

# 다중데이터 시각화 분석

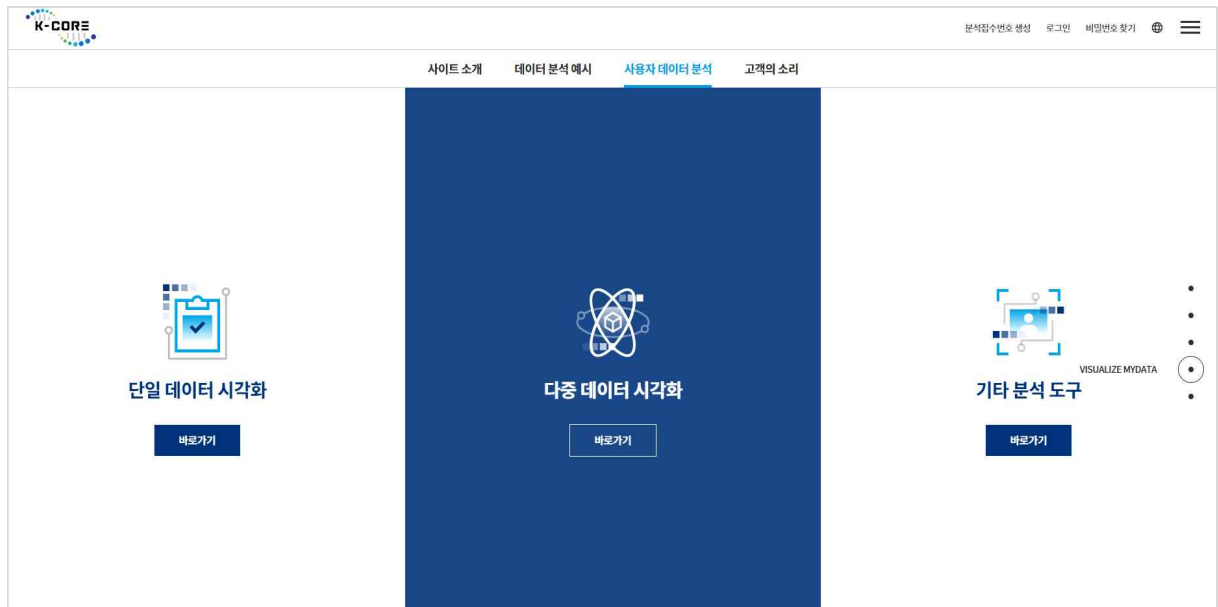
- 다중 데이터 시각화 메뉴는 암(질환) 임상 및 오믹스 데이터를 심층적으로 통계분석, 시각화할 수 있는 서비스를 제공합니다.
- 다중 데이터 시각화는 사용자가 업로드한 임상정보 및 오믹스 데이터 세트에 대해 프로젝트 형식으로 계정에 저장되며, 해당 프로젝트에 대해 조회/삭제를 할 수 있습니다.
- 전체 시각화 서비스에 대하여 사용자가 업로드한 데이터 세트에 대해 이용 가능한 기능만 활성화 됩니다.

다중 데이터 시각화 별 이용 가능한 데이터 종류

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

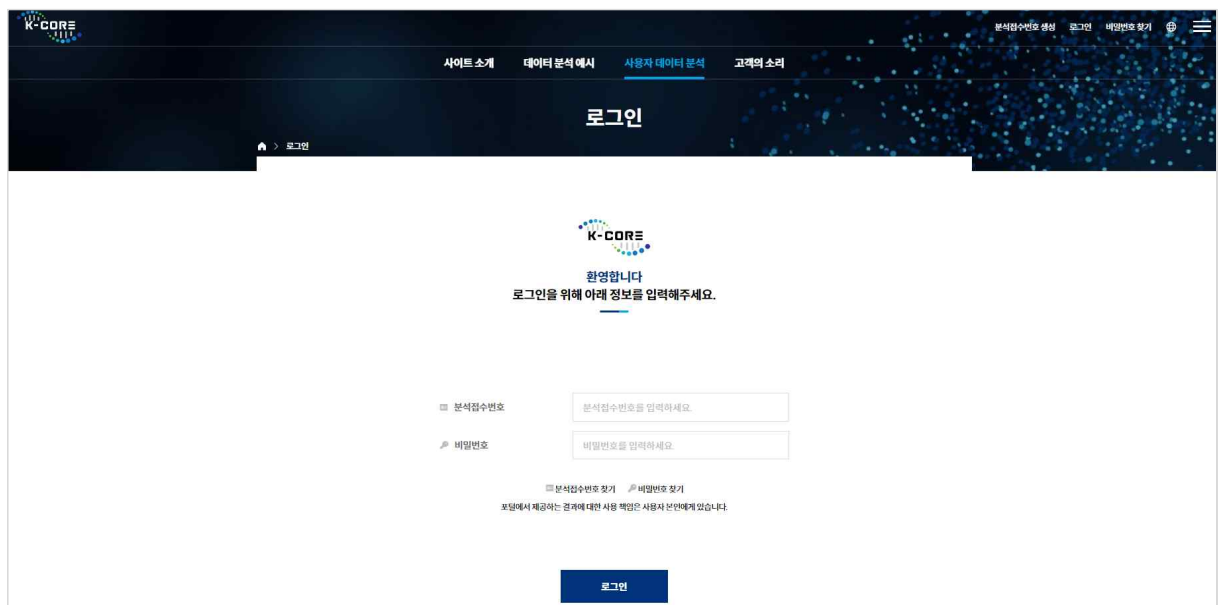
## 다중 데이터 시각화 프로젝트 생성 절차

### ① [사용자 데이터 분석] 메뉴의 다중 데이터 시각화



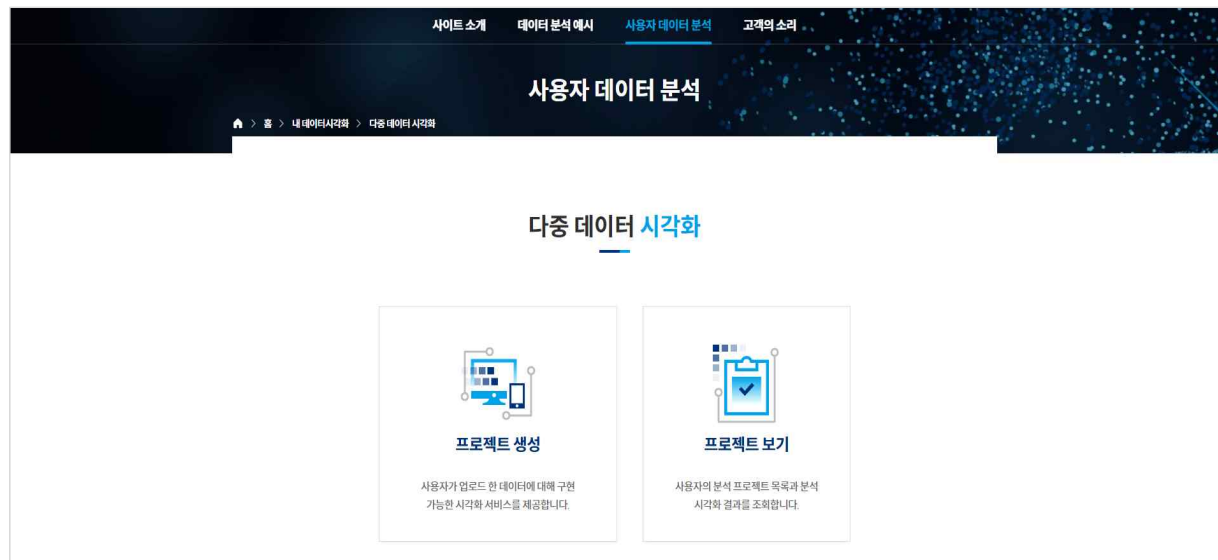
### ② 서비스 이용을 위한 로그인 페이지로 이동

. 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.



### ④ 다중 데이터 업로드

- **프로젝트 생성**을 통해 사용자 데이터 세트에 대한 업로드 페이지로 이동합니다.

- 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.

