

# Proteomics-Based Identification of Differentially Abundant Proteins from Human Keratinocytes Exposed to Arsenic Trioxide

Udensi K Udensi<sup>1\*</sup>, Alan J Tackett<sup>2</sup>, Stephanie Byrum<sup>2</sup>, Nathan L Avaritt<sup>2</sup>, Deepanwita Sengupta<sup>2</sup>, Linley W Moreland<sup>2</sup>, Paul B Tchounwou<sup>1</sup> and Raphael D Isokpehi<sup>1,3</sup>

<sup>1</sup>RCMI Center for Environmental Health, College of Science, Engineering and Technology, Jackson State University, Jackson Mississippi 39217, USA

<sup>2</sup>Proteomics Facility, University of Arkansas for Medical Sciences, Department of Biochemistry and Molecular Biology, Little Rock, AR 72205, USA

<sup>3</sup>Department of Biology, School of Science, Engineering and Mathematics, Bethune-Cookman University, Daytona Beach FL 32114, USA

## Abstract

**Introduction:** Arsenic is a widely distributed environmental toxicant that can cause multi-tissue pathologies. Proteomic assays allow for the identification of biological processes modulated by arsenic in diverse tissue types.

**Method:** The altered abundance of proteins from HaCaT human keratinocyte cell line exposed to arsenic was quantified using a label-free LC-MS/MS mass spectrometry workflow. Selected proteomics results were validated using western blot and RT-PCR. A functional annotation analytics strategy that included visual analytical integration of heterogeneous data sets was developed to elucidate functional categories. The annotations integrated were mainly tissue localization, biological process and gene family.

**Result:** The abundance of 173 proteins was altered in keratinocytes exposed to arsenic; in which 96 proteins had increased abundance while 77 proteins had decreased abundance. These proteins were also classified into 69 Gene Ontology biological process terms. The increased abundance of transferrin receptor protein (TFRC) was validated and also annotated to participate in response to hypoxia. A total of 33 proteins (11 increased abundance and 22 decreased abundance) were associated with 18 metabolic process terms. The Glutamate-cysteine ligase catalytic subunit (GCLC), the only protein annotated with the term sulfur amino acid metabolism process, had increased abundance while succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial precursor (SDHB), a tumor suppressor, had decreased abundance.

**Conclusion:** A list of 173 differentially abundant proteins in response to arsenic trioxide was grouped using three major functional annotations covering tissue localization, biological process and protein families. A possible explanation for hyperpigmentation pathologies observed in arsenic toxicity is that arsenic exposure leads to increased iron uptake in the normally hypoxic human skin. The proteins mapped to metabolic process terms and differentially abundant are candidates for evaluating metabolic pathways perturbed by arsenicals.

**Keywords:** Arsenic; HaCaT keratinocytes; Skin cancer; TFRC; Proteomics; Visual analytics

## Background

Arsenic (As) is a widely distributed environmental toxicant that causes skin cancer [1]. The types of skin cancer associated with arsenic include intraepidermal carcinomas (Bowen disease) [2], squamous cell carcinomas (SCC), basal cell carcinomas (BCC) [3], Merkel cell carcinoma (MCC) [4] and head and neck cancers [5]. Microarrays gene expression has revealed that arsenicals impact the function of diverse tissue types including the skin, [6,7], bladder and kidney [8], liver [9], prostate and lung [10], peripheral lymphocytes [11], neural tube [12], and urogenital cells [13]. Arsenic also is useful for the treatment of relapsed or refractory acute promyelocytic leukemia as well as solid tumors [14-17]. Clearly, diverse tissue types are affected by arsenic. However, new knowledge is needed to understand the molecular mechanisms of arsenic-induced cancers and arsenic-treatment of cancers in diverse tissue types.

The objective of the research reported here was to determine proteins in HaCaT keratinocyte that are altered in response to arsenic exposure. The application of toxicogenomics methods including microarray gene expression has facilitated the understanding of molecular pathology, and increased the understanding of toxicant responses [18,19]. Toxicoproteomics uses high-throughput technologies to identify critical proteins and pathways in biological systems that are affected by adverse chemical and environmental exposures. Therefore, to reliably identify proteins perturbed by arsenic trioxide in HaCaT keratinocyte

model, we have used a combination of LC-MS/MS with a Thermo LTQ-XL mass spectrometer coupled to an Eksigent nanoLC-2D [20].

The results from proteomics experiments provides list of proteins with altered abundance that have statistical significant changes in their abundance between two conditions. The list of proteins presents data set for discovery of functional categories that could be of biological significance to the arsenic-induced pathologies. For example, skin is a target organ for several pathological manifestations of arsenic toxicity including both hypo- and hyperpigmentation [21]. Several bioinformatics tools enable the annotation of the list of proteins with functional categories [22-36]. A challenge with multiple heterogeneous functional annotations is availability of informatics resources for conducting a variety of tasks on the data sets. These tasks include

**\*Corresponding author:** Udensi K Udensi, RCMI Center for Environmental Health, College of Science, Engineering and Technology, Jackson State University, Jackson Mississippi 39217, USA, Tel: +1-214-773-5598; E-mail: [udensi.k.udensi@jsums.edu](mailto:udensi.k.udensi@jsums.edu)

**Received** May 08, 2014; **Accepted** June 24, 2014; **Published** June 28, 2014

**Citation:** Udensi UK, Tackett AJ, Byrum S, Avaritt NL, Sengupta D, et al. (2014) Proteomics-Based Identification of Differentially Abundant Proteins from Human Keratinocytes Exposed to Arsenic Trioxide. J Proteomics Bioinform 7: 166-178. doi:10.4172/0974-276X.1000317

**Copyright:** © 2014 Udensi UK, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

identifying groups of altered proteins with identical functional categories.

In our research, the data obtained from proteomics were annotated for tissue types [22,26-28], Gene Ontology [28,32], and gene family membership [33]. A visual analytics environment was developed to facilitate several tasks designed to construct knowledge on potentially arsenic-modulated (i) tissue types; (ii) biological processes; and (iii) gene families. Visual analytics environments provide software and hardware to interact with data as well as make sense of relationships and patterns in collection of data sets [34].

In this report, arsenic-regulated proteins refer to proteins from the proteomics experiments with altered abundance (spectral counts) in the HaCaT keratinocyte exposed to arsenic relative to the control condition (no exposure to arsenic). The terms for describing proteins with increased abundance and proteins with decreased abundance were up-regulation and down-regulation respectively. The abundance of 173 proteins was altered in keratinocytes exposed to arsenic; in which 96 proteins had increased abundance while 77 proteins had decreased abundance. These proteins with altered abundance were also classified into 69 Gene Ontology biological process terms. Based on our prior research [35,36], we expected the up-regulation of the human transferrin receptor protein (TFRC) in the proteomics experiment. TFRC is involved in cellular uptake of iron via receptor-mediated endocytosis of ligand-occupied transferrin receptor into specialized endosomes [37-39]. We have identified a need for additional research on TRFC involvement in the hyperpigmentation observed in arsenic induced diseases [36]. Indeed, an increased abundance of transferrin receptor protein (TFRC) was observed and validated with RT-PCR and western blot. Further, TFRC was annotated to participate in response to hypoxia. A possible explanation for hyperpigmentation pathologies observed in arsenic toxicity is that arsenic exposure leads to an increased iron uptake and accumulation in the normally hypoxic human skin [37]. A total of 33 proteins (11 increased abundance and 22 decreased abundance) were associated with 18 metabolic process terms. The proteins mapped to metabolic process terms and differentially abundant are candidates for evaluating metabolic pathways perturbed by arsenicals.

## Method

### Cell culture

HaCaT cells (obtained from Dr. Van Wilson, Microbial & Molecular Pathogenesis, College of Medicine, Texas A&M Health Science Centre, College Station, Texas, USA) were cultured in Dulbecco's Modified Eagle's medium (DMEM) media supplemented with 10% fetal bovine serum (FBS) and 100 µg/mL penicillin-streptomycin at 37°C with 5% CO<sub>2</sub> in a humidified incubator. The cell lines were divided into two experimental groups; Exposed and Unexposed. Each group had three biological replicates. The exposed cells were exposed to 0.5 ppm of arsenic trioxide and passaged up to 8 passages. The unexposed controls were also passaged (passage control) alongside the exposed groups but without any exposure to arsenic trioxide up to passage 8 [40].

### Cytotoxicity assay (MTT Assay)

Dose response relationship was determined by measuring the survival rate (% viability) of cells after treatment with arsenic trioxide using MTT (tetrazolium salts) assay. Arsenic trioxide (arsenic trioxide dissolved in dilute nitric acid, 1 mL=1 mg AS; Reference Standard Solution, 1000 ppm ± 1%/certified, 99.9% purity, Fisher Scientific) dilutions were prepared using a complete DMEM medium as the

diluent. Using a 96 well plate, a total of 200 µL complete medium only was added to the first column, then 100 µL of HaCaT cell suspension in complete medium containing approximately 20,000 cells/well was added to columns 3-8 wells of the 96-well plates, 100 µL of medium was added to column 2, and 100 µL of arsenic trioxide concentrations was added to columns 3-8, ensuring that the final concentrations were 0, 1.0, 5.0, 10, 15, 20, and 25 ppm, respectively. They were incubated in a humidified incubator at 37°C in a 5% CO<sub>2</sub> for a period of 48 hours. 10 µL of the yellow MTT reagent was added to each of the wells and incubated at 37°C for 4 hours until purple formazan precipitate was visible. Then 100 µL of the detergent reagent was added to each of the wells and kept in the dark at room temperature for 2 hours. The absorbance was measured at 570 nm wave length using a microtiter plate reader (Fluoroskan II microplate reader, Helsinki, Finland). LD<sub>10</sub> and LD<sub>50</sub> were determined from the dose-response curve. Acute cytotoxicity experimental results were used to determine tolerated dose (LD<sub>10</sub>) in all the cell lines for chronic assays as used by Trouba et al. [41].

