

단일데이터 시각화 분석

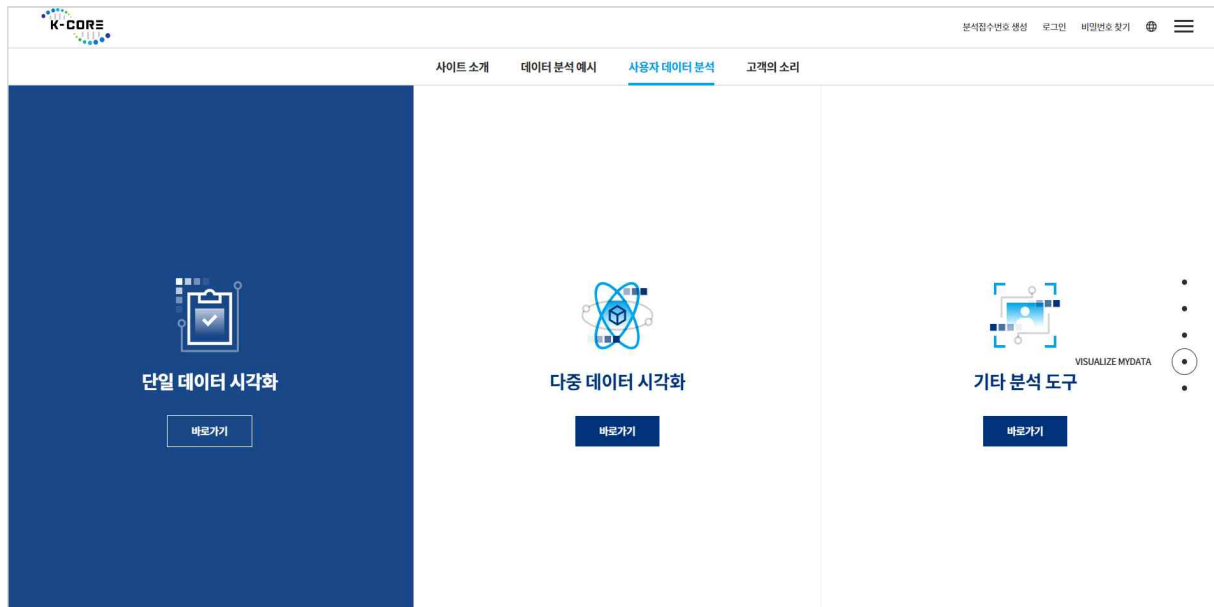
- 단일 데이터 시각화는 데이터 종류에 따라 하나의 임상 및 오믹스 데이터를 심층적으로 통계분석, 시각화할 수 있는 서비스를 제공합니다.
- 단일 데이터 시각화는 일회성 분석 서비스로 시각화 결과 화면에서 이동하면 해당 결과는 저장되지 않습니다.
- 시각화 결과를 프로젝트 형식으로 저장하고 열람(조회) 기능을 사용하려면 [다중 데이터 시각화] 서비스를 사용하여야 합니다.
- 각각의 단일 데이터 시각화 분석에 따라 사용 가능한 데이터가 다르게 구성됩니다.

단일 데이터 시각화 별 이용 가능한 데이터 종류

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

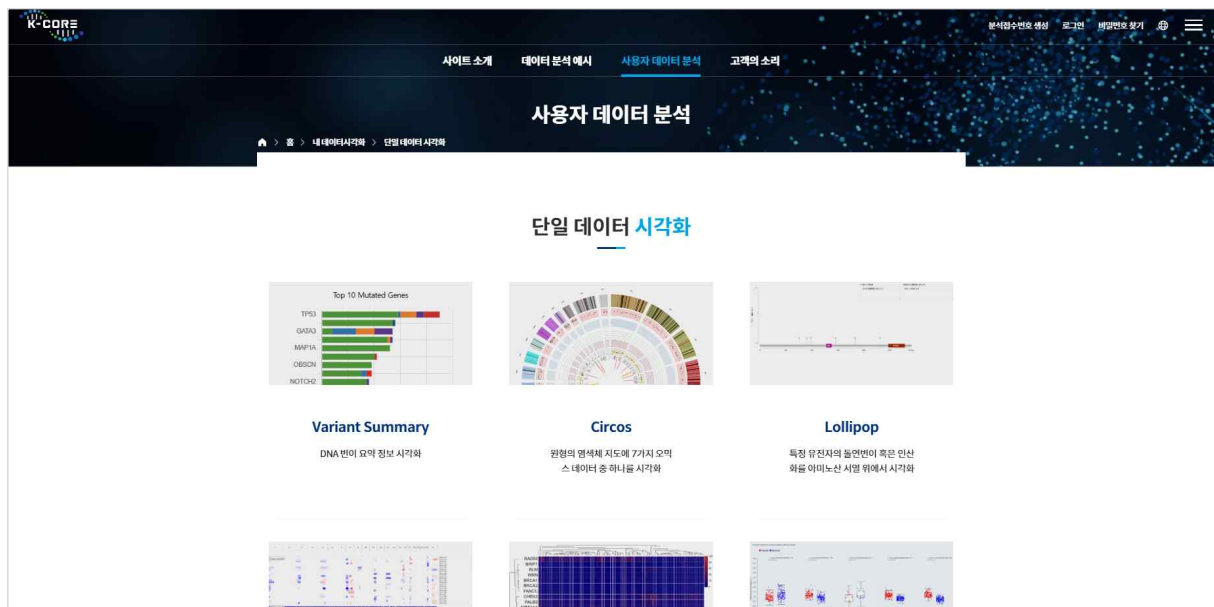
단일 데이터 시각화 절차

① [사용자 데이터 분석] 메뉴의 단일 데이터 시각화



② 카드섹션 페이지에서 원하는 시각화 서비스를 선택

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



③ 서비스 이용을 위한 로그인 페이지로 이동

- . 사용자 데이터 분석 메뉴는 로그인 후 사용 가능한 서비스 (이미 로그인된 상태라면 바로 데이터 업로드 화면으로 이동)

