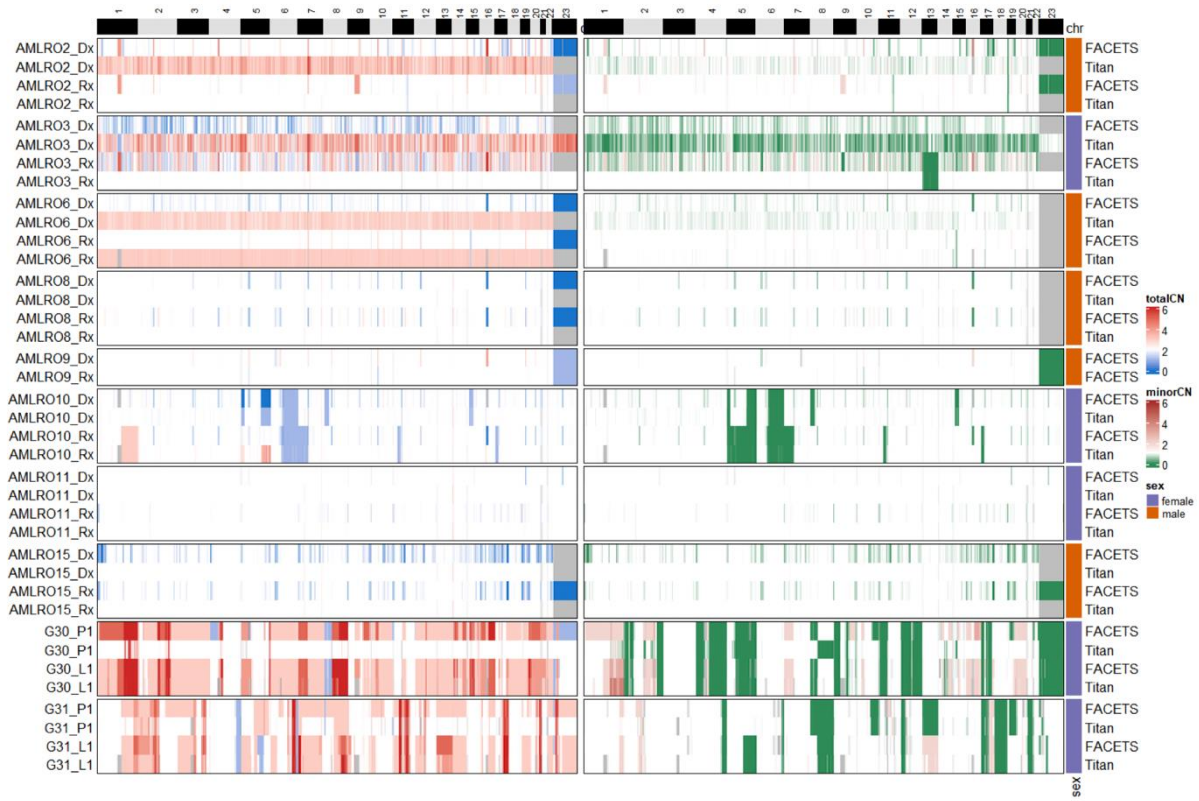


Fig. 12.2. Comparison of CNV calls obtained from FACETS and Titan. The left heatmap presents total CN changes, the right one changes in the minor allele CN. Gray – regions not reported by a caller. CNV callers are annotated on the right. Sample namings for AML: Dx – diagnosis, Rx – relapse; for BRCA: P1 – primary tumor, L1 – lymph node metastasis. Titan results for AMLRO-9 are missing, as it could not choose the optimal solution. Values above 6 were trimmed

Rys. 12.2. Porównanie wyników detekcji mutacji CNV otrzymanych za pomocą algorytmów FACETS i Titan

### 12.2.2. CNV callers - comparison of purity estimates

FACETS and Titan estimations of sample purities (Fig. 12.3) were significantly discordant and often incomplete or misleading. FACETS estimated purity of 14 out of 20 samples with values within the range from 0.16 to 0.81. In turn, Titan returned purity estimations of 18/20 samples, from which 8 were equal or almost equal to 0. These misleading values were associated with the nearly diploid state of samples and may impact the further analysis if propagated through the fully automated pipelines.

