

Collaborate. Innovate. Accelerate.

Reference Resources and Standards Development for Biomedical Genomics and Cancer Research



Michael Baudis

Professor of Bioinformatics University of Zürich







- Swiss Institute of Bioinformatics **SIB**
- GA4GH Workstream Co-lead DISCOVERY
- Co-lead ELIXIR Beacon API Development





Department of Molecular Life Sciences

Theoretical Cytogenetics and Oncogenomics

Our work @ UZH:

- cancer genome repositories
- biocuration
- protocols & formats







Michael Baudis :: 2023-10-10



Department of Molecular Life Sciences

Genome screening at the core of "Personalised Health"

- Genome analyses (including transcriptome, metagenomics) are core technologies for Personalised Health[™] applications
- The unexpectedly large amount of sequence variants in human genomes - germline and somatic/cancer - requires huge analysis efforts and creation of reference repositories
- Standardized data formats and exchange protocols are needed to connect these resources throughout the world, for reciprocal, international data sharing and biocuration efforts





0-10 Michael Baudis :: 2023



Global Genomic Data Sharing Can...



Demonstrate patterns in health & disease



Increase statistical significance of analyses



Lead to "stronger" variant interpretations



Increase accurate diagnosis



Advance precision medicine



Different Approaches to Data Sharing









Centralized Genomic Knowledge Bases

Data Commons

Trusted, controlled repository of multiple datasets



Hub and Spoke

Common data elements, access, and usage rules





Linkage of distributed and disparate datasets



The Swiss Personalized Health Network (SPHN)

Creation of a scalable and sustainable data-enabling environment

- Including routine health data, molecular / omics data, registry data, clinical research data, and other health-related data types
- Research infrastructure initiative funded 2017-2024 by the Swiss Government with CHF135 million;
- Operating under a common Ethical Framework and one Information Security Policy, incl. the setup of a Trusted Research Environment
- Foreseen consolidation of data coordination efforts with CHF 21 million 2025-2028
- Enable **institutions** to responsibly share interoperable health data \rightarrow
- Enable researchers to access, integrate, and analyze data \rightarrow





October 2023









The Swiss Personalized Health Network













Different Approaches to Data Sharing



Centralized Genomic Knowledge Bases





Data Commons

Trusted, controlled repository of multiple datasets









Hub and Spoke

Common data elements, access, and usage rules

Linkage of distributed and disparate datasets



progenetix.org

Cancer Genomics Reference Resource

- open resource for oncogenomic profiles
- over **116'000 cancer CNV profiles**
- more than 800 diagnostic types
- inclusion of reference datasets (e.g. TCGA)
- standardized encodings (e.g. NClt, ICD-O 3)
- identifier mapping for PMID, GEO, Cellosaurus, TCGA, cBioPortal where appropriate
- core clinical data (TNM, sex, survival ...)
- data mapping services
- recent addition of SNV data for some series







Regional CNV Frequencies for >800 Cancer Types



Progenetix R API using Beacon handover objects



Visualization of CNV features using the pgxRpi R package. Aggregated CNV data for cancer typesis displayed using Circos or frequency plots in a local R environment. The R package relies on the Beacon v2 API to communicate with Progenetix.











Number of tumor samples for each publication across the years



Cancer CNV Profiles

Search Samples

Studies & Cohorts

arrayMap

TCGA Samples

DIPG Samples

Gao & Baudis, 2021

Cancer Cell Lines

Publication DB

Services

NCIt Mappings

Progenetix Publication Collection

The current page lists articles describing whole genome screening (WGS, WES, aCGH, cCGH) experiments in cancer, registered in the Progenetix publication collection. For each publication the table indicates the numbers of samples analysed with a given technology and if sample profiles are available in Progenetix.

Please contact us to alert us about additional articles you are aware of. The inclusion criteria are described in the documentation \mathscr{O} .

Filter 🕕	City 🕕		
	Type to search		



Map of the geographic distribution (by affiliation) of the 104'543 genomic array, 36'766 chromosomal CGH and 15'409 whole genome/exome based cancer genome datasets. The numbers are derived from the 3'240 publications in Progenetix.

pg x



The Progenetix oncogenomic resource in 2021

Qingyao Huang^{1,2}, Paula Carrio-Cordo^{1,2}, Bo Gao^{1,2}, Rahel Paloots^{1,2} and Michael Baudis^{1,2,*}

¹Department of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland ²Swiss Institute of Bioinformatics, Winterthurerstrasse 190, Zurich 8057, Switzerland

*Corresponding author: Tel: +41 44 635 34 86; Email: michael.baudis@mls.uzh.ch

Citation details: Huang, Q., Carrio-Cordo, P., Gao, B. et al. The Progenetix oncogenomic resource in 2021. Database (2021) Vol. 2021: article ID baab043; DOI: https://doi.org/10.1093/database/baab043

Abstract

In cancer, copy number aberrations (CNAs) represent a type of nearly ubiquitous and frequently extensive structural genome variations. To disentangle the molecular mechanisms underlying tumorigenesis as well as identify and characterize molecular subtypes, the comparative and meta-analysis of large genomic variant collections can be of immense importance. Over the last decades, cancer genomic profiling projects have resulted in a large amount of somatic genome variation profiles, however segregated in a multitude of individual studies and datasets. The Progenetix project, initiated in 2001, curates individual cancer CNA profiles and associated metadata from published oncogenomic studies and data repositories with the aim to empower integrative analyses spanning all different cancer biologies. During the last few years, the fields of genomics and cancer research have seen significant advancement in terms of molecular genetics technology, disease concepts, data standard harmonization as well as data availability, in an increasingly structured and systematic manner. For the Progenetix resource, continuous data integration, curation and maintenance have resulted in the most comprehensive representation of cancer genome CNA profiling data with 138 663 (including 115 357 tumor) copy number variation (CNV) profiles. In this article, we report a 4.5-fold increase in sample number since 2013, improvements in data quality, ontology representation with a CNV landscape summary over 51 distinctive National Cancer Institute Thesaurus cancer terms as well as updates in database schemas, and data access including new web front-end and programmatic data access.

