

**Aplicação de metodologia *Next Generation Sequencing* na pesquisa de determinantes genéticos da febre Q crónica na via de sinalização IL12-IFN $\gamma$**   
(Versão final após defesa)

**Ulisses Emanuel Rodrigues Gaspar**

Dissertação para obtenção do Grau de Mestre em  
**Ciências Biomédicas**  
2<sup>o</sup> ciclo de estudos

Orientadora: Doutora Susana Correia de Matos David  
Coorientadora: Prof<sup>a</sup>. Doutora Maria Petronila Jorge Frade Rocha-Pereira

**Abril de 2021**

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crónica na via de sinalização IL12-IFN $\gamma$***

## **Dedicatória**

The pictures tell the story  
This life has many shades  
I'd wake up every morning and before I'd start each day  
I'd take a drag from last night's cigarette  
That smoldered in it's tray  
Down a little something and then be on my way

I traveled far and wide  
And laid this head in many ports  
I was guided by a compass  
I saw beauty to the north  
I drew the tales of many lives  
And wore the faces of my own  
I had these memories all around me  
So I wouldn't be alone

These scares, some may be from showing up  
Others are from growing up  
Sometimes I was so messed up and didn't have a clue  
I ain't winning no one over  
I wear it just for you  
I've got **my** name written here  
like a rose tattoo

Adaptado de *Dropkick murphys - Rose tattoo*

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crónica na via de sinalização IL12-IFN $\gamma$***

## **Agradecimentos**

Desde já gostaria de agradecer à Dra. Glória Isidro, Coordenadora do Departamento de Genética Humana (DGH), e ao Professor Doutor Fernando de Almeida, Presidente do Conselho Directivo<sup>1</sup> do INSA, IP (Instituto Nacional de Saúde Dr. Ricardo Jorge, Instituto Público), pela oportunidade de realização deste trabalho.

À minha orientadora Doutora Susana David, agradeço todo o empenho e dedicação em ensinar-me, a sua sensibilidade e compreensão, a sua garra e força contra as adversidades. Agradeço também pela oportunidade e voto de confiança em me incluir na realização de uma parte deste trabalho. que se insere no seu projeto científico “Desenvolvimento e avaliação da distribuição na população de preditores genéticos precoces de doença infecciosa”.

Agradeço à Doutora Ana Santos, do Centro de Estudos de Vetores e Doenças Infecciosas (CEVDI), detentora do projeto “Febre Q: do diagnóstico à pesquisa eco epidemiológica de *Coxiella burnetii* no contexto da infeção humana”, financiado pela Fundação de Ciência e Tecnologia (financiamento: PTDC/SAU-SAP/115266/2009), no qual também se insere este trabalho. Agradeço também por todo o apoio que me deu através do seu vasto conhecimento sobre *Coxiella burnetii*, o agente da febre Q.

À Professora Doutora Petronila Rocha-Pereira, que considero a “faísca explosiva” desta iniciativa, por me ensinar, motivar, empurrar e encorajar. Dentro do que lhe foi possível esteve sempre disposta a ajudar-me, inclusive por ser a pioneira em contactar o INSA, IP para desenvolver um projecto académico. O seu voto de confiança motivou-me a mostrar a qualidade da Universidade da Beira Interior.

Em conjunto agradeço ainda no INSA, IP, ao Dr. Luís Vieira, responsável da Unidade de Tecnologia e Inovação, e à sua equipa Joana Mendonça no laboratório, e ao Miguel Machado e José Ferrão no processo bioinformático em me deixarem acompanhar todo o processo da realização da sequenciação de nova geração; À equipa da Unidade de Genética Molecular pelo apoio laboratorial nomeadamente nas extrações de DNA das amostras biológicas, incluindo a Susana Gomes e Rita Certã por me deixarem acompanhar o processo.

---

<sup>1</sup> Conforme acordo ortográfico Português de 1990

Agradeço bastante aos Serviços de Ação Social tanto da Universidade da Beira interior na pessoa da Dra. Margarida de Sousa e Dr. Fernando Fernandes, como da Universidade de Lisboa na pessoa da Dra. Rita Casquilho e Dra. Maria Macau, sem a sua dedicação, profissionalismo e cuidado pelos alunos eu nunca teria conseguido conquistar este grau, onde quer que os alunos cheguem vão sempre lembrar que sem o vosso empenho não seria possível alcançar outros patamares. O meu sincero obrigado por mim e por todos os alunos, pelo serviço que prestam em assegurar a qualidade do nosso ensino.

Aos serviços administrativos, da Universidade da Beira Interior e do INSA, IP por me ajudarem a tornar possível esta parceria tão desafiante, espero que deste acordo nasçam outros projectos que façam avançar a investigação nacional.

Duma perspectiva pessoal, em primeiro lugar, agradeço a Deus por todas as graças e desafios que pôs no meu caminho, ao Espírito Santo em me munir de força e ponderação todos os dias, inclusive por reforçar que desistir nunca seria uma opção.

Agradeço à minha família por, mesmo desconfiando, confiar que eu iria cuidar bem de mim mesmo. Especialmente agradeço por saber que no pior dos cenários teria sempre um sítio seguro onde voltar, com uma porta aberta, uma sopa quente em cima da mesa e uns braços de mãe para me receber. Agradeço ainda aos meus primos Neuza e Roberto por todas as vezes que me aconselharam, arrancaram de casa, proporcionaram experiências únicas e apresentaram a cidade de Lisboa na sua melhor vertente para que esta experiência fosse mais que um “Pip”.

Dos amigos que seguiram comigo, e das amigadas que cruzei nesta etapa, agradeço a todos aqueles que me guardaram, se preocuparam comigo, com chamadas, mensagens e com “quando voltas cá?”, ou “vem jantar ao 3º!”, e tantos outros momentos que guardo e levo de pessoas tão especiais; por vezes a frase mais simples torna-se a mais motivadora por ser verdadeira.

Estendo ainda estes agradecimentos a todos os que contribuíram da forma mais singela para a realização desta dissertação e contribuíram para o meu crescimento académico.

“Um estômago com fome, um bolso vazio e um coração partido ensinam as melhores lições da vida”

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$***

## Resumo

A febre Q é uma zoonose causada por *Coxiella burnetii*, na qual cerca de 60% dos doentes apresentam quadros clínicos de infecção aguda, com recuperação após intervenção clínica. No entanto, nos casos de febre Q crónica, a infecção culmina em complicações graves, frequentemente endocardites. Mesmo com medidas preventivas, a incidência da febre Q tem vindo a aumentar, principalmente na Península Ibérica. A fisiopatologia da infecção por *C. burnetii* assemelha-se à do *Mycobacterium tuberculosis*. Em ambas as infecções, a via interleucina12—interferão gama (IL12-IFN $\gamma$ ) desempenha um papel importante na sobrevivência e replicação intracelular bacteriana. As doenças causadas por micobactérias apresentam homogeneidade fisiológica inerente à deficiência de interferão gama (IFN $\gamma$ ), o que é fundamentado pela identificação de determinantes genéticos dessas infecções na via IL12-IFN $\gamma$ . No entanto, esta via permanece praticamente inexplorada em termos de associação genética com a febre Q crónica, enaltecendo a necessidade destes estudos. Assim, nesta investigação procurou-se identificar variantes associadas a febre Q crónica através de uma abordagem por estudos populacionais de associação genótipo fenótipo com base em genes candidatos (*Candidate Gene Association Studies* - CGAS).

Esta abordagem insere-se numa arquitetura de estudo caso-controlo. Selecionaram-se genes candidatos e optou-se pela utilização de uma estratégia de sequenciação de nova geração (*Next Generation Sequencing* – NGS) para a genotipagem das regiões 5'UTR e os exões dos genes candidatos. Após consentimento informado, foram recrutados 34 indivíduos saudáveis e 60 doentes com febre Q (43 formas agudas e 17 formas crónicas). Os genes do receptor  $\beta$ 1 de interleucina-12 (*IL12RB1*), receptor 1 de interferão gama (*IFN $\gamma$ R1*) e receptor 2 de interferão gama (*IFN $\gamma$ R2*), com fragmentos de 2kb a 17kb, foram amplificados por polimerase de reação em cadeia (PCR) de longo alcance com ácido desoxirribonucleico (ADN) polimerase Takara LA Taq® para construção de bibliotecas Nextera e sequenciação com tecnologia da Illumina®. Uma vez que a NGS trabalha com milhares de milhões de dados em simultâneo requer a aplicação de várias métricas de controlo de qualidade de modo a garantir resultados fidedignos. Para além destas métricas foram ainda implementadas métricas específicas para CGAS que têm por objectivo diminuir a obtenção de erros por falsos positivos na análise estatística.

Do presente trabalho de investigação resultou uma lista de polimorfismos de nucleótido único (*Single Nucleotide Polymorphism* - SNP) dos genes *IL12RB1*, *IFN $\gamma$ R1* e *IFN $\gamma$ R2* em indivíduos com febre Q crónica, aguda e numa amostra de conveniência de indivíduos

saudáveis. Estes dados têm a necessária qualidade para constituir uma mais valia, podendo ser aplicados a diferentes estudos.

Futuramente, esta lista de SNPs será usada em estudos de associação genótipo-fenótipo para compreender se existem variantes na via IL12-IFN $\gamma$  associadas à apresentação clínica doença. Assim, estes resultados poderão vir a ser um importante avanço clínico não só para antecipar formas crónicas da doença como para o desenvolvimento de estratégias de prevenção e para prestação de melhores cuidados de saúde a estes doentes.

## **Palavras-chave**

Febre Q; *Coxiella burnetii*; variação genética do hospedeiro humano; imunidade inata; genética humana da suscetibilidade à infeção; via biológica IL12-IFN $\gamma$ ; sequenciação de nova geração; controlo de qualidade de dados genotípicos; estudos populacionais de associação genótipo-fenótipo com base em genes candidatos

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crónica na via de sinalização IL12-IFN $\gamma$***

## **Abstract**

Q fever is a zoonosis caused by *Coxiella burnetii*, in which about 60% of the affected patients develop acute Q fever, recovering after clinical intervention. However, in chronic Q fever cases, the disease is accompanied with serious complications, mainly manifested as endocarditis. Even with preventive measures, the incidence of Q fever has been increasing, namely in the Iberian Peninsula. The pathophysiology of *C. burnetii* infection resembles that of *Mycobacterium tuberculosis*, both characterized by granuloma formation. In this context, the interleukin 12 - interferon  $\gamma$  (IL12-IFN $\gamma$ ) pathway plays a major role in intracellular bacterial survival and replication. Diseases caused by mycobacteria display physiological homogeneity inherent to interferon  $\gamma$  (IFN- $\gamma$ ) deficiency, since genetic determinants of these infections have been identified in the IL12-IFN $\gamma$  pathway. Despite of the above mentioned, the association between genetic variants this pathway and chronic Q fever remains practically unexplored. Thus, in this investigation we seek to establish a list of variants within the IL12-IFN $\gamma$  pathway with pathogenic potential associated with the development of chronic Q fever. therefore, we sought to identify variants associated with chronic Q fever resorting to candidate-gene, population association analysis in a Candidate Gene Association Studies (CGAS) approach.

This approach is part of a case-control study architecture. Candidate genes were selected and a Next Generation Sequencing (NGS) strategy was chosen for the genotyping of the 5'UTR region and exons of candidate genes. After informed consent, 34 healthy individuals and 60 Q fever patients (43 acute and 17 chronic forms) were recruited. The genes interleukin-12 receptor  $\beta$ 1 (*IL12RB1*), interferon  $\gamma$  receptor 1 (*IFN $\gamma$ R1*) and interferon  $\gamma$  receptor 2 (*IFN $\gamma$ R2*), were amplified in 2kb to 17kb fragments, by long-range polymerase chain reaction (PCR) with Takara LA Taq® deoxyribonucleic acid (DNA) polymerase for Nextera library construction and Illumina® sequencing. Since NGS works simultaneously with billions of data points, it requires the application of several quality control metrics in order to guarantee reliable results. In addition to these metrics, specific metrics for CGAS were also implemented to reduce possible false positives in the statistical analysis of the data.

The present research work provided a list of single nucleotide polymorphisms (SNPs) of the *IL12RB1*, *IFN $\gamma$ R1* and *IFN $\gamma$ R2* genes, observed in patients with chronic Q fever, acute Q fever and in a convenience sample of healthy individuals. These data have the necessary quality to constitute an asset and can be applied to different studies.

In the future, this list of SNPs will be used in genotype-phenotype association studies to understand whether there are variants in the IL12-IFN $\gamma$  pathway associated with the clinical presentation of the disease. Thus, these results could prove to be an important clinical advance not only to anticipate chronic forms of the disease but also to develop prevention strategies and to provide better health care to these patients.

## **Keywords**

Q Fever; *Coxiella burnetii*; Genetic variation of the human host; Innate immunity; Human genetics of susceptibility to infection; IL12-IFN $\gamma$  biological pathway; Next generation sequencing; Quality control of genotyping data; Candidate gene association studies (CGAS)

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$***

## **Lista de publicações**

### **Poster:**

U Gaspar, AS Santos, JC Robles-Fuentes, J Lavinha, S David, IL12-IFN $\gamma$  pathway, a possible genetic determinant to chronic Q fever? 19th Portugaliae Genetica Phenotypic Manifestation of Genetic Heterogeneity, I3S, Porto, Portugal – Evento cancelado devido à pandemia de COVID-19.

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$***

# Índice

Dedicatória .....	iii
Agradecimentos.....	v
Resumo.....	viii
Abstract .....	xi
Lista de publicações .....	xiv
Índice.....	xvi
Lista de figuras .....	xviii
Lista de tabelas.....	xx
Lista de acrónimos .....	xxii
Capítulo I.....	1
1. Febre Q.....	1
1.1. Etiopatologia.....	2
1.2. Epidemiologia .....	3
1.3. Vias de transmissão .....	4
1.4. Aspetos clínicos.....	4
1.5. Diagnóstico laboratorial .....	6
1.6. Factores de risco .....	7
2. Resposta imunológica .....	8
2.1. Imunidade inata.....	8
2.2. Imunidade adaptativa.....	9
2.3. Eixo IL-12 IFN $\gamma$ .....	10
3. Modificadores genéticos.....	13
4. Projeto de estudo.....	14
4.1. Estudo populacional de associação genótipo-fenótipo.....	15
4.2. Sequenciação de nova geração.....	15
Capítulo II .....	20
Capítulo III.....	22
1. Materiais.....	22
2. Métodos.....	23
2.1. Amostras biológicas .....	23
2.2. Preparação NGS .....	24
2.3. Sequenciação NGS.....	27
2.4. Lista de dados - Preparação dos dados de Fase III para associação genótipo-fenótipo por CGAS .....	31
Capítulo IV .....	35
1. Preparação NGS .....	35
2. Fase I e II - Controlo de qualidade.....	38
2.1. <i>Base call</i> .....	38
2.2. <i>Variant call</i> .....	40
3. Fase III - Controlo de qualidade .....	41
3.1. Controlo de qualidade para CGAS .....	41
3.2. Lista de dados para associação .....	46
Conclusão e perspectivas futuras .....	49
Referências.....	51
Anexos .....	64

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$***

## **Lista de figuras**

Figura 1. Cascata de sinalização do receptor de interleucina (IL)-12. ....	11
Figura 2. Cascata de sinalização do receptor de interferão gama (IFN $\gamma$ ). ....	12
Figura 3. Distribuição dos exões por fragmento. ....	24
Figura 4. Organização dos índices. ....	28
Figura 5. Fluxo da metodologia NGS em fases (I, II, III). ....	30
Figura 6. Resultado eletroforese. ....	35
Figura 7. Gráfico de dispersão da percentagem de identificação das leituras em função do índice. ....	38
Figura 8. Dados de qScore. ....	39
Figura 9. Proporção de genótipos em falta por amostra versus taxa de heterozigotia. ...	43
Figura 10. Diagrama de controlo de qualidade para associação genótipo-fenótipo. ....	45
Figura 11. Evolução do número de SNPs ao longo do estudo. ....	46
Figura 12. Exemplo de Anderson e colaboradores. ....	64
Figura 13. Protocolo científico - fluxo de trabalho geral. ....	65
Figura 14. Protocolo AMPure XP. ....	66

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$***

## **Lista de tabelas**

Tabela 1. Resultados de estudos de Seroprevalência de febre Q (IgG) na população geral por vários países.....	4
Tabela 2. Manifestações clínicas da febre Q. ....	6
Tabela 3. Técnicas de diagnóstico de febre Q diretas e indiretas. ....	6
Tabela 4. Fatores de risco associados ao desenvolvimento de febre Q crónica. ....	8
Tabela 5. Protocolo PCR.....	25
Tabela 6. Condições do termociclador consoante o fragmento. ....	26
Tabela 7. Volume dos amplicões adicionados ao agrupamento da amostra de modo a padronizar as concentrações dos produtos PCR.....	27
Tabela 8. Lista de comandos utilizados na ferramenta PLINK para a análise exploratória e controlo de qualidade da lista de dados de NGS e níveis de aceitação (thresholds) correspondentes.....	31
Tabela 9. Dados do gene e amplicão segundo o sistema de coordenadas cromossómicas disponíveis em Ensembl.....	37
Tabela 10. Resultados da genotipagem, em número de SNPs. ....	40
Tabela 11. Resultado da aplicação do tratamento missing.....	42
Tabela 12. SNPs retirados pelo equilíbrio de Hardy-Weinberg. ....	44
Tabela 13. Resultado em número de SNPs após controlo de qualidade para CGAS nos genes da via IL12- IFN $\gamma$ .....	47
Tabela 14. Identificação dos resultados de dispersão de leitura. ....	66
Tabela 15. Lista de acesso processamento via Docker. ....	68

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$***

## Lista de acrónimos

$\mu$ L	Microlitro
ADN	Ácido desoxirribonucleico
CGAS	Estudo de associação de gene candidato; sigla Adaptada do inglês <i>Candidate Gene Association Studies</i>
CR3	Receptor de complemento 3; sigla adaptada do inglês: <i>Complement Receptor 3</i>
CXCL	Citosina C-X-C <i>motif</i> ligando
DNTP	Fosfatados Desoxirribonucleotídeos
DP	Profundidade de leitura
EDTA	Ácido etilenodiaminotetracético do inglês <i>Ethylenediaminetetraacetic acid</i>
HWE	Equilíbrio de Hardy-Weinberg
ELISA	Ensaio de imunoabsorção enzimática; sigla adaptada do inglês <i>enzyme linked immunosorbent assay</i>
g	Gramas
GVCF	<i>Genomic Variant Call Format</i>
IFA	Ensaio de imunofluorescência indireta; sigla adaptada do inglês <i>Immunofluorescence assay</i>
IFN $\gamma$	Interferão gama
IFN $\gamma$ R1	Receptor 1 de interferão gama
IFN $\gamma$ R2	Receptor 2 de interferão gama
Ig	Imunoglobulina
IL	Interleucina

*Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$*

IL12-IFN $\gamma$	Interleucina12 – interferão gama
IL12RB1	Receptor $\beta$ 1 de interleucina 12
INSA, IP	Instituto Nacional de Saúde Dr. Ricardo Jorge, Instituto Público
Jak	Janus cinase
LPS	Lipopolisacárido
mA	Miliampere
MAF	Frequência do alelo minoritário
mL	Mililitro
MMP	Metaloproteinase membrana
MPM	Marcador de peso molecular
NCBI	Centro nacional para informação biotecnológica; sigla adaptada do inglês <i>National Center for Biotechnology Information</i>
NGS	Sequenciação de nova geração; sigla adaptada do inglês: <i>Next Generation Sequencing</i>
NK	Do inglês: <i>Natural Killer</i>
nm	Nanómetro
°C	Graus Celsius
Pb	Par de base
PCR	Polimerase de reação em cadeia; sigla adaptada do inglês <i>Polymerase Chain Reaction</i>
PRR	Receptores de Reconhecimento de Padrões; sigla adaptada do inglês: <i>Pattern Recognition Receptors</i>
qPCR	Polimerização em cadeia em tempo real
qScore	<i>Quality Score</i>

rpm	Rotações por minuto
SCID	Imunodeficiência severa combinada; sigla adaptada do inglês <i>Severe Combined Immunodeficiency</i>
SNP	Polimorfismo de nucleótido único; sigla adaptada do inglês: <i>Single Nucleotide Polymorphism</i>
STAT	Proteína transdutora de sinal e ativadora de transcrição; sigla adaptada do inglês <i>Signal Transducer and Activator of Transcription</i>
TBE	Tris Borato EDTA
TGF	Fator de crescimento transformador; sigla adaptada do inglês <i>Transforming growth factor</i>
Th1	Linfócito T auxiliar; sigla adaptada do inglês <i>T helper 1</i>
TIMP	Inibidor tecidual de metaloproteinase; sigla adaptada do inglês <i>Tissue Inhibitor of Metalloproteinase</i>
TLR	Receptores <i>Toll</i> ; sigla adaptada do inglês: <i>Toll like receptor</i>
TNF	Fator de necrose tumoral; sigla adaptada do inglês <i>Tumor Necrosis Factor</i>
TNF $\alpha$	Fator de necrose tumoral alfa; sigla adaptada do inglês <i>Tumor Necrosis Factor alpha</i>
Tyk	Tirosina cinase
V	Volt

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$***



# Capítulo I

## Introdução

A presente dissertação aborda a investigação da febre Q em Portugal. Esta surge da necessidade de aprofundar o conhecimento desta doença infetocontagiosa, por ser pouco discutida entre a comunidade clínica e científica. Além disso, a febre Q já provou ter capacidade de provocar surtos significativos capazes de interferir com a saúde, economia, produção agropecuária, bem-estar e segurança das populações.

### 1. Febre Q

A febre Q é uma zoonose de distribuição mundial, tem sido associada a surtos marcantes que afetaram animais de interesse pecuário e o Homem. No hospedeiro humano, provoca uma infeção assintomática, aguda ou crónica, com diferentes manifestações clínicas. Embora a infeção não tenha taxas de mortalidade altas, causa danos persistentes e difíceis de reverter particularmente em indivíduos com doença valvular ou vascular e imunocomprometidos.

