UCSF Bakar Computational Health Sciences Institute

UCSF Helen Diller Family Comprehensive Cancer Center

UCSF Cancer Commons

NIC Containers and Workflows Interest Group

June 10th, 2022

Gundolf Schenk, PhD

Principal Data Scientist

June 2022

Collaboration:

Alan Ashworth

Alejandro Sweet-Cordero

Brinda Venkatesh

Michelle Turski

Yongmei Shi

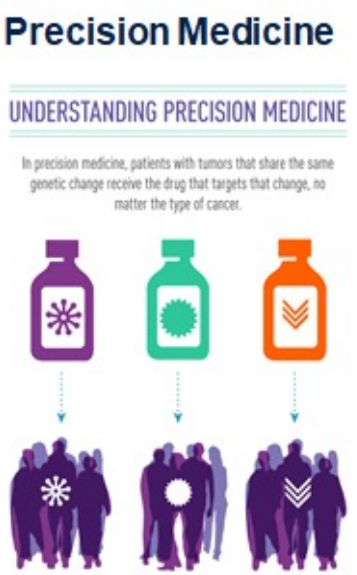
Carlos Espinosa Mendez

Outline

- Deid dataflow cancer genomics
- How it comes together: cancergenomics + clinical + notes/reports + radimages + cancerregistry + pathology
- cBioPortal
- Path concepts

CANCER GENOMICS DATA - SIGNIFICANCE

- Goal: Provide internal access to de-identified genomics panel testing data linked to other de-identified data assets
- Precision medicine is critical in cancer and an essential component of institutional goals
- Puts UCSF at the forefront



Molecular profiling for precision cancer therapies

Eoghan R. Malone, Marc Oliva, Peter J. B. Sabatini, Tracy L. Stockley & Lillian L. Siu

Genome Medicine **12**, Article number: 8 (2020) | Cite this article

An integrated functional and clinical genomics approach reveals genes driving aggressive metastatic prostate cancer

Rajdeep Das, Martin Sjöström, Raunak Shrestha, Christopher Yogodzinski, Emily A. Egusa, I
William S. Chen, Jonathan Chou, Donna K. Dang, Jason T. Swinderman, Alex Ge, Junjie T. H
Kabir, David A. Quigley, Eric J. Small, Alan Ashworth, Felix Y. Feng & Luke A. Gilbert

