Cancer genome sequencing analysis, storage, discovery and delivery

J. Zachary Sanborn[1], Charles M. Vaske[1], Stephen C. Benz[1]
[1]Five3 Genomics, Santa Cruz, CA

Abstract #3982

Introduction

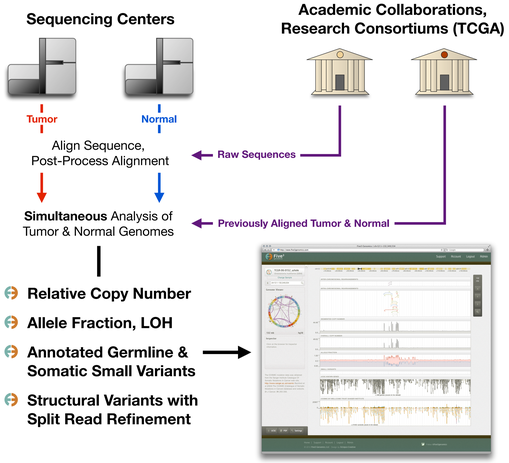
Genome instability and structural rearrangement is a distinctive hallmark of the cancer genome. With next-generation sequencing technologies, our ability to measure structural rearrangements that occur through tumorigenesis and progression has significantly improved, while creating an urgent need for rearrangement discovery, analysis, and visualization methods to aid our comprehension of these events.

Our sequencing analysis pipeline streamlines the discovery of individual tumor's mutations, small indels, copy number alterations, allele-specific amplifications and deletions, and genomic rearrangements. Rearrangements are refined to breakpoint precision using unmapped, putative split reads found in the vicinity of the breakpoint when available. Results are presented in an interactive, web-based genome browser that provides analysis and visualization of both high-level, processed results as well as the raw data from which they were derived.

The sequencing analysis pipeline was used to discover high-confident, small- and large-scale somatic events in 17 whole genome glioblastoma multiforme (GBM) tumor samples from The Cancer Genome Atlas (TCGA) project, using their matched normal sequences to identify somatic rearrangements. Among many interesting structural aberrations identified in these samples, we found two tumors with complicated rearrangement patterns in regions of extreme amplification that could be assembled to construct circular double minute chromosomes at base-level precision. Evidence of breakpoints specific to the double minute were found in blood sequencing data, raising the possibility that patient-specific PCR-based assays could be developed to quantify the presence of somatic rearrangements to use as a proxy in monitoring the progression of brain tumors.

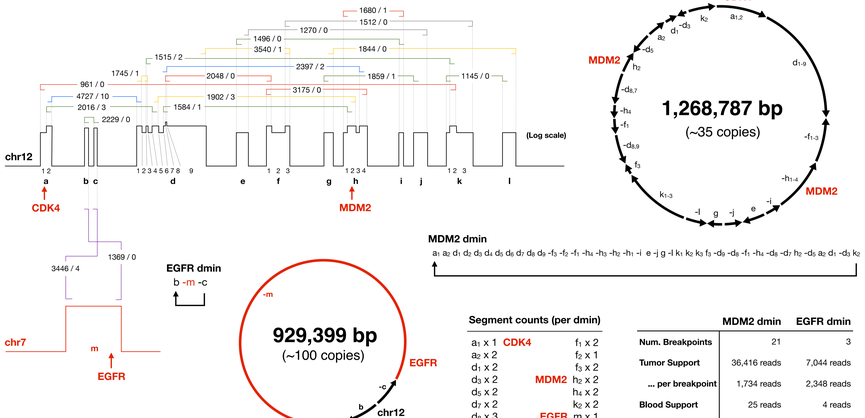
Also, four GBM tumors were found exhibiting EGFR amplifications and rearrangements indicating the presence of the EGFRvIII mutant gene, whereby exons 2-6 of EGFR are deleted. Comparing the read support of the EGFRvIII-associated breakpoints to the amount of normally-mapped reads in the neighborhood suggest that the EGFRvIII mutant emerges after the amplification of wild-type EGFR, existing as a fraction of the total number of EGFR copies in the tumor.

Sequencing Analysis Pipeline



Double Minute Assembly and Detection in Blood in TCGA Glioblastoma Multiforme Samples

