Current Genomics, 2017, *18*, 287-297

RESEARCH ARTICLE

Identification of Sets of Cytoskeletal Related and Adhesion-related Coding Region Mutations in the TCGA Melanoma Dataset that Correlate with a Negative Outcome

John M. Yavorski[1], Rebecca J. Stoll[1], Mohammad D. Samy[1], James A. Mauro[1] and George Blanck[1,2,*]

[1]*Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, Florida, USA;*
[2]*Immunology Program, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA*

ARTICLE HISTORY

Received: August 30, 2016
Revised: November 03, 2016
Accepted: November 05, 2016

DOI:
10.2174/1389202918666170105093953

Abstract: ***Background:*** Relatively little cancer genome atlas data has been associated with clinically relevant stratifications of individual cancers.

Results: Mutations in two subsets of a cytoskeletal related and adhesion-related protein coding region set (CAPCRs) were determined to have strong associations with a negative outcome for melanoma, including a subset constituted by: DSCAM, FAT3, MUC17 and PCDHGC5 ($p < 0.0001$).

Conclusion: Roles for CAPCR mutations in cancer progression raise a question about the potential dominant negative impact of these mutations for multi-meric subcellular and extra-cellular protein structures.

Keywords: Cytoskeletal and cell-adhesion genomics, Cancer genome atlas, Extra-cellular matrix, Melanoma stratification, Melanoma metastasis.

1. INTRODUCTION

One of the most important goals for understanding somatic mutations in cancer is to readily be able to identify the mutations, particularly in circulating tumor cells, which are associated with metastasis and negative prognoses. Once a reasonable subset of mutations is available for such an association, relatively simple, cost-effective and very rapid tests could be developed to assay for these mutations. Availability of these tests would thereby provide for regular monitoring routines, more accurate prognoses, and more refined treatment courses. In addition, certain mutations may represent new targets for drug development to limit metastasis.

There is a very high mutation rate among cytoskeletal related and adhesion-related coding regions (CAPCRs) in a variety of cancers [1], and in particular, mutations in this set occur at a higher mutational density, as do oncoproteins, and in rare cases, tumor suppressor proteins [2]. The disproportionately higher level of CAPCR mutations, as well oncoprotein mutations, in cancer samples with relatively few mutations, indicates that the CAPCR and oncoprotein mutations are candidate driver mutations [2]. In addition, in both cases, a wealth of empirical literature has supported a role for these two classes of coding regions in oncogenesis.

The cancer genome atlas (TCGA) has provided a large amount of conveniently accessible primary tumor mutations, facilitating the above conclusions regarding driver candidates, among many related conclusions. However mutations from metastatic samples have been less available, most likely because in the case of most cancers, accessing metastatic lesions would require surgery that is not indicated or would be difficult in other ways. However, there have likely been other, paradigm-oriented hindrances to understanding "metastasis driver" mutations. For example, so-called metastasis suppressor proteins differ little if at all, from the standpoint of biochemistry or molecular mechanisms, from classical tumor suppressor proteins [3]. Metastasis suppressor proteins are significantly smaller than classical tumor suppressor proteins, which is consistent with their inactivation late in the tumorigenesis process, *i.e*, inactivation that is occurring largely due to the stochastic process of mutagenesis that "hits" small targets less frequently [1]. Other work favors signal amplification as a feature of poor cancer prognoses [4], but this work has not been well developed, with the on/off switch of cancer signaling pathways a more commonly researched paradigm.

TCGA has established numerous raw sequence files representing melanoma metastases, an efficient goal, given the accessible nature of many melanoma metastases. We have analyzed these files, along with patient outcome data, and determined that CAPCRs are more heavily mutated in metastatic melanoma samples and are very likely associated with a negative patient outcome.

METHODS

Analysis of the Mutations in the TCGA Processed Data

The basic processing approach is described in Fig. (**1**). From the TCGA data portal (https://tcga-data.nci.nih.gov/),

*Address correspondence to this author at the Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, USA; Tel: 813-974-9585; E-mail: gblanck@health.usf.edu

1875-5488/17 $58.00+.00

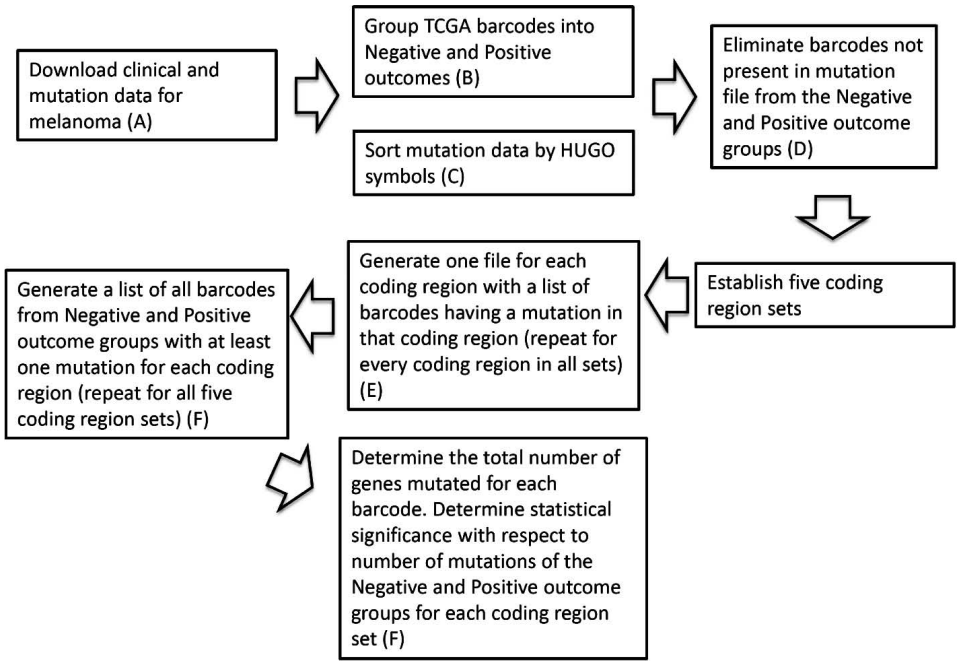


Fig. (1). Flow chart representing the processing steps for generating Tables **1-4**. The upper case letters in parentheses refer to Excel files in the Supporting Online Material (SOM) produced by the indicated step in the flow chart. However, (**A**) and (**C**), in the SOM, represent truncated files, due to size constraints; and (**E**) in the SOM represents an example set, *i.e.*, the oncoprotein set only, also due to size constraints. The members of the five coding region sets noted (established) above are in Table **3**. TCGA, the cancer genome atlas; HUGO, human genome organization.

