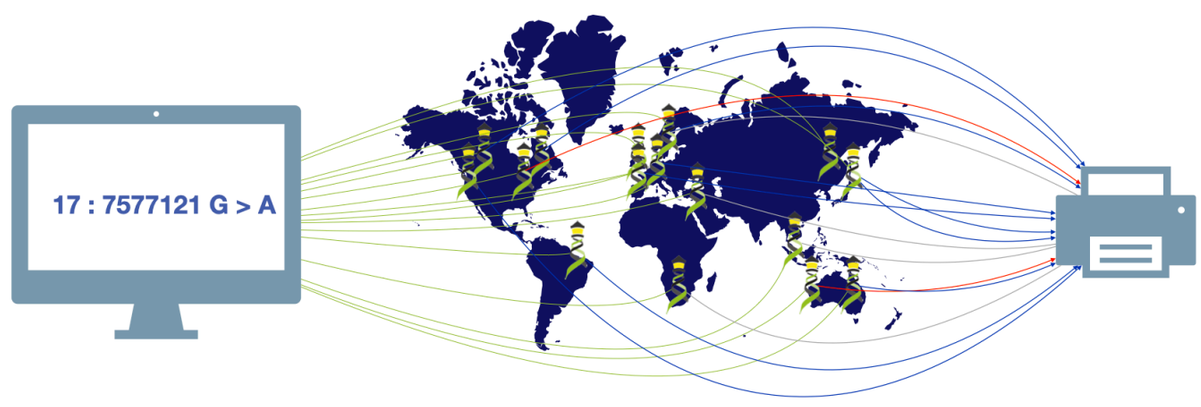
cancercelllines.org - a new curated resource for cancer cell line variants

Rahel Paloots[1], Ellery Smith[2], Dimitri Giagkos[3], Kurt Stockinger[2], Michael Baudis[1]

1 University of Zurich, Switzerland 2 Zurich University of Applied Sciences, Switzerland 3 Infili Technologies, Greece

A Beacon-based Framework for the Representation and Queries of Cancer Cell Line Variants

- The Beacon framework was created for global and federated queries [1]
- cancercelllines.org is the daughter of progenetix [2] (a cancer copy number variant (CNV) database), including over 5600 cell line CNVs
- We have mapped known cancer cell line variants from resources like CCLE and ClinVar to our database, resulting in 16178 cell lines from 400 different disease classifications (NCIt)



Have you seen this variant? It came up in my patient and we don't know if this is a common SNP or worth following up.

A Beacon network federates *genome variant queries* across databases that support the ***Beacon API***

Here: The variant has been found in **few** resources, and those are from **disease** specific **collections**

Source	Number of cell lines
ClinVar	15384
CCLE	1417
CNV	2138

Depiction of CNVs and SNVs of Cancer Cell Lines

- All cell lines are mapped to to Cellosaurus [3] IDs and have hierarchical representation
- SNVs mapped from ClinVar [4] show known severity of the variant and disease ontologies
- CCLE [5] variants are all known SNVs per cell line and include information about variant effect

