



Interactome of vertebrate GAF/ThPOK reveals its diverse functions in gene regulation and DNA repair

AVINASH SRIVASTAVA and RAKESH K MISHRA*

CSIR-Centre for Cellular and Molecular Biology (CCMB), Uppal Road, Hyderabad 500 007, India

*Corresponding author (Email, mishra@ccmb.res.in)

MS received 28 September 2019; accepted 23 January 2020; published online 10 February 2020

GAGA associated factor (GAF) is a sequence-specific DNA binding transcription factor that is evolutionarily conserved from flies to humans. Emerging evidence shows a context-dependent function of vertebrate GAF (vGAF, a.k.a. ThPOK) in multiple processes like gene activation, repression, and enhancer-blocking. We hypothesize that context-dependent interaction of vGAF with a diverse set of proteins forms the basis for the multifunctional nature of vGAF. To this end, we deciphered the protein–protein interactome of vGAF and show that vGAF interacts with chromatin remodelers, RNA metabolic machinery, transcriptional activators/repressors, and components of DNA repair machinery. We further validated the biological significance of our protein–protein interaction data with functional studies and established a novel role of vGAF in DNA repair and cell-survival after UV-induced DNA damage. One of the major risk factors for skin cutaneous melanoma is prolonged exposure of UV and subsequent DNA damage. vGAF is highly expressed in normal skin tissue. Interestingly, our analysis of high-throughput RNA-sequencing data shows that vGAF is heavily downregulated across all major stages of skin cutaneous melanoma suggesting its potential as a diagnostic biomarker. Taken together, our study provides a plausible explanation for the diverse gene regulatory functions of vGAF and unravels its novel role in DNA repair.

Keywords. Vertebrate GAF; ThPOK; DNA repair; protein-interactome; skin cutaneous melanoma

1. Introduction

GAGA associated factor (GAF) of *Drosophila melanogaster*; DmGAF, is a developmentally important transcription factor that has been implicated in diverse nuclear processes like gene activation, polycomb-mediated silencing, enhancer-blocking, position effect variegation and chromosomal segregation (Schweinsberg *et al.* 2004; Mishra *et al.* 2001; Bhat *et al.* 1996; Farkas *et al.* 1994; Lu *et al.* 1993; Biggin and Tjian 1988). Vertebrate GAF (vGAF), the evolutionarily conserved counterpart of DmGAF in vertebrates, is encoded by the *zbtb7b* gene (Matharu *et al.* 2010). vGAF/ThPOK/ZBTB7B has been shown to be essential for the proper development of the hematopoietic and adipose tissues (Li *et al.* 2017; Egawa and Littman 2008; Muroi *et al.* 2008). A recent study also points out the role of vGAF in the onset of lactation and in the

production of milk lipids (Zhang *et al.* 2018). The molecular functions of vGAF that translate into these biological processes are diverse in nature. vGAF was first identified as a tissue-specific activator of collagen genes in mouse dermis (Galera *et al.* 1994, 1996). Several other studies also point the role of vGAF in transcriptional activation of genes involved in cytokine signaling like *Socs1*, *Cish* and *TNF-alpha* (Stratigi *et al.* 2015; Luckey *et al.* 2014). Paradoxically, multiple other studies report that vGAF has a rather repressive effect on the transcription of another set of genes like *UDP glucose dehydrogenase*, *eomesodermin*, and *cd8* (Li *et al.* 2013; Rui *et al.* 2012; Beauchef *et al.* 2005). These are not, however, the only nuclear processes where vGAF has been shown to be involved. vGAF has been reported to bind at the enhancer-blocker elements in the murine *Hox* clusters and is essential for tissue-specific nuclear-lamina proximal

Electronic supplementary material: The online version of this article (<https://doi.org/10.1007/s12038-020-0014-0>) contains supplementary material, which is available to authorized users.

positioning of *IgH* and *cyp3a* loci, suggesting an architectural role of vGAF in chromatin organization (Srivastava *et al.* 2013; Zullo *et al.* 2012). Moreover, a recent study has shown a developmentally important association of vGAF with ribonucleoprotein complexes containing a lincRNA (long intergenic non-coding RNA) (Li *et al.* 2017). The remarkable functional versatility of vGAF, at least in part, may stem from the structural domains of the protein (Srivastava *et al.* 2018). The zinc-finger domains present at the C-terminal of vGAF allow sequence-specific tethering of vGAF molecules at numerous loci in the genome while the N-terminal BTB/POZ domain being a protein-protein interaction domain endows GAF with an ability to recruit protein-complexes of diverse functionalities at these loci. Several pieces of evidence suggest the importance of protein-protein interactions in the multifunctional roles of GAF proteins (Lomaev *et al.* 2017; Rui *et al.* 2012; Zullo *et al.* 2012; Chopra *et al.* 2008; Mishra *et al.* 2003). We hypothesize that, like DmGAF, the functional versatility of vGAF depends on its interacting partners, which themselves are functionally diverse.

In the present study, we sought to decipher a comprehensive protein-interaction network of vGAF. We used mass spectrometry and identified 314 proteins that co-immunoprecipitate with vGAF. These proteins belong to diverse functional classes, like chromatin remodelers, transcriptional activators/repressors, RNA processing factors, and components of DNA repair machinery. We show an essential role of vGAF in DNA repair and cell-survival after UV induced DNA damage. The diversity of the vGAF-associated protein complexes observed in this study explains the molecular basis for the diverse functions of this protein, and also implicate its function in multiple other nuclear processes. We next show that vGAF is heavily down-regulated across all major stages of skin cutaneous melanoma (SCM) and can be used as a potential biomarker for SCM.

2. Materials and methods

2.1 Plasmids and antibodies

ThPOK cDNA (IMAGE ID- 6309645) was obtained from Open-Biosystems (Dharmacon). The N-terminal 3X-FLAG-tagged ThPOK expression plasmid was generated by cloning the ThPOK coding DNA sequence (CDS) in the backbone of pEGFPC-3 (Clontech) plasmid after replacing the CDS of EGFP

with 3X-FLAG. CDS of MBD3, Cbx5, HDAC1, and RBM14 were amplified from C2C12 myoblast cell cDNA using Q5[®] High-Fidelity 2X Master Mix. Expression plasmids of EGFP-tagged MBD3, Cbx5, HDAC1, and RBM14 were generated by cloning the respective CDS of proteins into the pEGFPC-1 plasmid (See supplementary table S3 for primer and restriction site details). For Western blotting, anti-FLAG M2 mouse monoclonal antibody was procured from Sigma, and anti-GFP (ab290) rabbit polyclonal was from Abcam. p-Histone H2A.X (Ser-139)/ γ -H2A.X (SC-10196) was obtained from SantaCruz. Secondary antibodies anti-mouse HRP (ab6820) and anti-rabbit HRP (ab6802) were obtained from Abcam. For immunostaining, p-Histone γ -H2A.X (Ser-139) (clone JBW301, cat. 05-363) was from Millipore while anti-FLAG rabbit polyclonal (F7425) antibody was procured from Sigma. Fluorochrome-conjugated secondary antibodies anti-rabbit IgG (H + L) Alexa-fluor 647 (cat. 711-605-152), and anti-mouse IgG (H + L) Cy3 (cat. 115-165-062) were obtained from Jacksons laboratories.

2.2 RNA isolation and cDNA preparation

Total RNA was prepared from C2C12 myoblast cells using TRIzol (Ambion, Invitrogen) following the manufacturer's instructions. One microgram of total RNA was used for preparing cDNA using Prime-Script[™] 1st strand cDNA Synthesis Kit (Clontech) following the manufacturer's protocol.

2.3 Cell culture and transfection

C2C12 myoblast cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 20% fetal bovine serum (FBS) and 1X GlutaMax (Invitrogen). HEK293 cells were maintained in DMEM supplemented with 10% FBS. C2C12/HEK293 cells were grown in 100 mm tissues culture dishes and were transfected with 12 μ g plasmid DNA using Lipofectamine-LTX (C2C12 cells) or Lipofectamine-3000 (HEK293) following the manufacture's instruction. Primary fibroblast cell cultures were established using depilated skin and lung tissue of vGAF knockout mice (The Jacksons laboratory: Stock No: 027663) and wild type sex-matched littermates using a previously published protocol (Seluanov *et al.* 2010; Egawa and Littman 2008). All cell cultures were maintained at 37°C with 5% CO₂ in a humidified chamber. All

animal experiments were approved by the Institutional Animal Ethics Committee (IAEC) of the Center for Cellular and Molecular Biology (CCMB). All animal experiments were performed in accordance to the guidelines of IAEC of CCMB. Animals were housed in CCMB transgenic mice resource facility.