clinical and somatic mutation data were downloaded for the SKCM dataset, with NIH approval (request #27073-3 for project #6300). The clinical drug file from the clinical data (Biotab) was used to sort barcodes into negative and positive outcome categories based on best treatment response. Clinical progressive disease and stable disease were assigned to the Negative outcome group, while partial response and complete response were assigned to the Positive outcome group. These assignments were made on the basis that a recorded response to treatment, either complete or partial, was a better outcome than either progressive disease or lack of any detectable response to treatment (*i.e.*, stable disease). These groupings do not rule out the possibility that long-term results, not available *via* the TCGA clinical data files, would not align would negative or positive outcome. The BI_IlluminaGA_DNASeq file was used from the somatic mutation data and was sorted by HUGO symbol. The tumor sample barcodes were truncated to contain only the following characters, TCGA-##-####. The clinical drug and mutation files were then compared based on the clinical barcodes and truncated tumor sample barcodes. Barcodes in the Negative and Positive outcome groups, respectively, that did not appear in the mutation file were eliminated. A list of 42 barcodes was generated; with 28 barcodes in the Negative outcome group and 14 barcodes in the Positive outcome group. The grouping of the barcodes into Negative outcome, from progressive disease and stabile disease, and into positive outcome, from partial and complete response, was necessitated by the low number of samples for the indicated subcategories, *e.g.*, progressive disease.

Five sets of coding regions, for analyses of mutation occurrence in the Negative and Positive outcome groups, were established: (i) Cytoskeletal related and adhesion-related proteins coding regions (CAPCRs), established as a group because of previous indications of cancer driver status and because coding regions in this group fall into the top 25 most mutated coding regions for five specific, solid-tissue cancers, particularly indicated in refs [1, 2, 5]. Also, three additional cytoskeletal related and adhesion-related proteins coding regions (DNAH7, GPR98, DSCAM) were added due to their occurrence in among the top 25 most mutated coding regions in SKCM (melanoma). (ii) "other CR (coding regions), commonly mutated", which represents some proteins involved in cytoskeletal related and adhesion-related functions but which are not part of the cancer-frequent CAPCR set, because these "commonly mutated" coding regions are not among the top most 25 mutated coding regions in the original five cancers in ref. [1] or in SKCM. The other CR, commonly mutated set also includes many other coding regions with no relationship to the cytoskeleton or to adhesion. (iii) A set of large coding regions equal in number and approximately equal to the overall coding region size of the CAPCRs. The members of this latter set (termed, "CAPCR-coding region size control") have no specific connection to the TCGA SKCM mutation dataset and were selected randomly from a Uniprot list of coding regions and their AA lengths. As with the other CR, commonly mutated set, the CAPCR-coding region size control set also represents proteins involved in cytoskeletal related and adhesion-related functions, but these proteins are not part of the cancer-frequent CAPCR set, *i.e.*, are not found heavily mutated in the five cancers of ref. [1] or in SKCM. Additional gene sets for this study are: (iv) oncoproteins; and (v) tumor suppressor proteins, lists established in refs. [1, 2].

As noted in the previous paragraph, the CAPCR set was previously established in ref. [1], based on the occurrence of

each coding region in the top 25 most commonly mutated coding regions among the five TCGA datasets studied in ref. [1], namely, BLCA, COAD, LUAD, GBM, STAD. And, for this project three additional CAPCRs were included in the set due to their occurrence among the top 25 most commonly mutated coding regions in SKCM. The HUGO symbols for the coding regions for all sets are in (Table **3**). See also SOM file, "SOM Table **3** source file".

A file for each HUGO symbol in each of the above indicated, five coding region sets was created, *i.e.* the information, from the comprehensive mutation file, for each of the HUGO symbols was copied to a newly generated "individual gene file". This newly generated individual gene file (examples of which are indicated as "SOM file Fig. (**1E**)" in the SOM) includes the truncated tumor sample barcode and mutation type (amino acid altering or silent). The above indicated list of 42 barcodes, representing negative outcomes (28 barcodes) and positive (14 barcodes) outcomes, was copied into each individual gene file and a COUNTIF function was used to determine the number of mutations in the individual coding region for any one barcode in the Negative and Positive outcome groups' barcode list.

A final results file was created for each coding region set. The TCGA barcodes from each individual gene file, that contained at least one mutation that represented the coding region for that gene file, were copied and transferred to a single column in the final results file. The Negative and Positive outcome groups' barcode lists indicated above were copied into (Column F of) the file used to generate the statistical analyses, termed "SOM Fig. (**1F**)". There is a "SOM Fig. (**1F**)" file for each coding region set (in the SOM). A COUNTIF function was used to determine the number of HUGO symbols, from the particular coding region set, that were mutated for each barcode in the Negative and Positive outcome groups. The TTEST function was used to determine the statistical significance of coding region mutations for the Negative and Positive outcome groups (for each coding region set).

Analysis of the Mutations in the Raw Exome Files

The basic processing approach is indicated in Fig. (**2**), and the detailed steps are provided in the SOM file, "SOM Fig. (**2**), detailed protocol". The original programming code and several output files are also provided in the SOM. Note, two SNP databases (All SNPs (142) and 1000G) were used to filter the originally identified variants in the raw sequence files representing the metastatic and primary melanoma samples, as indicated in (Fig. **2**). Because there is no SNP information in the 1000G SNP database for GRP98, this coding region is not included in any analyses using the mutations representing the metastatic melanoma, TCGA barcodes.

RESULTS

Association of CAPCR and Oncoprotein Coding Regions with Negative Outcome

The TCGA designations, (i) progressive and stable disease and (ii) complete and partial remission, were grouped as Negative and Positive outcome, respectively (Table **1**), and the total number of mutations for each barcode in each of the two categories were determined. The Negative outcome group had more mutations, but without a statistically significant difference from the positive outcome group (Table **2**).

The algorithm for establishing the assessment of the mutations in the coding region sets discussed below is shown as a flow chart in Fig. (**1**), with the legend to Fig. (**1**) indicating the availability of related documents in the supporting online material (SOM).

