RESEARCH   **Open Access**

Characterization of the major human STAG3 variants using some proteomics and bioinformatics assays

Check for updates

Inam J. Lafta[1*], Bassam K. Kudhair[2] and Noralhuda N. Alabid[3]

Abstract

Background: STAG3 is the meiotic component of cohesin and a member of the Cancer Testis Antigen (CTA) family. This gene has been found to be overexpressed in many types of cancer, and recently, its variants have been implicated in other disorders and many human diseases. Therefore, this study aimed to analyze the major variants of STAG3. Western blot (WB) and immunoprecipitation (IP) assays were performed using two different anti-STAG3 antibodies that targeted the relevant protein in MCF-7, T-47D, MDA-MB-468, and MDA-MB-231 breast cancer cells with Jurkat and MCF-10A cells as positive and negative controls, respectively. In silico analyses were searched to study the major isoforms.

Results: WB and IP assays revealed two abundant polypeptides < 191 kDa and ~ 75 kDa in size. Specific bioinformatics tools successfully determined the three-dimensional (3-D) structure, the subcellular localization, and the secondary structures of the isoforms. Furthermore, some of the physicochemical properties of the STAG3 proteins were also determined.

Conclusions: The results of this study revealed the power of applying the biological techniques (WB and IP) with the bioinformatics assays and the possibility of their exploitation in understanding diseased genes. Exploring the major variants of STAG3 at the protein level could help decipher some disorders associated with their occurrence, along with designing drugs effective at least for some relevant diseases.

Keywords: STAG3 gene, Variants, Immunoblotting, Immunoprecipitation, Bioinformatics

Background

STAG3/SCC3 homolog3/stromalin-3/cohesin subunit SA-3 is the meiotic component of cohesin, which is a highly conserved and universally expressed multi-subunit protein complex [1]. Cohesin proteins are involved in many biological processes, mainly sister chromatid cohesion (SCC), the maintenance of chromatin structure, gene expression, DNA repair [2, 3], and positive regulation of the transcription of genes, e.g., *Myc*, *Runx1*, and *Runx3*, known to be disregulated in cancer [4].

In addition to being a meiotic cohesin component, STAG3 is a member of the Cancer Testis Antigen (CTA) family, which includes any gene that expresses exclusively in the testis as well as in neoplastic cells [5].

STAG3 cDNA was first identified by Pezzi and his colleagues (2000) in human and mouse as a new member of mammalian stromalin of the synaptonemal complex (SC), which is a protein structure that stabilizes homologous chromosomes pairing in prophase stage of the cell cycle [6]. The *Homo sapiens STAG3* gene was mapped to the 7q22 region of chromosome 7, in which six genes related to *STAG3* were mapped, including two at 7q22 near the functional gene, three at 7q11.23, and one at 7q11.22 [6]. It has been proposed that cohesin complex containing STAG3 is functional at the centromeres from the early stages of prophase I till metaphase I [7]. However, during metaphase I and the last meiosis stages, STAG3 was suggested to be located at the interchromatid domains, but not at the chiasmata areas, and was found to be necessary for meiotic sister chromatid cohesion at chromosome arms [8]. Thus, STAG3 is implicated in pairing of chromosomes and is necessary for

* Correspondence: ijlafta@gmail.com; inam.j@covm.uobaghdad.edu.iq
[1]Department of Microbiology, College of Veterinary Medicine, University of Baghdad, Baghdad, Iraq
Full list of author information is available at the end of the article

their correct segregation during meiotic division [9]. It has been shown to associate with the SC [10] and is required for the normal formation of this complex between homologous chromosomes [11].

In cancer, the first somatic mutations of cohesin components, including STAG3 were reported by Barber et al., who identified heterozygous somatic missense mutations in colon cancers [12]. Later, genes encoding cohesin subunits have been demonstrated to be mutated in a wide variety of human neoplasms [13]. Over- and underexpression of cohesin genes have been found to contribute to cancer by causing aneuploidy or chromosome instability [4]. The *STAG3* gene has been suggested to be implicated in the development of epithelial ovarian cancer owing to one common allele of *STAG3* responsible for loss of heterozygosity for one SNP (single-nucleotide polymorphism) [14]. On the other hand, *STAG3* has been verified to cause many diseases other than cancer. It has been considered as a strong candidate for men infertility due to the role it plays in gametogenesis [7]. Similarly, 30 single-nucleotide variations were found in the coding regions and intron boundaries of *STAG3* in patients with nonobstructive azoospermia [15]. Despite the presence of rare variants in *STAG3*, six truncated variants have been reported to be associated with premature ovarian insufficiency till now; these involved one splicing variant, two nonsense variants, and three frameshift variants [16, 17]. Furthermore, the *STAG3* gene has been confirmed to be the causative agent of primary ovarian insufficiency [18], and its role in this disease has been attributed to the presence of two rare heterozygous pathogenic variants in this gene [17]. Another recent study have identified two novel homozygous in-frame variants in *STAG3* in two sisters from a consanguineous Han Chinese family suffering from premature ovarian insufficiency; these variants were verified to be pathogenic [19].

The presence of variants of unknown or unclear significance can impose enormous problems in the existing genetic variation screening approaches and in gene therapy adding to the dilemmas for clinicians regarding patient advice [20]. Moreover, the interpretation of rare genetic variants of unknown clinical importance constitutes one of the major confrontations that encounter human molecular genetics [21]. A definite diagnosis is essential for the patient to know the reason of the disease, for the physician to offer appropriate care and for disease course prediction, and for the geneticist to provide genetic advice to the patient [21]. Analysis of unknown variants in novel disease genes is not only of diagnostic value, but might also be of a scientific impact [21]. At present, analyzing genetic variations in human relies on the detection of a pathogenic variant in individuals under high-risk and a bigger possibility for a certain inheritable disease, such as breast cancer and ovarian

cancer or particular types of monogenic disorders. Although a single genetic variant could offer worthy genetic information for scarce monogenic diseases [20], it has been suggested that different rare variants in the same gene can be responsible for a disease [22].

