
Institute of
Medical Genetics and
Applied Genomics
UNIVERSITÄTS
KLINIKUM
TÜBINGEN
Welcome to the Pan-Cancer Atlas
German Demidov
german.demidov@med.uni-tuebingen.de
SPb Bioinformatics Institute
Summer School 2018
23-28th June 2018
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN
CRG
Centre
for Genomic
Regulation
upf.
Universitat
Pompeu Fabra
Barcelona

- Disclaimer: all rights for the pictures/plots belong to the authors and/or Cell Press.
- For any further information:

https://www.cell.com/pb-assets/consortium/pancanceratlas/pancani3/index.html

Who am I to lecture you?

CRG Centre for Genomic Regulation

upf. Universitat Pompeu Fabra Barcelona


Brandschutz-
zone

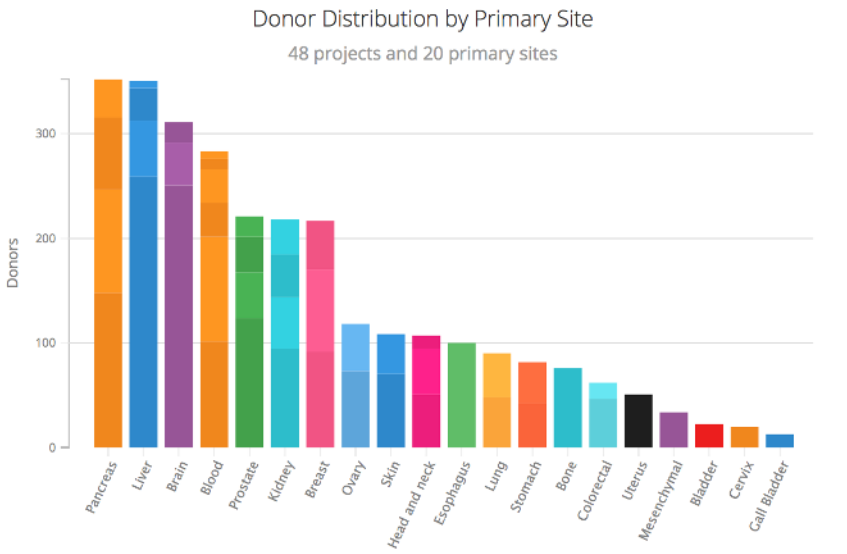
WG8: Germline Cancer Genome

PCAWG - PANCANCER ANALYSIS OF WHOLE GENOMES

The Pancancer Analysis of Whole Genomes (PCAWG) study is an international collaboration to identify common patterns of mutation in more than 2,800 cancer whole genomes from the International Cancer Genome Consortium. Building upon previous work which examined cancer coding regions (Cancer Genome Atlas Research Network, The Cancer Genome Atlas Pan-Cancer analysis project, Nat. Genet. 2013 45:1113), this project is exploring the nature and consequences of somatic and germline variations in both coding and non-coding regions, with specific emphasis on cis-regulatory sites, non-coding RNAs, and large-scale structural alterations.

In order to facilitate the comparison among diverse tumor types, all tumor and matched normal genomes have been subjected to a uniform set of alignment and variant calling algorithms, and must pass a rigorous set of qualilty control tests. The research activities are coordinated by a series of working groups comprising more than 700 scientists and covering the following themes:

1. Novel somatic mutation calling methods
2. Analysis of mutations in regulatory regions
3. Integration of the transcriptome and genome
4. Integration of the epigenome and genome
5. Consequences of somatic mutations on pathway and network activity
6. Patterns of structural variations, signatures, genomic correlations, retrotransposons and mobile elements
7. Mutation signatures and processes
8. Germline cancer genome
9. Inferring driver mutations and identifying cancer genes and pathways
10. Translating cancer genomes to the clinic
11. Evolution and heterogeneity
12. Portals, visualization and software infrastructure
13. Molecular subtypes and classification



2,834 Donors **72,888** Files **802.96 TB**

Data Type	# Donors	# Files	Format	Size
SGV	2,834	8,865	VCF	539.37 GB
StGV	2,834	5,908	VCF	7.58 GB
Aligned Reads	2,834	11,139	BAM	801.09 TB
Simple Somatic Mutations	2,834	26,239	VCF	198.09 GB
Copy Number Somatic Mutations	2,834	5,911	VCF	138.14 MB
Structural Somatic Mutations	2,834	14,743	VCF	1.70 GB

CRG Centre for Genomic Regulation

upf. Universitat Pompeu Fabra Barcelona

The goals of today's lecture

- Познакомить вас с крупнейшим на сегодня исследованием в геномике/ эпигеномике рака (WES-based и другие платформы) с небольшими отступлениями
- Дать представление о том, как работает большой консорциум (кто знает, не придется ли вам работать с чем-то подобным?)

Stats

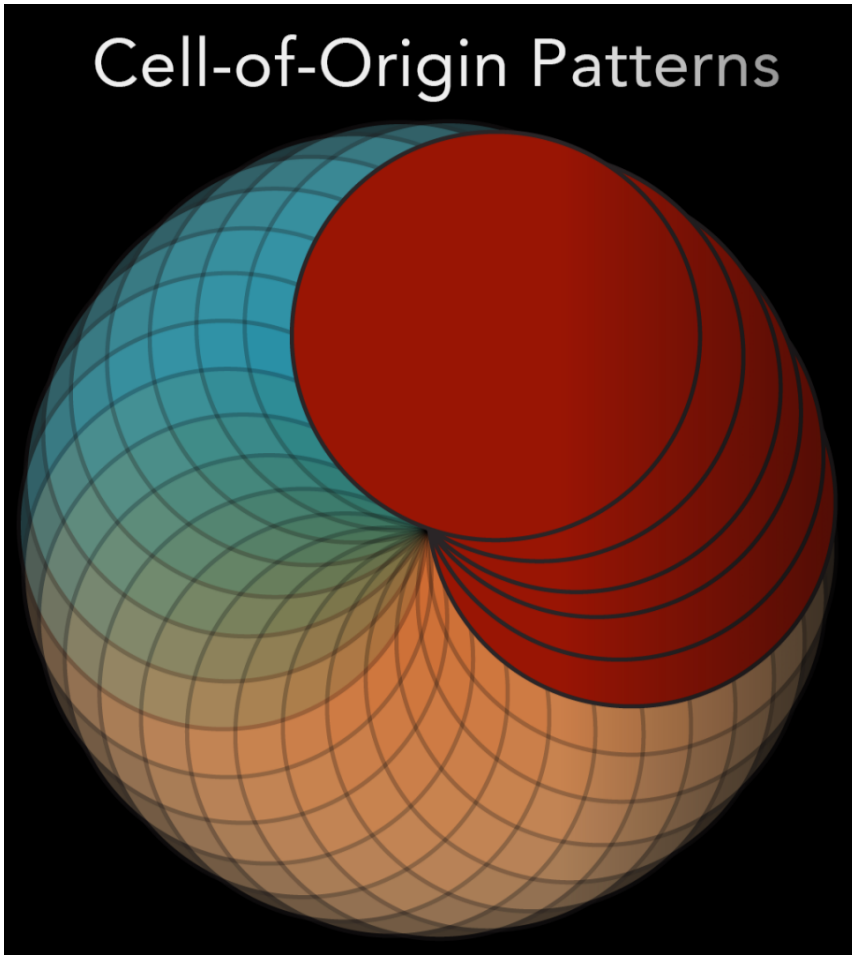
- 11.286 tumors
- 9.759 had complete data for 4 platforms: aneuploidy, DNA methylation, mRNA and miRNA. Reverse phase protein array data were available for a subset of samples (7.858). ~10.300 WES data
- from 33 of the most prevalent forms of cancer
- 27 papers in Cell

Types of cancer (Germline risk paper)

C

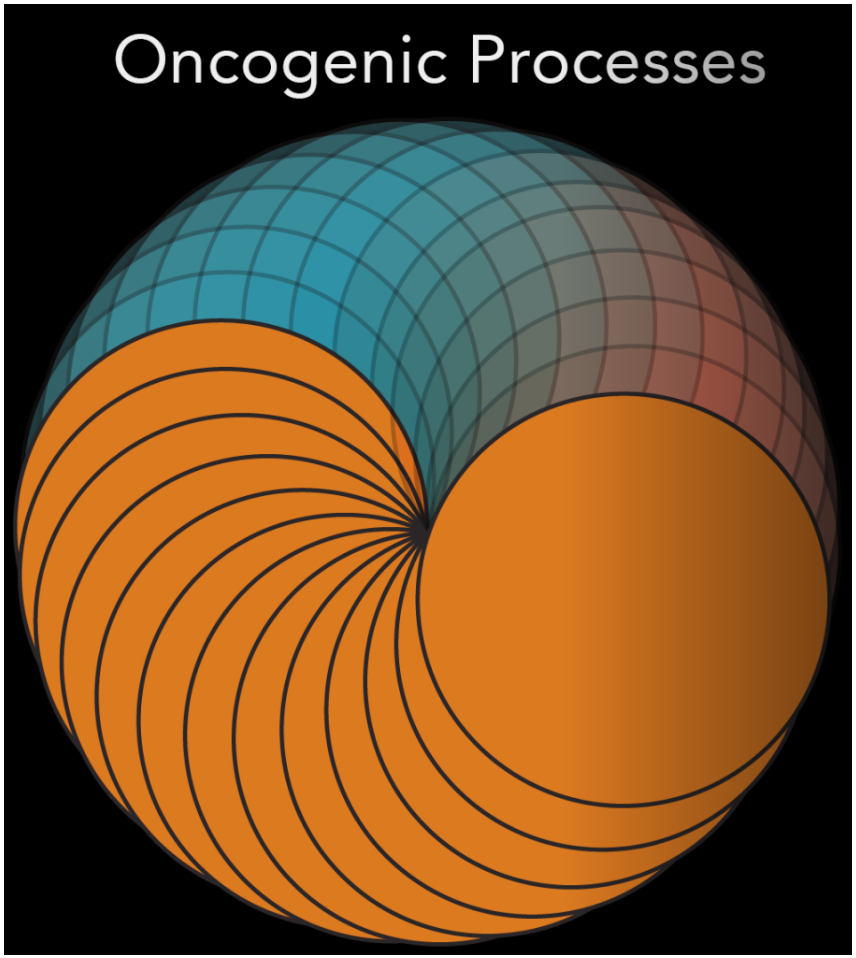
	Cancer	Sample size	Female ratio	Age at onset
ACC	Adrenocortical Carcinoma	92	65%	47.2 +/- 16.3
BLCA	Bladder Urothelial Carcinoma	412	26%	68.1 +/- 10.6
BRCA	Breast Invasive Carcinoma	1076	99%	58.5 +/- 13.2
CESC	Cervical Squamous Cell Carcinoma & Endocervical Adenocarcinoma	305	100%	48.2 +/- 13.8
CHOL	Cholangiocarcinoma	45	56%	63.6 +/- 12.2
COAD	Colon Adenocarcinoma	419	48%	66.7 +/- 13.2
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	41	54%	56.5 +/- 14.3
ESCA	Esophageal Carcinoma	184	15%	62.4 +/- 11.9
GBM	Glioblastoma Multiforme	393	37%	59.8 +/- 13.6
HNSC	Head and Neck Squamous Cell Carcinoma	526	27%	60.9 +/- 11.9
KICH	Kidney Chromophobe	66	41%	51.5 +/- 14.3
KIRC	Kidney Renal Clear Cell Carcinoma	387	36%	60.1 +/- 12.2
KIRP	Kidney Renal Papillary Cell Carcinoma	289	27%	61.4 +/- 12.1
LAML	Acute Myeloid Leukemia	142	46%	56.2 +/- 15.4
LGG	Brain Lower Grade Glioma	515	45%	42.9 +/- 13.4
LIHC	Liver Hepatocellular Carcinoma	375	32%	59.4 +/- 13.5
LUAD	Lung Adenocarcinoma	518	54%	65.3 +/- 10
LUSC	Lung Squamous Cell Carcinoma	499	26%	67.3 +/- 8.6
MESO	Mesothelioma	82	18%	63 +/- 9.9
OV	Ovarian Serous Cystadenocarcinoma	412	100%	59.6 +/- 11.6
PAAD	Pancreatic Adenocarcinoma	185	45%	64.9 +/- 11.1
PCPG	Pheochromocytoma and Paraganglioma	179	56%	47.3 +/- 15.1
PRAD	Prostate Adenocarcinoma	498	0%	61 +/- 6.8
READ	Rectum Adenocarcinoma	145	47%	63.7 +/- 12.2
SARC	Sarcoma	255	54%	60.7 +/- 14.8
SKCM	Skin Cutaneous Melanoma	470	38%	58.2 +/- 15.7
STAD	Stomach Adenocarcinoma	443	36%	65.7 +/- 10.8
TGCT	Testicular Germ Cell Tumors	134	0%	32 +/- 9.3
THCA	Thyroid Carcinoma	499	73%	47.3 +/- 15.8
THYM	Thymoma	123	48%	58.3 +/- 13
UCEC	Uterine Corpus Endometrial Carcinoma	543	100%	64 +/- 11.2
UCS	Uterine Carcinosarcoma	57	100%	69.7 +/- 9.3
UVM	Uveal Melanoma	80	44%	61.6 +/- 13.9
All	**All 33 Cancers Combined**	**10389**	**52%**	**59.2 +/- 14.4**

Are cells-of-origin determining the tumor?



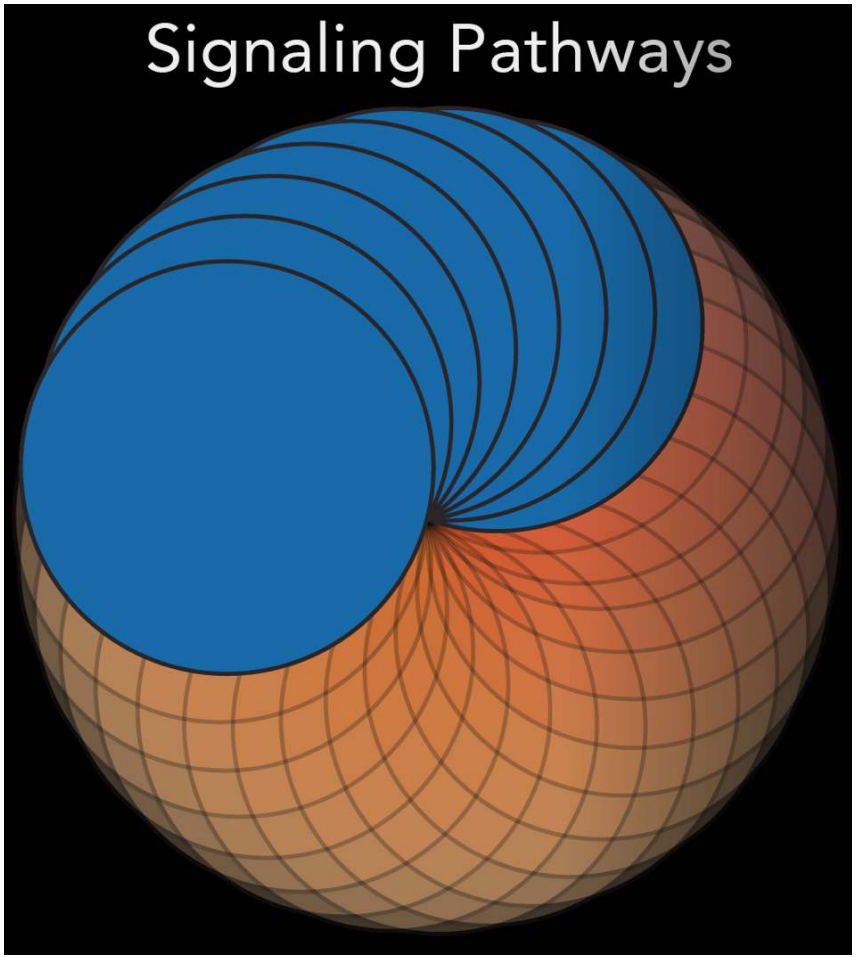
- The Pan-Cancer Atlas reclassifies human tumor types based on molecular similarity
- The cell of origin influences but does not fully determine tumor classification
- => future clinical trial design and interpretation

Oncogenic Processes



- Panoramic view of the oncogenic processes that contribute to human cancer
- Germline genetic variants and somatic mutations collaborate in cancer progression
- The influence of mutation on cell signaling and immune cell composition

Pathways in Cancer –
NOT COVERED



- In its comprehensive analysis of tumor signaling pathways, the Pan-Cancer Atlas reveals patterns of vulnerabilities that will aid in the development of personalized treatments and new combination therapies.

Terms

- **Missense mutation** = меняет кодон и аминокислоту (несинонимично)
- **Non-sense mutation** = обрезанный/ неполный белок
- **Squamous cells** ~ эпителий
- **Adeno-** = железа

Main types of cancer

- Epithelial cells (cells originating in the endodermal, mesodermal or ectodermal germ layer during embryogenesis) ⇨ **carcinoma**
- Non-hematopoietic mesenchymal cells ⇨ **sarcoma**
- Hematopoietic cells: Bone marrow-derived cells that normally mature in the bloodstream ⇨ **leukemia**, bone marrow-derived cells that normally mature in the lymphatic system ⇨ **lymphoma**
- Melanocytes ⇨ **melanoma**

Other types?

CNS	Neuroepithelial (brain tumors, spinal tumors)	Glioma	Astrocyte	Astrocytoma (Pilocytic astrocytoma · Pleomorphic xanthoastrocytoma · Subependymal giant cell astrocytoma · Fibrillary astrocytoma · Anaplastic astrocytoma · Glioblastoma multiforme)
			Oligodendrocyte	Oligodendroglioma
			Ependyma	Ependymoma · Subependymoma
			Choroid plexus	Choroid plexus tumor (Choroid plexus papilloma · Choroid plexus carcinoma)
			Multiple/unknown	Oligoastrocytoma · Gliomatosis cerebri · Gliosarcoma
		Mature neuron		Ganglioneuroma: Ganglioglioma · Retinoblastoma · Neurocytoma · Dysembryoplastic neuroepithelial tumour · Lhermitte–Duclos disease
		PNET		Neuroblastoma (Esthesioneuroblastoma · Ganglioneuroblastoma) · Medulloblastoma · Atypical teratoid rhabdoid tumor
		Primitive		Medulloepithelioma
	Meningeal (Meninges)			Meningioma · Hemangiopericytoma
	Hematopoietic			Primary central nervous system lymphoma

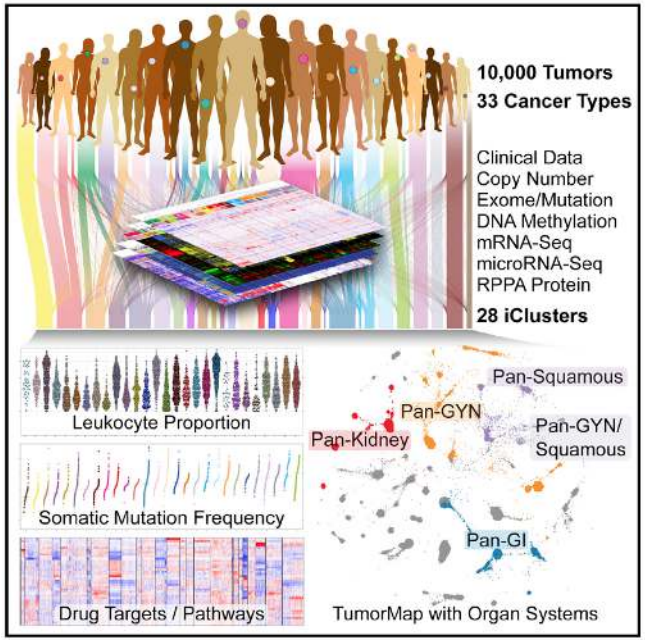
Are cells-of-origin determining the tumor?

Article

Cell

Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer

Graphical Abstract



Authors

Katherine A. Hoadley, Christina Yau, Toshinori Hinoue, ..., Joshua M. Stuart, Christopher C. Benz, Peter W. Laird

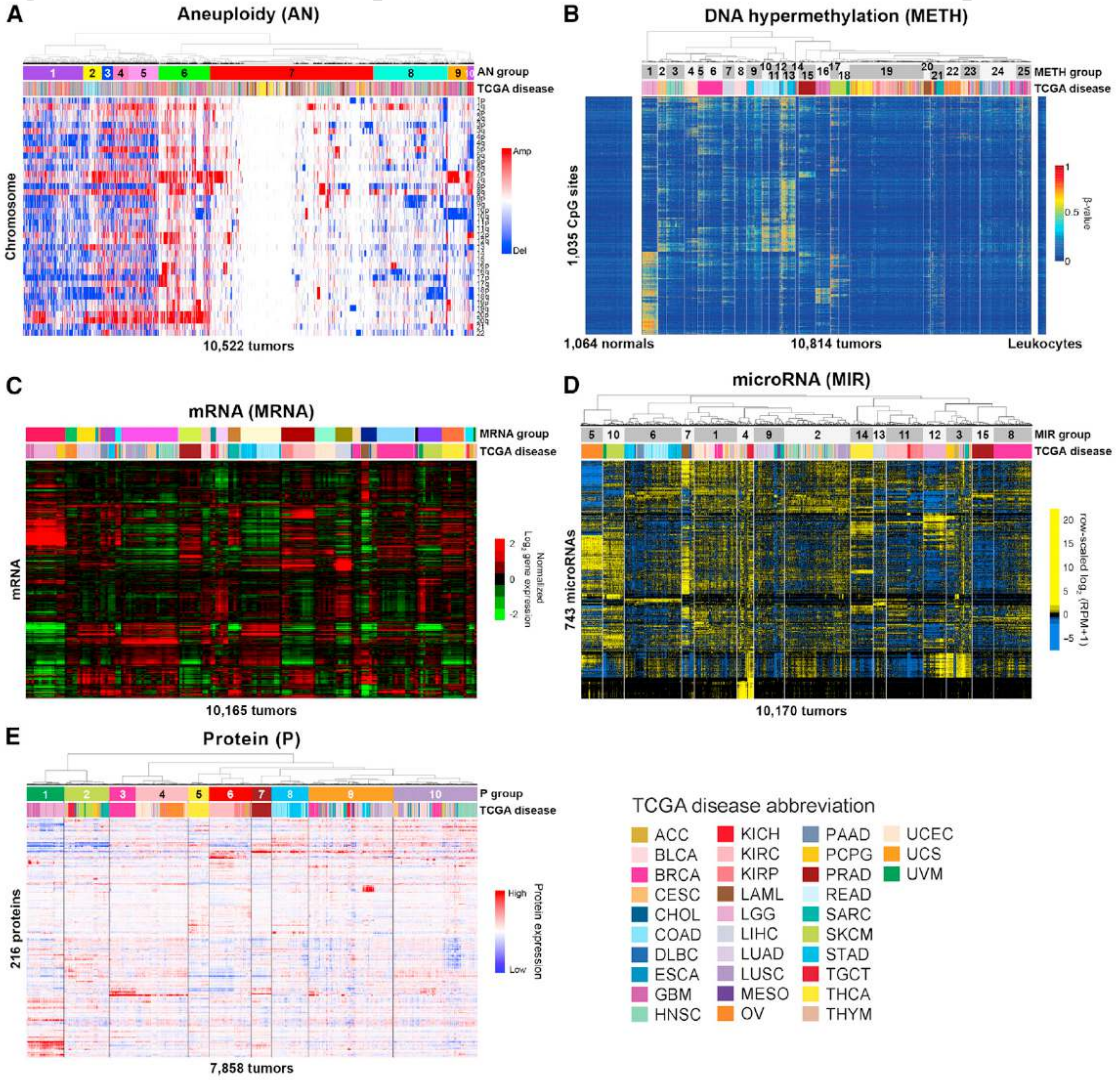
Correspondence

hoadley@med.unc.edu (K.A.H.), peter.laird@vai.org (P.W.L.)

