Heliyon

ELSEVIER

Received:
4 November 2016
Revised:
8 March 2017
Accepted:
23 March 2017

Heliyon 3 (2017) e00277

Progression inference for somatic mutations in cancer

Leif E. Peterson [a,b,c,d,e,*], **Tatiana Kovyrshina** [a,f]

[a] *Center for Biostatistics, Houston Methodist Research Institute, Houston, TX 77030, USA*

[b] *Dept. of Healthcare Policy and Research, Weill Cornell Medical College, Cornell University, New York, NY 10065, USA*

[c] *Dept. of Biostatistics, School of Public Health, University of Texas – Health Science Center, Houston, TX 77030, USA*

[d] *Dept. of Medicine, Baylor College of Medicine, Houston, TX 77030, USA*

[e] *Dept. of Neuroscience and Experimental Therapeutics, Texas A&M University Health Science Center, College Station, TX 77843, USA*

[f] *Dept. of Mathematics and Statistics, University of Houston – Downtown, Houston, TX 77002, USA*

* Corresponding author.
E-mail address: lepeterson@houstonmethodist.org (L.E. Peterson).

Abstract

Computational methods were employed to determine progression inference of genomic alterations in commonly occurring cancers. Using cross-sectional TCGA data, we computed evolutionary trajectories involving selectivity relationships among pairs of gene-specific genomic alterations such as somatic mutations, deletions, amplifications, downregulation, and upregulation among the top 20 driver genes associated with each cancer. Results indicate that the majority of hierarchies involved TP53, PIK3CA, ERBB2, APC, KRAS, EGFR, IDH1, VHL, etc. Research into the order and accumulation of genomic alterations among cancer driver genes will ever-increase as the costs of nextgen sequencing subside, and personalized/precision medicine incorporates whole-genome scans into the diagnosis and treatment of cancer.

Keywords: Oncology, Cancer research, Genetics, Computational biology

1. Introduction

Tumors evolve through a multi-step mutagenic process for which cells acquire resistance to apoptotic and antiproliferative signals, self-sufficiency in growth

http://dx.doi.org/10.1016/j.heliyon.2017.e00277

signals, and unlimited proliferation potential [1]. Tumor cells also activate glycolysis and deactivate oxidative phosphorylation (Warburg effect), evade immune detection by decreasing pH with lactic acid, increase ROS, exhibit chromosome aberrations and telomere shortening, negotiate life support from stromal cells, activate invasion and motility, and coordinate neovascularization [2, 3]. Cancer avoids extinction through genetic selection because its onset commonly occurs at older ages – beyond reproductive years. If cancer mortality rates decreased with age, cancer would be selected out due to zero fitness within several generations. Genomic instability and the seeding of a mutator phenotype is another hallmark, which can receive positive feedback from clonal expansion of a single pathogenic mutation in key stability pathways, such as DNA repair, replication, or cell-cycle checkpoints [4]. Within several DNA replications, a cascade of mutations ensues, sacrificing stability in the entire genome.

While tumors evolve as a consequence of the accumulation of somatic lesions, it is unclear how mutated genes interact to generate the phenotypic hallmarks of cancer. The accumulation of somatic mutations in tumors is rarely detectable during early stages of development, and difficult to detect at later stages because of the high mutational load [5]. From a population genetics standpoint, homozygous mutations in both alleles of mismatch-repair genes result in a 100-fold increase in mutation rates [6, 7, 8, 9], but have zero fitness and won't survive [10]. On the other hand, heterozygosity confers a 5–10 fold increase in mutations [11, 12], which in the presence of activated DNA repair pathways will result in a tolerable selective advantage.

The foundation of computational biology for tumor progression was introduced by Vogelstein et al., who introduced the notion of trajectories of progression during which tumor progression involved activation of mutations which progressed toward subpopulations of cells having clonal origin [13]. Attolini et al. [14] introduced a bioinformatic approach called RESIC (Retracing the Evolutionary Steps in Cancer) for deducing the temporal sequence of genetic events during tumorigenesis from cross-sectional genomic data of tumors. Several nextgen sequencing datasets were employed consisting of 70 advanced colorectal cancers, 91 primary human glioblastomas, and 57 acute myelogenous leukemias. In the colorectal cancers, RESIC accurately predicted the temporal sequence of APC, KRAS, and TP53 mutations, which was in agreement with order determined through analyzing tumors at different stages of colon cancer formation. For GBM tumors, it was observed that TP53 was the first gene showing selective pressure for somatic mutations, and in the AML samples, JAK2 and TET2 were the first genes to exhibit selective pressure. Youn and Simon [15] employed a highly-parameterized likelihood-based approach for inferring order of mutational steps in genes, using nextgen sequencing data for 188 lung cancers [16] and 133 colorectal tumors [17]. Results indicated that

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

KRAS, EGFR, and TP53 were among the first genes showing selective pressure in the lung tumors, while for colorectal tumors selective pressure first appeared in APC, KRAS, and TP53. Lecca et al. [18] describe the TO-DAG (Timed Oncogenetic Directed Acyclic Graph) algorithm applied to 74 human prostate cancer samples that include point mutations, copy number losses and gains, and rearrangements. Gerstung et al. [19] developed a computational approach to infer TO-DAGs from human tumor mutation data, and determined that TO-DAG shows high performance scores on synthetic data and recognizes mutations in gatekeeper tumor suppressor genes as a trigger for several downstream mutational events in the human tumor data. The models generated by TO-DAG have been extensively compared with the trees and the graphs inferred by most recent tools representative of the RESIC [14] and CT-CBN [19]. Kang et al. [20] introduced a parametric approach to estimate the sequential order of gene mutations during tumorigenesis from genome sequencing data based on a Markov chain model as TOMC (Temporal Order based on Markov Chain). TOMC revealed that tumor suppressor genes tend to be mutated ahead of oncogenes, which are considered as important events for key functional loss and gain during tumorigenesis. A larger workflow approach was used to develop CAPRI [21], which generates acyclic graphs to capture branched, independent and confluent evolution via bootstrapping, shrinkage, maximum likelihood, and regularization. Caravagna et al. [22] reported on the PiCnIc pipeline, which incorporates CAPRI, and is a versatile, modular, and customizable pipeline to extract ensemble-level progression models from cross-sectional sequenced cancer genomes.

This investigation focuses on development of cancer progression models derived from cross-sectional genomic data. The models employed focus on selectivity relationships between mutational events, such that event j selects for event k, resulting in a weighted directed acyclic graph (WDAG) of alterations representing the accumulation of events under selective pressure during cancer progression. We also introduce a permutation-based affinity metric (PBAM) approach, which is an iterative learning method that combines multi-tumor co-occurrence event statistics and within-tumor order permutations to extract affinity relationships between all possible pairs of events. The affinity relationships are then filtered by probabilistic causation conditions based on temporality and probability raising, which are not admissible. The temporality condition assumes that, if event j occurs earlier than event k, then event j will occur more frequently than event k, that is $P(j) > P(k)$. Whereas, the probability raising condition assumes that the occurrence of event k increases the occurrence of event j, i.e., $P(k|j) > P(k|\bar{j})$. Selectivity relationships stem from the idea that, during tumor clonal evolution, there is a selective advantage of certain genomic alterations which increase the probability of subsequent downstream events. The pattern that emerges is one for which greater frequencies of driver events subsume a mixture of later events, above some background noise level. Estimating selectivity relationships among

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

genomic alterations is important because early events may represent important therapeutic targets, while late mutations may play a role in metastases. Investigation of selectivity relationships can also deepen our understanding of tumorigenesis at the genome sequence level, and may help to elucidate functional roles of genomic alterations.

2. Methods

Cancer data

The cross-sectional data used in this investigation were derived from nextgen sequencing of tumors in the The Cancer Genome Atlas (TCGA) [23]. We investigated mutational events in 10 cancers (Table 1) for which nextgen sequencing and RNA-seq expression data were available from cBio-Portal (http://www.cbioportal.org) [24, 25].

Consensus driver genes

We obtained a consensus of the top 20 driver genes for each cancer considered from the DriverDB database [26], based on identification by at least 2 tools (default) for each cancer, since requesting a higher consensus could result in fewer than 20 driver genes for some cancers. DriverDB assembles together lists of the top ranked driver genes determined from the use of 15 packages, including ActiveDriver, Dendrix, MDPFinder, Simon, NetBox, OncodriveFM, MutSigCV, MEMo, CoMDP, DawnRank, DriverNet, e-Driver, iPAC, MSEA, and OncodriveCLUST. Table 1 lists the cancer types, sample size, and descriptions of the driver genes used, including chromosome location, gene ontology nomenclature, and molecular pathway.

Driver events: mutations, deletions, amplifications, downregulation, and upregulation

Data for all somatic mutations were obtained directly from cBio-Portal. We also acquired high-confidence deletions and amplifications from cBio-Portal, where a deletion was defined as full homozygous loss with a GISTIC score [27] of -2, and an amplification was defined as high-level gain with a GISTIC score of 2. Low-level deletions (heterozygous loss) and low-level gain (low-level amplifications) with GISTIC scores of -1 and 1, respectively, were not used. Downregulation and upregulation of RNA-Seq based expression was also obtained from cBio-Portal, where Z-scores less than -1.96 were assumed to represent downregulation, and Z-scores greater than 1.96 were assumed to represent upregulation. Since there were 20 driver genes considered per cancer and 5 "driver events" considered per gene

Table 1. Genomic annotation for driver genes employed, including Gene Ontology nomenclature and molecular pathway.

Cancer	Symbol	Chr	GO:Biol. Process	GO:Cell. Component	GO:Mol. Function	Pathway
Acute Myeloid Leukemia	CEBPA	19	urea cycle	nucleus	RNA polymerase II core promoter proximal region sequence-specific DNA binding	Non-alcoholic fatty liver disease (NAFLD)
$n = 200$	DIS3	13	rRNA processing	nuclear exosome (RNase complex)	3'–5'-exoribonuclease activity	RNA degradation
	DNMT3A	2	negative regulation of transcription from RNA polymerase II promoter	chromosome	DNA binding	Cysteine and methionine metabolism
	KIT	4	MAPK cascade	acrosomal vesicle	protease binding	Ras signaling pathway
	KRAS	12	MAPK cascade	intracellular	GTPase activity	MAPK signaling pathway
	RAD21	8	double-strand break repair	nuclear chromosome	transcriptional activator activity	Cell cycle
	SUZ12	17	negative regulation of transcription from RNA polymerase II promoter	sex chromatin	RNA polymerase II core promoter sequence-specific DNA binding	
	U2AF1	21	mRNA splicing	nucleoplasm	nucleotide binding	Spliceosome
	WT1	11	negative regulation of transcription from RNA polymerase II promoter	nucleus	transcriptional activator activity	Transcriptional misregulation in cancer
	FLT3	13	leukocyte homeostasis	nucleus	transmembrane receptor protein tyrosine kinase activity	Cytokine–cytokine receptor interaction
	IDH1	2	glyoxylate cycle	cytoplasm	magnesium ion binding	Citrate cycle (TCA cycle)
	IDH2	15	carbohydrate metabolic process	mitochondrion	magnesium ion binding	Citrate cycle (TCA cycle)
	NRAS	1	MAPK cascade	Golgi membrane	GTP binding	MAPK signaling pathway
	NPM1	1, 10, 15, 2, 5, 8	DNA repair	nucleus	nucleic acid binding	
	PHACTR1	6	actomyosin structure organization	nucleus	actin binding	
	PTPN11	12, 3, 4	DNA damage checkpoint	nucleus	phosphoprotein phosphatase activity	Ras signaling pathway
	PTPRT	20	protein dephosphorylation	plasma membrane	protein tyrosine phosphatase activity	

(*continued on next page*)

Table 1. (*continued*)

Cancer	Symbol	Chr	GO:Biol. Process	GO:Cell. Component	GO:Mol. Function	Pathway
	RUNX1	21	ossification	nucleus	regulatory region DNA binding	Pathways in cancer
	TET2	4	kidney development	nucleus	sulfonate dioxygenase activity	
	TP53	17	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II regulatory region sequence-specific DNA binding	MAPK signaling pathway
Brain Lower Grade Glioma	ATRX	X	DNA repair	nuclear chromosome	DNA binding	
n = 530	SETD2	3	angiogenesis	nucleus	protein binding	Lysine degradation
	TMEM189-UBE2V1	20, 3	protein polyubiquitination	ubiquitin ligase complex	ubiquitin protein ligase binding	
	CACNA1S	1	calcium ion transport	cytoplasm	voltage-gated calcium channel activity	MAPK signaling pathway
	CIC	19	negative regulation of transcription from RNA polymerase II promoter	nucleus	DNA binding	
	CDC27	17	metaphase/anaphase transition of mitotic cell cycle	nucleus	protein binding	Cell cycle
	CHEK2	15, 22, Y	DNA damage checkpoint	chromosome	protein kinase activity	Cell cycle
	EGFR	7	MAPK cascade	Golgi membrane	glycoprotein binding	MAPK signaling pathway
	IDH1	2	glyoxylate cycle	cytoplasm	magnesium ion binding	Citrate cycle (TCA cycle)
	IDH2	15	carbohydrate metabolic process	mitochondrion	magnesium ion binding	Citrate cycle (TCA cycle)
	KRTAP1-5	17		keratin filament		
	NF1	17	MAPK cascade	nucleus	GTPase activator activity	MAPK signaling pathway
	NOTCH1	9	negative regulation of transcription from RNA polymerase II promoter	Golgi membrane	core promoter binding	Dorso-ventral axis formation
	PTEN	10, 9	regulation of cyclin-dependent protein serine/threonine kinase activity	extracellular region	magnesium ion binding	Inositol phosphate metabolism

(*continued on next page*)

Table 1. (*continued*)

Cancer	Symbol	Chr	GO:Biol. Process	GO:Cell. Component	GO:Mol. Function	Pathway
	PIK3CA	3	angiogenesis	intracellular	protein serine/threonine kinase activity	Inositol phosphate metabolism
	PIK3R1	5	cellular glucose homeostasis	nucleus	transmembrane receptor protein tyrosine kinase adaptor activity	ErbB signaling pathway
	PLCG1	20	activation of MAPKK activity	ruffle	phosphatidylinositol phospholipase C activity	Inositol phosphate metabolism
	PTPN11	12, 3, 4	DNA damage checkpoint	nucleus	phosphoprotein phosphatase activity	Ras signaling pathway
	STK19	6	protein phosphorylation	nucleus	protein serine/threonine kinase activity	
	TP53	17	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II regulatory region sequence-specific DNA binding	MAPK signaling pathway
Breast Invasive Carcinoma	AKT1	14	protein import into nucleus	nucleus	protein kinase activity	MAPK signaling pathway
$n = 1105$	ARID1A	1	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	DNA binding	
	CTCF	16	negative regulation of transcription from RNA polymerase II promoter	chromosome	RNA polymerase II core promoter proximal region sequence-specific DNA binding	
	GATA3	10	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	transcription regulatory region sequence-specific DNA binding	
	SH3PXD2A	10	superoxide metabolic process	podosome	protein binding	
	CDH1	16	homophilic cell adhesion via plasma membrane adhesion molecules	extracellular region	glycoprotein binding	Rap1 signaling pathway
	ERBB2	17	MAPK cascade	nucleus	RNA polymerase I core binding	ErbB signaling pathway
	FOXA1	14	negative regulation of transcription from RNA polymerase II promoter	nucleus	RNA polymerase II transcription factor activity	
	ITPR1	3	response to hypoxia	nuclear inner membrane	ion channel activity	Calcium signaling pathway

(*continued on next page*)

http://dx.doi.org/10.1016/j.heliyon.2017.e00277

Table 1. (*continued*)

Cancer	Symbol	Chr	GO:Biol. Process	GO:Cell. Component	GO:Mol. Function	Pathway
	KMT2C	7	transcription	nucleus	DNA binding	Lysine degradation
	MAP2K4	17	apoptotic process	nucleus	protein kinase activity	MAPK signaling pathway
	MAP3K1	5	MAPK cascade	cytoplasm	protein kinase activity	MAPK signaling pathway
	NR1H2	19	negative regulation of transcription from RNA polymerase II promoter	nucleus	RNA polymerase II core promoter proximal region sequence-specific DNA binding	Insulin resistance
	OR5P2	11	G-protein coupled receptor signaling pathway	plasma membrane	G-protein coupled receptor activity	Olfactory transduction
	PTEN	10, 9	regulation of cyclin-dependent protein serine/threonine kinase activity	extracellular region	magnesium ion binding	Inositol phosphate metabolism
	PIK3CA	3	angiogenesis	intracellular	protein serine/threonine kinase activity	Inositol phosphate metabolism
	PIK3R1	5	cellular glucose homeostasis	nucleus	transmembrane receptor protein tyrosine kinase adaptor activity	ErbB signaling pathway
	RUNX1	21	ossification	nucleus	regulatory region DNA binding	Pathways in cancer
	TPRX1	19	regulation of transcription	nucleus	DNA binding	
	TP53	17	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II regulatory region sequence-specific DNA binding	MAPK signaling pathway
Colorectal Adenocarcinoma	APC	5	mitotic cytokinesis	kinetochore	protein binding	Wnt signaling pathway
$n = 633$	BRAF	7	MAPK cascade	nucleus	protein kinase activity	MAPK signaling pathway
	FBXW7	4	protein polyubiquitination	nucleoplasm	ubiquitin-protein transferase activity	Ubiquitin mediated proteolysis
	KRAS	12	MAPK cascade	intracellular	GTPase activity	MAPK signaling pathway
	SMAD4	18	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II regulatory region sequence-specific DNA binding	FoxO signaling pathway

(*continued on next page*)

Table 1. (*continued*)

Cancer	Symbol	Chr	GO:Biol. Process	GO:Cell. Component	GO:Mol. Function	Pathway
	TBP	6	DNA-templated transcription	nuclear chromatin	RNA polymerase II core promoter proximal region sequence-specific DNA binding	Basal transcription factors
	AR	X	in utero embryonic development	nuclear chromatin	RNA polymerase II core promoter proximal region sequence-specific DNA binding	Oocyte meiosis
	ATXN1	6	transcription	nucleus	DNA binding	
	CACNA1B	9	transport	plasma membrane	voltage-gated calcium channel activity	MAPK signaling pathway
	CTNNB1	3	negative regulation of transcription from RNA polymerase II promoter	spindle pole	RNA polymerase II transcription factor binding	Rap1 signaling pathway
	GRIA2	4	signal transduction	endoplasmic reticulum membrane	ionotropic glutamate receptor activity	cAMP signaling pathway
	IRF5	7	transcription	nucleus	regulatory region DNA binding	Toll-like receptor signaling pathway
	KRT1	12	complement activation	extracellular space	receptor activity	
	LAMC3	9	cell morphogenesis involved in differentiation	extracellular region	structural molecule activity	PI3K-Akt signaling pathway
	NRAS	1	MAPK cascade	Golgi membrane	GTP binding	MAPK signaling pathway
	NEFH	22	microtubule cytoskeleton organization	cytoplasm	structural molecule activity	Amyotrophic lateral sclerosis (ALS)
	OPRD1	1	protein import into nucleus	cytoplasm	opioid receptor activity	cGMP-PKG signaling pathway
	PIK3CA	3	angiogenesis	intracellular	protein serine/threonine kinase activity	Inositol phosphate metabolism
	PPM1E	17	negative regulation of protein kinase activity	nucleus	protein serine/threonine phosphatase activity	
	TP53	17	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II regulatory region sequence-specific DNA binding	MAPK signaling pathway
Renal Clear Cell	ARAP3	5	cytoskeleton organization	ruffle	GTPase activator activity	Rap1 signaling pathway
$n = 538$	BAP1	3	regulation of cell growth	intracellular	chromatin binding	

(*continued on next page*)

http://dx.doi.org/10.1016/j.heliyon.2017.e00277

Table 1. (*continued*)

Cancer	Symbol	Chr	GO:Biol. Process	GO:Cell. Component	GO:Mol. Function	Pathway
	GPR32	19	complement receptor mediated signaling pathway	plasma membrane	complement receptor activity	
	SETD2	3	angiogenesis	nucleus	protein binding	Lysine degradation
	SRGAP3	3	signal transduction	cytoplasm	GTPase activator activity	Axon guidance
	SSX3	X	transcription	intracellular	nucleic acid binding	
	ACACA	17	tissue homeostasis	nucleolus	acetyl-CoA carboxylase activity	Fatty acid biosynthesis
	BTRC	10	G2/M transition of mitotic cell cycle	nucleus	ubiquitin-protein transferase activity	Oocyte meiosis
	CDC27	17	metaphase/anaphase transition of mitotic cell cycle	nucleus	protein binding	Cell cycle
	FAM104A	17			protein binding	
	FAM151A	1		membrane		
	HEBP1	12	circadian rhythm	extracellular region	heme binding	
	KLK1	19	proteolysis	nucleus	serine-type endopeptidase activity	Renin-angiotensin system
	OVGP1	1	carbohydrate metabolic process	extracellular region	chitinase activity	
	PTEN	10, 9	regulation of cyclin-dependent protein serine/threonine kinase activity	extracellular region	magnesium ion binding	Inositol phosphate metabolism
	PABPC1	12, 8	nuclear-transcribed mRNA catabolic process	nucleus	nucleotide binding	RNA transport
	PBRM1	3	chromatin remodeling	nuclear chromosome	DNA binding	
	TP53	17	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II regulatory region sequence-specific DNA binding	MAPK signaling pathway
	VHL	3	negative regulation of transcription from RNA polymerase II promoter	nucleus	ubiquitin-protein transferase activity	HIF-1 signaling pathway
	ZNF175	19	transcription	intracellular	nucleic acid binding	
Lung Adenocarcinoma	ATM	11	DNA damage checkpoint	chromosome	DNA binding	NF-kappa B signaling pathway
$n = 522$	BRAF	7	MAPK cascade	nucleus	protein kinase activity	MAPK signaling pathway

(*continued on next page*)

Table 1. (*continued*)

Cancer	Symbol	Chr	GO:Biol. Process	GO:Cell. Component	GO:Mol. Function	Pathway
	CREBBP	16	negative regulation of transcription from RNA polymerase II promoter	histone acetyltransferase complex	core promoter proximal region sequence-specific DNA binding	cAMP signaling pathway
	KRAS	12	MAPK cascade	intracellular	GTPase activity	MAPK signaling pathway
	LRP1B	2	receptor-mediated endocytosis	integral component of membrane	calcium ion binding	
	SOS1	2	MAPK cascade	intracellular	DNA binding	MAPK signaling pathway
	U2AF1	21	mRNA splicing	nucleoplasm	nucleotide binding	Spliceosome
	CHEK2	15, 22, Y	DNA damage checkpoint	chromosome	protein kinase activity	Cell cycle
	DMD	X	positive regulation of cell-matrix adhesion	nucleus	dystroglycan binding	Hypertrophic cardiomyopathy (HCM)
	EGFR	7	MAPK cascade	Golgi membrane	glycoprotein binding	MAPK signaling pathway
	FLG	1	multicellular organism development	nucleus	structural molecule activity	
	KEAP1	19	in utero embryonic development	nucleoplasm	ubiquitin-protein transferase activity	Ubiquitin mediated proteolysis
	NF1	17	MAPK cascade	nucleus	GTPase activator activity	MAPK signaling pathway
	PIK3CA	3	angiogenesis	intracellular	protein serine/threonine kinase activity	Inositol phosphate metabolism
	RYR2	1	response to hypoxia	cell	ryanodine-sensitive calcium-release channel activity	Calcium signaling pathway
	STK11	19	regulation of cell growth	nucleus	magnesium ion binding	FoxO signaling pathway
	SPTA1	1	MAPK cascade	cytosol	Ras guanyl-nucleotide exchange factor activity	
	TNN	1	cell-matrix adhesion	proteinaceous extracellular matrix	integrin binding	PI3K-Akt signaling pathway
	TP53	17	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II regulatory region sequence-specific DNA binding	MAPK signaling pathway
	USH2A	1	visual perception	photoreceptor inner segment	protein binding	

(*continued on next page*)

http://dx.doi.org/10.1016/j.heliyon.2017.e00277

Table 1. (*continued*)

Cancer	Symbol	Chr	GO:Biol. Process	GO:Cell. Component	GO:Mol. Function	Pathway
Ovarian Serous Carcinoma	KIT	4	MAPK cascade	acrosomal vesicle	protease binding	Ras signaling pathway
$n = 603$	ACACB	12	acetyl-CoA metabolic process	nucleus	acetyl-CoA carboxylase activity	Fatty acid biosynthesis
	ANK1	8	exocytosis	nucleus	structural molecule activity	Proteoglycans in cancer
	CDH11	16	skeletal system development	cytoplasm	calcium ion binding	
	COL4A4	2	extracellular matrix organization	extracellular region	extracellular matrix structural constituent	PI3K-Akt signaling pathway
	COL6A6	3	cell adhesion	extracellular region		PI3K-Akt signaling pathway
	CYP4A11	1	long-chain fatty acid metabolic process	cytoplasm	monooxygenase activity	Fatty acid degradation
	EGFR	7	MAPK cascade	Golgi membrane	glycoprotein binding	MAPK signaling pathway
	GRIN2B	12	MAPK cascade	intracellular	NMDA glutamate receptor activity	Ras signaling pathway
	GNPAT	1	glycerophospholipid metabolic process	mitochondrion	receptor binding	Glycerophospholipid metabolism
	IL21R	16	natural killer cell activation	integral component of membrane	interleukin-21 receptor activity	Cytokine–cytokine receptor interaction
	KAT6B	10	nucleosome assembly	nucleosome	DNA binding	
	MYH13	17	muscle contraction	muscle myosin complex	microfilament motor activity	Tight junction
	MYH2	17	plasma membrane repair	Golgi apparatus	microfilament motor activity	Tight junction
	NF2	22	mesoderm formation	ruffle	actin binding	Hippo signaling pathway
	PLCH1	3	lipid catabolic process	cytoplasm	phosphatidylinositol phospholipase C activity	Inositol phosphate metabolism
	KCNQ5	6	protein complex assembly	plasma membrane	inward rectifier potassium channel activity	Cholinergic synapse
	SNTG1	8	cell communication	nucleus	actin binding	
	TP53	17	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II regulatory region sequence-specific DNA binding	MAPK signaling pathway
	ZAN	7	binding of sperm to zona pellucida	plasma membrane		

(*continued on next page*)

Table 1. (*continued*)