TCGA-06-0152

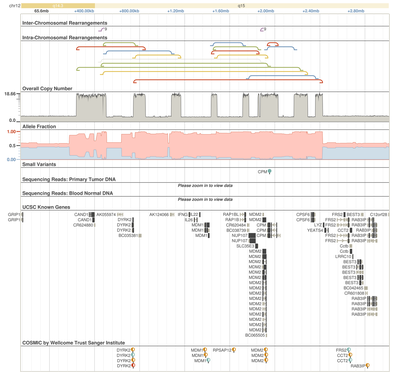


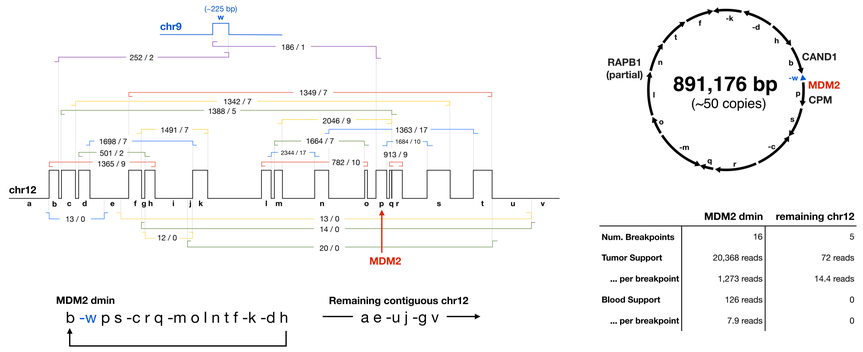
Segment counts (per dmin)

a_1 x 1	CDK4	f_1 x 2
a_2 x 2		f_2 x 1
d_1 x 2		f_3 x 2
d_3 x 2	MDM2	h_2 x 2
d_5 x 2		h_3 x 2
d_7 x 2		k_2 x 2
d_8 x 3	EGFR	m x 1
d_9 x 2	(All others at single copy)	

	MDM2 dmin	EGFR dmin
Num. Breakpoints	21	3
Tumor Support	36,416 reads	7,044 reads
... per breakpoint	1,734 reads	2,348 reads
Blood Support	25 reads	4 reads
... per breakpoint	1.19 reads	1.33 reads

TCGA-06-0648





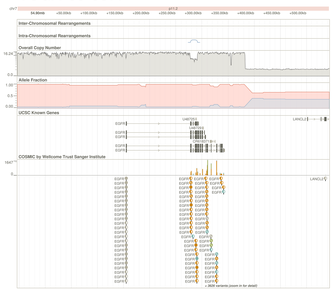
	MDM2 dmin	remaining chr12
Num. Breakpoints	16	5
Tumor Support	20,368 reads	72 reads
... per breakpoint	1,273 reads	14.4 reads
Blood Support	126 reads	0
... per breakpoint	7.9 reads	0

Probable Double Minutes Identified in Other Cancers

Sample	Cancer Type	Data Type	Amplified Oncogene(s)
TCGA-06-0152	Glioblastoma Multiforme (GBM)	Whole Genome Shotgun	MDM2, CDK4, EGFR
TCGA-06-0648	Glioblastoma Multiforme (GBM)	Whole Genome Shotgun	MDM2
TCGA-14-0786	Glioblastoma Multiforme (GBM)	Whole Genome Shotgun	EGFR
TCGA-CS-4941	Lower Grade Glioma (LGG)	Exome Capture	MDM2, CDK4, EGFR
TCDA-DU-6403	Lower Grade Glioma (LGG)	Exome Capture	MDM4, EGFR
TCGA-CS-5395	Lower Grade Glioma (LGG)	Exome Capture	MDM4
TCGA-66-2756	Lung Squamous Cell Carcinoma (LUSC)	Whole Genome Shotgun	TEK
TCGA-CG-4300	Stomach Adenocarcinoma (STAD)	Exome Capture	ERBB2, MAPK1
TCGA-CG-4466	Stomach Adenocarcinoma (STAD)	Exome Capture	ERBB2, CDK6
TCGA-CG-4475	Stomach Adenocarcinoma (STAD)	Exome Capture	MDM2, CDK6
TCGA-C4-A0F0	Bladder Urothelial Carcinoma (BLCA)	Exome Capture	MYC, KRAS, BIRC2, YAP1
TCGA-DK-A1A5	Bladder Urothelial Carcinoma (BLCA)	Exome Capture	ERBB2 (mutant)
TCGA-C4-A0F6	Bladder Urothelial Carcinoma (BLCA)	Exome Capture	MDM2
Unpublished	Melanoma	Low Pass Whole Genome	YAP1

Quantifying EGFRvIII in Glioblastoma Multiforme

TCGA-06-0877



Sample	EGFRvIII Support	Amplicon Coverage	EGFRvIII Percentage
TCGA-06-0877	331	~950	35%
TCGA-06-0145	1275 (848+427)	~5900	22%
TCGA-06-0214	321 (187+102+32)	~500	64%
TCGA-14-0786	8	~3400	0.2%

Of the 11 TCGA GBMs featuring high level amplification of EGFR, 4 tumors exhibit the EGFRvIII mutant. Interestingly, two tumors have multiple breakpoints, where each could independently create the EGFRvIII mutant.

The read support of the EGFRvIII breakpoints relative to the average coverage of the EGFR amplicon suggests that the EGFRvIII mutant emerged post-amplification.

Acknowledgments

The results published here are in whole or part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at “http://cancergenome.nih.gov”.

© 2012 Five3 Genomics. All Rights Reserved.