

Representação do conhecimento, grafos de conhecimento, ontologias e suas aplicações

Aplicações de Ontologias

Relatando experiências prévias

Patrícia Rufino Oliveira

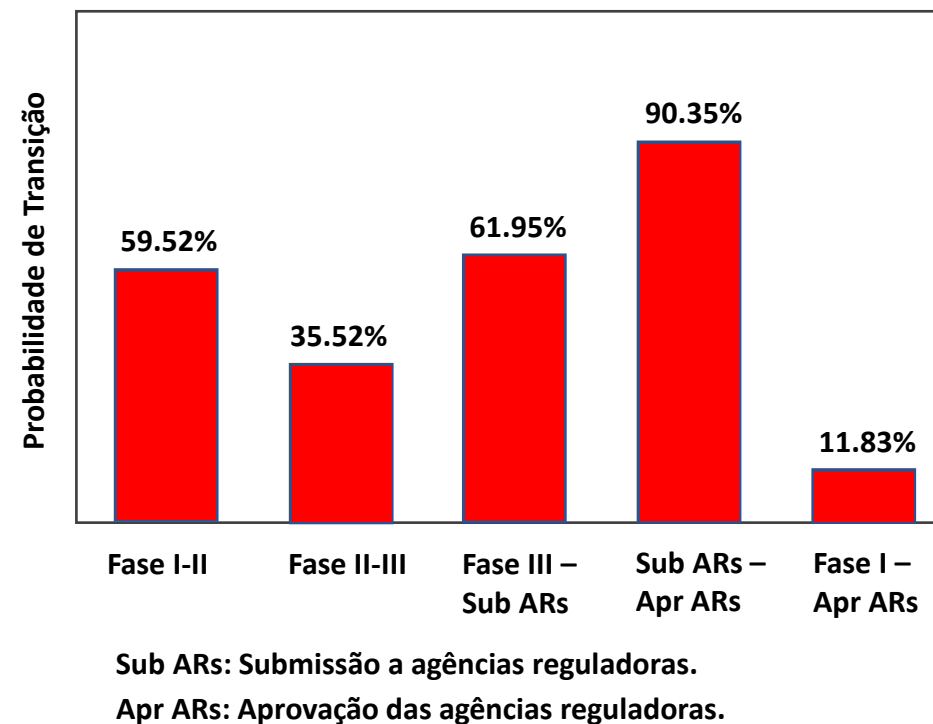
Fevereiro - 2023

Roteiro

- Introdução
- Metodologia baseada em IA para o suporte à descoberta de medicamentos
- OntoQSAR - Uma ontologia para apoio a descoberta de medicamentos
- Análise de agrupamento de expressões gênicas
- Transferência de aprendizado
- Referências

Contexto

- O custo médio para o desenvolvimento de um novo medicamento é estimado em US\$ 2,6 bilhões e leva, em média, pelo menos 10 anos.
- A probabilidade de um novo medicamento conseguir entrar no mercado foi estimada em menos de 12% (DIMASIA; GRABOWSKIB; HANSEN, 2016).

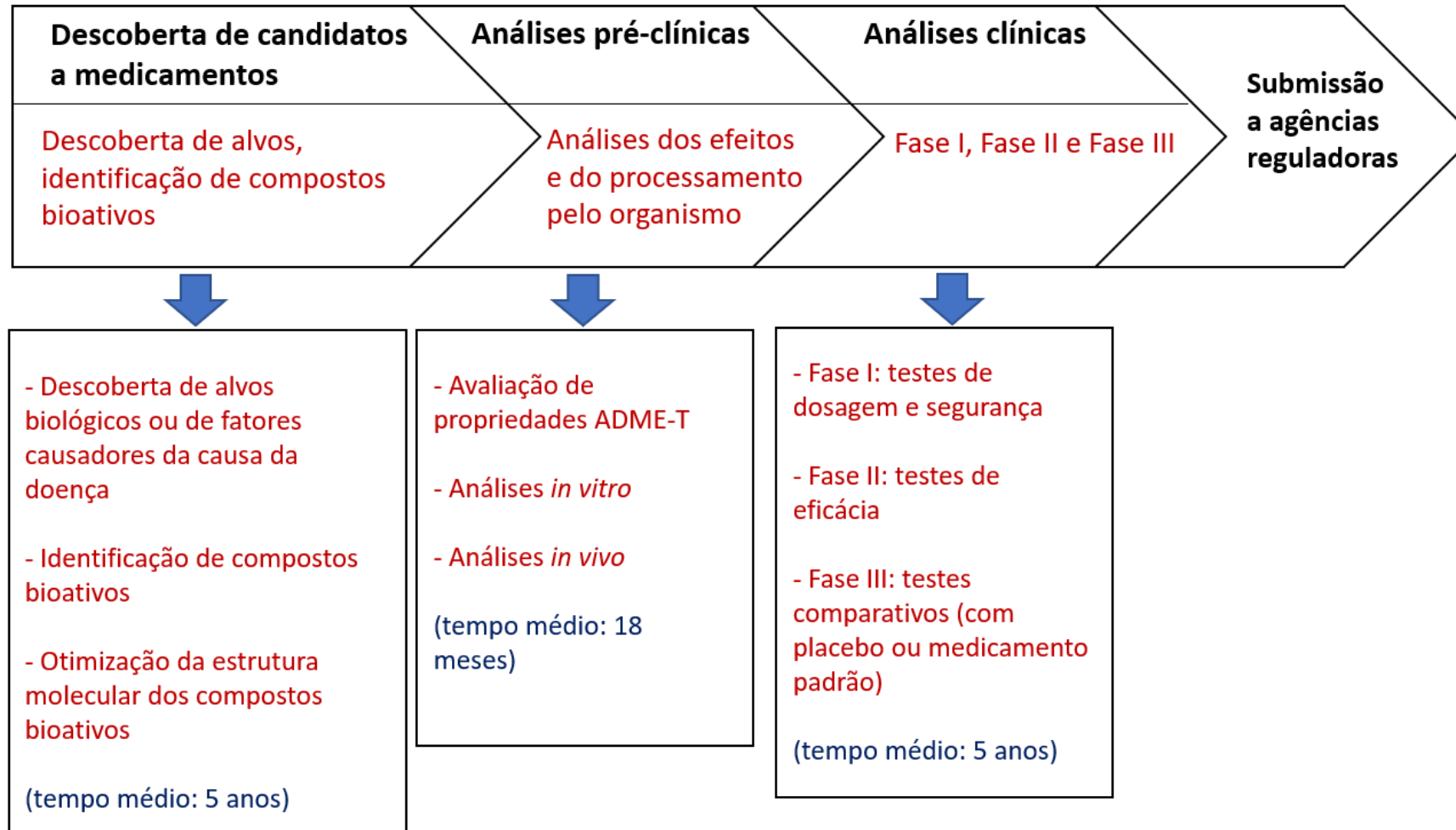


Fonte: (DIMASIA; GRABOWSKIB; HANSEN, 2016).

Motivação

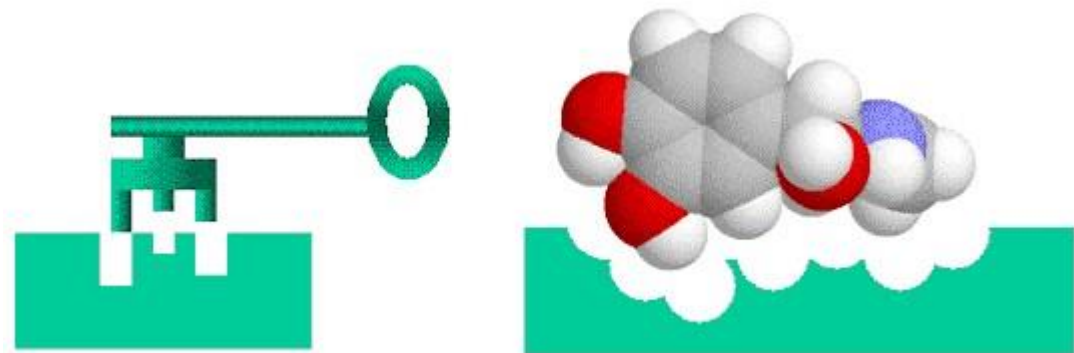
- Nesse contexto, técnicas computacionais são ferramentas poderosas, que podem reduzir o número de compostos químicos a ser analisados em ensaios experimentais, funcionando como um “atalho” para a descoberta um novo medicamento.

Etapas do desenvolvimento de um medicamento



Descoberta de novos medicamentos

- Pesquisas na área de descoberta de medicamentos buscam identificar pequenas moléculas com potencial para **interagir com alvos terapêuticos** no organismo, combatendo uma determinada doença.
- Um **alvo biológico** pode ser uma proteína ou um organismo patogênico, como um vírus ou uma bactéria.



fonte: Brokenhaven National Laboratory (bnl.gov)

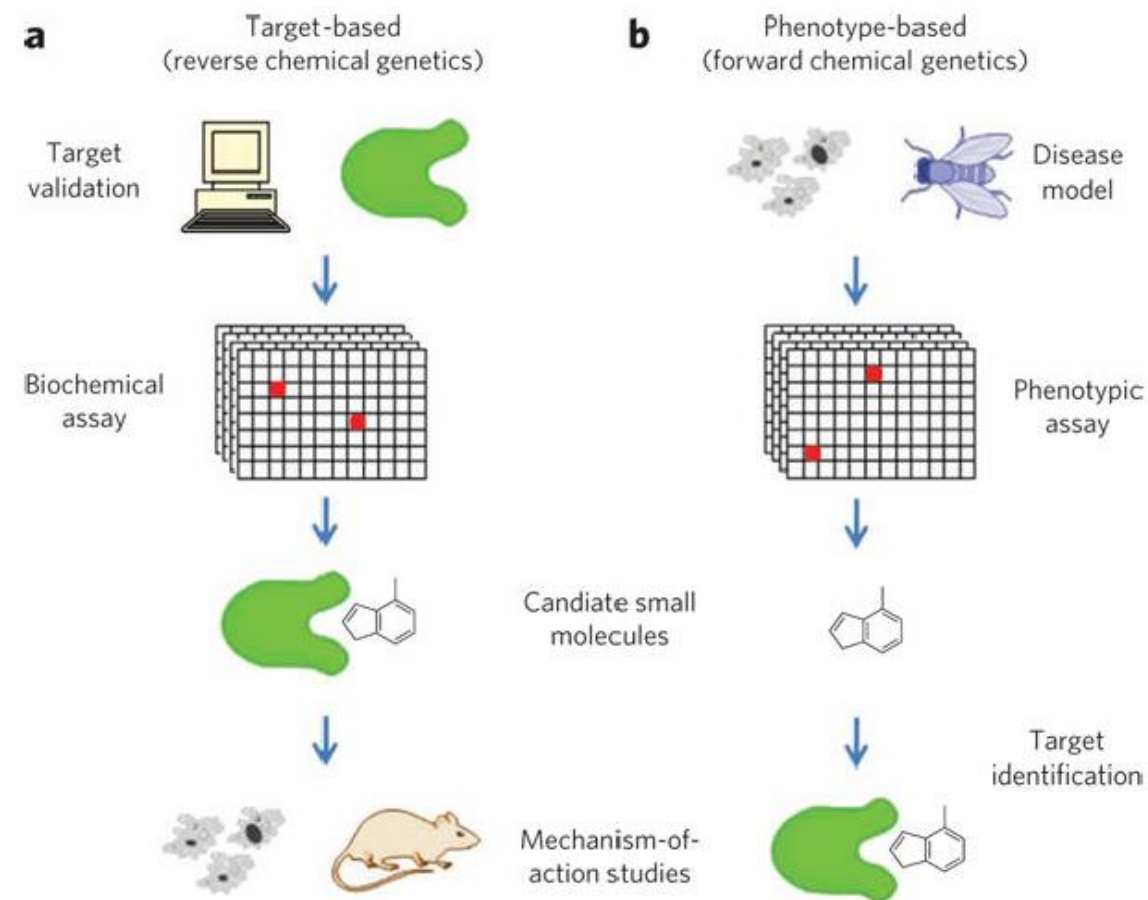
Identificação e validação do alvo biológico

- **Abordagem baseada no alvo**

- Um alvo é selecionado a priori e submetido a um processo de validação.