A febre Q conquistou particular relevância em 2007 após o aparecimento de um surto de grande escala na Holanda, o maior de que há registo, que resultou em 4026 casos de febre Q. Após estudos de coorte foram identificados oficialmente 284 casos com desenvolvimento crónico (Schimme, 2008; Delsing et al., 2010; van der Hoek et al., 2012; Honarmand, 2012). O caótico surto Holandês provou que a febre Q pode tornar-se uma ameaça séria para a saúde pública. Ocorreram ainda outros surtos significativos em números de casos, em países como Espanha (Errasti et al., 1984), Suíça (Dupuis & Vouilloz. et al, 1987), Inglaterra (Guigno et al., 1992), Alemanha (Schneider et al., 1993), Israel (Amitai et al., 2010), Eslováquia (Kováčová et al. 1998), nos trópicos (Angelakis et al., 2014) e igualmente durante o conflito armado no Médio Oriente (White et al, 2013). Estes episódios revelaram a presença de febre Q por todo o mundo. Geralmente o tratamento de febre Q crónica inclui um período de 3 a 4 anos com antibióticos, embora devastadora, atualmente é a única terapia eficaz. Para além da atenuação dos sintomas, continua a ser possível encontrar vestígios da *C. burnetii* no portador (Van Roeden et al., 2018).

Em Portugal, como em vários outros países a febre Q é uma doença de notificação obrigatória, mas defende-se que a sua incidência real possa estar subestimada. É ainda hoje uma doença pouco estudada que requer consciencialização da comunidade clínica e científica, podendo assim evitar a disseminação da doença.

## **1.1. Etiopatologia**

A febre Q é causada pela bactéria *Coxiella burnetii*, um cocobacilo de dimensões pequenas (compreendidas entre 0,2 a 0,4  $\mu\text{m}$  de largura e 1  $\mu\text{m}$  de comprimento). Pertence à subdivisão gama das proteobacterias na ordem *Legionellales*, pertence ainda à categoria B dos potenciais agentes biológicos do bioterrorismo (Baddour et al, 2007). Durante o ciclo de replicação intracelular ocorre formação esporos que conferem resistência no meio ambiente. A alta infecciosidade desta bactéria é conseguida por: apenas serem necessários menos de 10 germes para causar a infeção (Santos, Bacellar, & Franca, 2007); existirem quatro plasmídeos associados à *C. burnetii* (*QpH1*, *QpRS*, *QpDG* e *QpDV*) (Dragan & Voth, 2020).

Das estirpes de *C. burnetii*, destacam-se: *Nine Mile* (primeira a ser sequenciada nos Estados Unidos da América), a Z3055 (identificada como a origem dos surtos ocorridos na Alemanha e Holanda) e a Cb175 (associada ao surto da Guiana Francesa descrita como a estirpe mais virulenta) (Seshadri et al., 2003; D'Amato et al., 2014, 2015;)

A *C. burnetii* é ainda um microrganismo intracelular obrigatório, com variação de fase ligada à presença do LPS de membrana. Uma vez na célula alvo, a *C. burnetii* é internalizada num lisossoma das células fagocíticas, formando um vacúolo com características semelhantes a um fagolisossoma (Burette & Bonazzi, 2020). A *C. burnetii* é uma bactéria acidófila; é exatamente dentro do vacúolo de destruição que encontra as condições ideais para proliferar (Samanta & Gilk, 2017).

Verifica-se a presença de lipopolisacárido (LPS) de fase I e de fase II (Narasaki & Toman, 2012). O LPS em fase I é o mais presente na natureza, caracterizada pela presença de O-antígenos (açúcares como virenose e dihidrohidroxistreptose). O LPS de fase II é mais imunogénico com subsequente eliminação dos microrganismos (Narasaki & Toman, 2012). No decorrer da infeção existe modificação da fase do LPS (Alves et al., 2017).

Em Portugal, o genótipo da bactéria é semelhante entre os países mediterrâneos (Santos et al., 2012). Através da primeira genotipagem de diversidade de *C. burnetii* foram detectados os plasmídeos *QpH1*, *QpRS* e *QpDV* (Santos et al., 2012). Os plasmídeos

*QpDV* e *QpRS* foram identificados nos distritos de Lisboa, Setúbal, Évora e Faro, clinicamente associados à febre Q aguda; os plasmídeos *QpRS* e *QpH1* foram encontrados nos distritos de Lisboa e Castelo Branco, respetivamente, sendo estes associados à forma crónica da febre Q.

É de destacar a semelhança fisiopatológica da febre Q com a da tuberculose. O *Mycobacterium tuberculosis* (agente da tuberculose), é também uma bactéria intracelular. A fisiopatologia da infeção por *M. tuberculosis* e a resposta imunitária é em muito semelhante à da *C. burnetii*.

## **1.2. Epidemiologia**

Considera-se que tanto a incidência como a prevalência da febre Q real possa estar subestimada. Esta realidade pode ser em parte imputada à dificuldade em clinicamente reconhecer a infeção.

Em animais a infeção é comum, especialmente em ruminantes domésticos. É estimado que nas zonas endémicas, 15 a 20% dos animais em rebanho sejam seropositivos (United States Department of Agriculture, 2013). Nas espécies não ruminantes a estimativa de seropositivos oscila entre 10% a 66% (Panel & Ahaw, 2010). A febre Q provoca abortos esporádicos com uma frequência entre 0,5 a 4%, no entanto no decorrer de um surto esta estimativa ascende para 5 a 50% em rebanhos de ovelhas e 90% em rebanhos de cabras (United States Department of Agriculture, 2013).

Na população humana é estimado que esta doença tenha uma incidência global entre 1 a 30 casos por milhão de habitante (United States Department of Agriculture, 2013). No continente Europeu e Americano, verificam-se pequenos surtos entre o final da Primavera e início do Verão. Nas zonas tropicais existem surtos meses depois da época das chuvas.

Estudos retrospectivos de surtos e casos isolados apontam um aumento de casos desde 1999, com tendência crescente, contudo na Europa esta tendência têm sido controlada (Centre for Disease Prevention and Control, 2019). A reemergência ainda se verifica na Península Ibérica, Estados Unidos da América e outras partes do planeta (Nathalie. et al, 2005). A prevalência real de febre Q não é conhecida. A seroprevalência está estimada entre 5 a 25% de casos por todo o mundo (United States Department of Agriculture, 2013). Na tabela 1 encontram-se reunidos estudos de seroprevalência de imunoglobulina G (IgG) para a febre Q na população geral de vários países. Não existem evidências da

presença da febre Q na Nova Zelândia, Noruega, Islândia e Polinésia Francesa (Asamoah et al, 2020).

Tabela 1. Resultados de estudos de Seroprevalência de febre Q (IgG) na população geral por vários países.

<b>País</b>	<b>Seroprevalência</b>	<b>Amostra</b>	<b>Referência</b>
Mali	40%	156	(Schelling et al., 2005)
Malásia	9.8%	887	(Khor et al., 2018)
Ilha Réunion	6.81%	241	(Jaubert et al., 2019)
Austrália	5.6%	2785	(Gidding et al., 2020)
Somália	27%	190	(Ibrahim et al., 2020)

### **1.3. Vias de transmissão**

A maioria dos casos de febre Q é causada pela inalação de aerossóis contaminados por *C. burnetii* (Dupont & Raoult, 2008). A infecção começa após exposição a produtos animais infetados como: placenta, fetos, lãs, dejetos e poeiras que permanecem contaminadas por longos períodos. A ingestão de carne ou leite infetado é um fator de contaminação controverso, pois ainda não existem estudos conclusivos sobre esta questão (Rodolakis, 2009). A transfusão de fluídos, terapias celulares e xenotransplantes de dadores infetados promovem a infecção (Kuehn, 2018). A transmissão sexual da infecção por febre Q é pouco conhecida; de facto a *C. burnetii* parasita os espermatozoides e a placenta (Milazzo et al., 2001) como também pode estar presente em fluídos biológicos (saliva, urina, sangue, fluido amniótico e líquido cefalorraquidiano). O funcionamento desta via de transmissão ainda é desconhecida, no entanto a troca de fluídos entre os intervenientes dissemina a infecção (Miceli et al., 2010). Em obstetrícia os fluidos biológicos infetados são uma via de transmissão na forma de aerossóis excretados durante o parto (Dupont et al., 1994).

### **1.4. Aspetos clínicos**

#### **1.4.1. Febre Q aguda**

A febre Q aguda apresenta manifestações variáveis e pouco específicas. O processo de incubação da *C. burnetii* no hospedeiro demora entre 2 a 3 semanas, a infecção primária é assintomática em 60% dos infetados.

As manifestações da doença aguda são essencialmente o resultado do processo inflamatório com formação de imunocomplexos, e não tanto da proliferação do agente

nos tecidos. Os imunocomplexos definem-se por complexos antígeno-anticorpo que se depositam nos tecidos causando inflamação. A nível celular os imunocomplexos podem ligar-se a plaquetas, eritrócitos e neutrófilos causando trombocitopenia, anemia hemolítica e neutropenia, respetivamente. Outras manifestações clínicas de febre Q são as lesões cutâneas (Galache et al., 2004). Também podem ser detectados no sangue marcadores de autoimunidade (Fever et al., 2010).

Um dos factos curiosos da febre Q é que a mesma infeção no continente europeu é marcada maioritariamente por hepatites enquanto que na América do Norte por pneumonias (Aguilar-Olivos et al., 2013; Marrie, 2010).

Na verdade, nas infeções agudas a *C. burnetii* raramente é detectada noutra órgão ou tecido para além do sangue. O mesmo não se verifica nas manifestações crónicas, em que existe proliferação de *C. burnetii* nos tecidos afetados (Santos et al, 2007; Melenotte et al, 2020). As manifestações clínicas mais comuns estão descritas na tabela 2.

#### 1.4.2. Febre Q crónica

O quadro crónico de febre Q é caracterizado por infeções persistentes. A cronicidade da febre Q tem um intervalo de evolução de 1 a 20 anos após a infeção primária (Bossi et al, 2005). Os mecanismos deste processo ainda não foram elucidados pela comunidade científica; clinicamente, a forma crónica é identificada pela presença de LPS de fase II em maior concentração que o de fase I (Kampschreur et al., 2015). Empiricamente, considera-se febre Q crónica após persistência de 6 meses com sintomatologia.

São ainda detectadas alterações imunológicas como presença do factor-reumatoide, imunocomplexos, crioglobulinas e auto-anticorpos (Marrie & Raoult, 2002).

A persistência do processo inflamatório acaba por causar alterações hematológicas acentuadas. Em mais de metade dos casos verifica-se anemia, trombocitopenia, e alteração do perfil hepático e velocidade de sedimentação aumentada (Santos et al., 2007). As manifestações clínicas estão descritas na tabela 2.

**Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crónica na via de sinalização IL12-IFN $\gamma$**

Tabela 2. Manifestações clínicas da febre Q (adaptada de Santos et al., 2007).

<b>Manifestações</b>	<b>Febre Q aguda</b>	<b>Febre Q crónica</b>
Cutâneas	Exantema maculopapular ou purpúrico, eritema nodoso;	Vasculite cutânea, granuloma
Hematológicas	Anemia hemolítica, tromboflebite, rutura espontânea do baço;	Angeíte
Imunológicas	Linfadenopatia, formação de imunocomplexos, hemofagocitose histolítica, tiroidite, paniculite mesentérica ou lobular, epididimite, orquite, prostatite aguda, osteoartrite;	Infeção de aneurismas e próteses vasculares, osteomielite, osteoartrite, endometrite com aborto, amiloide, hipergamaglobulinémia;
Neurológicas	Síndrome febril, polineuropatia, síndrome de Guillain-Barré;	Síndrome de fadiga crónica
Cardiorrespiratórias	Pneumonia, miocardite, pericardite, derrame pleural eosinófilo, bronquiolites agudas	Endocardite, fibrose pulmonar, pseudo-tumor inflamatório do pulmão, bronquite
Gastrointestinais	Hepatite, colecistite, pancreatite, gastroenterite	Hepatite crónica por fibrose hepática ou cirrose

## 1.5. Diagnóstico laboratorial

O diagnóstico laboratorial de uma infeção por *C. burnetii* é realizado com diferentes técnicas e métodos de referência (Anderson et al., 2013), segundo a tabela 3.

Tabela 3. Técnicas de diagnóstico de febre Q diretas e indiretas (adaptada de Santos et al., 2007).

	<b>Técnica</b>	<b>Método</b>			
<b>Direto</b>	Imunocitoquímica	Observação			
	PCR	Biologia molecular			
	Isolamento <i>in vitro</i>	Cultura			
<b>Indireto</b>	IFA	Serologia	IgM	IgG	Fase
		Febre Q aguda	≥50	≥200	II
		Febre Q crónica		≥800	I

Direto: deteção da bactéria *C. burnetii* ou dos seus componentes mais específicos;  
Indireto: deteção da reação do sistema imunitário do hospedeiro à bactéria; IFA:  
Imunofluorescência Indireta; PCR: Reação de Polimerase em cadeia. Observação:  
Coloração de Gimenez (Mori, Mertens, Cutler, & Santos, 2017).

A presença de anticorpos anti-fase II IgG e IgM indiciam uma infecção primária, sendo detetáveis entre 7 a 15 dias após o início da sintomatologia (Frankel et al., 2011). Os anticorpos anti-fase I IgG estão relacionados com a progressão da infecção para cronicidade, a designada febre Q crónica (Frankel et al., 2011). No quadro crónico o aumento destes anticorpos mantém-se nos 6 meses após a infecção primária. No entanto, existem dados que revelam que alguns doentes desenvolveram febre Q crónica sem aumento de IgG fase I - a razão deste fenómeno ainda é desconhecida (Edouard et al., 2013).

Na infecção primária formam-se granulomas fibrosos em forma de anel, comparados visualmente com *donuts*, são visíveis em biópsias de tecido hepático, ósseo-medular e de válvulas cardíacas (Argov et al., 2008; Lee et al., 2012). No estudo de Lepidi e colaboradores, a análise histológica revelava fibrose, calcificação, inflamação com aumento do calibre vascular e vegetações, todos estes processos são infecciosos, lentos e degenerativos (Lepidi et al 2003).

A quantificação das concentrações de interferão gama (IFN $\gamma$ ) pelo método de imunoabsorção enzimática (ELISA) é considerado um novo parâmetro de diagnóstico. A produção de IFN $\gamma$  é descrita como significativamente elevada em doentes com infecções persistentes (Schoffelen et al., 2013), no entanto esta forma de diagnóstico ainda é controversa.

## **1.6. Factores de risco**

O desenrolar da infecção é multifactorial, partindo de factores ambientais, do agente infeccioso e do hospedeiro. As variáveis ambientais agregam o contacto com animais, a influência das estações e demografia. Relativamente ao agente, os factores de risco estão associados à virulência da estirpe infecciosa e à presença de plasmídeos. Já as variáveis do hospedeiro categorizam a suscetibilidade à infecção, num espectro individual (tabela 4) estes agrupam-se em factores:

- 1) Gerais que representam um risco ligeiramente aumentado (Maltezou & Raoult, 2002; Maltezou et al., 2004).

- 2) Focalizados que estão maioritariamente envolvidos no desenvolvimento crónico ou outros resultados adversos (Mboussou et al., 2019).
- 3) Imunológicos que são o ponto fulcral no desenvolvimento crónico, tais como imunodeficiências adquiridas devido à utilização de terapias imunossupressoras, doenças oncológicas, transplante renal, infecção pelo vírus da imunodeficiência humana (Raijmakers et al., 2019).

No entanto, estes factores clássicos não explicam, por si só, a totalidade do desenvolvimento crónico, o que sugere que possam existir também factores genéticos de predisposição.

Tabela 4. Fatores de risco associados ao desenvolvimento de febre Q crónica (adaptada de Kampschreur et al., 2012).

<b>Categoria</b>	<b>Fator de risco</b>
Gerais	Sexo masculino
	Tabagismo
	Idade > 60 > 15 anos
Focalizados	Insuficiência renal
	Valvulopatia cardíaca
	Vasculopatia
	Próteses valvulares e vasculares
Imunológicos	Gravidez
	Deficiência imunológica

## **2. Resposta imunológica**

De um ponto de vista estratégico, a defesa do organismo é proporcional à resistência da bactéria. No caso da febre Q a resposta à *C. burnetii* chega a ser violenta, com efeitos colaterais no hospedeiro.

### **2.1. Imunidade inata**

No hospedeiro, a *C. burnetii* ultrapassa as primeiras barreiras defensivas do corpo; consegue sobreviver no muco nasal, flora do trato urogenital e às enzimas dos fluidos corporais. Uma vez dentro do organismo, a *C. burnetii* é intercetada, principalmente por monócitos, macrófagos, células exterminadoras (*Natural Killer* - NK) e células dendríticas (Ghigo et al., 2009b).

A interação da *C. burnetii* com macrófagos difere consoante a fase do LPS (Hotta et al., 2002). O LPS de fase I é reconhecido pelos receptores *Toll* (*Toll Like Receptor* - TLR) 2 e 4 nos macrófagos, esta ligação, em particular, causa um rearranjo no citoesqueleto celular criando “rufles”<sup>2</sup> que dificultam a fagocitose da *C. burnetii* (Omsland et al., 2009). O LPS de fase II é reconhecido com o apoio do receptor de complemento 3 (CR3) promovendo uma internalização com maior eficiência (Romano et al., 2007).

A internalização da *C. burnetii* é feita através da formação de um fagossoma ácido (Howe et al., 2010) com propósito de destruir a bactéria em meio ácido – característica ideal para activação da *C. burnetii* - a *C. burnetii* de fase II consegue ser destruída após fusão do fagolisossoma, no entanto a *C. burnetii* de fase I sobrevive ao inibir o recrutamento das proteínas necessárias à conversão do fagossoma em fagolisossoma (Rab7 e catepsina D) (Romano et al., 2007; Medkour et al., 2020). A sobrevivência da *C. burnetii* desencadeia um novo alarme no sistema imunitário, a colonização das células protectoras é uma falha grave que compromete a viabilidade do organismo. Nesta fase as células secretam interleucina 12 (IL-12) para o meio extracelular (Schoffelen et al., 2013).

O papel do sistema complemento e células NK é pouco descrito na febre Q. Um estudo realizado com dois grupos de ratos com Imunodeficiência Severa Combinada (*Severe Combined Immunodeficiency* - SCID) concluiu que ratos SCID  $\beta\gamma$  desprovidos da linhagem NK apresentavam lesões com maior gravidade e extensão que ratos SCID padrão.

As células dendríticas fazem a ponte entre a imunidade inata e adaptativa, estas são células sentinela, com função bactericida e um elemento-chave na activação de linfócitos T. Estas células têm um papel essencial na apresentação de antígenos da *C. burnetii* (Chen et al., 2009).

## **2.2. Imunidade adaptativa**

A resposta adaptativa é acionada de forma a criar memória, activando novos mecanismos e aumentando a eficiência de resposta à infeção por *C. burnetii*.

Nos linfócitos T imaturos, a estimulação com IL-12 leva à diferenciação e proliferação de linfócitos auxiliares tipo 1 (*Thelper 1* - Th1) (Benoit et al., 2008), adquirindo capacidade de produzir factor de necrose tumoral alfa (TNF $\alpha$ ) e interferão gama (Xiong et al., 2014).

---

<sup>2</sup> Textura crespa e engelhada

Os linfócitos B têm um papel controverso na febre Q, são especializados em produzir anticorpos IgM contra o LPS fase I e IgM e IgG para o LPS fase II para marcar o patógeno (Melenotte et al., 2016). Estes anticorpos não têm um papel relevante na resolução de febre Q (Teunis et al., 2013), ainda assim são fundamentais na regulação da resposta imunitária (Shannon et al., 2009).

## **2.3. Eixo IL-12 IFN $\gamma$**

A via IL-12 IFN $\gamma$  parece ter um papel fulcral na resolução da febre Q (Schoffelen et al., 2017). Esta via é iniciada maioritariamente pelos macrófagos e células dendríticas. Esta via estimula simultaneamente as respostas imunitárias inata e adaptativa (Burke & Young, 2019). A via IL-12 IFN $\gamma$  é extremamente importante na resolução de infecções por vírus e bactérias, incluindo *C. burnetii* e *M. tuberculosis*. Em casos complicados e desde que esteja completamente funcional, a via IL-12 IFN $\gamma$  dita a viabilidade do organismo (Holland, 2005). A análise destas duas citocinas desvenda a importância desta via. A nível molecular, os mecanismos da cascata de sinalização revelam a complexidade dos fenómenos imunitários nesta guerra imunológica.

### **2.3.1. Interleucina 12**

Molecularmente, a ativação da via IL-12 IFN $\gamma$  inicia-se pela secreção de IL-12 para o meio extracelular, estimulando a diferenciação de linfócitos T imaturos em linfócitos Th1. As células dendríticas e macrófagos quando estimulados por antígenos libertam IL-12, sendo solúvel em meio extracelular, liga-se ao receptor de IL-12 presente na membrana das células dendríticas, NK e linfócitos T. Ao serem detectados macrófagos parasitados por *C. burnetii* outros macrófagos libertam interleucina (IL)-12 – precursor da ativação da via interleucina12-interferão gama (IL12-IFN $\gamma$ ), em simultâneo há atividade das células dendríticas e exterminadoras naturais (NK), a libertação de IL-12 potencia a libertação de IFN $\gamma$  levando a uma retroamplificação de respostas. O receptor de IL12 tem dois domínios o  $\beta$ 1 e  $\beta$ 2 (Vignali, 2014). Ambos os domínios são transmembranares, com tirosina cinase 2 (Tyk2) acoplada ao domínio intracelular do  $\beta$ 1 e janus cinase 2 (Jak2) acoplada ao domínio intracelular do  $\beta$ 2. Quando a IL-12 se liga a este receptor os domínios  $\beta$ 1 e  $\beta$ 2 desencadeiam uma cascata de sinalização Jak/STAT, no espaço intracelular (figura 1). As cinases Tyk2 e Jak2 são fosforiladas para ligar a proteína transdutora de sinal e activadora de transcrição 4 (*Signal Transducer and Activator of Transcription* - STAT4), que é também fosforilada e desacoplada em forma de dímero

para o núcleo, onde actua como activador de transcrição e provoca a maturação específica do linfócito T para Th1 (Vignali, 2014).

