BMC Bioinformatics

SOFTWARE Open Access

Mutational Signatures in Cancer (MuSiCa): a web application to implement mutational signatures analysis in cancer samples

Marcos Díaz-Gay[1†], Maria Vila-Casadesús[2,3†], Sebastià Franch-Expósito[1†], Eva Hernández-Illán[1], Juan José Lozano[2] and Sergi Castellví-Bel[1*]

Abstract

Background: Mutational signatures have been proved as a valuable pattern in somatic genomics, mainly regarding cancer, with a potential application as a biomarker in clinical practice. Up to now, several bioinformatic packages to address this topic have been developed in different languages/platforms. MutationalPatterns has arisen as the most efficient tool for the comparison with the signatures currently reported in the Catalogue of Somatic Mutations in Cancer (COSMIC) database. However, the analysis of mutational signatures is nowadays restricted to a small community of bioinformatic experts.

Results: In this work we present Mutational Signatures in Cancer (MuSiCa), a new web tool based on MutationalPatterns and built using the Shiny framework in R language. By means of a simple interface suited to non-specialized researchers, it provides a comprehensive analysis of the somatic mutational status of the supplied cancer samples. It permits characterizing the profile and burden of mutations, as well as quantifying COSMIC-reported mutational signatures. It also allows classifying samples according to the above signature contributions.

Conclusions: MuSiCa is a helpful web application to characterize mutational signatures in cancer samples. It is accessible online at http://bioinfo.ciberehd.org/GPtoCRC/en/tools.html and source code is freely available at https://github.com/marcos-diazg/musica.

Keywords: Mutational signatures, COSMIC database, Single nucleotide variants, Cancer genomics, Web tool, Shiny, R language

Background

Mutational processes in somatic cells are mainly led by endogenous or exogenous mutagenic agents, as well as errors in DNA replication or repair machineries. Any type of agent or defect is responsible for a specific footprint in the form of a different burden and pattern of mutations. Some of them are historically well-known, as in the case of ultraviolet light exposure and its association with C > T and CC > TT substitutions caused by pyrimidine dimers [1].

In recent years, a new methodology has arisen on this field. Mutational signatures framework enables the association of patterns of mutations with cellular processes and external agents causing them [2]. Since all cancers are caused by somatic mutations, this methodology has the potential to provide insight into their underlying biological processes and become a biomarker in clinical practice [3]. It is based on a computational implementation of non-negative matrix factorization (NMF) considering more than 10,000 cancer samples [4, 5]. Using the information of somatic single nucleotide variants (SNVs), a series of mutational profiles are extracted. These profiles take into account not only substituted nucleotides (all replacements are referred to by the pyrimidine of the

* Correspondence: sbel@clinic.cat
†Marcos Díaz-Gay, Maria Vila-Casadesús and Sebastià Franch-Expósito contributed equally to this work.
[1]Gastroenterology Department, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD, University of Barcelona, Barcelona, Spain
Full list of author information is available at the end of the article

mutated Watson-Crick base pair) but also the 5′ and 3′ adjacent bases. A total of 96 possibilities are evaluated, allowing to detect processes responsible for the same substitutions but in different contexts. According to the current information of the Catalogue of Somatic Mutations in Cancer (COSMIC) database [6], thirty mutational signatures have already been identified across 40 different types of human cancer. This methodology has the potential to reconstruct the mutational spectrum of any cancer sample with sufficient accuracy. This reconstruction is based on the combination of the different signatures contributions. Thus, it constitutes the imprint on the genome of specific mutagenic agents or genetic defects, each represented by a specific signature.

Several bioinformatic approaches have been developed to address mutational signature analysis using different platforms and programming languages. Including some commonalities such as the 96-mutation profile plotting (6 different nucleotide substitutions * 16 different 3-mer contexts), different packages have been recently developed for de novo signature extraction and contribution of known signatures. This is extremely important regarding the possibility of using this methodology in clinical practice. In this context, it would be convenient to perform the analysis at sample resolution, and this is only achievable by comparison with a set of established signatures. MutationalPatterns is an R/Bioconductor package that covers the whole spectrum of functionalities required for mutational signatures framework implementation [7]. It allows the extraction of de novo signatures using the original NMF algorithm, like former R packages pmsignature [8] and Somatic Signatures [9], and Galaxy tool MutSpec [10]. In addition, it also permits the quantification of COSMIC-reported signatures by finding their optimal linear combination. This process is performed approximately 400 times faster than deconstructSigs [11], the only package also covering this functionality [7]. MutationalPatterns has been proved useful in recent studies, both in the identification of somatic mutational profiles [12] and in the characterization of known mutational signatures in human stem cells [13].

However, analysis of somatic mutational signatures remains currently inaccessible for a substantial proportion of the scientific community. Developed software is only useful for bioinformatic experts that should adapt it to their somatic analysis pipelines. Computational resources are also a big challenge, especially when the number of samples to handle is considerably high. In this regard, we have developed a web application to overcome these challenges. Mutational Signatures in Cancer (MuSiCa) allows an easy and quick analysis of mutational signatures in cancer samples, based on a user-friendly web environment adapted to the whole research community. It is mainly built on top of the MutationalPatterns package, so benefiting from its functionalities but also adding a graphical interface designed for non-specialized researchers. MuSiCa also presents some extra features specially designed for cancer samples characterization. Its main aim is to quantify known mutational signatures contribution at sample level, therefore facilitating the identification of the underlying mutational processes. The application also permits to perform an analysis for a complete cohort of cancer patients.

Implementation

MuSiCa was developed using the Shiny framework, which enables the straightforward building of interactive web applications directly from R code [14]. It integrates different publicly available R packages in order to generate a convenient interface and computational efficiency to fluently handle somatic mutation data. Regarding mutational signature framework, MuSiCa uses the data available in COSMIC. Hitherto, the 30 signatures that have been validated and reported in this database have been considered, with the prospect of a future update which would also be transferred to the application. MuSiCa can be easily run online at http://bioinfo.ciberehd.org/GPtoCRC/en/tools.html. Source R code is freely available to download at https://github.com/marcos--diazg/musica, where the required dependencies to install the application are indicated.