- **Abordagem baseada no fenótipo (ou na fisiologia da doença)**

- testa moléculas quanto ao seu impacto em processos biológicos ou modelo de doença.
 - experimentos em células, tecidos isolados, ou animais.
- mede a função celular, sem impor noções pré-concebidas sobre alvos ditos relevantes
 - possibilidade de descobrir novos alvos terapêuticos.



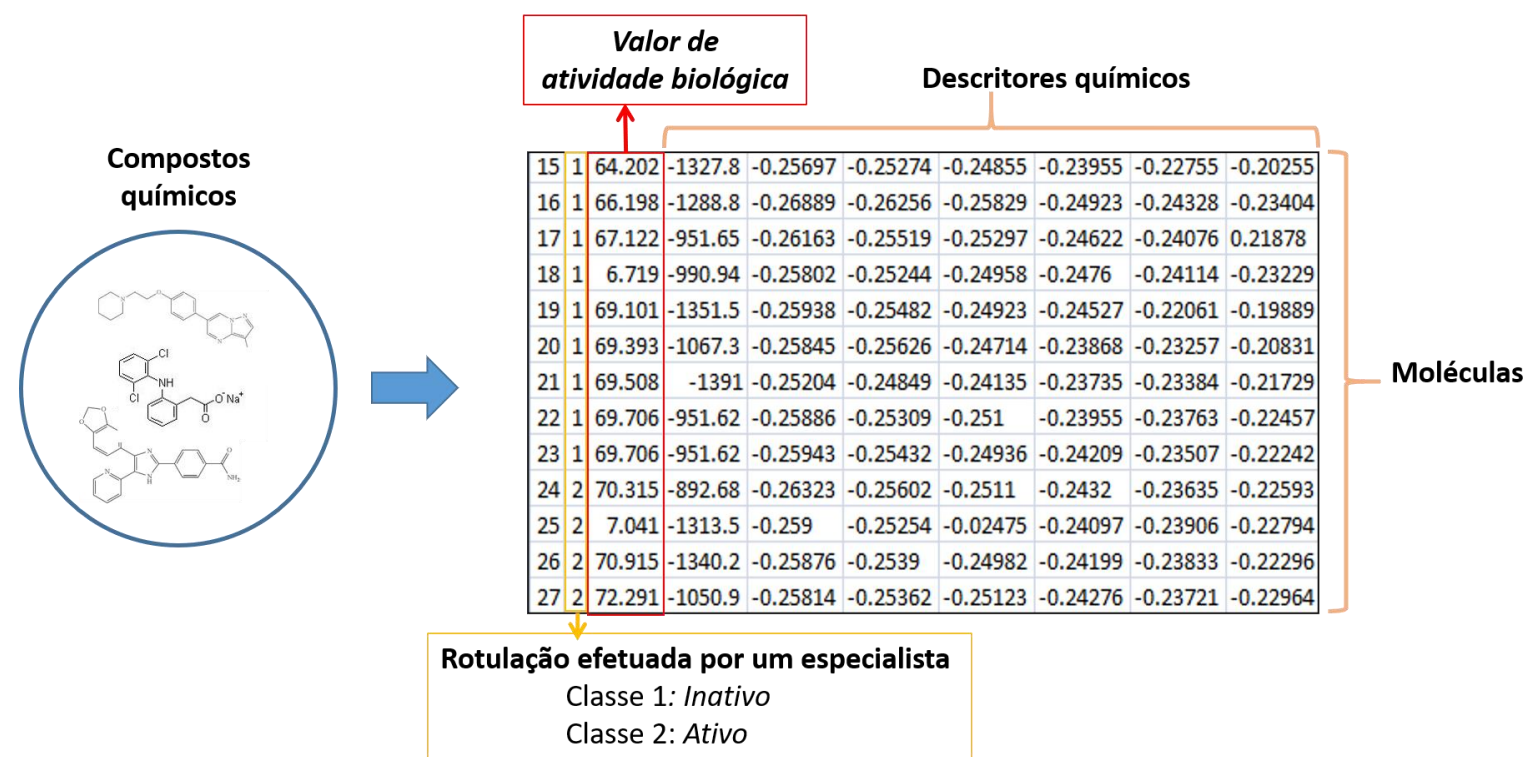
O processo de descoberta de medicamentos

- Para a identificação de candidatos a medicamentos, realiza-se a **triagem** de um grande número de moléculas.
- Objetivo: encontrar compostos que provocam uma resposta biológica desejável ao interagir com um alvo terapêutico no organismo.
- Conta-se com investimentos em síntese orgânica e em **testes biológicos**.



Estudos de relações quantitativas estrutura-atividade (QSAR)

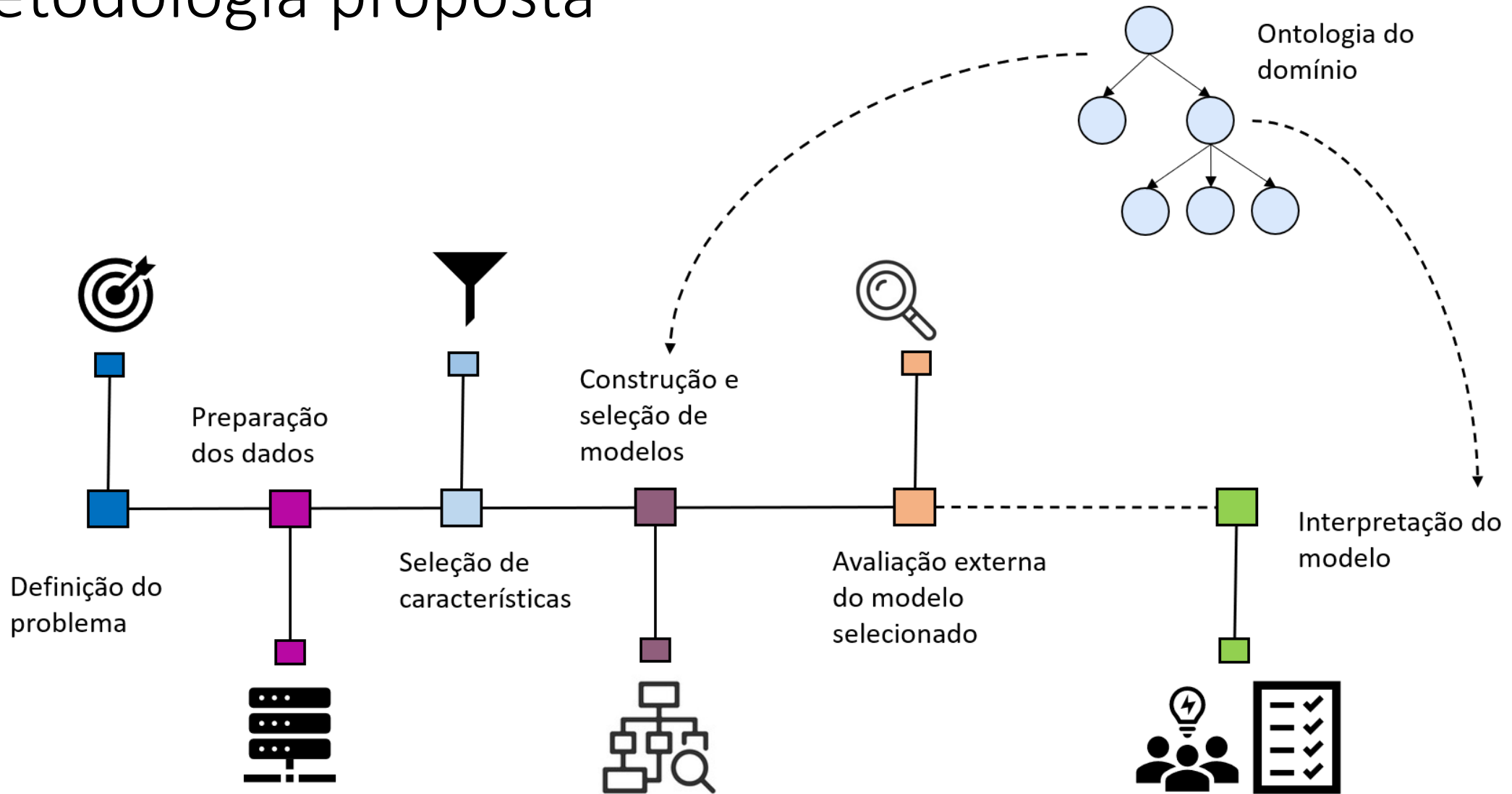
- Relacionam descritores químicos a valores de atividade biológica
- Conjunto de dados para análises QSAR:



Repositórios de dados e ferramentas de apoio

- **Bases de dados de ensaios biológicos:**
 - ChEMBL (BENTO et al., 2014)
 - BindingDB (GILSON et al., 2015)
 - PubChem (KIM et al., 2019)
- **Software para gerar descritores químicos (numéricos):**
 - Gaussian (FRISCH et al., 2009)
 - Padel (YAP, 2011)
 - Mordred (MORIWAKI et al., 2018)

Metodologia proposta



Principais estágios em que técnicas de AM são aplicadas

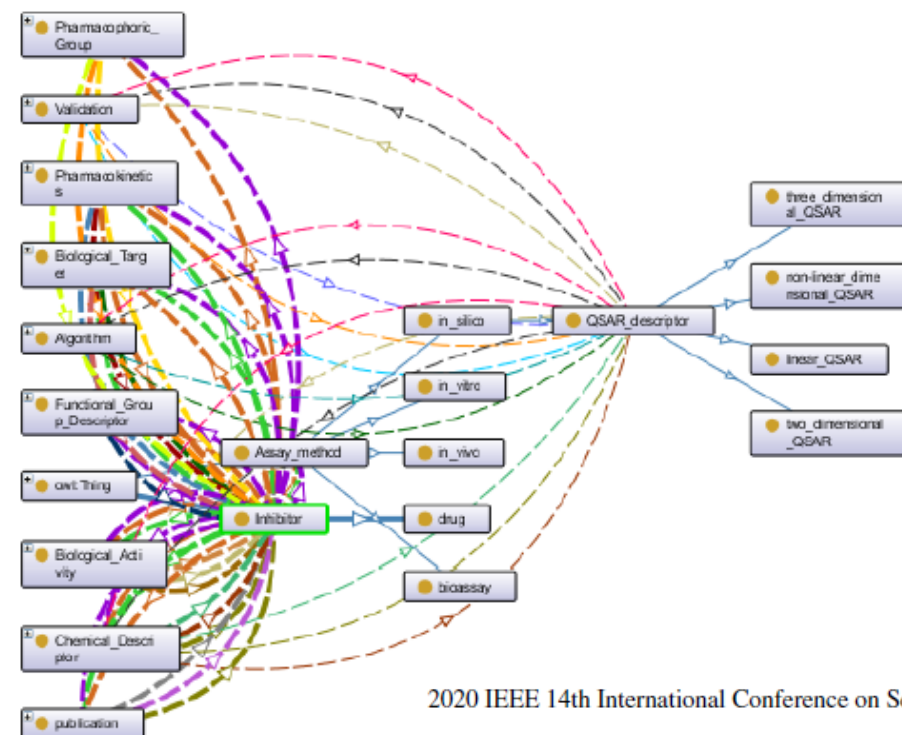
- Descoberta de alvos biológicos
- Identificação e otimização de compostos bioativos
- Análises pré-clínicas

Motivação para o desenvolvimento de uma ontologia para estudos QSAR

- Duas lacunas foram encontradas na literatura:
 - a falta de metadados para relacionar variáveis em diferentes conjuntos de dados de natureza química (GIRSCHICK; RUCKERT; KRAMER, 2013).
 - a falta de interpretabilidade de aplicações de AM na área de química medicinal (GRIFFEN et al., 2018).
 - é difícil identificar quais fatores (em termos químicos, quais subestruturas) estão conduzindo a estimativa gerada pelo algoritmo.

