Genome data access

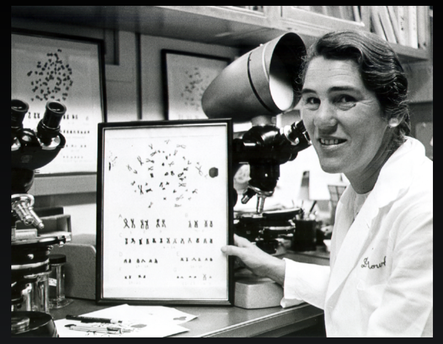
Application to personalized health and cancer research

Michael Baudis
Computational Oncogenomics

SIB

University of Zurich[UZH]

Global Alliance for Genomics & Health



Janet Rowley (1972/73)

Chromosomal translocations in cancer

- Recurrent chromosomal translocations in leukemias and lymphomas
- "Philadelphia chromosome" in CML (Nowell & Hungerford, 1960) represents a reciprocal translocation between chromosomes 9 and 22
- 1972: t(8;21) ALL manuscript rejected by NEJM
- 1973: t(9;22) manuscript rejected by *Nature* "*with some reasonable comments and some truly wrong*"
- Clinical implications: **Tyrosine Kinase inhibitors** as standard first-line therapy in CML
  - first trials in 1998 (STI-571; Imatinib/Gleevec)
  - cf. Druker BJ, Lydon NB (2000). Lessons learned from the development of an Abl tyrosine kinase inhibitor… J Clin Invest 2000;105:3-7)

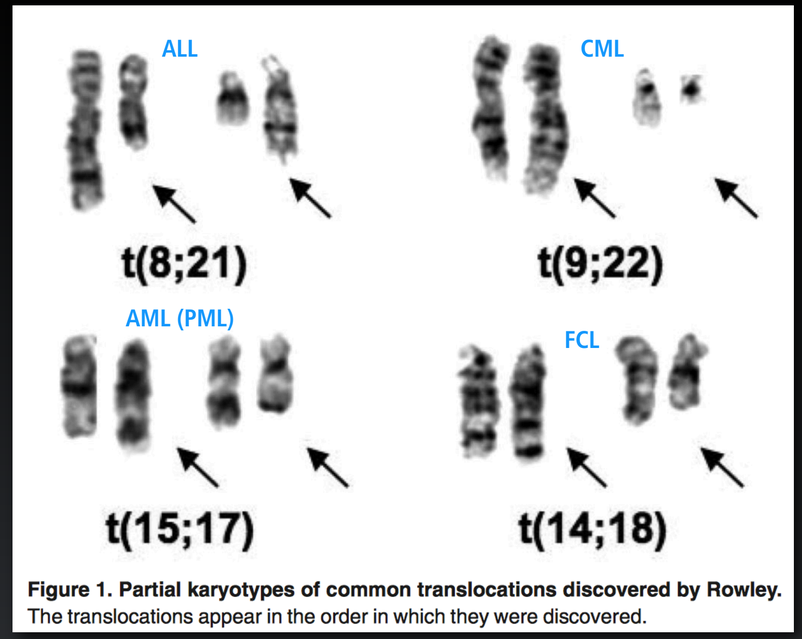
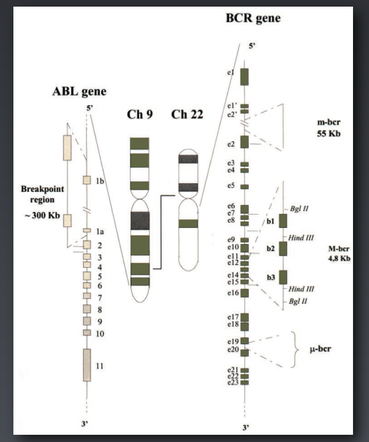
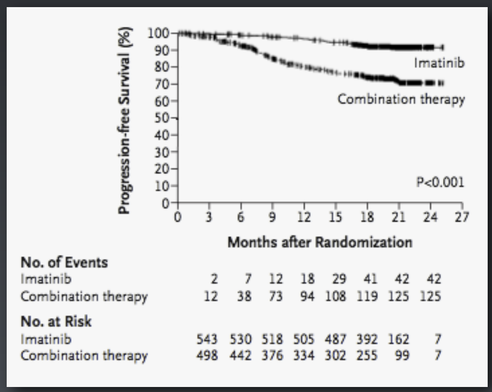


Figure 1. Partial karyotypes of common translocations discovered by Rowley. The translocations appear in the order in which they were discovered.

Janet D Rowley. Chromosomal translocations: revisited yet again *Blood* (2008), 112(6)



Pane et al. BCR/ABL genes …. *Oncogene* (2002), 21 (56)

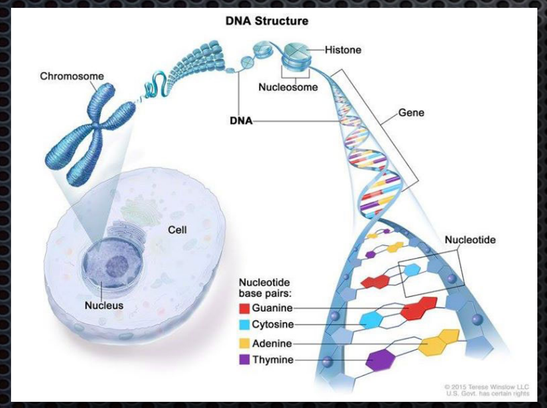


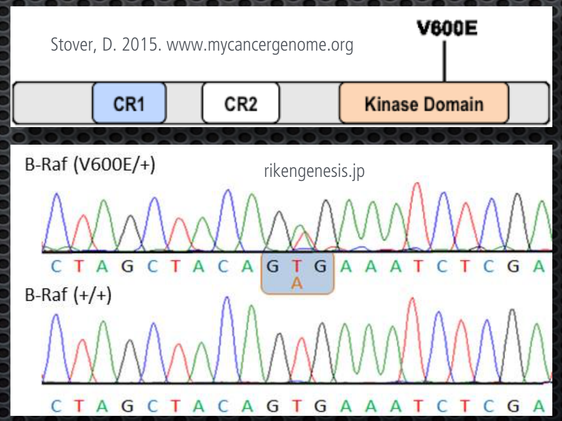
Event free Survival in first large Imatinib Trials

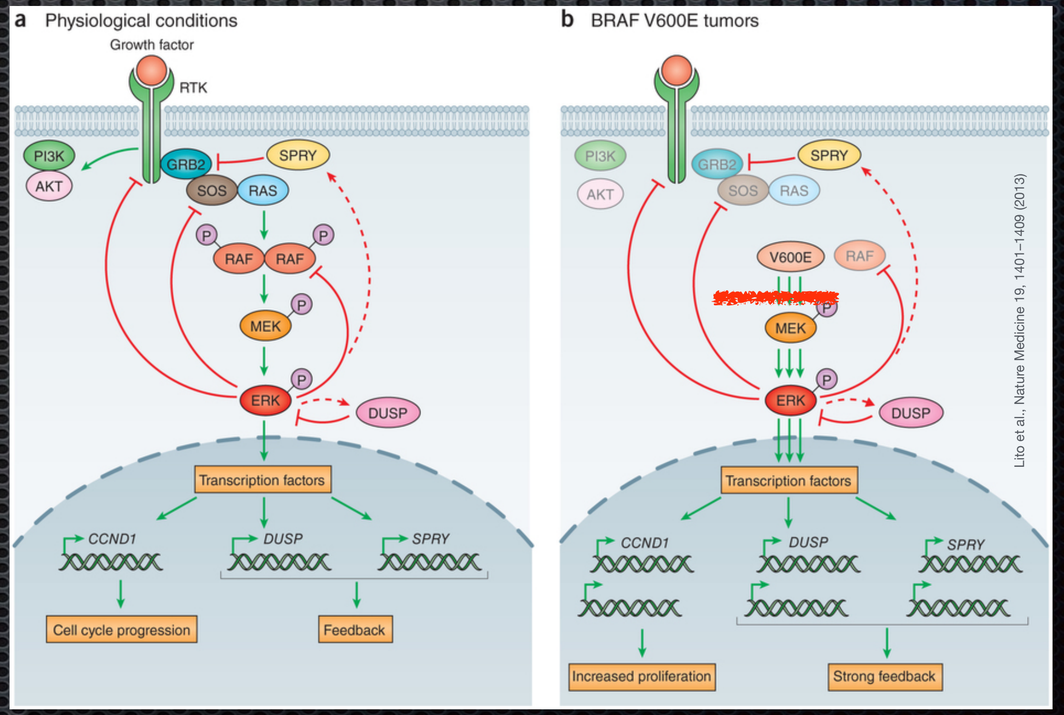
O'Brien et al. Imatinib compared with interferon and low-dose cytarabine… NEJM (2003) vol. 348 (11)

BRAF V600E (c.1799T>A) Mutation
Oncogene Activation by Single Nucleotide Alteration

- a single nucleotide exchange **T**hymidine > **A**denine leads to continuous RAF based activation of the MEK-ERK pathway
- BRAF V600E is a frequent mutation in >50% of malignant melanomas, but also CRC, lung ADC …
- pharmacologic block of B-Raf (e.g. through ***Vemurafenib***)

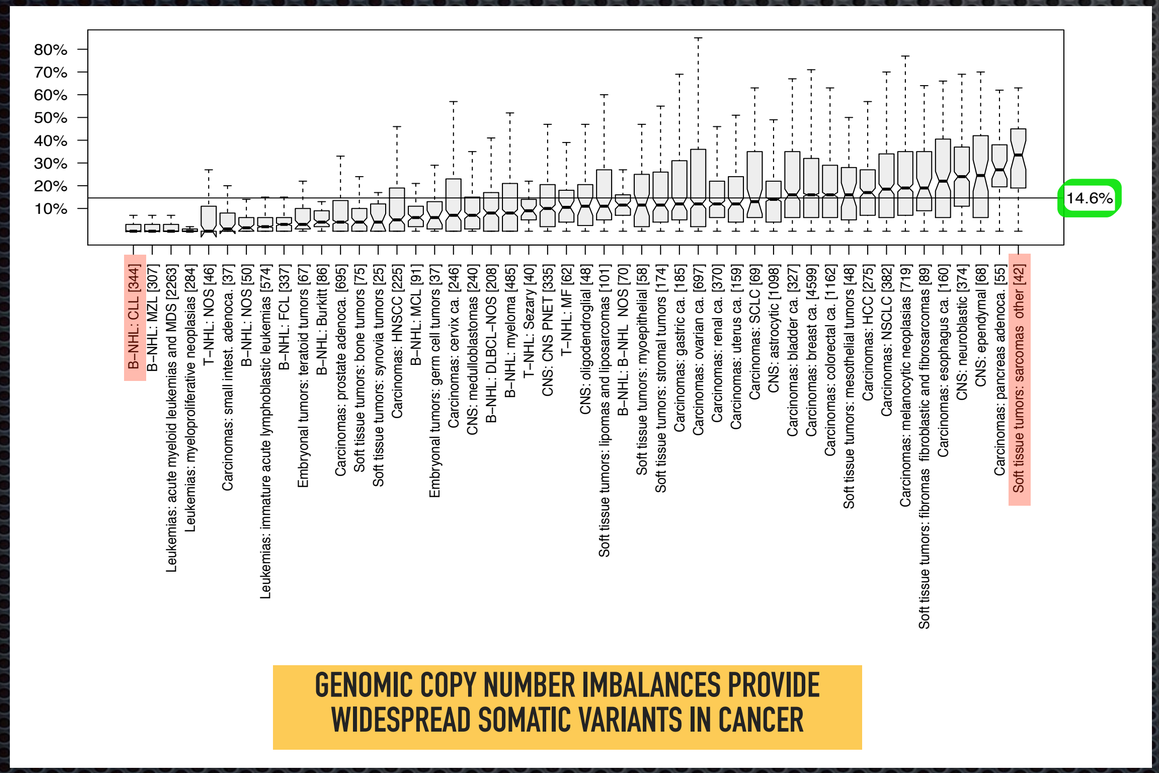






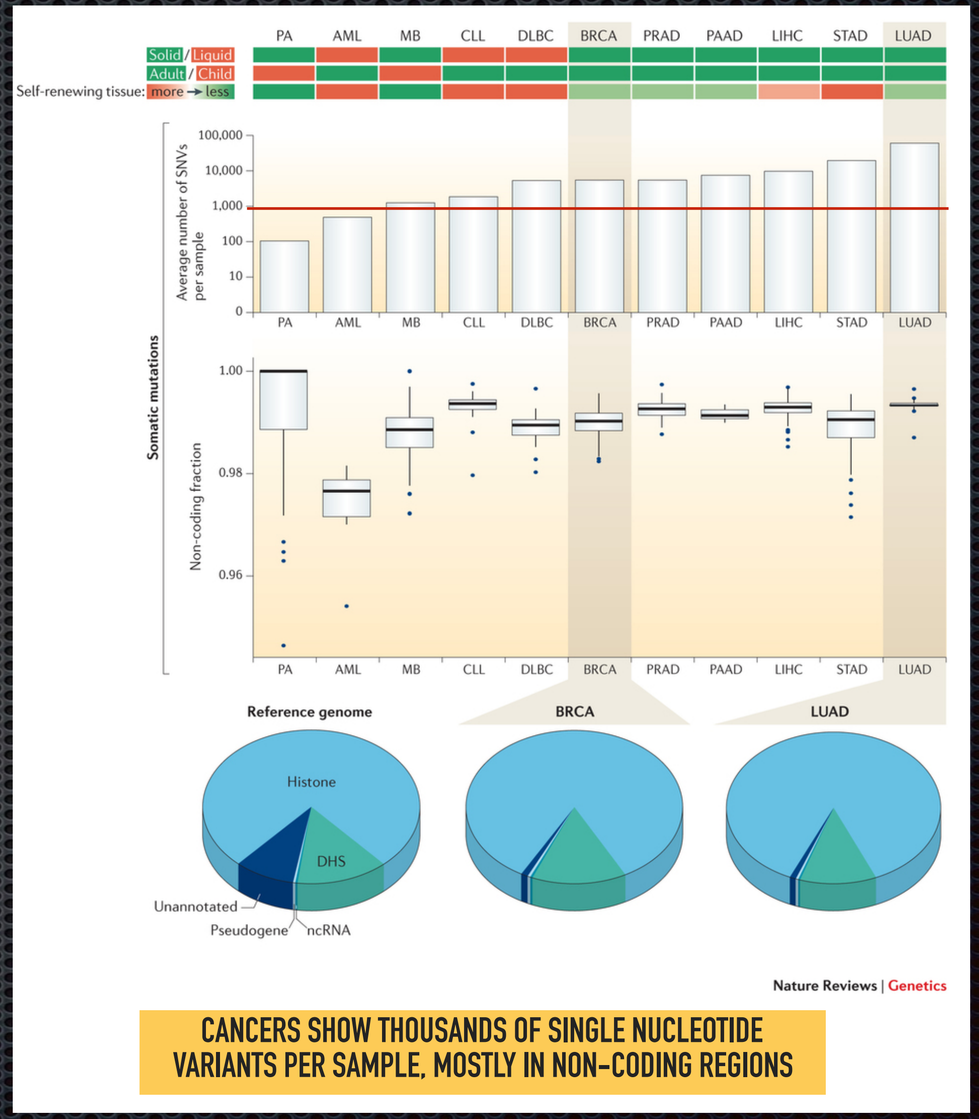
The BRAF V600E mutation leads to continuos phosphorylation of MEK, without the need for receptor based activation of the upstream pathway and loss of inhibitory feedback control.