Database URL: progenetix.org

Data source	GEO	ArrayExpress	cBioPortal	TCGA	
No. of studies	898	51	38	33	
No. of samples Tumor Normal	63 568 52 090 11 478	4351 3887 464	19 712 19 712 0	22 142 11 090 11 052	
Classifications ICD-O (Topography) ICD-O (Morphology) NCIt	100 246 346	54 908 148	88 265 422	157 140 182	
Collections Individuals Biosamples Callsets ^a Variants	63 568 63 568 63 568 5 514 126	4351 4351 4351 1184170	19 712 19 712 19 712 1 778 096	10 995 22 142 22 376 2 654 065	

Table 1. Statistics of samples from various data resources

^aset of variants from one genotyping experiment; ICD-O, International Classification of Diseases for Oncology; NCIt, National Cancer Institute Thesaurus.



ataset					and the second									CRORCHIDH 2 CT
progenetix ×					XIV		Contraction of the local distance of the loc			- Ouroury		A CONTRACTOR	Vidreme	A and the
eference name		(Structural) Variant Type				+	1 Same			50 50 m		Latvij	an the	IEAOCHL E
8	í.	V DUP (Duplication)			1.0	-				Danmark		Sun	22	~ ~
		- Our (ouplication)			1.*	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	United Ki	ngdom		København		Locals	5 to	notice of
tart or Position		End (Range or Structural Var.	.)			- 5	Belfast	Great Britain		11250	Gdansk	Калининград		and the
124000000-127736593		127740957-130000000				(and a	Alsie of M	an Leeds	- F	Holstein Wecklenburg Vorpommern	- 1 - L	Гродно	Белару	Ch Mor
ancer Classification(s)	•					Eire / Ire	eland Marci	ester Sheffield	Groningen	Hamburg Szczecin	Bydgoszcz	Bualyst	tok	5-6
icdom-80463: Non-small cell carcin	noma (221 B				× ~	100	A Gen	Birmingham	Nederland	dersachsen Berlin	Poznań	e Warszawa, 06	ecmicana Snocme	Comencian
iosample Type						6	- Gi	ardiff London	Dusseldort-	Magdeburg	Mil	Lubin	The	mater
					1 ~			and	Belgie O Fra	ankfurt Dresden	Wroclaw	Start I	Пуцья	- Ser
ilters 0		Filter Logic 0					an	and the state of t	Belgien	Nurnberg Ces	sko Krak	ON Diss	житомир	Kutin
6		AND			14			Paris	Grand fat	anna f	- Clause	and a	Bao	Harts
					1.20			Regnes	6.	Stutigart	Wien	and in	A RON	Кропивницы
ity 🔽						Sample De	Details	*						
Select		× .				DOX AM	RE 2207	GEA of duta 11	004 (progon	(ativ)				
			start	end		PGA_AMI_		0054-pii-uki2_11	_004 (progen	letix)				
			124000000	12000	00000	Description								
			124000000	13000		Glioblastoma								
Query Beacon						Diagnostic C	Classification	s						
						• icdot-C71	1.0: Cerebrum	21						
						· icdom-944	4403: Glioblast	ma						
Search Samples 📑						 NCIT:C30! 	058: Glioblastor	na						
						Clinical Data	a							
ssembly: GRCh38 Chro: 8	Start: 12400000-127736593 End:	127740957-130000000 Type:	DUP Filters: icdom-804	163		• Death: 0 (4	(at 2.0833333	333 months)						
						Provenance	,							
progenetix	A					Material: n	neoplastic sam	ble						
progenetix	•					Material: nOrigin: Hei	neoplastic sam eidelberg, Germ	ble any						
progenetix Samples: 135 Variante: 120	Phenopackets D UCSC Collector Verienter D UCSC	Cregion C Visual	lization options			Material: n Origin: Hei External Refe	neoplastic sam eidelberg, Germ ferences	ole any						
progenetix Samples: 135 Variants: 120 Calls: 135	Phenopackets & UCSC Callsets Variants & JSON Variants in UCSC &	C region 🗹 Visuali	ization options			Material: n Origin: Hei External Refe PMID:230	neoplastic sam eidelberg, Germ ferences 079654	ole any						
progenetix Samples: 135 Variants: 120 Calls: 135 f _{alleles} : 0.000972	Phenopackets & UCSC Callsets Variants & JSON Variants in UCSC &	C region 년 I Response 년	lization options			Material: n Origin: Hei External Refe PMID:230 CNV Profile(s)	neoplastic sam eidelberg, Germ ferences 079654 (s)	ole any						
progenetix Samples: 135 Variants: 120 Calls: 135 falleles: 0.000972 Results Biosamples B	Phenopackets & UCSC Callsets Variants & JSON Variants in UCSC & JSON	C region 🗹 Response 🗹	lization options			Material: n Origin: Hei External Refe PMID:230 CNV Profile(:	neoplastic sam eidelberg, Germ ferences 079654 (\$)	any any			10 11			17 18
progenetix Samples: 135 Variants: 120 Calls: 135 f _{alletes} : 0.000972 Results Biosamples E	Phenopackets C UCSC Callsets Variants C UCSC Variants in UCSC C UCSC Biosamples Map Variants	Cregion 🗹 Visuali	lization options			Material: n Origin: Hei External Refe PMID:230 CNV Profile(:	neoplastic sam eidelberg, Germ ferences 079654 (s)	ole any Alla u uu <mark>a</mark> uuu uuuu	angan dang dang dang dang dang dang dang	1 min min yên	10 11 11	12 13 14	15 N.	17 18 111 11
progenetix Samples: 135 Variants: 120 Calls: 135 faileles ² 0.000972 Results Biosamples B	Phenopackets C UCSC Callsets Variants C JSON Variants in UCSC C	C region & Visual Response &	lization options			Material: n Origin: Hei External Refe PMID:230 CNV Profile(:	neoplastic sam, eidelberg, Germ ferences 079654 (\$) 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3	ole any 111 1000 1000 1000 1000	andra andra	r mán nám yén	²⁶ 17	12 13 14	5	
progenetix Samples: 135 Variants: 120 Calls: 135 faileles: 0.000972 Results Biosamples Biosamples	Phenopackets & UCSC Callsets Variants & JSON Variants in UCSC & JSON	C region & Visual Response & Visual	12 13 14 15 16	n ola olla	19 20 21 22 10 10 10 10 10 10	Material: n Origin: Hei External Refe PMID:230 CNV Profile(: per: AM201	neoplastic sam, eidelberg, Germ ferences 079654 (s) 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3		uning uning	man Dec. 201	13 (GPC)	1. h38/ha39	н ійн ійн 8) Асса	an particular and a second sec
progenetix Samples: 135 Variants: 120 Calls: 135 falleles: 0.000972 Results Biosamples E	Phenopackets & UCSC Callsets Variants & UCSC Variants in UCSC & JSON Biosamples Map Variants	C region & Visual I Response &	1/2 13 14 15 19	n ata ana	9 20 21 22 10 10 10 10 20 20	Material: n Origin: Hei External Refe PMID:230 CNV Profile(: per:: AUGO per:: AUGO CNV	neoplastic sam, eidelberg, Germ ferences 079654 (s) 2 2 3 4 4 5 5 5 5 5 6 6 6 7 6 7 6 7 7 7 7 7 7 7 7	nome Brow	ser on Hu	uman Dec. 201	13 (GRC	138/hg38	8) Asse	embly
progenetix Samples: 135 Variants: 120 Calls: 135 falleles: 0.000972 Results Biosamples E	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants	C region C Visual	lization options			Material: n Origin: Hei External Refe PMID:230 CNV Profile(pres: AUGO pres: AUGO UC	neoplastic sam eidelberg, Germ ferences 079654 (s) CSC Ge move <	nome Brow	ser on Hu	Iman Dec. 201 min 1.5x 3x 10x	13 (GRC	h38/hg38	B) Asse	uíu uúu embly ∝