A ontologia OntoQSAR (ANGELO et al., 2020)

- Ontologia que integra conceitos químicos e biológicos para análises QSAR.
- Validação em um estudo de caso com compostos para o alvo ALK-5 (*activin receptor-like kinase 5*).
 - Um dos alvos mais investigados no combate ao câncer.



OntoQSAR: an ontology for interpreting chemical and biological data in quantitative structure-activity relationship studies

Motivação para o desenvolvimento de uma ontologia para estudos QSAR

- Limitações na manipulação de dados QSAR:
 - repositórios de dados públicos são quase sempre projetados de forma isolada.
 - dificuldades na combinação (e análise conjunta) de dados provenientes de diferentes fontes.
 - compostos químicos podem ser representados de modo diferente em repositórios distintos.
 - falta de um vocabulário controlado para descritores químicos.

Motivação para o desenvolvimento de uma ontologia para estudos QSAR

- **Ontologias nas áreas de Química e Biologia:**

- **Bioassay Ontology (BAO) - (VEMPATI et al., 2012)**

- representa o domínio do conhecimento sobre testes biológicos envolvendo compostos químicos.

- **Chemical Information Ontology (CHEMINF) - (HASTINGS et al, 2011)**

- representa estruturas químicas e descreve suas propriedades.

- **Chemical Entities of Biological Interest (ChEBI) Ontology - (DEGTYARENKO et al., 2008)**

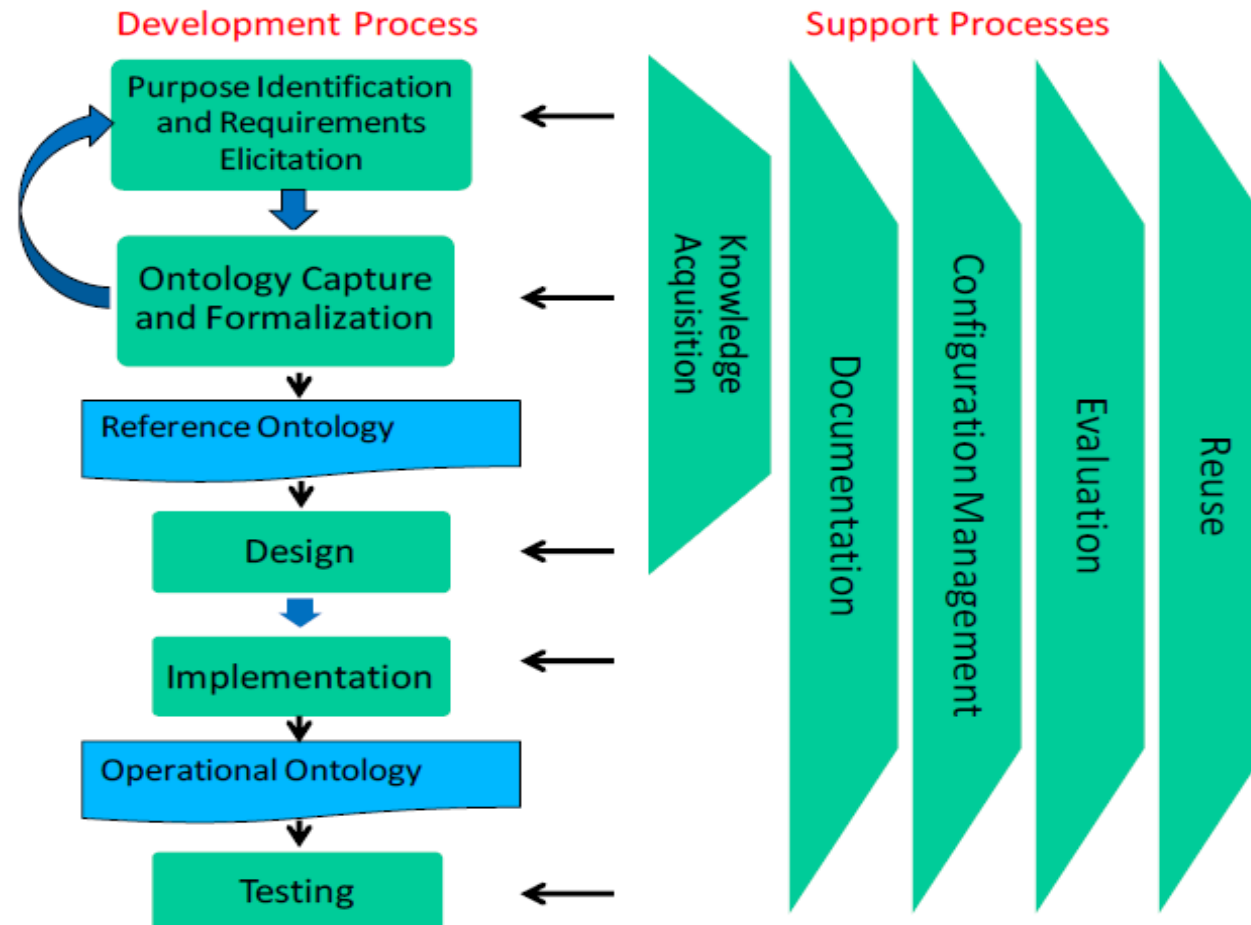
- contém conceitos acerca de dados estruturais para produtos naturais.

Motivação para o desenvolvimento de uma ontologia para estudos QSAR

- Todavia, o escopo dessas ontologias não incluem conceitos químicos e biológicos simultaneamente.
 - requerido para estudos QSAR.
- Por exemplo, CHEMINF e ChEBI não classificam compostos químicos de acordo com a atividade biológica que estes provocam em um alvo no organismo.

Metodologia SABiO

- SABiO: Systematic Approach for Building Ontologies (FALBO, 2014).



Metodologia SABiO

- Identificação dos propósitos da ontologia e levantamento de requisitos.
 - Visa a identificar os objetivos e possíveis aplicações da ontologia.
 - As **questões de competência** (CQs) são ferramentas essenciais para determinar esses requisitos.
- **Objetivos da ontologia OntoQSAR:**
 - representar o domínio QSAR, integrando várias bases de conhecimento;
 - criar um ambiente em que os pesquisadores possam obter facilmente informações para serem utilizadas em seus experimentos.
- A especialista no domínio QSAR identificou 60 CQs.
 - permitiu estruturar o primeiro desenho da ontologia envolvendo conceitos e propriedades relacionadas ao alvo ALK-5.

Metodologia SABiO

- **Questões de competência (CQs):**
 - Os requisitos de uma ontologia referem-se ao conhecimento (conteúdo) a ser representado e podem ser enunciados como questões de competência, ou seja, as **questões que a ontologia deve ser capaz de responder**.
 - As CQs são uma maneira eficaz de determinar o que é relevante para a ontologia e o que não é, definindo o seu **escopo**.
 - Além disso, fornecem um meio para **avaliação** da ontologia.
- A especialista no domínio QSAR identificou 60 CQs.
 - permitiu estruturar o primeiro desenho da ontologia envolvendo conceitos e propriedades relacionadas ao alvo ALK-5.

Desenvolvimento

- Especificação de requisitos da ontologia:
 - Uma especialista em QSAR identificou 60 Questões de Competência (CQs).

Questões simples

CQ1: Quais inibidores atravessam a barreira hematoencefálica?

CQ2: Quais inibidores já estão disponíveis comercialmente?

CQ3: Quais inibidores possuem o grupo funcional cetona?

CQ4: Quais inibidores foram estudados em modelagens QSAR 2D?

Complex Questions

CQ5: Em quais publicações encontra-se resultados de testes *in vivo* com propriedades farmacocinéticas calculadas?

CQ6: Quais inibidores foram testados no alvo biológico ALK-5 e estão na fase II?

CQ7: Quais inibidores têm dados sobre docking e testes *in vitro*?

CQ8: Quais inibidores violam a regra de Lipinski?

Desenvolvimento (observação)

- **Regra dos cinco de Lipinski:** uma molécula para ser um bom fármaco deve apresentar valores para 4 parâmetros múltiplos de 5:
 - log P maior ou igual a 5;
 - massa molecular menor ou igual a 500;
 - receptores de ligação de hidrogênio menor ou igual a 10;
 - doadores de ligação de hidrogênio menor ou igual a 5.

- SWRL (*Semantic Web Rule Language*) é a linguagem padrão para regras na web semântica).
- Todas as regras são expressas em termos de conceitos OWL (classes, propriedades, indivíduos).
- (<https://www.w3.org/Submission/SWRL/>)

Metodologia SABiO

- Coleta e formalização da ontologia
 - Define os termos que são considerados relevantes para o domínio e os categorizam em classes, subclasses, propriedades e indivíduos.
 - Essa etapa pode ser vista como uma modelagem conceitual que visa a representar o domínio com base nas CQs.
 - A utilização de linguagens expressivas, como a **OntoUML** ([https://ontouml.org/.](https://ontouml.org/)), é crucial para tornar a representação o mais próximo possível da realidade.

Metodologia SABiO

- Coleta e formalização da ontologia
 - **No caso da OntoQSAR:**
 - A construção da ontologia foi iniciada por meio da definição de conceitos relevantes no domínio de modelagem QSAR.
 - Pesquisando a literatura na área, foi criada uma lista de termos que representam conceitos, relações entre os termos e quaisquer propriedades que os conceitos possam apresentar.
 - Alguns desses termos foram incluídos na OntoQSAR pela reutilização de ontologias existentes, enquanto outros termos foram inseridos pelos próprios autores.

- 106 classes.
- 80 classes foram reutilizadas de ontologias existentes.

Ontology name (ID)	Number of reused classes
Bioassay Ontology (BAO)	19
Chemical Entities of Biological Interest (ChEBI)	13
Chemical Information Ontology (CHEMINF)	1
eNanoMapper Ontology (ENM)	31
Eagle-i resource ontology (ERO)	8
Gene Ontology (GO)	1
Information Artifact Ontology (IAO)	1
Kinetic Simulation Algorithm Ontology (KISAO)	1
Ontology for Biomedical Investigations (OBI)	1
Ontology for MIRNA Target (OMIT)	1
Physico-chemical methods and properties (FIX)	1
Physico-chemical process (REX)	1
Software Ontology (SWO)	1
Systems Biology Ontology (SBO)	1

Metodologia SABiO

- Coleta e formalização da ontologia
 - No caso da OntoQSAR:
 - A ontologia foi formalizada inicialmente utilizando a linguagem OntoUML e o software OLED (*OntoUML lightweight editor*), que é um ambiente para desenvolvimento, avaliação e implementação de ontologias.
 - Informações sobre essa ferramenta:
<https://ontouml.org/ontouml/tooling/>
 - Ao final desse processo, pode-se dizer que está criada uma **ontologia de referência**.

Metodologia SABiO

- Projeto da ontologia

- Etapa de desenho operacional da ontologia e implementação em uma linguagem computacional (por exemplo, OWL).
- A especificação conceitual da ontologia de referência é transformada em uma especificação de projeto:
 - questões de arquitetura e requisitos não funcionais.
- A mesma ontologia de referência pode ser usada para produzir vários projetos diferentes.

