

# Bancos de dados biológicos: uma investigação médica para familiarizar-se com a bioinformática

**Christina Pacheco<sup>1</sup>, Lucas Lima Vieira<sup>2</sup>, Juliana Osório Alves<sup>2</sup>, Paula Matias Soares<sup>3</sup>, Vânia Marilande Ceccatto<sup>3</sup>**

<sup>1</sup> Doutoranda em Biotecnologia, Universidade Estadual do Ceará

<sup>2</sup> Mestrando em Ciências Fisiológicas, Universidade Estadual do Ceará

<sup>3</sup> Laboratório de Bioquímica e Expressão Gênica, Curso de Licenciatura em Educação Física, Universidade Estadual do Ceará

Autor para correspondência: [christinaosvaldo@yahoo.com.br](mailto:christinaosvaldo@yahoo.com.br)

**Palavras-chave:** bioinformática, fotonovela, bancos de dados biológicos, ensino de biologia

Historicamente, a Biologia sempre necessitou de bancos de dados. O desenvolvimento da internet permitiu que uma grande quantidade de informações biológicas armazenadas, sobretudo no nível molecular, estejam disponíveis gratuitamente para a utilização, tanto por profissionais da saúde, como por educadores. Nesse contexto, apresentamos uma proposta, direcionada a universitários de cursos ligados às ciências biológicas e biomédicas, em que os participantes são motivados a desvendar um mistério médico a partir de uma fotonovela, e a utilizar as ferramentas de busca e análise bioinformática para desvendá-lo. A atividade contém um roteiro que visa orientar a navegação em alguns dos principais bancos de dados biológicos – *OMIM*, *NCBI gene*, *PubMed*, *dbSNP* e *Ensembl* – fornecendo *links* e caminhos até as informações necessárias para a finalização da tarefa. Para ilustrar os procedimentos, utilizamos uma simulação envolvendo a síndrome de Berardinelli-Seip, escolhida por ser uma doença ligada a poucos polimorfismos, mas a atividade pode ser adaptada para outras síndromes genéticas e patologias.

## BANCOS DE DADOS BIOLÓGICOS

Para os pesquisadores das ciências da vida, os bancos de dados biológicos são imprescindíveis, sejam eles coleções botânicas ou zoológicas em museus ou coleções moleculares, como os bancos genômicos. Grandes repositórios de dados gênicos, como o *National Center for Biotechnology Information* (NCBI, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) e o *Ensembl genome browser* (*Ensembl*, [www.ensembl.org](http://www.ensembl.org)), disponibilizam informações de forma organizada provenientes de toda a literatura científica. Esses dados estão disponíveis, bastando o acesso à internet e o conhecimento para usar as ferramentas de busca e análise. As informações contidas nos bancos incluem as seqüências de DNA, informações sobre os genes e transcritos, seqüências de aminoácidos e estruturas protéicas, dados sobre variabilidade genética, expressão gênica, vias metabólicas e ligações de genes/proteínas com doenças e síndromes genéticas. É nesse contexto que apresentamos uma atividade que visa facilitar um primeiro contato com os bancos de dados e suas ferramentas para os estudantes, futuros profissionais de saúde.

## DESAFIOS PARA OS PROFISSIONAIS E EDUCADORES NA ERA PÓS-GENÔMICA

O uso da genômica, atualmente, é corriqueiro na pesquisa científica, o que deve se refletir nos próximos anos nas bases curriculares dos cursos universitários das áreas biológicas e biomédicas. Para que haja o uso adequado

dos bancos de dados e para que essa experiência possa ser vivenciada em toda a sua potencialidade, é necessária uma aproximação cuidadosa, já que os bancos de dados podem ser complexos e pouco didáticos, dificultando a utilização por usuários não familiarizados.

Dentre as dificuldades existem a barreira da língua, o excesso de informações e a complexidade dos conceitos que estão envolvidos na compreensão do que significam os dados depositados e as ferramentas para analisá-los. Apesar desses desafios, a apropriação de novas tecnologias pode ampliar o leque de possibilidades pedagógicas para o educador.