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

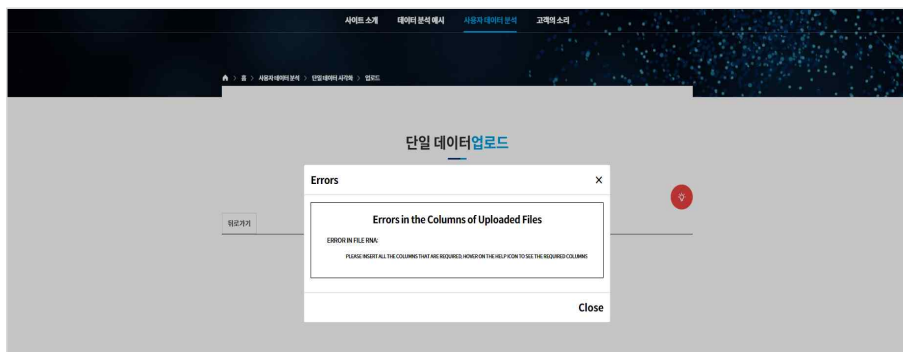
- . 업로드한 데이터에 대한 검증을 마치면 시각화 전 해당 데이터를 테이블 형태로 나타내어 줍니다.
- . 업로드한 파일에 이상이 없을 경우 [시각화] 버튼이 활성화 되고, 문제가 있을 경우 잘못된 데이터에 대해 안내 메시지를 제공합니다.

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

sample_id	time_vf	drink_yn	smoke_yn	fresh_liquid_yn	wt_score
sample_1	26.88	false	false		2
sample_2	19.17	false	false	true	3
sample_3	17.74	false	false	false	2
sample_4	23.77	false	false	false	2
sample_5	22.09	false	false	false	2
sample_6	23.28	false	false	false	3
sample_7	30.64	true	false	true	3
sample_8	22.89	false	false		3
sample_9	22.03				
sample_10	25.54	false	false	false	2

##### 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



- 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	value
11	10
11	10
11	10
11	10
11	10
11	10
11	10
11	10
11	10
11	10

