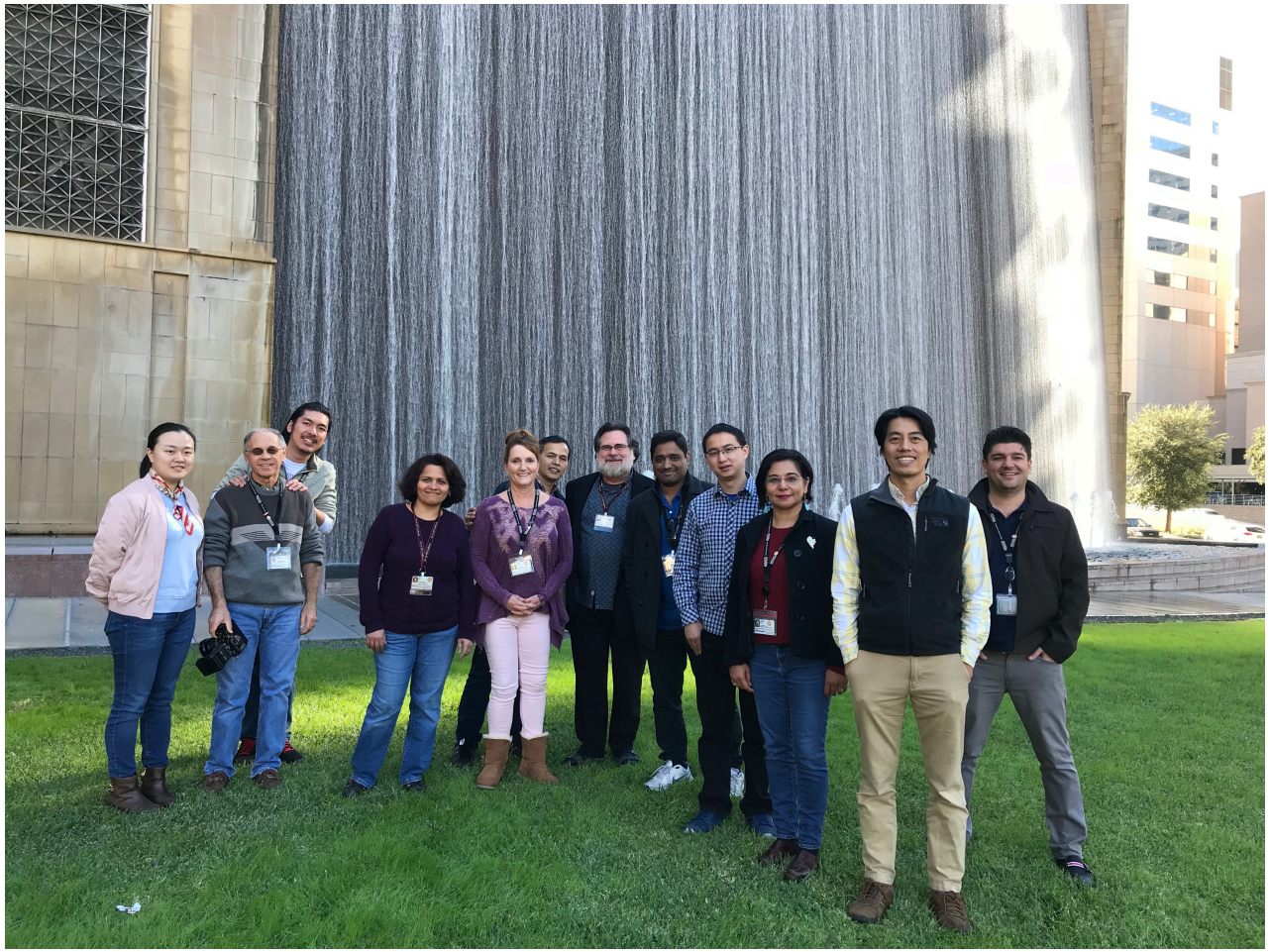
Using supercomputing resources in genomic research

THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center



Albino Bacolla

Research Investigator
MD Anderson Cancer Center
Dept. Mol. Cell. Oncology
Houston, TX – USA

