

# Leveraging public pan-cancer transcriptomic data for immuno-oncology: systematic collection and a graphical tool for analysis

**Marco Barreca, PhD student**

**IMMUNO-model Third Annual Conference**

University of Warsaw Biological and Chemical  
Research Centre, Warsaw

13/05/2025

# Overview of the project

## Background



Development of omics technologies has generated a growing amount of publicly available molecular profiles which could help to define the most appropriate models to study *immunotherapy efficacy* and *toxicity* in tumours and *validate biomarkers* to monitor these effects.

## Aims



The project aims to:

- Collect a large *amount of publicly available omics datasets* from immuno-oncology related experiments, and
- Make the access to and the analysis of such resource user-friendly.

## Methods

We developed:



- a *Systematic Search Strategy* for dataset collection, processing, curation and biobanking and
- a *graphical interface* for the visualization and the analysis using Shiny app.



## Impact

This project will support:

- the *development* and *validation* of *biomarkers*,
  - the *identification* of *new targets*,
- fostering translational research in immuno-oncology.

## An International Project

- The project involved a total of 37 researchers from 15 different countries establishing an international taskforce.

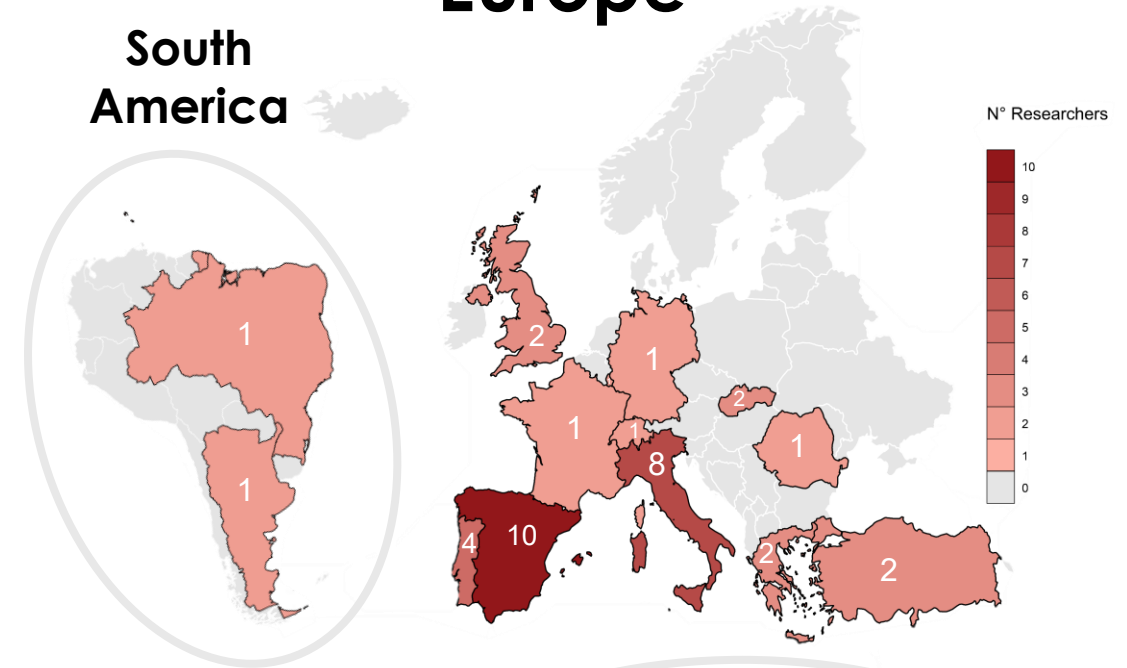
Scan the QR code to discover the Data Gathering Taskforce!