### Cell harvest

At approximately 80% confluence in DMEM/10% FBS at 37°C with 5% CO<sub>2</sub>, the media was discarded and washed three times with 5 mL DPBS. About 2 mL of 0.25% trypsin EDTA (1x) (Gibco, Life Technologies) was added to dissociate the cells and 4 mL of whole media was added to stop the trypsinization reaction. The supernatant with the dissociated cells were completely aspirated into 15 mL centrifuge tubes and centrifuged at 300 g for 5 minutes. The media was aspirated out and the pellet was washed 3 times with 5 mL DPBS.

### Cell lysis

The pellets were resuspended in 150 µL of 2X BME (β-mercaptoethanol) (155.8 mM tris-HCL, 0.006% bromophenol blue, 49.8 mM tri base, 12.5 mM TCEP-HCL, 25% glycerol, 5% SDS) in 1.5 mL microcentrifuge tube. It was sonicated for 5 minutes using a Bioruptor UCD 200 (Diagenode) on high power with a 30 second on/off cycle and heated at 95°C for 5 minutes and finally centrifuged at max for 1 minute.

### Determination of protein concentration

The protein concentration in the cell lysate was measured by spotting 2 µl of each sample onto read plate and was read with Take Three BioTek microplate spectrophotometer. It was analyzed using Take Three interface and the 280 nm reading was used to approximate loading volumes.

### Determination of proteomic responses

Protein was resolved using SDS-PAGE. Each SDS-PAGE gel lane was sliced into 24 bands of 3 mm segments in a laminar flow hood. Enzymatic digestion of proteins was performed by (i) reduction of disulfide bonds using Tris (2-carboxyethyl) phosphine hydrochloride (TCEP), (ii) alkylation of reduced Cys residues with iodoacetamide (IA) and (iii) digestion of proteins with trypsin. The reaction was stopped by adding formic acid. Tryptic peptides from the 144 gel bands were analyzed by nanoflow LC-MS/MS with a Thermo Orbitrap Velos mass spectrometer equipped with a Waters nanoACQUITY LC system. Tryptic peptides were separated by reverse phase Jupiter Proteo resin (Phenomenex) on a 100×0.1 mm column using a nanoAcquity UPLC system (Waters). Peptides were eluted using a 30 min gradient from 98:2 to 40:60 buffer A:B ratio. [Buffer A=0.1% formic acid, 0.05% acetonitrile; buffer B=0.1% formic acid, 75% acetonitrile.]

Eluted peptides were ionized by electrospray (2.0 kV) followed by MS/MS analysis using collision induced dissociation on a Thermo LTQ Orbitrap Velos mass spectrometer. MS data were acquired using the FTMS analyzer in profile mode at a resolution of 60,000 over a range of 375 to 1500 m/z. MS/MS data were acquired for the top 15 peaks from each MS scan using the ion trap analyzer in centroid mode and normal mass range with normalized collision energy of 35.0. Proteins were identified by a Mascot (version 2.3.01) database search with the following parameters: precursor ion tolerance 5 ppm, fragment ion tolerance 0.65 Da, fixed modification of carbamidomethyl on cysteine, variable modification of oxidation on methionine and acetylation on n-terminus, and 2 missed cleavages possible with trypsin. We searched the human 'Con\_IPI\_Hum.v3.34a' database (International Protein Index for *Homo sapiens*) and we additionally used reversed sequences for higher confidence identifications. The Mascot results were uploaded into Scaffold 3 (version 3.5.1) for viewing the proteins and peptide information. A protein threshold of 95%, minimum of 2 peptides, and a peptide threshold of 50% were used to export spectral counts to spreadsheet file for further data analysis. The general rule is to explain the spectral data with the smallest set of proteins when using scaffold to determine peptides that could be assigned to different proteins, hence our application of a 50% peptide threshold. For example, if protein A and protein B each have one unique peptide, they will be listed separately only if the peptide probability is >50% [42].

### Identification of arsenic regulated proteins

A label-free approach based on spectral counting was used [43,44] to determine if a protein was differentially regulated between arsenic treated and the untreated control cells. Spectral count by definition is the number of tandem mass spectra assigned to a given protein in all bands from a single gel lane. To determine the relative amount of a protein in a given gel lane, a normalized spectral abundance factor (NSAF) was calculated [44]. NSAF is calculated as the number of spectral counts (SpC) identifying a protein, divided by the protein's length (L), divided by the sum of SpC/L for all proteins in the experiment. The abundance of individual proteins in multiple independent samples can be compared using NSAF, which also allows quantification of expression changes in different complexes [45,46].

Prior to statistical analysis, the spectral count data were first normalized in order to compare between samples, account for heteroscedasticity (log transformation), and to account for the relative amount of proteins in a given gel lane by calculating a Normalized Spectral Abundance Factor (NSAF) [44,47].

$$(NSAF)_k = \frac{\left(\frac{SpC}{MW}\right)_k}{\sum_{i=1}^N \left(\frac{SpC}{MW}\right)_i}$$

The NSAF for a protein *k* is the number of Spectral Counts (SpC) identifying a protein, *k*, divided by the protein's Molecular Weight (MW), divided by the sum of SpC/MW for all *N* proteins in the gel lane [20].

Different proteins were identified among the arsenic exposed and unexposed keratinocyte cells. Therefore, some proteins will contain a spectral count of zero for a particular sample. In order to allow for log transformation, Zybailov et al. [44] described a method for processing spectral counts of zero in the data set by replacing the zero values with a fractional value prior to the NSAF calculation. In the present study, we replaced zero spectral count values with 0.5. As is commonly done

in microarray data sets, the data was scaled to ensure the medians of all distributions were equal and centered to ensure the standard deviations of all distributions were equal, allowing for more robust statistical testing [29]. The significance of the differences in spectral abundance was determined with t-tests.

### Real Time RT-PCR and western blots of transferrin receptor (TFRC)

In our prior bioinformatics research, we observed that arsenic increased the expression of the genes for transferrin receptor [35,36]. To confirm this prediction, we conducted RT-PCR and Western Blot experiments. Total RNA from cultured cells was extracted using RNeasy Mini kit (Qiagen, Valencia, CA) following the manufacturer's specifications. RNA was treated with RNase-free DNase I Amplification Grade (Invitrogen, Grand Island, NY) to remove any DNA contamination. About 1 µg of purified RNA was used for synthesis of cDNA with oligo dT primers using Superscript First-Strand Synthesis System (Invitrogen, Grand Island, NY) following manufacturer's instructions. The cDNA was purified by RNase H (Invitrogen, Grand Island, NY). Quantitative real-time PCR was performed using a reaction mixture containing 1X Advanced SYBR Green Supermix (Bio-Rad, Hercules, CA), 0.33 µM of each primer. The cyclic conditions used were 95°C for 2 min, followed by 45 cycles at 95°C for 15 s and 50°C for 20 s. The following primer pairs were used: TFRC; Forward Primer: 5'-GCGATAACAGTCATGTGGAGATG-3'; Reverse Primer: 5'-CTGTTGCAGCCTTACTATACGCC-3'; β-Actin: forward: 5'-CTGGCACCCAGCACAATG-3', and reverse 5'-GCCGATCCACACGGAGTACT-3'. The expression levels were standardized by one housekeeping gene, β-Actin in order to validate the result [48]. The comparative CT method was used to quantify mRNA expression and expressed as arbitrary units [49].

The protein lysates used for mass spectrometry analysis were resolved by SDS-PAGE (12% SDS-Tris-glycine gel). The gels were run under reducing conditions, and the resolved proteins were visualized with Coomassie stain. In addition, a portion of the protein lysates were resolved by SDS-PAGE and transferred to a polyvinylidene difluoride (PVDF) membrane for western blot analysis.

The proteins transferred were confirmed by staining the PVDF with 2 ml of amido black stain for 1 minute, and then washed 3 times with water. The membrane was washed thoroughly with TBS and was blocked with 5% milk in Tris Buffered Saline with Tween 20 (TBST) overnight with rocking at 4°C. The primary antibody TFRC, (Invitrogen 13-6800) was used at a dilution of 1:10000 for a time period of 1hour with rocking at room temperature. The secondary antibody, goat/anti-mouse IgG peroxidase conjugate was used appropriately at a dilution of 1:10000 in block solution and was incubated for 1 hour with rocking at room temperature. The ECL Perking Elmer detection kit system was used to visualize the reactive proteins on the blots. The quantification of TFRC protein levels was determined by normalization to glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

### Functional annotation analytics

The functional annotations of the differentially abundant proteins were determined with bioinformatics tools. Subsequently, visual analytics tasks were conducted on the data sets obtained from the bioinformatics tools. A visual analytics tool allows the user to perform several human cognitive tasks on datasets through the integration of automated analysis, interaction and visualization [23,24,36,50]. The visual analytics approach will also enable user specific tasks or questions to be accomplished.

