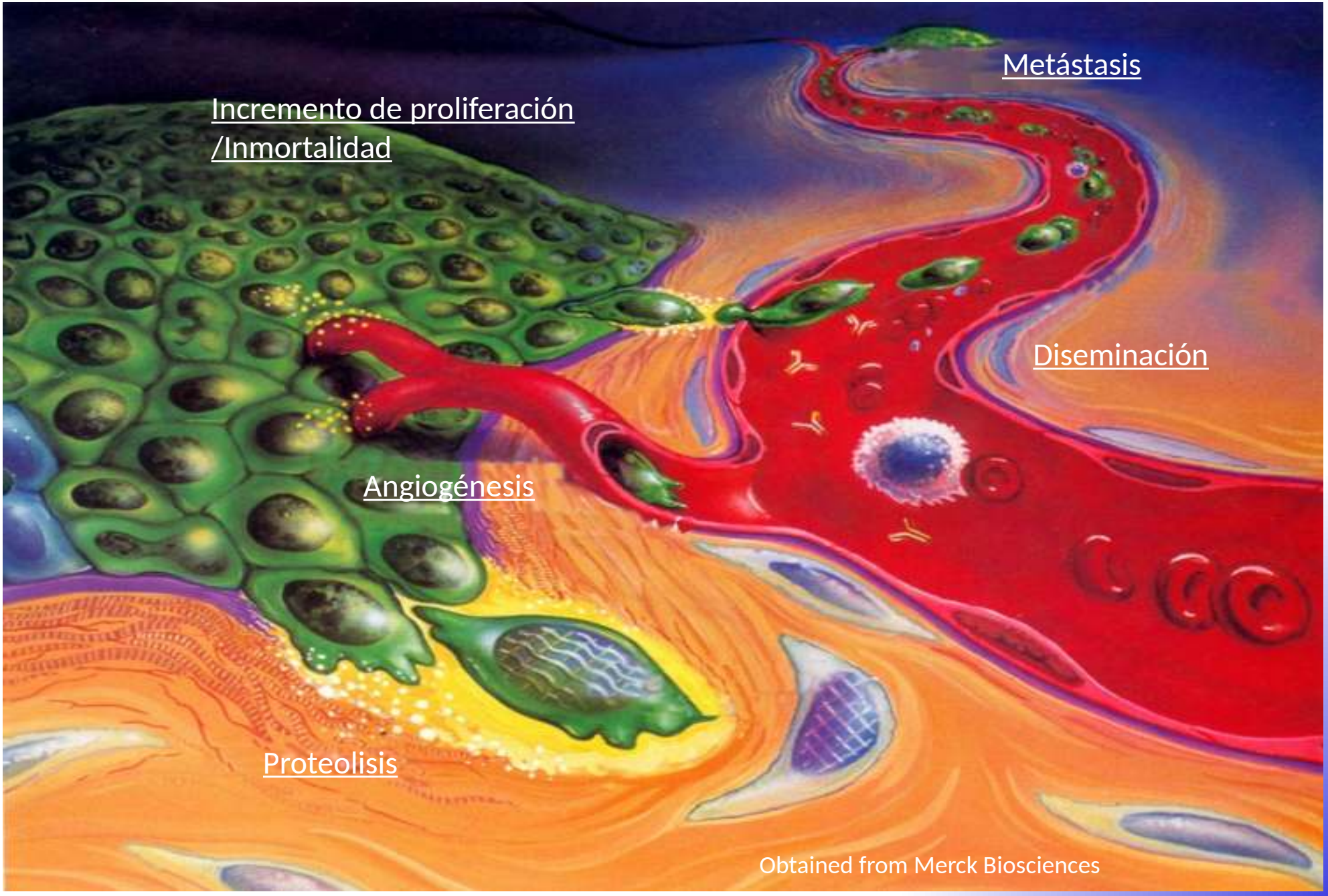
Análisis de datos biológicos en genómica del cáncer

Ignacio Varela Egocheaga
ignacio.varela@unican.es

Naturaleza del cáncer



El ADN, el libro de instrucciones de la célula

Courtesy of http://www.dreamstime.com

Courtesy of http://www.freepik.es

Éxitos derivados del estudio del ADN de células tumorales

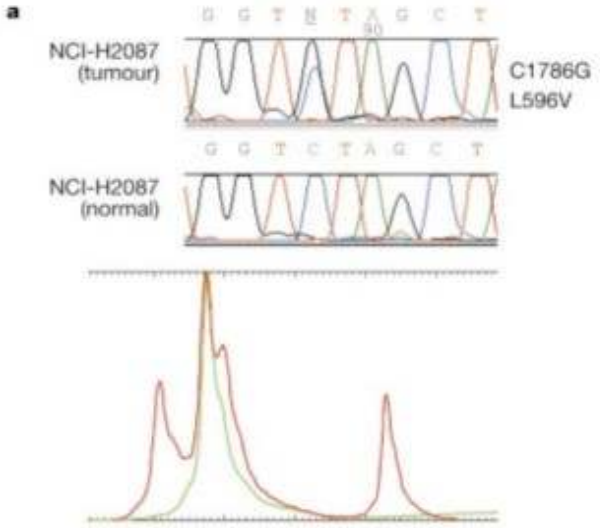
letters to nature

Mutations of the *BRAF* gene in human cancer

Helen Davies[1,2], Graham R. Bignell[1,2], Charles Cox[1,2], Philip Stephens[1,2], Sarah Edkins[1], Sheila Clegg[1], Jon Teague[1], Hayley Woffendin[1], Mathew J. Garnett[3], William Bottomley[1], Neil Davis[1], Ed Dicks[1], Rebecca Ewing[1], Yvonne Floyd[1], Kristian Gray[1], Sarah Hall[1], Rachel Hawes[1], Jaime Hughes[1], Vivian Kosmidou[1], Andrew Menzies[1], Catherine Mould[1], Adrian Parker[1], Claire Stevens[1], Stephen Watt[1], Steven Hooper[3], Rebecca Wilson[3], Hiran Jayatilake[4], Barry A. Gusterson[5], Colin Cooper[6], Janet Shipley[6], Darren Hargrave[7], Katherine Pritchard-Jones[7], Norman Maitland[8], Georgia Chenevix-Trench[9], Gregory J. Riggins[10], Darell D. Bigner[10], Giuseppe Palmieri[11], Antonio Cossu[12], Adrienne Flanagan[13], Andrew Nicholson[14] Judy W. C. Ho[15], Suet Y. Leung[16], Siu T. Yuen[16], Barbara L. Weber[17], Hilliard F. Seigler[18], Timothy L. Darrow[18], Hugh Paterson[3], Richard Marais[3], Christopher J. Marshall[3], Richard Wooster[1,6], Michael R. Stratton[1,4] & P. Andrew Futreal[1]

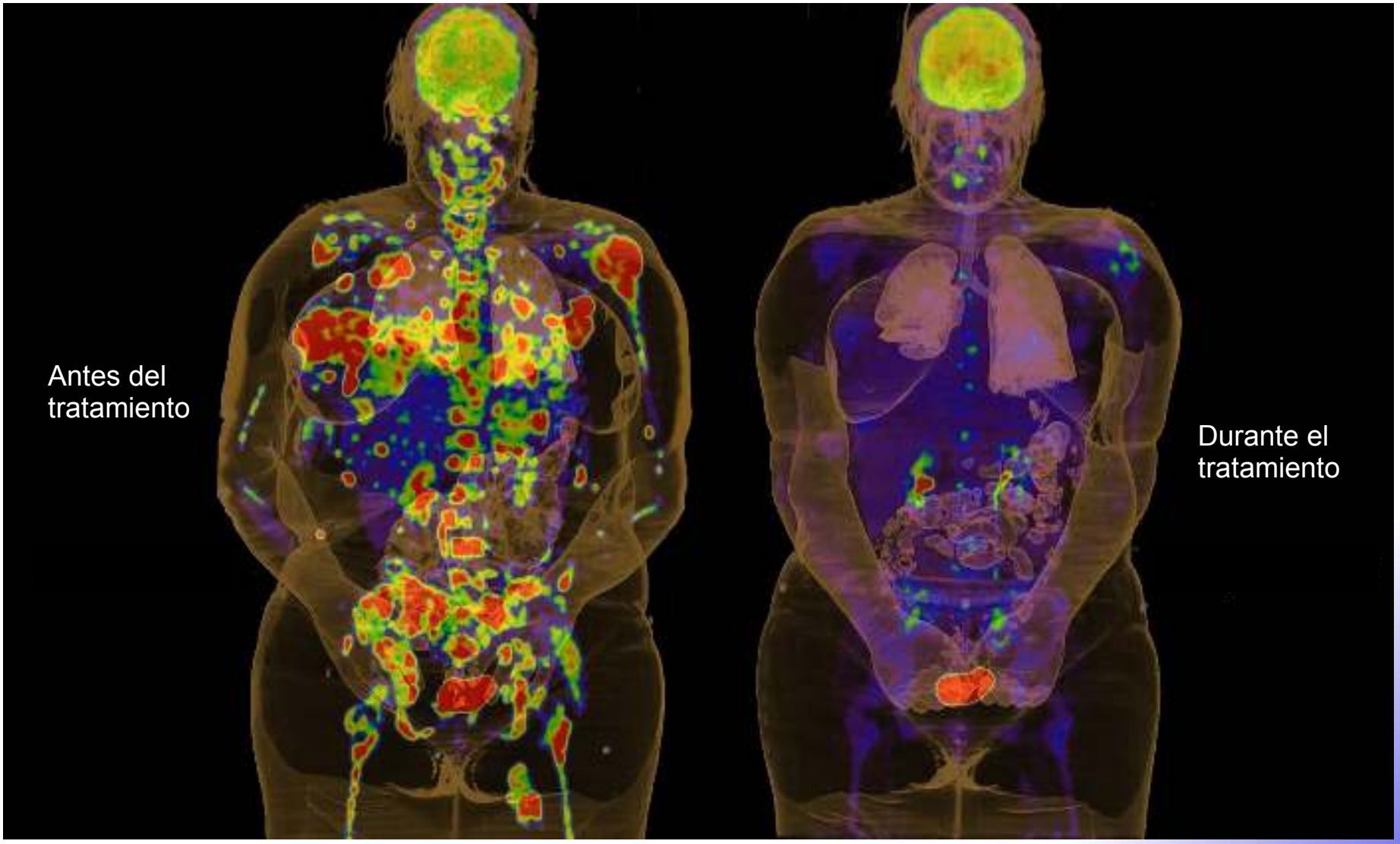
[1]*Cancer Genome Project, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, UK*
[3]*Cancer Research UK Centre for Cell and Molecular Biology, Chester Beatty Labs, Institute of Cancer Research, London SW3 6JB, UK*
[4]*Section of Cancer Genetics;* [6]*Section of Molecular Carcinogenesis; and* [7]*Section of Paediatrics, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK*
[5]*Department of Pathology, Western Infirmary, University of Glasgow, S11 6NT, UK*
[8]*Department of Biology, YCR Cancer Research Unit, University of York, York YO10 5YW, UK*
[9]*Queensland Institute of Medical Research, RBH Post Office Herston, Queensland 4029, Australia*
[10]*Department of Pathology, and* [18]*Department of Surgery, Duke University Medical Centre, Durham, North Carolina 27710, USA*
[11]*Institute of Molecular Genetics, C.N.R., Loc. Tramariglio, Alghero 07040, Italy*
[12]*Department of Pathology, University of Sassari, Azienda USL1, Sassari 07100, Italy*
[13]*Royal Free & University College Medical School, London WC1E 6JJ, UK*
[14]*Royal Brompton Hospital, London SW3 6NP, UK*
[15]*Department of Surgery, and* [16]*Department of Pathology, The University of Hong Kong, Queen Mary Hospital, Hong Kong*

phoblastoid cell lines from the same individuals were screened for sequence variants through the coding exons and intron–exon junctions of the *BRAF* gene using a capillary-based modified heteroduplex method followed by direct sequencing of polymerase chain reaction products. (Exon 1, containing 135 base pairs (bp) of coding sequence, failed to amplify despite the use of five different primer sets.) Three single-base substitutions were detected. Two were in *BRAF* exon 15: T1796A leading to a substitution of valine by glutamic acid at position 599 (V599E) in the melanoma cell line Colo-829, and C1786G leading to L596V in the NSCLC cell line NCI-H2087 (Fig. 1). A further mutation was found in exon 11: G1403C leading to G468A in the NSCLC cell line NCI-H1395. None of the three changes were present in the lymphoblastoid cell lines from the same individuals, indicating that the variants were somatically acquired mutations.



