Centro Nacional de Análisis Genómico

High content genome analysis in the age of personalized medicine

Ivo Glynne Gut

5.12.2013

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

cnag


81
250
50

History in PEDz ..

☒ Personal profile :

1- Name	2- Age	3- sex
4- address	5- admission through	6- time of admission
7- source of info		

☒ Chief Complaint

☒ Hx of present illness

☒ Systemic Review :

General 1- Wt change 2- Fever 3- Sweating 4- Activity 5- Sleeping 6- General condition 7- appetite	**Eye** 1- redness 2- discharge 3- diplopia **Ear** 1- discharge 2- pain **Neck** 1- swelling 2- limitation move	**CVS** 1- cyanosis 2- rapid breathing 3- excessive sweating during feeding **Infectious** 1- fever 2- chills 3- rigor	**Respiratory** 1- nasal discharge 2- epistaxis 3- cough 4- wheeze 5- cyanosis 6- rapid breathing 7- strider **Mouth** 1- ulceration 2- gum bleeding
CNS 1- abnormal movements 2- headache 3- gait 4- vision 5- hearing 6- level of consciousness 7- neck stiffness 8- diplopia 9- photophobia	**MSS** 1- joints swelling 2- joints pain 3- limitation movements 4- joints stiffness **Skin** 1- rash 2- hyper- , hypo pigmentation 3- hemangioma 4- mongolian spot	**UGS** 1- dysuria 2- urgency 3- enuresis 4- loins pain 5- # wet nappies 6- scrotal swelling 7- frequency	**GI** 1- vomiting 2- diarrhea 3- abd pain 4- constipation 5- color of stool 6- abd distention 7- sticky stool 8- blood in stool 9- melena

☒ Past Medical Hx

1- Previous admission	2- Previous surgeries	3- Hx of blood transfusion
4- Hx of allergy	5- Accidents , injuries	6- Previous investigations

☒ Perinatal Hx

a- Prenatal :

1- Mother's age	2- Maternal fever or rash	3- Maternal diabetes	4- Drugs taking
5- Exposure to radiation	6- Antenatal care	7- Prolonged rupture of membrane > 18-24 h	8- Duration of pregnancy

b- Natal :

1- Delivery mode	2- Crying	3- Birth wt

c- Post-natal :

1- Neonatal admissions	2- Neonatal jaundice

☒ Nutrition Hx

1- Breast feeding (frequency, duration)	2- Bottle feeding (frequency, amount, reason for bottle)	3- Wt gain
4- Weaning	5- Table food (type, amount)	

☒ Developmental Hx

1- Gross motor	2- Fine motor	3- Language
4- Social	5- School performance	

☒ Vaccination Hx

1- Type of vaccine	2- Age of vaccine
3- According JNP, UNRWA	4- complication

☒ Family Hx

1- Mother's age	2- Father's age	3- Consanguinity
4- # children and similar conditions	5- Genetic diseases	6- Early deaths

☒ Social Hx

1- House (rooms, ventilation)	2- Occupation of father	3- Income
4- Insurance	5- Animals	6- Smoking

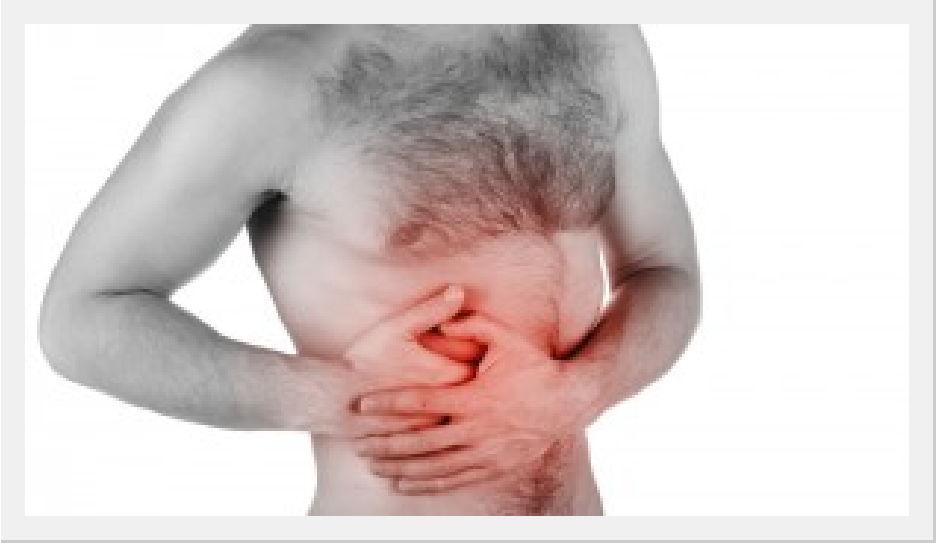
☒ Psychiatric Hx

1- anxiety	2- emotional problem	3- temper	4- school refusal
5- mental problem	6- behavioral problem	7- mood disorders	

Pediatrics SAWA

Symptoms of Early Stages of Colon Cancer

Unfortunately, during the early stages of colon cancer there are very few symptoms. When symptoms do appear, they are often ignored until they are intolerable, and at this point the colon cancer tends to be more advanced and more difficult to treat. When a polyp first forms in the colon, there are no symptoms. By the time symptoms do appear, the cancer has already advanced somewhat so it is best to see a doctor as soon as possible. Symptoms can include stomach or abdominal pain, dark or bloody stool, difficulty completing a bowel movement, general tiredness, an sudden weight loss. Although these symptoms may seem vague and can be caused by a number of medical issues, it is best not to ignore them, especially if they are prolonged.



Screening During Early Stages of Colon Cancer

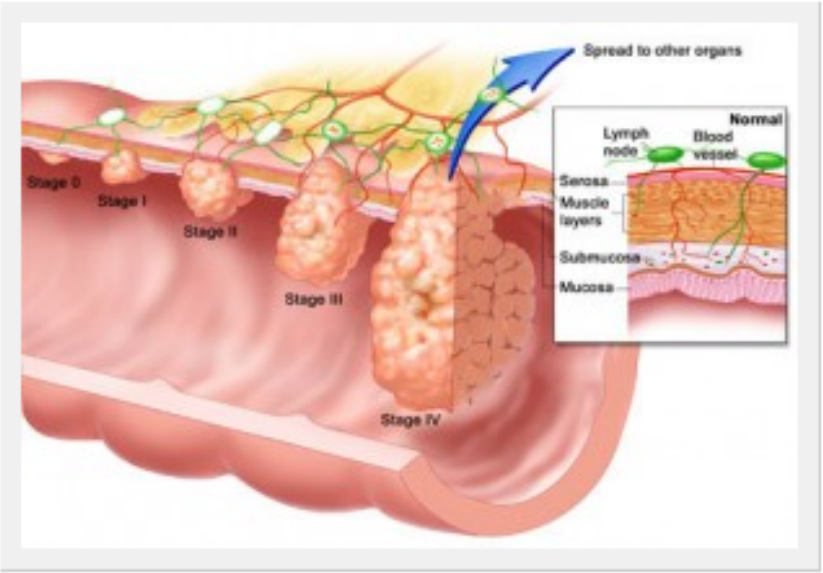
Series of tests

- Ultrasound
- Colonoscopy
- Biopsies
- Blood tests (10s of measurements)
- (DNA tests – 3 markers)
- ....

- Establish diagnosis

Colon Cancer Stages Survival Rates

After diagnosis, some patients may wish to know the historical data concerning the survivability of their disease. Since the main factor affecting the prognosis for colon cancer is stage, it is most helpful to view the statistics for each stage separately, rather than lumping all the cases of colon cancer together. The colon cancer stages survival rates are sourced from the 2010 staging manual of the American Joint Committee on Cancer.



Stage	Survival Rate
I	74%
IIA	67%
IIB	59%
IIC	37%
IIIA	73%
IIIB	46%
IIIC	28%
IV	6%

Treatment

- Operation
- Chemotherapy
- Radiation
- ....

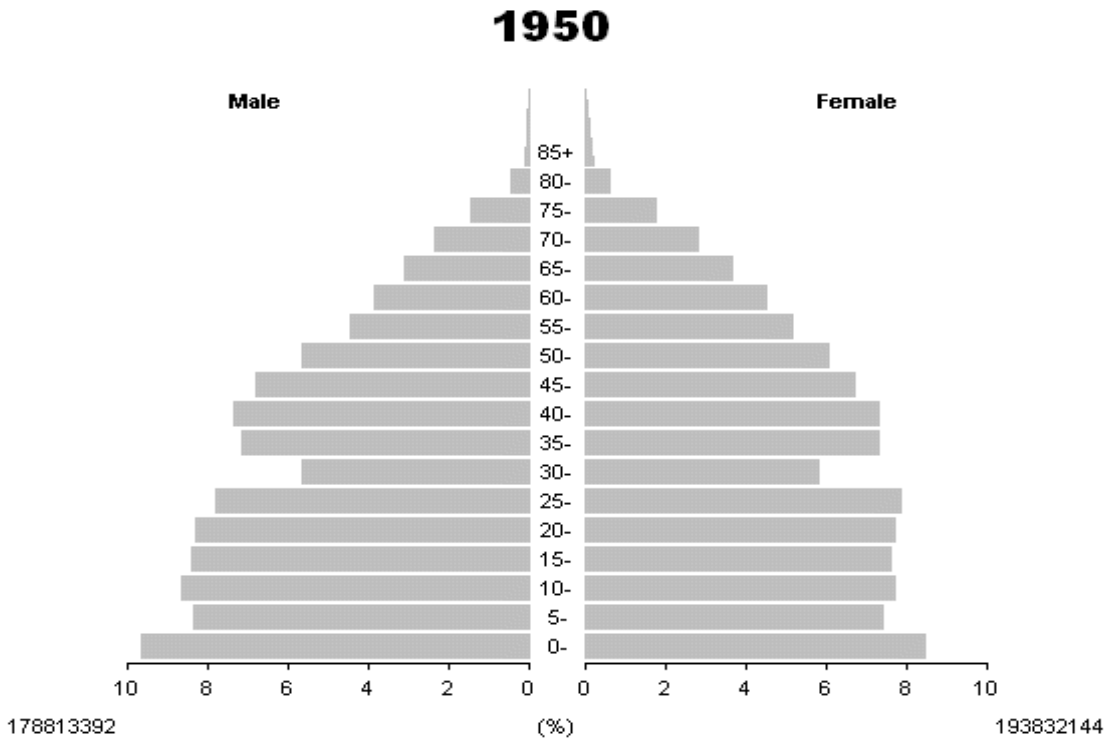
Basis for diagnosis/treatment decision is data-poor