Cancer	Symbol	Chr	GO:Biol. Process	GO:Cell. Component	GO:Mol. Function	Pathway
Prostate Adenocarcinoma	APC	5	mitotic cytokinesis	kinetochore	protein binding	Wnt signaling pathway
$n = 499$	BRAF	7	MAPK cascade	nucleus	protein kinase activity	MAPK signaling pathway
	POLI	18	DNA replication	intracellular	damaged DNA binding	Fanconi anemia pathway
	EP300	22	negative regulation of transcription from RNA polymerase II promoter	histone acetyltransferase complex	RNA polymerase II core promoter sequence-specific DNA binding	cAMP signaling pathway
	RGPD8	2	protein targeting to Golgi	intracellular	Ran GTPase binding	RNA transport
	CACNA1A	19	sulfur amino acid metabolic process	nucleus	ion channel activity	MAPK signaling pathway
	CDC27	17	metaphase/anaphase transition of mitotic cell cycle	nucleus	protein binding	Cell cycle
	CHEK2	15, 22, Y	DNA damage checkpoint	chromosome	protein kinase activity	Cell cycle
	FOXA1	14	negative regulation of transcription from RNA polymerase II promoter	nucleus	RNA polymerase II transcription factor activity	
	GRIK3	1	adenylate cyclase-inhibiting G-protein coupled glutamate receptor signaling pathway	plasma membrane	adenylate cyclase inhibiting G-protein coupled glutamate receptor activity	Neuroactive ligand-receptor interaction
	IDH1	2	glyoxylate cycle	cytoplasm	magnesium ion binding	Citrate cycle (TCA cycle)
	KRTAP4-11	17		keratin filament	protein binding	
	MSH3	5	meiotic mismatch repair	nuclear chromosome	damaged DNA binding	Mismatch repair
	PTEN	10, 9	regulation of cyclin-dependent protein serine/threonine kinase activity	extracellular region	magnesium ion binding	Inositol phosphate metabolism
	PIK3CA	3	angiogenesis	intracellular	protein serine/threonine kinase activity	Inositol phosphate metabolism
	SPOP	17	regulation of proteolysis	nucleus	protein binding	
	SYNE1	6	nucleus organization	nucleus	actin binding	
	TP53	17	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II regulatory region sequence-specific DNA binding	MAPK signaling pathway
	FRG1B	4				

(*continued on next page*)

Table 1. (*continued*)

Cancer	Symbol	Chr	GO:Biol. Process	GO:Cell. Component	GO:Mol. Function	Pathway
	KMT2C	7	transcription	nucleus	DNA binding	Lysine degradation
Stomach Adenocarcinoma	ARID1A	1	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	DNA binding	
$n = 478$	FAT4	4	branching involved in ureteric bud morphogenesis	intracellular	calcium ion binding	
	KRAS	12	MAPK cascade	intracellular	GTPase activity	MAPK signaling pathway
	LRP1B	2	receptor-mediated endocytosis	integral component of membrane	calcium ion binding	
	RGPD4	2	protein targeting to Golgi	intracellular		RNA transport
	SMAD4	18	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II regulatory region sequence-specific DNA binding	FoxO signaling pathway
	SMARCA4	19	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II core promoter proximal region sequence-specific DNA binding	
	CDH1	16	homophilic cell adhesion via plasma membrane adhesion molecules	extracellular region	glycoprotein binding	Rap1 signaling pathway
	CTNNB1	3	negative regulation of transcription from RNA polymerase II promoter	spindle pole	RNA polymerase II transcription factor binding	Rap1 signaling pathway
	CDC27	17	metaphase/anaphase transition of mitotic cell cycle	nucleus	protein binding	Cell cycle
	CHEK2	15, 22, Y	DNA damage checkpoint	chromosome	protein kinase activity	Cell cycle
	FLG	1	multicellular organism development	nucleus	structural molecule activity	
	MUC6	11	O-glycan processing	extracellular region	extracellular matrix structural constituent	
	OBSCN	1	protein phosphorylation	cytosol	protein kinase activity	
	PIK3CA	3	angiogenesis	intracellular	protein serine/threonine kinase activity	Inositol phosphate metabolism
	PCDHA3	5	cell adhesion	plasma membrane	calcium ion binding	

(*continued on next page*)

Table 1. (*continued*)

Cancer	Symbol	Chr	GO:Biol. Process	GO:Cell. Component	GO:Mol. Function	Pathway
	RHOA	3	transforming growth factor beta receptor signaling pathway	intracellular	GTPase activity	Ras signaling pathway
	SPTA1	1	MAPK cascade	cytosol	Ras guanyl-nucleotide exchange factor activity	
	TP53	17	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II regulatory region sequence-specific DNA binding	MAPK signaling pathway
	USH2A	1	visual perception	photoreceptor inner segment	protein binding	
Uterine Corpus Endometrial Carcinoma	AKT1	14	protein import into nucleus	nucleus	protein kinase activity	MAPK signaling pathway
n = 548	ARID1A	1	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	DNA binding	
	CTCF	16	negative regulation of transcription from RNA polymerase II promoter	chromosome	RNA polymerase II core promoter proximal region sequence-specific DNA binding	
	EP300	22	negative regulation of transcription from RNA polymerase II promoter	histone acetyltransferase complex	RNA polymerase II core promoter sequence-specific DNA binding	cAMP signaling pathway
	FBXW7	4	protein polyubiquitination	nucleoplasm	ubiquitin-protein transferase activity	Ubiquitin mediated proteolysis
	KRAS	12	MAPK cascade	intracellular	GTPase activity	MAPK signaling pathway
	TIAM1	21	cardiac muscle hypertrophy	nucleus	receptor signaling protein activity	Ras signaling pathway
	CTNNB1	3	negative regulation of transcription from RNA polymerase II promoter	spindle pole	RNA polymerase II transcription factor binding	Rap1 signaling pathway
	CHD4	12	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II core promoter proximal region sequence-specific DNA binding	Viral carcinogenesis

(*continued on next page*)

http://dx.doi.org/10.1016/j.heliyon.2017.e00277

Table 1. (*continued*)

Cancer	Symbol	Chr	GO:Biol. Process	GO:Cell. Component	GO:Mol. Function	Pathway
	ESR1	6	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II core promoter proximal region sequence-specific DNA binding	Estrogen signaling pathway
	FGFR2	10	negative regulation of transcription from RNA polymerase II promoter	extracellular region	protein tyrosine kinase activity	MAPK signaling pathway
	FLNA	X	platelet degranulation	extracellular region	G-protein coupled receptor binding	MAPK signaling pathway
	FOXA2	20	positive regulation of transcription from RNA polymerase II promoter by glucose	nucleus	RNA polymerase II core promoter proximal region sequence-specific DNA binding	Maturity onset diabetes of the young
	PTEN	10, 9	regulation of cyclin-dependent protein serine/threonine kinase activity	extracellular region	magnesium ion binding	Inositol phosphate metabolism
	PIK3CA	3	angiogenesis	intracellular	protein serine/threonine kinase activity	Inositol phosphate metabolism
	PIK3R1	5	cellular glucose homeostasis	nucleus	transmembrane receptor protein tyrosine kinase adaptor activity	ErbB signaling pathway
	PRPF8	17	spliceosomal tri-snRNP complex assembly	nucleus	second spliceosomal transesterification activity	Spliceosome
	PPP2R1A	19	G2/M transition of mitotic cell cycle	protein phosphatase type 2A complex	antigen binding	mRNA surveillance pathway
	SPOP	17	regulation of proteolysis	nucleus	protein binding	
	TP53	17	negative regulation of transcription from RNA polymerase II promoter	nuclear chromatin	RNA polymerase II regulatory region sequence-specific DNA binding	MAPK signaling pathway

(mutations, deletions, amplifications, downregulation, upregulation), each datafile for n tumor samples consisted of 100 binary $(0, 1)$ variables, where 20 columns represented the presence of somatic mutations (1:yes, 0:no), 20 columns represented presence of deletions (1:GISTIC $= -2$, 0:otherwise), 20 columns represented presence of amplifications (1:GISTIC $= 2$, 0:otherwise), 20 columns represented downregulation (1:Z-score ≤ -1.96, 0:Z-score > -1.96), and 20 columns represented presence of upregulation (1:Z-score ≥ 1.96, 0:Z-score < 1.96).

Permutation-based affinity matrix (PBAM)

The methods described in current and following sections are collectively called "PRogression Inference of Somatic Mutations in CAncer" (PRISM-CA). A permutation-based affinity matrix (PBAM) approach was employed for identifying the order of events among the top 20 driver genes used. The PBAM approach to infer sequential order of genomic alterations is non-parametric and independent of fitness, waiting times, and baseline somatic mutation rates. Define an event as a boolean outcome of true (or binary 1 vs. 0), for either a somatic mutation, deletion, amplification, downregulation, or upregulation of a driver gene. Thus, for 20 top driver genes and 5 possible outcomes per gene (mutation, deletion, amplification, downregulation, or upregulation), there are $p = 100$ potential events for each tumor sample based on the 20 driver genes. Let n be the total number of tumor samples $(i = 1, 2, \ldots, n)$ of a given cancer histological subtype. The assumed data file for each cancer employed will therefore have $p = 100$ columns (binary variables) for the 100 events and n rows representing tumor samples. For each jth event, calculate the frequency, f_j, of the event among the n tumor samples. For a pair of events j and k, the affinity between event j and k is defined as $c_{jk} = (f_{jk}/f_j)(f_{jk}/f_k)/d_{jk}^2$, where f_{jk} is the co-occurrence frequency of events j and k among the n tumor samples, and d_{jk} is the between-event distance for events j and k defined as

$$d_{jk} = \frac{1}{n_{jk}} \sum_{i \in n_{jk}} \left(1 + \sum_{l \neq j} I(f_{lk} > f_{jk} | i) \right), \tag{1}$$

where n_{jk} is the cardinality of tumors with co-occurring events j and k, and f_{lk} is the co-occurrence frequency for all other events along with event k (within the ith sample). The co-occurrence frequency f_{jk} is initially set to $g_{jk}/2$, where g_{jk} is the calculated frequency based on the data. Note that f_{jk} and g_{jk} matrices have dimensions $p \times p$, where p is the total number of events for a particular cancer type. The summation on the right side of (1) is performed over all pairs of events in each ith sample, while the summation on the left is over all tumor samples with events j and k. Since each ith tumor sample with events j and k can have up to m_i total events, the number of permutations of event labels, $\mathcal{G}_j$, for the ith tumor sample is $m_i!$ For

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

each permutation of event labels representing events in the ith tumor sample, we have

$$A_i^{perm} = \frac{1}{\sum_{l=1}^{m_i-1}\left(\frac{1}{\sum_{p=1}^{l} c_{\mathcal{G}(p),\mathcal{G}(l+1)}}\right)}, \tag{2}$$

where $\mathcal{G}(p)$ and $\mathcal{G}(l+1)$ are the subscripts for c_{jk} represented by events j and k, respectively, within a permutation. Once values of A_i^{perm} are determined for the ith tumor sample, calculate the permutation-specific probability for the ith tumor sample as

$$P_i^{perm} = \frac{A_i^{perm}}{\sum_{l=1}^{m_i!} A_i^l}. \tag{3}$$

A permutation, or specific order of events, was assumed to be significant if its observed probability, P_i^{perm}, was greater than the expected probability of $1/m_i!$ After looping through all of the $m_i!$ permutations per tumor sample with genes j and k mutated, we updated the co-occurrence frequency for events j and k using the relationship

$$f_{jk} = \frac{g_{jk}\sum_i^{n_{jk}}\sum_{perm}^{m_i!} P_i^{perm} I(j \to k)_{perm}}{\sum_i^{n_{jk}}\sum_{perm}^{m_i!} P_i^{perm} I(j \to k)_{perm} + \sum_i^{n_{jk}}\sum_{perm}^{m_i!} P_i^{perm} I(k \to j)_{perm}}, \tag{4}$$

where $I(j \to k)$ represents cases where j precedes k within a permutation, and $I(k \to j)$ represents cases where k precedes j within the same permutation. The above steps were repeated 10 times to iteratively update values of f_{jk}, which were used to update values of d_{jk} and c_{jk} at the beginning of each iteration. As we looped through all possible pairwise values of j and k during the last iteration, we compiled a list based on (only) the most significant value of P_i^{perm} for each tumor sample for each pairwise combination of j and k. From this list of the most significant permutations over all values of j and k, we determined the frequency that each event occupies the 1st, 2nd, 3rd, 4th, or 5th element in all of the output permutations. At convergence, $\mathbf{C}$ is not a symmetric matrix, since non-zero elements in the jth row provide the out-links of the jth node (from node j to others), while non-zero elements in the kth column provide its in-links (from others to node k). This leads to a dichotomy for each event, where the in-link can be compared with its out-link, essentially revealing whether event j is more of an information provider than an information receiver based on the relative values of c_{jk} and c_{kj}. We first filtered $\mathbf{C}$ by zeroing the smaller of each pair of c_{jk} and c_{kj} in the lower and upper triangulars. The next section describes an additional filtration method applied to $\mathbf{C}$. Figure 1 lists the computational steps involved in equations (1)–(4). During all runs for all histological subtypes considered, we monitored $||\mathbf{C}||$, which increased with decreasing increments during the iterations.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

Data: TCGA data
Result: Genes and mutation order
Select cancer histological subtype
Determine n total tumor samples with > 1 nextgen-based pathogenic mutations
Determine # genes with mutations among the n tumor samples
Define $\mathbf{X}(n, \#genes)$ as byte using binary values (0,1) for mutations (n/y)
Determine f_j, g_{jk}
Initialize $f_{jk} = g_{jk}/2$
for $iter \leftarrow 1$ **to** 10 **do**
  for $j \leftarrow 1$ **to** $\#genes$ **do**
    for $k \leftarrow 1$ **to** $\#genes$ **do**
      $n_{jk} = 0, sum2 = 0$
      for $i \leftarrow 1$ **to** n **do**
        if $x(i,j) = 1$ **and** $x(i,k) = 1$ **then**
          $n_{jk} += 1$
          $sum1 = 0$
          for $l \leftarrow 1$ **to** $\#genes$ **do**
            if $l \neq k$ **and** $f_{lk} > f_{jk}$ **then**
              $sum1 += 1$
            end
          end
          $sum1 += 1$
          $sum2 += sum1$
        end
      end
      $d_{jk} = sum2/n_{jk}$
      $C_{jk} = (f_{jk}/f_j)(f_{jk}/f_k)/d_{jk}^2$
    end
  end
  for $j \leftarrow 1$ **to** $\#genes$ **do**
    for $k \leftarrow 1$ **to** $\#genes$ **do**
      $sum1 = 0, sum2 = 0$
      for $i \leftarrow 1$ **to** n **do**
        if $x(i,j) = 1$ **and** $x(i,k) = 1$ **then**
          Determine number of mutations, m_i, in tumor sample i
          Generate $m_i!$ permutations, $\mathcal{G}$, of gene id labels for tumor sample i
          for $perm \leftarrow 1$ **to** $m_i!$ **do**
            $A_i^{perm} = 0$
            for $l \leftarrow 1$ **to** $m_i - 1$ **do**
              sum=0
              for $p \leftarrow 1$ **to** l **do**
                $sum += C_{\mathcal{G}(p),\mathcal{G}(l+1)}$
              end
              $A_i^{perm} += 1/sum$
            end
            $A_i^{perm} = 1/A_i^{perm}$
            for $l \leftarrow 1$ **to** m_i **do**
              if $\mathcal{G}(l) = j$ **then**
                $jpos = j$
              end
              if $\mathcal{G}(l) = k$ **then**
                $kpos = k$
              end
            end
            $jbeforek(perm) = 0$
            $kbeforej(perm) = 0$
            if $jpos < kpos$ **then**
              $jbeforek(perm) = 1$
            end
            if $kpos < jpos$ **then**
              $kbeforej(perm) = 1$
            end
          end
          $sumperm = 0$
          for $perm \leftarrow 1$ **to** $m_i!$ **do**
            $sumperm += A_i^{perm}$
          end
          for $perm \leftarrow 1$ **to** $m_i!$ **do**
            $P_i^{perm} = A_i^{perm}/sumperm$
          end
          if $iter = 10$ **then**
            $l =$ argmax$\{P_i^l > 1/m_i!\}$
            output $\mathcal{G}_l$
          end
          for $perm \leftarrow 1$ **to** $m_i!$ **do**
            if $jbeforek(perm) = 1$ **then**
              $sum1 += P_i^{perm}$
            end
          end
          for $perm \leftarrow 1$ **to** $m_i!$ **do**
            if $kbeforej(perm) = 1$ **then**
              $sum2 += P_i^{perm}$
            end
          end
        end
      end
      $f_{jk} = g_{jk} sum1/(sum1 + sum2)$
    end
  end
  Monitor increasing values of $||\mathbf{C}||$.
end

Figure 1. Computational algorithm for Permutation-based Affinity Matrix (PBAM) method.

Matrix filtration with Suppes probabilistic causation

Given a pair of events j and k, we can causally infer whether event j is likely to be causal of event k, or vice versa. Since the event data are binary(boolean), we first determined the probabilities $P(j) = f_j/n$, $P(k) = f_k/n$, $P(k|j) = f_{jk}/f_j$, and $P(k|\bar{j}) = ((f_k - f_{jk})/n)/(1 - P(j))$. Suppes definition of probabilistic causation [28] states that j is a prima facie cause of k, $j \rhd k$, if there is *temporality* such that event j occurs more frequently than event k, i.e., $P(j) > P(k)$, and k is a *probability raiser* of j, i.e., $P(k|j) > P(k|\bar{j})$ [29]. Taken collectively, the temporality condition and probability raising condition are amenable for describing selective advantage characteristics of the accumulation of genomic alterations during tumor progression. Temporality assumes that earlier occurring genomic events occur more frequently, and probability raising assumes that the probability of observing j raises the probability of observing k. These conditions were used for filtering $\mathbf{C}$, for which elements c_{jk} of $\mathbf{C}$ were set to zero if $P(j) \leq P(k)$ or $P(k|j) \leq P(k|\bar{j})$.

Polytree generation with weighted directed acyclic graphs

Once the affinity matrix $\mathbf{C}$ was filtered for Suppes probabilistic causation, we retained only the greatest column values of $\mathbf{C}$, and zeroed out the remaining column entries. This was done to prevent child nodes from having multiple parent nodes. Matrix $\mathbf{C}$ was then employed as an adjacency matrix $\mathbf{A}$ to generate a weighted directed acyclic graph (WDAG), $G = (V, E)$, with vertices (nodes) V_j and edges E_{jk} between vertices j and k. This essentially resulted in $\mathbf{A} \leftarrow \mathbf{C}$. A depth first search of G was then performed in order to uncover the structure of all possible depth-first trees with separate roots. Plots of tree forests were generated with edges which were colored according to their value of c_{jk}. We also outlined graph vertices (nodes) with colors representing the number of Pubmed publications found whose titles included the common gene symbol and "mutation," "deletion," "amplification," "downregulated" or "downregulation," or "upregulated" or "upregulation" for the gene's relevant event.

Mutual information Bayesian network (MIBN)

A Bayesian network approach was also developed using the Chow–Liu algorithm [30], for which between-event mutual information was determined in the form

$$MI(x, y) = \frac{1}{n} \sum_x \sum_y n_{xy} \log_e \left(\frac{n_{xy}}{n_x n_y} \right), \tag{5}$$

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

where $x, y \in \{0, 1\}$, n_{xy} is the number of tumor samples with co-occurrence of events x and y, and n_x and n_y are the number of tumor samples having singleton events. Prim's algorithm [31] was then applied to the WDAG to remove unconnected edges in order to construct the forest of trees. It warrants noting that the **MI** event-by-event matrix in MIBN is symmetric, while PBAM's **C** matrix is not symmetric. The adjacency matrix for tree construction was based on connected edge weights, which were also filtered using Suppes definition of probabilistic causation.

Randomization tests for temporality and probability raising

We developed a randomization test to determine the significance of temporality and probability raising between all possible pairs of events j and k represented by elements of **C** and **MI** which met the temporality and probability raising conditions during previous filtration. The p-value for temporality was based on the number of times the difference $P(j)^{(b)} - P(k)^{(b)}$ $(b = 1, 2, \ldots, 500)$ exceeded $P(j) - P(k)$, divided by 500. The p-value for the test of significance for probability raising was based on the number of times the difference $P(k|j)^{(b)} - P(k|\bar{j})^{(b)}$ exceeded $P(k|j) - P(k|\bar{j})$ divided by 500. During each bth iteration, binary outcomes $(0, 1)$ were randomly shuffled between events (binary variables) j and k. Graph edges for pairs of nodes with significant p-values ($P < 0.05$) were dashed using various line styles to represent full prima facie causation (—) where both temporality and probability raising conditions were met, only the temporality condition was met (- - -), only the probability raising condition existed (-.-.), or neither (....).

3. Results

Figures 2 and 3 illustrate the distribution of between-event PBAM values of c_{jk} and MIBN values of $MI(x, y)$ at convergence, which were employed for weighted directed acyclic graph generation. Figure 2 reveals that the distribution of c_{jk} values for the cancers investigated is right-skewed with the median typically falling near-zero. Renal clear cell carcinoma resulted in the greatest interquartile range (IQR) of c_{jk} values, and ovarian serous resulted in the smallest IQR of c_{jk} values. Driver genes having the greatest degree of "communication", that is, greatest extreme values of c_{jk} in the right whiskers of the boxes, will be dominated by only a few drivers such as TP53, PICK3CA, etc. Whereas the bulk of the remainder genes will have their c_{jk} values within the IQRs. This would imply that renal clear cell cancer has a greater proportion of drivers sharing information (greatest IQR), while ovarian serous cancers with a more narrow IQR of c_{jk} will share information to a lesser extent. Alternatively, Figure 3 reveals a left-skewed distribution of $MI(x, y)$ values for the cancers considered, showing much more variation in median values.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

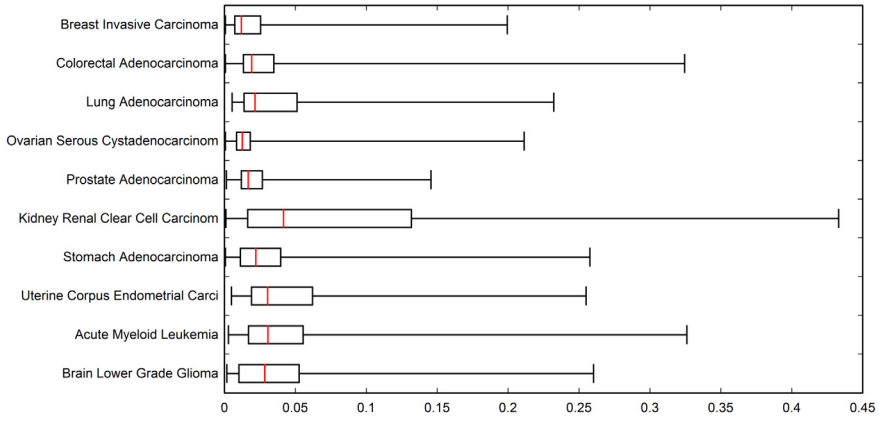

Figure 2. Box plot showing distribution of between-event PBAM c_{jk} matrix element values used for weighted directed acyclic graph generation.

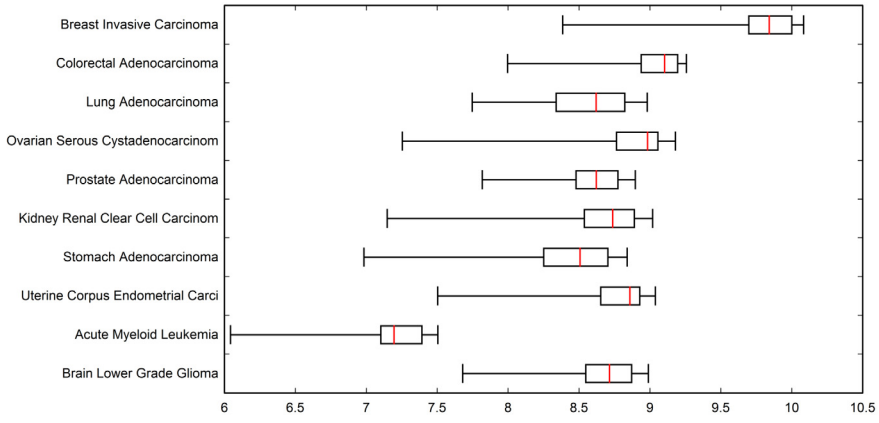

Figure 3. Box plot showing distribution of between-event MIBN $MI(x, y)$ mutual information values used for weighted directed acyclic graph generation.

Acute myeloid leukemia exhibits the lowest median value of roughly 7.2, and breast carcinoma exhibits the highest median of 9.8. Overall, there was not as much variation in range of the $MI(x, y)$ values across the cancers when compared with c_{jk} values. When compared with cancer-specific IQRs for c_{jk}, the IQR values of $MI(x, y)$ are approximately the same; however, the median values of $MI(x, y)$ are drastically different, especially between acute myeloid leukemia and breast invasive cancer. Thus, the between-cancer differences in information sharing represented by c_{jk} tend to be revealed by the extreme values, tail length, and width of IQR, whereas between-cancer differences in $MI(x, y)$ tend to depend mostly on median values.

The clinical impact of driver genes can best be portrayed by their impact on overall survival (OS) of patients. Table 2 lists the significant hazard ratios (HR) for OS, adjusted for age at diagnosis, and the 95% confidence interval. The majority of gene alterations were deleterious rather than protective, and we observed the

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

Table 2. Clinical relevance of genomic events evaluated in the context of survival analysis. Hazard ratios (HR) based on Cox proportional hazards regression with adjustment for age at diagnosis. HR compares the mortality rate among subjects with a given event with the average mortality rate among all subjects with the given cancer. Genes whose HR > 1 are termed "deleterious," while genes whose HR < 1 are called "protective." Survival time variable: overall survival in months (TCGA field: OS_MONTH), censoring variable: overall survival status (TCGA field: OS_STATUS). Gene symbol subscripts are: _m – mutation, _d – deletion, _a – amplification, _s – suppressed expression, _e – enhanced expression. Results sorted by p-values.