2.4 Immuno-precipitation, protein-extraction and western blotting

C2C12 myoblast/HEK293 cells were harvested 36 h post-transfection and washed twice with ice-cold PBS. Cells were resuspended in cell lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.5 mM MgCl₂, 1% Triton-X 100, 10% glycerol, 1X protease inhibitor cocktail (Roche), 1 mM PMSF, 270 U/ml DNAase-I (Sigma)) and incubated on ice for 45 min with intermittent mixing. The cell debris was removed by spinning the cell lysate at 14000 rpm/4°C for 10 min. ANTI-FLAG M2 (Sigma) mouse monoclonal or anti-GFP rabbit polyclonal (ab290) antibody was incubated with Protein-G/A magnetic beads (Dynabeads, Invitrogen) in PBST for 2 h at 4°C. Subsequently, Antibody-Protein-G/A magnetic bead complex was incubated with total cell lysate at 4°C for overnight. The immune-complexes captured on beads were washed thrice with cell-lysis buffer (without DNAase-I) and eluted in the laemmli buffer.

Protein extracts were prepared from skin explants or primary cells. Skin explants were pulverized in liquid nitrogen. Pulverized skin or cells were resuspended in protein-extraction buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% Triton-X 100, 1% SDS, 10% glycerol, 10 mM DTT, 1X protease inhibitor cocktail (Roche), 1 mM PMSF) and incubated on ice for 45 min with intermittent mixing. The cell and tissue debris were removed by spinning the extracts at 14000 rpm/4°C for 10 min. Protein samples were resolved on SDS-PAGE and transferred onto PVDF membrane followed by antibody-mediated detection using enhanced chemiluminescence method.

2.5 UV treatment

UV treatments were carried out using CL-1000 UV cross-linker (UVP). Exponentially growing cultures of primary cells were exposed to UVC (254 nm) at a dose of 05 J/m² (lung cells) or 30 J/m² (skin cells). C2C12 cells were grown on glass coverslips, and 36 h after the transfection were UVC treated at a dose of 5 J/m².

Depilated skin from vGAF knockout and wild type mice was cut into 1 mm² size and was kept with their dorsal side up in 6 well tissue culture plates. Skin explants were exposed to UVC at a dose of 60 J/m². Prior to the protein extract preparation, tissues and cells were allowed to recover for 3 h in DMEM medium supplemented with 10% FBS.

2.6 MTT assay

Skin primary cells were UV treated in 24 well tissue culture plates after 1-day post-seeding at a density of 5000 cells/well. Cells were allowed to grow for another 96 h and were washed with PBS before adding 100 ul of MTT solution (0.5 mg/ml MTT in PBS) to each well. Cells were incubated with MTT solution for another 4 h before adding 150 ul of DMSO to each well for the solubilization of formazan crystals. Finally, the absorbance was measured at 570 nm using Multiskan microplate photometer (Thermo Scientific, USA).

2.7 Immunostaining and microscopy

Cells were grown over glass coverslips and fixed with 4% paraformaldehyde. Next to the fixation, cells were permeabilized in PBST (0.5% Triton-X100) and were kept in blocking solution (1X PBS, 0.1% Tween-20, 1%BSA) for 1 h. Cells were incubated with anti- γ -H2AX (1:1000) and anti-FLAG (1:1000) antibodies for 1 h at room temperature. Fluorochrome-conjugated secondary antibodies used for subsequent detection of primary antibody binding were Anti-rabbit IgG (H + L) Alexa-fluor 647 (1:1000) and anti-mouse IgG (H+L) Cy3 (1:500). Finally, coverslips were mounted on glass slides using VectashieldTM mounting media (with DAPI).

The optical sections of immunostained cell nuclei were captured using Zeiss LSM 880 confocal microscope with 63X objective at 3X zoom. Single plane optical sections of 5 μ m thickness were used for colocalization analysis using the 'Colocalization Finder' plugin of ImageJ (<http://rsb.info.nih.gov/ij/plugins/colocalization-finder.html>) (Schindelin *et al.* 2012). We limited our analysis to only those pixels with a minimum 70% intensity ratio to get rid of irrelevant low-intensity pixels (Dunn *et al.* 2011). Pearson correlation coefficient and Overlap coefficient are reported as calculated by 'Colocalization Finder' alongside the overlay image.

2.8 Mass spectrometry and protein identification

Immuno-precipitated samples were fractionated by 12% SDS-PAGE. The gels were stained with Imperial™ protein stain (Thermo-Scientific), destained, and washed with Milli-Q water several times. Each gel lane was sliced into six pieces, and these gel-pieces were trypsin digested using a previously published protocol (Kallappagoudar *et al.* 2010). The resulting tryptic peptides were resolved using reverse phase chromatography. 18 ul of the sample was loaded on reverse phase C18-Biobasic PicoFrit™ chromatography columns using Easy nLC-II (Thermo-Scientific) with a flow rate of 0.4 ul/min; peptides were eluted on 60 min gradient (3% to 70% acetonitrile). Chromatographically separated peptides were analyzed using Q-exactive mass spectrometer. The top 10 peptide precursors were selected for MS/MS analysis. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD012890 (Perez-Riverol *et al.* 2019). The mass spectra obtained were searched against mouse proteome from UniProt using the SEQUEST HT algorithm incorporated in Thermo Proteome Discoverer (version 1.4.0.288). Enzyme specificity was set to full trypsin digestion with maximum of 2 missed cleavages. Precursor mass tolerance was set to 10 ppm, and fragment mass tolerance was 0.6 Da. Peptide identifications were accepted only if they pass the following criteria: Maximum Delta Cn value—0.1, Minimum Xcorr value—1.92, peptide confidence—High. Proteins identifications were accepted only if they cross the Percolator score—10. We identified a non-redundant and cumulative list of 582 proteins from three FLAG-vGAF pull-downs considering protein, which were identified in at least one of the replicate with $PSM \geq 3$. We used a less stringent criterion for two negative control pull-downs and considered proteins that were identified in at least one of the replicate with $PSM > 1$, yielding a total of 935 non-redundant proteins. We removed all the proteins that were common between vGAF pull-down and negative control pull down to identify 314 interactors of vGAF.

2.9 Gene ontology analysis of protein interactome

The list of protein interactors was analyzed using Gene ontology tool Panther (pantherdb.org) (Thomas *et al.* 2003). In order to identify the overrepresented Gene Ontology (GO) terms, we performed a statistical

overrepresentation test. Annotation datasets used were ‘GO molecular function complete’, ‘GO Biological process complete,’ and ‘GO Cellular compartment complete’. Fisher’s exact test was used for statistical significance using Bonferroni correction for multiple testing. GO terms were considered significant only at $p \leq 0.05$. The resultant significant terms that were overrepresented more than four folds were visualized in a bar chart.

The interaction network of DNA repair protein was constructed using publically available protein-protein interaction database STRING (string.db.org). In brief, DNA repair proteins from the vGAF protein interactome were submitted to STRING. Under the settings section, active interaction source was selected as ‘Experiment’ and the minimum required interaction score was set to ‘medium confidence (0.4), before generating the final interaction map.

2.10 Bioinformatic analysis of gene expression data

We used the Expression-DIY module of GEPIA (Gene Expression Profiling Interactive Analysis) to analyze RNA-seq data for skin cutaneous melanoma (SCM) and normal skin tissue (NST) from TCGA (The Cancer Genome Atlas) and GTEx (Genotype-Tissue Expression) databases. Box-plot function was used for plotting the differential expression of genes in SCM and NST samples using default settings. We used the stage-plot function for plotting the expression of vGAF across major stages of skin cutaneous melanoma using the default setting.