### South America



### Europe

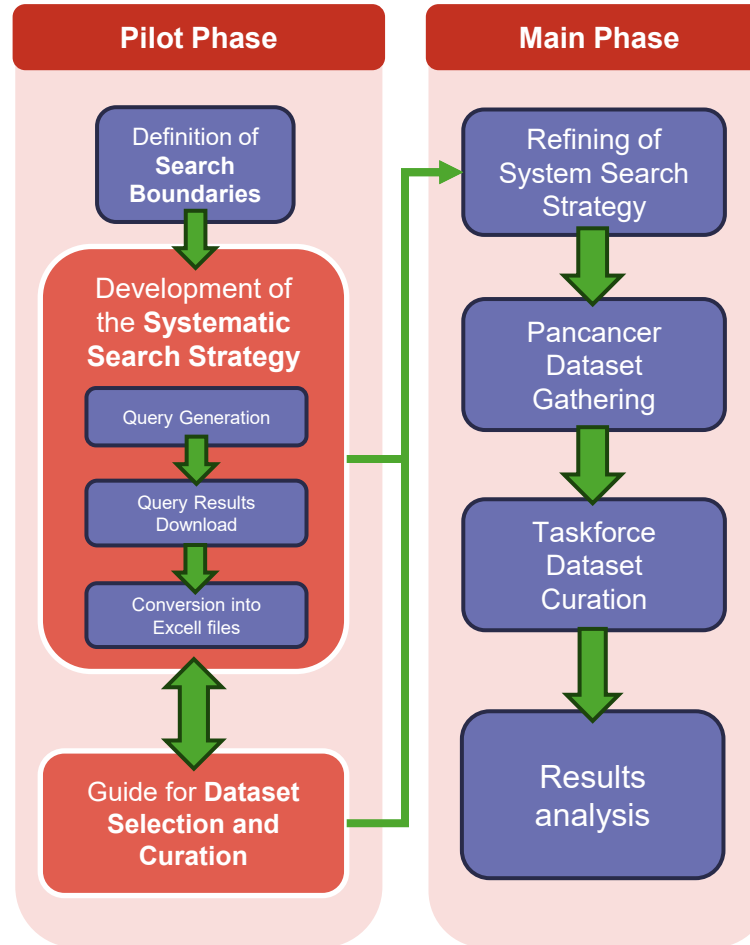


### South East Asia



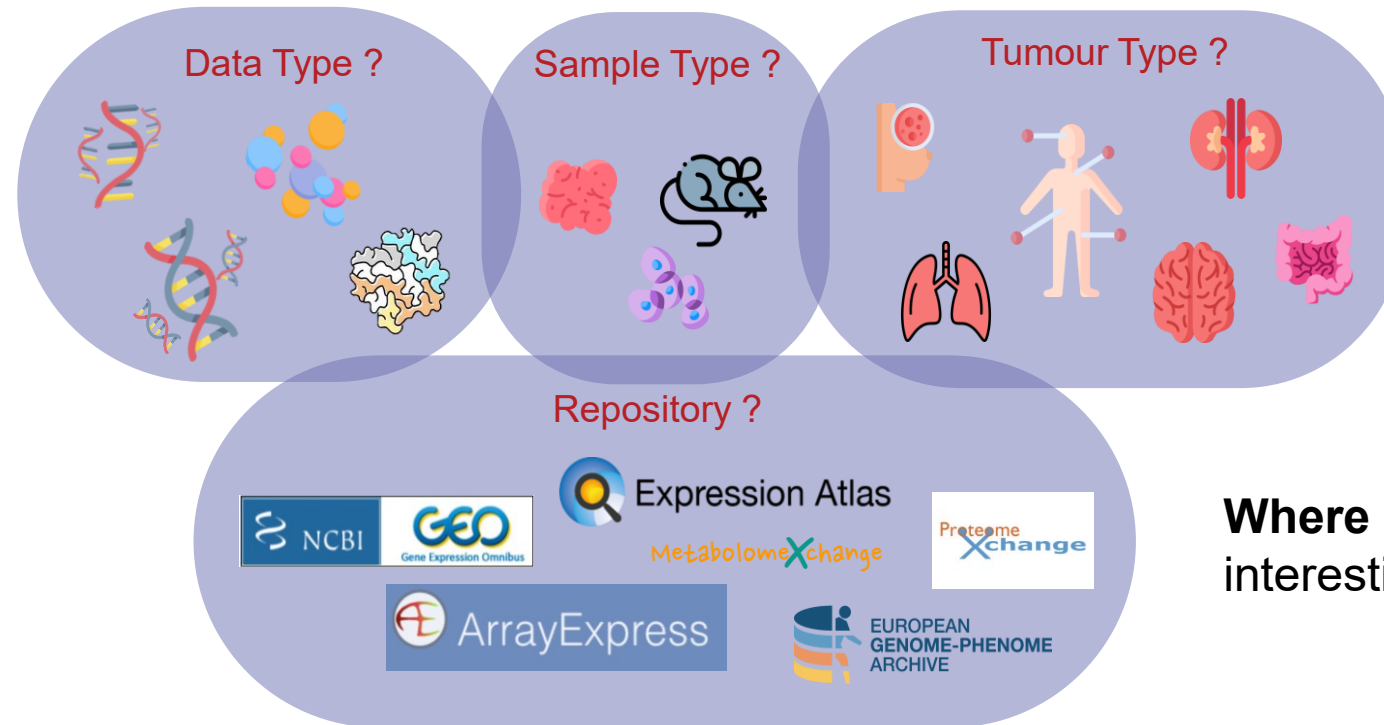
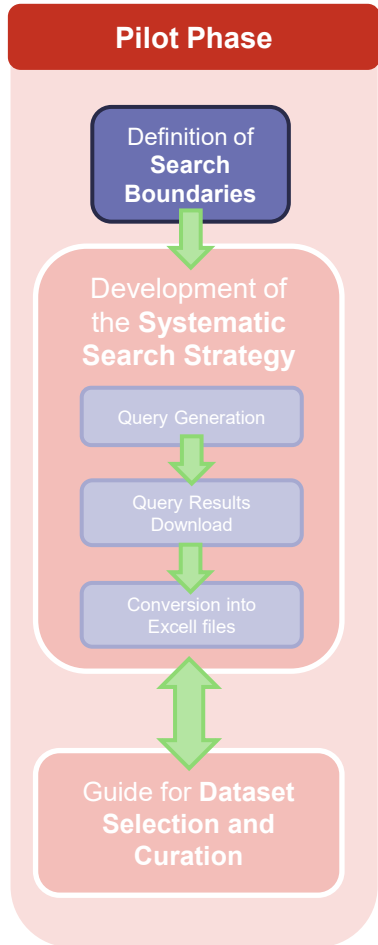
# Workflow

- **Feasibility Assessment**
- **Method Refinement**
- **Problem Identification**
- **Participant Recruitment**



- **Application of the Strategy**
- **Ongoing phase...**

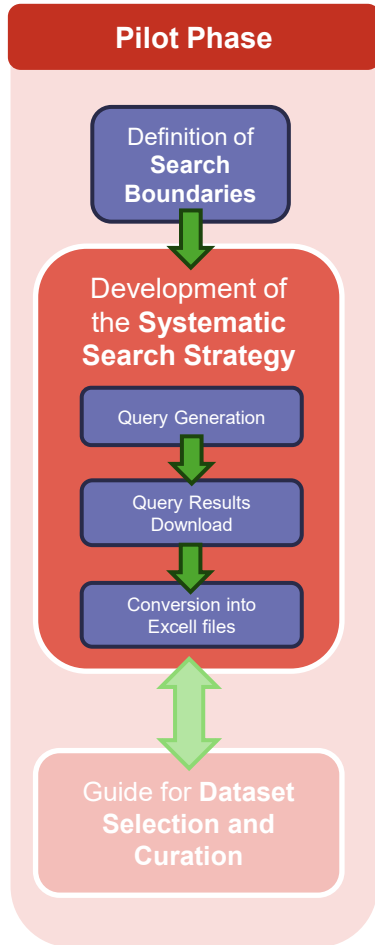
## Workflow – Definition of Search Boundaries



**What** kind of data are we interested in?

**Where** can we find the interesting data?

# Workflow – Systematic Search Strategy



## Query Structure

[Technology] AND [Organism] AND [Tumour type] AND [Treatment]

Automatic conversion

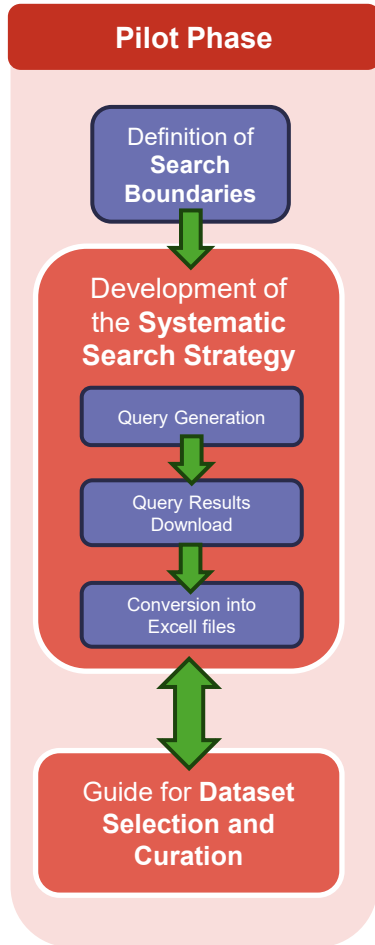
.txt to .xlsx file by R coding



Repository	Accession Number	Title	Description	Query ID	Dataset_Lin	Platform	Ty	Organism	Samples	Selected	Tumour_Ty	Setting	Biological_A	Treatment	Treatment_note
GEO	GSE225066	Transcrip RNASeq pro 2.1			https://www	Expression p	Homo sapi		26	<input type="checkbox"/>					
GEO	GSE262761	PIN1 iso Bladder can 2.1			https://www	Expression p	Homo sapi		12	<input type="checkbox"/>					
GEO	GSE268243	Role of F Antagonism 2.1			https://www	Expression p	Homo sapi		12	<input type="checkbox"/>					
GEO	GSE276612	Inhibitio Currently, oi 2.1			https://www	Expression p	Homo sapi		2	<input type="checkbox"/>					
GEO	GSE247185	Impact c Purpose: Ne 2.1			https://www	Expression p	Homo sapi		13	<input type="checkbox"/>					
GEO	GSE269520	Combin Glutamine ii 2.1			https://www	Expression p	Homo sapi		6	<input type="checkbox"/>					
GEO	GSE176493	Compres This SuperSi 2.1			https://www	Expression p	Homo sapi		252	<input type="checkbox"/>					
GEO	GSE176485	Xenogra Advanced bi 2.1			https://www	Expression p	Homo sapi		21	<input type="checkbox"/>					
GEO	GSE248167	Tumor ir Molecular si 2.1			https://www	Expression p	Homo sapi		51	<input type="checkbox"/>					
GEO	GSE217413	Cisplatin Bladder can 2.1			https://www	Expression p	Homo sapi		4	<input type="checkbox"/>					
GEO	GSE224248	Develop Although re 2.1			https://www	Expression p	Homo sapi		49	<input type="checkbox"/>					
GEO	GSE245122	Modulat Alterations i 2.1			https://www	Expression p	Homo sapi		16	<input type="checkbox"/>					
GEO	GSE245748	Prognost Chemoradia 2.1			https://www	Expression p	Homo sapi		73	<input type="checkbox"/>					
GEO	GSE235066	Immuno The Phase 3 2.1			https://www	Expression p	Homo sapi		32	<input type="checkbox"/>					
GEO	GSE244957	RNA-seq Analyses of 2.1			https://www	Expression p	Homo sapi		87	<input type="checkbox"/>					
GEO	GSE224099	A Popule We previous 2.1			https://www	Expression p	Homo sapi		12	<input type="checkbox"/>					
GEO	GSE222934	A Phase This SuperSi 2.1			https://www	Expression p	Homo sapi		76	<input type="checkbox"/>					
GEO	GSE222932	A Phase We designe 2.1			https://www	Expression p	Homo sapi		51	<input type="checkbox"/>					
GEO	GSE217785	Upreguli Methylenet 2.1			https://www	Expression p	Homo sapi		6	<input type="checkbox"/>					



# Workflow – Guide for Dataset Selection and Curation



## Guide for selection and curation process

This document provides guidelines for selecting and curating transcriptomic datasets from the **GEO** and **ArrayExpress** repositories, specifically focusing on **immuno-oncology experiments**. The goal is to ensure consistency and reliability, facilitating reproducible research.

