



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Criação de repositório de dados contendo RNAs não-codificadores envolvidos em câncer colorretal

Jaqueline Gutierri Coelho

Monografia apresentada como requisito parcial
para conclusão do Bacharelado em Ciência da Computação

Orientadora

Prof^ª. Dra. Maria Emilia Machado Telles Walter

Brasília
2021

Dedicatória

Dedico esta pesquisa a minha mãe, Maria Perciliana Gutierri Coelho, minha maior inspiração e por ser a melhor pessoa do mundo. Obrigada de toda a minha alma por ser o maior exemplo do sentido do amor na minha vida!

Agradecimentos

Agradeço à minha orientadora Maria Emília por toda dedicação e paciência ao me compartilhar a sua sabedoria e me apresentar a área de pesquisa mais fascinante da minha graduação. E também agradeço ao professor Li Weigang, que me incentivou a ficar no curso quando eu era caloura e me mostrou que computação é um mundo maravilhoso.

Resumo

Atualmente, existe um grande volume de informações sobre câncer colorretal (CRC) na literatura, porém, dispersas em diversas páginas *web* e artigos publicados em diferentes veículos. Além disso, os RNAs não-codificadores (ncRNAs) envolvidos em CRC ainda vêm sendo identificados, sendo descritos em periódicos e também se encontram disponíveis em bancos de dados, de acesso público. Tanto quanto temos conhecimento, essas informações não estão disponibilizadas em um único repositório de dados, que possa ser utilizado publicamente. A coleta dessas informações e o acesso a elas de maneira centralizada e confiável permite contribuir para as pesquisas em CRC, em particular para o entendimento dos pesquisadores sobre mecanismos que podem auxiliar a conhecer mecanismos celulares que contribuem para o surgimento de tumores, a progressão do câncer, ou mesmo, facilitar seu prognóstico. Neste contexto, este projeto visa coletar informações disponíveis de ncRNAs envolvidos em câncer colorretal e disponibilizá-los, de forma organizada, para consulta pública. Para isso, propusemos e implementamos um repositório de dados com informações sobre cinco subtipos de CRC e três tipos específicos de ncRNAs. Os tipos de câncer incluídos são câncer colorretal, câncer de colon, adenocarcinoma, neoplasmas e metástase colorretal hepática. Os três tipos de ncRNAs envolvidos em CRC são ncRNAs longos (em inglês, *long ncRNAs - lncRNAs*), ncRNAs longos circulares (em inglês, *circ ncRNAs*) e microRNAs (em inglês, *miRNAs*), junto com algumas características genômicas. Construímos ainda uma página *web*, através da qual podemos acessar o repositório (<https://www.percidatabase.com>). Perguntas como "Quais são os lncRNAs que estão envolvidos no adenocarcinoma?", "Quais são os lncRNAs e os miRNAs que estão envolvidos no câncer de cólon?" ou "Dado um certo circRNA, quais são os subtipos de câncer que têm esse circRNA envolvido?" podem ser respondidas com nosso repositório.

Palavras-chave: Câncer colorretal, RNAs não-codificadores, banco de dados, repositório de dados

Abstract

Currently, there is a large amount of information on colorectal cancer (CRC) in the literature, but it is scattered in several web pages and articles published in different vehicles. In addition, non-coding RNAs (ncRNAs) involved in CRC are still being identified, described in journals, and are also available in publicly accessible databases. To the best of our knowledge, this information is not available in a single, publicly available data repository. The collection of this information and the access to it in a centralized and reliable way allows contributing to CRC research, in particular to the researchers' understanding of mechanisms that may help to know cellular mechanisms that contribute to the appearance of tumors, cancer progression, or even, facilitate its prognosis. In this context, this project aims to collect available information of ncRNAs involved in colorectal cancer and make them available, in an organized way, for public consultation. To this end, we proposed and implemented a data repository with information on five CRC subtypes and three specific types of ncRNAs. The cancer types included are colorectal cancer, colon cancer, adenocarcinoma, neoplasms, and hepatic colorectal metastasis. The three types of ncRNAs involved in CRC are long ncRNAs (lncRNAs), long circular ncRNAs (circ ncRNAs), and microRNAs (miRNAs), along with some genomic features. We have also built a web page, through which we can access the repository (<https://www.percidatabase.com>). Questions like "Which lncRNAs are involved in adenocarcinoma?", "Which lncRNAs and miRNAs are involved in colon cancer?" or "Given a certain circRNA, which cancer subtypes have that circRNA involved?" can be answered with our repository.

Keywords: Colorectal cancer, Non-coding RNAs, databases, data repository

Sumário

1	Introdução	1
1.1	Motivação	3
1.2	Problema	3
1.3	Objetivos	3
1.4	Descrição dos capítulos	4
2	Noções básicas de RNAs não-codificadores	5
2.1	Ácidos nucleicos	5
2.1.1	DNA	6
2.1.2	RNA	8
2.1.3	DNA vs RNA	11
2.2	Proteínas	12
2.3	Dogma Central da Biologia Molecular	15
2.4	RNAs não-codificadores	17
2.4.1	NcRNAs pequenos	18
2.4.2	NcRNAs longos	19
3	Noções básicas de câncer colorretal	22
3.1	Noções gerais do câncer	22
3.2	Câncer Colorretal	26
3.2.1	Tipos de câncer colorretal	27
3.3	RNAs não-codificadores em câncer colorretal	29
4	Noções básicas de Banco de Dados	31
4.1	Informações, dados e bancos de dados	31
4.2	Conceitos básicos de Banco de Dados	33
4.3	Modelo Conceitual	37
5	Descrição do projeto	42
5.1	Dados	42

5.2	Modelo conceitual Perci	45
5.2.1	Modelo ER Perci	45
5.2.2	Modelo relacional Perci	46
5.3	Implementação do banco de dados Perci	47
5.4	Descrição da página <i>web</i>	49
5.5	Ligação entre a página <i>web</i> e o banco de dados Perci	51
6	Resultados e discussão	53
6.1	Informações da página <i>web</i>	53
6.1.1	Tipos de ncRNAs	54
6.1.2	Tipos de câncer	54
6.1.3	Pesquisar	55
6.1.4	Download e contato	55
6.2	Experimentos	56
7	Conclusões	61
7.1	Contribuições	61
7.2	Limitações	62
7.3	Trabalhos futuros	62
	Referências	63

Lista de Figuras

2.1	Composição do DNA e do RNA [1]	6
2.2	Uma porção do DNA, composto por uma sequência de nucleotídeos [2] . .	7
2.3	(a) Açúcar presente no DNA, a desoxirribose. (b) Açúcar presente no RNA, RNA, a ribose [3]	7
2.4	Estrutura molecular de uma fita de DNA [4]	8
2.5	Exemplos de RNAs: mRNA e dois ncRNAs (tRNA e rRNA) [5]	9
2.6	RNA mensageiro [6]	9
2.7	RNA ribossomal [7]	10
2.8	RNA transportador [8]	11
2.9	Tabela de códons [9]	12
2.10	Composição química de alguns aminoácidos [10]	13
2.11	Uma proteína formada por 2 aminoácidos, junto com os demais componen- tes [11]	13
2.12	As quatro estruturas da proteína. (a) Estrutura primária. (b) Estrutura secundária. (c) Estrutura terciária. (d) Estrutura quaternária [4]	14
2.13	Dogma Central da Biologia Molecular, que mostra os processos de trans- formação dos ácidos nucleicos em proteínas, ou seja, é o processo da síntese de proteínas [12, 13]	15
2.14	De forma genérica, as duas grandes classes de ncRNAs são os pequenos e os longos [14].	17
3.1	Os três estágios da formação do câncer [15]	23
3.2	Incidência estimada conforme a localização primária do tumor e sexo [15]. .	25
3.3	Estatísticas retiradas do site do INCA, para CRC, no Brasil [15].	26
4.1	Visão de consulta a uma estação GPS [16]	32
4.2	Banco de dados genético, relacionado a identificação criminal [17].	32
4.3	Sistema gerenciador de banco de dados [18].	34
4.4	Exemplos de Entidades [19].	37
4.5	Exemplos de Relacionamento [19].	38

4.6	Exemplos de grau dos relacionamentos [19].	39
4.7	Exemplo de cardinalidades [19].	40
4.8	Exemplo de atributos [19].	41
5.1	Número de dados coletados por tipo de CRC.	43
5.2	Existem diferentes denominações para um mesmo subtipo de CRC. Assim, os 10 subtipos das fontes primárias foram transformados em 5 subtipos para o banco de dados Perci.	44
5.3	Logo do banco de dados Perci.	44
5.4	Modelo ER de lncRNAs e circRNAs para o repositório de dados Perci formalizado pela plataforma BRModelo [20].	45
5.5	Modelo ER de miRNAs para o repositório de dados Perci formalizado pela plataforma BRModelo [20].	46
5.6	Modelo relacional do lncRNAs e circRNAs do projeto Perci.	46
5.7	Modelo relacional do miRNAs do projeto Perci.	47
5.8	O código do banco de dados criado.	47
5.9	Principais buscas do banco de dados Perci	48
5.10	Visualização de dados do banco Perci, quando o usuário acessa diretamente por meio de terminais de comando.	49
5.11	Página inicial do <i>site</i>	50
5.12	Ligação da plataforma Google Cloud com o banco de dados Perci.	51
5.13	Plataforma <i>Google Cloud</i> conectada à plataforma <i>Public Tableau</i>	51
5.14	Tabelas disponibilizadas de forma <i>online</i> na plataforma <i>Public Tableau</i>	52
6.1	Informações da página inicial do banco de dados Perci.	53
6.2	Tipos de ncRNAs que compõe o bando de dados Perci.	54
6.3	Tipos de CRC com informações armazenadas no banco de dados Perci.	54
6.4	Pesquisa Inicial.	55
6.5	Download	55
6.6	Pesquisa por tipo de ncRNAs.	56
6.7	Comportamento da busca por tipo de ncRNAs no banco de dados Perci.	56
6.8	Resultado da busca final em tabela na página <i>web</i>	57
6.9	Pesquisa por tipo de câncer.	57
6.10	Comportamento da busca por tipo de câncer no banco de dados Perci.	58
6.11	Resultado da busca final em tabela na página <i>web</i>	58
6.12	Pesquisa por tipo de gene.	59
6.13	Comportamento da busca por tipo de gene no banco de dados Perci.	59

6.14 Resultado da busca final em tabela na página *web* 60

Lista de Tabelas

2.1	Diferenças entre DNA e RNA	11
4.1	Principais Características	35
5.1	Descrição dos dados obtidos nos repositórios.	43
5.2	Número de dados por tipos de ncRNA.	43

Capítulo 1

Introdução

Na natureza, podemos encontrar dois tipos de ácidos nucleicos – DNA (do inglês, (*Deoxy-riboNucleic Acid*, ou ácido desoxirribonucleico) e o RNA (do inglês, (*Ribonucleic Acid*, ou ácido ribonucleico).

O DNA armazena informação para gerar diversos aminoácidos e moléculas de RNA. O DNA é característico da espécie, genes, cromossomos, as características hereditárias que passam de ancestrais para descendentes garantem a continuidade da sobrevivência das espécies. Mutações no DNA podem gerar novas espécies ao longo da evolução. Por esses motivos, o DNA, às vezes, é chamado de "o plano da vida".

O RNA é uma molécula que participa de vários processos celulares. Entre os RNAs, temos aqueles que são expressos em proteínas e outros que não geram proteínas, mas estão envolvidos em várias funções importantes na célula. Esse último grupo é conhecido como RNAs não-codificadores (do inglês, *non-coding RNAs* - ncRNAs). É conhecido que os ncRNAs realizam papéis celulares importantes, por exemplo, estão envolvidos em catalisação de reações químicas, expressão gênica e regulação de cromatina [21]. Menos de 2% do genoma humano é transcrito em proteínas [22], sendo que os demais RNAs não apresentam funções relacionadas à função de codificação de proteínas [23].

Proteínas são macromoléculas compostas por uma cadeia de moléculas mais simples, os aminoácidos [4]. As proteínas realizam diversos papéis vitais no organismo, como aceleração de reações químicas, transporte de nutrientes e construção de estruturas complexas. Diversos estudos visam descobrir os papéis de proteínas específicas no câncer. Por exemplo, a Her-2 é uma proteína relacionada ao aumento de proliferação e malignidade na célula, ao mesmo tempo inibe a diferenciação e apoptose das células [24]. As proteínas são formadas, a partir das informações do DNA, por moléculas específicas de RNA, tanto codificadoras de proteínas quanto ncRNAs.

Na literatura [21], os ncRNAs estão classificados como: ncRNAs pequenos, que têm características conhecidas e tamanho pequeno (20 a 300 nucleotídeos), e ncRNAs longos

(do inglês, *long ncRNAs* - *lncRNAs*), com tamanhos acima de 200 nucleotídeos e com capacidade muito baixa de sintetizar proteínas, sendo esses os transcritos menos conhecidos [25, 26].

Em anos passados, os ncRNAs eram considerados como lixo da evolução ou ruído transcricional [23], no entanto centenas de pesquisas sobre ncRNAs permitiram identificar seus papéis em diversos organismos, realizando diferentes papéis biológicos, como regulação de cromatina [27], regulação de ciclo de desenvolvimento em maturação de medula óssea [28] e supressão de tumor [29].

Por outro lado, os miRNAs constituem uma classe de ncRNAs pequenos, que são responsáveis por regulação no processo de tradução, mas também são conhecidos pois regulam a expressão de genes que estão envolvidos em mecanismos de desenvolvimento, metabolismo, proliferação celular, diferenciação e apoptose [30], que podem participar no início e na progressão de câncer.

Entre os papéis conhecidos de lncRNAs, um importante é o de regulação de doenças, especialmente em desenvolvimento de câncer [31, 32].

Com relação aos lncRNAs, [32, 33, 34] mostraram, respectivamente, a correlação de lncRNAs em células de câncer renal, sua participação em processos relacionados à epigenética do câncer e seu potencial como indicadores de tumores no diagnóstico e prognóstico de pacientes com câncer.

É importante ressaltar que essas informações estão dispersas na literatura, e a construção de repositórios com informações confiáveis, para consultas da comunidade científica, são essenciais. Existem bancos de dados com informações de boa qualidade, como o TCGA (acessado em 2021). O TCGA provê diversos tipos de informação associadas com câncer em humanos, como informação genômica, prognose e impacto na saúde pública em geral.

Neste trabalho, temos como foco o câncer colorretal. O câncer é uma doença bastante complexa e uma das maiores causas de óbitos no mundo [35]. O surgimento da doença pode estar relacionado a diversos fatores, incluindo mutações genéticas e epigenéticas [36, 37]. Em particular, o câncer colorretal (do inglês, colorectal cancer - CRC) é um dos mais comuns, e um dos tipos mais letais no mundo [38]. Embora existam avanços no diagnóstico, prevenção e tratamento, dado o ainda crescente número de pessoas com diagnóstico de CRC, anualmente, é muito importante entender os mecanismos de desenvolvimento e progressão do CRC [39, 40]. Existem ainda pesquisas contendo informações sobre interações entre proteínas, miRNAs e lncRNAs [41, 42, 43, 44].

Neste contexto, este projeto foca na construção de um repositório de dados sobre informações genômicas de CRC.

1.1 Motivação

Atualmente, existe um grande volume de informações sobre CRC na literatura, porém, não organizadas. A disponibilização de informações, de maneira centralizada e confiável vai permitir contribuir para as pesquisas em CRC, em particular para o entendimento dos pesquisadores sobre mecanismos que podem auxiliar a prevenir o surgimento de tumores ou o desenvolvimento do câncer, ou mesmo, facilitar sua identificação [45, 44, 43, 46].

1.2 Problema

Os ncRNAs envolvidos em CRC estão dispersos na literatura e também se encontram disponíveis em bancos de dados, de acesso público. Tanto quanto temos conhecimento, essas informações não estão coletadas em um único repositório de dados, que possa ser utilizado publicamente.

1.3 Objetivos

De forma geral, este projeto visa coletar as informações disponíveis de ncRNAs envolvidos em câncer colorretal e disponibilizá-los, de forma organizada, para consulta pública. Assim, o objetivo geral deste projeto é propor e implementar um repositório de dados que contenha informações sobre os tipos de ncRNAs envolvidos em CRC, e suas características genômicas, de modo a contribuir para as pesquisas dessa doença.

Os objetivos específicos deste plano de trabalho são:

- Realizar revisão bibliográfica para identificar ncRNAs envolvidos em CRC, tanto pequenos (especialmente miRNAs) quanto longos (os lncRNAs);
- Coletar informações sobre ncRNAs associados a CRC, disponíveis em bancos de dados públicos;
- Propor um repositório confiável de informações contendo: ncRNAs (sobretudo miRNAs e lncRNAs) envolvidos em CRC; e
- Criar uma página web para que esses dados possam ser acessados publicamente.

1.4 Descrição dos capítulos

No Capítulo 2, apresentaremos os conceitos básicos da Biologia Molecular necessários ao entendimento deste trabalho. Em particular, falaremos sobre ácidos nucleicos e proteínas, utilizando esses conceitos para apresentar o Dogma Central da Biologia Molecular. Em seguida, iremos falar sobre RNAs não-codificadores, que serão utilizados neste projeto.

No Capítulo 3, apresentamos os conceitos básicos do câncer colorretal (em inglês, *colorectal cancer* - CRC), necessários ao entendimento deste trabalho. Após falarmos sobre alguns conceitos básicos de câncer, especificamente alguns detalhes de CRC, vamos discutir a atuação dos RNAs não-codificadores em CRC.

No Capítulo 4, apresentaremos noções básicas de Banco de Dados, utilizados neste trabalho. Falaremos sobre algumas características fundamentais de um banco de dados, descreveremos o modelo Entidade Relacionamento e como ele é capaz de iniciar a formalização dos dados em um banco e depois caracterizaremos modelo relacional. Todos esses conceitos foram utilizados no projeto.

No Capítulo 5, apresentamos os métodos utilizados neste trabalho. Inicialmente, faremos uma breve descrição das características gerais da construção do banco de dados Perci. Em seguida, detalharemos o modelo conceitual do projeto, através da apresentação do modelo ER e do Modelo Relacional. Depois, descreveremos a página web e como foi feita a conexão da página *web* ao banco de dados Perci.

No Capítulo 6, apresentaremos os resultados obtidos a partir da implementação do banco de dados Perci e da página *web*, além de mostrarmos exemplos de perguntas relacionadas à CRC e ncRNAs, que podem ser respondidas com as informações armazenadas no banco de dados.

Finalmente, no Capítulo 7, apresentaremos as contribuições deste projeto e suas principais limitações, além de sugerir trabalhos futuros.

Capítulo 2

Noções básicas de RNAs não-codificadores

Neste capítulo, apresentamos os conceitos básicos da Biologia Molecular necessários ao entendimento deste trabalho. Na Seção 2.1, abordaremos sobre os ácidos nucleicos sobre suas diferenças e especificando cada um em Subseções separadas. Na Seção 2.2, demonstraremos o funcionamento das proteínas. Na Seção 2.3 será conceituado o Dogma Central da Biologia Molecular e por fim iremos abordar na Seção 2.4 os RNAs não codificadores, que é o foco desta monografia.

2.1 Ácidos nucleicos

Os organismos são estruturas em constante evolução. Em termos geológicos, após cerca de 3,5 bilhões de anos da formação da Terra, originou-se a vida em organismos simples. Através da hereditariedade (condição genética em que um ser vivo recebe e transmite informações genéticas através da reprodução) a evolução progrediu e diversificou de tal modo que foram-se criados organismos complexos. Em cada organismo existem estruturas simples como células ou complexas como órgãos. Com o avanço científico, a quisamos uma área denominada Biologia Molecular, junção dos ramos da genética, bioquímica e biologia celular. Do ponto de vista molecular, a ciência busca compreender as estruturas e funções das moléculas chamados ácidos nucleicos e proteínas.

Monômeros são capazes de criar macromoléculas chamadas de polímeros, que, por sua vez, se construídos por nucleotídeos, podem ser chamados de ácidos nucleicos. Existem dois tipos de ácidos nucleicos em todos os organismos vivos, sendo eles: ácido desoxirribonucleico (DNA) e ácido ribonucleico (RNA). Informações genéticas são passadas de formas hereditária através de um codificador chamado DNA e RNA, este último também participando da construção de proteínas e outros processos celulares diversos. Na

Figura 2.1 é ilustrado a composição molecular de ambos DNA e RNA, bem como as bases nitrogenadas que compõem cada uma dessas moléculas.

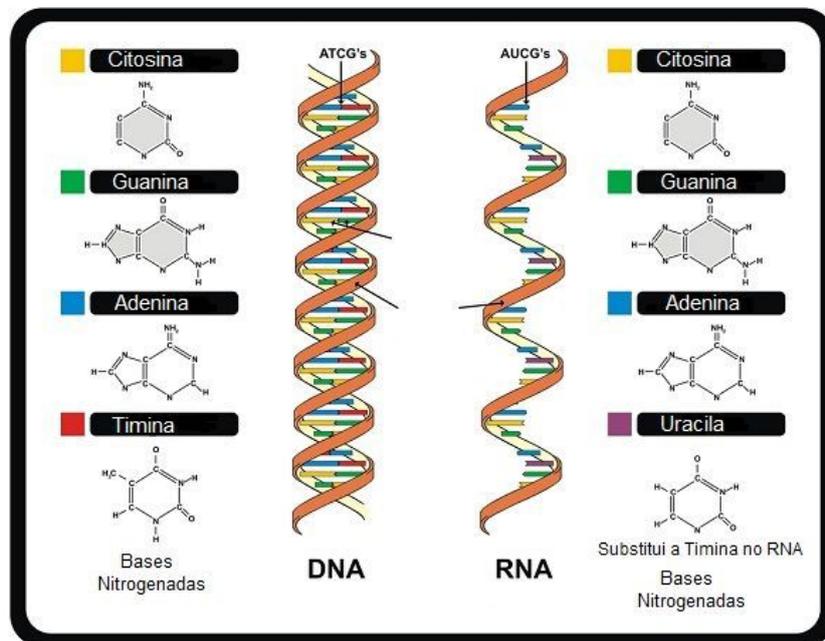


Figura 2.1: Composição do DNA e do RNA [1]

Já as proteínas são formados pela ligação de dois ou mais polipetídeos (cadeias de aminoácidos que contêm dez ou mais desses peptídeos) durante os mecanismos de transcrição. As proteínas são produzidas a partir da informação genômica contida no DNA, por meio de vários tipos de RNAs, e atuam em diversas funções vitais no organismo.

2.1.1 DNA

Nesta seção, iremos falar sobre o ácido desoxirribonucleico, mais conhecido como DNA. É possível encontrar o DNA em todas as formas de vida do nosso planeta, com exceção de alguns vírus. Nos organismos que têm núcleo celular (eucariotos), o DNA é encontrado nos cromossomos e nas mitocôndrias [47]. Nos organismos que não têm núcleo celular (procariotos), o DNA está disperso na célula.

As características hereditárias que passam de ancestrais para descendentes garantem a continuidade da sobrevivência das espécies. Mutações no DNA podem gerar novas espécies ao longo da evolução. Por esses motivos, o DNA, às vezes, é chamado de "o plano da vida".

O DNA é composto por uma sequência de unidades básicas formadas por uma molécula de açúcar, chamada 2'-desoxirribose, ligada a um resíduo de fosfato e a uma base.

Essa unidade básica é denominada de nucleotídeo (veja Figura 2.2).