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

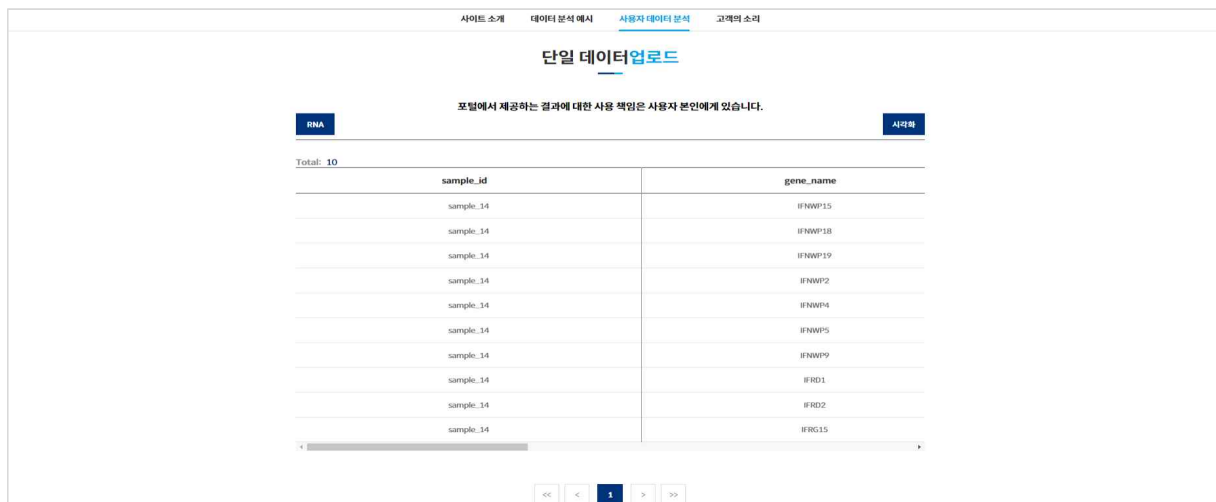
- . 데이터 업로드 박스 우측 상단의 전구모양 버튼을 누르면 해당 데이터에 대한 안내 팝업이 나타남
- . 안내 팝업 하단의 **Download**를 통해 해당 데이터에 대한 .tsv 형식의 샘플 파일 다운로드 가능

⑤ 데이터 업로드 후 [업로드] 버튼을 누르면 **QC 화면**으로 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

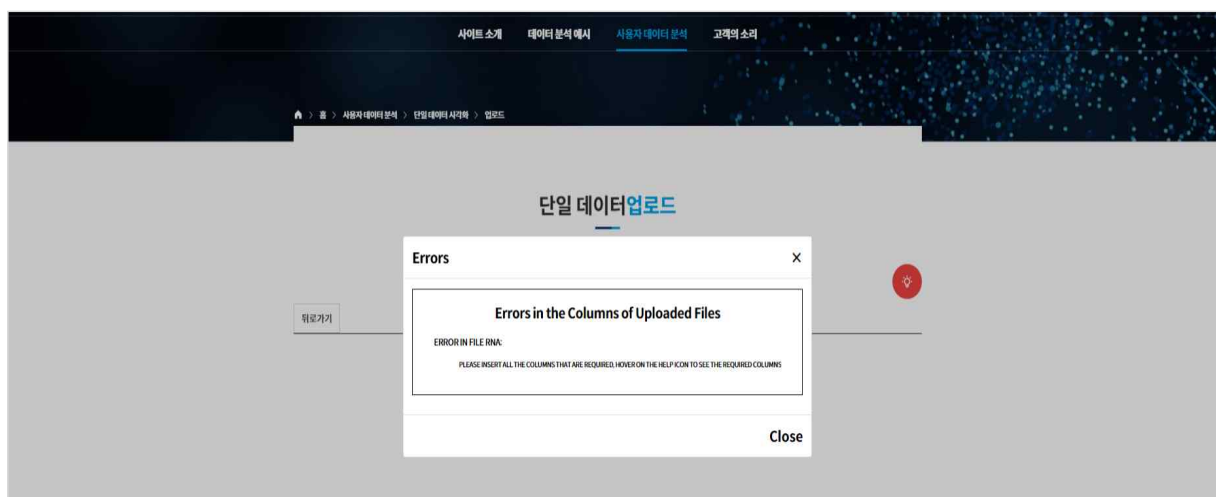


sample_id	gene_name
sample_14	IFNWP15
sample_14	IFNWP18
sample_14	IFNWP19
sample_14	IFNWP2
sample_14	IFNWP4
sample_14	IFNWP5
sample_14	IFNWP9
sample_14	IFRD1
sample_14	IFRD2
sample_14	IFRG15

- 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

사이트 소개 데이터 분석 예시 **사용자 데이터 분석** 고객문의 소리

단일 데이터 업로드

잘못된 데이터에 대해 빨간색으로 표시합니다.
포털에서 제공하는 결과에 대한 사용 책임은 사용자 본인에게 있습니다.