### Selection Criteria

The criteria outlined below define which datasets should be considered as the best practices for curating query results.

### Inclusion criteria

We are looking for datasets that include the **transcriptomic preclinical experiments** involving various tumour types:

- **Platform:** Any RNA-seq and scRNA-seq clinical/preclinical
- **Tumour type:** Only datasets that include the following
  - o TCGA cancer types
- **Tissue type:** In clinical datasets, **primary tumour, population** samples are eligible.
- **Preclinical Model:** In preclinical datasets, **cell line** and **PDX** are eligible.
- **Treatment:** Only datasets that include any **immunotherapy** should be selected. The combination of immunotherapies (including radiotherapy) are eligible.
- **Sample size:** The minimum number of samples required is:
  - o **>5** for bulk RNA-seq preclinical experiments;
  - o **>10** for bulk RNA-seq clinical experiments;
  - o **No minimum** number of samples for scRNA-seq
- **Other:** The effective cancer type does not need to be specified.
  - o For example, if a query for "breast cancer" returns meta-analyses included if they meet the inclusion criteria.

### Exclusion criteria

- **Platform:** **Microarray**-based experiments won't be included.
- **Preclinical Model:** In preclinical datasets, **knock-out** won't be included.

### Query results curation

Before curating a query result, a new sheet of the starting Excel should be created, and the template on the first page should be copied and pasted into the new sheet.

Please follow the next points in the order they are shown:

- Open the file "Query\_result\_#.xlsx" where the first number is the query ID connected to the file "Queries.xlsx" and the second is the page of query results (each one has a maximum of 50 datasets);
- Create a new page **INSIDE** the Excel and rename that with your name;
- Copy the template you find on the page "Datasets" and paste it to the page with your name;
- Start to curate!

Checkboxes and drop-down menus are provided to curate the query results, reducing the compilation time and standardising the information format.

Please notice that:

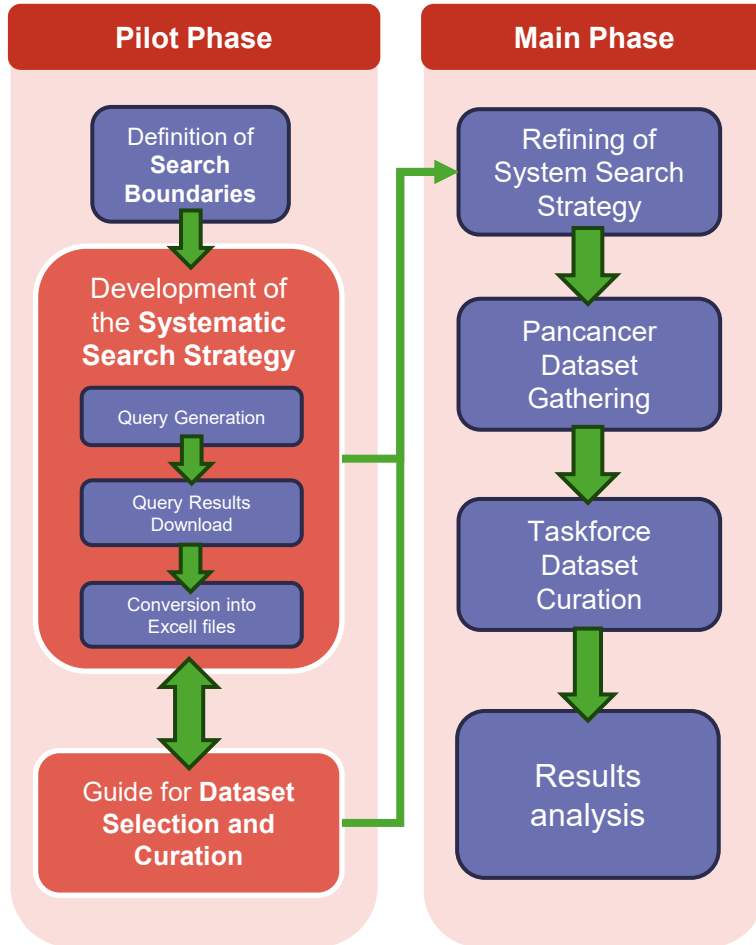
- Additional info requested by the form should be filled **only for selected datasets** (tick the checkbox if the dataset is selected and continue the curation process);
- Multiple choice option is active; if needed, select more than one term;
- At least **2 researchers per query** should be involved in the query results curation process;
- Find how the columns are expected to be filled out in the following table:

Column names	Information
Tumour type	The drop-down menu contains the tumour types, referring to the organ where the tumour occurs. For preclinical experiments, the original tissue of the model should be considered. All the subtypes are reconnected to the general classification. Use "other" only if the organ where the tumour occurs is not available.
Setting	The drop-down menu contains the major categories: "clinical" and "preclinical". The "preclinical" is used if the datasets contain bulk or single-cell samples collected from patients and directly analysed (primary tumours, metastasis, sorted population). The "preclinical" setting is used for datasets containing PDX, cell lines, organoids, mouse models and single-cell samples collected from pre-clinical studies (organoids or PDX, for example). The "Other" option should be used only for exceptions and studies that could not be classified as pre-clinical or clinical.
Biological Model	The drop-down menu contains some of the most common biological model categories. The "Sorted populations" are used for single-cell experiments or immune populations extracted from the tumour or the biological system. The "Other" option should be used only if the biological model is not cited from the menu.
Treatment	Only immunotherapy treatment and their combination will be considered. The term immunotherapy considers: immune checkpoint inhibitors, monoclonal antibodies (MABs), vaccines, cytokines, CAR (chimeric antigen receptor) therapies (CAR-T, CAR-NK, CAR-CIK), TIL (tumour-infiltrating lymphocytes) therapy, tumor-infecting viruses (or oncolytic viruses), and adoptive cell transfer.
Treatment Details	Please specify the type of treatment and/or the prescribed drugs and the clinical setting (neoadjuvant or adjuvant) if suitable.
Note	This is a space for you and your thoughts, use it freely!

## Guidelines

- Provide direction and clarity on how to achieve the goal;
- Promote best practices, ensure consistency, and enhance quality;
- Help new collaborators to keep up.

# Workflow



# Workflow – Refining of System Search Strategy

## Main Phase

Refining of System Search Strategy

Pancancer Dataset Gathering

Taskforce Dataset Curation

Results analysis

## Search Boundaries

Repository	Organism	Tumour type	Data type	Treatment
GEO	Homo sapiens	TCGA types	RNA-sequencing	immunotherapy
ArrayExpress	Mus musculus		scRNA	checkpoint inhibitors
	Rattus norvegicus			ICI
				ICB
				CTLA-4
				PD-1
				PD-L1
				Atezolizumab
				Pembrolizumab
				Nivolumab
				Ipilimumab
				Durvalumab
				TIM-3
				LAG-3
				VISTA
				OX40
				ACT (adoptive cell therapy)
				CD40
				IL-2
				GITR