progenetix Samples: 135 Variants: 120 Calls: 135 failetes: 0.000972 Results Biosamples E	Phenopackets & UCSC Callsets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants	C region C Visual Response C	Ization options		19 20 21 22 10 1	 Material: n Origin: Hei External Refe PMID:230 CNV Profile(: 	neoplastic sam eidelberg, Germ ferences 079654 (s) CSC Ge move chr9:21,531	any any any any any any any any	ser on Hu >>>>> zoor 261,586 bp. en	Iman Dec. 201 n in 1.5x 3x 10x ter position, gene symbol, H0	13 (GRC base zoom	h38/hg38 out 1.5x 3x	B) Asse	win min embly ∝) ∞
progenetix Samples: 135 Variants: 120 Calls: 135 faileles: 0.000972 Results Biosamples E	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants	C region C Visual	ization options		19 20 21 22 20 21 22 20 21 22 20 21 22 20 21 22 20 21 20 20 20 21 20 20 21 20 20 21 20 20 21 20 20 21 20 20 20 20 20 20 20 20 20 20 20 20	 Material: n Origin: Hei External Refe PMID:230 CNV Profile(s CNV Profile(s) CNV Profile(s) 	neoplastic sam, eidelberg, Germ ferences 079654 (s) 2 CSC Ge move chr9:21,53* (rre (s21.5)	eneme Brow (< << > > ,306-22,492,891 §	ser on Hu >> >>> zoor 961,586 bp. en	Iman Dec. 201 m in 1.5x 3x 10x Iter position, gene symbol, He	10 11 13 (GRC base zoom GVS or search ter	h38/hg38 out 1.5x 3x ms	B) Asse	embly
progenetix Samples: 135 Variants: 120 Calls: 135 Jaileles: 0.000972 Results Biosamples E	Phenopackets & UCSC Callsets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants	Cregion C Visual Response C	ization options		19 20 21 22 2021 20 2021 2021 202 2021 properties arg	 Material: n Origin: Hei External Refe PMID:230 CNV Profile(: per: AUXes 	neoplastic sam, eidelberg, Germ ferences 079654 (s) 2 CSC Ge move < chr9:21,531 crr9 (s21.5) scale	ele any nome Brow (< << < > ,306-22,492,891 \$	ser on Hu >> >>> zoor 261,586 bp. en	Iman Dec. 201 min 1.5x 3x 10x ter position, gene symbol, Hi	13 (GRC base zoom GVS or search ter	h38/hg38 out 1.5x 3x ms	B) Asse 10x 10	embly
progenetix Samples: 135 Variants: 120 Calls: 135 faileles: 0.000972 Results Biosamples B	Phenopackets C Calisets Variants C Variants in UCSC C Biosamples Map Variants	Cregion & Visual Response & Visual	Subset Match Frequencies		19 20 21 22 19 20 20 20 21 20 20 21 propendix.org	Material: n Origin: Hei External Refe PMID:230 CNV Profile(: per: AMOP	neoplastic sam, eidelberg, Germ ferences 079654 (s) CSC Ge move < chr9:21,531 crr9 (s21.3) scate	any any any any any any any any	Ser on Hu >> >>> zoor 61,586 bp. en	Iman Dec. 201 m in 1.5x 3x 10x ter position, gene symbol, H0 9912 10 280 top: 1000 1000 1000 1000 1000 1000 1000 10	13 (GRC base zoom GVS or search ter 22, 100, 000 122	h38/hg38 h38/hg38 out 1.5x 3x ms EIHE 3955 39 22,3	B) Asse 10x 10 22 24 24 24 24 24 24 24 24 24 24 24 24	embly
progenetix Samples: 135 Variants: 120 Calls: 135 faileles: 0.000972 Results Biosamples E	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants	Cregion & Visual Response & Visual	Subset Match Frequencies		10 20 21 22 10 10 10 10 10 20 21 22 20 21 propretingent	Material: n Origin: Hei External Refe PMID:230 CNV Profile(: per:: AUGO (neoplastic sam eidelberg, Germ ferences 079654 (s) CSCC Ge move < chr9:21,53* chr9:(e21.3) scale	any any any any any any any any	Ser on Hu >> >>> zoor 2001 2102 21,000,000 21,9	Iman Dec. 201 min 1.5x 3x 10x ter position, gene symbol, Ho ee, evel 22, ever, evel 2 cot. vier land a set of cot. vier land a set of	13 (GRC) base zoom GVS or search ter	12 13 14 h38/hg38 out 1.5x 3x ms b100 3225 30 ,224, eeel 22, 3	B) Asse 10x 10 200,000 22	embly
progenetix Samples: 135 Variants: 120 Calls: 135 falleles: 0.000972 Results Biosamples E	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples 2217 7037	Cregion & Visual Response & Visual Query Matches 135 135	Subset Match Frequencies		9 20 21 22 10 10 10 10 20 20 20 20 20 20 20 20 20 20	Material: n Origin: Hei External Refe PMID:230 CNV Profile(: per:: MODE CNV	neoplastic sam eidelberg, Germ ferences 079654 (s) CSCC Ge move < chr9:21,537 chr9:(e21.3) Scale	any any any any any any any any	Ser on Hu >> >>> zoor 961,586 bp. en	Iman Dec. 201 min 1.5x 3x 10x ter position, gene symbol, Ho ee, eeo 22, eeo, eeol 2 cott var lands and on	13 (GRC) base zoom GVS or search ter	h38/hg38 out 1.5x 3x ms baba 3322	B) Asse 10x 10 Dee, eee] 22	=====================================
progenetix Samples: 135 Variants: 120 Calls: 135 falleles: 0.000972 Results Biosamples E 2 2 7 7 7 7 7 7 7 7 7 7 7 7 7	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples 2217 7037 9483	Cregion & Visual Response & Visual Query Matches 135 135	Subset Match Frequencies		9 20 21 22 10 10 10 10 20 20 20 21 properties.org	Material: n Origin: Hei External Refe PMID:230 CNV Profile(: per:: MORE	neoplastic sam eidelberg, Germ ferences 079654 (s) CSCC Ge move < chr9:21,537 chr9:(e21.3) Scale	any any any any any any any any	Ser on Hu >> >>> zoor 961,586 bp. en	Iman Dec. 201 min 1.5x 3x 10x ter position, gene symbol, Ho ee, eeol 22, eeo, eeel 2 cott var lands and on	13 (GRCI base zoom GVS or search ter	h38/hg38 out 1.5x 3x ms baba 3355	B) Asse 10x 10 Dee, eee] 22	embly
progenetix Samples: 135 Variants: 120 Calls: 135 falleles: 0.000972 Results Biosamples E 2 2 2 2 2 2 2 2 2 2 2 2 2	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples 2217 7037 9483	Cregion & Visual Response & Visual Query Matches 135 135	Subset Match Frequencies		9 20 21 22 10 10 10 10 10 20 20 20 20 20 20 programmin.org	Material: n Origin: Hei External Refe PMID:230 CNV Profile(: per:: MORE	neoplastic sam eidelberg, Germ ferences 079654 (s) CSCC Ge move < chr9:21,537 chr9:(e21.3) Scale	any any any any any any any any	Ser on Hu >> >>> zoor 961,586 bp. en	Iman Dec. 201 min 1.5x 3x 10x ter position, gene symbol, Ho ee, eeol 22, eeo, eeol 2 cott var lands and on	13 (GRCI base zoom GVS or search ter	h38/hg38 out 1.5x 3x ms base 22,3	B) Asse 10x 10 Dee, eee] 22	embly