In Brief

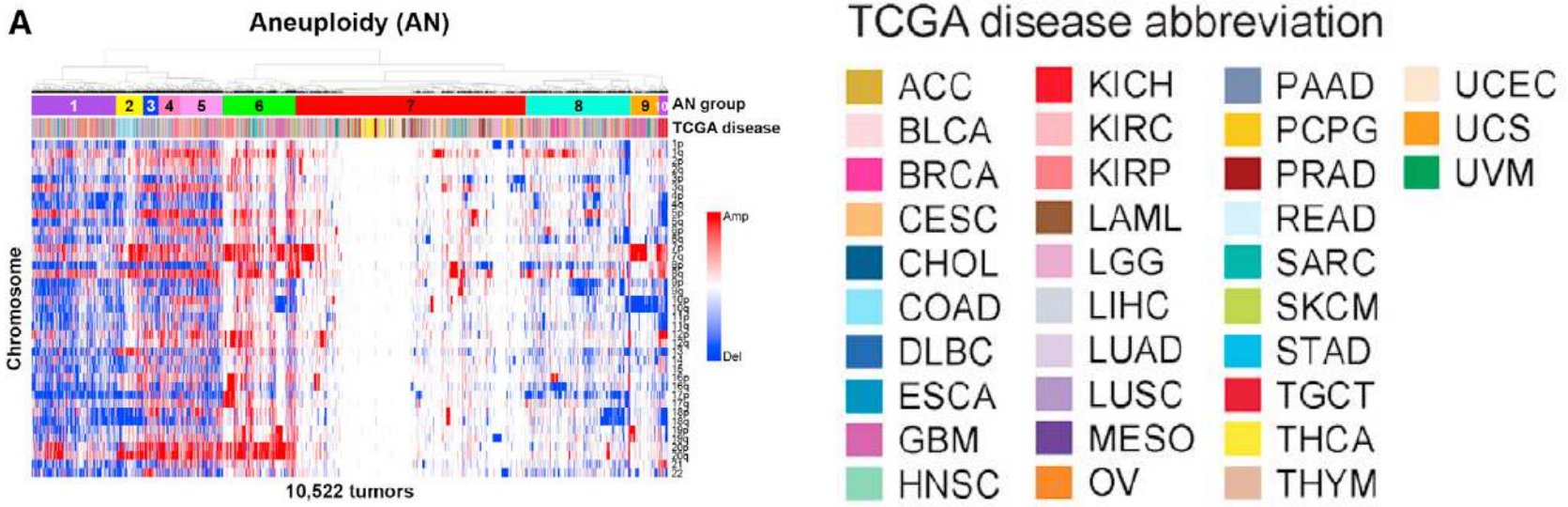
Comprehensive, integrated molecular analysis identifies molecular relationships across a large diverse set of human cancers, suggesting future directions for exploring clinical actionability in cancer treatment.

1D analysis: 10-25 groups (>40 samples)



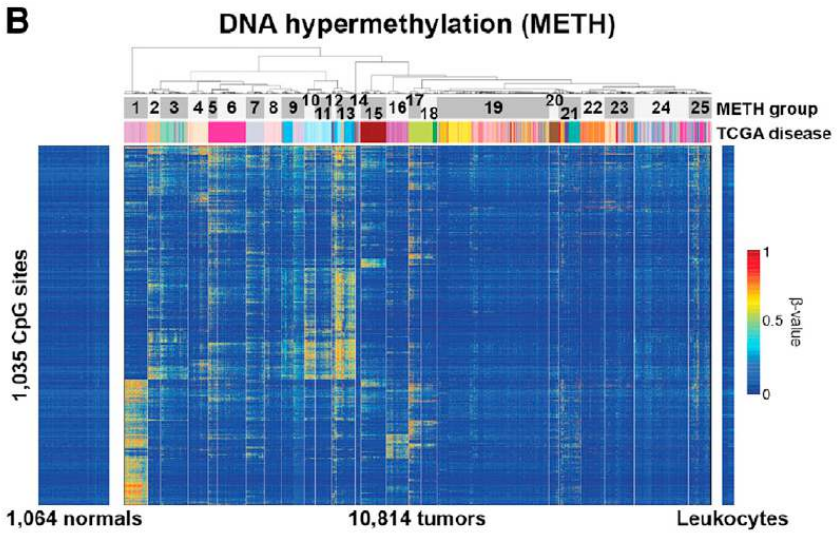
1D analysis: Aneuploidy

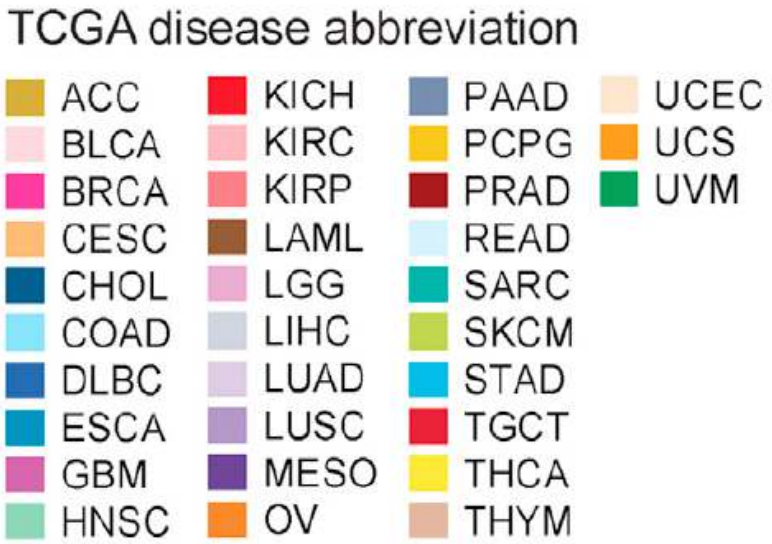
Aneuploidy classifications were weakly consistent with other classifications, in part due to low numbers of arm-level copy-number events in one-third of the tumors. Samples were split mainly by those with few alterations (AN7), those with moderate alterations (AN6,8-10), and those with many alterations (AN1-5).



1D analysis: DNA methylation

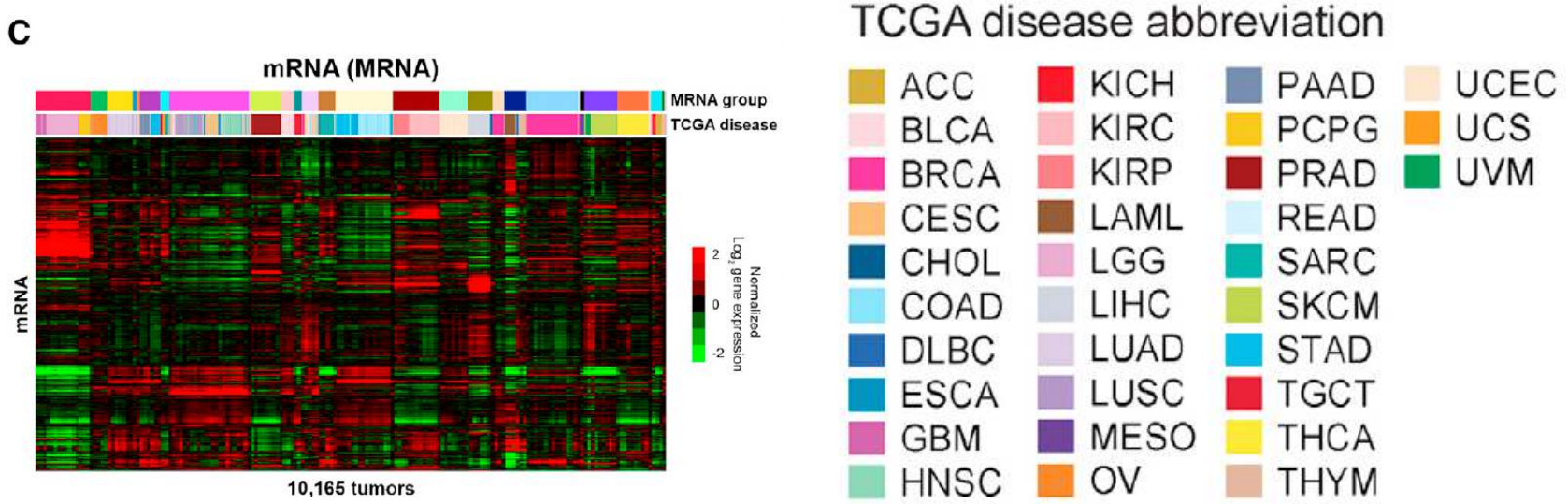
Despite the exclusion of loci known to be involved in tissue-specific DNA Methylation, tumors originating from the same organ often aggregated by cancer-type-specific hypermethylation => cancer-associated DNA Hypermethylation is influenced by pre-existing cell-type-specific chromatin marks or transcriptional programs, and not just by cell-type-specific DNA methylation patterns.





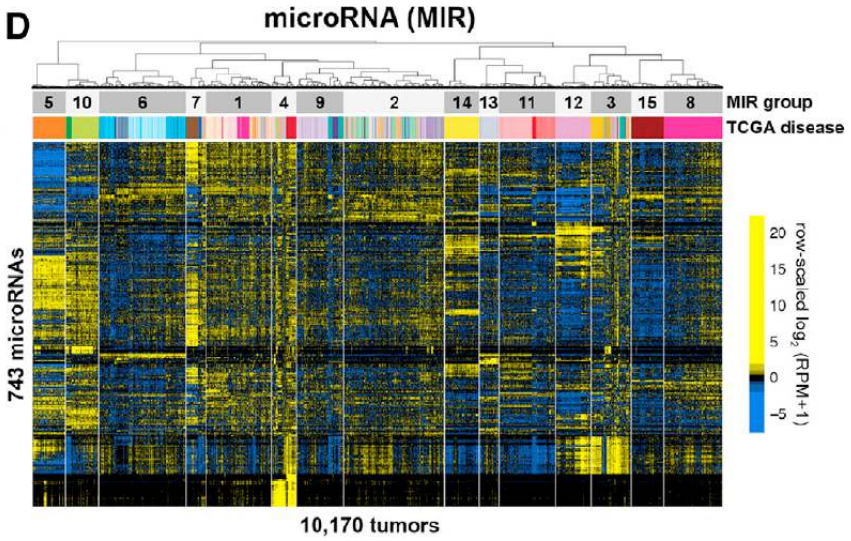
1D analysis: mRNA

While tumor type was a driving feature for many groups, several groups were comprised of tumors from different organ types. Samples with squamous morphology components grouped together. Similarly, tumors with tissue or organ similarities or proximity also grouped together.



1D analysis: miRNA

While six groups contained only a single cancer type, the remaining nine groups each represented a mix of cancer types.

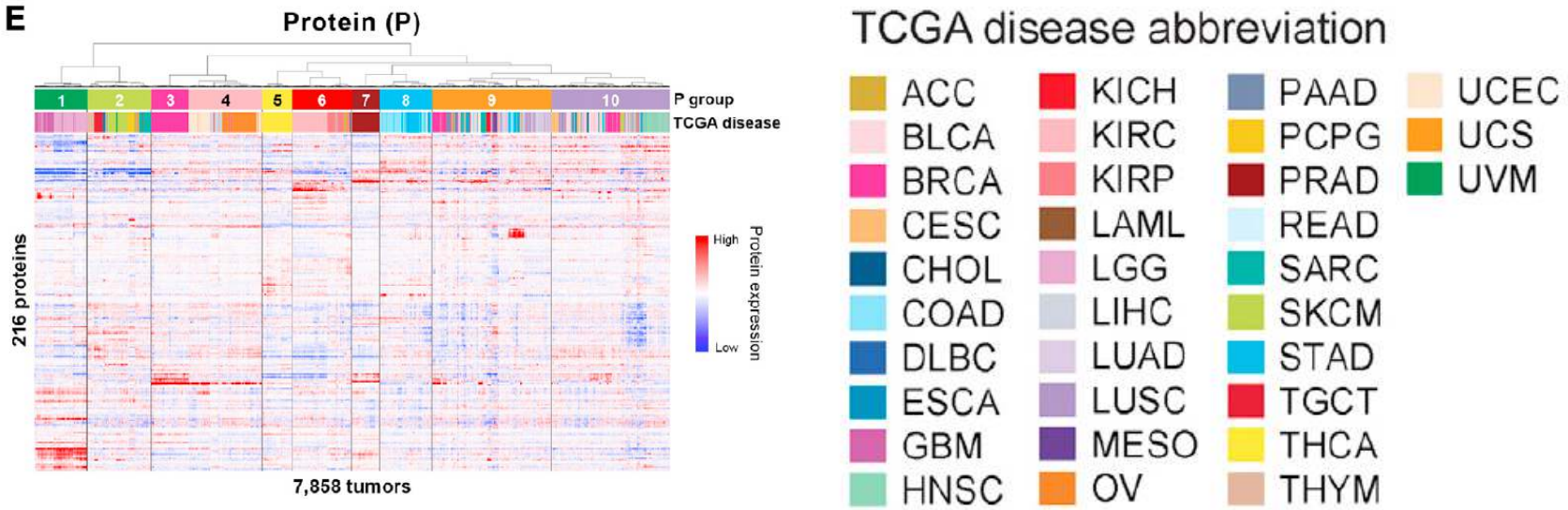


TCGA disease abbreviation

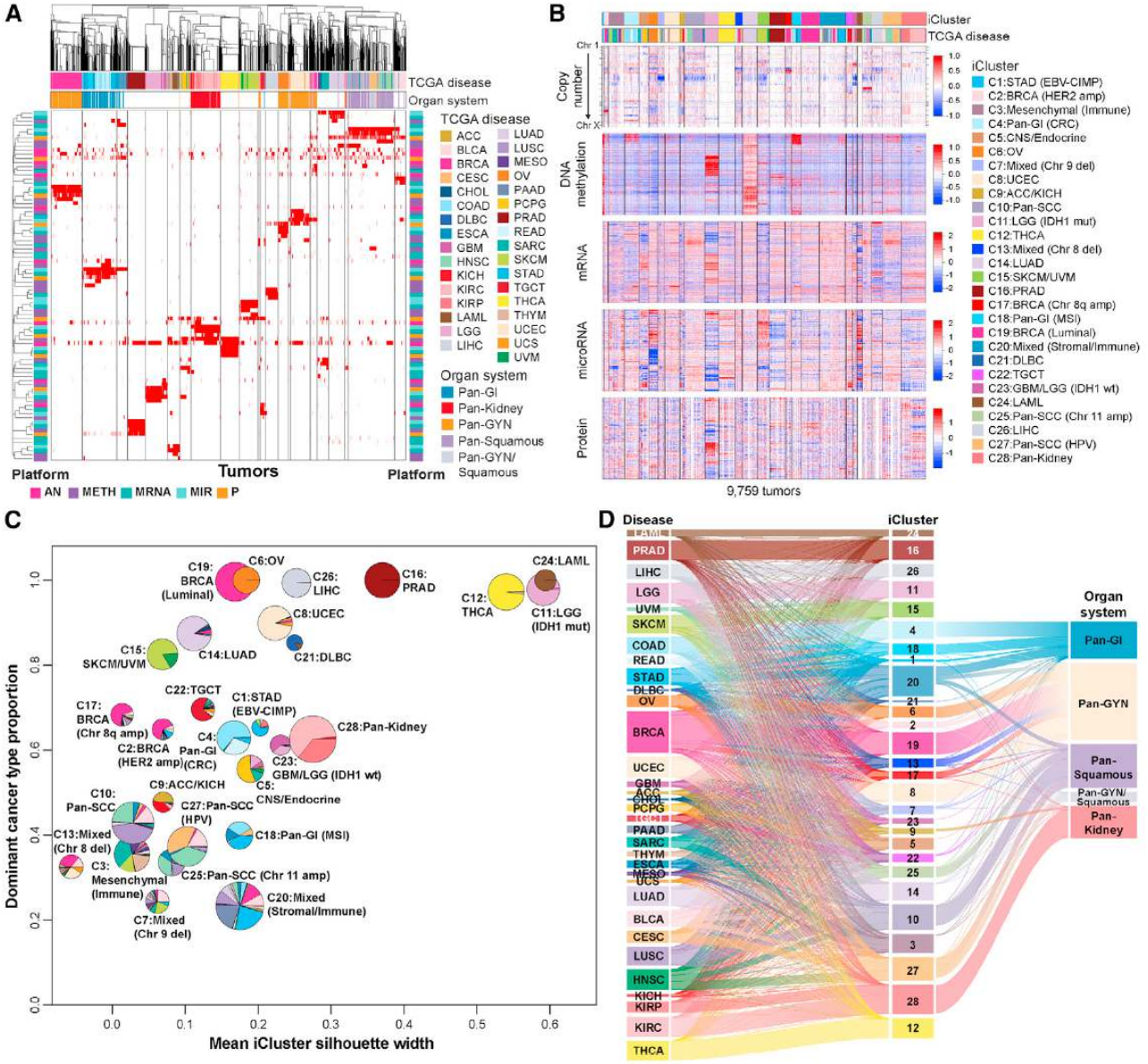
ACC	KICH	PAAD	UCEC
BLCA	KIRC	PCPG	UCS
BRCA	KIRP	PRAD	UVM
CESC	LAML	READ	
CHOL	LGG	SARC	
COAD	LIHC	SKCM	
DLBC	LUAD	STAD	
ESCA	LUSC	TGCT	
GBM	MESO	THCA	
HNSC	OV	THYM	

1D analysis: Protein Arrays

P1 and P2 were distinguished from the remaining 8 groups, largely corresponding to mesenchymal like tumor types with high EMT (epithelial-mesenchymal transition) signatures.
Similar to the other individual data platforms,
samples from related organ systems grouped together



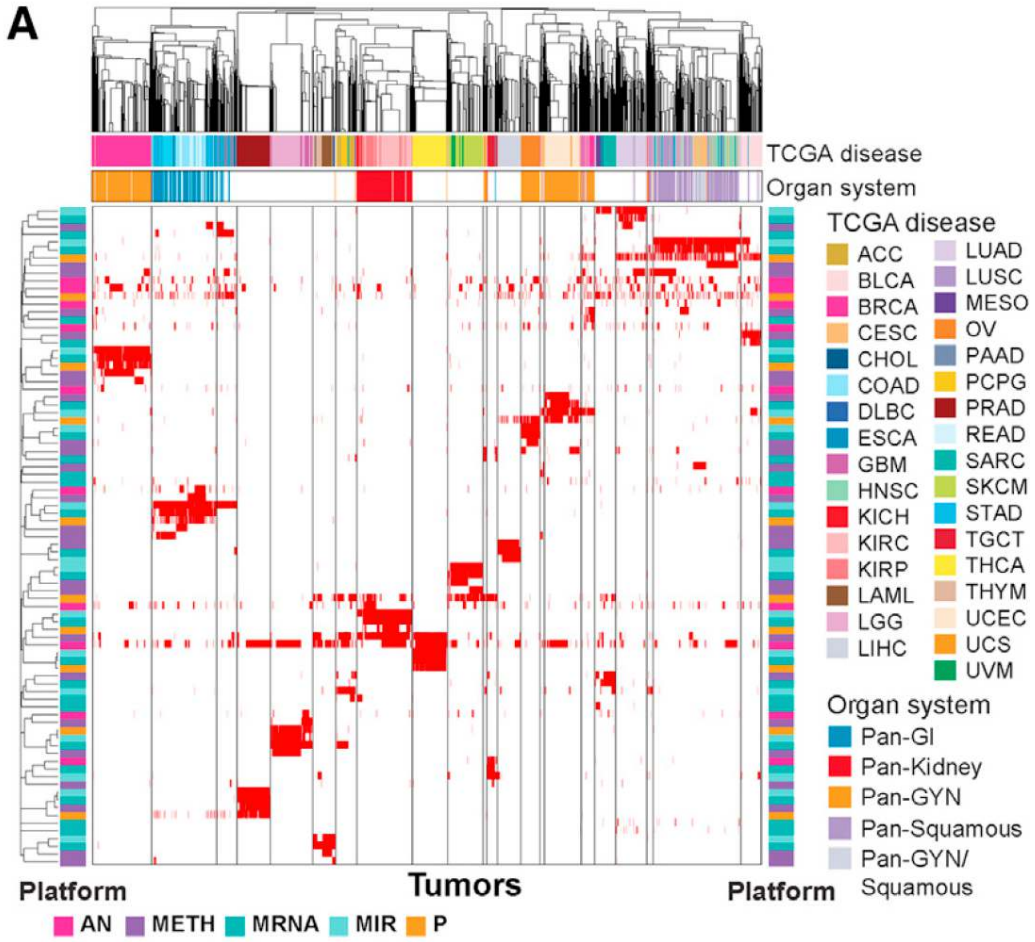
Cross Platform Analysis



Centre for Genomic Regulation

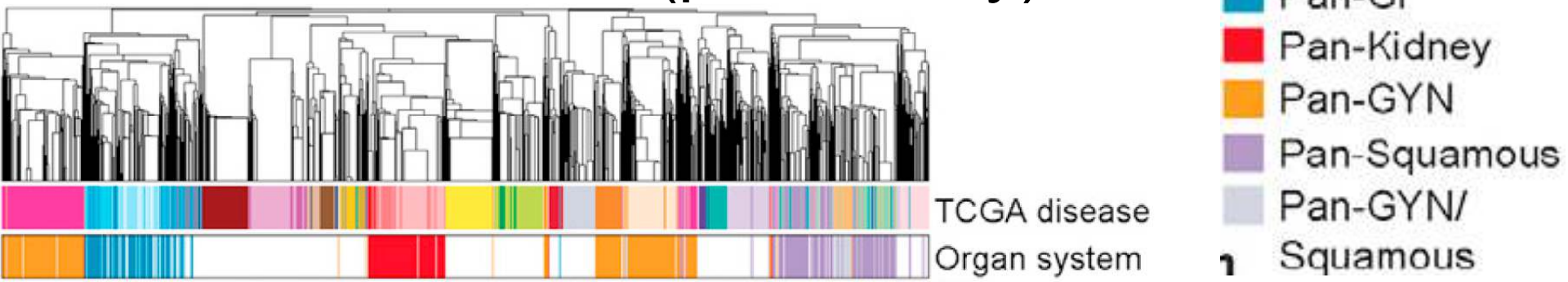
Universitat Pompeu Fabra Barcelona

Clustering of cluster assignments (COCA)



