

Exploring Current Challenges and Perspectives for Automatic Reconstruction of Clonal Evolution

SARAH SANDMANN¹, SILJA RICHTER¹, XIAOYI JIANG² and JULIAN VARGHESE¹

¹*Institute of Medical Informatics, University of Münster, Münster, Germany;*

²*Department of Computer Science, University of Münster, Münster, Germany*

Abstract. *Background/Aim: In the field of cancer research, reconstructing clonal evolution is of major interest. The technique provides new insights for analysis and prediction of tumor development. However, reconstruction based on mutational data is characterized by several challenges. Materials and Methods: By performing extensive literature research, we identified 51 currently available tools for reconstructing clonal evolution. By analyzing two cancer data sets (n=21), we investigated the applicability and performance of each tool. Results: Seventeen out of 51 tools could be applied to our data. Correct clustering of variants can be observed for 4 patients in the presence of ≤ 3 clusters and ≥ 5 time points. Correct phylogenetic trees are determined for 10 patients. Accurate visualization is possible, by applying adjustments to the original algorithms. Conclusion: Despite bearing considerable potential, automatic reconstruction of clonal evolution remains challenging. To replace tedious manual reconstruction, further research including systematic error analyses using simulation tools needs to be conducted.*

According to the International Agency for Research on Cancer, more than 19 million new cases of cancer were registered in 2020. Furthermore, almost 10 million deaths were recorded (1). For many types of cancer, precise mutational characterization is of major interest: For Burkitt lymphoma (BL), for example, translocations affecting the *MYC* gene are considered a biological hallmark and essential in terms of diagnosis (2). Furthermore, recent data shows that relapse may be associated with a deficiency of *TP53* (3).

This article is freely accessible online.

Correspondence to: Sarah Sandmann, Institute of Medical Informatics, Albert Schweitzer-Campus 1, Building A11, 48149 Münster, Germany. Tel: +49 2518355270, Fax: +49 2518352259, e-mail: sarah.sandmann@uni-muenster.de

Key Words: Clonal evolution, clustering, tree reconstruction, variant integration, visualization.

Another example is myeloid neoplasia, including acute myeloid leukemia (AML) and myelodysplastic syndromes (MDS). These heterogeneous hematopoietic stem cell disorders are characterized by the acquisition of somatic alterations and clonal evolution (4-6). To optimize treatment selection, patients are stratified, among others, according to their genetic alterations (7). Recent data even shows clusters within lower risk MDS patients based on their mutational profiles and association of these clusters with clinical endpoints (8).

To determine genetic alterations, a variety of laboratory techniques is currently available. Small variants are usually investigated via bulk DNA analysis using Next-generation sequencing (NGS) techniques. Over the last 20 years, the costs for sequencing have been reduced significantly (9), allowing for not only targeted, but whole-exome sequencing (WES) and even whole-genome sequencing (WGS) data to be generated for one patient at several time points in the course of his/her disease. For the detection of large variants, like copy-number variants (CNVs), a variety of techniques is available: karyotyping, fluorescent in situ hybridization (FISH), SNP arrays, array-CGH and – with limitations – also NGS.

Over the last years, the detection of variants has been greatly optimized (10-12). However, for the combined evaluation of mutations and their development over time in terms of clonal evolution, many studies still rely on manual analyses (3, 13).

Currently, a wide set of algorithms is available performing automatic analysis of clonal evolution. These algorithms can be split according to the three main tasks in reconstructing clonal evolution: 1) clustering variants by cell fraction, 2) reconstructing the clonal evolution tree, 3) visualizing clonal evolution.

Analyzing two sets of well-characterized, publicly available real data, we performed a systematic evaluation of currently available tools for automatic reconstruction of clonal evolution, considering all three categories. Data set 1 covers 10 BL patients – 5 with and 5 without relapse (3). Data set 2 covers 11 MDS patients – 6 patients receiving supportive care and 5 patients treated with lenalidomide (13). In both data sets, cases of linear as well as branched evolution can be observed.

Table I. Overview of the two cancer data sets analyzed and their main characteristics.

	Data set 1	Data set 2
Disease entity	Burkitt lymphoma (BL)	Myelodysplastic syndromes (MDS)
Number of samples	10	11
Number of time points (median [min-max])	1.5 [1-2]	5 [5-30]
Number of clones (median [min-max])	9 [7-17]	5 [1-9]
Number of variants (median [min-max])	32.5 [17-117]	19 [11-28]
Clonal evolution mechanisms	Linear, branched dependent	Linear, branched dependent, branched independent

Materials and Methods

Data sets analyzed. Two cancer data sets were analyzed. Their main characteristics are summarized in Table I. Data set 1 covers 10 patients with BL – 5 patients with relapse (germline + 2 additional time points) and 5 patients without relapse (germline + 1 time point). For all patients, clonal evolution based on WES, targeted sequencing, Sanger sequencing, SNP array analysis (Infinium OmniExpressExome-8v1.3kit) and FISH has been reconstructed manually. Results on clonal evolution, as well as detailed variant calling information (including coverage and variant allele frequency – VAF), have been published (3), original FASTQ files of the WES experiments are available at the Sequence Read Archive (PRJNA561490).

The data set can be considered a typical, representative use-case for the application of a pipeline performing automatic reconstruction of clonal evolution: a variety of data sources is available. However, data has only been collected at a few time points. The number of clones is relatively high (up to 17). In 1 out of 10 cases, manual reconstruction of clonal evolution did not lead to a unique result on the basis of available data.

Data set 2 covers 11 patients with MDS – 6 patients receiving supportive care and 5 patients treated with lenalidomide. To reconstruct clonal evolution WES, targeted deep sequencing, karyotyping, CytoScan HD Array (Affymetrix) and FISH data have been integrated (13). Results on clonal evolution, as well as variant calling information (including coverage and VAF), are available along with the publication, FASTQ files of the NGS experiments are available at the Sequence Read Archive (PRJNA355124).

This second data set poses another potential challenge for automatic reconstruction of clonal evolution: A considerably higher number of time points (up to 30) is available compared to data set 1. Fewer clones are detected (up to 9), however, clonal evolution itself is highly complex. Cases with parallel dependent evolution, including several branches, as well as branched independent evolution are present.

Algorithms evaluated. We searched PubMed for algorithms connected to clonal evolution (clonal evolution AND algorithm). Additionally, we considered algorithms referenced in papers identified by PubMed search. We focused on approaches handling bulk DNA-seq data and – ideally – additionally considering information on CNVs. Any approach requiring single-cell data was not taken into account. Furthermore, algorithms just analyzing CNVs detected in *e.g.* array data were not considered.

Altogether, we identified 51 algorithms (Figure 1). These can be split into three categories according to the three main tasks in

reconstructing clonal evolution: 1) clustering of variants, 2) reconstructing the clonal evolution tree, 3) visualizing clonal evolution.

