AMBAR - Interactive Alteration Annotations for Molecular Tumor Boards

Axel Fürstberger
Universitat Ulm
Nensi Ikonomi
Universitat Ulm
Angelika M.R. Kestler
Universitatsklinikum Ulm
Ralf Marienfeld
Universitatsklinikum Ulm
Thomas Seufferlein
Universitatsklinikum Ulm
Hans A. Kestler (hans.kestler@uni-ulm.de)
Universitat Ulm https://orcid.org/0000-0002-4759-5254

Software

Keywords: alteration, annotation, tumor boards, R shiny, FHIR, visualization, interactive

Posted Date: August 25th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-62823/v1

SOFTWARE

AMBAR - Interactive alteration annotations for molecular tumor boards

Axel Fürstberger[1], Nensi Ikonomi[1], Angelika MR Kestler[2], Ralf Marienfeld[3], Thomas Seufferlein[2] and Hans A Kestler[1*]

Abstract

Background: Providing suitable treatments strategies that take into account cancer specific alterations is a crucial task for successful cancer treatment. To this end, molecular tumor boards (MTBs), that bring together clinicians as well as scientists with diverse expertise, are increasingly established in the clinical routine for therapeutic interventions. Molecular profiling from sequencing data is an integral part of the decision making process of an MTB. To debate variant calling results from next generation sequencing NGS analyses, detailed information about the detected mutations are mandatory. Further, these results need to be combined with knowledge and up to date evidence from databases. At the moment, few tools are available that aim at managing this amount of required information. As a result, the whole process of analysis and documentation of patients data becomes time consuming and difficult to manage for MTBs.

Results: To overcome these limitations, we developed an interactive web application AMBAR (Alteration annotations for Molecular tumor BoARds) to visualize not only annotated mutations, but also evidence for possible therapeutic drug targets. Found mutations can be evaluated, discussed and exported to clinical information systems. The application is based on R shiny and allows customization, interactive filtering and visualization.

Conclusion: AMBAR is an interactive application to not only support MTBs in decision making, but to act as interface between results of NGS analyses, result visualization and export into clinical information systems.

Keywords: alteration; annotation; tumor boards; R shiny; FHIR; visualization; interactive

Background

Despite decades of research, cancer is still one of the leading causes of death worldwide [1]. Understanding the underlying mechanisms of carcinogenesis and finding suitable treatments is the most challenging question for successful cancer therapy [2]. In this context, analysis of next generation sequencing (NGS) data is becoming more and more common practice in oncology all over the world especially in light of personalized treatment approaches [3]. Moreover, lower sequencing costs and secure high performance compute servers have ensured the feasibility of using even large gene panels for targeted sequencing and in house analysis with NGS pipelines [4, 5, 6, 7]. Due to these improvements, an increasing number of databases for genes and molecular targets as possible drug targets in cancer therapy have been published [8, 9, 10, 11].

In NGS analysis, after sequencing alignment, variant calling and / or copy number variation analysis are necessary steps to search for mutations and aberrations. Annotation of alterations like single nucleotide variations (SNVs) as well as copy number variations (CNVs) of patients are then combined with databases for cancer treatments. Taken together, this process is crucial for personalized cancer treatment [12, 13]. Nevertheless, all the information driven by these analyses has to be revised and interpreted. Therefore, a group of clinicians, scientists, pathologists and geneticists discuss in molecular tumor boards (MTB) these patients in the context of clinical data, see Figure 1. Clinical interpretation of genetic variants is often done manually in a very time-consuming process.

Moreover, available databases and tools are queried often one by one and results are assembled for a report. Hence, even if the primary focus of molecular diagnostics is to propose a suitable personalized therapy, time consuming analyses and limited staff make this process costly and inefficient. Also printed stan-

*Correspondence: hans.kestler@uni-ulm.de
[1]Institute of Medical Systems Biology, Ulm University, Albert-Einstein-Allee 11, 89081 Ulm, Germany
Full list of author information is available at the end of the article

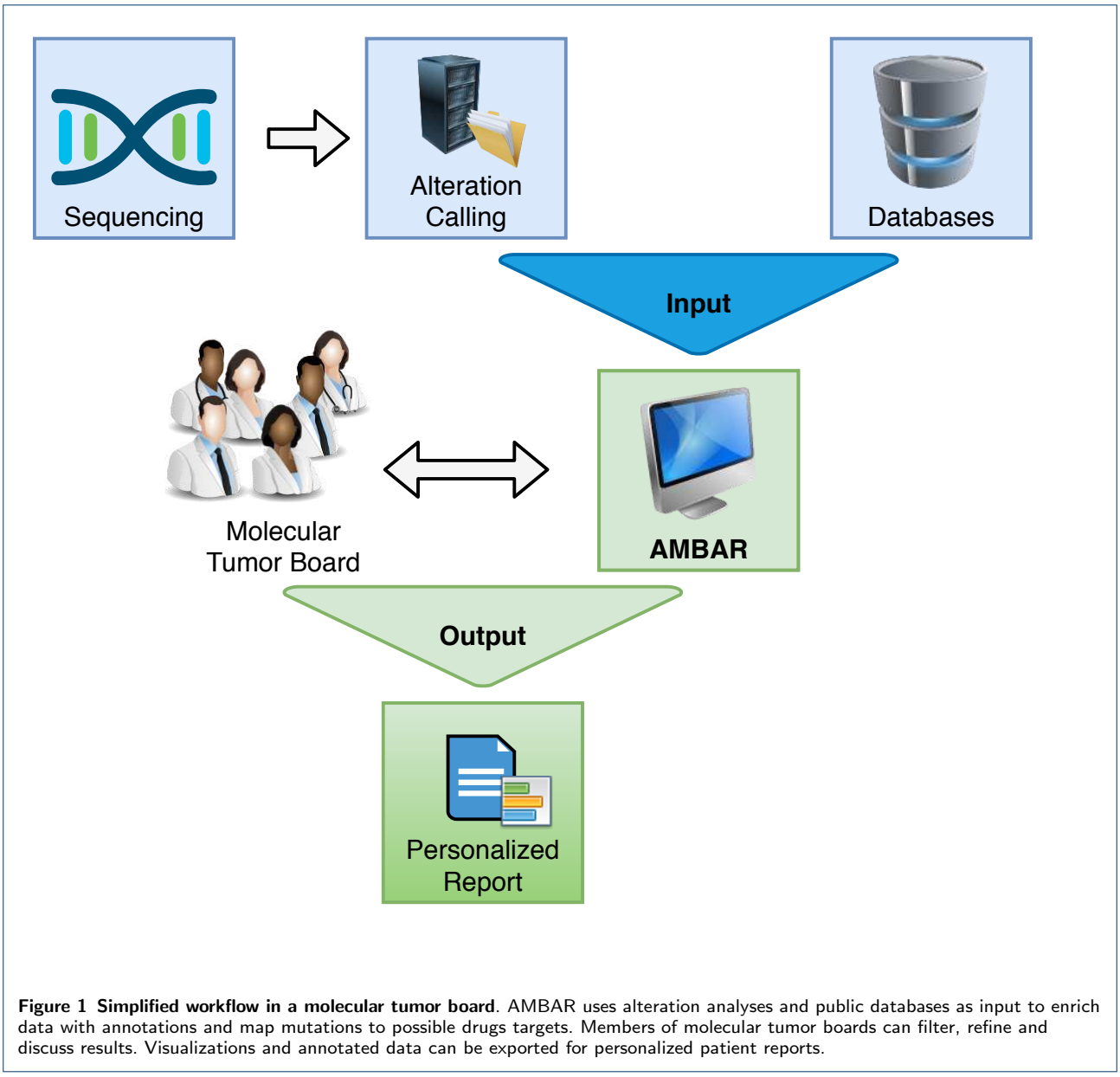


Figure 1 Simplified workflow in a molecular tumor board. AMBAR uses alteration analyses and public databases as input to enrich data with annotations and map mutations to possible drugs targets. Members of molecular tumor boards can filter, refine and discuss results. Visualizations and annotated data can be exported for personalized patient reports.

dardized forms have to be documented and transferred into digital format to the clinical information system. As a result, documentation and progress are prone to mistakes.

Existing tools to generate alteration reports are often not open source [14] or only generate static (PDF) reports [15], that are not interactive and cannot be transferred into clinical information system entries right away.

To tackle these problems, we developed in close cooperation with clinicians and pathologists an interactive application AMBAR (Alteration annotations for Molecular tumor BoARds) to enable meaningful and feature rich alteration annotation analysis and visualization with various export formats. Hence, our tool will crucially help MTBs in discussing and analyzing patients data with the final aim of providing personalized treatment suggestions.

Implementation

We developed an interactive R shiny application AMBAR (Alteration annotations for Molecular tumor BoARds) to visualize, annotate, enrich and refine variants of patients with publicly available databases. AMBAR consists of a frontend module, a processing module, a visualization module and an export module. All modules can be customized and extended. Especially scripts to import and update databases are available.

Evidence-driven treatment options in molecular tumor boards can be taken from databases like Gene Drug Knowledge Database (GDKD) [8], Clinical Interpretation of Variants in Cancer database (CIViC) [9], and Tumor Alterations Relevant for Genomics-driven Therapy database (TARGET) [10]. By linking and integrating different layers of information from these databases we form a basis for variant annotations in cancer context. By using functions introduced by [15] in the MTB-Report (`https://github.com/jperera-bel/MTB-Report`) we preprocess these public databases to use in our application.

High-throughput sequencing data is processed by different tools, for example quality control (like ClinQC [16], alignment (like bwa [17]) and variant calling (like MuTect2 [18]) or copy number alteration (like SeqCNV [19]). Variant calling data is uploaded to analysis in Variant Call Format (VCF), processed by annovar [20] and parsed by maftools [21]. Copy Number Variations (CNV) are processed by pureCN [22]. To use these NGS pipeline results in AMBAR, we integrated parser and methods to process and prepare data (Table)

Besides representations of data in tables we also implemented methods for visualising alterations. Principles from [23] were used. An overview of positions and regions where mutations appear can be visualised in an ideogram plot [24]. Using packages `ggbio` [25], `GenomicRanges` [26], `BSgenome` [27] and `VariantAnnotation` [28] we generate marker for positions and regions within the human genome. `generate_visu_variants_ideo()` returns a `ggplot` object, containing a bar per chromosome with markers for centromeres and alterations. This allows a quick outline of mutation hot spots on a chromosomal level.

Large-scale analyses have shown a spectrum of mutational signatures across human cancer types. Such a mutation signature provides information about underlying causes for mutations [29]. We integrated this analysis in AMBAR by using `MutationalPatterns` [30] and Mutational Signatures (v3 - May 2019) from COSMIC [31]. In `generate_mutsig()` we construct an object containing different mutation signature plots and a mutation contribution table. For visualization in the application four parts of out mutation signature object are used: `mutsigspectrumplot` generates as combined bar plot for the single nucleotide substitution landscape. In a profile bar plot `mutsigprofileplot` trinucleotide mutation counts are displayed. `contribution_table` contains a table of the different contributions to the COSMIC Mutational Signatures (v3). A bar plot for contributions to different mutation signatures is build and saved into `contribution_barplot`.