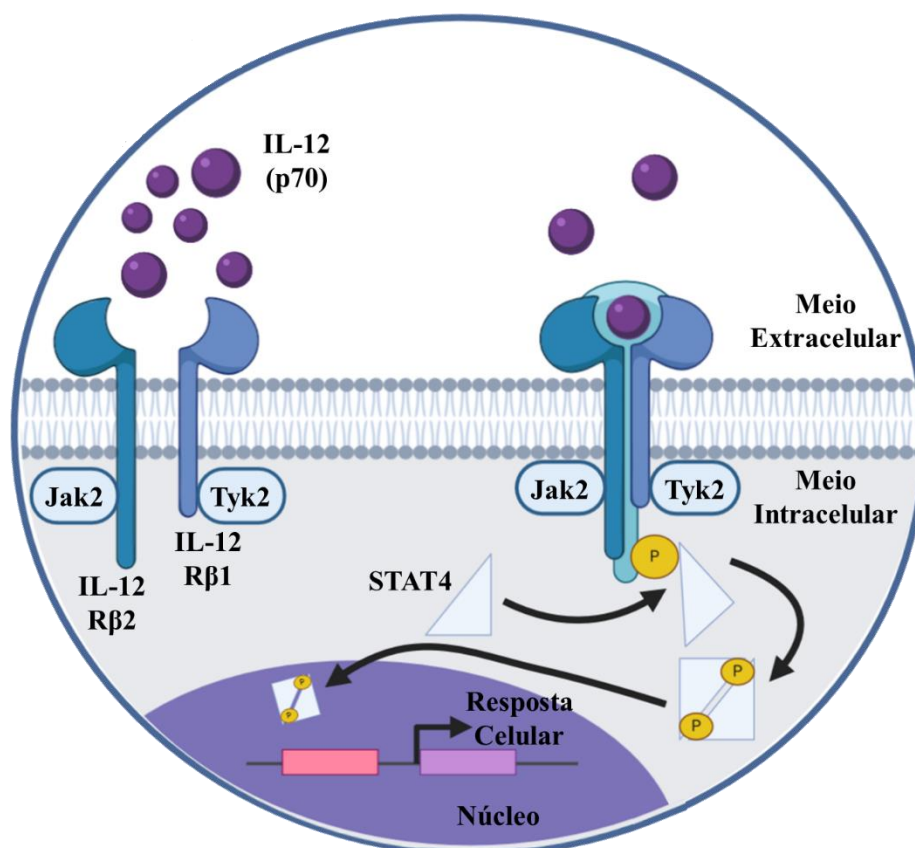


Figura 1. Cascata de sinalização do receptor de interleucina (IL)-12; a molécula de IL-12 ativa (p70) ao ligar no receptor de IL-12 (IL-12r) fosforila (P) dímeros da proteína transdutora de sinal e ativadora de transcrição 4 (STAT4) levando à resposta celular. Jak: Janus cinase; Tyk: Tirosina cinase (imagem criada a partir de biorender.com).

### 2.3.2. Interferão Gama

O IFN $\gamma$  é uma citocina da classe II de interferões, como o próprio nome indica, tem o propósito de interferir na replicação viral ou bacteriana dentro das células do hospedeiro. O IFN $\gamma$  é expresso por macrófagos, células dendríticas, monócitos e linfócitos NK, B e T (CD8, CD4 e gama delta) (Cooper et al., 2000; Gao et al., 2003; Robinson et al, 2009).

O receptor de IFN $\gamma$  é composto por dois domínios o receptor 1 e o receptor 2 (IFN $\gamma$ R1 e IFN $\gamma$ R2, respectivamente), ambos os receptores têm duas cadeias (Burke & Young, 2019). Quando a molécula de IFN $\gamma$  se liga no receptor, as cadeias do IFN $\gamma$ R1 e IFN $\gamma$ R2 dimerizam. No espaço intracelular, o conjunto das subunidades do receptor leva à transfosforilação e activação das cinases Jak1 e Jak2. Estas fosforilam resíduos de tirosina formando um conjunto de ancoragem para a STAT1. Duas moléculas de STAT1 ligam-se nestes locais de acoplamento, sendo também fosforiladas. As moléculas STAT1

quando fosforiladas dissociam-se do receptor e formam complexos homodiméricos no citosol. Os homodímeros STAT1 deslocam-se para o núcleo onde se ligam a sequências específicas, iniciando a transcrição de genes (Burke & Young, 2019). Esta cascata de sinalização, representada na figura 2, desencadeia diversas respostas do IFN $\gamma$ , tais como:

- 1) Regulação expressão de antígenos, citocinas inflamatórias e quimiotáticas (Moran, 2017);
- 2) Activação e polarização de leucócitos (Zhang et al., 2020);
- 3) Indução de expressão de IL-12 em macrófagos (Moschen et al., 2019);
- 4) Regulação da gênese de óxido nítrico e espécies reactivas de oxigênio (Wang et al., 2018);
- 5) Indução de quimiocinas angioestáticas (Darakhshan et al., 2019);
- 6) Supressão da expressão de IL-10 e TGF- $\beta$  (Kang et al., 2017);
- 7) Indução apoptose pela activação das caspases 1 e 8 (Takeda et al., 2002) e activação de Fas e ligando Fas (Lee et al., 2000);
- 8) Restauo da fusão do fagolisossoma (Ghigo et al., 2009a).

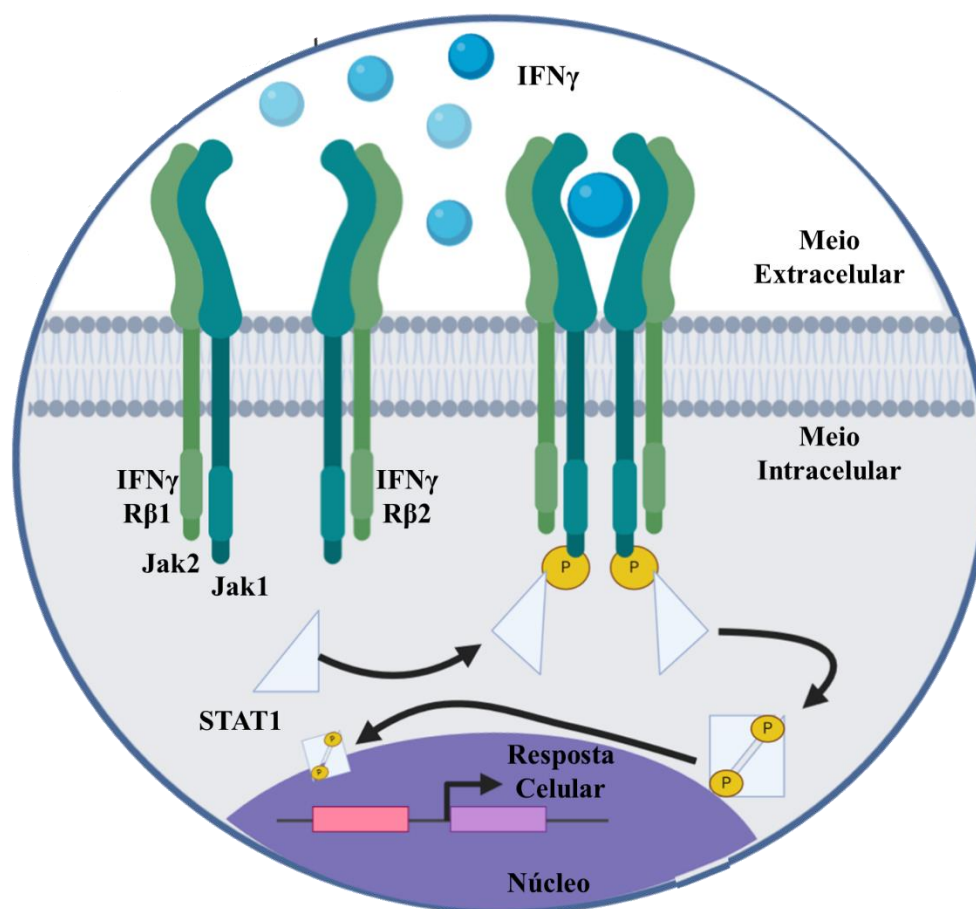


Figura 2. Cascata de sinalização do receptor de interferência gama (IFN $\gamma$ ); a molécula de IFN $\gamma$  ao ligar no receptor e fosforila dímeros de proteína transdutora de sinal e ativadora de transcrição (STAT1),

desencadeia a resposta celular. IFN $\gamma$ R: Receptor de interferão; Jak: Janus cinase; gama (imagem criada a partir de biorender.com).

### **3. Modificadores genéticos**

Vários estudos prévios demonstraram que variantes genéticas relacionadas com a resposta do sistema imunitário à infeção por *C. burnetii* estariam associados a um aumento de risco para a febre Q crónica. Estes incluem: genes que codificam as metaloproteinases da matriz (*Matrix Metalloproteinases* - MMP); genes da via fagolisossomal; gene de IFN $\gamma$ ; genes dos receptores *Toll*.

As metaloproteinases<sup>3</sup> membranares clivam a matriz extracelular, estão envolvidas em infecções e manutenção de aneurismas (Rodríguez et al., 2010). O envolvimento das MMPs na febre Q crónica é expectável pois ambos estão fortemente associados a endocardites (Jansen et al., 2017). Além do mais, a MMP7 encontra-se mais concentrada no soro de doentes crónicos (em média 33.5 ng/mL) quando comparada com o grupo de indivíduos saudáveis (em média 20,6 ng/mL). No seu estudo, Jansen e colaboradores (Jansen et al., 2017), com 139 doentes com febre Q crónica e 220 saudáveis após febre Q aguda, concluem que o polimorfismo rs11568810 no gene *MMP7* e o rs17576 no gene *MMP9* são mais comuns em doentes crónicos. Alterações nos genes inibidores tecidulares de metaloproteinases (*Tissue Inhibitor of Metalloproteinase* - TIMPs), nomeadamente o *TIMP1*, *TIMP2* e *TIMP3* não atingiram significância nesta população (Jansen et al., 2017).

Do estudo de Jansen e colaboradores (Jansen et al., 2019), sobre genes associados à destruição da *C. burnetii* no meio extracelular, resultaram novos dados relevantes. Foram analisados a 173 doentes de febre Q crónica mais 184 indivíduos com factores de risco de desenvolver cronicidade. Foram estudados os grupos de genes da via fagolisossomal (*RAB5A*, *RAB7A*, *LAMP1*, *LAMP2*, *EEA1*, *CTSD*, *STX17*, *VPS35*, *TNF*), vias microbidas (*CYBA*, *CYBB*, *NCF2*, *NCF4*, *RAC2*, *NOX1*, *NOX4*, *NOS2*, *IDO1*, *P2RX7*) e da via autofágica (*ATG10*, *ATG16L1*, *ATG5*, *EREG*, *ATG2B*, *LAMP3*, *IRGM*, *ATG9B*, *ATG18*, *MAP1LC3A*). Os resultados revelaram uma presença acrescida em doentes crónicos das variantes rs13081864 e rs3751143 nos genes *RAB7A* e *P2RX7*, respetivamente. Neste estudo, foram ainda encontradas variantes que indiciam um papel protector contra febre Q, nomeadamente *RAB5A* (rs8682), *P2RX7* (rs1718119),

---

<sup>3</sup> Família de protéases dependentes de cálcio e zinco que atuam a nível da matriz extracelular na degradação de várias proteínas

*MAP1LC3A* (rs1040747) e *ATG5* (rs2245214). Os genes restantes não foram significativos na análise (Jansen et al., 2019).

Os TLR são uma família importante de receptores de reconhecimento de padrões (PRR) que reconhecem padrões moleculares associados aos agentes patogénicos (pathogen-associated molecular patterns - PAMPs), como os dos LPS. O envolvimento da família dos receptores TLR sugere um papel importante na internalização da *C. burnetii*. Schoffelen e colaboradores (Schoffelen et al., 2015), num grupo de 139 doentes de febre Q crónica e 220 indivíduos com histórico de exposição a febre Q e presença de factores de risco cardiovasculares, foi averiguada a existência de variantes associadas ao desenvolvimento crónico. A análise focou um painel de genes dos PRR e proteínas adaptadoras associadas. As variantes dos genes *TLR1* (rs5743611), *NOD2* (rs2066847) e *MYD88* (rs4988453) estão significativamente presentes nos doentes crónicos, o que não se verificou no grupo controlo. Nesta investigação foram ainda analisados outros genes que não atingiram resultados significativos, sendo eles o *TLR2*, *TLR4*, *TLR6*, *TLR8*, *ITGAV*, *ITGB3*, *ITCAM* e *TIRAP* (Schoffelen et al., 2015).

As variantes rs2430561 e rs1861493 do *IFN $\gamma$* , rs1914408 do *STAT1* e rs2228570 do *VDR* foram investigadas em duas populações holandesas por Wielders e colaboradores (Wielders et al., 2015). A primeira população concentrou 85 indivíduos do surto holandês de 2007 provocado pela *C. burnetii*; a segunda população composta com 123 indivíduos provenientes do surto holandês de 2009. Foi averiguado que os portadores da variante rs2430561 do *IFN $\gamma$*  apresentam sintomatologia mais evidente que os não portadores. A variante rs1861493 do *IFN $\gamma$*  segue a mesma tendência embora com menos significância. As restantes variantes dos genes *STAT1* e *VDR* não obtiveram correlações significativas (Wielders et al., 2015).

## **4. Projeto de estudo**

Os genes *IL12RB1*, *IFN $\gamma$ R1* e *IFN $\gamma$ R2* pertencem à lista de genes associados a susceptibilidade mendeliana a infecções por micobactérias (Ramirez-Alejo & Santos-Argumedo, 2014). Nestes genes, existem variantes genéticas associadas ao desenvolvimento de formas graves de tuberculose, cuja patofisiologia tem semelhanças com a febre Q (formação de granuloma, existência forma aguda e crónica, implicação da via IL-12 IFN $\gamma$ ) (Langan et al., 2020).

Os genes da via IL-12 IFN $\gamma$  também têm um papel importante na febre Q, no entanto, fora o IFN $\gamma$ , variantes nestes genes ainda não foram estudadas por métodos de associação genótipo-fenótipo, levantando a questão: será que existem determinantes genéticos nos receptores desta via que promovam o desenvolvimento crónico de febre Q?

#### **4.1. Estudo populacional de associação genótipo-fenótipo**

Os estudos populacionais de associação genótipo-fenótipo com base em genes candidatos (*Candidate Gene Association Studies* – CGAS) têm vindo a contribuir significativamente para a identificação de variantes genéticas causais e para o aumento da nossa compreensão da arquitectura genética das doenças humanas. Com os avanços tecnológicos, houve uma explosão na quantidade de ferramentas de análise e detecção. Actualmente é possível identificar variantes genéticas que predispõem a doenças infecciosas ou a formas graves das mesmas muito antes de estas se manifestarem (David, 2020).

As abordagens CGAS são relativamente baratas e rápidas de executar, estão focadas na selecção de genes que foram de alguma forma relacionadas anteriormente com a doença e, portanto, vêm com conhecimento prévio sobre a função do gene (Esparragón et al., 2009). O CGAS começa com a selecção de um gene candidato putativo, cujo factor relevância se destaca no mecanismo da doença (característica) (Kwon & Goate, 2000). As abordagens CGAS são valiosas na identificação de variantes de risco, podendo estar na origem de deficiências no produto do gene, que possam afectar o fenótipo clínico ou aumentar o risco de desenvolvimento de formas graves da doença. A finalidade destes estudos é poder canalizar os doentes em risco para cuidados especiais (médicos, hospitalares, vacinação) e desenvolver novas opções terapêuticas capazes de superar essas deficiências.

#### **4.2. Sequenciação de nova geração**

Para associar determinantes genéticos como factores de risco para o desenvolvimento de febre Q crónica é necessário aceder à informação genética (Sheng & Christopher, 2016), isto requer a selecção de metodologias de genotipagem dos genes candidatos do hospedeiro e da análise desses dados (Yohe & Thyagarajan, 2017). A tecnologia NGS tem diversas aplicações, nos estudos CGAS aplica-se para sequenciação massiva das amostras e devolve um relatório de dados com as variantes encontradas (sejam elas: polimorfismos de nucleótido único (*Single Nucleotide Polymorphism* - SNP), inserções

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crónica na via de sinalização IL12-IFN $\gamma$***

ou deleções (INDEL), variantes em número de cópias (CNV) ou variantes estruturais (SV)).

Até à data, os genes da via IL12-IFN $\gamma$  não foram contemplados em estudos de associação genética com a febre Q, ou seja, não existem variantes conhecidas que possam guiar uma selecção de genes ou variantes (Gargis et al., 2015), assim como não existem estudos descritos que abordem esta questão com sequenciação de nova geração (NGS)

A NGS tornou-se uma metodologia de genotipagem acessível técnica e economicamente aos laboratórios de biologia molecular. Através do NGS obtém-se um número elevado de variantes, aumentando a possibilidade de analisar um maior número de genes e variantes. Outra mais valia desta metodologia é a robustez implícita na análise e a fiabilidade dos resultados (Gargis et al., 2015). Por estas razões, e por não terem sido exploradas até á data variantes candidatas nestes genes, é usada a metodologia NGS.

A metodologia NGS é tripartida:

I. A fase primária decorre no aparelho de sequenciação. Nesta fase são geradas leituras (*reads*) através da amplificação por ponte. A sequenciação por síntese ocorre após a formação da ponte, para serem obtidas as leituras das bases (*base call*). Após a corrida é gerado um relatório geral com os parâmetros de qualidade da sequenciação. Neste relatório são descritos:

- A **dispersão da percentagem de leitura** é um dado que expõe a homogeneidade da sequenciação;
- **Cluster** é um conjunto de fragmentos clonados na célula de fluxo (*flow cell*). Estes resultam da amplificação em ponte dos fragmentos de ADN. Quando existe demasiada sobreposição de luminescência dos fragmentos pode gerar erros de leitura. O *clustering* é a unidade que representa o total número de *clusters*.

São geradas ainda métricas de qualidade nomeadamente os índices de qualidade (*Phred quality scores*), que medem a qualidade de identificação das *reads*, são exemplo:

- **Quality Score 30 (Q30)**, significa que, durante o *base call*, a probabilidade de erro é de 1 em 1000 bases identificadas, ou seja, uma precisão de 99,9%. Este dado é indicado no ficheiro FASTQ (formato do ficheiro de saída da corrida) Representa-se pela equação  $Q = -10 \log_{10} P$  (onde Q representa qualidade e P a probabilidade de uma base ser mal identificada).
- **Cobertura** inclui métricas como a cobertura de leitura que define a percentagem em que o genoma de referência ou uma região alvo do genoma é coberto por leituras; e a cobertura de profundidade, uma métrica de eficiência, contabiliza a quantidade de leituras geradas num determinado locus.

II. A fase secundária consiste na separação das leituras consoante os índices correspondentes a cada amostra (desmultiplexação). São feitos o mapeamento e o alinhamento das *reads* ao genoma de referência, a chamada das variantes, com recalibração dos parâmetros. Esta fase termina com a identificação de variantes (*variant call*). Nesta fase ainda se pode incluir a priorização, isto é, filtração com ajuste de parâmetros segundo o tipo de dados genotipados, por análise computadorizada;

III. O objetivo desta fase terciária final é a análise da lista de dados NGS gerada pelas fases anteriores, para identificar variantes relevantes e fornecer o respectivo relatório, seja para fins clínicos ou para estudos de associação genótipo-fenótipo.

Uma vez que a NGS trabalha com milhares de milhões de dados em simultâneo, eleva-se a probabilidade em ser lida uma base incorreta, para ultrapassar esta limitação e garantir resultados fidedignos existem parâmetros e métricas de qualidade que permitem validar as fases I e II. Para além destas, são ainda aplicadas métricas de qualidade específicas para CGAS que têm por objetivo diminuir a obtenção de erros por falsos positivos na análise de associação.

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$***

## Capítulo II

### Objectivos

O objectivo principal desta dissertação é genotipar as regiões 5'UTR e os exões dos genes *IL12RB1*, *IFN $\gamma$ R1*, *IFN $\gamma$ R2*, recorrendo a tecnologias de NGS, e destacar desses resultados variantes do tipo SNPs, com a qualidade necessária na sequenciação e representatividade a nível da amostra para estudos de associação genótipo-fenótipo. O controlo de qualidade por amostra é fulcral na garantia da fiabilidade dos resultados de um CGAS.

São apontados os seguintes objectivos laboratoriais:

- 1) Desenhar um estudo caso controlo;
- 2) Seleccionar os genes candidatos e fragmentos alvo;
- 3) Genotipar por NGS, utilizando a tecnologia NGS da *NexteraXT* da Illumina®:
  - a. Amplificação por PCR dos fragmentos alvo dos genes;
  - b. Sequenciação paralela dos amplicões:
    - i. NGS Fase I e controlo de qualidade;
    - ii. NGS Fase II e controlo de qualidade;
- 4) Preparar os dados NGS de Fase III para análise de associação genótipo-fenótipo pela imposição de limiares (*thresholds*) de controlo de qualidade.

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$***

## Capítulo III

### Materiais e métodos

#### 1. Materiais

Amostras: neste estudo constam 94 indivíduos portugueses, separados em três grupos: caso crónico (17 indivíduos, grupo caso); caso agudo (43 indivíduos, grupo controlo); população geral (34 indivíduos). Para extrair ADN genómico foi usado o aparelho *MagNA Pure nucleic acid extractor* da Roche®.

Kit PCR: os *primers* foram adquiridos através da *Thermo Fisher Scientific*®. Para a PCR foi usada água bidestilada (H<sub>2</sub>O), betaína, solução com trifosfato de nucleosídeos (dNTP's), *primers Forward* e *Reverse*. A polimerase usada foi a Takara Taq polimerase do kit de *Long-Range PCR* (TaKara Holdings®). O termociclador T *Gradient Biometra*® foi usado para executar a PCR. O gel de eletroforese foi constituído por agarose da Seakem® LE Agarose, o brometo de etídio da Sigma®, e o marcador de peso molecular 1kb *Ladder* da Biolabs®. Para a corrida do gel foi usada a fonte de tensão BIORAD® PowerPac300. A purificação das amostras foi conseguida com esferas magnéticas *AMPure XP* (Beckman Coulter®).

*Kit* sequenciação: na tagmentação do ADN genómico foi utilizada a enzima *Nextera Transposon*. Para a sequenciação foi usado o equipamento MiSeq™ da Illumina®. Para o fluxo de trabalho o *kit* bibliotecas Nextera XT combinado com o *kit* indexação Nextera XT Index v2 (Illumina®).

Análise de dados: Para plotagem e construção de dados foi usado o programa Excel do Microsoft Office 365, com licença de uso pela Universidade da Beira Interior. Para tratamento de dados e controlo de qualidade foi usado o programa PLINK, versões 1.07 e 1.9 (Purcell et al., 2007).

## **2. Métodos**

O plano geral do fluxo de trabalho inclui a obtenção de amostras biológicas através do recrutamento de participantes para grupos em estudo, recolha de amostras de sangue e isolamento de ADN. Preparação da NGS com selecção e amplificação dos fragmentos alvos dos genes candidatos, agrupamento por amostra e purificação dos produtos PCR. Sequenciação NGS com preparação das bibliotecas, aplicação da metodologia Nextera XT (fase I NGS no aparelho MiSeq) e genotipagem (fase II NGS). E, por fim, a análise de dados da sequenciação para o controlo de qualidade referente a CGAS. O esquema geral de todo o processo está ilustrado no anexo 2 -Fluxo de trabalho.