TCGA tumour types		
Study Abbreviation	Study Name	Organ
ACC	Adrenocortical carcinoma	Adrenal gland
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	B cells
BLCA	Bladder Urothelial Carcinoma	Bladder
L0ML	Chronic Myelogenous Leukemia	Bone marrow cells
PCPG	Pheochromocytoma and Paraganglioma	Brain
LGG	Brain Lower Grade Glioma	Brain
GBM	Glioblastoma multiforme	Brain
BRCA	Breast invasive carcinoma	Breast
COAD	Colon adenocarcinoma	Colon
ESCA	Esophageal carcinoma	Esophagus
UVM	Uveal Melanoma	Eye
HNSC	Head and Neck squamous cell carcinoma	Head and neck
KICH	Kidney Chromophobe	Kidney
KIRC	Kidney renal clear cell carcinoma	Kidney
KIRP	Kidney renal papillary cell carcinoma	Kidney
CHOL	Cholangiocarcinoma	Liver
LIHC	Liver hepatocellular carcinoma	Liver
LUAD	Lung adenocarcinoma	Lung
LUSC	Lung squamous cell carcinoma	Lung
MESO	Mesothelioma	Mesothelium
LAML	Acute Myeloid Leukemia	Myeloid cells
FPFP	FFPE Pilot Phase II	None
MISC	Miscellaneous	Other
CNTL	Controls	None
OV	Ovarian serous cystadenocarcinoma	Ovary
PAAD	Pancreatic adenocarcinoma	Pancreas
PRAD	Prostate adenocarcinoma	Prostate
READ	Rectum adenocarcinoma	Rectum
SARC	Sarcoma	Other
SKCM	Skin Cutaneous Melanoma	Skin
STAD	Stomach adenocarcinoma	Stomach
TGCT	Testicular Germ Cell Tumors	Testicle
THYM	Thymoma	Thymus
THCA	Thyroid carcinoma	Thyroid
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma	Uterus
UCS	Uterine Carcinosarcoma	Uterus
UCEC	Uterine Corpus Endometrial Carcinoma	Uterus

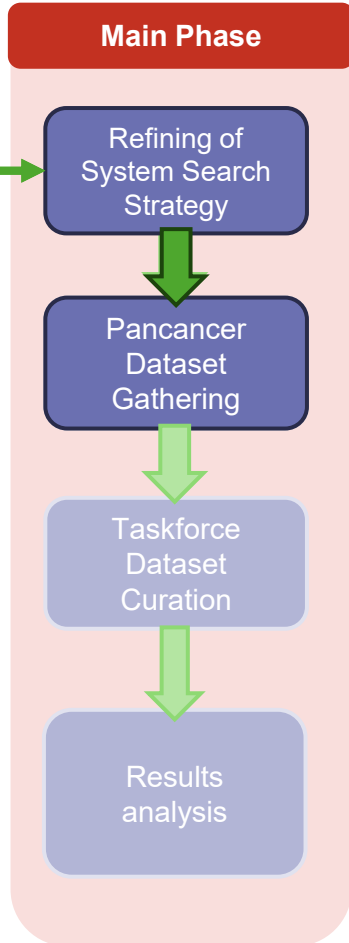
We looked for :

1. **RNA-sequencing datasets** providing **human samples**
2. More than **20 cancer types** and **20 treatment terms** were used to create the queries.

## Query Generated

Query_ID	Repo	Query
1	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND ("adrenal gland neoplasm"[MeSH] OR "adrenocortical carcinoma"[MeSH] OR "lymphoid neoplasm, diffuse large b-cell type"[MeSH] OR "diffuse large b-cell lymphoma"[MeSH] OR "bladder neoplasm"[MeSH] OR "urothelial carcinoma of bladder"[MeSH] OR "chronic myelogenous leukemia"[MeSH] OR "bone marrow cells"[MeSH] OR "pheochromocytoma and paraganglioma"[MeSH] OR "brain neoplasm, lower grade glioma"[MeSH] OR "glioblastoma multiforme"[MeSH] OR "breast neoplasm, invasive carcinoma"[MeSH] OR "colon adenocarcinoma"[MeSH] OR "esophageal carcinoma"[MeSH] OR "uveal melanoma"[MeSH] OR "head and neck squamous cell carcinoma"[MeSH] OR "kidney chromophobe tumor"[MeSH] OR "kidney renal clear cell carcinoma"[MeSH] OR "kidney renal papillary cell carcinoma"[MeSH] OR "cholangiocarcinoma"[MeSH] OR "liver hepatocellular carcinoma"[MeSH] OR "lung adenocarcinoma"[MeSH] OR "lung squamous cell carcinoma"[MeSH] OR "mesothelioma"[MeSH] OR "acute myeloid leukemia"[MeSH] OR "miscellaneous"[MeSH] OR "controls"[MeSH] OR "ovarian serous cystadenocarcinoma"[MeSH] OR "pancreatic adenocarcinoma"[MeSH] OR "prostate adenocarcinoma"[MeSH] OR "rectum adenocarcinoma"[MeSH] OR "sarcoma"[MeSH] OR "skin cutaneous melanoma"[MeSH] OR "stomach adenocarcinoma"[MeSH] OR "testicular germ cell tumor"[MeSH] OR "thymoma"[MeSH] OR "thyroid carcinoma"[MeSH] OR "cervical squamous cell carcinoma and endocervical adenocarcinoma"[MeSH] OR "uterine carcinosarcoma"[MeSH] OR "uterine corpus endometrial carcinoma"[MeSH])
2	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND ("carcinoma, transitional cell"[MeSH] OR "transitional cell carcinoma"[MeSH] OR "brain neoplasm, glioma, transitional cell type"[MeSH] OR "glioblastoma multiforme"[MeSH] OR "breast neoplasm, invasive carcinoma"[MeSH] OR "colon adenocarcinoma"[MeSH] OR "esophageal carcinoma"[MeSH] OR "uveal melanoma"[MeSH] OR "head and neck squamous cell carcinoma"[MeSH] OR "kidney chromophobe tumor"[MeSH] OR "kidney renal clear cell carcinoma"[MeSH] OR "kidney renal papillary cell carcinoma"[MeSH] OR "cholangiocarcinoma"[MeSH] OR "liver hepatocellular carcinoma"[MeSH] OR "lung adenocarcinoma"[MeSH] OR "lung squamous cell carcinoma"[MeSH] OR "mesothelioma"[MeSH] OR "acute myeloid leukemia"[MeSH] OR "miscellaneous"[MeSH] OR "controls"[MeSH] OR "ovarian serous cystadenocarcinoma"[MeSH] OR "pancreatic adenocarcinoma"[MeSH] OR "prostate adenocarcinoma"[MeSH] OR "rectum adenocarcinoma"[MeSH] OR "sarcoma"[MeSH] OR "skin cutaneous melanoma"[MeSH] OR "stomach adenocarcinoma"[MeSH] OR "testicular germ cell tumor"[MeSH] OR "thymoma"[MeSH] OR "thyroid carcinoma"[MeSH] OR "cervical squamous cell carcinoma and endocervical adenocarcinoma"[MeSH] OR "uterine carcinosarcoma"[MeSH] OR "uterine corpus endometrial carcinoma"[MeSH])
3	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "brain neoplasms"[MeSH]
4	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "breast neoplasms"[MeSH]
5	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "colorectal neoplasms"[MeSH]
6	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "head and neck neoplasms"[MeSH]
7	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "liver neoplasms"[MeSH]
8	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "lung neoplasms"[MeSH]
9	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "skin neoplasms"[MeSH]
10	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "sarcoma"[MeSH]
11	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "ovarian neoplasms"[MeSH]
12	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "pancreatic neoplasms"[MeSH]
13	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "prostatic neoplasms"[MeSH]
14	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "kidney neoplasms"[MeSH]
15	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "stomach neoplasms"[MeSH]
16	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "uterine neoplasms"[MeSH]
17	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "leukemia, myeloid"[MeSH]
18	GEO	expression profiling by high throughput sequencing[DataSet Type] AND "human"[Organism] AND "lymphoma, b-cell"[MeSH]

## Workflow - Pancancer Dataset Gathering



Query_ID	Rep	Query	Number of results	Date	Organism	Tumour type	Tissue type
1	GEO	expression profiling	36	28/3/2025	human	Adrenal Cancer Est ...	Solid Tumour
2	GEO	expression profiling	34	28/3/2025	human	Bladder Cancer	Solid Tumour
3	GEO	expression profiling	40	28/3/2025	human	Brain Cancer	Solid Tumour
4	GEO	expression profiling	115	28/3/2025	human	Breast Cancer	Solid Tumour
5	GEO	expression profiling	93	28/3/2025	human	Colorectal Cancer	Solid Tumour
6	GEO	expression profiling	35	28/3/2025	human	Head and Neck Canc ...	Solid Tumour
7	GEO	expression profiling	92	28/3/2025	human	Liver Cancer	Solid Tumour
8	GEO	expression profiling	152	28/3/2025	human	Lung Cancer	Solid Tumour
9	GEO	expression profiling	214	28/3/2025	human	Melanoma	Solid Tumour
10	GEO	expression profiling	49	28/3/2025	human	Other	Solid Tumour
11	GEO	expression profiling	60	28/3/2025	human	Ovarian Cancer	Solid Tumour
12	GEO	expression profiling	38	28/3/2025	human	Pancreatic Cancer	Solid Tumour
13	GEO	expression profiling	44	28/3/2025	human	Prostatic Cancer	Solid Tumour
14	GEO	expression profiling	54	28/3/2025	human	Renal Cancer	Solid Tumour
15	GEO	expression profiling	32	28/3/2025	human	Stomach Cancer	Solid Tumour
16	GEO	expression profiling	22	28/3/2025	human	Uterine Cancer	Solid Tumour
17	GEO	expression profiling	55	28/3/2025	human	Myeloid Leukemia	Liquid Tum...
18	GEO	expression profiling	51	28/3/2025	human	B cell Lymphoma	Liquid Tum...