## 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

**Box**

선택된 유전자에 유전정보 통계량  
박스 플롯 형태로 표현

**CNV**

동일 유전자 영역에 적용된 복제수  
변이 데이터 시각화

**Circos**

환형의 영역에 지도에 최대 7가지  
오믹스 데이터를 시각화

**Fusion**

선택된 샘플 그룹에 대한 Fusion  
gene 개수 및 엑스 정보 시각화

**Sankey**

유전자의 역할 연계 정보 시각화

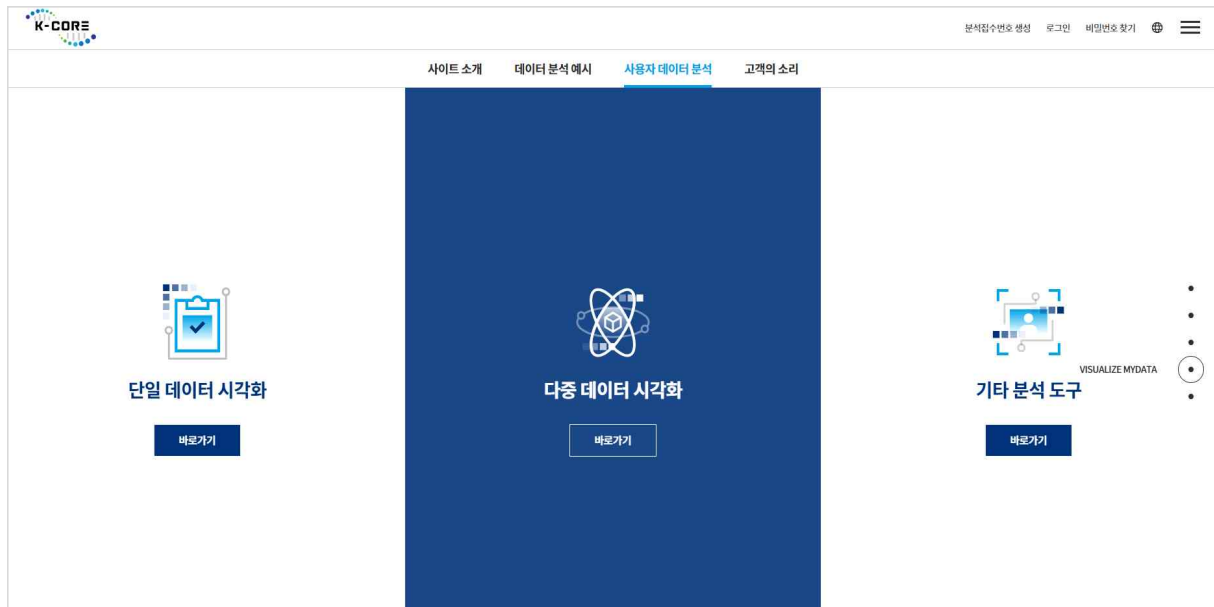
**Heatmap**

다양한 유전체 정보의 값을 색상  
다이어그램 형태로 표현

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

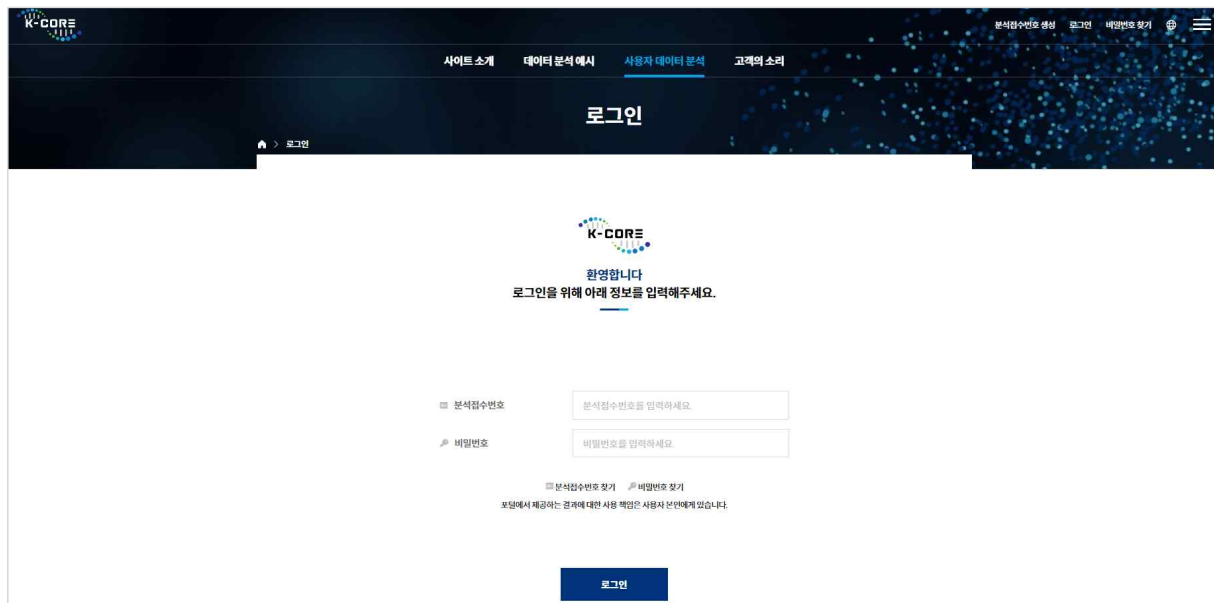
## 다중 데이터 시각화 프로젝트 보기

### ① [사용자 데이터 분석] 메뉴의 다중 데이터 시각화



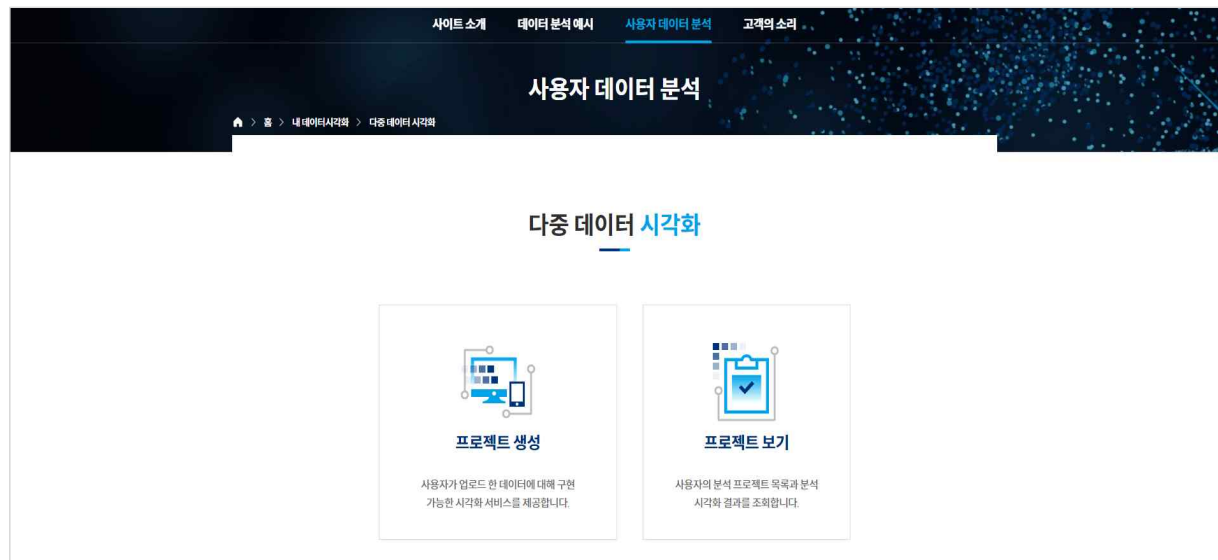
### ② 서비스 이용을 위한 로그인 페이지로 이동

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



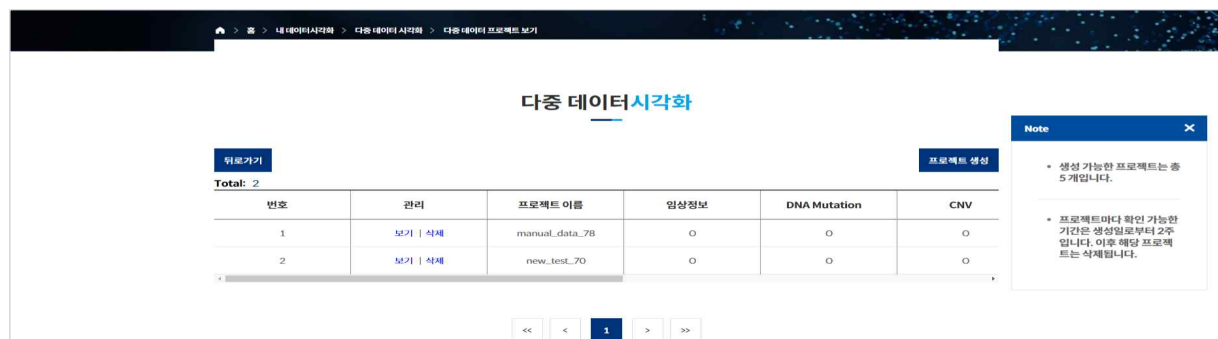
### ③ 프로젝트 보기 페이지 이동

- 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.



### ④ 프로젝트 보기 페이지 이동

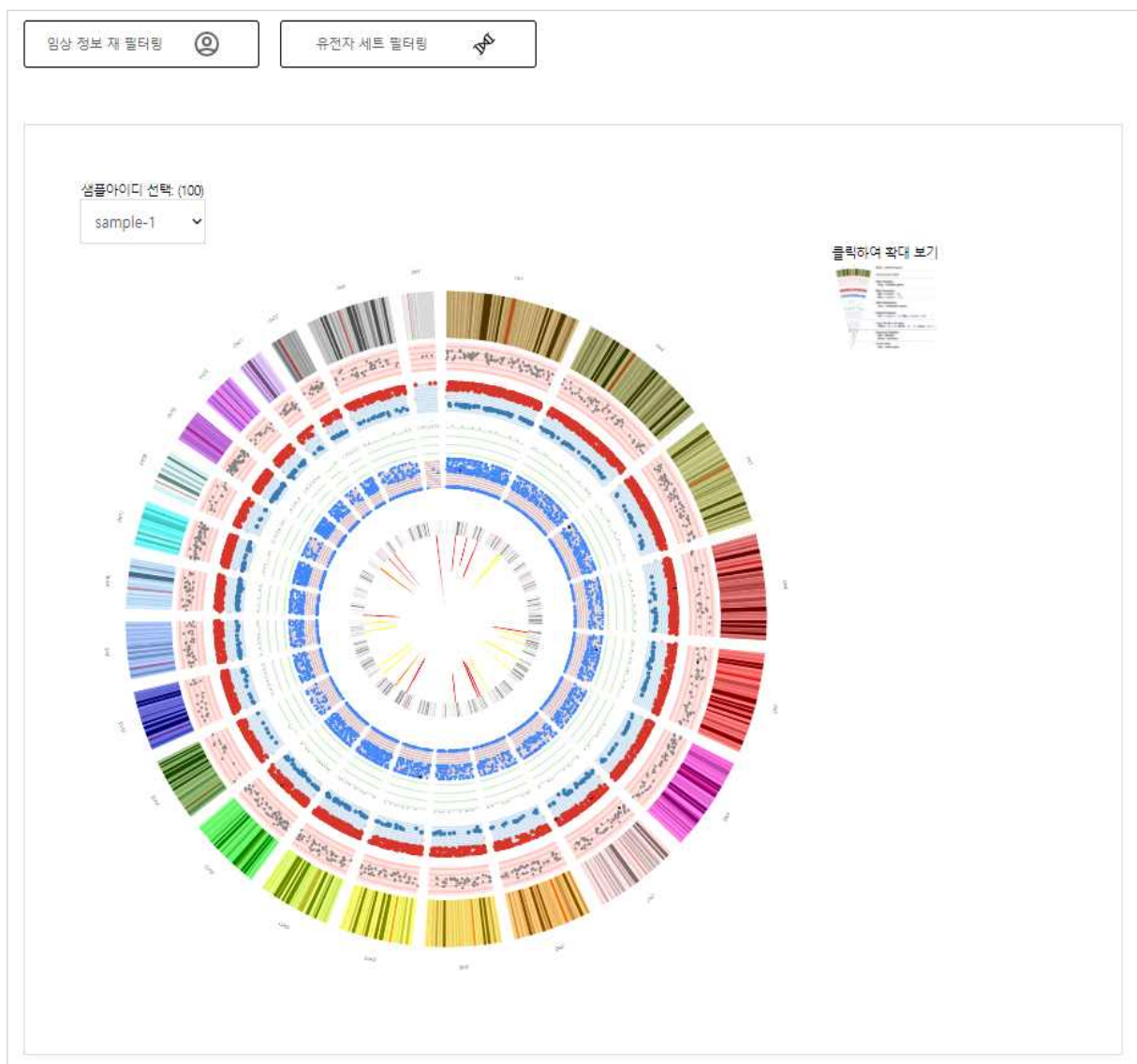
- 계정에서 생성한 프로젝트에 대해 조회 및 삭제할 수 있는 페이지입니다.
- 프로젝트 테이블에서 각 프로젝트에 대해 업로드한 데이터 정보를 확인할 수 있고, [관리] 컬럼을 통해 원하는 프로젝트를 보기/삭제 할 수 있습니다.
- 테이블 우측 위쪽에 [프로젝트 생성] 버튼을 통해 새로운 프로젝트를 생성할 수 있습니다.
  - 한 계정에서 생성할 수 있는 프로젝트는 최대 5개입니다.
  - 한 계정에서 5개의 프로젝트를 생성한 상태라면 최소 하나의 프로젝트를 삭제해야 새로운 프로젝트 생성이 가능
- 프로젝트마다 확인 가능한 기간은 생성일로부터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



## 다중 데이터 시각화 분석 서비스 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.



< 전체 오믹스 데이터에 대한 Circos Plot 예시 화면 >



. [임상 정보 재 필터링]을 통해 원하는 샘플 조건으로 필터링할 수 있습니다.

