ANALYSIS OF GLOBAL PROTEIN PROFILE OF HEART FROM CLONED PIGLETS

Thesis submitted in partial fulfillment of the requirements for the degree of

Masters in Technology

in

BIOTECHNOLOGY

by

Lipsa Patnaik

Under the guidance of **Prof. Mukesh Kumar Gupta**

Department of Biotechnology and Medical Engineering

National Institute of Technology, Rourkela

Odisha

Department of Biotechnology and Medical Engineering

National Institute of Technology, Rourkela

Odisha

CERTIFICATE

This is to certify that the work in the thesis entitled "ANALYSIS OF GLOBAL PROTEIN PROFILE OF HEART FROM CLONED PIGLETS**"** submitted by Ms. Lipsa Patnaik, in partial fulfillment of the requirements for the award of M. Tech (Biotechnology) at the National Institute of Technology Rourkela, is an authentic work performed by her under my supervision and guidance.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any University/Institute for the award of any Degree or Diploma.

Prof. Mukesh Kumar Gupta Date: 01 June 2013

Department of Biotechnology and Medical Engineering

National Institute of Technology, Rourkela

Department of Biotechnology and Medical Engineering

National Institute of Technology, Rourkela

Odisha

DECLARATION

I hereby declare that the project work titled "ANALYSIS OF GLOBAL PROTEIN PROFILE OF HEART FROM CLONED PIGLETS**"**, submitted to the National Institute of Technology, Rourkela, is a record of an original work done by me under the guidance of Prof. Mukesh Kumar Gupta, and that this project work has not performed the basis for the award of any other degree or diploma, if any.

Lipsa Patnaik

Roll No. 211BM2011

Department of Biotechnology and Medical Engineering

National Institute of Technology, Rourkela

ACKNOWLEDGEMENT

The author would like to acknowledge Dr. Sang Jun Uhm (Konkuk University, South Korea) for providing the tissue samples, Prof. Kwang Pyo Kim (Institute of Bio-Medical Science and Technology, South Korea) for mass spectrometry and Prof. Hoon Taek Lee (Bio-Organ Research Center, South Korea) for RT-PCR analysis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

.

ABBREVIATIONS

2D-DIGE	Two-dimensional Differential Gel Electrophoresis
DAVID	Database for Annotation, Visualization and Integrated Discovery
EASE	Expression Analysis Systematic Explorer
ESI	Electron Spray Ionization
GO	Gene Ontology
KDD	Knowledge Discovery in Databases
LC	Liquid Chromatography
MALDI-TOF	Matrix-Assisted Laser Desorption/ Ionization – Time Of Flight
MS	Mass Spectrometry
NCBI	National Centre for Biotechnology Information
NIAID	National Institute of Allergy and Infectious Disease
NIH	National Institute of Health
PERL	Practical Extraction and Report Language
PIMA	Pattern-Induced Multi-Sequence Alignment
RT-PCR	Reverse- Transcriptase Polymerase Chain Reaction
SCNT	Somatic Cell Nuclear Transfer
SELDI	Surface-Enhanced Laser Desorption/ Ionization
SQL	Structured Query Language
TCA	Tri-Carboxylic Acid

ABSTRACT

Somatic cell nuclear transfer (SCNT) has wide application in transgenesis, conservation of biodiversity, generation of patient-specific autologous stem cells and biomedical engineering. This technology, however, could not be exploited to its full potential due to the nuclear reprogramming errors and high incidence of phenotypic abnormalities and perinatal deaths in SCNT clones. Although investigations on these deaths, chiefly peri-natal and early post-natal, hypothesize aberrant epigenetic programming as the underlying cause, no confirmatory studies are available. Many reports on the deaths have pointed out abnormalities in cardiovascular system in the cloned species. This work aimed at comparative protein expression of heart tissues obtained from three piglet samples, namely those born out of normal conception (NC), the live SCNT clones (CL) and the dead SCNT clones (CD). Global protein profiling data from these three sample sets were obtained from high-throughput proteomic approach of LC-MS/MS and differentially expressed proteins were analyzed. The protein expression data were annotated and analyzed using data mining approaches, and a pattern in proteins showing abnormal expression was hypothesized. Analysis revealed that most of the proteins involved in basal metabolism were down-regulated in the cloned species. This may cause aberrant metabolism, leading to death. A number of proteins pertaining to signaling pathways involved in heart formation were also found to be either down-regulated or the expression level was undetectable in cloned samples. Thus, it could be hypothesized that incomplete or aberrant signaling processes could affect the formation of the heart and therefore, result in peri-natal deaths. Structural proteins of the heart, and proteins involved in heart function were also differentially expressed, pointing at loss of structural integrity and/or functional competency of the heart resulting in early deaths. Taken together, this study revealed that heart of SCNT clones was both developmentally and functionally insufficient

with gross abnormalities of basal metabolism. Further studies should validate these data to decipher novel ways to circumvent these issues and improve the efficiency of the SCNT technology.

Key Words: SCNT, proteomics, LC-MS/MS, data mining, KDD, DAVID, mass spectrometry.

Chapter One

INTRODUCTION

Somatic cell nuclear transfer (SCNT) technology has been a major breakthrough in biomedical engineering for the derivation of patient-specific stem cells, transgenesis, conservation of biodiversity and animal cloning [1, 2]. The technology has been widely used for rapid propagation of genetically superior animals, resurrection of endangered species, production of pharmaceutically important proteins and preservation of germplasm through somatic cell banking. In humans, SCNT has been used to evaluate the development potential of gamete and embryos. Another major application of SCNT is the study of reprogramming events occurring during cell differentiation [3, 4]. However, despite successful production of healthy clones in a number of species, including mice, sheep, cattle, goats, pigs, rabbits, cats and dogs, the efficiency of the process has been very low due to incomplete or aberrant nuclear reprogramming [5,6]. Even though SCNT embryos are capable of normal embryonic development machinery, yet they have much lower rates of development, ultimately leading to embryonic and foetal deaths [7, 8]. Furthermore, clones occasionally show phenotypic abnormalities and high incidence of perinatal and early post-natal deaths accompanied with gestational abnormalities such as placental anomalies, large offspring syndrome and anomalies of umbilical cord [9, 10, 11]. In particular, organs of cardiovascular and respiratory systems commonly show incomplete organogenesis and functional inefficiency [12]. Despite numerous speculations, hypotheses and analytical research, the underlying causes of these abnormalities and deaths are, however, largely unknown.

Proteomic studies deal with large-scale analysis of all proteins expressed in a biological system in a given space and time [13]. It can not only provide qualitative and quantitative information of the global protein expression, but also allows the detailed analysis of complex biological network of interconnected signaling pathways in different biological phenomena. The ultimate goal of

proteomics is to create a complete three-dimensional map of cells in a system, indicating protein location, functions and interactions. Large-scale global protein profiling through various high-throughput techniques has also enabled the simultaneous evaluation of multiple proteins in SCNT clones and has been useful in divulging the aberrancy during embryonic and fetal growth. Application of mass spectrometry (MS) methods such as matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) and MS/MS combined with multi-dimensional liquid chromatography (LC-MS/MS) and gel-based fluorescence two-dimensional differential gel electrophoresis (2D-DIGE) methods on cellular lysate of SCNT clones have revealed the role of mitochondrial protein levels in developmental losses [14], and the role of spindle composition and formation [15], placental dysfunction [16], etc. in developmental abnormalities. However, a large scale sequencing, identification and expression analysis of global proteins in heart tissue of SCNT clones has not been investigated. Furthermore, most studies have utilized MALDI-TOF or gel-based proteomic approaches which suffer from certain shortcomings. While in MALDI-TOF, the relationship between the amount of analyte present and measured signal intensity is complex and therefore, signal intensity of a peptide ion does not directly indicate the absolute amount of protein present and it can be used to identify relatively simple peptide mixtures [17], gel-based proteomics remain a low throughput approach that requires a relatively large amount of sample. LC-MS/MS allows sequencing of complex peptide samples and enables multi-dimensional protein identification. However, data on LC-MS/MS analysis of heart tissue are scarce.

A requirement of large scale proteomic approaches is to analyze, annotate and mine the voluminous data to decipher the useful information. Data mining refers to the process of exploration and analysis by computational or semi-computational means of large data sets with

an aim to determine or predict implicit, previously unknown potentially useful correlations or discover some hidden pattern or trend in the existing data, without any *a priori* knowledge of the biological interpretation of the data. Although data mining is one of the major steps of Knowledge Discovery in Databases (KDD), the two terms are often used interchangeably. Data mining involves various forms of interpretation and information harvesting, including predictive classification, descriptive clustering and cluster analysis, deviation/ anomaly detection, predictive regression, sequential pattern discovery and association rule discovery. The objective of data mining is to classify large amounts of data and form corresponding hypotheses in order to be able to build a predictive model of the underlying biological process to reduce the need for experimentation. The application of data mining techniques is now being extended to the prediction of protein interactions and protein modifications, including their dynamics, in addition to the conventional study of protein interaction, abundance and activation.

This study was designed for the data mining and bioinformatics analysis of protein expression data experimentally obtained by reverse phase LC-MS/MS combined with 1-D SDS-PAGE. The protein profiling was obtained from the heart of three sample sets, namely live piglets born out of normal conception, live SCNT piglets and dead SCNT piglets. The objective of the study was to compare the global protein profiling and examine the pattern, if any, among the differentially expressed proteins to draw a plausible explanation for early peri-natal death of SCNT clones. Various data mining approaches were followed to attain the aforementioned objective.

Chapter Two

REVIEW OF LITERATURE

2.1 Foetal abnormalities in SCNT clones

SCNT technology has unveiled the capabilities of an oocyte to reprogram the somatic genome from a differentiated to a totipotent state. After the first successful cloning of the sheep Dolly by Dr. Ian Wilmut in 1996, the technique has been successfully replicated in several species, including mice, sheep, cattle, goats, pigs, rabbits, cats and dogs [18-24]. However, the success rate of producing healthy, live animals through the process of cloning across various animal species is less than 5% [24-26], owing to the incomplete and/or aberrant nuclear programming in SCNT. Specifically, pig oocytes are highly sensitive to stressors and chemical factors associated with nuclear transfer technology. Studies show that the lipids in pig oocytes affect embryonic development following SCNT [27]. Numerous instances of perinatal and early post-natal deaths observed in majority of SCNT cloning cases [28-30] have restricted the exploration of this technique for various applications, and specifically for reproductive cloning purposes. Many accounts of incomplete organogenesis and functional inefficiency have been reported in various organs, primarily in the organs of cardiovascular and respiratory systems [12]. The deaths have been majorly attributed to foetal and embryonic abnormalities, including large offspring syndrome, low birth weight and abnormal gestation length [11, 31, 32]. Many of the surviving clones succumb to a number of other abnormalities potentially arising out of inappropriate or incomplete epigenetic reprogramming, resulting in a plethora of abnormalities, such as malformed organs, weakened immune system, internal hemorrhaging, metabolic abnormalities, cardiopulmonary abnormality, lymphoid hypoplasia, behavioral abnormalities, neonatal respiratory distress and multiple organ failure [12, 33]. Studies have shown that the low efficiency of the SCNT process could be due to aberrant and/or incomplete epigenetic reprogramming, including post-translational modifications like DNA methylation and histone modification [34-36].

Epigenetic reprogramming consists of different processes like genomic imprinting [37], X-chromosome inactivation [38] and gene silencing [39, 40]. It also encompasses embryonic [41-43] and placental [44-46] development, carcinogenesis [47-49] and reprogramming in SCNT embryos [36, 50]. Faulty epigenetic patterns are found to vary in a species-specific and cell type-specific manner [51]. Morphological modification of the donor nucleus by the recipient cell is essential for successful nuclear reprogramming, that leads to transcriptional silencing, removal of epigenetic marks and appropriate gene expression [52]. However, in SCNT embryos, reversal of epigenetic modifications needs to occur before the zygotic genome activation [53], which usually remains incomplete and/or aberrant. Aberrant DNA methylation does not follow a defined pattern, and is species-specific. While in pig embryos, increasing DNA methylation is observed from the one-cell stage to the blastocyst stage [54, 55], limited demethylation has been observed in sheep, with no remethylation [56]. Aberrant histone modification is much more complex, and hence no general pattern has been established. Incomplete remodeling of H4K8ac and H4K12ac has been observed in the somatic nucleus after SCNT [57]. Some aberrant histone methylation patterns have also been observed [58]. Transcription in cloned animals is affected by the relationship between these two epigenetic abnormalities [59, 60]. Study of epigenetic reprogramming errors associated with SCNT is of particular importance since reprogramming events in the induced pluripotent stem (iPS) cells follows a similar pattern to that of SCNT clones. Hence, this information may have application in the iPS cell research.

