KSBi-BIML 2024

Bioinformatics & Machine Learning(BIML)
Workshop for Life and Medical Scientists

생명정보학 & 머신러닝 워크샵 (온라인)

Mutational signatures in cancer genomes

주영석 _ KAIST

KSBI
KOREAN SOCIETY FOR BIOINFORMATICS
한국생명정보학회

⚠

본 강의 자료는 한국생명정보학회가 주관하는 BIML 2024 워크샵 온라인 수업을 목적으로 제작된 것으로 해당 목적 이외의 다른 용도로 사용할 수 없음을 분명하게 알립니다.

이를 다른 사람과 공유하거나 복제, 배포, 전송할 수 없으며 만약 이러한 사항을 위반할 경우 발생하는 **모든 법적 책임은 전적으로 불법 행위자 본인에게 있음을 경고**합니다.

KSBi-BIML 2024

Bioinformatics & Machine Learning(BIML)
Workshop for Life and Medical Scientists

안녕하십니까?

한국생명정보학회가 개최하는 동계 교육 워크샵인 BIML-2024에 여러분을 초대합니다. 생명정보학 분야의 연구자들에게 최신 동향의 데이터 분석기술을 이론과 실습을 겸비해 전달하고자 도입한 전문 교육 프로그램인 BIML 워크샵은 2015년에 시작하여 올해로 벌써 10년 차를 맞이하게 되었습니다. BIML 워크샵은 국내 생명정보학 분야의 최초이자 최고 수준의 교육프로그램으로 크게 인공지능과 생명정보분석 두 개의 분야로 구성되어 있습니다. 올해 인공지능 분야에서는 최근 생명정보 분석에서도 응용이 확대되고 있는 다양한 인공지능 기반 자료모델링 기법들에 대한 현장 강의가 진행될 예정이며, 관련하여 심층학습을 이용한 단백질구조예측, 유전체분석, 신약개발에 대한 이론과 실습 강의가 함께 제공될 예정입니다. 또한 단일세포오믹스, 공간오믹스, 메타오믹스, 그리고 롱리드염기서열 자료 분석에 대한 현장 강의는 많은 연구자의 연구 수월성 확보에 큰 도움을 줄 것으로 기대하고 있습니다.

올해 BIML의 가장 큰 변화는 최근 연구 수요가 급증하고 있는 의료정보자료 분석에 대한 현장 강의를 추가하였다는 것입니다. 특히 의료정보자료 분석을 많이 수행하시는 의과학자 및 의료정보 연구자들께서 본 강좌를 통해 많은 도움을 받으실 수 있기를 기대하고 있습니다. 또한 다양한 생명정보학 분야에 대한 온라인 강좌 프로그램도 점차 증가하고 있는 생명정보 분석기술의 다양화에 발맞추기 위해 작년과 비교해 5강좌 이상을 신규로 추가했습니다. 올해는 무료 강좌 5개를 포함하여 35개 이상의 온라인 강좌가 개설되어 제공되며, 연구 주제에 따른 연관된 강좌 추천 및 강연료 할인 프로그램도 제공되며, 온라인을 통한 Q&A 세션도 마련될 예정입니다. BIML-2024는 국내 주요 연구중심 대학의 전임 교원이자 각 분야 최고 전문가들의 강의로 구성되었기에 해당 분야의 기초부터 최신 연구 동향까지 포함하는 수준 높은 내용의 강의가 될 것이라 확신합니다.

BIML-2024을 준비하기까지 너무나 많은 수고를 해주신 운영위원회의 정성원, 우현구, 백대현, 김태민, 김준일, 김상우, 장혜식, 박종은 교수님과 KOBIC 이병욱 박사님께 커다란 감사를 드립니다. 마지막으로 부족한 시간에도 불구하고 강의 부탁을 흔쾌히 허락하시고 훌륭한 현장 강의와 온라인 강의를 준비하시는데 노고를 아끼지 않으신 모든 강사분들께 깊은 감사를 드립니다.

2024년 2월

한국생명정보학회장 이 인 석

강의개요

Mutational signatures in cancer genomes

Cancer genome sequencing을 이용하면 우리는 무엇을 배울 수 있을까? 1차적으로는 최적화 약제를 선별하기 위한 cancer driver mutation을 찾기 위한 목표로 쓰인다. 하지만 Cancer genome에서 나오는 수 많은 돌연변이의 pattern, 즉 mutational signature를 체계적으로 분석하면 정상세포에서 암 세포로 돌변하는 과정중에서 돌연변이들을 만들어낸 기전을 이해할 수 있다.

본 강의에서는 암 세포에서 발견한 돌연변이로부터 mutational signature를 빠르게 추출하고 분석하는 방법을 설명한다. Mutational signature의 개념, signature를 calling하는 알고리즘 및 툴을 소개하며, 이를 실제 암 유전체 데이터에 적용하여 효율적이고 효과적인 분석을 할 수 있는 핵심 역량을 갖추는 것을 목표로 한다.

강의는 다음의 내용을 포함한다:

- Mutational signature 의 개념
- Mutational signature 의 calling algorithm 및 tools

* 강의 난이도: 초급

* 강의: 주영석 교수 (KAIST 의과학대학원)

Curriculum Vitae

Speaker Name: Young Seok Ju, M.D. Ph.D.

▶ Personal Info

Name	Young Seok Ju
Title	Associate Professor
Affiliation	Grad School of Medical Science and Engineering, KAIST

▶ Contact Information

Address	291 Daehak-ro Yuseong-gu, Daejeon 34141
Email	ysju@kaist.ac.kr
Phone Number	042-350-4237

Research Interest

Somatic mutation, somatic mosaicism, bioinformatics, mutational process

Educational Experience

2007	M.D. in Medicine, Seoul Nat'l Univ College of Medicine, Seoul, Korea
2010	Ph.D. in Genomic Medicine, Seoul Nat'l Univ College of Medicine, Seoul, Korea

Professional Experience

2013-2015	Post-doc, Wellcome Sanger Institute, Daejeon, Korea
2015-	Associate/Assistant Professor, KAIST

Selected Publications (5 maximum)

1. Park S, Mali NM, Kim R, Choi JW, Lee J*,...,Oh J#, , **Ju YS#**. Clonal dynamics in early human embryogenesis inferred from somatic mutation. ***Nature*** 2021
2. Youk J*, Kim T*, Evans KV*, Jeong Y-I*, Hur Y*, Hong SP*, ..., Kim YT#, Koh GY#, Choi B-S#, **Ju YS#**, Lee JH#. Three-dimensional human alveolar stem cell culture models reveal infection response to SARS-CoV-2. ***Cell Stem Cell.*** 2020
3. Lee JS, An Y, Yoon CJ, Kim JY, Kim KH, ... , Lee EY# & **Ju YS#**. Germline gain-of-function mutation of STAT1 rescued by somatic mosaicism in immune dysregulation-polyendocrinopathy-enteropathy-X-linked-like disorder. ***J Allergy Clin Immunol.*** 2020
4. Lee JJ-K, Park S, Park H, Kim S, Lee J, ... , **Ju YS#** & Kim YT#. Tracing oncogene rearrangements in the mutational history of lung adenocarcinoma. ***Cell.*** 2019
5. Lee JK., Lee J, Kim S, Kim S, Youk J, ..., Kim TM# & **Ju YS#**. Clonal history and genetic predictors of transformation into small cell carcinomas from lung adenocarcinomas. ***Journal of Clinical Oncology*** 2017 Sep 10;35(26):3065-3074. PMID:28498782

KSBi-BIML

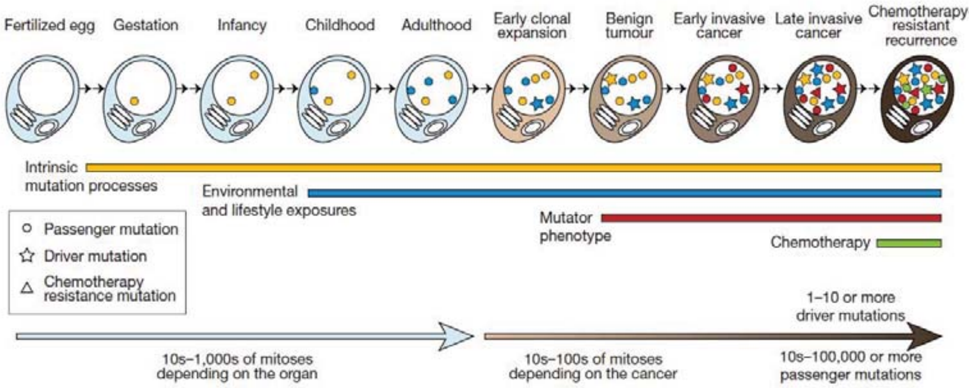
Mutational signatures in cancer genomes

주영석 (KAIST) ysju@kaist.ac.kr

분석의 목적: 왜 암 유전체를 분석하는가?

- 목적에 따라 다양한 접근법을 이용할 수 있음
  - 임상 의사: 환자 암에서 clinically actionable target을 발굴, 진료에 응용
    (EGFR activating mutation 발굴)
  - Genomics, Bioinformatics 에 관심이 있는 학부생, 대학원생, 박사 후 연구원 등
    새로운 돌연변이 발굴, technology/bioinformatics 개발, 논문 출판
  - 회사나 연구소의 전문 연구원
    Pipeline 구축 등

암 유전체 분석의 시작: 돌연변이의 검출



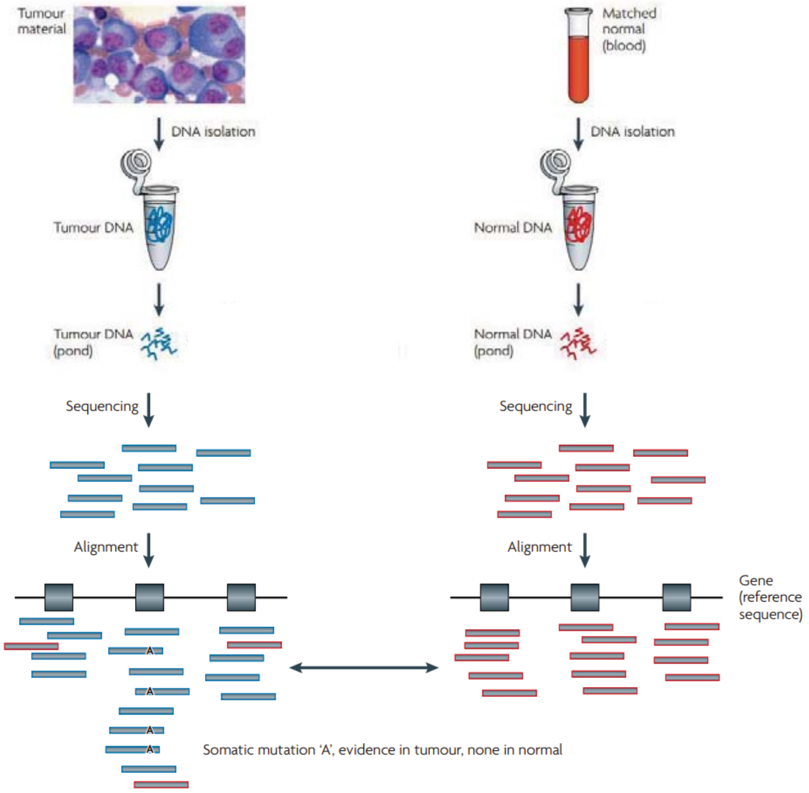
Fertilized egg
Gestation
Infancy
Childhood
Adulthood
Early clonal expansion
Benign tumour
Early invasive cancer
Late invasive cancer
Chemotherapy-resistant recurrence
Intrinsic mutation processes
Environmental and lifestyle exposures
Mutator phenotype
Chemotherapy
Passenger mutation
Driver mutation
Chemotherapy resistance mutation
1–10 or more driver mutations
10s–1,000s of mitoses depending on the organ
10s–100s of mitoses depending on the cancer
10s–100,000 or more passenger mutations

Figure 1 | The lineage of mitotic cell divisions from the fertilized egg to a single cell within a cancer showing the timing of the somatic mutations acquired by the cancer cell and the processes that contribute to them. Mutations may be acquired while the cell lineage is phenotypically normal, reflecting both the intrinsic mutations acquired during normal cell division and the effects of exogenous mutagens. During the development of the cancer other processes, for example DNA repair defects, may contribute to the mutational burden. Passenger mutations do not have any effect on the cancer cell, but driver mutations will cause a clonal expansion. Relapse after chemotherapy can be associated with resistance mutations that often predate the initiation of treatment.

- 대부분의 산발성 암 (sporadic cancer) 의 원인은 체세포 돌연변이이다

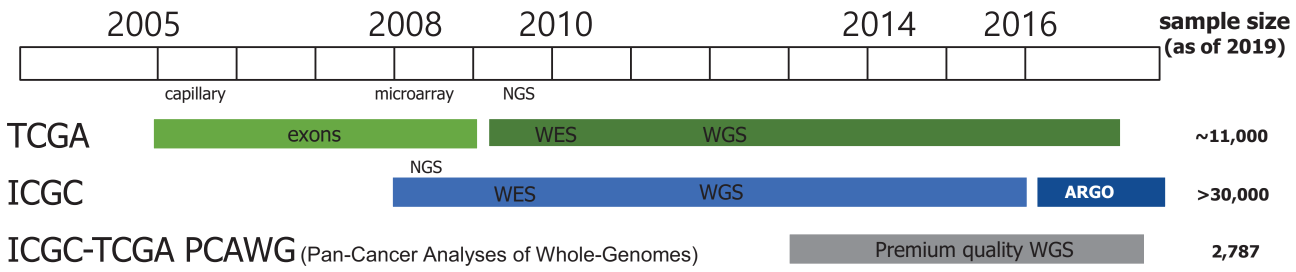
Stratton *et al.*, ***Nature*** (2009)

돌연변이의 검출을 위한 전략



Meyerson M *et al.*, ***Nat Rev Cancer*** (2010)

International consortia for cancer genome analyses



2005 2008 2010 2014 2016

sample size (as of 2019)

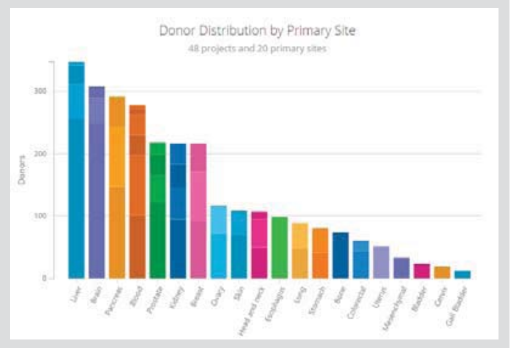
capillary microarray NGS

TCGA — exons — WES WGS — ~11,000

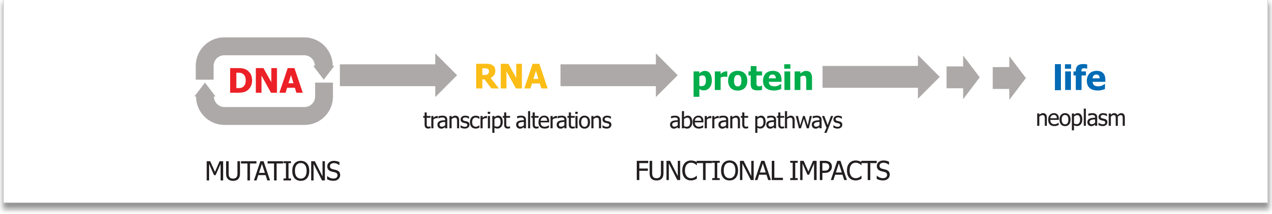
NGS

ICGC — WES WGS — ARGO — >30,000

ICGC-TCGA PCAWG (Pan-Cancer Analyses of Whole-Genomes) — Premium quality WGS — 2,787



Driver mutation을 찾는 것이 암유전체 분석의 한 목적



DNA → RNA → protein → life

MUTATIONS

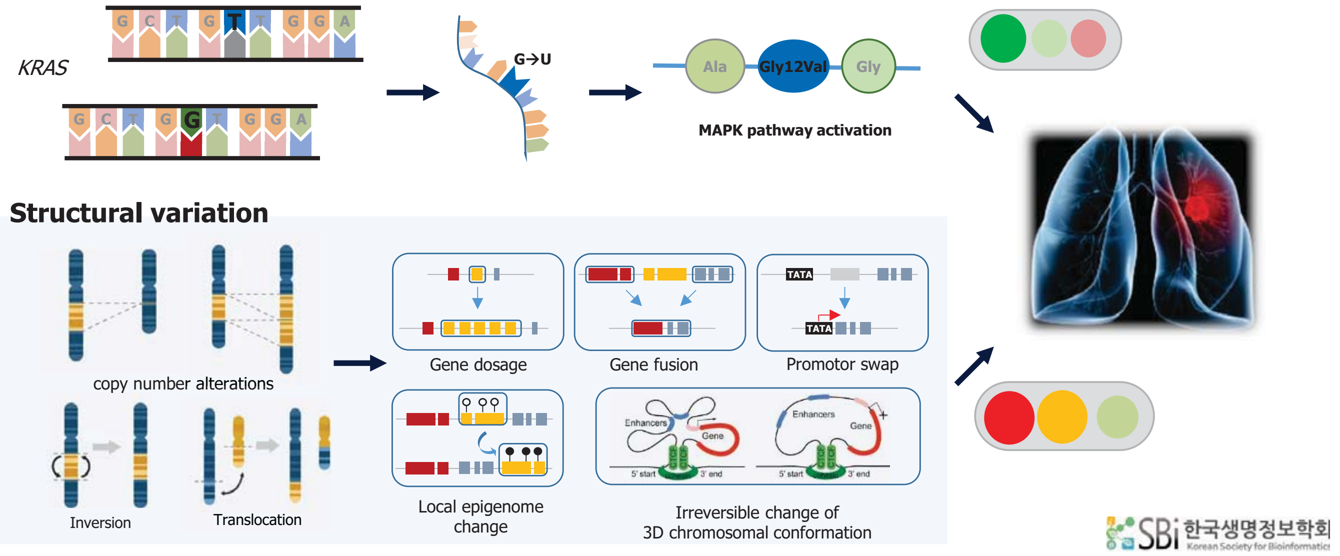
RNA: transcript alterations

protein: aberrant pathways

life: neoplasm

FUNCTIONAL IMPACTS

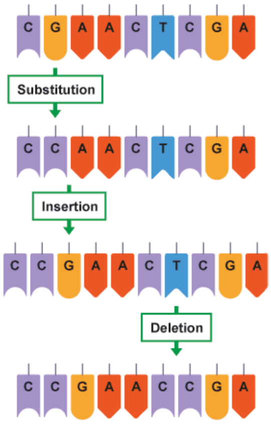
Point mutation

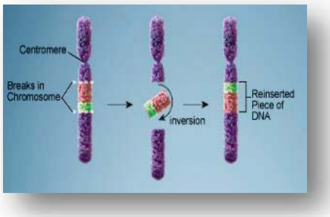


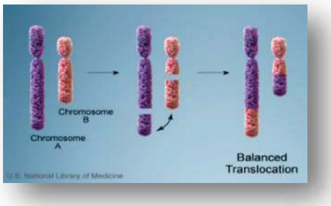
Structural variation

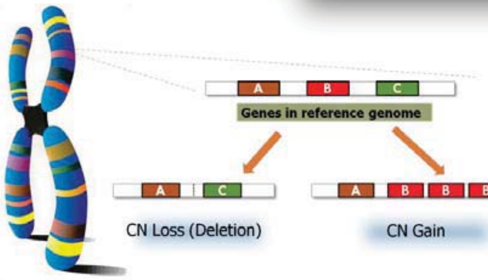
어떤 돌연변이가 있는가?

- 크기에 의한 분류
  - Small (point-mutation):
    - base substitution (SNV, SNP), short-indel
  - Large:
    - Copy number variation, genome rearrangements, SV
- 유전체 위치에 의한 분류
  - Coding mutation (in the protein coding region)
    - Non-sense/frameshift (truncating, stop-gain)
    - Missense (non-synonymous)
    - Silent (synonymous)
  - UTR, intronic, splicing-junction
  - intergenic (between two genes)

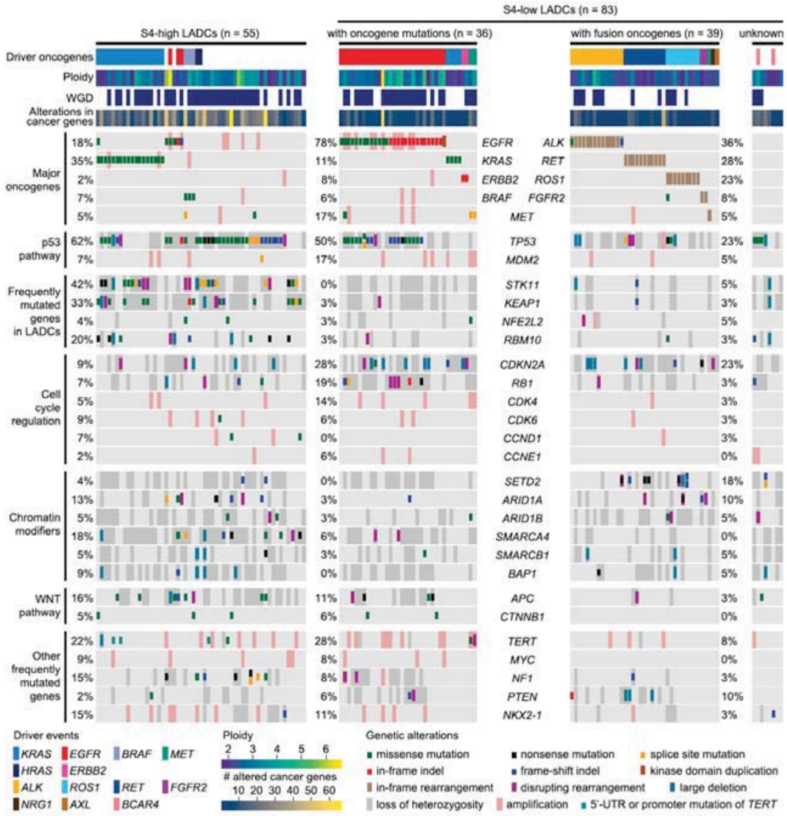






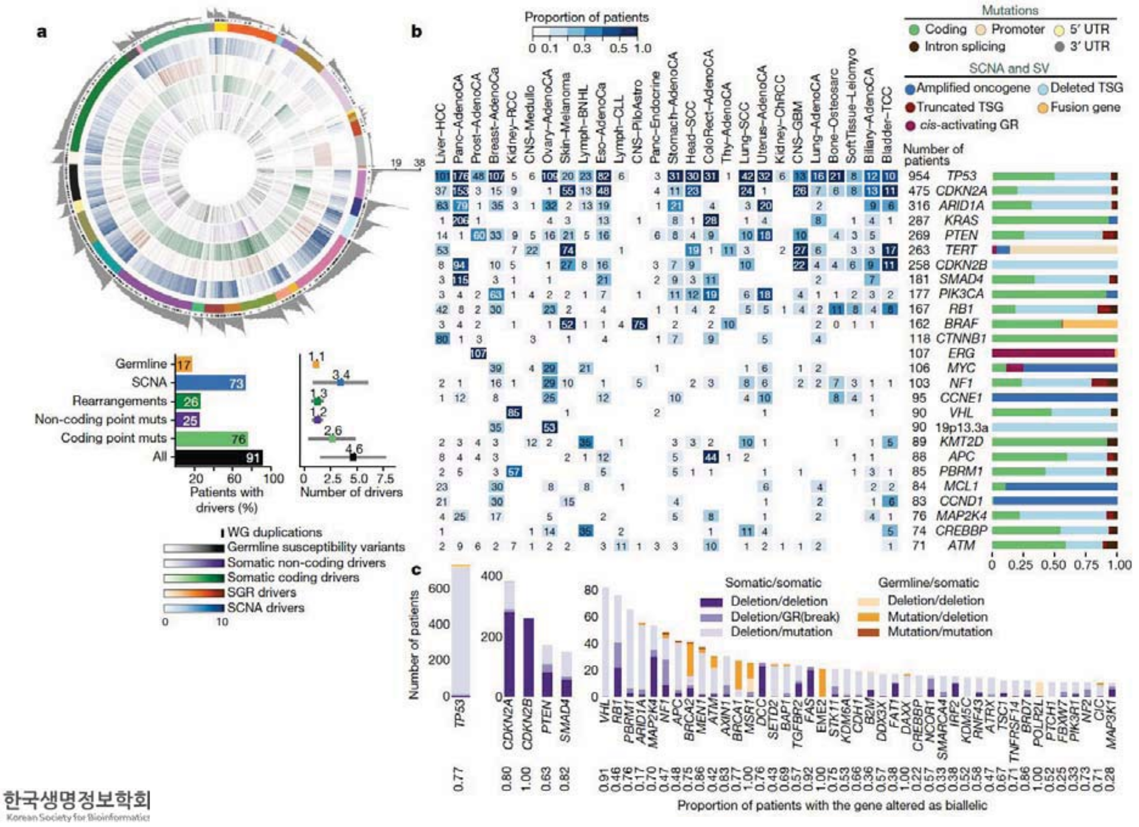


Cancer genome에서 driver mutation의 분포



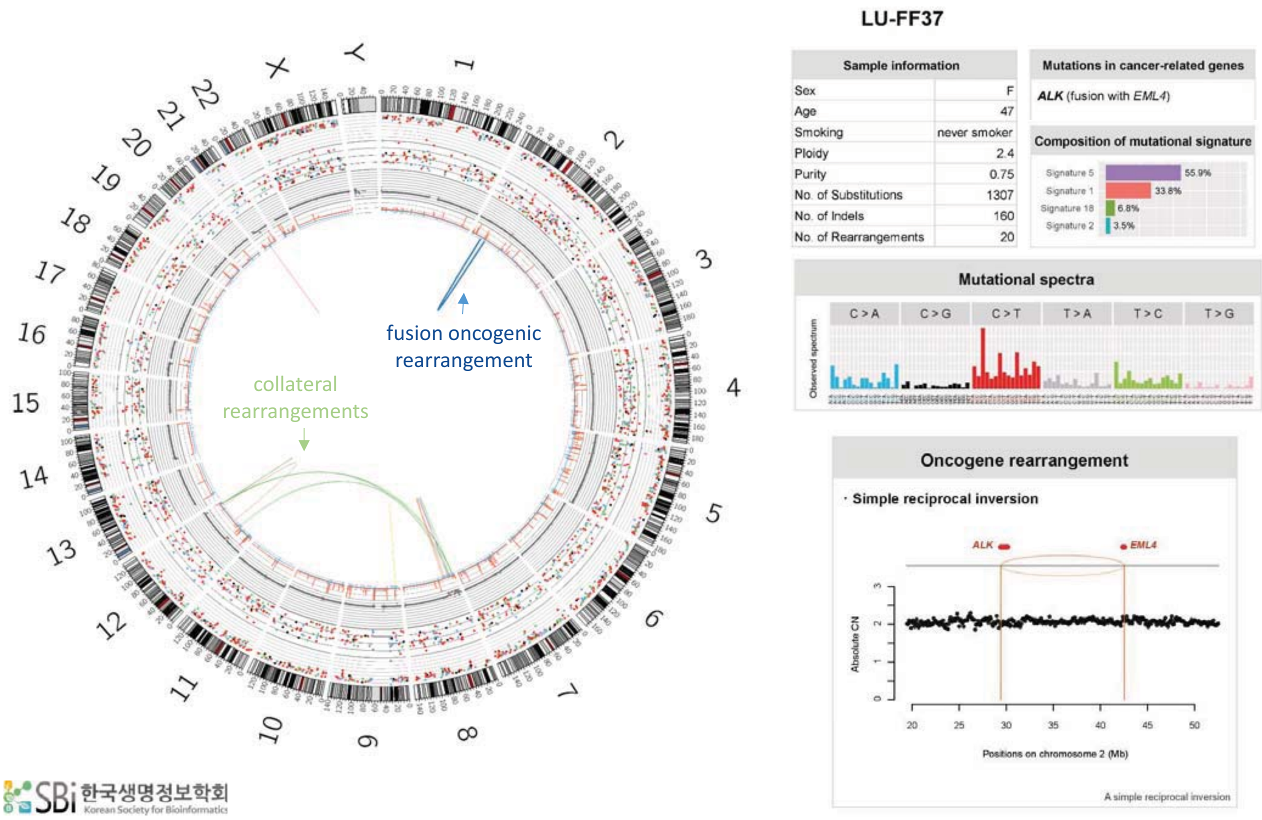
Lee JJ *et al.*, ***Cell*** (2019)

Driver mutations in pan-cancer genomes



PCAWG consortium, ***Nature*** (2020)

An example of genome-wide sequencing of a cancer genome



fusion oncogenic rearrangement

collateral rearrangements

LU-FF37

Sample information	
Sex	F
Age	47
Smoking	never smoker
Ploidy	2.4
Purity	0.75
No. of Substitutions	1307
No. of Indels	160
No. of Rearrangements	20

Mutations in cancer-related genes

ALK (fusion with *EML4*)