As a first step, variants detected in genomic data are clustered. Thereby, information on variants can originate from different sources, *e.g.*, NGS, array, FISH or karyotyping. Variants detected by NGS are characterized by coverage and VAF. Thus, the cancer cell fraction (CCF), *i.e.*, the percentage of cells affected by a mutation, is not directly available and has to be inferred, considering overlapping copy number variants (CNVs) and the most likely genotype. Variants detected by array data are – in case of SNP-arrays – characterized by rough estimates of the CCF based on observed changes in the Log R Ratio (LRR) and the beta allele frequency (BAF). These calls are lacking information on coverage. Variants detected by FISH and/or karyotyping are characterized by estimates of CCF, based on the number of evaluated cells.

Algorithms performing clustering ideally support variants characterized by VAF as well as variants characterized by CCF as input and automatically perform all necessary transformations. Altogether, we identified 17 algorithms performing clustering of variants. Ten out of 17 were, however, excluded due to reasons of unavailability, unresolvable error messages or having the wrong scope (see Figure 1, see Supplementary Information, section 1.1 for information on the precise reasons and – if applicable – the precise error messages). The remaining tools are CloneHD (14), clonosGP (15), DeCiFer (16), PyClone (17), PyClone-VI (18), QuantumClone (19) and sciClone (20). Additionally, we identified 23 algorithms performing both clustering and tree reconstruction. Nineteen out of 23 were excluded due to reasons of unavailability, unresolvable error messages, having the wrong scope or missing documentation (see Figure 1, see Supplementary Information, section 1.2). The remaining 4 [Canopy (21), Cloe (22), LICHeE (23) and SPRUCE (24)] were, together with the 7 tools only performing clustering, considered for detailed evaluation.

As a second step, subsequent to clustering, actual tree reconstruction is performed: Clustering results are taken as input. Every cluster represents one clone/node in the clonal evolution tree. By analyzing the development of the different clones over time, the most likely underlying clonal evolution tree is determined.

Altogether, we identified 7 tools with the main functionality of clonal evolution tree reconstruction. Three out of 7 were excluded due to reasons of unresolvable error messages or having the wrong scope (see Figure 1, see Supplementary information, section 1.3). The remaining tools are ClonalTree (25), ClonEvol (26), SCHISM (27) and TRaP (28). Additionally, tools providing a full analysis pipeline, performing clustering as well as tree reconstruction, were taken into account: Canopy, Cloe, LICHeE and SPRUCE. Thus, a total of 8 tools was considered for detailed evaluation.

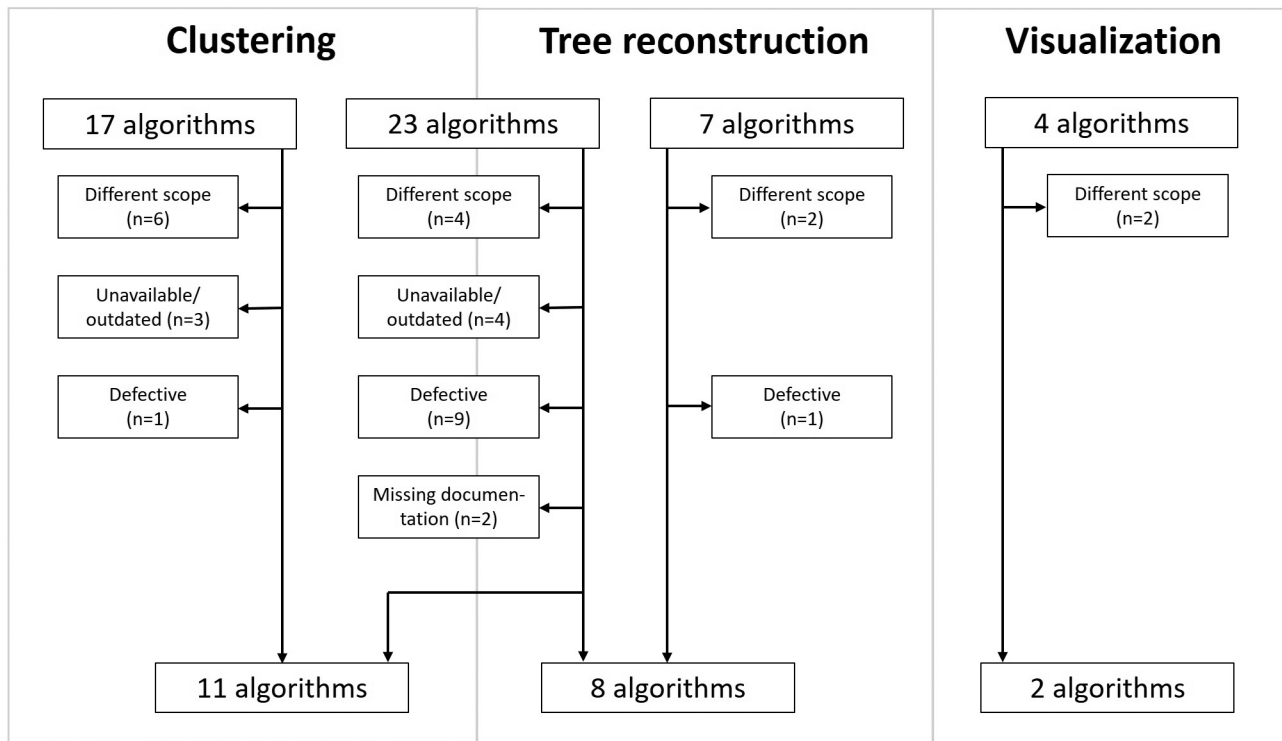


Figure 1. Overview of algorithms identified for clonal evolution analysis. Approaches are classified according to the three main tasks analyzing clonal evolution: clustering of variants, reconstruction of clonal evolution trees and visualization.

As a third and final step, we consider visualization of clonal evolution. Most algorithms performing clonal evolution tree reconstruction provide a basic visualization of the determined tree. However, we concentrate on specialized tools for visualization in the field of clonal evolution. We require an approach to display the development of all clones over time, to show the precise tumor load at every time point, and to provide an option for labeling clones according to their precise mutations. To visualize clonal evolution, we identified 4 tools. However, performing the required tasks, 2 out of 4 tools were excluded (see Figure 1, see Supplementary Information, section 1.4). For the remaining 2 tools, fishplot (29) and timescape (30), a detailed evaluation was performed.

In Table II an overview of all algorithms considered for detailed evaluation on the basis of two real data sets is provided. For every tool, the precise command executing analysis is provided in Supplementary Information, sections 1.1 to 1.4, exemplarily considering data set 1, patient 1.

Results

Clustering. As an essential prerequisite to clonal evolution tree reconstruction, clustering of variants is performed. All variants with similar CCFs are expected to be clustered together, assuming that they represent one clone. We focus our analysis on algorithms clustering SNVs and small indels, ideally also taking the underlying copy number into account. Algorithms clustering both, small variants originating from

NGS data and non-overlapping CNVs originating from array data, FISH and/or karyotyping could not be identified.

We evaluate 11 tools for variant clustering: CloneHD, clonosGP, DeCiFer, PyClone, PyClone-VI, QuantumClone, sciClone, Canopy, Cloe, LICHeE and SPRUCE. Tools show considerable differences in the approach applied for clustering (see Table II) and, as a result, the required input data.

