

## Análise genómica no Serviço Nacional de Saúde: modelo colaborativo INSA–ULSSM para implementação da análise de variantes patogénicas no exoma

*Genomic analysis in the National Health Service: an INSA–ULSSM collaborative model for implementing the analysis of pathogenic variants in the exome*

José Ferrão<sup>1</sup>, Catarina Macedo<sup>2</sup>, Lara Neto<sup>2</sup>, Joana Mendonça<sup>1</sup>, Sara Rangel<sup>1</sup>, Hugo Martiniano<sup>3</sup>, Marta Soares<sup>2</sup>, Sónia Custódio<sup>2</sup>, Maria Rosário Santos<sup>2</sup>, Ana Cristina Sousa<sup>2</sup>, Ana Berta Sousa<sup>2</sup>, Luís Vieira<sup>1</sup>

luis.vieira@insa.min-saude.pt

(1) Unidade de Tecnologia e Inovação. Departamento de Genética Humana, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal

(2) Serviço de Genética. Departamento de Pediatria, Unidade Local de Saúde de Santa Maria, Lisboa, Portugal

(3) Unidade de Investigação. Departamento de Promoção da Saúde e Prevenção de Doenças Não Transmissíveis, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal

### Resumo

A sequenciação do exoma é, hoje em dia, um teste de primeira linha para um espectro alargado de doenças genéticas. No entanto, a natureza multidisciplinar deste teste requer o envolvimento de técnicos especializados em sequenciação, bioinformáticos, especialistas de genética laboratorial clínica e médicos geneticistas, que raramente se encontram numa mesma instituição de saúde. Neste sentido, desenvolveu-se um modelo de colaboração entre o Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA) e a Unidade Local de Saúde de Santa Maria (ULSSM), com o objetivo de implementar um fluxo de trabalho completo para a análise rápida das variantes patogénicas no exoma.

A validação deste modelo envolveu a sequenciação e análise do exoma completo de 6 casos de doença genética, cujas alterações patogénicas tinham sido previamente determinadas por laboratórios externos acreditados.

Os resultados agora obtidos, sem conhecimento prévio dos anteriores, mostraram uma concordância total. Este modelo incorpora uma análise bioinformática totalmente automatizada e um algoritmo de seleção rápida das variantes priorizadas, o que aumenta a eficiência do processo de análise de dados e reduz o tempo de resposta.

A junção da capacidade de sequenciação massiva e bioinformática do INSA aos recursos especializados de interpretação e validação clínica de variantes da ULSSM, constitui um protótipo para uma integração ampla da sequenciação do exoma no Serviço Nacional de Saúde.

### Abstract

Exome sequencing is currently a first-line test for a broad spectrum of genetic diseases. However, the multidisciplinary nature of this test requires the involvement of specialized sequencing technicians, bioinformaticians, clinical laboratory genetics specialists, and medical geneticists, who are rarely found in the same healthcare institution. To this end, a collaborative model was developed between the National Institute of Health Doutor Ricardo Jorge (INSA) and the Local Health Unit of Santa Maria (ULSSM), aiming to implement a complete workflow for the rapid analysis of pathogenic variants in the exome.

Validation of this model involved the sequencing and analysis of the whole exome of six cases of genetic disease, whose pathogenic alterations had been previously determined by accredited external laboratories.

The results obtained now, without prior knowledge of the previous results, showed complete agreement. This model incorporates fully automated bioinformatic analysis and a rapid selection algorithm for prioritized variants, which increase the efficiency of the data analysis process and reduce turnaround time.

The combination of INSA's massive sequencing and bioinformatics capabilities with specialized resources for interpretation and clinical validation of variants by ULSSM constitutes a prototype for broad integration of exome sequencing into the National Health Service.

### Introdução

Durante as últimas duas décadas e meia, os métodos moleculares utilizados para o diagnóstico de patologias de base genética sofreram uma evolução muito significativa. Isto deveu-se ao desenvolvimento de tecnologias que possibilitaram transitar da análise de genes individuais para uma análise genómica global. A sequenciação do exoma é uma das principais aplicações da tecnologia de sequenciação de nova geração, e é atualmente utilizada como teste genético de primeira linha em condições clínicas muito diversas (1-3).

O exoma é formado pelo conjunto de genes que contêm informação para a produção de proteínas. Apesar da sequência completa do exoma humano constituir somente 1-2% da sequência genómica total, estima-se que contenha cerca de 85% das variantes genéticas associadas a doença. No entanto, atualmente, mais de 50% dos casos de suspeita de doença Mendeliana carecem de um diagnóstico molecular mesmo após a sequenciação do exoma. Alguns dos fatores

que contribuem para este insucesso parcial incluem limitações técnicas e metodológicas, compreensão incompleta da patogenicidade das variantes, ausência de associações genótipo-fenótipo, e a existência de interações complexas entre genes e ambiente (4).