Various export formats are provided to save, archive and version-control results. If provided, annotated variants based on the uploaded vcf file can be downloaded as tab separated value (TVS) file. Copy number variations are splitted by gene symbol and can be downloaded in comma separated value (CSV) format. Due to latest developments in health care records, Fast Healthcare Interoperability Resources (FHIR) become more and more standard as part of health care systems. These allow linkage and integration of results from different sites and centers for exchanging electronic health records [32]. We implemented an export function `generate_fhir()` for FHIR's genetic variant assessment in Extensible Markup Language (XML) format. We implemented subfunctions to build different parts of the FHIR xml scheme entries. Besides core data information, non-synonymous variants and their annotations are parsed and converted to xml entries. IDs, symbols and names are extracted from HUGO Gene Nomenclature Committee [33, 34]. For every variant Logical Observation Identifiers Names and Codes (LOINC) numbers and descriptions are appended.

To make all functions, results and visualisations public available we developed a `R shiny` web application. It is responsive, interactive and customizable. Our user interface allows input of patient's core data and upload of variant calling file (VCF) as well as copy number variations (CNV). VCFs in standard 4.1 or 4.2 format can be parsed and - with optional genotype fields - for example depth and allelic frequency can be annotated as well. CNVs are parsed from OncoScan and get annotations for gene symbols, gain, loss or loss of heterozygosity (LOH) and Cytoband as well. A table of non-synonymous variants and a table for CNVs are displayed. Combined with a linked knowledge databases (based on [15]) a third table with evidence level of possible drug targets, their effect prediction, status and publication is shown. Marking, filtering, sorting, searching, and export is possible in all result tables. Additionally SNVs and CNVs are visualized in a separate tab as ideogram plot for a quick overview of patients mutational landscape. The program also generates in a third tab a Mutation Signature (v3) of patient data [31]. The user interface is responsive, so even using it on a tablet or smart phone is possible. Results can be exported and downloaded in various formats (for example Excel, csv, tsv, or FHIR xml). Customisation of the user interface and highlighting of data fields as well as integration of other knowledge sources is possible.

Results

To illustrate the capabilities of AMBAR we added test data (`examples`-folder in AMBAR). A modified version of `colon2-sample.vcf` from the MutationalPatterns package is used as input for SNP variants

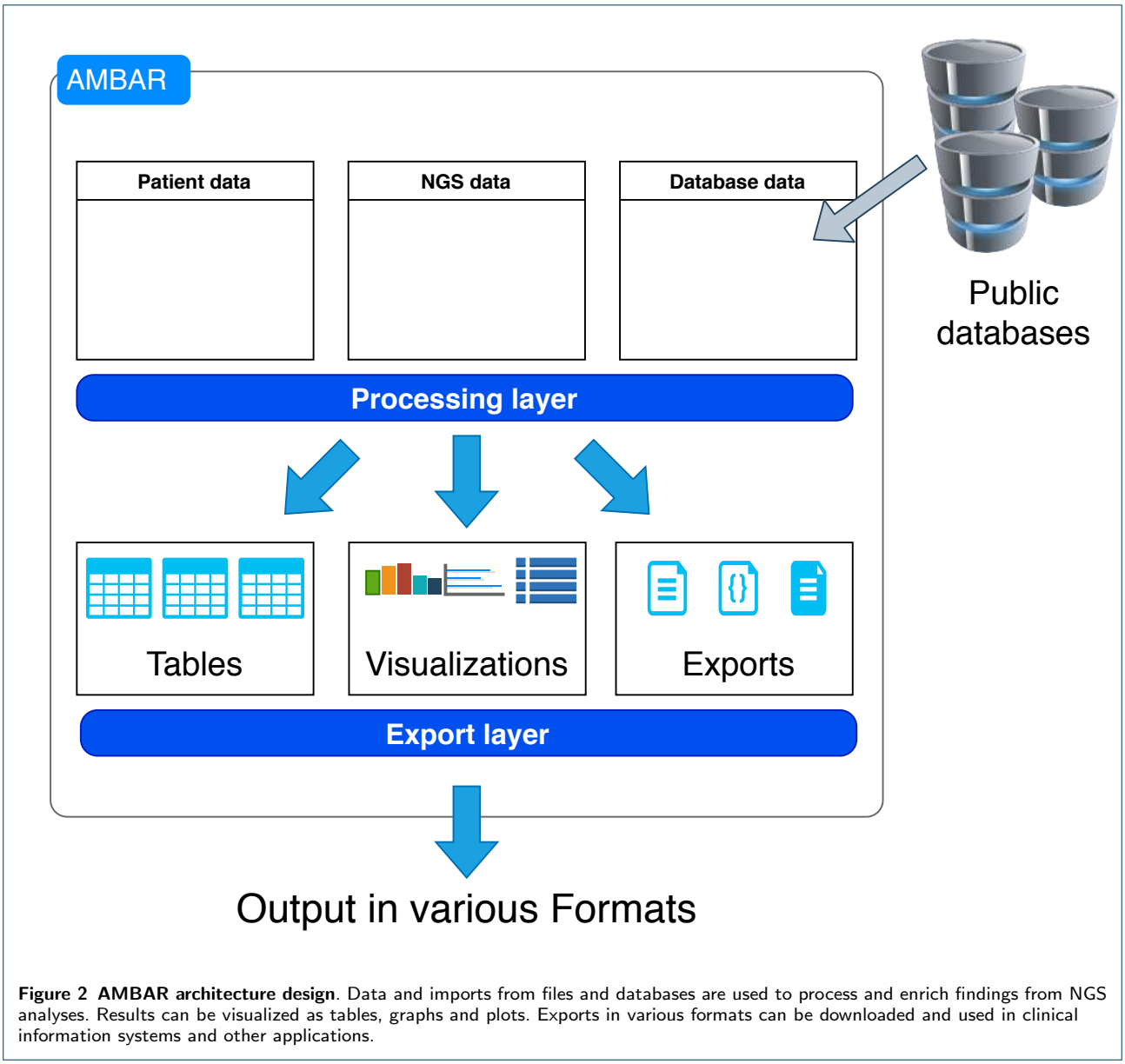


Figure 2 AMBAR architecture design. Data and imports from files and databases are used to process and enrich findings from NGS analyses. Results can be visualized as tables, graphs and plots. Exports in various formats can be downloaded and used in clinical information systems and other applications.

Table 1 AMBAR parser and methods to process input data

Method name	Functionality	R package / skript
`annovarToMaf`	import annovar output in Mutation Annotation Format (MAF)	maftools
`generate_fhir`	generates a XML FHIR file from patient data and variants	AMBAR
`generate_mtb_variants`	returns a datatable with missense variants and annotations	AMBAR
`generate_mtb_cnvs`	returns a datatable with copy number alterations per gene	AMBAR
`generate_mtb_table`	returns a datatable with drug treatment evidence per mutaion and alteration	AMBAR
`generate_mutsig`	generates different plot and tables with a Mutation Signature	MutationalPatterns
`generate_visu_variants_ideo`	generates a ideogram plot with mutations	AMBAR
`read.csv`	read comma and tab separated value files of CNVs	base
`renderDataTable`	generates a table with interactive elements for a web interface	shiny

(`example.vcf`). For CNV we generated a TSV file `examplecna.txt`. After starting the application on command line by `Rscript ambar.R` (or running it in RStudio) a `R shiny server` is started. The url with port to 127.0.0.1 (localhost) is displayed. Open a modern web browser and point it to the generated url, for example `http://127.0.0.1:3214`. An user interface to input basic patient data is shown (see Figure 3).

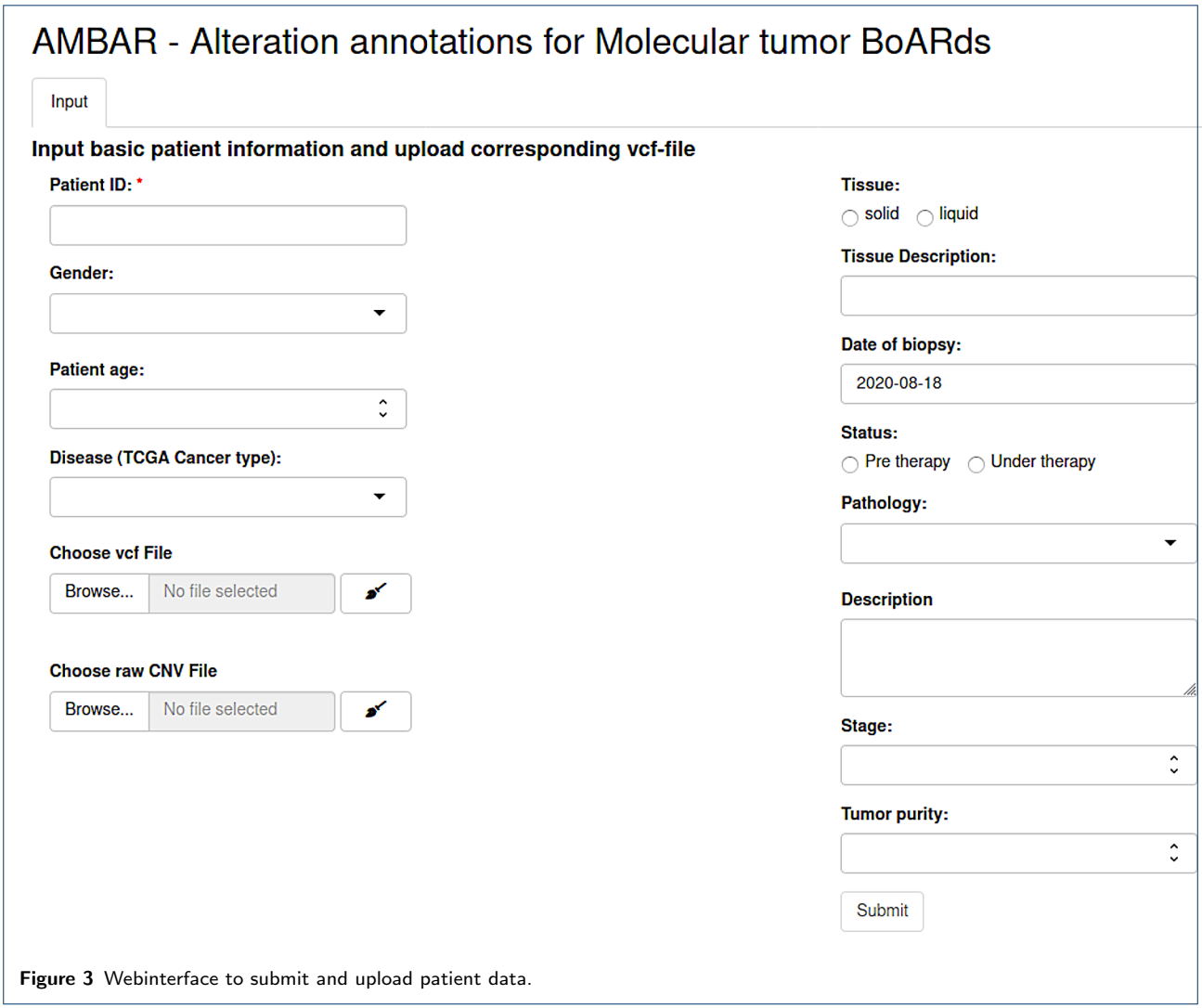


Figure 3 Webinterface to submit and upload patient data.