Quantifying Somatic Mutations In Cancer



GENOMIC COPY NUMBER IMBALANCES PROVIDE WIDESPREAD SOMATIC VARIANTS IN CANCER

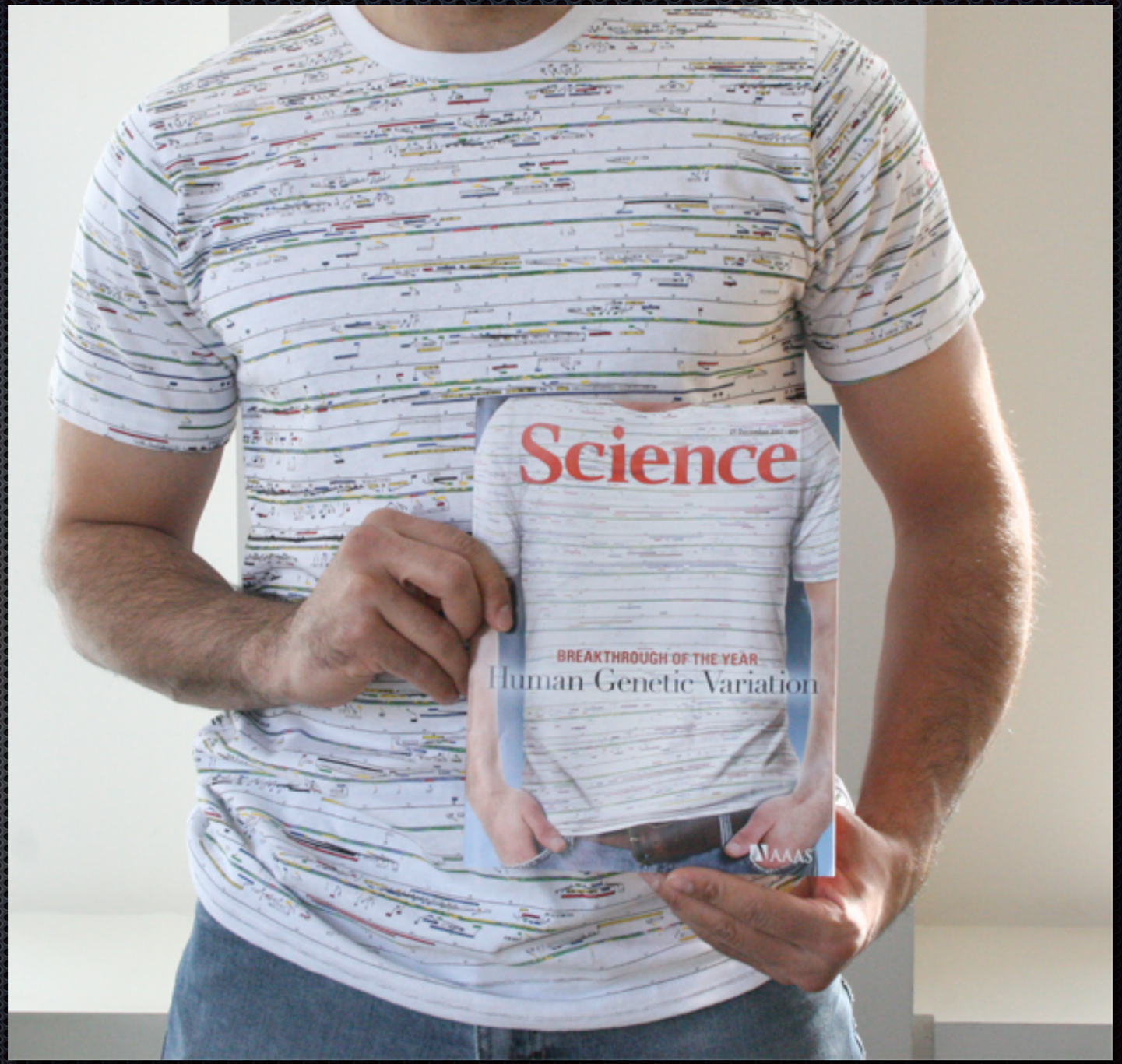
On average ~15% of a cancer genome are in an imbalanced state (more/less than 2 alleles); Original data based on >30'000 cancer genomes from arraymap.org



CANCERS SHOW THOUSANDS OF SINGLE NUCLEOTIDE VARIANTS PER SAMPLE, MOSTLY IN NON-CODING REGIONS

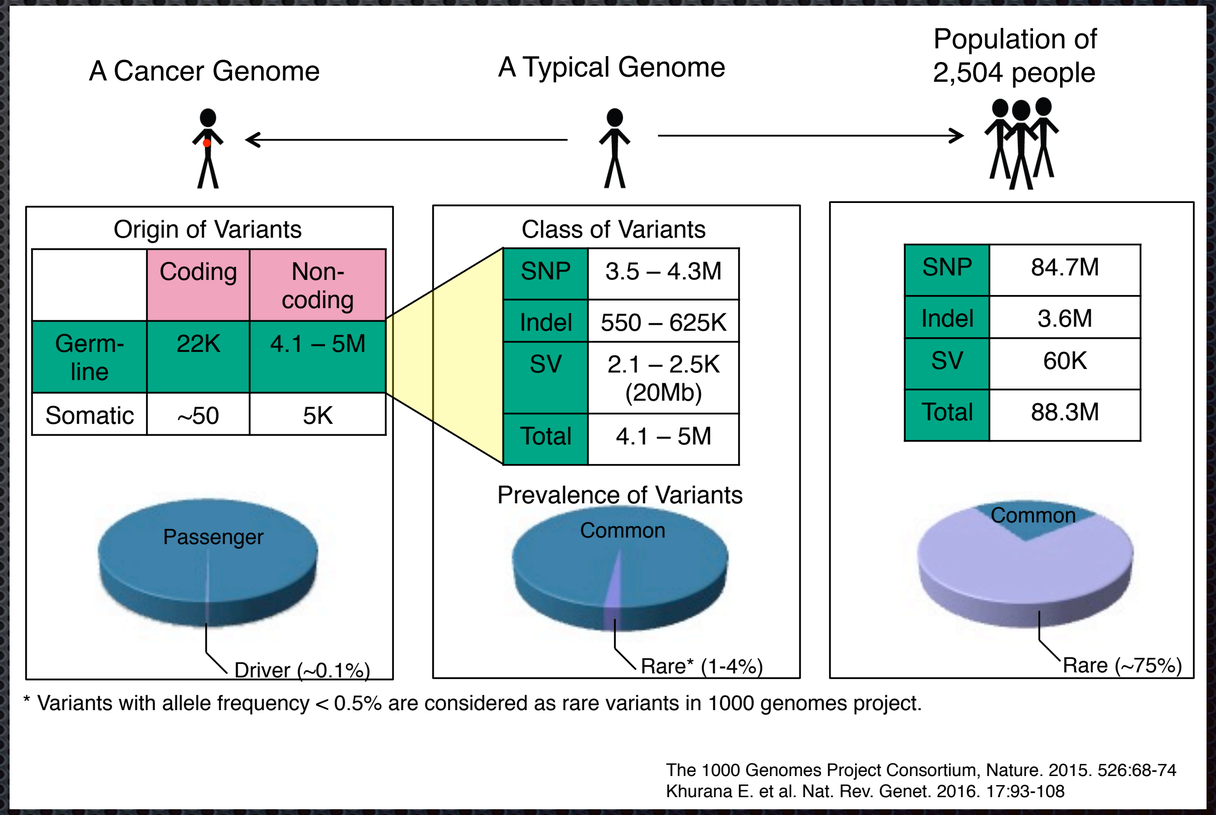
Pan-Cancer Analysis of Whole Genomes (PCAWG) data show widespread mutations in non-coding regions of cancer genomes (Khurana et al., Nat. Rev. Genet. (2016)

The trouble with human genetic variation



Finding Somatic Mutations In Cancer:
Many Needles in a Large Haystack

- a typical humen genome (~3 billion base pairs) has ~5 million variants
- most of them are "rare"; i.e. can only be identified as recurring when sequencing thousands of people
- cancer cells accumulate additional variants, only **few** of which ("**drivers**") are relevant for the disease

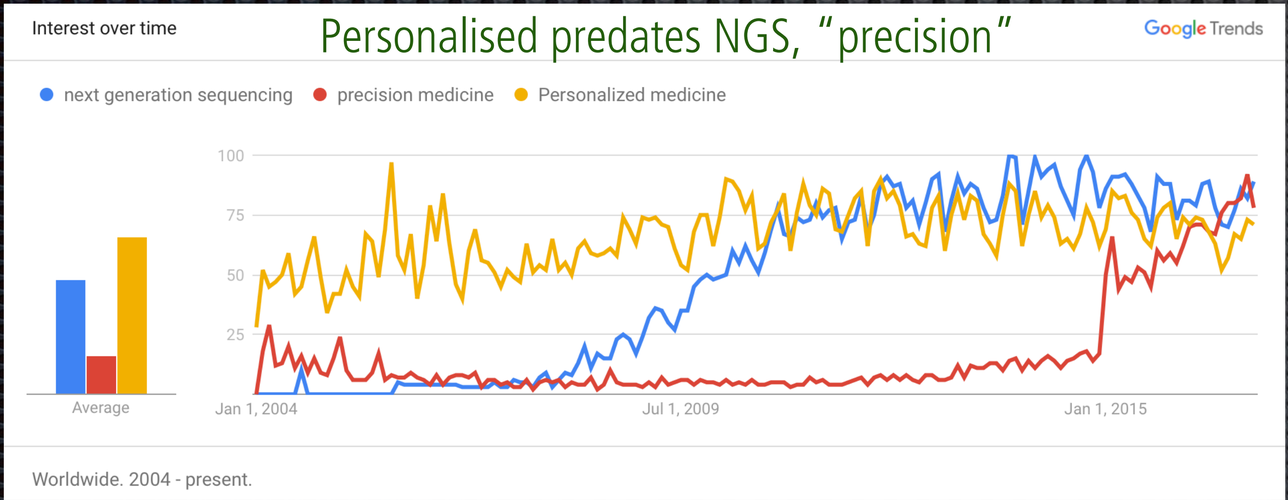


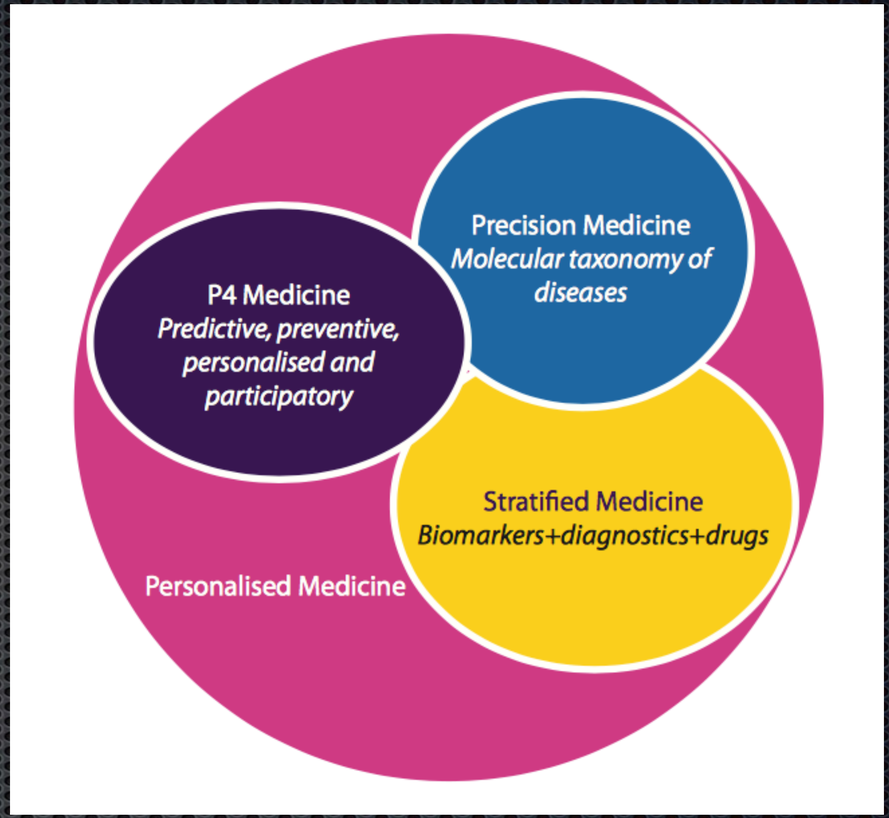
The 1000 Genomes Project Consortium, Nature. 2015. 526:68-74
Khurana E. et al. Nat. Rev. Genet. 2016. 17:93-108

Graphic adapted from Mark Gerstein (GersteinLab.org; @markgerstein)

Genomic Background + Disease Parameters

Personalised Medicine
Precision Medicine
…



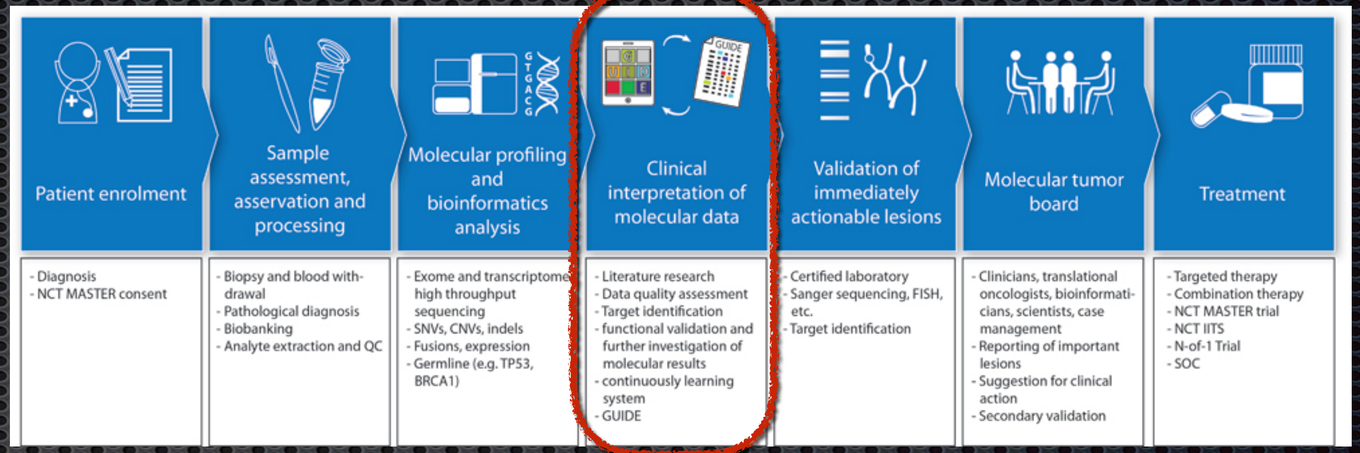


Source: PHG Foundation

While medicine has always been "personal" and "precise" in the given context of available knowledge and technologies, the concept of "**Personalised Medicine**" describes the use of individual genome information and individually targeted therapies.

Personalised Medicine in Cancer - A Genome Based Approach

- personalized cancer therapy uses information about the **individual genetic background** and **tumor sequence analysis** for the identification of somatic variants



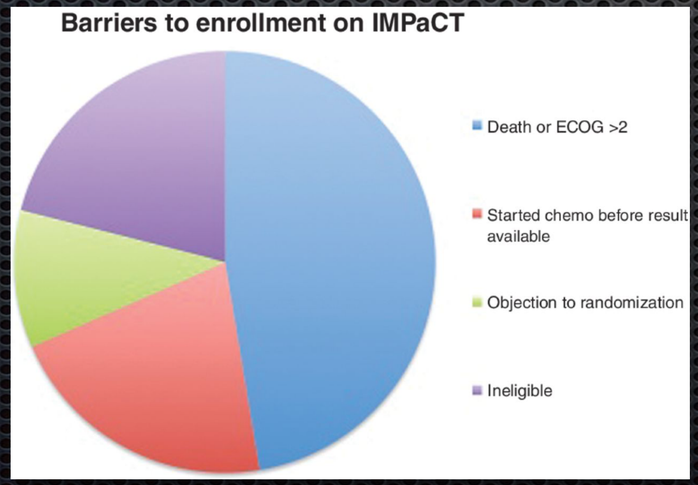
Workflow of a cancer treatment protocol based on "personalized" assessment of actionable genomic lesions (source: NCT Heidelberg).