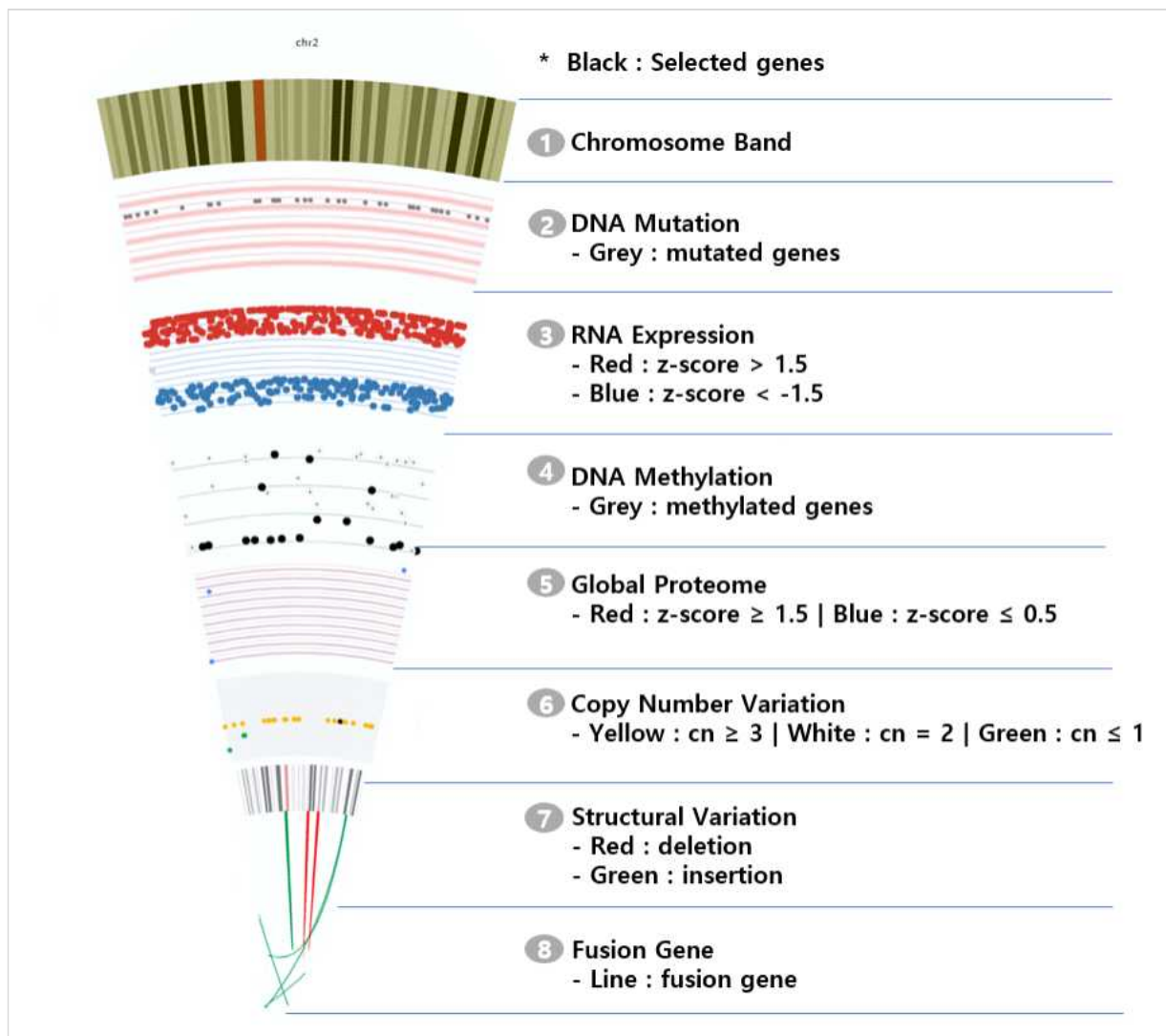
. [유전자 세트 필터링]을 통해 원하는 유전자를 선택할 수 있습니다.

- 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

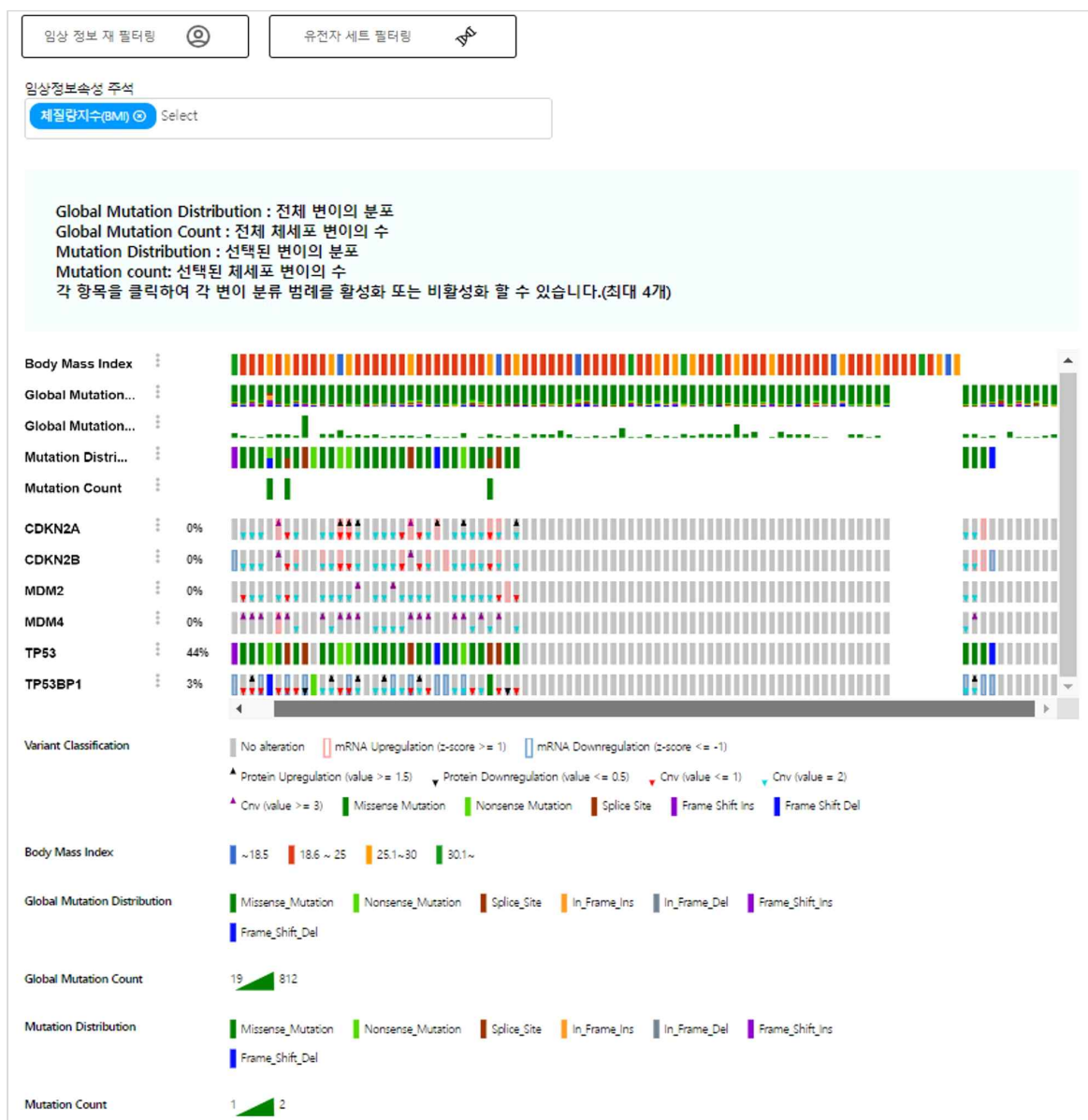
. OncoPrint는 각 환자의 유전자 내 DNA 돌연변이와 여러 오믹스 정보를 칸, 색상, 기호 등으로 시각화합니다.

. 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

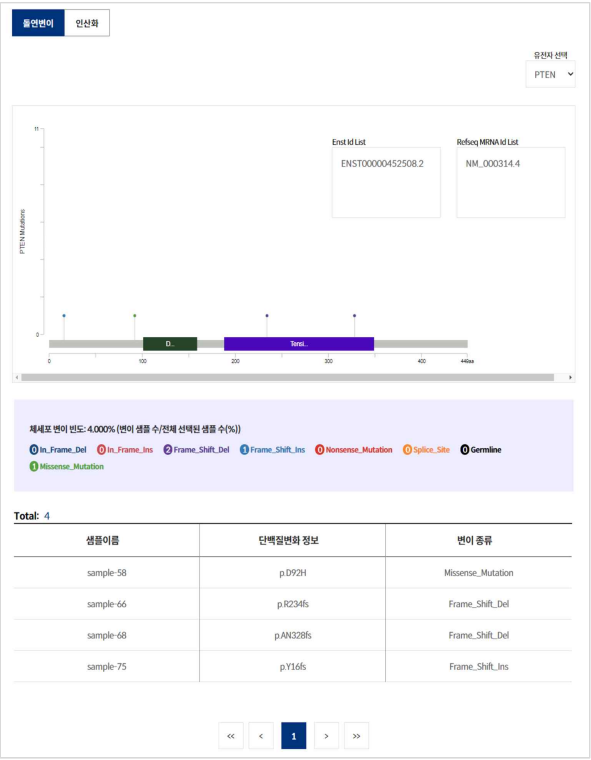
- . 시각화 가능한 오믹스 데이터 : DNA Mutation, Phosphorylation
- . Lollipop plot을 통해 특정 유전자의 돌연변이나 인산화를 아미노산 서열 위에서 파악할 수 있습니다.
- . [임상 정보 재 필터링]을 통해 원하는 샘플 조건으로 필터링할 수 있습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자를 선택할 수 있습니다.
- . [유전자 선택] 드롭다운 메뉴에서 원하는 유전자 하나를 선택합니다.
- . Lollipop plot은 돌연변이(Mutation)와 인산화(Phosphorylation)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