progenetix Samples: 135 Variants: 120 Calls: 135 falleles: 0.000972 Results Biosamples E	Phenopackets & UCSC Callsets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples 2217 7037 9483 osamples Map Variants	Cregion & Visual Response & Visual Query Matches 135 135	Subset Match Frequencies		9 20 21 22 757 200 200 200 200 200 200 200 200 200 20	Material: n Origin: Hei External Refe PMID:230 CNV Profile(pre:: XXXX	neoplastic sam eidelberg, Germ ferences 079654 (s) CSCC Ge move < chr9:21,531 crr9 (s21.3) Scale	any any any any any any any any	Ser on Hu >> >>> zoor 961,586 bp. en	Iman Dec. 201 min 1.5x 3x 10x ter position, gene symbol, Ho 2012 13 2014 22, 000, evel 2 2014 23 2014 23 2014 23 2014 24 2014 24 24 24 24 24 24 24 24 24 24 24 24 24 2	13 (GRC) base zoom GVS or search ter	h38/hg38 out 1.5x 3x ms bbb 33255	B) Asse 10x 10 200 10 200 10	ambly
progenetix Samples: 135 Variants: 120 Calls: 135 faileles: 0.000972 Results Biosamples E 2 2 7 7 7 7 7 7 7 7 7 7 7 7 7	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples 2217 7037 9483 osamples Map Variants	Cregion & Visual Response & Visual Query Matches 135 135	Subset Match Frequencies 0.061 0.019 0.014	n 170 olin 1 170 olin 1	19 20 21 22 19 20 21 22 257 257 257 257 257 257 257 2	Material: n Origin: Hei External Refe PMID:230 CNV Profile(pres: AMOS	neoplastic sam eidelberg, Germ ferences 079654 (s) CSCC Ge move < chr9:21,531 chr9:(p21.5) scate	any any any any any any any any	Ser on Hu >> >>> zoor 201,586 bp. en	Iman Dec. 201 min 1.5x 3x 10x ter position, gene symbol, Hi 9010 10 2010 10 20	13 (GRC) base zoom GVS or search ter	h38/hg38 out 1.5x 3x ms b2001 93/222	B) Asse 10x 10 20 20 20 20 20 20 20 20 20 20 20 20 20	and in the second secon
progenetix Samples: 135 Variants: 120 Calls: 135 Jaileles: 0.000972 Results Biosamples E 2 2 2 2 2 2 2 2 2 2 2 2 2	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples 2217 7037 9483 osamples Map Variants	Cregion & Visual Response & Visual Query Matches 135 135	Subset Match Frequencies 0.061 0.019 0.014	n 1 ⁷ 1 olin , , , , , , , , , , , , , , , , , , ,	19 20 21 22 20 21 20 21	Material: n Origin: Hei External Refe PMID:230 CNV Profile(pre:: AMOE	neoplastic sam eidelberg, Germ ferences 079654 (s) CSC Ge move < chr9:21,531 crr9 (p21.5) scale	any any any any any any any any	Ser on Hu >> >>> zoor 061,586 bp. en	Iman Dec. 201 n in 1.5x 3x 10x ter position, gene symbol, Hi 2012 13 200 to ar (art of ant or) 200 to ar (art of ant or)	13 (GRC) base zoom GVS or search ter	h38/hg38 out 1.5x 3x ms 22001 3222	B) Asse 10x 10 22 24 24 24 24 24 24 24 24 24	aviii (1111) embly or) or ,,
progenetix Samples: 135 Variants: 120 Calls: 135 faileles: 0.000972 Results Biosamples E 2 2 2 2 2 2 2 2 2 2 2 2 2	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples 2217 7037 9483 osamples Map Variants	Cregion & Visual Response & Visual Query Matches 135 135	Subset Match Frequencies		19 20 21 22 2011 propensitik.org	Material: n Origin: Hei External Refe PMID:230 CNV Profile(pre: AMOR	neoplastic sam eidelberg, Germ ferences 079654 (s) CSCC Ge move < chr9:21,531 crr9 (s21.9) scate	any any any any any any any any	Ser on Hu >> >>> zoor 061,586 bp. en	Iman Dec. 201 min 1.5x 3x 10x Iter position, gene symbol, H4	13 (GRC) base zoom GVS or search ter	h38/hg38 out 1.5x 3x ms 2001 9922	B) Asse 10x 10 Dee, eeel 22	a a a a a a a a a a a a a a a a a a a
progenetix Samples: 135 Variants: 120 Calls: 135 Jaileles: 0.000972 Results Biosamples E 1 2 2 2 2 2 2 2 2 2 2 2 2 2	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples 2217 7037 9483 osamples Map Variants Description	Cregion & Visual Response & Visual Query Matches 135 135 135	Subset Match Frequencies 0.061 0.019 0.014	DEL	19 20 21 22 10 10 10 20 20 21 22 20 20 21 22 20 20 21 22 20 20 21 22 20 20 21 22 20 20 21 22 20 20 21 22 20 DUP CNV 20 20	Material: n Origin: Hei External Refe PMID:230 CNV Profile(: per: AM201	neoplastic sam eidelberg, Germ ferences 079654 (s) CSCCGE move < chr9:21,531 crr9 (s21.9) scale	ale any enome Brow exect (1),004 enome Brow (< << < > ,306-22,492,891 (22,7**,*** 1 21,7**,*** 1 21,7**,***	Ser on Hu >> >>> zoor 201,586 bp. en 21,0**,*** 21,9	Iman Dec. 201 m in 1.5x 3x 10x Iter position, gene symbol, Hi ee, eeel 22, eee, eeel 2 Det. Ver Hant Batton	13 (GRC) base zoom GVS or search ter	h38/hg38 out 1.5x 3x ms DIRE 3955 00 ,200, 000 22, 3	B) Asse 10x 10 20	ambly
progenetix Samples: 135 Variants: 120 Calls: 135 faileles: 0.000972 Results Biosamples E Calls: 135 faileles: 0.000972 Results Biosamples E Subsets Icdom-80463 NCIT:C2926 Icdot-C34.9 Results Biosamples Bio JSON Download Respon Id PGX_AM_BS_20164920_H23	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples 2217 7037 9483 osamples Map Variants Description non-small cell lung carcinoma [cell line H23]	Cregion & Visual Response & Visual Query Matches 135 135 135 135 135 135	Subset Match Frequencies 0.061 0.019 0.014	DEL 0.072	19 20 21 22 10 10 20 20 20 10 10 10 20 20 20271 propenetik.org 20 20 20 DUP CNV 0.129 0.2022	Material: n Origin: Hei External Refe PMID:230 CNV Profile(: per: AMORE per: AMORE COMPOSITION COMPOSITION COMPOSITION COMPOSITION COMPOSITION	neoplastic sam eidelberg, Germ ferences 079654 (s) 2 CSCCGE move < chr9:21,531 crr9 (s21.3) scale	ale any Ender: 11,004 Enome Brow (< < < > ,306-22,492,891 S 2220 E 1 21,7**,**el 2 1 21,7***,**el 2 1 21,7***,**e	Ser on Hu >> >>> zoor 161,586 bp. en 11,6**,*** 21,9	Iman Dec. 201 m in 1.5x 3x 10x Iter position, gene symbol, Hi ee, seel 22, see, seel 2 Cott. Vier Levit autor	13 (GRC base zoom GVS or search ter	h38/hg38 out 1.5x 3x ms 200,200,0001 22,3	15 16 10x 10 20 10	ambly
progenetix Samples: 135 Variants: 120 Calls: 135 faileles: 0.000972 Results Biosamples E 1 7 7 7 7 7 7 7 7 7 7 7 7 7	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples 2217 7037 9483 osamples Map Variants Description non-small cell lung carcinoma [cell line H23]	Cregion & Visual Response & Visual Query Matches 135 135 135 135 135	Subset Match Frequencies 0.061 0.019 0.014	DEL 0.072	19 20 21 22 10 10 10 35 10 10 10 35 2021 properation org 35 DUP CNV 0.2022	Material: n Origin: Hei External Refe PMID:230 CNV Profile(: oper: AMOR	neoplastic sam eidelberg, Germ ferences 079654 (s) CSCC Ge move < chr9:21,53: chr9:(p21,3) chr9:	ale any Entropy of the second secon	Ser on Hu >> >>> Zoor 2015 21,586 bp. en 21,9**,*** 21,9 11,6**,*** 21,9	Iman Dec. 201 m in 1.5x 3x 10x iter position, gene symbol, H0 ee, eeg 22, eee, eeel 2 oot. vier land and and oot.	13 (GRC) base zoom GVS or search ter	h38/hg38 out 1.5x 3x ms	B) Asse 10x 100	embly