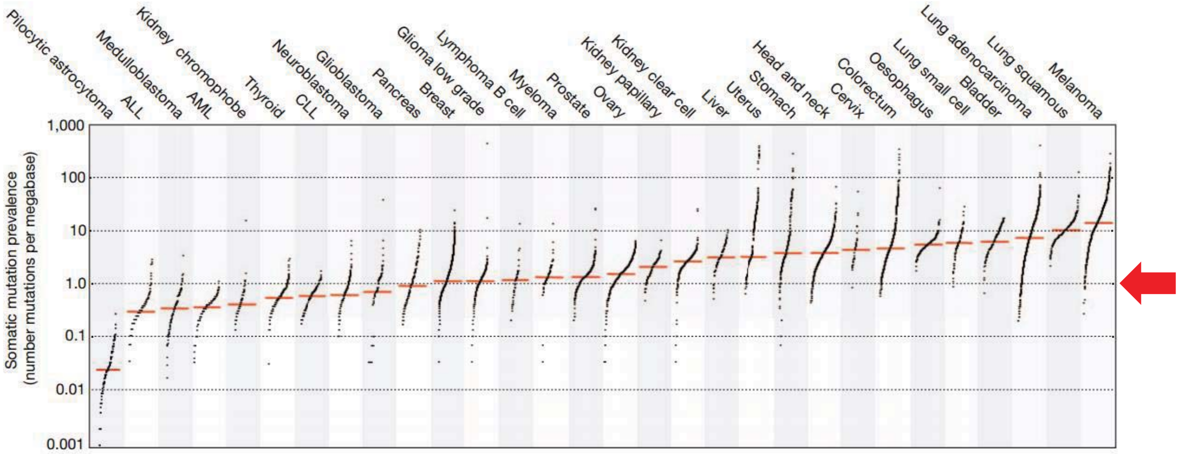
Composition of mutational signature

Signature 5	55.9%
Signature 1	33.8%
Signature 18	6.8%
Signature 2	3.5%

· Simple reciprocal inversion

Lee JJ *et al.*, ***Cell*** (2019)

암유전체의 돌연변이들

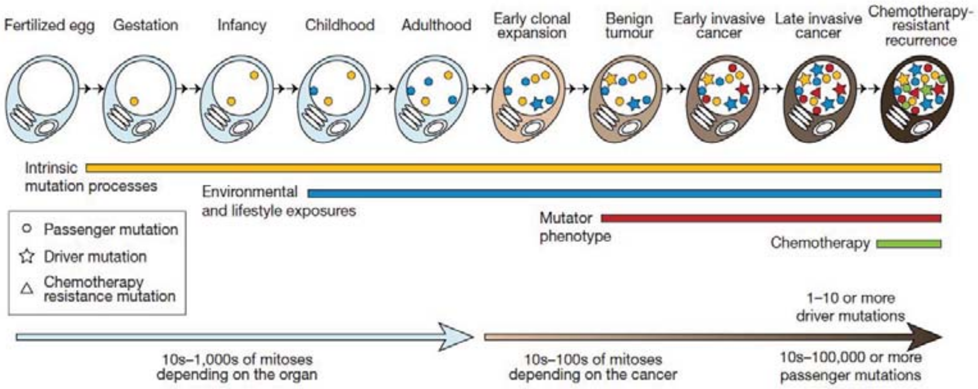


Average
(1.0/Mb)

Alexandrov L *et al.*, ***Nature*** (2013)

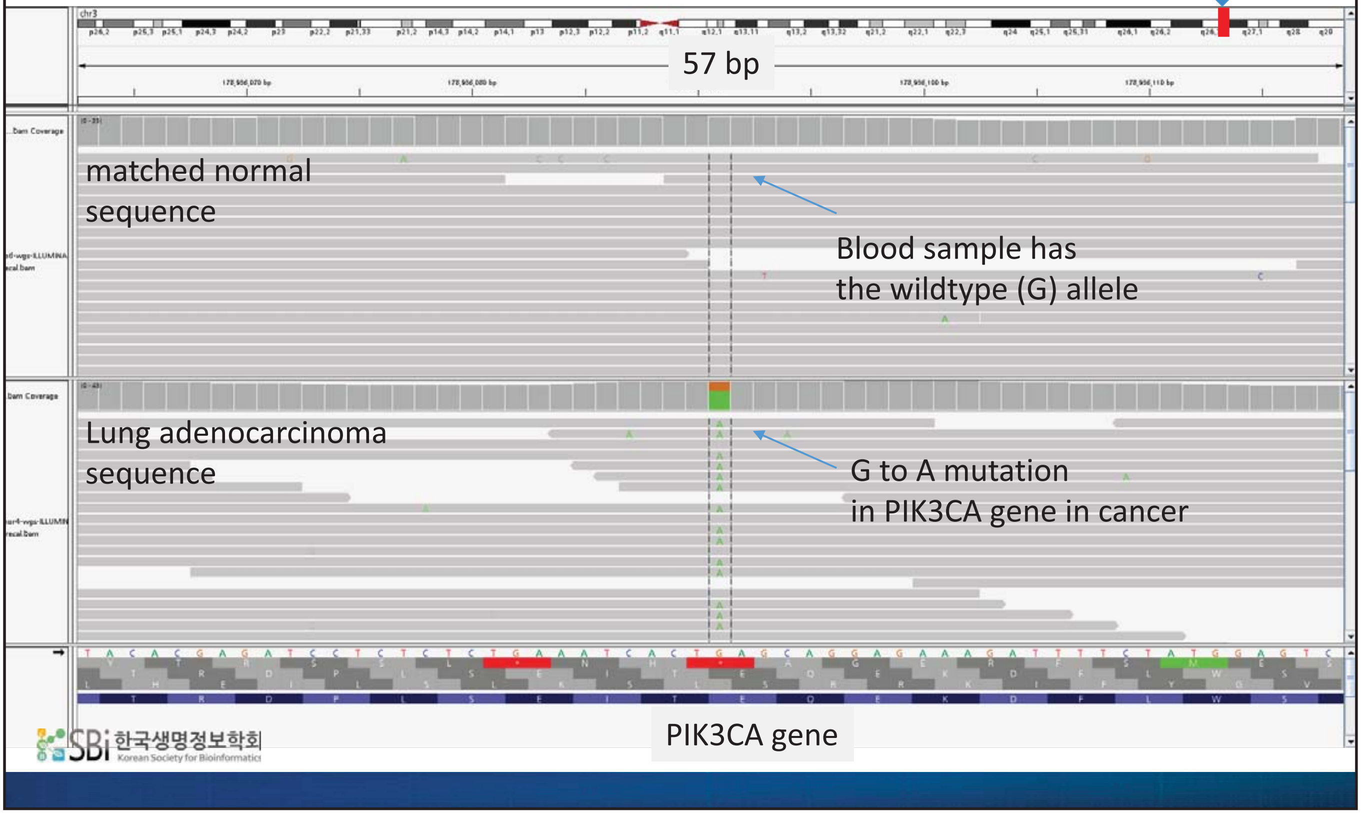
- WGS (3,000 Mb) ➔ 3,000 (1,000 – 100,000 substitutions)
- WES (~50 Mb) ➔ 50 (10 – 1,000 substitutions)
- Targeted-gene seq. (covers ~1 Mb) ➔ 10 (1 – 100 substitutions)

Cancer genomics에서 passenger mutation은 쓸모가 없는가?


Fertilized egg
Gestation
Infancy
Childhood
Adulthood
Early clonal expansion
Benign tumour
Early invasive cancer
Late invasive cancer
Chemotherapy-resistant recurrence
Intrinsic mutation processes
Environmental and lifestyle exposures
Mutator phenotype
Chemotherapy
○ Passenger mutation
☆ Driver mutation
△ Chemotherapy resistance mutation
1–10 or more driver mutations
10s–1,000s of mitoses depending on the organ
10s–100s of mitoses depending on the cancer
10s–100,000 or more passenger mutations

Stratton *et al.*, ***Nature*** (2009)

An example of base substitution



Chr3

We are looking at this position (3q26.32)

57 bp

matched normal sequence

Blood sample has the wildtype (G) allele

Lung adenocarcinoma sequence

G to A mutation in PIK3CA gene in cancer

PIK3CA gene

Mutational signature 개념을 접하다

Genome Research (2012.3)

Research

A transforming *KIF5B* and *RET* gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing

Young Seok Ju,[1,2] Won-Chul Lee,[1,3] Jong-Yeon Shin,[1,4] Seungbok Lee,[1,3] Thomas Bleazard,[1] Jae-Kyung Won,[5] Young Tae Kim,[6,7] Jong-Il Kim,[1,3,4,8] Jin-Hyoung Kang,[9] and Jeong-Sun Seo[1,2,3,4,8,10]

[1]Genomic Medicine Institute (GMI), Medical Research Center, Seoul National University, Seoul 110-799, Korea; [2]Macrogen Inc., Seoul 153-781, Korea; [3]Department of Biomedical Sciences, Seoul National University Graduate School, Seoul 110-799, Korea; [4]Psoma Therapeutics Inc., Seoul 153-781, Korea; [5]Molecular Pathology Center, Seoul National University Cancer Hospital, Seoul 110-744, Korea; [6]Department of Thoracic and Cardiovascular Surgery, Clinical Research Institute, Seoul National University Hospital, Seoul 110-799, Korea; [7]Cancer Research Institute, Seoul National University College of Medicine, Seoul 110-799, Korea, [8]Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul 110-799, Korea; [9]Department of Internal Medicine, Seoul St. Mary's Hospital, The Catholic University, Seoul 137-040, Korea

The identification of the molecular events that drive cancer transformation is essential to the development of targeted agents that improve the clinical outcome of lung cancer. Many studies have reported genomic driver mutations in non-small-cell lung cancers (NSCLCs) over the past decade; however, the molecular pathogenesis of >40% of NSCLCs is still unknown. To identify new molecular targets in NSCLCs, we performed the combined analysis of massively parallel whole-genome and transcriptome sequencing for cancer and paired normal tissue of a 33-yr-old lung adenocarcinoma patient, who is a never-smoker and has no familial cancer history. The cancer showed no known driver mutation in *EGFR* or *KRAS* and no *EML4-ALK* fusion. Here we report a novel fusion gene between *KIF5B* and the *RET* proto-oncogene caused by a pericentric inversion of 10p11.22–q11.21. This fusion gene overexpresses chimeric RET receptor tyrosine kinase, which could spontaneously induce cellular transformation. We identified the *KIF5B-RET* fusion in two more cases out of 20 primary lung adenocarcinomas in the replication study. Our data demonstrate that a subset of NSCLCs could be caused by a fusion of *KIF5B* and *RET*, and suggest the chimeric oncogene as a promising molecular target for the personalized diagnosis and treatment of lung cancer.

[Supplemental material is available for this article.]

Lung cancer remains a leading cause of mortality in cancer, with around 1.38 million deaths worldwide annually (Ferlay et al. 2010). With a conventional chemotherapeutic regimen, the median survival time for lung cancer patients in advanced stages is <1 yr from diagnosis (Schiller et al. 2002). Tobacco smoking is known to be a major risk factor of lung cancer in Western countries, where 85%–90% of all lung cancers were attributed to smoking (Toh et al. 2006). However, ~25% of lung cancer patients worldwide are "never-smokers" (Lee et al. 2011). Data from many Asian countries have shown that never-smokers constitute 30%–40% of non-small-cell lung cancer (NSCLC). NSCLC accounts for ~80% of lung cancer cases (Subramanian and Govindan 2007), and the dominant histological type is adenocarcinoma (>50%) (Pao and Girard 2011).

Lung cancer of never-smokers tends to be driven by single somatic mutation events, rather than global genetic and epigenetic changes (Lee et al. 2011). A subset of somatic mutations has been reported in NSCLCs in the past few years, such as *EGFR*, *KRAS*, and *EML4-ALK* genes (which are conventionally called the triple-markers) (Pao and Girard 2011). Mutations in the tyrosine kinase domain of *EGFR*, which are associated preferentially with NSCLCs of non-smokers and Asians, are sensitive to *EGFR*-targeted therapy, such as gefitinib (Paez et al. 2004). Missense mutations in *KRAS* are common in the lung adenocarcinomas of smokers and induce resistance to *EGFR* inhibitors (Riely et al. 2008). More recently, the *EML4-ALK* fusion gene was identified in NSCLC (Soda et al. 2007), which is generated by inversion in chromosome 2. This fusion gene, formed by chromosomal rearrangement, is more frequently detected in the lung adenocarcinoma of young patients, regardless of ethnicity, with no or little history of cigarette smoking (Wong et al. 2009). *ALK*-positive lung cancer constitutes ~5% of NSCLCs and is highly sensitive to *ALK* inhibitors, such as crizotinib (Pao and Girard 2011).

Although several genetic mutations have been reported previously, a large proportion of lung cancer patients have been observed to have none of them in their cancer genome. More than 40% of NSCLCs appear to be driven by unknown genetic events (Harris 2010; Pao and Girard 2011).

Here we report a novel fusion gene generated by a chromosomal inversion event in a young, never-smoker lung adenocarcinoma patient, whose cancer was negative for the triple-markers, using massively parallel DNA and RNA sequencing. The patient, known as AK55, was healthy until he was 33 yr of age, when a poorly differentiated adenocarcinoma developed in the right upper lobe of a lung (Fig. 1A). He had no known family history of cancers

[10]Corresponding author.
E-mail jeongsun@snu.ac.kr.

Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.133645.111. Freely available online through the Genome Research Open Access option.

22:436–445 © 2012 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/12; www.genome.org

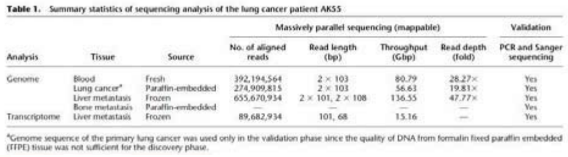
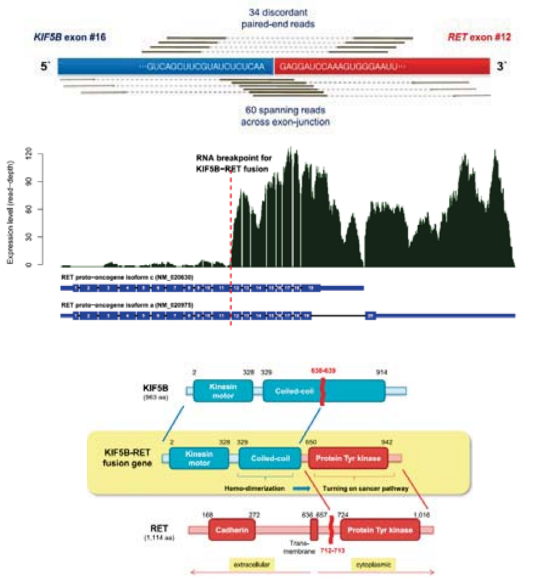


Table 1. Summary statistics of sequencing analysis of the lung cancer patient AK55

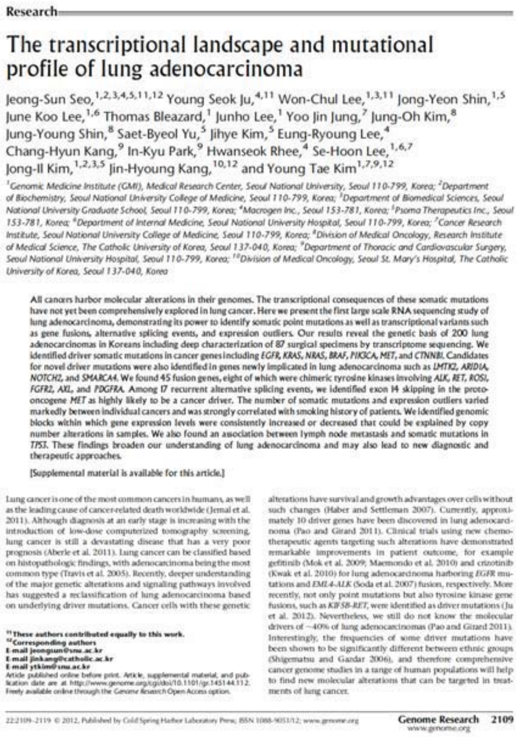
			Massively parallel sequencing (mappable)				Validation
Analysis	Tissue	Source	No. of aligned reads	Read length (bp)	Throughput (Gbp)	Read depth (fold)	PCR and Sanger sequencing
Genome	Blood	Fresh	392,194,564	2 × 103	80.79	28.27×	Yes
	Lung cancer[a]	Paraffin-embedded	274,909,815	2 × 103	56.63	19.81×	Yes
	Liver metastasis	Frozen	655,670,934	2 × 101, 2 × 108	136.55	47.77×	Yes
	Bone metastasis	Paraffin-embedded	—	—	—	—	Yes
Transcriptome	Liver metastasis	Frozen	89,682,934	101, 68	15.16	—	Yes

[a]Genome sequence of the primary lung cancer was used only in the validation phase since the quality of DNA from formalin fixed paraffin embedded (FFPE) tissue was not sufficient for the discovery phase.



Ju YS *et al.*, Genome Res (2012a)

Mutational signature 개념을 접하다 (2)



Research

The transcriptional landscape and mutational profile of lung adenocarcinoma

Jeong-Sun Seo,[1,2,3,4,5,11,12] Young Seok Ju,[4,11] Won-Chul Lee,[1,3,11] Jong-Yeon Shin,[1,5] June Koo Lee,[1,6] Thomas Bleazard,[1] Junho Lee,[1] Yoo Jin Jung,[7] Jung-Oh Kim,[8] Jung-Young Shin,[8] Saet-Byeol Yu,[5] Jihye Kim,[5] Eung-Ryoung Lee,[4] Chang-Hyun Kang,[9] In-Kyu Park,[9] Hwanseok Rhee,[4] Se-Hoon Lee,[1,6,7] Jong-Il Kim,[1,2,3,5] Jin-Hyoung Kang,[10,12] and Young Tae Kim[1,7,9,12]

[1]Genomic Medicine Institute (GMI), Medical Research Center, Seoul National University, Seoul 110-799, Korea; [2]Department of Biochemistry, Seoul National University College of Medicine, Seoul 110-799, Korea; [3]Department of Biomedical Sciences, Seoul National University Graduate School, Seoul 110-799, Korea; [4]Macrogen Inc., Seoul 153-781, Korea; [5]Psoma Therapeutics Inc., Seoul 153-781, Korea; [6]Department of Internal Medicine, Seoul National University Hospital, Seoul 110-799, Korea; [7]Cancer Research Institute, Seoul National University College of Medicine, Seoul 110-799, Korea; [8]Division of Medical Oncology, Research Institute of Medical Science, The Catholic University of Korea, Seoul 137-040, Korea; [9]Department of Thoracic and Cardiovascular Surgery, Seoul National University Hospital, Seoul 110-799, Korea; [10]Division of Medical Oncology, Seoul St. Mary's Hospital, The Catholic University of Korea, Seoul 137-040, Korea

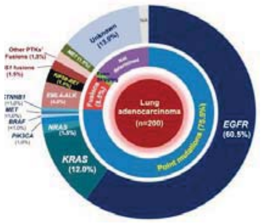
All cancers harbor molecular alterations in their genomes. The transcriptional consequences of these somatic mutations have not yet been comprehensively explored in lung cancer. Here we present the first large scale RNA sequencing study of lung adenocarcinoma, demonstrating its power to identify somatic point mutations as well as transcriptional variants such as gene fusions, alternative splicing events, and expression outliers. Our results reveal the genetic basis of 200 lung adenocarcinomas in Koreans including deep characterization of 87 surgical specimens by transcriptome sequencing. We identified driver somatic mutations in cancer genes including *EGFR, KRAS, NRAS, BRAF, PIK3CA, MET*, and *CTNNB1*. Candidates for novel driver mutations were also identified in genes newly implicated in lung adenocarcinoma such as *LMTK2, ARID1A, NOTCH2*, and *SMARCA4*. We found 45 fusion genes, eight of which were chimeric tyrosine kinases involving *ALK, RET, ROS1, FGFR2, AXL*, and *PDGFRA*. Among 17 recurrent alternative splicing events, we identified exon 14 skipping in the proto-oncogene *MET* as highly likely to be a cancer driver. The number of somatic mutations and expression outliers varied markedly between individual cancers and was strongly correlated with smoking history of patients. We identified genomic blocks within which gene expression levels were consistently increased or decreased that could be explained by copy number alterations in samples. We also found an association between lymph node metastasis and somatic mutations in *TP53*. These findings broaden our understanding of lung adenocarcinoma and may also lead to new diagnostic and therapeutic approaches.

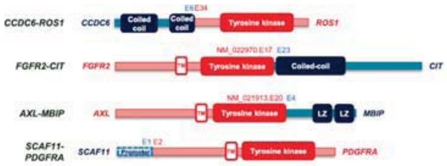
[Supplemental material is available for this article.]

Lung cancer is one of the most common cancers in humans, as well as the leading cause of cancer-related death worldwide (Jemal et al. 2011). Although diagnosis at an early stage is increasing with the introduction of low-dose computerized tomography screening, lung cancer is still a devastating disease that has a very poor prognosis (Aberle et al. 2011). Lung cancer can be classified based on histopathologic findings, with adenocarcinoma being the most common type (Travis et al. 2005). Recently, deeper understanding of the major genetic alterations and signaling pathways involved has suggested a reclassification of lung adenocarcinoma based on underlying driver mutations. Cancer cells with these genetic alterations have survival and growth advantages over cells without such changes (Haber and Settleman 2007). Currently, approximately 10 driver genes have been discovered in lung adenocarcinoma (Pao and Girard 2011). Clinical trials using new chemotherapeutic agents targeting such alterations have demonstrated remarkable improvements in patient outcome, for example gefitinib (Mok et al. 2009; Maemondo et al. 2010) and crizotinib (Kwak et al. 2010) for lung adenocarcinoma harboring *EGFR* mutations and *EML4-ALK* (Soda et al. 2007) fusion, respectively. More recently, not only point mutations but also tyrosine kinase gene fusions, such as *KIF5B-RET*, were identified as driver mutations (Ju et al. 2012). Nevertheless, we still do not know the molecular drivers of ~40% of lung adenocarcinomas (Pao and Girard 2011). Interestingly, the frequencies of some driver mutations have been shown to be significantly different between ethnic groups (Shigematsu and Gazdar 2006), and therefore comprehensive cancer genome studies in a range of human populations will help to find new molecular alterations that can be targeted in treatments of lung cancer.

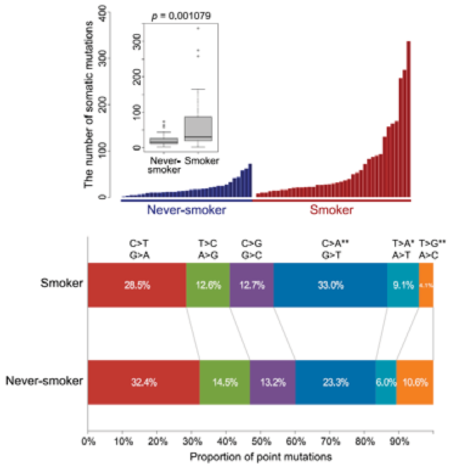
[11]These authors contributed equally to this work.
[12]Corresponding authors
E-mail jeongsun@snu.ac.kr
E-mail jinkang@catholic.ac.kr
E-mail ytkim@snu.ac.kr
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.145144.112. Freely available online through the Genome Research Open Access option.

22:2109–2119 © 2012, Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/12; www.genome.org





Dr. Myles Axton
(former Chief Editor
@ Nature Genetics)



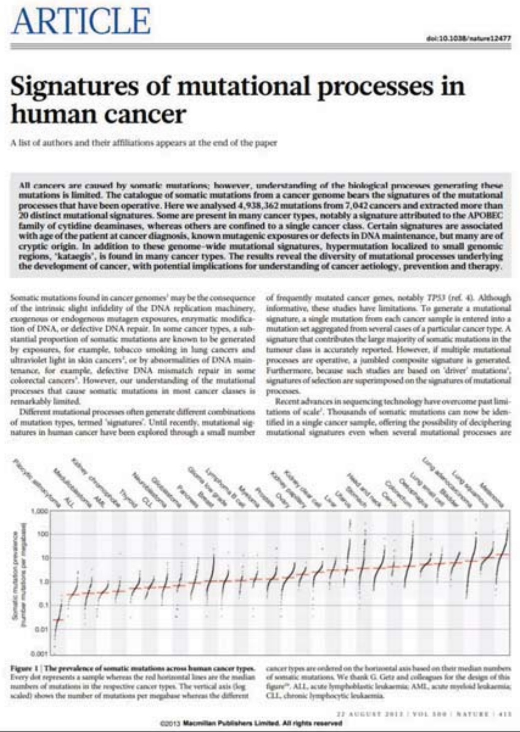
SBi 한국생명정보학회 Korean Society for Bioinformatics

Seo JS *et al.*, Genome Res (2012b)

Mutational signature 개념을 접하다 (3)

Ludmil B Alexandrov

Elizabeth P Murchison



ARTICLE

doi:10.1038/nature12477

Signatures of mutational processes in human cancer

A list of authors and their affiliations appears at the end of the paper

All cancers are caused by somatic mutations; however, understanding of the biological processes generating these mutations is limited. The catalogue of somatic mutations from a cancer genome bears the signatures of the mutational processes that have been operative. Here we analysed 4,938,362 mutations from 7,042 cancers and extracted more than 20 distinct mutational signatures. Some are present in many cancer types, notably a signature attributed to the APOBEC family of cytidine deaminases, whereas others are confined to a single cancer class. Certain signatures are associated with age of the patient at cancer diagnosis, known mutagenic exposures or defects in DNA maintenance, but many are of cryptic origin. In addition to these genome-wide mutational signatures, hypermutation localized to small genomic regions, 'kataegis', is found in many cancer types. The results reveal the diversity of mutational processes underlying the development of cancer, with potential implications for understanding of cancer aetiology, prevention and therapy.

Somatic mutations found in cancer genomes[1] may be the consequence of the intrinsic slight infidelity of the DNA replication machinery, exogenous or endogenous mutagen exposures, enzymatic modification of DNA, or defective DNA repair. In some cancer types, a substantial proportion of somatic mutations are known to be generated by exposures, for example, tobacco smoking in lung cancers and ultraviolet light in skin cancers[2], or by abnormalities of DNA maintenance, for example, defective DNA mismatch repair in some colorectal cancers[3]. However, our understanding of the mutational processes that cause somatic mutations in most cancer classes is remarkably limited.

Different mutational processes often generate different combinations of mutation types, termed 'signatures'. Until recently, mutational signatures in human cancer have been explored through a small number of frequently mutated cancer genes, notably *TP53* (ref. 4). Although informative, these studies have limitations. To generate a mutational signature, a single mutation from each cancer sample is entered into a mutation set aggregated from several cases of a particular cancer type. A signature that contributes the large majority of somatic mutations in the tumour class is accurately reported. However, if multiple mutational processes are operative, a jumbled composite signature is generated. Furthermore, because such studies are based on 'driver' mutations[5], signatures of selection are superimposed on the signatures of mutational processes.

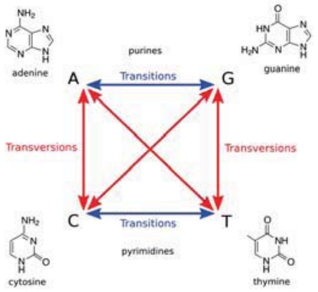
Recent advances in sequencing technology have overcome past limitations of scale[1]. Thousands of somatic mutations can now be identified in a single cancer sample, offering the possibility of deciphering mutational signatures even when several mutational processes are

Figure 1 | The prevalence of somatic mutations across human cancer types. Every dot represents a sample whereas the red horizontal lines are the median numbers of mutations in the respective cancer types. The vertical axis (log scaled) shows the number of mutations per megabase whereas the different cancer types are ordered on the horizontal axis based on their median numbers of somatic mutations. We thank G. Getz and colleagues for the design of this figure[16]. ALL, acute lymphoblastic leukaemia; AML, acute myeloid leukaemia; CLL, chronic lymphocytic leukaemia.