After submitting data to an AMBAR server, annotation and post processing steps takes place. Therefore the external tool `annovar` [20] has to be installed and a database for human genome (hg19) has to be generated beforehand. After download of the necessary files use the following commands to build the database:

```
cd <to_extracted_annovar_folder>
./annotate_variation.pl -buildver hg19 -downdb -webfrom annovar refGene humandb/
./annotate_variation.pl -buildver hg19 -downdb cytoBand humandb/
./annotate_variation.pl -buildver hg19 -downdb genomicSuperDups humandb/
./annotate_variation.pl -buildver hg19 -downdb -webfrom annovar esp6500siv2_all humandb/
./annotate_variation.pl -buildver hg19 -downdb -webfrom annovar 1000g2015aug humandb/
./annotate_variation.pl -buildver hg19 -downdb -webfrom annovar exac03 humandb/
./annotate_variation.pl -buildver hg19 -downdb -webfrom annovar avsnp147 humandb/
./annotate_variation.pl -buildver hg19 -downdb -webfrom annovar dbnsfp30a humandb/

./prepare_annovar_user.pl -dbtype cosmic CosmicMutantExport.tsv \
-vcf CosmicCodingMuts.vcf > humandb/hg19_cosmic89_coding.txt
./prepare_annovar_user.pl -dbtype cosmic CosmicNCV.tsv \
-vcf CosmicNonCodingVariants.vcf > humandb/hg19_cosmic89_noncoding.txt
```

To run the annotation from inside the web application paths to `annovar` and the `annovar` database have to be configured inside AMBAR:

```
annopath="/program/annovar"
annodb="/database/annovar/humandb"
```

Depending on uploaded data annotation process may take some time.

The application automatically switches to the result table tab called `Summary`. Depending on input basic patient data, SNP variant table, CNV table and Found Evidence table are generated and shown (Figure 4, 5, 6). All tables allow pagination, filtering, searching, sorting and exporting of all and filtered data.

Uploaded data is visualized by an ideogram on the tab `Ideogram`. Using methods from the `GenomicRanges` package, annotated variants and copy number variations are converted into `GRanges` objects. These objects are passed into the ideogram plot function. Each chromosome is marked with a centromere region in gray. Positions of SNPs are denoted as blue line inside the corresponding chromosome and regions of copy number alterations are filled in green for gain, in red

Patient SNP Variants

Information:

Variants are filtered for Proteinchange and non Silent mutations. Variant depth smaller 50 is marked in red, larger 100 in green.

Column visibility | Copy | CSV | Excel | Show 10 entries | Search:

	Chr	Gene	Type	ProteinChange	DBSNP(avsnp150)	esp6500siv2_all	1000g2015_08_ALL	Depth	VAF
			All	All	All	All	All		
1	1	ADORA1	Missense_Mutation	P222S	.	.	.	232	0.143
2	2	COL3A1	Missense_Mutation	R413W	.	.	.	221	0.143
3	11	KDM2A	Missense_Mutation	G441S	.	.	.	217	0.143
4	11	OPCML	Missense_Mutation	T146I	.	.	.	212	0.071
5	6	SYNE1	Missense_Mutation	L4300F	.	.	.	254	0.143
6	22	TBC1D10A	Missense_Mutation	A466T	rs746334856	.	.	229	0.071
7	17	TP53	Missense_Mutation	Y88S	rs121912666	.	.	236	0.143
8	17	TP53	Missense_Mutation	Y102C	rs587780073	.	.	217	0.143
9	16	TRIM72	Missense_Mutation	V311M	.	.	.	243	0.143

Showing 1 to 9 of 9 entries | Previous 1 Next

Figure 4 Table of non-synonymous mutations from uploaded vcf patient data sorted by gene symbol.

Patient CNVs

Information:

CNVs are parsed and split by gene symbol.

Column visibility | Copy | CSV | Excel | Show 10 entries | Search:

	Chr	Gene	Cytoband.Start	Cytoband.End	CN.State	Type	Full.Location	Median.Log2Ratio	Size.in.kbp.
		All	All	All	.		All	All	All
302	chr7	C7orf50	p22.3	p22.3	3	Gain	chr7:1116246-1243525	0.31802	127.279
303	chr7	GPER1	p22.3	p22.3	3	Gain	chr7:1116246-1243525	0.31802	127.279
304	chr7	ZFAND2A	p22.3	p22.3	3	Gain	chr7:1116246-1243525	0.31802	127.279
305	chr7	LOC101927021	p22.3	p22.3	3	Gain	chr7:1116246-1243525	0.31802	127.279
297	chr22	DRG1	q12.2	q12.2	1	Loss	chr22:31772776-32057501	-0.283293	284.725
298	chr22	EIF4ENIF1	q12.2	q12.2	1	Loss	chr22:31772776-32057501	-0.283293	284.725
299	chr22	SFI1	q12.2	q12.2	1	Loss	chr22:31772776-32057501	-0.283293	284.725
300	chr22	PISD	q12.2	q12.2	1	Loss	chr22:31772776-32057501	-0.283293	284.725
301	chr22	MIR7109	q12.2	q12.2	1	Loss	chr22:31772776-32057501	-0.283293	284.725
1	chr21	MIR3648-1	p11.2	q22.3	3	Gain	chr21:9648314-45121987	0.2727594	35473.673

Showing 1 to 10 of 305 entries | Previous 1 2 3 4 5 … 31 Next

Figure 5 Table of copy number variations from uploaded cnv patient data sorted by chromosome.

for loss, and in orange for loss of heterozygosity (see Figure 7).

Information about underlying causes for mutations can be described and summarised by a mutational signature. As this can influence treatment drug choices

Found Evidence

Evidence Level:

L01-L05 = same cancer with Evidence FDA-approved / NCCN guidelines, late trials, early trials, case report and preclinical

L06-L10 = different cancer with Evidence FDA-approved / NCCN guidelines, late trials, early trials, case report and preclinical

Column visibility | Copy | CSV | Excel | Show 10 entries | Search:

	Gene	Pat. Var	Cancer	Known Var	Predicts	Drugs	Evidence	Citation	Source	Level
	All	All	All	All	All	All	All	All	All	All
4	KRAS		colorectal	wild type	response	cetuximab, panitumumab	FDA-approved	FDA	FDA	L01
5	KIT		gastric (stromal)	wild type	decreased sensitivity	imatinib	late trials	18955458	PubMed	L02
7	KIT		gastric (stromal)	wild type	response	sunitinib	late trials	18955458	response	L02
6	PDGFRA		gastric (stromal)	wild type	decreased sensitivity	imatinib	late trials	14645423, 18955458	PubMed	L02
8	PDGFRA		gastric (stromal)	wild type	response	sunitinib	late trials	18955458	response	L02
2	KIT		gastric (stromal)	wild type	response	sorafenib	early trials	ASCO 2011 (abstr 10009)	sensitivity	L03
3	PDGFRA		gastric (stromal)	wild type	response	sorafenib	early trials	ASCO 2011 (abstr 10009)	sensitivity	L03
1	TP53		AML	wild type (LoF)	response	HDM2 inhibitor	early trials	AACR 2017 (abstr CT152)	AACR 2017 (abstr CT152)	L03
9	KIT		gastric (stromal)	wild type	sensitivity	dasatinib	preclinical	16397263		L05
10	PDGFRA		gastric (stromal)	wild type	sensitivity	dasatinib	preclinical	16397263		L05

Figure 6 Table of found evidence calculated from SNVs and CNVs based on selected cancer type and sorted by evidence level.

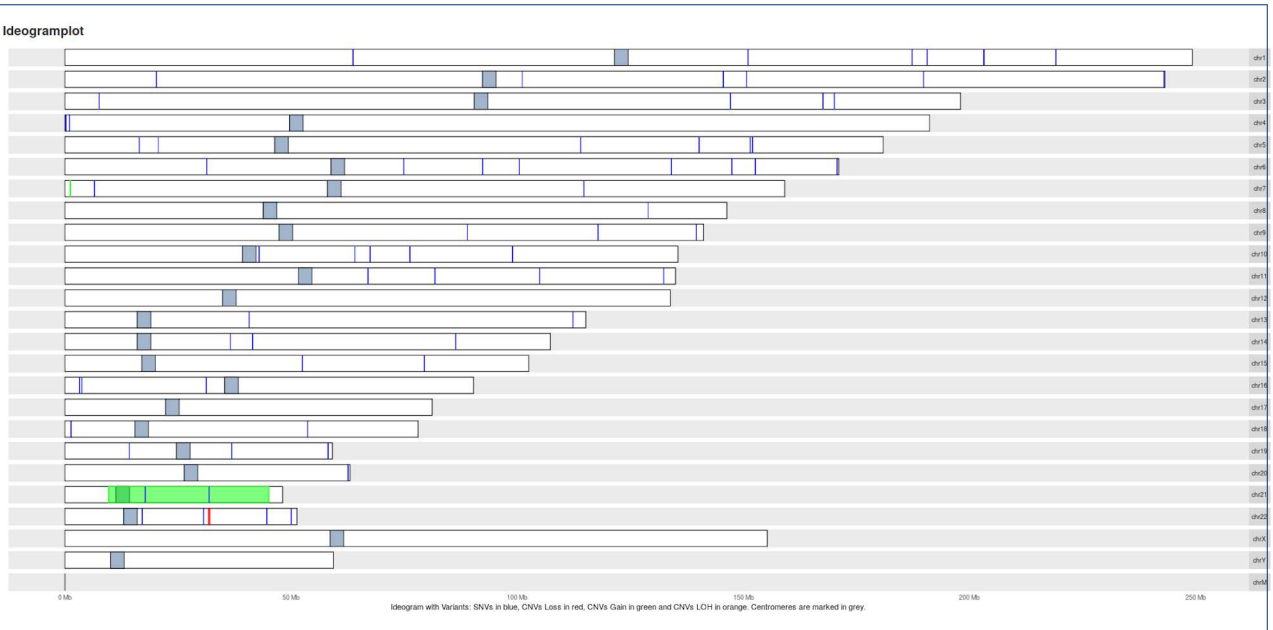


Figure 7 Ideogram with markers for centromere (gray), SNPs (blue), and copy number alterations (green = gain, red = loss, orange = loss of heterozygosity.

we show overview plots and a signature contribution table in the tab `Mutation Signature`. The first plot in the tab shows a spectrum plot as stacked bar plot with relative contributions of single nucleotide mutations (see Figure 8). The second bar plot displays the relative contributions of trinucleotide mutation counts of the patient's variants (see Figure 9). The last two objects on the web tab are a table and a stacked bar plot with the signature contributions (`>0`), based on COSMIC Mutational Signatures (v3, May 2019) (see Figure 10 and 11).