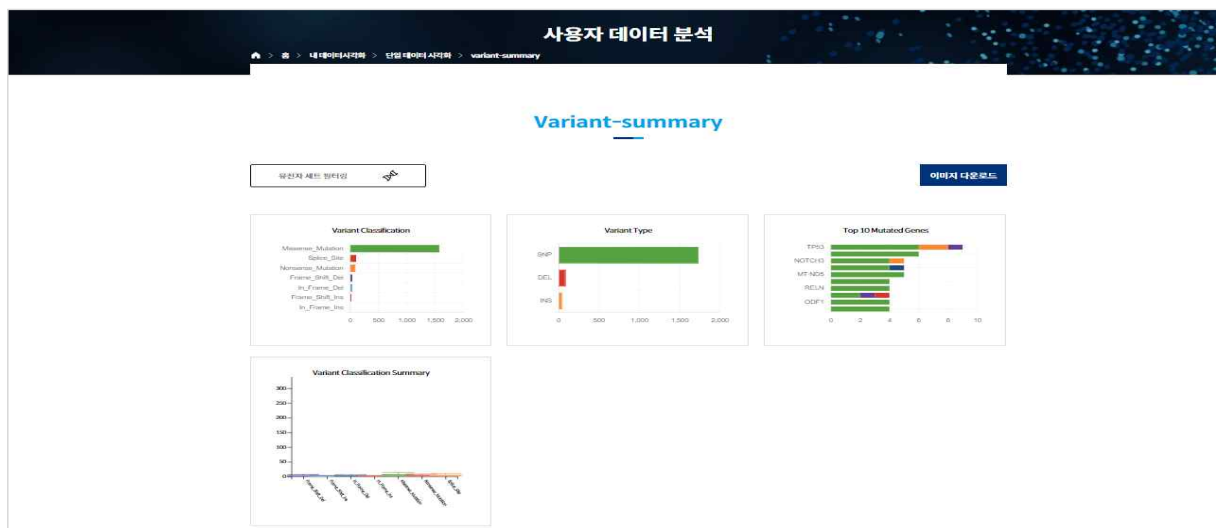
뒤로가기 RNA

gene_name	RAW
II	e
II	a
JJ	b
JJ	c
KK	d
KK	e
LL	f
LL	a
MM	b
MM	c

<< < 1 2 3 > >>

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 분석 서비스 안내

① Variant Summary

. 시각화 가능한 오믹스 데이터 : DNA Mutation

. Variant Summary 도구를 통해 샘플에 대한 변이의 요약 통계 정보를 얻을 수 있습니다.

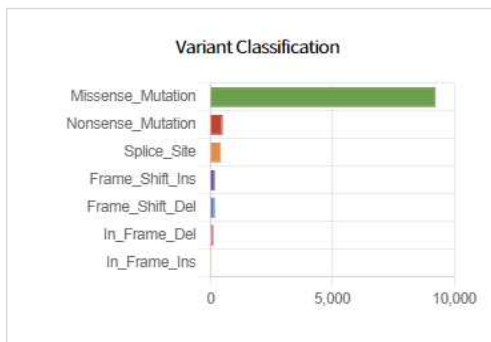
. 사용자 DNA Mutation 오믹스 데이터를 통해 변이 분류(Classification), 종류(Type)에 대한 요약된 통계 정보를 시각화하여 제공합니다. (차트에 대해 마우스 오버시, 해당 정보에 대한 라벨을 확인 가능)

1) **Variant Classification** : 전체 DNA Mutation 오믹스 데이터에 대한 주요 7가지 변이 통계 그래프

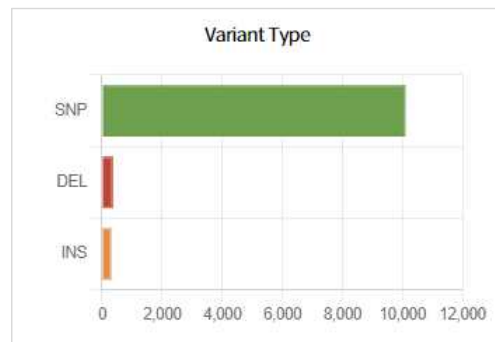
2) **Variant Type** : 전체 DNA Mutation 오믹스 데이터에 대한 변이 타입 통계 그래프

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)

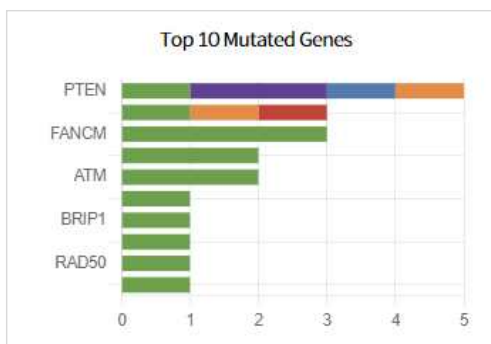
4) **Variant Classification Summary** : 전체 DNA Mutation 오믹스 데이터에 대한 주요 7가지 변이에 대한 박스형 그래프