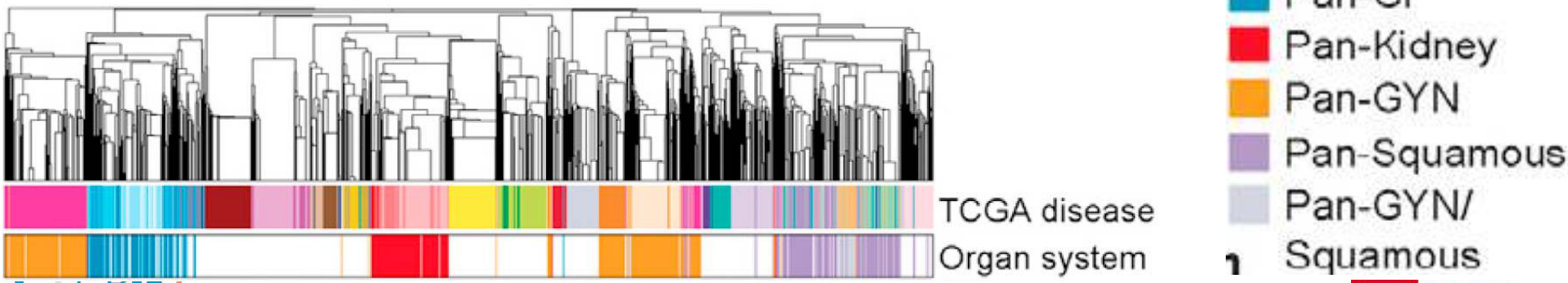
Clustering of cluster assignments (COCA)

- Many samples similarly grouped together by multiple platform-specific cluster memberships, both in groups that were defined by a single tumor type and in tumor types that co-clustered, such as KIRC and KIRP (pan-kidney).



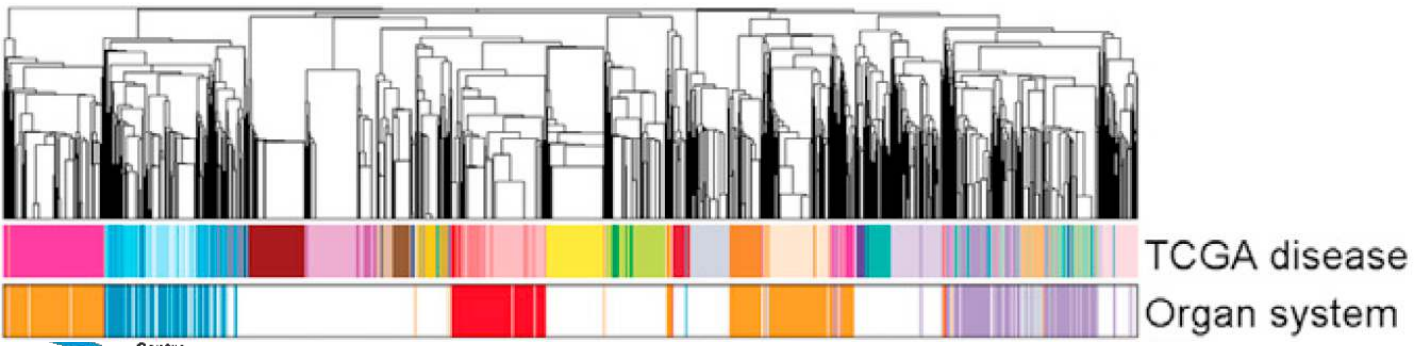
Clustering of cluster assignments (COCA)

- Gastrointestinal tumors (COAD, READ, STAD, and ESCA adenocarcinomas)
- co-clustered in the mRNA, miRNA, and RPPA platforms
- several distinct DNA methylation clusters.



Clustering of cluster assignments (COCA)

- Squamous histology cancers (LUSC, HNSC, CESC, ESCA, and BLCA) were similarly classified by the miRNA, mRNA and RPPA data
- divided by the aneuploidy and DNA methylation data.



Organ system
Pan-GI
Pan-Kidney
Pan-GYN
Pan-Squamous
Pan-GYN/
Squamous

Centre for Genomic Regulation

upf. Pompeu Fabra Barcelona

Clustering of cluster assignments (COCA)

- Within pan-gyn cancers (BRCA, ovarian serous cystadenocarcinoma, Uterine Corpus Endometrial Carcinoma, and Uterine Carcinosarcoma),
- RPPA data suggested that OV and UCEC shared similarities at the protein level
- miRNA, mRNA, and DNA methylation data were grouped by their organ sites.

Clustering of cluster assignments (COCA)

- Also of note, 13% of BRCA formed a subtype distinct from the majority of other BRCA, influenced by the mRNA and DNA methylation platforms.

iCluster

Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* 25, 2906-2912.

Package 'iCluster'

February 20, 2015

Title Integrative clustering of multiple genomic data types

Version 2.1.0

Date 2012-05-01

Depends R (>= 2.15.0), lattice, caTools, gdata, gtools, gplots, parallel

Author Ronglai Shen

Maintainer Ronglai Shen <shenr@mskcc.org>

Description Integrative clustering of multiple genomic data types using a joint latent variable model.

LazyData yes

License GPL (>= 2)

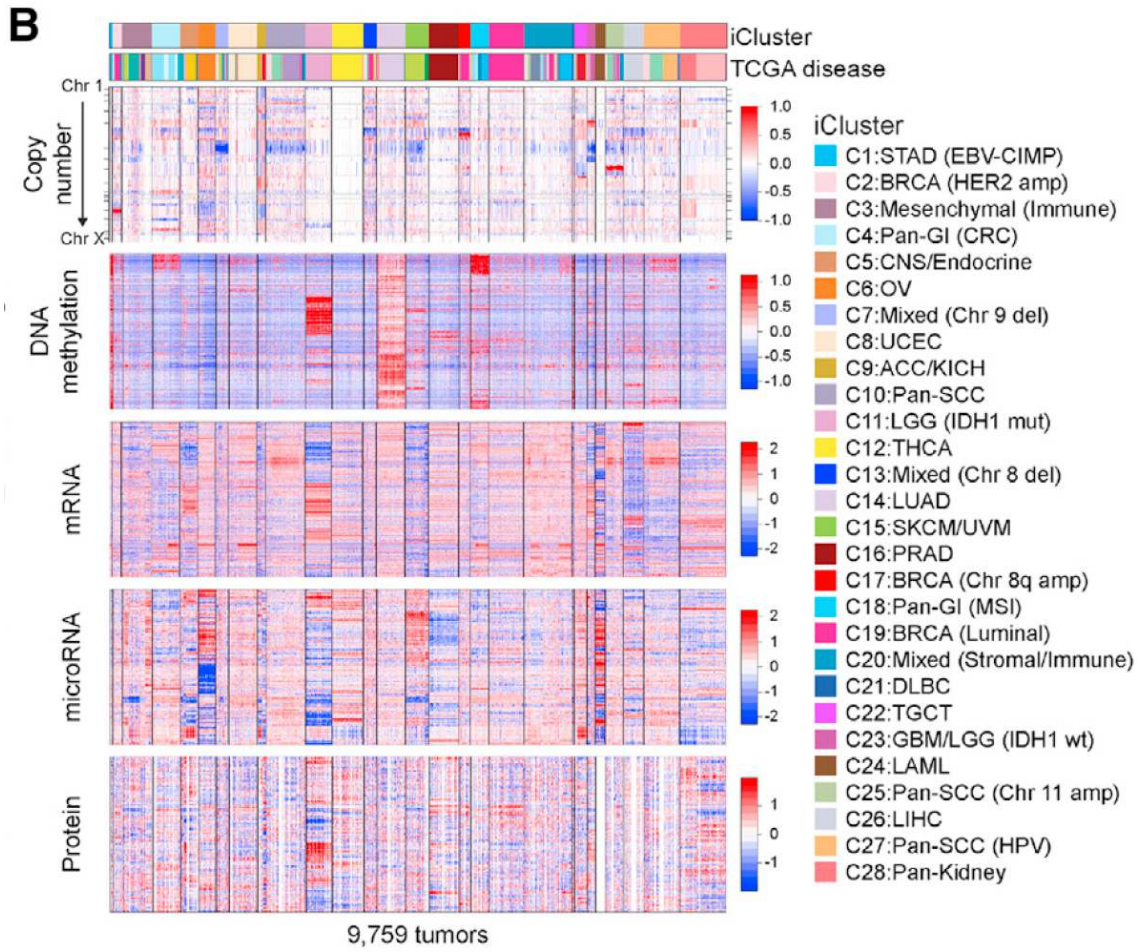
biocViews Integrated omic data, Bioinformatics

Repository CRAN

Date/Publication 2012-05-08 04:07:00

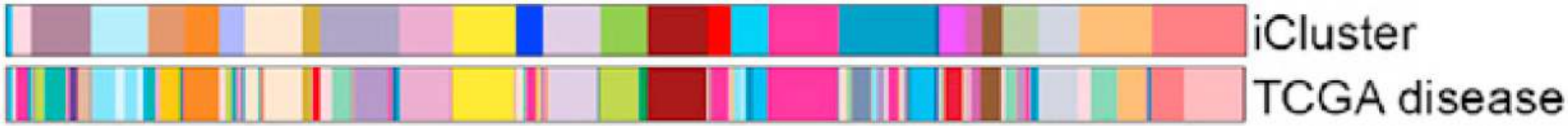
NeedsCompilation no

iCluster Map



iCluster Structure

- While a third of iClusters weremostly homogeneous for a single tumor type, the other two-thirds showed varying degrees of heterogeneity.
- The most diverse group, C20:mixed (stromal/immune), contained a remarkable 25 tumor types).



Concordance of methods

- Individual cluster assignments, COCA, and iCluster-determined molecular subsets were concordant
- Multiplatform co-clustering of different kidney malignancies (pankidney), various gastrointestinal malignancies (pan-GI), diverse squamous cell malignancies (pan-squamous) and most gynecological malignancies (pan-gyn)

iCluster: Unknown Primary

- These iCluster assignments have potential clinical utility, and their multi-platform basis suggests that this new subclassification system might further improve the management of the 1%–3%of all cancer patients newly diagnosed with **cancer of unknown primary (CUP)**

Usage of different data types

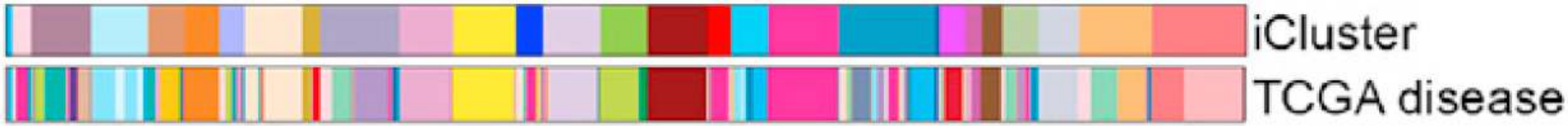
- Using either RNA or DNA methylation profiling has recently led to improved patient outcomes by better defining the tissues of origin for this diverse group of life-threatening malignancies.

iCluster: Data Types "Variance Explained"

- The relative contribution of each platform to the overall clustering was quantified by summing the different platform feature weights on the iCluster latent variables.
- Copy-number alterations contributed 47% to the overall integrated clustering results, followed by the transcriptome(mRNA and miRNA) at 42%, and DNA methylation at 11%.

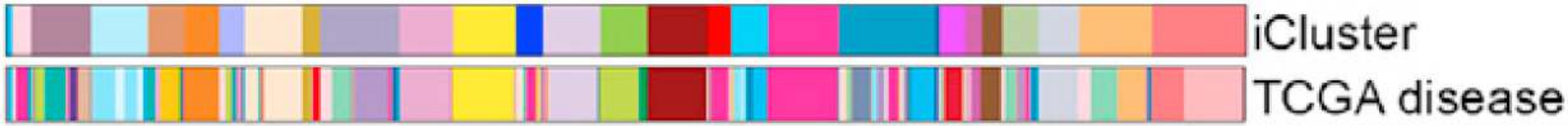
iCluster: Homogenity

- For 16 of the tumor types, over 80% of samples grouped together in the same iCluster.
- Eight iClusters were dominated by a single tumor type
- Others contained tumors from similar or related cells or tissues



iCluster: Tumor Types diversity

- Six tumor types had more diverse iCluster membership, with less than 50% of tumors represented in a given iCluster
- The pan-GI cohort separated into three iClusters (C1, C4, and C18), primarily driven by differences in DNA methylation profiles.

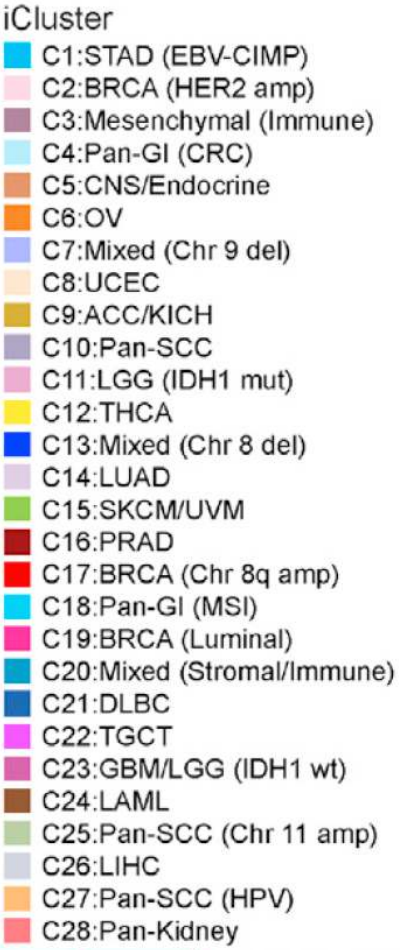


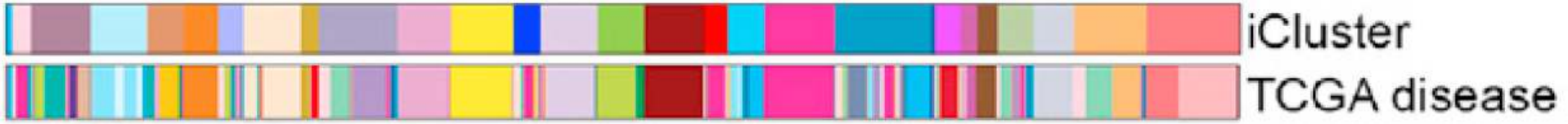
iCluster: Separation by Oncogenic Process

- C1:STAD (Epstein-Barr virus [EBV]-CIMP) consisted of hypermethylated EBV-associated tumors, and C18:pan-GI (MSI) consisted mostly of microsatellite instability (MSI) tumors of STAD and COAD.
- C4:pan-GI (CRC) was predominantly COAD and READ with chromosomal instability (CIN) and a distinct aneuploidy profile.

iCluster: Squamous Cancers

- The pan-squamous cohort formed three iClusters (C10, C25, and C27). The majority of Lung SC fell into C10:pan-Squamous Cell Carcinoma, and nearly all Cervical SC fell into C27:pan-SCC (human papillomavirus [HPV]).



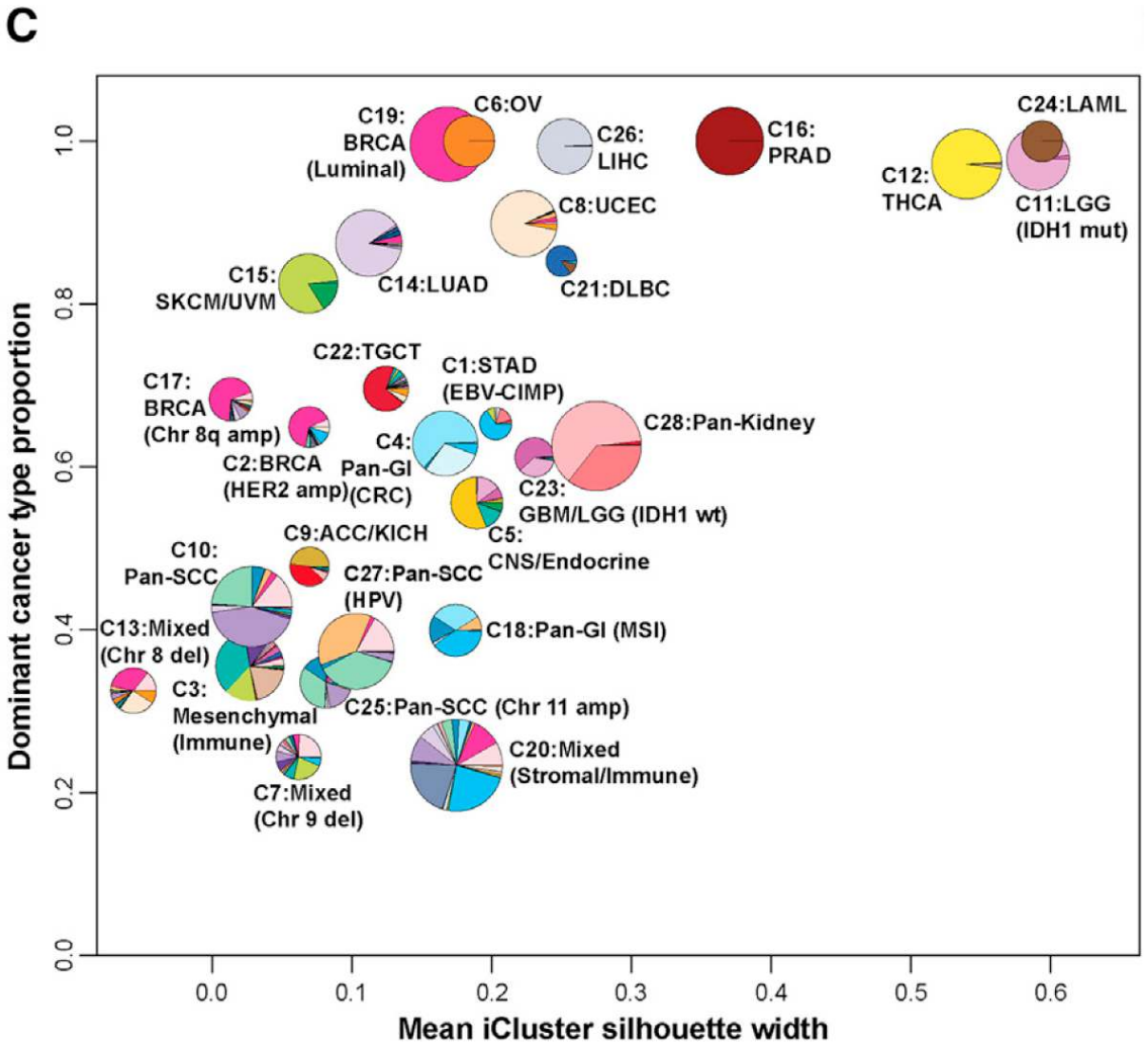


iCluster: Mixed Clusters

- Among mixed tumor type iClusters, three were defined by copy-number alterations. C7:mixed was characterized by **chr9 deletion**, C2:BRCA mainly **ERBB2-amplified tumors**, and C13:mixed (Chr8 del) contained **highly aneuploid tumors**. C3 and C20 were **defined by their non-tumor cell components** including immune and stromal features.