2.2 Proteomics

Proteomic approach has been gaining increasing importance in the scientific world since it generates huge amounts of data that, on proper analysis, could provide massive information and multi-faceted interpretation that could be extended to a detailed understanding of the working of

the living system, including problems associated with it and developing potential solutions to those problems. This field has come a long way from large-scale separation and identification of proteins to all possible kinds of analyses of protein sets as well as the interactions among them [17]. Proteomics can broadly be classified into three major divisions, namely clinical proteomics, structural proteomics and functional proteomics. Clinical proteomics deals with the investigation of disease biomarkers. Structural proteomics deals with the analysis and understanding of cellular proteins. Functional proteomics deals with the investigation of cell signaling mechanisms. Until recently, two-dimensional polyacrylamide gel electrophoresis (2-DE) was primarily used for proteomic profiling of cellular systems [61, 62]. This approach involved solubilization of the entire protein content of a system and their subsequent separation on 2D gels. 2-DE, though most preferable, is associated with major limitations in the profiling of proteins with very low or high molecular masses, low abundance, extreme *pI* or high hydrophobicity [63, 64]. Advances in large scale global protein profiling include 2D-DIGE, MS, protein microarray, in silico proteomics and activity-based profiling. The 2D-DIGE involves the mixing of two pools of proteins labeled with different fluorescent dyes, and separation in the same gel. On the other hand, MS approach derives information on the basis of mass-to-charge (m/z) ratio of the proteins. This technique involves fractionation of protein samples by 1D SDS-PAGE or 2-DE, digestion of the proteins into peptides by sequence-specific proteases such as trypsin, followed by separation of the peptides by HPLC or ion-exchange chromatography. Subsequently, ionization of the sample is done by electrospray ionization (ESI) or MALDI. Finally, mass analysis is done by one of the mass analyzers viz. Quadrupole or TOF or ion trap Fourier Transform ion-cyclotron (FT-MS) mass analyzer. MALDI-TOF is commonly used to identify proteins by peptide-mass fingerprinting. Further advancements in high-throughput

technologies have introduced the technique of surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) [65 - 67]. This approach has enabled the proteomic expression profiling of single human embryos [68] and of proteins secreted during in vitro culture of human and mouse embryos [69]. Currently preferred multi-dimensional protein profiling technology is LC-MS/MS [70, 71]. Further advances introduce LC-MS/MS with tandem mass tags [72], isotope-coded affinity tags (ICAT) [73] or isobaric tag for relative and absolute quantification (iTRAQ) [74, 75]. Alternatively, technology combining one-dimensional (1-D) protein separation with reverse phase LC-MS/MS is also used [76].

The proteomic approach has also been used for the discovery of novel biomarkers for the diagnosis of diseases, prediction of susceptibility, monitoring disease progression, therapeutic targeting and patient-tailored therapy [77-79]. In addition, it has many other promising applications, including establishment of protein-protein interactions, determination of post-translational modifications by phosphoproteomics and glycoproteomics [80, 81] and identification and designing of potential drugs. Proteomic approach has been increasingly used for the study of various diseases and their corresponding drug discovery studies [82], including different types of cancer [83], kidney-related diseases [84], cardiovascular diseases like myocardial infarction and neural diseases like Alzheimer's disease [85] and multiple sclerosis [86].

MS methods such as MALDI-TOF and LC-MS/MS, and gel-based 2D-DIGE methods have been widely used to analyze cellular lysates of SCNT clones in order to investigate the foetal abnormalities. 2D-DIGE has been used to study the variations in mitochondrial protein levels between control, SCNT-derived and artificial insemination-derived cattle to analyze the mitochondrial-related gene expression in developmental losses of SCNT embryos [14]. Proteomic analyses have also been conducted on spindles isolated from normal oocytes and SCNT-derived

oocytes in metaphase II stage in order to associate low term developmental potential of SCNT embryos with deficiencies in spindle composition and assembly [15]. 2-DE has been used for the comparative proteomic analysis of malformed umbilical cords in SCNT-derived piglets [87]. A global proteomic approach using 2-DE and MS has been used to analyze differential protein profiling of different placentae samples to determine the role of placental dysfunction as a cause of post-natal death [9]. 2DE and MALDI-TOF MS have also been used for the proteomic analysis of extra-embryonic tissues from cloned porcine embryos [88]. In other experiments, MALDI-TOF MS peptide mass fingerprinting and MALDI-TOF MS/MS peptide sequence analysis have been used to examine the differential abundance of bovine conceptus fluid proteins [89].

2.3 Data Mining

The huge amount of data generated by current proteomic approaches demands for efficient and robust computational tools to successfully mine the data and accurately extract meaningful information. KDD is basically an iterative process, consisting of seven basic steps [90] (Figure 1). The first step is Data Cleaning, wherein noise and irrelevant data is removed from the data set. The second step is Data Integration, in which the data set is analyzed across multiple data sources or databases combined in a common source. This step is followed by Data Selection, in which relevant data is shortlisted and retrieved separately. The next step is Data Transformation, otherwise known as data consolidation, in which the selected data is transformed into different forms that are friendly to the subsequent mining procedure. The fifth step is Data Mining, which is the most relevant and crucial step of the entire KDD process. This process involves the recognition or discovery of hidden patterns in the given data set in order to deduce meaningful hypotheses or models. This step is followed by Pattern Evaluation, in which interesting and plausible patterns corresponding to the given requirements are identified. The last step in the

KDD process is Data Representation, which deals with the visual and/or graphical representation of the acquired information for better understanding and further analysis. One or more of these steps may be consolidated, suiting various requirements. There are various approaches to data mining, depending on the source and procedure of obtaining the data, and the kind of interpretation required. It specifically depends on the data type, i.e. whether the data to be analyzed is a sequence data, structure data, spectroscopic data or image data. Data mining may either be descriptive or predictive, each determining the kind of pattern that can be discovered. The criticality of the discovered patterns depends on the validity when tested on new data, degree of certainty, understandability, novelty and usefulness of the patterns.

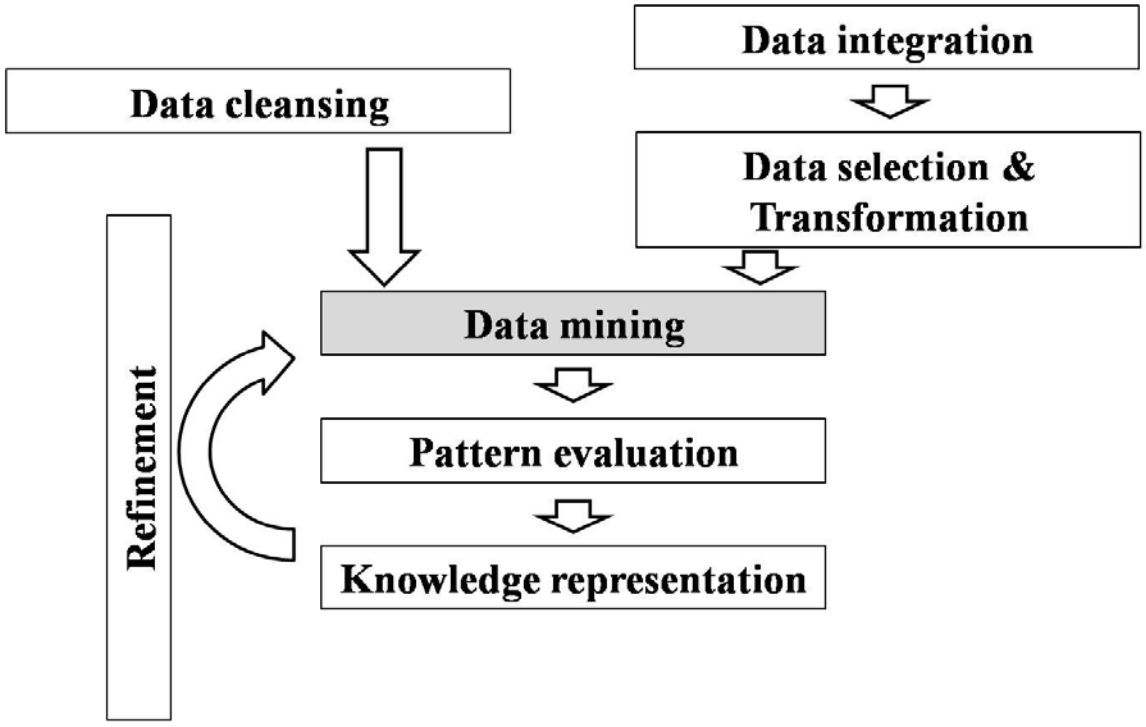


Figure 1: Strategy for data mining of proteomic data

Constant progress is seen in this direction, with a steady increase in the number and kind of databases to accommodate the data and the development of superior tools for efficient mining of

the data. Some examples of such tools developed over the past decade include Database for Annotation, Visualization and Integrated Discovery (DAVID), MAPPFinder, GeneMerge, etc. [91 - 93]. Data mining is currently used for a majority of proteomic assessment studies and subsequent applications like validating the role of certain proteins or groups of proteins in diseases [94], identification of biomarkers [78, 79, 95], mapping of regulatory mechanisms pertaining to physiological processes [96], identification of drug targets and drug delivery approaches [82], etc. Certain visual mining tools directly make use of information generated by high-throughput sequencing and proteomics based on mass spectrometry. Some examples of such tools are mspecLINE, CircAtlas and Cytoscape [97, 98]. Proteomic analysis of a single system or comparative analyses between two systems, differing on one or more defined parameters, use bioinformatics and data mining approach. For example, comparative proteomic analysis of parthenogenetically developed and in-vitro fertilized embryo [99], comparison between nuclear reprogramming via transcription factors and during SCNT [100], etc. The interactive exploration and linking of complementary proteomic and structural data is highly focused on, and is facilitated by a very less number of bioinformatics tools. Two of such tools that are widely used for complementary study purposes are STRING and structureViz [101].

2.4 Bioinformatics Tools for Data Mining

Data mining involves the use of various general-purpose and bioinformatics tools, as well as high-level languages. Three of the most common programming languages used are Structured Query Language (SQL), Practical Extraction and Report Language (PERL) and Python. All three of these are interpreted scripting languages, and hence slow in execution. PERL and Python are flexible and powerful languages for mining of sequence and textual data, while SQL is useful only for querying a relational database and isn't a stand-alone application. However, it has the

advantages of high performance, portability between databases and vendor-independence. A large number of general-purpose web-based as well as stand-alone data mining applications are also available, including Angoss, Clustran, Daisy, DataMind, Miner, Magnify, MatLab, Oracle Darwin, Spotfire, etc. Some of these applications are database-specific (e.g. Oracle Darwin), while others are non-specific and can be used with any database system (e.g. SAS). MatLab is a widely used commercial application that can be extended through add-ons. Bioinformatics-specific tools include Multiple Em for Motif Elicitation (MEME), Pattern-Induced Multi-sequence Alignment (PIMA), Pratt and Sequence Pattern EXhaustive Search (SPEXS). MEME is used for motif discovery. PIMA is used for multiple sequence alignment of a sequence set. Pratt is used as a pattern discovery tool in a set of unaligned protein sequences. SPEXS is used for discovery of sequence patterns.