CloneHD is evaluating coverage information (number of reads with the alternative and sequencing depth) of SNVs/indels, optionally also considering read depth and B-allele count data. However, as read depth profiles are commonly considered in the context of WGS data and our data is mainly characterized by targeted sequencing and partly WES data, we focus analysis with CloneHD on SNV/indel information only.

Canopy allows for considering coverage information on SNVs/indels and underlying CNVs. However, as the corresponding function leads to errors when applied to our data, we perform analysis on the basis of SNV/indel data only (see Supplementary Information, section 1.2.1).

Cloe considers, to our knowledge, only coverage information on SNVs/indels. Similarly, LICHeE is just taking information on the variants' VAFs as input.

ClonosGP, PyClone, PyClone-VI, QuantumClone and sciClone evaluate information on coverage of SNVs/indels as well as the underlying copy number in tumor vs normal

Table II. Overview of all tools considered for reconstructing clonal evolution based on two real data sets. Tools are sorted according to their main functionality: clustering, tree reconstruction and visualization. C: Clustering; T: tree reconstruction; V: visualization.

Program	Approach	Implementation	Functionality
cloneHD	Probabilistic algorithm for subclone reconstruction	C++	C
clonosGP	Clonal deconvolution method incorporating temporal spacing of longitudinally sampled tumors	Python	C
DeCiFer	Clustering SNVs by descendant cell fraction	Python, C++	C
PyClone	Bayesian clustering method for grouping sets of somatic mutations	Python	C
PyClone-VI	Bayesian clustering method for inferring clonal population structure	Python	C
QuantumClone	Variant clustering based on VAFs, genotypes and tumor purity	R	C
sciClone	Identify subclones using variational Bayesian mixture modeling of beta distributions	R	C
Canopy	Bayesian framework to reconstruct tumor phylogeny by NGS data	R	C+T
Cloe	Bayesian method employing a phylogenetic latent feature model	R	C+T
LICHeE	Evaluation of VAFs to infer subclonal composition and cell lineage trees	Java	C+T
SPRUCE	Multi-state perfect phylogeny mixture deconvolution using combinatorial enumeration	C++	C+T
ClonalTree	Maximum likelihood framework modelling occurrence of mutations	Python	T
ClonEvol	Bootstrap resampling approach to estimate CCFs and probabilistically model clonal ordering constraints	R	T
SCHISM	Statistical hypothesis tests for phylogenetic tree reconstruction	Python	T
TRaP	Evolutionary framework for deconvolving data from single genome-wide experiment	Java	T
fishplot	Display changes in clonal structure over time	R	V
timescape	Visualizing temporal clonal evolution data	R	V

samples (deletions, amplifications and loss of heterozygosity – LOH). This approach also allows for the correct handling of variants on the X- and Y-chromosome.

DeCiFer and SPRUCE evaluation not only covers information of SNVs/indels and the underlying copy number, but also the fraction of cells in which this copy number is observed. Thereby, DeCiFer and SPRUCE are the only tools not just modeling CNVs as binary events (present in all or no cells), but the precise circumstances.

Clustering results for all tools and samples are summed up in Table III. Detailed results on each variant and its assigned cluster are available in Supplementary Data 1 (data set 1) and Supplementary Data 2 (data set 2).

It can be observed that clusters reported by the 11 approaches show major differences compared to true clustering results. For data set 1, no tool succeeds in clustering variants correctly for any sample. For data set 2, three tools – clonosGP, Canopy and LICHeE – report correct clustering for 2 out of 11 samples. Five additional tools – PyClone, PyClone-VI, QuantumClone, sciClone and Cloe – perform correct clustering for 1 out of 11 samples.

Data set 1 is characterized by a high number of clusters (7 to 17) and a low number of time points (1 to 2). It is expected that this situation is difficult to handle for all clustering algorithms. In general, the number of clusters is underestimated by the applied tools. SPRUCE fails to report any results for any of the samples. CloneHD reports the number of clusters, but output files do not allow for the assignment of the variants to a cluster.

Detailed evaluation of the clustering results shows that clonosGP, DeCiFer, PyClone, PyClone-VI, Cloe and LICHeE tend to cluster variants similar to biological truth, but with higher gradation. However, exceptions from this observation can also be made. For patient 4, for example, all tools except for clonosGP report completely mixed-up clusters. QuantumClone and Canopy, on the contrary, show considerable differences compared to biological truth for all patients. Partly, variants from most distant original clusters are put together, while variants from the same cluster are split. SciClone reports only one cluster for a majority of patients. However, most variants are excluded by the algorithm as they do not fit this clustering. The single cluster reported for patient 8, for example, contains only 1 out of 17 variants.

Detailed evaluation of data set 2 shows similar results. Again, ClonosGP, DeCiFer, PyClone, PyClone-VI, Cloe and LICHeE show clusters similar, but less finely graduated compared to true clusters. However, the performance of DeCiFer appears slightly inferior as several mixed-up clusters can be observed (UPN04, UPN09, UPN11 and to a lower extend UPN02, UPN03, UPN06, UPN07). Like in case of data set 1, QuantumClone reports mixed-up clusters for several samples. However, for five samples, only a single cluster is reported. Against the background of partly high differences in CCF (e.g., UPN01 time point 3: $CCF_{TP53}=0\%$, $CCF_{PALLD}=75\%$) this observation appears astonishing. Of note, improved performance of Canopy can be observed compared to data set 1. Clustering results are close to true

Table III. Number of clusters for 11 tools performing clustering of variants in comparison to true clusters. Results marked in bold indicate perfect results, not just concerning the number of clusters, but also including assignment of the right variant to the right cluster.

	Data set 1										Data set 2										
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	11
Patient	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	11
Timepoints	2	2	2	2	2	1	1	1	1	1	15	5	5	5	5	5	6	12	8	30	5
Cluster	17	13	7	10	17	8	8	7	11	7	2	7	5	1	7	3	6	4	8	9	3
CloneHD	3	3	2	2	3	2	2	1	1	1	2	3	3	2	1	2	3	2	3	3	2
clonosGP	5	3	3	2	5	2	2	1	1	2	1	3	4	1	2	2	5	2	4	4	3
DeCiFer	9	9	5	6	8	5	5	5	5	2	1	4	4	2	4	4	2	2	2	3	5
PyClone	6	8	3	4	11	4	6	4	3	2	2	5	5	2	6	4	6	2	3	2	3
PyClone-VI	6	5	2	3	5	2	3	2	2	/	3	6	4	1	5	4	5	3	5	5	3
QuantumClone	5	4	3	5	5	4	2	3	3	2	1	5	4	1	1	5	1	2	1	4	3
sciClone	4	2	1	1	1	2	1	1	1	2	1	2	1	1	2	/	2	1	1	/	1
Canopy	10	11	7	7	15	6	6	3	5	4	2	7	5	/	7	3	6	4	8	9	3
Cloe	3	3	1	3	5	2	1	2	2	1	3	5	6	1	7	6	6	5	8	8	4
LICHeE	3	4	1	2	3	2	2	1	1	1	2	6	4	1	3	2	4	4	7	6	3
SPRUCE	/	/	/	/	/	/	/	/	/	/	4	13	12	8	12	5	16	6	6	10	9

clustering and – for UPN01 and UPN06 – perfectly match biological truth.