Davies H et al. *Nature* 417:949 (2002)

Éxitos derivados del estudio del ADN de células tumorales



Éxitos derivados del estudio del ADN de células tumorales

ASH 50th anniversary review

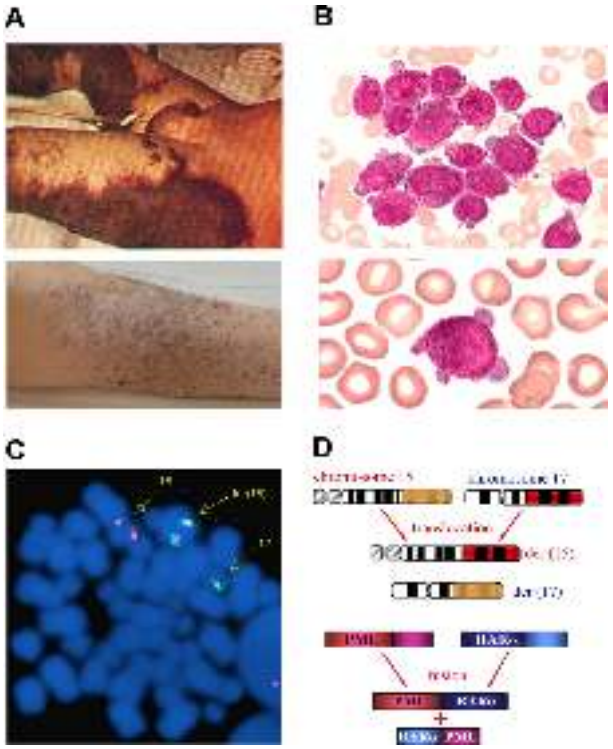
Acute promyelocytic leukemia: from highly fatal to highly curable

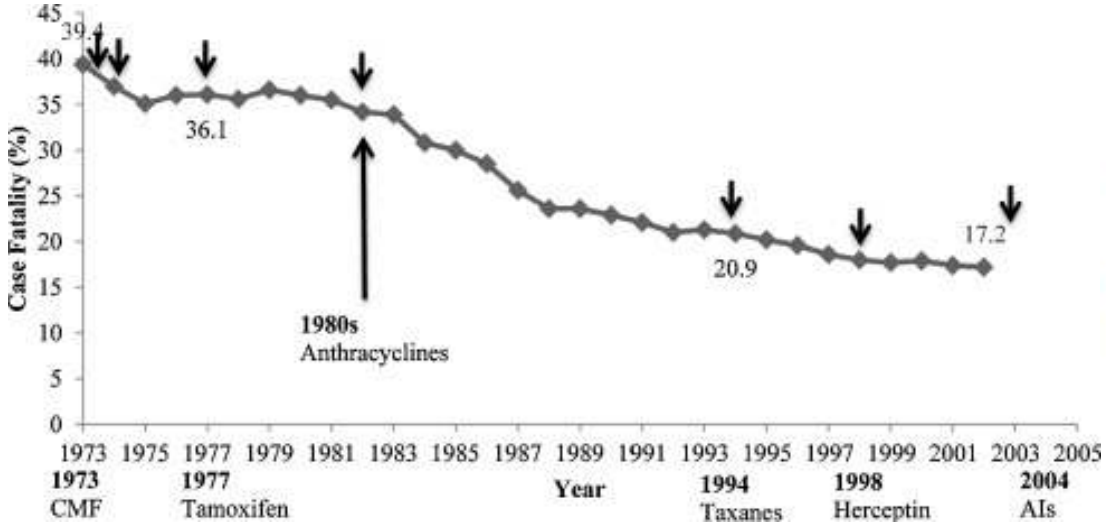
Zhen-Yi Wang[1] and Zhu Chen[1,2]

[1]Shanghai Institute of Hematology and State Key Laboratory of Medical Genomics, Rui Jin Hospital affiliated to the Shanghai Jiao Tong University (SJTU School of Medicine, Shanghai; and [2]Shanghai Center for Systems Biomedicine at SJTU, Shanghai, China

Acute promyelocytic leukemia (APL) is a distinct subtype of acute myeloid leukemia. Morphologically, it is identified as the M3 subtype of acute myeloid leukemia by the French-American-British classification and cytogenetically is characterized by a balanced reciprocal trans- bleeding tendency and a fatal course of only weeks. Chemotherapy (CT; daunorubicin, idarubicin and cytosine arabinoside) was the front-line treatment of APL with a complete remission (CR) rate of 75% to 80% in newly diagnosed patients. Despite all these progresses, the median was raised up to 90% to 95% and disease free survival (DFS) to 74 use of arsenic trioxide (ATO) sinc 1990s further improved the clinic come of refractory or relapsed as newly diagnosed APL. In this arti review the history of introduction o

Wang Z et al. *Blood* 111:2505 (2008)





Original Research

Why have breast cancer mortality rates declined?

Steven A. Narod[a,b,*], Javaid Iqbal[a], Anthony B. Miller[b]

[a] Women's College Research Institute, Women's College Hospital, Toronto, ON, Canada
[b] Dalla Lana School of Public Health, Toronto, ON, Canada

Narod S et al. *J Cancer Pol* 5:8 (2015)

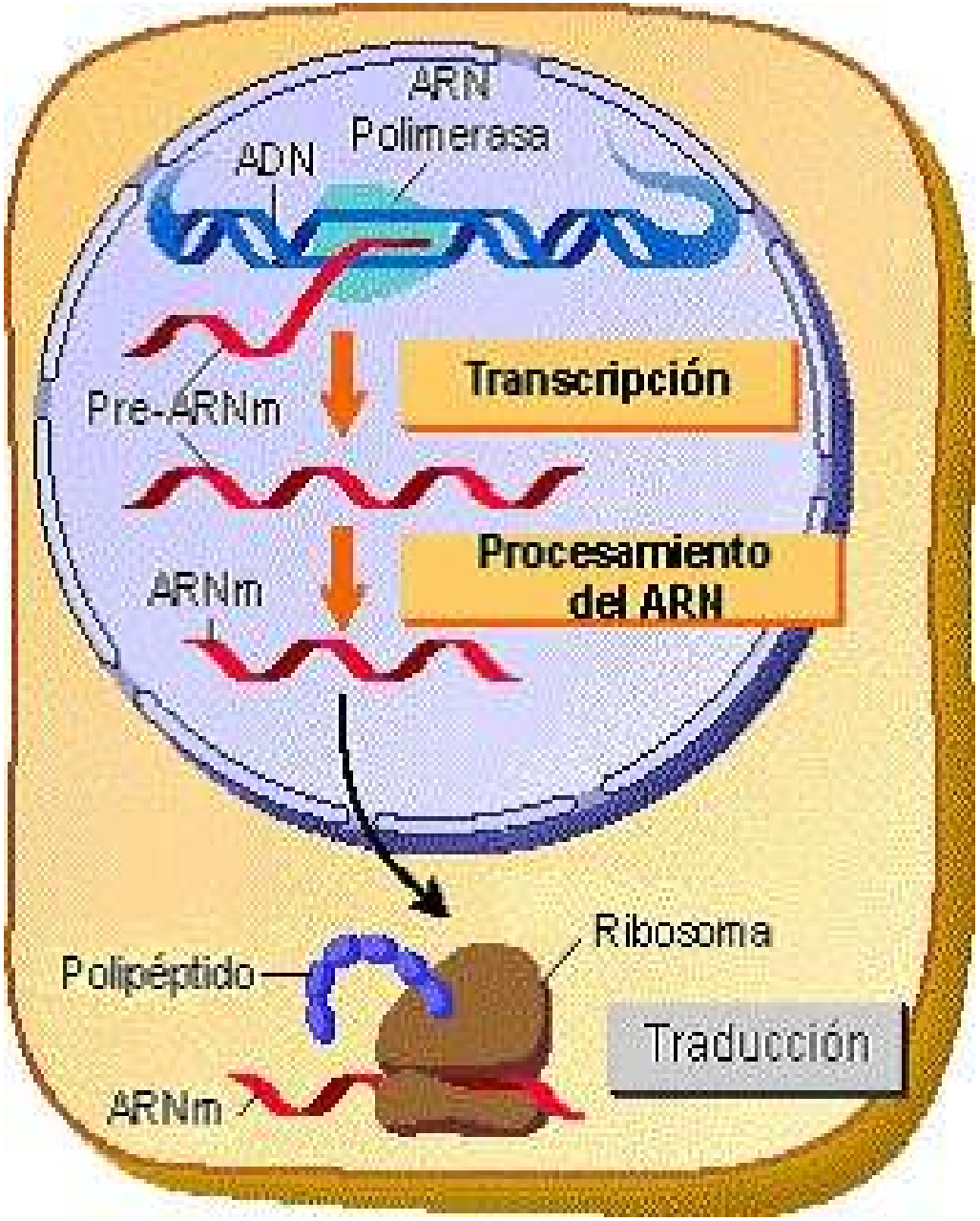
*El ADN, **la biblioteca** de instrucciones de la célula*

Courtesy of http://www.dreamstime.com

Courtesy of http://www.morguefile.com

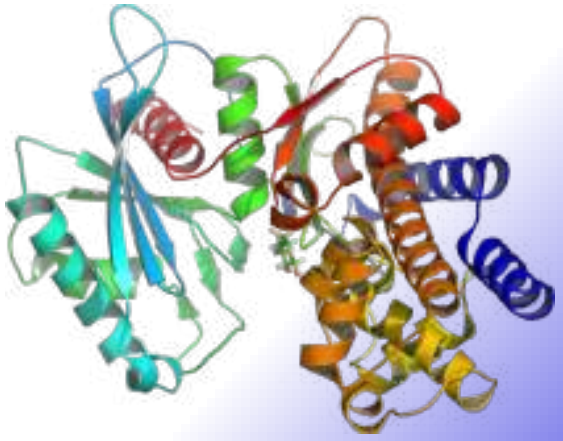
Nuestro genoma está compuesto por 3.000.000.000 letras (nt) ~> 2.000 libros

El DNA contiene las instrucciones para crear todos los compenentes celulares



ADN: Una biblioteca de información

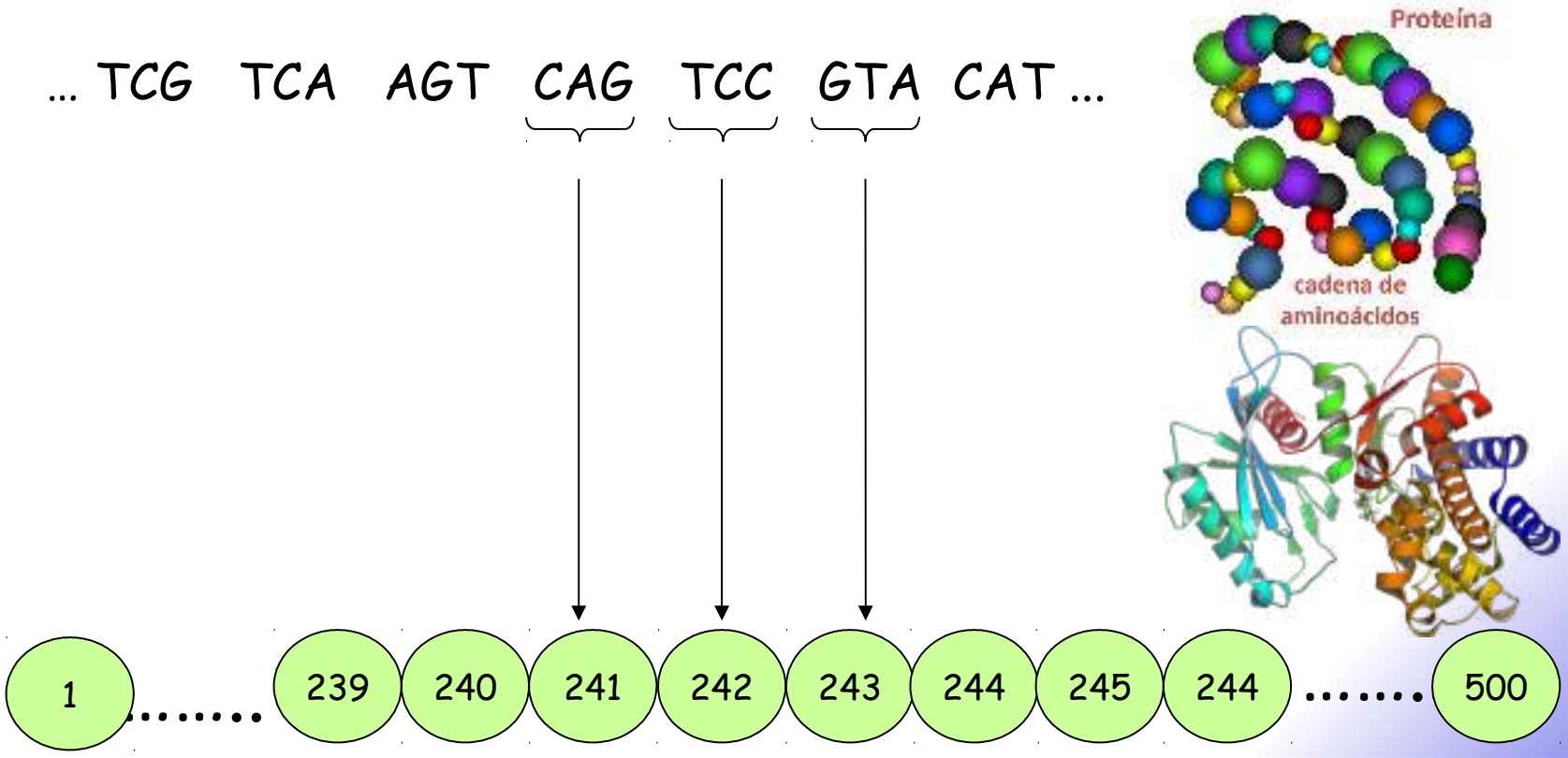
Gen: Un manual de instrucciones



Proteínas: Unidades funcionales básicas

La secuencia del ADN copia a RNA mensajero se TRADUCE a amionácidos para generar las proteínas