For the arsenic-regulated proteins, the Database for Annotation, Visualization and Integrated Discovery (DAVID) [27,28] was used to determine (i) the Universal Protein Resource (UniProt) tissue annotation [26]; and (ii) Gene Ontology annotation [31,51]. Data sets from (i) proteomic analysis; (ii) DAVID bioinformatics resources; and (ii) the Comparative Toxicogenomics Database were uploaded in Tableau Visual Analytics Software (Tableau Software Inc. Seattle, WA). The visualization and analysis designs were developed for several tasks that require blending of selected fields from multiple data sources. The International Protein Identifiers (IPI) served as the shared field.

Three tasks that were conducted using the visual analytics tools are reported in this article. The tasks were based on results from the DAVID Functional Annotation Tool. The results obtained from the Functional Annotation Tool were uploaded as a data source in Tableau Software. Subsequently, blending of the data sources was performed by selecting the appropriate fields including the shared field.

The first task was to identify arsenic regulated proteins that are annotated with UniProt tissue term “keratinocyte” or “Skin”. This task was accomplished by blending the tissue annotations from DAVID with the proteomics data set that we produced. In the second task, the proteomics data set was annotated and grouped by Gene Ontology (GO) terms for biological process. When the GO annotation is present, one or more GO terms could be mapped to the protein. For example: “GO:0006508~proteolysis,GO:0006511~ubiquitin-dependent protein catabolic process,GO:0009057~macromolecule catabolic process”. In our analytics, only the first GO identifier was selected to accomplish the annotation and grouping of the proteomics data set. To have access to the entire Gene Ontology terms for this research, the GO Terms and GO IDs were extracted from the full ontology file of the Gene Ontology Consortium (<http://www.geneontology.org/GO.downloads.ontology.shtml>). Providing the data set in a visual analytics tool will allow us also to identify GO terms that contain the word “arsenic”. In the third task, the tissue annotations for members of protein families in the proteomics data set were compared. The mapping of gene symbols to gene family descriptions was downloaded from the website of the Human Genome Organization Gene Nomenclature Committee (HGNC [www.genenames.org](http://www.genenames.org)). The gene family description data was uploaded to the visual analytics tool to group proteins by the tissue annotation profiles.

## Results

### Cytotoxicity assay (MTT Assay)

MTT assay shows that acute exposure of HaCaT cells to arsenic had LD10 and LD50 values of 1mg/L and 10mg/L respectively. The LD10 dose result obtained from the cytotoxicity assay (Figure 1) was arbitrarily divided by 1/2 to get 0.5 mg/L (0.5 ppm) of Arsenic that was used for chronic dosage treatment.

### Identification of proteins with differential abundance in HaCaT keratinocytes exposed to arsenic trioxide

A total of 2164 proteins were identified from the arsenic exposed cells (1.0% false discovery rate using a decoy database from 192940 spectra). The control group yielded 2268 (0.8% false discovery rate using a decoy database from 198762 spectra). The Mascot results were uploaded into Scaffold 3 (version 3.00.01) for viewing the proteins and peptide information. A total of 1633 proteins were further analyzed after normalization (NSAF). A total of 173 unique proteins had significant increase or decrease in protein abundance when the HaCaT keratinocytes exposed to arsenic were compared to the

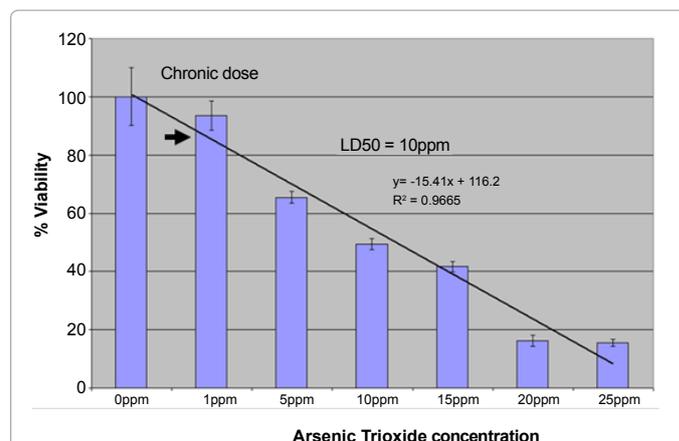


Figure 1: Cytotoxicity assay (MTT Assay) result

To determine the LD10/LD 50 and estimate the minimal cytotoxic concentration that would be used to establish a chronic cell exposure condition. The exposure of HaCaT cells to arsenic had LD10 and LD50 values of 1ppm and 10ppm respectively. The LD10 dose result obtained from the initial cytotoxicity assay was arbitrarily divided by 1/2 to get 0.5ppm of Arsenic that was used for chronic dosage treatment.

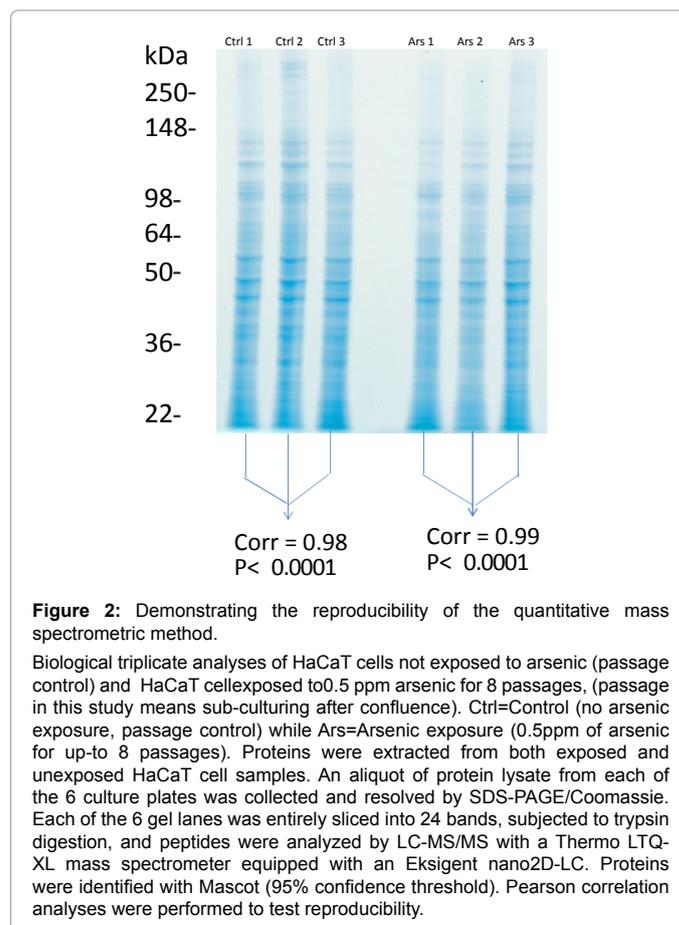


Figure 2: Demonstrating the reproducibility of the quantitative mass spectrometric method.

Biological triplicate analyses of HaCaT cells not exposed to arsenic (passage control) and HaCaT cells exposed to 0.5 ppm arsenic for 8 passages, (passage in this study means sub-culturing after confluence). Ctrl=Control (no arsenic exposure, passage control) while Ars=Arsenic exposure (0.5ppm of arsenic for up-to 8 passages). Proteins were extracted from both exposed and unexposed HaCaT cell samples. An aliquot of protein lysate from each of the 6 culture plates was collected and resolved by SDS-PAGE/Coomassie. Each of the 6 gel lanes was entirely sliced into 24 bands, subjected to trypsin digestion, and peptides were analyzed by LC-MS/MS with a Thermo LTQ-XL mass spectrometer equipped with an Eksigent nano2D-LC. Proteins were identified with Mascot (95% confidence threshold). Pearson correlation analyses were performed to test reproducibility.

HaCaT keratinocytes not exposed to arsenic (control). Transferrin receptor protein 1 (TFRC) had increased abundance in the cells. The reproducibility of the quantitative mass spectrometric method is demonstrated in Figure 2.

The availability (presence or absence) of raw spectra data was

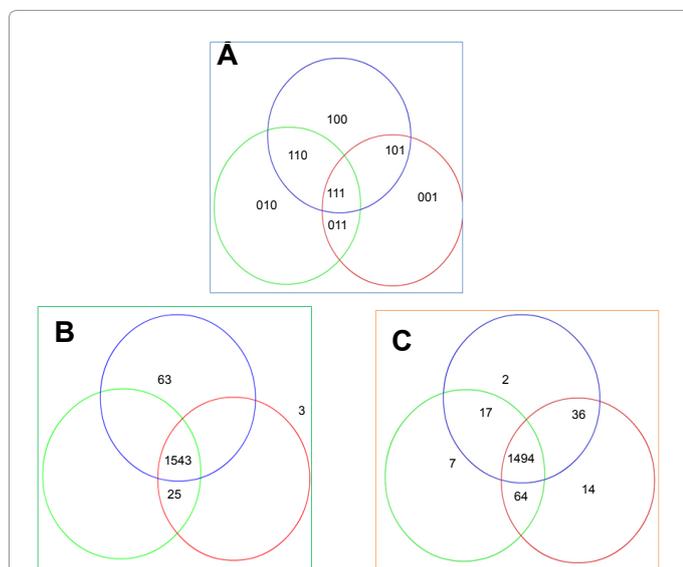
determined for each biological replicate experiment. Since three biological replicates were done, a 3-digit binary code was used to encode the availability for the three exposed set and the three unexposed set. The data obtained from the two sets was then used to determine the agreement of availability of raw spectra data for all the six replicates. Of the total 1634 protein spots analyzed, 91.4% in exposed set and 94.4% unexposed set had presence of spectra in all replicates (Table 1). Within the 173 differentially abundant proteins only two proteins did not agree. The reproducibility of MS analysis is illustrated in Figure 3. The normalized MS result data and calculation formula are shown in the supplementary file.