Cancer site	Event	Coef. (β_j)	s.e. (β_j)	Z	Prob.	HR(95%CI)
Acute Myeloid Leukemia	TP53_m	0.9981	0.3093	3.2268	0.0012	2.71(1.480, 4.975)
	RUNX1_a	1.8958	0.6300	3.0089	0.0026	6.66(1.937, 22.893)
	DNMT3A_m	0.4935	0.2173	2.2708	0.0231	1.64(1.070, 2.508)
	U2AF1_a	0.9604	0.4729	2.0307	0.0422	2.61(1.034, 6.602)
	NPM1_m	0.4309	0.2137	2.0167	0.0437	1.54(1.012, 2.339)
Brain Lower Grade Glioma	EGFR_a	1.4690	0.2461	5.9676	0.0000	4.34(2.682, 7.039)
	PTEN_s	1.3346	0.2641	5.0522	0.0000	3.80(2.263, 6.375)
	EGFR_m	1.7053	0.3470	4.9133	0.0000	5.50(2.787, 10.866)
	NF1_s	1.8884	0.4042	4.6710	0.0000	6.61(2.992, 14.598)
	NF1_m	1.6238	0.3512	4.6232	0.0000	5.07(2.548, 10.097)
	EGFR_e	0.8338	0.2034	4.0982	0.0000	2.30(1.545, 3.430)
	IDH1_e	1.1112	0.2828	3.9286	0.0000	3.04(1.745, 5.289)
	PTEN_m	1.5066	0.3858	3.9042	0.0000	4.51(2.118, 9.611)
	CHEK2_e	0.9563	0.2592	3.6889	0.0002	2.60(1.566, 4.326)
	PLCG1_e	0.9060	0.2575	3.5185	0.0004	2.47(1.494, 4.099)
	IDH1_m	−0.50696	0.1881	−2.69438	0.0070	0.60(0.417, 0.871)
	CACNA1S_m	1.7076	0.7172	2.3809	0.0172	5.52(1.352, 22.498)
	PTEN_d	1.2200	0.5213	2.3401	0.0192	3.39(1.219, 9.411)
	CIC_m	−0.78141	0.3463	−2.25623	0.0240	0.46(0.232, 0.902)
	CDC27_e	0.7492	0.3464	2.1625	0.0305	2.12(1.073, 4.172)
	TP53_e	0.6866	0.3306	2.0763	0.0378	1.99(1.039, 3.799)
	CIC_s	−0.49682	0.2500	−1.98657	0.0469	0.61(0.373, 0.993)
Breast Invasive Carcinoma	FOXA1_m	1.2905	0.3919	3.2931	0.0009	3.63(1.686, 7.836)
	CTCF_e	1.2368	0.3910	3.1627	0.0015	3.44(1.601, 7.413)
	MAP2K4_d	1.1677	0.3928	2.9726	0.0029	3.21(1.489, 6.943)
	ITPR1_a	1.0064	0.3457	2.9104	0.0036	2.74(1.389, 5.388)
	NR1H2_a	1.0615	0.4240	2.5030	0.0123	2.89(1.259, 6.638)
	ERBB2_m	1.1357	0.4571	2.4843	0.0129	3.11(1.271, 7.628)
	ARID1A_d	1.5909	0.7159	2.2221	0.0262	4.91(1.206, 19.969)
	PIK3R1_m	1.1007	0.5085	2.1648	0.0304	3.01(1.110, 8.145)
	KMT2C_d	1.2208	0.5870	2.0796	0.0375	3.39(1.073, 10.713)
	KMT2C_s	2.0576	1.0102	2.0367	0.0416	7.83(1.081, 56.699)
	PIK3CA_e	0.5588	0.2826	1.9774	0.0479	1.75(1.005, 3.043)
	CDH1_m	−0.59829	0.3037	−1.96998	0.0488	0.55(0.303, 0.997)
Colorectal Adenocarcinoma	GRIA2_m	1.4665	0.5148	2.8485	0.0043	4.33(1.580, 11.889)
	FBXW7_e	1.0294	0.4200	2.4505	0.0142	2.80(1.229, 6.378)
	KRT1_e	1.4667	0.7181	2.0423	0.0411	4.34(1.061, 17.714)
	FBXW7_s	1.4227	0.7250	1.9623	0.0497	4.15(1.002, 17.180)
Lung Squamous Cell Carcinoma	PIK3CA_e	−0.35430	0.1420	−2.49367	0.0126	0.70(0.531, 0.927)
	APC_m	0.9155	0.4191	2.1843	0.0289	2.50(1.099, 5.681)
	CTNNA2_a	1.0102	0.5091	1.9842	0.0472	2.75(1.012, 7.450)

(*continued on next page*)

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

Table 2. (*continued*)

Cancer site	Event	Coef. (β_j)	s.e. (β_j)	Z	Prob.	HR(95%CI)
Ovarian Serous Cystadenocarcinoma	PLCH1_m	2.1810	0.5868	3.7163	0.0002	8.86(2.803, 27.976)
	GRIN2B_a	0.4998	0.1932	2.5865	0.0096	1.65(1.129, 2.407)
Prostate Adenocarcinoma	IDH1_e	1.9711	0.7292	2.7031	0.0068	7.18(1.719, 29.977)
	POLI_e	1.8943	0.8191	2.3124	0.0207	6.65(1.335, 33.113)
Renal Clear Cell Carcinoma	ACACA_e	1.1876	0.2354	5.0436	0.0000	3.28(2.067, 5.203)
	PABPC1_e	1.1618	0.2350	4.9438	0.0000	3.20(2.016, 5.065)
	TP53_e	0.9174	0.2377	3.8594	0.0001	2.50(1.571, 3.988)
	SSX3_e	1.0375	0.2807	3.6964	0.0002	2.82(1.628, 4.893)
	CDC27_e	1.0453	0.3446	3.0328	0.0024	2.84(1.447, 5.590)
	FAM104A_e	0.6320	0.2553	2.4754	0.0133	1.88(1.141, 3.103)
	BAP1_m	0.4940	0.2430	2.0326	0.0420	1.64(1.018, 2.639)
	ACACA_a	2.0403	1.0130	2.0141	0.0440	7.69(1.056, 56.026)
Stomach Adenocarcinoma	USH2A_e	1.1331	0.2741	4.1341	0.0000	3.11(1.815, 5.314)
	OBSCN_a	1.6007	0.3905	4.0990	0.0000	4.96(2.306, 10.657)
	USH2A_a	1.7512	0.5906	2.9651	0.0030	5.76(1.811, 18.337)
	FAT4_m	−0.64513	0.2350	−2.74459	0.0060	0.52(0.331, 0.832)
	CTNNB1_m	−1.19370	0.5067	−2.35545	0.0185	0.30(0.112, 0.818)
	SMAD_d	0.6054	0.2638	2.2946	0.0217	1.83(1.092, 3.073)
	TP53_m	−0.34194	0.1622	−2.10777	0.0350	0.71(0.517, 0.976)
	SMARCA4_m	−1.03735	0.5085	−2.03990	0.0413	0.35(0.131, 0.960)
Uterine Corpus Endometrial Carcinoma	PIK3CA_m	−1.06895	0.3110	−3.43620	0.0005	0.34(0.187, 0.632)
	PTEN_m	−0.92347	0.2926	−3.15592	0.0016	0.40(0.224, 0.705)
	AKT1_a	1.4284	0.4629	3.0856	0.0020	4.17(1.684, 10.339)
	ARID1A_m	−1.78160	0.5895	−3.02214	0.0025	0.17(0.053, 0.535)
	ESR1_d	1.8821	0.7272	2.5879	0.0096	6.57(1.579, 27.317)
	PTEN_s	1.4300	0.5980	2.3913	0.0167	4.18(1.294, 13.494)
	PTEN_d	0.9293	0.3965	2.3435	0.0191	2.53(1.164, 5.510)
	PPP2R1A_a	1.3301	0.5894	2.2566	0.0240	3.78(1.191, 12.006)
	AKT1_e	0.8895	0.4253	2.0911	0.0365	2.43(1.057, 5.603)
	ESR1_a	0.8182	0.3986	2.0528	0.0400	2.27(1.038, 4.951)
	FBXW7_e	1.2182	0.5935	2.0525	0.0401	3.38(1.056, 10.821)
	KRAS_m	−1.21089	0.5911	−2.04842	0.0405	0.30(0.094, 0.949)

greatest HR of 8.86 (95% CI, 2.80–27.98) from the mutation of PLCH1 in ovarian serous cancer. This is followed by the suppressed expression of KMT2C in breast adenocarcinoma, resulting in an HR of 7.83 (95% CI, 1.08–56.67), amplification of ACACA in renal clear cell carcinoma with an HR of 7.69 (95% CI, 1.06–56.03), and enhanced expression of IDH1 in prostate adenocarcinoma, resulting in an HR of 7.18(1.72, 29.98). We also observed a protective effect from several genes, including the mutations of CTNNB1 and SMARCA4 in stomach adenocarcinoma, which resulted in HRs of 0.30 (95% CI, 0.11–0.82) and 0.35 (95% CI, 0.13–0.96), and mutations of ARID1A, KRAS, and PTEN in uterine cancer with an HR of 0.17 (95% CI, 0.05–0.54), 0.30 (95% CI, 0.09–0.95), and 0.40 (95% CI, 0.22–0.71).

http://dx.doi.org/10.1016/j.heliyon.2017.e00277

Interpreting Figures 4–23

The following sections describe results observed for the 10 cancers considered, and first cover PBAM results followed by MIBN results for each cancer. Each figure illustrates the degree of information sharing between events by use of gradient color scales for (a) values of between-event c_{jk} for PBAM plots or $MI(x, y)$ for MIBN plots, and (b) the amount of Pubmed hits for each node (event) based on reports containing the gene symbol and either "somatic mutation", "deletion", "amplification", "downregulation", or "upregulation", depending on the event being searched for. The line colors used between nodes (events) are based on the value of the between-event c_{jk} or $MI(x, y)$, which also refers to the color gradient scale used. There are five node colors representing the various gene-related events: mutation, deletion, amplification, downregulation, or upregulation in expression. Since there are 20 driver genes used per cancer, there are 100 potential nodes that can appear in each plot. Genes whose events never co-occur with other events are not shown in the plots, and therefore, co-occurrence of events is required in order for an event to appear in a plot. The number of samples harboring the event along with the percentage out of the total sample size (in parentheses) for each cancer is also listed above the relevant node. Varying line styles are also used to reflect the existence of between-event prima facie causality, probability raising, or neither.

3.1. Breast invasive carcinoma

PBAM-based results for invasive breast carcinoma shown in Figure 4 illustrate that mutations in TP53, PIK3CA, and GATA3, and upregulation of ERBB2 were the main driver events observed in the data [33, 37, 38, 47, 48, 51, 52]. The events which followed TP53 mutations were comprised of a mixture of deletions, amplifications, downregulation and upregulation, with few mutational events, which include PTEN deletions [40, 49], PIK3CA amplifications [38, 50], downregulation of ERBB2 [51, 52], PTEN [44, 51], AKT1 [55, 56], TP53 [53, 54], ARID1A [57, 58], and CTCF [59, 60] and upregulation of TP53 [53, 54], AKT1 [55, 56], PIK3CA [61, 62], and ARID1A [57, 58]. Secondary events following mutations in PIK3CA included upregulation of PTEN [41, 44], RUNX1 [45, 46], and mutations in PTEN [40, 41], ERBB2 [42, 43] and CDH1 [39]. Events downstream of ERBB2 upregulation included amplification of ERBB2 [42, 63] and upregulation of FOXA1 [66, 67] and CDH1 [64, 65]. Events that were likely to be causal after GATA3 mutations included AKT1 mutations [34] and upregulation of GATA3 [35, 36] and CTCF [59, 60].

The MIBN-derived WDAG shown in Figure 5 illustrates that in invasive breast carcinoma, causal inference indicates three main root nodes involving mutations in TP53, PIK3CA, and GATA3. Interestingly, while PIK3CA and GATA3 mutations

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

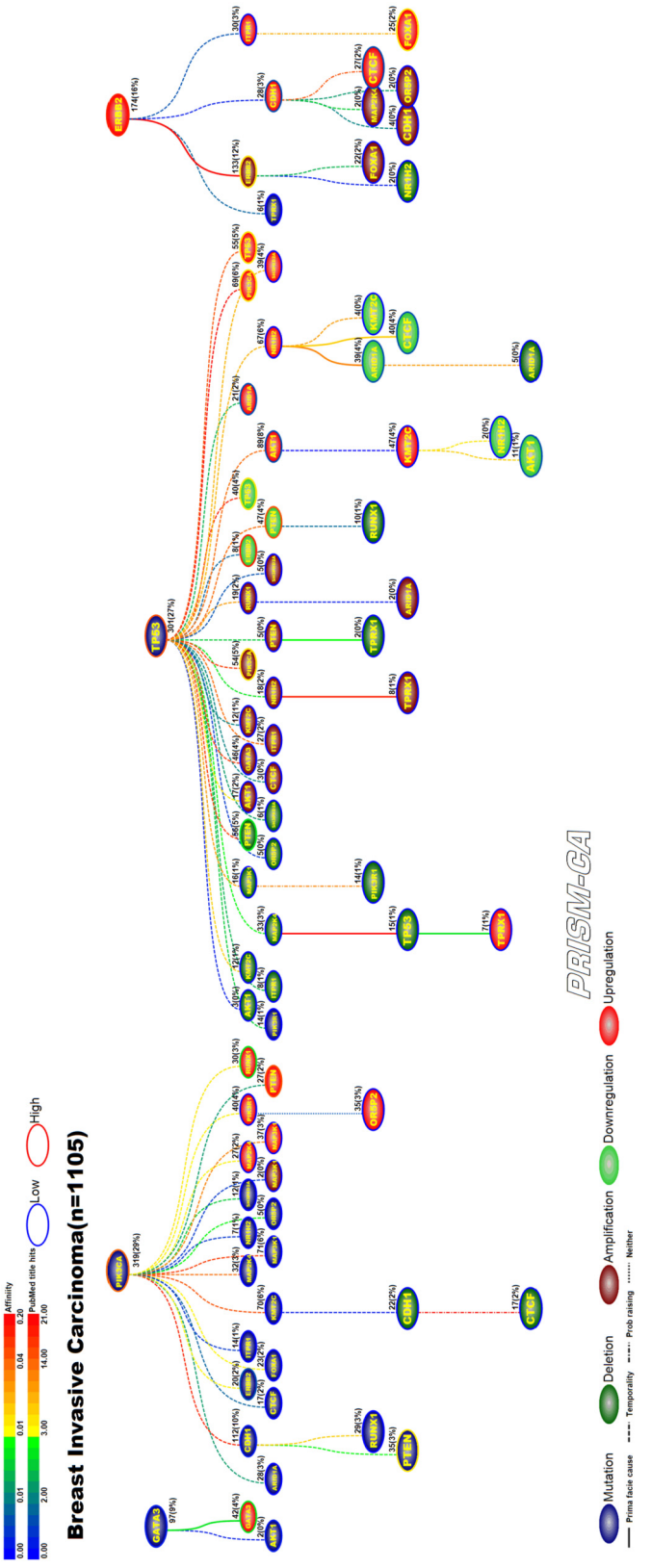

Figure 4. PBAM analysis results for Invasive Breast Carcinoma.

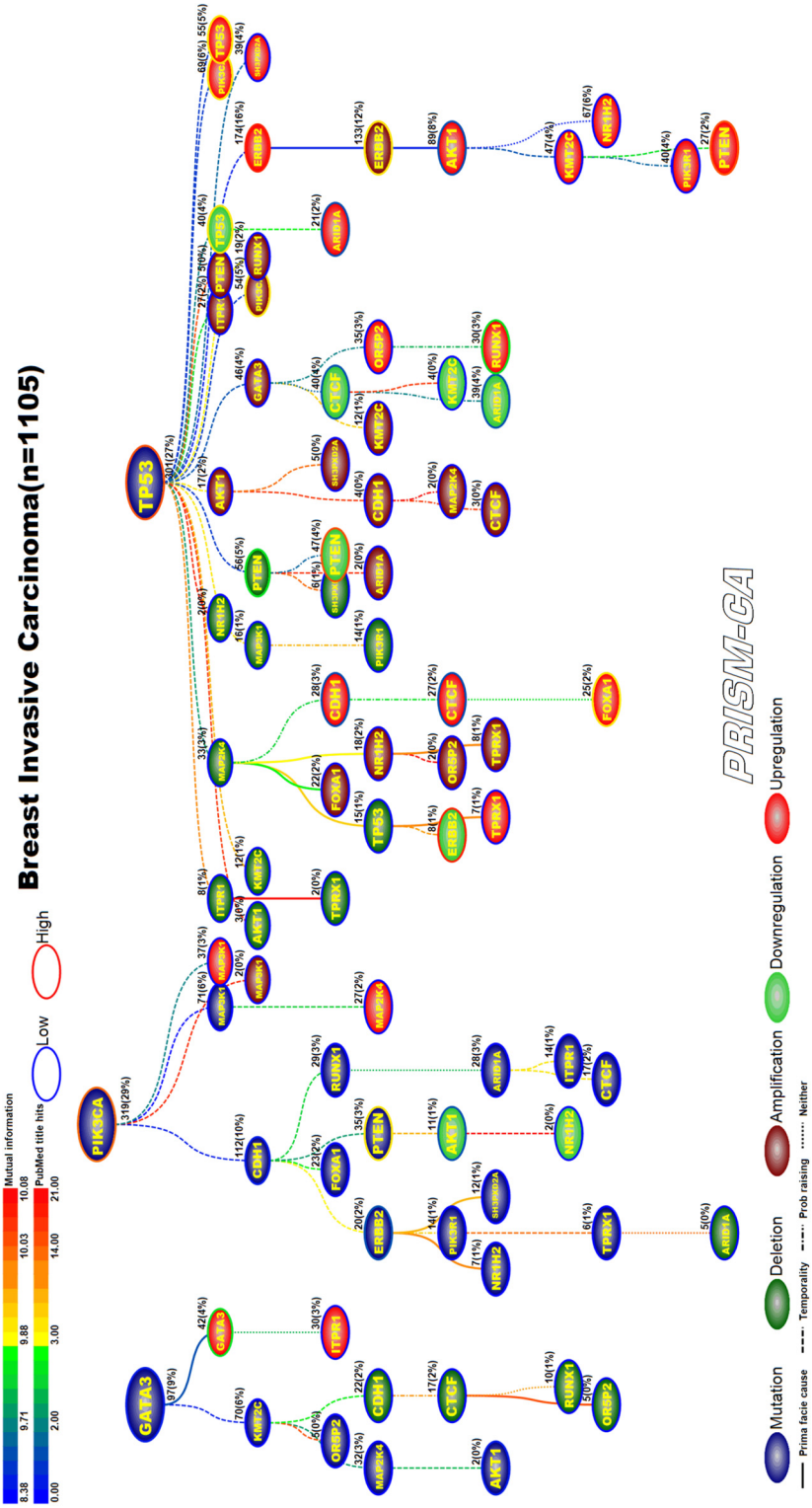

Figure 5. MIBN analysis results for Invasive Breast Carcinoma.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

seem to precede mostly mutations, TP53 seemed to precede deletions, amplifications, and transcriptional alterations. The tree starting with mutations in GATA3 contained one node involving expression changes in GATA3, which have been reported and another involving mutations in AKT1. Within the PIK3CA tree, there were nodes representing mutations in CDH1, PTEN, and ERBB2. The tree rooted by TP53 contained the remaining alterations involving deletions in PTEN, amplifications in PIK3CA and ERBB2, and gene expression changes in AKT1, ERBB2, CDH1, CTCF, FOXA1, PTEN, ARID1A, RUNX1, TP53, and PIK3CA.

3.2. Colorectal adenocarcinoma

PBAM results for colorectal adenocarcinoma (Figure 6) indicate that mutations in APC [68, 69] and downregulation of SMAD4 [81, 82] were the main driver events. Although the literature commonly reports mutations in BRAF, PIK3CA, and KRAS as key driver events in colorectal cancer, our results show that mutations in APC are the driver for events in these three genes. Driver events which were secondary to APC mutations included mutations in BRAF [70, 71], KRAS [72, 73], NRAS [74, 75], PIK3CA [76, 77], SMAD4 [78, 79], and TP53 [75, 80]. Whereas the secondary driver events observed to follow SMAD4 downregulation were SMAD4 deletions [83, 84], KRAS amplifications [95, 96], downregulation of TP53 [85, 86], APC [89, 90], KRAS [72, 91], PIK3CA [93, 94], FBXW7 [97, 98], and upregulation of KRAS [72, 91], BRAF [87, 88], CTNNB1 [92], APC [89, 90], PIK3CA [93, 94], FBXW7 [97, 98], SMAD4 [81, 82], and TP53 [85, 86]. The main driver events for MIBN-based analysis (Figure 7) were also mutations in APC and downregulation of SMAD4. Furthermore, we observed that KRAS-induced mutations in SMAD4, upregulation and downregulation in APC, downregulation in PIK3CA, and downregulation in KRAS. A variety of events in the APC tree also include mutations in NRAS and TP53, and upregulation of FBXW7. An apparent pattern in Figure 7 is that the tree rooted by APC is mostly populated with daughter nodes representing deletions, amplifications, and transcriptional changes, while the primary daughter events in the SMAD4 tree are transcriptionally related. Transcriptional changes visible within the SMAD4 tree are downregulation of CTNNB1, FBXW7, TP53 and upregulation of BRAF, CTNNB1, TP53, KRAS, SMAD4, and PIK3CA.

3.3. Lung adenocarcinoma

PBAM results for lung adenocarcinoma in Figure 8 indicate that mutations in TP53 [99, 100] were the single main driver event, followed by three separate clusters of events. One cluster (left side) included a series of child nodes events representing

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

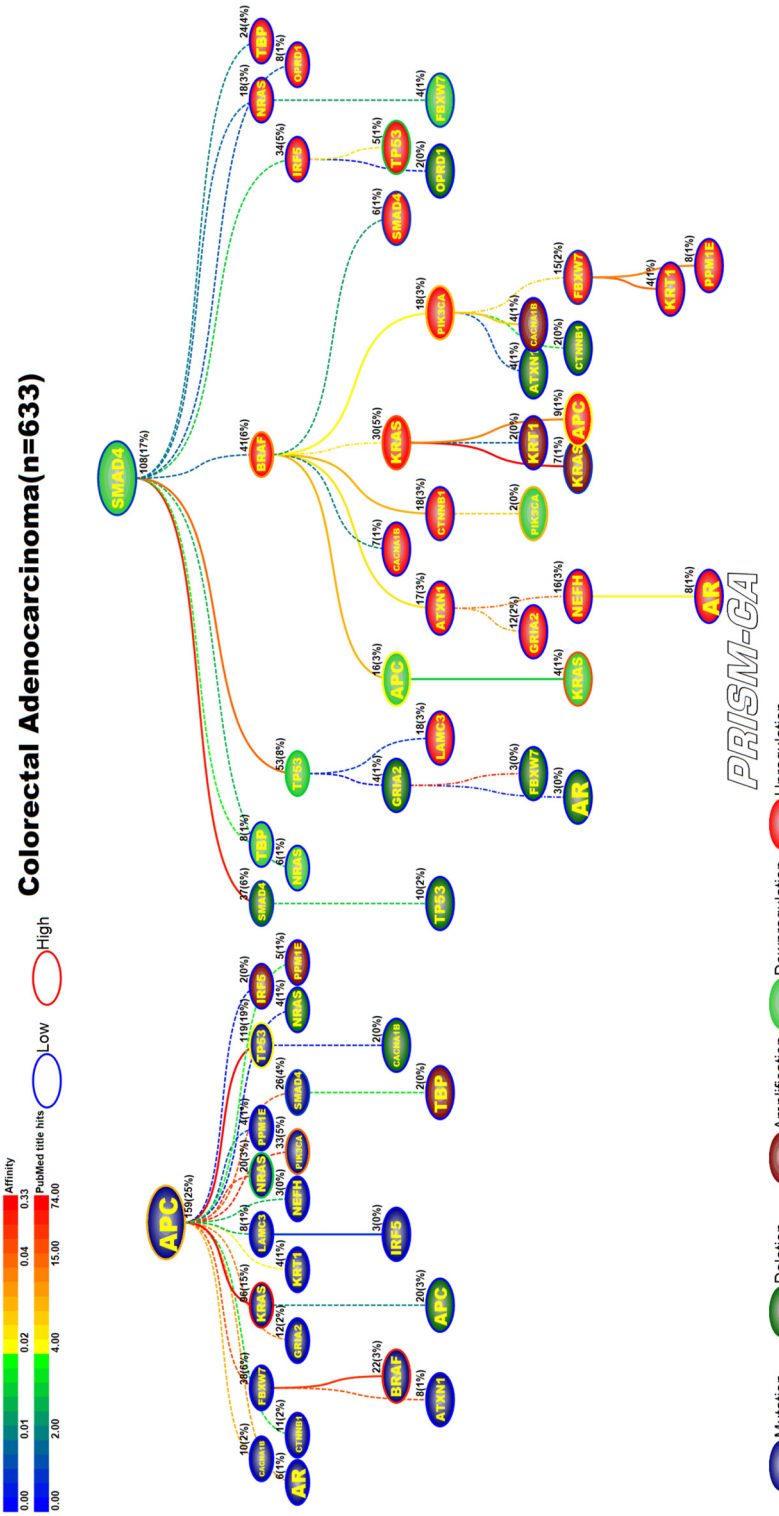

Figure 6. PBAM analysis results for Colorectal Adenocarcinoma.

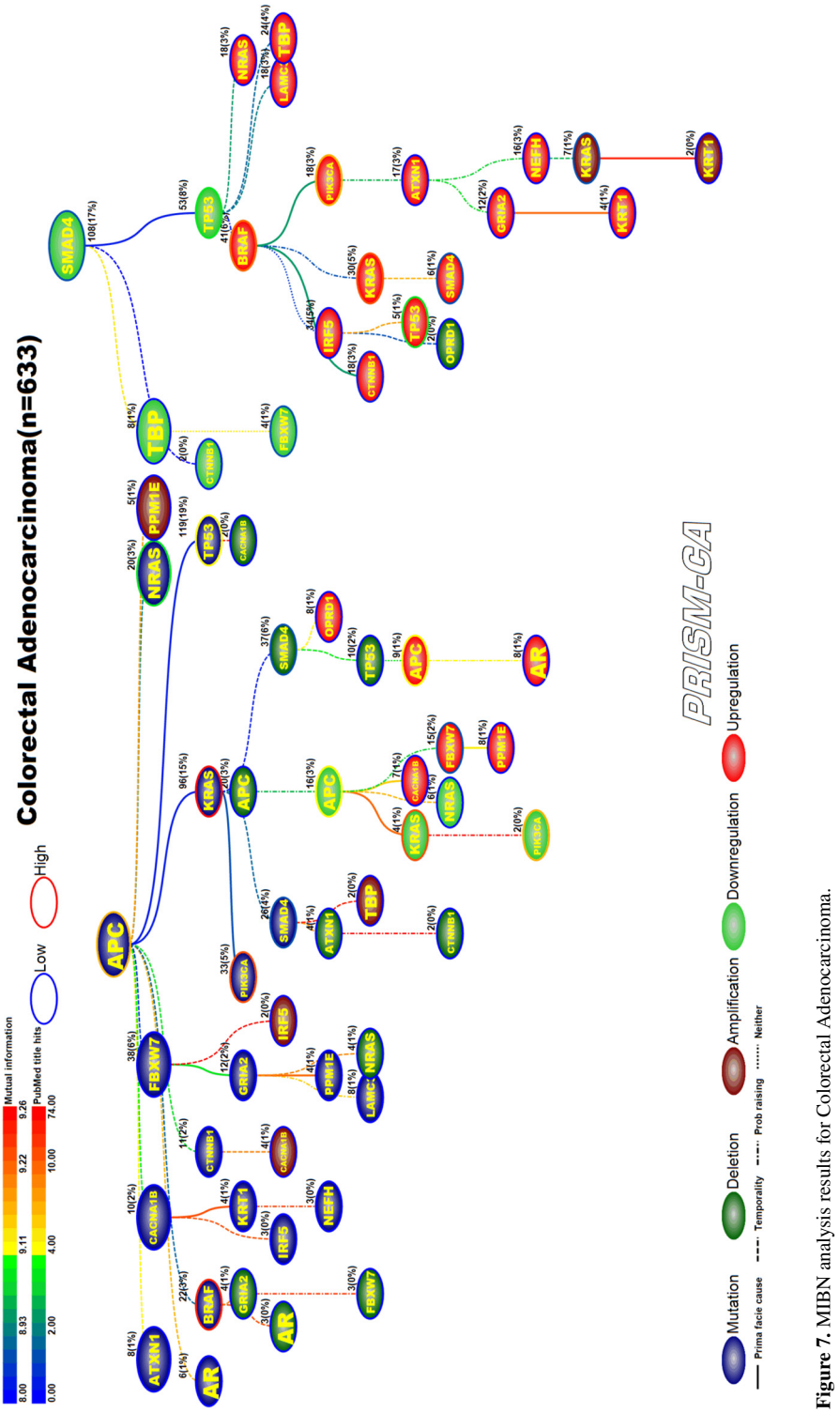

Figure 7. MIBN analysis results for Colorectal Adenocarcinoma.