3. Result

3.1 Identification of vGAF associated protein interactome

To define the protein interactome of vGAF, we used immuno-affinity based purification of protein complexes. We subsequently identified the proteins using liquid chromatography coupled with mass-spectrometry (LC-MS/MS). We tagged vGAF with an N-terminal 3X FLAG-tag and expressed the recombinant protein in C2C12 myoblast cells. Immuno-staining of FLAG-vGAF transfected cells with anti-FLAG antibody shows a very specific nuclear localization of FLAG-vGAF (supplementary figure S1). The whole-cell lysate was used to immuno-precipitate the vGAF containing

protein complexes using a monoclonal anti-FLAG antibody or a non-specific IgG (figure 1A). Whole-cell lysate prepared from empty-vector transfected C2C12 cells was also used for immunoprecipitation using the

anti-FLAG antibody to serve as the negative control in the mass-spectrometry experiment. Immuno-precipitation experiments were done in triplicates (FLAG-vGAF transfected cells) or in duplicates (empty-vector

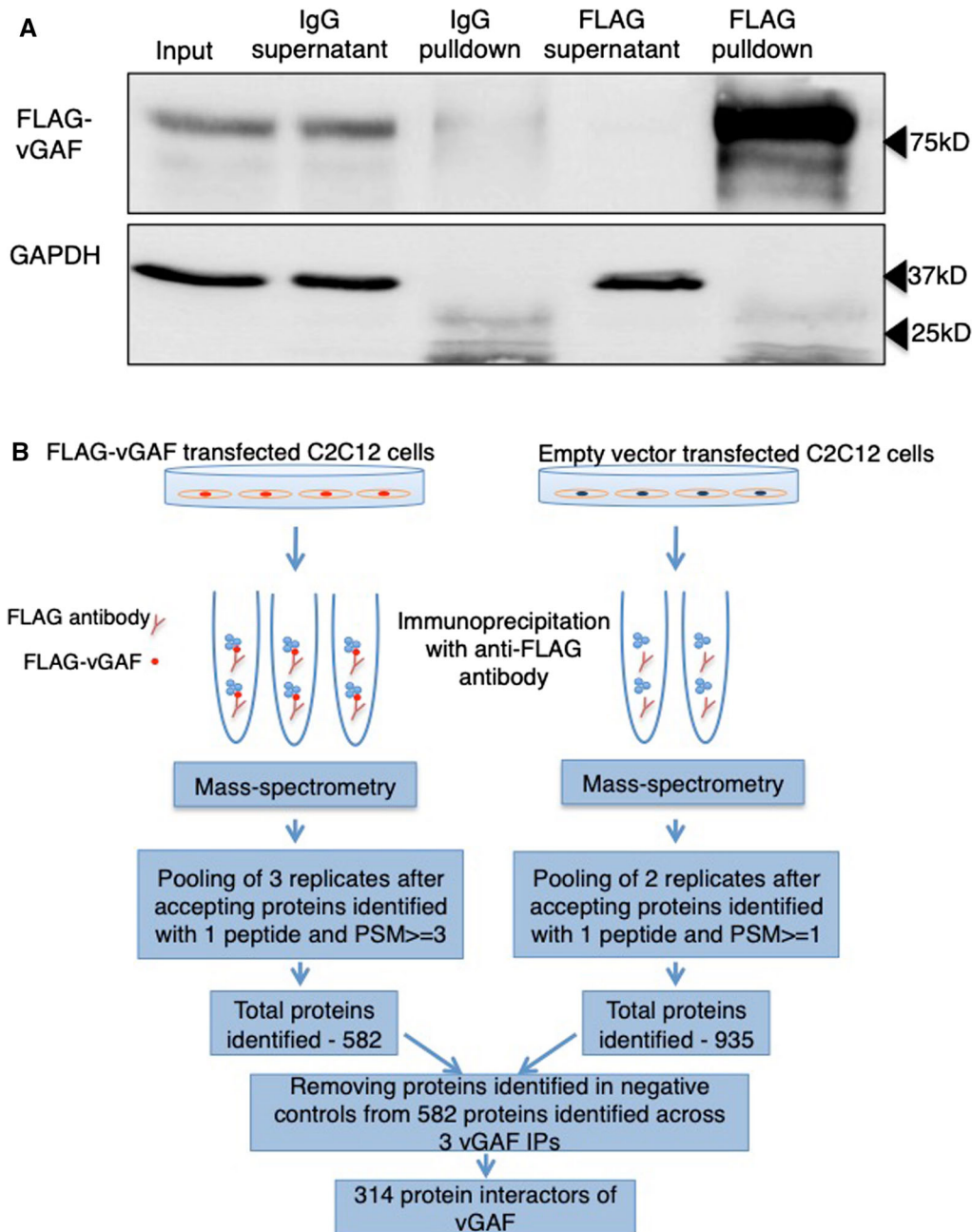


Figure 1. Immuno-precipitation of FLAG-vGAF. (A) Western blot analysis of immunoprecipitated samples. FLAG antibody was used to assess the efficiency of the immuno-pull-downs. Blots were probed with GAPDH antibody to check for non-specificity. Lane 1—total cell extract (Input), Lane 2—unbound fraction (supernatant) after immuno-pull-down using IgG, Lane 3—Immuno-pull-down using IgG, Lane 4—unbound fraction (supernatant) after immuno-pull-down using FLAG antibody, Lane 5—immuno-pull-down using FLAG antibody. (B) The workflow of IP-MS experiment. Protein extracts from FLAG-vGAF or mock transfected cells were immuno-precipitated using anti-FLAG antibody. Proteins identified with at least 1 peptide and $PSM \geq 3$ were pooled from all three FLAG-vGAF pull-down experiments. Proteins identified with at least 1 peptide and $PSM \geq 1$ were pooled from two duplicate control pull-down experiments. 314 protein interactors of vGAF were listed after proteins identified in negative controls were removed from the 582 proteins identified across 3 vGAF IPs.

transfected cells). Finally, the immuno-precipitated protein samples were analyzed using LC-MS/MS (figure 1B). Altogether, we could identify 314 proteins that specifically interact with vGAF across our three replicates (supplementary table S1 and S2). Although we have used the total cell extracts for all our experiments, a cellular component ontology for these 314 proteins showed a statistical overrepresentation of nuclear proteins, suggesting the specificity of FLAG-vGAF immuno-precipitation (supplementary figure S2 and supplementary table S3). We further validated our mass spectrometry data by reverse Co-IPs followed by western blots. To this end, we selected four candidate proteins from the list of vGAF protein interactome like HDAC1, MBD3, (both are the parts of NuRD chromatin remodeling complex) (Clapier and Cairns 2009), RBM14 (a protein that is involved in both DNA double-strand break repair and transcription-coupled alternative splicing) (Simon *et al.* 2017; Auboeuf *et al.* 2004), CBX5 (Protein associated with heterochromatin) (Maeng *et al.* 2015). These proteins were N-terminally GFP-tagged in pEGFPC-1 plasmid and were co-transfected with 3X-FLAG-tagged vGAF plasmid in HEK293 cells (supplementary figure S3A, B). We chose HEK293 cells for co-transfection experiments owing to the high transfection efficiency of these cells. Immuno-precipitation of GFP-tagged candidate proteins from co-transfected HEK293 cells shows enrichment of vGAF compared to IgG controls suggesting an *in-vivo* interaction between vGAF and candidate proteins (figure 2A, B). These results validate the efficacy of our experiment and the identification of novel interacting partners of vGAF.

3.2 Gene ontology analysis of vGAF protein interactome

Gene ontology analysis of the vGAF protein interactome enabled us to identify the functional diversity among vGAF interacting partners. Figure 3 shows a bar plot of significantly enriched ‘Molecular function’ and ‘Biological process’ gene ontology (GO) terms that are at least 4 fold overrepresented in vGAF interactome. The same results with additional GO terms and details are presented in supplementary table 3. The most overrepresented Biological process associated GO terms (GOBP) in our analysis are ‘double-strand break repair via nonhomologous end joining’ and ‘non-recombinational repair’, suggesting a novel interaction of vGAF with components of DNA repair machinery. Another set of terms like ‘transcription by RNA polymerase II’ and