### **2.1. Amostras biológicas**

O recrutamento dos participantes e colheita de amostras para este estudo foi autorizado pelo comité de ética do Instituto Nacional de Saúde Dr. Ricardo Jorge Instituto Público (INSA/IP). A declaração de consentimento informado foi de preenchimento obrigatório para inclusão no estudo, nos casos em que não foi possível, as amostras foram utilizadas à guarda do Artigo 19º da Lei nº 12/2005, de 26 de janeiro. A codificação dos participantes foi respeitada de acordo com as normas homologadas para a investigação no INSA, IP, em concordância com a declaração de Helsínquia (Association., 2001).

Como critérios de elegibilidade, a inclusão para o grupo crónico requereu: indivíduos expostos a *C. burnetii*, febre Q crónica diagnosticada e título isolado de IgG  $\geq$  800 (através de serologia por técnica de IFA - conforme a tabela 3); para o grupo agudo: indivíduos expostos a *C. burnetii*, febre Q aguda diagnosticada e título isolado de IgM  $\geq$  50 e IgG  $\geq$  200 (através de serologia por técnica de IFA - conforme a tabela 3); para o grupo da população geral: amostra de conveniência com exposição à *C. burnetii* desconhecida. Amostras são independentes, os indivíduos não eram relacionados.

Por indivíduo, foram extraídos 3 mL de sangue periférico em tubos com ácido etilenodiaminotetracético (EDTA). A partir destes tubos, o ADN genómico foi extraído automaticamente. As amostras biológicas de sangue foram armazenadas com registo de utilização e código atribuído. O ADN foi armazenado a  $-4^{\circ}\text{C}$  e o sangue periférico a  $-20^{\circ}\text{C}$ . Foram reunidas 94 amostras, das quais: 17 de indivíduos com febre Q crónica, 43 com febre Q aguda e 34 da população geral.

## 2.2. Preparação NGS

### 2.2.1. Selecção dos fragmentos dos genes candidatos

Em cada amostra foram amplificados, por PCR, fragmentos dos genes *IL12RB1*, o *IFN $\gamma$ R1* e o *IFN $\gamma$ R2*. Usando o principal transcrito disponível no Ensemble 96 da versão genómica GRCh38.p12 (Ensembl, 2019), os fragmentos para amplificação foram desenhados para incluir os exões, algumas regiões intrónicas e 1000 pares de base (pb) na região 5'UTR (figura 3).

Os *primers* foram desenhados a partir da ferramenta informática *Primer Blast*, disponibilizada gratuitamente no *National Center for Biotechnology Information* (NCBI). Os dados sobre os *primers* estão disponíveis na tabela 9.

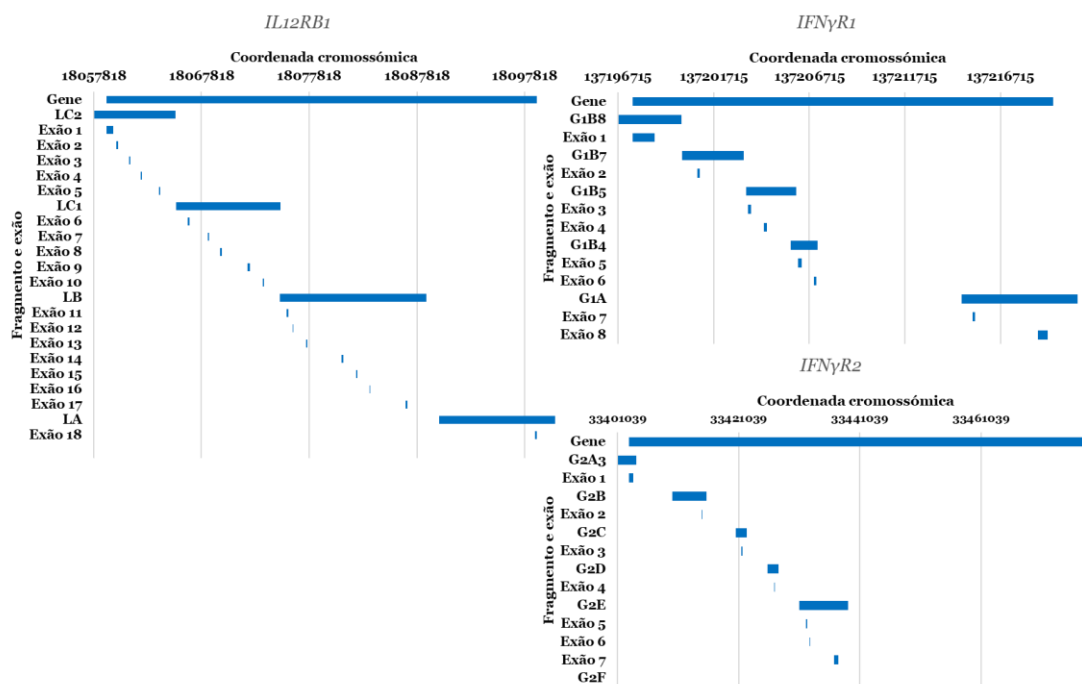


Figura 3. Distribuição dos exões por fragmento amplificado ao longo do gene (transcrito principal – Ensemble 96, versão genómica GRCh38.p12) (Ensembl, 2019). *IL12RB1*: receptor 1 de interleucina 12; *IFN $\gamma$ R1*: receptor 1 de interferão gama; *IFN $\gamma$ R2*: receptor 2 de interferão gama.

### 2.2.2. Amplificação dos fragmentos alvo

Para garantir o máximo rendimento, a PCR foi otimizada com ensaios consecutivos. Diferentes concentrações de magnésio e betaína (potenciador de amplificação) foram testadas até obter o melhor resultado de amplificação. No termociclador foram testadas

temperaturas num intervalo de 65°C a 70°C (vector de gradiente). O número de ciclos foi posteriormente testado num intervalo de 30 a 40 ciclos. Todo o processo foi repetido e registado para cada fragmento até gerar o protocolo de execução. Para as misturas reacionais, foi destacado o máximo rendimento e gerado o protocolo PCR (misturas - tabela 5; termociclador – tabela 6), os fragmentos foram agrupados em três misturas:

1. LA e LB;
2. G2C, G1B4, LC1 e G2A3;
3. G1B8, G1B7, G2E, G1B5, G2B, G2D, G1A, LC2 e G2F.

Tabela 5. Protocolo PCR. Secção Mistura reúne os reagentes com quantidades expressas em  $\mu$ L, num volume de reação final de 25  $\mu$ L por fragmento a amplificar

<b>Mistura Mestre</b>	<b>Reagente</b>	<b>Ci (mM)</b>	<b>Vi (<math>\mu</math>L)</b>
	Tampão <i>takara</i>	10x*	2,5
	dNTP	2,5	4
	Betaina	5	5
	ADN	NA**	1
	Primer	0,025	2,5
<b>Mistura 1</b>			
	Takara polimerase	5 (U/mL)	0,25
	H <sub>2</sub> O	NA	7,25
<b>Mistura 2</b>			
	Takara polimerase	5 (U/mL)	0,125
	H <sub>2</sub> O	NA	7,375
<b>Mistura 3</b>			
	Takara polimerase	5 (U/mL)	0,125
	H <sub>2</sub> O	NA	6,375
	Mg	25	1

Mistura mestre comum a todas; Mistura 1 para os fragmentos LA e LB; Mistura 2 para os fragmentos G2C, G1B4, LC1 e G2A3; Mistura 3 para os fragmentos G1B8, G1B7, G2E, G1B5, G2B, G2D, G1A, LC2 e G2F. Ci: concentração inicial, Vi: volume inicial, Vf: volume final, NA: Não aplicável, mM: milimolar,  $\mu$ L: microlitro; \*10 vezes, \*\*A quantidade de ADN utilizada por reação foi de 30-100 ng.

O programa PCR comum a todos os fragmentos, começou com um passo de desnaturação durante 1 minuto a 94°C. Após este passo foram repetidos por um determinado número de ciclos os passos de desnaturação a 98°C durante 10 segundos, e de hibridação com uma temperatura consoante ao fragmento durante 20 minutos (tabela 6). Por esta ser uma polimerase alta fidelidade de longo alcance, o passo de alongamento foi feito após a repetição dos ciclos a 72°C durante 10 minutos, conforme as indicações do fabricante (Jia et al., 2014).

**Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crónica na via de sinalização IL12-IFN $\gamma$**

Tabela 6. Condições do termociclador consoante o fragmento.

<b>Temperatura</b>	65°C	66°C	67°C	69.7°C	67.6°C	68°C	67°C
<b>Ciclos</b>	30	30	30	30	35	35	40
<b>Fragmento</b>	G1B8 G1B7 G2E	G1B5	G2B G2D G2C G1B4	G2F	G2A3	G1A LC2	LC1 LA LB

A verificação das amplificações foi feita por electroforese em gel de agarose a 1% (1,5 g de agarose em 150 mL de tampão tris borato EDTA - TBE) e 12  $\mu$ L de brometo de etídio (10 mg/mL). Para carregamento do gel foram usados 6  $\mu$ L de amostra mais 1  $\mu$ L de *Loading Dye* por poço. Para a corrida, a fonte de tensão foi regulada para 400mA, 80V contínuos durante 90 minutos, na tina de electroforese foi usado o tampão TBE a 10%. Durante a análise ao gel, ao ser visível uma banda nítida com o peso molecular esperado, a execução da PCR foi validada e os produtos armazenados a -20°C.

### 2.2.3. Agrupamento por amostra

Após o PCR dos diferentes fragmentos, foram obtidas diferentes concentrações de amplicões, o que se verificou pela intensidade das bandas durante a electroforese. Segundo o fabricante, o protocolo Nextera XT requer uma padronização através da quantificação objectiva das concentrações dos produtos PCR, porém esta estratégia iria esgotar o produto PCR para as próximas fases.

Uma vez que o método de amplificação foi constante (incluindo quantidades e concentração de reagentes utilizados no PCR, quantidade de amostra depositada para visualização no gel de agarose etc.), considera-se que as intensidades das bandas no gel de agarose reflectem as concentrações dos amplicões. Assim, foi desenvolvida uma estratégia de quantificação visual. A intensidade das bandas no gel foi classificada e anotada entre 1x, 3x, 5x e 10x concentrado (sendo 1 a classificação mais baixa, e 10 a classificação mais elevada - logo a mais concentrada).

Através desta classificação foi adicionado produto PCR àqueles que tiveram uma classificação baixa para padronizar as concentrações (tabela 7). De seguida, os produtos PCR foram agrupados por amostra e verificados por electroforese sob as mesmas condições anteriores.

Tabela 7. Volume dos amplicões adicionados ao agrupamento da amostra de modo a padronizar as concentrações dos produtos PCR

<b>Classificação</b>	<b>Volume (<math>\mu</math>L)</b>
1	*Repetição
3	10
5	5
10	1

\*Amostra insuficiente requerendo a repetição da PCR

#### 2.2.4. Purificação dos produtos de PCR

Das amplificações por PCR resultam excedentes como restos de *primers* e reagentes (dDNTPs, sais e enzimas) que podem afectar a *base calling*, para evitar erros de leitura NGS os produtos de PCR foram purificados com recurso às *esferas* magnéticas AMPure XP (anexo 3- AMPure XP).

A cada produto PCR foi adicionado 1,8  $\mu$ L da solução com esferas magnéticas por 1  $\mu$ L de amostra de ADN, com incubação de 5 minutos à temperatura ambiente (para as esferas ligarem ao ADN). A placa com os produtos PCR foi colocada num suporte magnético para permitir a separação magnética das esferas ligadas ao ADN. O sobrenadante foi removido e descartado com o cuidado de não perturbar as esferas. As esferas foram ainda lavadas duas vezes com etanol a 70%. No passo final o ADN purificado foi eluído em 40  $\mu$ L de tampão de eluição para desagregar o ADN das esferas magnéticas (Beckman Coulter, 2002).

### 2.3. Sequenciação NGS

A sequenciação NGS segue um fluxo de trabalho característico por passos. Neste é ainda incluída a preparação das bibliotecas que antecipam a NGS.

#### 2.3.1. Preparação das bibliotecas de ADN

A preparação das bibliotecas foi realizada com o *kit* Nextera XT da Illumina®, segundo as instruções do fabricante, todo o processo foi realizado numa placa de 96 poços com disposição conforme a figura 4. Este procedimento incluiu os passos:

Tagmentação: o ADN foi aleatoriamente fragmentado e adicionados os adaptadores nas extremidades. Estes adaptadores servem de ligação entre as bibliotecas geradas e a *flow*

**Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$**

cell (superfície de ancoragem onde ocorre a sequenciação). As amostras foram quantificadas e purificadas pela plataforma Qubit, e diluídas para 0,2 ng/ $\mu$ L.

**Amplificação de bibliotecas:** foram adicionados aos adaptadores, 2 índices diferentes: um índice 1 (i7) e um índice 2 e (i5) (um de topo outro de base, respetivamente) - necessários para a gerar *clusters* e sequenciação e a identificação da amostra (figura 4). A organização dos índices foi escolhida segundo o *software Illumina Experiment Manager*, segundo configuração da figura 4. Foi executado o programa de PCR *Nextera* para amplificação, conforme o protocolo *Nextera XT DNA Library Prep* (Illumina, 2016a).

**Clean Up de bibliotecas:** para purificar as bibliotecas amplificadas, foram utilizadas novamente as esferas magnéticas *AMPure XP*, utilizando a metodologia de separação magnética. As esferas purificaram a biblioteca de ADN, além de removerem fragmentos muito pequenos. Foi repetido o procedimento de purificação segundo o protocolo *Instructions For Use AMPure XP* (Beckman Coulter, 2002).

**Normalização das bibliotecas:** os volumes de cada biblioteca foram normalizados, garantindo uma representação equitativa das bibliotecas quando agrupadas.

**Diluição e desnaturação de bibliotecas:** após o processo de normalização foi feito um *pool* das bibliotecas, com 10  $\mu$ L cada. A execução seguiu o protocolo *MiSeq System - Denature and Dilute Libraries Guide* (Kahl, 2015). Foi utilizado o genoma do fago Phix a 12,5 pM como controlo interno de sequenciação. A biblioteca PhiX fornece um controlo de qualidade para geração de *clusters*, sequenciação, alinhamento e calibração de leituras.

	1	2	3	4	5	6	7	8	9	10	11	12
A	SMD1_C1 S902 N716	SMD9_C10 S902 N716	SMD17_C18 S902 N719	SMD25_A15 S902 N720	SMD33_A27 S902 N721	SMD41_A37 S902 N722	SMD49_A45 S902 N723	SMD57_A56 S902 N724	SMD65_T5 S902 N725	SMD73_T13 S902 N727	SMD81_T21 S902 N728	SMD89_T29 S902 N729
B	SMD2_C2 S903 N716	SMD10_C11 S903 N716	SMD18_A1 S903 N719	SMD26_A16 S903 N720	SMD34_A28 S903 N721	SMD42_A38 S903 N722	SMD50_A46 S903 N723	SMD58_A57 S903 N724	SMD66_T6 S903 N725	SMD74_T14 S903 N727	SMD82_T22 S903 N728	SMD90_T30 S903 N729
C	SMD3_C3 S905 N716	SMD11_C12 S905 N716	SMD19_A2 S905 N719	SMD27_A17 S905 N720	SMD35_A29 S905 N721	SMD43_A39 S905 N722	SMD51_A47 S905 N723	SMD59_A58 S905 N724	SMD67_T7 S905 N725	SMD75_T15 S905 N727	SMD83_T23 S905 N728	SMD91_T31 S905 N729
D	SMD4_C4 S906 N716	SMD12_C13 S906 N716	SMD20_A4 S906 N719	SMD28_A19 S906 N720	SMD36_A30 S906 N721	SMD44_A40 S906 N722	SMD52_A51 S906 N723	SMD60_A59 S906 N724	SMD68_T8 S906 N725	SMD76_T16 S906 N727	SMD84_T24 S906 N728	SMD92_T32 S906 N729
E	SMD5_C5 S907 N716	SMD13_C14 S907 N716	SMD21_A8 S907 N719	SMD29_A20 S907 N720	SMD37_A31 S907 N721	SMD45_A41 S907 N722	SMD53_A52 S907 N723	SMD61_T1 S907 N724	SMD69_T9 S907 N725	SMD77_T17 S907 N727	SMD85_T25 S907 N728	SMD93_T34 S907 N729
F	SMD6_C6 S908 N716	SMD14_C15 S908 N716	SMD22_A10 S908 N719	SMD30_A21 S908 N720	SMD38_A34 S908 N721	SMD46_A42 S908 N722	SMD54_A53 S908 N723	SMD62_T2 S908 N724	SMD70_T10 S908 N725	SMD78_T18 S908 N727	SMD86_T26 S908 N728	SMD94_T44 S908 N729
G	SMD7_C7 S910 N716	SMD15_C16 S910 N716	SMD23_A11 S910 N719	SMD31_A24 S910 N720	SMD39_A35 S910 N721	SMD47_A43 S910 N722	SMD55_A54 S910 N723	SMD63_T3 S910 N724	SMD71_T11 S910 N725	SMD79_T19 S910 N727	SMD87_T27 S910 N728	
H	SMD8_C8 S911 N716	SMD16_C17 S911 N716	SMD24_A14 S911 N719	SMD32_A26 S911 N720	SMD40_A36 S911 N721	SMD48_A44 S911 N722	SMD56_A55 S911 N723	SMD64_T4 S911 N724	SMD72_T12 S911 N725	SMD80_T20 S911 N727	SMD88_T28 S911 N728	

Figura 4. Organização dos índices, visualização superior da placa de 96 poços. Número do índice do topo (S, correspondente ao i7) e do fundo (N, correspondente ao i5). Indicação da organização das amostras por

poço, C - crônico, A - agudo e T – controlo (imagem do sistema informático *Illumina experiment manager*<sup>TM</sup>).

### 2.3.2 Fase I da NGS

O procedimento de sequenciação começou por transferir o produto de desnaturação das bibliotecas para o cartucho do MiSeq e carregados os reagentes necessários para o aparelho, segundo o fabricante (Illumina®, 2019). A *flow cell* foi limpa para retirar o excesso de tampão de armazenamento e colocada no compartimento respetivo. Todos os passos seguiram as instruções conforme o guia do sistema (Illumina®, 2019).

Este passo permitiu a sequenciação massiva por síntese para identificação das sequências de ADN. A sequência dos índices foi reconhecida pelo *software* do equipamento MiSeq, apenas foram consideradas, como leituras, as a sequência entre os índices (i5 e i7) (figura 5- Fase I). Foram indicados pelo MiSeq os parâmetros de controlo de qualidade: dispersão de leituras, *Cluster* e Q30.

### 2.3.3 Fase II da NGS

Para reconhecimento do ADN e identificação dos SNPs, iniciou-se o processo de *variant call* recorrendo a metodologias bioinformáticas (anexo 5 - Imagem *Docker*), estas foram executadas pelos seguintes passos:

Mapeamento: as sequências desmultiplexadas<sup>4</sup> foram mapeadas no genoma de referência através do Bowtie 2 v2.3.4 (Langmead et al., 2019) (figura 5 - Fase II: Bowtie). Foi considerado, como referência, o genoma de humano versão GRCh38.p12, disponível em Ensembl.com (Ensembl, 2019). As leituras mapeadas foram agrupadas por coordenadas genómicas pelo software SAMtools v1.9 (figura 5– Fase II: SAM) (Li et al., 2009). As leituras duplicadas foram identificadas e marcadas pela ferramenta GATK v4.1.5.0 com o módulo *MarkDuplicates* (figura 5 – Fase II: GATK) (Van der Auwera et al., 2013; Broad Institute, 2020).

Variant call: os valores de qualidade do *base call* foram corrigidos com o *software* GATK (ferramentas *BaseRecalibrator* e *ApplyRecalibration*) removendo “o”, entradas sem alelos alternativos e alelos alternativos ambíguos ou repetidos (erros sistemáticos) usando a base de dados *Germline variations* (transferida do Ensembl *Release 96*) (figura 5 - Fase II: GATK). O arquivo de alinhamento resultante foi usado para

---

<sup>4</sup> Sequências separadas por amostra sem adaptadores ou índices

determinação de variantes. Posições variáveis em comparação com a sequência de referência foram inicialmente chamadas para cada amostra usando a ferramenta GATK HaplotypeCaller no formato *Genomic Variant Call Format* (GVCF) (figura 5- Fase II: GATK).

Anotação: Usando a base de dados *Ensembl Release 96 germline variants*, as variantes conhecidas foram anotadas. Para as variantes com *minimum call confidence* de 20 pelo *HaplotypeCaller* a genotipagem conjunta foi feita com a ferramenta GATK *Genotype GVCFs* (figura 5 - Fase II: GATK).

Filtragem: Os genótipos das amostras e posições variáveis foram anotados e filtrados usando GATK (módulos *SelectVariants* and *VariantFiltration*), *software* BCFtools v1.9 e *software* PLINK v1.9 (Li, 2011; Chang et al., 2015). Em primeiro lugar, os genótipos das amostras foram filtrados (figura 5 - Fase II: *BCF Tools*) quando a profundidade de leitura (DP) foi inferior a 10x e um alto desequilíbrio de DP alélico na posição heterozigótica putativa foi observado ( $30\% \leq$  equilíbrio alélico aceite  $\leq 70\%$ , figura 5). Em seguida, várias posições foram filtradas sequencialmente quando:

- 1) a variante foi encontrada numa região com mais de 8 unidades repetitivas;
- 2) a variante SNP estava proximal em 2 bp de um INDEL<sup>5</sup>;
- 3) não foi possível genotipar com segurança pelo menos 50% das amostras.

