



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

RONMILSON ALVES MARQUES

**METABONÔMICA APLICADA À AVALIAÇÃO DA QUALIDADE DO FLUIDO
SEMINAL: Identificação de Infertilidade Associada à Varicocele & Prognóstico
de Melhora dos Padrões Seminais Pós-Varicocelectomia e Impacto no
Sucesso Gestacional**

Recife

2022

RONMILSON ALVES MARQUES

**METABONÔMICA APLICADA À AVALIAÇÃO DA QUALIDADE DO FLUIDO
SEMINAL: Identificação de Infertilidade Associada à Varicocele & Prognóstico
de Melhora dos Padrões Seminais Pós-Varicocelectomia e Impacto no
Sucesso Gestacional**

Tese apresentada ao Programa de Pós-Graduação em Química da Universidade Federal de Pernambuco como requisito parcial para a obtenção do título de Doutor em Química. Área de concentração: Química Analítica

Orientador: Professor Dr. Ricardo Oliveira da Silva

Recife

2022

Catálogo na fonte
Bibliotecária Luiza de Oliveira, CRB4-1316

M357m Marques, Ronmilson Alves

Metabonômica aplicada à avaliação da qualidade do fluido seminal : identificação de infertilidade associada à varicocele & prognóstico de melhora nos padrões seminais pós-varicocelectomia e impacto no sucesso gestacional / Ronmilson Alves Marques. – 2022.
132 f.: il., tab.

Orientador: Ricardo Oliveira da Silva.

Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Química Analítica, Recife, 2022.

Inclui referências. Inclui apêndice.

1. Metabolômica. 2. Fertilidade. 3. Urologia. 4. Fluido Seminal. I. Silva, Ricardo Oliveira da (orientador). II. Título.

543

CDD (23. ed.)

UFPE - CCEN 2022-144

RONMILSON ALVES MARQUES

Metabonômica Aplicada à Avaliação da Qualidade do Fluido Seminal: Identificação de Infertilidade Associada à Varicocele & Prognóstico de Melhora dos Padrões Seminais Pós-Varicocelectomia e Impacto no Sucesso Gestacional

Tese apresentada ao Programa de Pós-Graduação no Departamento de Química Fundamental da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Química.

Aprovada em: 23/02/2022

BANCA EXAMINADORA

Prof. Ricardo Oliveira da Silva (Orientador)
Universidade Federal de Pernambuco

Prof. Thiago Mendonça de Aquino
Universidade Federal de Alagoas

Profa. Gisele André Baptista Canuto
Universidade Federal da Bahia

Prof. Ricardo Pimenta Bertolla
Universidade Federal de São Paulo

Profa. Betânia Lucena Domingues Hatzlhofer
Universidade Federal de Pernambuco

Dedico este trabalho ao Deus da minha vida, que desde quando criança me concede graça e permite que os meus sonhos sejam realizados.

AGRADECIMENTOS

Agradeço de todo coração ao Rei Eterno, Deus único, imortal e invisível, mas cristalino em minha vida. A Ele sejam a honra e a glória para todo o sempre. Amém!

À pessoa mais importante na minha vida, Silvia Rafaelli Marques. Já temos mais tempo juntos do que separados. Desde o início desse projeto me incentivou a não desistir, a manter a fé e esperança para a concretização deste sonho. Além disso, é peça fundamental em nossa linda família. Falando em família, mesmo sem entenderem bem o porquê do pai trabalhar tanto, agradeço muito aos meus filhotes Isaac e Esther, não apenas por existirem, mas também pela motivação que me transmitem diariamente nesta caminhada.

Sou bastante grato à minha mãe Maria Severina Alves. Nesta conquista, parcela significativa lhe pode ser atribuída por todo cuidado e orações à Deus por minha vida. Infelizmente, na vida real, nossos heróis morrem... agradeço ao meu pai Severino Damasceno Marques, embora não se encontre presente neste momento ímpar, porém seus ensinamentos, valores e ações estão empregados tanto na minha vida quanto na dos meus irmãos. Agradeço a todos os familiares e amigos que sempre torceram e acreditaram na concretização deste doutorado.

Agradeço a todos os professores e servidores públicos que fazem parte do Departamento de Química Fundamental, do Programa de Pós-Graduação em Química e da Central Analítica. Assim também, agradeço ao Hospital das Clínicas e ao Departamento de Cirurgia Geral, na figura do Prof. Dr. Salvador Villar e do Msc. Filipe Tenório pela colaboração, fornecimento das amostras de sêmen utilizadas neste trabalho e discussão dos achados da pesquisa. Da mesma forma, agradeço aos professores Dr. Licarion Pinto e Dr. Leslie Clifford Noronha Araújo pelos conhecimentos repassados.

Em especial, sou grato aos componentes do nosso grupo de pesquisa LabMeQ Marcio, Tatiane, Milena, Luiza, Fernanda e Letícia pelas trocas de conhecimentos, aprendizado e momentos de descontração vivenciados. Da mesma forma, sou grato ao Professor Dr. Ronaldo Dionísio pelos conhecimentos transferidos e orientações na execução deste projeto.

Por fim, quero agradecer, com enorme admiração, o Professor Dr. Ricardo Oliveira da Silva por todos os ensinamentos, sejam eles técnicos ou de vida, transmitidos a mim. Acredito que durante todo esse período de desenvolvimento acadêmico pude aprender com um dos melhores pesquisadores em metabonômica do nosso País. Vou mais além, os ensinamentos não verbalizados, mas observados e vivenciados no dia a dia, revelam o exemplar professor, servidor público e cidadão que devo me espelhar no exercício da pesquisa bem como da docência, se assim Deus me proporcionar carreira parecida. Muito obrigado Professor Ricardo Oliveira da Silva, o senhor influenciou bastante na minha formação acadêmica, profissional e de vida.

RESUMO

É estimado que aproximadamente 15% dos casais, em idade reprodutiva, tenham um diagnóstico de infertilidade. Os fatores masculinos respondem por 40% a 50% desses casos. A varicocele é uma condição que pode estar relacionada com os casos de infertilidade masculina, sendo prevalente em 12% dos indivíduos e acomete aproximadamente 25% dos homens inférteis. O presente trabalho teve por finalidade a construção de modelos metabonômicos baseados nos dados espectrais de RMN de ^1H obtidos de soro de sêmen para o diagnóstico da varicocele e da infertilidade associada a dilatação das veias que drenam o sangue testicular. Além disso, foram elaborados modelos metabonômicos com fins de prognósticos da melhoria da qualidade do sêmen após a varicocelectomia. Para fins didáticos, separamos o trabalho em quatro estudos: no primeiro, avaliamos a discriminação entre os pacientes com varicocele dos pacientes saudáveis; no segundo, buscamos discriminar infertilidade associada à varicocele; no terceiro estudo procuramos prever quais pacientes teriam melhora dos parâmetros seminais pós-varicocelectomia; e, no quarto estudo, foram elaborados modelos metabonômicos para identificar quais os voluntários com varicocele e infertilidade teriam aumentadas as chances de gestação após o reparo da mesma. A primeira e segunda parte do estudo foram realizadas com 80 voluntários, sendo 24 do grupo controle, 21 do grupo Varicocele Fértil e 35 do grupo Varicocele Infértil. Enquanto as duas últimas etapas, foram executadas com amostras de 43 voluntários que foram submetidos à varicocelectomia, visando aumentar as chances de gravidez. Foram obtidos espectros de RMN de ^1H e os dados espectrais foram tratados usando a plataforma online MetaboAnalyst e os softwares Matlab, Statistica e Excel. Os modelos foram criados usando os formalismos PLS-DA, OPLS-DA e LDA. Nos Estudos I e II, o formalismo GA-LDA foi o que apresentou melhores resultados. No primeiro, foram observados valores de exatidão, sensibilidade e especificidade iguais a 95,8%, 100% e 85,7%, respectivamente, para o diagnóstico de varicocele. No estudo II, foram encontrados valores de exatidão, sensibilidade e especificidade iguais a 94,1%, 100% e 83,3%, respectivamente, para o diagnóstico diferencial de infertilidade associada à varicocele. Nos Estudos III e IV, os melhores resultados foram obtidos com o formalismo LDA. No Estudo III, os valores de exatidão, sensibilidade e especificidade são iguais a 93,1%, 94,4% e 90,9%, respectivamente, para

prognóstico de melhora da qualidade do sêmen. E no Estudo IV, o modelo conseguiu discriminar, as amostras de voluntários que relataram gravidez em até 24 meses após varicocelectomia com a extensão de 95,6% na validação cruzada. Os achados deste trabalho de tese confirmam a hipótese admitida de que condições clínicas como infertilidade e varicocele, bem como a intervenção cirúrgica, alteram o perfil de metabólitos endógenos contidos no fluido seminal, permitindo a construção de modelos metabonômicos capazes de prever os casos em que a varicocelectomia pode melhorar as chances de gravidez para pacientes com infertilidade associada à varicocele.

Palavras-chave: metabolômica; fertilidade; urologia; fluido seminal.

ABSTRACT

Infertility is a condition present in about 15% of couple in reproductive age. Male factors are responsible by 40%-50% of infertility cases. In world, about 12% of men have been diagnosed with varicocele. This number increase to 25% when we observed only men with some level of infertility. In this thesis, we built ^1H NMR-based metabonomics models using seminal sera for varicocele and male infertility diagnosis, as well as for prognostic of seminal parameters improving after varicocelectomy. The study was structured in four steps: (I) varicocele diagnosis; (II) differential diagnosis of male infertility associated to varicocele; (III) prognostic of seminal quality improving after repair varicocele surgery; and (IV) prognostic of the surgery impact on the pregnancy probability. The I and II studies were performed using samples from 80 volunteers (Control group, 24; Fertile Varicocele group, 21; and Infertile Varicocele group 35), while III and IV studies were performed using 43 samples from infertile men who were submitted to varicocelectomy aiming upper the probability of have a pregnancy. ^1H NMR spectra from seminal sera were processed using online platform MetaboAnalyst and Matlab, Statistica and Excel software. Metabonomics models were built using PLS-DA, OPLS-DA and LDA formalisms. In the I and II studies, GA-LDA formalism presented the best results, with 95.8% accuracy, 100% sensibility and 85.7% specificity for varicocele diagnosis. For differential diagnosis of infertility associated to varicocele, GA-LDA metabonomics model was 94.1% accuracy, 100% sensibility and 83,3% specificity. In the III and IV studies, LDA formalism using Wilk's lambda as variable selection tool presented the best results with 93.1% accuracy, 94.4% sensibility and 90.9% specificity for seminal quality improving prognostic. Metabonomics model for assessment of the surgery impact on the pregnancy probability correctly classified 95.6% of the samples. The admitted hypothesis in this thesis was confirmed by findings here presented, since that it was possible built metabonomics models from ^1H NMR spectra of seminal plasma aiming male infertility assessment.

Keywords: metabolomics; fertility; urology; seminal serum.

LISTA DE FIGURAS

Figura 1 –	Regulação endócrina através do eixo Hipotálamo-Hipófise.	31
Figura 2 –	Anatomia do sistema venoso testicular.	35
Figura 3 –	Estratégias ômicas e suas composições.	38
Figura 4 –	Espectro de RMN de ¹ H do plasma seminal humano	43
Figura 5 –	Estrutura química dos principais metabólitos encontrados no sêmen.	44
Figura 6 –	Estrutura química da Gliceril Fosforil Colina, da Colina e da Creatina.	44
Figura 7 –	Exemplo de uma PCA. O gráfico à esquerda representa a distribuição dos escores em função das PC1 e PC2; à direita temos a representação do gráfico dos loadings.	51
Figura 8 –	Exemplo de gráfico de escore de uma LDA.	53
Figura 9 –	Fluxograma que representa o Algoritmo Genético.	58
Figura 10 –	Exemplo de PLS-DA. À esquerda, gráfico de escores; à direita, VIP escore.	59
Figura 11 –	Exemplo de gráfico de escores na OPLS-DA.	60
Figura 12 –	Distribuição dos grupos investigados em cada estudo desenvolvido.	69
Figura 13 –	Espectro de RMN de ¹ H (400 MHz, D ₂ O, CPMG-Presat) de uma amostra de soro de sêmen.	78
Figura 14 –	Em azul, a variância explicada por cada PC; e em verde, a variância acumulada.	79
Figura 15 –	Gráficos de escores da PCA realizada usando a matriz de covariância. Em verde, amostras do grupo controle; em vermelho, amostras dos pacientes diagnosticados com varicocele.	80
Figura 16 –	Modelagem PLS-DA Controle versus Varicocele. Na parte superior à direita, teste de permutação; à esquerda, figuras de mérito; inferior, gráfico de escores entre as duas primeiras variáveis latentes.	81
Figura 17 –	Resultado da OPLS-DA Controle versus Varicocele. À esquerda, gráfico de escores; à direita, teste de permutação.	82
Figura 18 –	Resultados GA-LDA Controle versus Varicocele. À esquerda, gráfico de escores; à direita, a distribuição das variáveis que melhor discriminaram o modelo.	84
Figura 19 –	Resultado da PLS-DA Controle versus Varicocele Fértil versus Varicocele Infértil. Superior, gráfico de escores; à esquerda, figuras de mérito; e à direita, teste de permutação.	86
Figura 20 –	Escores das Variáveis Importantes na Projeção (VIP scores).	87

Figura 21 –	Metabólitos com níveis séricos mais elevados no grupo Controle.	88
Figura 22 –	Metabólitos com níveis séricos mais elevados no grupo Varicocele Fértil.	89
Figura 23 –	Resultado da GA-LDA para discriminar varicocele infértil (azul), varicocele fértil (vermelho), grupos de controle (preto). Amostras de treinamento (bolas) e amostras de teste (pontos). Em (b) tem-se as regiões selecionadas, enquanto (c) e (d) apresentam o gráfico de pesos para cada função discriminante.	90
Figura 24 –	À esquerda, Gráfico de escores da OPLS-DA Varicocele Fértil versus Varicocele Infértil. Em vermelho e verde, estão as amostras dos grupos VI e VF, respectivamente. À direita, teste de permutação.	92
Figura 25 –	(a) Gráfico de escores da GA-LDA. Em azul e vermelho, estão as amostras dos grupos VI e VF, respectivamente. (b) As variáveis espectrais selecionadas. (c) Gráfico de pesos da GA-LDA.	93
Figura 26 –	Resultado da PCA Melhora versus Estabilidade dos parâmetros seminais pós-varicocelectomia. À esquerda, Gráfico de escores; e à direita, Screen plot com a variância explicada por cada componente principal, em azul, e acumulada, em verde.	98
Figura 27 –	Resultado da PLS-DA Melhora versus Estabilidade dos parâmetros seminais pós-varicocelectomia. À esquerda, Gráfico de escores; e à direita, VIP escore com as variáveis mais importantes para a discriminação.	98
Figura 28 –	Figuras de mérito da PLS-DA Melhora versus Estabilidade dos parâmetros seminais pós-varicocelectomia. À esquerda, valores de exatidão, R ² e Q ² ; e à direita, teste de permutação.	99
Figura 29 –	Figuras de mérito da OPLS-DA Melhora versus Estabilidade dos parâmetros seminais pós-varicocelectomia. À esquerda, Gráfico de escores; e à direita, VIP escores com as variáveis mais importantes.	100
Figura 30 –	Figuras de mérito da OPLS-DA Melhora versus Estabilidade dos parâmetros seminais pós-varicocelectomia. À esquerda, valores de R ² e Q ² ; e à direita, teste de permutação.	100
Figura 31 –	Gráfico de escores LDA Melhora (quadrados vermelhos) versus Estabilidade (quadrados azuis) dos parâmetros seminais.	101
Figura 32 –	Espectros de RMN de ¹ H (D ₂ O, 400 MHz, CPMG-Presat) de fluidos seminais coletados pré-varicocelectomia de voluntários que apresentaram melhora (A) e estabilidade	104

	(B) dos parâmetros seminais pós intervenção cirúrgica. Variáveis selecionadas na LDA.	
Figura 33 –	Gráfico de escores com os valores preditos pelo modelo LDA versus informação de gravidez.	107
Figura 34 –	Gráfico de escores das funções discriminantes. Os círculos vermelhos são amostras de voluntários que não conseguiram gravidez em até 24 meses da cirurgia; os triângulos azuis são amostras de voluntários que conseguiram engravidar a companheira; os quadrados dourados são 05 amostras para as quais não há informação sobre gravidez.	108
Figura 35 –	Gráfico de escores nas funções discriminantes para correlação com gravidez. Círculos Azuis = Gravidez; Quadrados Vermelhos = Não Gravidez; Triângulos Pretos = sem informação de gravidez.	110

LISTA DE TABELAS

Tabela 1 –	Principais fatores relacionados com a infertilidade masculina.	22
Tabela 2 –	Aspectos envolvidos na produção e maturação dos espermatozoides.	26
Tabela 3 –	Limites de referência mais baixos (5º percentis e seus intervalos de confiança de 95%) para as características do sêmen.	29
Tabela 4 –	Nomenclatura relacionada à qualidade do sêmen.	30
Tabela 5 –	Exemplo de Tabela de Contingência.	62
Tabela 6 –	Características clínicas e laboratoriais dos participantes dos estudos I e II.	76
Tabela 7 –	Matriz de Contingência - Modelo OPLS-DA. Grupo controle (24 amostras) versus grupo varicocele (56 amostras). Desempenho baseado após a Validação Cruzada.	82
Tabela 8 –	Matriz de Contingência - Modelo GA-LDA. Conjunto treinamento: grupo controle (17 amostras) versus grupo varicocele (39 amostras).	83
Tabela 9 –	Matriz de Contingência - Modelo GA-LDA. Conjunto Teste: grupo controle (7 amostras) versus grupo varicocele (17 amostras).	83
Tabela 10 –	Matriz de Contingência para o modelo usando GA-LDA – Controle versus Varicocele Fértil versus Varicocele Infértil. Conjunto de treinamento com 17, 15 e 24 voluntários, respectivamente.	90
Tabela 11 –	Matriz de Contingência para o modelo usando GA-LDA – Controle versus Varicocele Fértil versus Varicocele Infértil. Conjunto de teste com 7, 6 e 11 voluntários, respectivamente.	91
Tabela 12 –	Matriz de Contingência – Modelo OPLS-DA. Grupo Fértil (21 amostras) versus grupo Infértil (35 amostras). Desempenho baseado após LOOCV.	92
Tabela 13 –	Matriz de Contingência - Modelo GA-LDA. Conjunto treinamento: grupo Infértil (24 amostras) versus grupo Fértil (15 amostras).	93
Tabela 14 –	Matriz de Contingência - Modelo GA-LDA. Conjunto teste: grupo Infértil (11 amostras) versus grupo Fértil (6 amostras).	94
Tabela 15 –	Descrição dos parâmetros seminais pré e pós varicocelectomia dos voluntários participantes do estudo III.	96
Tabela 16 –	Matriz de Contingência para o modelo usando LDA – Melhora versus Estabilidade dos parâmetros seminais.	102

Tabela 17 –	Atribuição dos sinais responsáveis pela discriminação nas funções discriminantes para predição de melhora dos parâmetros seminais pós-varicocelectomia em até 24 meses pós-varicocelectomia.	102
Tabela 18 –	Parâmetros de qualidade seminal dos grupos investigados. Foram utilizados dados pós-varicocelectomia.	106
Tabela 19 –	Atribuição dos sinais responsáveis pela discriminação nas funções discriminantes de correlação com casos de gravidez, em até 24 meses pós-varicocelectomia.	107
Tabela 20 –	Distribuição das amostras com base na qualidade do sêmen e da condição de gravidez.	108
Tabela 21 –	Matriz de Contingência para o modelo usando LDA – Correlação com gravidez.	109
Tabela 22 –	Matriz de Contingência para o modelo usando LDA – SIM versus NÃO sucesso gestacional obtida a partir das funções discriminantes construída usando os escores nas funções discriminantes Melhora dos Parâmetros Seminais e Correlação com Gravidez.	110

LISTA DE ABREVIATURAS E SIGLAS

ANOVA	Análise de Variância
ATP	Trifosfato de adenosina
ATR-FTIR	<i>Attenuated Total Reflection Fourier-Transform Infrared</i>
CG-EM	Cromatografia em fase Gasosa acoplada à Espectrometria de Massa
CLAE-EM	Cromatografia Líquida de Alta Eficiência acoplada à Espectrometria de Massa
CPMG	Carr-Purcell-Meiboom-Gill (sequência de pulsos de radiofrequência)
DA	Análise Discriminante (do inglês, <i>Discriminants Analysis</i>)
DF	Função Discriminante (do inglês, <i>Discriminant Function</i>)
DNA	Ácido Desoxirribonucleico
EM	Espectrometria de Massa
EPZ	Espermatozoide
EROs	Espécies Reativas do Oxigênio
FSH	Hormônio Folículo Estimulante
GA	Algoritmo Genético (do inglês, <i>Genetic Algorithms</i>)
GABA	Ácido Gama-Aminobutírico
GnRH	Hormônio Liberador de Gonadotrofina
GPC	Glicerofosforilcolina
HMDB	<i>Human metabolome database</i>
IC	Intervalo de Confiança
LDA	Análise de Discriminantes Lineares (do inglês, <i>Linear Discriminants Analysis</i>)
LH	Hormônio Luteinizante
LOOCV	<i>Leave-One-Out Cross Validation</i>
MAD	<i>Median Absolute Deviation</i>
OAT	Oligoastenoteratoespermia
OMS	Organização Mundial da Saúde
OPLS-DA	<i>Orthogonal Partial Least Square Discriminants Analysis</i>
PCA	Análise de Componentes Principais
PCs	Componentes Principais
PLS-DA	<i>Partial Least Square Discriminants Analysis</i>
PPD	Pré-Processamento dos Dados
RMN	Ressonância Magnética Nuclear
S/R	Razão Sinal/Ruído

SNV	<i>Standard Normal Variables</i>
TCLE	Termo de Consentimento Livre e Esclarecido
TEARs	Tecnologias Experimentais de Alto Rendimento
USG	Ultrassonografia
USGD	Ultrassonografia com Doppler
VIP	<i>Variable Importance in the Projection</i>
VPN	Valor Preditivo Negativo
VPP	Valor Preditivo Positivo
ZP	Zona pelúcida

SUMÁRIO

1	INTRODUÇÃO	19
2	INFERTILIDADE MASCULINA	22
2.1	SÊMEN	23
2.2	ASPECTOS GERAIS DOS ESPERMATOZOIDES	25
2.3	DIAGNÓSTICO DA INFERTILIDADE MASCULINA	27
2.3.1	Análise do Sêmen	27
2.3.2	Investigação endócrina	31
2.3.3	USG de bolsa escrotal	33
3	VARICOCELE	34
4	ESTRATÉGIAS ÔMICAS	37
5	EMPREGO DAS TÉCNICAS ESPECTROMÉTRICAS NO ESTUDO DA INFERTILIDADE MASCULINA	42
6	ANÁLISES ESTATÍSTICAS MULTIVARIADAS	46
6.1	PRÉ-PROCESSAMENTO DOS DADOS (PPD)	46
6.2	ANÁLISE EXPLORATÓRIA DOS DADOS: ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)	48
6.3	ANÁLISE DISCRIMINANTE	52
6.3.1	Análise de Discriminante Linear (LDA)	53
6.3.2	PLS-DA / OPLS-DA	58
6.3.3	Figuras de Mérito e Avaliação dos Modelos Multivariados	60
7	HIPÓTESE	66
8	OBJETIVOS	67
8.1	OBJETIVO GERAL	67
8.2	OBJETIVOS ESPECIFICOS	67
9	MATERIAIS E MÉTODOS	68
9.1	SELEÇÃO DOS PACIENTES E GRUPOS DE ESTUDO E TRATAMENTOS ESTATÍSTICOS UNIVARIADOS	68
9.2	PREPARAÇÃO DAS AMOSTRAS DE SÊMEN, ANÁLISE POR RMN DE ¹ H E TRATAMENTO ESTATÍSTICO PARA ANÁLISES MULTIVARIADAS	72
10	RESULTADOS E DISCUSSÃO	75

10.1	DIAGNÓSTICO DE VARICOCELE & DIAGNÓSTICO DE INFERTILIDADE ASSOCIADA À VARICOCELE (ESTUDOS I E II)	75
10.1.1	Resultados demográficos clínico-laboratoriais	75
10.1.2	Diagnóstico da varicocele (estudo I)	78
10.1.3	Infertilidade associada à varicocele (estudo II)	86
10.1.4	Predição de melhora da qualidade dos parametros seminais após a varicocelectomia (estudo III)	94
10.1.5	Avaliação da taxa de gravidez após varicocelectomia	104
11	CONCLUSÃO	111
12	PRESPECTIVAS	114
	REFERENCIAS	115
APÊNDICE A	ESPECTROS DE RMN DE ALGUMAS AMOSTRAS	128
APÊNDICE B	ARTIGO PUBLICADO NO JOURNAL OF ASSISTED REPRODUCTION AND GENETICS	131
ANEXO A	PARECER CONSUBSTANCIADO DO COMITÊ DE ÉTICA	132

1 INTRODUÇÃO

A infertilidade é uma doença associada a disfunção do sistema reprodutor. É estimado que aproximadamente 15% dos casais, em idade reprodutiva, sejam inférteis.^{1,2} Esta doença é caracterizada quando o casal não atinge a concepção após um ano de atividade sexual regular sem o uso de métodos contraceptivos.² Esta é uma condição que, de modo geral, não é esperada pelos casais, pois a possibilidade de gerar filhos é um processo natural de manutenção da espécie humana. Depois que o casal é diagnosticado com infertilidade, os envolvidos passam a vivenciar situações desconfortáveis em seu relacionamento que podem afetar a qualidade de vida, o emocional e a atividade sexual.³

Quando o casal procura ajuda de um profissional especializado, ambos devem ser investigados. A doença pode ter sua etiologia por fatores estritamente masculinos ou femininos, contudo a infertilidade pode ser abordada como uma dificuldade do casal. Dos aspectos relacionados à problemática, o componente feminino, é responsável de 50% a 60% dos casos, enquanto os fatores masculinos correspondem por 40% a 50%. Aproximadamente 10% dos casos de infertilidade são por condições inexplicáveis.⁴

No que diz respeito ao sistema reprodutor masculino, a infertilidade idiopática é a disfunção mais comum, sendo identificada quando um ou mais parâmetros anormais estão presentes no sêmen com causa não identificável.⁵ Além do fator idiopático, alguns trabalhos indicam uma correlação entre a infertilidade masculina e a varicocele, visto que essa última tem uma frequência aproximadamente de duas a três vezes maior na população masculina infértil que entre os férteis. De fato, dados apontados na literatura indicam que esta doença é prevalente em 12% da população

masculina e acomete aproximadamente 25% dos homens inférteis. Apesar dos significativos esforços para o esclarecimento sobre a fisiopatologia, acredita-se que, em quase 50% dos casos de infertilidade, os mecanismos bioquímicos ainda não foram identificados. ^{1,6-8}

De modo geral, o diagnóstico das doenças é um dos primeiros passos para o correto tratamento. Desse modo, a medicina laboratorial é considerada, em diversos casos, como indispensável para a determinação do quadro clínico-patológico de um indivíduo. Neste contexto, o principal objetivo dos exames laboratoriais é disponibilizar ao médico, bem como ao paciente, informações precisas e confiáveis de modo a auxiliar na conduta terapêutica. Estima-se que aproximadamente 70% das decisões médicas são embasadas na medicina laboratorial. ^{9,10}

As análises clínicas são ferramentas eficazes para o diagnóstico, prognóstico e profilaxia de inúmeras doenças. Comumente, uma ou um conjunto de biomoléculas são empregadas com tais finalidades. Por exemplo, a creatinina é empregada para averiguar a função renal, enquanto o dano ao sistema hepatobiliar pode ser acompanhado pela determinação das concentrações das transaminases, bilirrubinas, fosfatase alcalina, lactato desidrogenase e o tempo de protrombina ativada. Para tais fins, as análises clínicas têm como material de trabalho as amostras biológicas, dentre as quais podemos destacar: urina, sangue, sêmen entre outros. ^{11,12}

As determinações para os diferentes biofluidos podem abranger o estudo das alterações no aspecto físico, como cor e turbidez da urina; a pesquisa de modificações nos padrões morfológicos celulares, como ocorrem no hemograma e no espermograma; e investigação de alterações nas concentrações dos componentes bioquímicos. ¹³ Considerando que os sistemas vivos, de modo geral, apresentam elevada complexidade e variabilidade molecular, e que os compostos endógenos

fazem parte do universo das reações metabólicas, a interpretação conjunta dos metabólitos pode fornecer informações singulares para mapeamento da homeostasia, assim como também das desordens físicas, químicas e biológicas.¹⁴

Os metabólitos são caracterizados como pequenas moléculas que são produzidas durante os distintos processos metabólicos, que podem oferecer indícios do estado celular momentâneo. Diferentemente das proteínas, que sofrem regulação genética e efeitos pós-tradução, os metabólitos delineiam a atividade bioquímica atual e direta do organismo. Dessa forma, sua correlação com as desordens celulares é mais fácil e pode ser empregada com a finalidade diagnóstica.¹⁵ O conjunto de metabólitos contidos em um determinado biofluido, tecido ou célula é denominado de metaboloma. Uma vez que a composição bioquímica dos fluidos corpóreos reflete o *status* metabólico do indivíduo, a compreensão minuciosa da composição e das concentrações dos mesmos, pode fornecer aos profissionais de saúde subsídios importantes quanto à conduta a ser adotada em relação ao tratamento e/ou acompanhamento das doenças.¹⁶⁻¹⁸

Entretanto, o surgimento das estratégias ômicas no final do século XX indicou uma mudança de paradigma nos exames laboratoriais. Novos procedimentos têm sido desenvolvidos, principalmente pelo domínio da estatística multivariada, que permitiram as interpretações dos dados biológicos de maneira sistemática. Assim sendo, novos campos de pesquisa em análises clínicas e biológicas foram desenvolvidos e aprimorados. Como exemplo, destacamos a metabonômica, que é a estratégia metodológica empregada na presente tese para o diagnóstico de varicocele e de infertilidade masculina associada à essa condição, bem como para fazer o prognóstico de melhoria dos parâmetros seminais pós-varicolectomia.

2 INFERTILIDADE MASCULINA

Estima-se que 7% dos homens em idade fértil sejam acometidos com problemas de fertilidade. A etiologia desta disfunção do sistema reprodutor masculino pode ser conferida tanto aos aspectos congênitos quanto aos adquiridos. Dessa forma, a infertilidade masculina pode ter origem pré-testiculares, pós-testiculares e danos testiculares. Contudo, há relatos de quadros de infertilidade em homens com causa não explicada.¹⁹⁻²¹ Na tabela 1, são apresentados os principais fatores relacionados com a infertilidade masculina.

Tabela 1 – Principais fatores relacionados com a infertilidade masculina.

Fatores congênitos	Fatores adquiridos
✓ <i>Anorquia testicular – Não formação dos testículos</i>	✓ Traumas testiculares
✓ <i>Criptoquidismo – Testículos que não desceram</i>	✓ Processos inflamatórios (orquites, epididimites)
✓ <i>Ausência de vasos deferentes - Condição anatômica</i>	✓ Obstrução parcial ou total do trato urogenital
✓ <i>Alterações genéticas – Síndrome de Klinefelter; Síndrome de Kallmann</i>	✓ Infecções urogenitais recorrentes (prostatites, uretrites; infecções urinárias)
	✓ Fatores exógenos (medicamentos, radioterapia, quimioterapia)
	✓ Doenças sistêmicas – Diabetes, cirrose hepática, doenças renais
	✓ Varicocele

Fonte: Autoria própria.