### 12.2.3. Clonal reconstruction - comparison number of clones

We found only the PhyloWGS to give a stable number of clusters, regardless of the dataset, CNV caller used, and the usage of estimated purity values or assuming them to be perfectly pure. In contrast, the number of clones returned by PyClone depended on all of these factors and in some cases differed by up to two orders of magnitude: starting from the few for AML dataset with FACETS calls and purity estimations,through dozens for the same dataset with other methods' combinations, up to the hundreds for BRCA dataset. Interestingly, when all purities were set to 1 the number of clones grew for the AML and decreased for the BRCA dataset. Therefore, we found only the PhyloWGS to be detecting a reasonable and predictable number of clones irrespective of the other methods used.

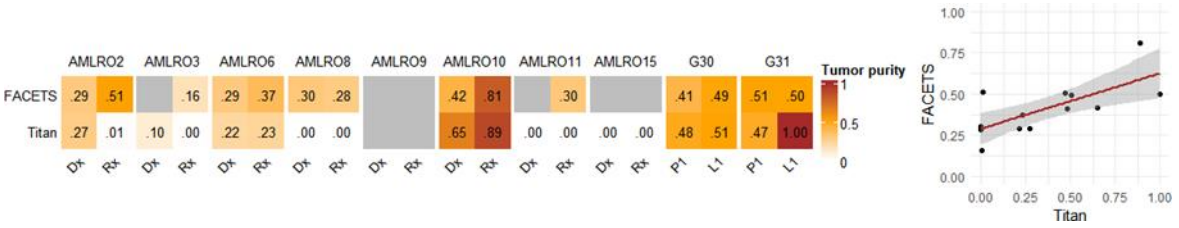


Fig. 12.3. Comparison of purity estimations from FACETS and Titan. Gray - missing estimation. Sample namings as in fig. 12.2. Pearson correlation coef. = 0.72, p = 0.0039

Rys. 12.3. Porównanie czystości próbek estymowanych za pomocą algorytmów FACETS i Titan

### 12.2.4. Clonal reconstruction - comparison of mutation to clone assignments

Pairwise similarities of the results of selected pipelines were measured using two different similarity measures: Adjusted Rand Index (ARI) and Adjusted Mutual Information (AMI), both returning negative or close to 0 values for independent classifications and equal to 1 for the perfect match (Fig. 12.4). The range of obtained

values varied among the patients,lower among such cases like G30,G31,and AMLRO2,and significantly visibly higher for others, e.g. AMLRO15 (Fig. 12.4a). Therefore, average index values might correspond to the clonal distinguishability or detectability feature of given tumors.

Pipelines were more convergent when the purity estimates from CNV callers were not used but all samples were assumed to be 100% pure (Fig. 12.4a vs 12.4b). In association with the differences in the number of detected clones (Sec. 2.3) it shows that both algorithms significantly differ in how they deal with impure samples. Interestingly, under the assumption of the perfect purities, PyClone results were less affected by the choice of the CNV caller than the results of PhyloWGS.

Direct comparison of the results of the same pipelines obtained with and without the purity estimates from CNV callers showed significant differences between PhyloWGS and PyClone (Fig. 12.4c). The former showed a moderate agreement of both runs, even for those cases where purity estimates were missing or faulty and were assumed equal to 1 in both runs. The results returned by the latter matched nearly perfectly in such cases but were almost independent for some other cases.Again,a potential explanation for that might be related to the different number of detected clones which we cannot explain.

Table 12.1

The numbers of clones detected by sets of algorithms. FAC - FACETS, NA – pipeline did not completed within 4 days or was not run due to missing CNV calls, * – Only the purities above 0.1 were used in the analysis. Purities below this threshold, as well as the missing values, were replaced with 1

| with purity estimation* | | | | | all tumor purities set 1 | | | |
|---|---|---|---|---|---|---|---|---|
| PyClone | | | PhyloWGS | | PyClone | | PhyloWGS | |
| patient | FAC. | Titan | FAC. | Titan | FAC. | Titan | FAC. | Titan |
| AMLRO2 | 3 | 3 | 7 | 6 | 33 | 21 | 6 | 6 |
| AMLRO3 | 2 | 13 | NA | NA | 14 | 16 | 8 | 6 |
| AMLRO6 | 2 | 2 | 6 | 4 | 7 | 6 | 7 | 7 |
| AMLRO8 | 2 | 16 | 7 | 7 | 18 | 16 | 5 | 5 |
| AMLRO9 | 4 | NA | 7 | NA | 4 | NA | 5 | NA |
| AMLRO10 | 2 | 5 | NA | 7 | 17 | 21 | 6 | 6 |
| AMLRO11 | 3 | 16 | 5 | 5 | 25 | 17 | 8 | 5 |
| AMLRO15 | 14 | 15 | 9 | 6 | 13 | 15 | 9 | 6 |
| G30 | 217 | 109 | 13 | 9 | 66 | 83 | 8 | 8 |
| G31 | 91 | 122 | 17 | 10 | 66 | 66 | 13 | 10 |

