

RESEARCH ARTICLE

Open Access



Putative human sperm Interactome: a networks study

Alessandra Ordinelli^{1†}, Nicola Bernabò^{1*†} , Massimiliano Orsini², Mauro Mattioli^{1,2} and Barbara Barboni¹

Abstract

Background: For over sixty years, it has been known that mammalian spermatozoa immediately after ejaculation are virtually infertile. They became able to fertilize only after they reside for long time (hours to days) within female genital tract where they complete their functional maturation, the capacitation. This process is finely regulated by the interaction with the female environment and involves, in spermatozoa, a myriad of molecules as messengers and target of signals. Since, to date, a model able to represent the molecular interaction that characterize sperm physiology does not exist, we realized the Human Sperm Interactome Network3.0 (HSIN3.0) and its main component (HSIN3.0_MC), starting from the pathway active in male germ cells.

Results: HSIN3.0 and HSIN3.0_MC are scale free networks, adherent to the Barabasi-Albert model, and are characterised by an ultra-small world topology. We found that they are resistant to random attacks and that are designed to respond quickly and specifically to external inputs. In addition, it has been possible to identify the most connected nodes (the hubs) and the bottlenecks nodes. This result allowed us to explore the control mechanisms active in driving sperm biochemical machinery and to verify the different levels of controls: party vs. date hubs and hubs vs. bottlenecks, thanks the availability of data from KO mice. Finally, we found that several key nodes represent molecules specifically involved in function that are thought to be not present or not active in sperm cells, such as control of cell cycle, proteins synthesis, nuclear trafficking, and immune response, thus potentially open new perspectives on the study of sperm biology.

Conclusions: For the first time we present a network representing putative human sperm interactome. This result gives very intriguing biological information and could contribute to the knowledge of spermatozoa, either in physiological or pathological conditions.

Keywords: Spermatozoa, Interactome, Biological networks, Scale free, Small world, Male infertility, Systems biology

Background

Mammalian spermatozoa are known to acquire their fertilizing ability only after a long and complex series of biochemical modification that recognises different steps. It starts within the testis, with spermatogenesis, and continues in epididymis, where spermatozoa undergo a series of important chemical modifications, and then it culminates with a process, called capacitation, that occurs within the female genital tract. Capacitation was discovered in early fifties, of the twentieth century independently by Austin [1, 2] and

Chang [3] and in next sixty-five years, it has attracted the attention of groups of researchers from around the world. Nowadays we know several molecular events involved in spermatozoa physiology and in signal transduction pathways. It is clear that peculiar biological features characterize male gametes: virtually they are transcriptionally silent, their cytosol is absent, and they are the only cellular typology that exert its function in an organism other than that produced them. Consequently, they adapted their biochemical machinery to these conditions and evolved a complex control strategy. Their function is finely modulated by an uninterrupted dialogue with the surrounding environment, constituted by the cells and the fluids of female genital tract. By this way, they are able to physiologically responding to physical and chemical

* Correspondence: nbernabo@unite.it

[†]Equal contributors

¹Faculty of Bioscience and Technology for Food, Agriculture and Environment, University of Teramo, Teramo, Italy

Full list of author information is available at the end of the article



signals that arise from the neuro-endocrine activity of female. It has been found that sperm membrane and cytoskeleton play a key role in this process. Indeed, both of them are highly dynamical structures and undergo a deep reorganization during post-ejaculatory life of male gametes. The plasma membrane (PM) and the outer acrosome membrane (OAM) lose their asymmetry and become more fluid and isotropic, thus acquiring the ability to fuse (fusogenicity). The actin, first, polymerizes forming a network that separates PM and OAM, then it depolymerizes, allowing their contact and favoring the start of acrosome reaction (AR). All these events are caused and cause important modifications in the concentration of messengers, such as intracellular calcium concentration, $[Ca^{2+}]_i$; intracellular pH (pH_i); cAMP, nitric oxide (NO), and intracellular concentration of bicarbonate ($[HCO_3^-]_i$). Consequently, several target (kinases, phosphatases, lipases, cyclases, etc.) are activated originating a specific and regulated metabolic response [4–6].

Until a few years ago, the experimental works of researchers approaching the study of these events were limited for technical reasons. More recently, the adoption of high-throughput technologies allowed obtaining a myriad of data. This had great positive consequences, as the rapid increase in the number of molecules and molecular event studied and the great amount of functional data obtained. In particular, the studies carried out by using proteomic approaches are giving very interesting results and offer important suggestion of new phenomena to be investigated [7, 8]. On the other side, unfortunately, we still lack of the ability to take inference from the data, in part due to the unavailability of models designed to take into account the biological complexity of these events. Indeed, the data per se are useless without the possibility to aggregate and interpret them. The proof is that in a large number of patients with fertility problems and undergoing assisted reproductive technology (ART), it is impossible to reach an etiological diagnosis [9–11].

To overcome this limitation, here, we explored the possibility to apply a computational modelling approach, based on networks theory, to the definition and the study of spermatozoa interactome. Briefly, we used the data from published proteomic studies on pathways expressed in spermatozoa, to build a model in which the molecules are represented as nodes and the interactions among them as links, thus originating a network. We had already used this approach for studying specific events of sperm biology, such as the epididymal maturation [12], the capacitation [13–15], the cytoskeleton dynamics [16], or the sperm-egg binding and recognition [17]. Here, we propose a more global view of mature spermatozoa physiology, and the study of network architecture, the so-called

topology, has allowed us to infer important biological features and to lay the foundations for further studies.

Results

Human spermatozoa interactome network (HSIN)

As result of our data collection and refinement activity, we obtained different networks: HSIN1.0 composed by 225 connected components (data not shown), HSIN2.0 composed by 104 connected components (data not shown), and HSIN 3.0 composed by 39 connected components (whose topological parameters are represented in Table 1). The final step of our work was the realization of HSIN3.0, composed by 25 connected components, 7891 nodes and 14,712 links, which represent the most complete evolution of HSIN. Since the larger connected component has 7758 nodes, thus representing 98.3% of HSIN 3.0, all the further analysis were carried out on this network, called HSIN 3.0 Main Component, HSIN3.0_MC (Table 1 and Table 2; Fig. 1).