SBi 한국생명정보학회 Korean Society for Bioinformatics

Alexandrov L *et al.*, ***Nature*** (2013)

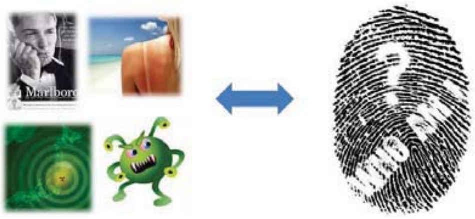
mutational origin: Mutation은 랜덤하게 생기는 것이 아니다



6 classes of base substitutions

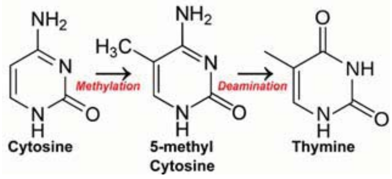
C>A (G>T), C>G (G>C), C>T (G>A)

T>A (A>T), T>C (A>G), T>G (A>C)

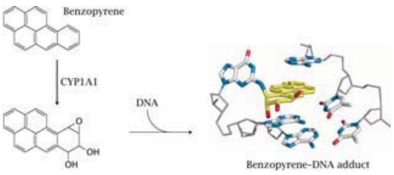


돌연변이는 “랜덤” 이 아니라 DNA damage x DNA repair 과정

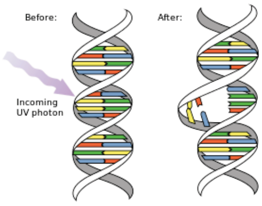
Spontaneous cytosine deamination
C>T substitutions
(mostly at CpG context)



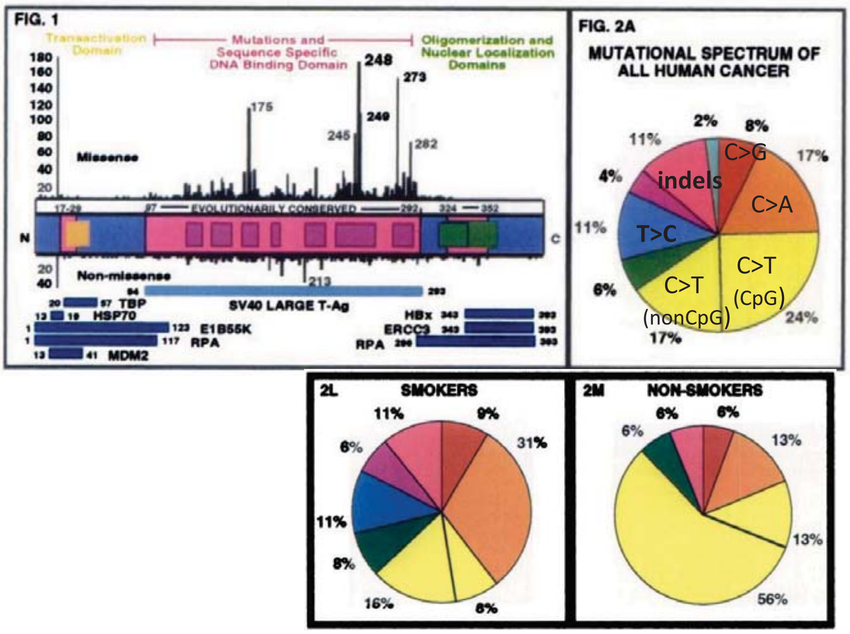
Tobacco smoking
C>A substitutions



Ultraviolet (UV) light
C>T substitutions
(CC>TT)



Classical observation



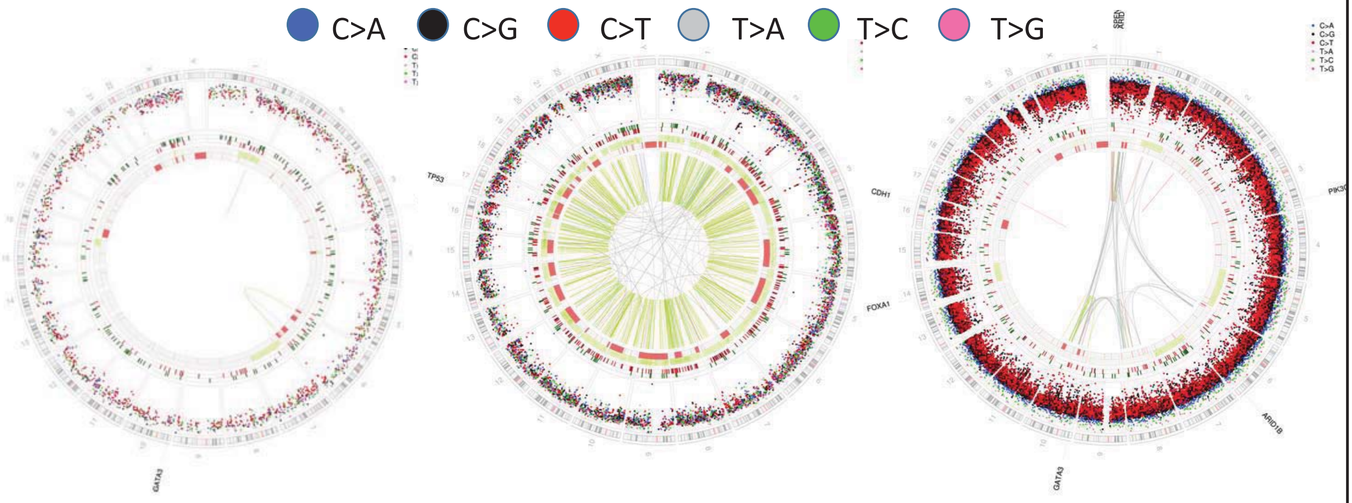
- 암종마다, 그리고 발암물질의 노출에 따라서 TP53 유전자에 생기는 돌연변이 패턴이 상이하다

Greenblatt *et al.*, *Cancer Research* (1994)

Mutational signature의 예시

- 폐암의 전장 유전체 분석에서 20,000개의 base substitution 발견
  - 이 가운데 80%가 C>A mutations. 주된 돌연변이 발생기전은?
  - (흡연에 노출)
- 흑색종의 전장 유전체 분석에서 20,000 개의 base substitution 발견
  - 이 가운데 90%가 C>T mutations 이고 수백개의 CC>TT 도 같이 발견
    주된 돌연변이 발생기전은?
  - (UV에 노출)
- 실제로는 하나의 암 유전체에서 발생하는 돌연변이들이
  위와 같은 단일 돌연변이 발생 기전이 아니라,
  여러 돌연변이 기전의 '조합' 으로 만들어지는 일이 훨씬 흔하다

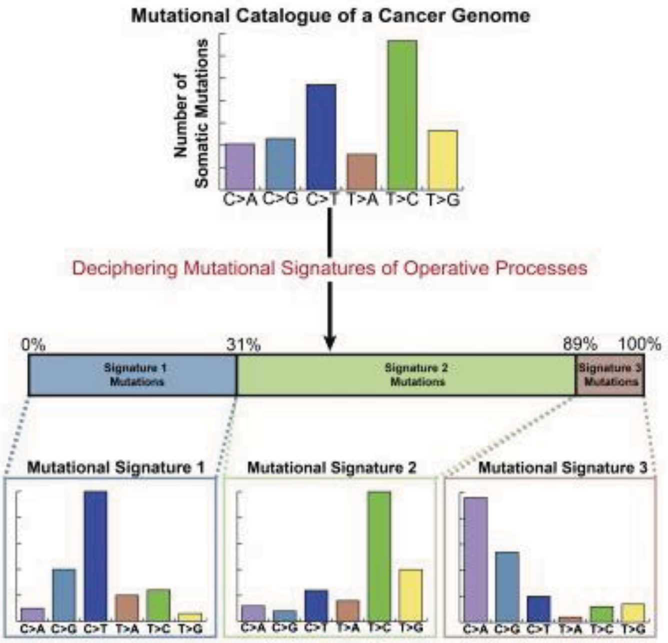
실제 3개의 breast cancer whole-genome sequencing 결과



3000 base substitutions ~50% C>T ~30% T>C	20,000 base substitutions ~15% C>A ~20% C>G ~20% C>T ~15% T>A ~20% T>C ~10% T>G	100,000 base substitutions ~40% C>G ~50% C>T
일반적인 ER+ breast cancer	BRCA1 mutant breast cancer	APOBEC hyperactivated breast cancer

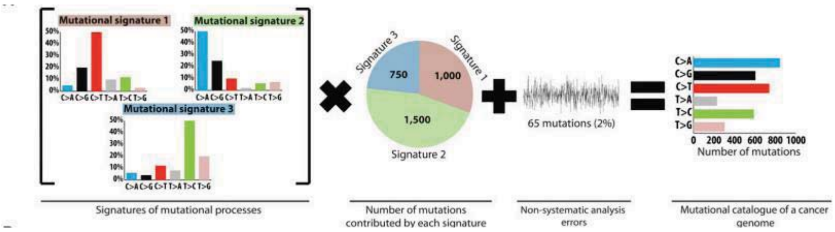
Tumor의 돌연변이 스펙트럼은 이론적으로 n개의 서로 다른 Mutational process로 설명된다

하지만 우리는 n이 얼마인지도, 각각의 process의 spectrum도 알고있지 못한다



Alexandrov L *et al.*, ***Cell reports*** (2013)

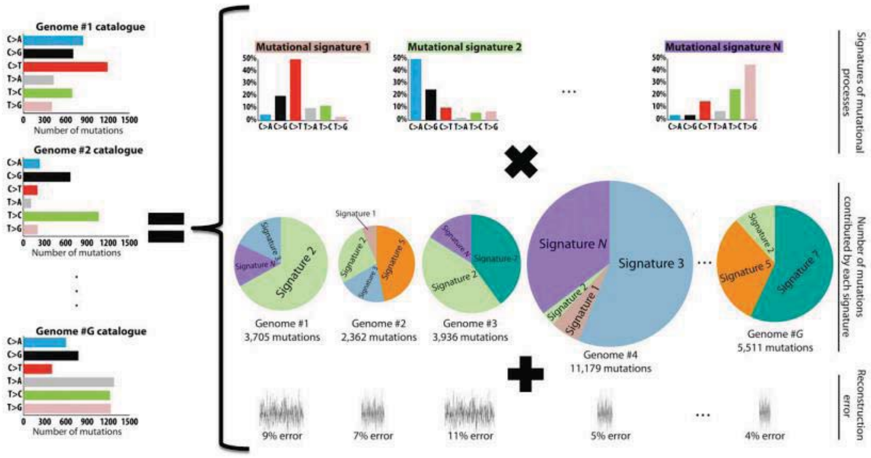
Understanding mutational processes from mutational spectrum: a blind-source separation problem



Somatic mutations explored in a sample can be explained by linear sum of different exposures

With genome big-dataset

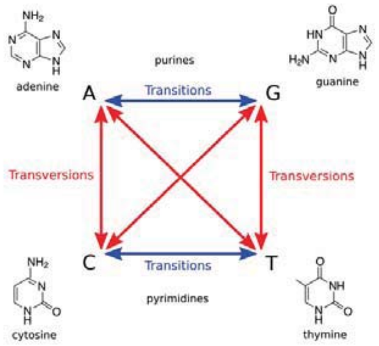
& using NMF
(or other equivalent algorithms)



Alexandrov L *et al.*, ***Cell reports*** (2013)

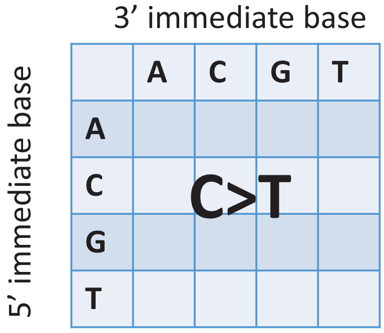
Single base substitutions (SBS) into 96 subclasses

- C>A (G>T)
- C>G (G>C)
- C>T (G>A)
- T>A (A>T)
- T>C (A>G)
- T>G (A>C)



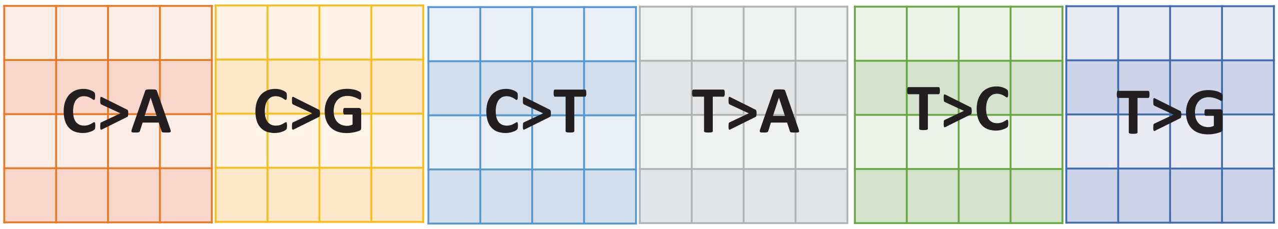
sequence context

5’B - Wt > Var- 3’B

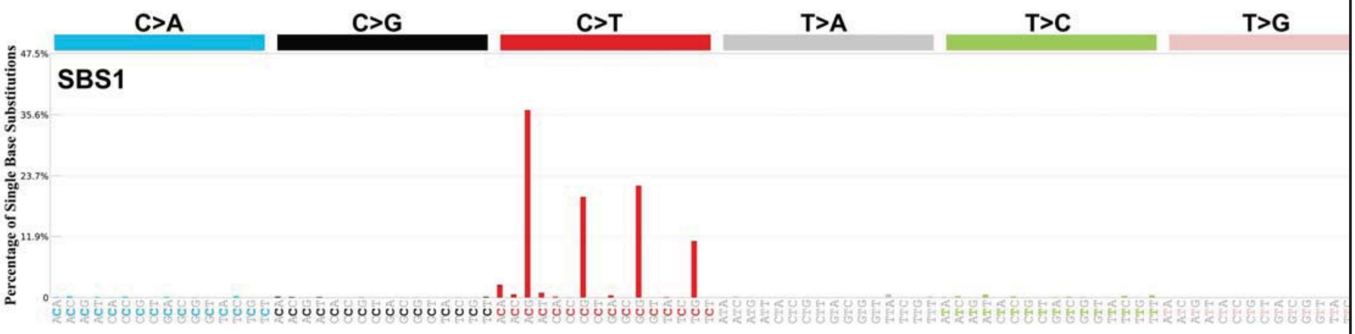


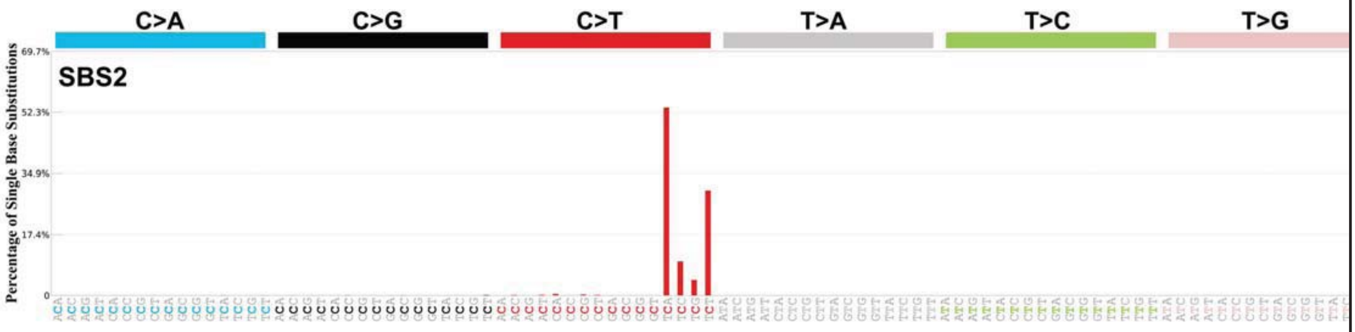
4 x 6 types x 4

= 96 subtypes

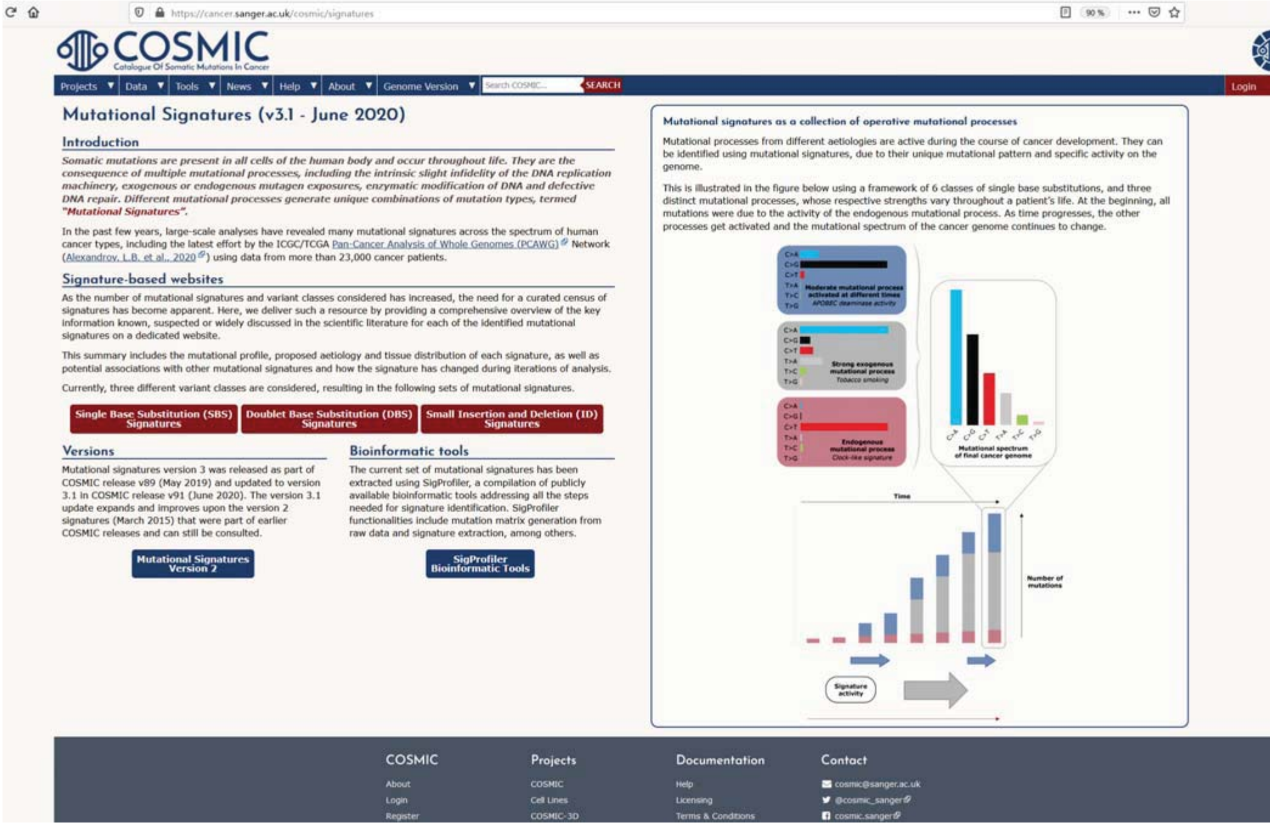


SBS Signature 1 and Signature 2





Dictionary for mutational signatures: COSMIC



Mutational Signatures (v3.1 - June 2020)

Introduction

Somatic mutations are present in all cells of the human body and occur throughout life. They are the consequence of multiple mutational processes, including the intrinsic slight infidelity of the DNA replication machinery, exogenous or endogenous mutagen exposures, enzymatic modification of DNA and defective DNA repair. Different mutational processes generate unique combinations of mutation types, termed "Mutational Signatures".

In the past few years, large-scale analyses have revealed many mutational signatures across the spectrum of human cancer types, including the latest effort by the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes (PCAWG) Network (Alexandrov, L.B. et al., 2020) using data from more than 23,000 cancer patients.

Signature-based websites

As the number of mutational signatures and variant classes considered has increased, the need for a curated census of signatures has become apparent. Here, we deliver such a resource by providing a comprehensive overview of the key information known, suspected or widely discussed in the scientific literature for each of the identified mutational signatures on a dedicated website.

This summary includes the mutational profile, proposed aetiology and tissue distribution of each signature, as well as potential associations with other mutational signatures and how the signature has changed during iterations of analysis.

Currently, three different variant classes are considered, resulting in the following sets of mutational signatures.

Single Base Substitution (SBS) Signatures
Doublet Base Substitution (DBS) Signatures
Small Insertion and Deletion (ID) Signatures

Versions

Mutational signatures version 3 was released as part of COSMIC release v89 (May 2019) and updated to version 3.1 in COSMIC release v91 (June 2020). The version 3.1 update expands and improves upon the version 2 signatures (March 2015) that were part of earlier COSMIC releases and can still be consulted.

Bioinformatic tools

The current set of mutational signatures has been extracted using SigProfiler, a compilation of publicly available bioinformatic tools addressing all the steps needed for signature identification. SigProfiler functionalities include mutation matrix generation from raw data and signature extraction, among others.

Mutational signatures as a collection of operative mutational processes

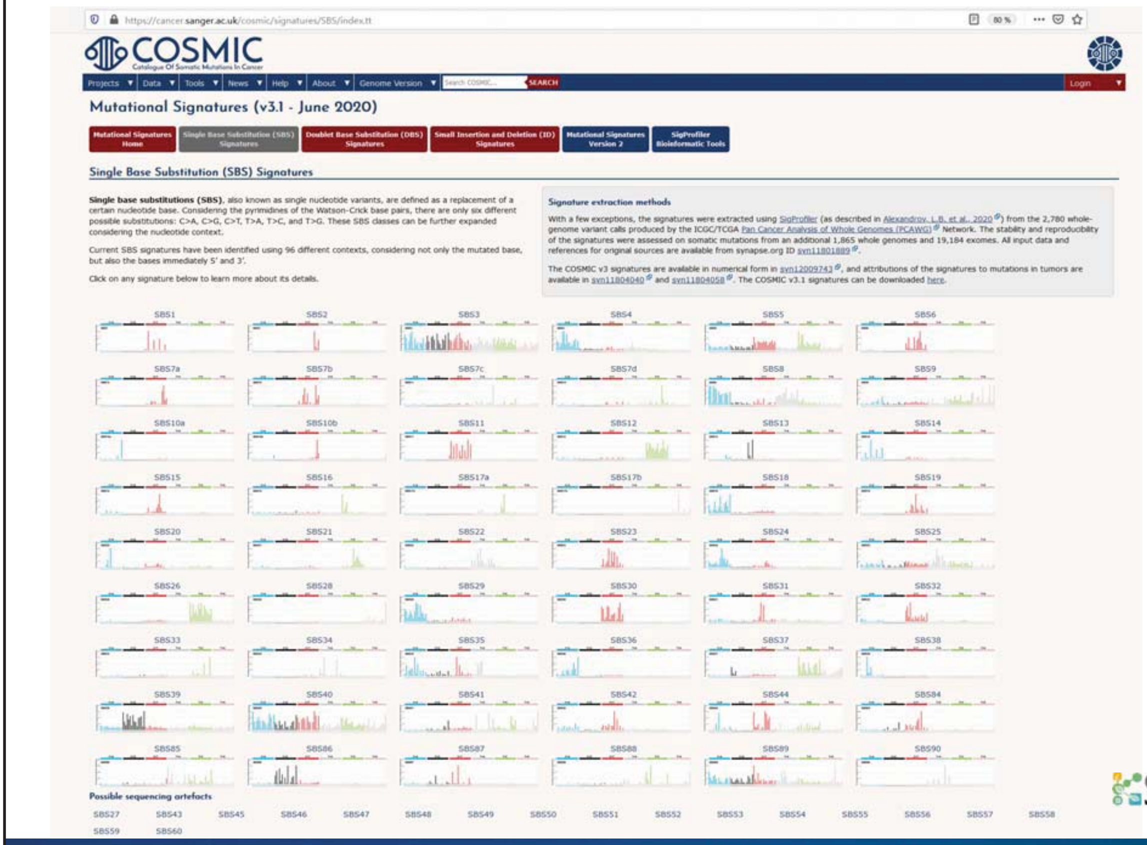
Mutational processes from different aetiologies are active during the course of cancer development. They can be identified using mutational signatures, due to their unique mutational pattern and specific activity on the genome.

This is illustrated in the figure below using a framework of 6 classes of single base substitutions, and three distinct mutational processes, whose respective strengths vary throughout a patient's life. At the beginning, all mutations were due to the activity of the endogenous mutational process. As time progresses, the other processes get activated and the mutational spectrum of the cancer genome continues to change.

C>A
C>G
C>T
T>A
T>C
T>G
Moderate mutational process activated at different times
APOBEC deaminase activity
Strong exogenous mutational process
Tobacco smoking
Endogenous mutational process
Clock-like signature
Mutational spectrum of final cancer genome
Time
Number of mutations
Signature activity

cancer.sanger.ac.uk/cosmic/signatures

49 +5 biologic signatures in SBS mutations (v3)


https://cancer.sanger.ac.uk/cosmic/signatures/SBS/index.tt
80 %
COSMIC
Catalogue Of Somatic Mutations In Cancer
Projects ▼
Data ▼
Tools ▼
News ▼
Help ▼
About ▼
Genome Version ▼
Search COSMIC...
SEARCH
Login ▼

Mutational Signatures (v3.1 - June 2020)

Mutational Signatures Home
Single Base Substitution (SBS) Signatures
Doublet Base Substitution (DBS) Signatures
Small Insertion and Deletion (ID) Signatures
Mutational Signatures Version 2
SigProfiler Bioinformatic Tools

Single Base Substitution (SBS) Signatures

Single base substitutions (SBS), also known as single nucleotide variants, are defined as a replacement of a certain nucleotide base. Considering the pyrimidines of the Watson-Crick base pairs, there are only six different possible substitutions: C>A, C>G, C>T, T>A, T>C, and T>G. These SBS classes can be further expanded considering the nucleotide context.

Current SBS signatures have been identified using 96 different contexts, considering not only the mutated base, but also the bases immediately 5' and 3'.

Click on any signature below to learn more about its details.

Signature extraction methods

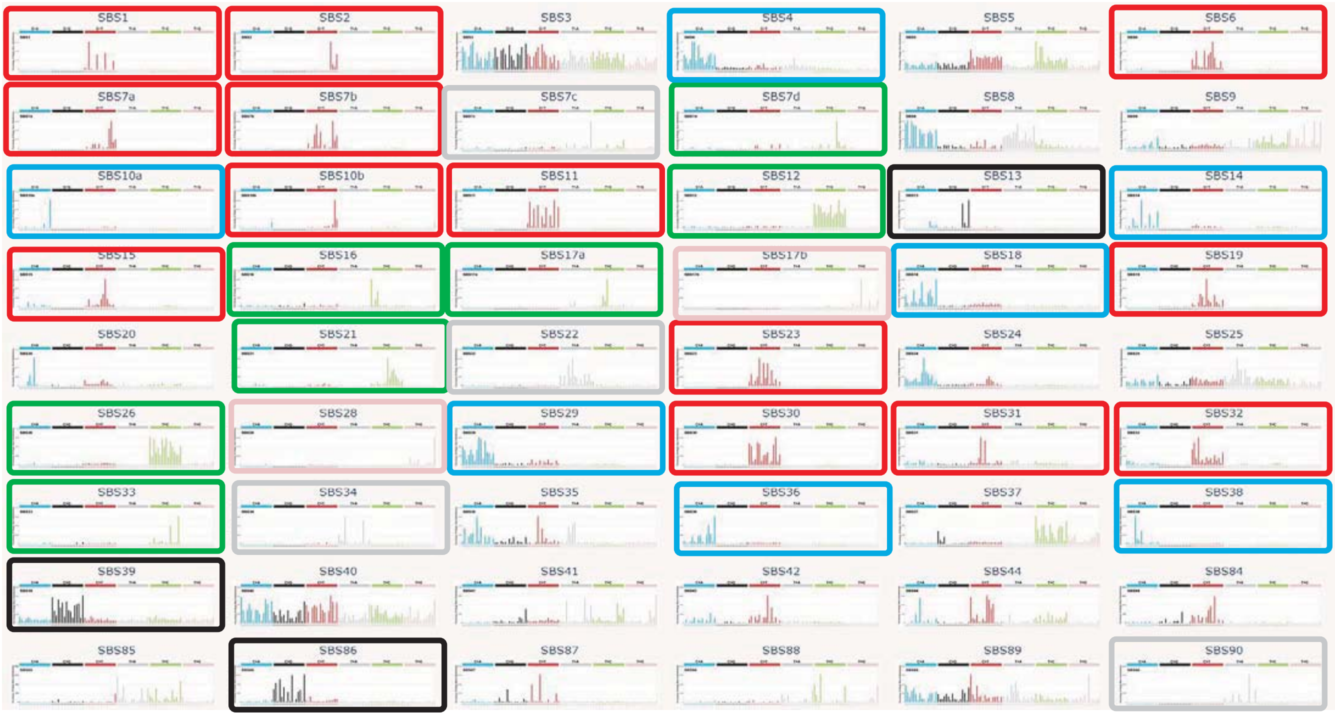
With a few exceptions, the signatures were extracted using SigProfiler (as described in Alexandrov, L.B. et al., 2020) from the 2,780 whole-genome variant calls produced by the ICGC/TCGA Pan Cancer Analysis of Whole Genomes (PCAWG) Network. The stability and reproducibility of the signatures were assessed on somatic mutations from an additional 1,865 whole genomes and 19,184 exomes. All input data and references for original sources are available from synapse.org ID syn11801889.

The COSMIC v3 signatures are available in numerical form in syn12009743, and attributions of the signatures to mutations in tumors are available in syn11804040 and syn11804058. The COSMIC v3.1 signatures can be downloaded here.