progenetix Samples: 135 Variants: 120 Calls: 135 faileles: 0.000972 Results Biosamples E 1 2 7 7 7 7 7 7 7 7 7 7 7 7 7	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples 2217 7037 9483 osamples Map Variants Description non-small cell lung carcinoma (cell line H23)	Cregion & Visual Response & Visual Query Matches 135 135 135 135 135 135 135 135 135	Ization options	DEL 0.072	10 20 21 22 10 10 10 25 10 10 10 25 20271 properents.org 35 36 DUP CNV 0.2022 0.2022 0.0063	Material: n Origin: Hei External Refe PMID:230 CNV Profile (r Over More CNV Profile (r Over More E UC Over More State Dispect	neoplastic sam eidelberg, Germ ferences 079654 (s) CSC Ge move < chr9:21,53* chr9:(e21.3) scale chr9:21,53* chr9:(e21.3) scale chr9:21,53* chr9:(e21.3) scale chr9:21,53* chr9:(e21.3) scale chr9:21,53* chr9:(e21.3) scale chr9:chr9:(e21.3) scale chr9:chr9:(e21.3) scale chr9:chr9:(e21.3) scale chr9:chr9:(e21.3) scale chr9:chr9:(e21.3) scale chr9:chr9:(e21.3) scale chr9:(e21.3) scale chr9:chr9:(e21.3) scale chr9:chr9:(e21.3) scale chr9:chr9:(e21.3) scale chr9:chr9:(e21.3) scale chr9:(e21	ale any any any any any any any any	Ser on Hu >> >>> zoor 21001 • • • • • • • 21,9• 11586 bp. en	Iman Dec. 201 min 1.5x 3x 10x ter position, gene symbol, Ho 2012 13 200 100 100 100 2010 100 100 100 2010 100 100 100 2010 100 100 100	13 (GRCI	12 13 14 h38/hg38 out 1.5x 3x ms 22,3 00 ,22+, eeel 22,3	B) Asse 10x 10 10x 10 10e, eee] 22	embly 0x 900 1, 400, 000
progenetix Samples: 135 Variants: 120 Calls: 135 faileles: 0.000972 Results Biosamples E 1 2 7 7 7 7 7 7 7 7 7 7 7 7 7	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples Subset Samples 2217 7037 9483 osamples Map Variants Description non-small cell lung carcinoma [cell line H23]	Cregion & Visual Response & Visual Query Matches 135 135 135 135 135 135 135 135	Ization options	DEL 0.072 0.01	20 21 22 20 21 22 20 21 23 20 21 23 20 21 23 20 21 23 20 21 23 20 21 23 20 21 23 20 21 23 20 21 23 20 21 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 21 23 23 23 23 23 24 24 24 25 24 24 25 24 24 25 24 24 25 24 24 26 24 24	Material: n Origin: Hei External Refe PMID:230 CNV Profile(: Over Profile(: Results Bio JSON V Dc Digest	neoplastic sam eidelberg, Germ ferences 079654 (s) CSCC Ge move < chr9:21,53* chr9:(e21.3) chr9:	ale any **** *****************************	Ser on Hu >> >>> zoor 2001	Iman Dec. 201 min 1.5x 3x 10x ter position, gene symbol, Ho 2012 10 2012 10 2012 10 2012 10 2012 10 2014 10 20	13 (GRCI	12 13 14 h38/hg38 out 1.5x 3x ms 52252 32252 50 52252 </td <td>B) Asse 10x 10 10x 10 10x 22 10x 22 10x 10 10x 10x 10x 10 10x 10x 10x 10 10x 10x 10x 10x 10x 10x 10x 10x 10x 10x</td> <td>ambly am</td>	B) Asse 10x 10 10x 10 10x 22 10x 22 10x 10 10x 10x 10x 10 10x 10x 10x 10 10x 10x 10x 10x 10x 10x 10x 10x 10x 10x	ambly am
progenetix Samples: 135 Variants: 120 Calls: 135 faileles: 0.000972 Results Biosamples E 1 2 7 7 7 7 7 7 7 7 7 7 7 7 7	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples Subset Samples 2217 7037 9483 osamples Map Variants Description non-small cell lung carcinoma [cell line H23] non-small cell lung carcinoma	Cregion & Visual Response & Visual Cuery Matches 135 135 135 135 135 135 135 135	Identifiers PMID:20164920 geogse-GSE19399 PMID:20164920	DEL 0.072 0.01	20 21 22 20 21 22 20 23 23 2027 proprestik.org 2027 proprestik.org 0.129 0.2022 0.0553 0.0663	 Material: n Origin: Hei External Refe PMID:230 CNV Profile (s per: 2000 CNV Profile (s per: 2000 CNV Profile (s Disper: 2000 CNV Profile (s <	neoplastic sam eidelberg, Germ ferences 079654 (s) CSCC Ge move < chr9:21,53 chr9:(e21.3) chr9:(ale any mome Brow cc << < > mome Brow cc << < > 1 21,700,000 1 21,700,0000 1 21,700,000 1 21,700,0000 1 21,700,000 1 2	Ser on Hu >> >>> zoor 061,586 bp. en 21,0**,*** 21,9* 115	Iman Dec. 201 min 1.5x 3x 10x ter position, gene symbol, Ho 2012 80 2012 80 2014 80 20	13 (GRCI base zoom GVS or search ter	12 12 13 14 h38/hg38 out 1.5x 3x ms 5255 3255 30 5255 3255 <	B) Asse 10x 10 10x 10 10x 22 10x 22 10x 10 10x 10x 10x 10 10x 10x 10x 10 10x 10x 10x 10x 10x 10x 10x 10x 10x 10x	Alt. Base(s)
progenetix Samples: 135 Variants: 120 Calls: 135 faileles: 0.000972 Results Biosamples E 1 2 2 7 7 7 7 7 7 7 7 7 7 7 7 7	Phenopackets & UCSC Calisets Variants & UCSC Variants in UCSC & UCSC Biosamples Map Variants Subset Samples 2217 7037 9483 osamples Map Variants Description non-small cell lung carcinoma [cell line H23] non-small cell lung carcinoma	Cregion & Visual Response & Visual Cuery Matches 135 135 135 135 135 135 135 135	Identifiers PMID:20164920 PMID:201649 PMID:201649 PMID:201649 PMID:201649 PMID:201649 PMID:201649 PMID:201649 PMID:201649 PMID:201649 PMID:201	DEL 0.072 0.01	20 21 22 20 21 22 20 21 23 20 21 23 20 21 23 20 21 23 20 21 23 20 23 21 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23 20 23 23	 Material: n Origin: Hei External Refe PMID:230 CNV Profile (per: 2005 CNV Profile (Digest 9:21548871- 219995:DEL 	neoplastic sam eidelberg, Germ ferences 079654 (s) CSCC Ge move < chr9:21,531 chr9:21,531 chr9:(e21.3) Scale over9:(e21.3) scale chr9:21,531 chr9:(e21.3) scale chr9:21,531 chr9:(e21.3) ch	ale any nome Brow c < < > 306-22,492,891 1 21,700,000 1 21,700,0000 1 21,700	Ser on Hu >> >>> zoor 2001 21,000,000 21,90 21,000,000 21,90 11,000,000 21,90 3079654-pfi-	Iman Dec. 201 min 1.5x 3x 10x ter position, gene symbol, Hi 2012 33 200 22, 400, e001 2 2010 2010 2010 2 2010 2010 2010 2 2010 2010	13 (GRCI base zoom GVS or search ter transformed and the transformed and the transform	12 12 14 h 38/hg38 out 1.5x 3x ms 3325 3325 22,3 3325 3325 30 22,3 3325 30 22,3 3325 30 22,3 3325 30 22,3 3325 30 22,3 3325 30 22,3 3325 30 24,0 3425 30 24,0 3425 30 3445 3425 30 3445 3425 30 3445 3425 30 3445 3425 30 3445 3445 30 3445 3445 30 3445 3445 31 3445 3445 32 3445 3445 34 3445 3445 34 3445 3445 34 3445 3445 34 3445 3445 34 3445 3445	B) Asse 10x 10 10x 10 10x 22 1000, 000 22 Ref. Base(s)	Alt. Base(s)