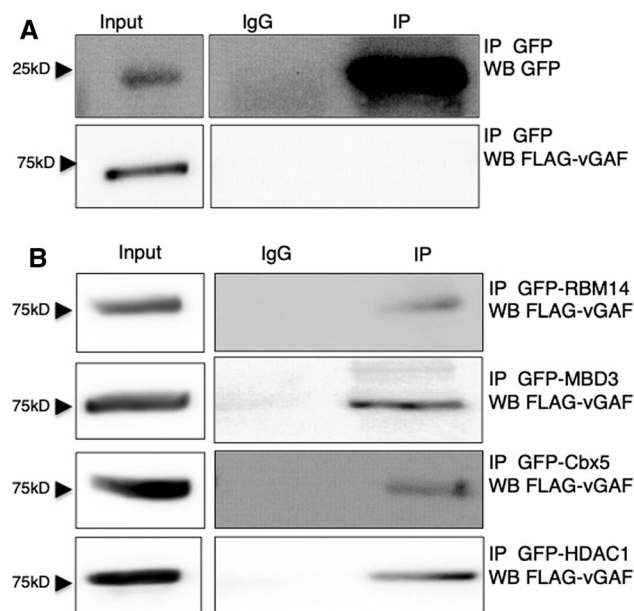
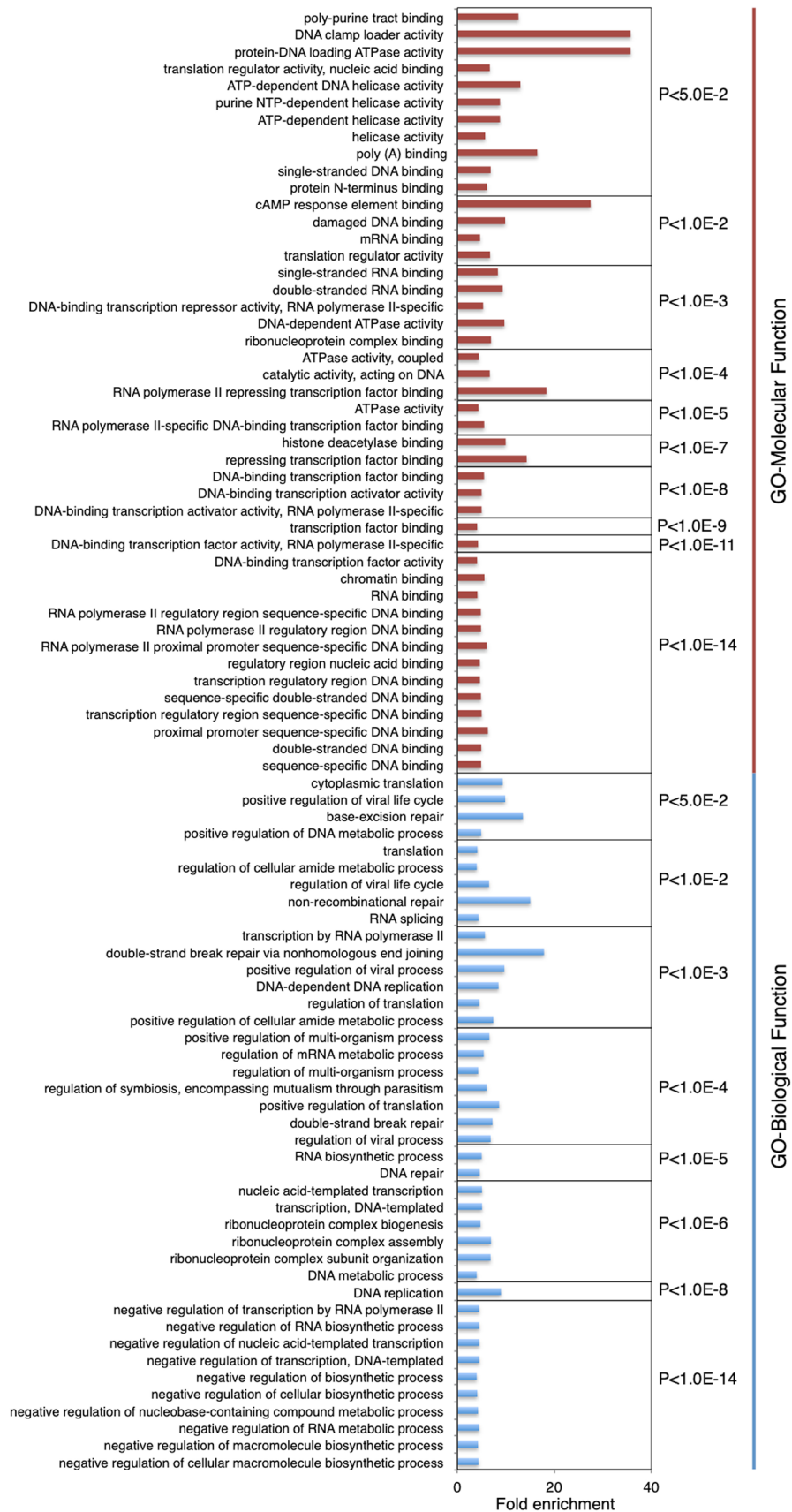


Figure 2. Validation of IP proteins with reverse Co-IP. (A) FLAG-vGAF protein does not interact with EGFP. Immuno-precipitation (IP) was done using anti-GFP antibody from HEK293 cells co-transfected with FLAG-vGAF and EGFP expression plasmids. Western blots were done with anti-GFP antibody or anti-FLAG antibody. (B) FLAG-vGAF immuno-precipitates with N-terminal GFP tagged RBM14, MBD3, Cbx5 and HDAC1. HEK293 cells were co-transfected with FLAG-vGAF and one of the N-terminal GFP tagged proteins (RBM14, MBD3, Cbx5, HDAC1) constructs. IP was done with anti-GFP antibody and subsequently, anti-FLAG antibody was used to detect the presence of FLAG-vGAF in immuno-precipitate using western blot (WB).

‘proximal promoter sequence-specific DNA binding’ delineate the association of vGAF with RNA-polymerase-II associated macromolecular transcriptional complexes. Our analysis suggests the interaction of vGAF with components of RNA metabolic machinery through the presence of GO terms like ‘RNA metabolic process’ and ‘mRNA splicing, via spliceosome.’ Collectively, GO analysis of vGAF protein interactome clearly shows that vGAF interacts with protein complexes of diverse functionalities and nature.

3.3 GAF associates with ATP dependent chromatin remodelers

ATP dependent chromatin remodelers (ADCRs) are critical players in the regulation of chromatin packaging and thereby essential for various DNA dependent nuclear functions (Bartholomew 2014). There are four different families of chromatin remodeling complexes



◀ **Figure 3.** Analysis of statistically significantly gene ontology (GO) terms enriched in vGAF protein interactome. Panther Gene ontology analysis (Molecular function and Biological process) was performed for 314 proteins identified in vGAF protein interactome. The X-axis represents the fold enrichment. Bar plots for GO categories showing similar p-values are placed together on Y-axis, and p-values are indicated next to the rectangle encircling the bars.

like SWI/SNF, ISWI, CHD, and INO80 (Clapier and Cairns 2009). All these complexes have an ATPase subunit and several non-catalytic subunits, which further diversify these complexes into sub-families (Clapier and Cairns 2009). In our analysis, we identify all the components of the NuRD complex that belongs to the CHD family of ADCRs (supplementary table S4). We identify Mi-2beta/CHD4 that is the ATPase subunit of this complex, along with that, we also identify all other non-catalytic subunits like MBD3, HDAC1, HDAC2, RbAp46, and p66-alpha/beta. Similarly, we identify both the components of the ACF subfamily of ISWI complexes, like SNF2H (ATPase subunit) and BAZ1A (non-catalytic subunit) (Clapier and Cairns 2009). On the contrary, we do not identify most of the components of SWI/SNF and INO80 complexes, other than BAF53a (SWI/SNF and INO80) & RUVBL1 (INO80) (Clapier and Cairns 2009). These results suggest that vGAF interacts with whole ACF and NuRD remodeling complexes.