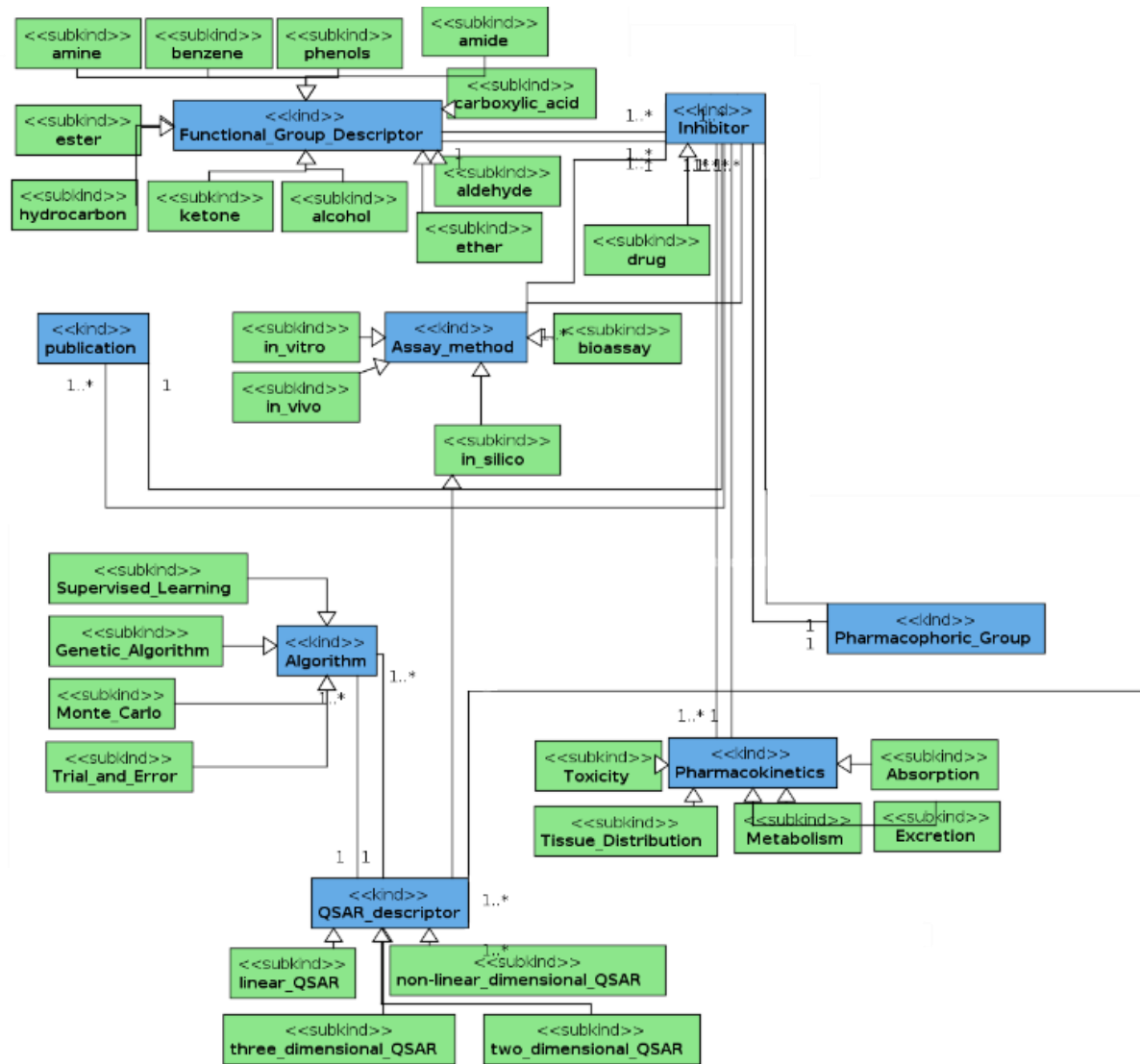
Metodologia SABiO

- Projeto da ontologia

- No caso da OntoQSAR:

- A ontologia de referência, construída em OntoUML, foi exportada para uma extensão OWL.
 - O software Protégé foi utilizado para manipular a ontologia a partir de então.
 - Por fim, as cardinalidades entre as classes e subclasses foram incluídos na ontologia.

OntoQSAR – Ontologia de referência



OntoQSAR – restrições de cardinalidade

- Restrições de cardinalidade para algumas relações entre conceitos:

Um inibidor possui pelo menos 1 e no máximo n grupos funcionais
Um grupo funcional pode estar presente em n inibidores

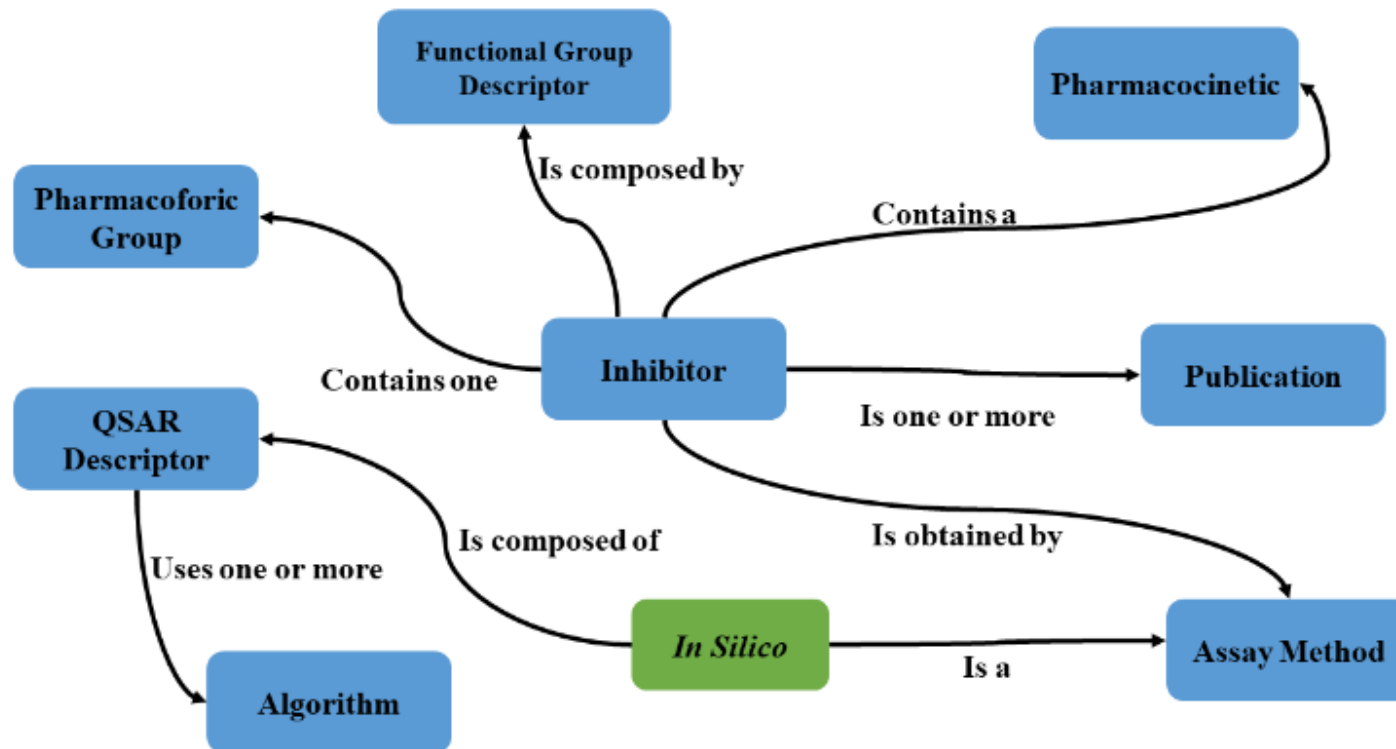
Um inibidor tem pelo menos 1 e no máximo n testes experimentais
Um teste experimental pode ter n inibidores

Um inibidor tem pelo menos 1 e no máximo n publicações
Uma publicação pode ter n inibidores

Um grupo farmacofórico possui pelo menos um e no máximo n inibidores
Um inibidor possui exatamente um grupo farmacofórico

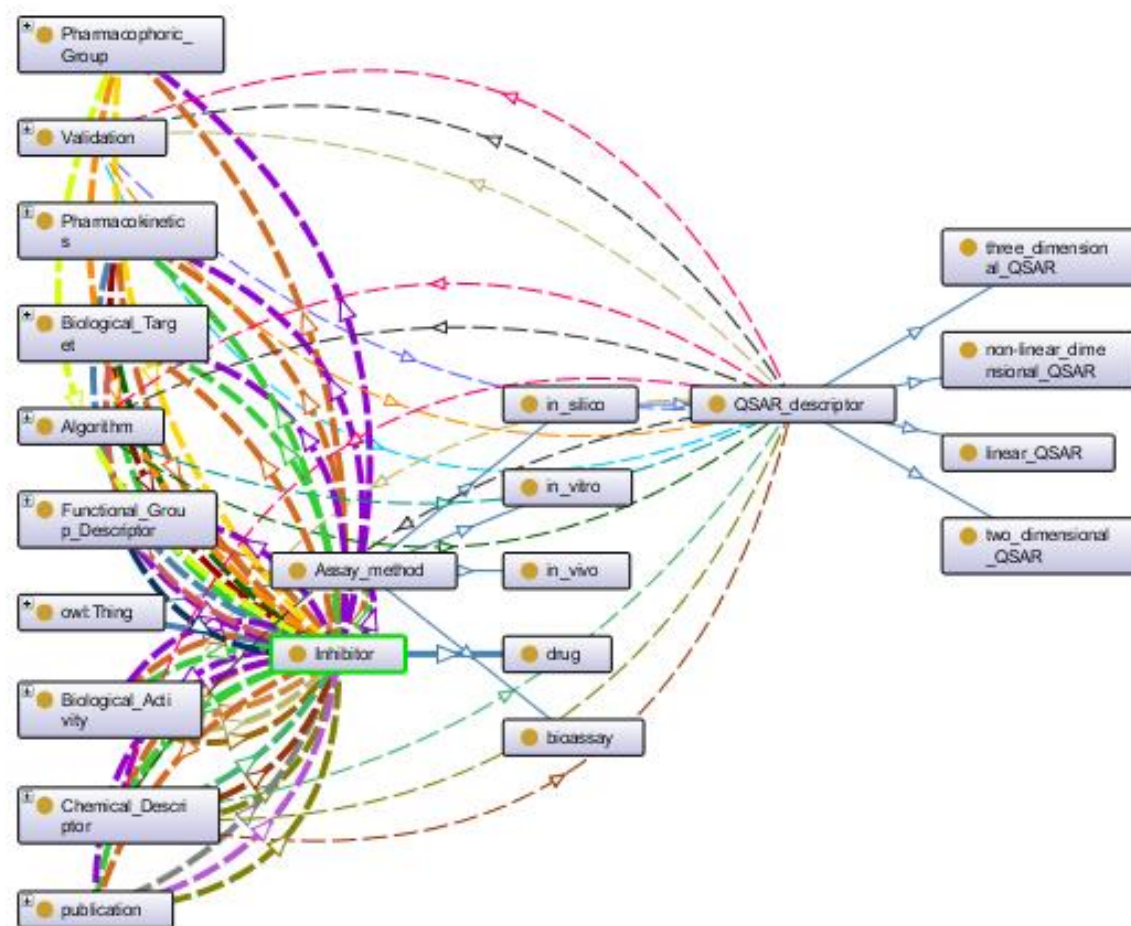
OntoQSAR – relacionamentos

- Alguns relacionamentos importantes:



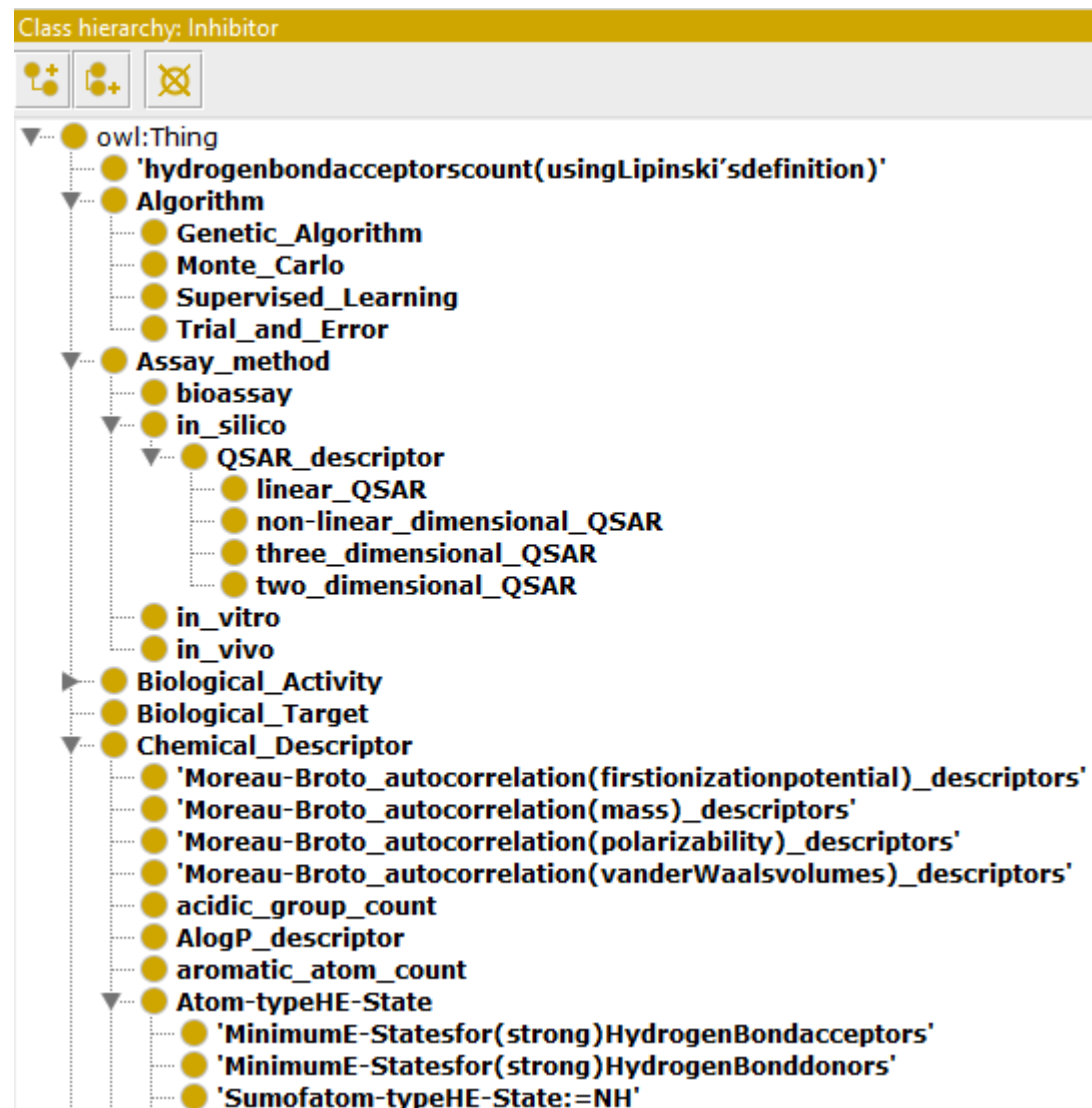
OntoQSAR – relacionamentos

- Um fragmento da ontologia OntoQSAR produzido pela ferramenta Protégé:



OntoQSAR – classes e subclasses

- Parte da estrutura de classes/subclasses produzida pela ferramenta Protégé:



OntoQSAR – validação

- O estudo de caso considerado teve como foco compostos químicos com impacto na atividade biológica do **receptor ALK-5**.
 - um dos alvos mais investigados em terapias de combate ao câncer.
- Inicialmente, foi realizada uma revisão de literatura para identificar estudos que apresentam dados sobre moléculas inibidoras do alvo ALK-5.
 - Identificação de conceitos químicos e biológicos relevantes, entre outros relacionados ao domínio QSAR.
 - Essa investigação resultou em **54 artigos** e **750 inibidores** (candidatos a medicamentos), encontrados nas bases de dados Web of Science (<https://www.webofknowledge.com.>) e PubChem (<https://pubchem.ncbi.nlm.nih.gov/>).

OntoQSAR – validação

- O estudo de caso considerado teve como foco compostos químicos com impacto na atividade biológica do **receptor ALK-5**.
 - um dos alvos mais investigados em terapias de combate ao câncer.
- Inicialmente, foi realizada uma revisão de literatura para identificar estudos que apresentam dados sobre moléculas inibidoras do alvo ALK-5.
 - Identificação de conceitos químicos e biológicos relevantes, entre outros relacionados ao domínio QSAR.
 - Essa investigação resultou em **54 artigos** e **750 inibidores** (candidatos a medicamentos), encontrados nas bases de dados Web of Science (<https://www.webofknowledge.com.>) e PubChem (<https://pubchem.ncbi.nlm.nih.gov/>).

OntoQSAR – validação (cont.)

- Em seguida, as estruturas de todos os inibidores encontrados foram construídas utilizando as ferramentas Avogadro Chems sketch e Open Babel.
- Os resultados foram registrados nos formatos 2D, 3D e SMILES para fosse possível usá-los como entrada para as ferramentas que calculam descritores químicos:
 - Gaussian, Padel, Mordred, pkCSM e SwissADME .
- No próximo passo, os dados de cada molécula foram associados ao valor correspondente de atividade biológica no alvo (estimado experimentalmente).
- Por fim, informações adicionais foram extraídas dos testes biológicos disponíveis, juntamente com as propriedades das técnicas aplicadas nas análises correspondentes.

OntoQSAR – validação (cont.)

- A ontologia OntoQSAR foi então populada com os dados gerados e avaliada com base em um conjunto de consultas em **linguagem SPARQL** para as CQs previamente definidas.
- **Exemplos:**

CQ1: Quais inibidores têm o valor de massa molecular inferior a 500?

CQ2: Quais inibidores violam a regra de Lipinski?

CQ3: Quais inibidores atravessam a barreira hematoencefálica?

OntoQSAR – validação (cont.)

- Consulta SPARQL para o item CQ1:

PREFIX

```
OntoQSAR:<http://www.semanticweb.org/matheus19/ontologies/2019/1/untitled-ontology-14#>
```

```
SELECT ?inhibitor ?chemical_descriptor ?subject  
?value_molecularMass
```

```
WHERE {?subject OntoQSAR:molecular mass ?value_molecularMass.  
?chemical_descriptor OntoQSAR:sub has the ?subject. ?inhibitor  
OntoQSAR:has_the ?chemical_descriptor.
```

```
FILTER (?value_molecularMass < 500.0)}
```

- Os resultados das consultas foram considerados **válidos e consistentes** por uma especialista do domínio.

Como escolher o método apropriado para agrupar dados de expressão gênica? (Santos, 2022)

- **Aplicação: obter insights sobre a morte celular induzida por progesterona.**
 - Acredita-se que a diabetes gestacional pode estar associada a esse fenômeno.
- **Cinco métodos de agrupamento foram testados.**
 - Dentre estes, o algoritmo **k-médias** obteve o melhor desempenho.

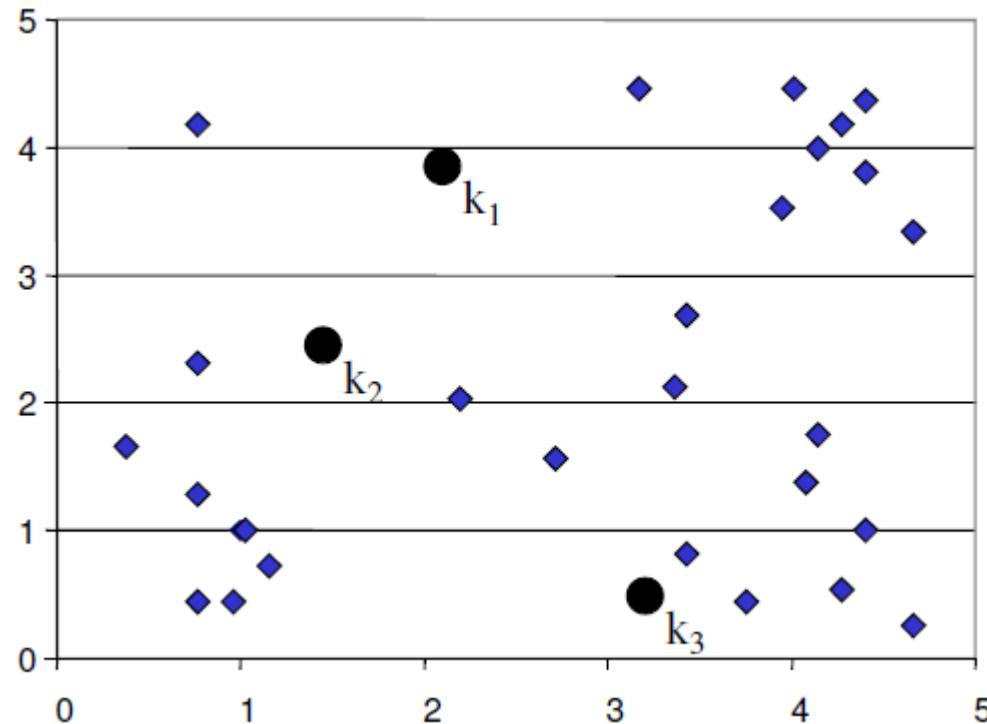
Como escolher o método apropriado para agrupar dados de expressão gênica? (Santos, 2022)

- **Experimentos de microarray:**

- 89 genes relacionados ao estresse oxidativo.
- células submetidas a concentrações de 0.1 μM , 1 μM e 100 μM e sendo os dados coletados em 6h e 24h a partir do início dos experimentos.
- grupo controle foi usado como referência.