A typical workflow of MuSiCa application is presented in Fig. 1. It starts with the uploading of the files containing the somatic SNVs of the samples to analyze. Samples may be derived from international studies as ICGC/TCGA or directly provided by the users. The minimum information required is the chromosome and genomic position according to the human reference genome (UCSC GRCh38/hg38, GRCh37/hg19 and 1000genomes hs37d5 builds are supported), as well as the reference and alternative alleles for every mutation. Different file formats are permitted including the default for this kind of data, the Variant Call Format (VCF). Tab-Separated Values (TSV), Excel and Mutation Annotation Format (MAF) are also allowed. MAF format is commonly used for packing multi-sample data from the Genomics Data Commons projects. Multiple file uploading is allowed in the case of VCF, TSV and Excel formats, each containing the somatic mutations of one sample at a time. For MAF format, only one multi-sample file is allowed. A help modal is present in the MuSiCa website to clarify input format options to the users. The human reference genome build and the type of genomic study performed also need to be provided in order to correctly calculate the prevalence of somatic mutations (i.e. the number of mutations per megabase).

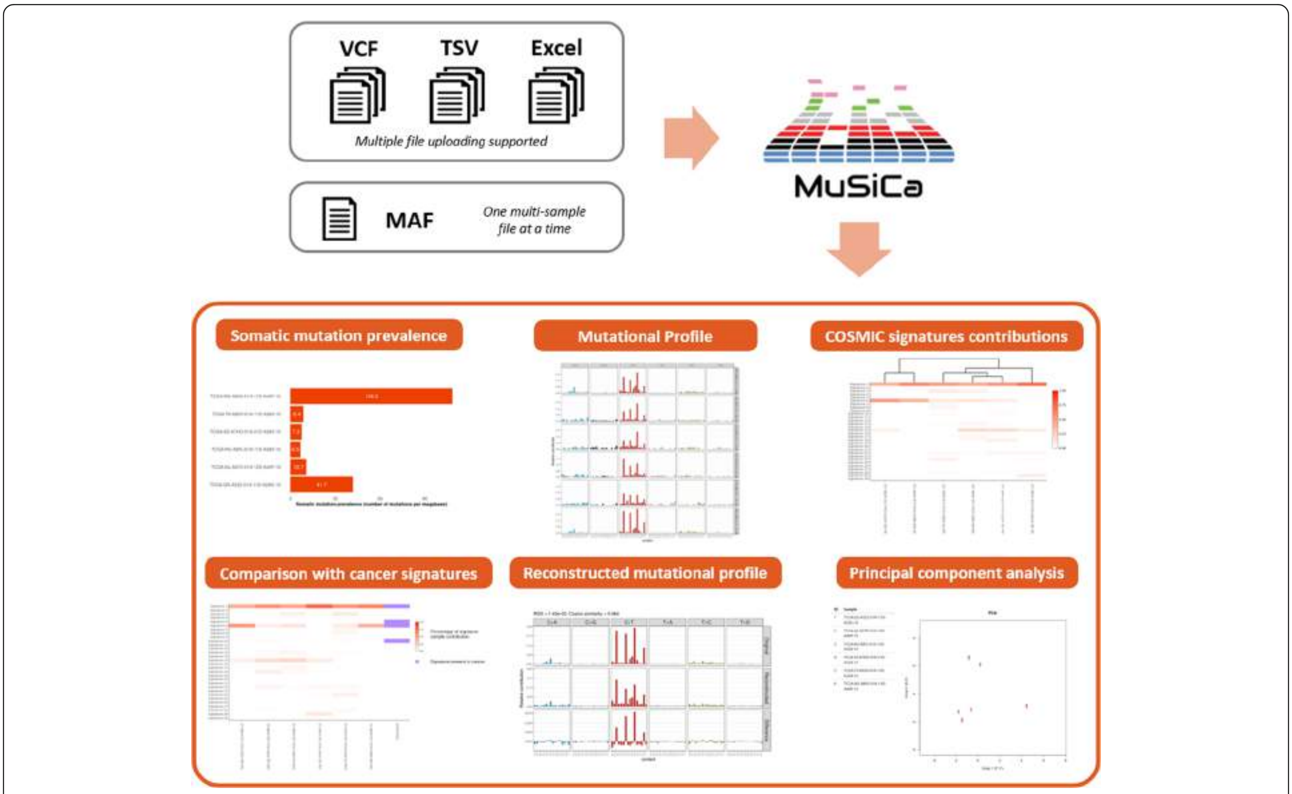

VCF
TSV
Excel
Multiple file uploading supported
MAF
One multi-sample file at a time
MuSiCa
Somatic mutation prevalence
Mutational Profile
COSMIC signatures contributions
Comparison with cancer signatures
Reconstructed mutational profile
Principal component analysis

Fig. 1 Analysis workflow of MuSiCa. Input options supported by the web application and output elements available as plots and data tables ready to download by the users

Output elements are displayed in six different tabs. They are presented in the form of publication-ready figures and tables that can be directly downloaded by the users in different formats. Firstly, mutation prevalence and profiling are presented for somatic mutation characterization. Regarding profiles, all possible SNVs considering the substituted base and the 5′ and 3′ adjacent nucleotides are depicted. Regarding the mutational signatures pattern, it is possible to visualize the contribution of COSMIC-reported signatures, as well as those associated with the distinct cancer types present in this database. The application also permits clustering samples and signatures according to the contributions using a distance measure based on Pearson correlation (1 – correlation value), as well as selecting which samples and cancer types are represented. A principal component analysis (PCA) plot is also presented when more than three samples are uploaded. Both clustering and PCA enable the classification of provided samples according to their quantification regarding known mutational signatures.