Users of AMBAR can save the results in different formats. In the `|Downloads|`-tab of the web interface, results of the annotation step for vcf data as tsv file, and the split-by-gene results for cnv data as csv file is available (`downloadRawVAR` and `downloadRawCNV`). An export `generate_fhir()` for FHIR's genetic variant assessment in Extensible Markup Language (XML) format is available as well. As this resource is still in development, we added generic functions to generate xml entries. The basic structure consists of following:

```
fhirxml <- read_xml("add/base.xml")

subNode <- xml_child(fhirxml, "d1:subject")
refNode <- xml_child(subNode, "d1:reference")
xml_set_attr(refNode, "value", paste("Patient:",pat_ID))
subNode <- xml_child(fhirxml, "d1:performer")
refNode <- xml_child(subNode, "d1:reference")
xml_set_attr(refNode, "value", "Institution")

subNode <- xml_child(fhirxml, "d1:specimen")
idNode <- xml_child(subNode, "d1:identifier")
sysNode <- xml_child(idNode, "d1:system")
valNode <- xml_child(idNode, "d1:value")
xml_set_attr(sysNode, "value", "url_as_institution_identifier")
xml_set_attr(valNode, "value", "TISSUEID")

variants <- xml_add_child(fhirxml, "variants")
```

For every variant a block of function calls builds the xml FHIR entry, containing Logical Observation Identifiers Names and Codes (LOINC), numbers and descriptions. Exported xml files in FHIR format can be imported into data integration systems or FHIR stores to enrich patient data.

Conclusions

We developed an interactive, highly customizable and configurable R-shiny application AMBAR to support molecular tumor boards in analyzing patients molecular data and to optimize personalized cancer treatment. AMBAR allows annotation, visualization and refinement of imported alterations and export of results in various formats. Integration in clinical information systems is possible, especially adaption of new FHIR standards for genetic variants. As AMBAR is a web-based application, it is platform-independent. In addition, modern web technologies make AMBAR responsive and reactive. Hence, displaying and sharing results on monitors or tablets is easy and can be helpful during interdisciplinary meetings of MTBs. As no external connection to web resources after the preparation steps is necessary, self-hosting AMBAR within a restricted and secure network of a hospital is possible. Different issues have to be faced when personalized medicine approaches are implemented. First, when analyzing patients alterations data, available databases and tools are queried often one by one. Moreover, the alteration report is frequently written by hand or filled out in standardized forms that then have to be further digitalized. With AMBAR, we provide a unique tool that allows annotation and visualization of NGS patients data together with clinical and molecular parameters taking into account information available in the databases. Moreover, results in AMBAR can be explored interactively and further filtered. Finally, alteration reports are exported and presented in a digitalized form. Altogether, we presented here our R-shiny application, AMBAR, that tackles significant issues encountered by Molecular Tumor Boards in discussing and have the goal of expanding the state of precision medicine treatment approaches.

Availability and requirements

- Project name: AMBAR
- Project home page: `https://sysbio.uni-ulm.de/?Software:Ambar`
- Operating system(s): Platform independent
- Programming language: R (version 4.0.0 (2020-04-24)), `x86_64-pc-linux-gnu`
- Other requirements: Perl (v5.30.0)
- License: e.g. GNU GPL Version 3
- Any restrictions to use by non-academics: none
- Base R packages: base, datasets, graphics, grDevices, methods, parallel, stats, stats4, utils
- Other R packages: AnnotationDbi 1.50.0, Biobase 2.48.0, BiocGenerics 0.34.0, Biostrings 2.56.0, BSgenome 1.56.0, BSgenome.Hsapiens.UCSC.hg19 1.4.3, Cairo 1.5-12, cluster 2.1.0, data.table 1.12.8, DelayedArray 0.14.0, DNAcopy 1.62.0, dplyr 0.8.5, DT 0.13, formattable 0.2.0.1, g3viz 1.1.2, GenomeInfoDb 1.24.0, GenomicFeatures 1.40.0, GenomicRanges 1.40.0, ggbio 1.36.0, ggplot2 3.3.0, GO.db 3.11.1, Homo.sapiens 1.3.1, IRanges 2.22.1, knitr 1.28, maftools 2.4.0, matrixStats 0.56.0, MutationalPatterns 2.0.0, NMF 0.22.0, org.Hs.eg.db 3.11.1, OrganismDbi 1.30.0, pander 0.6.3, pkgmaker 0.31.1, PureCN 1.18.0, registry 0.5-1, rngtools 1.5, Rsamtools 2.4.0, rtracklayer 1.48.0, S4Vectors 0.26.1, shiny 1.4.0.2, shinycssloaders 0.3,

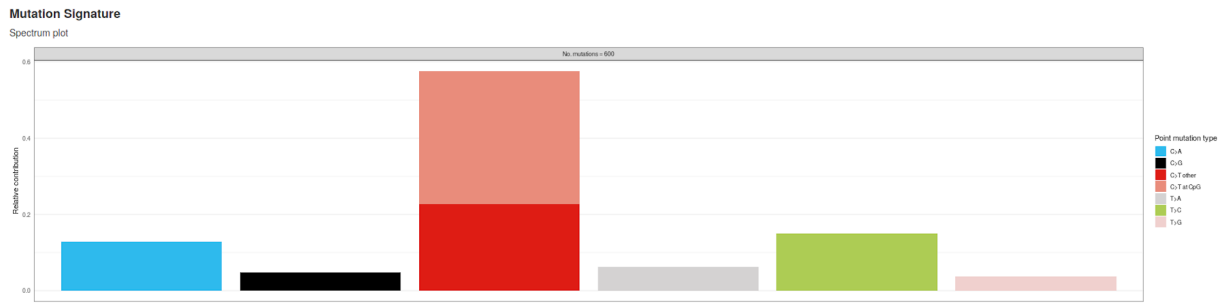


Figure 8 Relative contribution of single nucleotide substitutions are shown. Substitutions of cytosine (C) with thymine (T) C→T are splitted into CpG islands and other positions. The number of point mutations is shown as well.

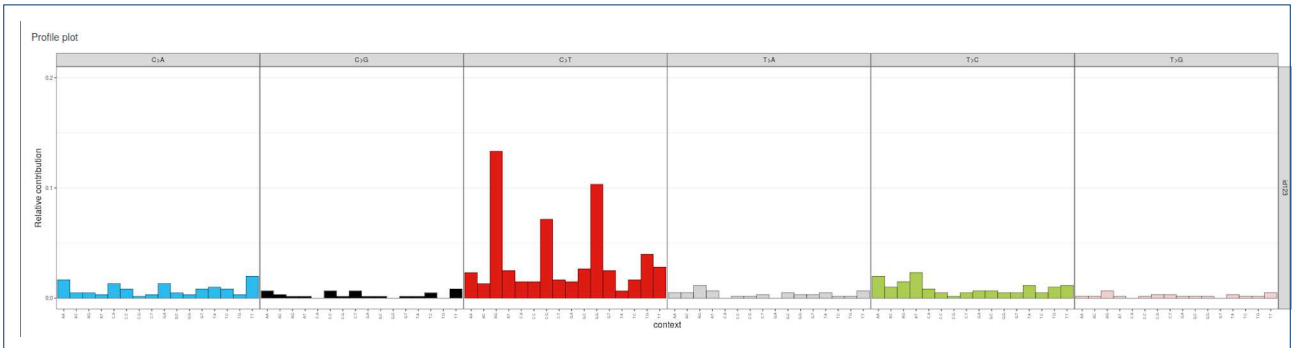


Figure 9 The profile plot shows the relative contribution of single nucleotide mutations on the context of trinucleotides.

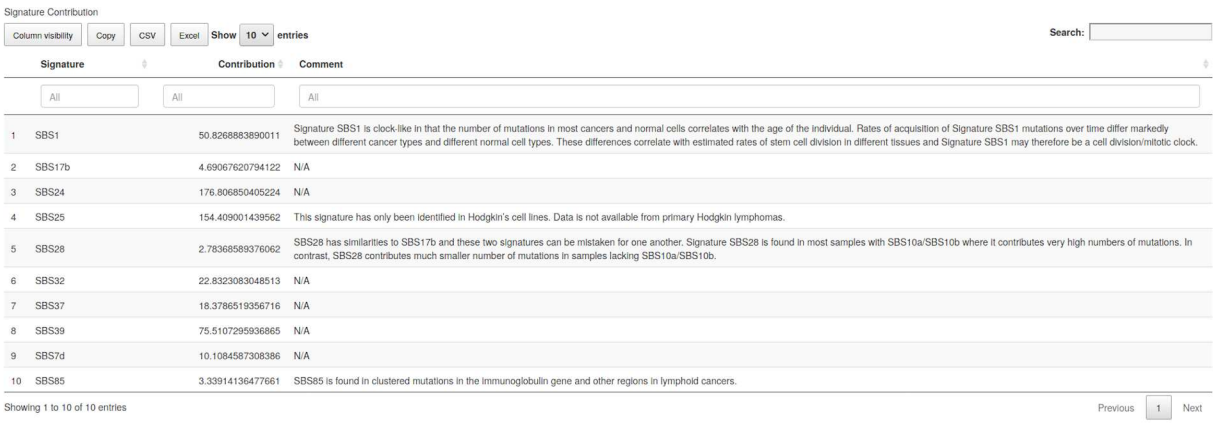


Figure 10 Signature contribution: The table shows the calculated signature contribution in comparison to the 67 Single Base Substitution (SBS) Mutational Signatures from COSMIC v3 (May 2019) with their comments. The table is filtered for signatures with zero fit to a signature.

shinycustomloader 0.9.0, shinyjs 1.1, shinythemes 1.1.2, shinyWidgets 0.5.2, stringr 1.4.0, SummarizedExperiment 1.18.1, tidyr 1.0.3, TxDb.Hsapiens.UCSC.hg19.knownGene 3.2.2, VariantAnnotation 1.34.0, vcfR 1.10.0, xml2 1.3.2, xtable 1.8-4, XVector 0.28.0

List of abbreviations

AMBAR: Alteration annotations for Molecular tumor BoARds; CIViC: Clinical interpretation of variants in cancer database; CNV: Copy number variation; COSMIC: Catalogue of somatic mutations in cancer; CSV: comma separated value; FHIR: Fast healthcare interoperability resources; GDKD: Gene drug knowledge

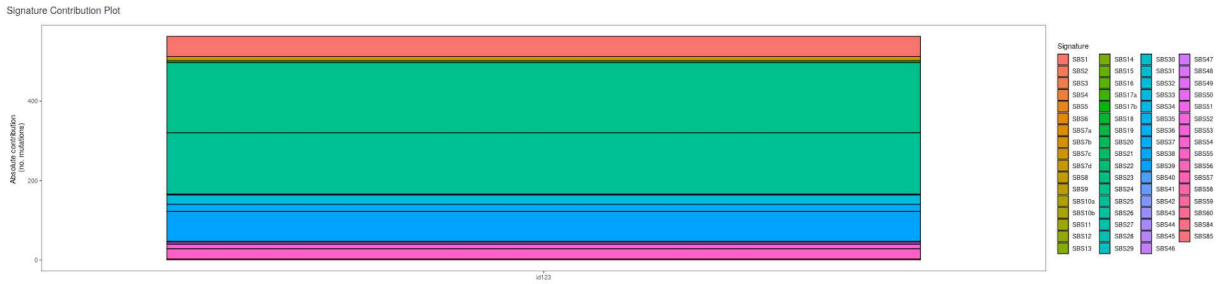


Figure 11 Analog to the signature contribution table, a stacked bar plot is generated, showing the absolute contributions to every Single Base Substitution (SBS) Mutational Signature from COSMIC v3 (May 2019). The plotted stack is sorted ascending by the SBS numbers.

database; HUGO:Human genome organisation; LOH: Loss of heterozygosity; LOINC: Logical observation identifiers names and codes; MAF: Mutation annotation format; MTB: Molecular tumor boards; NGS: Next generation sequencing; PDF: Portable document format; SNV: Single nucleotide variation; SNP: Single nucleotide polymorphism; TARGET: Tumor alterations relevant for genomics driven therapy database; TSV: Tab separated value; VCF: Variant call format; XML: Extensible Markup Language

Declarations

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Availability of data and materials
All data used in the application is freey available in the ZIP-archive of AMBAR at `https://sysbio.uni-ulm.de/?Software:Ambar`

Funding
HAK acknowledges funding from the Germany Federal Ministry of Education and Research (BMBF) as part of the DIFUTURE project (Medical Informatics Initiative, grant number 01ZZ1804I). HAK acknowledges funding from the Ministry of Science and Art Baden-Württemberg (Zentrum für Innovative Versorgung, ZIV). TS acknowledges funding from the Ministry of Social Affairs of Baden-Württemberg (Zentren für Personalisierte Medizin, ZPM).