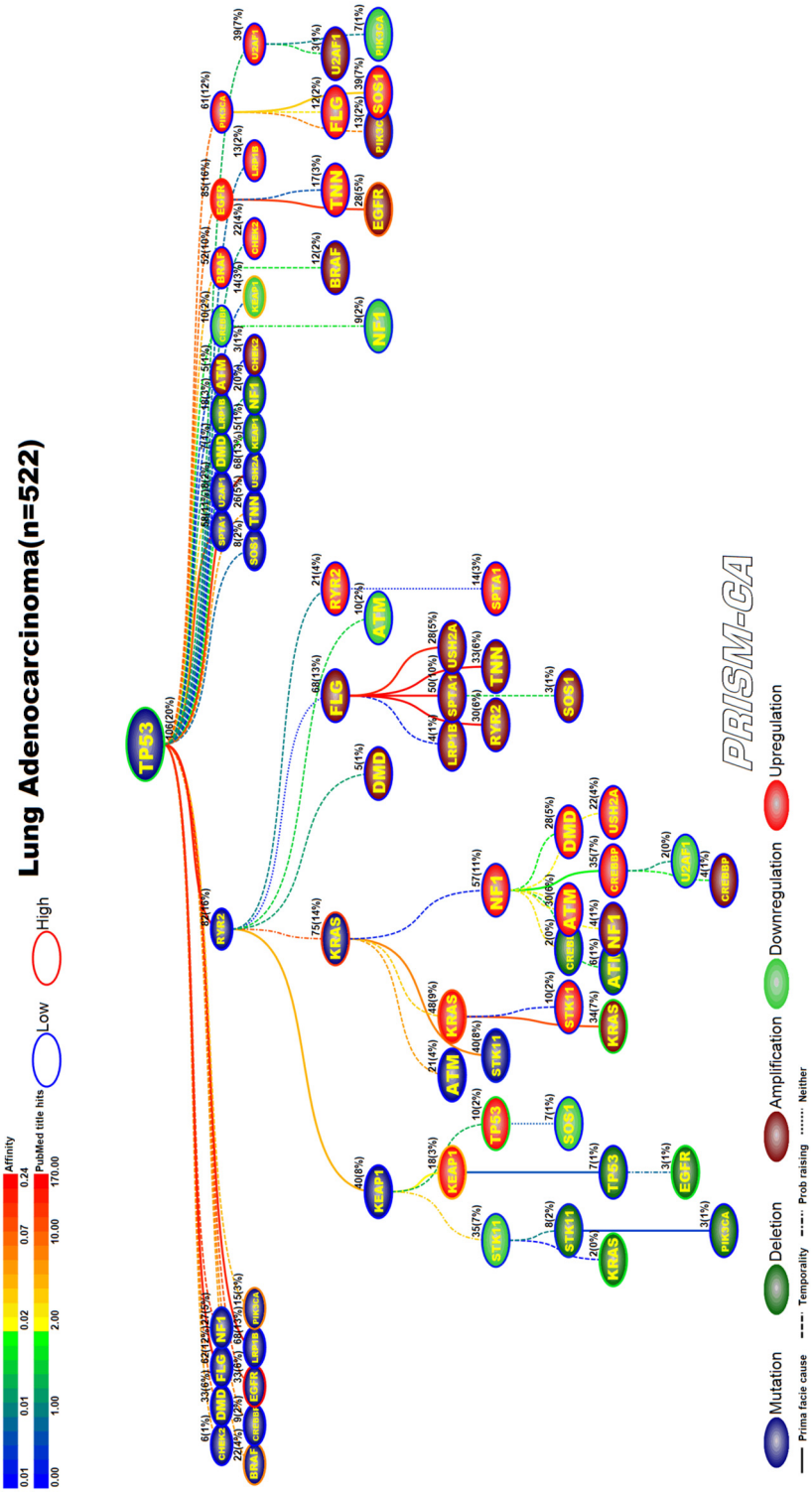

Figure 8. PBAM analysis results for Lung Adenocarcinoma.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

mostly mutations in BRAF [101, 102], EGFR [103, 104], NF1 [105], and PIK3CA [106, 107]. A second cluster was comprised of a polytree rooted by mutations in RYR2, followed by mutations in KRAS [116, 117], STK11 [100], deletions in KRAS [108, 109], STK11 [110], EGFR [113, 114], amplifications in KRAS [119, 120], downregulation of STK11 [100], and upregulation of KEAP1 [111, 112], TP53 [100, 115], KRAS [100, 118], STK11 [100], NF1 [121]. The third cluster of events was merely an agglomeration of child events consisting of deletions in KEAP1 [122], amplifications of EGFR [125, 126] and PIK3CA [127], downregulation of NF1 [121] and KEAP1 [111, 112], and upregulation of EGFR [123, 124]. Regarding the MIBN-based results for lung adenocarcinoma (Figure 9), there was one tree identified which was rooted by mutations in TP53, and contained mutations in BRAF, EGFR, KRAS, NF1, STK11, and PIK3CA. We also identified reports for deletions in KEAP1, KRAS, STK11, EGFR, and amplifications in EGFR, KRAS, and PIK3CA, and downregulation of NF1, STK11, KEAP1, EGFR, and upregulation of KEAP1, KRAS, STK11, TP53, and NF1.

3.4. Ovarian serous cystadenocarcinoma

Our PBAM results for ovarian cancer, shown in Figure 10, identified one major tree rooted by TP53 mutation [129, 130]. This tree contains clusters of alterations in gene amplification, mutation, and upregulation. The upregulation of EGFR and KIT in ovarian cancer has been widely reported in literature [131, 132, 133, 134]. Our results also identified two smaller trees, one rooted by amplification in GRIN2B, which appears to precede upregulation in TP53 [135, 136], and an even smaller tree rooted by amplification of PLCH1. Our MIBN results, shown in Figure 11, identified TP53 as the single precursor event of all alterations in ovarian cancer.

3.5. Prostate adenocarcinoma

Figure 12 illustrates results of the PBAM run for prostate adenocarcinoma. The tree has four main root nodes involving mutations in FRG1B, SPOP [139, 140], downregulation in PTEN [147, 148], and upregulation in APC [141]. FRG1B precedes mutations in CHEK2, which has been reported in literature [137, 138]. The tree rooted by SPOP causes deletion and downregulation of genes, including downregulation in APC [141]. The third root with downregulation in PTEN causes a variety of aberrations, including mutations in PTEN [144], downregulation in PTEN [142, 143] and CHEK2 [149], and upregulation in FOXA1 [145, 146]. The fourth tree rooted by upregulation in APC causes mostly other upregulations, such as in the gene PTEN [142, 143], and a few deletion, such as in the gene FOXA1 [145, 146]. By comparison, the MIBN results (Figure 13) have three main root nodes, one of

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

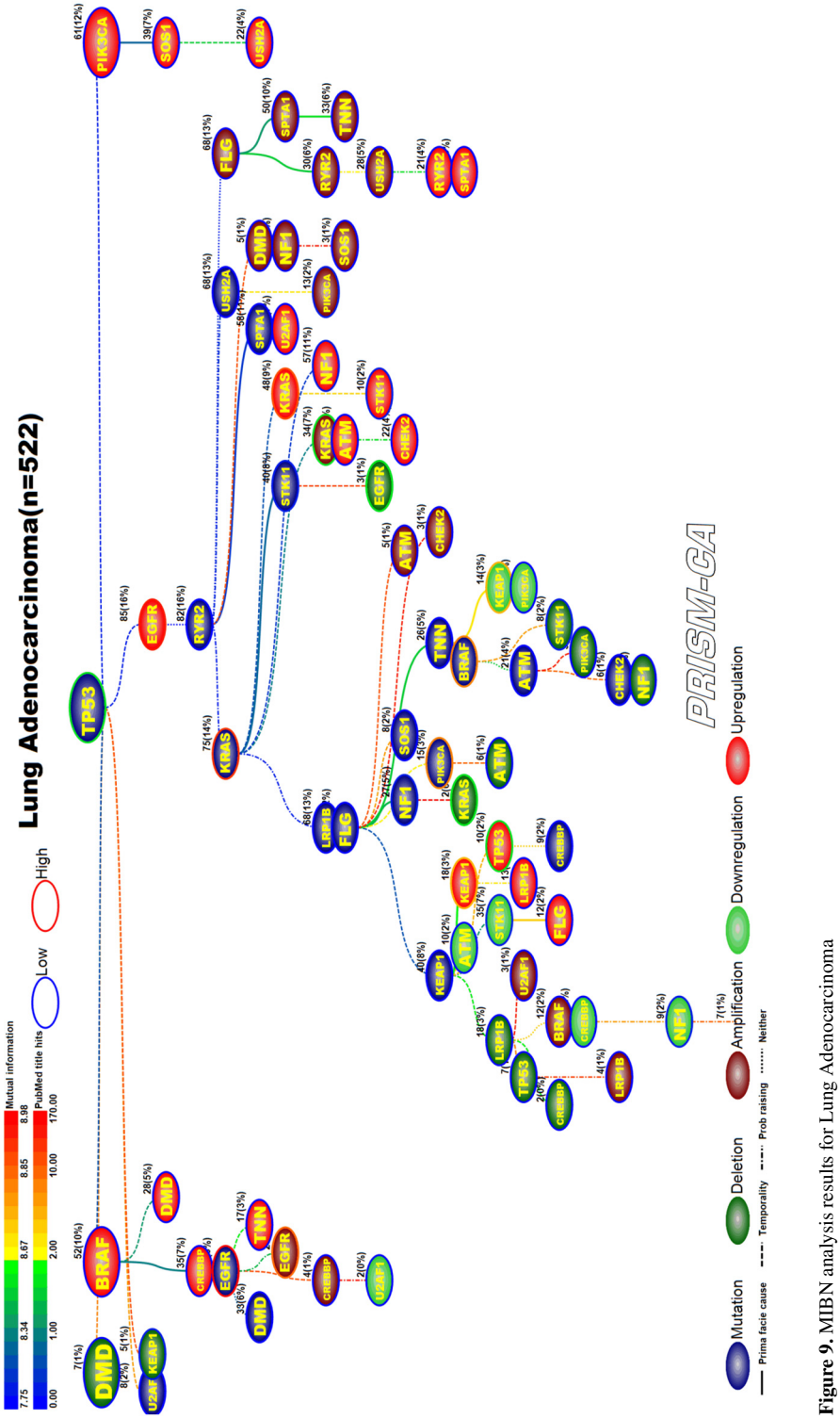

Figure 9. MIBN analysis results for Lung Adenocarcinoma

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

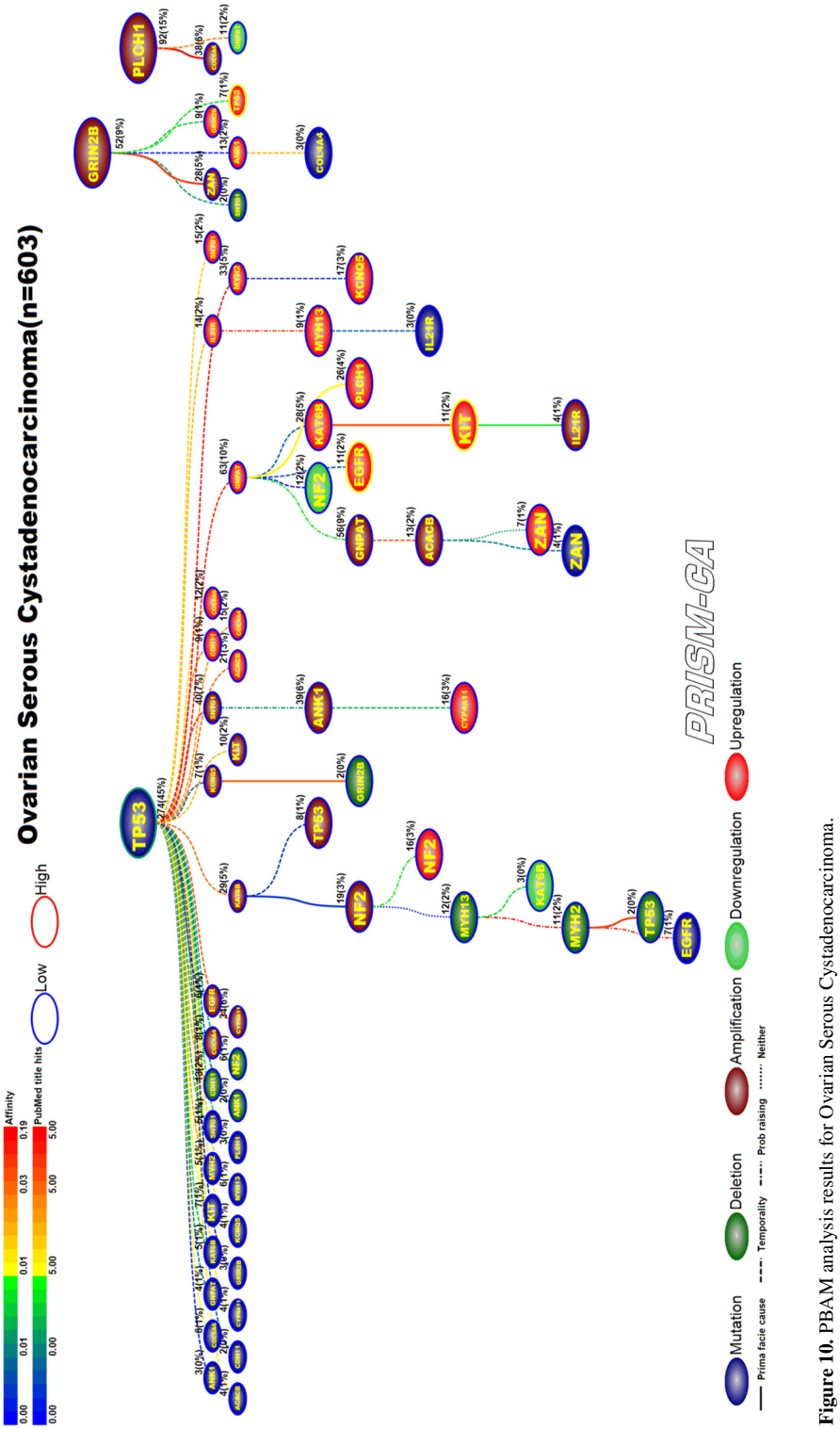

Figure 10. PBAM analysis results for Ovarian Serous Cystadenocarcinoma.

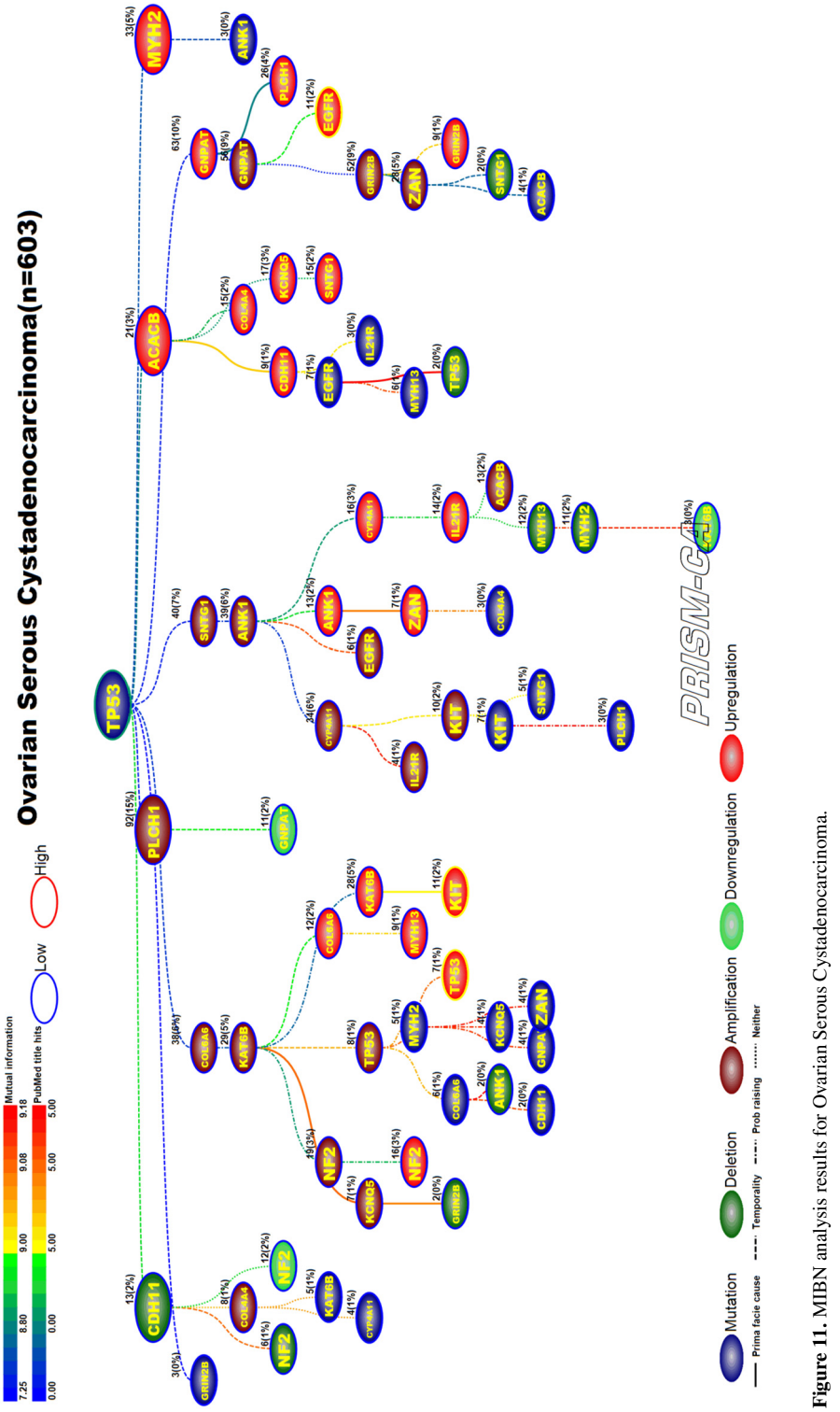

Figure 11. MIBN analysis results for Ovarian Serous Cystadenocarcinoma.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

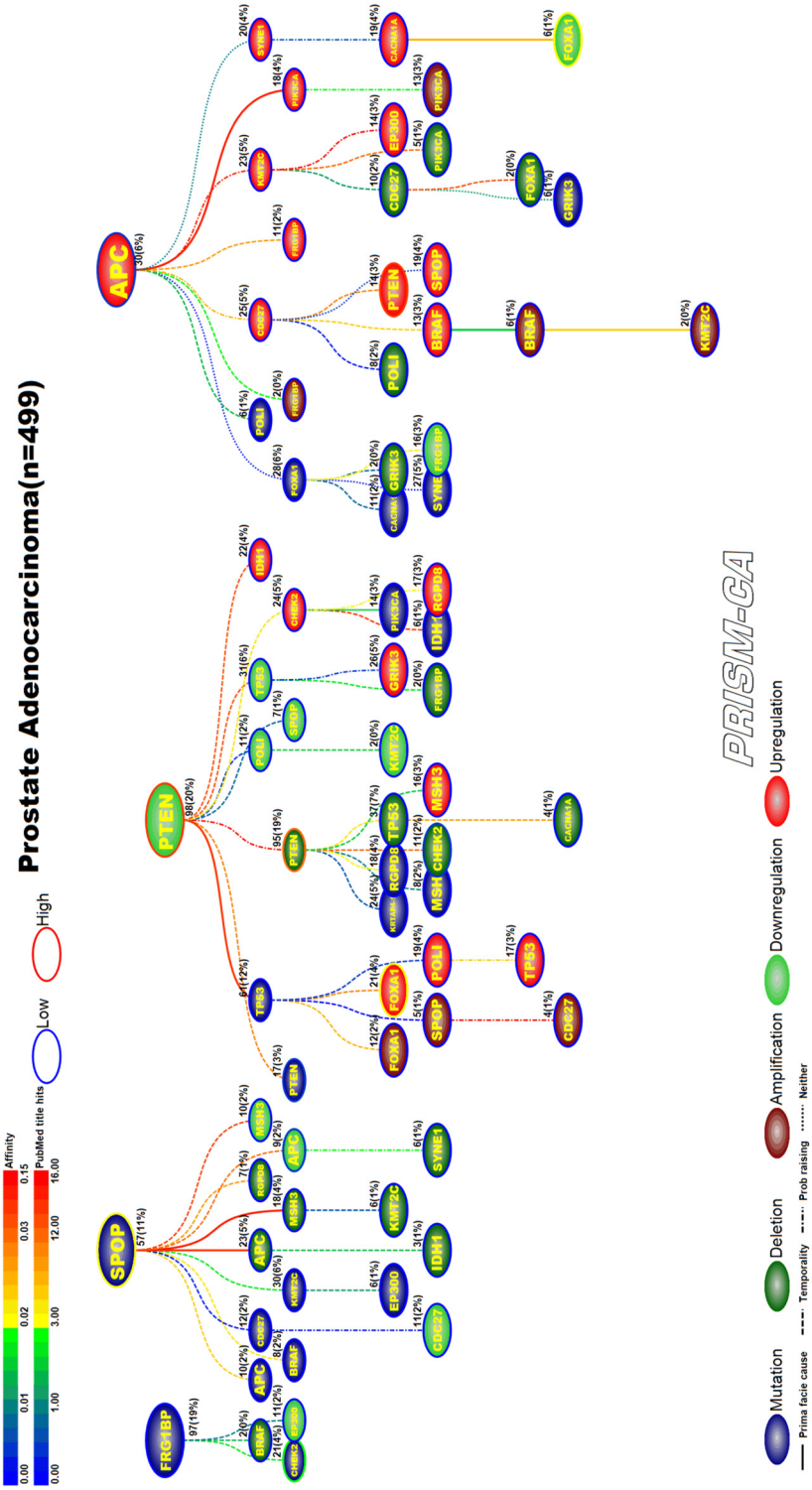

Figure 12. PBAM analysis results for Prostate Adenocarcinoma.

which is mutations of TP53. In this result, the mutation of TP53 precedes mutation in PTEN, deletion of CHEK2, downregulation of FOXA1, and upregulation of PTEN and FOXA1. TP53 was only causal to upregulation and amplification of FOXA1 in the first model. The second root, downregulation of PTEN, drives the deletion of PTEN and mutation of CHEK2. The third root, FRG1BP mutation, is causal to clusters of mutations, including SPOP, deletions, and downregulation of APC. By contrast, we did not observe FRG1B to be such a major parent node in our PBAM results.

3.6. Renal clear cell cancer, RCC

From our PBAM run of renal clear cell (RCC) cancer in Figure 14, we observed that alterations in the BAP1 gene appear the most. Our results show six trees, generally broken up by alteration. The first tree, rooted by ARAP3, is causally linked to a multitude of mutations and an upregulation of BAP1 [150, 151]. The second tree is rooted by VHL [152, 153] and drives a mixture of aberrations, including mutation of BAP1 [154], downregulation of BAP1 [150, 151], and upregulation of VHL [155, 200]. Two small trees contained only deletions, none of which have been previously reported in RCC. The tree of mostly downregulations is rooted by PBRM1 [150, 156] and precedes downregulation of SETD2 [157]. Upregulation of ACACA was the parent root of most upregulations, including PBRM1 [150, 156] and SETD2 [157]. This was quite different from our tree results of the MIBN run in Figure 15, for which we had one tree rooted by VHL mutations which was likely causal to all the other alterations in RCC.

3.7. Stomach adenocarcinoma

PBAM results for stomach adenocarcinoma can be found in Figure 16. We identified that mutations in ARID1A, LRP1B, and TP53 [172, 173] formed three major parent nodes. Mutation in ARID1A appears to be causal to chains of alterations within the same gene. For example, we observed that ARID1A mutation precedes mutation, downregulation, and deletion within the CDH1 gene [158, 159, 160, 161, 162]. Similarly, we also observed a causal link to upregulation and deletion of TP53 [163, 164]. ARID1A also had several daughter nodes with mutations, including the genes PIK3CA [165, 166, 168] and RHOA [167, 169, 170]. The tree rooted by LRP1B mutation contains a small cluster of amplifications, which includes PIK3CA [168] and KRAS [169, 170], and a cluster of upregulation, which includes PIK3CA [160, 171]. The third tree, rooted by TP53 mutation, is likely causal of groups of mutations, amplifications, downregulations, and upregulations. Downregulation was observed in ARID1A and TP53, and upregulation was observed in ARID1A, CDH1, and

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

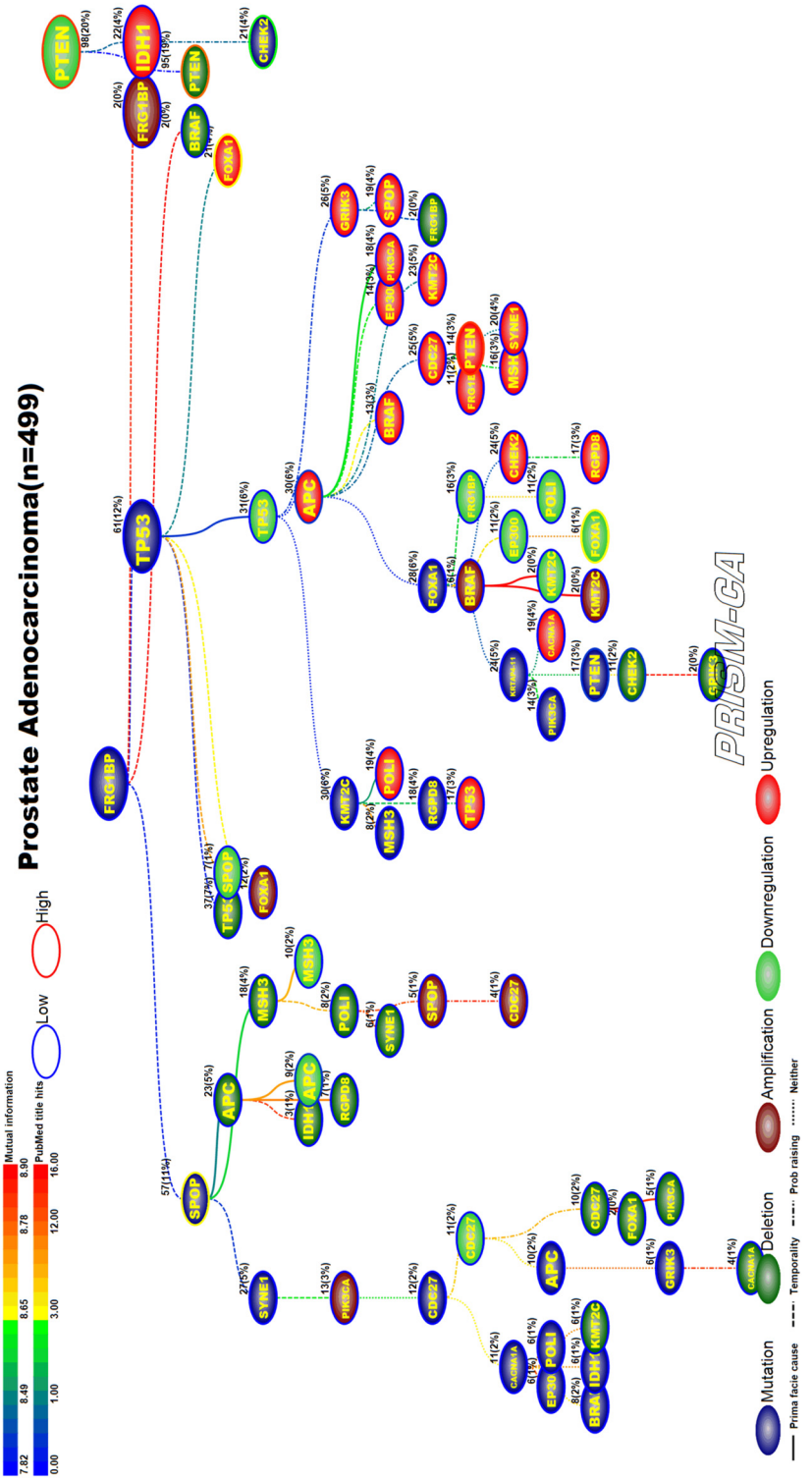

Figure 13. MIBN analysis results for Prostate Adenocarcinoma.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