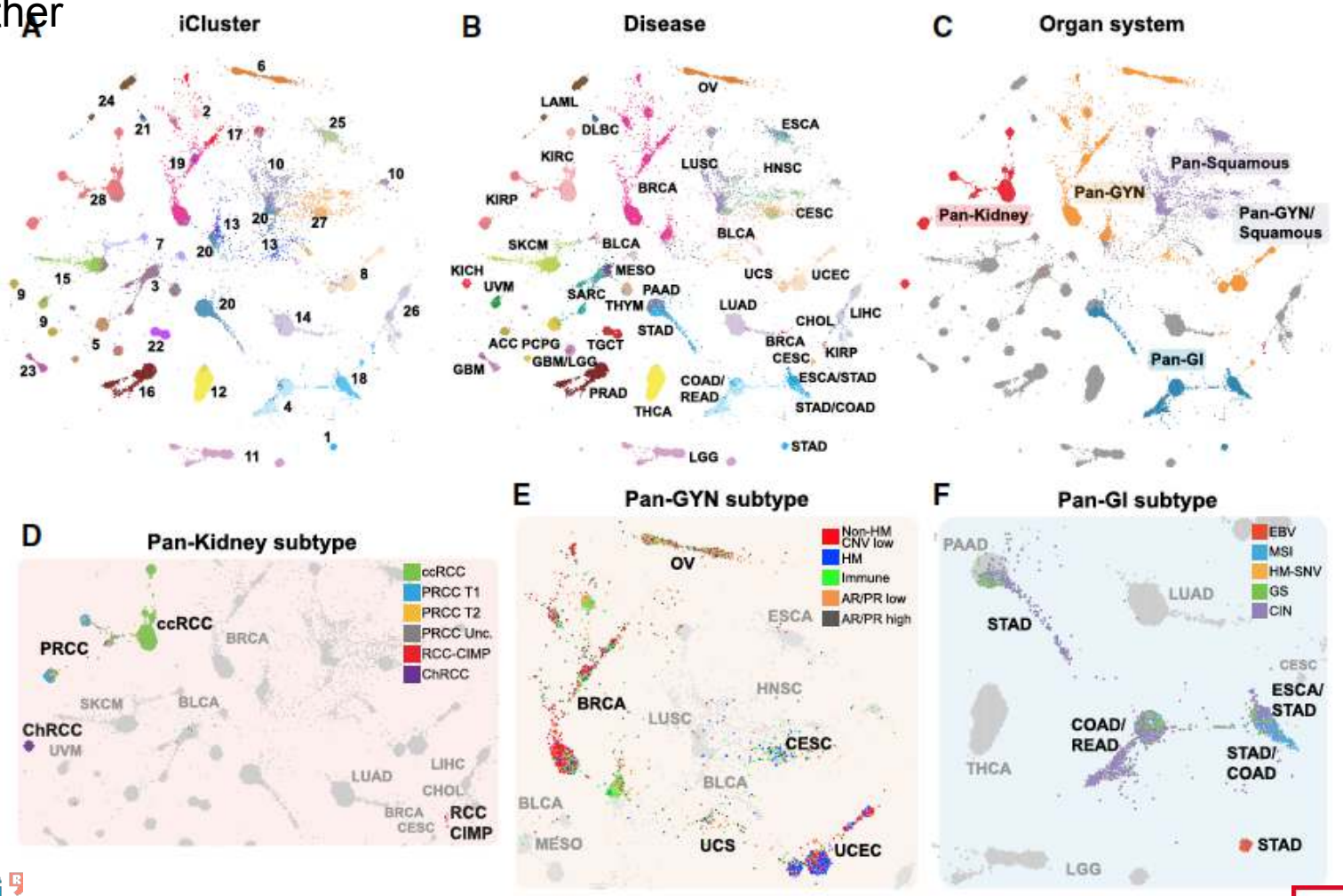
iCluster siluette

Siluette width = measure of within group homogenity

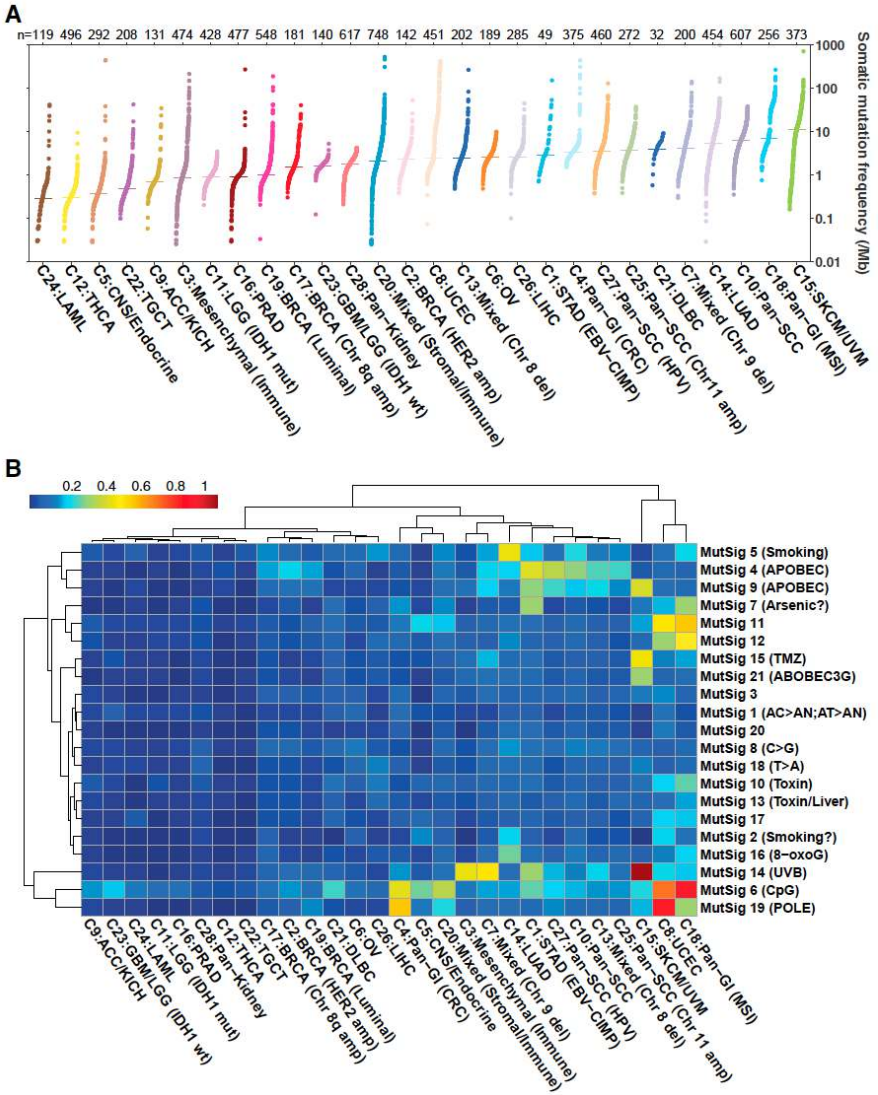


iCluster tumor map

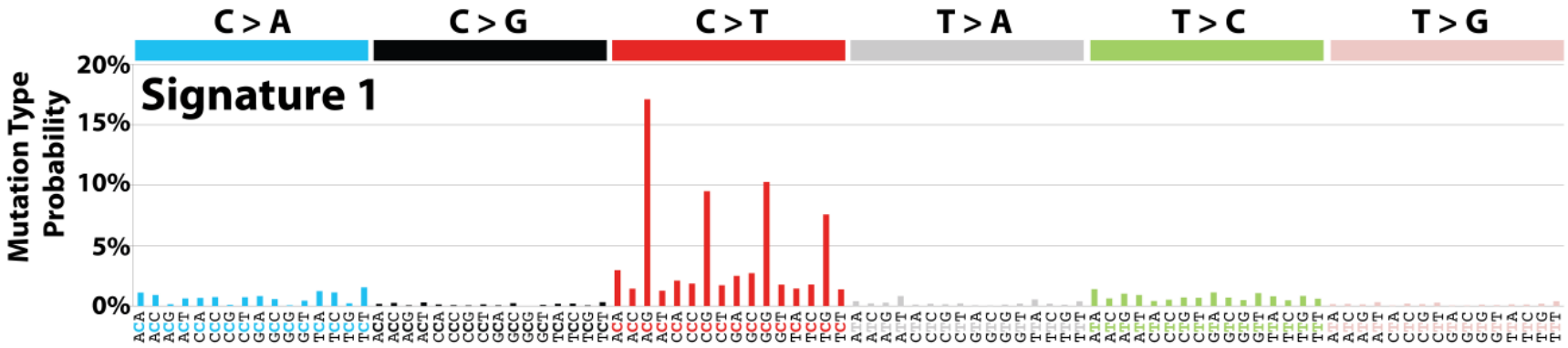
The map layout was computed from sample Euclidean similarity in the iCluster latent space, and similar samples are positioned in close proximity to each other



Mutation Burden and signatures



Mutational Signatures

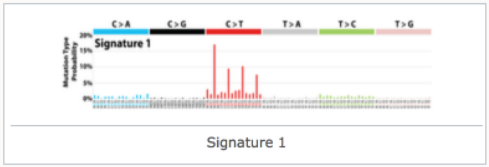


Mutational Signatures

https://cancer.sanger.ac.uk/cosmic/signatures

as the number of mutational signatures grows the need for a curated census of signatures has become apparent. Here, we deliver such a resource by providing the profiles of, and additional information about, known mutational signatures.

Signature 1



Signature 1

Cancer types: Signature 1 has been found in all cancer types and in most cancer samples.

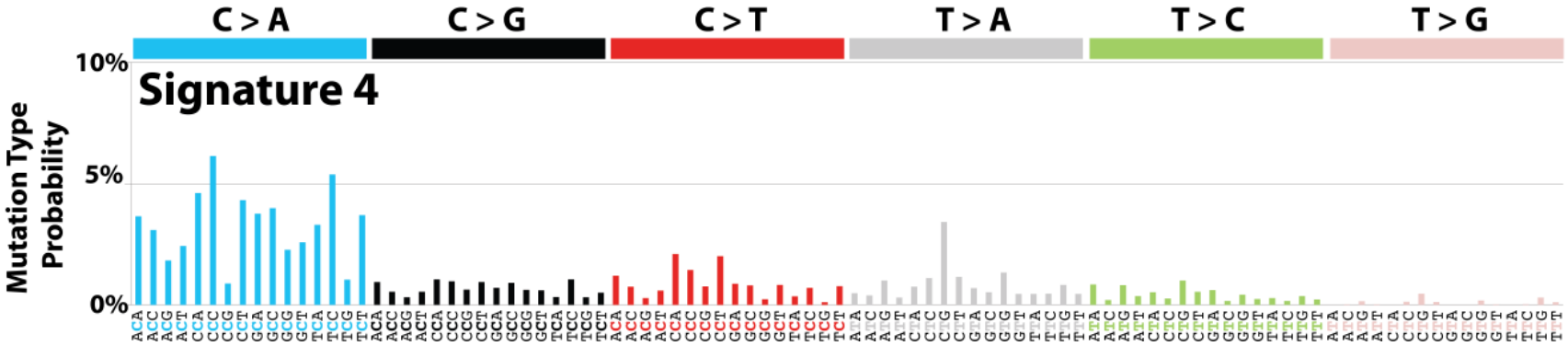
Proposed aetiology: Signature 1 is the result of an endogenous mutational process initiated by spontaneous deamination of 5-methylcytosine.

Additional mutational features: Signature 1 is associated with small numbers of small insertions and deletions in most tissue types.

Comments: The number of Signature 1 mutations correlates with age of cancer diagnosis.

Signatures are needed for understanding the processes that lead to particular tumor type

Mutational Signatures



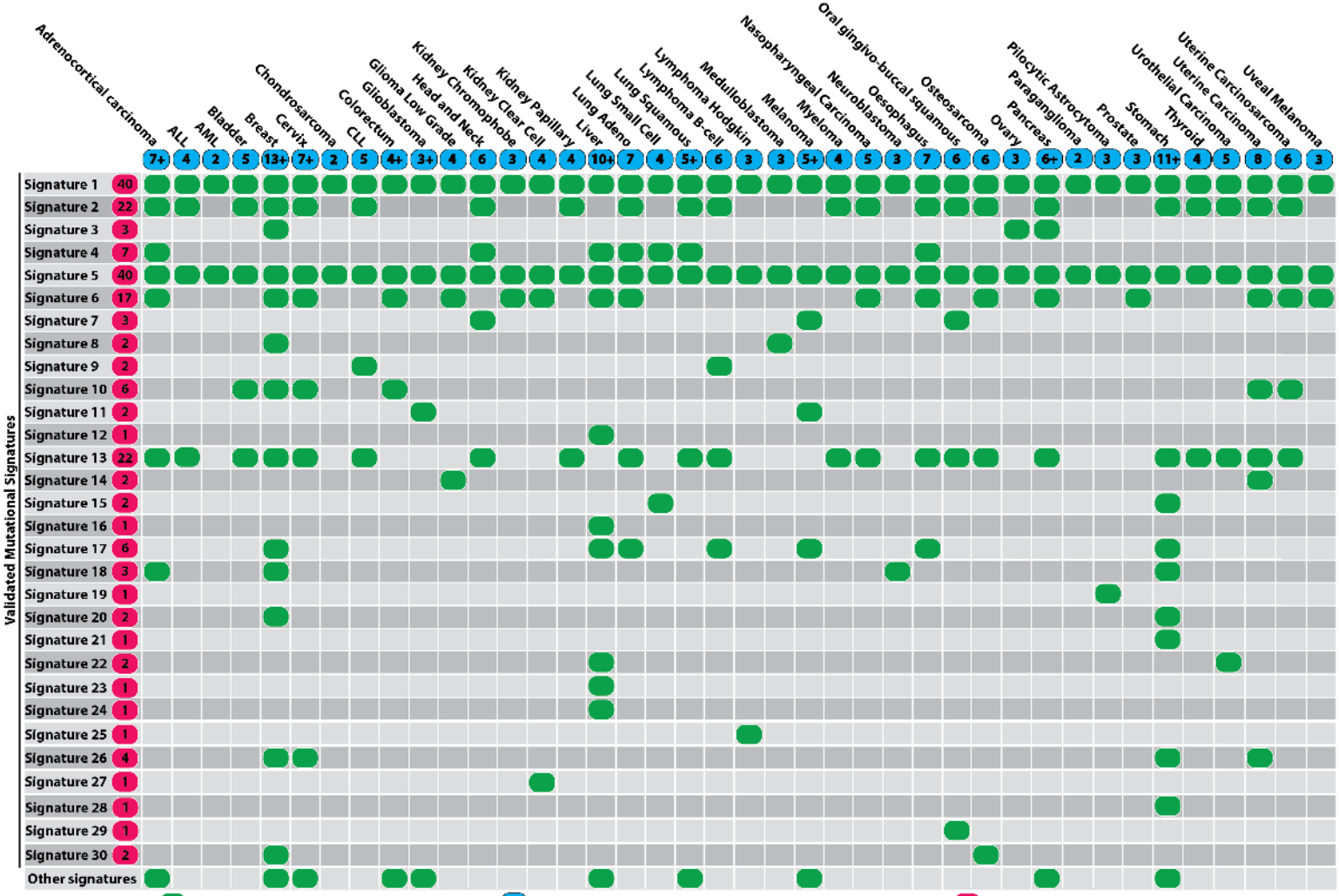
Cancer types: Signature 4 has been found in head and neck cancer, liver cancer, lung adenocarcinoma, lung squamous carcinoma, small cell lung carcinoma, and oesophageal cancer.

Proposed aetiology: Signature 4 is associated with smoking and its profile is similar to the mutational pattern observed in experimental systems exposed to tobacco carcinogens (e.g., benzo[a]pyrene). Signature 4 is likely due to tobacco mutagens.

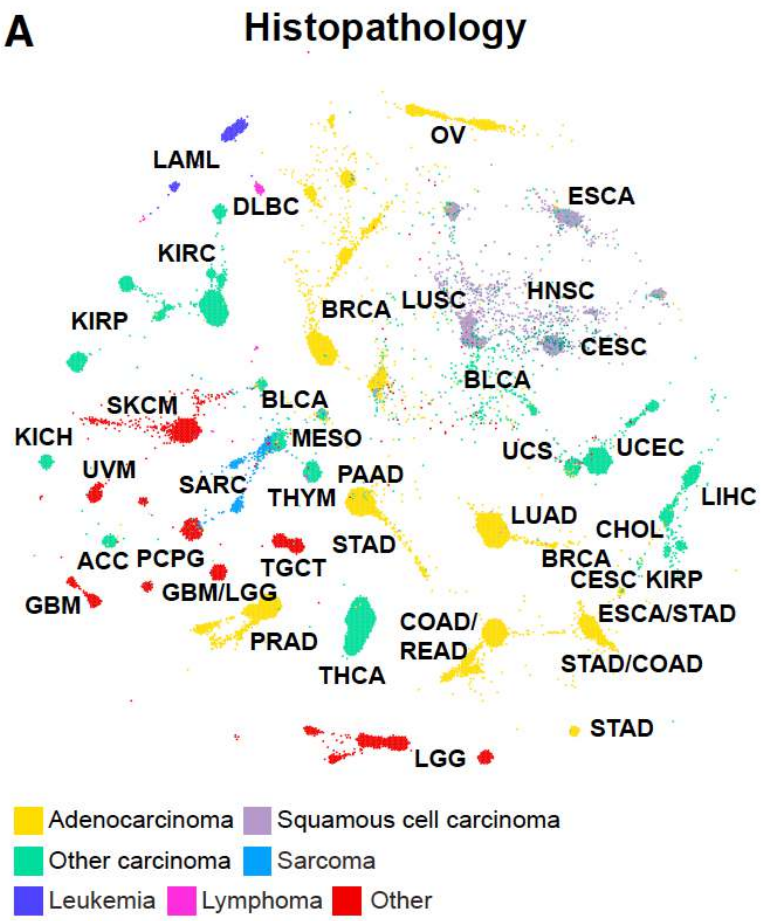
Additional mutational features: Signature 4 exhibits transcriptional strand bias for C>A mutations, compatible with the notion that damage to guanine is repaired by transcription-coupled nucleotide excision repair. Signature 4 is also associated with CC>AA dinucleotide substitutions.

Comments: Signature 29 is found in cancers associated with tobacco chewing and appears different from Signature 4.

Mutational Signatures



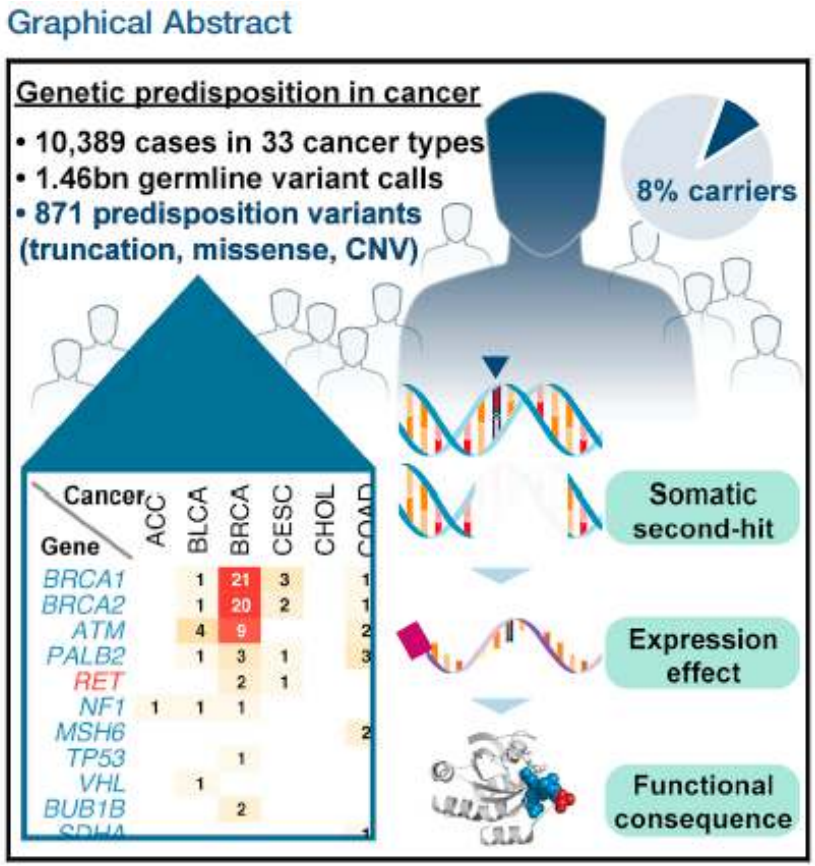
Histopathology



Conclusion

- As drugs become increasingly clinically available to target such cancer-driving pathway targets as ALK, EGFR, ERBB2, ER**a**, KIT, BRAF, and ABL1, the traditional system of anatomic cancer classification should be supplemented by a classification system based on molecular alterations shared by tumors across different tissue types

Pathogenic Germline Variants (Adult cancers)

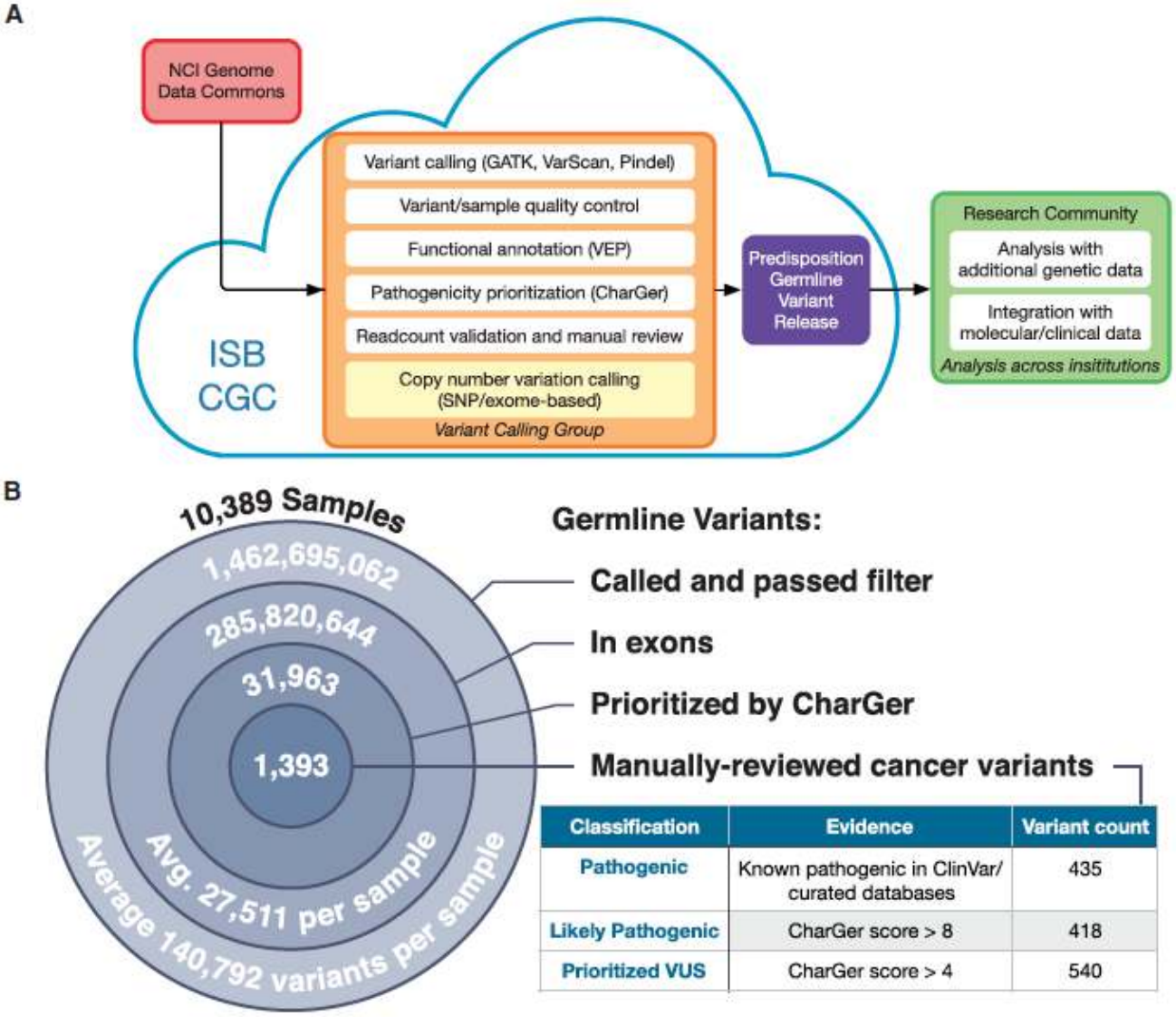


- 871 predisposition variants/CNVs discovered in 8% of 10,389 cases of 33 cancers
- Pan-cancer approach identified shared variants and genes across cancers
- 33 variants affecting activating domains of oncogenes showed high expression
- 47 VUSs prioritized using cancer enrichment, LOH, expression and other evidence

2-hit hypothesis

- Tumor supressor in most of the cases remains functional even if it was mutated in 1 copy
- Thus, 2 loss-of-function mutations have to occur in a supressor to start the oncogenic process
- Loss of heterozygosity leads to inactivation of 2nd copy of tumor supressor

Abstract



Classification	Evidence	Variant count
Pathogenic	Known pathogenic in ClinVar/ curated databases	435
Likely Pathogenic	CharGer score > 8	418
Prioritized VUS	CharGer score > 4	540

The landscape of pathogenic variants

- 8% of cases carrying pathogenic or likely pathogenic germline variants, ranging in prevalence from a striking 22.9% in [Pheochromocytoma and Paraganglioma] to a scarce 2.2% in Cholangiocarcinoma.
- Notably, we identified 33 such variants within oncogenes.