Entretanto, a infertilidade idiopática é o tipo mais comum da disfunção do sistema reprodutor masculino, sendo identificada quando um ou mais parâmetros anormais estão presentes no sêmen com causa não identificável.¹⁹ Além do fator idiopático, alguns trabalhos indicam uma forte correlação entre a infertilidade masculina e a varicocele, uma vez que essa última tem uma frequência duas a três vezes maior na população masculina infértil do que nos homens férteis. De fato, alguns estudos corroboram com esse dado, indicando que essa doença é prevalente em 12% da

população masculina e acomete aproximadamente 25% dos homens inférteis. Apesar dos significativos esforços para o esclarecimento sobre a fisiopatologia, acredita-se que, em quase 50% dos casos de infertilidade, os mecanismos bioquímicos ainda não foram identificados.¹⁹⁻²²

2.1 SÊMEN

A avaliação da quantidade, qualidade e composição do sêmen da espécie humana é motivo de inúmeros estudos por razões que decorrem de processos inflamatórios, como infecções na próstata, tumores e avaliação da fertilidade masculina. Segundo a Organização Mundial da Saúde (OMS), o sêmen é composto por um concentrado de espermatozoides, diluídos por secreções líquidas dos órgãos sexuais acessórios. Estima-se que 90% do líquido seminal seja formado por secreções dos órgãos acessórios, sobretudo das vesículas seminais, próstata e, em menor proporção, das glândulas bulbouretrais.²

Em relação ao material ejaculado, não se observa uma homogeneidade em toda sua extensão, o que indica que frações são incorporadas ao ejaculado por órgãos diferentes e em momentos distintos do processo. Sua primeira porção, que corresponde à aproximadamente 5% do total, é formada por uma secreção ligeiramente mucosa oriunda das glândulas bulbouretrais que tem por finalidade a lubrificação da uretra antes da ejaculação; a segunda parte, que corresponde a cerca de 15% a 30% do volume, é formada a partir das secreções prostáticas; por fim, o constituinte principal do ejaculado é formado a partir das vesículas seminais.^{2,23}

As glândulas sexuais acessórias fornecem componentes distintos ao líquido seminal. Por exemplo, a próstata é o principal fornecedor da fosfatase ácida, nitrato, inositol, cálcio, zinco e magnésio. Enquanto a secreção das vesículas seminais,

relativamente viscosa, é rica em carboidratos, com destaque para a frutose, substrato energético para os espermatozoides ejaculados. Além disso, encontram-se nesta secreção seminal, as prostaglandinas e o ácido ascórbico. Por último, o epidídimo produz e libera a carnitina livre, a enzima α -glicosidase neutra e a glicerofosforilcolina.^{2,23}

A determinação do *pH* seminal é um dos parâmetros analisados no exame de rotina deste fluido. A OMS, em seu manual para laboratórios mais recente, afirma que não existe um valor de referência de *pH* para homens férteis, entretanto é orientado considerar 7,2 para o limite inferior. O *pH* do sêmen tem características básicas devido às características alcalinas das secreções das vesículas seminais, bem como das secreções prostáticas.^{38,40} Uma interessante característica do sêmen é o seu elevado poder de tamponamento. Esta propriedade é comprovada, uma vez que, o valor do *pH* se mantém próximo da neutralidade mesmo no ambiente ácido da vagina, possibilitando aos espermatozoides a sua chegada à região do muco cervical, que tem *pH* neutro.²³⁻²⁴

Quanto à participação dos carboidratos, a frutose e a glicose são os principais representantes contidos no sêmen. Ambos têm como principal função o fornecimento energético para os espermatozoides. A frutose é produzida principalmente pelas vesículas seminais, inclusive a determinação da concentração da mesma no sêmen é um dos parâmetros para a avaliação da função das vesículas seminais, uma vez que a frutose é produzida neste lugar. Porém, as concentrações dos carboidratos podem alterar com o tempo devido ao processo de frutólise e, por conseguinte, a formação de ácido láctico, fato este que pode ser observado com a redução do *pH* do material.²³⁻

O plasma seminal é um fluido bastante complexo composto por diferentes componentes bioquímicos, como proteínas, lipídeos, enzimas, macro e microelementos. Dados recentes da literatura apresentam mais de 300 metabólitos contidos no plasma seminal.²⁷⁻²⁹ Contudo, o nitrato pode ser considerado como um dos mais importantes, pois tem significativa afinidade por cálcio, magnésio e zinco. Estima-se que o nitrato é o principal regulador dos níveis de cálcio no líquido seminal humano. É válido ressaltar que, os níveis de cálcio estão relacionados com a motilidade dos espermatozoides, metabolismo, reação acrossomal e, conseqüentemente, a fertilidade.²³

No sêmen, aminoácidos livres são encontrados, em abundância, os quais são produzidos a partir de biossíntese, nas células do trato genital masculino. Todavia, uma fração é produto da degradação das proteínas seminais, após a ejaculação mediante a ação de enzimas proteolíticas seminais. Ainda não está bem esclarecido a função dos aminoácidos livres no sêmen, porém estão associados às rotas bioenergéticas e processos de sobrevivência dos espermatozoides.^{28,29} Além disso, dados da literatura sugerem forte correlação entre diversos aminoácidos com concentração espermática.³⁰

2.2 ASPECTOS GERAIS DOS ESPERMATOZOIDEOS

Os espermatozoides (EPZs) são células haploides com elevado grau de diferenciação e altamente especializados. São produzidos nos testículos através da espermatogênese, e sua principal função está associada com a fertilidade, sendo o componente masculino no processo de fecundação. Em outras palavras, a função do EPZ é percorrer o trato reprodutor feminino, penetrar no oócito ocorrendo assim, a formação do zigoto. Do ponto de vista morfológico, o gameta masculino é dividido em

duas partes: cabeça e cauda (flagelo). A cabeça dos EPZs é composta pelo núcleo, acrossoma e membrana plasmática. O DNA espermático tem elevado grau de condensação ao ponto que não existir expressão gênica visto que o processo de transcrição não é possível. Por outro lado, a cauda é composta por um conjunto de microtúbulos e um axonema flagelar espermático que confere a força de motilidade aos EPZs. A cauda é ainda dividida em quatro porções: peça conectora, peça intermediária, peça principal e peça final.³¹

Todavia, os gametas masculinos não estão funcionalmente maduros após a espermatogênese. Para que ocorra a fecundação do oócito pelo espermatozoide, é necessário que o mesmo seja submetido a um conjunto de reações bioquímicas que envolvem o epidídimo e o trato genital feminino. Os aspectos envolvidos na produção e maturação dos EPZ são resumidos na Tabela 2.

Tabela 2 – Aspectos envolvidos na produção e maturação dos espermatozoides.

Processo	Localização	Principais aspectos adquiridos
Espermatogênese	Testículos	Biossíntese dos espermatozoides
Maturação	Epidídimo	Mobilidade progressiva
Capacitação	Trato Reprodutor Feminino	Hiperativação da motilidade Modificação da Membrana plasmática
Reação Acrossomal	Zona Pelúcida do Oócito	Reconhecimento e aderência do espermatozoide capacitado

Fonte: Autoria própria.

No epidídimo, a maturação espermática promove modificações morfológicas e bioquímicas que conferem aos espermatozoides motilidade progressiva e aumenta o potencial de sobrevivência no trato genital feminino. Os processos de capacitação promovem a alteração no padrão de mobilidade denominada de hiperativação, além disso, esta etapa remove proteínas/glicoproteínas do plasma seminal que foram adsorvidas na superfície dos EPZs contidos no ejaculado. A passagem pelo trato

reprodutor feminino também tem a função de modificar e reorganizar moléculas de superfície da membrana plasmática para favorecer o aumento da afinidade pela zona pelúcida (ZP). A reação acrossômica é caracterizada pelo reconhecimento e ligação do EPZ à ZP, com posterior liberação do conteúdo acrossômico. As enzimas hidrolíticas contidas no acrossoma auxilia na interação do EPZ com o oócito.³¹

2.3 DIAGNÓSTICO DA INFERTILIDADE MASCULINA

O diagnóstico da infertilidade masculina deve ser realizado a partir do histórico médico e reprodutor do paciente, visando identificar possíveis fatores de risco, bem como padrões comportamentais que podem influenciar na fertilidade.⁸

2.3.1 Análise do sêmen

A análise do sêmen, exame laboratorial conhecido como espermograma, é um dos primeiros exames empregados na avaliação da infertilidade masculina, sendo de fundamental importância na descrição da severidade do fator masculino. Porém, quando se trata do levantamento dos fatores masculinos, os parâmetros avaliados no sêmen são baseados nos dados de normalidade para este fluido biológico, dessa forma, outros fatores não podem ser desconsiderados, principalmente em pacientes com resultados normais obtidos a partir do espermograma.¹⁹⁻²⁰ Ressalta-se ainda que é possível haver a concepção da gestação mesmo quando o homem apresenta valores anormais no espermograma. Esta subfertilidade masculina pode ser compensada com bons parâmetros de fertilidade feminina.¹⁹

Na investigação seminal, são avaliados os aspectos físico-químicos do sêmen, através da análise macroscópica mediante a determinação dos seguintes parâmetros: volume, viscosidade, *pH*, cor e aspecto. Em seguida, é avaliada a qualidade do

material mediante a determinação da concentração de espermatozoides no ejaculado bem como às alterações morfológicas, motilidade e, quando necessário, a vitalidade.

Esse exame requer profissionais especializados para sua execução, além disso, fatores pré-analíticos, analíticos e pós-analíticos podem interferir na confiabilidade dos resultados. Sendo assim, tais fatores devem ser considerados ao ser interpretado o referido exame.^{1,2} Outra característica que deve ser considerada ao analisar e interpretar o líquido seminal, é a variação intraindividual da composição do sêmen. Assim, é recomendado não caracterizar a qualidade do sêmen do paciente a partir da avaliação de uma única amostra do paciente. O manual da OMS (2010) orienta a realização do espermograma com duas ou três amostras coletadas entre um intervalo de 2 a 3 meses. Este espaço temporal é necessário para garantir que ocorra a espermatogênese, além de possibilitar o correto acompanhamento dos tratamentos.²

O espermograma, infelizmente, não é uma técnica perfeita, visto que existem algumas variáveis incontrolláveis que podem influenciar na execução do exame, a saber: amostragem adequada, o local da coleta do material, o tempo desde a última atividade sexual e o tamanho do testículo. Sendo assim, não pode ser considerada padrão ouro para o diagnóstico de infertilidade no sistema reprodutor masculino. No entanto, é empregada rotineiramente, sendo a base para a pesquisa da infertilidade nos homens, pois auxilia na definição da gravidade do fator masculino.¹⁹

É válido ressaltar que os intervalos de referência estabelecidos pela OMS (2010), apresentados na tabela 3, tem o propósito de auxiliar a classificação dos homens como férteis ou subférteis. Além disso, tais valores foram determinados como referência e não como padronização de normalidade, baseados em um estudo

populacional com 1953 homens que tiveram filhos nos últimos 12 meses, à época do estudo.³²

Tabela 3 – Limites de referência mais baixos (5º percentis e seus intervalos de confiança de 95%) para as características do sêmen.

Parâmetros	Limite inferior de referência – IC 95%
Volume do sêmen (mL)	1,5 (1,4 – 1,7)
Concentração de espermatozoides (10 ⁶ U/mL)	15 (12 – 16)
Número total de espermatozoides (10 ⁶ U/ejaculado)	39 (33 – 46)
Motilidade total (%)	40 (38 – 42)
Motilidade progressiva (%)	32 (31 – 34)
Vitalidade (%)	58 (55 – 63)
Morfologia espermática (%)	4 (3 – 4)
<i>pH</i>	≥ 7,2
Concentração de leucócitos (10 ⁶ U/mL)	≤ 1,0

Fonte: Manual de processamento e investigação do sêmen humano da OMS (2010).²

Portanto, pode-se inferir que os valores atribuídos a partir do estudo de Cooper et al. (2010), os quais foram adotados pela OMS, são indicadores de fertilidade masculina. Como os limites foram determinados a partir do quinto percentil, logo é possível afirmar que os homens que apresentam dados seminais abaixo do quinto percentil (limite inferior) não são necessariamente inférteis.^{2,32} Na tabela 4 podemos observar as principais nomenclaturas que estão relacionadas a qualidade do sêmen.

Tabela 4 – Nomenclatura relacionada à qualidade do sêmen.

Nomenclatura	Definição
Aspermia	Ausência de sêmen (sem ou ejaculação retrógrada)
Astenozoospermia	Porcentagem de espermatozoides com motilidade progressiva abaixo do limite inferior de referência
Astenoteratozoospermia	Porcentagens de espermatozoides progressivamente móveis e morfológicamente normais abaixo dos limites de referência mais baixos
Azoospermia	Sem espermatozoides no ejaculado
Criptoospermia	Espermatozoides ausentes nas preparações frescas, mas observados no sedimento centrifugado
Hemospermia (hematospermia)	Presença de hemácias no ejaculado
Leucospermia (leucocitospermia, piospermia)	Presença de leucócitos acima do valor limite no ejaculado
Necrozoospermia	Baixa porcentagem de espermatozoides vivos e alta porcentagem de espermatozoides imóveis no ejaculado
Normozoospermia	Número total de espermatozoides e percentual de espermatozoides progressivamente móveis e morfológicamente normais, igual ou acima dos limites de referência mais baixos
Oligoastenozoospermia	Número total de espermatozoides e percentual de espermatozoides progressivamente móveis, abaixo dos limites de referência mais baixos
Oligoastenoteratozoospermia	Número total de espermatozoides e percentual de ambos os espermatozoides progressivamente móveis e morfológicamente normais, abaixo dos limites de referência mais baixos
Oligoteratozoospermia	Número total de espermatozoides e percentual de espermatozoides morfológicamente normais, abaixo dos limites de referência mais baixos
Oligozoospermia	Número total de espermatozoides abaixo dos limites de referência mais baixos
Teratozoospermia	Porcentagens de espermatozoides morfológicamente normais abaixo dos limites de referência mais baixos

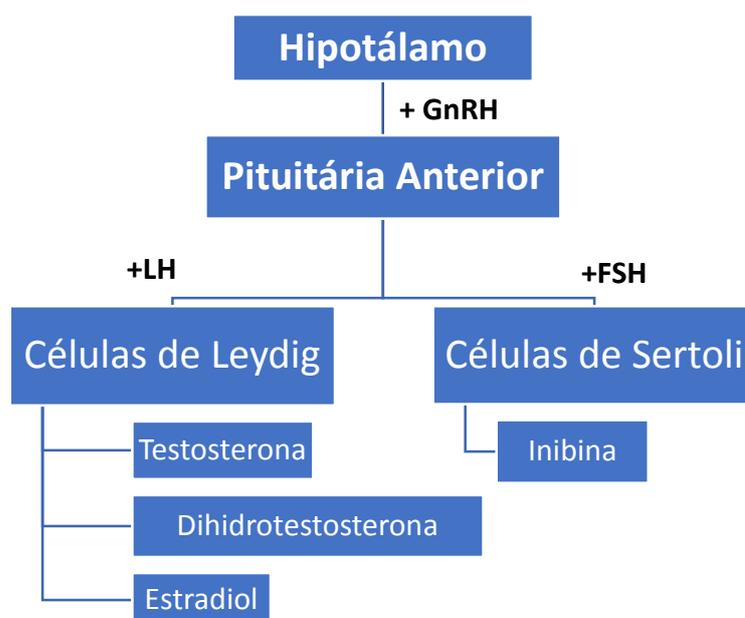
Fonte: Fonte: Manual de processamento e investigação do sêmen humano da OMS (2010).²

2.3.2 Investigação endócrina

A fertilidade masculina pode ser impactada negativamente por fatores que envolve a qualidade de vida (tabagismo, etilismo, alimentação e prática de atividades física). Entretanto, alguns quadros de infertilidade seriam em decorrência das desordens observadas do sêmen. Tais irregularidades nos padrões seminais podem estar correlacionados com alterações hormonais. As disfunções hormonais masculinas têm maior prevalência em homens inférteis quando comparado à população geral, contudo são fenômenos raros. A incidência de desordens endócrinas primárias como causa da infertilidade masculina varia de 1-3% dos casos.³³⁻³⁵

O eixo hipotalâmico-hipofisário-gonadal masculino tem grande relevância no que se refere a regulação endócrina da espermatogênese e assim como também sobre a função testicular. Em outras palavras, a fertilidade em homens tem uma significativa relação de dependência com a correta atividade do referido eixo.

Figura 1 – Regulação endócrina através do eixo Hipotálamo-Hipófise



Fonte: Mcgriff, Sarah C. et al. (2020).³⁴

A secreção de hormônios testiculares é controlada pela ação de duas gonadotrofinas hipofisárias: hormônio luteinizante (LH), que tem como função regular a atividade das células de Leydig e, conseqüentemente, a síntese dos hormônios sexuais, em nível testicular; e o Hormônio Folículo-Estimulante (FSH). Este último, em conjunto com a testosterona, atua nos túbulos seminíferos, através das células de Sertoli, induzindo e mantendo a espermatogênese. O controle da liberação do LH e do FSH é responsabilidade do hormônio liberador de gonadotrofina (GnRH), que é sintetizado no hipotálamo. A regulação hormonal da espermatogênese também é realizada pela ação androgênica da testosterona, enquanto o FSH estimula a divisão, diferenciação, e inibição da apoptose dos espermagônios. Além de estimular as meioses, a testosterona promove o alongamento e transformação das espermatídes e, por fim, favorece sua adesão às células de Sertoli.³⁴

De modo geral, quando é observado padrões seminais alterados, a investigação hormonal se baseia na determinação dos níveis séricos de FSH, LH e testosterona. Nos casos de azoospermia ou OAT (oligoastenoteratoespermia) acentuada, é importante distinguir as causas obstrutivas daquelas não obstrutivas. Para os cenários de não obstrução, os valores preditivos positivos aceitáveis podem ser considerados a partir de valores normais de FSH associado ao volume testicular bilateral normal.³³⁻³⁶

O quadro de hipogonadismo hipergonadotrófico causa a insuficiência testicular primária, que pode estar associada à níveis elevados de gonadotrofinas (FSH e LH) e com a diminuição dos níveis de testosterona, estando relacionada a causas congênitas (Síndrome de Klinefelter, anorquia ou criptorquidia, por exemplo) ou adquiridas (pós-orquite, torção testicular ou tumor testicular, por exemplo). No outro extremo, podemos encontrar o hipogonadismo hipogonadotrófico, que provoca a

insuficiência testicular secundária. Esse quadro é caracterizado pelos baixos níveis de testosterona associados à deficiência dos hormônios FSH e LH. Contudo, é um evento raro que pode ocorrer a partir de alterações congênitas (Síndrome de Kallmann ou Síndrome de Prader-Willi) ou anomalias adquiridas (tumores no eixo Hipotálamo-Hipófise).³³

As recomendações mais recentes para a avaliação endócrina em homens inférteis ocorrem quando é observada concentração espermática menor que 10×10^6 espermatozoides/mL, presença de disfunção erétil, e indicativos de níveis de testosterona baixos. A investigação inicial deve incluir a determinação dos níveis séricos da testosterona e FSH. Nos quadros que apresentem níveis baixos de testosterona, é interessante avaliar a testosterona total e livre, além dos níveis séricos de LH, globulina de ligação de hormônios sexuais e da prolactina.^{33,36}

2.3.3 USG de bolsa escrotal

A ultrassonografia de bolsa escrotal com Doppler (USGD) é um instrumento importante para determinar defeitos intraescrotais. Aproximadamente 30% dos homens subférteis são diagnosticados com varicocele com o auxílio do USGD da região escrotal. Além disso, neoplasias testiculares podem ser descobertas em 0,5% dos casos. Em pacientes com histórico de subfertilidade, a USG é indicada haja vista que os quadros de microcalcificações testiculares (quadro indicativo de lesão pré-maligna) são detectados em cerca de 2-5% desse público.³³

É recomendado que a USGD seja realizada com o paciente em pé, deitado e durante repouso e manobra de Valsalva. Este exame consegue avaliar a presença de refluxo venoso, além de avaliar com precisão o diâmetro das veias do plexo pampiniforme.^{33,37}

3 VARICOCELE

Clinicamente, a varicocele é definida como uma enfermidade decorrente da dilatação do plexo venoso pampiniforme que drena os testículos. A varicocele é reconhecida como a causa passível de correção mais comum de infertilidade em homens. Além disso, estudos indicam que essa doença esteja relacionada a efeitos negativos sobre a qualidade seminal, função espermática, histologia testicular e sobre os hormônios reprodutores.³⁸⁻⁴²

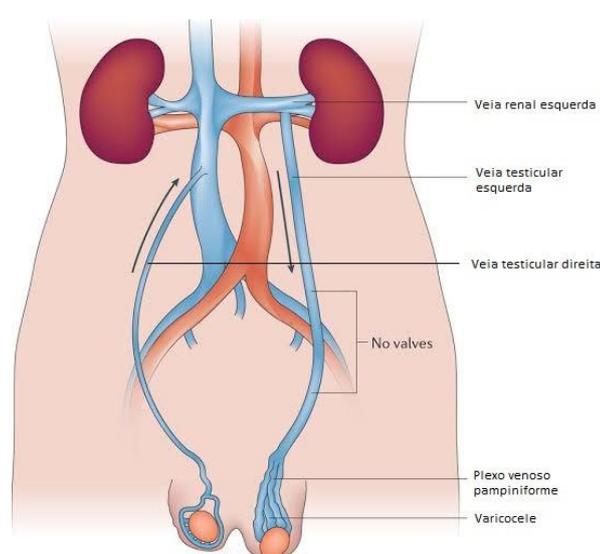
As primeiras evidências da associação da varicocele com a infertilidade masculina foram relatadas por Tulloch et al, em 1952. Os autores mencionam que o reparo da varicocele, em homens como azoospermia, culminou com a concepção natural.⁴³ Do ponto de vista epistemológico, a incidência de varicocele em homens com infertilidade primária varia de 35-44%; enquanto nos casos de infertilidade secundária varia de 45-81%.⁴⁴⁻⁴⁵ A incidência da doença tem uma grande variação quando é considerada a faixa etária do paciente. Estima-se que em meninos e pré-adolescente, com idades variando de 2-10 anos, a incidência é inferior a 1%, contudo, entre os adolescentes com idade de 15-19 anos, esse número é de aproximadamente 14%. Considerando a população total de homens adultos, a varicocele tem uma incidência de 15%.⁴⁶ Embora alguns achados epidemiológicos ofereçam suporte para a associação entre a varicocele e a infertilidade masculina, nem todos os homens com varicocele são inférteis. Além disso, é possível encontrar homens que tem varicocele com diferentes graus, apresentando parâmetros seminais normais.⁴⁰ A instalação da doença ocorre em função do retorno venoso sobre a veia espermática interna, culminando com a dilatação do plexo pampiniforme. Este refluxo do sangue pode ser devido a fatores congênitos ou ainda pela ausência de válvulas que impede o retorno sanguíneo. Acomete com maior frequência o testículo esquerdo devido a pressão

hidrostática ser maior sobre o mesmo, fato este que é explicado pela disposição da veia espermática interna esquerda que se encontra inserida praticamente perpendicular a veia renal esquerda (Figura 2).^{40,47}

Estima-se que em 90% dos homens com dilatação dos vasos testiculares e diagnosticados como sendo varicocele unilateral, isso ocorra no testículo esquerdo.⁴⁷ Gat et al (2004) investigaram 286 homens inférteis, usando exame físico, USGD e venografia, constatando que 89,2% (255 pacientes) apresentavam varicocele. Destes, 206 apresentavam bilateralidade, enquanto 49 apresentavam varicocele unilateral. Entre os que apresentaram varicocele unilateral, 91,8% (45 pacientes) apresentaram varicocele no lado esquerdo.⁴⁸

O mecanismo que desencadeia a doença e como ela afeta a espermatogênese ainda não estão totalmente esclarecidos. Porém, existem algumas hipóteses para determinar sua causa e seus efeitos sobre a espermatogênese, bem como sobre a fertilidade. Acredita-se que não existe um único fator responsável pelos efeitos negativos sobre os testículos. O que de fato existem são as evidências de que a doença tem aspecto multifatorial com possíveis atuações conjuntas de diferentes mecanismos desencadeadores da varicocele. Dentre as teorias mais conhecidas atribuídas a gênese da doença, destacam-se a hipertermia, hipóxia, diminuição do fluxo sanguíneo intratesticular e epididimário, alterações hormonais intratesticulares, estresse oxidativo e refluxo de metabólitos renal e adrenal.³⁹⁻⁴⁰

Figura 2 – Anatomia do sistema venoso.



Fonte: Jensen et al (2017)⁴⁰.

Neste contexto de complexidade e com características multifatoriais, o estresse oxidativo se apresenta como uma rota central de convergência entre as diferentes teorias propostas para a varicocele. O estresse oxidativo é decorrente do aumento da concentração das espécies reativas do oxigênio (EROs) e da diminuição da capacidade antioxidativa em nível seminal. As EROs são classificadas como subprodutos das vias metabólicas e sua produção em excesso está associada à redução da motilidade dos espermatozoides; alterações morfológicas; diminuição da produção de trifosfato de adenosina (ATP).^{39,40} Esses três fatores podem impactar o processo de fecundação, visto que a motilidade é uma característica importante para a migração dos espermatozoides até o óvulo; as modificações morfológicas inviabilizam o correto funcionamento do espermatozoide; e por fim, com a produção inadequada de ATP, a célula não terá energia suficiente para produção de metabólitos além do comprometimento da motilidade.^{39,40} O desequilíbrio entre os EROs e a capacidade antioxidante favorece também a oxidação dos ácidos graxos das membranas dos espermatozoides promovendo alterações em sua morfologia, motilidade e, por fim, o comprometimento da capacidade fertilizantes dos mesmos.⁵¹

Alguns trabalhos relatam elevados níveis de estresse oxidativo no sêmen dos pacientes com varicocele e que após a cirurgia de varicocelectomia é observado a diminuição dos biomarcadores do estresse oxidativo, acompanhados da melhora dos parâmetros seminais.⁴⁹⁻⁵⁰ Entretanto, níveis anormalmente altos de EROs podem indicar aumento na fragmentação do DNA o que, em última análise, poderia induzir uma diminuição da fertilidade. Dessa forma, esse mecanismo seria uma possível ligação da infertilidade com a varicocele.⁴⁰

4 ESTRATÉGIAS ÔMICAS

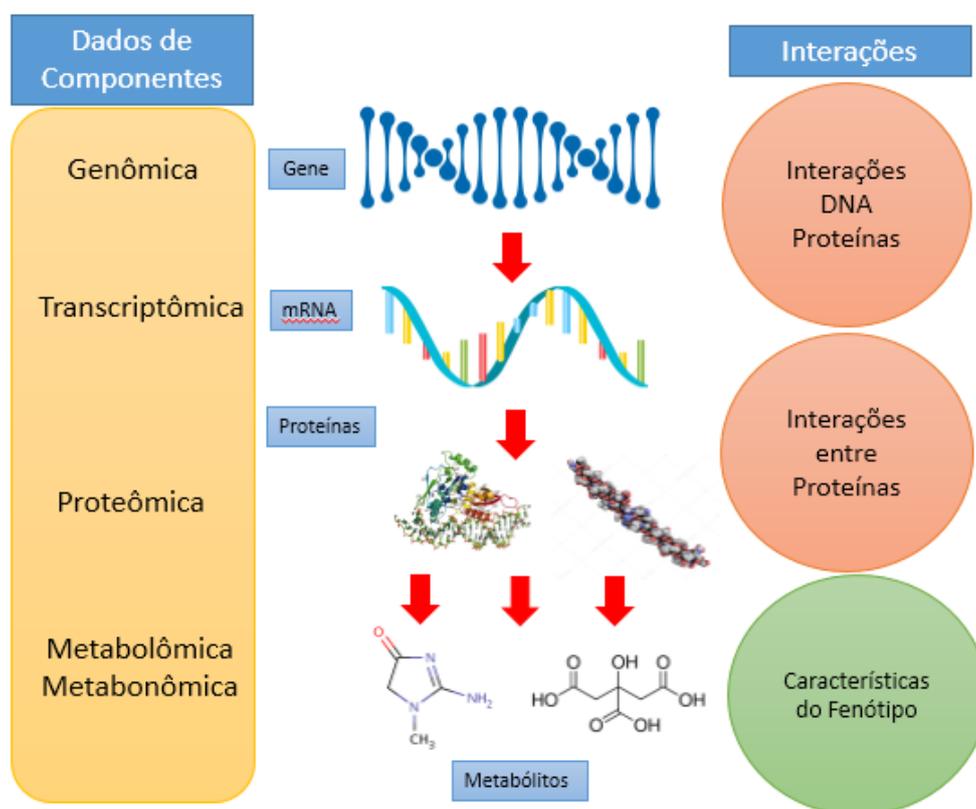
A década de 90 foi um divisor de águas na pesquisa científica tanto para as ciências biológicas quanto para a saúde, visto que em 1995, o genoma do *Haemophilus Influenza* foi totalmente sequenciado. Alguns anos depois, em 2003, o Projeto Genoma Humano concluiu o mapeamento genético da nossa espécie, fato que possibilitou avanços expressivos na compreensão da diversidade e complexidade molecular envolvida na homeostasia.⁵²⁻⁵⁴

A partir do enfoque genético das doenças, as engrenagens dos sistemas biológicos passaram a ser, até certo ponto, reveladas e novas oportunidades de diagnóstico, tratamento e prevenção estão sendo discutidas. Sendo assim, condições e táticas foram desenvolvidas com a finalidade de compreender e quantificar o fenótipo molecular a partir dos produtos da cadeia gênica, a saber: mRNAs, proteínas e metabólitos.⁵⁵

O emprego das tecnologias experimentais de alto rendimento (TEAR) é outro marco importante para as análises dos sistemas vivos. Tais instrumentos transformaram significativamente a pesquisa biológica, visto que ricas informações podem ser extraídas e devidamente interpretadas. Como exemplos de TEARs podemos citar a espectroscopia de Ressonância Magnética Nuclear (RMN) e a Cromatografia, em fase gasosa ou fase líquida, acoplada à Espectrometria de Massa (EM). Dessa forma, inúmeros alvos e marcadores biológicos passaram a ser relatados e, de acordo com o aspecto molecular, foram criadas linhas de pesquisa distintas que, em conjunto, foram denominadas de estratégias Ômicas. Com o exame do conjunto de dados ômicos, informações mais abrangentes foram expostas tornando visíveis o estado celular para um determinado tempo e/ou condição.^{14,55-57}

De modo geral, os dados ômicos podem ser dispostos em três categorias: dados de componentes, integração e funcional (Figura 3). O conteúdo molecular característico da célula é extraído a partir dos dados de componente, enquanto os dados relativos às interações delineiam as conexões entre os componentes moleculares. Por último, os dados de função descrevem o desempenho geral do sistema, fato este expresso no fenótipo celular.⁵⁸

Figura 3 – Estratégias ômicas e suas composições.



Fonte: Joyce et al (2006)⁵⁸.

A proteômica ampliou o conhecimento sobre as proteínas, que são as principais unidades moleculares funcionais das células. É de amplo conhecimento a importância dessas biomoléculas em inúmeros processos metabólicos e, em decorrência disto, a proteômica tem como finalidade a identificação e quantificação das proteínas codificadas a partir do genoma. Além do mais, essa estratégia ômica incorpora

atributos que podem ser empregados ao diagnóstico de algumas enfermidades, como por exemplo, alguns tipos de cânceres.⁵⁸⁻⁶² Assim como as proteínas podem fornecer dados relevantes para a proteômica, os metabólitos, em geral, proporcionam informações para duas outras áreas das estratégias ômicas: a metabolômica e a metabonômica.^{14,15,63}

Uma característica importante que tanto à metabolômica quanto à metabonômica possuem é o emprego das TEARs para a análise do metaboloma. Além disso, comumente essas duas metodologias utilizam biofluidos que podem ser obtidos de modo não-invasivo, por exemplo a urina, ou minimamente invasivos como o sangue. A metabonômica foi descrita pela primeira vez em 1999, enquanto a metabolômica teve seu marco formal em 2001.⁶⁴⁻⁶⁶ A Metabonômica foi definida por Lindon et al. (1999) como a avaliação quantitativa da resposta metabólica multiparamétrica dos estímulos fisiopatológicos ou das modificações gênicas, estando associada com a análise do perfil metabólico frente às doenças, alterações ambientais e variações genéticas.⁶⁵ De modo semelhante, Dunn e Ellis (2005) definiram a estratégia metabonômica como sendo uma avaliação dos tecidos e fluidos biológicos mediante às alterações dos níveis dos metabólitos endógenos resultados de doenças ou exposição à tratamentos terapêuticos.⁵⁶ Portanto, considerando que os organismos vivos respondem de modo complexo e imprevisível aos estímulos que impactam negativamente a saúde, podemos apresentar a metabonômica como sendo uma ferramenta que faz uso de técnicas analíticas modernas, que permite mensurar e construir modelos matemáticos a partir de biofluidos oferecendo uma visão panorâmica sobre o sistema vivo.⁶⁷

Por outro lado, a abordagem metabolômica tem como objetivo a identificação e quantificação de todos os metabólitos envolvidos em todo sistema biológico.

Enfatizamos que, além da caracterização, a quantificação é uma etapa de fundamental importância para a perspectiva metabolômica, visto que alterações em níveis metabólicos são os reflexos instantâneos das condições do fenótipo, deste modo as quantificações dos mesmos podem ser empregadas para fins de padronização do sistema biológico.^{55,68} Fiehn et al. (2001) prognosticaram que, assim como a proteômica, a metabolômica seria um elemento importante para a compreensão do fenótipo. Este fato se observa nos excelentes resultados dos estudos metabolômicos recentemente alcançados.^{14,55,58,69}

A distinção entre os dois termos sempre foi motivo de questionamentos. Nicholson (2008) apresenta a distinção entre metabolômica e metabonômica como mais filosófica do que necessariamente técnica. Além disso, o autor afirma que a metabonômica tem por finalidade a determinação da resposta dinâmica metabólica dos sistemas vivos a partir dos estímulos biológicos, enquanto a metabolômica busca uma descrição analítica, tanto do ponto de vista da identificação quanto da quantificação de todos os metabólitos na amostra biológica. Apesar disso, nesse mesmo trabalho, o autor evidencia que, na prática, frequentemente os termos são tratados como sinônimos, visto que os procedimentos e modelagem analítica são semelhantes.⁶⁷ Dunn e Ellis (2005), por sua vez, associaram as definições às técnicas analíticas empregadas. A necessidade de identificação e quantificação absoluta para a estratégia metabolômica implica no uso de ferramentas de separação de misturas e de técnicas espectrométricas sensíveis, como CG-EM (Cromatografia em fase Gasosa acoplada à Espectrometria de Massas) e CLAE-EM (Cromatografia Líquida de Alta Eficiência acoplada à Espectrometria de Massas); enquanto para a estratégia metabonômica, não há necessidade de quantificação absoluta, dispensando as ferramentas de separação de misturas, empregando a ferramenta espectrométrica de

forma direta. Nesse caso, destaca-se o uso da espectroscopia de Ressonância Magnética Nuclear (RMN). A diferença filosófica citada por Nicholson (2008) resulta numa diferença metodológica evidenciada por Dunn e Ellis (2005).^{55, 67}

No nosso caso, optamos por usar o termo metabonômica, haja vista que empregamos a espectroscopia de RMN em nossas análises e buscamos a identificação de padrões sem fazer determinações quantitativas absolutas. Seja na estratégia metabonômica ou metabolômica, a escolha do biofluido a ser utilizado no estudo depende do objeto de investigação e da relação que o biofluido tem com o mesmo. Estudaremos assim, casos de varicocele e de infertilidade masculina, por isso, o sêmen foi escolhido, mais especificamente, o fluido seminal.