Acknowledgements
None

Competing interests
The authors declare that they have no competing interests.

Author's contributions
AF is the lead software designer, developer, and implementer of AMBAR; NI, AMRK, RM and TS contributed to the UI of the software; HAK supervised and revised the manuscript; All authors approved the manuscript and agreed to be accountable for their corresponding part of the work.

Author details
[1]Institute of Medical Systems Biology, Ulm University, Albert-Einstein-Allee 11, 89081 Ulm, Germany. [2]Department of Internal Medicine I, Ulm University Hospital, Albert-Einstein-Allee 23, 89081 Ulm, Germany. [3]Department of Pathology, Ulm University Hospital, Albert-Einstein-Allee 23, 89081 Ulm, Germany.

References

1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin D, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. International Journal of Cancer. 2019;144(8):1941–1953.
2. Sonnenschein C, Soto AM. Carcinogenesis explained within the context of a theory of organisms. Progress in Biophysics and Molecular Biology. 2016;122(1):70–76.
3. Shin SH, Bode AM, Dong Z. Addressing the challenges of applying precision oncology. NPJ Precision Oncology. 2017;1(1):1–10.
4. Hovelson DH, McDaniel AS, Cani AK, Johnson B, Rhodes K, Williams PD, et al. Development and Validation of a Scalable Next-Generation Sequencing System for Assessing Relevant Somatic Variants in Solid Tumors. Neoplasia. 2015;17(4):385–399.
5. Loomans-Kropp HA, Umar A. Cancer prevention and screening: the next step in the era of precision medicine. NPJ Precision Oncology. 2019;3(1):1–8.
6. Bode AM, Dong Z. Recent advances in precision oncology research. NPJ Precision Oncology. 2018;2(1):11.
7. Singer J, Ruscheweyh HJ, Hofmann AL, Thurnherr T, Singer F, Toussaint NC, et al. NGS-pipe: a flexible, easily extendable and highly configurable framework for NGS analysis. Bioinformatics. 2018;34(1):107–108.
8. Dienstmann R, Jang IS, Bot B, Friend S, Guinney J. Database of genomic biomarkers for cancer drugs and clinical targetability in solid tumors. Cancer Discovery. 2015;5(2):118–123.
9. Griffith M, Spies CN, Krysiak K, McMichael FJ, Coffman CA, Danos MA, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nature Genetics. 2017;49(2):170–174.
10. Le Tourneau C, Delord JP, Gonçalves A, Gavoille C, Dubot C, Isambert N, et al. Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. The Lancet Oncology. 2015;16(13):1324–1334.
11. Tong D, Tian Y, Zhou T, Ye Q, Li J, Ding K, et al. Improving prediction performance of colon cancer prognosis based on the integration of clinical and multi-omics data. BMC Medical Informatics and Decision Making. 2020;20(1):22.
12. Kraus JM, Lausser L, Kuhn P, Jobst F, Bock M, Halanke C, et al. Big data and precision medicine: challenges and strategies with healthcare data. International Journal of Data Science and Analytics. 2018;6(3):241–249.
13. Mahato K, Srivastava A, Chandra P. Paper based diagnostics for personalized health care: Emerging technologies and commercial aspects. Biosensors and Bioelectronics. 2017;96:246–259.
14. Singer F, Irmisch A, Toussaint CN, Grob L, Singer J, Thurnherr T, et al. SwissMTB: establishing comprehensive molecular cancer diagnostics in Swiss clinics. BMC Medical Informatics and Decision Making. 2018;18(1):89.
15. Perera-Bel J, Hutter B, Heining C, Bleckmann A, Fröhlich M, Fröhling

S, et al. From somatic variants towards precision oncology: Evidence-driven reporting of treatment options in molecular tumor boards. Genome Medicine. 2018;10(1):1–15.
16. Pandey RV, Pabinger S, Kriegner A, Weinhäusel A. ClinQC: a tool for quality control and cleaning of Sanger and NGS data in clinical research. BMC Bioinformatics. 2016;17(1):56.
17. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010;26(5):589–595.
18. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling Somatic SNVs and Indels with Mutect2. BioRxiv. 2019;p. 861054.
19. Chen Y, Zhao L, Wang Y, Cao M, Gelowani V, Xu M, et al. SeqCNV: a novel method for identification of copy number variations in targeted next-generation sequencing data. BMC Bioinformatics. 2017;18(1):1–9.
20. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research. 2010 07;38(16):e164–e164.
21. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Research. 2018;28(11):1747–1756.
22. Riester M, Singh AP, Brannon AR, Yu K, Campbell CD, Chiang DY, et al. PureCN: copy number calling and SNV classification using targeted short read sequencing. Source Code for Biology and Medicine. 2016;11(1):13.
23. Nolan D, Lang DT. Interactive and Animated Scalable Vector Graphics and R Data Displays. Journal of Statistical Software, Articles. 2012;46(1):1–88.
24. Müller A, Holzmann K, Kestler HA. Visualization of genomic aberrations using Affymetrix SNP arrays. Bioinformatics. 2006 11;23(4):496–497.
25. Yin T, Cook D, Lawrence M. ggbio: an R package for extending the grammar of graphics for genomic data. Genome Biology. 2012;13(8):R77.
26. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. PLoS Computational Biology. 2013;9(8).
27. Carlson MR, Pagès H, Arora S, Obenchain V, Morgan M. Genomic annotation resources in R/Bioconductor. In: Mathé E, Davis S, editors. Statistical Genomics. Methods in Molecular Biology. vol. 1418. Humana Press; 2016. p. 67–90.
28. Obenchain V, Lawrence M, Carey V, Gogarten S, Shannon P, Morgan M. VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. Bioinformatics. 2014;30(14):2076–2078.
29. Wojtowicz D, Sason I, Huang X, Kim YA, Leiserson MMD, Przytycka MT, et al. Hidden Markov models lead to higher resolution maps of mutation signature activity in cancer. Genome Medicine. 2019;11(1).
30. Blokzijl F, Janssen R, Van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Medicine. 2018;10(1):33.
31. Alexandrov BL, Kim J, Haradhvala JN, Huang NM, Tian Ng WA, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94–101.
32. Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. In: Proceedings of the 26th IEEE International Symposium on Computer-based Medical Systems. IEEE; 2013. p. 326–331.
33. Povey S, Lovering R, Bruford E, Wright M, Lush M, Wain H. The HUGO gene nomenclature committee (HGNC). Human Genetics. 2001;109(6):678–680.
34. Eyre TA, Ducluzeau F, Sneddon TP, Povey S, Bruford EA, Lush MJ. The HUGO gene nomenclature database, 2006 updates. Nucleic Acids Research. 2006;34(suppl_1):D319–D321.

Figures

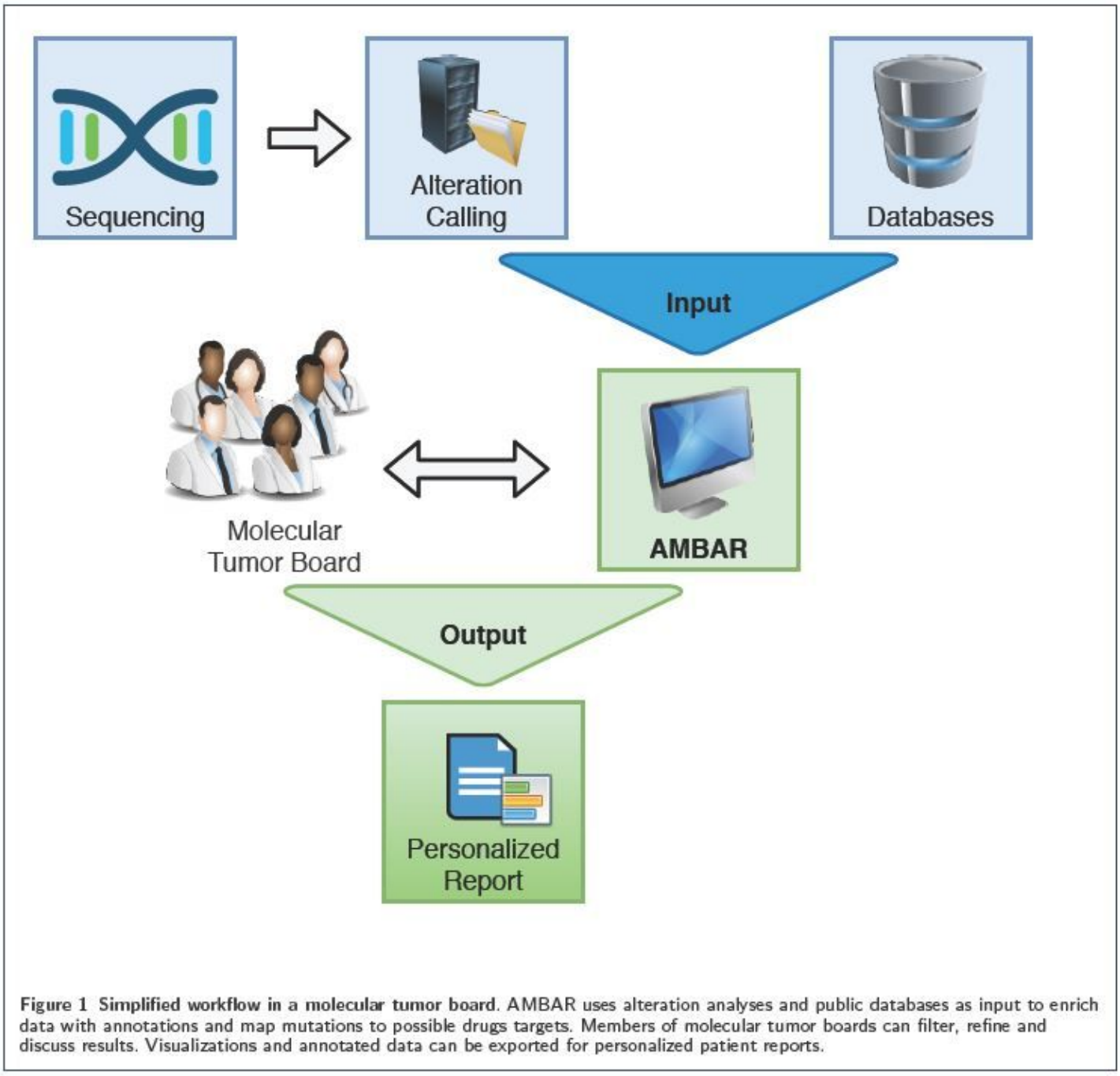


Figure 1 Simplified workflow in a molecular tumor board. AMBAR uses alteration analyses and public databases as input to enrich data with annotations and map mutations to possible drugs targets. Members of molecular tumor boards can filter, refine and discuss results. Visualizations and annotated data can be exported for personalized patient reports.