3.4 vGAF interacts with a spectrum of DNA repair proteins and is essential for efficient DNA repair and cell survival after UV induced DNA damage

Analysis of vGAF protein interactome shows the presence of several DNA repair proteins that are involved in DNA break repair, including Ku60 and Ku80, suggesting a physical association between vGAF and DNA repair machinery (figure 4 and supplementary table S5). The phosphorylated form of histone variant H2AX (γ -H2AX) quickly accumulates at the site of double-strand DNA breaks within minutes after the induction of DNA damage (Kuo and Yang 2008). It has extensively been used both as a spatial and quantitative marker for DNA damage (Kuo and Yang 2008). To understand the functional consequence of physical association between vGAF and DNA repair proteins, we used tissues and cells derived from vGAF knock out (KO) mice and asked if vGAF could affect the process of DNA-damage repair. We measured the levels of γ -H2AX using western blotting and used it as

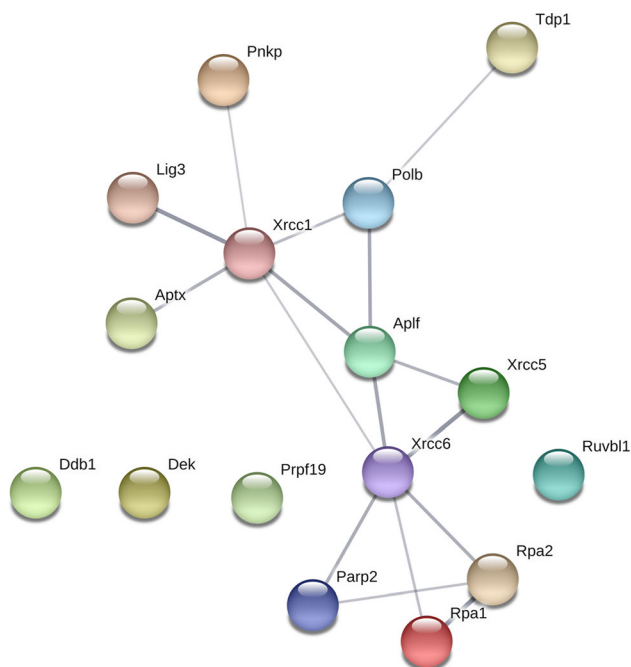


Figure 4. Interaction network of DNA repair protein identified in vGAF protein interactome. Protein-protein interaction network of DNA repair proteins identified in vGAF protein interactome was generated using STRING DB. Nodes represent the DNA repair proteins. Source of interaction from the database was selected as ‘Experiment’ and edges connecting the nodes represent the protein-protein association. Thickness of the edge measures the strength of the publicly available data supporting the association.

a proxy for the level of DNA damage in the tissues or cells used for protein extract preparation. vGAF has been shown to express predominantly in adult mouse skin (Galera *et al.* 1994). Protein extracts prepared 3-h after the UV treatment of the skin explants shows a higher accumulation of γ -H2AX in vGAF KO skin compared to the wild-type controls, suggesting a delayed repair of DNA-damage in KO skin tissue compared to wild-type tissue (figure 5A). We further checked the survival of skin primary fibroblast 4-days after the UV treatment using MTT assay. Here again, we see that primary fibroblast cells from vGAF KO skin are more sensitive to UV treatment compared to the control (figure 5B). To further confirm our results, we derived primary fibroblast cells from the lung that is another organ where vGAF show a significant expression (Michaloski *et al.* 2011). Protein extracts prepared 3-h after UV treatment of primary lung fibroblasts from KO mice show an increased signal of γ -H2AX in western blots compared to the protein extract from the wild-type lung primary fibroblast cells (figure 5C). To determine whether vGAF localizes to

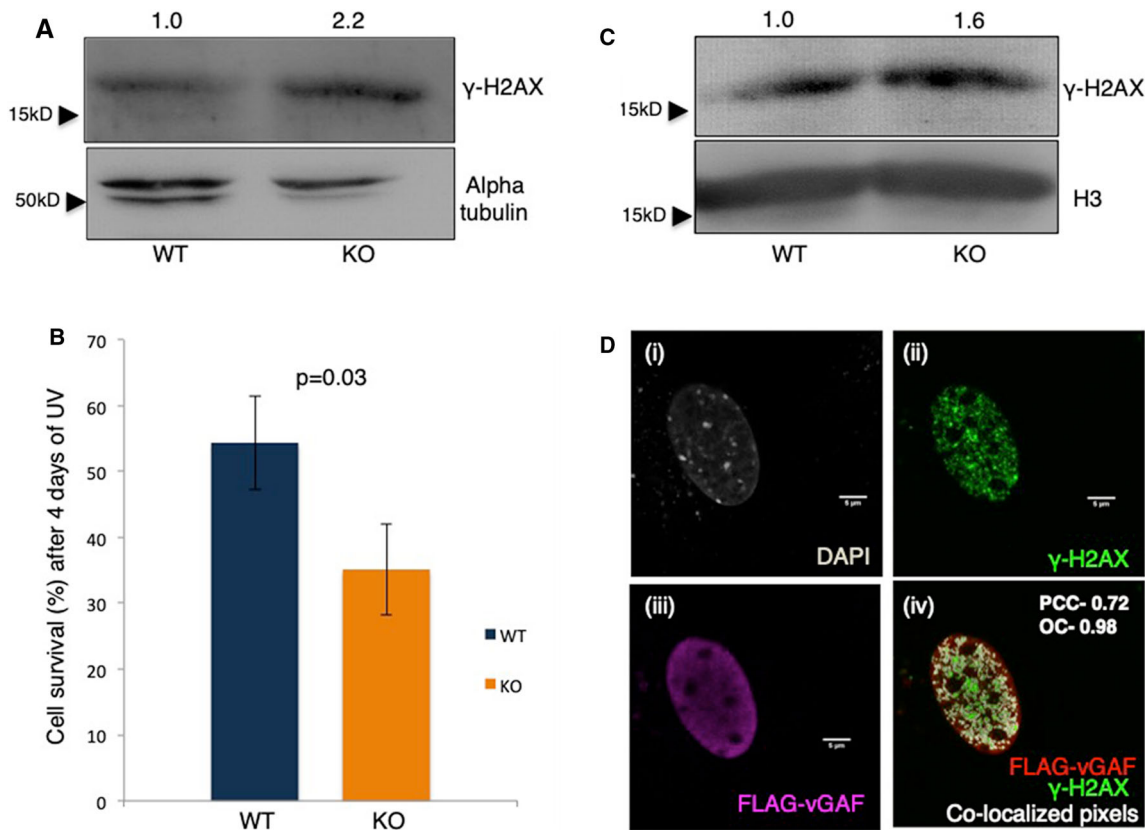


Figure 5. vGAF is essential for efficient DNA repair and cell survival after UV induced DNA damage. (A) vGAF KO mice skin explants show higher levels of DNA damage (increased γ -H2A.X) after UV induced DNA damage. Skin explants were treated with UVC (60 J/m^2) and were allowed to recover for 3 h. An increase in γ -H2A.X was detected in protein extracts prepared from vGAF KO mice skin explants. (B) Skin primary fibroblast cells derived from vGAF KO mice are sensitive to UVC treatment. Skin primary fibroblast cells were treated with UVC (30 J/m^2). Cell survival assay shows a significant decrease in the survival of primary fibroblasts derived from the skin of KO mice. (C) Lung primary fibroblasts derived from vGAF KO mice show higher levels of DNA damage (increase in γ -H2A.X) after UV induced DNA damage. Lung primary fibroblasts were treated with UVC (5 J/m^2) and were allowed to recover for 3 h. Increased γ -H2A.X was detected in protein extracts prepared from cells derived from vGAF KO. (D) C2C12 myoblast cells show co-localization of FLAG-vGAF with γ -H2A.X. C2C12 myoblast cells were transfected with FLAG-vGAF, and 36 h post-transfection cells were treated with UVC (5 J/m^2) and allowed to recover for 1 h. Cells were co-immunostained with anti- γ -H2A.X and anti-FLAG antibodies. Single optical section from the center of the nucleus show the staining of (i) DAPI (grey), (ii) FLAG-vGAF (magenta) (iii) γ -H2A.X (green) and (iv) co-localization of FLAG-vGAF and γ -H2A.X (white pixels). ImageJ was used to identify and highlight pixels where FLAG-vGAF and γ -H2A.X staining signals were present together. Scale bars are 5 μm .

the sites of DNA double-strand breaks, we performed immunostaining of γ -H2AX 1- h after the UV treatment of C2C12 myoblast cells expressing FLAG-vGAF. We chose C2C12 myoblast cells for this experiment as these cells were originally used to decipher the vGAF protein interactome that is enriched in DNA repair protein. As shown in figure 5D, co-immunostaining of γ -H2AX (green) and FLAG-vGAF (magenta) shows several regions where these two proteins co-localize (white pixels) with each other, suggesting an association of vGAF with doubled strand DNA break lesions. Collectively, these results show that vGAF is essential for efficient DNA repair after

UVC induced DNA damage. Furthermore, these results serve as a biological validation of our IP mass-spec data that suggest a physical association between vGAF and DNA repair machinery.