A total of **18 different queries** were generated  
 A total of **1216 datasets** were found.

# Workflow – Taskforce Dataset Curation

## Main Phase

Refining of System Search Strategy

Pancancer Dataset Gathering

Taskforce Dataset Curation

Results analysis

Organism	Samples	Selected	Tumour_Type	Setting	Biological_Model	Treatment	Treatment_   note
Homo sapie	20	<input type="checkbox"/>					glucocortic
Mus muscul	117	<input checked="" type="checkbox"/>	Breast Cancer	Clinical	Primary tumor ...	Chemo + Imm ...	paclitaxel, n scRNA
Homo sapie	12	<input type="checkbox"/>					Non-treat
Homo sapie	1254	<input type="checkbox"/>					No immunc
Homo sapie	81	<input type="checkbox"/>					No immunc
Homo sapie	14	<input checked="" type="checkbox"/>	Breast Cancer	Clinical	Primary tumor		d, celecoxib,
Homo sapie	144	<input type="checkbox"/>					RNA-Seq an
Homo sapie	12	<input type="checkbox"/>					scRNAseq o
Homo sapie	3	<input type="checkbox"/>					small sampl
Homo sapie	12	<input type="checkbox"/>					pre-treatm
Homo sapie	15	<input type="checkbox"/>					no immuno
Homo sapie	6	<input type="checkbox"/>					no immuno
Homo sapie	18	<input type="checkbox"/>					no immuno

Curation process

Repository	Accession_Num	Title	Description	Query_ID	Dataset_Link	Platform_Type	Organism	Samples	Selected	Tumour_Type	Setting	Biological_Model	Treatment	Treatment_Data_note
GEO	GSE225066	Transcriptomic p	Transcriptomic p RNASeq profiling 2.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE225066	RNA-Seq	Homo sapiens	20	<input checked="" type="checkbox"/>	Bladder ...	Clinical	Primary ...	Immuno ...	neoadjuvancy treatment with bo
GEO	GSE262761	RN11 isomerase	RN11 isomerase Bladder cancer, 2.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE262761	RNA-Seq	Homo sapiens	12	<input checked="" type="checkbox"/>					no immunother
GEO	GSE262761	RN11 isomerase	RN11 isomerase Bladder cancer, 2.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE262761	RNA-Seq	Homo sapiens	12	<input checked="" type="checkbox"/>					no immunother
GEO	GSE278724	De novo methylation	De novo methylation at various combin...		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE278724	RNA-Seq	Homo sapiens	15	<input checked="" type="checkbox"/>	Lung Car...	Prediclin...	PDX	Other + ...	De novo methylation plus pembrolizumab
GEO	GSE278968	SMARCA4-mut	SMARCA4-mut Genomic studies 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE278968	RNA-Seq	Homo sapiens	6	<input type="checkbox"/>					
GEO	GSE269930	SMARCA4-mut	SMARCA4-mut Genomic studies 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE269930	RNA-Seq	Homo sapiens	4	<input type="checkbox"/>					
GEO	GSE269618	SMARCA4-mut	SMARCA4-mut Genomic studies 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE269618	RNA-Seq	Homo sapiens	8	<input type="checkbox"/>					
GEO	GSE269551	SMARCA4-mut	SMARCA4-mut Genomic studies 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE269551	RNA-Seq	Homo sapiens	8	<input type="checkbox"/>					
GEO	GSE266035	Divergent gene	Divergent gene s single-cell RNAf 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE266035	RNA-Seq	Homo sapiens	9	<input checked="" type="checkbox"/>	Lung Car...	Clinical	Other ...	Immuno ...	anti-PD-1/CTLA-RNA-seq on T ce
GEO	GSE230489	RNA-sequencing	RNA-sequencing We used a probe 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE230489	RNA-Seq	Homo sapiens	38	<input type="checkbox"/>					
GEO	GSE233820	PARP inhibitor	PARP inhibitor r Immunotherapy 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE233820	RNA-Seq	Homo sapiens	12	<input type="checkbox"/>					
GEO	GSE234934	Single-cell multi	Single-cell multi Analytical: smg 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE234934	RNA-Seq	Homo sapiens	23	<input type="checkbox"/>					
GEO	GSE243013	A single-cell	A single-cell atfa Anti-PD1/L1 with 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE243013	RNA-Seq	Homo sapiens	243	<input checked="" type="checkbox"/>	Lung Car...	Clinical	Primary ...	Chemo + ...	neoadjuvant anti-PD1 agents cor
GEO	GSE283829	In situ detection	In situ detection Background: lmv 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE283829	RNA-Seq	Homo sapiens	27	<input checked="" type="checkbox"/>	Lung Car...	Clinical	Primary ...	Immuno ...	ICI
GEO	GSE285029	Cell-intrinsic	Cell-intrinsic PD- Background: Th 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE285029	RNA-Seq	Homo sapiens	234	<input checked="" type="checkbox"/>	Lung Car...	Clinical	Primary ...	Immuno ...	PD-1 or PD-L1 blockade)
GEO	GSE235144	PD-1 and TRV	PD-1 and TRV Non-small cell li 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE235144	RNA-Seq	Homo sapiens	6	<input checked="" type="checkbox"/>	Lung Car...	Prediclin...	Cell line	Immuno ...	ICI primary cell line
GEO	GSE234223	Transposable	Transposable ele mRNA splicing e 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE234223	RNA-Seq	Homo sapiens	106	<input type="checkbox"/>					
GEO	GSE259407	Anti-EGFR	Anti-EGFR aptan Non-small cell li 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE259407	RNA-Seq	Homo sapiens	36	<input type="checkbox"/>					
GEO	GSE272174	Characteristics	Characteristics o Samples are fro 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE272174	RNA-Seq	Homo sapiens	9	<input type="checkbox"/>					
GEO	GSE273406	Uncovering the	Uncovering the Harmonic sig 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE273406	RNA-Seq	Homo sapiens	6	<input type="checkbox"/>					
GEO	GSE266364	Immunogenic	Immunogenic ce Immunogenic ce 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE266364	RNA-Seq	Homo sapiens	33	<input type="checkbox"/>					
GEO	GSE280232	Divergent clin	Divergent clin Co-mutations of 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE280232	RNA-Seq	Homo sapiens	42	<input checked="" type="checkbox"/>	Lung Car...	Clinical	Primary ...	Immuno ...	neoadjuvant immunotherapy
GEO	GSE272512	Targeting TRX1	Targeting TRX1 Small cell lung c 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE272512	RNA-Seq	Homo sapiens	6	<input type="checkbox"/>					
GEO	GSE232529	Immune niches	Immune niches We analyzed lmv 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE232529	RNA-Seq	Homo sapiens	58	<input type="checkbox"/>					
GEO	GSE242780	Single-cell	Single-cell RNA s The purpose of 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE242780	RNA-Seq	Homo sapiens	6	<input checked="" type="checkbox"/>	Lung Car...	Clinical	Other ...	Immuno ...	scRNA-seq of PB
GEO	GSE271689	Enhancing imm	Enhancing imm Non-small cell li 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE271689	RNA-Seq	Homo sapiens	586	<input checked="" type="checkbox"/>	Lung Car...	Clinical	Primary ...	Immuno ...	PD-1 based immunotherapies
GEO	GSE260899	Spatially	Spatially resolve The expression 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE260899	RNA-Seq	Homo sapiens	95	<input type="checkbox"/>					
GEO	GSE234070	Tumor-intrinsic	Tumor-intrinsic To better elucid 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE234070	RNA-Seq	Homo sapiens	16	<input type="checkbox"/>					
GEO	GSE273436	Voluntary exer	Voluntary exer Physical activity 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE273436	RNA-Seq	Homo sapiens	8	<input type="checkbox"/>					
GEO	GSE269942	YAP1 status	YAP1 status defn Purpose: Large c 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE269942	RNA-Seq	Homo sapiens	12	<input type="checkbox"/>					
GEO	GSE237818	Enhancing Tumo	Enhancing Tumo In this study, ou 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE237818	RNA-Seq	Homo sapiens	6	<input type="checkbox"/>					
GEO	GSE232620	Blood-based	Blood-based Geo: We integrated 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE232620	RNA-Seq	Homo sapiens	43	<input type="checkbox"/>	Lung Car...	Clinical	Other ...	Immuno ...	Tisliozumab (RNA-seq PBMCs
GEO	GSE269440	Bulk RNA-seq	Bulk RNA-seq of In the evolving 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE269440	RNA-Seq	Homo sapiens	6	<input type="checkbox"/>					
GEO	GSE241977	IFN -induced	IFN -induced by While immunotf 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE241977	RNA-Seq	Homo sapiens	6	<input type="checkbox"/>					
GEO	GSE267108	Single-cell	Single-cell seqe Lung cancer is tr 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE267108	RNA-Seq	Homo sapiens	16	<input type="checkbox"/>					
GEO	GSE241944	Neoadjuvant	Neoadjuvant im We launched an 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE241944	RNA-Seq	Homo sapiens	88	<input type="checkbox"/>	Lung Car...	Clinical	Sorted P...	Chemo + ...	simtlimab and platinum base
GEO	GSE254545	RNA sequencing	RNA sequencing The tumor micr 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE254545	RNA-Seq	Homo sapiens	8	<input checked="" type="checkbox"/>	Lung Car...	Prediclin...	Mouse f...	Immuno ...	CAR-T RNA-seq of CAR-
GEO	GSE266219	Clinical and	Clinical and imm Persistent inflar 8.1		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE266219	RNA-Seq	Homo sapiens	79	<input checked="" type="checkbox"/>	Lung Car...	Clinical	Sorted P...	Chemo + ...	anti-PD1 and JAI sorted cell popu

