



Regulación génica y predicción

Guillermo Cruz Rojas
Santiago Buey Utrilla



Regulación de la expresión génica

Mecanismos principales mediante los cuales los genes regulan su actividad

Tres mecanismos biológicos fundamentales a nivel celular:

- Replicación del material genético y su transferencia a las siguientes generaciones.
- Síntesis de proteínas a partir de la información genética.
- Expresión de los genes en los cromosomas.



Definiciones

Expresión:

Proceso mediante el cual las células generan proteínas a partir de genes

Regulación:

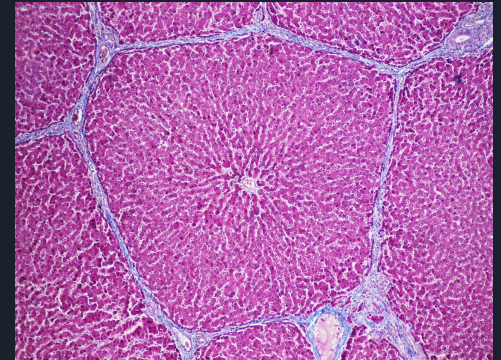
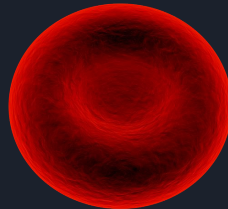
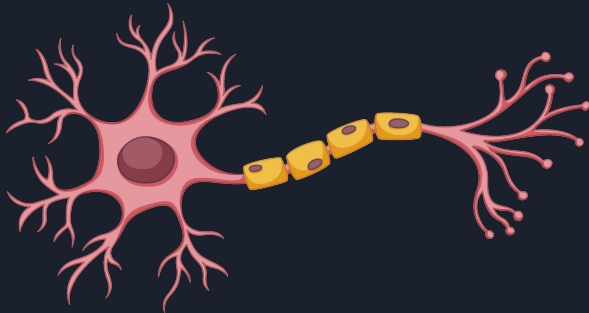
Proceso que controla qué genes en el ADN de una célula se expresan.

Regulación de la expresión génica

Todas las células contienen el mismo ADN

Proceso de activación y desactivación de genes

Decide que una célula sea:





Regulación de la expresión génica



Distintos tipos de regulación

Influencias externas: Cáncer

Mecanismos varían entre especies: Estudio de bacterias

Regulación: Células procariotas

Bacteria *Escherichia coli*

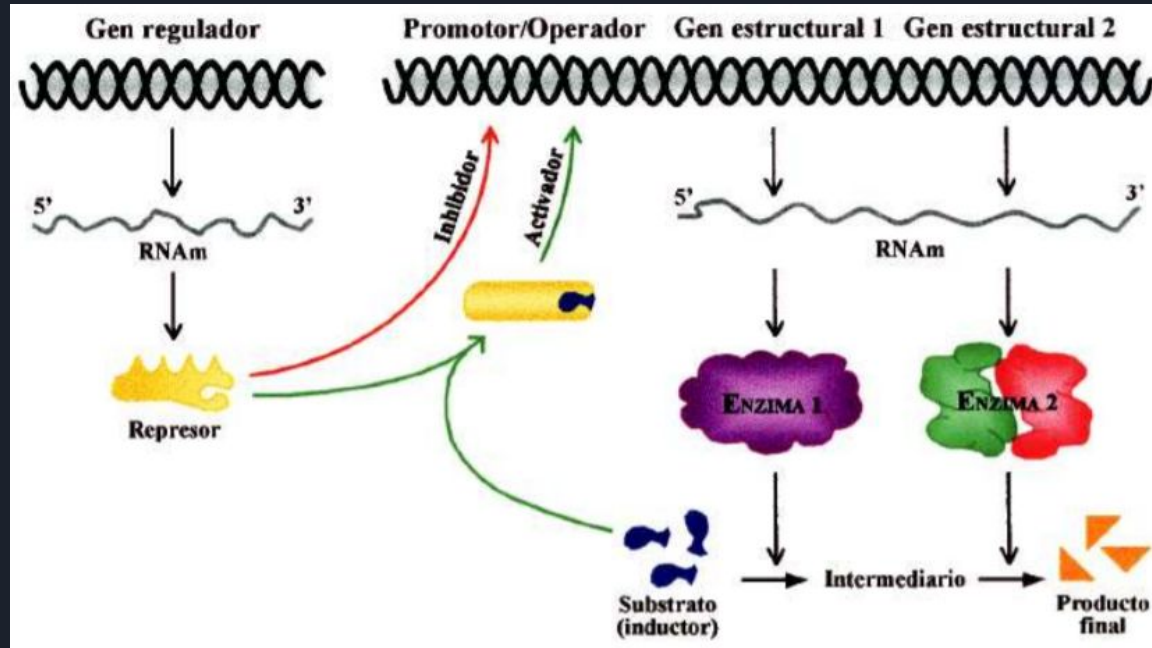
Genes procariotas se pueden transcribir como
ARNm

Expresión de los genes regulada a nivel de síntesis o
transcripción de ARNm



Regulación: Células procariotas

Mecanismos de regulación génica a nivel de transcripción: represión, inducción, activación, represión catabólica, terminación, antiterminación y atenuación.