< Lollipop Plot - 돌연변이 화면>



< Lollipop Plot - 인산화 화면>

#### ④ Volcano Plot

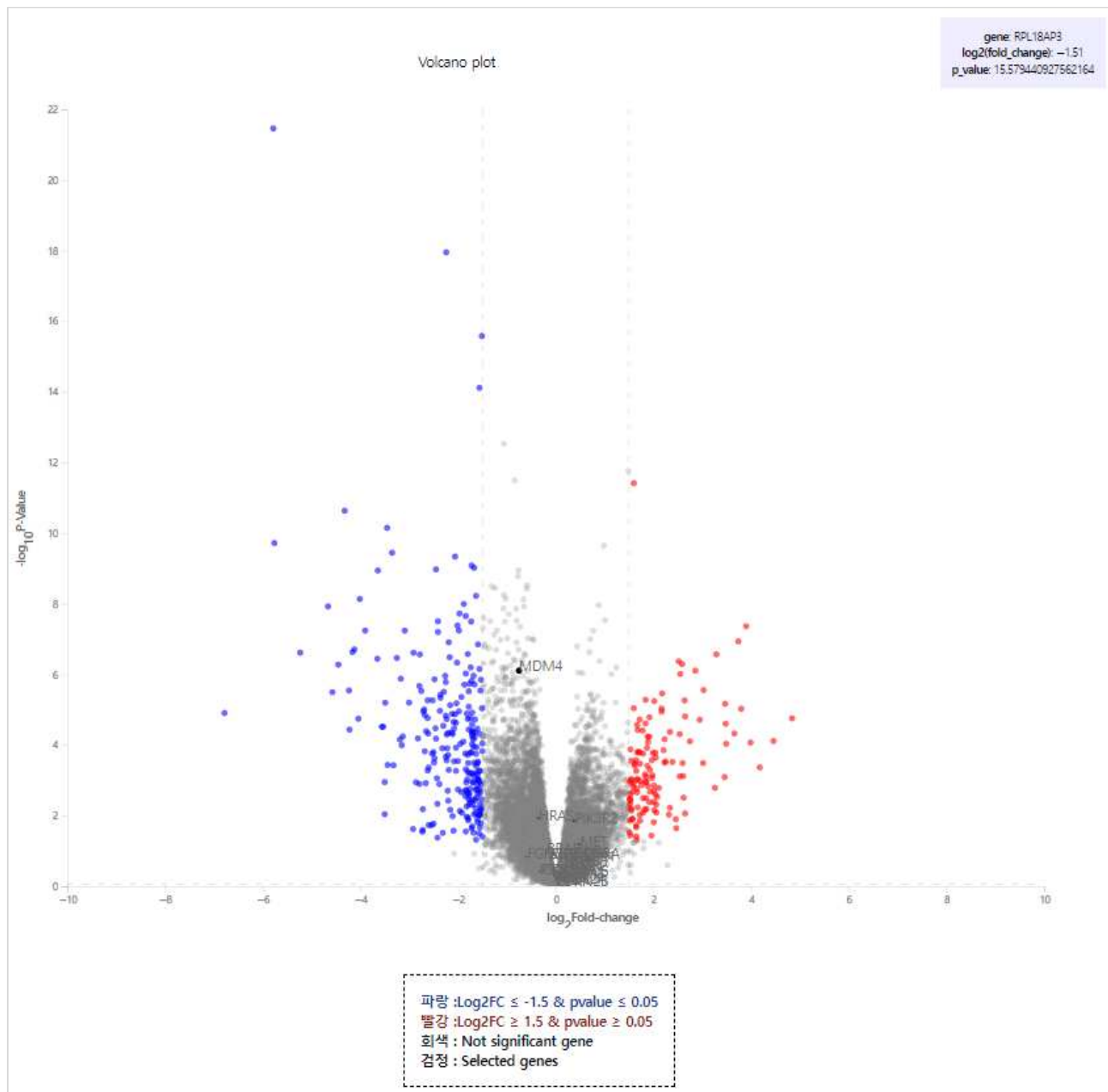
- . 시각화 가능한 오믹스 데이터 : Clinical Information, RNA, Proteome
- . Volcano plot은 임상 정보 조건에 연관된 유의미한 바이오마커 유전자를 파악할 때 유용합니다.
- . Volcano plot은 임상 정보 조건에 따라 구분된 두 그룹 간에 발현 차이를 보이는 유의미한 유전자 (Differentially Expressed Genes, DEG)를 시각화합니다.
- . [유전자 세트 필터링]을 통해 선택한 유전자는 플롯에서 **검은색 점**으로 표시됩니다.
- . 설정한 조건에 대해 UP/DOWN level 유전자가 표시됩니다.
  - **UP** regulation level gene :  $\text{Log2FC} \geq 1.5$  &  $\text{P-value} \leq 0.05$  (**빨간색 점**)
  - **Down** regulation level gene :  $\text{Log2FC} \leq -1.5$  &  $\text{P-value} \leq 0.05$  (**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 단백질체 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) 단백질체(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.



	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	CHRD12	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	SFTPD	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	RUNDC3A	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 단백질 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.

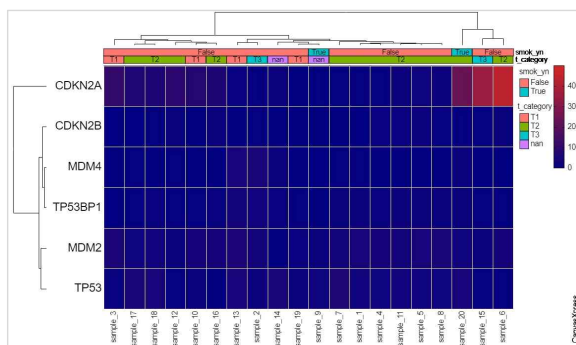
. Heatmap은 RNA, Methylation, Proteome, Phospho 네 가지 탭으로 구성됩니다.

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

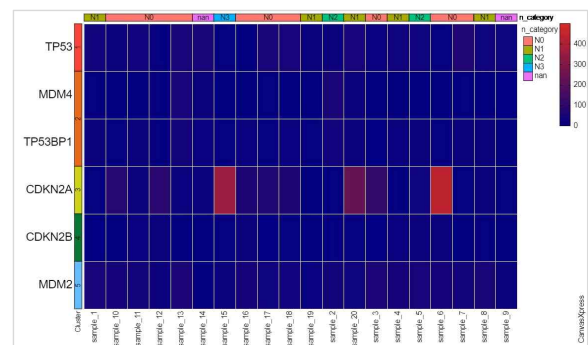
- [히트맵 종류 선택]은 RNA를 선택

- X축 : 샘플 ID

- Y축 : 선택한 유전자(세트)