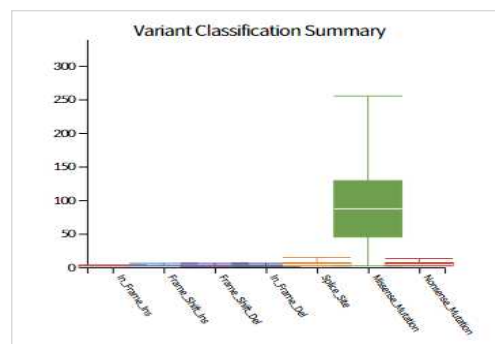
< Variant Classification >



< Variant Type >



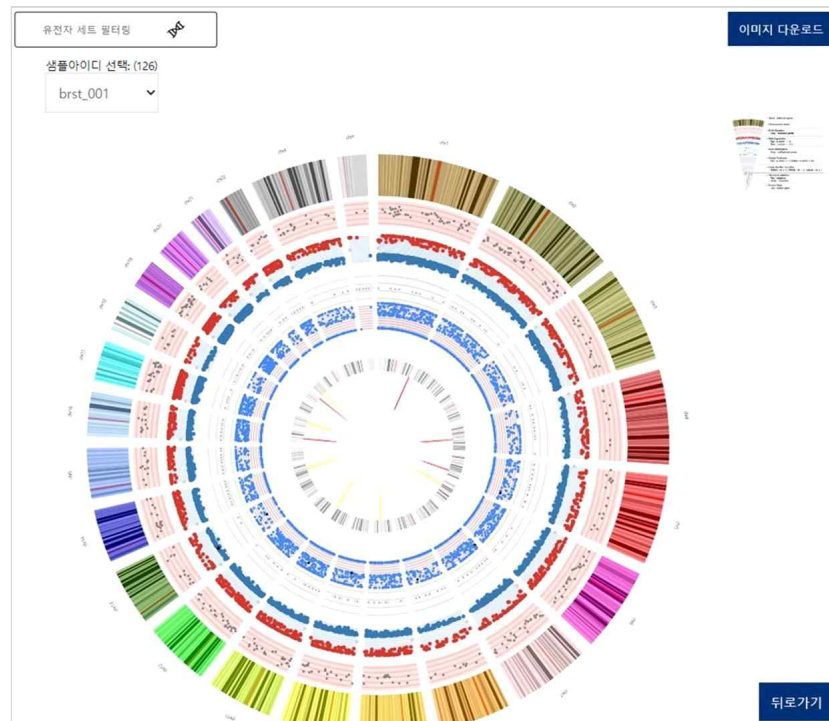
< Top 10 Mutated Genes >



< Variant Classification Summary >

② Circos Plot

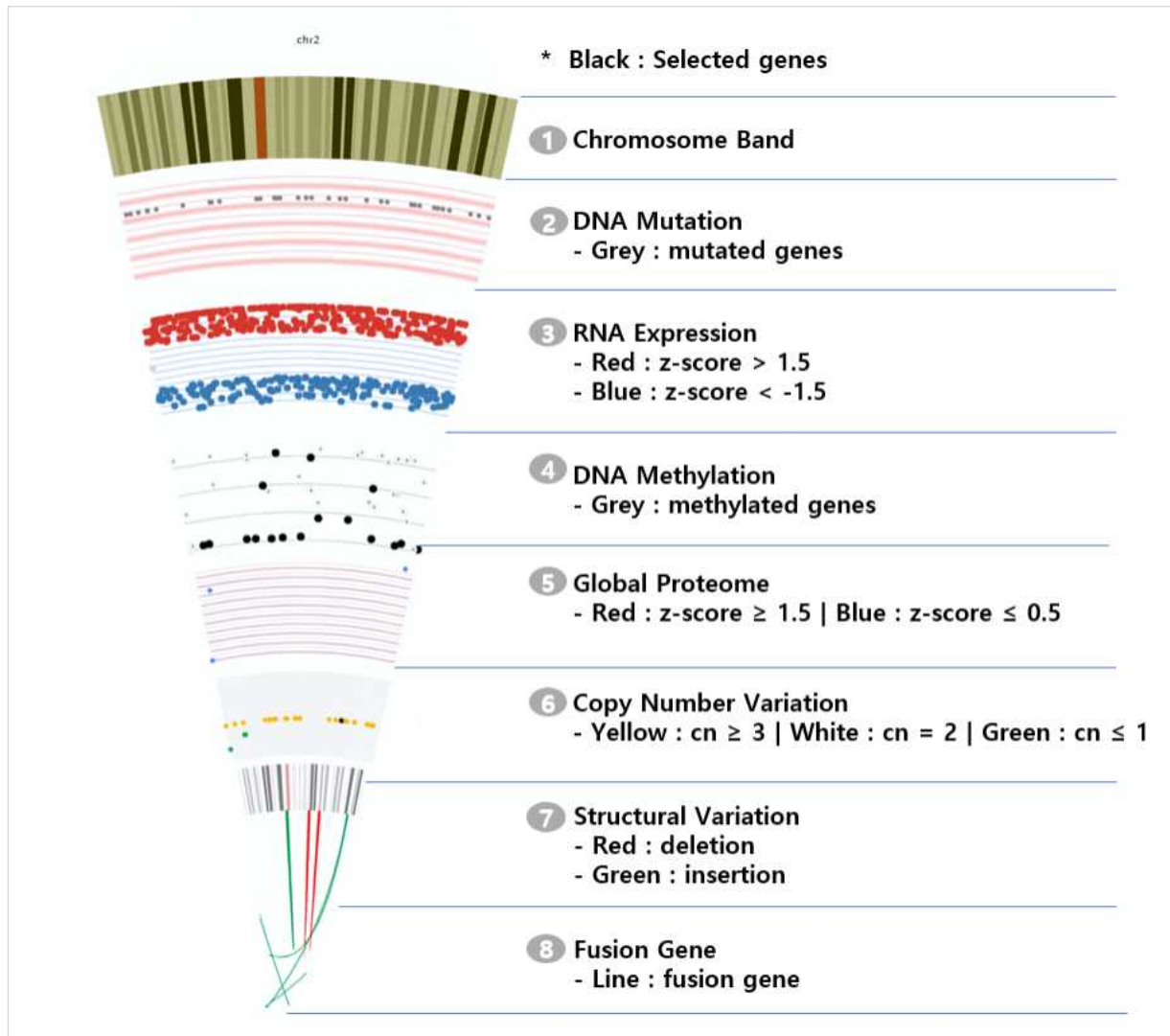
- . 시각화 가능한 오믹스 데이터 : DNA Mutation, CNV, Methylation, RNA, Fusion, Proteome
- . Circos plot은 원형의 염색체 지도 위에 7가지 유전체 데이터를 다수의 레이어로 나타내는 플롯이며, 각 유전체 데이터를 각 레이어 위에 서로 다른 색상의 점이나 선으로 표시합니다.
- . 단일 데이터 Circos plot 시각화 도구는 사용자가 업로드한 데이터에 대해서만 레이어가 표시됩니다. (실제 사용자 시각화 결과와 아래 예시 화면 결과가 다를 수 있습니다.)