Total

1939

138 663 115 357 23 306

> Figure 3. Beacon-style guery using fuzzy ranges to identify biosamples with variants matching the CNA range This example gueries for a continuous, focal duplication covering the complete MYC gene's coding region with < = 6 Mb in size. A: Filter for dataset; B: filter for cancer classification (NCIt and ICD-O-3 ontology terms available); C: additional filter, e.g. Cellosaurus; D: additional filter for geographic location; E: external link to UCSC browser to view the alignment of matched variants; F: cancer type classification sorted by frequency of the matched biosamples present in the subset; G: list of matched biosamples with description, statistics and reference. More detailed biosample information can be viewed through 'id' link to the sample detail page; H: matched variants with reference to biosamples can be downloaded in json or csv format.



cancercellines.org

Cancer Cell Line Genomics Resource

- cancercelllines.org built on Progenetix platform
- includes over 5600 cell line CNV profiles
- cancer cell line variants ()SNV, INDELs ...) for
 16178 cell lines from 400 different disease classifications
- mapped to to Cellosaurus
- hierarchical representation ("derived from"...)
- SNVs mapped from ClinVar with variant severity and disease ontologies
- CCLE per cell line include variant effect
- CNV profiles allow temporal stability estimates and tumor type similarity matching





cancercelllines

Cancer Cell Lines⁰

Cell Line Listing

Search Cell Lines

CNV Profiles by

Cancer Type

NCIT Codes ICD-O 3 Morphologies

Documentation

Progenetix

Progenetix Data

Progenetix Documentation

Baudisgroup @ UZH

Cancer Cell Line Genomics



The *cancercelllines.org* genomic information resource contains genome profiling data, somatic mutation information and associated metadata for thousands of human cancer cell lines. It has its origins in genomic copy number variation (CNV) profiling data of cell lines originally collected as part of the more than 100'000 individual datasets in the Progenetix *O* oncogenomic resource. However, by providing genome mapped, annotated data for many types of genomic mutations, together with CNV profiles for a subset of the overall more than 16'000 cell lines, *cancercelllines.org* provides a unique entry point for the comparative analysis of genomic variants in cell lines as well as for the exploration of related publications.