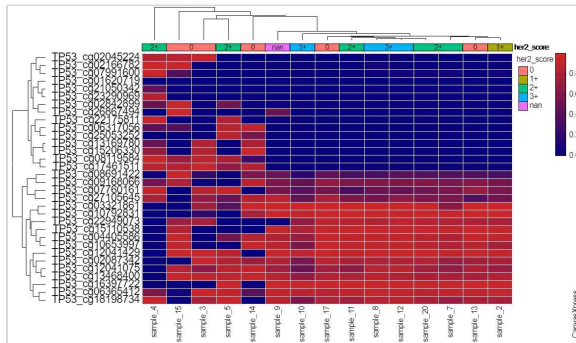
< RNA - Heatmap >



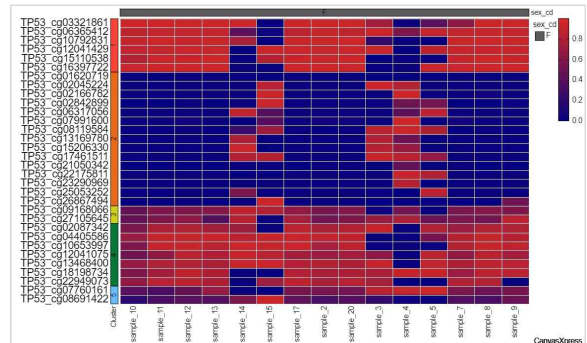
< RNA - K-mean >

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

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



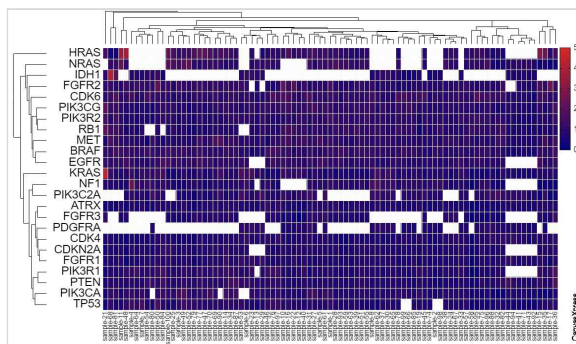
< Methylation - K-mean >



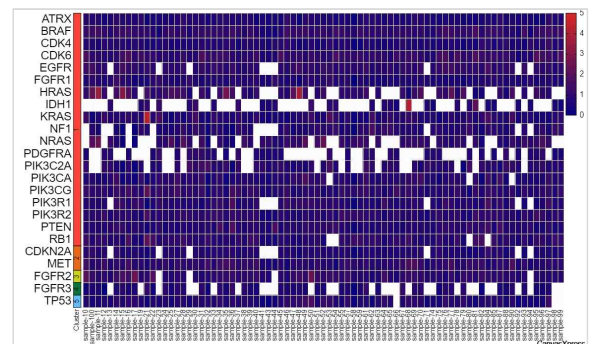
< Methylation - K-mean >

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

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



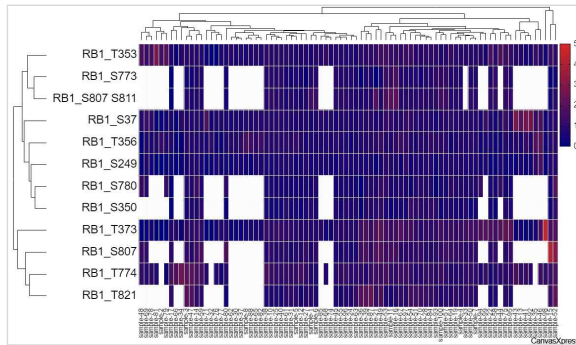
< Proteome - Heatmap >



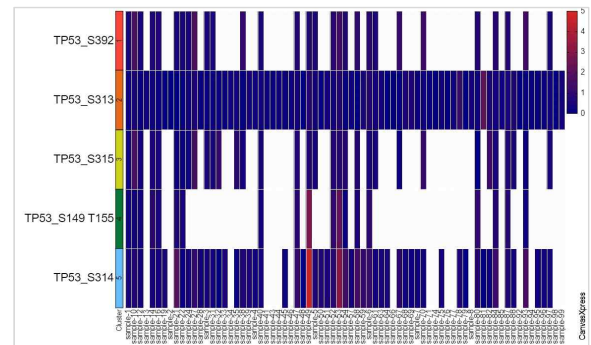
< Proteome - K-mean >

#### 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 >



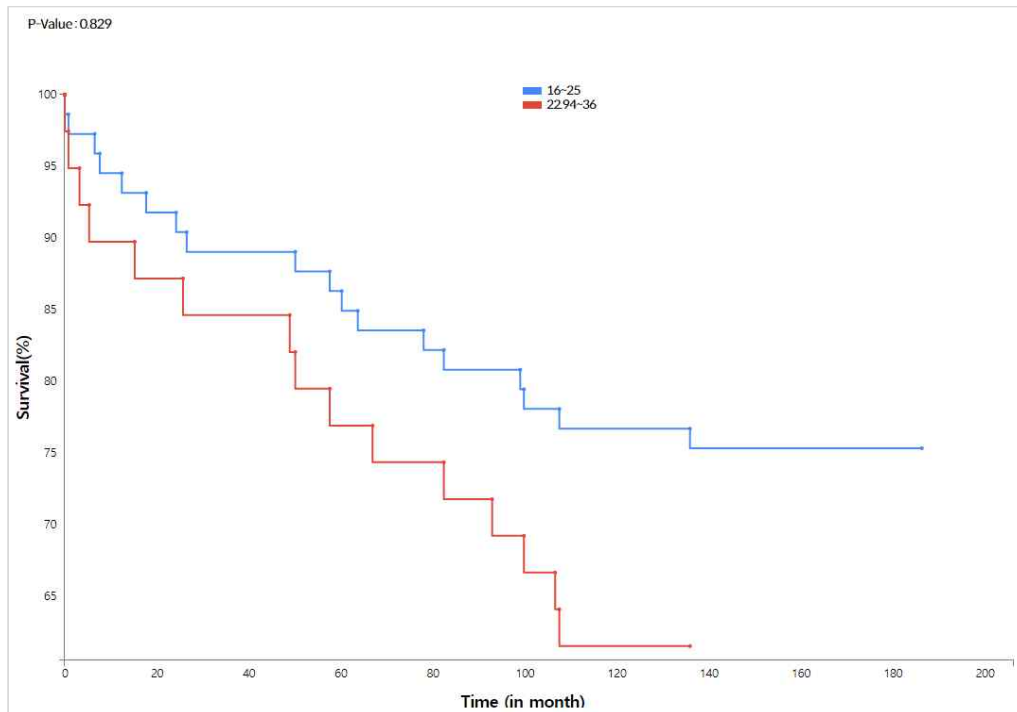
< Phosphorylation - K-mean >

## 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(rps\_yn)**, **재발시간(rps\_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



< 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 > >>

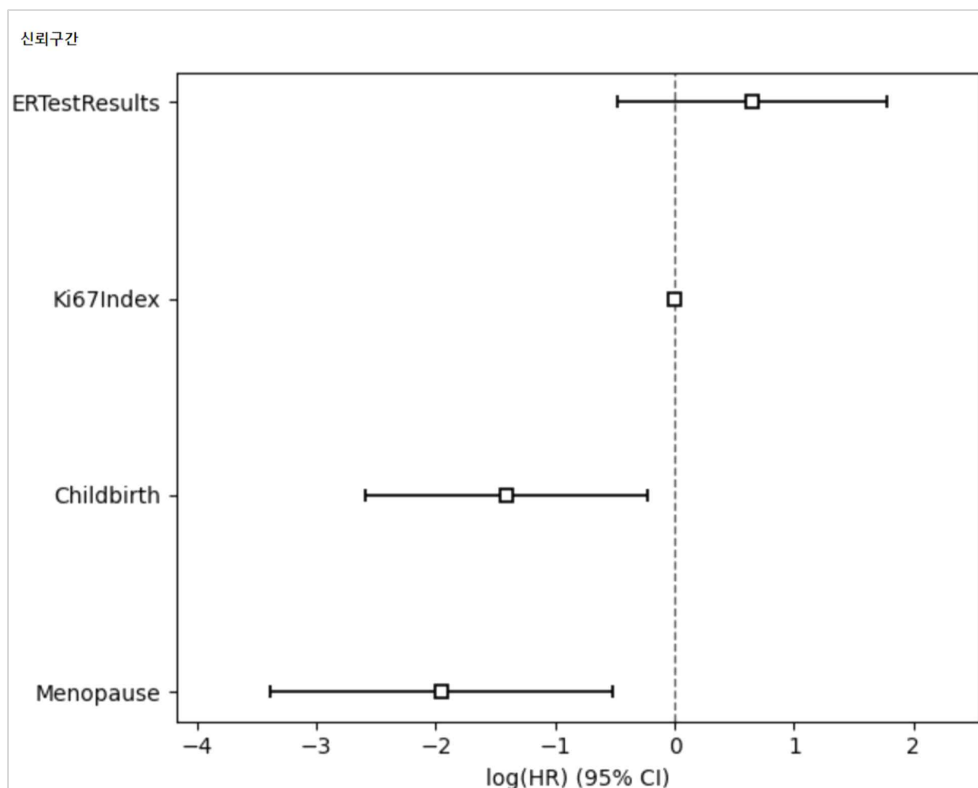
< Survival Plot - 그룹 정보 테이블>

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) lower 95%	exp(coef) upper 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

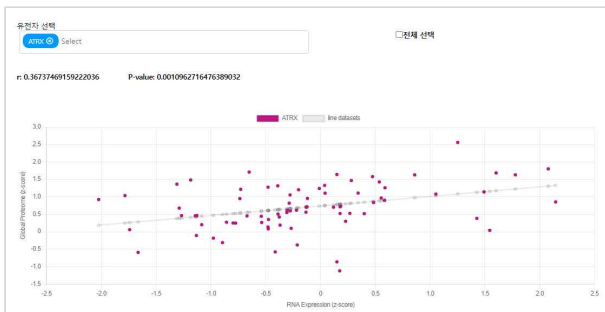
< Cox Regression. Co-efficient table >



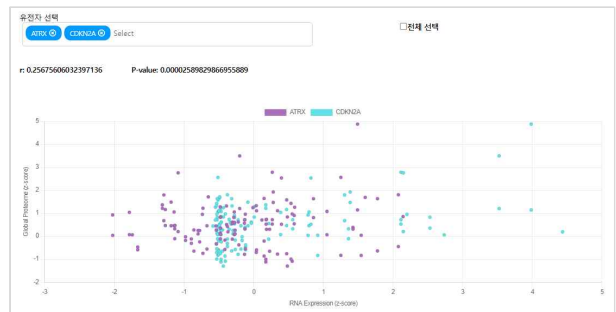
< Cox Regression. Confidence Interval, 95% CI >

