

RESEARCH ARTICLE

Somatic Variations in Cervical Cancers in Indian Patients

Poulami Das^{1#a}, Akanksha Bansal¹, Sudha Narayan Rao², Kedar Deodhar³, Umesh Mahantshetty⁴, Shyam K. Shrivastava⁴, Karthikeyan Sivaraman⁵, Rita Mulherkar^{1#b,*}

1 Mulherkar Lab, ACTREC, Tata Memorial Centre, Kharghar, Navi Mumbai, 410210, India, **2** Genotypic Technology Pvt. Ltd., #259, Apoorva 4th Cross, 80 Ft Road, RMV II Stage, Bengaluru, 560094, India, **3** Department of Pathology, Tata Memorial Hospital, Parel, Mumbai, 400012, India, **4** Department of Radiation Oncology, Tata Memorial Hospital, Parel, Mumbai, 400012, India, **5** Dhitiomics Technologies Pvt. Ltd. #2/10, 5th Floor, 80 Feet Road, RMV 2nd Stage, Poojari Layout, Bengaluru, Karnataka, 560094, India

^{#a} Current address: Department of Molecular Biosciences, Centre for Infectious Disease, University of Texas, Austin, TX, United States of America

^{#b} Current address: A 103, Park Dew, Sector 20, Kharghar, Navi Mumbai, 410210, India

* rmulherkar@gmail.com



OPEN ACCESS

Citation: Das P, Bansal A, Rao SN, Deodhar K, Mahantshetty U, Shrivastava SK, et al. (2016) Somatic Variations in Cervical Cancers in Indian Patients. PLoS ONE 11(11): e0165878. doi:10.1371/journal.pone.0165878

Editor: Obul Reddy Bandapalli, German Cancer Research Center (DKFZ), GERMANY

Received: May 4, 2016

Accepted: October 19, 2016

Published: November 9, 2016

Copyright: © 2016 Das et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The work was funded by Department of Science and Technology (DST No: SR/SO/HS-22/2009), a grant sanctioned to RM. Genotypic Technology Pvt. Ltd. and Dhitiomics Technologies Pvt. Ltd. provided support in the form of salaries/consultation fees for authors SNR and KS respectively, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. SNR provided services which included

Abstract

There are very few reports that describe the mutational landscape of cervical cancer, one of the leading cancers in Indian women. The aim of the present study was to investigate the somatic mutations that occur in cervical cancer. Whole exome sequencing of 10 treatment naïve tumour biopsies with matched blood samples, from a cohort of Indian patients with locally advanced disease, was performed. The data revealed missense mutations across 1282 genes, out of 1831 genes harbouring somatic mutations. These missense mutations (nonsynonymous + stop-gained) when compared with pre-existing mutations in the COSMIC database showed that 272 mutations in 250 genes were already reported although from cancers other than cervical cancer. More than 1000 novel somatic variations were obtained in matched tumour samples. Pathways / genes that are frequently mutated in various other cancers were found to be mutated in cervical cancers. A significant enrichment of somatic mutations in the MAPK pathway was observed, some of which could be potentially targetable. This is the first report of whole exome sequencing of well annotated cervical cancer samples from Indian women and helps identify trends in mutation profiles that are found in an Indian cohort of cervical cancer.

Introduction

With 122800 new cases and 67400 deaths annually, cervical cancer significantly contributes to the cancer burden in India [1]. Although infection by HPV has been established as a major etiological factor for the genesis of the disease, only a small proportion of HPV infected women develop cervical cancer; [2]. This suggests possible involvement of genetic aberrations in development of cervical cancer. Over the years, studies have described direct role of various genetic alterations in disease development. Somatic mutations in PIK3CA, PTEN, TP53, STK11 and

exome sequencing using Illumina platform and providing the raw data. KS, an independent consultant, was specifically brought in by Dhitiomics Technologies Pvt. Ltd., for handling the data analysis part. The specific roles of these authors are articulated in the 'author contributions' section.

Competing Interests: Sudha Narayan Rao is employed by Genotypic Technology Pvt. Ltd. Karthikeyan Sivaraman is a consultant for Dhitiomics Technologies Pvt. Ltd. There are no patents, products in development or marketed products to declare. This does not alter our adherence to all the PLOS ONE policies on sharing data and materials.

KRAS as well as copy number alterations have been reported [2–5]. In addition, recurrent mutations in EP300, MAPKI, FBXW7, HLAB, NFE2L2 have been reported from a large study from Meyerson's group [6].

Integration of HPV in the human genome could also contribute to genetic aberrations and alter gene expression, leading to cervical carcinogenesis. We and others have reported that the site of HPV integration is not totally random [7, 8]. Meyerson's group has reported that gene expression levels at sites of HPV integration were higher in tumours with integration than level of expression in other tumours without integration at that site [6]. HPV integration may result in loss of function of tumour suppressors or critical DNA repair gene or gain of function of important oncogenes [9]. For instance, integration in the DNA repair gene *RAD51B* results in amplification of an intronic segment and expression of alternative *RAD51B* transcripts which are non-functional [10]. Similarly, integration in the tumor suppressor gene *ETS2* is accompanied by deletion of exons at the integration site, resulting in truncated forms of the protein [10]. The integration process has also been reported to alter the copy number of potential oncogenes such as MYC, FOXE1, etc [9].

In the present study, we report the application of whole-exome sequencing technology for identifying trends in mutation profiles found in an Indian cohort of cervical cancer.

Materials and Methods

Clinical Samples

Ten pretreatment cervical tumour biopsies along with the matched blood samples from Indian patients with advanced stage cervical cancer (FIGO stage IIIB) were taken for the study. A generic consent for basic research was obtained prior to obtaining the biopsies. Since this was a retrospective study, patient informed consent for this particular project could not be obtained and permission for waiver from Institute Ethics Committee (IEC) was obtained. The study was approved by IEC. All procedures pertaining to sample collection and processing were in conformity with the IEC. The biopsies were obtained and stored in liquid Nitrogen as described earlier [8], from histologically proven, primary cervical tumour, before the start of radical radiation therapy. The samples were coded for identification by the physician prior to testing and were assigned a laboratory code to maintain confidentiality. Buffy coat was separated from blood samples by centrifugation.

Processing of tumour samples and isolation of genomic DNA

Cryosectioning of the cervical biopsies was done for hematoxylin-eosin (H&E) staining as well for isolation of DNA. The H&E stained sections were examined by the pathologist to assess the percentage of tumour in the sample. Samples with >70% tumour were taken for whole exome sequencing study. Genomic DNA was isolated from the tumour sections and buffy coat from matched blood samples, using DNA Mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. RNase treatment of the DNA samples was done using RNase A (Sigma, St. Louis, MO, USA).