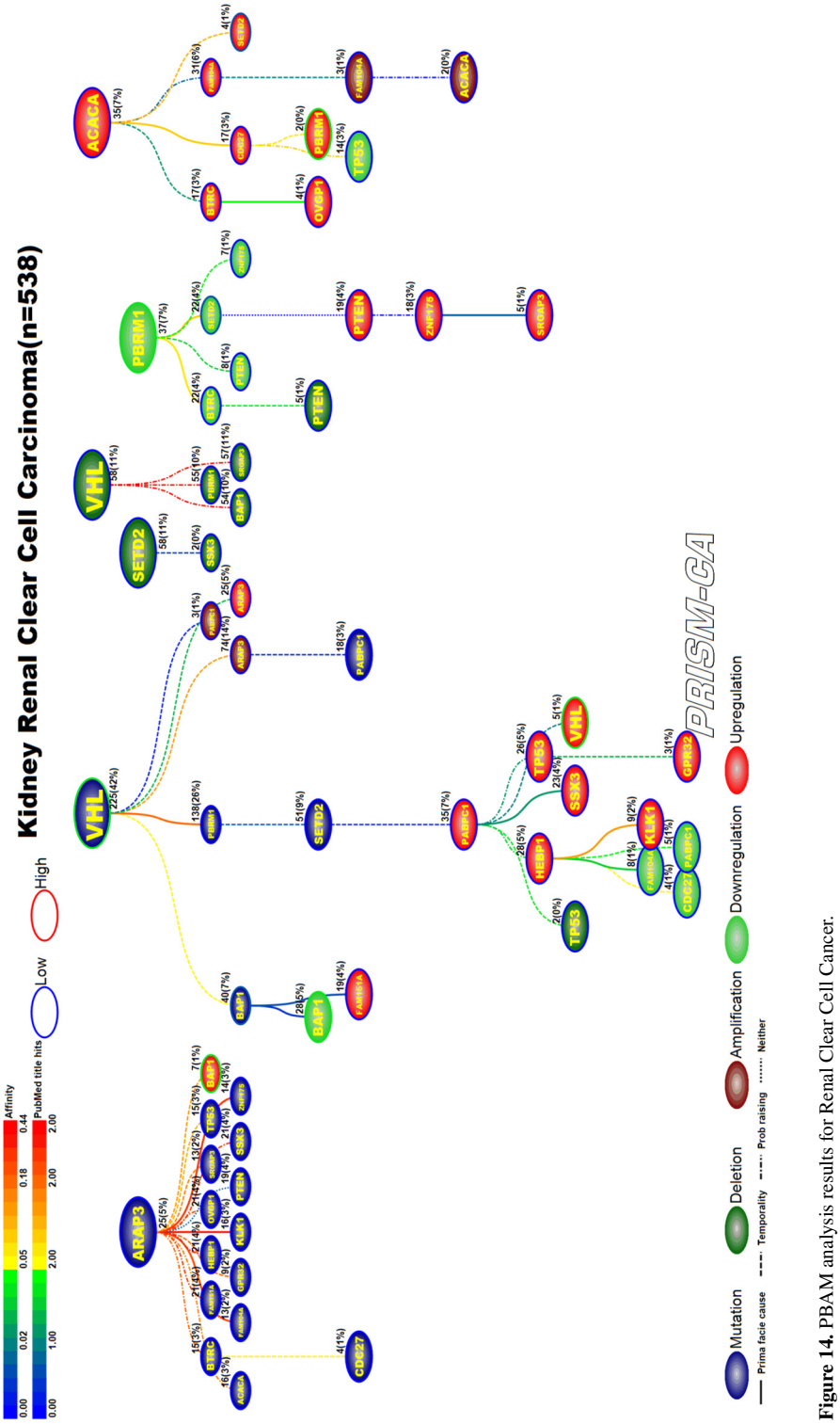

Figure 14. PBAM analysis results for Renal Clear Cell Cancer.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

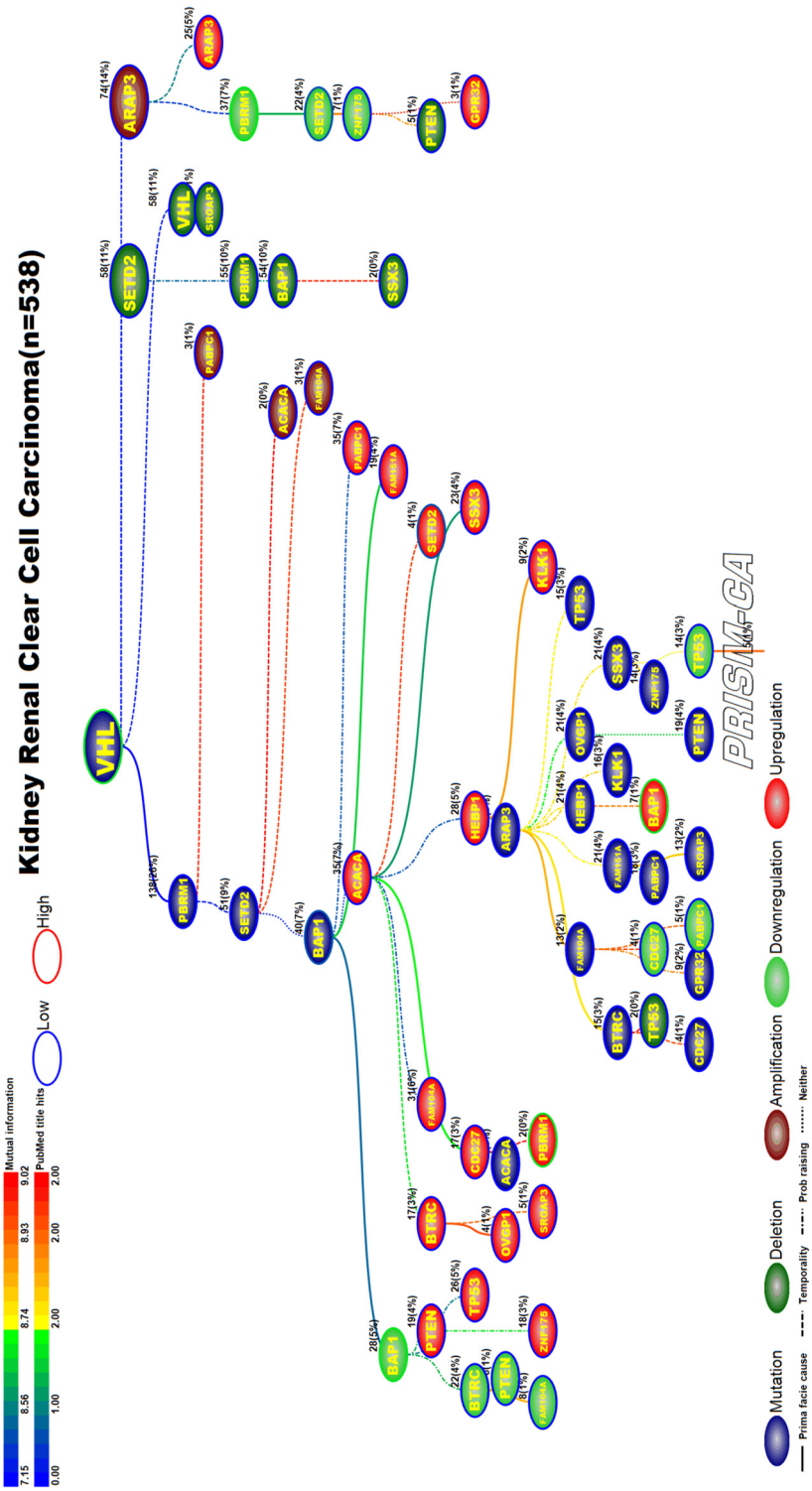

Figure 15. MIBN analysis results for Renal Clear Cell Cancer.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

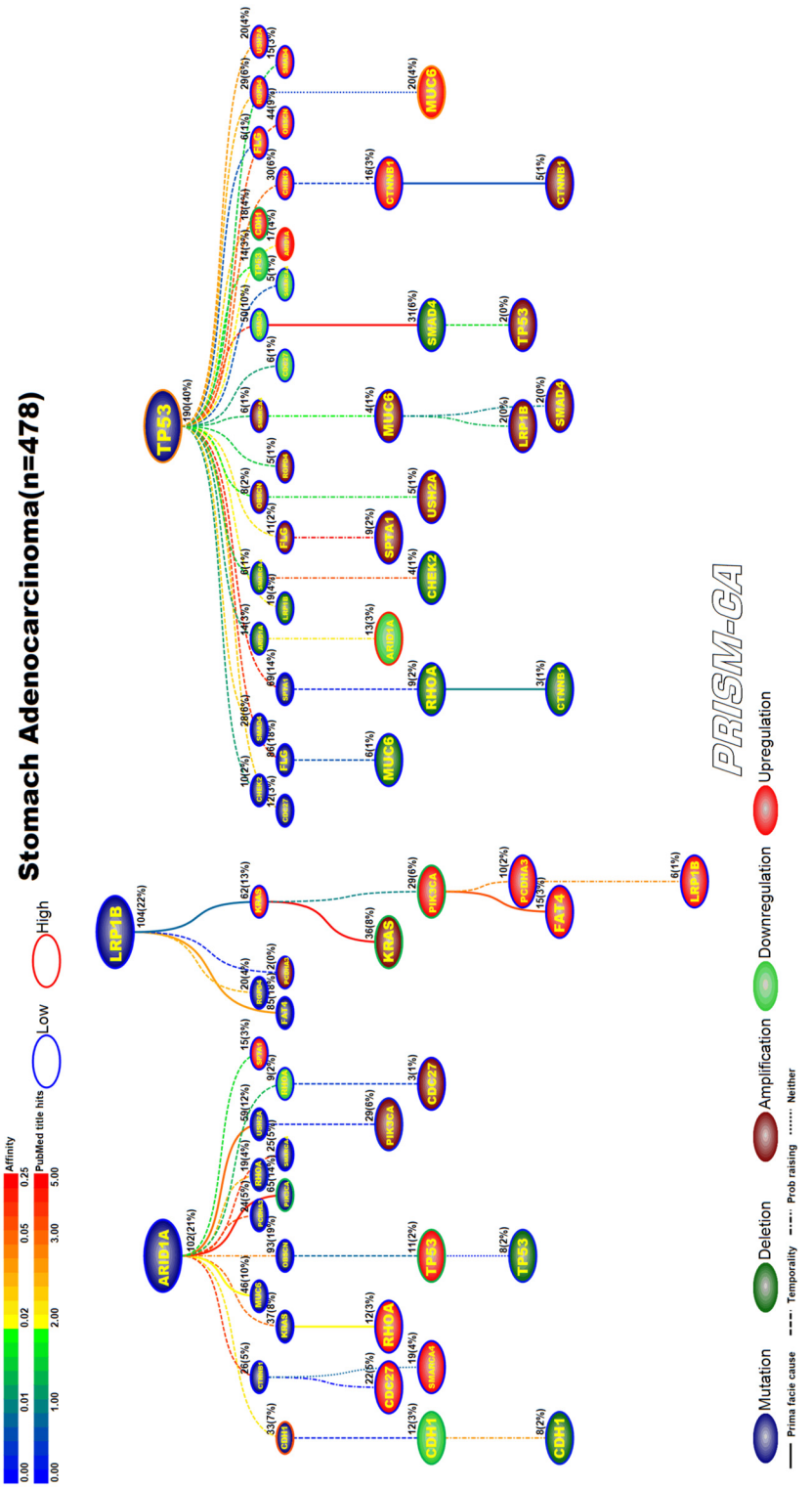

Figure 16. PBAM analysis results for Stomach Adenocarcinoma.

http://dx.doi.org/10.1016/j.heliyon.2017.e00277

MUC6, which have been reported in stomach cancer [160, 161, 163, 164, 174, 175, 176]. Our MIBN results, shown in Figure 17, identified only one parent node consisting of TP53 mutations.

3.8. Uterine corpus endometrial carcinoma

For uterine corpus endometrial carcinoma we identified two parent roots (Figure 18) with the PBAM method. One tree root, mutation of PTEN [177, 178], is causally linked to only other mutations, among which were the genes ARID1A, CTNNB1, FGFR2, KRAS, AKT1 [177, 179, 180, 181, 182, 183]. The second tree root, mutation of PIK3CA [183], is causal to the mutation of TP53, which is the driver for a cluster of amplifications, including KRAS [181], and a cluster of upregulations which include KRAS, ARID1A, and PTEN [182, 184, 185, 186, 187]. PIK3CA also drove the deletion and downregulation of PTEN [187, 188, 189, 190]. This was similar to our MIBN results (Figure 19) which show PTEN mutation as the only parent node and PIK3CA mutation as a daughter node but with similar causality as seen in the first run.

3.9. Acute myelogenous leukemia, AML

Figure 20 shows our results for a PBAM analysis of AML. There were five main root nodes observed for mutations in FLT3, IDH2, NPM1, TP53, and upregulation of RAD21. Most of these mutations have been observed previously in AML [191, 192, 201, 210, 211]. Our first tree shows that mutations in FLT3 will precede mutations in WT1 [193]. In the second tree we see that mutations in IDH2 cause further mutation in RUNX1 [194], as well as a host of upregulations in genes such as FLT3, RUNX1, and WT1 [195, 196, 197, 198, 199, 200]. The tree rooted by NPM1 drives mostly mutations, including the genes DNMT3A, TET2, and KIT [194, 202, 203, 204, 207], however, there is upregulation in genes such as CEBPA and NPM1 [205, 206, 208, 209]. The tree rooted by TP53 drives a mixture of gene aberrations, but nothing previously reported in literature for AML. The final tree shows upregulation of RAD21 as a driver for upregulation in other genes, including KIT [212, 213]. By comparison, Figure 21 shows the causal MIBN-based WDAG for AML mutations. Our results typically showed fewer trees and more daughters resulting from WDAGs, which we can see is true for AML. Our WDAG for AML had two main root nodes observed for mutations in FLT3 and TP53. FLT3 mutations also precede additional mutations which were observed in NPM1, DNT3A, KIT, RUNX1, WT1, and TET2. Our WDAG results also identified a causal inference for transcriptional alterations in genes such as FLT3, RUNX1, CEBPA, KIT, NPM1, and WT1.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

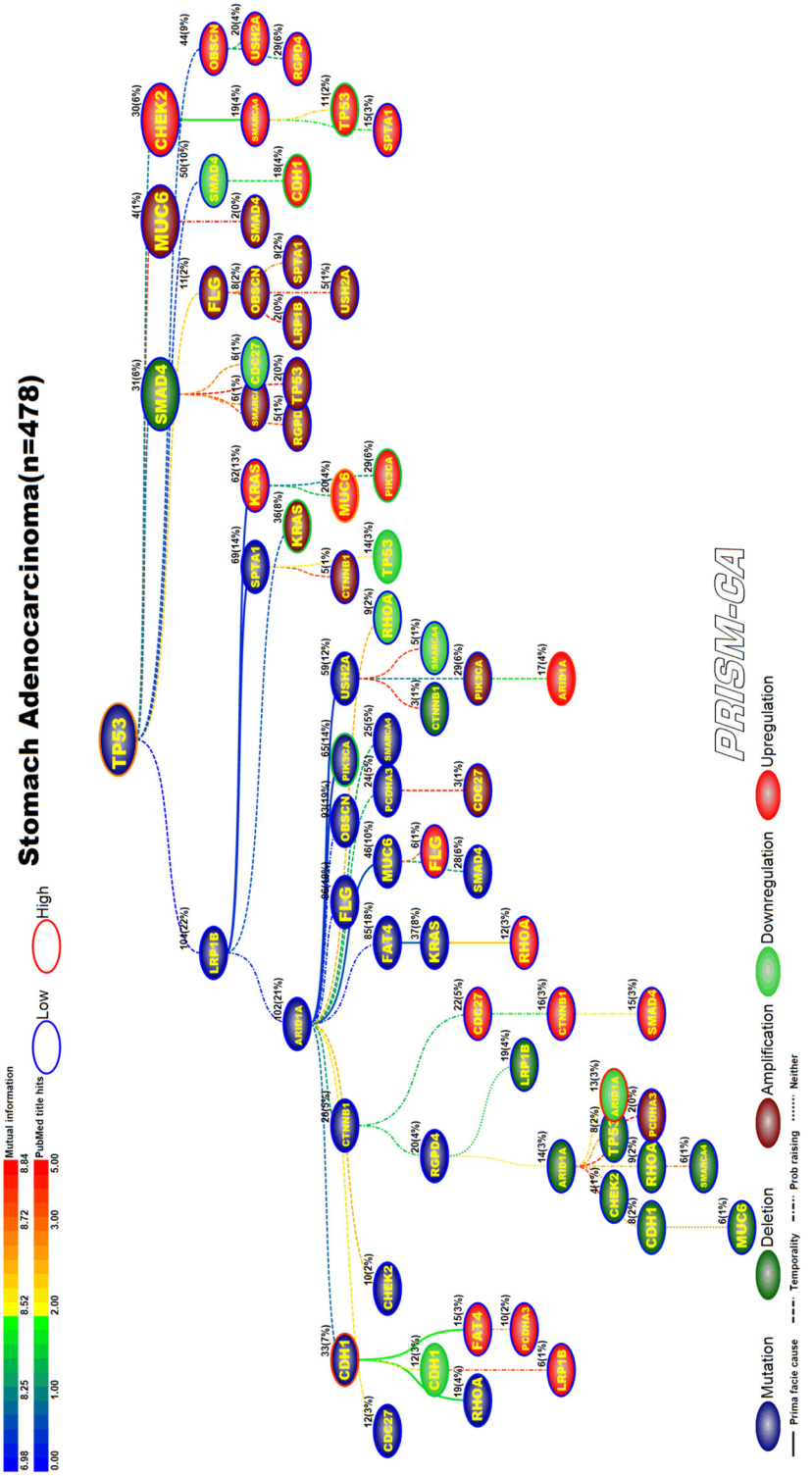

Figure 17. MIBN analysis results for Stomach Adenocarcinoma.

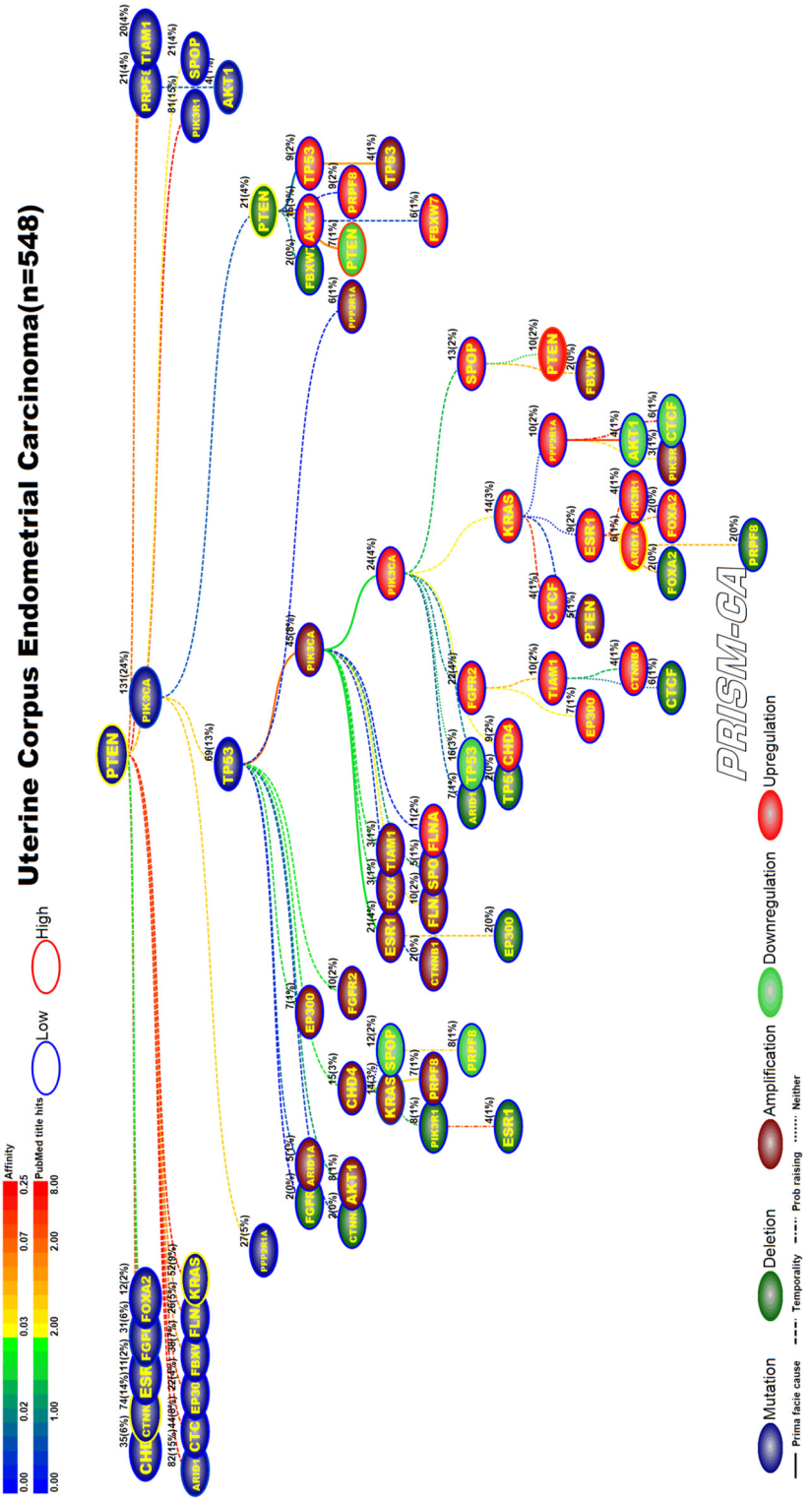

Figure 18. PBAM analysis results for Uterine Corpus Endometrial Carcinoma.

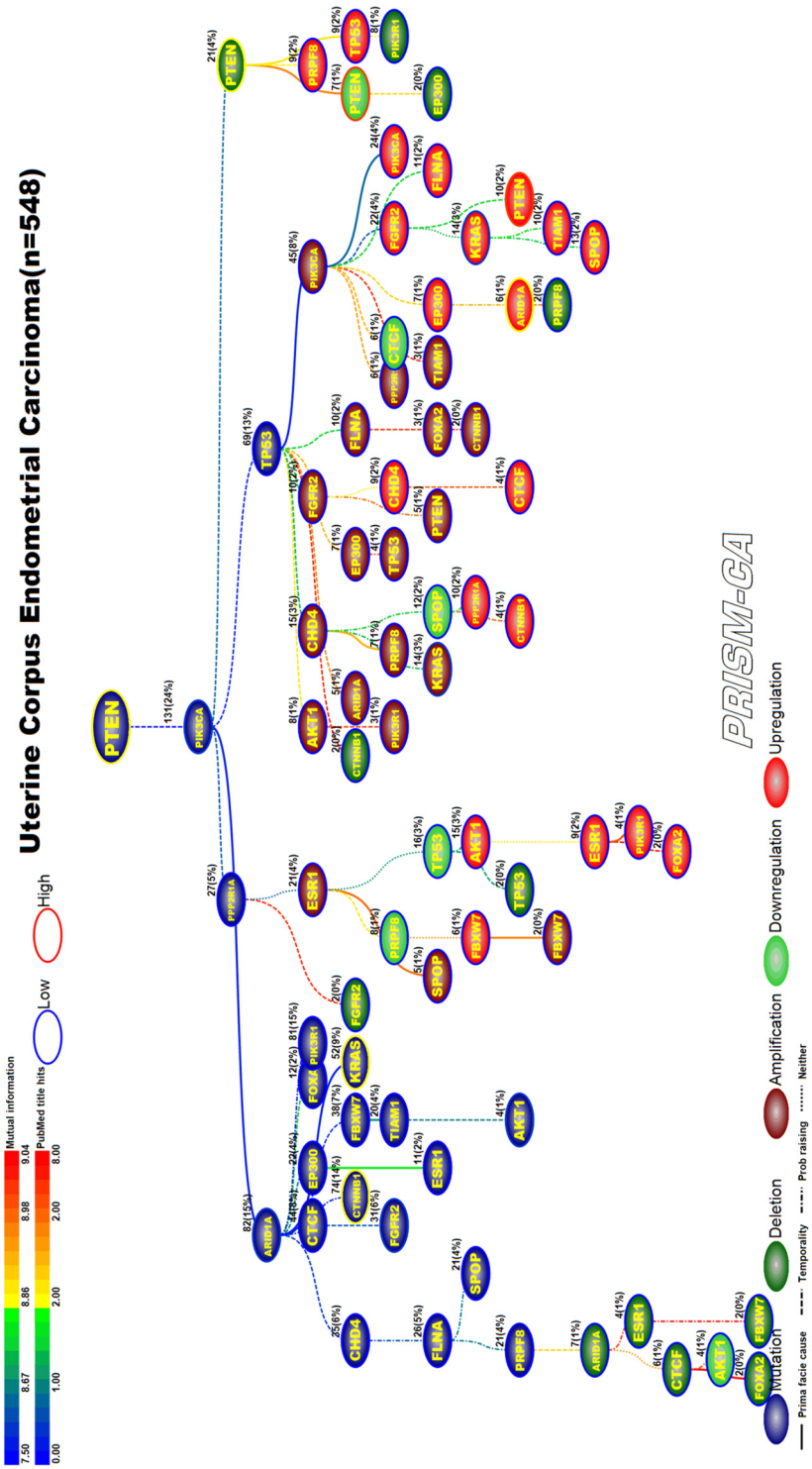

Figure 19. MIBN analysis results for Uterine Corpus Endometrial Carcinoma.

http://dx.doi.org/10.1016/j.heliyon.2017.e00277

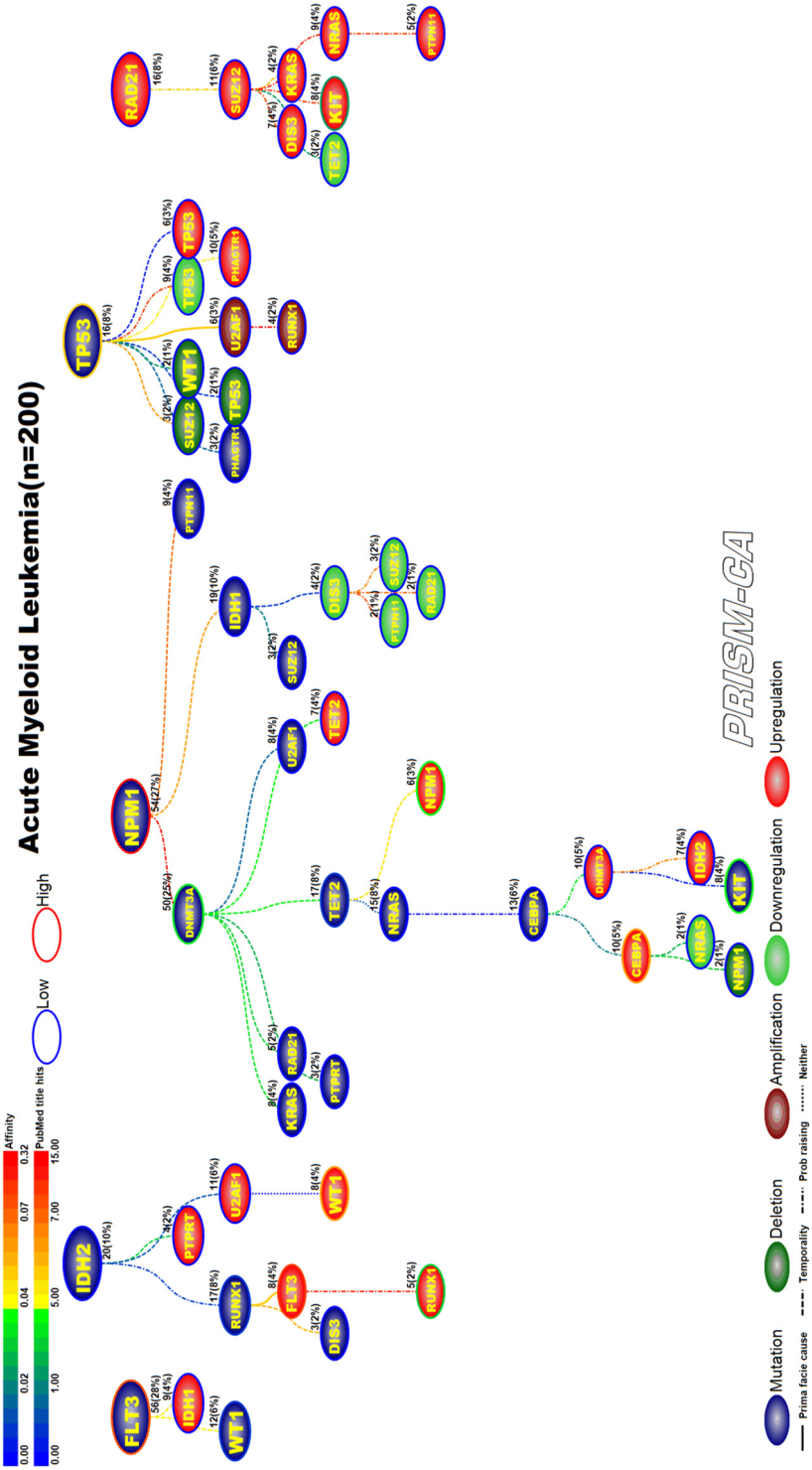

Figure 20. PBAM analysis results for Acute Myelogenous Leukemia.