- Complexity of cancer is high
- Number of measurements is low

Needs

- **World wide ~12 million new cancer cases/year**
- **Cure rates for most common forms of cancer have hardly changed over the last decades**
- **Even the most advanced targeted therapies are typically only effective for a small fraction of the patients**
- **Pharma development costs have dramatically increased, while the number of new drugs keeps dropping**

Age-Structure of the European Union (27 Member States) 1950-2050



Source: United Nations, World Population Prospects, The 2006 Revision

World population 2011 – 7 billion, prediction 2050 – 9 billion

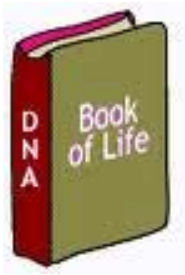
Personalized Medicine

Right treatment for the right patient

Precision Medicine

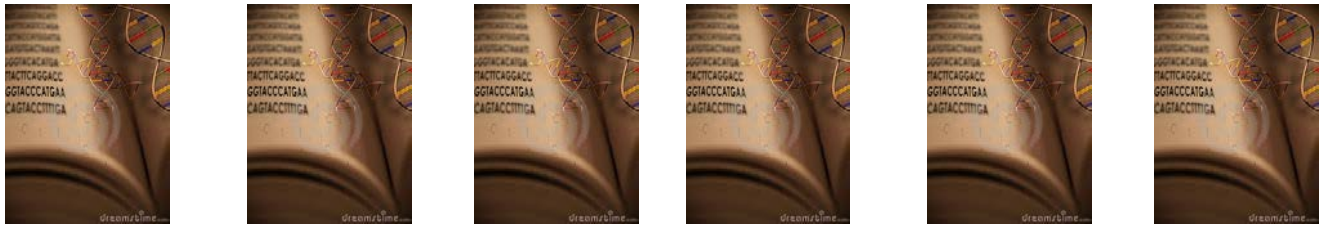
Innovation in Genomics

GENOME is the genetic information of a living organism coded in the **DNA**, a long, thin, helical thread



GENOME

1 **2** **3** **23 chapters**



CHROMOSOMES

48-250M letters **48-250M letters** **48-250M letters**

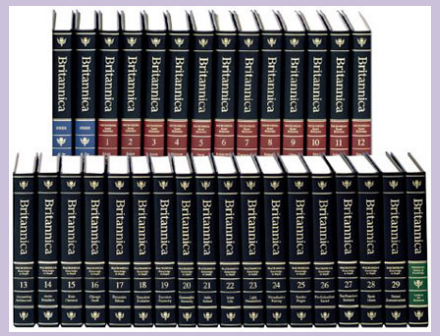
TGCTACGAT...
TGCTACGAT...
TGCTACGAT...

NUCLEOTIDES

The sequence of these letters (nucleotides) carries information

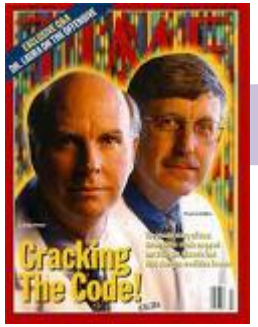
The book (Human Genome) contains over 3.2 billion letters (ACGT) total

= 11 x



Innovation in Genomics

1989



2003

2010

Human Genome Project

- First complete sequence of a Human Genome
- 13-year project
- Cost: roughly $ 3 billion
- Hundreds of researchers involved
- US, UK Japan, France, Germany, China, India

1.000 Genomes Project Consortium

- 1.000 human genomes sequenced
- 15-month project (2 human gen x day)
- Cost: $50M
- US, UK and China

- Ten years later, we can study routinely our individual genomes thanks to the **new generation sequencing technologies**.

- **Genomics research** is leading innovation in medicine, energy, environment or agriculture, bringing improvements in people's health and quality of life.

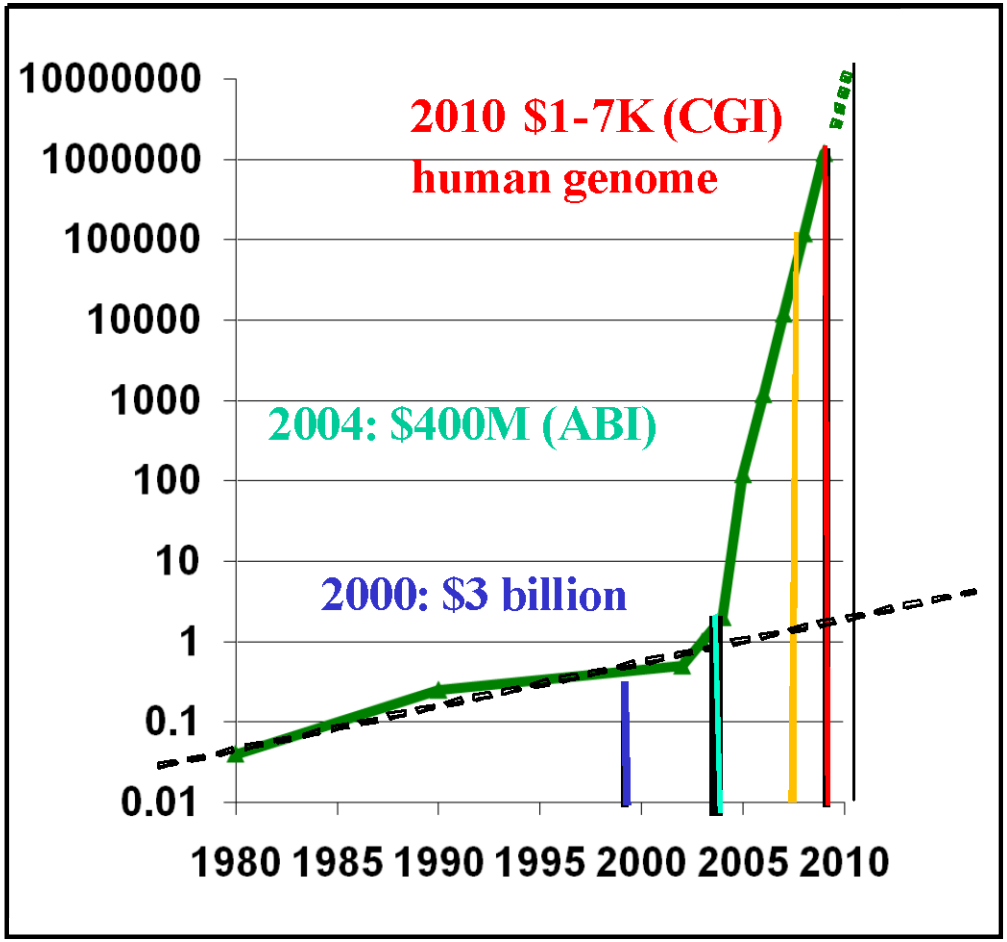
Life is the translation of the information in the genome
into the phenotype of the organism:

The organism ‚computes' this phenotype from its genotype,
given a specific environment

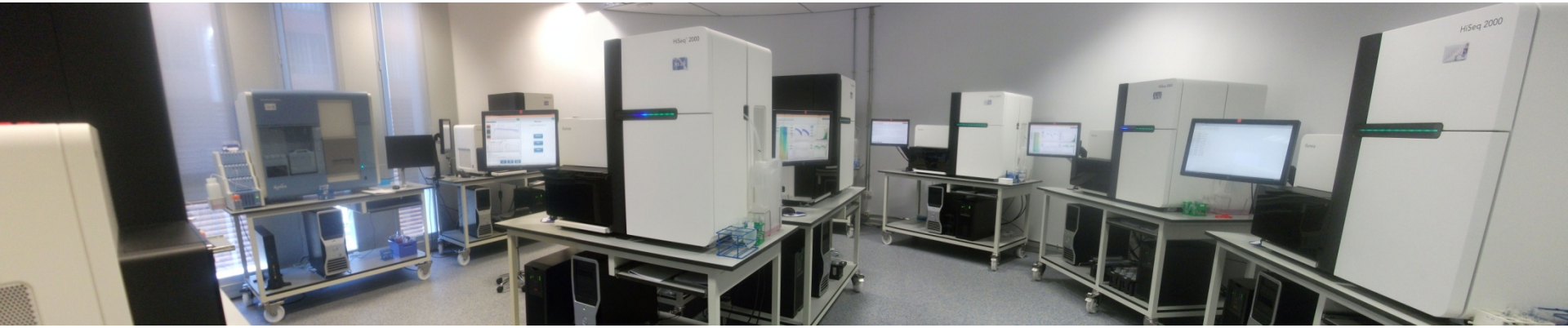
The omics Era and the sequencing hockey stick