- currently mostly use of **targeted / panel sequencing** for identification of tens - hundreds of most common "actionable" mutations

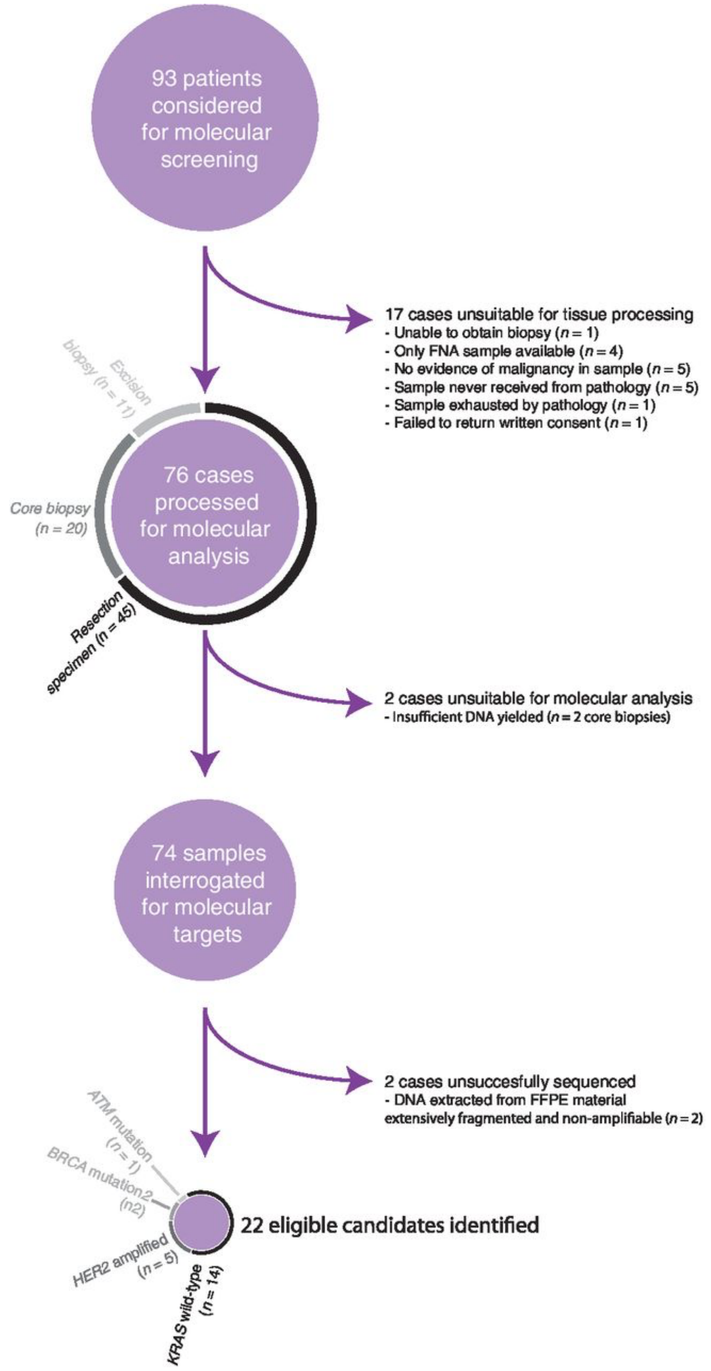
- knowledge resources and literature search for interpretation of non-standard variants

Personalised Cancer Therapy Needs More Targets

- with the current knowledge, targeted molecular analysis will not lead to the identification of actionable interventions in a majority of cancer cases



Precision Medicine for Advanced Pancreas Cancer: The Individualized Molecular Pancreatic Cancer Therapy (IMPaCT) Trial. Chantrill *et al.*, Clin Cancer Res. 2015 May 1;21(9):2029-37

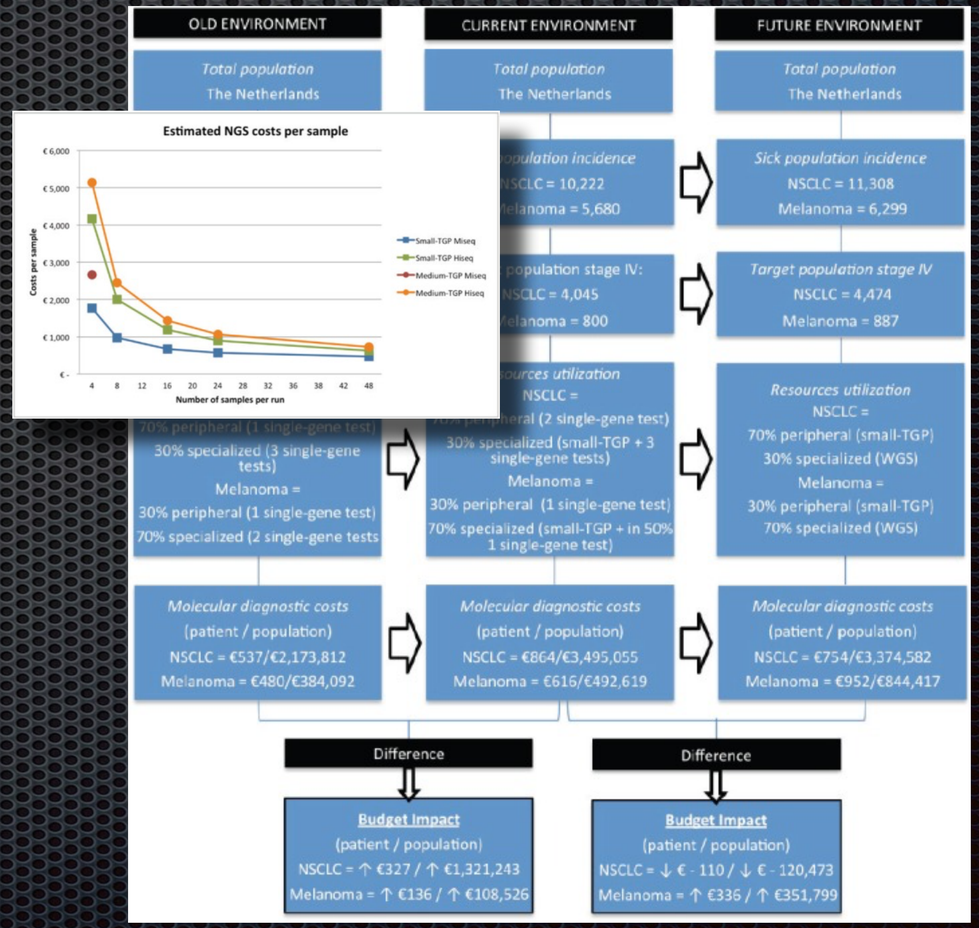


Next-generation sequencing strategies in malignancies

	Targeted panels	Whole exome	Whole genome
Pro	▸ High depth of coverage ▸ Readily standardisable ▸ Rapid interpretation for clinical use ▸ Low costs ▸ Easy clinical implementation	▸ Detection of unknown variants ▸ Detection of CNVs ▸ Research applications ▸ Feasible in clinical routine ▸ Low price/performance ratio	▸ Comprehensive assessment of cancer genomes ▸ Highest resolution of genomic alterations ▸ SNVs in enhancer/promoter and ncRNA regions ▸ Decreasing costs ▸ Subject to future studies
Contra	▸ Limited, 'peephole' observations ▸ Limited value for research ▸ Limited assessment of complex aberrations	▸ Not fully comprehensive ▸ Lower CNV resolution ▸ Amplification or exon capture necessary ▸ High bioinformatic effort ▸ Demanding clinical interpretation ▸ Time-consuming workflow	▸ Uncertain value for clinical interpretation ▸ Most expensive

CNV, copy number variant; ncRNA, non-coding RNA; SNV, single nucleotide variant.

Horak P, et al. ESMO Open 2016;1:e000094.



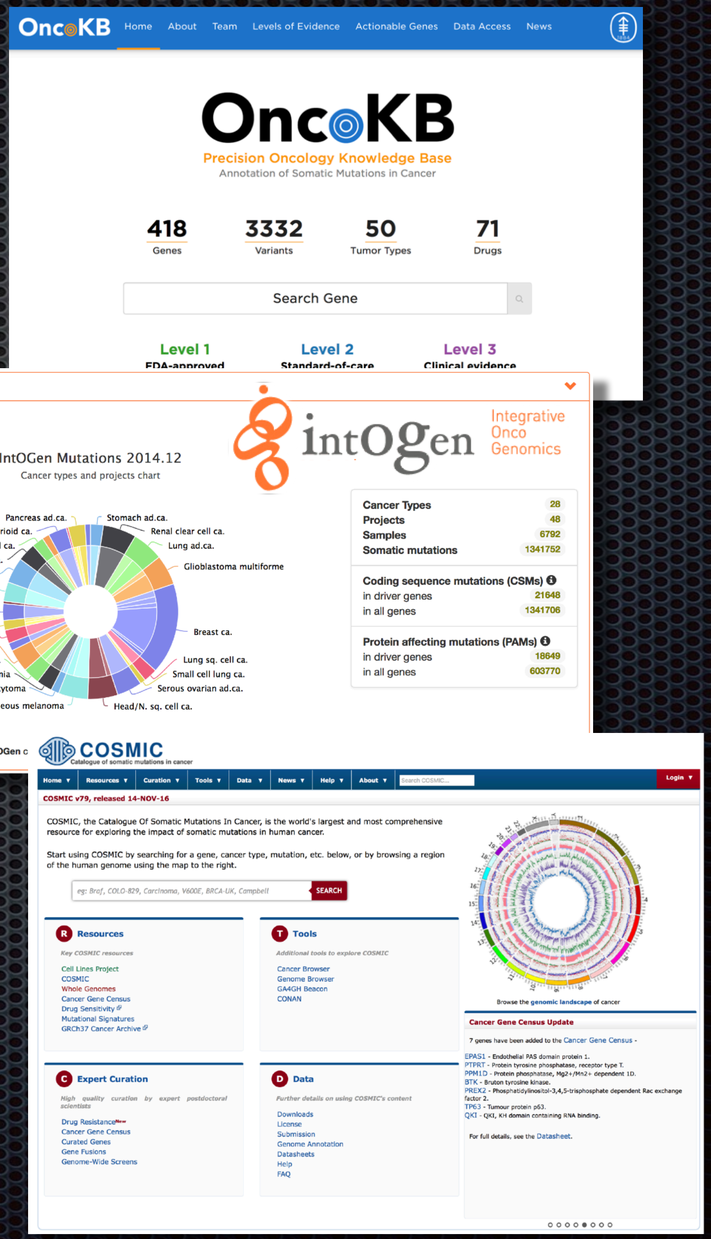
WGS use is expected to reduce the cost of additional tests

Estimated costs per sample for small- and medium TGP and WGS; van Amerongen et al., Ecancermedicalscience. 2016

Curated Variant Data Resources as Backbone of Personalised Cancer Therapy

- cancer variant interpretation resources apply manual **data curation** and **bioinformatics** methods to provide information about putative targets and possible interventions

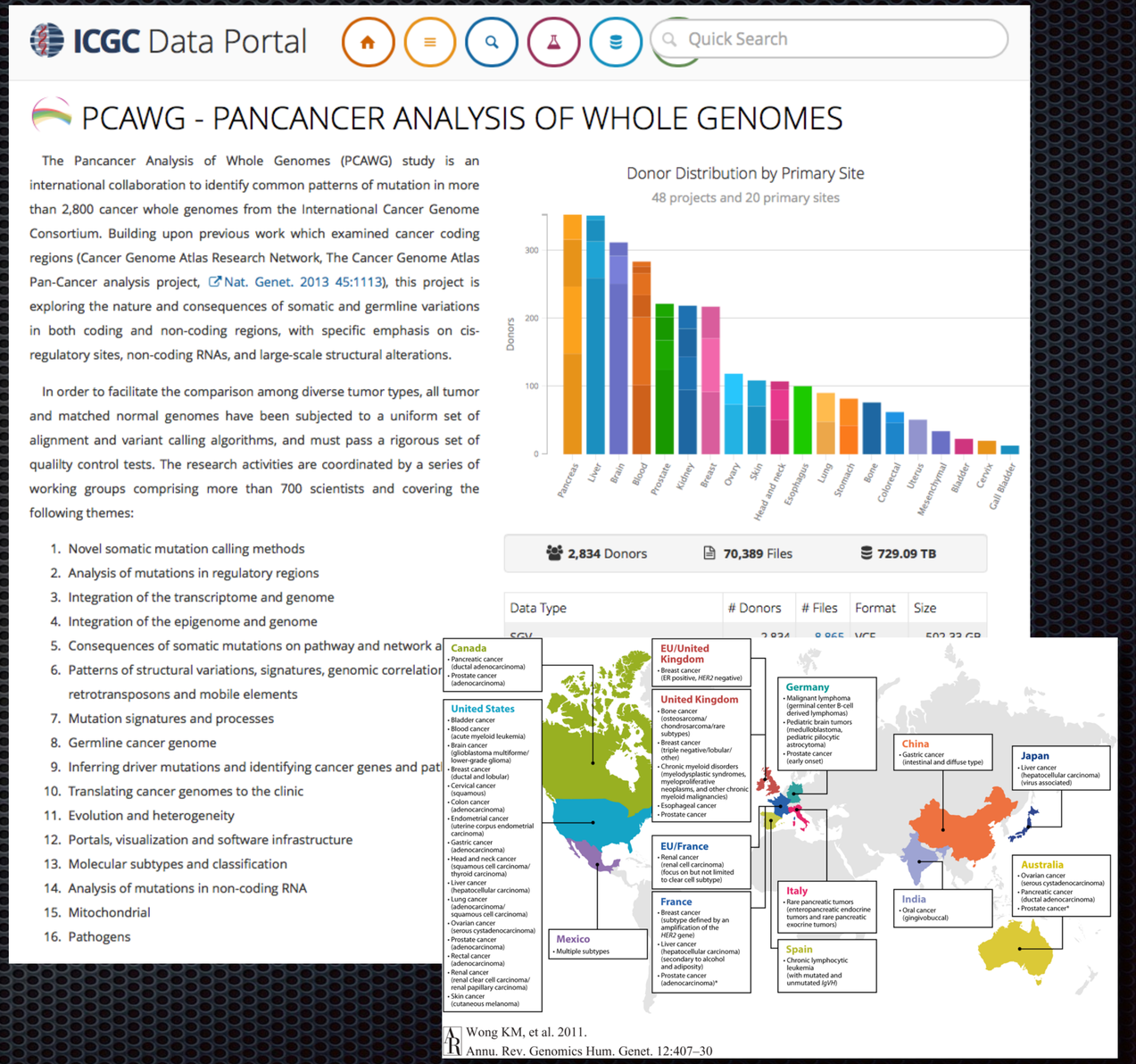
Database	Institute	Organized by
TARGET	BROAD	Gene
PCT	MD Anderson	Gene
cBioPortal / OncoKB	MSK	TCGA diseases
COSMIC	Sanger	Gene
IntOGen	University Pompeu Fabra	Gene
My Cancer Genome	Vanderbilt	Disease
CIViC	Washington University	Variant
DGIdb	Washington University	Drug/gene interaction



Genome Data Access in Cancer
Mining for **New** Knowledge

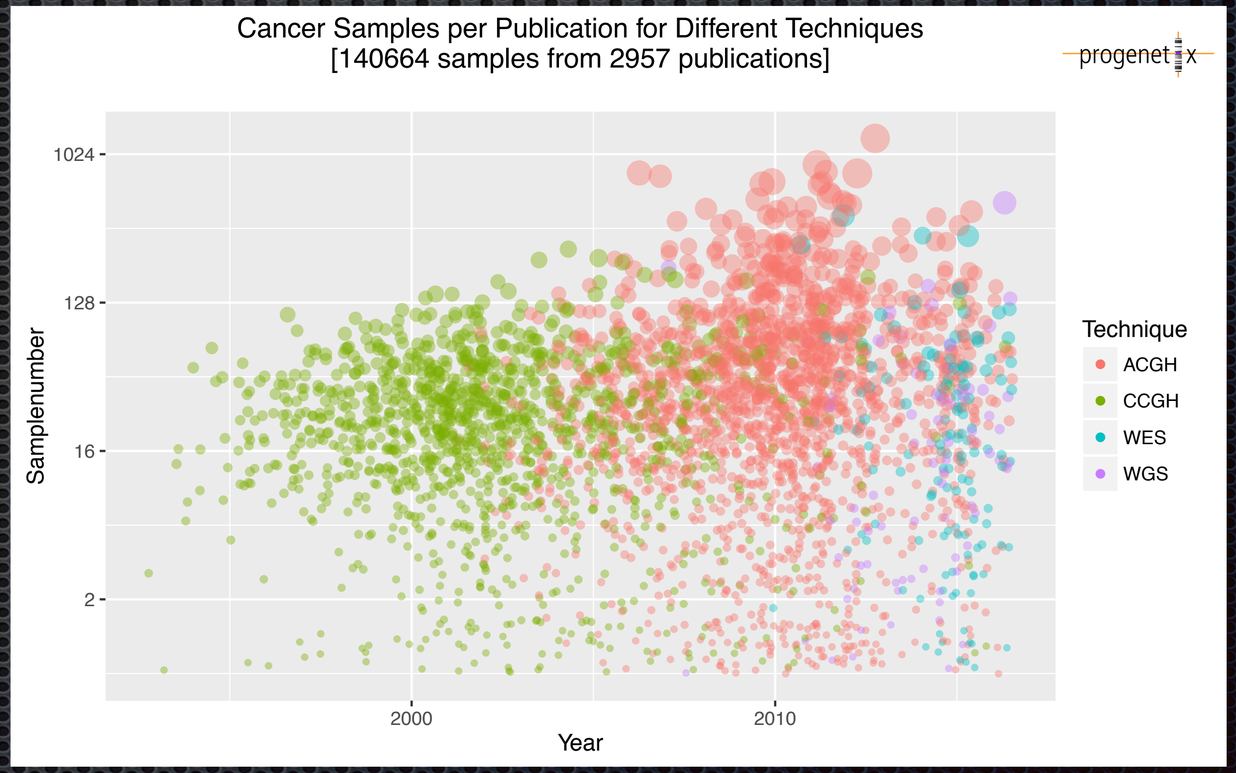
Genome-wide multi"*omics*" data generation for understanding tumor biology