Possible sequencing artefacts

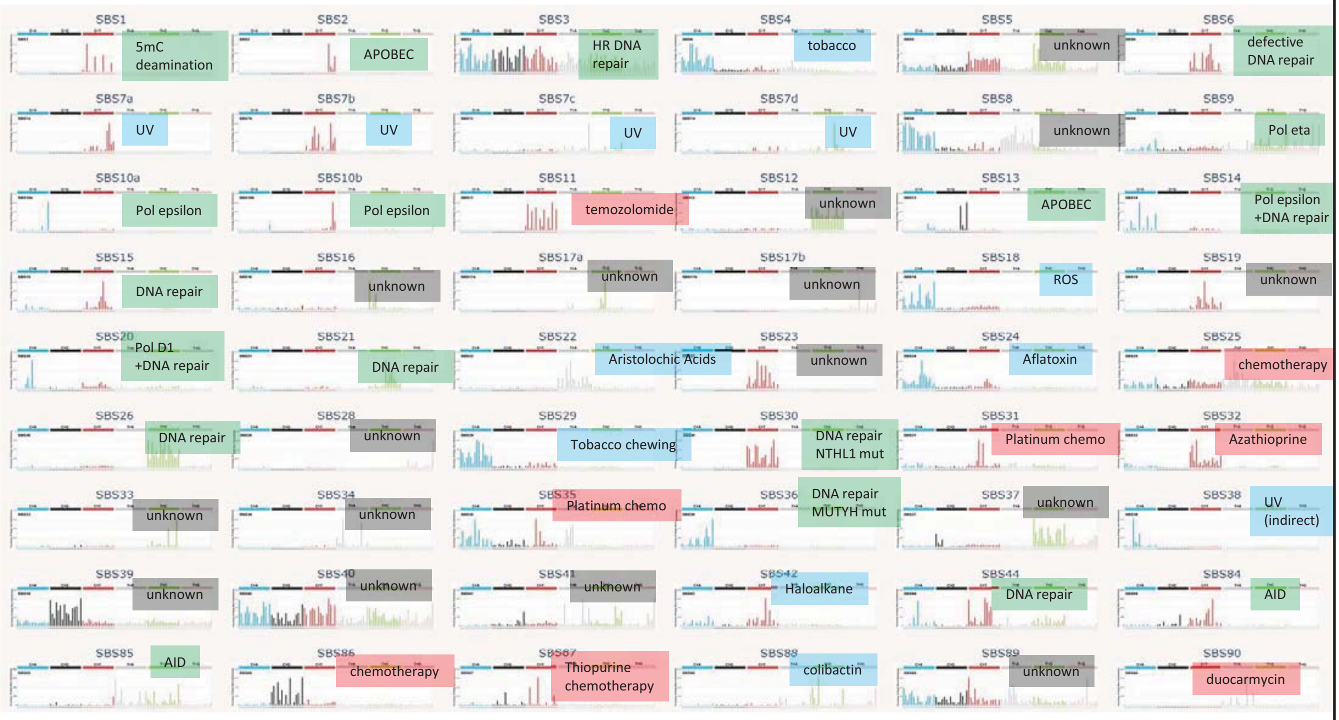
SBS27 SBS43 SBS45 SBS46 SBS47 SBS48 SBS49 SBS50 SBS51 SBS52 SBS53 SBS54 SBS55 SBS56 SBS57 SBS58 SBS59 SBS60

Signatures by patterns



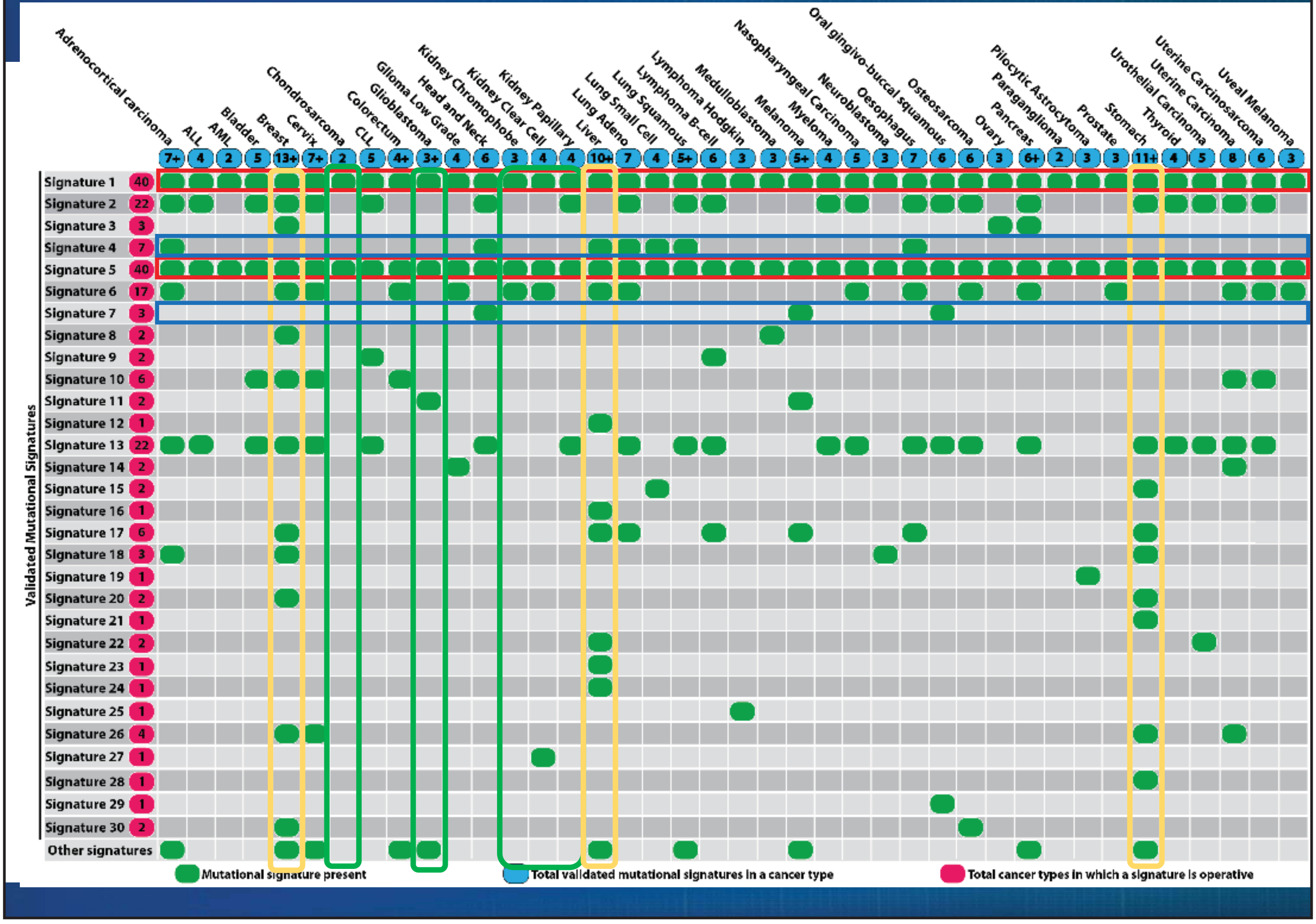
SBi 한국생명정보학회 Korean Society for Bioinformatics

Signatures by etiology



SBi 한국생명정보학회 Korean Society for Bioinformatics

Extensive cell type specificity



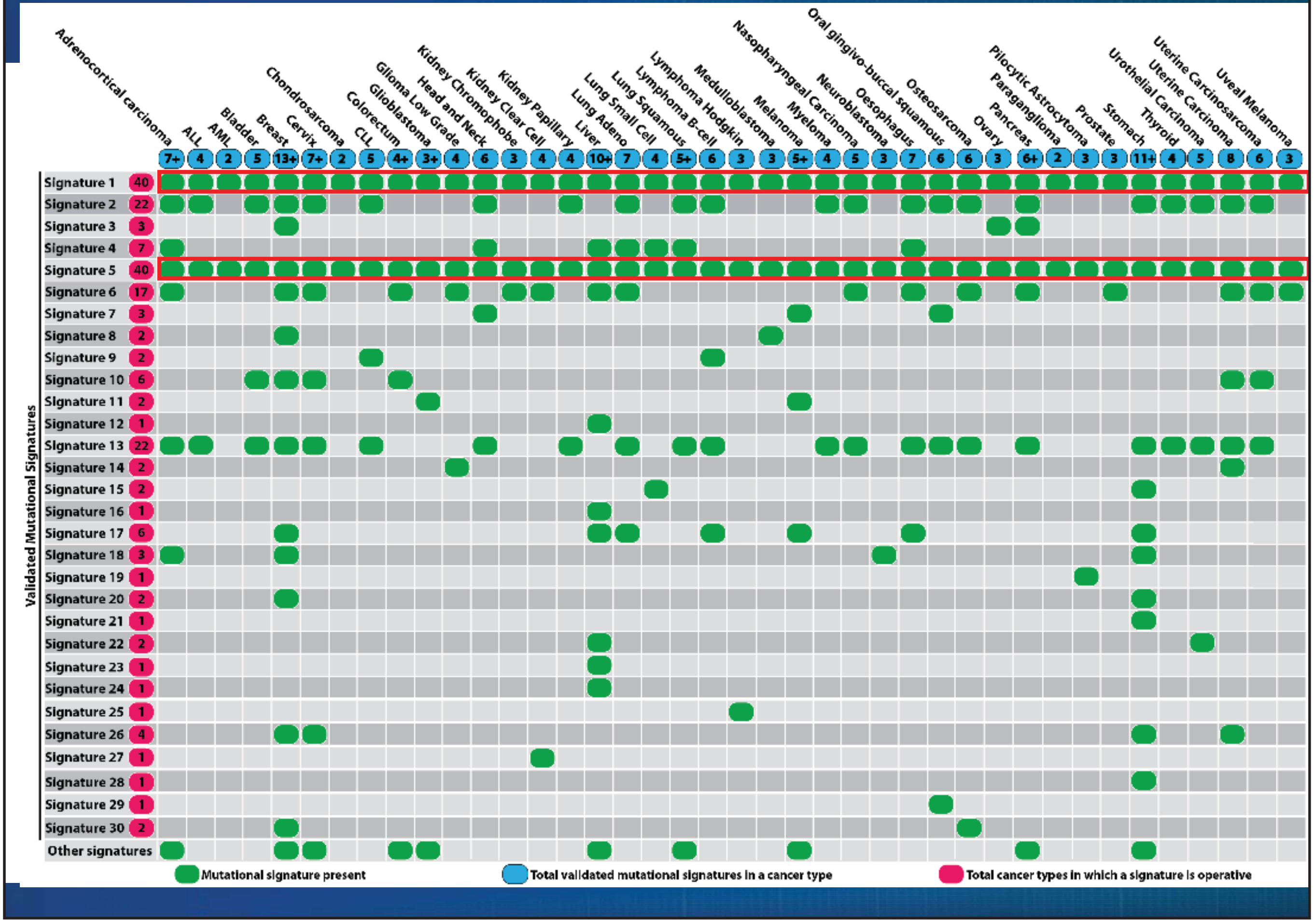
Validated Mutational Signatures

		Adrenocortical carcinoma	ALL	AML	Bladder	Breast	Cervix	Chondrosarcoma	CLL	Colorectum	Glioblastoma	Glioma Low Grade	Head and Neck	Kidney Chromophobe	Kidney Clear Cell	Kidney Papillary	Liver	Lung Adeno	Lung Small Cell	Lung Squamous	Lymphoma B-cell
		7+	4	2	5	13+	7+	2	5	4+	3+	4	6	3	4	4	10+	7	4	5+	6
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩
Signature 3	3					🟩															
Signature 4	7	🟩											🟩				🟩	🟩	🟩	🟩	
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17	🟩				🟩	🟩			🟩		🟩		🟩	🟩		🟩	🟩			
Signature 7	3												🟩								
Signature 8	2					🟩															
Signature 9	2								🟩												🟩
Signature 10	6				🟩	🟩	🟩			🟩											
Signature 11	2										🟩										
Signature 12	1																🟩				
Signature 13	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩
Signature 14	2											🟩									
Signature 15	2																		🟩		
Signature 16	1																🟩				
Signature 17	6					🟩											🟩	🟩			🟩
Signature 18	3	🟩				🟩															
Signature 19	1																				
Signature 20	2					🟩															
Signature 21	1																				
Signature 22	2																🟩				
Signature 23	1																🟩				
Signature 24	1																🟩				
Signature 25	1																				
Signature 26	4					🟩	🟩														
Signature 27	1														🟩						
Signature 28	1																				
Signature 29	1																				
Signature 30	2					🟩															
Other signatures		🟩				🟩	🟩			🟩	🟩						🟩			🟩	

		Lymphoma Hodgkin	Medulloblastoma	Melanoma	Myeloma	Nasopharyngeal Carcinoma	Neuroblastoma	Oesophagus	Oral gingivo-buccal squamous	Osteosarcoma	Ovary	Pancreas	Paraganglioma	Pilocytic Astrocytoma	Prostate	Stomach	Thyroid	Urothelial Carcinoma	Uterine Carcinoma	Uterine Carcinosarcoma	Uveal Melanoma
		3	3	5+	4	5	3	7	6	6	3	6+	2	3	3	11+	4	5	8	6	3
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 3	3										🟩	🟩									
Signature 4	7							🟩													
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17					🟩		🟩		🟩		🟩			🟩				🟩	🟩	🟩
Signature 7	3			🟩					🟩												
Signature 8	2		🟩																		
Signature 9	2																				
Signature 10	6																		🟩	🟩	
Signature 11	2			🟩																	
Signature 12	1																				
Signature 13	22				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 14	2																		🟩		
Signature 15	2															🟩					
Signature 16	1																				
Signature 17	6			🟩				🟩								🟩					
Signature 18	3						🟩									🟩					
Signature 19	1													🟩							
Signature 20	2															🟩					
Signature 21	1															🟩					
Signature 22	2																	🟩			
Signature 23	1																				
Signature 24	1																				
Signature 25	1	🟩																			
Signature 26	4															🟩			🟩		
Signature 27	1																				
Signature 28	1															🟩					
Signature 29	1								🟩												
Signature 30	2									🟩											
Other signatures				🟩								🟩				🟩					

🟩 Mutational signature present 🟦 Total validated mutational signatures in a cancer type 🟥 Total cancer types in which a signature is operative

Signature 1 and 5: basal, cellular intrinsic mutagenesis



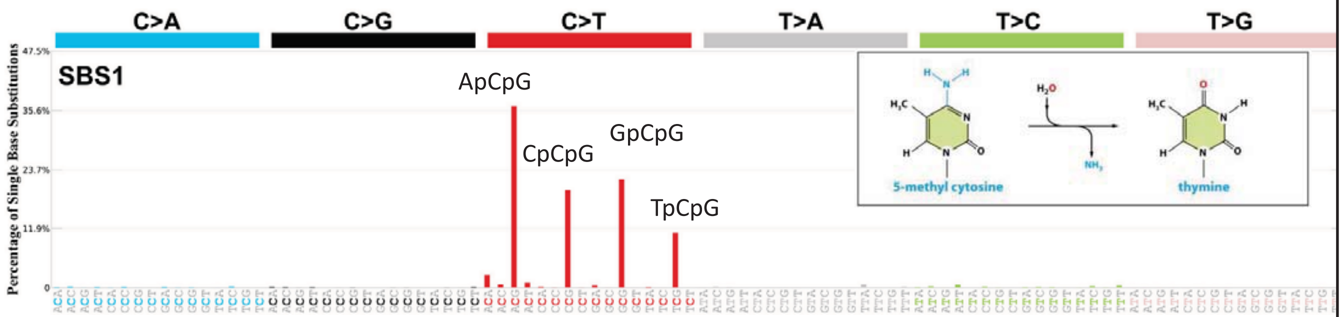
Validated Mutational Signatures

		Adrenocortical carcinoma	ALL	AML	Bladder	Breast	Cervix	Chondrosarcoma	CLL	Colorectum	Glioblastoma	Glioma Low Grade	Head and Neck	Kidney Chromophobe	Kidney Clear Cell	Kidney Papillary	Liver	Lung Adeno	Lung Small Cell	Lung Squamous	Lymphoma B-cell
		7+	4	2	5	13+	7+	2	5	4+	3+	4	6	3	4	4	10+	7	4	5+	6
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩
Signature 3	3					🟩															
Signature 4	7	🟩											🟩				🟩	🟩	🟩	🟩	
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17	🟩				🟩	🟩			🟩		🟩		🟩	🟩		🟩	🟩			
Signature 7	3												🟩								
Signature 8	2					🟩															
Signature 9	2								🟩												🟩
Signature 10	6				🟩	🟩	🟩			🟩											
Signature 11	2										🟩										
Signature 12	1																🟩				
Signature 13	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩
Signature 14	2											🟩									
Signature 15	2																		🟩		
Signature 16	1																🟩				
Signature 17	6					🟩											🟩	🟩			🟩
Signature 18	3	🟩				🟩															
Signature 19	1																				
Signature 20	2					🟩															
Signature 21	1																				
Signature 22	2																🟩				
Signature 23	1																🟩				
Signature 24	1																🟩				
Signature 25	1																				
Signature 26	4					🟩	🟩														
Signature 27	1														🟩						
Signature 28	1																				
Signature 29	1																				
Signature 30	2					🟩															
Other signatures		🟩				🟩	🟩			🟩	🟩						🟩			🟩	

		Lymphoma Hodgkin	Medulloblastoma	Melanoma	Myeloma	Nasopharyngeal Carcinoma	Neuroblastoma	Oesophagus	Oral gingivo-buccal squamous	Osteosarcoma	Ovary	Pancreas	Paraganglioma	Pilocytic Astrocytoma	Prostate	Stomach	Thyroid	Urothelial Carcinoma	Uterine Carcinoma	Uterine Carcinosarcoma	Uveal Melanoma
		3	3	5+	4	5	3	7	6	6	3	6+	2	3	3	11+	4	5	8	6	3
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 3	3										🟩	🟩									
Signature 4	7							🟩													
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17					🟩		🟩		🟩		🟩			🟩				🟩	🟩	🟩
Signature 7	3			🟩					🟩												
Signature 8	2		🟩																		
Signature 9	2																				
Signature 10	6																		🟩	🟩	
Signature 11	2			🟩																	
Signature 12	1																				
Signature 13	22				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 14	2																		🟩		
Signature 15	2															🟩					
Signature 16	1																				
Signature 17	6			🟩				🟩								🟩					
Signature 18	3						🟩									🟩					
Signature 19	1													🟩							
Signature 20	2															🟩					
Signature 21	1															🟩					
Signature 22	2																	🟩			
Signature 23	1																				
Signature 24	1																				
Signature 25	1	🟩																			
Signature 26	4															🟩			🟩		
Signature 27	1																				
Signature 28	1															🟩					
Signature 29	1								🟩												
Signature 30	2									🟩											
Other signatures				🟩								🟩				🟩					

🟩 Mutational signature present 🟦 Total validated mutational signatures in a cancer type 🟥 Total cancer types in which a signature is operative

(1) SBS Signature 1: 5mC deamination



Cancer types:

Signature 1 has been found in all cancer types and in most cancer samples.

Proposed etiology:

Signature 1 is the result of an endogenous mutational process initiated by spontaneous deamination of 5-methylcytosine.

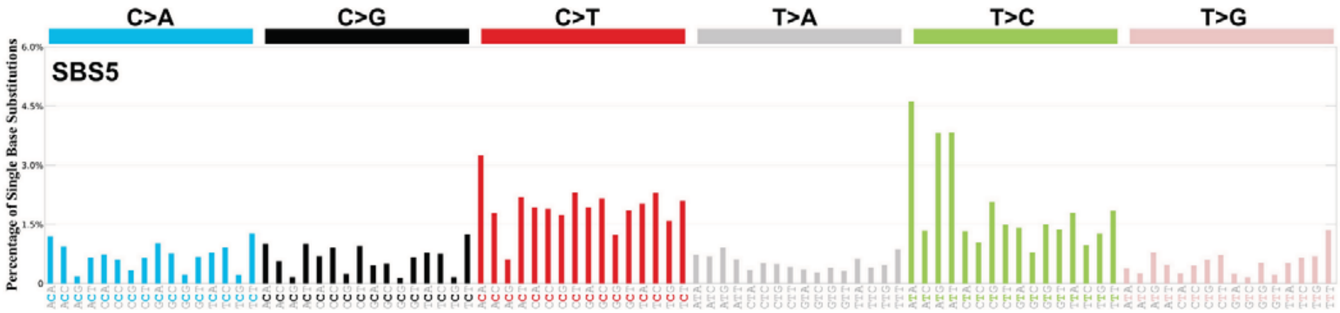
Additional mutational features:

Signature 1 is associated with small numbers of small insertions and deletions in most tissue types.

Comments:

The number of Signature 1 mutations correlates with age of cancer diagnosis.

(1) SBS Signature 5: unknown mechanism



Cancer types:

Signature 5 has been found in all cancer types and most cancer samples.

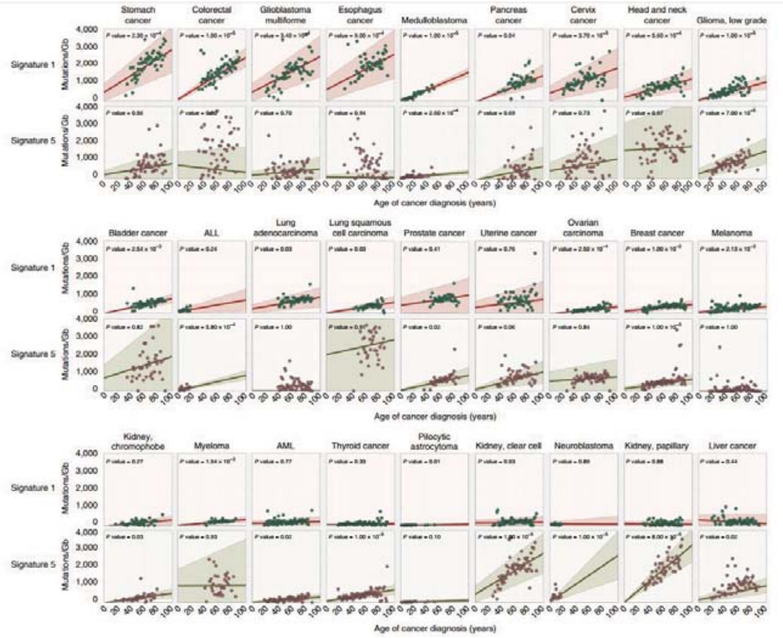
Proposed etiology:

The aetiology of Signature 5 is unknown.

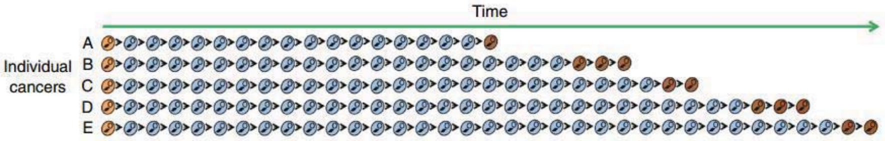
Additional mutational features:

Signature 5 exhibits transcriptional strand bias for T>C substitutions at ApTpN context.

(1) SBS Signatures 1, 5; clock-like property

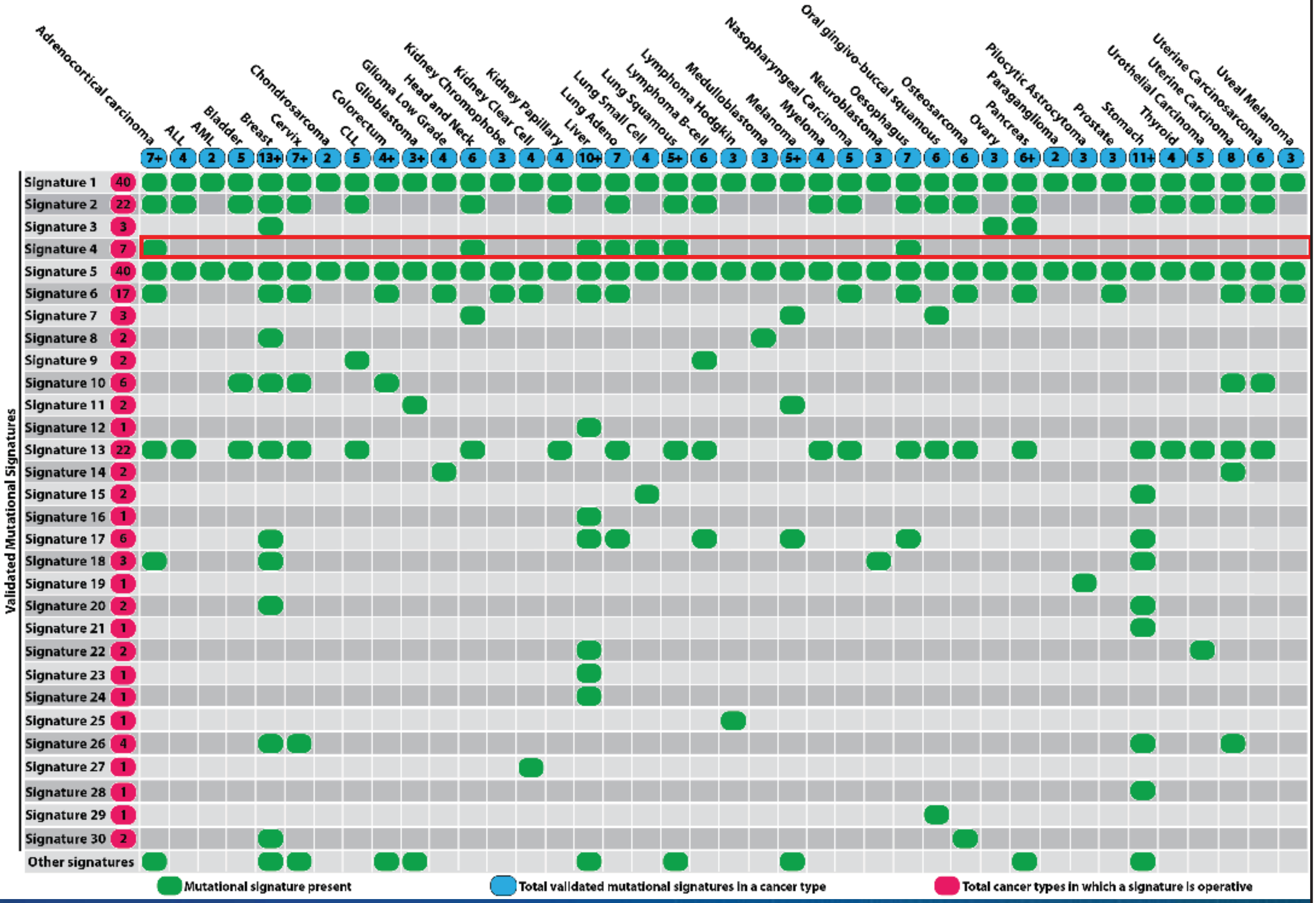


Sig 1, 5 mutations accumulate over time with similar rate



Alexandrov L *et al.*, *Nature Genet* (2015)

(2) Signature 4: due to direct smoke exposure

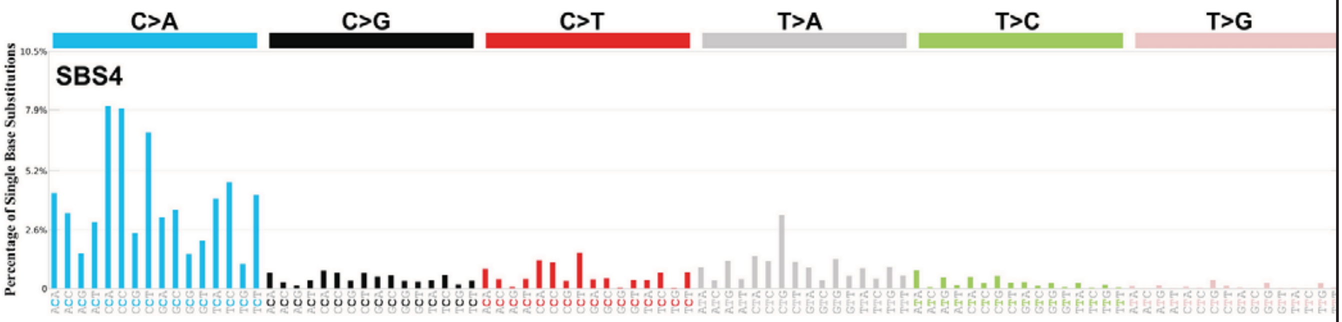


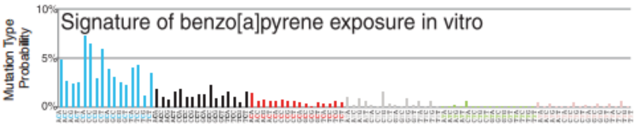
Validated Mutational Signatures		Adrenocortical carcinoma	ALL	AML	Bladder	Breast	Cervix	Chondrosarcoma	CLL	Colorectum	Glioblastoma	Glioma Low Grade	Head and Neck	Kidney Chromophobe	Kidney Clear Cell	Kidney Papillary	Liver	Lung Adeno	Lung Small Cell	Lung Squamous	Lymphoma B-cell
		7+	4	2	5	13+	7+	2	5	4+	3+	4	6	3	4	4	10+	7	4	5+	6
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩
Signature 3	3					🟩															
Signature 4	7	🟩											🟩				🟩	🟩	🟩	🟩	
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17	🟩				🟩	🟩			🟩		🟩		🟩	🟩		🟩	🟩			
Signature 7	3												🟩								
Signature 8	2					🟩															
Signature 9	2								🟩												🟩
Signature 10	6				🟩	🟩	🟩			🟩											
Signature 11	2										🟩										
Signature 12	1																🟩				
Signature 13	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩
Signature 14	2											🟩									
Signature 15	2																		🟩		
Signature 16	1																🟩				
Signature 17	6					🟩											🟩	🟩			🟩
Signature 18	3	🟩				🟩															
Signature 19	1																				
Signature 20	2					🟩															
Signature 21	1																				
Signature 22	2																🟩				
Signature 23	1																🟩				
Signature 24	1																🟩				
Signature 25	1																				
Signature 26	4					🟩	🟩														
Signature 27	1														🟩						
Signature 28	1																				
Signature 29	1																				
Signature 30	2					🟩															
Other signatures		🟩				🟩	🟩			🟩	🟩						🟩			🟩	