## ⑦ Correlation

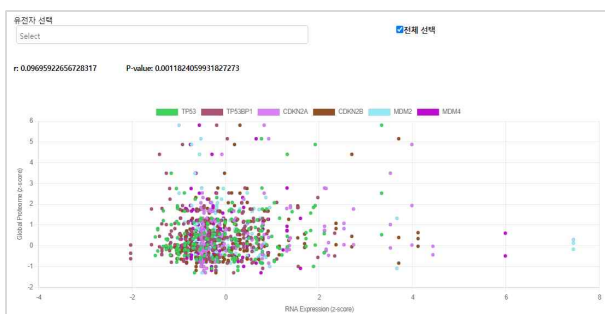
- . 시각화 가능한 오믹스 데이터 : RNA, Proteome
- . Correlation Plot은 선택된 유전자에 대하여 RNA 발현 값과 Proteome 풍부도 값의 상관관계를 점그래프 (Scatter Plot) 형태로 시각화합니다.
- . Correlation Plot을 통해 특정 유전자에 대한 RNA 발현과 단백질 발현 간의 상관관계 확인할 수 있습니다.
- . [임상 정보 재 필터링]을 통해 원하는 샘플 조건으로 필터링할 수 있습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자에 대한 Correlation Plot을 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 >



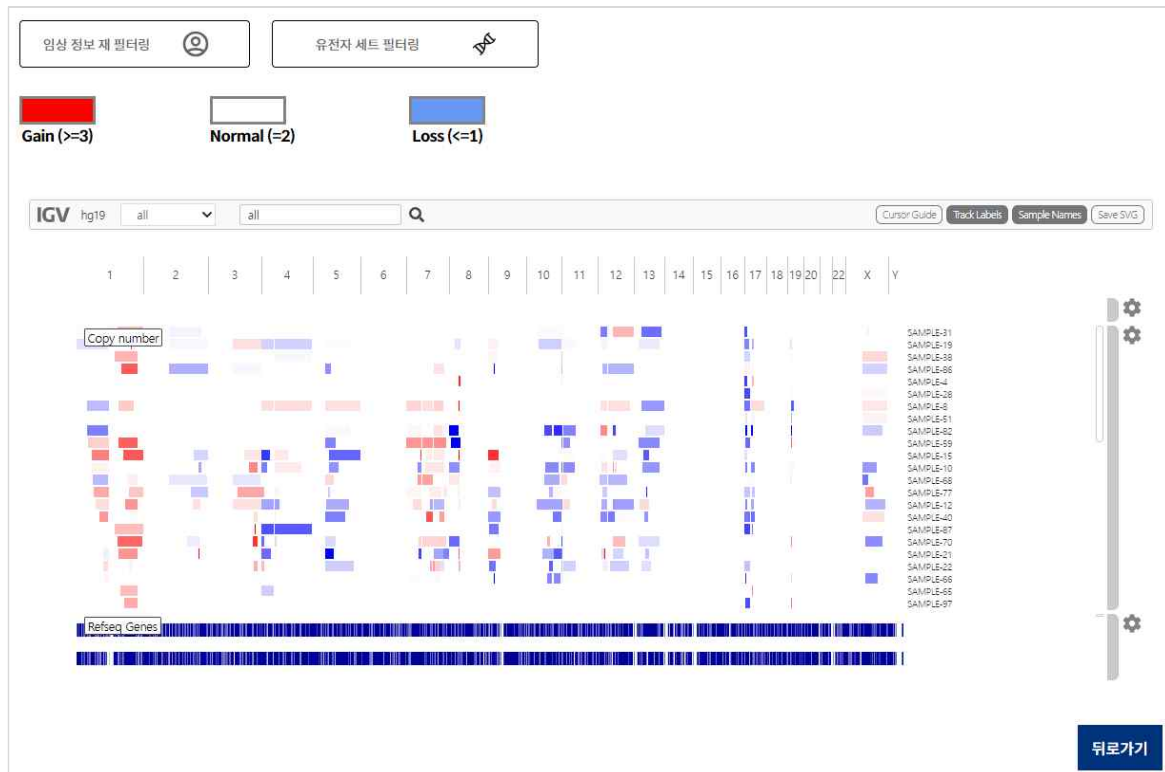
< Correlation. 다중 유전자 선택 >



< 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 ( $CN \leq 1$ )



. RefSeq Genes 트랙 (\*확대 시, 상세 보기 가능)

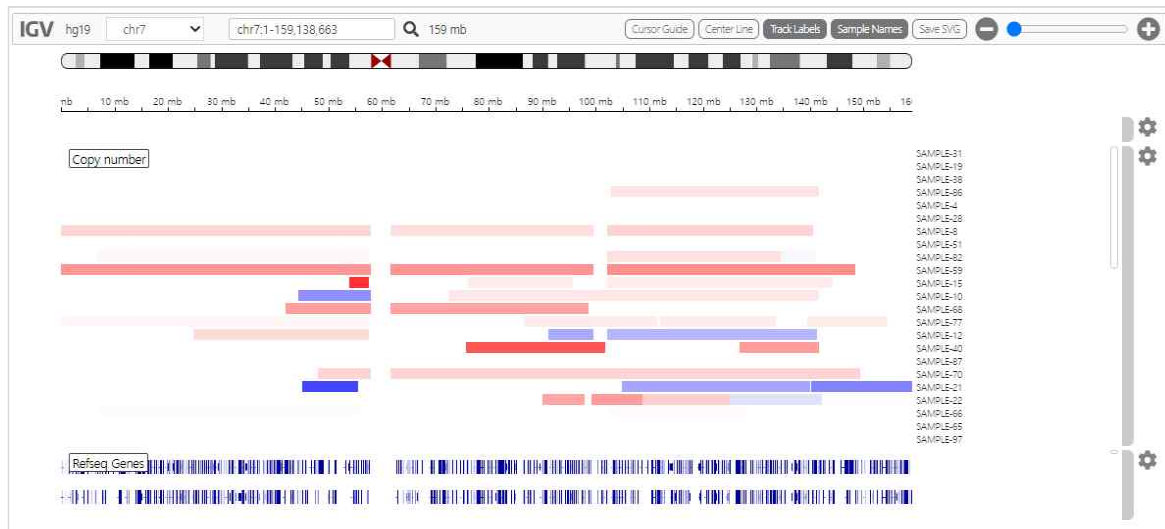
. 옵션 기능들

- 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

## ⑨ Box Plot

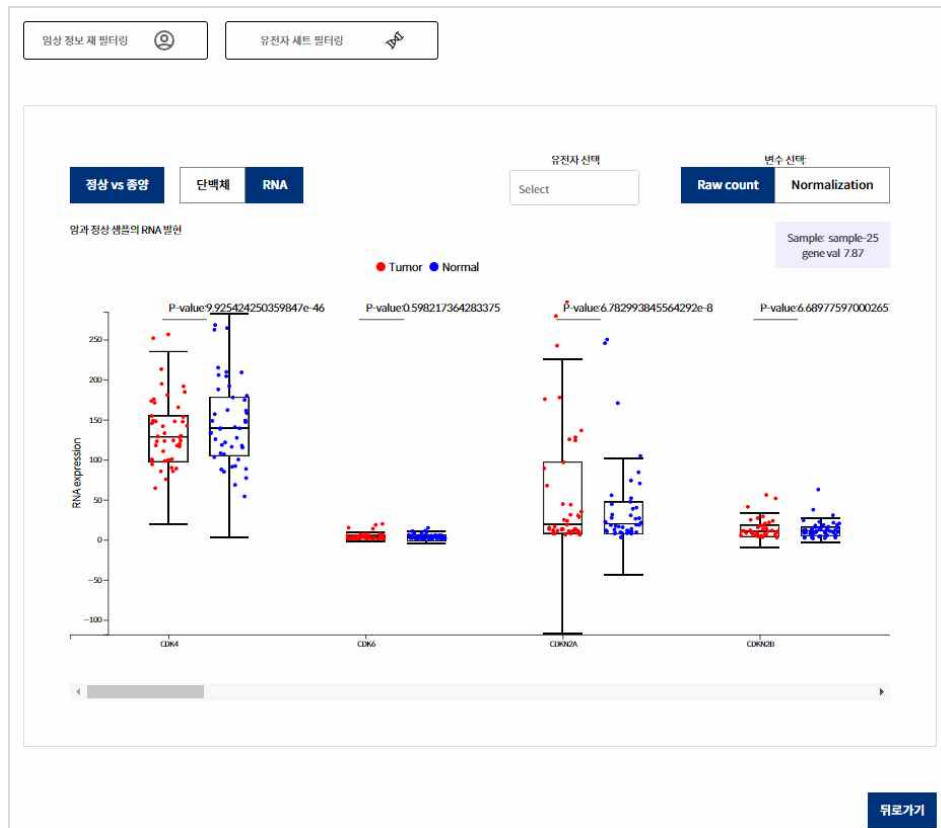
- . 시각화 가능한 오믹스 데이터 : RNA, Proteome
- . Box plot은 선택된 유전자의 단백질 풍부도 또는 RNA 발현 값 데이터의 기본 통계를 Box plot 형태로 보여줍니다. 최솟값, 최댓값, 1 사분위 수, 3 사분위 수, 이상치(outlier) 등을 나타냅니다.
- . [임상 정보 재 필터링]을 통해 원하는 샘플 조건으로 필터링할 수 있습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자에 대한 Box Plot을 그릴 수 있습니다.
- . 유전자 세트 필터링한 데이터 중 [유전자 선택]을 통해 원하는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