Data Generation

- More than 121,000 virtual machines running for over 600,000 hr on the ISB-CGC during the course of the project.
- Precision of variants >99%
- Variant calls from Genome Analysis Toolkit, VarScan2, and Pindel were merged, filtered, and annotated, resulting in 286,657,499 total exonic variants

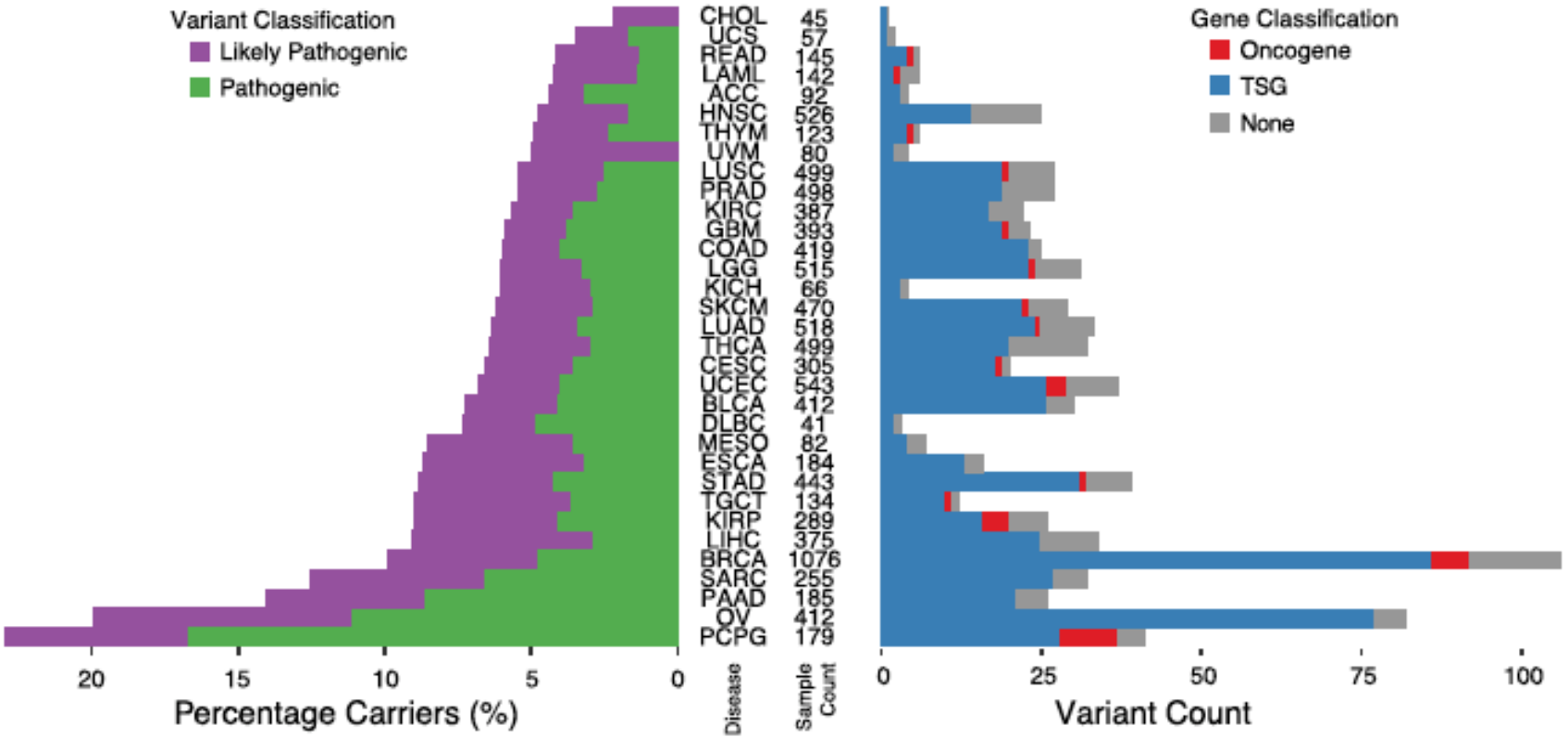
Frequency of pathogenic variants

- Across all cancer types, 4.1% of cases (n = 428) harbored pathogenic variants, and another 3.8% (n = 390) carried likely pathogenic variants.
- The frequencies of pathogenic or likely pathogenic variants vary greatly across cancer types, with the expected high rates in OV (19.9%) and BRCA (9.9%).

Which genes

- 659 pathogenic or likely pathogenic variants in **66 tumor suppressor genes**
- 33 pathogenic or likely pathogenic variants in **five oncogenes**: RET, AR, PTPN11, MET, and CBL.
- 21 RET variants were found across 11 cancer types. Others appear as cancer specific

Histogram

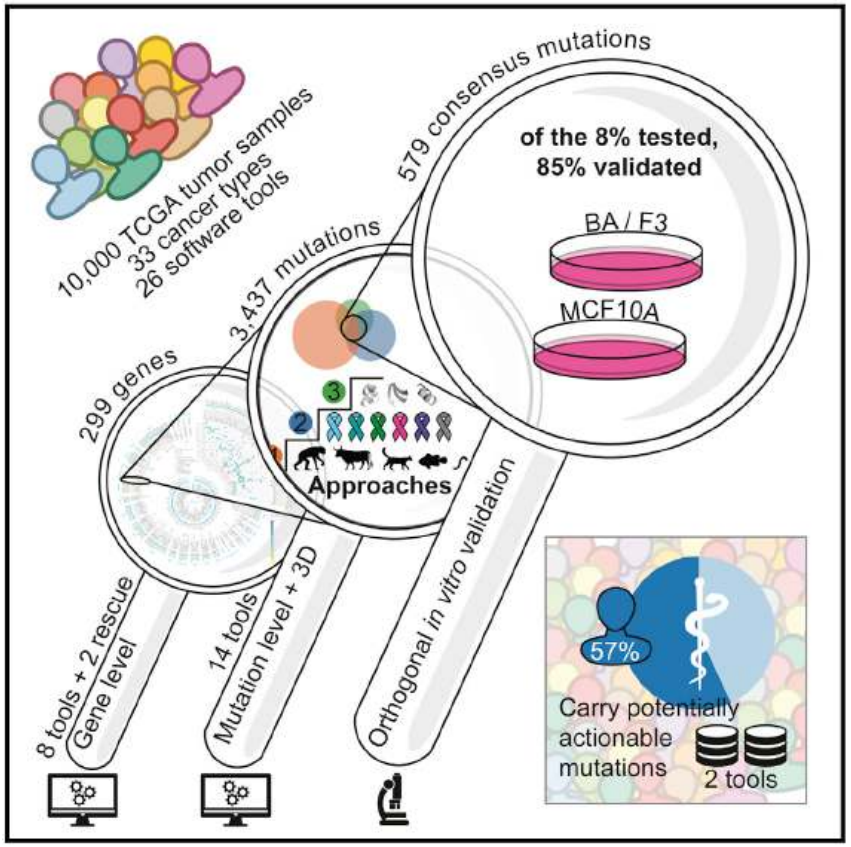

Variant Classification
Likely Pathogenic
Pathogenic
CHOL 45
UCS 57
READ 145
LAML 142
ACC 92
HNSC 526
THYM 123
UVM 80
LUSC 499
PRAD 498
KIRC 387
GBM 393
COAD 419
LGG 515
KICH 66
SKCM 470
LUAD 518
THCA 499
CESC 305
UCEC 543
BLCA 412
DLBC 41
MESO 82
ESCA 184
STAD 443
TGCT 134
KIRP 289
LIHC 375
BRCA 1076
SARC 255
PAAD 185
OV 412
PCPG 179
20
15
10
5
0
Percentage Carriers (%)
Disease
Sample Count
Gene Classification
Oncogene
TSG
None
0
25
50
75
100
Variant Count

Germline Predisposition

CNVs

- 42,208 rare (AF < 0.6% considering 50% overlaps) CNVs using SNP-array data and 53,726 using XHMM on the WES data. 3584 were identified by both platforms
- 18 events (2 were jointly identified using both WES and SNP array) that marked copy number deletions of 11 tumor suppressor genes

Cancer driver genes and mutations



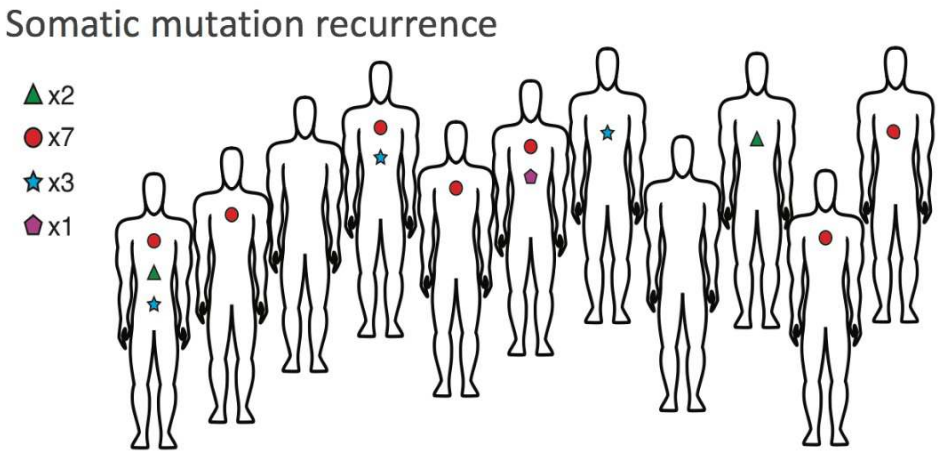
- 299 cancer driver genes
- Drivers and mutations are shared across types
- ~3400 driver mutations
- 57% of tumors have potentially actionable mutations

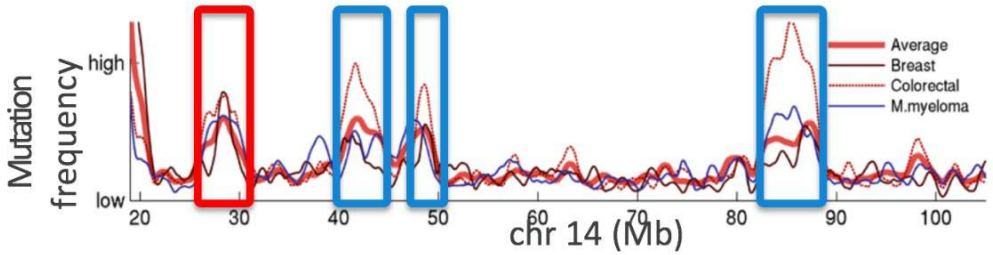
Cancer driver and cancer passenger

- **Cancer driver** = push cell towards cancer state
- **Cancer passenger** = appear as a consequence of oncogenic process
- Only small part of cancer mutations are driver mutations
- There are "*fragile*" sites in human genome − deletion of genetic material there is most likely passenger

How to search for cancer driver genes

Driver genes identification

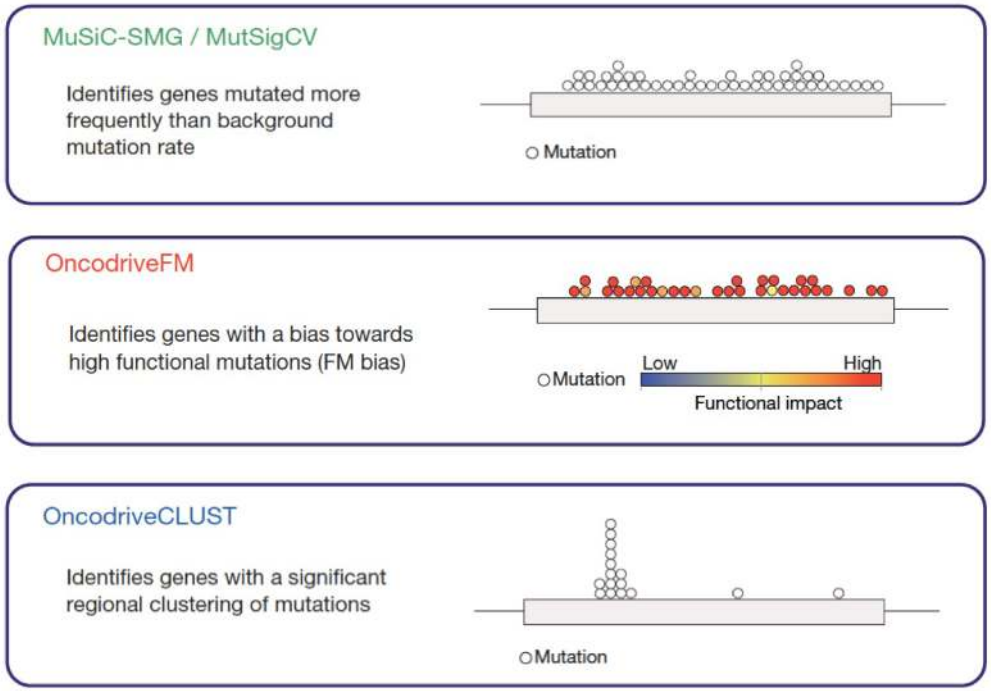




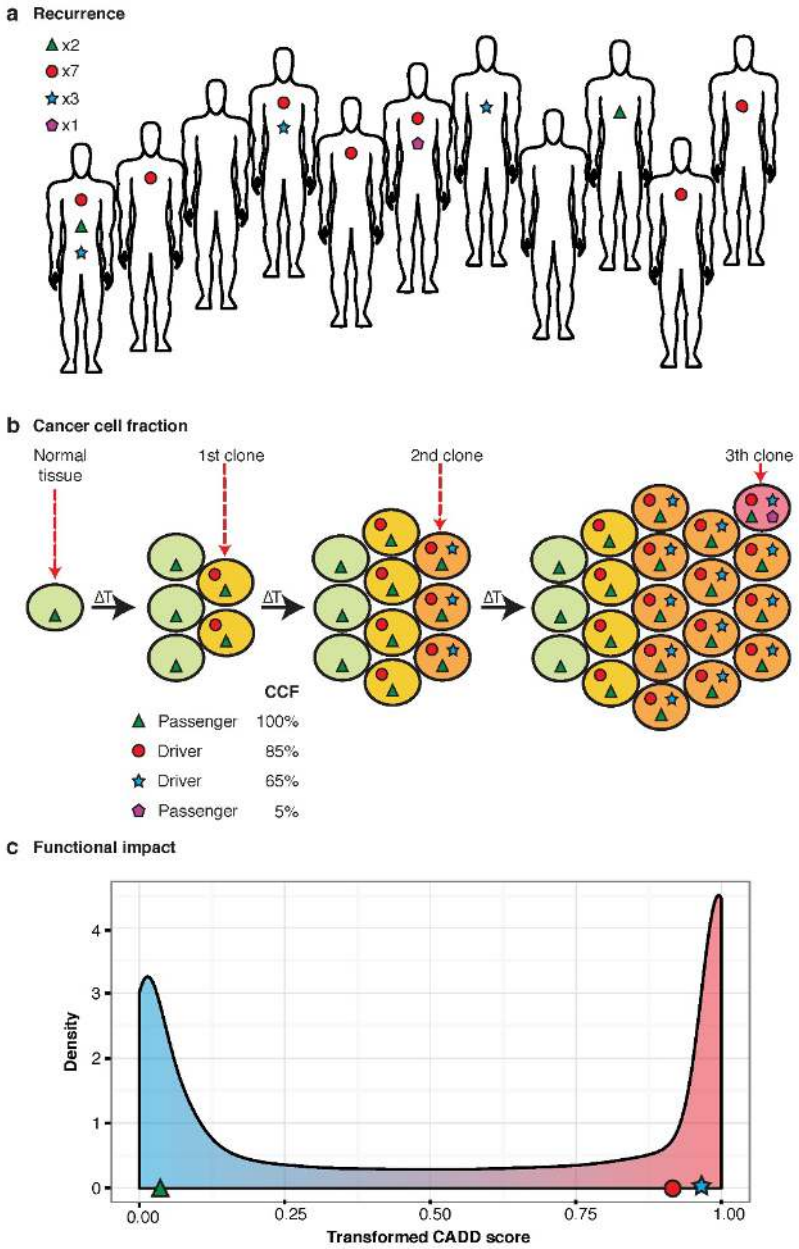
Thanks to Hana Susak for plot!

How to search for cancer driver genes

Driver genes identification



Thanks to Hana Susak for plot!



Thanks to Hana Susak for plot!

