GENOMICS, PROTEOMICS AND *IN SILICO* ANALYSIS OF SPONTANEOUS ABORTIONS

A. Carvalho[1,2,*], J. Ferreira[1,2,*], R. Pinto Leite[3], M. Souto[3], P. Botelho[3], O. Moutinho[3], J. Capelo[4,5], H. Santos[4,5], L. Pinto[1,2], G. Igrejas[1,2,4,5,†]

[1] *Genetics and Biotechnology Department, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal*
[2] *Proteomics and Functional Genomics Unit, University of Trás -os-Montes and Alto Douro, Vila Real, Portugal*
[3] *Cytogenetics Laboratory, Genetics Service, Hospitalar Centre of Trás -os-Montes and Alto Douro, Vila Real, Portugal*
[4] *Bioscope group. Chemistry Department. Faculty of Sciences and Technology. New University of Lisbon. Campus de Caparica. Caparica. Portugal*
[5] *ProteoMass Scientific Society. Faculty of Sciences and Technology. Campus de Caparica. Portugal*
[*] *These authors contributed equally to this work*
[†] *Gilberto Igrejas, Genetics and Biotechnology Department, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal, Quinta dos Prados, 5000-801 Vila Real, Portugal; E-Mail: gigrejas@utad.pt; Tel: +351 259 350 530/Fax: +351 259 350 480.*

INTRODUCTION

It has been estimated that about 20% of all recognized clinical pregnancies end in a spontaneous abortion, mainly in the first trimester [1]. Risk factors associated with the occurrence of a sudden miscarriage have been established and are related with different etiology [2]. There also been reported possible causes, which can be genetics, structural, infectious, endocrine, immune or others that are still inexplicable and under discussion. Genetic factors are the ones that possess a greater prevalence, being highlighted the numeric chromosomal abnormalities, aberrant gene expression, mutations and the Single Nucleotide Polymorphisms (SNPs) [2, 3]. This is a problem that affect innumerous couples and sometimes without explanation for their occurrence, so it is crucial to add new techniques to the conventional ones, to increase the quality of prognosis and diagnosis. Genomics, bioinformatics and proteomics techniques can be used as powerful tools to provide an integrated molecular analysis of genotypic and phenotypic factors potentially related to sudden miscarriage [4]. One of the aims of this study was to detect specific SNPs in genes that encode mediators that play roles in the reproductive process and to predict if their existence could cause an impact on the occurrence of a spontaneous abortion. It was also an additional objective the identification of adverse proteins to the human embryonic development, such as proteins with post transcriptional modifications, in order to elucidate the possible causes related to the studied miscarriages.

MATERIAL AND METHODS

There were analysed 9 samples of spontaneous abortions with normal karyotype, with no relevant risk factors associated to the parents. In order to understand the possible causes of the occurrence of those miscarriages, this work consisted in to three phases: Genomics, *in silico* prediction and Proteomics (Figure 1).

- Genomics: There were selected 9 genes involved in the angiogenesis (VEGF, LEPR, PAI-1and TGFB1) and apoptosis (BID, Caspases 3, 8, 9 and 10) processes, since they are crucial to embryonic development, and it was performed the study of intronic, exonic and 3' untranslated region (3'UTR).
- *In silico* prediction: There were used bioinformatics tools to predict the impact of the detected modifications in *splicing* mechanism and in protein functions and structure.
- Proteomics: Identification of adverse proteins that could affect the correct embryonic development.