For data set 2, systematic underestimation of the true number of clusters can no longer be observed except for sciClone. The highest number of clusters is reported by SPRUCE. Detailed evaluation of the results shows that only part of the SNVs/indels present in the data is actually considered by the approach. Furthermore, no clustering is performed. Instead, every variant represents a single cluster. It appears likely that due to the high number of mutations characterizing patients in data set 1, SPRUCE fails to generate any results.

For Canopy and Cloe we provide the true number of clusters as an input parameter. It can be observed that Cloe fails to report the pre-defined number of clusters in all but 4 cases (all in data set 2). For Canopy, the determined number of clusters corresponds to the pre-defined one for 1 out of 10 samples in data set 1 and for 10 out of 11 samples in data set 2. Of note, the actual label of the highest clone, *e.g.*, clone 17 for data set 1, patient 1, is always reported by Canopy. However, several intermediate clones are missing.

Tree reconstruction. Based on clustered variant calls, a clonal evolution tree can be determined. We distinguish between algorithms providing a full pipeline for variant clustering and subsequent tree reconstruction based on the tools' own clustering results and algorithms only performing tree reconstruction. For each type, we identified four approaches: Canopy, Cloe, LICHeE and SPRUCE provide a full pipeline, while ClonalTree, ClonEvol, SCHISM and TRaP only provide an approach for tree reconstruction.

For Canopy, Cloe, LICHeE and SPRUCE no additional input has to be defined to allow for tree reconstruction.

Analyses are performed on the initially provided input characterizing SNVs/indels and information on the underlying copy number in the case of SPRUCE.

ClonalTree and TRaP do not take any information on detected variants as input. Instead, clusters and their CCFs at every time point are directly evaluated. TRaP additionally provides an option to specify the estimated CCFs' errors.

By contrast, ClonEvol analyzes information on every variant, their assigned cluster and their CCF at every time point. Similarly, SCHISM requires information on every SNV/indel, their coverage, the underlying copy number and the assigned cluster as input.

The mechanisms of clonal evolution, determined by the 8 tools, are – for all samples – summed up in Table IV. Detailed results on each clone and its determined parent are available in Supplementary Data 3 (data set 1) and Supplementary Data 4 (data set 2).

For all tools just providing a tree reconstruction approach – ClonalTree, ClonEvol, SCHISM and TRaP – we provide the correct clustering of variants as input. Still, the correct mechanism of clonal evolution cannot be determined in all cases. ClonalTree succeeds in 5 out of 22 (counting data set 1, patient 8 twice due to two models being possible; no tree can be determined for 7 samples), ClonEvol in 8 (no tree for 12 samples), SCHISM in 5 (no tree for 1 sample) and TRaP in 9 (no tree for 1 sample).

It can be observed that all tools except for SCHISM show especially good performance considering only a single time point (data set 1, patients 6 to 10). ClonEvol and TRaP even succeed in identifying both possible models for patient 8 – linear as well as branched dependent evolution. SCHISM, however, mistakes the ordering of clones in 4 out of 5 patients.

Table IV. Mechanism of clonal evolution determined by 8 tools in comparison to true clonal evolution. Results marked in bold indicate perfect results, not just concerning the general mechanism, but also including correct ordering of the clones. *D*: Branched dependent evolution; *I*: branched independent evolution; *L*: linear evolution; *N*: no evolution; *two to three versions reported; **more than three versions reported; ***more than 100 versions reported; *A*: all possible versions reported.

	Data set 1										Data set 2										
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	11
Patient	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	11
Timepoints	2	2	2	2	2	1	1	1	1	1	15	5	5	5	5	5	6	12	8	30	5
Evolution	D	D	D	L	D	L	L	L/D	L	L	L	D	D	N	D	L	D	I	I	I	L
ClonalTree	/	/	D	I	/	L	L	D	/	L	I	I	N	/	/	I	I	I	/	I	
ClonEvol	/	/	D	/	/	L	L	L/D	L	L	/	D*	D*	N	/	/	/	/	/	/	
SCHISM	D**	D*	L	D	D**	L	L	L	L	L	L	D*	D*	/	D*	A	D*	D*	A	D	L
TRaP	/	D***	D	D***	D	L	L	L/D	L	L	L	D***	D**	N	D***	L/D	D***	D**	D***	D***	L/D
Canopy	/	/	L	/	/	/	/	/	/	/	L	L	D	/	L	L	D	D	L	L	L
Cloe	L	L	N	L	D	L	N	L	L	N	L	D	D	N	L	D	D	I	I	I	L
LICHeE	D	D	N	L	D	L	L	N	N	N	L	D	D	N	L	D	D	I	I	D	L
SPRUCE	/	/	/	/	/	/	/	/	/	/	L**	D**	D**	L**	D**	L	D***	D**	D***	D**	L**

For data set 1, patients 1 to 5, the correct clonal evolution tree is not reconstructed by any tool. For data set 2, correct results can partly be observed for UPN01, UPN02, UPN03, UPN04, UPN06 and UPN11. However, no tool succeeds in reporting the correct results for all 6 patients. Of note, branched independent evolution (UPN08, UPN09, UPN10) is never determined correctly.

When reconstructing clonal evolution manually, the presence of a high number of time points eases the process considerably. Usually, it is possible to retrieve unique solutions. When determining clonal evolution trees automatically, however, data indicates that many time points rather lead to inferior results. TRaP tends to report a high number of possible solutions, partly >100. SCHISM similarly reports several possible solutions or – in two cases – even reconstructs a fully connected graph. By contrast, ClonalTree and ClonEvol report no solution as no converging tree can be found.

For tools providing a full pipeline for reconstruction of clonal evolution – Canopy, Cloe, LICHeE and SPRUCE – the true clonal evolution can only be reconstructed for two patients: data set 2, UPN01 and UPN04. Of note, UPN04 actually does not show any clonal evolution as all variants form a single cluster, showing minor changes in frequency over time. UPN01 is characterized by the presence of only 2 clones, 15 time points and linear clonal evolution. Canopy and LICHeE report the correct tree. For UPN06 Canopy reported the correct clustering. The tool correctly determines linear clonal evolution. However, ordering of the clones does not match biological truth. For all other patients and tools, clustering does not lead to the correct result, which is why the correct tree cannot be determined.

Visualization. Numerous approaches exist for visualizing clonal evolution. A basic approach considers common trees,

with nodes representing clones and edges indicating their evolutionary relation. In this work, however, we focus on plots displaying detailed information on clonal evolution, including the development of all clones over time and clonal prevalence. We identified and evaluated two tools performing this task: fishplot and timescape. The clonal evolution plots generated with fishplot and timescape for a representative example of linear and branched dependent evolution each, evaluating data set 1, are provided in Figure 2. Results for all samples are available in Supplementary Information, section 2.1.