Nature Communications **12**, Article number: 4601 (2021) | Cite this article

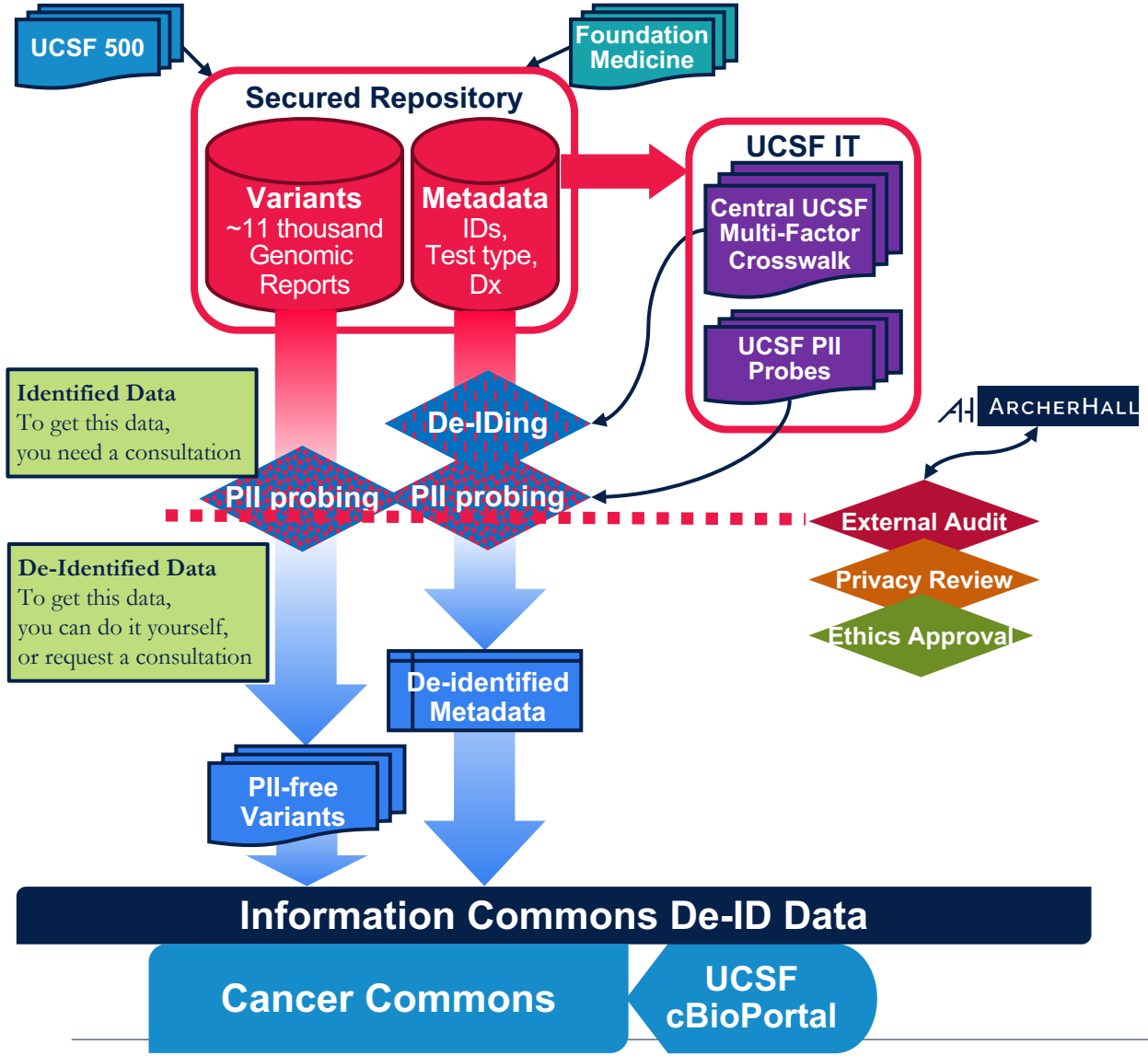
Clinical, radiological and genomic features and targeted therapy in *BRAF* V600E mutant adult glioblastoma

Mary Jane Lim-Fat, Kun Wei Song, J. Bryan Iorgulescu, Brian M. Andersen, Deborah A. Forst, Justin T. Jordan, Elizabeth R. Gerstner, David A. Reardon, Patrick Y. Wen & Isabel Arrillaga-Romany