Exome Capture and Next-Generation Sequencing

Matched tumour and blood samples from 10 cervical cancer patients were subjected to exome sequencing using Illumina platform. Exome capture and paired end sequencing of DNA was done using Paired-End Sample Preparation kit (Illumina, San Diego, CA, USA) followed by in-solution capture of genomic DNA using SureSelect Human All Exon kit (Agilent Technologies, Santa Clara CA, USA), targeting 37 Mb sequence from exons. For each sample, captured

DNA was sequenced by multiplexed paired end sequencing using 2 pools of samples on Illumina GAIIX platform (Illumina, San Diego, CA, USA)

Bioinformatic analysis outline

The data were QC filtered and mapped to Human Genome version 37 (hg37.p13) using Bowtie. Variants were called using *lofreq* [11], which calls SNVs from each dataset (normal and tumour) and identifies somatic variations—including low frequency, rare alleles. The SNVs called by *lofreq* were subjected to quality filtering based on a) read depth for each variation; b) strand bias; c) end mapping bias; d) mapping quality; e) base RMS quality; and f) variant phred score. A given SNV was considered to be novel if it was not in the dbSNP nor in the 1000 genome database. The resultant variations were submitted to Variant Effect Predictor server (<http://www.ensembl.org/info/docs/tools/vep/index.html>) and the effect of coding site variants identified. Further analyses were performed for the variations that resulted in a missense or nonsense mutation in coding regions (non-synonymous). The genes containing these SNVs were analyzed for their functional enrichment using GO classification (AMIGO2; <http://amigo.geneontology.org>). These were also subjected to KEGG pathway analysis (www.genome.jp/kegg/tool/search_pathway.html) and compared to those previously reported in the COSMIC database (cancer.sanger.ac.uk/cancergenome/projects/cosmic/).

Validation of the data by Sanger sequencing

A subset of variations was validated by PCR amplification of the region followed by Sanger sequencing of the amplicon. Primers were designed flanking the variation ~250 bp upstream and 250 bp downstream using NCBI primer Blast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Standard PCR conditions were followed with the annealing temperature in the range between 55–65°C. The PCR products were purified using QIAquick PCR Purification kit (Qiagen, Hilden, Germany) and sequenced on DNA Sequencer (3100 Avant Genetic analyzer, Applied Biosystems, Foster City, CA, USA).

Results

Clinical material

In the present study, 10 tumour biopsies along with matched blood from treatment naïve cervical cancers from a cohort of Indian women were included. The patients were in FIGO stage IIIB and their age ranged between 35–58 years. The HPV genotype and viral genome status in these patients was determined from an earlier study [8, 10]. Majority of the samples were positive for HPV16, although HPV18, 45 and 56 were also present [12]. In some of the samples, the virus was integrated in the genome [8]. The demographic details of the patients are given in [Table 1](#).

Exome sequencing and somatic variations

The matched tumour and blood samples from 10 cervical cancer patients were subjected to exome sequencing using Illumina platform. On an average 97% of the reads could be mapped to the human genome (hg37.p13) and the average read depth was >20X. The matched blood control in each case was used to filter out the germline variations. Further quality filtering by methods as described in Materials and Methods, identified 2281 somatic variations falling within coding regions of which 1233 (54%) were of missense type, 68 (3%) stop gained, and the rest were synonymous coding and stop (retained) variations. The frequency of mutations in each of the 10 cases is shown in [Fig 1](#). A total of 1831 genes were found to harbour somatic

Table 1. Demographic details of the 10 patients included in the whole exome sequencing study

Sample ID	Histology	FIGO stage	Age (yrs.)	Disease status	HPV genotype*	Viral genome status [#]	Integration site [#]
1123	SCC	IIIB	40	NED	16	Integrated	Unknown
999	SCC	IIIB	49	NED	16	Integrated	4q13.3
938	SCC	IIIB	50	LR + DM	16, 18	Not known	
1099	SCC	IIIB	38	NED	16	Not known	
1103	SCC	IIIB	50	DM	16	Not known	
755	SCC	IIIB	44	LR	16	Integrated	3q23
785	SCC	IIIB	43	LR	16, 18	Integrated	3q23, 20q11.21
937	SCC	IIIB	50	LR	Not detected	-	-
1023	SCC	IIIB			16	Not known	
1094	SCC	IIIB	58	DM	16	Integrated	13q

Key: SCC- Squamous cell carcinoma; NED-No Evidence of Disease; LR-Local Recurrence; DM-Distant Metastasis; HPV-Human Papilloma Virus

*HPV was determined by Luminex Bead Array and HPV type specific PCR

[#]Viral genome status and integration site were determined by APOT assay as described in Ref. 8

doi:10.1371/journal.pone.0165878.t001

variations. A total of 299 genes carried at least 2 variations (either in same tumour or across tumours). A panel of 87 genes were found to be mutated in at least 2 tumours (Fig 2).

Two genes, PDE4DIP (mutated in 4 tumours) and MUC16 (mutated in 2 tumours) carried 11 somatic variations each, making them the most mutated genes in the dataset. MUC17 showed a high frequency of mutations in a single sample. Apart from this, 80 ZNF family members representing 4.5% of the total number of genes harbouring somatic mutations were also found to be mutated. ZNF proteins had an overrepresentation in the list of somatically mutated genes, when compared to their overall frequency in the human genes (538 out of 32620 genes: 1.6%). This overrepresentation was statistically significant (Fisher’s $p = 2.6 \times 10^{-13}$; OR = 2.72; 95% CI = 2.11–3.47).