The 173 significant proteins, according to the lowest *p*-value from the t-test, were visually inspected by hierarchical clustering (Figure 4). Both RT-PCR and Western Blot experiments showed that arsenic induced a significant increase in abundance of TFRC in HaCaT cells treated with arsenic. In Figure 5, the relative quantification relates the PCR signal of TFRC transcript in arsenic exposed HaCaT cell compared to that of unexposed control HaCaT cells. The western

Binary Code for Unexposed	Binary Code for Exposed						
	001	010	011	100	101	110	111
000							3
011	3		4		6	3	9
100	1	1	13		3	2	43
111	10	6	47	2	27	12	1439

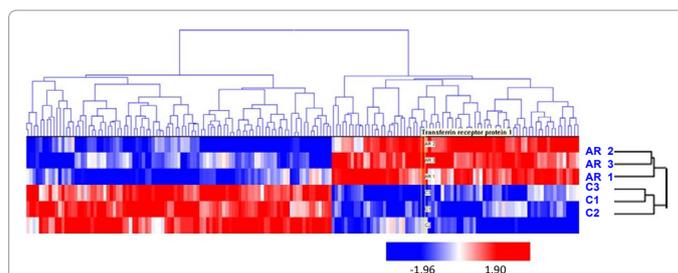
A 3-digit binary code was used to encode the availability for the three exposed replicates set and the three unexposed replicates set using the raw spectra data. Presence and absence of spectra are encoded as 1 and 0 respectively.

**Table 1:** Agreement on availability of raw spectra data in biological replicates.



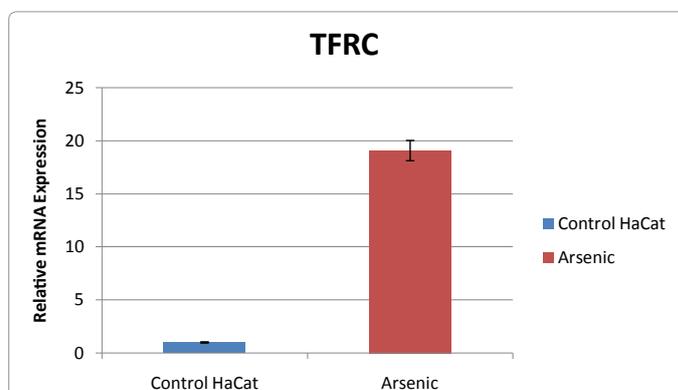
**Figure 3.** The reproducibility of mass spectrophotometer analysis for cells exposed to arsenic and those not exposed to arsenic

A 3-digit binary code (A) was used to encode the availability for the three unexposed experiment set (B) and the three exposed experiment set (C). The data obtained from the two sets were then used to determine the agreement of availability of raw spectra data for all the six replicates. Of the total 1634 protein spots analyzed, 94.4% unexposed set and 91.4% in exposed set had presence of spectra in all replicates. In the case of the 173 differentially abundant proteins, there was presence of spectra in all the unexposed experiments while in two protein spots no spectra were detected. The probability that this distribution of presence/absence for the differentially abundant proteins was due to chance was tested using the Sign test. The probability was less than 0.0001.

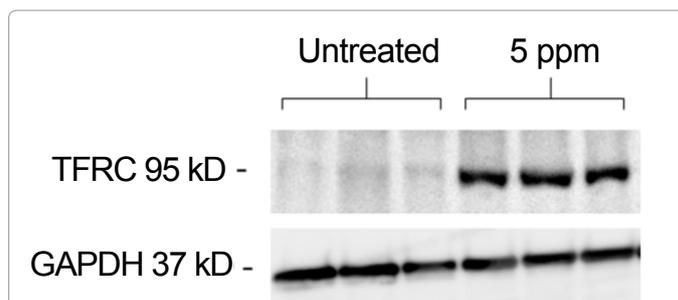


**Figure 4:** Hierarchical clustering of differentially abundant proteins in HaCaT keratinocytes exposed to arsenic.

Clustered heat map of the 173 proteins with significant changes in abundance between arsenic treated (AR1, AR2, AR3) and untreated HaCaT cells (C1, C2, C3) as determined by t-test ( $p < 0.05$ ). An unsupervised cluster of significant proteins indicates a difference between the arsenic treated cells and the control cells. Red data points indicate increase protein abundance, while blue indicates decrease abundance.



**Figure 5:** Relative mRNA expression induced by Arsenic in HaCaT cells. The relative quantification relates the PCR signal of TFRC transcript in arsenic exposed HaCaT cell compared to that of unexposed control HaCaT cells. Actin was used as the house keeping gene. TFRC was up-regulated by arsenic in the exposed HaCaT cells.



**Figure 6:** Western blotting analysis for TFRC proteins separated on a 12% reducing SDS gel. The proteins were transferred to a PVDF membrane and the size of the proteins seen 95 kDa and 37 kDa for GAPDH (House-keeping gene).

blot result is illustrated in Figure 6. Two distinct protein bands are noticeable with sizes of 95 kDa and 37 kDa which correspond with the TFRC and GAPDH (House-keeping gene) respectively.

In the treated HaCaT keratinocytes, 96 proteins showed increase in abundance while 77 proteins showed decrease in abundance. The complete list of proteins is found in the supplementary file. Additionally, a visual analytics resource for exploring and user-defined analysis is available at <http://public.tableausoftware.com/profile/#/>

vizhome/arsenic\_hacat/abstract. We recognize that users of our dataset could prefer to set other cutoff ratio when comparing the relative abundance of proteins between untreated and treated cells. Therefore, we have included filter with sliding bar for values of both the t-test and the cut-off ratio. The interaction with the dataset is available at [http://public.tableausoftware.com/profile/#!/vizhome/arsenic\\_hacat/all\\_reg\\_proteins](http://public.tableausoftware.com/profile/#!/vizhome/arsenic_hacat/all_reg_proteins)

**Identification of arsenic regulated proteins annotated with terms: keratinocyte or skin**

Protein annotation performed using DAVID with proteins annotated with skin tissue as the filter revealed two clusters of 20 up-regulated and 34 down-regulated proteins (Table 2). Two groups of regulated proteins were 5 up-regulated and 5 down-regulated proteins (Table 3). The up-regulated proteins include Calnexin precursor, Isoform 7 of Filamin-B, heterogeneous nuclear, ribonucleoprotein L isoform A, Keratin, type II cytoskeletal 7, Keratin, type I cytoskeletal 19. The down-regulated protein groups were Involucrin, Protein

S100-A2, Protein S100-A8, Protein S100-A9, Isoform Long of 14-3-3 protein beta/alpha.

**Grouping of arsenic regulated proteins by gene ontology biological process terms**

A total of 69 Gene Ontology Biological Process terms were associated with 148 arsenic regulated proteins (Figure 7). The visualization allows from comparison of the number of proteins with increase in abundance versus proteins with decrease in abundance. Numerous decisions can be made from the visualization. However, we decided to identify other arsenic regulated proteins that are grouped in the same biological process as TFRC (IPI00022462). ATPase, Na+/K+ transporting, beta 1 polypeptide (ATP1B1, IPI00747849) was the only other protein with “response to hypoxia” Gene Ontology biological process.

**Proposing major pathways for arsenic toxicity response**

The Gene Ontology Biological Process terms were searched for

	Up-Regulated		Down Regulated
ID_IPI	Protein Name	ID_IPI	Protein Name
IPI00003377	Isoform 1 of Splicing factor, arginine/serine-rich 7	IPI00001580	Isoform 1 of FYVE and coiled-coil domain-containing protein 1
IPI00009407	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1	IPI00003783	Dual specificity mitogen-activated protein kinase kinase 2
IPI00010271	Isoform A of Ras-related C3 botulinum toxin substrate 1 precursor	IPI00005162	Actin-related protein 2/3 complex subunit 3
IPI00011118	Ribonucleoside-diphosphate reductase M2 subunit	IPI00007694	Isoform 1 of Protein phosphatase methylesterase 1
IPI00012750	40S ribosomal protein S25	IPI00009841	CDNA FLJ31747 fis, clone NT2RI2007377, highly similar to RNA-BINDING PROTEIN EWS
IPI00015954	GTP-binding protein SAR1a	IPI00011107	Isocitrate dehydrogenase [NADP], mitochondrial precursor
IPI00021840	40S ribosomal protein S6	IPI00011578	Isoform 1 of Neuroplastin precursor
IPI00023048	Elongation factor 1-delta	IPI00011692	Involucrin
IPI00023101	Protein RCD1 homolog	IPI00013933	Isoform DPI of Desmoplakin
IPI00056334	protein kinase C, delta binding protein	IPI00025084	Calpain small subunit 1
IPI00215768	Glutamate--cysteine ligase catalytic subunit	IPI00216318	Isoform Long of 14-3-3 protein beta/alpha
IPI00219365	Moesin	IPI00219301	Myristoylated alanine-rich C-kinase substrate
IPI00221088	40S ribosomal protein S9	IPI00219526	Isoform 1 of Phosphoglucomutase-1
IPI00290279	Isoform Long of Adenosine kinase	IPI00240675	programmed cell death 4 isoform 2
IPI00293260	Isoform 1 of DnaJ homolog subfamily C member 10 precursor	IPI00293464	DNA damage-binding protein 1
IPI00295992	Isoform 2 of ATPase family AAA domain-containing protein 3A	IPI00307200	Switch-associated protein 70
IPI00329629	DnaJ homolog subfamily C member 7	IPI00412880	Isoform 1 of Histone-arginine methyltransferase CARM1
IPI00384444	Keratin, type I cytoskeletal 14	IPI00001580	Isoform 1 of FYVE and coiled-coil domain-containing protein 1
IPI00414676	Heat shock protein HSP 90-beta	IPI00003783	Dual specificity mitogen-activated protein kinase kinase 2
IPI00449049	Poly [ADP-ribose] polymerase 1	IPI00005162	Actin-related protein 2/3 complex subunit 3
		IPI00007694	Isoform 1 of Protein phosphatase methylesterase 1
		IPI00009841	CDNA FLJ31747 fis, clone NT2RI2007377, highly similar to RNA-BINDING PROTEIN EWS
		IPI00011107	Isocitrate dehydrogenase [NADP], mitochondrial precursor
		IPI00011578	Isoform 1 of Neuroplastin precursor
		IPI00011692	Involucrin
		IPI00013933	Isoform DPI of Desmoplakin
		IPI00025084	Calpain small subunit 1
		IPI00216318	Isoform Long of 14-3-3 protein beta/alpha
		IPI00219301	Myristoylated alanine-rich C-kinase substrate
		IPI00219526	Isoform 1 of Phosphoglucomutase-1
		IPI00240675	programmed cell death 4 isoform 2
		IPI00293464	DNA damage-binding protein 1
		IPI00307200	Switch-associated protein 70
		IPI00412880	Isoform 1 of Histone-arginine methyltransferase CARM1