A taxa de sucesso da sequenciação do exoma depende também do grau de otimização das várias etapas do fluxo de trabalho, nomeadamente dos procedimentos laboratoriais, das análises de dados (bioinformática) e da interpretação das variantes genéticas no contexto clínico. Além disso, o tempo de resposta é um aspeto de grande relevância no contexto assistencial, pelo que a automatização de algumas tarefas de elevada complexidade, como as relativas à análise bioinformática, é fundamental para encurtar o tempo necessário para um diagnóstico.

### \_Objetivo

Neste artigo descrevemos um modelo de trabalho colaborativo para a sequenciação, análise e interpretação de dados do exoma completo, e a sua validação através do estudo de 6 casos distintos de doença genética.

### \_Material e métodos

Neste trabalho foram incluídos 6 casos clínicos (HSM1-6) com etiologias genéticas distintas, acompanhados no Serviço de Genética da Unidade Local de Saúde de Santa Maria (ULSSM), para os quais tinham sido anteriormente determinadas as respetivas variantes causais por laboratórios externos acreditados. Para cada caso clínico foi definido um conjunto de termos da *Human Phenotype Ontology* (HPO), por um clínico sénior, especialista em Genética Médica, e independente ao estudo.

As tarefas de preparação de bibliotecas de exoma, sequenciação e bioinformática, foram realizadas na Unidade de Tecnologia e Inovação do Departamento de Genética Humana, do Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA). As bibliotecas de DNA foram preparadas utilizando um *kit* da *Twist Bioscience* e sequenciadas no sistema *NextSeq 2000*

(*Illumina*). A análise de dados foi automatizada e estruturada em análise primária (qualidade dos dados), secundária (chamada de variantes) e terciária (anotação e priorização de variantes), conforme representado na **figura 1**. A chamada de variantes foi efetuada usando o DRAGEN (*Illumina*) e a priorização foi efetuada utilizando o *Exomiser* (5), com inclusão dos termos HPO para cada caso clínico. Os ficheiros das diferentes etapas da análise bioinformática foram carregados no portal *INSA\_genetics*, de acesso restrito aos autores deste estudo.

A interpretação e validação das variantes foi efetuada pela equipa do Serviço de Genética da ULSSM. Para tal, foi utilizada uma *checklist* metodológica estruturada para filtrar a lista de variantes priorizadas pelo *Exomiser*, tendo por base a qualidade de genótipo, o *gene combined score*, os valores dos preditores de patogenicidade das variantes, e o modo de hereditariedade. Finalmente, procedeu-se à curadoria clínica para correlacionar os fenótipos previamente associados às variantes identificadas com o fenótipo clínico dos 6 doentes.

Na **figura 2** ilustram-se as várias etapas do fluxo de trabalho e alguns dos principais parâmetros laboratoriais, métricas de qualidade dos dados, e critérios de interpretação de variantes e clínicos, que foram analisados por ambas as equipas em cada etapa.

Figura 1: Representação esquemática da *pipeline* de análise bioinformática automatizada, desde a obtenção das leituras de sequenciação até à disponibilização dos resultados no portal *INSA\_genetics*.

Os programas utilizados e tipos de ficheiros principais estão indicados à direita.

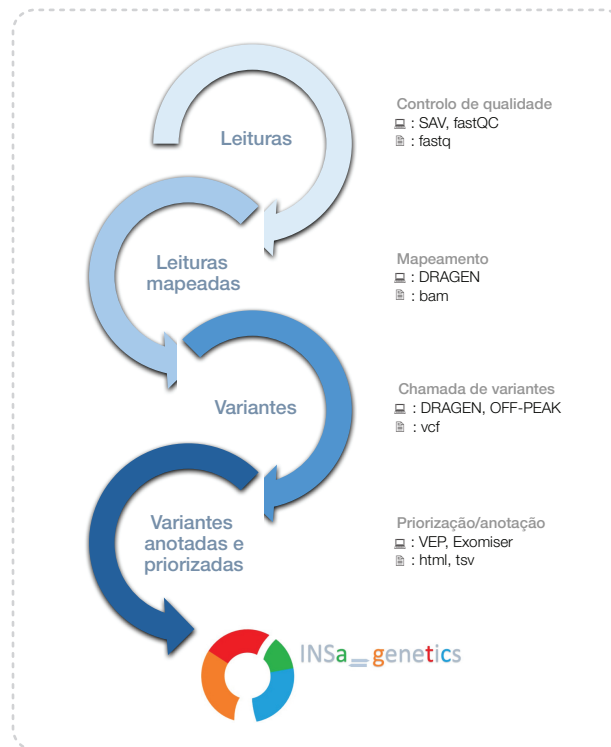


Figura 2: Diagrama representativo das etapas do modelo de trabalho colaborativo para a sequenciação e análise do exoma.