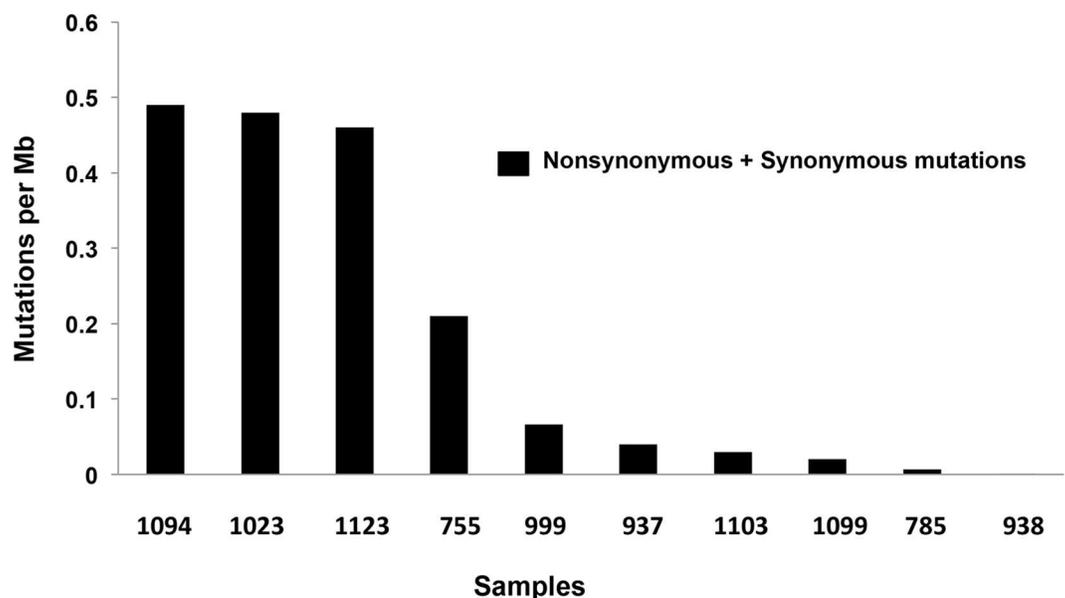


Fig 1. Frequency of somatic mutations in each of the 10 cases. The frequency of all the mutations (synonymous and nonsynonymous) per Mb of genome is depicted. Each bar represents an individual tumour sample and the number below the bar is the designated laboratory number of each sample.

doi:10.1371/journal.pone.0165878.g001

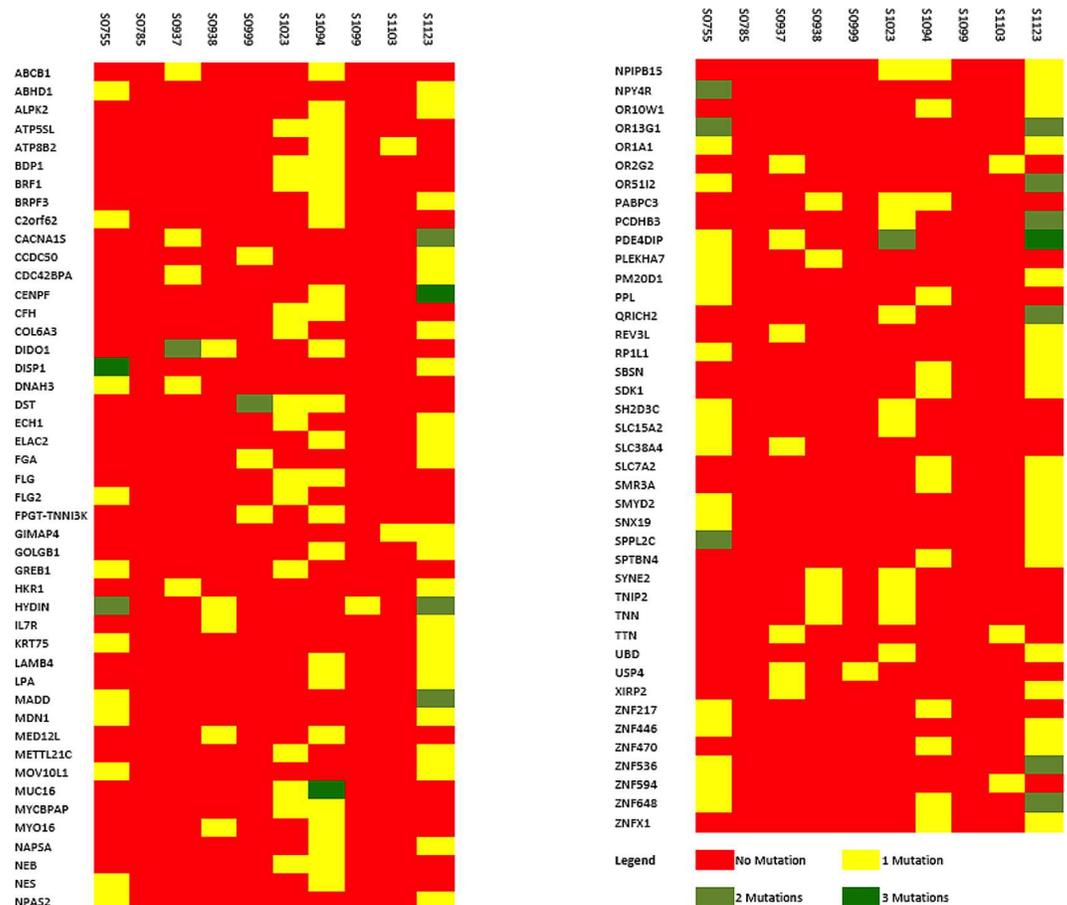


Fig 2. Heat map of frequently mutated genes. A panel of 87 genes was found to be mutated in at least 2 tumour samples. Some of the genes had 3 mutations in the same sample.

doi:10.1371/journal.pone.0165878.g002

MAPK signal transduction pathway that is well known to play a key role in the pathogenesis of cancer, surfaced as one with most mutated genes in this dataset. KEGG lists 257 genes as part of the MAPK signal transduction pathway, including the core signal transduction cascade, the upstream ligands and other associated factors. Twelve of the 257 genes were identified to have somatic variations across patients. In order to assess the significance of this occurrence, we generated 1000 random gene sets of 257 each and calculated the frequency of observed somatic mutations in these sets; Z-scores were used as a measure of significance. The Z-score for finding 12 genes with somatic mutations out of 257 random genes was 2.69 (P-value = 0.004). This indicates a significant enrichment for somatic mutations in the MAPK pathway of these tumour samples

(Fig 3).

Missense somatic variations from each paired tumour sample were used to analyze functional enrichment under the GO molecular function categories. All GO category enrichments with an adjusted p = 0.05 or less were considered to be significant. We did not find any negative enrichment (underrepresentation) of GO functional categories. In a combined GO analysis across multiple samples olfactory receptors, ion binding, signalling cascade receptor and signal transduction genes were found to be enriched.