3.5 Expression of vGAF is downregulated in skin cancer tissue

UV induced DNA damage is the primary cause of skin cancers, and that prompted us to compare the expression data of vGAF (ZBTB7B) and DNA repair proteins that interact with it (supplementary table S5), in skin

cutaneous melanoma (SCM) and normal skin tissue (NST). GEPIA (Gene Expression Profiling Interactive Analysis) is a web-based tool for the analysis of TCGA (The Cancer Genome Atlas) and GTEx (Genotype-Tissue Expression) database (Tang *et al.* 2017). We analyzed gene expression data from these two databases using GEPIA and show that SCM tissue samples express vGAF at a much lower level than the NST samples (figure 6A). In contrast, other DNA repair proteins that interact with vGAF (figure 4 and supplementary table S5) are either upregulated or show no significant difference in expression in SCM vs. NST (supplementary figure S4). Further analysis also shows that vGAF expression does not change significantly across all the stages of SCM, suggesting a consistent downregulation of vGAF even in the stage-0 of SCM (figure 6B). Collectively, these results suggest that vGAF is heavily downregulated in SCM even at stage-0, and thus low levels of vGAF in skin biopsies can possibly be used as a diagnostic biomarker for SCM.

4. Discussion

Gene-knockout mouse model and extensive cell culture studies have revealed functions of vGAF in a variety of biological processes like development of CD4+ T-cells, brown-fat development, lactation and

thermogenesis (Zhang *et al.* 2018; Li *et al.* 2017; Egawa and Littman 2008; Muroi *et al.* 2008). Intriguingly, the molecular functions of vGAF that form the basis for these biological processes are confounding in nature. Studies suggest that vGAF is a repressor of *cd8* gene in CD4+ T-cells while vGAF forms a ribonucleoprotein complex with lincRNA to activate thermogenic gene expression program in adipocytes (Li *et al.* 2017; Rui *et al.* 2012). Moreover, several cell-culture based reports on vGAF indicate an even a broader array of molecular activities associated with vGAF like nuclear lamina mediated repression of essential developmental genes and enhancer-blocking activity in mammalian *Hox* clusters (Srivastava *et al.* 2013; Zullo *et al.* 2012). Studies in both *Drosophila melanogaster* and mammals have suggested the importance of protein-protein interactions in the molecular processes associated with GAF proteins (Lomaev *et al.* 2017; Beauchef *et al.* 2012; Rui *et al.* 2012; Zullo *et al.* 2012; Chopra *et al.* 2008; Kypriotou *et al.* 2007; Mishra *et al.* 2003; Widom *et al.* 2001). Here, we have identified a comprehensive protein interactome of vGAF to understand the diversity of vGAF protein-interacting complexes.

We used a high-throughput IP-mass spec approach to identify 314 interacting partners of vGAF across three replicates. Many of these proteins could be only be identified in one of the replicates. A possible reason for

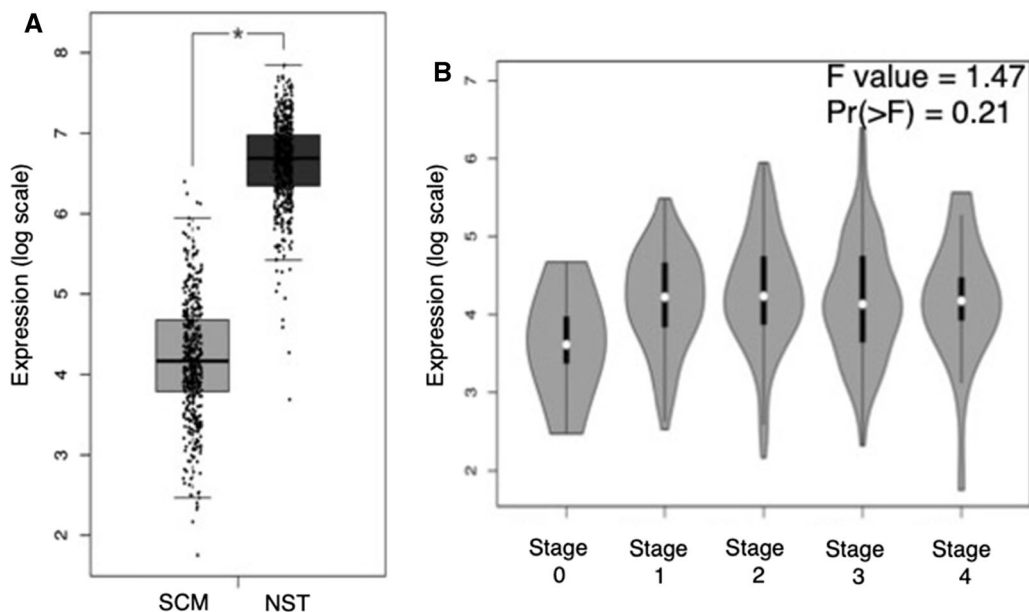


Figure 6. Analysis of vGAF expression in skin cutaneous melanoma (SCM) samples from human subjects. (A) vGAF expression is downregulated in SCM samples compared to normal skin tissue (NST). GEPIA is used to analyze the expression data from 461 SCM and 558 NST samples. The analysis shows a significant decrease in vGAF expression in SCM ($p \leq 0.05$). (B) vGAF expression levels does not change significantly across different stages of SCM. Violin plots of vGAF expression data across pathological stages of SCM shows a consistent downregulation, including stage 0 of SCM.

this could be the transient or labile nature of vGAF containing protein complexes. Our High throughput approach could result in several non-specific hits. To ensure that the proteins identified in only one of the replicates are not contaminants, we validated the interaction of a few candidate proteins, including Cbx5 (identified only in replicate 1) with vGAF through reverse Co-IP followed by western blot. The specificity of the proteins identified in our interactome is further indicated by the gene-ontology cellular compartment (GOCC) analysis. We have used a monoclonal anti-FLAG antibody for both vGAF and negative control pull-down experiments. GO cellular compartment analysis of proteins identified in negative control experiments show that most the significant terms ('myelin sheath', 'ribonucleoprotein complex' and 'cytosolic part') represent cytosolic or membrane localization of these proteins (supplementary figure S2). On the other hand, vGAF is a nuclear protein, and GOCC analysis of proteins identified in the vGAF pull-down experiment shows the overrepresentation of proteins localized in chromosome/chromatin (supplementary figure S2).

The biological significance of the vGAF protein interactome is exemplified by the presence of functional classes of proteins that mediate the biological processes where vGAF is already known to be involved. We identify several proteins that are involved in RNA metabolic processes like alternative splicing (AS) & mRNA maturation (figure 3 and supplementary table S3). The functional significance of this association between vGAF and components of AS is exemplified by a high-throughput molecular screen which identifies vGAF as one of the factors that affect developmentally important AS events (Han *et al.* 2017). Gene ontology analysis of vGAF protein interactome also suggests the interaction of vGAF with proteins involved in the regulation of viral life cycle. A recent report shows that knockdown of vGAF in A549 cells results in almost 50% reduction of Influenza A virus (IAV) titers 24 h post-infection (hpi) compared to the mock treatment, suggesting the functional relevance of the proteins identified in our analysis (Chen *et al.* 2017). Existing literature on vGAF shows that it binds on the promoter of genes like *UDP glucose dehydrogenase*, *omesodermin*, *Coll1a1*, *Socs1* and *Cish* and regulates the gene expression (Luckey *et al.* 2014; Li *et al.* 2013; Kyriotou *et al.* 2007; Beauchef *et al.* 2005). In our study, we also identify a number of transcription factors and components of RNA-polymerase macromolecular complexes. The interaction of vGAF with these protein complexes can modulate the expression of genes where it binds on promoters.

Our analysis shows an association of vGAF with whole ACF and NuRD chromatin remodeling complexes. Apart from their gene regulatory functions, both ACF and NuRD complexes are involved in DNA replication and repair (Aydin *et al.* 2014; Smeenk *et al.* 2010). We could not identify most of the components of SWI/SNF and INO80 except BAF53a and RUVBL1. It might suggest that vGAF does not interact with whole SWI/SNF and INO80 chromatin remodeling complexes, and possibly BAF53a and RUVBL1 are also part of protein complexes other than the canonical chromatin remodeling complexes. In our analysis, the DNA repair associated functions of ACF and NuRD chromatin remodeling complexes are of vital importance as we also identify a number of other DNA repair proteins that interact with vGAF. We identify both Ku60 and Ku80, the two major proteins that form a complex with DNA ends immediately after a double-strand DNA break. We show that vGAF associates with several γ -H2AX foci 1 h after UV induced DNA damage. In these experiments, we have taken only one-time point, and more detailed experiments are required in the future to better understand the temporal dynamics of the association between vGAF and damaged DNA lesions. Our analysis of vGAF protein interactome shows the presence of several cNHEJ (classical non-homologous end joining) accessory proteins (PNKP, Aprataxin, APLF) and ligase3, a component of alternative NHEJ (Lieber 2010). We also identify the components of base excision DNA repair, like XRCC1, Polb, and ligase3 (Giglia-Mari *et al.* 2011). These results show a physical association of vGAF with components of DNA repair machinery, The functional aspects of this physical association between vGAF and DNA repair machinery is further supported by our results, which show that vGAF is necessary for efficient DNA repair and cell survival after DNA damage. The present study does not explore the specific DNA repair pathways where vGAF is functionally implicated, and the exact molecular role of vGAF in the DNA double-strand break repair. The presence of several NHEJ proteins and absence of proteins involved in homologous recombination DNA repair possibly suggest the role of vGAF in NHEJ DNA repair pathways.