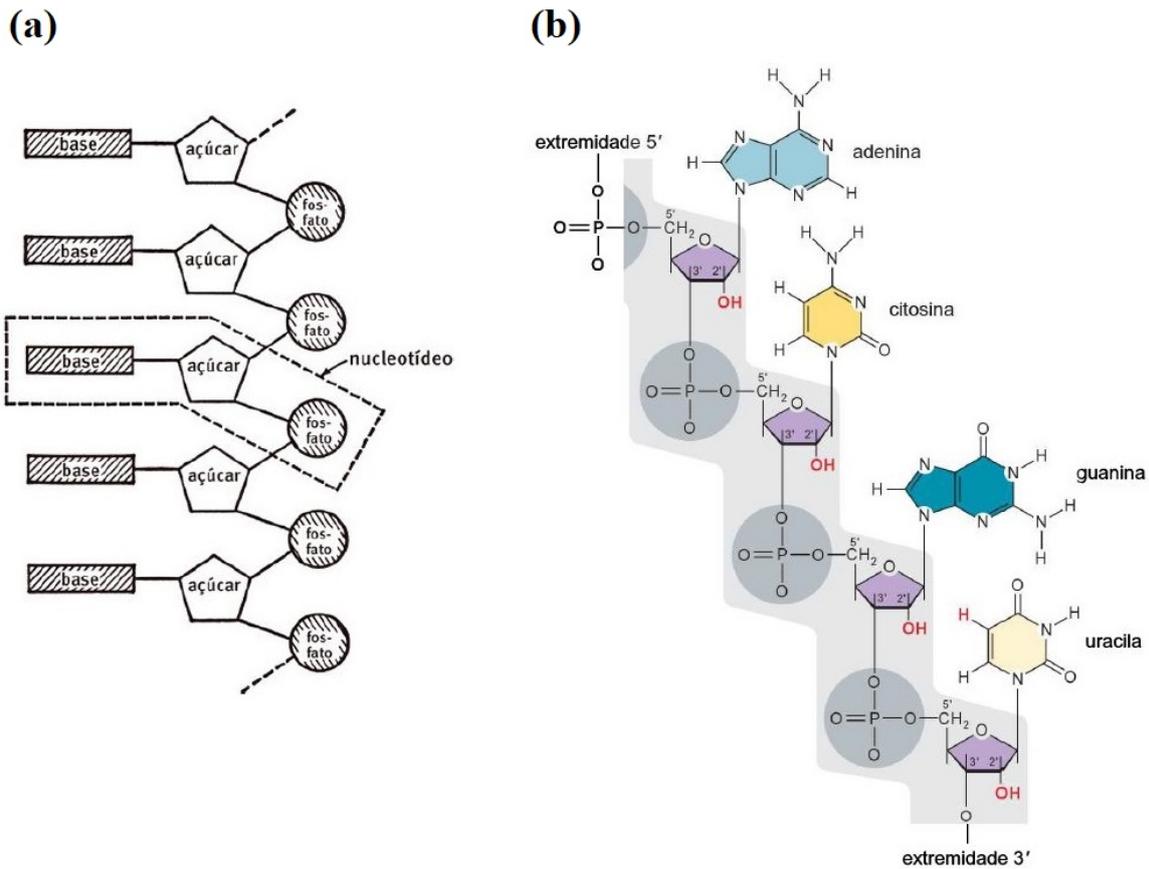


Figura 2.2: Uma porção do DNA, composto por uma sequência de nucleotídeos [2] e Ligação mostrando o padrão da molécula de DNA [48]

A molécula de açúcar, chamada 2'-desoxirribose, é constituída por 5 átomos de carbono - 1', 2', 3', 4', 5' (veja na Figura 2.3 (a)).

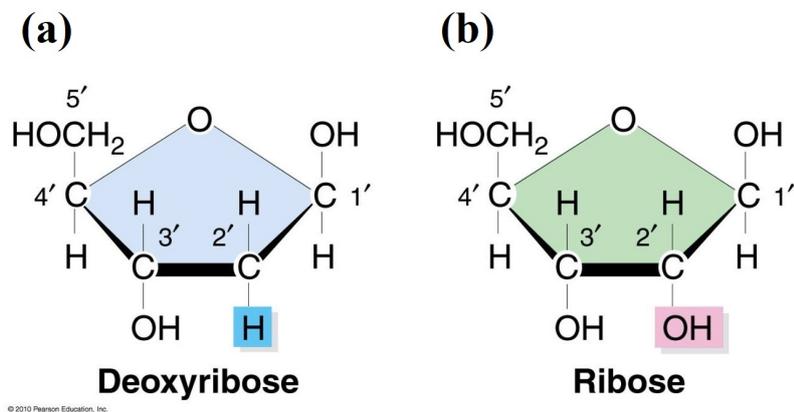


Figura 2.3: (a) Açúcar presente no DNA, a desoxirribose. (b) Açúcar presente no RNA, RNA, a ribose [3]

No DNA, existem quatro tipos de bases, que são moléculas ligadas a cada carbono na espinha dorsal, são elas adenina (A), guanina (G), citosina (C) e timina (T) (veja Figura 2.1). As bases purinas possuem dois anéis (um hexagonal e um pentagonal) de carbono e nitrogênio. Fazem parte desse grupo as bases A e G. Já as bases pirimidinas são menores que as púricas e formadas por um anel (hexagonal) de carbono e nitrogênio. Fazem parte desse grupo as bases C e T.

Uma cadeia de DNA, chamada de espinha dorsal, é criada pela ligação entre o carbono 3' de um nucleotídeo, o fosfato resíduo e o carbono 5' do próximo nucleotídeo. Essa composição leva a um padrão de orientação da molécula de DNA, que começa na extremidade 5' e termina na extremidade 3' (veja Figura 2.4).

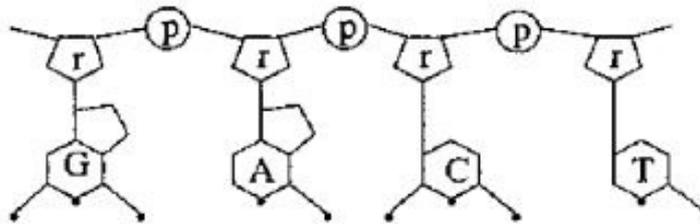


Figura 2.4: Estrutura molecular de uma fita de DNA [4]

O DNA é uma cadeia dupla de moléculas, mantida por ligações de hidrogênio entre pares de bases das cadeias opostas. Devido ao fato de serem agrupadas de forma helicoidal, Watson e Crick a chamaram, em 1953, de dupla hélice de DNA [2]. Portanto, a base A acopla-se à base T e a base C acopla-se à base G. Dizemos que as bases A e T são complementares (chama-se par de bases complementares), e a mesma regra é aplicada às bases C e G. Esses pares são conhecidos como pares de bases (pb) Watson-Crick.

2.1.2 RNA

RNA é uma molécula formada por ribonucleotídeos. Cada ribonucleotídeo é ligado entre si por ligações fosfodiéster, formando uma fita simples. Os nucleotídeos encontrados no DNA diferenciam-se dos encontrados no RNA pelo fato de que, no RNA, o açúcar é a ribose (veja Figura 2.3 (b)) e uma das bases nitrogenadas é a uracila (U) em vez da timina (T) (veja Figura 2.1).

Existem duas grandes classes de RNAs:

1. RNAs mensageiros (em inglês, *messenger RNA* - mRNA): codificam proteínas.
2. RNAs não-codificadores (em inglês, *non-coding RNAs* - ncRNAs): não codificam proteínas.

Na figura 2.5 possuímos exemplos de ncRNAs, temos o RNA transportador (em inglês, transfer RNA - tRNA), o RNA ribossômico (em inglês, *ribosomal RNA* - rRNA) e os pequenos RNA nucleares (em inglês, *small nuclear RNAs* - snRNA).

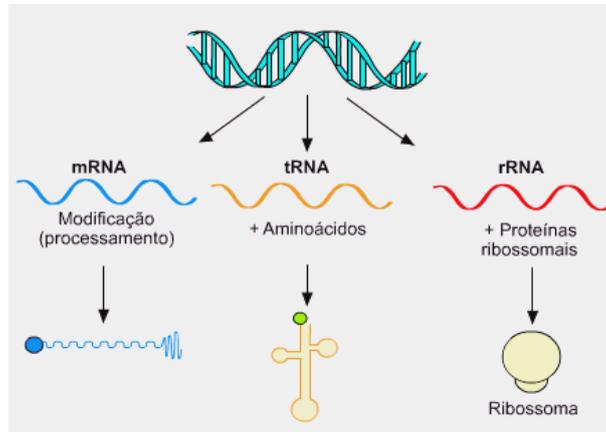


Figura 2.5: Exemplos de RNAs: mRNA e dois ncRNAs (tRNA e rRNA) [5]

Vamos agora explorar separadamente cada um desses RNAs.

RNA mensageiro (mRNA)

Essa classe de RNA é responsável por codificar as proteínas. É responsável por levar do núcleo celular, no caso dos organismos eucariotos, até o citoplasma as informações genéticas armazenadas no DNA.

É uma classe heterogênea, podendo apresentar estruturas que variam de 500 até mais de 6.000 nucleotídeos (Veja a figura 2.6).

Nos procariotos (organismos que não possuem núcleo celular, por exemplo, bactérias), o mRNA é transcrito diretamente do DNA. Mas, nos eucariotos, é necessário que haja um transcrito primário (pré-mRNA), formado a partir do DNA, que é processado para gerar o mRNA maduro.

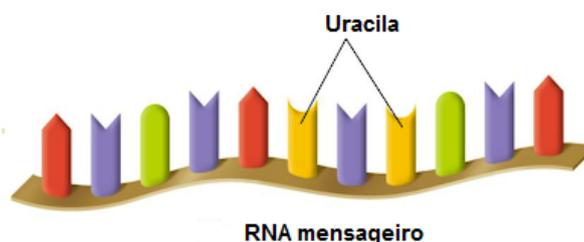


Figura 2.6: RNA mensageiro [6]

RNA ribossômico ou ribossomal (rRNA)

Recebe esse nome pois é o principal constituinte dos ribossomos (Veja figura 2.7). Ele é o principal responsável pela síntese de proteínas. O ribossomo é uma organela onde ocorre a síntese de proteínas. Cada ribossomo é formado por duas subunidades e cada uma dessas subunidades é constituída por cerca de três moléculas de rRNA e várias proteínas associadas. Assim que são sintetizados, os rRNAs acumulam-se, formando regiões conhecidas como nucléolos. Nesses locais, um rRNA combina-se com proteínas e origina os ribossomos.

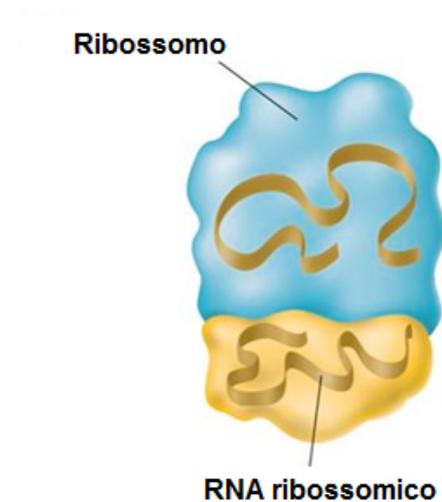


Figura 2.7: RNA ribossomal [7]

RNA transportador (tRNA)

Esse RNA é responsável por fazer a ligação códon (sequência de três nucleotídeos) e aminoácido. Sua função é transportar as moléculas de aminoácidos que serão utilizados na síntese de proteínas. O tRNA transporta essas moléculas até os ribossomos, local em que se unem e formam as proteínas. Comparado com os outros, este possui o menor peso.

O tRNA é pequeno, possui entre 74 e 95 nucleotídeos, e essa classe de RNA é semelhante em procariotos e eucariotos. Apresenta uma estrutura semelhante à de uma folha de trevo que possui quatro alças (veja Figura 2.8). Duas dessas alças possuem grande importância para que esse RNA realize suas funções: a alça que se liga a um aminoácido específico; e a alça do anticódon apresenta uma sequência complementar ao códon do mRNA. As outras duas alças são importantes para manter a estrutura dessa molécula de RNA.

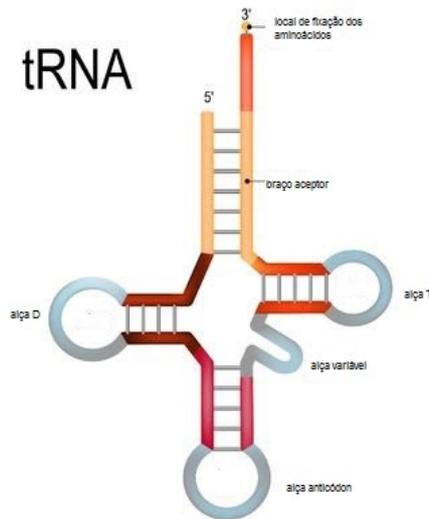


Figura 2.8: RNA transportador [8]

2.1.3 DNA vs RNA

Nas seções anteriores falamos a respeito das características do DNA e do RNA, mas afinal qual a diferença entre DNA e RNA? A Tabela 2.1 apresenta as características diferentes entre o DNA e o RNA.

Tabela 2.1: Diferenças entre DNA e RNA

Características	DNA	RNA
Pentose	Desoxirribose	Ribose
Bases Púricas	Adenina e Guanina	Adenina e Guanina
Bases Pirimidicas	Citosina e Timina	Citosina e Uracila
Estruturas	2 cadeias helicoidais	1 cadeia helicoidal
Origem	Replicação	Transcrição
Enzima	DNA-polimerase	RNA-polimerase
Função	Informação Genética	Síntese de Proteínas e outras funções

Com essas diferenças especificadas (Tabela 2.1), podemos definir outros componentes celulares. Por exemplo, o gene é um trecho contíguo de DNA que contém informações necessárias para construir uma proteína ou uma molécula de RNA. Os comprimentos dos genes variam, mas no caso dos humanos, um gene pode ter 10.000 bp. Certos mecanismos celulares são capazes de reconhecer no DNA os pontos precisos nos quais um gene começa e termina.

Na Tabela 2.9, temos a correspondência entre cada possível tripla de nucleotídeos (códon) e seu aminoácido correspondente (chamado também de código genético). Nota-se que os códon são dados usando bases de RNA, em vez de bases de DNA.

		2.ª BASE				
		U	C	A	G	
1.ª BASE	U	UUU } Fenilalanina (Fen) UUC } UUA } Leucina (Leu) UUG }	UCU } UCC } Serina (Ser) UCA } UCG }	UAU } Tirosina (Tir) UAC } UAA } Codão de finalização UAG } Codão de finalização	UGU } Cisteína (Cis) UGC } UGA } Codão de finalização UGG } Triptofano (Trp)	3.ª BASE U C A G
	C	CUU } CUC } Leucina (Leu) CUA } CUG }	CCU } CCC } Prolina (Pro) CCA } CCG }	CAU } Histidina (His) CAC } CAA } Glutamina (Glu) CAG }	CGU } CGC } Arginina (Arg) CGA } CGG }	
	A	AUU } Isoleucina (Ile) AUC } AUA } AUG } Metionina (Met) codão de iniciação	ACU } ACC } Treonina (Tre) ACA } ACG }	AAU } Asparagina (Asn) AAC } AAA } Lisina (Lis) AAG }	AGU } Serina (Ser) AGC } AGA } Arginina (Arg) AGG }	
	G	GUU } GUC } Valina (Val) GUA } GUG }	GCU } GCC } Alanina (Ala) GCA } GCG }	GAU } Ácido aspártico (Asp) GAC } GAA } Ácido glutâmico (Glu) GAG }	GGU } GGC } Glicina (Gli) GGA } GGG }	

Figura 2.9: Tabela de códons [9]

Observamos que diferentes códons correspondem ao mesmo aminoácido. Por exemplo, os códons AAG e AAA correspondem ao aminoácido lisina. Por outro lado, três dos possíveis códons não codificam nenhum aminoácido e são usados para sinalizar o final de um gene. Esses códons de terminação especiais são identificados na figura com a palavra STOP escrito na posição correspondente. O código genético mostrado na Figura 2.9 é usado pela grande maioria dos organismos vivos, mas alguns organismos usam um código ligeiramente modificado.

2.2 Proteínas

Nesta seção, descrevemos as proteínas, que são formadas por aminoácidos (Veja figura 2.10). Aminoácidos são compostos orgânicos que contêm grupos funcionais amino (NH₂) e carboxil (COOH), junto com uma cadeia lateral (grupo R) específica para cada aminoácido. Os principais elementos de um aminoácido são carbono (C), hidrogênio (H), oxigênio (O) e nitrogênio (N), embora outros elementos sejam encontrados nas cadeias laterais de certos aminoácidos. Na forma de proteínas, os resíduos de aminoácidos formam o segundo maior componente (a água é o maior) dos músculos humanos e outros tecidos. Além de seu papel como resíduos em proteínas, os aminoácidos participam de vários processos, como transporte de neurotransmissores e biosíntese.

Existem muitas proteínas diferentes que contêm funções distintas relacionadas ao número e classificação dos aminoácidos presentes, assim como sua estrutura tridimensional.

Uma cadeia linear de resíduos de aminoácidos é chamada de polipeptídeo. Uma proteína contém pelo menos um polipeptídeo longo. Polipeptídeos curtos, contendo menos

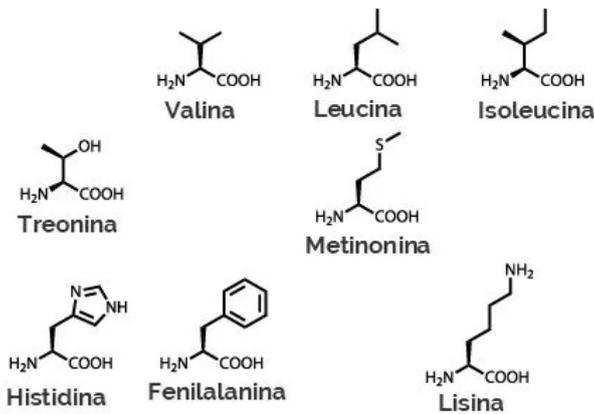


Figura 2.10: Composição química de alguns aminoácidos [10]

de 20-30 resíduos, raramente são considerados proteínas e são comumente chamados de peptídeos ou, às vezes, de oligopeptídeos. Os resíduos de aminoácidos individuais são ligados entre si por ligações peptídicas e resíduos de aminoácidos adjacentes. A sequência de resíduos de aminoácidos em uma proteína é definida pela sequência de um gene, que é codificado no código genético. Algumas proteínas têm grupos não peptídicos anexados, que podem ser chamados de grupos proteicos ou cofatores. As proteínas também podem trabalhar juntas para atingir uma função específica e frequentemente se associam para formar complexos de proteínas estáveis (Veja figura 2.11).

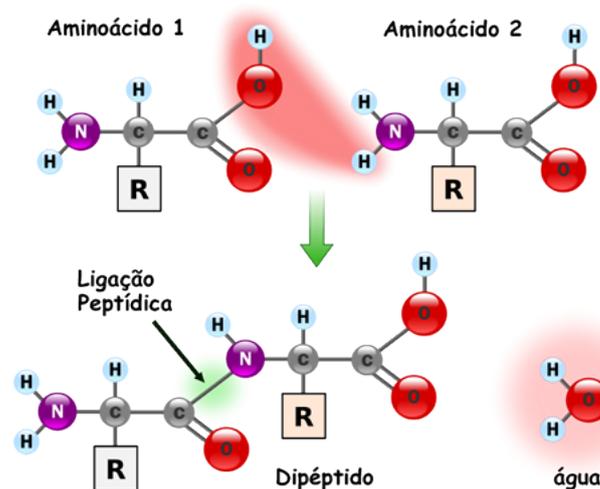


Figura 2.11: Uma proteína formada por 2 aminoácidos, junto com os demais componentes [11]

As proteínas podem ser classificadas em dois grupos básicos: proteínas fibrosas e proteínas globosas. O primeiro tipo se classifica em cadeias de aminoácidos que estão disponibilizadas de maneira a formar fios e, normalmente, são insolúveis em água. O

segundo tipo se classifica em cadeias que se enrolam sobre elas mesmas, formando assim uma estrutura esférica e normalmente solúvel em água.

Podemos definir estruturas de proteínas, de quatro formas diferentes (Figura 2.12):

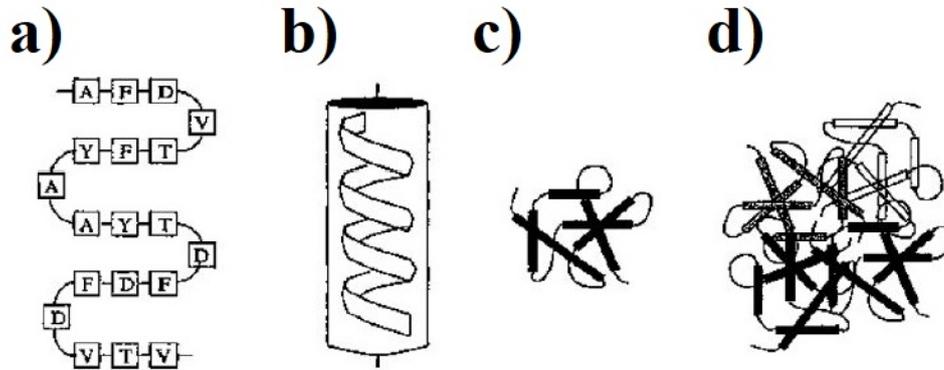


Figura 2.12: As quatro estruturas da proteína. (a) Estrutura primária. (b) Estrutura secundária. (c) Estrutura terciária. (d) Estrutura quaternária [4]

Estrutura primária

Podemos observar uma proteína de forma sequencial, ou seja, através da sequência de aminoácidos que a forma.

Estrutura secundária

Podemos observar uma proteína em um plano bidimensional, que é uma projeção no plano da estrutura tridimensional encontrada na natureza.

Estrutura terciária

Podemos observar o uma proteína em um plano tridimensional, ou seja, a modelagem da proteína da forma mais próxima possível à encontrada na natureza.

Estrutura quaternária

Podemos observar várias proteínas enoveladas, ou seja, como são as ligações das proteínas em um grupo.

2.3 Dogma Central da Biologia Molecular

O Dogma Central da Biologia Molecular, proposto por Crick em 1958 [2], esclarece o funcionamento do fluxo de informações do DNA para formar RNAs e proteínas. Na figura 2.13 de acordo com o dogma, o fluxo da informação genética segue o seguinte sentido:

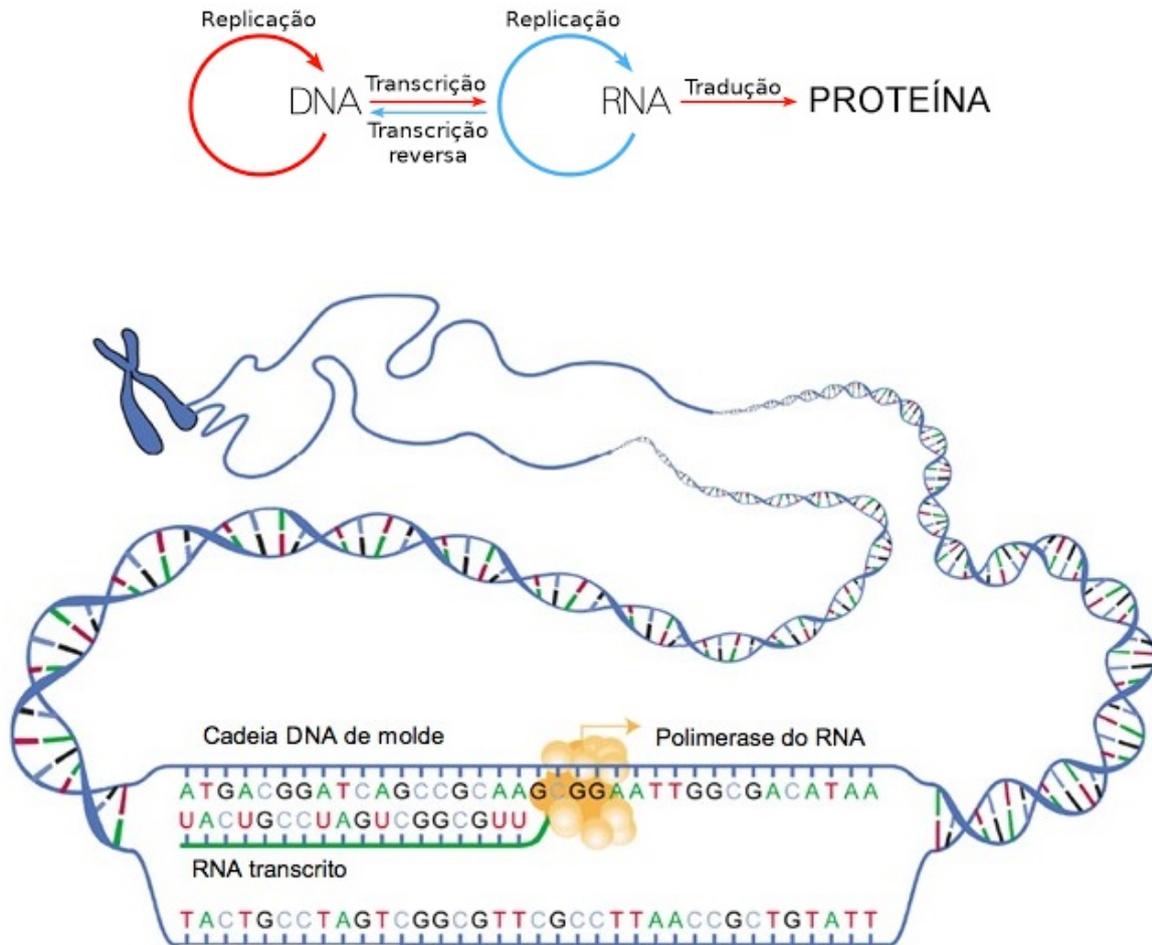


Figura 2.13: Dogma Central da Biologia Molecular, que mostra os processos de transformação dos ácidos nucleicos em proteínas, ou seja, é o processo da síntese de proteínas [12, 13]

A síntese de proteínas é um processo que ocorre dentro das células, sendo um processo muito semelhante para procarionotos e eucarionotos, com algumas diferenças. A síntese de proteínas pode ser dividida em duas grandes fases - transcrição e tradução. Durante a transcrição, uma seção de DNA que codifica uma proteína, conhecida como gene, é convertida em uma molécula modelo chamada RNA mensageiro (mRNA). Essa conversão é realizada por enzimas, conhecidas como RNA polimerases. Em eucarionotos, a transcrição ocorre no núcleo da célula. Nesse caso, o mRNA é inicialmente produzido em uma forma

inicial (pré-mRNA), que sofre modificações pós-transcricionais para produzir mRNA maduro. O mRNA maduro é exportado do núcleo da célula através de poros nucleares para o citoplasma da célula para que ocorra a tradução [2].