Keeping in mind the functional degeneracy of many mutations and proteins facilitating cancer [6-8], we established the following five sets of coding regions for analyses of

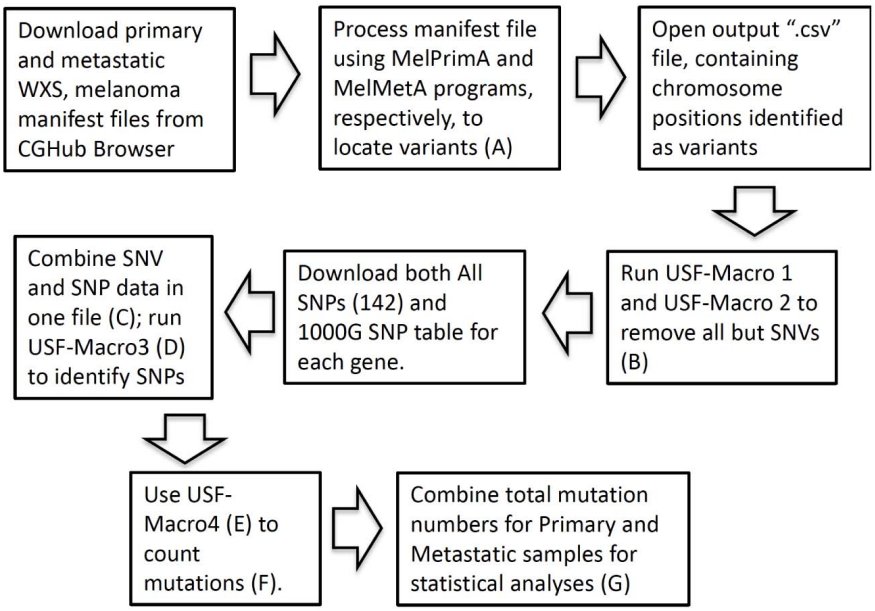


Fig. (2). Flow chart representing the processing steps for generating Table **6**. The upper case letters in parentheses refer to SOM files, either: (i) PDFs with program code; or (ii) Excel files with program or related output. WXS, whole exome sequence; SNV, single nucleotide variant; SNP, single nucleotide polymorphism.

Table 1. Establishment of clinical outcome categories from the cancer genome atlas (TCGA), skin cancer cutaneous melanoma (SKCM), clinical data (detailed in the SOM labeled "SOM file Fig1D, eliminate barcodes").

Clinical Outcome Categories	Best Treatment Response Categories from TCGA SKCM Clinical Drug File	Number of Barcodes in Each Group
Negative outcome	Stable disease	4
	Progressive disease	24
	SubTotal	28
Positive outcome	Partial response	1
	Complete response	13
	SubTotal	14
	Total number of barcodes	42

Table 2. Overview of mutations in the negative and positive, cancer genome atlas, skin cancer cutaneous melanoma, outcome groups (detailed in the supporting online material file labeled "SOM file Fig1D, eliminate barcodes").

Total number of mutations for the Negative outcome group	15088
Total number of mutations for the Positive outcome group	4479
Average number of mutations for the Negative outcome group	538.86
Average number of mutations for the Positive outcome group	319.93
p-value	0.1111 (Not significant)

mutation occurrence in the Negative and Positive outcome groups: (i) cytoskeletal related and adhesion-related proteins coding regions (CAPCRs); (ii) other CR (coding regions), commonly mutated; (iii) and a set of large coding regions equal to the coding region size of the cytoskeletal related and adhesion-related proteins but otherwise having no specific, functional connection to the TCGA SKCM dataset (CAPCR-coding region size control); (iv) oncoproteins; (v) tumor suppressor proteins. The HUGO symbols for these five sets are listed in Table **3**.

The tumor suppressor set includes both classical tumor suppressor protein coding regions, such as p16, and BRCA1, and coding regions representing what have been previously termed metastasis suppressor proteins. However, there is little biochemical, mechanistic distinction between classical tumor suppressor and metastasis suppressor proteins [3], and thus, for the purpose of applying the functional degeneracy principle, the tumor suppressor set used here includes both types of coding regions. (For further details on the establishment of the above, five coding regions sets, see **Methods**.)

CAPCRs have a long but contradictory history in cancer development [9-15]. In addition, CAPCRs are often very large and are thus subject to extensive mutation, presumably traceable to the highly stochastic aspect of mutagenesis, as indicated by a number of different types of analyses [1, 16]. Thus, the CAPCR-coding region size control was established in an attempt to appreciate the specificity of the CAPCR mutations in the TCGA SKCM data set.

As indicated in Table **4**, the CAPCR mutations (in the TCGA processed data) are significantly associated with negative outcome while the CAPCR-coding region size control is not. In the Table **4** analysis, the mutated, coding region-member of a set is counted one time, regardless of the actual number of mutations in that coding region. However, if the Negative and Positive outcome groups are distinguished based on differences in the total number of mutations in their respective CAPCR sets, the results are the same, *i.e*, CAPCR mutations are significantly associated with the Negative outcome group (data not shown).

We also determined that the oncoprotein coding region mutations, and the other CR, commonly mutated set were associated with negative outcome. However, no such association was detected for the tumor suppressor set.

The above associations hold with or without consideration of silent mutations in the coding regions, not surprising given the fact that silent mutations consistently represent about 25% of the total mutations in cancer cells, even when only a relatively small number of mutations has occurred [1]. To be certain this was the case for the current study, the ratios of silent mutations to total mutations was determined for the entire SKCM dataset, for negative and positive outcome groups, and for the CAPCR mutations for the preceding two groups, respectively, with the results indicated in Table **5**. The ratios of silent mutations, as predicted, were found to be consistent across all datasets accounting for about 30% of the total mutations. Therefore, the distinction of the functional impact of the mutation type is statistically unnecessary in distinguishing the Negative and Positive outcome groups. However, each gene set for the two groups was compared with the silent mutations removed, with the results indicated in the final row of Table **4**. As expected, given previous

Table 3. HUGO symbols for the indicated gene sets (additional information regarding categorization of individual coding regions in methods; more detailed gene information in the supporting online material file "SOM Table 3 source file".) CAPCR, cytoskeletal-related and adhesion-related protein coding region.

