

# K-CORE Portal

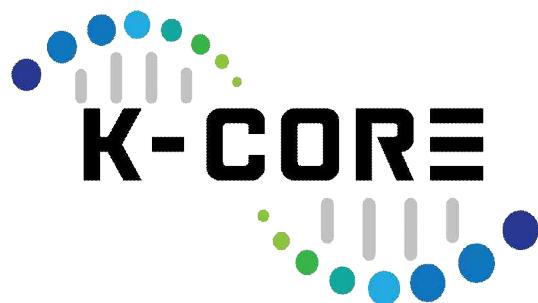
## User Manual

---

*Korean Cancer Omics  
Research Portal*

2023년 9월

VER 1.2



## - 목 차 -

### I. 시작하기

사이트 접속	3
사이트 소개	4
분석 접수번호 생성	5
로그인	7

### II. 데이터 분석 예시

데이터 분석 예시 화면구성	8
단일 데이터 시각화 안내	10
① Variant Summary	10
② Circos Plot	11
③ Lollipop Plot	14
④ CNV	16
⑤ Heatmap	18
⑥ Box Plot	21
⑦ Survival Plot	23
다중 데이터 시각화 안내	25
① Circos Plot	25
② OncoPrint	27
③ Lollipop Plot	29
④ Volcano Plot	31
⑤ Heatmap	34
⑥ Survival Plot	37
⑦ Correlation	40
⑧ CNV	41
⑨ Box Plot	43
⑩ Fusion Plot	45
⑪ Sankey Plot	48

### III. 사용자 데이터 분석

단일 데이터 시각화	51
다중 데이터 시각화	57
기타 분석 도구	64

---

---

## 부록 - 사용자 임상정보 및 오믹스 데이터 안내

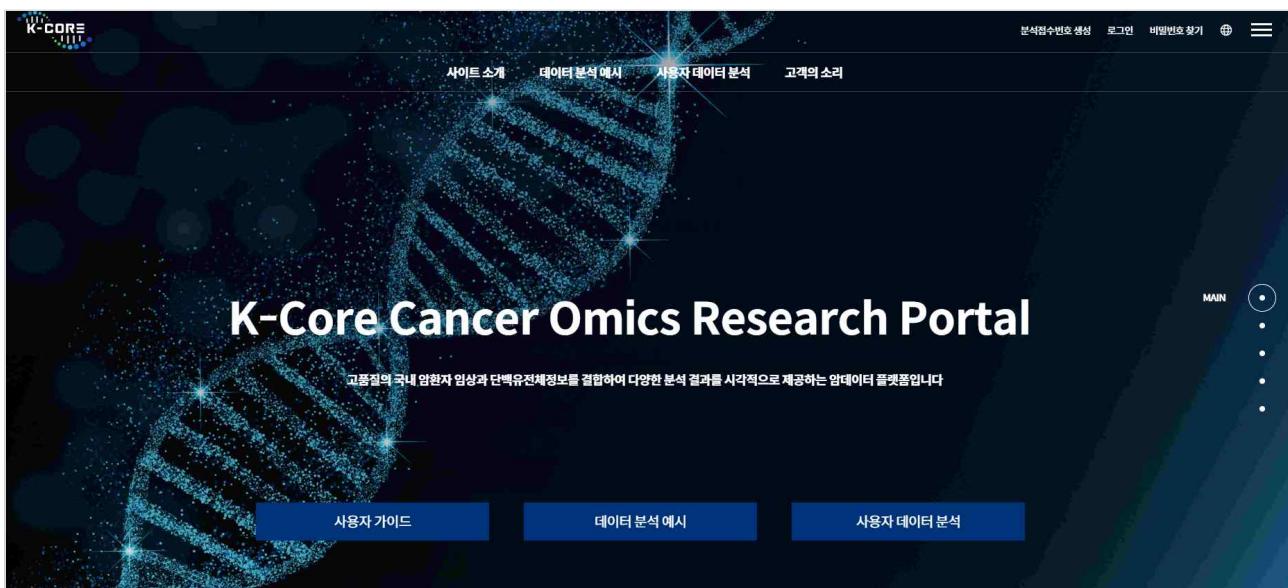
임상정보(Clinical information) .....	67
오믹스 데이터(Omics data) .....	69
① DNA Mutation .....	69
② CNV .....	71
③ Methylation .....	72
④ RNA .....	73
⑤ Fusion .....	74
⑥ Proteome .....	75
⑦ Phosphorylation .....	76

## I . 시작하기

### K-CORE Portal 시작은 이렇게 합니다.

국립암센터에서 제공하는 K-CORE Portal 플랫폼을 통해 고품질의 국내 암 환자 임상 정보와 단백유전체 정보를 연계한 통계 그래프를 생성하여 볼 수 있습니다. 현재 제공하는 사용자 데이터 분석 서비스에서는 사용자가 가진 임상 및 오믹스 데이터를 사용자가 원하는 대로 필터링하여 분석할 수 있게 구성되었습니다.

#### 1. 사이트 접속



- [HOME] 화면 구성은 다음과 같습니다.
  - . 메뉴는 사이트 소개, 데이터 분석 예시, 사용자 데이터 분석, 고객의 소리로 구성됩니다.
  - . 분석접수번호 생성, 로그인, 비밀번호 찾기, 언어선택, 펼친 메뉴[...] 기능을 지원합니다.
  - . [사용자 가이드] 연동 버튼을 통해 홈페이지 가이드.pdf 파일을 다운받을 수 있습니다.
  - . [데이터 분석 예시], [사용자 데이터 분석] 연동 버튼을 통해 각 메뉴로 이동을 지원합니다.

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

### 2. 사이트 소개

- 메인 메뉴의 [사이트 소개]를 클릭하거나, 마우스 스크롤을 UP/DOWN을 통해 사이트 소개, 서비스 소개, 운영 조직, 유관 사이트를 보실 수 있습니다.

The screenshot shows the 'Site Introduction' section of the K-CORE website. It features a large image of a modern building complex, likely the headquarters. Below the image, there are three bullet points: '국가암데이터센터 사이트의 하위 서비스로써 암 유전체 데이터를 시각화하여 제공하는 웹 기반 분석 포털입니다.', '단일 데이터 분석 및 일상정보를 연계한 월드오믹스 분석 기능', '시각화 분석 결과 디문포드 기능', and '액션 단계 정보 제공 기능'.

< 사이트 소개 >

The screenshot shows the 'Service Introduction' section of the K-CORE website. It displays three service icons: '단일 데이터 시각화' (Single Data Visualization), '다중 데이터 시각화' (Multi-Data Visualization), and '기타 분석 도구' (Other Analysis Tools). Each icon has a brief description below it.

< 서비스 소개 >

The screenshot shows the 'Operation Organization' section of the K-CORE website. It features a large image of a person in a lab coat. Below the image, there are four categories: '데이터 분석', '데이터 관리', '데이터 분석', and '데이터 관리'. Each category has a list of sub-categories or specific data sets.

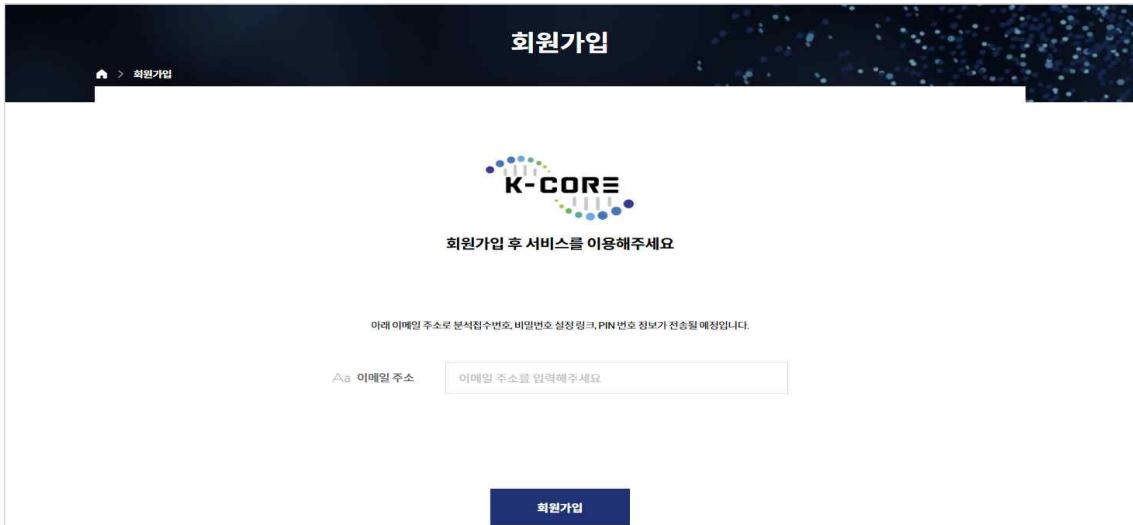
< 운영 조직 >

The screenshot shows the 'Related Websites' section of the K-CORE website. It lists several external databases: 'dbiPortal', 'COSMIC', 'TCGA', 'CCLE', 'IGC', and 'EuropDX'. Below the list, there is a brief description of dbiPortal and two bullet points about its features.

< 유관 사이트 >

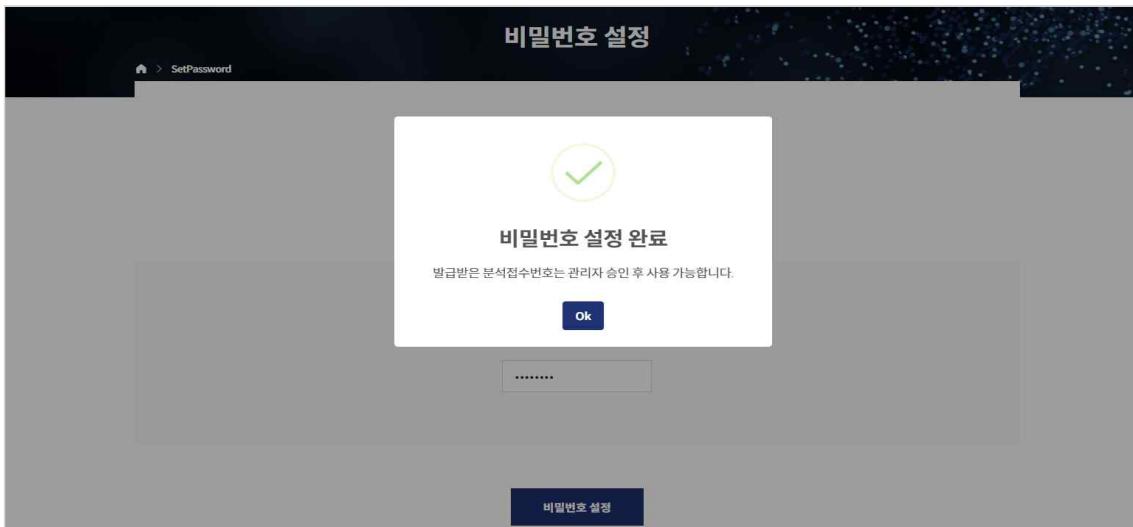
### 3. 분석 접수번호 생성

- 사이트 상단 오른쪽의 [분석접수번호 생성]을 클릭하여 아이디를 발급받을 수 있습니다. 아이디를 발급받을 이메일 주소를 입력하여 [회원가입]을 누릅니다.



< 분석접수번호 생성 화면 >

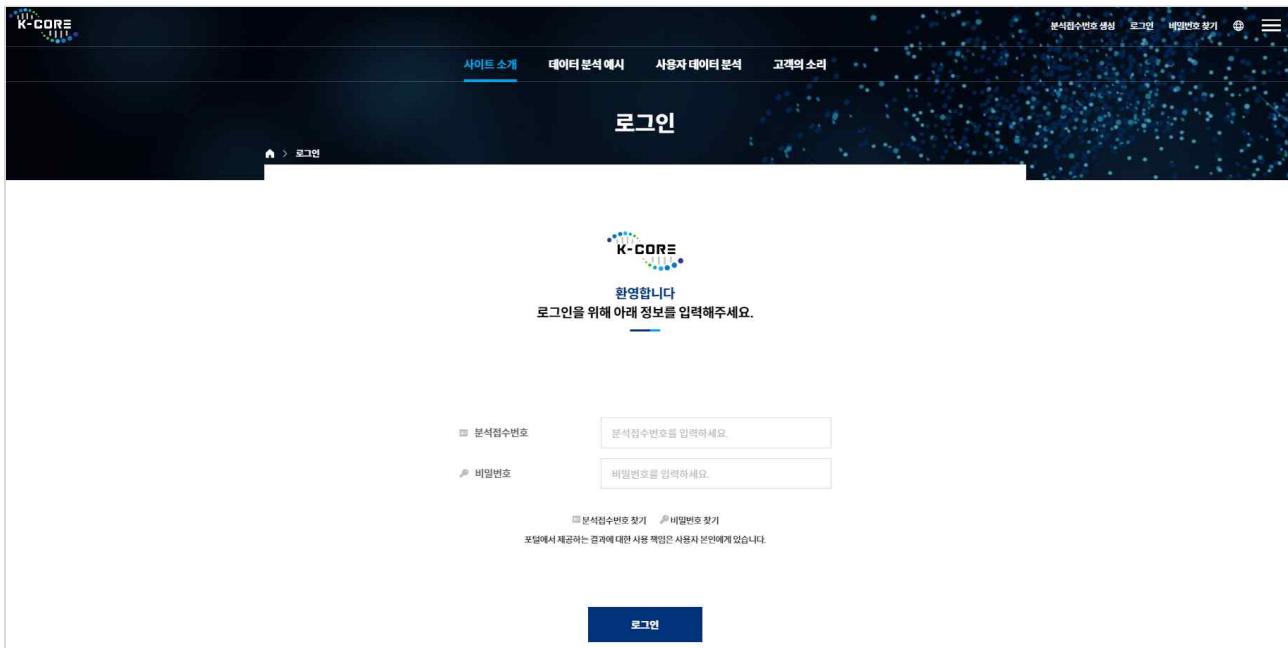
- 입력한 이메일을 통해 분석접수번호, PIN번호 및 비밀번호 설정 링크를 받습니다. 비밀번호 설정 링크를 누르면 발급 받은 분석접수번호에 대한 비밀번호 설정 화면으로 이동합니다. 비밀번호 설정을 완료하면 관리자의 승인 후, 로그인하여 사용자 데이터 시각화를 비롯하여 포털의 모든 기능을 이용할 수 있습니다.



< 비밀번호 설정 화면 >

### 4. 로그인

- 상단의 로그인 버튼을 클릭하여 로그인 화면으로 이동한 뒤, 아이디와 비밀번호를 입력하여 로그인할 수 있습니다.



< 로그인 화면 >

- . 상단의 [로그인] 버튼을 클릭합니다.
- . 아이디와 비밀번호를 입력한 후, [로그인] 버튼을 클릭하여 로그인합니다.
- . 아이디를 잊어버렸을 시, [아이디 찾기] 버튼을 클릭하여 ID를 찾습니다.
- . 비밀번호를 잊어버렸을 시, [비밀번호 찾기] 버튼을 클릭하여 비밀번호를 찾습니다.

## II. 데이터 분석 예시

### 암 환자 데이터를 연계, 분석합니다

데이터 분석 예시 페이지는 국내 암 환자 100명의 임상 정보와 단백·유전체 정보를 연계하여 수행할 수 있는 여러 가지 종류의 통계분석 및 시각화 기능을 제공합니다. [데이터 분석 예시] 메뉴에선 사용자 선택에 따라 단일 데이터 시각화 및 다중 데이터 시각화 분석, 기타 분석 도구에 대한 안내를 볼 수 있습니다. [예시 페이지로 이동]을 통해 예제 데이터에 대한 시각화 분석을 할 수 있고, [메뉴얼 다운로드]를 통해 사용자 가이드 매뉴얼을 다운받을 수 있습니다.

### 데이터 분석 예시 화면구성

#### ① 단일 데이터 시각화 가이드

- . 사용 가능한 분석 도구에 대한 안내
- . [예시 페이지로 이동] : 100명의 암 환자 예제 데이터에 대한 예시시각화 페이지로 이동
- . [메뉴얼 다운로드] : 분석 가이드 매뉴얼 다운로드
- . 단일 데이터로 시각화 가능한 플롯 별 필요한 오믹스 데이터 안내

Data Type	Variant Summary	Circos	Lollipop	CNV	Heatmap	Box	Survival
Clinical Information							
DNA Mutation	○	○	○				○
CNV		○		○			
Methylation	○				○		
RNA	○				○	○	
Fusion	○						
Proteome	○		○		○	○	
Phosphorylation					○		

< 단일 데이터 시각화 가이드 화면 >

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

### ② 다중 데이터 시각화 가이드

- . 사용 가능한 분석 도구에 대한 안내
- . [예시 페이지로 이동] : 100명의 암 환자 예제 데이터에 대한 예시시각화 페이지로 이동
- . [메뉴얼 다운로드] : 분석 가이드 매뉴얼 다운로드
- . 다중 데이터로 시각화 가능한 플롯 별 필요한 오믹스 데이터 안내

The screenshot shows the K-CORE web application. At the top, there's a navigation bar with tabs: '사이트 소개' (Site Introduction), '데이터 분석 예시' (Data Analysis Examples), '사용자 데이터 분석' (User Data Analysis), and '고객의 소리' (Customer Feedback). Below the navigation is a search bar and user authentication links ('분석집수번호 생성', '로그인', '비밀번호 찾기').  
The main content area has three tabs at the top: '단일 데이터 시각화 가이드' (Single Data Visualization Guide), '다중 데이터 시각화 가이드' (Multiple Data Visualization Guide) (which is selected), and '기타 분석 도구 가이드' (Other Analysis Tools Guide).  
A 'Contents' section lists several visualization methods:

- Circos: 원형의 연색체 지도에 최대 7가지 오믹스 데이터를 시각화
- Oncoprint: 각 환자의 유전자 내 DNA 변이와 여러 오믹스 정보를 한 세션, 기호 등으로 시각화
- Lollipop: 특정 유전자의 돌연변이 혹은 인산화물을 아미노산 서열 위에서 시각화
- Volcano: 임상조건에 따라 나는 환자 그룹 간 유의한 발현 차이를 보이는 유전자(DEG) 시각화
- Correlation: 선택된 유전자에 대해 RNA 발현 갑과 Proteome(질부도) 갈의 상관관계를 시각화
- CNV: 통합 유전체 뷰어를 사용한 목제수 변이 데이터 시각화
- Box (Tumor vs Normal): 선택된 유전자의 유전정보 통계를 박스 플롯 형태로 표현

A 'Data Type' table maps these tools to different data types:

Data Type	Circos	Oncoprint	Lollipop	Volcano	Heatmap	Survival	Correlation	CNV	Box	Fusion	Sankey
Clinical Information											
DNA Mutation	○	○	○			○					
CNV	○							○			
Methylation	○				○						
RNA	○			○	○	○	○		○		
Fusion	○				○					○	
Proteome	○				○	○	○	○		○	
Phosphorylation					○						

On the right side, there are buttons for '예시 페이지로 이동' (Move to Example Page) and '메뉴얼 다운로드' (Manual Download). A 'VIZUALIZE EXAMPLE DATA' button is also present.

< 다중 데이터 시각화 가이드 화면 >

### ③ 기타 분석 도구 가이드

- . 사용 가능한 분석 도구에 대한 안내
- . [메뉴얼 다운로드] : 분석 가이드 매뉴얼 다운로드

The screenshot shows the K-CORE web application. The layout is similar to the previous one, with a navigation bar and search bar at the top.  
The main content area has three tabs at the top: '단일 데이터 시각화 가이드' (Single Data Visualization Guide), '다중 데이터 시각화 가이드' (Multiple Data Visualization Guide), and '기타 분석 도구 가이드' (Other Analysis Tools Guide) (selected).  
A 'Contents' section lists several file conversion tools:

- VCF to MAF: VCF 파일로 자동 변환 하는 도구 (VCF 파일 필요)
- Interpro: 단백질 서열정보를 통해 단백질을 분류하고, 도메인과 주요 사이트를 예측하여 기능 분석을 제공하는 도구 (Protein FASTA 파일 필요)
- BLAST: 단백질의 아미노산 서열 또는 DNA/RNA 서열 정보를 통해 1차 생물학적 서열 정보를 비교하는 도구 (DNA/RNA Protein 서열 정보 또는 해당 FASTA 파일 필요)

A '메뉴얼 다운로드' button is located on the right side of the contents area. A 'VIZUALIZE EXAMPLE DATA' button is also present.

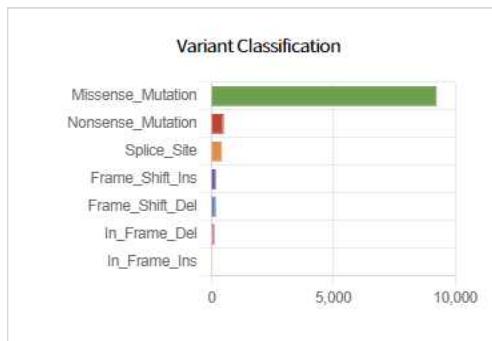
< 기타 분석 도구 가이드 화면 >

## 1. 단일 데이터 시각화 안내

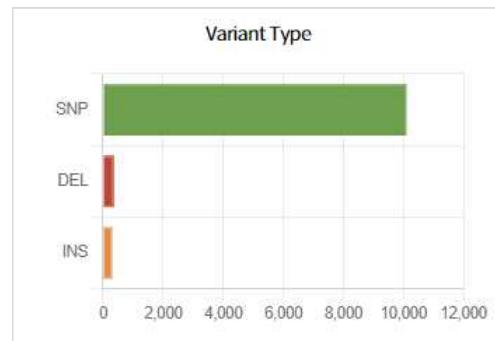
### 시각화 가능한 도구 안내

#### ① Variant Summary

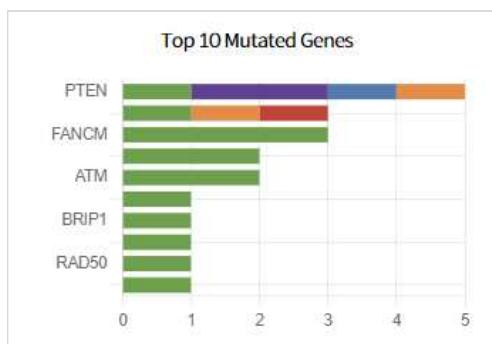
- . 시각화 가능한 오믹스 데이터 : DNA Mutation
- . Variant Summary 도구를 통해 샘플에 대한 변이의 요약 통계 정보를 얻을 수 있습니다.
- . 사용자 DNA Mutation 오믹스 데이터를 통해 변이 분류(Classification), 종류(Type)에 대한 요약된 통계 정보를 시각화하여 제공합니다. (차트에 대해 마우스 오버시, 해당 정보에 대한 라벨을 확인 가능)
  - 1) Variant Classification : 전체 DNA Mutation 오믹스 데이터에 대한 주요 7가지 변이 통계 그래프
  - 2) Variant Type : 전체 DNA Mutation 오믹스 데이터에 대한 변이 타입 통계 그래프
  - 3) Top 10 Mutated Genes : [유전자 세트 필터링]을 통해 선택한 유전자에 대해 주요 7가지 변이를 가지는 상위 10개 유전자에 대한 통계 그래프 (선택한 유전자가 10개 미만인 경우 선택한 유전자만 그래프로 나타냄)
  - 4) Variant Classification Summary : 전체 DNA Mutation 오믹스 데이터에 대한 주요 7가지 변이에 대한 박스형 그래프



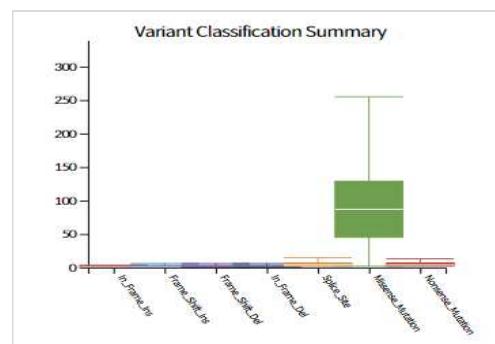
< Variant Classification >



< Variant Type >



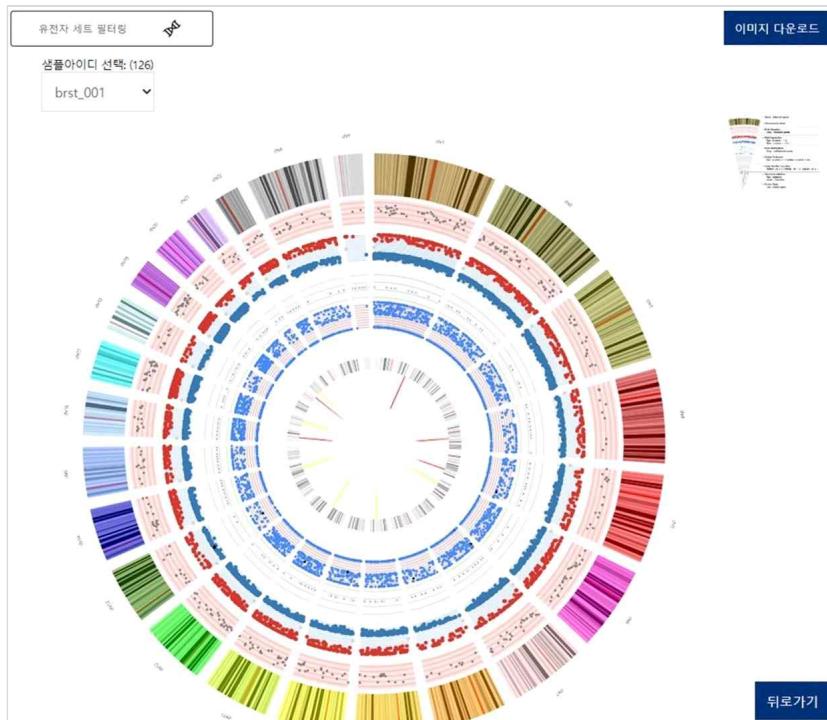
< Top 10 Mutated Genes >



< Variant Classification Summary >

### ② Circos Plot

- . 시각화 가능한 오믹스 데이터 : DNA Mutation, CNV, Methylation, RNA, Fusion, Proteome
- . Circos plot은 원형의 염색체 지도 위에 7가지 유전체 데이터를 다수의 레이어로 나타내는 플롯이며, 각 유전체 데이터를 각 레이어 위에 서로 다른 색상의 점이나 선으로 표시합니다.
- . 단일 데이터 Circos plot 시각화 도구는 사용자가 업로드한 데이터에 대해서만 레이어가 표시됩니다.  
(실제 사용자 시각화 결과와 아래 예시 화면 결과가 다를 수 있습니다.)