Protein annotation was performed using DAVID with proteins annotated with skin tissue as the filter, and two clusters of 20 up-regulated and 34 down-regulated proteins were obtained

**Table 2:** Arsenic regulated proteins annotated with the skin tissue.

CLUSTERID	Symbol	ID_IPI	Protein Name	Tissue Annotation
CLUSTER1_Down	IVL	IPI00011692	Involucrin	Keratinocyte, Skin
	S100A2	IPI00019869	Protein S100-A2	Keratinocyte, Mammary epithelium, Ovary
	S100A8	IPI00007047	Protein S100-A8	Ascites, Keratinocyte, Neutrophil, Skeletal muscle, Tongue
	S100A9	IPI00027462	Protein S100-A9	Keratinocyte, Liver, Lung
	YWHAB	IPI00216318	Isoform Long of 14-3-3 protein beta/alpha	Brain, Colon carcinoma, Keratinocyte, Platelet, Skin
CLUSTER2_Up	CANX	IPI00020984	Calnexin precursor	B-cell lymphoma, Epithelium, Fibroblast, Keratinocyte, Kidney, Liver, Lymph, Placenta, Platelet
	FLNB	IPI00477536	Isoform 7 of Filamin-B	Aortic endothelium, Endometrial tumor, Endometrium carcinoma cell line, Endothelial cell, Epithelium, Fetal brain, Keratinocyte, Placenta, Skeletal muscle, Thyroid,
	HNRNPL	IPI00027834	heterogeneous nuclear ribonucleoprotein L isoform a	Brain, Cajal-Retzius cell, Fetal brain cortex, Keratinocyte, Pancreatic cancer, Synovial membrane tissue, Uterus
	KRT7	IPI00306959	Keratin, type II cytoskeletal 7	Colon carcinoma, Epithelium, Keratinocyte, Mammary cancer, Mesothelium, Ovarian carcinoma, Pancreas, PCR rescued clones, Placenta,
	KRT19	IPI00479145	Keratin, type I cytoskeletal 19	Keratinocyte, Lymph node, Mammary gland, Pancreas, Peripheral blood leukocyte, Placenta

Using DAVID, categorization for annotation with keratinocytes yielded two clusters of 5 up-regulated and 5 down-regulated proteins.

**Table 3:** Arsenic regulated proteins annotated with keratinocyte term.

CLUSTERID	GO_TERM	GO_ID	ID_IPI	Protein Name
CLUSTER2_Up	DNA metabolic process	GO:0006259	IPI00011118	Ribonucleoside-diphosphate reductase M2 subunit
CLUSTER2_Up	DNA metabolic process	GO:0006259	IPI00306959	Keratin, type II cytoskeletal 7
CLUSTER2_Up	DNA metabolic process	GO:0006259	IPI00026087	Barrier-to-autointegration factor
CLUSTER2_Up	L-serine metabolic process	GO:0006563	IPI00019178	Phosphoserine phosphatase
CLUSTER2_Up	monosaccharide metabolic process	GO:0005996	IPI00219616	Ribose-phosphate pyrophosphokinase 1
CLUSTER2_Up	monosaccharide metabolic process	GO:0005996	IPI00000684	Isoform AGX2 of UDP-N-acetylhexosamine pyrophosphorylase
CLUSTER2_Up	monosaccharide metabolic process	GO:0005996	IPI00009790	6-phosphofructokinase type C
CLUSTER2_Up	purine nucleotide metabolic process	GO:0006163	IPI00290279	Isoform Long of Adenosine kinase
CLUSTER2_Up	pyrimidine nucleoside metabolic process	GO:0006213	IPI00029631	Enhancer of rudimentary homolog
CLUSTER2_Up	sulfur amino acid metabolic process	GO:0000096	IPI00215768	Glutamate-cysteine ligase catalytic subunit
CLUSTER2_Up	xenobiotic metabolic process	GO:0006805	IPI00012069	NAD

**Table 4:** Proteins in metabolic processes with increased abundance to arsenic trioxide.

CLUSTERID	GO_TERM	GO_ID	ID_IPI	Protein Name
CLUSTER1_Down	acetyl-CoA metabolic process	GO:0006084	IPI00096066	Succinyl-CoA ligase [GDP-forming] beta-chain, mitochondrial precursor
CLUSTER1_Down	acetyl-CoA metabolic process	GO:0006084	IPI00294911	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial precursor
CLUSTER1_Down	cellular aldehyde metabolic process	GO:0006081	IPI00011107	Isocitrate dehydrogenase [NADP], mitochondrial precursor
CLUSTER1_Down	coenzyme metabolic process	GO:0006732	IPI00184821	Bifunctional coenzyme A synthase
CLUSTER1_Down	DNA metabolic process	GO:0006259	IPI00013871	Ribonucleoside-diphosphate reductase large subunit
CLUSTER1_Down	DNA metabolic process	GO:0006259	IPI00018349	DNA replication licensing factor MCM4
CLUSTER1_Down	DNA metabolic process	GO:0006259	IPI00184330	DNA replication licensing factor MCM2
CLUSTER1_Down	fatty acid metabolic process	GO:0006631	IPI00024993	Enoyl-CoA hydratase, mitochondrial precursor
CLUSTER1_Down	membrane lipid metabolic process	GO:0006643	IPI00005745	Serine palmitoyltransferase 1
CLUSTER1_Down	monosaccharide metabolic process	GO:0005996	IPI00550364	Phosphoglucomutase-2
CLUSTER1_Down	monosaccharide metabolic process	GO:0005996	IPI00479186	Isoform M2 of Pyruvate kinase isozymes M1/M2
CLUSTER1_Down	monosaccharide metabolic process	GO:0005996	IPI00219526	Isoform 1 of Phosphoglucomutase-1
CLUSTER1_Down	monosaccharide metabolic process	GO:0005996	IPI00073772	Fructose-1,6-bisphosphatase 1
CLUSTER1_Down	monosaccharide metabolic process	GO:0005996	IPI00063408	dehydrogenase E1 and transketolase domain containing protein 1
CLUSTER1_Down	nucleoside metabolic process	GO:0009116	IPI00219617	Isoform 1 of Ribose-phosphate pyrophosphokinase 2
CLUSTER1_Down	purine nucleobase metabolic process	GO:0006144	IPI00025273	Isoform Long of Trifunctional purine biosynthetic protein adenosine-3
CLUSTER1_Down	purine nucleobase metabolic process	GO:0006144	IPI00029079	GMP synthase
CLUSTER1_Down	regulation of nucleotide metabolic process	GO:0006140	IPI00220578	Guanine nucleotide-binding protein G
CLUSTER1_Down	retinoid metabolic process	GO:0001523	IPI00026663	Aldehyde dehydrogenase 1A3
CLUSTER1_Down	tRNA metabolic process	GO:0006399	IPI00004860	Isoform Complexed of Arginyl-tRNA synthetase, cytoplasmic
CLUSTER1_Down	tRNA metabolic process	GO:0006399	IPI00216951	Aspartyl-tRNA synthetase, cytoplasmic
CLUSTER1_Down	very long-chain fatty acid metabolic process	GO:0000038	IPI00019912	Peroxisomal multifunctional enzyme type 2

A search for candidate proteins involved in metabolic pathways of arsenic toxicity response yielded a total of 33 proteins (11 increased abundance and 22 decreased abundance) associated with 18 metabolic process terms were identified

**Table 5:** Proteins in metabolic processes with decreased abundance to arsenic trioxide

terms that include the word “metabolic” so as to identify major pathways and 22 decreased abundance) associated with 18 metabolic process for arsenic toxicity. A total of 33 proteins (11 increased abundance terms were identified (Tables 4 and 5). Twenty-five of the protein





down) is provided to enable follow-up research on biological processes (Figure 7). We have also provided supplementary resources to aid the several cognitive tasks of our datasets; including a visual analytics resource for exploring the Gene Ontology terms associated with 40,909 Gene Ontology Identifiers. We identified a need to expand the number of genes annotated with the GO term response to arsenic-containing substance (GO:0046685).