Regulación: Células eucariotas

Diferencia principal con bacterias: Heterogeneidad

Regulación más complicada

La fisiología de cualquier célula está bajo el control de redes que regulan la expresión génica.

Proyectos de secuenciación



A pipeline to create predictive functional networks: application to the tumor progression of hepatocellular carcinoma.

M. Folschette et al. BMC Bioinformatics 21 (2020) [1]

- Flujo de trabajo para extraer el conocimiento existente en bases de datos biológicas.
- Generar predicciones computacionales a partir de dicho conocimiento relacionadas con el estado de expresión de las moléculas biológicas.
- Estudio de la progresión tumoral del carcinoma hepatocelular.



Motivación

Activación de la transición epitelio-mesénquima



Aumento de la agresividad del tumor



Motivación

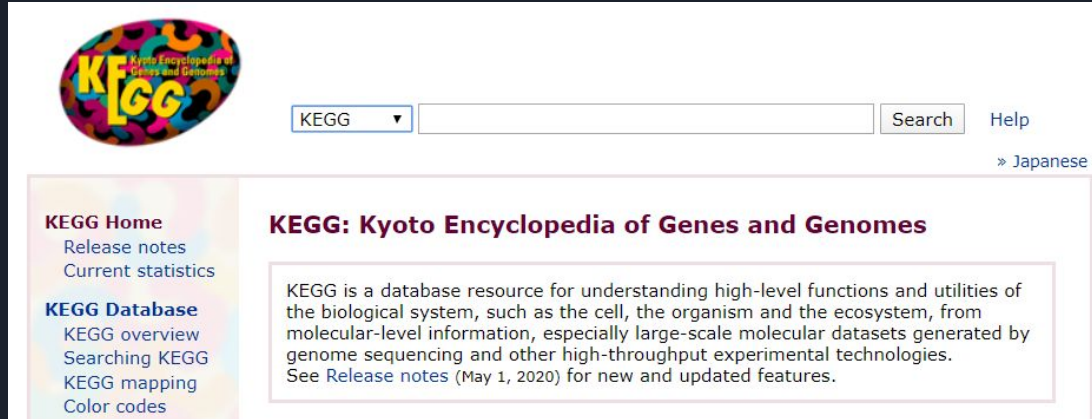
- ¿Cómo predecir el comportamiento de la enfermedad?
 - Rutas o caminos biológicos en las denominadas *pathway databases*.
- Problema:
 - No se ofrece un razonamiento lógico que permita aplicar inferencia.



Motivación

- Solución: Método *sign-consistency*.
- Modelo: 2 entradas.
 - Datos expresión génica (*International Cancer Genome Consortium [2]*)
 - Red causal de eventos regulatorios de distintos genes (Enciclopedia de Kyoto de genes y genomas [3])
- Resultados:
 - Grafo con 3.383 nodos y 13.771 relaciones. Predicción de 146 nodos.

