



Inteligencia Artificial () en la atención sanitaria del futuro

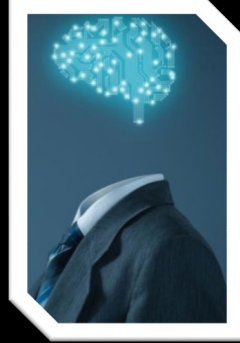


dr. florentino fdez-riverola
riverola@uvigo.es
<http://sing-group.org/>





Agenda



■ Informática Médica

- Definición, ámbitos
- Conocimiento + Capacidad de procesamiento (inteligente)

■ Inteligencia Artificial

- Definición de IA
- **strong** (sustituir) VS *weak* IA (facilitar)
- Técnicas y métodos comúnmente empleados

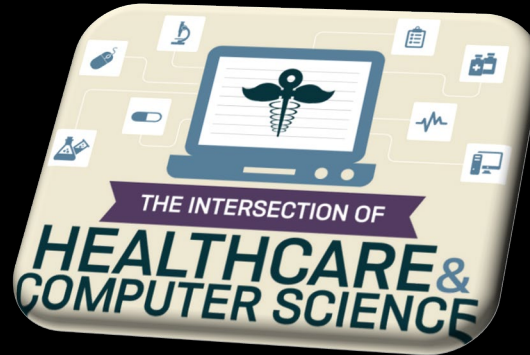
■ De la teoría a la práctica (IA en medicina)

- Informática clínica: *InNoCBR*, *PolyDeep*
- Informática para investigación clínica: *PathJam*, *WhichGenes*
- Bioinformática traslacional: *Genómica* (ALTER), *Proteómica* (DPD, BI, BS, Mlibrary, Mass-Up) e integración de datos ÓMICOS (PunDrugs)
- Informática en salud pública: *CURMIS4th*

■ Futuro

- IA interpretable y Human-In-The-Loop

■ Conclusiones



Informática Médica

As defined by the U.S. National Library of Medicine, **health informatics** is *the interdisciplinary study of the design, development, adoption, and application of IT-based innovations in healthcare services delivery, management, and planning.*



Informática médica

ámbitos

INFORMÁTICA APLICADA A LA SALUD DE LOS CONSUMIDORES

1 CONSUMER HEALTH INFORMATICS

Addresses informatics from the viewpoint of the consumer and patient.

FOCUSES ON:

- ✓ Personal health records
- ✓ Health information literacy
- ✓ Easily accessible resources that patients can both find and understand
- ✓ Patient advocacy and understanding

INFORMÁTICA EN SALUD PÚBLICA

2 PUBLIC HEALTH INFORMATICS

Applies population data and findings to public health.

FOCUSES ON:

- ✓ Prevention
- ✓ Preparedness
- ✓ Public health promotion
- ✓ Population health management

INFORMÁTICA CLÍNICA / OPERACIONAL

3 CLINICAL INFORMATICS

Also referred to as operational informatics, this area helps streamline the delivery of healthcare services.

FOCUSES ON:

- ✓ Clinical decision-making
- ✓ Healthcare delivery
- ✓ All technologies that "touch" the patient
- ✓ Content from clinical disciplines

INFORMÁTICA APLICADA A INVESTIGACIÓN CLÍNICA

4 CLINICAL RESEARCH INFORMATICS

Uses new data and discoveries to improve healthcare.

FOCUSES ON:

- ✓ Clinical trials
- ✓ Translational research
- ✓ New medical studies
- ✓ Providing high quality, good outcomes, and efficient care delivery

BIOINFORMÁTICA TRASLACIONAL

5 TRANSLATIONAL BIOINFORMATICS

Focuses on the use and storage of biomedical and genomic data.

FOCUSES ON:

- ✓ Understanding differences across seemingly similar patient populations
- ✓ Interpreting biological and clinical data to suggest efficient therapies
- ✓ Using clinical informatics as a foundation
- ✓ Using genomic data to predict and prevent certain health outcomes



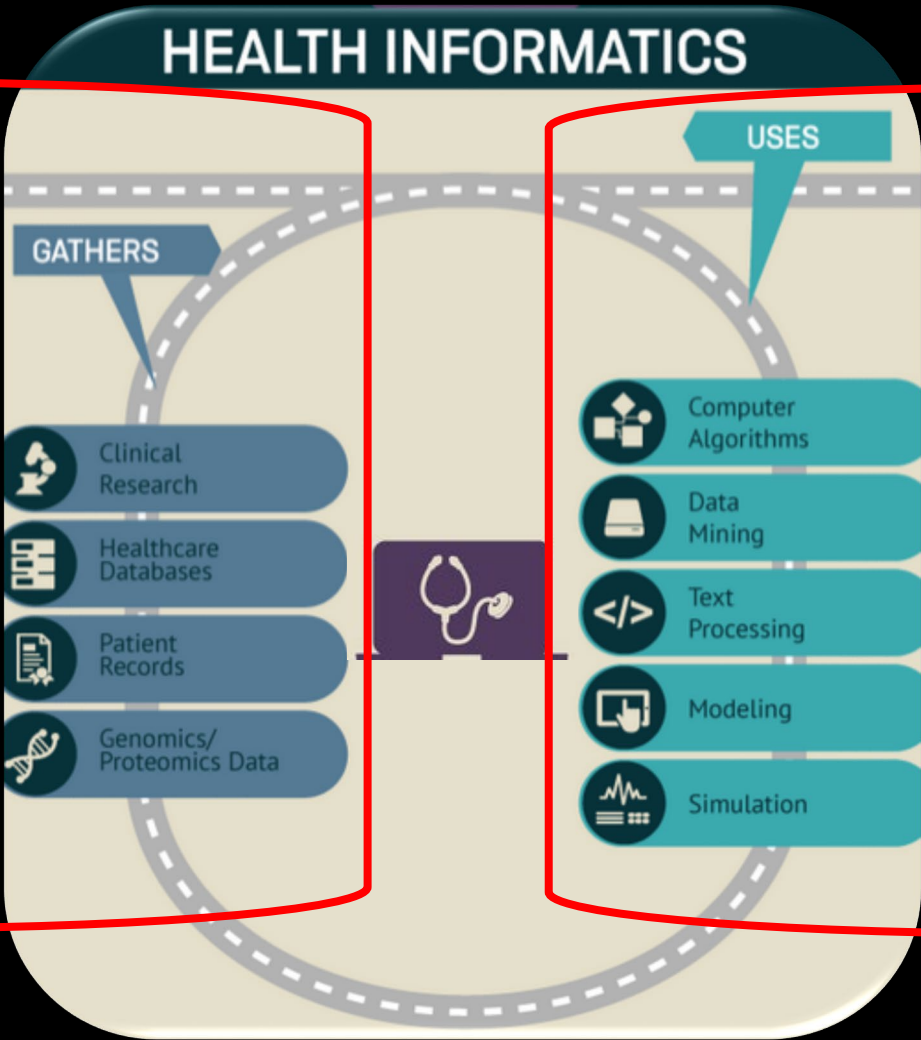
Informática médica

elementos constituyentes

INFORMÁTICA APLICADA
A INVESTIGACIÓN
CLÍNICA

INFORMÁTICA
CLÍNICA

BIOINFORMÁTICA
TRASLACIONAL



Conocimiento

bases de datos públicas

Genómicas

Secuencias

Entrez Nucleotide

DDBJ DNA Data Bank of Japan

EMBL NUCLEOTIDE SEQUENCE DATABASE

Genomas completos

Entrez Genome

e! Ensembl

UCSC

TIGR

Centradas en genes

Entrez Gene

GeneCards®

Bibliográficas (*biblioma*)

PubMed

Proteómicas

Proteínas

UniProt the universal protein resource

Entrez Protein

swissprot

TrEMBL TRANSLATED EMBL

PIR

Estructura de proteínas

PDB PROTEIN DATA BANK

SCOP STRUCTURAL CLASSIFICATION OF PROTEINS

CATH Protein Structure Classification

Dominios de proteínas

proSite

InterPro

Conserved Domains

Pfam

SMART

Datos experimentales

GEO

Gene Expression Omnibus

AE

ArrayExpress

Redes de interacción

Proteína-Proteína

STRING

IntAct

Rutas metabólicas

KEGG

Biocarta

R

PathwayInteractionDatabase

GenMAPP Gene Map Annotator and Pathway Profiler

Ontologías

the Gene Ontology

MeSH



Conocimiento

Historia Clínica Electrónica

Growth in healthcare data

1 exabyte = 1 billion gigabytes

2013
153
EXABYTES

2020
2,314
EXABYTES

Source: Stanford Medicine 2007, IDC 2014

To put that into perspective, data centers globally will only have enough room for an estimated **985 exabytes by 2020**—meaning that almost two and a half times this capacity would be required to house all the healthcare data.

Source: Cisco Global Cloud Index 2016

DATA STORAGE VS MEDICAL DATA (2020)

STORAGE CAPACITY	MEDICAL DATA GENERATED
985 EXABYTES	2,314 EXABYTES





Conocimiento

retos

Big Data

Big data refers to the mass of structured and unstructured data generated worldwide. In healthcare, this encompasses everything from electronic medical records, to internet-connected (IoT) devices.

It spans four dimensions:

VOLUME

Life sciences/pharma businesses are awash with patient-related data of all types, every day.

Artificial intelligence enables healthcare providers to parse through large amount of data and perform complex analytical tasks quicker and with greater accuracy.

VELOCITY

Waiting for answers from data troves tends to slow things down in a **time-sensitive industry**.

To maximize the value of big data, meaningful insights must be extrapolated as it streams in.

VERACITY

Establishing **trust and accuracy** in big data is imperative in the healthcare industry. Data assurance includes guaranteeing that data analytics are error-free and credible.

VARIETY

Big data comes from a myriad of sources, such as **IoT devices and social media**.

New and actionable insights can be found when various data types are analyzed together.



Source: IBM Industry Insights



Inteligencia Artificial

"AI is whatever hasn't been done yet."

Tesler's Theorem

strong | *weak*



IA

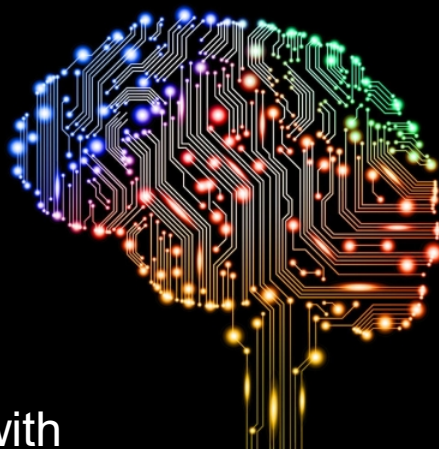
definición

"The automation of activities that we associate with human thinking, activities such as decision making, problem solving, learning."

Bellman, 1978

"The study of mental facilities through the use of computational models."

Charniak and McDermott, 1985



"The branch of computer science that is concerned with the automation of intelligent behaviour."

Luger and Stubblefield, 1993

"The study of how to make computers do things at which, at the moment, people are better."

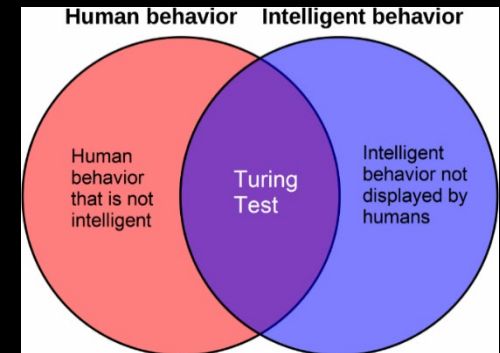
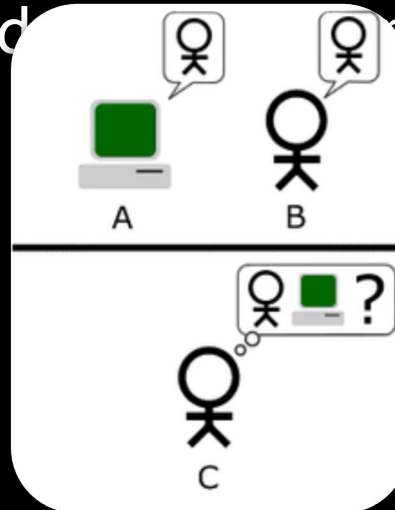
Rick and Knight, 1991

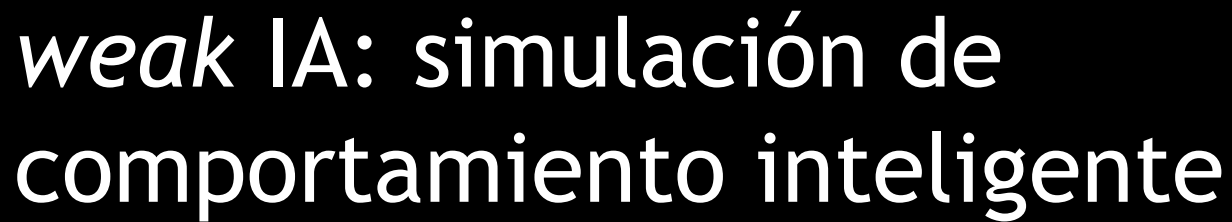


strong IA: aprendizaje y adaptación

trata de dar respuesta a ...

- ¿Puede una máquina actuar de forma inteligente?, ¿puede abordar/resolver *cualquier* problema que una persona resolvería pensando?
- ¿Son la inteligencia humana y computacional lo mismo?, ¿el cerebro humano se comporta esencialmente como un computador?
- ¿Puede una máquina tener ... mente, estados mentales y consciencia del mismo modo que los humanos?, ¿puede *sentir* cómo son las cosas?





■ Representación del conocimiento + Búsqueda

- [illegible]



IA

técnicas y métodos



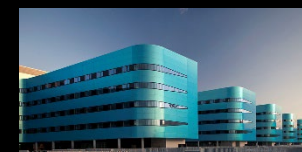
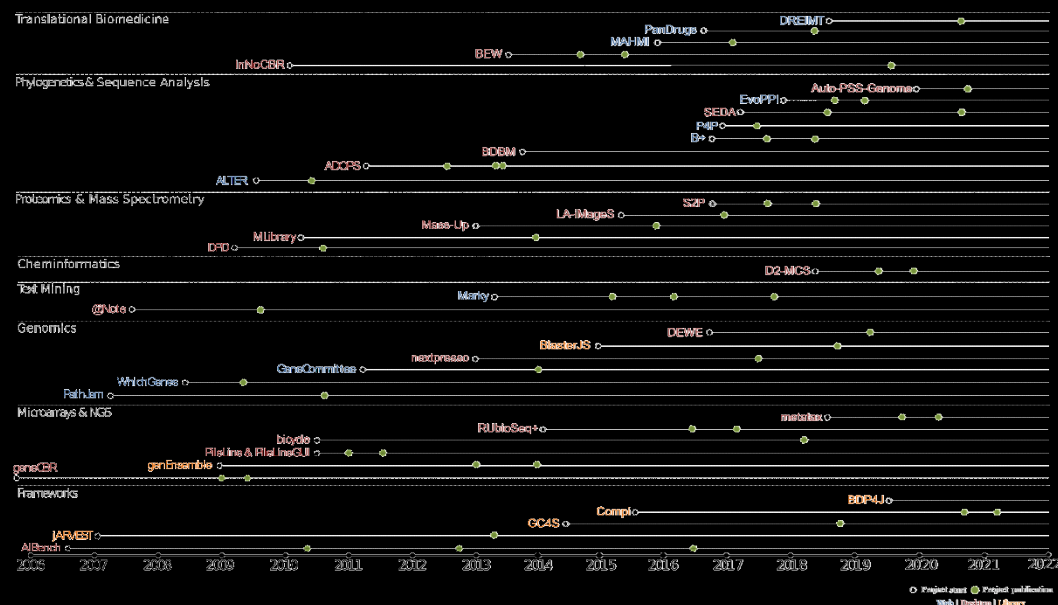
- **ANN: Artificial Neural Networks** - **DL: Deep Learning**
 - Problemas de predicción, clasificación y agrupamiento
- **CBR: Case-Based Reasoning**
 - Sistemas de apoyo al diagnóstico clínico y toma de decisiones
- **MAS: MultiAgent Systems**
 - Inteligencia ambiental y simulación de entornos sociales
- **FL: Fuzzy Logic** - **BN: Bayesian Networks** - **RS: Rough Sets**
 - Manejo de incertidumbre. Información incompleta y/o imprecisa
- **DW: Data Warehousing** - **DM: Data Mining**
 - Almacenamiento y análisis de grandes cantidades de datos (genómicos, proteómicos, etc.)
- **IR: Information Retrieval** - **TM: Text Mining**
 - Procesamiento automático de grandes corpus de documentos

Casos de estudio

de la teoría a la práctica



Next Generation Computer Systems Group
Sistemas Informáticos de Nueva Generación



INFORMÁTICA
CLÍNICA

3
**CLINICAL
INFORMATICS**

Also referred to as operational informatics, this area helps streamline the delivery of healthcare services.