An integrated biological knowledgebase and a set of analytical tools for functional annotation of genomic or proteomic information is available in DAVID [91, 102]. This platform was developed by the National Institute of Allergy and Infection Disease (NIAID), National Institute of Health (NIH) for the enhancement of data analysis in functional annotation, especially for microarray and proteomic studies. The algorithm and tools of DAVID provide a high-throughput data mining environment for systematic extraction of biological meaning from large input datasets [ref] by means of batch annotations and gene ontology (GO) term enrichment analysis. It provides extensive annotation coverage over nearly 40 annotation categories containing one or more databases, including disease databases like OMIM, Gene Ontology, functional databases like PIR, COG, etc., pathway databases like KEGG, BIOCARTA, etc., literature databases like PUBMED, etc., as well as other databases for protein domains, protein interactions and tissue expression. The tools permit the selection of GO parameters for clustering of proteins based on

their cellular localization, molecular function and involvement in biological processes. Cellular component pertains to the sub-cellular or extra-cellular location of the protein, or a gene product group that it might be a part of. The domain of molecular function explores the elemental activities of the protein at the molecular level. Biological process, on the other hand, defines the function of the protein in terms of a series of events involving one or more molecular functions with a defined beginning and end. DAVID displays results in three formats, namely, as cluster, chart and table. Gene enrichment is determined by Expression Analysis Systematic Explorer (EASE) score, which uses Fisher Exact test. In addition to function annotation, DAVID also provides much other functionality like functional classification of genes, gene ID conversion into various formats and also a batch viewer of gene names. The result of analysis is represented in three forms, namely cluster, chart and table. Functional analysis clustering report is a non-redundant representation of the data, clustered using the integrated techniques of Kappa statistics, used to measure the degree of similarity between two annotations, and fuzzy heuristic clustering, used for classification based on kappa values. It offers various filtering parameters, namely stringency ranging from lowest to highest, similarity term overlap, similarity threshold, initial group membership, final group membership and multiple linkage thresholds. Functional annotation chart is a listing of annotation terms with their associated genes. It uses EASE score threshold, which represents a modified Fisher Exact p-value that ranges from 0 to 1 (0 representing perfect enrichment). A p-value of 0.05 or lower is considered to be a strong enrichment value. Functional annotation table is a gene-centric view that involves non-statistical listing of genes and their associated annotation terms.

Cytoscape is yet another open source software platform that facilitates the analysis and visualization of complex network data and biological pathways, and integration of this data with

various kinds of annotations and gene expression profiles. It is widely considered a standard tool for integrated analysis and visualization of networks in biological research [103]. Cytoscape organizes data in terms of network graphs, with biological entities depicted as nodes, and the interactions between them depicted as edges. Attribute values help in mapping of nodes and edges to specific values, and are used to control their visual aspects and perform network searches. This tool has a total of 66 built-in functions. Cytoscape provides a common platform for the analysis of diverse types of data against multiple public databases at a time, including Reactome, String, Uniprot, ChEMBL, DIP, etc. It also has the provision of merging the output data for a set of input data from different databases into one network in order to facilitate easy analysis of inter-connections between the networks. Cytoscape is widely used for visualization and analysis of biological networks [104], visualization of protein interaction networks [105], computational construction of specialized biological networks [106], construction of computational network models for various visualization of biological phenomena [107], visualization of multi-dimensional cancer genomics data [108] and integration of genomic and proteomic information with biological networks [109].

Chapter 3

MATERIALS AND METHODS

3.1 Discovery Proteomics

Heart tissue samples from piglets were kindly donated by Dr. Sang Jun Uhm, Konkuk University, South Korea. The tissues were collected from the live piglets born out of normal conception (n = 3) or SCNT (n = 3) and from the dead piglets created by SCNT (n = 3). Proteins were extracted from each sample, separated on 1D-SDS PAGE, digested by trypsin, purified by column chromatography using C_{18} reversed phase resin and analyzed by reverse phase LC-nano ESI-MS/MS using LTQ ion-trap mass spectrometer essentially as described previously [99] and was performed at the Institute of Biomedical Science and Technology, South Korea. The acquired MS/MS spectra were searched against the non-redundant National Center for Biotechnology Information (NCBI) protein database for mammalian and *Sus scrofa* proteins in both forward and reverse in order to strike a compromise between minimizing the number of false positives and avoiding missing real protein hits [110]. Mammalian database was searched because protein database of *S. scrofa* is still incomplete.

Three data sets were analyzed, one comprising of protein expression profile of piglets born out of normal conception, hereafter referred to as NC1 for mammalian and NC2 for *S. scrofa*, another comprising of protein expression profile of live piglets created by SCNT, hereafter referred to as CL1 for mammalian and CL2 for *S. scrofa*, and the last comprising of protein expression profile of dead piglets created by SCNT, hereafter referred to as CD1 for mammalian and CD2 for *S. scrofa*. A dataset for the proteins that were differentially expressed across different categories (normal conception, cloned live and cloned dead) was also compiled, hereafter referred to as DE1 for mammalian and DE2 for *S. scrofa*.

3.2 Data Mining (DM) / Knowledge Discovery in Database (KDD)

Following filtering techniques were used for removing the noisy and irrelevant data from the data sets. On matching with the NCBI database, only those protein sequences were considered that had a match of at least two peptides with the database. The rest of the data was neglected. In addition, data was cleansed based on the intensity of the spectra obtained from mass spectrometry results. The sequences were mined both in forward and reverse directions, and a true match was considered only if the false positive rate, defined as the number of unique peptides identified from the reversed database search (number of false positives) divided by the number of unique peptides identified from the normal database search number of true positives plus the number of false positives) [111], was less than 0.05. In case of multiple protein database entries for the same set of peptides, only a single entry (highest molecular weight) was considered. Quantification of the proteomic data was done by measuring peak heights and peptide counts. The protein abundance was estimated by considering the coverage of unique peptides as well as the total number of repeat observations of peptides.

The identified proteins were annotated by the functional annotation tool suite of the DAVID Bioinformatics Resources version 6.7 (URL: http://david.abcc.ncifcrf.gov). Three major gene ontology parameters viz. cellular component, molecular function and biological process [112] were analyzed and proteomic data from all the three datasets were clustered using different stringency parameters. The resulting patterns were manually evaluated and refined to finally determine following stringency parameters: Multiple mammalian species were selected as the background, including *Homo sapiens, Sus scrofa, Bos taurus, Canis lupus, Rattus norvegicus, Mus musculus* and *Macaca mulatta.* Medium stringency was used for the classification process. Kappa similarity search was used with a similarity term overlap value of 3, and similarity

threshold of 0.50. For the classification, both initial and final group membership values were taken as 3, with a multiple linkage threshold of 0.50. Expression Analysis Systematic Explorer (EASE) was used for enrichment with a threshold value of 1.0.

Based on the information obtained from the analysis of the proteomic data, hypotheses were formed as to the nature of the proteins affected in cloned piglets that result in the SCNT abnormalities. Hypothesis testing was performed using the bioinformatics tool Cytoscape version 3.0.1 (URL: http://www.cytoscape.org). Five databases, namely Reactome, GeneMANIA, IntAct, String and UniProt were searched. Reactome (URL: http://www.reactome.org/ReactomeGWT/entrypoint .html) is an open-source manually curated pathway database. IntAct (URL: http://www.ebi.ac.uk/intact/) is an open-source database system and tools for analysis of molecular interaction data. String (URL: http://string-db.org/) is a database of known and predicted protein associations, including both physical and functional associations. UniProt (URL: http://www.uniprot.org/ is an open-source database of protein sequence and functional information.

3.3 Data Validation

Expression level of selected 13 genes that were found to be differentially expressed among the three sample types viz. NC, CD and CL, were validated by reverse transcriptase polymerase chain reaction (RT-PCR) of total RNA isolated from the respective samples. The validation was performed by Bio-Organ Research Center, South Korea using the procedure described earlier [113].

Chapter 4

RESULTS AND DISCUSSION

4.1 Protein Profiling of Heart

4.1.1 Heart proteins in piglets

Analysis of these peptide sequences, generative from LC-MS/MS spectra, against non-redundant NCBI protein database led to the identification of 507 mammalian proteins and 447 *S. scrofa* proteins with confidence level of 95%. The number of proteins identified in each dataset, i.e. NC, CL and CD for both mammalian and *S. scrofa* proteins is given in Table 1.

Table 1: Total number of proteins identified in each dataset, i.e. NC, CL and CD for mammalian and *S. scrofa* proteins

Normal Conception (NC)		**SCNT Cloned Live (CL)**		**SCNT Cloned Dead (CD)**	
NC1	**NC2**	**CL1**	**CL2**	**CD1**	**CD2**
155	203	135	14	70	90

DB: non-redundant NCBI database

Numerical classification of the differentially expressed proteins in heart samples of cloned piglets according to the number of unique peptides matching each identified protein is given in Table 2.

Among these proteins, 26 mammalian proteins and 44 *S. scrofa* proteins had sequence coverage of > 30%. The distribution of proteins according to their sequence coverage is shown in Figure 2. The twenty most abundant proteins in cloned embryos identified by LC-MS/MS are presented in Table 3.

Table 2: Number of differentially expressed proteins of heart samples from piglets classified according to the number of unique peptides matching each identified protein

No. of unique peptides matching per protein	Normal Piglet		Cloned Piglet born Live		Cloned Piglet born Dead	
	Mammalian DB	*S. scrofa* DB	Mammalian DB	*S. scrofa* DB	Mammalian DB	*S. scrofa* DB
0	10	1	15	99	77	19
2~4	75	67	78	34	50	78
5~9	46	40	40	3	15	28
>10	16	32	14	4	5	15
Total	147	140	147	140	147	140

a) Acceptance Spectrum Mill score > 11; scored peak intensity >60%; confidence level > 95%.
b) DB : non-redundant NCBI database

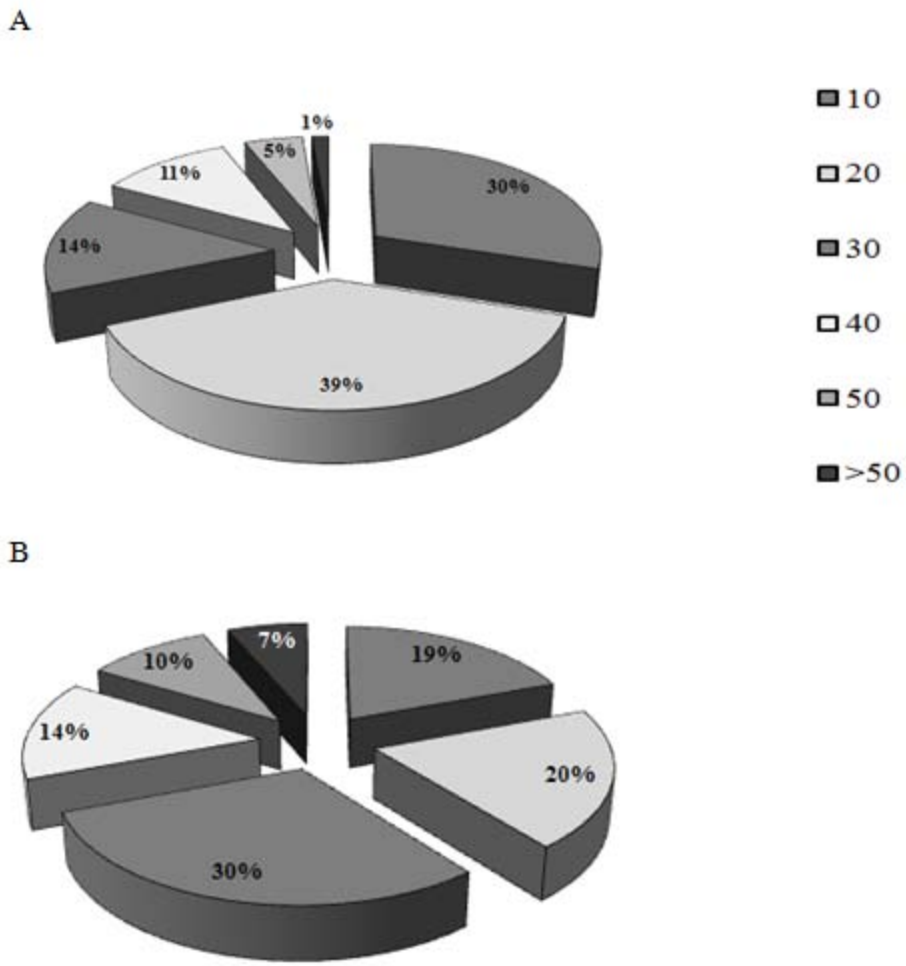


Figure 2: Pie charts depicting the sequence coverage of mammalian (A) and *S. scrofa* (B) proteins identified by LC-MS/MS

Further analysis of the data revealed that among the identified 507 mammalian proteins, only 45 overlapped between the three datasets, whereas in case of 447 *S. scrofa* proteins, 21 proteins overlapped. In case of mammalian datasets, out of 507 proteins, the number of proteins uniquely expressed in NC1, CL1 and CD1 were 55, 135 and 71 respectively. In case of porcine datasets, out of 447 proteins, the number of proteins uniquely expressed in NC2, CL2 and CD2 were 201, 14 and 91 respectively. Comparison of fold changes in the levels of identified proteins revealed that 147 mammalian and 140 *S. scrofa* proteins were differentially expressed, corresponding to

29% and 31% of the total proteins respectively. This indicates a similarity between the data from mammalian and *S. scrofa* databases.

