CONSTRUCTION AND ANALYSIS OF PROTEIN-PROTEIN INTERACTION NETWORK ASSOCIATED WITH HUMAN SPERM-EGG INTERACTION

SOUDABEH SABETIAN FARD JAHROMI

A thesis submitted in fulfilment of the requirements for the award of the degree of
Doctor of Philosophy (Bioscience)

Faculty of Bioscience and Medical Engineering
Universiti Teknologi Malaysia

JUNE 2015
To my beloved family, especially my husband

“Alireza Valipour”
ACKNOWLEDGEMENT

In the name of God, the Most Gracious and the Most Merciful. All praises to God for the strengths and His blessing in completing this thesis.

Special appreciation goes to my supervisor, Assoc. Prof. Dr. Mohd. Shahir Shamsir Bin Omar, for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study. I would like to express my appreciation to my co-supervisor, Dr. Mohammed Abu Naser for his support and knowledge regarding this topic.

I would like to thank all of the BIRG team members especially Chew Teong Han for his kind support and help.

I wish to thank all the staff of Faculty of Biosciences and Medical Engineering and also Research Management Centre for solving the administrative problems regarding of this research.

Sincere thanks to my dear husband “Alireza V” for his kindness and moral support during my study.

Last but not least, my deepest gratitude goes to my beloved parents; Mr. Mahmoud S and Mrs. Nasrin V and also to my dear siblings; Sara, Golnar and Goodarz for their endless love, prayers and encouragement. Also not forgetting my friends, Sepideh, Faezeh, Ashraf and Bashir for all the emotional support, camaraderie, entertainment, and caring they provided.
ABSTRACT

Complete elucidation of sperm-egg interaction at molecular level is one of the unresolved challenges in sexual reproduction studies, and the understanding the molecular mechanism is crucial in overcoming difficulties in infertility and unsuccessful \textit{in-vitro} fertilization. Numerous molecular interactions in the form of protein-protein interactions mediate the sperm-egg membrane interaction process. Due to the various limitations of materials and difficulties in analyzing \textit{vivo} membrane protein-protein interaction, many efforts have failed to comprehensively consider the interaction mechanism at molecular level which mediates sperm-egg membrane interaction. The main purpose of this study was to identify the possible protein interaction in human sperm-egg interaction using the protein-protein interaction network approach. Datasets from varying databases containing information on membrane-associated proteins in sperm-egg binding and fusion process have been utilized to construct human sperm-egg interaction network using Cytoscape Software. The analyzing network represented new interactions in binding and fusion of sperm to egg. CD151 and CD9 in human oocyte had interaction with CD49 in sperm, and CD49 and ITGA4 in sperm interacted with CD63 and CD81 respectively in the oocyte. These results showed that the different integrins in sperm may be involved in human sperm-egg interaction. It also revealed novel interactions between acrosomal proteins in sperm and membrane proteins in oocyte. The Fn1, PDIa6 and ARSA from acrosome content of sperm have been shown to interact with ITGA9, IGSF8 and BMPR2 in oocyte, respectively. Topological analysis of the sperm-egg interaction network identified ITGB1, FN1, EGFR, ITGA3, ITGAV, ITGB3 and COL1A1 as putative drug targets to future study on sperm-egg interaction disorder. Using functional analysis of the network by ClueGo and ClueGo Pedia (two Cytoscape Plugins), the major molecular functions in sperm-egg interaction protein network was identified. The Interleukin-4 receptor activity, receptor signaling protein tyrosine kinase activity, manganese ion transmembrane transport activity were identified as the major molecular function group in sperm-egg interaction protein network. The disease association analysis using Database for Annotation Visualization and Integrated Discovery (DAVID) analysis represented the associated diseases with sperm-egg interaction disorder and indicated that sperm-egg interaction defects possess significant association with diseases such as cardiovascular, hematological and breast cancer. The Ingenuity Pathway Analysis (IPA) of the putative drug targets showed that the drugs ocriplasmin (Jetrea©), gefitinib (Iressa©), erlotinib hydrochloride (Tarceva©), clotitide, cetuximab (Erbitux©) and panitumumab (Vectibix©) are possible candidates for efficacy testing for the treatment of infertility cases. In conclusion, the result of this study has generated new knowledge regarding sperm-egg interaction that is for future proteomic studies.
ABSTRAK

### TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DECLARATION</td>
<td>ii</td>
</tr>
<tr>
<td></td>
<td>DEDICATION</td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGEMENT</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td></td>
<td>ABSTRAK</td>
<td>vi</td>
</tr>
<tr>
<td></td>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>xiii</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>xv</td>
</tr>
<tr>
<td></td>
<td>LIST OF SYMBOLS</td>
<td>xvii</td>
</tr>
<tr>
<td></td>
<td>LIST OF ABBREVIATIONS</td>
<td>xviii</td>
</tr>
<tr>
<td></td>
<td>LIST OF APPENDICES</td>
<td>xxi</td>
</tr>
</tbody>
</table>

1 INTRODUCTION

1.1 Background of Study 1
1.2 Problem Statement 4
1.3 Objectives 5
1.4 Scope of study 5
1.5 Significance of study 6

2 LITERATURE REVIEW

2.1 Introduction 7
2.2 Fertilization 8
2.3 Infertility 11
2.3.1 Male infertility
2.3.2 Unexplained Male Infertility
2.4 Fertilization Defects
   2.4.1 Sperm-ZP Binding Defect
   2.4.2 Acrosome Reaction Defect
   2.4.3 Fusion Defect of the Acrosome- Reacted Sperm with the Oocyte Plasma Membrane
2.5 Sperm – Oocyte Interactions
   2.5.1 Approaches to Identify Protein Candidates In Sperm- Oocyte Interaction
   2.5.2 Candidate Sperm Proteins in Sperm–Egg Interaction
      2.5.2.1 IZUMO1 on Sperm
      2.5.2.2 ADAMs on Sperm
      2.5.2.3 Other Sperm Proteins
   2.5.3 Candidate Oocyte Proteins in Sperm– Egg Interaction
      2.5.3.1 Integrin on Eggs
      2.5.3.2 Tetraspanins on Eggs
      2.5.3.3 Glycosyl Phosphatidylinositol- Anchored Protein on Eggs
2.6 Experimental Methodologies to Detect PPI
2.7 Computational Methods to Detect PPI
   2.7.1 PPI datasets
2.8 PPI network
   2.8.1 Visualization Tools
   2.8.2 Network Topological Analysis to Discover Essential Proteins in PINs
   2.8.3 Functional Analysis, Clustering and Drug Discovery
2.9 Assess the Quality of Molecular Interaction
   2.9.1 Confident Experimental and Physical
3 RESEARCH METHODOLOGY

3.1 Introduction

3.2 Initial Collection of Associated Proteins with Human Sperm and Egg

3.3 Mapping of Initial Proteins with Different ID to used ID

3.4 Data Mining Protein-Protein Interactions

3.4.1 Sources of Experimental and Physical Interactions

3.4.2 Sources of Predicted and Functional Interactions

3.5 Query and Download Format for Different PPIs

3.6 Construction of Sperm-Egg Interaction Network

3.6.1 Calculation of Overlapping Two Protein Interaction Network (PIN)

3.6.2 Removal of the Duplicated Protein-Protein Interaction (PPI)

3.6.3 Identification of Membrane Proteins Involved in Sperm-Egg Interaction Network

3.6.4 Selection of High Confident Protein Interactions

3.7 Network Validation

3.7.1 Network Validation Based on the Definition of Clustering and Function
Agreement with Annotated Protein Databases 62

3.7.2 Subcellular Location of the Proteins in Sperm--Egg Network 63

3.8 Identification of New Protein Interaction in Human Sperm-Egg Interaction 63

3.9 Topological Analysis of Protein Interaction Network 64

3.9.1 Degree Exponent Distribution and Coefficient of Determination 64

3.9.2 Node Degree 64

3.9.3 Betweenness Centrality (BC) 64

3.9.4 Closeness Centrality (CC) 65

3.9.5 Shortest Path 65

3.10 Molecular Function Analysis 65

3.10.1 Functional Network 66

3.11 Associated Diseases, Associated Drug Targets and Drugs Analysis 66

3.11.1 DAVID Analysis 66

3.11.2 IPA Analysis 67

3.12 Construction of the Protein-Protein Interaction Network Associated with Azoospermia and Teratozoospermia 67

3.13 Comparison Sperm-Egg Interaction Network with Azoospermia and Teratozoospermia PINs 67

3.14 Summary of methods 68

4 RESULTS AND DISCUSSION 69

4.1 Introduction 69

4.2 Mining of Associated Proteins with Human Sperm and Egg 69

4.3 Mining of Protein-Protein Interaction of Human Sperm and Egg 70
4.4 Construction of the Sperm and Egg Protein-Protein Interaction Networks
  4.4.1 Construction of Sperm-Egg Interaction Network