Hubs: identification and subpopulations

We found 190 hubs in HSIN3.0_MC (for hubs list and topological parameters see Additional file 1). Interestingly the KDE analysis we carried out allowed to identifying the presence of hubs subpopulations when analysing the clustering coefficient, as it is evident from the data depicted in Fig. 2.

Identification of bottleneck nodes

The analysis of HSIN3.0_MC for bottleneck carried out with cytoHubba allowed to identify the nodes characterized by a higher score (see Additional file 2).

Effect of the deletion of gene relative to protein listed in HSIN3.0_MC in KO mice on male fertility

We found that the genes coding for about 11% (248 on 2130) of the proteins present in HSIN3.0_MC have been studied in terms of fertility on mice KO models. More in detail, we investigated four different KO phenotypes: “male infertile”, “male hypofertile”, “male fertility affected”,

Table 1 main topological parameter computed on HSI 1.0, HSI 2.0, HSI 3.0, MC_HSI 3.0 networks

Parameter	HSI 3.0	MC_HSI 3.0
Number of nodes	7892	7758
Number of edges	14,712	14,534
Clustering coefficient	0.073	0.070
Connected components	25	1
Network diameter	37	37
Shortest paths	27,139,995 (43%)	27,139,415 (45%)
Characteristic path length	10.320	10.321
Avg. number of neighbours	3.727	3.746

Table 2 Node degree distribution and correlation of node degree with clustering coefficient in MC_HSI 3.0

Node degree distribution	In-degree	Out-degree
- γ	-1.850	-1.764
R	0.985	0.963
R ²	0.781	0.814
Clustering coefficient vs. node degree		
- γ	-1.628	
r	0.223	
R ²	0.620	

and “male fertility unaffected/normal phenotype”. Interestingly, as it shown in Table 3, the hubs are more studied than the non-hubs genes (the data available are 35.6% for hubs and 19.6% for non-hubs), and the percentage of protein whose genes are related with fertility did not differ in hubs in comparison with non-hubs ($p = 0.21991$, Chi² test). On the contrary, the deletion of genes codifying for the bottleneck protein in 100% of cases caused infertility.

Discussion

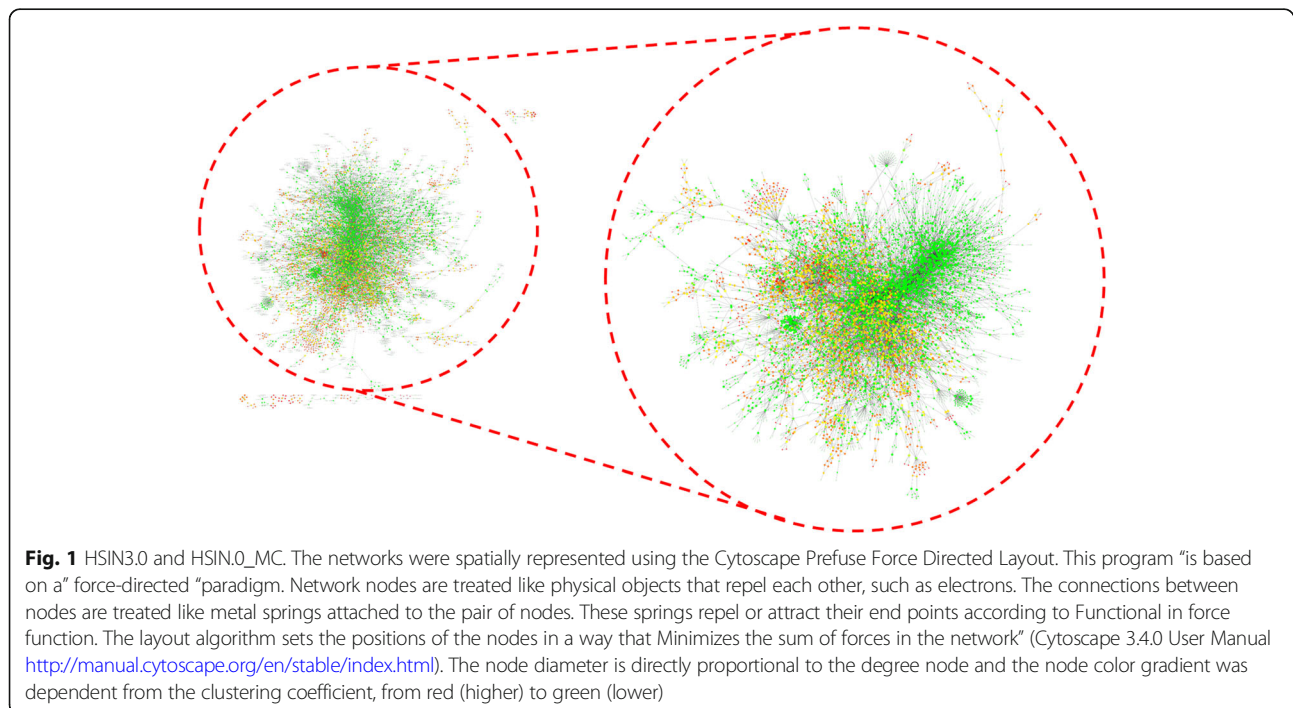
In this work, we realized the networks representing the putative Human Spermatozoa Interactome (HSIN3.0 and HSIN3.0_MC), starting from published proteomic data. Once obtained the network model, we assessed its topology to infer important biological information. As first, it is evident that HSIN3.0 and HSIN3.0_MC are characterised by a specific topology: they are scale free