Table 3: Twenty most abundant proteins in heart of cloned piglets identified by LC-MS/MS

Protein	Percent coverage	Number of unique peptides	Number of total peptides
Desmin	41.8	14	43
Malate dehydrogenase	18.0	4	23
Histone 1	12.4	2	18
Annexin A2	20.1	6	13
Calsequestrin	8.3	4	10
Synaptopodin 2	8.2	3	10
Phosphoglycerate kinase 1	23.5	7	9
SERPINA3-6	27.1	6	9
Troponin I	19.4	3	9
Pyruvate dehydrogenase	14.9	5	8
Spectrin	6.1	7	7
Talin 1	4.9	6	7
Clathrin	4.2	5	7
Cadherin 2	11.6	4	7
Elongation factor 1 - alpha 2	20.3	4	7
Ubiquinol cytochrome C reductase	10.2	2	7
Apoptosis-inducing factor	15.2	6	6
Calcium/calmodulin-dependent protein kinase	13.2	4	6
RAS oncogene family isoform 3	31.8	4	6
Myomesin 2	3.8	4	6

Peptide sequences were matched with *S. scrofa* protein sequences in nrNCBI database

MS methods such as MALDI-TOF and LC-MS/MS, and gel-based 2D-DIGE methods have been widely used to analyze cellular lysates of SCNT clones in order to investigate the foetal abnormalities. One such study analyzed spindle composition and assembly in SCNT clones to elucidate the role of spindles in decreased developmental potential of SCNT embryos [15]. A global proteomic approach using 2-DE and MS has been used to analyze differential protein profiling of different placentae samples to determine the role of placental dysfunction as a cause of post-natal death [9]. However, a large scale sequencing, identification and expression analysis of global proteins in heart tissue on SCNT clones has not yet been investigated. In addition, MALDI-TOF or gel-based proteomic approaches used for most of the studies suffer from certain shortcomings, restricting the global profiling of tissue samples. This study was the first record of global protein profiling of heart samples. It applied reverse phase LC-nano ESI-MS/MS combined with 1-D SDS-PAGE strategy for peptide sequencing in heart samples of piglets born out of SCNT technology. No other report of using LC-MS/MS for global proteomic analysis has been found.

4.1.2 Normal versus cloned piglets

Among the differentially expressed mammalian proteins, 52 proteins were up-regulated in cloned piglets as compared to normal piglets, out of which 15 were up-regulated by two or more folds. 54 proteins were down-regulated in cloned piglets as compared to normal piglets, out of which 20 were down-regulated by two or more folds. 15 proteins were unique to the normal samples, with undetected expression levels in cloned samples. On the other hand, 10 proteins were unique to the cloned samples, with undetected expression levels in normal samples.

In *S. scrofa*, only two proteins were up-regulated in cloned piglets as compared to normal piglets by less than two-fold. 37 proteins were down-regulated in cloned piglets as compared to normal

piglets, out of which 33 were down-regulated by two or more folds. 19 of the proteins were down-regulated by five-fold or more. 99 proteins were unique to the normal samples, with undetected expression levels in cloned samples. On the other hand, only one protein was unique to the cloned samples, with undetected expression levels in normal samples.

The number of up-regulated and down-regulated proteins corresponding to the fold changes in heart samples from cloned piglets in comparison to normal piglets is shown in Table 4.

Differential expression of proteins in normal and SCNT clones have been studied to identify the proteins that could be responsible for the abnormalities and deaths in SCNT clones. Variations in mitochondrial protein levels have been analyzed between control and SCNT clones using 2D-DIGE [14]. 2-DE has also been used for the comparative proteomic analysis of malformed umbilical cords in normal and SCNT-derived piglets. A limited number of proteins were considered for this study. Another approach was the use of 2-DE and MALDI-TOF for proteomic analysis of extra-embryonic tissue from cloned porcine embryos [88]. However, the gel-based proteomics require very large amounts of samples, and are low throughput technologies.

4.1.3 Live versus Dead cloned piglets

Among the differentially expressed mammalian proteins, 52 proteins were up-regulated in CD1 as compared to CL1, out of which 15 were up-regulated by two or more folds. 26 proteins were down-regulated in CD1 as compared to CL1, out of which 20 were down-regulated by two or more folds. 77 proteins were unique to CL1 and had undetected expression levels in CD1 On the other hand, 15 proteins were unique to CD1 with undetected expression levels in CL1.

In *S. scrofa*, 19 proteins were up-regulated in CD2 as compared to CL2, out of which 12 were up-regulated by two or more folds. Three proteins were down-regulated in CD2 as compared to CL2, within a range of 15 to 3 folds. 99 proteins were unique to the CD2, with undetected expression levels in CL2. On the other hand, 19 proteins were unique to the CL2, with undetected expression levels in CD2.

The number of up-regulated and down-regulated proteins corresponding to the fold changes in heart samples from cloned dead piglets in comparison to cloned live piglets is shown in Table 5.

Majority of the studies have compared the protein profiles of normally born animals with those developed by SCNT. Not much work has been done on the comparative proteomic analysis between live and dead SCNT clones.

Table 4: Number of proteins up- or down-regulated in heart samples from cloned piglets in comparison to normal piglets as revealed by reverse phase LC-MS/MS

Fold change	**Mammalian DB**		***S. scrofa* DB**	
	Up-regulated	**Down-regulated**	**Up-regulated**	**Down-regulated**
<2	41	39	2	5
2~5	11	14	0	16
5~10	0	1	0	14
>10	0	0	0	3
Total	52	54	2	38

DB: non-redundant NCBI database

Table 5: Number of proteins up- or down-regulated in heart samples from cloned dead piglets in comparison to cloned live piglets as revealed by reverse phase LC-MS/MS

Fold change	Mammalian DB		*S. scrofa* DB	
	Up-regulated	**Down-regulated**	**Up-regulated**	**Down-regulated**
< 2	15	16	7	2
2~5	0	17	9	1
5~10	0	2	0	0
>10	0	1	0	0
Total	15	36	16	3

DB: non-redundant NCBI database

4.2 Functional annotation and cellular location

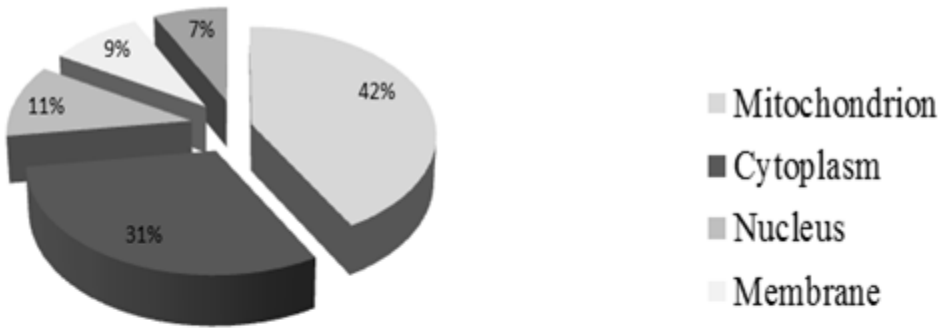
4.2.1 Heart proteins in cloned piglets

The LC-MS/MS analysis could identify both cellular and membrane proteins. The distribution of the identified proteins across different cellular locations in mammalian and *S. scrofa* is shown in Figure 3.

To further analyze the probable cellular processes and signaling pathways that might be disrupted in cloned piglets compared to normal piglets, the differentially expressed proteins between the two groups were assigned functional annotations in order to cluster them into groups in accordance with their biological and molecular functions. The analysis revealed that binding activity and catalytic activity were the main functions of a majority of the differentially

expressed proteins. The differentially expressed proteins were majorly involved in basal metabolism, including glycolysis and tricarboxylic acid cycle (44% in mammalian and 43% in *S. scrofa*). A classification of the differentially expressed proteins across major biological processes is given in Figure 4.

A



B

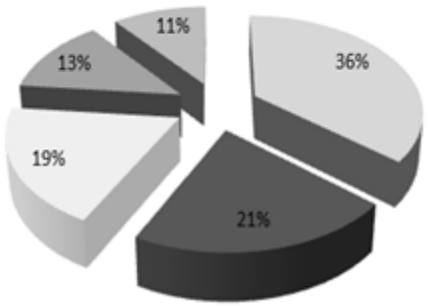


Figure 3: Pie charts depicting the distribution of mammalian (A) and *S. scrofa* (B) proteins across different cellular components

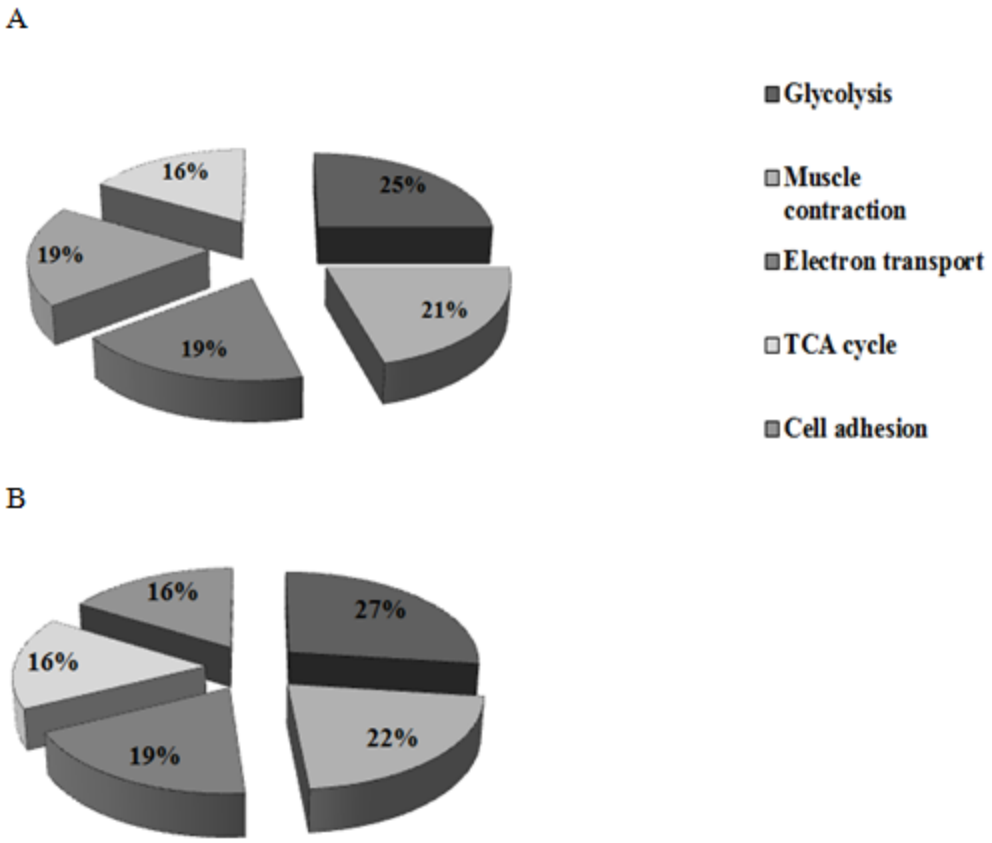


Figure 4: Pie charts depicting the classification of differentially expressed mammalian (A) and *S. scrofa* (B) proteins on basis of biological process involved

4.2.2 Normal versus cloned piglets

Among the 60 annotated mammalian proteins that were up-regulated in cloned piglets as compared to normal piglets, majority of the proteins (30.9%) were involved in specific molecular functions related to basal metabolism, including glycolysis and tricarboxylic acid (TCA) cycle. Among the 50 annotated mammalian proteins that were down-regulated in cloned piglets as compared to normal piglets, 25.7% were involved in molecular functions related to basal metabolism. Other proteins had structural activity, ion binding activity, transporter activity, nucleotide binding activity and other enzyme activities. A classification of the up-regulated and

down-regulated proteins on basis of molecular functions in cloned piglets as compared to normal piglets is shown in Figure 5.