Cost of one human genome sequence is 1/1000000 of initial sequence

Moore's law – doubling of density of transistors on a computer ship is 1.5 years

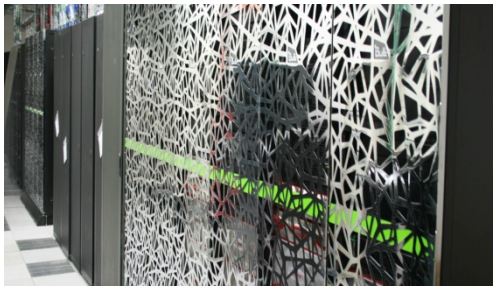


The genomehenge



Sequencing capacity

- >1 Tbase/day = 8-9 human genomes per day at 30x coverage



Equipment

- 11 Illumina HiSeq2000/2500
- 1 Illumina MiSeq
- 4 Illumina cBots
- Caliper liquid handling robotics

- Bull 950 core cluster super computer
- Maxeler Data Flow Engine
- 2.2 petabyte disc space
- Barcelona Supercomputing Center (10 x 10 Gb/s)

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

cnag

1 Terabase = 1.000.000.000.000.000 bases

Workflow

BIOREPOSITORY

LABORATORY

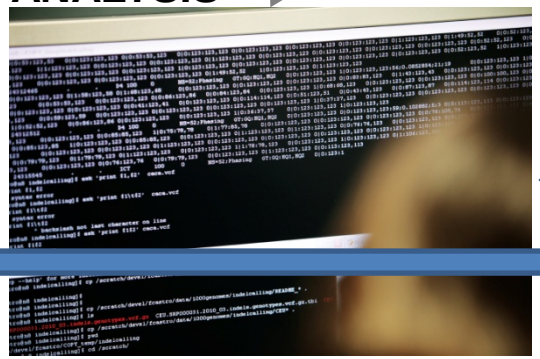
SEQUENCING

LIMS

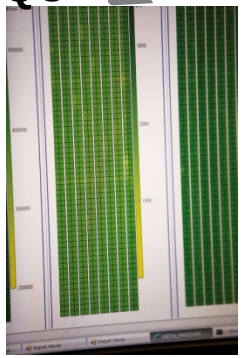
Subproject	Status	Deadline (Countdown)	Details by Sample	Details by Library	Analysis	Sample Type	Application	Expected Capture Protocol	Numbers of Samples	Ready for Sequence	Samples	Sample Reception Dates	Aliquots	Libraries pass / total	Libraries Barcodes	Samples with a pass Library	Flowcells & Dates	Samples above Coverage	Samples above pfmr	Data Transfed by Sample
_11	Open (Experimental)	2012-10-01 (26 days)	Details by Sample; Sample Aggregate View	Details by Library; Library Aggregate View	BAG	gDNA	ExomeCapture-Seq	Agilent Human All Exon 51Mb v4	6 8 8 6 4	yes	E061 E058 E055 E059 E056 E060 E057 E054	2012-07-02 2012-07-02	861C 948C 862C 863C 864C	3 / 4	861C 862C 863C 864C	3				0 sent, of 4 to be seq
_01	Open (Experimental)	2012-10-02 (27 days)	Details by Sample; Sample Aggregate View	Details by Library; Library Aggregate View	BAG	gDNA	WG-Seq		2 4 4 4 2	yes	H367 H368 H368 H367	2012-07-03 2012-07-03	580C 581C	2 / 2	580C 581C	4	C0YHRACXX - 2012-07-27 C11CBACXX - 2012-08-09 C11D3ACXX - 2012-08-09			0 sent, of 2 to be seq
_03	Open (Experimental)	2012-10-03 (28 days)	Details by Sample; Sample Aggregate View	Details by Library; Library Aggregate View	No	Exome library	ExomeCapture-Seq	Nimblegen SeqCap 64 Mb v3	16 80 80 0 80	yes	H349 H365 H346 H362 H343 H359 H340 H356 H337	2012-06-18 2012-07-04		80 / 80	I231 I229 I230 I227 I228 H342 H344 H346 H347	80	C0TRHACXX - 2012-06-19 C0TVGACXX - 2012-06-26 D113HACXX - 2012-06-26 D14PPACXX - 2012-07-25 C0YHRACXX - 2012-07-27	80		80 sent, of 80 to be seq

TRANSFER

ANALYSIS



QC



Informatics Resources

Production Bioinformatics

- Primary run analysis and verification
- QC systems and LIMS

Analysis Production

- Data analysis and interpretation

Statistics

- Alignment and variant calling with high performance CNAG pipeline
- Proprietary GEM/BFAST alignment and SNAPE variant calling

Annotation

- Proprietary pipeline for genome annotation
- Annotation against genome databases with Ensembl API (mirror)

Algorithm Development

- Development and improvement of alignment and assembly methods

Functional Bioinformatics

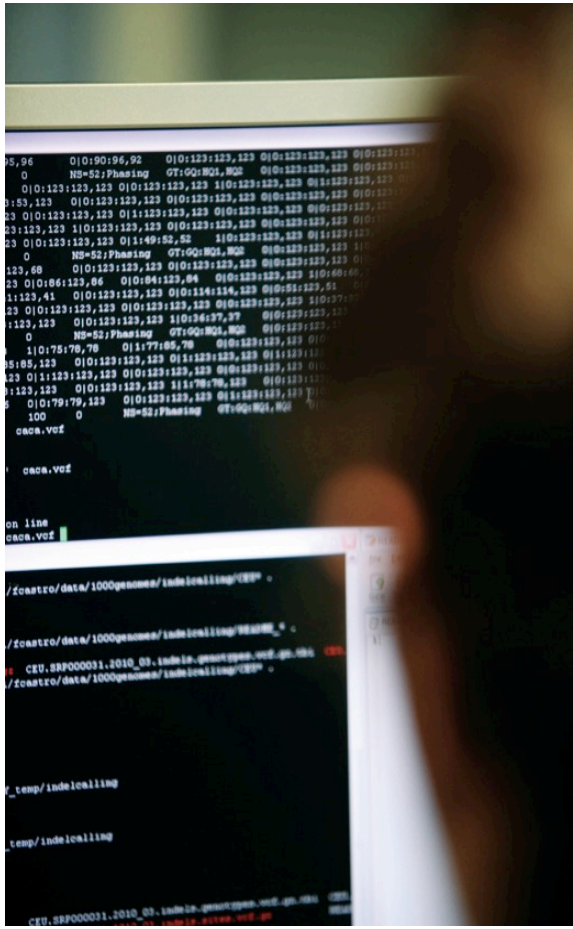
- Establish models of functional effects of variants
- Establishment of networks

Genome Biology - Structural Genomics

- 3-d structure of genomes

Databases

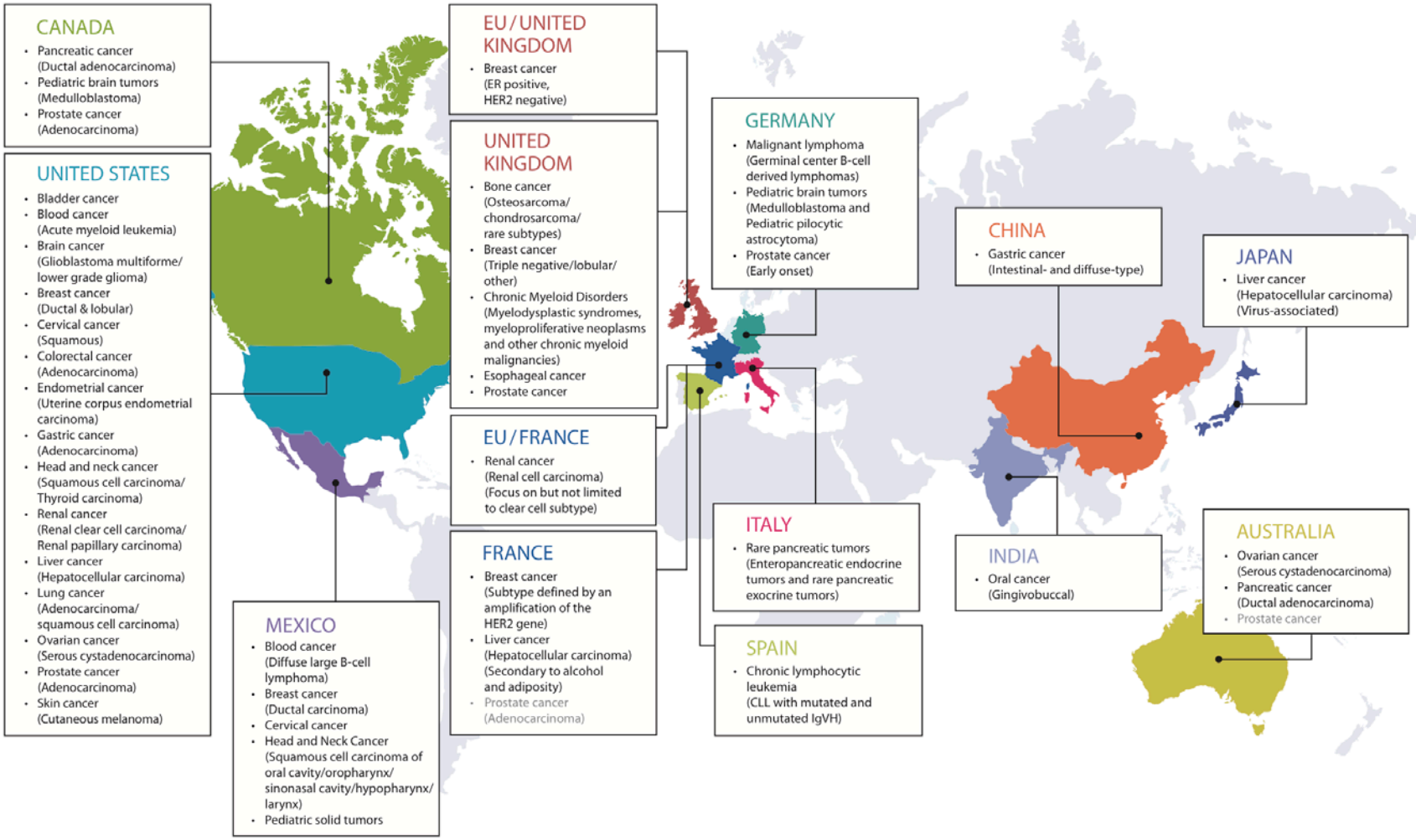
- Storage and distribution of data in collaboration with the BSC