< Box plot - 단백질체(Proteome) >

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

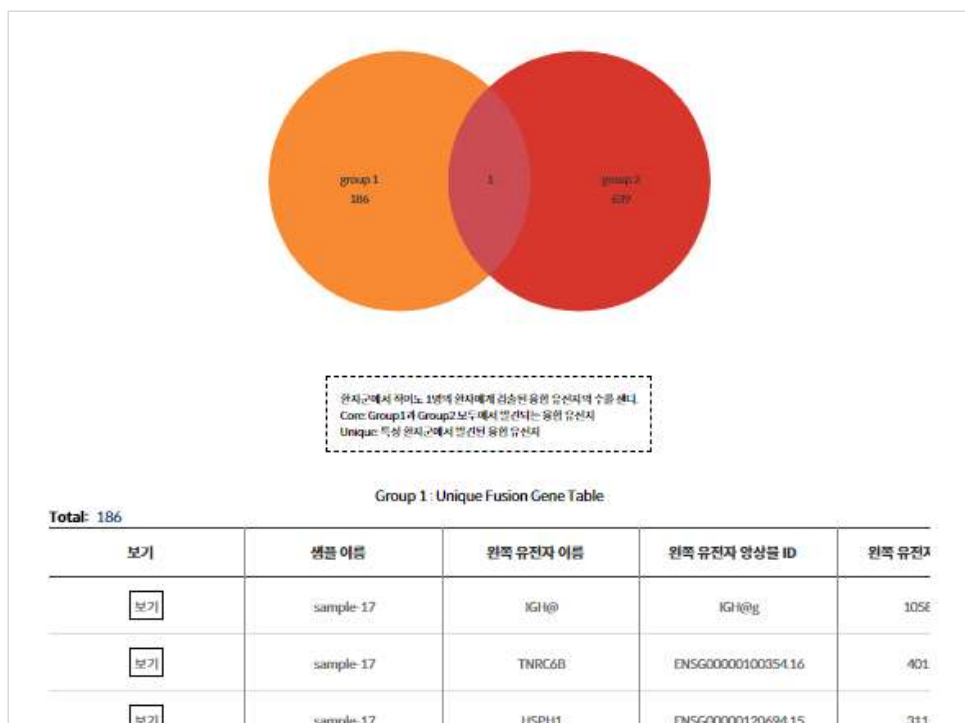
- [플롯 종류]는 RNA를 선택
- X축 : 선택한 유전자
- Y축 : 선택한 유전자의 샘플 타입 별 유전자



< Box plot - 전사체(RNA) >

## 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

→ 영역 구분 기준 : Left Gene Name, Right Gene Name, Left Breakpoint, Right Breakpoint가 일치할 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 오른쪽 유전자 이름

- Right Ensembl ID : 오른쪽 유전자의 Ensembl ID

- Right Breakpoint : 오른쪽 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

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-score  $\geq 1$ , Low : z-score  $\leq -1$ , Intermediate : 나머지

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

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

Patient Summary
Drug Relation Plot

Choose a Sample: (100)  
sample-10

Basic Information

Total: 5

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

<<
<
1
>
>>

Genomic Information

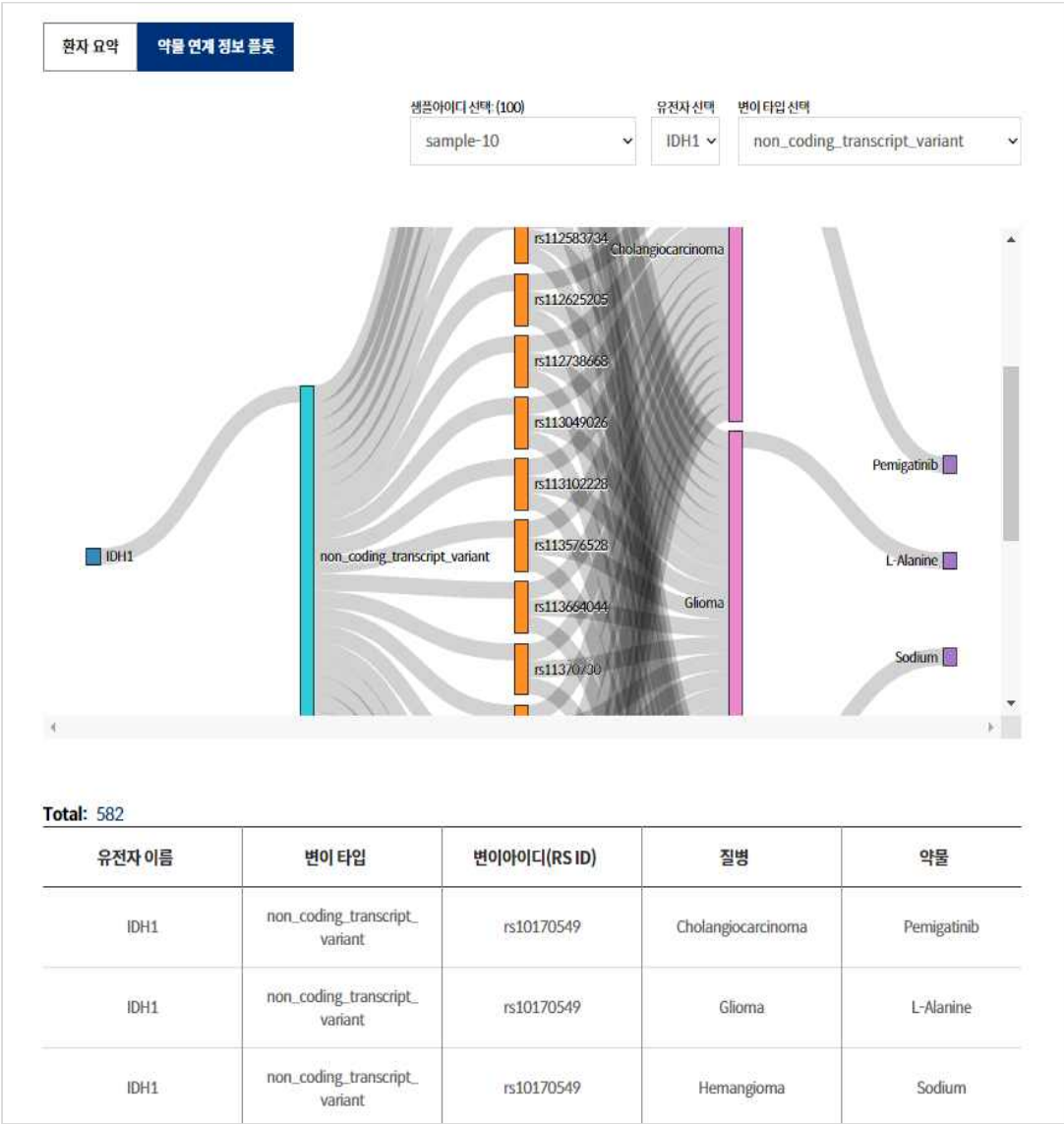
Total: 28

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)





< 약물 연계 정보 플롯(Drug Relation Plot) >