- the International Cancer Genome Consortium (ICGC) as leading example of deep analysis of multiple cancer entities
- international collbaoration of leadinc research centers for each of ~20 tumor types
- limitations:
  - focus on prominant cancer types w/ limited representation of rare entities
  - data access policies influenced by national regulations and legal frameworks
  - technical heterogeneity



Molecular Cytogenetic & Sequencing Studies for **Whole Genome Profiling** in Cancer

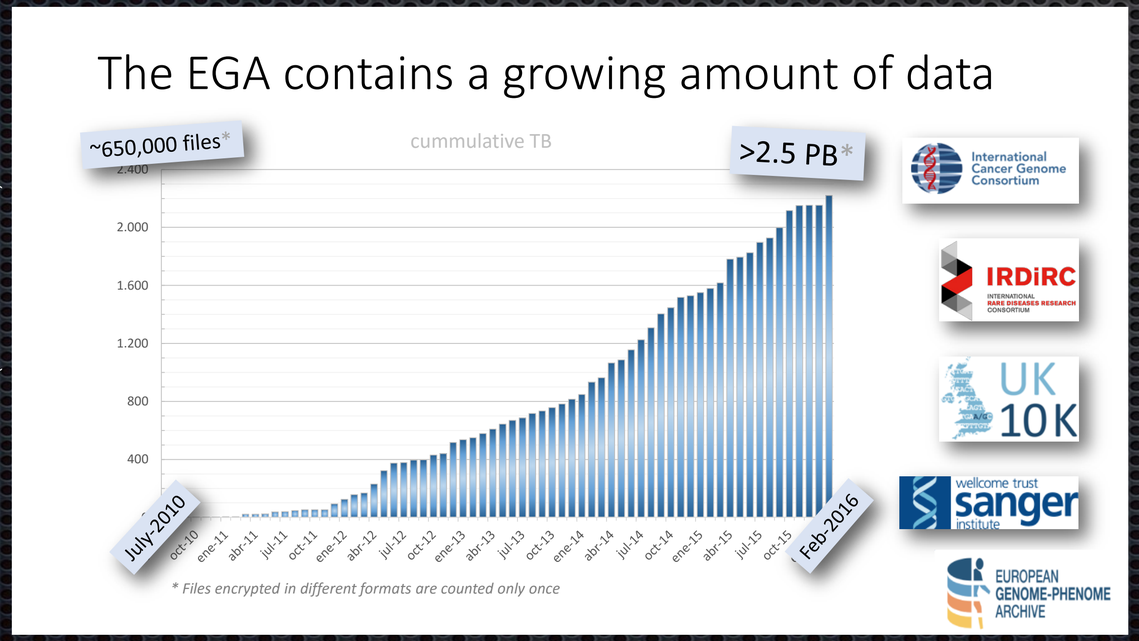
- genome screening to identify mutations in cancer samples
- for diagnostic purposes and therapeutic target identification
  - karyotyping (~1968)
  - **C**omparative **G**enomic **H**ybridization (1992)
  - genome **microarrays** (aCGH, SNP arrays ...; 1997)
  - **W**hole **E**xome **S**equencing (2010)
  - **W**hole **G**enome **S**equencing (2011)



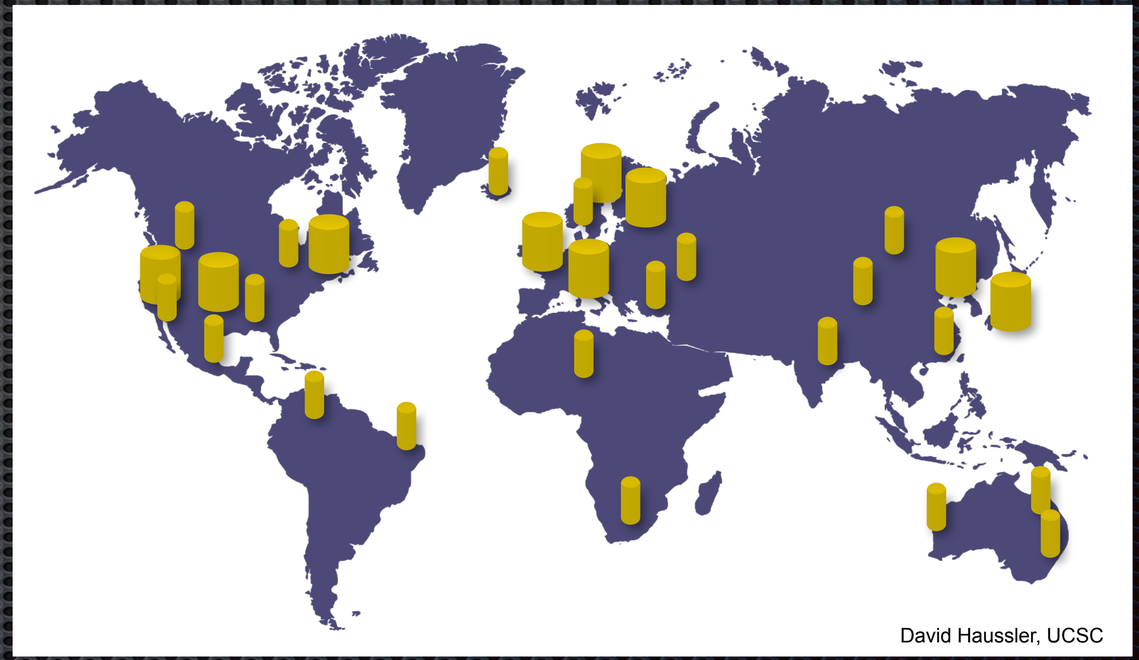
Overview of publications reporting whole-genome screening analysis of cancer samples, by molecular-cytogenic or genome sequencing methods. The data represents articles assessed for the progenetix.org cancer genome data resource (M. Baudis, 2001-2016)

Genome Datasets: Rapid Growth, Limited Access

population based and cancer research studies produce a rapidly increasing amount of genome sequence data



Source: Jordi Rambla (CRG Barcelona) 2016



genome data is stored in an increasing number of institutional and core repositories, with **incompatible data** structures and **access** policies

Genomes Everywhere

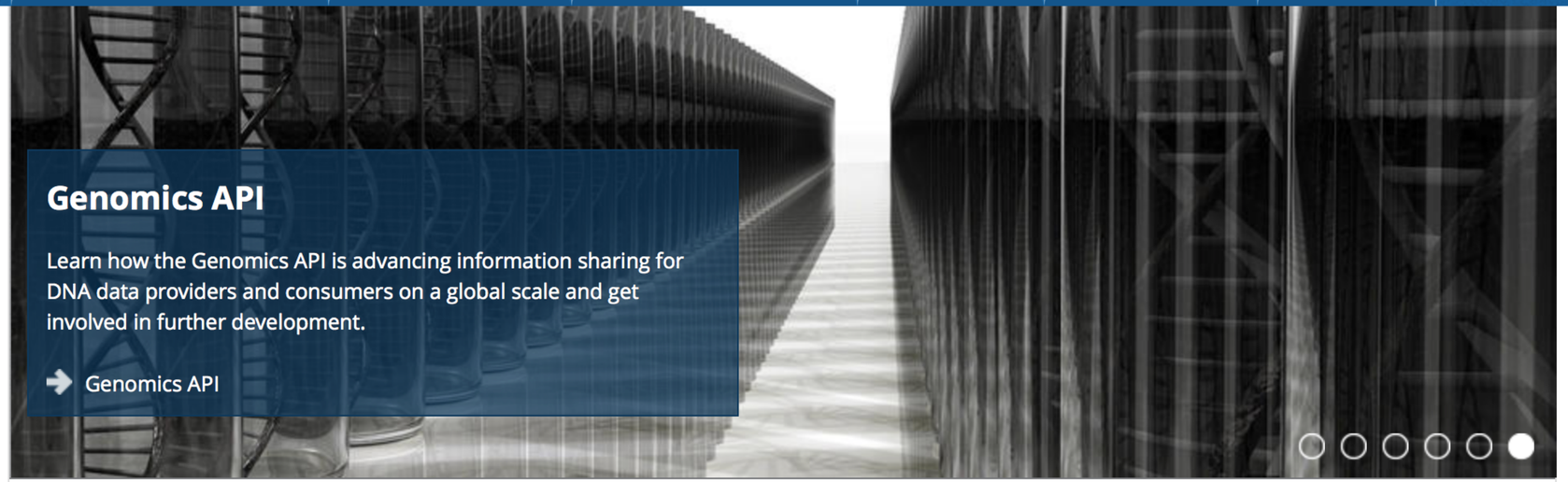
Large Genome Data Generation, Analysis & Sharing Initiatives

Organization / Initiative: Name	Organization / Initiative: Category	Cohort
100K Wellness Project	Research Project	107 unaffected individuals (scaling up to 100,000)
23andMe	Organization	>1 million customers (>80% consented to research)
Actionable Cancer Genome Initiative (ACGI)	Data-Sharing Project	Goal: 100,000 individuals
Ancestry.com	Organization	1.4 million customer DNA samples (what % consented to research?)
BioBank Japan	Repository	Specimens from >200,000 patients and unaffected controls
Cancer Moonshot2020	Consortium	Phase 1: 20,000 cancer patients
Children's Hospital of Philadelphia Biorepository	Repository	Capacity for 8.6 million samples
China Kadoorie Biobank	Repository	>512,000 participants (general population, China). Genotyping data available for ~100,000.
CIMBA	Consortium	>15,000 BRCA1 carriers, >8,000 BRCA2 carriers
Clinical Sequencing Exploratory Research (CSER)	Consortium	~4,000 patients and healthy controls
DECIPHER	Repository	19,014 patients (international)
deCode Genetics	Organization	500,000 participants (international)
East London Genes & Health	Research Project	100,000 unaffected individuals (East London, Pakistani or Bangladeshi heritage)
Electronic Medical Records and Genomics (eMERGE) Network	Repository, Consortium, Research Project	55,028 patients
European Network for Genetic and Genomic Epidemiology (ENGAGE)	Research Project	80,000 GWAS scans, and DNA and serum/plasma from >600,000 individuals
Exome Aggregation Consortium (ExAC)	Consortium	60,706 individuals
GENIE/AACR	Data-Sharing Project	>17,000 cancer patients (international)
Genome Asia 100K	Consortium	Goal: 100,000 individuals (Asia)
Genomics England	Organization	Goal: 100,000 genomes from 70,000 individuals (rare disease & cancer patients, and their relatives)
GoT2D	Consortium, Data-Sharing Project	Multiple case-control cohorts
International Cancer Genome Consortium (ICGC)	Consortium	currently data from >16'000 samples
International Genomics of Alzheimer's Project (IGAP)	Consortium	40,000 patients with Alzheimer's disease
International Multiple Sclerosis Genetics (IMSG) Consortium	Consortium	Goal: >50,000 patients with MS
Kaiser Permanente: Genes, Environment, and Health (RPGEH)	Repository, Research Project	200,000 DNA samples (scaling up to 500,000)
Leiden Open Variation Database (LOVD)	Repository	>170,000 individuals
Million Veteran Program	Research Project	Goal: 1 million individuals; first 200,000 is complete.
MyCode® Community Health Initiative	Repository, Research Project	Goal: >250,000 patients
Precision Medicine Initiative	Research Project	Goal: >1 million participants, starting in 2016 (US)
Psychiatric Genomics Consortium (PGC)	Consortium	>170,000 subjects
Resilience Project	Research Project	589,306 individuals
Saudi Human Genome Program	Research Project	Goal: ~100,000 patients and controls (Saudi Arabia)
Scottish Genomes Partnership (SGP)	Research Project	>3,000 individuals (Scotland)
T2D-GENES	Consortium, Data-Sharing Project	10,000 patients and controls (five ethnicities); 600 individuals (Mexican American)
TBResist	Consortium	>2,600 samples
UK Biobank	Repository, Consortium, Research Project	500,000 individuals (age 40-69 years; UK)
UK10K	Research Project	10,000 participants (6,000 patients and 4,000 controls)
Vanderbilt's BioVU	Repository	>215,000 samples

Global Alliance
for Genomics & Health

Become a Member

ABOUT GLOBAL ALLIANCE | WORKING GROUPS | PRODUCTS & PROJECTS | MEMBERSHIP | NEWS & EVENTS | CONTACT US



Our Work

The diverse members of the Global Alliance are working together to create interoperable approaches to catalyze projects that will help unlock the great potential of genomic data. Our four Working Groups advance Initiatives that develop key Work Products.

Clinical »

Aims to enable compatible, readily accessible, and scalable approaches for sharing clinical data and linking it with genomic data.

Data »

Concentrates on data representation, storage, and analysis of genomic data to develop approaches that facilitate interoperability.

Regulatory and Ethics »

Focuses on ethics and the legal and social implications of the Global Alliance, including harmonizing policies and standards.

Security »

Leads the thinking on the technology aspects of data security, user access control, audit functions, and developing or adopting data security standards.

GA4GH MEMBERSHIP

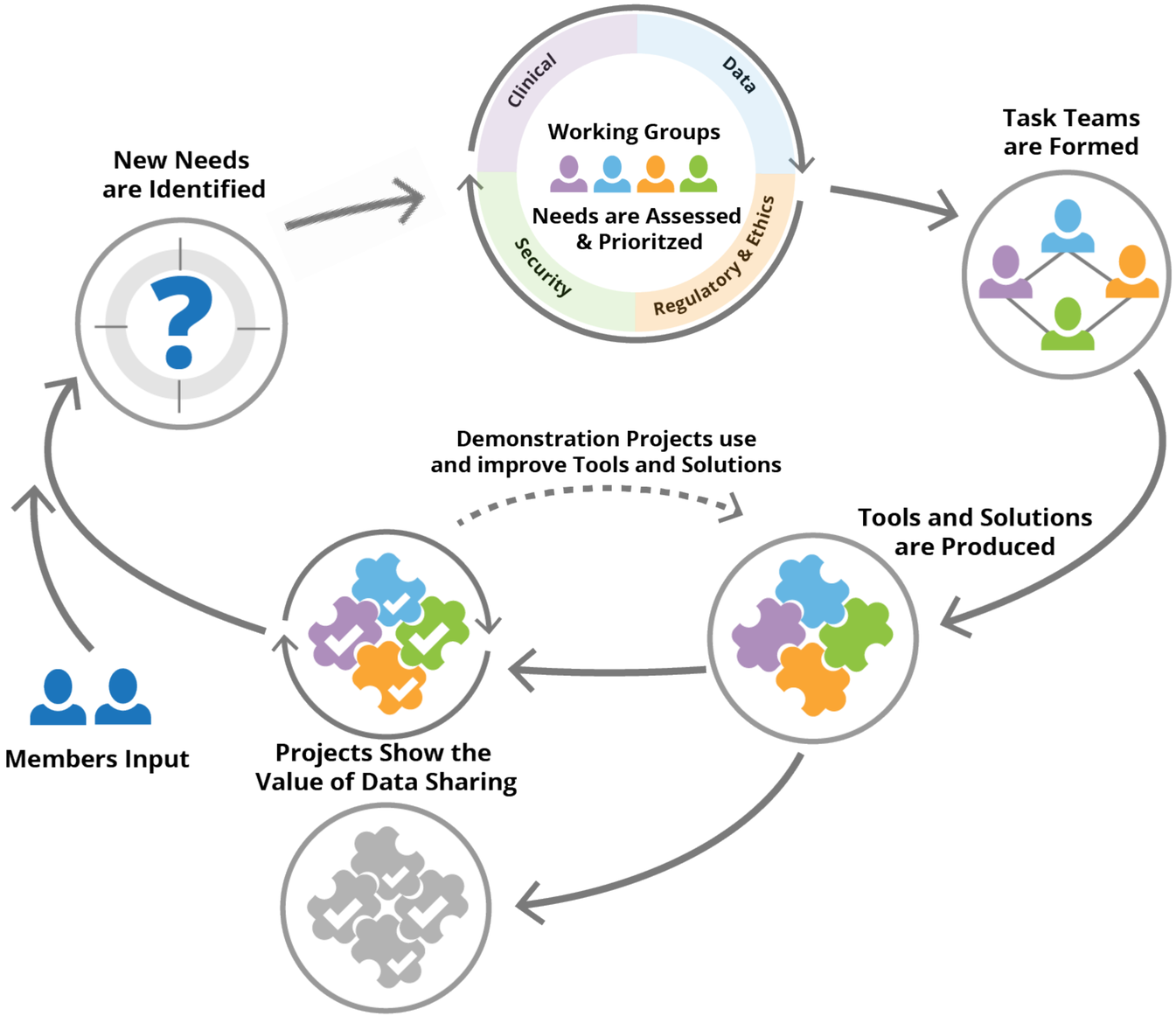
- host institutions
  - OICR, Broad Institute, Sanger Institute
  - Peter Goodhand, OICR, exec. director
- funding
  - CanSHARE, NIH, Wellcome Trust
- founding members
  - 229 partner organizations based in 30 countries as of October 3, 2014
  - SIB and UZH represented Switzerland
- membership status
  - 433 organizational members as of September 2016
  - open to individual registrations & participation
  - no financial commitment necessary; however, calls for support for events

Swiss Institute of Bioinformatics

Universität Zürich[UZH]

Global Alliance for Genomics & Health

How We Work



Global Alliance for Genomics & Health

ga4gh / schemas

Unwatch 115 Star 196 Fork 110

Code Issues 152 Pull requests 29 Projects 1 Wiki Pulse Graphs

Work on data models and APIs for Genomic data. http://ga4gh.org/#/api

1,102 commits **17** branches **16** releases **46** contributors Apache-2.0

Branch: **metadata-integ...** New pull request Create new file Upload files Find file Clone or download

This branch is 15 commits ahead, 3 commits behind master. Pull request Compare

mbaudis Merge branch 'master' into metadata-integration Latest commit `077c2c7` 2 days ago

doc	Merge branch 'master' into metadata-integration	2 days ago
python	Add constraints file	2 days ago
scripts	Utilize new common methods in schemas	2 days ago
src/main/proto	Merge branch 'master' into metadata-integration	13 days ago
tests	Utilize new common methods in schemas	2 days ago
tools	Merge branch 'master' into metadata-integration	13 days ago
.gitignore	Remove protoc call from install path (#781)	7 days ago
.travis.yml	Add constraints file	2 days ago
CONTRIBUTING.rst	Convert Avro -> proto3.	10 months ago

Global Alliance for Genomics & Health

GA4GH Driver Projects

BRCA Challenge
The BRCA Challenge aims to advance understanding of the genetic basis of breast and other cancers using data from around the world.