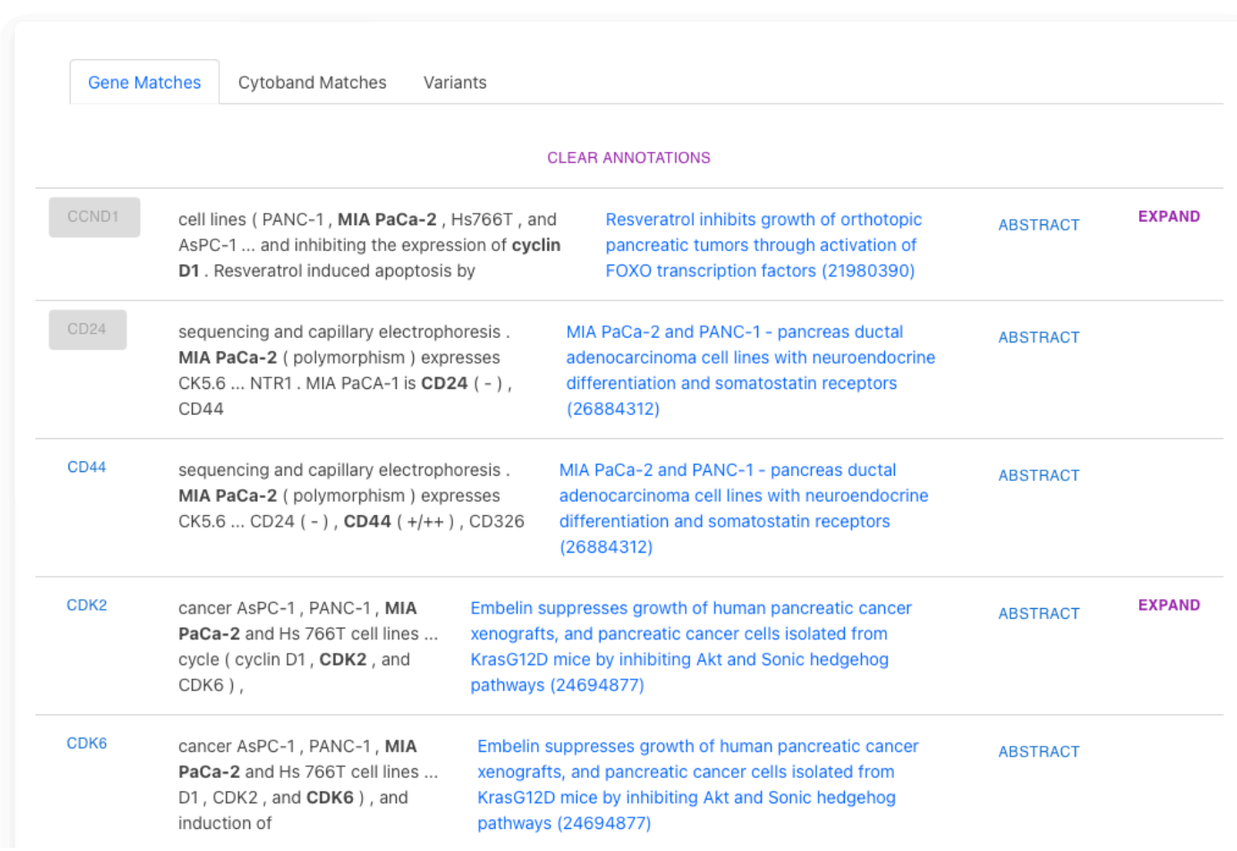
Annotated Variants for cellosaurus:CVCL_0428

Digest	Gene	Pathogenicity	Variant Effect	Variant Instances
17:7674220-7674221:G>A	TP53	Pathogenic		V: pgxvar-63ce6a913319d2172d27d545 B: pgxbs-8B88A6e3 V: pgxvar-63ce6a913319d2172d27d546 B: pgxbs-06a58B1c V: pgxvar-63ce6a913319d2172d27d54a B: pgxbs-DEcFEDB4
17:81914368-81914369:C>A	SIRT7		Missense variant	V: pgxvar-63ce6ad3a24c83054b8e3876 B: pgxbs-8B88A6e3
10:11870140-11870141:C>T	PROSER2		Missense variant	V: pgxvar-63ce6ad3a24c83054b8e37f9 B: pgxbs-8B88A6e3
1:42154814-42154815:C>T	GUCA2B		Missense variant	V: pgxvar-63ce6ad3a24c83054b8e37d1 B: pgxbs-8B88A6e3
13:110186511-110186512:T>G	COL4A1		Silent mutation	V: pgxvar-63ce6ad3a24c83054b8e3833 B: pgxbs-8B88A6e3
2:98236431-98236432:C>T	VWA3B		Missense variant	V: pgxvar-63ce6ad3a24c83054b8e38bb B: pgxbs-8B88A6e3

Information Enrichment of Genes in Cancer Cell Lines

We performed information extraction using natural language processing from abstracts [6] and enriched cell lines in our database with additional information about gene and cytoband matches.

- Display selected genes on CNV plots
- Expand and display abstracts
- Links to publications



[1] Rambla J, Baudis M, Ariosa R, Beck T, Fromont LA, Navarro A, Paloots R, Rueda M, Saunders G, Singh B, Spalding JD, Törnroos J, Vasallo C, Veal CD, Brookes AJ. Beacon v2 and Beacon networks: A ”lingua franca” for federated data discovery in biomedical genomics, and beyond. Hum Mutat. 2022 Mar 17. doi: 10.1002/humu.24369. PMID: 35297548. [2] Huang Q, Carrio-Cordo P, Gao B, Paloots R, Baudis M. The Progenetix oncogenomic resource in 2021. Database (Oxford). 2021 Jul 17; 2021:baab043.doi:10.1093/database/baab043. PMID: 34272855; PMCID: PMC8285936. [3] https://www.cellosaurus.org/ [4] https://www.ncbi.nlm.nih.gov/clinvar/ [5] https://sites.broadinstitute.org/ccle/ [6] Ellery Smith, Dimitris Papadopoulos, Martin Braschler, and Kurt Stockinger. Lillie: Information extraction and database integration using linguistics and learning-based algorithms. Information Systems, 105:101938, 2022.