CAPCRs		Other-CR, Commonly Mutated		CAPCR-coding Region Size Control		Oncoprotein	Tumor Suppressor	
ANK2	MUC4	abParts	MYH1	ABCA13	KIAA1109	ACVR1	AKAP12	KISS1R
APC	NEB	ANK3	MYH2	AHNAK	LPA	ALK	AXIN1	KLF6
COL11A1	NEFH	ANKRD30A	MYO18B	AHNAK2	LRP1	ARAF	BMP2	LATS2
DNAH10	NF1	APOB	OBSCN	APOB	LRP1B	BRAF	BMPR1B	LIMD1
DNAH11	PCDH15	C15orf2	PAPPA2	BIRC6	MACF1	CTNNB1	BMPR2	MAP2K4
DNAH3	PCDHAC2	CACNA1E	PKHD1L1	DNAH17	MDN1	EGFR	BRCA1	MED23
DNAH5	PCDHGC5	CSMD1	PTPRT	DNAH2	MUC2	FGFR2	BRCA2	PBRM1
DNAH7	PCLO	CSMD2	RP1	DNAH9	MUC5AC	FLT3	BRMS1	PEBP1
DNAH8	PKHD1	CSMD3	RYR1	DNHD1	MUC5B	FRK	CASZ1	PPAPDC1B
DSCAM	PLEC	DNAH17	SCN10A	DYNC1H1	MYCBP2	HRAS	CDKN2A	PRDM2
DST	RELN	DNAH9	SCN11A	EPPK1	OBSCN	JAK2	CHD5	PTEN
FAT3	SPTA1	DSP	SPHKAP	FAT1	RNF213	KRAS	CHEK2	RB1
FAT4	SPTAN1	GRIN2A	THSD7B	FCGBP	RYR1	MTOR	CTCF	RECK
FBN2	SSPO	HYDIN	TNXB	FSIP2	RYR2	NRAS	DLC1	SMAD4
FLG	SYNE1	LRP1B	TRANK1	HERC1	RYR3	PRKACA	DOK2	SMAD7
GPR98	SYNE2	MAGEC1	UNC13C	HERC2	SACS	RAF1	FLCN	SMARCB1
MUC16	TTN	MGAM	USH2A	HMCN1	UBR4		FOXP3	SP100
MUC17	XIRP2	MXRA5	ZFHX4	HYDIN	USH2A		GPR68	TFPI2
							ING1	TMPRSS11A
							ING4	TXNIP
							INPP4B	VHL
							KISS1	WWOX

Table 4. Association of mutations in the indicated gene sets with TCGA SKCM negative outcome. (The source data for each coding region set (column) is present in the following SOM files: (i) "SOM file Fig1F, CAPCR, neg outcome stats", (ii) "SOM file Fig1F, Non-CAPCR, common mut., neg outcome stats", (iii) "SOM file Fig1F, Non-CAPCR, size control, neg outcome stats", (iv) "SOM file Fig1F, Oncoprotein, neg outcome stats", (v) "SOM file Fig1F, Tumor Suppressor, neg outcome stats". The source data for the "silent removed" p-values are present in SOM files: (i) "SOM file Fig1F, CAPCR, neg outcome stats, silents removed", (ii) "SOM file Fig1F, Non-CAPCR, common mut., neg outcome stats, silents removed", (iii) "SOM file Fig1F, Non-CAPCR, size control, neg outcome stats, silents removed", (iv) "SOM file Fig1F, Oncoprotein, neg outcome stats, silents removed", (v) "SOM file Fig1F, Tumor Suppressor, neg outcome stats, silents removed"). NS = not significant.

	Table 4. Coding Region Sets				
	CAPCRs	Other-CR, Commonly Mutated	CAPCR-coding Region Size Control	Oncoprotein	Tumor Suppressor
Total number of coding regions in set	36	36	36	16	44
Total number of amino acids in coding region set	191226	107584	186706	13058	34582

(Table 4) contd….

	Table 4. Coding Region Sets				
	CAPCRs	**Other-CR, Commonly Mutated**	**CAPCR-coding Region Size Control**	**Oncoprotein**	**Tumor Suppressor**
Average amino acid length per coding region	5313	2988	5186	816	786
Average number of coding regions per barcode with at least one mutation (Negative outcome group)	12.93	14.93	9.04	1.64	1.86
Average number of coding regions per barcode with at least one mutation (Positive outcome group)	8.86	9.29	6.14	1	1.07
p-Value for association of gene set mutations with negative outcome	**0.0466**	**0.0444**	**0.1175**(NS)	**0.0166**	**0.1127**(NS)
p-Value for association of gene set mutations with negative outcome (silent mutations removed)	0.0455	0.0474	0.2044 (NS)	0.0353	0.0642 (NS)

studies [1] and the data of Table **5**, the removal of silent mutations does not impact the statistical significance of the association, or lack thereof, of any gene set with the Negative and Positive outcome group. In short, there is a strong stochastic aspect to mutagenesis as represented by the TCGA datasets, as indicated by the very consistent ratio of silent to non-silent mutations over a wide range of TCGA datasets and mutation frequencies represented by the different samples within those datasets [1]. Thus, the silent mutations do not interfere with observing a statistically significant association of "mutation biomarkers" of the above indicated gene sets with Negative outcome.

Table 5. Ratios of silent mutations to total number of mutations (Source data present in "SOM Table 5, source file"). The data below are in very close agreement with Parry *et al.* [1] and support the conclusion that mutagenesis in cancer samples has a large stochastic component and the rate of silent mutations largely tracks the rate of amino acid replacement mutations.

	Ratio of Silent Mutations to Total Mutations
Entire SKCM dataset	0.323
Negative outcome group mutations	0.319
Positive outcome group mutations	0.304
CAPCR mutations negative outcome	0.274
CAPCR mutations positive outcome	0.297

More CAPCR Mutations in TCGA, Metastatic Melanoma Samples, Compared with Primary Melanoma Samples

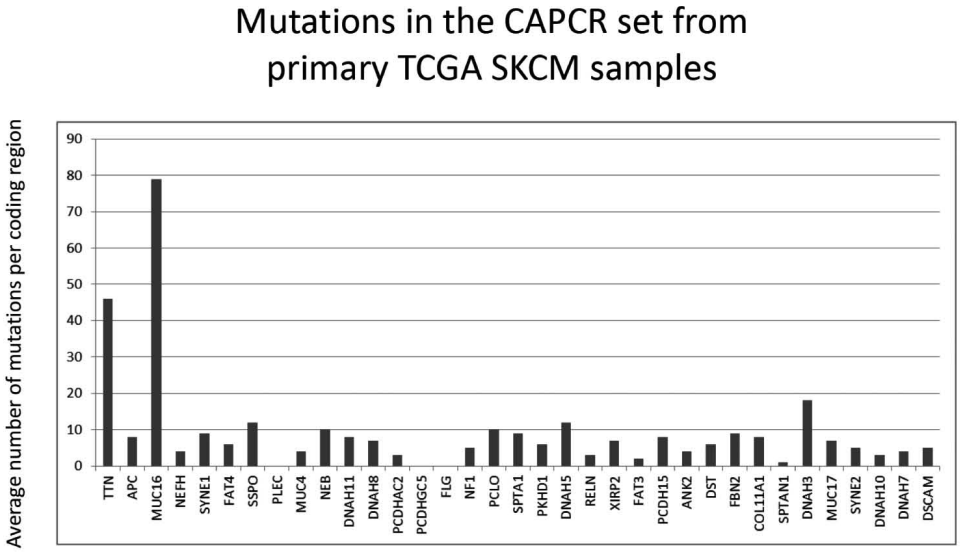
Because of the association of CAPCR mutations with the Negative outcome group, we determined whether there were more CAPCR mutations in metastatic melanoma samples *vs*. primary melanoma samples. To accomplish this goal, we downloaded raw exome sequence files representing both primary and metastatic melanoma samples from the CGHub browser (https://cghub.ucsc.edu/) as described in **Methods**; in Fig. (**2**); and in the SOM file, "SOM Fig. **2**, detailed protocol". The exome sequence files were processed as indicated in Fig. (**2**). The range of mutation occurrences is indicated in Table **6** (totaling all mutations for every coding region within the respective CAPCR sets), and a summary of the results for each member of the CAPCR set is indicated in Fig. (**3**). There are indeed about twice as many CAPCR mutations in the metastatic samples ($p < 0.0008$), consistent with more CAPCR mutations being associated with the Negative outcome group. It is likely that this result is due to an increase in the overall number of mutations in the metastatic samples and the highly stochastic nature of mutagenesis as revealed by previous studies, for example, ref. [1].