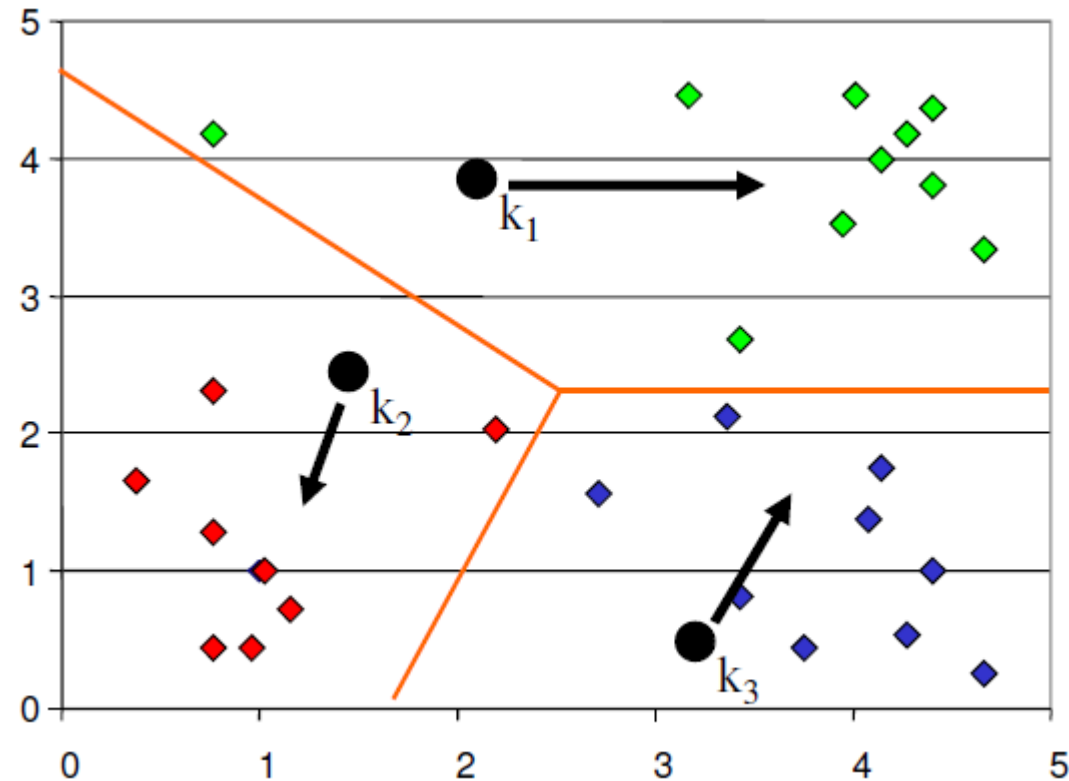
Algoritmo k-médias: inicialização

- Estabeleça o número de clusters (k) e escolha os protótipos (centróides) de cada cluster aleatoriamente.



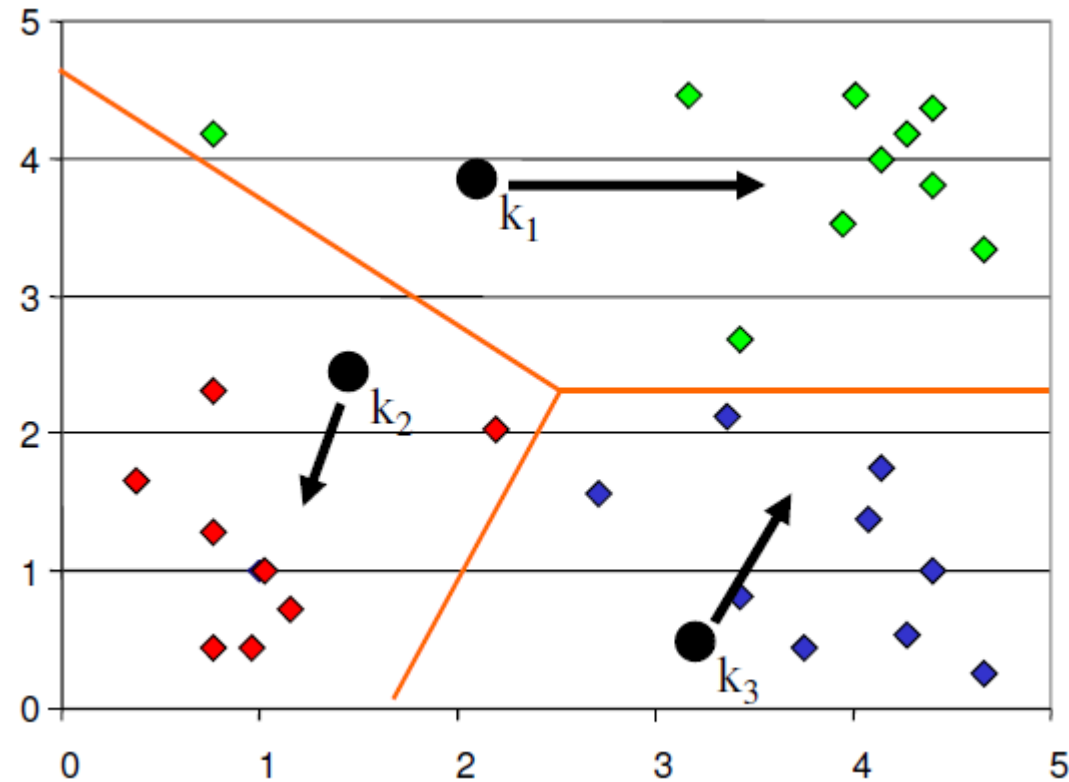
Algoritmo k-médias: iteração 1

- Calcule as distâncias entre objetos e os protótipos (k_1, k_2, k_3), encontrando clusters iniciais pela regra do vizinho mais próximo:



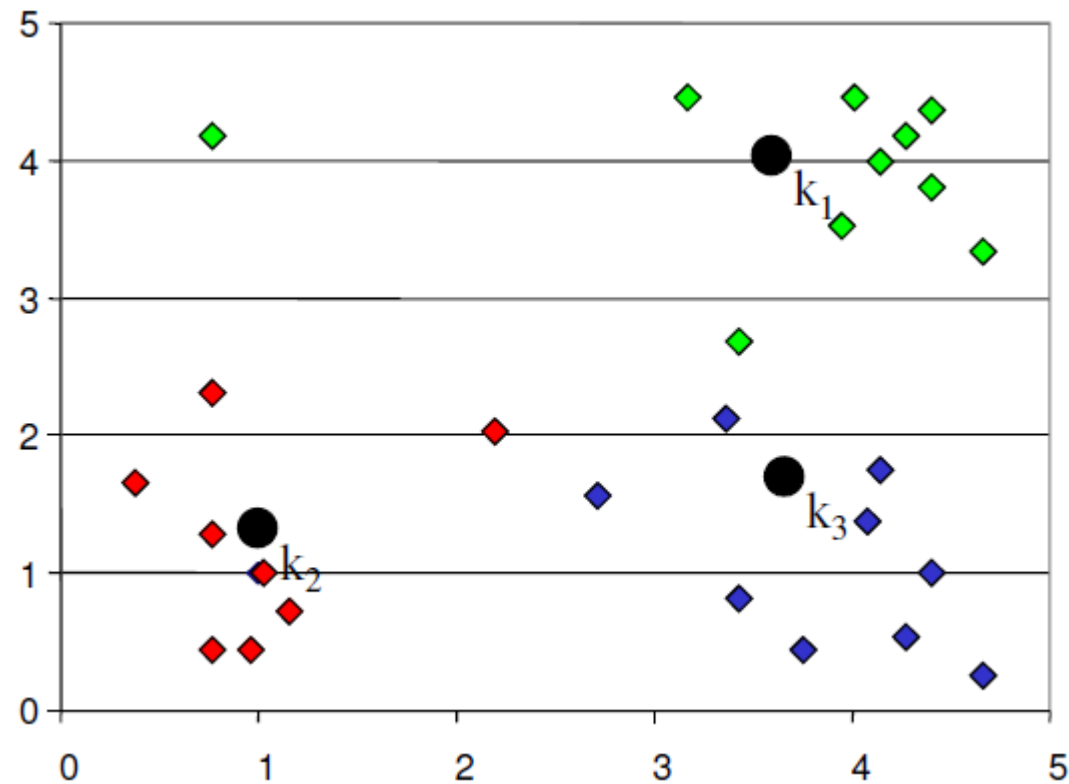
Algoritmo k-médias: iteração 1

- Depois, atualize os protótipos (k_1, k_2, k_3), calculando o vetor de médias para os objetos alocados em cada cluster.



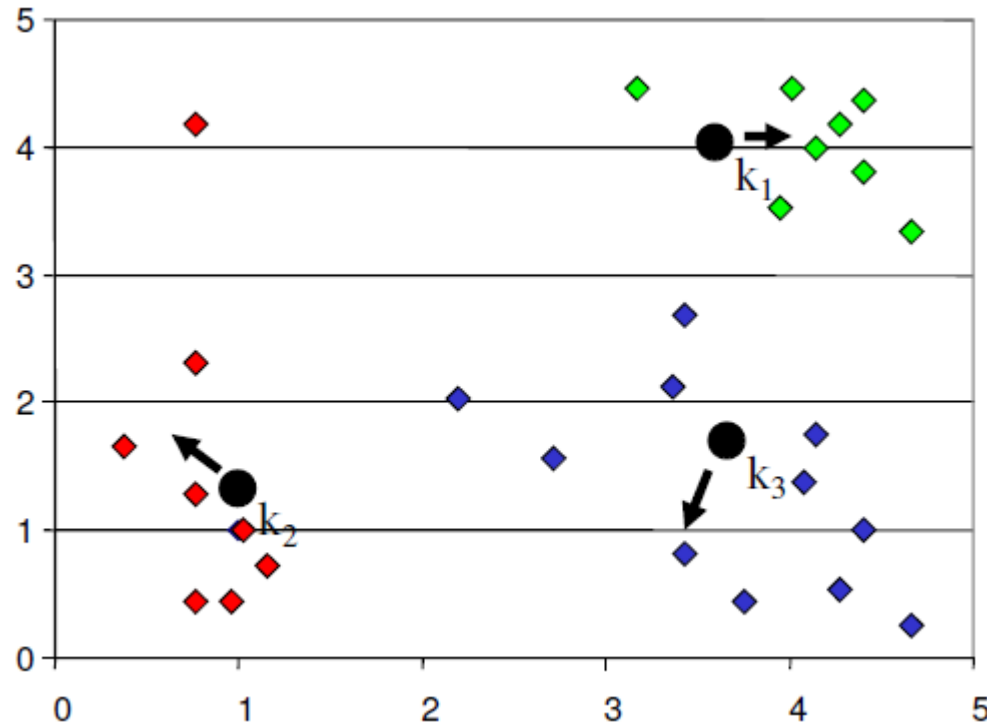
Algoritmo k-médias: iteração 2

- Realoque os objetos de acordo com a menor distância aos protótipos atualizados.



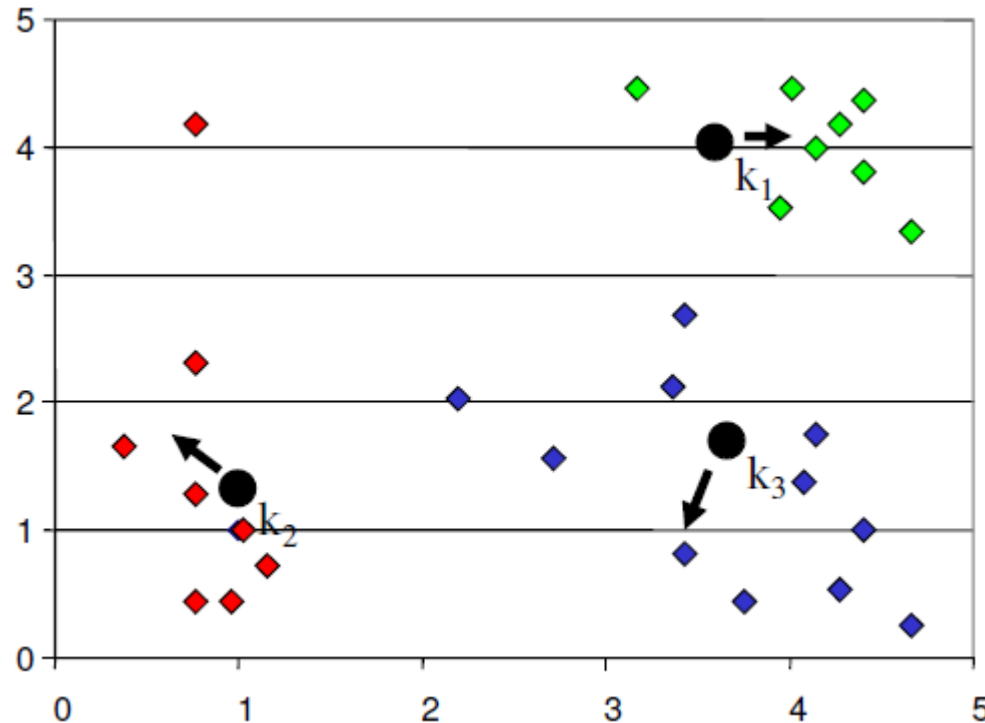
Algoritmo k-médias: iteração 2

- Depois, atualize os protótipos, calculando o vetor de médias para os objetos alocados em cada novo cluster.