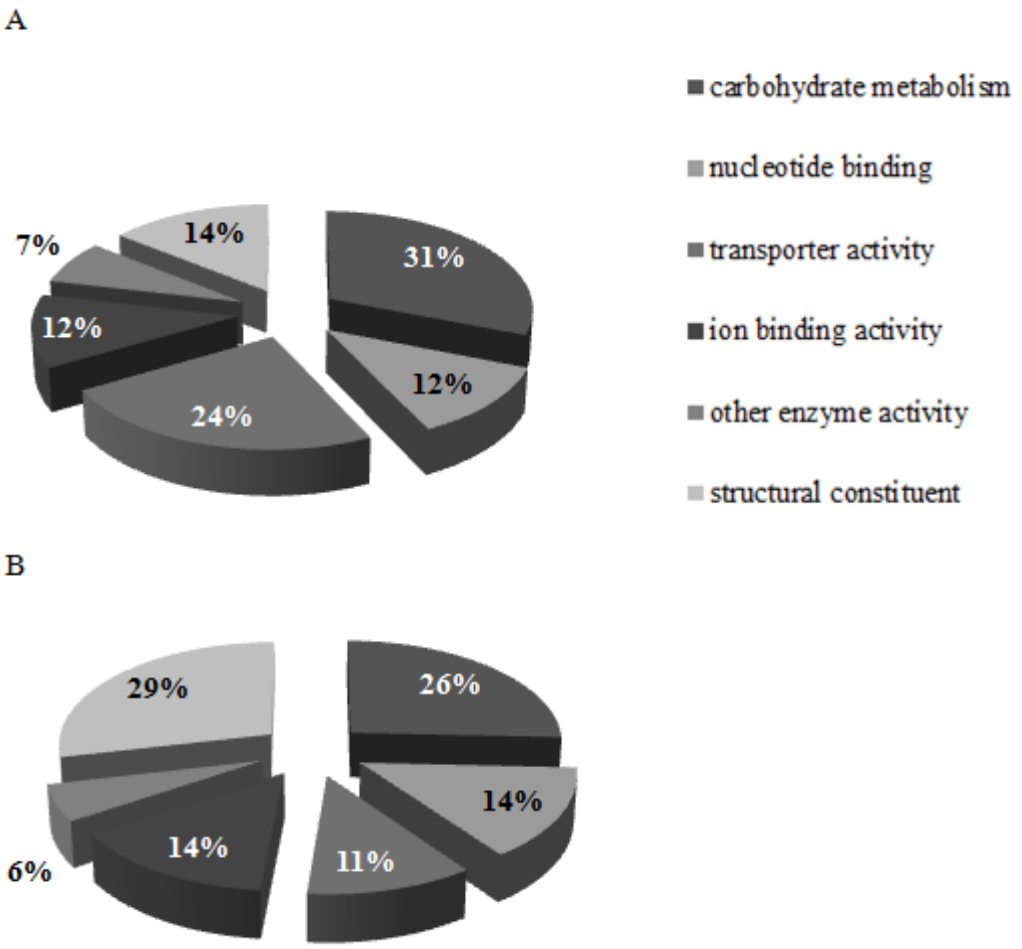


Figure 5: Pie charts depicting the functional classification on basis of molecular function of proteins up-regulated (A) and down-regulated (B) in cloned piglets as compared to normal piglets.

4.2.3 Live versus Dead cloned piglets

Among the 15 annotated mammalian proteins that were up-regulated in cloned dead piglets as compared to cloned live piglets, 20.7% were involved in specific molecular functions related to basal metabolism, including glycolysis and TCA acid cycle. Among the 36 annotated mammalian proteins that were down-regulated in cloned dead piglets as compared to cloned live piglets, 25% were involved in molecular functions related to basal metabolism. Other proteins had structural activity, ion binding activity, transporter activity, nucleotide binding activity and other enzyme activities. A classification of the up-regulated and down-regulated proteins on basis of molecular functions in cloned piglets as compared to normal piglets is shown in Figure 6.

4.3 Proteins related to basal metabolism

We found at least 22 differentially expressed proteins in mammalian and 19 in *S. scrofa* that are known to be related to basal metabolism, including the processes of glycolysis, TCA cycle, glucose metabolism, glycogen metabolism and gluconeogenesis. This included almost all enzymes of glycolysis and TCA cycle, and some enzymes of the other metabolic processes. The affected proteins and their corresponding role in glycolysis and TCA cycle are depicted in Figures 7 and 8 respectively. All proteins of glycolysis, and most of the proteins of TCA cycle were down-regulated. These cycles are crucial for basic cell survival. Hence, basal metabolism could be targeted in cloned species to study the aberrance and investigate any relationship with the abnormal deaths, and measures may be developed to circumvent the issues and facilitate the healthy development and metabolism of cloned piglets.

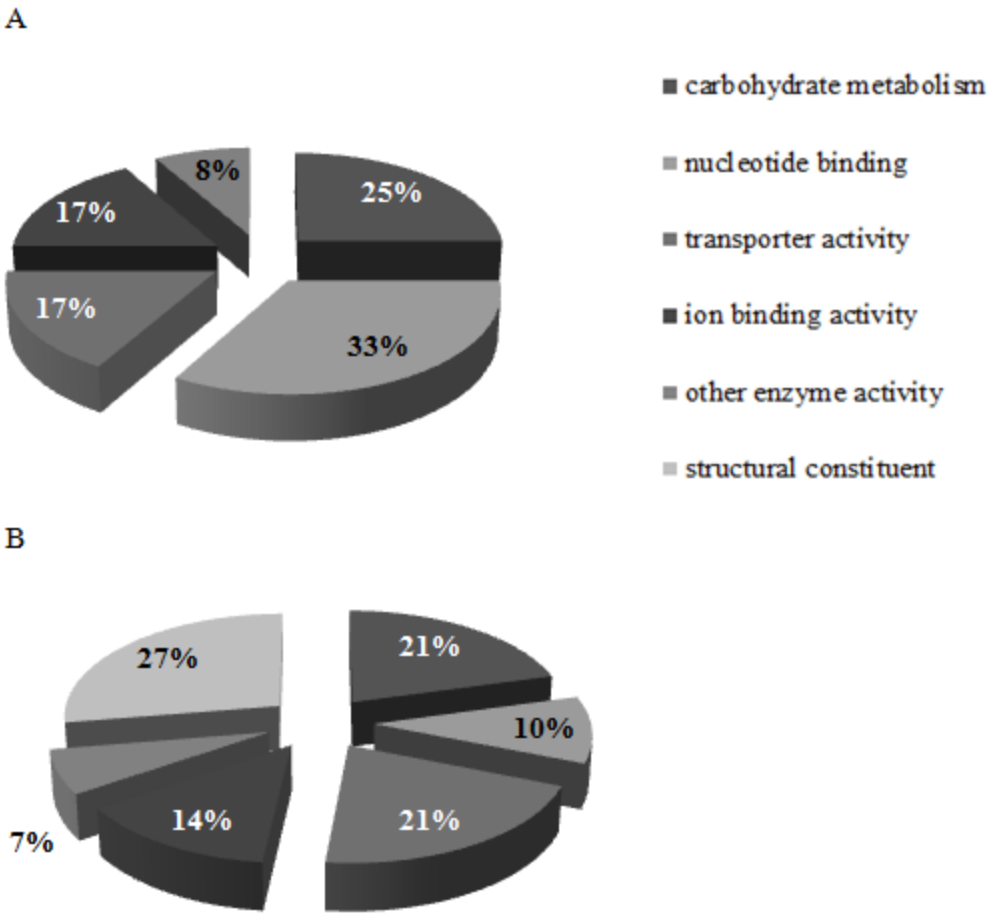


Figure 6: Pie charts depicting the functional classification on basis of molecular function of proteins up-regulated (A) and down-regulated (B) in cloned dead piglets as compared to cloned live piglets

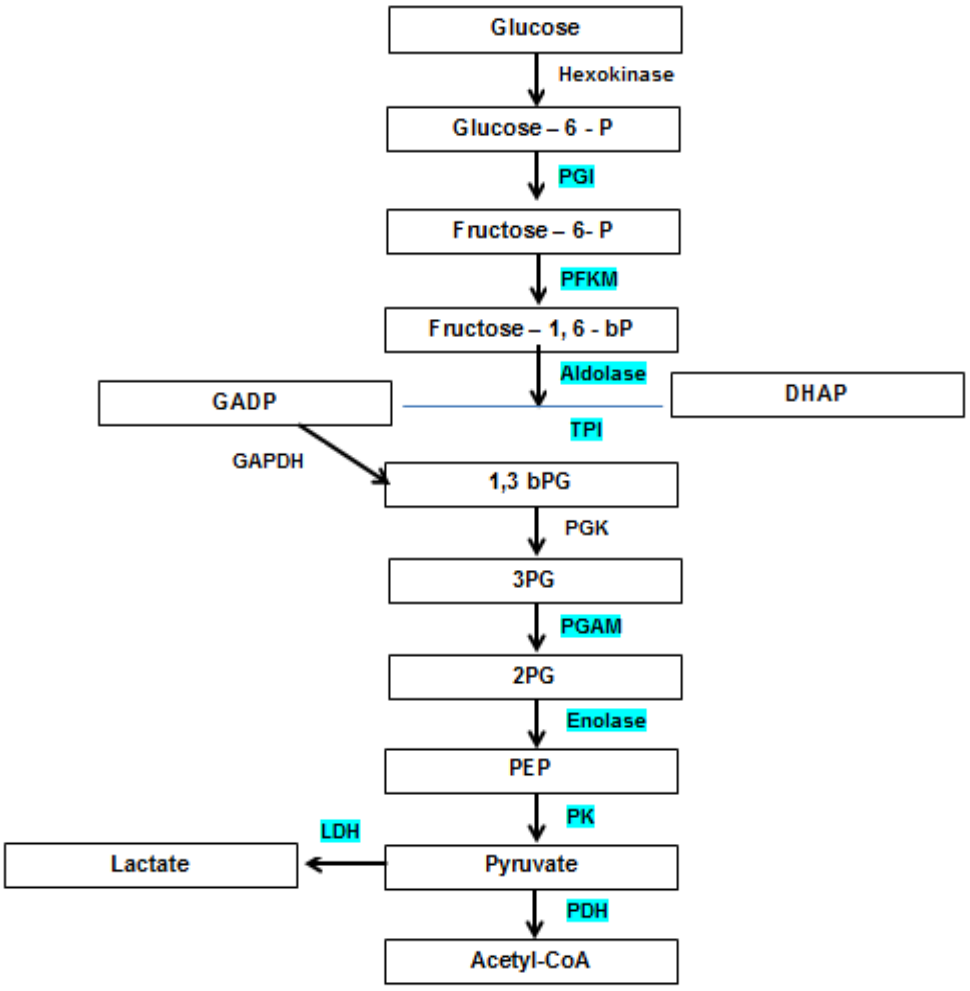


Figure 7: Differentially expressed proteins in cloned piglets as compared to normal piglets in glycolysis, and their role in the process. The grayed out proteins are differentially expressed in cloned piglets.

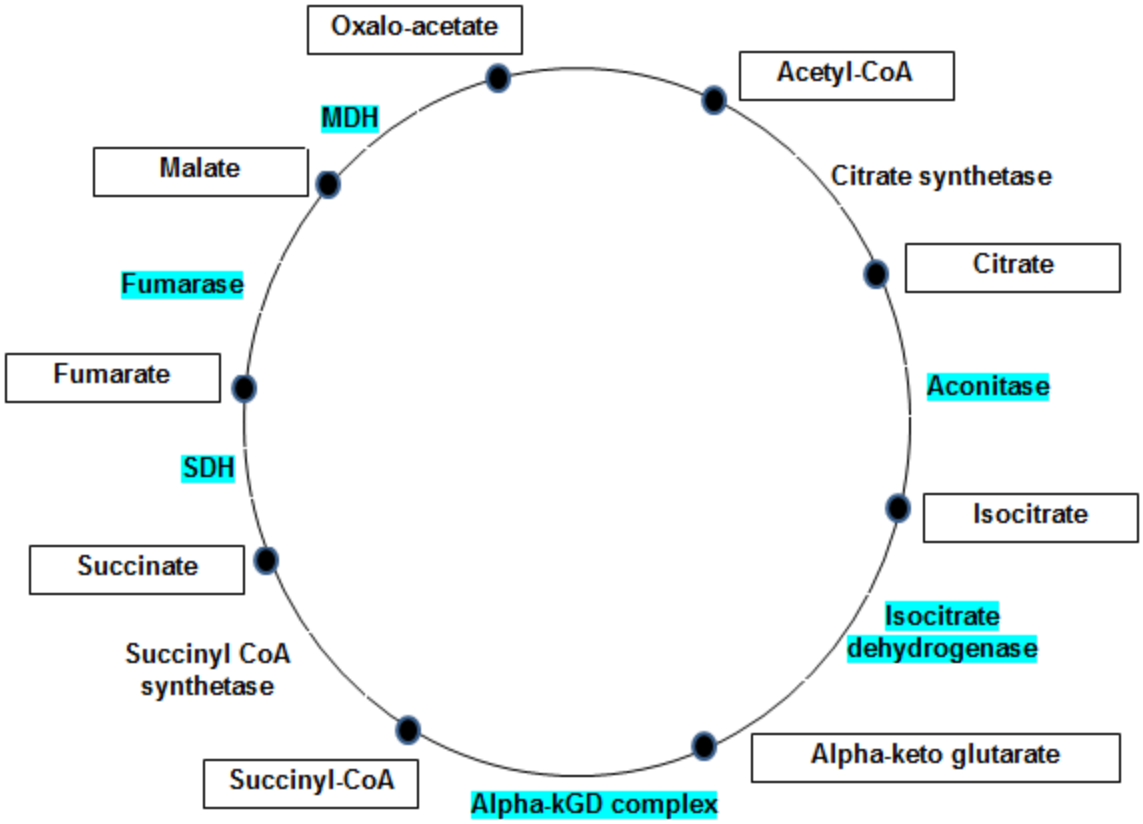


Figure 8: Differentially expressed proteins in cloned piglets as compared to normal piglets in TCA cycle, and their role in the process. The grayed out proteins are differentially expressed in cloned piglets.