- El mensaje se lee en palabras de tres letras (tripletes de bases)
- Cada triplete de bases codifica para un aminoácido (tenemos 20 diferentes)
  - La secuencia de bases determina la secuencia de aminoácidos



El desciframiento del código genético nos ha dado un premio Nobel. Este conocimiento tiene menos de 60 años.

N	AMINOACIDOS	TRIPLETES DE BASES DEL mRNA					
	Ácido aspártico	GAC	GAU				
	Ácido glutámico	GAA	GAG				
	Alanina	GCA	GCC	GCG	GCU		
	Arginina	AGA	AGG	CGA	CGC	CGG	CGU
	Asparagina	AAC	AAU				
	Cisteína	UGC	UGU				
	Fenilalanina	UUC	UUU				
	Glutamina	CAA	CAG				
	Glicina	GGA	GGC	GGG	GGU		
	Histidina	CAC	CAU				
	Isoleucina	AUA	AUC	AUU			
	Leucina	CUA	CUC	CUG	CUU	UUA	UUG
	Lisina	AAA	AAG				
	Metionina	AUG					
	Prolina	CCA	CCC	CCG	CCU		
	Serina	AGC	AGU	UCA	UCC	UCG	UCU
	Treonina	ACA	ACC	ACG	ACU		
	Triptófano	UGG					
	Tirosina	UAC	UAU				
	Valina	GUA	GUC	GUG	GUU		

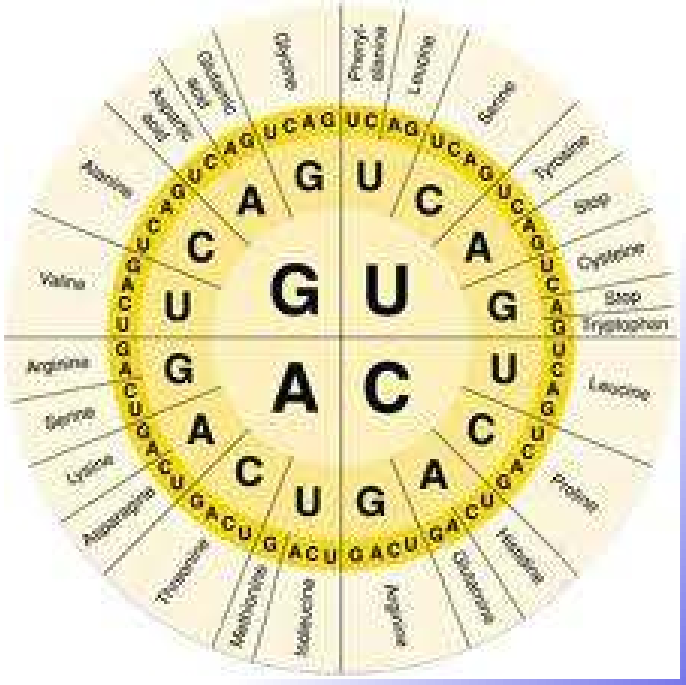
GTA ATC GAC GCC

V I D A

Premio Nobel de Medicina 1959

Severo Ochoa

Arthur Kornberg



ibbtec

Estudiar todo el ADN de una persona no es una tarea fácil

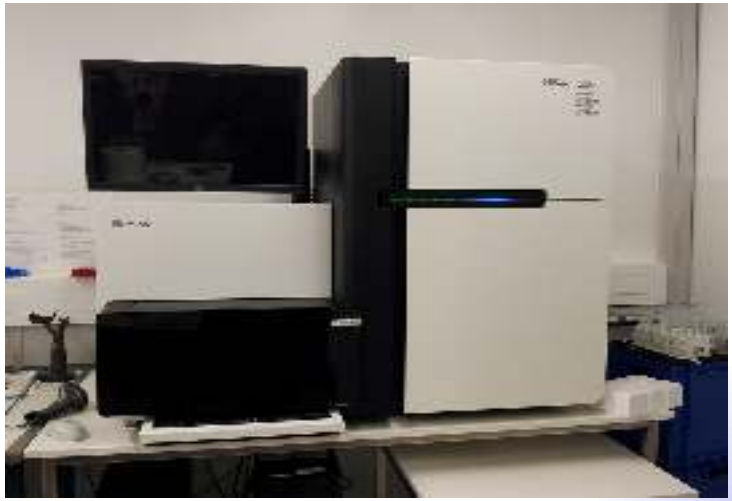
(1990-2003)

http://www.blog.gurukpo.com/wp-content/uploads/2012/08/metal-gear-solid-human-genome-project-big.jpg

- Costó ~2.700 millones de dólares
- Se tardaron 13-15 años
- Se llevó a cabo gracias a la colaboración de cientos de investigadores de al menos 6 países distintos.

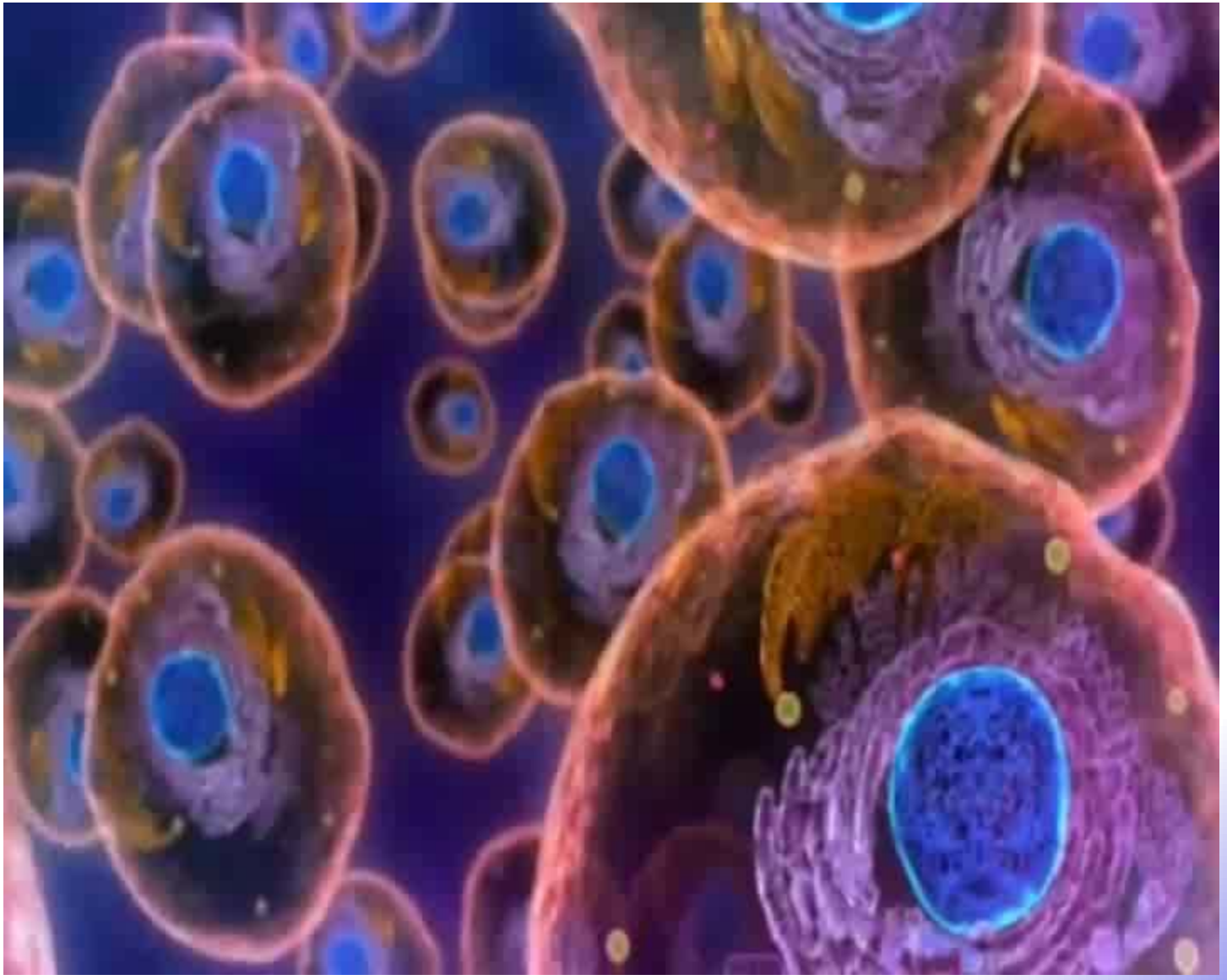
(2010...)

- Coste ~3.000 €.
- Un secuenciador necesita 3 días.



http://en.wikipedia.org/wiki/DNA_sequencing

Estudiar todo el ADN de una persona no es una tarea fácil



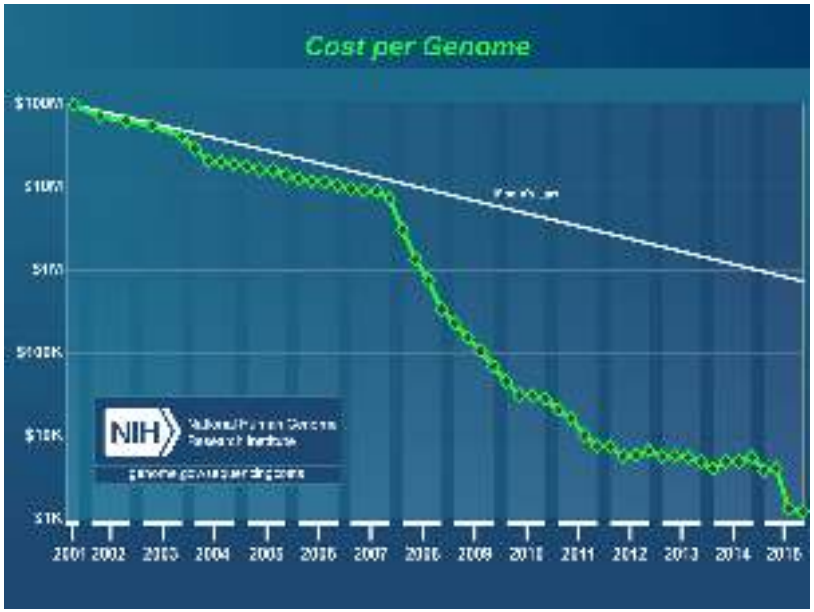
Estudiar todo el ADN cada vez más fácil

https://nanoporetech.com/

Problemas técnicos asociados a la NGS

Cantidad de datos

Los datos de secuenciación brutos de un paciente ocupan típicamente unos 20 Gb de texto. Esto genera problemas en el almacenamiento y transferencia de los datos entre colaboradores.



Capacidad de procesamiento

El alineamiento/reorganización de los datos para identificar el origen de cada una de las cientos de millones de secuencias pequeñas requiere típicamente varias horas de procesamiento. En proyectos con una gran cantidad de muestras es necesario el acceso a sistemas de computación masiva paralela.