The role of vGAF in the DNA repair process is novel considering the established functions of this protein as a zinc-finger transcription factor with sequence-specific DNA binding ability (Srivastava *et al.* 2018). However, we have recently started understanding the role of sequence-specific zinc-finger transcription factors in the maintenance of genome integrity (Vilas *et al.* 2018).

ZBTB7A/LRF, a paralogue of vGAF in mammals, has recently been shown to be involved in cNHEJ DNA repair pathway (Liu *et al.* 2015). LRF is a reported interacting partner of vGAF and it is present in our vGAF protein interaction data as well (Widom *et al.* 2001). DNA damage and repair processes have closely been linked to tumorigenesis (Halazonetis *et al.* 2008). The UV-induced DNA damage is the major cause of tumor formation in skin tissue where vGAF also expresses predominantly (Armstrong and Krickler 2001). Our results show that vGAF is heavily downregulated in SCM and its expression does not change significantly across major stages of melanoma progression, suggesting downregulation even in stage 0 of SCM. The similar analysis done with other DNA repair proteins does not show a significant change in expression levels across NST and SCM, except RUVBL1. Our study reports a link between SCM and the downregulation of vGAF. We do not yet understand if the low levels of vGAF are causally linked to the occurrence of SCM and more detailed experiments are required to better understand the relationship between vGAF downregulation, DNA double-strand break repair and tumorigenesis.

Recent studies have established that DNA repair is closely linked with two other major nuclear processes like transcription and alternative splicing (Naro *et al.* 2015; Hanawalt and Spivak 2008). Noticeably, our vGAF protein interactome data includes several proteins, which are uniquely or redundantly involved in these three processes. For example, vGAF interacts with factors (like RBM14, RBMX, and PRPF19) that are involved in both splicing/RNA processing and DNA damage repair (Simon *et al.* 2017; Mahajan 2016; Adamson *et al.* 2012; Song *et al.* 2010; Heinrich *et al.* 2009; Auboeuf *et al.* 2004). Likewise, our analysis shows that vGAF associates with transcription factors which are involved in DNA damage repair (LRF, ZNF281) and RNA processing (LRF, YY1) (Han *et al.* 2017; Pieraccioni *et al.* 2016; Wai *et al.* 2016; Liu *et al.* 2015; Sigova *et al.* 2015; Bielli *et al.* 2014). Altogether, these observations suggest that vGAF can influence all three major nuclear processes (Transcription, Splicing and DNA repair) and stands a position in nuclear-proteome-wide interaction network where molecular circuits of transcription, splicing and DNA repair criss-cross each other.

Acknowledgements

AS acknowledge GATE-SRF fellowship from CSIR. The authors thank CCMB-Proteomics facility members

for performing the mass-spectrometry. Research in RKM lab is supported by CSIR and DBT, India.

References

- Adamson B, Smogorzewska A, Sigoillot FD, King RW and Elledge SJ 2012 A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nat. Cell Biol.* **14** 318–328
- Armstrong BK and Krickler A 2001 The epidemiology of UV induced skin cancer. *J. Photochem. Photobiol. B* **63** 8–18
- Auboeuf D, Dowhan DH, Li X, Larkin K, Ko L, Berget SM and O'Malley BW 2004 CoAA, a nuclear receptor coactivator protein at the interface of transcriptional coactivation and RNA splicing. *Mol. Cell Biol.* **24** 442–453
- Aydin OZ, Vermeulen W and Lans H 2014 ISWI chromatin remodeling complexes in the DNA damage response. *Cell Cycle* **13** 3016–3025
- Bartholomew B 2014 Regulating the chromatin landscape: structural and mechanistic perspectives. *Annu. Rev. Biochem.* **83** 671–696
- Beauchef G, Bigot N, Kypriotou M, Renard E, Poree B, Widom R, Domp martin-Blanchere A, Oddos T, Maquart FX, Demoor M, *et al.* 2012 The p65 subunit of NF-kappaB inhibits COL1A1 gene transcription in human dermal and scleroderma fibroblasts through its recruitment on promoter by protein interaction with transcriptional activators (c-Krox, Sp1, and Sp3). *J. Biol. Chem.* **287** 3462–3478
- Beauchef G, Kypriotou M, Chadjichristos C, Widom RL, Poree B, Renard E, Moslemi S, Wegrowski Y, Maquart FX, Pujol JP, *et al.* 2005 c-Krox down-regulates the expression of UDP-glucose dehydrogenase in chondrocytes. *Biochem. Biophys. Res. Commun.* **333** 1123–1131
- Bhat KM, Farkas G, Karch F, Gyurkovics H, Gausz J and Schedl P 1996 The GAGA factor is required in the early *Drosophila* embryo not only for transcriptional regulation but also for nuclear division. *Development* **122** 1113–1124
- Bielli P, Busa R, Di Stasi SM, Munoz MJ, Botti F, Kornblihtt AR and Sette C 2014 The transcription factor FBI-1 inhibits SAM68-mediated BCL-X alternative splicing and apoptosis. *EMBO Rep.* **15** 419–427
- Biggin MD and Tjian R 1988 Transcription factors that activate the Ultrabithorax promoter in developmentally staged extracts. *Cell* **53** 699–711
- Chen SC, Jeng KS and Lai MMC 2017 Zinc finger-containing cellular transcription corepressor ZBTB25 promotes influenza virus RNA transcription and is a target for zinc ejector drugs. *J. Virol.* <https://doi.org/10.1128/jvi.00842-17>