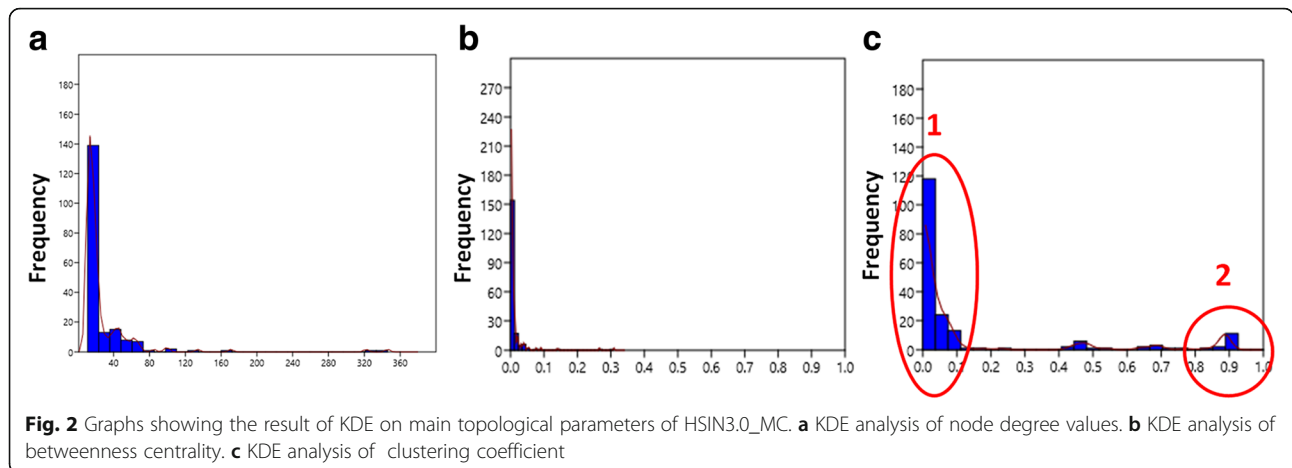
networks in keeping with the Barabasi-Albert (BA) model. They are characterised by a wide heterogeneity of node degree of nodes. In particular, in BA networks a low number of highly connected nodes (the “hubs”) co-exists with a higher number of scarcely connected nodes [18]. From a dynamical point of view, these networks grow over the time, due to the attachment of new nodes preferentially to the hubs: the more one node is connected, the higher is the probability that it attracts a new node (preferential attachment) [19]. Thus, the probability p_i that the new node is connected to node i is:

$$p_i = k_i \sum_j k_j$$

where k_i is the degree of node i and the sum is made over all pre-existing nodes j [19, 20]. Finally, the node degree of nodes within the network follows a power-law distribution, i.e. the probability that a node has k links follows $P(k) \sim k^{-\gamma}$, where the degree exponent γ usually ranges between 2 and 3. Differently from other models of scale free networks such as the hierarchical networks, in BA networks the clustering coefficient, $C(k)$, is independent of k and the average path length follows $l \sim \log \log N$ [20, 21].

In this context, we looked for the most connected nodes, to identify biologically relevant molecules. The analysis of hubs offers very intriguing information and very thought-provoking data. Indeed, together with molecules well known as ubiquitous cell components (as it is the case of ATP) or as modulators of sperm





physiology (see for instance Ca^{2+}) we found several unsuspected nodes. They represent molecules specifically involved in function that are thought to be not present or not active in sperm cells, such as control of cell cycle, proteins synthesis, nuclear trafficking, and immune response. In all these cases, it will be very important to explore these functions in spermatozoa to verify the real biological meaning of this result. Anyway, this study, with other proteomic studies (see for reference [7]), could potentially open new perspectives on the study of sperm biology.

To date not all the researchers agree in definition of hubs and a univocal opinion on their topological and functional proprieties does not exist [22]. Anyway, our data seem to suggest that in sperm Interactome network it will be possible to differentiate hubs involved in control of a wide variety of functions and hubs devoted in controlling a specific function. In keeping with Han et al. [23] we analysed the clustering coefficient values of the most connected nodes within the network (the hubs) carrying out a kernel density estimation. As a result, we identified different subpopulations of nodes (see Additional file 3). The bigger one is characterised by very low values of this parameter, ranging from 0 to about 0.17, while the other one

is smaller and its cluster coefficient is significantly higher, from 0.83 to 0.89. It is possible to speculate that the different clusterization of hubs could be related to a different functional meaning of the corresponding molecule in sperm physiology. This difference seems to be in agreement with the hypothesis of Han and colleagues, which defined two different types of hubs “party’ hubs, which interact with most of their partners simultaneously [and that are characterized by lower values of clustering coefficient], and ‘date’ hubs, which bind their different partners at different times or locations [and that display higher values of clustering coefficient]” [23].

For instance, here, the most important (in term of node degree) party hub is ATP, which is obviously involved in a myriad of chemical reactions in different times and sub-cellular locations, thus behaving as general hub. As energy substrate, in male germ cells ATP is produced by glycolysis and mitochondrial oxidative phosphorylation, the two major metabolic pathways expressed in spermatozoa that are localized in distinct cellular sub-compartments: glycolysis occurs mainly in the fibrous sheath of the flagellum where glycolytic enzymes are tightly anchored, while oxidative phosphorylation occurs in mitochondria, in the sperm midpiece [24]. The preferential use of one pathway is highly species specific [25].

The second subpopulation (date hubs) is characterized by a scarce clusterization and by the presence of members of the interferon family. As it is known in somatic cells, interferon is involved in immune response, being released in presence of pathogens such as viruses, bacteria, and parasites. In addition, interferon activates immune cells, such as natural killer cells and macrophages, and is able to increase host defences by up-regulating antigen presentation. Its function in spermatozoa has poorly studied, and there are only a few reports about its biological activity in mature male germ cells. During spermatogenesis, it has been supposed that INF could play a key role, as suggested by Hibi et al. [26] that found that the administration of α -IFN

Table 3 phenotypical effects of the deletion of gene relative to protein listed in HSIN3.0_MC in KO mice on male fertility

	Hubs N (%)	Non-hubs N (%)
Male infertility	5 (23.8)	96 (27.8)
Male hypofertility	3 (14.3)	60 (17.4)
Male fertility affected	3 (14.3)	95 (27.5)
Male fertility unaffected/normal phenotype	10 (47.6)	94 (27.2)
Total	21 (100)	345 (100)