Alternative splicing of RNA can result in production of multiple versions of a protein called isoforms from a single gene. The term " proteome" indicates proteins encoded by the genome as well as the alterations resulting from posttranslational modifications, in which other chemical elements, such as phosphates, sugars, fats, and even other proteins can be added [23], whereas the term "Proteomics" is a comprehensive study that deals with the structure and function of proteins [23], and it was included in this study to investigate some of STAG3 features.

Due to the importance of STAG3 protein in humans, and because little is known about its variants so far, the present study was conducted. This study aimed to detecting STAG3 isoforms by using bioinformatics along with investigating the abundant variants using western blotting and immunoprecipitation techniques. Discovering major STAG3 isoforms, in turn, might aid in finding a potential novel biomarker for disease diagnosis, in discovering new therapeutic targets and could be of a great value to the scientific community.

Methods
Cell culture
The cell lines used in this study involved the negative control MCF-10A (non-tumorigenic epithelial breast cell line) and the breast cancer cell lines MCF-7, T-47D, MDA-MB-468, and MDA-MB-231. The aforementioned cell lines were purchased from Sigma-Aldrich, UK. In addition, Jurkat cells (leukemia T-lymphocytes) were used as a positive control and presented as a gift from Professor Matthew Holley, Department of Biomedical Science, The University of Sheffield, Sheffield, UK. The MCF-10A cells were grown in Dulbecco's modified Eagle's medium (DMEM; Lonza) containing 4.5 g/L glucose with L-glutamine, and supplemented with 1× non-essential amino acids (NEAAs; Bio Whittaker), 10 μg/mL epidermal growth factor (Sigma-Aldrich), 50 μM hydrocortisone (Sigma-Aldrich), 10 μg/mL insulin (Sigma-Aldrich), 0.1 μg/mL cholera toxin (Calbiochem), and 5% horse serum (Invitrogen).

The breast cancer cell lines were grown in DMEM (Lonza) containing L-glutamine with 4.5 g/L glucose and supplemented with 10% fetal calf serum (FCS; Seralab) and 1× NEAAs (Bio Whittaker). Regarding the positive control, Jurkat cells, they were grown in RPMI 1640 (Roswell Park Memorial Institute medium; Lonza) containing L-glutamine, and provided with 1× NEAAs and 10% FCS. Before use, the above media, Trypsin-Versene

(EDTA) and phosphate-buffered saline (PBS) were warmed in a 37 °C water bath for at least half an hour.

Protein lysates preparation and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Whole protein lysates were extracted from the cell pellets of MCF-10A, Jurkat, MCF-7, T-47D, MDA-MB-231, and MDA-MB-468, using 1× lysis solution (5× RIPA buffer diluted with ddH_2O), which was supplemented with phenylmethanesulfonyl fluoride (PMSF; Sigma-Aldrich; 1 mM), nuclease (25 U/μL; Novagen), 1× each of two phosphatase inhibitor cocktails 2 and 3 (Sigma-Aldrich), and protease inhibitors (Calbiochem). Cell pellets were re-suspended in 1-pellet volume lysis solution, pipetted up and down, incubated on ice for half an hour, vortexed each 10 min, and spun down for 10 min at 13, 400 rpm using Mini Spin placed in 4 °C. Finally, the concentration of the supernatant, which represented the whole cell protein lysate, was determined, aliquoted, and kept in – 80 °C for long-term storage.

The protein concentration of cell lysates was determined by using Bio-Rad Protein Assay as per the manufacturer's instructions. The assay involved preparing different dilutions of bovine serum albumin (BSA; stock 0.1 mg/ mL) to make protein standard curve. A set of Eppendorf tubes was prepared to contain five different amounts of ddH_2O (800, 790, 750, 700, 650, and 600 μL) and mixed with standards' amounts of freshly made BSA (0, 10, 50, 100, 150, and 200 μL, respectively). The BSA protein concentration in these tubes ranged from 0, 1, 5, 10, and 15 to 20 μL, respectively. Simultaneously, another set of Eppendorf tubes containing 800 μL of ddH_2O each was used and mixed with 1 μL of each protein sample. Then, 200 μL of Bio-Rad dye reagent concentrate was mixed with all BSA standards and protein samples. After 5-min incubation at room temperature, 200 μL of each standard and sample was moved to 96-well plate to be read using plate reader. The optical density (OD) of all proteins was measured at 595 nm, and a standard curve was made using Microsoft Excel by plotting the OD of the standards against their concentrations. Finally, the protein concentration of each sample was calculated based on the equation of the standard curve.

Subsequently, the same concentration (30 μg) of each protein lysate was mixed with 5× sample loading buffer and diluted with ddH_2O up to 22 μL, and each was boiled at 95 °C for 5 min by using heat block (Grant Instruments). Then, the protein samples along with pre-stained protein ladder (Geneflow) were spun briefly and loaded onto Sodium dodecyl sulfate polyacrylamide gel to run an electrophoresis called SDS-PAGE, which was performed to separate the protein samples. The gel was made up of 8% resolving gel and 5% stacking gel. Then, the electrophoresis was run at 130 V for 1 h at room temperature using 1× running buffer.

Western blotting (WB)

Immunoblotting (WB) was carried out to analyze the protein level of STAG3 in the cancer cells. Following SDS-PAGE, the samples were transferred from the gel to a nitrocellulose membrane (Amersham) assembled with blotting papers inside Mini Trans-Blot® Electrophoresis Transfer cell system (Bio-Rad) filled with Towbin transfer buffer. Ponceau S stain (Sigma-Aldrich) was used to check for the transfer efficiency from the gel to the membrane. In this step, the membrane was cut into suitable parts to enable each part to be probed later with the proper antibody. Later, the membrane was rinsed with tap water to get rid of the stain. Following blocking with 5% milk/PBST at room temperature for 1 h, the appropriate part of the cut blot was probed with either Abcam (Cat. No. ab69928), Sigma (Cat. No. HPA049106) anti-STAG3 antibody, or the control anti-β-actin (Abcam ab8226) or β-tubulin (Sigma-Aldrich T8328) antibodies, which were used as a loading control to confirm that the same amount of protein was loaded into each well of the gel. These primary antibodies were diluted in 5% milk/PBST and incubated with the blots overnight on a shaker at 4 °C. After washing three times with 0.1% PBST for 8 min each, a suitable secondary IgG horseradish peroxidase-linked antibody diluted in 5% milk/PBST was used. The secondary antibody was incubated with the membrane on a shaker at room temperature for 1 h. Washing was done as mentioned before, and the blots were incubated with 2 mL ECL detection reagents (GE Healthcare) for 1 min. Eventually, the membrane was exposed to X-ray films (Fuji Medical X-ray film), developed and fixed by Konica SRX 101A Processor to visualize the protein bands.