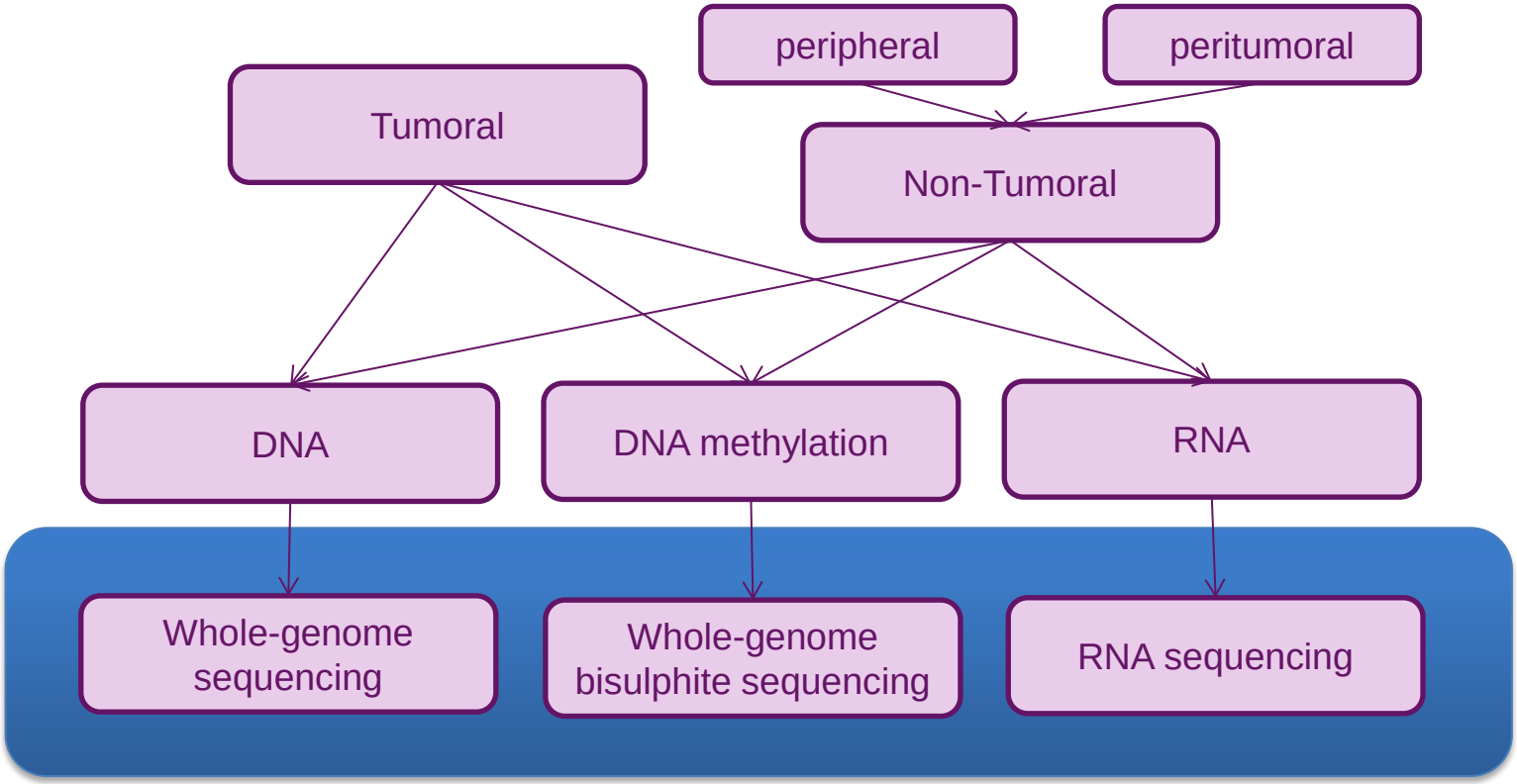
What we do?

International Cancer Genome Consortium

52 Projects/20 Countries



Sampling structure



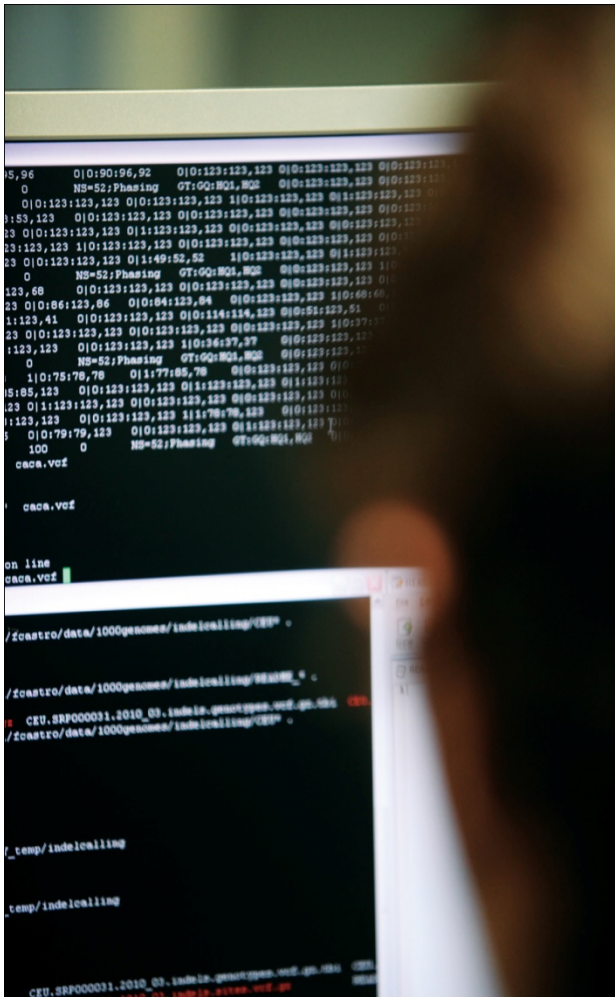
Compute Elements of Sequencing

- Human genome sequence 3.000.000.000 bases
    - 30x coverage – 100.000.000.000 bases – 1.000.000.000 reads

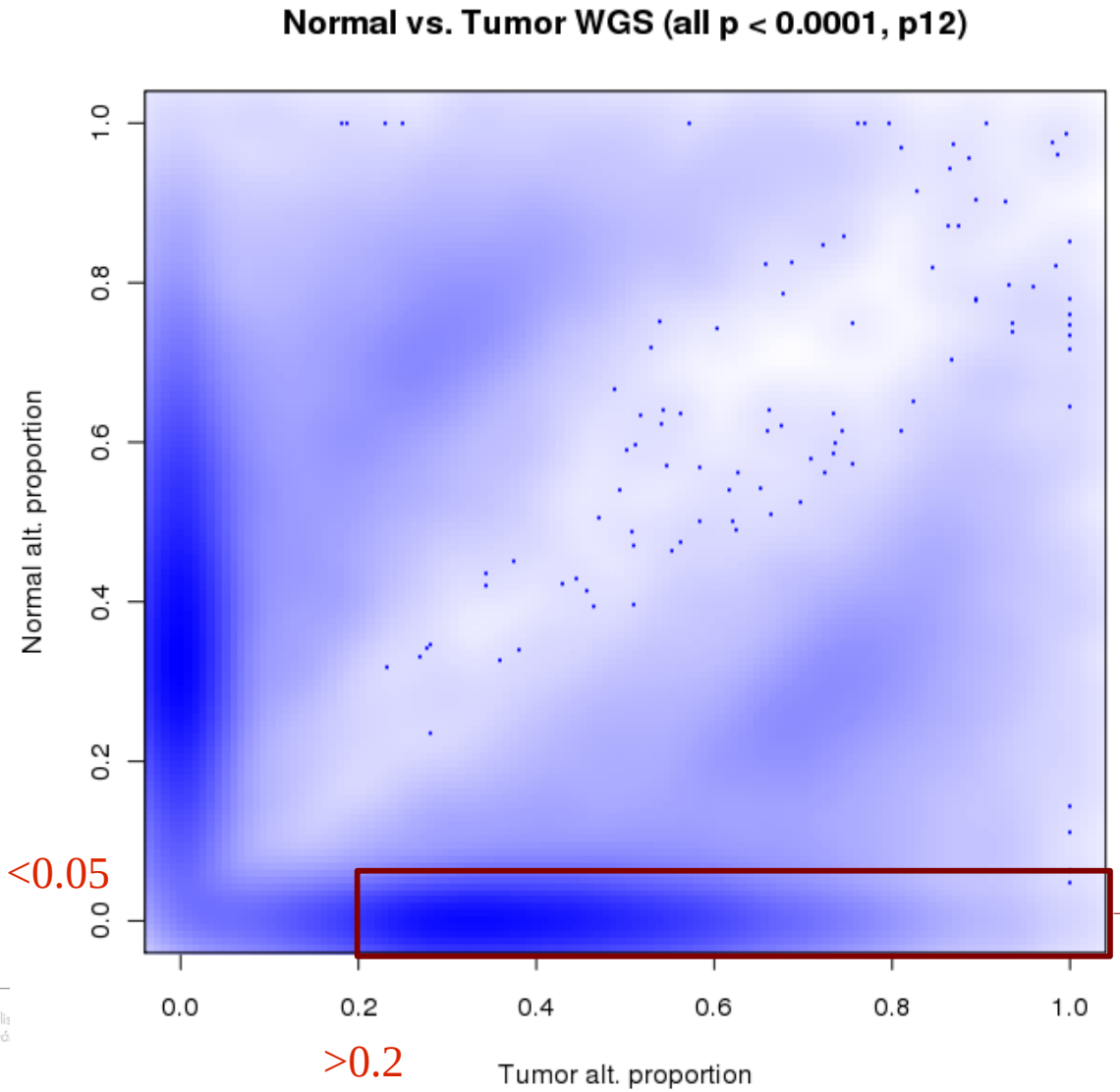
- Base calling (Illumina)
- Alignment to reference
- Variant calling
- Joining data for interpretation of study
- Presentation of the data
- Verification of results
- Interpretation of results