5 EMPREGO DAS TÉCNICAS ESPECTROMÉTRICAS NO ESTUDO DA INFERTILIDADE MASCULINA

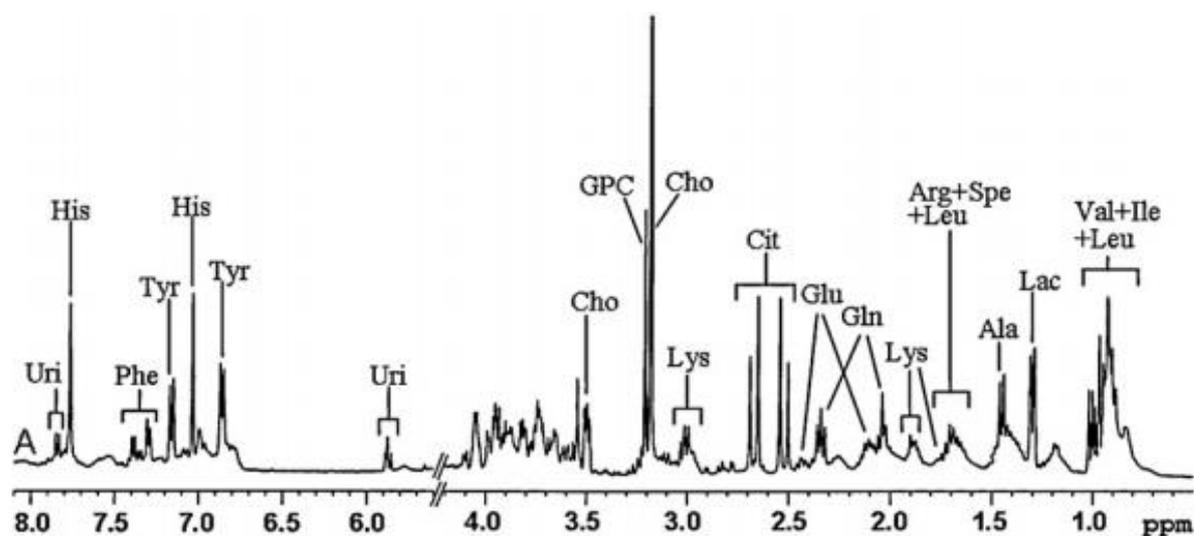
Estudos com métodos espectrométricos para a avaliação da infertilidade masculina são encontradas com facilidade na literatura. Jayaraman et. al. (2014) enfatizaram que o uso de técnicas metabolômicas podem ser a chave para o diagnóstico e tratamento da infertilidade nos homens. No mesmo trabalho, os autores descrevem que os estudos para identificação dos biomarcadores com o emprego das estratégias ômicas ainda estão em processo de amadurecimento, porém as perspectivas são o oferecimento de um diagnóstico e tratamento mais precisos da infertilidade.⁷⁰ Alguns estudos utilizando espectroscopia de RMN demonstraram bons resultados na identificação dos constituintes do plasma seminal e no diagnóstico da infertilidade. No entanto, não encontramos trabalhos associando varicocele e infertilidade, usando RMN.^{71,72}

Embora as mudanças observadas no soro seminal em estudos metabolômicos/metabonômicos não foram totalmente compreendidas, destacamos que esse biofluido contém as informações dos constituintes químicos do ejaculado e, em última análise, dos espermatozoides. Além do mais, esse fluido é um material que é extraído com facilidade por métodos não invasivos, fornecendo informações diretas das desordens reprodutivas nos homens.⁷²

Como já mencionado, o líquido seminal contém centenas de metabólitos e o espectro de RMN de ^1H contém sinais que podem ser atribuídos a esses metabólitos. Isso implica em um espectro com várias sobreposições. Vale salientar que um mesmo composto, em geral, apresenta mais de um sinal no espectro de RMN de ^1H devido aos diferentes ambientes químicos dos núcleos de ^1H presentes na estrutura química.

A Figura 4 apresenta um típico espectro de RMN de ^1H do líquido seminal descrito na literatura, contendo algumas atribuições⁷³.

Figura 4 – Espectro de RMN de H do plasma seminal humano.

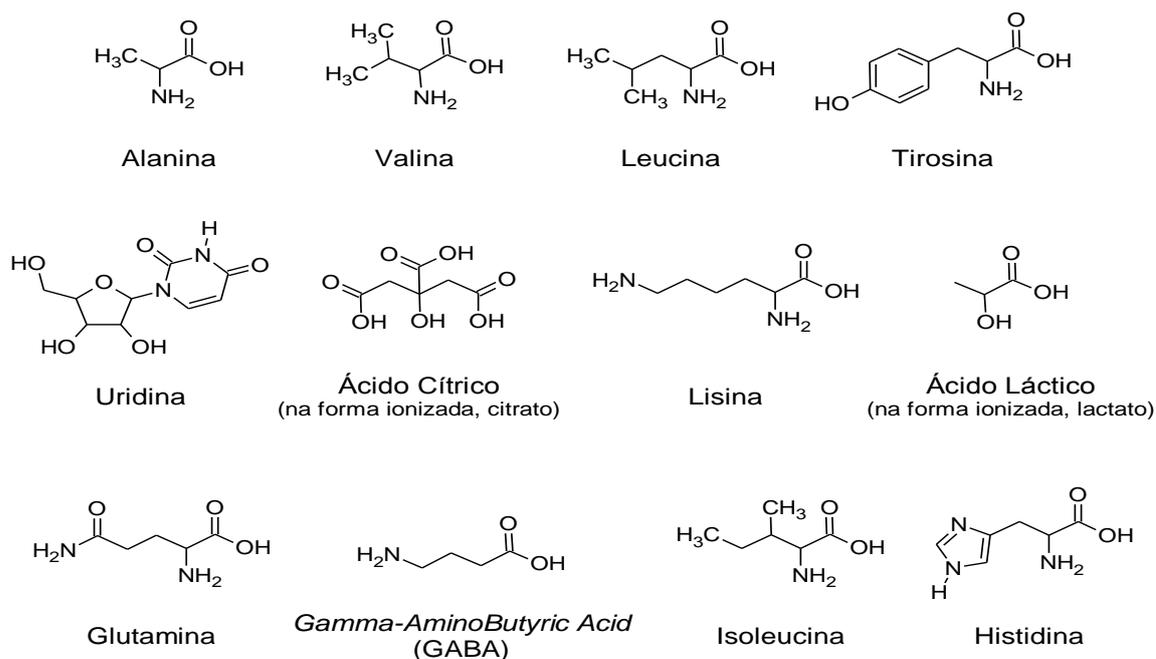


Fonte: Gupta et. Al (2011)⁷³.

O principal constituinte do líquido seminal é a água, que domina o espectro. Assim, a região que contém o sinal da água (entre δ 4,4 e 5,0 ppm) é excluída da apresentação. Sinais de grupos olefínicos e aromáticos apresentam-se com deslocamento químico (δ) superiores a 5,5 ppm, enquanto grupos alifáticos se apresentam em δ abaixo de 2,0 ppm. Em tese, quanto maior a densidade eletrônica sobre o núcleo, mais blindado e perto de δ igual a zero; quando menor a densidade eletrônica, mais desblindado e mais distante do δ igual a zero. As Figuras 5 e 6 apresentam as estruturas de alguns metabólitos contidos no fluido seminal e importantes no espectro de RMN de ^1H . No espectro da Figura 4, estão identificadas as metilas ($-\text{CH}_3$) da alanina, valina, leucina e lactato, todas se apresentando como dupletos com δ abaixo de 1,5 ppm. Também estão identificados os sinais atribuídos aos grupos aromáticos presentes na tirosina, histidina e uridina. Há de se destacar o sinal atribuído ao grupo metileno ($-\text{CH}_2-$) do citrato, que se apresenta no espectro

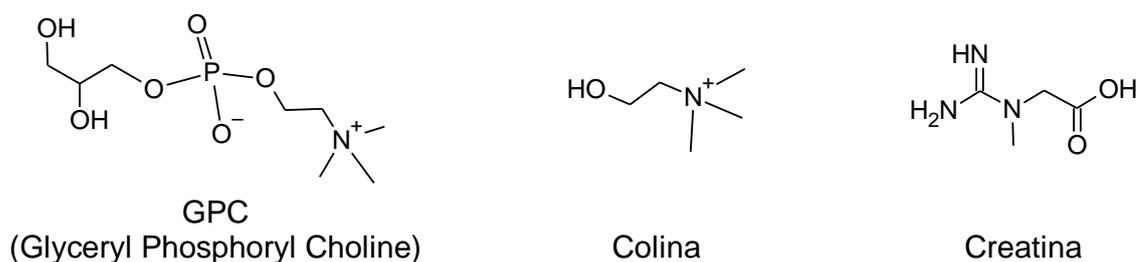
como dois dupletos, uma vez que eles são diastereotópicos; os sinais atribuídos às metilas da colina e da glicerofosforilcolina, simpletos em δ acima de 3,0 ppm; e um simpleto em δ 3,90 ppm atribuído ao grupo metileno da creatina.

Figura 5 – Estrutura química dos principais metabólitos encontrados no sêmen.



Fonte: Autoria própria.

Figura 6 – Estrutura química da Gliceril Fosforil Colina, da Colina e da Creatina.



Fonte: Autoria própria.

Hamamah et. al. (1998) empregaram a RMN de ^1H para distinguir a insuficiência espermatozônica da azoospermia (ausência de espermatozoides no ejaculado). Foram identificados quatro compostos significantes para a discriminação dos grupos: glicerofosforilcolina (GPC), colina, lactato e o nitrato. A relação entre a

área do sinal da colina/nitrato e da colina/lactato foram significativamente diferentes ($p < 0,01$) entre os grupos com falência espermatogênica e azoospermia obstrutiva. Os pesquisadores ainda observaram importante diferença na relação GPC/colina ($p < 0,01$) para diferenciar os grupos em questão, quando os valores séricos do FSH (hormônio folículo estimulante) são normais.⁷¹

Gupta et. al. (2011) com a proposta de fornecer de modo rápido, simples, e preciso o diagnóstico para diferentes casos de infertilidade, avaliaram quantitativamente, a partir dos dados de RMN de ^1H , a concentração absoluta dos possíveis metabólitos (biomarcadores) candidatos a descritores da infertilidade usando a análise de discriminantes lineares.⁷³ Neste trabalho, inicialmente foi observado, a partir da análise estatística univariada, que o nitrato, GPC e a fenilalanina poderiam ser usados como biomarcadores para a distinção do grupo controle dos inférteis, porém a sensibilidade, especificidade e a probabilidade preditiva para a discriminação dos subgrupos de infertilidade foram discretas. O problema foi contornado com o emprego da análise estatística multivariada, com o emprego para análise discriminante linear. Foram identificados que o grupo controle foi segregado do grupo infértil pela determinação das concentrações de alanina, nitrato, GPC, tirosina e fenilalanina. Além disso, foi possível distinguir os casos de normozoospermia dos oligozoospermia (pouca quantidade de espermatozoides no ejaculado) a partir das concentrações de lactato, uridina, histidina, alanina, fenilalanina, nitrato. Por fim, os autores concluíram que os aminoácidos impactam diretamente ou indiretamente na infertilidade masculina.⁷³

6 ANÁLISES ESTATÍSTICAS MULTIVARIADAS

A Quimiometria é um campo da Química Analítica que usa ferramentas matemáticas e de estatística multivariada para analisar dados químicos. Beebe e colaboradores definem Quimiometria como processos nos quais os dados químicos (números em uma tabela, por exemplo) são transformados em informações que são úteis para uma tomada de decisão ⁷⁴. A Quimiometria pode ser dividida em três grandes áreas: planejamento de experimentos; reconhecimento de padrões; e calibração multivariada ⁷⁵. A estratégia metabonômica busca essencialmente a identificação de padrões, associando-os aos status bioquímicos da amostra analisada. Nesse sentido, a avaliação estatística multivariada é um dos pilares que ampara a estratégia metabonômica. As áreas de integração associadas aos deslocamentos químicos, no espectro de RMN de ¹H, são as variáveis empregadas nos estudos. Assim, é construída uma matriz de dados, de forma que cada linha corresponde a amostra do paciente (casos) e cada coluna (variáveis) corresponde a uma região espectral. É essa matriz bidimensional que será utilizada nas modelagens quimiométricas. Antes da realização propriamente dita das análises quimiométricas, uma etapa determinante é requerida: pré-processamento dos dados.^{76,77}

6.1 PRÉ-PROCESSAMENTO DOS DADOS (PPD)

De modo geral, esta etapa é um procedimento importante para diversos tipos de análises, não sendo diferente para os dados espectrais. O objetivo principal do pré-processamento é restringir variações indesejáveis que poderão influenciar os resultados obtidos ao final das análises. O analista, ao dispor deste instrumento, deve ter atenção e ser bastante criterioso, uma vez que o pré-tratamento pode promover

distorções das informações contidas naquele conjunto de dados que, em última análise, pode gerar interpretações equivocadas a respeito dos dados analisados. Portanto, é interessante que vários procedimentos possam ser testados de modo que assegure que o método empregado foi o mais adequado possível⁷⁷.

Para melhor compreensão da necessidade do pré-processamento, é imprescindível entender que os sinais captados pelos instrumentos têm duas contribuições: uma determinística (sinal verdadeiro) e outra estocástica (ruídos aleatórios). Enquanto a contribuição determinística contém informações relevantes a respeito do estudo que está sendo avaliado, a componente estocástica acomoda as variações indesejáveis que são constituídas, em grande parte, pelos erros experimentais de medida, de amostragem, além de outras fontes de variabilidade.⁷⁵ Porém, não podemos afirmar que a contribuição determinística do sinal reflète, em sua totalidade, as informações relevantes; pelo contrário, dependendo do instrumento utilizado e das condições experimentais, as informações sistemáticas que não estão atreladas ao estudo podem estar associadas aos sinais, dessa forma, devem ser avaliadas e, se necessário for, corrigidas e ajustadas na etapa.⁷⁷

O PPD pode ser efetuado nas amostras (nas linhas) e/ou nas variáveis (nas colunas). Os métodos que são aplicados às linhas da matriz (amostras) têm por objetivo promover a minimização das variações aleatórias (ruídos experimentais). Alguns procedimentos são aplicados para tal finalidade, dentre os quais destacamos o alisamento (*smoothing*), que tem por finalidade reduzir os ruídos conseqüentemente aumenta a razão sinal/ruído (S/R). Na espectroscopia de RMN, esse procedimento é denominado de *linebroadening*, que é da ordem de 0,3 Hz. Por outro lado, as informações sistemáticas indesejáveis podem ser corrigidas pelas técnicas de ajuste de linha de base e de normalização das variáveis das amostras, por exemplo. Os

equipamentos analíticos modernos são integrados (interfaceados) aos computadores, de modo que, alguns tratamentos dos dados são automaticamente aplicados ou podem ser selecionados pelo analista no próprio software de interfaceamento do equipamento.^{76,77}

Após a aquisição da informação espectral a partir do biofluido, comumente outra etapa de tratamento dos dados é requisitada: intervenções matemáticas para padronização da informação. Desta forma, algumas estratégias são utilizadas como pré-processamentos, as quais podem ser aplicadas às variáveis (mais comum) e/ou aos casos (amostras). Em algumas circunstâncias, é possível a aplicação tratamentos matemáticos com finalidade de reduzir possíveis influências que venham impactar a etapa de modelagem dos dados, por exemplo temos o uso de filtros que eliminam ruídos, evidenciando os sinais que de fato são importantes na modelagem.

Dentre as operações possíveis, o SNV (do inglês *Standard Normal Variables*) é o recurso mais empregado quando se trata de pré-processamento das amostras. Em relação as variáveis, outros operadores têm seu uso bastante difundido na quimiometria, como exemplos podemos descrever o autoescalonamento, e a estratégia de centrar os dados na média.

6.2 ANÁLISE EXPLORATÓRIA DOS DADOS: ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

Os métodos comumente empregados na identificação de semelhanças e diferenças em determinado conjunto de amostras a fim de se obter um padrão de classificação são reunidos em dois grupos: métodos supervisionados e os não-

supervisionados. Independente da abordagem a ser aplicada, ambos estão alicerçados na premissa que as amostras do mesmo tipo são parecidas; havendo alterações significativas entre as amostras de grupos distintos; e, por fim, é possível detectar tais similaridades e diferenças a partir do conjunto de medidas que foram extraídas das amostras.⁷⁶⁻⁷⁸

Os métodos não-supervisionados são aqueles que não fazem uso das informações de classe. Esta etapa faz parte da etapa da análise exploratória dos dados em que serão observados a ocorrência natural de agrupamentos entre as amostras, bem como a investigação da ocorrência de amostras anômalas. Para dados espectrais de RMN, a análise exploratória frequentemente empregada é a Análise de Componentes Principais (*PCA – Principal Component Analysis*). Este método de tratamento de dados multivariados propõe a redução da dimensionalidade do conjunto de dados originais, produzindo um novo sistema de coordenadas formado por eixos ortogonais, que são combinações lineares das variáveis originais (componentes principais), de modo que as diferentes fontes de variação do conjunto de dados sejam isoladas em diferentes componentes principais, que são ortogonais entre si. Ou seja, a variância contida numa componente principal não está contida nas outras. Em outras palavras, a PCA executa a compressão dos dados (redução no número de variáveis) através da projeção dos mesmos em um espaço de dimensão menor.⁷⁸

Matematicamente, a compressão dos dados, é possível a partir da combinação linear das variáveis originais, de tal forma que as que transportam informações similares sejam agrupadas. A consequência desta operação será a obtenção de um conjunto de variáveis denominadas de Componente Principais (PCs).⁷⁶

Para entender o efeito da redução da dimensionalidade promovida pela PCA, os dados devem ser dispostos numa matriz X , com i linhas ($i=1, 2, \dots, i$); chamadas de

amostras ou casos, e j colunas ($j = 1, 2, \dots, j$), que são as variáveis. As variáveis individuais desta matriz X são expressas por x_j ($j = 1, 2, \dots, j$) e podem ser interpretadas como todos os vetores no espaço dimensional- i , ou seja, as variáveis individuais correspondem a todos os vetores para cada caso (i). Considerando que a matriz X carrega as variações relevantes para a identificação de grupos, é possível o raciocínio que mediante a redução da dimensionalidade, a partir da combinação linear das variáveis originais x_j , tais variações sejam ainda mantidas, mas, tendo como incremento a possibilidade de o conjunto de dados ser avaliado com menor complexidade, além de permitir a identificação das variáveis que influenciam na modelagem e como elas impactam nos diferentes grupos. Dessa forma, podemos expressar a combinação linear das x variáveis da matriz X de acordo com a equação 1.

$$t = w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_j \cdot x_j \quad (eq. 1)$$

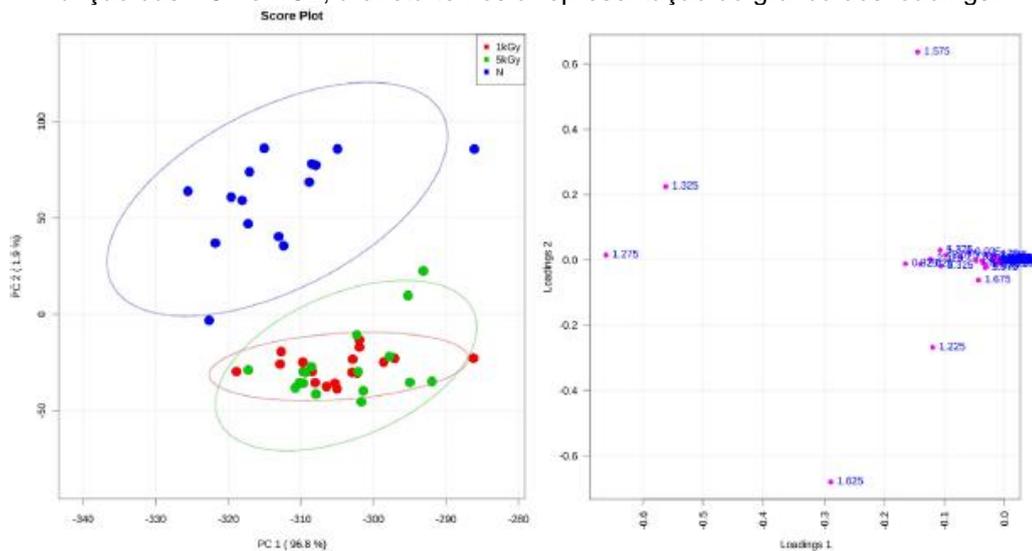
onde t corresponde ao novo vetor (componente principal) obtido, sendo w os pesos de cada variável nessa componente. A variação em t pode ser estimada a partir da análise de variância. Neste tipo de abordagem é desejável obter a variância máxima possível em t , para que isto seja alcançado deverá ser escolhido os pesos ideais w_1, w_2, \dots, w_j ⁷⁶. Podem ser construídas $(i-1)$ vetores t , de forma que o somatório da variância contida (v_n) em cada vetor seja igual a 100% (equação 2).

$$\sum_{n=1}^{(i-1)} v_n = 100\% \quad (eq. 2)$$

Considerando que os diferentes vetores t construídos são ortogonais, entre si, na prática, tem-se um novo sistema de coordenadas. Normalmente, o resultado da Análise de Componentes Principais é apresentado através da projeção das amostras (gráfico de escores) ou das variáveis (gráfico de pesos) nesse novo sistema de

coordenadas. Com isso, uma das vantagens da PCA se faz presente, pois pode-se visualizar o agrupamento das amostras e identificar quais variáveis originais são mais relevantes para segregar os grupos. Além disso, podem ser visualizadas amostras com comportamentos anômalos (*outliers*), visto que na projeção dos dados em um novo sistema de coordenadas as diferenças entre as amostras tendem a se amplificar.⁷⁷ A Figura 7 apresenta um exemplo típico de apresentação de resultados de uma PCA, com os gráficos de escores (esquerda) e pesos (direita).

Figura 7 – Exemplo de uma PCA. O gráfico à esquerda representa a distribuição dos escores em função das PC1 e PC2; à direita temos a representação do gráfico dos loadings.



Fonte: Ribeiro et al (2014)⁹².

6.3 ANÁLISE DISCRIMINANTE

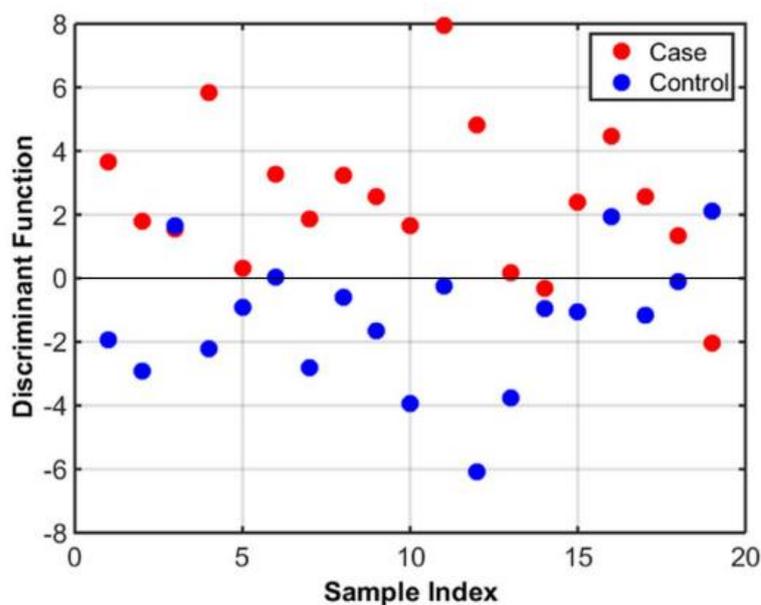
Entre os métodos supervisionados usados na estratégia metabonômica, destaca-se a Análise Discriminante (DA, do inglês *Discriminants Analysis*) que é, normalmente, empregada quando a variável de resposta é uma variável de classe. Aqui, investigamos diagnóstico clínico, portanto, uma variável de classe, o que torna os formalismos DA candidatos naturais para a modelagem. A DA é uma ferramenta de reconhecimento de padrões que se baseia na construção de combinações lineares das variáveis originais, visando a obtenção de um escore (ponto de corte) que discrimine as amostras nos grupos de interesse. Nos estudos classificatórios, cada amostra é descrita por um conjunto de medidas experimentais. A grande vantagem da DA é permitir a identificação das variáveis que mais se sobressaem na discriminação das classes. Isso permite, por um lado, separar as amostras e, por outro, inferir sobre a natureza química dessa discriminação. A seleção das variáveis é realizada usando formalismos de estatística multivariada, como o lambda de Wilks, o algoritmo genético, o teste qui-quadrado e/ou correlação canônica.⁷⁷

Atualmente, estão disponíveis na literatura vários formalismos que podem ser empregados para o processamento de problemas que envolvem a identificação de grupos. Aqui, empregamos os formalismos LDA (do inglês *Linear Discriminants Analysis*), PLS-DA (do inglês *Partial Least Square Discriminants Analysis*) e o OPLS-DA (do inglês *Orthogonal Partial Least Square Discriminants Analysis*).

6.3.1 Análise de Discriminante Linear (LDA)

A Análise de Discriminante Linear (LDA) permite a determinação de uma superfície (classificador) ou limiar de classificação no espaço, para que as amostras de uma classe sejam segregadas de um lado do eixo de decisão, assim como as amostras da outra classe sejam agrupadas no outro. Em outras palavras, pode-se entender a LDA como um método estatístico que pode ser empregado em reconhecimento de padrões. Para tal é construído uma combinação linear de variáveis previamente selecionadas, visando caracterizar e separar duas ou mais classes (casos) ⁷⁷. A figura 10 apresenta um exemplo de gráfico de escores de uma LDA desenvolvida para avaliar um possível método para diagnóstico de fibromialgia a partir da análise do plasma sanguíneo através da ATR-FTIR (do inglês, *attenuated total reflection Fourier-transform infrared*). Podemos observar a função de classificação dividindo o grupo controle (em azul) dos casos (vermelho).⁷⁸

Figura 8 – Exemplo de gráfico de escore de uma LDA.



Fonte: Passos et al. (2020)⁷⁸.

Evidentemente que o ideal é que as amostras de cada classe sejam ao máximo parecidas entre si, e estejam relativamente distantes das amostras consideradas da outra classe. Todavia, nem sempre é possível obter tal resultado. Como lidamos com variáveis de classe, podemos arbitrariamente atribuir o valor 1 (um), para amostras de um grupo, e o valor -1, para as amostras do outro grupo, no caso de dois grupos apenas. Matematicamente, a superfície de decisão ou classificador é o plano que corresponde à equação 3:

$$k(x) = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n = 0 \quad (eq. 3)$$

Em que $k(x)$ representa a variável dependente; a_n , os coeficientes discriminantes para cada variável; e x_n as variáveis explicativas do formalismo LDA⁷⁷.

Considerando a figura 8, a interpretação mais simples para os resultados da equação 1 é que as amostras Caso (em vermelho), que estão dispostas acima do plano, devem assumir coordenadas que satisfaçam a condição $k(x) > 0$. Prontamente, as amostras Controle (em azul), localizadas abaixo do plano, terão coordenadas tais que $k(x) < 0$ ⁷⁸. Portanto, a superfície de decisão descrita na eq.1 é consequência da combinação linear das variáveis preditivas que melhor descrevem a LDA. A equação apresentada é, em outras palavras, a função discriminante do modelo.

Quanto ao critério empregado para a escolha das variáveis, o lambda de Wilks (Λ) é um dos mais utilizados, e que, de modo sucinto, permite a averiguação da existência de diferenças de médias entre as classes para cada variável. O valor do Λ pode variar de 0 a 1, sendo expresso pela equação 4.

$$\Lambda = \frac{SQ_{dg}}{SQ_T} \quad (eq. 4)$$

Em que SQ_{dg} corresponde à soma quadrática dos erros dentro do grupo, e SQ_T é a soma quadrática residual total. O conjunto de variáveis preditivas é selecionado em função do Λ , que deve ser o menor possível, haja vista que a combinação de variáveis,

que não têm relação com o objeto de estudo, resultará em resíduos (SQ_{dg}) grandes, enquanto combinação entre variáveis preditivas apresentam desvios menores. Assim, o conjunto de variáveis selecionada para a função discriminante é aquele que apresenta o menor valor de Λ .¹⁰⁶

Outra técnica para seleção de variáveis que pode ser associada à LDA é o Algoritmo Genético (do inglês, *Genetic Algorithms* – GA). Os GAs são operadores matemáticos desenvolvidos pelo John H. Holland, na década de 60, baseados nos processos de seleção natural e recombinação genética para utilização em sistemas complexos. Esta estratégia almejava simular, matematicamente, o mecanismo da evolução biológica. Nas décadas seguintes, mediante os avanços tecnológicos dos computadores, observou-se um impulsionamento do uso do GA em diversas áreas como processamento de imagem; modelagem e identificação de sistemas; filtros para cancelamento de ruídos; robótica; e seleção de parâmetros de redes neurais.⁸⁰

Assim como em outras áreas, a química envolvida nos sistemas biológicos, apresenta problemas de alta complexidade, e o desenvolvimento do GA disponibilizou mais uma ferramenta na busca por respostas às situações de difícil resolução. Os GAs são empregados para a otimização da seleção de variáveis que melhor classifica os grupos, ou seja, considerando a existência de duas classes em um conjunto de dados, este algoritmo é capaz de identificar as variáveis chaves que maximizam a separação entre as classes em questão⁸⁰. Além disso, o GA oferece como grande vantagem, em relação aos métodos tradicionais de seleção de variáveis, a capacidade de avaliar as relações existentes entre as variáveis. Assim sendo, os algoritmos genéticos buscam avaliar os resultados das combinações das variáveis e seu impacto sobre a função classificatória.

A seleção proposta pelo GA está baseada na premissa que as informações referentes a um determinado sistema ou conjunto de dados são codificadas de maneira análoga aos cromossomos. Dessa forma, se assemelhando ao processo de seleção natural das espécies proposta por Charles Darwin. Esta operação matemática se processa em cinco etapas: codificação das variáveis, criação da população inicial, avaliação da resposta, cruzamento e mutação.⁸⁰

A primeira fase consiste na confecção de um cromossomo “artificial” estruturado a partir das variáveis presente no estudo. Esta analogia aos cromossomos biológicos é possível de ser idealizada, visto que os cromossomos são estruturados a partir do DNA (ácido dioxirribonucleico) e as regiões específicas que carregam informações referente as características fenotípicas (genes) que representará, na modelagem matemática, os parâmetros que nosso conjunto de dados (variáveis). O modo mais fácil de representar cada uma das variáveis que serão possivelmente selecionadas ao final do processo é transformando-as em códigos binários.

A etapa seguinte é a criação de uma população inicial que nada mais é do que gerar uma população aleatória de cromossomos artificiais, assegurando a não existência de tendências oriundas do meio externo. Possivelmente, este processo exigirá um esforço computacional significativo. Então, em determinadas situações, informações adicionais podem ser empregadas com o objetivo de minimizar este tempo de análise.

No processo evolutivo, a avaliação da resposta adaptativa é uma característica intrínseca do indivíduo que reflete diretamente em suas habilidades de como irá se adequar as mudanças. No caso do GA, a avaliação da resposta terá como métrica as variáveis que apresentaram melhor resposta. Esta será interpretada segundo o erro de cálculo entre o valor previsto e o valor esperado. A finalidade aqui será identificar

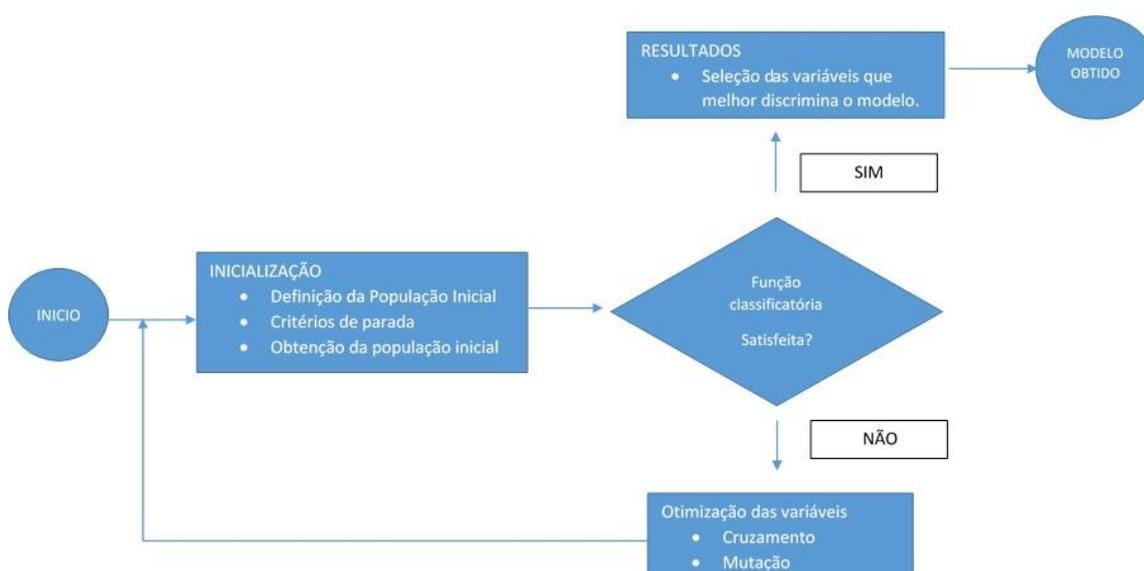
o menor erro possível, e este será o responsável direto pela “vida ou morte” daquele indivíduo.

As etapas de cruzamento e mutação são importantes para a seleção de variáveis dominantes. Especificamente, é a fase de cruzamento responsável pela convergência para a otimização do processo. A técnica de cruzamento ocorre pela escolha fortuita dos indivíduos (casos) e da posição onde ocorrerá a ruptura seguida pela troca de material genético. Essas rupturas geralmente variam entre 1 e 2, pois uma maior fragmentação promovera menor semelhança entre as gerações e por conseguinte, dificultando a convergência.

Estas etapas gerais do algoritmo genético básico são ininterruptamente repetidas até que o número máximo de gerações seja adquirido ou até que um erro mínimo desejado seja alcançado⁸⁰. No que diz respeito ao emprego dos GAs na seleção de variáveis, a sua implementação difere-se de outras aplicações quanto à codificação do problema bem como a avaliação da função de resposta. Entretanto, as outras etapas permanecem inalteradas. Na codificação para efeitos de seleção de variáveis espectrais, considera-se que o conjunto de genes do cromossomo artificial são todas as variáveis espectrais, isto quer dizer que, cada gene será representado por uma das variáveis espectrais ou derivações destas (neste trabalho utilizamos o deslocamento químico). Assim sendo, o cromossomo apresentará o mesmo número de variáveis contidas no espectro. Inicialmente, o software irá codificar aleatoriamente o conjunto de dados espectrais em zero e um. Cada gene poderá assumir o valor um ou zero, dito isto, que quando a posição assumida referente a uma determinada variável for determinada, aleatoriamente, igual a um, isto implicará na seleção desta variável. Por outro lado, se a posição for atribuída o valor zero, indicará que a variável não será selecionada. Em seguida serão realizados os procedimentos de

cruzamentos e a resposta serão avaliadas a partir do formalismo empregado para a averiguação da resposta. É importante destacar que o resultado é avaliado de acordo com o risco de avaliação incorreta, ou seja, quanto menor o risco de classificação incorreta, melhor será o conjunto de variáveis avaliadas e a partir dessas, novas combinações serão testadas de modo que valores otimizados de exatidão sejam atingidos.