This process of signatures quantification is based on the least squares method. This method permits to find the optimal linear combination of the 30 signatures that minimize the residual sum of squares (RSS). Therefore, RSS is a measure of the efficiency of the original mutational profile reconstruction. MuSiCa presents an output tab where original and reconstructed profiles are depicted. RSS is also shown, as well as cosine similarity between both profiles. This value presents instead a direct measure of the correspondence between the two depicted profiles in a 0–1 range (identical profiles would have a value of 1). A value above 0.9 is considered as sufficient accuracy.

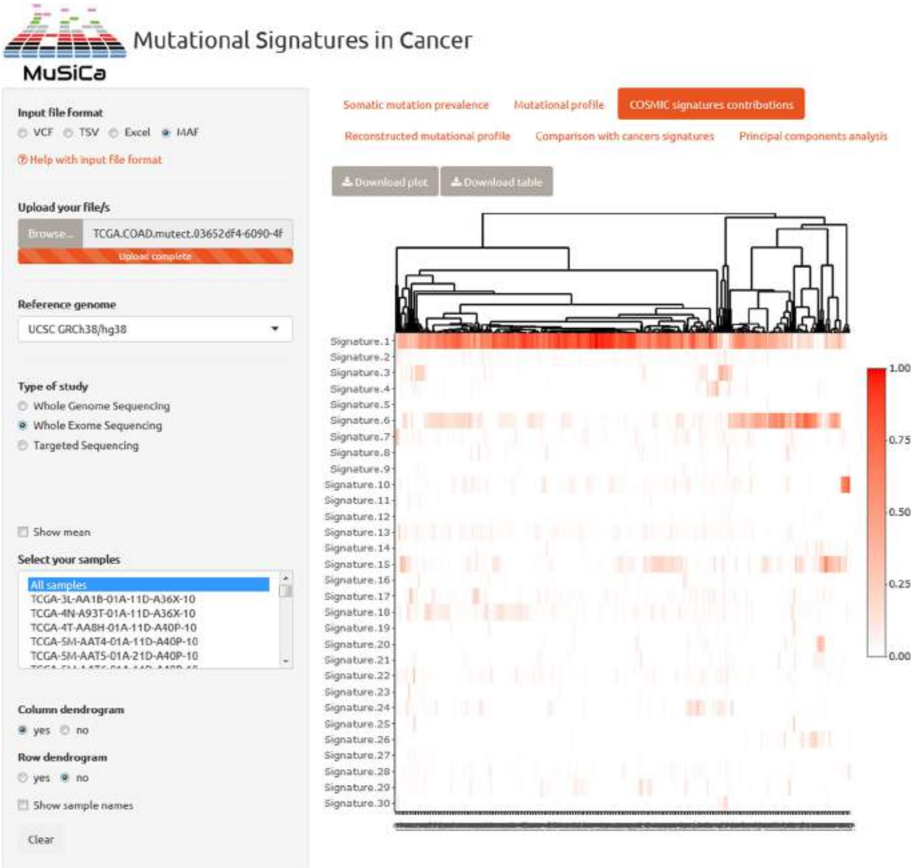
Results and discussion

To assess the usability of the application, colon cancer SNV data from the NCI Genomic Data Commons was used. Four hundred thirty-three samples of this neoplasia were analyzed. They corresponded to the TCGA-COAD project. Somatic mutation data derived from TCGA projects was freely available in MAF format. As this is one of the supported input formats in MuSiCa, the application permitted to directly analyze this publicly available repository. Different upstream analysis workflows were available, using different somatic variant callers. MuTect2-derived data was selected in this example in accordance with GATK Best Practices [15].

Colorectal cancer is one of leading neoplasms worldwide considering mortality and morbidity. Regarding mutagenic agents, effects of environmental factors such as smoking are well-known. However, defects in key molecular pathways, especially those related with DNA repair, have been established as key factors in this neoplasm. Both malfunctioning of mismatch repair (MMR) genes and polymerases δ and ε are reported to affect colorectal carcinogenesis [16]. This is particularly important in the case of hypermutated tumors, defined as those having a mutation rate above 12 per 10^6. This malfunctioning could be caused by somatic but also germline genetic alterations. Indeed, Lynch syndrome and Polymerase proofreading-associated polyposis are both hereditary colorectal cancer syndromes related to malfunctioning of previously indicated DNA repair pathways [17].

Results of the analysis of colorectal cancer TCGA samples with MuSiCa are presented in Fig. 2 and Additional file 1. Regarding the quantification of COSMIC signatures, clustering discriminated at least three different subsets of colon cancer samples in this cohort. The group on the left, accounting for more than half of the samples, was mainly characterized by signature 1. This profile has been found in all cancer types and has been correlated with the age of cancer diagnosis. It is produced as an endogenous process derived from spontaneous deamination of 5-methylcytosine. The other two groups presented a higher level of signatures predominantly associated with MMR deficiency (signatures 6, 15 and 20) and defects in polymerase ε (signature 10). This is in agreement with microsatellite-unstable and *POLE*-mutated colon cancers [16]. However, they also showed the impact of age-associated signature 1. Therefore, this is a good example to realize how mutational signatures reconstruction highlighted the impact of the different underlying causes of mutations present in specific cancer samples. This fact could be a key evidence connecting to the carcinogenic process and even germline susceptibility to the neoplasm.