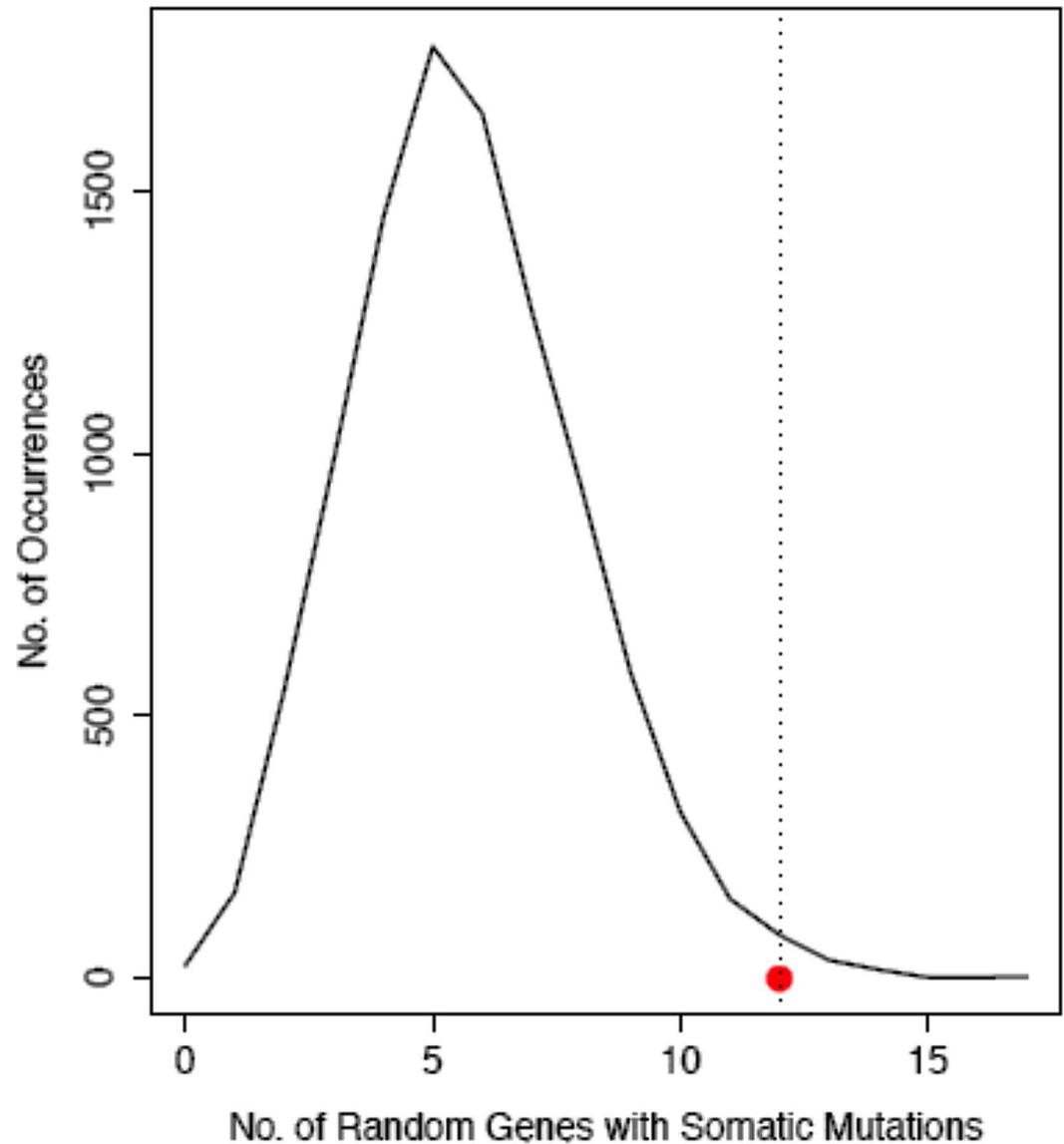


Fig 3. Randomization test for somatic mutations. The curve represents the distribution of number of genes that belonged to the MAPK pathway in the 1000 sets of random genes. The mean of the data set is 5.7 somatic mutations in 257 genes with a standard deviation of 2.34. The red dot along the X-axis depicts a value of 12—the number of MAPK genes having somatic mutations. The P-value was calculated based on the number of random sets that had less than 12 genes belonging to MAPK pathway. The number of somatic mutations observed in MAPK pathway is significantly higher than the mean ($P = 0.0046$).

doi:10.1371/journal.pone.0165878.g003

Validation of selected variations by Sanger sequencing

A total of 53 somatic variations were randomly selected for variation by Sanger sequencing. Out of this, 25 were validated as true somatic and 10 turned out to be germline (Table 2).

Comparison with COSMIC database

From the present dataset missense mutations were catalogued across 1282 genes, out of 1831 genes harbouring somatic mutations. These missense variations (nonsynonymous + stop-gained) were compared with pre-existing mutations in the COSMIC database. 272 mutations

Table 2. Variations from exome sequencing data validated by Sanger sequencing.