Muitos genes eucarióticos consistem em partes alternadas chamadas íntrons e exons. Os íntrons e exons são sequências de nucleotídeos dentro de um gene. Conforme o RNA amadurece, os íntrons serão removidos por splicing de RNA, o que significa que não serão expressos no produto final do RNA mensageiro (mRNA), enquanto os exons serão covalentemente ligados entre si para produzir mRNA maduro. Os íntrons podem ser considerados sequências de intervenção e os exons podem ser considerados sequências de expressão. Os exons são sequências de nucleotídeos em DNA e RNA, que são conservadas na produção de RNA maduro. O processo de usar o DNA como um modelo para criar mRNA é chamado de transcrição. Os íntrons são sequências de nucleotídeos no DNA e no RNA que não codificam diretamente as proteínas. Eles são removidos no estágio pré-mRNA (pré-mRNA) do splicing do mRNA no splicing do RNA [49].

Durante a tradução, o mRNA é lido por ribossomos, que usam a sequência de nucleotídeos do mRNA para determinar a sequência de aminoácidos. Os ribossomos catalisam a formação de ligações peptídicas entre os aminoácidos produzidos, para formar uma cadeia polipeptídica. Após a tradução, a cadeia polipeptídica deve dobrar para formar uma proteína funcional. Para adotar uma forma tridimensional funcional (3D), a cadeia polipeptídica deve primeiro formar uma série de estruturas subjacentes menores chamadas estruturas secundárias. A cadeia polipeptídica nessas estruturas secundárias então se dobra para produzir a estrutura terciária 3D. Uma vez dobrada corretamente, a proteína pode sofrer maturação adicional por meio de diferentes modificações pós-tradução. Modificações pós-tradução podem alterar a capacidade da proteína de funcionar, o local onde ela está localizada dentro da célula e a capacidade da proteína de interagir com outras proteínas [4].

A síntese de proteína tem um papel fundamental em doenças, pois alterações e erros nesse processo, por meio de mutações de DNA subjacentes ou dobra incorreto da proteína, costumam ser as causas subjacentes de uma doença. As mutações de DNA alteram a sequência de mRNA subsequente, que então altera a sequência de aminoácidos codificada pelo mRNA. As mutações podem fazer com que a cadeia polipeptídica seja mais curta, gerando uma sequência de parada que causa o término precoce da tradução. Alternativamente, uma mutação na sequência de mRNA altera o aminoácido específico codificado naquela posição na cadeia polipeptídica [2].

2.4 RNAs não-codificadores

De acordo com John S. Mattick [50], o termo RNA não codificante (ncRNA) é normalmente empregado para RNA que não codifica uma proteína, mas isto não significa que tais RNAs não contêm informação nem tenham função. Embora tenha sido geralmente assumido que a maioria da informação genética é transaccionada por proteínas, evidências recentes sugerem que a maioria dos genomas de mamíferos e outros organismos complexos é de facto transcrita em ncRNAs, muitos dos quais são alternativamente emendados e/ou processados em produtos mais pequenos. Os RNAs não-codificadores (ncRNAs) são responsáveis por desempenhar funções regulatórias e estruturais na célula, sendo classificados em: ncRNAs pequenos e ncRNAs longos (Veja Figura 2.14).

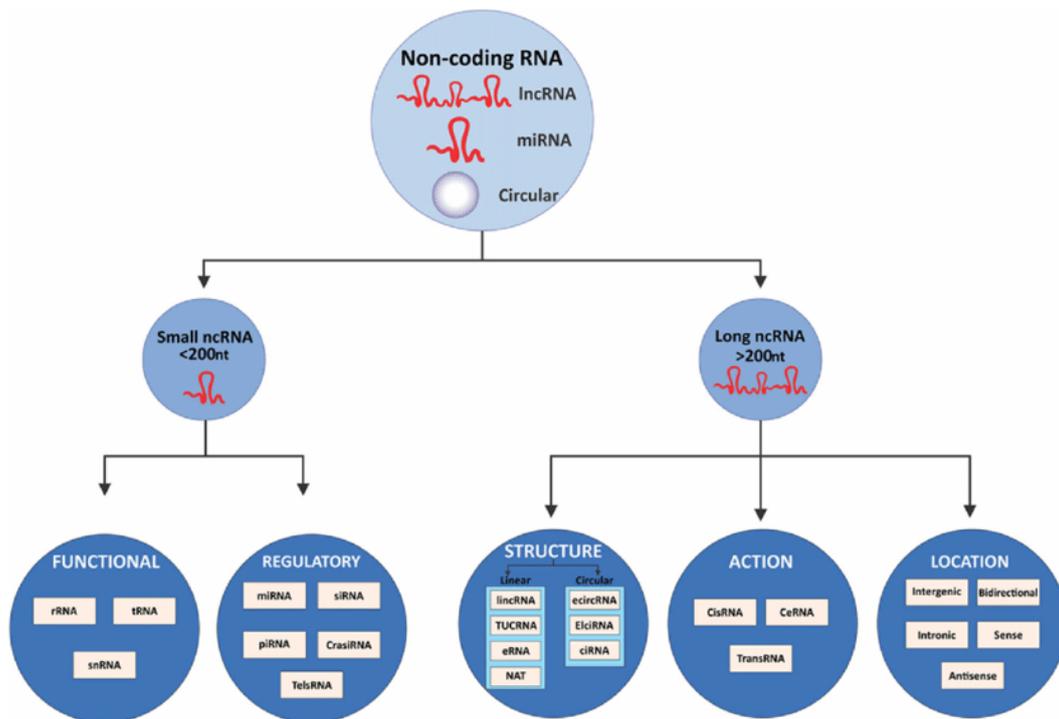


Figura 2.14: De forma genérica, as duas grandes classes de ncRNAs são os pequenos e os longos [14].

Estes RNAs (incluindo os derivados de introns) parecem compreender uma camada oculta de sinais internos que controlam vários níveis de expressão genética na fisiologia e no desenvolvimento, incluindo a arquitectura cromatina/memória hepigenética, transcrição, emenda de RNA, edição, tradução e rotação. As redes reguladoras do RNA podem determinar a maioria das nossas características complexas, desempenhar um papel significativo na doença e constituir um mundo inexplorado de variação genética, tanto dentro como entre espécies [50].

2.4.1 NcRNAs pequenos

Os ncRNAs pequenos possuem classificações de acordo com as suas funções em um organismo, sendo suas funções associados às suas estruturas. Um banco de dados chamado de RFAM [51] indica 3.016 classes de ncRNAs.

Dentre elas, podemos destacar: tRNAs, que transportam aminoácidos para ajudar na processo de síntese de proteínas; rRNAs, responsáveis pela catálise da síntese de proteínas; siRNAs (em inglês, *small interfering RNAs*), que podem causar interferência na tradução de proteínas, separando e promovendo a degradação de trechos de mRNAs; snoRNAs (em inglês, *small nucleolar RNAs*), que pode modificar rRNAs, tRNAs e snRNAs; e miRNAs (em inglês, *microRNAs*), que regulam o processo de tradução [52].

Com a percepção da diversidade de funções desempenhadas pelos ncRNAs pequenos, é necessário citar o desenvolvimento e supressão de doenças. Os siRNAs podem ser utilizados no tratamento de diversos tipos de câncer [53]. Com relação ao câncer, podemos também citar os miRNAs, que possuem a capacidade de regular a expressão de genes responsáveis por mecanismos de desenvolvimento, metabolismo, proliferação celular, diferenciação e apoptose [30], que podem participar do início e da progressão do câncer.

MicroRNAs

Os MicroRNAs (ou miRNAs) compreendem uma classe de pequenos RNAs endógenos não codificadores que regulam a expressão gênica por inibição da tradução ou degradação de seus mRNAs alvo na pós-transcrição [30]. A sua descoberta acrescentou uma nova dimensão à compreensão das complexas redes reguladoras de genes, tanto em humanos como em animais. Processos como proliferação celular, diferenciação e apoptose também envolvem determinadas interações de miRNAs [54]. Mudanças na expressão de miRNAs podem causar doenças, em particular a iniciação de tumor, progressão e metástase. Vários mecanismos como locus gênico, amplificação, deleção cromossômica, mutação e silenciamento epigenético foram identificados como responsáveis pela desregulação da expressão de miRNAs em alguns tipos de câncer [30, 55].

MicroRNAs (miRNAs) são ncRNAs pequenos, de fita simples, normalmente com cerca de 22 nucleotídeos de comprimento. As 22 sequências de RNA altamente conservadas regulam a expressão dos genes através da ligação às regiões 3'-untraduzidas (3'-UTR) de mRNAs específicos. Um conjunto crescente de provas mostra que os mRNAs são um dos principais actores na diferenciação e crescimento celular, mobilidade e apoptose (morte celular programada) [56, 57].

Diferenciar os miRNAs de outras classes de pequenos RNAs que estão presentes na célula é frequentemente incômodo - particularmente a distinção dos pequenos RNAs endó-

genos interferentes (siRNAs), a distinção mais significativa entre os miRNAs e os siRNAs é se eles silenciam a sua própria expressão. Quase todos os siRNAs (independentemente da sua origem viral ou outra) silenciam o mesmo locus a partir do qual foram derivados. Por outro lado, a maioria dos miRNAs não silencia os seus próprios loci, mas sim outros genes [58, 59]. Os miRNAs regulam diversos aspectos do desenvolvimento e da fisiologia, por isso a compreensão do seu papel biológico revela-se cada vez mais importante. A análise da expressão miRNA pode fornecer informação valiosa, uma vez que a desregulação da sua função pode levar a doenças humanas tais como o cancro, doenças cardiovasculares e metabólicas, condições hepáticas e disfunções imunitárias.

2.4.2 NcRNAs longos

Os ncRNAs longos (em inglês, *long ncRNAs* -lncRNAs) possuem mais de 200 nucleotídeos de comprimento, sendo classificados em seis categorias[60] :

1. *Sense*: quando o lncRNA se sobrepõe à região de transcrição de um ou mais exons de outro gene, na mesma fita;
2. *Antisense*: quando o lncRNA se sobrepõe à região de transcrição de um ou mais exons de outro gene, na fita oposta;
3. *Bidirectional*: quando o início da transcrição do lncRNA e outro gene na fita oposta estão próximos;
4. *Intronic*: quando os lncRNAs são derivados inteiramente em íntrons;
5. *Enhancer*: quando os lncRNAs estão localizados entre dois genes;
6. *Intergenic*: quando o lncRNA está localizado no intervalo entre dois genes.

Os lncRNAs regulam a expressão gênica, alterando a estrutura da cromatina, silenciando ou ativando um gene ou família de genes e, em certos casos, inteiros cromossomos via métodos cis ou trans. Eles podem regular como [60]:

1. *Decoy Effect*: ligando-se a outros RNAs e proteínas para alterar suas funções;
2. *Scaffold Effect*: conectando proteínas modificadas com cromatina e regiões de DNA para formar conexões de sinal;
3. *Post-Transcriptional Effect*: formando dímeros de RNA com sequências de mRNA para bloquear locais associados à transcrição.

Com a diversidade das funções desempenhadas pelos ncRNAs longos, podemos incluir a participação no desenvolvimento e na supressão de doenças. Assim, várias referências bibliográficas, tanto em Biologia Molecular quanto em Bioinformática, relacionam lncRNAs como biomarcadores para entender diversos tipos de câncer [61, 62, 63, 64, 65]. Esses estudos mostram as funções dos lncRNAs na epigenética do câncer e também a interação entre os lncRNAs e os miRNAs no desenvolvimento do câncer.

As ferramentas computacionais não são muito eficazes para prever lncRNAs. Normalmente, os lncRNAs são identificados filtrando os transcritos produzidos por RNASeq, utilizando características biológicas importantes, tais como: comprimento maior que 200 pares de bases; comprimento de ORF maior ou igual a 100 aminoácidos; e indicadores de expressão de proteínas. Algumas ferramentas de bioinformática, por exemplo, BLAST para anotação e CPAT [66] e CPC [67] para identificação de potencial codificador de proteínas [68, 69, 70, 71, 72, 73] são utilizadas nessas ferramentas, além de métodos de aprendizado de máquina.

Para lncRNA, o método que tem mostrado bons resultados é o aprendizado de máquina [60]. Por exemplo, existem algumas ferramentas que podem distinguir lncRNA de PCT (em inglês *transcrição de codificação de proteína*) e identificar certos tipos de lncRNA, como indicado na citação [74]. Métodos baseados em algoritmos de aprendizado de máquina têm sido propostos, por exemplo, SVM (*Support Vector Machine*), regressão logística (LR), aprendizado profundo (DL) e floresta aleatória (RF). Existem muitos métodos para identificar e classificar lncRNAs, [75, 76, 77, 74, 78, 79, 80], e alguns métodos são usados principalmente para prever a categoria específica de lncRNA, como iSeeRNA [81] baseado em SVM. Também temos o linc-SF [82] (classificador baseado em características selecionadas e algoritmo genético, chamado GA-SVM (linc-SF) [82]). Apesar de termos alguns métodos para distinguir lncRNAs de PCTs, faltam métodos que tentem prever o papel ou relação dessas moléculas no surgimento de doenças.

NcRNAs circulares

Normalmente formados por exons e íntrons, circRNAs são classificados como sendo ligados a RNAs não codificantes, formando um círculo. Sua forma original permite evitar a degradação pela exonuclease de RNA e contribui para sua abundância e estabilidade no citoplasma das células eucarióticas.

CiRNA e ecRNA são resistentes a RNase R (ou Ribonuclease R, é uma exoribonuclease 3'→5', que pertence à superfamília RNase II, um grupo de enzimas que hidrolisam o RNA na direção 3' - 5'. O RNase R demonstrou estar envolvido na degradação selectiva do mRNA, particularmente de mRNAs não parados em bactérias [83, 84]). Enquanto o ecRNA é encontrado no citoplasma, o ciRNA está localizado nos núcleos das células e

é constituído de íntrons derivados do RNA linear. Durante o salto de exon, ecRNA foi encontrado no pré-mRNA. Assim como o ciRNA, o elciRNA também existe no núcleo das células, originando-se de íntrons e exons. CircRNA é chamado de esponjas de micro RNA (miRNA) devido à sua capacidade de se ligar a moléculas de miRNA e afetar sua função. Eles desempenham um papel fundamental na regulação do genes alvo de miRNA. Devido à sua estabilidade, o circRNA também pode atuar como iscas de proteínas, ligando-se e interagindo com muitas proteínas. CircRNAs são reguladores transcricionais e podem ser traduzidos [85, 86, 87] .

Os circRNAs podem ser categorizado em três grupos:

1. CircRNAs intrônicos (em inglês, *intronic circRNAs (ciRNAs)*);
2. CircRNAs exônicos (em inglês, *Exonic circRNAs (ecRNAs)*);
3. CircRNAs exônicos-intrônicos (em inglês, *exonic-intronic circRNAs (elciRNA)*).

A diversidade entre eles é enorme (mais de 30.000 tipos de circRNAs foram identificados), e espera-se que o número de circRNAs descobertos cresça de forma constante [85]. A descoberta de novos circRNAs pode ser auxiliada pela utilização de bases de dados que usam algoritmos preditivos. As bases de dados podem também prever a função de circRNAs específicos, permitindo que os circRNAs sejam associados a estados de doenças relacionados ao sistema cardiovascular e doenças cancerígenas, sabe-se que a expressão de circRNAs específicos muda em várias fases de desenvolvimento dessas doenças, os circRNAs são conhecidos por desempenharem papéis no desenvolvimento, progressão e regulação dos estados dessas doenças [88, 89, 90].

O papel regulador desempenhado pelos circRNAs seria melhor compreendido se os mecanismos por detrás da biogênese dos circRNAs fossem totalmente caracterizados. Uma vez que a investigação do circRNA é relativamente nova, investigações futuras sobre o seu potencial parece promissor e deverá reunir conhecimentos interessantes sobre a sua utilização como biomarcadores e alvos terapêuticos. É possível desenvolver abordagem individualizada para o tratamento de câncer, devido à capacidade dos circRNAs de se adaptarem a diferenças nos biomarcadores expressos em tecido canceroso e tecido normal. Desta maneira, os circRNAs têm potencial para serem utilizados como biomarcadores de câncer, sendo essa uma área interessante de investigação [86].

Capítulo 3

Noções básicas de câncer colorretal

Neste capítulo, apresentamos de forma breve os conceitos básicos do Câncer colorretal (CRC), necessários ao entendimento deste trabalho. Na Seção 3.1, falaremos sobre conceitos básicos do câncer. Na Seção 3.2, abordaremos especificamente algumas características do câncer colorretal. Na Seção 3.3, descreveremos o envolvimento de RNAs não-codificadores (ncRNAs) em CRC.

3.1 Noções gerais do câncer

Câncer é o nome genérico para um grupo de mais de 100 doenças malignas, que possuem em comum o crescimento desordenado de células, que podem invadir inclusive tecidos ou órgãos mais distantes [15]. Dividindo-se muito rapidamente, essas células tendem a ser muito agressivas e incontroláveis, determinando a formação de tumores, que podem espalhar-se para outras regiões do corpo [91]. A enfermidade também é conhecida como neoplasia e a ciência médica que estuda o câncer denomina-se Oncologia, enquanto que oncologista é o médico que trata a doença [15, 92, 93].

Os diferentes tipos de câncer correspondem aos vários tipos de células do corpo (Veja Figura 3.2). Quando começam em tecidos epiteliais, como pele ou mucosas, são denominados carcinomas. Se o ponto em que se inicia o câncer são os tecidos conjuntivos, como osso, músculo ou cartilagem, são chamados sarcomas [94]. Outras características que diferenciam os diversos tipos de câncer entre si são a velocidade de multiplicação das células e a capacidade de invadir tecidos e órgãos vizinhos ou distantes, conhecida como metástase [15, 94].

De acordo com o INCA(Instituto Nacional de Câncer) [15], o câncer surge a partir de uma mutação genética, ou seja, de uma alteração no DNA da célula, que passa a receber instruções erradas para as suas atividades. As alterações podem ocorrer em genes especiais, denominados proto-oncogenes, que a princípio são inativos em células normais.

Quando ativados, os proto-oncogenes tornam-se oncogenes, responsáveis por transformar as células normais em células cancerosas. O processo de formação do câncer (chamado de carcinogênese ou oncogênese), em geral, acontece lentamente, podendo levar vários anos para que uma célula cancerosa prolifere-se e dê origem a um tumor visível. Esse processo é composto por três estágios (Veja na Figura 3.1).



Figura 3.1: Os três estágios da formação do câncer [15]

Estágio de Iniciação

Os genes sofrem ação dos agentes cancerígenos, que provocam modificações em alguns de seus genes. Nessa fase, as células se encontram geneticamente alteradas, porém ainda não é possível se detectar um tumor clinicamente. Elas encontram-se "preparadas", ou seja, "iniciadas" para a ação de um segundo grupo de agentes que atuará no próximo estágio.

Estágio de Promoção

As células geneticamente alteradas, ou seja, "iniciadas", sofrem o efeito dos agentes cancerígenos classificados como oncopromotores. A célula iniciada é transformada em célula maligna, de forma lenta e gradual. Para que ocorra essa transformação, é necessário um longo e continuado contato com o agente cancerígeno promotor. A suspensão do contato com agentes promotores muitas vezes interrompe o processo nesse estágio. Alguns componentes da alimentação e a exposição excessiva e prolongada a hormônios são exemplos de fatores que promovem a transformação de células iniciadas em malignas.

Estágio de Progressão

Este estágio é caracterizado pela multiplicação descontrolada e irreversível das células alteradas. Nesse estágio, o câncer já está instalado, evoluindo até o surgimento das primeiras manifestações clínicas da doença. Os fatores que promovem a iniciação ou

progressão da carcinogênese são chamados agentes oncoaceleradores ou carcinógenos. O fumo é um agente carcinógeno completo, pois possui componentes que atuam nos três estágios da carcinogênese.

As opções de tratamento dependem do tipo de câncer, da sua fase, se o câncer se espalhou e da sua saúde em geral. O objetivo do tratamento é matar o maior número possível de células cancerosas, reduzindo ao mesmo tempo os danos às células normais nas proximidades[95]. Os avanços da tecnologia tornam isto possível.

Os três tratamentos principais são:

- Cirurgia: remoção directa do tumor.
- Quimioterapia: utilização de produtos químicos para matar células cancerosas.
- Radioterapia: utilização de raios X para matar células cancerosas.

O mesmo tipo de câncer em um indivíduo X é muito diferente do câncer em outro indivíduo. Dentro de um único tipo de câncer, como o câncer de mama, os pesquisadores descobrem subtipos que requerem uma abordagem de tratamento diferente [95, 15]. Os tipos de câncer [96] podem ser agrupados em categorias mais amplas. As principais categorias incluem:

- Carcinomas: começam na pele ou nos tecidos que revestem ou cobrem os órgãos internos. Existe um número de subtipos de carcinoma, incluindo adenocarcinoma, carcinoma de células basais, carcinoma de células escamosas e carcinoma de células de transição.
- Sarcomas: começam no osso, cartilagem, gordura, músculo, vasos sanguíneos ou outro tecido conjuntivo ou de suporte;
- Leucemias: começam no tecido que produz o sangue, como a medula óssea, o que provoca um grande número de células anormais que entram na circulação sanguínea;
- Linfomas e Mielomas: começam nas células do sistema imunológico;
- Cânceres do Sistema Nervoso Central: começam nos tecidos do cérebro e da medula espinhal.

O futuro do tratamento do câncer reside em proporcionar aos doentes um nível de personalização ainda maior. Os médicos começaram a oferecer opções de tratamento com base nas alterações genéticas que ocorrem num tumor específico. Uma nova ferramenta inovadora de diagnóstico, a avaliação genómica do tumor, examina geneticamente o tumor de um paciente para identificar o mecanismo que causou o câncer [97].