Si-STAG3 transfection

To explore which band was STAG3 protein on WB, transfection experiments were performed to knockdown *STAG3*. Small interfering RNA (siRNA) specific for *STAG3* mRNA (si-STAG3) was used to transfect the breast cancer cell line MCF-7 cells. In this study, specific si-STAG3#1 sense GCGCAAGACCCAAGCCGAU and si-STAG3#2 sense UGACUAUGGUGACAUUAUC (Eurofins) were used. As a control, small-interfering ribonucleic acid (siRNA) non-specific for any gene termed scrambled (sense UAAUGU AUUGGAACGCAUA) from Eurofins was used to transfect the cells. Transfection conditions were optimized for the above cell line. Standard transfection in six-well tissue culture plates was performed. The optimized cell numbers (2 × 10^5 in 2 mL media) were plated overnight using complete medium without antibiotic. The cells were approximately 50–70% confluent at the time of transfection. The next day, if necessary, the media was substituted with fresh complete

media. Otherwise, the transfection continued by the addition of the optimized amount of siRNA to DMEM serum-free medium (SFM) in an Eppendorf tube, which was left for 5 min at room temperature. Then, appropriate amount of Dharmafect® #4 (transfection reagent; Thermo Scientific) was added to SFM in an Eppendorf tube and left for the same time. After that, siRNA-SFM was mixed by pipetting with the Dharmafect-SFM and kept for approximately 25 min at room temperature. A suitable amount of the mixture (siRNA-DharmaFECT-SFM) was added drop-wise to each well containing the growing cells. Finally, the plates were incubated at 37 °C, 5% CO_2 in a humid incubator for 24 or 48 h of transfection before collecting cell pellets to be used in western blot.

Protein lysates preparation and immunoprecipitation (IP)

For the IP assay, the cells were grown to 85% confluence in 10 cm cell culture dishes. One dish was used for the control and other two dishes for each primary antibody to be used. Later on, the media was removed from all plates, and the cells were rinsed gently twice in ice-cold PBS. After discarding all of the PBS, 800 μL of lysis buffer (1% Triton-X100, 50 mM Tris pH 7.6, 200 mM NaCl, 1 mM EDTA, protease inhibitor, phosphatase inhibitor, benzonase) was added to each dish. The cells were scraped into lysis buffer, collected, and transferred to a 1.5-mL Eppendorf tube, which was incubated on ice for 20 min. At 13,000 rpm, cell lysates were spun down for 20 min. Next, protein G beads were washed in a 1.5-mL Eppendorf tube in 1 mL lysis buffer three times (spinning down at 3000 rpm for 30 s each time). One more spinning was done to remove the remaining liquid. Subsequently, an equal volume of lysis buffer was mixed with the beads. To set up the IP, aliquots of 40 μL of bead solution were dispersed into individual tubes, which were kept on ice. When the lysates were spun down, 50 μL of each lysate was used as an input sample, this would show whether the protein is present in the lysate and allow lining up any immunoprecipitated band. In order to perform WB analysis, an appropriate loading buffer was added in a suitable concentration to the lysates. The mixture was boiled at 95°C for 5 min and put in freezer at − 80°C till use. Then, the rest of the cell lysates were put in the appropriately labeled tubes containing protein G beads. While the tubes were kept on ice, 2–5 μg of IgG protein was added to the control sample and an equal mass of STAG3 antibody for the IP. Finally, all the samples were incubated on the cold room rotator at 4 °C and 20 rpm overnight. The next day, the samples were washed by spinning at 3000 rpm for 30 s, removing supernatant without touching the beads and replacing with 1 mL lysis buffer. This step was repeated four times, and the last time included spinning down and removal of liquid as much as possible

without disturbing beads. At the end, 1× WB loading buffer (50 μL) was added and boiled for 5 min at 95 °C, which was followed by spinning down at 3000 rpm for 30 s. The supernatant was spun again, and 30 μL from which were loaded onto SDS-PAGE. With every sample to be loaded on the gel, inputs as well as the IgG control and STAG3 IP samples were also loaded. Both SDS-PAGE and WB were performed as described above.

Bioinformatics analysis of STAG3 isoforms

Using the online NCBI (National Center for Biotechnology Information) and Ensembl software, the *Homo sapiens* STAG3 isoforms were searched. The three-dimensional (3-D) structure of the isoforms was detected using Phyre2 software. Using PSORTII and SOPMA tools, the subcellular localization and the secondary structures of the isoforms were studied, respectively. Some of the physicochemical properties of the STAG3 proteins were determined by ProtParam software.

Results

Western blot

As shown in the immunoblotting, numerous protein sizes were detected by the anti-STAG3 antibodies. Using the antibody produced by Sigma, based on the positive control (Jurkat cells) and the negative control (MCF-10A), STAG3 was postulated to have the size of < 180 kDa (Fig. 1a). In contrast, by using the antibody manufactured by Abcam, a clear band of ~ 135 kDa was produced by Jurkat and the breast cancer cells relative to the normal cell line (Fig. 1b).

Figure 2 shows the same order of the cells run on the left and right parts of the gel with the protein marker in the middle. According to the positive control (Jurkat cells) and the negative control (MCF-10A), the bands assumed to be STAG3 were indicated with arrows. While each anti-STAG3 antibody recognized different protein bands, a band of ~ 75 kDa was detected by both antibodies, and this band may be another variant of STAG3.