Como público alvo para essa atividade, sugerimos estudantes de nível universitário das áreas de ciências biológicas e saúde. Ela pode ser inserida em diversas disciplinas, como Biologia molecular, Genética, Genética médica, Bioinformática, Patologia e outras afins. A simulação aqui apresentada tem como tema a síndrome de Berardinelli-Seip, também conhecida como lipodistrofia congênita generalizada, mas a atividade pode ser adaptada para outras síndromes genéticas, contanto que sejam fenótipos ligados a poucos genes, sobre os quais existam informações nos bancos de dados. Por meio do estudo dessa síndrome, a atividade visa familiarizar os estudantes com os bancos de dados biológicos, de modo que percebam o potencial das ferramentas de bioinformática para resolver problemas biomédicos. A atividade propõe ainda que os estudantes apliquem seus conhecimentos sobre Genética e Biologia molecular em um contexto mais próximo da prática dos cientistas.

### BERARDINELLI-SEIP

A lipodistrofia congênita generalizada, ou Berardinelli-Seip, é uma doença genética rara de herança autossômica recessiva. Sua taxa de ocorrência é de aproximadamente de 1 em 10.000.000 habitantes. Essa síndrome pode ser dividida em quatro tipos de acordo com a alteração genética envolvida, que, apesar de apresentarem similaridades, variam na severidade dos sintomas.

A síndrome está ligada ao metabolismo dos lipídeos. Dentre as diversas implicações que a causa no corpo humano, a mais comum é redução significativa ou a ausência de tecido adiposo subcutâneo e visceral.

## ROTEIRO DA ATIVIDADE

Uma fotonovela que traz uma investigação médica foi criada para instigar a curiosidade dos alunos sobre os aspectos genéticos de uma doença, neste caso a síndrome de Berardinelli-Seip. A fotonovela apresenta algumas lacunas que os alunos devem preencher, o que os levará a pesquisar sobre a síndrome. Após a leitura da história, o professor pode descrever as atividades, apresentar o roteiro para a busca dos dados necessários (Procedimento) e demonstrar como as informações podem ser encontradas nos bancos de dados.

Nesse caso, as buscas por informações genéticas iniciam-se através do nome da patologia no banco de dados *Online Mendelian Inheritance in Man* (OMIM, [www.ncbi.nlm.nih.gov/omim](http://www.ncbi.nlm.nih.gov/omim)), através do portal do NCBI. Quatro subtipos da síndrome são encontrados, cada subtipo ligado a um determinado gene. Esses genes devem ser inseridos no roteiro. A síndrome em estudo também pode ser pesquisada em artigos científicos, usando

igualmente nos campos de busca dos portais *PubMed* ([www.ncbi.nlm.nih.gov/pubmed](http://www.ncbi.nlm.nih.gov/pubmed)) e *PubMed Central* ([www.ncbi.nlm.nih.gov/pmc](http://www.ncbi.nlm.nih.gov/pmc)), os quais também se encontram no site do NCBI. Neste primeiro momento, são obtidos os nomes dos quatro genes envolvidos na síndrome. A partir daí, direcionamos as investigações sugeridas no roteiro, por vários links de bancos de dados, como o NCBI *gene* ([www.ncbi.nlm.nih.gov/gene](http://www.ncbi.nlm.nih.gov/gene)), o que promove a busca de informações sobre cada um dos quatro genes. Da página do NCBI *gene*, um link (*SNP: geneview*) leva o estudante ao *dbSNP*, onde informações sobre polimorfismos de base única (*SNPs*) nos genes poderão ser adquiridas. Os alunos também devem buscar o gene no *Ensembl* ([www.ensembl.org](http://www.ensembl.org)), anotando no roteiro outras informações. Na parte prática, os participantes podem buscar as informações, preenchendo as lacunas que existem na história. Durante essa etapa, é aconselhável o acompanhamento pelo professor ou por monitores treinados, para que os alunos esclareçam quaisquer dúvidas que possam surgir.