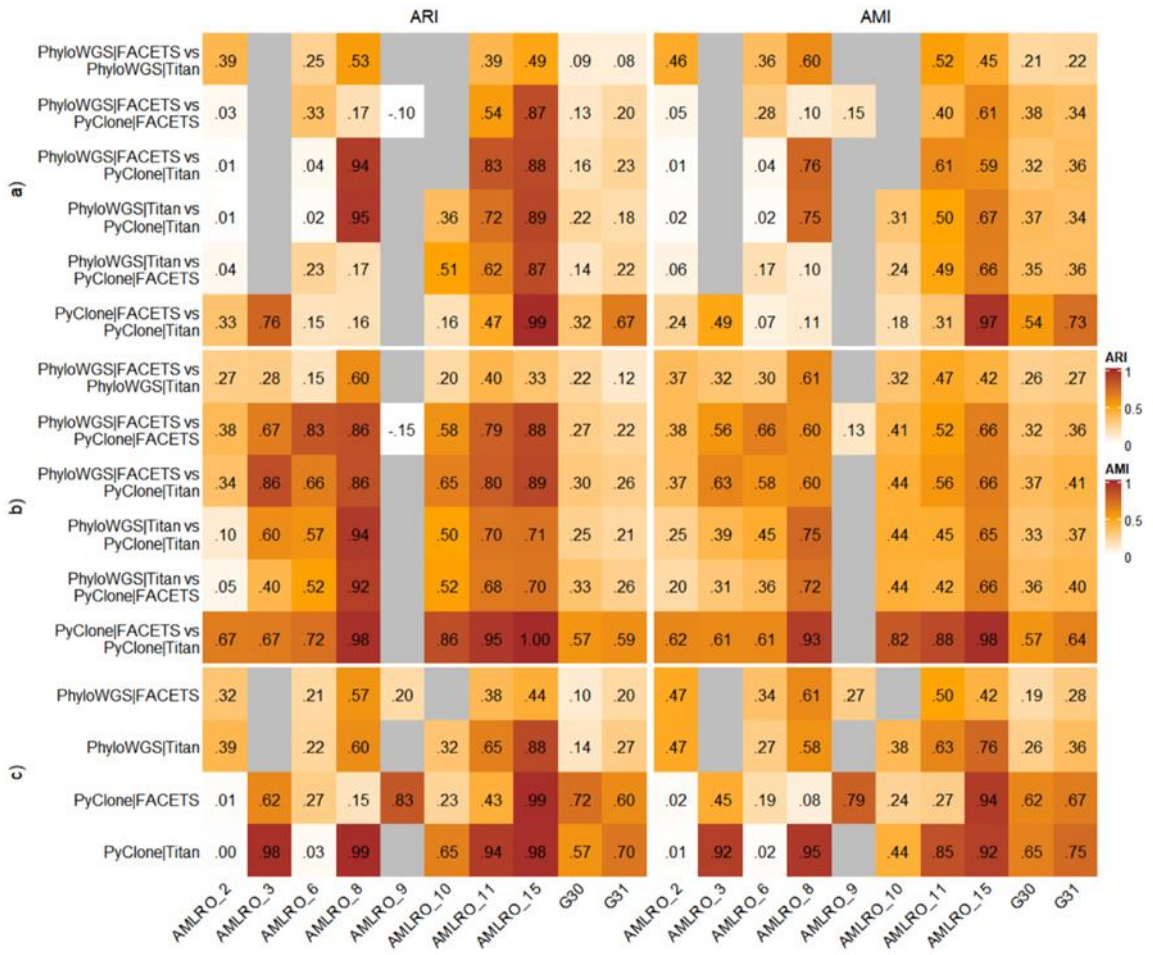


Fig. 12.4. The similarity of mutation-to-clone assignments by pipelines. Similarity measured by Adjusted Rand Index (ARI) and Adjusted Mutual Information (AMI). Panel a) presents results obtained with purity estimations from CNV corresponding, where only the values missing or below 0.1 were replaced with 1 (run 1). Results on panel b) were obtained with all required purity values set to 1 (run 2). Panel c) compares the results from both runs of the same combination of methods. Pearson coefficient of correlation of ARI and AMI = 0.92, $p < 2.2e\text{-}16$

Rys. 12.4. Heatmapa zgodności przypisań mutacji do klonów uzyskanych przez różne kombinacje metod

## 12.3. Discussion

In this study, we established and compared the results of 4 different pipelines consisting of Mutect2 for SNV/Indel calling, two CNV callers: FACETS and Titan, and two methods of clonal reconstruction: PhyloWGS and PyClone. In addition, all pipelines were tested with and without the use of purity estimations derived from CNV callers.