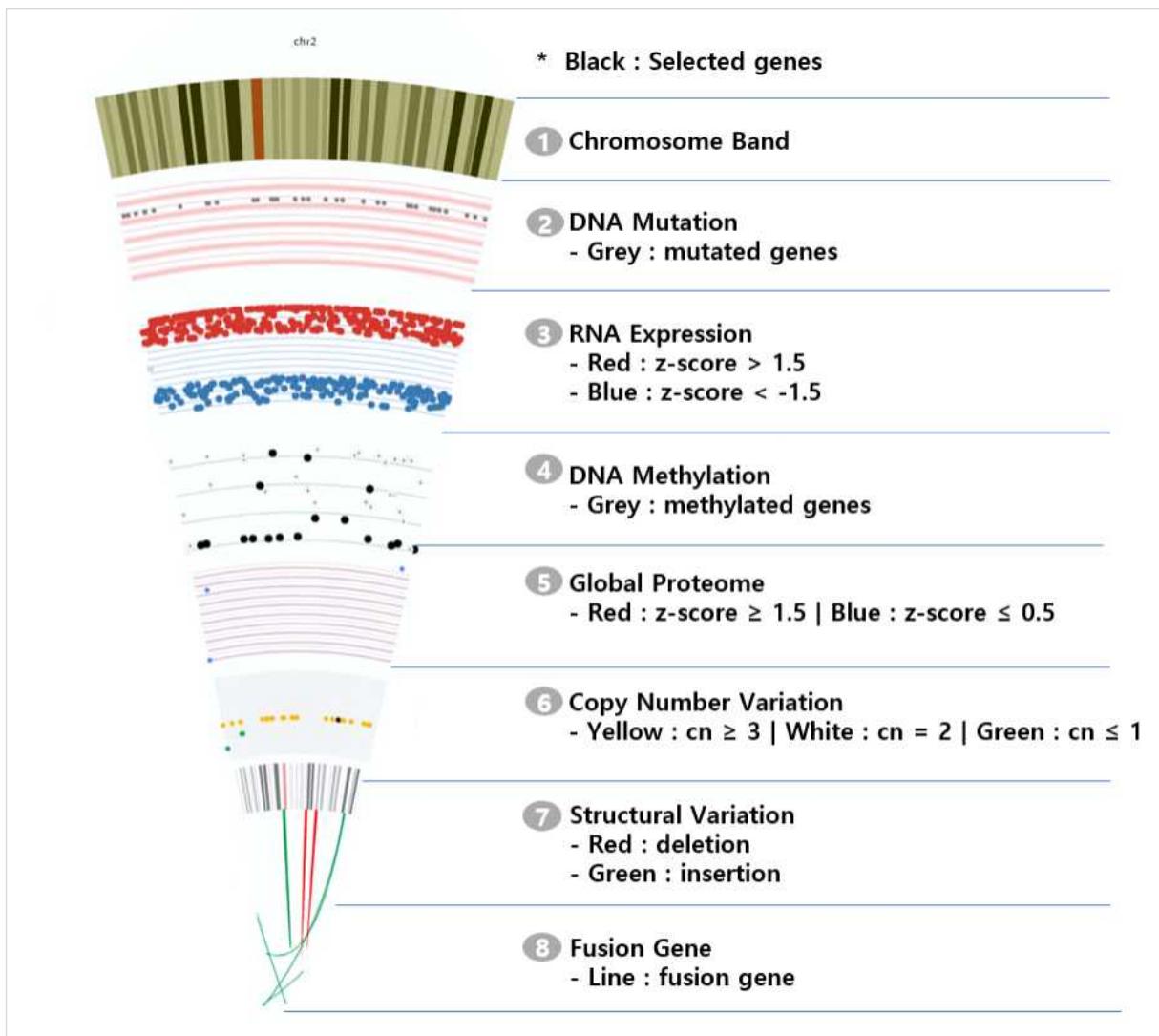


< 전체 오믹스 데이터에 대한 Circos Plot 예시 화면 >

- . 단일 데이터 시각화는 오직 하나의 오믹스 데이터를 시각화하기 때문에 임상정보 필터 기능이 없습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자를 선택할 수 있습니다.
  - 선택한 유전자 세트는 Circos Plot에서 검은색 점으로 표시됩니다.
- . [범례]에서 Circos plot 구성요소에 대한 정보를 확인할 수 있습니다.
  - 범례 그림을 누르면 팝업 안내를 통해 더 자세히 볼 수 있습니다.

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

. Circos plot의 각 레이어는 다음과 같은 정보를 표시합니다.



< Circos plot 범례 >

## 암 임상-옴릭스 분석 시각화 시스템(K-CORE)

---

. Circos plot 레이어 상세 정보

Layer	Omics data	Information
1	Chromosome(염색체)	<ul style="list-style-type: none"> <li>Chromosome band (염색체 밴드)</li> </ul>
2	DNA Mutation	<ul style="list-style-type: none"> <li>Mutation의 염색체상 위치 표시</li> <li>값이 클수록 바깥쪽 레이어에 위치 표시</li> <li>회색 점으로 표시</li> <li>마우스 오버 시, '염색체:위치, Hugo gene symbol' 표시</li> </ul>
3	RNA Expression	<ul style="list-style-type: none"> <li>발현된 유전자의 염색체상 위치 표시</li> <li>값이 클수록 바깥쪽 레이어에 위치 표시</li> <li><b>빨간색 점</b> : Z-score <math>\geq 1.5</math></li> <li><b>푸른색 점</b> : Z-score <math>\leq -1.5</math></li> <li>마우스 오버 시, 'Hugo gene symbol, 염색체:위치, Z-score' 표시</li> </ul>
4	DNA Methylation	<ul style="list-style-type: none"> <li>Methylation의 염색체상 위치 표시</li> <li>값이 클수록 바깥쪽 레이어에 위치 표시</li> <li>회색 점으로 표시</li> <li>마우스 오버 시, 'Hugo gene symbol, 염색체:위치, gene_vl'</li> </ul>
5	Global Proteome	<ul style="list-style-type: none"> <li>Proteome 발현을 염색체상 위치 표시</li> <li>값이 클수록 바깥쪽 레이어에 위치 표시</li> <li><b>푸른색 점</b>으로 표시</li> <li>마우스 오버 시, 'Hugo gene symbol, 염색체:위치, gene_vl'</li> </ul>
6	Copy Number Variation	<ul style="list-style-type: none"> <li>Copy number variation의 염색체상 위치 표시</li> <li><b>노란색 점</b> : cn <math>\geq 3</math></li> <li>검정색 점 : cn = 2</li> <li><b>초록색 점</b> : cn <math>\leq 1</math></li> </ul>
7	Structural Variation	<ul style="list-style-type: none"> <li>Structural variation 보유한 유전자의 염색체상 위치 표시</li> <li>값이 클수록 바깥쪽 레이어에 위치 표시</li> <li><b>빨간색 선</b> : Deletion</li> <li><b>노란색 선</b> : Insertion</li> <li>마우스 오버 시, 'Insertion/Deletion, 염색체, Hugo gene symbol' 표시</li> </ul>
8	Fusion gene	<ul style="list-style-type: none"> <li>융합(fusion)된 두 유전자의 위치를 선으로 연결하여 표시</li> <li>마우스 오버 시, 'Left Gene - Right Gene' 표시</li> </ul>

< 표. Circos plot 레이어 상세 정보 >

### ③ Lollipop Plot

- . 시각화 가능한 오믹스 데이터 : DNA Mutation, Phosphorylation
- . Lollipop plot을 통해 특정 유전자의 돌연변이나 인산화를 아미노산 서열 위에서 파악할 수 있습니다.
- . 단일 데이터 시각화는 오직 하나의 오믹스 데이터를 시각화하기 때문에 임상정보 필터 기능이 없습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자를 선택할 수 있습니다.
- . [유전자 선택] 드롭다운 메뉴에서 원하는 유전자 하나를 선택합니다.
- . 업로드한 오믹스 데이터에 따라 돌연변이(Mutation)와 인산화(Phosphorylation) 탭으로 구성됩니다.

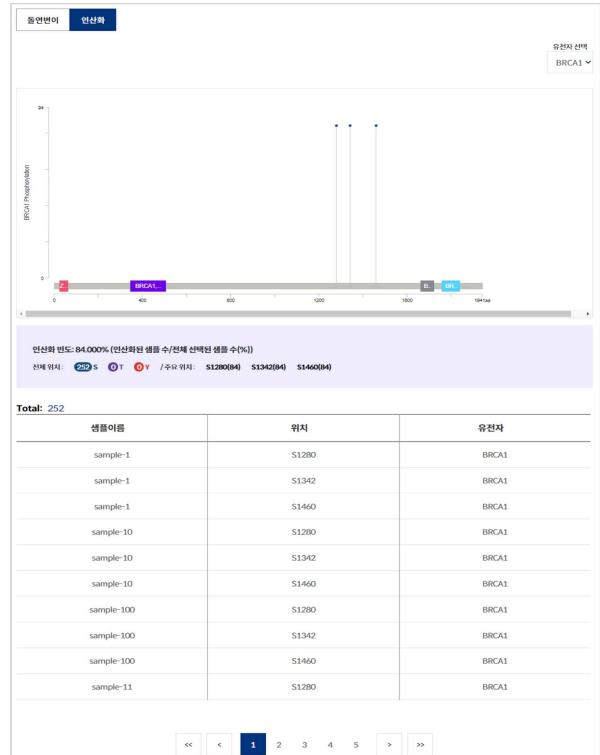
#### 1) 돌연변이 (Mutation) : 아미노산 서열이 바뀐 돌연변이 위치, 도메인 정보, 개수를 시각화

- 선택된 유전자의 mRNA ID 목록을 표시 (ENST ID 목록, RefSeq ID 목록)
- X축 : 아미노산 위치 (aa)
- X축 색상 박스 : 도메인 이름 (마우스 오버 시, Full domain name 이름 볼 수 있음)
- Y축 : Mutation 개수 (Lollipop plot 머리에 마우스 오버 시, 개수 볼 수 있음)
- Lollipop 머리 색상 : Mutation 종류
- 체세포 변이 빈도(%) : 전체 샘플 중 선택한 유전자의 돌연변이를 보유한 샘플의 비율
- 범례 : Variant classification 색상 (= Lollipop plot 머리 색상)과 개수

#### 2) 인산화 (Phosphorylation) : 단백질 내 인산화 위치, 도메인 정보, 개수를 시각화

- X축 : 아미노산 위치 (aa)
- X축 색상 박스 : 도메인 이름 (마우스 오버 시, Full domain name 이름 볼 수 있음)
- Y축 : Phosphorylation site 개수 (Lollipop plot 머리에 마우스 오버 시, 개수 볼 수 있음)
- Lollipop 머리 색상 : Phosphorylation site 종류
- 범례 :
  - . Total Site : S, T, Y Site 색상 (= Lollipop plot 머리 색상)과 개수
  - . Major Site : 가장 많이 검출된 Site 이름과 수 (여러 개 가능)
- 인산화 빈도(%) : 전체 샘플 중 선택한 유전자의 인산화를 보유한 샘플의 비율
- Phosphorylation 표 : 각 Phosphophorylation의 샘플 ID, Site, 유전자

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

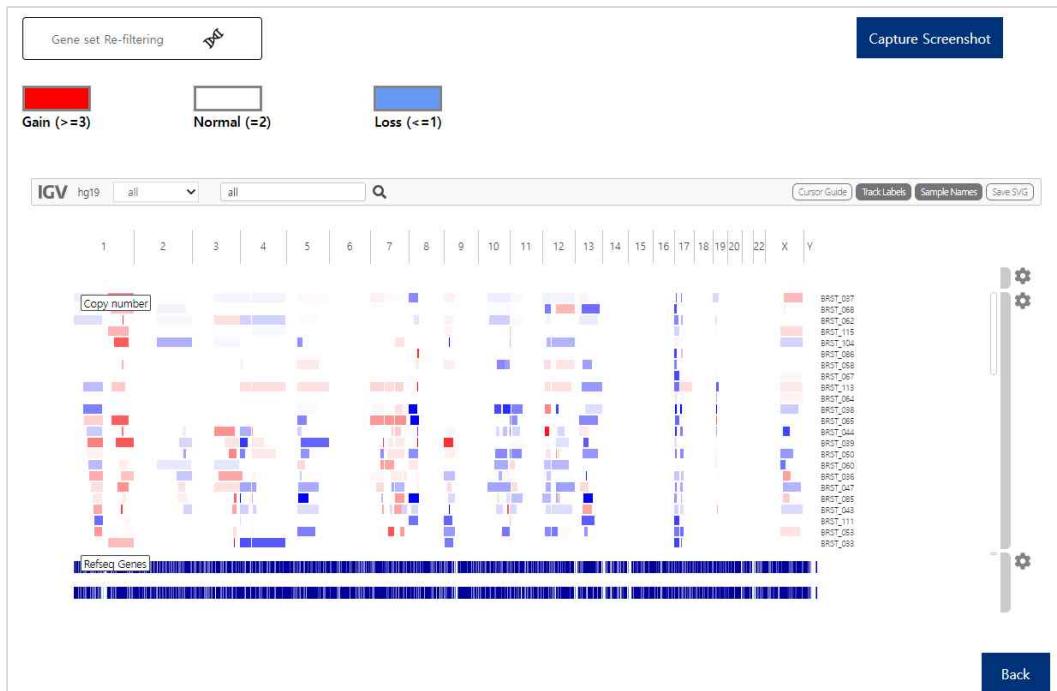


< Lollipop Plot - 돌연변이 화면 >

< Lollipop Plot - 인산화 화면 >

### ④ CNV

- . 시각화 가능한 오믹스 데이터 : CNV
- . CNV는 복제 수 변이 정보를 Zoom-in/out 가능한 유전체 뷰어에서 시각화합니다.
  - 마우스 스크롤 UP/DOWN을 통해 확대 가능
- . CNV plot을 통해 전체 복제 수 변이 데이터를 시각적으로 파악할 수 있습니다.



〈 CNV 시각화 화면〉

- . 단일 데이터 시각화는 오직 하나의 오믹스 데이터를 시각화하기 때문에 임상정보 필터 기능이 없습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자를 선택할 수 있습니다.
- . IGV(Integrated Genome Viewer)에서 염색체와 위치(base-pair) 범위를 검색하면 해당 범위를 확대하여 시각화합니다.  
예시) IGV의 첫 번째 상자에서 드롭 다운 기능을 통해 원하는 위치의 염색체로 이동 가능  
IGV의 두 번째 상자에서 [chr번호]:[위치 범위]를 통해 해당 위치로 이동 가능 (chr10:87800-87971)
- . Copy Number 시각화를 지원합니다.
  - X축 : 염색체 및 위치 (확대 시, base 단위로 볼 수 있음)
  - Y축 : 샘플 이름
  - 붉은색 : Copy number gain ( $CN \geq 3$ )
  - 흰색 : Copy number normal ( $CN = 2$ )

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

---

---

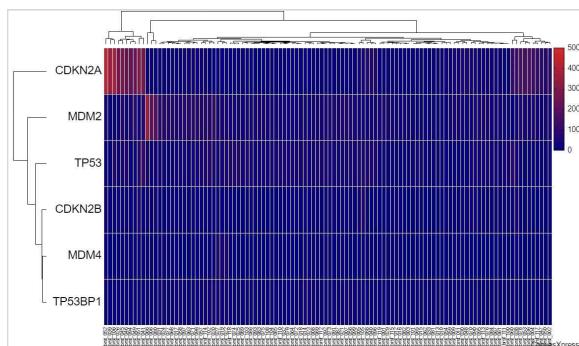
- 푸른색 : Copy number loss ( $CN \leq 1$ )
- . RefSeq Genes 트랙을 지원합니다. (확대 시, 상세 보기 가능)
- . 그 외 옵션 가능
  - Cursor Guide : 마우스 커서 위치 강조
  - Track labels : 트랙 레이블 보이게 설정
  - Sample names : 샘플 이름 보이게 설정
  - Save SVG : SVG 형식 다운로드

### ⑤ Heatmap

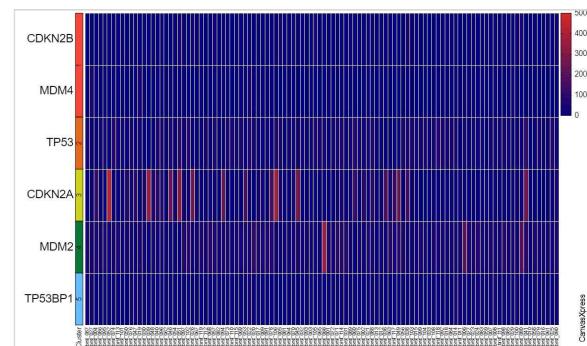
- . 시각화 가능한 오믹스 데이터 : Methylation, RNA, Proteome, Phosphorylation
- . Heatmap은 단백 유전체 정보를 색상으로 표현하여 일정한 이미지 위에 열(Heat) 분포 형태로 나타내는 시각화 기법으로, 패턴이 유사한 샘플을 클러스터링(Clustering)하여 한눈에 파악할 수 있습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자에 대한 Heatmap을 그릴 수 있습니다.
- . 클러스터링 알고리즘을 선택할 수 있습니다.
  - Heatmap : Hierarchical clustering(계층 분석)
  - K-mean : K-means clustering
- . [변수 선택]에서 입력데이터를 Raw count 혹은 Normalization 중 하나를 선택할 수 있습니다.
- . [Color] 및 [Spectrum]을 통해 Heatmap 색상 및 스케일을 조정할 수 있습니다.
- . Heatmap은 RNA, Methylation, Proteome, Phospho 네 가지 탭으로 구성됩니다.

#### 1) RNA Heatmap : RNA 유전자 발현 값을 히트맵 형태로 시각화

- [히트맵 종류 선택]은 RNA를 선택
- X축 : 샘플 ID
- Y축 : 선택한 유전자(세트)



〈 RNA - Heatmap 〉

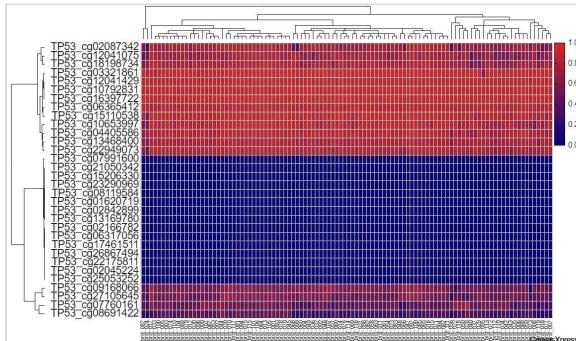


〈 RNA - K-mean 〉

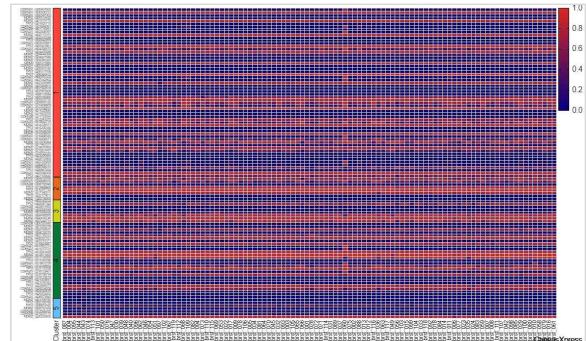
## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

### 2) Methylation Heatmap : DNA 메틸화 값을 히트맵 형태로 시각화

- [히트맵 종류 선택]은 Methylation을 선택
- X축 : 샘플 ID
- Y축 : 선택한 유전자(세트)의 Methylation site



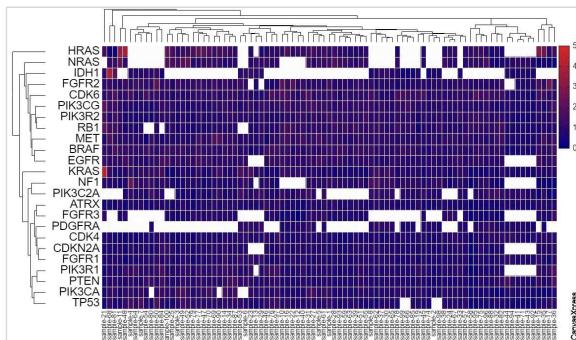
< Methylation - Heatmap >



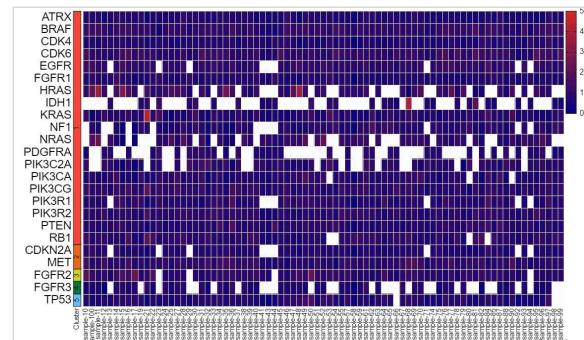
< Methylation - K-mean >

### 3) Proteome Heatmap : 단백질 풍부도를 히트맵 형태로 시각화

- [히트맵 종류 선택]은 Proteome을 선택
- X축 : 샘플 ID
- Y축 : 선택한 유전자 - Protein ID



< Proteome - Heatmap >

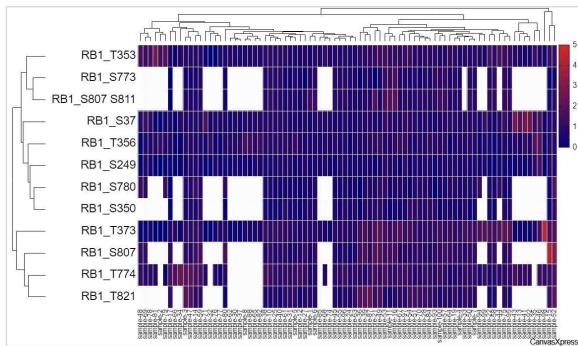


< Proteome - K-mean >

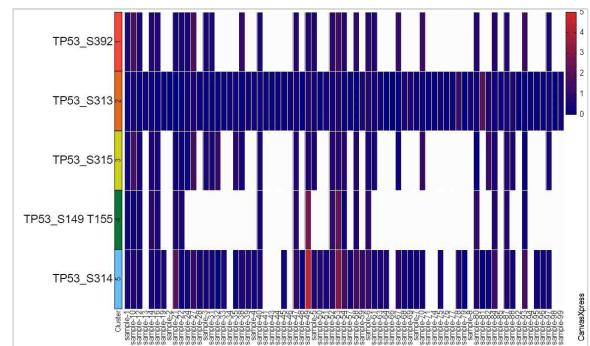
### 4) Phosphorylation Heatmap : 단백질 인산화를 히트맵 형태로 시각화

- [히트맵 종류 선택]은 Phosphorylation을 선택
- X축 : 샘플 ID
- Y축 : 선택한 유전자 - 인산화 위치(S, T, Y Site)

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)



< Phosphorylation - Heatmap >



< Phosphorylation - K-mean >

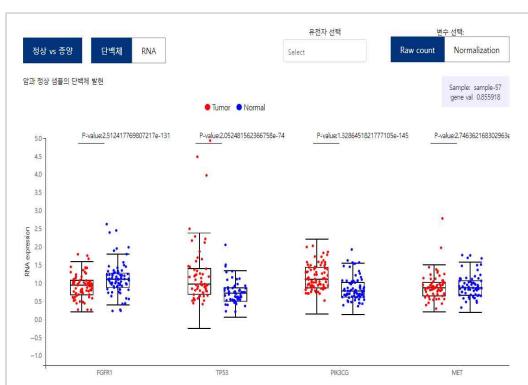
## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

### ⑥ Box Plot

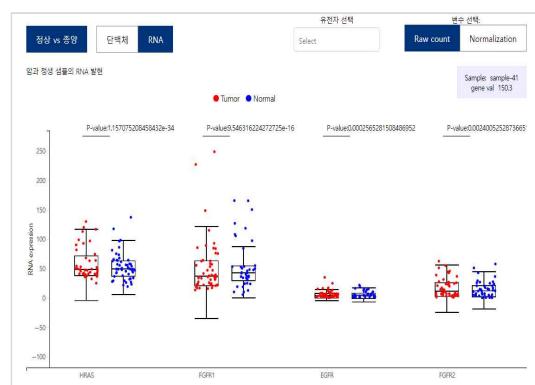
- . 시각화 가능한 오믹스 데이터 : RNA, Proteome
  - . Box plot은 선택된 유전자의 단백질 풍부도 또는 RNA 발현 값 데이터의 기본 통계를 Box plot 형태로 보여줍니다. 최솟값, 최댓값, 1 사분위 수, 3 사분위 수, 이상치(outlier) 등을 나타냅니다.
  - . 선택한 유전자에 대해 정상(Normal)과 종양(Tumor) 타입으로 나눠 시각화합니다.
  - . 유전자 세트 필터링한 데이터 중 [유전자 선택]을 통해 원하는 유전자만 선택할 수 있습니다.
  - . [변수 선택]을 통해 데이터를 Raw count 또는 Normalization 값으로 선택할 수 있습니다.
  - . Box plot은 단백체(Proteome)와 RNA 두 가지 탭으로 구성됩니다.
- 1) 단백체(Proteome) : Tumor 및 Normal 샘플에 대하여, 선택된 유전자의 단백질 풍부도를 Box plot 형태로 시각화 합니다.
- [플롯 종류]는 단백체(Proteome)을 선택
  - X축 : 선택한 유전자
  - Y축 : 선택한 유전자의 샘플 타입 별 단백질 발현 값

2) RNA : Tumor 및 Normal 샘플에 대하여, 선택된 유전자의 RNA 유전자 발현 값을 Box plot 형태로 시각화 합니다.