4.5 Selection of High Confident Predicted Protein Interactions

4.6 Subcellular Location of the Proteins in Sperm-Egg Network

4.7 New Protein Interaction in Human Sperm-Egg Interaction

4.8 Topological Analysis of Protein Interaction Network
  4.8.1 Important Nodes of PIN and Identifying Putative Drug Targets
  4.8.2 Experimental Evidence for the Essentiality of Predicted Potential Targets

4.9 Network Validation Based on the Definition of Clustering and Agreement with Annotated Protein Function Databases
  4.9.1 Clustering of Sperm-Egg Protein Interaction Network
  4.9.2 Gene Ontology (GO) Enrichment Analysis

4.10 ClueGO Analysis to Identify Function Groups

4.11 Functional Network and Disease Correlations
  4.11.1 David Analysis

4.12 Identifying Drug Targets in PPI Network

4.13 Comparison Sperm-Egg Interaction Network with Azoospermia and Teratozoospermia PINs
  4.13.1 Construction of the Protein-Protein Interaction Network Associated with Azoospermia
  4.13.2 Comparison Azoospermia PIN and Sperm-Egg Interaction Network
4.13.3 Construction of the Protein-Protein Interaction Network Associated with Teratozoospermia

4.13.4 Comparison Teratozoospermia PIN and Sperm-Egg Interaction Network

4.14 Summary of Results and Discussion

5 CONCLUSION AND RECOMMENDATION

5.1 Conclusion

5.1.1 To Mine, Identify and Map Protein-Protein Interaction Network for Human Sperm-Egg interaction

5.1.2 To Find New Protein-Protein Interaction Associated with Human Sperm-Egg Interaction

5.1.3 To Identify Putative Drug Targets in Sperm-Egg Interaction Network

5.1.4 To Predict The Functional Network for Interaction between Sperm and Egg

5.1.5 To Investigate Associated Diseases with the Sperm-Egg Interaction Disorder

5.2 Research Contributions

5.3 Limitation of Study

5.4 Recommendation

REFERENCES

Appendices A-I
CHAPTER 1

INTRODUCTION

1.1 Background of Study

Sperm-egg interaction is a unique cell-cell connection process in sexual reproduction that involves two gametes recognize, bind and eventually fuse with each other (Wortzman et al., 2006). The potential intermediaries of molecular process in sperm– oocyte fusion and binding have been studied over the past 20 years (Kaji and Kudo, 2004; Primakoff and Myles, 2002; Stein et al., 2004). However, the molecular adhesions that intercede sperm– egg membrane interaction event is still poorly understood due to limitations in studying the process in-vitro and the analysis of membrane associated sperm-egg interactions in human (Brewis et al., 2005).

Fertilization is the process in which sperm and egg recognize, bind and fuse with each other (Inoue et al., 2005; Inoue et al., 2010; Kaji and Kudo, 2004). During this process, many molecular interactions in the form of protein-protein interactions will mediate the sperm-egg binding process. The elucidation of sperm-egg interaction at the molecular level is one of the unresolved problems in sexual reproduction, and understanding the molecular mechanism is crucial in solving problems in infertility and in vitro fertilization failure (Evans, 2012).
Infertility is a current medical problem affecting 10–15% of reproduction-aged couples in the world (Hotaling et al., 2011). The percentage of infertility is being higher in underdeveloped countries in which limited resources for diagnosis and treatment exist (Hamada et al., 2011b).

Male factor problems, mainly defined by abnormal semen analysis, are the single most conventionally diagnosed reason of infertility. In vitro fertilization (IVF) is implemented for men with no sperm dysfunction. Nevertheless it is surprising that the complete fertilization fail is still prevalent event in the process of IVF. This suggests that the sperm dysfunction is not certain even with common semen analyzing while is a considerable cause of fertilization failure. This phenomenon is called unexplained male infertility and remains an unknown syndrome as researchers have limited information regarding the clinical nature of the sperm dysfunction (Brewis et al., 2005; Hamada et al., 2012). Recent studies have suggested that sperm deficiencies such as zona binding, the zona-induced AR (Acrosome Reaction) or oocyte plasma membrane fusion defects are significant causes of reduced fertilization and total fertilization fail in assisted reproductive technologies. Therefore the major cause of fertilization failure in conventional IVF of unexplained male infertile is due to abnormalities of sperm–oocyte interaction and penetration (Brewis et al., 2005; Hamada et al., 2012; Liu and Baker, 2000).