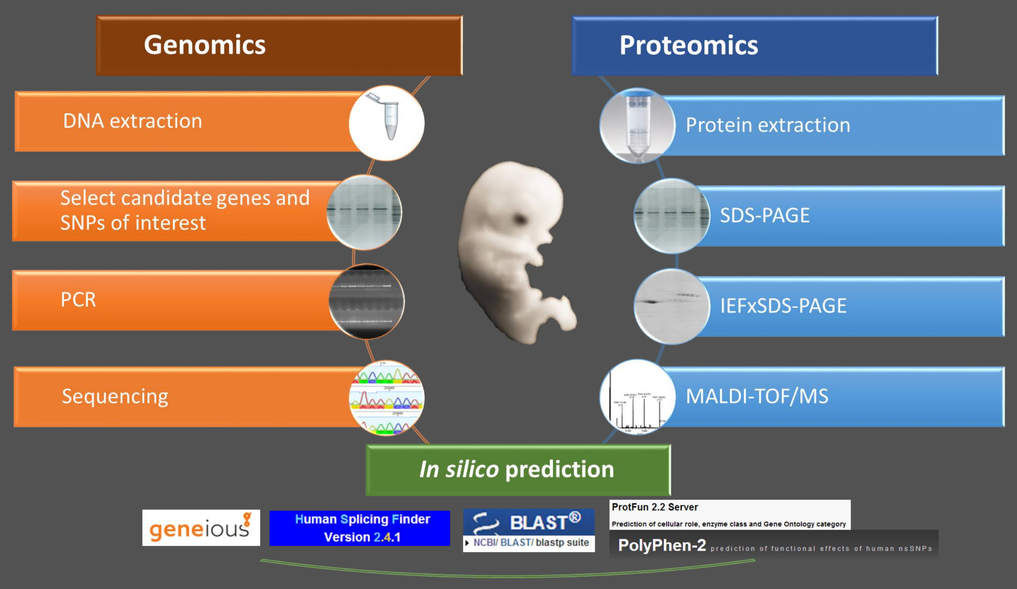


Figure 1. Workflow representing the experimental design conceived for genomics and proteomics approaches.

RESULTS

Figure 2 indicates the studied regions on the selected genes (intronic, exonic and 3'UTR), the number of analysed sequences, the detected modifications and their possible impact. Table 1 is a summary table of all detected modifications (nucleotidic and aminoacidic) on the studied gene regions.

The proteins were separated by two-dimensional electrophoresis (2DE) according their isoelectric point and molecular weight (IEF x SDS PAGE) and it was possible to identify 23 different proteins by *Peptide Mass Finger Print* technique (Figure 3). The *ratio* between those proteins and their principal functions can be visualized in Figure 4.

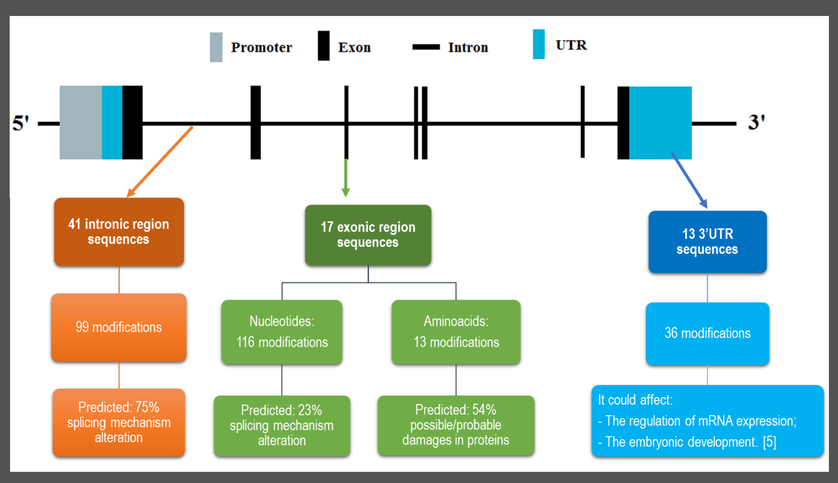


Figure 2. Schematic figure of a gene and its regions (promoter, exons, introns, 5'UTR and 3'UTR). It is also represented the studied regions, the number of detected modifications and their possible impact, in some biological mechanisms, protein functions and structure (predicted by *in silico* analysis and/or literature).

Table 1. Summary of all detected mutations of nucleotides and aminoacids (nucleotide conversion). Yellow represent the modifications of nucleotides on intronic region, orange represent the ones on the exonic regions and 3'UTR and light orange symbolize the alterations of the coding aminoacids. (N.F means non-functional proteins due to the many predicted modifications, as well codon-stop formation).