- [플롯 종류]는 RNA를 선택
- X축 : 선택한 유전자
- Y축 : 선택한 유전자의 샘플 타입 별 유전자



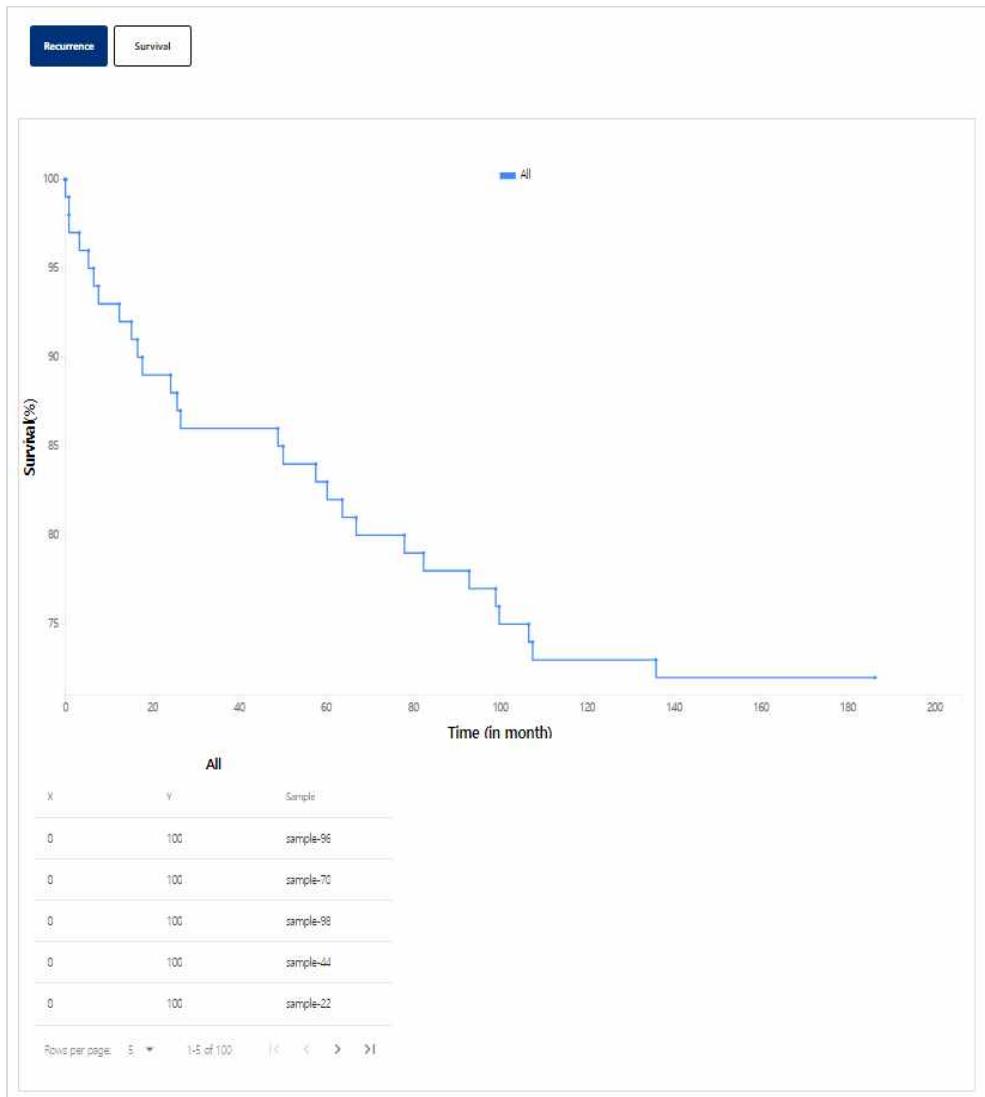
< Box plot - 단백체(Proteome) >



< Box plot - 전사체(RNA) >

### ⑦ Survival Plot

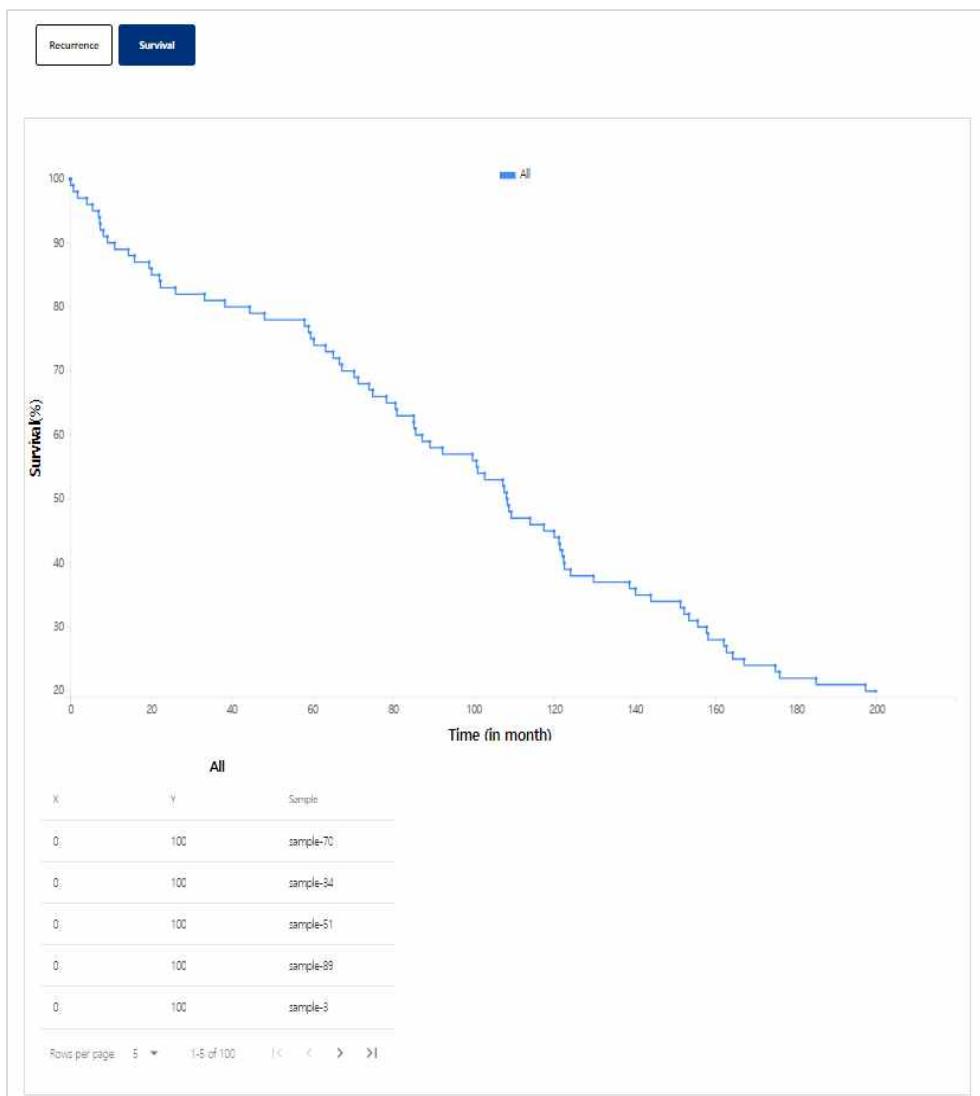
- . 시각화 가능한 오믹스 데이터 : Clinical Information
  - . Survival plot은 시간에 따른 환자들의 재발(Recurrence) 또는 생존(Survival) 확률을 보여주는 플롯입니다.
  - . 2가지 임상조건(Recurrence, Survival) 중 1가지를 선택하여 시각화 할 수 있습니다.
- 1) Recurrence : 시간에 따른 환자의 재발여부를 시각화합니다. 재발 확률을 보여주는 플롯을 그리기 위해 선 임상정보 파일에서 재발 여부 정보를 담고 있는 [rlps\_yn]와 재발이 확정되기까지 걸린 시간 정보인 [rlps\_cnfr\_drtm] 컬럼이 필수입니다.
- [플롯 종류]는 Recurrence를 선택
  - X축 : 시간 (Month)
  - Y축 : 전체 샘플에 대해 해당 기간동안 재발되지 않은 확률 (%)



< Survival plot - Recurrence >

2) Survival : 시간에 따른 환자의 생존여부를 시각화합니다. 생존 확률을 보여주는 플롯을 그리기 위해선 임상정보 파일에서 생존 여부 정보를 담고 있는 [death\_yn]와 생존시간 정보인 [death\_cnfr\_drtm] 컬럼이 필수입니다.

- [플롯 종류]는 Survival을 선택
- X축 : 생존 기간 (Month)
- Y축 : 전체 샘플에 대해 해당 기간동안 생존한 확률 (%)



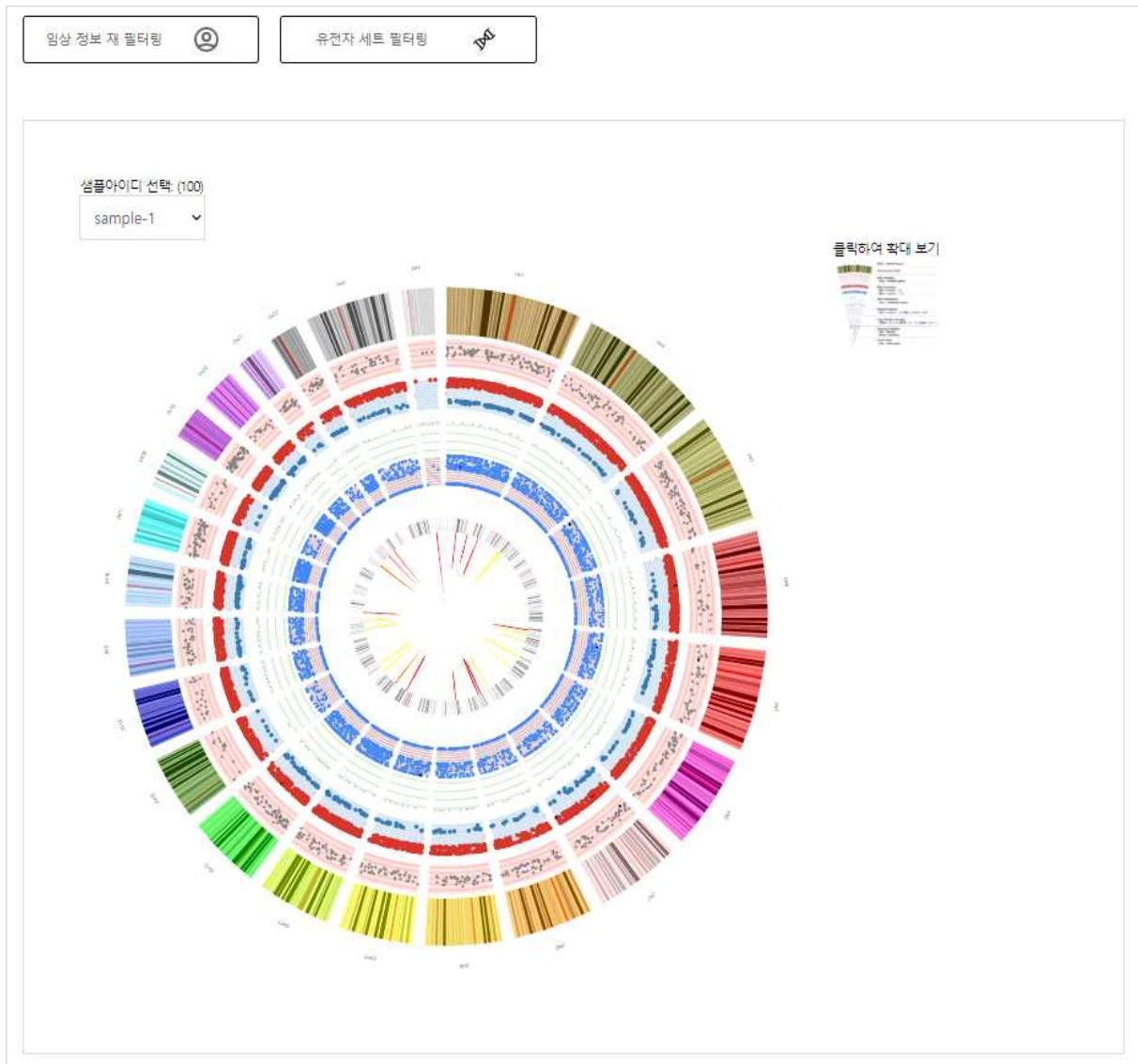
< Survival plot - Recurrence >

## 2. 다중 데이터 시각화 안내

### 시각화 가능한 도구 안내

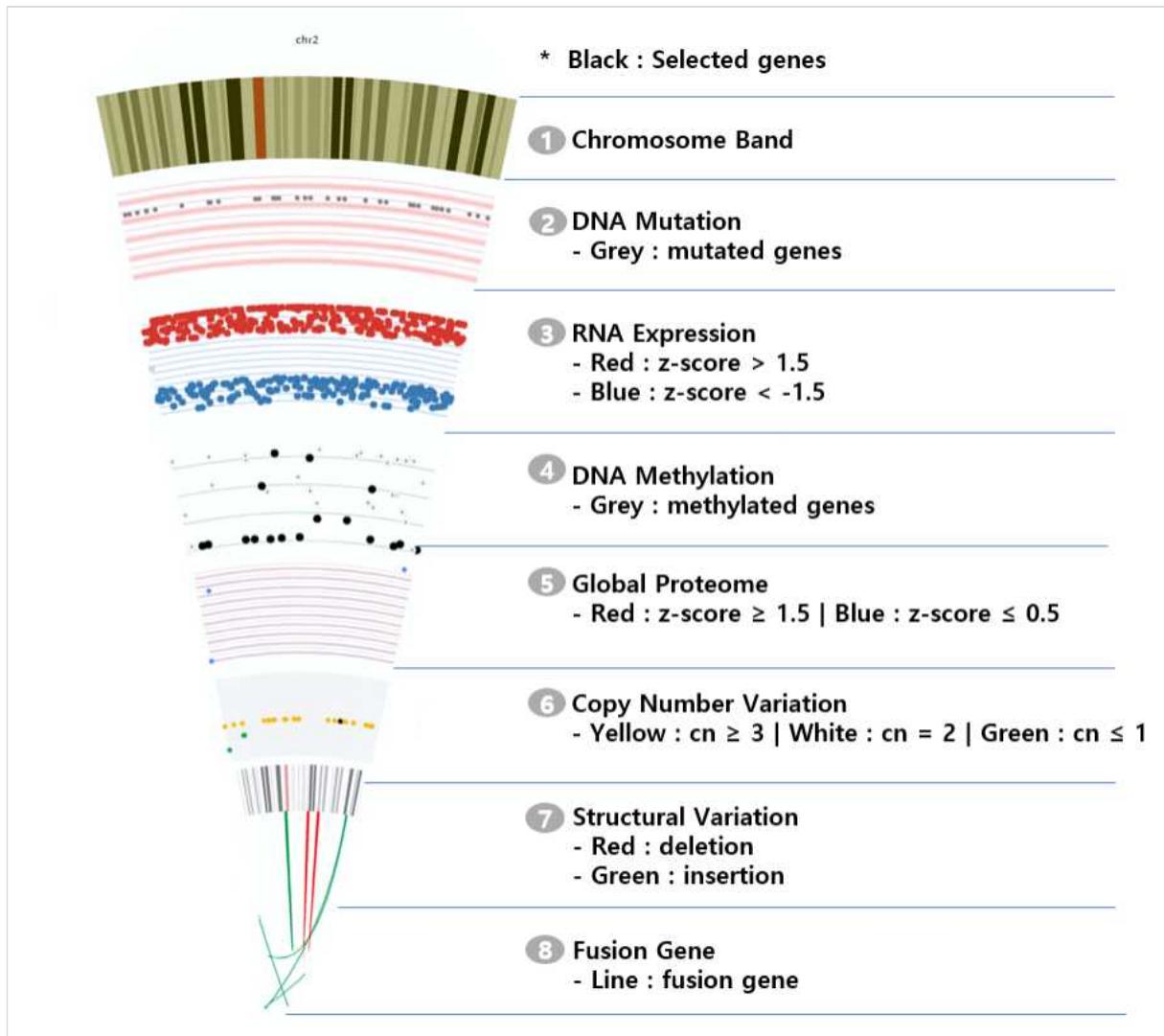
#### ① Circos Plot

- . 시각화 가능한 오믹스 데이터 : DNA Mutation, CNV, Methylation, RNA, Fusion, Proteome
- . Circos plot은 원형의 염색체 지도 위에 7가지 유전체 데이터를 다수의 레이어로 나타내는 플롯이며, 각 유전체 데이터를 각 레이어 위에 서로 다른 색상의 점이나 선으로 표시합니다.
- . Circos plot을 통해서 특정 환자의 모든 유전체 데이터를 한눈에 파악할 수 있습니다.
- . 사용자가 업로드한 오믹스 데이터에 대해서만 레이어가 표시됩니다.



## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

- . [임상 정보 재 필터링]을 통해 원하는 샘플 조건으로 필터링할 수 있습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자를 선택할 수 있습니다.
  - 선택한 유전자는 Circos Plot에서 검은색 점으로 표시됩니다.
- . [범례]에서 Circos plot 구성요소에 대한 정보를 확인할 수 있습니다.
  - 범례 그림을 누르면 팝업 안내를 통해 더 자세히 볼 수 있습니다.
- . Circos plot의 각 레이어는 다음과 같은 정보를 표시합니다.

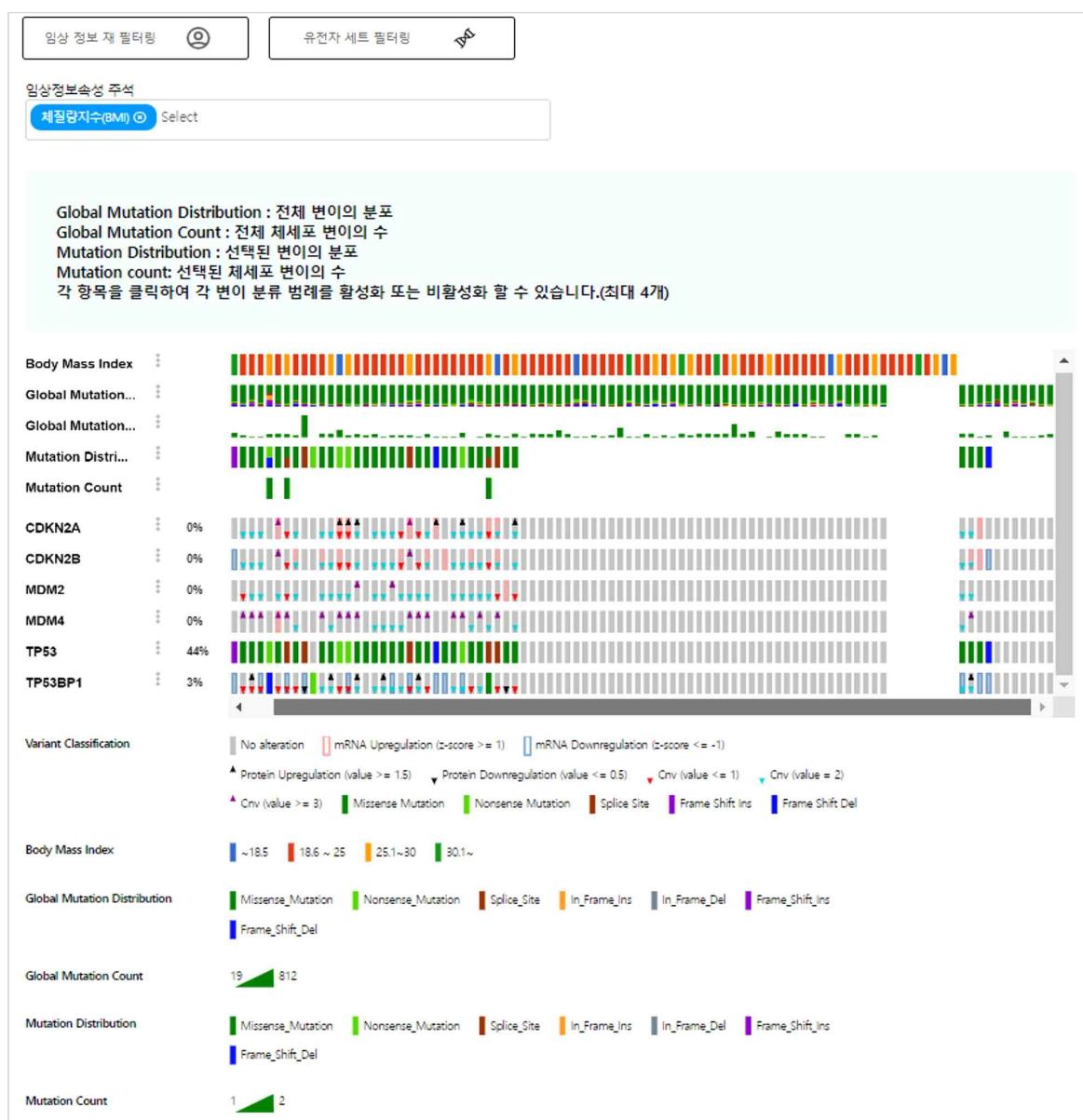


< Circos plot 범례 >

### ② OncoPrint

- . 시각화 가능한 오믹스 데이터 : Clinical Information, DNA Mutation
- . Oncoprint는 각 환자의 유전자 내 DNA 돌연변이와 여러 오믹스 정보를 칸, 색상, 기호 등으로 시각화 합니다.
- . Oncoprint를 통해 특정 유전자에 대한 환자들의 유전체 데이터의 통계를 파악할 수 있습니다.
  - X축 : 각 샘플 정보(하나의 샘플을 하나의 막대로 표현)
  - Y축 : 선택된 유전자

→ 하나의 막대그래프에 다양한 유전체 정보를 표시합니다.