Journal of Neuro-Oncology **152**, 515–522 (2021) | Cite this article

UCSF

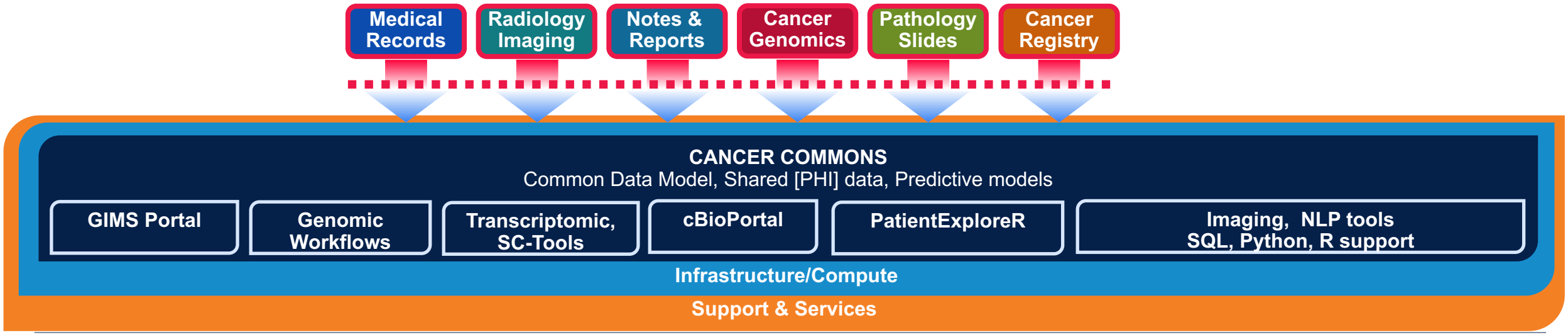
CANCER GENOMICS DATA - DEIDENTIFICATION



- At the present time, we have more than 11,000 Foundation Medicine and UCSF500 reports that have been aggregated into a database
- Metadata has been de-identified using the same protocol that has been approved for Information Commons
  - Data has been de-identified through the application of specific probes that look for personal identifying information (PII)
    - Fields have been removed/surrogated if PII has been found
    - Database is monitored regularly for potential PII
- Genomics data elements are variants retrieved from panel testing and are considered low risk according UC Health external legal counsel