STAG3 knockdown

Knockdown of STAG3 in MCF-7 cells, especially when transfected with si-STAG3#2, successfully depleted this protein. Figure 3 (left panel) shows disappearance of a band of less than 180 kDa in the cells transfected with si-STAG3#2 compared to the scrambled control when Sigma antibody was used. However, in the right panel of Fig. 3 where the blot was probed with Abcam antibody, a band of approximately 135 kDa in the scrambled control appeared, but it was absent from the lane containing cells transfected with si-STAG3#2. This implies that these bands might be different splice variants of STAG3.
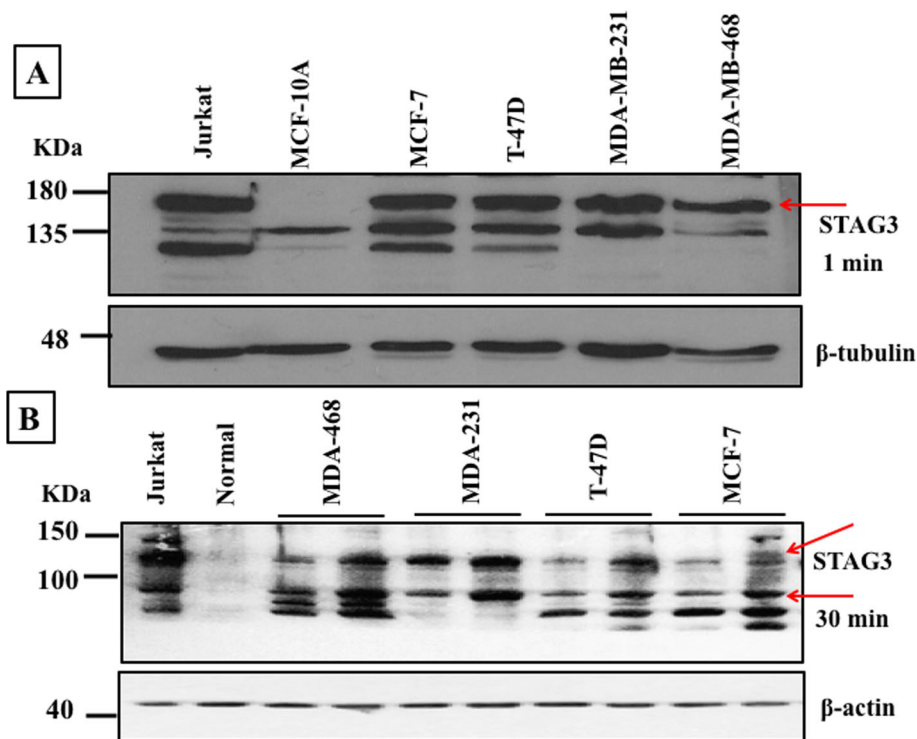
Fig. 1 STAG3 protein levels in breast cancer cell lines relative to the normal. Protein lysates were extracted from the cell pellets of normal breast and cancer cell lines to study STAG3 protein levels by Western blot. From each cell line, four samples from different flasks were used. **a** Western blot analysis was performed using Sigma-Aldrich anti-STAG3 antibody to determine the putative protein level of STAG3 in the breast cancer cell lines compared to the positive control, Jurkat cells, and the negative control, MCF-10A. β-Tubulin was used as a loading control. **b** Western blot analysis was done on two separated protein lysates of each breast cancer cell line and compared to protein lysate extracts of the normal breast and Jurkat cell line. Abcam anti-STAG3 antibody was used to show the putative STAG3 protein level. β-Actin was the loading control. Data shown in **a** and **b** represent four experiments for each

Immunoprecipitation

Analysis of the immunoprecipitated STAG3 protein (Fig. 4) showed the presence of different bands when immuno-blotted against Sigma antibody. Mainly, a band of < 191 kDa and another < 97 kDa were abundant in the cancer cells where the STAG3 protein was immunoprecipitated with Abcam or Sigma antibody. However, only the band of < 97 kDa was noticed in the lane loaded with Jurkat cell lysate compared with the other cell lines because of its high amount and relatively short exposure time (10 s) to X-ray film. This band does exist in the other cells lysates as shown in Figs. 1 and 2, but here in the IP experiment when the blot was incubated with the X-ray film for relatively longer time (30 s), this led to its appearance with other bands in the lane containing lysates only, along with darkening of the other lanes loaded with the immunoprecipitated proteins (data not shown). On the other hand, when Abcam anti-STAG3 antibody was used in WB to detect the immunoprecipitated bands, the finding was consistent with that obtained when Sigma antibody was used; however, Abcam antibody needed higher concentration and longer time to see the bands leading to darkening of the X-ray film (data not shown).

Bioinformatics analysis
The STAG3 proteins sizes

Table 1 shows the sizes and accession numbers of the STAG3 isoforms based on NCBI as well as Ensembl software.

Structure prediction and the subcellular localization of the isoforms

The online software tool SOPMA predicted the secondary structure of the STAG3 encoding products (Table 2), while the PSORTII tool analyzed the subcellular localization of them (Table 2).

The analytic results of Phyre2 software indicated that the secondary structure of all STAG3 isoforms showed ∼ 74% similarity with the crystal structure of human stromal antigen 2 (SA2) in complex with two sister chromatid cohesion protein 1 (SCC1). The 3-D structure of the isoforms was successfully analyzed by Phyre2 tool as exemplified by Fig. 5.