Beacon Project
Beacon Project is an open web service that tests the willingness of international sites to share genetic data. It is being implemented on the websites of the world's top genomic research organizations.

Matchmaker Exchange
Matchmaker Exchange is a federated network of databases whose goal is to find genetic causes of rare diseases by matching similar phenotypic and genotypic profiles.

The Cancer Gene Trust
proposes to aggregate somatic cancer mutation data and some clinical data in order to improve the genomic landscape of actionability in some cancers and to enable greater personalized clinical care for individuals with rare cancer mutations.

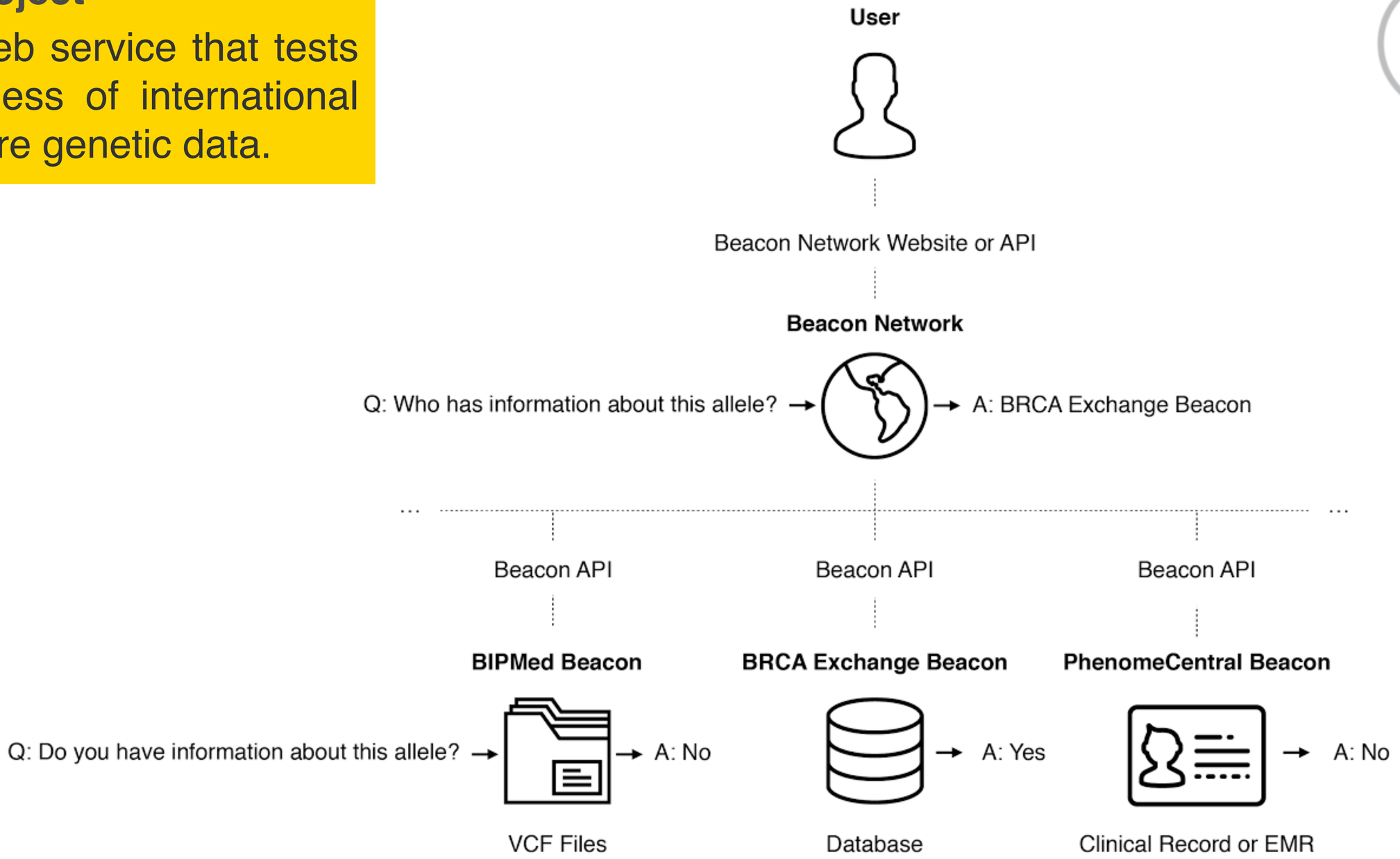
BRCA CHALLENGE

Matchmaker Exchange

Cancer Gene Trust

Beacon Project

An open web service that tests the willingness of international sites to share genetic data.



Global Alliance for Genomics & Health

GA4GH API promotes sharing

GENOMICS

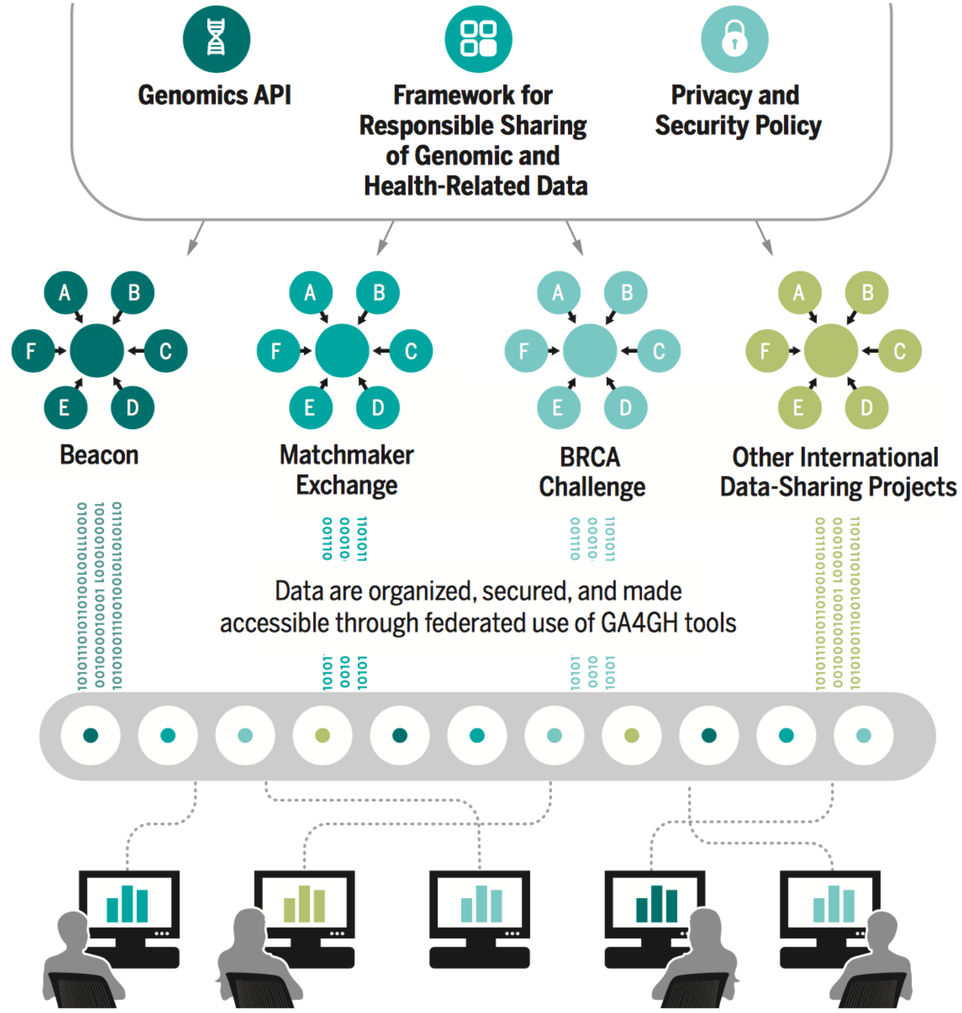
A federated ecosystem for sharing genomic, clinical data

Silos of genome data collection are being transformed into seamlessly connected, independent systems

The Global Alliance for Genomics and Health*

SCIENCE 10 JUNE 2016 • VOL 352 ISSUE 6291

A federated data ecosystem. To share genomic data globally, this approach furthers medical research without requiring compatible data sets or compromising patient identity.



Global Alliance for Genomics & Health

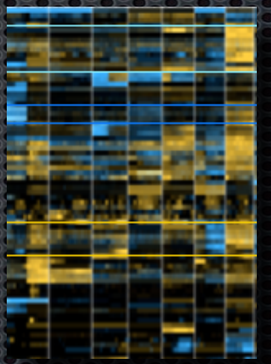
Cancer Genome Data & Beyond

Contributing our share

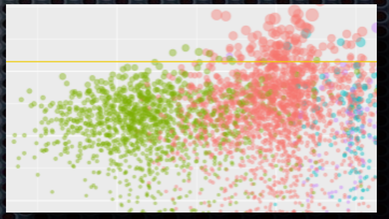
- Cancer genome data resources
- Software tools for data analysis & visualisation
- Parsing the cancer genome landscape: Patterns & targets
- GA4GH: Standards & Beacons
- Quantifying cancer genomic research
- The Swiss Personalized Health Network initiative
- Collaborations!

progenetix

arrayMap



Global Alliance for Genomics & Health



SAMW ASSM

akademien der wissenschaften schweiz

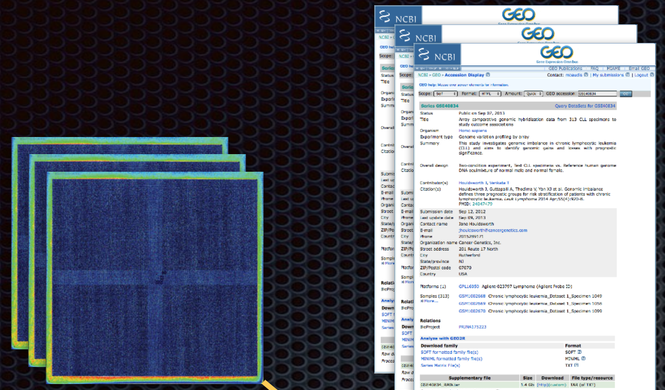
SIB Swiss Institute of Bioinformatics

Reference Resources for Cancer Genome Profiling

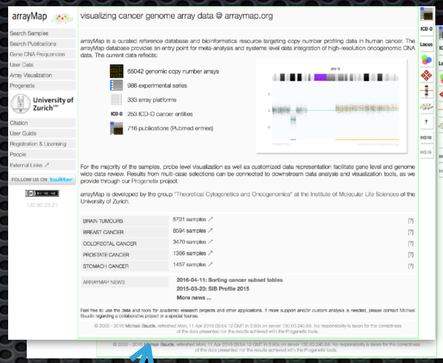
- continuously updated reference resources for cancer genome profiling data and related information
- basis for own research activities, collaborative projects and external use
- structured information serves for implementing GA4GH concepts

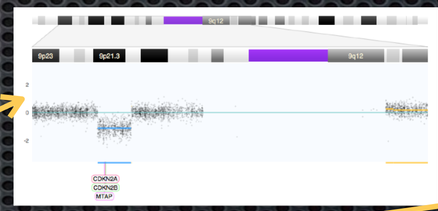
	progenetix	arrayMap
techniques	cCGH, aCGH, WES, WGS	aCGH (+?)
scope	sample (e.g. combination of several experiments)	experiment
content	**>31000** samples	**>60000** arrays
raw data presentation	no (link to sources if available)	yes (raw, log2, segmentation if available)
per sample re-analysis	no; supervised result (mostly as provided through publication)	yes (re-segmentation, thresholding, size filters ...)
final data	annotated/interpreted CN status for GP and cytogenetic regions	unsupervised CN status for GP and cytogenetic regions
main purposes	• Distribution of CNA target regions in most tumor types (>350 ICD-O) • Cancer classification	• Gene specific hits • Genome feature correlation (fragile sites ...)

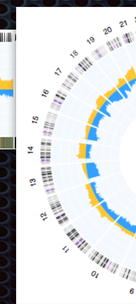
DATA PIPELINE

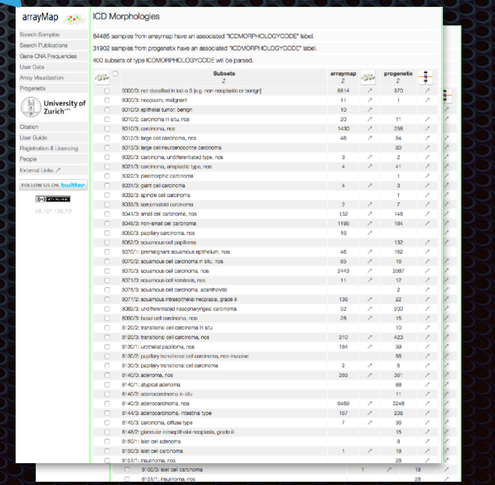


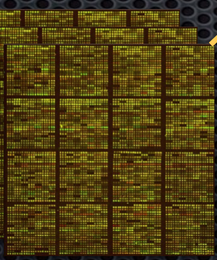
Automagical Processing Engine™

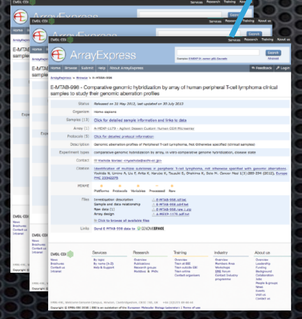






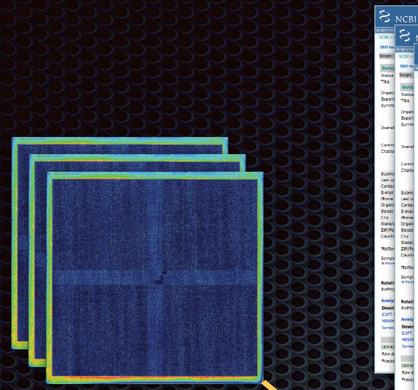


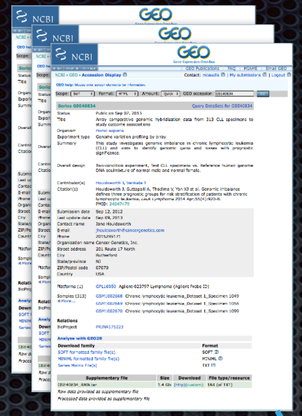


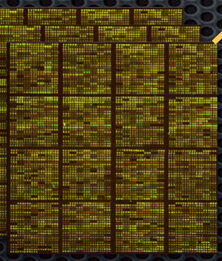


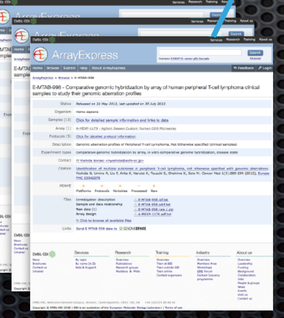
arrayMap

DATA PIPELINE



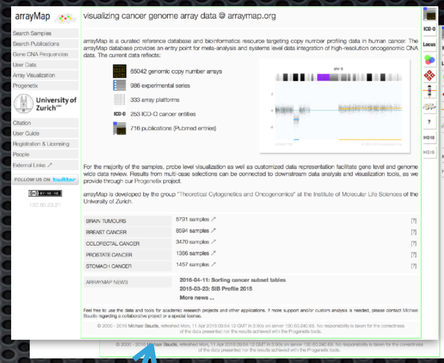


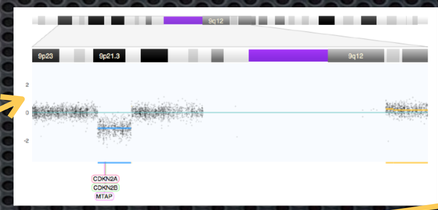


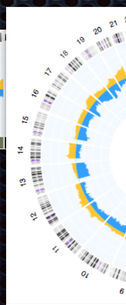


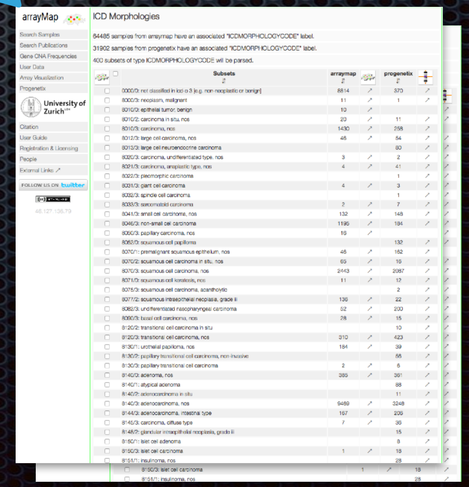
BIOCURATION

BIOINFORMATICS









arrayMap

ARRAYMAP DATA "PIPELINE"

GEO
ArrayExpress

amDownload

Metadata
(GEO soft files…)

amParseMeta

metadata.ods
(multi-sample)

amUpdateMeta

db.metadata

Array Probe Data
(log2
tables, .CEL)

Publications
Direct
Submissions

Array
Processing
Pipeline

Metadata Review
Selection
Entity Mapping
(ICD-O, Biosample,
Individual, clinical...)