SK-MEL-1 (cellosaurus:CVCL_0068)



© CC-BY 2001 - 2023 progenetix.org

Cell Line Data CNV Frequency Plot The CNV histogram above represents CNV data from a randomly selected set of samples - either instances of a common cell line or with a shared diagnosis. In this example the frequencies of regional gains and losses in 5 samples from cellosaurus:CVCL_0068 (SK-MEL-1) are on display.

Download SVG | Go to cellosaurus:CVCL_0068 | Download CNV Frequencies

In *cancercelllines.org* genomic variation data collected from a variaty of external resources and from original data (re-) analyses has been mapped to GRCh38 genome coordinates and is queryable using the Beacon v2 API 2. The resource contains data of **16340** individual cancer cell lines from **382** different cancer types (NCIt neoplasm classification).

A large amount of the cancer cell line data has been collected based on annotations and pointers from Cellosaurus 🗹 , a reference knowledge resource on cell lines.

Citation

- · cancercelllines.org: Cancer cell line oncogenomic online resource (2023)
- Huang Q, Carrio-Cordo P, Gao B, Paloots R, Baudis M. (2021): The Progenetix oncogenomic resource in 2021. Database (Oxford). 2021 Jul 17



Maintaining some Standards CNV Term Use in Computational (File/Schema) Formats

tandards/ S i on -.org/resources/CNV-annotat ar > //cn • • https

				1	
EFO	Beacon	VCF	SO	GA4GH VRS ⇒ VRS proposal ¹	Notes
EF0:0030070 copy number gain	DUP ² or EF0:0030070	DUP SVCLAIM=D ³	S0:0001742 copy_number_gain	low-level gain (implicit) ⇒ EF0:0030070 copy number gain	a sequence alteration whereby the copy number of a given genomic region is greater than the reference sequence
EF0:0030071 low-level copy number gain	DUP ² or EF0:0030071	DUP SVCLAIM=D ³	S0:0001742 copy_number_gain	low-level gain ⇒ EF0:0030071 low-level copy number gain	
EF0:0030072 high-level copy number gain	DUP ² or EF0:0030072	DUP SVCLAIM=D ³	S0:0001742 copy_number_gain	high-level gain ⇒ EF0:0030072 high-level copy number gain	commonly but not consistently used for >=5 copies on a bi-allelic genome region
EF0:0030073 focal genome amplification	EF0:0030073	DOF SVCLAIM=D ³	Sociology copy_number_gain	high-level gain ⇒ EF0:0030073 focal genome amplification	commonly but not consistently used for >=5 copies on a bi-allelic genome region, of limited size (operationally max. 1-5Mb)
EF0:0030067 copy number loss	DEL ² or EF0:0030067	DEL SVCLAIM=D ³	S0:0001743 copy_number_loss	partial loss (implicit) ⇒ EF0:0030067 copy number loss	a sequence alteration whereby the copy number of a given genomic region is smaller than the reference sequence
EF0:0030068 low-level copy number loss	DEL ² or EF0:0030068	DEL SVCLAIM=D ³	S0:0001743 copy_number_loss	partial loss \Rightarrow EF0:0030068 low-level copy number loss	
EF0:0020073 high-level copy number loss	DEL ² or EF0:0020073	DEL SVCLAIM=D ³	S0:0001743 copy_number_loss	partial loss \Rightarrow EF0:0020073 high-level copy number loss	a loss of several copies; also used in cases where a complete genomic deletion cannot be asserted
EF0:0030069 complete genomic deletion	DEL ² or EF0:0030069	DEL SVCLAIM=D ³	S0:0001743 copy_number_loss	<pre>complete loss ⇒ EF0:0030069 complete genomic deletion</pre>	complete genomic deletion (e.g. homozygous deletion on a bi- allelic genome region)



Hangjia Zhao Michael Baudis (& the VRS group!)





Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

GA4GH Standards for Federated Genomic Data Discovery



Different Approaches to Data Sharing









Centralized Genomic Knowledge Bases

Data Commons

Trusted, controlled repository of multiple datasets



Hub and Spoke

Common data elements, access, and usage rules



Linkage of distributed and disparate datasets





Cell Genomics



Technology

The GA4GH Variation Representation Specification: A computational framework for variation representation and federated identification

Alex H. Wagner,^{1,2,25,*} Lawrence Babb,^{3,*} Gil Alterovitz,^{4,5} Michael Baudis,⁶ Matthew Brush,⁷ Daniel L. Cameron,^{8,9} Melissa Cline,¹⁰ Malachi Griffith,¹¹ Obi L. Griffith,¹¹ Sarah E. Hunt,¹² David Kreda,¹³ Jennifer M. Lee,¹⁴ Stephanie Li,¹⁵ Javier Lopez,¹⁶ Eric Moyer,¹⁷ Tristan Nelson,¹⁸ Ronak Y. Patel,¹⁹ Kevin Riehle,¹⁹ Peter N. Robinson,²⁰ Shawn Rynearson,²¹ Helen Schullenburg,¹² Kirill Tsukanov,¹² Brian Walsh,⁷ Melissa Konopko,¹⁵ Heidi L. Rehm,^{3,22} Andrew D. Yates,¹² Robert R. Freimuth,²³ and Reece K. Hart^{3,24,*}





International federation of genomic medicine databases using GA4GH standards

Adrian Thorogood,^{1,2,*} Heidi L. Rehm,^{3,4} Peter Goodhand,^{5,6} Angela J.H. Page,^{4,5} Yann Joly,² Michael Baudis,⁷ Jordi Rambla,^{8,9} Arcadi Navarro,^{8,10,11,12} Tommi H. Nyronen,^{13,14} Mikael Linden,^{13,14} Edward S. Dove,¹⁵ Marc Fiume,¹⁶



Overview of GA4GH standards and frameworks







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.





2014 GA4GH founding event; Jim Ostell proposes Beacon concept with "more features... version 2"

2015	 beacon-network.org aggregator created by DNAstack 	
	 Beacon v0.3 release 	B
2016	 work on queries for structural variants (brackets for fuzzy start and end) 	p c
	parameters)	e e
2017	 OpenAPI implementation integrating CNV parameters (e.g. "startMin, statMax") 	• B
2018	 Beacon v0.4 release in January; feature release for GA4GH approval process GA4GH Beacon v1 approved at Oct plenary 	• B
2019	 ELIXIR Beacon Network 	a • ",
2020		fi • d • fr

2022

2021

Beacon v2 Development

- Beacon⁺ concept implemented on
- rogenetix.org
- oncepts from GA4GH Metadata (ontologies...)
- ntity-scoped query parameters
- 'individual.age")
- Beacon+ demos "handover" concept
- Beacon hackathon Stockholm; settling or "filters" Barcelona goes Zurich developers meeting eacon API v2 Kick off
- dopting "handover" concept
- Scouts" teams working on different aspects -Iters, genomic variants, compliance ... iscussions w/ clinical stakeholders
- ramework + models concept implemented range and bracket queries, variant length starting of GA4GH review process
- further changes esp. in default model, aligning with Phenopackets and VRS unified beacon-v2 code & docs repository Beacon v2 approved at Apr GA4GH Connect

Related ...