Regarding developed software for mutational signature analysis, some other tools were already available. In reference to bioinformatic packages, some different options were available as previously mentioned. MutationalPatterns has arisen as the most efficient tool enabling the comparison with the currently reported signatures. In recent years, some web applications have also been published in order to improve the accessibility to this methodology to the whole research community. Pmsignature was the first online application ready to apply mutational signatures framework [8]. However, it was intended just to extract new mutational signatures derived from the supplied samples, not allowing the comparison with known signatures. More recent examples include MutaGene, providing a huge computational framework



Input file format

○ VCF ○ TSV ○ Excel ● MAF

Help with input file format

Upload your file/s

Browse... TCGA.COAD.mutect.03652dF4-6090-4f

Upload complete

Reference genome

UCSC GRCh38/hg38

Type of study

○ Whole Genome Sequencing
● Whole Exome Sequencing
○ Targeted Sequencing

□ Show mean

Select your samples

All samples
TCGA-3L-AA1B-01A-11D-A36X-10
TCGA-4N-A93T-01A-11D-A36X-10
TCGA-4T-AA8H-01A-11D-A40P-10
TCGA-5M-AAT4-01A-11D-A40P-10
TCGA-5M-AAT5-01A-21D-A40P-10

Column dendrogram

● yes ○ no

Row dendrogram

○ yes ● no

□ Show sample names

Clear

Fig. 2 Capture of MuSiCa web application. Known signatures contribution (reported in COSMIC) of 433 TCGA-COAD colon cancer samples is shown. Input options are on the left-hand side and results appear on the main panel when the analysis is performed

regarding somatic cancer mutations [18]. It includes a large repository regarding mutational signatures, but it is more focused on the analysis of publicly available datasets than samples directly provided by the users. In fact, regarding this last point, it permits analyzing a set of samples but cannot generate analysis reports on a single sample level. mSignatureDB is a recent web implementation that allows for the first time to perform signature analysis on datasets directly uploaded by users [19]. Although it permits to quantify known mutational signatures contributions, it lacks some functionalities regarding sample classification, as clustering or PCA analysis. In addition, quantification process is based on deconstructSigs package, with the mentioned weakness on computational efficiency. To the best of our knowledge, no web application is able to characterize the burden of mutation of different cancer samples, as well as cluster and classify them according to their COSMIC-signatures quantification. Thus, MuSiCa becomes the most comprehensive tool available online for somatic characterization of cancer samples datasets directly provided by users.

Conclusions

Our study shows the potential of the mutational signature framework as a biomarker in cancer and the simplicity and usefulness of our implementation. It is also remarkable that MuSiCa allows the analysis at sample level, which is mandatory regarding future clinical implementation of this methodology. Direct accessibility via web, user-friendly environment and computational performance are key factors of our application.

Availability and requirements

- Project name: MuSiCa
- Project home page: https://github.com/marcos-diazg/musica
- Operating system(s): Platform-independent
- Programming language: R, Shiny
- Other requirements: Internet connectivity
- License: MIT License
- Any restrictions to use by non-academics: No

Additional file

Additional file 1: Figure S1. Somatic mutational prevalence in MuSiCa web app. **Figure S2**. Mutational profile representation in MuSiCa web app. **Figure S3**. Reconstruction of mutational profile in MuSiCa web app. **Figure S4**. Comparison with cancer signatures in MuSiCa web app. **Figure S5**. Principal component analysis in MuSiCa web app. (PDF 3039 kb)

Abbreviations

COSMIC: Catalogue of somatic mutations in cancer; MAF: Mutation annotation format; MMR: Mismatch repair; NMF: Non-negative matrix factorization; PCA: Principal component analysis; RSS: Residual sum of squares; SNV: Single nucleotide variant; TSV: Tab-separated values; VCF: Variant call format

Acknowledgements

We are sincerely grateful to Pau Erola, Hadrián Villar and ATIC-UPC for technical support. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

Funding

MDG and SFE are supported by contracts from FI 2017 (B00619, AGAUR, Generalitat de Catalunya) and CIBEREHD, respectively. CIBEREHD is funded by the Instituto de Salud Carlos III. This work was supported by grants from Fondo de Investigación Sanitaria/FEDER (14/00173, 17/00878), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), COST Action CA17118, PERIS (SLT002/16/00398, Generalitat de Catalunya), CERCA Programme (Generalitat de Catalunya) and Agència de Gestió d'Ajuts Universitaris i de Recerca (Generalitat de Catalunya, 2014SGR255, GRPRE 2017SGR21). The funding bodies did not play any role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

All source code has been made publicly available on Github at: https://github.com/marcos-diazg/musica.

Authors' contributions

MDG, MVC, SFE and SCB conceived the idea. MDG, MVC, SFE, EHI and JJL developed the application. MDG and SCB wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Gastroenterology Department, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), University of Barcelona, Barcelona, Spain. [2]Bioinformatics Platform, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Barcelona, Spain. [3]Present Address: Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.

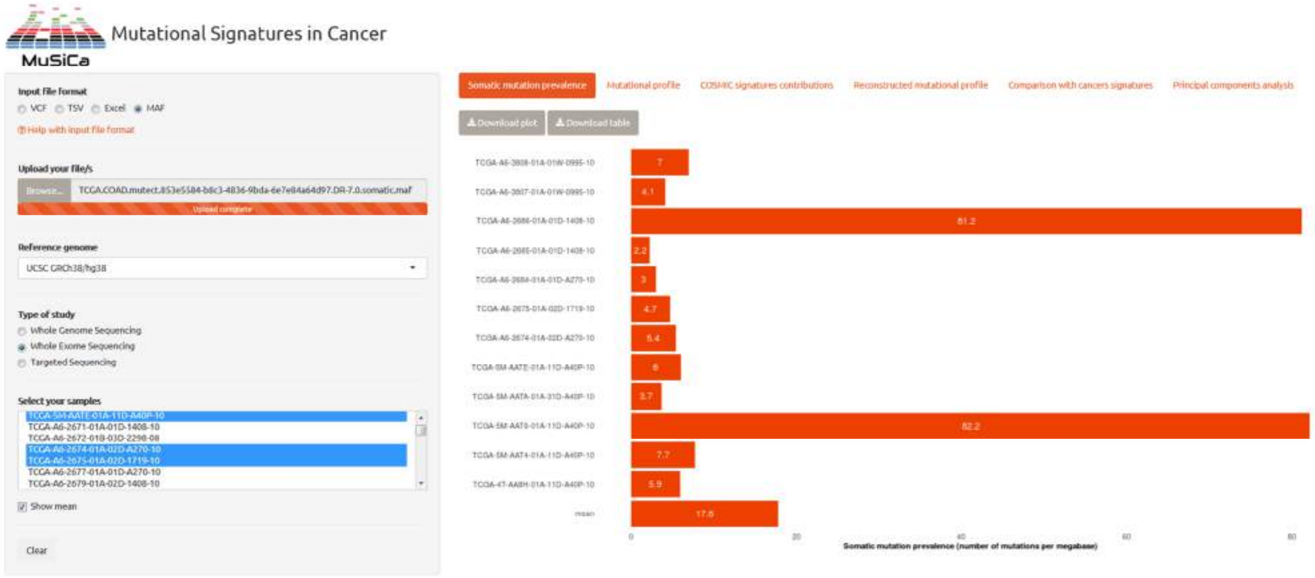
Received: 5 February 2018 Accepted: 4 June 2018
Published online: 14 June 2018