- Em homens, Brasil, 2020

Localização Primária	Casos Novos	%
Próstata	65.840	29,2
Cólon e Reto	20.540	9,1
Traqueia, Brônquio e Pulmão	17.760	7,9
Estômago	13.360	5,9
Cavidade Oral	11.200	5,0
Esôfago	8.690	3,9
Bexiga	7.590	3,4
Laringe	6.470	2,9
Leucemias	5.920	2,6
Sistema Nervoso Central	5.870	2,6
Todas as Neoplasias, exceto pele não melanoma	225.980	100,0
Todas as Neoplasias	309.750	

- Em mulheres, Brasil, 2020

Localização Primária	Casos Novos	%
Mama feminina	66.280	29,7
Cólon e Reto	20.470	9,2
Colo do útero	16.710	7,5
Traqueia, Brônquio e Pulmão	12.440	5,6
Glândula Tireoide	11.950	5,4
Estômago	7.870	3,5
Ovário	6.650	3,0
Corpo do útero	6.540	2,9
Linfoma não-Hodgkin	5.450	2,4
Sistema Nervoso Central	5.230	2,3
Todas as Neoplasias, exceto pele não melanoma	223.110	100,0
Todas as Neoplasias	316.280	

Fonte:

- MS / INCA / Estimativa de Câncer no Brasil, 2020
- MS / INCA / Coordenação de Prevenção e Vigilância / Divisão de Vigilância e Análise de Situação

Figura 3.2: Incidência estimada conforme a localização primária do tumor e sexo [15].

3.2 Câncer Colorretal

O câncer colorretal, também chamado de câncer de cólon e reto ou câncer do intestino, abrange os tumores que se iniciam na parte do intestino grosso chamada cólon e no reto (final do intestino, imediatamente antes do ânus) e ânus [98, 99]. É tratável e, na maioria dos casos, curável, ao ser detectado precocemente, quando ainda não se espalhou para outros órgãos. Grande parte desses tumores se inicia a partir de pólipos, lesões benignas que podem crescer na parede interna do intestino grosso. A Figura 3.3 traz estatísticas do INCA, relativas ao CRC, no Brasil.

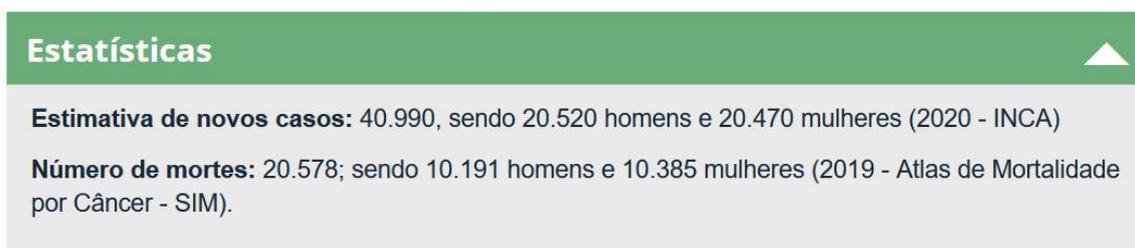


Figura 3.3: Estatísticas retiradas do site do INCA, para CRC, no Brasil [15].

Os principais fatores relacionados a riscos altos de uma pessoa desenvolver câncer do intestino são: idade igual ou acima de 50 anos; excesso de peso corporal; e alimentação não saudável, ou seja, pobre em frutas, vegetais e outros alimentos que contenham fibras. O consumo de carnes processadas (salsicha, mortadela, linguiça, presunto, bacon, peito de peru e salame) e a ingestão excessiva de carne vermelha (acima de 500 gramas de carne cozida por semana) também aumentam o risco para este tipo de câncer [100].

Outros fatores relacionados a maior chance de desenvolvimento da doença são história familiar de câncer de intestino, história pessoal de câncer de intestino, ovário, útero ou mama, além de tabagismo e consumo de bebidas alcoólicas. Doenças inflamatórias do intestino, como retocolite ulcerativa crônica e doença de Crohn, também aumentam o risco de câncer do intestino, bem como doenças hereditárias, como polipose adenomatosa familiar (FAP) e câncer colorretal hereditário sem polipose (HNPCC). A exposição ocupacional à radiação ionizante, como aos raios X e gama, pode aumentar o risco para câncer de cólon [101].

A detecção pode ser feita por meio da investigação com exames clínicos, laboratoriais ou radiológicos, de pessoas com sinais e sintomas sugestivos da doença (diagnóstico precoce), ou com o uso de exames em pessoas sem sinais ou sintomas (rastreamento) mas pertencentes a grupos com maior chance de ter a doença. Os tumores de cólon e reto (ou colorretal) podem ser detectados precocemente através de dois exames principais: pesquisa de sangue oculto nas fezes e endoscopias (colonoscopia ou retossigmoidoscopias) [102].

Os sintomas mais frequentemente associados ao câncer do intestino são:

- Sangramento nas fezes;
- Alteração na forma das fezes (fezes muito finas e comprimidas);
- Massa (tumoração) abdominal;
- Dor abdominal;
- Perda de peso e anemia;
- Mudança de hábito intestinal.

Além do diagnóstico precoce, a Organização Mundial da Saúde (OMS) [103] preconiza que os países com condições de garantir a confirmação diagnóstica, referência e tratamento, realizem o rastreamento do câncer do colon e reto em pessoas acima de 50 anos, por meio do exame de sangue oculto de fezes. Os casos positivos neste exame deverão fazer uma colonoscopia ou retossigmoidoscopia, onde o médico visualizará a parte interna do intestino buscando o câncer ou pólipos que possam vir a se transformar em câncer.

Quando apanhados cedo, o câncer colorrectal pode ser tratado com uma colonoscopia para remover pólipos ou células cancerosas do revestimento do cólon [102]. As fases avançadas da doença podem requerer cirurgia para remover parte do intestino afetado e os gânglios linfáticos (pequenas estruturas que fazem parte do sistema de defesa do corpo) dentro do abdome, ou em em casos raros, todo o cólon[101]. Outras opções de tratamento incluem imunoterapia(pode ser uma opção para pacientes cujo câncer tem características genômicas específicas), terapia orientada(utiliza medicamentos de engenharia biológica que visam proteínas específicas encontradas em células cancerígenas. Estes medicamentos podem ser utilizados isoladamente ou em combinação com outros tratamentos), quimioterapia e radioterapia para diminuir a possibilidade de recidiva (retorno) do tumor. O tratamento depende principalmente do tamanho, localização e extensão do tumor. Quando a doença está espalhada, com metástases para o fígado, pulmão ou outros órgãos, as chances de cura ficam reduzidas. Após o tratamento, é importante realizar o acompanhamento médico para monitoramento de recidivas ou novos tumores [98, 99, 104, 101, 100, 102].

3.2.1 Tipos de câncer colorretal

Os adenocarcinomas do cólon e do reto constituem 95% de todos os casos de câncer colorretal. No tracto (conjunto de canais, cordões, vias, feixes ou órgãos que pertencem a um mesmo sistema anatomofisiológico) gastrointestinal, os adenocarcinomas do cólon e do recto desenvolvem-se nas células do revestimento no interior do intestino grosso. Os

adenocarcinomas do cólon e do reto começam tipicamente como um crescimento de tecido chamado pólipó. Um tipo particular de pólipó chamado adenoma pode desenvolver-se em câncer. Os pólipos são frequentemente removidos durante uma colonoscopia de rotina antes de poderem desenvolver-se em câncer [105, 106, 107].

Adenocarcinoma colorretal

"Adeno" é um prefixo que significa glândulas. "Carcinoma" é um tipo de cancro que cresce em células epiteliais que alinham as superfícies dentro e fora do corpo. Os adenocarcinomas do cólon ou do recto desenvolvem-se no revestimento do intestino grosso. Começam frequentemente no forro interior e depois propagam-se a outras camadas. Existem dois subtipos menos comuns de adenocarcinoma: O adenocarcinoma mucoso é constituído por cerca de 60% de muco. O muco pode fazer com que as células cancerosas se espalhem mais rapidamente e se tornem mais agressivas do que os típicos adenocarcinomas. Os adenocarcinomas mucosos representam 10 a 15 por cento de todos os adenocarcinomas retais e do cólon. O adenocarcinoma de células anelares Signet é responsável por menos de 1 por cento dos adenocarcinomas. Nomeado pelo seu aparecimento sob um microscópio, o adenocarcinoma de células anelares signet é tipicamente agressivo e pode ser mais difícil de tratar [108, 109, 110].

Tumores de carcinoides gastrointestinais

Os tumores carcinoides desenvolvem-se em células nervosas chamadas células neuroendócrinas, que ajudam a regular a produção hormonal. Estes tumores estão entre um grupo de cancros chamados tumores neuroendócrinos (NETs). Os tumores carcinoides são de crescimento lento e podem desenvolver-se nos pulmões e/ou no tracto gastrointestinal. São responsáveis por 1% de todos os cancros colorrectais e metade de todos os cancros encontrados no intestino delgado [111, 112, 113].

Outros tipos de cancros colorretais raros combinados representam menos de 5 por cento de todos os casos e incluem os seguintes tipos.

Linfomas colorretais primários

Um tipo de linfoma não-Hodgkin, este tipo de cancro desenvolve-se no sistema linfático, especificamente em células chamadas linfócitos. Os linfócitos são um tipo de glóbulos brancos que ajudam o corpo a combater as infecções. O linfoma pode desenvolver-se em muitas partes do corpo, incluindo os gânglios linfáticos, a medula óssea, o baço, o timo e o tracto digestivo. Os linfomas colorrectais primários representam apenas 0,5 por cento de todos os cancros colorrectais e cerca de 5 por cento de todos os linfomas. Este tipo de

cancro colorrectal ocorre geralmente mais tarde na vida, e é mais comum nos homens do que nas mulheres [114, 115, 116].

Tumores estromais gastrintestinais

Os GISTs são um tipo raro de câncer colorretal que se forma numa célula especial encontrada no revestimento do tracto gastrointestinal (GI) chamado células intersticiais de Cajal (ICCs). Mais de 50 por cento dos GIST desenvolvem-se no estômago. Enquanto a maioria dos outros GISTs se formam no intestino delgado, o reto é o terceiro local mais comum. Os GISTs são classificados como sarcomas, ou cancros que começam nos tecidos conjuntivos, que incluem gordura, músculo, vasos sanguíneos, tecidos profundos da pele, nervos, ossos e cartilagem [117, 118, 119].

Leiomiiossarcomas

Outra forma de sarcoma, o leiomiiossarcoma significa essencialmente "cancêr do músculo liso". O cólon e o reto têm três camadas do tipo de músculo afetado pelo leiomiiossarcoma, e os três trabalham em conjunto para orientar os resíduos através do tracto digestivo. Este tipo raro de câncer colorrectal é responsável por cerca de 0,1% de todos os casos colorrectais [120, 121].

Melanomas

Mais frequentemente associados ao câncêr de pele, estes podem ocorrer em qualquer lugar, incluindo o cólon ou o reto [122, 123, 124].

3.3 RNAs não-codificadores em câncêr colorretal

A expressão dos genes nas células e tecidos de cada organismo complexo é precisamente controlada e depende em grande parte de diferentes condições (tais como desenvolvimento, alterações no ambiente, doenças ou drogas). Várias células e sistemas de órgãos dentro desse organismo (incluindo humanos) contêm diferentes perfis de expressão genética, pelo que a compreensão adequada dos mecanismos reguladores envolvidos nessa expressão representa uma das questões-chave da medicina genómica.

As moléculas não codificadoras de RNA têm um papel na multiplicidade de eventos reguladores - desde o controle do número de cópias na divisão bacteriana até à inactivação do cromossoma X nos mamíferos. Análises recentes dos genomas humano e animal mostraram que a maioria das transcrições de RNA não codificam proteínas (ou seja, são RNAs mensageiros ou mRNAs), mas sim RNAs não codificadores (ncRNAs).

Já são conhecidos um conjunto relativamente grande de RNAs não-codificadores associados a CRC.

Os interesses de investigação em microRNAs têm aumentado, sobretudo na última década. Muitos estudos mostraram que os microRNAs têm relações estreitas com vários tipos de câncer humanos, e poderiam potencialmente ser utilizados como indicadores no diagnóstico de câncer ou como supressores, para fins de tratamento.

Existem várias bases de dados que contêm microRNAs associados a câncer, previstos por métodos computacionais, mas poucos a partir de resultados empíricos. Apesar de terem sido realizadas muitas experiências de investigação de expressões de microRNA em células cancerígenas, os resultados continuam dispersos na literatura. Hoje em dia, podemos realizar a extração de associações de microRNAs em câncer por mineração de texto, por exemplo, na base de dados chamada miRCancer. O miRCancer armazena atualmente 878 relações entre 236 microRNAs e 79 tipos de câncer humanos, extraídos de mais de 2.000 artigos publicados [125].

O banco de dados LncRNA Disease [126] tem como objetivo analisar, prever e modelar associações de doenças com RNAs. Esse banco de dados apresenta métodos e ferramentas de bioinformática para a predição de associações de doenças RNA (principalmente miRNAs e lncRNAs). As informações trazem os principais miRNAs e lncRNAs envolvidos em doenças importantes e complexas, tais como doenças cardiovasculares, câncer e diabetes.

Capítulo 4

Noções básicas de Banco de Dados

Neste capítulo, apresentamos noções básicas de Banco de Dados, necessários a metodologia deste trabalho. Na Seção 4.1, faremos uma breve introdução para caracterizar o que é um banco de dados. Na Seção 4.2, Falaremos a respeito de quais características fundamentais um banco de dados necessita. Na Seção 4.3, conceituaremos o que é o modelo Entidade-Relacionamento (modelo ER) e como ele formaliza os dados em um banco de dados. Por fim, iremos abordar na Seção 4.4, a estrutura de um modelo relacional. Neste projeto, o modelo utilizado é o de banco de dados relacional.

4.1 Informações, dados e bancos de dados

Com o avanço da tecnologia, passou-se a notar a necessidade de manusear os dados de forma mais prática e com a possibilidade de interligar os dados pertencentes aquele conjunto de informações, para isso foi criado o Banco de Dados. O Bancos de dados e sistemas de banco de dados são fundamentais na vida da nossa sociedade atualmente. A maioria das atividades realizadas hoje em dia abrangem interações com um banco de dados, por exemplo quando um cliente de uma instituição financeira consulta as seguinte opções: saldo, produtos e serviços disponibilizados ele está acessando um banco de dados com a finalidade de verificação de possíveis negócios ou linhas de créditos disponíveis [127]. Também acessamos banco de dados quando vamos comprar uma passagem, ou consultar se há debitos a serem quitados no site do Detran, ou até mesmo uma simples matrícula das matérias da faculdade, todas essa interações podem ser chamadas de aplicações de banco de dados tradicionais, onde as informações armazenadas se correlacionam e onde as informações acessadas são de forma textual ou numérica [128].

Com a necessidade de facilitar cada vez mais o cotidiano, foram desenvolvidas novas tecnologias onde podemos armazenar mídias (imagens, vídeos e áudios), damos o nome dessas estruturas de banco de dados de multimídia. Temos como exemplo o GPS na

Figura 4.1, um sistema de informações geográficas onde é possível armazenar e analisar mapas e imagens de satélites e traçar o melhor caminho para o usuário.

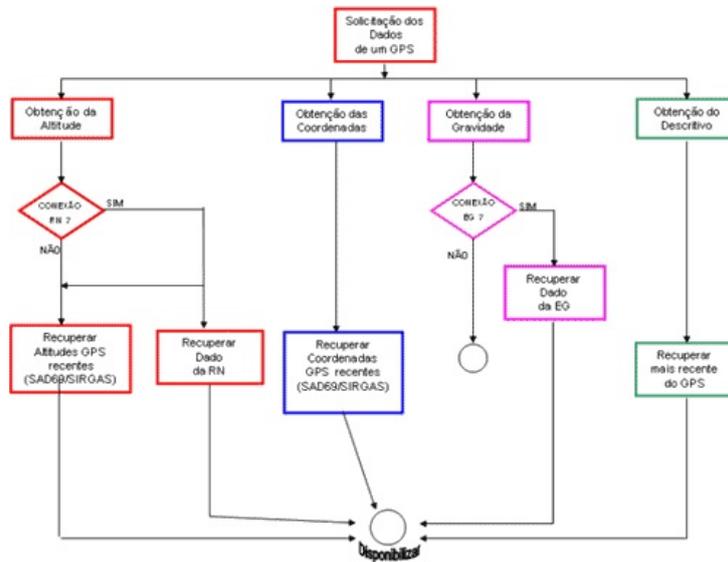


Figura 4.1: Visão de consulta a uma estação GPS [16]

Podemos afirmar que toda a tecnologia composta em um banco de dados impactou diretamente o aumento do uso dos computadores e dispositivos eletrônicos como celulares e tablets. E não somente influenciou no aumento da utilização dos equipamentos como também desenvolveu um papel importante na maioria das áreas em que a tecnologia apresenta-se como aliada, como por exemplo negócios, comércio eletrônico, engenharias, direito, educação, medicina, genética e muitos outros (Veja um exemplo na Figura 4.2 onde é apresentado um banco de dados genético para identificação criminal).

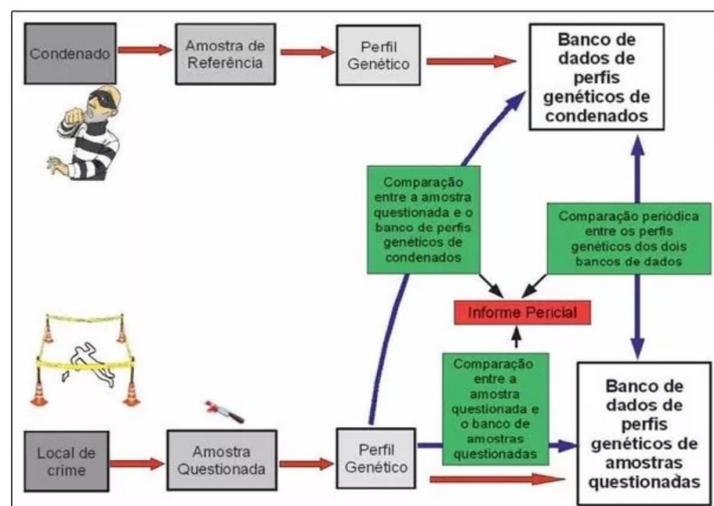


Figura 4.2: Banco de dados genético, relacionado a identificação criminal [17].

4.2 Conceitos básicos de Banco de Dados

Podemos definir um banco de dados como um acervo de dados que se relacionam. Dados referem-se a fatos que podem ser registrados, como textos, números e mídias. Quando falamos que os dados se relacionam, nos referimos ao fato que os dados possuem um significado implícito e têm certos tipos de vinculação entre eles. Por exemplo, uma coleção de dados relacionados à matrícula dos cursos de graduação de uma universidade, onde se encontram registros das matérias, professores e turmas existentes. Nesse caso, podemos declarar que esse conjunto é um banco de dados [129].

Com isso podemos dizer que um banco de dados é uma coleção logicamente coerente de dados com uma significância própria. Logo, podemos afirmar que uma variedade de dados de forma aleatória não podem ser considerado um banco de dados. É observável que a definição acima é genérica e pode haver confusões de interpretação, visto que usando a definição em seu sentido literal podemos considerar que esta Subseção seria um banco de dados, visto que é constituída de palavras que se relacionam. Para não ocorrer possíveis erros de significação, foi criado para o termo Banco de Dados algumas propriedades implícitas que tornam a nossa definição muito mais restrita [129].

Um banco de dados é projetado, construído e populado com dados para uma finalidade específica. Ele possui um grupo definido de usuários e algumas aplicações previamente concebidas nas quais esses usuários estão interessados. O que significa que um banco de dados tem alguma fonte da qual o dado é derivado, algum grau de interação com eventos no mundo real e um público que está ativamente interessado em seu conteúdo. Os usuários finais de um banco de dados podem realizar transações comerciais ou eventos de organização (um banco de dados de informações dos funcionários de uma determinada empresa), fazendo que a informação no banco de dados mude. Para que um banco de dados seja preciso e confiável o tempo todo, ele precisa ser um reflexo verdadeiro do minimundo que representa; portanto, as mudanças precisam ser refletidas no banco de dados o mais breve possível [130].

As vantagens de um banco de dados nos permite ter qualquer tamanho e complexidade. Ele pode ser de forma manual ou de forma computadorizada. De forma manual ele pode ser criado e mantido armazenando fisicamente, é o caso dos arquivos físicos de um consultório que mantem os dados armazenados em papéis. Já no caso de um banco de dados computadorizados podemos criar um grupo de programas que gerencia esse banco de dados, chamado também de sistema gerenciador de banco de dados [131, 132].

Um sistema gerenciador de banco de dados (SGBD — Database Management System) é uma conjunto de programas que permite aos usuários criar e manter um banco de dados. O SGBD é um sistema de software de uso geral que facilita o processo de definição, construção, manipulação e compartilhamento de bancos de dados entre diversos usuários

e aplicações. Definir um banco de dados envolve especificar os tipos, estruturas e restrições dos dados a serem armazenados [129, 133].

A definição ou informação descritiva do banco de dados também é armazenada pelo SGBD na forma de um catálogo ou dicionário, chamado de metadados. A construção do banco de dados é o processo de armazenar os dados em algum meio controlado pelo SGBD. A manipulação de um banco de dados inclui funções como consulta ao banco de dados para recuperar dados específicos, atualização do banco de dados para refletir mudanças no mundo e geração de relatórios com base nos dados. O compartilhamento de um banco de dados permite que diversos usuários e programas acessem-no simultaneamente [131, 132, 129, 133].

Um programa de aplicação acessa o banco de dados ao enviar consultas ou solicitações de dados ao SGBD.



Figura 4.3: Sistema gerenciador de banco de dados [18].

Outras funções importantes fornecidas pelo SGBD incluem proteção do banco de dados e sua manutenção por um longo período. A proteção inclui proteção do sistema contra defeitos (ou falhas) de hardware ou software e proteção de segurança contra acesso não autorizado ou malicioso. Um banco de dados grande pode ter um ciclo de vida de muitos anos, de modo que o SGBD precisa ser capaz de manter o sistema, permitindo

que ele evolua à medida que os requisitos mudam com o tempo. Não é absolutamente necessário utilizar software de SGBD de uso geral para implementar um banco de dados computadorizado. Poderíamos escrever nosso próprio conjunto de programas para criar e manter o banco de dados, com efeito criando nosso próprio software de SGBD de uso especial. Em ambos os casos (se usarmos um SGBD de uso geral ou não), em geral temos de implementar uma quantidade considerável de software complexo. É fato que a maioria dos SGBDs é constituída de sistemas de software muito complexos [131, 132, 129, 133].

Há muitas características que distinguem a abordagem de um banco de dados da abordagem muito mais antiga de programação com arquivos. No processamento de arquivo tradicional, cada usuário define e implementa os arquivos necessários para uma aplicação de software específica como parte da programação da aplicação [129].

Por exemplo, um usuário, o departamento de registro acadêmico, pode manter arquivos sobre os alunos e suas notas. Os programas para imprimir o histórico escolar de um aluno e inserir novas notas são implementados como parte da aplicação. Um segundo usuário, o departamento de finanças, pode registrar as mensalidades dos alunos e seus pagamentos. Embora ambos os usuários estejam interessados em dados sobre alunos, cada um mantém arquivos separados — e programas para manipular esses arquivos, pois cada usuário requer dados não disponíveis nos arquivos do outro. Essa redundância na definição e no armazenamento de dados resulta em desperdício no espaço de armazenamento e em esforços redundantes para manter os dados comuns atualizados [129].

Na abordagem de banco de dados, um único repositório mantém dados que são definidos uma vez e depois acessados por vários usuários. Nos sistemas de arquivo, cada aplicação é livre para nomear os elementos de dados independentemente. Ao contrário, em um banco de dados, os nomes ou rótulos de dados são definidos uma vez, e usados repetidamente por consultas, transações e aplicações [134, 135, 136].

As principais características da abordagem de banco de dados versus a abordagem de processamento de arquivo são mostradas de forma resumida na Tabela 4.1.

Tabela 4.1: Principais Características Banco de Dados x Processamento de Arquivos
Natureza de autodescrição de um sistema de banco de dados
Isolamento entre programas e dados, e abstração de dados
Suporte de múltiplas visões dos dados
Compartilhamento de dados e processamento de transação multiusuário

Todas essas características são importantes para distinguir um SGBD de um software tradicional de processamento de arquivos [129]. Quando falamos do SGBD podemos enumerar variadas vantagens ao utilizar um sistema tão eficaz:

- Controle de redundância de dados
- Restrição de acesso não autorizado
- Armazenamento persistente para objetos do programa
- Oferece estruturas de armazenamento e técnicas de pesquisa para o processamento eficiente de consulta.
- Oferece backup e recuperação.
- Oferece múltiplas interfaces do usuário
- Representa relacionamentos complexos entre dados.
- Impõe restrições de integridade.
- Permite dedução e ações usando regras.
- Implica adicionais do uso da abordagem de banco de dados.