Output - CLL

	P12 WGS			P12 EXOME		
	Normal	Tumor	Somatic *	Normal	Tumor	Somatic *
All variants	3650523	3686288	3900	86108	83483	104
SNV	3140618	3166245	3719	73965	71833	102
Short Indels	509905	520043	181	12143	11650	2
Entry in dbSNP	3102387	3126985	201	71713	69542	1
No entry in dbSNP	548136	559303	3699	14395	13941	103
N. Homoz. Alternative	1169685	1177930	14	27509	26768	1
N. Herez. Ref./Altern.	1970933	1988315	3705	46456	45065	101
Promoter	684739	699156	566	42109	39394	50
Exon	48731	50754	67	51016	48632	62
Intron	4802570	4866290	5258	207535	201268	283
3PRIME_UTR	52865	53943	62	9661	9218	18
5PRIME_UTR	6410	6718	4	3914	3539	1
COMPLEX_INDEL	62	44		29	55	
DOWNSTREAM	712934	726972	656	58475	55576	101
ESSENTIAL_SPLICE_SITE	316	311	2	183	196	
FRAMESHIFT_CODING	1538	1793		849	730	
INTERGENIC	1693060	1703480	1854	2030	2174	1
INTRONIC	4802570	4866290	5258	207535	201268	283
NMD_TRANSCRIPT	290819	294122	292	16767	16253	31
NON_SYNONYMOUS_CODING	19553	20260	58	21106	20288	57
PARTIAL_CODON	4	2		2	2	
REGULATORY_REGION	7069	7303	5	1239	1147	
SPLICE_SITE	6621	6813	1	7034	6891	1
STOP_GAINED	220	199		157	133	
STOP_LOST	49	47		28	28	
SYNONYMOUS_CODING	20517	21336	3	24719	23606	4
UPSTREAM	677670	691853	561	40870	38247	50
WITHIN_MATURE_miRNA	38	34	1	9	11	
WITHIN_NON_CODING_GENE	1336226	1350359	1428	65979	63801	108



* p < 0.0001 (Fisher exact test) + in-house post-filtering



Somatic mutations, > heteroz. in tumor, homoz. in healthy

CONSORTIUM FOR THE STUDY OF THE CLL GENOME
CHRONIC LYMPHOCYTIC LEUKEMIA GENOME PROJECT

centre nacional d'anàlisi
centro nacional de aná

cnag

Variant analysis example: stepwise filtering

VCF

Filtering parameters

Pedigree

Multiple samples

Somatic detection

Inheritance model

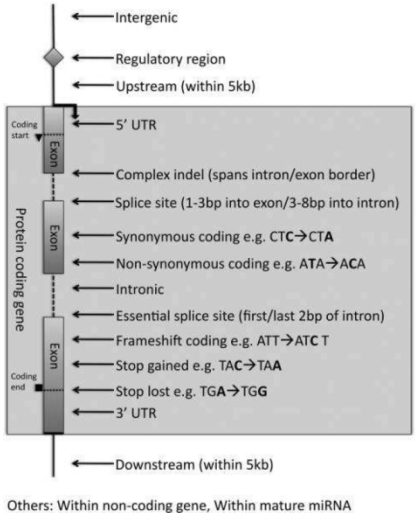
Shared vs non-shared

Fisher test allele counts

dbSNP

dbCNAG

Functional annotation



Others: Within non-coding gene, Within mature miRNA

Variant analysis example: stepwise filtering

VCF

Filtering parameters

Pedigree

Multiple samples

Somatic detection

Inheritance model

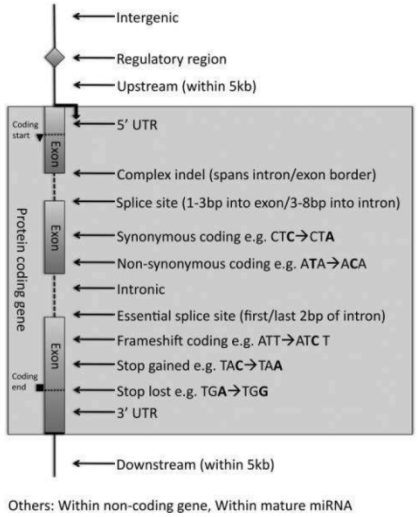
Shared vs non-shared

Fisher test allele counts

dbSNP

dbCNAG

Functional annotation

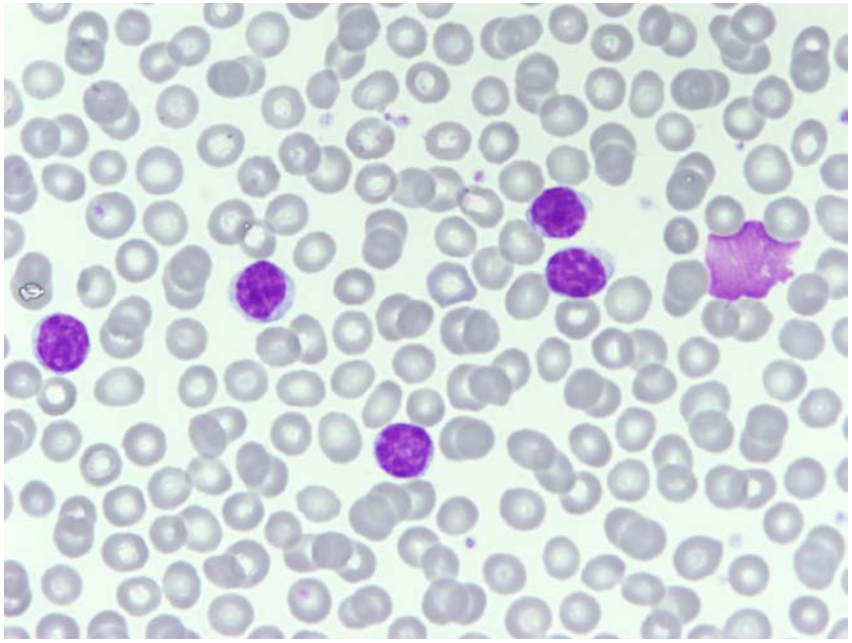


Others: Within non-coding gene, Within mature miRNA

CONSORTIUM FOR THE STUDY OF THE CLL GENOME
CHRONIC LYMPHOCYTIC LEUKEMIA GENOME PROJECT

International Cancer Genome Consortium

The Genomic Study of Chronic Lymphocytic Leukaemia



The International Cancer Genome Consortium is an **international effort to establish 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.

Spain's contribution to the ICGC is on chronic lymphocytic leukaemia (CLL).