Male infertility: male KO is infertile; Male hypofertility: male KO has reduced fertility in comparison with wild type; Male fertility affected: KO mice have a damage in reproductive function with unclear or unknown effects on male fertility; Male fertility unaffected/normal phenotype: the KO mice have the same fertility of wild type

improved testicular spermatogenesis and increased the epididymal sperm concentration in the rat. Interestingly, they found also that progressive motility of the spermatozoa was unaffected by α -IFN treatment, while serum LH levels were decreased and serum testosterone and FSH levels were significantly increased. More recently, Satie et al. [27] compared a mouse strain transgenic for IFN $_{\gamma}$ (Tg10) and a sister strain lacking the IFNAR1 subunit of IFNAR(Tg10-*Ifnar1*^{-/-}), both strains expressing the transgene in the testis. The Tg10 mice, but not the double mutant Tg10-*Ifnar1*^{-/-}, showed altered spermatogenesis. The most important IFNAR-dependent histological alteration was a higher apoptosis index in all germ cell categories that occurred 3 weeks after the onset of IFN production, at post-natal day 20. In addition, the Authors found that several known interferon-stimulated genes were up-regulated in Tg10 Sertoli cells and in prepachytene germ cells. At day 60, Tg10 males were sterile and Sertoli cells showed increased concentration of anti-Mullerian hormone and decreased production of inhibin B. Based on these findings, they concluded that type I interferon signalling could be involved in etiopathogenesis of male idiopathic infertilities by affecting the interplay between germ cells and Sertoli cells.

The distinction between date and party hubs could be important for designing drugs or diagnostic tools in clinical andrology [28].

A second class of nodes that exert a strong control is that of bottlenecks, mainly involved in control of information flow within the network. As for hubs, several molecules known to be involved in sperm physiology represent the bottlenecks (see Additional files 1 and 2). Interestingly, if compared to hubs, in bottlenecks there are less represented proteins (28 vs. 59, respectively) and more represented reactions (81 vs. 9, respectively), thus indicating a stronger relevance of control on information flow (see above, the results of KO mice analysis).

To assess the reliability of our theoretical data, we evaluated the results of the analysis of reproductive phenotypes of KO mice. In other words, we compared the data regarding HSIN3.0_MC hubs and bottlenecks after removal of correspondent genes and their effects on fertility. It is very interesting that we found that the percentage of the frequency of fertility phenotypes in hubs and in non-hubs nodes did not differ significantly ($p = 0.21991$) (see Table 3), while the deletion of genes codifying for the bottleneck protein in 100% of cases caused infertility.

This finding lead us to conclude that in our network the bottlenecks play a more important role than the hubs. The reason of this different biological function could be found in the specific topology of the network. Indeed HSIN3.0_MC is constituted by direct links (it is the union of several pathways) so it could be considered a regulatory network. As elegantly demonstrated by Yu et al. [29] in that case the bottleneck are more essential

than hubs, while in interaction networks the hub are the main controller of networks.

Obviously, this experiment has important limitations due to the small number of bottlenecks, to the species-specific differences, and to the impossibility to test non-protein nodes. Anyway, in our opinion, this finding could have important implications because it confirms the difference in term of function between regulatory and interaction networks and, from a more general point of view, it allows to explore a complex biological function by using an in silico (network)-in vivo (KO mice) by controlling their complexity [30].

Once identified the hubs and the bottlenecks in HSIN3.0_MC, using the function “intersection” of Advanced Network Merge (in Cytoscape Tools menu) we computed the intersection of the network of most hubs with that of bottleneck, both of them obtained with cytoHubba. By this way, we obtained the network (Human Sperm Interactome Control Network, HSICN, see Fig. 3) that represents the backbone of control mechanism involved in spermatozoa physiology. Its analysis highlighted the idea that a few of nodes act as controllers of whole the network, allowing inferring two important biological implications. As first, we demonstrated that it is possible to create a map of control mechanisms of male gametes function, with enormous potential applications in the study of sperm physiology and pathology. Then, in several cases, the controllers are molecules poorly or not at all studied in spermatozoa. This, in one hand, could explain the inability to perform an etiological diagnosis in a large number of patients (unexplained infertility of male origin) and, on the other hand, could suggest several new targets for new studies.

Still looking at mechanisms active in leading sperm physiology, the results of protein characteristics analysis carried out using Uniprot are very interesting. As search parameters, we used factors known or supposed to be involved in sperm physiology and biochemistry, with the purpose to define the regulatory nodes within the network. We found that the protein present in HSIN3.0_MC are regulated by metal binding (12.3%), calcium binding (2.12%), pH (1.29%), and temperature (0.32%) (see Additional files 4, 5, 6 and 7) (we reported the percentages indicated by Uniprot, that are underestimated).

Metal binding and calcium binding

It is known that spermatozoa biochemical machinery is under the control of metallic ions. Here we found that Mg²⁺, Ca²⁺, Zn²⁺, and Fe²⁺ are hubs of HSIN3.0_MC with 41, 37, 21, and 14 links. This datum is very interesting in the light of the presence in oviduct of metals (either in physiological conditions or in response to neuro-endocrine signals), in fact all of them are enzyme cofactors, thus concur in modulating whole