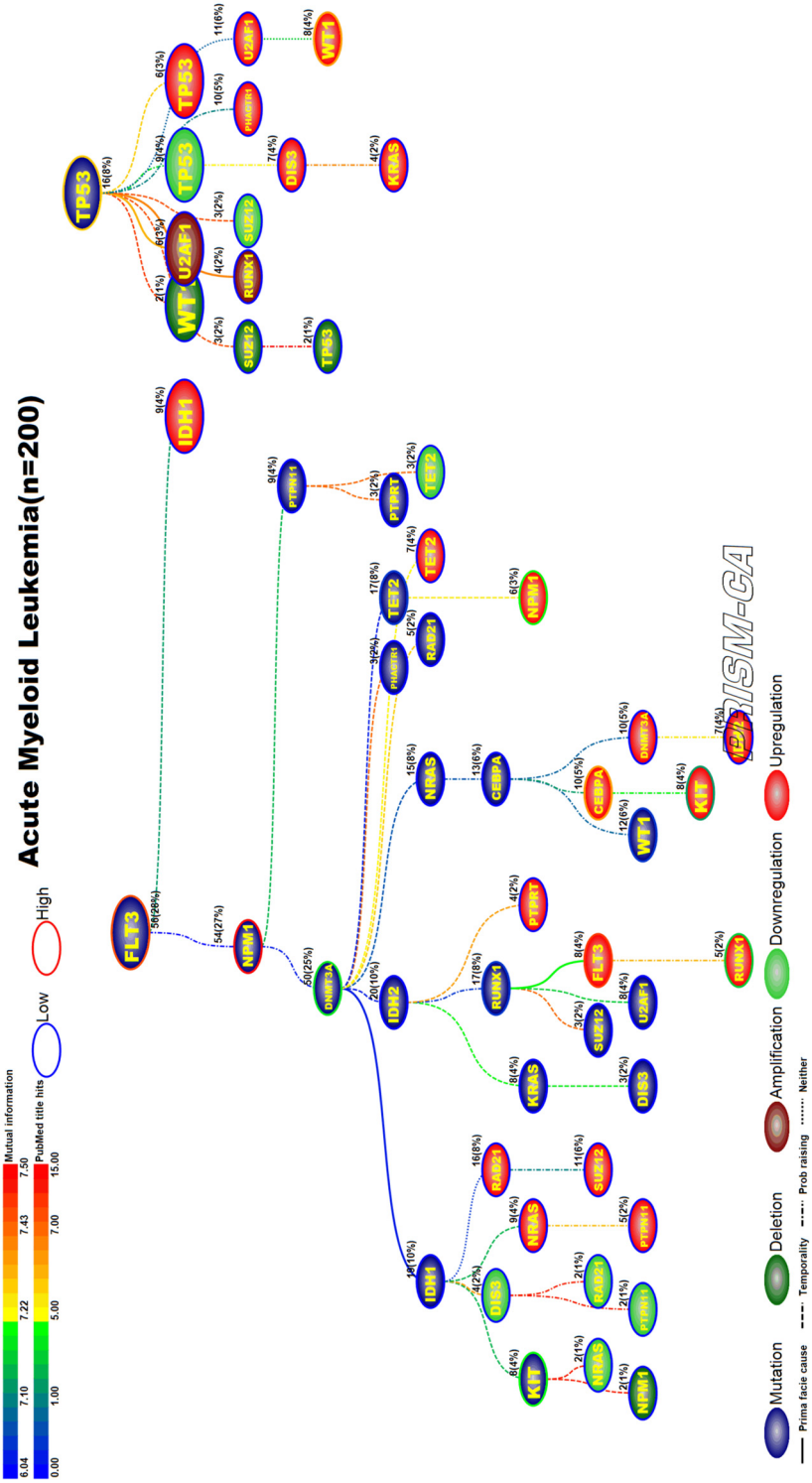

Figure 21. MIBN analysis results for Acute Myelogenous Leukemia.

http://dx.doi.org/10.1016/j.heliyon.2017.e00277

3.10. Brain lower grade gliomas

Low grade brain gliomas (Figure 22 and 23) was one of the few cancers that we did not observe any major difference between the PBAM and MIBN runs. Both runs identified two main root nodes involving mutations in IDH1 [214, 215] and upregulation of EGFR [216, 217, 218, 219]. In both results mutations in IDH1 preceded mutations in ATRX [216] as well as a variety of deletions, amplifications, and upregulations. The upregulation of NOTCH1 has been previously reported in low grade glioma [217]. Upregulation of EGFR seemed to be causally linked to amplifications in EGFR [220, 221], mutations in NF1 [222], upregulation of TP53 [128], and transcriptional alterations in other genes.

4. Discussion

Cancer is a relatively short-term evolutionary process involving initiation and progression [223]. Genetic variation is the key to evolutionary existence and selection, and in cancer this variation is generated via somatic mutations. Natural selection determines the fate of somatic mutations by purifying selection, or reducing the likelihood that deleterious mutations persist, and by positive selection, for which functionally advantageous mutations persist. Tumor mutations persisting until the point of nextgen sequencing are referred to as substitutions, which are responsible, in part, for transformation, growth and progression, drug resistance, invasion and neovascularization, and metastasis. In addition to the viewpoint concerning mutations in tumor suppressor and oncogenes, genetic adaptation of tumors is commonly driven through somatic mutations in genes with basic cellular functions [224]. Therefore, somatic mutations among genes expressed globally throughout many tissues cannot be refuted as a major determinant of cancer. To date, the large-scale nextgen sequencing studies have revealed a high degree of prevalence of somatic mutations in human cancers. As more tumor DNA sequences are analyzed, there will continue to be new information available regarding driver and passenger genes, mutations in housekeeping genes, and the role of somatic mutations in cancer.

Our PBAM approach employed iterative calculations to determine gene pairwise selectivity relationships using data on gene-specific alterations and their permutations. The results support tumor progression inference models of evolutionary trajectories, which can be applied to longitudinal studies and clinical trials for the purpose of stratifying patients based on somatic driver events and gene expression. The MIBN method was used as an alternative to PBAM to reveal results hinged to mutual information via a straightforward BN approach. Overall, our models allowed us to make inferences about probabilistic causal models to draw

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

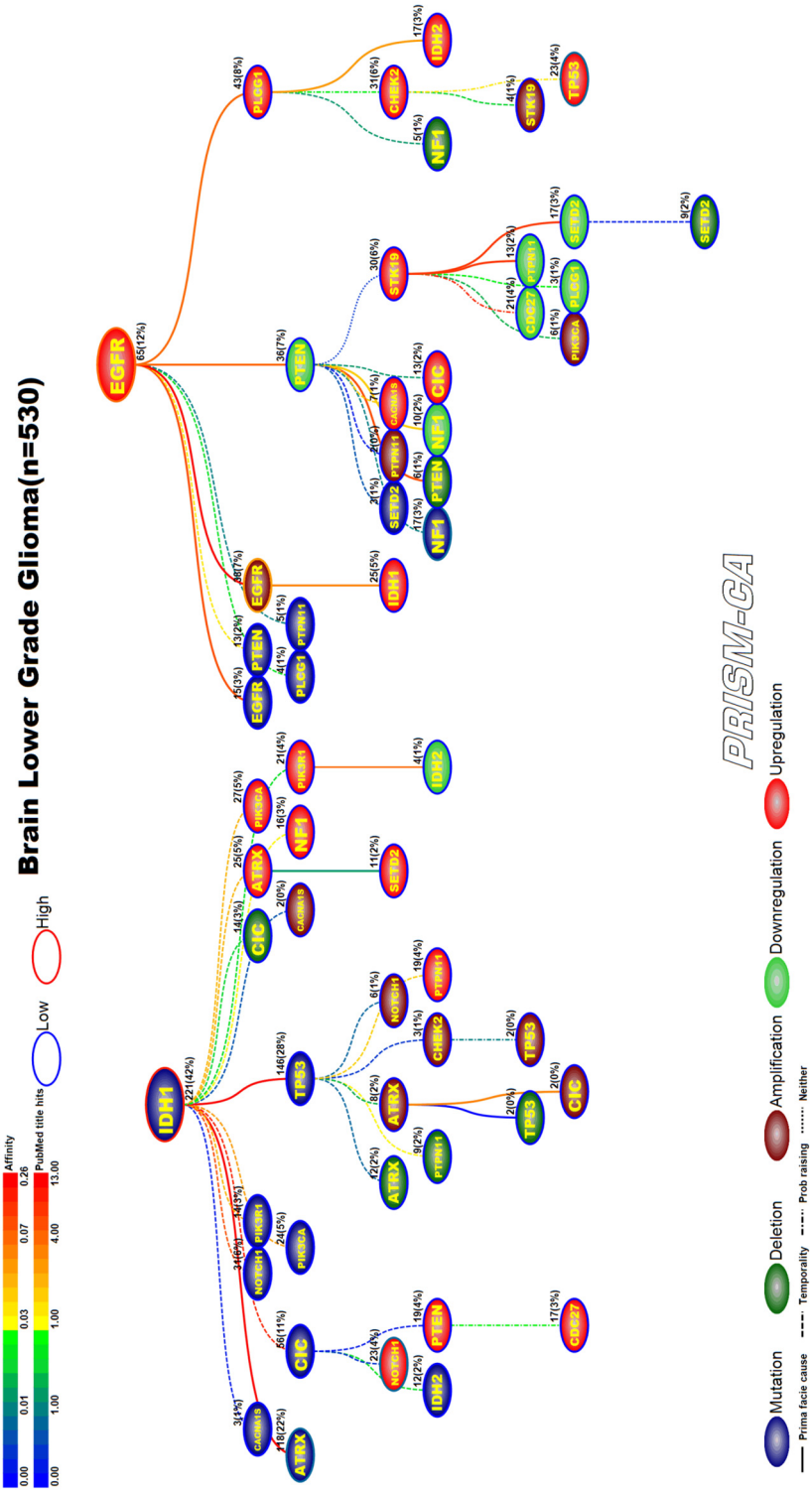

Figure 22. PBAM analysis results for Brain Lower Grade Gliomas.

http://dx.doi.org/10.1016/j.heliyon.2017.e00277

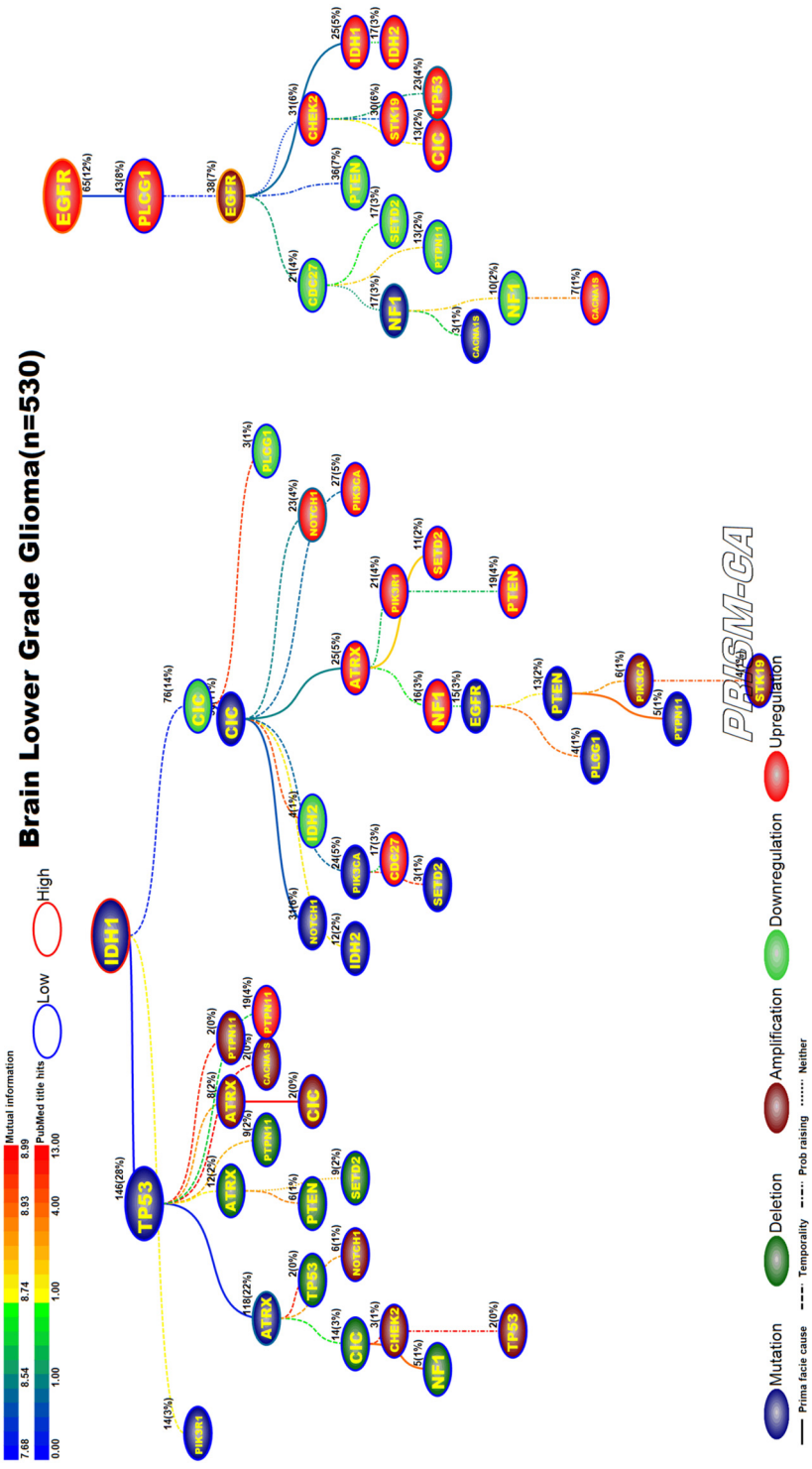

Figure 23. MIBN analysis results for Brain Lower Grade Gliomas.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

conclusions about the relationships between genomic alterations among key driver genes for common cancers.

For the cancers investigated, PBAM and MIBN were able to partition events into hierarchical trees representing groups of patients with similar events. Altogether, our results indicate that driver mutations in TP53 were the most common across the cancers considered, which support prima facie causal inference for a host of other aberrations such as deletions, amplifications, downregulation, and upregulation. The major driver events in root nodes of trees were often TP53, APC, PICK3CA, PTEN, SMAD4, KRAS, BRAF, EGFR, IDH1, SPOP, VHL, etc., which revealed the importance of these driver genes and their alterations. A major difference between PBAM and MIBN was that while PBAM partitioned the events to distinct clusters of patients, MIBN tended to agglomerate the events into a single hierarchy – mostly as a result of mutual information. We tend to favor the results of PBAM over MIBN because it partitions events(patients) into distinct evolutionary trajectories of events, for which there is a major driver event. MIBN results would suggest that there is commonly a single event with a selective advantage over all subsets of events. Thus, the partitioning characteristic of PBAM is more amenable to the clustering of patients with distinct histories of evolutionary trajectories that parallels patient diagnosis, treatment, and follow-up.

The translational value of our results are established by the potential identification of novel patient genotypes, which could prove useful in future studies of molecular markers of therapy and metastatic prediction. Our future investigations will link TCGA clinical data for recurrence, survival, and metastasis to the trees generated to identify whether certain patterns of events confer various levels of risk. Application of the PBAM and MIBN approaches to TCGA data has enabled us to view cancer from a distant perspective based on high-granularity genomic alterations which occurred in major driver genes. This view will hopefully enforce an appreciation among oncologists and biologists for the translational value of diagnostically partitioning patients according to their deleterious genomic alterations, enrolling such patients in trials involving single- or multi-label treatments associated with prolonged survival, and pursuing longitudinal studies to improve therapeutic strategies.

We did not comparatively assess numerous techniques for their computational efficiency, scalability, or differences in selectivity relationships. We also did not employ a gold-standard to establish false positive and false negative rates. Rather, we highlighted differences in the polytrees portrayed by PBAM and MIBN. Estimation of the hierarchical structure of gene selectivity events may help identify the major partitions of patients having various mixtures of genomic alterations, as well as identify early and late driver events which may confer a positive selection advantage during tumor progression. The work presented here suggests that investigation of the

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

selectivity relationships among genes can provide new insights into the development of human cancer and can establish new leads for future research on molecular diagnosis and therapeutics for cancer.

There are several challenging issues surrounding development of tumor progression models using TCGA data. First, there is the problem of unknown upstream effects of germline polymorphisms which may result in a variety of results including deletions and amplifications. In low grade gliomas, we did observe a tree rooted by upregulation of EGFR which has been reported [32], and in breast cancer, we observed a tree rooted by upregulation of ERBB2 (HER2), which has been reported to be the result of amplification in 18–20% of breast cancer cases [225]. Second, there were cases of primary events involving gene loss, for which we observed downregulation of SMAD4 as a root node in colorectal adenocarcinoma, which confers worse survival in stage I-II patients [226] and early recurrence after therapy [227]. Third, tumor heterogeneity is another hallmark of cancer that cannot be easily overcome when constructing models of tumor progression based on genomic alterations. TCGA data are not based on DNA and RNA extraction from single-cells, which would be helpful for elucidating heterogeneity; however, the large variation in alleles identified throughout all the TCGA samples used would exacerbate the complexity surrounding our attempt to portray tumor progression via a single picture. We also did not consider DNA methylation status, chromosome aberrations, and microsatellite instability, which would overlay more complexity on the models developed.

In conclusion, our data-driven approach to infer tumor progression models from deleterious genomic alterations in the absence of germline polymorphisms should be cautiously interpreted. Although 90–95% of cancers are sporadic, there is nevertheless great importance in the inherited component of cancer, for which the long-term lifetime chronicity of exposure to genetic polymorphisms can seed genomic alterations which were not considered in this investigation.

5. Conclusions

We employed computational methods to derive selectivity relationships between deleterious genomic alterations available in the The Cancer Genome Atlas. Results of our computational methods translate to portraits of evolutionary trajectories of events among major cancer driver genes. The utility of our results can be realized by oncologists and biologists who envision partitions of clusters of patients for which selectivity relationships confer certain outcomes. We conclude that the portraits which emerged during construction of graphs presented can be employed for longitudinal studies of cancer patients, to fuse genotype data with prognostic indicators of recurrence, metastasis, and survival.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

Declarations

Author contribution statement

Leif E. Peterson: Conceived and designed the experiments; Wrote the paper. Tatiana Kovyrshina: Analyzed and interpreted the data.

Funding statement

This work was funded by NASA (grant NNX12AO52A).

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

[1] J. Luo, N.L. Solimini, S.J. Elledge, Principles of cancer therapy: oncogene and non-oncogene addiction, Cell 136 (2009) 823–837.

[2] D. Hanahan, R.A. Weinberg, The hallmarks of cancer, Cell 100 (2000) 57–70.

[3] G. Kroemer, J. Pouyssegur, Tumor cell metabolism: cancer's Achilles' heel, Cancer Cell 13 (2008) 472–482.

[4] A.L. Jackson, L.A. Loeb, The mutation rate and cancer, Genetics 148 (1998) 1483–1490.

[5] M.S. Lawrence, et al., Mutational heterogeneity in cancer and the search for new cancer – associated genes, Nature 499 (2013) 214–218.

[6] M. Lynch, The cellular, developmental, and population-genetic determinants of mutation-rate evolution, Genetics 180 (2008) 933–943.

[7] V.S. Borgdorff, G.M. van Hees-Stuivenberg, N. Meyers, Spontaneous and mutagen-induced loss of DNA mismatch repair in MSH2-heterozygous mammalian cell, Mutat. Res. 574 (2005) 50–57.

[8] D.C. Hegan, L. Narayana, W. Jirik, R.M. Edelmann, R.M. Liskay, P.M. Glazer, Differing patterns of genetic instability in mice deficient in the mismatch

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

repair genes PMS2, MLH1, MSH2, MSH3, and MSH6, Carcinogenesis 27 (2006) 2402–2408.

[9] K.L. Burr, P. van-Duyn-Goedhardt, K. Hickenbotham, P.P. Monger, et al., The effects of MSH2 deficiency on spontaneous and radiation-induced mutation rates in the mouse germline, Mutat. Res. 617 (2007) 147–151.

[10] E.C. Chao, S.M. Lipkin, Molecular models for the tissue specificity of DNA mismatch repair-deficient carcinogenesis, Nucleic Acids Res. 34 (2006) 840–852.

[11] M.A. Coolbaugh-Murphy, L. Maleki, M. Ramagli, B. Frazier, et al., Estimating mutant microsatellite allele frequencies in somatic cells by small-pool PCR, Genomics 84 (2004) 419–430.

[12] H.E. Alazzouzi, S. Domingo, I. González, M. Blanco, et al., Low levels of microsatellite instability characterize MLH1 and MSH2 HNPCC carriers before tumor diagnosis, Hum. Mol. Genet. 14 (2005) 235–239.

[13] B. Vogelstein, N. Papadopoulos, V.E. Velculescu, S. Zhou, L.A. Diaz Jr., K.W. Kinzler, Cancer genome landscapes, Science 339 (2013) 1546–1558.

[14] C.S. Attolini, Y.K. Cheng, R. Beroukhim, et al., A mathematical framework to determine the temporal sequence of somatic genetic events in cancer, Proc. Natl. Acad. Sci. USA 107 (41) (2010) 17604–17609.

[15] A. Youn, R. Simon, Identifying cancer driver genes in tumor genome sequencing studies, Bioinformatics 27 (2) (2011) 175–181.

[16] L. Ding, et al., Somatic mutations affect key pathways in lung adenocarcinoma, Nature 455 (7216) (2008) 1069–1075.

[17] L.D. Wood, D.W. Parsons, S. Jones, J. Lin, T. Sjöblom, R.J. Leary, et al., The genomic landscapes of human breast and colorectal cancers, Science 318 (5853) (2007) 1108–1113.

[18] P. Lecca, N. Casiraghi, F. Demichelis, Defining order and timing of mutations during cancer progression: the TO-DAG probabilistic graphical model, Front. Genet. 6 (2015) 309.

[19] M. Gerstung, N. Eriksson, J. Lin, B. Vogelstein, N. Beerenwinkel, The temporal order of genetic and pathway alterations in tumorigenesis, PLoS ONE 6 (11) (2011) e27136.

[20] H. Kang, K.H. Cho, X.D. Zhang, T. Zeng, L. Chen, Inferring sequential order of somatic mutations during tumorigenesis based on Markov chain model, IEEE/ACM Trans. Comput. Biol. Bioinform. 12 (5) (2015) 1094–1103.

[21] D. Ramazzotti, G. Caravagna, L. Olde Loohuis, A. Graudenzi, I. Korsunsky, G. Mauri, et al., CAPRI: efficient inference of cancer progression models from cross-sectional data, Bioinformatics 31 (18) (2015) 3016–3026.

[22] G. Caravagna, A. Graudenzia, D. Ramazzotti, R. Sanz-Pamplona, L. De Sano, G. Mauri, et al., Algorithmic methods to infer the evolutionary trajectories in cancer progression, Proc. Natl. Acad. Sci. USA 113 (28) (2016) E4025–E4034.

[23] NCI and the NHGRI, The cancer genome atlas, 2005.

[24] E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, et al., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, Cancer Discov. 2 (2012) 401.

[25] J. Gao, et al., Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal, Sci. Signal. 6 (2013) 11.

[26] W.C. Cheng, I.F. Chung, C.Y. Chen, H.J. Sun, J.J. Fen, W.C. Tang, et al., DriverDB: an exome sequencing database for cancer driver gene identification, Nucleic Acids Res. 42 (Database issue) (2014) D1048-54.

[27] C.H. Mermel, S.E. Schumacher, B. Hill, M.L. Meyerson, R. Beroukhim, G. Getz, GISTIC 2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers, Genome Biol. 12 (4) (2011) R41.

[28] P. Suppes, A Probabilistic Theory of Causality, North-Holland Publishing, Amsterdam, 1970.

[29] L.O. Loohuis, G. Caravagna, A. Graudenzi, D. Ramazzotti, G. Mauri, M. Antoniotti, et al., Inferring tree causal models of cancer progression with probability raising, PLoS ONE 9 (10) (2014).

[30] C.K. Chow, C.N. Liu, Approximating discrete probability distributions with dependence trees, IEEE Trans. Inf. Theory IT-14 (1968) 3462–3467.

[31] R.C. Prim, Shortest connection networks and some generalizations, Bell Syst. Tech. J. 36 (6) (1957) 1389–1401.