Variant Calling

- 7 algorithms for calling
- 344 hypermutator samples were excluded (sensitivity to background mutation rate)
- Less stringent filers applied to ovarian serous cystadenocarcinoma (OV) and acute myeloid leukemia (LAML) – WES data from them have different characteristics

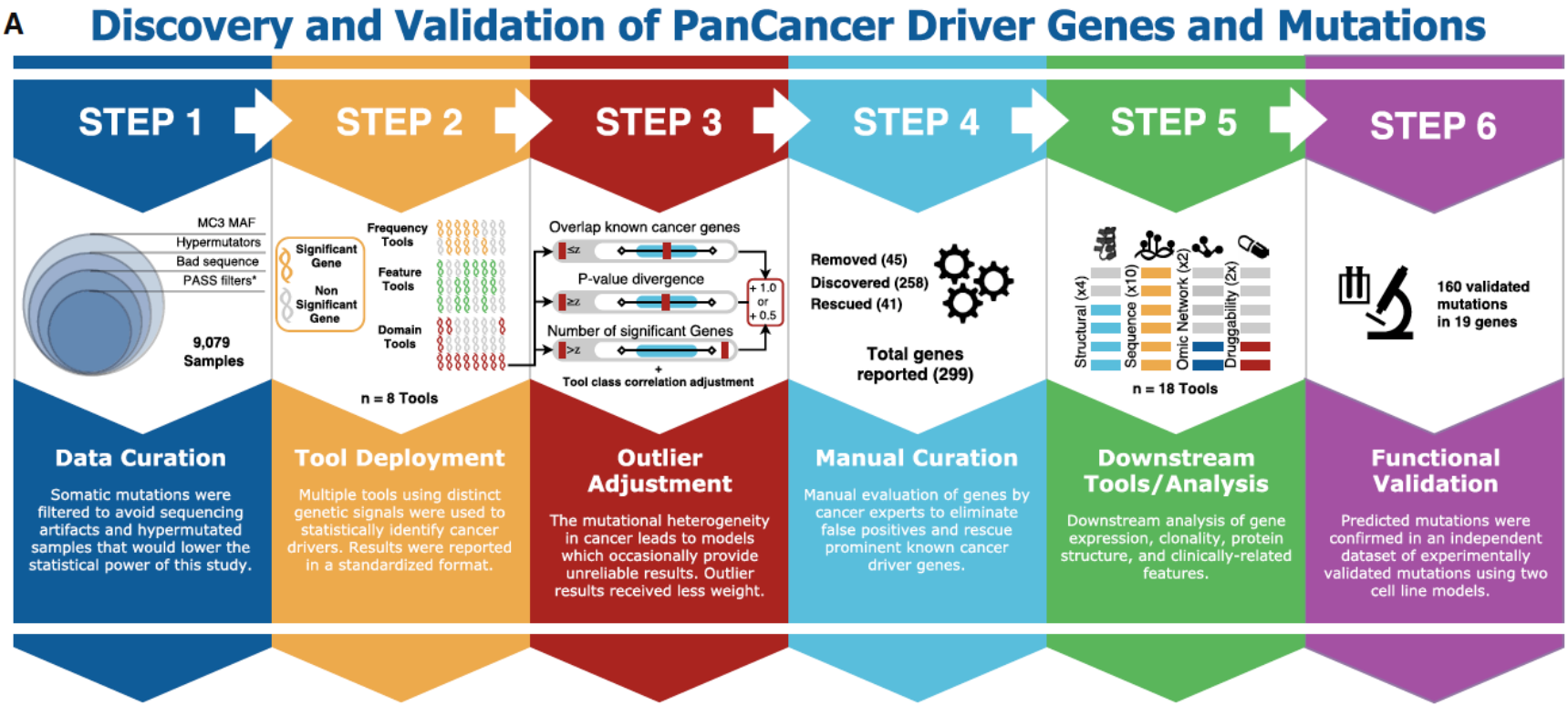
Power of Detection

- lymphoid neoplasm diffuse large B cell lymphoma (25.5%, n = 37), cholangiocarcinoma (20.5%, n = 34), and uterine carcinosarcoma (14.9%, n = 55)
- breast invasive carcinoma (BRCA) (2.3%, n = 779), brain lower grade glioma (2.8%, n = 510), and thyroid carcinoma (2.3%, n = 491).
- Median: 6.1%

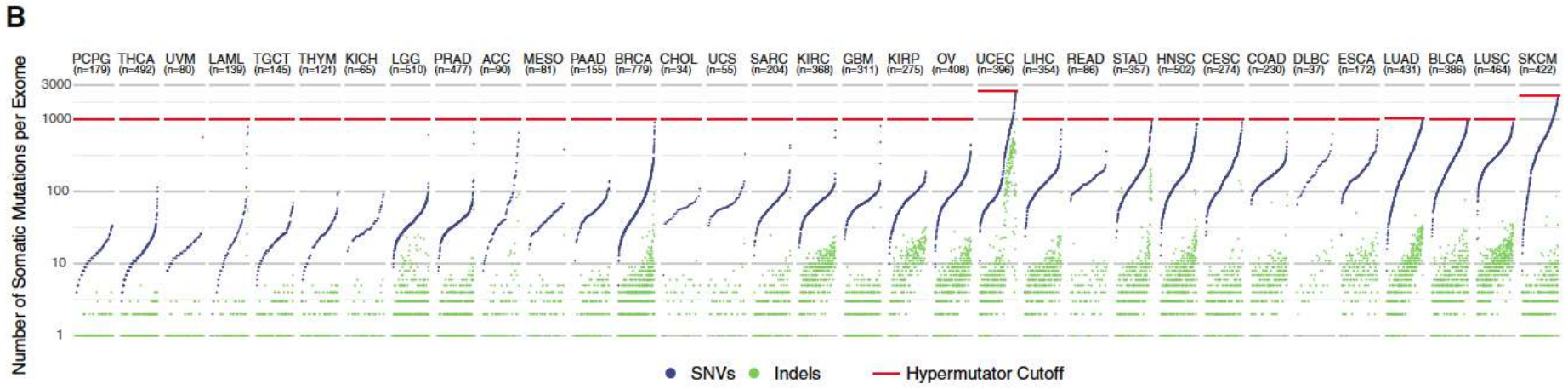
2 approaches

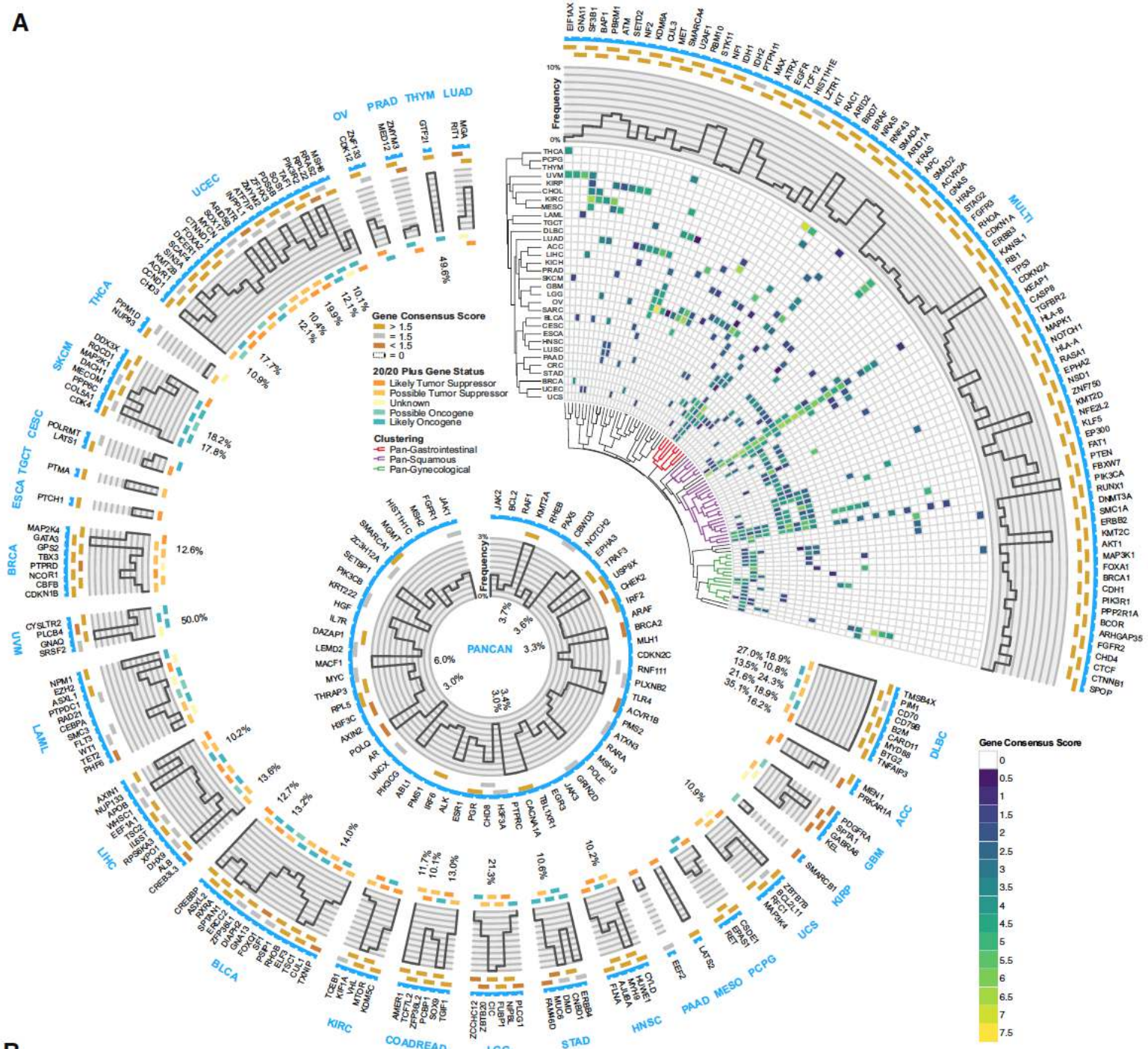
- The final consensus list consists of 299 unique genes: 258 genes obtained from a systematic approach and 41 additional genes recovered after manual curation of previous TCGA marker papers
- 142 out of 258 associated with single cancer

Pipeline


A
Discovery and Validation of PanCancer Driver Genes and Mutations
STEP 1
STEP 2
STEP 3
STEP 4
STEP 5
STEP 6
MC3 MAF
Hypermutators
Bad sequence
PASS filters*
9,079 Samples
Significant Gene
Non Significant Gene
Frequency Tools
Feature Tools
Domain Tools
n = 8 Tools
Overlap known cancer genes
P-value divergence
Number of significant Genes
+ 1.0 or + 0.5
Tool class correlation adjustment
Removed (45)
Discovered (258)
Rescued (41)
Total genes reported (299)
Structural (x4)
Sequence (x10)
Omic Network (x2)
Druggability (2x)
n = 18 Tools
160 validated mutations in 19 genes
Data Curation
Somatic mutations were filtered to avoid sequencing artifacts and hypermutated samples that would lower the statistical power of this study.
Tool Deployment
Multiple tools using distinct genetic signals were used to statistically identify cancer drivers. Results were reported in a standardized format.
Outlier Adjustment
The mutational heterogeneity in cancer leads to models which occasionally provide unreliable results. Outlier results received less weight.
Manual Curation
Manual evaluation of genes by cancer experts to eliminate false positives and rescue prominent known cancer driver genes.
Downstream Tools/Analysis
Downstream analysis of gene expression, clonality, protein structure, and clinically-related features.
Functional Validation
Predicted mutations were confirmed in an independent dataset of experimentally validated mutations using two cell line models.

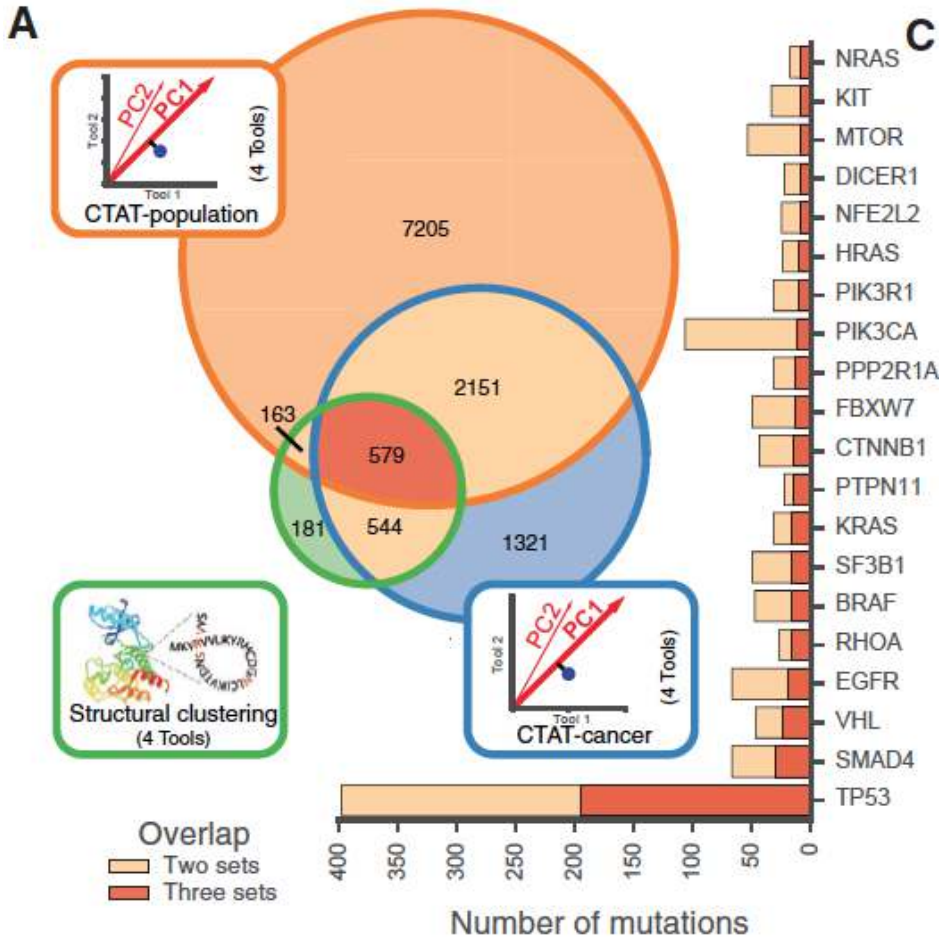
Mutational Burden



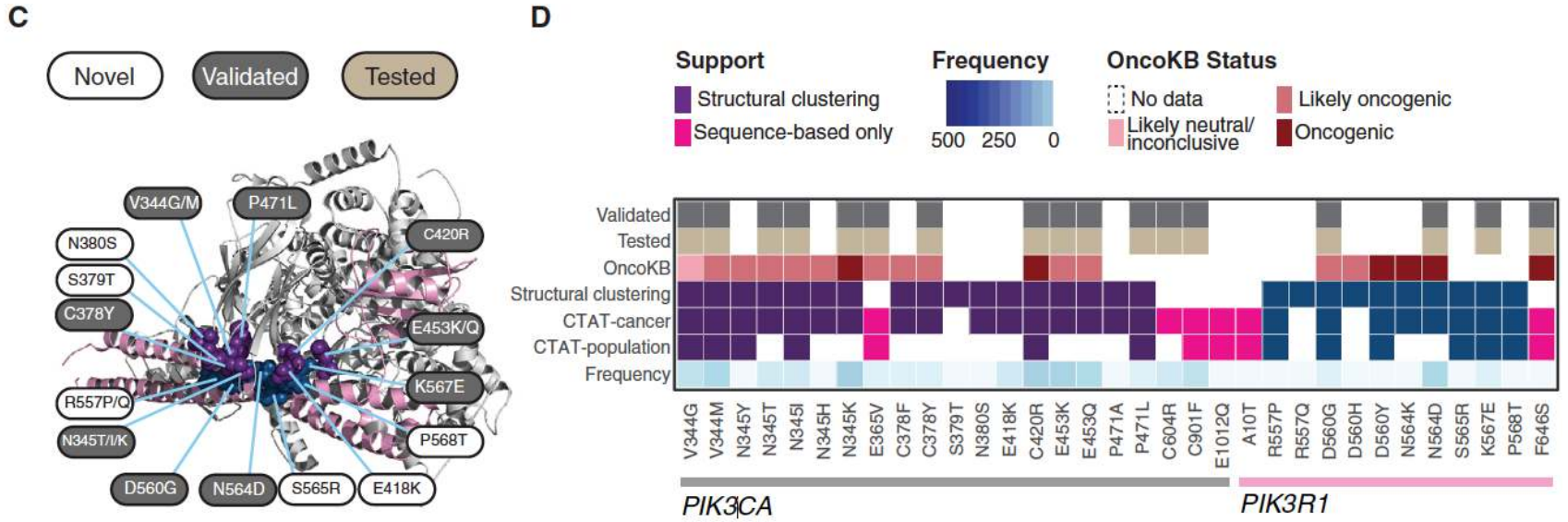


Universitat Pompeu Fabra Barcelona

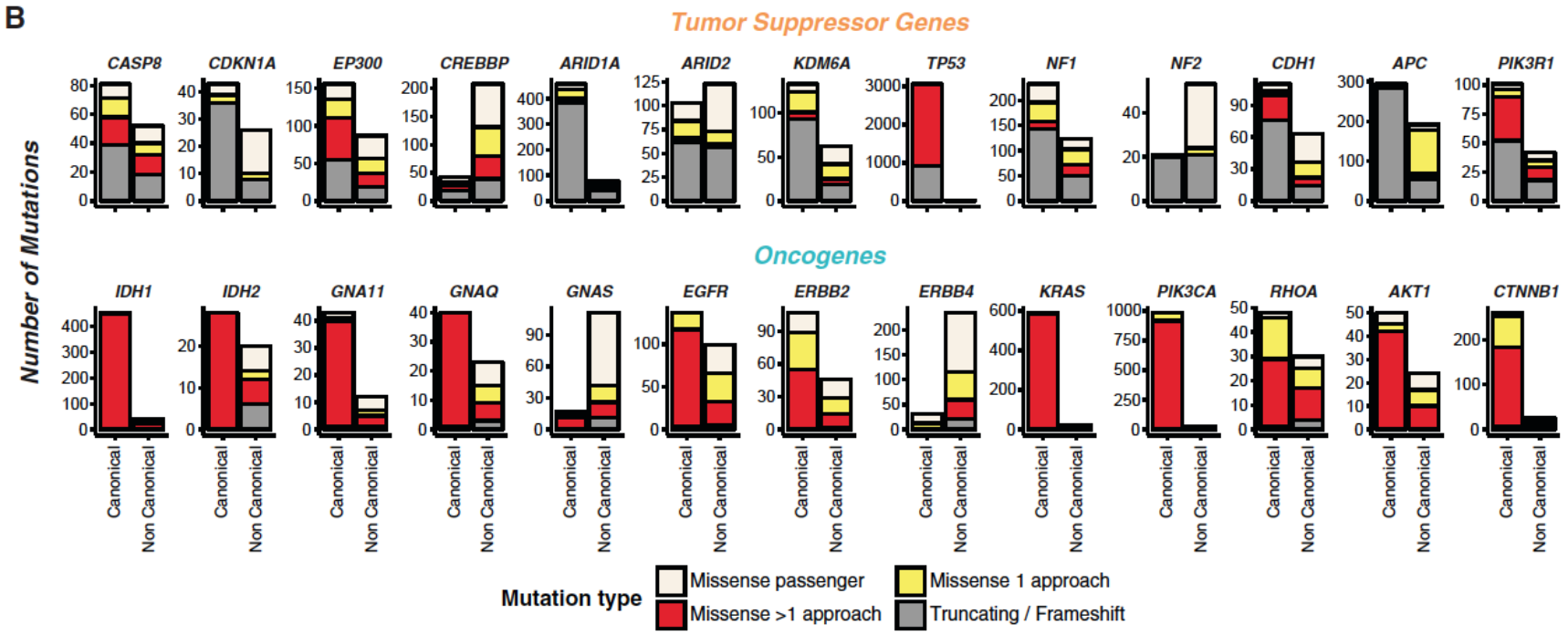
Approaches for Calling



Example: Proteins PIK3CA/ PIK3R1 (PDB: 4OVU)



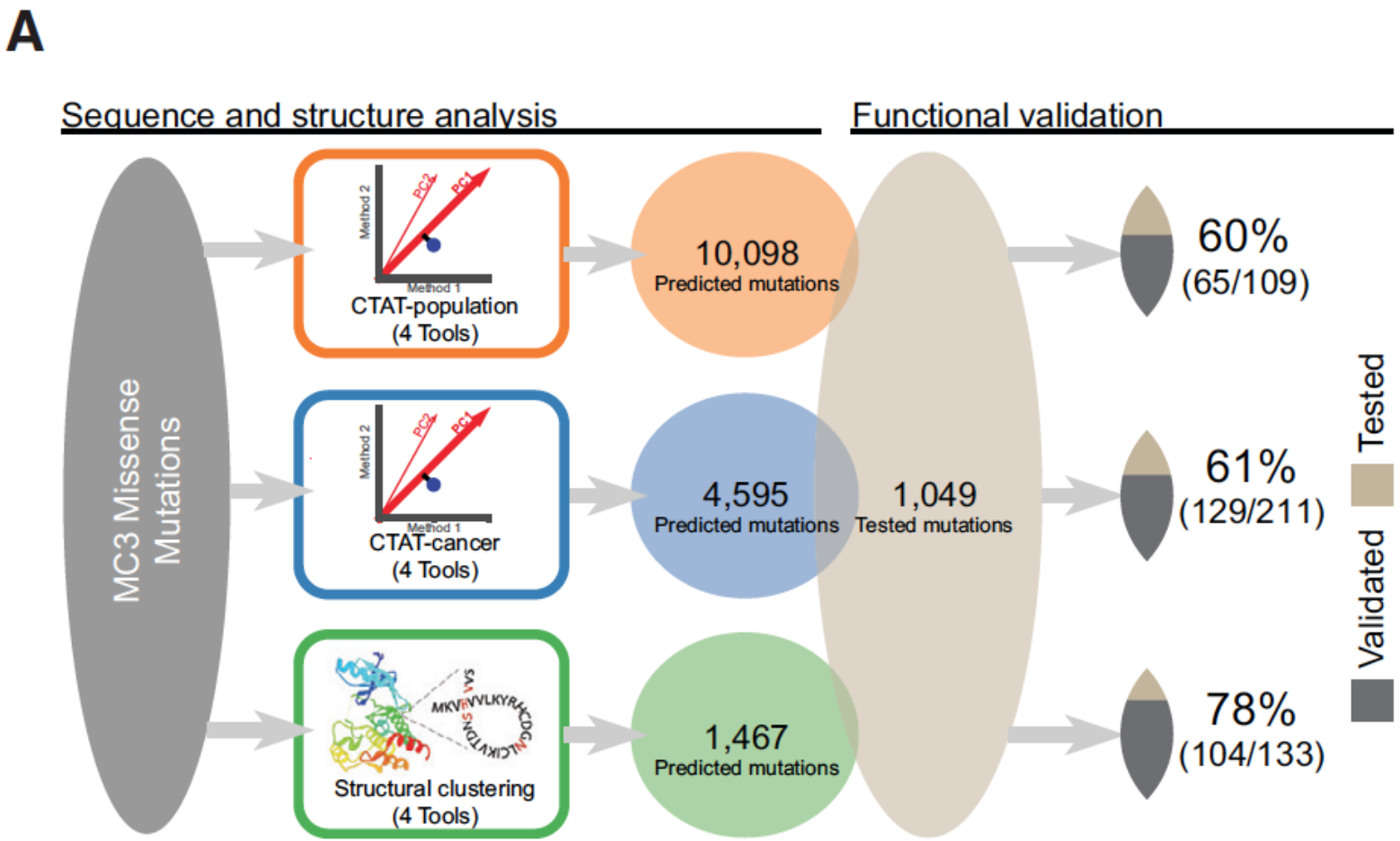
Types of Mutations



How many driver mutations found?