< 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 점**으로 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 >

. Circos plot 레이어 상세 정보

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

< 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 Phosphorylation의 샘플 ID, Site, 유전자



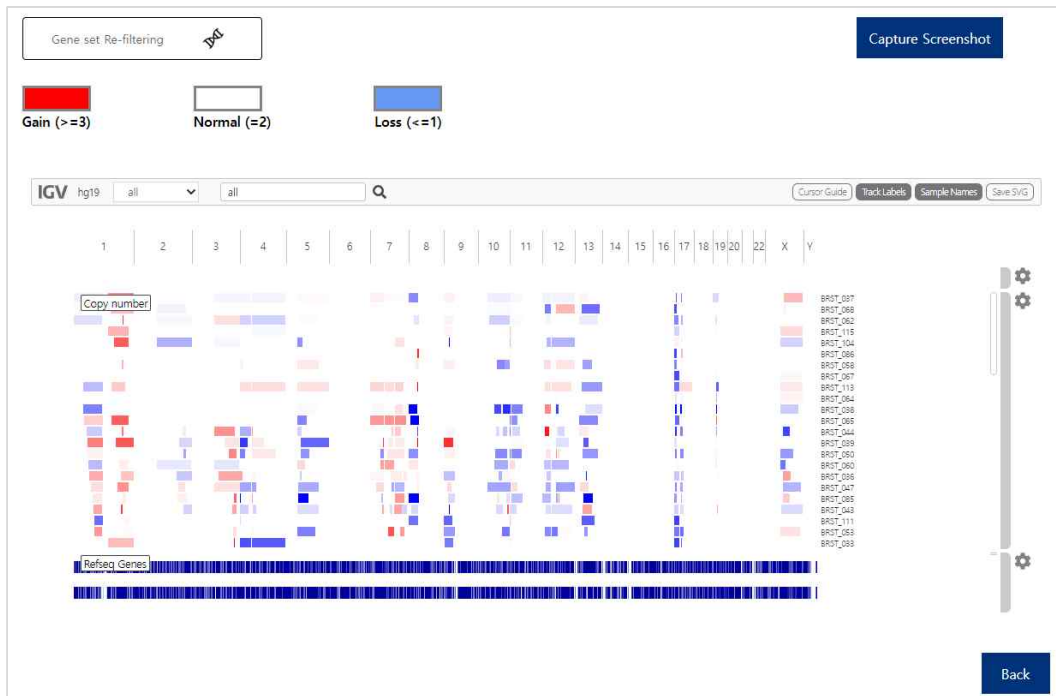
< Lollipop Plot - 돌연변이 화면>



< 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 ≥ 3)
 - 흰색 : Copy number normal (CN = 2)

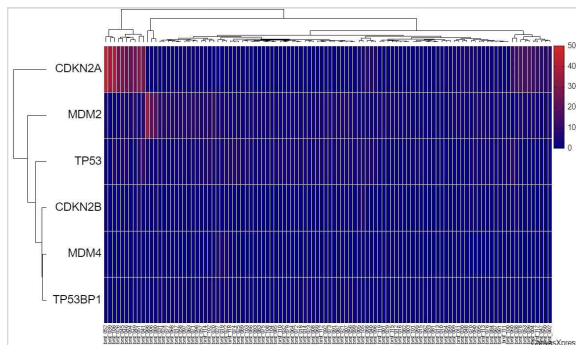
- 푸른색 : Copy number loss ($CN \leq 1$)
- . RefSeq Genes 트랙을 지원합니다. (확대 시, 상세 보기 가능)
- . 그 외 옵션 기능
 - Cursor Guide : 마우스 커서 위치 강조
 - Track labels : 트랙 레이블 보이게 설정
 - Sample names : 샘플 이름 보이게 설정
 - Save SVG : SVG 형식 다운로드

⑤ Heatmap

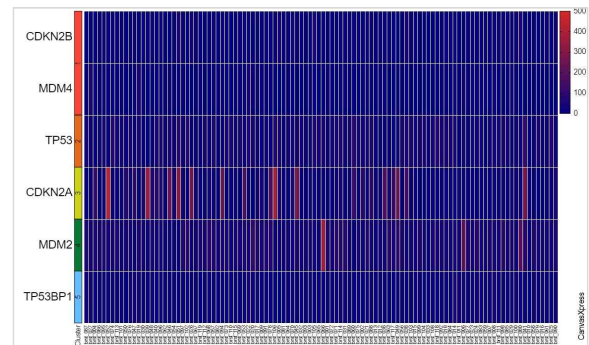
- . 시각화 가능한 오믹스 데이터 : Methylation, RNA, Proteome, Phosphorylation
- . Heatmap은 단백질 유전체 정보를 색상으로 표현하여 일정한 이미지 위에 열(Heat) 분포 형태로 나타내는 시각화 기법으로, 패턴이 유사한 샘플을 클러스터링(Clustering)하여 한눈에 파악할 수 있습니다.
- . [유전자 세트 필터링]을 통해 원하는 유전자에 대한 Heatmap을 그릴 수 있습니다.
- . 클러스터링 알고리즘을 선택할 수 있습니다.
 - Heatmap : Hierarchical clustering(계층 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)