4.4 Proteins related to heart development

The study found proteins specific for signaling pathways for heart development among the differentially expressed proteins in cloned piglets. These proteins exert complex functions in signaling pathways include Wnt signaling, TGF-β signaling and Notch signaling and their deranged expression levels may reduce in abnormal deaths. Signaling proteins control a number of other proteins involved in vital roles of cell development and regulation. A depiction of such proteins found to be grossly down-regulated in SCNT clones, and the large number of proteins they control and affect, is shown in Figure 9. In addition, we found at least 14 differentially

expressed proteins in mammalian and 10 in *S. scrofa* that are related to heart development and muscle contraction. These include different isoforms of actin, myosin, troponin, tropomyosin, actinin, titin, nebulin, keratin, profilin, calsarcin, lamin, collagen, sarcalumenin and cofilin. Among these proteins, some like actinin, myosin, lamin and titin have been frequently detected among the heart samples, indicating that faulty functioning of the heart might be a cause of abnormal deaths in cloned animals. Therefore, these proteins can be used as internal markers of a successful proteome identification process during batch analysis. As expected, these proteins were represented in our results.

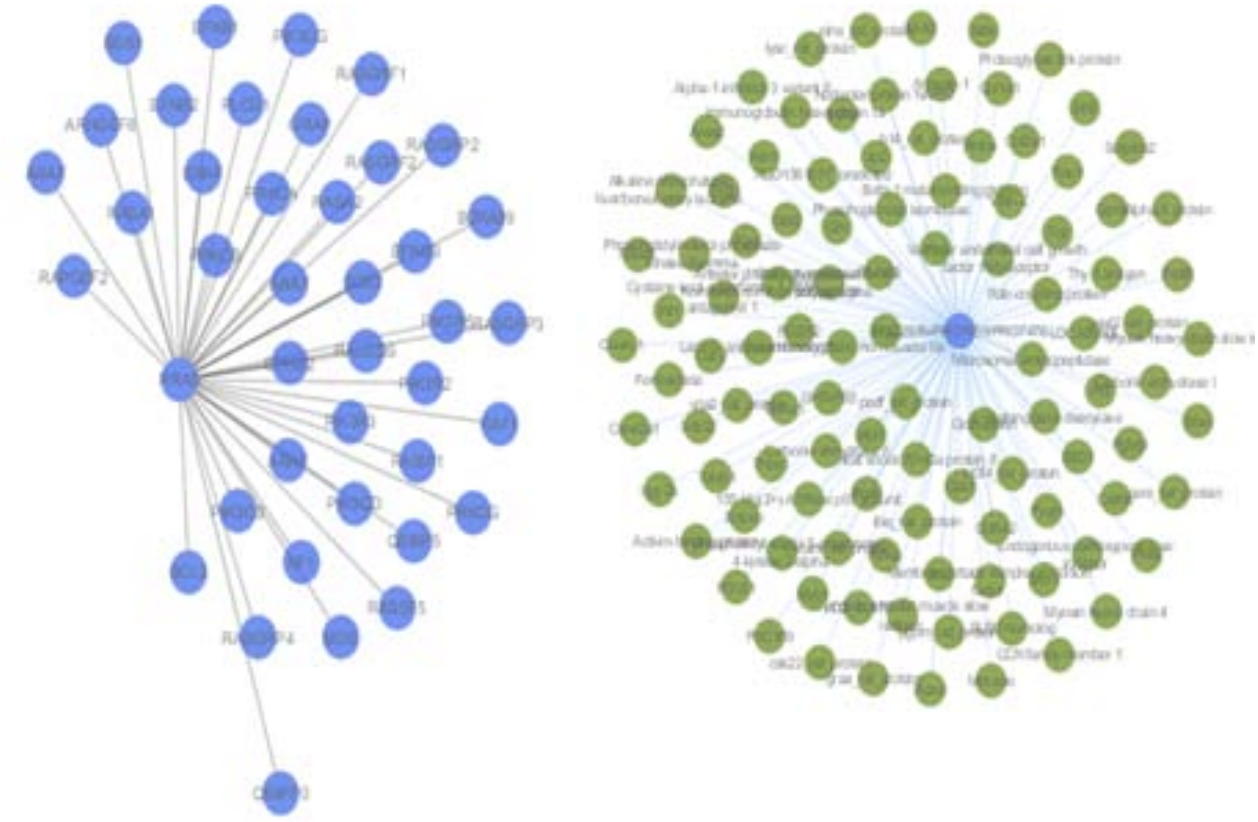


Figure 9: Differentially expressed proteins involved in signaling pathways grossly down-regulated in SCNT clones, and the related proteins they affect. The images were drawn using Cytoscape 3.0.1

147 differentially expressed proteins were found among normal and cloned piglets. The proteins belonged to a variety of biological processes and signaling pathways. However, these proteins were also inter-connected via a number of pathways. Hence, the up- or down-regulation of a protein could affect other pathways in which the protein is not directly involved. Seven differentially expressed proteins that were highly abundant were considered and their interactions analyzed. The proteins were succinate dehydrogenase and pyruvate dehydrogenase (involved in basal metabolism), adenosylhomocysteinase (involved in control of methylations), electron transfer flavoprotein (involved in cellular respiration), cationic amino acid transporter 2 (plays regulatory role in activation of macrophages), superoxide dismutase 2 (involved in removal of toxic superoxide anion radicals from the cell) and dystrophin (involved in structural integrity and signaling events). The inter-relationship between these proteins is depicted in Figure 10.

Another abundant protein with gross differential expression was 40S ribosomal protein SA (rpsA). This protein plays a pivotal role in the adhesion of cell to the basement membrane and in the consequent activation of signaling pathways. This protein also plays a role in cell fate determination and morphogenesis. The differential expression of this protein points towards a probable cause of aberrant and/or incomplete development and often subsequent death of SCNT clones. rpsA also affects the function of many other proteins involved in various pathways. A depiction of the number of pathways affected by rpsA is given in Figure 11.

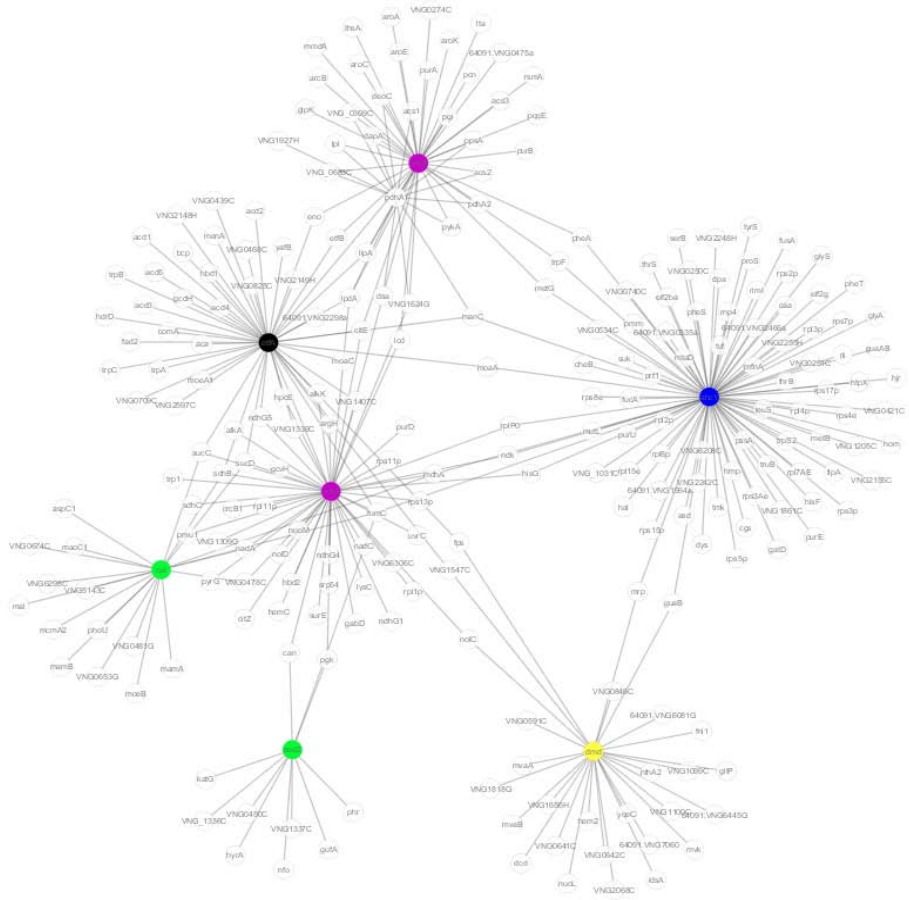


Figure 10: Interconnection between differentially expressed proteins involved in different biological processes. The images were drawn using Cytoscape 3.0.1

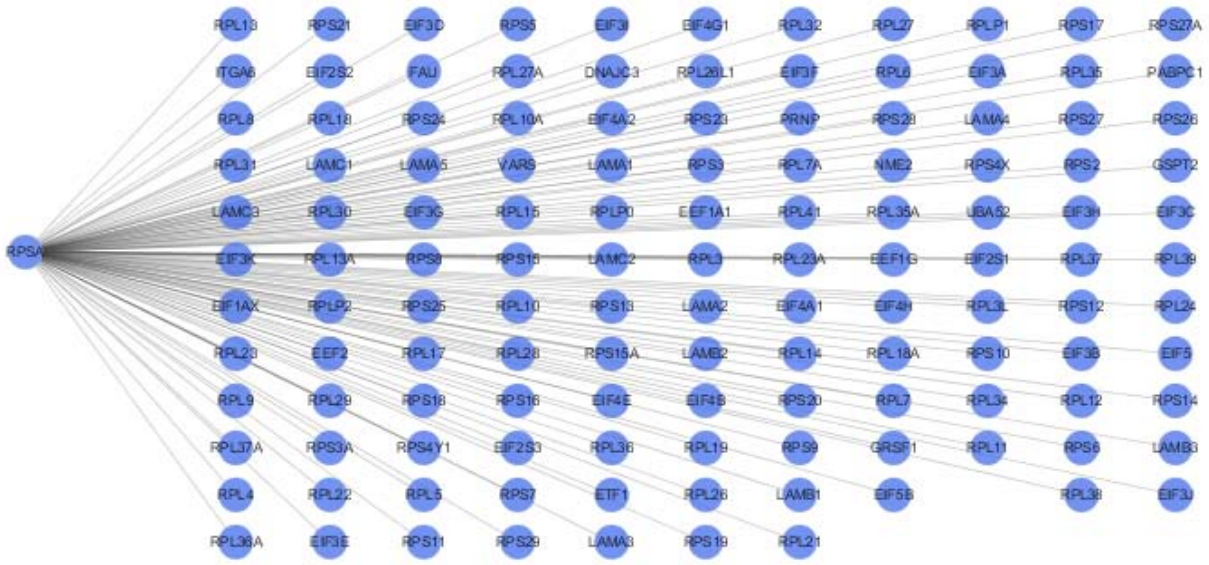


Figure 11: Number of signaling proteins inter-connected to rpsA, which is majorly differentially regulated in cloned piglets. The images were drawn using Cytoscape 3.0.1

Chapter 5

CONCLUSION

This study provides the first exhaustive analysis of global proteins in heart of pigs. To the best of my knowledge, no previous study has performed global profiling of heart proteins and hence, the results of this study may be used as a reference map for future proteomic studies. Furthermore, the study showed that basal metabolism is grossly abnormal and several proteins related to heart organogenesis and functionality were either up- or down-regulated in the cloned piglets. These abnormalities were more pronounced in the dead SCNT clones and thus, may imply that the heart of SCNT clones is developmentally and functionally insufficient and hence, may be the cause of high peri-natal death. Future studies should decipher the mechanism of these abnormalities and novel ways to circumvent these abnormalities to increase the SCNT efficiency.