• ELIXIR starts Beacon project support

- GA4GH re-structuring (workstreams...)
- Beacon part of Discovery WS
- new Beacon website (March)
- Beacon publication at Nature Biotechnology

docs.genomebeacons.org









docs.genomebeacons.org



Progenetix & Beacon

Implementation driven standards development

- Progenetix Beacon+ has served as implementation driver since 2016
- prototyping of advanced Beacon features such as
 - → structural variant queries
 - → data handovers
 - Phenopackets integration









	Beacon v2 GA4	GH Approval Registry	
	Beacons: European GENOME-PHENOME ARCHIVE	enet x- Cnag 🐺 📖	IVERSITY OF ICESTER
EUROPEAN GENOME-PHENOME ARCHIVE Visit us Contact us	European Genome-Phenome Archive (EGA) GA4GH Approval Beacon Test This <u>Beacon</u> is based on the GA4GH Beacon <u>v2.0</u>	progenet x X Wisit us Beacon UI Beacon API Contact us	Theoretical Cytogenetics and Oncogenomics group at UZH and SIB Progenetix Cancer Genomics Beacon+ Beacon+ provides a forward looking implementation of the Beacon v2 API, with focus on structural genome variants and metadata based on the
BeaconMap Bioinformatics analysis Biological Sample Cohort Configuration Dataset EntryTypes Genomic Variants Individual Info Sequencing run		BeaconMap Bioinformatics analysis Biological Sample Cohort Configuration Dataset EntryTypes Genomic Variants Individual Info Sequencing run	
CINCUS Wisit us Beacon API Contact us	Centre Nacional Analisis Genomica (CNAG-CRG) Beacon @ RD-Connect This <u>Beacon</u> is based on the GA4GH Beacon <u>v2.0</u>	UNIVERSITY OF LEICESTER E Beacon UI E Beacon API Contact us	University of Leicester Cafe Variome Beacon v2 This <u>Beacon</u> is based on the GA4GH Beacon <u>v2.0</u>
BeaconMap Bioinformatics analysis Biological Sample Cohort Configuration Dataset EntryTypes Genomic Variants Individual Info Sequencing run		BeaconMap Bioinformatics analysis Biological Sample Cohort Configuration Dataset EntryTypes Genomic Variants Individual Info Sequencing run	







Can you provide data about focal deletions in CDKN2A in Glioblastomas from juvenile patients with unrestricted access?





Beacon v2 API

The Beacon API v2 represents a simple but powerful genomics API for *federated* data discovery and retrieval



The GA4GH Phenopackets v2 Standard A Computable Representation of Clinical Data

The GA4GH Phenopacket schema consists of several optional elements, each containing information about a certain topic, such as phenotype, variant or pedigree. An element can contain other elements, which allows a hierarchical representation of data.

For instance, Phenopacket contains elements of type *Individual, PhenotypicFeature, Biosample* and so on. Individual elements can therefore be regarded as **building blocks** of larger structures.

Jacobsen JOB, Baudis M, Baynam GS, Beckmann JS, Beltran S, Buske OJ, Callahan TJ, *et al.* 2022.

The GA4GH Phenopacket Schema Defines a Computable Representation of Clinical Data. *Nature Biotechnology* 40 (6): 817–20.







The GA4GH Phenopackets v2 Standard **A Computable Representation of Clinical Data**

The GA4GH Phenopacket schema

consists c elements informatio such as p pedigree. other elen hierarchic For instan contains e Individual Biosample elements regarded larger stru

Jacobsen Beckmann . TJ, et al. 202 The GA4GH



Family

$\leftarrow \mathsf{ICS} \leftarrow \mathsf{35} \leftarrow \mathsf{35.240} \leftarrow \mathsf{35.240.80}$

ISO 4454:2022 Genomics informatics — Phenopackets: data exchange

Abstract

🖅 Preview

This document specifies a uniform, machine-readable, phenotypic description of an individual, patient or sample in the context of rare disease, common/complex disease or cancer.

It is applicable to academic, clinical and commercial research, as well as clinical diagnostics. While intended for human data collection, it can be used in other areas (e.g. mouse research). It does not define the phenotypic information that needs to be collected for a particular use but represents that information in an appropriately descriptive manner that allows it to be computationally exchanged between systems.

General information [™]

Status : 🕑 Published

Publication date : 2022-07

Edition : 1

Number of pages : 86

Technical Committee : ISO/TC 215/SC 1 Genomics Informatics Computable Representation of Clinical Data

Nature Biotechnology 40 (6): 817-20.



		ACMG, American College Genetics	
	Search		 HTS, high-throughput scr VCF, variant call format. Metadata Medical action
: A fc	ormat for Buy th	phenotypic his standard	Co-hosted by Smart Health Standards Forum, IS
	Format	Language	TC215 SC1 Korear
	✓ PDF + ePub	English 🗸	mirror committee
	Paper	English 🗸	
	снғ 19	98 □ Buy	Age Interval High-level the Person / far Disease / in Genomic in Biosample
rec	ord context	Variation	Medical act Phenotypic Multi-purpo











Theoretical Cytogenetics and **Oncogenomics @baudisgroup**

- curated resources, patterns & markers in cancer genomics, especially somatic structural genome variants
- bioinformatics in **collaborative studies**
- bioinformatics **tools** & methods
- standards and implementations for data sharing in genomics, personalized health
- open research data "ambassadoring"















Jordi Rambla Arcadi Navarro **Roberto** Ariosa Manuel Rueda Lauren Fromont Mauricio Moldes Claudia Vasallo Babita Singh Sabela de la Torre Marta Ferri Fred Haziza



Juha Törnroos Teemu Kataja Ilkka Lappalainen **Dylan Spalding**





Tony Brookes Tim Beck Colin Veal Tom Shorter Personalized SPHN

University of Zurich

Michael Baudis Rahel Paloots Hangjia Zhao Ziying Yang Bo Gao



Augusto Rendon Ignacio Medina Javier López Jacobo Coll Antonio Rueda

The Beacon team through the ages





Fundación Progreso y Salud CONSEJERÍA DE SALUD



Marc Fiume Miro Cupak







GA4GH Phenopackets Peter Robinson Jules Jacobsen



GA4GH VRS Alex Wagner Reece Hart

Beacon PRC

Alex Wagner Jonathan Dursi Mamana Mbiyavanga

Alice Mann Neerjah Skantharajah



docs.genomebeacons.org





Universität Zürich^{UZH}



Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.



eliír

Swiss Institute of Bioinformatics

SPHN