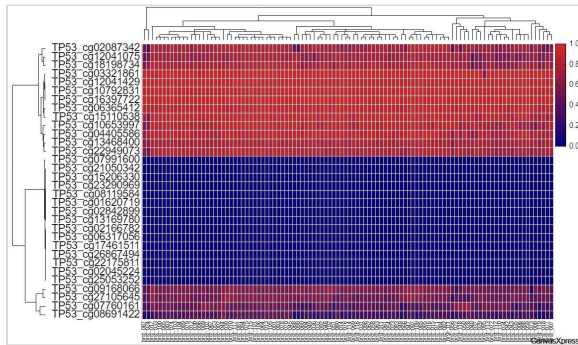
< RNA - Heatmap >



< RNA - K-mean >

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



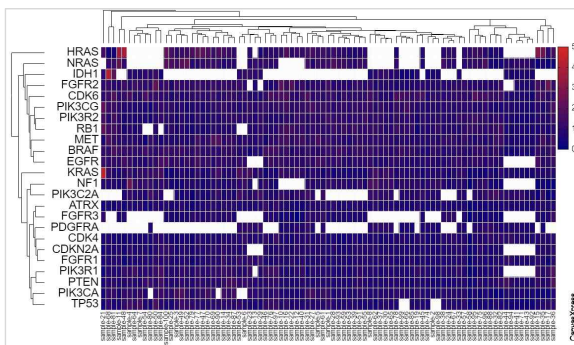
< Methylation - Heatmap >



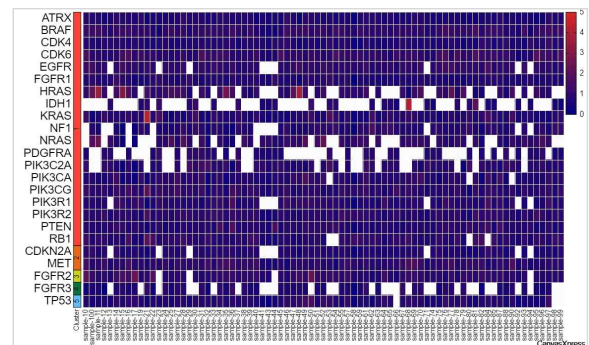
< Methylation - K-mean >

3) Proteome Heatmap : 단백질 풍부도를 히트맵 형태로 시각화

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



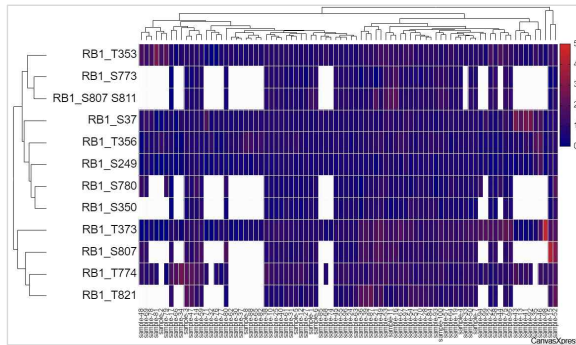
< Proteome - Heatmap >



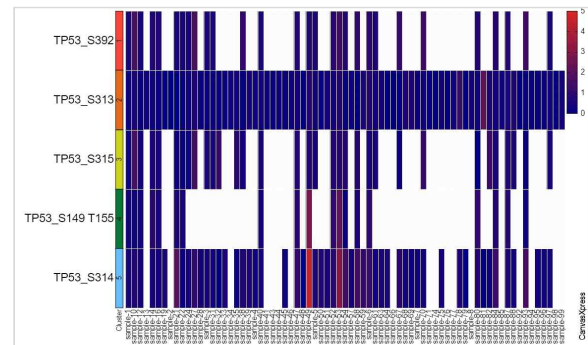
< Proteome - K-mean >

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)



< Phosphorylation - Heatmap >



< Phosphorylation - K-mean >

⑥ Box Plot

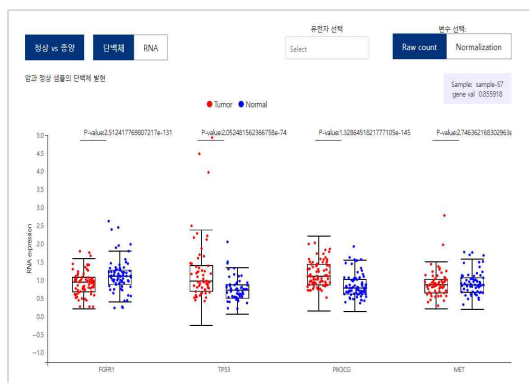
- . 시각화 가능한 오믹스 데이터 : RNA, Proteome
- . Box plot은 선택된 유전자의 단백질 풍부도 또는 RNA 발현 값 데이터의 기본 통계를 Box plot 형태로 보여줍니다. 최솟값, 최댓값, 1 사분위 수, 3 사분위 수, 이상치(outlier) 등을 나타냅니다.
- . 선택한 유전자에 대해 정상(Normal)과 종양(Tumor) 타입으로 나눠 시각화합니다.
- . 유전자 세트 필터링한 데이터 중 **[유전자 선택]**을 통해 원하는 유전자만 선택할 수 있습니다.
- . **[변수 선택]**을 통해 데이터를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

1) **단백체(Proteome)** : Tumor 및 Normal 샘플에 대하여, 선택된 유전자의 단백질 풍부도를 Box plot 형태로 시각화 합니다.