Validated Mutational Signatures		Lymphoma Hodgkin	Medulloblastoma	Melanoma	Myeloma	Nasopharyngeal Carcinoma	Neuroblastoma	Oesophagus	Oral gingivo-buccal squamous	Osteosarcoma	Ovary	Pancreas	Paraganglioma	Pilocytic Astrocytoma	Prostate	Stomach	Thyroid	Urothelial Carcinoma	Uterine Carcinoma	Uterine Carcinosarcoma	Uveal Melanoma
		3	3	5+	4	5	3	7	6	6	3	6+	2	3	3	11+	4	5	8	6	3
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 3	3										🟩	🟩									
Signature 4	7							🟩													
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17					🟩		🟩		🟩		🟩			🟩				🟩	🟩	🟩
Signature 7	3			🟩					🟩												
Signature 8	2		🟩																		
Signature 9	2																				
Signature 10	6																		🟩	🟩	
Signature 11	2			🟩																	
Signature 12	1																				
Signature 13	22				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 14	2																		🟩		
Signature 15	2															🟩					
Signature 16	1																				
Signature 17	6			🟩				🟩								🟩					
Signature 18	3						🟩									🟩					
Signature 19	1													🟩							
Signature 20	2															🟩					
Signature 21	1															🟩					
Signature 22	2																	🟩			
Signature 23	1																				
Signature 24	1																				
Signature 25	1	🟩																			
Signature 26	4															🟩			🟩		
Signature 27	1																				
Signature 28	1															🟩					
Signature 29	1								🟩												
Signature 30	2									🟩											
Other signatures				🟩								🟩				🟩					

🟩 Mutational signature present 🟦 Total validated mutational signatures in a cancer type 🟥 Total cancer types in which a signature is operative

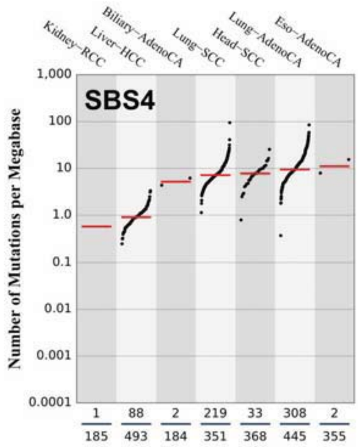
(2) SBS Signature 4: tobacco smoking

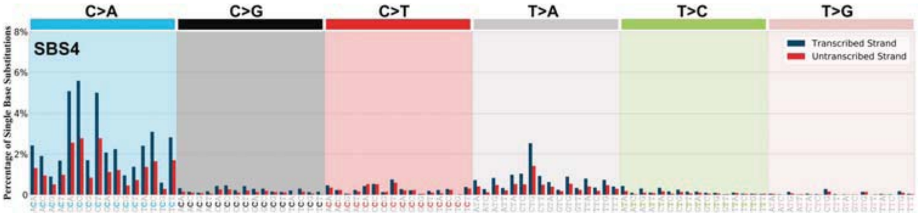


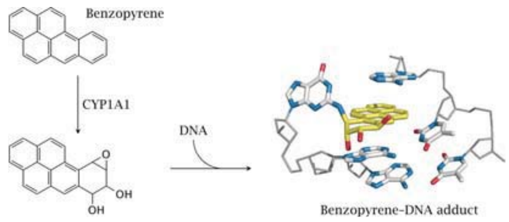


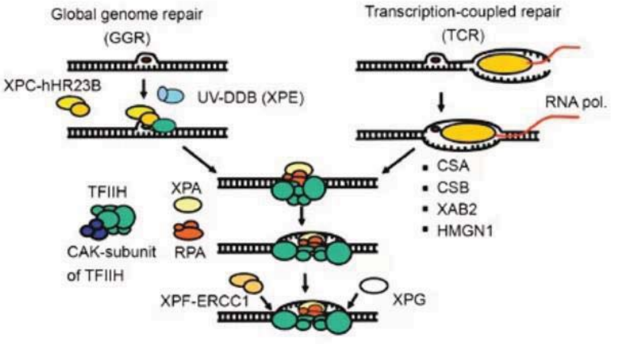
Cancer types:
Signature 4 has been found in head and neck cancer, liver cancer, lung adenocarcinoma, lung squamous carcinoma, small cell lung carcinoma, and esophageal cancer.
Proposed etiology:
Signature 4 is associated with smoking and its profile is similar to the mutational pattern observed in experimental systems exposed to tobacco carcinogens (e.g., benzo[a]pyrene).
Signature 4 is likely due to tobacco mutagens.
Additional mutational features:
Signature 4 exhibits transcriptional strand bias for C>A mutations, compatible with the notion that damage to guanine is repaired by transcription-coupled nucleotide excision repair. Signature 4 is also associated with CC>AA dinucleotide substitutions.

(2) SBS Signature 4: mutational burden and strand bias



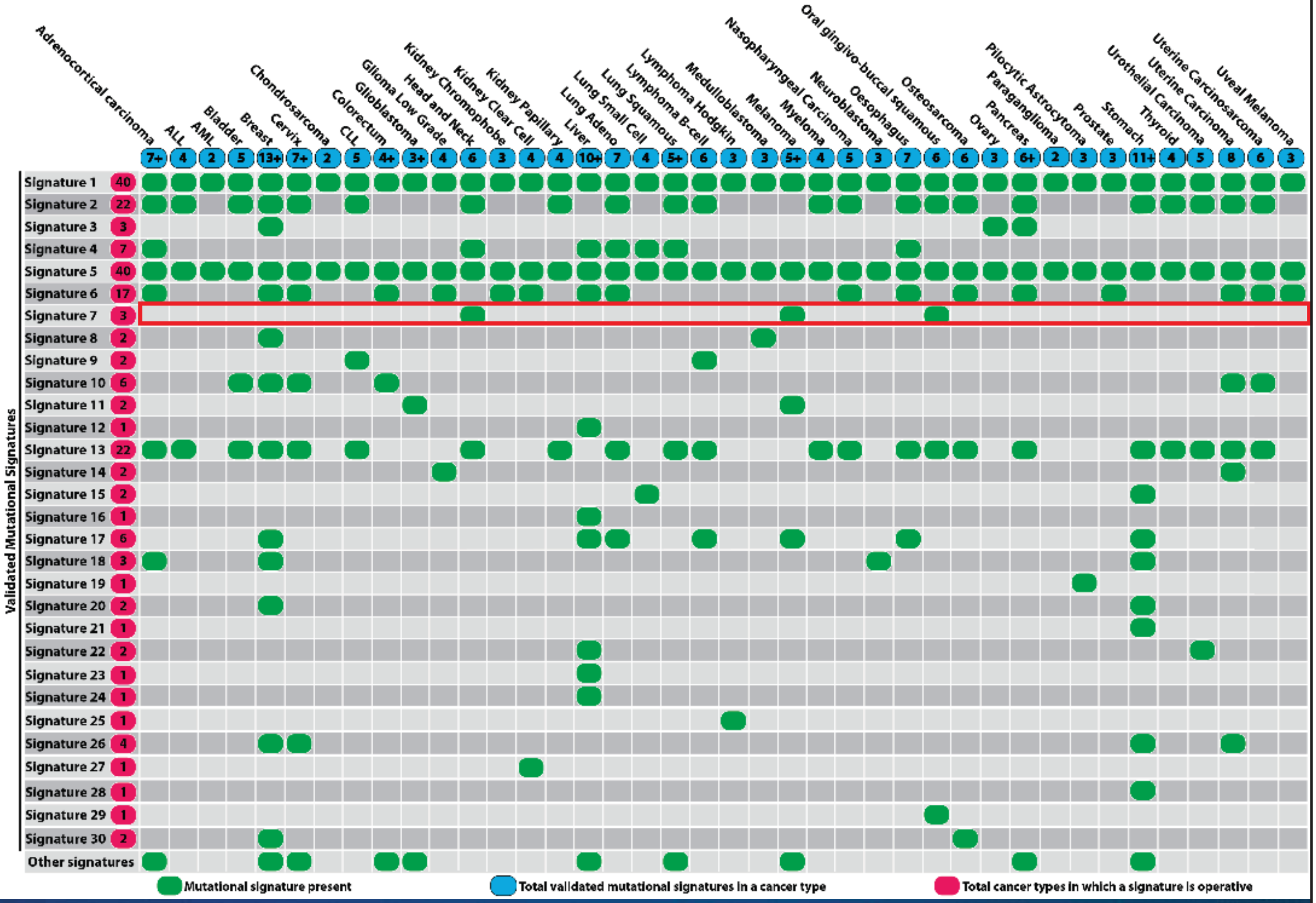






Fousteri and Mullenders (2008)

(3) SBS Signature 7s: due to ultraviolet-light



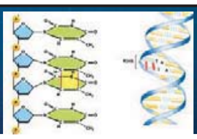
Validated Mutational Signatures		Adrenocortical carcinoma	ALL	AML	Bladder	Breast	Cervix	Chondrosarcoma	CLL	Colorectum	Glioblastoma	Glioma Low Grade	Head and Neck	Kidney Chromophobe	Kidney Clear Cell	Kidney Papillary	Liver	Lung Adeno	Lung Small Cell	Lung Squamous	Lymphoma B-cell	Lymphoma Hodgkin	Medulloblastoma	Melanoma	Myeloma	Nasopharyngeal Carcinoma	Neuroblastoma	Oesophagus	Oral gingivo-buccal squamous	Osteosarcoma	Ovary	Pancreas	Paraganglioma	Pilocytic Astrocytoma	Prostate	Stomach	Thyroid	Urothelial Carcinoma	Uterine Carcinoma	Uterine Carcinosarcoma	Uveal Melanoma
		7+	4	2	5	13+	7+	2	5	4+	3+	4	6	3	4	4	10+	7	4	5+	6	3	3	5+	4	5	3	7	6	6	3	6+	2	3	3	11+	4	5	8	6	3
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 3	3					🟩																									🟩	🟩									
Signature 4	7	🟩											🟩				🟩	🟩	🟩	🟩								🟩													
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17	🟩				🟩	🟩			🟩		🟩		🟩	🟩		🟩	🟩								🟩		🟩		🟩		🟩			🟩				🟩	🟩	🟩
Signature 7	3												🟩											🟩					🟩												
Signature 8	2					🟩																	🟩																		
Signature 9	2								🟩												🟩																				
Signature 10	6				🟩	🟩	🟩			🟩																													🟩	🟩	
Signature 11	2										🟩													🟩																	
Signature 12	1																🟩																								
Signature 13	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 14	2											🟩																											🟩		
Signature 15	2																		🟩																	🟩					
Signature 16	1																🟩																								
Signature 17	6					🟩											🟩	🟩			🟩			🟩				🟩								🟩					
Signature 18	3	🟩				🟩																					🟩									🟩					
Signature 19	1																																	🟩							
Signature 20	2					🟩																														🟩					
Signature 21	1																																			🟩					
Signature 22	2																🟩																					🟩			
Signature 23	1																🟩																								
Signature 24	1																🟩																								
Signature 25	1																					🟩																			
Signature 26	4					🟩	🟩																													🟩			🟩		
Signature 27	1														🟩																										
Signature 28	1																																			🟩					
Signature 29	1																												🟩												
Signature 30	2					🟩																								🟩											
Other signatures		🟩				🟩	🟩			🟩	🟩						🟩			🟩				🟩								🟩				🟩					

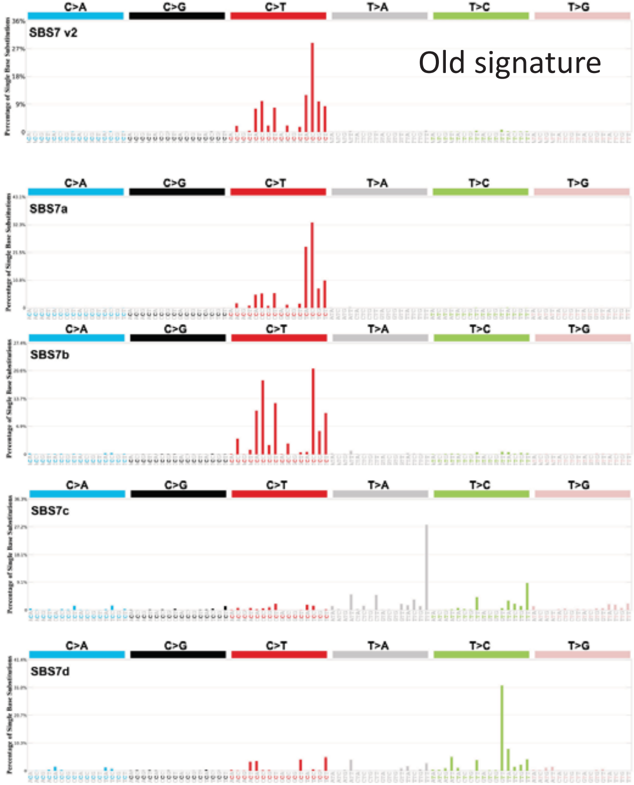
🟩 Mutational signature present

🔵 Total validated mutational signatures in a cancer type

🟥 Total cancer types in which a signature is operative

(3) Signature 7: ultraviolet-light damage





Old signature

Cancer types:

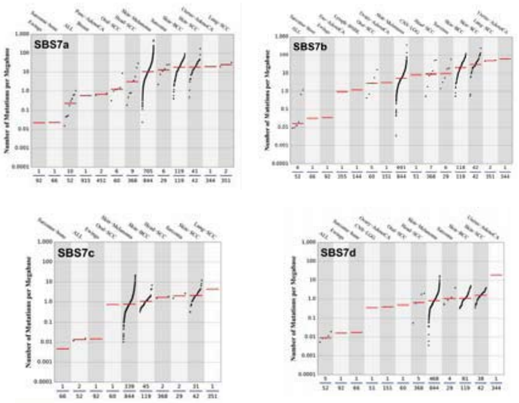
Signature 7 has been found predominantly in skin cancers and in cancers of the lip categorized as head and neck or oral squamous cancers.

Proposed etiology:

Based on its prevalence in ultraviolet exposed areas and the similarity of the mutational pattern to tha observed in experimental systems exposed to ultraviolet light Signature 7 is likely due to ultraviolet light exposure.

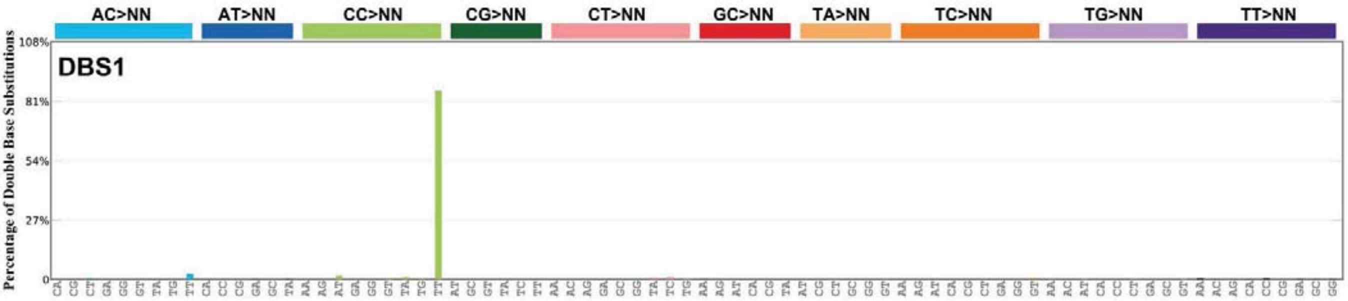
Additional mutational features:

Signature 7 is associated with large numbers of CC>TT dinucleotide mutations at dipyrimidines. Additionally, Signature 7 exhibits a strong transcriptional strand-bias indicating that mutations occur a pyrimidines (viz., by formation of pyrimidine-pyrimidine photodimers) and these mutations are being repaired by transcription-coupled nucleotide excision repair.

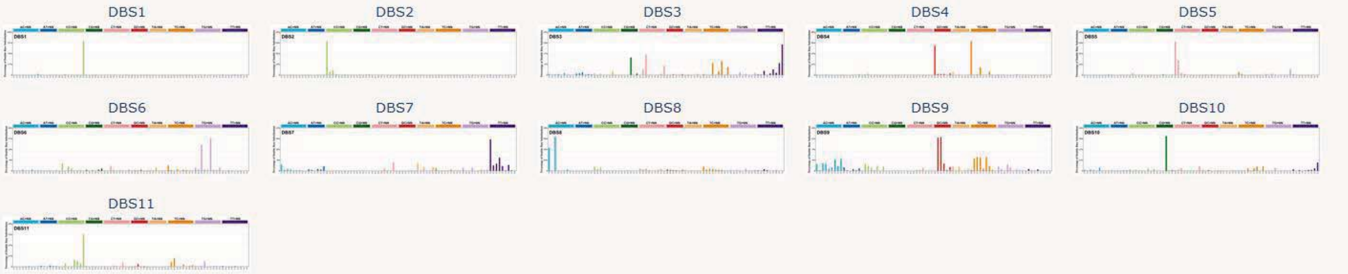


SBi 한국생명정보학회 Korean Society for Bioinformatics

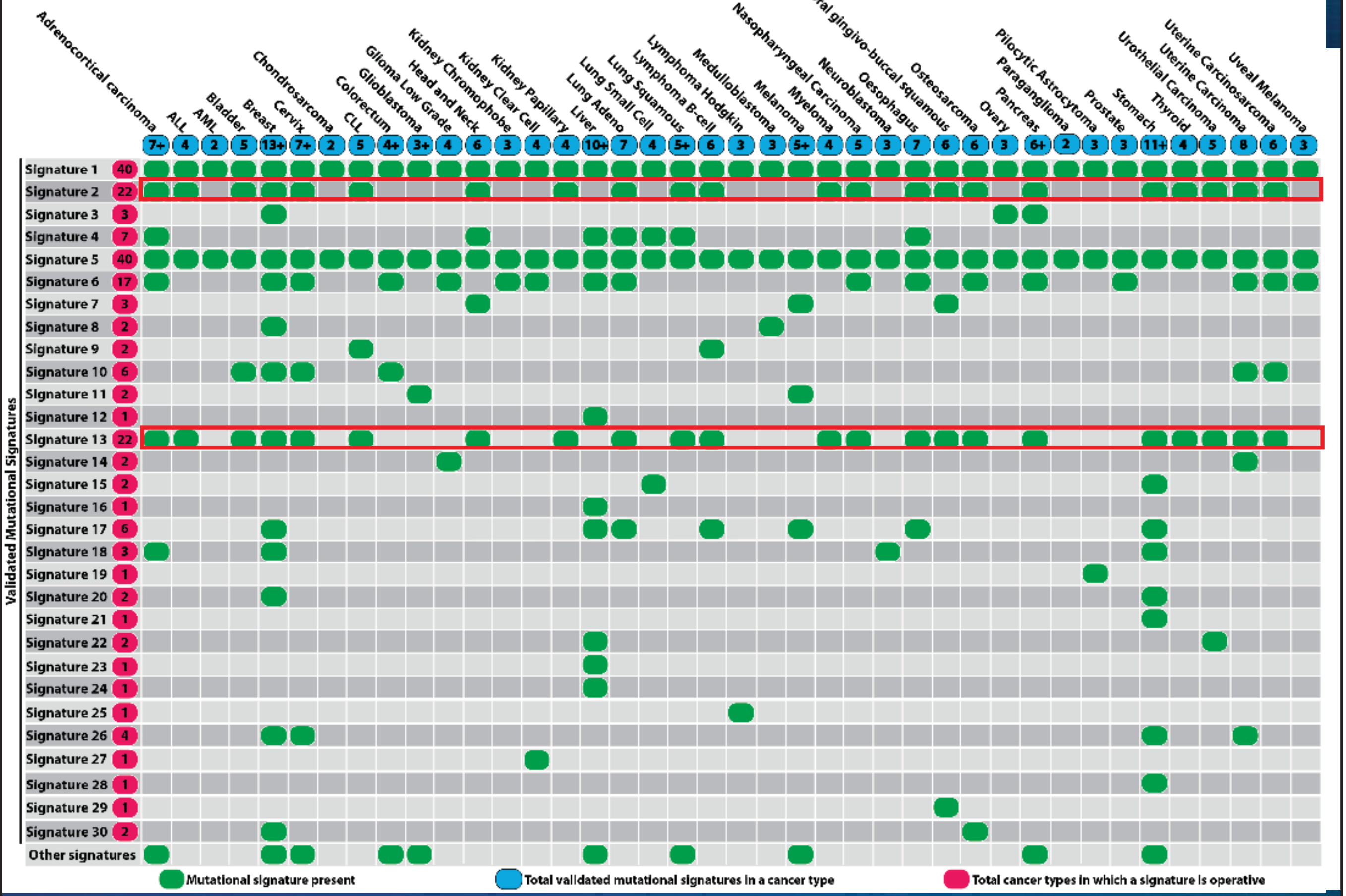
(3) Double base substitution (DBS1)



78 mutation classes



(4) SBS Signatures 2 and 13: APOBEC-mediated mutagenesis



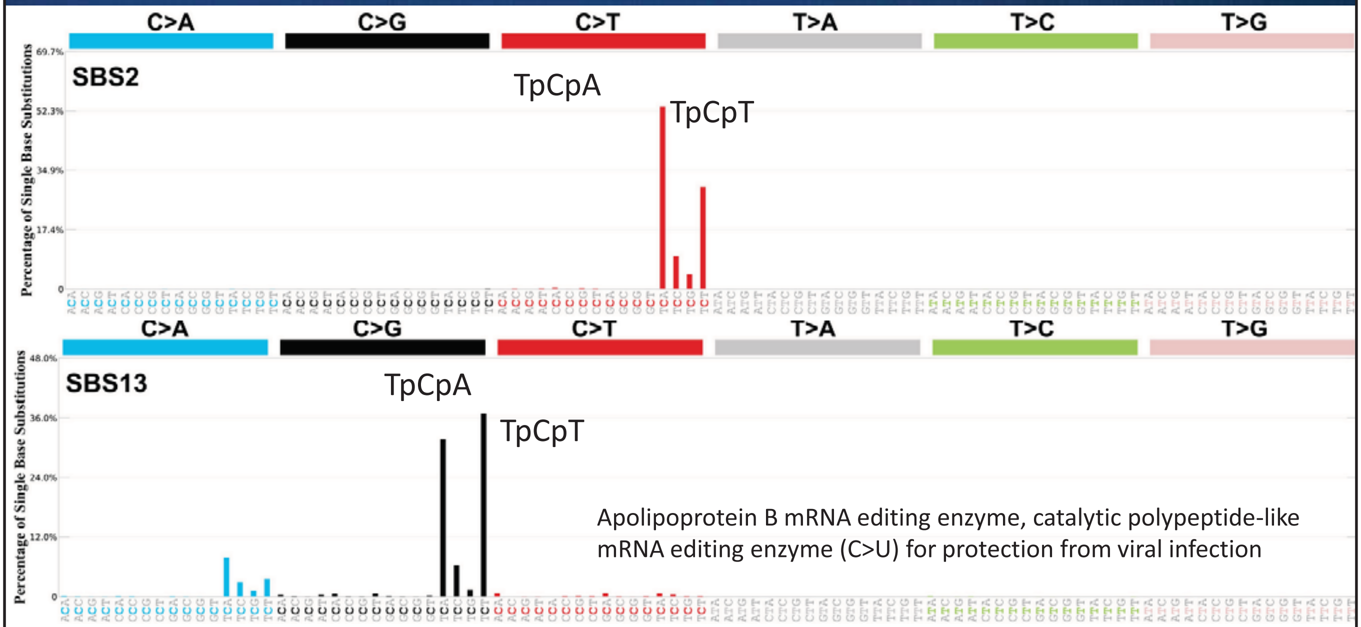
Validated Mutational Signatures		Adrenocortical carcinoma	ALL	AML	Bladder	Breast	Cervix	Chondrosarcoma	CLL	Colorectum	Glioblastoma	Glioma Low Grade	Head and Neck	Kidney Chromophobe	Kidney Clear Cell	Kidney Papillary	Liver	Lung Adeno	Lung Small Cell	Lung Squamous	Lymphoma B-cell	Lymphoma Hodgkin	Medulloblastoma	Melanoma	Myeloma	Nasopharyngeal Carcinoma	Neuroblastoma	Oesophagus	Oral gingivo-buccal squamous	Osteosarcoma	Ovary	Pancreas	Paraganglioma	Pilocytic Astrocytoma	Prostate	Stomach	Thyroid	Urothelial Carcinoma	Uterine Carcinoma	Uterine Carcinosarcoma	Uveal Melanoma
		7+	4	2	5	13+	7+	2	5	4+	3+	4	6	3	4	4	10+	7	4	5+	6	3	3	5+	4	5	3	7	6	6	3	6+	2	3	3	11+	4	5	8	6	3
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 3	3					🟩																									🟩	🟩									
Signature 4	7	🟩											🟩				🟩	🟩	🟩	🟩								🟩													
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17	🟩				🟩	🟩			🟩		🟩		🟩	🟩		🟩	🟩								🟩		🟩		🟩		🟩			🟩				🟩	🟩	🟩
Signature 7	3												🟩											🟩					🟩												
Signature 8	2					🟩																	🟩																		
Signature 9	2								🟩												🟩																				
Signature 10	6				🟩	🟩	🟩			🟩																													🟩	🟩	
Signature 11	2										🟩													🟩																	
Signature 12	1																🟩																								
Signature 13	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 14	2											🟩																											🟩		
Signature 15	2																		🟩																	🟩					
Signature 16	1																🟩																								
Signature 17	6					🟩											🟩	🟩			🟩			🟩				🟩								🟩					
Signature 18	3	🟩				🟩																					🟩									🟩					
Signature 19	1																																	🟩							
Signature 20	2					🟩																														🟩					
Signature 21	1																																			🟩					
Signature 22	2																🟩																					🟩			
Signature 23	1																🟩																								
Signature 24	1																🟩																								
Signature 25	1																					🟩																			
Signature 26	4					🟩	🟩																													🟩			🟩		
Signature 27	1														🟩																										
Signature 28	1																																			🟩					
Signature 29	1																												🟩												
Signature 30	2					🟩																								🟩											
Other signatures		🟩				🟩	🟩			🟩	🟩						🟩			🟩				🟩								🟩				🟩					

🟩 Mutational signature present

● Total validated mutational signatures in a cancer type

● Total cancer types in which a signature is operative

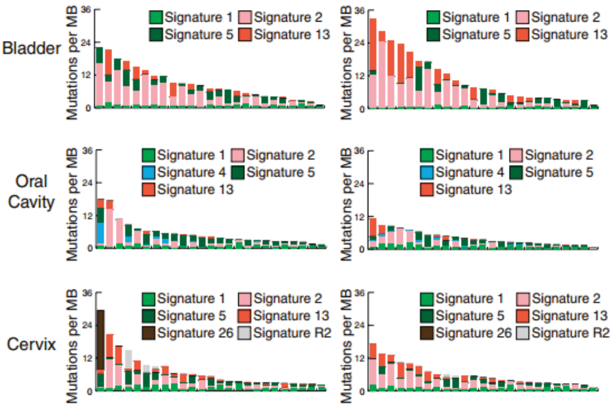
(4) SBS Signatures 2 and 13: APOBEC-mediated mutagenesis



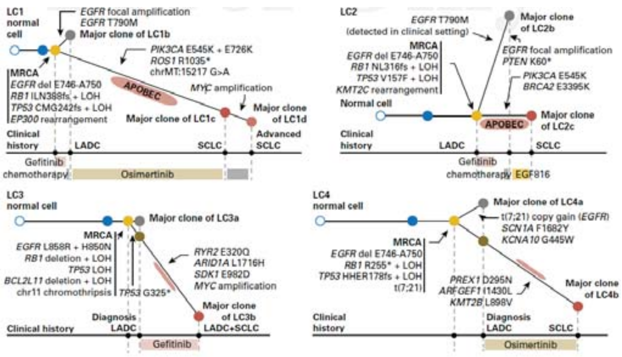
Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
mRNA editing enzyme (C>U) for protection from viral infection

Cancer types:
Signature 2 has been found in 22 cancer types. Dominant processes in cervical and bladder cancers.
Proposed etiology:
Signature 2 has been attributed to activity of the **AID/APOBEC family of cytidine deaminases**.
On the basis of similarities in the sequence context of cytosine mutations caused by APOBEC enzymes in experimental systems, a role for APOBEC1, APOBEC3A and/or APOBEC3B in human cancer appears more likely than for other members of the family.

APOBEC-mediated mutations



Alexandrov L *et al.*, *Science* (2016)



Lee J *et al.*, *J Clin Oncol* (2017)

Activated in many cancer types
including cervical, bladder, breast and lung cancers.

