

CHALLENGE



Table 1. List of Cancer Patient Cards. There are four cards per cancer type, each card representing a different patient sample. In total, 51 genes are associated with the cancers listed on the 32 cards.

Cancer Type	Sample	Genes					
Lung cancer	1	<i>NF1</i>	<i>SETBP1</i>	<i>TP53</i>			
	2	<i>EGFR</i>	<i>MLL2</i>				
	3	<i>CTNNB1</i>	<i>KRAS</i>	<i>NF2</i>	<i>TP53</i>		
	4	<i>KT</i>	<i>MEN1</i>	<i>MLL3</i>	<i>TP53</i>		
Breast cancer	1	<i>BRCA1</i>	<i>TP53</i>				
	2	<i>CDH1</i>	<i>PIK3CA</i>				
	3	<i>ARID1B</i>	<i>TP53</i>				
	4	<i>FGFR2</i>	<i>GATA3</i>				
Colorectal cancer	1	<i>APC</i>	<i>TP53</i>				
	2	<i>KDM6A</i>	<i>KRAS</i>	<i>PIK3CA</i>	<i>SMAD4</i>		
	3	<i>APC</i>	<i>ATM</i>				
	4	<i>BRAF</i>	<i>CARD11</i>	<i>GNAS</i>	<i>PIK3CA</i>	<i>SMAD4</i>	<i>TP53</i>
Glioma	1	<i>CIC</i>	<i>IDH1</i>	<i>PIK3C5</i>			
	2	<i>CBL</i>	<i>TP53</i>				
	3	<i>ALK</i>	<i>ATM</i>	<i>BRCA1</i>	<i>TP53</i>		
	4	<i>HNFI1A</i>	<i>PTEN</i>				
Melanoma	1	<i>BRAF</i>	<i>CREBBP</i>	<i>EP300</i>			
	2	<i>FGFR2</i>	<i>MLL3</i>	<i>NRAS</i>	<i>PTCH1</i>		
	3	<i>BRAF</i>	<i>APC</i>	<i>BCOR</i>	<i>JAK3</i>	<i>MLL2</i>	
	4	<i>BRAF</i>	<i>CDKN2A</i>	<i>MLL3</i>	<i>ATM</i>	<i>PAX5</i>	
Hepatic cancer	1	<i>ARID1A</i>	<i>ARID2</i>	<i>BRCA1</i>	<i>RB1</i>	<i>TP53</i>	
	2	<i>CTNNB1</i>	<i>MED12</i>				
	3	<i>HNFI1A</i>	<i>TP53</i>				
	4	<i>ARID2</i>	<i>AXIN1</i>	<i>JAK2</i>	<i>TP53</i>		
Pancreatic cancer	1	<i>APC</i>	<i>GNAS</i>	<i>KRAS</i>	<i>RNF43</i>		
	2	<i>KRAS</i>	<i>TRAF7</i>	<i>TP53</i>			
	3	<i>CDKN2A</i>	<i>KRAS</i>	<i>MPL</i>	<i>TP53</i>		
	4	<i>KRAS</i>	<i>SMAD4</i>	<i>TP53</i>			
Leukemia	1	<i>MYD88</i>	<i>SETD2</i>				
	2	<i>BCOR</i>	<i>NOTCH1</i>				
	3	<i>BRAF</i>	<i>FAM123B</i>	<i>KRAS</i>			
	4	<i>NOTCH1</i>	<i>PIK3CA</i>	<i>TP53</i>			

Analisi 2:

- Utilizza il portale di dati genomici *BioMart* per ottenere una tabella di coordinate genomiche relative ai geni tumorali elencati nella tabella fornita
- Il portale *BioMart* portale è accessibile al seguente sito:
<http://www.ensembl.org/biomart/martview/>

Suggerimenti per svolgere l'analisi:

- Per coordinate genomiche di un gene si intende la sua localizzazione sul genoma: Nome del cromosoma su cui risiede il dato gene; posizione di inizio; posizione di fine.
- Segui la guida passo-passo nelle diapositive seguenti per vedere come ottenere queste informazioni per la lista di geni tumorali di nostro interesse

Guida passo-passo per ottenere le coordinate genomiche di una lista di geni con BioMart:

1. Apri una finestra di browser sul sito: <http://www.ensembl.org/biomart/martview>
2. Seleziona “Ensembl Genes 91” come database e “Human Genes” come dataset

The screenshot shows the Ensembl BioMart interface. At the top, the Ensembl logo is on the left, and navigation links (BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog) and a search bar are on the right. Below the header, there are tabs for 'New', 'Count', and 'Results'. A secondary row of buttons includes 'URL', 'XML', 'Perl', and 'Help'. The main configuration area on the left has sections for 'Dataset' (Human genes (GRCh38.p10)), 'Filters' ([None selected]), 'Attributes' (Gene stable ID, Transcript stable ID), and another 'Dataset' section ([None Selected]). The main panel on the right contains two dropdown menus: the first is set to 'Ensembl Genes 91' and the second is set to 'Human genes (GRCh38.p10)'.

Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog Search all species... Login/Register

New **Count** **Results** **URL** **XML** **Perl** **Help**

Dataset
Human genes (GRCh38.p10)

Filters
[None selected]

Attributes
Gene stable ID
Transcript stable ID

Dataset
[None Selected]

Ensembl Genes 91

Human genes (GRCh38.p10)

3. Seleziona la voce "Attributes" dal menu' di sinistra
4. Espandi la sezione "GENE" e seleziona "Input external references ID list" come metodo di filtro
5. Dal menu' a tendina corrispondente seleziona "HGNC symbols"
6. copia / incolla la lista di geni tumorali di interesse nell'apposito spazio (uno per riga), oppure Copiali in un file di testo (sempre uno per riga) e carica il file
7. Premi sul tasto counts dal menu' in alto per vedere quanti geni hai selezionato. Il numero Corrisponde a quello che ti aspettavi?

New **Count** **Results**

URL **XML** **Perl** **Help**

Dataset
 Human genes (GRCh38.p10)

Please restrict your query using criteria below

(If filter values are truncated in any lists, hover over the list item to see the full text)

☒ REGION:

☒ GENE:

☐ Limit to genes (external references)...

With CCDS ID(s) ▼
☒ Only
☐ Excluded

☒ Input external references ID list [Max 500 advised]

HGNC symbol(s) [e.g. A1BG] ▼

NF1
GATA3
KRAS
CDKN2A
SMAD4

Sfoglia... Nessun file selezionato.


☐ Limit to genes (microarray probes/probesets)...

With AFFY HC G110 probe ID(s) ▼
☒ Only
☐ Excluded

☐ Input microarray probes/probesets ID list [Max 500 advised]

AFFY HC G110 probe ID(s) [e.g. 266_s_at] ▼

8. Seleziona la voce “Filters” dal menu’ di sinistra
9. Espandi la sezione “GENE” e deselecta eventuali voci selezionate. Poi seleziona:
“Chromosome/scaffold name”
“Gene start (bp)”
“Gene end (bp)”
10. Espandi la sezione “EXTERNAL” e seleziona:
“HGNC symbol”

 BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Login/Register

Search all species...

NewCountResultsURLXMLPerlHelp

Dataset 5 / 63967 Genes
Human genes (GRCh38.p10)

Filters
HGNC symbol(s) [e.g. A1BG]: [ID-list specified]

Attributes
Chromosome/scaffold name
Gene start (bp)
Gene end (bp)
HGNC symbol

Dataset
[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

Missing non coding genes in your mart query output, please check the following [FAQ](#)

☒ Features

☐ Variant (Germline)

☐ Structures

☐ Variant (Somatic)

☐ Homologues

☐ Sequences

GENE:

EXTERNAL:

GO

☐ GO term accession

☐ GO term evidence code

☐ GO term name

☐ GO domain

☐ GO term definition

GOSlim GOA

☐ GOSlim GOA Accession(s)

☐ GOSlim GOA Description

External References (max 3)

☐ CCDS ID

☐ MIM morbid accession


☐ ChEMBL ID

☐ miRBase accession

☐ Clone-based (Ensembl) gene ID

☐ miRBase ID

11. Seleziona la voce “Results” dal menu’ in alto a sinistra
12. Come opzioni di output seleziona “Export all results to” file “TSV”;
13. Spunta l’opzione “Unique results only” e poi premi il tasto “Go”
14. Apri un foglio di calcolo Excel e importaci il file scaricato da BioMart


Login/Register

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog
Search all species...

New Count Results
URL XML Perl Help

Dataset 5 / 63967 Genes

Human genes (GRCh38.p10)

Filters

HGNC symbol(s) [e.g. A1BG]: [ID-list specified]

Attributes

Chromosome/scaffold name

Gene start (bp)

Gene end (bp)

HGNC symbol

Dataset

[None Selected]

Export all results to

Email notification to

File

TSV

☐ Unique results only

Go

View

All rows as TSV ☐ Unique results only

Chromosome/scaffold name	Gene start (bp)	Gene end (bp)	HGNC symbol
9	21967753	21995301	CDKN2A
10	8053604	8075198	GATA3
12	25204789	25250936	KRAS
17	31094927	31382116	NF1
18	51028394	51085045	SMAD4

Analisi 3:

- Crea un file BED di coordinate genomiche per i geni tumorali di nostro interesse e importala nel UCSC Genome Browser <http://genome.ucsc.edu/> facendo in modo che venga visualizzata come traccia di colore rosso, con nome e descrizione opportuni (vedi immagine qui sotto)

Suggerimenti per svolgere l'analisi:

- Se non ricordi come caricare un file BED nel UCSC Genome Browser, o come modificarne il colore / nome / descrizione, ripercorri i passaggi che lo illustrano dal tutorial #2 fatto in classe la volta scorsa
- Per quanto riguarda la creazione del file BED, utilizza il file excel appena creato nell'Analisi 2 eliminando il nome delle colonne e riarrangiandone il contenuto come mostrato qui sotto. Salva questo file come file di testo semplice separato da tabulazione.