〈 OncoPrint 〉

- . [임상 정보 재 필터링]에서 원하는 임상 정보 조건으로 환자를 필터링합니다.
  - . [유전자 세트 필터링]을 통해 원하는 유전자를 선택할 수 있습니다.
  - . [임상정보속성 주석] 상자를 통해 드롭다운 메뉴에서 원하는 임상 정보 레이블을 선택 가능합니다.
    - Oncoprint 맨 위에 해당 레이블이 생성
    - 하단 범례에 해당 임상 정보의 범례가 생성
  - . 변이와 관련된 네 가지 통계 정보를 플롯 상단 파란색 박스에 안내하고, 하단 범례에 해당 임상 정보 범례가 생성됩니다.
- 1) Global Mutation Distribution : 전체 변이의 분포
- 2) Global Mutation Counts : 전체 체세포 변이의 수
- 3) Mutation Distribution : 선택된 변이의 분포
- 4) Mutation Counts : 선택된 체세포 변이의 수
- . 7가지 주요 돌연변이 종류 : Missense Mutation, Nonsense Mutation, Splice Site, In Frame Insertion, In Frame Deletion, Frame Shift Insertion, Frame Shift Deletion

### ③ Lollipop Plot

- . 시각화 가능한 오믹스 데이터 : DNA Mutation, Phosphorylation
- . Lollipop plot을 통해 특정 유전자의 돌연변이나 인산화를 아미노산 서열 위에서 파악할 수 있습니다.
- . [임상 정보 재 필터링]을 통해 원하는 샘플 조건으로 필터링할 수 있습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자를 선택할 수 있습니다.
- . [유전자 선택] 드롭다운 메뉴에서 원하는 유전자 하나를 선택합니다.
- . Lollipop plot은 돌연변이(Mutation)와 인산화(Phosphorylation) 탭으로 구성됩니다.

#### 1) Mutation : 선택된 유전자의 아미노산 서열이 바뀐 돌연변이 위치, 도메인 정보, 개수를 시각화

- 선택된 유전자의 mRNA ID 목록을 표시 (ENST ID 목록, RefSeq ID 목록)
- X축 : 아미노산 위치 (aa)
- X축 색상 박스 : 도메인 이름 (마우스 오버 시, Full domain name 이름 볼 수 있음)
- Y축 : Mutation 개수 (Lollipop plot 머리에 마우스 오버 시, 개수 볼 수 있음)
- Lollipop 머리 색상 : Mutation 종류
- 체세포 변이 빈도(%) : 전체 샘플 중 선택한 유전자의 돌연변이를 보유한 샘플의 비율
- 범례 : Variant classification 색상 (= Lollipop plot 머리 색상)과 개수

#### 2) Phosphorylation : 선택된 유전자에 대해 단백질 상의 인산화 위치, 도메인 정보, 개수를 시각화

- X축 : 아미노산 위치 (aa)
- X축 색상 박스 : 도메인 이름 (마우스 오버 시, Full domain name 이름 볼 수 있음)
- Y축 : Phosphorylation site 개수 (Lollipop plot 머리에 마우스 오버 시, 개수 볼 수 있음)
- Lollipop 머리 색상 : Phosphorylation site 종류
- 범례 :
  - . Total Site : S, T, Y Site 색상 (= Lollipop plot 머리 색상)과 개수
  - . Major Site : 가장 많이 검출된 Site 이름과 수 (여러 개 가능)
- 인산화 빈도(%) : 전체 샘플 중 선택한 유전자의 인산화를 보유한 샘플의 비율
- Phosphorylation 표 : 각 Phosphophorylation의 샘플 ID, Site, 유전자

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)



< Lollipop Plot - 돌연변이 화면 >

< Lollipop Plot - 인산화 화면 >

### ④ Volcano Plot

- . 시각화 가능한 오믹스 데이터 : Clinical Information, RNA, Proteome
- . Volcano plot은 임상 정보 조건에 연관된 유의미한 바이오마커 유전자를 파악할 때 유용합니다.
- . Volcano plot은 임상 정보 조건에 따라 구분된 두 그룹 간에 발현 차이를 보이는 유의미한 유전자 (Differentially Expressed Genes, DEG)를 시각화합니다.
- . [유전자 세트 필터링]을 통해 선택한 유전자는 플롯에서 검은색 점으로 표시됩니다.
- . 설정한 조건에 대해 UP/DOWN level 유전자가 표시됩니다.
  - UP regulation level gene :  $\text{Log2FC} \geq 1.5$  & P-value  $\leq 0.05$  (**빨간색 점**)
  - Down regulation level gene :  $\text{Log2FC} \leq -1.5$  & P-value  $\leq 0.05$  (**파란색 점**)
- . Plot의 점에 마우스 오버 시, 우측 상단에 라벨이 표시됩니다.
- . [Download]를 통해 UP/DOWN level 유전자 세트 또는 전체 데이터 세트를 다운로드 받을 수 있습니다.
- . Volcano Plot은 데이터에 따라 전사체 또는 단백체 데이터로 시각화를 진행할 수 있습니다.

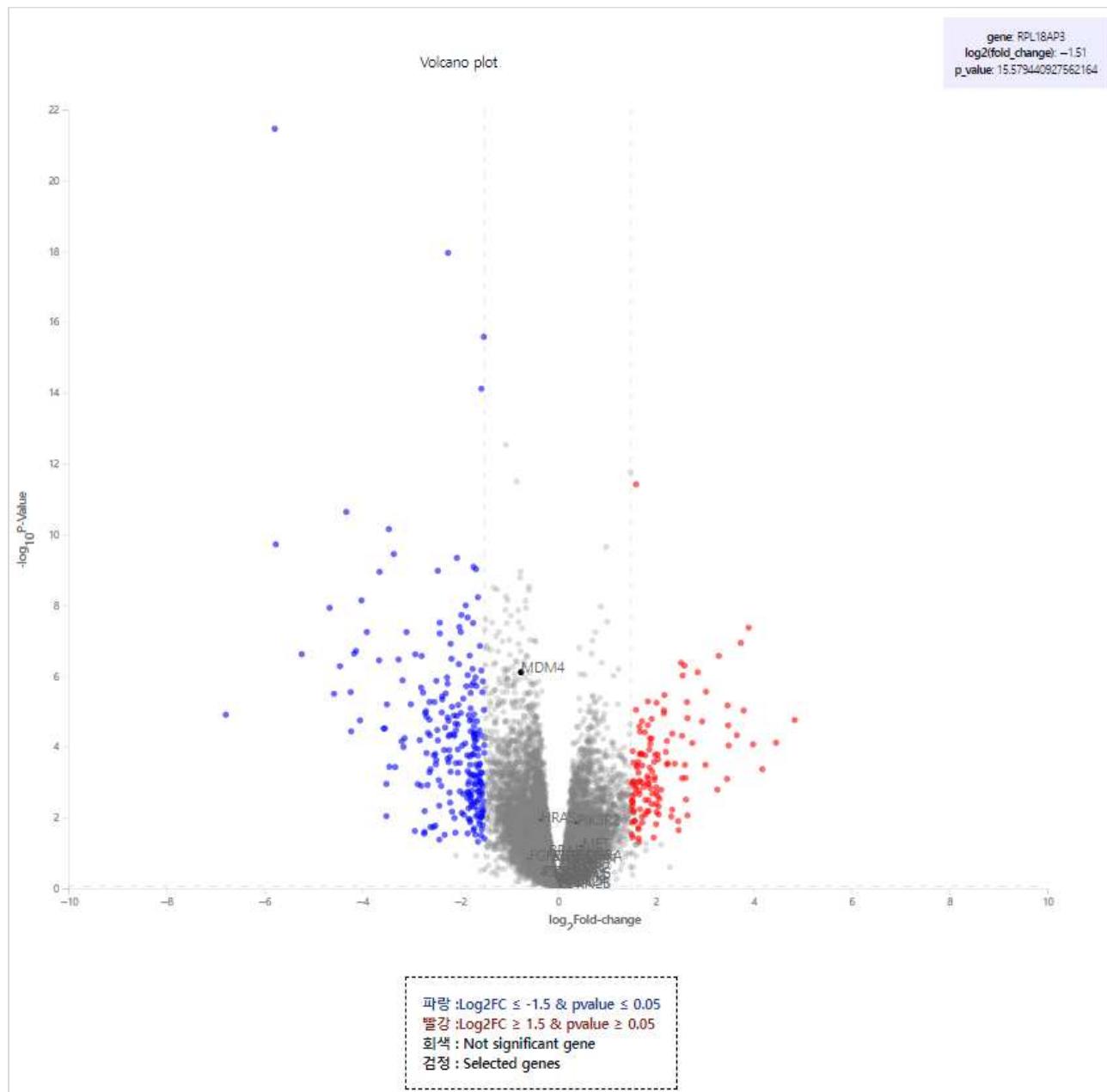
#### 1) 전사체(Transcriptome)

- [임상정보 그룹화]를 통해 [전사체] 데이터를 선택할 수 있습니다.
- [임상정보 그룹화] - [임상정보 속성]에서 원하는 임상정보를 두 그룹으로 지정해 데이터를 시각화합니다.
- 업로드한 임상정보 컬럼의 타입에 따라 다른 형태로 그룹 지정을 합니다.
  - . 연속형 변수 : 선택한 임상정보의 최대, 최소 값 범위 내에서 사용자 지정
  - . 범주(그룹)형 변수 : 각각의 데이터를 Group A, B로 사용자 지정
  - . 부울형 변수 : TRUE / FALSE에 따라 자동으로 두 그룹 지정

#### 2) 단백체(Proteome)

- [임상정보 그룹화]를 통해 [단백체] 데이터를 선택할 수 있습니다.
- 샘플 타입을 설정해 분석을 진행합니다.
  - . 정상(Normal) : 업로드한 샘플의 정상(Normal, N) 타입에 대해 분석 수행
  - . 종양(Tumor) : 업로드한 샘플의 종양(Tumor, T) 타입에 대해 분석 수행
  - . 정상 vs 종양 : 정상세포와 종양세포 간 비교 분석 수행
- 샘플 타입을 [정상] 또는 [종양]으로 선택한 경우 전사체 데이터와 마찬가지로 원하는 임상정보를 두 그룹으로 지정해 데이터를 시각화합니다.
- 샘플 타입을 [정상 vs 종양]으로 선택한 경우 이미 샘플 타입에 대해 그룹이 지정되었기 때문에 임상 정보 설정 과정 없이 정상 및 종양 그룹에 대해 바로 시각화합니다.

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)



< Volcano Plot 화면 >

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

---

	A	B	C
1	GENE NAME	LOG2FC	LOG(PVALUE) negative
2	DLK1	-5.8	0.0000168
3	NTN4	-1.5	0.0000172
4	LINC00578	-3.09	0.0000194
5	SLC40A1	-1.68	0.0000813
6	C8orf34	-3.76	0.0000923
7	CYP4Z1	-3.95	0.0001
8	CA2	-2.58	0.00011
9	MSL1	-1.5	0.00012
10	SCGB1D2	-3.6	0.00015
11	CITED1	-2.77	0.0002
12	SSTR2	-1.85	0.00024
13	SCGB2A2	-3.54	0.00024
14	RP11-114M1.1	-3.85	0.00025
15	CBLN2	-3.67	0.00025
16	HPGD	-2.9	0.00026
17	PTGER3	-1.58	0.00027
18	ZBTB16	-1.84	0.00028
19	KLK12	-4.17	0.00029
20	SCIN	-1.86	0.00039
21	RP13-401N8.3	-1.6	0.00039
22	IGF2	-2	0.00039
23	C2orf40	-2.47	0.0004
24	MYBPC1	-3.37	0.00045
25	SLC26A3	-3.29	0.00051
26	GRIK1	-3.61	0.00056
27	KLK11	-3	0.00084
28	IGLV4-3	-4.66	0.00085
29	CXCL13	-2.56	0.00087
30	RP11-143K11.1	-1.8	0.00089

< Down level gene 다운로드 파일 >

	A	B	C
1	GENE NAME	LOG2FC	LOG(PVALUE) negative
2	NOTUM	4.72	2.59E-18
3	SCG2	3.59	2.33E-16
4	SNORD3A	4.33	6.00E-14
5	DKK4	5.5	2.75E-13
6	CHRDL2	3.99	1.23E-12
7	COL9A1	4.27	1.88E-11
8	SNAI1	2.24	2.98E-11
9	MPZ	2.98	1.17E-10
10	RP11-329B9.4	2.66	5.73E-10
11	HLA-G	3.7	6.99E-10
12	AC016739.2	2.32	1.27E-09
13	TNNI2	2.59	1.33E-09
14	IGKV1-33	3.39	2.94E-09
15	SFTP D	2.73	5.91E-09
16	CHGB	5.36	6.31E-09
17	C6orf15	3.37	6.90E-09
18	RP1-111B22.3	2.91	1.07E-08
19	RUND C3A	2.7	1.27E-08
20	RP11-351J23.1	3.6	2.13E-08
21	TF	2.43	2.22E-08
22	RP11-689K5.3	2.46	2.99E-08
23	HTN1	8.23	3.36E-08
24	SLC38A3	2.45	4.20E-08
25	LINC00707	4.5	7.15E-08
26	RP11-758M4.4	3.21	7.96E-08
27	RP11-329B9.3	2.43	1.01E-07
28	PTGFR	2.5	1.06E-07
29	ACHE	2.14	1.15E-07
30	FAM180A	1.72	1.20E-07

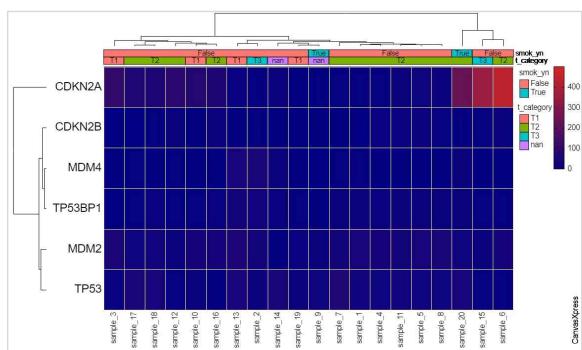
< UP level gene 다운로드 파일 >

### ⑤ Heatmap

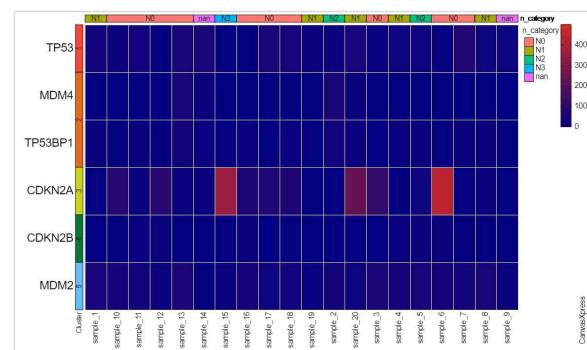
- . 시각화 가능한 오믹스 데이터 : Clinical Information, Methylation, RNA, Proteome, Phosphorylation
- . Heatmap은 단백 유전체 정보를 색상으로 표현하여 일정한 이미지 위에 열(Heat) 분포 형태로 나타내는 시각화 기법으로, 패턴이 유사한 샘플을 클러스터링(Clustering)하여 한눈에 파악할 수 있습니다.
- . [임상 정보 재 필터링]을 통해 원하는 샘플 조건으로 필터링할 수 있습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자에 대한 Heatmap을 그릴 수 있습니다.
- . 클러스터링 알고리즘을 선택할 수 있습니다.
  - Heatmap : Hierarchical clustering(계층 분석)
  - K-mean : K-means clustering
- . [임상정보속성 주석]에서 선택한 임상정보에 대해 히트맵에 주석을 나타낼 수 있습니다. 여러개의 임상정보를 선택해 나타내는 것이 가능합니다.
- . [변수 선택]에서 입력데이터를 Raw count 혹은 Normalization 중 하나를 선택할 수 있습니다.
- . [Color] 및 [Spectrum]을 통해 Heatmap 색상 및 스케일을 조정할 수 있습니다.
- . Heatmap은 RNA, Methylation, Proteome, Phospho 네 가지 탭으로 구성됩니다.

#### 1) RNA Heatmap : RNA 유전자 발현 값을 히트맵 형태로 시각화

- [히트맵 종류 선택]은 RNA를 선택
- X축 : 샘플 ID
- Y축 : 선택한 유전자(세트)



〈 RNA - Heatmap 〉

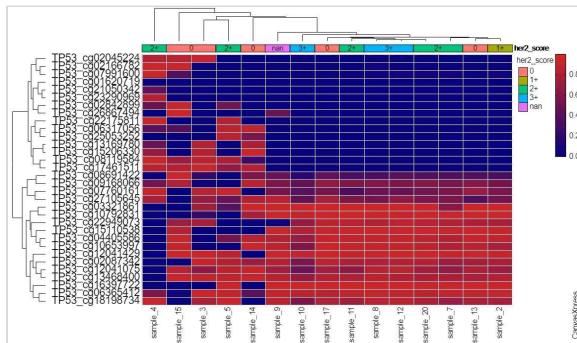


〈 RNA - K-mean 〉

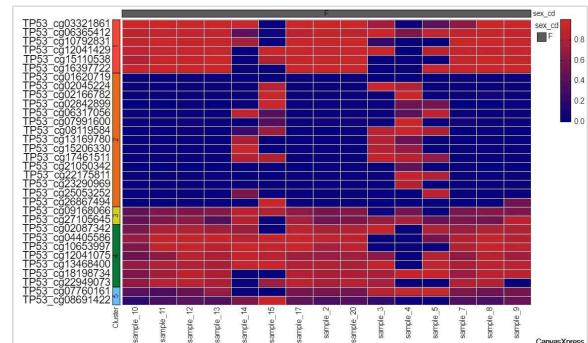
## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

### 2) Methylation Heatmap : DNA 메틸화 값을 히트맵 형태로 시각화

- [히트맵 종류 선택]은 Methylation을 선택
- X축 : 샘플 ID
- Y축 : 선택한 유전자(세트)의 Methylation site



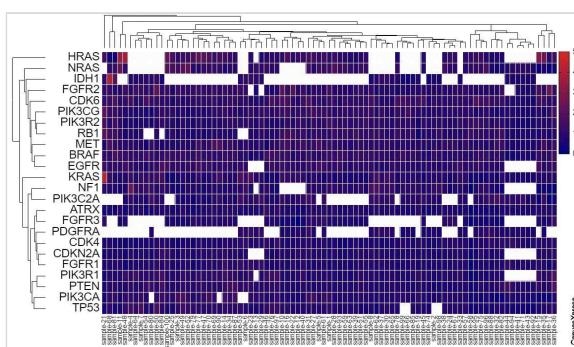
< Methylation - K-mean >



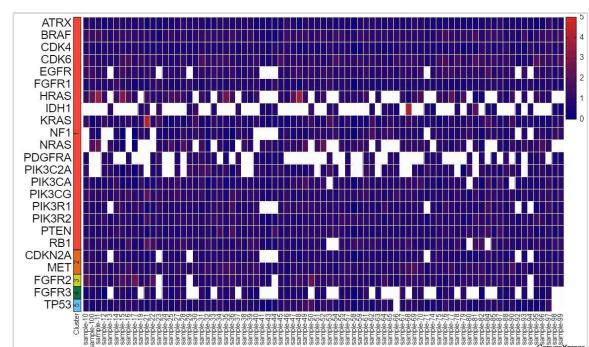
< Methylation - K-mean >

### 3) Proteome Heatmap : 단백질 풍부도를 히트맵 형태로 시각화

- [히트맵 종류 선택]은 Proteome을 선택
- X축 : 샘플 ID
- Y축 : 선택한 유전자 - Protein ID



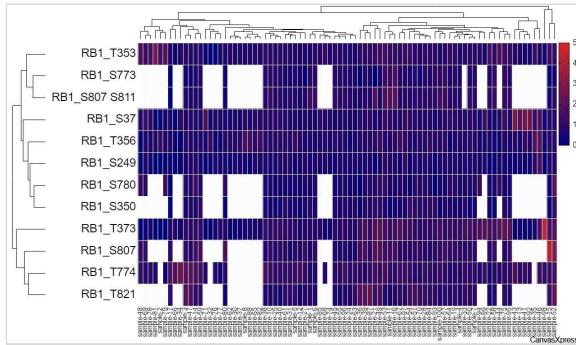
< Proteome - Heatmap >



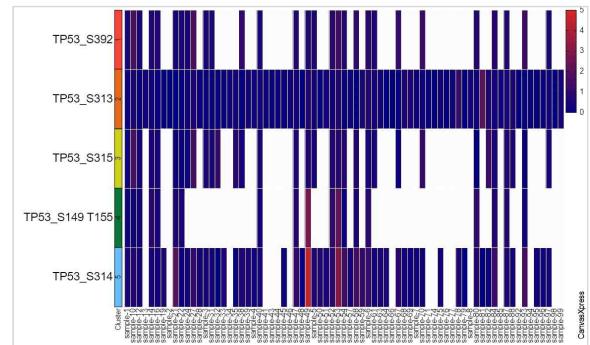
< Proteome - K-mean >

4) Phosphorylation Heatmap : 단백질 인산화를 히트맵 형태로 시각화

- [히트맵 종류 선택]은 Phosphorylation을 선택
- X축 : 샘플 ID
- Y축 : 선택한 유전자 - 인산화 위치(S, T, Y Site)



< Phosphorylation - Heatmap >

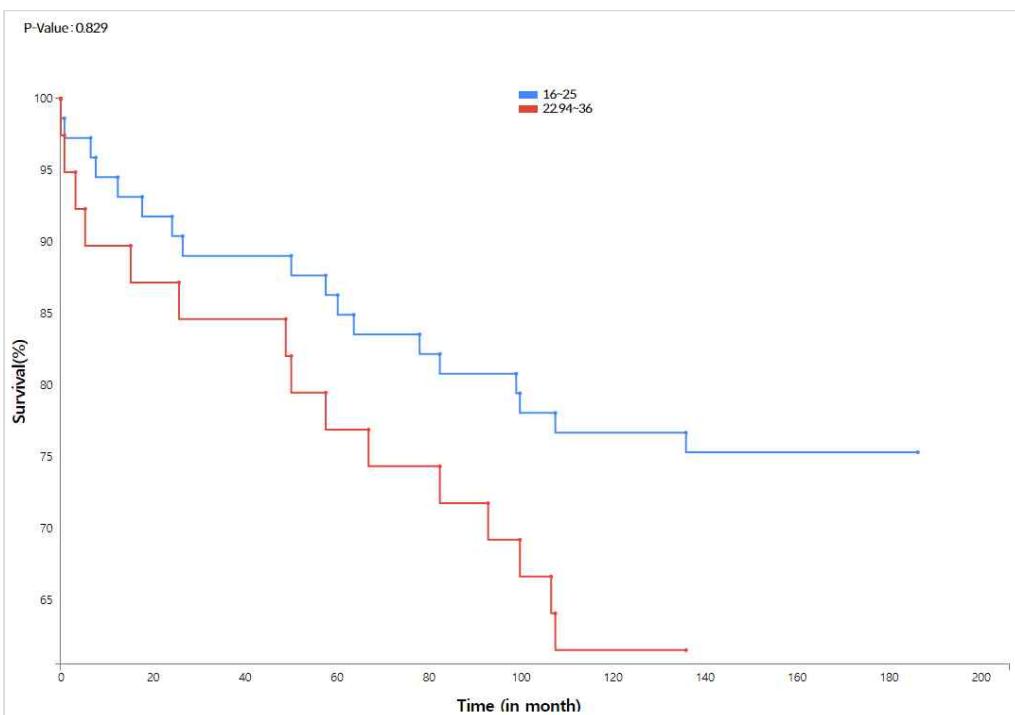


< Phosphorylation - K-mean >

### ⑥ Survival Plot

- . 시각화 가능한 오믹스 데이터 : Clinical Information, DNA Mutation, RNA, Proteome
  - . Survival plot은 시간에 따른 환자들의 재발(Recurrence) 또는 생존(Survival) 확률을 보여주는 플롯입니다.
  - . 두 가지 통계 모델(Recurrence or Survival, Cox Regression) 중 하나를 선택하여 시각화 할 수 있습니다.
- 1) Recurrence or Survival : 시간에 따른 환자의 재발/생존 확률을 시각화
- 임상정보 혹은 유전체 정보 조건으로 환자군을 나누어 시간에 따른 재발/생존률 변화를 시각화
  - 환자군 간의 재발/생존률을 비교하여 암(질병) 재발과 연관된 중요한 임상정보 혹은 유전체 바이오마커 식별 가능
  - [임상정보 그룹화] - [모델선택]에서 Recurrence 또는 Survival 모델을 선택
  - 재발확률 시각화(Recurrence)를 위해선 임상정보 파일에서 재발여부(rlps\_yn), 재발시간(rlps\_cnfr\_drtn) 컬럼이 필수
  - 생존확률 시각화(Survival)를 위해선 임상정보 파일에서 생존여부(death\_yn), 생존시간(death\_cnfr\_drtn) 컬럼이 필수
  - 정보 타입에 따라 임상정보 또는 오믹스정보를 통해 시각화 가능
    - . 임상정보 : 선택한 임상정보의 그룹 지정을 통해 재발/생존 정보를 시각화
      - 연속형 변수 : 선택한 임상정보의 최대, 최소 값 범위 내에서 사용자 지정
      - 범주(그룹)형 변수 : 환자 그룹의 구성원을 직접 선택하여 각 그룹을 지정
      - 부울형 변수 : TRUE / FALSE에 따라 자동으로 두 그룹 지정
    - . 오믹스정보 : 선택한 유전자의 DNA Mutation, RNA Expression, Proteome Expression 정보와 샘플의 재발/생존 정보를 이용해 시각화 (선택한 유전자의 오믹스 정보가 없다면 시각화 되지 않음)
      - DNA Mutation : 7가지 주요 변이 중 하나라도 선택한 유전자에 있는 샘플과 그렇지 않은 샘플로 그룹이 지정되어 시각화
      - RNA Expression : 선택한 유전자의 RNA expression UP/DOWN regulation level에 따라 샘플 그룹이 지정되어 시각화 (UP level : z-score  $\geq 1.5$ , DOWN level : z-score  $\leq -1.5$ )
      - Proteome Expression : 선택한 유전자의 Proteome expression UP/DOWN regulation level에 따라 샘플 그룹이 지정되어 시각화 (UP level : z-score  $\geq 1.5$ , DOWN level : z-score  $\leq 0.5$ )
  - X축 : 시간 (Month)
  - Y축 : 전체 샘플에 대해 해당 기간동안 재발되지 않은 확률 또는 생존 확률 (%)
  - 각 그룹에 대한 범례는 우측 상단에 표시
  - 각 그룹에 있는 샘플의 시간에 따른 재발/생존 정보는 플롯 하단의 테이블에 안내