Physicochemical characteristics

Some of the important physicochemical properties of STAG3 proteins including relative molecular weight,
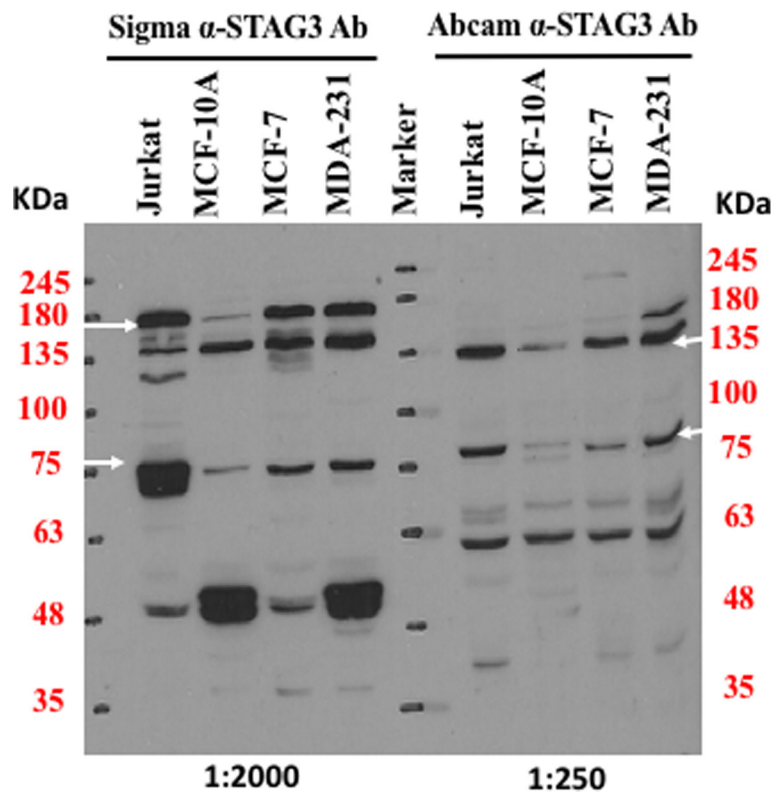
Fig. 2 Using Abcam and Sigma anti-STAG3 antibodies for detecting STAG3 protein levels in different cell lines. Following protein lysates extraction from the normal and cancer cell lines, they were run on the same SDS-PAGE gel with the protein size marker loaded in the middle of the gel. Jurkat cells were the positive control and the normal MCF-10A was the negative control. After transfer to a nitrocellulose membrane, it was cut into two pieces from the region of the protein marker. Each part of the membrane was incubated with either Abcam (diluted 1:250) or Sigma (1:2000 dilution) anti-STAG3 antibody
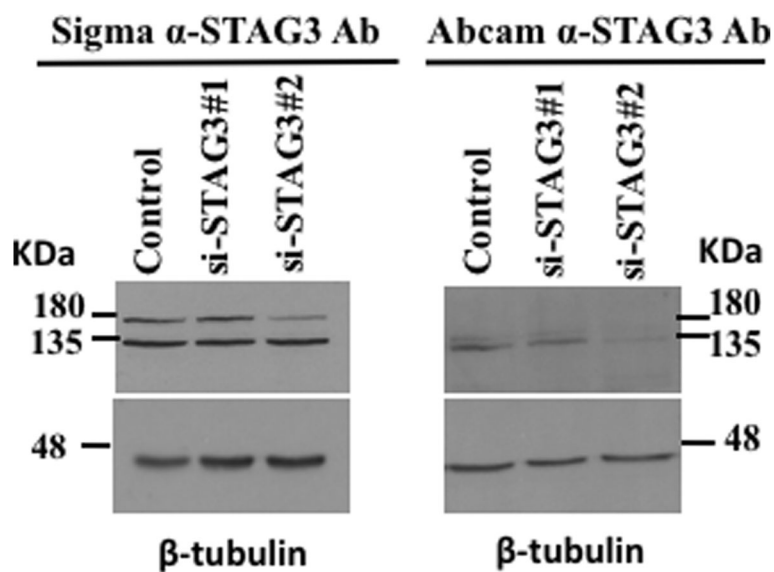

Fig. 3 STAG3 knockdown in MCF-7 cells using specific si-STAG3. To help identify which band represented STAG3, lysates of MCF-7 cells already transfected with scrambled (control) siRNA, si-STAG3#1, or si-STAG3#2 were run on SDS-PAGE and then Western blotted. The labeled bands represent the putative STAG3 proteins
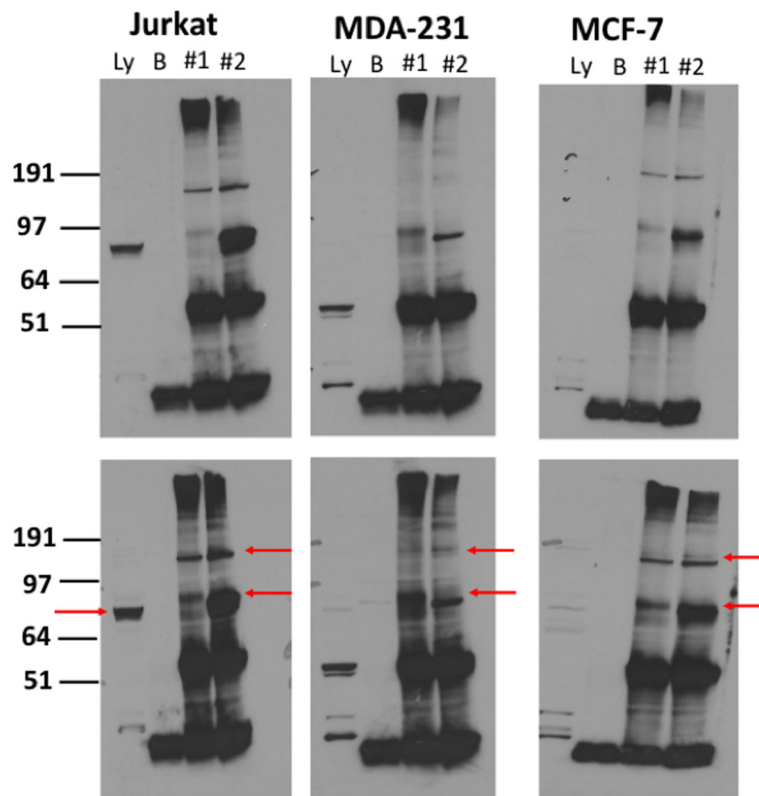
Fig. 4 Western blot analysis for the immunoprecipitated STAG3 protein. Three cancer cell lines, including Jurkat (the positive control), MCF-7, and MDA-MB-231 were used. The STAG3 protein in the lysates (Ly) and those treated with beads and precipitated with either Abcam anti-STAG3 antibody (termed #1) or precipitated with the antibody produced by Sigma (designated #2) as well as beads (B) only was analyzed by Western blot against the antibody manufactured by Sigma to localize protein bands. The upper panel shows blots of the cancer cells exposed to X-ray film for 10 s, while the lower panel refers to the same blot exposed to the film for 30 s. The red arrows represent the postulated STAG3 isoforms