Sr. No	Gene name	Mutation	Amino acid change	Variation Type	Nature of variation	COSMIC overlap	Importance	Domain
1.	FGF7	tAc/tGc	Y76C	Missense	Somatic	No	MAPKinase signaling pathway; pathways in cancer	FGF domain
2.	TGFBR2	Cag/Tag	Q220*	Stop gain	Somatic	No	MAPKinase pathway; pathways in cancer	
3.	TAF5	Gag/Cag	E383Q	Missense	Somatic	No	Basal transcription factor	
4.	MAP3K3	Gag/Aag	E645K	Missense	Somatic	No	MAPKinase pathway	
5.	DIDO1	tCg/tTg	R303Q	Missense	Somatic	Yes	Involved in apoptosis	PHD Zinc finger domain
6.	DIDO1	ctC/ctG	E301Q	Missense	Somatic	No	Involved in apoptosis	PHD Zinc finger domain
7.	SOS2	tGa/tAa	S1049L	Missense	Somatic	No	MAPKinase pathway; pathways in cancer.	
8.	RASA1	tGg/tAg	W490*	Stop gain	Somatic	No	MAPKinase pathway	PH domain
9.	EP300	Gat/Aat	D1399N	Missense	Somatic	Yes (stomach and cervix)	Signaling pathway; Play a role in epithelial cancer	
10.	RNF111	Cag/Tag	Q605*	Stop gain	Somatic	Yes (acute myeloid leukaemia)	Regulation of transcription	
11.	COL15A1	Gga/Tga	G990*	Stop gain	Somatic	No	Signal transduction; Role in angiogenesis	
12.	NCOR2	aCc/aAc	G1289V	Missense	Somatic	No	Notch signaling pathway	
13.	EDNRB	Caa/Gaa	L233F	Missense	Somatic	No	Regulation of transcription; phospholipase C-activating G-protein coupled receptor signaling pathway; etc.	
14.	MYD88	Gac/Cac	D113H	Missense	Somatic	No	regulation of I-kappaB kinase/NF-kappaB cascade and NF-kappaB transcription factor activity	
15.	CD274	Gaa/Caa	E152Q	Missense	Somatic	No	Cell adhesion molecules (CAMs); signal transduction	
16.	CDCA7L	GTa/gCa	Y64C	Missense	Somatic	No	positive regulation of cell proliferation; regulation of transcription	
17.	PRPF4B	cGt/cAt	R856H	Missense	Somatic	No	protein serine/threonine kinase activity	
18.	ABCG5	cGt/cAt	T489M	Missense	Somatic	No	ATPase activity; ABC transporters	Transmembrane domain
19.	NOL9	acC/acG	G153R	Missense	Somatic	No	polynucleotide 5'-hydroxyl-kinase activity	
20.	NOL9	ctC/ctG	E185Q	Missense	Somatic	No	polynucleotide 5'-hydroxyl-kinase activity	
21.	PTPDC1	Gag/Aag	E639K	Missense	Somatic	No	protein tyrosine/serine/threonine phosphatase activity	
22.	ODZ1	ctG/ctA	Q957*	Stop gain	Somatic	No	negative regulation of cell proliferation	
23.	ALS2CR8	Gaa/Aaa	E688K	Missense	Somatic	No	sequence-specific DNA binding transcription factor activity	
24.	MACF1	Cgg/Tgg	R2091W	Missense	Somatic	No	cell cycle arrest	Spectrin domain
25.	TTN	gaC/gaT	V2387I	Missense	Somatic	Yes (breast cancer)	protein serine/threonine kinase activity; protein tyrosine kinase activity	
26.	G6PD	caC/caT	V390M	Missense	Germline	No	Metabolic pathways	
27.	FGF22	Cgg/Tgg	R100W	Missense	Germline	No	MAPKinase Pathway; Pathways in cancer	FgGF domain
28.	FMOD	gGa/gAa	S105F	Missense	Germline	No	transforming growth factor beta receptor complex assembly	LRR domain

(Continued)

Table 2. (Continued)

Sr. No	Gene name	Mutation	Amino acid change	Variation Type	Nature of variation	COSMIC overlap	Importance	Domain
29.	HIVEP3	ctC/ctT	E155K	Missense	Germline	No	positive regulation of transcription, may be a previously unidentified tumor-suppressor gene	
30.	AIM1	Cat/Aat	H219N	Missense	Germline	No		
31.	SEPN1	Agc/Ggc	S388G	Missense	Germline	No		
32.	SEPN1	Gtg/Atg	V534M	Missense	Germline	No		
33.	HAGH	gaC/gaA	V207F	Missense	Germline	No	Metabolic pathways	
34.	GRAMD1A	cGc/cAg	R117Q	Missense	Germline	No		Gram Domain
35.	GRAMD1A	cCc/cGc	P554R	Missense	Germline	No		Gram Domain

Key:

* indicates termination of the amino acid chain

doi:10.1371/journal.pone.0165878.t002

in 250 genes were common with those reported in COSMIC database (see [S1 Table](#)). However, the data did not match any known cervical cancer mutations.

Discussion

Although HPV is the major causative factor for development of cervical cancer, the role of several genetic components in the disease has long been recognized. While allelic variation of HLA class II antigen [13, 14], and p53 codon 72 polymorphism [15] may increase the risk of developing the disease, mutations in Ras oncogene, PI3KA, STK11, are reported to directly cause or promote progression of cervical carcinoma [2, 3, 5]. Apart from this, a number of chromosomal aberrations have also been associated with the disease [4]. However, till date there are very few high throughput studies that describe the genetic landscape of the disease. The study by Ojesina et al. involving 100 Norwegian and 15 Mexican cervical carcinoma cases is the most significant among them. They identified recurrent mutations in MAPK, and mutations in genes such as HLA-B, EP300, FBXW7, NFE2L2, TP53 and ERBB2 in 79 primary squamous cell carcinomas and that of ELF3 and CBF3 in 24 adenocarcinomas [6]. A study from Latin America reported mutations in PIK3CA to be a key player both in squamous and adeno cervical carcinomas [16]. Another study on cervical cancer cases of European origin identified PIK3CA, KRAS and FBXW7 to be the most mutated [17]. Chung et al have reported mutations in FAT1, ARID1A, ERBB2 and PIK3CA in 15 cervical adenocarcinomas cases of Chinese origin [18]. However, there are no reports on such studies on the Indian population. In the present study, using whole exome sequencing technology we have attempted to catalogue the somatic aberrations in a cohort of Indian cervical cancer cases. Initially identified as an effective tool for identification of a wide variety of novel variations especially in the field of Mendelian disorders [19], over past few years the technology has been quiet successful in discovering novel mutations in diseases characterized by marked genetic heterogeneity like cancer including the rare forms [20, 21].

Whole exome sequencing of 10 paired advanced stage cervical cancer biopsies from Indian cohort identified 1029 novel, non-synonymous variations. High frequency of mutations in two genes—PDE4DIP (a number of mutations in the same gene across samples) and MUC17 (a number of mutations in the same gene in only one sample) were found. Known for its role in anchoring phosphodiesterase 4D to the Golgi/centrosome region of the cell, deleterious mutations of PDE4DIP have recently been demonstrated to be involved in breast cancer [22].

A number of membrane-associated mucin genes such as MUC2, 4, 5B, 12, 16, 17, 20, 21 were found to be mutated in our dataset. Similar to our results, whole exome sequencing study of HPV negative versus HPV positive head and neck tumours identified MUC4, 12 and 16 as mutated [23]. We found MUC16 to be recurrently mutated across the samples. A number of MUC mutations have been reported in the COSMIC database for several cancer types. In general, the membrane associated mucins are also known to play role in invasion and metastasis by modulating the adhesive properties of cells [24]. On the other hand, it is possible that mutations in Muc16 as well as Muc4 are false positives as these are large genes which are low expressed and late replicating and therefore show a high frequency of mutations in the coding region [25].