FOCUSES ON:

- ✓ Clinical decision-making
- ✓ Healthcare delivery
- ✓ All technologies that "touch" the patient
- ✓ Content from clinical disciplines

INFORMÁTICA APLICADA
A INVESTIGACIÓN
CLÍNICA

2
**PUBLIC
HEALTH
INFORMATICS**

Applies population data and findings to public health.

FOCUSES ON:

- ✓ Prevention
- ✓ Preparedness
- ✓ Public health promotion
- ✓ Population health management

4
**CLINICAL
RESEARCH
INFORMATICS**

Uses new data and discoveries to improve healthcare.

FOCUSES ON:

- ✓ Clinical trials
- ✓ Translational research
- ✓ New medical studies
- ✓ Providing high quality, good outcomes, and efficient care delivery

BIOINFORMÁTICA
TRASLACIONAL

5
**TRANSLATIONAL
BIOINFORMATICS**

Focuses on the use and storage of biomedical and genomic data.

FOCUSES ON:

- ✓ Understanding differences across seemingly similar patient populations
- ✓ Interpreting biological and clinical data to suggest efficient therapies
- ✓ Using clinical informatics as a foundation
- ✓ Using genomic data to predict and prevent certain health outcomes

"By better integrating big data analytics into healthcare, the industry could save \$300bn a year – that's the equivalent of reducing the healthcare costs of every man, woman and child by \$1,000 a year."



Informática clínica





Finanzamento:



XUNTA
DE GALICIA



Fondo Europeo de
Desenvolvemento
Rexional
"Un selo de facer Europa"

UNIÓN EUROPEA

Institucións participantes:

Universidade de Vigo

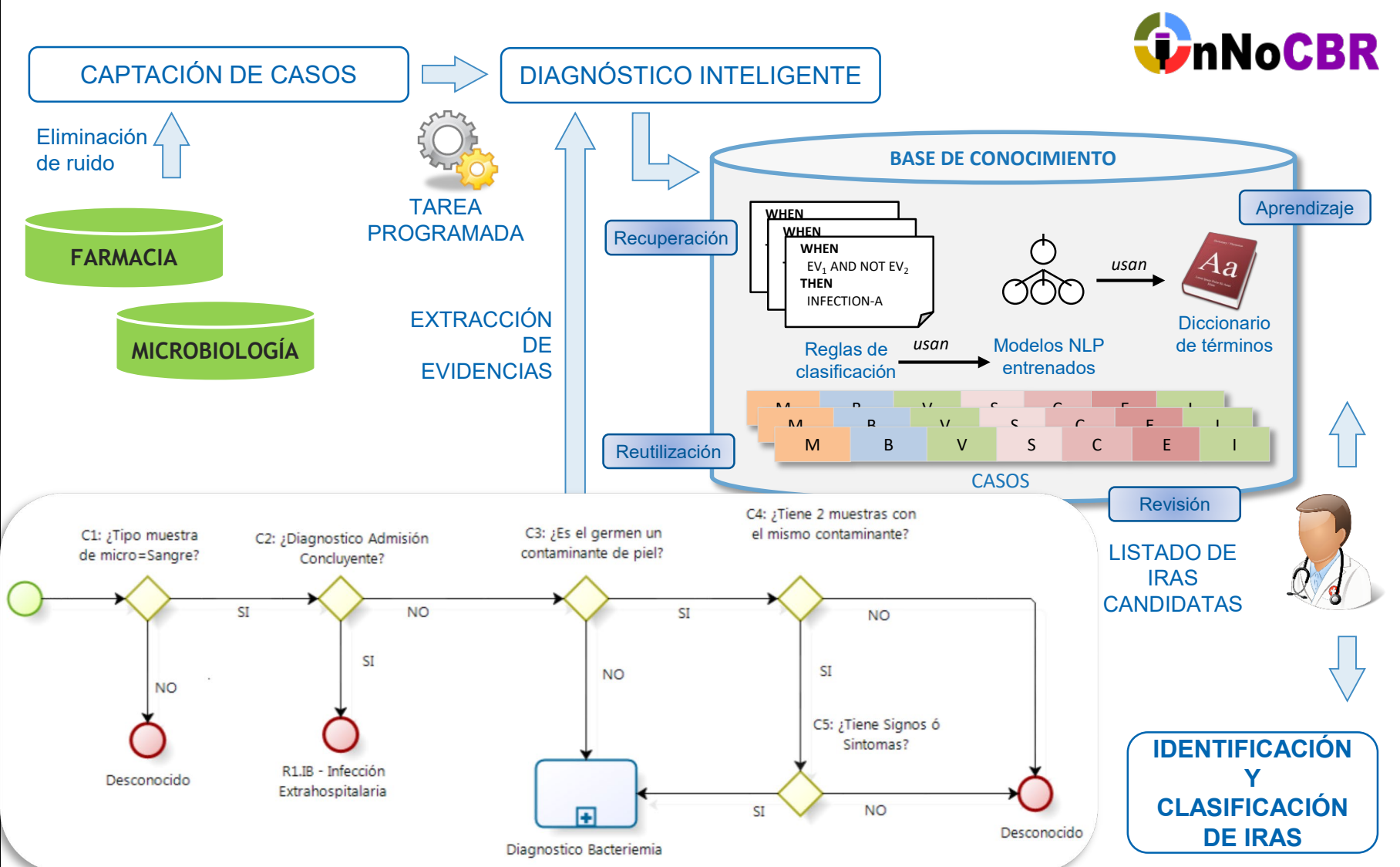


SERVIZO
GALEGO
de SAÚDE

Complexo Hospitalario Universitario
de Ourense



- Sistema inteligente de búsqueda y clasificación de casos de Infección Nosocomial (IRAS)
- Proyecto coordinado UVigo-CHUO financiado por la Xunta de Galicia [2010-2013] con 143.602,25 €
- Después de 4 años de desarrollo y 6 de implantación y mejora, InNoCBR es marca comercial (2017) y forma parte de la cartera de aplicaciones corporativas del SERGAS
- CHUO - alto impacto económico, sanitario y social:
 - Capacidad media de 800 pacientes
 - 1122 análisis de microbiología / día (2.5 millones en total)
 - 181 prescripciones de antibioticos / día (3.5 millones en total)
 - 3000 comentarios de enfermería / día (12.5 millones en total)
- Validación 2016: Gold Standard de 938 posibles casos de IRAS
 - Automático: 70.83% Sen. - 97.76% Esp. - 77.24% VPP - 0.67 Kappa
 - Semi-automatico: 81.73% Sen. / 99.47% Esp. - 94.33% VPP





- Sistema CBR (razonamiento basado en casos) que combina
 - motor de reglas (casos donde el experto proporciona conocimiento)
 - extractor automático de reglas (obtiene conocimiento de casos previos)
- Utiliza NLP (procesamiento de lenguaje natural)
 - análisis automático de comentarios de enfermería para valorar si contienen indicios de infección
- Algoritmos de clasificación
 - Naïve Bayes, J48, k -NN
- Capacidad de aprendizaje
 - a partir de los diagnósticos del personal especializado
- Automatización total de captación de casos y evidencias
- Unificar en todos los hospitales intensidad de búsqueda y criterios
 - comparación de indicadores de forma fiable

MEDICINA
Los españoles sufren dolor crónico

El 23% de los españoles afirman que sufren dolor crónico diagnosticado por un médico, según una encuesta realizada por la Organización de Consumidores y Usuarios entre más de 2.400 personas. De las respuestas obtenidas se desprende que el dolor es una parte frecuente o constante diaria del 45 % de la población de este país. EFE

GERIATRÍA
Aumentan los abusos a ancianos en Japón

El número de casos de abusos físicos y verbales a ancianos registrados en Japón aumentó el año pasado en más de mil nuevas denuncias, hasta la cifra récord de 16.700 nuevas denuncias, según el Ministerio de Sanidad del país asiático. Del total de casos registrados, el 63 % fueron de carácter físico y el 39 % verbal. En muchos casos, las víctimas sufren demencia. EFE



El santuario de la Virgen de Lourdes. **DESEO VILLAR**

LOURDES
Otras dos curaciones inexplicables

El santuario mariano de Lourdes probó dos nuevas curaciones que el comité médico de ese lugar de culto declaró «inexplicables» en el actual estado del conocimiento científico y que datan de 1963 y 1968. Estos nuevos casos son de clínic a hipertensión arterial. Hasta ahora, la Iglesia ha reconocido oficialmente 68 curaciones milagrosas. EFE

ESPAÑA
El 42 % tienen más de un factor de riesgo cardiovascular

El 42 % de los españoles padecen más de un factor de riesgo cardiovascular (hipertensión, sobrepeso, diabetes, tabaquismo o sedentarismo) y el 57 % sufren una de ellas. La combinación de diversos factores hace que las posibilidades de sufrir una cardiopatía aumenten enormemente, advierte la Federación Española del Corazón. EFE

Un programa pionero reducirá las infecciones hospitalarias

El CHUO ourensano sufre hasta ocho episodios por cada cien pacientes

PERO NI NÚMERO GARCÍA
CIENCIA / LA VOZ

Uno de los grandes problemas de la hospitalización son las infecciones nosocomiales (del latín nosocomium, hospital de enfermos). Muchos pacientes las sufren durante su ingreso, pese a que son evitables. En ocasiones tienen graves consecuencias.

De ahí que el servicio de medicina preventiva del CHUO ourensano y la Escuela Superior de Ingeniería Informática de la Universidad de Vigo se hayan puesto de acuerdo para desarrollar un sistema inteligente capaz de identificar todos los tipos y generar una base de datos que las clasifique.

Ese sistema accederá a todas las fuentes de datos para recoger e integrar las múltiples variables e identificar las evidencias. Tras el proceso de clasificación ofrecerá resultados de una manera clara y exhaustiva. Este proyecto es inédito en España y supondrá una revolución por el ahorro de dinero y de tiempo que implica.

Desde el servicio de medicina preventiva del CHUO, su responsable, Berta Uriel y La Torre, dice que las infecciones hospitalarias afectan aproximadamente a 5 o a un 6 % de los pacientes que ingresan. «Este tipo de infección supone casi el 30 % de los efectos adversos que puede implicar la hospitalización», explica la doctora, que añade: «Las consecuencias van desde prolongar la estancia entre 5 y 6 días de media (de 48 horas a varios meses, según los casos) a tratamientos complejos, reintervenciones quirúrgicas e incluso la muerte. El coste es de millones de euros ca año».

Las infecciones derivan en 60 millones de euros al año en el CHUO.

Campañas preventivas

Explica Uriel que solo con vigilar las cifras de infección y comunicárselas de manera periódica a los sanitarios que atienden a los pacientes los casos disminuyen un 20 y un 30 %. «Pero si además se hacen programas de control y campañas preventivas, infecciones como las septicemias por el uso de catéteres pueden erradicarse», precisa. Estas infecciones hospitalarias se producen en los centros gallegos con una frecuencia de 6 a 8 episodios por cada 100 pacientes ingresados. En el caso del CHUO, que tiene 700 camas,



Las infecciones derivan en 60 millones de euros al año en el CHUO.

LAS CIFRAS

1.500 millones de euros al año en el CHUO.

60 millones de euros al año en el CHUO.

300 % Porcentaje en que aumenta el gasto por infección nosocomial, sobre el coste de una hospitalización normal.

80 % Ahorro estimado en recursos humanos mediante el nuevo programa.

Se dan al año hasta 1.500 infecciones hospitalarias o nosocomiales; fallecen entre 50 y personas por este motivo y producen 6.000 ingresos en el CHUO.

Si a ello se suman los primeros episodios de infección, se origina un aumento de hasta un 300 % del coste total de una hospitalización en el mismo centro no derive en infección.

fundación
Tecnología y Salud

CON EL SOPORTE DE
la federación española
de empresas de
fenin **TECNOLOGÍA SANITARIA**

PROPUESTA DE PLAN DE LUCHA contra las Infecciones Relacionadas con la Asistencia Sanitaria



TECNOLOGÍA

Tecnología

Organiza: **DIARIO MEDICO**
ESCUELA DE FORMACIÓN

Patrocina: **Allergan**

Su contribución ha sido decisiva para que estas infecciones disminuyeran en el CHUO un 27%

Galicia implanta un programa de inteligencia artificial que reduce las nosocomiales

El CHUO creó hace años un sistema de inteligencia artificial que, rastreando las bases de datos hospitalarias, detecta infecciones nosocomiales (IRAS). Ahora se está extendiendo en todos los hospitales gallegos.



OURENSE



Pereiro Fernández, junto a parte de su equipo. Le acompañan Eva Lorenzo, Rosalía Laza y Reyes Pardo.

Ingeniería informática contra las infecciones hospitalarias

colabora con el CHUO y espe

SEGÚN EL PROFESOR FERNÁNDEZ RIVERO, LOS GRUPOS DE INVESTIGACIÓN BUSCAN QUE SUS ESTUDIOS SEAN UTILIZADOS Y AYUDAN A LA SOCIEDAD.

FINA ULLOA
OURENSE / LA VOZ

En el año 2010 un equipo de profesionales de los servicios de Medicina Preventiva, Informática y Microbiología del CHUO, junto con la Universidad de Vigo, comenzaron a trabajar en un sistema que permitiera detectar con rapidez cualquier indicio de infección intrahospitalaria para evitar riesgos a los pacientes. Se desarrolló, mediante procesos de minería de datos e inteligencia artificial, una aplicación



«Nos gustaría que se aprovechara, sobre todo porque el trabajo y la inversión

DMJ

millones de euros. Son infecciones que se contraen en los hospitales y que no tienen nada que ver con la causa de ingreso del paciente ni estaban en proceso de incubación cuando llegó. Pueden ser causadas por bacterias o gérmenes a causa de una deficiente asepsia o esterilización, o por contacto con microorganismos que presentan una alta resistencia a los tratamientos antibióticos y se comportan de forma «oportunistas» aprovechando la debilidad de un enfermo para propagarse por su organismo.

El sistema InnoCBR, que es como se bautizó el proyecto del CHUO, es una especie de policía anti-infección que actúa por la noche. Busca en distintas bases de datos, que analiza y relaciona con apuntes en informes médicos o de enfermería para detectar observaciones que pueden estar relacionadas. También recaba datos de la actividad quirúrgica y de registros de los propios pacientes, como puede ser una anotación de fiebre. Combinando todas estas fuentes genera alertas que los profesionales se encuentran a primera hora de la mañana y que les indican si hay pacientes con riesgo de contagio, o incluso sospechosos de patología contagiosa que sea necesario tratar, revisar o aislar para evitar la propagación. «Es capaz incluso de marcar los riesgos por servicios, como quirúrgico, incluso por facultativo, y hacer diagnósticos y propuestas para minimizarlos», señala Berta Uriel.

CLINICAL INFORMATICS

Also referred to as operational informatics, this area helps streamline the delivery of healthcare services.