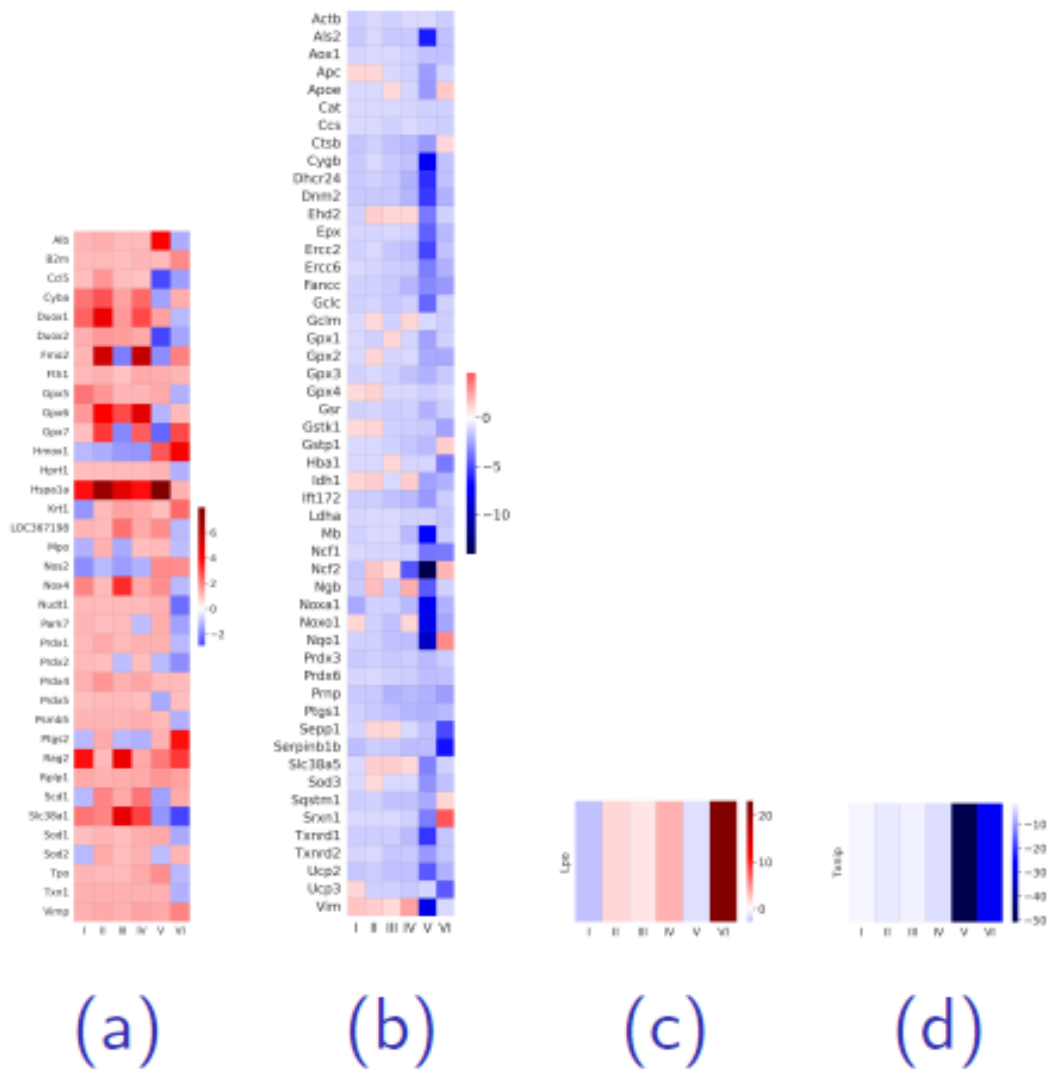
Algoritmo k-médias: término

- Prossiga, realocando os objetos e atualizando os protótipos até que não haja mais mudança na configuração dos clusters.



Como escolher o método apropriado para agrupar dados de expressão gênica? (Santos, 2022)

- Módulos identificados no experimento com melhor desempenho (k-médias com $k = 4$):



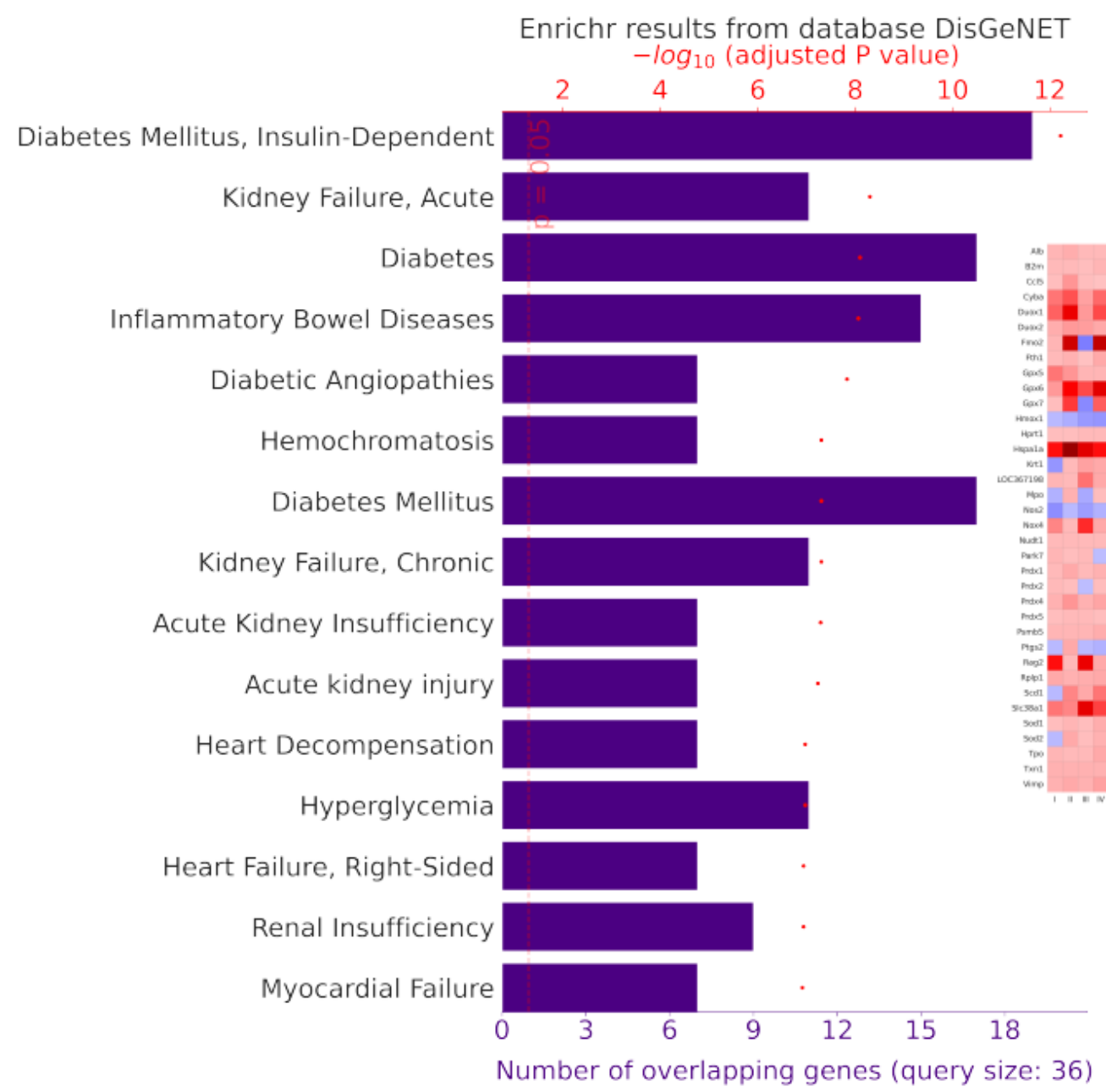
Análise de enriquecimento funcional

- **Anotações funcionais** transformam sequências genômicas em informações biológicas, integrando análises computacionais e dados biológicos auxiliares.
- **Análises de enriquecimento funcional** utilizam as anotações funcionais disponíveis em bases de dados e ontologias.
- Espera-se que genes com funções similares sejam alocados no mesmo grupo.
 - isso pode ser verificado comparando o resultado do agrupamento com o conhecimento disponível nas ontologias.
 - genes com funções desconhecidas provavelmente possuem papel similar aos dos genes (com funções conhecidas) no mesmo grupo.
 - permite obter insights sobre o fenômeno investigado.

Análise de enriquecimento funcional

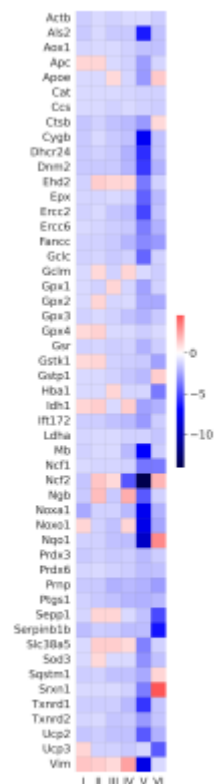
- **Ontologias investigadas em (Santos, 2022):**
 - Gene ontology
 - produz um vocabulário controlado sobre genes (e seus produtos e atributos) para todas as espécies.
 - DisGeNet ontology (<https://www.disgenet.org/rdf>)
 - ontologia para associações entre genes e doenças.
- **Enrichr** (<http://amp.pharm.mssm.edu/Enrichr>)
 - ferramenta web para análise de enriquecimento funcional.
 - fornece vários tipos de visualização para anotações funcionais de genes.

Resultados – módulo (a)



- **DisGeNet ontology:**
 - apresenta termos relacionados ao Diabetes tipo I e tipo II.
 - relação do estresse oxidativo em células beta-pancreáticas com a **Diabetes**.
- **Gene ontology:**
 - retornou termos como “resposta celular ao estresse oxidativo” e “resposta com via de sinalização mediada por citocinas”.
 - pode-se inferir que os genes deste módulo têm um papel na resposta ao estímulo da progesterona, ou um papel na **defesa antioxidante**.

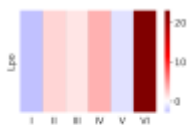
Resultados – módulo (b)



(b)

- DisGeNet ontology:
 - não retornou termos relacionados ao problema estudado.
- Gene ontology:
 - retornou termos como “regulação negativa do processo apoptótico”, o qual para, previne ou reduz a frequência, taxa ou extensão da morte celular.
 - Como os genes nesse módulo foram suprimidos, os mesmos não são capazes de evitar a morte celular.