Comparing visualization of linear clonal evolution in the presence of few time points and many clones with fishplot (Figure 2A) to timescape (Figure 2B), major differences can be observed. The patient is characterized by 10 clones, one only present at the second time point. In the plot generated with fishplot, however, only 7 clones are visible. The time between two emerging clones continuously decreases. According to the figure, the first gray clone is exclusively present for a long time. Then, within short time, clones 2 to 9 appear. With respect to time point 2, the newly emerging clone 10 is not visible at all.

Different from fishplot's visualization, clones are well separated in the plot generated by timescape. Instead of decreasing time spans, separating two emerging clones, equidistant development is assumed. As a result, all clones present at time points 1 and 2 can be identified in the plot. Additionally, a basic tree visualizing clonal evolution, using matching colors, is provided next to the main plot.

Similar results can also be observed when considering an exemplary case of branched dependent evolution in the presence of a few time points and many clones. The visualization by fishplot (Figure 2C) does not allow for the identification of all 17 clones. Instead, only 9 clones can be

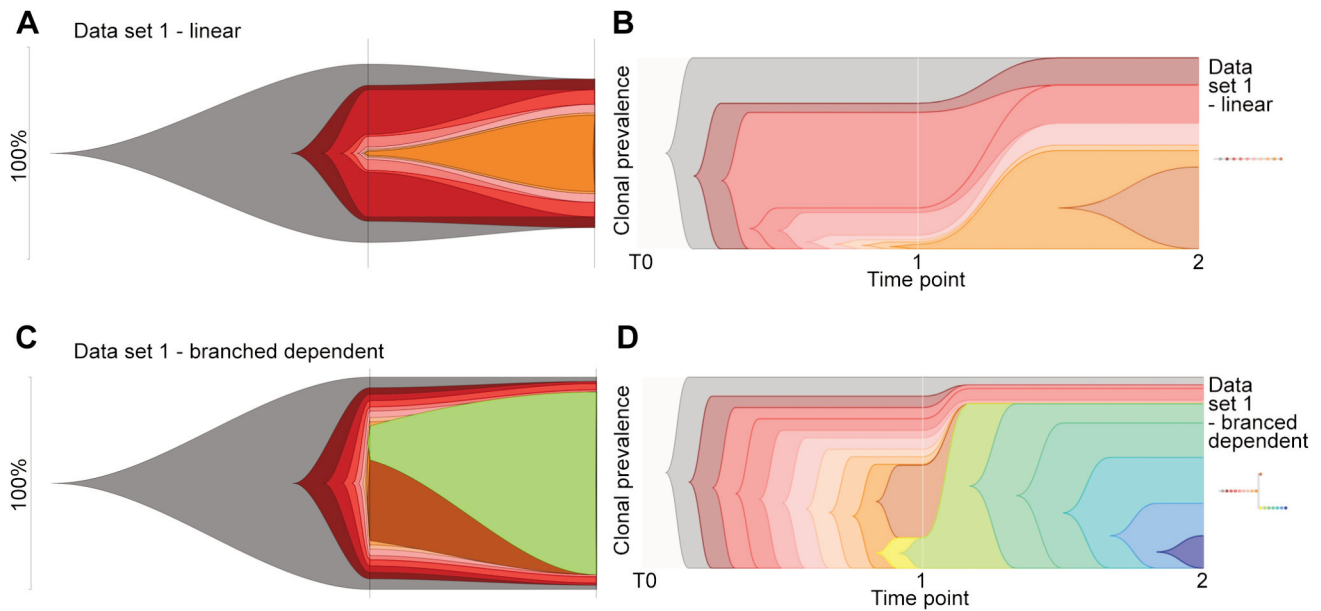


Figure 2. Visualization of clonal evolution analyzing two representative examples of data set 1 (linear: patient 4; branched dependent: patient 5). (A) Visualization of linear evolution using fishplot. (B) Visualization of linear evolution using timescape. (C) Visualization of branched dependent evolution using fishplot. (D) Visualization of branched dependent evolution using timescape.

identified. None of the 5 clones characterizing time point 2 are visible. In the plot generated by timescape, however, all 17 clones are distinguishable. Additionally, the tree plot allows for quick identification of the underlying evolutionary model.

Despite this disadvantage of timescape, it is possible to manually improve visualization using additional auxiliary time points, previously described by Reutter *et al.* (3). These time points can be set to force fishplot towards visualization of equidistant development. Exemplary output for all patients in data set 1 is provided in Supplementary Information, Figures S1 to S11.

Considering visualization of clonal prevalence, a disadvantage of timescape can be observed: it is known that the tumor load at time point 1 is at ~85%, while it decreases to ~70% at time point 2, which can be clearly observed in Figure 2A generated by fishplot. By contrast, Figure 2B generated by timescape suggests a tumor load of 100% for both time points.

The relative clonal prevalence as displayed by timescape, however, matches the actual prevalence. As plots generated by timescape are interactive, it is possible to receive information on the exact underlying clonal prevalence of each clone by hovering over it. However, it should be noted that timescape does not take the CCF as input, but the difference in CCF of related clones.

Similar to fishplot, it is possible to manually improve visualization with timescape. An additional clone can be defined, representing normal cells. If no relation is given for

this clone, it is not displayed in the plot, but empty space is included instead, leading to a correct representation of absolute clonal prevalence. Exemplary output for all patients in data set 1 with CCFs of the outer-most clone <100% are provided in Supplementary Information, Figures S2 to S4 and S8 to S11.

Results for data set 2, considering a representative example of linear evolution, branched dependent and branched independent evolution each is provided in Figure 3. Results for all samples are available in Supplementary Information, section 2.1.

Data set 2 is characterized by a relatively high number of time points (5 to 30) and low number of clones (1 to 9). As a result, visualization by fishplot is considerably improved. Manually adding auxiliary time points does not appear necessary. In Figure 3A, showing linear clonal evolution of 2 clones, both clones can be distinguished easily. The same is true for the exemplary branched dependent evolution in Figure 3C (5 clones) and – with minor constraints – also the branched independent evolution in Figure 3E (8 clones). As the orange subclone is dominated by the green and blue/yellow subclone, it is expected to be observed with certain difficulties. Also in case of plots generated with timescape, all clones can be identified easily. Of note, branched independent evolution cannot be visualized using timescape. The tool mandatorily requires trees with a single root.