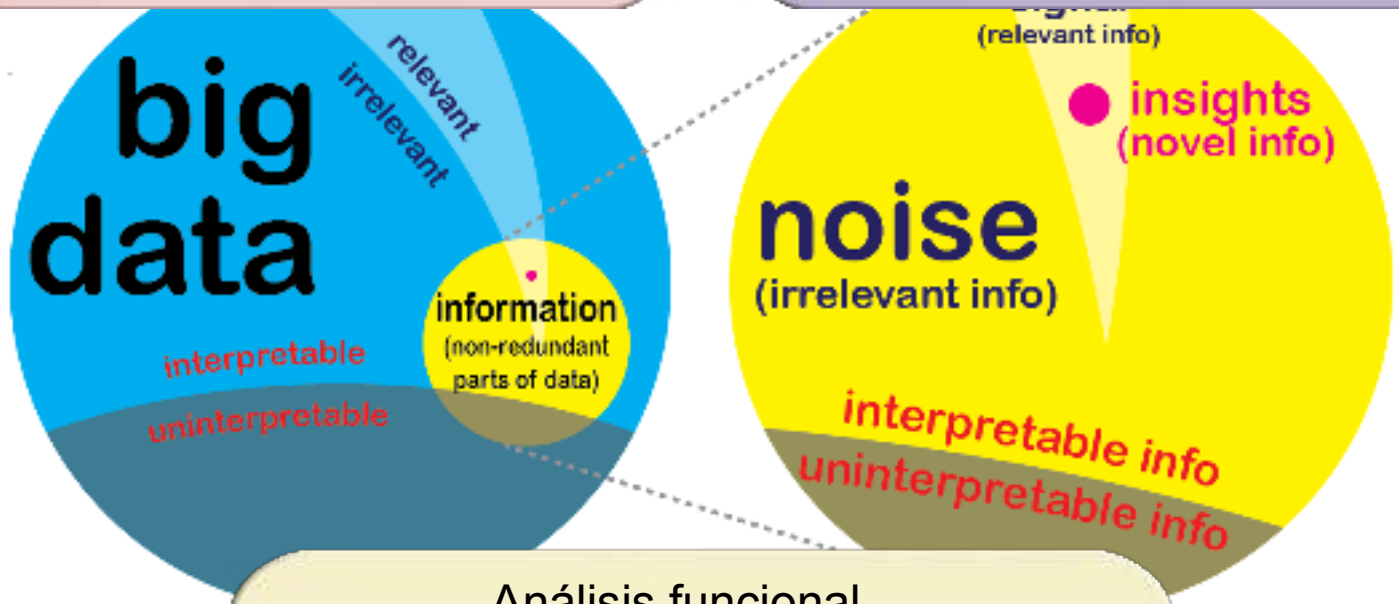
Problemas analíticos asociados a la NGS

Alineamiento

Las técnicas de secuenciación masiva generan típicamente cientos de millones de secuencias cortas con las que hay que reconstruir el genoma secuenciado

Identificación de variantes

El proceso de secuenciación está asociado a la generación de artefactos en la generación de librería, secuenciación y alineamiento que hay que computar para identificar variantes reales.



Análisis funcional

La interpretación de las posibles consecuencias funcionales de las variantes requiere un amplio conocimiento biológico y diversas herramientas de comparativa entre distintas bases de datos públicas.

Proyecto genoma del cáncer

International Cancer Genome Consortium

Login | Create new account with OpenID

Search

Overview | Cancer Genome Projects | Committees | Policies and Guidelines | Media | Contacts

International Cancer Genome Consortium

Brain Cancer
United States

Breast Cancer
European Union / United Kingdom

Breast Cancer
France

Breast Cancer
United Kingdom

Chronic Lymphocytic Leukemia
Spain

Colon Cancer
United States

Gastric Cancer
China

Leukemia
United States

Liver Cancer
France

Liver Cancer
Japan

(Show)

ICGC Goal: To obtain a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types and/or subtypes which are of clinical and societal importance across the globe.

Lung Cancer
United States

Lung Cancer
United States

Oral Cancer
India

Ovarian Cancer
Australia

Ovarian Cancer
United States

Pancreatic Cancer
Australia

Pancreatic Cancer
Canada

Pediatric Brain Tumors
Germany

Renal Cancer
European Union / France

Test
Italy

Policies and Guidelines
International Cancer Genome Consortium

Access to Data
Data Coordination Center

Apply for Access to Controlled Data
Data Access Compliance Office

info@icgc.org | dcc@icgc.org | daco@icgc.org | Contacts

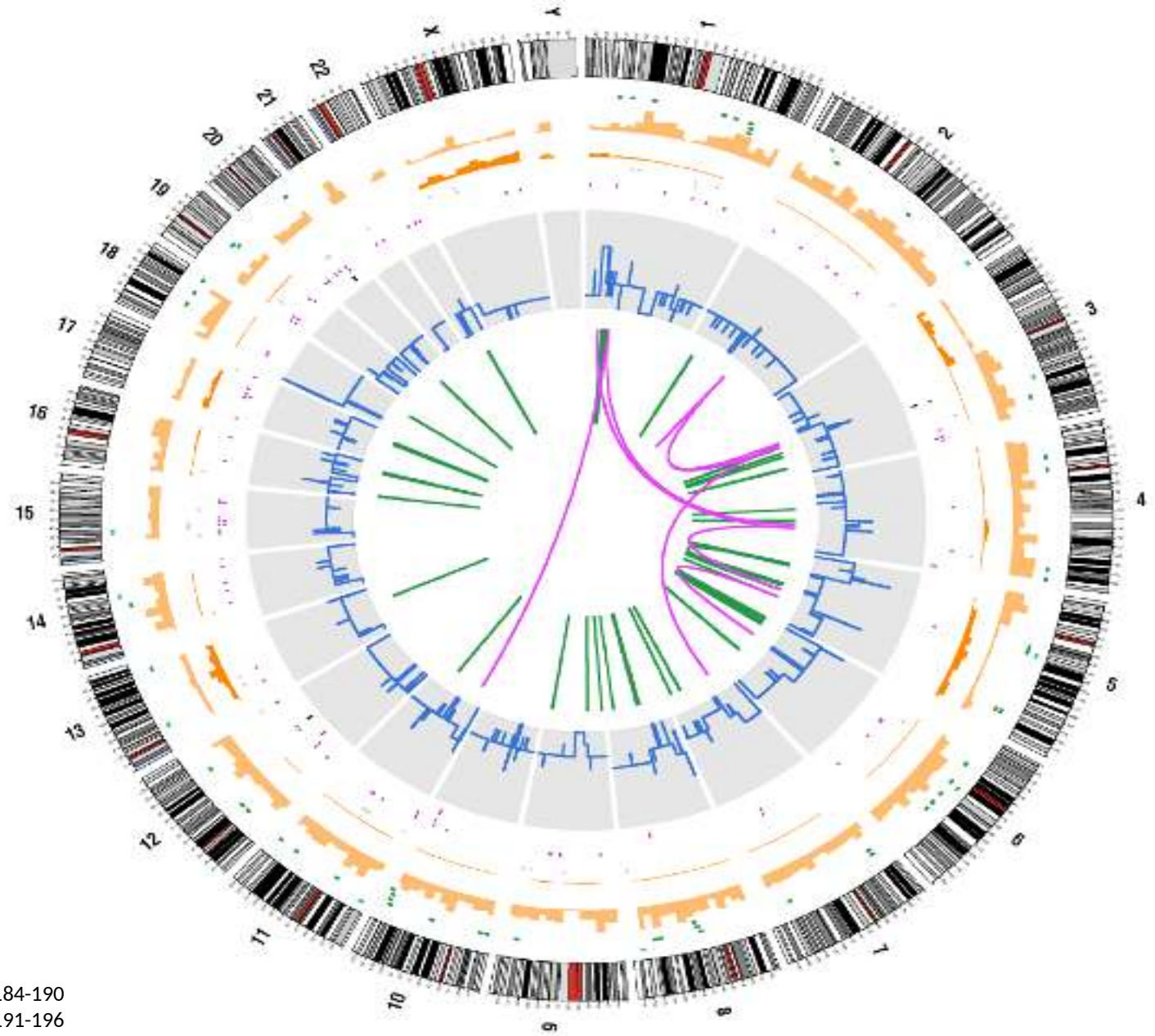
© 2010 International Cancer Genome Consortium. All rights reserved. Terms & Conditions | Privacy Policy

Last updated on: 23 January 2010

Los tumores tienen muchas alteraciones en el ADN (mutaciones)

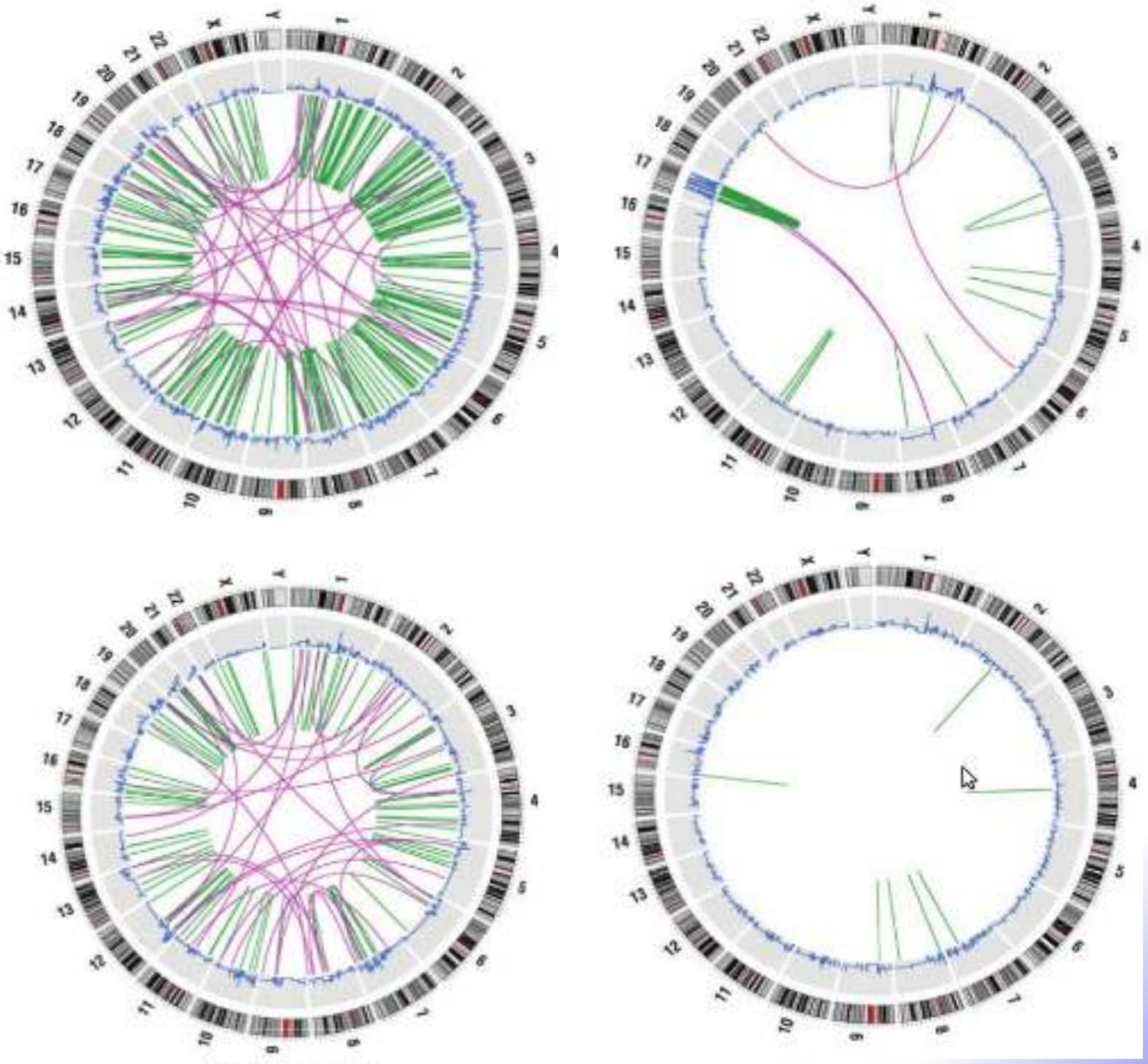
~ cada persona tiene de media unos 3.000.000 de cambios en su ADN.

~ 25.000 cambios únicos del tumor no presentes en las células normales.