P.I.
Prof. John A. Tainer

Analyzing cancer and genomic disease data

THE UNIVERSITY OF TEXAS MD Anderson Cancer Center

TCGA | COSMIC | Patients

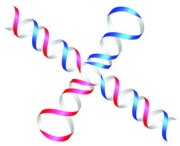
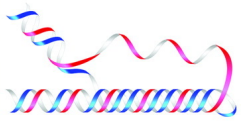
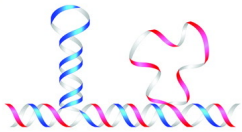
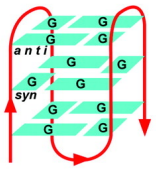
Transcriptome | Genome sequencing | Clinical data

TACC

Genomic features involved in disease etiology

https://github.com/abacolla

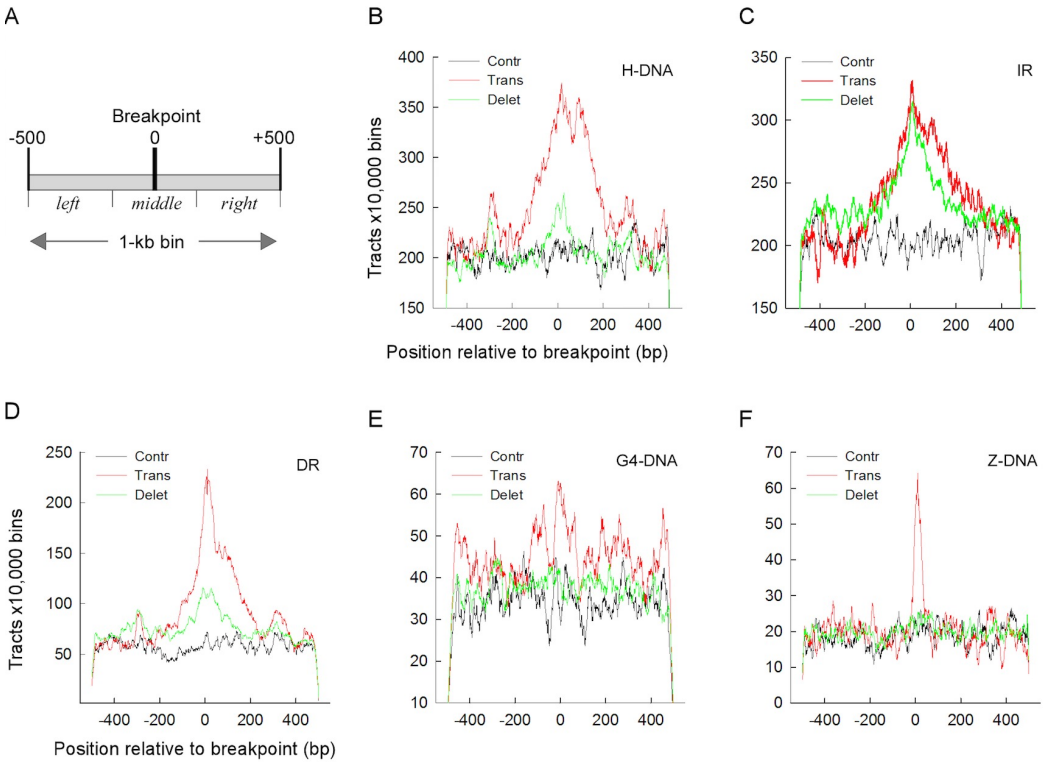
nonB DNA

Name	Conformation	General Seq. Requirements	Sequence
Cruciform		Inverted Repeats	TCGGTACCGA AGCCATGGCT
Triplex		$(R•Y)_n$ Mirror Repeats	AAGAGGGGAGAA TTCTCCCCTCTT
Slipped (Hairpin) Structure		Direct Repeats	TCGGTTCGGT AGCCAAGCCA
Tetraplex		Oligo $(G)_n$ Tracts	$AG_3(T_2AG_3)_3$ single strand
Left-handed Z - DNA		$(YR•YR)_n$	CGCGTGCGTGTG GCGCACGCACAC