Figura 9 – Fluxograma que representa o Algoritmo Genético.



Fonte: Autoria própria.

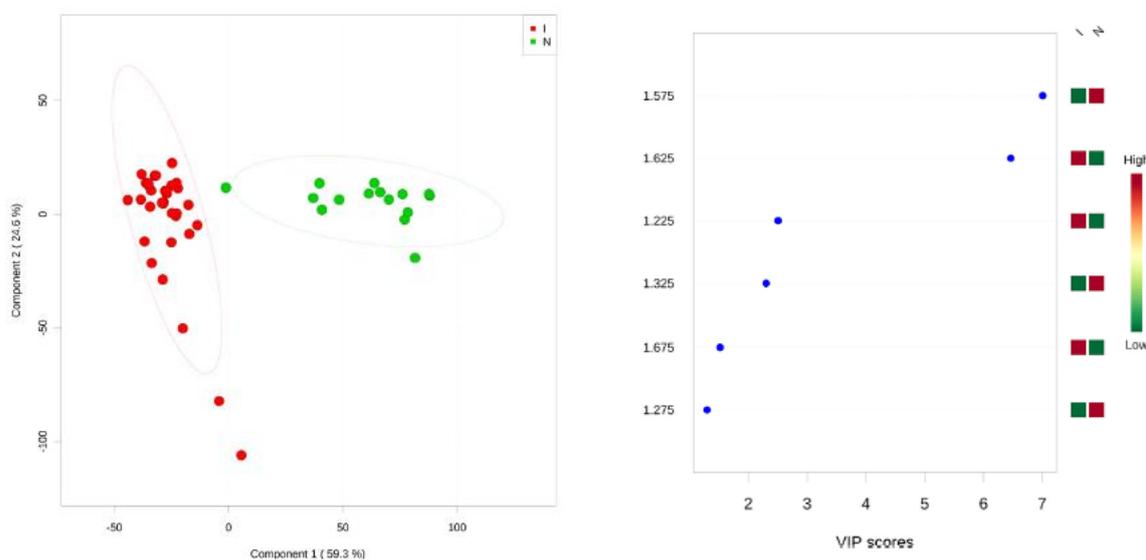
6.3.2 PLS-DA / OPLS-DA

A PLS-DA é um método supervisionado comumente utilizado na otimização da separação entre duas classes de amostras. O modelo é construído a partir de duas matrizes: uma matriz \mathbf{X} (em nosso caso, dados espectrais) e uma matriz de classe \mathbf{Y} , que contém a classificação das amostras. De modo geral, o fundamento desse formalismo está na maximização da covariância entre as variáveis independentes (matriz \mathbf{X}), e as variáveis dependentes (as informações de classe), contidas na matriz

Y, de um conjunto de dados multidimensional, determinando o subespaço linear das variáveis explicativas (latentes), componentes PLS⁸¹.

Dentre as vantagens fornecidas pelo formalismo PLS-DA para dados espectrais, destacamos a possibilidade de avaliação de dados altamente correlacionados e extração de elementos estatísticos que avaliam: a contribuição de cada variável para a construção dos modelos (os pesos) e os valores VIP (do inglês, *variance importante on projection*). Finalmente, o formalismo PLS-DA permite uma fácil e simples identificação gráfica dos dados, em uma dimensão reduzida, destacando a separação entre classes distintas.^{81,82} A figura 10 apresenta um exemplo de PLS-DA, com os gráficos de escores e VIP escores, permitindo identificar as variáveis mais importantes para a discriminação.

Figura 10 – Exemplo de PLS-DA. À esquerda, gráfico de escores; à direita, VIP escore.

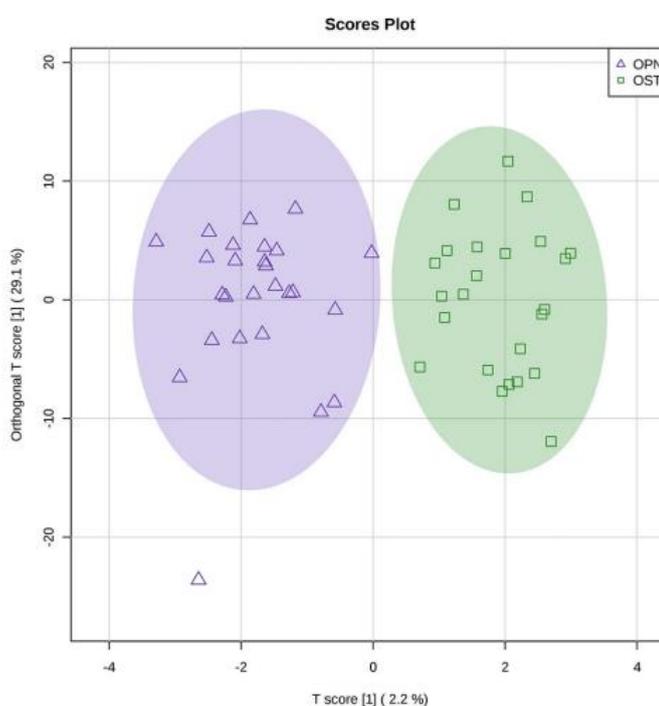


Fonte: Ribeiro et al. (2014)⁹².

Em alguns casos, a variância intraclasse é tão grande que dificulta a observação da discriminação desejada, usando o formalismo PLS-DA. Nesses casos, uma alternativa seria concentrar a variância intraclasse numa componente ortogonal a variável latente. Esse formalismo é chamado de OPLS-DA. A grande vantagem da

OPLS-DA frente à PLS-DA é que uma única componente preditiva é empregada na discriminação da classe, pois os demais componentes difundem suas fontes de variações na representante ortogonal à primeira componente (a qual explica a maior variância dos dados associada a matriz resposta). A componente responsável pela capacidade preditiva do formalismo OPLS-DA reflete a diferença entre os valores médio da classe A e as médias da classe B. Então, todas as amostras são projetadas nessa componente para ser estimado o valor dos escores.^{82,83}

Figura 11 – Exemplo de gráfico de escores na OPLS-DA.



Fonte: Pontes et al. (2019)⁹⁴.

6.3.3 Figuras de Mérito e Avaliação dos Modelos Multivariados

Após a elaboração dos modelos de classificação, os mesmos devem ser avaliados a partir da sua capacidade assertiva de enquadrar as amostras nos seus respectivos grupos. O conjunto de parâmetros utilizados para esse avaliar o desempenho de um modelo é chamado de figuras de mérito. E, em geral, o

desempenho do modelo é comparado às figuras de mérito alcançada por um padrão-ouro. Este apresenta-se como um teste referência empregado para o diagnóstico que, via de regra, tem elevada capacidade de determinar quem é portador ou não de uma doença, com baixos índices de falsos positivos e falsos negativos.¹⁰⁷

A ausência de erros de classificação dos modelos é um dos indicadores de qualidade da modelagem que já deve ser avaliada durante a construções dos modelos para que limites aceitáveis sejam atingidos ainda nesta etapa. Porém, é interessante que os modelos devam ser submetidos à avaliações de amostras externas que não foram empregadas na sua construção, empregando as figuras de mérito como indicadores de desempenho dos modelos.⁷⁷

O objetivo é sempre a construção de um modelo que represente bem o universo e que erre o mínimo possível. Para isso, o ideal é ter um grupo para a modelagem e outro, independente, para validação. Quando isso não é possível, uma alternativa é usar Validação Cruzada. Nesse caso, as n amostras utilizadas para selecionar as variáveis discriminatórias, e construir o modelo propriamente dito, são utilizadas para validar o modelo. Retira-se uma amostra por vez e constrói-se modelos com $(n - 1)$ amostras e testa-se a amostra retirada. Após testar todas as amostras com os n modelos criados, constrói-se uma matriz de contingência com esses dados. As figuras de mérito são extraídas a partir dessa matriz.⁷⁷

A elaboração das métricas mais simples inicia-se pela identificação dos casos Verdadeiro-Positivo (VP), Falso Positivo (FP), Verdadeiro-Negativo (VN) e Falso-Negativo (FN). Tais parâmetros podem ser dispostos em uma tabela conhecida como matriz de contigência (tabela 5).

Tabela 5 – Exemplo de Tabela de Contigência.

		PADRÃO-OURO	
		Positivo	Negativo
TESTE DIAGNÓSTICO	Positivo	VP	FP
	Negativo	FN	VN

Fonte: Autoria própria.

Dependendo da análise que será realizada, os valores de FP ou FN podem inviabilizar o emprego do teste. Todavia, o primeiro parâmetro que será apresentado é o de Exatidão, que é a expressão da fração das amostras que foram classificadas corretamente pelo teste que está em avaliação (Equação 5).

$$Exatidão = \frac{(VP + VN)}{(VP + VN + FP + FN)} \quad (eq. 5)$$

Outras duas métricas importantes rotineiramente empregadas para a avaliação dos testes diagnósticos são Sensibilidade (S) e Especificidade (E). A primeira objetiva refletir a porcentagem das amostras positivas, geralmente associadas aos pacientes doentes, que foram classificadas pelo método do conjunto total de amostras positivas determinadas pelo padrão-ouro (Equação 6). A Especificidade (E) representa a porcentagem de amostras que foram atribuídas como negativo (“pacientes sadios”) pelo modelo testado do conjunto total de amostras negativas notadamente atribuídas como negativas pelo padrão-ouro (Equação 7).

$$S = \frac{VP}{(VP + FN)} \quad (Eq. 6)$$

$$E = \frac{VN}{(VN + FP)} \quad (Eq. 7)$$

A interpretação clínica para essas métricas diz que a Sensibilidade representa a capacidade do teste diagnóstico identificar corretamente os pacientes que, via de regra, são portadores da doença quando esta condição se faz presente; enquanto a Especificidade exprime a competência do teste diagnóstico de encontrar as amostras

negativas no universo dos pacientes com ausência de doença. Sendo assim, para estas duas métricas os resultados errados, sejam eles FP (falso-positivo) ou FN (falso-negativo), comprometem o desempenho do teste. Contudo, esses erros de classificação podem variar para cada modelagem, de modo que precisam ser analisados isoladamente. Para isso, dois outros parâmetros são empregados para avaliar o modelo em questão, são eles: o Valor Preditivo Positivo (VPP) e o Valor Preditivo Negativo (VPN). O VPP indica a probabilidade da amostra classificada como positiva realmente corresponder a esta classificação (Equação 8). O VPN representa a probabilidade de as amostras negativas de fato serem negativas para o método em questão (Equação 9). Portanto, em última análise, o VPP e o VPN também fornecem informações sobre a precisão do modelo e esses cinco parâmetros devem ser considerados em conjunto.^{108,109}

$$VPP = \frac{VP}{(VP + FP)} \quad (Eq. 8)$$

$$VPN = \frac{VN}{(VN + FN)} \quad (Eq. 9)$$

Na avaliação da qualidade dos modelos obtidos, espera-se que toda variância contida no conjunto de dados seja integralmente modelada. Contudo, isto só seria possível para um modelo hipoteticamente perfeito. Na prática, uma parte da variação não é modelada, visto que existem diferenças entre os valores preditos e as medidas estimadas a partir da modelagem dos dados. Considerando que a modelagem é uma regressão, é comum usar o coeficiente de determinação (R^2) e o valor de F (teste de Fischer) para avaliar a qualidade da modelagem. Matematicamente, R^2 é a razão entre a variância explicada (SQ_{REG} – soma quadrática da regressão) pelo modelo e a variância total do conjunto de dados (SQ_T – soma quadrática total). Em outras palavras, este parâmetro descreve o quanto o modelo está ajustado aos dados.⁷⁷

$$R^2 = \frac{SQ_{REG}}{SQ_T} \quad (Eq. 10)$$

Assim, uma modelagem é tanto melhor quanto mais perto de 1 (um) for o valor de R^2 , pois significa que a soma quadrática da regressão se aproxima do valor da soma quadrática total. No caso do teste F (de Fischer), ele testa a hipótese de que a regressão construída se ajusta bem ao conjunto de dados utilizados na modelagem, considerando as variâncias explicada e não explicadas pelo modelo. Matematicamente, o valor de F é a razão entre a variância da regressão (SQ_{REG}/k), porção explicada pelo modelo, e a variância residual ($SQ_{RES}/(n-k-1)$), parte não explicada pelo modelo, conforme descrito pela equação 11.

$$F_{p,(n-k-1)} = \frac{(SQ_{REG}/k)}{[SQ_{RES}/(n-k-1)]} \quad (eq. 11)$$

Onde n é o número de amostras e k é o número de variáveis utilizadas na modelagem.

Se a regressão explica bem a variância contida no conjunto de dados, o valor de F será alto, pois a variância explicada será muito maior que a variância não explicada pelo modelo. Utiliza-se a tabela de probabilidade para uma distribuição F , considerando os graus de liberdade utilizados para determinar a probabilidade (valor de p) para rejeitar a hipótese nula (o modelo não explicar a variância contida no conjunto de dados).⁷⁷

De acordo com os dados amostrais em análise, alguns algoritmos de classificação podem influenciar significativamente a modelagem. Por exemplo, nos casos em que a matriz de dados possui um número elevado de variáveis e poucas amostras, os modelos supervisionados podem estar superajustados. Quando isso acontece, o algoritmo de classificação irá discriminar os grupos a partir de variáveis para as quais as diferenças entre as classes ocorrem ao acaso. Assim, quando novas amostras são avaliadas, a função classificatória construída não é capaz de qualificar

corretamente, visto que o padrão de detecção utilizado não foi em função das variáveis de interesse, mas sim devido ao componente aleatório. Para investigar casos de superajustes do modelo, alguns testes podem ser empregados de modo a avaliar se o classificador obtido é uma consequência direta das variáveis que discriminam o modelo ou não.

O teste de permutações é frequentemente utilizado para esse fim. O mecanismo deste teste é avaliar o desempenho dos modelos obtidos a partir de permutações aleatórias das classes originais, empregando o mesmo conjunto de dados. No caso da PLS-DA, a execução do teste é realizada usando a capacidade de predição ou distância de separação entre os grupos. Para cada permutação, um modelo PLS-DA é construído entre bloco das variáveis e a matriz resposta permutada usando o mesmo número de variáveis latentes previamente identificadas. O valor p reflete a proporção das vezes em que a separação de classe atribuída aleatoriamente é melhor que a baseada nos dados originais. Espera-se que os modelos obtidos a partir de permutações de classes tenham figuras de mérito piores que os valores observados para a modelagem feitas com a classificação original.

7 HIPÓTESE

Varicocele, infertilidade e intervenção cirúrgica (varicocelectomia) são condições que, devido à homeostase, alteram o conjunto dos metabólitos endógenos, tornando possível a associação entre o perfil metabólico e o *status* bioquímico do paciente. Dessa forma, admite-se que os espectros de ressonância magnética nuclear de ^1H , obtidos a partir do plasma seminal, contenham informações acerca das alterações dos perfis metabólicos em função da doença ou da melhora (ou não) dos padrões seminais pós-varicocelectomia.

8 OBJETIVOS

8.1 OBJETIVO GERAL

✓ Desenvolver modelos metabonômicos, utilizando espectros de RMN de ^1H a partir de amostra biológica, fluido seminal, para o diagnóstico da varicocele e da infertilidade associada à varicocele, bem como fazer prognóstico da melhora dos parâmetros seminais pós varicocelectomia.

8.2 OBJETIVOS ESPECÍFICOS

✓ Obter os espectros de RMN de ^1H a partir do plasma seminal de voluntários férteis e inférteis, diagnosticados com varicocele, bem como de voluntários saudáveis;

✓ Investigar quais ferramentas de análise multivariada melhor se aplicam aos dados coletados – PCA, PLS-DA, OPLS-DA e GA-LDA, e construir modelos metabonômicos para o diagnóstico da varicocele e da infertilidade associada à varicocele;

✓ Identificar diferenças nos perfis metabonômicos de pacientes diagnosticados clinicamente com varicocele dos pacientes saudáveis;

✓ Identificar diferenças nos perfis metabonômicos de pacientes férteis e inférteis, ambos diagnosticados clinicamente com varicocele;

✓ Selecionar as variáveis discriminantes em cada modelo metabonômico e identificar quais metabólitos endógenos estão associados às possíveis discriminações a serem observadas no presente estudo.

9 MATERIAIS E MÉTODOS

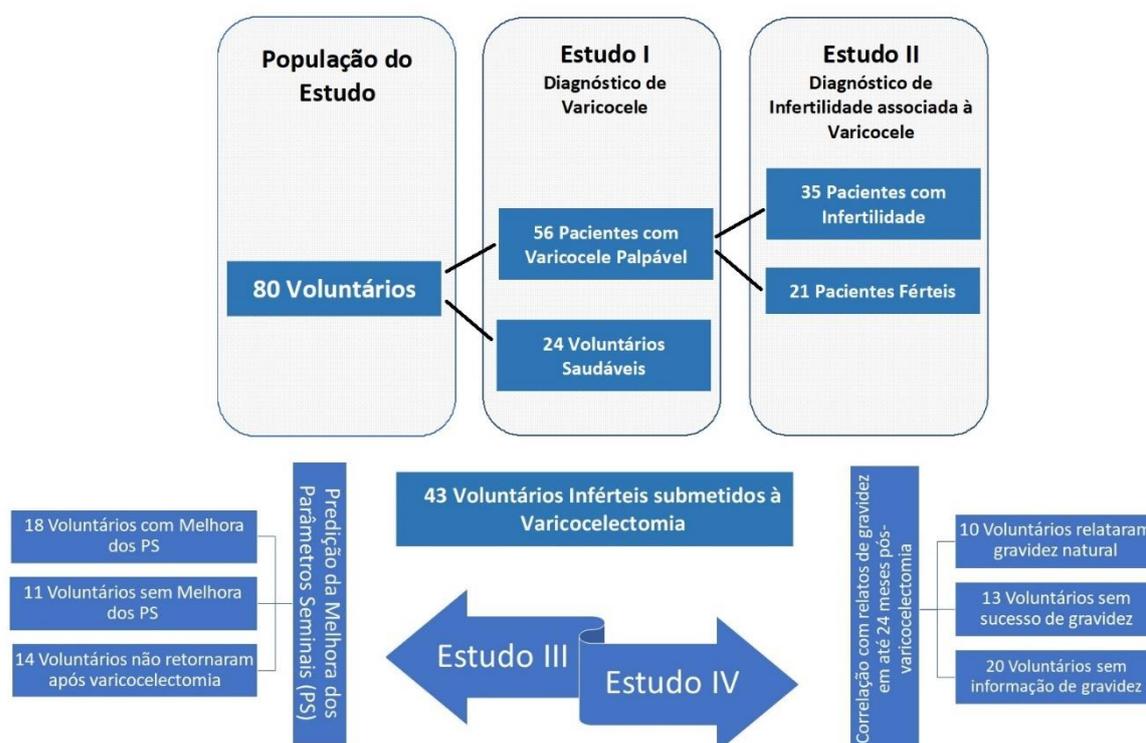
9.1 SELEÇÃO DOS PACIENTES E GRUPOS DE ESTUDO E TRATAMENTOS ESTATÍSTICOS UNIVARIADOS

Este estudo foi realizado através da parceria entre pesquisadores dos programas de pós-graduação em Química e de Cirurgia da Universidade Federal de Pernambuco (UFPE). Foram selecionados 80 homens voluntários, na faixa etária de 18 a 50 anos, atendidos na clínica de Fertilidade Masculina do Instituto de Medicina Integral Professor Fernando Figueira (IMIP) para efetuar vasectomia ou tratar varicocele.

A avaliação clínica foi executada por um único urologista e doutorando do PPG em Cirurgia, do Centro de Ciências da Saúde da UFPE. Foram realizados quatro estudos e a distribuição dos voluntários se deu respeitando esse cenário. O estudo I contou com 80 voluntários, que foram assim distribuídos: Grupo Controle (C) – 24 voluntários saudáveis que tinham pelo menos 1 filho nos últimos 12 meses, sem histórico familiar de infertilidade, que não fizeram tratamento de fertilidade, nem possuíam diagnóstico de varicocele; e Grupo Varicocele (V) – 56 pacientes com varicocele palpável, podendo ser infértil ou não, sem histórico de tratamentos para fertilidade. Para o segundo estudo, o grupo V foi dividido em dois subgrupos: Varicocele Fértil (VF) – 21 voluntários férteis com varicocele palpável que tinham pelo menos 1 filho nos últimos 12 meses, sem histórico familiar de infertilidade, que não fizeram tratamento de fertilidade; e o grupo Varicocele Infértil (VI) – 35 voluntários inférteis com varicocele palpável que não obtiveram êxito na concepção ao final de 12 meses com atividade sexual regular e sem emprego de métodos contraceptivos, sem histórico de outras causas de infertilidade ou em tratamento de infertilidade e com causas de infertilidade feminina desconhecidas por parte da parceira. O terceiro e

quarto estudos foram realizados com amostras de 43 voluntários, diagnosticados com varicocele e submetidos à varicocelectomia. Os dados espectrais usados foram de amostras coletadas antes da cirurgia e a matriz resposta foi construída usando dados da análise seminal pré e pós varicocelectomia (Estudo III) e de sucesso gestacional em até 24 meses pós cirurgia (Estudo IV). As amostras dos voluntários para os quais não tínhamos informações pós-cirurgia foram utilizadas apenas para predição. A Figura 12 apresenta a população de cada estudo.

Figura12 – Distribuição dos grupos investigados em cada estudo desenvolvido.



Fonte: Autoria Própria.

Quanto aos critérios de exclusão, foram adotados os seguintes parâmetros: evidências de infecção urogenital, presença de doenças urológicas diagnosticadas pela avaliação andrológica, defeitos genéticos, histórico de criptoquidismo, uso atual ou recente (< 12 meses) de testosterona ou qualquer esteroide anabolizante, histórico

de quimioterapia e radioterapia, histórico de trauma escrotal, histórico de cirurgia escrotal. Todos os participantes foram submetidos as anamneses andrológicas e exames físicos. A idade de puberdade foi considerada como sendo a idade que o participante identificou o aumento peniano ou do tamanho testicular, ou aparecimento de pelos pubianos. Os exames físicos foram realizados no ambiente bem iluminado e com temperatura adequada, com os participantes na posição supino e vertical. O tamanho do testículo foi mensurado usando o orquidômetro e o grau de varicocele foi classificado baseado nos critérios de Dubin e Amelar¹⁰⁴. Quanto à parceira, foi perguntado sobre idade, histórico de fertilidade no passado e/ou tratamento para fertilidade, histórico de cirurgias pélvicas ou patologias pélvicas, além de informação sobre a intensidade e regularidade do ciclo menstrual. Contudo, a averiguação da fertilidade feminina não foi realizada.

Todos os voluntários assinaram o Termo de Consentimento Livre e Esclarecido (TCLE) antes de sua inclusão no estudo. Todos os voluntários foram submetidos à análise padrão de sêmen de acordo com as diretrizes da OMS (2010)². A ultrassonografia escrotal com Doppler colorido (GE Logiq S8 GE Healthcare, Wauwaosa,WI, USA), com probe linear de alta frequência (SL 15-8MHz GE Healthcare, Wauwaosa,WI, USA) foi empregada para avaliar o tamanho testicular, diâmetro da varicocele e o retorno venoso. Para a avaliação hormonal empregou-se o doseamento séricos dos hormônios, sendo a coleta do sangue realizada por punção venosa entre as 07:00 e 11:00 da manhã, após o jejum noturno. Os parâmetros hormonais avaliados foram: testosterona total (intervalo de referência: 1,4 – 18,2 mUI/mL), estradiol (intervalo de referência: 11,6 – 41,2 pg/dL), hormônio folículo estimulante (FSH) (intervalo de referência: 206 – 1200 ng/dL), hormônio luteinizante (LH) (intervalo de referência: 1,5 – 9,3 mUI/mL), albumina e globulina de ligação aos

hormônios sexuais (SHBG) (intervalo de referência: 10 – 57 nmol/L). Os níveis séricos dos hormônios foram obtidos através do imunoensaio em fase-sólida, de quimiluminescência em tempo real, com o auxílio de analisador automático ADAVIA Centaur XP, Siemens Healthcare Diagnostics.

Quanto às dosagens da albumina (intervalo de referência 3,4 - 4,8 g/dL) foram determinadas usando o sistema automatizado (Architect c16000, Abbott Diagnostics). Os níveis de testosterona livre (intervalo de referência 49,9 – 199,9 pg/mL) foram obtidos a partir da fórmula de Vermeulen et al.¹⁰⁵. A execução desse trabalho foi aprovada pelo Comitê de Ética em Pesquisa da UFPE e do IMIP - nº 2.075.028.

Para investigar a distribuição de dados demográficos e clínicos entre os grupos, testes univariados foram realizados usando Software GraphPad Prism 6 (GraphPad Software, Inc., La Jolla, CA). O teste de Shapiro-Wilk foi usado para analisar a normalidade da distribuição dos dados. Análise de variância (ANOVA) juntamente com o teste de Tukey, quando necessário, foi usado para comparar as seguintes variáveis entre os três grupos: idade, idade do parceiro, idade da puberdade, tamanhos testiculares por exame físico, IMC, motilidade progressiva do esperma, morfologia do esperma, volume do ejaculado, testosterona total, testosterona livre, estradiol, LH e tamanhos testiculares por USGD. O teste de Kruskal-Wallis acoplado com o teste de Dunn, quando necessário, foi usado para comparar as seguintes variáveis entre os três grupos: concentração de esperma, contagem total de esperma, contagem progressiva total de esperma, FSH e tamanhos de veia testicular. O teste exato de Fisher foi usado para comparar grau de varicocele e a proporção de varicocele bilateral entre os grupos VF e VI. Um valor $p < 0,05$ foi definido como o nível de significância estatística.

9.2 PREPARAÇÃO DAS AMOSTRAS DE SÊMEN, ANÁLISE POR RMN DE ^1H E TRATAMENTO ESTATÍSTICO PARA ANÁLISES MULTIVARIADAS

As amostras de sêmen foram centrifugadas a 3000 G, por 15 minutos, para separar os espermatozoides do plasma seminal. Após este processamento, as amostras contendo apenas o líquido seminal foram mantidas congeladas a -40°C até serem analisadas por RMN de ^1H . Após serem descongeladas, foram adicionados 400 μL da amostra de soro do sêmen em um tubo de RMN, com 5 mm de diâmetro interno, em seguida, foi acrescentada 200 μL de água deuterada (D_2O) ao tubo e a mistura foi homogeneizada. Os espectros de RMN de ^1H foram obtidos utilizando uma sequência de pulsos CPMG (Carr-Purcell-Meiboom-Gill) com pré-saturação (presat) do sinal da água, no equipamento Agilent VNMRS 400, operando a 400 MHz e a temperatura de 298 K. Os parâmetros experimentais utilizados foram os seguintes: tempo de aquisição de 1,704 s, tempo de pré-saturação igual a 2,0 s, pulso de RF de 90° igual a 13,2 μs , janela espectral de 4,8 kHz, número de ciclos de 88, tau igual a 0,0004s e bigtau igual a 0,07s e 128 transientes. Cada espectro foi adquirido utilizando um tempo total de 10,0 minutos e 4 segundos. Os espectros foram processados utilizando *line broadening* de 0,3 Hz, a linha de base e distorções de fase foram ajustadas manualmente. O sinal do grupo metileno do nitrato (δ 2.65 ppm) foi atribuído como deslocamento químico referência. Um segundo processamento foi efetuado com a assistência do software MestReNova 9.0 em que foi novamente reavaliada e ajustada automaticamente a linha de base, pelo método de polinômios de Bernstein, e a fase dos espectros foi ajustada manualmente. Para a construção da matriz de dados, utilizamos apenas a região do espectro entre δ (ppm) 0,69 e 8,01, que foi dividida em regiões (bins) de 0,04 ppm. A região entre os δ 4,5 – 5,5 foi excluída do espectro, visto que é entre tais deslocamentos que se encontra o sinal da água, que mesmo após a

etapa de pré-saturação ainda dificulta a análise. Dessa forma foi obtida uma matriz com 80 linhas (amostras) e 159 variáveis (bins do RMN de ^1H mais a variável de classe) que posteriormente foi submetida a análise estatística multivariada. De acordo com o estudo realizado, o número de linhas da matriz foi alterado, de forma que nos estudos II, III e IV tivemos 56, 29 e 23 linhas (amostras).

O tratamento quimiométrico dos dados espectrais foi realizado partindo da análise exploratória dos mesmos, através da Análise de Componentes Principais (PCA). Também foram empregados formalismos de Análise Discriminante (DA), como: Análise Discriminante Linear usando Algoritmo Genético (GA-LDA) e lambda de Wilks (LDA), como ferramenta de seleção de variáveis; Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) e Análise Discriminante por Projeções Ortogonais à estrutura latente (OPLS-DA). PCA, PLS-DA e o OPLS-DA foram realizadas usando a plataforma on-line disponibilizada para estudos metabonômicos, denominada de MetaboAnalyst 5.0, enquanto os modelos determinados pela GA-LDA foram obtidos a partir do MATLAB®2010^a. Os modelos usando o formalismo LDA foram construídos usando o software Statistica 12. Para a validação do modelo OPLS-DA, foi utilizado o teste de permutação, fazendo 2000 permutações. O LOOCV (*Leave-One-Out Cross Validation*) foi usado para determinar os valores de Q², R² e a exatidão de classificação dos modelos.

O formalismo empregado para construção dos modelos GA-LDA consiste na divisão das amostras em dois conjuntos: Conjunto treinamento com 70% das amostras e o conjunto teste com 30% restante das amostras. O mecanismo de seleção das amostras foi o Algoritmo de Kennard Stone. As amostras, individualmente, foram normalizadas pela média, usando a expressão $\frac{(x_i - \bar{x})}{\sigma}$, onde x_i é a área de integração para a variável i , \bar{x} é a média das áreas de integração para

aquela amostra e σ é o desvio-padrão; na sequência, cada variável foi auto escalonada, usando expressão semelhante.

Os modelos desenvolvidos pela LDA, no estudo III, usaram uma matriz de entrada contendo 43 amostras sendo 29 delas tinha o diagnóstico fechado de melhora ou não dos padrões seminais pós-varicocelectomia. Desses, 18 (62,1%) apresentaram melhora, enquanto 11 (37,9%) não foi verificado melhora no espermograma. O modelo metabonômico foi construído usando 5 variáveis, que foram selecionadas empregando o lambda de Wilks. Esse modelo foi validado usando *Leave-One-Out-Cross Validation* (LOOCV) e teste-*F* para a função discriminante que foi realizado usando 22 e 5 graus de liberdade. Após a etapa de validação, calculou-se as seguintes figura de mérito: sensibilidade, especificidade, valor preditivo positivo (VPP), valor preditivo negativo (VPN) e exatidão do modelo. O modelo também apresentou uma predição de classificação para os voluntários que não tinham até o momento o diagnóstico definido. Estratégia semelhante foi adotada no estudo IV, sendo que a matriz utilizada na modelagem tem 23 amostras, sendo que 10 (43,48%) voluntários conseguiram engravidar a companheira, enquanto 13 (56,52%) não conseguiram sucesso gestacional em até 24 meses após a cirurgia de varicocelectomia. Nos estudos III e IV, a matriz-resposta utilizou dados pós-cirúrgicos – melhora dos parâmetros seminais (Estudo III) ou sucesso gestacional (Estudo IV). Considerando o universo de 43 amostras utilizadas nos estudos III e IV, tinha-se pelo menos uma informação pós-cirúrgica de 29 pacientes, enquanto para 14 voluntários nenhuma informação foi disponibilizada.

10 RESULTADOS E DISCUSSÃO

Como descrito na seção Materiais e Métodos, foram desenvolvidos quatro estudos: (I) Diagnóstico de Varicocele; (II) Diagnóstico de Infertilidade associada à Varicocele; (III) Diagnóstico de Melhoria Pós-Varicocelectomia; e (IV) Avaliação de Sucesso Gestacional. Nesse sentido, a apresentação e discussão desses resultados será feita considerando essa distribuição. Será feita uma breve exposição do problema, seguida da apresentação dos achados da investigação.

10.1 DIAGNÓSTICO DE VARICOCELE & DIAGNÓSTICO DE INFERTILIDADE ASSOCIADA À VARICOCELE (ESTUDOS I E II)

10.1.1 Resultados demográficos clínico-laboratoriais

Inicialmente, foram avaliadas estatisticamente os parâmetros clínico-laboratoriais de cada um dos pacientes incluídos na pesquisa com intenção de averiguar se existiria diferenças amostrais importantes que poderiam influenciar nos modelos metabonômicos. Para tal, foram executados testes estatísticos a partir dos dados demográficos clínico-laboratoriais dos pacientes inseridos no estudo (Tabela 5)

Tabela 6 – Características clínicas e laboratoriais dos participantes dos estudos I e II.