Number of detected modifications of nucleotides and coding aminoacids on the sequences of the studied genes																
	Apoptosis					Angiogenesis										
	Caspase 3	Caspase 9	Caspase 10		Total	PAI-1	VEGF-1,2,3,4	VEGF-5,6	VEGF-7,8	VEGF-11,12	VEGF-13	VEGF-15	LEPR		Total	Global total
Sample 1	3	1	1	0	5	0	1	2		2			2	N.F	7	12
Sample 2	2	2	11	5	15	0	4	1	4	1	1		1	0	12	27
Sample 3		1	4		5	1				1		3	1	N.F	6	11
Sample 4	1	2	24	2	27	4	4	2		2			0	0	12	39
Sample 5	1	0	9	4	10	4	4	4		4			1	0	17	27
Sample 6	2	1			3	6	4	5		4		2	0	0	21	24
Sample 7	2	2	20	2	24	5	2	2		6		1	0	0	16	40
Sample 8	0	1	21		22	5	2	1		6		1	0	0	15	37
Sample 9		2	20		22	0				7		4	1	0	12	34
Global total on nucleotides	11	12	110		133	25	21	17	4	33	1	11	6		118	251
Global total on aminoacids				13	13										13	13

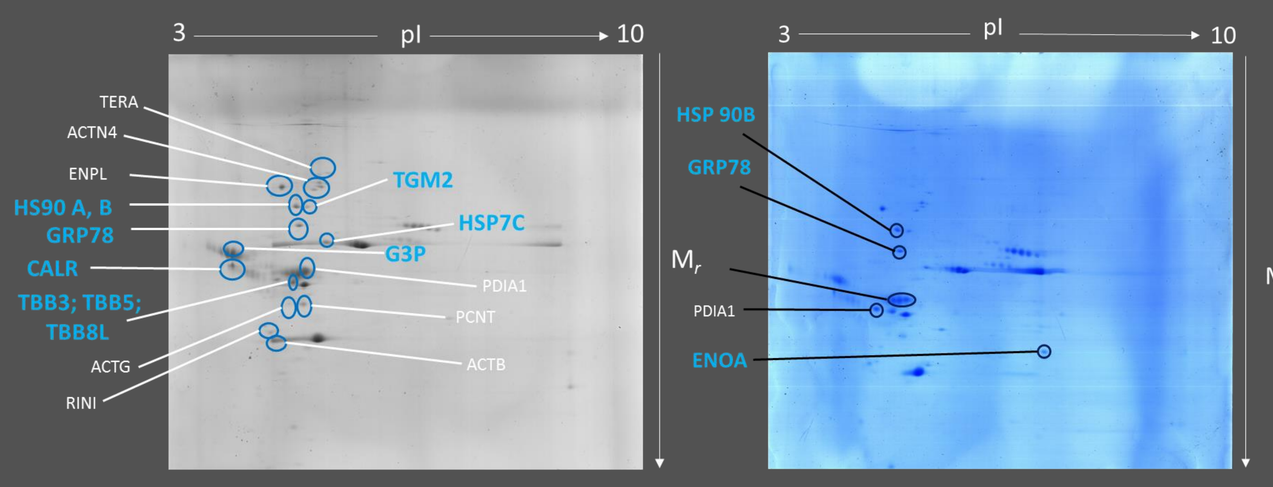


Figure 3. Two 2D-gels (Silver staining) in which are represented the identified proteins and their respective spots. The most relevant identified proteins to this study are written in blue.