One of our interests was to identify proteins annotated with tissue term keratinocyte. Therefore, this discussion includes the relevance of two groups of regulated proteins consisting of 5 up-regulated and 5 down-regulated proteins (Table 3). The up-regulated proteins include Calnexin precursor; Isoform 7 of Filamin-B; heterogeneous nuclear; ribonucleoprotein L isoform A; Keratin, type II cytoskeletal 7; and Keratin, type I cytoskeletal 19. The down-regulated proteins were Involucrin, Protein S100-A2, Protein S100-A8, Protein S100-A9, and Isoform Long of 14-3-3 protein beta/alpha.

Calnexin is an endoplasmic reticulum (ER) transmembrane chaperone involved in translocation, in protein folding, and in the quality control of newly synthesized polypeptides [56-58]. There are no reported studies on direct calnexin-arsenic interaction and since calnexin plays critical cellular roles, any perturbation of its expression may have detrimental consequences [59,60]. Filamins (FLNs) are large actin-binding, cross-linking and scaffolding proteins. They stabilize the membrane by anchoring the transmembrane proteins and are implicated in signal transduction, through interaction with ion channels, receptors and signaling proteins [61,62]. Mutations and perturbation in the FLNA and FLNB genes are known to cause different developmental disorders in humans, such as bone anomalies, periventricular heterotopia, aortic dissection and aneurysm [62,63]. Tumor marker proteins including stress-activated protein kinase/c-Jun N-terminal kinase (SAPK/JNK) that belongs to the mitogen activated protein kinase (MAPK) are known to be up-regulated in HaCaT cells exposed to arsenic [63]. SAPK/JNK and MAPK are associated with reactive oxygen species induction and other signal transduction. Filamin links two MAPKs together and promote JNK activation [64].

Heterogeneous nuclear ribonucleoprotein L isoform A (HNRNPL) is a splicing regulatory protein [65] which is up-regulated in our work. It has also been implicated in other cancers such as esophageal squamous cell carcinoma [66]. Understanding the molecular mechanisms of HNRNPL induction and its role during keratinocyte transformation by arsenic will be important since they function as a docking protein for DNA, RNA, and transcriptional molecules [67].

Keratin, type II cytoskeletal 7 and Keratin, type I cytoskeletal 19 were up-regulated in this study. Keratins are major components of structural proteins of the epidermis and are one of the most abundant proteins in epithelial cells. Alterations in keratin expression are associated with the development of skin pathology, including hyperkeratosis, and skin cancers [68,69]. Keratinocytes in the epidermis express a differentiation-specific set of type I and type III keratins which form a stable network and contribute to keratinocyte mechanical properties [70]. Mapping of the cytokeratin profile of the special type of skin pathology may be crucial in typing of the stages of the disease as has been attempted with breast carcinoma [71].

Involucrin (IVL) is a marker of keratinocyte differentiation and its down regulation as observed in our study has been related to various disease conditions. Involucrin function was perturbed in skin squamous cell carcinoma (SCC) [72], and cervical cancer [73]. The

protein levels were down-regulated by human papilloma virus (HPV) oncogenes in human foreskin keratinocyte (HFK) cells [74]. Arsenate suppresses IVL expression, and, keratinocyte programming, probably by altering the transcription factors AP1 (activating protein-1) and AP2 (activating protein-2) [75].

S100 proteins are a group of proteins that are involved in the regulation of a number of cellular processes such as cell cycle progression and differentiation. They are found in the cytoplasm and sometimes in the nucleus of a wide range of cells. The tissue annotation profile in Figure 6 revealed that the down-regulated S100 calcium binding proteins, especially A100A2 and S100A8 were associated with keratinocytes. We observed down-regulation of Protein S100-A2, Protein S100-A8, and Protein S100-A9. Protein S100-A8 and Protein S100-A9 are suggested as markers for early diagnosis of necrotizing enterocolitis in neonates [76]. Protein S100-A2, is an EF hand calcium-binding protein, that is perturbed in several cancers and is also a TGF- $\beta$  (transforming growth factor- $\beta$ )-regulated gene in melanoma and lung cancer cells [77]. S100A2 was identified as putative tumor suppressor gene in human breast cancer cell lines [78]. In other words its suppression by arsenic will allow tumor cell to replicate unabated. Altered expression and chromosomal rearrangements of S100 proteins have been implicated in breast cancer [79].

Isoform Long of 14-3-3 protein beta/alpha, though down-regulated in this study was previously observed to be over expressed in cyst fluid from cyst-associated renal cell carcinoma. Its differential expression makes it a potential urinary biomarker for renal cell carcinoma [80]. The perturbation of Isoform Long of 14-3-3 protein beta/alpha may suggest its involvement in HaCaT cells carcinogenesis or response to arsenic and it will be very interesting to understand its roles more.

We previously predicted that arsenic will interact with TFRC based on the presence of vicinal cysteines [80]. Thus, after the observed up-regulation of Transferrin receptor protein 1 (TFRC) in the proteomics data, we conducted western blot assay (Figure 4) to build additional evidence that arsenic exposure increases the abundance of TFRC and possibly increase iron transport into keratinocytes. A comparison of tissue annotation profiles and integration of datasets on proteomic response, tissue annotation (Figure 6) showed that TFRC, an up-regulated Cluster of Differentiation (CD) molecule, is associated with the brain, erythroleukemia, eye, human endometrium carcinoma cell line, kidney, pancreas, placenta and prostatic carcinoma. A previous study had reported increased expression of transferrin receptor in cancer cells than in normal cells [81]. TFRC has been suggested as a potential biomarker in the diagnosis of early onset of colorectal cancer [82]. A proteomics analysis showed that TFRC is highly expressed in inflammatory breast cancer [83].

We observed in our previous study that low dose arsenic triggers proliferation of HaCaT cells [84]. The cell proliferation could be linked to the expression of TFRC. This observation was corroborated by Ryschich et al. who reported TFRC as a membrane-bound protein expressed in larger amounts in proliferating, human pancreatic cancer cells than in quiescent cells [85]. We have demonstrated increased expression of TFRC in HaCaT cells exposed to arsenic trioxide using proteomic, transcriptomics and western blot. Arsenic induces skin color disorders such as hyper and hypopigmentation [86]. TFRC is a carrier protein for iron which is a crucial component of melanin. TFRC may also be involved in the transfer of melanin from melanocytes to other keratinocytes using the transferrin receptor-mediated endocytosis pathway [87,88]. Increased melanocyte dendricity is correlated with hyperpigmentation [38]. Disruption of the iron transfer

system by arsenic could lead to skin discoloration pathologies such as hyperpigmentation and hypopigmentation.

The transferrin receptor expression is modulated by hypoxia. In rats, hypoxia increases expression of transferrin receptor leading to increased cellular uptake of iron in the pineal gland [89]. Also in humans, hypoxia is a side effect of arsenic trioxide treatment [90]. The skin is normally mildly hypoxic ranging from 1.5% to 5% oxygen [90]. The modulation of hypoxic levels in the keratinocytes by arsenic leading to accumulation of iron species warrants further research.

A total of 33 proteins (11 increased abundance and 22 decreased abundance) associated with 18 metabolic process terms were identified (Table 4). The inclusion of proteins involved in the mitochondrial metabolic function is consistent with arsenic suppressing mitochondrion function through binding with thiol-containing enzymes on the mitochondrion membrane [91]. The Glutamate--cysteine ligase catalytic subunit (GCLC) is one of the oxidative stress defense genes regulated by Transcription factor NF-E2-related factor 2 (Nrf2) [92]. In HaCaT keratinocytes, *Gclc* gene expression is induced by arsenite [93]. This prior finding provided independent confirmation of the differentially abundant gene list reported here. Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial precursor (SDHB) had decreased abundance after arsenic exposure compared to unexposed cells (Table 4). The SDHB is part of the catalytic domain of succinate dehydrogenase, an enzyme complex that functions in the citric acid cycle and the electron transport chain [94]. We inferred from the presence of sulfur in the subunit that arsenic could bind to SDHB and interfere with its function. In fact, defect in SDHB induces tumor formation due to succinate accumulation [95,96].

In conclusion, a list of 173 protein altered by arsenic trioxide were grouped using three major functional annotations covering tissue localization, biological process and protein family. A possible explanation for hyperpigmentation pathologies observed in arsenic toxicity is that arsenic exposure leads to increased iron uptake in the normally hypoxic human skin. The proteins mapped to metabolic process terms and differentially abundant are candidates for evaluating metabolic pathways perturbed by arsenicals.

## Funding

Funding was provided by RCMI Center for Environmental Health [National Center for Research Resources (5 G12 RR013459) & National Institute on Minority Health and Health Disparities (8 G12 MD007581)], National Institutes of Health grants R01DA025755, 1 T36 GM095335; and the National Science Foundation EPSCoR EPS-0903787. AJT acknowledges the University of Arkansas Medical Sciences (UAMS) Proteomics Facility for mass spectrometric support and funding from the National Institutes of Health grants R01GM106024, R33CA173264, P30GM103450, P20GM103429 and UL1TR000039. RDI acknowledges funding support from the Office of the Provost, Bethune-Cookman University, Daytona Beach, Florida.