Figure 1

Simplified work ow in a molecular tumor board. AMBAR uses alteration analyses and public databases as input to enrich data with annotations and map mutations to possible drugs targets. Members of molecular tumor boards can filter, refine and discuss results. Visualizations and annotated data can be exported for personalized patient reports.

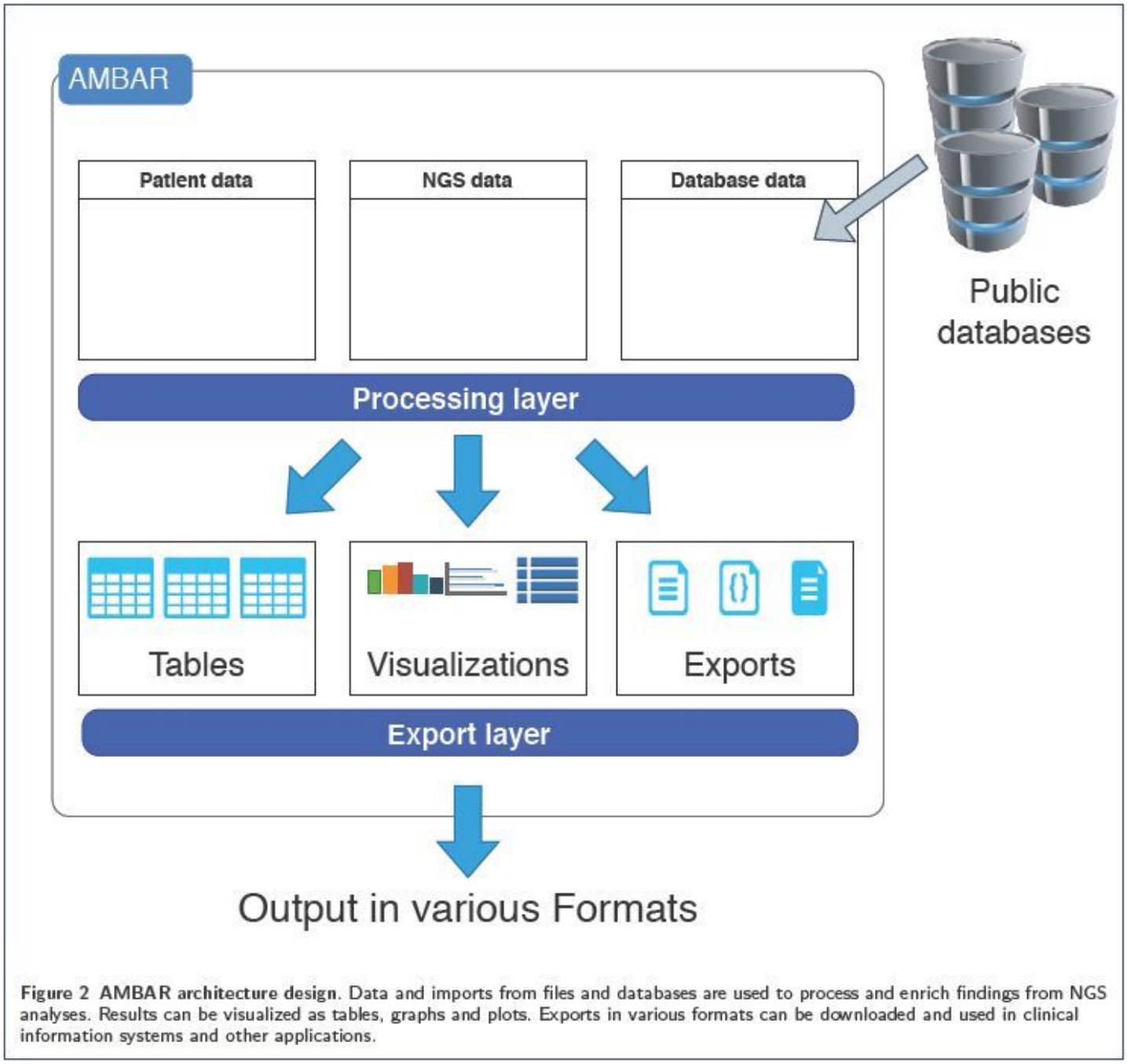


Figure 2 AMBAR architecture design. Data and imports from files and databases are used to process and enrich findings from NGS analyses. Results can be visualized as tables, graphs and plots. Exports in various formats can be downloaded and used in clinical information systems and other applications.

Figure 2

AMBAR architecture design. Data and imports from files and databases are used to process and enrich findings from NGS analyses. Results can be visualized as tables, graphs and plots. Exports in various formats can be downloaded and used in clinical information systems and other applications.

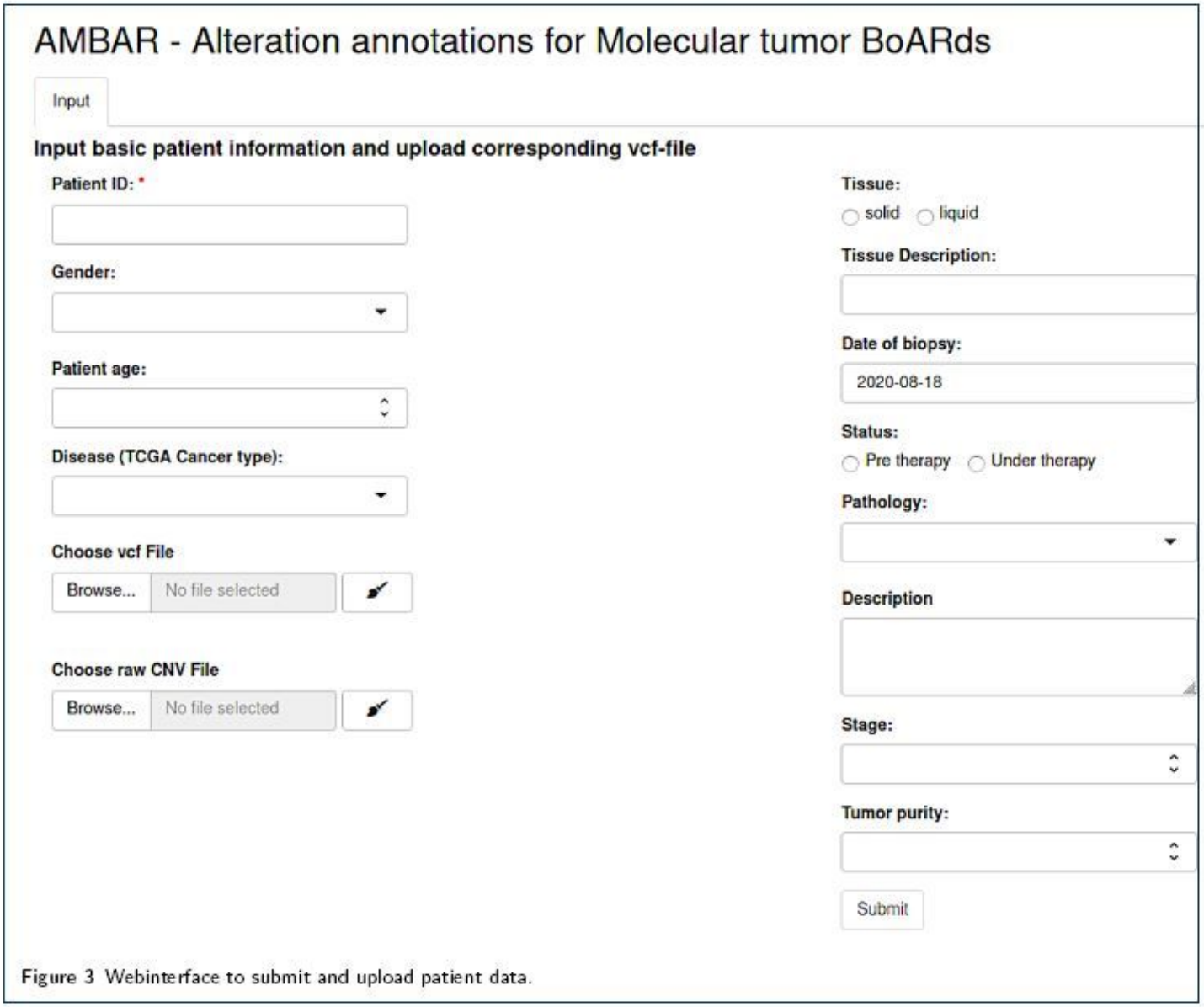
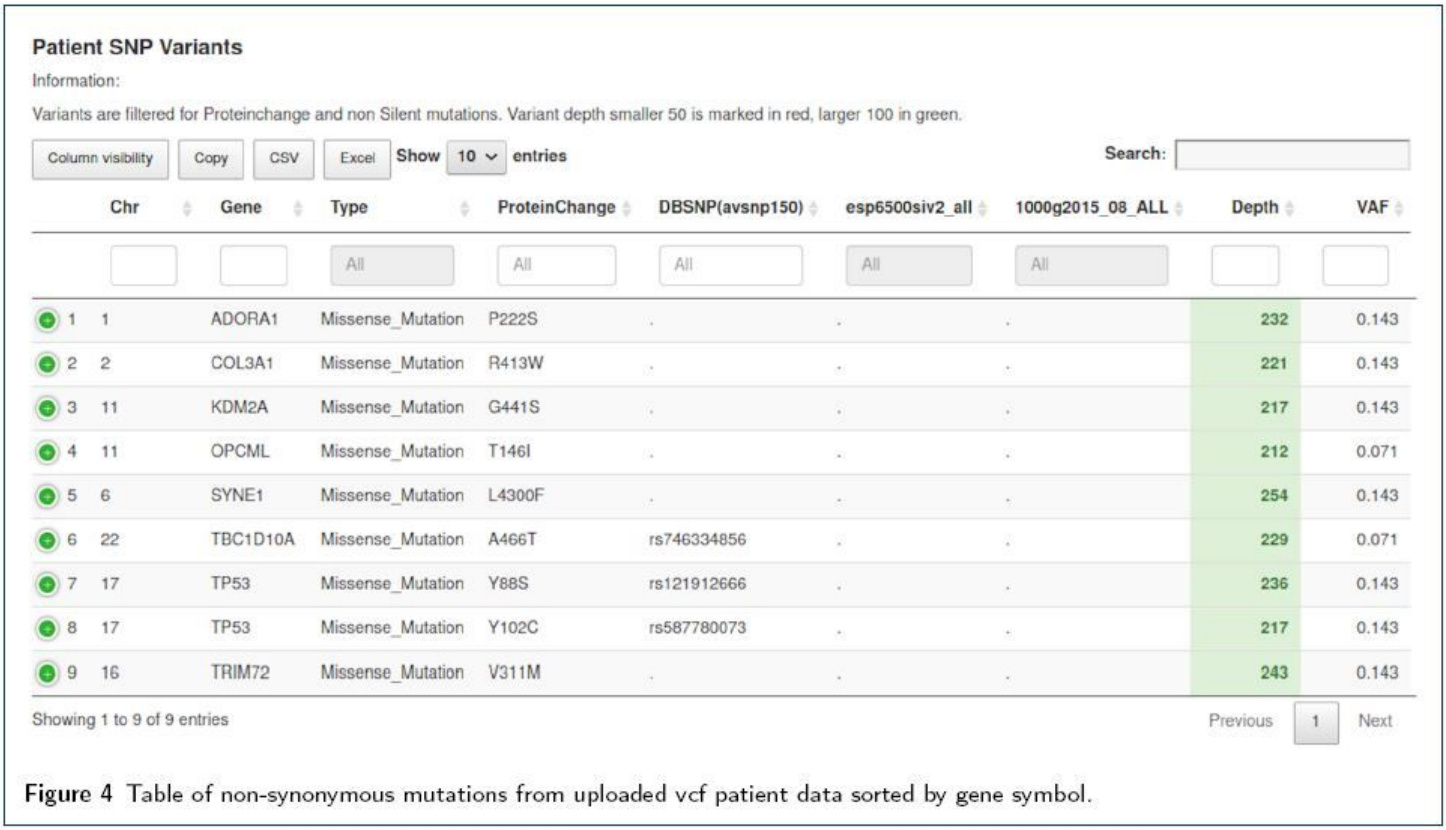