- Overall 9.919 cancer driver muations in 5,782 tumor samples (3.437 sites) identified by >= 2 approaches (population − benign vs pathogenic, cancer − driver vs passenger, stucture − 3D clusters of missence mutations)

Predicted / Validated



Validation

- 1.049 mutations were experimentally tested to validate driver prediction
- They were introduced in 2 cancer cell lines Ba/F3 and MCF10A and evaluated for oncogenicity

Hypermutation

- UV, tobacco, microsatellite instability leads to hypermutator phenotype
- 90% of the samples labeled as hypermutator had MSI, UV, POLE (polymerase), APOBEC (enzyme protecting from viruses) or smoking as primary signature

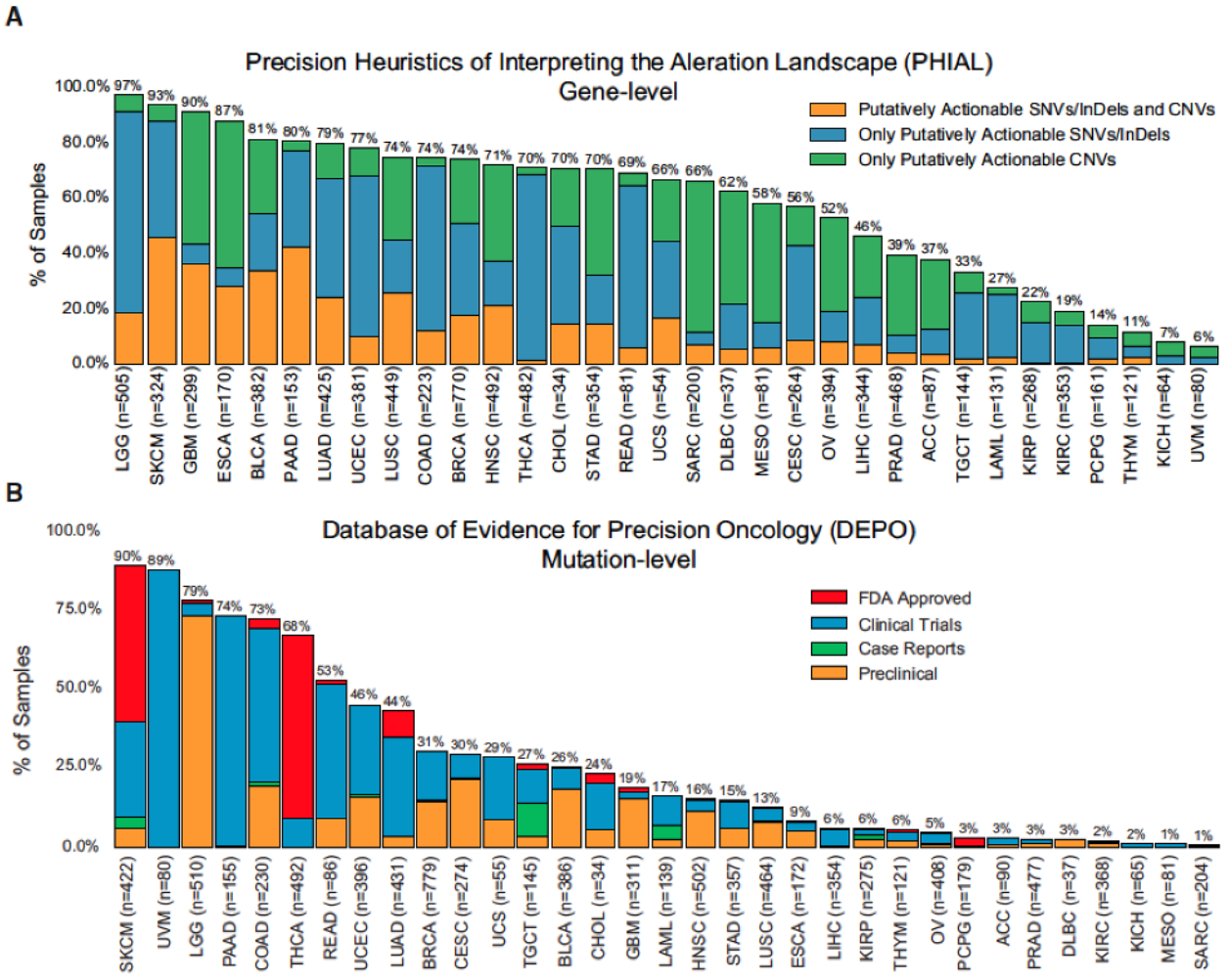
MSI signature

- Canoncial types: Uterine Corpus Endometrial Carcinoma, colon/stomach adenocarcinoma
- 2% of OV, 2% of Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (CESC)
- Improved response to immune therapy (no matter histology)

Therapeutic Implications of molecular events

- 2 databases: Precision Heuristics for Interpreting the Alteration Landscape (PHIAL), Database of Evidence for Precision Oncology (DEPO)
- Both contain therapeutic projections based on FDA-approved therapies, clinical trials, published clinical evidence, and some other databases.
- PHIAL works at the gene level, whereas DEPO focuses on specific mutations

Therapeutic Implications of molecular events



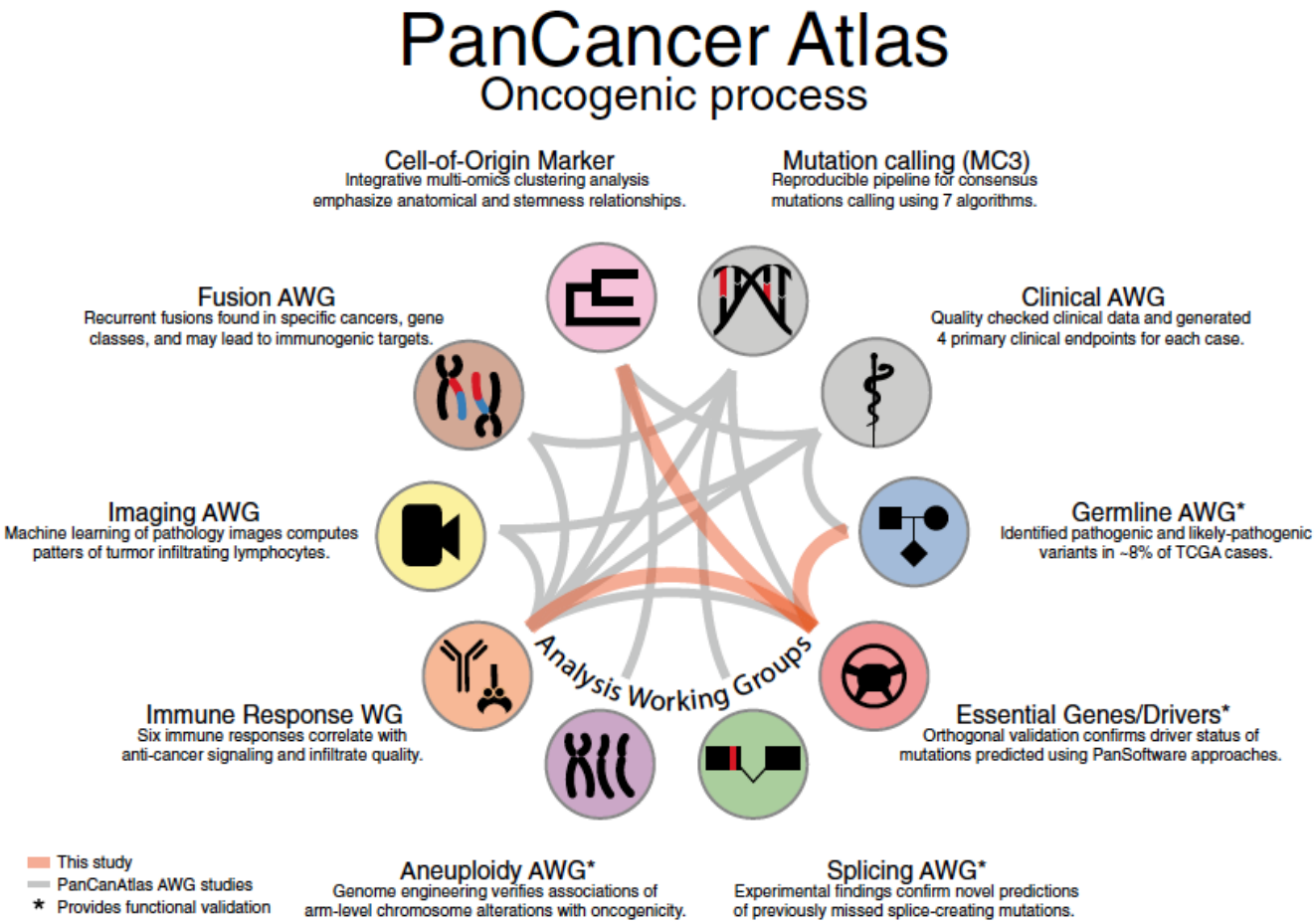
Therapeutic Implications of molecular events

- Using PHIAL, the most common putatively actionable alterations across the entire dataset were CDKN2A deletions (13%), PIK3CA mutations (12%), MYC amplifications (8%), BRAF mutations and amplifications (8%), and KRAS mutations (7%).

Therapeutic Implications of molecular events

- Similar to PHIAL, PIK3CA, BRAF, and KRAS contributed to the most number of samples with potentially actionable alterations from DEPO.

Oncogenic Processes



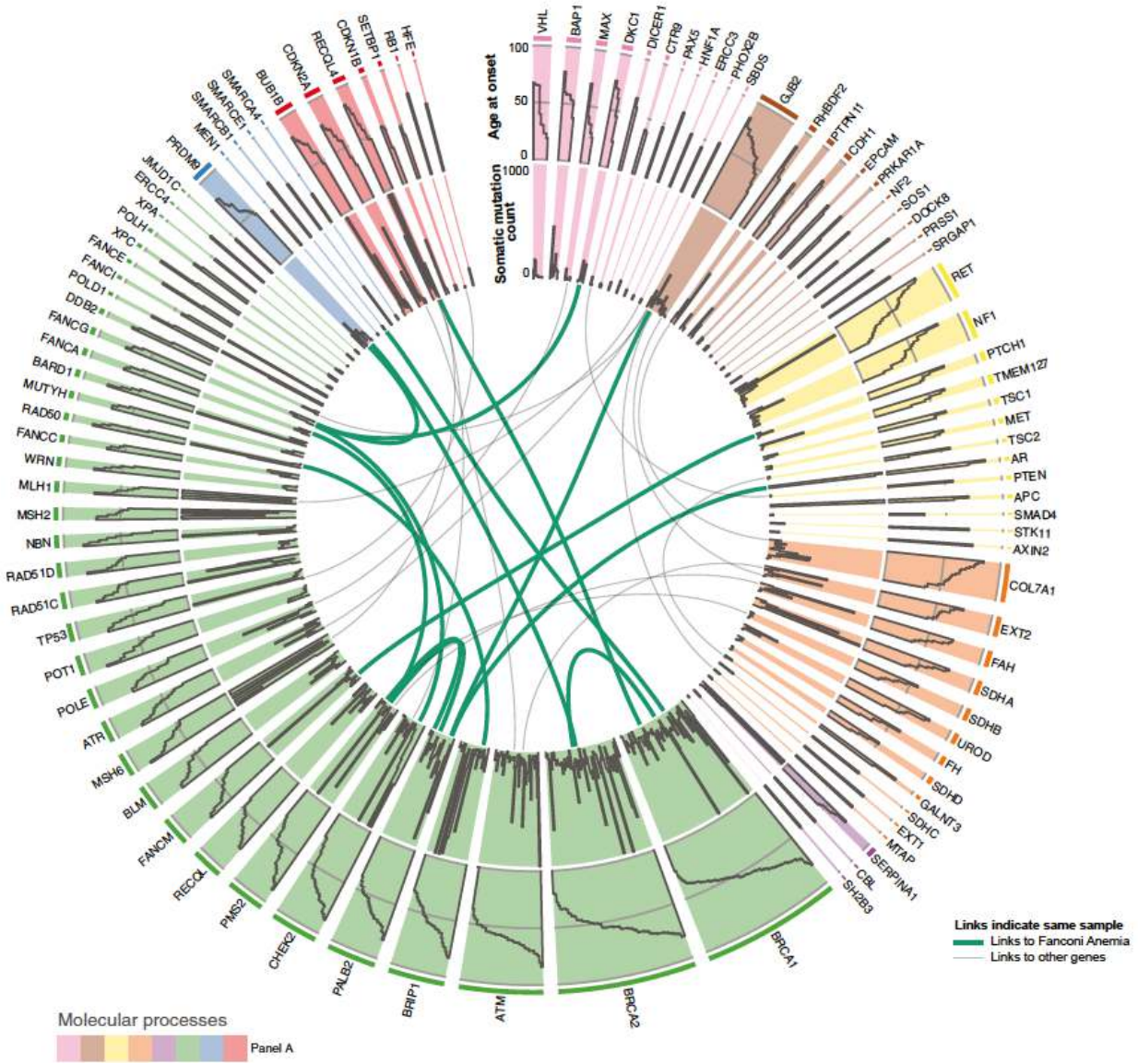
Oncogenic Processes

- (1) interactions between somatic drivers and germline pathogenic variants;
- (2) links across genomic substrates, i.e., methylome, transcriptome, and proteome;
- (3) tumor microenvironment and implications for targeted and immune therapies.

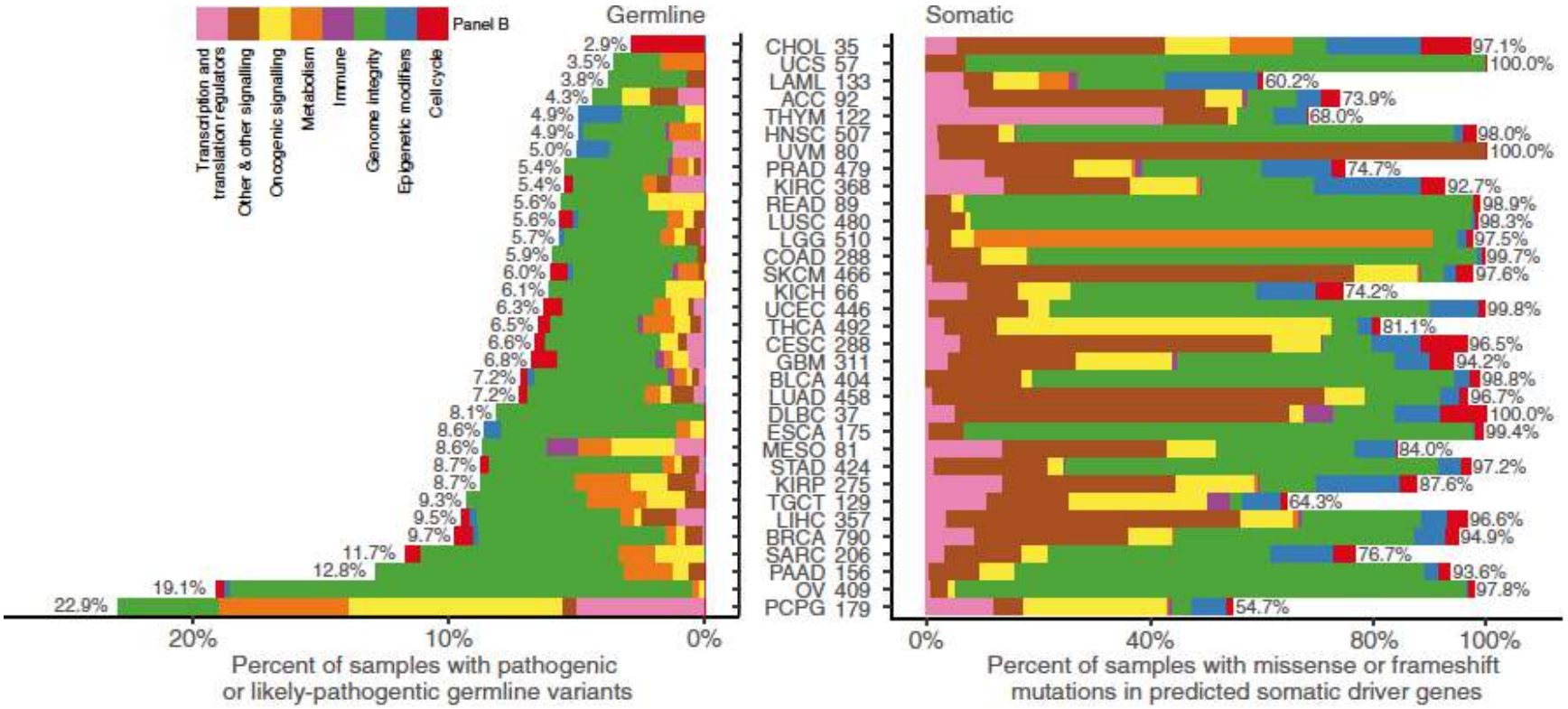
Variant Callsets

- Aneuploidy (focal + chromosomal-level)
- Gene fusions
- Germline and somatic variants

Predisposition Cancer Genes



Germline & Somatic



Predisposition genes

- Many predisposition genes play roles in genome integrity.
- Alterations in these genes represent a higher fraction of germline variants (63%, 490/769) versus somatic drivers (14%, 8850/75825), highlighting the role of genome integrity in cancer predisposition.

Genome integrity

- Dominates germline and somatic alterations in ovarian serous cystadenocarcinoma due to BRCA1 or BRCA2 predisposition variants and a high fraction of TP53 mutations.
- Other cancers are further skewed with respect to percent of cases carrying mutations involved in genome integrity; 4% of samples in lung squamous cell carcinoma have germline compared to 89% somatic

DNA Damage Response Pathway

- Most predisposition genes affecting genome integrity (64%, 23/36) belong to the Core DDR (DNA damage response) genes
- Several show high germline variant counts, including BRCA1, BRCA2, CHEK2, ATM, BRIP1, PALB2, and PMS2.

DNA Damage Response Pathway

- Germline and somatic jointly: the most frequently mutated genes are BRCA1 and BRCA2, together having 854 (571 samples) somatic and 153 (152 samples) germline mutations.

Microsatellite instability phenotypes

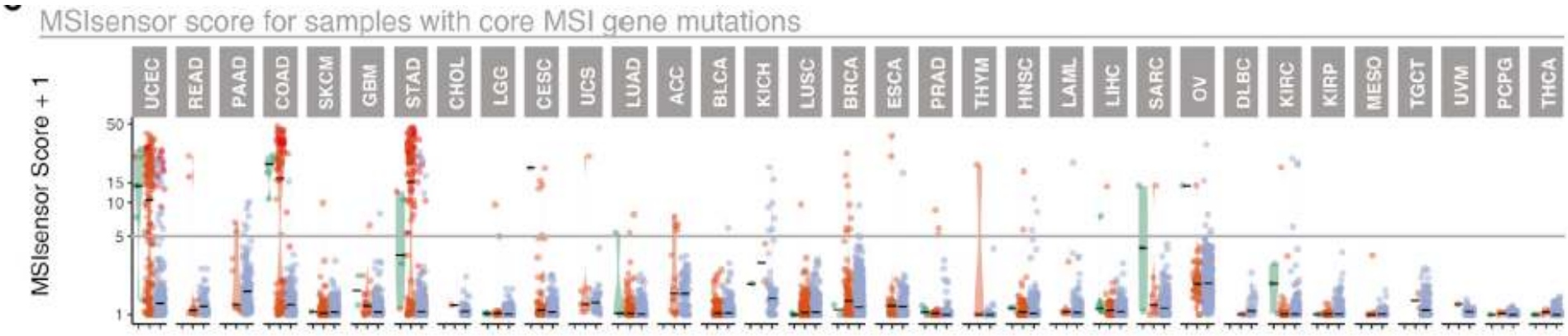
- Microsatellites are repeated sequences of DNA. These sequences can be made of repeating units of one to six base pairs in length.
- When the MMR proteins do not function normally, as in the case of MSI, this loop results in frame-shift mutations, either through insertions or deletions, yielding non-functioning proteins

Microsatellite instability phenotypes

- Many samples (250 out of 1,464) with nonsynonymous somatic mutations in DNA mismatch repair (MMR) genes have high microsatellite instability (MSI) status (MSIsensor score **>4**)

Microsatellite instability phenotypes

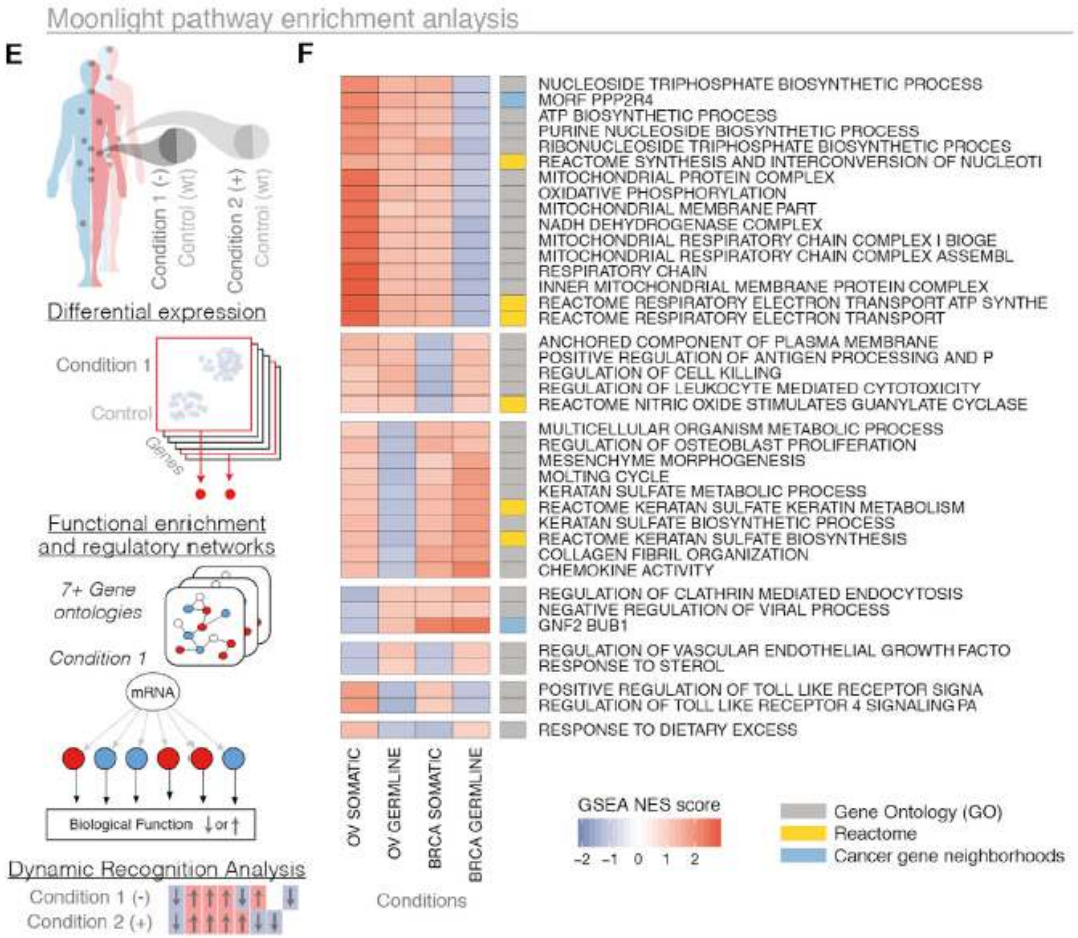
- Samples with germline pathogenic variants in MMR genes (18 out of 60) also have high MSI status.