nonB DNA triggers genetic instability

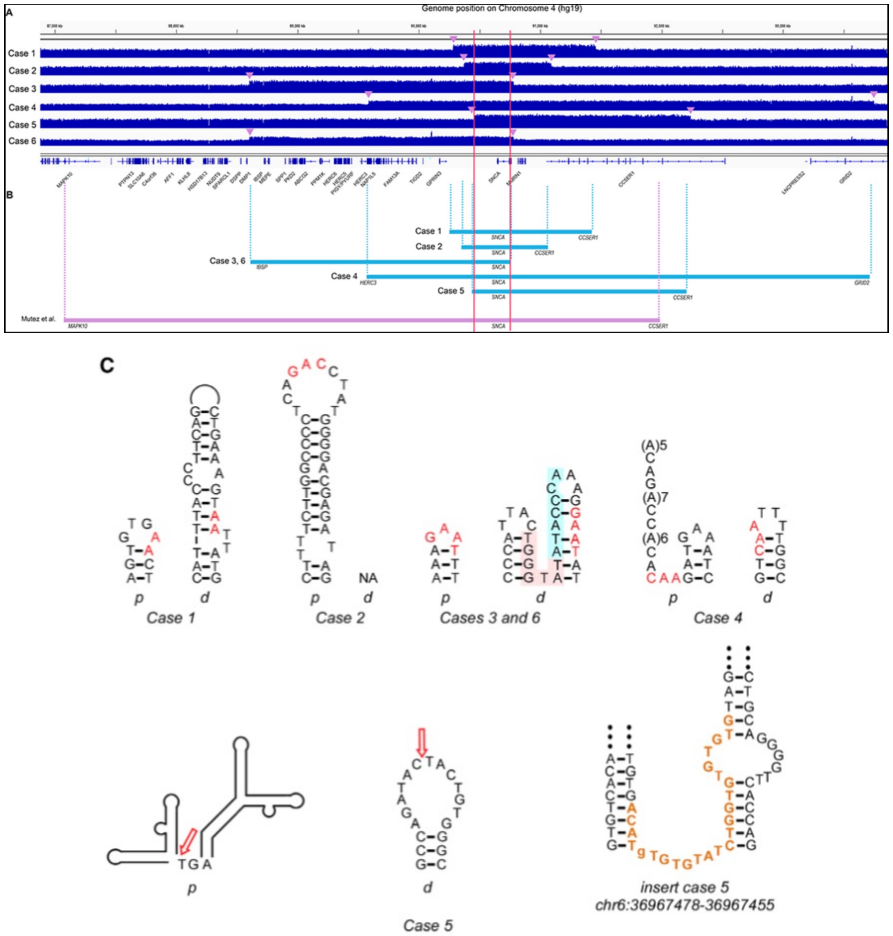
THE UNIVERSITY OF TEXAS MD Anderson Cancer Center

nonB DNA-forming repeats are associated with translocations in cancer genomes



Nucleic Acids Res 2016

Hairpin-loops are sites of duplications predisposing to Parkinson's disease



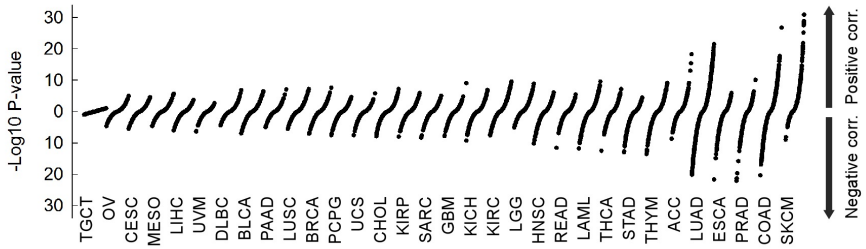
Mov Disord 2020

TACCSTER 2021

Top genes associated with poor cancer survival

THE UNIVERSITY OF TEXAS MD Anderson Cancer Center

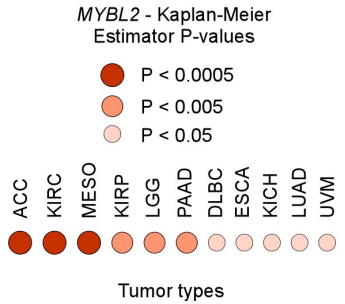
Gene expression – mutation load profiles identify tumor susceptibility genes