References

1. Miller JH. Mutagenic specificity of ultraviolet light. J Mol Biol. 1985;182:45–65.
2. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415–21.
3. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 2017;23:517–25.
4. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 2013;3:246–59.
5. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nat Genet. 2015;47:1402–7.
6. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017;45:D777–83.

7. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10:33.
8. Shiraishi Y, Tremmel G, Miyano S, Stephens M. A simple model-based approach to inferring and visualizing Cancer mutation signatures. Marchini J, editor. PLOS Genet. 2015;11:e1005657.
9. Gehring JS, Fischer B, Lawrence M, Huber W. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. Bioinformatics. 2015;31:3673–5.
10. Ardin M, Cahais V, Castells X, Bouaoun L, Byrnes G, Herceg Z, et al. MutSpec: a galaxy toolbox for streamlined analyses of somatic mutation spectra in human and mouse cancer genomes. BMC Bioinformatics. 2016;17:170.
11. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.
12. Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, et al. Tissue-specific mutation accumulation in human adult stem cells during life. Nature. 2016;538:260–4.
13. Drost J, van Boxtel R, Blokzijl F, Mizutani T, Sasaki N, Sasselli V, et al. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. Science. 2017;358:234–8.
14. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. Shiny: Web Application Framework for R [Internet]. 2017. Available from: https://cran.r-project.org/package=shiny
15. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.
16. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, et al. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487:330–7.
17. Valle L. Recent discoveries in the genetics of familial colorectal Cancer and polyposis. Clin Gastroenterol Hepatol. 2017;15:809–19.
18. Goncearenco A, Rager SL, Li M, Sang Q-X, Rogozin IB, Panchenko AR. Exploring background mutational processes to decipher cancer genetic heterogeneity. Nucleic Acids Res. 2017;45:W514–22.
19. Huang P-J, Chiu L-Y, Lee C-C, Yeh Y-M, Huang K-Y, Chiu C-H, et al. mSignatureDB: a database for deciphering mutational signatures in human cancers. Nucleic Acids Res. 2018;46:D964–70.

Supplementary Material


Mutational Signatures in Cancer
MuSiCa
Somatic mutation prevalence
Mutational profile
COSMIC signatures contributions
Reconstructed mutational profile
Comparison with cancers signatures
Principal components analysis
Download plot
Download table
Browse...
Upload complete
Clear

Input File format
○ VCF ○ TSV ○ Excel ● MAF
Help with input file format

Upload your file/s
TCGA.COAD.mutect.853e5584-b8c3-4836-9bda-6e7e84a64d97.DR-7.0.somatic.maf

Reference genome
UCSC GRCh38/hg38

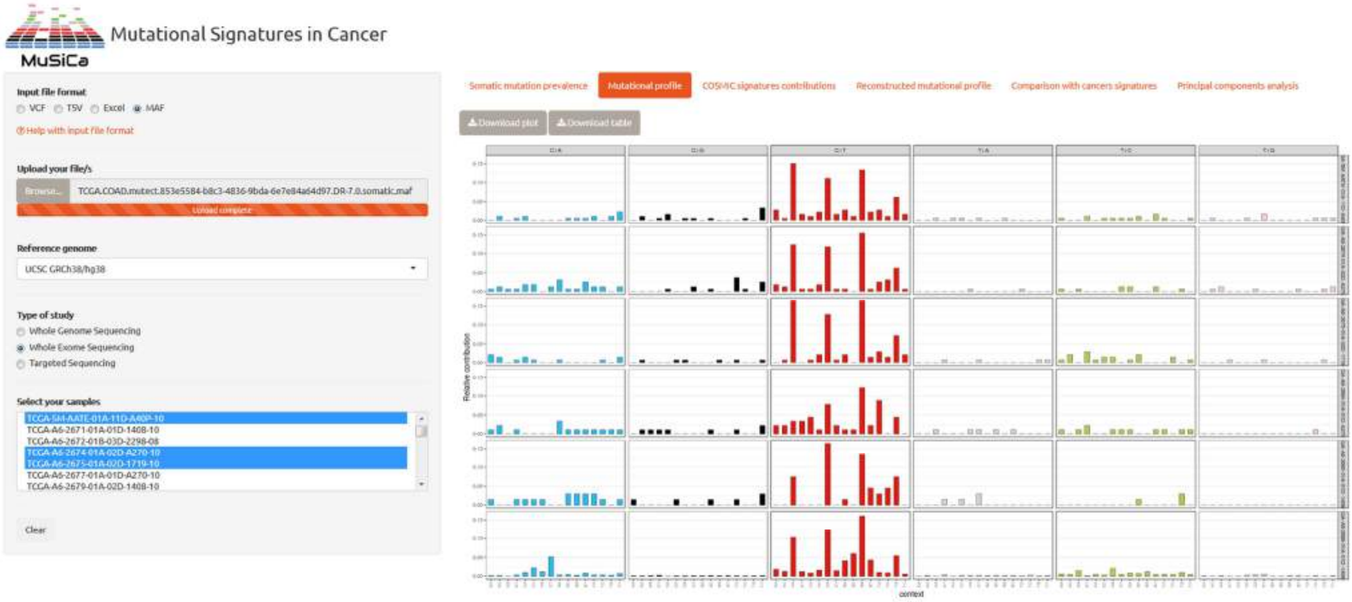
Type of study
○ Whole Genome Sequencing
● Whole Exome Sequencing
○ Targeted Sequencing