Activated in the late branch in lung cancers.
(Episodically activating?)

SBi 한국생명정보학회 Korean Society for Bioinformatics

(5) Signature 11: temozolomide-driven

Validated Mutational Signatures

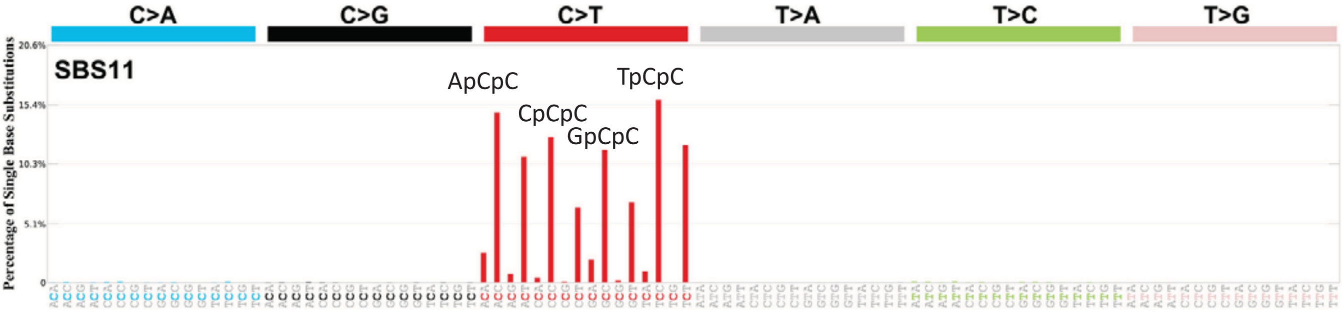
		Adrenocortical carcinoma	ALL	AML	Bladder	Breast	Cervix	Chondrosarcoma	CLL	Colorectum	Glioblastoma	Glioma Low Grade	Head and Neck	Kidney Chromophobe	Kidney Clear Cell	Kidney Papillary	Liver	Lung Adeno	Lung Small Cell	Lung Squamous	Lymphoma B-cell	Lymphoma Hodgkin	Medulloblastoma	Melanoma	Myeloma	Nasopharyngeal Carcinoma	Neuroblastoma	Oesophagus	Oral gingivo-buccal squamous	Osteosarcoma	Ovary	Pancreas	Paraganglioma	Pilocytic Astrocytoma	Prostate	Stomach	Thyroid	Urothelial Carcinoma	Uterine Carcinoma	Uterine Carcinosarcoma	Uveal Melanoma
		7+	4	2	5	13+	7+	2	5	4+	3+	4	6	3	4	4	10+	7	4	5+	6	3	3	5+	4	5	3	7	6	6	3	6+	2	3	3	11+	4	5	8	6	3
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 3	3					🟩																									🟩	🟩									
Signature 4	7	🟩											🟩				🟩	🟩	🟩	🟩								🟩													
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17	🟩				🟩	🟩			🟩		🟩		🟩	🟩		🟩	🟩								🟩		🟩		🟩		🟩			🟩				🟩	🟩	🟩
Signature 7	3												🟩											🟩					🟩												
Signature 8	2					🟩																	🟩																		
Signature 9	2								🟩												🟩																				
Signature 10	6				🟩	🟩	🟩			🟩																													🟩	🟩	
Signature 11	2										🟩													🟩																	
Signature 12	1																🟩																								
Signature 13	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 14	2											🟩																											🟩		
Signature 15	2																		🟩																	🟩					
Signature 16	1																🟩																								
Signature 17	6					🟩											🟩	🟩			🟩			🟩				🟩								🟩					
Signature 18	3	🟩				🟩																					🟩									🟩					
Signature 19	1																																	🟩							
Signature 20	2					🟩																														🟩					
Signature 21	1																																			🟩					
Signature 22	2																🟩																					🟩			
Signature 23	1																🟩																								
Signature 24	1																🟩																								
Signature 25	1																					🟩																			
Signature 26	4					🟩	🟩																													🟩			🟩		
Signature 27	1														🟩																										
Signature 28	1																																			🟩					
Signature 29	1																												🟩												
Signature 30	2					🟩																								🟩											
Other signatures		🟩				🟩	🟩			🟩	🟩						🟩			🟩				🟩								🟩				🟩					

🟩 Mutational signature present

Total validated mutational signatures in a cancer type

Total cancer types in which a signature is operative

(5) SBS Signature 11: alkylating agent



Cancer types:

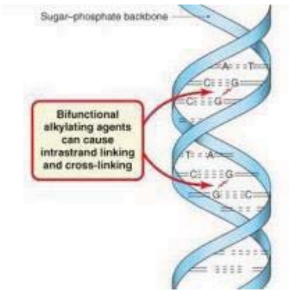
Signature 11 has been found in melanoma and glioblastoma.

Proposed etiology:

Signature 11 exhibits a mutational pattern resembling that of alkylating agents. Patient histories have revealed an association between treatments with the **alkylating agent temozolomide** and Signature 11 mutations.

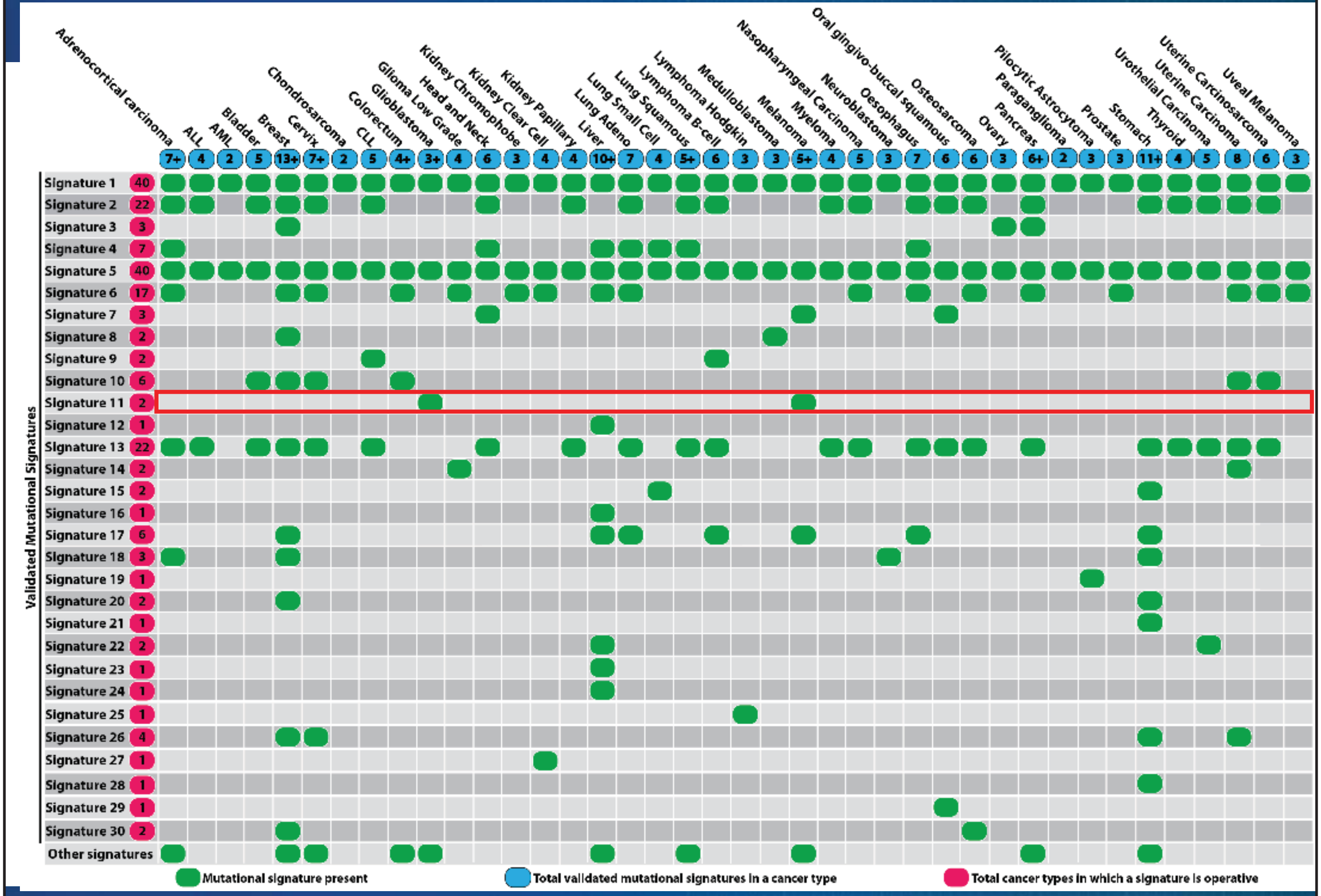
Additional mutational features:

Signature 11 exhibits a strong transcriptional strand-bias for C>T substitutions indicating that mutations occur on guanine and that these mutations are effectively repaired by transcription-coupled nucleotide excision repair.


Sugar-phosphate backbone
Bifunctional alkylating agents can cause intrastrand linking and cross-linking
A T
C G
C G
T A
C G
G C

SBi 한국생명정보학회 Korean Society for Bioinformatics

(6) SBS Signature 22: aristolochic acid driven



Validated Mutational Signatures

		Adrenocortical carcinoma	ALL	AML	Bladder	Breast	Cervix	Chondrosarcoma	CLL	Colorectum	Glioblastoma	Glioma Low Grade	Head and Neck	Kidney Chromophobe	Kidney Clear Cell	Kidney Papillary	Liver	Lung Adeno	Lung Small Cell	Lung Squamous	Lymphoma B-cell
		7+	4	2	5	13+	7+	2	5	4+	3+	4	6	3	4	4	10+	7	4	5+	6
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩
Signature 3	3					🟩															
Signature 4	7	🟩											🟩				🟩	🟩	🟩	🟩	
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17	🟩				🟩	🟩			🟩		🟩		🟩	🟩		🟩	🟩			
Signature 7	3												🟩								
Signature 8	2					🟩															
Signature 9	2								🟩												🟩
Signature 10	6				🟩	🟩	🟩			🟩											
Signature 11	2										🟩										
Signature 12	1																🟩				
Signature 13	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩
Signature 14	2											🟩									
Signature 15	2																		🟩		
Signature 16	1																🟩				
Signature 17	6					🟩											🟩	🟩			🟩
Signature 18	3	🟩				🟩															
Signature 19	1																				
Signature 20	2					🟩															
Signature 21	1																				
Signature 22	2																🟩				
Signature 23	1																🟩				
Signature 24	1																🟩				
Signature 25	1																				
Signature 26	4					🟩	🟩														
Signature 27	1														🟩						
Signature 28	1																				
Signature 29	1																				
Signature 30	2					🟩															
Other signatures		🟩				🟩	🟩			🟩	🟩						🟩			🟩	

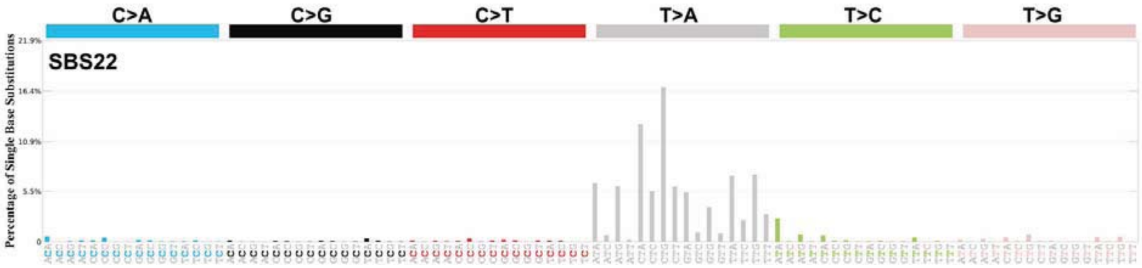
		Lymphoma Hodgkin	Medulloblastoma	Melanoma	Myeloma	Nasopharyngeal Carcinoma	Neuroblastoma	Oesophagus	Oral gingivo-buccal squamous	Osteosarcoma	Ovary	Pancreas	Paraganglioma	Pilocytic Astrocytoma	Prostate	Stomach	Thyroid	Urothelial Carcinoma	Uterine Carcinoma	Uterine Carcinosarcoma	Uveal Melanoma
		3	3	5+	4	5	3	7	6	6	3	6+	2	3	3	11+	4	5	8	6	3
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 3	3										🟩	🟩									
Signature 4	7							🟩													
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17					🟩		🟩		🟩		🟩			🟩				🟩	🟩	🟩
Signature 7	3			🟩					🟩												
Signature 8	2		🟩																		
Signature 9	2																				
Signature 10	6																		🟩	🟩	
Signature 11	2			🟩																	
Signature 12	1																				
Signature 13	22				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 14	2																		🟩		
Signature 15	2															🟩					
Signature 16	1																				
Signature 17	6			🟩				🟩								🟩					
Signature 18	3						🟩									🟩					
Signature 19	1													🟩							
Signature 20	2															🟩					
Signature 21	1															🟩					
Signature 22	2																	🟩			
Signature 23	1																				
Signature 24	1																				
Signature 25	1	🟩																			
Signature 26	4															🟩			🟩		
Signature 27	1																				
Signature 28	1															🟩					
Signature 29	1								🟩												
Signature 30	2									🟩											
Other signatures				🟩								🟩				🟩					

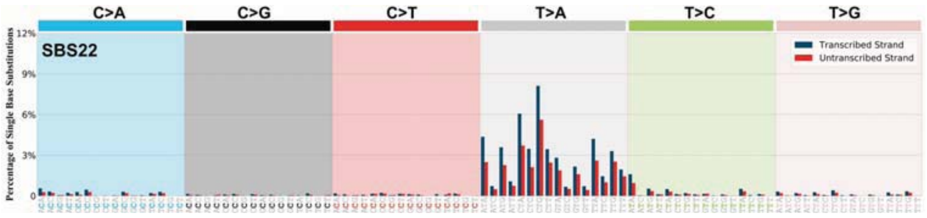
🟩 Mutational signature present

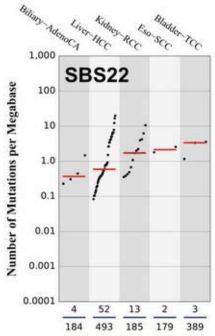
Total validated mutational signatures in a cancer type

Total cancer types in which a signature is operative

(6) SBS Signature 22: aristolochic acids







Cancer types:

Signature 22 has been found in urothelial (renal pelvis) carcinoma and liver cancers.

Proposed aetiology:

Signature 22 has been found in cancer samples with known exposures to aristolochic acid. Additionally, the pattern of mutations exhibited by the signature is consistent with the one previous observed in experimental systems exposed to aristolochic acid.

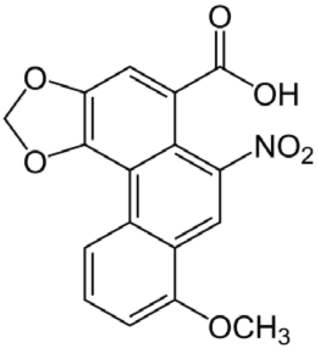
Additional mutational features:

Signature 22 exhibits a very strong transcriptional strand bias for T>A mutations indicating adenine damage that is being repaired by transcription-coupled nucleotide excision repair.

Comments:

Signature 22 has a very high mutational burden in urothelial carcinoma; however, its mutational burden is much lower in liver cancers.

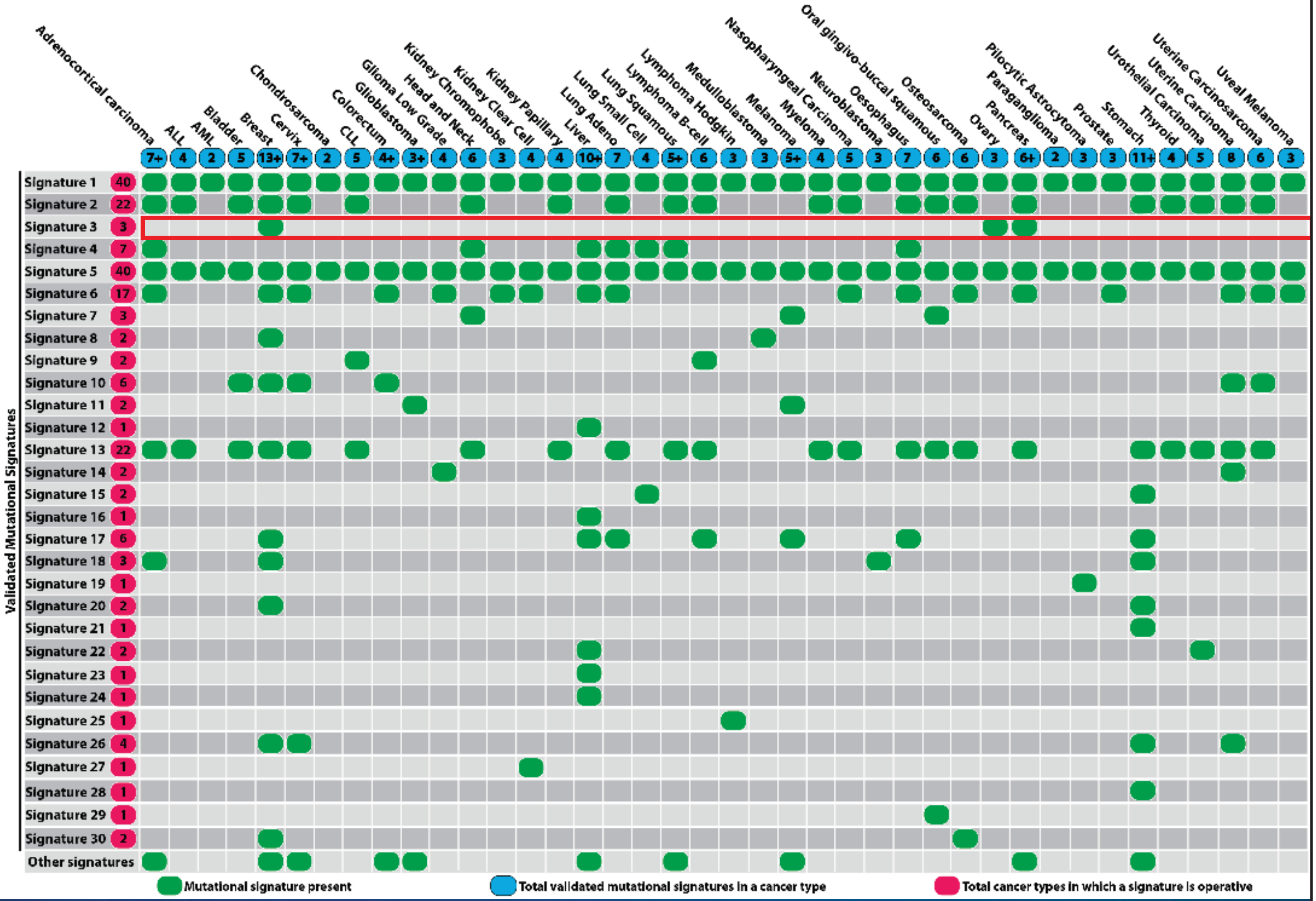
(6) SBS signature 22: aristolochic acids



Aristolochia clematitis
(쥐방울덩굴, 동북마두령, 관목통)

https://www.accessdata.fda.gov/cms_ia/importalert_141.html

(7) SBS Signature 3: HR-based DNA repair



Validated Mutational Signatures

		Adrenocortical carcinoma	ALL	AML	Bladder	Breast	Cervix	Chondrosarcoma	CLL	Colorectum	Glioblastoma	Glioma Low Grade	Head and Neck	Kidney Chromophobe	Kidney Clear Cell	Kidney Papillary	Liver	Lung Adeno	Lung Small Cell	Lung Squamous	Lymphoma B-cell
		7+	4	2	5	13+	7+	2	5	4+	3+	4	6	3	4	4	10+	7	4	5+	6
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩
Signature 3	3					🟩															
Signature 4	7	🟩											🟩				🟩	🟩	🟩	🟩	
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17	🟩				🟩	🟩			🟩		🟩		🟩	🟩		🟩	🟩			
Signature 7	3												🟩								
Signature 8	2					🟩															
Signature 9	2								🟩												🟩
Signature 10	6				🟩	🟩	🟩			🟩											
Signature 11	2										🟩										
Signature 12	1																🟩				
Signature 13	22	🟩	🟩		🟩	🟩	🟩		🟩				🟩			🟩		🟩		🟩	🟩
Signature 14	2											🟩									
Signature 15	2																		🟩		
Signature 16	1																🟩				
Signature 17	6					🟩											🟩	🟩			🟩
Signature 18	3	🟩				🟩															
Signature 19	1																				
Signature 20	2					🟩															
Signature 21	1																				
Signature 22	2																🟩				
Signature 23	1																🟩				
Signature 24	1																🟩				
Signature 25	1																				
Signature 26	4					🟩	🟩														
Signature 27	1															🟩					
Signature 28	1																				
Signature 29	1																				
Signature 30	2					🟩															
Other signatures		🟩				🟩	🟩			🟩	🟩						🟩			🟩	

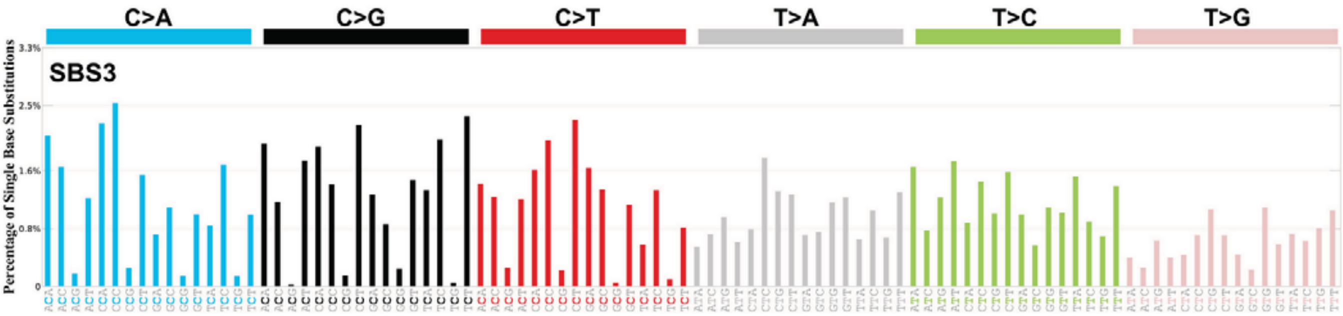
		Lymphoma Hodgkin	Medulloblastoma	Melanoma	Myeloma	Nasopharyngeal Carcinoma	Neuroblastoma	Oesophagus	Oral gingivo-buccal squamous	Osteosarcoma	Ovary	Pancreas	Paraganglioma	Pilocytic Astrocytoma	Prostate	Stomach	Thyroid	Urothelial Carcinoma	Uterine Carcinoma	Uterine Carcinosarcoma	Uveal Melanoma
		3	3	5+	4	5	3	7	6	6	3	6+	2	3	3	11+	4	5	8	6	3
Signature 1	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 2	22				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 3	3										🟩	🟩									
Signature 4	7							🟩													
Signature 5	40	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Signature 6	17					🟩		🟩		🟩		🟩			🟩				🟩	🟩	🟩
Signature 7	3			🟩					🟩												
Signature 8	2		🟩																		
Signature 9	2																				
Signature 10	6																		🟩	🟩	
Signature 11	2			🟩																	
Signature 12	1																				
Signature 13	22				🟩	🟩		🟩	🟩	🟩		🟩				🟩	🟩	🟩	🟩	🟩	
Signature 14	2																		🟩		
Signature 15	2															🟩					
Signature 16	1																				
Signature 17	6			🟩				🟩								🟩					
Signature 18	3						🟩									🟩					
Signature 19	1													🟩							
Signature 20	2															🟩					
Signature 21	1															🟩					
Signature 22	2																	🟩			
Signature 23	1																				
Signature 24	1																				
Signature 25	1	🟩																			
Signature 26	4															🟩			🟩		
Signature 27	1																				
Signature 28	1															🟩					
Signature 29	1								🟩												
Signature 30	2									🟩											
Other signatures				🟩								🟩				🟩					

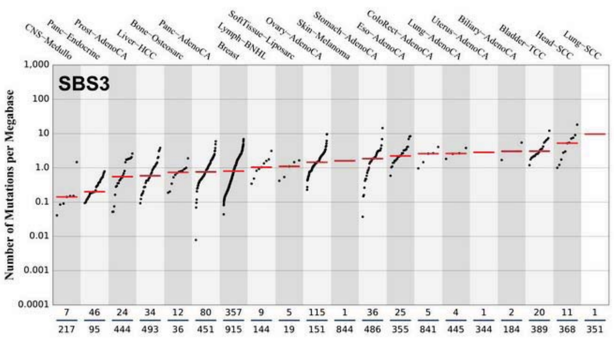
🟩 Mutational signature present

Total validated mutational signatures in a cancer type

Total cancer types in which a signature is operative

(7) SBS Signature 3: HR-based DNA repair





Proposed aetiology

Defective homologous recombination-based DNA damage repair which manifests predominantly as small indels and genome rearrangements due to abnormal double strand break repair but also in the form of this base substitution signature.

Comments

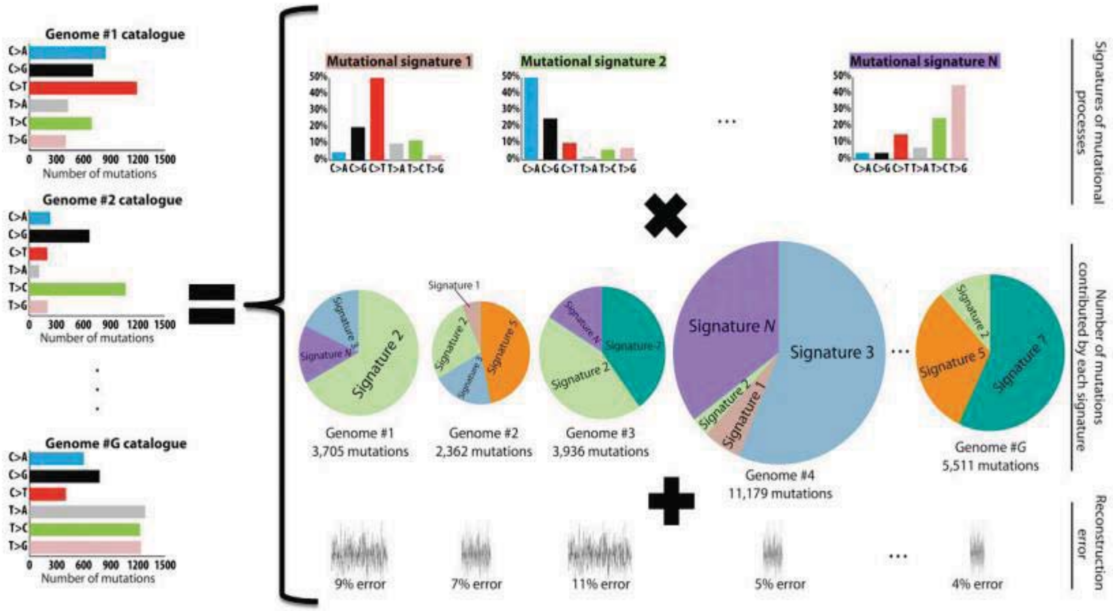
SBS3 is strongly associated with germline and somatic BRCA1 and **BRCA2 mutations and BRCA1** promoter methylation in breast, pancreatic, and ovarian cancers. In pancreatic cancer, responders to platinum therapy usually exhibit SBS3 mutations. Together with associated indel and rearrangement signatures, SBS3 has been proposed as a predictor of defective homologous recombination-based repair and thus of response to therapies exploiting this repair defect.

SBi 한국생명정보학회 Korean Society for Bioinformatics

어떻게 mutational signature를 구할 것인가?

SBi 한국생명정보학회 Korean Society for Bioinformatics

몇 개의 sample, 몇 개의 돌연변이가 필요할까?



Alexandrov L *et al.*, ***Cell reports*** (2013)

(1) For inferring *de novo* mutational signature: many whole-genome sequences
(2) For fitting known signatures: whole-genome, (exome?)