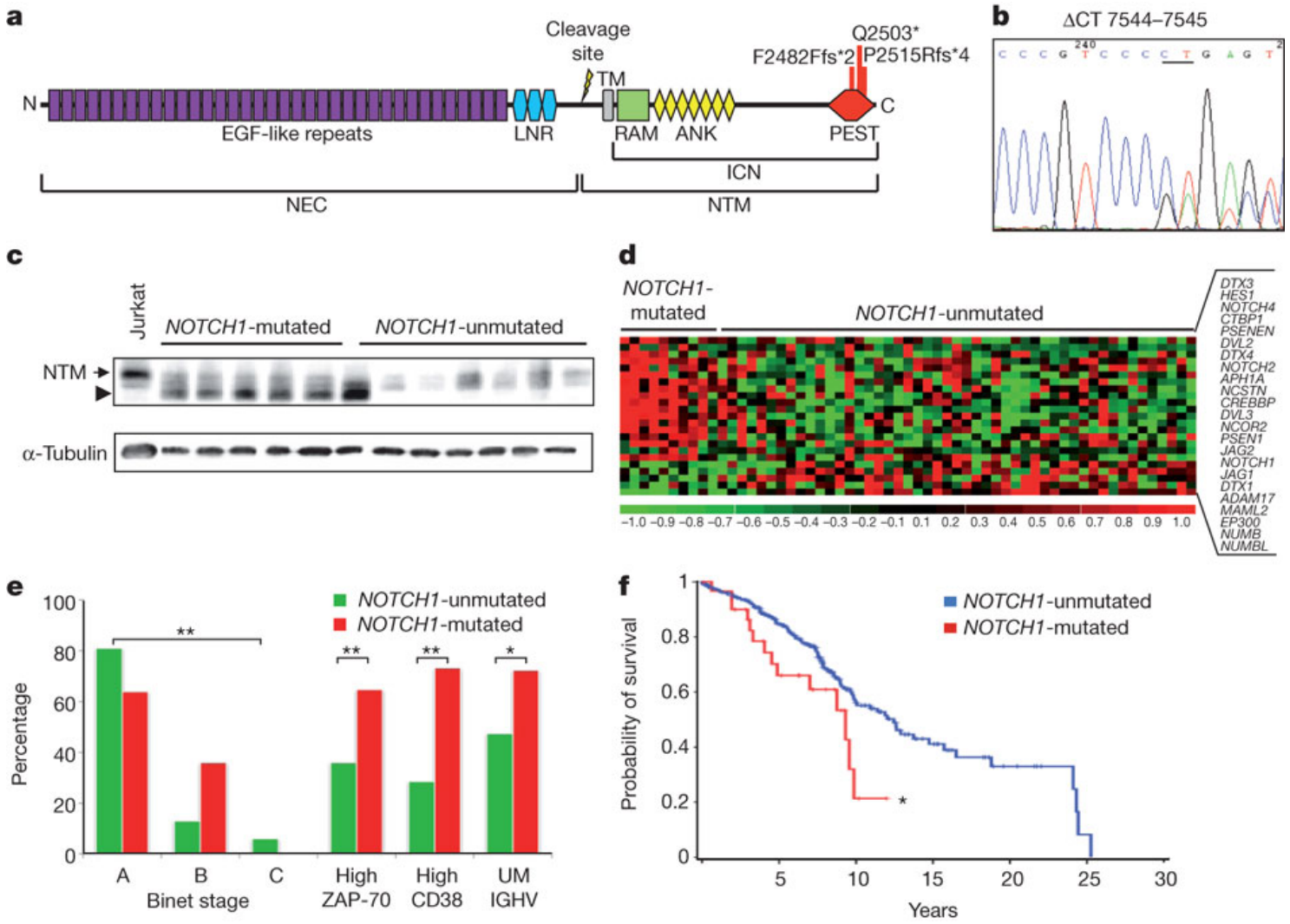
FUNDING SOURCE: Ministerio de Ciencia e Innovación

COORDINATOR: Prof. Elias Campo, Hospital Clínic Barcelona

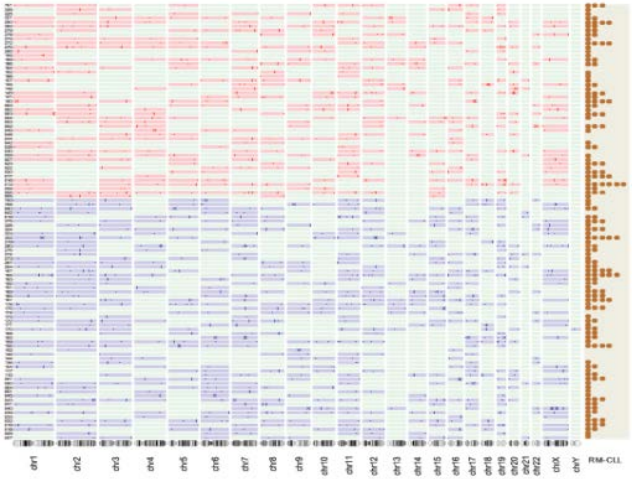
DURATION: May 2009 – May 2014

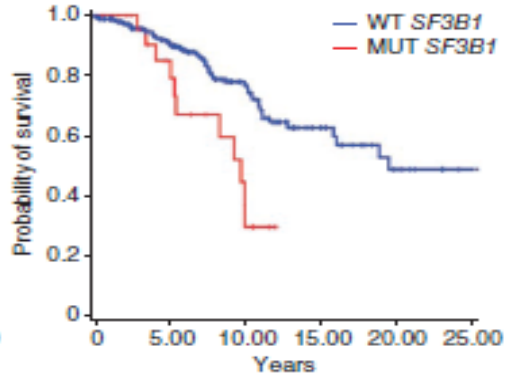
PROJECT FUNDING: €10 million

Mutational and functional analysis of *NOTCH1* in CLL



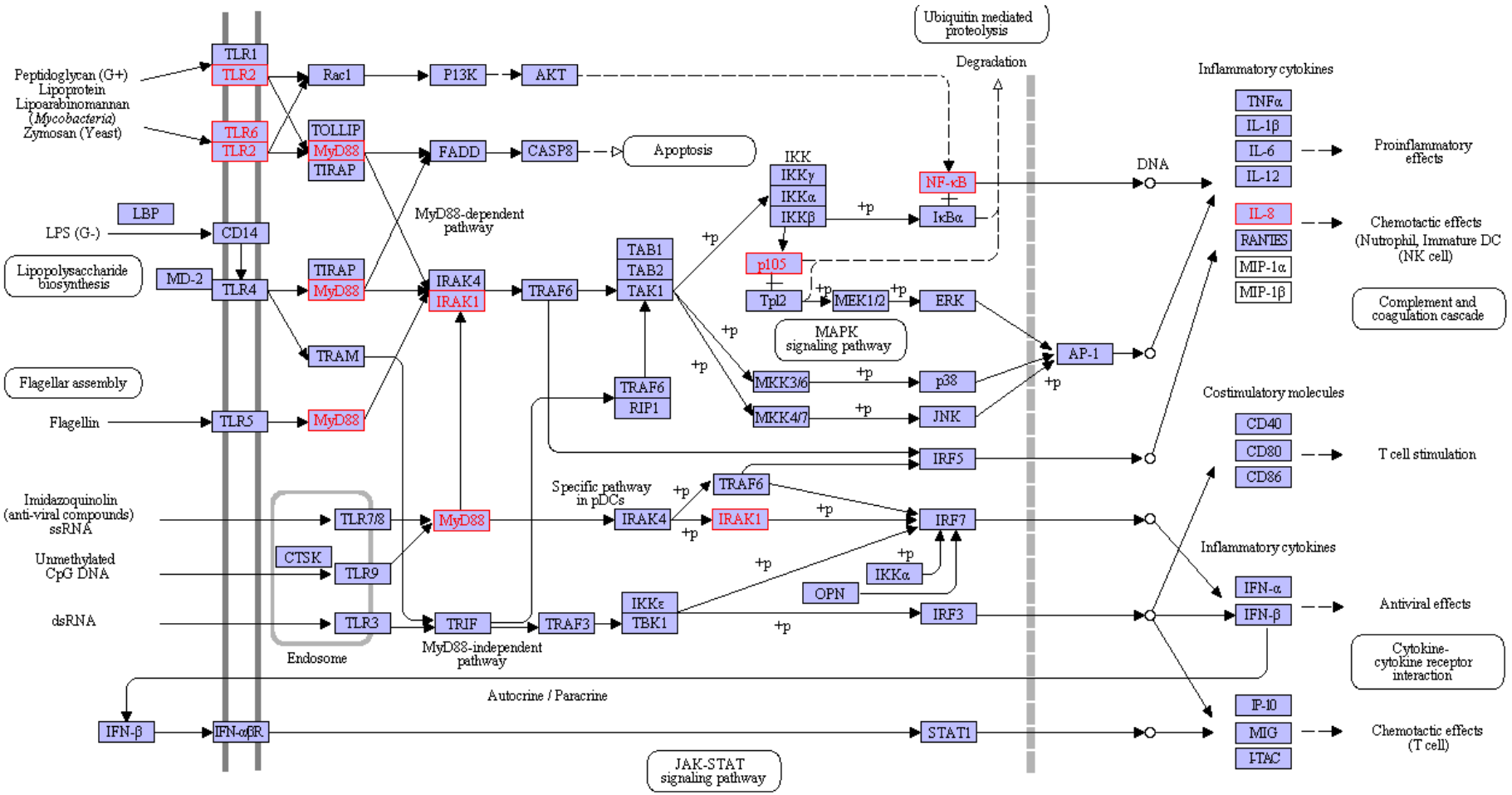
XS Puente *et al. Nature* **000**, 1-5 (2011) doi:10.1038/nature10113