FOCUSES ON:

- ✓ Clinical decision-making
- ✓ Healthcare delivery
- ✓ All technologies that "touch" the patient
- ✓ Content from clinical disciplines

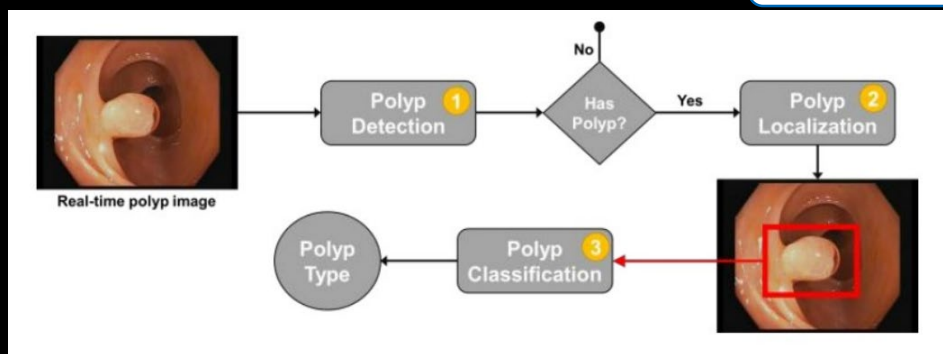
El CHUO exporta un «vigilante» contra infecciones hospitalarias

La Fundación Tecnología y Salud incorporó como ejemplo de buenas prácticas un proyecto del hospital ourensano

Proyecto

Code	DPI2017-87494-R
Funding Institution	Ministerio de Economía, Industria y Competitividad
Programme	National Programme for Research Aimed at the Challenges of Society. Challenge: Health, demographic change and well-being
Beneficiary Institution	Universidad de Vigo
Budget	127.171€
Scheduling	January 2018 – December 2020

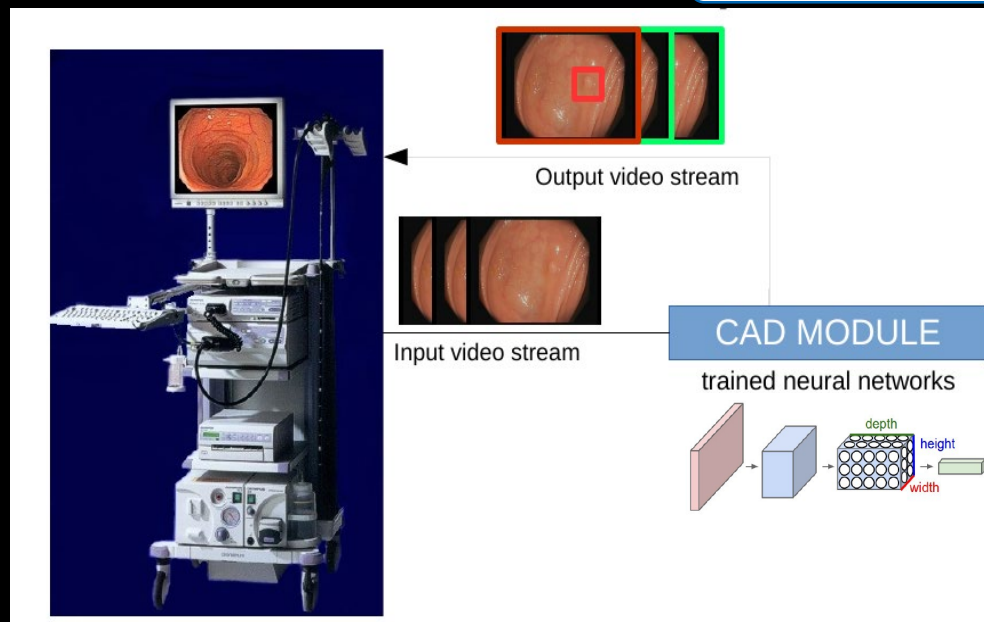
Tareas



Equipo



Implantación



[sistema inteligente de detección y clasificación en tiempo real de lesiones colorrectales mediante deep learning]

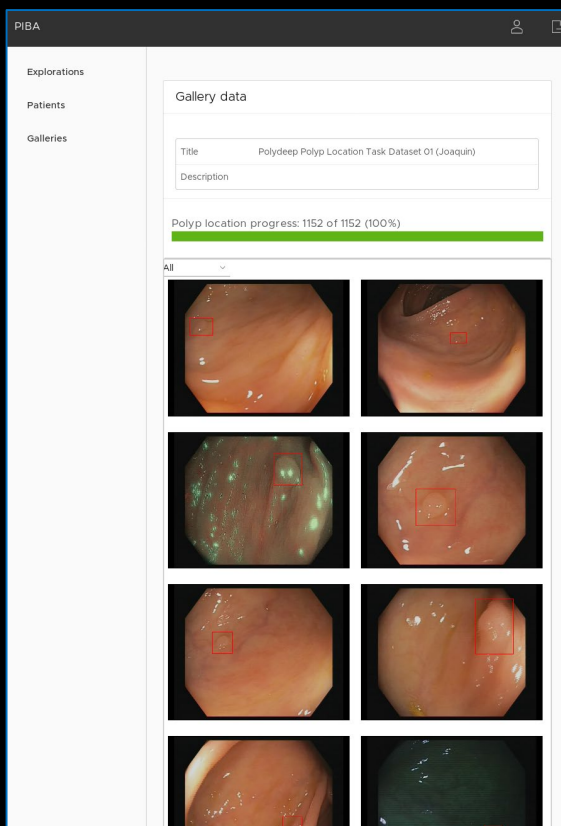
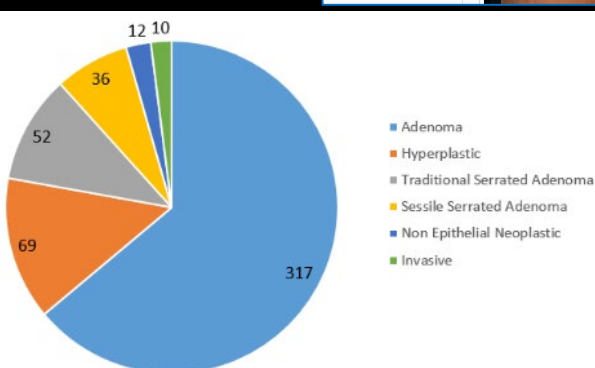
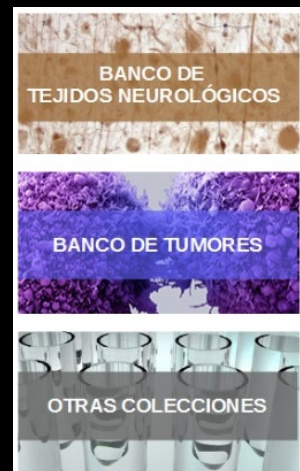


CLINICAL INFORMATICS

Also referred to as operational informatics, this area helps streamline the delivery of healthcare services.

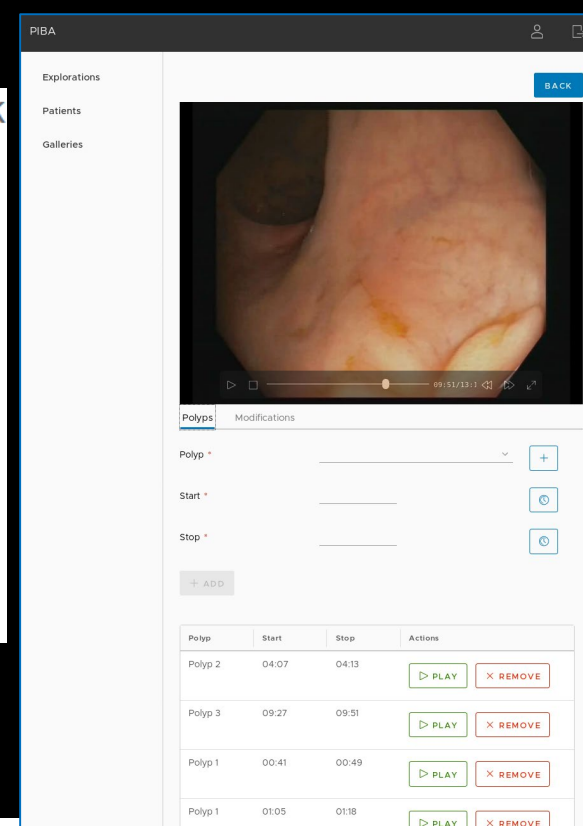
FOCUSES ON:

- ✓ Clinical decision-making
- ✓ Healthcare delivery
- ✓ All technologies that "touch" the patient
- ✓ Content from clinical disciplines



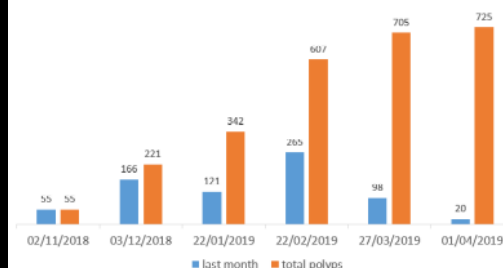
PIBA: Polyp Image Bank

- Polyp database
 - Histology
 - NICE, WASP
- Video Annotation
 - Polyp regions
 - Not polyp regions
 - Instrumental, NBI regions
- Image Localization
 - Done by endoscopists



Explorations (patients): **225**

Polyps: **725** (496 with stored histology)



Database annotation

- Polyp: **312** minutes
- Not polyp: **202** minutes
- Polyp location: **4320** localized images



[sistema inteligente de detección y clasificación en tiempo real de lesiones colorrectales mediante deep learning]

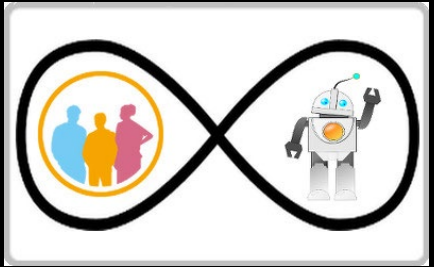
CLINICAL INFORMATICS
Also referred to as operational informatics, this area helps streamline the delivery of healthcare services.

FOCUSES ON:

- ✓ Clinical decision-making
- ✓ Healthcare delivery
- ✓ All technologies that "touch" the patient
- ✓ Content from clinical disciplines



POLYDEEP DETECTION
13/02/2020, SING GROUP
YOLOv3
Dataset: PASCAL VOC (transfer) + PIBA localized polyps
confidence threshold: 50





Informática aplicada a investigación clínica



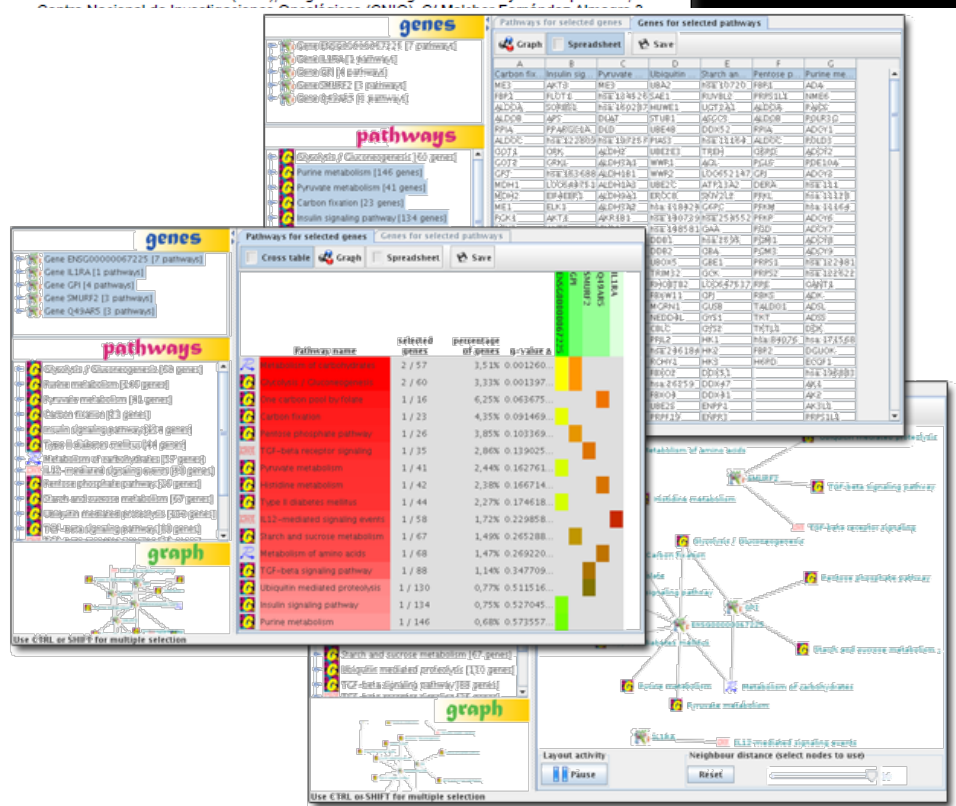
PathJam: a new service for integrating biological pathway information

Daniel Glez-Peña¹, Miguel Reboiro-Jato¹, Rubén Domínguez², Gonzalo Gómez-López², David G. Pisano³, Florentino Fdez-Riverola^{1*}

¹Escuela Superior de Ingeniería Informática University of Vigo, Edificio Politécnico, Campus Universitario As Lagoas s/n, 32004 Ourense, Spain

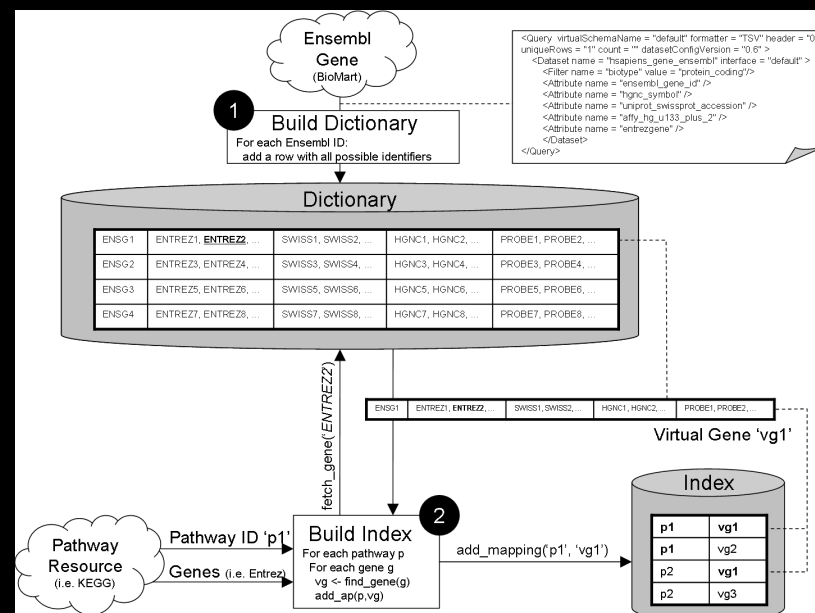
²Informatics Unit, Complexo Hospitalario de Ourense, C/ Ramón Puga 54, 32005, Ourense, Spain

³Unidad de Bioinformática (UBio), Programa de Biología Estructural y Biocomputación,



■ *Análisis de listas de genes de interés en pathways de múltiples fuentes*

- Colaboración con el CNIO
- Reconocimiento de varios espacios de nombres para genes (ensembl, entrez, swiss-prot, probesets)
- Data-warehousing de los pathways de KEGG, Reactome, NCI-PID y Biocarta
- Grafos interactivos y exportables
- Análisis estadístico funcional (estilo DAVID o FatiGO+)



WhichGenes

[www.whichgenes.org]

CLINICAL RESEARCH INFORMATICS
Uses new data and discoveries to improve healthcare.

FOCUS ON:

- ✓ Clinical trials
- ✓ Translational research
- ✓ New medical studies
- ✓ Providing high quality good outcomes, and efficient care delivery

Published online 30 April 2009

Nucleic Acids Research, 2009, Vol. 37, Web Server issue W329–W334
doi:10.1093/nar/gkp263

WhichGenes: a web-based tool for gathering, building, storing and exporting gene sets with application in gene set enrichment analysis

Daniel Glez-Peña¹, Gonzalo Gómez-López², David G. Pisano² and Florentino Fdez-Riverola^{1,3,*}

¹Higher Technical School of Computer Engineering, University of Vigo, Ourense, ²Bioinformatics Unit (UBIO), Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), Madrid and ³Informatics Department, University of Vigo, Vigo, Pontevedra, Spain

Received January 2009

ABSTRACT

WhichGenes building tool extract always databases sources. Wh of interest b the tool is a Every time a added to the notified by a new set stor functionalities sets as well ones by com tion, union a export his s selecting ar tion to the u allows progr a programm State Transf is freely av WhichGenes genes.org/et

INTRODUCTION

During the ps tools have pl contributing t lists (ranging i for various big

From the large amount of tools that are currently available in the community, two widely used approaches can be