Manual interpretation
of associated
information (article,
series descriprtion...)

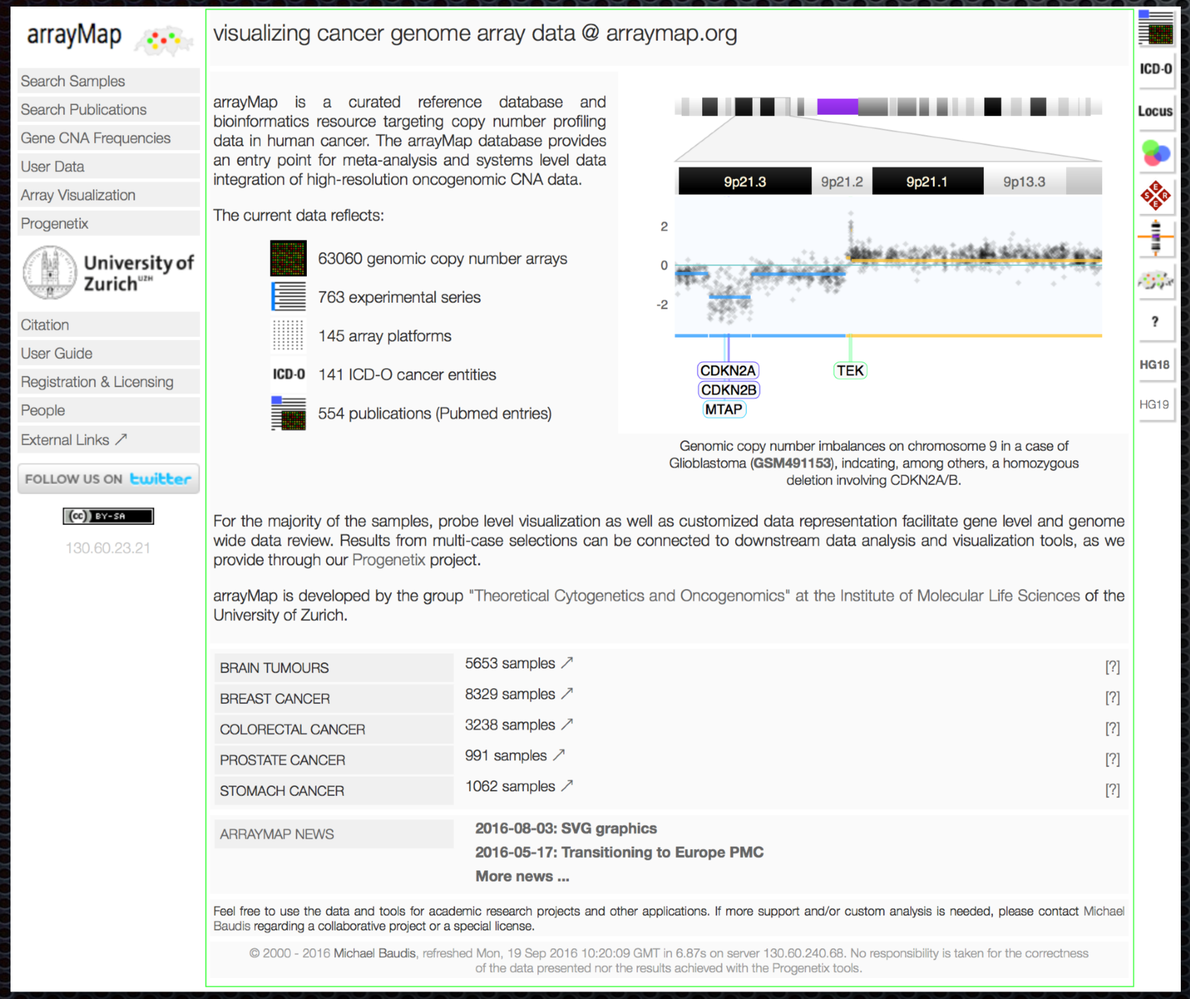
db.experiments

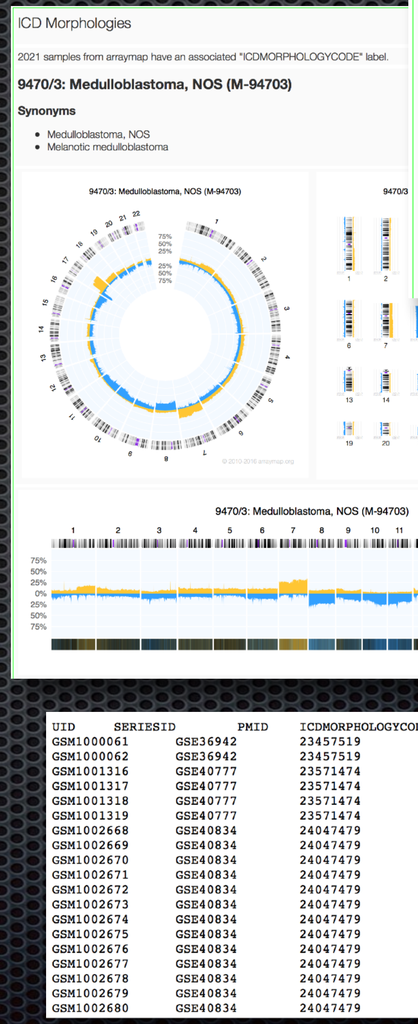
Array
Processing
Pipeline

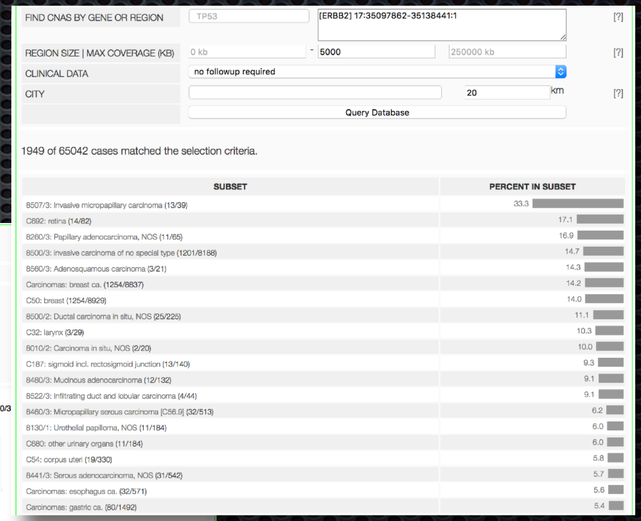
arrayMap

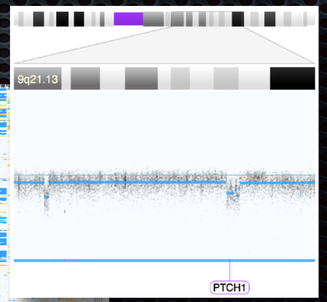
arrayMap

Resource for copy number variation data in cancer





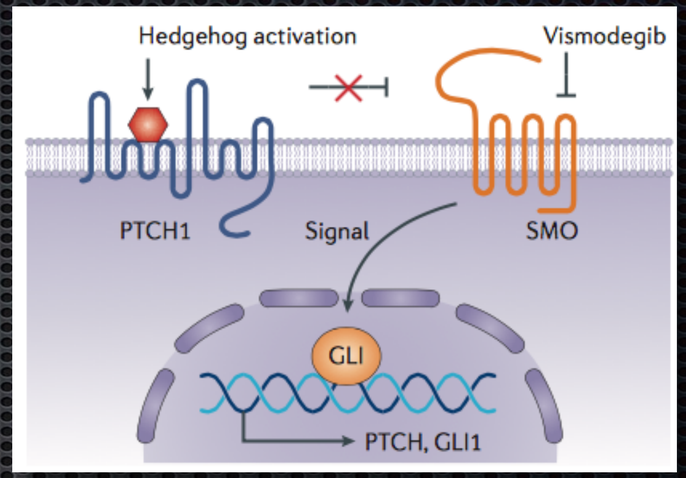




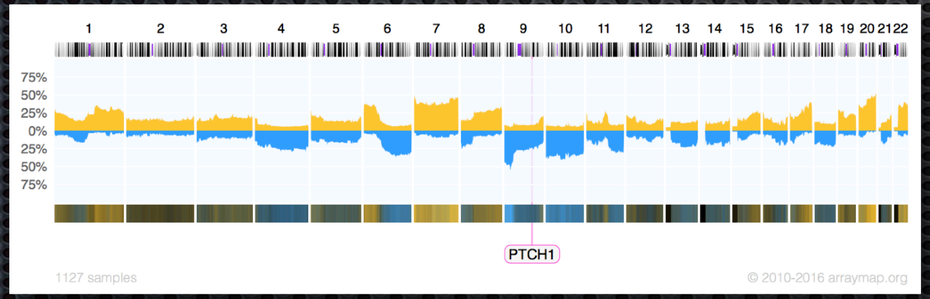
UID	SERIESID	PMID	ICDMORPHOLOGYCODE	ICDTOPOGRAPHYCODE
GSM1000061	GSE36942	23457519	8070/3	C10
GSM1000062	GSE36942	23457519	8070/3	C10
GSM1001316	GSE40777	23571474	8070/3	C53
GSM1001317	GSE40777	23571474	8010/3	C34
GSM1001318	GSE40777	23571474	8070/3	C09
GSM1001319	GSE40777	23571474	8010/3	C34
GSM1002668	GSE40834	24047479	9823/3	C42
GSM1002669	GSE40834	24047479	9823/3	C42
GSM1002670	GSE40834	24047479	9823/3	C42
GSM1002671	GSE40834	24047479	9823/3	C42
GSM1002672	GSE40834	24047479	9823/3	C42
GSM1002673	GSE40834	24047479	9823/3	C42
GSM1002674	GSE40834	24047479	9823/3	C42
GSM1002675	GSE40834	24047479	9823/3	C42
GSM1002676	GSE40834	24047479	9823/3	C42
GSM1002677	GSE40834	24047479	9823/3	C42
GSM1002678	GSE40834	24047479	9823/3	C42
GSM1002679	GSE40834	24047479	9823/3	C42
GSM1002680	GSE40834	24047479	9823/3	C42

Rare Events & Hidden Therapeutic Options?
Example: PTCH1 deletions in malignant melanomas

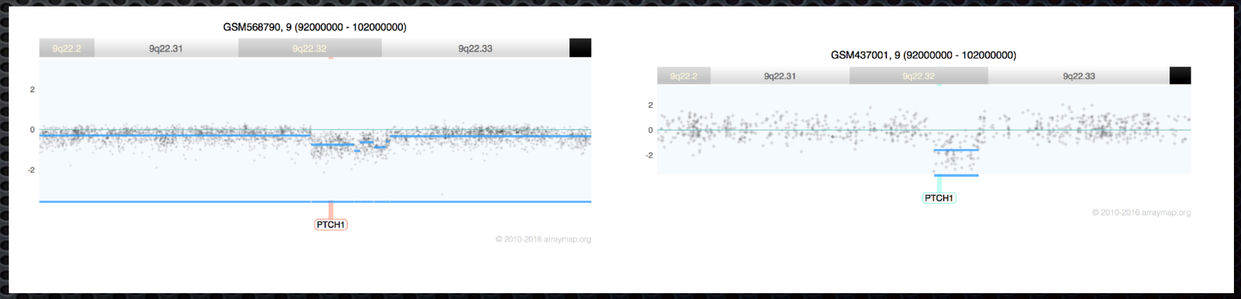
- PTCH1 is a actionable tumor suppressor gene, which has been demonstrated in e.g. basaliomas and medulloblastomas
- analysis of 1127 samples from 26 different publications could identify **focal** deletions in 4 samples
- a current project addresses the focal involvement of all mapped genes, in >50'000 cancer genome profiles



In its normal function, PTCH1 is a tumor suppressor gene in the sonic hedgehog pathway and inhibits SMO driven transcriptional activation. A loss of PTCH1 function (mutation, deletion) can be mitigated through drugs antagonistic to SMO activation.



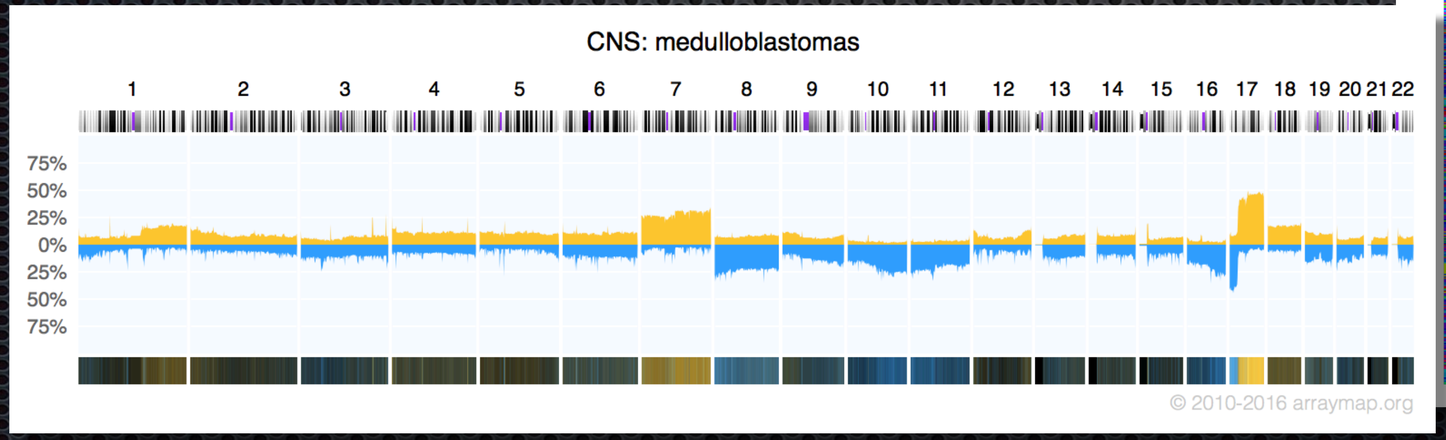
Summary of somatic copy number aberrations from the analysis of 1127 genome profiles of malignant melanomas, collected in our arraymap.org cancer genome resource. While PTCH1 does not represent a deletion hotspot, the genomic locus is part of larger deletions in ~25% of melanoma samples.