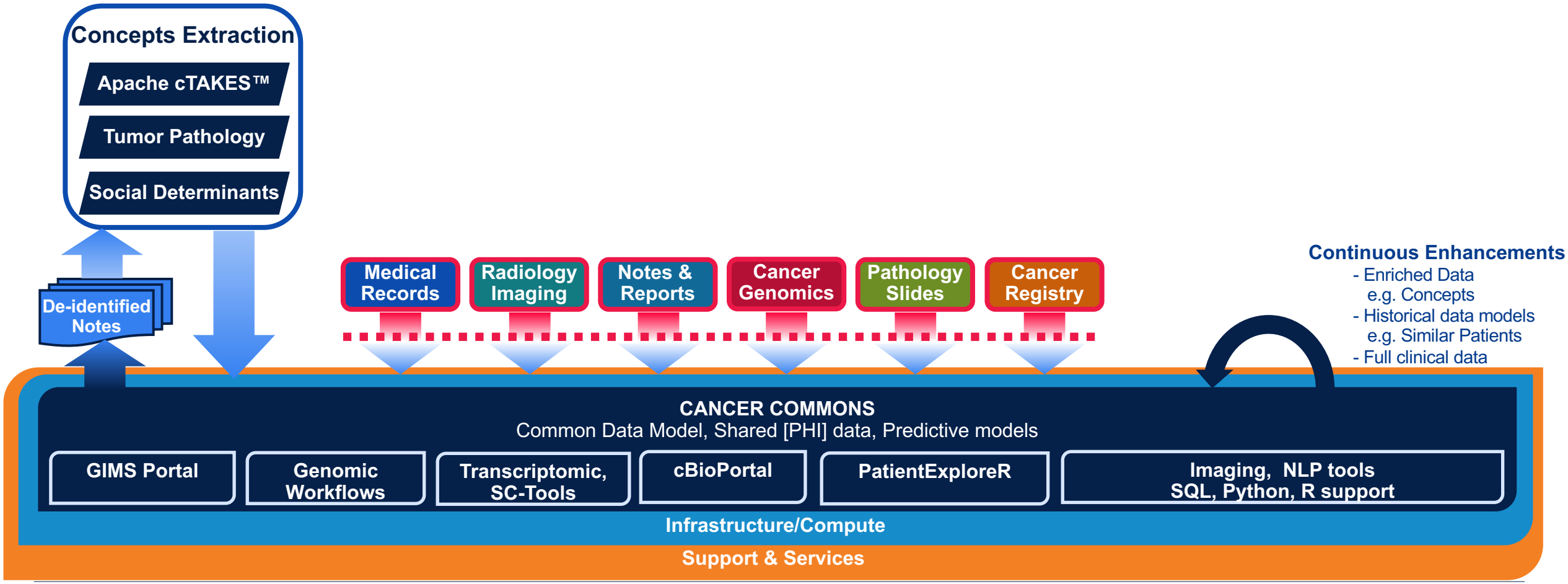
UCSF

Cancer Commons Data Flow



UCSF

Cancer Commons Data Flow



UCSF

Cancer Cor

Ms. ___ also stated that she had just ended a three week course of prednisone four days ago, which she had started about a month ago at 60 mg and tapered herself down over a few days by 10 mg. She began the course of prednisone last month because she felt as if she was about to have a Crohn's flare at the time. Ms. ___ was last hospitalized at UNC for Crohn's disease exacerbation in March 2007. She denies any recent hemoptysis, constipation, hematochezia, melena, and changes in her bowel habits since Wednesday. She has been compliant with taking her medications for Crohn's and has been stable on her mesalamine, mercaptopurine, and omega-3-acid supplement.

41: Fentanyl

Upon arrival to the ED, Ms. ___ was put on IV fluids, given fentanyl 50 mcg IVP†, phenegran 12.5 mg diluted with 10 mL NS IVP†, and Mg sulfate†† IVP†. Radiological images were obtained through an abdominal CT scan†, ultrasound†, and 2V XR. Ms. ___ was not given any other narcotics for her pain because of a past violation of a pain contract after a positive toxicology screen† for cocaine resulted in her discharge from her family medicine provider and due to suspicions that she was narcotic-seeking.

Past Medical History

Crohn's disease, diagnosed 1998
Adenocarcinoma of terminal ileum 1998 - s/p resection of terminal ileum, rad and chemo, no mets.
hx of small bowel obstruction† secondary to Crohn's Disease
DM

text	negated	location	domain	hx	conditional	cui	location snomed id	
Crohn Disease	✗		diseases	✗	✗	C0010346		34
Crohn Disease	✗		diseases	✗	✗	C0010346		34
Hemoptysis	✓		symptoms	✗	✗	C0019079		66
Constipation	✓		symptoms	✗	✗	C0009806		14
Hematochezia	✗		symptoms	✗	✗	C0018932		40
Melena	✓		symptoms	✗	✗	C0025222		29
Defecation	✗	Intestines	symptoms	✗	✗	C0011135	261093006	39
Crohn Disease	✗		diseases	✗	✗	C0010346		34
mesalamine	✗		medications	✗	✗	C0127615		52
mercaptopurine	✗		medications	✗	✗	C0000618		10
Omega-3 Fatty Acids	✗		medications	✗	✗	C0015689		43
Fentanyl	✗		medications	✗	✗	C0015846		37
Intravenous pyelogram	✗		procedures	✗	✗	C0203108		32
Intravenous pyelogram	✗		labs	✗	✗	C0203108		32
Intravenous pyelogram	✗		procedures	✗	✗	C0203108		32

Concepts Extraction

- Apache cTAKES™
- Tumor Pathology
- Social Determinants

De-identified Notes

Medical Records | Radiology Imaging | Notes & Reports | Cancer Genomics | Pathology Slides | Cancer Registry

Continuous Enhancements

- Enriched Data
  e.g. Concepts
- Historical data models
  e.g. Similar Patients
- Full clinical data