ECM-related Proteins are Significantly Mutated in the Negative Outcome Group Representing the Processed TCGA Mutation Data

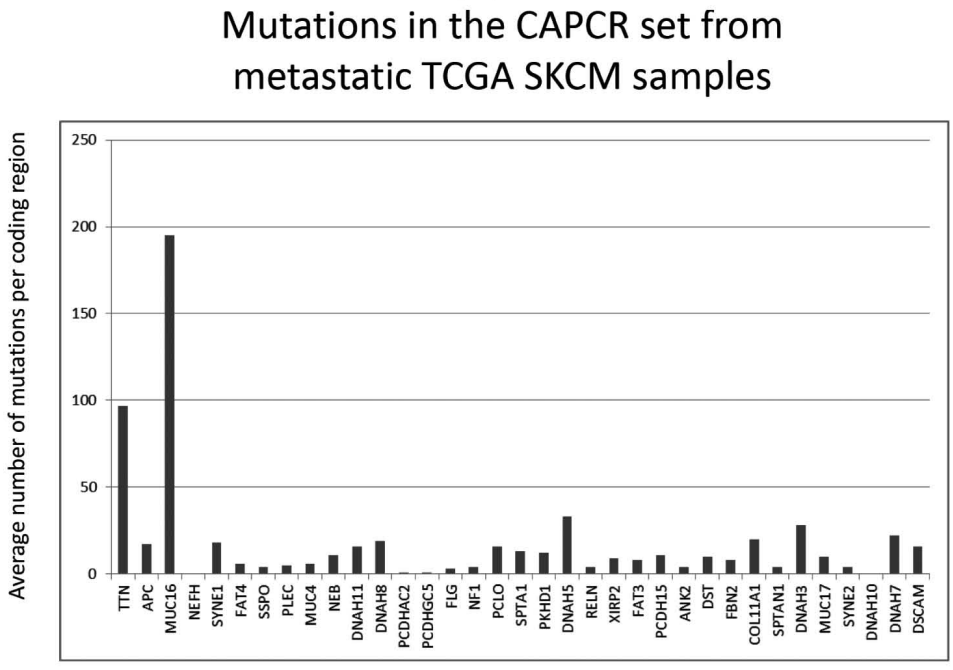
In the course of the above analyses of the metastatic samples, we noticed a range of ratios of mutation occurrence to number of amino acids in the coding region (Fig. **4**). We reasoned that, more mutations normalized to coding regions size could suggest cancer driver mutations, *i.e.*, could suggest a slight skewing away from the random impact of the likely, overall increase in mutations for the metastatic samples. Thus, we ranked the CAPCRs with respect to mutations per amino acid lengths (Fig. **4**) and observed the shift in rank order between the primary and metastatic samples. The CAPCRs with the ten largest rank shifts were found to have significantly higher mutation rates in the Negative outcome group defined and studied above Table **7**. However, coding regions with a decreased rank, from primary to metastatic, showed no statistically significant difference between the Positive and Negative outcome groups.

Table 6. Range of CAPCR mutations in primary and metastatic, TCGA SKCM raw exome sequence files. (Additional detail, including the TCGA barcodes for each sample, is present in the SOM file labeled "Figure 2G -Total mutations, prim and Met").

Sample number →	Primary 1	Primary 2	Primary 3	Primary 4	Primary 5	Primary 6	Primary 7	Primary 8	Primary 9	Primary 10
Number of mutations	24	29	15	82	29	53	16	33	14	33
Sample number →	Metastatic 1	Metastatic 2	Metastatic 3	Metastatic 4	Metastatic 5	Metastatic 6	Metastatic 7	Metastatic 8	Metastatic 9	Metastatic 10
Number of Mutations	67	43	97	154	79	40	21	28	72	34

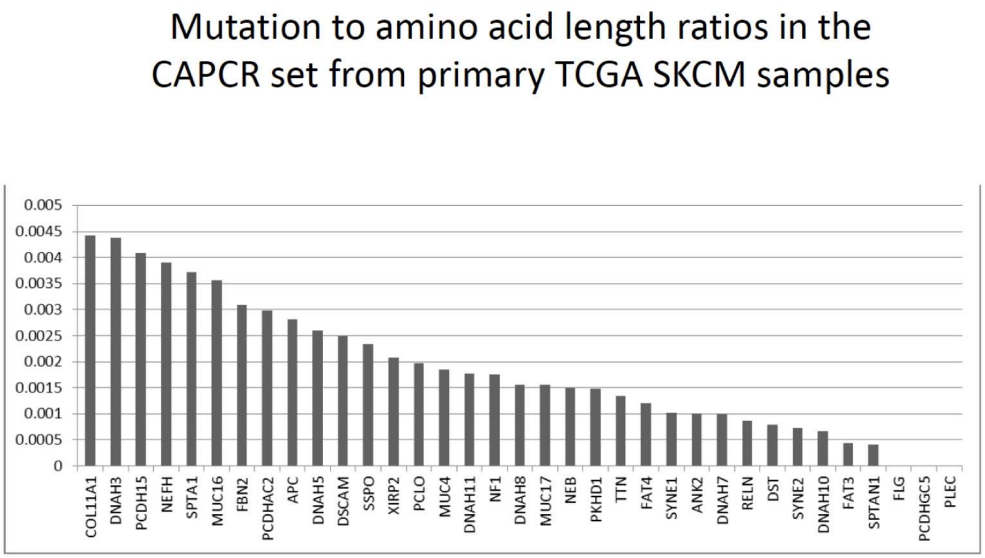


(3A)