	Controle	Varicocele Fértil	Varicocele Infértil	Valor de <i>p</i>
Demografia e Exame Físico				
Número de voluntários	24	21	35	
Idade (ano)	34,2 ± 4,3	33,7 ± 4,9	32,0 ± 6,1	0,22*
Idade da parceira (ano)	29,2 ± 3,8	32,3 ± 4,4	30,7 ± 5,1	0,14*
Idade da puberdade (ano)	14,2 ± 1,5	14,8 ± 1,8	13,8 ± 1,4	0,08*
Volume do testículo direito (mL)	22,7 ± 5,0 ^a	19,0 ± 3,5 ^a	15,7 ± 5,4 ^b	<0,05*
Volume do testículo esquerdo (mL)	22,0 ± 5,6 ^a	17,4 ± 5,0 ^b	12,5 ± 5,6 ^c	<0,05*
IMC (kg/m ²)	28,0 ± 4,0	26,3 ± 4,2	26,2 ± 5,0	0,23*
Varicocele		67%	48%	
Testículo direito (%)				
	Grau 1			
	Grau 2	N/A	33%	0.53 [†]
	Grau 3		43%	
	Grau 1		0%	
	Grau 2		9%	
Varicocele		24%	17%	
Testículo esquerdo (%)				
	Grau 1	N/A	38%	0.38 [†]
	Grau 2		26%	
	Grau 3		38%	
Varicocele bilateral (%)	NA	14%	60%	<0,05 [†]
Ultrassonografia com Doppler				
Volume do testículo direito (mL)	16,0 ± 3,9 ^a	14,7 ± 4,7 ^{a,b}	12,5 ± 4,5 ^b	<0,05*
Volume do testículo esquerdo (mL)	14,6 ± 4,3 ^a	11,8 ± 4,3 ^{a,b}	11,1 ± 4,3 ^b	<0,05*
Volume da veia testicular direita (cm), mediana ± IQR	1,9 ± 0,7 ^a	2,0 ± 0,9 ^{a,b}	2,5 ± 0,8 ^b	<0,05 [§]
Volume da veia testicular esquerda (cm), mediana ± IQR	2,0 ± 0,7 ^a	3,2 ± 1,0 ^b	3,8 ± 1,3 ^b	<0,05 [§]
Hormônios Sexuais				
Testosterona (ng/dL)	345 ± 135 ^a	422 ± 158 ^{a,b}	446 ± 179 ^b	<0,05*
Testosterona livre (ng/dL)	7,5 ± 2,3	8,4 ± 2,6	8,8 ± 3,4	0,25*
Estradiol (pg/mL)	33,9 ± 12,0	39,5 ± 12,3	31,7 ± 12,3	0,16*
FSH (mUI/mL), mediana ± IQR	3,9 ± 2,4 ^a	3,6 ± 2,3 ^a	6,1 ± 4,9 ^b	<0,05 [§]
LH (mUI/mL)	3,5 ± 1,3 ^a	3,9 ± 1,6 ^{a,b}	4,9 ± 2,3 ^b	<0,05*
Análise Seminal				
Concentração espermática (10 ⁶ /mL), mediana ± IQR	38,9 ± 05,4 ^a	40,0 ± 56,9 ^a	5,0 ± 12,6 ^b	<0,05 [§]
Contagem total de espermatozoides (10 ⁶ /mL), mediana ± IQR	133 ± 312 ^a	105 ± 136 ^a	14 ± 31 ^b	<0,05 [§]
Motilidade progressiva (%)	65 ± 20 ^a	55 ± 22 ^a	30 ± 15 ^b	<0,05*
Contagem total de espermatozoides progressivos (10 ⁶ /mL), mediana ± IQR	74 ± 241 ^a	55 ± 89 ^a	2 ± 10 ^b	<0,05 [§]
pH, mediana ± IQR	8,0 ± 0,5	7,5 ± 1,0	7,3 ± 1,0	NA [¶]
Morfologia (%)	10,0 ± 6,1	8,6 ± 3,5	7,8 ± 4,5	0,30*
Volume (mL)	4,0 ± 1,5	3,0 ± 1,1	3,2 ± 1,6	0,06*

Quando não mencionados, os dados estão apresentados como média ± desvio padrão.

IMC = Índice de massa corpórea; FSH = Hormônio Folículo Estimulante; LH = Hormônio Luteinizante;

IQR = Intervalo Interquartil * Análise de Variância; [†] Teste exato de Fisher; [§] Teste de Kruskal-Wallis.

^{a,b,c} Foram usados para identificar onde ocorreu a diferença estatisticamente significativa (*p* < 0,05)

[¶] Teste não aplicável devido à baixa precisão da ferramenta de medição de pH (tiras testes)

Fonte: Autoria Própria.

No que diz respeito a idade, foram considerados tanto as dos homens quanto as de suas parceiras, não sendo observado diferenças entre os três grupos. De modo

similar, a idade do início da puberdade nos homens não mostrou ser discriminante entre os grupos em questão. Os pacientes do grupo VI apresentaram testículos menores comparados aos grupos VF e C com significância estatística ($p < 0,05$). O grau da varicocele também não foi caracterizado como um discriminador da fertilidade visto que os grupos VF e VI não apresentaram diferenças tanto no lado esquerdo quanto no direito. Contudo, a manifestação bilateral da doença teve uma alta proporção de homens do grupo VI, quadro semelhante também foi identificado na USGD em que os pacientes VI exibiram diâmetros venoso maiores em ambos os lados.

Em relação às análises do espermograma não foram identificadas diferenças entre os grupos C e VF. Contudo, os homens com varicocele e inférteis foram evidenciados as menores concentrações e contagens de espermatozoides assim como também da motilidade e o número total de espermatozoides progressivos quando comparado aos outros dois casos. Quanto aos efeitos hormonais, observa-se que o grupo VI exibiu níveis do hormônio FSH mais elevados do que o C e VF grupos, e níveis mais elevados de LH do que o grupo C. Interessantemente, o grupo C tinha níveis mais baixos de testosterona total sérica do que o grupo VI, mas os níveis de testosterona livre não foram diferentes entre os grupos.

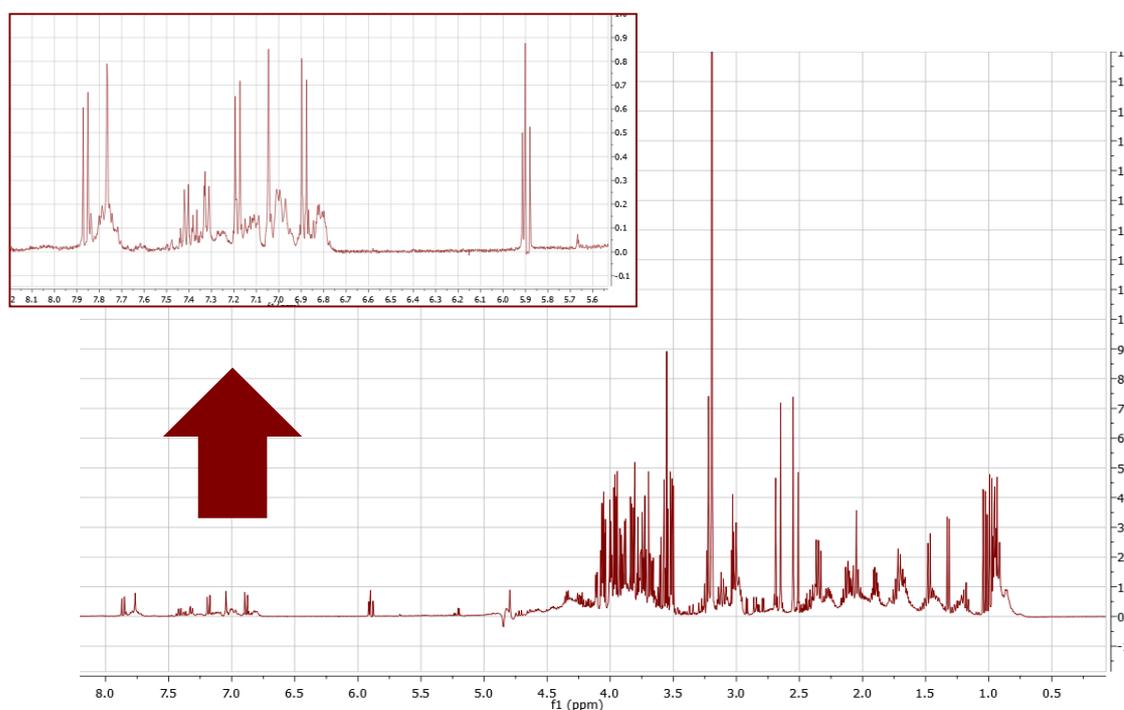
O presente trabalho evidenciou a existência de padrões de fertilidade díspares para os homens que tem varicocele. O grupo VI apresentou alterações clínicas devido à varicocele, como diminuição do volume testicular, diminuição da concentração de espermatozoides, diminuição do total de espermatozoides contagem, diminuição da motilidade total e progressiva e diminuição do volume do sêmen, quando comparados com controles saudáveis. Embora que foram identificados os níveis maiores de testosteronas, LH e FSH no grupo VI quando comparado ao grupo controle,

clínicamente esta informação não se apresenta relevante, visto que os valores obtidos estão dentro dos intervalos de referência atribuídos para pacientes normais.

10.1.2 Diagnóstico da varicocele (estudo I)

A Figura 13 apresenta um exemplo dos espectros que foram obtidos na etapa de aquisição dos dados. Não existem diferenças visuais entre os grupos avaliados.

Figura 13 – Espectro de RMN de ^1H (400 MHz, D_2O , CPMG-Presat) de uma amostra de soro de sêmen.



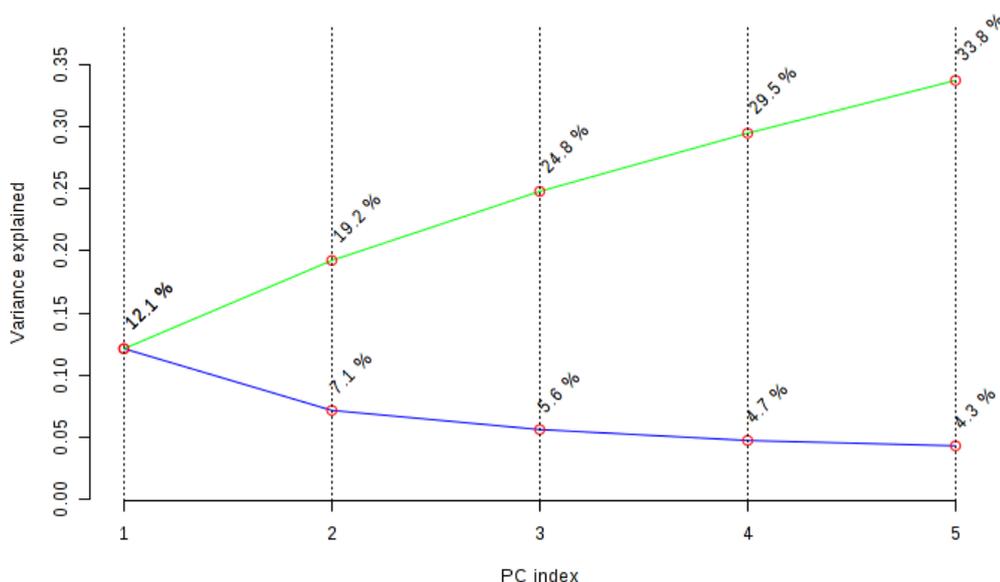
Fonte: Autoria própria.

Com o auxílio da plataforma MetaboAnalytic 5.0, filtramos os dados a fim de identificar e remover as variáveis que, provavelmente, não serão úteis na modelagem dos dados. O MetaboAnalytic 5.0 emprega um filtro de 5% do conjunto de dados com menos de 250 variáveis, o qual foi utilizado neste estudo visto que nosso conjunto de dados possui 159 variáveis espectrais. Esta etapa é interessante e recomendado nos estudos metabolômicos e metabonômicos não direcionados (*untargeted*) visto que

possui um grande número de variáveis das quais algumas são ruídos na linha de base.⁹⁵ Em seguida foi realizado o pré-processamento nas variáveis aplicando o autoescalamento, enquanto as amostras foram centradas na média.

A partir da exploração não supervisionada dos dados não foi observado a existência de distribuições naturais entre os grupos no espaço multidimensional. A PCA foi aplicada usando a matriz de covariância. A Figura 14 apresenta o *Scree plot*, indicando que as cinco primeiras componentes principais explicam apenas 33,8% da variância contida no conjunto de dados.

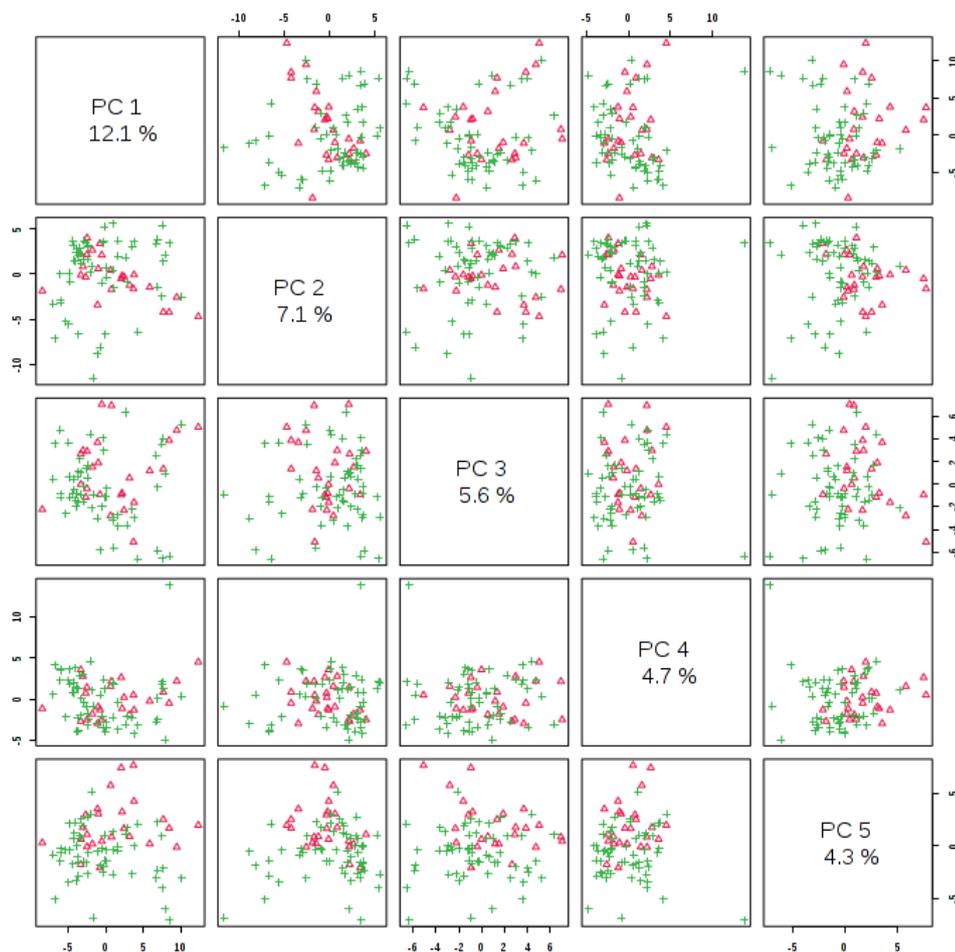
Figura 14 – PCA. Em azul, a variância explicada por cada PC; e em verde, a variância acumulada.



Fonte: Autoria própria.

A Figura 15 apresenta os gráficos de escores da PCA, usando PC1 *versus* PC2, PC3, PC4 e PC5. Observa-se que não há discriminação natural entre os grupos de interesse. Isso indica que a variância devida ao status bioquímico das amostras é muito pequena, sendo recomendado o uso de métodos supervisionados para a modelagem.

Figura 15 – Gráficos de escores da PCA realizada usando a matriz de covariância. Em verde, amostras do grupo controle; em vermelho, amostras dos pacientes diagnosticados com varicocele.

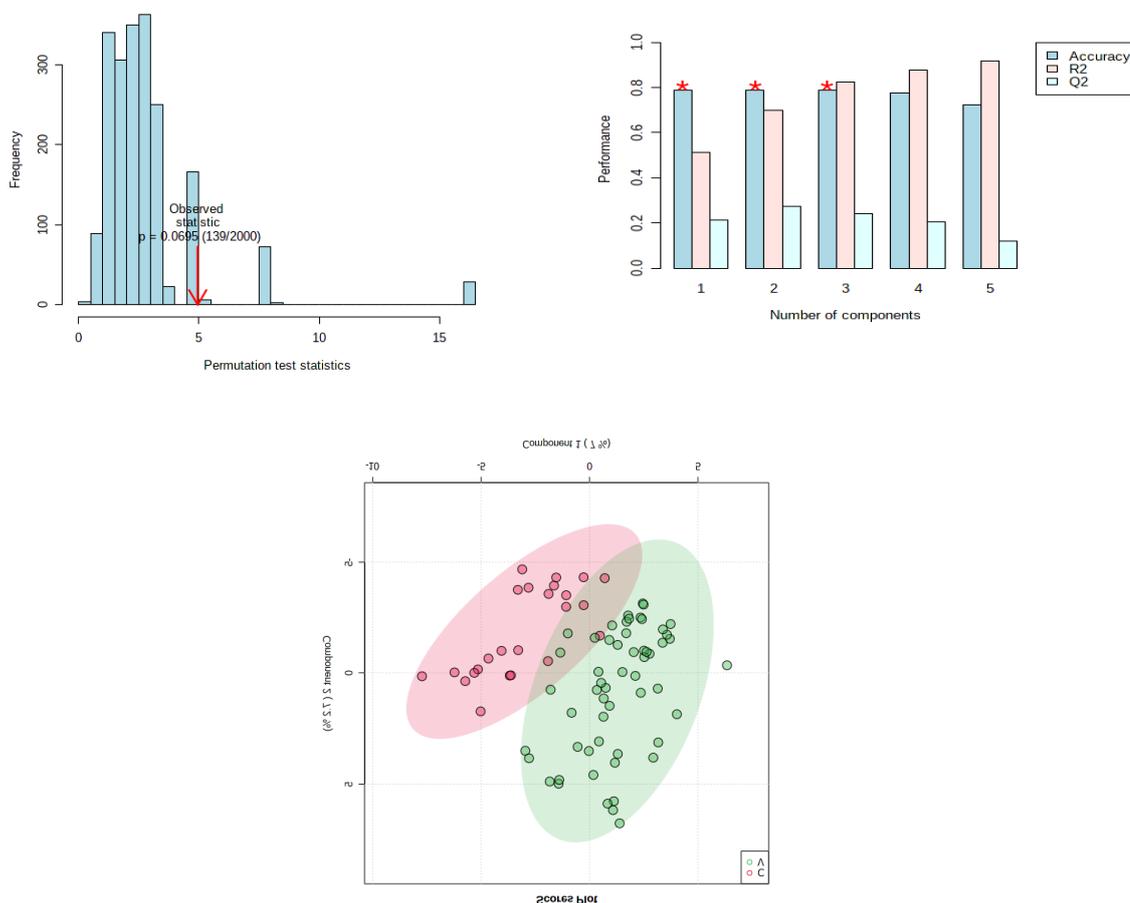


Fonte: Autoria própria.

A primeira modelagem supervisionada empregada foi a PLS-DA. Neste cenário, foi selecionado o filtro do intervalo interquartil. Como podemos observar na Figura 16, o gráfico de escore entre a primeira e segunda componente latente indicou a existência de separação entre os grupos controle (C) e varicocele (V). No teste da validação cruzada foi observado uma exatidão de aproximadamente 80%, com valores positivos tanto de R2 quanto de Q2. Este último parâmetro é uma estimativa da capacidade preditiva do modelo sendo calculado durante a validação cruzada em

que cada amostra predita pelo modelo é comparada à classificação original da mesma, deste modo é calculado a soma dos quadrados dos erros⁹⁶.

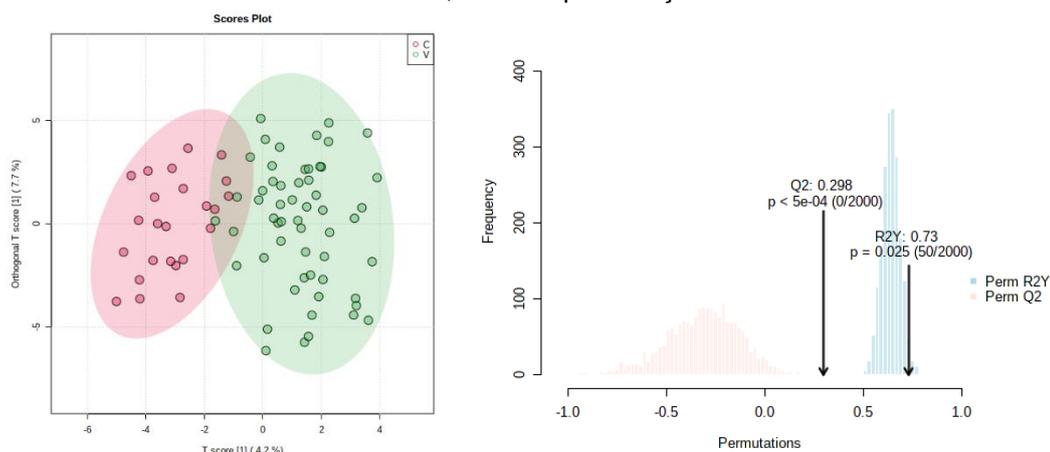
Figura 16 – Modelagem PLS-DA Controle versus Varicocele. Na parte superior à direita, teste de permutação; à esquerda, figuras de mérito; inferior, gráfico de escores entre as duas primeiras variáveis latentes.



Fonte: Autoria própria.

Entretanto, ao ser realizado o teste de permutação, o valor de p calculado foi de 0,065, sugerindo que o modelo pode estar superajustado. Então, optamos por usar o formalismo OPLS-DA. Foram realizados estudos quanto aos diferentes tipos de filtro que poderíamos aplicar em nossa matriz de dados, a fim de obter melhores condições para a modelagem. Foi escolhido o filtro baseado no valor da intensidade média. A Figura 17 apresenta o resultado da OPLS-DA.

Figura 17 – Resultado da OPLS-DA Controle versus Varicocele. À esquerda, gráfico de escores; à direita, teste de permutação.



Fonte: Autoria própria.

Pode-se observar no gráfico de escore formado entre a primeira componente e a sua ortogonal que de fato existe uma tendência de separação entre os dois grupos, como já tinha sido evidenciado no PLS-DA. Entretanto, o OPLSDA apresentou dados satisfatórios no teste de permutações com um p de 0,025 (50/2000). Quanto ao valor de exatidão foi calculado um valor de 90%. Outras figuras de mérito, que foram consideradas como relevantes para fins de discriminação entre os dois grupos, considerando o intervalo de confiança de 95%, são apresentadas na Tabela 7.

Tabela 7 – Matriz de Contingência - Modelo OPLS-DA. Grupo controle (24 amostras) versus grupo varicocele (56 amostras). Desempenho baseado após a Validação Cruzada.

		Diagnóstico Clínico		
		Controle	Varicocele	p
Modelo OPLS-DA	Controle	53	4	0,025
	Varicocele	3	20	

Exatidão 91,2%, Sensibilidade 83,3%, Especificidade 94,6%, Valor preditivo positivo 86,9%, Valor preditivo negativo 93,0%

Fonte: Autoria própria.

É conveniente ressaltar que, usualmente, o diagnóstico da varicocele pode ser realizado mediante o exame físico através da técnica de Valsava que, dependendo do

grau, pode ter uma significativa variação intraobservador e interobservador. Frequentemente, os médicos associam este exame físico à ultrassonografia (US) da região escrotal para fechar o diagnóstico de varicocele. De fato, a US é uma ferramenta amplamente utilizada no diagnóstico da varicocele, contudo, as informações podem representar uma limitada relevância clínica, visto que há uma baixa correlação com parâmetros de sêmen anormais o que, em última análise, melhor refletem a condição da espermatogênese prejudicada⁹⁷. Desta forma, reiteramos que a abordagem metabonômica poderá se constituir em uma ferramenta para o inquérito clínico dos pacientes com suspeita de varicocele.

Ainda neste estudo, um último formalismo supervisionado foi aplicado, o GA-LDA, visando melhorar o poder preditivo do modelo. As figuras de mérito para o conjunto treinamento e para o modelo de validação (conjunto teste) podem ser visualizados nas Tabelas 8 e 9, respectivamente.

Tabela 8 – Matriz de Contingência - Modelo GA-LDA. Conjunto treinamento: grupo controle (17 amostras) versus grupo varicocele (39 amostras).

Modelo GA-LDA	Diagnóstico Clínico		Valor de <i>p</i>
	Controle	Varicocele	
Controle	16	0	< 0,05
Varicocele	1	39	

Exatidão = 98,2%; Sensibilidade = 100%; Especificidade = 94,1%; VPP = 97,5%; VPN= 100%

Fonte: Autoria própria.

Tabela 9 – Matriz de Contingência - Modelo GA-LDA. Conjunto Teste: grupo controle (7 amostras) versus grupo varicocele (17 amostras).

Modelo GA-LDA	Diagnóstico Clínico		Valor de <i>p</i>
	Controle	Varicocele	
Controle	6	0	< 0,05
Varicocele	1	17	

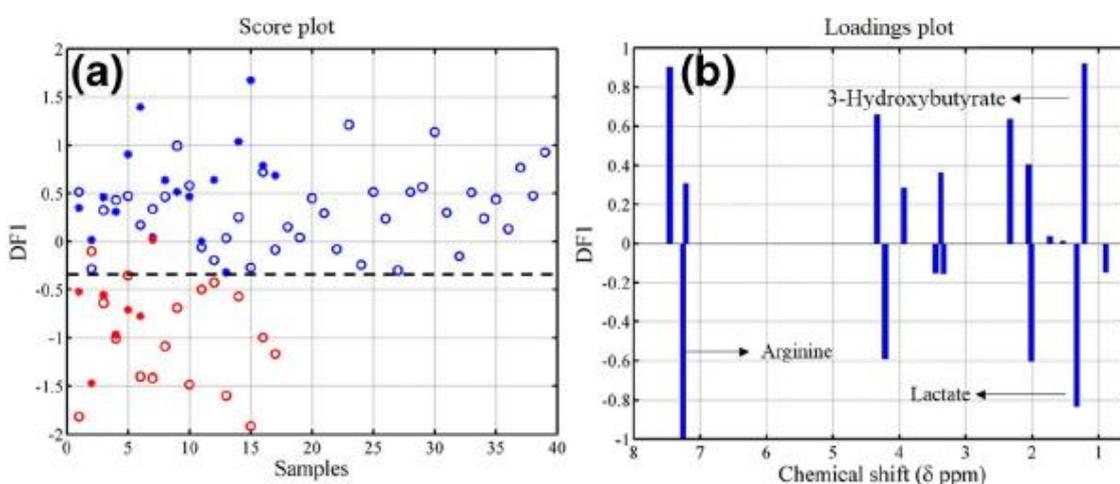
Exatidão = 95,8%; Sensibilidade = 100%; Especificidade = 85,7%; VPP = 94,4%; VPN= 100%

Fonte: Autoria própria.

Na etapa de validação do modelo, foi identificado apenas um caso de falso-positivo, isto é, apenas uma amostra do grupo controle foi classificada como tendo varicocele, ou seja, o GA-LDA demonstrou uma exatidão de 95,8%. Outro resultado importante a ser destacado, é que essa modelagem conseguiu identificar todos os pacientes com varicocele, isto confere uma sensibilidade de 100%. Sendo assim, o formalismo GA-LDA apresenta-se bastante útil para testes de triagem na identificação de pacientes com varicocele.

Na Figura 18 (a), tem-se o gráfico de escores na função discriminante. As bolas vermelhas (○) representam as amostras do grupo controle, na etapa de treinamento; enquanto os pontos vermelhos (▪) são as amostras do grupo Controle usadas na etapa de validação (teste). Do mesmo modo, as bolas azuis (○) são as amostras do grupo Varicocele usadas na etapa de treinamento, ao passo que os pontos azuis (▪) compõem as amostras desse grupo usadas na etapa de validação (teste).

Figura 18 – Resultados GA-LDA Controle versus Varicocele. À esquerda, gráfico de escores; à direita, a distribuição das variáveis que melhor discriminaram o modelo.



Fonte: Autoria própria.

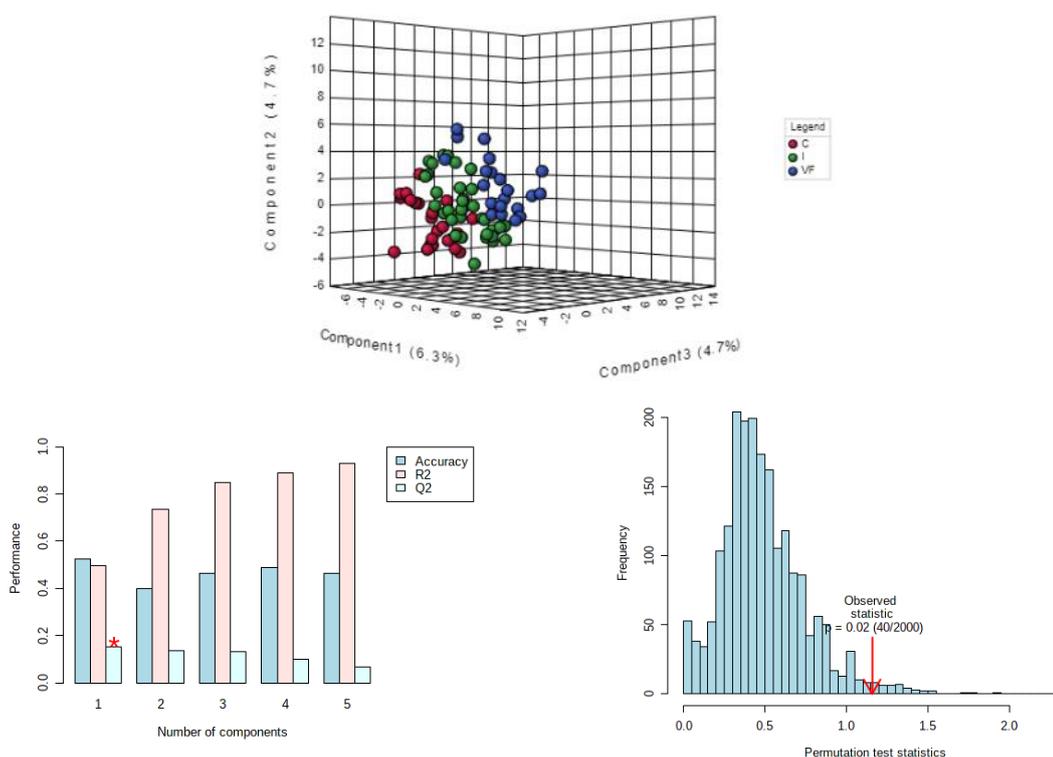
É interessante apresentar que importância de cada variável para o modelo GALDA é apontada na Figura 18 (b) pela intensidade das barras. Na figura b, a direção das barras indica se a variável tem peso positivo ou negativo na combinação linear da função discriminante. Os metabólitos que tiveram a maior influência neste modelo foram 3-hidroxibutirato, lactato e arginina.

A capacidade destes modelos metabonômicos para discriminar homens com varicocele dos controles saudáveis com alta precisão corrobora a relação estreita entre metaboloma e fenótipos. Além disso, alguns achados estão em conformidade com os resultados proteômicos avaliados nas mudanças do sêmen dos adolescentes e homens saudáveis e com varicocele. ZYLBERSZTEJN et al (2013) e DEL GIUDICE et al. (2016) compararam o material seminal de adolescentes sem varicocele (grupo controle) com o sêmen dos adolescentes com varicocele os quais apresentaram alta expressão de proteínas envolvidas na resposta imunológica e inflamatória, mesmo quando não foram observadas mudanças nos parâmetros seminais.^{98,99} A hipótese que estamos investigando é que a presença de alterações no metaboloma e no proteoma seminal de indivíduos com varicocele, mesmo sem apresentar anormalidades aparentes no sêmen, sugere que a varicocele é uma condição com aspectos heterogêneos e que as implicações vão além do impacto sobre os parâmetros clássicos avaliados no espermograma.

10.1.3 Infertilidade associada à varicocele (estudo II)

Com a mesma matriz de dados espectrais do sêmen utilizada no estudo I, foram construídos modelos metabonômicos para a averiguação da infertilidade associada à varicocele. Para tal, as amostras foram subdivididas em três categorias: Controle (C), Varicocele Fértil (VF) e Varicocele Infértil (VI). A matriz dos dados foi devidamente pré-processada empregando, inicialmente, o filtro de intensidade do valor médio e posteriormente as amostras foram normalizadas pela média e as variáveis foram autoescaladas. Assim como no Estudo I, a análise exploratória não mostrou agrupamento natural nas classes de interesse e, portanto, partimos para os métodos supervisionados. A Figura 19 exibe o resultado da PLS-DA, considerando os três grupos investigados. Usando três variáveis latentes, tem-se uma discriminação entre os grupos, porém com baixa capacidade preditiva.

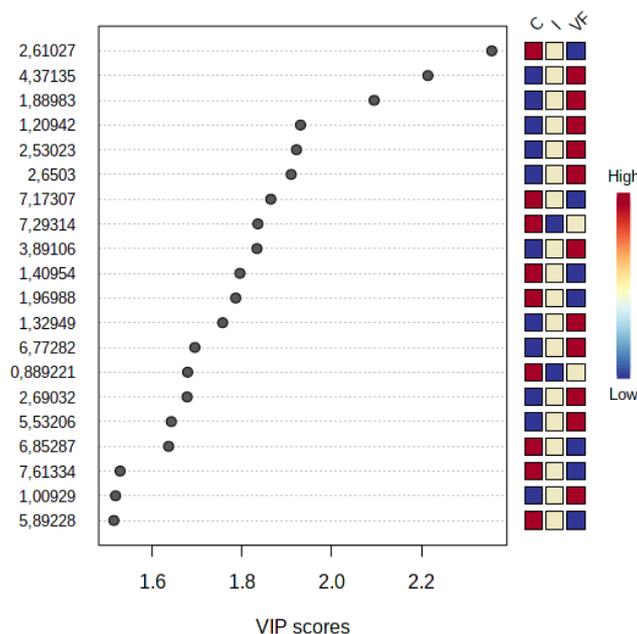
Figura 19 – Resultado da PLS-DA Controle versus Varicocele Fértil versus Varicocele Infértil. Superior, gráfico de escores; à esquerda, figuras de mérito; e à direita, teste de permutação.



Fonte: Autoria própria.

O resultado obtido foi bastante interessante não apenas pela clara tendência de separação entre os grupos no gráfico dos escores, mas também pelo desempenho apresentado nos testes de validação. Considerando o número de amostras por grupo, foi adotado a validação cruzada pelo método *leave-one-out* (LOOCV) no qual uma amostra é retirada do conjunto dos dados e em seguida é construído um modelo e a amostra excluída é testada. No teste de permutação o valor de p foi igual a 0,02 (40/2000), refletindo bom desempenho estatístico (Figura 19). Outro achado importante, determinado a partir da PLSDA, é a identificação das variáveis que são mais importantes para a discriminação dos diferentes grupos. Esta informação é extraída a partir do gráfico dos VIP (Variable Importance in the Projection) escores que é apresentada na Figura 20.