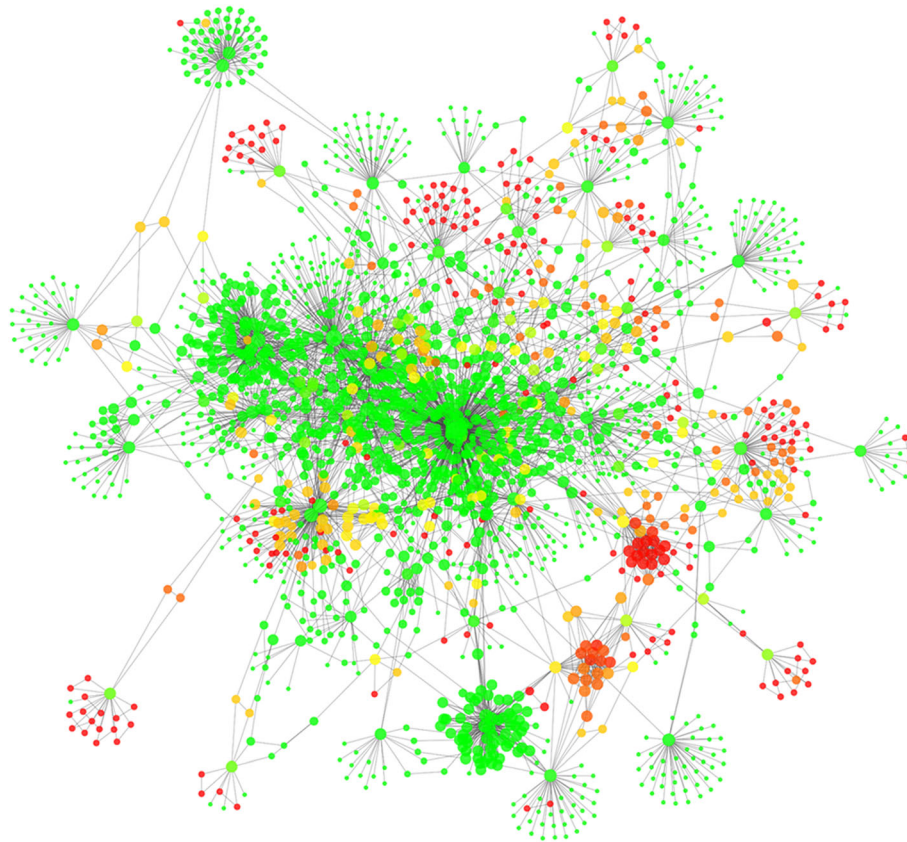


Fig. 3 HSICN. The networks was spatially represented using the Cytoscape Prefuse Force Directed Layout. This program "is based on a" force-directed "paradigm. Network nodes are treated like physical objects that repel each other, such as electrons. The connections between nodes are treated like metal springs attached to the pair of nodes. These springs repel or attract their end points according to functional in force function. The layout algorithm sets the positions of the nodes in a way that minimizes the sum of forces in the network "(Cytoscape 3.4.0 User Manual <http://manual.cytoscape.org/en/stable/index.html>)). The node diameter is directly proportional to the degree node and the node color gradient was dependent from the clustering coefficient, from red (higher) to green (lower)

biochemical machinery expressed by spermatozoa. For instance, it is known that the concentrations of Ca^{2+} in mammalian oviductal fluid immediately before ovulation are under hormonal control, and are involved in modulating the homeostasis and the hyperactivation of sperm motility [31], as well as Ca^{2+} is a key element in sperm capacitation and acrosome reaction [13, 16]. Indeed, spermatozoa are classified as excitable cells, such as neurons, i.e. as cells whose biological function is intimately calcium-dependent.

Moreover, zinc has been studied in a wide variety of experiments on different mammalian and non-mammalian models. In human andrology it has been found that seminal plasma zinc concentration is negatively associated with male infertility [32] and it has been suggested that its beneficial activity could be due to its antioxidant proprieties [33–35]. In addition this ion is known to interact with several enzymes such as acrosin, thus modulating their function.

pH: is spermatozoa intracellular pH is an important controller of cell function. As it has been recently

reviewed, it "is essential for spermatozoa function. Notably, several unique sperm ion transporters and enzymes whose elimination causes infertility are either pH_i dependent or somehow related to pH_i regulation" [36]. In particular, the molecules that are involved in pH_i control are CatSper, a Ca^{2+} channel; Slo, a K^+ channel; the sperm-specific Na^+/H^+ exchanger and the soluble adenylyl cyclase, and are active in fundamental sperm functions such as motility, capacitation, and acrosome reaction.

Temperature

The spermatozoa during their journey within the female genital tract experience site-specific differences in environmental temperature that originate a temperature gradient. To date, data concerning the temperature in human fallopian tube are not available; while in mammalian models (particularly in rabbit and sow) it has been found that the fertilization site in the oviduct is 1–2 °C warmer than the sperm reservoir [37]. In rabbit, it has been shown this temperature difference is time

dependent and that it increases from 0.8 ± 0.2 °C before ovulation to 1.6 ± 0.1 °C after ovulation [38].

The mechanisms involved in this phenomenon are not yet completely understood, but three different determinants have been proposed:

- 1) The hormone-dependent release of a macromolecule, the acid mucus glycoprotein, that undergoes extensive endothermic hydration [39];
- 2) The counter-current heat exchange of blood: the cold blood from the ovarian vein cools the blood that directed to the oviduct (as it happens in testicular pampiniform plexus) [40];
- 3) The change in the source of blood supply: the warmer ovarian artery before ovulation and the cooler uterine artery after the ovulation [41].

Interestingly, to date, the specific signalling pathways involved in spermatozoa thermotaxis are unknown, while it has been understood that this prove is capacitation-dependent: only capacitated spermatozoa are thermotactically responsive [37].

Conclusions

In conclusion, the network topology could explain important biological characteristics of the sperm cells biology. For instance, it confers to these cells a high robustness against random damages. In fact the relatively low number of nodes (in our model we estimate that the hub are 196 on 7758 nodes) that exert a strong control on network limits the probability that a random damage could have important consequences on the network itself (the probability is about 2.5%) [20]. Reasonably this specific behaviour could offer an important evolutionary advantage in terms of adaptation [42]. In addition, the existence of a non-democratic structure of control systems allows the cells to limit the energetic expenses: they could maintain the homeostasis controlling a small number (10^1 – 10^2) instead a myriad (10^3 – 10^4) of molecules, with an evident saving of time and energy [14]. Without considering that an increase in the number of molecules to be controlled will correspond to a non-linear increase in interactions, with a further relevant amount of energy consumption.