Select your samples
TCGA-5M-AATE-01A-11D-A40P-10
TCGA-A6-2671-01A-01D-1408-10
TCGA-A6-2672-01B-03D-2298-08
TCGA-A6-2674-01A-02D-A270-10
TCGA-A6-2675-01A-02D-1719-10
TCGA-A6-2677-01A-01D-A270-10
TCGA-A6-2679-01A-02D-1408-10

☑ Show mean

TCGA-A6-3808-01A-01W-0995-10
7
TCGA-A6-3807-01A-01W-0995-10
4.1
TCGA-A6-2686-01A-01D-1408-10
81.2
TCGA-A6-2685-01A-01D-1408-10
2.2
TCGA-A6-2684-01A-01D-A270-10
3
TCGA-A6-2675-01A-02D-1719-10
4.7
TCGA-A6-2674-01A-02D-A270-10
5.4
TCGA-5M-AATE-01A-11D-A40P-10
6
TCGA-5M-AATA-01A-31D-A40P-10
3.7
TCGA-5M-AAT6-01A-11D-A40P-10
82.2
TCGA-5M-AAT4-01A-11D-A40P-10
7.7
TCGA-4T-AA8H-01A-11D-A40P-10
5.9
mean
17.6
0
20
40
60
80
Somatic mutation prevalence (number of mutations per megabase)

Supplementary Fig. 1. Somatic mutational prevalence in MuSiCa web app. Mutational burden output tab showing a subset of the TCGA-COAD project samples and its mean value.


Mutational Signatures in Cancer
MuSiCa
Somatic mutation prevalence
Mutational profile
COSMIC signatures contributions
Reconstructed mutational profile
Comparison with cancers signatures
Principal components analysis
Download plot
Download table
Browse...
Upload complete
Clear
C>A
C>G
C>T
T>A
T>C
T>G
0.00
0.05
0.10
0.15
Relative contribution
context

Input File format
○ VCF ○ TSV ○ Excel ● MAF
Help with input file format

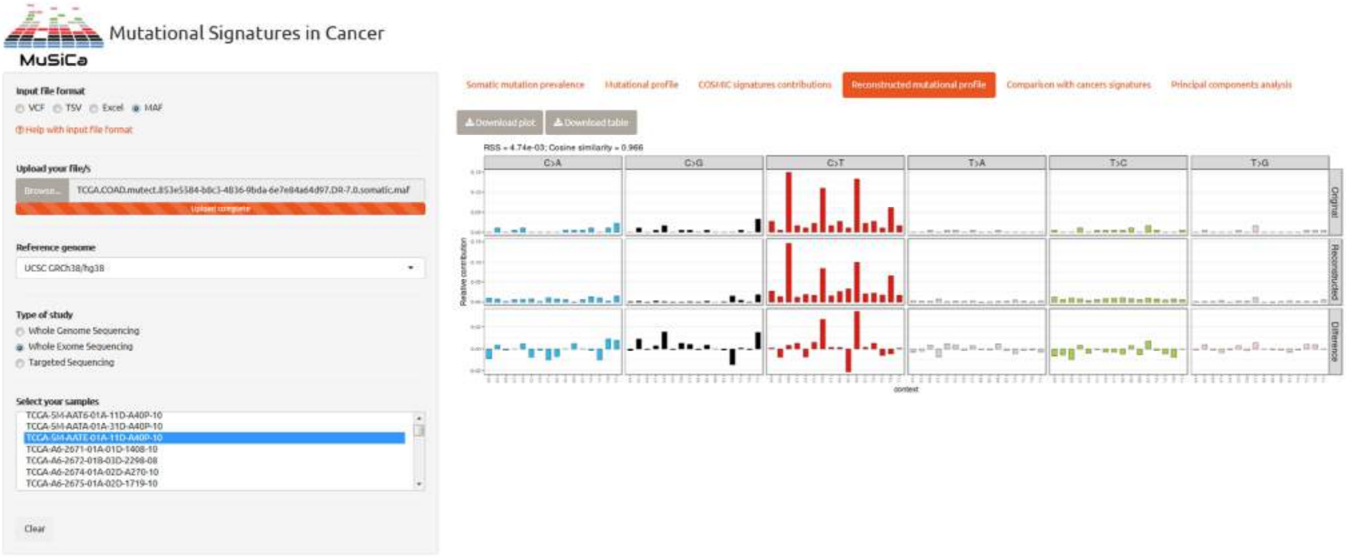
Upload your file/s
TCGA.COAD.mutect.853e5584-b8c3-4836-9bda-6e7e84a64d97.DR-7.0.somatic.maf

Reference genome
UCSC GRCh38/hg38

Type of study
○ Whole Genome Sequencing
● Whole Exome Sequencing
○ Targeted Sequencing

Select your samples
TCGA-5M-AATE-01A-11D-A40P-10
TCGA-A6-2671-01A-01D-1408-10
TCGA-A6-2672-01B-03D-2298-08
TCGA-A6-2674-01A-02D-A270-10
TCGA-A6-2675-01A-02D-1719-10
TCGA-A6-2677-01A-01D-A270-10
TCGA-A6-2679-01A-02D-1408-10

Supplementary Fig. 2. Mutational profile representation in MuSiCa web app. Mutational profile output tab showing a subset of the TCGA-COAD samples.