An interesting finding from the present study was the overrepresentation of mutated members of the ZNF family. A total of 80 ZNF genes were found to be mutated. Zinc Finger members have been implicated in tumorigenesis and acquisition of resistance to drugs. Members of these family interact with nucleic acids, proteins and small molecules and are involved in a variety of crucial molecular processes in cells at replication, transcriptional and translational levels. Mutations of ZNF family members are associated with various diseases including cancer. Zinc Finger genes map to regions often deleted in solid tumours and involved in recurrent chromosomal rearrangements in hematological malignancies [26]. Deleterious mutation of one of the ZNF family members, PRDM1 has been associated with primary central nervous system lymphoma [27]. Corroborating our findings, a recent study has reported multiple mutations of the ZNF family genes in metastatic hepatocellular carcinoma. [28]. As reported in the study by Nicholas et. al., for HPV positive and negative head and neck squamous cancer, ZNF family members including ZNF10, ZNF470 and ZNF716 were also found to be mutated in our dataset [23]

GO analysis identified a large number of olfactory receptor gene mutations. Mutations in olfactory receptors are now being associated with cancer [29] [30], tumour invasion and metastasis [31]. However, according to a recent report, high frequency of somatic mutations in these genes may be red herrings during high throughput sequencing of cancer genomes. According to this study, genes such as the olfactory receptors that have low expression, are late in replicating and display a high regional noncoding mutation rate, are more prone to somatic mutations in their coding regions and hence are seen in many cancer genomics data [25]. Therefore, in order to conclude whether mutations in these receptors play any role in cervical carcinogenesis, more samples need to be sequenced as well as functional analysis of these genes in cell lines and patient samples is required.

Of the mutations that have occurred in single samples and have been identified and validated from the dataset, there were several significant candidates known for their role in the process of tumorigenesis. For instance FGF7 is known for its mitogenic and cell survival activities [32]; MAP3K3 is involved in direct activation of stress-activated protein kinase (SAPK) and extracellular signal-regulated protein kinase (ERK) pathways [33] as well as plays a critical role in TNF-induced NF-kappaB activation [34]. RASA1 helps in control of cellular proliferation and its mutation is associated with basal cell carcinoma [35]. RNF111 acts as a tumour suppressor in preventing development and progression of colorectal cancer [36]. CDCA7L interacts with c-Myc in bringing about cellular transformation in medulloblastoma [37]. Mutation of CD274 leading to its overexpression have been associated with gastric carcinoma [38]. Apart from these, mutations in ODZ1, TTN and EP300 are already reported in cervical cancer. The variation in the transcriptional regulator, EP300 (D1399N) was identical to that reported in COSMIC database. EP300 was also reported to be one of the genes with significant recurrent somatic mutations in the study by Ojesina *et al* [6].

Despite the important role played by each of these genes, majority of the mutations would be 'passengers' and identifying potential 'drivers' with a cardinal role in cervical carcinogenesis

from this data would be the actual challenge. Several methods have been described that can predict potential driver mutations in a cancer genome [39, 40]. These, combined with functional characterization of the variants and validation studies involving larger sample size might help in identifying potential 'drivers' in cervical carcinogenesis.

A significant proportion of the genes found to be mutated such as MAP3K3, SOS2, FGF7, TGFBR2, TRAF2, RASA1 and RPS6KA6 belong to the MAPKinase signal transduction pathway that is known to play a significant role in cell proliferation and apoptosis. This observation is in concordance with the study by Ojesina *et al* [6]. This indicates that the MAPKinase pathway is one of the highly mutated pathways in cervical carcinogenesis and hence might be more vulnerable to therapeutic intervention.

The major highlight of our study was to identify non-synonymous, somatic variations (mutations) that might play a potential role in cervical carcinogenesis. We could not find any significant driver mutation that could lead to cervical carcinogenesis. This might be due to the fact that all (except one) of the samples were positive for high risk HPV which are known to target the two key cell cycle regulators-p53 and Rb for proteasomal degradation, without incurring mutation. In our study, HPV infection in the samples and its integration was determined by Luminex Bead Array [8]. Presence of HPV and its integration status was also confirmed in the exome sequencing data of the samples [12]. Thus, our treatment naïve primary cancers could be HPV driven and hence no driver mutations are seen.

Our results suggest that there is considerable genetic and mechanistic heterogeneity in cervical cancers. Pathways / genes that are frequently mutated in various other cancers are also mutated in cervical cancers indicating that specific mutational signatures notwithstanding, distinct cancers based on their tissue of origin, are offshoots of the same mechanistic process. Thus, a tumour driving pathway in prostate or lung cancer may well be the culprit even in cervical cancer. Given this convergence of pathways / mechanisms across cancer types, one can explore the possibility of cross-treating cancers, for example, the same drug may be effective against two divergent cancer types if they carry the same mutational signature at the pathway level.

To the best of our knowledge, this is the first study where whole exome sequencing has been applied to understanding the mutational landscape of the disease in an Indian cohort. However, there might be several polymorphisms that predispose an individual to the disease or aid the HPV in causing cervical cancer, which we have not dealt with in this study. An additional aspect of the study is that it provides an insight into the genome of Indian individuals by identifying a number of novel variations which could enrich the existing databases.

Supporting Information

S1 Table. List of somatic mutations in common with COSMIC database
(XLSX)

Acknowledgments

The authors thank all individuals from the Gynaecology Disease Management Group, Tata Memorial Hospital, Mumbai who were involved in judiciously collecting, storing the biopsies from cervical cancer patients and compiling the clinical history.

Author Contributions

Conceptualization: RM PD.

Data curation: RM.

Formal analysis: KS.

Funding acquisition: RM.

Investigation: RM PD AB UM SKS KD.

Methodology: PD AB UM SKS KD SNR.

Project administration: RM.

Resources: SNR RM.

Software: SNR KS.

Supervision: RM.

Validation: PD AB.

Visualization: RM PD KS.

Writing – original draft: PD.

Writing – review & editing: RM PD KS UM.