Pleasance et al. *Nature* (2010) 463:184-190
Pleasance et al. *Nature* (2010) 463:191-196

Cada tumor es diferente: Medicina personalizada

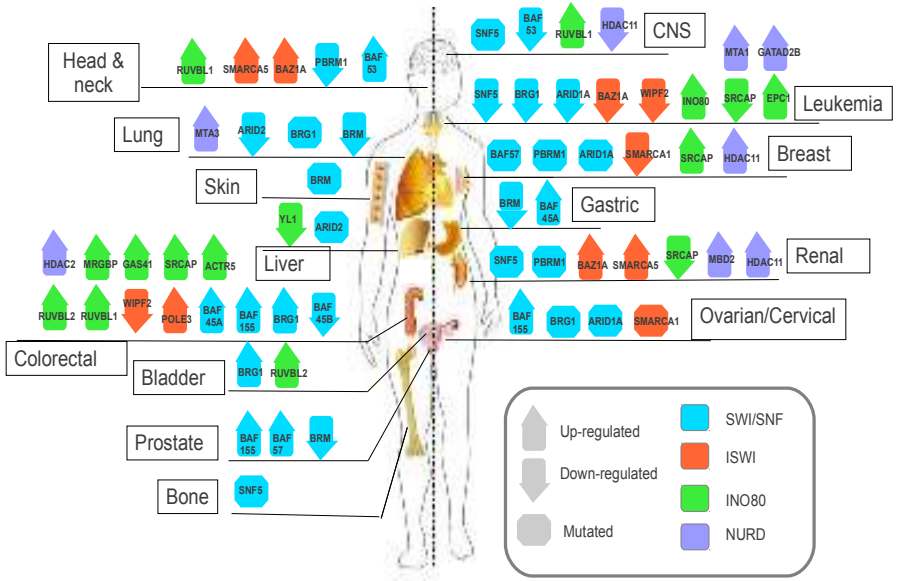


Stephens et al. *Nature* (2009) 462:1005-1010

Proyectos de investigación en marcha en nuestro laboratorio.

GOBIERNO DE ESPAÑA | MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Cambios en la estructura del ADN en cáncer.



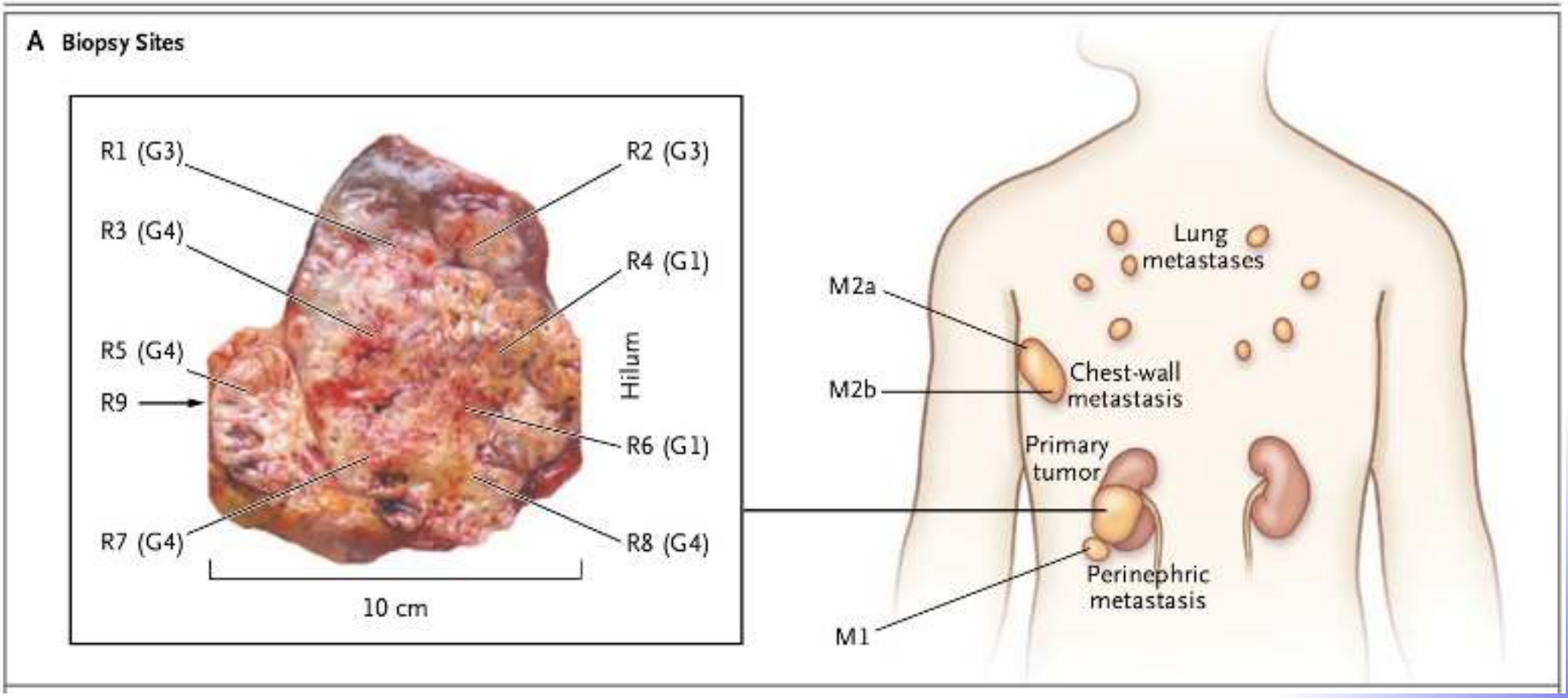
INSTITUTO DE INVESTIGACIÓN CONTRA LA LEUCEMIA JOSEP CARRERAS

Dr. Pablo Menéndez Buiján

Estudio genético de la leucemia infantil.

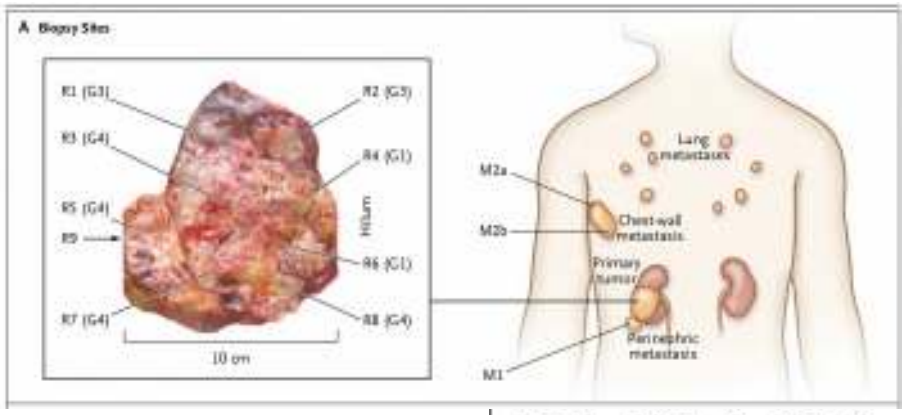
http://trialx.com/curetalk/wp-content/blogs.dir/7/files/2012/10/childhood-ALL.png

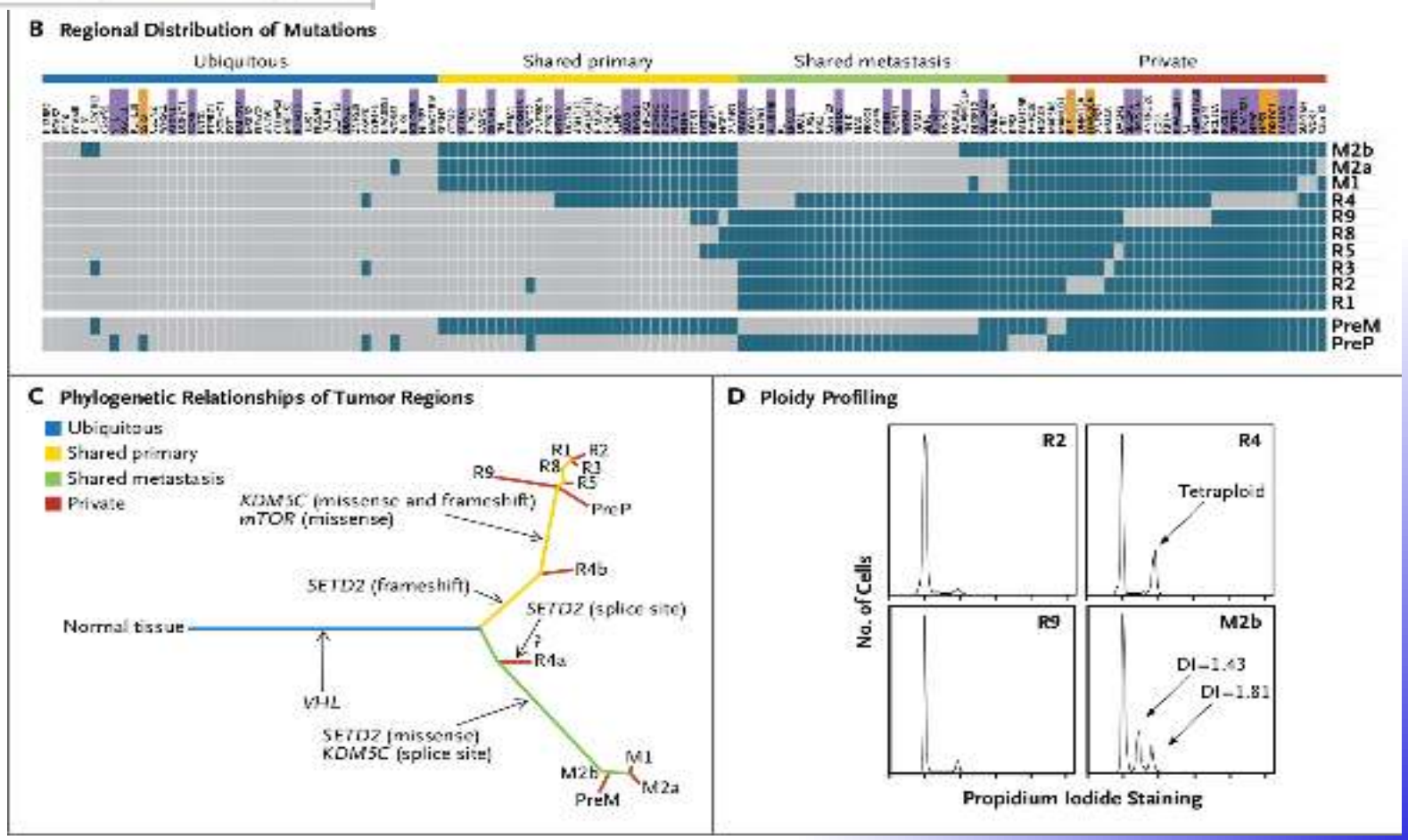
Variabilidad interna del tumor.



Marco et al. *NEJM* (2012) 366:886

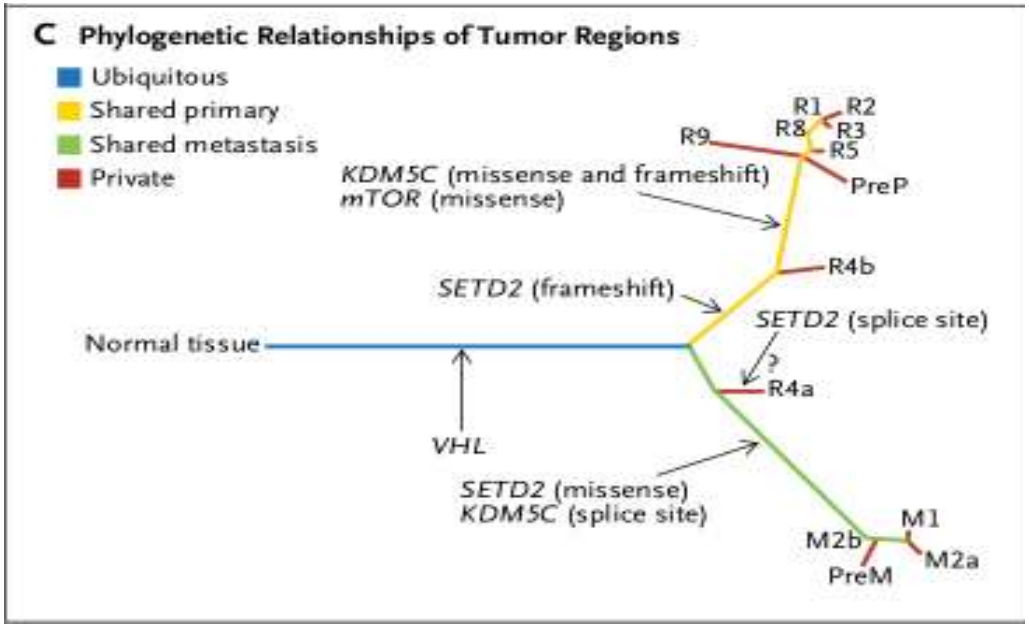
Variabilidad interna del tumor.