LIST OF REFERENCES

1. Eyestone WH, Campbell KH (1999) Nuclear transfer from somatic cells: applications in farm animal species. *J Reprod Fertil Suppl 54: 489–497.*

2. Polejaeva IA (2001) Cloning pigs: advances and applications. *Reprod Suppl 58:293–300*

3. Prather RS. Nuclear remodeling and nuclear reprogramming for making transgenic pigs by nuclear transfer. *Adv Exp Med Biol 2007 591:1–13*

4. Tamada H, Kikyo N. Nuclear reprogramming in mammalian somatic cell nuclear cloning. *Cytogenet Genome Res 2004. 105:285–291*

5. Young et al., 2001 L.E. Young, K. Fernandes, T.G. McEvoy, S.C. Butterwith, et al. Epigenetic change in IGF2R is associated with fetal overgrowth after sheep embryo culture *Nat Genet, 27 (2001), pp. 153–154*

6. Fairburn et al., 2002 H.R. Fairburn, L.E. Young, B.D. Hendrich Epigenetic reprogramming: how now, cloned cow? *Curr Biol, 12 (2002), pp. R68–R70*

7. Vajta G, Gjerris M Science and technology of farm animal cloning: state of the art. *Anim Reprod Sci 2..6. 92:211–230*

8. Yang X, Smith SL, Tian XC, Lewin HA, Renard JP, Wakayama T, Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning. *Nat Genet 2007. 39:295–302*

9. Constant, F.; Guillomot, M.; Heyman, Y.; Vignon, X.; Laigre, P.; Servely, J.L.; Renard, J.P.Chavatte-Palmer, P. Large offspring or large placenta syndrome? Morphometric analysis of late gestation bovine placentomes from somatic nuclear transfer pregnancies complicated by hydrallantois. *Biol. Reprod. 2006, 75, 122–130.*

10. Tamashiro K.L.K.; Wakayama T.; Blanchard R.J.; Blanchard D.C.; Yanagimachi R. Postnatal growth and behavioral development of mice cloned from adult cumulus cells. *Biol. Reprod. 2000, 63, 328–334*

11. Curchoe C.L, Zhang S, Yang L, Page R, Tian X.C. Hypomethylation trends in the intergenic region of the imprinted IGF2 and H19 genes in cloned cattle. *Anim. Reprod. Sci. 2009, 116, 213–225*

12. Chavatte-Palmer P, Remy D, Cordonnier N, Richard C, Issenman H, Laigre P, Heyman Y & Mialot JP 2004 Health status of cloned cattle at different ages. *Cloning and Stem Cells 6 94–100*

13. Alberts B (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell 92(3):291–294*

14. Takeda K, Tasai M, Akagi S, Watanabe S et al. Comparison of liver mitochondrial proteins derived from newborn cloned calves and from cloned adult cattle by two-dimensional differential gel electrophoresis. *Mol Reprod Dev. 2011 Apr;78(4):263-73.*

15. Han Z, Liang CG, Cheng Y, Duan X, Zhong Z et al. Oocyte spindle proteomics analysis leading to rescue of chromosome congression defects in cloned embryos. *J Proteome Res. 2010 Nov 5;9(11):6025-32.*

16. Kim HR, Kang JK, Yoon JT, Seong HH, Jung JK, Lee HM, Sik Park C, Jin DI. Protein profiles of bovine placenta derived from somatic cell nuclear transfer. *Proteomics. 2005 Nov;5(16):4264-73.*

17. Thiele H, Heldmann S, Trede D, Strehlow J, 2D and 3D MALDI-imaging: *Conceptual strategies for visualization and data mining. Elsevier*

18. Betthauser J, Forsberg E, Augenstein M, Childs L, Eilertsen K, Enos J, Forsythe T, Golueke P, Jurgella G, Koppang R, Lesmeister T, Mallon K, Mell G, Misica P, Pace M, Pfister-Genskow M, Strelchenko N, Voelker G, Watt S, Thompson S, Bishop M. Production of cloned pigs from in vitro systems. *Nat Biotechnol 2000;18(10):1055–1059.*

19. Galli C, Lagutina I, Crotti G, Colleoni S, Turini P, Ponderato N, Duchi R, Lazzari G. Pregnancy: a cloned horse born to its dam twin. *Nature 2003;424(6949):635*

20. Kato Y, Tani T, Sotomaru Y, Kurokawa K, Kato J, Doguchi H, Yasue H, Tsunoda Y. Eight calves cloned from somatic cells of a single adult. *Science 1998;282(5396):2095–2098.*

21. Onishi A, Iwamoto M, Akita T, Mikawa S, Takeda K, Awata T, Hanada H, Perry AC. Pig cloning by microinjection of fetal fibroblast nuclei. *Science 2000;289(5482):1188–1190.*

22. Polejaeva IA, Chen SH, Vaught TD, Page RL, Mullins J, Ball S, Dai Y, Boone J, Walker S, Ayares DL, Colman A, Campbell KH. Cloned pigs produced by nuclear transfer from adult somatic cells. *Nature 2000;407(6800):86–90.*

23. Shin T, Kraemer D, Pryor J, Liu L, Rugila J, Howe L, Buck S, Murphy K, Lyons L, Westhusin M. A cat cloned by nuclear transplantation. *Nature 2002;415(6874):859.*

24. Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KH. Viable offspring derived from fetal and adult mammalian cells. *Nature 1997;385(6619):810–813.*

25. Cibelli JB, Stice SL, Golueke PJ, Kane JJ, Jerry J, Blackwell C, Ponce de Leon FA, Robl JM. Cloned transgenic calves produced from nonquiescent fetal fibroblasts. *Science 1998;280(5367):1256–1258.*

26. De Sousa PA, King T, Harkness L, Young LE, Walker SK & Wilmut I 2001 Evaluation of gestational deficiencies in cloned sheep fetuses and placentae. *Biology of Reproduction 65 23–30.*

27. Nagashima H, Kashiwazaki N, Ashman RJ, Grupen C, Nottle, MB Cryopreservation of porcine embryos. *1995. Nature 374, 416*

28. Hill JR, Burghardt RC, Jones K, Long CR, Looney CR, Shin T, Spencer TE, Thompson JA, Winger QA, Westhusin ME: Evidence for placental abnormality as the major cause of mortality in firsttrimester somatic cell cloned bovine fetuses. *Biol Reprod 2000, 63:1787-1794.*
29. Chae JI, Cho SK, Seo JW, Yoon TS, Lee KS, Kim JH, Lee KK, Han YM, Yu K: Proteomic analysis of the extraembryonic tissue from cloned porcine embryos. *Mol Cell Proteomics 2006, 5:1559-1566.*

30. Wakisaka-Saito N, Kohda T, Inoue K, Ogonuki N, Miki H, Hikichi T, Mizutani E, Wakayama T, Kaneko-Ishino T, Ogura A, Ishino F: Chorioallantoic placenta defects in cloned mice. *Biochem Biophys Res Commun 2006, 349:106-114.*

31. Renard J, Chastant S, Chesne P, Richard C, Marchal J, Cordonnier N, Chavatte P & Vignon X 1999 Lymphoid hypoplasia and somatic cloning. *Lancet 353 1489–1491*.

32. Wilson JM,Williams JD, Bondioli KR, Looney CR,Westhusin ME, McCalla DF. Comparison of birthweight and growth characteristics of bovine calves produced by nuclear transfer (cloning), embryo transfer and natural mating. *Anim Reprod Sci 1995;38:73–83*

33. Rodriguez-Osorio N, Wang Z, Kasinathan P, Page GP, Robl JM, Memili E Transcriptional reprogramming of gene expression in bovine somatic cell chromatin transfer embryos. *BMC Genomics 2009. 10:190*

34. Li E, Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet 2002. 3:662–673*

35. Morgan HD, Santos F, Green K, Dean W, Reik W Epigenetic reprogramming in mammals. *Hum Mol Genet 2005. 14:R47–R58*

36. Gronbaek, K, Hother C, Jones P.A. Epigenetic changes in cancer. *APMIS 2007, 115, 1039–1059.*

37. Hore T.A, Rapkins R.W, Graves J.A.M. Construction and evolution of imprinted loci in mammals. *Trends Genet. 2007, 23, 440–448.*

38. Yen Z.C, Meyer I.M, Karalic S, Brown C.J. A cross-species comparison of X-chromosome inactivation in Eutheria. *Genomics 2007, 90, 453–463.*

39. Miranda T.B, Jones P.A. DNA methylation: The nuts and bolts of repression. *J. Cell. Physiol. 2007, 213, 384–390.*

40. Lande-Diner L, Zhang J, Ben-Porath I, Amariglio N, Keshet I, Hecht M, Azuara V, Fisher A.G, Rechavi G, Cedar H. Role of DNA methylation in stable gene repression. *J. Biol. Chem. 2007, 282, 12194–12200.*

41. Shi L.J, Wu J. Epigenetic regulation in mammalian preimplantation embryo development. *Reprod. Biol. Endocrinol. 2009, 7, doi:10.1186/1477-7827-7-59.*

42. Wang J.L, Zhang M, Zhang Y Kou Z.H, Han Z.M, Chen D.Y, Sun Q.Y, Gao S.R. The histone demethylase JMJD2C is stage-specifically expressed in preimplantation mouse embryos and is required for embryonic development. *Biol. Reprod. 2010, 82, 105–111.*

43. Badr H, Bongioni G, Abdoon A.S.S, Kandil O, Puglisi R. Gene expression in the in vitro-produced preimplantation bovine embryos. *Zygote 2007, 15, 355–367.*

44. Coan P.M, Burton G.J, Ferguson-Smith A.C. Imprinted genes in the placenta—A review. *Placenta 2005, 26, S10–S20.*

45. Fowden A.L, Coan P.M, Angiolini E, Burton G.J, Constancia M. Imprinted genes and the epigenetic regulation of placental phenotype. *Prog. Biophys. Mol. Biol. 2011, 106, 281–288.*

46. Ng H.K, Novakovic B, Hiendleder S, Craig J.M, Roberts C.T, Saffery R. Distinct patterns of gene-specific methylation in mammalian placentas: Implications for placental evolution and function. *Placenta 2010, 31, 259–268.*

47. Yang X, Smith S.L, Tian X.C, Lewin H.A, Renard, J.P, Wakayama T. Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning. *Nat. Genet. 2007, 39, 295–302.*

48. Yamanaka S, Blau H.M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature 2010, 465, 704–712*

49. Meissner A, Jaenisch R: Mammalian nuclear transfer. *Dev Dyn 235:2460–2469, 2006*
50. Schär P, Fritsch O. DNA Repair and the Control of DNA Methylation. In Epigenetics and Disease; Gasser, S.M., Li, E., Eds.; *Springer: Basel, Switzerland, 2011; Volume 67, pp. 51–68.*

51. Bonk AJ, Cheong H-T, Li R, Lai L, Hao Y, Liu Z, Samuel M, et al. Correlation of developmental differences of nuclear transfer embryo cells to the methylation profiles of nuclear transfer doner cells in swine. *Epigenetics. 2007. 2:179–186*

52. Dean W, Santos F, Reik W Epigenetic reprogramming in early mammalian development and following somatic nuclear transfer. *Semin Cell Dev Biol. 2003. 14:93–100*

53. Zuccotti M, Garagna S, Redi CA Nuclear transfer, genome reprogramming and novel opportunities in cell therapy. *J Endocrinol Invest 2000. 23:623–629*

54. Bonk AJ, Li R, Lai L, Hao Y, Liu Z, Samuel M, et al. Aberrant DNA methylation in porcine in vitro-, parthenogenetic-, and somatic cell nuclear transfer-produced embryos. *Mol Reprod Dev 2007. 75:250–264*

55. Fulka J, Fulka H, Slavik T, Okada K, Fulka J Jr. DNA methylation pattern in pig in vivo produced embryos. *Histochem Cell Biol 2003. 126:213–217*

56. Young L, Beaujean N. DNA methylation in the preimplantaiton sheep embryo: the differing stories of the mouse and sheep. *Anim Reprod Sci 2004. 82–83:61–78*