References

1. GLOBOCAN 2012 (IARC) Section of Cancer Surveillance (17/12/2015). 2015.
2. Cui B, Zheng B, Zhang X, Stendahl U, Andersson S, Wallin KL. Mutation of PIK3CA: possible risk factor for cervical carcinogenesis in older women. *Int J Oncol.* 2009 Feb; 34(2):409–16. PMID: [19148475](#)
3. Wong YF, Chung TK, Cheung TH, Lam SK, Xu YG, Chang AM. Frequent ras gene mutations in squamous cell cervical cancer. *Cancer Lett.* 1995 Aug 16; 95(1–2):29–32. PMID: [7656240](#)
4. Narayan G, Murty VV. Integrative genomic approaches in cervical cancer: implications for molecular pathogenesis. *Future Oncol.* 2010 Oct; 6(10):1643–52. doi: [10.2217/fo.10.114](#) PMID: [21062161](#)
5. Wingo SN, Gallardo TD, Akbay EA, Liang MC, Contreras CM, Boren T, et al. Somatic LKB1 mutations promote cervical cancer progression. *PLoS One.* 2009; 4(4):e5137. doi: [10.1371/journal.pone.0005137](#) PMID: [19340305](#)
6. Ojesina AI, Lichtenstein L, Freeman SS, Pedamallu CS, Imaz-Rosshandler I, Pugh TJ, et al. Landscape of genomic alterations in cervical carcinomas. *Nature.* 2014 Feb 20; 506(7488):371–5. doi: [10.1038/nature12881](#) PMID: [24390348](#)
7. Christiansen IK, Sandve GK, Schmitz M, Durst M, Hovig E. Transcriptionally active regions are the preferred targets for chromosomal HPV integration in cervical carcinogenesis. *PLoS One.* 2015; 10(3): e0119566. doi: [10.1371/journal.pone.0119566](#) PMID: [25793388](#)
8. Das P, Thomas A, Mahantshetty U, Shrivastava SK, Deodhar K, Mulherkar R. HPV genotyping and site of viral integration in cervical cancers in Indian women. *PLoS One.* 2012; 7(7):e41012. doi: [10.1371/journal.pone.0041012](#) PMID: [22815898](#)
9. Rusan M, Li YY, Hammerman PS. Genomic landscape of human papillomavirus-associated cancers. *Clin Cancer Res.* 2015 May 1; 21(9):2009–19. doi: [10.1158/1078-0432.CCR-14-1101](#) PMID: [25779941](#)
10. Parfenov M, Pedamallu CS, Gehlenborg N, Freeman SS, Danilova L, Bristow CA, et al. Characterization of HPV and host genome interactions in primary head and neck cancers. *Proc Natl Acad Sci U S A.* 2014 Oct 28; 111(43):15544–9. doi: [10.1073/pnas.1416074111](#) PMID: [25313082](#)
11. Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* 2012 Dec; 40(22):11189–201. doi: [10.1093/nar/gks918](#) PMID: [23066108](#)
12. Chandrani P, Kulkarni V, Iyer P, Upadhyay P, Chaubal R, Das P, et al. NGS-based approach to determine the presence of HPV and their sites of integration in human cancer genome. *Br J Cancer.* 2015 Jun 9; 112(12):1958–65. doi: [10.1038/bjc.2015.121](#) PMID: [25973533](#)
13. Hildesheim A, Schiffman M, Scott DR, Marti D, Kissner T, Sherman ME, et al. Human leukocyte antigen class I/II alleles and development of human papillomavirus-related cervical neoplasia: results from a case-control study conducted in the United States. *Cancer Epidemiol Biomarkers Prev.* 1998 Nov; 7(11):1035–41. PMID: [9829713](#)

14. Madeleine MM, Brumback B, Cushing-Haugen KL, Schwartz SM, Daling JR, Smith AG, et al. Human leukocyte antigen class II and cervical cancer risk: a population-based study. *J Infect Dis*. 2002 Dec 1; 186(11):1565–74. doi: [10.1086/345285](https://doi.org/10.1086/345285) PMID: [12447731](https://pubmed.ncbi.nlm.nih.gov/12447731/)
15. Storey A, Thomas M, Kalita A, Harwood C, Gardiol D, Mantovani F, et al. Role of a p53 polymorphism in the development of human papillomavirus-associated cancer. *Nature*. [Research Support, Non-U.S. Gov't]. 1998 May 21; 393(6682):229–34. doi: [10.1038/30400](https://doi.org/10.1038/30400) PMID: [9607760](https://pubmed.ncbi.nlm.nih.gov/9607760/)
16. Lou H, Villagran G, Boland JF, Im KM, Polo S, Zhou W, et al. Genome Analysis of Latin American Cervical Cancer: Frequent Activation of the PIK3CA Pathway. *Clin Cancer Res*. 2015 Dec 1; 21(23):5360–70. doi: [10.1158/1078-0432.CCR-14-1837](https://doi.org/10.1158/1078-0432.CCR-14-1837) PMID: [26080840](https://pubmed.ncbi.nlm.nih.gov/26080840/)
17. Muller E, Brault B, Holmes A, Legros A, Jeannot E, Campitelli M, et al. Genetic profiles of cervical tumors by high-throughput sequencing for personalized medical care. *Cancer Med*. 2015 Oct; 4(10):1484–93. doi: [10.1002/cam4.492](https://doi.org/10.1002/cam4.492) PMID: [26155992](https://pubmed.ncbi.nlm.nih.gov/26155992/)
18. Chung TK, Van Hummelen P, Chan PK, Cheung TH, Yim SF, Yu MY, et al. Genomic aberrations in cervical adenocarcinomas in Hong Kong Chinese women. *Int J Cancer*. 2015 Jan 27.
19. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*. 2011 Nov; 12(11):745–55. doi: [10.1038/nrg3031](https://doi.org/10.1038/nrg3031) PMID: [21946919](https://pubmed.ncbi.nlm.nih.gov/21946919/)
20. Harbour JW, Onken MD, Roberson ED, Duan S, Cao L, Worley LA, et al. Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science*. 2010 Dec 3; 330(6009):1410–3. doi: [10.1126/science.1194472](https://doi.org/10.1126/science.1194472) PMID: [21051595](https://pubmed.ncbi.nlm.nih.gov/21051595/)
21. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011 Oct 6; 478(7367):64–9. doi: [10.1038/nature10496](https://doi.org/10.1038/nature10496) PMID: [21909114](https://pubmed.ncbi.nlm.nih.gov/21909114/)
22. Horvath A, Pakala SB, Mudvari P, Reddy SD, Ohshiro K, Casimiro S, et al. Novel insights into breast cancer genetic variance through RNA sequencing. *Sci Rep*. 2013; 3:2256. doi: [10.1038/srep02256](https://doi.org/10.1038/srep02256) PMID: [23884293](https://pubmed.ncbi.nlm.nih.gov/23884293/)
23. Nichols AC, Chan-Seng-Yue M, Yoo J, Xu W, Dhaliwal S, Basmaji J, et al. A Pilot Study Comparing HPV-Positive and HPV-Negative Head and Neck Squamous Cell Carcinomas by Whole Exome Sequencing. *ISRN Oncol*. 2012; 2012:809370. doi: [10.5402/2012/809370](https://doi.org/10.5402/2012/809370) PMID: [23304554](https://pubmed.ncbi.nlm.nih.gov/23304554/)
24. Hilkens J, Ligtenberg MJ, Vos HL, Litvinov SV. Cell membrane-associated mucins and their adhesion-modulating property. *Trends Biochem Sci*. 1992 Sep; 17(9):359–63. PMID: [1412714](https://pubmed.ncbi.nlm.nih.gov/1412714/)
25. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013 Jul 11; 499(7457):214–8. doi: [10.1038/nature12213](https://doi.org/10.1038/nature12213) PMID: [23770567](https://pubmed.ncbi.nlm.nih.gov/23770567/)
26. Tommerup N, Vissing H. Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. *Genomics*. 1995 May 20; 27(2):259–64. doi: [10.1006/geno.1995.1040](https://doi.org/10.1006/geno.1995.1040) PMID: [7557990](https://pubmed.ncbi.nlm.nih.gov/7557990/)
27. Courts C, Montesinos-Rongen M, Brunn A, Bug S, Siemer D, Hans V, et al. Recurrent inactivation of the PRDM1 gene in primary central nervous system lymphoma. *J Neuropathol Exp Neurol*. 2008 Jul; 67(7):720–7. doi: [10.1097/NEN.0b013e31817dd02d](https://doi.org/10.1097/NEN.0b013e31817dd02d) PMID: [18596541](https://pubmed.ncbi.nlm.nih.gov/18596541/)
28. Ouyang L, Lee J, Park CK, Mao M, Shi Y, Gong Z, et al. Whole-genome sequencing of matched primary and metastatic hepatocellular carcinomas. *BMC Med Genomics*. 2014; 7:2. doi: [10.1186/1755-8794-7-2](https://doi.org/10.1186/1755-8794-7-2) PMID: [24405831](https://pubmed.ncbi.nlm.nih.gov/24405831/)
29. Xu L, Tang H, Chen DW, El-Naggar AK, Wei P, Sturgis EM. Genome-wide association study identifies common genetic variants associated with salivary gland carcinoma and its subtypes. *Cancer*. 2015 Jul 15; 121(14):2367–74. doi: [10.1002/cncr.29381](https://doi.org/10.1002/cncr.29381) PMID: [25823930](https://pubmed.ncbi.nlm.nih.gov/25823930/)
30. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*. 2008 Sep 26; 321(5897):1807–12. doi: [10.1126/science.1164382](https://doi.org/10.1126/science.1164382) PMID: [18772396](https://pubmed.ncbi.nlm.nih.gov/18772396/)
31. Sanz G, Leray I, Dewaele A, Sobilo J, Lerondel S, Bouet S, et al. Promotion of cancer cell invasiveness and metastasis emergence caused by olfactory receptor stimulation. *PLoS One*. 2014; 9(1):e85110. doi: [10.1371/journal.pone.0085110](https://doi.org/10.1371/journal.pone.0085110) PMID: [24416348](https://pubmed.ncbi.nlm.nih.gov/24416348/)
32. Ornitz DM, Xu J, Colvin JS, McEwen DG, MacArthur CA, Coulier F, et al. Receptor specificity of the fibroblast growth factor family. *J Biol Chem*. 1996 Jun 21; 271(25):15292–7. PMID: [8663044](https://pubmed.ncbi.nlm.nih.gov/8663044/)
33. Ellinger-Ziegelbauer H, Brown K, Kelly K, Siebenlist U. Direct activation of the stress-activated protein kinase (SAPK) and extracellular signal-regulated protein kinase (ERK) pathways by an inducible mitogen-activated protein Kinase/ERK kinase kinase 3 (MEKK) derivative. *J Biol Chem*. 1997 Jan 31; 272(5):2668–74. PMID: [9006902](https://pubmed.ncbi.nlm.nih.gov/9006902/)