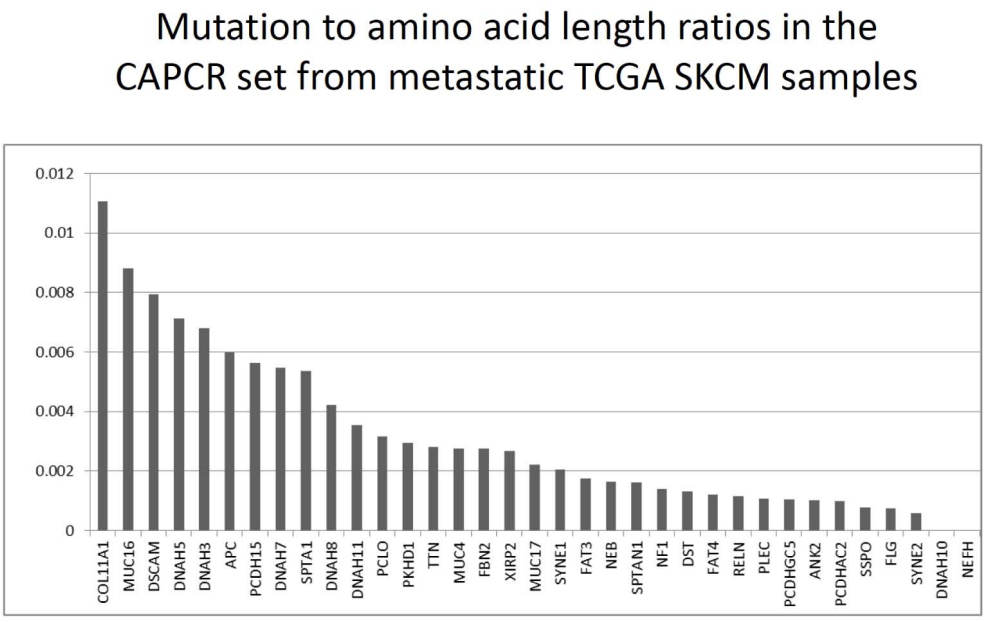


(3B)

Fig. (3). Histogram representing the mutation occurrences in the individual coding regions of the cytoskeletal-related protein coding region (CAPCR) set in primary (**A**) and metastatic (**B**) melanoma samples. The coding regions are ordered from largest to smallest.



(4A)



(4B)

Fig. (4). Histogram representing the range of ratios of mutations to amino acid lengths.

The two groups defined by the rank order shifts also represented different numbers of deleterious amino acids (AA), as determined by use of the PROVEAN tool (http:// provean.jcvi.org/index.php), as follows: the average number of deleterious AA replacements for the Negative outcome group was 4.0 per barcode *versus* 2.29 deleterious AA replacements for the Positive outcome group. However, for the eleven CAPCR coding regions with a decrease in rank (going from primary to metastatic; Table **8**), the average number of deleterious AA was 0.79 and 0.50 for the Negative and Positive outcome groups, respectively. The use of the PROVEAN tool and the calculation of the averages are available in "SOM file deleterious AA, source info".

A second subset of CAPCR's was tentatively identified by inspection of the frequencies of individual, mutated CAPCRs in the Negative and Positive outcome groups. This subset was subsequently defined such that a CAPCR was included in this subset if the CAPCR was mutated in a minimum of 50% of the Negative outcome samples and a maximum of 50% in the Positive outcome samples. This subset of four CAPCRs showed a very strong association with the Negative outcome group (Table **8**, $p < 1.0$ E-4). Furthermore, the average number of deleterious AA replacements, in this CAPCR subset (Table **8**), was 9.96 for the Negative outcome group and 5.93 for the Positive outcome group.

DISCUSSION

The above data demonstrate an opportunity to establish an association of a subset of cytoskeletal and adhesion-related coding regions with subcategories of melanoma. The subcategory clinical definitions in this case are minimal, due

Table 7. Association of a reduced CAPCR set, based on ranking shift between the mutations per amino acid ratios of the primary and metastatic samples, with the negative outcome group. (Additional detail is present in the SOM file labeled "SOM Table 7, source file").

	Increased Rank Shift in mut/AA Ratio From Primary to Metastatic Sample	Decreased Rank Shift in mut/AA Ratio From Primary to Metastatic Sample
	DNAH7	**NEFH**
	FAT3	**PCDHAC2**
	SPTAN1	**SSPO**
	DNAH8	**FBN2**
	DSCAM	**NF1**
	PKHD1	**XIRP2**
	PLEC	**SYNE2**
	TTN	**SPTA1**
	DNAH5	**PCDH15**
	PCDHGC5	**DNAH10**
		ANK2
p-value for association with negative outcome	0.0012	0.0602

to the TCGA limitations on clinical information, and the approach does not provide a cause and effect connection between cytoskeletal and adhesion-related coding region mutations and a specific course of disease for melanoma. However, the results do strongly suggest that the CAPCR sets and subsets indicated are biomarkers of melanoma subcategories. Furthermore, the CAPCR coding regions studied in this report represent a specificity of association in comparison to control sets, although it is possible, even likely, that other groupings of melanoma would reveal other types of associations, presumably in comparison to other types of control coding region sets. Also, the other-CR, commonly mutated set, effectively and unexpectedly representing a second large set of coding region mutations associated with negative outcome (Table **4**), includes additional cytoskeletal related and adhesion related coding regions that did not emerge as among the top 25 most mutated coding regions in SKCM. However, it is possible that these coding regions, if mutated, can have a driver effect in SKCM. This is not unexpected, in that certain oncoproteins are rarely mutated in certain cancer types, but can readily replace original oncoproteins as a result of designer drug resistance [8]. We would hypothesize a similar phenomenon here, whereby certain CAPCRs are commonly mutated in SKCM for unknown reasons, perhaps related to exposure of specific segments of DNA to mutagens. And, that less common mutations in the coding regions of other cytoskeletal and adhesion-related proteins could be selected during tumor evolution or during drug treatment and essentially substitute for the cancer driving impact of more commonly mutated CAPCRs.

Further specificity will come from rotating CAPCR set members in and out of the above CAPCR group when studying even more detailed clinical parameters not yet available through TCGA. In particular, Guan *et al.* identified specific, single driver mutations in melanoma, as opposed to the above "either/or" (or subset) approach, and identified one CAPCR not included in any of the sets defined above, as a driver, namely CDH9, a cadherin involved in cell-cell adhesion [17]. However, given the nature of polymer formation and the functional importance of protein-protein interaction in the proper functioning of the cytoskeleton and in cell adhesion, we would expect that the opportunities to identify single mutations as drivers among CAPCR proteins to require an unusually high sample size and to be rarely detectable. And as a consequence, we would expect such mutations to appear as drivers in very, very few cases. However, cooperative protein-protein function does imply successful opportunities to identify driver groups, particularly among CAPCRs [2]. Indeed, it is important to note that CAPCRs are not the only category where grouping can be of use. In particular, both oncoproteins and tumor suppressor proteins can be represented by alternatives within the same cancer datasets [2-4], as has been recently confirmed by the very common selection for alternative oncogenic signaling pathways upon the application of designer drugs [8], as noted above.