Apesar das vantagens de usar um SGBD, existem algumas situações em que esse sistema pode envolver custos adicionais desnecessários, que não aconteceriam no processamento de arquivos tradicional [129]. Os custos adicionais do uso de um SGBD devem-se aos seguintes fatores:

- Alto investimento inicial em hardware, software e treinamento;
- A generalidade que um SGBD oferece para a definição e o processamento de dados;
- Esforço adicional para oferecer funções de segurança, controle de concorrência, recuperação e integridade. Portanto, pode ser mais desejável usar arquivos comuns sob as seguintes circunstâncias:
- Aplicações de banco de dados simples e bem definidas, para as quais não se espera muitas mudanças;
- Requisitos rigorosos, de tempo real, para alguns programas de aplicação, que podem não ser atendidos devido as operações extras executadas pelo SGBD;
- Sistemas embarcados com capacidade de armazenamento limitada, onde um SGBD de uso geral não seria apropriado;

- Nenhum acesso de múltiplos usuários aos dados.

Em alguns setores e aplicações utilizam sistemas mais adequados para finalidades específicas como por exemplo no caso de implementações dos sistemas de informações geográficas(SIG).

4.3 Modelo Conceitual

A finalidade do modelo conceitual de dados é capturar os requisitos de informação e regras do produto sob o ponto de vista do produto. Para isto, torna-se necessário o entendimento e a correta aplicação dos mecanismos de abstração, utilizados na modelagem conceitual de dados. Para realizar a criação do modelo entidade relacionamento é necessário primeiro exemplificar os mecanismos de abstração utilizados na modelagem conceitual dos dados [19].

Entidades

As entidades são representadas através de um retângulo com o nome da entidade escrito em seu centro conforme Figura 4.4:

- Entidades Fortes: possuem um alto grau de independência de existência de identificação. Geralmente, outras entidades podem depender dela para serem identificadas. Podemos tomar como exemplo a entidade “BANCO”, onde a existência da mesma não depende de nenhuma outra entidade para ser identificada [137];
- Entidades Fracas: possuem dependência de existência e/ou identificação. São sempre ligadas a outras tabelas através de relacionamentos. Podemos tomar como exemplo a entidade “AGENCIA”, onde a existência e identificação da mesma estão vinculadas a outra entidade forte, no caso o “BANCO” [128].

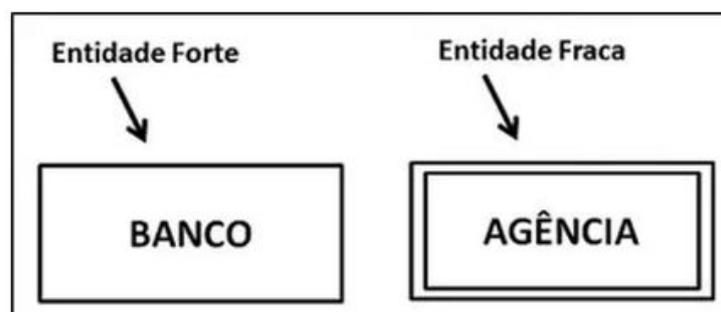


Figura 4.4: Exemplos de Entidades [19].

Relacionamentos

Relacionamentos são associações entre entidades com um significado específico dentro do mundo real. Os objetos do mundo real não ocorrem de forma isolada, eles se associam ou se vinculam. A figura de um relacionamento é representada através de um losango, tal como as entidades os relacionamentos são classificados em fortes ou fracos. Tal como as entidades, os relacionamentos também possuem nome e devem expressar o real significado dentro do contexto modelado. A Figura 4.5 a seguir mostra como os relacionamentos são representados em um modelo conceitual de dados [138, 139].

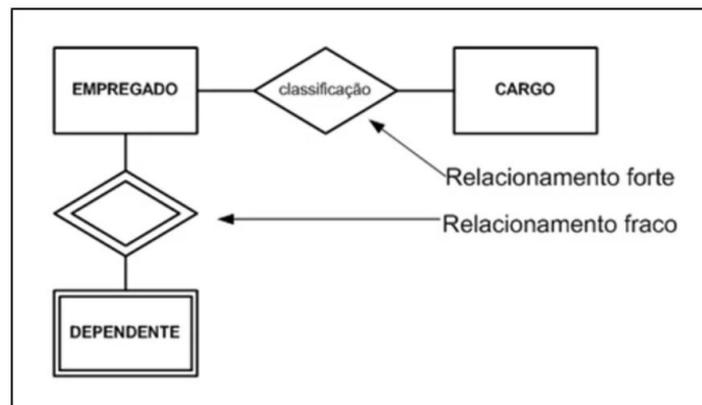


Figura 4.5: Exemplos de Relacionamento [19].

Na Figura 4.5 DEPENDENTE é uma entidade fraca em relação ao EMPREGADO, sempre que esta relação existir de forma fraca, o relacionamento também será fraco, por esta razão o losango desta relação está representado com uma linha dupla. Já na relação entre EMPREGADO e CARGO não há dependência de existência ou identificação, pois um CARGO não depende de um EMPREGADO para existir e ser identificado e vice-versa. Quando tratamos de relacionamentos, devemos ter em mente três conceitos importantes que influenciam diretamente na modelagem e entendimento de um modelo conceitual. Os conceitos são o grau, cardinalidade e tipo do relacionamento [19].

Grau dos Relacionamentos

O grau de um relacionamento corresponde ao número de entidades envolvidas na mesma relação (veja Figura 4.6) [138, 129, 140]. O grau de um relacionamento pode ser:

- Binário: Onde duas entidades participam de um relacionamento. Este é o grau utilizado na maioria dos relacionamentos.
- Ternário: Onde três entidades participam de um relacionamento. Muito se discute sobre o uso e aplicabilidade de relacionamentos com grau maior que dois (terná-

rios e n-ários) em modelos de dados. Alguns autores sugerem inclusive que esses relacionamentos não sejam adotados.

- N-ário: Onde quatro ou mais entidades participam de um relacionamento.

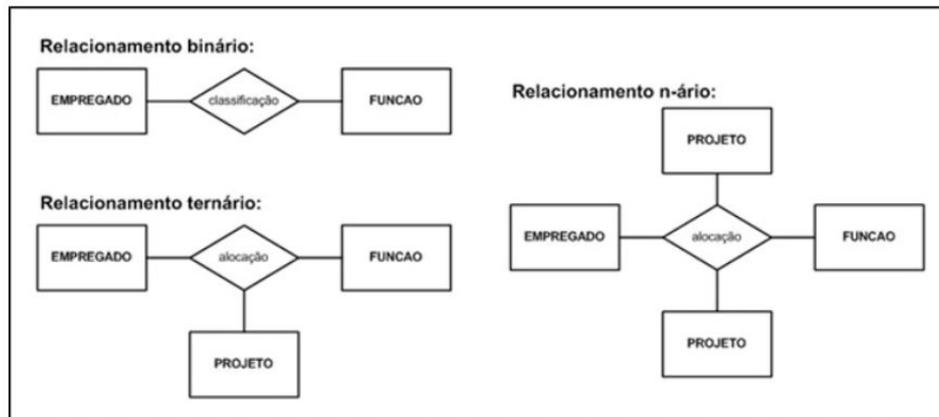


Figura 4.6: Exemplos de grau dos relacionamentos [19].

Cardinalidade

A cardinalidade representa a quantidade de vezes que um elemento de um conjunto de entidades pode, em um determinado instante, estar associado em um dado relacionamento, a outros elementos de outras entidades. A cardinalidade de uma relação é definida em cada um dos sentidos do relacionamento por um conjunto (x,y) onde x representa a cardinalidade mínima e y representa a cardinalidade máxima. A cardinalidade mínima é responsável por orientar a obrigatoriedade (opcionalidade) do relacionamento. Já a cardinalidade máxima é responsável por definir a quantidade máxima de vezes que um elemento pode estar associado no relacionamento [141, 142, 143].

Na Figura 4.7, um EMPREGADO trabalha em uma e somente uma EMPRESA e, em uma EMPRESA trabalham nenhum ou vários EMPREGADOS. Ou seja, dentro do contexto que foi modelado, é impossível existir um EMPREGADO sem uma EMPRESA associada, porém é totalmente viável criar uma EMPRESA e não associar inicialmente algum EMPREGADO.

Tipos de Relacionamentos

O simples fato de associar duas entidades através de um relacionamento com suas cardinalidades às vezes não são suficientes para representar todas as regras de negócio existentes dentro dessas relações [144, 145, 146]. Para isto, podemos usar mecanismos de

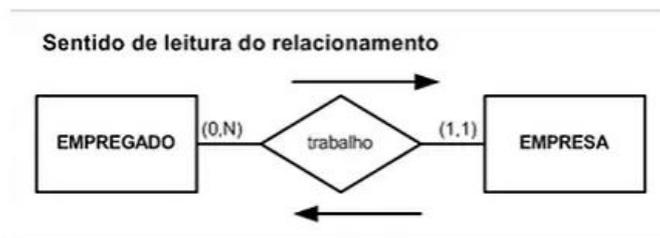


Figura 4.7: Exemplo de cardinalidades [19].

representação um pouco mais detalhados. Sob esta ótica, podemos ainda classificar os relacionamentos em três tipos:

- **Relacionamentos Independentes:** Tipo de relacionamento presente na maioria das relações. Não há necessidade de interpretação simultânea de outro relacionamento. Ou seja, é independente, não depende de ninguém para existir ou influenciar o seu comportamento.
- **Relacionamentos Contingentes:** Estabelecem associações simultâneas entre os elementos envolvidos. Ou seja, mais de um relacionamento deve ocorrer em um mesmo instante. Sua representação é recomendada, pois envolvem regras de negócio específicas que se não mapeadas neste momento, fatalmente serão esquecidas mais adiante no decorrer do projeto.
- **Relacionamentos Mutuamente Exclusivos:** Estabelecem associações onde, se um relacionamento ocorre, os outros não deverão ocorrer em relação a um determinado objeto. Sua representação é recomendada, pois envolvem regras de negócio específicas que se não mapeadas neste momento, fatalmente serão esquecidas mais adiante no decorrer do projeto.

Atributos

Os atributos são informações que caracterizam as entidades e os relacionamentos. Um atributo pode: identificar, descrever, qualificar, quantificar ou registrar o estado (ou a situação ou a ocorrência) de uma entidade [147]. No modelo conceitual são representados através de “bolinhas com uma aste” colocados juntos as entidades. Os atributos podem ser classificados em quatro tipos:

- **Atributo identificador:** Representado através de uma bola cheia na extremidade do atributo. Atributos identificadores identificam ou compõe a identificação única de uma ocorrência em uma entidade. Vale ressaltar que uma entidade e/ou relacio-

namento pode possuir mais de um atributo identificador, desde que os mesmos em conjunto componham a identificação única.

- **Atributo não identificado:** Representado através de uma bola vazia na extremidade do atributo. Corresponde a maioria das ocorrências de uma entidade. Além disso, atributos não identificados podem ser opcionais, ou seja, em algumas instâncias de entidade, alguns atributos poderão conter valores nulos.
- **Atributos multivalorados:** Representado através de uma flor ou asterisco na extremidade do atributo. Atributos multivalorados são utilizados para representar mais de uma ocorrência de valor de um atributo dentro de uma mesma instância de uma entidade. Exemplo: Geralmente, uma pessoa possui mais de um número de telefone. Como o objetivo do modelo conceitual é capturar a essência do negócio sem levar em conta aspectos de implementação, este tipo de abordagem é utilizado para representar todas essas instâncias em um único atributo, porém deve-se ter em mente que este tipo de abordagem não deve ser utilizado a partir da modelagem lógica de dados, onde entrarão em cena os conceitos de normalização.
- **Atributos compostos:** Representados através de uma oval com vários nós na extremidade do atributo. Atributos compostos são utilizados para representar mais de um tipo de informação (qualificação) em um atributo. Tal como o atributo multivalorado, seu uso é recomendado somente no modelo conceitual de dados.

A Figura 4.8 mostra os tipos de atributos utilizados em um modelo conceitual de dados. O exemplo mostra atributos comuns aos alunos de qualquer instituição. O número da matrícula é um atributo identificador, enquanto o nome do aluno é um atributo não identificador. Já o atributo telefones é um atributo multivalorado, onde representa os diversos telefones que um aluno possui. Endereço é considerado um atributo composto, pois é formado pela composição da UF, Cidade e Logradouro [19].

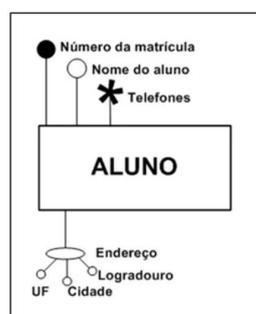


Figura 4.8: Exemplo de atributos [19].

Capítulo 5

Descrição do projeto

Neste capítulo, apresentamos os métodos utilizados para a criação do banco de dados Perci. Na Seção 5.1, faremos uma breve descrição dos dados primários, junto com os locais em que foram coletados, e dos filtros que foram utilizados para uniformizar esses dados. Na Seção 5.2, apresentaremos o modelo conceitual do projeto, tanto o modelo ER quanto o Modelo Relacional. Na seção 5.3, descreveremos a página *web*. Por fim, na Seção 5.4, detalharemos como foi realizada a conexão da página *web* com banco de dados Perci.

5.1 Dados

De acordo com as buscas realizadas na literatura, os principais bancos de dados com as informações moleculares de CRC e ncRNAs são:

1. LncRNADisease (<http://www.cuilab.cn/lncrnadisease/>);
2. Lnc2Cancer (<http://www.bio-bigdata.com/lnc2cancer/>);
3. MirCancer (<http://mircancer.ecu.edu/>).

Filtro dos Dados

Os dados foram encontrados em dois formatos: em tabelas; e em formato de texto. Para uniformizar os dados, foi feita inicialmente uma transcrição dos dados encontrados nos repositórios (indicados acima) para tabelas Excel. Depois, foi realizada buscas dos principais dados usando três categorias, respectivamente, *ncRNA Category*, *Species* e *Disease Name* foi padronizado as colunas das tabelas para utilização no banco de dados (Veja a Tabela 5.1). Por fim, foi contabilizado o número total de dados, para verificar se de fato a criação de um banco de dados seria justificada (Veja a Figura 5.1 e a Tabela 5.2).

Tabela 5.1: Descrição dos dados obtidos nos repositórios.

Tipo de dado coletado	Descrição
Tipos de RNAs	lncRNAs, circRNAs, miRNAs
Nomenclatura	Nomenclatura do gene
Tipos de câncer	Variante do câncer colorretal
Métodos de Validação	Laboratório ou <i>in Silico</i>
Amostra	Especificação do tecido ou da região do organismo
Padrão de disfunção	Qual foi a disfunção do gene
Informação associada	Detalhamento do gene
Pubmed ID	Referência da publicação do artigo no Pubmed
Sequência genômica	Sequência do gene

LncRNA

```
[False, False, True, True, True, True, False]
Tipo de Cancer
COLON ADENOCARCINOMA      16
COLON CANCER              930
COLON CARCINOMA           23
COLONIC NEOPLASMS         6
COLORECTAL ADENOCARCINOMA 22
COLORECTAL CANCER        13193
COLORECTAL CARCINOMA      11
COLORECTAL NEOPLASIA     460
COLORECTAL NEOPLASMS      3
HEPATIC COLORECTAL METASTASIS 462
dtype: int64
```

CircRNA

```
[False, False, True, True, True, True, False]
Tipo de Cancer
COLON CANCER      68
COLORECTAL CANCER 338
dtype: int64
```

MiRNA

```
[False, False, False, True]
Tipo de Cancer
COLON CANCER      152
COLON CARCINOMA   6
COLONIC ADENOCARCINOMA 6
COLORECTAL ADENOCARCINOMA 2
COLORECTAL CANCER 713
COLORECTAL CARCINOMA 25
dtype: int64
```

Figura 5.1: Número de dados coletados por tipo de CRC.

Tabela 5.2: Número de dados por tipos de ncRNA.

Tipos de ncRNA	Quantidade
lncRNA	15.532
circRNA	406
miRNA	904
Total	16.436

Variantes do câncer colorretal

Inicialmente, foi realizada uma revisão bibliográfica sobre os tipos de câncer colorretal, e solicitado o apoio do Prof. João Batista de Sousa, para realizar a classificação das 10 variantes do câncer colorretal encontradas. Foram criados, então, cinco subtipos de CRC, conforme demonstrado na Figura 5.2.

10 subtipos de variantes encontradas		5 principais variantes do câncer colorretal		
Nº	Nome das variantes encontradas	Nome das variantes encontradas	Nº	
1	Colon Cancer	Colon Cancer	1	
2	Colorectal Cancer	Colon Carcinoma	7	
3	Colonic Neoplasms	Colorectal Cancer	2	
4	Colorectal Carcinoma	Colorectal Carcinoma	4	
5	Colon Adenocarcinoma	Colorectal Adenocarcinoma	8	
6	Colorectal Neoplasms	Adenocarcinoma	5	
7	Colon Carcinoma	Colon Adenocarcinoma	10	
8	Colorectal Adenocarcinoma	Neoplasms	3	
9	Hepatic Colorectal Metastasis	Colonic Neoplasms	6	
10	Colonic Adenocarcinoma	Colorectal Neoplasms	6	
		Hepatic Colorectal Metastasis	9	

Figura 5.2: Existem diferentes denominações para um mesmo subtipo de CRC. Assim, os 10 subtipos das fontes primárias foram transformados em 5 subtipos para o banco de dados Perci.

Nome e logo do banco de dados

Por fim, apresentamos o nome e o logo do repositório criado neste projeto, como mostrado na Figura 5.3.



Figura 5.3: Logo do banco de dados Perci.

5.2 Modelo conceitual Perci

A implementação de um banco de dados integrando os dados coletados necessita inicialmente da modelagem dos dados. A base conceitual é fundamental para execução, funcionamento e manutenção do banco de dados.

5.2.1 Modelo ER Perci

Com a análise dos dados foi possível decidir que a entidade principal é o Gene (seja de lncRNA/circRNA ou miRNA). A partir daí, criamos o modelo ER com seus relacionamentos, atributos e cardinalidades, analisando os dados coletados durante a revisão bibliográfica.

As informações que não eram relevantes para buscas por parte do usuário foram modeladas como atributos da entidade principal Gene. Já as informações que podem ser usadas como buscas foram modeladas como entidades, com as informações associadas à respectiva entidade principal.

Para este projeto, foram necessários dois modelos ER, um para lncRNA e circRNA e outro para miRNA, conforme mostrado nas Figuras 5.4 e 5.5.

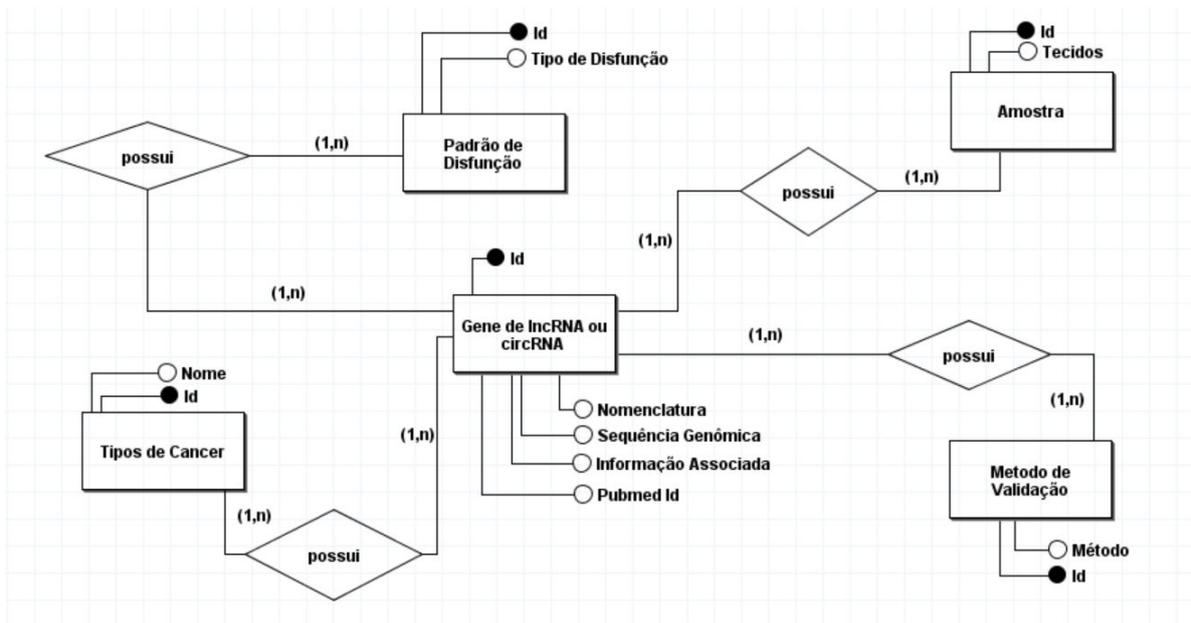


Figura 5.4: Modelo ER de lncRNAs e circRNAs para o repositório de dados Perci formalizado pela plataforma BRModelo [20].

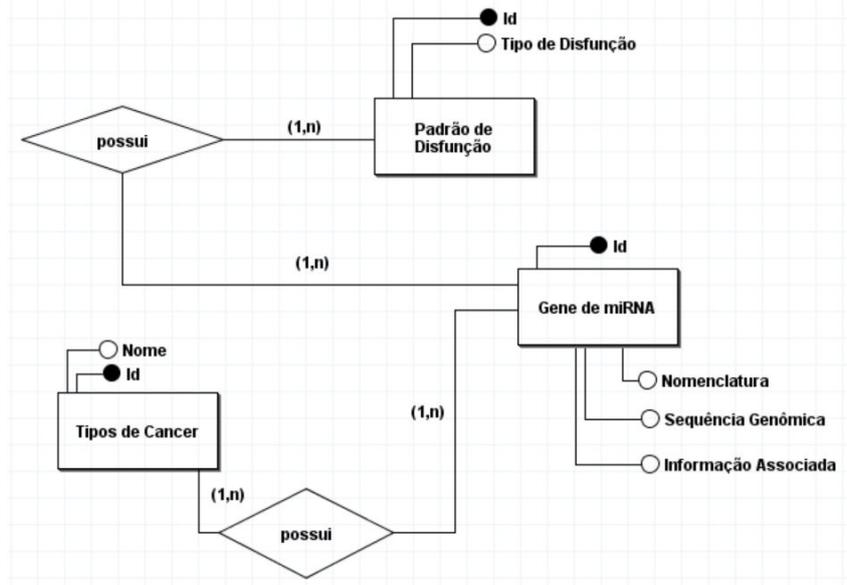


Figura 5.5: Modelo ER de miRNAs para o repositório de dados Perci formalizado pela plataforma BRModelo [20].

5.2.2 Modelo relacional Perci

As tabelas do modelo relacional (veja as Figuras 5.6 e 5.7) foram geradas a partir do modelo ER, de forma automática, pela mesma plataforma utilizada para formalizar o modelo ER (plataforma BRModelo [20]). Estas tabelas são utilizadas para criar a estrutura computacional (tabelas) do banco de dados a ser implementado, diferentemente da abstração do modelo ER.