We found that change at any step of the analysis causes significant changes in the clonal reconstruction results. The level of similarity of different pipelines' results depended

more on the patient being analyzed than the pair of pipelines being compared. We found that pipelines using the same clonal reconstruction algorithm but different CNV caller were the most similar ones when PyClone was used and the most divergent in the case of PhyloWGS. This tendency was more visible when all samples were assumed to be perfectly pure than when we used the purity estimations from CNV callers.

Interestingly, we identified the significant effect of the purity estimation values used in the analysis, which we did not find reported in the previous studies. Especially in the pipelines involving PyClone change in the purity value had an unpredictable effect on the results, leading to nearly independent clusterization among some patients and only slightly affecting the results for the others. The number of clones detected by PyClone significantly differed if the purity estimations changed, in contrast to PhyloWGS which identified a similar and reasonable number of clones regardless of the used purity values. Limitations of this study involve the limited number of methods used and the lack of alternative SNV/Indel callers, which have been reported to be an additional source of variation [12, 13]. Another limitation of our work is the automatization of the workflow running the selected pipelines. The obtained results possibly could be improved by the manual inspection of eg. CNV calls, which could reduce the number of the noise and false CN calls. Unfortunately, the evaluation of accuracy provided by the wide spectra of available methods still needs to be done. To perform such evaluation the knowledge of the ground truth is needed, which requires the simulation of the synthetic data or usage of the real bulk sequencing data supported by single-cell sequencing.

## 12.4. Matherials and methods

**Materials.** Whole Genome Sequencing results for 7 men and 4 women with AML were downloaded as hg19-aligned BAM files from the EGA (ID: EGAD00001003234). One normal and two tumor samples (diagnosis/relapse) were sequenced from each patient. BRCA cohort consisted of 4 female patients diagnosed with HER2+ breast cancer, from whom one normal and two tumor samples (primary tumor and lymph node metastasis) were obtained. Libraries for in cohort were prepared using the Agilent SureSelect v6 kit and subjected to the Whole Exome Sequencing.

**Processing and SNV/Indel calling for AML data.** Aligned BAM files were processed using MarkDuplicates/BaseRecalibrator from GATK toolkit (v4.1.1.0) [17]. Somatic mutations were identified with MuTect2 and annotated using VEP [18]. Variants from both tumor samples were combined using GATK's CombineVariants tool. Finally, we removed the variants that failed to pass at least one of the filters in both samples.

**Processing and SNV/Indel calling for BRCA data.** Raw reads were aligned against human genome hg38 using BWA-MEM (v0.7.17) [15] and processed using MarkDuplicates/BaseRecalibrator from GATK (v4.1.4.0) [17]. Somatic mutations were called with Mutect2, filtered using FilterMutectCalls, and annotated using VEP.

**Detection of CNV changes.** Integer, allele-spicefic copy numbers were detected using FACETS [10] and Titan [11]. FACETS input files were prepared with the command *snppileup -g -q10 -Q10 -P100 -r25,0 -d10000 snp.vcf.gz pileup.csv.gz normal.bam tumor.bam, where* SNP file for *hg38 w*as downloaded from *ftp.ncbi.nlm.nih.gov/ snp/organisms/human_9606/VCF/00-common_all.vcf.gz* and for *hg19* from *ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606_b151_GRCh37p13/VCF/00-common_all.vcf.gz.* Then FACETS was run using run_facets.R script from aleighbrown/facets_snakemake GitHub repository [19] with nhet = 15. Titan was run using the snakemake workflow for Standard WGS/WES analysis available through the Titan GitHub repository.

**SNVs/Indels and cases filtering.** bcftools were used to remove the synonymous/intron/intergenic variants, variants with MMQ below 50 or average coverage in tumor samples below 10. Then only the patients with less than 3000 mutations were kept to shorten the computational time (<4 days for PhyloWGS).

**Analysis of tumor structure.** CNV calls and filtered SNVs/Indels were parsed to the PyClone input format using custom scripts. For PhyloWGS, snakemake workflow created by A.L Brown [14, 20] was used to prepare input files and run the algorithm with 20 MCMC chains.PhyloWGS write_results.py script was run with–include-multiprimary –max-multiprimary 1 options to prevent it from throwing an error when only multiprimary trees were created. The best tree was extracted using an approach of the default browser of PhyloWGS results (selection of the tree with the highest density).

