

Obtención de conocimiento a partir de información génica utilizando procesos de minería de datos (Data Mining)

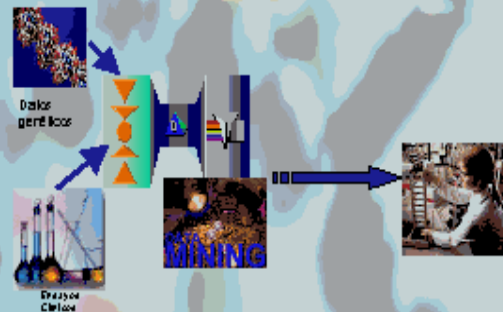
Drozdzowskyj A., Municio M. M., Sánchez J., Pueyo A.

La investigación genómica tiene una gran implicación en la Industria Farmacéutica, para el desarrollo de nuevos medicamentos. Uno de los mayores problemas al que se enfrentan es la habilidad para utilizar todos los datos disponibles ya que en el área genómica el volumen de datos es del orden de millones de registros. Por ello es crítico que los investigadores dispongan de las herramientas bio-informáticas capaces de transformar los datos en conocimiento útil.

De los datos al conocimiento

Un **Data Warehouse** específico para la investigación clínica permitirá el acceso rápido a toda la información disponible para el diseño de fármacos específicos adaptados al perfil genético.

Así mismo es necesaria la utilización de técnicas y herramientas analíticas capaces de trabajar sobre esa gran cantidad de datos.



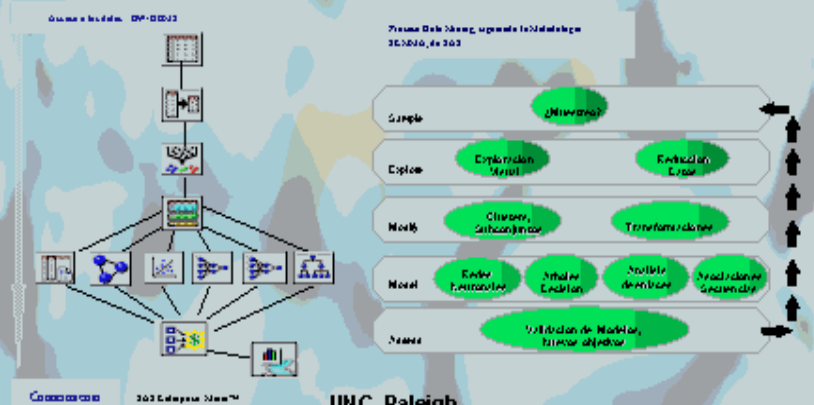
Data Mining es uno de los procesos más adecuados para el análisis y la exploración del genoma humano. La esperanza de esta investigación es que revele qué patrones genéticos causan enfermedades, con la posibilidad de encontrar como consecuencia tratamientos curativos aplicables a pacientes determinados y predecir el comportamiento de un fármaco en función del perfil genético de un individuo.

Metodología

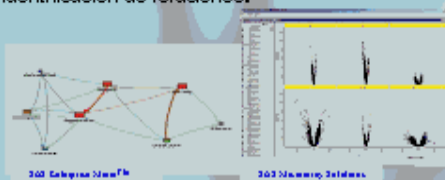
En particular, los científicos necesitan una **metodología** que les ayude en el proceso de Data Mining y un conjunto de herramientas analíticas específicas para el estudio del genoma humano.

El proceso Data Mining proporciona un método sistemático para el análisis de la secuencia genética lo que facilita la identificación de diferentes genes o marcadores de genes lo que puede sugerir por qué algunos individuos tienen mayor riesgo de desarrollar una determinada enfermedad o sufrir efectos adversos.

También es necesario disponer de técnicas analíticas que den lugar a **representaciones gráficas** de los datos, facilitando la identificación de relaciones.



Algunos resultados



UNC, Raleigh

El análisis de los datos de los SNP de pacientes afectados por **Alzheimer**, con procesos Data Mining ha permitido descubrir patrones genéticos asociados con la enfermedad.

Clasificación de Leucemia Aguda

La leucemia linfoblástica aguda y la leucemia mieloide aguda, presentan sintomatología similar. La distinción de los dos cánceres es crítica para un tratamiento eficaz. Se ha utilizado Data Mining para identificar los genes que diferencian claramente los dos cánceres – *asegurando un tratamiento adecuado de los pacientes*

Conclusiones: La investigación génica dispone en la actualidad de herramientas software y metodología de análisis que posibilitan un nuevo enfoque en el tratamiento de las enfermedades. Es fundamental que los investigadores y analistas de datos dispongan de soluciones Data Warehouse y Data Mining para el desarrollo de nuevas moléculas y compuestos que tengan un prometedor potencial terapéutico.