Tools for extracting mutational signatures

Inferring *de novo* signatures

Alexandrov, MatLab (***Nature*** 2013)
EMu (***Genome Biology*** 2013)
Maftools (***Genome Res*** 2018)
MutationalPatterns (***Genome Med*** 2018)
MutSpec (***BMC Bioinformatics*** 2016)
SigFit (*BioRxiv* 2020)
SigMiner (*medRxiv* 2020)
SignatureAnalyzer (***Nature Commun*** 2015)
SignatureToolsLib (***Nat Cancer*** 2020)
SigneR (***Bioinformatics*** 2017)
SomaticSignatures (***Bioinformatics*** 2015)
SigProfiler (COSMIC)

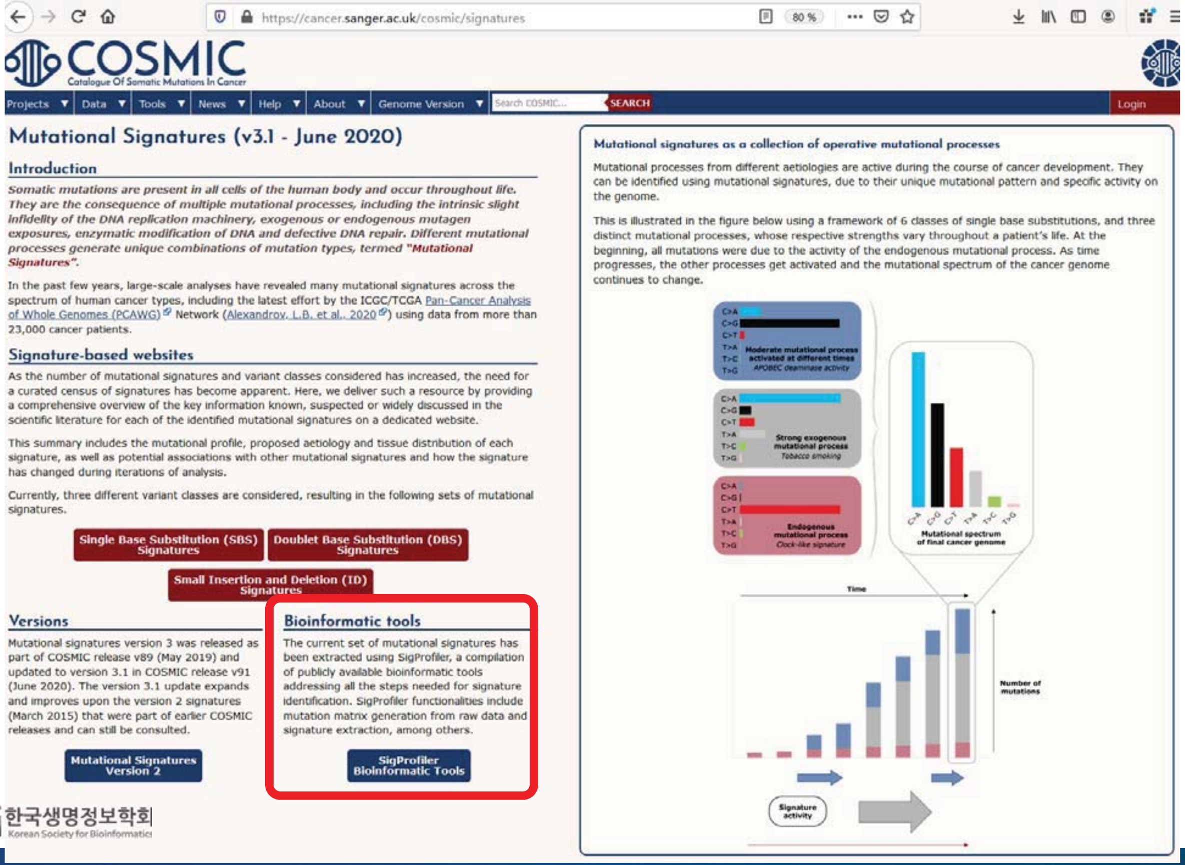
Fitting known signatures

deconsructSigs (***Genome Biology*** 2016)
SignatureEstimation (***Bioinformatics*** 2018)
YAPSA (R Package v 1.16.0)

Web interfaces

MutaGene (***NAR*** 2017)
mSignatureDB (***NAR*** 2018)
MuSiCa (***BMC Bioinformatics*** 2018)
Mutalisk (***NAR*** 2018)

(1) Sigprofilers


https://cancer.sanger.ac.uk/cosmic/signatures
80 %
COSMIC
Catalogue Of Somatic Mutations In Cancer
Projects ▼ Data ▼ Tools ▼ News ▼ Help ▼ About ▼ Genome Version ▼
Search COSMIC...
SEARCH
Login

Mutational Signatures (v3.1 - June 2020)

Introduction

*Somatic mutations are present in all cells of the human body and occur throughout life. They are the consequence of multiple mutational processes, including the intrinsic slight infidelity of the DNA replication machinery, exogenous or endogenous mutagen exposures, enzymatic modification of DNA and defective DNA repair. Different mutational processes generate unique combinations of mutation types, termed **"Mutational Signatures"**.*

In the past few years, large-scale analyses have revealed many mutational signatures across the spectrum of human cancer types, including the latest effort by the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes (PCAWG) Network (Alexandrov, L.B. et al., 2020) using data from more than 23,000 cancer patients.

Signature-based websites

As the number of mutational signatures and variant classes considered has increased, the need for a curated census of signatures has become apparent. Here, we deliver such a resource by providing a comprehensive overview of the key information known, suspected or widely discussed in the scientific literature for each of the identified mutational signatures on a dedicated website.

This summary includes the mutational profile, proposed aetiology and tissue distribution of each signature, as well as potential associations with other mutational signatures and how the signature has changed during iterations of analysis.

Currently, three different variant classes are considered, resulting in the following sets of mutational signatures.

Single Base Substitution (SBS) Signatures
Doublet Base Substitution (DBS) Signatures
Small Insertion and Deletion (ID) Signatures

Versions

Mutational signatures version 3 was released as part of COSMIC release v89 (May 2019) and updated to version 3.1 in COSMIC release v91 (June 2020). The version 3.1 update expands and improves upon the version 2 signatures (March 2015) that were part of earlier COSMIC releases and can still be consulted.

Bioinformatic tools

The current set of mutational signatures has been extracted using SigProfiler, a compilation of publicly available bioinformatic tools addressing all the steps needed for signature identification. SigProfiler functionalities include mutation matrix generation from raw data and signature extraction, among others.

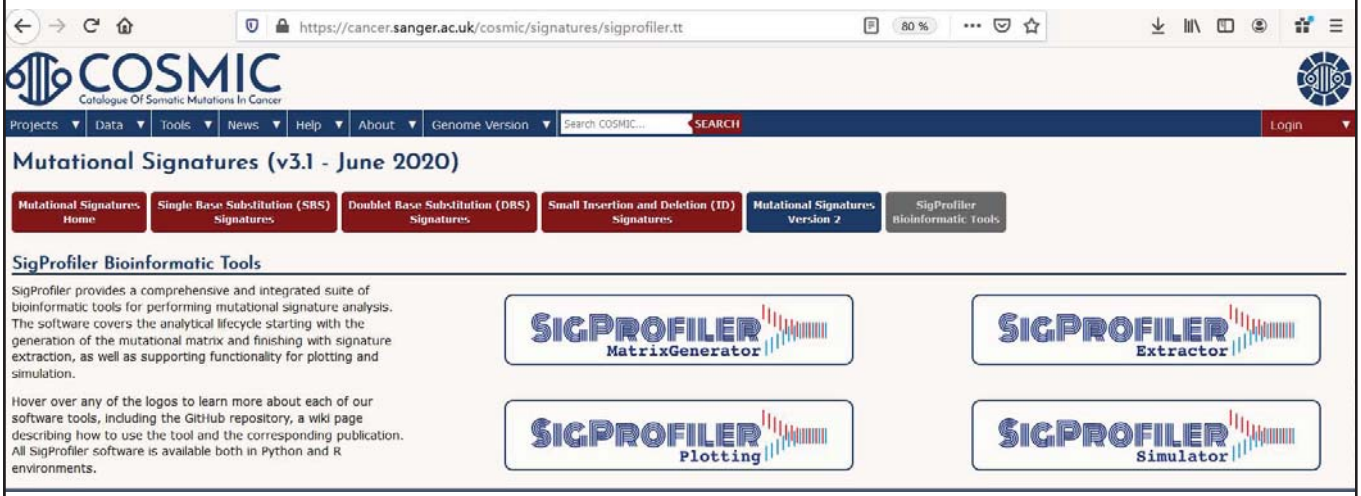
Mutational signatures as a collection of operative mutational processes

Mutational processes from different aetiologies are active during the course of cancer development. They can be identified using mutational signatures, due to their unique mutational pattern and specific activity on the genome.

This is illustrated in the figure below using a framework of 6 classes of single base substitutions, and three distinct mutational processes, whose respective strengths vary throughout a patient's life. At the beginning, all mutations were due to the activity of the endogenous mutational process. As time progresses, the other processes get activated and the mutational spectrum of the cancer genome continues to change.

C>A
C>G
C>T
T>A
T>C
T>G
Moderate mutational process activated at different times
APOBEC deaminase activity
Strong exogenous mutational process
Tobacco smoking
Endogenous mutational process
Clock-like signature
Mutational spectrum of final cancer genome
Time
Number of mutations
Signature activity

(1) Sigprofiler tools


https://cancer.sanger.ac.uk/cosmic/signatures/sigprofiler.tt
80 %
COSMIC
Catalogue Of Somatic Mutations In Cancer
Projects ▼ Data ▼ Tools ▼ News ▼ Help ▼ About ▼ Genome Version ▼
Search COSMIC...
SEARCH
Login ▼

Mutational Signatures (v3.1 - June 2020)

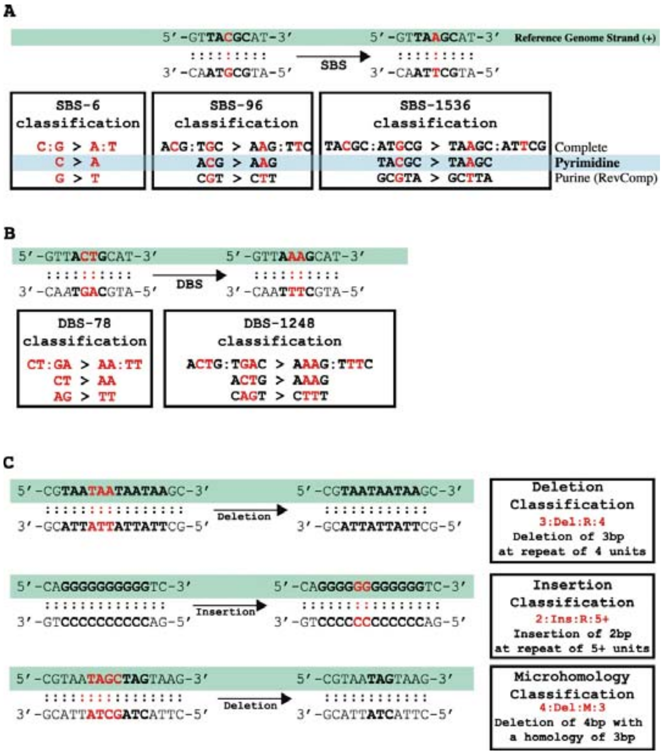
Mutational Signatures Home
Single Base Substitution (SBS) Signatures
Doublet Base Substitution (DBS) Signatures
Small Insertion and Deletion (ID) Signatures
Mutational Signatures Version 2
SigProfiler Bioinformatic Tools

SigProfiler Bioinformatic Tools

SigProfiler provides a comprehensive and integrated suite of bioinformatic tools for performing mutational signature analysis. The software covers the analytical lifecycle starting with the generation of the mutational matrix and finishing with signature extraction, as well as supporting functionality for plotting and simulation.

Hover over any of the logos to learn more about each of our software tools, including the GitHub repository, a wiki page describing how to use the tool and the corresponding publication. All SigProfiler software is available both in Python and R environments.

(1-1) Sigprofiler matrix generator



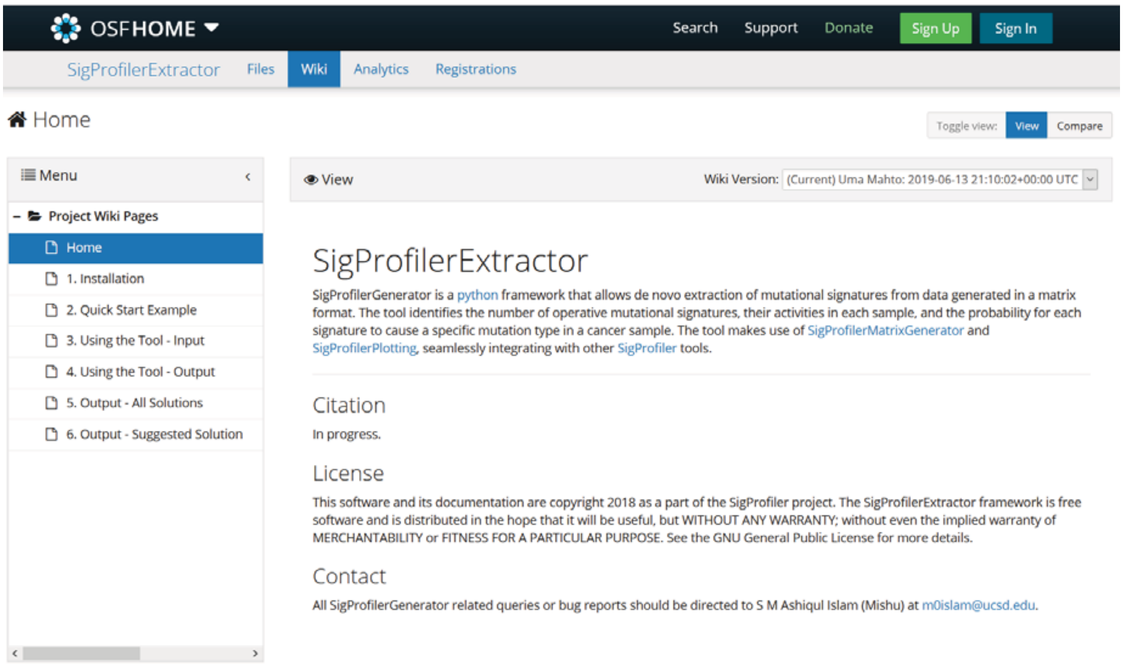
Python and R

(input)
VCF/MAF

(output)
Matrices with
Sequencing context
Transcriptional strand bias

Bergstrom et al., ***BMC Genomics*** (2019)

(1-2) Sigprofiler Extractor



Wiki Version: (Current) Uma Mahto: 2019-06-13 21:10:02+00:00 UTC

SigProfilerExtractor

SigProfilerGenerator is a python framework that allows de novo extraction of mutational signatures from data generated in a matrix format. The tool identifies the number of operative mutational signatures, their activities in each sample, and the probability for each signature to cause a specific mutation type in a cancer sample. The tool makes use of SigProfilerMatrixGenerator and SigProfilerPlotting, seamlessly integrating with other SigProfiler tools.

Citation

In progress.

License

This software and its documentation are copyright 2018 as a part of the SigProfiler project. The SigProfilerExtractor framework is free software and is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

Contact

All SigProfilerGenerator related queries or bug reports should be directed to S M Ashiqul Islam (Mishu) at m0islam@ucsd.edu.

Somatic mutation matrix ➔ NMF ➔ model selection (# of signatures and stability)
➔ Detection of de novo mutational signatures ➔ comparison with known signatures

Tools for extracting mutational signatures

Inferring *de novo* signatures

Alexandrov, MatLab (***Nature*** 2013)
EMu (***Genome Biology*** 2013)
Maftools (***Genome Res*** 2018)
MutationalPatterns (***Genome Med*** 2018)
MutSpec (***BMC Bioinformatics*** 2016)
SigFit (*BioRxiv* 2020)
SigMiner (*medRxiv* 2020)
SignatureAnalyzer (***Nature Commun*** 2015)
SignatureToolsLib (***Nat Cancer*** 2020)
SigneR (***Bioinformatics*** 2017)
SomaticSignatures (***Bioinformatics*** 2015)
SigProfiler (COSMIC)

Fitting known signatures

deconsructSigs (***Genome Biology*** 2016)
SignatureEstimation (***Bioinformatics*** 2018)
YAPSA (R Package v 1.16.0)

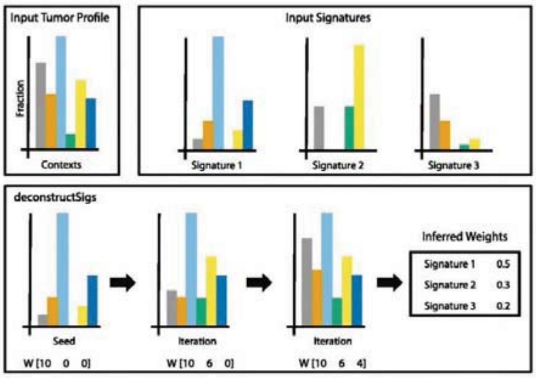
Web interfaces

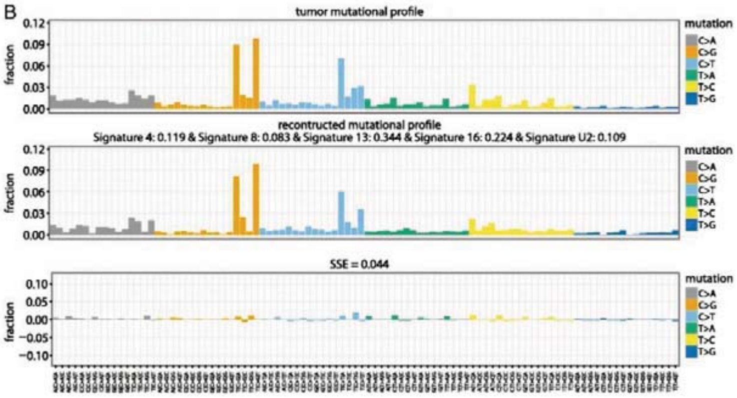
MutaGene (***NAR*** 2017)
mSignatureDB (***NAR*** 2018)
MuSiCa (***BMC Bioinformatics*** 2018)
Mutalisk (***NAR*** 2018)

SBi 한국생명정보학회 Korean Society for Bioinformatics

(2) deconstructSigs

R based package. Mutation matrix as an input (sample, chr, pos, ref, alt)





Rosenthal et al., ***Genome Biology*** (2016)

SBi 한국생명정보학회 Korean Society for Bioinformatics

Tools for extracting mutational signatures

Inferring *de novo* signatures

Alexandrov, MatLab (***Nature*** 2013)
EMu (***Genome Biology*** 2013)
Maftools (***Genome Res*** 2018)
MutationalPatterns (***Genome Med*** 2018)
MutSpec (***BMC Bioinformatics*** 2016)
SigFit (*BioRxiv* 2020)
SigMiner (*medRxiv* 2020)
SignatureAnalyzer (***Nature Commun*** 2015)
SignatureToolsLib (***Nat Cancer*** 2020)
SigneR (***Bioinformatics*** 2017)
SomaticSignatures (***Bioinformatics*** 2015)
SigProfiler (COSMIC)

Fitting known signatures

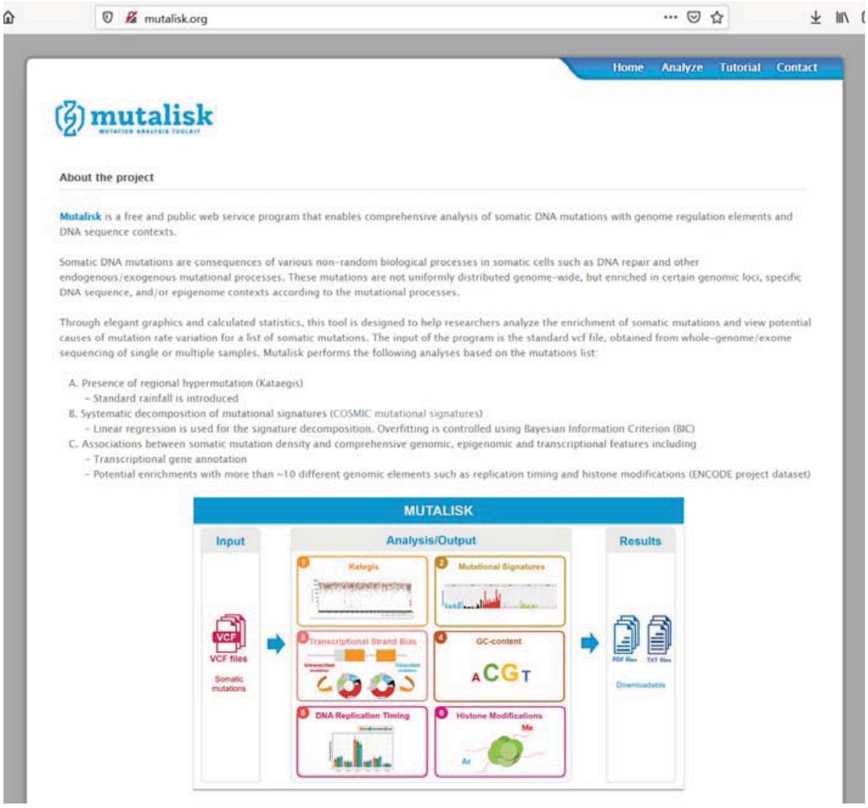
deconsructSigs (***Genome Biology*** 2016)
SignatureEstimation (***Bioinformatics*** 2018)
YAPSA (R Package v 1.16.0)

Web interfaces

MutaGene (***NAR*** 2017)
mSignatureDB (***NAR*** 2018)
MuSiCa (***BMC Bioinformatics*** 2018)
Mutalisk (***NAR*** 2018)

(3) Web interfaces: Mutalisk

http://mutalisk.org



About the project

Mutalisk is a free and public web service program that enables comprehensive analysis of somatic DNA mutations with genome regulation elements and DNA sequence contexts.

Somatic DNA mutations are consequences of various non-random biological processes in somatic cells such as DNA repair and other endogenous/exogenous mutational processes. These mutations are not uniformly distributed genome-wide, but enriched in certain genomic loci, specific DNA sequence, and/or epigenome contexts according to the mutational processes.

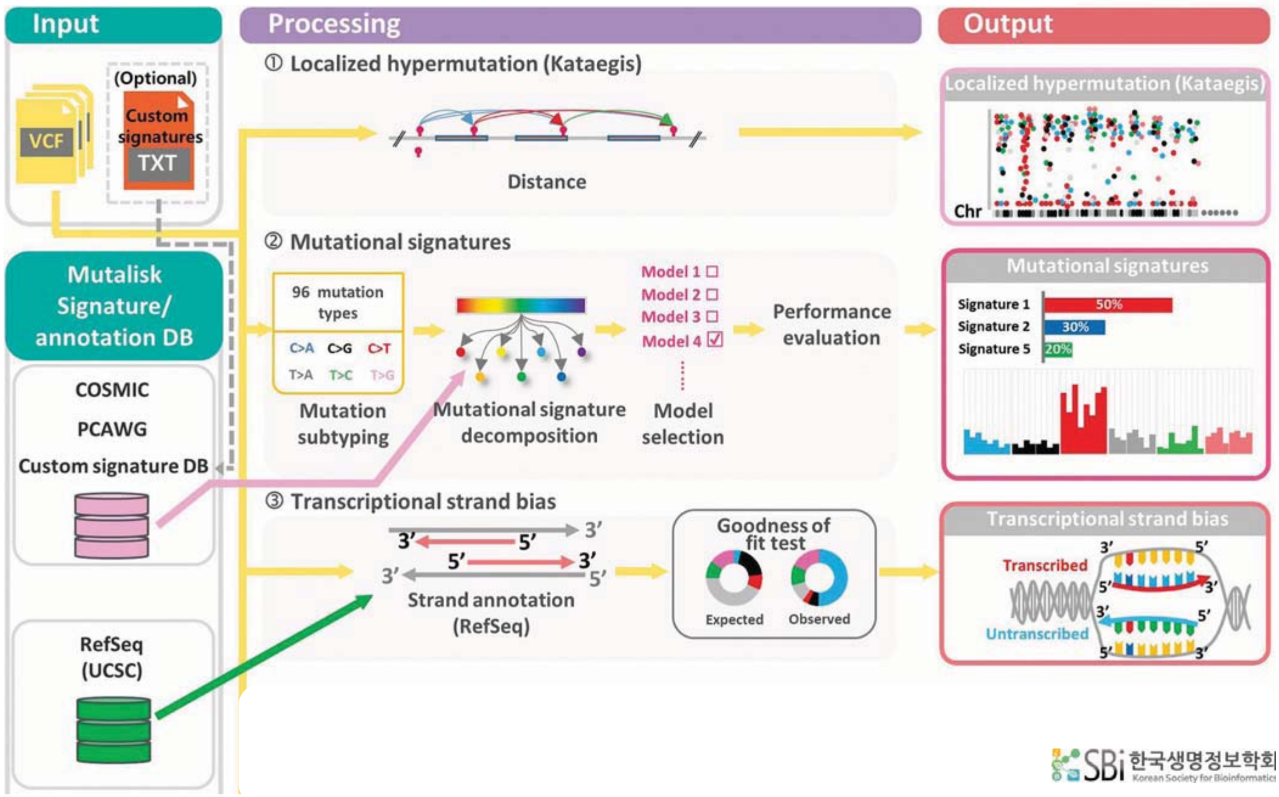
Through elegant graphics and calculated statistics, this tool is designed to help researchers analyze the enrichment of somatic mutations and view potential causes of mutation rate variation for a list of somatic mutations. The input of the program is the standard vcf file, obtained from whole-genome/exome sequencing of single or multiple samples. Mutalisk performs the following analyses based on the mutations list:

A. Presence of regional hypermutation (Kataegis)
  - Standard rainfall is introduced
B. Systematic decomposition of mutational signatures (COSMIC mutational signatures)
  - Linear regression is used for the signature decomposition. Overfitting is controlled using Bayesian Information Criterion (BIC)
C. Associations between somatic mutation density and comprehensive genomic, epigenomic and transcriptional features including
  - Transcriptional gene annotation
  - Potential enrichments with more than ~10 different genomic elements such as replication timing and histone modifications (ENCODE project dataset)

Mutalisk: a web-based somatic MUTation AnaLyIS toolKit for genomic, transcription and epigenomic signatures

Lee JK et al., NAR 2018

(3) Workflow in Mutalisk



Input

(Optional)

VCF

Custom signatures TXT

Mutalisk Signature/ annotation DB

COSMIC

PCAWG

Custom signature DB

RefSeq (UCSC)

Processing

① Localized hypermutation (Kataegis)

Distance

② Mutational signatures

96 mutation types

C>A C>G C>T T>A T>C T>G

Mutation subtyping

Mutational signature decomposition

Model 1 □

Model 2 □

Model 3 □

Model 4 ☑

Model selection

Performance evaluation

③ Transcriptional strand bias

3′ 5′ 3′

5′ 3′

3′ 5′

Strand annotation (RefSeq)

Goodness of fit test

Expected

Observed

Output

Localized hypermutation (Kataegis)

Chr

Mutational signatures

Signature 1 50%

Signature 2 30%

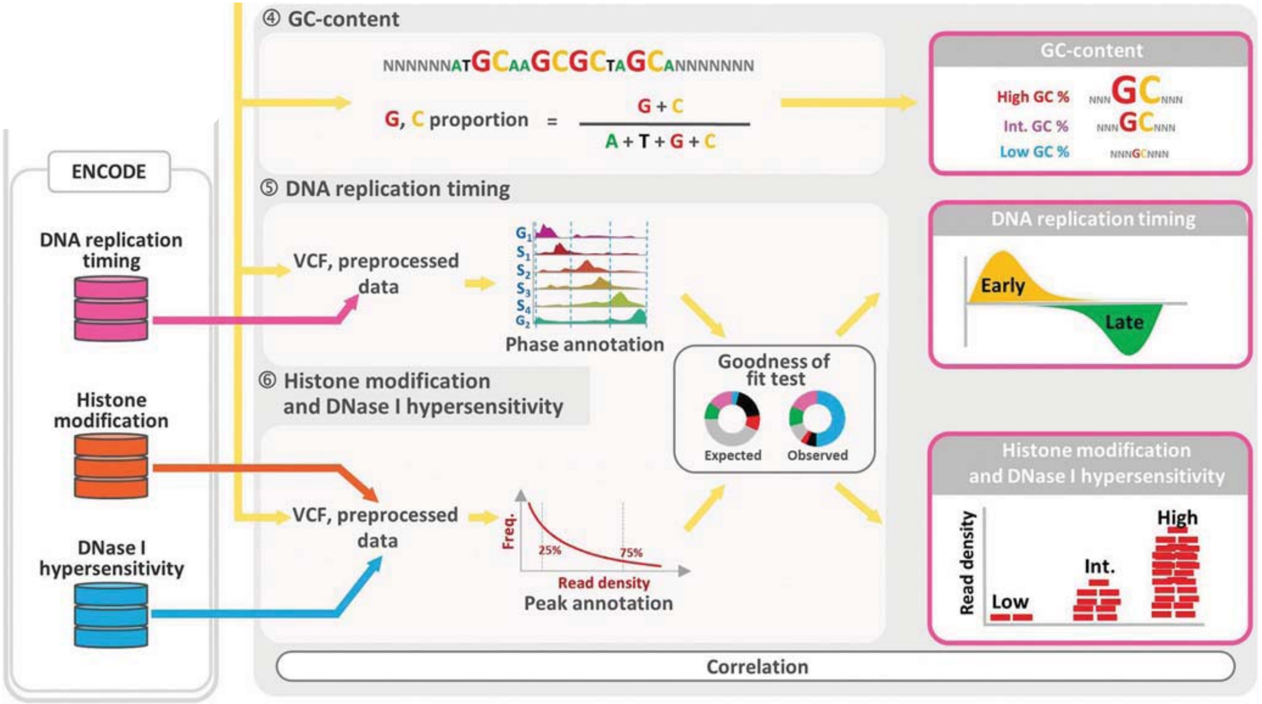
Signature 5 20%

Transcriptional strand bias

Transcribed

Untranscribed

(3) Workflow in Mutalisk (2)



ENCODE

DNA replication timing

Histone modification

DNase I hypersensitivity

④ GC-content

NNNNNNATGCAAGCGCTAGCANNNNNNN

G, C proportion = $\frac{G+C}{A+T+G+C}$

⑤ DNA replication timing

VCF, preprocessed data

G_1 S_1 S_2 S_3 S_4 G_2

Phase annotation

⑥ Histone modification and DNase I hypersensitivity

VCF, preprocessed data

Freq.