CANCER COMMONS

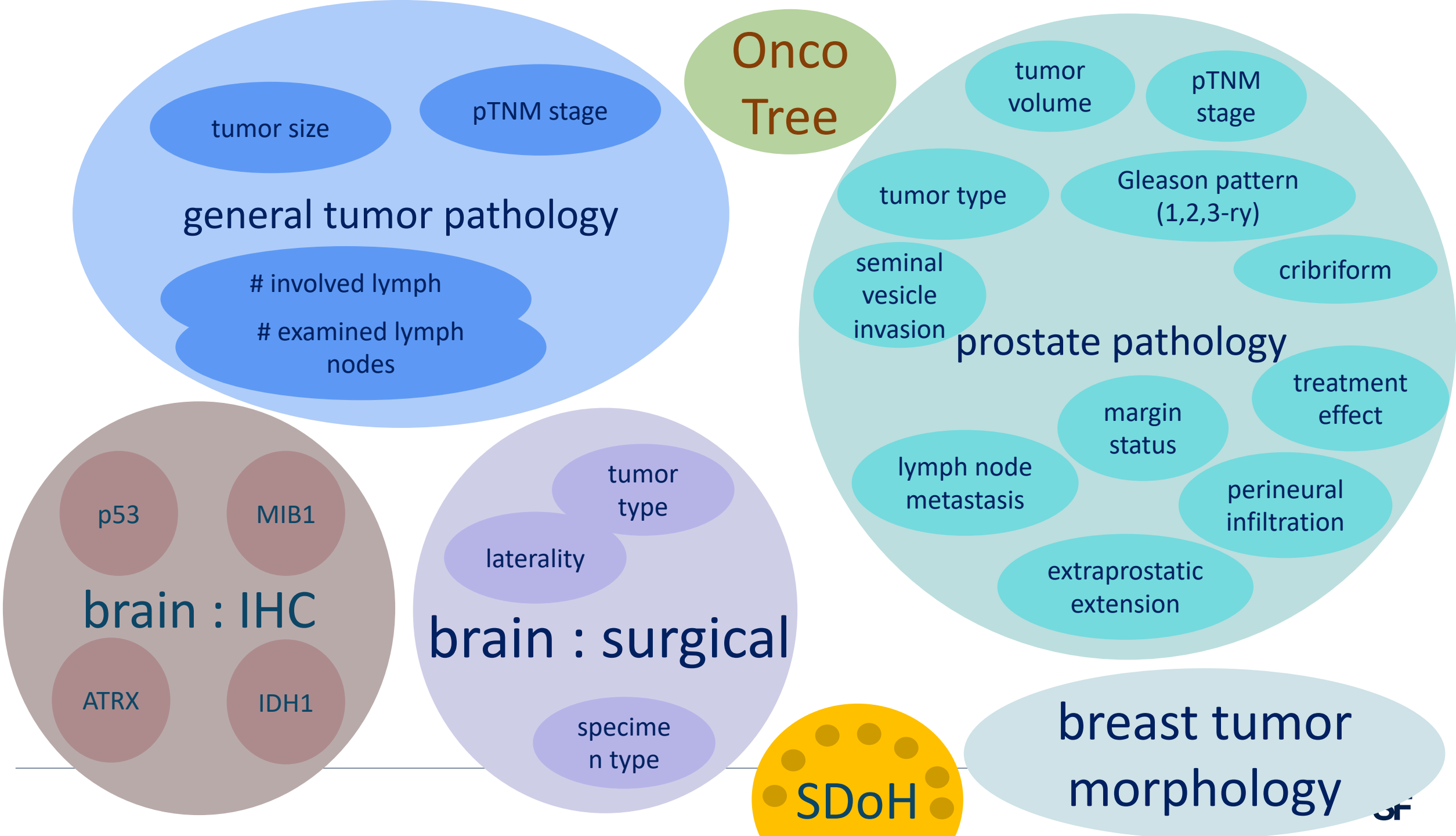
Common Data Model, Shared [PHI] data, Predictive models

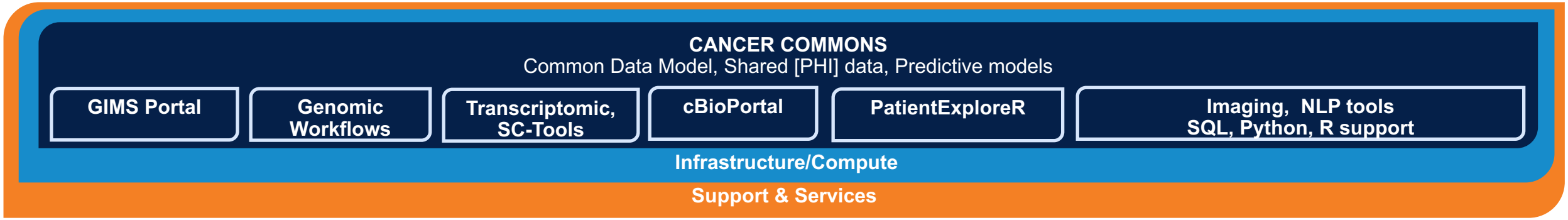
GIMS Portal | Genomic Workflows | Transcriptomic, SC-Tools | cBioPortal | PatientExploreR | Imaging, NLP tools SQL, Python, R support