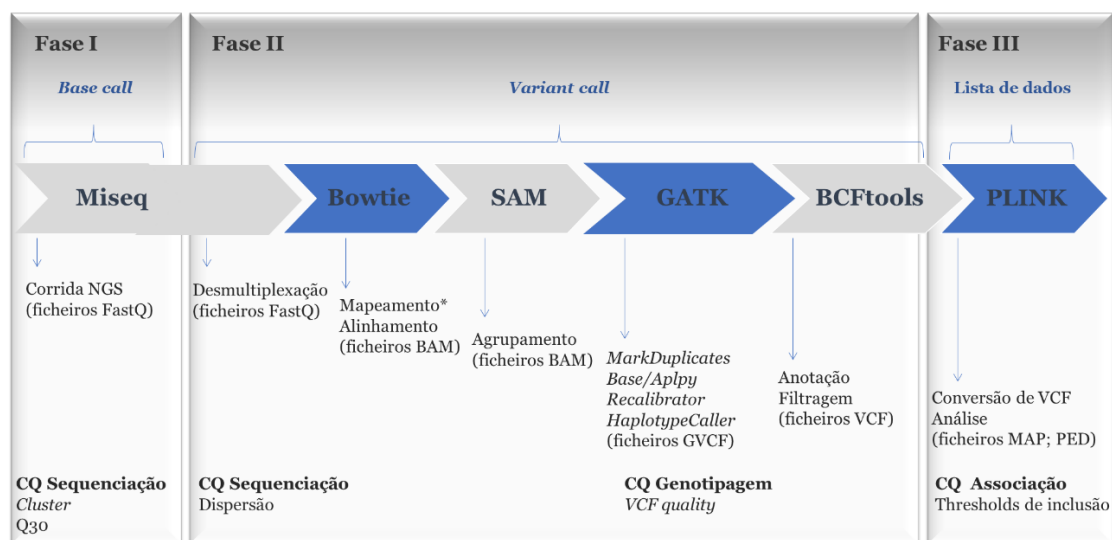


Figura 5. Fluxo da metodologia NGS em fases (I, II, III), indicação das ferramentas bioinformáticas usada em cada passo com os respetivos parâmetros/métricas de qualidade. Formatos, FASTQ: formato baseado caracteres ASCII; BAM: ficheiro binário de armazenamento de dados; GVCF: *Genomic Variant Call Format*; VCF: *Variant Call Format*; MAP: mapa dos SNPs; PED: *Pedigree Data*; CQ: Controlo de Qualidade. \*no genoma de referência; Instruções dos comandos presente no anexo “Imagem Docker”.

<sup>5</sup> Termo que descreve a inserção ou deleção de bases num gene.

## 2.4. Lista de dados - Preparação dos dados de Fase III para associação genótipo-fenótipo por CGAS

Nos estudos de associação genética é recorrente existirem vieses, como inclusão de etnias diferentes, cruzamentos, migrações, perfis genéticos entre outros o que pode induzir vieses. Para diminuir o risco de ocorrerem vieses foi necessário aplicar critérios de ajustes específicos, uma vez que os dados finais são dirigidos para um CGAS.

Foram incluídos controles de qualidade: proporção de genótipos ausentes por indivíduo e proporção de genótipos ausentes por marcador; verificando o equilíbrio de Hardy-Weinberg no grupo controle e a frequência do alelo minoritário.

Os estudos de controle de qualidade foram realizados com a ferramenta PLINK. O PLINK permite realizar uma análise exploratória (*summary statistics*) sobre os dados da genotipagem através da aplicação de níveis de aceitação (*thresholds*) sobre parâmetros de qualidade específicos que incidem, tanto nas amostras como nos marcadores. Os *thresholds* valores *threshold* aplicados foram os recomendados para os CGAS (Anderson et al., 2010), conforme consta na tabela 8.

Tabela 8. Lista de comandos utilizados na ferramenta PLINK para a análise exploratória e controle de qualidade da lista de dados da NGS e níveis de aceitação (*thresholds*) correspondentes

Ordem	Função	Análise exploratória	Comando <i>threshold</i>
1 <sup>o</sup>	Ausência por amostra <sup>1</sup>	--missing	--mind 0,1
2 <sup>o</sup>	Ausência por marcador <sup>2</sup>	--missing	--geno 0,05
3 <sup>o</sup>	Equilíbrio Hardy-Weinberg <sup>3</sup>	--hardy	--hwe 0,00001
4 <sup>o</sup>	Frequência alélica <sup>4</sup>	--freq	--maf 0,01

<sup>1</sup> Proporção de genótipos ausentes por indivíduo, *threshold* retido < 0,1

<sup>2</sup> Proporção de genótipos ausentes por marcador *threshold* retido < 0,05

<sup>3</sup> Significância do Equilíbrio Hardy-Weinberg (aplicados aos controles), *threshold* retido < 0,00001

<sup>4</sup> Frequência do alelo minoritário, *threshold* retido > 0,01

### 2.5.1. Ausência de genótipo

Da genotipagem resultam grandes variações na qualidade de ADN por amostra, este efeito pode causar vieses na taxa de inclusão ou na identificação precisa do genótipo.

Para apreciação dos dados ausentes foi aplicado o comando *missing*. Este comando devolve duas listas uma de dados ausentes por amostra e uma de dados ausentes por marcador (por exemplo SNPs). Para remoção de dados com baixo nível de genotipagem

foi aplicado o comando *mind* com o *threshold* 0,1 (removeu os indivíduos com uma proporção de genótipos ausentes superior a 10%) e o comando *geno* com *threshold* 0,05 (removeu os marcadores com uma proporção na ausência de genotipagem superior a 5%).

O efeito da proporção de genótipos ausentes por indivíduo pode afectar a taxa de heterozigotia que por sua vez afecta os resultados em estudos de associação genótipo fenótipo. Para determinar o *threshold* aceitável para a proporção de genótipos ausentes por indivíduo foi examinado cuidadosamente a taxa de genotipagem ausente em todas as amostras através da taxa de heterozigotia observada por indivíduo (obtido pela equação 1) contra a proporção de SNPs ausentes.

$$O(RHet) = \frac{N(NM) - O(Hom)}{N(NM)}$$

Equação 1. Cálculo da taxa de heterozigotia observada por indivíduo. *O(RHet)*: taxa de heterozigotia observada por indivíduo; *O(Hom)*: número genótipos homozigóticos observados; *N(NM)*: número de genótipos presentes (Anderson et al., 2010)

## 2.5.2. Equilíbrio de Hardy-Weinberg

O equilíbrio de Hardy-Weinberg (HW) é um princípio que afirma que a variação genética numa população permanecerá constante entre gerações na ausência de factores perturbadores. Podendo estes factores tratar-se de evolução, selecção natural, estratificação da população, ou até mesmo associação real com a característica em estudo.

Para garantir a fiabilidade dos resultados foi testado o equilíbrio de HW apenas no grupo controlo através da apreciação estatística com o comando *hardy*. As variantes que desviam dos pressupostos de HW, foram eliminadas com aplicação do comando *hwe*. No grupo de casos, os desvios do equilíbrio de HW podem resultar de associações com o fenótipo clínico e a eliminação das variantes em desequilíbrio seria contraproducente. Segundo as recomendações para CGAS, foi aplicado o *threshold* com valor de significância menor ou igual a 0,00001 (*p-value*). Este *threshold* é recomendado para detectar desvios de HW devidos a erros de genotipagem com o comando descrito na tabela 8 (Wigginton et al., 2005; Anderson et al., 2010).

Remover SNPs com diferenças significativas na taxa de genotipagem entre casos e controlos é outro meio de reduzir a discrepância e remover SNPs mal genotipados (Moskvina et al., 2006).

### 2.5.3. Frequência do alelo minoritário

Neste estudo é necessário determinar quais os SNPs mais relevantes. O poder de detectar efeitos genéticos é influenciado pela frequência de alelo minoritário (*Minor Allele Frequency* - MAF), este pode torna-se significativo ao desviar resultados concretos se a frequência for demasiado baixa. Na lista de SNPs existe uma ampla distribuição de MAFs, portanto, é crítico entender o efeito do MAF para remover dados que não sejam SNPs significativos, isto é, diminuir a taxa de falsos-positivos. Para tal foi feita a apreciação estatística com o comando *freq* e aplicado o comando de remoção com os critérios recomendados para este efeito, conforme a tabela 8.

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crónica na via de sinalização IL12-IFN $\gamma$***

## Capítulo IV

### Resultados e discussão

A finalização deste projecto agregou resultados pertinentes para a construção de um fluxo de trabalho para aplicar medidas de controlo de qualidade em estudos CGAS. Durante a execução do protocolo experimental destacaram-se pontos chave que determinaram a qualidade dos dados obtidos. Estes pontos incluem a preparação da NGS pela optimização da PCR e quantificação dos amplicões; os controlos de qualidade durante as fases I e II da NGS pela *base calling* e *variant calling*, respetivamente; e, por fim, os controlos de qualidade da fase III da NGS pelo controlo de qualidade da lista de dados para associação genótipo-fenótipo.

#### 1. Preparação NGS

Todos os fragmentos foram amplificados com sucesso conforme a figura 6. O mesmo processo de amplificação e verificação por gel de agarose foi executado nas 94 amostras, de forma a garantir a amplificação específica de cada fragmento.

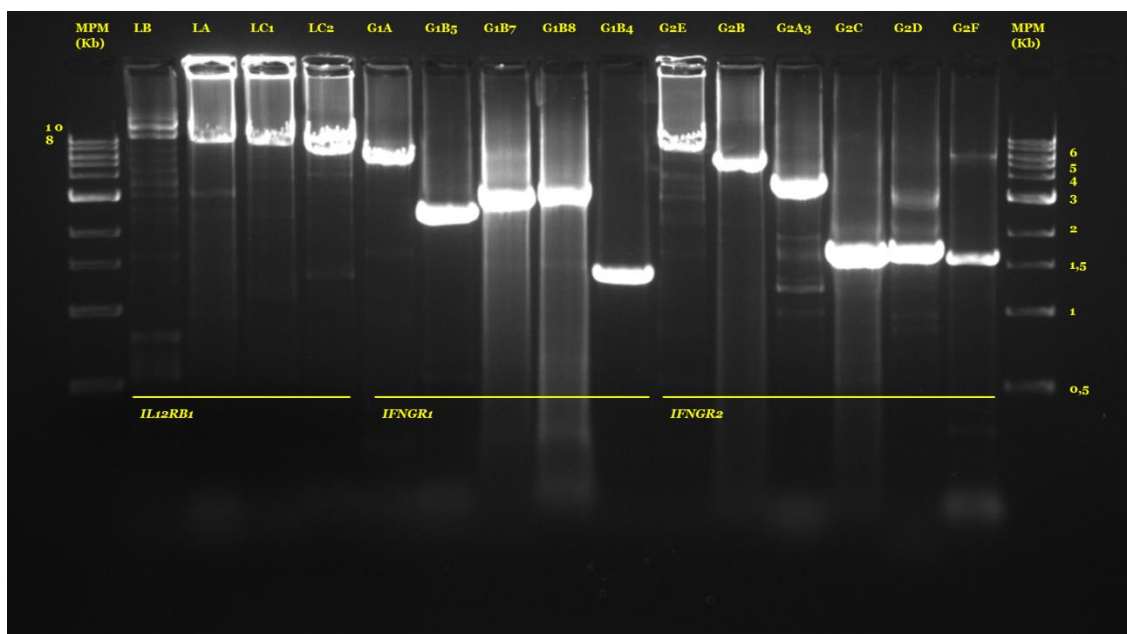


Figura 6. Resultado eletroforese de cada fragmento dos genes *IL12RB1*, *IFN $\gamma$ R1* e *IFN $\gamma$ R2*. Com condições de corrida 80v durante 90 minutos em tampão TBE. MPM: Marcador de peso molecular, *BIOLabs* 1Kb.

O número elevado em pb de cada gene obrigou à construção dos fragmentos, para ser possível amplificar por PCR, consoante a tabela 9. Esta construção foi ainda projectada para ser possível estabelecer os efeitos das variantes na sequência da síntese proteica, fusão nos locais de disrupção do ADN (*splicing*) e ligação dos factores de transcrição ao ADN (Esparragón et al., 2009).

A optimização do protocolo PCR permitiu maior eficiência e especificidade nos fragmentos amplificados (amplicões). Ainda assim, para algumas das amostras foram necessárias repetições das amplificações, isto pode ter-se devido a vários factores, desde a degradação do ADN, a degradação dos *primers*, à execução da PCR, ou até algum erro durante a preparação do gel. A amplificação dos fragmentos revelou ser um processo moroso, mas necessário para os resultados.

Como para cada amostra foram obtidos 20 amplicões, foi necessário o passo de agrupamento dos amplicões por amostra. Este agrupamento segue habitualmente a metodologia proposta *Nextera XT* com uma quantificação numérica procurando estabelecer uma quantidade homogénea entre cada um dos amplicões (Seth-Smith et al., 2019). A estratégia de agrupamento executada neste projecto baseou-se numa apreciação visual da fluorescência e verificação em gel de agarose dos grupos de amplicões. Esta quantificação visual revelou ser uma alternativa fiável neste projecto ao sugerido para a metodologia. Esta fiabilidade na qualidade dos amplicões reflectiu-se na qualidade dos dados após a sequenciação dos fragmentos.

Uma curiosidade neste trabalho foi a necessidade de dividir o par de *primers* do amplicão G2A3 em alíquotas de uso único. Foi verificado que este *primer* deixou de amplificar, depois de nova construção e diversas optimizações específicas para este fragmento, foi concluído que este *primer* era demasiado instável. Este obstáculo foi ultrapassado pela alíquotação do *primers* para usar em alíquotas de uso único.

Tabela 9. Dados do gene e amplificação segundo o sistema de coordenadas cromossômicas disponíveis em Ensembl, versão do genoma GRCh38.p12, conforme o primeiro transcrito.

Gene	Fragmento				Fim				
	Localização Genômica	Orientação	Coordenadas Cromossômicas	Tamanho	Início Sequencia	Coordenada Sequencia			
IL12RB1	19p13.11	Reversa	18058995-18098944	LA	10764	AATAGCACTGACCACTCGGG	18089867	CTGACCTCATAGCCCTGCTGG	18100631
				LB	13557	TTTAGAGGCTTTGCCGACAGT	18075127	CAATGTGGCATCCGGTCTTG	18088684
				LC1	9747	AGGAGAAATGGGTGCTCAA	18065426	ACTTGAGGGTCTCGATCACC	18075173
				LC2	7575	CCAGCCAAGTGTCTATGCTG	18057818	TTGGGTGCGAAGTAGGAG	18065393
IFNGR1	6q23.3	Reversa	137197484-137219449	G1A	6068	TGTCTACTCCATAGAGGCC	137214674	ATGCCAGACCAACCCAACT	137220742
				G1B4	1414	GGGCAATGTGGCATCTTACAA	137205750	TCCTCCCTTCTTTCAAAGCCC	137207164
				G1B5	2635	ACAATGCTTAAGAAACACTTGG			
				G1B7	3205	CCCTTGACCAGGAAAGAAA	137200077	TGCCATTTACACGGCTAG	137206039
				G1B8	3329	ACCAGCAGTAGCATAACGGG	137196715	GCACACATTAGAAATGGGAAAAGCA	137200044
				G2A3	3789	TCTAACTGTATGCCCGTGTCT	33401040	ACCACGTTATGAGCCAAAC	33404829
				G2B	5587	AGGCAAGAAAGTTACAGCAAGT	33410124	GTCCAGTGTGTGAGATGGG	33415711
				G2C	1819	TGGAAGAAAACTCGGCACCA	33420534	CCGTGTCTGGTTTATGCACAC	3342353
IFNGR2	21q22.11	Frontal	33402896-33479348	G2D	1794	GAGAAATCCAGCCCTGCTGTT	33425848	TCAGCCTCGGGGAGTATAGG	33427642
				G2E	8033	GCAGAGGAGACACAGAACCTT	33431071	GAAAAACAAAGGCCCACTCCC	33439104
				G2F	1635	CCCAGTACTTTCAGAGACCGAG	33478213	GGACTCACCCAGCTTCCCTCA	33479848

## 2. Fase I e II - Controlo de qualidade

Existem diversas metodologias científicas para aplicar controlos de qualidades em estudos CGAS. É importante ter em conta que a NGS deste trabalho incluiu outros genes durante a sequenciação para outros estudos.

### 2.1. Base call

Da sequenciação massiva por NGS, resultaram milhares de milhões de leituras. A percentagem global de leituras identificadas alcançou os 97,094%, este dado revela uma altíssima percentagem de reconhecimento dos amplicões, conseguido pelas estratégias de optimização, amplificação e agrupamento que foram aplicadas.

Dentro da percentagem global foram ainda identificadas as amostras com maior e menor sucesso, sendo elas a T5 (contribuiu com 1,7009%) e a T32 (contribuiu com 0,0015%), respectivamente (figura 7). A associação dos pontos, índices e percentagens de leitura estão presentes no anexo 4 - Tabela de dispersão de leitura. Nestes resultados é possível verificar uma concentração homogénea no intervalo 0,6% a 1,4% com variação de poucos pontos. Estes pontos podem ter sido amostras cujo produto PCR degradou ou ocorreu um erro de sequenciação, pois, todos os produtos foram validados por electroforese antes da sequenciação NGS.

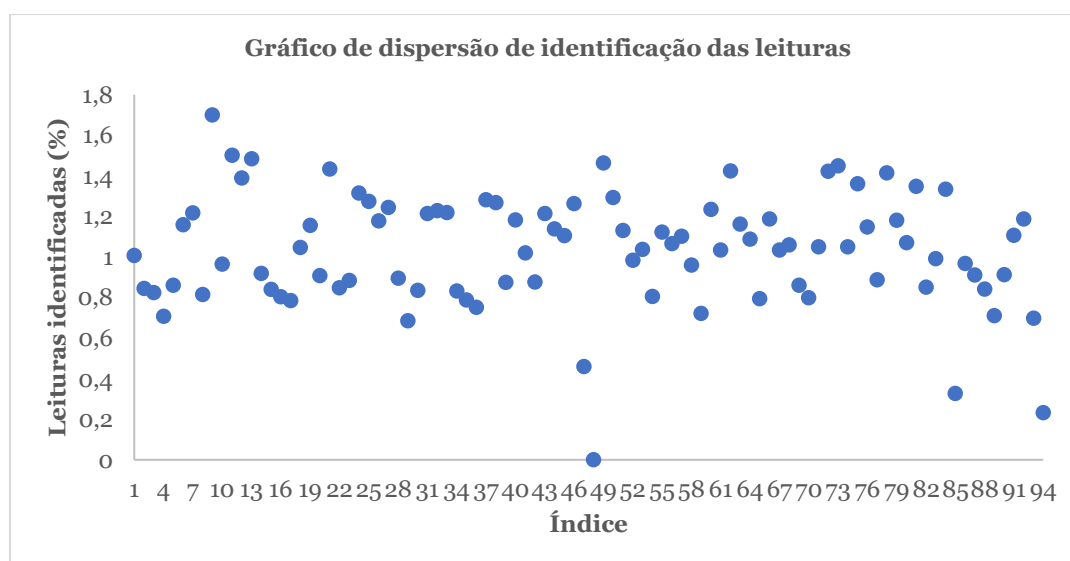


Figura 7. Gráfico de dispersão da percentagem de identificação das leituras em função do índice. Dados numéricos no anexo “Tabela de dispersão de leitura”.

### 2.1.1. Cluster

Nesta sequenciação foi obtida uma densidade média dos *clusters* de 1549 K/mm<sup>2</sup>, o filtro de densidade do *cluster* permitiu a inclusão 86,30% das leituras na lista de dados. Este determina-se como um resultado excelente, segundo os parâmetros de qualidade da Illumina® (Illumina, 2016b).

Uma vez que as leituras são feitas por luminescência, estas podem-se sobrepor, induzindo um erro na identificação da base. Se o *clustering* for alto, vão existir *clusters* sobrepostos (provenientes de fragmentos originais de ADN diferentes) que não permitem a detecção correcta da base a ler para cada um dos *clusters* (e consequentemente de cada um dos fragmentos de ADN originais).

### 2.1.2. Q30

A sequenciação por NGS tem margens de erro associadas, no meio de milhares de milhões de leituras é possível que ocorram trocas na identificação de uma base. Para identificar este problema foi tido em conta o *qScore*, cujo resultado foi 89.8% de Q30 (figura 8).

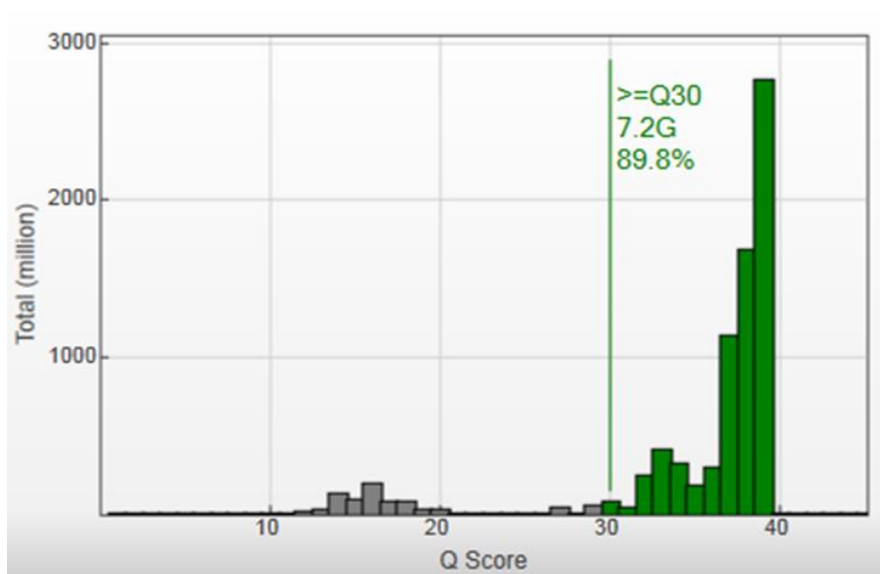


Figura 8. Dados de *qScore* após sequenciação, gerados pelo aparelho MiSeq.

Após o controlo de qualidade interno do MiSeq e filtragens de dados, contabilizaram-se 129253693 leituras. Alguns destes representam erros de: leitura, cobertura do aparelho ou outros fatores imprevisíveis. A NGS permitiu uma rápida sequenciação em larga escala das amostras. A escolha pela metodologia *Nextera XT* assentou na rotina laboratorial e qualidade de sequenciação.

Ainda que a tecnologia de genotipagem tenha tido um grande avanço tecnológico na última década, a sequenciação por NGS continua a ser altamente dispendiosa. Uma vez que esta dissertação se insere num projecto de maior escala, esta corrida NGS foi incluída num painel de genes a serem aplicados em outros estudos, para assim rentabilizar tempo e fundos monetários.

Em corrida NGS, dentro do mesmo painel de amostras podem existir leituras mais robustas que outras. No controlo interno do MiSeq, ao ser verificada uma grande dispersão nos dados são impostas dúvidas, tornando os resultados dúbios (Pfeifer, 2017).

## **2.2. Variant call**

Durante a fase II da NGS, após anotação e filtração dos dados, totalizam-se na via IL12-IFN $\gamma$  393 SNPs. Na tabela 10 estão presentes os números de SNPs identificados por cada gene.

Tabela 10. Resultados da genotipagem, em número de SNPs, nas 94 amostras.

Gene	Número de SNPs
<i>IL12RB1</i>	249
<i>IFN<math>\gamma</math>R1</i>	52
<i>IFN<math>\gamma</math>R2</i>	92

Os parâmetros usados na genotipagem partiram da experiência e conhecimento da equipa da Unidade de Tecnologia e Inovação do INSA, IP, responsável. Foi estipulada uma cobertura de profundidade de 10 vezes, ou seja, num determinado locus teriam de existir pelo menos 10 leituras igualmente representadas para estas serem validadas, isto garante que leituras pouco específicas não sejam incluídas aquando do reconhecimento pelo fragmento de referência (genoma de referência).