As regards visualization of clonal prevalence, correct scaling can be observed in the case of fishplot. For timescape

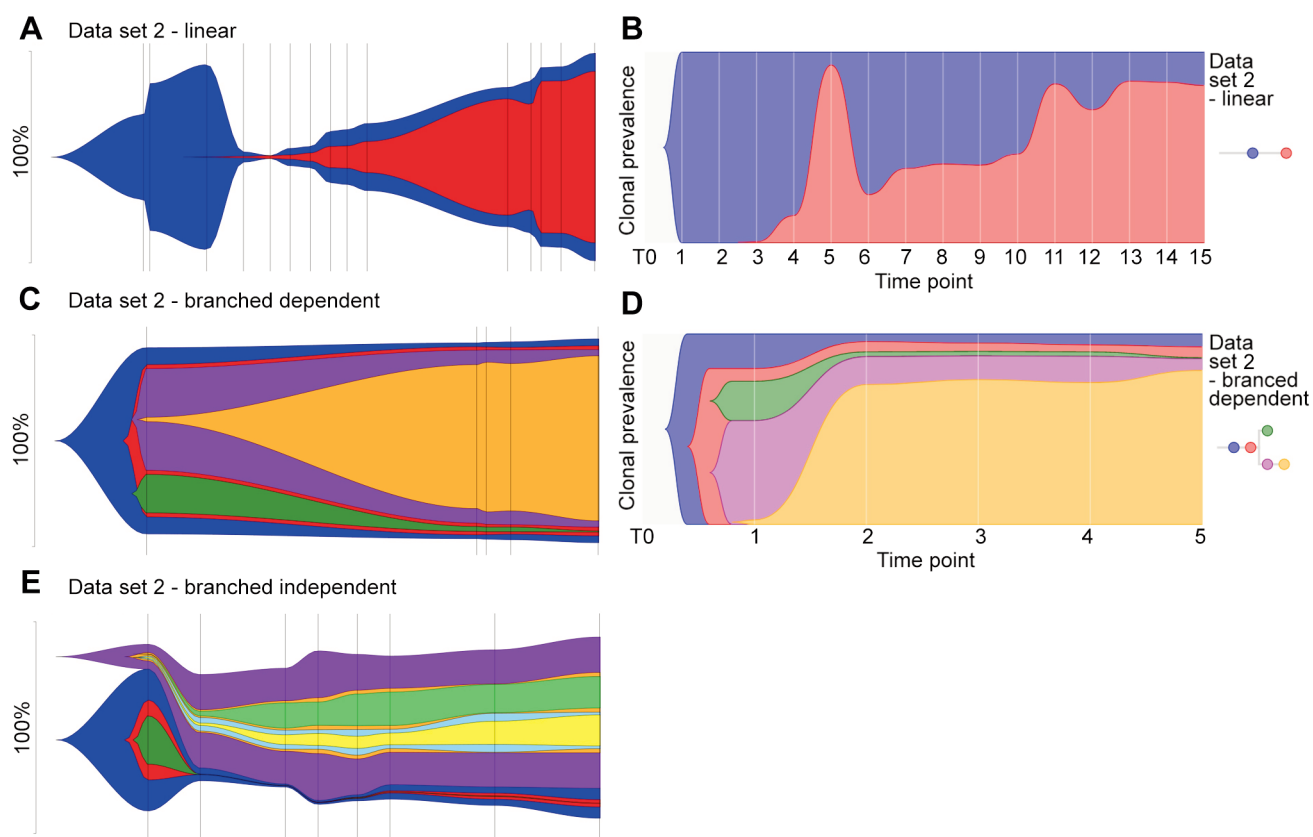


Figure 3. Visualization of clonal evolution analyzing three representative examples of data set 2 (linear: UPN01; branched dependently: UPN03; branched independent: UPN09). (A) Visualization of linear evolution using fishplot. (B) Visualization of linear evolution using timescape. (C) Visualization of branched dependent evolution using fishplot. (D) Visualization of branched dependent evolution using timescape. (E) Visualization of branched independent evolution using fishplot. Of note, timescape does not support visualization of branched independent evolution.

the relative clonal prevalence is displayed correctly. However, to visualize the correct absolute scaling, normal cells have to be added manually (see Supplementary Information, Figures S12 to S14, S16 to S18 and S22).

Of note, timescape fails at visualizing the development of UPN04, which is characterized by a single clone. Strictly speaking, no clonal evolution can be observed in this case. However, visualization of clonal prevalence over time might still be interesting. With respect to timescape, this is not possible as no information on tree edges can be provided.

Discussion

In this work we evaluated tools performing the three main tasks in the process of reconstructing clonal evolution: clustering of variants, reconstructing clonal evolution trees and visualizing clonal evolution. We identified 51 approaches and performed a detailed evaluation of 17. We analyzed two publically available data sets, covering altogether 21 patients.

When clustering variants we focused on SNVs and indels. Most tools performing clustering consider the underlying copy number. Even in case of CNVs being absent, this approach appears essential to model small variants located on the X- and Y-chromosome correctly. However, only two tools – DeCiFer and SPRUCE – allow for the definition of the precise cellular fraction of a CNV. Unexpectedly, these two tools do not perform best on our data sets. Instead, they never report the correct clustering. Performance of the remaining 9 tools is greatly comparable: the correct clustering is – at a maximum – reported for 2 out of 21 patients per tool. For data set 1, the correct clustering is never reported by any tool.

It may be argued that manually determined clusters in data set 1, taken as biological truth, are defined too finely graduated. In the presence of only 1 or 2 time points it is possible that the true number of clusters is manually overestimated. Natural variation in the VAF of variants may be misinterpreted to originate from different clones. Though, data is characterized by high sequencing depth (at every variant position commonly

>100×). Furthermore, clustering results determined by the tools do usually not just indicate 2 or 3 clusters being merged, but instead completely different clustering.

Our results show the general deficiencies of currently available algorithms with respect to clustering variants. To our knowledge, not a single algorithm accepts variant calls from different sources as input, automatically integrating NGS, array, FISH and/or karyotyping data. In theory, it is possible to define CNVs detected by FISH or karyotyping just like SNVs/indels. The number of evaluated cells could be defined as coverage, the number of cells with the CNV divided by 2 as the number of reads carrying a heterozygous variant. However, for variants only detected by array analyses, i.e., LOH and smaller CNVs, this is impossible. Furthermore, equating the number of evaluated cells with read counts does not reflect validity of the variant calls correctly. A CNV that is microscopically detected in 1 out of 10 cells indicates a true variant with a CCF of 10%. A mismatch present in 1 out of 10 reads, on the contrary, indicates a sequencing artifact.

With respect to reconstructing clonal evolution trees we consider both, tools only performing tree reconstruction based on user-defined clustering as well as tools performing reconstruction based on their own clustering results. Even when providing perfect clustering as input, the correct tree is not reported by any of the tools for 11 out of 21 patients. The best results can be observed for data available at a single time point. It remains unclear why the tools do not show superior performance in the presence of more time points.

For tools providing a full analysis pipeline, results are even worse as the initial clustering is usually not correct. LICHeE, as well as SPRUCE, do not provide any option for interference: Clustering results cannot be manipulated manually as the main analysis is executed by a single command. Cloe is working with specific data structures within R, including a “cloe_input” object and “meta-mutations”. Thus, manipulation of the clustering appears equally impossible. Just in the case of Canopy it may be possible to change clusters determined by the tool to improve tree reconstruction.