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)



< Survival Plot - 생존곡선 >

16~25

Total: 73

시간(월)	재발되지 않은 확률	Sample
0	100	sample-2
0	100	sample-4
0	100	sample-7
0	100	sample-10
0	100	sample-11
0	100	sample-13
0	100	sample-16
0	100	sample-20
0	100	sample-22
0	100	sample-23

<< < 1 2 3 4 > >>

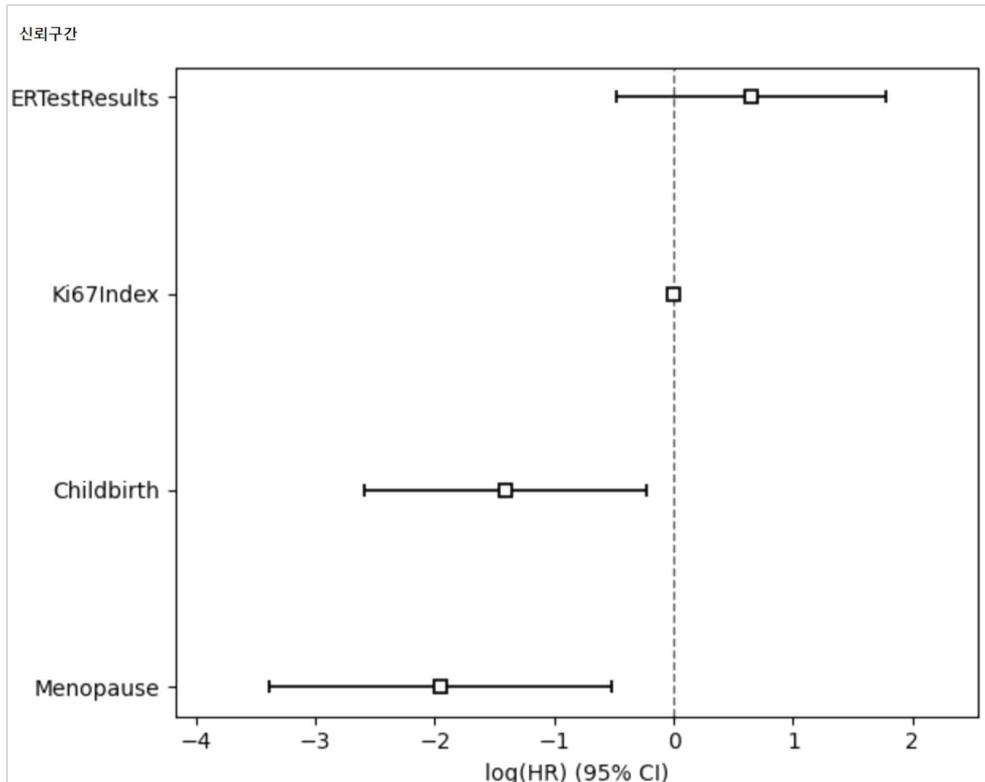
< Survival Plot - 그룹 정보 테이블>

2) Cox Regression : 특정 사건(예. 환자의 죽음 혹은 암 재발)이 발생하는데 걸리는 시간에 대한 여러 변수의 영향을 시각화

- Cox Regression 분석을 통해 생존률에 영향을 미치는 위험인자 식별 가능
- [임상정보 그룹화] - [모델선택]에서 Cox Regression 모델을 선택
- 업로드한 임상정보에 대해 시각화
  - . [임상정보속성]에서 원하는 임상변수 선택 가능(단일, 다중, 전체 선택 모두 가능)
- 선택한 임상정보에 대한 상관계수표(Co-efficient Table)와 신뢰구간(Confidence Interval Plot, 95% CI)에 대한 정보 표시

상관계수표										
	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) l ower 95%	exp(coef) u pper 95%	z	p	-log2(p)
Menopause	-1.96	0.14	0.73	-3.39	-0.52	0.03	0.59	-2.67	0.01	7.04
Childbirth	-1.41	0.24	0.60	-2.59	-0.23	0.07	0.80	-2.34	0.02	5.69
ERTestResults	0.65	1.91	0.57	-0.48	1.77	0.62	5.88	1.12	0.26	1.94
Ki67Index	-0.00	1.00	0.01	-0.02	0.01	0.98	1.01	-0.34	0.73	0.45

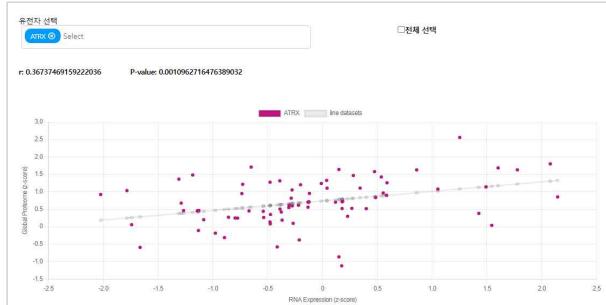
〈 Cox Regression. Co-efficient table 〉



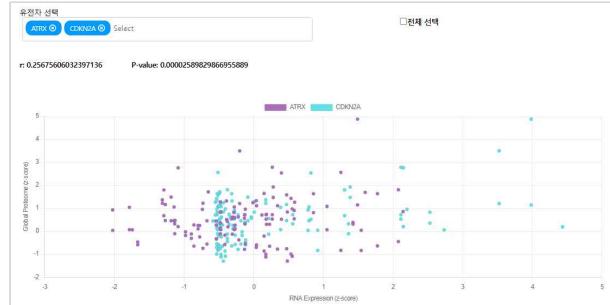
〈 Cox Regression. Confidence Interval, 95% CI 〉

### ⑦ Correlation

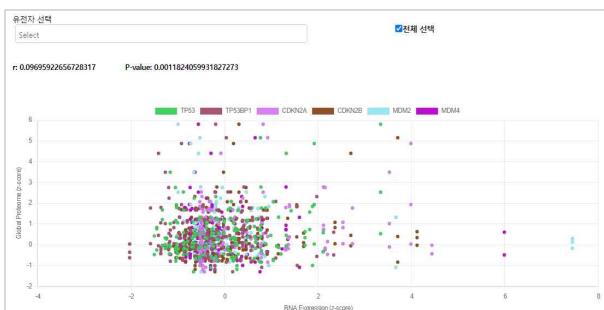
- . 시각화 가능한 오믹스 데이터 : RNA, Proteome
- . Correlation Plot은 선택된 유전자에 대하여 RNA 발현 값과 Proteome 풍부도 값의 상관관계를 점그라프(Scatter Plot) 형태로 시각화합니다.
- . Correlation Plot을 통해 특정 유전자에 대한 RNA 발현과 단백질 발현 간의 상관관계 확인할 수 있습니다.
- . [임상 정보 재 필터링]을 통해 원하는 샘플 조건으로 필터링할 수 있습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자에 대한 Correlation Plot을 그릴 수 있습니다.
- X축 : RNA Expression (z-score)
- Y축 : Global Proteome (Z-score)
- 그래프 상단에 R squared, P-value 표시
- [유전자 선택]에서 원하는 유전자 선택 가능(여러 개 또는 전체 유전자 선택 가능)
- Regression line 표시 (하나의 유전자만 선택했을 경우 표시)



< Correlation. 단일 유전자 선택 >



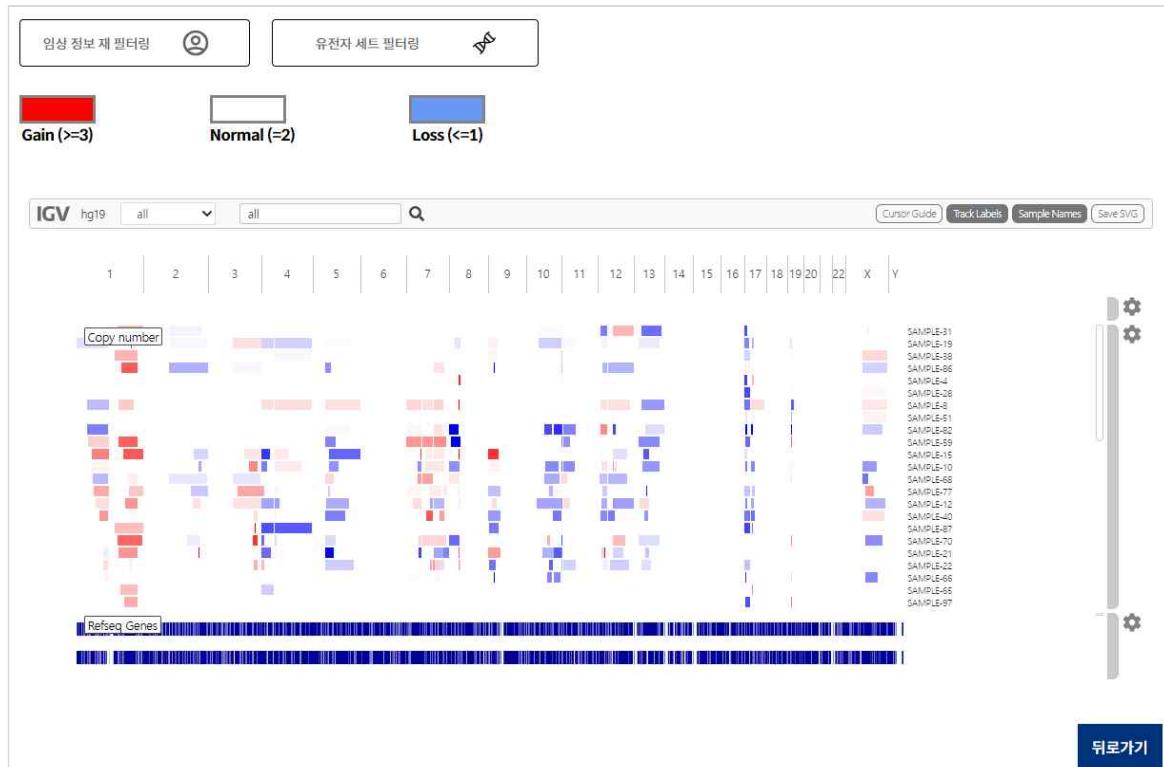
< Correlation. 다중 유전자 선택 >



< Correlation. 필터링 유전자 전체 선택 >

### ⑧ CNV

- . 시각화 가능한 오믹스 데이터 : Clinical Information, CNV
- . CNV plot은 복제 수 변이(Copy Number Variation) 데이터를 Zoom-in, Zoom-out 가능한 유전체 뷰어에서 시각화합니다.
- . CNV plot을 통해 전체 복제 수 변이 데이터를 시각적으로 파악할 수 있습니다.



< CNV >

- . [임상 정보 재 필터링]을 통해 원하는 샘플 조건으로 필터링할 수 있습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자에 대한 CNV Plot을 그릴 수 있습니다.
- . IGV(Integrated Genome Viewer)에서 염색체와 위치(base-pair) 범위를 검색하면 해당 범위를 확대하여 시각화합니다.
- . Copy number 창
  - X축 : 염색체 및 위치(\*확대 시, base 단위로 볼 수 있음)
  - Y축 : 샘플 이름
  - 붉은색 : Copy number gain ( $CN \geq 3$ )
  - 흰색 : Copy number normal ( $CN = 2$ )
  - 푸른색 : Copy number loss ( $CN \leq 1$ )

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

- . RefSeq Genes 트랙 (\*확대 시, 상세 보기 가능)
- . 옵션 기능들
  - Cursor Guide : 마우스 커서 위치 강조
  - Track labels : 트랙 레이블 보이게 설정
  - Sample names : 샘플 이름 보이게 설정
  - Save SVG : SVG 형식 다운로드
- . IGV(Integrated Genome Viewer)에서 염색체와 위치(base-pair) 범위를 검색하면 해당 범위를 확대하여 시각화합니다.

예시) IGV의 첫 번째 상자에서 드롭 다운 기능을 통해 원하는 위치의 염색체로 이동 가능

IGV의 두 번째 상자에서 [chr번호]:[위치 범위]를 통해 해당 위치로 이동 가능 (chr10:87800-87971)



< CNV, Chromosome - Position 설정 >

- X축 : 염색체 밴드 (Ideogram) 및 염기서열
- Y축 : 샘플 이름
- 마우스로 화면을 왼쪽, 오른쪽으로 드래그하여 이동 가능

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

### ⑨ Box Plot

- . 시각화 가능한 오믹스 데이터 : RNA, Proteome
  - . Box plot은 선택된 유전자의 단백질 풍부도 또는 RNA 발현 값 데이터의 기본 통계를 Box plot 형태로 보여줍니다. 최솟값, 최댓값, 1 사분위 수, 3 사분위 수, 이상치(outlier) 등을 나타냅니다.
  - . [임상 정보 재 필터링]을 통해 원하는 샘플 조건으로 필터링할 수 있습니다.
  - . [유전자 세트 필터링]을 통해 원하는 유전자에 대한 Box Plot을 그릴 수 있습니다.
  - . 유전자 세트 필터링한 데이터 중 [유전자 선택]을 통해 원하는 유전자만 선택할 수 있습니다.
  - . 선택한 유전자에 대해 정상(Normal)과 종양(Tumor) 타입으로 나눠 시각화합니다.
  - . [변수 선택]을 통해 데이터를 Raw count 또는 Normalization 값으로 선택할 수 있습니다.
  - . Box plot은 단백체(Proteome)와 RNA 두 가지 탭으로 구성됩니다.
- 1) 단백체(Proteome) : Tumor 및 Normal 샘플에 대하여, 선택된 유전자의 단백질 풍부도를 Box plot 형태로 시각화
- [플롯 종류]는 단백체(Proteome)을 선택
  - X축 : 선택한 유전자
  - Y축 : 선택한 유전자의 샘플 타입 별 단백질 발현 값



< Box plot - 단백체(Proteome) >

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

2) RNA : Tumor 및 Normal 샘플에 대하여, 선택된 유전자의 RNA 유전자 발현 값을 Box plot 형태로 시각화

- [플롯 종류]는 RNA를 선택
- X축 : 선택한 유전자
- Y축 : 선택한 유전자의 샘플 타입 별 유전자

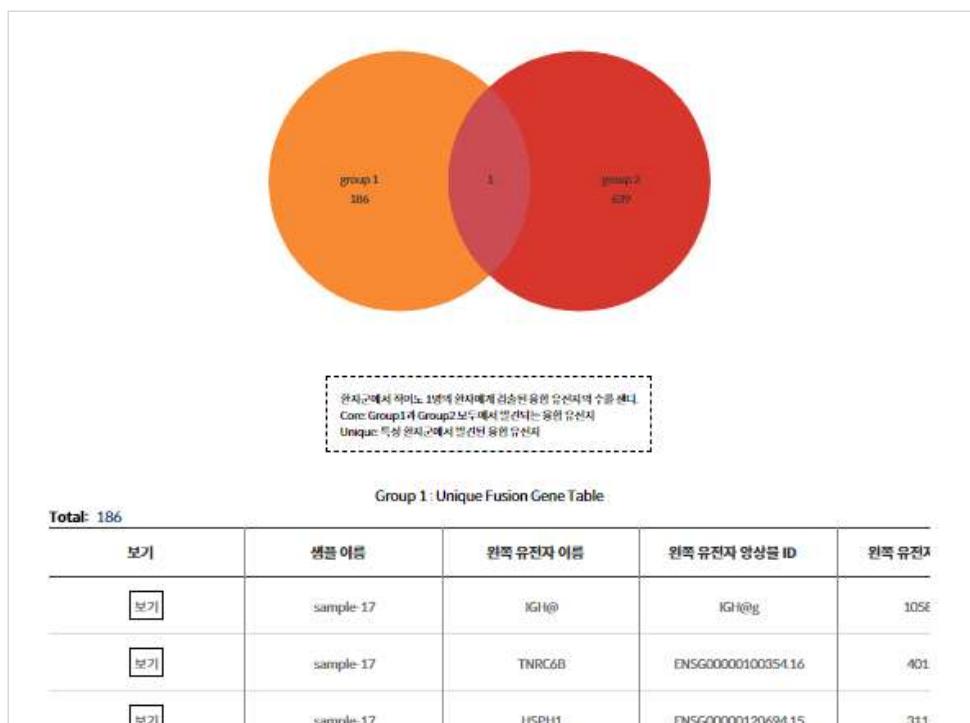


< Box plot - 전사체(RNA) >

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

### ⑩ Fusion Plot

- . 시각화 가능한 오믹스 데이터 : Clinical Information, Fusion
- . Fusion plot은 선택된 샘플 그룹에 대한 Fusion genes 개수 정보를 벤 다이어그램으로 나타내고, 개별 Fusion gene을 시각화합니다.
- . [임상정보 그룹화]를 통해 사용자가 원하는 임상 정보를 선택하여 샘플에 대해 지정된 그룹으로 나눌 수 있습니다.
  - 연속형 변수 : 사용자가 [Add group] 버튼으로 최대 3개 그룹까지 그룹을 추가하여 값 범위를 직접 입력해 지정
  - 범주형 변수 : 사용자가 환자 그룹의 구성원을 체크박스를 통해 직접 선택하여 각 그룹을 지정
  - 부울형 변수 : YES(TRUE) or NO(FALSE) 데이터에 따라 자동으로 그룹 지정
- . [유전자 세트 필터링]을 통해 원하는 유전자에 대한 Fusion Plot을 그릴 수 있습니다.
- . 그룹 지정을 완료하면 벤 다이어그램과 fusion gene 목록이 화면에 나타납니다.



< Fusion plot - 벤다이어그램 시각화 >

- . 벤 다이어그램의 각 영역을 [클릭]하면 하단에서 해당하는 fusion gene 목록을 볼 수 있습니다.
  - group1 : 그룹 1 샘플만 지닌 fusion gene 개수
  - group2 : 그룹 2 샘플만 지닌 fusion gene 개수
  - 교집합 영역 : 그룹 1, 그룹 2의 환자가 공통으로 지닌 fusion gene 개수

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

→ 영역 구분 기준 : Left Gene Name, Right Gene Name, Left Breakpoint, Right Breakpoint가 일치 할 시, 공통된 fusion gene으로 계수

- . 각 영역을 클릭하면 해당 영역의 fusion gene 목록 테이블을 볼 수 있습니다.

- Sample Name : 샘플 이름
- Left Gene Name : fusion gene을 구성하는 왼쪽 유전자 이름
- Left Ensembl ID : 왼쪽 유전자의 Ensembl ID
- Left Breakpoint : 왼쪽 유전자가 잘리는(융합되는) 지점
- Right Gene Name : fusion gene을 구성하는 왼쪽 유전자 이름
- Right Ensembl ID : 왼쪽 유전자의 Ensembl ID
- Right Breakpoint : 왼쪽 유전자가 잘리는(융합되는) 지점
- Junction Read Count : Splice junction에 존재하는 read 수
- Spanning Frage Count : Spanning read 수

→ Spanning read : single-end 또는 paired-end reads 중 하나의 read가 fusion gene의 두 유전자에 일부분씩 정렬된 것(split)

- Splice Type : 변이 타입

- . Fusion gene 목록 테이블에서 View 컬럼의 [View] 버튼을 통해 각 fusion gene의 그래프를 볼 수 있습니다.



< Fusion plot - Fusion gene 시각화 >

- Left gene transcript ID, Right gene transcript ID의 각 드롭다운 메뉴에서 원하는 transcript를

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

---

---

선택하여 시각화

- . 첫 번째 줄 : 엑손(exon) 영역은 색상 박스로 표시되며, Left/Right gene transcript에 따라 다른 색상으로 나타냄(각 엑손 영역에 마우스 오버 시 exon 번호 확인 가능)
- . 두 번째 줄 : Left/Right gene transcript의 breakpoint 기준으로 바깥쪽 영역이 융합되는 모습이 표시
- . 세 번째 줄 : 융합된 유전자에서 인트론(intron)을 제외한 엑손(exon) 영역만 표시

### ⑪ Sankey Plot

- . 시각화 가능한 오믹스 데이터 : Clinical Information
- . Sankey 시각화 도구를 통해 선택된 샘플의 질병 연계 정보, 약물 연계 정보를 볼 수 있습니다.
- . [임상 정보 재 필터링]을 통해 원하는 샘플 조건으로 필터링할 수 있습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자를 선택할 수 있습니다.
- . Sankey plot은 환자 요약(Patient Summary)와 약물 연계 정보 플롯(Drug Relation Plot) 탭으로 구성됩니다.