34. Yang J, Lin Y, Guo Z, Cheng J, Huang J, Deng L, et al. The essential role of MEKK3 in TNF-induced NF-kappaB activation. *Nat Immunol.* 2001 Jul; 2(7):620–4. doi: [10.1038/89769](https://doi.org/10.1038/89769) PMID: [11429546](https://pubmed.ncbi.nlm.nih.gov/11429546/)
35. Friedman E, Gejman PV, Martin GA, McCormick F. Nonsense mutations in the C-terminal SH2 region of the GTPase activating protein (GAP) gene in human tumours. *Nat Genet.* 1993 Nov; 5(3):242–7. doi: [10.1038/ng1193-242](https://doi.org/10.1038/ng1193-242) PMID: [8275088](https://pubmed.ncbi.nlm.nih.gov/8275088/)
36. Sharma V, Antonacopoulou AG, Tanaka S, Panoutsopoulos AA, Bravou V, Kalofonos HP, et al. Enhancement of TGF-beta signaling responses by the E3 ubiquitin ligase Arkadia provides tumor suppression in colorectal cancer. *Cancer Res.* 2011 Oct 15; 71(20):6438–49. doi: [10.1158/0008-5472.CAN-11-1645](https://doi.org/10.1158/0008-5472.CAN-11-1645) PMID: [21998011](https://pubmed.ncbi.nlm.nih.gov/21998011/)
37. Huang A, Ho CS, Ponzielli R, Barsyte-Lovejoy D, Bouffet E, Picard D, et al. Identification of a novel c-Myc protein interactor, JPO2, with transforming activity in medulloblastoma cells. *Cancer Res.* 2005 Jul 1; 65(13):5607–19. doi: [10.1158/0008-5472.CAN-05-0500](https://doi.org/10.1158/0008-5472.CAN-05-0500) PMID: [15994933](https://pubmed.ncbi.nlm.nih.gov/15994933/)
38. Wang W, Sun J, Li F, Li R, Gu Y, Liu C, et al. A frequent somatic mutation in CD274 3'-UTR leads to protein over-expression in gastric cancer by disrupting miR-570 binding. *Hum Mutat.* 2012 Mar; 33(3):480–4. doi: [10.1002/humu.22014](https://doi.org/10.1002/humu.22014) PMID: [22190470](https://pubmed.ncbi.nlm.nih.gov/22190470/)
39. Torkamani A, Schork NJ. Prediction of cancer driver mutations in protein kinases. *Cancer Res.* 2008 Mar 15; 68(6):1675–82. doi: [10.1158/0008-5472.CAN-07-5283](https://doi.org/10.1158/0008-5472.CAN-07-5283) PMID: [18339846](https://pubmed.ncbi.nlm.nih.gov/18339846/)
40. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science.* 2006 Oct 13; 314(5797):268–74. doi: [10.1126/science.1133427](https://doi.org/10.1126/science.1133427) PMID: [16959974](https://pubmed.ncbi.nlm.nih.gov/16959974/)