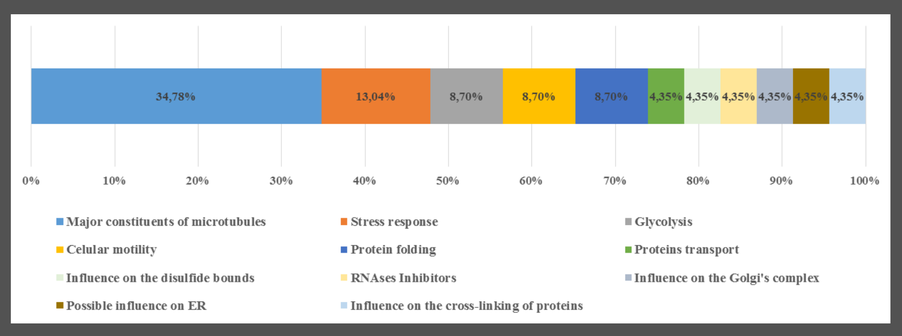
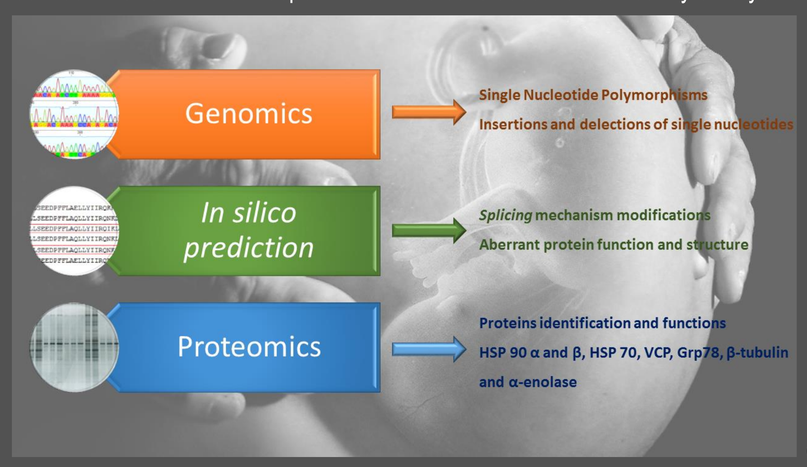


Figure 4. Representative diagram of the ratio between the number of identified proteins and their principal functions.

DISCUSSION AND CONCLUSIONS

The most prevalent mutations on the sequences of the studied gene regions, were SNPs, followed by deletions and insertions of a single nucleotide. Using the *Human Splicing Finder* [6] and *PolyPhen-2* [7] tools it was possible to predict the probable impact of those modifications in *splicing* mechanism and in the function and structure of some proteins. It was predicted the formation of new cryptic *splice* sites, as well ESEs/ESSs broken or created, which could silence the wild type *splice* sites or enhance cryptic sites, leading to intron retention, exon skipping and formation of truncated proteins. These findings along with some essential biological processes to development, such as angiogenesis and apoptosis, could highlight us to some potential causes of those miscarriages.

In the Matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF/MS) analysis, it was possible to identify 23 different proteins related to glycolysis, regulation of the cell cycle, transcription and angiogenic mechanisms, and stress responses. HSP 7C and 90 were also identified and we can hypothesize the presence of an undetected bacterial infection (Figure 5). Anti-HSPs produced in excess may have affected fetal growth as the balance between HSPs and their respective antibodies is essential for healthy embryonic development.



With the obtained results we emphasise again the power that this tools can give us as routine prog- and diagnosis techniques in order to answer to the problematic of spontaneous abortion.

Figure 5. Main conclusions of the three techniques carried out for the analysis of spontaneous abortions.

REFERENCES

[1] Lentz, G.M., et al., *Spontaneous and Recurrent Abortion: Etiology, Diagnosis, Treatment*, in *Comprehensive Gynecology*, E. Mosby, Editor. 2012. p. 335-359.

[2] Griebel, C.P., et al., *Management of spontaneous abortion*. Am Fam Physician, 2005. **72**(7): p. 1243-50.

[3] Rai, R. and L. Regan, *Recurrent miscarriage*. Lancet, 2006. **368**(9535): p. 601-11.

[4] Rull, K., L. Nagirnaja, and M. Laan, *Genetics of recurrent miscarriage: challenges, current knowledge, future directions.* Front Genet, 2012. **3**: p. 34.

[5] Kuersten, S. and E.B. Goodwin, *The power of the 3' UTR: translational control and development.* Nat Rev Genet, 2003. **4**(8): p. 626-37

[6] Desmet, F.O., et al., *Human Splicing Finder: an online bioinformatics tool to predict splicing signals*. Nucleic Acids Res, 2009. **37**(9): p. e67.

[7] Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations*. Nat Methods, 2010. **7**(4): p. 248-9.