1) 환자 요약(Patient Summary) : 선택한 샘플에 대한 임상정보와 선택한 유전자에 대한 해당 샘플의 유전 정보를 테이블 형태로 표시

- [샘플아이디 선택] 박스를 통해 샘플 선택
- 사용자 임상 정보에 대한 선택한 샘플 정보를 Basic Information 테이블에 표시
- 선택한 유전자에 대한 해당 샘플의 DNA Mutation, RNA, Proteome 정보를 Genomic Information 테이블에 표시

a. DNA Mutation : 사용자가 선택(필터링)한 유전자에 대해 주요 7가지 변이에 대해 해당되는 개수를 표시

예시) 7가지 주요 변이에 대해 3가지가 해당하는 경우 [DNA Mutation - YES] 컬럼에 O(3)로 표시되고, 표시된 곳에 마우스를 올리면 해당하는 돌연변이 타입이 나타남

→ 7가지 주요 변이 종류 : Missense Mutation, Nonsense Mutation, Splice Site, In Frame Insertion, In Frame Deletion, Frame Shift Insertion, Frame Shift Deletion

b. RNA : 사용자가 선택(필터링)한 유전자에 대해 RNA expression level에 따라 표에 O 표시

→ High :  $z\text{-score} \geq 1$ , Low :  $z\text{-score} \leq -1$ , Intermediate : 나머지

c. Proteome : 사용자가 선택(필터링)한 유전자에 대해 Proteome expression level에 따라 표에 O 표시

→ High :  $z\text{-score} \geq 1.5$ , Low :  $z\text{-score} \leq 0.5$ , Intermediate : 나머지

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

Patient Summary
Drug Relation Plot

Choose a Sample (100)

**Basic Information**

Total: 5

Clinical Attribute	Value
Sex	F
Age Of Diagnosis (1st Day Of Diagnosis)	39
BMI (1st Day Of Diagnosis)	23.76
Alcohol Consumption	No
Smoking Status	No

<< < **1** > >>

**Genomic Information**

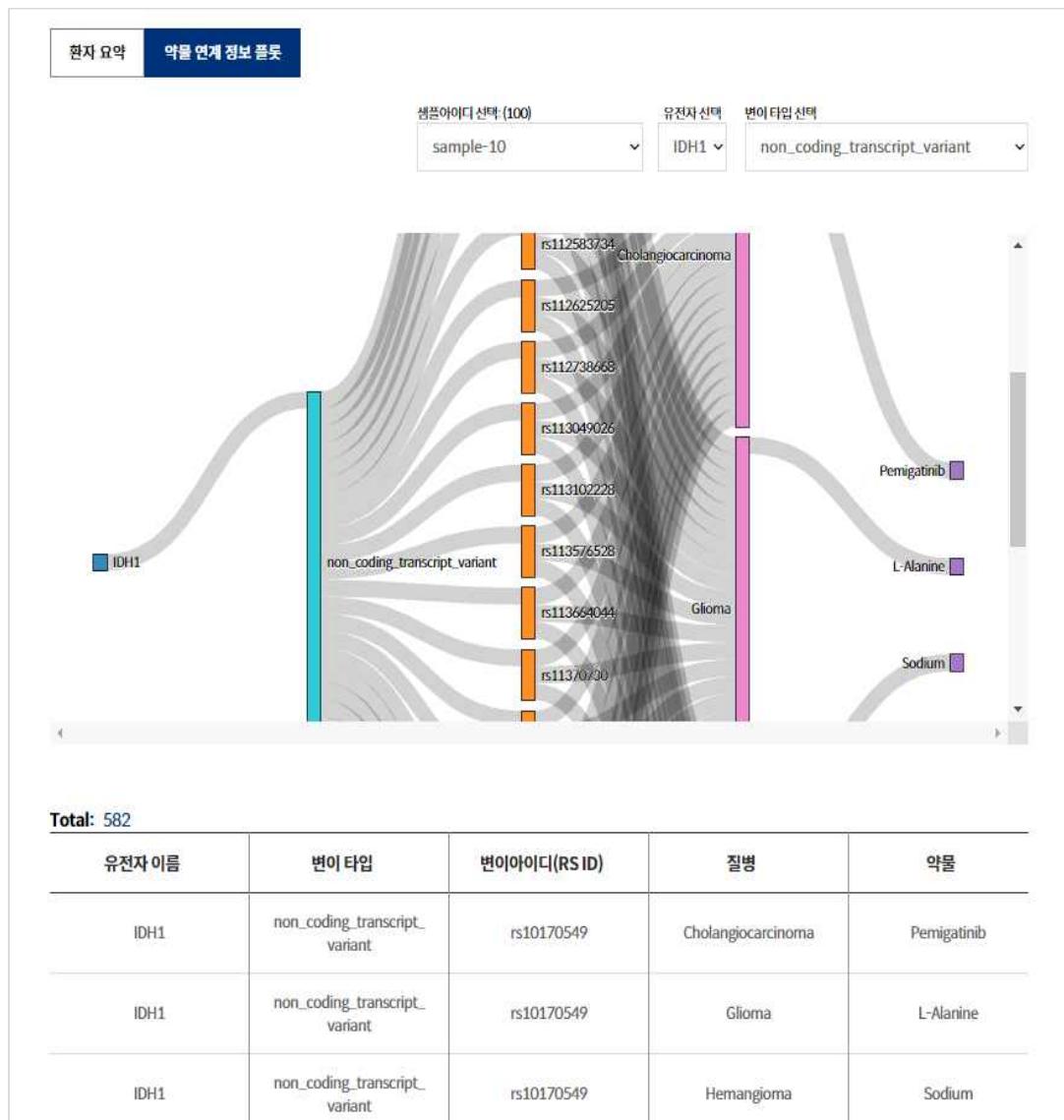
Total: 28

Gene Name	DNA Mutation		RNA		
	Yes	No	High	Intermediate	Low
ATRX		NO			O
BRAF		NO		O	
CDK4		NO		O	
CDK6		NO	O		
CDKN2A		NO		O	

< 환자 요약(Patient Summary) 정보 >

- 2) 약물 연계 정보 플롯(Drug Relation Plot) : 샘플, 유전자, 변이 타입 선택을 통해 해당 조건에 대한 [유전자 - 변이 타입 - 변이아이디(RS ID) - 질병 - 약물 연계 정보]를 Sankey plot 형태로 출력
- [샘플아이디 선택] 박스를 통해 샘플 선택
  - [유전자 선택] 박스를 통해 유전자 선택
  - [변이 타입 선택] 박스를 통해 변이 정보 선택
  - 현재는 DB상 구축된 Reference Drug Relation Information을 통해 선택한 유전자에 대한 Reference 정보를 시각화하여 출력  
→ Reference DB : Variant Effect Predictor (VEP)

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)



< 약물 연계 정보 플롯(Drug Relation Plot) >

### III. 사용자 데이터 분석

#### 사용자의 데이터를 분석합니다.

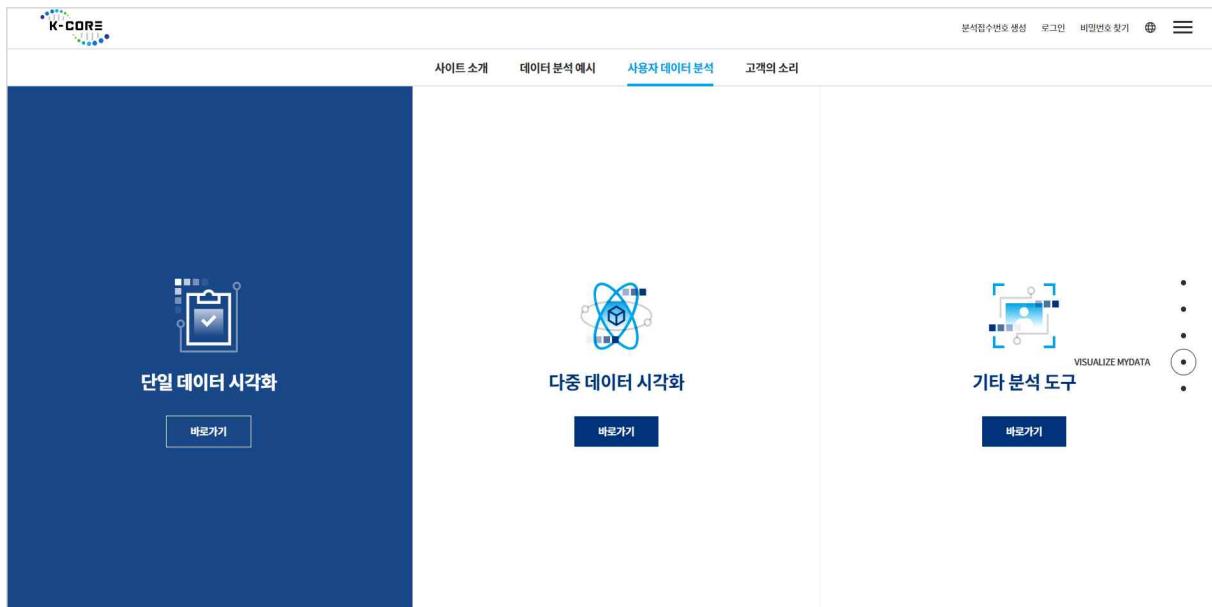
사용자 데이터 분석 메뉴는 데이터 종류에 따라 여러 가지 사용자 분석을 지원합니다. 단일 및 다중 데이터 시각화 메뉴는 암(질환) 임상 및 오믹스 데이터를 심층적으로 통계분석, 시각화할 수 있는 서비스를 제공합니다. 기타 분석 도구 메뉴는 생물학적 서열 정보를 분석을 지원하고, VCF 형식의 파일을 MAF 형식으로 변환시켜주는 서비스를 제공합니다.

#### 1. 단일 데이터 시각화

- 단일 데이터 시각화는 일회성 분석 서비스로 시각화 결과 화면에서 이동하면 해당 결과는 저장되지 않습니다.
- 시각화 결과를 프로젝트 형식으로 저장하고 열람(조회) 기능을 사용하려면 [다중 데이터 시각화] 서비스를 사용하여야 합니다.

#### 단일 데이터 시각화 절차

##### ① [사용자 데이터 분석] 메뉴의 단일 데이터 시각화

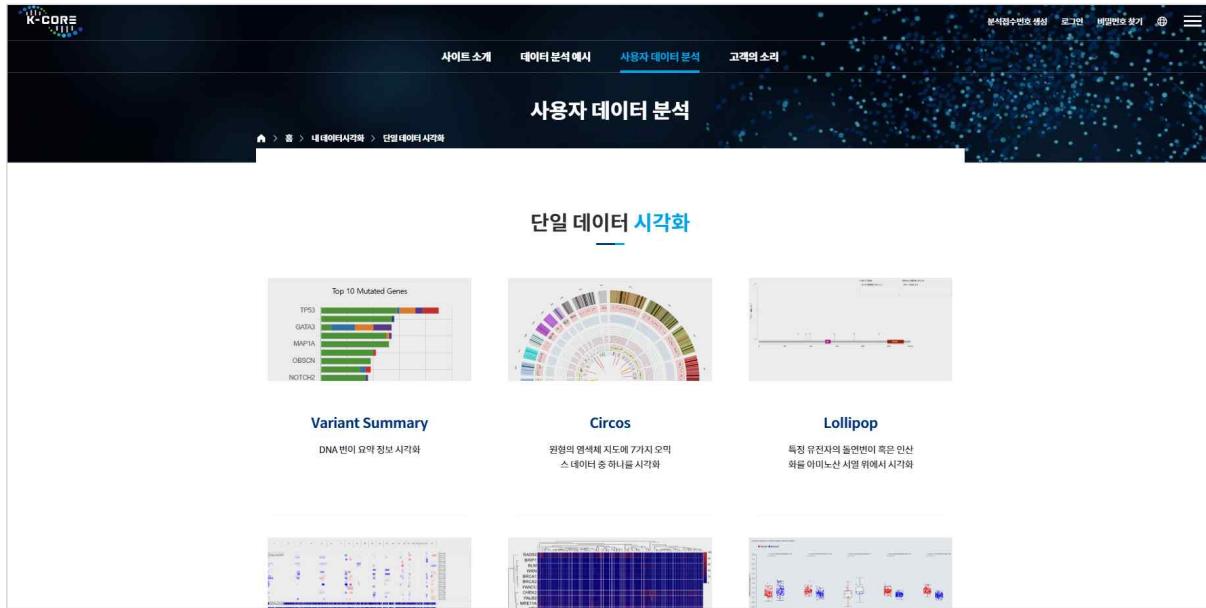


< 사용자 데이터 분석 - 단일 데이터 시각화 >

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

### ② 카드섹션 페이지에서 원하는 시각화 서비스를 선택

- . 단일 데이터 시각화 종류 : Variant Summary, Circos, Lollipop, CNV, Heatmap, Box, Survival



< 단일 데이터 시각화 도구 선택 화면 >

### ③ 서비스 이용을 위한 로그인 페이지로 이동

- . 사용자 데이터 분석 메뉴는 로그인 후 사용 가능한 서비스로, 이미 로그인된 상태라면 바로 데이터 업로드 화면으로 이동합니다.

The screenshot shows the 'Login' page of the K-CORE interface. At the top, there are tabs for '사이트 소개', '데이터 분석 예시', '사용자 데이터 분석' (selected), and '고객의 소리'. Below the tabs, the title '로그인' is displayed. A logo for 'K-CORE' is shown with the text '환영합니다' and '로그인을 위해 아래 정보를 입력해주세요.' Below this, there are two input fields:

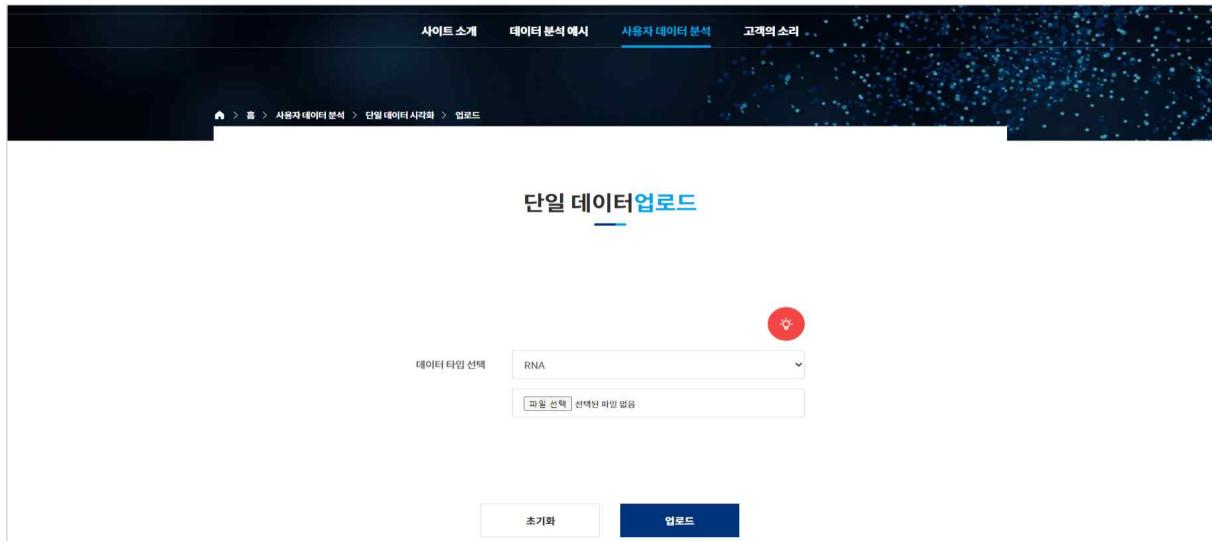
- 분석접수번호: [입력 필드]
- 비밀번호: [입력 필드]

At the bottom of the form, there is a note: '분석접수번호 찾기' and '비밀번호 찾기' with the text '모임에서 제공하는 경우에 대한 사용 책임은 사용자 본인에게 있습니다.' A blue '로그인' button is located at the bottom right of the form.

< 로그인 화면 >

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

④ 로그인 후 선택한 단일 데이터 시각화 서비스에 대한 데이터 업로드 화면으로 이동



< 단일 데이터 업로드 화면 >

Data Type	Variant Summary	Circos	Lollipop	CNV	Heatmap	Box	Survival
Clinical Information							○
DNA Mutation	○	○	○				
CNV		○		○			
Methylation		○			○		
RNA		○			○	○	
Fusion		○					
Proteome		○			○	○	
Phosphorylation		○			○		

< 표. 단일 데이터 시각화 별 이용 가능한 데이터 종류>

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

- . 각각의 단일 데이터 시각화 종류에 따라 업로드할 수 있는 데이터가 다르게 구성됩니다.
  - Variant Summary : DNA Mutation
  - Circos : DNA Mutation, CNV, Methylation, RNA, Fusion, Proteome, Phosphorylation
  - Lollipop : DNA Mutation, Phosphorylation
  - CNV : CNV
  - Heatmap : Methylation, RNA, Proteome, Phosphorylation
  - Box : RNA, Proteome
  - Survival : Clinical Information
- . 데이터 업로드 박스 우측 상단의 전구모양 버튼을 누르면 해당 데이터에 대한 안내 팝업이 나타납니다.

Sample File Download						X
sample_id	gene_name	raw	norm	type		
sample_14	IFNWP15	0	-0.138123	T		
sample_14	IFNWP18	0	-0.148884	N		
sample_14	IFNWP19	0.29	-0.164395	T		
sample_14	IFNWP2	0	-0.298402	T		
sample_14	IFNWP4	0	0	T		
sample_14	IFNWP5	0	0	T		
sample_14	IFNWP9	0	-0.156845	N		
sample_14	IFRD1	16.82	-0.577061	T		
sample_14	IFRD2	30.63	-0.852351	N		

Click on the link to download the sample file [Download](#)

[Close](#)

< 사용자 데이터 설명 팝업 >

- . 데이터 안내 팝업 하단의 [Download](#)를 누르면 해당 데이터에 대한 .tsv 형식의 샘플 파일을 다운받을 수 있습니다.

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

### ⑤ 데이터 업로드 후 [업로드] 버튼을 누르면 QC 화면으로 이동

- . 업로드한 데이터에 대한 검증을 마치면 시각화 전 해당 데이터를 테이블 형태로 나타내어 줍니다.
- . 업로드한 파일에 이상이 없을 경우 [시각화] 버튼이 활성화 되고, 문제가 있을 경우 잘못된 데이터에 대해 안내 메시지를 제공합니다.

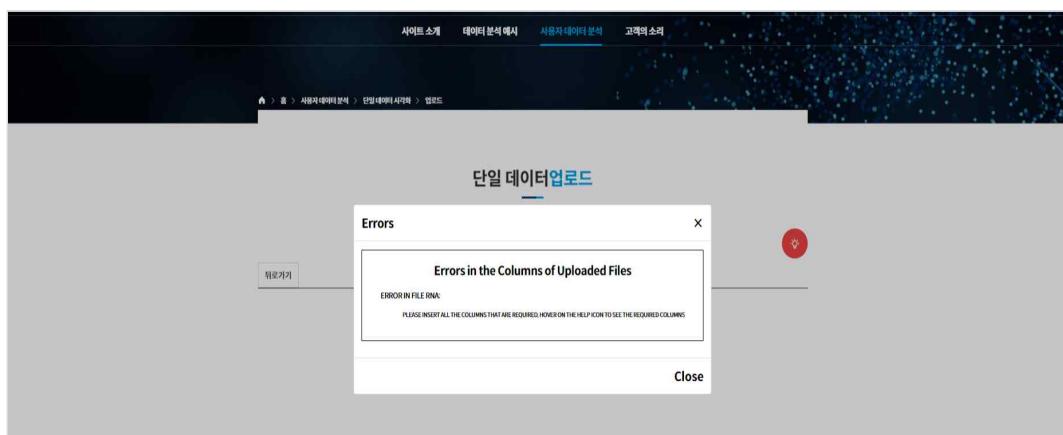
**1) 업로드 데이터에 문제가 없는 경우 :** 업로드한 데이터의 처음 10개 정보를 테이블 형식으로 제공하고, [시각화] 버튼 활성화

sample_id	gene_name
sample_14	IFNW15
sample_14	IFNW18
sample_14	IFNW19
sample_14	IFNW2
sample_14	IFNW4
sample_14	IFNW5
sample_14	IFNW9
sample_14	IFRD1
sample_14	IFRD2
sample_14	IFNG15

< QC 화면 - 정상적인 입력 데이터 파일을 업로드 한 경우 >

**2) 업로드 데이터에 문제가 있는 경우 :** 잘못된 형식 또는 데이터 값(타입)에 대해 안내

- 입력 데이터의 파일형식 자체가 잘못된 경우 안내 팝업 출력  
→ 필수적인 컬럼을 모두 포함하라는 안내 팝업 출력



< 데이터 형식 자체가 잘못된 파일을 업로드한 경우 >

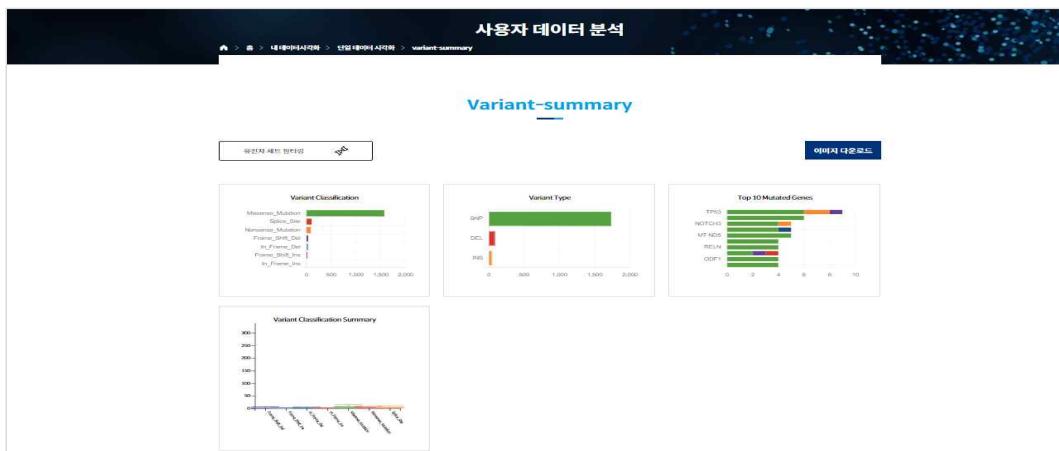
## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

- 입력 데이터의 파일 형식은 문제가 없지만 데이터 값의 타입이 잘못된 경우 테이블을 통해 잘못된 데이터 안내
  - 입력파일의 잘못된 값에 대한 부분이 빨간색으로 표시
  - 우측 상단의 전구모양 버튼을 누르면 업로드 데이터의 몇 번째 행이 잘못되었는지 안내

The screenshot shows a table titled 'Variant-summary' with two columns: 'gene\_name' and 'raw'. The 'raw' column contains values 'e', 'a', 'b', 'c', 'd', 'e', 'g', 'a', 'b', and 'c', which are highlighted in red. Below the table is a navigation bar with buttons for '<<', '<', '1', '2', '3', '>', and '>>'. A red circle with a question mark icon is located at the top right of the table area.

< 잘못된 데이터 탑입을 입력한 파일을 업로드한 경우 >

- ⑥ 데이터 업로드 - QC 과정까지 마친 후 활성화된 [시각화] 버튼을 통해 해당 결과 출력
- . 업로드한 데이터에 대한 [유전자 세트 필터링]을 통해 유전자 선택 가능
  - . [이미지 다운로드] 버튼을 통해 결과 다운로드 가능
  - . 단일 데이터 시각화는 하나의 오믹스 데이터를 사용한 일회성 분석 도구로, 분석 결과에 대해 프로젝트 형태로 계정에 저장하고 싶다면 다중 데이터 시각화를 이용해야함



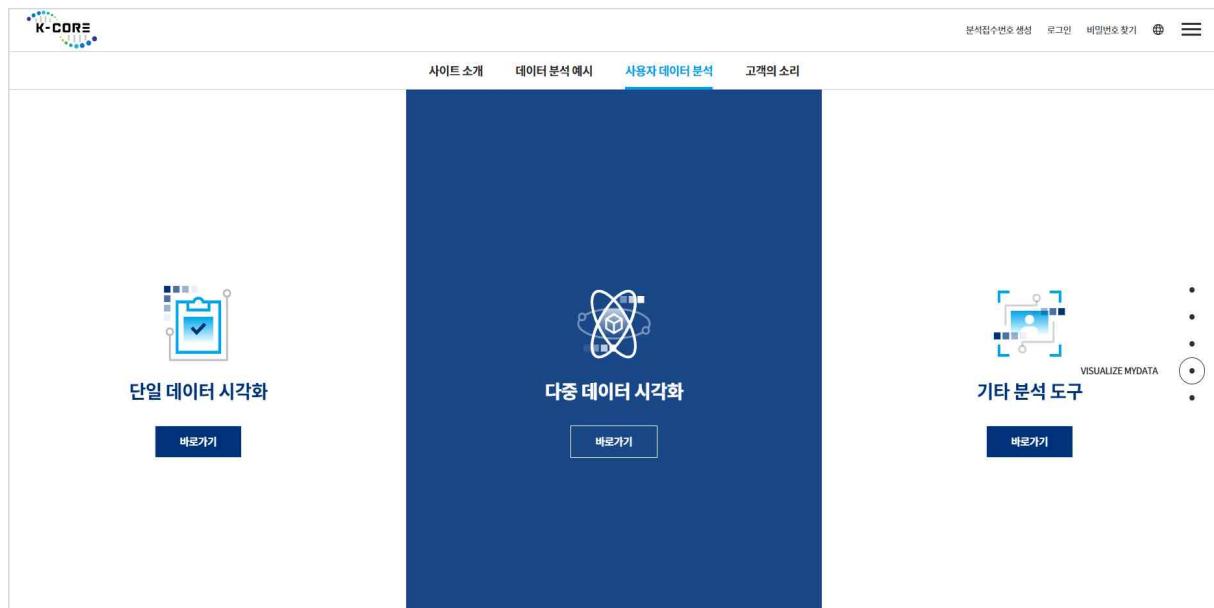
< 단일 데이터 시각화 결과 예시 >

## 2. 다중 데이터 시각화

- 다중 데이터 시각화는 사용자가 업로드한 임상정보 및 오믹스 데이터 세트에 대해 프로젝트 형식으로 계정에 저장되며, 해당 프로젝트에 대해 조회/삭제를 할 수 있습니다.
- 전체 시각화 서비스에 대하여 사용자가 업로드한 데이터 세트에 대해 이용 가능한 기능만 활성화 됩니다.

### 다중 데이터 시각화 프로젝트 생성 절차

#### ① [사용자 데이터 분석] 메뉴의 **다중 데이터 시각화**



< 사용자 데이터 분석 - 다중 데이터 시각화 >

#### ② 서비스 이용을 위한 로그인 페이지로 이동

- . 사용자 데이터 분석 메뉴는 로그인 후 사용 가능한 서비스로, 이미 로그인된 상태라면 바로 데이터 업로드 화면으로 이동합니다.

#### ③ 프로젝트 생성 페이지 이동

- . **프로젝트 생성** : 사용자 데이터에 대한 프로젝트 생성하는 과정으로 업로드 화면으로 이동합니다.
- . **프로젝트 보기** : 이미 분석 완료된 프로젝트에 대해 조회/삭제 할 수 있는 화면으로 이동합니다.

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)



< 다중 데이터 시각화의 프로젝트 생성/보기 선택 화면 >

### ④ 다중 데이터 업로드

- 프로젝트 생성을 통해 사용자 데이터 세트에 대한 업로드 페이지로 이동합니다.

The screenshot shows the 'Multiple Data Upload' page. At the top, there are tabs: '사이트 소개' (Site Introduction), '데이터 분석 예시' (Data Analysis Examples), '사용자 데이터 분석' (User Data Analysis, currently selected), and '고객의 소리' (Customer Feedback). Below the tabs, the page title is '다중 데이터 업로드'. The main content area has a form for uploading data files. On the left, there is a '프로젝트 이름' (Project Name) input field. The form is divided into sections for different types of data:

- Clinical Information file [Required]: Includes a note that '파일선택' (File Selection) is required and a '초기화' (Reset) button.
- DNA Mutation: Includes a note that '파일선택' (File Selection) is required and a '초기화' (Reset) button.
- CNV: Includes a note that '파일선택' (File Selection) is required and a '초기화' (Reset) button.
- methylation: Includes a note that '파일선택' (File Selection) is required and a '초기화' (Reset) button.
- RNA File: Includes a note that '파일선택' (File Selection) is required and a '초기화' (Reset) button.
- Fusion: Includes a note that '파일선택' (File Selection) is required and a '초기화' (Reset) button.
- Proteome: Includes a note that '파일선택' (File Selection) is required and a '초기화' (Reset) button.
- Phosphorylation: Includes a note that '파일선택' (File Selection) is required and a '초기화' (Reset) button.