## Selection Criteria

Primary tumours, metastases, cell lines, mouse models and PDX are eligible for the selection.

Datasets including any **immunotherapy treatments** with at least 10 different samples (5 for single cell experiments).

## Curation Guideline

Each query result should be curated by at least two researchers.

## Workflow – Results Analysis

### Main Phase

Refining of  
System Search  
Strategy

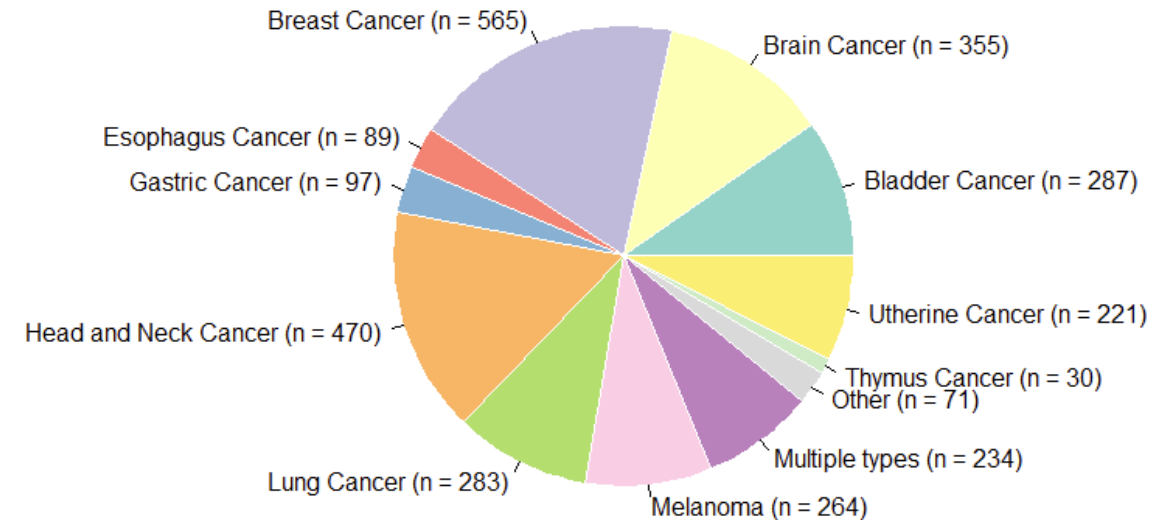
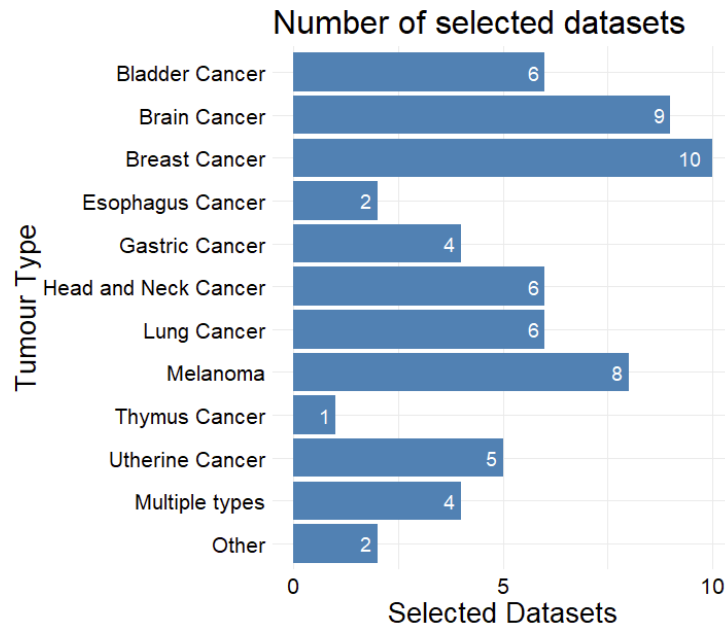
Pancancer  
Dataset  
Gathering

Taskforce  
Dataset  
Curation

Results  
analysis

Only **7 out of 18 queries** were curated by at least 2 researchers.

At the moment, a total of **63 datasets** and **2966 samples** were selected and collected