Infrastructure/Compute

Support & Services

UCSF


Onco Tree
tumor size
pTNM stage
general tumor pathology
involved lymph
examined lymph nodes
tumor volume
pTNM stage
tumor type
Gleason pattern (1,2,3-ry)
seminal vesicle invasion
cribriform
prostate pathology
treatment effect
margin status
lymph node metastasis
perineural infiltration
extraprostatic extension
p53
MIB1
brain : IHC
ATRX
IDH1
tumor type
laterality
brain : surgical
specime n type
SDoH
breast tumor morphology


CANCER COMMONS
Common Data Model, Shared [PHI] data, Predictive models
GIMS Portal
Genomic Workflows
Transcriptomic, SC-Tools
cBioPortal
PatientExploreR
Imaging, NLP tools
SQL, Python, R support
Infrastructure/Compute
Support & Services

cBioPortal FOR CANCER GENOMICS

Data Sets | Tutorials/Webinars | FAQ | News | Visualize Your Data | About

Logged in as gundolf.schenk@ucsf.edu

UCSF

Combined Study

This combined study contains samples from 2 studies

Click gene symbols below or enter here | Query

Summary | Clinical Data | CN Segments

Selected: 10,781 patients | 11,523 samples

Custom Selection | Charts | Groups

Cancer Type

	#	Freq
Glioma	2,343	20.3%
Non-Small Cell Lung Cancer	1,001	8.7%
Other Cancer	752	6.5%
Cancer of Unknown Primary	609	5.3%
Colorectal Cancer	502	4.4%
Prostate Cancer	445	3.9%
Melanoma	420	3.6%
Pancreatic Cancer	366	3.2%
Breast Cancer	357	3.1%
CNS Cancer	339	2.9%
Bladder Cancer	336	2.9%

Search...

Genomic Profile Sample Counts

Molecular Profile	#	Freq
FM Fusions	11,523	100.0%
Mutations	11,523	100.0%
Putative copy-number alterations	11,476	99.6%

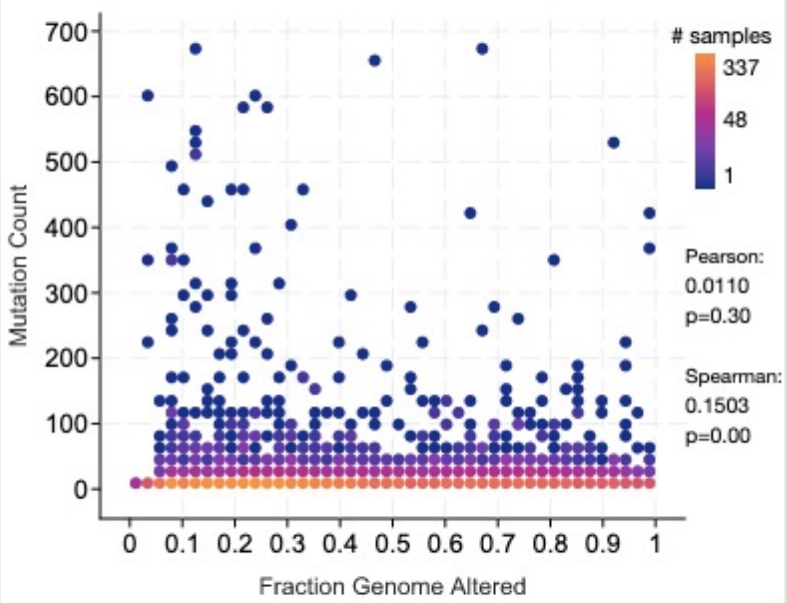
Search...

Mutated Genes (11523 profiled samples)

Gene	# Mut	#	Freq
TP53	4,643	4,033	35.0%
TERT	4,445	3,059	26.8%
KMT2D	2,120	1,543	13.5%
SYNE1	914	551	13.1%

Search...