O uso de metodologias baseadas em NGS, têm atingido um equilíbrio custo-benefício que torna os CGAS mais acessíveis para estudar genes candidatos a causar doenças. Esta

tecnologia permite a inclusão de mais indivíduos por grupo, estudar sequências alvos nos genes (sejam elas regiões codificantes, não-codificantes ou flanqueadoras, e locais de *splicing* ou regulação) (Jorgensen et al., 2009). Para além de outras vantagens como aplicação de investigações clínicas nos hospitais para variantes de baixa frequência, e a disponibilidade de bases de dados e ferramentas bioinformáticas (Patnala et al., 2013; David, 2021).

### **3. Fase III - Controlo de qualidade**

A fase III utilizou os ficheiros decorrentes das análises anteriores, após aplicação dos vários controlos de qualidade associados à fase I e fase II. Os critérios de controlo de qualidade utilizados dependem da análise subsequente (clínica ou associação), no caso desta dissertação são próprios à preparação da lista de dados para fins exploratórios por CGAS. Estas métricas de controlo de qualidade devem ser aplicadas para reduzir ao máximo o risco de obter resultados espúrios (David, 2020).

Nesta dissertação, a fase III da NGS ficou pela análise e implementação dos controlos de qualidade. As próximas fases seriam a análise de associação genótipo-fenótipo e a análise destes resultados nas bases de dados disponíveis (estudos *in silico*) e por eventuais estudos funcionais (laboratoriais) para se compreender as consequências destes SNPs.

#### **3.1. Controlo de qualidade para CGAS**

Nos CGAS existem diversos factores que elevam a margem de erro, sendo estes factores tanto internos ao estudo como externos. Para assegurar um estudo fiável e reprodutível os passos de controlo de qualidade são imprescindíveis para se obterem resultados concretos e com relevância científica (Border et al., 2019).

Os problemas causados por falsas deduções conseguem ser minimizadas por um controlo de qualidade apropriado (Anderson et al., 2010). A garantia de qualidade também desempenha um papel importante na identificação de vieses ao serem reduzidos e até eliminados durante a fase de análise do CGAS (Morin et al., 2010).

Este controlo de qualidade incidiu sobre 393 SNPs e, após remoção dos SNPs que não passaram os *thresholds* de inclusão, finalizou com 209 (figura 10). Em cada controlo foi contabilizada a remoção de SNPs e analisada a aplicação de cada *threshold*.

### 3.1.1. Missing

Do controlo de ausência de dados ou baixa qualidade de genotipagem (*missing*), resultaram 11 remoções de indivíduos (*threshold* 0,1) e a remoção de 22 SNPs (*threshold* 0,05).

Referente às amostras, na tabela 11 estão presentes os indivíduos retirados com o número de SNPs ausentes e a proporção dos mesmos sobre os genes da via da via IL12-IFN $\gamma$ .

Tabela 11. Resultado da aplicação do tratamento *missing*. Lista de indivíduos retirados do estudo por baixa qualidade de genotipagem. Amostra de indivíduo: C- crónico, A- agudo, T-população geral.

Indivíduo	Número de SNPs ausentes	Proporção de SNPs ausentes
A11	44	0,112
A47	46	0,117
T27	48	0,1221
A21	49	0,1247
A15	57	0,145
T18	65	0,1654
C12	75	0,1908
A10	80	0,2036
T28	91	0,2316
C16	142	0,3613
T32	393	1

Referente aos marcadores, o *threshold* aplicado pode alterar significativamente a taxa de inclusão de genótipos, logo altera também a qualidade da lista de dados final para os resultados em CGAS.

Se o *threshold* estiver muito baixo, serão incluídas amostras com baixa qualidade nas genotipagens ou até mesmo genotipagens erradas. O que se iria reflectir nos resultados com baixa associação entre SNPs e a doença. Ou mesmo, associações erradas entre os grupos de estudo.

Se o *threshold* estiver muito alto, serão incluídos apenas genótipos com elevada certeza, o que pode resultar na exclusão excessiva de dados, além de que, um *threshold* alto neste controlo irá produzir uma alta proporção de SNPs ausentes, reduzindo assim a cobertura genómica e novamente a capacidade estabelecer associação.

A escolha do *threshold* adequado para este estudo fundamentou-se também na verificação da proporção dos genótipos ausentes contra a taxa de heterozigotia (figura 9). Na figura 9, não se verifica uma maior incidência da ausência de genótipo por amostra sobre o rácio da heterozigotia. Por esta razão considerámos suficiente aplicar um critério de inclusão apenas sobre a amostra (proporção de genótipos em falta > 0,1), sem a necessidade de aplicar um critério de inclusão sobre o rácio da heterozigotia como no exemplo no anexo 1 - Exemplo de Anderson, onde se verifica em algumas amostras o efeito importante da ausência de genótipo sobre a dispersão no rácio da heterozigotia.

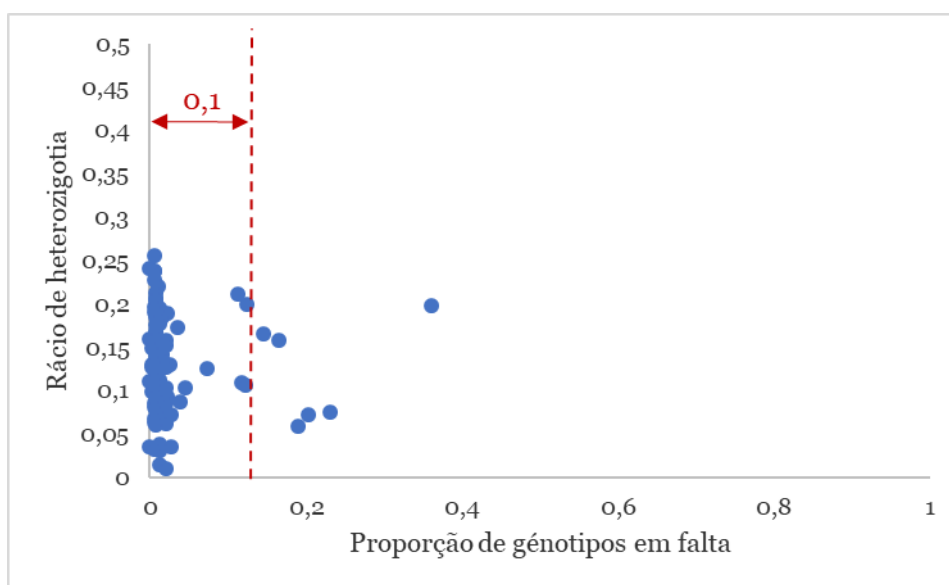


Figura 9. Proporção de genótipos em falta por amostra *versus* taxa de heterozigotia. Cada ponto indica a proporção de genótipos ausentes por indivíduo, as linhas vermelhas indicam os *threshold* definidos: 0,1 para indivíduos; 0,05 para SNPs.

Não foram encontradas diferenças na taxa de genotipagem entre casos e controlos para todos os SNPs analisados. Os indivíduos não eram relacionados, pois caso acontecesse poderiam ocorrer uma vez que estas metodologias de análise assumem implicitamente a independência das amostras (Burton et al., 2007).

### 3.1.2. HWE

A análise ao HWE destacou a remoção de 3 SNPs com *p-value* do teste menor que 0,00001, estes SNPs, presentes na tabela 12, foram excluídos.

Tabela 12. SNPs retirados pelo equilíbrio de Hardy-Weinberg com *threshold* 0,00001.

Cromossoma	SNP	Fenótipo	A1	A2	<i>p-Value</i>
19	rs439409	Agudo	A	G	0,00000401000
6	rs11914	Agudo	C	A	0,00000004850
6	rs9389480	Agudo	A	G	0,00000951000

A1: Alelo maioritário; A2: Alelo minoritário.

No grupo de casos o desvio deste equilíbrio pode indicar evolução, selecção ou estratificação da população e, até mesmo, associação real com a característica em estudo (Cox, 2005). Mas neste nível muito baixo de *threshold* é utilizado apenas para verificar eventuais erros de genotipagem.

Ao aplicar este teste no grupo casos, na presença de uma alteração causal também podiam ser eliminados os *loci* associados à doença, isto seria obviamente contraproducente (Wittke-Thompson et al., 2005). Portanto, apenas amostras do grupo controlo são usadas ao testar desvios para o equilíbrio HW. O limite de significância para declarar SNPs em equilíbrio de Hardy-Weinberg varia entre vários estudos, *p-value* entre 0,001 e  $5,7 \times 10^{-7}$  foram relatados na literatura (Burton et al., 2007; Meyre et al., 2009).

A maioria dos estudos de associação genética opta por excluir marcadores que mostrem um desvio do equilíbrio de Hardy-Weinberg (HWE) a um nível extremo do *p-value* do teste porque, uma vez não esperadas alterações a nível populacional, restam a possibilidade destes desvios serem resultado de erros de genotipagem ou inclusão (Anderson et al., 2010).

### 3.1.3. MAF

Nesta etapa foram removidos 167 SNPs por apresentarem MAF inferior a 0,01. Os sinais de associação vistos nestes SNPs raros são menos robustos, porque o tamanho da amostra do estudo é pequeno. Sendo esse poder de detectar associação em variantes raras tão baixas, também a sua remoção não causa impacto excessivo no estudo (Morris & Zeggini, 2010).

A etapa final ao conduzir o CQ é remover todos os SNPs com um MAF muito baixo. Normalmente é aplicado um *threshold* no MAF de 1-2% , mas estudos com pequeno tamanho de amostras podem precisar definir um *threshold* mais alto (Morris & Zeggini, 2010). O *threshold* de 0,01, aplicado neste estudo, é o sugerido por Anderson e colaboradores para CGAS, uma vez que o objectivo fundamental de um estudo de associação de caso-controlo é comparar a diferença de frequência alélica entre casos e controlos, a inclusão de SNPs com baixa frequência alélica poderia causar vieses durante a associação dos SNPs entre casos e controlos, acabando por reduzir a capacidade para encontrar SNPs causadores de suscetibilidade à doença (Anderson et al., 2010).

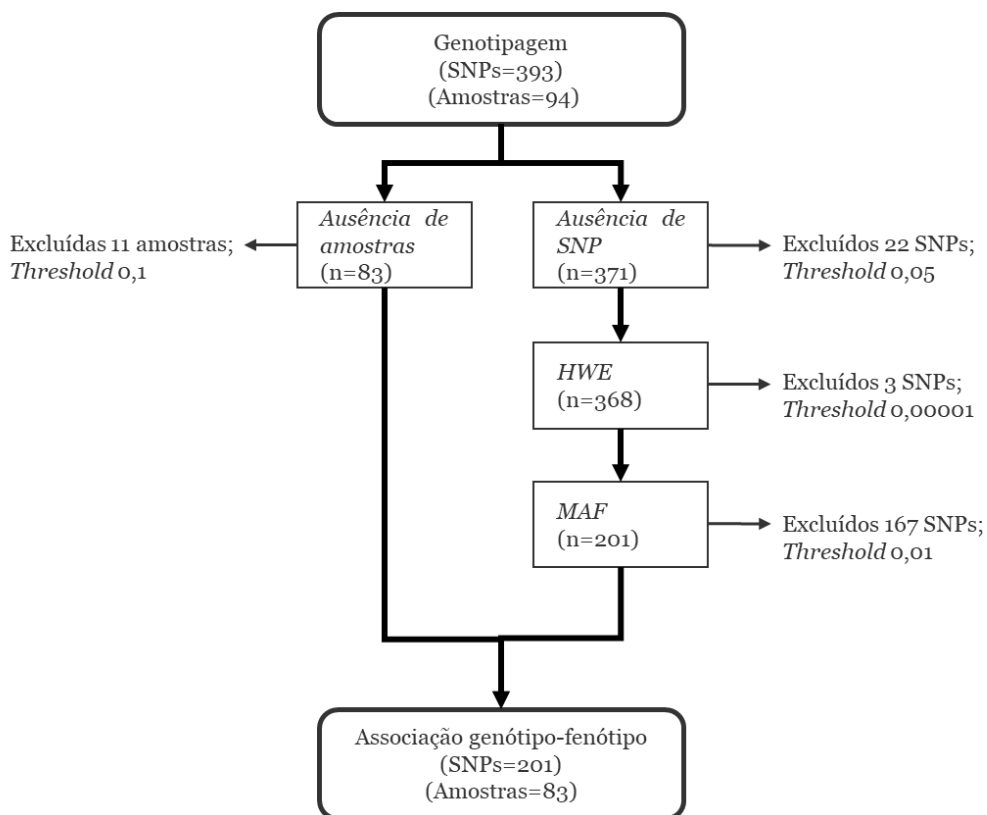


Figura 10. Diagrama de controlo de qualidade para associação genótipo-fenótipo (fase III); SNP: *Single Nucleotide Polymorphism*, MAF: *Minor Allele Frequency*, HWE: *Hardy-Weinberg Equilibrium*.

### 3.2. Lista de dados para associação

Ao contrário dos estudos de genoma completo, os estudos CGAS requerem um forte e rigoroso controlo de qualidade, caso não seja feito, é fortemente reduzida a eficiência em identificar e remover SNPs ou indivíduos que desviam a associação. Esta é talvez outra razão pela qual os estudos de genes candidatos tipicamente não rendem associações suficientes ou reproduzíveis entre genes e doenças (além do uso de amostras de tamanho reduzido, cobertura insuficiente da variação genética e escolha inadequada de candidatos).

Repetir genotipagem dos erros considera-se uma boa prática para recuperar dados. No entanto, requer mais tempo de trabalho com custos avultados para resequenciar, com uma baixa probabilidade de sucesso. A possibilidade de encontrar dados relevantes nos indivíduos que falharam foi considerada baixa, pelo que não foram resequenciados. No entanto, as amostras foram arquivadas com a possibilidade de se fazer sequenciação automática de *Sanger* em locus que se demonstrem importantes. Em todo o processo de controlo de qualidade foram reunidos os SNPs relevantes para o estudo de associação na via IL12-IFN $\gamma$  a evolução está ilustrada na figura 11.

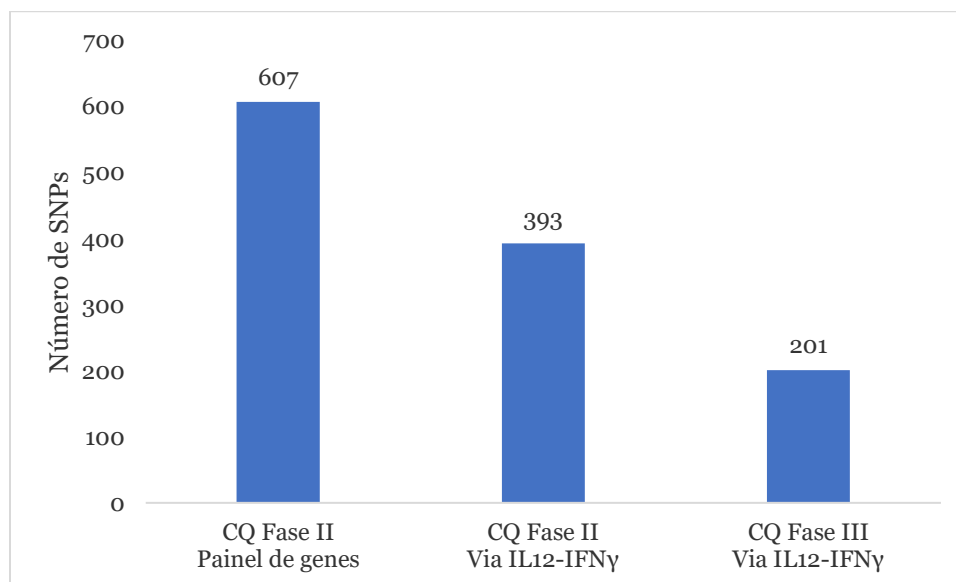


Figura 11. Evolução do número de SNPs ao longo do estudo. Número de SNPs após controlo de qualidade da Fase II da NGS no \*painel de genes total e nos genes da via IL12-IFN $\gamma$ , e número de SNPs após controlo de qualidade da Fase III dos genes da via IL12-IFN $\gamma$ ; CQ: Controlo de Qualidade.

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$***

Do controlo de qualidade resultou então uma lista de dados com 83 amostras e 209 SNPs prontos para o CGAS (tabela 13).

Tabela 13. Resultado em número de SNPs após controlo de qualidade para CGAS nos genes da via IL12-IFN $\gamma$ .

Gene	Número de SNPs
<i>IL12RB1</i>	144
<i>IFN<math>\gamma</math>R1</i>	16
<i>IFN<math>\gamma</math>R2</i>	41

Mesmo nos estudos bem projectados, com arquitectura caso-control, podem existir subestratificações genéticas escondidas na população (ou a inclusão inadvertida de indivíduos de outra população), é uma possibilidade a ter em conta. A confusão ocorre quando a população subestruturada não é igualmente distribuída entre os grupos de caso e controlo. Neste cenário, um sinal de associação surgirá para um SNP ancestralmente informativo, não por causa de uma associação com o risco de doença, mas por causa das diferenças de frequência de alelos entre as populações que compõem os casos e controlos (Anderson et al., 2010).

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crónica na via de sinalização IL12-IFN $\gamma$***

## **Conclusão e perspectivas futuras**

O tratamento da febre Q crónica inclui um período de antibioterapia longo, de 3 a 4 anos. Esta terapia é extremamente agressiva para o doente, sendo reduzidas as chances de sucesso terapêutico. Terapêuticas baseadas em antibióticos são, cada vez mais, um factor agravante na resistência bacteriana, isto leva ao desenvolvimento de antibióticos mais agressivos o que acaba por avançar a resistência em simultâneo. As terapias com administração de factores imunológicos recombinantes, ao invés da administração de antibióticos, começam a ser um caminho explorado para escapar à resistência bacteriana. A febre Q, com o seu potencial de agente bioterrorista, é uma ameaça latente pouco caracterizada em Portugal. O conhecimento sobre os genes envolvidos no desenvolvimento da febre Q crónica é muito reduzido, embora se saiba que a administração exógena de IFN $\gamma$  tenha efeitos terapêuticos noutras patologias, não foram estudados SNPs nos receptores desta e outras interleucinas na febre Q. A falta de dados e conhecimento básico é uma limitação séria, por essa razão este estudo resume-se a um estudo pré-exploratório destas possíveis alternativas.

O estudo de determinantes genéticos da febre Q crónica iniciou há dois anos. Através do meu trabalho de dissertação foi gerada por NGS uma lista de SNPs em genes candidatos previamente selecionados, com a qualidade requerida para estudos de associação genótipo-fenótipo. O meu papel neste projecto foi desempenhado na finalização das amplificações dos fragmentos, verificação e preparação dos amplicões como também desenho estratégico, seguimento das fases I e II da NGS e aplicação das medidas de controlo de qualidade de fase III da NGS. Dediquei particular atenção à justificação da seleção dos genes da via IL12- IFN $\gamma$ , *IL12RB1*, *IFN $\gamma$ R1* e *IFN $\gamma$ R2*. Foram aplicados todos os controlos de qualidade requeridos para que esta lista de dados possa ser utilizada em CGAS.

Estes resultados contribuirão para iniciar uma nova vertente de investigação. Os mesmos dados genéticos podem ser aplicados e transformados em diferentes bases de dados.

No futuro, este estudo poderá sustentar novas linhas de investigação para gerar novos marcadores de diagnóstico precoce. Este estudo poderá ainda potenciar a prevenção de surtos ao compreender se existem variantes associadas à doença na via IL12-IFN $\gamma$  que promovem a febre Q crónica. Antecipar a cronicidade desta doença seria um importante avanço clínico para prevenção e prestação de melhores cuidados de saúde.