The proteins-protein interactions (PPI) can interact in various ways, ranging from direct physical relationships amongst proteins in a complex, to indirect interactions that happen amid components of specific protein pathways. These links between the proteins in a protein set can make a protein-protein interaction network (PIN) and all proteins which involved in PINs are spatially or temporally engaged to interact with other proteins within the process as well as functioning as an indirect interacting members of the same pathway (Brewis et al., 2005; Strong and Eisenberg, 2007). Currently, the discovery of protein connections has been assisted by developments in both biochemical (Brown and Botstein, 1999; Ho et al., 2002; Ito et al., 2001; Uetz et al., 2000) and computational methods (Janga and Moreno-Hagelsieb, 2004; Salgado et al., 2000; Strong et al., 2003), which have produced precious awareness into the fundamental building of protein interactions in cellular
networks (Jeong et al., 2000; Marcotte et al., 1999; Ravasz et al., 2002; Rives and Galitski, 2003; Wuchty, 2002).

Until now, the original experimental methods, for instance, mass spectrometry and proteomics approaches have been used to identify protein-protein interaction between human sperm and egg (Evans, 2012). However, due to the complexity of the problems like difficulties in analyzing in-vivo membrane PPIs, denaturing proteins and possible disruption in the formation of protein complexes, many efforts have failed to comprehensively elucidate the fusion mechanism and the molecular interactions that facilitate sperm-egg membrane fusion, leaving the molecular mechanism during sperm-egg interaction still poorly understood (Brewis et al., 2005; Evans, 2012; Kaji and Kudo, 2004).

Existing computational approaches, in addition to experimental methods, can assist our understanding of PPIs at various levels. The computational approaches may be utilised for comprehensive examination or perform a wide scale analysis across large datasets (Lu et al., 2002; Salwinski and Eisenberg, 2003). This approach signifies the multifaceted association of proteins with PPI links, in a protein interaction network and would help to comprehend how signalling pathways linked with a disease are connected (Xu and Li, 2006). By using computational methods, it is possible to identify a comprehensive information about complex diseases not discernible in laboratory methods such as the putative disease target genes, putative drug targets, new prognostic biomarkers and understanding the mechanism of complex disease by network analysis (Devaux et al., 2010; Flórez et al., 2010; Yao et al., 2010). Thus, computational approach denotes a novel technique for investigating the complex impacts of candidate genes that are connected to complex diseases, and is also worthwhile in recognizing important drug targets and genes in a disease and in complex biological systems (Hwang et al., 2008; Kim and Kim, 2009).


1.2 Problem Statement

Sperm–oocyte interaction is one of the most remarkable events in fertilization process and its surrounding molecular events have been studied in various attempts over the last 20 years. Due to detection the molecules that immediate human membrane sperm-oocyte interaction, different techniques have been used and represented a low number of sperm and oocyte proteins which have a fusogenic role in sperm–egg interaction event (Evans, 2012).

Because of difficulties in vitro analyzing membrane protein-protein interactions and the limitation of technical materials, the molecular mechanisms involved in sperm–egg membrane adhesion are still poorly understood (Inoue et al., 2010; Kaji and Kudo, 2004). Moreover, the focus of the wide majority of the previous studies was on utilizing animal models to understand subsequent related molecular mechanisms, but in order to understand the human infertility and fertilization mechanisms, the emphasis must be on the human model. The study on human model in this conception could be predominantly a difficult task because of the modicum of human oocytes or embryos (Ola et al., 2001; Tournaye et al., 2002). Elucidation of sperm-egg interaction at molecular level is one of the unresolved problems in sexual reproduction and the major cause of fertilization failure in conventional IVF and unexplained male infertile is due to defect in sperm–oocyte interaction and penetration (Brewis et al., 2005; Hamada et al., 2012).