In addition, these networks are characterised by an ultra-small world topology and by a scarce clusterization. It means that the information will spread within the networks in a very low number of steps. In mean, all the messages starting from one node in HSIN3.0_MC will reach all the others 7757 nodes in 7.4 steps. This assures the maximal fastness and the minimal loss of messages integrity without loops or feed-backs, thus maximizing the performance of the system [14]. Finally, it allows to identify the nodes acting a higher degree of control on the network (hubs and bottlenecks) and to

study their topology and their biological proprieties. Interestingly, the data from KO mice model seem to support the biological relevance of our findings and suggest a more important role of bottlenecks than hubs in our network.

In our opinion, these results could have important implication for the study of sperm biology and physiopathology, even more in light of the high percentage of cases of unexplained infertility of male origin [43].

Methods

Data collection and networks construction and analysis

To realize HISN we used the data concerning the pathways expressed in human mature spermatozoa found in recent (< 5 years) peer-reviewed literature on PubMed database, referred to the sperm physiology. Once identified the pathways (See Table 4 for the list), in keeping with Bernabò et al. [17, 44], we downloaded them from Reactome (<http://www.reactome.org/>), which is a free, open-source, curated and peer-reviewed pathway database, whose aim is to offer bioinformatics tools for the visualization, interpretation and analysis of biochemical pathways involved in relevant biological events (see Table 4).

The obtained network was called Human Sperm Interactome 1.0. To aggregate the same molecules with different ID an ad-hoc Python script has been realized to harmonize molecule IDs and their aliases.

After this step, we obtained HSIN 2.0. To complete the network with all the possible interaction among the proteins that compose it, we used Search Tool for the Retrieval of Interacting Genes/Proteins, STRING (<http://string-db.org/>). It is a database of known and predicted protein interactions, which includes direct (physical) and indirect (functional) associations (Franceschini et al., 2013). The interaction provided by the STRING search were filtered based on experimentally determined interaction, database annotated, automated text mining, the species (*Homo sapiens*), and confidence score > 900. After this enrichment step, we obtained HSIN 3.0.

The different networks have been realized and analyzed using Cytoscape 3.4.0 (<http://www.cytoscape.org>) and its plugins Network Analyzer and CytoHubba. To assess the networks topology, we automatically measured the parameters listed in Table 5, considering the networks as directed.

Identification of HSIN 3.0 hubs

In keeping with Bernabò et al. [14, 15], we identified the hubs within HSIN 3.0, as the nodes with a degree at least one standard deviation above the network mean.

Identification of subpopulation in hubs, by kernel density estimation (KDE)

To identify eventual subpopulation of hubs based on their topological parameters, we assessed the probability density function of node degree, of clustering coefficient,

Table 4 pathways downloaded from Reactome, to realize HSIN3.0

Pathway	Pathway ID (Reactome)
Apoptosis	R-HSA-109581
Signalling to RAS	R-HSA-167044
Signaling by Wnt	R-HSA-195721
Signaling by EGFR	R-HSA-177929
Cell Cycle, Mitotic	R-HSA-69278
Regulation of mitotic cell cycle	R-HSA-453276
Cell Cycle Checkpoints	R-HSA-69620
Axon guidance	R-HSA-422475
Regulation of insulin secretion	R-HSA-422356
Signaling by Insulin receptor	R-HSA-74752
Protein folding	R-HSA-391251
Translation	R-HSA-72766
tRNA Aminoacylation	R-HSA-379724
L1CAM interactions	R-HSA-373760
Metabolism of amino acids and derivatives	R-HSA-71291
Triglyceride Biosynthesis	R-HSA-75109
Metabolism of nucleotides	R-HSA-15869
Gluconeogenesis	R-HSA-70263
Glycolysis	R-HSA-70171
Hexose transport	R-HSA-189200
Hemostasis	R-HSA-109582
Membrane Trafficking	R-HSA-199991
Nucleosome assembly	R-HSA-774815
Post-translational protein modification	R-HSA-597592
ABC transporters in lipid homeostasis	R-HSA-1369062
Signaling by Interleukins	R-HSA-449147
Cholesterol biosynthesis	R-HSA-191273
Metabolism of vitamins and cofactors	R-HSA-196854
Mitochondrial fatty acid beta-oxidation of unsaturated fatty acids	R-HSA-77288
Mitochondrial fatty acid beta-oxidation of saturated fatty acids	R-HSA-77286
Peroxisomal lipid metabolism	R-HSA-390918
Formation of ATP by chemiosmotic coupling	R-HSA-163210

and of betweenness centrality was estimated by kernel density estimation (KDE) [2]. It is a nonparametric analysis used to estimate the probability density function of a variable, thus allowing inferring if it constituted by subpopulations of data.

Identification of bottleneck nodes within HSIN 3.0

We carried out the identification of bottlenecks within the network by using the Cytoscape plugin CytoHubba. It implements the following algorithm for bottleneck

Table 5 Topological parameters assessed in this study

Parameter	Definition
Connected Components	It is the number of networks in which any two vertices are connected to each other by links, and which is connected to no additional vertices in the network.
Number of nodes	It is the total number of molecules involved.
Number of edges	It is the total number of interactions found.
Clustering coefficient	It is calculated as $C_l = 2n_l / (k_l(k_l - 1))$, where n_l is the number of links connecting the k_l neighbors of node l to each other. It is a measure of how the nodes tend to form clusters.
Network diameter	It is the longest of all the calculated shortest paths in a network.
Shortest paths	The length of the shortest path between two nodes n and m is $L(n,m)$. The shortest path length distribution gives the number of node pairs (n,m) with $L(n,m) = k$ for $k = 1, 2, \dots$
Characteristic path length	It is the expected distance between two connected nodes.
Averaged number of neighbors	It is the mean number of connections of each node.
Node degree	It is the number of interaction of each node.
Node degree distribution	It represents the probability that a selected nodes has k links.
γ	Exponent of node degree equation.
R^2	Coefficient of determination of node degree vs. number of nodes, on logarithmized data.

calculation: Let T_s be a shortest path tree rooted at node s . $BN(v) = \sum_{s \in V} ps(v)$ where $ps(v) = 1$ if more than $|V(T_s)| / 4$ paths from node s to other nodes in T_s meet at the vertex v ; otherwise $ps(v) = 0$ (Chin CH et al., 2014).