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crónica na via de sinalização IL12-IFN $\gamma$***

## Referências

- A.Nathalie, R. A. ; (2005). Is Q Fever an emerging or re-emerging zoonosis?, *36*, 327–349. doi:10.1051/vetres
- Aguilar-Olivos, N., del Carmen Manzano-Robleda, M., Gutiérrez-Grobe, Y., Chablé-Montero, F., Albores-Saavedra, J., & López-Méndez, E. (2013). Granulomatous hepatitis caused by Q fever: a differential diagnosis of fever of unknown origin. *Annals of Hepatology*, *12*(1), 138–141. Retrieved from <http://europepmc.org/abstract/MED/23293205>
- Alves, J., Almeida, F., Duro, R., Ferraz, R., Silva, S., Sobrinho-Simões, J., & Sarmiento, A. (2017). Presentation and diagnosis of acute Q fever in Portugal – A case series. *IDCases*, *7*, 34–37. doi:10.1016/j.idcr.2016.11.002
- Amitai, Z., Bromberg, M., Bernstein, M., Raveh, D., Keysary, A., David, D., ... Shohat, T. (2010). A Large Q Fever Outbreak in an Urban School in Central Israel. *Clinical Infectious Diseases*, *50*(11), 1433–1438. doi:10.1086/652442
- Anderson, A., Bijlmer, H., Fournier, P.-E., Graves, S., Hartzell, J., Kersh, G. J., ... Sexton, D. J. (2013). Diagnosis and management of Q fever--United States, 2013. *MMWR. Recommendations and Reports : Morbidity and Mortality Weekly Report. Recommendations and Reports / Centers for Disease Control*, *62*(3), 1–30. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23535757>
- Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., & Zondervan, K. T. (2010). Data quality control in genetic case-control association studies. *Nature Protocols*, *5*(9), 1564–1573. doi:10.1038/nprot.2010.116
- Angelakis, E., Mediannikov, O., Socolovschi, C., Mouffok, N., Bassene, H., Tall, A., ... Raoult, D. (2014). Coxiella burnetii-positive PCR in febrile patients in rural and urban Africa. *International Journal of Infectious Diseases*, *28*, 107–110. doi:10.1016/j.ijid.2014.05.029
- Argov, O., Weintraub, M., & Charach, G. (2008). ‘Doughnut’ granulomas from erythema nodosum in acute Q fever. *Israel Medical Association Journal*, *10*(3), 241–242.
- Asamoah, J. K. K., Jin, Z., Sun, G. Q., & Li, M. Y. (2020). A Deterministic Model for Q Fever Transmission Dynamics within Dairy Cattle Herds: Using Sensitivity Analysis and Optimal Controls. *Computational and Mathematical Methods in Medicine*, 2020. doi:10.1155/2020/6820608
- Association., W. M. (2001). World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization*, *79*(4), 373–374. Retrieved from

<https://apps.who.int/iris/handle/10665/268312>

- Baddour, L. M., Zheng, Z.-J., Labarthe, D. R., & O'Connor, S. (2007). ACCF/AHA/CDC Conference report on emerging infectious diseases and biological terrorism threats. Task force I: direct cardiovascular implications of emerging infectious diseases and biological terrorism threats. *Journal of the American College of Cardiology*, 49(12), 1380–1389. doi:10.1016/j.jacc.2007.01.019
- Beckman Coulter. (2002). *Agencourt AMPure XP PCR Purification. AMPure XP*. Retrieved from <http://autodiscover.josam.com/pdf/josam/catalog/Instructions.pdf>
- Benoit, M., Barbarat, B., Bernard, A., Olive, D., & Mege, J. L. (2008). Coxiella burnetii, the agent of Q fever, stimulates an atypical M2 activation program in human macrophages. *European Journal of Immunology*, 38(4), 1065–1070. doi:10.1002/eji.200738067
- Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *American Journal of Psychiatry*, 176(5), 376–387. doi:10.1176/appi.ajp.2018.18070881
- Bossi, P., Loock, F., Tegnell, A., & Gouvras, G. (2005). BICHAT clinical guidelines for bioterrorist agents. *Euro Surveillance : Bulletin Européen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, 9, E1-2. doi:10.2807/esm.09.12.00499-en
- Broad Institute. (2020). Somatic short variant discovery (SNVs + Indels). Retrieved 13 August 2020, from <https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels>
- Burette, M., & Bonazzi, M. (2020). From neglected to dissected: How technological advances are leading the way to the study of Coxiella burnetii pathogenesis. *Cellular Microbiology*, 22(4), 1–8. doi:10.1111/cmi.13180
- Burke, J. D., & Young, H. A. (2019). IFN- $\Gamma$ : A cytokine at the right time, is in the right place. *Seminars in Immunology*, 43(June), 101280. doi:10.1016/j.smim.2019.05.002
- Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., ... Investigators, P. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145), 661–678. doi:10.1038/nature05911
- Caitlin A. Moran, M. N. W. and I. O. (2017). 乳鼠心肌提取 HHS Public Access. *Physiology & Behavior*, 176(12), 139–148. doi:10.1016/j.physbeh.2017.03.040

- Centre for Disease Prevention and Control, E. (2019). *SURVEILLANCE REPORT Q fever Annual Epidemiological Report for 2018 Key facts*.
- Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1). doi:10.1186/s13742-015-0047-8
- Chen, C., Bouman, T. J., Beare, P. A., Mertens, K., Zhang, G. Q., Russell-Lodrigue, K. E., ... Samuel, J. E. (2009). A systematic approach to evaluate humoral and cellular immune responses to *Coxiella burnetii* immunoreactive antigens. *Clinical Microbiology and Infection*, *15*(SUPPL. 2), 156–157. doi:10.1111/j.1469-0691.2008.02206.x
- Cooper, M. A., Fehniger, T. A., Turner, S. C., Chen, K. S., Ghaheri, B. A., Carson, W. E., & Caligiuri, M. A. (2000). Human natural killer cells: A unique innate immunoregulatory role for the CD56BRIGHT SUBSET. *Blood*, *96*(11 PART II), 3146–3151.
- Cox. (2005). Rational Inferences about Departures from HWE. *American Journal of Human Genetics*, *76*(6), 967–986. doi:10.1086/430507
- D'Amato, F., Eldin, C., Georgiades, K., Edouard, S., Delerce, J., Labas, N., & Raoult, D. (2015). Loss of TSS1 in hypervirulent *Coxiella burnetii* 175, the causative agent of Q fever in French Guiana. *Comparative Immunology, Microbiology and Infectious Diseases*, *41*, 35–41. doi:10.1016/j.cimid.2015.04.003
- D'Amato, F., Rouli, L., Edouard, S., Tyczka, J., Million, M., Robert, C., ... Raoult, D. (2014). The genome of *Coxiella burnetii* Z3055, a clone linked to the Netherlands Q fever outbreaks, provides evidence for the role of drift in the emergence of epidemic clones. *Comparative Immunology, Microbiology and Infectious Diseases*, *37*(5–6), 281–288. doi:10.1016/j.cimid.2014.08.003
- Darakhshan, S., Hassanshahi, G., Mofidifar, Z., Soltani, B., & Karimabad, M. N. (2019). CXCL9/CXCL10 angiostasis CXC-chemokines in parallel with the CXCL12 as an angiogenesis CXC-chemokine are variously expressed in pre-eclamptic-women and their neonates. *Pregnancy Hypertension*, *17*(December 2018), 36–42. doi:10.1016/j.preghy.2019.05.001
- David, S. (2004). General principals of candidate gene, population association analysis in the genetic epidemiology of common diseases and complex traits. *Seven*, (December), 1–5.
- David, S. (2020). \_ Métodos para estudos populacionais de associação genótipo-fenótipo com base em genes candidatos, (7), 33–37.
- Delsing, C. E., Kullberg, B. J., & Bleeker-Rovers, C. P. (2010). Q fever in the Netherlands from 2007 to 2010. *The Netherlands Journal of Medicine*, *68*(12),

382–387.

- Dragan, A. L., & Voth, D. E. (2020). Coxiella burnetii: international pathogen of mystery. *Microbes and Infection*, 22(3), 100–110.  
doi:10.1016/j.micinf.2019.09.001
- Dupont, W., PAGE, D., Parl, F., Vnencak-Jones, C., \*Plummer, W. J., & Rados, M. (1994). The New England Journal of Medicine Downloaded from nejm.org at KUOPIO UNIVERSITY LIBRARY on August 13, 2014. For personal use only. No other uses without permission. Copyright © 1994 Massachusetts Medical Society. All rights reserved. *N Engl J Med.*, 297(15), 800–802. doi:10.1007/s00280-004-0998-4
- DUPUIS, G., PETITE, J., PÉTER, O., & VOUILLOZ, M. (1987). An Important Outbreak of Human Q Fever in a Swiss Alpine Valley. *International Journal of Epidemiology*, 16(2), 282–287. doi:10.1093/ije/16.2.282
- Edouard, S., Million, M., Lepidi, H., Rolain, J. M., Fournier, P. E., La Scola, B., ... Raoult, D. (2013). Persistence of DNA in a cured patient and positive culture in cases with low antibody levels bring into question diagnosis of Q fever endocarditis. *Journal of Clinical Microbiology*, 51(9), 3012–3017.  
doi:10.1128/JCM.00812-13
- Ensembl. (2019). Homo sapiens - Ensembl genome browser 96. Retrieved 21 October 2020, from <http://apr2019.archive.ensembl.org/index.html>
- Errasti, C., Baranda, M., Almaraz, J., de la Hoz, C., Gutierrez, E., Villate, J., & Peña, V. (1984). An outbreak of Q fever in the Basque country. *Canadian Medical Association Journal*, 131, 48–49.
- Esparragón, F., Rodríguez Pérez, J. C., & García Bello, M. A. (2009). Practical guide to genetic association studies. Considerations on their clinical usefulness. *Nefrología*, 29(6), 582–588. doi:10.3265/Nefrologia.2009.29.6.5483.en.full
- Fever, I. Q., Lefebvre, M., Grossi, O., & Agard, C. (2010). Systemic Immune Presentations of. *YSARH*, 39(5), 405–409. doi:10.1016/j.semarthrit.2008.10.004
- Frankel, D., Richet, H., Renvoisé, A., & Raoult, D. (2011). Q fever in France, 1985-2009. *Emerging Infectious Diseases*, 17(3), 350–356. doi:10.3201/eid1703.100882
- Galache, C., Blanco, S., Rodr, E., & Nez, I. (2004). CASE REPORT Q fever : a new cause of ‘ doughnut ’ granulomatous lobular panniculitis, 685–687. doi:10.1111/j.1365-2133.2004.06125.x
- Gao, Y., Yang, W., Pan, M., Scully, E., Girardi, M., Augenlicht, L. H., ... Yin, Z. (2003).  $\gamma\delta$  T cells provide an early source of interferon  $\gamma$  in tumor immunity. *Journal of Experimental Medicine*, 198(3), 433–442. doi:10.1084/jem.20030584
- Gargis, A. S., Kalman, L., Bick, D. P., Da Silva, C., Dimmock, D. P., Funke, B. H., ...

- Lubin, I. M. (2015). Good laboratory practice for clinical next-generation sequencing informatics pipelines. *Nature Biotechnology*, 33(7), 689–693. doi:10.1038/nbt.3237
- Ghigo, E., Pretat, L., Desnues, B., Capo, C., Raoult, D., & Mege, J.-L. (2009a). Intracellular Life of Coxiella burnetii in Macrophages. *Annals of the New York Academy of Sciences*, 1166, 55–66. doi:10.1111/j.1749-6632.2009.04515.x
- Ghigo, E., Pretat, L., Desnues, B., Capo, C., Raoult, D., & Mege, J. L. (2009b). Intracellular life of coxiella burnetii in macrophages: An update. *Annals of the New York Academy of Sciences*, 1166, 55–66. doi:10.1111/j.1749-6632.2009.04515.x
- Gidding, H. F., Peng, C. Q., Graves, S., Massey, P. D., Nguyen, C., Stenos, J., ... Wood, N. (2020). Q fever seroprevalence in Australia suggests one in twenty people have been exposed. *Epidemiology and Infection*, 3–7. doi:10.1017/S0950268820000084
- Guigno, D., Coupland, B., Smith, E. G., Farrell, I. D., Desselberger, U., & Caul, E. O. (1992). Primary humoral antibody response to Coxiella burnetii, the causative agent of Q fever. *Journal of Clinical Microbiology*, 30(8), 1958 LP – 1967. Retrieved from <http://jcm.asm.org/content/30/8/1958.abstract>
- Holland, S. M. (2005). Defects in the interferon-  $\gamma$  and interleukin-12 pathways, 203, 38–47.
- Honarmand, H. (2012). Q fever: An old but still a poorly understood disease. *Interdisciplinary Perspectives on Infectious Diseases*, 2012. doi:10.1155/2012/131932
- Hotta, A., Kawamura, M., To, H., Andoh, M., Yamaguchi, T., Fukushi, H., & Hirai, K. (2002). Phase variation analysis of Coxiella burnetii during serial passage in cell culture by use of monoclonal antibodies. *Infection and Immunity*, 70(8), 4747–4749. doi:10.1128/IAI.70.8.4747-4749.2002
- Howe, D., Shannon, J. G., Winfree, S., Dorward, D. W., & Heinzen, R. A. (2010). Coxiella burnetii phase I and II variants replicate with similar kinetics in degradative phagolysosome-like compartments of human macrophages. *Infection and Immunity*, 78(8), 3465–3474. doi:10.1128/IAI.00406-10
- Ibrahim, M., Schelling, E., Zinsstag, J., Hattendorf, J., Andargie, E., & Tschopp, R. (2020). Sero-prevalence of brucellosis, Q-fever and Rift Valley Fever in humans and livestock in Somali region, Ethiopia. doi:10.1101/2020.01.31.928374
- Illumina®. (2019). *MiSeq-SystemGuide*.
- Illumina. (2016a). *Nextera XT DNA Library Prep Reference Guide (15031942)*.
- Illumina. (2016b). *Optimizing Cluster Density on Illumina Sequencing Systems Table*

of Contents, 1–12.

- Jansen, A. F. M., Schoffelen, T., Bleeker-Rovers, C. P., Wever, P. C., Jaeger, M., Oosting, M., ... van de Vosse, E. (2019). Genetic variations in innate immunity genes affect response to *Coxiella burnetii* and are associated with susceptibility to chronic Q fever. *Clinical Microbiology and Infection*, 25(5), 631.e11–631.e15. doi:10.1016/j.cmi.2018.08.011
- Jansen, A. F. M., Schoffelen, T., Textoris, J., Mege, J. L., Bleeker-Rovers, C. P., Roest, H. I. J., ... van Deuren, M. (2017). Involvement of matrix metalloproteinases in chronic Q fever. *Clinical Microbiology and Infection*, 23(7), 487.e7–487.e13. doi:10.1016/j.cmi.2017.01.022
- Jaubert, J., Naze, F., Camuset, G., Larrieu, S., Pascalis, H., Guernier, V., ... Gérardin, P. (2019). Seroprevalence of *Coxiella burnetii* (Q fever) Exposure in Humans on Reunion Island. *Open Forum Infectious Diseases*, 6(7), 1–4. doi:10.1093/ofid/ofz227
- Jia, H., Guo, Y., Zhao, W., & Wang, K. (2014). Long-range PCR in next-generation sequencing: Comparison of six enzymes and evaluation on the MiSeq sequencer. *Scientific Reports*, 4, 1–8. doi:10.1038/srep05737
- Jorgensen, T. J., Ruczinski, I., Kessing, B., Smith, M. W., Shugart, Y. Y., & Alberg, A. J. (2009). Hypothesis-driven candidate gene association studies: Practical design and analytical considerations. *American Journal of Epidemiology*, 170(8), 986–993. doi:10.1093/aje/kwp242
- Kahl, G. (2015). *MiSeq™ system*. *The Dictionary of Genomics, Transcriptomics and Proteomics*. <https://doi.org/10.1002/9783527678679.dg07722>
- Kampschreur, L. M., Dekker, S., Hagens, J. C. J. P., Lestrade, P. J., Renders, N. H. M., de Jager-Leclercq, M. G. L., ... Oosterheert, J. J. (2012). Identification of risk factors for chronic Q fever, the Netherlands. *Emerging Infectious Diseases*, 18(4), 563–570. doi:10.3201/eid1804.111478
- Kampschreur, L. M., Wegdam-Blans, M. C. A., Wever, P. C., Renders, N. H. M., Delsing, C. E., Sprong, T., ... Group, D. Q. F. C. (2015). Chronic Q fever diagnosis—consensus guideline versus expert opinion. *Emerging Infectious Diseases*, 21(7), 1183–1188. doi:10.3201/eid2107.130955
- Kang, K., Park, S. H., Chen, J., Qiao, Y., Giannopoulou, E., Berg, K., ... Ivashkiv, L. B. (2017). Interferon- $\gamma$  Represses M2 Gene Expression in Human Macrophages by Disassembling Enhancers Bound by the Transcription Factor MAF. *Immunity*, 47(2), 235–250.e4. doi:10.1016/j.immuni.2017.07.017
- Kasahara, T., Hooks, J. J., Dougherty, S. F., & Oppenheim, J. J. (1983). Interleukin 2-mediated immune interferon (IFN-gamma) production by human T cells and T

- cell subsets. *Journal of Immunology (Baltimore, Md. : 1950)*, 130(4), 1784–1789.
- Khor, C. S., Mohd-Rahim, N. F., Hassan, H., Chandren, J. R., Nore, S. S., Johari, J., ... Abubakar, S. (2018). Seroprevalence of Q Fever among the Indigenous People (Orang Asli) of Peninsular Malaysia. *Vector-Borne and Zoonotic Diseases*, 18(3), 131–137. doi:10.1089/vbz.2017.2153
- Kováčová, E., Kazár, J., & Šimková, A. (1998). Clinical and Serological Analysis of a Q Fever Outbreak in Western Slovakia with Four-Year Follow-Up. *European Journal of Clinical Microbiology and Infectious Diseases*, 17(12), 867–869. doi:10.1007/s100960050209
- Kuehn, B. M. (2018). News from the centers for disease control and prevention. *JAMA - Journal of the American Medical Association*, 319(24), 2471. doi:10.1001/jama.2018.7441
- Kwon, J. M., & Goate, A. M. (2000). The candidate gene approach. *Alcohol Research and Health*, 24(3), 164–168.
- L. Sheng, M. Christopher, A. M. (2016). 乳鼠心肌提取 HHS Public Access. *Physiology & Behavior*, 176(1), 100–106. doi:10.1016/j.gde.2016.03.011
- Langan, E. A., Graetz, V., Allerheiligen, J., Zillikens, D., Rupp, J., & Terheyden, P. (2020). Immune checkpoint inhibitors and tuberculosis: an old disease in a new context. *The Lancet Oncology*, 21(1), e55–e65. doi:https://doi.org/10.1016/S1470-2045(19)30674-6
- Langmead, B., Wilks, C., Antonescu, V., & Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics*, 35(3), 421–432. doi:10.1093/bioinformatics/bty648
- Lee, J.-K., Sayers, T. J., Brooks, A. D., Back, T. C., Young, H. A., Komschlies, K. L., ... Wiltrot, R. H. (2000). IFN- $\gamma$ -Dependent Delay of In Vivo Tumor Progression by Fas Overexpression on Murine Renal Cancer Cells. *The Journal of Immunology*, 164(1), 231–239. doi:10.4049/jimmunol.164.1.231
- Lee, M., Jang, J. J., Kim, Y. S., Lee, S. O., Choi, S. H., Kim, S. H., & Yu, E. (2012). Clinicopathologic features of Q fever patients with acute hepatitis. *Korean Journal of Pathology*, 46(1), 10–14. doi:10.4132/KoreanJPathol.2012.46.1.10
- Lepidi, H., Houpikian, P., Liang, Z., & Raoult, D. (2003). Cardiac Valves in Patients with Q Fever Endocarditis: Microbiological, Molecular, and Histologic Studies. *The Journal of Infectious Diseases*, 187(7), 1097–1106. doi:10.1086/368219
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)*, 27(21), 2987–2993. doi:10.1093/bioinformatics/btr509

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. doi:10.1093/bioinformatics/btp352
- Maltezou, H. C., Constantopoulou, I., Kallergi, C., Vlahou, V., Georgakopoulos, D., Kafetzis, D. A., & Raoult, D. (2004). Q fever in children in Greece. *The American Journal of Tropical Medicine and Hygiene*, 70(5), 540–544.
- Maltezou, H. C., & Raoult, D. (2002). Q fever in children. *Lancet Infectious Diseases*, 2(11), 686–691. doi:10.1016/S1473-3099(02)00440-1
- Marrie, T. J. (2010). Q fever pneumonia. *Infectious Disease Clinics of North America*, 24(1), 27–41. doi:10.1016/j.idc.2009.10.004
- Marrie, T. J., & Raoult, D. (2002). Update on Q fever, including Q fever endocarditis. *Current Clinical Topics in Infectious Diseases*, 22, 97–124. Retrieved from <http://europepmc.org/abstract/MED/12520650>
- Mboussou, Y., Jaubert, J., Larrieu, S., Atiana, L., Naze, F., Folio, C., ... Gérardin, P. (2019). Pregnancy outcomes of Q fever: Prospective follow-up study on Reunion island. *BMC Infectious Diseases*, 19(1), 1–8. doi:10.1186/s12879-019-4619-6
- Medkour, H., Davoust, B., Angelakis, M., Thiéry, R., Raoult, D., Rousset, E., ... Eldin, C. (2020). A sporadic case of acute Q fever and identification of the animal source of the infection. *Folia Microbiologica*, 65(5), 797–800. doi:10.1007/s12223-020-00788-3
- Melenotte, Clea, Million, M., Audoly, G., Gorse, A., Dutronc, H., Roland, G., ... Raoult, D. (2016). B-cell non-Hodgkin lymphoma linked to *Coxiella burnetii*. *Blood*, 127(1), 113–121. doi:10.1182/blood-2015-04-639617
- Melenotte, Cléa, Million, M., & Raoult, D. (2020). New insights in *Coxiella burnetii* infection: diagnosis and therapeutic update. *Expert Review of Anti-Infective Therapy*, 18(1), 75–86. doi:10.1080/14787210.2020.1699055
- Meyre, D., Delplanque, J., Chèvre, J.-C., Lecoœur, C., Lobbens, S., Gallina, S., ... Froguel, P. (2009). Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nature Genetics*, 41(2), 157–159. doi:10.1038/ng.301
- Miceli, M. H., Veryser, A. K., Anderson, A. D., Hofinger, D., Lee, S. A., & Tancik, C. (2010). A case of person-to-person transmission of Q fever from an active duty serviceman to his spouse. *Vector-Borne and Zoonotic Diseases*, 10(5), 539–541. doi:10.1089/vbz.2009.0101
- Milazzo, A., Hall, R., Storm, P. A., Harris, R. J., Winslow, W., & Marmion, B. P. (2001). Sexually Transmitted Q Fever. *Clinical Infectious Diseases*, 33(3), 399–402. doi:10.1086/321878



- axis in susceptibility to infections by mycobacteria and salmonella. *Journal of Interferon and Cytokine Research*, 34(5), 307–317. doi:10.1089/jir.2013.0050
- Robinson, C. M., O’Dee, D., Hamilton, T., & Nau, G. J. (2009). Cytokines involved in interferon- $\gamma$  production by human macrophages. *Journal of Innate Immunity*, 2(1), 56–65. doi:10.1159/000247156
- Rodolakis, A. (2009). Q fever in dairy animals. *Annals of the New York Academy of Sciences*, 1166, 90–93. doi:10.1111/j.1749-6632.2009.04511.x
- Rodríguez, D., Morrison, C. J., & Overall, C. M. (2010). Matrix metalloproteinases: What do they not do? New substrates and biological roles identified by murine models and proteomics. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1803(1), 39–54. doi:10.1016/j.bbamcr.2009.09.015
- Romano, P. S., Gutierrez, M. G., Berón, W., Rabinovitch, M., & Colombo, M. I. (2007). The autophagic pathway is actively modulated by phase II Coxiella burnetii to efficiently replicate in the host cell. *Cellular Microbiology*, 9(4), 891–909. doi:10.1111/j.1462-5822.2006.00838.x
- Samanta, D., & Gilk, S. D. (2017). Measuring pH of the coxiella burnetii parasitophorous vacuole. *Current Protocols in Microbiology*, 2017, 6C.3.1-6C.3.11. doi:10.1002/cpmc.38
- Santos, A. S., Bacellar, F., & Franca, A. (2007). Febre Q: Revisão de conceitos. *Medicina Interna*, 14, 90–99.
- Santos, A. S., Tilburg, J., Botelho, A., Barahona, M., Nuncio, M. S., Nabuurs-Franssen, M., & Klaassen, C. (2012). Genotypic diversity of clinical Coxiella burnetii isolates from Portugal based on MST and MLVA typing. *International Journal of Medical Microbiology : IJMM*, 302, 253–256. doi:10.1016/j.ijmm.2012.08.003
- Schelling, E., & Traore, M. (2005). Seroprevalence of Q-fever in febrile individuals in Mali, 10(6), 612–617.
- Schimme, B. (2008). R a p i d c o m m u n i c a t i o n s Large ongoing Q fever N etherlands , 2008 outbreak in the south of. *Euro Surveill*, 13(C), 2005–2006.
- Schneider, T., Jahn, H., Steinhoff, D., Guschoreck, H., Liesenfeld, O., Mäter-Böhm, H., ... Dissmann, T. (1993). [A Q fever epidemic in Berlin. The epidemiological and clinical aspects]. *Deutsche Medizinische Wochenschrift (1946)*, 118, 689–695. doi:10.1055/s-2008-1059379
- Schoffelen, T., Textoris, J., Bleeker-Rovers, C. P., Ben Amara, A., van der Meer, J. W. M., Netea, M. G., ... van de Vosse, E. (2017). Intact interferon- $\gamma$  response against Coxiella burnetii by peripheral blood mononuclear cells in chronic Q fever. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, 23(3), 209.e9-209.e15.