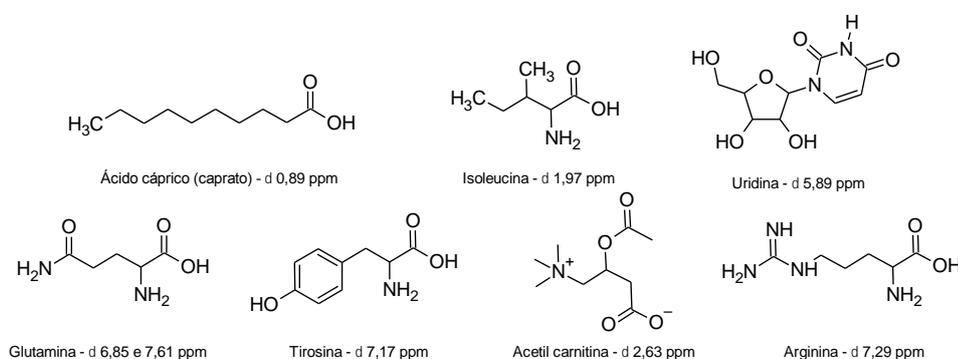
Figura 20 – Escores das Variáveis Importantes na Projeção (VIP scores).



Fonte: Autoria própria.

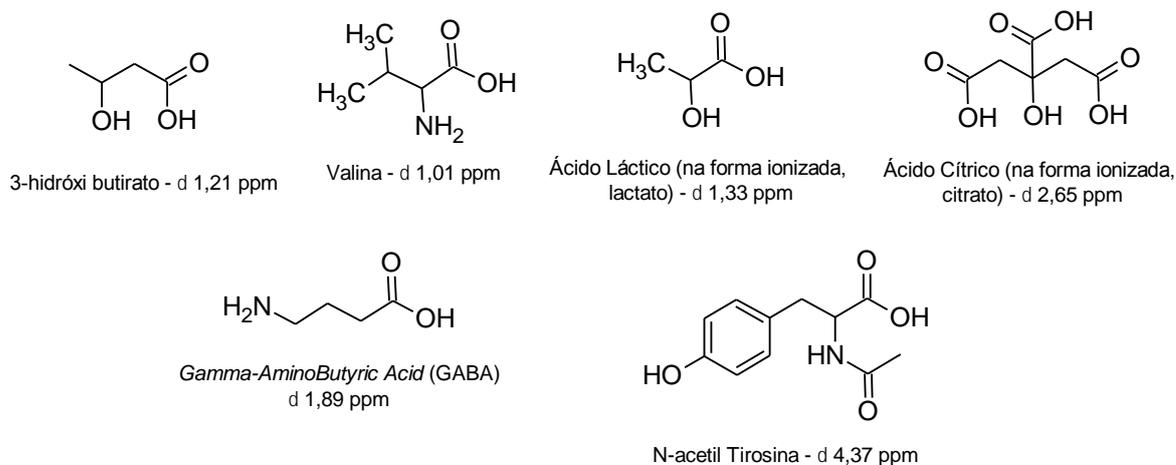
Baseado no escore VIP, foram identificados 13 descolamentos químicos importantes para a discriminação dos grupos. Para a identificação dos deslocamentos químicos das variáveis mais importantes para a discriminação dos grupos com os respectivos metabólitos foi empregado o *Human metabolome database (HMDB)*¹⁰⁰ e o trabalho de Paiva et al. (2015)²⁶. Infelizmente para 7 deslocamentos químicos não foram identificadas substâncias endógenas associadas aos seguintes deslocamentos (ppm), são eles: 1,41, 2,53, 2,69, 5,53 e 6,77. Curiosamente, no gráfico de escores, os extremos são ocupados pelas amostras dos grupos Controle e Varicocele Fértil, com as amostras do grupo Varicocele Infértil estando entre os dois grupos. Isso indica que as principais diferenças nos níveis séricos dos metabólitos devem ser observadas entre Controle e Varicocele Fértil. Entre os metabólitos identificados, foram observados níveis séricos mais elevados no grupo Controle para 08 (oito) metabólitos: caprato, isoleucina, O-acetil carnitina, uridina, glutamina, tirosina, arginina e glutamina (Figura 21). Por outro lado, foram identificados níveis mais elevados no grupo Varicocele Fértil para 06 (seis) metabólitos: valina, 3-hidróxi-butilato, lactato, GABA, nitrato, N-acetiltirosina. O nível sérico de glicosídeos também estava elevado no grupo Varicocele Fértil (Figura 22).

Figura 21 – Metabólitos com níveis séricos mais elevados no grupo Controle.



Fonte: Autoria própria.

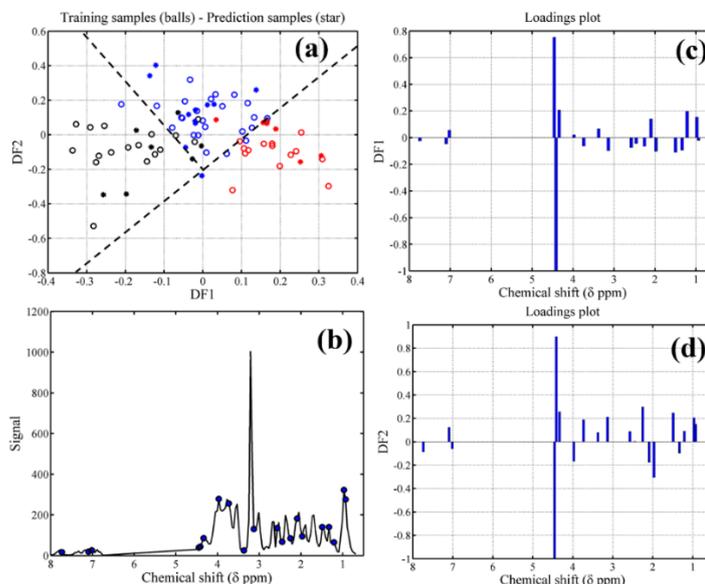
Figura 22 – Metabólitos com níveis séricos mais elevados no grupo Varicocele Fértil.



Fonte: Autoria própria.

Modelo semelhante foi construído, usando o formalismo GA-LDA, pois poderíamos separar o conjunto de dados em grupos de treinamento e de teste, uma vez que a validação do modelo PLS-DA foi realizada usando LOOCV. A Figura 23 apresenta o resultado da GA-LDA, considerando a construção de duas funções discriminantes (DF1 e DF2).

Figura 23 – Resultado da GA-LDA para discriminar varicocele infértil (azul), varicocele fértil (vermelho), grupos de controle (preto). Amostras de treinamento (bolas) e amostras de teste (pontos). Em (b) tem-se as regiões selecionadas, enquanto (c) e (d) apresentam o gráfico de pesos para cada função discriminante.



Fonte: Autoria própria.

Este modelo foi criado usando 56 amostras (conjunto de treinamento) e validado usando 24 amostras (conjunto de teste). A importância de cada variável no modelo pode ser visualizada no gráfico de pesos (*loadings*) e cada uma delas pode ser identificada nas suas respectivas funções discriminantes. Como são três grupos então o GA-LDA constrói duas funções discriminantes (DF1 e DF2). A Tabela 10 apresenta a matriz de contingência para essa modelagem.

Tabela 10 – Matriz de Contingência para o modelo usando GA-LDA – Controle *versus* Varicocele Fértil *versus* Varicocele Infértil. Conjunto de treinamento, com 17, 15 e 24 voluntários, respectivamente.

		Diagnóstico Clínico			<i>p</i>
		Controle ©	Varicocele Fértil (VF)	Varicocele Infértil (VI)	
Modelo GA-LDA	Controle	14	0	1	< 0,05
	Varicocele Fértil (VF)	0	15	2	
	Varicocele Infértil (VI)	3	0	21	

Fonte: Autoria própria.

O modelo classifica corretamente 50 amostras, resultando numa exatidão de 89,3%. Para o cálculo de sensibilidade, especificidade, valor preditivo positivo e valor preditivo negativo, vamos considerar o quadro de fertilidade. Então, temos três casos classificados como Falso-Positivo e 3 casos como Falso-Negativo. Logo, tem-se 90,6% de sensibilidade e 87,5% de especificidade. Valor preditivo positivo igual a 87,5% e valor preditivo negativo igual a 90,6%.

Em seguida, o modelo GA-LDA foi validado com base nas amostras testes. A Tabela 11 apresenta esses dados resumidamente. No teste, a exatidão foi igual a 83,3% (20 acertos em 24 casos). Os valores de especificidade e sensibilidade foram iguais a 76,9% e 90,9%, respectivamente. Esses valores discretos, em relação ao observado para o diagnóstico de varicocele, deve-se a similaridade entre as amostras dos voluntários férteis e inférteis com varicocele.

Tabela 11– Matriz de Contingência para o modelo usando GA-LDA – Controle *versus* Varicocele Fértil *versus* Varicocele Infértil. Conjunto de teste, com 7, 6 e 11 voluntários, respectivamente.

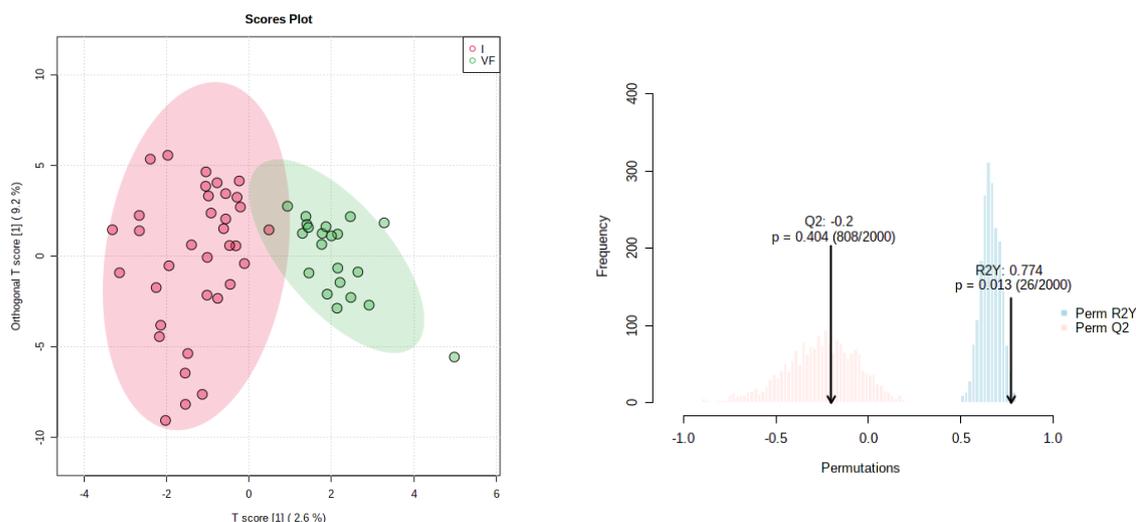
		Diagnóstico Clínico			<i>p</i>
		Controle (C)	Varicocele Fértil (VF)	Varicocele Infértil (VI)	
Modelo GA-LDA	Controle	5	0	0	
	Varicocele Fértil (VF)	0	5	1	0,01
	Varicocele Infértil (VI)	2	1	10	

Fonte: Autoria própria.

Como o foco é o diagnóstico diferencial, visando identificar os casos em que a varicocele está associada ao quadro de infertilidade, novos modelos metabonômicos foram desenvolvidos, usando apenas as amostras dos voluntários com varicocele, distribuídos em dois grupos: Varicocele Fértil (21) e Varicocele Infértil (35). Neste cenário criamos inicialmente um modelo empregando o formalismo OPLS-DA. Para este conjunto de dados aplicamos o filtro de intervalo interquartis, em seguida as

amostras foram normalizadas pela soma e por fim as variáveis foram autoescaladas. Na Figura 24 pode-se observar uma separação entre as duas classes com significância estatística no teste de permutações com valor de p igual a 0,013 (26/2000). O desempenho do modelo pode ser observado na Tabela 12.

Figura 24 – À esquerda, Gráfico de escores da OPLS-DA Varicocele Fértil versus Varicocele Infértil. Em vermelho e verde, estão as amostras dos grupos VI e VF, respectivamente. À direita, teste de permutação.



Fonte: Autoria própria.

Tabela 12 – Matriz de Contingência – Modelo OPLS-DA. Grupo Fértil (21 amostras) versus grupo Infértil (35 amostras). Desempenho baseada após LOOCV.

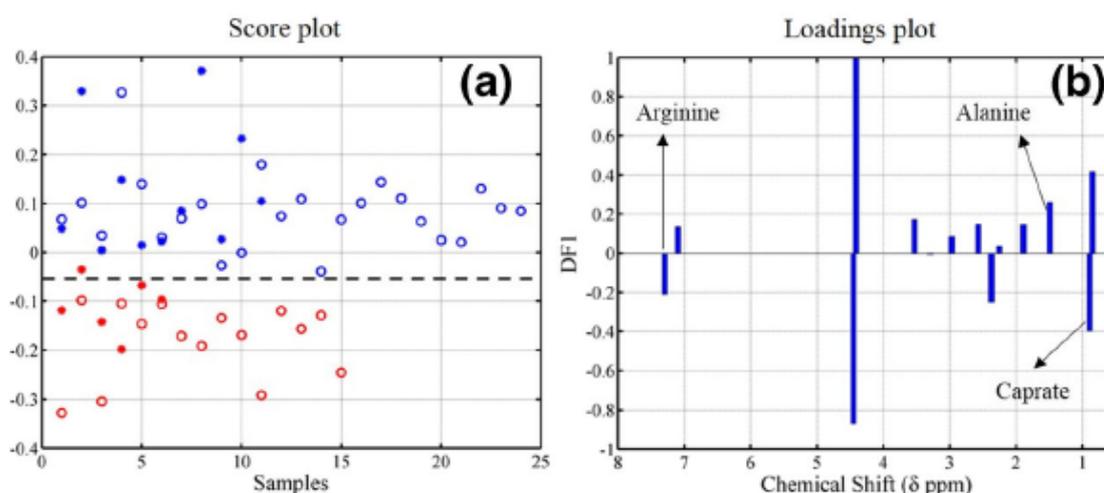
		Diagnóstico Clínico			
		Fértil	Infértil	p	
Modelo	OPLS-DA	Fértil	20	2	< 0,01
	Infértil	1	33		

Fonte: Autoria própria

Para o formalismo OPLS-DA, apesar da boa discriminação visual no gráfico de escores (Figura 24), o teste de permutação sugere que o modelo tem problemas de predição para novas amostras, em decorrência do valor de Q^2 negativo. Então, optamos pela modelagem usando o formalismo GA-LDA. Os modelos metabonômicos

foram construídos com amostras de treinamento e, em seguida, validados por um conjunto de teste. A Figura 25 apresenta a modelagem usando GA-LDA.

Figura 25 – (a) Gráfico de escores da GA-LDA. Em azul e vermelho, estão as amostras dos grupos VI e VF, respectivamente. (b) Gráfico de pesos da GA-LDA.



Fonte: Autoria própria

Como podemos observar na Figura 25 (b), o caprato, a alanina e arginina foram os metabólitos mais importantes identificados para este modelo. As figuras de mérito em relação à modelagem do conjunto treinamento são apresentadas na Tabela 13.

Tabela 13 – Matriz de Contingência - Modelo GA-LDA. Conjunto treinamento: grupo Infértil (24 amostras) versus grupo Fértil (15 amostras).

		Diagnóstico Clínico			
		Fértil	Infértil	<i>p</i>	
Modelo	GA-LDA	Fértil	15	0	< 0,05
		Infértil	0	24	

Fonte: Autoria própria.

Na etapa de validação deste modelo GA-LDA, apenas uma amostra fértil foi classificada erroneamente como infértil, dessa forma o modelo apresentou uma exatidão de 94,1% (Tabela 14).

Tabela 14 – Matriz de Contingência - Modelo GA-LDA. Conjunto teste: grupo Infértil (11 amostras) versus grupo Fértil (6 amostras).

		<i>Diagnóstico Clínico</i>		
		Fértil	Infértil	<i>p</i>
<i>Modelo GA-LDA</i>	Fértil	5	0	< 0,05
	Infértil	1	11	

Fonte: Autoria própria.

Os estudos I e II demonstraram que a hipótese admitida inicialmente tem fundamento, pois foi possível discriminar, usando a metabonômica baseada em RMN de ^1H do fluido seminal, voluntários portadores de varicocele daqueles saudáveis. Também foi possível discriminar, entre os portadores de varicocele, quais também apresentam um quadro de infertilidade. Esses resultados foram publicados no periódico *Journal of Assisted Reproduction and Genetics*, sendo o primeiro relato do emprego da estratégia ômica, visando diagnóstico diferencial de infertilidade masculina associada à varicocele¹⁰¹.

10.1.4 Predição de melhora da qualidade dos parametros seminais após a varicocelectomia (estudo III)

A terceira parte da presente tese foi concretizada a partir da construção de modelos metabonômicos com a finalidade de identificar quais os pacientes obtiveram melhorar nos parâmetros seminais após a cirurgia de varicocelectomia. A cirurgia de correção da varicocele não é um procedimento indicado para todos os pacientes acometidos pela doença.

Geralmente, o procedimento cirúrgico é indicado para pacientes que apresenta grau elevados da doença, impactando negativamente na sua rotina. No entanto, há relatos que a varicocelectomia apresenta impacto positivo no que diz respeito à qualidade dos parâmetros seminais e estes têm relação direta com a capacidade de

fecundação. Nesse sentido, um modelo que possa indicar as chances de uma intervenção cirúrgica resultar em melhora da qualidade do sêmen é salutar.

Neste estudo, foram utilizadas amostras apenas de pacientes diagnosticados com varicocele e infertilidade, que foram submetidos à cirurgia de correção da varicocele, totalizando 43 pacientes. No entanto, apenas 29 pacientes retornaram após a intervenção cirúrgica. Então, o modelo foi construído com 29 amostras, as quais foram segregadas em dois grupos, sendo 18 e 11 voluntários nos grupos Melhora (M) e Estabilidade (E) dos parâmetros seminais pós-varicocelectomia, respectivamente. Os dados seminais de cada grupo estão exibidos na Tabela 15.

Tabela 15 – Descrição dos parâmetros seminais pré e pós varicocelectomia dos voluntários participantes do estudo III.

	Limite Inferior de Referência (IC 95%) ¹	Grupo Estabilidade (n=11)			Grupo Melhora (n=18)		
		Pré	Pós	Valor de p [§]	Pré	Pós	Valor de p [§]
Concentração dos EZP (10 ⁶ /mL), mediana ± IQR	15 (12 – 16)	5.0 (0.0-11.0)	1.0 (0.1-8.0)	<0.05	3.0 (0.5-11.5)	16.0 (4.5-23.5)	<0.05
Contagem total de EZP (10 ⁶ /mL), mediana ± IQR	39 (33 – 46)	12.5 (0.0-36.0)	5.0 (0.1-30)	<0.05	8.5 (2.1-21.8)	41.0 (13.8-67.8)	<0.05
Motilidade Progressiva (%), média ± SD	32 (31 – 34)	26 (13-42)	26 (3-36)	0,46	26 (9-36)	39 (27-50)	<0.05
Contagem total de EZP com motilidade progressiva (10 ⁶ /mL), mediana ± IQR	4,8 (3,7 – 5,4)	1.9 (0.0-23.0)	0.2 (0.0-9.7)	<0.05	1.7 (0.1-7.1)	14.7 (5.5-34.2)	<0.05[§]
Morfologia (%), média ± SD	4 (3 – 4)	5.5 (4.0-10.0)	5.0 (2.5-8.5)	0.68	9.0 (4.5-11.0)	9.0 (5.0-10.0)	0.24
Volume (mL), média ± SD	1,5 (1,4 – 1,7)	3.5 (2.5-4.0)	2.5 (1.7-5.0)	0.27	2.3 (1.5-4.3)	2.5 (2.0-4.5)	0.33
pH, média ± SD	≥ 7,2	7.5 (7.0-8.0)	8.0 (7.0-8.0)	NA	7.0 (7.0-8.0)	7.5 (7.0-8.0)	NA

¹ OMS. [§]Teste Wilcoxon; NA – Não se aplica

Fonte: Autoria própria.

Os portadores de varicocele podem apresentar diferentes graus da doença. Alguns trabalhos indicam que maiores prejuízos à qualidade do sêmen podem estar relacionados tanto com o grau da doença, quanto com sua manifestação bilateral. Dentre os parâmetros seminais que são analisados durante o espermograma, a concentração de espermatozoides e a contagem total de espermatozoides progressivos são empregados na avaliação do sucesso da correção da varicocele mediante a cirurgia.⁸⁸⁻⁹⁰ Os resultados apresentados na Tabela 9 indicam clara diferença entre os grupos, quando consideramos os resultados pós-varicocelectomia. Há aumento expressivo na contagem de espermatozoides (4,8 vezes) e na contagem daqueles com motilidade progressiva (8,6 vezes). Isso indica sucesso do procedimento para pelo menos 62% dos pacientes.

Denise et al., a partir de uma extensa revisão da literatura sobre o impacto do grau da varicocele sobre a resposta da intervenção cirúrgica, analisaram 20 estudos envolvendo 2001 homens inférteis com varicocele, e concluíram que o manejo cirúrgico promove melhora na concentração de espermatozoides/ml no sêmen assim como também sobre a motilidade. Além disso, os autores observaram que a magnitude da melhora estaria diretamente ligada ao grau da doença, sendo os melhores resultados foram alcançados nos pacientes com maior grau da doença.⁸⁷

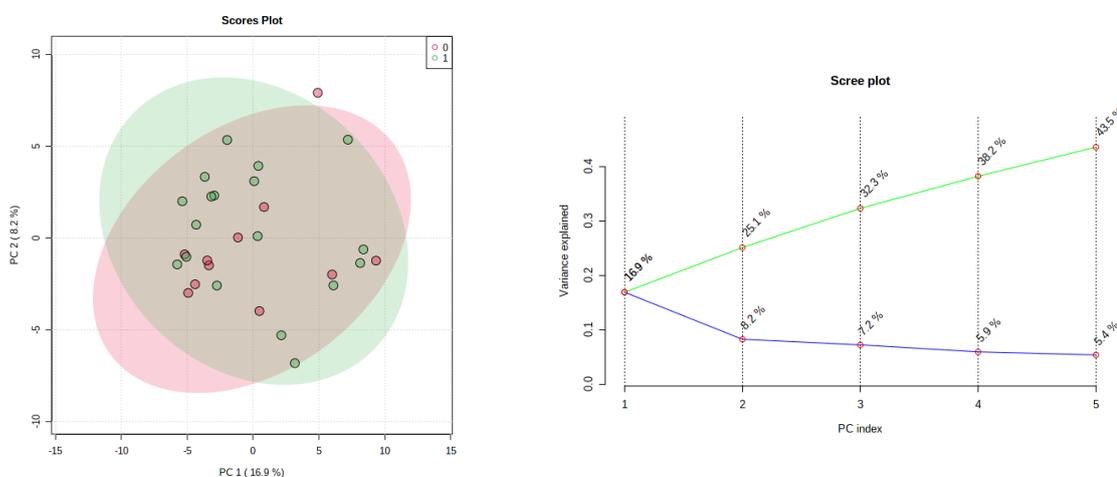
Graham Luke et al. demonstraram que a melhora da Contagem Total de Espermatozoides Progressivos ocorreu em 69,4% dos homens avaliados, a maioria dos resultados já poderia ser visualizada com três meses após a varicocelectomia (78,8%). Os autores consideraram como critério de melhora significativa os pacientes aqueles que apresentaram pelo menos 50% de melhora no total de espermatozoides progressivamente móveis⁹¹. Nosso estudo apresenta números semelhantes.

Ressalta-se que o procedimento cirúrgico é invasivo, requer infraestrutura e profissionais médicos especializados para este tipo de cirurgia. Além disso, o primeiro exame para a avaliação da melhoria dos padrões seminais pós-cirurgia deverá ser feito de 8 a 16 semanas. Portanto, a elaboração de modelos metabonômicos com finalidade de prognóstico da melhoria ou não da qualidade do sêmen após cirurgia poderá ser mais uma informação importante na decisão do médico para a realização de tal procedimento, principalmente quando o objetivo principal do paciente é obter o sucesso gestacional.

Dispondo da plataforma *MetaboAnalyst* 5.0, iniciamos a investigação não-supervisionada dos nossos dados. Para isto, na etapa de pré-processamento dos dados, foi empregando o filtro MAD (do inglês, *median absolute deviation*) e posteriormente as variáveis foram devidamente autoescaladas. Ao analisar o

gráfico de escores da PCA, não foi possível identificar agrupamentos naturais nas classes de interesse (Figura 26). Além disso, mesmo empregando 5 PCs, apenas 43,5% da variância foi explicada. Isso indica que há múltiplas fontes de variância no conjunto de dados.

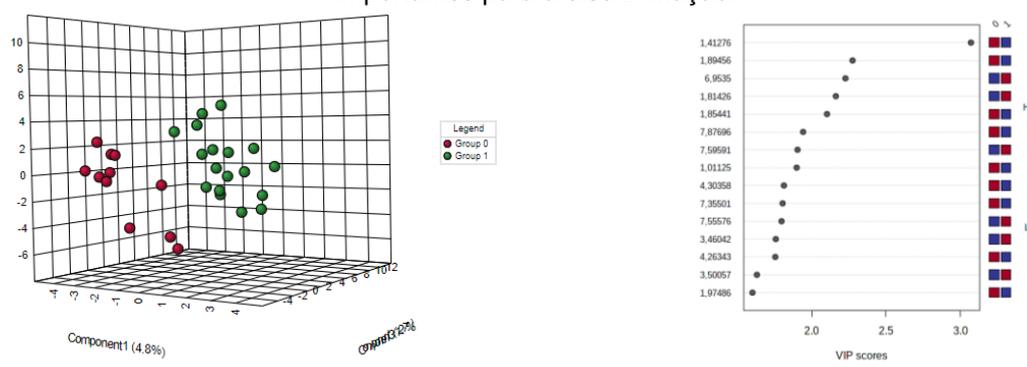
Figura 26 – Resultado da PCA Melhora versus Estabilidade dos parâmetros seminais pós-varicolectomia. À esquerda, Gráfico de escores; e à direita, Screen plot com a variância explicada por cada componente principal, em azul, e acumulada, em verde.



Fonte: Autoria própria

Em seguida, os formalismos supervisionados foram empregados no conjunto dos dados em análise. O primeiro a ser analisado foi a PLS-DA que demonstrou separação entre os dois grupos, como pode-se visualizar na Figura 27.

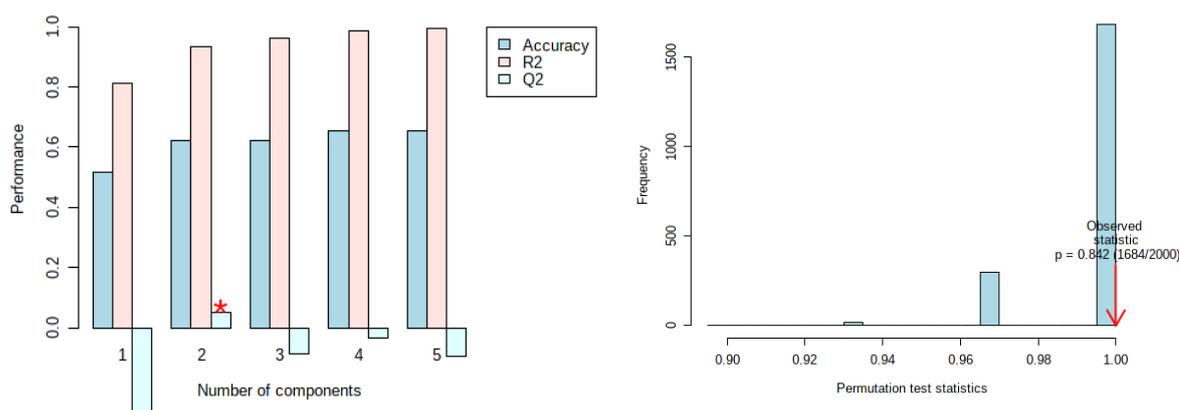
Figura 27 – Resultado da PLS-DA Melhora versus Estabilidade dos parâmetros seminais pós-varicolectomia. À esquerda, Gráfico de escores; e à direita, VIP score com as variáveis mais importantes para a discriminação.



Fonte: Autoria própria

Os gráficos de escore e o VIP da PLS-DA (Figura 27), permite demonstrar a nítida separação dos grupos e identificar as variáveis mais importantes que foram empregadas para discriminar os dois grupos, respectivamente. Entretanto, ao analisar as figuras de mérito do modelo obtido evidenciamos as suas limitações (Figura 28). A primeira dela é que o modelo apresenta um baixo valor preditivo ($Q^2 < 0$) e segundo que no teste com 2000 permutações o valor $p = 0,842$. Estes indicadores sugerem que o modelo está superajustado.

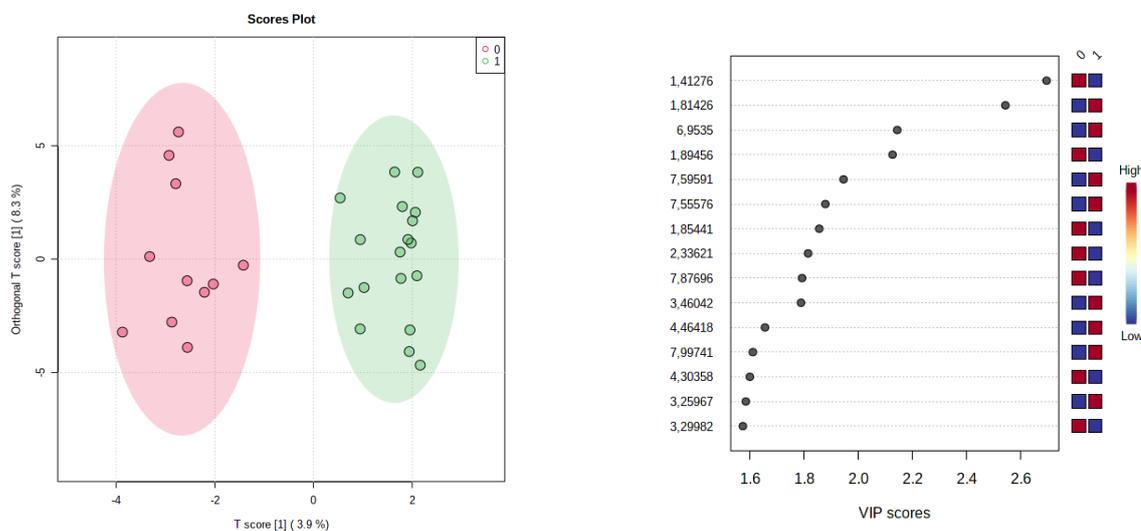
Figura 28 – Figuras de mérito da PLS-DA Melhora versus Estabilidade dos parâmetros seminais pós-varicolectomia. À esquerda, valores de exatidão, R2 e Q2; e à direita, teste de permutação.



Fonte: Autoria própria

Assim, optamos por avaliar o comportamento dos nossos dados frente a OPLS-DA. No que pese a representação gráfica da separação dos dois grupos, o OPLS-DA manteve o comportamento previsto na PLS-DA, assim como também foi possível, a partir da análise do VIP Score, averiguar que as 4 primeiras variáveis mais importantes para a construção deste modelo eram as mesmas que tinham sido identificadas PLS-DA, contudo em ordens de importância distintas (Figura 29).

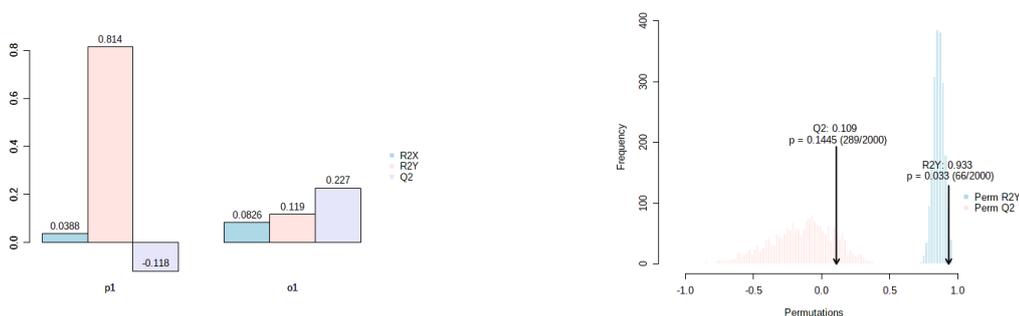
Figura 29 – Figuras de mérito da OPLS-DA Melhora versus Estabilidade dos parâmetros seminais pós-varicolectomia. À esquerda, Gráfico de escores; e à direita, VIP escores com as varáveis mais importantes.



Fonte: Autoria própria.

Porém, o grande diferencial da OPLS-DA foi a capacidade de separar os dois grupos, com valores significativos de exatidão, além do valor de p no teste com 2000 permutações ser inferior a 0,05 (Figura 30).

Figura 30 – Figuras de mérito da OPLS-DA Melhora versus Estabilidade dos parâmetros seminais pós-varicolectomia. À esquerda, valores de R2 e Q2; e à direita, teste de permutação.

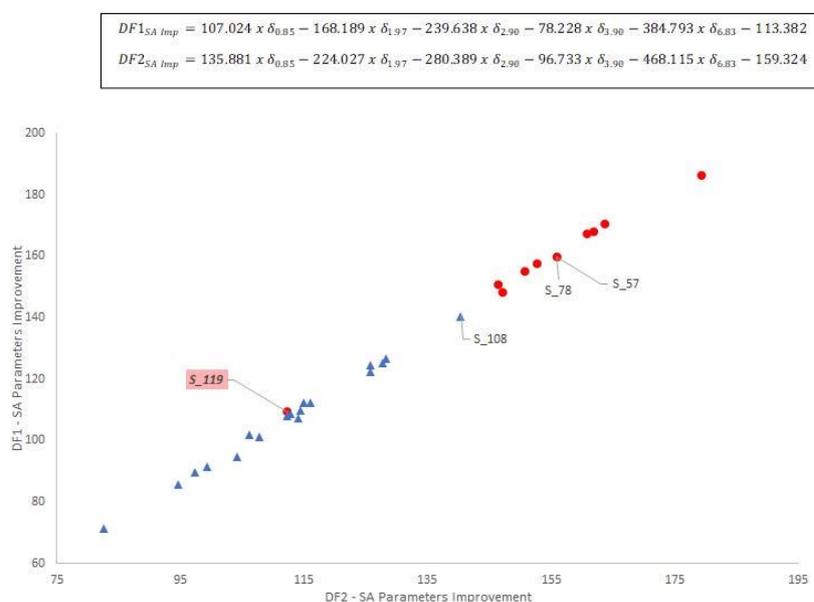


Fonte: Autoria própria.