Nasce um bebê, mas parece haver algo errado.



Dr. Casagrande e sua equipe analisam o caso, mas é complexo...



A equipe estuda os sintomas: ausência de gordura corporal, esteatose hepática, resistência à insulina.



Dr. Casagrande: Você usou drogas pesadas na gravidez? Tem diabetes na família?



Mãe: Não doutor, não que eu saiba.



Dr. Casagrande: Pense direito, conte tudo sobre seu histórico familiar de doenças.



Mãe: Existem alguns casos de uma doença rara na família, Berardinelli-Seip.



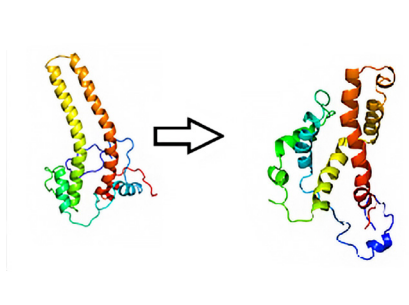
A equipe estuda a doença no NCBI, OMIM, dbSNP e Ensembl.



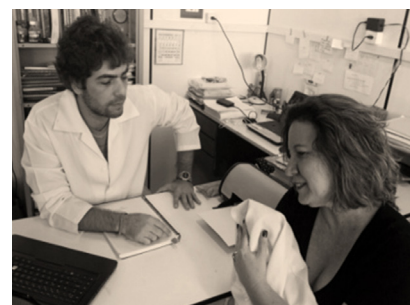
Berardinelli-Seip é uma síndrome autossômica recessiva, que pode envolver mutações em 4 genes, \_\_\_\_\_ e \_\_\_\_\_.



Os 4 genes são sequenciados no bebê. Um polimorfismo G-\_\_\_\_\_ é detectado.



O SNP está no cromossomo 7, no gene \_\_\_\_\_, o que gera uma proteína truncada e inativa.



Dr. Casagrande: O bebê precisará acompanhar a glicemia e de intervenções nutricionais.

## PROCEDIMENTO

Siga os passos deste roteiro para saber mais sobre fatores genéticos associados à Síndrome de Berardinelli-Seip. Use as informações que encontrar para preencher a tabela a seguir e as lacunas da fonovela.

- 1) Acessar o site do *OMIM* ([www.ncbi.nlm.nih.gov/omim](http://www.ncbi.nlm.nih.gov/omim)) e buscar a doença (Berardinelli-Seip ou *congenital generalized lipodystrophy*). Procurar genes ligados à desordem e a descrição da doença (sintomas). Nesse mesmo site, buscar também os quatro genes envolvidos na síndrome por meio do *OMIM* e descobrir o número de variantes alélicas em cada gene.
- 2) Acessar o *PubMed* e *PubMed Central* ([www.ncbi.nlm.nih.gov/pubmed](http://www.ncbi.nlm.nih.gov/pubmed) e [www.ncbi.nlm.nih.gov/pmc](http://www.ncbi.nlm.nih.gov/pmc)) e neles pesquisar informações sobre a síndrome de Berardinelli-Seip em artigos científicos. Procurar por detalhes que enriqueçam a descrição da síndrome e seus sintomas.
- 3) Acessar o site do *NCBI gene* (<http://www.ncbi.nlm.nih.gov/gene>) e procurar informações sobre os genes (nome oficial, sinônimos, localização) e seus produtos (tipo de produto gênico, proteína codificada). Para se encontrar o tamanho do gene, pode-se entrar na sequência do gene (no formato *FASTA*) e observar quantos quilobases (1 kb = 1000 bases) estão sendo mostrados, no início da página.
- 4) A partir da página do *NCBI gene*, seguir o link “*SNP: geneview*”, que se encontra no lado direito da página. Buscar o número de *cSNPs* (*SNPs* nas regiões dos éxons).
- 5) No site do *Ensembl* ([www.ensembl.org](http://www.ensembl.org)), procurar informações sobre os genes (nome oficial, sinônimos, localização, número de transcritos alternativos).