Cytoskeletal and adhesion-related proteins have long been a topic of cancer research, with early reports indicating that a disorganized cytoskeleton is associated with tumorigenesis [14, 18]. However, later reports have led to contradictory conclusions, particularly with an emphasis on the idea that cytoskeletal organization is a crucial aspect of cell migration and metastasis [13]. Understanding the role of cytoskeletal disorganization in prognoses will likely require a better understanding of distinct sub-functions of the cell cytoskeleton. For example, metastasis has been associated with the cell's ability to form, Rho-regulated invadopodia, to migrate in experimental settings [11], and to form the com-

Table 8. Association of reduced CAPCR set from inspection of negative outcome group. CAPCR coding regions were determined using selection criteria of any coding region with greater than 50% of negative outcome barcodes and less than 50% of positive outcome barcodes having a mutation in that coding region. (Source data present in "SOM Table 8, source file")

Description	CAPCR Coding Regions with Greater Than 50% of Negative Outcome Barcodes and Less Than 50% of Positive Outcome Barcodes with a Mutation
	DSCAM
	FAT3
	MUC17
	PCDHGC5
p-value for association with negative outcome	8.65E-05

ponents of the cytoskeleton required for migration. On the other hand, the complete lack of dystrophin [19, 20] can lead to a tumor phenotype in mice.

There is virtually no understanding of the impact of CAPCR mutations, and ensuing mutant CAPCR proteins, on patient cancer prognosis, not surprising, given the contradictions discussed above and the challenges of studying mutant forms of very large proteins in many experimental settings. Nevertheless, CAPCRs are among the most commonly mutated coding regions in cancer, likely owing to their large sizes [1]. And, protein products of CAPCRs in many cases interact, making it possible that any number of mutant proteins could have a dominant negative effect, akin to the dominant negative effect seen in many other, better studied polymeric or multimeric protein complex-related disorders, *e.g.*, the dominant negative effect seen in osteogenesis imperfecta, due to collagen protein mutations. Thus, in addition to signaling pathway connections to the cytoskeleton, and other broad-based analysis approaches, such as microscopic analyses of the staining of cytoskeletal components, understanding mutated protein components of the cytoskeleton is likely to shed new light on the connection between the cytoskeletal function and cancer.

Due to a lack of clinical information overlapping molecular information, TCGA has not facilitated many connections between the two sets of parameters, particularly within a given cancer type. However, in the case of the above study and results, there is an indication that increased numbers of mutations of CAPCRs are inconsistent with a positive outcome in melanoma. Although as presented, and given the available quantity of information, the results are consistent, the results are based on modest clinical information available *via* TCGA databases and would need to be further investigated with a more controlled clinical setting. For example, it is possible that patients representing the Negative outcome group were treated more aggressively from the beginning of cancer detection with alkylating agents, likely to mutate more, and in particular mutate large coding regions more. In this case, the association between mutation of large coding regions and a negative outcome could be a by-product of the treatment, rather than a biomarker or functional aspect of cancer progression. However, it remains possible that disruption of the cytoskeleton or cell adhesion *via* expression of mutant cytoskeletal or adhesion related proteins could have effects that lead to more aggressive cancers.

While the regulated, or abnormally regulated disassembly and re-formation of the cytoskeleton or ECM may provide explanations for migration or other aspects of cell function, the disruption of the cytoskeleton, particularly *via* mutant proteins, raises the question of what cell-shape changes could be physically associated with cancer development? Cells that lack the capacity to form an extensive cytoskeleton are likely to be more spheroid in shape, and due to physical attributes alone, less likely to remain attached to surrounding tissue. Also, it remains intriguing that almost all cells selected for drug-resistance in culture are spherical, a result that is often connected to epithelial-mesenchymal transition. However, the very common CAPCR mutations raise the question of whether cell shape changes lead to reduced surface to volume ratios, increased diffusion co-efficients, and therefore less sensitivity to drugs?

In the above results, we observed a trend towards a higher number of mutations, normalized to coding region size, in the metastatic samples, for many CAPCRs that function in adhesion, *i.e.*, as extra-cellular matrix (ECM) proteins, *e.g*, DSCAM and PCDHGC5. Mutation of the ECM subclass of the CAPCRs has a statistically significant association with the Negative outcome group (Table **7**). A further reduced subset of mutated CAPCRs (DSCAM, FAT3, MUC17 and PCDHGC5; Table **8**) also had a statistically significant association with the Negative outcome group. This subset also represented ECM components, and overlapped with the subset defined by increased mutation rates in the metastatic samples, with the exception of MUC17 being included in the smaller subset.

Interestingly, the deleterious AA rates, as determined by use of the PROVEAN tool, track the mutation rates almost exactly. As would be expected for a functional impact, there are more deleterious AA in the ECM coding regions associated with the Negative outcome group, although there is no apparent selection for deleterious AA independently of the mutation rates (unpublished observations), *i.e.*, there is no apparent increase in the number of deleterious AA per mutation. However, there is an apparent selection for a higher mutation rate in the indicated CAPCR sets (Tables **4**, **7**, **8**). This mutation rate is apparently largely, random, leading to deleterious AA replacements at the same rate, whether there is a high rate of mutagenesis or a low rate of mutagenesis. This randomness of "deleterious AA related mutagenesis" is consistent with previous work indicating almost no differ-

ences between the ratio of silent mutations to AA replacement mutations in cancer samples with high rates of mutations *versus* low rates of mutation [1], and is consistent with verification of that analysis in this report (Table **5**). Presumably, with a high enough rate of mutagenesis, there is a high enough accumulation of deleterious AA to have a functional, in this case, negative impact.

The increased number of deleterious AA among the ECM subset of the CAPCRs is consistent with the possibility that tissue detachment or tissue instability facilitating metastasis is at the heart of the association of the mutated CAPCRs with a negative outcome. The result also raises questions about the overlap of mutations in ECM proteins and other ECM signatures reported as associated with metastasis, such as the higher level of expression of particular ECM components [21]. For example, is it possible that metastasis selects for the higher level expression of mutant ECM proteins because the increased level of expression of the mutant form would further enhance the disruption of the ECM caused by a structurally defective subunit?