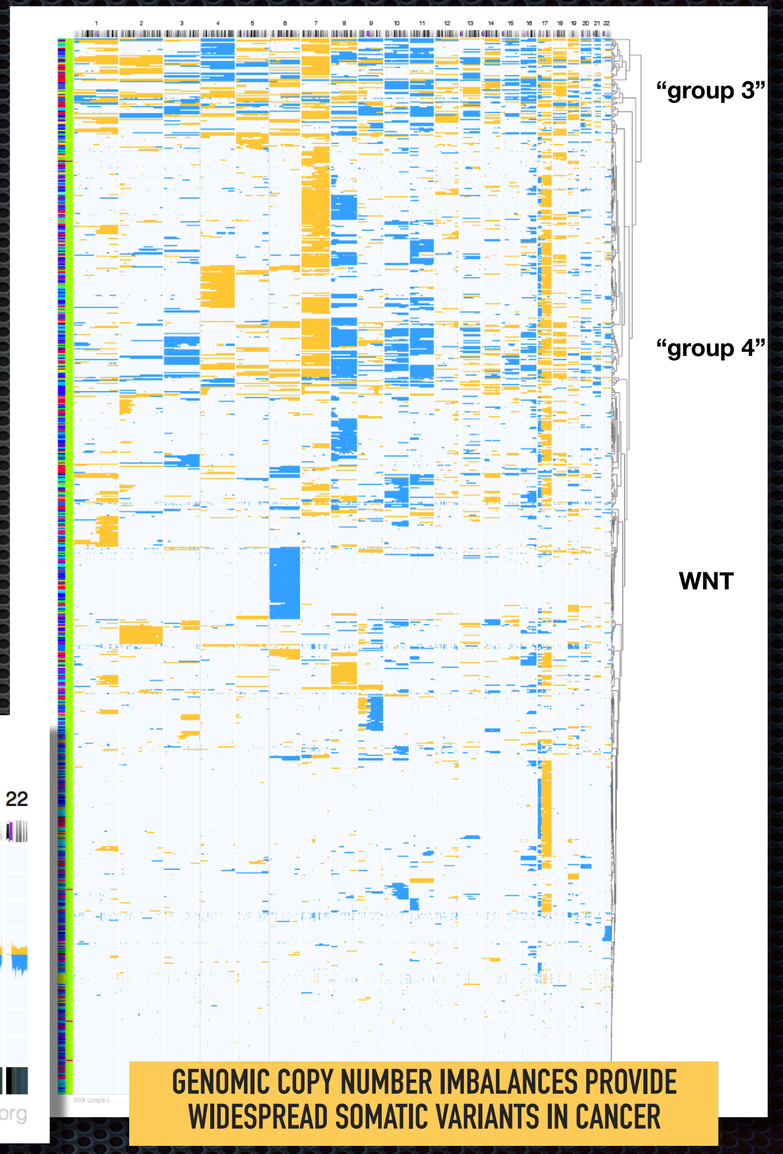


Examples of focal / homozygous PTCH1 deletions detected in the analysis of 1127 genomic array datasets. Focal somatic imbalance events are considered an indicator for oncogenic involvement of the affected target genes.

Somatic Mutations In Cancer: Patterns

- many tumor types express **recurrent mutation patterns**
- **How can** those patterns be used for classification and determination of biological mechanisms?



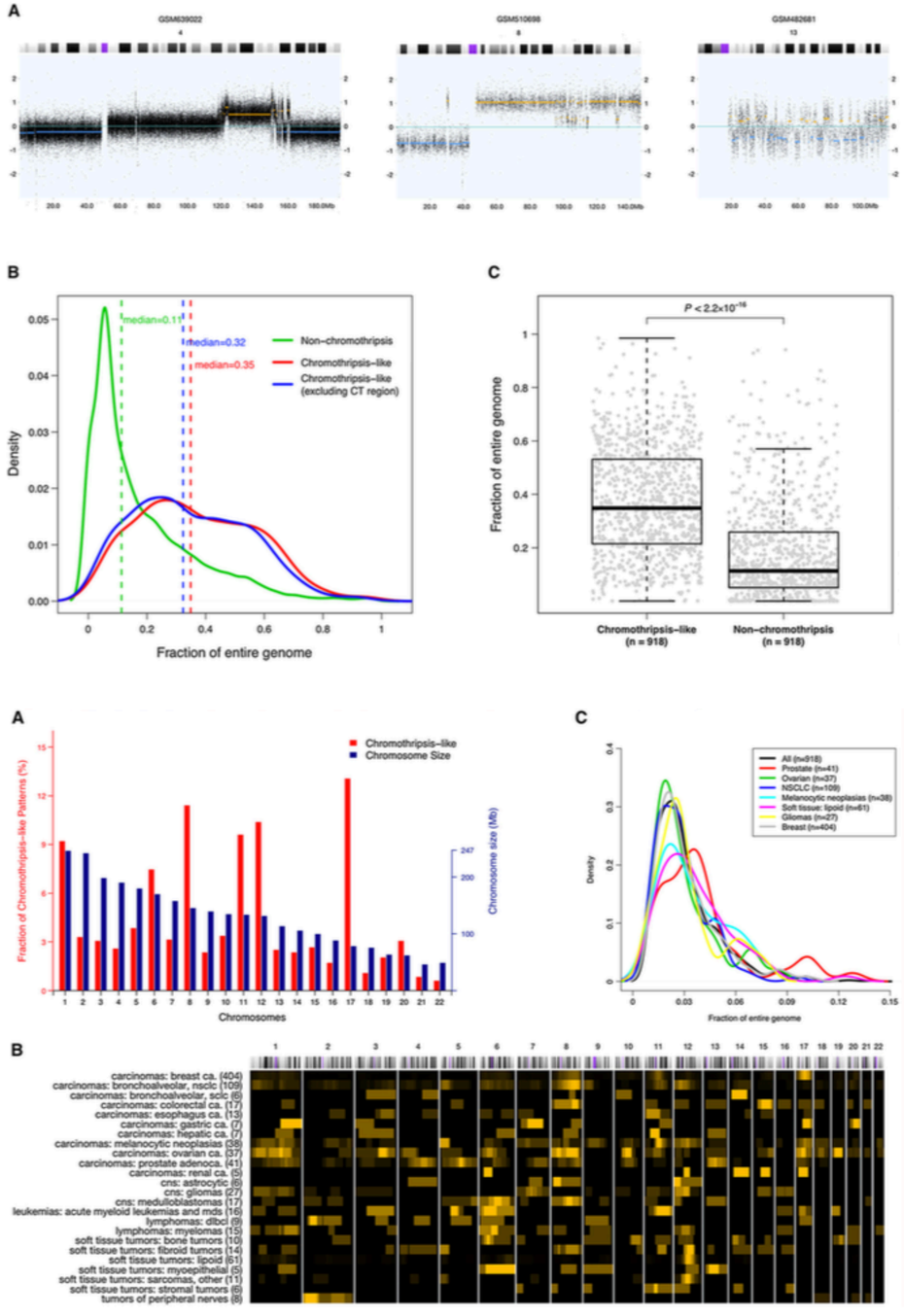


A genomic copy number histogram for malignant medulloblastomas, the most frquent type of pediatric brain tumors, displaying regions of genomic duplications and deletions. These can be decomposed into individual tumor profiles which segregate into several clusters of related mutation patterns with functional relevance and clinical correlation. From arraymap.org

Somatic Mutations In Cancer: Patterns II

Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens

- “Chromothripsis” is a recently described mechanism in which hundreds of genomic fragments are re-assembled in a single rearrangement event, which may oblivate gradual accumulation of mutations
- many cancer genome profiles haven been attributed to chromothripsis
- our analysis of >22’000 cancer genome profiles pointed to heterogeneity and relation to predominant overall genome instability in cancers with such chromothripsis-like genome patterns



Cai et al. BMC Genomics 2014, 15:82

Somatic Mutations In Cancer: Patterns III
Making the case for genomic classifications
Some related cancer entities show similar copy number profiles

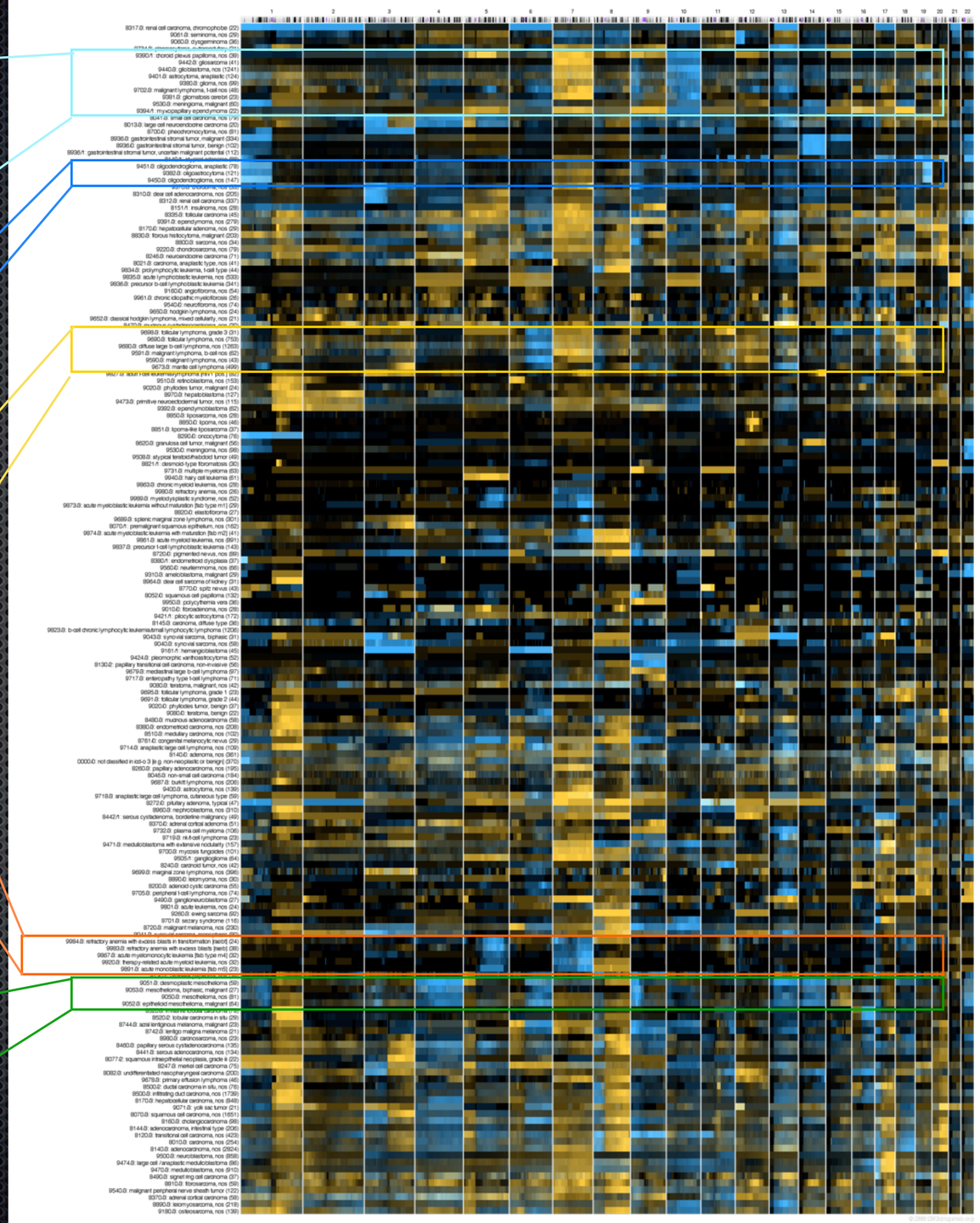
9390/1: choroid plexus papilloma, nos (39)
9442/3: gliosarcoma (41)
9440/3: glioblastoma, nos (1241)
9401/3: astrocytoma, anaplastic (124)
9380/3: glioma, nos (99)
9702/3: malignant lymphoma, t-cell nos (48)
9381/3: gliomatosis cerebri (23)
9530/3: meningioma, malignant (60)
9394/1: myxopapillary ependymoma (22)

9451/3: oligodendroglioma, anaplastic (78)
9382/3: oligoastrocytoma (121)
9450/3: oligodendroglioma, nos (147)

9698/3: follicular lymphoma, grade 3 (31)
9690/3: follicular lymphoma, nos (753)
9680/3: diffuse large b-cell lymphoma, nos (1263)
9591/3: malignant lymphoma, b-cell nos (62)
9590/3: malignant lymphoma, nos (43)
9673/3: mantle cell lymphoma (499)

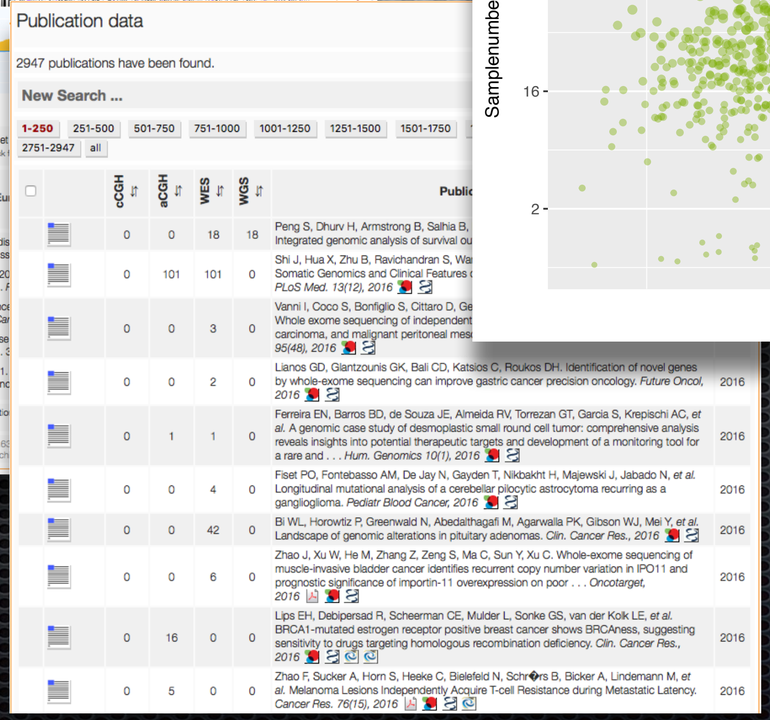
9984/3: refractory anemia with excess blasts in transformation [raebt] (24)
9983/3: refractory anemia with excess blasts [raeb] (38)
9867/3: acute myelomonocytic leukemia [fab type m4] (32)
9920/3: therapy-related acute myeloid leukemia, nos (32)
9891/3: acute monoblastic leukemia [fab m5] (23)

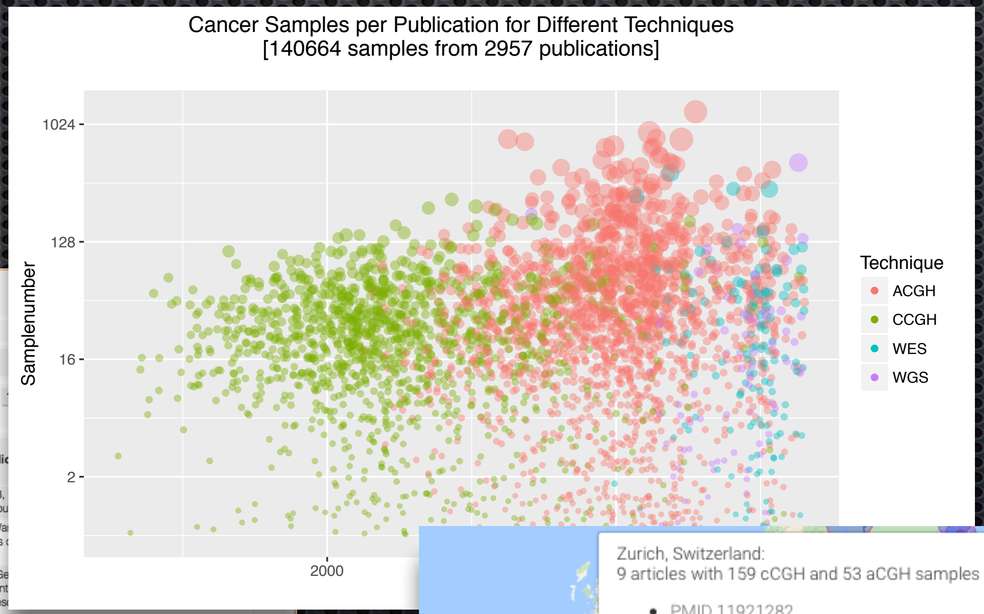
9051/3: desmoplastic mesothelioma (59)
9053/3: mesothelioma, biphasic, malignant (27)
9050/3: mesothelioma, nos (81)
9052/3: epithelioid mesothelioma, malignant (64)