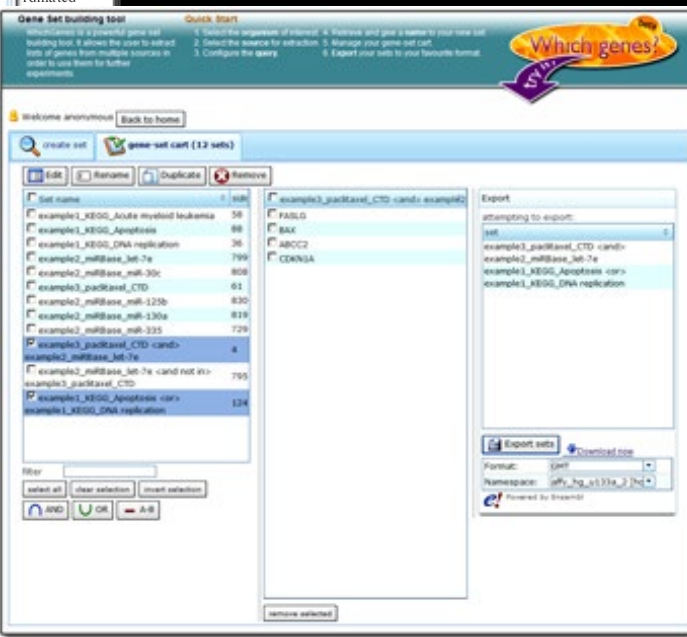
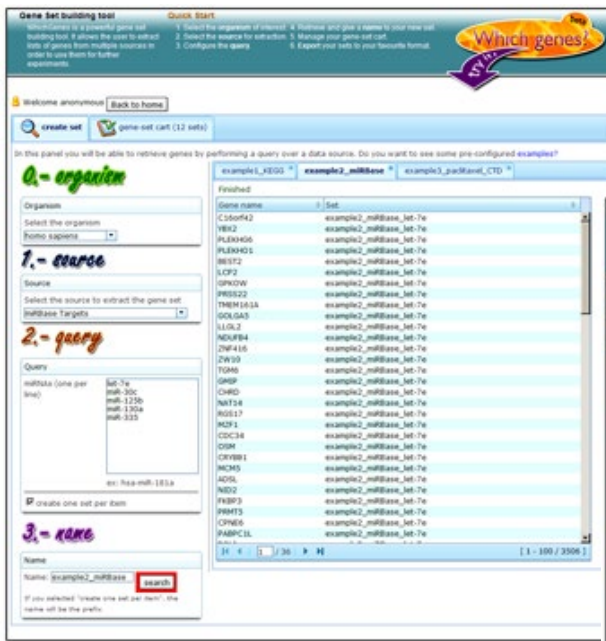
Decipher, Diseases CTD, TargetScan, miRBase CTD, AmiGO and IntAct. Generated gene

ich eval-
een two
analysis
d' meth-

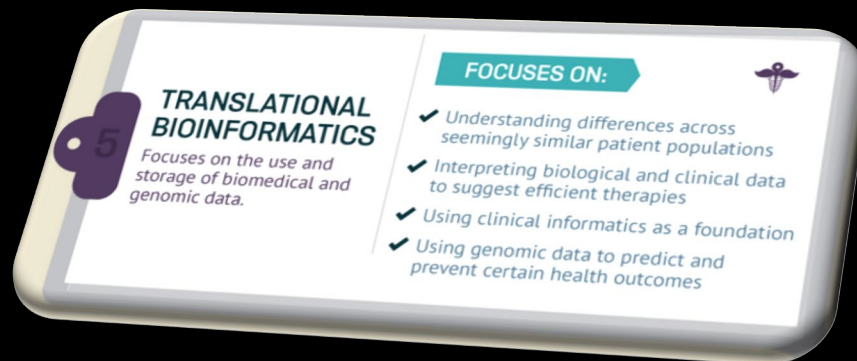
a great
spective,
rdinated

■ Extracción/construcción de listas de genes a partir de múltiples fuentes

- Colaboración con el CNIO
- Unificación de identificadores
- Soporta operadores sobre conjuntos
- Posibilidad de exportar los conjuntos generados en múltiples formatos
- Enlace directo con WhichModel
- Extracción directa de la fuente en el momento de la solicitud



*To whom correspondence should be addressed. Tel: +34 988 387015; Fax: +34 988 387001; Email: riverola@uvigo.es



Bioinformática traslacional



ALTER

[sing-group.org/ALTER]

TRANSLATIONAL BIOINFORMATICS

Focuses on the use and storage of biomedical and genomic data.

FOCUSES ON:

- ✓ Understanding differences across seemingly similar patient populations
- ✓ Interpreting biological and clinical data to suggest efficient therapies
- ✓ Using clinical informatics as a foundation
- ✓ Using genomic data to predict and prevent certain health outcomes

W14-W18 Nucleic Acids Research, 2010, Vol. 38, Web Server issue
doi:10.1093/nar/gkq321

Published online 3 May 2010

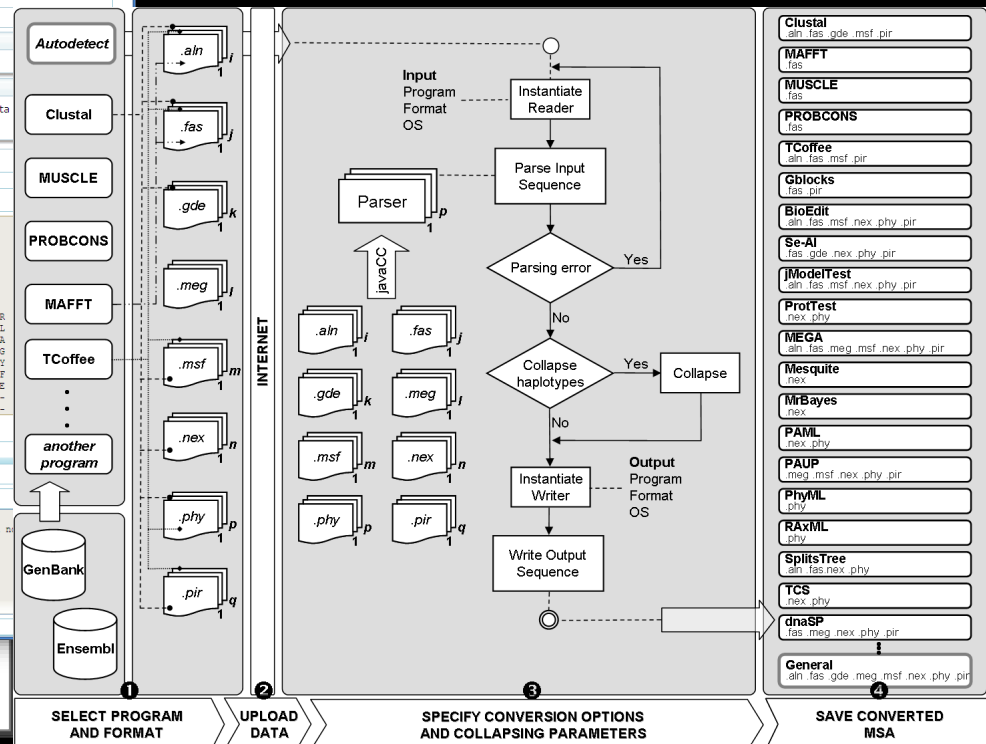
ALTER: program-oriented conversion of DNA and protein alignments

Daniel Glez-Peña¹, Daniel Gómez-Blanco¹, Miguel Reboiro-Jato¹,
Florentino Fdez-Riverola^{1,2*} and David Posada^{2,3*}

The screenshot shows the ALTER web interface. The 'select format' section has 'Autodetect' selected. The 'upload or paste MSA' section has a file upload area and a 'use sample MSA' button. The 'sample MSA' section shows a list of sequences with their lengths and weights. The 'Log' section shows the conversion process from MSA to NEXUS format.

Estudio de árboles filogenéticos

- Conversión de ficheros de alineamiento multiple de secuencias (MSA) e identificación de haplotipos
- Flujo de trabajo en 4 pasos guiado por asistente
- Multiplataforma: Internet Explorer 7, Firefox 3, Opera 9.62, Google Chrome 3.0.195.38 y Safari 3 bajo Windows XP/Vista, Ubuntu Linux 8.04 version and Mac OSX 10.6



Herramientas de soporte en proteómica

TRANSLATIONAL BIOINFORMATICS
Focuses on the use and storage of biomedical and genomic data.

FOCUSES ON:

- ✓ Understanding differences across seemingly similar patient populations
- ✓ Interpreting biological and clinical data to suggest efficient therapies
- ✓ Using clinical informatics as a foundation
- ✓ Using genomic data to predict and prevent certain health outcomes

28

La Voz de Galicia

Sociedad

Viernes, 8 de enero del 2010

Con su método aceleran las reacciones enzimáticas de la orina, pasando de cuatro horas a solo 20 minutos

Científicos de Ourense logran reducir el tiempo del análisis antidopaje

Jesús Manuel García

OURENSE | El análisis de orina para detectar esteroides anabólicos y hormonas en los deportistas y ver si se dopan, ya puede hacerse en pocos minutos. Tanto es así que el grupo Bioscope, en colaboración con el equipo de investigación de nutrición y bromatología AAI de la Facultad de Ciencias de Ourense y el Laboratorio Antidopaje de Roma, de la Federación Médica Deportiva Italiana, ha desarrollado un método que rebaja de aproximadamente cuatro horas a veinte minutos el tiempo del análisis.

«Comenzamos una colaboración con el Laboratorio Antidopaje de Portugal aunque hoy trabajamos con el de Roma», señala José Luis Capelo, de Bioscope, porque los tratamientos de muestras para detectar esteroides en deportistas utilizan metodologías muy lentas y caras. Por ello, estos científicos ourenses desarrollaron un método muy novedoso en colaboración con sus colegas italianos, «que reduce el análisis de una muestra de orina para detectar

cualquiera de los 60 esteroides prohibidos por la Organización Deportiva Mundial de unas cinco horas que se tardaba hasta ahora hasta veinte minutos», señala el citado doctor.

La investigación ya dio resultado y va a permitir saber con mayor seguridad si un deportista se está dopando o no. En este trabajo tiene una parte importante el investigador Marco Galesio.

Biblioteca

El siguiente paso será crear una biblioteca de estos esteroides para hacer la identificación más rápida y barata de modo que una búsqueda ágil de esteroides prohibidos por la Organización Deportiva Mundial llevará no más de 30 minutos. «Habremos modificado completamente la metodología de análisis de esteroides a nivel mundial con este sistema», señala el científico ourense.

Esta investigación llegó a Bioscope «porque es un proyecto que presentamos y ganamos como tesis doctoral, ya que nosotros descubrimos y tenemos la



Los integrantes de Bioscope son pioneros en acelerar reacciones de enzimas con ultrasonidos | LUDERO

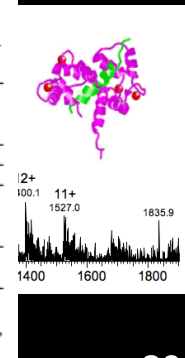
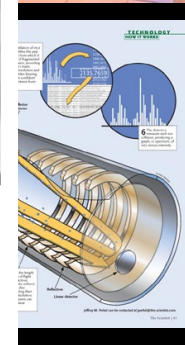
patente, de que con ultrasonidos conseguimos acelerar reacciones enzimáticas. El tratamiento que tenemos aquí es químico y utiliza enzimas», señala José Luis Capelo. Para que esas enzimas hicieran su trabajo se les aplicaba calor a 55 grados una hora «pero lo mismo que consigues con eso, lo logras en 5 minutos y hacemos todo en veinte», asegura el responsable del grupo investigador. A la muestra le aceleran la reacción enzimática que descompone los esteroides en compuestos que se pueden detectar. «Aceleramos el proceso en el tiempo y el análisis final necesita una búsqueda rápida y lo que tarda dos horas lo haremos en segundos con ayuda de los informáticos del campus», señala Capelo.

El peligro de consumir productos con anabolizantes prohibidos como suplementos nutricionales

Entre 5 y 10 años, dicen los responsables de Bioscope, las autoridades deportivas pueden llevar unos aparatos portátiles, los *maldis*, al laboratorio y hacer el análisis de orina allí al momento.

Los esteroides anabolizantes androgénicos son el mayor grupo incluido en la lista de sustancias prohibidas por la agencia mundial antidopaje y por las autoridades deportivas mundiales. Pero curiosamente el abuso de estas sustancias abunda mucho más

en atletas no competitivos. A mayores, tal abuso se produce fuera del mundo deportivo, como reflejo de un mejor estilo de vida. Sin saberlo, podemos tomar productos que contienen esas sustancias como suplementos nutricionales, sustancias que sin una indicación adecuada pueden empeorar el problema. «Como la gente toma esos productos sin parar se encuentra con efectos muy diversos como una disminución del tamaño de los testículos», dice Capelo.



- ✓ Understanding differences across seemingly similar patient populations
- ✓ Interpreting biological and clinical data to suggest efficient therapies
- ✓ Using clinical informatics as a foundation
- ✓ Using genomic data to predict and prevent certain health outcomes

Piñero-Yáñez et al. *Genome Medicine* (2018) 10:41
<https://doi.org/10.1186/s13073-018-0546-1>

Genome Medicine

SOFTWARE

Open Access



PanDrugs: a novel method to prioritize anticancer drug treatments according to individual genomic data

Elena Piñero-Yáñez¹, Miguel Reboiro-Jato^{2,3,†}, Gonzalo Gómez-López¹, Javier Perales-Patón¹, Kevin Troulé¹, José Manuel Rodríguez⁴, Héctor Tejero⁵, Takeshi Shimamura⁶, Pedro Pablo López-Casas¹, Julián Carretero⁵, Alfonso Valencia⁴, Manuel Hidalgo^{1,7}, Daniel Glez-Peña^{2,3} and Fátima Al-Shahrour^{1*}

Abstract

Background: Large sequencing cancer genome projects have shown that tumors have thousands of molecular alterations and their frequency is highly heterogeneous. In such scenarios, physicians and oncologists routinely face lists of cancer genomic alterations where only a minority of them are relevant biomarkers to drive clinical decision-making. For this reason, the medical community agrees on the urgent need of methodologies to establish the relevance of tumor alterations, assisting in genomic profile interpretation; and, more importantly, to prioritize those that could be clinically actionable for cancer therapy.

Results: We present PanDrugs, a new computational methodology to guide the selection of personalized treatments in cancer patients using the variant lists provided by genome-wide sequencing analyses. PanDrugs offers the largest database of drug-target associations available from well-known targeted therapies to preclinical drugs. Scoring data-driven gene cancer relevance and drug feasibility PanDrugs interprets genomic alterations and provides a prioritized evidence-based list of anticancer therapies. Our tool represents the first drug prescription strategy applying a rational based on pathway context, multi-gene markers impact and information provided by functional experiments. Our approach has been systematically applied to TCGA patients and successfully validated in a cancer case study with a xenograft mouse model demonstrating its utility.

Conclusions: PanDrugs is a feasible method to identify potentially druggable molecular alterations and prioritize drugs to facilitate the interpretation of genomic biomarkers and clinical decision making in cancer patients. Our research

Mediante Microarrays o NGS se obtienen listas de alteraciones moleculares:

- Cuáles dirigen el tumor? y, de ellas...
- Cuáles son accionables? y, de ellas...
- Cómo priorizar la lista de posibles fármacos?

What is PanDrugs?

PanDrugs provides a bioinformatics platform to **prioritize anticancer drug treatments according to individual genomic data**. PanDrugs current version integrates data from **24** primary sources and supports **56297** drug-target associations obtained from **4804** genes and **9092** unique compounds.

Data input: standard VCF file, RNK file, gene lists and drug query.

Please note the PanDrugs terminology for druggable genes:

- I. Direct targets:** Genes that contribute to disease phenotype and can be directly targeted by a drug (e.g. BRAF is a direct target for vemurafenib).
- II. Biomarkers:** Genes showing a genetic status associated with drug response which protein product is not the drug target itself (e.g. BRCA-mutated cancers responding to PARP inhibitors).
- III. Pathway members:** Genes located downstream in the biological pathway of a given undruggable gene (e.g. patients with mutations in TSC1/2 respond to a downstream inhibition of the mTOR pathway).