Table 1 The sizes and accession numbers of the STAG3 protein variants based on NCBI and Ensembl

NCBI			Ensembl	
Variant	Size/amino acid (aa)	Protein accession no.	STAG3-ID	Peptide size/aa
1	1225	NP_036579.2 NP_001269645.1	STAG3-206	1225
2	1226	NP_001269646.1	STAG3-218	1226
3	1167	NP_001269647.1	STAG3-219	1167
4	/	/	STAG3-201	1225
X1	1239	XP_016867172.1 XP_016867173.1 XP_016867174.1	STAG3-202	1167
X2	1238	XP_016867175.1	STAG3-203	184
X3	1226	XP_011514044.1	STAG3-207	188
X4	1180	XMP_016867176.1	STAG3-204	144
X5	/	/	STAG3-205	112
X6	/	/	/	/

Table 2 The secondary structure details and the subcellular localization of the STAG3 peptides

Isoform ID	Secondary structure details			
	α-Helix	Random coils	Extended strands	β-Turns
1, 201, 206	60.00%	34.37%	3.76%	1.88%
2, X3, 218	59.95%	34.34%	3.92%	1.79%
3, 202, 219	61.18%	33.08%	4.37%	1.37%
X1	59.16%	34.30%	4.44%	2.10%
X2	59.94%	34.25%	3.96%	1.86%
X4	61.44%	32.80%	4.66%	1.10%
Predicted subcellular localization				
Isoform/possibility %				Subcellular localization
1, 2, X1, X2, X3, 206, 201, 218		3, X4, 202, 219		
22.2%		17.4%		Vacuolar
11.1%		4.3%		Cytoplasmic
11.1%		4.3%		Cytoskeletal
11.1%		4.3%		Mitochondrial
11.1%		21.7%1		Endoplasmic reticulum
11.1%		4.3%		Golgi
11.1%		17.4%		Nuclear
11.1%		26.1%		Plasma membrane

and the amino acid composition were determined by ExPASy online software ProtParam as described below.

Isoform 1.   The molecular weight (mwt) was 139,034 kDa. The most prevalent amino acids (aa) were Leu (L) of 13.9%, Ser (S) of 8.9%, and Glu (E) of 8.2%. All features of isoform 1 were similar to those of STAG3-201 and STAG3-206 (they had similar aa sequences).
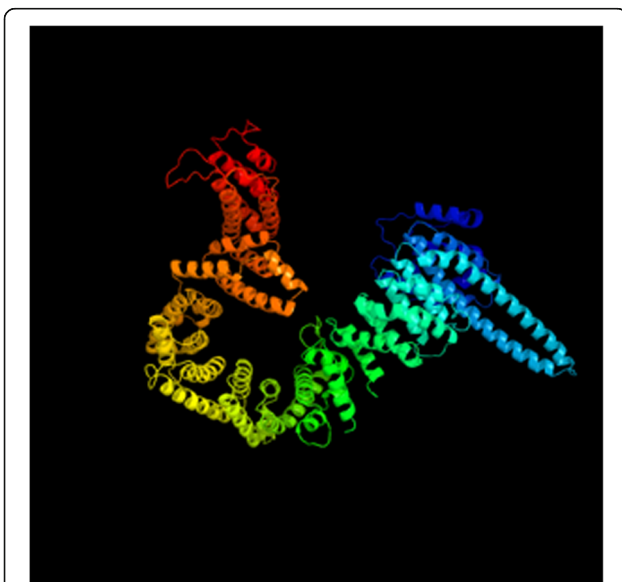
Isoform 2.   The most abundant aa were L (13.9%) and S (9.0%). This protein had mwt of 139,121 kDa, All features of this isoform were similar to those of STAG3-218 and isoform X3 (with similar aa sequences).

Isoform 3.   The most common aa were L (14.1%) and S (8.9%). The protein had a mwt of 132,319 kDa. The variants STAG3-202 and STAG3-219 had the same characteristics as isoform 3.

Isoform X1. It had mwt of 140,441 kDa and pI of 5.25. The aa composition was comprised mainly of L (14.0%) and S (9.0%).

Isoform X2. This protein had mwt of 140,354 kDa. Similar to the other splice variants, the prevalent aa were L and S containing 14.0% and 8.9%, respectively.

Isoform X3. It had the same features as those of variant STAG3-218 and isoform 2 mentioned above.

Isoform X4. It had mwt of 133,640 kDa and pI of 5.34. The aa composition of the protein was comprised mainly of L (14.2%) and S (8.9%).

Fig. 5 Example of the 3-D structure of the STAG3 isoform 1 using Phyre2 software

In addition to the above variants, some short STAG3 isoforms exist in Ensembl, such as STAG3-207, which contained 188 aa of 20,621 kDa. The most prevalent aa were 26 S (13.8%) and 18 L (6.9%). The last short protein coding STAG3 isoform was STAG3-203, which had

mwt of 20,438 kDa. The most abundant aa was Glu (E) (13.6%) and S (13.0%).

Discussion

As no paper investigating human STAG3 protein in cancer cell lines has been published so far, most of the reports have looked at STAG3 in either mouse embryonic fibroblasts or human testis. Therefore, this study is the first to undertake STAG3 analysis at the protein level using immunoblotting and immunoprecipitation in cancer cell lines paralleled with in silico investigation. The *STAG3 mRNA* has been found to be overexpressed in many types of cancer [24, 25]. Furthermore, it is overexpressed in many datasets in Oncomine, with some cancer types exhibit upregulation of *STAG3* over 2-fold in over 50% of samples [25].