Nas respetivas etapas estão identificados alguns dos principais parâmetros laboratoriais, métricas de qualidade dos dados e critérios de interpretação de variantes e clínicos, cuja avaliação é relevante para garantir o sucesso final do diagnóstico.



## Resultados

O tempo de duração do fluxo de trabalho foi inferior a 4 dias desde o início da etapa de preparação das bibliotecas, até à disponibilização de todos os resultados da análise bioinformática no portal *INSA\_genetics*. As várias fases da análise de dados foram totalmente automatizadas, desde a geração dos ficheiros de leituras (*fastq*) até à obtenção das variantes priorizadas e anotadas, o que permitiu completar este processo num período de tempo inferior a 24 horas para todas as amostras em simultâneo. A **tabela 1** sumariza algumas das métricas de qualidade e de chamada de variantes obtidas nas 6 amostras.

O *Exomiser* priorizou uma média de 1482 variantes por caso clínico (intervalo: 1372-1821). A aplicação da *checklist* metodológica estruturada permitiu obter uma lista constituída por apenas 1 variante em 4 casos, 3 variantes num caso e 4 variantes noutra caso. Nos casos HSM1 e HSM2 foi identificada uma variante, em heterozigotia, com significado clínico incerto, associada a um gene com modo de hereditariedade autossómica dominante (AD). Nos casos HSM3 e HSM5, foi identificada uma variante, em heterozigotia, classificada como patogénica, num gene com modo de hereditariedade AD. No caso HSM6, foi identificada uma variante, em heterozigotia, com significado clínico incerto, associada a um

gene com modo de hereditariedade ligada ao X dominante. No caso HSM4, assumiu-se a possibilidade de um fenótipo composto, reportando-se uma variante, em heterozigotia, com significado clínico incerto, associada a um modo de hereditariedade AD e duas variantes em heterozigotia, no mesmo gene, com hereditariedade autossómica recessiva (AR), uma com significado clínico incerto e outra patogénica.

Estes resultados tiveram uma correspondência de 100% com o que havia sido reportado previamente pelos laboratórios externos acreditados. Em quatro casos, a variante identificada correspondeu àquela priorizada em primeiro lugar pelo *Exomiser*. Num quinto caso, a variante reportada encontrava-se priorizada na posição 2. No caso com fenótipo composto, a variante associada a hereditariedade AD foi priorizada em primeiro lugar, enquanto as variantes com modo de hereditariedade AR apenas surgiram na posição 27.

Tabela 1: Descrição das principais métricas de qualidade e de chamada de variantes nos 6 casos estudados.

	HSM1	HSM2	HSM3	HSM4	HSM5	HSM6
Número inicial de leituras (milhões)	132,9	95,7	98,7	119,6	106,3	98,5
Leituras mapeadas (%)	99,2	99,1	99,2	99,1	99,2	99,2
Enriquecimento de sequências-alvo (X)	35	33	33	33	34	34
Cobertura média nas regiões-alvo dos autossomas (X)	174	125	128	154	137	127
Número de leituras para a chamada de variantes (milhões)	93	69,6	69,7	83,5	76,4	70,5
Rácio transições/transversões (Ti/Tv)	2,5	2,6	2,6	2,6	2,5	2,6
Rácio variantes heterozigotia/homozigotia (Het/Hom)	1,6	2,1	1,6	1,7	1,7	1,6
Total de variantes do tipo SNV e indel	47842	54280	47329	47263	48129	47225
Total de variantes estruturais - DRAGEN	289	324	282	288	267	249
Total de variantes do tipo CNV - OFF-PEAK	109	213	104	111	124	95

SNV - single nucleotide variant ; CNV - copy number variant

## \_Discussão

O modelo de trabalho colaborativo para a sequenciação do exoma envolveu 3 etapas principais, que incluíram (i) a preparação de bibliotecas e sequenciação, (ii) o tratamento dos dados desde os ficheiros primários até à disponibilização dos resultados no portal *INSA\_genetics*, e (iii) a interpretação clínico-patológica das variantes priorizadas. As primeiras duas etapas foram concluídas em apenas 4 dias, incluindo um período inferior a 24 horas para a obtenção de todos os ficheiros de resultados. A rapidez desta última etapa deveu-se à automatização completa da *pipeline* bioinformática, que usou os recursos de *hardware* e *software* do sequenciador para produzir os resultados da análise secundária, os quais foram automaticamente transferidos para um servidor local que completou a análise terciária e disponibilizou todos os ficheiros de resultados num portal dedicado.