	Gene 1	Gene 2	Gene 3	Gene 4
Símbolo oficial do gene				
Outros nomes (Aliases) <i>NCBI / Ensembl</i>				
Localização (ex. 1q21.2) <i>NCBI / Ensembl</i>				
Número de SNPs nos éxons ( <i>cSNPs</i> ) <i>NCBI gene/dbSNP</i>				
Número de variantes alélicas ligadas a Berardinelli-Seip <i>OMIM</i>				
Tamanho (kb) <i>NCBI gene - Seq. FASTA</i>				
Fita (sense ou antisense) <i>NCBI / Ensembl</i>				
Tipo de produto gênico <i>NCBI / Ensembl</i>				
Nome da proteína codificada <i>NCBI gene</i>				
Número de transcritos alternativos <i>Ensembl</i>				

## VENCENDO OS OBSTÁCULOS

Para o desenvolvimento da atividade, pode ser necessário um estudo prévio dos principais conceitos genéticos envolvidos, além dos esclarecimentos relativos a algumas definições dos bancos de dados e de suas ferramentas. Alguns pontos são levantados

na tabela 1, mas outros poderão ser abordados. Os sites podem ser visitados pelo professor, para que ele próprio se familiarize previamente. Esta é uma oportunidade de revisão dos conceitos envolvidos, que pode ser aproveitada pelo professor que utilizar a atividade.

Banco de dados	Conhecimentos necessários <i>a priori</i>
OMIM	Padrões de herança genética Mendeliana, <i>SNPs</i> , mutações, <i>InDels</i> , expressão gênica, transcrição e tradução.
NCBI gene	Genes, nomenclatura gênica, éxons e íntrons, cromossomos, bandeamento cromossômico, fitas senso e anti-senso.
PubMed e PMC	Entendimento e leitura na língua inglesa.
Ensembl	Genes, éxons e íntrons, cromossomos, <i>splicing</i> , <i>splicing</i> alternativo, transcritos alternativos.
dbSNP	Variações e polimorfismos genéticos, <i>SNPs</i> , éxons e íntrons.

**Tabela 1.**

Conhecimentos prévios de genética necessários para um melhor aproveitamento dos recursos dos bancos de dados para essa atividade.

Entre os principais desafios apontados está o idioma, já que todos os grandes bancos de dados biológicos estão na língua inglesa, provavelmente, um dos principais obstáculos na aplicação desta atividade com os estudantes brasileiros. Com o direcionamento das buscas, o roteiro proposto facilita aos participantes encontrar algumas informações importantes, minimizando assim, a barreira linguística.

Apesar dos portais serem desenhados visando a facilidade do uso, é fácil perder-se no excesso de links e informações. O roteiro de buscas que propomos traz informações detalhadas sobre onde as informações requeridas se localizam, e pode facilitar no direcionamento da busca de informações, especialmente por principiantes na área da bioinformática. A fotonovela contextualiza o conteúdo e permite ao usuário assumir o papel de um pesquisador, desvendando uma ocorrência médica. Os alunos interessados devem, também, ser encorajados a expandir suas buscas, e seguirem sua curiosidade, permitindo que, por vezes, abram novas abas e verifiquem novas informações disponíveis.

Depois podem retornar ao roteiro e continuar as buscas dirigidas.

Devemos lembrar que os portais estão em contínua atualização, portanto, *links* já conhecidos e sua localização na página, são eventualmente modificados, bem como os resultados da busca. Sugerimos ao professor, verificar seu roteiro previamente e até mesmo adaptá-lo com novas propostas.

## AGRADECIMENTOS

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, à Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico - FUNCAP e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES. Agradecemos aos voluntários pelo apoio com as fotos. Agradecemos à Rede Nordeste de Biotecnologia - RENORBIO, ao Programa de Pós-Graduação em Ciências Fisiológicas – PPGCF/UECE, ao Instituto Superior de Ciências Biomédicas – ISCB/UECE e à Secretaria de Apoio às Tecnologias Educacionais – SATE/UECE.