57. Wang R, Kou Z, Zhang Y, Gao S Dynamic reprogramming of histone acetylation and methylation in the first cell cycle of cloned mouse embryos. *Biol Reprod 2007. 77:1007–1016*

58. Santos F, Zakhartchenko V, Stojkovic M, Peters A, Jenuwein T, Wolf E, Reik W, Dean W Epigenetic marking correlates with developmental potential in cloned bovine preimplantation embryos. *Curr Biol 2003. 13:1116–1121*

59. Cervoni N, Szyf M Demethylase activity is directed by histone acetylation. *J Biol Chem 2001. 276:40778–40787*

60. Geiman TM, Robertson KD Chromatin remodeling, histone modifications, and DNA methylation-how does it all fit together? *J Cell Biochem 2002. 87:117–125*

61. Craig-Schapiro R, Perrin RJ, Roe CM, Xiong C, Carter D, et al. (2010) YKL-40: A novel prognostic fluid biomarker for preclinical Alzheimer's disease. *Biol Psychiatry 68(10): 903–12*

62. Perrin RJ, Craig-Schapiro R, Malone JP, Shah AR, Gilmore P, et al. (2011) Identification and validation of novel cerebrospinal fluid biomarkers for staging early Alzheimer's disease. *PLoS One 6: e16032.*

63. Fountoulakis, M., Proteomics: current technologies and applications in neurological disorders and toxicology. *Amino Acids 2001, 21, 363–381.*
64. Lubec, G., Krapfenbauer, K., Fountoulakis, M., Proteomics in brain research: potentials and limitations. *Prog. Neurobiol. 2003, 69, 193–211.*

65. Carrette O, Demalte I, Scherl A, Yalkinoglu O, Corthals G, et al. (2003) A panel of cerebrospinal fluid potential biomarkers for the diagnosis of Alzheimer's disease. *Proteomics 3: 1486–1494*

66. Ranganathan S, Williams E, Ganchev P, Gopalakrishnan V, Lacomis D, et al. (2005) Proteomic profiling of cerebrospinal fluid identifies biomarkers for amyotrophic lateral sclerosis. *J Neurochem 95: 1461–1471.*

67. Simonsen AH, McGuire J, Podust VN, Hagnelius NO, Nilsson TK, et al. (2007) A novel panel of cerebrospinal fluid biomarkers for the differential diagnosis of Alzheimer's disease versus normal aging and frontotemporal dementia. *Dement Geriatr Cogn Disord 24: 434–440*

68. Katz-Jaffe, M. G., Linck, D. W., Schoolcraft, W. B., Gardner, D. K., A proteomic analysis of mammalian preimplantation embryonic development. *Reproduction 2005, 130, 899–905*

69. atz-Jaffe, M. G., Gardner, D. K., Schoolcraft, W. B., Proteomic analysis of individual human embryos to identify novel biomarkers of development and viability. Fertil. *Steril. 2006, 85, 101–107.*

70. Washburn, M. P., Wolters, D., Yates, J. R., III, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol. 2001, 19, 242–247.*

71. Wu, C. C., MacCoss, M. J., Howell, K. E., Yates, J. R., III, A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol. 2003, 21, 532–538.*

72. Dayon L, Hainard A, Licker V, Turck N, Kuhn K, et al. (2008) Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal Chem 80: 2921–2931.*

73. Zhang J, Goodlett DR, Quinn JF, Peskind E, Kaye JA, et al. (2005) Quantitative proteomics of cerebrospinal fluid from patients with Alzheimer disease. *J Alzheimers Dis 7: 125–133; discussion 173–180.*

74. Ogata Y, Charlesworth MC, Higgins L, Keegan BM, Vernino S, et al. (2007) Differential protein expression in male and female human lumbar cerebrospinal fluid using iTRAQ reagents after abundant protein depletion. *Proteomics 7: 3726–3734.*

75. Choe LDAM, Relkin NR, Pappin D, Ross P, Williamson B, et al. (2007) 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics 7: 3651–3660.*

76. Park, Y. M., Kim, J. Y., Kwon, K. H., Lee, S. K. et al., Profiling human brain proteome by multi-dimensional separations coupled with MS. *Proteomics 2006, 6, 4978–4986.*

77. Conrads T P, Zhou M, Liotta L and Veenstra T D, Cancer diagnosis using proteomic patternsExpert Rev. *Mol. Diagn. 3(4), 411–420 (2003)*

78. Principe S, Hui AB, Bruce J, Sinha A, Liu FF, Kislinger T. Tumor-derived exosomes and microvesicles in head and neck cancer: implications for tumor biology and biomarker discovery. *Proteomics. 2013 Mar 18. doi: 10.1002/pmic.201200533*

79. Kovac JR, Pastuszak AW, Lamb DJ. The use of genomics, proteomics, and metabolomics in identifying biomarkers of male infertility. *Fertil Steril. 2013 Mar 15;99(4):998-1007. doi: 10.1016/j.fertnstert.2013.01.111. Epub 2013 Feb 15.*

80. Engholm-Keller K, Larsen MR. Technologies and challenges in large-scale phosphoproteomics. *Proteomics. 2013 Mar;13(6):910-31*

81. Vojvodic M, Hansford LM, Morozova O, Blakely KM, Taylor P, Fathers KE, Moffat J, Marra M, Smith KM, Moran MF, Kaplan DR. A Phosphoproteomics Approach to Identify Candidate

Kinase Inhibitor Pathway Targets in Lymphoma-Like Primary Cell Lines Curr Drug Discov Technol. *2013 May 20.*

82. Schenone M, Dančík V, Wagner BK, Clemons PA. Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol. 2013 Apr;9(4):232-40.*

83. Alves G, Pereira DA, Sandim V, Ornellas AA, Escher N, Melle C, von Eggeling F. Urine screening by Seldi-Tof, followed by biomarker identification, in a Brazilian cohort of patients with Renal Cell Carcinoma (RCC). *Int Braz J Urol. 2013 Mar-Apr;39(2):228-39*
84. Sethi S, Vrana JA, Theis JD, Dogan A.. Mass spectrometry based proteomics in the diagnosis of kidney disease. *Curr Opin Nephrol Hypertens. 2013 Mar 10.*

85. Richard J. Perrin1,3,8,9*., Jacqueline E. Payton2,3., James P. Malone4, Petra Gilmore4, Alan E. Davis4, Chengjie Xiong6,8, Anne M. Fagan7,8,9, R. Reid Townsend4,5, David M. Holtzman4,7,8,9. Quantitative Label-Free Proteomics for Discovery of Biomarkers in Cerebrospinal Fluid: Assessment of Technical and Inter-Individual Variation. *PLOS ONE May 2013 | Volume 8 | Issue 5 | e64314*

86. Häggmark A, Byström S, Ayoglu B, Qundos U, Uhlén M, Khademi M, Olsson T, Schwenk JM, Nilsson P. Antibody-based profiling of cerebrospinal fluid within multiple sclerosis. *Proteomics. 2013 May 21. doi: 10.1002/pmic.201200580*

87. Jong-Yi Park†1,3, Jae-Hwan Kim†2, Yun-Jung Choi1, Kyu-Chan Hwang1, Seong-Keun Cho3, Ho-Hyun Park4, Seung-Sam Paik4, Teoan Kim5, ChanKyu Park1, Hoon Taek Lee1, Han Geuk Seo6, Soo-Bong Park7, Seongsoo Hwang and Jin-Hoi Kim. Comparative proteomic analysis of malformed umbilical cords from somatic cell nuclear transfer-derived piglets: *implications for early postnatal death BMC Genomics 2009, 10:511*

88. Chae JI, Cho SK, Seo JW, Yoon TS, Lee KS, Kim JH, Lee KK, Han YM, Yu K. Proteomic analysis of the extraembryonic tissue from cloned porcine embryos. *Mol Cell Proteomics. 2006 Sep;5(9):1559-66.*

89. Riding GA, Hill JR, Jones A, Holland MK, Josh PF, Lehnert SA. Differential proteomic analysis of bovine conceptus fluid proteins in pregnancies generated by assisted reproductive technologies. *Proteomics. 2008 Jul;8(14):2967-82.*

90. Zaiane O.CMPUT690 Principles of Knowledge Discovery in Databases. *Department of Computing Science, University of Alberta 1999*

91. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res. 2009;37(1):1-13*

92. Doniger SW, Salomonis N, Dahlquist KD, Vranizan K, Lawlor SC, Conklin BR. MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarraydata. *Genome Biol. 2003;4(1):R7. Epub 2003 Jan 6.*

93. Castillo-Davis CI, Hartl DL. GeneMerge--post-genomic analysis, data mining, and hypothesis testing. *Bioinformatics. 2003 May 1;19(7):891-2.*

94. Simona F, Laura S, Simona T, Riccardo A. Contribution of proteomics to understanding the role of tumor-derived exosomes in cancer progression: State of the art and new perspectives. *Proteomics. 2013 Feb 11.*

95. Tsuchida S, Satoh M, Kawashima Y, Sogawa K, Kado S, Sawai S, Nishimura M, Ogita M, Takeuchi Y, Kobyashi H, Aoki A, Kodera Y, Matsushita K, Izumi Y,Nomura F. Application of quantitative proteomic analysis using tandem mass tags for discovery and identification of novel biomarkers in periodontal disease. *Proteomics. 2013 May 21. doi: 10.1002/pmic.201200510.*

96. Wright B, Stanley RG, Kaiser WJ, Gibbins JM. The integration of proteomics and systems approaches to map regulatory mechanisms underpinning platelet function. *Proteomics Clin Appl. 2013 Jan;7(1-2):144-54. doi: 10.1002/prca.201200095.*

97. Handcock J, Deutsch EW, Boyle J: mspecLINE: bridging knowledge of human disease with the proteome. *BMC Med Genomics 2010, 3:7*

98. Shannon P, et al.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res 2003, 13(11):2498-2504*

99. Gupta M K et al., Proteomic analysis of parthenogenetic and in vitro fertilized porcine embryos. *J Proteomics 2009, 9, 1-15*

100. Wang S, Kou Z, Jing Z, et al. Proteome of mouse oocytes at different developmental stages. *PNA0053*

101. Morris J H, Meng E C, Ferrin T E. Computational tools for the interactive exploration of proteomic and structural data. *Mol Cell Proteomics. 2010 August; 9(8): 1703–1715.*

102. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc. 2009;4(1):44-57*

103. Smoot M E, Ono K, Ruscheinski J, Wang P, Ideker T, Cytoscape 2.8: new features for data integration and network visualization. *J Bioinformatics Vol. 27 no. 3 2011, page 431-432*

104. Millán PP. Visualization and analysis of biological networks. *Methods Mol Biol. 2013;1021:63-88*

105. Agapito et al., Visualization of protein interaction networks: problems and solutions. *BMC Bioinformatics 2013, 14(Suppl1):S1*

106. Lee HM, Dietz KJ, Hofestädt R. Computational construction of specialized biological networks. *J Bioinform Comput Biol. 2013 Feb;11(1):1340003*

107. Gebel S, LItchne R B, Schlage W K, et al, Construction of a Computable Network Model for DNA Damage, Autophagy, Cell Death, and Senescence. *Bioinformatics and Biology Insights 2013:7 97–117*

108. Schroeder M P, Gonzalez-Perez A, Lopez-Bigas N, Visualizing multidimensional cancer genomics data. *Schroeder et al. Genome Medicine 2013, 5:9*

109. Bauer-Mehren A. Integration of genomic information with biological networks using cytoscape. *Methods Mol Biol. 2013;1021:37-61*

110. Michel PE, Crettaz D, Morier P, Heller M, Gallot D, Tissot JD, Reymond F, Rossier JS. Proteome analysis of human plasma and amniotic fluid by Off-Gel isoelectric focusing followed by nano-LC-MS/MS. Electrophoresis. *2006 Mar;27(5-6):1169-81.*

111. Qian W J, Liu T, Monroe M E, Strittmatter E F et al., Probability-based evaluation of peptide and protein identifications from tandem mass spectroscopy and SEQUST analysis: the human proteome. *J Proteome Res. 2005, 4, 53-62*

112. Ashburner M et al. Gene Ontology: tool for the unification of biology. *Nat Genet. 2000 May ; 25(1): 25–29*

113. Jung YH, Gupta MK, Oh SH, Uhm SJ, Lee HT. Glial cell line-derived neurotrophic factor alters the growth characteristics and genomic imprinting of mouse multipotent adult germline stem cells. *Exp Cell Res. 2010 Mar 10;316(5):747-61*