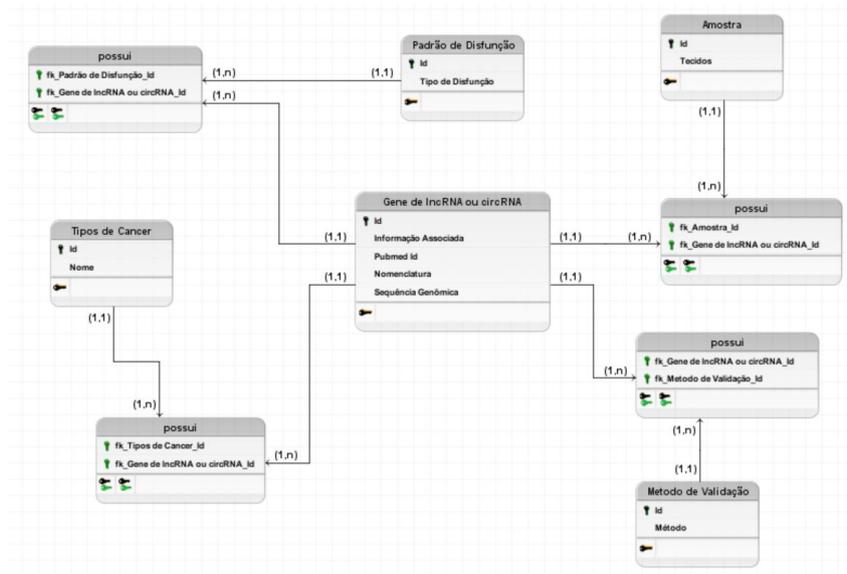


Figura 5.6: Modelo relacional do lncRNAs e circRNAs do projeto Perci.

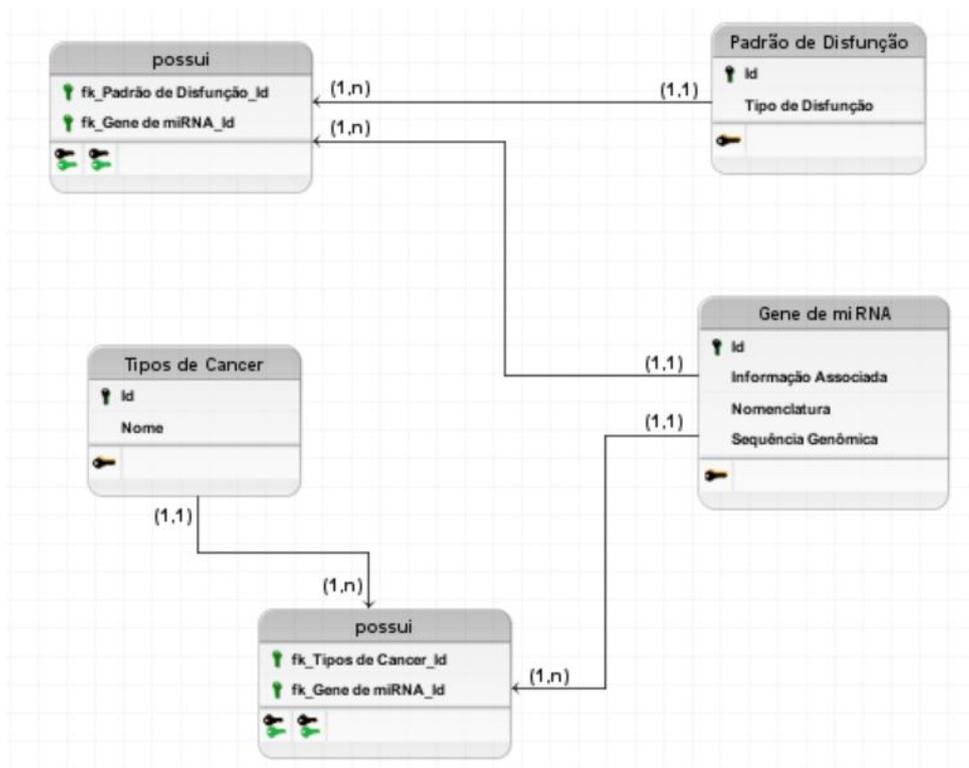


Figura 5.7: Modelo relacional do miRNAs do projeto Perci.

5.3 Implementação do banco de dados Perci

A implementação do banco de dados foi realizada em *mySQL* (veja Figura 5.8), a partir das informações coletadas, com base nos modelos ER e respectivos modelos relacionais, descritos nas seções anteriores.

```

1  ## Criando o banco de dados
2  create database dadosExtracao;
3
4  use dadosExtracao;
5
6  ## Criando a tabela
7  CREATE TABLE `rna` (
8    `id` int NOT NULL AUTO_INCREMENT,
9    `tipo_rna` varchar(25) DEFAULT NULL,
10   `nomenclatura` varchar(50) DEFAULT NULL,
11   `tipo_cancer` varchar(70) DEFAULT NULL,
12   `metodo_validacao` varchar(255) DEFAULT NULL,
13   `amostra` varchar(255) DEFAULT NULL,
14   `padrao_disfuncao` varchar(255) DEFAULT NULL,
15   `informacao_associada` varchar(255) DEFAULT NULL,
16   `pubmed_id` varchar(20) DEFAULT NULL,
17   `sequencia_genomica` varchar(255) DEFAULT NULL,
18   PRIMARY KEY (`id`)
19 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci
20
21
22 ## PERGUNTA 1
23 select DISTINCT * from rna WHERE tipo_rna = 'lncRNAs' and tipo_cancer = 'colon adenocarcinoma';
24
25 ## PERGUNTA 2
26 select DISTINCT * from rna WHERE (tipo_rna = 'lncRNAs' or tipo_rna = 'mirRNAs') and tipo_cancer = 'colon cancer';
27
28 ## PERGUNTA 3
29 select DISTINCT * from rna where nomenclatura = 'hsa_circRNA_104700';
30
31
32 #SCRIPT COM TODOS INSERTS
33 INSERT INTO dadosExtracao.rna (tipo_rna,nomenclatura,tipo_cancer,metodo_validacao,amostra,padrao_disfuncao,informacao_associada,pubmed_id,sequenci
34 ('lncRNAs','AOC4P','colon cancer','ChIP/Immunoblot/RIP/RNAi/qRT-PCR','N/A','Expression','UPAT and UHRF1 may be promising molecular target
35 ('lncRNAs','BACE1-AS','colon cancer','qPCR','cell line (SNU-C4R, SNU-C5R)','Regulation [down-regulated]','We selected three lncRNAs,snaR,BACE
36 ('lncRNAs','BCAR4','colon cancer','qRT-PCR','colon cancer tissues and paired normal tissues','Regulation [up-regulated]','In our research, we
37 ('lncRNAs','BCYRN1','colon cancer','qRT-PCR','colon cancer tissues and adjacent non-cancerous tissues','Regulation [up-regulated]','Here, we
38 ('lncRNAs','CCAT1','colon cancer','TSH/qPCR','colonic tissue','Regulation [up-regulated]','CCAT1 is up-regulated across the colon adenoma-ca

```

Figura 5.8: O código do banco de dados criado.

No banco de dados Perci, será possível realizar consultas do tipo:

- Por tipo de ncRNA: pesquisa sobre quais dados estão relacionados ao tipo específico de RNA selecionado;
- Por tipo de câncer: pesquisa sobre quais dados estão relacionados ao subtipo do câncer selecionado;
- Por gene: pesquisa sobre quais dados estão relacionados ao gene pesquisado.

O usuário poderá escolher diferentes combinações de acesso aos dados, entre as combinações o banco de dados possui 10 principais caminhos de busca no banco de dados, que podem ser conferidos de forma manual através do código do banco de dados conforme Figura 5.9.

```
### CAMINHO 1: Pesquisar/Por tipo de ncRNA/ lncRNAs
select DISTINCT* from rna WHERE tipo_rna = 'lncRNAs';

### CAMINHO 2: Pesquisar/Por tipo de câncer/Câncer de cólon
select DISTINCT * from rna WHERE tipo_cancer = 'colon cancer';

### CAMINHO 3: Pesquisar/Por tipo de gene
select DISTINCT * from rna where nomenclatura = '';

### CAMINHO 4: Pesquisar/Por tipo de ncRNA/ circRNAs
select DISTINCT* from rna WHERE tipo_rna = 'circRNAs';

### CAMINHO 5: Pesquisar/Por tipo de ncRNA/ miRNAs
select DISTINCT* from rna WHERE tipo_rna = 'miRNAs';

### CAMINHO 6: Pesquisar/Por tipo de ncRNA/ Todos os RNAs
select DISTINCT* from rna WHERE tipo_rna = '';

### CAMINHO 7: Pesquisar/Por tipo de câncer/Câncer Colorretal
select DISTINCT * from rna WHERE tipo_cancer = 'colorectal cancer';

### CAMINHO 8: Pesquisar/Por tipo de câncer/Neoplasmas
select DISTINCT * from rna WHERE tipo_cancer = 'neoplasms';

### CAMINHO 9: Pesquisar/Por tipo de câncer/Adenocarcinoma
select DISTINCT * from rna WHERE tipo_cancer = 'adenocarcinoma';

### CAMINHO 10: Pesquisar/Por tipo de câncer/Metástase Colorretal Hepática
select DISTINCT * from rna WHERE tipo_cancer = 'hepatic colorectal metastasis';
```

Figura 5.9: Principais buscas do banco de dados Perci

A página *web* ainda possibilitará fazer perguntas, conforme descrito na Seção 6.2, do tipo:

- "Quais são os lncRNAs envolvidos no adenocarcinoma?"
- "Quais são os lncRNAs e os miRNAs envolvidos no câncer de cólon?"
- "Dado um certo circRNA (hsa_circRNA_104700), quais são os subtipos de câncer que têm esse circRNA envolvido?"

Observamos que o usuário poderá acessar o banco de dados Perci por meio de métodos computacionais, como terminais de comando do sistema operacional. Neste caso, o banco de dados retornará para o usuário tabelas com as informações desejadas, como mostrado Figura 5.10.

id	tipo_rna	nomenclatura	tipo_cancer	metodo_validacao	amostra	padrao_disfuncao	informacao_associada	pubmed_id
1	lncRNAs	AOC4P	colon cancer	ChIP/immunoblot/RIP	N/A	Expression	UPAT and UHRF1 may be promising molecu	26768845
2	lncRNAs	BACE1-AS	colon cancer	qPCR	cell line (SNU-C4R, SNU-C5R)	Regulation [down-regulated]	We selected three lncRNAs, miR-BACE1-AS	25078450
3	lncRNAs	BCYRN1	colon cancer	qRT-PCR	colon cancer tissues and paired normal tissue	Regulation [up-regulated]	In our research, we found that the expressio	29190958
4	lncRNAs	CCAT1	colon cancer	ISH/qPCR	colonic tissue	Regulation [up-regulated]	Here, we found that expression of BC200 w	29625226
5	lncRNAs	CCAT1	colon cancer	ISH/qPCR	colonic tissue	Regulation [up-regulated]	CCAT1 is up-regulated across the colon aden	23594791
6	lncRNAs	CCAT1	colon cancer	ISH/qPCR	GC tissue, cell line (MGC-803, SGC-7901, AGS)	Regulation [up-regulated]	we provide the first evidence that CCAT1 reg	26825578
7	lncRNAs	CCAT1	colon cancer	ISH/qPCR	N/A	Regulation [up-regulated]	Deregulation of CCAT1 (colon cancer-associa	27134049
8	lncRNAs	CCAT2	colon cancer	MTT assay/Flow Cytom	colon cancer tissues	Regulation [up-regulated]	Results showed that CCAT1 and miR-410 wer	29190961
9	lncRNAs	CCAT2	colon cancer	qPCR	N/A	Regulation	CCAT2, a novel noncoding RNA mapping to 8i	23796952
10	lncRNAs	CCAT2	colon cancer	qPCR	colon tissue	Regulation [up-regulated]	Our results revealed that CCAT1 was significa	25185650
11	lncRNAs	CYTOR	colon cancer	Luciferase reporter gen	N/A	Expression	Linc00152 as a candidate prognostic indicat	27633443
12	lncRNAs	DACOR1	colon cancer	qPCR/RIP/Western blot	colon tissue, cell line (HCT116)	Regulation [down-regulated]	DACOR1 shows high,tissue-specific expressi	26307088
13	lncRNAs	FER1L4	colon cancer	qPCR	N/A	Expression	It is predictive of poor prognosis in colon ca	26224446
14	lncRNAs	GSEC	colon cancer	Luciferase reporter gen	N/A	Regulation	GSEC plays an important role in colon cancer	27797375
15	lncRNAs	H19	colon cancer	ISH/Luciferase reporte	N/A	Expression [differentially exp	DTA was preferentially expressed in liver me	15531051
16	lncRNAs	H19	colon cancer	ISH/Luciferase reporte	CRC tissue	Regulation [up-regulated]	By quantitative reverse transcription-polyme	19936638
17	lncRNAs	H19	colon cancer	ISH/Luciferase reporte	N/A	N/A	Regional therapy with DTA-H19 vector suppr	21874233
18	lncRNAs	H19	colon cancer	ISH/Luciferase reporte	N/A	Expression [differentially exp	Control of imprinting. Containing miRNA miR	22963375
19	lncRNAs	H19	colon cancer	RT-PCR/Northern blot	colon cancer tissues and cells	Regulation [up-regulated]	Interestingly, the lncH19 RNA was up-regulat	27981548
20	lncRNAs	H19	colon cancer	qRT-PCR/MTT assay	colon cancer tissue, cell lines	Regulation [up-regulated]	Together, these results suggested that H19 o	28189050
21	lncRNAs	H19	colon cancer	qPCR	N/A	Interaction [miR-675-5p]	The results suggested that VDR signaling was	28189050
22	lncRNAs	H19	colon cancer	qPCR	colon cancer tissue, cell lines	Regulation [up-regulated]	We found that H19 was overexpressed in col	28358427
23	lncRNAs	H19	colon cancer	qPCR	N/A	Interaction [miR-138]	Taken together, our research indicated an H1	28358427
24	lncRNAs	HNF1A-AS1	colon cancer	qRT-PCR	colon cancer tissues	Regulation [up-regulated]	In this study, we reported that HNF1A-AS1 w	28943452
25	lncRNAs	HOTAIR	colon cancer	Bioinformatic analysis	N/A	Expression [differentially exp	In approximately one-quarter of human brea	24667321
26	lncRNAs	HOTAIR	colon cancer	Bioinformatic analysis	colon tissue, cell line (SW480, HT29)	Regulation [up-regulated]	Increased HOTAIR expression was significant	24840737
27	lncRNAs	HOTAIR	colon cancer	Bioinformatic analysis	N/A	Regulation	MIR-125b-5p decreases after long non coding	24962687
28	lncRNAs	HOTAIR	colon cancer	Bioinformatic analysis	colon tissue	Regulation [up-regulated]	The expression of HOTAIR was significantly h	27285868
29	lncRNAs	HOXB-AS3	colon cancer	qRT-PCR	colon cancer tissues	Regulation	The HOXB-AS3 peptide, not lncRNA, suppres	18985503
30	lncRNAs	HNF1P-DBP	colon cancer	qPCR	colon tissue	Regulation [up-regulated]	In the present study, qRT-PCR was performed	30771467

Figura 5.10: Visualização de dados do banco Perci, quando o usuário acessa diretamente por meio de terminais de comando.

5.4 Descrição da página web

Descrevemos agora, de forma breve, a página web criada neste projeto.

Na página inicial, o usuário encontrará na abertura a imagem da Figura 5.11 e uma breve descrição da motivação de criar este banco de dados. Na coluna esquerda, o usuário encontrará cinco opções para navegar pelo site.

Na aba "Tipos de RNAs", é possível visualizar os dois tipos de RNAs que o banco de dados Perci possui. Já na aba "Tipos de câncer" é possível visualizar os dez subtipos de CRC, junto com o mapeamento desses dez subtipos, para cinco.

Na aba "Pesquisar", o usuário encontrará 3 opções:

1. Por tipo de RNA;
2. Por tipo de câncer;
3. Por gene.

Selecionando a opção desejada, o usuário será encaminhado conforme descrito na Seção 5.3.

Na primeira opção, o usuário será direcionado a uma página na página *web* para selecionar a escolha do RNA. Em seguida, o usuário será encaminhado para uma tabela com todos os RNAs do tipo escolhido. Caso o usuário escolha a opção "Todos os RNAs", será gerada a planilha "TOTAL" do arquivo PERCI disponibilizado.

Na segunda opção de consulta, o usuário será direcionado a uma página na página *web* para selecionar uma das cinco variações do CRC. Depois, será disponibilizado para o usuário uma tabela com todos os dados do tipo escolhido de câncer, independentemente de RNAs.

Na terceira opção de consulta, o usuário será encaminhado para uma página *web* onde já terá acesso automático ao banco de dados. O usuário encontrará uma barra de pesquisa onde ele poderá digitar o nome do gene. Notamos que o usuário não precisa digitar o nome inteiro do gene. Caso digite apenas uma parte do nome do gene, por exemplo "hsa-7", serão fornecidos todos os dados que possuem "hsa-7" como *substring* do nome. A pesquisa será feita pela coluna da tabela chamada "nomenclatura".

A aba "Download" terá as referências do banco de dados e um botão para o usuário clicar e realizar o *download* da tabela dos dados, disponibilizada em formato excel.

A criação do *site* foi realizada pela plataforma Webnode, que possui uma interface didática para o usuário acessar depois de pronta.



Figura 5.11: Página inicial do *site*.

5.5 Ligação entre a página *web* e o banco de dados Perci

Nesta seção, descrevemos como foi feita a conexão do banco de dados Perci com a página *web* criada neste projeto.

O banco de dados Perci foi disponibilizado de forma *online*, usando uma API (*Application Programming Interface*), pela plataforma da Google Cloud. Em mais detalhes, o código MySQL desenvolvido neste projeto foi importado para a plataforma da plataforma *Google Cloud*, conforme a Figura 5.12.

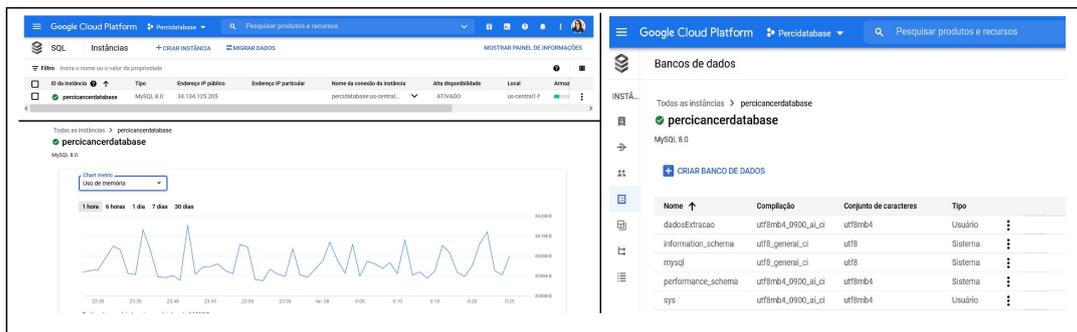


Figura 5.12: Ligação da plataforma Google Cloud com o banco de dados Perci.

Em seguida, a API foi interligada com a plataforma *Public Tableau*, uma plataforma de análise visual de dados que facilita a visualização dos dados pelos usuários. Após a conexão da plataforma *Public Tableau* com a API (veja Figura 5.13), é necessário extrair os dados de forma *online* e, em seguida, manipular e disponibilizar as informações extraídas no formato desejado (por exemplo, tabelas, gráficos ou mapas).

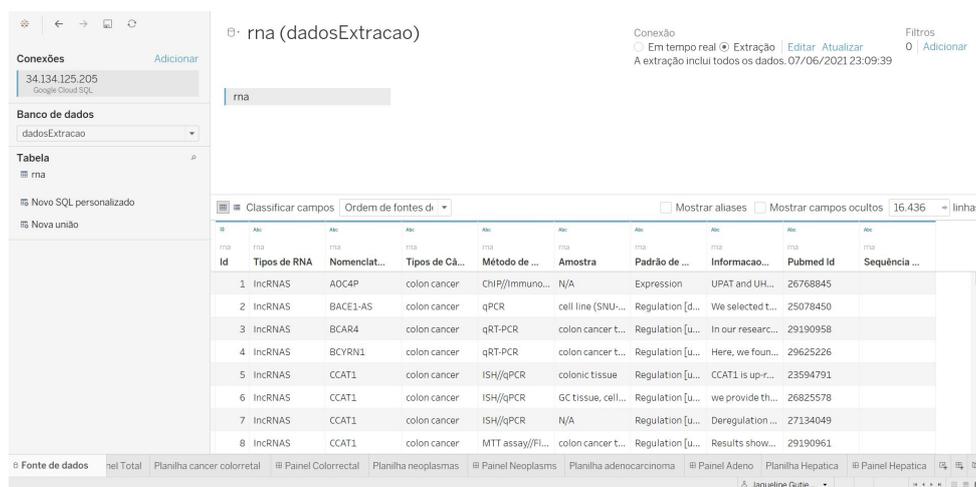


Figura 5.13: Plataforma *Google Cloud* conectada à plataforma *Public Tableau*.

Neste projeto, foi escolhido o formato "tabela" para visualizar os dados coletados. Após disponibilizar de forma *online* essas tabelas (veja Figura 5.14), realiza-se o acesso ao código *html* disponibilizado por essa plataforma para incorporar à página *web* do banco de dados Perci.

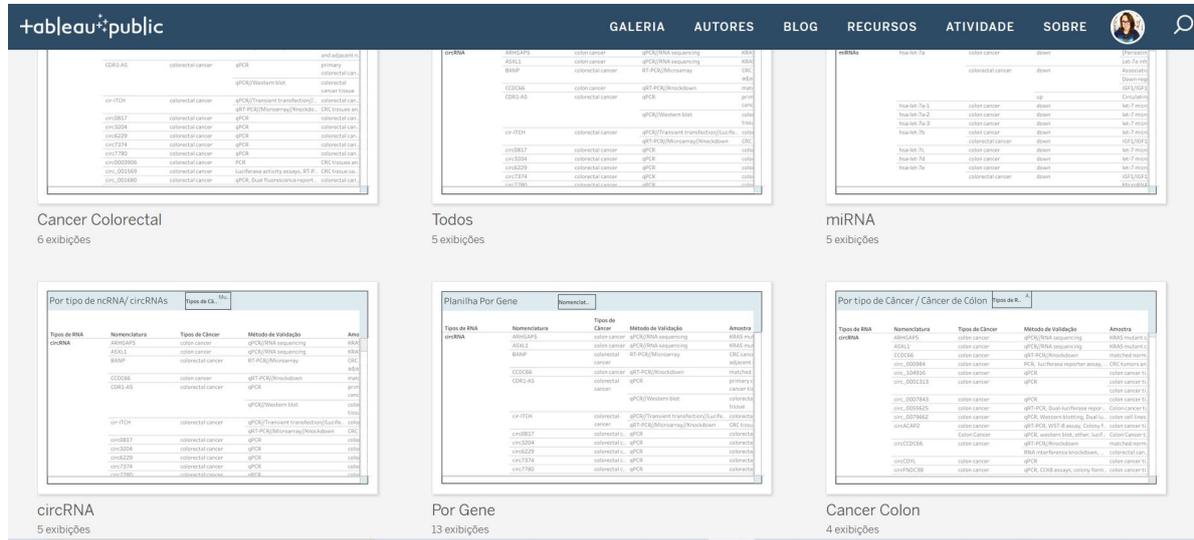


Figura 5.14: Tabelas disponibilizadas de forma *online* na plataforma *Public Tableau*.

Capítulo 6

Resultados e discussão

Neste capítulo, são apresentadas algumas informações da página web e discutidos alguns resultados de experimentos feitos para obter informações de ncRNAs ligados a subtipos de CRC. Na Seção 6.1, mostramos informações que podem ser obtidas da página *web*. Na Seção 6.2, exibimos resultados encontrados a partir de cada pesquisa realizada no banco de dados Perci para uma dada pergunta feita por um usuário.

6.1 Informações da página *web*

Nesta seção, apresentaremos informações da página *web*, a partir de cada janela de acesso. A Figura 6.1 mostra a página inicial da página *web*.

The image shows the homepage of the Perci database. On the left is a dark blue sidebar with a logo at the top, the text 'Perci' and 'Banco de Dados sobre Câncer Colorretal', and a vertical list of navigation links: 'Página inicial', 'Tipos de ncRNAs', 'Tipos de Câncer', 'Pesquisar', 'Download', and 'Contato'. The main content area has a light blue background with a photograph of pink cherry blossoms. The title 'PERCI' is displayed in large white letters, with the subtitle 'Banco de Dados sobre o câncer' below it. A 'Bem-vindo' section follows, with the subtitle 'Repositório de dados contendo RNAs não-codificadores envolvidos em câncer colorretal' and a paragraph of introductory text. To the right, there are sections for 'Autoria' (listing Jaqueline Gutierrez Coelho) and 'Colaboradores' (listing Dra Maria Emília Machado Telles and Walter).