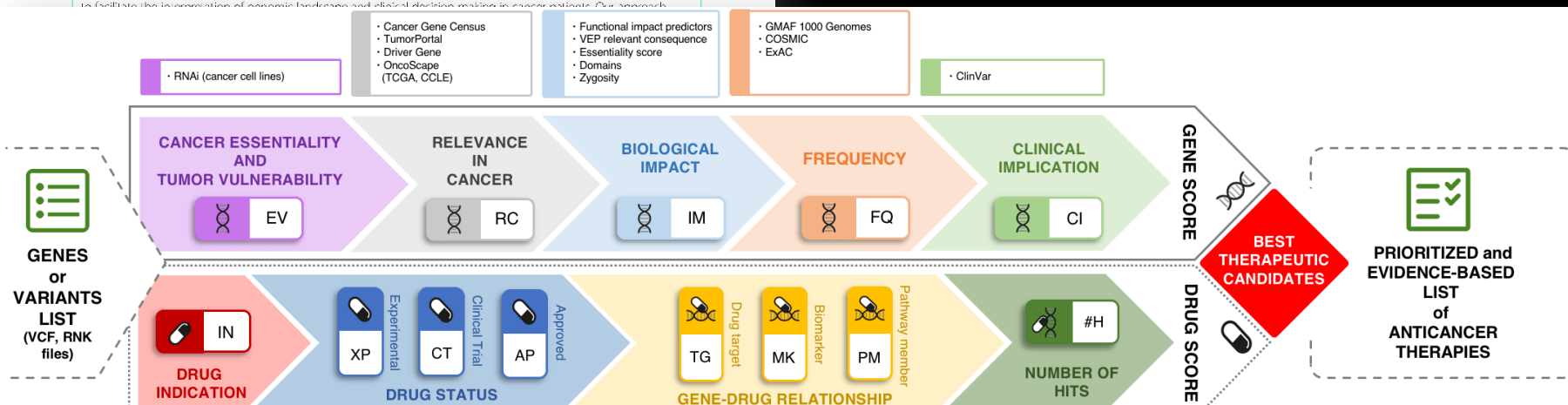
DIRECT TARGET



BIOMARKER



PATHWAY MEMBER





Informática en salud pública

CURMIS4th

CURMIS4th

[Machine learning-based models to detect and curb Health misinformation on social media]



Proyecto

Code	PID2020-113673RB-I00
Funding Institution	Ministerio de Ciencia e Innovación
Programme	National Programme for Research Aimed at the Challenges of Society. Challenge: Health, demographic change and well-being
Beneficiary Institution	Universidad de Vigo
Budget	60.137€
Scheduling	September 2021 – August 2024

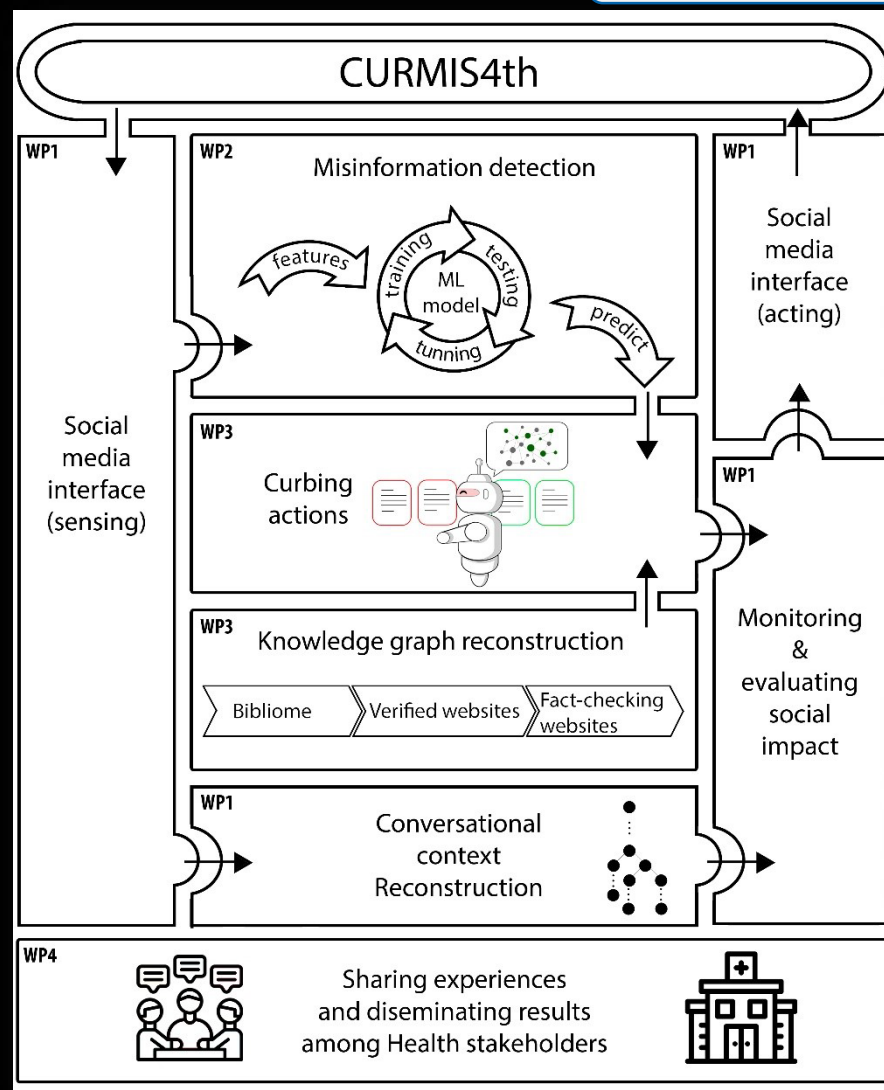
Equipo

- Dra. Beatriz Villamarín Bello
médica atención primaria - Área sanitaria de A Coruña y Cee
- Dra. Noelia Estévez Rionegro
doctora en lingüística

EPOs



Propuesta





IA interpretable

“The future of AI lies in enabling people to collaborate with machines to solve complex problems. Like any efficient collaboration, this requires good communication, trust and understanding.”

[Explainable AI Research Lead](#)

[Accenture Labs](#)

IA interpretable

introducción

■ En decisiones de sistemas basados en IA que afectan a vidas humanas se necesita transparencia

- medicina, defensa, justicia, etc.

■ Compromiso entre rendimiento y transparencia

- *black-box* (opaco) VS inteligible

■ El considerar la interpretabilidad como un objetivo de diseño aporta ventajas:

- garantiza imparcialidad en la toma de decisiones (*correct from bias*)
- incrementa robustez detectando inconsistencias
- posibilita que sólo se tengan en cuenta variables relevantes (*truthful causality*)

JCR Impact Factor

JCR Year	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE			COMPUTER SCIENCE, THEORY & METHODS		
	Rank	Quartile	JIF Percentile	Rank	Quartile	JIF Percentile
2018	3/134	Q1	98.134	2/105	Q1	98.571

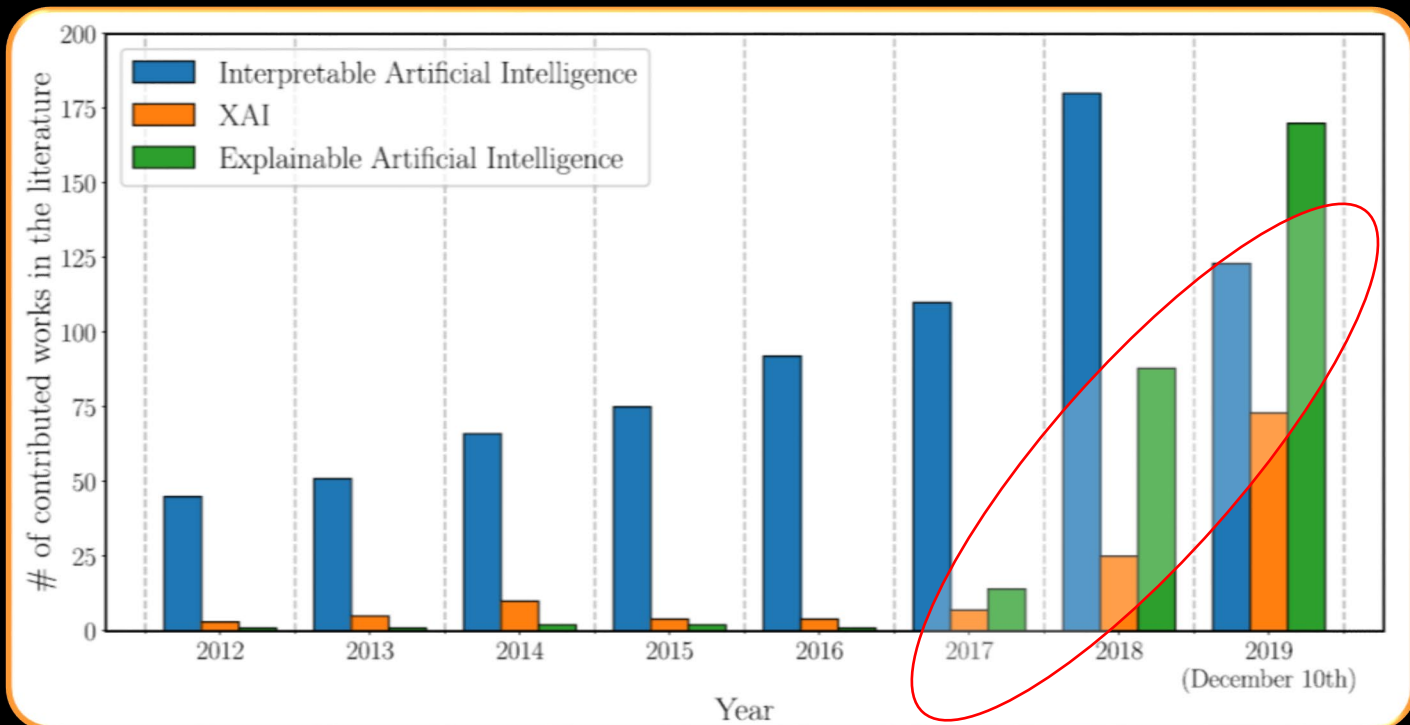




IA interpretable

evolución y objetivo

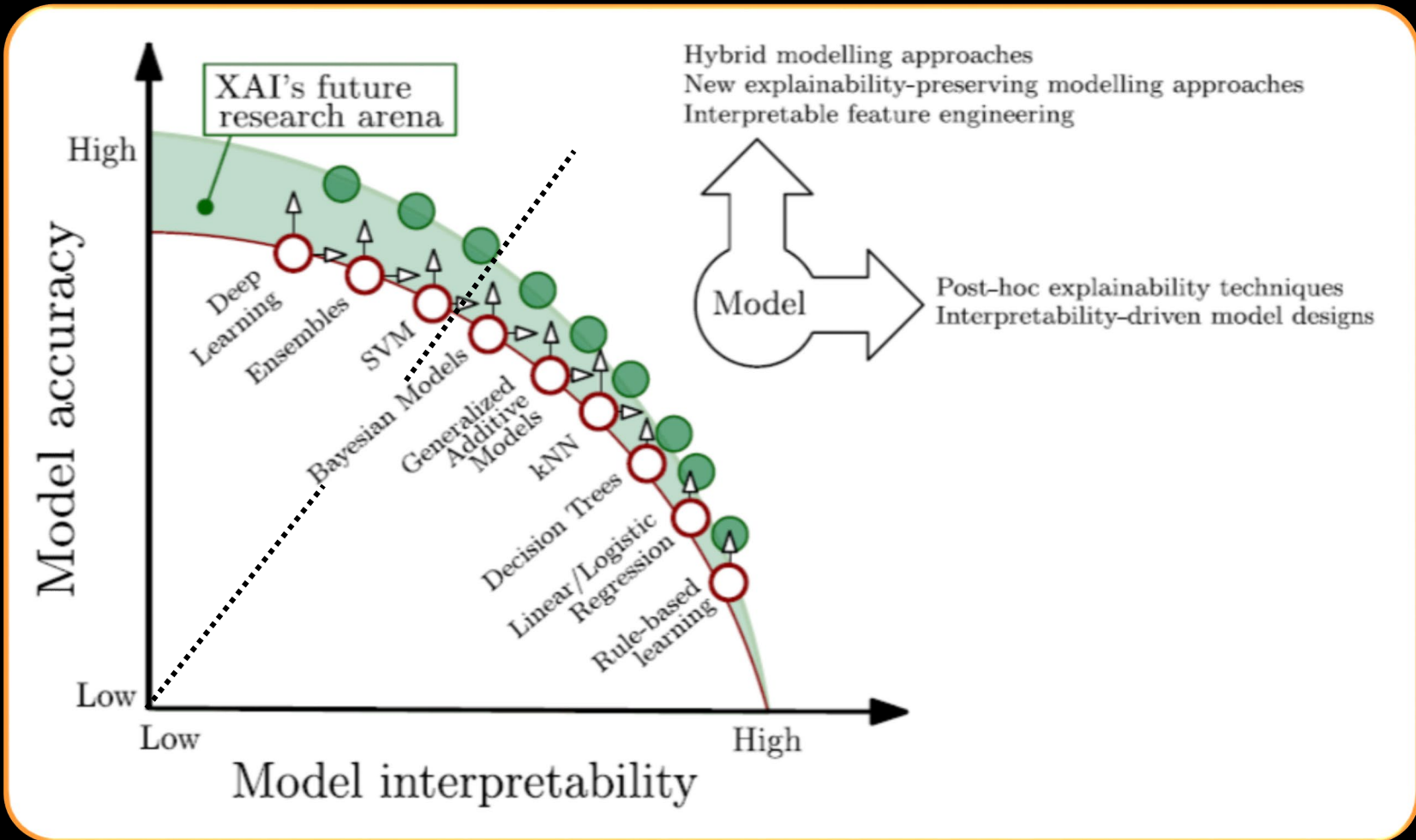
- Generar avances en las técnicas de ML para conseguir...
 - producir modelos fácilmente interpretables manteniendo un alto nivel de rendimiento en su aprendizaje (*high prediction accuracy*).
 - garantizar que los humanos puedan **entender fácilmente, confiar de forma segura y manejar de forma efectiva** la nueva generación de *partners* inteligentes





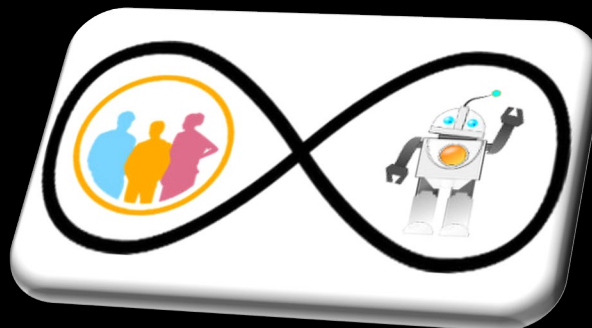
IA interpretable

¿hacia dónde se evoluciona?: áreas de mejora de los modelos de ML existentes





*“...The future of AI lies in **enabling people to collaborate with machines** to solve complex problems...*



Human in the loop

“Computers are incredibly fast, accurate, and stupid; humans are incredibly slow, inaccurate, and brilliant; together they are powerful beyond imagination.”

[Albert Einstein](#)

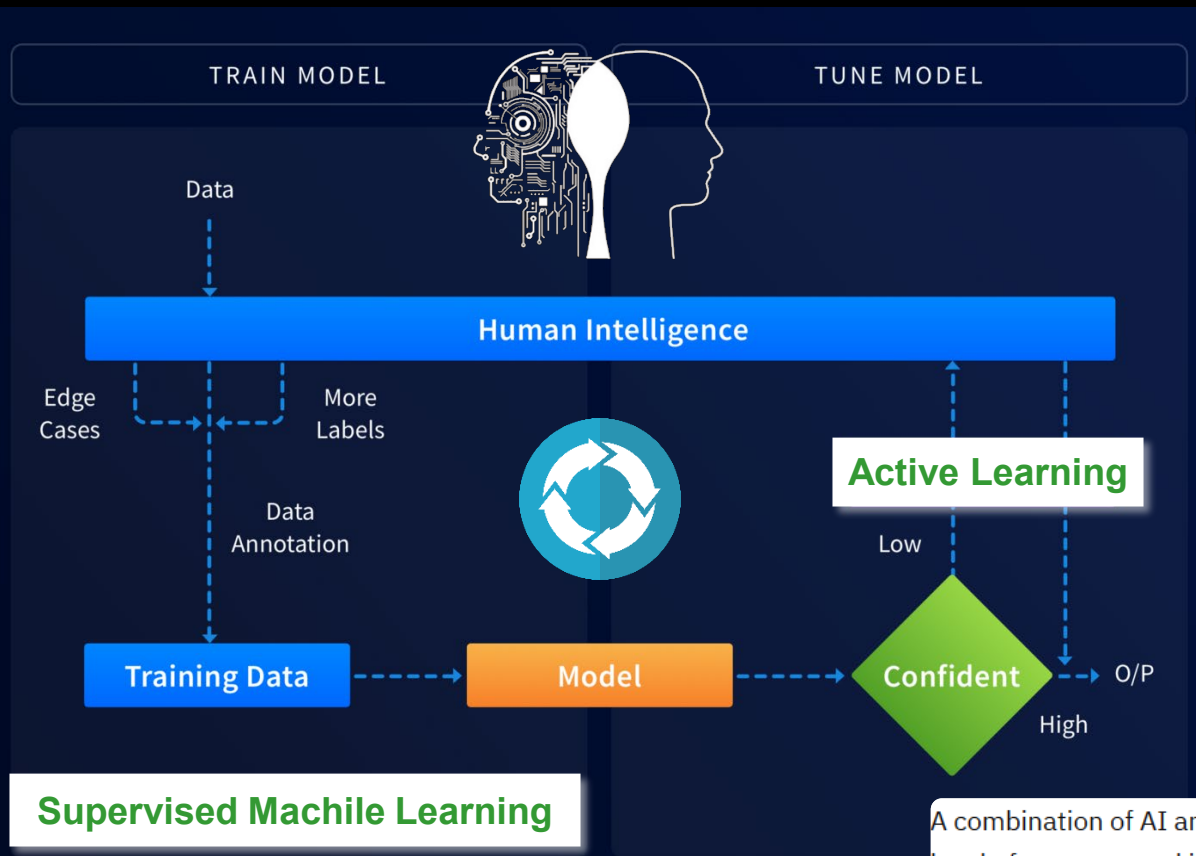


Humanistic Intelligence

definición, marco conceptual



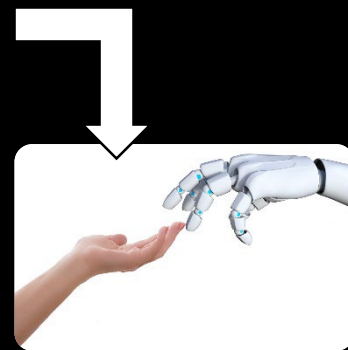
■ Se basa en la creación de inteligencia como resultado de un ciclo iterativo de reatrolimentación (*feedback*) entre un proceso computacional y un humano



Human In The Loop

=

SML + AL



A combination of AI and Human Intelligence gives rise to an extremely high level of accuracy and intelligence (Super Intelligence). This combination is powerful beyond imagination.