## Workflow – Results Analysis

### Main Phase

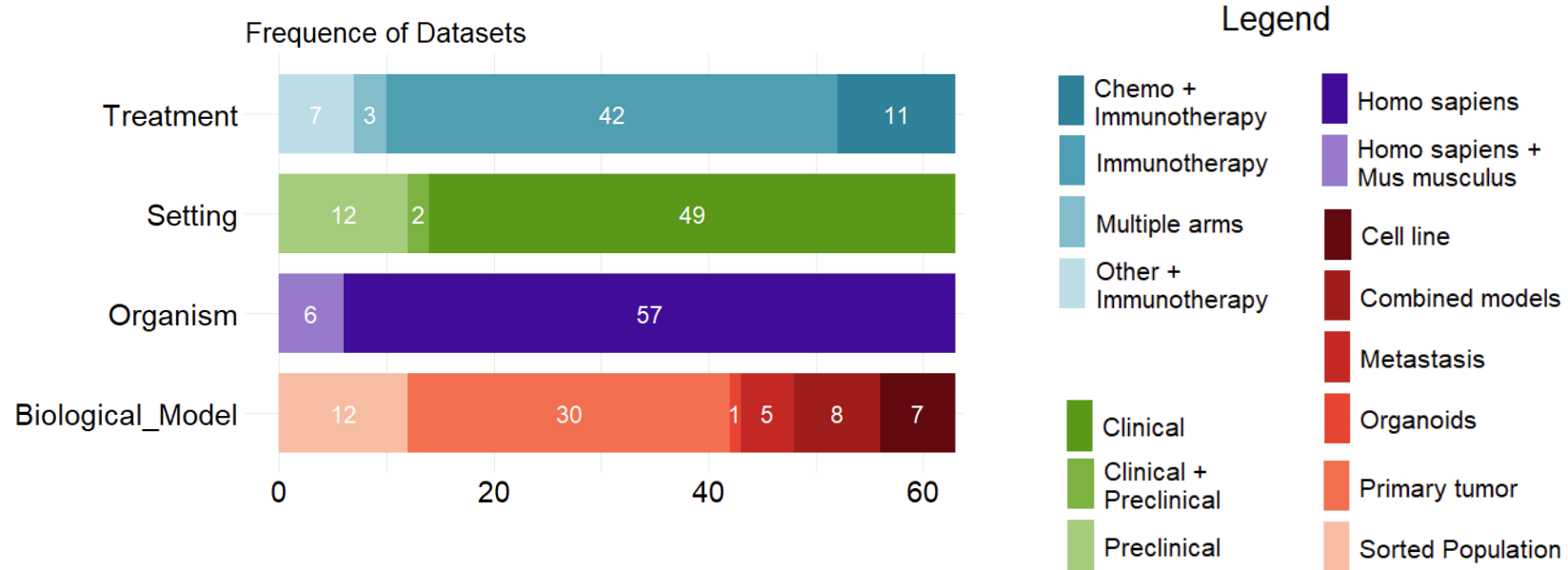
Refining of  
System Search  
Strategy

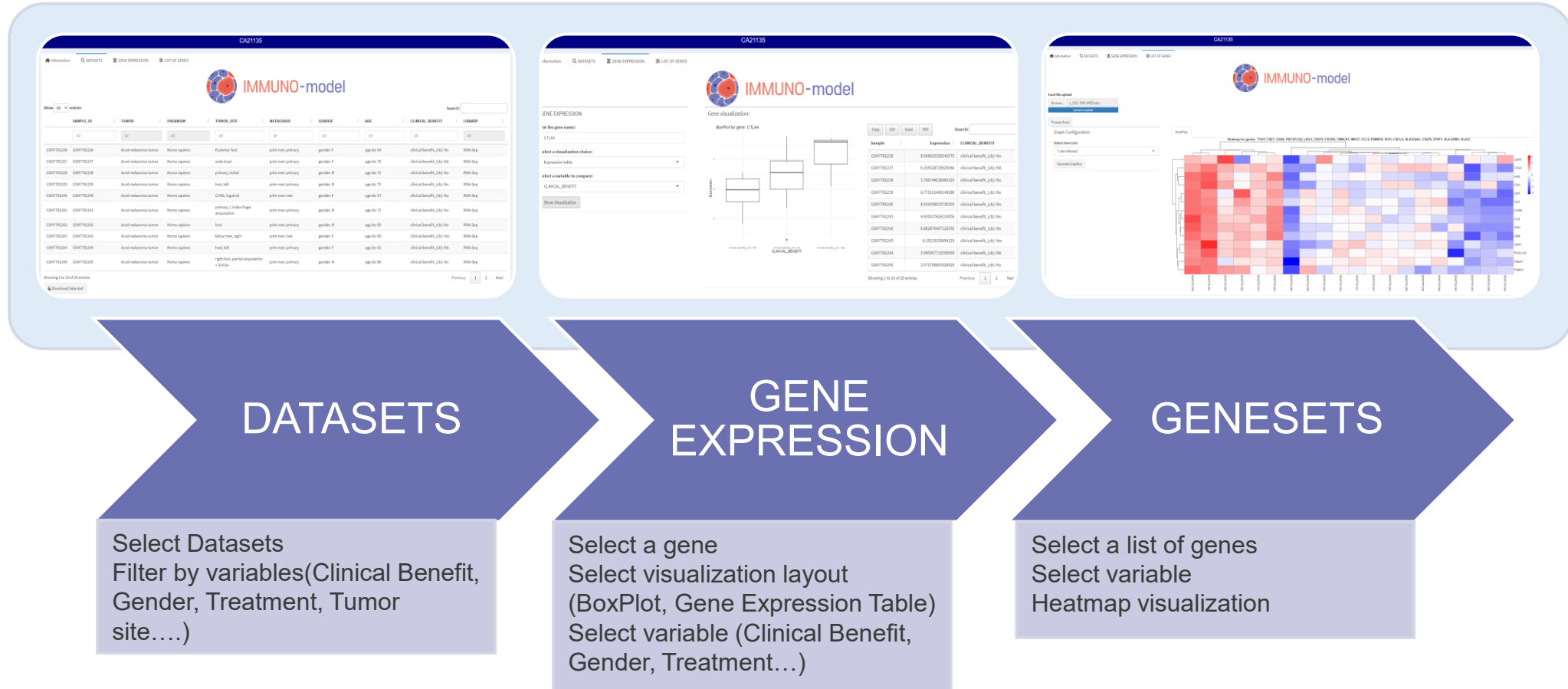
Pancancer  
Dataset  
Gathering

Taskforce  
Dataset  
Curation

Results  
analysis

The majority of datasets collected so far were derived from clinical samples, underlying the scarcity of preclinical models suitable for immunoncology studies.





DATASETS

GENE  
EXPRESSION

GENESETS

- Select Datasets

- Filter by variables  
(Clinical Benefit, Gender,  
treatment, Tumor site....)

CA21135

Information DATASETS GENE EXPRESSION GENESETS

IMMUNO-model

Show 10 entries Search:

SAMPLE_ID	TUMOR	ORGANISM	TUMOR_SITE	METASTASIS	GENDER	AGE	CLINICAL_BENEFIT	LIBRARY
GSM7781236	Acral melanoma tumor	Homo sapiens	R plantar foot	prim met: primary	gender: F	age dx: 54	clinical benefit_(cb): No	RNA-Seq
GSM7781237	Acral melanoma tumor	Homo sapiens	wide local	prim met: primary	gender: F	age dx: 79	clinical benefit_(cb): NA	RNA-Seq
GSM7781238	Acral melanoma tumor	Homo sapiens	primary, initial	prim met: primary	gender: M	age dx: 71	clinical benefit_(cb): No	RNA-Seq
GSM7781239	Acral melanoma tumor	Homo sapiens	foot, left	prim met: primary	gender: M	age dx: 70	clinical benefit_(cb): No	RNA-Seq
GSM7781240	Acral melanoma tumor	Homo sapiens	CLND, inguinal	prim met: met	gender: F	age dx: 57	clinical benefit_(cb): No	RNA-Seq
GSM7781241	Acral melanoma tumor	Homo sapiens	primary, L index finger amputation	prim met: primary	gender: M	age dx: 71	clinical benefit_(cb): No	RNA-Seq
GSM7781242	Acral melanoma tumor	Homo sapiens	foot	prim met: primary	gender: M	age dx: 69	clinical benefit_(cb): No	RNA-Seq
GSM7781243	Acral melanoma tumor	Homo sapiens	femur met, right	prim met: met	gender: F	age dx: 66	clinical benefit_(cb): Yes	RNA-Seq
GSM7781244	Acral melanoma tumor	Homo sapiens	heel, left	prim met: primary	gender: F	age dx: 61	clinical benefit_(cb): NA	RNA-Seq
GSM7781245	Acral melanoma tumor	Homo sapiens	right foot, partial amputation + SLN bx	prim met: primary	gender: M	age dx: 86	clinical benefit_(cb): No	RNA-Seq