On the right side, there is a 'Note' box with the following content:

- 업로드한 데이터와 관련된 경우에 대해서만 시각화 결과를 제공합니다.
- 플랫폼마다 필요한 오믹스 데이터 정보는 데이터 분석 예시 페이지를 참고 바랍니다.

At the bottom, there are two buttons: '모두 초기화' (Reset All) and a blue '업로드' (Upload) button.

< 다중 데이터 업로드 화면 >

- 화면 우측의 "Note" 팝업을 통해 다중 데이터 시각화 도구에 대해 간단히 안내합니다.
- (필수) 프로젝트 이름을 설정합니다.
- [파일 선택] 버튼을 통해 임상정보 및 각 오믹스 데이터 파일을 업로드 합니다.
  - (필수) 임상정보(Clinical Information) 파일은 사용자 필수 업로드 데이터입니다.
  - DNA, RNA, Protein 관련 오믹스 데이터는 전체 또는 사용자가 선택한 파일만 업로드할 수 있습니다.  
→ 단, 사용자가 업로드한 데이터에 따라 제공하지 않는 시각화 도구도 있습니다.

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

Data Type	Circos	Oncoprint	Lollipop	Volcano	Heatmap	Survival	Correlation	CNV	Box	Fusion	Sankey
Clinical Information		○		○	○	○		○		○	○
DNA Mutation	○	○	○			○					
CNV	○							○			
Methylation	○				○						
RNA	○			○	○	○	○	○	○		
Fusion	○									○	
Proteome	○			○	○	○	○	○	○		
Phosphorylation			○		○						

〈 표. 다중 데이터 시각화 별 이용 가능한 데이터 종류〉

### ⑤ 데이터 업로드 후 [업로드] 버튼을 누르면 QC 화면으로 이동

- . 업로드한 데이터에 대한 검증을 마치면 시각화 전 해당 데이터를 테이블 형태로 나타내어 줍니다.
- . 업로드한 파일에 이상이 없을 경우 [시각화] 버튼이 활성화 되고, 문제가 있을 경우 잘못된 데이터에 대해 안내 메시지를 제공합니다.

1) 업로드 데이터에 문제가 없는 경우 : 업로드한 데이터의 처음 10개 정보를 테이블 형식으로 제공하고, [시각화] 버튼 활성화

The screenshot shows the QC interface with a table of 10 samples. The columns are labeled: sample\_id, bmi\_vl, drnk\_yn, smok\_yn, fmhs\_brst\_yn, and er\_score. The data is as follows:

sample_id	bmi_vl	drnk_yn	smok_yn	fmhs_brst_yn	er_score
sample_1	26.88	false	false		2
sample_2	19.17	false	false	true	1
sample_3	17.74	false	false	false	2
sample_4	21.77	false	false	false	2
sample_5	22.09	false	false	false	2
sample_6	23.28	false	false	false	1
sample_7	30.64	true	false	true	1
sample_8	22.89	false	false		1
sample_9	22.03				
sample_10	25.54	false	false	false	2

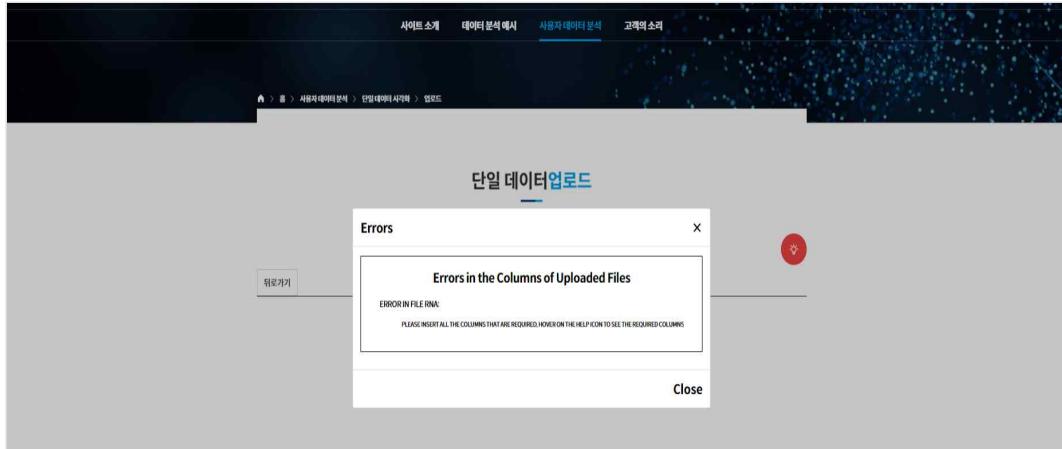
Below the table, there are navigation buttons: <<, <, 1, >, >>.

〈 QC 화면 - 정상적인 입력 데이터 파일 업로드 한 경우 〉

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

2) 업로드 데이터에 문제가 있는 경우 : 잘못된 형식 또는 데이터 값(타입)에 대해 안내

- 입력 데이터의 파일형식 자체가 잘못된 경우 안내 팝업 출력  
→ 필수적인 컬럼을 모두 포함하라는 안내 팝업 출력



< 데이터 형식 자체가 잘못된 파일을 업로드한 경우 >

- 입력 데이터의 파일 형식은 문제가 없지만 데이터 값의 타입이 잘못된 경우 테이블을 통해 잘못된 데이터 안내  
→ 입력파일의 잘못된 값에 대한 부분이 빨간색으로 표시  
→ 우측 상단의 전구모양 버튼을 누르면 업로드 데이터의 몇 번째 행이 잘못되었는지 안내

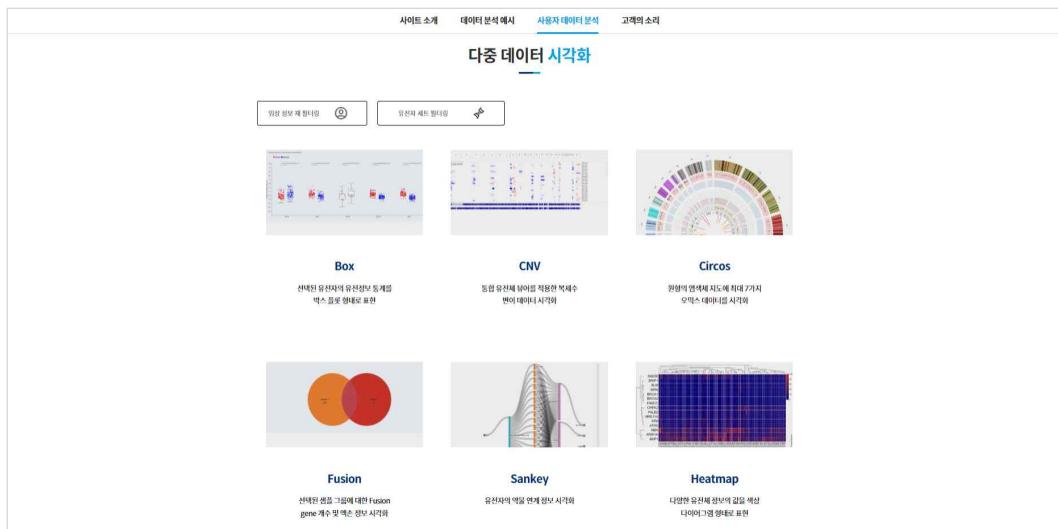
gene_name	raw
II	e
II	a
JJ	b
JJ	c
KK	d
KK	e
LL	g
LL	a
MM	b
MM	c

< 잘못된 데이터 타입을 입력한 파일을 업로드한 경우 >

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

### ⑥ QC 후 사용 가능한 시각화 서비스 리스트

- . 업로드한 사용자 데이터의 종류에 따라 사용 가능한 시각화 서비스를 카드섹션으로 나타냅니다.
- 전체 오믹스 데이터 업로드 기준 사용 가능한 다중 데이터 시각화 리스트 : Box, CNV, Circos, Fusion, Sankey, Heatmap, Correlation, Volcano, Lollipop, Survival, OncoPrint



< QC 후 다중 데이터 시각화 리스트 화면 >

- . 사용자가 원하는 시각화 서비스를 선택해 데이터 세트 필터링을 통한 결과 출력이 가능합니다.

### 다중 데이터 시각화 프로젝트 보기

#### ① [사용자 데이터 분석] 메뉴의 다중 데이터 시각화



< 사용자 데이터 분석 - 다중 데이터 시각화 >

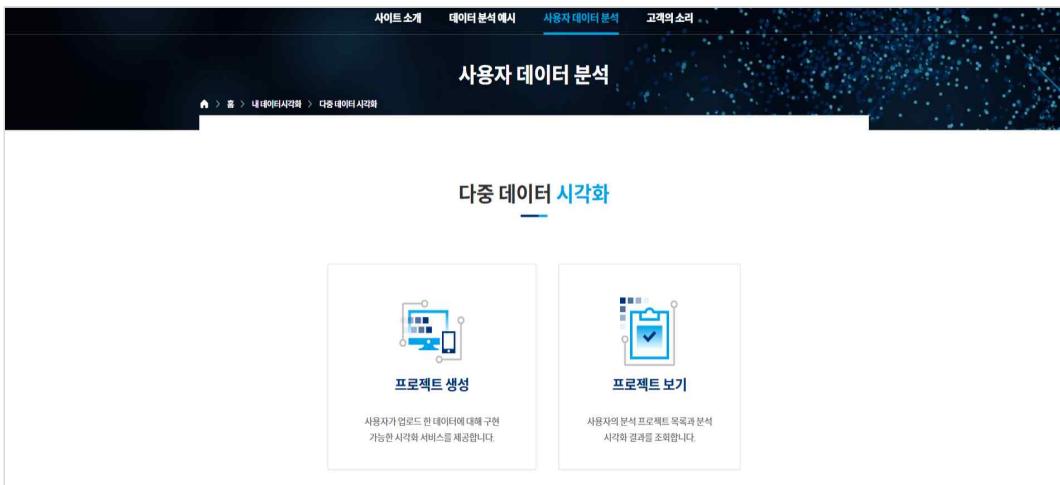
## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

### ② 서비스 이용을 위한 로그인 페이지로 이동

- . 사용자 데이터 분석 메뉴는 로그인 후 사용 가능한 서비스로, 이미 로그인된 상태라면 바로 데이터 업로드 화면으로 이동합니다.

### ③ 프로젝트 보기 페이지 이동

- . **프로젝트 생성** : 사용자 데이터에 대한 프로젝트 생성하는 과정으로 업로드 화면으로 이동합니다.
- . **프로젝트 보기** : 이미 분석 완료된 프로젝트에 대해 조회/삭제 할 수 있는 화면으로 이동합니다.



< 다중 데이터 시각화의 프로젝트 생성/보기 선택 화면 >

### ④ 프로젝트 보기 페이지 이동

- . 계정에서 생성한 프로젝트에 대해 조회 및 삭제할 수 있는 페이지입니다.
- . **프로젝트 테이블**에서 각 프로젝트에 대해 업로드한 데이터 정보를 확인할 수 있고, [관리] 컬럼을 통해 원하는 프로젝트를 보기/삭제 할 수 있습니다.
- . 테이블 우측 위쪽에 [프로젝트 생성] 버튼을 통해 새로운 프로젝트를 생성할 수 있습니다.
  - 한 계정에서 생성할 수 있는 프로젝트는 최대 5개입니다.
  - 한 계정에서 5개의 프로젝트를 생성한 상태라면 최소 하나의 프로젝트를 삭제해야 새로운 프로젝트 생성이 가능
- . 프로젝트마다 확인 가능한 기간은 생성일로부터 2주입니다. 이후 해당 프로젝트는 삭제됩니다.

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

The screenshot shows a web-based application interface for managing genomic data. At the top, there is a navigation bar with links to '사이트 소개', '데이터 분석 예시', '사용자 데이터 분석', and '고객의 소리'. Below the navigation bar, a breadcrumb trail indicates the current location: 'Home > 내 데이터시각화 > 다중 데이터 시각화 > 다중 데이터 프로젝트 보기'. The main content area is titled '다중 데이터 시각화' and displays a table of projects. The table has columns: 번호 (Number), 관리 (Management), 프로젝트 이름 (Project Name), 임상정보 (Clinical Information), DNA Mutation, and CNV. There are two entries in the table:

번호	관리	프로젝트 이름	임상정보	DNA Mutation	CNV
1	<a href="#">보기</a>   <a href="#">삭제</a>	manual_data_78	0	0	0
2	<a href="#">보기</a>   <a href="#">삭제</a>	new_test_70	0	0	0

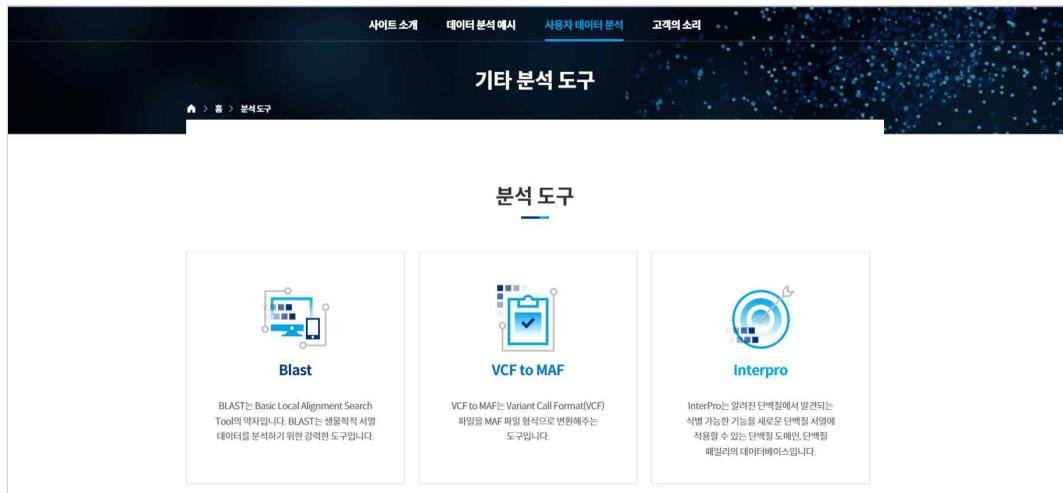
Below the table, there are navigation buttons: '<<', '<', '1' (highlighted in blue), '>', and '>>'. To the right of the table, there is a 'Note' section with a 'X' button. It contains two bullet points:

- 생성 가능한 프로젝트는 총 5개입니다.
- 프로젝트마다 확인 가능한 기간은 생성일로부터 2주입니다. 이후 해당 프로젝트는 삭제됩니다.

< 다중 데이터 시각화 - 프로젝트 보기 화면 >

### 3. 기타 분석 도구

- 기타 분석 도구 메뉴는 시각화 기능 외 생물학적 서열 데이터(FASTA) 또는 Variant Call Format(VCF) 파일을 분석하기 위한 메뉴입니다.
- 해당 메뉴는 생물학적 서열 데이터를 사용하는 BLAST, Interpro, VCF 파일을 사용하는 VCFtoMAF 도구로 구성됩니다.

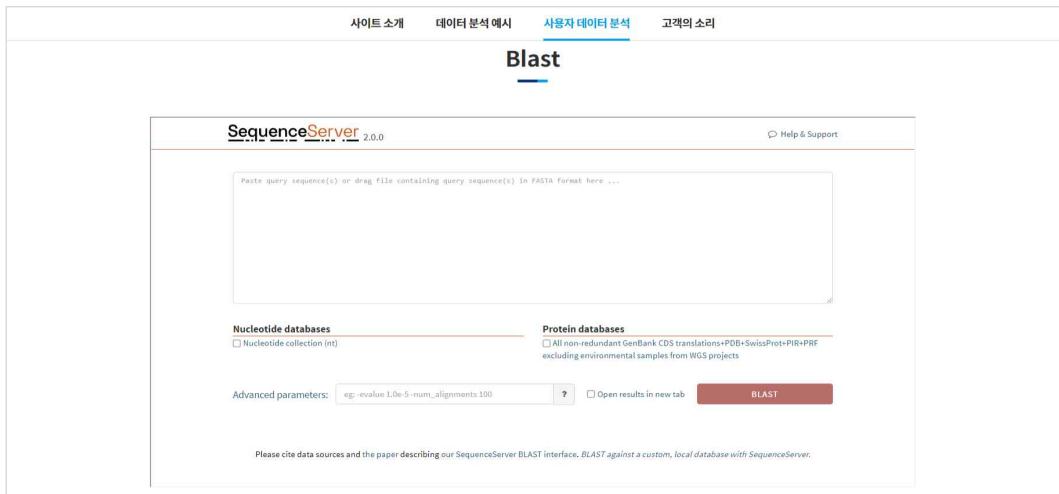


< 사용자 데이터 분석 - 기타 분석 도구 화면 >

#### BLAST

- BLAST는 Basic Local Alignment Search Tool의 약자로 생물학적 서열 데이터(FASTA 파일)를 위한 분석 도구입니다.
  - ① 서열 입력 상자에 직접 FASTA 서열 정보를 입력하거나 또는 FASTA 파일을 Drag & Drop 합니다.
  - ② 본인의 서열 데이터가 염기서열 기반이라면 Nucleotide databases를 선택하고, 아미노산 기반이라면 Protein databases를 선택합니다.
  - ③ Advanced parameters 상자에서 parameter를 설정([?] 박스를 누르면 parameter 종류 확인 가능)  
예시) -evaluate 1.0e-5 -num\_alignments 100  
(입력 정보가 없으면 default 값으로 자동 설정. 특별한 경우가 아니면 parameter는 default로 설정하는 것이 좋음)
  - ④ 입력정보를 모두 입력했으면 우측 하단의 BLAST 버튼 클릭
  - ⑤ 분석이 완료되었으면 결과 파일을 다운로드 할 수 있습니다.

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)



< 기타 분석 도구 화면- BLAST >

### VCF to MAF

- VCF to MAF는 Variant Call Format(VCF) 파일을 Minor Allele Frequency(MAF) 파일 형식으로 변환해 주는 도구입니다.

① [파일 선택] 버튼을 통해 VCF 파일을 업로드 합니다.

(\* Reference genome HG38 version을 사용하여 생성된 파일만 허용됩니다.)

② VCF 파일 업로드 후 [제출] 버튼을 누르면 변환이 시작됩니다.

③ 변환 후 결과 파일을 다운로드 할 수 있습니다.



< 기타 분석 도구 화면- VCF to MAF >

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

### Interpro

- Interpro는 알려진 단백질에서 발견되는 식별 가능한 기능을 새로운 단백질 서열에 적용할 수 있는 단백질 도메인, 단백질 패밀리 데이터베이스로써, 단백질 엔지니어링을 패밀리로 분류하고 도메인 및 중요 부위의 존재를 예측하여 기능 분석을 제공하는 리소스입니다.
- Interpro는 signatures 예측 모델을 사용합니다.
  - ① FASTA 형식의 서열정보를 입력 상자에 직접 입력하거나, [파일 선택] 버튼을 통해 아미노산 서열 정보를 담고 있는 FASTA 파일을 업로드 합니다.
  - ② 서열 정보 입력 또는 파일 업로드 후 [제출] 버튼을 누르면 분석이 시작됩니다.
  - ③ 분석 후 결과 파일을 다운로드 할 수 있습니다.

The screenshot shows the 'Interpro' analysis tool interface. At the top, there are tabs: '사이트 소개', '데이터 분석 예시', '사용자 데이터 분석' (which is underlined), and '고객의 소리'. Below the tabs, the 'Interpro' logo is centered. To the right of the logo is a red circular button with a white play icon. The main area has a light gray background. In the center, there is a large text input field labeled '시퀀스를 입력하세요' with placeholder text 'FASTA 형식의 서열'. Below this input field are two buttons: '업로드 파일' (Upload file) and '파일 선택' (Select file). To the right of these buttons is another button labeled '선택된 파일 없음' (No selected file). At the bottom of this section is a blue '제출' (Submit) button. At the very bottom of the interface, there is a small note: '참고: 단백질 서열은 fasta 포맷이어야 합니다.'

< 기타 분석 도구 화면- Interpro >

## 부록

### 사용자 임상정보 및 오믹스 데이터 안내

#### 임상정보 (Clinical information)

- 임상정보 파일의 각 컬럼은 **탭(tab)**으로 분리된 형식입니다.
- 첫 번째 행은 컬럼(column) 행으로 임상정보의 종류를 나타내는 행입니다.
  - . [sample\_id]와 사용자가 지정한 각 임상 변수의 이름으로 이루어집니다.
  - . 임상 정보 파일의 컬럼 개수는 15개를 초과할 수 없고, 반드시 [sample\_id] 컬럼이 필수적으로 포함되어 있어야 합니다. (샘플 아이디 열의 이름은 반드시 [sample\_id]로 지정)
  - . 컬럼 구성 예시 :

sample_id	BMI	T_category	smoke_yn	weight
example1	42	T1	TRUE	32.45
example2	23		FALSE	
example3	58	T3	TRUE	45.2

- 두 번째 행 부터는 각 컬럼에 맞는 데이터 값 정보입니다.
  - . 두 번째 행 부터는 각 컬럼에 맞는 값으로 구성됩니다.
  - . 해당 컬럼의 **데이터 값이 없는 경우 빈 칸**으로 비워둡니다.
  - . 각 컬럼 별 데이터 값은 같은 형식의 자료형으로 구성되어야 합니다. (범주형, 연속형, 논리형)
    - **범주형(Categorical)** : 범주형 타입은 그룹형 데이터에 사용되며 숫자는 연속형 데이터로 인식하기 때문에 문자형 데이터는 반드시 문자가 포함되어야 합니다.
      - 1, 2, 3과 같은 숫자로 그룹지정 시 시스템에서 연속형 타입으로 인식하기 때문에 반드시 Group1, Group2, Group3와 같은 문자를 포함해주어야 합니다.
      - **범주형 데이터는 그룹지정을 최대 5개까지 할 수 있습니다.**
    - **연속형(Numerical)** : 연속형 타입은 연속적인 수로 수령화가 가능한 데이터에 사용되며 정수 또는 실수로 이루어집니다.
    - **논리형(Boolean)** : True / False

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

---

- . 데이터(값) 구성 예시 :

sample_id	BMI	T_category	smoke_yn	weight
example1	42	T1	TRUE	32.45
example2	23		FALSE	45.2
example3	58	T3	TRUE	

- 임상정보(Clinical Information) 파일에는 **한글 및 공백 기호를 사용할 수 없습니다.**

- . 컬럼 및 데이터 값 모두 한글 및 공백 기호를 사용할 수 없습니다.

예시) T\_catergory (o), T catergory (x), group1(o), group 1(x)

- 생존분석(Survival analysis)을 위한 컬럼을 추가할 수 있습니다. (**Recurrence, Survival**)

- . 환자의 **재발확률**을 분석하기 위해 **[rlps\_yn], [rlps\_cnfr\_drtm]** 컬럼을 추가할 수 있습니다.

- rlps\_yn : 환자의 재발 여부를 나타내는 컬럼으로 논리형 타입 사용 (재발 : True, 재발x : False)
- rlps\_cnfr\_drtm : 질병(암)이 재발된 시간을 나타내는 데이터로 연속형 타입 사용

- . 환자의 **생존확률**을 분석하기 위해 **[rlps\_yn], [rlps\_cnfr\_drtm]** 컬럼을 추가할 수 있습니다.

- death\_yn : 환자의 생존 여부를 나타내는 컬럼으로 논리형 타입 사용 (생존 : True, 사망 : False)
- death\_cnfr\_drtm : 샘플(환자)의 사망까지 시간을 나타내는 데이터로 연속형 타입 사용

- . 임상정보 파일에서 생존분석(Survival analysis)을 위한 컬럼은 필수 컬럼은 아니지만, 재발 또는 생존확률을 분석하기 위해선 -\_yn, -\_cnfr\_drtm 컬럼을 항상 같이 사용해야 합니다.