Porém, continuamos com problemas de predição, uma vez que o valor de Q2 ainda é muito baixo. Por isso, usamos o formalismo LDA. No que diz respeito ao pré-tratamento dos dados, as amostras foram normalizadas pela soma (SNV), enquanto

as variáveis foram autoescaladas. Assim, o modelo foi construído usando cinco variáveis, as quais foram selecionadas usando o lambda de Wilks. As variáveis selecionadas para o modelo LDA apresentaram os seguintes deslocamentos químicos δ (ppm) 0,85, 1,97, 2,89, 3,90 e 6,83. A significância estatística do modelo foi avaliada pelo teste F (Fischer), com 5 e 23 graus de liberdade, resultando $F_{23,5} = 10,35$ ($p < 0,001$). Como podemos observar, o valor de F calculado pelo modelo foi demasiadamente significativo, visto que foi 3,4 vezes maior que o valor $F_{23,5}$ tabelado (4,47), apresentando $p < 0,001$. O modelo foi validado usando LOOCV. Basicamente, essa validação construiu 29 modelos distintos utilizando as mesmas amostras, com a exceção de uma amostra, por vez, a qual teve sua classificação predita pelo modelo construído com a sua ausência. Então, os modelos obtidos nesta validação apresentaram uma exatidão de 93,0%. As equações abaixo apresentam as funções discriminantes (DF) obtidas partir da LDA, enquanto a Figure 31 demonstra a projeção das amostras no espaço definido a partir das duas funções discriminantes.

Figura 31 – Gráfico de escores LDA Melhora (quadrados vermelhos) versus Estabilidade (quadrados azuis) dos parâmetros seminais.



Fonte: Autoria própria

Observa-se que há um erro de classificação para cada grupo. Isso indica que a exatidão do modelo é igual a 93,1%, enquanto especificidade e sensibilidade são iguais a 94,4% e 90,9%, respectivamente. Esses resultados estão resumidos na matriz de contingência para o modelo na Tabela 16.

Tabela 16 – Matriz de Contingência para o modelo usando LDA – Melhora *versus* Estabilidade dos parâmetros seminais.

		Diagnóstico Clínico		
		Melhora (M)	Estabilidade (E)	Valor de <i>p</i>
Modelo LDA	Melhora (M)	10	1	< 0,001
	Estabilidade (E)	1	17	

Fonte: Autoria própria.

Para a identificação dos deslocamentos químicos das cinco variáveis selecionadas na discriminação dos grupos associadas aos possíveis metabólitos, foi empregado o *Human metabolome database* (HMDB)¹⁰⁰ e o relato de Paiva et al.²⁶. Em relação ao deslocamento (δ) 0,85 ppm foram observados dois possíveis metabólitos: 2-hidroxi-3-metilvalerato (δ 0,87 ppm/HMDB) ou caprilato (δ 0,86 ppm/HMDB). O sinal em δ 1,97 ppm foi atribuído ao grupo metil da isoleucina, enquanto os sinais em δ 3,90 e 6,83 ppm foram atribuídos ao GPC e a N-acetiltirosina, respectivamente. Não conseguimos atribuir o sinal em δ 2,89 ppm.

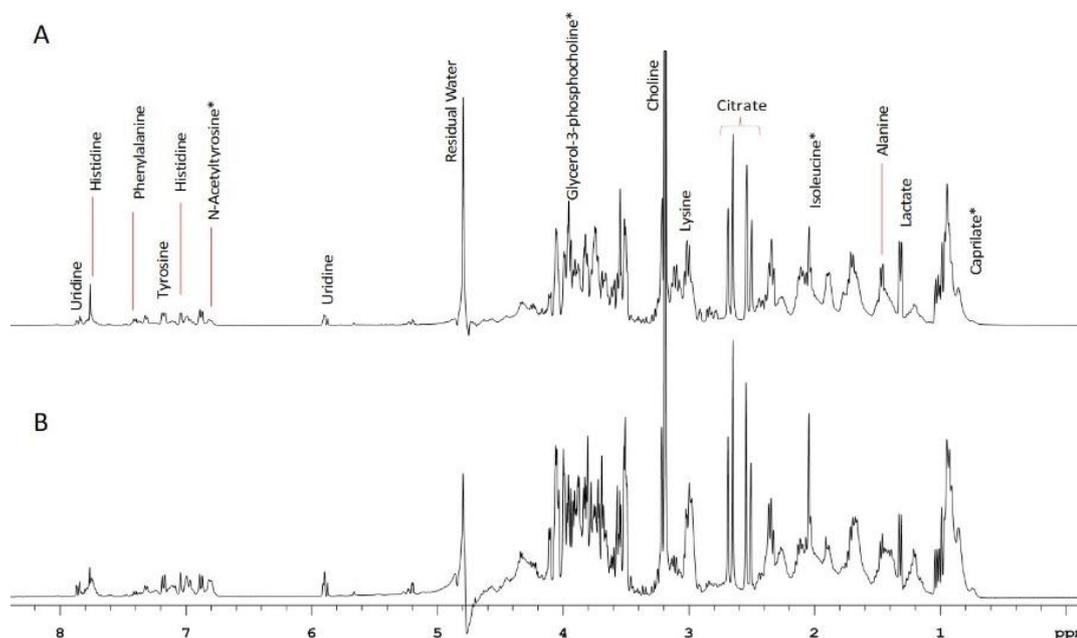
Tabela 17 – Atribuição dos sinais responsáveis pela discriminação nas funções discriminantes para predição de melhora dos parâmetros seminais pós-varicolectomia em até 24 meses pós-varicolectomia.

<i>FD Melhora dos Parâmetros Seminais</i>	
δ (ppm)	Metabólito
0,85	2-Hidroxi-3-Metil Valerato ou Caprilato
1,97	Isoleucina
2,89	Não atribuído
3,90	GPC – Glicerol-fosfocolina ou Creatina
6,83	N-Acetil tirosina

Fonte: Autoria própria

Como podemos observar, nas funções discriminantes construídas, altos escores nas funções discriminantes estão associados às amostras em que ocorreu melhora dos parâmetros seminais pós-varicocelectomia, enquanto baixos escores estão àquelas em que não foram observados acréscimos na qualidade dos parâmetros seminais. É possível avaliar, a partir dos pesos de cada variável selecionada, a influência dos metabólitos na predição. No entanto, é preciso estar atentos para considerar o aspecto multivariado da questão. A capacidade preditiva é obtida a partir da variação em conjunto desses metabólitos, não podendo, portanto, ser avaliada a variação isolada de cada metabólito. Dessa forma, as amostras que apresentem altos níveis séricos de caprilato, antes da intervenção cirúrgica, tendem a melhorar os padrões seminais pós-varicocelectomia. No entanto, essa classificação depende também de níveis séricos moderados de isoleucina, N-acetiltirosina e glicerol-3-fosfocolina, que possuem pesos negativos nas funções discriminantes. A Figura 32 apresenta dois espectros de RMN de ^1H de fluido seminal, pré-varicocelectomia, de voluntários que apresentaram melhora (A) e estabilidade (B) dos parâmetros seminais pós-varicocelectomia.

Figura 32 – Espectros de RMN de ^1H (D_2O , 400 MHz, CPMG-Presat) de fluidos seminais coletados pré-varicocelectomia de voluntários que apresentaram melhora (A) e estabilidade (B) dos parâmetros seminais pós-intervenção cirúrgica. * Variáveis selecionadas na LDA.



Fonte: Autoria própria

10.1.5 Avaliação da taxa de gravidez após varicocelectomia.

A qualidade seminal não é o melhor descritor sobre a infertilidade, entretanto há relatos indicando a relação da melhora da qualidade dos padrões seminais com o aumento das chances de gravidez. Mesmo assim, há uma parcela de homens para os quais, mesmo após a varicocelectomia, não são observadas melhoras significativas no espermograma⁸⁴⁻⁸⁷. Embora a melhora da qualidade dos parâmetros seminais pós-varicocelectomia seja uma sinalização boa, ela não oferece nenhuma garantia do sucesso gestacional. Neste caso, dois parâmetros importantes que devem ser considerados são as taxas de gravidez e de nascidos vivos. Abdel-Meguid et al. (2011) observaram, a partir de um estudo controlado randomizado com 145 homens inférteis com varicocele palpável, acompanhados por 12 meses, que a taxa de gravidez espontânea em homens inférteis com varicocele, sem intervenção cirúrgica, foi de

13,9% (10 de 72 pacientes); enquanto no grupo que foi submetido à varicocelectomia, a taxa de gravidez foi de 32,9% (24 de 73 pacientes). Isso sugere que a intervenção cirúrgica tem impacto positivo na taxa de gravidez¹⁰². A partir da meta-análise envolvendo 31 artigos, Birowo et al. (2020) observaram que o reparo da varicocele induz uma melhora na qualidade dos parâmetros seminais e que a taxa de gravidez no grupo que foi submetido à cirurgia é significativamente maior que a observada no grupo de não tratados – OR (*odds ratio*) igual a 1,82 considerando o intervalo de confiança de 95% se estendendo de 1,37 a 2,41 ($p = 0,0001$). Além disso, nesta revisão, os autores relataram também aumento na taxa de nascidos vivos no grupo de homens que foram submetidos à varicocelectomia, OR = 2,80; 95% IC: 1,67 – 4,72; $p = 0,0001$ ¹⁰³.

Nesse sentido, nós investigamos a taxa de gravidez entre os voluntários que foram submetidos à varicocelectomia. Utilizamos o mesmo grupo amostral empregado no estudo III, ou seja, 43 amostras de voluntários com varicocele e infertilidade, que foram submetidos à varicocelectomia. O tempo adotado para a avaliação do sucesso de gravidez foi de 24 meses após a data da cirurgia. Relato de gravidez, a qualquer tempo no intervalo de 24 meses, inclui o voluntário no grupo de Sucesso Gestacional; enquanto a inclusão no grupo de Insucesso Gestacional só se deu após 24 meses da cirurgia. A equipe médica conseguiu contactar 24 voluntários para obter informações quanto à ocorrência de gravidez. Desses, 10 relataram que conseguiram engravidar a companheira, enquanto 14 relataram insucesso. No entanto, o modelo foi construído com 23 amostras, pois um dos voluntários que relatou insucesso ainda não tinha completos 24 meses do procedimento cirúrgico. A Tabela 18 apresenta os dados coletados no espermograma pós-varicocelectomia dos participantes do estudo.

Tabela 18 – Parâmetros de qualidade seminal dos grupos investigados. Foram utilizados dados pós-varicocelectomia.

	Limite Inferior de Referência (IC 95%) ¹	Sucesso Gestacional (n=10)	Insucesso Gestacional (n=13)	Valor de p^*
Concentração EZP (106/mL), mediana (IQR)	15 (12 – 16)	16.0 (8.5-23.5)	6.0 (0.9-23.5)	0,21
Contagem total EZP (106U), mediana (IQR)	39 (33 – 46)	45.0 (33.0-67.8)	20.0 (3.7-50.3)	<0,05
Motilidade Progressiva (%), mediana (IQR)	32 (31 – 34)	43 (37-51)	28 (16-38)	<0,05
Contagem total EZP com motilidade progressiva (106/mL), mediana (IQR)	4,8 (3,7 – 5,4)	7.1 (13.8-34.2)	3.5 (0.2-13.3)	<0,05

1 OMS. *Teste Wilcoxon unicaudal

Fonte: Autoria própria.

Os dados do espermograma indicam que há uma diferença estatisticamente significativa entre os grupos, no número total de espermatozoides com motilidade progressiva, bem como no percentual de espermatozoides com motilidade progressiva. O grupo que conseguiu engravidar a companheira apresenta mediana superior ao observado no grupo que não conseguiu gravidez no intervalo de 24 meses.

O modelo metabonômico foi construído usando o formalismo LDA, sendo as funções discriminantes construídas com cinco variáveis selecionadas considerando o menor valor de lambda de Wilks. As variáveis selecionadas nesta modelagem têm os seguintes deslocamentos químicos δ (ppm): 3,26; 3,42; 3,86; 3,90 e 4,26. As duas funções discriminantes são apresentadas abaixo.

$$DF_1 = -204,33 - 125,5x\delta_{3,26} - 1140,63x\delta_{3,42} - 59,85x\delta_{3,86} - 142,61x\delta_{3,90} + 52,08x\delta_{4,26}$$

$$DF_2 = -134,486 - 75,576x\delta_{3,26} - 839,563x\delta_{3,42} - 21,329x\delta_{3,86} - 98,692x\delta_{3,90} + 5,279x\delta_{4,26}$$

A Tabela 19 apresenta os possíveis metabólitos associados à discriminação observada entre os grupos. Essa atribuição não é definitiva, uma vez que necessita de outros experimentos para tal.

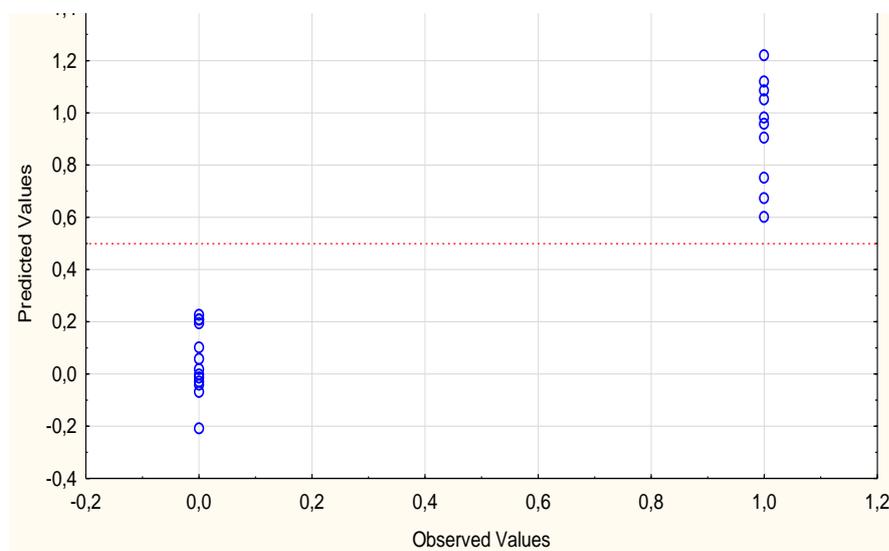
Tabela 19 – Atribuição dos sinais responsáveis pela discriminação nas funções discriminantes de correlação com casos de gravidez, em até 24 meses pós-varicocelectomia.

δ (ppm)	Possíveis Metabólitos
3,26	Arginina ou Glicose ou TMAO (Óxido de Trimetil Amônio)
3,42	Carnitina
3,86	GPC ou 2-Hidroxi-3-Metil Valerato ou O-Acetil Carnitina
3,90	Creatina
4,26	Treonina

Fonte: Autoria própria.

A Figura 33 apresenta o gráfico de escores com os valores preditos pelo modelo LDA *versus* informação de gravidez, para a qual foi arbitrariamente dado o valor 0 (zero), para os indivíduos que não conseguiram uma gravidez no período de 24 meses pós-varicocelectomia, e 1 (um), para aqueles que conseguiram uma gestação pós-varicocelectomia no período mencionado.

Figura 33 – Gráfico de escores com os valores preditos pelo modelo LDA *versus* informação de gravidez.



Fonte: Autoria própria.

O modelo desenvolvido não pode ser classificado como uma ferramenta prognóstica para gravidez, haja vista que não insere a contribuição feminina. Mas, pode ser um indicador importante para a tomada de decisão de indicar o procedimento cirúrgico, visando aumentar as chances de sucesso gestacional. Assim, comparamos os resultados obtidos pelas duas abordagens, que está resumido na Tabela 20.

Tabela 20 – Distribuição das amostras com base na qualidade do sêmen e da condição de gravidez.

		<i>Qualidade Seminal</i>		
		SIM	NÃO	S/ Classificação
<i>Sucesso Gestacional</i>	SIM	8	2	0
	NÃO	8	5	0
	S/ Classificação	2	4	14

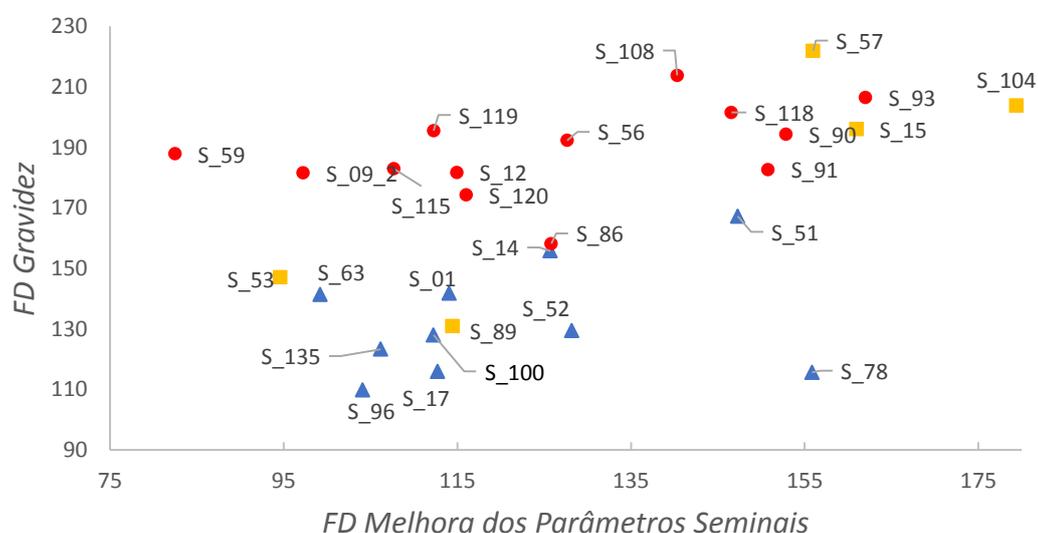
Fonte: Autoria própria.

A tabela 20 diz que tínhamos informação sobre a qualidade dos parâmetros seminais de 29 voluntários, mas sobre o status de gravidez só tínhamos dados de 23 voluntários. A Figura 34 apresenta a projeção das amostras em um sistema de coordenadas formado por uma das funções discriminantes construída no Estudo III (Predição da Melhora da Qualidade Seminal) e uma das funções construída no Estudo IV (Impacto sobre o Sucesso Gestacional). As funções discriminantes para predição de melhora dos parâmetros seminais pós-varicocelectomia e correlação com gravidez são, respectivamente:

$$FD = 107,024 \times \delta_{0,85} - 168,189 \times \delta_{1,97} - 239,638 \times \delta_{2,90} - 78,228 \times \delta_{3,90} - 384,793 \times \delta_{6,83} - 113,382$$

$$FD = -75,576 \times \delta_{3,26} - 839,563 \times \delta_{3,42} - 21,329 \times \delta_{3,86} - 98,679 \times \delta_{3,90} + 5,279 \times \delta_{4,26} - 134,486$$

Figura 34 – Gráfico de escores das funções discriminantes. Os círculos vermelhos são amostras de voluntários que não conseguiram gravidez em até 24 meses da cirurgia; os triângulos azuis são amostras de voluntários que conseguiram engravidar a companheira; os quadrados dourados são 05 amostras para as quais não há informação sobre gravidez.



Fonte: Autoria própria.

A significância estatística do modelo foi avaliada pelo teste F (Fischer), com 5 e 17 graus de liberdade, resultando $F_{17,5} = 32,56$ ($p < 0,001$). O modelo foi validado usando $LOOCV$ e a matriz de contingência após validação cruzada é apresentada na Tabela 21.

Tabela 21 – Matriz de Contingência para o modelo usando LDA – Correlação com gravidez.

		<i>Sucesso Gestacional</i>		Valor de p
		SIM (S)	NÃO (N)	
<i>Modelo LDA</i>	SIM (S)	9	0	< 0,001
	NÃO (N)	1	13	

Fonte: Autoria própria.

Podemos observar na Figura 34 que as amostras 14, 51 e 86 estão dispostas em uma região limítrofe entre os dois grupos. No caso das amostras 14 e 86, que pertencem a grupos distintos, elas praticamente se sobrepõem no sistema de coordenadas definido pelas duas funções discriminantes. Isso sugere uma classificação duvidosa. Na validação cruzada, observamos que o modelo construído apresenta apenas um erro na classificação das amostras, sendo a amostra 51 classificada erroneamente.

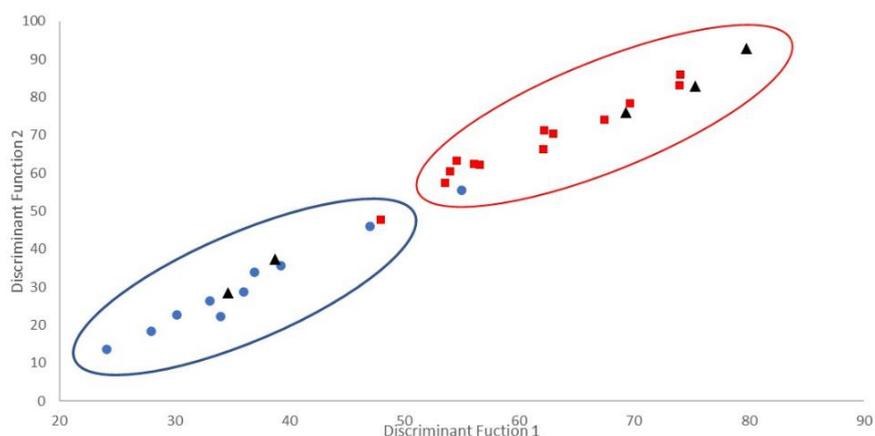
Visando uma melhor classificação, construímos uma nova função discriminante, desta vez usando os escores alcançados por cada amostra nas funções discriminantes construídas (Melhora da Qualidade Seminal e Correlação com Gravidez), como variável de entrada, e o relato de ocorrência (ou não) de gravidez como matriz resposta (matriz Y). As funções discriminantes construídas são apresentadas abaixo:

$$FD1 = 0,1428 x (\text{Escore na função Melhora}) + 0,4306 x (\text{Escore na função Gravidez})$$

$$FD2 = 0,0929 x (\text{Escore na função Melhora}) + 0,6628 x (\text{Escore na função Gravidez})$$

Após a validação cruzada para este novo modelo LDA, identificamos que as amostras que foram apresentadas na Figura 34 na região limítrofe do classificador para a LDA das variáveis originais apenas a 14 apresentou classificação correta nesta nova LDA. A amostra 51 foi novamente classificada no grupo errado, porém a amostra 86 que anteriormente teve sua classificação correta, neste novo modelo foi classificada erroneamente. A Figura 35 apresenta o gráfico de escores das amostras no sistema de coordenadas definido pelas funções discriminantes FD1 e FD2.

Figura 35 – Gráfico de escores nas funções discriminantes para correlação com gravidez. Círculos Azuis = Gravidez; Quadrados Vermelhos = Não Gravidez; Triângulos Pretos = sem informação de gravidez.



Fonte: Autoria própria

Como pode ser observado na Figura 35, há dois erros de classificação – amostras 51 e 86, resultando em 90,3% de exatidão. A Tabela 22 apresenta a matriz de contingência para esse novo modelo.

Tabela 22 – Matriz de Contingência para o modelo usando LDA – SIM versus NÃO sucesso gestacional obtida a partir das funções discriminantes construída usando os escores nas funções discriminantes Melhora dos Parâmetros Seminais e Correlação com Gravidez.

		<i>Sucesso Gestacional</i>		Valor de <i>p</i>
		SIM (S)	NÃO (N)	
Modelo LDA	SIM (S)	9	1	< 0,001
	NÃO (N)	1	12	

Fonte: Autoria própria.

11 CONCLUSÃO

O trabalho de tese aqui desenvolvido iniciou com a coleta de dados espectrais de ressonância magnética nuclear de amostras de fluido seminal de voluntários previamente selecionados e incluídos no estudo. Inicialmente, identificamos as melhores condições experimentais para a obtenção e processamento dos espectros, de forma a preservar as informações acerca do status bioquímico dos voluntários. Concluímos que a sequência de pulsos de RF CPMG com pré-saturação do sinal da água entrega os melhores dados para os estudos metabonômicos. Seguiu-se o tratamento quimiométrico visando identificar os formalismos de estatística multivariada que conseguem discriminar os grupos investigados. Foram realizados quatro estudos, resultando em quatro modelos metabonômicos:

- I. Diagnóstico de varicocele
- II. Diagnóstico diferencial de infertilidade masculina associada à varicocele
- III. Prognóstico de melhora dos parâmetros seminais pós-varicocelectomia
- IV. Predição do impacto da varicocelectomia sobre as chances de sucesso gravidez.

No Estudo I, diagnóstico de varicocele, os formalismos OPLS-DA e GA-LDA foram eficientes na identificação das amostras do grupo Controle, discriminando-as das amostras dos voluntários diagnosticado com varicocele, independente do quadro de fertilidade. O modelo usando o formalismo OPLS-DA obteve exatidão, sensibilidade e especificidade iguais a 90%, 92,8% e 83,3%, respectivamente; enquanto o modelo usando GA-LDA alcançou exatidão, sensibilidade e especificidade iguais a 92,86%, 100% e 85,71%. Em relação aos metabólitos que mais influenciaram no modelo destacamos arginina, lactato e 3-hidroxi-butirato.

No Estudo II, diagnóstico de infertilidade associada à varicocele, tanto a OPLS-DA quanto a GD-LDA demonstraram capacidades de discriminação das amostras do grupo varicocele fértil, das amostras dos voluntários diagnosticado com varicocele infértil. O modelo empregando o formalismo OPLS-DA obteve exatidão, sensibilidade e especificidade iguais a 94,64%, 95,24% e 95,24%, respectivamente; o modelo usando GA-LDA exibiu uma exatidão, sensibilidade e especificidade iguais a 91,67%, 100% e 83,33%. Neste estudo os principais metabolitos identificados foram o caprato, a alanina e arginina.

Modelando simultaneamente os 3 grupos (controle, varicocele fértil e varicocele infértil), o formalismo PLS-DA apresentou valor de R^2 igual a 0,82, e foi capaz de discriminar os grupos baseado no status metabólico. Foram identificados 13 metabolitos de acordo com sua importância na obtenção do modelo, são eles: caprato, valina, 3-hidroxi butirato, lactato, 4-aminobutirato (GABA), Isoleucina, nitrato, glicosídeos, N-acetiltirosina, glutamina, tirosina, arginina e uridina.

No Estudo III, prognóstico de melhora dos parâmetros seminais após varicocelectomia, o formalismo LDA apresentou os melhores resultados. Os valores de exatidão, sensibilidade e especificidade foram 93,1%, 90,9% e 94,4%, respectivamente. Em relação aos sinais/metabolitos, para o deslocamento (δ) 0,85 ppm foram observados dois possíveis metabolitos: 2-hidroxi-3-metilvalerato (δ 0,87 ppm/HMDB) ou caprilato (δ 0,86 ppm/HMDB). O sinal em δ 1,97 ppm foi atribuído ao grupo metil da isoleucina, enquanto os sinais em δ 3,90 e 6,83 ppm foram atribuídos à creatinina e à N-acetiltirosina, respectivamente.

Em relação ao estudo IV, correlação com os relatos de gravidez natural em até 24 meses após varicocelectomia, a LDA também foi o método quimiométrico mais efetivo. Dos pacientes que foram submetidos à cirurgia de varicocelectomia, 47,3%

conseguiram engravidar suas companheiras. Entre os homens com melhora na qualidade do líquido seminal pós-operatório e que alcançaram a gestação, o percentual corresponde a 56,2%. Entretanto, 43,8% dos participantes que apresentaram melhora na qualidade do sêmen pós-operatório, não conseguiram, no período de 24 meses, a gestação. Por outro lado, 28,6% dos voluntários relataram que conseguiram engravidar suas parceiras, mesmo que não tenhamos observado melhora nos parâmetros de qualidade do sêmen. Quanto à caracterização dos deslocamentos associados aos metabólitos, restam dúvidas em relação ao sinal em δ 1,37 ppm que pode ser atribuído tanto a grupo metileno da isoleucina, quanto ao mesmo grupo do 2-hidróxi-metil valerato; o sinal em δ 1,73 ppm pode ser atribuído ao grupo metileno do 2-metil glutarato ou a um grupo metino do 2-hidróxi-metil valerato. Além disso, o mesmo pode ser atribuído ao grupo metileno central da putrescina. O sinal em δ 3,86 ppm, que pode ser conferido tanto à acetilcarnitina, quanto ao GPC. Consequentemente, outros ensaios precisam ser executados para atribuir adequadamente esses sinais.

Os achados deste trabalho de tese confirmam a hipótese admitida de que, devido à homeostase, condições clínicas como infertilidade e varicocele, bem como a intervenção cirúrgica (varicocelectomia), alteram o perfil de metabólitos endógenos contidos no fluido seminal, permitindo a construção de modelos metabonômicos capazes de discriminar os pacientes em função do seu status bioquímico. É um achado relevante, pois permitiu a construção de ferramentas não-invasivas para diagnóstico diferencial de infertilidade masculina associada à varicocele e prognóstico de melhora da qualidade do sêmen do paciente que busca sucesso na tentativa de engravidar sua companheira.

12 PERSPECTIVAS

Mediante os resultados expostos nesta tese, pretendemos ampliar a população dos estudos, adquirir mais espectros do soro do sêmen de pacientes com infertilidade e varicocele para avaliar o desempenho dos modelos metabonômicos. Além disso, pretendemos incorporar também novas amostras ao modelo de predição da melhoria dos parâmetros seminais de modo a alcançar um conjunto teste para a validação dos modelos.

Nos estudos futuros, pretendemos avaliar as principais vias metabólicas associadas aos metabólitos discriminantes encontrados em cada um dos estudos avaliados nesta tese, e correlacioná-los, quando possível, à varicocele e/ou infertilidade masculina.

Por último, estamos construindo uma nova matriz de dados, utilizando dados espectrais de RMN de ^1H de soro de sêmen coletados pós-varicocelectomia, visando identificar entre os pacientes que não conseguiram gravidez aqueles que têm indicação e que poderão ter sucesso na extração de espermatozoides, que serão utilizados nas técnicas de fertilização assistida.

REFERÊNCIAS

1. Kantartzi, P. D., Goulis, C. D., Goulis, G. D., & Papadimas, I. Male infertility and varicocele: myths and reality. *Hippokratia*, 11.3 (2007): 99
2. World Health Organization et al. WHO laboratory manual for the examination and processing of human semen. 2010
3. Peterson, B. D. et al. The impact of partner coping in couples experiencing infertility. *Human Reproduction*, v. 23, n. 5, p. 1128-1137, 2008
4. Ferlin, Alberto; Foresta, Carlo. Infertility: practical clinical issues for routine investigation of the male partner. *Journal of Clinical Medicine*, v. 9, n. 6, p. 1644, 2020
5. Lara, L. A., Fuentealba-Torres, M., dos Reis, R. M., & Cartagena-Ramos, D. Impact of Infertility on the Sexuality of Couples: an Overview. *Current Sexual Health Reports*, 10.4 (2018): 353-359
6. Lindgren, Mark C. Male Infertility. *Physician Assistant Clinics*, 3.1 (2018) 139-147
7. Alsaikhan, B., Alrabeeah, K., Delouya, G., & Zini, A. Epidemiology of varicocele. *Asian journal of andrology*, 18.2 (2016):179
8. Krausz, Csilla, and Antoni Riera-Escamilla. Genetics of male infertility. *Nature Reviews Urology*, 1 (2018):1.
9. Lima-Oliveira, Gabriel et al. Pre-analytical phase management: a review of the procedures from patient preparation to laboratory analysis. *Scandinavian journal of clinical and laboratory investigation*, v. 77, n. 3, p. 153-163, 2017.
10. Campana, Gustavo Aguiar, Carmen Paz Oplustil, and Lorena Brito de Faro. Tendências em medicina laboratorial. *Jornal Brasileiro de Patologia e Medicina Laboratorial*, 47.4 (2011): 399-408

11. Sodré, Fábio L., Josete Conceição Barreto Costa, and José Carlos C. Lima. Avaliação da função e da lesão renal: um desafio laboratorial. *Jornal Brasileiro de Patologia e Medicina Laboratorial* 43.5 (2007): 329-337
12. Mincis, Moisés, and Ricardo Mincis. Doença hepática alcoólica. *Gastroenterologia e hepatologia* (1997): 695-716
13. Gladwin, Mark T., Jack H. Crawford, and Rakesh P. Patel. The biochemistry of nitric oxide, nitrite, and hemoglobin: role in blood flow regulation. *Free Radical Biology and Medicine*, 6.6 (2004): 707-717
14. Bingol, K., Bruschiweiler-Li, L., Li, D., Zhang, B., Xie, M., & Brüschiweiler, R. Emerging new strategies for successful metabolite identification in metabolomics. *Bioanalysis*, 8.6, (2016): 557-573
15. Mathew, A. K., and V. C. Padmanaban. Metabolomics: the apogee of the omics trilogy. *Int J Pharm Pharm Sci*, 5.2 (2013): 45-8
16. Silva, Ronaldo Dionísio da. Aplicações metabonômicas usando ressonância magnética nuclear de ¹H: diagnóstico não-invasivo de câncer de próstata e urológico & classificação de oliva extra virgem de produção orgânica. Tese de doutorado. UFPE. (2017).
17. Robertson, Donald G. Metabonomics in toxicology: a review. *Toxicological sciences*, 85.2 (2005): 809-822.
18. Tang, Huiru, and Yulan Wang. Metabonomics: a revolution in progress. *Sheng wu hua xue yu sheng wu wu li jin zhan*, 33.5 (2006): 401-417.
19. Krausz, Csilla. Male infertility: pathogenesis and clinical diagnosis. *Best practice & research Clinical endocrinology & metabolism*, v. 25, n. 2, p. 271-285, 2011
20. Teixeira, Maria Yasmin Paz et al. Componentes do estilo de vida associados à infertilidade masculina. *Nutr. clín. diet. hosp*, p. 179-184, 2018.