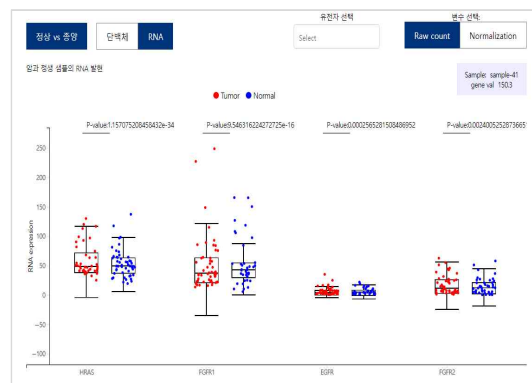
- [플롯 종류]는 단백질체(Proteome)을 선택
- X축 : 선택한 유전자
- Y축 : 선택한 유전자의 샘플 타입 별 단백질 발현 값

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 - 단백질체(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.

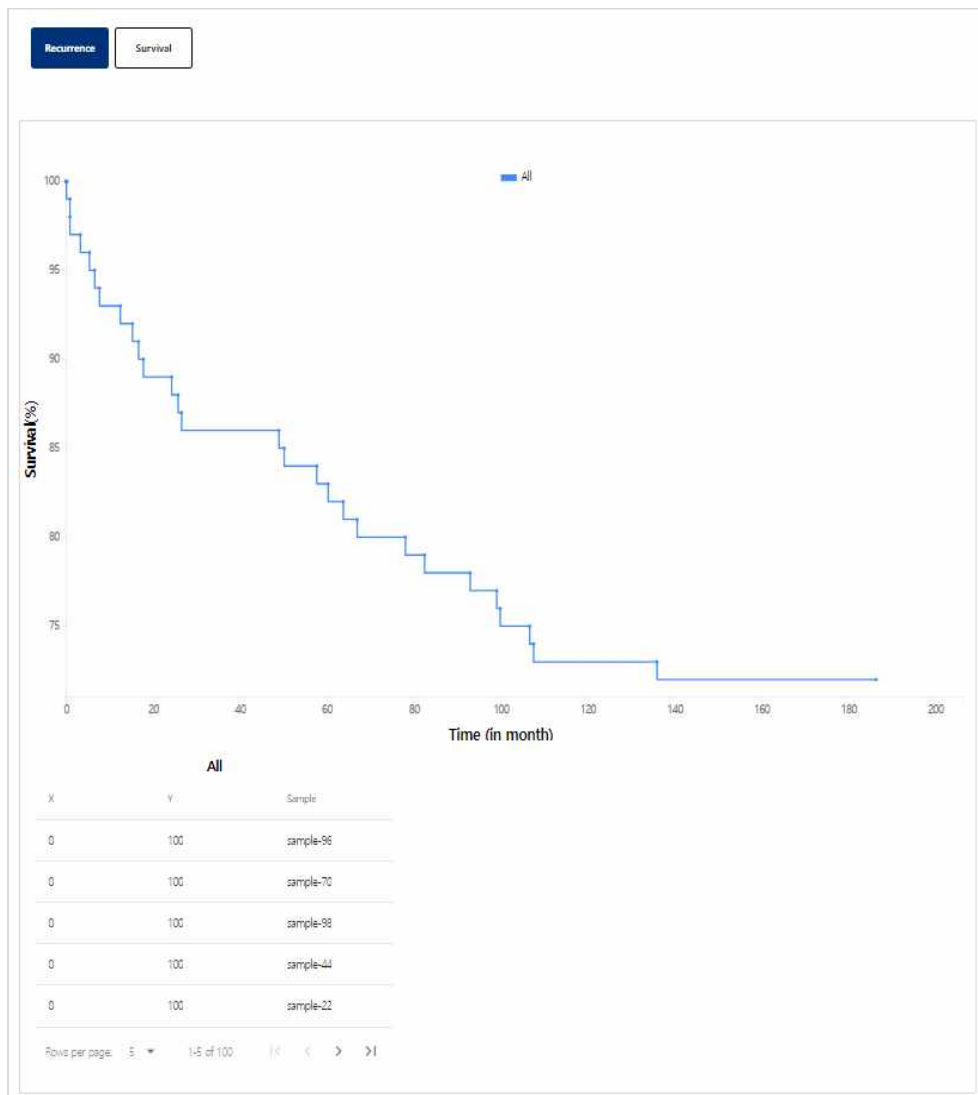
. 2가지 임상조건(Recurrence, Survival) 중 1가지를 선택하여 시각화 할 수 있습니다.

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_drtn]** 컬럼이 필수입니다.

- [플롯 종류]는 Recurrence를 선택

- X축 : 시간 (Month)