As regards visualizing clonal evolution, both available tools – fishplot and timescape – show certain deficiencies: while fishplot does not allow for identification of distinct clones in the context of many clones and few time points, timescape does not display the correct clonal prevalence in case of a tumor load <100%. However, it is possible to compensate for both deficiencies. Auxiliary time points can be added to improve visualization with fishplot, normal cells can be added to improve visualization with timescape.

Despite these two acceptable solutions being available, to our knowledge, no tool visualizing clonal evolution on an allelic level is currently available. In cancer research, the identification of “double-hit events” is particularly important

as it might be related to a higher risk for disease progression or relapse (3). If, for example, a patient has a deletion of chr17p and – on the remaining p-arm – an additional SNV in gene *TP53*, this double-hit of *TP53* cannot be visualized, neither by fishplot nor by timescape.

It may be questioned why we focus our work on variants detected in bulk DNA-seq and not single-cell DNA-seq (scDNA-seq) data. Despite the clear advantage of scDNA-seq to potentially determine the exact mutational profile of a single cell, the technique is, at present, still characterized by major disadvantages: Due to technical issues, like amplification bias, leading to allelic imbalance, and allelic dropout, sensitivity with respect to variant calling is still low. Data preparation for single-cell sequencing limits its application to either determining small variants or CNVs in a single cell – not both. Furthermore, analysis of retrospective samples is, conditioned by missing suitable material, not possible (31-34).

Concluding, currently available methods evaluated in this paper, performing reconstruction of clonal evolution, cannot yet replace manual reconstruction. No valid tool or combination of tools could be identified to perform clustering and tree reconstruction with comparable precision.

In the context of large sets of data collected in clinical routine, it appears highly beneficial to develop a reliable method replacing tedious manual reconstruction of clonal evolution. However, in order to develop such a method, systematic error analysis has to be performed first. While our analysis of two real data sets clearly shows the deficiencies of current approaches, the reasons for their limitations are not thoroughly clear. Therefore, it will be the next step to develop a simulation tool generating realistic user-definable input data for clonal evolution analysis. On the basis of this data, it can be investigated why certain algorithms generate mixed-up clusters of variants or fail to generate clonal evolution trees in the presence of >1 time point. Only then it may be possible to identify the best approach to develop a pipeline for thoroughly automatic reconstruction of clonal evolution.

Supplementary Material

Available at: <https://uni-muenster.sciebo.de/s/YTHAwXtvXcrRzp9>

Conflicts of Interest

The Authors declare that they have no conflicts of interest.

Authors' Contributions

SS performed analyses and wrote the manuscript. SR performed supporting analyses. JV and XJ supported the analyses and provided feedback. All Authors read and approved the final version of this manuscript.

Acknowledgements

We would like to thank Dr. Marius Wöste for his technical support.

References

- 1 International Agency for Research on Cancer: World Health Organization, 2020. Available at: <https://gco.iarc.fr/today/data/factsheets/cancers/39-All-cancers-fact-sheet.pdf> [Last accessed on October 21, 2021]
- 2 Aukema SM, Kreuz M, Kohler CW, Rosolowski M, Hasenclever D, Hummel M, Küppers R, Lenze D, Ott G, Pott C, Richter J, Rosenwald A, Szczepanowski M, Schwaenen C, Stein H, Trautmann H, Wessendorf S, Trümper L, Loeffler M, Spang R, Kluin PM, Klapper W, Siebert R and Molecular Mechanisms in Malignant Lymphomas Network Project: Biological characterization of adult MYC-translocation-positive mature B-cell lymphomas other than molecular Burkitt lymphoma. *Haematologica* 99(4): 726-735, 2014. PMID: 24179151. DOI: 10.3324/haematol.2013.091827
- 3 Reutter K, Sandmann S, Rohde J, Müller S, Wöste M, Khanam T, Michgehl U, Klapper W, Wöbmann W, Seggewiß J, Lenz G, Dugas M and Burkhardt B: Reconstructing clonal evolution in relapsed and non-relapsed Burkitt lymphoma. *Leukemia* 35(2): 639-643, 2021. PMID: 32404974. DOI: 10.1038/s41375-020-0862-5
- 4 Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, Schnittger S, Sanada M, Kon A, Alpermann T, Yoshida K, Roller A, Nadarajah N, Shiraishi Y, Shiozawa Y, Chiba K, Tanaka H, Koeffler HP, Klein HU, Dugas M, Aburatani H, Kohlmann A, Miyano S, Haferlach C, Kern W and Ogawa S: Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 28(2): 241-247, 2014. PMID: 24220272. DOI: 10.1038/leu.2013.336
- 5 Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, Yoon CJ, Ellis P, Wedge DC, Pellagatti A, Shlien A, Groves MJ, Forbes SA, Raine K, Hinton J, Mudie LJ, McLaren S, Hardy C, Latimer C, Della Porta MG, O'Meara S, Ambaglio I, Galli A, Butler AP, Walldin G, Teague JW, Quek L, Sternberg A, Gambacorti-Passerini C, Cross NC, Green AR, Boulton J, Vyas P, Hellstrom-Lindberg E, Bowen D, Cazzola M, Stratton MR, Campbell PJ and Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium: Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 122(22): 3616-27; quiz 3699, 2013. PMID: 24030381. DOI: 10.1182/blood-2013-08-518886
- 6 Doulatov S and Papapetrou EP: Studying clonal evolution of myeloid malignancies using induced pluripotent stem cells. *Curr Opin Hematol* 28(1): 50-56, 2021. PMID: 33264225. DOI: 10.1097/MOH.0000000000000620
- 7 Edited by Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H and Thiele J: WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. International Agency for Research on Cancer. Revised 4th edition, 2017.
- 8 Malcovati L, Crouch S, de Graaf AO, Sandmann S, Tobiaasson M, Kosmider O, van der Reijden BA, Painter D, van de Loosdrecht AA, Symeonidis A, Cermak J, Clappier E, Preudhomme C, Stauder R, Mittelman M, Germing U, Bowen D, Fenaux P, Van Marrewijk C, Smith A, Dugas M, Hellstrom Lindberg E, de Witte TM, Fontenay M and Jansen J: Mutation profiles identify distinct clusters of lower risk myelodysplastic syndromes with unique clinical and biological features and clinical endpoints. *Blood* 136(Suppl 1): 29, 2020. DOI: 10.1182/blood-2020-138549
- 9 National Human Genome Research Institute: The Cost of Sequencing a Human Genome, 2020. Available at: <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost> [Last accessed on October 21, 2021]
- 10 Sandmann S, de Graaf AO, Karimi M, van der Reijden BA, Hellström-Lindberg E, Jansen JH and Dugas M: Evaluating variant calling tools for non-matched next-generation sequencing data. *Sci Rep* 7: 43169, 2017. PMID: 28233799. DOI: 10.1038/srep43169
- 11 Sandmann S, Karimi M, de Graaf AO, Rohde C, Göllner S, Varghese J, Ernsting J, Walldin G, van der Reijden BA, Müller-Tidow C, Malcovati L, Hellström-Lindberg E, Jansen JH and Dugas M: appreci8: a pipeline for precise variant calling integrating 8 tools. *Bioinformatics* 34(24): 4205-4212, 2018. PMID: 29945233. DOI: 10.1093/bioinformatics/bty518
- 12 Sandmann S, Wöste M, de Graaf AO, Burkhardt B, Jansen JH and Dugas M: CopyDetective: Detection threshold-aware copy number variant calling in whole-exome sequencing data. *Gigascience* 9(11): giaa118, 2020. PMID: 33135740. DOI: 10.1093/gigascience/giaa118
- 13 da Silva-Coelho P, Kroeze LI, Yoshida K, Koorenhof-Scheele TN, Knops R, van de Locht LT, de Graaf AO, Massop M, Sandmann S, Dugas M, Stevens-Kroef MJ, Cermak J, Shiraishi Y, Chiba K, Tanaka H, Miyano S, de Witte T, Blijlevens NMA, Muus P, Huls G, van der Reijden BA, Ogawa S and Jansen JH: Clonal evolution in myelodysplastic syndromes. *Nat Commun* 8: 15099, 2017. PMID: 28429724. DOI: 10.1038/ncomms15099
- 14 Fischer A, Vázquez-García I, Illingworth CJR and Mustonen V: High-definition reconstruction of clonal composition in cancer. *Cell Rep* 7(5): 1740-1752, 2014. PMID: 24882004. DOI: 10.1016/j.celrep.2014.04.055
- 15 Vavoulis DV, Cutts A, Taylor JC and Schuh A: A statistical approach for tracking clonal dynamics in cancer using longitudinal next-generation sequencing data. *Bioinformatics* 37(2): 147-154, 2021. PMID: 32722772. DOI: 10.1093/bioinformatics/btaa672
- 16 Satas G, Zaccaria S, El-Kebir M and Raphael BJ: DeCifering the elusive cancer cell fraction in tumor heterogeneity and evolution. *Cell Syst* 12(10): 1004-1018.e10, 2021. PMID: 34416171. DOI: 10.1016/j.cels.2021.07.006
- 17 Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, Ha G, Aparicio S, Bouchard-Côté A and Shah SP: PyClone: statistical inference of clonal population structure in cancer. *Nat Methods* 11(4): 396-398, 2014. PMID: 24633410. DOI: 10.1038/nmeth.2883
- 18 Gillis S and Roth A: PyClone-VI: scalable inference of clonal population structures using whole genome data. *BMC Bioinformatics* 21(1): 571, 2020. PMID: 33302872. DOI: 10.1186/s12859-020-03919-2
- 19 Deveau P, Colmet Daage L, Oldridge D, Bernard V, Bellini A, Chicard M, Clement N, Lapouble E, Combaret V, Boland A, Meyer V, Deleuze JF, Janoueix-Lerosey I, Barillot E, Delattre O, Maris JM, Schleiermacher G and Boeva V: QuantumClone: clonal assessment of functional mutations in cancer based on a genotype-aware method for clonal reconstruction. *Bioinformatics* 34(11): 1808-1816, 2018. PMID: 29342233. DOI: 10.1093/bioinformatics/bty016