Construcción de un grafo de interacciones etiquetado



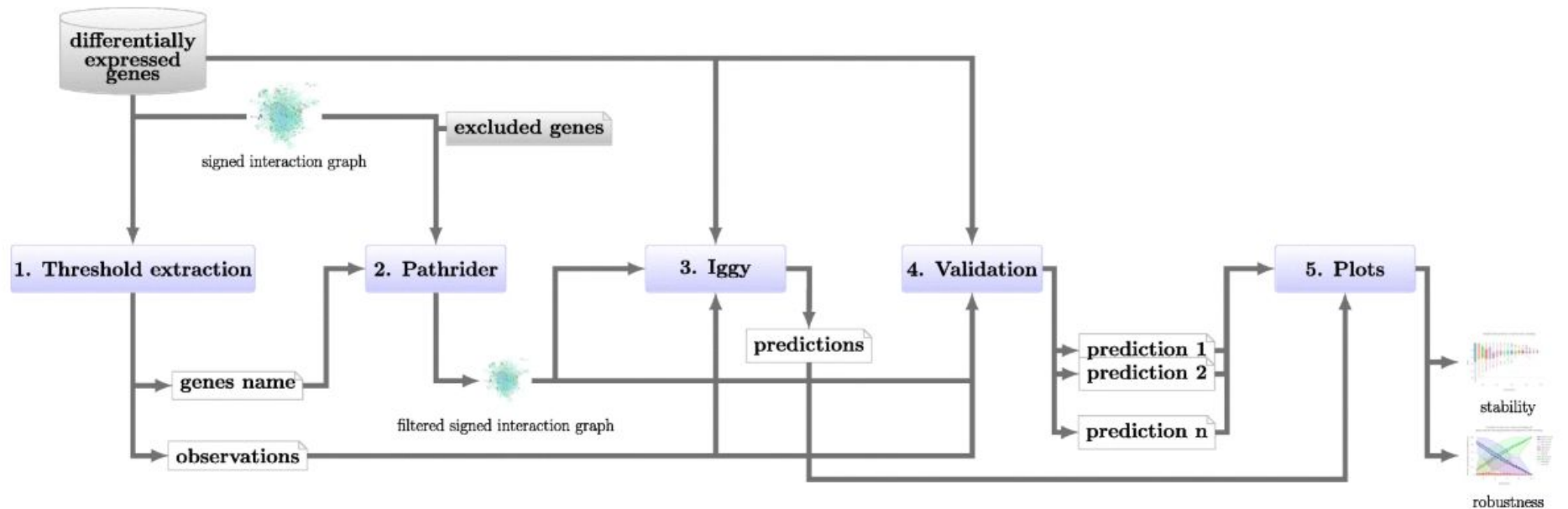
- Relaciones proteína-proteína. (BBDD)
- Relaciones de expresión génica. (BBDD)
- Relación gen-proteína. (Añadida)



Sign-consistency

- Relaciones: etiquetadas con signo $-$ o signo $+$.
- Nodos observados: etiquetados con $-$, 0 ó $+$.
- Principio *sign-consistency*:
 - Las observaciones deben mantener sus etiquetas iniciales.
 - Cada etiquetado $+$ ó $-$ debe estar justificado por al menos un predecesor.
 - Cada etiquetado 0 debe tener solo predecesores etiquetados como 0 o un par de predecesores etiquetados $+$ y $-$.
- Función de predicción a partir de etiquetas consistentes.

Resultados





Validación

- 146 predicciones.
 - 92 activaciones (+): 77 proteínas, 13 genes y 2 complejos proteicos.
 - 54 inhibiciones (-): 33 proteínas, 20 genes y 1 complejo proteico.
- 82 componentes predichos como (+) son coherentes con la base de datos *International Cancer Genome Consortium* y 8 no lo son.
- 44 componentes predichos como (-) son coherentes con la base de datos y 9 no lo son.
- Por tanto, 88% de predicciones correctas.



Sign-consistency. Función de predicción

- Relaciones: etiquetadas con signo - o signo +.
- Nodos observados: etiquetados con -, **0** ó +.
- Principio *sign-consistency*:
 - Las observaciones deben mantener sus etiquetas iniciales.
 - Cada etiquetado + ó - debe estar justificado por al menos un predecesor.
 - Cada etiquetado **0** debe tener solo predecesores etiquetados como **0** o un par de predecesores etiquetados + y -.



Discusión

Imperfección de los datos de los pacientes

Complejidad de los eventos a nivel molecular implicados en la progresión del carcinoma hepatocelular

Se requiere un modelo a gran escala





Discusión

Separación entre alta y baja agresividad

Predicción de los caminos regulatorios biológicos

Implementación de la herramienta Pathrider

Nodos del grafo etiquetados como gen o proteína



Discusión: Resultados

- Predicción del comportamiento de 146 componentes.
- 88% de las predicciones validadas por *Genome Consortium* y cross validation.





Conclusión

Modelo con resultados satisfactorios

Programación lógica

Aplicación del método a subconjuntos de pacientes

Generalizable a otros tipos de cáncer y enfermedades





Referencias

1. A pipeline to create predictive functional networks: application to the tumor progression of hepatocellular carcinoma. M. Folschette et al. BMC Bioinformatics 21 (2020).
<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3316-1>
2. International Cancer Genome Consortium.
<https://icgc.org/>
3. Enciclopedia de Kyoto de Genes y Genomas.
<https://www.genome.jp/kegg/>
4. Sistemas Regulatorios de la Expresión Génica. Daniel Mateos García. Universidad de Sevilla
<http://www.lsi.us.es/docs/doctorado/memorias/Memoria-Invest-DMateosGarcia.pdf>



FIN