Figura 6.1: Informações da página inicial do banco de dados Perci.

6.1.1 Tipos de ncRNAs

Nesta página do *site*¹ são apresentados os tipos de ncRNAs armazenados no banco de dados Perci apresentado na Figura 6.2, quando o usuário clica no tipo de ncRNA será encaminhado para uma breve descrição.



Figura 6.2: Tipos de ncRNAs que compõe o bando de dados Perci.

6.1.2 Tipos de câncer

Nesta página são apresentadas as variantes dos tipos de câncer colorretal, incluindo os 5 grandes conjuntos, além do detalhamento das variantes, para facilitar e direcionar a pesquisa realizada no *site*. Se o usuário desejar, poderá clicar na imagem e ela abrirá em formato maior para visualização.



Figura 6.3: Tipos de CRC com informações armazenadas no banco de dados Perci.

¹<https://www.percidatabase.com/>

6.1.3 Pesquisar

Nesta página, pode-se realizar o acesso ao banco de dados Perci, através da 3 botões como descrito na Seção 5.4.



Figura 6.4: Pesquisa Inicial.

6.1.4 Download e contato

Nesta página, pode-se realizar o *download* dos dados (veja Figura 6.5), que serão armazenados em tabela do Excel, disponibilizada em um arquivo do *Google drive*. Além disso, o usuário poderá ser acessar os bancos de dados que originaram o Perci. Por fim, os contatos estão descritos na página *web*.



Figura 6.5: Download

6.2 Experimentos

Os experimentos mostram funcionalidades do banco de dados através do *site* Perci ². A seguir, apresentamos exemplos de possíveis perguntas que o usuário poderá realizar, como as perguntas poderiam ser formuladas através do site e detalhes técnicos de como essas perguntas são respondidas.

Quais são os lncRNAs que estão envolvidos no adenocarcinoma?

O usuário irá acessar a aba "Pesquisar"(veja Figura 6.4), escolher a opção "Por tipo de ncRNA" e em seguida será encaminhado para uma nova página, onde poderá selecionar o tipo de câncer "adenocarcinoma" (veja a Figura 6.6).



Figura 6.6: Pesquisa por tipo de ncRNAs.

Após selecionar o tipo de RNA "ncRNAs longos", ele será encaminhado para a página onde está disponibilizado o banco de dados com a especificação selecionada. Na Figura 6.7 é demonstrado o comportamento do banco de dados Perci com a seleção desejada.

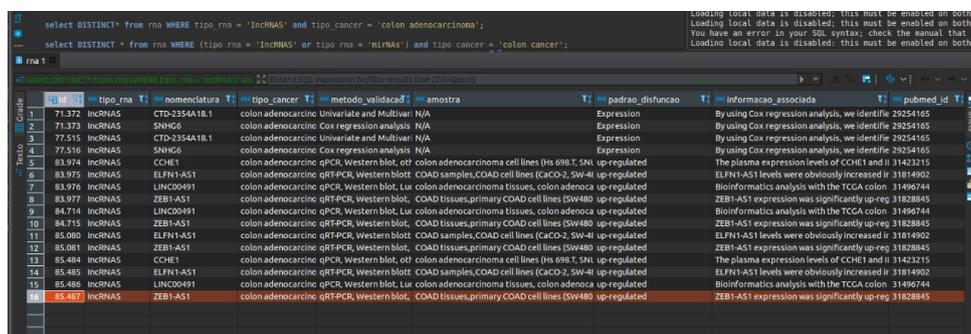


Figura 6.7: Comportamento da busca por tipo de ncRNAs no banco de dados Perci.

²<https://www.percidatabase.com/>

Nesta página, conforme demonstrado na Figura 6.8 o usuário poderá realizar a pesquisa com a tabela de todos os ncRNAs longos e selecionar na caixa de pesquisa "Tipos de Câncer" o câncer desejado (neste caso o adenocarcinoma). Após a busca, as informações desejadas são disponibilizadas em formato de tabela nesta mesma página *web*.

Tipos de RNA	Nomenclatura	Tipos de Câncer	Método de Validação	Amostra	Padrão de Disfunção	Ref
lncRNAs	CHE1	colon adenocarcinoma	qPCR, Western blot, other	colon adenocarcinoma	up-regulated	T
	CTD-2354A18.1	colon adenocarcinoma	Univariate and Multivariate Co	N/A	Expression	8
	ELFH1-AS1	colon adenocarcinoma	qRT-PCR, Western blotting	COAD samples,COAD	up-regulated	E
	LINC00491	colon adenocarcinoma	qPCR, Western blot, Luciferase reporter assay, etc.	colon adenocarcinoma	up-regulated	B
	SNHG6	colon adenocarcinoma	Cox regression analysis	N/A	Expression	B
	ZEB1-AS1	colon adenocarcinoma	qRT-PCR, Western blot,	COAD	up-regulated	Z

Figura 6.8: Resultado da busca final em tabela na página *web*

Quais são os lncRNAs e os miRNAs que estão envolvidos no câncer de cólon?

O usuário irá acessar a aba "Pesquisar"(veja Figura 6.4), escolher a opção "Por tipo de câncer" e em seguida será encaminhado para uma nova página, onde poderá selecionar o tipo de câncer de cólon (veja a Figura 6.9).

Variantes do Câncer Colorretal

Selecione uma das variantes do câncer colorretal para visualizar suas associações aos ncRNAs

[Câncer Colorretal](#)
[Câncer de Colon](#)

[Neoplasmas](#)
[Adenocarcinoma](#)

[Metástase Colorretal Hepática](#)

Figura 6.9: Pesquisa por tipo de câncer.

Após selecionar o tipo de câncer "câncer de cólon", ele será encaminhado para a página onde está disponibilizado o banco de dados com a especificação selecionada. Na Figura 6.10 é demonstrado o comportamento do banco de dados Perci com a seleção desejada.

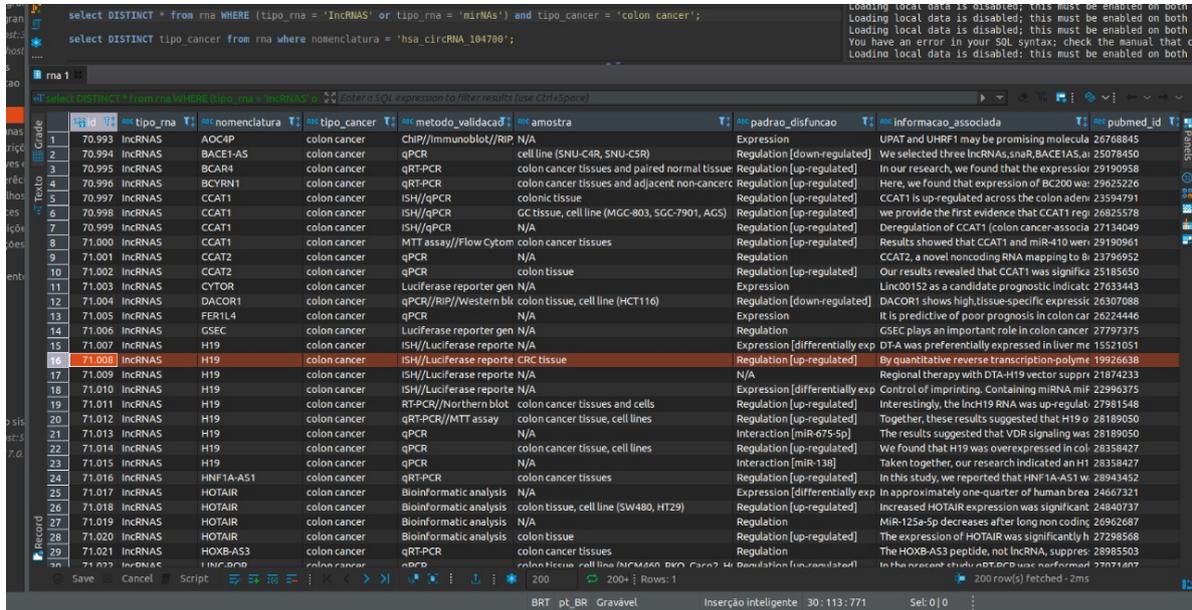


Figura 6.10: Comportamento da busca por tipo de câncer no banco de dados Perci.

Nesta página, conforme demonstrado na Figura 6.11 o usuário poderá realizar a pesquisa com a tabela de todos os dados de câncer de cólon e selecionar na caixa de pesquisa "Tipos de RNA" os tipos desejados (neste caso o lncRNA e o miRNA). Após a busca, as informações desejadas são disponibilizadas em formato de tabela nesta mesma página web.

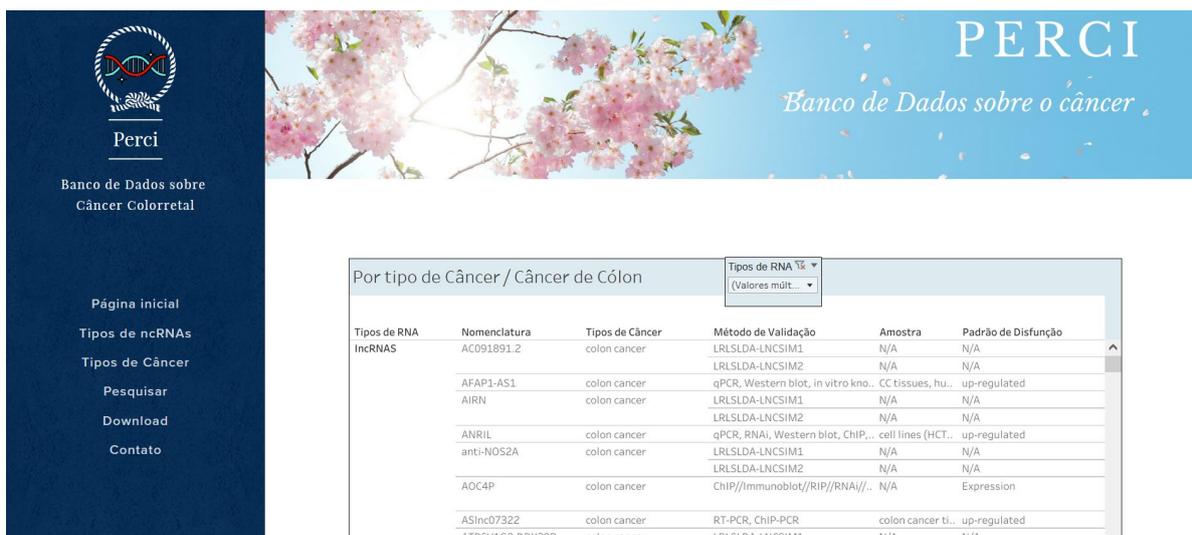


Figura 6.11: Resultado da busca final em tabela na página web

Dado um certo circRNA (hsa_circRNA_104700), quais são os subtipos de câncer que têm esse circRNA envolvido?

O usuário irá acessar a aba "Pesquisar" conforme Figura 6.4, escolher a opção "Por tipo de gene" e logo em seguida será encaminhado para uma nova página, onde poderá buscar o tipo de gene desejado (Veja a Figura 6.12).



Figura 6.12: Pesquisa por tipo de gene.

Neste caso, diferentemente dos outros experimentos, o usuário será encaminhado automaticamente para a página onde está disponibilizado o banco de dados com a especificação selecionada. Na Figura 6.13 é demonstrado o comportamento do banco de dados Perci com a seleção desejada.

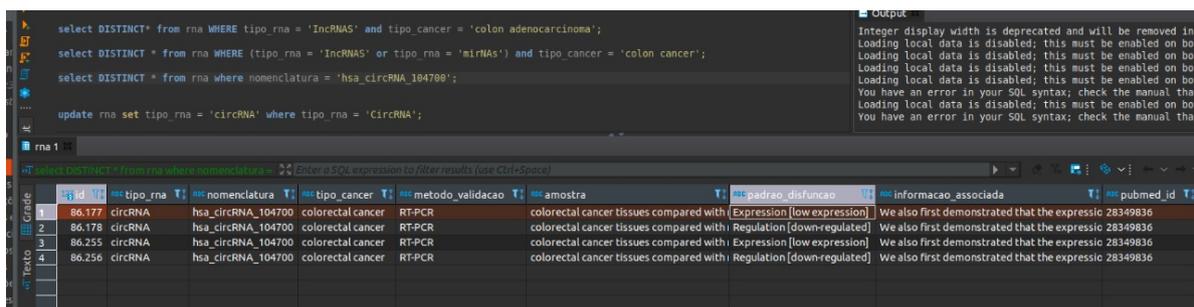


Figura 6.13: Comportamento da busca por tipo de gene no banco de dados Perci.

Após a realização da busca, os resultados são disponibilizados em formato de tabela na página *web*, conforme demonstrado na Figura 6.14 o usuário poderá realizar a pesquisa digitando o gene desejado na caixa de pesquisa "Nomenclatura". Após a busca, as informações desejadas são disponibilizadas em formato de tabela nesta mesma página *web*.

The screenshot shows the PERCI website interface. On the left is a dark blue sidebar with the PERCI logo and navigation links: 'Página inicial', 'Tipos de ncRNAs', 'Tipos de Câncer', 'Pesquisar', 'Download', and 'Contato'. The main header features the PERCI logo and the text 'Banco de Dados sobre o câncer' against a background of pink cherry blossoms. Below the header is a search box titled 'Pesquisa por Gene' with a dropdown menu showing 'Nomenclatura' and 'hsa_circRN'. The search results are displayed in a table with the following data:

Tipos de RNA	Nomenclatura	Tipos de Câncer	Método de Validação	Amostra	Padrão de Disfunç
circRNA	hsa_circRNA_086376	colorectal cancer	qPCR//High-throughput microarray a...	CRC tissues	Expression [differ
	hsa_circRNA_102761	colorectal cancer	qPCR//High-throughput microarray a...	CRC tissues from ...	Expression [differ
	hsa_circRNA_103809	colorectal cancer	RT-PCR	colorectal cancer	Expression [low ex
	hsa_circRNA_104700	colorectal cancer	RT-PCR	colorectal cancer	Expression [low ex

Figura 6.14: Resultado da busca final em tabela na página *web*

Capítulo 7

Conclusões

Neste projeto, a partir de informações de RNAs não-codificadores (ncRNAs) envolvidos em câncer colorretal (CRC), coletadas em diversas fontes, e os disponibilizamos para consulta pública, em um banco de dados. Em mais detalhes, propusemos e implementamos um repositório de dados que contém informações sobre tipos específicos de ncRNAs - longos, longos circulares e microRNAs, envolvidos em CRC, juntamente com suas características genômicas. Ainda, podem ser investigadas as interações conhecidas entre proteínas, miRNAs e lncRNAs relacionados ao CRC. por fim, criamos uma página *web*¹ para que esses dados possam ser consultados publicamente.

O banco de dados Perci permite responder perguntas como:

- Quais são os lncRNAs que estão envolvidos no adenocarcinoma?
- Quais são os lncRNAs e os miRNAs que estão envolvidos no câncer de cólon?
- Dado um certo miRNA, quais são os subtipos de câncer que têm esse miRNA envolvido?

Observamos que outra contribuição deste projeto é centralizar o acesso às informações, de forma organizada e confiável. O projeto do *site* teve como objetivo a facilidade do acesso a esses dados.

7.1 Contribuições

A criação de um repositório confiável de informações contendo ncRNAs específicos - lncRNAs(circRNAs) e sncRNAs(miRNAs), envolvidos em CRC, o que permite aos pesquisadores a obtenção, de forma rápida e confiável, de informações sobre vinculações entre determinados subtipos de CRC e tipos específicos de ncRNAs.

¹<https://www.percidatabase.com/>

7.2 Limitações

A produção de um projeto com escala relativamente grande, como este, traz limitações de como coletar um volume considerável de dados de sequências genômicas para os ncRNAs, pois esses encontram-se espalhados em diferentes tipos de fontes (páginas e suplementos de artigos, entre outros). Assim, ainda não temos disponível no banco essas sequências, para todos os ncRNAs. Outra limitação é a falta de dados clínicos, que poderiam contribuir para as pesquisas de detecção e prognóstico de CRC.

7.3 Trabalhos futuros

Os próximos passos para a evolução do banco de dados Perci é completar as sequências genômicas de CRC e incluir dados clínicos referentes aos pacientes portadores de CRC. Além disso, pretendemos incluir as redes ceRNAs [148], que ligam lncRNAs, miRNAs e proteínas específicas, ligadas ao CRC, que poderiam contribuir para encontrar e propor ceRNAs *in silico*, a serem posteriormente comprovadas em laboratório.

Referências

- [1] Magalhães, L.: *O que são os Ácidos nucleicos?*, 2015. <https://www.todamateria.com.br/que-sao-os-acidos-nucleicos/>. ix, 6
- [2] Watson, J. D.: *A Dupla Hélice*. Zahar, 2013. ix, 7, 8, 15, 16
- [3] Jesus, A. M. G., D. Badaró, D. Oliveira, E. S. Miyasato, F. Pereira e R. M. Aquino: *Ácidos nucleicos e nucleotídeos*, 2015. https://edisciplinas.usp.br/pluginfile.php/2311662/mod_resource/content/0/pdf_Apresent_14_Gr14.pdf. ix, 7
- [4] Setubal, J. e J. Meidanis: *Introduction to Computational Molecular Biology*. PWS PUBLISHING COMPANY, 1997. ix, 1, 8, 14, 16
- [5] Magalhães, M.: *DNA, RNA e proteínas*, 2014. <http://salabioquimica.blogspot.com/2014/05/bases-bioquimicas-da-hereditariedade.html>. ix, 9
- [6] Koenig, Ms.: *06. picture of 3 types of RNA*, 2009. <https://sites.google.com/site/biopeepsfiles/dna/063typesofrnaimage>. ix, 9
- [7] Koenig, Ms.: *Picture of 3 types of RNA*, 2009. <https://sites.google.com/site/biopeepsfiles/dna/063typesofrnaimage>. ix, 10
- [8] Magalhães, F. D.: *Vector tRNA*, 2014. encurtador.com.br/dLPU9. ix, 11
- [9] Bezerra, J. A.: *Ácidos nucléicos, transcrição e tradução*, 2010. <http://bioconhecer.blogspot.com.br/2010/04/acidos-nucleicos-transcricao-e-traducao.html>. ix, 12
- [10] Dias, D. L.: *Proteínas - manual da química*, 2018. <https://www.manualdaquimica.com/quimica-dos-alimentos/proteinas.htm>. ix, 13
- [11] Campbell, M. K.: *Proteínas: Classificação e características*, 2008. <http://www.explicatorium.com/quimica/proteinas.html>. ix, 13
- [12] Lima, A. P. O.: *Dogma central da biologia molecular*, 2013. <https://www.infoescola.com/biologia/dogma-central-da-biologia-molecular/>. ix, 15
- [13] Moreira, C.: *Transcrição*, 2015. <https://rce.casadasciencias.org/rceapp/art/2015/065/>. ix, 15

- [14] Bhat, A. A., S. N. Younes, S. S. Raza, L. N. Zarif, S. S. Nisar, I. Ahmed, R. Mir, S. Kumar, S. K. Sharawat, S. Hashem e et all: *Role of non-coding RNA networks in leukemia progression, metastasis and drug resistance*. *Molecular cancer*, 19(1):1–21, 2020. ix, 17
- [15] Inca: *Câncer*. "<https://www.inca.gov.br/o-que-e-cancer>", 2020. ix, 22, 23, 24, 25, 26
- [16] Freitas, E.: *Banco de dados geodésicos*, 2012. <https://mundogeo.com/2005/09/27/banco-de-dados-geodesicos/>. ix, 32
- [17] Silva, C. O. P.: *A identificação criminal e banco de dados genéticos: a constitucionalidade da lei.*, 2018. <https://carluskaoliveira.jusbrasil.com.br/artigos/660508335/a-identificacao-criminal-e-banco-de-dados-geneticos>. ix, 32
- [18] Alves, G. F. O.: *O que é um SGBD?*, 2013. <https://dicasdeprogramacao.com.br/o-que-e-um-sgbd/>. ix, 34
- [19] Lopes, B.: *Modelo conceitual de dados - aprenda a utilizar os principais mecanismos de abstração*. "encurtador.com.br/alv46", 2020. ix, x, 37, 38, 39, 40, 41
- [20] Cândido, C. H.: *Brmodelo 3.3*. "<http://www.sis4.com/brModelo/>", 2020. x, 45, 46
- [21] Eddy, S.: *Non-coding RNA genes and the modern RNA world*. *Nature Reviews Genetics*, 2(12):919–929, 2001. 1
- [22] Huang, B. e R. Zhang: *Regulatory non-coding RNAs: revolutionizing the RNA world*. *Molecular biology reports*, 41(6):3915–3923, 2014. 1
- [23] Li, J., M. Zhang, G. An e Q. Ma.: *lncRNA tug1 acts as a tumor suppressor in human glioma by promoting cell apoptosis*. *Experimental Biology and Medicine*, página 1535370215622708, 2016. 1, 2
- [24] Yang, Q., Z. L. Cui, Q. Wang e et al: *plncRNA-1 induces apoptosis through the her-2 pathway in prostate cancer cells*. *Asian journal of andrology*, 19(4):453, 2017. 1
- [25] Orom, U. A. e R. Shiekhattar: *Noncoding RNAs and enhancers: complications of a long-distance relationship*. *Trends in Genetics*, 27(10):433–439, 2011. 2
- [26] Ponting, C., P. Oliver e W. Reik: *Evolution and functions of long noncoding RNAs*. *Cell*, 136(4):629–641, February 2009. 2
- [27] Mercer, T., M. Dinger e J. Mattick: *Long non-coding RNAs: insights into functions*. *Nature Reviews Genetics*, 10(3):155–159, 2009. 2
- [28] Zhang, X., S. Weissman e P. Newburger: *Long intergenic non-coding RNA HO-TAIRM1 regulates cell cycle progression during myeloid maturation in nb4 human promyelocytic leukemia cells*. *RNA Biology*, 11(6):777–787, 2014. 2

- [29] Li, J., M. Zhang, G. An e Q. Ma.: *LncRNA TUG1 acts as a tumor suppressor in human glioma by promoting cell apoptosis*. Experimental Biology and Medicine, página 1535370215622708, 2016. 2
- [30] Paci, P., T. Colombo e L. Farina: *Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer*. BMC Systems Biology, 8(1):83, 2014. 2, 18
- [31] Prensner, J. R. e A. M. Chinnaiyan: *The emergence of lncRNAs in cancer biology*. Cancer Discovery, 1(5):391–407, 2011. 2
- [32] Fachel, A., A. Tahira, S. V. Arias, V. M. Coutinho, E. Gimba, G. Vignal, F. Campos, E. Reis e S. V. Almeida: *Expression analysis and in silico characterization of intronic long noncoding RNAs in renal cell carcinoma: emerging functional associations*. Mol Cancer, 12(140):10–1186, 2013. 2
- [33] Beckedorff, F. C., M. S. Amaral, C. D. Pereira e S. V. Almeida: *Long non-coding RNAs and their implications in cancer epigenetics*. Biosci Rep, 2013. 2
- [34] Reis, E. e S. V. Almeida: *Perspectives of long non-coding RNAs in cancer diagnostics*. Frontiers in Genetics, 3:32, 2012. 2
- [35] Siegel, R., J. Ma, Z. Zou e A. Jemal: *Cancer statistics*. CA: a Cancer Journal for Clinicians, 64(1):9–29, 2014. 2
- [36] Hanahan, D. e R. A Weinberg: *Hallmarks of cancer: the next generation*. Cell, 144(5):646–674, 2011. 2
- [37] Liao, Q., L. Chen, J. Liu, T. Yang, J. Li, X. Zhang e J. Zhao: *Comprehensive analysis of epigenetic pattern of long noncoding RNA loci in colorectal cancer*. Cell, 159(1):9–17, 2016. 2
- [38] Hu, T., L. Fei Li, J. Shen e et al: *Chronic inflammation and colorectal cancer: The role of vascular endothelial growth factor*. Current pharmaceutical design, 21(21):2960–2967, 2015. 2
- [39] Gao, R., C. Fang, J. Xu e et al: *LncRNA CACS15 contributes to oxaliplatin resistance in colorectal cancer by positively regulating abcc1 through sponging mir-145*. Archives of Biochemistry and Biophysics, 633:183–191, 2019. 2
- [40] Zhou, J., J. Lin, H. Zhang e et al: *LncRNA HAND2-AS1 sponging mir-1275 suppresses colorectal cancer progression by upregulating KLF14*. Biochemical and Biophysical Research Communications, 503(3):1848–1853, 2018. 2
- [41] Ding, J., J. Zhao, Z., Song e et al: *MiR-223 promotes the doxorubicin resistance of colorectal cancer cells via regulating epithelial–mesenchymal transition by targeting FBXW7*. Acta Biochimica et Biophysica Sinica, 50(6):597–604, 2018. 2
- [42] Thorenor, N., P. F. Vychytilova e et al S. Hombach: *Long non-coding RNA ZFAS1 interacts with CDK1 and is involved in p53-dependent cell cycle control and apoptosis in colorectal cancer*. Oncotarget, 7(1):622, 2016. 2