Thus, two different human anti-STAG3 antibodies targeting different aa sequences (either carboxy C-terminus or amino N-terminus) of the protein were used here. Analysis of the STAG3 protein with WB using the antibody produced by Sigma showed bands differ from those detected when Abcam antibody was used. In comparison with Jurkat cells and MCF-10A cells, the positive and negative controls, respectively, when using the first antibody (Sigma) STAG3 was hypothesized to have a band of < 180 kDa and another of ~ 75 kDa. Both of these bands were abundant in the positive control while faint in the negative control cells, but were produced by the other cancer cells. On the other hand, when the second antibody manufactured by Abcam was applied, also two abundant bands ~ 135 kDa and 75 kDa were produced by the positive control and the cancer cells, but faint in the negative control. Then, IP assay using the same antibodies was able to determine two abundant bands of STAG3 in the cancer cell lines; their sizes were < 191 kDa and < 97 kDa. The last band may be the same truncated 75 kDa protein determined using immunoblotting, while the first band of < 191 kDa on IP could be either ~ 180 kDa or 135 kDa seen on WB or both together. The transfection experiments using siRNA specific for *STAG3 mRNA* succeeded in depletion of STAG3 at the protein level, especially when using si-STAG3#2. From this experiment, it was also clear that each antibody recognized different STAG3 variants.

STAG3 protein size is well documented to be ~ 135 or 139 kDa in mice and human testis using WB. Nevertheless, this size may not be the same in human in cases of cancer. In comparison with mice, 75% homology is found between human and murine STAG3 protein [6] in normal situations. In the same context, 77% sequence identity exists between rat Stag3 protein and that of human, with the first one encodes for 1256 aa [26]. Upon using NCBI software to search for *Mus musculus Stag3* gene, its coding region was found to have 78.2% identity

with that of human. Moreover, this gene had four transcripts and four isoforms; the longest transcript encoded 1240 aa protein. There were other three predicted transcripts: X1 encoding 1240 aa, X2 encoding 652 aa, and X3 encoding 629 aa. The main difference between STAG3 of human and that of other species resides on N- and C-termini of the proteins [6], for which the anti-STAG3 antibodies used in this study were designed by the manufacturers. From another point of view, as *STAG3* is located on chromosome 7 [6], numerous studies implicate this chromosome as a cause of genetic diseases [27, 28]. This chromosome has been shown to be numerically and structurally influenced in breast cancer, where structural aberrations including deletions, duplications, and translocations have frequently been reported [27]. These findings might partly explain the causes of differences seen in this protein [28].

In support of the above notions, two unique homozygous truncating variants in STAG3 have been identified as the cause of primary ovarian insufficiency. This truncated protein resulted from a homozygous two base pair duplication, which in turn, resulted in a frameshift at amino acid position 650 followed by a premature stop codon, along with omission of exons 19–32 [18].

The second part of this work involved in silico analysis of the STAG3 major variants using the freely available bioinformatics tools. The widely used bioinformatics software, NCBI and Ensembl, showed some consistent results about this gene. However, some paradox is found concerning the numbers and lengths of the transcripts and isoforms of STAG3. Regarding the secondary structure of the STAG3 peptides, various isoforms showed little differences in their structure. However, the secondary structure of the isoforms was found to have 74% similarity with the human stromal antigen 2 (SA2) forming a complex with two sister chromatid cohesion protein 1 (SCC1). Concerning their subcellular localization, the STAG3 proteins were found to be dispersed in different regions of the cell.

Some physicochemical properties of the variants are presented in this research. According to NCBI, the mwt of the proteins ranged from 132,319 kDa to 140,441 kDa, and the number of amino acids ranged from 1167 to 1239. The amino acids count ranged from 112 to 1226 based on Ensembl. Although some short STAG3 variants with mwt of > 20 kDa do exist. However, the detection of adequate genetic variation via complete proteome analysis requires studying large population in order to verify the occurrence of tolerant and intolerant mutations [29]. Taken together, more studies are needed to understand the structure and pathogenicity of the *STAG3* gene and its variants and their relationship with diseases especially cancer.

Conclusions

STAG3 has numerous transcripts and isoforms that could be implicated in many human diseases. Further extensive work on the protein variants, both in vivo and in vitro, is needed to explore the dominant variants responsible for the disease conditions. This highlights the importance of STAG3 isoforms as potential novel biomarkers for certain disease diagnosis. Moreover, discovering the major STAG3 proteins could be a requisite to find suitable gene therapies that interfere with its functions.

Abbreviations
3-D: Three-dimentional; BSA: Bovine serum albumin; Cat. No.: Catalog number; CTA: Cancer testis antigen; DMEM: Dulbecco's modified Eagle's medium; E: Glutamic acid; FCS: Fetal calf serum; IP: Immunoprecipitation; L: Leucine; NCBI: National Center for Biotechnology Information; NEAAs: Non-essential amino acids; PAGE: Polyacrylamide gel electrophoresis; PBS: Phosphate-buffered saline; PMSF: Phenyl-methane-sulfonyl fluoride; RPMI: Roswell Park Memorial Institute medium; S: Serine; SA: Stromal antigen; SC: Synaptonemal complex; SCC: Sister chromatid cohesion; SDS: Sodium dodecyl sulfate; siRNA: Small interfering ribonucleic acid; WB: Western blot

Authors' contributions
IJL designed and supervised the experimental work. IJL, BKK, and NNA performed the biological experiments and the bioinformatics assays as well as analyzing the results. IJL wrote the manuscript. All authors have read and approved the final manuscript.

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Microbiology, College of Veterinary Medicine, University of Baghdad, Baghdad, Iraq. [2]Department of Laboratory Investigations, Faculty of Science, University of Kufa, Najaf, Iraq. [3]Department of Urban Planning, Faculty of Physical Planning, University of Kufa, Najaf, Iraq.