Los centros gallegos contarán con una nueva materia de Inteligencia Artificial el próximo curso en 1º de bachillerato



Presentación de la nueva materia de Iniciación á Intelixencia Artificial para centros escolares - XOÁN CRESPO

La asignatura, planteada con un enfoque práctico, será de libre configuración y se impartirá en dos horas semanales



¿Qué piensa sobre las
máquinas que piensan?

Conclusiones

*“Computers are incredibly fast, accurate, and stupid;
humans are incredibly slow, inaccurate, and brilliant;
together they are powerful beyond imagination.”*

[Albert Einstein](#)



Lecciones aprendidas...

después de 16 años de desarrollo [2006-2021]

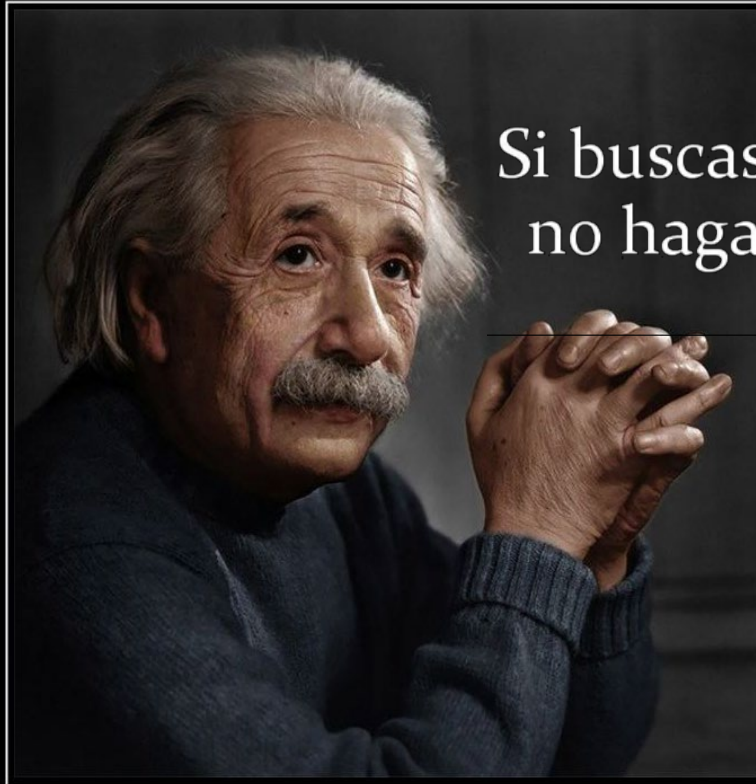
strong IA



weak IA



- Informática (tecnología) + IA (ciencia) tienen mucho potencial en el ámbito de la salud
 - proactividad: búsqueda de problemas y nuevos retos
- Fundamental la creación de equipos de trabajo de carácter multidisciplinar
 - todo el mundo tiene algo valioso que aportar y se necesita personal formado y motivado, como vosotros ...
 - SNS y la Universidad son un muy buen lugar para iniciarse en investigación
- No todo son casos de éxito, ni siempre existe recompensa
 - pero es parte de nuestro compromiso con la sociedad y merece la pena intentarlo



Si buscas resultados distintos
no hagas siempre lo mismo

Albert Einstein

¡gracias por
vuestra
curiosidad!





Informática clínica



Computational Intelligence, Volume 22, Number 3/4, 2006

GENE-CBR: A CASE-BASED REASONING TOOL FOR CANCER DIAGNOSIS USING MICROARRAY DATA SETS

FERNANDO DÍAZ

Escuela Universitaria de Informática, University of Valladolid, Spain

FLORENTINO FDEZ-RIVEROLA

Escuela Superior de Ingeniería Informática, University of Vigo, Spain

JUAN M. CORCHADO

Facultad de Ciencias, University of Salamanca, Spain

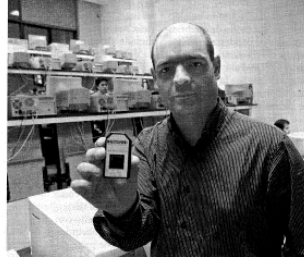
Martes, 18 de noviembre del 2008

La herramienta apoya la toma de decisiones y la utiliza el Clínico de Salamanca

Informática ideó un sistema que clasifica pacientes con cáncer desde muestras de ADN

Jesús Manuel García

OURENSE | La Escuela Superior de Ingeniería Informática de Ourense colabora con el Hospital Universitario de Salamanca y con la Fundación Biomédica del Complejo Hospitalario de Vigo para mejorar la clasificación de pacientes utilizando datos que provienen de unas fichas denominadas microarray. Estas permiten ver el estado de una célula desde el punto de vista genético. La técnica del microarray es novedosa, se hizo común hace unos cinco o seis años y permite obtener una fotografía a nivel genético de un paciente.



Fernández muestra un ejemplar de microarray | J. MANUEL VILAS

El director de este proyecto que ya se está desarrollando en el hospital salmantino, Florentino Fernández Rivero, explica que antes los análisis que se podían realizar a una persona se venían de sangre y ahí se podía ver cómo estaban algunas células y una determinada célula cancerígena se podía observar a través del microscopio. Sin embargo con esta técnica de los microarray es posible saber cómo está esa célula a nivel genético, dice.

Medición

Estos dispositivos, explica el ingeniero ourensano, permiten medir el valor de 22.000 genes y conocer si un gen está muy expresado o poco. «Hoy hay microarrays que en un solo experimento son capaces de medir la expresión de más de un millón de genes. En este campo se necesitan nuevos programas y técnicas capaces de analizar estos datos. Las técnicas clásicas de datos minús y las de inteligencia artificial como redes neuronales no están preparadas para trabajar con un volumen de información masivo como este y por ello es muy difícil procesar este tipo de información».

Al microarray se le introduce ADN de la célula que se quiere analizar, de un determinado paciente. «En informática el array es como un tablero de ajedrez, tiene un montón de celdas pequeñas y en cada una se va a medir la expresión de un gen, y se llama micro porque estamos trabajando con unas dimensiones muy pequeñas», explica Fernández.

Hay dos líneas de investigación. Una, consiste en adaptar la técnica del equipo ourensano para los nuevos microarrays, que en vez de tener 22.000 valores por experimento tienen más de un millón de valores. La otra línea de investigación pretende unir el conocimiento existente en la web en forma de publicaciones científicas en el campo de la medicina general.

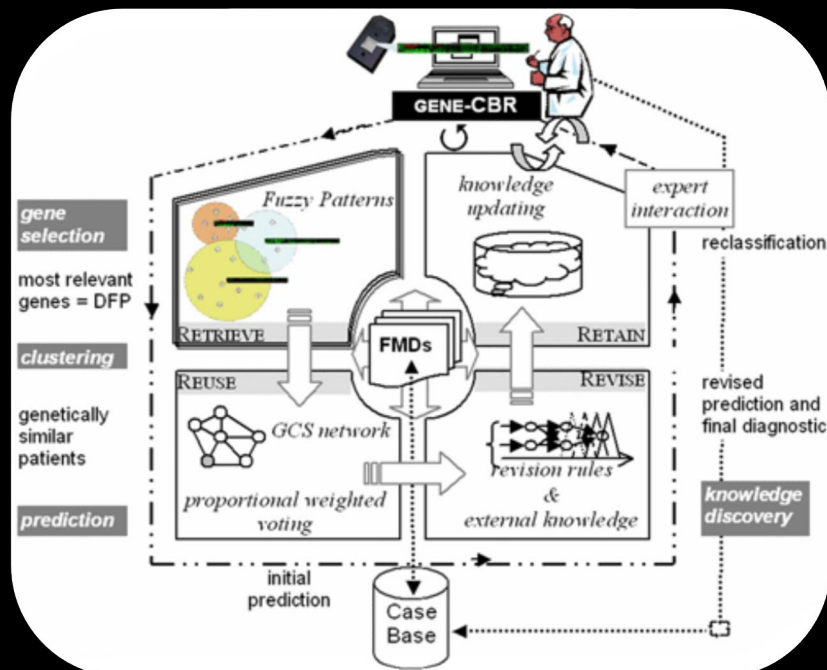
El microarray es como un porta de microscopios. Esto lo lee un escáner de alta resolución que mide colores. «Los genes que metemos ahí hibridan con otros genes base que vienen en esa porta. Los base están teñidos de un color y los que hibridan se teñen de otro y si hay tal hibridación se ve el nuevo color resultante, por ejemplo, de un gen verde y uno rojo. Si no hay hibridación está el color base que tenía el gen que se hallaba en ese huequillo. El escáner mira la intensidad cromática y es capaz de saber si el gen está mucho o poco expresado», explica Florentino. La aplicación desarrollada en el campus se puede descargar en la web www.genecbr.org. Permite clasificar pacientes con distintas enfermedades a partir de datos de microarrays de ADN. «El Hospital Universitario de Salamanca tenía un montón de microarrays sin forma de analizar toda esa información y no podían utilizar ningún programa o no se fiaban de los resultados de ciertos programas», señala el ingeniero. Este proyecto lo prueban en Salamanca y, según Fernández, es un sistema de apoyo a la toma de decisiones médicas. En este trabajo participaron Daniel González, Reyes Pavón, Rosalía Laza y Eva Lorenzo Iglesias en el equipo de Florentino Fernández, además del profesor Fernando Díaz, de la Universidad de Valladolid; el Hospital Clínico de Salamanca y la Fundación Biomédica.

Un diagnóstico más sencillo

Que el médico pueda diagnosticar de forma sencilla y colaborante en el caso de cáncer es fundamental, señala Florentino Fernández, «porque importa mucho el tratamiento que se dará a las personas en función del cáncer que tengan». El proyecto del campus se ha diseñado o se está empleando en casos de leucemias.

Sistema de apoyo al diagnóstico de cáncer basado en microarrays de ADN

- Colaboración con el CHUS y CIC
- 3 Modos de funcionamiento:
 - Modo diagnóstico. Para el oncólogo
 - Modo experto. Ajuste de técnicas de IA
 - Modo programación. Prueba de prototipos nuevos
- Técnicas de IA empleadas
 - DFP para selección de genes discriminantes
 - GCS para clasificación



BMC Bioinformatics

Software

DFP: a Bioconductor package for fuzzy profile id gene reduction of microarray data

Daniel Glez-Peña¹, Rodrigo Álvarez², Fernando Díaz³ and Riverola⁴*

Address: ¹Escuela Superior de Ingeniería Informática, Universidad de Salamanca, ²Departamento de Informática, Universidad de Salamanca, ³Escuela Superior de Ingeniería Informática, Universidad de Vigo, ⁴Escuela Superior de Ingeniería Informática, Universidad de Salamanca

Published: 29 January 2009

BMC Bioinformatics 2009, 10:137

This article is available from: <http://www.biomedcentral.com/1471-2105-10-137>

© 2009 Glez-Peña et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

BMC Bioinformatics

Software

geneCBR: a translational tool for integrative information retrieval research

Daniel Glez-Peña¹, Fernando Díaz² and Florentino Fdez-Riverola⁴*

Address: ¹Escuela Superior de Ingeniería Informática, Universidad de Salamanca, ²Departamento de Informática, Universidad de Salamanca, ³Escuela Superior de Ingeniería Informática, Universidad de Vigo, ⁴Escuela Superior de Ingeniería Informática, Universidad de Salamanca

Address: ¹Escuela Superior de Ingeniería Informática, Universidad de Salamanca, ²Departamento de Informática, Universidad de Salamanca, ³Escuela Superior de Ingeniería Informática, Universidad de Vigo, ⁴Escuela Superior de Ingeniería Informática, Universidad de Salamanca

Published: 18 June 2009

BMC Bioinformatics 2009, 10:187 doi:10.1186/1471-2105-10-187

This article is available from: <http://www.biomedcentral.com/1471-2105-10-187>

© 2009 Glez-Peña et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Bioinformatics and medical informatics communities. Both domains share the common goal of biomedical research, and contributing to the treatment have been successfully used to investigate useful information integration of existing methods into day-to-day clinical reasoning emerges as a suitable paradigm specially in decision support systems, given the support and the goals of removing barriers against multi-disciplinary knowledge to real practice, case-based reasoning system because their computational reasoning paradigm is similar in their own practice of clinical medicine.

Results: In addressing the issue of bridging the existing domain of cancer diagnosis, prognosis and treatment framework. Our geneCBR system implements a freely that can be applied to gene selection, clustering, knowledge discovery and decision support systems, given the support and the goals of removing barriers against multi-disciplinary knowledge to real practice, case-based reasoning system because their computational reasoning paradigm is similar in their own practice of clinical medicine.

Conclusion: geneCBR is a new translational tool for biomedical researchers and clinicians working together. GPL license and can be obtained at <http://www.genecbr.org>



- Muestras de médula osea pertenecientes a 43 pacientes adultos con Leucemia Mieloide Aguda + 6 individuos sanos
 - 10 patients with Acute Promyelocytic Leukemia [APL]
 - 4 patients with Acute Myeloid Leukemia with inv(16) [AML-inv(16)]
 - 7 patients with Acute Monocytic Leukemia [AML-mono]
 - 22 patients with Acute non-Monocytic Leukemia [AML-other]
 - 6 samples belonging to sane individuals [control samples]
- Volumen de la información manejada
 - Cada microarray contiene 22.283 ESTs ($\equiv$ genes)
 - 49 microarrays = 1.091.867 valores de expresión genética
- Datos disponibles en la actualidad
 - 150 microarrays (Human Genome 133A) + 210 microarrays (Human Genome - plus)

BioAnnote

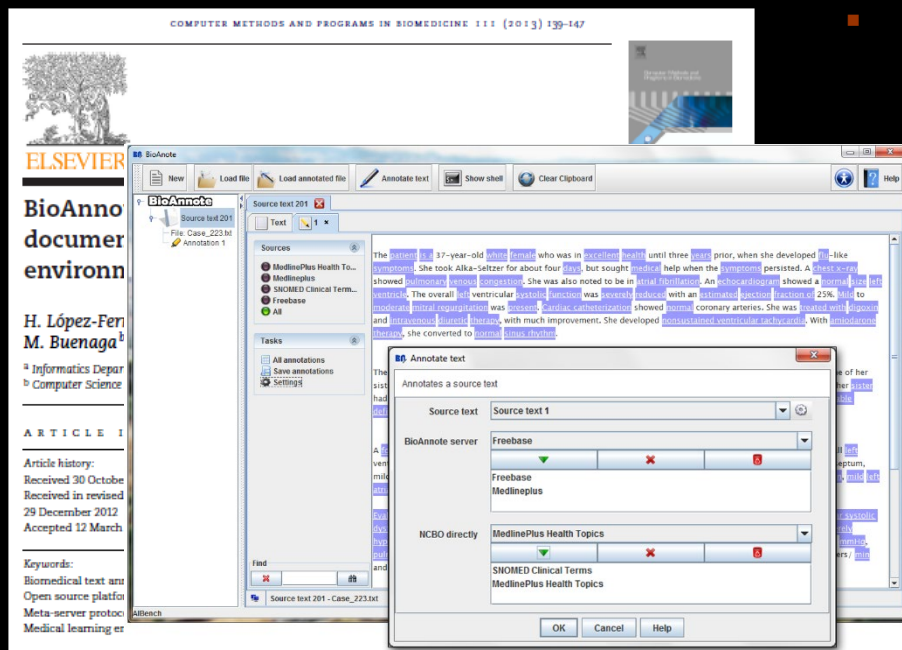
[sing-group.org/bioannotate]

CLINICAL INFORMATICS

Also referred to as operational informatics, this area helps streamline the delivery of healthcare services.