Mutation Count vs Fraction Genome Altered



Structural Variant Genes (11523 profiled samples)

Gene	# SV	#	Freq
XBP1	9	9	7.1%
TCF3	3	3	2.4%
IKZF3	3	3	2.4%
BRAF	266	265	2.3%

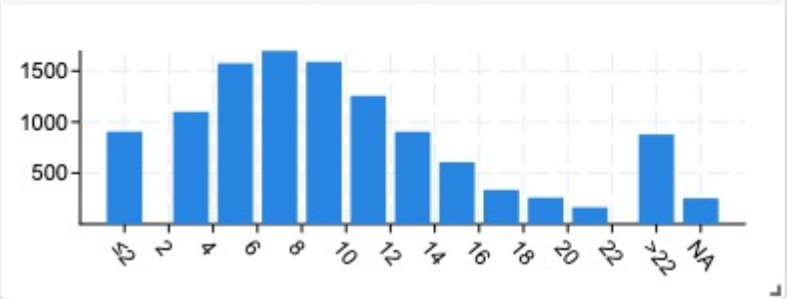
Search...

CNA Genes (11476 profiled samples)

Gene	Cytoband	CNA	#	Freq
CDKN2A	9p21.3	HOMD...	2,241	19.5%
CDKN2B	9p21.3	HOMD...	2,183	19.0%
HIP1	7q11.23	AMP	1,429	15.6%
MYC	8q24.21	AMP	1,778	15.5%

Search...

Mutation Count

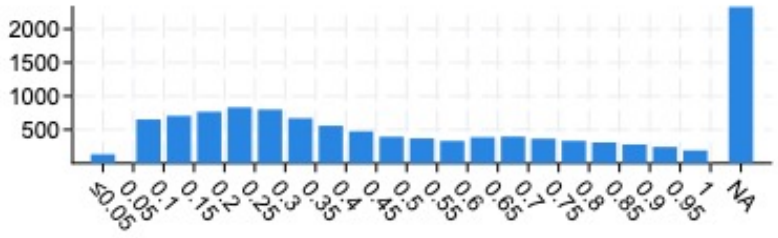


Primary Site

	#	Freq
Brain	1,020	8.9%
Lung	807	7.0%
Liver	748	6.5%
Lymph Node	638	5.5%
Soft Tissue	375	3.3%
Blood	221	1.9%
Colon	209	1.8%
Kidney	204	1.8%
Brain, right frontal	193	1.7%
Brain, posterior fossa	177	1.5%
Brain, left frontal	176	1.5%

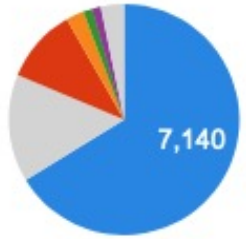
Search...

Fraction Genome Altered

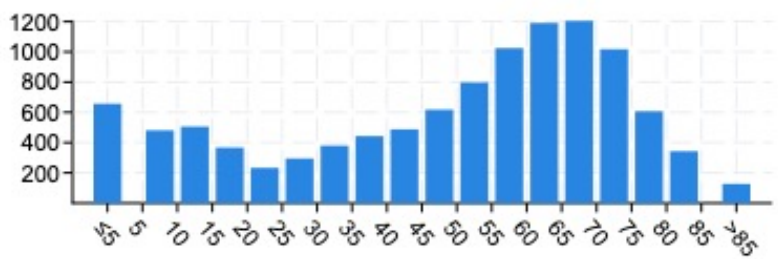


Overall Survival Status

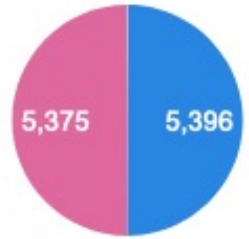
Ethnicity



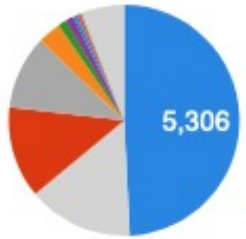
Age (Years)



Sex



Race



UCSF