CONCLUSION

Many CAPCRs are mutated in melanomas that have high numbers of overall mutations, consistent with the significant stochastic effect of mutagenesis and the vulnerability of large coding regions. The CAPCRs can be subdivided into groups of CAPCRs that have a strong association with a negative outcome and relatively higher rates of mutations in metastatic melanoma samples.

LIST OF ABBREVIATIONS

The gene names for the HUGO symbols used in this report are in the supporting online material.

AA	=	Amino Acid
CAPCR	=	Cytoskeletal Protein-related Coding Regions
ECM	=	Extra-cellular Matrix
SKCM	=	Skin, Cutaneous Melanoma
TCGA	=	The Cancer Genome Atlas

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Authors acknowledge support of the Anna Valentine Program and the taxpayers of the State of Florida.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

[1] Parry, M.L.; Ramsamooj, M.; Blanck, G.; Big genes are big mutagen targets: A connection to cancerous, spherical cells? *Cancer Lett.,* **2015**, *356*, 479-482.

[2] Fawcett, T.J.; Parry, M.L.; Blanck, G. A Novel Approach to Evaluating Cancer Driver Gene Mutation Densities: Cytoskeleton-related Gene Candidates. *Cancer Genom. Proteom.,* **2015**, *12*, 283-290.

[3] Long, K.; Abuelenen, T.; Pava, L.; Bastille, M.; Blanck, G. Size matters: Sequential mutations in tumorigenesis may reflect the stochastic effect of mutagen target sizes. *Genes Cancer,* **2011**, *2*, 927-931.

[4] Ford, S.A.; Blanck, G. Signal persistence and amplification in cancer development and possible, related opportunities for novel therapies. *Biochim. Biophys. Acta,* **2014**, *1855*, 18-23.

[5] Parry, M.L.; Blanck, G. Flat cells come full sphere: Are mutant cytoskeletal-related proteins oncoprotein-monsters or useful immunogens? *Hum. Vaccin. Immunother.,* **2016**, *12*, 120-123.

[6] Bucheit, A.D.; Davies, M.A. Emerging insights into resistance to BRAF inhibitors in melanoma. *Biochem. Pharmacol.,* **2014**, *87*, 381-389.

[7] Lito, P.; Rosen, N.; Solit, D.B. Tumor adaptation and resistance to RAF inhibitors. *Nat. Med.,* **2013**, *19*, 1401-1409.

[8] Frangione, M.L.; Lockhart, J.H.; Morton, D.T.; Pava, L.M.; Blanck, G. Anticipating designer drug-resistant cancer cells. *Drug Discov. Today*, **2015**, *20*(7), 790-793

[9] Guo, C.; Liu, S.; Wang, J.; Sun, M.Z.; Greenaway, F.T. ACTB in cancer. *Clin. Chim. Acta,* **2013**, *417*, 39-44.

[10] Xu, W.; Mezencev, R.; Kim, B.; Wang, L.; McDonald, J.; Sulchek, T. Cell stiffness is a biomarker of the metastatic potential of ovarian cancer cells. *PLoS One,* **2012**, *7*, e46609.

[11] Narumiya, S.; Tanji, M.; Ishizaki, T.; Rho signaling, ROCK and mDia1, in transformation, metastasis and invasion. *Cancer Metastasis Rev.,* **2009**, *28*, 65-76.

[12] Pokorna, E.; Jordan, P.W.; O'Neill, C.H.; Zicha, D.; Gilbert, C.S.; Vesely, P. Actin cytoskeleton and motility in rat sarcoma cell populations with different metastatic potential. *Cell Motil Cytoskeleton*, **1994**, *28*, 25-33.

[13] Zachary, J.M.; Cleveland, G.; Kwock, L.; Lawrence, T.; Weissman, R.M.; Nabell, L.; Fried, F.A.; Staab, E.V.; Risinger, M.A.; Lin, S. Actin filament organization of the Dunning R3327 rat prostatic adenocarcinoma system: correlation with metastatic potential. *Cancer Res.,* **1986**, *46*, 926-932.

[14] Verderame, M.; Alcorta, D.; Egnor, M.; Smith, K.; Pollack, R.; Cytoskeletal F-actin patterns quantitated with fluorescein isothiocyanate-phalloidin in normal and transformed cells. *Proc. Natl. Acad. Sci. U.S.A.,* **1980**, *77*, 6624-6628.

[15] Kopelovich, L.; Conlon, S.; Pollack, R. Defective organization of actin in cultured skin fibroblasts from patients with inherited adenocarcinoma. *Proc. Natl. Acad. Sci. U.S.A.,* **1977**, *74*, 3019-3022.

[16] Narsing, S.; Jelsovsky, Z.; Mbah, A.; Blanck, G. Genes that contribute to cancer fusion genes are large and evolutionarily conserved. *Cancer Genet. Cytogenet.,* **2009**, *191*, 78-84.

[17] Guan, J.; Gupta, R.; Filipp, F.V. Cancer systems biology of TCGA SKCM: efficient detection of genomic drivers in melanoma. *Sci. Rep.,* **2015**, *5*, 7857.

[18] Brinkley, B.R.; Beall, P.T.; Wible, L.J.; Mace, M.L.; Turner, D.S.; Cailleau, R.M. Variations in cell form and cytoskeleton in human breast carcinoma cells *in vitro*. *Cancer Res.,* **1980**, *40*, 3118-3129.

[19] Wang, Y.; Marino-Enriquez, A.; Bennett, R.R.; Zhu, M.; Shen, Y.; Eilers, G.; Lee, J.C.; Henze, J.; Fletcher, B.S.; Gu, Z.; Fox, E.A.; Antonescu, C.R.; Fletcher, C.D.; Guo, X.; Raut, C.P.; Demetri, G.D.; van de Rijn, M.; Ordog, T.; Kunkel, L.M.; Fletcher, J.A. Dystrophin is a tumor suppressor in human cancers with myogenic programs. *Nat. Genet.,* **2014**, *46*, 601-606.

[20] Fernandez, K.; Serinagaoglu, Y.; Hammond, S.; Martin, L.T.; Martin, P.T. Mice lacking dystrophin or alpha sarcoglycan spontaneously develop embryonal rhabdomyosarcoma with cancer-associated p53 mutations and alternatively spliced or mutant Mdm2 transcripts. *Am. J. Pathol.,* **2010**, *176*, 416-434.

[21] Naba, A.; Clauser, K.R.; Whittaker, C.A.; Carr, S.A.; Tanabe, K.K.; Hynes, R.O. Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver. *BMC Cancer,* **2014**, *14*, 518.