Figure 3 Webinterface to submit and upload patient data.

Figure 3

Webinterface to submit and upload patient data.



Patient SNP Variants

Information:

Variants are filtered for Proteinchange and non Silent mutations. Variant depth smaller 50 is marked in red, larger 100 in green.

Column visibility | Copy | CSV | Excel | Show 10 entries | Search:

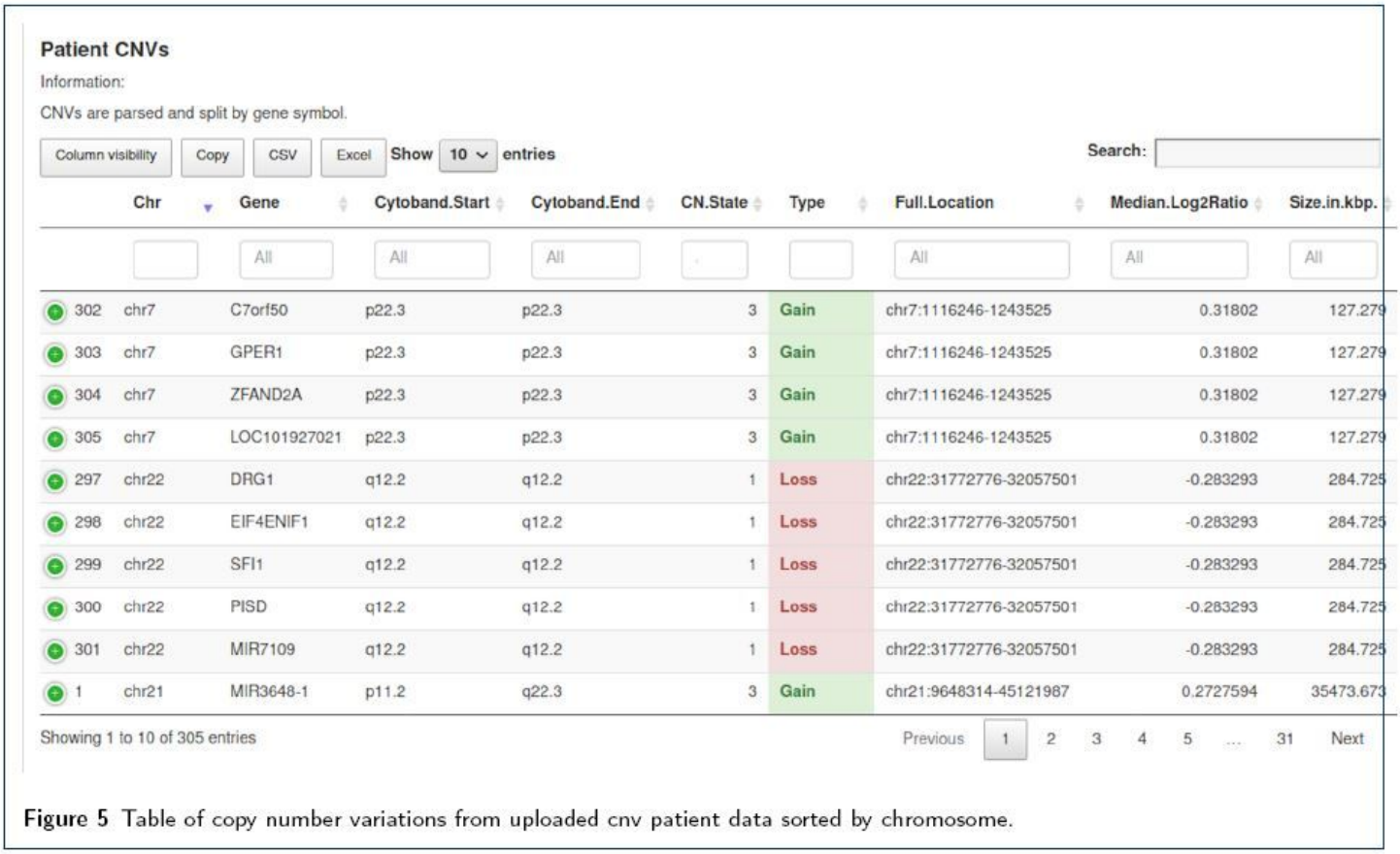
	Chr	Gene	Type	ProteinChange	DBSNP(avsnp150)	esp6500siv2_all	1000g2015_08_ALL	Depth	VAF
			All	All	All	All	All		
1	1	ADORA1	Missense_Mutation	P222S	.	.	.	232	0.143
2	2	COL3A1	Missense_Mutation	R413W	.	.	.	221	0.143
3	11	KDM2A	Missense_Mutation	G441S	.	.	.	217	0.143
4	11	OPCML	Missense_Mutation	T146I	.	.	.	212	0.071
5	6	SYNE1	Missense_Mutation	L4300F	.	.	.	254	0.143
6	22	TBC1D10A	Missense_Mutation	A466T	rs746334856	.	.	229	0.071
7	17	TP53	Missense_Mutation	Y88S	rs121912666	.	.	236	0.143
8	17	TP53	Missense_Mutation	Y102C	rs587780073	.	.	217	0.143
9	16	TRIM72	Missense_Mutation	V311M	.	.	.	243	0.143

Showing 1 to 9 of 9 entries | Previous | 1 | Next

Figure 4 Table of non-synonymous mutations from uploaded vcf patient data sorted by gene symbol.

Figure 4

Table of non-synonymous mutations from uploaded vcf patient data sorted by gene symbol.



Patient CNVs

Information:

CNVs are parsed and split by gene symbol.

Column visibility | Copy | CSV | Excel | Show 10 entries | Search:

	Chr	Gene	Cytoband.Start	Cytoband.End	CN.State	Type	Full.Location	Median.Log2Ratio	Size.in.kbp.
		All	All	All			All	All	All
302	chr7	C7orf50	p22.3	p22.3	3	Gain	chr7:1116246-1243525	0.31802	127.279
303	chr7	GPER1	p22.3	p22.3	3	Gain	chr7:1116246-1243525	0.31802	127.279
304	chr7	ZFAND2A	p22.3	p22.3	3	Gain	chr7:1116246-1243525	0.31802	127.279
305	chr7	LOC101927021	p22.3	p22.3	3	Gain	chr7:1116246-1243525	0.31802	127.279
297	chr22	DRG1	q12.2	q12.2	1	Loss	chr22:31772776-32057501	-0.283293	284.725
298	chr22	EIF4ENIF1	q12.2	q12.2	1	Loss	chr22:31772776-32057501	-0.283293	284.725
299	chr22	SFI1	q12.2	q12.2	1	Loss	chr22:31772776-32057501	-0.283293	284.725
300	chr22	PISD	q12.2	q12.2	1	Loss	chr22:31772776-32057501	-0.283293	284.725
301	chr22	MIR7109	q12.2	q12.2	1	Loss	chr22:31772776-32057501	-0.283293	284.725
1	chr21	MIR3648-1	p11.2	q22.3	3	Gain	chr21:9648314-45121987	0.2727594	35473.673

Showing 1 to 10 of 305 entries

Previous 1 2 3 4 5 … 31 Next

Figure 5 Table of copy number variations from uploaded cnv patient data sorted by chromosome.

Figure 5

Table of copy number variations from uploaded cnv patient data sorted by chromosome.

Found Evidence

Evidence Level:

L01-L05 = same cancer with Evidence FDA-approved / NCCN guidelines, late trials, early trials, case report and preclinical

L06-L10 = different cancer with Evidence FDA-approved / NCCN guidelines, late trials, early trials, case report and preclinical

Column visibility | Copy | CSV | Excel | Show 10 entries | Search:

	Gene	Pat. Var	Cancer	Known Var	Predicts	Drugs	Evidence	Citation	Source	Level
	All	All	All	All	All	All	All	All	All	All
4	KRAS		colorectal	wild type	response	cetuximab, panitumumab	FDA-approved	FDA	FDA	L01
5	KIT		gastric (stromal)	wild type	decreased sensitivity	imatinib	late trials	18955458	PubMed	L02
7	KIT		gastric (stromal)	wild type	response	sunitinib	late trials	18955458	response	L02
6	PDGFRA		gastric (stromal)	wild type	decreased sensitivity	imatinib	late trials	14645423, 18955458	PubMed	L02
8	PDGFRA		gastric (stromal)	wild type	response	sunitinib	late trials	18955458	response	L02
2	KIT		gastric (stromal)	wild type	response	sorafenib	early trials	ASCO 2011 (abstr 10009)	sensitivity	L03
3	PDGFRA		gastric (stromal)	wild type	response	sorafenib	early trials	ASCO 2011 (abstr 10009)	sensitivity	L03
1	TP53		AML	wild type (LoF)	response	HDM2 inhibitor	early trials	AACR 2017 (abstr CT152)	AACR 2017 (abstr CT152)	L03
9	KIT		gastric (stromal)	wild type	sensitivity	dasatinib	preclinical	16397263		L05
10	PDGFRA		gastric (stromal)	wild type	sensitivity	dasatinib	preclinical	16397263		L05

Figure 6 Table of found evidence calculated from SNVs and CNVs based on selected cancer type and sorted by evidence level.