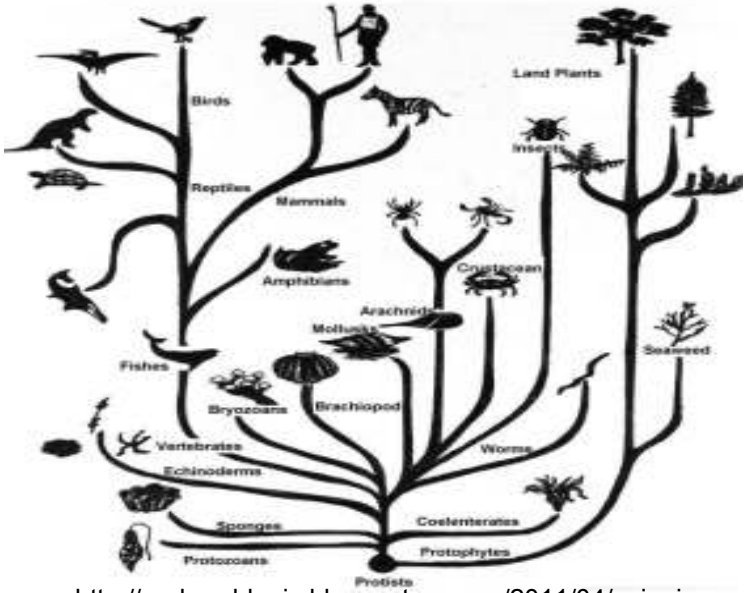


Marco et al. *NEJM* (2012) 366:886

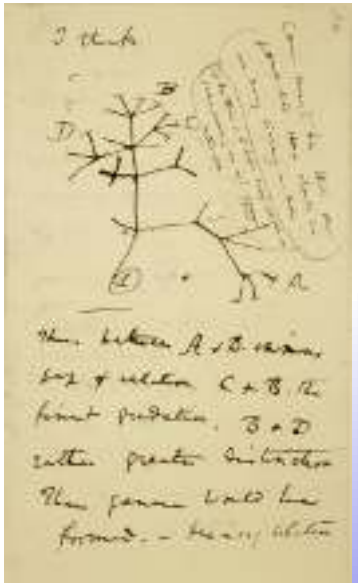
Existencia de diversas colonias celulares dentro del tumor. Evolución de Darwin.



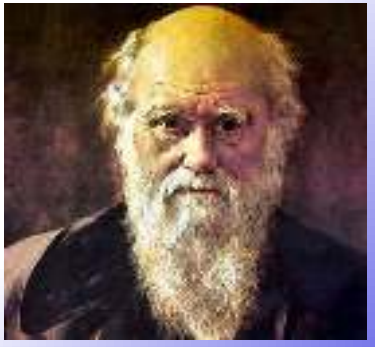
Marco et al. *NEJM* (2012) 366:886



http://god-and-logic.blogspot.com.es/2011/04/spinning-ancient-webs-of-deceit-or.html

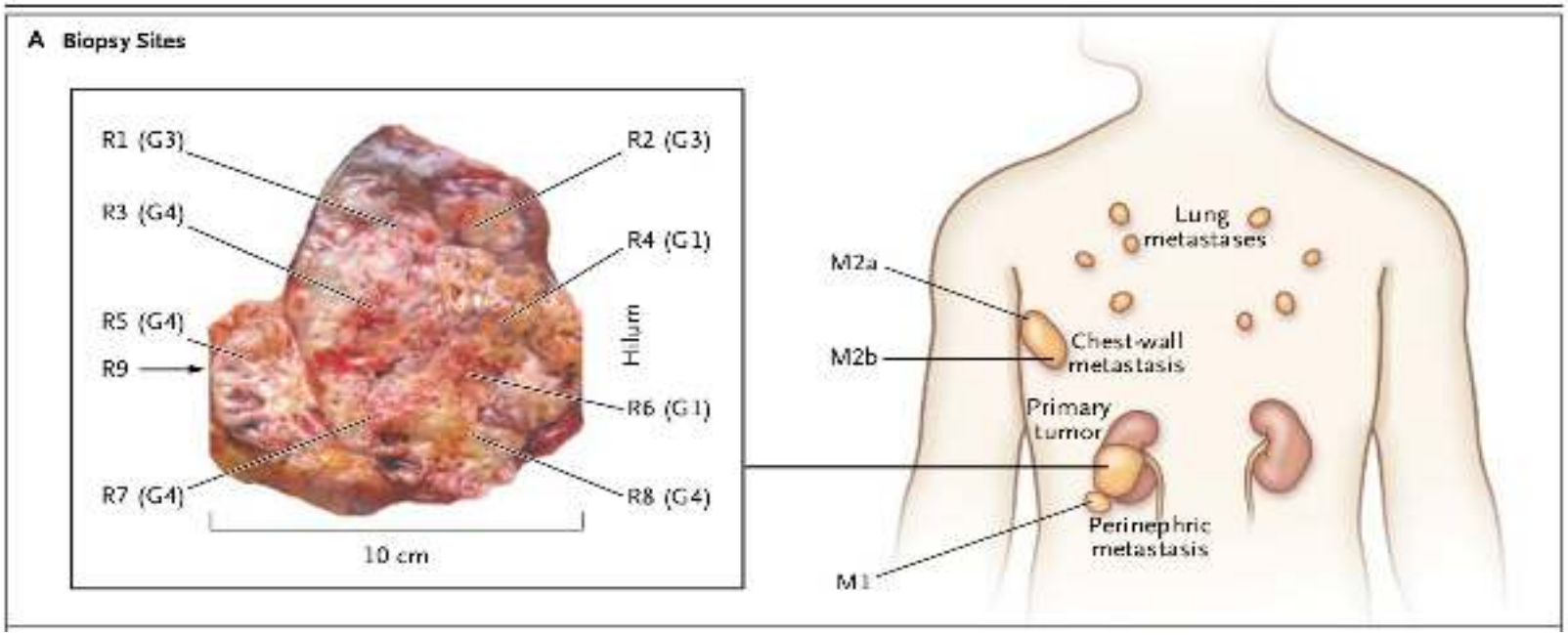


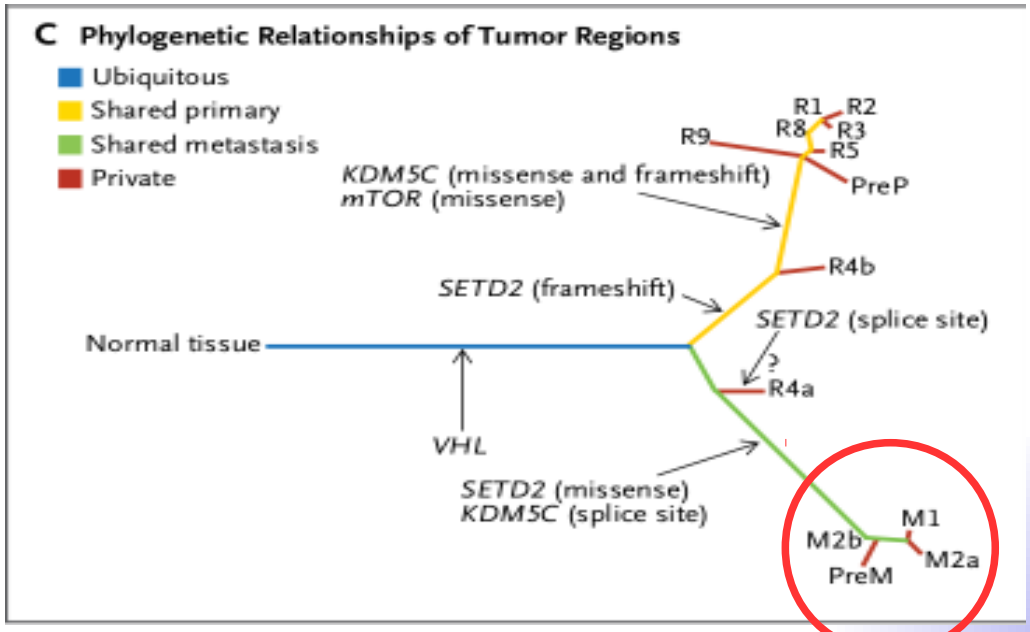
http://www.culturalhistoriesofthematerialworld.com/books/ways-of-making-and-knowing-the-material-culture-of-empirical-knowledge/excerpt/



http://www.biografiasyvidas.com/monografia/darwin/

Metástasis derivadas de colonias concretas.