- 20 Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, Vij R, Tomasson MH, Graubert TA, Walter MJ, Ellis MJ, Schierding W, DiPersio JF, Ley TJ, Mardis ER, Wilson RK and Ding L: SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput Biol* *10*(8): e1003665, 2014. PMID: 25102416. DOI: 10.1371/journal.pcbi.1003665
- 21 Jiang Y, Qiu Y, Minn AJ and Zhang NR: Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. *Proc Natl Acad Sci USA* *113*(37): E5528-E5537, 2016. PMID: 27573852. DOI: 10.1073/pnas.1522203113
- 22 Marass F, Mouliere F, Yuan K, Rosenfeld N and Markowitz F: A phylogenetic latent feature model for clonal deconvolution. *Ann Appl Stat* *10*(4): 2377-2404, 2016. DOI: 10.1214/16-AOAS986
- 23 Popic V, Salari R, Hajirasouliha I, Kashef-Haghighi D, West RB and Batzoglou S: Fast and scalable inference of multi-sample cancer lineages. *Genome Biol* *16*: 91, 2015. PMID: 25944252. DOI: 10.1186/s13059-015-0647-8
- 24 El-Kebir M, Satas G, Oesper L and Raphael BJ: Inferring the mutational history of a tumor using multi-state perfect phylogeny mixtures. *Cell Syst* *3*(1): 43-53, 2016. PMID: 27467246. DOI: 10.1016/j.cels.2016.07.004
- 25 Ismail WM and Tang H: Clonal reconstruction from time course genomic sequencing data. *BMC Genomics* *20*(Suppl 12): 1002, 2019. PMID: 31888455. DOI: 10.1186/s12864-019-6328-3
- 26 Dang HX, White BS, Foltz SM, Miller CA, Luo J, Fields RC and Maher CA: ClonEvol: clonal ordering and visualization in cancer sequencing. *Ann Oncol* *28*(12): 3076-3082, 2017. PMID: 28950321. DOI: 10.1093/annonc/mdx517
- 27 Niknafs N, Beleva-Guthrie V, Naiman DQ and Karchin R: SubClonal hierarchy inference from somatic mutations: Automatic reconstruction of cancer evolutionary trees from multi-region next generation sequencing. *PLoS Comput Biol* *11*(10): e1004416, 2015. PMID: 26436540. DOI: 10.1371/journal.pcbi.1004416
- 28 Strino F, Parisi F, Micsinai M and Kluger Y: TrAp: a tree approach for fingerprinting subclonal tumor composition. *Nucleic Acids Res* *41*(17): e165, 2013. PMID: 23892400. DOI: 10.1093/nar/gkt641
- 29 Miller CA, McMichael J, Dang HX, Maher CA, Ding L, Ley TJ, Mardis ER and Wilson RK: Visualizing tumor evolution with the fishplot package for R. *BMC Genomics* *17*(1): 880, 2016. PMID: 27821060. DOI: 10.1186/s12864-016-3195-z
- 30 Smith M: Timescape: Patient clonal timescapes. *Bioconductor*, 2021. DOI: 10.18129/B9.bioc.timescape
- 31 Navin NE: Cancer genomics: one cell at a time. *Genome Biol* *15*(8): 452, 2014. PMID: 25222669. DOI: 10.1186/s13059-014-0452-9
- 32 Luquette LJ, Bohrson CL, Sherman MA and Park PJ: Identification of somatic mutations in single cell DNA-seq using a spatial model of allelic imbalance. *Nat Commun* *10*(1): 3908, 2019. PMID: 31467286. DOI: 10.1038/s41467-019-11857-8
- 33 Gawad C, Koh W and Quake SR: Single-cell genome sequencing: current state of the science. *Nat Rev Genet* *17*(3): 175-188, 2016. PMID: 26806412. DOI: 10.1038/nrg.2015.16
- 34 Zaccaria S and Raphael BJ: Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. *Nat Commun* *11*(1): 4301, 2020. PMID: 32879317. DOI: 10.1038/s41467-020-17967-y

Received October 29, 2021

Revised December 11, 2021

Accepted December 17, 2021