No entanto, o tempo necessário para a interpretação e validação clínica dos resultados pode ser variável em função das variantes encontradas, da complexidade fenotípica e do grau de (in)certeza do diagnóstico clínico. Neste sentido, implementámos um algoritmo de seleção de variantes causais, após a priorização, com o objetivo de reduzir significativamente o número de variantes candidatas. Com base na correlação entre os fenótipos clínicos e as variantes filtradas, foi possível identificar as variantes causais nos 6 casos clínicos e demonstrar uma concordância total com os resultados obtidos previamente por laboratórios externos acreditados.

Contudo, apesar dos resultados promissores, é importante destacar que o *Exomiser* deve ser considerado como ferramenta complementar à análise clínica especializada, sendo indispensável a integração com dados fenotípicos detalhados, a revisão de bases de dados clínicas e da literatura científica, e a consulta da evidência funcional disponível.

Os casos incluídos neste estudo apresentam um viés de seleção, uma vez que a maioria (5/6) se enquadra numa categoria fenotípica, a das síndromes complexas (6), que na literatura está associada a um elevado rendimento diagnóstico, e beneficiou de uma caracterização clínica detalhada por um

especialista sénior em Genética Médica. Esta circunstância não é representativa da potencial população prescritora, habitualmente menos familiarizada com a utilização de termos HPO e com menor experiência na fenotipagem de síndromes complexas. Assim, perspetivamos alargar este modelo de trabalho a uma maior variedade de indicações clínicas e a um maior número de casos, fenotipados por diferentes clínicos.

## \_Conclusões

Este modelo colaborativo inédito entre o INSA e o Serviço de Genética da ULSSM constitui um marco na integração da sequenciação de exomas no Serviço Nacional de Saúde, aliando capacidade de sequenciação massiva e análise bioinformática e genómica aos recursos especializados de interpretação e validação clínica de variantes.

Representa, assim, um primeiro passo decisivo para tornar os diagnósticos genéticos mais rápidos, integrativos e multidisciplinares, em linha com os princípios da medicina de precisão e com a Estratégia Nacional para a Medicina Genómica.

### Referências bibliográficas:

- (1) Artech-López A, Gómez Rodríguez MJ, Sánchez Calvin MT, et al. Towards a Change in the Diagnostic Algorithm of Autism Spectrum Disorders: Evidence Supporting Whole Exome Sequencing as a First-Tier Test. *Genes (Basel)*. 2021 Apr 12;12(4):560. <https://doi.org/10.3390/genes12040560>
- (2) Boonsimma P, Ittiwut C, Kamolvisit W, et al. Exome sequencing as first-tier genetic testing in infantile-onset pharmacoresistant epilepsy: diagnostic yield and treatment impact. *Eur J Hum Genet*. 2023 Feb;31(2):179-87. Epub 2022 Oct 5. <https://doi.org/10.1038/s41431-022-01202-x>
- (3) Xiang L, Deng A, Zhou J, et al. Diagnostic yield of trio exome sequencing as a first-tier test for identifying genetic causes of pregnancy loss. *J Hum Genet*. 2025 Jul 24. (Epub ahead of print) <https://doi.org/10.1038/s10038-025-01373-7>
- (4) Wojcik MH, Reuter CM, Marwaha S, et al. ; Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) Consortium. Beyond the exome: What's next in diagnostic testing for Mendelian conditions. *Am J Hum Genet*. 2023 Aug 3;110(8):1229-48. <https://doi.org/10.1016/j.ajhg.2023.06.009>
- (5) Smedley D, Jacobsen JO, Jäger M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc*. 2015 Dec;10(12):2004-15. Epub 2015 Nov 12. <https://doi.org/10.1038/nprot.2015.124>
- (6) Andersen PF, Ek J, Karstensen HG, et al. Diagnostic yield of whole exome sequencing in a cohort of 825 patients. *Eur J Med Genet*. 2025 Sep 2;78:105043. (Epub ahead of print) <https://doi.org/10.1016/j.ejmg.2025.105043>