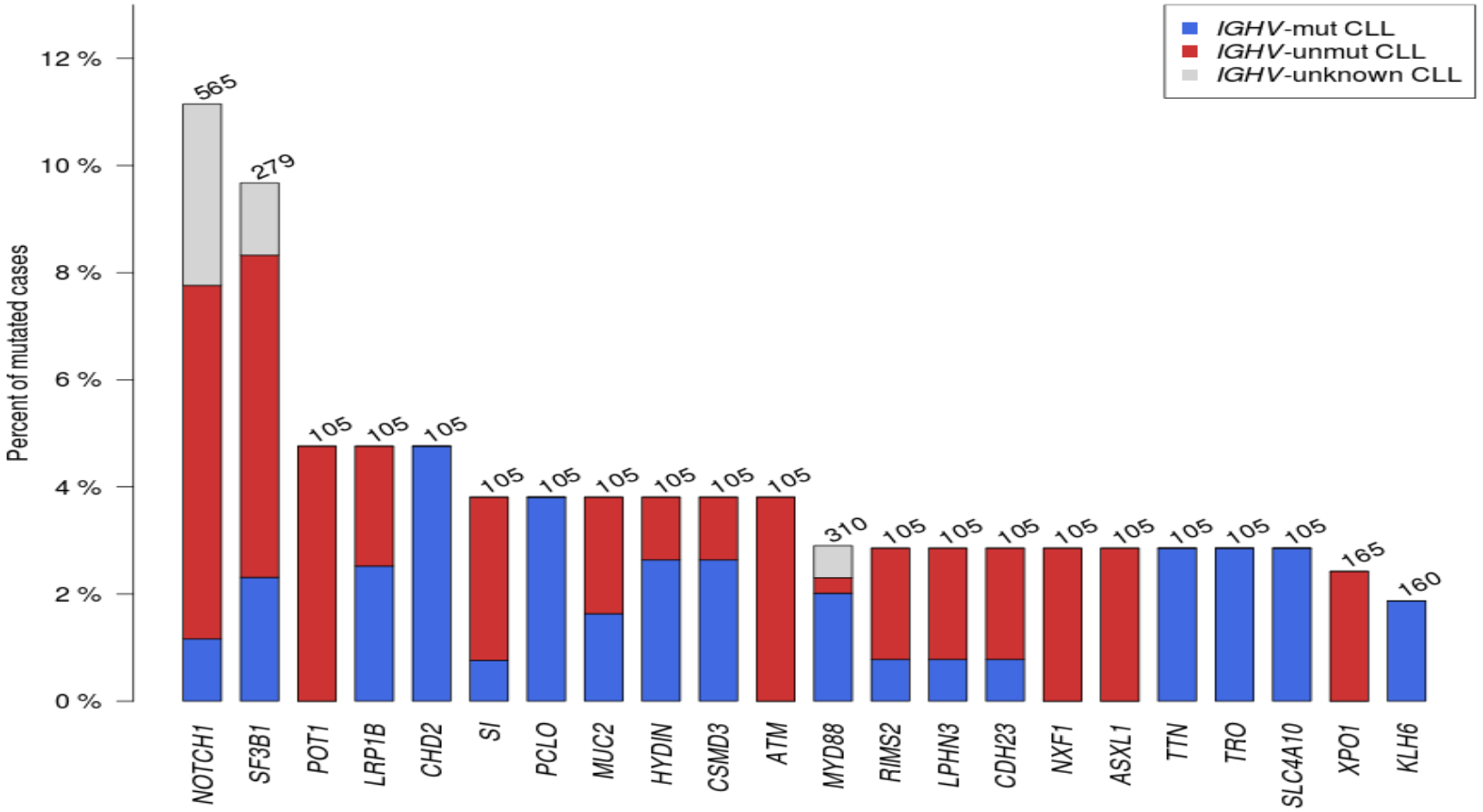
- **Sequencing of 105 CLL tumours**
- **Identification of 1246 somatic mutations in 1.100 genes**
- **78 genes with mutations in more than one patient**
- **Mutated genes cluster in pathways**
- **Distinct mutations are associated with clinical classifications**
- **Mutations in the SF3B1 gene identifies a group of patients with more aggressive disease**

- **SF3B1 encodes a subunit of the spliceosomal U2 small nuclear ribonucleoprotein (snRNP)**



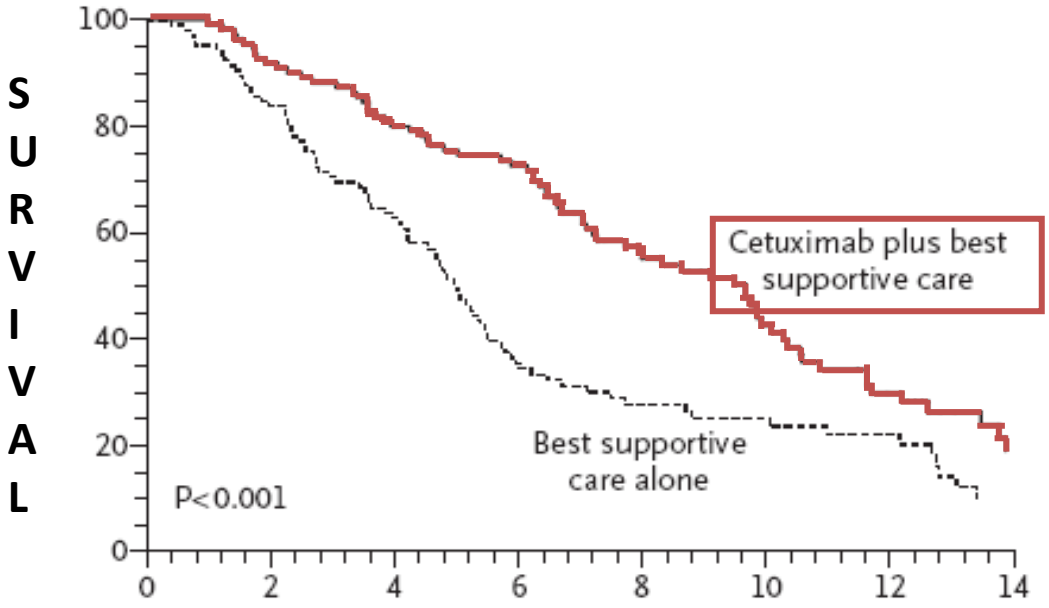
Supplementary Figure 1. Mutations in Toll-like receptor signaling pathways. The *KEGG* (http://www.genome.jp/kegg/) pathway including Toll-like receptor signaling is depicted. Somatically mutated genes in CLL are highlighted in red.

Distribution of Recurrent Mutations in Different Subtypes of CLL



Colon Cancer

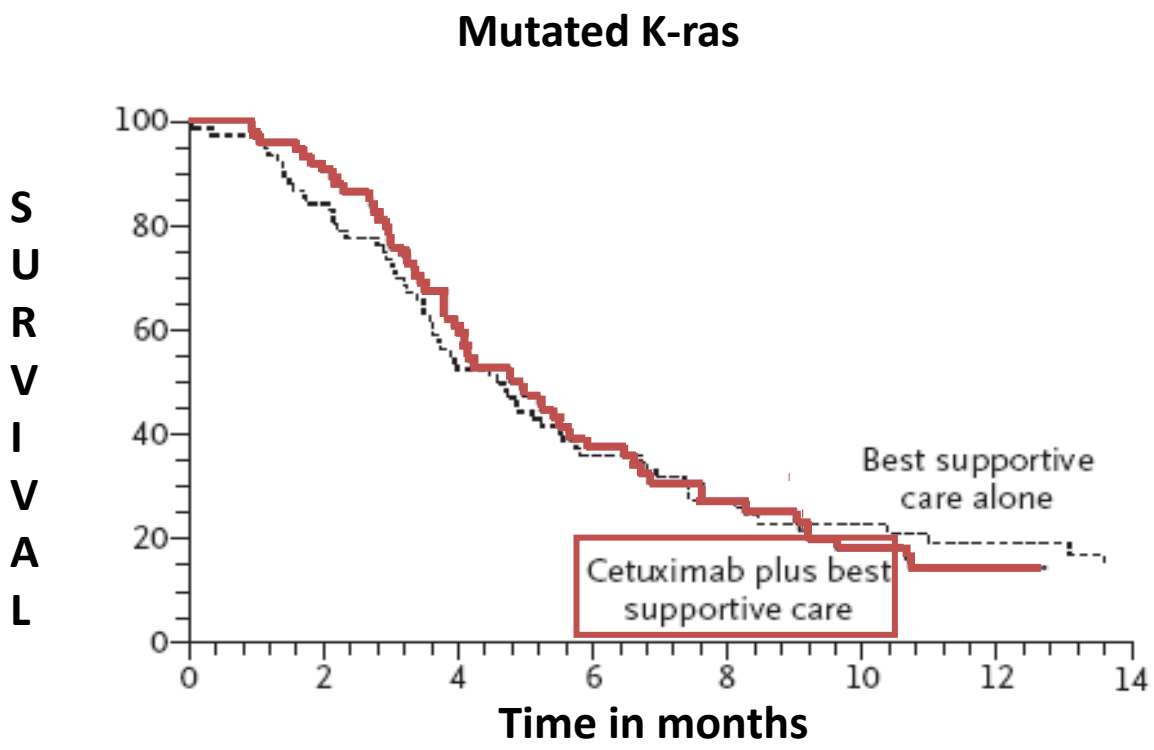
(K-ras wild type)



Zalcberg et al, NEJM 2008

Many colon cancer patients with tumors without K-ras wild type will respond positively to treatment with EGFR receptor anatagonists

Colon Cancer



Zalcberg et al, NEJM 2008

Patients with tumors with K-ras mutations do not: a 30000 Euro treatment will show no effect on the tumor (but will cause significant side effects on the patient)

FDA-approved targeted therapeutics

Tumour	Gene (mutation)	Prevalence of gene alteration (%)	FDA-approved drug	Therapeutic target	Response rate in mutant tumours (%)
Chronic myeloid leukaemia	BCR-ABL (translocation)	>95	Imatinib	ABL1	>95
Gastrointestinal stromal tumour	KIT (mutation), PDGFRA (mutation)	85 (KIT), 5-8 (PDGFRA)	Imatinib	KIT, PDGFRA	>80
Non-small cell lung cancer	EGFR (mutation)	10	Gefitinib, erlotinib	EGFR	70
Chronic myeloid leukaemia (imatinib-resistant)	BCR-ABL (translocation)	>95	Dasatanib	ABL1	>90
Breast cancer – node +ve	HER2 amplification	15-20	Trastuzumab	ERBB2	HR 0.48
Melanoma	BRAF (mutation)	40-70	Vemurafenib	BRAF	>50
Non-small cell lung cancer	EML4-ALK (translocation)	2-7	Crizotinib	ALK	57
Melanoma	BRAF (mutation)	40-70	Debrafenib	BRAF	52
Melanoma	BRAF (mutation)	40-70	Trametinib	MEK1	22
Non-small cell lung cancer	EGFR (mutation)	10	Afatinib	EGFR/ERBB2	50
Breast cancer (metastatic)	HER2 amplification	15-20	Trastuzumab	ERBB2	44