- Chopra VS, Srinivasan A, Kumar RP, Mishra K, Basquin D, Docquier M, Seum C, Pauli D and Mishra RK 2008 Transcriptional activation by GAGA factor is through its direct interaction with dmTAF3. *Dev. Biol.* **317** 660–670
- Clapier CR and Cairns BR 2009 The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78** 273–304
- Dunn KW, Kamocka MM and McDonald JH 2011 A practical guide to evaluating colocalization in biological microscopy. *Am. J. Physiol. Cell Physiol.* **300** C723–C742
- Egawa T and Littman DR 2008 ThPOK acts late in specification of the helper T cell lineage and suppresses Runx-mediated commitment to the cytotoxic T cell lineage. *Nat. Immunol.* **9** 1131–1139
- Farkas G, Gausz J, Galloni M, Reuter G, Gyurkovics H and Karch F 1994 The Trithorax-like gene encodes the Drosophila GAGA factor. *Nature* **371** 806–808
- Galera P, Musso M, Ducy P and Karsenty G 1994 c-Krox, a transcriptional regulator of type I collagen gene expression, is preferentially expressed in skin. *Proc. Natl. Acad. Sci. USA* **91** 9372–9376
- Galera P, Park RW, Ducy P, Mattei MG and Karsenty G 1996 c-Krox binds to several sites in the promoter of both mouse type I collagen genes. Structure/function study and developmental expression analysis. *J. Biol. Chem.* **271** 21331–21339
- Giglia-Mari G, Zotter A and Vermeulen W 2011 DNA damage response. *Cold Spring Harb. Perspect. Biol.* **3** a000745
- Halazonetis TD, Gorgoulis VG and Bartek J 2008 An oncogene-induced DNA damage model for cancer development. *Science* **319** 1352–1355
- Han H, Braunschweig U, Gonatopoulos-Pournatzis T, Weatheritt RJ, Hirsch CL, Ha KCH, Radovani E, Nabeel-Shah S, Sterne-Weiler T, Wang J, et al. 2017 Multilayered control of alternative splicing regulatory networks by transcription factors. *Mol. Cell* **65** 539–553 e7
- Hanawalt PC and Spivak G 2008 Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* **9** 958–970
- Heinrich B, Zhang Z, Raitskin O, Hiller M, Benderska N, Hartmann AM, Bracco L, Elliott D, Ben-Ari S, Soreq H, et al. 2009 Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regions in pre-mRNA. *J. Biol. Chem.* **284** 14303–14315
- Kallappagoudar S, Varma P, Pathak RU, Senthilkumar R and Mishra RK 2010 Nuclear matrix proteome analysis of *Drosophila melanogaster*. *Mol. Cell Proteomics* **9** 2005–2018
- Kuo LJ and Yang LX 2008 Gamma-H2AX - a novel biomarker for DNA double-strand breaks. *In Vivo* **22** 305–309
- Kypriotou M, Beauchef G, Chadjichristos C, Widom R, Renard E, Jimenez SA, Korn J, Maquart FX, Oddos T, Von Stetten O, et al. 2007 Human collagen Krox up-regulates type I collagen expression in normal and scleroderma fibroblasts through interaction with Sp1 and Sp3 transcription factors. *J. Biol. Chem.* **282** 32000–32014
- Li S, Mi L, Yu L, Yu Q, Liu T, Wang GX, Zhao XY, Wu J and Lin JD 2017 Zbtb7b engages the long noncoding RNA Blnc1 to drive brown and beige fat development and thermogenesis. *Proc. Natl. Acad. Sci. USA* **114** E7111–E7120
- Li Y, Tsun A, Gao Z, Han Z, Gao Y, Li Z, Lin F, Wang Y, Wei G, Yao Z, et al. 2013 60-kDa Tat-interactive protein (TIP60) positively regulates Th-inducing POK (ThPOK)-mediated repression of eomesodermin in human CD4+ T cells. *J. Biol. Chem.* **288** 15537–15546
- Lieber MR 2010 The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79** 181–211
- Liu XS, Chandramouly G, Rass E, Guan Y, Wang G, Hobbs RM, Rajendran A, Xie A, Shah JV, Davis AJ, et al. 2015 LRF maintains genome integrity by regulating the non-homologous end joining pathway of DNA repair. *Nat. Commun.* **6** 8325
- Lomaev D, Mikhailova A, Erokhin M, Shaposhnikov AV, Moresco JJ, Blokhina T, Wolle D, Aoki T, Ryabykh V, Yates JR, 3rd, et al. 2017 The GAGA factor regulatory network: Identification of GAGA factor associated proteins. *PLoS ONE* **12** e0173602
- Lu Q, Wallrath LL, Granok H and Elgin SC 1993 (CT)n (GA)n repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the Drosophila hsp26 gene. *Mol. Cell Biol.* **13** 2802–2814
- Luckey MA, Kimura MY, Waickman AT, Feigenbaum L, Singer A and Park JH 2014 The transcription factor ThPOK suppresses Runx3 and imposes CD4(+) lineage fate by inducing the SOCS suppressors of cytokine signaling. *Nat. Immunol.* **15** 638–645
- Maeng YS, Kwon JY, Kim EK and Kwon YG 2015 Heterochromatin Protein 1 Alpha (HP1alpha: CBX5) is a key regulator in differentiation of endothelial progenitor cells to endothelial cells. *Stem Cells* **33** 1512–1522
- Mahajan K 2016 hPso4/hPrp19: a critical component of DNA repair and DNA damage checkpoint complexes. *Oncogene* **35** 2279–2286
- Matharu NK, Hussain T, Sankaranarayanan R and Mishra RK 2010 Vertebrate homologue of Drosophila GAGA factor. *J. Mol. Biol.* **400** 434–447
- Michaloski JS, Galante PA, Nagai MH, Armelin-Correa L, Chien MS, Matsunami H and Malnic B 2011 Common promoter elements in odorant and vomeronasal receptor genes. *PLoS ONE* **6** e29065
- Mishra K, Chopra VS, Srinivasan A and Mishra RK 2003 Trl-GAGA directly interacts with lola like and both are part of the repressive complex of Polycomb group of genes. *Mech. Dev.* **120** 681–689

- Mishra RK, Mihaly J, Barges S, Spierer A, Karch F, Hagstrom K, Schweinsberg SE and Schedl P 2001 The iab-7 polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. *Mol. Cell Biol.* **21** 1311–1318
- Muroi S, Naoe Y, Miyamoto C, Akiyama K, Ikawa T, Masuda K, Kawamoto H and Taniuchi I 2008 Cascading suppression of transcriptional silencers by ThPOK seals helper T cell fate. *Nat. Immunol.* **9** 1113–1121
- Naro C, Bielli P, Pagliarini V and Sette C 2015 The interplay between DNA damage response and RNA processing: the unexpected role of splicing factors as gatekeepers of genome stability. *Front. Genet.* **6** 142
- Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, *et al.* 2019 The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47** D442–D450
- Pieraccioli M, Nicolai S, Antonov A, Somers J, Malewicz M, Melino G and Raschella G 2016 ZNF281 contributes to the DNA damage response by controlling the expression of XRCC2 and XRCC4. *Oncogene* **35** 2592–2601
- Rui J, Liu H, Zhu X, Cui Y and Liu X 2012 Epigenetic silencing of CD8 genes by ThPOK-mediated deacetylation during CD4 T cell differentiation. *J. Immunol.* **189** 1380–1390
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, *et al.* 2012 Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9** 676–682
- Schweinsberg S, Hagstrom K, Gohl D, Schedl P, Kumar RP, Mishra R and Karch F 2004 The enhancer-blocking activity of the Fab-7 boundary from the *Drosophila* bithorax complex requires GAGA-factor-binding sites. *Genetics* **168** 1371–1384
- Seluanov A, Vaidya A and Gorbunova V 2010 Establishing primary adult fibroblast cultures from rodents. *J. Vis. Exp.* <https://doi.org/10.3791/2033>
- Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, Jangi M, Giallourakis CC, Sharp PA and Young RA 2015 Transcription factor trapping by RNA in gene regulatory elements. *Science* **350** 978–981
- Simon NE, Yuan M and Kai M 2017 RNA-binding protein RBM14 regulates dissociation and association of non-homologous end joining proteins. *Cell Cycle* **16** 1175–1180
- Smeenk G, Wiegant WW, Vrolijk H, Solari AP, Pastink A and van Attikum H 2010 The NuRD chromatin-remodeling complex regulates signaling and repair of DNA damage. *J. Cell Biol.* **190** 741–749
- Song EJ, Werner SL, Neubauer J, Stegmeier F, Aspden J, Rio D, Harper JW, Elledge SJ, Kirschner MW and Rape M 2010 The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. *Genes Dev.* **24** 1434–1447
- Srivastava A, Kumar AS and Mishra RK 2018 Vertebrate GAF/ThPOK: emerging functions in chromatin architecture and transcriptional regulation. *Cell Mol. Life Sci.* **75** 623–633
- Srivastava S, Puri D, Garapati HS, Dhawan J and Mishra RK 2013 Vertebrate GAGA factor associated insulator elements demarcate homeotic genes in the HOX clusters. *Epigenet. Chromatin* **6** 8
- Stratigi K, Kapsetaki M, Aivaliotis M, Town T, Flavell RA and Spilianakis CG 2015 Spatial proximity of homologous alleles and long noncoding RNAs regulate a switch in allelic gene expression. *Proc. Natl. Acad. Sci. USA* **112** E1577–E1586
- Tang Z, Li C, Kang B, Gao G, Li C and Zhang Z 2017 GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45** W98–W102
- Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A and Narechania A 2003 PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13** 2129–2141
- Vilas CK, Emery LE, Denchi EL and Miller KM 2018 Caught with One's Zinc Fingers in the Genome Integrity Cookie Jar. *Trends Genet.* **34** 313–325
- Wai DC, Shihab M, Low JK and Mackay JP 2016 The zinc fingers of YY1 bind single-stranded RNA with low sequence specificity. *Nucleic Acids Res.* **44** 9153–9165
- Widom RL, Lee JY, Joseph C, Gordon-Froome I and Korn JH 2001 The hcKrox gene family regulates multiple extracellular matrix genes. *Matrix Biol.* **20** 451–462
- Zhang R, Ma H, Gao Y, Wu Y, Qiao Y, Geng A, Cai C, Han Y, Zeng YA, Liu X, *et al.* 2018 Th-POK regulates mammary gland lactation through mTOR-SREBP pathway. *PLoS Genet.* **14** e1007211
- Zullo JM, Demarco IA, Pique-Regi R, Gaffney DJ, Epstein CB, Spooner CJ, Luperchio TR, Bernstein BE, Pritchard JK, Reddy KL, *et al.* 2012 DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* **149** 1474–1487

Corresponding editor: BJ Rao