References
1. Duro E, Marston AL (2015) From equator to pole: splitting chromosomes in mitosis and meiosis. Genes Develop. 29:109–122
2. Peters JM, Nishiyama T (2012) Sister chromatid cohesion. Cold Spring Harb Perspect Biol. 4:a011130 [PubMed: 23043155]
3. Haarhuis JH, Elbatsh AM, Rowland BD (2014) Cohesin and its regulation: on the logic of X-shaped chromosomes. Dev Cell. 31:7–18 [PubMed: 25313959]
4. Rhodes JM, McEwan M, Horsfield JA (2011) Gene regulation by cohesin in cancer: Is the ring an unexpected party to proliferation? Mol Cancer Res. 9(12):1587–1607
5. Hofmann O, Caballero OL, Stevenson BJ, Chen YT, Cohen T, Chua R et al (2008) Genome-wide analysis of cancer/testis gene expression. Proc Natl Acad Sci U S A. 105:20422–20427
6. Pezzi N, Prieto I, Kremer L, Perez Jurado LA, Valero C, Del Mazo J et al (2000) STAG3, a novel gene encoding a protein involved in meiotic chromosome pairing and location of STAG3-related genes flanking the Williams-Beuren syndrome deletion. FASEB J. 14:581–592
7. Llano E, Gomez HL, Garcia-Tunon I, Sanchez-Martin M, Caburet S, Barbero JL (2014) STAG3 is a strong candidate gene for male infertility. Hum Mol Genet. 23:3421–3431
8. Prieto I, Suja JA, Pezzi N, Kremer L, Martinez AC, Rufas JS et al (2001) Mammalian STAG3 is a cohesin specific to sister chromatid arms in meiosis I. Nat Cell Biol. 3:761–766
9. Storre J, Schafer A, Reichert N, Barbero JL, Hauser S, Eilers M et al (2005) Silencing of the meiotic genes SMC1beta and STAG3 in somatic cells by E2F6. J Biol Chem. 280:41380–41386
10. Pelttari J, Hoja MR, Yuan L, Liu JG, Brundell E, Moens P et al (2001) A meiotic chromosomal core consisting of cohesin complex proteins recruits DNA recombination proteins and promotes synapsis in the absence of an axial element in mammalian meiotic cells. Mol Cell Biol. 21:5667–5677
11. Hopkins J, Hwang G, Jacob J, Sapp N, Bedigian R, Oka K et al (2014) Meiosis-specific cohesin component, Stag3 is essential for maintaining centromere chromatid cohesion, and required for DNA repair and synapsis between homologous chromosomes. PLoS Genet. 10:e1004413
12. Barber TD, McManus K, Yuen KW, Reis M, Parmigiani G, Shen D et al (2008) Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. Proc Natl Acad Sci USA. 105:3443–3448 [PubMed: 18299561]
13. Hill VK, Kim JS, Waldman T (2016) Cohesin mutations in human cancer. Biochim Biophys Acta. 1866(1):1–11. https://doi.org/10.1016/j.bbcan.2016.05.002
14. Notaridou M, Quaye L, Dafou D, Jones C, Song H, Hogdall E (2011) Common alleles in candidate susceptibility genes associated with risk and development of epithelial ovarian cancer. Int J Cancer. 128:2063–2074
15. Nam Y, Kang KM, Sung SR, Park JE, Shin YJ, Song SH (2019) The association of stromal antigen 3 (STAG3) sequence variations with spermatogenic impairment in the male Korean population. Asian J Androl. 21:1–6. https://doi.org/10.4103/aja.aja_28_19
16. Caburet S, Arbolenda VA, Llano E, Overbeek PA, Barbero JL, Oka K (2014) Mutant cohesin in premature ovarian failure. New Eng J Med. 370:943–949
17. Franca MM, Nishi MY, Funari MF, Lerario AM, Baracat EC, Hayashida SA et al (2019) Two rare loss-of-function variants in the STAG3 gene leading to primary ovarian insufficiency. Euro J Med Genet. 62(3):186–189. https://doi.org/10.1016/j.ejmg.2018.07.008
18. Quesne Stabej PL, Williams HJ, James C, Tekman M, Stanescu HC, Kleta R et al (2016) STAG3 truncating variant as the cause of primary ovarian insufficiency. Euro J Hum Genet. 24:135–138
19. Xiao WJ, He WB, Zhang YX, Meng LL, Lu GX, Lin G et al (2019) In-frame variants in STAG3 gene cause premature ovarian insufficiency. Frontiers in Genetics. 10:1016. https://doi.org/10.3389/fgene.2019.0101
20. Oulas A, Minadakis G, Zachariou M, Spyrou GM (2019) Selecting variants of unknown significance through network-based gene-association significantly improves risk prediction for disease-control cohorts. Scientific Reports. 9:3266. https://doi.org/10.1038/s41598-019-39796-w
21. Rodenburg RJ (2018) The functional genomics laboratory: functional validation of genetic variants. J Inherit Metab Dis. 41:297–307. https://doi.org/10.1007/s10545-018-0146-7
22. Cirulli ET (2016) The increasing importance of gene-based analyses. PLoS Genet. 12(4):e1005852. https://doi.org/10.1371/journal.pgen.1005852
23. Adams J (2008) The proteome: discovering the structure and function of proteins. Nat Edu. 1(3):6
24. Pohlers M, Truss M, Frede U, Scholz A, Strehle M, Kuban RJ (2005) A role for E2F6 in the restriction of male-germ-cell-specific gene expression. Curr Biol. 15:1051–1057
25. Strunnikov A (2013) Cohesin complexes with a potential to link mammalian meiosis to cancer. Cell regeneration (London, England). 2:4.
26. Bayes M, Prieto I, Noguchi J, Barbero JL, Perez Jurado LA (2001) Evaluation of the Stag3 gene and the synaptonemal complex in a rat model (as/as) for male infertility. Mol Reprod Dev. 60:414–417

27. Rondon-Lagos M, Verdun Di Cantongo L, Marchio C, Rangel N, Payan-Gomez C, Gugliotta P et al (2014) Differences and homologies of chromosomal alterations within and between breast cancer cell lines: a clustering analysis. Mol Cytogenet. 7:8
28. Aka JA, Lin SX (2012) Comparison of functional proteomic analyses of human breast cancer cell lines T47D and MCF7. PLoS One. 7:e31532
29. Hicks M, Bartha I, di Iulio J, Venter JC, Telenti A (2019) Functional characterization of 3D protein structures informed by human genetic diversity. PNAS. 116(18):8960–8965 www.pnas.org/cgi/doi/10.1073/pnas.1820813116

Publisher's Note