Figure 6

Table of found evidence calculated from SNVs and CNVs based on selected cancer type and sorted by evidence level.

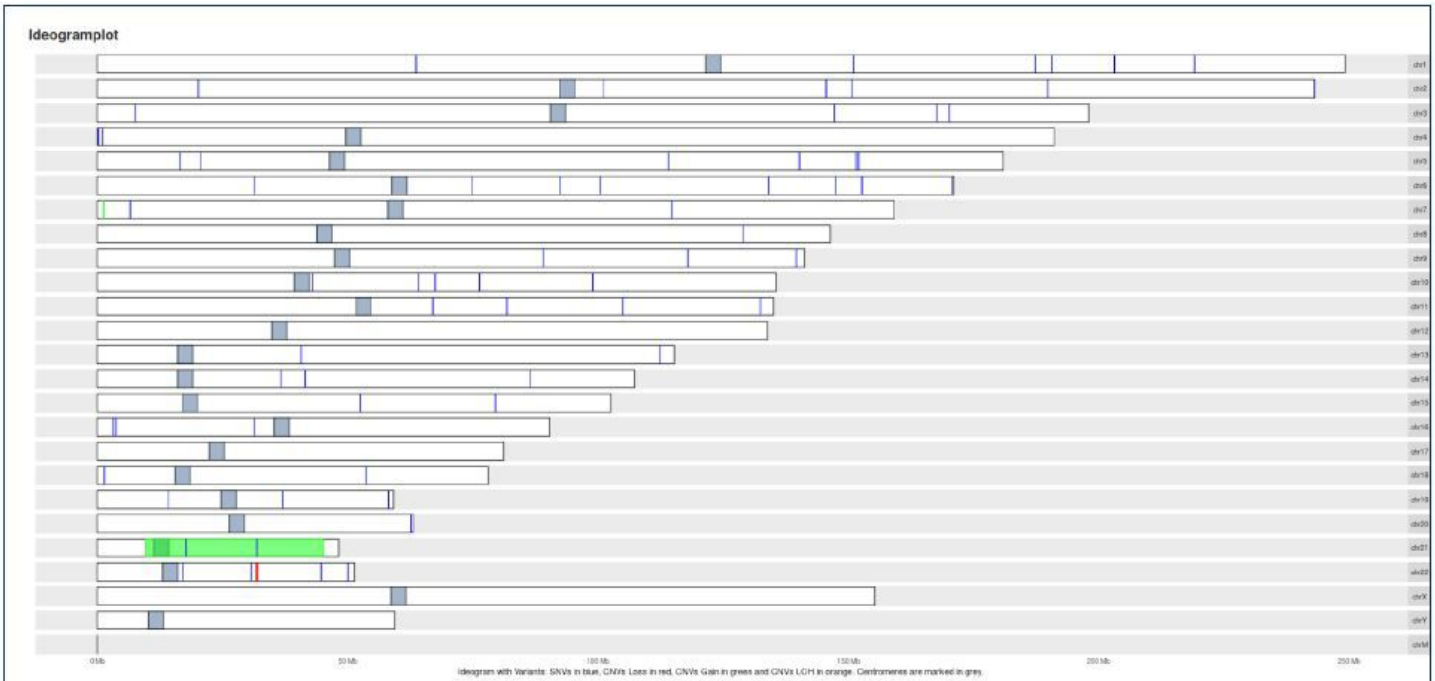


Figure 7 Ideogram with markers for centromere (gray), SNPs (blue), and copy number alterations (green = gain, red = loss, orange = loss of heterozygosity.

Figure 7

Ideogram with markers for centromere (gray), SNPs (blue), and copy number alterations (green = gain, red = loss, orange = loss of heterozygosity.

Figure 8 Relative contribution of single nucleotide substitutions are shown. Substitutions of cytosine (C) with thymine (T) C→T are splitted into CpG islands and other positions. The number of point mutations is shown as well.

Figure 8

Relative contribution of single nucleotide substitutions are shown. Substitutions of cytosine (C) with thymine (T) C -> T are splitted into CpG islands and other positions. The number of point mutations is shown as well.

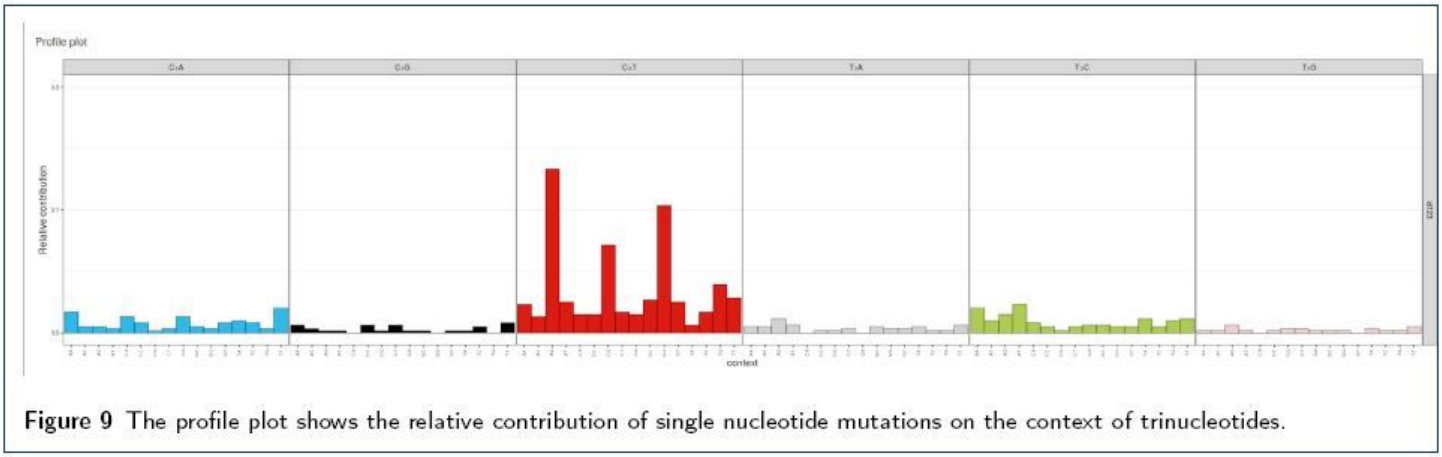


Figure 9 The profile plot shows the relative contribution of single nucleotide mutations on the context of trinucleotides.

Figure 9

The profile plot shows the relative contribution of single nucleotide mutations on the context of trinucleotides.

Signature Contribution

Column visibility | Copy | CSV | Excel | Show 10 entries | Search:

	Signature	Contribution	Comment
	All	All	All
1	SBS1	50.8288883890011	Signature SBS1 is clock-like in that the number of mutations in most cancers and normal cells correlates with the age of the individual. Rates of acquisition of Signature SBS1 mutations over time differ markedly between different cancer types and different normal cell types. These differences correlate with estimated rates of stem cell division in different tissues and Signature SBS1 may therefore be a cell division/mitotic clock.
2	SBS17b	4.69067620794122	N/A
3	SBS24	176.806850405224	N/A
4	SBS25	154.409001439562	This signature has only been identified in Hodgkin's cell lines. Data is not available from primary Hodgkin lymphomas.
5	SBS28	2.78368589376062	SBS28 has similarities to SBS17b and these two signatures can be mistaken for one another. Signature SBS28 is found in most samples with SBS10a/SBS10b where it contributes very high numbers of mutations. In contrast, SBS28 contributes much smaller number of mutations in samples lacking SBS10a/SBS10b.
6	SBS32	22.8323083046513	N/A
7	SBS37	18.3786519356716	N/A
8	SBS39	75.5107295936865	N/A
9	SBS7d	10.1084587308386	N/A
10	SBS85	3.33914136477661	SBS85 is found in clustered mutations in the immunoglobulin gene and other regions in lymphoid cancers.

Showing 1 to 10 of 10 entries | Previous | 1 | Next

Figure 10 Signature contribution: The table shows the calculated signature contribution in comparison to the 67 Single Base Substitution (SBS) Mutational Signatures from COSMIC v3 (May 2019) with their comments. The table is filtered for signatures with zero fit to a signature.

Figure 10

Signature contribution: The table shows the calculated signature contribution in comparison to the 67 Single Base Substitution (SBS) Mutational Signatures from COSMIC v3 (May 2019) with their comments. The table is filtered for signatures with zero fit to a signature.

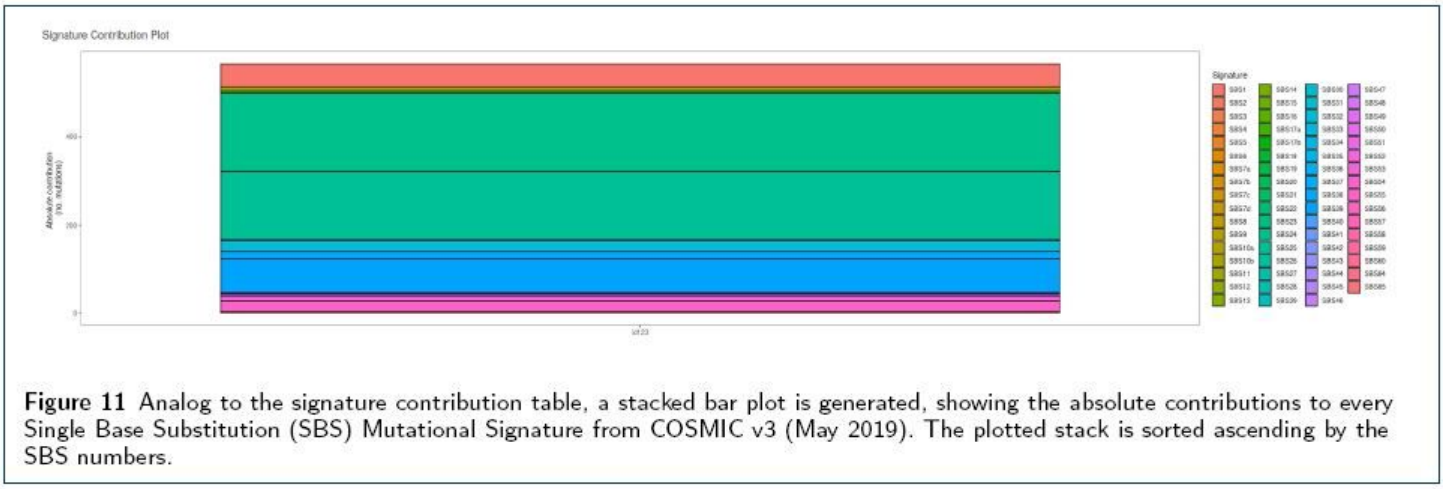


Figure 11 Analog to the signature contribution table, a stacked bar plot is generated, showing the absolute contributions to every Single Base Substitution (SBS) Mutational Signature from COSMIC v3 (May 2019). The plotted stack is sorted ascending by the SBS numbers.

Figure 11

Analog to the signature contribution table, a stacked bar plot is generated, showing the absolute contributions to every Single Base Substitution (SBS) Mutational Signature from COSMIC v3 (May 2019). The plotted stack is sorted ascending by the SBS numbers.