doi:10.1016/j.cmi.2016.11.008

- Schoffelen, Teske, Ammerdorffer, A., Hagenaars, J. C. J. P., Bleeker-Rovers, C. P., Wegdam-Blans, M. C., Wever, P. C., ... van de Vosse, E. (2015). Genetic Variation in Pattern Recognition Receptors and Adaptor Proteins Associated With Development of Chronic Q Fever. *The Journal of Infectious Diseases*, 212(5), 818–829. doi:10.1093/infdis/jiv113
- Schoffelen, Teske, Joosten, L. A. B., Herremans, T., De Haan, A. F. J., Ammerdorffer, A., Rümke, H. C., ... Van Deuren, M. (2013). Specific interferon  $\gamma$  detection for the diagnosis of previous Q fever. *Clinical Infectious Diseases*, 56(12), 1742–1751. doi:10.1093/cid/cit129
- Seshadri, R., Paulsen, I. T., Eisen, J. A., Read, T. D., Nelson, K. E., Nelson, W. C., ... Heidelberg, J. F. (2003). Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proceedings of the National Academy of Sciences of the United States of America*, 100(9), 5455–5460. doi:10.1073/pnas.0931379100
- Seth-Smith, H. M. B., Bonfiglio, F., Cuénod, A., Reist, J., Egli, A., & Wüthrich, D. (2019). Evaluation of rapid library preparation protocols for whole genome sequencing based outbreak investigation. *Frontiers in Public Health*, 7(AUG). doi:10.3389/fpubh.2019.00241
- Shannon, J. G., Cockrell, D. C., Takahashi, K., Stahl, G. L., & Heinzen, R. A. (2009). Antibody-mediated immunity to the obligate intracellular bacterial pathogen *Coxiella burnetii* is Fc receptor- and complement-independent. *BMC Immunology*, 10, 1–8. doi:10.1186/1471-2172-10-26
- Takeda, K., Smyth, M. J., Cretney, E., Hayakawa, Y., Kayagaki, N., Yagita, H., & Okumura, K. (2002). Critical role for tumor necrosis factor-related apoptosis-inducing ligand in immune surveillance against tumor development. *Journal of Experimental Medicine*, 195(2), 161–169. doi:10.1084/jem.20011171
- Teunis, P. F. M., Schimmer, B., Notermans, D. W., Leenders, A. C. A. P., Wever, P. C., Kretzschmar, M. E. E., & Schneeberger, P. M. (2013). Time-course of antibody responses against *Coxiella burnetii* following acute Q fever. *Epidemiology and Infection*, 141(1), 62–73. doi:10.1017/S0950268812000404
- Tissot-Dupont, H., & Raoult, D. (2008). Q Fever. *Infectious Disease Clinics of North America*, 22(3), 505–514. doi:10.1016/j.idc.2008.03.002
- United States Department of Agriculture. (2013). Evaluation of Factors that Would Initiate or Propagate Epidemic Coxiellosis in the U.S. Domesticated Goat Population, (June), 1–104.
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., ... DePristo, M. A. (2013). From FastQ data to high confidence

- variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43(1110), 11.10.1-11.10.33.  
doi:10.1002/0471250953.bi1110s43
- van der Hoek, W., Morroy, G., Renders, N. H. M., Wever, P. C., Hermans, M. H. A., Leenders, A. C. A. P., & Schneeberger, P. M. (2012). Epidemic Q fever in humans in the Netherlands. *Advances in Experimental Medicine and Biology*, 984, 329–364. doi:10.1007/978-94-007-4315-1\_17
- Van Roeden, S. E., Bleeker-Rovers, C. P., De Regt, M. J. A., Kampschreur, L. M., Hoepelman, A. I. M., Wever, P. C., & Oosterheert, J. J. (2018). Treatment of chronic Q fever: Clinical efficacy and toxicity of antibiotic regimens. *Clinical Infectious Diseases*, 66(5), 719–726. doi:10.1093/cid/cix886
- Vignali, D. A. A. (2014). Engineering Optimal Network Effects via Social Media.pdf, 13(8), 722–728. doi:10.1038/ni.2366.IL-12
- Wang, Q. S., Shen, S. Q., Sun, H. W., Xing, Z. X., & Yang, H. L. (2018). Interferon-gamma induces autophagy-associated apoptosis through induction of cPLA2-dependent mitochondrial ROS generation in colorectal cancer cells. *Biochemical and Biophysical Research Communications*, 498(4), 1058–1065.  
doi:10.1016/j.bbrc.2018.03.118
- White, B., Brooks, T., & Seaton, R. A. (2013). Q fever in military and paramilitary personnel in conflict zones: Case report and review. *Travel Medicine and Infectious Disease*, 11(2), 134–137. doi:10.1016/j.tmaid.2012.11.001
- Wielders, C. C. H., Hackert, V. H., Schimmer, B., Hodemaekers, H. M., de Klerk, A., Hoebe, C. J. P. A., ... Janssen, R. (2015). Single nucleotide polymorphisms in immune response genes in acute Q fever cases with differences in self-reported symptoms. *European Journal of Clinical Microbiology and Infectious Diseases*, 34(5), 943–950. doi:10.1007/s10096-014-2310-9
- Wigginton, J. E., Cutler, D. J., & Abecasis, G. R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, 76(5), 887–893.  
doi:10.1086/429864
- Wittke-Thompson, J. K., Pluzhnikov, A., & Cox, N. J. (2005). Rational inferences about departures from Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, 76(6), 967–986. doi:10.1086/430507
- Xiong, X., Qi, Y., Jiao, J., Gong, W., Duan, C., & Wen, B. (2014). Exploratory study on Th1 epitope-induced protective immunity against *Coxiella burnetii* infection. *PLoS ONE*, 9(1), 1–8. doi:10.1371/journal.pone.0087206
- Yohe, S., & Thyagarajan, B. (2017). Review of clinical next-generation sequencing. *Archives of Pathology and Laboratory Medicine*, 141(11), 1544–1557.

doi:10.5858/arpa.2016-0501-RA

Zhang, M., Huang, L., Ding, G., Huang, H., Cao, G., Sun, X., ... Yan, Q. (2020).

Interferon gamma inhibits CXCL8-CXCR2 axis mediated tumor-associated macrophages tumor trafficking and enhances anti-PD1 efficacy in pancreatic cancer. *Journal for Immunotherapy of Cancer*, 8(1), e000308. doi:10.1136/jitc-2019-000308

## Anexos

### 1- Exemplo de Anderson

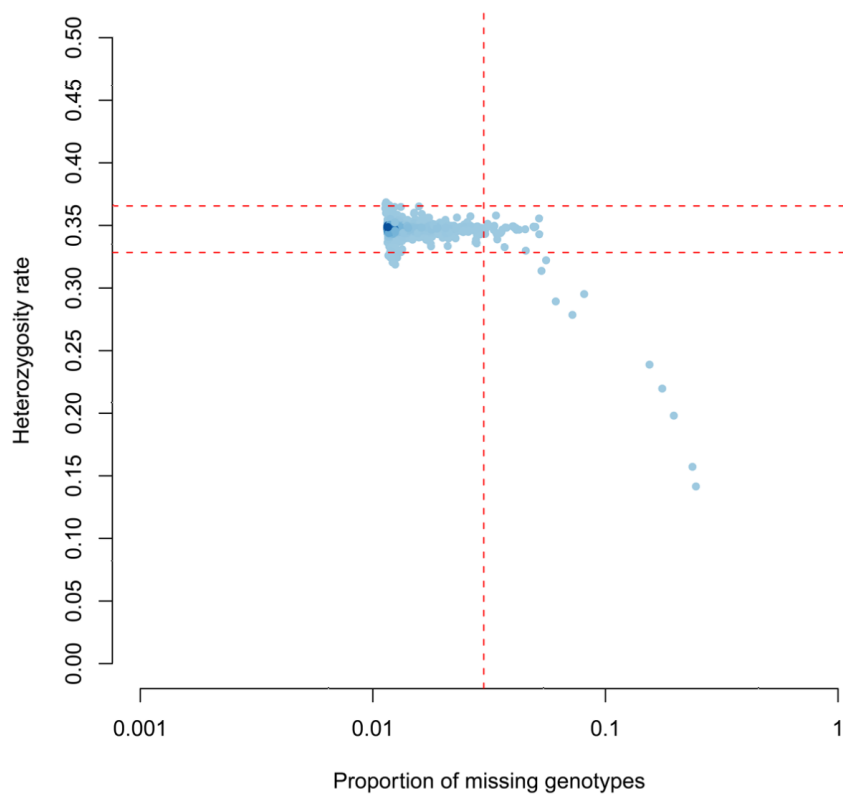


Figura 12. Exemplo de Anderson e colaboradores para estudos de associação genética. Taxa de falha do genótipo (*Genotype failure rate*) versus heterozigotia (*heterozygosity rate*) em todos os indivíduos do estudo. Sombreado indica densidade de amostra e linhas tracejadas denotam limites de controle de qualidade (Anderson et al., 2010).

## 2- Fluxo de trabalho

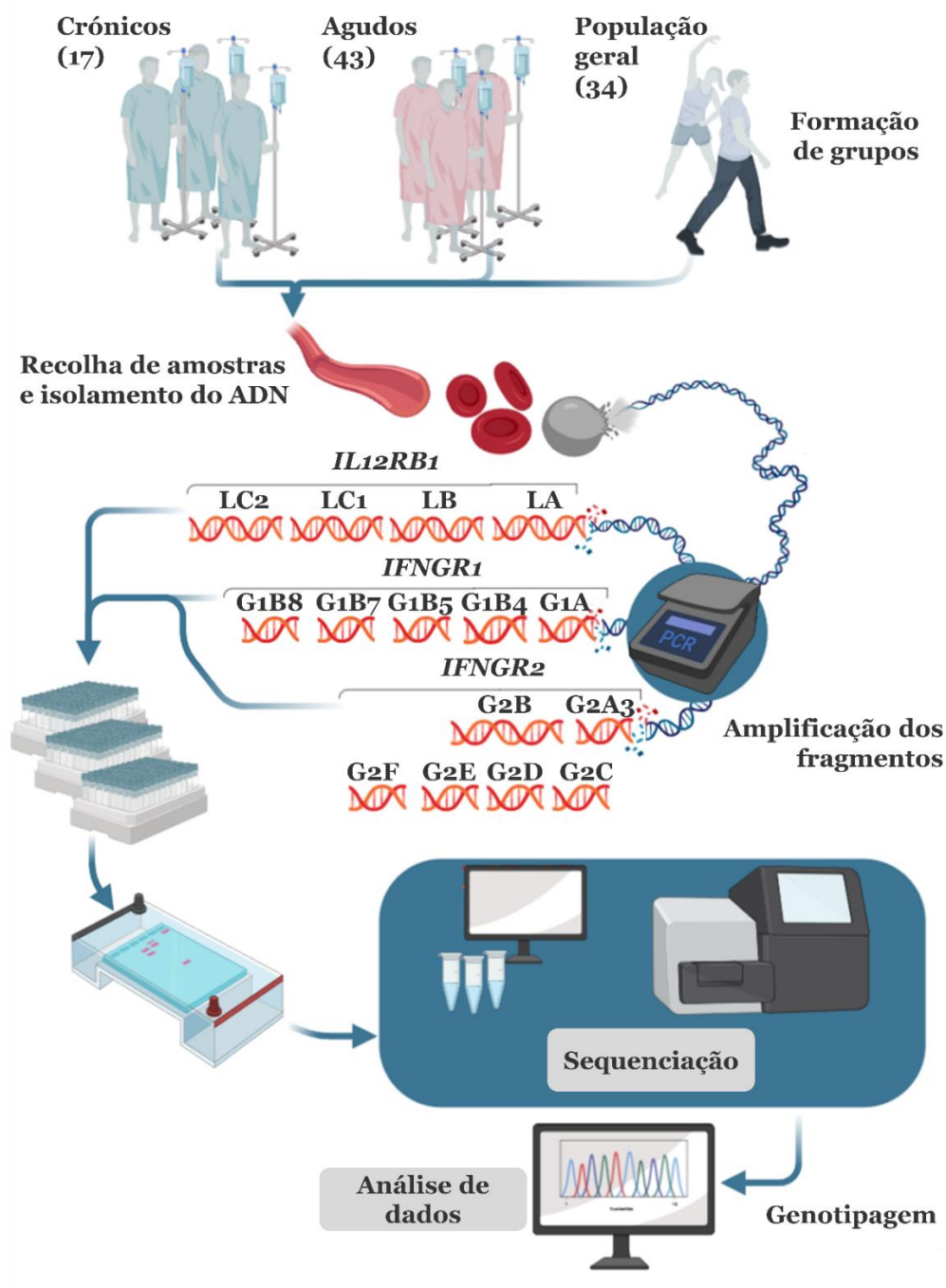


Figura 13. Protocolo científico - fluxo de trabalho geral. Formação de grupos: composição dos grupos de estudo; recolha de amostras e isolamento de ADN: extração de sangue de cada indivíduo; Amplificação dos genes: multiplicação do número de cópias de cada gene proposto por cada indivíduo; Sequenciação: sequenciar e multiplicar número de fragmentos de interesse; Genotipagem: reconhecimento de fragmentos de interesse (bioinformática); Análise de dados: análise de resultados e cruzamento de dados. (imagem criada a partir de biorender.com).

### 3- AMPure XP

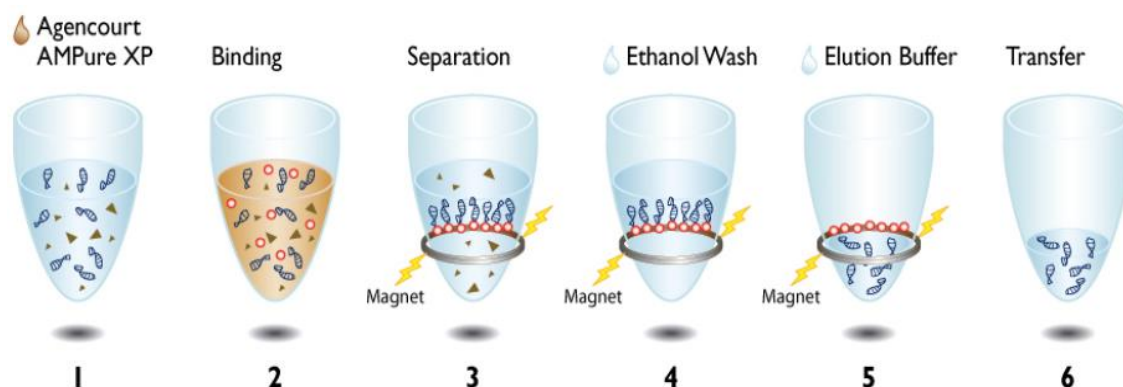


Figura 14. Protocolo *AMPure XP*. Em cada poço da placa, com produtos PCR (1), adicionam-se as esferas magnéticas (2), os ADNs ligam às esferas magnéticas sendo estas ancoradas no poço através por magnetismo (3), são feitas lavagens (4) e separadas as esferas dos ADNs (4 e 5) (imagem retirada do protocolo *Instructions For Use AMPure XP*).

### 4- Tabela de dispersão de leitura

Tabela 14. Identificação dos resultados de dispersão de leitura.

ID NGS	Amostra	Index 1 (i7)	Index 2 (i5)	% Leituras identificadas
1	C1	ACTCGCTA	CTCTCTAT	1.0084
2	C10	GGAGCTAC	CTCTCTAT	0.8456
3	C18	GCGTAGTA	CTCTCTAT	0.8239
4	A15	CGGAGCCT	CTCTCTAT	0.7085
5	A27	TACGCTGC	CTCTCTAT	0.8606
6	A37	ATGCGCAG	CTCTCTAT	1.1593
7	A45	TAGCGCTC	CTCTCTAT	1.2172
8	A56	ACTGAGCG	CTCTCTAT	0.8145
9	T5	CCTAAGAC	CTCTCTAT	1.7009
10	T13	CGATCAGT	CTCTCTAT	0.9653
11	T21	TGCAGCTA	CTCTCTAT	1.501
12	T29	TCGACGTC	CTCTCTAT	1.3895
13	C2	ACTCGCTA	TATCCTCT	1.4849
14	C11	GGAGCTAC	TATCCTCT	0.9191
15	A1	GCGTAGTA	TATCCTCT	0.8398
16	A16	CGGAGCCT	TATCCTCT	0.8044
17	A28	TACGCTGC	TATCCTCT	0.7857
18	A38	ATGCGCAG	TATCCTCT	1.0473
19	A46	TAGCGCTC	TATCCTCT	1.1563

***Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$***

20	A57	ACTGAGCG	TATCCTCT	0.9087
21	T6	CCTAAGAC	TATCCTCT	1.434
22	T14	CGATCAGT	TATCCTCT	0.8476
23	T22	TGCAGCTA	TATCCTCT	0.8847
24	T30	TCGACGTC	TATCCTCT	1.3148
25	C3	ACTCGCTA	GTAAGGAG	1.2741
26	C12	GGAGCTAC	GTAAGGAG	1.178
27	A2	GCGTAGTA	GTAAGGAG	1.2449
28	A17	CGGAGCCT	GTAAGGAG	0.8958
29	A29	TACGCTGC	GTAAGGAG	0.6863
30	A39	ATGCGCAG	GTAAGGAG	0.8351
31	A47	TAGCGCTC	GTAAGGAG	1.2144
32	A58	ACTGAGCG	GTAAGGAG	1.2282
33	T7	CCTAAGAC	GTAAGGAG	1.2185
34	T15	CGATCAGT	GTAAGGAG	0.8333
35	T23	TGCAGCTA	GTAAGGAG	0.7889
36	T31	TCGACGTC	GTAAGGAG	0.7514
37	C4	ACTCGCTA	ACTGCATA	1.283
38	C13	GGAGCTAC	ACTGCATA	1.2677
39	A4	GCGTAGTA	ACTGCATA	0.8752
40	A19	CGGAGCCT	ACTGCATA	1.1822
41	A30	TACGCTGC	ACTGCATA	1.0208
42	A40	ATGCGCAG	ACTGCATA	0.8762
43	A51	TAGCGCTC	ACTGCATA	1.2144
44	A59	ACTGAGCG	ACTGCATA	1.1381
45	T8	CCTAAGAC	ACTGCATA	1.1059
46	T16	CGATCAGT	ACTGCATA	1.2627
47	T24	TGCAGCTA	ACTGCATA	0.4593
48	T32	TCGACGTC	ACTGCATA	0.0015
49	C5	ACTCGCTA	AAGGAGTA	1.4638
50	C14	GGAGCTAC	AAGGAGTA	1.2935
51	A8	GCGTAGTA	AAGGAGTA	1.1314
52	A20	CGGAGCCT	AAGGAGTA	0.9838
53	A31	TACGCTGC	AAGGAGTA	1.0379
54	A41	ATGCGCAG	AAGGAGTA	0.8059
55	A52	TAGCGCTC	AAGGAGTA	1.1234
56	T1	ACTGAGCG	AAGGAGTA	1.0668
57	T9	CCTAAGAC	AAGGAGTA	1.1019
58	T17	CGATCAGT	AAGGAGTA	0.96
59	T25	TGCAGCTA	AAGGAGTA	0.7216
60	T34	TCGACGTC	AAGGAGTA	1.2346
61	C6	ACTCGCTA	CTAAGCCT	1.0348
62	C15	GGAGCTAC	CTAAGCCT	1.4249
63	A10	GCGTAGTA	CTAAGCCT	1.1621
64	A21	CGGAGCCT	CTAAGCCT	1.0883
65	A34	TACGCTGC	CTAAGCCT	0.794

**Aplicação de metodologia Next Generation Sequencing na pesquisa de determinantes genéticos da febre Q crônica na via de sinalização IL12-IFN $\gamma$**

66	A42	ATGCGCAG	CTAAGCCT	1.1879
67	A53	TAGCGCTC	CTAAGCCT	1.0347
68	T2	ACTGAGCG	CTAAGCCT	1.0594
69	T10	CCTAAGAC	CTAAGCCT	0.861
70	T18	CGATCAGT	CTAAGCCT	0.7995
71	T26	TGCAGCTA	CTAAGCCT	1.0501
72	T44	TCGACGTC	CTAAGCCT	1.4234
73	C7	ACTCGCTA	CGTCTAAT	1.4502
74	C16	GGAGCTAC	CGTCTAAT	1.05
75	A11	GCGTAGTA	CGTCTAAT	1.3608
76	A24	CGGAGCCT	CGTCTAAT	1.1477
77	A35	TACGCTGC	CGTCTAAT	0.8882
78	A43	ATGCGCAG	CGTCTAAT	1.4155
79	A54	TAGCGCTC	CGTCTAAT	1.1815
80	T3	ACTGAGCG	CGTCTAAT	1.0712
81	T11	CCTAAGAC	CGTCTAAT	1.3489
82	T19	CGATCAGT	CGTCTAAT	0.8508
83	T27	TGCAGCTA	CGTCTAAT	0.9919
84	C9	ACTCGCTA	TCTCTCCG	1.3349
85	C17	GGAGCTAC	TCTCTCCG	0.327
86	A14	GCGTAGTA	TCTCTCCG	0.9686
87	A26	CGGAGCCT	TCTCTCCG	0.9115
88	A36	TACGCTGC	TCTCTCCG	0.8425
89	A44	ATGCGCAG	TCTCTCCG	0.7115
90	A55	TAGCGCTC	TCTCTCCG	0.9125
91	T4	ACTGAGCG	TCTCTCCG	1.1067
92	T12	CCTAAGAC	TCTCTCCG	1.1883
93	T20	CGATCAGT	TCTCTCCG	0.6989
94	T28	TGCAGCTA	TCTCTCCG	0.2321

## 5- Imagem Docker

Tabela 15. Lista de acesso processamento via Docker.

Função	Ferramenta	Imagem Docker
Mapeamento	<i>Bowtie 2</i> v2.3.4	miguelpmachado / bowtie2: 2.3.4.3-02
Agrupamento	<i>SAM Tools</i>	miguelpmachado / samtools: 1.9
Identificar duplicados	<i>GATK</i>	broadinstitute / gatk: 4.1.5.0
Filtragem	<i>BCF Tools</i>	miguelpmachado / bcftools: 1.9-01