[32] R.G. Verhaak, K.A. Hoadley, E. Purdom, V. Wang, Y. Qi, M.D. Wilkerson, et al., Cancer genome atlas research network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1, Cancer Cell 17 (1) (2010) 98–110.

[33] A.B. Adomas, S.A. Grimm, C. Malone, M. Takaku, J.K. Sims, P.A. Wade, et al., Breast tumor specific mutation in GATA3 affects physiological

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

mechanisms regulating transcription factor turnover, BMC Cancer 14 (2014) 278.

[34] J. Lauring, D.P. Cosgrove, S. Fontana, J.P. Gustin, H. Konishi, A.M. Abukhdeir, et al., Knock in of the AKT1 E17K mutation in human breast epithelial cells does not recapitulate oncogenic PIK3CA mutations, Oncogene, Author manuscript, available in PMC 2011 Feb. 22.

[35] I.M. Chu, W. Lai, O. Aprelikova, L.H. Touny, H. Kouros-Mehr, J.E. Green, et al., Expression of GATA3 in MDA-MB-231 triple-negative breast cancer cells induces a growth inhibitory response to TGFß, PLoS ONE 8 (4) (2013) e61125.

[36] A. Cimino-Mathews, A.P. Subhawong, P.B. Illei, R. Sharma, M.K. Halushka, R. Vang, et al., GATA3 expression in breast carcinoma, Hum. Pathol., Author manuscript, available in PMC 2014 Apr. 18.

[37] Y. Zhou, C. Wang, H. Zhu, Y. Lin, B. Pan, X. Zhang, et al., Diagnostic accuracy of PIK3CA mutation detection by circulating free DNA in breast cancer, PLoS ONE 11 (6) (2016) e0158143.

[38] M. Janiszewska, L. Liu, V. Almendro, Y. Kuang, C. Paweletz, R.A. Sakr, et al., In situ single cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2+ breast cancer, Nat. Genet., Author manuscript, available in PMC 2016 Apr. 1.

[39] K.M. Kuusisto, A. Bebel, M. Vihinen, J. Schleutker, S. Sallinen, Screening for BRCA1, BRCA2, CHEK2, PALB2, BRIP1, RAD50, and CDH1 mutations in high-risk Finnish BRCA1/2-founder mutation-negative breast and/or ovarian cancer individuals, Breast Cancer Res. Treat. 13 (1) (2011) R20.

[40] S. Wang, J.C. Liu, D. Kim, A. Datti, E. Zacksenhaus, Targeted Pten deletion plus p53-R270H mutation in mouse mammary epithelium induces aggressive claudin-low and basal-like breast cancer, Breast Cancer Res. Treat. 18 (2016) 9.

[41] C.B. Morse, R.L. Garcia, K.E. Calhoun, E.M. Swisher, Premalignant alterations in breast and endometrium associated with a PTEN mutation in a woman with Cowden syndrome, Gynecol. Oncol. Rep. 12 (2015 Apr.) 13–16.

[42] J. Shih, B. Bashir, K.S. Gustafson, M. Andrake, R.L. Dunbrack, L.J. Goldstein, et al., CSI.

[43] M. Cizkova, A. Susini, S. Vacher, G. Cizeron-Clairac, C. Andrieu, K. Driouch, et al., PIK3CA mutation impact on survival in breast cancer patients and in ERa, PR and ERBB2-based subgroups, Breast Cancer Res. Treat. 14 (1) (2012) R28.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

[44] F.D. Amicis, S. Aquila, C. Morelli, C. Guido, M. Santoro, I. Perrotta, et al., Bergapten drives autophagy through the up-regulation of PTEN expression in breast cancer cells, Mol. Cancer 14 (2015) 130.

[45] M.S. Recouvreux, E.N. Grasso, P.C. Echeverria, L. Rocha-Viegas, L.H. Castilla, C. Schere-Levy, et al., RUNX1 and FOXP3 interplay regulates expression of breast cancer related genes, Oncotarget 7 (6) (2016 Feb. 9) 6552–6565.

[46] N. Ferrari, Z.M. Mohammed, C. Nixon, S.M. Mason, E. Mallon, D.C. McMillan, et al., Expression of RUNX1 correlates with poor patient prognosis in triple negative breast cancer, PLoS ONE 9 (6) (2014) e100759.

[47] K.C. Andrade, K.M. Santiago, F.P. Fortes, L.I. Mambelli, A.F. Nóbrega, M.I. Achatz, et al., Early-onset breast cancer patients in the South and Southeast of Brazil should be tested for the TP53 p.R337H mutation, Genet Mol Biol. 39 (2) (2016 Apr.–Jun.) 199–202.

[48] N. Shtraizent, H. Matsui, A. Polotskaia, J. Bargonetti, Hot spot mutation in TP53 (R248Q) causes oncogenic gain-of-function phenotypes in a breast cancer cell line derived from an African American patient, Int. J. Environ. Res. Public Health 13 (1) (2016 Jan.) 22.

[49] P. Lebok, V. Kopperschmidt, M. Kluth, C. Hube-Magg, C. Özden, B. Taskin, et al., Partial PTEN deletion is linked to poor prognosis in breast cancer, BMC Cancer 15 (2015) 963.

[50] M. Firoozinia, M.Z. Jahromi, S.Z. Moghadamtousi, S. Nikzad, H.A. Kadir, PIK3CA gene amplification and PI3K p110a protein expression in breast carcinoma, Int. J. Med. Sci. 11 (6) (2014) 620–625.

[51] B.C. Calhoun, B. Portier, Z. Wang, E.C. Minca, G.T. Budd, C. Lanigan, et al., MET and PTEN gene copy numbers and Ki-67 protein expression associate with pathologic complete response in ERBB2-positive breast carcinoma patients treated with neoadjuvant trastuzumab-based therapy, BMC Cancer 16 (1) (2016) 695.

[52] D.E. Carvajal-Hausdorf, K.A. Schalper, L. Pusztai, A. Psyrri, K.T. Kalogeras, V. Kotoula, et al., Measurement of domain-specific HER2 (ERBB2) expression may classify benefit from trastuzumab in breast cancer, J. Natl. Cancer Inst. 107 (8) (2015 Aug.) djv136.

[53] S. Knappskog, R. Chrisanthar, E. Lřkkevik, G. Anker, B. Řstenstad, S. Lundgren, et al., Low expression levels of ATM may substitute for CHEK2/TP53 mutations predicting resistance towards anthracycline and

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

mitomycin chemotherapy in breast cancer, Breast Cancer Res. Treat. 14 (2) (2012) R47.

[54] Y. Fu, H. Edvardsen, A. Kaushiva, J.P. Arhancet, T.M. Howe, I. Kohaar, et al., NOTCH2 in breast cancer, Mol. Cancer 9 (2010) 113.

[55] J. Liu, X. Wei, W. Huang, C. Chen, J. Bai, G. Zhang, et al., Correction, PLoS ONE 8 (7) (2013) 10.

[56] J. Liu, X. Wei, W. Huang, C. Chen, J. Bai, G. Zhang, et al., Cytoplasmic Skp2 expression is associated with p-Akt1 and predicts poor prognosis in human breast carcinomas, PLoS ONE 7 (12) (2012) e52675.

[57] H.D. Cho, J.E. Lee, H.Y. Jung, M. Oh, J. Lee, S. Jang, et al., Loss of tumor suppressor ARID1A protein expression correlates with poor prognosis in patients with primary breast cancer, Breast Cancer Res. Treat. 18 (4) (2015 Dec.) 339–346.

[58] X. Zhang, Q. Sun, M. Shan, M. Niu, T. Liu, B. Xia, et al., Promoter hypermethylation of ARID1A gene is responsible for its low mRNA expression in many invasive breast cancers, PLoS ONE 8 (1) (2013) e53931.

[59] D. Wang, C. Li, X. Zhang, The promoter methylation status and mRNA expression levels of CTCF and SIRT6 in sporadic breast cancer.

[60] E.A. Rakha, S.E. Pinder, C.E. Paish, I.O. Ellis, Expression of the transcription factor CTCF in invasive breast cancer, Br. J. Cancer 91 (8) (2004 Oct. 18) 1591–1596.

[61] W. Jacot, C. Mollevi, F. Fina, E. Lopez-Crapez, P. Martin, P. Colombo, et al., High EGFR protein expression and exon 9 PIK3CA mutations are independent prognostic factors in triple negative breast cancers, BMC Cancer 15 (2015) 986.

[62] A. Sueta, Y. Yamamoto, M. Yamamoto-Ibusuki, M. Hayashi, T. Takeshita, S. Yamamoto, et al., An integrative analysis of PIK3CA mutation, PTEN, and INPP4B expression in terms of trastuzumab efficacy in HER2-positive breast cancer, PLoS ONE 9 (12) (2014) e116054.

[63] M. Marotta, X. Chen, A. Inoshita, R. Stephens, G.T. Budd, J.P. Crowe, et al., A common copy-number breakpoint of ERBB2 amplification in breast cancer colocalizes with a complex block of segmental duplications, Breast Cancer Res. Treat. 14 (6) (2012) R150.

[64] J. Liu, X. Sun, S. Qin, H. Wang, N. Du, Y. Li, et al., CDH1 promoter methylation correlates with decreased gene expression and poor prognosis in patients with breast cancer, Oncol. Lett. 11 (4) (2016 Apr.) 2635–2643.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

[65] J.R. Caldeira, É.C. Prando, F.C. Quevedo, F.A. Neto, C.A. Rainho, S.R. Rogatto, et al., CDH1 promoter hypermethylation and E-cadherin protein expression in infiltrating breast cancer, BMC Cancer 6 (2006) 48.

[66] K. Tachi, A. Shiraishi, H. Bando, T. Yamashita, I. Tsuboi, T. Kato, et al., FOXA1 expression affects the proliferation activity of luminal breast cancer stem cell populations, Cancer Sci. 107 (3) (2016 Mar.) 281–289.

[67] L. Zheng, B. Qian, D. Tian, T. Tang, S. Wan, L. Wang, et al., FOXA1 positively regulates gene expression by changing gene methylation status in human breast cancer MCF-7 cells, Int. J. Clin. Exp. Pathol. 8 (1) (2015) 96–106.

[68] M.J. Schell, M. Yang, J.K. Teer, F.Y. Lo, A. Madan, D. Coppola, et al., A multigene mutation classification of 468 colorectal cancers reveals a prognostic role for APC, Nat. Commun. 7 (2016) 11743.

[69] D. Xu, L. Yuan, X. Liu, M. Li, F. Zhang, X. Gu, et al., EphB6 overexpression and Apc mutation together promote colorectal cancer, Oncotarget 7 (21) (2016 May 24) 31111–31121.

[70] L. Zhu, C. Dong, Y. Cao, X. Fang, C. Zhong, D. Li, et al., Prognostic role of BRAF mutation in stage II/III colorectal cancer receiving curative resection and adjuvant chemotherapy, PLoS ONE 11 (5) (2016) e0154795.

[71] L. Lin, L. Chen, Y. Wang, X. Meng, C. Liang, F. Zhou, et al., Efficacy of cetuximab-based chemotherapy in metastatic colorectal cancer according to RAS and BRAF mutation subgroups, Mol. Clin. Oncol. 4 (6) (2016 Jun.) 1017–1024.

[72] H. Niitsu, T. Hinoi, Y. Kawaguchi, K. Sentani, R. Yuge, Y. Kitadai, et al., KRAS mutation leads to decreased expression of regulator of calcineurin 2, resulting in tumor proliferation in colorectal cancer, Oncogenesis 5 (8) (2016 Aug.) e253.

[73] E. Mack, K. Stabla, J. Riera-Knorrenschild, R. Moll, A. Neubauer, C. Brendelet, et al., A rational two-step approach to KRAS mutation testing in colorectal cancer using high resolution melting analysis and pyrosequencing, BMC Cancer 16 (2016) 585.

[74] H. Taniguchi, K. Yamazaki, T. Yoshino, K. Muro, Y. Yatabe, T. Watanabe, et al., Japanese society of medical oncology clinical guidelines, Cancer Sci. 106 (3) (2015 Mar.) 324–327.

[75] B. Kleist, M. Kempa, M. Novy, C. Oberkanins, L. Xu, G. Li, et al., Comparison of neuroendocrine differentiation and KRAS/NRAS/BRAF/PIK3CA/TP53

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

mutation status in primary and metastatic colorectal cancer, Int. J. Clin. Exp. Pathol. 7 (9) (2014) 5927–5939.

[76] S.K. Nam, S. Yun, J. Koh, Y. Kwak, A.N. Seo, K.U. Park, et al., BRAF, PIK3CA, and HER2 oncogenic alterations according to KRAS mutation status in advanced colorectal cancers with distant metastasis, PLoS ONE 11 (3) (2016) e0151865.

[77] R. Stec, A. Semeniuk-Wojtas, R. Charkiewicz, L. Bodnar, J. Korniluk, M. Smoter, et al., Mutation of the PIK3CA gene as a prognostic factor in patients with colorectal cancer, Oncol. Lett. 10 (3) (2015 Sep.) 1423–1429.

[78] C. Mamot, G. Mild, J. Reuter, U. Laffer, U. Metzger, L. Terracciano, et al., Infrequent mutation of the tumour-suppressor gene Smad4 in early-stage colorectal cancer, Br. J. Cancer 88 (3) (2003 Feb. 10) 420–423.

[79] S. Roth, M. Johansson, A. Loukola, P. Peltomaki, H. Jarvinen, J. Mecklin, et al., Mutation analysis of SMAD2, SMAD3, and SMAD4 genes in hereditary non-polyposis colorectal cancer, Br. J. Cancer 88 (3) (2003 Feb. 10) 420–423.

[80] E. Sakai, M. Fukuyo, K. Matsusaka, K. Ohata, N. Doi, K. Takane, et al., TP53 mutation at early stage of colorectal cancer progression from two types of laterally spreading tumors, Cancer Sci. 107 (6) (2016 Jun.) 820–827.

[81] C. Wang, Y. Zhou, R. Ruan, M. Zheng, W. Han, L. Liao, et al., High expression of COUP-TF II cooperated with negative Smad4 expression predicts poor prognosis in patients with colorectal cancer, Int. J. Clin. Exp. Pathol. 8 (6) (2015) 7112–7121.

[82] L. Liu, J. Nie, L. Chen, G. Dong, X. Du, X. Wu, et al., The oncogenic role of microRNA-130a/301a/454 in human colorectal cancer via targeting Smad4 expression, PLoS ONE 8 (2) (2013) e55532.

[83] Y. Ma, F. Yan, L. Li, L. Liu, J. Sun, Deletion and down-regulation of SMAD4 gene in colorectal cancers in a Chinese population.

[84] S. Zhou, P. Buckhaults, L. Zawel, F. Bunz, G. Riggins, J.L. Dai, et al., Targeted deletion of Smad4 shows it is required for transforming growth factor ß and activin signaling in colorectal cancer cells.

[85] M. Zhang, F. Cui, S. Lu, H. Lu, T. Jiang, J. Chen, et al., Increased expression of prothymosin-a, independently or combined with TP53, correlates with poor prognosis in colorectal cancer, Int. J. Clin. Exp. Pathol. 7 (8) (2014) 4867–4876.

[86] L.S. Tillmans, R.A. Vierkant, A.H. Wang, N.J. Samadder, C.F. Lynch, K.E. Anderson, et al., Associations between cigarette smoking, hormone therapy

and folate intake with incident colorectal cancer by TP53 protein expression level in a population-based cohort of older women, Cancer Epidemiol. Biomark. Prev., Author manuscript, available in PMC 2015 Feb. 1.

[87] Y.W. Choi, Y.S. Song, H. Lee, K. Yi, Y. Kim, K.W. Suh, et al., MicroRNA expression signatures associated with BRAF-mutated versus KRAS-mutated colorectal cancers, Medicine (Baltimore) 95 (15) (2016 Apr.), e3321.

[88] H.O. Kim, B.G. Kim, S.J. Cha, Y.G. Park, T.J. Lee, Clinicopathologic significance of BRAF mutation and extracellular signal regulated kinase 1/2 expression in patients with a colorectal adenocarcinoma, Ann. Coloproctol. 31 (1) (2015 Feb.) 9–15.

[89] G.M. Aceto, F. Fantini, S.D. Iure, M.D. Nicola, G. Palka, R. Valanzano, et al., Correlation between mutations and mRNA expression of APC and MUTYH genes, J. Exp. Clin. Cancer Res. 34 (2015) 131.

[90] X. Fu, L. Shi, W. Zhang, X. Zhang, Y. Peng, X. Chen, et al., Expression of Indian hedgehog is negatively correlated with APC gene mutation in colorectal tumors, Int. J. Clin. Exp. Med. 7 (8) (2014) 2150–2155.

[91] S. Chen, Y. Wang, Y. Zhang, Y. Wan, Low expression of PKCa and high expression of KRAS predict poor prognosis in patients with colorectal cancer, Oncol. Lett. 12 (3) (2016 Sep.) 1655–1660.

[92] M. Lüchtenborg, M.P. Weijenberg, P.A. Wark, A.M. Saritas, G.M. Roemen, G.N. Muijen, et al., Mutations in APC, CTNNB1 and K-ras genes and expression of hMLH1 in sporadic colorectal carcinomas from the Netherlands cohort study, BMC Cancer 5 (2005) 160.

[93] C. Mao, X. Wu, Z. Yang, D.E. Threapleton, J. Yuan, et al., Concordant analysis of KRAS, BRAF, PIK3CA mutations, and PTEN expression between primary colorectal cancer and matched metastases, Sci. Rep. 5 (2015) 8065.

[94] F. Wen, S. He, C. Sun, T. Li, S. Wu, PIK3CA and PIK3CB expression and relationship with multidrug resistance in colorectal carcinoma, Int. J. Clin. Exp. Pathol. 7 (11) (2014) 8295–8303.

[95] J.K. Bar, A. Lis-Nawara, A. Lebioda, A. Jonkisz, M. Jelen, T. Dobosz, et al., KRAS mutation in relation to HER2 overexpression/amplification in colorectal cancer, Hered. Cancer. Clin. Pract. 10 (Suppl. 4) (2012) A25.

[96] L.S. Kristensen, T.E. Kjeldsen, H. Hager, L.L. Hansen, Competitive amplification of differentially melting amplicons (CADMA) improves KRAS hotspot mutation testing in colorectal cancer, BMC Cancer 12 (2012) 548.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

[97] P. Zhan, Y. Wang, S. Zhao, C. Liu, Y. Wang, M. Wen, et al., FBXW7 negatively regulates ENO1 expression and function in colorectal cancer, Labor Invest., Author manuscript, available in PMC 2016 Sep. 1.

[98] N. Li, F. Lorenzi, E. Kalakouti, M. Normatova, R. Babaei-Jadidi, I. Tomlinson, et al., FBXW7-mutated colorectal cancer cells exhibit aberrant expression of phosphorylated-p53 at serine-15, Oncotarget 6 (11) (2015 Apr. 20) 9240–9256.

[99] A.R. Halvorsen, L. Silwal-Pandit, L.A. Meza-Zepeda, D. Vodak, P. Vu, C. Sagerup, et al., TP53 mutation spectrum in smokers and never smoking lung cancer patients, Front. Genet. 7 (2016) 85.

[100] M.B. Schabath, E.A. Welsh, W.J. Fulp, L. Chen, J.K. Teer, Z.J. Thompson, et al., Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma, Oncogene, Author manuscript, available in PMC 2016 Jul. 8.

[101] L. Shan, T. Qiu, Y. Ling, L. Guo, B. Zheng, B. Wang, et al., Prevalence and clinicopathological characteristics of HER2 and BRAF mutation in Chinese patients with lung adenocarcinoma, PLoS ONE 10 (6) (2015) e0130447.

[102] W. Li, J. Qu, Z. Xu, Clinical features and mutation status of EGFR, KRAS, BRAF, EML4-ALK and ROS1 between surgical resection samples and non surgical resection samples in lung cancer, PLoS ONE 10 (6) (2015) e0130447.

[103] F. Imamura, T. Inoue, M. Kimura, K. Nishino, T. Kumagai, A long-term survivor of non-small-cell lung cancer harboring concomitant EGFR mutation and ALK translocation, Respir. Med. Case Rep. 19 (2016) 137–139.

[104] A.L. Riediger, S. Dietz, U. Schirmer, M. Meister, I. Heinzmann-Groth, M. Schneider, et al., Mutation analysis of circulating plasma DNA to determine response to EGFR tyrosine kinase inhibitor therapy of lung adenocarcinoma patients, Sci. Rep. 6 (2016) 33505.

[105] E.M. Beauchamp, B.A. Woods, A.M. Dulak, L. Tan, C. Xu, N.S. Gray, et al., Acquired resistance to dasatinib in lung cancer cell lines conferred by DDR2 gatekeeper mutation and NF1 loss, Mol. Cancer Ther., Author manuscript, available in PMC 2015 Feb. 1.

[106] S. Li, L. Li, Y. Zhu, C. Huang, Y. Qin, H. Liu, et al., Coexistence of EGFR with KRAS, or BRAF, or PIK3CA somatic mutations in lung cancer, Br. J. Cancer 110 (11) (2014 May 27) 2812–2820.

[107] D. Jelovac, J.A. Beaver, S. Balukrishna, H.Y. Wong, P.V. Toro, A. Cimino-Mathews, et al., A PIK3CA mutation detected in plasma from a patient

with synchronous primary breast and lung cancers, Hum. Pathol., Author manuscript, available in PMC 2015 Apr. 1.

[108] X. Tan, J. Carretero, Z. Chen, J. Zhang, Y. Wang, J. Chen, et al., Loss of p53 attenuates the contribution of IL-6 deletion on suppressed tumor progression and extended survival in Kras-driven murine lung cancer, PLoS ONE 8 (11) (2013) e80885.

[109] J. Mazumdar, M.M. Hickey, D.K. Pant, A.C. Durham, A. Sweet-Cordero, A. Vachani, et al., HIF-2a deletion promotes Kras-driven lung tumor development, Proc. Natl. Acad. Sci. USA 107 (32) (2010 Aug. 10) 14182–14187.

[110] R. Gill, S. Yang, D. Meerzaman, L. Mechanic, E. Bowman, H. Jeon, et al., Frequent homozygous deletion of the LKB1/STK11 gene in non-small cell lung cancer, Oncogene, Author manuscript, available in PMC 2013 Apr. 4.

[111] Y. Zhang, H. Fan, S. Fang, L. Wang, L. Chen, Y. Jin, et al., Mutations and expression of the NFE2L2/KEAP1/CUL3 pathway in Chinese patients with lung squamous cell carcinoma.

[112] J. Fourtounis, I. Wang, M. Mathieu, D. Claveau, T. Loo, A.L. Jackson, et al., Gene expression profiling following NRF2 and KEAP1 siRNA knockdown in human lung fibroblasts identifies CCL11/Eotaxin-1 as a novel NRF2 regulated gene, Respir. Res. 13 (1) (2012) 92.

[113] Y. Zhang, J. Sheng, S. Kang, W. Fang, Y. Yan, Z. Hu, et al., Patients with exon 19 deletion were associated with longer progression-free survival compared to those with L858R mutation after first-line EGFR-TKIs for advanced non-small cell lung cancer, PLoS ONE 9 (9) (2014) e107161.

[114] G.J. Weiss, A.K. Liman, J. Allen, P.Y. Cheung, R.N. Kukunoor, Squamous cell carcinoma of the lung with metastasis to the GI tract associated with EGFR exon 19 deletion, Case Rep. Med. 2013 (2013) 874836.

[115] J.S. Erdem, V. Skaug, P. Bakke, A. Gulsvik, A. Haugen, S. Zienolddiny, et al., Mutations in TP53 increase the risk of SOX2 copy number alterations and silencing of TP53 reduces SOX2 expression in non-small cell lung cancer, BMC Cancer 16 (2016) 28.

[116] H. Kodaz, E. Tastekin, B. Erdogan, I. Hacibekiroglu, H. Tozkir, H. Gürkan, et al., KRAS mutation in small cell lung carcinoma and extrapulmonary small cell cancer, Balkan Med. J. 33 (4) (2016 Jul.) 407–410.

[117] S.E. Lee, S. Lee, H. Park, S. Oh, H. Kim, K. Lee, et al., Detection of EGFR and KRAS mutation by pyrosequencing analysis in cytologic samples of non-small cell lung cancer, J. Korean Med. Sci. 31 (8) (2016 Aug.) 1224–1230.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

[118] Y. Takanashi, S. Tajima, T. Hayakawa, H. Neyatani, K. Funai, KRAS mutation-positive bronchial surface epithelium (BSE)-type lung adenocarcinoma with strong expression of TTF-1, Int. J. Clin. Exp. Pathol. 8 (11) (2015) 15338–15343.

[119] Y. Miyamae, K. Shimizu, Y. Mitani, T. Araki, Y. Kawai, M. Baba, et al., Mutation detection of epidermal growth factor receptor and KRAS genes using the smart amplification process version 2 from formalin-fixed, paraffin-embedded lung cancer tissue.

[120] T. Araki, K. Shimizu, K. Nakamura, T. Nakamura, Y. Mitani, K. Obayashi, et al., Usefulness of peptide nucleic acid (PNA)-clamp smart amplification process version 2 (SmartAmp2) for clinical diagnosis of KRAS Codon12 mutations in lung adenocarcinoma.

[121] E.C. Bruin, C. Cowell, P.H. Warne, M. Jiang, R.E. Saunders, M.A. Melnick, et al., Reduced NF1 expression confers resistance to EGFR inhibition in lung cancer, Cancer Discov., Author manuscript, available in PMC 2014 Nov. 1.

[122] D.J. Blake, A. Singh, P. Kombairaju, D. Malhotra, T.J. Mariani, R.M. Tuder, et al., Deletion of Keap1 in the lung attenuates acute cigarette smoke-induced oxidative stress and inflammation, Am. J. Respir. Cell. Mol. Biol. 42 (5) (2010 May) 524–536.

[123] B. Sandfeld-Paulsen, B.H. Folkersen, T.R. Rasmussen, P. Meldgaard, B.S. Sorensen, Gene expression of the EGF system—a prognostic model in non-small cell lung cancer patients without activating EGFR mutations, Transl. Oncol. 9 (4) (2016 Aug.) 306–312.

[124] H. Chang, J. Oh, X. Zhang, Y.J. Kim, J.H. Lee, C. Lee, et al., EGFR protein expression using a specific intracellular domain antibody and PTEN and clinical outcomes in squamous cell lung cancer patients with EGFR-tyrosine kinase inhibitor therapy, Onco. Targets Ther. 9 (2016) 5153–5162.

[125] S. Miyoshi, T. Kato, H. Katayama, R. Ito, Y. Mizuno, T. Okura, et al., A case of EGFR mutant lung adenocarcinoma that acquired resistance to EGFR-tyrosine kinase inhibitors with MET amplification and epithelial-to-mesenchymal transition, Onco. Targets Ther. 8 (2015) 783–787.

[126] M.A. Lewandowska, K. Czubak, K. Klonowska, W. Jozwicki, J. Kowalewski, P. Kozlowski, et al., The use of a two-tiered testing strategy for the simultaneous detection of small EGFR mutations and EGFR amplification in lung cancer, PLoS ONE 10 (2) (2015) e0117983.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

[127] M. Ji, H. Guan, C. Gao, B. Shi, P. Hou, Highly frequent promoter methylation and PIK3CA amplification in non-small cell lung cancer (NSCLC), BMC Cancer 11 (2011) 147.