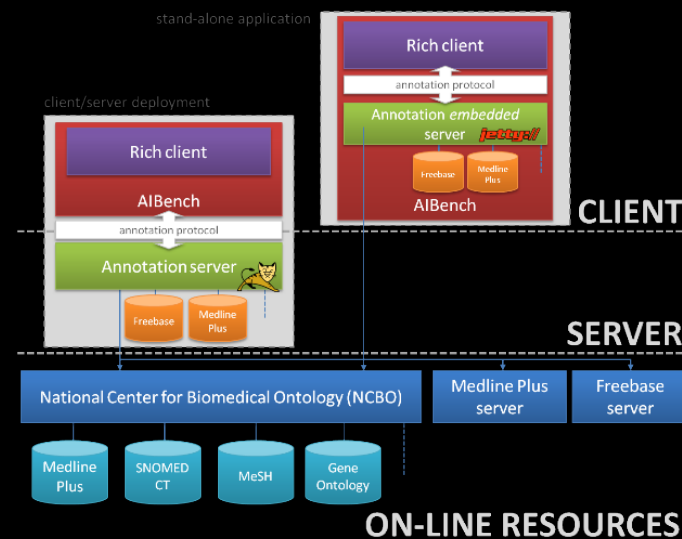
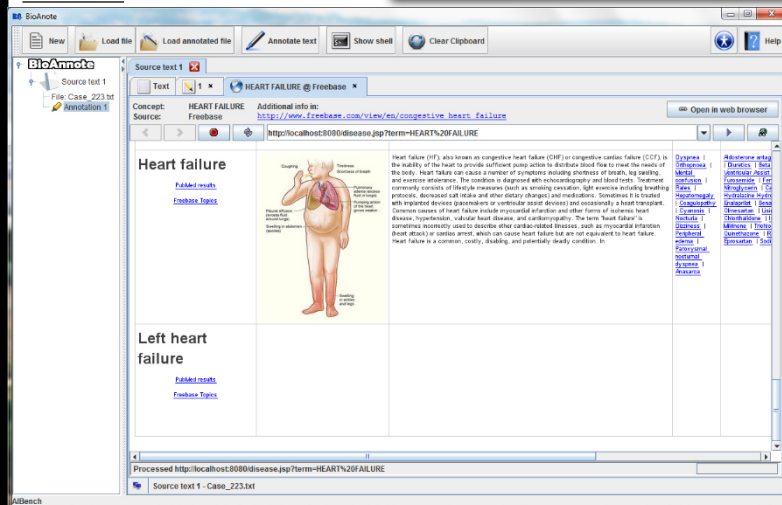
FOCUSES ON:

- ✓ Clinical decision-making
- ✓ Healthcare delivery
- ✓ All technologies that "touch" the patient
- ✓ Content from clinical disciplines



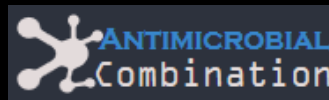
■ Aplicación capaz de anotar textos biomédicos utilizando diferentes recursos on-line como Medlineplus y Freebase

- Colaboración con UEM (Grado de Medicina) para el análisis de casos clínicos por parte de los estudiantes
- El usuario introduce documentos y obtiene anotación de términos relevantes (enfermedades, síntomas, tratamientos)
- Recuperación automática de información detallada sobre cada entidad reconocida (topics relacionados, enlaces externos, bibliografía de relevancia, resúmenes de artículos, etc.)





Informática aplicada a investigación clínica



WhichModel

[www.whichmodel.org]

CLINICAL
RESEARCH
INFORMATICS

Uses new data and discoveries to improve healthcare.

FOCUS ON:

- ✓ Clinical trials
- ✓ Translational research
- ✓ New medical studies
- ✓ Providing high quality good outcomes, and efficient care delivery

Estudio del poder discriminante en problemas de clasificación de hipótesis establecidas utilizando conjuntos de genes relevantes biológicamente

- Integrado con **WhichGenes?**
- Soporta Naïve Bayes, Decision trees, k-NN y SVM
- Asistente para la creación de experimentos en 2 sencillos pasos
- Ejecución paralela en el servidor de cálculo
- Notificación por e-mail con enlaces directos a los resultados en el servidor

Gene set	Naive Bayes Simple	J48 (C4.5)	IBk	LibSVM
OMIM_LEUKEMIA	0.6610	0.4643	0.3443	0.2453
OMIM_MYELOID	0.6970	0.5161	0.4737	0.2453
OMIM_MYELOMONOCYTIC	0.3939	0.5161	0.5000	0.2857
OMIM_PROMYELOCYTIC	0.3220	0.3056	0.3333	0.2453

GeneCommittee

[sing-group.org/GC]



CLINICAL RESEARCH INFORMATICS

Uses new data and discoveries to improve healthcare.

FOCUSES ON:

- ✓ Clinical trials
- ✓ Translational research
- ✓ New medical studies
- ✓ Providing high quality good outcomes, and efficient care delivery



Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc



A novel ensemble of classifiers that use biological relevant gene sets for microarray classification

Miguel Reboiro-Jato^a, Fernando Díaz^b, Daniel Glez-Peña^a, Florentino Fdez-Riverola^{a,*}

^a Departamento de Informática, Escuela Superior de Ingeniería Informática, University of Vigo, Ourense, Spain

^b Departamento de Informática, Escuela Universitaria de Informática, University of Valladolid, Segovia, Spain



- Mezcla de clasificadores ‘especialistas’ en el manejo de distintos conjuntos de genes relevantes biológicamente
 - Colaboración con Univ. Aveiro (Portugal)
 - Integrado con  **GENEBROWSER** (enriquecimiento)
 - Soporta datos de DNA microarray y RNA-Seq

GeneCommittee - Home

sing.ei.uvigo.es/GC/home.zul

guest_1423395328547

Home



Committee Training

A six steps wizard will guide you in the creation and training of a new committee. Select a training data set, a set of gene sets and your preferred classifiers and form a new committee with the best experts (classifier & gene set)



Diagnostic Mode

Use your committees to evaluate new patients. Each member of your committee will classify your patients using the knowledge acquired on its training. A final summary will be generated with useful diagnostic information



Data Management

Upload and manage your training data sets

Committee Training - Gene Set

sing.ei.uvigo.es/GC/committee/geneSet.zul

guest_1423395328547

GeneCommittee

Navigation

Select top 50 genes using Chi-square

Select Genes Continue

Advanced Configuration

☐ Binarize Numeric Attributes
 ☒ Missing Merge

Parameters Info

Selected Genes

Position	Gene	Ranking	Info
1	VCAN	63.291666666666664	
2	NRGN	61.000000000000000	
3	FLOT1		
4	ASNS		
5	CD14		
6	ITM2C		
7	CCND2		
8	FBP1		
9	SC5DL		
10	LILRB2		
11	RIN2		
12	FAM129A		
13	HEXB		
14	VNN2		

Select the best genes using a feature selection algorithm

1 / 4

Gene Explorer

0 - Gene Report for ASNS

Summary

Gene Name

NARS

Fullname

Asparagine--tRNA ligase, cytoplasmic

Synonyms

No data found

Function

No data found

Location

Cytoplasm

Uniprot Status

Swiss-Prot

Gene Ontology

Pathway

Homologies

Structure

Sequence

References

1 - Gene Report for ASNS

Summary

Gene Name

ASNS

Fullname

Asparagine synthetase [glutamine-hydrolyzing]

Synonyms

TS11

Function

No data found

Antimicrobial Combination Network

[sing-group.org/antimicrobialCombination/]

CLINICAL RESEARCH INFORMATICS

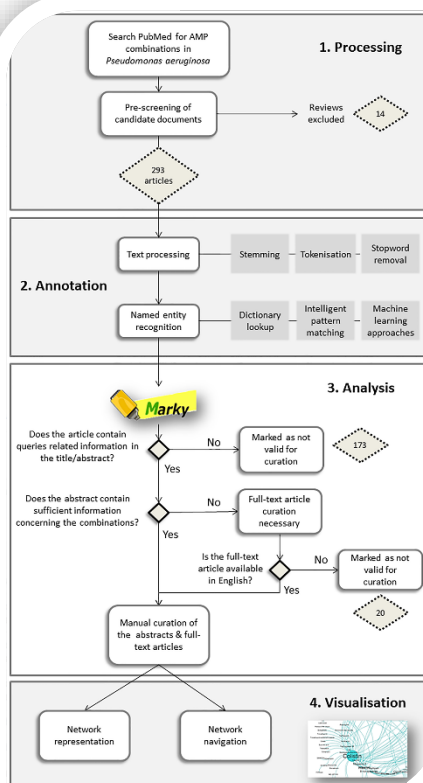
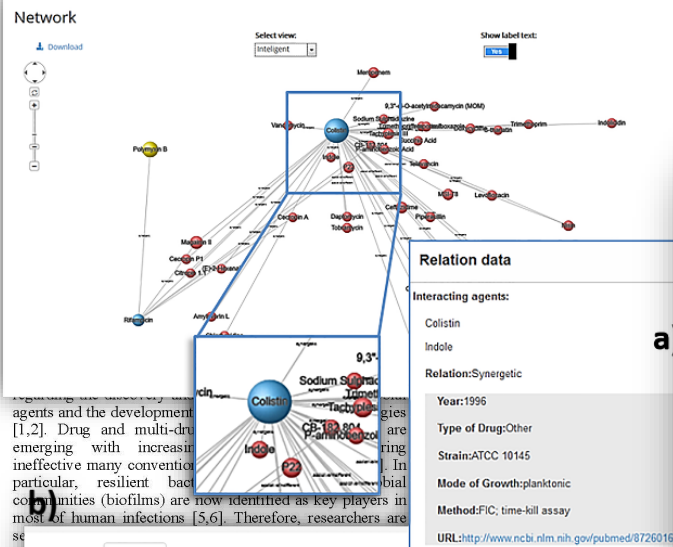
Uses new data and discoveries to improve healthcare.

FOCUS ON:

- ✓ Clinical trials
- ✓ Translational research
- ✓ New medical studies
- ✓ Providing high quality, good outcomes, and efficient care delivery

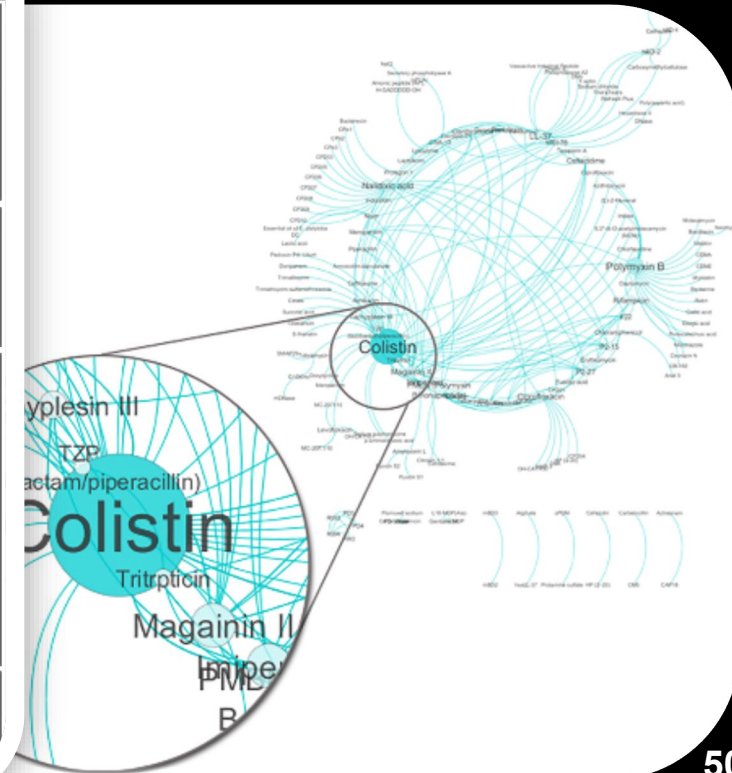
Reconstruction of the Network of Experimentally Validated AMP-Drug Combinations against *Pseudomonas aeruginosa* Infections

Paula Jorge^a, Martín Pérez-Pérez^b, Gael Pérez Rodríguez^b, Florentino Fdez-Riverola^b, Maria Olívia Pereira^a and Anália Lourenço^{a,b}



Red semántica de combinaciones antimicrobianas testeadas experimentalmente contra infecciones de *Pseudomonas aeruginosa*

- Colaboración con Univ. do Minho (Portugal)
- Incluye 239 combinaciones de AMPs y distintos agentes antimicrobianos
- Reconstrucción sistemática combina minería de textos, curación manual y análisis de grafos



BEW [Biofilms Experiment Workbench]

[\[sing-group.org/bew\]](http://sing-group.org/bew)

CLINICAL RESEARCH INFORMATICS
Uses new data and discoveries to improve healthcare.

FOCUSES ON:

- ✓ Clinical trials
- ✓ Translational research
- ✓ New medical studies
- ✓ Providing high quality good outcomes, and efficient care delivery

COMM-3881; No. of Pages 13

ARTICLE IN PRESS

COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE XXX (2015) XXX-XXX

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Enabling systematic, harmonised and large-scale biofilms data computation: The Biofilms Experiment Workbench

Gael Pérez-Rodríguez^a, Daniel Glez-Peña^a, Nuno F. Azevedo^b, Maria Olívia Pereira^c, Florentino Fdez-Riverola^a, Anália Lourenço^{a,c,*}

^a ESEI – Escuela Superior de Ingeniería Informática, Edificio Politécnico, Campus Universitario As Lagoas, s/n, Universidad de Vigo, 32004 Ourense, Spain¹
^b LEPAE – Department of Chemical Engineering, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
^c Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

ARTICLE INFO

Article history:
Received 24 July 2014
Received in revised form
26 December 2014
Accepted 30 December 2014

Keywords:
Clinical microbiology
Biofilms
Bioinformatics
Experimental data workbench
Interchange format
Harmonised vocabulary

ABSTRACT

Background and objective: Biofilms are receiving increasing attention from the biomedical community. Biofilm-like growth within human body is considered one of the key microbial strategies to augment resistance and persistence during infectious processes. The Biofilms Experiment Workbench is a novel software workbench for the operation and analysis of biofilms experimental data. The goal is to promote the interchange and comparison of data among laboratories, providing systematic, harmonised and large-scale data computation.

Method
cator
biom
tistic
exp
Res
inv
of
t

* Corresponding author at: ESEI – Escuela Superior de Ingeniería Informática, Edificio Politécnico, Campus Universitario As Lagoas, s/n, Universidad de Vigo, 32004 Ourense, Spain.
E-mail address: analua@uvigo.es
¹ <http://www.sing.es/uvigo/es/>
<http://dx.doi.org/10.1016/j.cmpb.2014.12.007>
0169-2607/© 2015 Elsevier Ireland Ltd.

- *Gestión de datos y experimentos de biofilms entre distintos laboratorios*
 - Colaboración con Univ. do Minho (Portugal)
 - Permite importar/exportar datos en .xls
 - Ejecución de tests estadísticos y generación de gráficas
 - Generación de informes HTML sobre los datos
 - Comparación de experimentos entre laboratorios → Inter Laboratory Experiment
 - Herramienta oficial de BiofOmics (banco de datos de biofilms pertenecientes a *Salmonella*, *Escherichia*, *Candida*, etc.)



Data input/Output

BEW manage data from:

- ✓ eXtensible Markup Language (XML)
- ✓ XLS
- ✓ Biofomics database

Fueled by the open-source community

BEW uses only open-source external software, including:

- ✓ JXL
- ✓ JFreeChart
- ✓ JSoup
- ✓ R
- ✓ AIBench Framework

Plots and Statistics

- ✓ Statistical bar charts
- ✓ Scatter plots
- ✓ Outliers
- ✓ Normality
- ✓ Heterogeneity
- ✓ Variance

BiofOmics webservices

- ✓ Harmonised terminology
- ✓ Access to public Experiments
- ✓ Automatic submission of Experiments

BEW is open source with license GPLv3

@Note

[sysbio.di.uminho.pt/anote/wiki]

Marky

[sing-group.org/marky]

**CLINICAL RESEARCH INFORMATICS**

Uses new data and discoveries to improve healthcare.