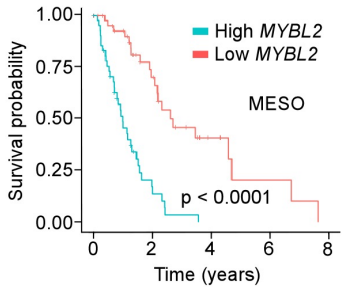


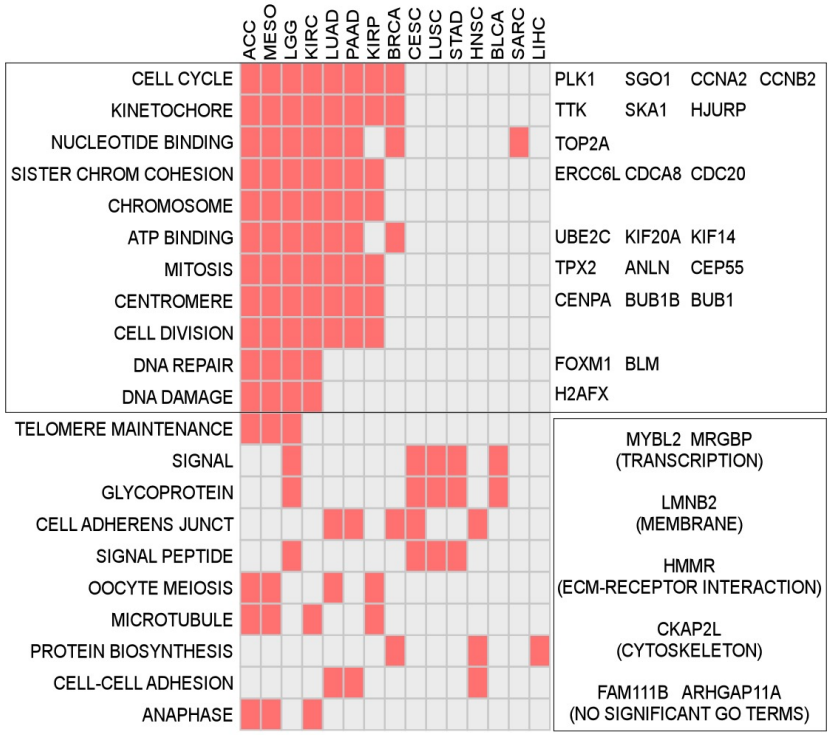
Tumor type	Gene	GO Term	Corr. P-value	KM P-value	COSMIC CGCC
Top 10 Positive Correlations					
LUAD	*BORA*	Cell Cycle	1.4E-20	0.026	N
LUAD	*CENPA*	Cell Cycle	4.6E-20	0.012	N
LUAD	*EME1*	Fanconi Anemia Pathway	3.6E-21	ns	N
LUAD	*HLTF*	BRCA1 Pathway	3.1E-22	ns	N
LUAD	*KIF2C*	Cell Cycle	2.4E-20	0.048	N
LUAD	*KIFC1*	Cytoskeletal Signaling	2.3E-21	0.026	N
LUAD	*MCM10*	Cell Cycle	2.6E-20	0.014	N
LUAD	*MYBL2*	EGF/EGFR Signaling	1.4E-20	0.024	N
LUAD	*SPAG5*	PI3K/Akt Signaling	9.5E-21	0.009	N
LUAD	*TPX2*	Cell Cycle	9.6E-21	0.009	N
Top 10 Negative Correlations					
COAD	*SPIRE2*	TGF-Beta Pathway	4.6E-21	ns	N
ESCA	*MLH1*	Cell Cycle	2.1E-22	ns	Y
LUAD	*KIAA0141*	--	7.7E-20	ns	N
LUAD	*KCNQ1*	SIDS Pathway	7.2E-21	ns	N
LUAD	*PHKB*	Metabolism	2.4E-20	ns	N
LUAD	*PIGR*	Innate Immunity	3.0E-20	0.034	N
LUAD	*SPATA18*	DNA Damage Response	7.5E-21	0.024	N
PRAD	*GCA*	Innate Immunity	1.9E-20	ns	N
PRAD	*MARK1*	Cytoskeletal Signaling	4.7E-22	ns	N
PRAD	*SDHAF3*	Complex II Assembly	7.1E-23	ns	N

Prog Biophys Mol Biol 2019

30 top genes associated with poor cancer survival when overexpressed







Mol Cell 2021

TACCSTER 2021

THANK YOU

THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center