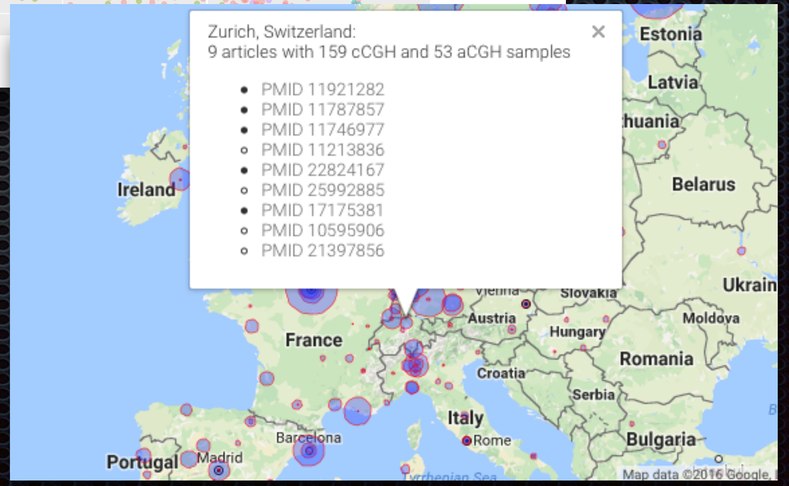


arrayMap

Progenetix: Cancer Genome Profiles, Article Metrics, Epistemology, **Resource Hub**



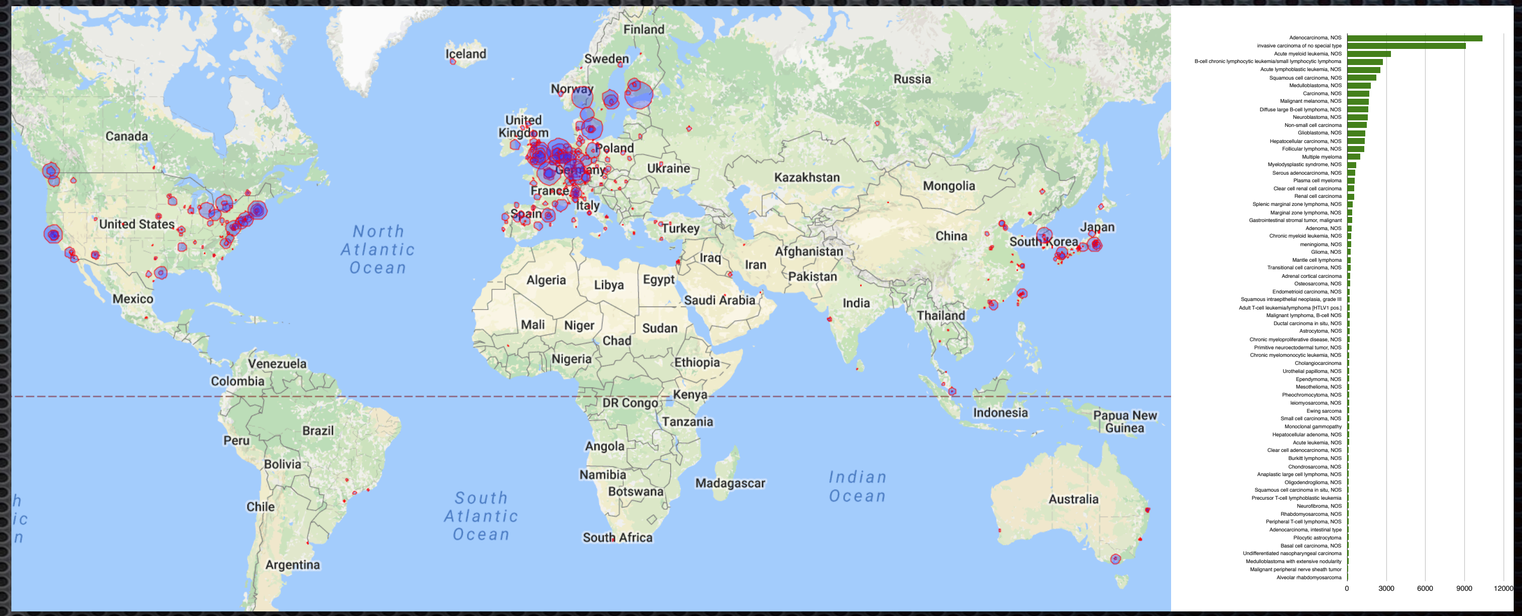




progenetix

Bias in Ascertainment / Background / Environment in Cancer Genome Studies

- the frequency of many genome variants depends on the genetic background
- cancer incidence & type can correlate to environmental factors
- geographic analysis can support interpretation and point to knowledge gaps

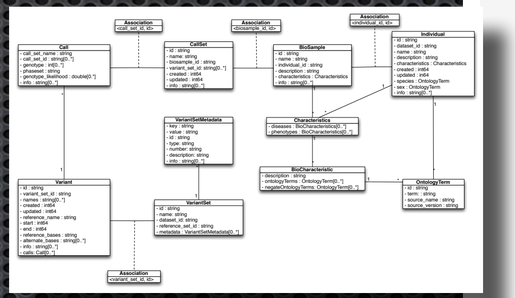


Geographic distribution of >140'000 cancer genome profiles reported in the literature. The numbers are derived from the 2947 publications registered in the Progenetix database.

progenetix

Developing the GA4GH **Metadata** Schema

- arrayMap for GA4GH
  - metadata schema development through implementation of arrayMap resource data
  - OntologyTerm objects for biodata
  - implementation w/ ontology services



Driving **Beacon** Development

- Beacon+
  - CNV/CNA as first type of structural variants
  - disease specific queries
  - quantitative reporting

Global Alliance for Genomics & Health

EMBL-EBI

arrayMap

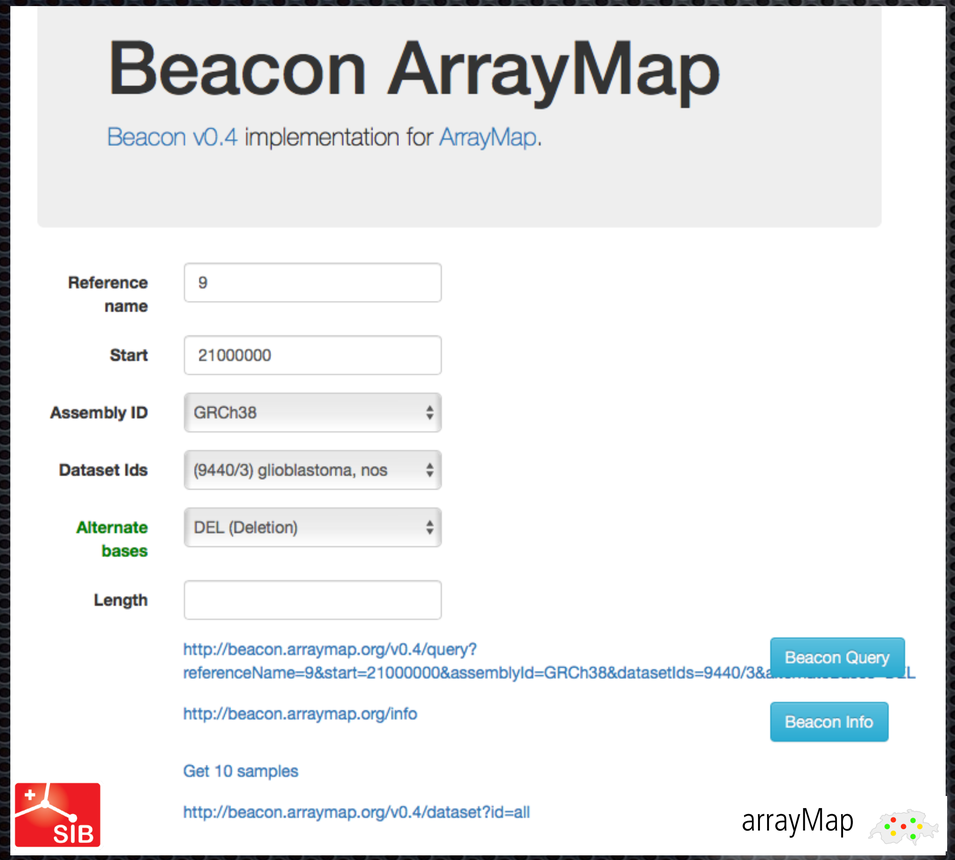
elixir

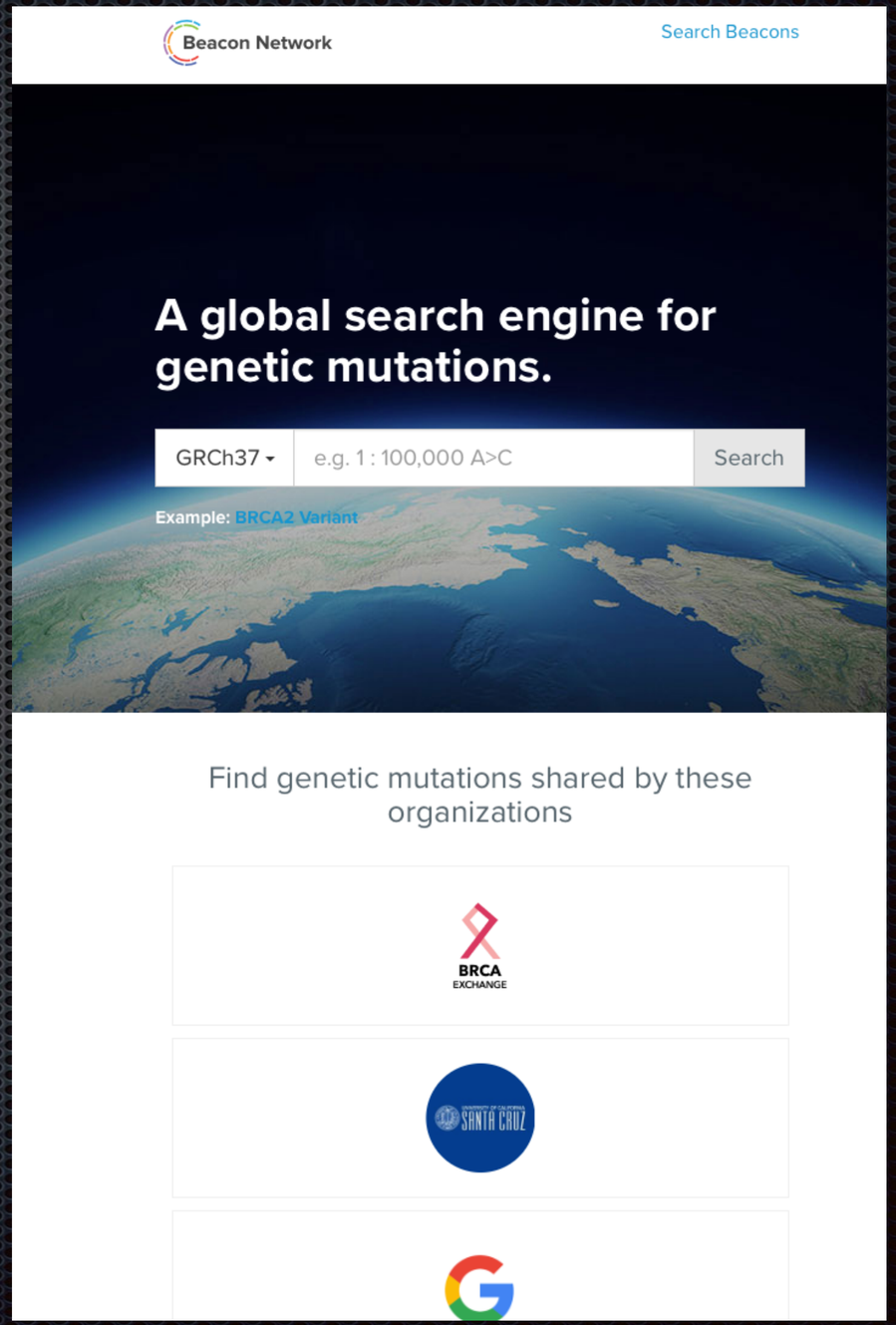
SIB Swiss Institute of Bioinformatics


```
{
        "_id" : ObjectId("58297ca32ca4591e5a0df054"),
        "id" : "AM_V_1778741",
        "variant_set_id" : "AM_VS_HG18",
        "reference_name" : "10"
        "start" : 579049,
        "end" : 17236099,
        "alternate_bases" : "DUP",
        "reference_bases" : ".",
        "info" : {
          "svlen":16657050,
          "cipos":[
              -1000,
              1000
          ],
          "ciend":[
              -1000,
              1000
          ]
        },
        "calls" : [
                {
                        "genotype" : [
                                ".",
                                "."
                        ],
                        "call_set_id" : "AM_CS_TCGA-61-1917-01A-01D-0648-01",
                        "info" : {
                                "segvalue" : 0.5491
                        }
                }
        ],
        "created" : ISODate("2016-11-14T08:33:58.202Z"),
        "updated" : ISODate("2016-11-14T08:33:58.202Z"),
}
```

Beacon+ Concept

- Implementation of cancer beacon prototype, backed by arrayMap data
- structural variation, quantitative queries, metadata



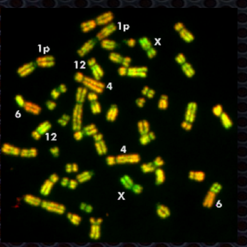


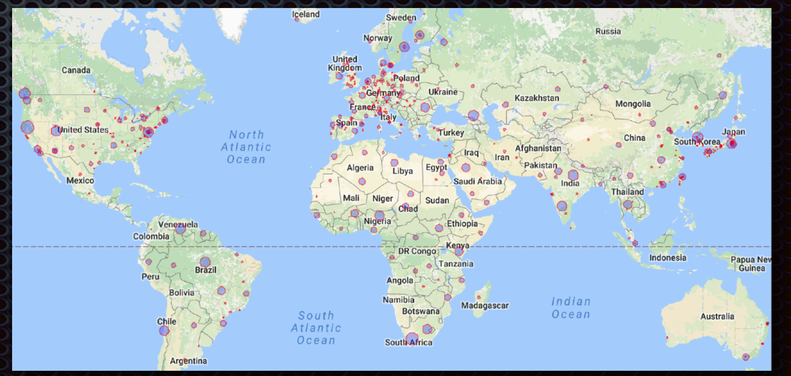
Heinz Stockinger, Séverine Duvaud & SIB Technology Group

University of Zurich[UZH]

Cancer Genome Data: Where Do We Need To Go

- balancing clinical medicine (**panel** sequencing of **actionable** cancer targets) and the necessary knowledge generation (**complete genomes**, multi"omes", genetic background)
- "genetic **awareness**" and regulatory **security**
- vastly increasing **data curation efforts** to make best use of existing data
- data **sharing frameworks** & **federated** analysis
- **everywhere in the world...**





BAUDISGROUP @ UZH

NI AI
MICHAEL BAUDIS
(HAOYANG CAI)
PAULA CARRIO CORDO
BO GAO
(LINDA GROB)
SAUMYA GUPTA)
(ROMAN HILLJE
(NITIN KUMAR)
(ALESSIO MILANESE)

SIB

HEINZ STOCKINGER
SÉVERINE DUVAUD
VASSILIOS IOANNIDIS
DANIEL TEIXEIRA

THOMAS EGGERMANN
ROSA NOGUERA
REINER SIEBERT
CAIUS SOLOVAN

University of Zurich[UZH]

SIB

elixir

Global Alliance for Genomics & Health

GA4GH DWG + CWG

JACQUI BECKMANN
ANTHONY BROOKES
MELANIE COURTOT
MARK DIEKHANS
MELISSA HAENDEL
DAVID HAUSSLER
SARAH HUNT
STEPHEN KEENAN
SUZY LEWIS
DAVID LLOYD
MICHAEL MILLER
HELEN PARKINSON
GUNNAR RÄTSCH
ELEANOR STANLEY
DAVID STEINBERG
JULIA WILSON

ELIXIR & CRG

JORDI RAMBLA DE ARGILA
S. DE LA TORRE PERNAS
SUSANNA REPO
SERENA SCOLLEN

Prof. Dr. Michael Baudis
Department of Molecular Life Sciences
University of Zurich
SIB I Swiss Institute of Bioinformatics
Winterthurerstrasse 190
CH-8057 Zurich
Switzerland

arraymap.org

progenetix.org

sib.swiss/baudis-michael

imls.uzh.ch/en/research/baudis

Cai H, Gupta S, Rath P, Ai N, Baudis M. arrayMap 2014: an updated cancer genome resource. Nucleic Acids Res. 2015 Jan;43 (Database issue).

Cai, H., Kumar, N., & Baudis, M. 2012. arrayMap: A Reference Resource for Genomic Copy Number Imbalances in Human Malignancies. PLoS One 7(5), e36944.

Baudis, M. 2007. Genomic imbalances in 5918 malignant epithelial tumors: An explorative meta-analysis of chromosomal CGH data. BMC Cancer 7:226.

Baudis, M, and ML Cleary. 2001. Progenetix.net: an online repository for molecular cytogenetic aberration data. Bioinformatics 12, no. 17: 1228-1229.