	A	B	C	D	E	F	G	H	I
1	sample_id	age	bmi	smok_yn	t_category	rlps_yn	rlps_cnfr_drtm	death_yn	death_cnfr_drtm
2	sample_1	36	26.88	FALSE	T2	TRUE	25.7	TRUE	23.2
3	sample_2	37		TRUE	T3	TRUE	17.7	FALSE	18.4
4	sample_3	39	17.74	TRUE	T1	FALSE	17.7	TRUE	18.4
5	sample_4		21.77	TRUE	T2	TRUE	7.6	FALSE	6.9
6	sample_5	40	22.09		T2	TRUE	0.8	TRUE	0.9
7	sample_6	40		FALSE	T2	FALSE	17.7	FALSE	18.4
8	sample_7		30.64	TRUE	T2	TRUE	92.9	TRUE	54.6
9	sample_8	40	22.89	FALSE	T2	TRUE	24.2	FALSE	20.1
10	sample_9	32	22.03			TRUE	12.4	TRUE	9.6
11	sample_10	39	25.54	FALSE	T1	TRUE	17.7	TRUE	18.4

< 임상정보(Clinical information) 파일 예시 >

### 오믹스 데이터(Omics data)

#### ① DNA Mutation

- . DNA Mutation 파일의 각 컬럼은 **탭(tab)**으로 분리된 형식입니다.
- . **데이터 값이 없는 경우 빈 칸**으로 비워둡니다.
- . DNA Mutation 파일의 컬럼은 아래와 같이 구성되며 해당 양식에 맞게 파일을 구성하여야 합니다.
  - 1) **sample\_id** : 임상정보(Clinical information)와 일치하는 샘플 이름
  - 2) **gene\_name** : 공식 유전자 심볼 형식 (HUGO)
  - 3) **chromosome** : 염색체 번호 (형식 : chr1, chr2, chr3, ...)
  - 4) **start\_position** : 변이 시작 위치
  - 5) **end\_position** : 변이 종료 위치
  - 6) **variant\_classification** : 변이 종류 (7가지 주요 변이는 아래와 같이 나타냄)
    - Missense mutation → Missense\_Mutation
    - Splice site mutatio → Splice\_Site
    - Inframe insertion mutatio → In\_Frame\_Ins
    - Inframe deletion mutatio → In\_Frame\_Del
    - Frameshift insertion mutatio → Frame\_Shift\_Ins
    - Frameshift deletion mutation → Frame\_Shift\_Del
    - Nonsense mutation → Nonsense
  - 7) **variant\_type** : 변이 유형에 따라 아래 세 가지로 분류 (SNP, INS, DEL)
    - 단일 뉴클레오티드 다형성 → SNP
    - 삽입 → INS
    - 결실 → DEL
  - 8) **protein\_change** : 아미노산 변화 (예. p.R586R)
  - 9) **swiss\_prot\_acc\_id** : SwissProt 단백질 ID (예. O60477)
  - 10) **annotation\_transcript** : Ensembl 유전자 ID (예. ENST00000449103.1)
  - 11) **gc\_content** : RefSeq mRNA ID (예. NM\_001128432.2)
  - 12) **refseq\_mrna\_id** : RefSeq mRNA ID (예. NM\_001128432.2)

## 암 임상-오믹스 분석 시각화 시스템(K-CORE)

---

sample_id	gene_name	chromosome	start_position	end_position	variant_classification	variant_type	protein_change	swiss_prot_acc_id	annotation_transcript	gc_content	refseq_mrna_id
sample_1	FOXD4	chr1	117998	117998	Missense_Mutation	SNP	p.A41E	Q12950	ENST00000382500.2	0.682	NM_207305.4
sample_1	DMRT3	chr1	990492	990492	Silent	SNP	p.P302P	Q9NQL9	ENST00000190165.2	0.577	NM_021240.2
sample_1	IDUA	chr2	996574	996574	Missense_Mutation	SNP	p.N415I	P35475	ENST00000247933.4	0.771	NM_000203.3
sample_2	CYP2W1	chr2	1024202	1024207	In_Frame_Del	DEL		Q8TAV3	ENST00000308919.7	0.723	NM_017781.2
sample_2	CNN2	chr3	1037812	1037846	Frame_Shift_Ins	INS	p.P293fs	Q99439	ENST00000263097.4	0.66	NM_004368.2
sample_2	RAD52	chr3	1039269	1039269	In_Frame_Del	DEL	p.N76K		ENST00000358495.3	0.453	NM_134424.2
sample_2	RAD52	chr5	1039269	1039269	In_Frame_Del	DEL	p.N76K	P43351	ENST00000358495.3	0.453	NM_134424.2
sample_2	HMHA1	chr5	1083228	1083228	Frame_Shift_Del	DEL	p.T94I	Q92619	ENST00000313093.2	0.657	NM_012292.3
sample_3	SLC12A7	chr5	1085567	1085567	Frame_Shift_Del	DEL	p.A233T	Q9Y666	ENST00000264930.5	0.697	NM_006598.2
sample_3	ID11	chr5	1088655	1088655	Missense_Mutation	SNP		Q13907	ENST00000381344.3	0.468	NM_004508.2
sample_3	MUC2	chr8	1092948	1092948	Silent	SNP	p.T1589T	Q02817	ENST00000441003.2	0.627	NM_002457.2
sample_3	KCNQ1	chr8	2683309	2683309	Silent	SNP	p.S504S	P51787	ENST00000155840.5	0.587	NM_000218.2
sample_3	KCNQ1	chr8	2683309	2683309	Silent	SNP	p.S504S		ENST00000155840.5	0.587	NM_000218.2
sample_4	KCTD5	chr8	2732701	2732701	In_Frame_Ins	INS	p.G51D	Q9NXV2	ENST00000301738.4	0.701	NM_018992.3
sample_4	PCED1A	chr9	2816196	2816196	Missense_Mutation	SNP	p.R426M	Q9H1Q7	ENST00000360652.2	0.612	NM_022760.4
sample_4	FOXM1	chr11	2975591	2975591	In_Frame_Ins	INS	p.N315H	Q08050	ENST00000359843.3	0.517	NM_021953.3
sample_5	ZNRF4	chr11	5456131	5456131	In_Frame_Ins	INS	p.P210L	Q8WWF5	ENST00000222033.4	0.682	NM_181710.3
sample_5	DNAJB2	chr11	220149485	220149485	In_Frame_Del	DEL	p.E251Q	P25686	ENST00000336576.5	0.657	NM_006736.5
sample_6	DISP1	chr11	223176649	223176649	Missense_Mutation	SNP	p.A637V	Q96F81	ENST00000284476.6	0.448	NM_032890.3
sample_7	DISP1	chr11	223177819	223177819	Splice_Site	SNP	p.L1027R	Q96F81	ENST00000284476.6	0.473	NM_032890.3
sample_7	ACBD3	chr13	226374240	226374240	Splice_Site	SNP	p.G46E	Q9H3P7	ENST00000366812.5	0.821	NM_022735.3
sample_7	NYAP2	chr13	226447164	226447164	Missense_Mutation	SNP	p.V344G	Q9P242	ENST00000272907.6	0.637	NM_020864.1
sample_7	CDC42BPA	chr17	227348265	227348265	Silent	SNP	p.L224L		ENST00000366769.3	0.299	NM_003607.3
sample_7	COL4A4	chr17	227953465	227953465	Silent	SNP	p.G509G	P53420	ENST00000396625.3	0.587	NM_000092.4
sample_8	NUP133	chr17	229623269	229623269	Missense_Mutation	SNP	p.S429Y	Q8WUM0	ENST00000261396.3	0.413	NM_018230.2
sample_8	CAPN9	chr17	230903401	230903401	Nonsense	SNP	p.Y217*	O14815	ENST00000271971.2	0.547	NM_006615.2
sample_9	C1orf198	chr18	231004010	231004010	Silent	SNP	p.R83R	Q9H425	ENST00000366663.5	0.766	NM_032800.2
sample_10	GPR55	chr18	231775126	231775126	Silent	SNP	p.P184P	Q9Y276	ENST00000392040.1	0.552	NM_005683.3
sample_10	ALPP	chr19	233244568	233244568	Splice_Site	SNP	p.D193D	P05187	ENST00000392027.2	0.657	NM_001632.3
sample_10	ALPP	chr19	233245439	233245439	Splice_Site	SNP	p.N324N	P05187	ENST00000392027.2	0.647	NM_001632.3
sample_10	ALPPL2	chr19	233273280	233273280	Missense_Mutation	SNP	p.M285L	P10696	ENST00000295453.3	0.657	NM_031313.2
sample_10	GIGYF2	chr2	233712227	233712232	In_Frame_Del	DEL		Q6Y7W6	ENST00000409547.1	0.544	NM_015575.3

< DNA Mutation 파일 예시 >

## ② CNV (Copy Number Variation)

- . CNV 파일의 각 컬럼은 탭(tab)으로 분리된 형식입니다.
  - . 데이터 값이 없는 경우 빈 칸으로 비워둡니다.
  - . CNV 파일의 컬럼은 아래와 같이 구성되며 해당 양식에 맞게 파일을 구성하여야 합니다.
- 1) **sample\_id** : 임상정보(Clinical information)와 일치하는 샘플 이름
  - 2) **gene** : 공식 유전자 심볼 형식 (HUGO)
  - 3) **chromosome** : 염색체 번호 (형식 : chr1, chr2, chr3, ...)
  - 4) **start\_pos** : 변이 시작 위치
  - 5) **end\_pos** : 변이 종료 위치
  - 6) **log2** : 실수(decimal) 형태의 값을 입력
  - 7) **cn** : 복제 수(copy numbers)로 0 이상의 정수를 입력
  - 8) **depth** : 실수(decimal) 형태의 값을 입력
  - 9) **probes** : 0 이상의 정수를 입력
  - 10) **weight** : 실수(decimal) 형태의 값을 입력

<b>sample_id</b>	<b>gene</b>	<b>chromosome</b>	<b>start_pos</b>	<b>end_pos</b>	<b>log2</b>	<b>cn</b>	<b>depth</b>	<b>probes</b>	<b>weight</b>
sample_1	HES4	chr1	117998	117998	0.585556	3	141.034	286	74.9187
sample_1	HEYL	chr1	990492	990492	-0.0972076	2	94.4764	6774	1878.09
sample_1	HES5	chr2	996574	996574	-0.0972076	2	94.4764	6774	1878.09
sample_2	MTOR	chr2	1024202	1024207	-0.0972076	2	94.4764	6774	1878.09
sample_2	HES2	chr3	1037812	1037846	-0.0972076	2	94.4764	6774	1878.09
sample_2	HDAC1	chr3	1039269	1039269	-0.0972076	2	94.4764	6774	1878.09
sample_2	RPS6KA1	chr5	1039269	1039269	-0.0972076	2	94.4764	6774	1878.09
sample_2	ARID1A	chr5	1083228	1083228	-0.0972076	2	94.4764	6774	1878.09
sample_3	CASP9	chr5	1085567	1085567	-0.0972076	2	94.4764	6774	1878.09
sample_3	E2F2	chr5	1088655	1088655	-0.0972076	2	94.4764	6774	1878.09
sample_3	MMP23B	chr8	1092948	1092948	-0.0972076	2	94.4764	6774	1878.09
sample_3	MTHFR	chr8	2683309	2683309	-0.0972076	2	94.4764	6774	1878.09
sample_3	SPEN	chr8	2683309	2683309	-0.0972076	2	94.4764	6774	1878.09
sample_4	RAD54L	chr8	2732701	2732701	0.0576209	2	86.9083	2567	743.576
sample_4	ZFYVE9	chr9	2816196	2816196	0.0576209	2	86.9083	2567	743.576
sample_4	JAK1	chr11	2975591	2975591	-0.110286	2	56.9119	3009	932.968
sample_5	TGFBR3	chr11	5456131	5456131	-0.110286	2	56.9119	3009	932.968
sample_5	DIRAS3	chr11	220149485	220149485	-0.110286	2	56.9119	3009	932.968
sample_6	CSF1	chr11	223176649	223176649	-0.0177537	2	80.9354	1673	487.063
sample_7	NRAS	chr11	223177819	223177819	-0.0177537	2	80.9354	1673	487.063

< CNV 파일 예시 >

### ③ Methylation

- . Methylation 파일의 각 컬럼은 텁(tab)으로 분리된 형식입니다.
- . 데이터 값이 없는 경우 빈 칸으로 비워둡니다.
- . Methylation 파일의 컬럼은 아래와 같이 구성되며 해당 양식에 맞게 파일을 구성하여야 합니다.
  - 1) **sample\_id** : 임상정보(Clinical information)와 일치하는 샘플 이름
  - 2) **gene\_name** : 공식 유전자 심볼 형식 (HUGO)
  - 3) **target\_id** : Methylation target ID (예. cg00030074)
  - 4) **target\_type** : 샘플의 타입으로 아래 중 하나의 타입으로 지정
    - N : 정상(Normal)
    - T : 종양(Tumor)
  - 5) **gene\_vl** : Methylation 유전자 값 (예. 0.962545)

<b>sample_id</b>	<b>gene_name</b>	<b>target_id</b>	<b>target_type</b>	<b>gene_vl</b>
sample_1	MAP2K4	cg00030074	T	0.962545
sample_1	CREBBP	cg00039463	T	0.220806
sample_1	RB1	cg00059930	T	0.798067
sample_2	ALK	cg00073794	T	0.907596
sample_2	IGF1R	cg00098799	T	0.44016
sample_2	IGF1R	cg00105080	T	0.815186
sample_2	MYC	cg00163372	T	0.345549
sample_2	PTEN	cg00164720	T	0.97395
sample_3	IGF1R	cg00169856	T	0.152389
sample_3	ARID3B	cg00242443	T	0.907199
sample_3	CCND1	cg00248856	T	0.702758
sample_3	CDKN2B-AS1	cg00264650	T	0.325241
sample_3	LPAR6	cg00286388	T	0.919669
sample_4	CCND1	cg00347938	T	0.664497
sample_4	IGF1R	cg00348254	T	0.0373413
sample_4	ARID1A	cg00371107	T	0.0693479
sample_5	ARID1B	cg00372369	T	0.507513

< Methylation 파일 예시 >

#### ④ RNA

- . RNA 파일의 각 컬럼은 탭(tab)으로 분리된 형식입니다.
- . 데이터 값이 없는 경우 빈 칸으로 비워둡니다.
- . RNA 파일의 컬럼은 아래와 같이 구성되며 해당 양식에 맞게 파일을 구성하여야 합니다.
  - 1) **sample\_id** : 임상정보(Clinical information)와 일치하는 샘플 이름
  - 2) **gene\_name** : 공식 유전자 심볼 형식 (HUGO)
  - 3) **raw** : RNA expression level raw count 값
  - 4) **norm** : CPM, RPKM, FPKM, TPM과 같은 정규화된 값
  - 5) **target\_type** : 샘플의 타입으로 아래 중 하나의 타입으로 지정
    - N : 정상(Normal)
    - T : 종양(Tumor)

<b>sample_id</b>	<b>gene_name</b>	<b>raw</b>	<b>norm</b>	<b>type</b>
sample_1	FOXD4	0	-0.138123	T
sample_1	DMRT3	0	-0.148884	N
sample_1	IDUA	0.29	-0.164395	T
sample_2	CYP2W1	0	-0.298402	T
sample_2	CNN2	0	0	T
sample_2	RAD52	0	0	T
sample_2	RAD52	0	-0.156845	N
sample_2	HMHA1	16.82	-0.577061	T
sample_3	SLC12A7	30.63	-0.852351	N
sample_3	IDI1	1.36	-1.09422	T
sample_3	MUC2	20.47	-0.248175	N
sample_3	KCNQ1	16.51	-0.0918038	N
sample_3	KCNQ1	24.55	0.127891	T
sample_4	KCTD5	46.41	0.0306975	T
sample_4	PCED1A	22.67	-0.47865	T
sample_4	FOXM1	34.26	-0.339922	T
sample_5	ZNRF4	13.25	-0.759815	N
sample_5	DNAJB2	26.41	-0.33654	T
sample_6	DISP1	18.02	-0.474546	T
sample_7	DISP1	9.37	-0.494514	N
sample_7	ACBD3	12.74	-0.262462	T

< RNA 파일 예시 >

### ⑤ Fusion

- . Fusion 파일의 각 컬럼은 **탭(tab)**으로 분리된 형식입니다.
  - . Fusion 데이터의 경우 [STAR-Fusion Tool](#)을 사용하여 나온 결과 데이터를 사용하는 것을 권장합니다.
  - . **데이터 값이 없는 경우 빈 칸으로** 비워둡니다.
  - . Fusion 파일의 컬럼은 아래와 같이 구성되며 해당 양식에 맞게 파일을 구성하여야 합니다.
- 1) **sample\_id** : 임상정보(Clinical information)와 일치하는 샘플 이름
  - 2) **junction\_read\_count** : Splice junction에 존재하는 read 수
  - 3) **spanning\_frag\_count** : Spanning read 수
  - 4) **splice\_type** : Breakpoint splice junction이 reference로 알려진 유무에 따라 세 가지 타입으로 지정
    - ONLY\_REF\_SPLICER : Breakpoint splice junction이 참조주석에 의해 알려진 경우
    - INCL\_NON\_REF\_SPLICER : Breakpoint splice junction이 참조주석에 의해 알려지지 않은 경우
    - NO\_JUNCTION\_READS\_IDENITIFIED : Breakpoint splice junction read가 식별되지 않는 경우
  - 5) **left\_gene\_name** : 공식 유전자 심볼 형식 (HUGO)
  - 6) **left\_gene\_ensmbl\_id** : 왼쪽 유전자의 Ensmbl ID (예. ENSG00000160785)
  - 7) **left\_gene\_chr** : 왼쪽 유전자의 염색체 이름 (형식 : chr1, chr2, chr3, ...)
  - 8) **left\_hg38\_pos** : 왼쪽 유전자의 위치 (Reference human genome hg38 버전 기준)
  - 9) **right\_gene\_name** : 공식 유전자 심볼 형식 (HUGO)
  - 10) **right\_gene\_ensmbl\_id** : 오른쪽 유전자의 Ensmbl ID (예. ENSG00000160785)
  - 11) **right\_gene\_chr** : 오른쪽 유전자의 염색체 이름 (형식 : chr1, chr2, chr3, ...)
  - 12) **right\_hg38\_pos** : 오른쪽 유전자의 위치 (Reference human genome hg38 버전 기준)

sample_id	junction_read_count	spanning_frag_count	splice_type	left_gene_name	left_gene_ensmbl_id	left_gene_chr	left_hg38_pos	right_gene_name	right_gene_ensmbl_id	right_gene_chr	right_hg38_pos
sample_1	182	136	ONLY_REF_SPLICER	SLC25A44	ENSG00000160785	chr1	156164038	DAP3	ENSG00000132676	chr1	155679564
sample_1	82	129	ONLY_REF_SPLICER	MBP	ENSG00000197971	chr18	74778254	RP11-723G8.2	ENSG00000265352	chr18	69400083
sample_1	48	27	INCL_NON_REF_SPLICER	FCHSD2	ENSG00000137478	chr11	72598515	FAM168A	ENSG00000054965	chr11	73179537
sample_2	45	36	ONLY_REF_SPLICER	CCDC6	ENSG00000108091	chr10	61665880	ANK3	ENSG00000151150	chr10	61898845
sample_2	43	18	ONLY_REF_SPLICER	STARD10	ENSG00000214530	chr11	72492020	FCHSD2	ENSG00000137478	chr11	72613683
sample_2	33	15	ONLY_REF_SPLICER	SLC38A1	ENSG00000111371	chr12	46622936	COL2A1	ENSG00000139219	chr12	48373845
sample_2	20	28	ONLY_REF_SPLICER	SVIL	ENSG00000197321	chr10	29946898	MPP7	ENSG00000150054	chr10	28378770
sample_2	19	24	ONLY_REF_SPLICER	EWSR1	ENSG00000182944	chr22	29669853	ZNRF3	ENSG00000183579	chr22	29383064
sample_3	19	16	ONLY_REF_SPLICER	ITSN2	ENSG00000198399	chr2	24550921	OTOF	ENSG00000115155	chr2	26739467
sample_3	19	0	INCL_NON_REF_SPLICER	TSC2	ENSG00000103197	chr16	2098072	MAPK8IP3	ENSG00000138834	chr16	1809953
sample_3	12	18	ONLY_REF_SPLICER	FBXO34	ENSG00000178974	chr14	55738255	CDKN3	ENSG00000100526	chr14	54866612
sample_3	13	13	ONLY_REF_SPLICER	CAMTA1	ENSG00000171735	chr1	6885270	KCNAB2	ENSG00000069424	chr1	6100577
sample_3	12	15	ONLY_REF_SPLICER	CACHD1	ENSG00000158966	chr1	65145439	ROR1	ENSG00000185483	chr1	64516358
sample_4	10	9	ONLY_REF_SPLICER	NUP98	ENSG00000110713	chr11	3818630	BEAN1	ENSG00000166546	chr16	66511463
sample_4	12	1	ONLY_REF_SPLICER	TIMM23	ENSG00000138297	chr10	51606988	RP11-574K11.28	ENSG00000271816	chr10	75473090
sample_4	7	18	ONLY_REF_SPLICER	FBXO34	ENSG00000178974	chr14	55738255	CDKN3	ENSG00000100526	chr14	54868242

〈 Fusion 파일 예시 〉

## ⑥ Proteome

- . Proteome 파일의 각 컬럼은 탭(tab)으로 분리된 형식입니다.
- . 데이터 값이 없는 경우 빈 칸으로 비워둡니다.
- . Proteome 파일의 컬럼은 아래와 같이 구성되며 해당 양식에 맞게 파일을 구성하여야 합니다.
  - 1) **sample\_id** : 임상정보(Clinical information)와 일치하는 샘플 이름
  - 2) **gene\_name** : 공식 유전자 심볼 형식 (HUGO)
  - 3) **p\_name** : 단백질 이름 (예. Q9UIB8)
  - 4) **gene\_vl** : Proteome expression level raw count 값
  - 5) **z\_score** : CPM, RPKM, FPKM, TPM과 같은 정규화된 값
  - 6) **type** : 샘플의 타입으로 아래 중 하나의 타입으로 지정
    - N : 정상(Normal)
    - T : 종양(Tumor)

<b>sample_id</b>	<b>gene_name</b>	<b>p_name</b>	<b>gene_vl</b>	<b>z_score</b>	<b>type</b>
sample_1	CD84	Q9UIB8	1.1393	0.5998	N
sample_1	CD84	Q9UIB8	1.162	0.6827	T
sample_1	EID2	Q8N6I1			T
sample_2	PHF11	Q9UIL8	0.4881	-0.8251	N
sample_2	PHF11	Q9UIL8	0.6364	-0.6189	T
sample_3	FKBPL	Q9UIM3			N
sample_3	FKBPL	Q9UIM3			T
sample_3	GGT7	Q9UJ14	1.0235	-0.1713	N
sample_3	GGT7	Q9UJ14	1.6183	1.6373	T
sample_4	TSR3	Q9UJK0	2.2977	2.1824	N
sample_4	TSR3	Q9UJK0	1.691	1.0286	T
sample_4	ING1	Q9UK53	0.9283	-0.3197	N
sample_4	ING1	Q9UK53	0.9723	-0.2072	T
sample_5	FEM1B	Q9UK73	0.8406	-0.6165	N
sample_5	FEM1B	Q9UK73	1.3198	0.9298	T
sample_5	AKAP11	Q9UKA4	0.8926	-0.3825	N
sample_5	AKAP11	Q9UKA4	1.1959	0.3565	T

< Proteome 파일 예시 >

## ⑦ Phosphorylation

- . Phosphorylation 파일의 각 컬럼은 텐(tab)으로 분리된 형식입니다.
  - . 데이터 값이 없는 경우 빈 칸으로 비워둡니다.
  - . Phosphorylation 파일의 컬럼은 아래와 같이 구성되며 해당 양식에 맞게 파일을 구성하여야 합니다.
- 1) **sample\_id** : 임상정보(Clinical information)와 일치하는 샘플 이름
  - 2) **gene\_name** : 공식 유전자 심볼 형식 (HUGO)
  - 3) **site** : 인산화(phosphorylation)가 일어난 위치 (예. S462)
  - 4) **gene\_vl** : Proteome expression level raw count 값
  - 5) **z\_score** : CPM, RPKM, FPKM, TPM과 같은 정규화된 값
  - 6) **type** : 샘플의 타입으로 아래 중 하나의 타입으로 지정
    - N : 정상(Normal)
    - T : 종양(Tumor)

<b>sample_id</b>	<b>gene_name</b>	<b>site</b>	<b>gene_vl</b>	<b>z_score</b>	<b>type</b>
sample_1	CFL1	S24	0.0838	-1.4169	N
sample_1	CFL1	S24	1.1512	0.8161	T
sample_1	EIF4B	S462	0.8981	-0.539	N
sample_2	EIF4B	S462	0.8644	-0.587	T
sample_2	EIF4B	S462	0.643	-1.0705	N
sample_2	EIF4B	S462	0.7945	-0.5989	T
sample_2	EIF4B	Y609	0.8981	-0.539	N
sample_2	EIF4B	Y609	0.8644	-0.587	T
sample_3	EIF4B	T205	0.643	-1.0705	N
sample_3	EIF4B	T205	0.7945	-0.5989	T
sample_3	EIF4B	S207	0.8357	-0.5339	N
sample_3	EIF4B	S207	0.8904	-0.3523	T
sample_3	EIF4B	S497	0.6787	-0.8336	N
sample_4	EIF4B	S497	0.9062	-0.4061	T
sample_4	EIF4B	S498 S504	0.3511	-1.5972	N
sample_4	EIF4B	S498 S504	0.88	-0.216	T
sample_5	EIF4B	S498	0.7849	-0.5659	N
sample_5	EIF4B	S498	0.9126	-0.1953	T

< Phosphorylation 파일 예시 >