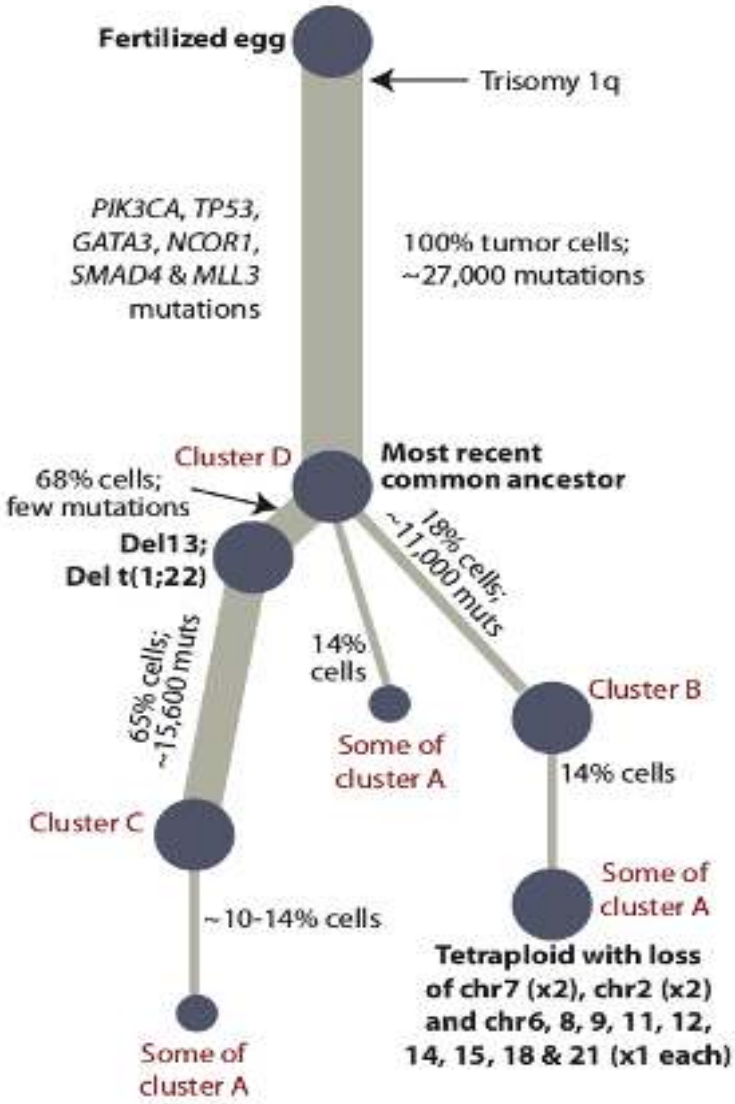




Marco et al. *NEJM* (2012) 366:886

Heterogeneidad en otros tumores

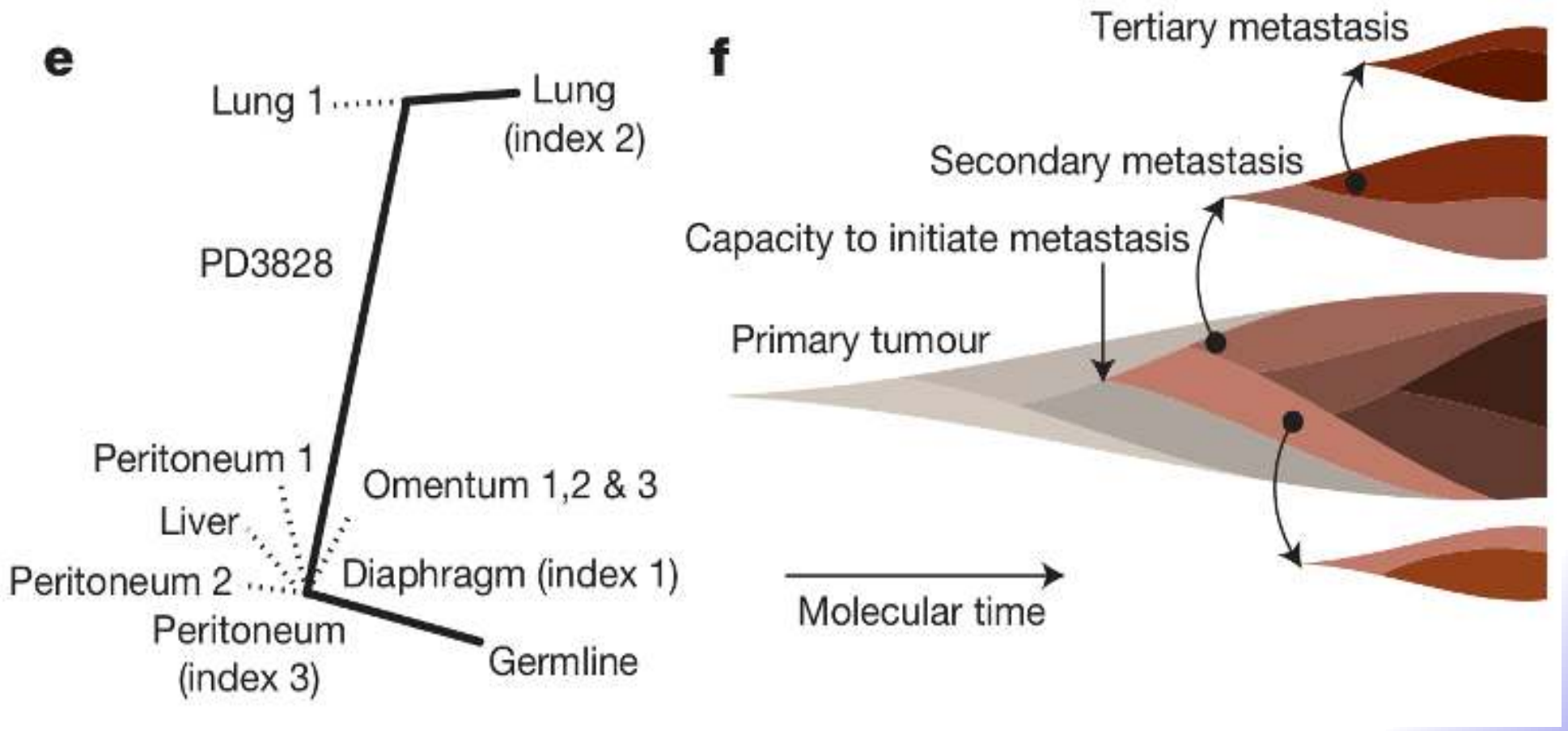
Cáncer de mama



Nik-Zainal et al. *Cell* (2012) 149:994-1007

Heterogeneidad en otros tumores

Cáncer de páncreas

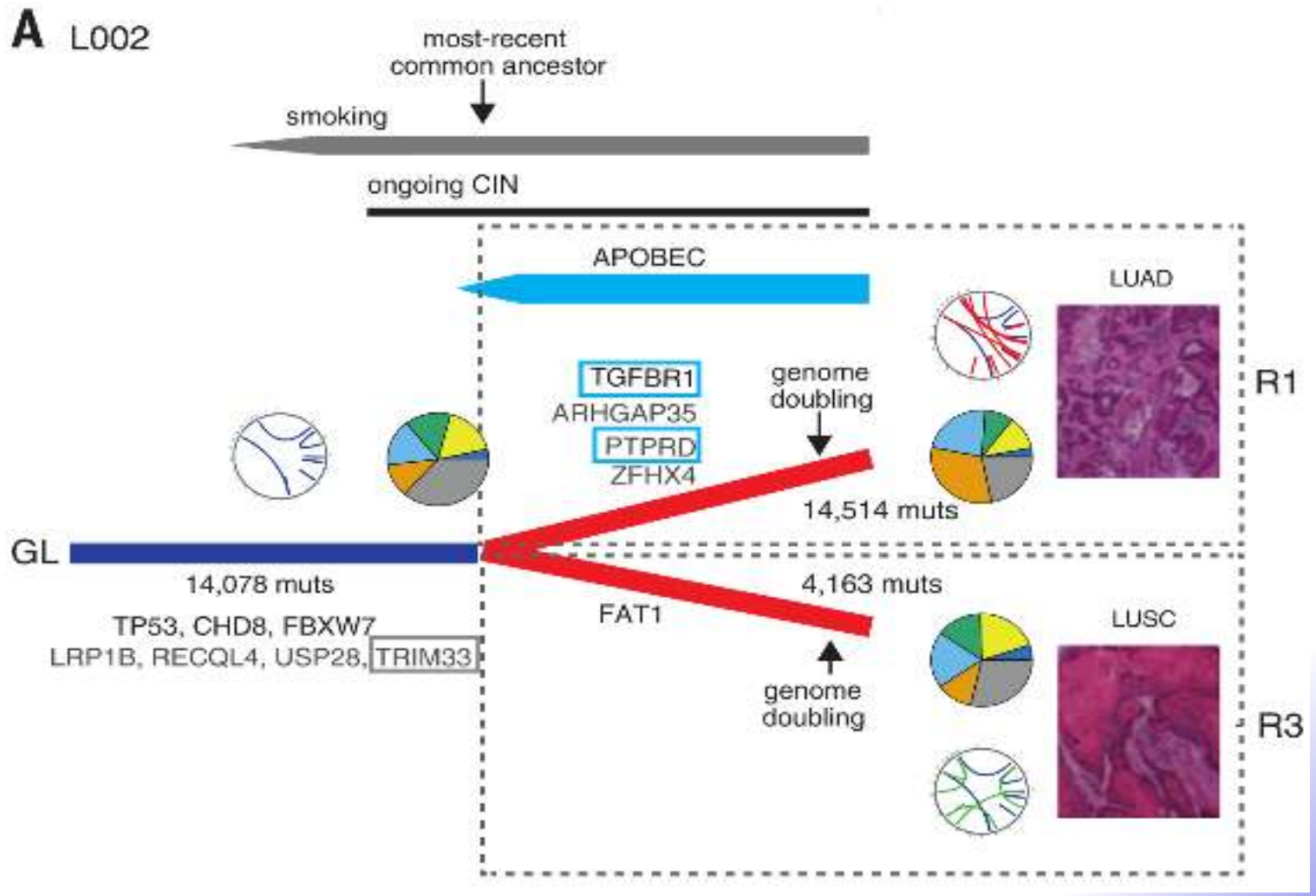


Campbell et al. *Nature* (2010) 467:1109-1113

¡Metástasis generadas a partir de otras metástasis!

Heterogeneidad en otros tumores

Cáncer de pulmón

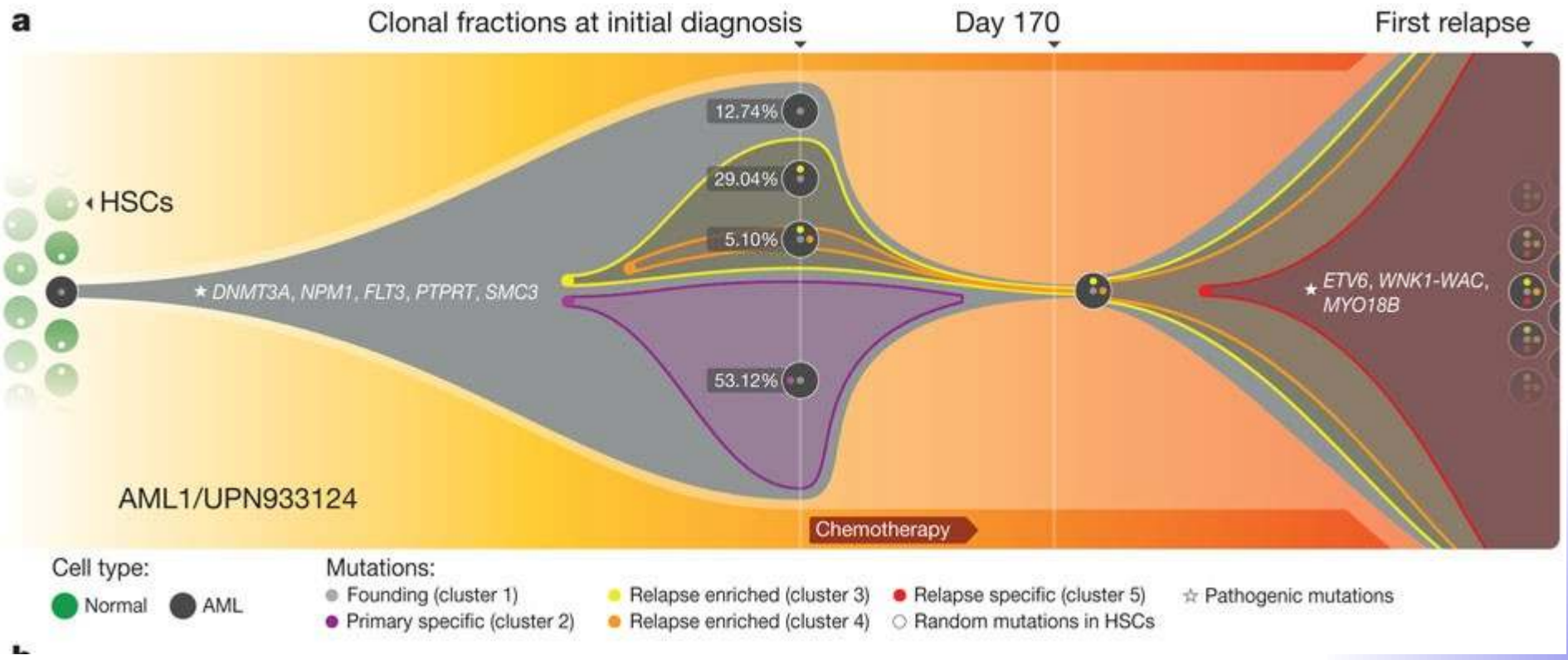


De Bruin et al. *Science* (2014) 346:251-256

¡Distintos subtipos moleculares dentro del mismo tumor!

Heterogeneidad en otros tumores

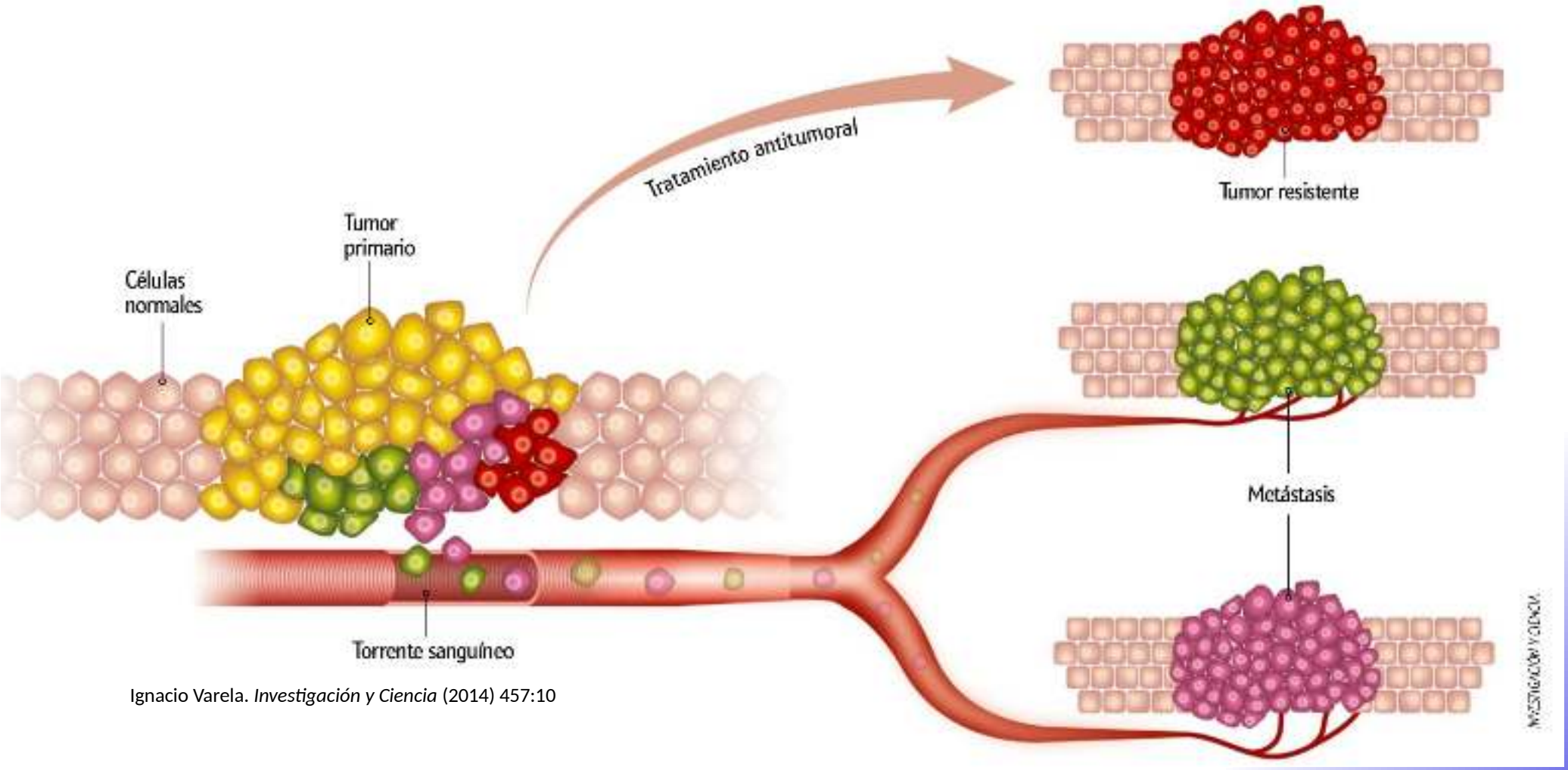
Leucemia mieloide aguda



Dentro del tumor primario ya hay colonias celulares resistentes al tratamiento

Ding et al. *Nature* 481, 506 (2012)

Modelo de evolución tumoral.