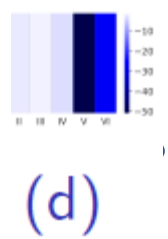
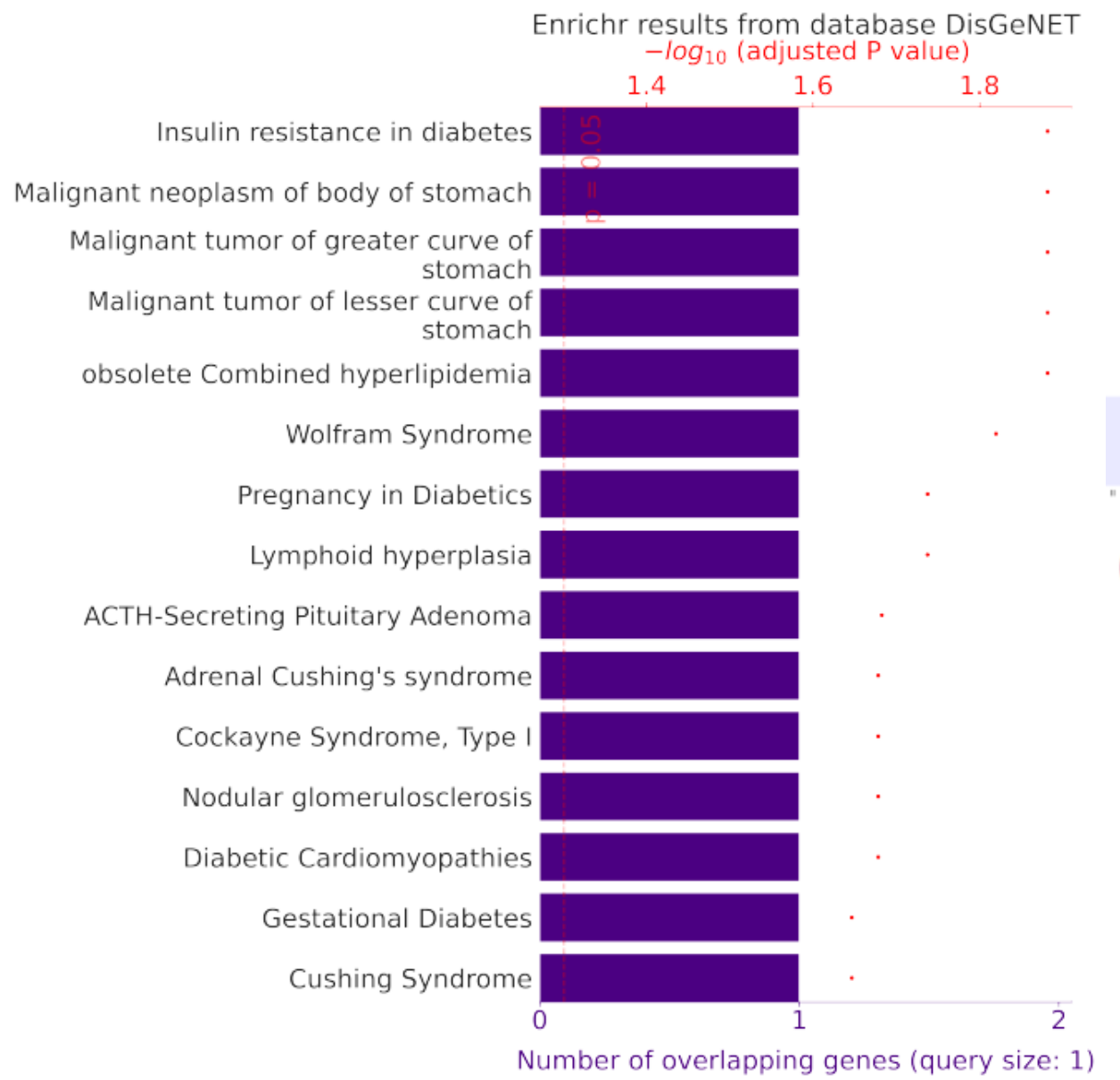
Resultados – módulo (c)



(c)

- DisGeNet ontology:
 - não retornou termos relacionados ao problema estudado.
- Gene ontology:
 - retornou termos como “regulação negativa da divisão celular”, que está relacionada a processos que interrompem, impedem ou reduzem a frequência da divisão celular.
 - Como esse gene está superexpresso, esse processo pode gerar desequilíbrio na massa de células beta-pancreáticas, fator que pode contribuir para o Diabetes Gestacional.

Resultados – módulo (d)



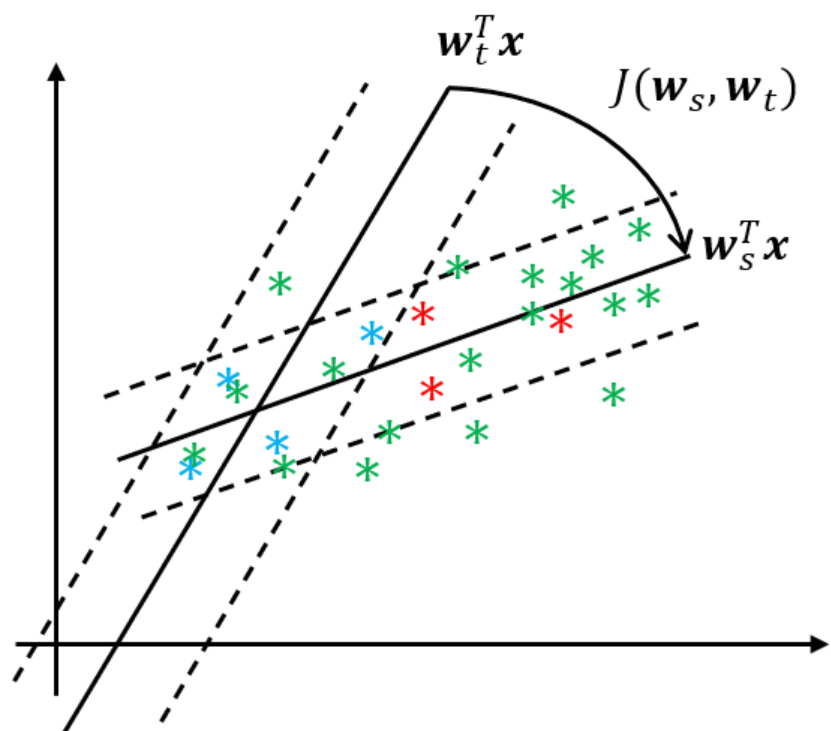
- DisGeNet ontology:
 - apresenta termos, como “gravidez no diabetes” e “diabetes gestacional”.
- Gene ontology:
 - demonstrou que este módulo também está associado à regulação negativa da divisão celular.
 - Isso pode contribuir para a preservação da massa de células beta-pancreáticas.

Motivação para a transferência de aprendizado em modelos QSAR

- A disponibilidade de uma grande quantidade de dados para construção de modelos QSAR é um desafio.
- Problemas quanto à qualidade das fontes de dados públicas, incluindo a representação imprecisa de estruturas químicas e informações incompletas sobre os ensaios biológicos de interesse.
- A natureza de diferentes protocolos experimentais pode levar a dados com diferentes distribuições de probabilidade, o que torna o uso de técnicas tradicionais de AM impraticável.

O método SVR-ATL (SIMÕES et al., 2018)

O termo regularizador $J(\mathbf{w}_s, \mathbf{w}_t)$ torna o modelo da tarefa de destino (\mathbf{w}_t) mais semelhante ao modelo da tarefa de origem (\mathbf{w}_s).



- * dados no domínio de destino
- * dados desconhecidos no domínio de destino
- * dados no domínio de origem



O método SVR-ATL (SIMÕES et al., 2018)

- A técnica SVR-ATL foi empregada para diferentes alvos biológicos e doenças, considerando pares (fonte, destino) de conjuntos de dados:
 - **Receptor ALK-5**: alvo muito utilizado para o tratamento do câncer.
 - **Receptor PPAR-alfa**: relacionado ao metabolismo de lipídios e carboidratos.
 - **Receptor PPAR-delta**: regula o metabolismo de lipídios e a resistência à insulina.
 - **Receptor DHFR**: associado ao tratamento do câncer e doenças inflamatórias.
 - **Enzima ciclooxygenase (COX)**: associada ao tratamento de doenças inflamatórias.
 - **Receptores do estrogênio (ER)**: relacionados ao tratamento de distúrbios hormonais.

Motivação para o desenvolvimento de uma ontologia para estudos QSAR

- Duas lacunas foram encontradas na literatura:
 - a falta de metadados para relacionar variáveis em diferentes conjuntos de dados de natureza química (GIRSCHICK; RUCKERT; KRAMER, 2013).
 - a falta de interpretabilidade de aplicações de AM na área de química medicinal (GRIFFEN et al., 2018).
 - é difícil identificar quais fatores (em termos químicos, quais subestruturas) estão conduzindo a estimativa gerada pelo algoritmo.

Referências

- ANGELO, R. M. et al. OntoQSAR: an ontology for interpreting chemical and biological data in quantitative structure-activity relationship studies. In: The 14th IEEE International Conference on Semantic Computing (ICSC). San Diego, USA. p. 203–206, 2020.
- BENTO, A. P. et al. The ChEMBL bioactivity database: an update. *Nucleic Acids Research*, v. 42, p. D1083 – D1090, 2014.
- FALBO, R.A.. SABiO: Systematic Approach for Building Ontologies. In: Joint Workshop ONTO.COM / ODISE on Ontologies in Conceptual Modeling and Information Systems Engineering (Co-located with 8th International Conference on Formal Ontology in Information Systems, FOIS 2014). Rio de Janeiro – RJ, Brazil. 2014.
- CARRARA, L. et al. Current mathematical models for cancer drug discovery. *Expert Opinion on Drug Discovery*, v. 12, n. 8, p. 785–799, 2017.
- DIMASIA, J. A.; GRABOWSKIB, H. G.; HANSEN, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, v. 47, p. 20–33, 2016.
- FRISCH, M. et al. Gaussian 09 (revision a02). Gaussian Inc. Wallingford CT, 2009.
- GILSON, M. et al. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic acids research*, v. 44, 2015.
- KIM, S. et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Research*, v. 47, n. D1, p. 102–109, 2019.
- LIPINSKI, C. F. et al. Advances and perspectives in applying deep learning for drug design and discovery. *Frontiers in Robotics and AI*, v. 6, p. 108, 2019.
- LIPINSKI, C. F. et al. A molecular modeling study of combretastatin-like chalcones as anticancer agents using PLS, ANN and consensus models. *Structural Chemistry*, v. 29, 2018.

Referências (cont.)

- MALTAROLLO, V. G. et al. Applying machine learning techniques for ADME-Tox prediction: a review. *Expert Opinion on Drug Metabolism & Toxicology*, v. 11, n. 2, p. 259–271, 2015.
- MALTAROLLO, V. G. et al. Advances with support vector machines for novel drug discovery. *Expert Opinion on Drug Discovery*, v. 14, p. 1–11, 2018.
- MORIWAKI, H. et al. Mordred: A molecular descriptor calculator. *Journal of Cheminformatics*, v. 10, 2018.
- OLIVEIRA, A. A. et al. New consensus multivariate models based on PLS and ANN studies of sigma-1 receptor antagonists. *Journal of molecular modeling*, v. 23, n. 10, 2017.
- RéDA, C.; KAUFMANN, E.; DELAHAYE-DURIEZ, A. Machine learning applications in drug development. *Computational and Structural Biotechnology Journal*, v. 18, p. 241–252, 2020.
- SERAFIM, M. et al. The application of machine learning techniques to innovative antibacterial discovery and development. *Expert Opinion on Drug Discovery*, v. 15, p. 1–15, 2020.
- SERAFIM, M. et al. Knowing and combating the enemy: a brief review on SARS-CoV-2 and computational approaches applied to the discovery of drug candidates. *Bioscience Reports*, v. 41, 2021.
- SHENONE, M. et al. Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol*. v. 9, n.4, p. 232–240, 2013.
- SIMÕES, R. et al. Transfer and multi-task learning in QSAR modeling: Advances and challenges. *Frontiers in Pharmacology*, v. 9, 2018.
- Santos, L. M. D. Análise de expressão de dados de expressão gênica e sua aplicação para o entendimento entre progesterone e diabetes gestacional. Dissertação de mestrado. EACH/USP, 2022.
- YAP, C. W. Padel-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry*, v. 32, p. 1466–74, 2011

Obrigada!