21. Hamada A, Esteves SC, Nizza M, Agarwal A. Unexplained male infertility: diagnosis and management. *Int Braz J Urol.* 2012 Sep-Oct;38(5):576-94.
22. Ferlin, Alberto; Foresta, Carlo. Infertility: practical clinical issues for routine investigation of the male partner. *Journal of Clinical Medicine*, v. 9, n. 6, p. 1644, 2020
23. Owen, Derek H., and David F. Katz. A review of the physical and chemical properties of human semen and the formulation of a semen simulant. *Journal of andrology* 26.4 (2005): 459-469
24. Harraway, Carolyn, Nathan G. Berger, and Norman H. Dubin. Semen pH in patients with normal versus abnormal sperm characteristics. *American journal of obstetrics and gynecology* 182.5 (2000): 1045-1047
25. Practice Committee of the American Society for Reproductive Medicine. Diagnostic evaluation of the infertile male: a committee opinion. *Fertility and sterility* 103.3 (2015): e18-e25
26. Paiva, C. et al. Identification of endogenous metabolites in human sperm cells using proton nuclear magnetic resonance (¹H-NMR) spectroscopy and gas chromatography-mass spectrometry (GC-MS). *Andrology*, v. 3, n. 3, p. 496-505, 2015.
27. Qiao, Shanlei et al. Seminal plasma metabolomics approach for the diagnosis of unexplained male infertility. *PLoS One*, v. 12, n. 8, p. e0181115, 2017.
28. Saleem, Tahia H. et al. Biochemical assessments of seminal plasma zinc, testis-expressed sequence 101 and free amino acids and their correlations with reproductive hormones in male infertility. *Biological trace element research*, p. 1-14, 2020.

29. COLLODEL, Giulia et al. Associations between biochemical components of human semen with seminal conditions. *Systems biology in reproductive medicine*, v. 65, n. 2, p. 155-163, 2019.
30. Engel, Kathrin M. et al. Metabolomic profiling reveals correlations between spermogram parameters and the metabolites present in human spermatozoa and seminal plasma. *PloS one*, v. 14, n. 2, p. e0211679, 2019.
31. Cruz, Daniel Filipe et al. Vias de sinalização reguladoras das funções do espermatozoide. *Revista Internacional de Andrología*, v. 12, n. 3, p. 104-111, 2014.
32. Cooper, Trevor G. et al. World Health Organization reference values for human semen characteristics. *Human reproduction update*, v. 16, n. 3, p. 231-245, 2010
33. Jungwirth, A. et al. *EAU Guidelines on*. 2017
34. Mcgriff, Sarah C. et al. Optimal Endocrine Evaluation and Treatment of Male Infertility. *Urol Clin North Am*, v. 47, n. 2, p. 139-146, 2020
35. Meeker, John D.; Godfrey-bailey, Linda; Hauser, Russ. Relationships between serum hormone levels and semen quality among men from an infertility clinic. *Journal of andrology*, v. 28, n. 3, p. 397-406, 2007.
36. Barbonetti, A. et al. The use of follicle stimulating hormone (FSH) for the treatment of the infertile man: position statement from the Italian Society of Andrology and Sexual Medicine (SIAMS). *Journal of endocrinological investigation*, v. 41, n. 9, p. 1107-1122, 2018.
37. Belay RE, Huang GO, Shen JK, Ko EY. Diagnosis of clinical and subclinical varicocele: how has it evolved? *Asian journal of andrology*. 2016;18(2):182-5
38. SU, Johnny S.; Farber, Nicholas J.; VIJ, Sarah C. Pathophysiology and treatment options of varicocele: An overview. *Andrologia*, v. 53, n. 1, p. e13576,

2021. Damsgaard, J. et al. Varicocele is associated with impaired semen quality and reproductive hormone levels: a study of 7035 healthy young men from six European countries. *Eur. Urol.* 70, 1019–1029 (2016)
- 39.** Clavijo, Raul I., Robert Carrasquillo, and Ranjith Ramasamy. Varicoceles: prevalence and pathogenesis in adult men. *Fertility and sterility* 108.3 (2017): 364-369
- 40.** Jensen, Christian Fuglesang S. et al. Varicocele and male infertility. *Nature Reviews Urology*, v. 14, n. 9, p. 523, 2017.
- 41.** Tanaka, Takashi et al. Seminal oxidation–reduction potential and sperm DNA fragmentation index increase among infertile men with varicocele. *Human Fertility*, p. 1-5, 2020.
- 42.** Dave, Priya; Farber, Nicholas; Vij, Sarah. Conventional semen analysis and advanced sperm function tests in diagnosis and management of varicocele. *Andrologia*, v. 53, n. 2, p. e13629, 2021.
- 43.** Tulloch, W. Selby. A consideration of sterility factors in the light of subsequent pregnancies. II. Subfertility in the male. *Edinburgh medical journal*, v. 59, n. 3, p. T29, 1952.
- 44.** Alsaikhan, Bader et al. Epidemiology of varicocele. *Asian journal of andrology*, v. 18, n. 2, p. 179, 2016.
- 45.** Lundy, Scott D.; Sabanegh Jr, Edmund S. Varicocele management for infertility and pain: a systematic review. *Arab journal of urology*, v. 16, n. 1, p. 157-170, 2018.
- 46.** Alharbi, Mohannad; ZINI, Armand. Epidemiology of Varicocele in Pediatric, Adolescent, and Adult Populations. In: *Varicocele and Male Infertility*. Springer, Cham, 2019. p. 97-106.

47. Hamada, A., Esteves, S. C. & Agarwal, A. *Varicocele and Male Infertility: Current Concepts, Controversies and Consensus* (Springer International Publishing, 2016)
48. Gat, Y., Bachar, G. N., Zukerman, Z., Belenky, A. & Gornish, M. Varicocele: a bilateral disease. *Fertil. Steril.* 81, 424–429 (2004)
49. Allamaneni, S. S., Naughton, C. K., Sharma, R. K., Thomas, A. J. Jr & Agarwal, A. Increased seminal reactive oxygen species levels in patients with varicoceles correlate with varicocele grade but not with testis size. *Fertil. Steril.* 82, 1684–1686 (2004).
50. Hurtado de Catalfo, G. E., Ranieri-Casilla, A., Marra, F. A., de Alaniz, M. J. & Marra, C. A. Oxidative stress biomarkers and hormonal profile in human patients undergoing varicocelectomy. *Int. J. Androl.* 30, 519–530 (2007).
51. Romeo, C. & Santoro, G. Free radicals in adolescent varicocele testis. *Oxid. Med. Cell. Longev.* 2014, 912878 (2014)
52. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., ... & Merrick, J. M. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269.5223 (1995): 496-512.
53. Cerami, Ethan, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. (2012): 401-404.
54. Brenner, Steven E. A tour of structural genomics. *Nature Reviews Genetics*, 2.10 (2001): 801.
55. Fiehn, Oliver, Sebastian Kloska, and Thomas Altmann. Integrated studies on plant biology using multiparallel techniques. *Current opinion in Biotechnology*, 12.1 (2001): 82-86
56. Dunn, Warwick B., and David I. Ellis. Metabolomics: current analytical platforms and methodologies. *TrAC Trends in Analytical Chemistry*, 24.4 (2005): 285-294.

57. Goodwin, Sara, John D. McPherson, and W. Richard McCombie. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17.6 (2016): 333.
58. Joyce, Andrew R., and Bernhard Ø. Palsson. The model organism as a system: integrating 'omics' data sets. *Nature reviews Molecular cell biology*, 7.3 (2006): 198.
59. López, E., Madero, L., López-Pascual, J., & Latterich, M, Clinical proteomics and OMICS clues useful in translational medicine research. *Proteome Science*, 10.1 (2012): 35
60. Barbosa, E. B., Vidotto, A., Polachini, G. M., Henrique, T., de Marqui, A. B. T., & Tajara, E. H. Proteômica: metodologias e aplicações no estudo de doenças humanas. *Revista da Associação Médica Brasileira*, 58.3, (2012). 366-375.
61. Huang, Zhao, et al. Proteomic profiling of human plasma for cancer biomarker discovery. *Proteomics* 17.6 (2017): 1600240.
62. Borrebaeck, Carl AK. Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer. *Nature Reviews Cancer* 17.3 (2017): 199
63. Bollard, M. E., Stanley, E. G., Lindon, J. C., Nicholson, J. K., & Holmes, E. . NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In vivo*. 18.3 (2005): 143-162
64. Godoy, M. M. G., Lopes, E. P. A., Silva, R. O., Hallwass, F., Koury, L. C. A., Moura, I. M., Simas, A. M. Hepatitis C virus infection diagnosis using metabonomics. *Journal of viral hepatitis*, 17.12, (2010). 854-858.
65. Nicholson, Jeremy K., John C. Lindon, and Elaine Holmes. "'Metabonomics': understanding the metabolic responses of living systems to pathophysiological

- stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*. 9.11 (1999): 1181-1189
66. Lindon, J. C., Nicholson, J. K., Holmes, E., & Everett, J. R. Metabonomics: metabolic processes studied by NMR spectroscopy of biofluids. *Concepts in Magnetic Resonance: An Educational Journal*, 12.5, (2000) 289-320
67. Nicholson, Jeremy K., and John C. Lindon. Systems biology: metabonomics. *Nature* 455.7216 (2008): 1054
68. Fiehn, Oliver. Metabolomics—the link between genotypes and phenotypes. *Functional genomics*. Springer, Dordrecht, (2002). 155-171
69. Cerami, Ethan, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. (2012): 401-404
70. Jayaraman, V., Ghosh, S., Sengupta, A., Srivastava, S., Sonawat, H. M., & Narayan, P. K. Identification of biochemical differences between different forms of male infertility by nuclear magnetic resonance (NMR) spectroscopy. *Journal of assisted reproduction and genetics*, 31.9, (2014): 1195-1204.
71. Hamamah, F. Seguin, L. Bujan, C. Barthelemy, R. Mieusset, J. Lansac, Quantification by magnetic resonance spectroscopy of metabolites in seminal plasma able to differentiate different forms of azoospermia, *Hum. Reprod.* 13 (1998) :132–135
72. QIAO, S., Wu, W., Chen, M., Tang, Q., Xia, Y., Jia, W., & Wang, X. Seminal plasma metabolomics approach for the diagnosis of unexplained male infertility. *PloS one*. 12.8 (2017): e 0181115
73. Gupta, Ashish, et al. ¹H NMR spectroscopic studies on human seminal plasma: a probative discriminant function analysis classification model. *Journal of pharmaceutical and biomedical analysis*. 54.1 (2011): 106-113.

74. BEEBE, Kenneth R.; PELL, Randy J.; SEASHOLTZ, Mary Beth. Chemometrics a practical guide. 1998.
75. NETO, Benício Barros; SCARMINIO, Ieda Spacino; BRUNS, Roy Edward. Como Fazer Experimentos-: Pesquisa e Desenvolvimento na Ciência e na Indústria. Bookman Editora, 2010.
76. BLAISE, Benjamin J. et al. Statistical analysis in metabolic phenotyping. Nature Protocols, p. 1-28, 2021.
77. FERREIRA, Marcia Miguel Castro. Quimiometria: conceitos, métodos e aplicações. Editora da UNICAMP, 2015.
78. PASSOS, João Octávio Sales et al. Spectrochemical analysis in blood plasma combined with subsequent chemometrics for fibromyalgia detection. Scientific Reports, v. 10, n. 1, p. 1-8, 2020.
79. ROBOTTI, Elisa; MARENGO, Emilio. Chemometric multivariate tools for candidate biomarker identification: LDA, PLS-DA, SIMCA, Ranking-PCA. In: 2-D PAGE Map Analysis. Humana Press, New York, NY, 2016. p. 237-267.
80. COSTA FILHO, Paulo Augusto da; POPPI, Ronei Jesus. Algoritmo genético em química. Química Nova, v. 22, p. 405-411, 1999. Genetic algorithms combined with discriminant analysis for key variable identification
81. GROMSKI, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., & Goodacre, R. A tutorial review: Metabolomics and partial least squares-discriminant analysis—a marriage of convenience or a shotgun wedding. Analytica chimica acta, 879, (2015): 10-23
82. WORLEY, Bradley, and Robert Powers. Multivariate analysis in metabolomics. Current Metabolomics 1.1 (2013): 92-107.

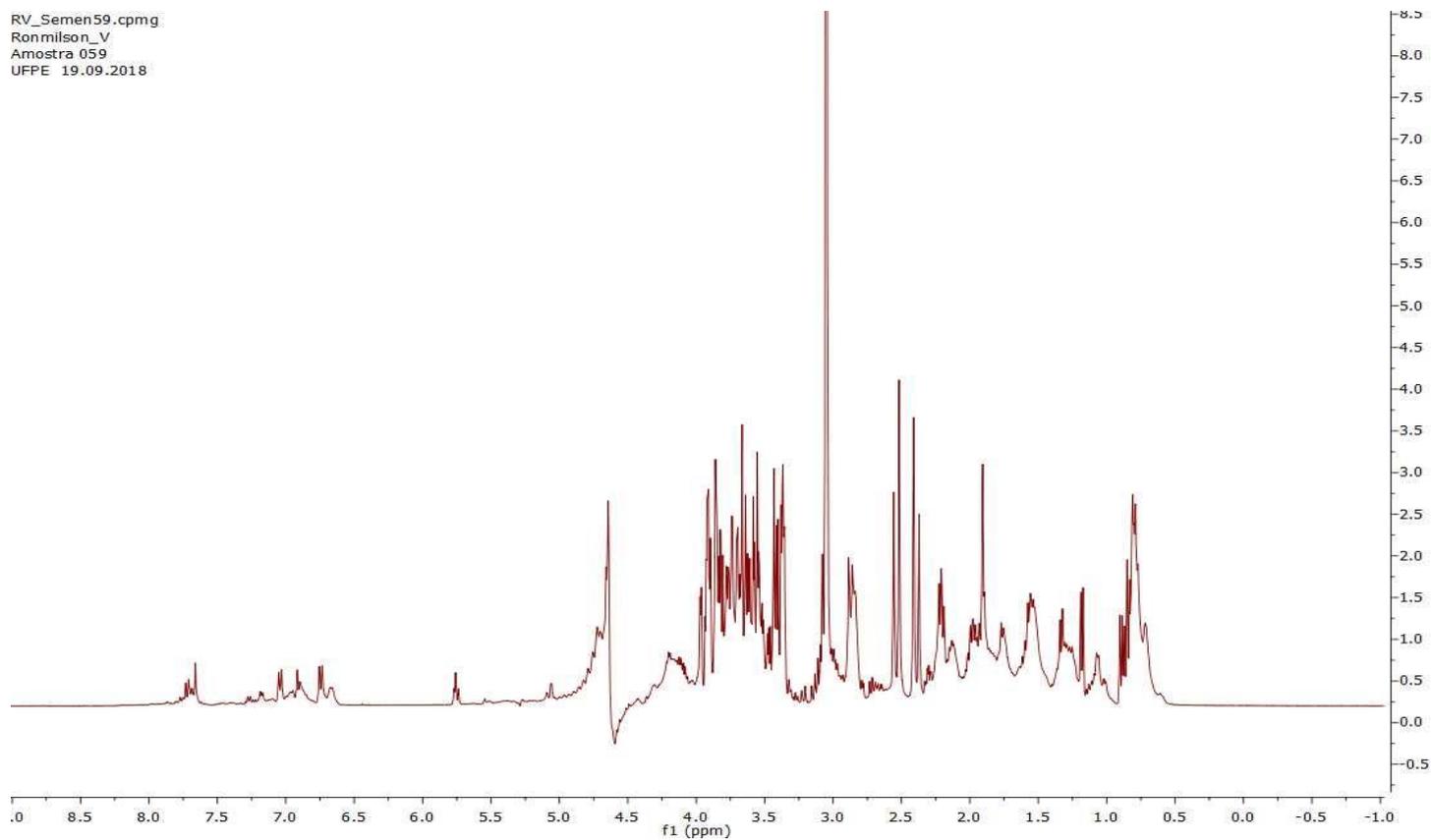
83. WESTERHUIS, J. A., Hoefsloot, H. C., Smit, S., Vis, D. J., Smilde, A. K., van Velzen, E. J., ... & van Dorsten, F. A. Assessment of PLSDA cross validation. *Metabolomics*. 4.1 (2008): 81-89
84. KIZILAY, Fuat; ALTAY, Barış. Evaluation of the effects of antioxidant treatment on sperm parameters and pregnancy rates in infertile patients after varicocelelectomy: a randomized controlled trial. *International journal of impotence research*, v. 31, n. 6, p. 424-431, 2019.
85. LARA-CERRILLO, Sandra et al. Microsurgical varicocelelectomy effect on sperm telomere length, DNA fragmentation and seminal parameters. *Human fertility*, p. 1-7, 2020.
86. ÖZTEKİN, Ünal et al. Evaluation of the influence of subinguinal varicocelelectomy procedure on seminal parameters, reproductive hormones and testosterone/estradiol ratio. *Archivio Italiano di Urologia e Andrologia*, v. 92, n. 2, 2020.
87. ASAFU-ADJEI, Denise et al. Systematic review of the impact of varicocele grade on response to surgical management. *The Journal of urology*, v. 203, n. 1, p. 48-56, 2020.
88. PALLOTTI, F. et al. Varicocele and semen quality: a retrospective case–control study of 4230 patients from a single centre. *Journal of endocrinological investigation*, v. 41, n. 2, p. 185-192, 2018.
89. VIVAS-ACEVEDO G, Lozano JR, Camejo MI (2010) Effect of varicocele grade and age on seminal parameters. *Urol Int* 85:194–199
90. ALKAN, İlter et al. Superoxide anion production by the spermatozoa of men with varicocele: relationship with Varicocele grade and semen parameters. *The world journal of men's health*, v. 36, n. 3, p. 255-262, 2018.

91. MACHEN, Graham Luke et al. Time to improvement of semen parameters after microscopic varicocelectomy: When it occurs and its effects on fertility. *Andrologia*, v. 52, n. 2, p. e13500, 2020.
92. RIBEIRO, Abene Silva et al. Discriminating gamma-irradiated soybean seeds by ¹H NMR-based metabolomics. *Food control*, v. 36, n. 1, p. 266-272, 2014.
93. GOUVEIA, Liana Ribeiro et al. Diagnosis of coinfection by schistosomiasis and viral hepatitis B or C using ¹H NMR-based metabolomics. *Plos one*, v. 12, n. 8, p. e0182196, 2017.
94. PONTES, T. A. et al. Osteopenia-osteoporosis discrimination in postmenopausal women by ¹H NMR-based metabolomics. *PloS one*, v. 14, n. 5, p. e0217348, 2019.
95. HACKSTADT, Amber J.; HESS, Ann M. Filtering for increased power for microarray data analysis. *BMC bioinformatics*, v. 10, n. 1, p. 1-12, 2009.
96. SZYMAŃSKA, Ewa et al. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics*, v. 8, n. 1, p. 3-16, 2012.
97. RYU, Young Jin et al. A preliminary study of shear-wave elastography for the evaluation of varicocele in adolescents and young adults. *Ultrasonography*, v. 41, n. 1, p. 131, 2022.
98. ZYLBERSZTEJN, Daniel Suslik et al. Proteomic analysis of seminal plasma in adolescents with and without varicocele. *Fertility and sterility*, v. 99, n. 1, p. 92-98, 2013.
99. DEL GIUDICE, Paula Toni et al. Determination of testicular function in adolescents with varicocoele—a proteomics approach. *Andrology*, v. 4, n. 3, p. 447-455, 2016.
100. WISHART, David S. et al. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic acids research*, v. 50, n. D1, p. D622-D631, 2022.

101. NETO, Filipe Tenorio Lira et al. ¹H NMR-based metabonomics for infertility diagnosis in men with varicocele. *Journal of assisted reproduction and genetics*, v. 37, n. 9, p. 2233-2247, 2020.
102. ABDELBAKI, Shabieb A. et al. The impact of coexisting sperm DNA fragmentation and seminal oxidative stress on the outcome of varicocelectomy in infertile patients: A prospective controlled study. *Arab journal of urology*, v. 15, n. 2, p. 131-139, 2017.
103. BIROWO, Ponco et al. The effects of varicocelectomy on the DNA fragmentation index and other sperm parameters: a meta-analysis. *Basic and Clinical Andrology*, v. 30, n. 1, p. 1-9, 2020.
104. DUBIN, Lawrence; AMELAR, Richard D. Varicocelectomy: 986 cases in a twelve-year study. *Urology*, v. 10, n. 5, p. 446-449, 1977.
105. VERMEULEN, Alex; VERDONCK, Lieve; KAUFMAN, Jean M. A critical evaluation of simple methods for the estimation of free testosterone in serum. *The Journal of Clinical Endocrinology & Metabolism*, v. 84, n. 10, p. 3666-3672, 1999.
106. EL OUARTIGHI, A.; EL AKADI, A.; ABOUTAJDINE, D. Feature selection on supervised classification using Wilks lambda statistic. In: 2007 International Symposium on Computational Intelligence and Intelligent Informatics. IEEE, 2007. p. 51-55.
107. VERMIGLIO, Andrew J. The gold standard and auditory processing disorder. *Perspectives of the ASHA Special Interest Groups*, v. 3, n. 6, p. 6-17, 2018.
108. LUQUE, Amalia et al. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, v. 91, p. 216-231, 2019.

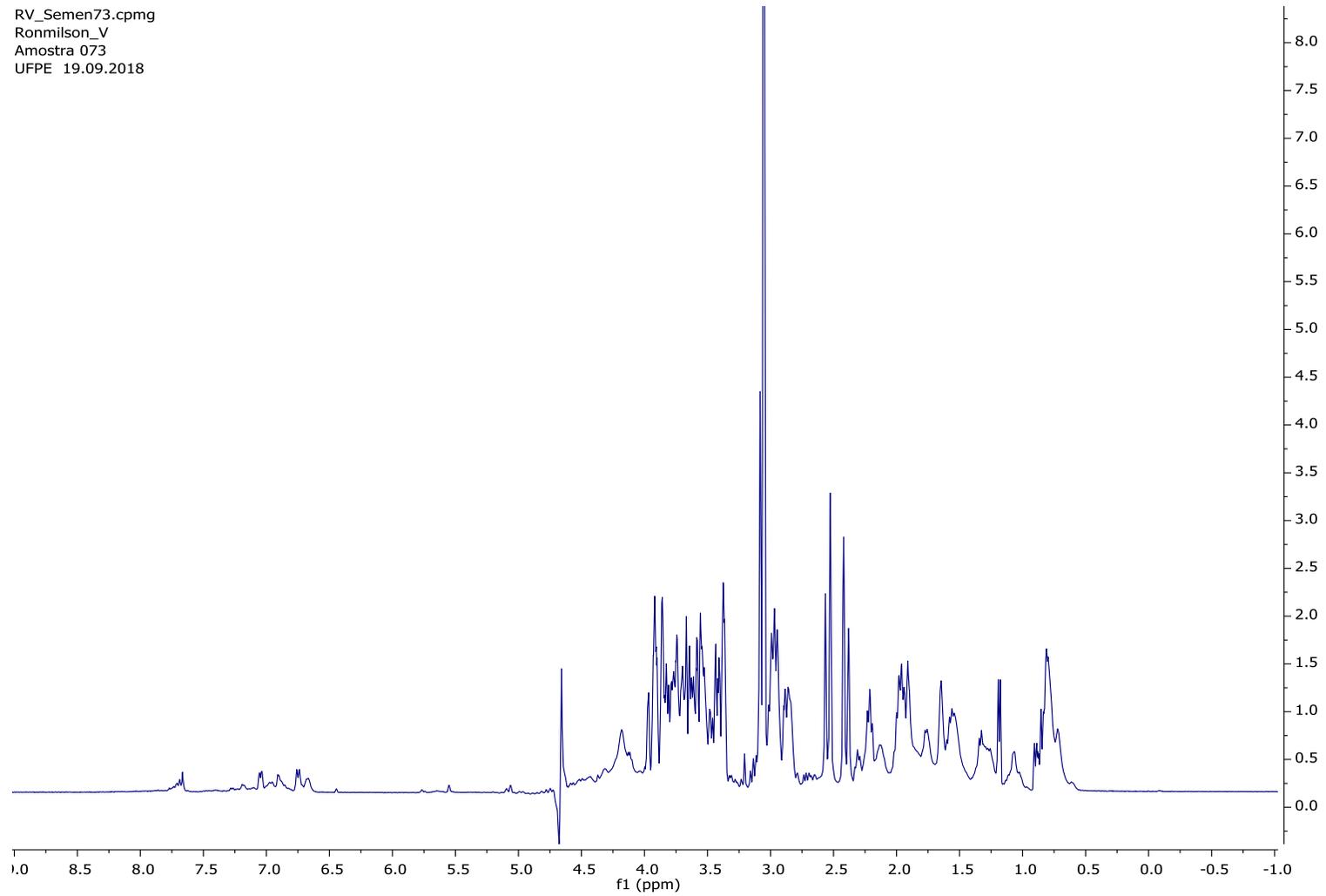
- 109.** THARWAT, Alaa. Classification assessment methods. Applied Computing and Informatics, 2020.

APÊNDICE A – ESPECTROS DE RMN DE ALGUMAS AMOSTRAS



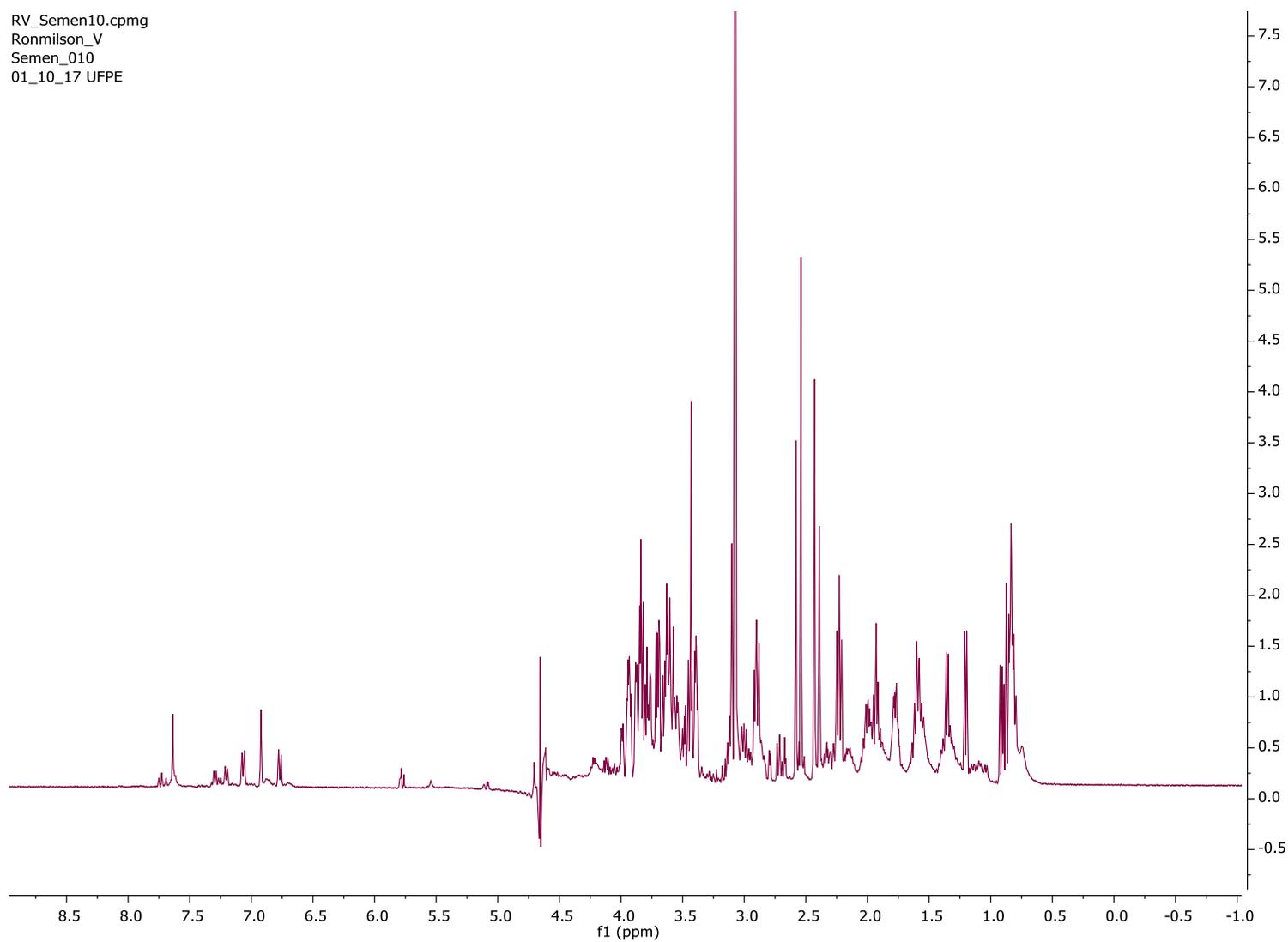
Espectro de RMN de ^1H (D_2O , CPMG-Presat, 400 MHz) da amostra 59 (Grupo Varicocele Infértil)

RV_Semen73.cpmg
Ronmilson_V
Amostra 073
UFPE 19.09.2018



Espectro de RMN de ¹H (D₂O, CPMG-Presat, 400 MHz) da amostra 73 (Grupo Varicocele Fértil)

RV_Semen10.cpmg
Ronmilson_V
Semen_010
01_10_17 UFPE



Espectro de RMN de ^1H (D_2O , CPMG-Presat, 400 MHz) da amostra 10 (Grupo Controle)

APÊNDICE B – ARTIGO PUBLICADO NO JOURNAL OF ASSISTED REPRODUCTION AND GENETICS

Journal of Assisted Reproduction and Genetics (2020) 37:2233–2247
<https://doi.org/10.1007/s10815-020-01896-2>

REPRODUCTIVE PHYSIOLOGY AND DISEASE



1H NMR-based metabonomics for infertility diagnosis in men with varicocele

Filipe Tenorio Lira Neto^{1,2}  · Ronmilson Alves Marques³ · Alexandre de Freitas Cavalcanti Filho² ·
Leslie Clifford Noronha Araujo⁴ · Salvador Vilar Correia Lima⁵ · Licarion Pinto⁶ · Ricardo Oliveira Silva⁶

Received: 15 October 2019 / Accepted: 16 July 2020 / Published online: 26 July 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Purpose “Omics” techniques have been used to understand and to identify biomarkers of male infertility. We report on the first metabonomics models created to diagnose varicocele and infertility among men with varicocele.

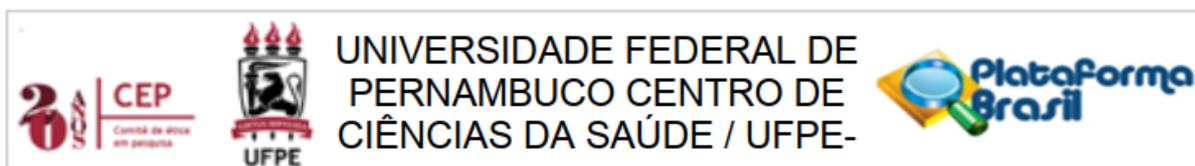
Methods We recruited 35 infertile men with varicocele (VI group), 21 fertile men with varicocele (VF group) and 24 fertile men without varicocele (C group). All men underwent standard semen analysis, scrotal duplex ultrasonography, and sexual hormone level measurement. Hydrogen-1 nuclear magnetic resonance (¹H NMR) spectra of seminal plasma were used to create metabonomics models to discriminate between men with and without varicocele, and between fertile and infertile men with varicocele.

Results Using the statistical formalisms partial least square discriminants analysis and genetic algorithm–based linear discriminant analysis (GA-LDA), we created two models that discriminated the three groups from each other with accuracy of 92.17%. We also created metabonomics models using orthogonal partial least square discriminants analysis and GA-LDA that discriminated VF group from VI group, with an accuracy of 94.64% and 100% respectively. We identified 19 metabolites that were important in group segregation: caprate, 2-hydroxy-3-methylvalerate, leucine, valine, 3-hydroxybutyrate, lactate, alanine, 4-aminobutyrate, isoleucine, citrate, methanol, glucose, glycosides, glycerol-3-phosphocoline, n-acetyltyrosine, glutamine, tyrosine, arginine, and uridine.

Conclusions ¹HNMR-based metabonomics of seminal plasma can be used to create metabonomics models to discriminate between men with varicocele from those without varicocele, and between fertile men with varicocele from those infertile with varicocele. Furthermore, the most important metabolites for group segregation are involved in the oxidative stress caused by varicocele.

Keywords Metabolome · Metabolomics · Fertility · Semen · NMR spectroscopy

ANEXO A – PARECER CONSUBSTANCIADO DO COMITÊ DE ÉTICA



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Investigação da Varicocele Através de Análise Metabonômica Baseada na Espectroscopia de RMN de Hidrogênio de Soro de Sêmen, Soro de Sangue, e de Urina

Pesquisador: Filipe Tenório Lira Neto

Área Temática:

Versão: 2

CAAE: 64211917.1.0000.5208

Instituição Proponente: CENTRO DE CIÊNCIAS DA SAÚDE

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 1.978.735

Apresentação do Projeto:

Trata-se de um Projeto de Pesquisa do Mestrado em Cirurgia do Programa de Pós-graduação em Cirurgia apresentado pelo aluno Felipe Tenório Lira Neto, sob a orientação do Prof.Dr. Salvador Vilar Correia Lima, cujo objetivo principal é proceder a Investigação da Varicocele Através de Análise Metabonômica Baseada na Espectroscopia de RMN de Hidrogênio de Soro de Sêmen, Soro de Sangue e de Urina.

A varicocele é citada como causa de 35% a 80% dos casos de infertilidade masculina, no entanto, a relação de causa-efeito entre varicocele e infertilidade masculina ainda não está esclarecida não obstante os estudos revelando melhora da fertilidade quando homens inférteis são submetidos à varicocelectomia.

Desde o ano 2000, estudos metabonômicos têm sido utilizados para criar modelos que sejam capazes de auxiliar o diagnóstico e prognóstico das mais diferentes doenças. A construção de um modelo metabonômico por um meio não invasivo como a RMN de H-1 (avalia as alterações dos metabólitos

Endereço: Av. da Engenharia s/nº - 1º andar, sala 4, Prédio do Centro de Ciências da Saúde
Bairro: Cidade Universitária **CEP:** 50.740-600
UF: PE **Município:** RECIFE
Telefone: (81)2126-8588 **E-mail:** cepccs@ufpe.br