## References

- Schulz EJ (1967) Arsenic as a cause of skin cancer, with notes on its occurrence in Pretoria. *S Afr Med J* 41: 819-822.
- Cöl M, Cöl C, Soran A, Sayli BS, Oztürk S (1999) Arsenic-related Bowen's disease, palmar keratosis, and skin cancer. *Environ Health Perspect* 107: 687-689.
- Yu HS, Chen GS, Sheu HM, Kao JS, Chang KL, et al. (1992) Alterations of skin-associated lymphoid tissue in the carcinogenesis of arsenical skin cancer. *Proc Natl Sci Counc Repub China B* 16: 17-22.
- Ho SY, Tsai YC, Lee MC, Guo HR (2005) Merkel cell carcinoma in patients with long-term ingestion of arsenic. *J Occup Health* 47: 188-192.
- Bao L, Shi H (2010) Potential molecular mechanisms for combined toxicity of arsenic and alcohol. *J Inorg Biochem* 104: 1229-1233.
- McNeely SC, Xu X, Taylor BF, Zacharias W, McCabe MJ Jr, et al. (2006) Exit from arsenite-induced mitotic arrest is p53 dependent. *Environ Health Perspect* 114: 1401-1406.
- Udensi UK, Cohly HH, Graham-Evans BE, Ndebele K, Garcia-Reyero N, et al. (2011) Aberrantly Expressed Genes in HaCaT Keratinocytes Chronically Exposed to Arsenic Trioxide. *BiomarkInsights* 6: 7-16.
- Sasaki A, Oshima Y, Fujimura A (2007) An approach to elucidate potential mechanism of renal toxicity of arsenic trioxide. *Exp Hematol* 35: 252-262.
- Liu J, Xie Y, Ducharme DM, Shen J, Diwan BA, et al. (2006) Global gene expression associated with hepatocarcinogenesis in adult male mice induced by in utero arsenic exposure. *Environ Health Perspect* 114: 404-411.
- Chen CJ, Chiou HY, Chiang MH, Lin LJ, Tai TY (1996) Dose-response relationship between ischemic heart disease mortality and long-term arsenic exposure. *Arterioscler Thromb Vasc Biol* 16: 504-510.
- Argos M, Kibriya MG, Parvez F, Jasmine F, Rakibuz-Zaman M, et al. (2006) Gene expression profiles in peripheral lymphocytes by arsenic exposure and skin lesion status in a Bangladeshi population. *Cancer Epidemiol Biomarkers Prev* 15: 1367-1375.
- Wlodarczyk BJ, Cabrera RM, Hill DS, Bozinov D, Zhu H, et al. (2006) Arsenic-induced gene expression changes in the neural tube of folate transport defective mouse embryos. *Neurotoxicology* 27: 547-557.
- Su PF, Hu YJ, Ho IC, Cheng YM, Lee TC (2006) Distinct gene expression profiles in immortalized human urothelial cells exposed to inorganic arsenite and its methylated trivalent metabolites. *Environ Health Perspect* 114: 394-403.
- Kuo CC, Liu TW, Chen LT, Shiah HS, Wu CM, et al. (2011) Combination of arsenic trioxide and BCNU synergistically triggers redox-mediated autophagic cell death in human solid tumors. *Free Radic Biol Med* 51: 2195-2209.
- Rehman K, Naranmandura H (2013) Double-edged effects of arsenic compounds: anticancer and carcinogenic effects. *Curr Drug Metab* 14: 1029-1041.
- Rogers CS, Yedjou CG, Sutton DJ, Tchounwou PB (2014) Vitamin D3 potentiates the antitumorigenic effects of arsenic trioxide in human leukemia (HL-60) cells. *Exp Hematol Oncol* 3: 9.
- Yedjou C, Tchounwou P, Jenkins J, McMurray R (2010) Basic mechanisms of arsenic trioxide (ATO)-induced apoptosis in human leukemia (HL-60) cells. *J Hematol Oncol* 3: 28.
- Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, et al. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292: 929-934.
- van Ommen B, Stierum R (2002) Nutrigenomics: exploiting systems biology in the nutrition and health arena. *Curr Opin Biotechnol* 13: 517-521.
- Byrum S, Avaritt NL, Mackintosh SG, Munkberg JM, Badgwell BD, et al. (2011) A quantitative proteomic analysis of FFPE melanoma. *J Cutan Pathol* 38: 933-936.
- Hamada T, Horiguchi S (1976) [Occupational chronic arsenical poisoning. On the cutaneous manifestations (author's transl)]. *Sangyo Igaku* 18: 103-115.
- Huang da W, Sherman BT, Tan Q, Kir J, Liu D, et al. (2007) DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* 35: W169-W175.
- Thomas JJ, Cook KA (2006) A visual analytics agenda. *IEEE Comput Graph Appl* 26: 10-13.
- Keim DA, Mansmann FaS, Ziegler H (2008) Visual Analytics: Definition, Process, and Challenges. Lecture Notes. In: Kerren A, Stasko JT, Fekete JD, North C (Eds) Information Visualization. Springer-Verlag Berlin, Heidelberg.
- (2014) Oral presentations. *Luminescence* 29 Suppl 1: 6-55.
- Magrane M, Consortium U (2011) Uni Prot Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011: bar009.
- Huang da W, Sherman BT, Zheng X, Yang J, Imamichi T, et al. (2009)

- Extracting biological meaning from large gene lists with DAVID. *Curr Protoc Bioinformatics* Chapter 13: Unit 13.
28. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, et al. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4: P3.
29. Stekel D (2003) *Microarray Bioinformatics*. Cambridge University Press, New York.
30. UniProt Consortium (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res*. 42(Database issue): D191-D198.
31. Gene Ontology Consortium, Blake JA, Dolan M, Drabkin H, Hill DP, et al. (2013) Gene Ontology annotations and resources. *Nucleic Acids Res* 41: D530-D535.
32. Huntley RP, Sawford T, Martin MJ, O'Donovan C (2014) Understanding how and why the Gene Ontology and its annotations evolve: the GO within UniProt. *Gigascience* 3: 4.
33. Gray KA, Daugherty LC, Gordon SM, Seal RL, Wright MW, et al. (2013) Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res* 41: D545-552.
34. Keim D, Andrienko G, Fekete JD, Görg C, Kohlhammer J, et al. (2008) Visual Analytics: Definition, Process, and Challenges. *Information Visualization Lecture Notes in Computer Science* 4950: 154-175.
35. Isokpehi RD, Cohly HH, Anyanwu MN, Rajnarayanan RV, Tchounwou PB, et al. (2010) Candidate single nucleotide polymorphism markers for arsenic responsiveness of protein targets. *Bioinform Biol Insights* 4: 99-111.
36. Isokpehi RD, Udensi UK, Anyanwu MN, Mbah AN, Johnson MO, et al. (2012) Knowledge building insights on biomarkers of arsenic toxicity to keratinocytes and melanocytes. *Biomark Insights* 7: 127-141.
37. McClelland A, Kühn LC, Ruddle FH (1984) The human transferrin receptor gene: genomic organization, and the complete primary structure of the receptor deduced from a cDNA sequence. *Cell* 39: 267-274.
38. Schneider C, Owen MJ, Banville D, Williams JG (1984) Primary structure of human transferrin receptor deduced from the mRNA sequence. *Nature* 311: 675-678.
39. Steere AN, Chasteen ND, Miller BF, Smith VC, MacGillivray RT, et al. (2012) Structure-based mutagenesis reveals critical residues in the transferrin receptor participating in the mechanism of pH-induced release of iron from human serum transferrin. *Biochemistry* 51: 2113-2121.
40. Bedogni B, Welford SM, Cassarino DS, Nickoloff BJ, Giaccia AJ, et al. (2005) The hypoxic microenvironment of the skin contributes to Akt-mediated melanocyte transformation. *Cancer Cell* 8: 443-454.
41. Trouba KJ, Glanzer JG, Vorce RL (1999) Wild-type and Ras-transformed fibroblasts display differential mitogenic responses to transient sodium arsenite exposure. *Toxicol Sci* 50: 72-81.
42. Analysis of Complex Proteomic Dataset Using Scaffold (2014).
43. Chaerkady R, Pandey A (2007) Quantitative proteomics for identification of cancer biomarkers. *Proteomics Clin Appl* 1: 1080-1089.
44. Zybailov B, Mosley AL, Sardu ME, Coleman MK, Florens L, et al. (2006) Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res* 5: 2339-2347.
45. Florens L, Carozza MJ, Swanson SK, Fournier M, Coleman MK, et al. (2006) Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* 40: 303-311.
46. Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, et al. (2006) Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci U S A* 103: 18928-18933.
47. Wu Q, Zhao Q, Liang Z, Qu Y, Zhang L, et al. (2012) NSI and NSMT: usages of MS/MS fragment ion intensity for sensitive differential proteome detection and accurate protein fold change calculation in relative label-free proteome quantification. *Analyst* 137: 3146-3153.
48. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, et al. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 55: 611-622.
49. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-ΔΔC<sub>T</sub></sup> Method. *Methods* 25: 402-408.
50. Johnson MO, Cohly HH, Isokpehi RD, Awofolu OR (2010) The case for visual analytics of arsenic concentrations in foods. *Int J Environ Res Public Health* 7: 1970-1983.
51. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25-29.
52. Berglund SR, Santana AR, Li D, Rice RH, Rocke DM, et al. (2009) Proteomic analysis of low dose arsenic and ionizing radiation exposure on keratinocytes. *Proteomics* 9: 1925-1938.
53. Bernstam L, Lan CH, Lee J, Nriagu JO (2002) Effects of arsenic on human keratinocytes: morphological, physiological, and precursor incorporation studies. *Environ Res* 89: 220-235.
54. Balakrishnan R, Harris MA, Huntley R, Van Auken K, Cherry JM (2013) A guide to best practices for Gene Ontology (GO) manual annotation. *Database (Oxford)* 2013: bat054.
55. Davis AP, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, et al. (2013) The Comparative Toxicogenomics Database: update 2013. *Nucleic Acids Res* 41: D1104-1114.
56. Guérin R, Beaugard PB, Leroux A, Rokeach LA (2009) Calnexin regulates apoptosis induced by inositol starvation in fission yeast. *PLoS One* 4: e6244.
57. Williams DB (2006) Beyond lectins: the calnexin/calreticulin chaperone system of the endoplasmic reticulum. *J Cell Sci* 119: 615-623.
58. Bukau B, Deuring E, Pfund C, Craig EA (2000) Getting newly synthesized proteins into shape. *Cell* 101: 119-122.
59. Jannatipour M, Rokeach LA (1995) The *Schizosaccharomyces pombe* homologue of the chaperone calnexin is essential for viability. *J Biol Chem* 270: 4845-4853.
60. Guérin R, Arseneault G, Dumont S, Rokeach LA (2008) Calnexin is involved in apoptosis induced by endoplasmic reticulum stress in the fission yeast. *Mol Biol Cell* 19: 4404-4420.
61. Wang Q, Dai XQ, Li Q, Wang Z, Cantero Mdel R, et al. (2012) Structural interaction and functional regulation of polycystin-2 by filamin. *PLoS One* 7: e40448.
62. Stosel TP, Condeelis J, Cooley L, Hartwig JH, Noegel A, et al. (2001) Filamins as integrators of cell mechanics and signalling. *Nat Rev Mol Cell Biol* 2: 138-145.
63. Pilop C, Aregger F, Gorman RC, Brunisholz R, Gerrits B, et al. (2009) Proteomic analysis in aortic media of patients with Marfan syndrome reveals increased activity of calpain 2 in aortic aneurysms. *Circulation* 120: 983-991.
64. Nakagawa K, Sugahara M, Yamasaki T, Kajiho H, Takahashi S, et al. (2010) Filamin associates with stress signalling kinases MKK7 and MKK4 and regulates JNK activation. *Biochem J* 427: 237-245.
65. Shankarling G, Lynch KW (2013) Minimal functional domains of paralogues hnRNP L and hnRNP LL exhibit mechanistic differences in exonic splicing repression. *Biochem J* 453: 271-279.
66. Qi YJ, He QY, Ma YF, Du YW, Liu GC, et al. (2008) Proteomic identification of malignant transformation-related proteins in esophageal squamous cell carcinoma. *J Cell Biochem* 104: 1625-1635.
67. Gao FH, Wu YL, Zhao M, Liu CX, Wang LS, et al. (2009) Protein kinase C-delta mediates down-regulation of heterogeneous nuclear ribonucleoprotein K protein: involvement in apoptosis induction. *Exp Cell Res* 315: 3250-3258.
68. Sullivan JA, Dachelet CZ, Sultz HA, Henry M (1978) The rural nurse practitioner: a challenge and a response. *Am J Public Health* 68: 972-976.
69. Tseng WP, Chu HM, How SW, Fong JM, Lin CS, et al. (1968) Prevalence of skin cancer in an endemic area of chronic arsenicism in Taiwan. *J Natl Cancer Inst* 40: 453-463.
70. Ramms L, Fabris G, Windoffer R, Schwarz N, Springer R, et al. (2013) Keratins as the main component for the mechanical integrity of keratinocytes. *Proc Natl Acad Sci U S A* 110: 18513-18518.
71. Toth T (1999) Cytokeratins 7 and 14 in special types of invasive breast carcinomas. *Med Pregl* 52: 227-232.