25%

75%

Read density

Peak annotation

Goodness of fit test

Expected

Observed

Correlation

GC-content

High GC % NNNGCNNN

Int. GC % NNNGCNNN

Low GC % NNNGCNNN

DNA replication timing

Early

Late

Histone modification and DNase I hypersensitivity

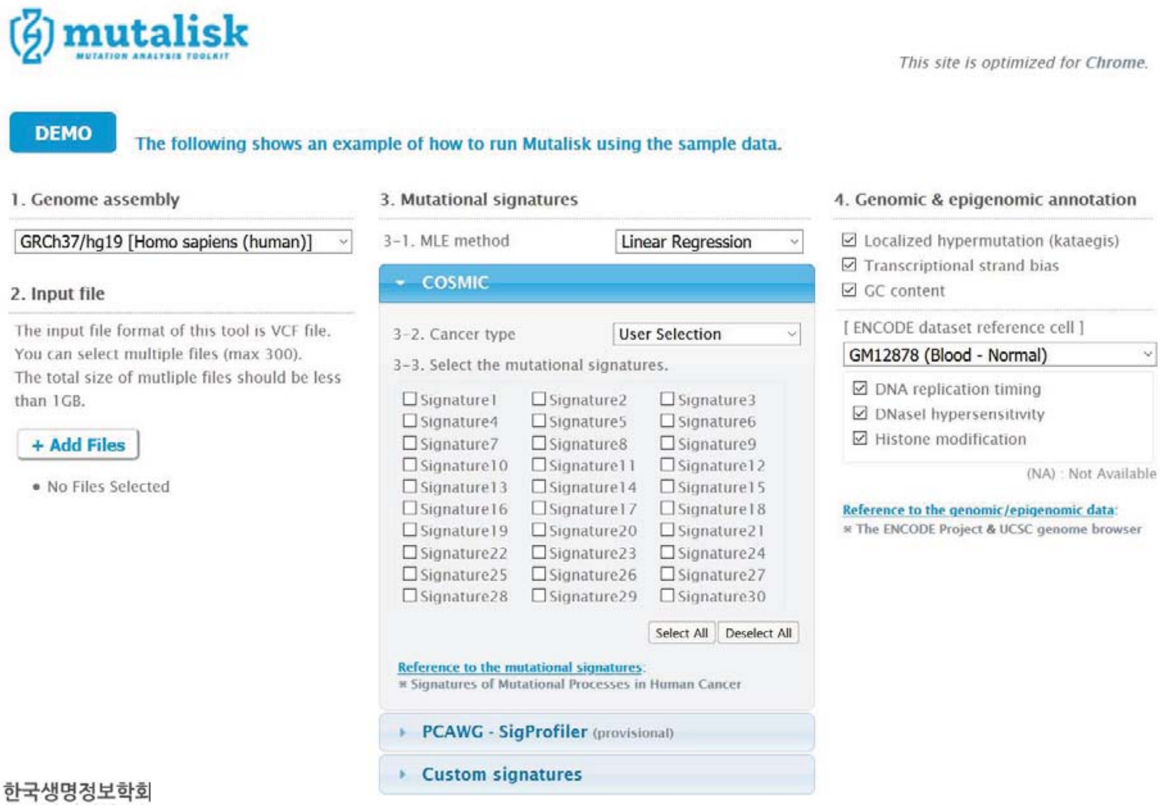
Read density

Low

Int.

High

(3) Input for Mutalisk



This site is optimized for Chrome.

DEMO The following shows an example of how to run Mutalisk using the sample data.

1. Genome assembly

GRCh37/hg19 [Homo sapiens (human)]

2. Input file

The input file format of this tool is VCF file.
You can select multiple files (max 300).
The total size of mutliple files should be less than 1GB.

+ Add Files

- No Files Selected

3. Mutational signatures

3-1. MLE method: Linear Regression

COSMIC

3-2. Cancer type: User Selection

3-3. Select the mutational signatures.

□ Signature1	□ Signature2	□ Signature3
□ Signature4	□ Signature5	□ Signature6
□ Signature7	□ Signature8	□ Signature9
□ Signature10	□ Signature11	□ Signature12
□ Signature13	□ Signature14	□ Signature15
□ Signature16	□ Signature17	□ Signature18
□ Signature19	□ Signature20	□ Signature21
□ Signature22	□ Signature23	□ Signature24
□ Signature25	□ Signature26	□ Signature27
□ Signature28	□ Signature29	□ Signature30

Select All | Deselect All

Reference to the mutational signatures:
※ Signatures of Mutational Processes in Human Cancer

PCAWG - SigProfiler (provisional)

Custom signatures

4. Genomic & epigenomic annotation

☑ Localized hypermutation (kataegis)
☑ Transcriptional strand bias
☑ GC content

[ENCODE dataset reference cell]

GM12878 (Blood - Normal)

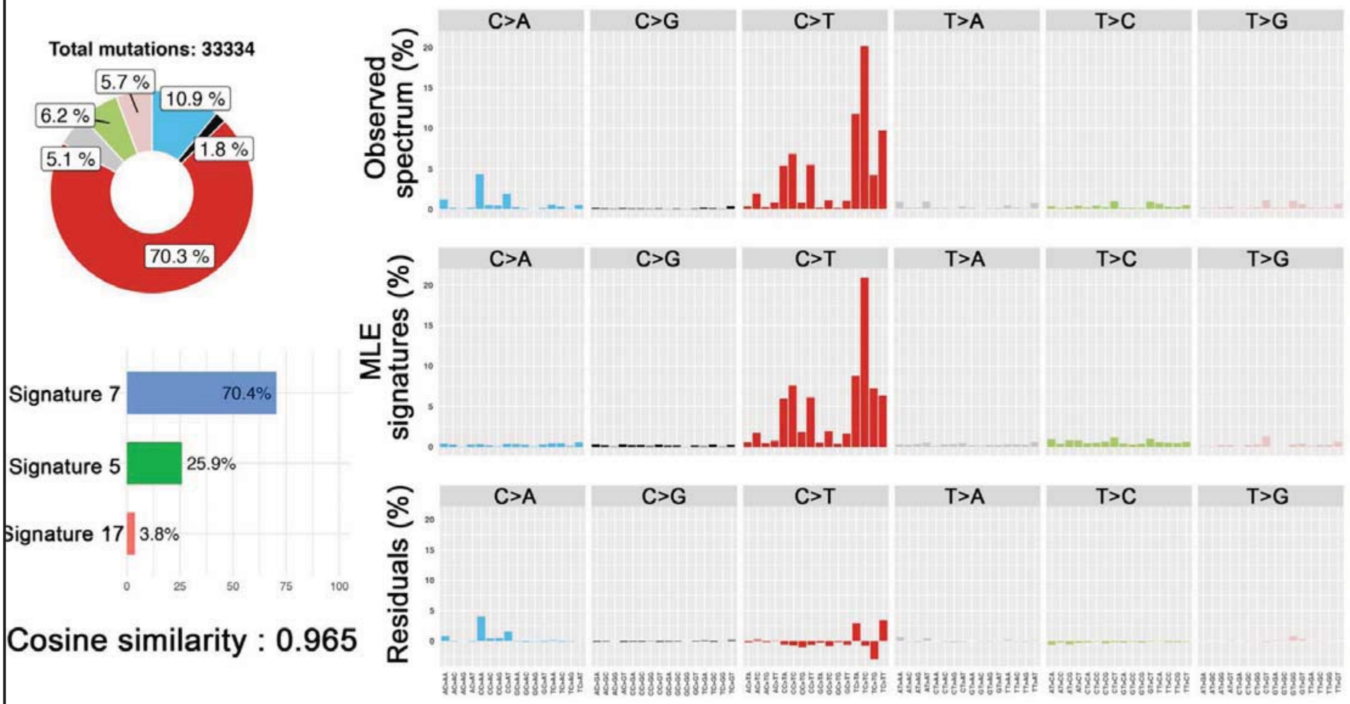
☑ DNA replication timing
☑ DNaseI hypersensitivity
☑ Histone modification

(NA) : Not Available

Reference to the genomic/epigenomic data:
※ The ENCODE Project & UCSC genome browser

SBi 한국생명정보학회 Korean Society for Bioinformatics

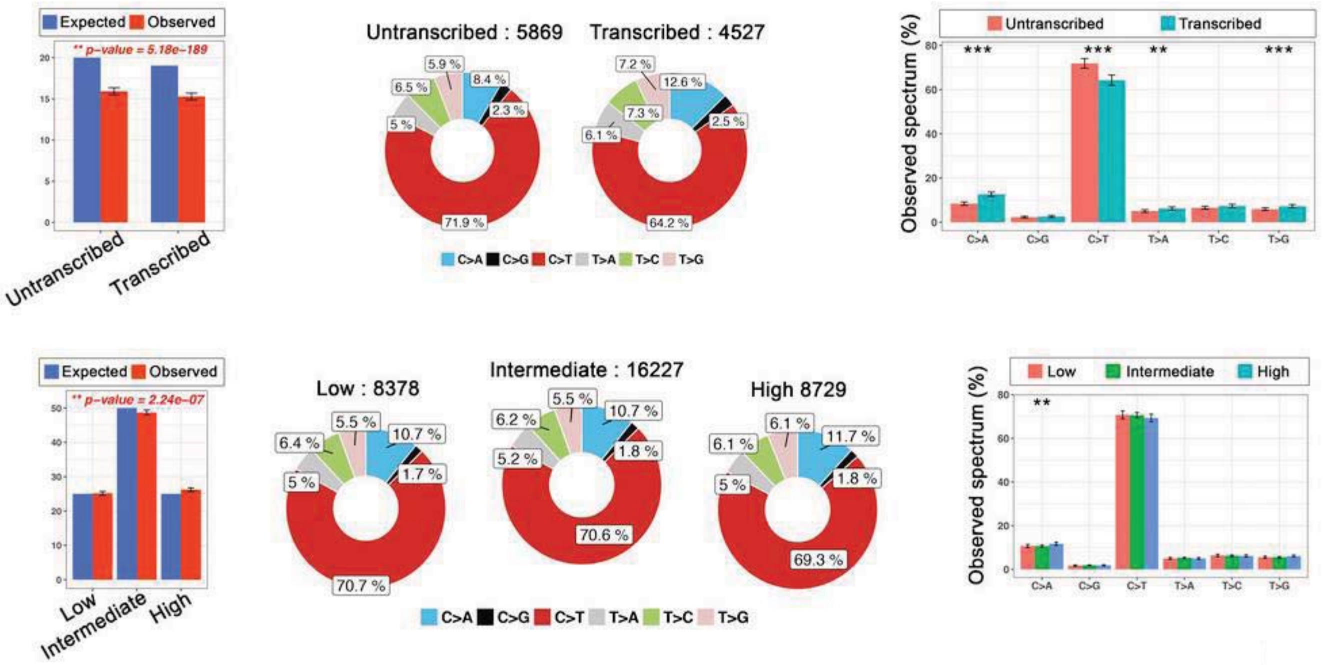
(3) Output in Mutalisk (1)



Cosine similarity : 0.965

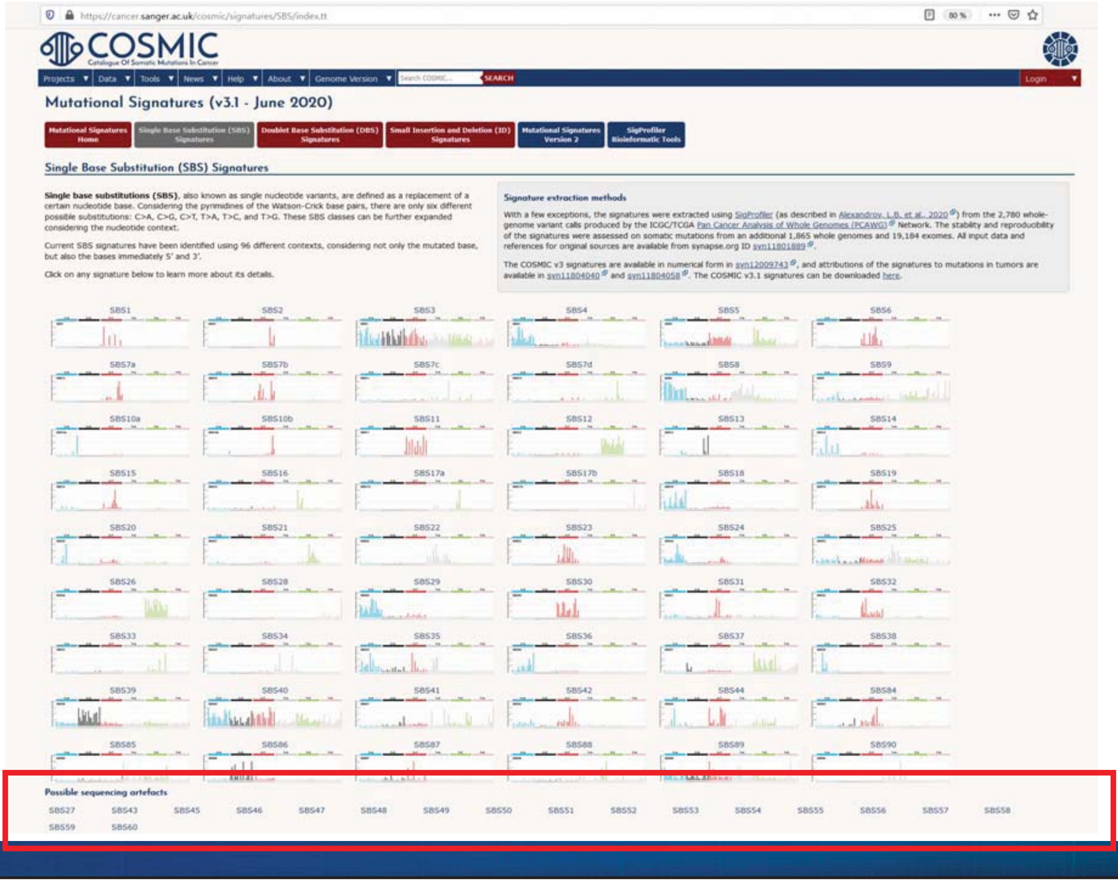
SBi 한국생명정보학회 Korean Society for Bioinformatics

(3) Output in Mutalisk (2)



Genome QC with mutational signatures

Amplification/sequencing artifacts make unique signatures



Mutational Signatures (v3.1 - June 2020)

Single Base Substitution (SBS) Signatures

Single base substitutions (SBS), also known as single nucleotide variants, are defined as a replacement of a certain nucleotide base. Considering the pyrimidines of the Watson-Crick base pairs, there are only six different possible substitutions: C>A, C>G, C>T, T>A, T>C, and T>G. These SBS classes can be further expanded considering the nucleotide context.

Current SBS signatures have been identified using 96 different contexts, considering not only the mutated base, but also the bases immediately 5' and 3'.

Click on any signature below to learn more about its details.

Signature extraction methods

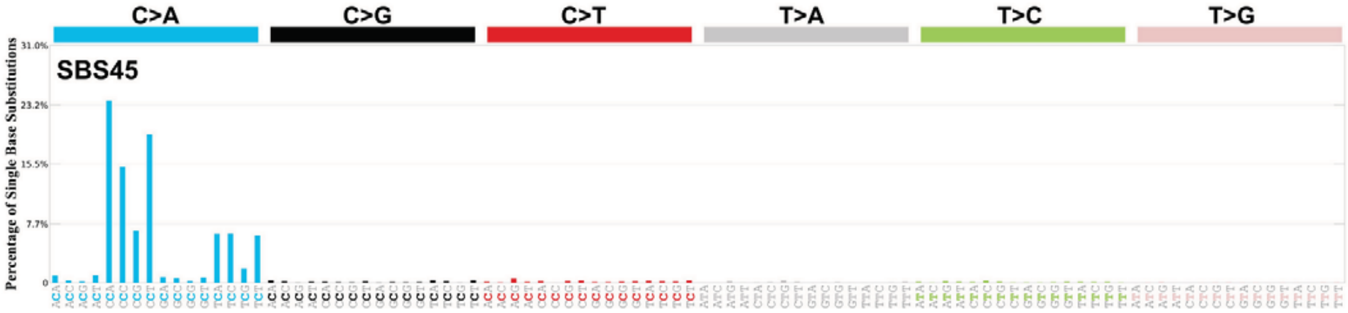
With a few exceptions, the signatures were extracted using SigProfiler (as described in Alexandrov, L.B. et al., 2020) from the 2,780 whole-genome variant calls produced by the ICGC/TCGA Pan Cancer Analysis of Whole Genomes (PCAWG) Network. The stability and reproducibility of the signatures were assessed on somatic mutations from an additional 1,865 whole genomes and 19,184 exomes. All input data and references for original sources are available from synapse.org ID syn11801889.

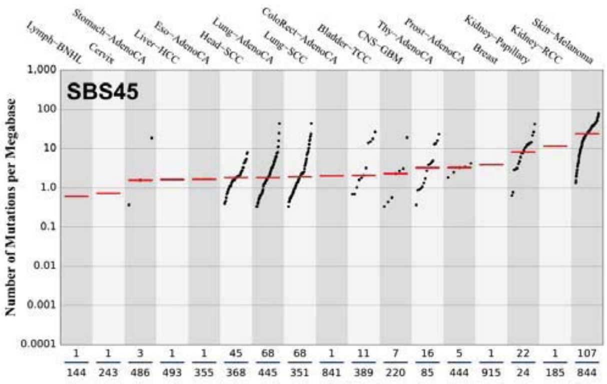
The COSMIC v3 signatures are available in numerical form in syn12009743, and attributions of the signatures to mutations in tumors are available in syn11804040 and syn11804058. The COSMIC v3.1 signatures can be downloaded here.

Possible sequencing artefacts

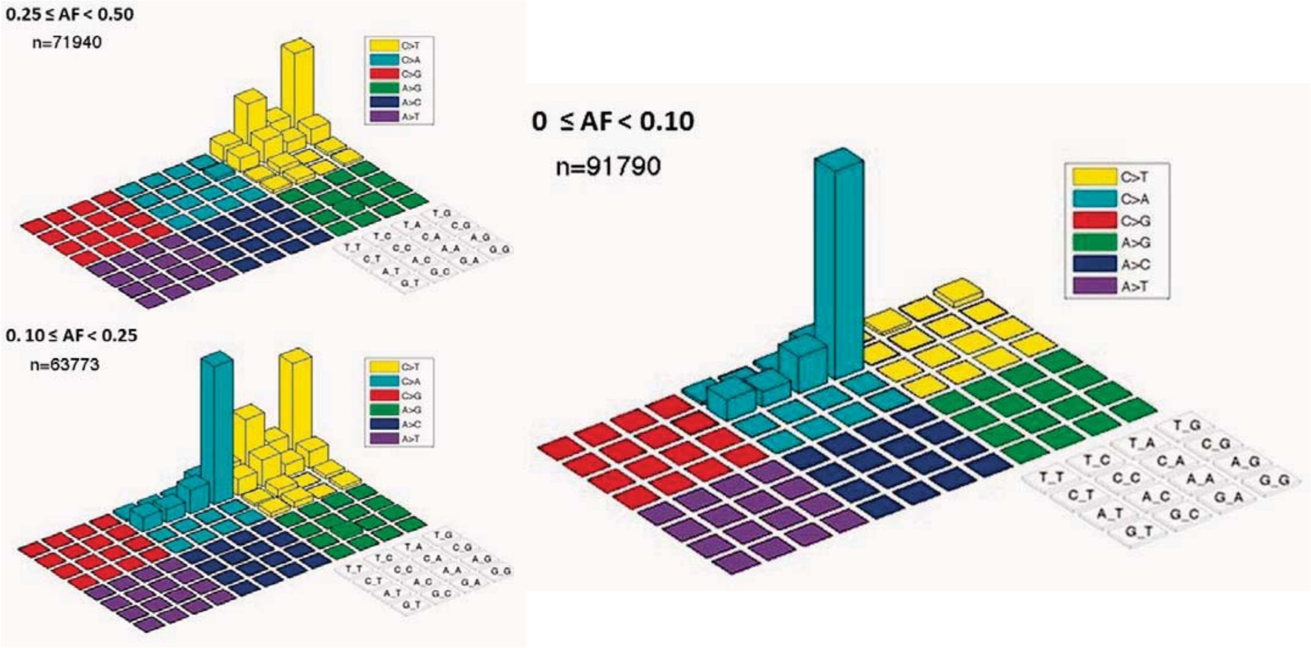
SBS27 SBS43 SBS45 SBS46 SBS47 SBS48 SBS49 SBS50 SBS51 SBS52 SBS53 SBS54 SBS55 SBS56 SBS57 SBS58
SBS59 SBS60

SBS45, a signature of 8-oxoG artifact



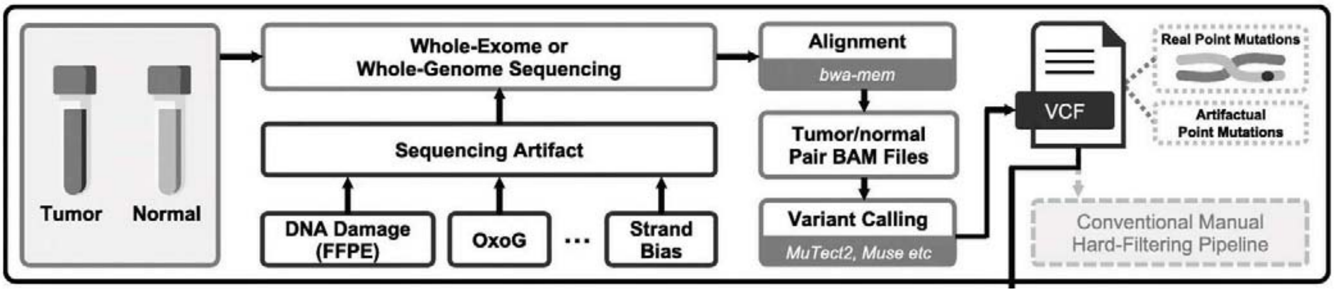


First report for 8-oxoG artificial signature



Costello et al., ***NAR*** (2012)

A typical pipeline for cancer genome analyses

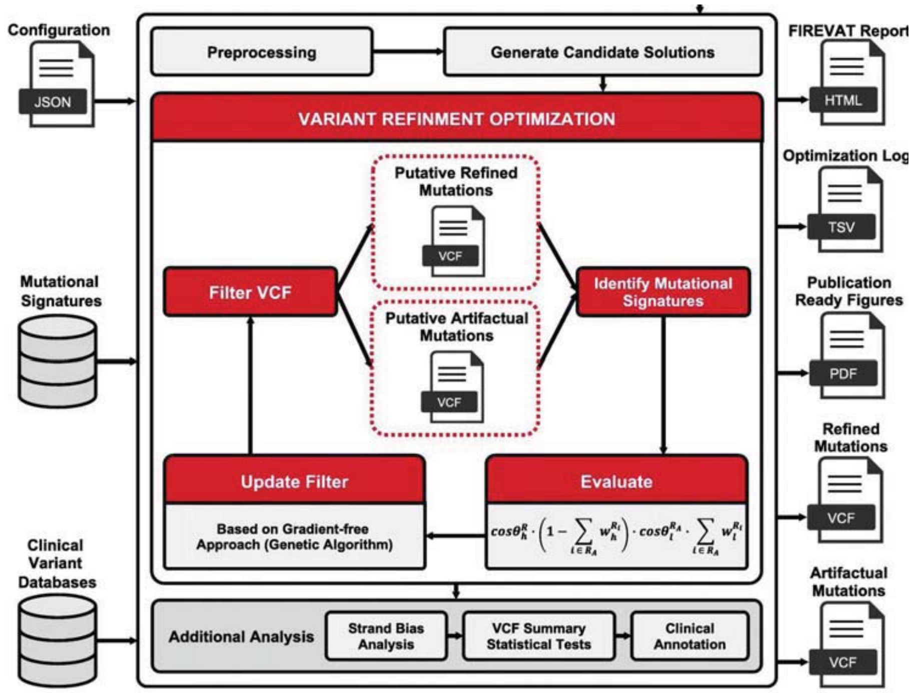


Workflow in FIREVAT, a software for filtering artifacts

Software | Open Access | Published: 17 December 2019

FIREVAT: finding reliable variants without artifacts in human cancer samples using etiologically relevant mutational signatures

Kim et al., Genome Medicine 2019



Configuration

JSON

Mutational Signatures

Clinical Variant Databases

Preprocessing

Generate Candidate Solutions

VARIANT REFINMENT OPTIMIZATION

Putative Refined Mutations

VCF

Filter VCF

Identify Mutational Signatures

Putative Artifactual Mutations

VCF

Update Filter

Based on Gradient-free Approach (Genetic Algorithm)

Evaluate

$\cos\theta_h^R \cdot \left(1 - \sum_{i \in R_A} w_h^{R_i}\right) \cdot \cos\theta_i^{R_A} \cdot \sum_{i \in R_A} w_i^{R_i}$

Additional Analysis

Strand Bias Analysis

VCF Summary Statistical Tests

Clinical Annotation

FIREVAT Report

HTML

Optimization Log

TSV

Publication Ready Figures

PDF

Refined Mutations

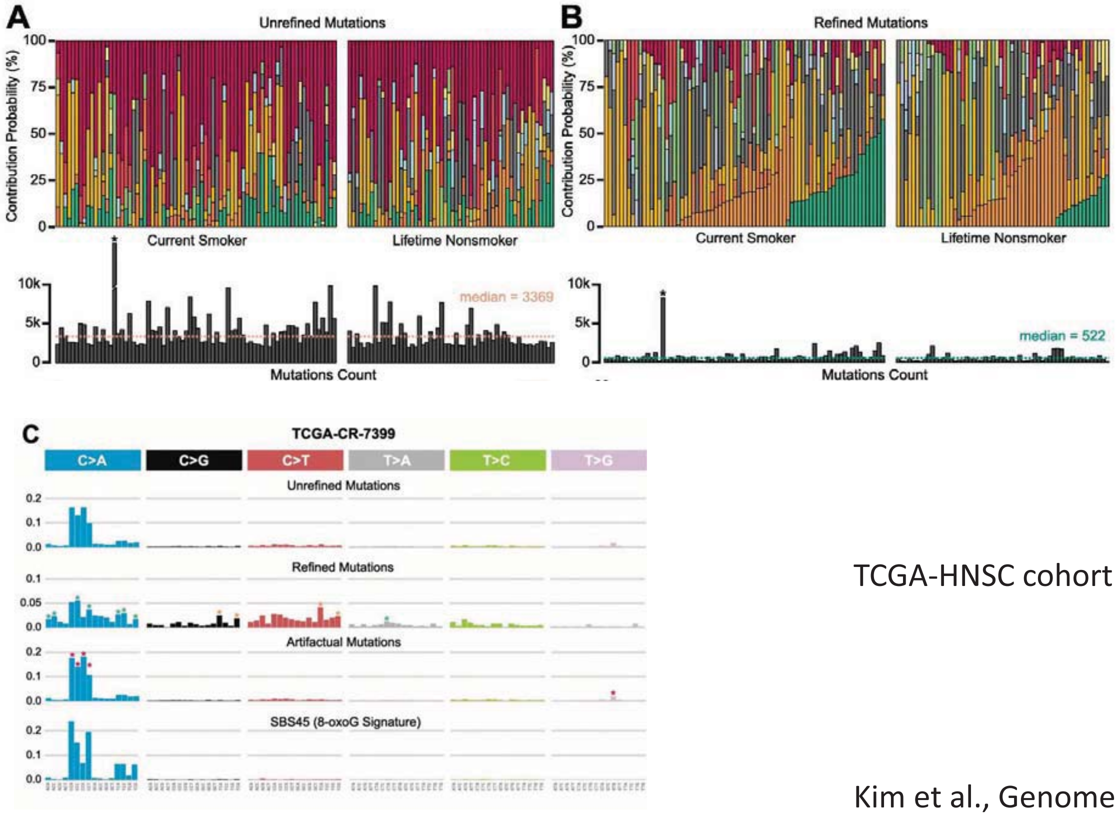
VCF

Artifactual Mutations

VCF

SBi 한국생명정보학회 Korean Society for Bioinformatics

Filtering mutations using FIREVAT



TCGA-HNSC cohort

Kim et al., Genome Med 2019

전망

- 돌연변이 signature 개수는 총 몇 개가 될까?
- 돌연변이 signature 각각의 원인을 규명할 수 있을까?
- Structural variation의 signature는 무엇이 있을까?

Summary

- 돌연변이는 random 하게 생기지 않는다
- Mutational signature 개념을 이용하여 정확한 variant calling 을 할 수 있다
- Mutational signature 개념을 이용하여 돌연변이가 만들어진 원인을 추적할 수 있다
- Mutational signature를 구하는 tool을 이해하고 사용할 수 있다.