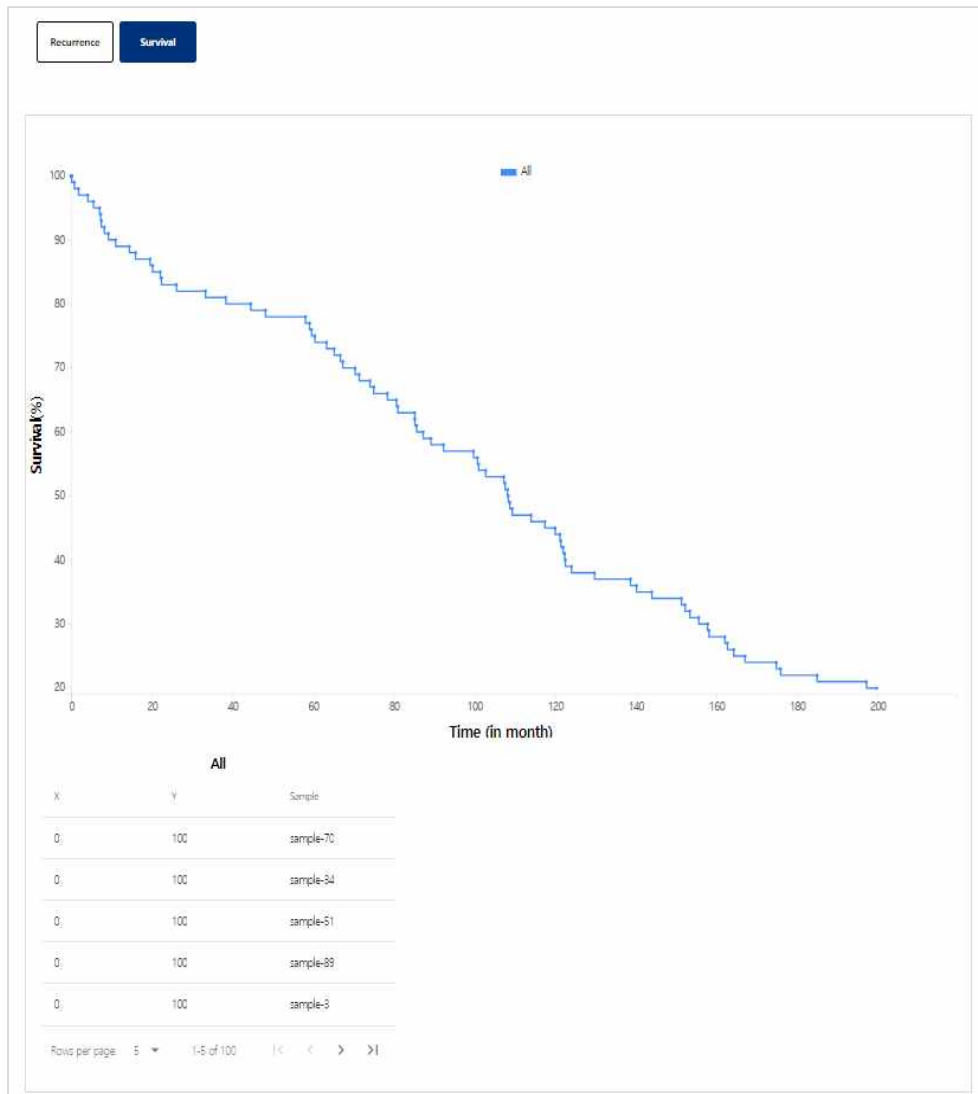
- Y축 : 전체 샘플에 대해 해당 기간동안 재발되지 않은 확률 (%)



< 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_drtn]** 컬럼이 필수입니다.

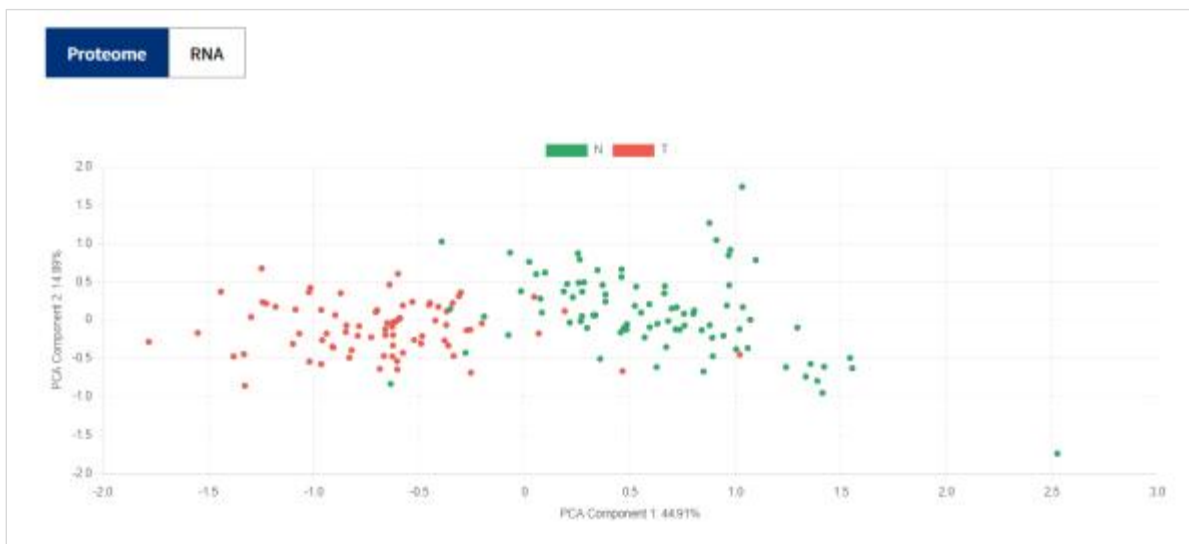
- [플롯 종류]는 Survival을 선택
- X축 : 생존 기간 (Month)
- Y축 : 전체 샘플에 대해 해당 기간동안 생존한 확률 (%)



< Survival plot - Survival >

⑧ PCA(Principal Component Analysis)

- . 시각화 가능한 오믹스 데이터 : RNA, Proteome
- . PCA는 2개 이상의 선택한 유전자 세트를 기준으로 고차원의 데이터를 저차원으로 환원시키는 기법입니다.
- . 단일 데이터분석의 PCA plot은 주성분을 2개만 사용하여 2차원 평면(X축, Y축)에 시각화합니다.
- . 2가지 오믹스 데이터(RNA, Proteome) 중 1가지를 선택하여 시각화 할 수 있습니다.
- . PCA plot은 업로드한 [유전자 세트 필터링] 박스를 통해 RNA 또는 Proteome 데이터의 모든 유전자를 선택하여 시각화를 할 수 있습니다.
- . 업로드한 데이터에 대해 Tumor, Normal 샘플을 그룹화하여 시각화합니다.



< PCA >