Therefore understanding the molecular mechanism surrounding human sperm–egg interaction is crucial in solving problems in infertility and in vitro fertilization failure. To date, no clear candidate interaction proteins have identified and the molecular interaction events still poorly understood. There is a lack of a comprehensive protein-protein interaction network for human sperm-egg interaction and also a lack of a protein links network of human sperm-egg interaction process with other cellular pathways (Evans, 2012).
1.3 **Objectives**

The main purpose of this study was to reveal possible protein interaction and associated molecular function during sperm-egg interaction using protein interaction network approach. The objectives of this research are:

1. To mine, identify and map protein-protein interaction network for human sperm-egg interaction
2. To find new protein-protein interaction associated with human sperm-egg interaction
3. To identify putative drug targets in sperm-egg interaction network
4. To predict the functional network for interaction between sperm and egg
5. To investigate associated diseases with the sperm-egg interaction disorder

1.4 **Scope of study**

In this research, the focus was on the proteins that involved in human sperm-egg interaction. These proteins were mined from PubMed research articles, UniProt database and protein-protein interaction databases which store the physical and functional interactions. It has also included the predicted interactions from predicted database. The results were then mapped using Cytoscape software as a general platform for complex network analysis and visualization. The created network has been analyzed using Cytoscape plugins: Allegro-Mcode, Bingo, ClueGO and ClueGO Pedia. Then the diseases associated with sperm-egg interaction disorder would be discovered using DAVID software and GAD database. In order to find out known drug targets in sperm-egg interaction network, the analysis using IPA (Ingenuity Pathway Analysis) software have been carried out.
1.5 Significance of study

The elucidation of sperm-egg interaction at the molecular level is one of the main unresolved problems in sexual reproduction, and it is also crucial to understand the molecular mechanism in solving problems in infertility and failed IVF. Many molecular interactions in the form of protein-protein interactions (PPI’s) mediate the sperm-egg membrane interaction process. But, due to the complexity of the problems such as difficulties in analyzing \textit{vivo} membrane PPI’s, many experimental efforts have failed to comprehensively elucidate the fusion mechanism and the molecular events that mediate sperm-egg membrane interaction (Brewis \textit{et al.}, 2005; Evans, 2012; Hamada \textit{et al.}, 2012; Kaji and Kudo, 2004).

In addition to experimental methods, computational methods can deal with PPIs at various levels. Computational approach for constructing protein interaction networks, utilizing genomic and protein sequence information contain a study of the absence or presence of genes in associated species, co-occurrence of sequence domains, gene fusion occurrences, conservation of gene neighborhood, interrelated mutations on protein surfaces, the resemblance of phylogenetic trees, co-expression and functional annotations (Salwinski and Eisenberg, 2003). Sometimes, combinations of these aspects are implemented to foresee novel interactions or to appraise the trustworthiness of PPIs measured experimentally (Brown and Jurisica, 2005; Jensen \textit{et al.}, 2009).

Therefore, this study was carried out to promote a protein-protein interaction map that reveals the possible protein interaction, major molecular functions and disease association genes in sperm-egg interaction protein network. This study facilitates future research in biomarker detection and new drug designs for diagnosis and treatment of the infertile cases, who suffer from failures in assisted reproductive technology.


Batagelj, V. and Mrvar, A. (2004). Pajek — Analysis and Visualization of Large Networks Graph Drawing Software. In M. Jünger & P. Mutzel (Eds.), *Graph Drawing Software* (pp. 77-103): Springer Berlin Heidelberg.


Berruti, G., Ripolone, M. and Ceriani, M. (2010). USP8, a regulator of endosomal sorting, is involved in mouse acrosome biogenesis through interaction with
the spermatid ESCRT-0 complex and microtubules. Biology of reproduction, 82(5), 930-939.


Coonrod, S. A., Naaby-Hansen, S., Shetty, J., Shibahara, H., Chen, M., White, J. M., et al. (1999). Treatment of Mouse Oocytes with PI-PLC Releases 70-kDa (p< i> 5) and 35-to 45-kDa (p< i> 5.5) Protein Clusters from the Egg


Diaz, E. S., Kong, M. and Morales, P. (2007). Effect of fibronectin on proteasome activity, acrosome reaction, tyrosine phosphorylation and intracellular...


novel ones, and analysis of integrins as candidate FN1 receptors in bovine preimplantation embryos. *BMC developmental biology*, 9(1), 1.


He, Z.-Y., Brakebusch, C., Fässler, R., Kreidberg, J. A., Primakoff, P. and Myles, D. G. (2003). None of the integrins known to be present on the mouse egg or to be ADAM receptors are essential for sperm–egg binding and fusion. *Developmental biology*, 254(2), 226-237.


male reproductive tract and gametes and evidence of its participation in fertilization. *Molecular human reproduction*, 14(10), 561-571.


randomized controlled trial and a meta-analysis of the literature. *Fertility and sterility*, 78(5), 1030-1037.