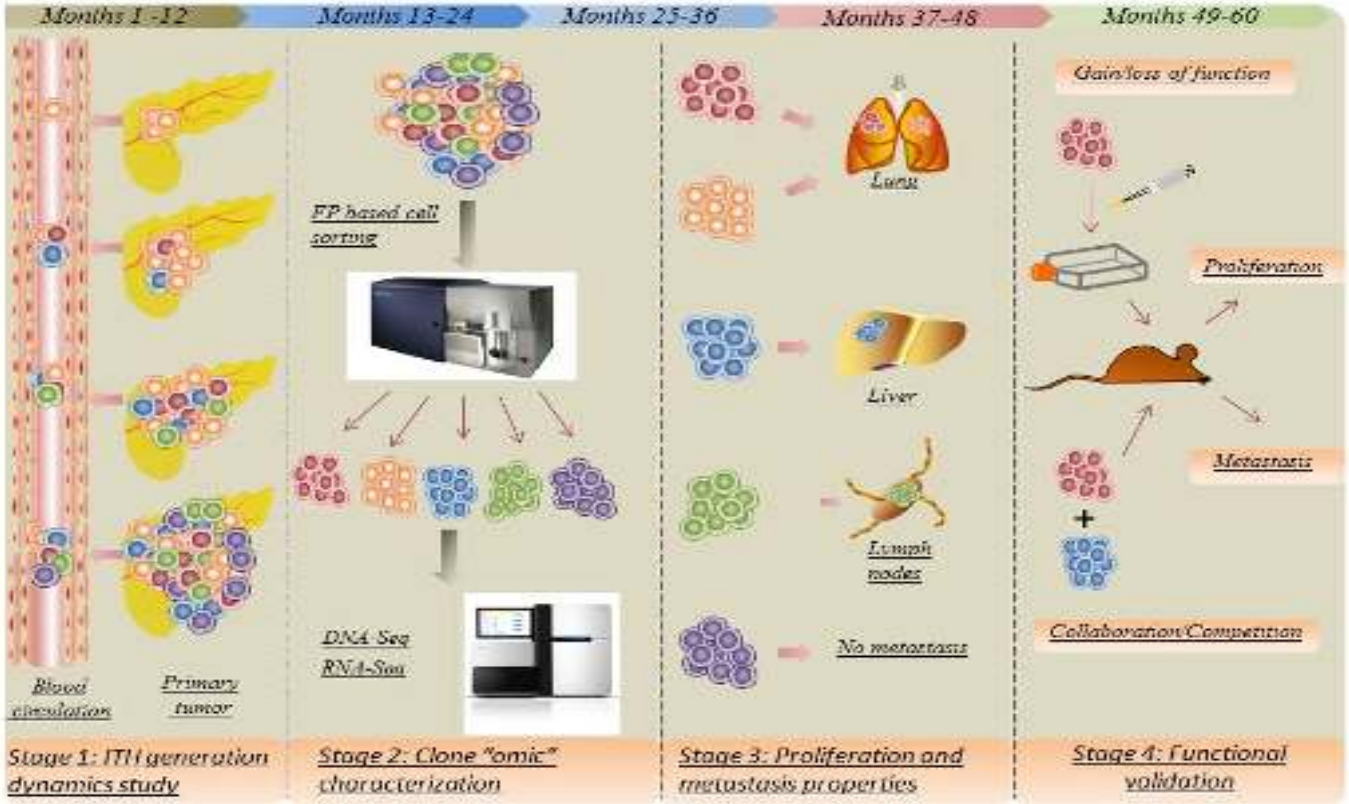
Ignacio Varela. *Investigación y Ciencia* (2014) 457:10

Proyecto en marcha en el IBBTEC

1,5 millones para entender mejor los tumores

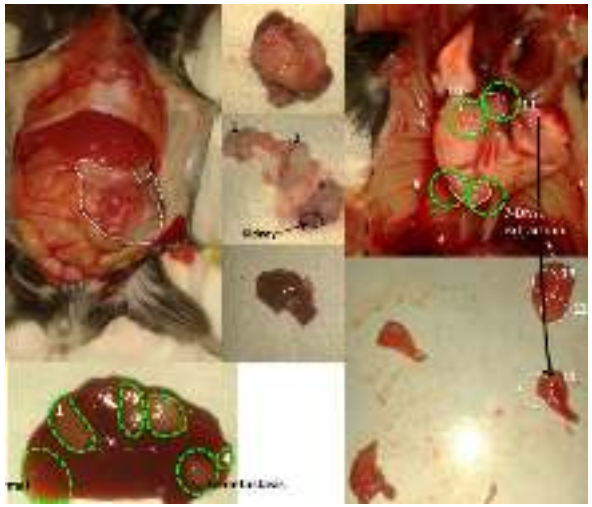
El investigador del IBBTEC Ignacio Varela recibe una de las prestigiosas ERC Starting Grants, primera que consigue la Universidad de Cantabria, para desarrollar un proyecto sobre el cáncer de páncreas

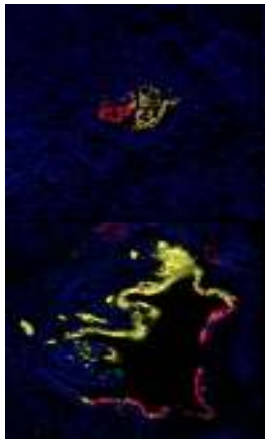
http://web.unican.es/noticias/Paginas/2014/diciembre/1,5-millones-para-entender-mejor-los-tumores.aspx

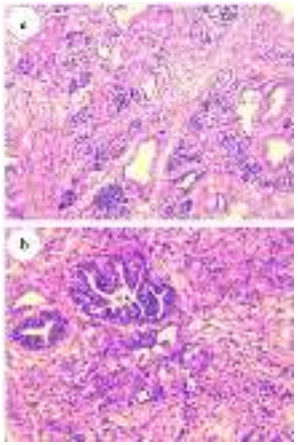


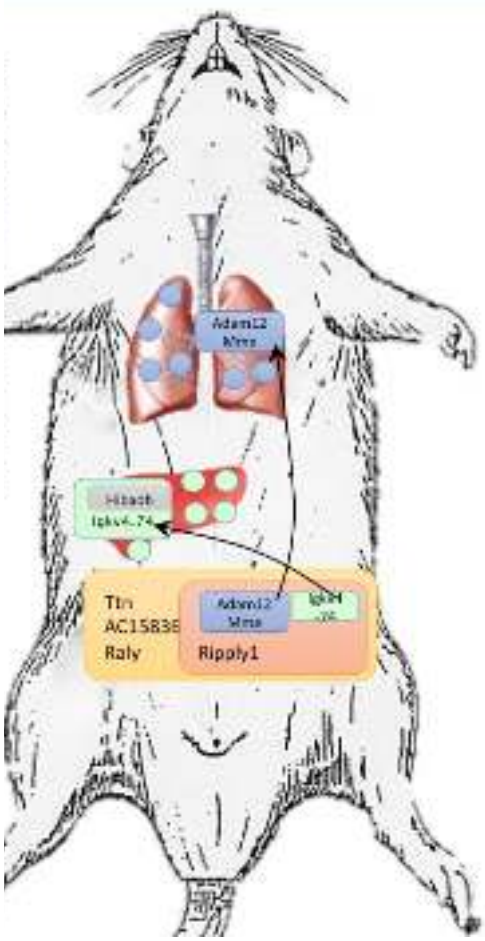
Proyecto en marcha en el IBBTEC

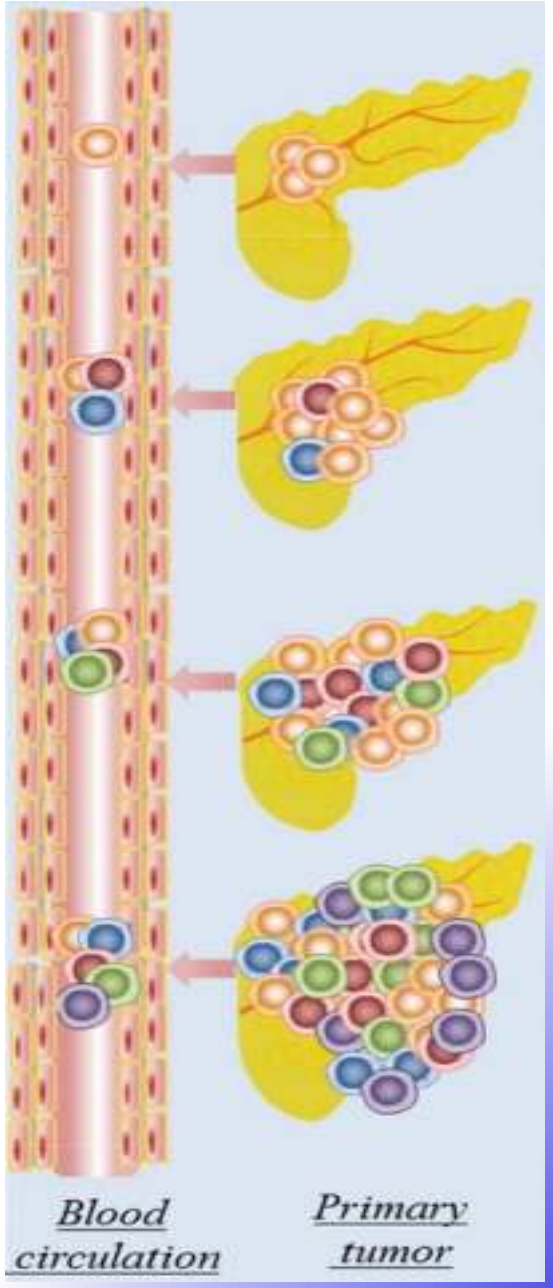
"..Tú ves un ratón, yo una posible cura para el cáncer..."











Bases de datos de interés biológico

-Ensembl: Información sobre localización de genes, transcritos:
https://www.ensembl.org/index.html

-Variant Effect: Herramienta de anotación funcional de mutaciones genómicas:
https://www.ensembl.org/info/docs/tools/index.html

-COSMIC: Bases de datos de mutaciones somáticas encontradas en cáncer:
http://cancer.sanger.ac.uk/cosmic

-ICGC,TCGA: Bases de datos de los proyectos internacionales de caracterización de mutaciones encontradas en cáncer:
https://dcc.icgc.org/
https://cancergenome.nih.gov/

-KEGG Pathway: Base de datos de rutas moleculares descritas
http://www.genome.jp/kegg-bin/show_pathway?map01100

-INTOGEN: Herramienta para detectar genes importantes en cáncer:
https://www.intogen.org/search?gene=BRAF

Análisis de datos biológicos en genómica del cáncer

Ignacio Varela Egocheaga
ignacio.varela@unican.es