[128] S.Y. Lee, S.K. Luk, C.P. Chuang, S.P. Yip, S.S. To, Y.M. Yung, et al., TP53 regulates human AlkB homologue 2 expression in glioma resistance to Photofrin-mediated photodynamic therapy, Br. J. Cancer 103 (3) (2010 Jul. 27) 362–369.

[129] C.G. Woo, D.S. Suh, J.Y. Kim, C.O. Sung, J. Choi, K Kim, et al., Peritoneal carcinosarcoma and ovarian papillary serous carcinoma are the same origin, Korean J. Pathol. 48 (6) (2014 Dec.) 449–453.

[130] K.Q. Cai, H. Wu, A.J. Klein-Szanto, X. Xu, Acquisition of a second mutation of the Tp53 alleles immediately precedes epithelial morphological transformation in ovarian tumorigenicity, Gynecol. Oncol., Author manuscript, available in PMC 2009 Sep. 8.

[131] K. Wang, D. Li, L. Sun, High levels of EGFR expression in tumor stroma are associated with aggressive clinical features in epithelial ovarian cancer, Onco. Targets Ther. 9 (2016) 377–386.

[132] R. Ranjbar, F. Nejatollahi, A.S. Ahmadi, H. Hafezi, A. Safaie, Expression of vascular endothelial growth factor (VEGF) and epidermal growth factor receptor (EGFR) in patients with serous ovarian carcinoma and their clinical significance, Iran J. Cancer Prev. 8 (4) (2015 Aug.) e3428.

[133] C. Yi, L. Zhang, L. Li, X. Liu, S. Ling, F. Zhang, et al., Establishment of an orthotopic transplantation tumor model in nude mice using a drug-resistant human ovarian cancer cell line with a high expression of c-Kit, Oncol. Lett. 8 (6) (2014 Dec.) 2611–2615.

[134] C. Yi, L. Li, K. Chen, S. Lin, X. Liu, Expression of c-Kit and PDGFRa in epithelial ovarian tumors and tumor stroma, Oncol. Lett. 3 (2) (2012 Feb.) 369–372.

[135] B.A. Davidson, J.M. Rubatt, D.L. Corcoran, D.K. Teoh, M.Q. Bernardini, L.A. Grace, et al., Differential angiogenic gene expression in TP53 wild-type and mutant ovarian cancer cell lines, Front. Oncol. 4 (2014) 163.

[136] M.Q. Bernardini, T. Baba, P.S. Lee, J.C. Barnett, G.P. Sfakianos, A.A. Secord, et al., Expression signatures of TP53 mutations in serous ovarian cancers, BMC Cancer 10 (2010) 237.

[137] Y. Wang, B. Dai, D. Ye, CHEK2 mutation and risk of prostate cancer, Int. J. Clin. Exp. Med. 8 (9) (2015) 15708–15715.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

[138] V. Hale, M. Weischer, J.Y. Park, CHEK2 *1100delC mutation and risk of prostate cancer, Prostate Cancer 2014 (2014) 294575.

[139] G. Boysen, C.E. Barbieri, D. Prandi, M. Blattner, S. Chae, A. Dahija, et al., SPOP mutation leads to genomic instability in prostate cancer, eLife 4 (2015) e09207.

[140] M. Saar, H. Zhao, R. Nolley, S.R. Young, I. Coleman, P.S. Nelson, et al., Spheroid culture of LuCaP 147 as an authentic preclinical model of prostate cancer subtype with SPOP mutation and hypermutator phenotype, Cancer Lett., Author, manuscript, available in PMC 2015 Sep. 1.

[141] W. Zhang, H. Jiao, X. Zhang, R. Zhao, F. Wang, W. He, et al., Correlation between the expression of DNMT1, and GSTP1 and APC, and the methylation status of GSTP1 and APC in association with their clinical significance in prostate cancer, Mol. Med. Rep. 12 (1) (2015 Jul.) 141–146.

[142] V. Nodouzi, M. Nowroozi, M. Hashemi, G. Javadi, R. Mahdian, Concurrent down-regulation of PTEN and NKX3.1 expression in Iranian patients with prostate cancer.

[143] S. Dhar, A. Kumar, A.M. Rimando, X. Zhang, A.S. Levenson, Resveratrol and pterostilbene epigenetically restore PTEN expression by targeting oncomiRs of the miR-17 family in prostate cancer, Oncotarget 6 (29) (2015 Sep. 29) 27214–27226.

[144] I.C. Gray, L.M. Stewart, S.M. Phillips, J.A. Hamilton, N.E. Gray, G.J. Watson, et al., Mutation and expression analysis of the putative prostate tumour-suppressor gene PTEN.

[145] M.M. Grabowska, A.D. Elliott, D.J. DeGraff, P.D. Anderson, G. Anumanthan, H. Yamashita, et al., NFI transcription factors interact with FOXA1 to regulate prostate-specific gene expression, Mol. Endocrinol. 28 (6) (2014 Jun.) 949–964.

[146] R.P. McMullin, A. Dobi, L.N. Mutton, A. Orosz, S. Maheshwari, C.S. Shashikant, et al., A FOXA1-binding enhancer regulates Hoxb13 expression in the prostate gland, Proc. Natl. Acad. Sci. USA 107 (1) (2010 Jan. 5) 98–103.

[147] C. Burdelski, D. Menan, M.C. Tsourlakis, M. Kluth, C. Hube-Magg, N. Melling, et al., The prognostic value of SUMO1/Sentrin specific peptidase 1 (SENP1) in prostate cancer is limited to ERG-fusion positive tumors lacking PTEN deletion, BMC Cancer 15 (2015) 538.

[148] D.A. Troyer, T. Jamaspishvili, W. Wei, Z. Feng, J. Good, S. Hawley, et al., A multicenter study shows PTEN deletion is strongly associated with seminal

vesicle involvement and extracapsular extension in localized prostate cancer, Prostate 75 (11) (2015 Aug.) 1206–1215.

[149] C. Cybulski, D. Wokolorczyk, T. Huzarski, T. Byrski, J. Gronwald, B. Górski, et al., A large germline deletion in the Chek2 kinase gene is associated with an increased risk of prostate cancer.

[150] T.H. Ho, P. Kapur, R.W. Joseph, D.J. Serie, J.E. Eckel-Passow, M. Parasramka, et al., Loss of PBRM1 and BAP1 expression is less common in non-clear cell renal cell carcinoma than in clear cell renal cell carcinoma, Urol. Oncol., Author manuscript, available in PMC 2016 Jan. 1.

[151] R.W. Joseph, P. Kapur, D.J. Serie, J.E. Eckel-Passow, M. Parasramka, T. Ho, et al., Loss of BAP1 protein expression is an independent marker of poor prognosis in patients with low risk clear cell renal cell carcinoma, Cancer, Author manuscript, available in PMC 2015 Apr. 1.

[152] J. Patard, P. Fergelot, P.I. Karakiewicz, T. Klatte, Q. Trinh, N. Rioux-Leclercq, et al., Low CAIX expression and absence of VHL gene mutation are associated with tumor aggressiveness and poor survival of clear cell renal cell carcinoma, J. Cancer 2 (2011) 515–526.

[153] E. Woodward, S. Clifford, D. Astuti, N. Affara, E. Maher, Familial clear cell renal cell carcinoma (FCRC).

[154] M.N. Farley, L.S. Schmidt, J.L. Mester, S. Pena-Llopis, A. Pavia-Jimenez, A. Christie, et al., A novel germline mutation in BAP1 predisposes to familial clear-cell renal cell carcinoma, Mol. Cancer Res., Author manuscript, available in PMC 2014 Oct. 28.

[155] V.A. Valera, B.A. Walter, W.M. Linehan, M.J. Merino, Regulatory effects of microRNA-92 (miR-92) on VHL gene expression and the hypoxic activation of miR-210 in clear cell renal cell carcinoma, J. Cancer 2 (2011) 515–526.

[156] B. Chowdhury, E.G. Porter, J.C. Stewart, C.R. Ferreira, M.J. Schipma, E.C. Dykhuizen, et al., PBRM1 regulates the expression of genes involved in metabolism and cell adhesion in renal clear cell carcinoma, PLoS ONE 11 (4) (2016) e0153718.

[157] W. Liu, Q. Fu, H. An, Y. Chang, W. Zhang, Y. Zhu, et al., Decreased expression of SETD2 predicts unfavorable prognosis in patients with nonmetastatic clear-cell renal cell carcinoma, Medicine (Baltimore) 94 (45) (2015 Nov.), e2004.

[158] M. López, C. Cervera-Acedo, P. Santibáñez, R. Salazar, J. Sola, E. Domínguez-Garrido, et al., A novel mutation in the CDH1 gene in a Spanish family with hereditary diffuse gastric cancer, SpringerPlus 5 (1) (2016) 1181.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

[159] R.S. Post, I.P. Vogelaar, F. Carneiro, P. Guilford, D. Huntsman, N. Hoogerbrugge, et al., Hereditary diffuse gastric cancer, J. Med. Genet. 52 (6) (2015 Jun.) 361–374.

[160] M. Ibarrola-Villava, M.J. Llorca-Cardeńosa, N. Tarazona, C. Mongort, T. Fleitas, J.A. Perez-Fidalgo, et al., Deregulation of ARID1A, CDH1, cMET and PIK3CA and target-related microRNA expression in gastric cancer, Oncotarget 6 (29) (2015 Sep. 29) 26935–26945.

[161] L. Wang, F. Zhang, P. Wu, X. Jiang, L. Zheng, Y Yu, et al., Disordered beta-catenin expression and E-cadherin/CDH1 promoter methylation in gastric carcinoma, World J. Gastroenterol. 12 (26) (2006 Jul. 14) 4228–4231.

[162] S. Sugimoto, H. Yamada, M. Takahashi, Y. Morohoshi, N. Yamaguchi, Y. Tsunoda, et al., Early-onset diffuse gastric cancer associated with a de novo large genomic deletion of CDH1 gene, Gastric Cancer 17 (4) (2014) 745–749.

[163] D.Q. Calcagno, V.M. Freitas, M.F. Leal, C.R. Souza, S. Demachki, R. Montenegro, et al., MYC, FBXW7 and TP53 copy number variation and expression in gastric cancer, BMC Gastroenterol. 13 (2013) 141.

[164] A.S. Khayat, A.C. Guimarães, D.Q. Calcagno, A.D. Seabra, E.M. Lima, M.F. Leal, et al., Interrelationship between TP53 gene deletion, protein expression and chromosome 17 aneusomy in gastric adenocarcinoma, BMC Gastroenterol. 9 (2009) 55.

[165] K. Harada, Y. Baba, H. Shigaki, T. Ishimoto, K. Miyake, K. Kosumi, et al., Prognostic and clinical impact of PIK3CA mutation in gastric cancer, BMC Cancer 16 (2016) 400.

[166] J.H. Park, M. Ryu, Y.S. Park, S.R. Park, Y. Na, B. Rhoo, et al., Successful control of heavily pretreated metastatic gastric cancer with the mTOR inhibitor everolimus (RAD001) in a patient with PIK3CA mutation and pS6 overexpression, BMC Cancer 15 (2015) 119.

[167] T. Ushiku, S. Ishikawa, M. Kakiuchi, A. Tanaka, H. Katoh, H. Aburatani, et al., RHOA mutation in diffuse-type gastric cancer, Gastric Cancer 19 (2016) 403–411.

[168] J. Shi, D. Yao, W. Liu, N. Wang, H. Lv, G. Zhang, et al., Highly frequent PIK3CA amplification is associated with poor prognosis in gastric cancer, BMC Cancer 12 (2012) 50.

[169] S. Matsusaka, T. Kobunai, N. Yamamoto, K. Chin, M. Ogura, G. Tanaka, et al., Prognostic impact of KRAS mutant type and MET amplification in metastatic and recurrent gastric cancer patients treated with first-line S-1 plus cisplatin chemotherapy.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

[170] H. Mita, M. Toyota, F. Aoki, H. Akashi, R. Maruyama, Y. Sasaki, et al., A novel method, digital genome scanning detects KRAS gene amplification in gastric cancers, BMC Cancer 9 (2009) 198.

[171] S. Jang, K. Kim, M. Oh, J. Lee, H.J. Lee, H.D. Cho, et al., Clinicopathological significance of elevated PIK3CA expression in gastric cancer, J. Gastric Cancer 16 (2) (2016 Jun.) 85–92.

[172] H. Ikeda, T. Kukitsu, W. Johmen, H. Nakamura, N. Yamauchi, K. Ishikawa, et al., Gastric invasive micropapillary carcinoma with intestinal phenotypes harboring a TP53 R175H mutation, Case Rep. Oncol. 7 (3) (2014 Sep.–Dec.) 611–620.

[173] E.M. Silva, M.I. Achatz, G. Martel-Planche, A.L. Montagnini, M. Olivier, P.A. Prolla, et al., TP53 mutation p.R337H in gastric cancer tissues of a 12-year-old male child – evidence for chimerism involving a common mutant founder haplotype, BMC Cancer 11 (2011) 449.

[174] L. Yang, S. Wei, R. Zhao, Y. Wu, H. Qiu, H. Xiong, et al., Loss of ARID1A expression predicts poor survival prognosis in gastric cancer, Sci. Rep. 6 (2016) 28919.

[175] K. Kim, H.Y. Jung, M. Oh, H. Cho, J. Lee, H.J. Lee, et al., Loss of ARID1A expression in gastric cancer, J. Gastric Cancer 15 (3) (2015 Sep.) 201–208.

[176] A.N. Bartley, P.A. Thompson, J.A. Buckmeier, C.Y. Kepler, C. Hsu, M.S. Snyder, et al., Expression of gastric pyloric mucin MUC6 in colorectal serrated polyps, Mod. Pathol., Author manuscript, available in PMC 2010 Aug. 1.

[177] M.K. McConechy, J. Ding, J. Senz, W. Yang, N. Melnyk, A.A. Tone, et al., Ovarian and endometrial endometrioid carcinomas have distinct CTNNB1 and PTEN mutation profiles, Mod. Pathol., Author manuscript, available in PMC 2014 Feb. 6.

[178] T.H. Kim, J. Wang, K.Y. Lee, H.L. Franco, R.R. Broaddus, J.P. Lydon, et al., The synergistic effect of conditional Pten loss and oncogenic K-ras mutation on endometrial cancer development occurs via decreased progesterone receptor action, J. Oncol. 2010 (2010) 139087.

[179] T. Takeda, K. Banno, R. Okawa, M. Yanokura, M. Iijima, H. Irie-Kunitomi, et al., ARID1A gene mutation in ovarian and endometrial cancers (Review), Oncol. Rep. 35 (2) (2016 Feb.) 607–613.

[180] Y. Liu, L. Patel, G.B. Mills, K.H. Lu, A.K. Sood, L. Ding, et al., Clinical significance of CTNNB1 mutation and Wnt pathway activation in

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

endometrioid endometrial carcinoma, J. Natl. Cancer Inst. 106 (9) (2014 Sep.) dju245.

[181] C. Krakstad, E. Birkeland, D. Seidel, K. Kusonmano, K. Petersen, S. Mjřs, et al., High-throughput mutation profiling of primary and metastatic endometrial cancers identifies KRAS, FGFR2 and PIK3CA to be frequently mutated, PLoS ONE 7 (12) (2012) e52795.

[182] E. Birkeland, E. Wik, S. Mjřs, E.A. Hoivik, J. Trovik, H.M. Werner, et al., KRAS gene amplification and overexpression but not mutation associates with aggressive and metastatic endometrial cancer, Br. J. Cancer 107 (12) (2012 Dec. 4) 1997–2004.

[183] K. Shoji, K. Oda, S. Nakagawa, S. Hosokawa, G. Nagae, Y. Uehara, et al., The oncogenic mutation in the pleckstrin homology domain of AKT1 in endometrial carcinomas, Br. J. Cancer 101 (1) (2009 Jul. 7) 145–148.

[184] L.J. Lee, E. Ratner, M. Uduman, K. Winter, M. Boeke, K.M. Greven, et al., The KRAS-variant and miRNA expression in RTOG endometrial cancer clinical trials 9708 and 9905, PLoS ONE 9 (4) (2014) e94167.

[185] G. Allo, M.Q. Bernardini, R. Wu, I. Shih, S. Kalloger, A. Pollett, et al., ARID1A loss correlates with mismatch repair deficiency and intact p53 expression in high-grade endometrial carcinomas, Mod. Pathol., Author manuscript, available in PMC 2015 Oct. 13.

[186] O. Fadare, I.L. Renshaw, S.X. Liang, Does the Loss of ARID1A (BAF-250a) expression in endometrial clear cell carcinomas have any clinicopathologic significance? A pilot assessment, J. Cancer 3 (2012) 129–136.

[187] C. Guo, W. Song, P. Sun, L. Jin, H. Dai, LncRNA-GAS5 induces PTEN expression through inhibiting miR-103 in endometrial cancer cells, J. Biomed. Sci. 22 (2015) 100.

[188] P. Fong, L. Meng, Effect of mTOR inhibitors in nude mice with endometrial carcinoma and variable PTEN expression status, Med. Sci. Monit. Basic Res. 20 (2014) 146–152.

[189] A. Joshi, C. Miller Jr., S.J. Baker, L.H. Ellenson, Activated mutant p110a causes endometrial carcinoma in the setting of biallelic Pten deletion.

[190] A. Joshi, L.H. Ellenson, Adenovirus mediated homozyogous endometrial epithelial Pten deletion results in aggressive endometrial carcinoma, Exp. Cell Res., Author manuscript, available in PMC 2012 Jul. 1.

[191] C. Li, B. Zhu, J. Chen, X. Huang, Feature genes predicting the FLT3/ITD mutation in acute myeloid leukemia, Mol. Med. Rep. 14 (1) (2016 Jul.) 89–94.

[192] M.J. Kim, S. Ahn, S. Jeong, J. Jang, J.H. Han, J.R. Choi, et al., Minor BCR-ABL1-positive acute myeloid leukemia associated with the NPM1 mutation and FLT3 internal tandem duplication, Ann. Lab. Med. 36 (3) (2016 May) 263–265.

[193] A. Busse, A. Letsch, C. Scheibenbogen, A. Nonnenmacher, S. Ochsenreither, E. Thiel, et al., Mutation or loss of Wilms' tumor gene 1 (WT1) are not major reasons for immune escape in patients with AML receiving WT1 peptide vaccination, J. Transl. Med. 8 (2010) 5.

[194] J.H. Lee, C. Park, S. Kim, M. Shin, A novel KIT INDEL mutation in acute myeloid leukemia with t(8, 21) (q22, q22), RUNX1-RUNX1T1, Ann. Lab. Med. 36 (4) (2016 Jul.) 371–374.

[195] R.A. Chougule, J.U. Kazi, L. Rönnstrand, FYN expression potentiates FLT3-ITD induced STAT5 signaling in acute myeloid leukemia, Oncotarget 7 (9) (2016 Mar. 1) 9964–9974.

[196] S. Sironi, M. Wagner, A. Kuett, H. Drolle, H. Polzer, K. Spiekermann, et al., Microenvironmental hypoxia regulates FLT3 expression and biology in AML, Sci. Rep. 5 (2015) 17550.

[197] L. Fu, H. Fu, L. Tian, K. Xu, K. Hu, J. Wang, et al., High expression of RUNX1 is associated with poorer outcomes in cytogenetically normal acute myeloid leukemia, Oncotarget 7 (13) (2016 Mar. 29) 15828–15839.

[198] V. Grossmann, U. Bacher, A. Kohlmann, K. Butschalowski, A. Roller, S. Jeromin, et al., Expression of CEBPA is reduced in RUNX1-mutated acute myeloid leukemia, Blood Cancer J. 2 (8) (2012 Aug.) e86.

[199] L. Xiang, J. Zhou, W. Gu, R. Wang, J. Wei, G. Qiu, et al., Changes in expression of WT1 during induced differentiation of the acute myeloid leukemia cell lines by treatment with 5-aza-2'-deoxycytidine and all-trans retinoic acid, Oncol. Lett. 11 (2) (2016 Feb.) 1521–1526.

[200] G. McCarty, D.M. Loeb, Hypoxia-sensitive epigenetic regulation of an antisense-oriented lncRNA controls WT1 expression in myeloid leukemia cells, PLoS ONE 10 (3) (2015) e0119837.

[201] S.Y. Jo, S.H. Park, I. Kim, J. Yi, H. Kim, C.L. Chang, et al., Correlation of NPM1 type A mutation burden with clinical status and outcomes in acute myeloid leukemia patients with mutated NPM1 type A, Ann. Lab. Med. 36 (5) (2016 Sep.) 399–404.

[202] H. Kao, D. Liang, M. Kuo, J. Wu, P. Dunn, P. Wang, et al., High frequency of additional gene mutations in acute myeloid leukemia with MLL partial tandem duplication, Oncotarget 6 (32) (2015 Oct. 20) 33217–33225.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

[203] R. Berenstein, I.W. Blau, N. Suckert, C. Baldus, A. Pezzutto, B. Dörken, et al., Quantitative detection of DNMT3A R882H mutation in acute myeloid leukemia, J. Exp. Clin. Cancer Res. 34 (1) (2015) 55.

[204] J. Ahn, H. Kim, Y. Kim, S. Jung, D. Yang, J. Lee, et al., Adverse prognostic effect of homozygous TET2 mutation on the relapse risk of acute myeloid leukemia in patients of normal karyotype.

[205] Y. Kuo, H. Hou, Y. Chen, L. Li, P. Chen, M. Tseng, et al., The N-terminal CEBPA mutant in acute myeloid leukemia impairs CXCR4 expression.

[206] E.A. Federzoni, M. Humbert, B.E. Torbett, G. Behre, M.F. Fey, M.P. Tschan, et al., CEBPA-dependent HK3 and KLF5 expression in primary AML and during AML differentiation, Sci. Rep. 4 (2014) 4261.

[207] W. Jang, J. Yoon, J. Park, G.D. Lee, J. Kim, A. Kwon, et al., Significance of KIT exon 17 mutation depends on mutant level rather than positivity in core-binding factor acute myeloid leukemia, Blood Cancer J. 6 (1) (2016 Jan.) e387.

[208] D. Monteferrario, S.M. Noordermeer, S. Bergevoet, G. Huls, J.H. Jansen, B.A. Reijden, et al., High DNA-methyltransferase 3B expression predicts poor outcome in acute myeloid leukemia, especially among patients with co-occurring NPM1 and FLT3 mutations, Blood Cancer J. 4 (8) (2014 Aug.) e233.

[209] S. Konoplev, P. Lin, C.C. Yin, E. Lin, G.M. González, H.M. Kantarjian, et al., CXCR4 expression, CXCR4 activation, and wild type NPM1 are independently associated with unfavorable prognosis in patients with acute myeloid leukemia, Clin. Lymphoma Myeloma Leuk., Author manuscript, available in PMC 2014 Dec. 1.

[210] K.L. McGraw, J. Nguyen, R.S. Komrokji, D. Sallman, N.H. Ali, E. Padron, et al., Immunohistochemical pattern of p53 is a measure of TP53 mutation burden and adverse clinical outcome in myelodysplastic syndromes and secondary acute myeloid leukemia.

[211] H. Hou, W. Chou, Y. Kuo, C. Liu, L. Lin, M. Tseng, et al., TP53 mutations in de novo acute myeloid leukemia patients, Blood Cancer J. 5 (7) (2015 Jul.) e331.

[212] T. Hirano, R. Yoshikawa, H. Harada, Y. Harada, A. Ishida, T. Yamazaki, et al., Long noncoding RNA, CCDC26, controls myeloid leukemia cell growth through regulation of KIT expression, Mol. Cancer 14 (2015) 90.

[213] X. Gao, J. Lin, L. Gao, A. Deng, X. Lu, Y. Li, et al., High expression of c-kit mRNA predicts unfavorable outcome in adult patients with t(8, 21) acute myeloid leukemia, PLoS ONE 10 (4) (2015) e0124241.

[214] J. Wang, Y. Zhao, J. Li, C. Guo, F. Chen, H. Su, et al., IDH1 mutation detection by droplet digital PCR in glioma, Oncotarget 6 (37) (2015 Nov. 24) 39651–39660.

[215] A. Zeng, Q. Hu, Y. Liu, Z. Wang, X. Cui, R. Li, et al., IDH1/2 mutation status combined with Ki-67 labeling index defines distinct prognostic groups in glioma, Oncotarget 6 (30) (2015 Oct. 6) 30232–30238.

[216] K. Kannan, A. Inagaki, J. Silber, D. Gorovets, J. Zhang, E.R. Kastenhuber, et al., Whole exome sequencing identifies ATRX mutation as a key molecular determinant in lower-grade glioma, Oncotarget 3 (10) (2012 Oct.) 1194–1203.

[217] P. Dell'Albani, M. Rodolico, R. Pellitteri, E. Tricarichi, S.A. Torrisi, S. D'Antoni, et al., Differential patterns of NOTCH1-4 receptor expression are markers of glioma cell differentiation, Neuro-Oncol. 16 (2) (2014 Jan.) 204–216.

[218] C.A. Fedrigo, I. Grivicich, D.P. Schunemann, I.M. Chemale, D.d. Santos, T. Jacovas, et al., Radioresistance of human glioma spheroids and expression of HSP70, p53 and EGFr, Radiat. Oncol. 6 (2011) 156.

[219] S. Beck, X. Jin, Y. Sohn, J. Kim, S. Kim, J. Yin, et al., Telomerase activity-independent function of TERT allows glioma cells to attain cancer stem cell characteristics by inducing EGFR expression, Radiat. Oncol. 6 (2011) 156.

[220] C.N. Paxton, L.R. Rowe, S.T. South, Observations of the genomic landscape beyond 1p19q deletions and EGFR amplification in glioma, Mol. Cytogenet. 8 (2015) 60.

[221] A. Mishra, J. Kong, D. Brat, CS-23 – defining the networks and intersection of EGFR amplification and NFKBIA deletion in human glioma, Mol. Cytogenet. 8 (2015) 60.

[222] M. Kocova, E. Kochova, E. Sukarova-Angelovska, Optic glioma and precocious puberty in a girl with neurofibromatosis type 1 carrying an R681X mutation of NF1, BMC Endocr. Disord. 15 (2015) 82.

[223] L.M. Merlo, J.W. Pepper, B.J. Reid, C.C. Maley, Cancer as an evolutionary and ecological process, Nat. Rev. Cancer 6 (2006) 924–935.

[224] S.L. Ostrow, R. Barshir, J. DeGregori, E. Yeger-Lotem, R. Hershberg, Cancer evolution is associated with pervasive positive selection on globally expressed genes, PLoS Genet. 10 (3) (2014) e1004239.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00277

[225] M.A. Owens, B.C. Horten, M.M. Da Silva, HER2 amplification ratios by fluorescence in situ hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues, Clin. Breast Cancer 5 (1) (2004) 63–69.

[226] W.E. Mesker, G.J. Liefers, J.M. Junggeburt, G.W. van Pelt, P. Alberici, P.J. Kuppen, et al., Presence of a high amount of stroma and downregulation of SMAD4 predict for worse survival for stage I-II colon cancer patients, Cell. Oncol. 31 (3) (2009) 169–178.

[227] B.K. Ahn, S.H. Jang, S.S. Paik, K.H. Lee, SMAD4 may help to identify a subset of colorectal cancer patients with early recurrence after curative therapy, Hepatogastroenterology 58 (112) (2011) 1933–1936.