Showing 1 to 10 of 20 entries

Download Selected

Previous 1 2 Next

DATASETS

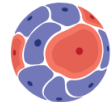
GENE  
EXPRESSION

GENESETS

- Select a gene
- Select visualization layout
- Select variable

CA21135

Information DATASETS GENE EXPRESSION GENESETS

 IMMUNO-model

GENE EXPRESSION

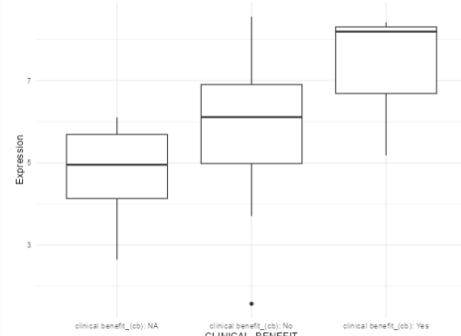
Put the gene name:

Select a visualization choice:  
 Expression table

Select a variable to compare:

Show Visualization

Gene visualization:  
 BoxPlot for gene: CTLA4



Copy CSV Excel PDF Search:

Sample	Expression	CLINICAL_BENEFIT
GSM7781236	8.066620328345575	clinical benefit_(cb): No
GSM7781237	6.105528729625046	clinical benefit_(cb): NA
GSM7781238	3.706744038985325	clinical benefit_(cb): No
GSM7781239	6.775161466148188	clinical benefit_(cb): No
GSM7781240	8.554258533730303	clinical benefit_(cb): No
GSM7781241	4.910027626210005	clinical benefit_(cb): No
GSM7781242	6.882876467128356	clinical benefit_(cb): No
GSM7781243	8.19219159694129	clinical benefit_(cb): Yes
GSM7781244	3.990367731595959	clinical benefit_(cb): NA
GSM7781245	1.571799865926929	clinical benefit_(cb): No

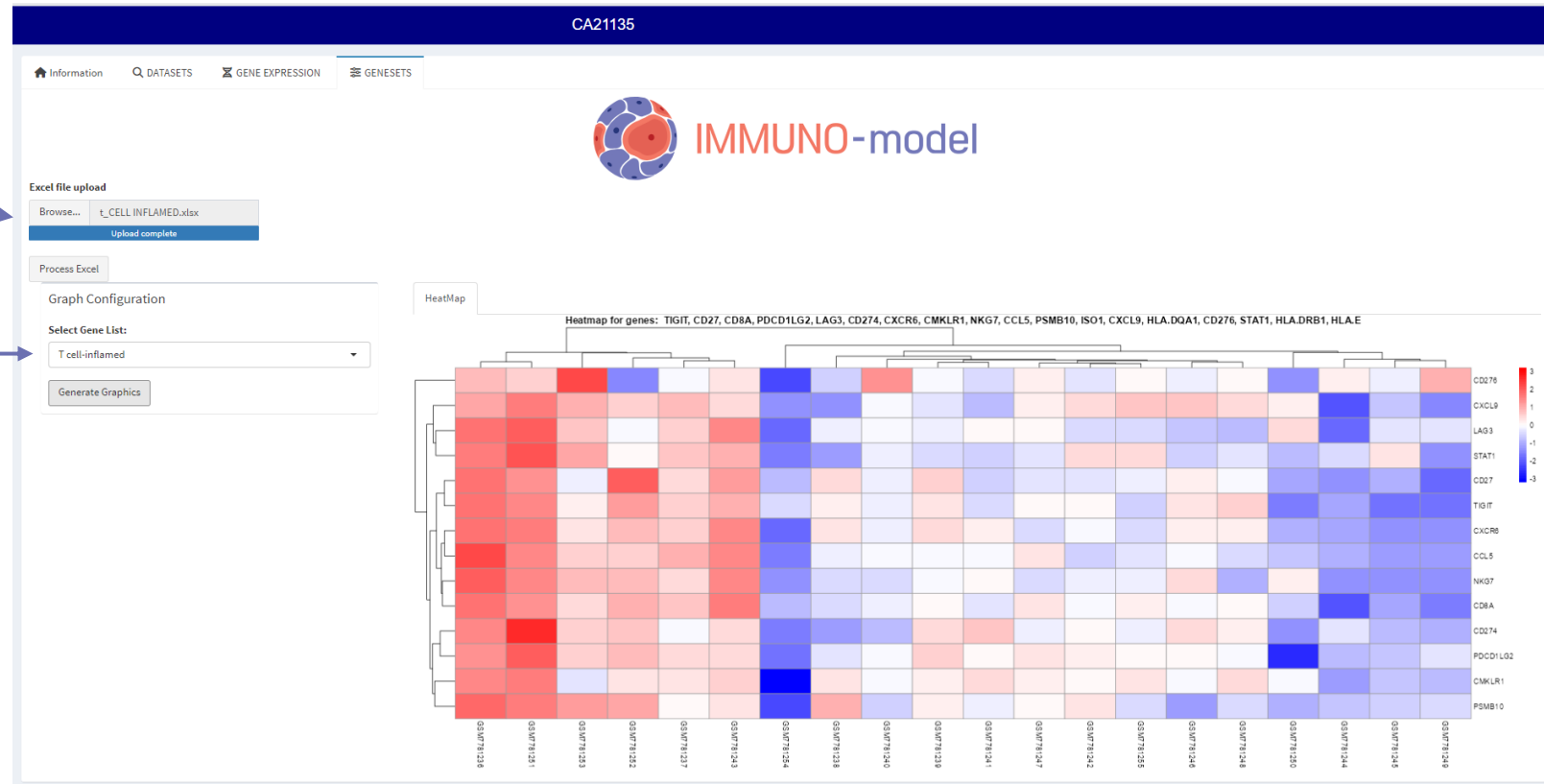
Showing 1 to 10 of 20 entries Previous 1 2 Next

DATASETS

GENE  
EXPRESSION

GENESETS

- Select a list of genes
- Select visualization layout



## Conclusion

- We are carrying out a systematic identification of bulk and single cell transcriptomic data from pre-clinical and clinical samples treated with immunotherapy
- We developed a Systematic Search Strategy to streamline the dataset search, curation and annotation workflow
- 63 informative datasets have been identified so far, providing transcriptomic data from a range of clinical and preclinical samples
- We established knowledge-sharing channels that could foster the connection between basic, translational, and clinical investigators;
- Parallel development of a graphical interface will provide easy access to the collected resources

## Next steps

- To complete the curation process of all the queries
- To continue the query generation process
  - to include mouse and rat as organisms
  - to interrogate ArrayExpress repository.
- To include as many datasets as possible into the graphic tool
- To expand the functionalities of the tool

# Acknowledgement

Eduardo Aranga Cañada, Istituto de Investigación Sanitaria Aragón

Rebeca Sanz-Pamplona, Istituto de Investigación Sanitaria Aragón

Maurizio Callari, Fondazione Michelangelo

Daniela Besozzi, University of Milano-Bicocca

Luca Gianni, Fondazione Michelangelo



FONDAZIONE MICHELANGELO  
avanzamento dello studio e cura dei tumori

## Curators

Ester Raigón Gómez

Ida Lettiero

Jessica Dal Col

Daniel Josef Lindeger

Juan Carlos Rodriguez-Manzaneque

Lenka Trnkova

Oriana Barros

Ahmed Hassan

Marilina Garcia Aranda

Vinton Cheng

Anna Falanga

Eduardo Candeal Nuñez

## Collaborators

Abolaji Olagunju

Andrea Moreno Manuel

Andreia Brandão

Anguraj Sadanandam

Bozena Smolkova

Catarina A. Rodrigues

Chun-Wai Mai

Daniel Sur

Filipa Reis

Hamed Allahverdi

Hatim Boughanem

Marco Russo

María Julia Lamberti

Mustafa Samet Pir

Nafsika Rossopoulou

Noel Gómez Molina

Rosalinda Sorrentino

Sandra Garcia-Mulero

Theodora Katsila

Ugur Sezerman

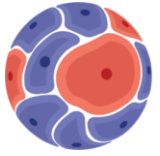
Yanis Saidani

Scan the QR code to  
discover the Data  
Gathering Taskforce!



Funded by  
the European Union

This work has been supported by the **Virtual Mobility Grant** guaranteed by the **European Cooperation in Science & Tehnology**, action number **CA21135**



# Thank you for your attention