72. Lan YJ, Chen H, Chen JQ, Lei QH, Zheng M, et al. (2013) Immunolocalization of Vimentin, Keratin 17, Ki-67, Involucrin,  $\beta$ -Catenin and E-Cadherin in Cutaneous Squamous Cell Carcinoma. *Pathol Oncol Res* .
73. Santin AD, Zhan F, Bignotti E, Siegel ER, Cané S, et al. (2005) Gene expression profiles of primary HPV16- and HPV18-infected early stage cervical cancers and normal cervical epithelium: identification of novel candidate molecular markers for cervical cancer diagnosis and therapy. *Virology* 331: 269-291.
74. Gyöngyösi E, Szalmás A, Ferenczi A, Kónya J, Gergely L, et al. (2012) Effects of human papillomavirus (HPV) type 16 oncoproteins on the expression of involucrin in human keratinocytes. *Viol J* 9: 36.
75. Kachinskas DJ, Qin Q, Phillips MA, Rice RH (1997) Arsenate suppression of human keratinocyte programming. *Mutat Res* 386: 253-261.
76. Terrin G, Passariello A, De Curtis M, Paludetto R, Berni Canani R (2012) S100 A8/A9 protein as a marker for early diagnosis of necrotising enterocolitis in neonates. *Arch Dis Child* 97: 1102.
77. Naz S, Ranganathan P, Bodapati P, Shastry AH, Mishra LN, et al. (2012) Regulation of S100A2 expression by TGF- $\beta$ -induced MEK/ERK signalling and its role in cell migration/invasion. *Biochem J* 447: 81-91.
78. Lee SW, Tomasetto C, Swisshelm K, Keyomarsi K, Sager R (1992) Down-regulation of a member of the S100 gene family in mammary carcinoma cells and reexpression by azadeoxycytidine treatment. *Proc Natl Acad Sci U S A* 89: 2504-2508.
79. Zimmer DB, Cornwall EH, Landar A, Song W (1995) The S100 protein family: history, function, and expression. *Brain Res Bull* 37: 417-429.
80. Minamida S, Iwamura M, Kodera Y, Kawashima Y, Tabata K, et al. (2011) 14-3-3 protein beta/alpha as a urinary biomarker for renal cell carcinoma: proteomic analysis of cyst fluid. *Anal Bioanal Chem* 401: 245-252.
81. Gatter KC, Brown G, Trowbridge IS, Woolston RE, Mason DY (1983) Transferrin receptors in human tissues: their distribution and possible clinical relevance. *J Clin Pathol* 36: 539-545.
82. Thorsen SB, Lundberg M, Villablanca A, Christensen SL, Belling KC, et al. (2013) Detection of serological biomarkers by proximity extension assay for detection of colorectal neoplasias in symptomatic individuals. *J Transl Med* 11: 253.
83. Zhang EY, Cristofanilli M, Robertson F, Reuben JM, Mu Z, et al. (2013) Genome wide proteomics of ERBB2 and EGFR and other oncogenic pathways in inflammatory breast cancer. *J Proteome Res* 12: 2805-2817.
84. Udensi UK, Graham-Evans BE, Rogers C, Isokpehi RD (2011) Cytotoxicity patterns of arsenic trioxide exposure on HaCaT keratinocytes. *Clin Cosmet Investig Dermatol* 4: 183-190.
85. De Leon-Conti G (1966) Administrative problems concerning the orientation of students. *St Tomas Nurs J* 4: 314-318.
86. Otleş S, Çağindi O (2010) Health importance of arsenic in drinking water and food. *Environ Geochem Health* 32: 367-371.
87. Qian ZM, Li H, Sun H, Ho K (2002) Targeted drug delivery via the transferrin receptor-mediated endocytosis pathway. *Pharmacol Rev* 54: 561-587.
88. Ando H, Niki Y, Ito M, Akiyama K, Matsui MS, et al. (2012) Melanosomes are transferred from melanocytes to keratinocytes through the processes of packaging, release, uptake, and dispersion. *J Invest Dermatol* 132: 1222-1229.
89. Kaur C, Sivakumar V, Ling EA (2007) Expression of transferrin receptors in the pineal gland of postnatal and adult rats and its alteration in hypoxia and melatonin treatment. *Glia* 55: 263-273.
90. Zhang TD, Chen GQ, Wang ZG, Wang ZY, Chen SJ, et al. (2001) Arsenic trioxide, a therapeutic agent for APL. *Oncogene* 20: 7146-7153.
91. Ramanathan K, Shila S, Kumaran S, Panneerselvam C (2003) Ascorbic acid and alpha-tocopherol as potent modulators on arsenic induced toxicity in mitochondria. *J Nutr Biochem* 14: 416-420.
92. Pi J, Diwan BA, Sun Y, Liu J, Qu W, et al. (2008) Arsenic-induced malignant transformation of human keratinocytes: involvement of Nrf2. *Free Radic Biol Med* 45: 651-658.
93. Schuliga M, Chouchane S, Snow ET (2002) Upregulation of glutathione-related genes and enzyme activities in cultured human cells by sublethal concentrations of inorganic arsenic. *Toxicol Sci* 70: 183-192.
94. Bayley JP, Devilee P, Taschner PE (2005) The SDH mutation database: an online resource for succinate dehydrogenase sequence variants involved in pheochromocytoma, paraganglioma and mitochondrial complex II deficiency. *BMC Med Genet* 6: 39.
95. Brière JJ, Favier J, Bénil P, El Ghouzzi V, Lorenzato A, et al. (2005) Mitochondrial succinate is instrumental for HIF1alpha nuclear translocation in SDHA-mutant fibroblasts under normoxic conditions. *Human Molecular Genetics* 14: 3263-3269.
96. Hong Y, Piao F, Zhao Y, Li S, Wang Y, et al. (2009) Subchronic exposure to arsenic decreased Sdha expression in the brain of mice. *Neurotoxicology* 30: 538-543.