**Results analysis and visualization.** All results were analyzed and visualized using R (v4.0.3) and RStudio. Adjusted Rand Index and Adjusted Mutual Information for pairs

of results were calculated using ARI and AMI functions from the aricode package (v1.0.0). Visualizations were prepared using ggplot2 and ComplexHeatmap packages.

**Acknowledgements**

This work was co-financed by the European Union through the European Social Fund (grant POWR.03.02.00-00-I029) and by Polish National Science Centre grant 2018/29/B/ST7/02550. Calculations were carried out using the computer cluster Ziemowit (https://www.ziemowit.hpc.polsl.pl) funded by the Silesian BIO-FARMA project No. POIG.02.01.00-00-166/08.

## Bibliography

1. Ch. Kim, R. Gao, E. Sei, N. Navin et al., Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing, Cell (2018).
2. K. Polyak, A. Marusyk, Cancer: Clonal cooperation, Nature (2014).
3 S. Haider, P.C. Boutros et al., Systematic Assessment of Tumor Purity and Its Clinical Implications, JCO Precision Oncology (2020).
4 L. Shlush, A. Mitchell, L. Heisler et al., Tracing the origins of relapse in acute myeloid leukaemia to stem cells Nature (2017) 547, 104–108.
5. A.G. Deshwar, C.K. Yung et al., PhyloWGS: Reconstructing subclonal composition and evolution from whole-genome sequencing of tumors, Genome Biol (2015).
6. A. Roth, J. Khattra, D. Yap et al., PyClone: statistical inference of clonal population structure in cancer, Nat Methods 11, 396–398 (2014).
7. H.X. Dang, C.A. Maher et al. ClonEvol: clonal ordering and visualization in cancer sequencing, Annals of Oncology (2017).
8. C.A. Miller et al., SciClone: Inferring Clonal Architecture and Tracking the Spatial and Temporal Patterns of Tumor Evolution, PLoS Computational Biology (2014).
9. K. Yu, S.J. Shin, H. Zhu, W. Wang et al., CliP: fast subclonal architecture reconstruction from whole-genome sequencing data (2018).

10. R. Shen, VE. Seshan, FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing, Nucleic Acids Res. (2016).
11. G. Ha, SP. Shah et al., TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data, Genome Res. (2014).
12. L.Y. Liu, V. Bhandari, A. Salcedo et al., Quantifying the influence of mutation detection on tumour subclonal reconstruction, Nat Commun (2020) 11, 6247.
13. A. Salcedo, S.M.G. Espiritu et al., A community effort to create standards for evaluating tumor subclonal reconstruction, Nat Biotechnol (2020).
14. N. Roper et al. Clonal Evolution and Heterogeneity of Osimertinib Acquired Resistance Mechanisms in EGFR Mutant Lung Cancer, Cell Reports Medicine (2020).
15. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv.org: p. arXiv:1303.3997 (2013).
16. K. Cibulskis, G. Getz. et al., Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples, Nature biotechnology (2013).
17. MA. DePristo, KV Garimella et al., A framework for variation discovery and genotyping using next-generation DNA sequencing data, Nature Genetics (2011).
18. W. McLaren, L. Gil, SE. Hunt, HS Riat, GR. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The Ensembl Variant Effect Predictor, Genome biology (2016).
19. GitHub, FACETS snakemake workflow: github.com/aleighbrown/facets_snakemake, accessed: 11 March, 2021.
20. GitHub, PhyloWGS snakemake workflow: github.com/aleighbrown/pwgs_snakemake, accessed: 11 March, 2021.

## TUMOR SUBCLONAL RECONSTRUCTION PIPELINES – COMPARISON OF RESULTS

### Abstract

Tumors arise in a complex evolutionary process, which leads to the development of a heterogeneous population of malignant cells, composed of distinct cell subpopulations called clones. This clonal heterogeneity has been associated with cancer relapse and

drug resistance, both leading to poor prognosis. Among the methods of clonal heterogeneity analysis important place belongs to the algorithms that use bulk DNA sequencing data to identify subclonal tumor populations. Unfortunately, although many such tools have been developed during the last decade, no gold standard of analysis has been yet established. Here we compare the results of 4 analysis pipelines consisting of two clonal reconstruction algorithms: PyClone and PhyloWGS, and two CNV callers: FACETS and Titan. We found, that aside from the known factors affecting results like the choice of the CNV detection method, inaccuracies of the sample purity estimations are another significant source of variation.

**Keywords:** clonal evolution, tumor heterogeneity, subclones, clonal reconstruction.