FOCUS ON:

- ✓ Clinical trials
- ✓ Translational research
- ✓ New medical studies
- ✓ Providing high quality good outcomes, and efficient care delivery

Journal of Biomedical Informatics 42 (2009) 710–720

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

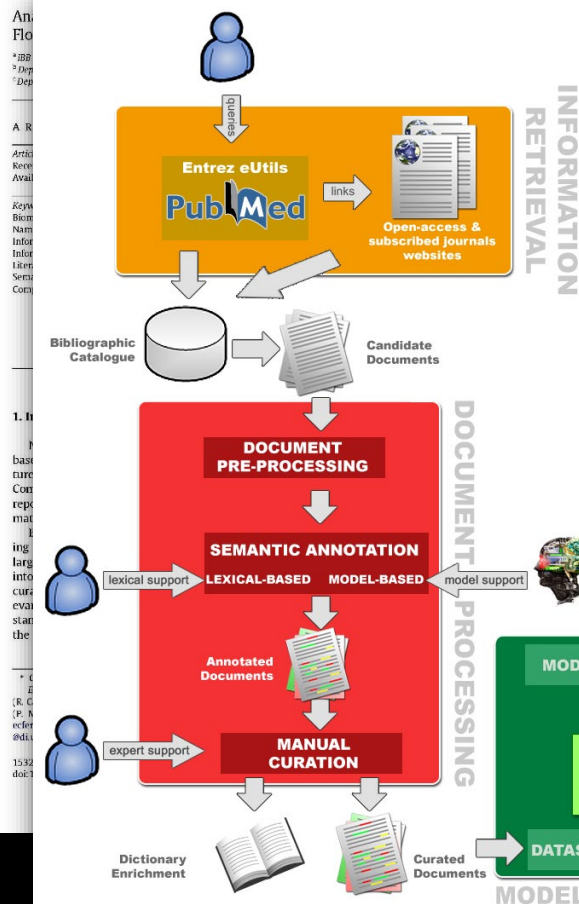
journal homepage: www.elsevier.com/locate/yjbin

COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 118 (2015) 240–251

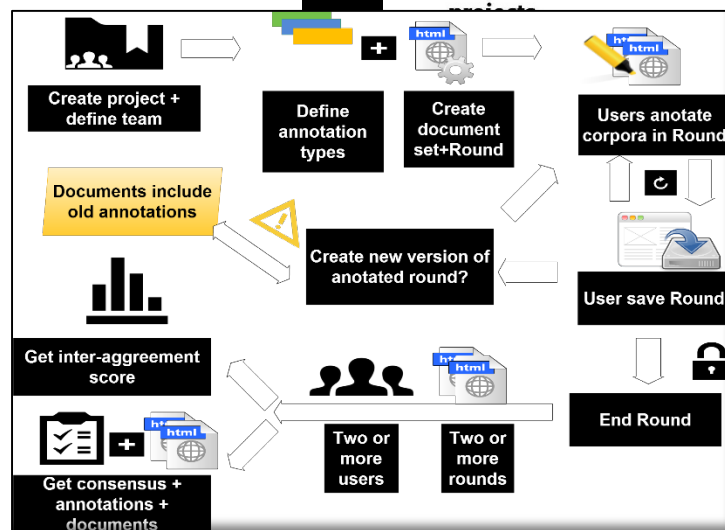
ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

@Note: A workbench for Biomedical Text Mining



Marky: A tool supporting annotation consistency in multi-user and iterative document annotation



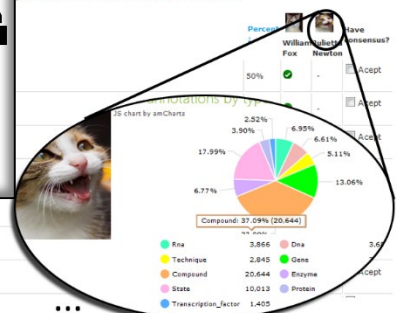
^a, Florentino Fdez-Riverola^a,

^aPolitécnico, Campus Universitario As Lagoas s/n,
4710-057 Braga, Portugal

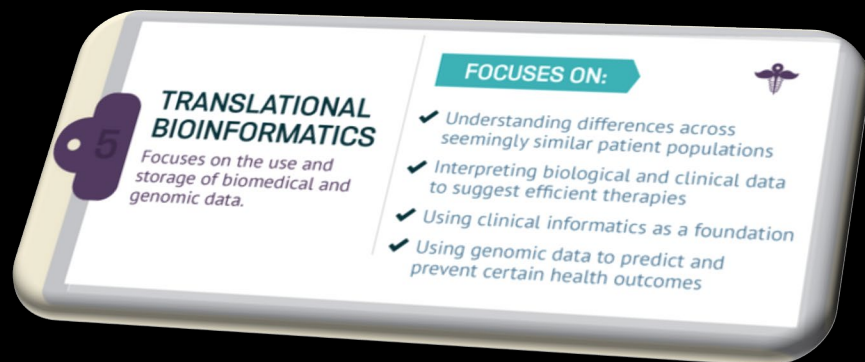
A B S T R A C T

Background and objectives: Document annotation is a key task in the development of text tools and applications. High quality annotated corpora are invaluable, but their creation requires a considerable amount of resources and time. Although the existing tools offer good user interaction interfaces to domain experts, project management and control abilities are still limited. Therefore, the current work introduces a Web-based document annotation tool equipped to manage multi-user and iterative document annotation.

Keywords: Document annotation; Multi-user; Iterative; Web-based



State	9701518	amino acid starvation							
State	7629155	steady state growth							
State	1829119	amino acid - starved							
Compound	2981798	cyclic amp	100%	✓	✓	✓	✓	✓	Accept
Compound	8798619	guanosine tetraphosphate	100%	✓	✓	✓	✓	✓	Accept
Enzyme	19072927	rna polymerases	100%	✓	✓	✓	✓	✓	Accept
Rna	6169010	stable rna	100%	✓	✓	✓	✓	✓	Accept
Protein	6160468	ribosome	100%	✓	✓	✓	✓	✓	Accept
Gene	2981798	relA	100%	✓	✓	✓	✓	✓	Accept
Compound	7629155	ppppp	100%	✓	✓	✓	✓	✓	Accept



Bioinformática traslacional

Pile|Line

 ADOPS

PileLine

[sing-group.org/pileline]

TRANSLATIONAL BIOINFORMATICS

Focuses on the use and storage of biomedical and genomic data.

FOCUS ON:

- Understanding differences across seemingly similar patient populations
- Interpreting biological and clinical data to suggest efficient therapies
- Using clinical informatics as a foundation
- Using genomic data to predict and prevent certain health outcomes

Glez-Peña et al. BMC Bioinformatics 2011, 12:31
http://www.biomedcentral.com/1471-2105/12/31



SOFTWARE

Open Access

PileLine: a toolbox to handle genome position information in next-generation sequencing studies

Daniel Glez-Peña^{1*}, Gonzalo Gómez-López², Miguel Reboiro-Jato¹, Florentino Fdez-Riverola¹, David G Pisano²

Nucleic Acids Research Advance Access published June 6, 2011

Nucleic Acids Research, 2011, 1-5
doi:10.1093/nar/gkr439

PileLineGUI: a desktop environment for handling genome position files in next-generation sequencing studies

Hugo López-Fernández¹, Daniel Glez-Peña^{1*}, Miguel Reboiro-Jato¹, Gonzalo Gómez-López², David G. Pisano² and Florentino Fdez-Riverola¹

¹Higher Technical School of Computer Engineering, University of Vigo, Ourense and ²Bioinformatics Unit (UBio), Structural Biology and Biocomputing Programme, Spanish National Cancer Research Center (CNIO), Madrid, Spain

Received February 25, 2011; Revised April 27, 2011; Accepted May 13, 2011

ABSTRACT

Next-generation sequencing (NGS) technologies are making sequence data available on an unprecedented scale. In this context, new catalogs of Single Nucleotide Polymorphism and mutations generated by resequencing studies are usually stored in genome position files (e.g. Variant Call Format, SAMtools pileup, BED, GFF) comprising of large lists of genomic positions, which are difficult to handle by researchers. Here, we present PileLineGUI, a novel desktop application primarily designed for manipulating, browsing and analysing genome position files (GPFs), with specific support to somatic mutation finding studies. The developed tool also integrates a new genome browser module specially designed for inspecting GPFs. PileLineGUI is free, multiplatform and designed to be intuitively used by biomedical researchers. PileLineGUI is available at: <http://sing.eluvigo.es/pileline/pilelinegui.html>.

INTRODUCTION

Nowadays, next-generation sequencing (NGS) technologies allow the generation of sequence data on an unprecedented scale. The versatility of NGS-based techniques together with the remarkable reduction in per-base sequencing cost, have encouraged researchers to produce a growing number of DNA resequencing studies. The 1000 Genomes Project leads this trend and represents a notable effort to understand the human variation at a genome-wide scale (1). The goal is to report a comprehensive catalogue of DNA variants corresponding to a

number of human populations, expanding the concept of reference human genome from the current haploid and mosaic sequence to a frequency-based model built from population-specific variants. Along the same lines, the detection of novel DNA variants is also an objective for the International Cancer Genome Consortium (ICGC) (<http://www.icgc.org>). This project is currently resequencing more than 50 types of cancer providing new collections of tumour-associated genomic mutations with the aim of relating them to the oncogenic processes, tumour evolution and clinical prognosis (2).

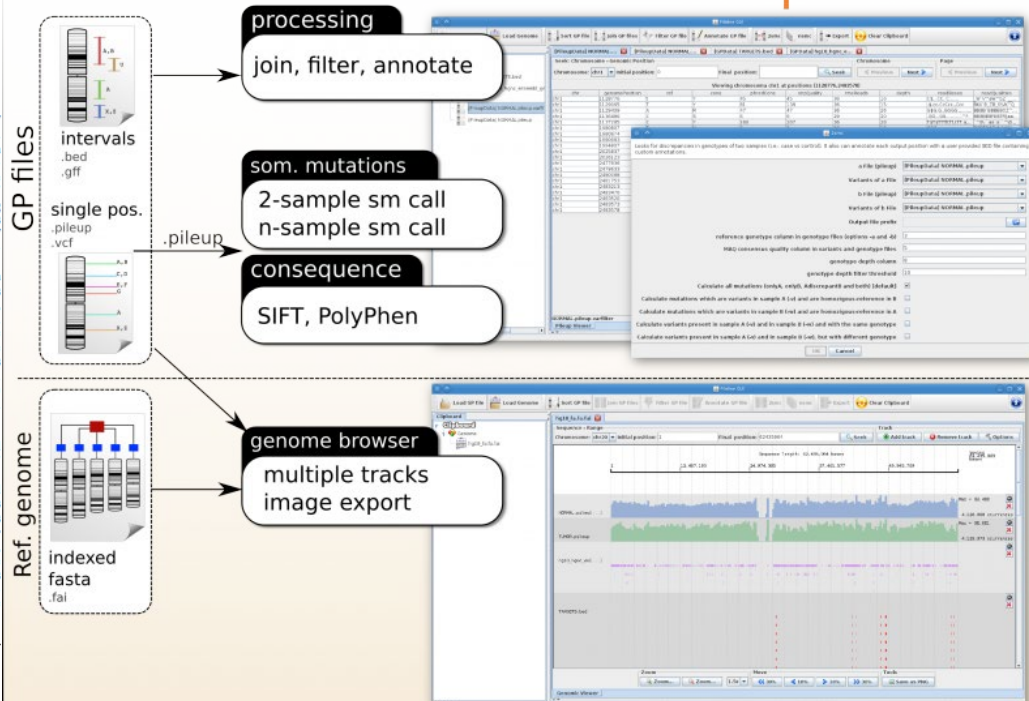
The new catalogues of Single Nucleotide Polymorphism (SNPs) and mutations generated by resequencing studies are usually stored in genome position files. Basically, genome position files (GPFs) consist of tab-delimited text files containing information regarding the base reported by the NGS experiment, defined by chromosome name and genomic coordinates. Several standard formats have been recently proposed to organize the information contained in GPFs, such as the Variant Call Format (vcf) adopted by the 1000 Genomes Project (<http://www.1000genomes.org/wiki/Analysis/Variant%20Call%20Format/vcf-variant-call-format-version-40>), and the .pileup format (3). More traditional GPFs also include the .bed and .gff formats (<http://genome.ucsc.edu/FAQ/FAQformat>).

In such a situation, GPFs analysis becomes a critical task for a deeper understanding of the genome and its relation to human diseases. However, given that GPFs usually include large lists of genomic positions, their handling is cumbersome. This situation calls for efficient, practical and intuitive tools to explore, analyse, visualize and annotate DNA variants. To this end, some authors have recently introduced novel, free and useful applications to deal with standard GPFs. For instance, the

Manejo de datos procedentes de análisis de Ultra-Secuenciación

- Colaboración con CNIO
- Procesado rápido de nuevos ficheros de posición genómica
 - Búsqueda ultra-rápida
 - Comparación de dos muestras para encontrar mutaciones somáticas
- Incorpora un visor genómico de escritorio

Pile | Line GUI



*To whom correspondence should be addressed. Tel: +34 988 387015; Fax: +34 988 387001; Email: dglezpena@uvigo.es

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2011. Published by Oxford University Press.

ADOPS

[sing-group.org/ADOPS]

TRANSLATIONAL BIOINFORMATICS

- FOCUS ON:
- Understanding differences across seemingly similar patient populations
 - Interpreting biological and clinical data to suggest efficient therapies
 - Using clinical informatics as a foundation
 - Using genomic data to predict and prevent certain health outcomes

OPEN ACCESS Freely available online



The *Drosophila melanogaster methuselah* Gene: A Novel Gene with Ancient Functions

Ana Rita Araújo^{1,3}, Micael Reis^{1,3}, Helder Rocha¹, Bruno Aguiar¹, Ramiro Morales-Hojas¹, Sandra Macedo-Ribeiro¹, Nuno A. Fonseca^{2,3}, David Reboiro-Jato⁴, Miguel Reboiro-Jato⁴, Florentino Fdez-Riverola⁴, Cristina P. Vieira^{1,4}

¹Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Porto, Portugal, ²Center of Research in Advanced Computing Systems (CRACS-NRSC), Universidade do Porto, Porto, Portugal, ³FEML: European Brainformatics Institute, Hinxton, Cambridge, United Kingdom, ⁴Departamento de Informática, Universidade de Vigo, Ourense, Spain

Abstract

The *Drosophila melanogaster* G protein less than 10 million years old. Nevertheless, this gene has been implicated in larval development. *meth* belongs to a gene subfamily with this subfamily. Therefore, it is surprising hypothesis that *meth* is an old gene. Using gene with features similar to those of expression analyses we show that the *meth* gene. We also show that, in polymorphism at the *G12490* ortholog differences. Our results imply that G differences in *Drosophila* in general.

Citation: Araújo AR, Reis M, Rocha H, Aguiar B, Morales-Hojas R, Macedo-Ribeiro S, et al. (2013) The *Drosophila melanogaster methuselah* Gene: A Novel Gene with Ancient Functions. PLoS ONE 8(3): e61747. doi:10.1371/journal.pone.0061747

Editor: Suzanne Rutherford, Fred Hutchinson Cancer Research Center

Copyright: © 2013 Araújo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by FEDER Funds, Fundação para a Ciência e a Tecnologia under the PTDC/BIA-BIO/10887/2008, and the project PCOMP-RII/10887/2008. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: jvleira@ibmc.up.pt

† These authors contributed equally to this work.

Introduction

G protein-coupled receptors (GPCRs) are a group of receptor proteins involved in signal transduction. They can be classified into five large families: rhodopsin-like receptors, Adenylate cyclase-coupled receptors [1], and are known to be involved in a variety of biological processes, from light vision to

Journal of Experimental Botany, Vol. 64, No. 8, pp. 2423–2434, 2013

doi:10.1093/jxb/erl096 Advance Access publication 18 April, 2013
This paper is available online free of access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)



RESEARCH PAPER

Patterns of evolution at the gametophytic self-incompatibility *Sorbus aucuparia* (Pyrinae) S pollen genes support the non-self recognition by multiple factors model

Bruno Aguiar¹, Jorge Vieira¹, Ana E. Cunha¹, Nuno A. Fonseca², David Reboiro-Jato⁴, Miguel Reboiro-Jato⁴, Florentino Fdez-Riverola⁴, Olivier Raspé³ and Cristina P. Vieira^{1,4}*

¹ Instituto de Biologia Molecular e Celular (IBMC), University of Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal

² CRACS-INESC Porto, Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal

³ National Botanic Garden of Belgium, Domaine van Bouchout, B-1860 Meise, Belgium

⁴ Escuela Superior de Ingeniería Informática, Edificio Politécnico, Campus Universitario s/n, University of Vigo, 32004 Ourense, Spain

* To whom correspondence should be addressed. Email: cgvieira@ibmc.up.pt

Received 12 December 2012; Revised 5 March 2013; Accepted 11 March 2013

Abstract

S-RNase-based gametophytic self-incompatibility evolved once before the split of the Asteridae and Rosidae. In *Prunus* (tribe Amygdaloideae of Rosaceae), the self-incompatibility *S*-pollen is a single F-box gene that presents the expected evolutionary signatures. In *Malus* and *Pyrus* (subtribe Pyrinae of Rosaceae), however, clusters of F-box genes (called *SBFB*) have been described in pollen only and are linked to the *S-RNase* gene. Although polymorphic, *SBFB* genes present levels of diversity lower than those of the *S-RNase* gene. They have been suggested as putative *S*-pollen genes, in a system of non-self recognition by multiple factors. Subsets of allelic products of the different *SBFB* genes interact with non-self *S-RNases*, marking them for degradation, and allowing compatible pollination. This study performed a detailed characterization of *SBFB* genes in *Sorbus aucuparia* (Rosa

Automatiza un flujo de trabajo para la detección automática de sitios seleccionados positivamente (detección de mutaciones)

- Colaboración con IBMC
- Integra T-Coffee, MrBayes y CodeML
- Utiliza ALTER para la conversion de formatos entre distintos programas
- Proporciona múltiples resúmenes y vistas de los resultados