Identification of phenotypical effects of the deletion of gene relative to protein listed in HSIN3.0_MC in KO mice on male fertility

We studied the phenotypical effect on male fertility of removing the genes codifying for proteins present in HSIN 3.0 using Mammalian Phenotype Browser by Mouse Genome Informatics (MGI) (<http://www.informatics.jax.org/>), by selecting the MP: 0005389 (reproductive system phenotype).

Classification of HSNI 3.0 proteins

The classification of HSIN 3.0 proteins was carried out using UniProt (<http://www.uniprot.org/>). In particular, we have filtered the data for:

Function:

- Calcium binding: denotes the position(s) of calcium binding region(s) within the protein;
- Metal binding: binding site for a metal ion
- pH dependence: the optimum pH for protein activity

- Temperature dependence: indicates the optimal temperature for enzyme activity and/or the variation of enzyme activity with temperature variation; the thermostability of the enzyme is also mentioned in this subsection, when known.

Data analysis

All the statistical analysis have been carried out using Past3 (<https://folk.uio.no/ohammer/past/>).

Additional files

- Additional file 1:** List of hubs. (XLSX 18 kb)
Additional file 2: List of bottleneck nodes. (XLSX 16 kb)
Additional file 3: Party and date hubs. (XLSX 10 kb)
Additional file 4: Nodes controlled by calcium binding. (XLSX 12 kb)
Additional file 5: Nodes controlled by temperature. (XLSX 8 kb)
Additional file 6: Nodes controlled by metal binding. (XLSX 39 kb)
Additional file 7: Nodes controlled by pH. (XLSX 11 kb)

Acknowledgements

Not applicable.

Funding

This work has been funded by Project PRIN: "Environment, lifestyle and iatrogenic effects on sperm DNA genetic and epigenetic modifications: consequences for male reproductive fitness" (MM and NB).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

AO: performed collected the data, realized and analysed the networks, analysed the KO data; MO realized the ad-hoc Python script to harmonize molecule IDs and their aliases; NB conceived the study, wrote the manuscript, carried out the data analysis; BB, MM, MO were involved in critically revising the manuscript, with important intellectual contributions. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Faculty of Bioscience and Technology for Food, Agriculture and Environment, University of Teramo, Teramo, Italy. ²Istituto Zooprofilattico Sperimentale "G. Caporale", Teramo, Italy.

Received: 21 March 2017 Accepted: 27 March 2018

Published online: 11 April 2018

References

1. Austin C. Observations on the penetration of the sperm in the mammalian egg. *Aust J Sci Res B*. 1951;4:581–96.

2. Austin CR. The capacitation of the mammalian sperm. *Nature*. 1952;326.
3. Chang MC. Fertilizing capacity of spermatozoa deposited into the fallopian tubes. *Nature*. 1951;168:697–8. Available from: <http://www.nature.com.ezp-prod1.hul.harvard.edu/nature/journal/v168/n4277/abs/168697b0.html%5Cnhttp://www.nature.com/doi/10.1038/168697b0>.
4. Breitbart H, Finkelstein M. Regulation of sperm capacitation and the acrosome reaction by PIP 2 and actin modulation. *Asian J Androl*. 2015;17:597. Available from: <http://www.ajandrology.com/article.asp?issn=1008-682X;year=2015;volume=17;issue=4;spage=597;epage=600;aualast=Breitbart>.
5. Gadella BM, Luna C. Cell biology and functional dynamics of the mammalian sperm surface. *Theriogenology*. 2014;81:74–84. Available from: <https://doi.org/10.1016/j.theriogenology.2013.09.005>.
6. Itach SBS, Finklestein M, Etkovitz N, Breitbart H. Hyper-activated motility in sperm capacitation is mediated by phospholipase D-dependent actin polymerization. *Dev Biol*. 2012;362:154–61. Elsevier Inc.; Available from: <https://doi.org/10.1016/j.ydbio.2011.12.002>.
7. Amaral A, Castillo J, Ramalho-Santos J, Oliva R. The combined human sperm proteome: cellular pathways and implications for basic and clinical science. *Hum Reprod Update*. 2014;20:40–62.
8. Jodar M, Sendler E, Krawetz SA. The protein and transcript profiles of human semen. *Cell Tissue Res*. 2016;363:85–96.
9. Pandian Z, Gibreel A, Bhattacharya S. In vitro fertilisation for unexplained subfertility. *Cochrane Database Syst Rev*. 2015;11:CD003357.
10. Somigliana E, Paffoni A, Busnelli A, Filippi F, Pagliardini L, Viganò P, et al. Age-related infertility and unexplained infertility: an intricate clinical dilemma. *Hum Reprod*. 2016;31:1390–6.
11. Messinis IE, Messini CI, Daponte A, Garas A, Mahmood T. The current situation of infertility services provision in Europe. *Eur J Obstet Gynecol Reprod Biol*. 2016. Elsevier Ireland Ltd.; Available from: <http://www.sciencedirect.com/science/article/pii/S0301211516309526>.
12. Bernabò N, Di Agostino R, Ordinelli A, Mattioli M, Barboni B. The maturation of murine spermatozoa membranes within the epididymis, a computational biology perspective. *Syst Biol Reprod Med*. 2016;207:200–4.
13. Bernabò N, Mattioli M, Barboni B. The spermatozoa caught in the net: the biological networks to study the male gametes post-ejaculatory life. *BMC Syst Biol*. 2010;4:87. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2905340&tool=pmcentrez&rendertype=abstract>.
14. Bernabò N, Mattioli M, Barboni B. Signal transduction in the activation of spermatozoa compared to other signalling pathways: a biological networks study. *Int J Data Min Bioinform*. 2015;12(1):59–69.
15. Bernabò N, Greco L, Ordinelli A, Mattioli M, Barboni B. Capacitation-related lipid remodeling of mammalian spermatozoa membrane determines the final fate of male gametes: a computational biology study. *OMICS*. 2015;19:712–21. [cited 2015 Nov 18]. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26565605>.
16. Bernabò N, Berardinelli P, Mauro A, Russo V, Lucidi P, Mattioli M, et al. The role of actin in capacitation-related signaling: an in silico and in vitro study. *BMC Syst Biol*. 2011;5:47. BioMed Central Ltd; [cited 2014 Dec 4] Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3079638&tool=pmcentrez&rendertype=abstract>.
17. Bernabò N, Ordinelli A, Di Agostino R, Mattioli M, Barboni B. Network analyses of sperm-egg recognition and binding: ready to rethink fertility mechanisms? *OMICS*. 2014;18:740–53. [cited 2014 Dec 4]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25454512>.
18. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet*. 2004;5:101–13. [cited 2014 Jul 9] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14735121>.
19. Barabási A-L, Albert R. Emergence of scaling in random networks. *Science* (80-). 1999;286:509–12.
20. Albert R, Barabasi AL. Statistical mechanics of complex networks. *Rev Mod Phys*. 2002;74:47–97. Available from: <https://doi.org/10.1103/RevModPhys.74.47%5Cnpapers2//publication/doi/10.1103/RevModPhys.74.47>.
21. Ravasz E, Barabási A-L. Hierarchical organization in complex networks. *Phys Rev E Stat Nonlinear Soft Matter Phys*. 2003;67:26112.
22. Agarwal S, Deane CM, Porter MA, Jones NS. Revisiting date and party hubs: novel approaches to role assignment in protein interaction networks. *PLoS Comput Biol*. 2010;6:1–12.
23. Han J-DJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*. 2004;430:88–93.