Mutational Signatures in Cancer
MuSiCa
Somatic mutation prevalence
Mutational profile
COSMIC signatures contributions
Reconstructed mutational profile
Comparison with cancers signatures
Principal components analysis
Download plot
Download table
C>A
C>G
C>T
T>A
T>C
T>G
Original
Reconstructed
Difference
Relative contribution
context
Browse...
Upload complete
Clear

Input file format
○ VCF ○ TSV ○ Excel ● MAF

Help with input file format

Upload your file/s
TCGA.COAD.mutect.853e5584-b8c3-4836-9bda-6e7e84a64d97.DR-7.0.somatic.maf

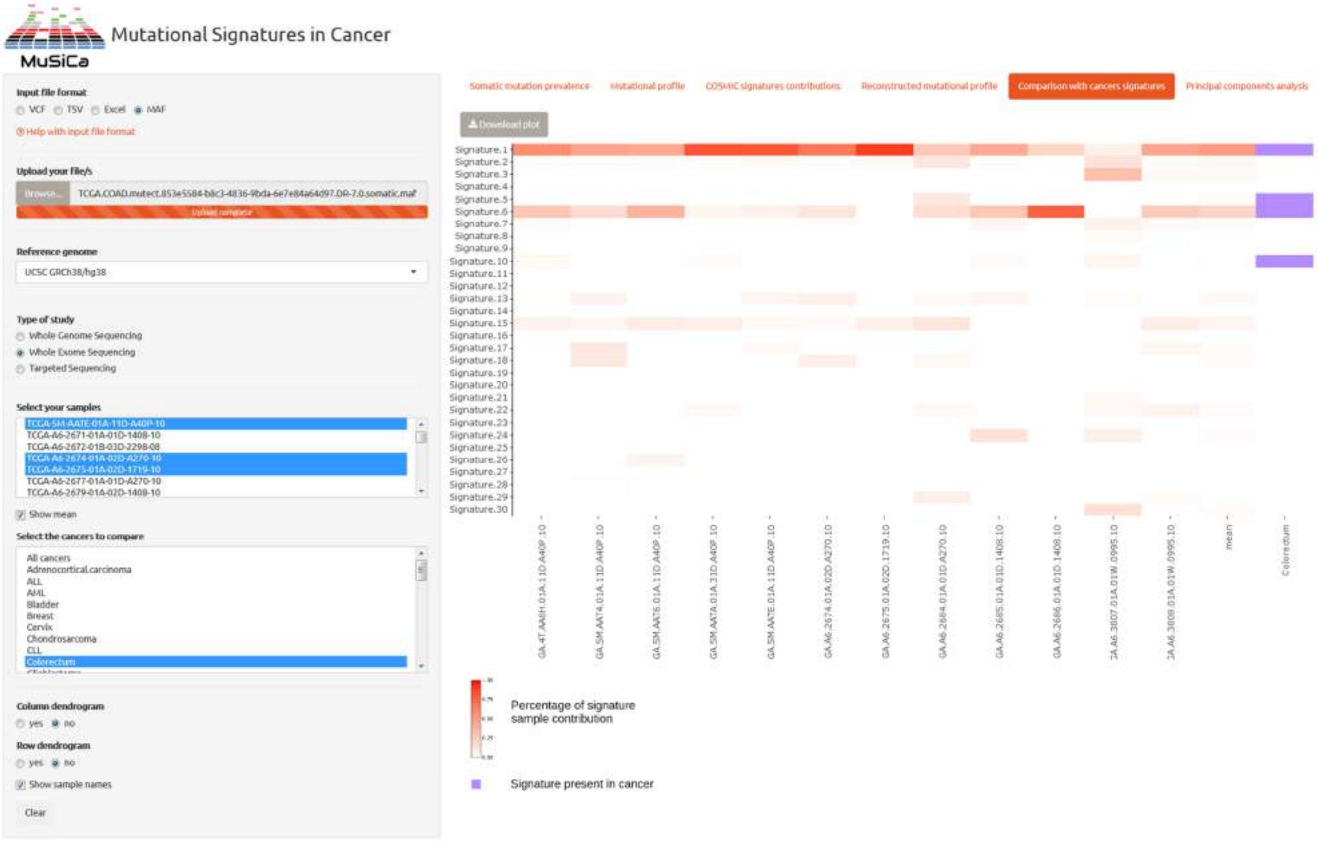
Reference genome
UCSC GRCh38/hg38

Type of study
○ Whole Genome Sequencing
● Whole Exome Sequencing
○ Targeted Sequencing

Select your samples
TCGA-5M-AAT6-01A-11D-A40P-10
TCGA-5M-AATA-01A-31D-A40P-10
TCGA-5M-AATE-01A-11D-A40P-10
TCGA-A6-2671-01A-01D-1408-10
TCGA-A6-2672-01B-03D-2298-08
TCGA-A6-2674-01A-02D-A270-10
TCGA-A6-2675-01A-02D-1719-10

RSS = 4.74e-03; Cosine similarity = 0.966

Supplementary Fig. 3. Reconstruction of mutational profile in MuSiCa web app. Reconstructed mutational profile output tab showing a specific sample of the TCGA-COAD project.


Mutational Signatures in Cancer
MuSiCa
Somatic mutation prevalence
Mutational profile
COSMIC signatures contributions
Reconstructed mutational profile
Comparison with cancers signatures
Principal components analysis
Download plot
Browse...
Upload complete
Clear
Signature.1
Signature.2
Signature.3
Signature.4
Signature.5
Signature.6
Signature.7
Signature.8
Signature.9
Signature.10
Signature.11
Signature.12
Signature.13
Signature.14
Signature.15
Signature.16
Signature.17
Signature.18
Signature.19
Signature.20
Signature.21
Signature.22
Signature.23
Signature.24
Signature.25
Signature.26
Signature.27
Signature.28
Signature.29
Signature.30
GA.4T.AA8H.01A.11D.A40P.10
GA.5M.AAT4.01A.11D.A40P.10
GA.5M.AAT6.01A.11D.A40P.10
GA.5M.AATA.01A.31D.A40P.10
GA.5M.AATE.01A.11D.A40P.10
GA.A6.2674.01A.02D.A270.10
GA.A6.2675.01A.02D.1719.10
GA.A6.2684.01A.01D.A270.10
GA.A6.2685.01A.01D.1408.10
GA.A6.2686.01A.01D.1408.10
3A.A6.3807.01A.01W.0995.10
3A.A6.3808.01A.01W.0995.10
mean
Colorectum
Percentage of signature sample contribution
Signature present in cancer