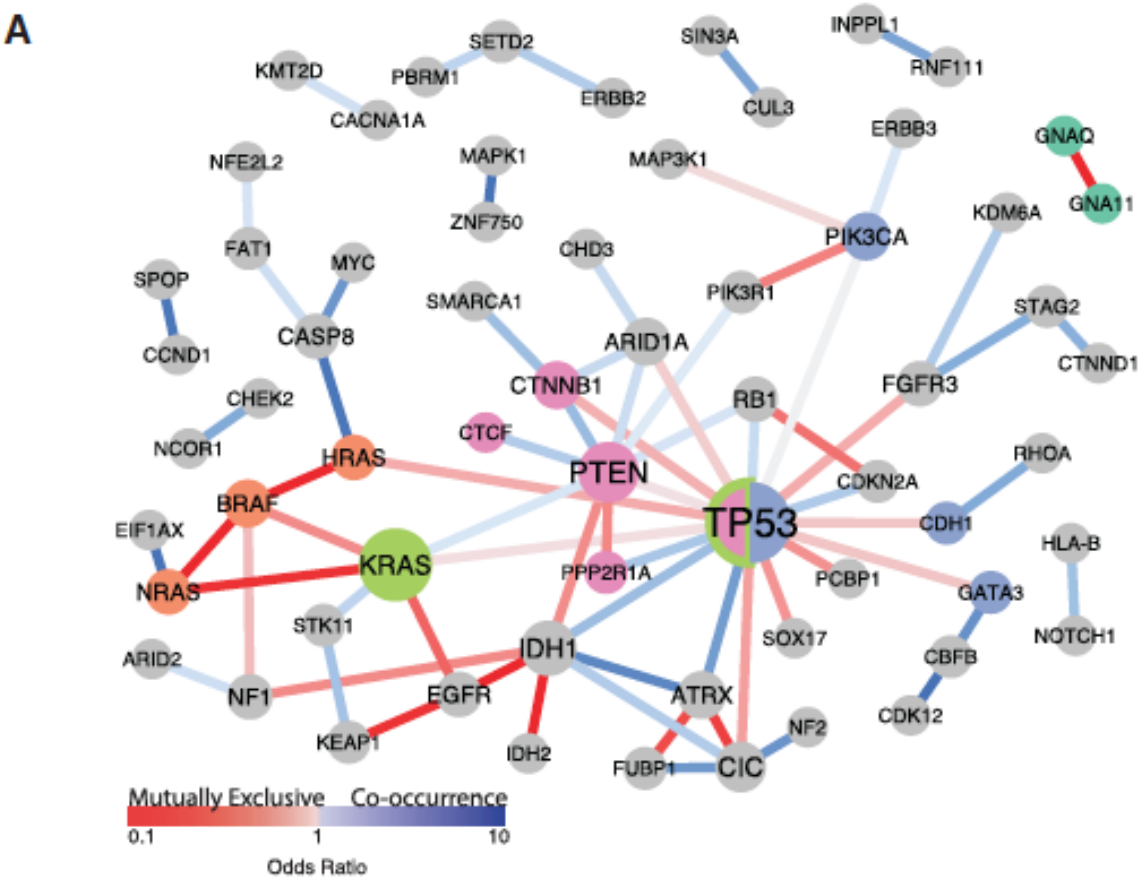
Diff Expression in Germline/Somatic BRCA

- several pathways that are differentially expressed depending on whether the mutations affecting BRCA1 and/or BRCA2 are somatic or germline

Diff Expression in Germline/Somatic BRCA



Somatic-Somatic interactions



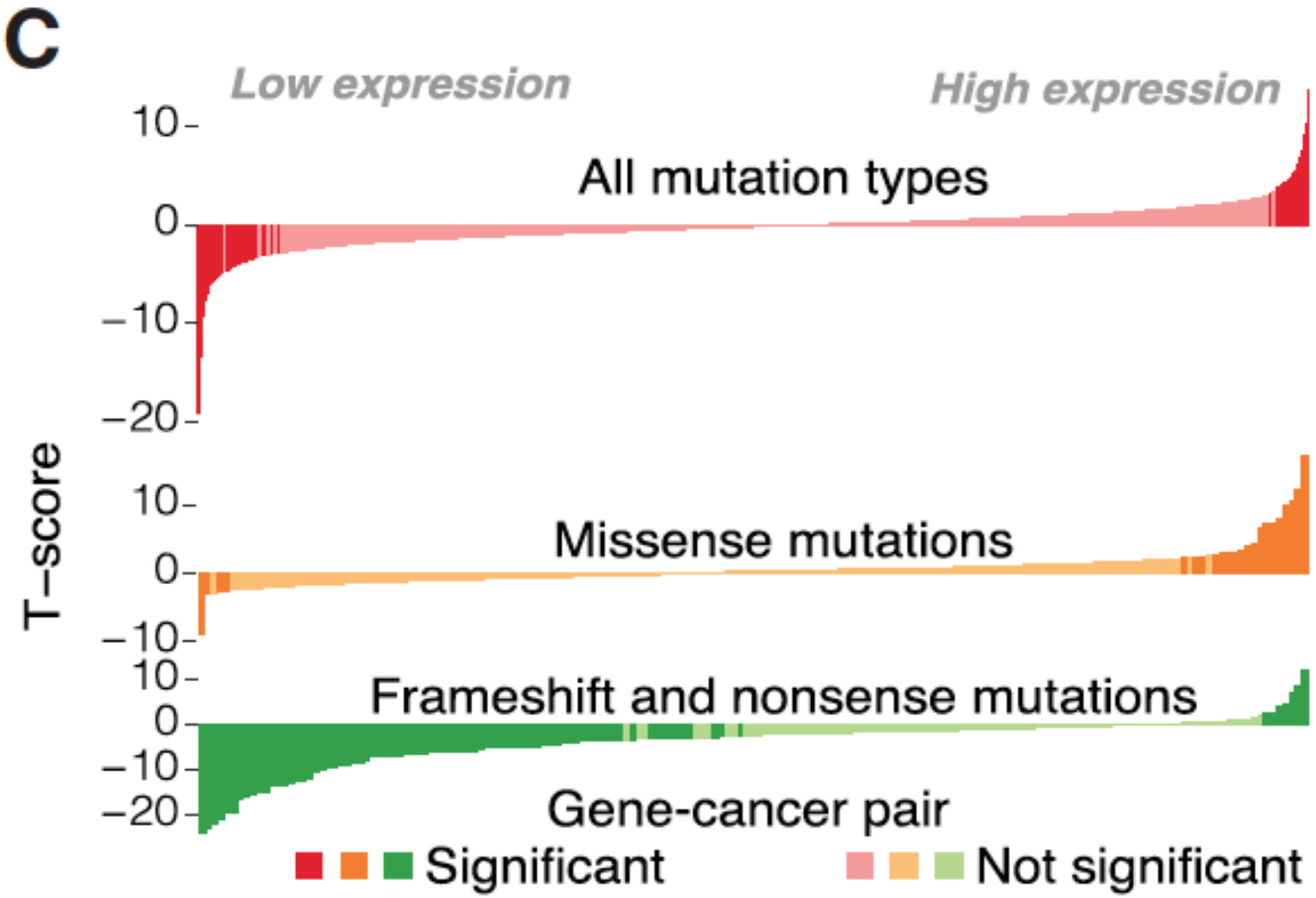
Somatic-Somatic interactions

- Some cancers require cooperation between gene networks − eg UCEC there are 2 mutually exclusive networks

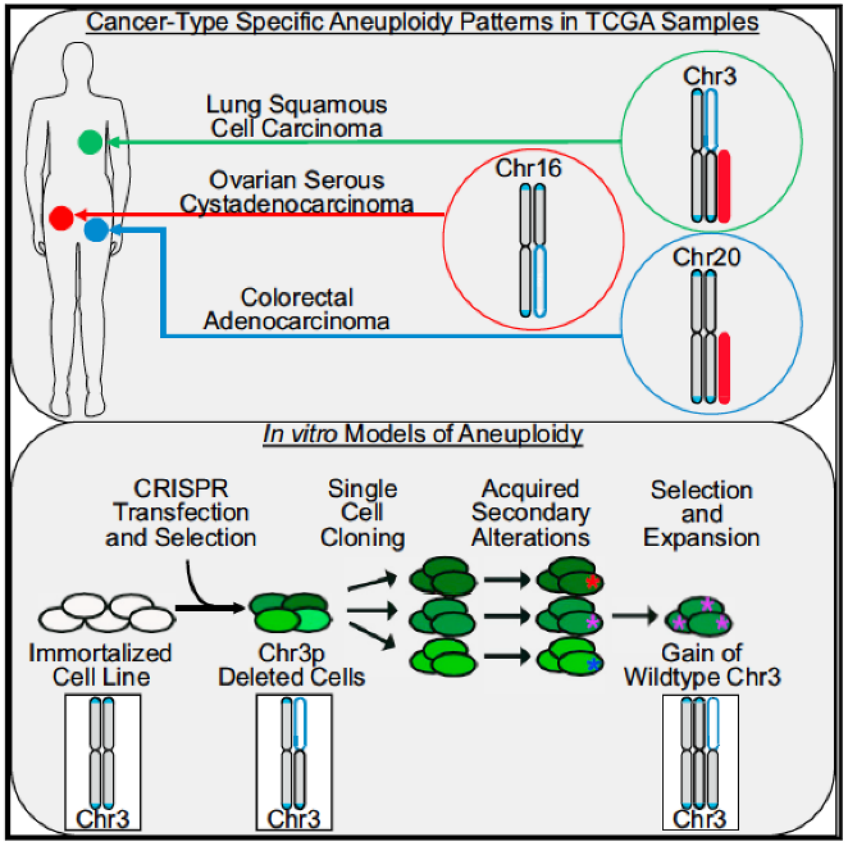
1) TP53 + PPP2R1A (a lot of copy number changes)

2) CTNNB1, PTEN, CTCF (CN low, hypermutated)

Effect of cis-mutations on expression

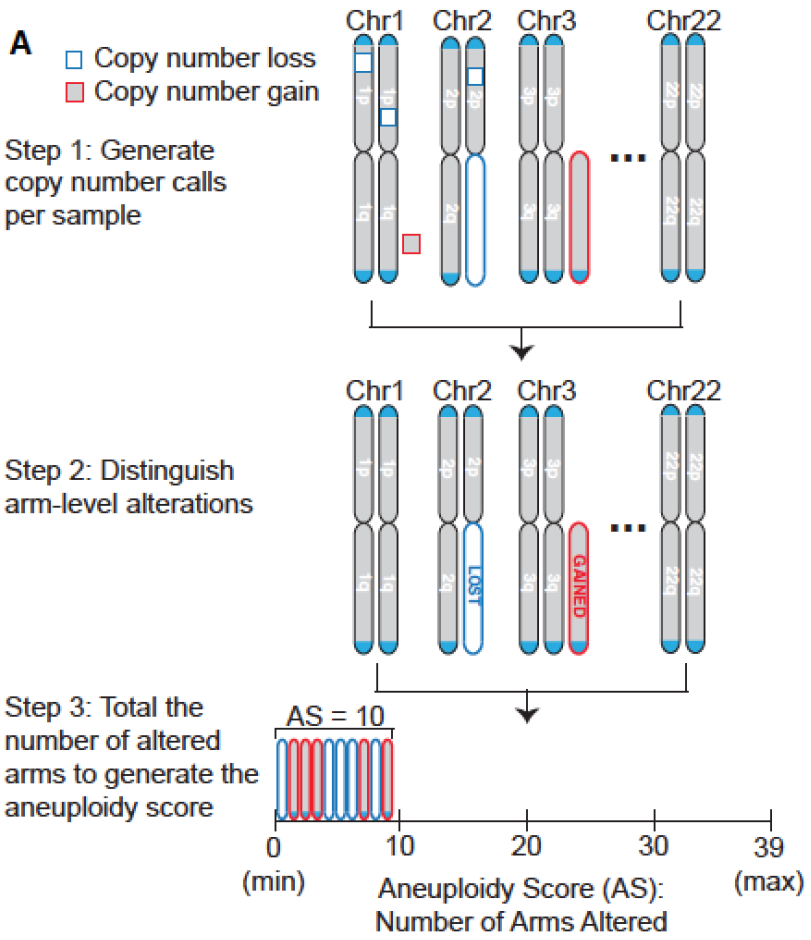


Cancer aneuploidy

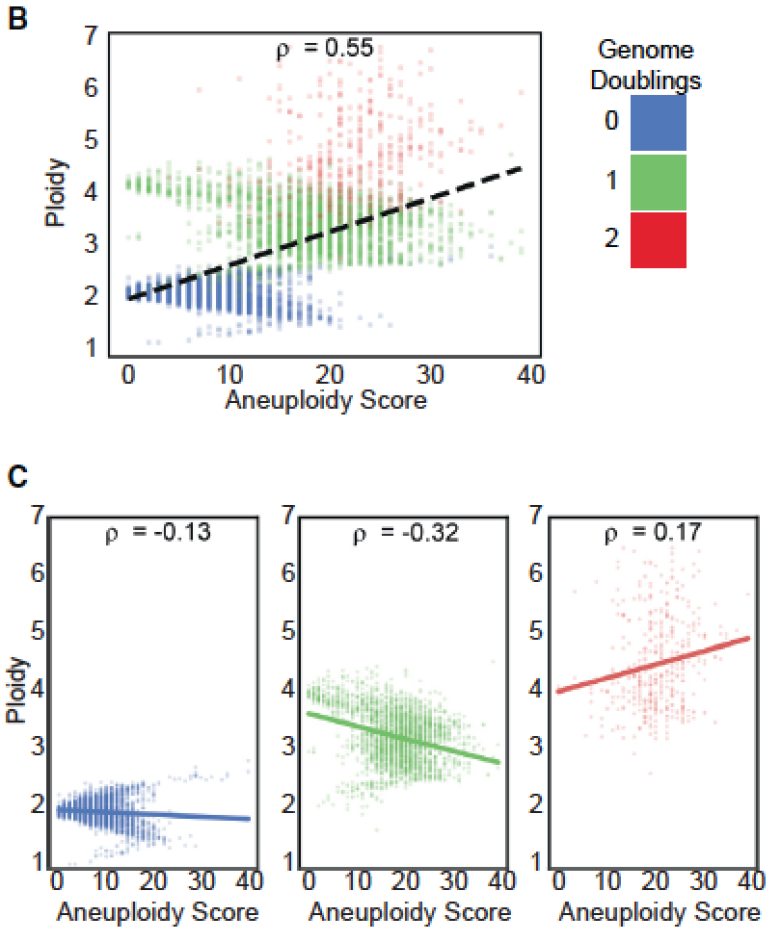


- Aneuploidy, whole chromosome or chromosome-arm imbalance, occurs in 88% of cancers
- Aneuploidy correlates with cell-cycle genes and anticorrelates with immune levels
- Patterns of aneuploidy alterations are tumor-type specific

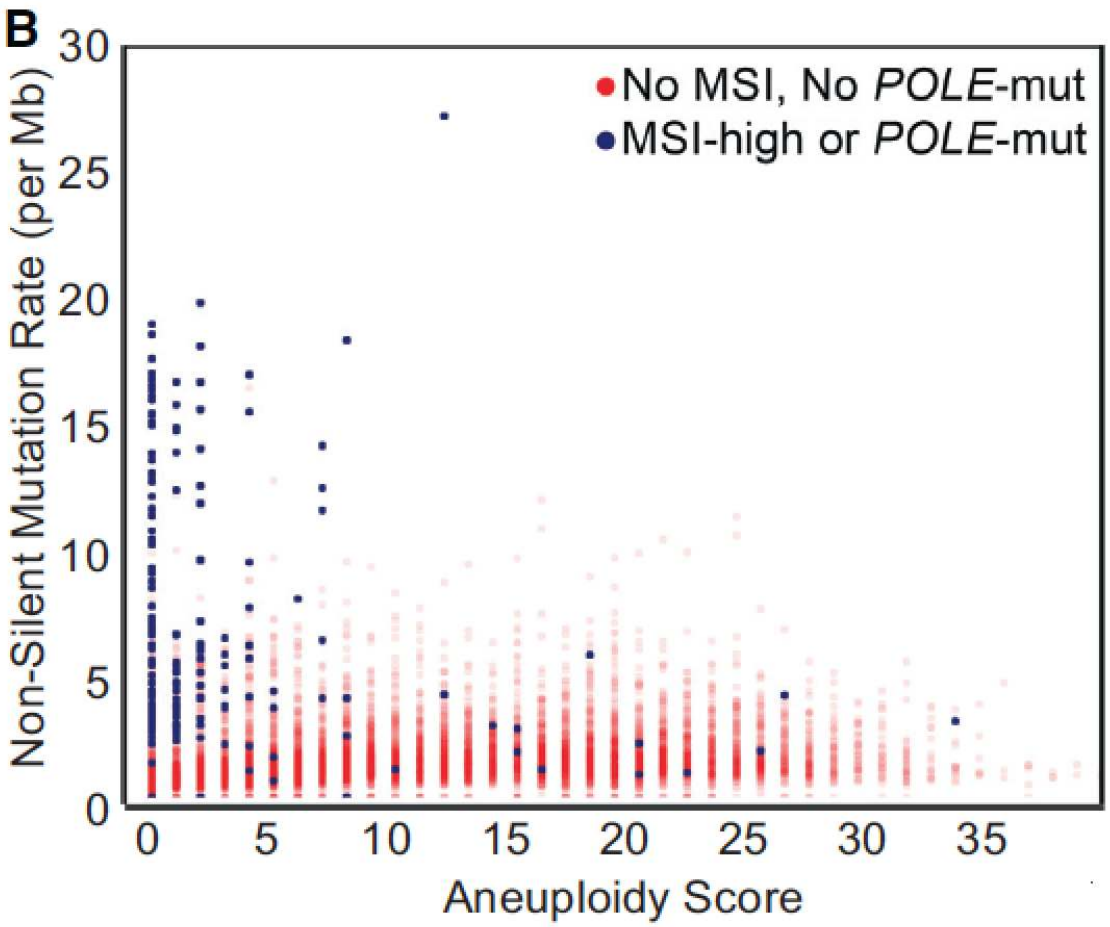
Cancer aneuploidy



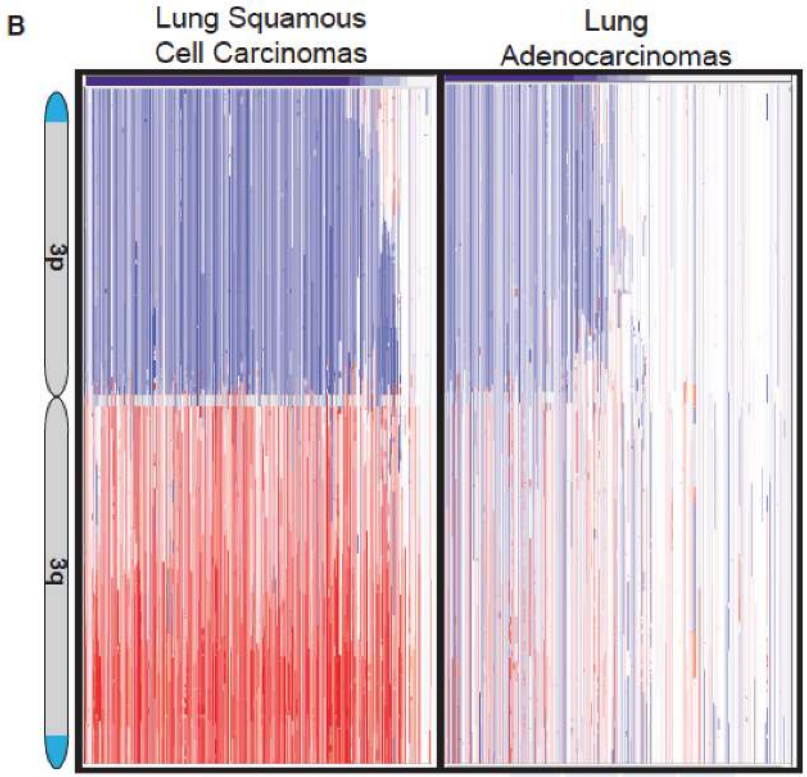
Cancer aneuploidy



Aneuploidy vs Somatic mutation rate



Different Aneuploidy profiles for lung cancers



	3q-	3q	3q+
3p-	-	31	245
3p	-	22	12
3p+	-	-	-

*p = .039

	3q-	3q	3q+
3p-	85	55	34
3p	6	177	11
3p+	1	3	9

N.S. p = .063

Working in consortium

Working in consortium

- *When your Server gets down or all your data were accidentally removed*
- *Deadlines – add 3-6 months to expected date!*
- *Communication: teleconferences*
- *Passwords renewal, permissions to access*
- *Efficient data sharing – speed, reliability, confidentiality*

Working in consortium

- *Different naming of the same samples in different working groups / labs*
- *Wrong/Missing Identifiers (eg wrong cancer type or population) – case: normal and somatic were actually swapped*
- *The same, but from clinicians*
- *Different labs - different library preparation (eg coverage depths after PCR-free and PCR-based WGS)*
- *Several tools can be used for the analysis – establishment of the best tool or generation of joint callset*
- *Multiple blacklist or outlier lists (every lab/group has its own and they do not completely overlap)*

Acknowledgements

- **Tobias Rausch, Jan Korbel** (EMBL)
- Ossowski's group, IMGAG: Axel Gschwind, **Marc Sturm**, Jakob Matthes, Francesc Muyas
- UKT: Franz Hilke, Christopher Schroeder