- [43] Wang, Q., H. Zhang, X. Shen e S. Ju: *Serum microRNA-135a-5p as an auxiliary diagnostic biomarker for colorectal cancer*. *Annals of Clinical Biochemistry*, 54(1):76–85, 2017. 2, 3
- [44] Inoue, A., H. Yamamoto, M. Uemura e et al: *MicroRNA-29b is a novel prognostic marker in colorectal cancer*. *Annals of Surgical Oncology*, 22(3):1410–1418, 2015. 2, 3
- [45] Igarashi, H., H. Kurihara, K. Mitsuhashi e et al: *Association of microRNA-31-5p with clinical efficacy of anti-EGFR therapy in patients with metastatic colorectal cancer*. *Annals of Surgical Oncology*, 22(8):2640–2648, 2015. 3
- [46] Yuan, W., X. Li, L. Liu e et al: *Comprehensive analysis of lncrna associated ceRNA network in colorectal cancer*. *Biochemical and biophysical research communications*, 508(2):374–379, 2019. 3
- [47] Amabis, J. M. e G. R. Martho: *Fundamentos da biologia moderna*. Moderna, 1991. 6
- [48] Watson, J. D., T. A. Becker, S. P. Bell, A. Gann, M. Levine e R. Losick: *Biologia Molecular do Gene*. Artmed, 2015. 7
- [49] Michael Greenwood, M.Sc.: *Que são introns e exons?* [https://www.news-medical.net/life-sciences/What-are-introns-and-exons-\(Portuguese\).aspx](https://www.news-medical.net/life-sciences/What-are-introns-and-exons-(Portuguese).aspx), 2018. 16
- [50] Mattick, J. S. e I. V. Makunin: *Non-coding RNA*. *Human molecular genetics*, 15(suppl_1):R17–R29, 2006. 17
- [51] Kalvari, I., J. Argasinska, N. Quinones-Olvera, E. P. Nawrocki e et a: *Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families*. *Nucleic Acids Research*, 46(D1):D335–D342, 2017. 18
- [52] Oliveira, J. V. A.: *Identificação de snoRNAs usando aprendizagem de máquina*. Master's thesis, Universidade de Brasília, 2016. 18
- [53] Oh, Y. K. e T. G. Park: *siRNA delivery systems for cancer treatment. advanced drug delivery reviews*. Master's thesis, Universidade de Brasília, 61(10):850–862, 2009. 18
- [54] Brennecke, J., D. R. Hipfner, A. Stark, R. B. Russell, e S. M. Cohen: *bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in drosophila*. *Cell*, 113(1):25–36, 2003. 18
- [55] Beyret, E., H. Lin e K. Appasani: *MicroRNAs—from Basic Science to Disease Biology*. cambridge, 2008. 18
- [56] He, L. e G. J. Hannon: *MicroRNAs: small RNAs with a big role in gene regulation*. *Nature reviews genetics*, 5(7):522–531, 2004. 18
- [57] Zhang, C.: *MicroRNomics: a newly emerging approach for disease biology*. *Physiological genomics*, 33(2):139–147, 2008. 18

- [58] Bartel, D. P.: *MicroRNAs: genomics, biogenesis, mechanism, and function*. *cell*, 116(2):281–297, 2004. 19
- [59] Ambros, V., B. Bartel, D. P. Bartel, C. B. Burge, J. C. Carrington, X. Chen, G. Dreyfuss, S. R. Eddy, S. G. Jones, M. Marshall e et all: *A uniform system for microRNA annotation*. *RNA*, 9(3):277–279, 2003. 19
- [60] Vieira, Lucas Maciel: *Non-coding RNAs in colorectal cancer: a bioinformatics perspective*. Doctorate text. Departamento de Ciência da Computação. Universidade de Brasília, página 9, 2019. 19, 20
- [61] Tahira, A. C., M. S. Kubrusly, M. F. Faria, B. Dazzani, R. S. Fonseca, V. M. Coutinho, S. V. Almeida, M. Machado e E. M. Reis: *Long noncoding intronic rnas are differentially expressed in primary and metastatic pancreatic cancer*. *Molecular Cancer*, 10(1):141, 2011. 20
- [62] Peng, W. X., P. Koirala e Y.Y. Mo: *LncRNA-mediated regulation of cell signaling in cancer*. *Oncogene*, 36(41):5661, 2017. 20
- [63] Ma, C., K. Nong, H. Zhu, W.Wang, X. Huang, Z. Yuan e K. Ai: *H19 promotes pancreatic cancer metastasis by derepressing let-7's suppression on its target hmga2-mediated emt*. *Tumor Biology*, 35(9):9163– 9169, 2014. 20
- [64] Zheng, H. T., D. B. Shi, Y. W. Wang, X. X. Li, Y. Xu, P. Tripathi, W. L. Gu, G.X. Cai e S. J. Cai: *High expression of lncRNA malat1 suggests a biomarker of poor prognosis in colorectal cancer*. *International Journal of Clinical and Experimental Pathology*, 7(6):3174, 2014. 20
- [65] Zhang, R., L. Q. Xia, W. W. Lu, J. Zhang e J. S. Zhu: *LncRNAs and cancer*. *Oncology Letters*, 12(2):1233–1239, 2016. 20
- [66] Wang, L., H. Jung Park, S. Dasari, S. Wang, J. Kocher e W. Li: *CPAT: Coding-potential assessment tool using an alignment-free logistic regression model*. *Nucleic Acids Research*, 41(6):e74–e75, 2013. 20
- [67] Kong, L., Y. Zhang, Z. Q. Ye, X. Q. Liu, S. Q. Zhao, L. Wei, e G. Gao: *CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine*. *Nucleic Acids Research*, 35:W345–9, 2007. 20
- [68] W. Hu, H. Zhang a nd, J. Hao, S. Lv, C. Wang, W. Tong, Y. Wang, Y. Wang, X. Liu e W. Ji: *Genome-wide identification and functional prediction of novel and fungi-responsive lincrnas in triticum aestivum*. *BMC Genomics*, 17(1):238, 2016. 20
- [69] Zhao, W., Y. Mu, L. Ma, C. Wang, Z. Tang, S. Yang, R. Zhou, X. Hu, M. H. Li e K. Li: *Systematic identification and characterization of long intergenic non-coding RNAs in fetal porcine skeletal muscle development*. *Scientific Reports*, 5, 2015. 20
- [70] Etebari, K., M. J. Furlong e S. Asgari: *Genome wide discovery of long intergenic non-coding RNAs in diamondback moth (plutella xylostella) and their expression in insecticide resistant strains*. *Scientific Reports*, 5, 2015. 20

- [71] Tong, C., Q. Chen, L. Zhao, J. Ma, E. Ibeagha-Awemu e X. Zhao: *Identification and characterization of long intergenic noncoding RNAs in bovine mammary glands*. BMC Genomics, 18(1):468, 2017. 20
- [72] Wang, L., X. Ma, X. Xu e Y. Zhang: *Systematic identification and characterization of cardiac long intergenic noncoding rnas in zebrafish*. Scientific Reports, 7(1):1250, 2017. 20
- [73] Etebari, K., S. Asad, G. Zhang e S. Asgari: *Identification of aedes aegypti long intergenic non-coding rnas and their association with wolbachia and dengue virus infection*. PLoS Neglected Tropical Diseases, 10(10):e0005069, 2016. 20
- [74] Han, S., Y. Liang, Y. Li e W. Du: *Long non-coding RNA identification: comparing machine learning based tools for long non-coding transcripts discrimination*. BioMed Research Internationals, 2016. 20
- [75] Sun, L., H. Liu, L. Zhang e J. Meng: *lncscan-svm: A tool for predicting long non-coding rnas using support vector machine*. PloS One, 10(10):e0139654, 2015. 20
- [76] Fan, X. e S. Zhang: *lncRNA-mfdl: identification of human long non-coding RNAs by fusing multiple features and using deep learning*. Molecular Biosystems, 11(3):892–897, 2015. 20
- [77] Pian, C., G. Zhang, Z. Chen, Y. Chen, J. Zhang, T. Yang e L. Zhang: *LncRNAPred: Classification of long non-coding rnas and proteincoding transcripts by the ensemble algorithm with a new hybrid feature*. PloS One, 11(5):e0154567, 2016. 20
- [78] Sun, L., H. Luo, D. Bu, G. Zhao, K. Yu, C. Zhang, Y. Liu, R. Chen e Y. Zhao: *Utilizing sequence intrinsic composition to classify proteincoding and long non-coding transcripts*. Nucleic Acids Research, página gkt646, 2013. 20
- [79] Li, A., J. Zhang e Z. Zhou: *PLEK: a tool for predicting long non-coding rnas and messenger RNAs based on an improved k-mer scheme*. BMC Bioinformatics, 15(1):1, 2014. 20
- [80] Schneider, H.: *Identification of long non-coding RNAs using SVM*. Doctorate text. Departamento de Ciência da Computação. Universidade de Brasília, 15(1):1, 2016. 20
- [81] Sun, K., X. Chen, P. Jiang, X. Song, H. Wang e H. Sun: *IseeRNA: identification of long intergenic non-coding RNA transcripts from transcriptome sequencing data*. BMC Genomics, 13:Suppl 2:S7, 2013. 20
- [82] Wang, Y., Y. Li, Q. Wang, Y. Lv, S. Wang, X. Chen, X. Yu, W. Jiang, e X. Li: *Computational identification of human long intergenic noncoding rnas using a GA-SVM algorithm*. Gene, 533(1):94–99, 2014. 20
- [83] Cheng, Z. F. e M. P. Deutscher: *An important role for RNase r in mRNA decay*. Molecular cell, 17(2):313–318, 2005. 20

- [84] Venkataraman, K., M. G. Diaz, W. Karzai e et all: *Non-stop mRNA decay: a special attribute of trans-translation mediated ribosome rescue*. *Frontiers in microbiology*, 5:93, 2014. 20
- [85] Zhang, Miaoci e Yan Xin: *Circular RNAs: a new frontier for cancer diagnosis and therapy*. *Journal of hematology & oncology*, 11(1):1–9, 2018. 21
- [86] Greene, J., A. Baird, L. Brady, M. Lim, S. Gray, R. McDermott e S. Finn: *Circular RNAs: biogenesis, function and role in human diseases*. *Frontiers in molecular biosciences*, 4:38, 2017. 21
- [87] Dong, Y., D. He, Z. Peng, W. Peng, W. Shi, J. Wang, B. Li, C. Zhang e C. Duan: *Circular rnas in cancer: an emerging key player*. *Journal of hematology & oncology*, 10(1):1–8, 2017. 21
- [88] Organization, World Health et al.: *Global action plan for the prevention and control of noncommunicable diseases 2013-2020*. World Health Organization, 2013. 21
- [89] Gomes, C., A. Somoza, E. Creemers, C. Dieterich, M. Lustrek, Y. Devaux e et all: *Circular RNAs in the cardiovascular system*. *Non-coding RNA research*, 3(1):1–11, 2018. 21
- [90] Kristensen, LS, TB Hansen, MT Venø e J Kjems: *Circular RNAs in cancer: opportunities and challenges in the field*. *Oncogene*, 37(5):555–565, 2018. 21
- [91] Pinto, M. F. S.: *Avaliação de novos ciclotideos com potencial anticâncer nanoestruturação associadas*. Tese apresentada ao Programa de Pós-Graduação Stricto Sensu em Ciências Genômicas e Biotecnologia da Universidade Católica de Brasília, 2017. 22
- [92] Daniel, C. L. V.: *Estudo sobre os cânceres*. Omnis Scientia, 2021. 22
- [93] Santos, Z. M. S. A., J. M. P. Caldas, M. H. A. G. Jardim e A. P. Soares: *Câncer—da prevenção dos fatores de risco, às condutas de controle com foco na integralidade*. *TRATAMENTO MULTIDISCIPLINAR EM PACIENTES ONCOLÓGICOS*, página 28, 2010. 22
- [94] SABRINA, V. S. e L. R. M. SANTOS: *Papel do enfermeiro no cuidado ao paciente vítima de lesão por extravasamento de quimioterápico vesicante*. SEMESP, 2020. 22
- [95] Bailar, J. C. e H. L. Gornik: *Cancer undefeated*. *New England Journal of Medicine*, 336(22):1569–1574, 1997. 24
- [96] Oncoguia, Equipe: *O que é câncer?* "<http://www.oncoguia.org.br/conteudo/cancer/12/1/>", 2017. 24
- [97] Verma, M.: *Personalized medicine and cancer*. *Journal of personalized medicine*, 2(1):1–14, 2012. 24

- [98] Inca: *Câncer do intestino*. "https://www.inca.gov.br/tipos-de-cancer/cancer-de-intestino", 2020. 26, 27
- [99] Aw, M. S., L. Paniwnyk e D. Losic: *The progressive role of acoustic cavitation for non-invasive therapies, contrast imaging and blood-tumor permeability enhancement*. Expert opinion on drug delivery, 13(10):1383–1396, 2016. 26, 27
- [100] Galsky, M. D.: *Everything You Need to Know About Cancer in Language You Can Understand*. Jones & Bartlett Learning, 2009. 26, 27
- [101] Pisani, P., M. Airoidi, A. Allais, P. A. Valletti, M. Battista, M. Benazzo, R. Briatore, S. Cacciola, S. Cocuzza, A. Colombo e et all: *Metastatic disease in head & neck oncology*. Acta Otorhinolaryngologica Italica, 40(Suppl 1):S1, 2020. 26, 27
- [102] Polomano, R., M. Ashburn e J. Farrar: *Pain syndromes in cancer survivors*. Em *Cancer pain: Assessment and management*, páginas 145–163. Cambridge University Press New York, 2010. 26, 27
- [103] Saúde, Organização Mundial da: *Cancer colorectal*, 2021. <https://www.who.int/>. 27
- [104] Srivastava, M. Raghvendra, Purohit, A. Tanaya, Chan e A. Timothy: *Diverse neoantigens and the development of cancer therapies*. Em *Seminars in radiation oncology*, volume 30, páginas 113–128. Elsevier, 2020. 27
- [105] Williams, A. R., B. A. Balasooriya e D. W. Day: *Polyps and cancer of the large bowel: a necropsy study in liverpool*. Gut, 23(10):835–842, 1982. 28
- [106] Shia, J., L. H. Tang, M. R. Weiser, B. Brenner, N. V. Adsay, E. B. Stelow, L. B. Saltz, J. Qin, R. Landmann, G. D. Leonard e et all: *Is non-small cell type high-grade neuroendocrine carcinoma of the tubular gastrointestinal tract a distinct disease entity?* The American journal of surgical pathology, 32(5):719–731, 2008. 28
- [107] Thomas, R. M. e L. H. Sobin: *Gastrointestinal cancer*. Cancer, 75(S1):154–170, 1995. 28
- [108] Preiser, C. M. F, R. R. Pascal e K. Perzin: *Tumors of the intestines*. Armed Forces Institute of Pathology under the auspices of Universities . . . , 1990. 28
- [109] Moll, R., A. Löwe, J. Laufer e W. W. Franke: *Cytokeratin 20 in human carcinomas. a new histodiagnostic marker detected by monoclonal antibodies*. The American journal of pathology, 140(2):427, 1992. 28
- [110] Hatch, Q. M. e E. W. Gilbert: *“miscellaneous” tumors of the small bowel, colon, and rectum: Appendiceal neoplasms*. Clinics in colon and rectal surgery, 31(5):278, 2018. 28
- [111] Oronsky, B., P. C. Ma, D. Morgensztern e C. A. Carter: *Nothing but net: a review of neuroendocrine tumors and carcinomas*. Neoplasia, 19(12):991–1002, 2017. 28

- [112] Modlin, I. M., K. Oberg, D. C. Chung, R. T. Jensen, W. W. Herder, R. V. Thakker, M. Caplin, F. G. Delle Fave, G. A. Kaltsas, E. P. Krenning e et all: *Gastroenteropancreatic neuroendocrine tumours*. *The lancet oncology*, 9(1):61–72, 2008. 28
- [113] Gardner, G. J., D. R. Lagunes e P. A. Gehrig: *Neuroendocrine tumors of the gynecologic tract: A society of gynecologic oncology (sgo) clinical document*. *Gynecologic oncology*, 122(1):190–198, 2011. 28
- [114] Hassan, A. A. A.: *Biological Analysis of Exon 7 in P53 Gene among Non-Hodgkin Lymphoma Patients using DNA Sequencing Technique, the National Cancer Institute, University of Gezira, Sudan (2017)*. Tese de Doutorado, University of Gezira, 2018. 29
- [115] Mahoney, K. M., P. D. Rennert e G. J. Freeman: *Combination cancer immunotherapy and new immunomodulatory targets*. *Nature reviews Drug discovery*, 14(8):561–584, 2015. 29
- [116] Abbott, M. e Y. Ustoyev: *Cancer and the immune system: The history and background of immunotherapy*. Em *Seminars in oncology nursing*, volume 35, página 150923. Elsevier, 2019. 29
- [117] Coffey, R. J., M. K. Washington, C. L. Corless, M. C. Heinrich e et all: *Menetrier disease and gastrointestinal stromal tumors: hyperproliferative disorders of the stomach*. *The Journal of clinical investigation*, 117(1):70–80, 2007. 29
- [118] Preiser, C. F., F. Carneiro, P. Correa, P. Guilford, R. Lambert, F. Megraud, N. Munoz, S. M. Powell, M. Rugge, M. Sasako e et all: *Gastric carcinoma*. *Pathology and genetics of tumours of the digestive system*, 1:35–52, 2000. 29
- [119] Sanders, K. M., S. M. Ward e S. D. Koh: *Interstitial cells: regulators of smooth muscle function*. *Physiological reviews*, 2014. 29
- [120] Bishnupuri, K. S. e M. K. Mishra: *Epigenetics of colorectal cancer*. Em *Epigenetic Advancements in Cancer*, páginas 97–121. Springer, 2016. 29
- [121] Islam, S. M. A.: *The Role of Atypical Protein Kinase C in Colorectal Cancer Cells Carcinogenesis*. Tese de Doutorado, University of South Florida, 2019. 29
- [122] Miller, K. D., L. Nogueira, A. B. Mariotto, J. H. Rowland, K. R. Yabroff, C. M. Alfano, A. Jemal, J. L. Kramer e R. L. Siegel: *Cancer treatment and survivorship statistics, 2019*. *CA: a cancer journal for clinicians*, 69(5):363–385, 2019. 29
- [123] Scotto, J., A. W. Kopf e F. Urbach: *Non-melanoma skin cancer among caucasians in four areas of the united states*. *Cancer*, 34(4):1333–1338, 1974. 29
- [124] Johnson, M. M., S. A. Leachman, L. S. Aspinwall, L. D. Cranmer, C. C. Lewandrowski, V. K. Sondak, C. E. Stemwedel, S. M. Swetter, J. Vetto, T. Bowles e et all: *Skin cancer screening: recommendations for data-driven screening guidelines and a review of the us preventive services task force controversy*. *Melanoma management*, 4(1):13–37, 2017. 29

- [125] Xie, B., Q. Ding, H. Han e D. Wu: *mirncancer: a microRNA-cancer association database constructed by text mining on literature*. *Bioinformatics*, 29(5):638–644, 2013. 30
- [126] Chen, G., Z. Wang, D. Wang, C. Qiu, M. Liu, X. Chen, Q. Zhang, G. Yan e Q. Cui: *Lncrnadisease: a database for long-non-coding rna-associated diseases*. *Nucleic acids research*, 41(D1):D983–D986, 2012. 30
- [127] ALVES, W. P.: *Banco de dados*. Saraiva Educação SA, 2014. 31
- [128] Fusco, E.: *Modelos conceituais de dados como parte do processo da catalogação: perspectiva de uso dos FRBR no desenvolvimento de catálogos bibliográficos digitais*. Univem, 2012. 31, 37
- [129] Elmasri, R., S. B. Navathe, M. G. Pinheiro e et all: *Sistemas de banco de dados*. Pearson Addison Wesley São Paulo, 2005. 33, 34, 35, 36, 38
- [130] Pimentel, C. C.: *Bancos de dados relacionais: uma análise comparativa entre ferramentas sgbd livre e proprietária*. *Tecnologia em Gestão da Tecnologia da Informação- Unisul Virtual*, 2019. 33
- [131] Date, C. J.: *Introdução a sistemas de bancos de dados*. Elsevier Brasil, 2004. 33, 34, 35
- [132] Mozzato, A. R., D. Grzybovski e A. N. Teixeira: *Análises qualitativas nos estudos organizacionais: as vantagens no uso do software nvivo®*. *Revista Alcance (Online)*, 23(4):578, 2016. 33, 34, 35
- [133] Ramakrishnan, R. e J. Gehrke: *Sistemas de gerenciamento de banco de dados-3*. AMGH Editora, 2008. 34, 35
- [134] CARVALHO, J. R.: *MODELO DE METADADOS PARA SISTEMAS DE DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS UTILIZANDO XML*. Universidade Estadual de Maringá, 2007. 35
- [135] Sayão, L., L. M. B. B. Toutain, F. G. M. G. Rosa e C. H. Marcondes: *Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação*. EDUFBA, 2009. 35
- [136] Barcellos, J. S. e V. B. Goedert: *Simplificando a administração de bancos de dados: Uma abordagem no modelo relacional com postgresql*. *Sistemas de Informação- Florianópolis*, 2019. 35
- [137] Booch, G., J. Rumbaugh e I. Jacobson: *UML: guia do usuário*. Elsevier Brasil, 2006. 37
- [138] Heuser, C. A.: *Projeto de banco de dados: Volume 4 da Série Livros didáticos informática UFRGS*. Bookman Editora, 2009. 38
- [139] Furgeri, S. e et all: *Representação de informação e conhecimento: estudo das diferentes abordagens entre a ciência da informação e a ciência da computação*. Pontifícia Universidade Católica de Campinas, 2006. 38

- [140] Lima, J. A. O.: *Modelo genérico de relacionamentos na organização da informação jurídica e legislativa*. FURB, 2008. 38
- [141] Machado, F. N. R.: *Banco de Dados Projeto e Implementação*. Saraiva Educação SA, 2020. 39
- [142] Alexandruk, M.: *Modelagem de banco de dados*. São Paulo, SP, Brasil, 2007. 39
- [143] Filho, J. L. e C. Iochpe: *Modelagem de Bancos de Dados Geográficos*. Revista IP – Informática Pública, 2001. 39
- [144] Lopes, M. M. G.: *Modelagem conceitual de regras de negócio baseada em ontologia de fundamentação*. Tese de Mestrado, PUCPR, 2011. 39
- [145] Pádua, S. I. D.: *Investigação do processo de desenvolvimento de software a partir da modelagem organizacional, enfatizando regras do negócio*. Tese de Doutorado, Universidade de São Paulo, 2001. 39
- [146] Campos, M. L. A., L. M. Campos e J. S. Medeiros: *A representação de domínios de conhecimento e uma teoria de representação: a ontologia de fundamentação*. Informação & Informação, 16(2):140–164, 2011. 39
- [147] Lazzaletti, A. T.: *Csdc—uma ferramenta de conversão de script sql em diagrama de classes uml*. INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA SUL-RIO, 2012. 40
- [148] JH, Li, Liu S, Zhou H, Qu LH e Yang JH: *starbase v2.0 release*, 2013. <https://starbase.sysu.edu.cn/starbase2/index.php>. 62