24. Ferramosca A, Zara V. Bioenergetics of mammalian sperm capacitation. *Biomed Res Int.* 2014;2014.
25. Piomboni P, Focarelli R, Stendardi A, Ferramosca A, Zara V. The role of mitochondria in energy production for human sperm motility. *Int J Androl.* 2012;35:109–24.
26. Hibi H, Yokoi K, Yamamoto M. Effects of alpha-interferon on sperm production, concentration, and motility in the rat. *Int J Urol.* 1997;4:603–7. Available from: 9477192.
27. Satie A-P, Mazaud-Guittot S, Seif I, Mahé D, He Z, Jouve G, et al. Excess type I interferon signaling in the mouse seminiferous tubules leads to germ cell loss and sterility. *J Biol Chem.* 2011;286:23280–95. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3123094&tool=pmcentrez&rendertype=abstract>.
28. Wagner GP, Pavlicev M, Cheverud JM. The road to modularity. *Nat. Rev. Genet.* 2007;8:921–31. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18007649>.
29. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol.* 2007;3:713–20.
30. Schughart K, Libert C, Kas MJ. Controlling complexity: the clinical relevance of mouse complex genetics. *Eur J Hum Genet.* 2013;21:1191–6. Nature ePublishing Group. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3798854&tool=pmcentrez&rendertype=abstract%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/23632795>.
31. Coy P, García-Vázquez FA, Visconti PE, Avilés M. Roles of the oviduct in mammalian fertilization. *Reproduction.* 2012;144:649–60.
32. Taravati A, Tohidi F. Association between seminal plasma zinc level and asthenozoospermia: a meta-analysis study. *Andrologia.* 2016;48:646–53.
33. Ahmadi S, Bashiri R, Ghadiri-Anari A, Nadjarzadeh A. Antioxidant supplements and semen parameters: an evidence based review. *Int J Reprod Biomed (Yazd, Iran).* 2016;14:729–36. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28066832%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5203687>.
34. Ajina T, Sallem A, Haouas Z, Mehdi M. Total antioxidant status and lipid peroxidation with and without in vitro zinc supplementation in infertile men. *Andrologia.* 2017;49(7). <https://doi.org/10.1111/and.12703>. Epub 2016 Sep 29.
35. Gholirad S, Razi M, Hassani Bafrani H. Tracing of zinc and iron in experimentally induced varicocele: correlation with oxidative, nitrosative and carbonyl stress. *Andrologia* 2016;1–11.
36. Nishigaki T, José O, González-Cota AL, Romero F, Treviño CL, Darszon A. Intracellular pH in sperm physiology. *Biochem Biophys Res Commun.* 2014;450:1149–58. Elsevier Inc. Available from: <https://doi.org/10.1016/j.bbrc.2014.05.100>.
37. Eisenbach M, Giojalas LC. Sperm guidance in mammals - an unpaved road to the egg. *Nat Rev Mol Cell Biol.* 2006;7:276–85.
38. Bahat A, Eisenbach M, Tur-Kaspa I. Perioovulatory increase in temperature difference within the rabbit oviduct. *Hum Reprod.* 2005;20:2118–21.
39. Luck MR, Griffiths S, Gregson K, Watson E, Nutley M, Cooper A. Follicular fluid responds endothermally to aqueous dilution. *Hum Reprod.* 2001;16:2508–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11726567>.
40. Leese HJ. The formation and function of oviduct fluid. *J Reprod Fertil.* 1988;82:843–56.
41. Cicinelli E, Einer-Jensen N, Barba B, Luisi D, Alfonso R, Tartagni M. Blood to the cornual area of the uterus is mainly supplied from the ovarian artery in the follicular phase and from the uterine artery in the luteal phase. *Hum Reprod.* 2004;19:1003–8.
42. Bernabò N, Barboni B, Maccarrone M. The biological networks in studying cell signal transduction complexity: the examples of sperm capacitation and of endocannabinoid system. *Comput Struct Biotechnol J.* 2014;11:11–21. Elsevier B.V. Available from: <https://doi.org/10.1016/j.csbj.2014.09.002>.
43. Bernabò N, Mattioli M, Barboni B. A systems biology approach to understanding male infertility. 1994.
44. Bernabò N, Barboni B, Maccarrone M. Systems biology analysis of the endocannabinoid system reveals a scale-free network with distinct roles for anandamide and 2-arachidonoylglycerol. *OMICS.* 2013;17:646–54. [cited 2016 Jan 25]; Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3837438&tool=pmcentrez&rendertype=abstract>.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