Input file format
○ VCF ○ TSV ○ Excel ● MAF

Help with input file format

Upload your file/s
TCGA.COAD.mutect.853e5584-b8c3-4836-9bda-6e7e84a64d97.DR-7.0.somatic.maf

Reference genome
UCSC GRCh38/hg38

Type of study
○ Whole Genome Sequencing
● Whole Exome Sequencing
○ Targeted Sequencing

Select your samples
TCGA-5M-AATE-01A-11D-A40P-10
TCGA-A6-2671-01A-01D-1408-10
TCGA-A6-2672-01B-03D-2298-08
TCGA-A6-2674-01A-02D-A270-10
TCGA-A6-2675-01A-02D-1719-10
TCGA-A6-2677-01A-01D-A270-10
TCGA-A6-2679-01A-02D-1408-10

☑ Show mean

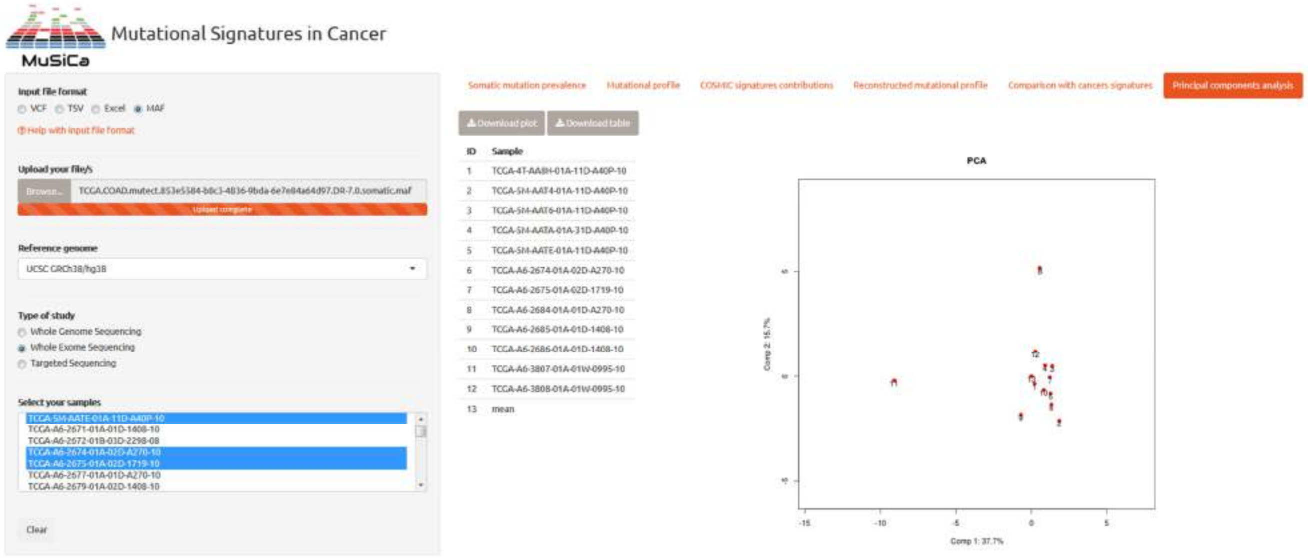
Select the cancers to compare
All cancers
Adrenocortical.carcinoma
ALL
AML
Bladder
Breast
Cervix
Chondrosarcoma
CLL
Colorectum

Column dendrogram
○ yes ● no

Row dendrogram
○ yes ● no

☑ Show sample names

Supplementary Fig. 4. Comparison with cancer signatures in MuSiCa web app. Output tab presenting a comparison of known signatures contributions with mutational signatures reported in different human cancer types in a subset of the TCGA-COAD samples and its mean value.


Mutational Signatures in Cancer
MuSiCa
Somatic mutation prevalence
Mutational profile
COSMIC signatures contributions
Reconstructed mutational profile
Comparison with cancers signatures
Principal components analysis
Download plot
Download table
Help with input file format
Browse...
Upload complete
Clear
PCA
Comp 2: 15.7%
Comp 1: 37.7%
5
0
-5
-15
-10
-5
0
5

Input file format
○ VCF ○ TSV ○ Excel ● MAF

Upload your file/s
TCGA.COAD.mutect.853e5384-b8c3-4836-9bda-6e7e84a64d97.DR-7.0.somatic.maf

Reference genome
UCSC GRCh38/hg38

Type of study
○ Whole Genome Sequencing
● Whole Exome Sequencing
○ Targeted Sequencing

Select your samples
TCGA-5M-AATE-01A-11D-A40P-10
TCGA-A6-2671-01A-01D-1408-10
TCGA-A6-2672-01B-03D-2298-08
TCGA-A6-2674-01A-02D-A270-10
TCGA-A6-2675-01A-02D-1719-10
TCGA-A6-2677-01A-01D-A270-10
TCGA-A6-2679-01A-02D-1408-10

ID	Sample
1	TCGA-4T-AA8H-01A-11D-A40P-10
2	TCGA-5M-AAT4-01A-11D-A40P-10
3	TCGA-5M-AAT6-01A-11D-A40P-10
4	TCGA-5M-AATA-01A-31D-A40P-10
5	TCGA-5M-AATE-01A-11D-A40P-10
6	TCGA-A6-2674-01A-02D-A270-10
7	TCGA-A6-2675-01A-02D-1719-10
8	TCGA-A6-2684-01A-01D-A270-10
9	TCGA-A6-2685-01A-01D-1408-10
10	TCGA-A6-2686-01A-01D-1408-10
11	TCGA-A6-3807-01A-01W-0995-10
12	TCGA-A6-3808-01A-01W-0995-10
13	mean

Supplementary Fig. 5. Principal component analysis in MuSiCa web app. Principal component analysis output tab presenting a classification of a specific subset of the TCGA-COAD samples according to the quantification of known signatures contributions.