Companion Diagnostic

- Gleevec (imatinib mesylate)
  For the treatment of gastrointestinal stromal tumors (GISTs)

- Gleevec (imatinib mesylate)
  Oral therapy for the treatment of chronic myeloid leukemia

- Herceptin
  Treatment for metastatic breast cancer

- Herceptin (trastuzumab)
  For the treatment of gastric cancer

illumina

Log in to get personalized account information. Quick Order View Cart

Contact Us MyIllumina Tools

APPLICATIONS SYSTEMS CLINICAL SERVICES SCIENCE SUPPORT COMPANY Search

Clinical / Molecular Diagnostics / IVD Assay Development /
MiSeqDx Instrument

Subscribe | Follow us: | Select Language

Molecular Diagnostics

- Overview
- Cystic Fibrosis
- IVD Assay Development
  - MiSeqDx Instrument
  - MiSeqDx Universal Kit
  - Workflow
  - Technology
  - Literature
- Cancer

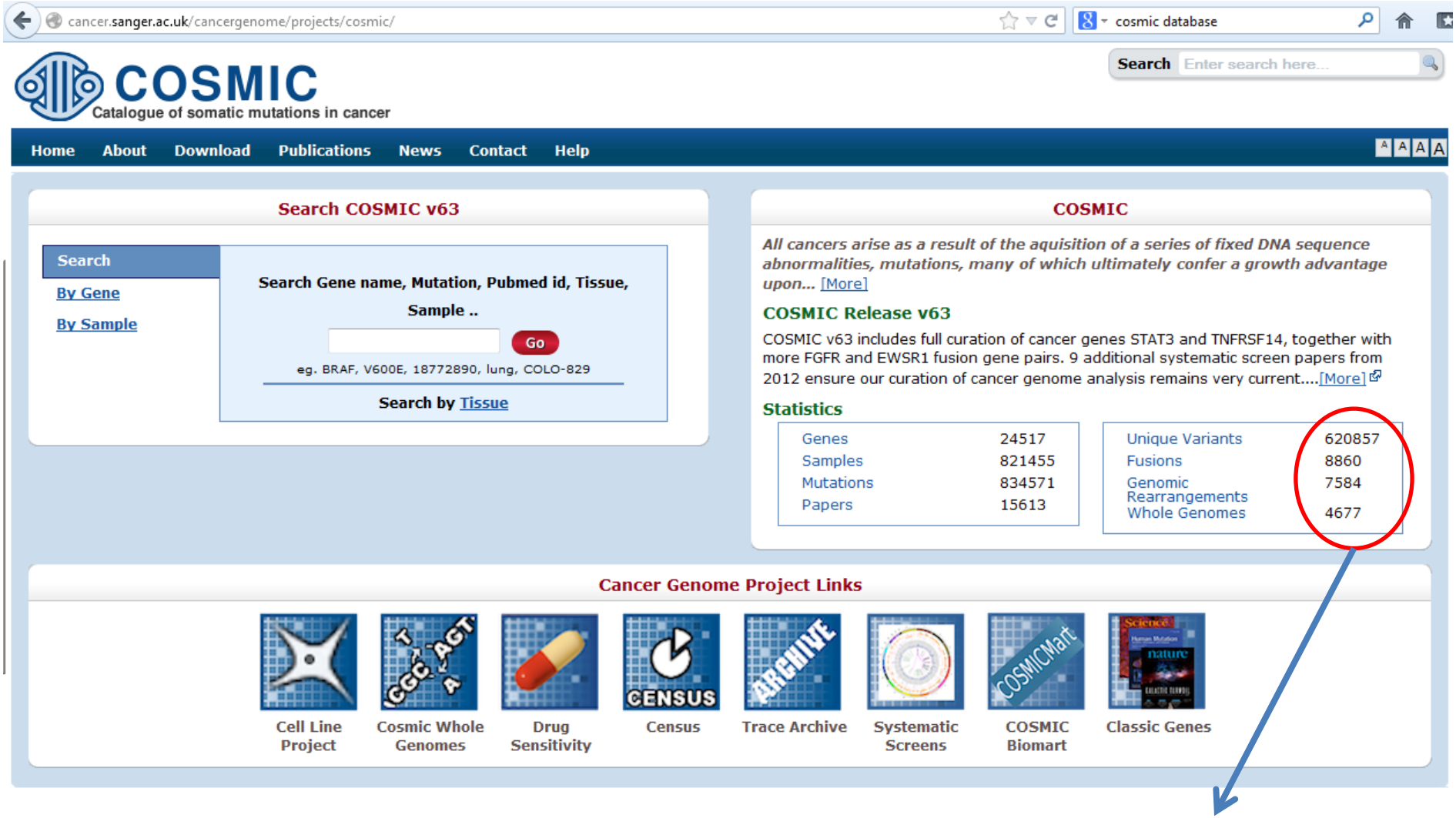
QUESTIONS

The MiSeqDx Instrument

The MiSeqDx instrument is the first and only FDA-cleared *in vitro* diagnostic (IVD) next-generation sequencing (NGS) system. Designed specifically for the clinical laboratory environment, the MiSeqDx instrument offers a small, approximately 4 square feet (0.3 square meters) footprint, an easy-to-follow workflow, and data output tailored to the needs of clinical labs. In addition, the integrated software enables sample tracking, user traceability, and results interpretation*. Taking advantage of proven Illumina sequencing by synthesis (SBS) chemistry, the MiSeqDx instrument provides accurate, reliable screening, and diagnostic testing.

Targeted panels

Standards – e.g. Whole Exome Sequencing, Whole Genome Sequencing



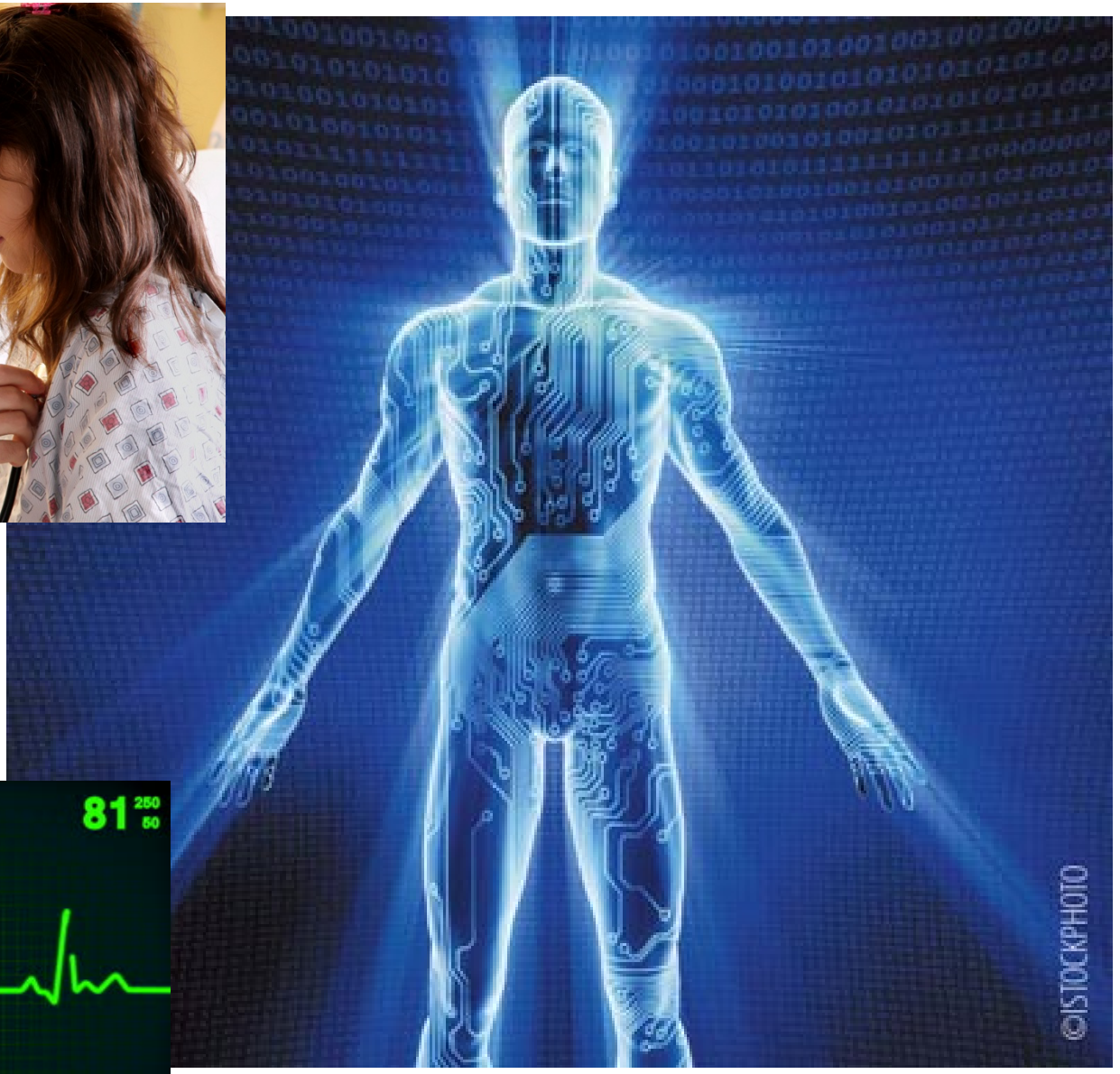
>> 1000 recurrent genomic lesions – gene-centric

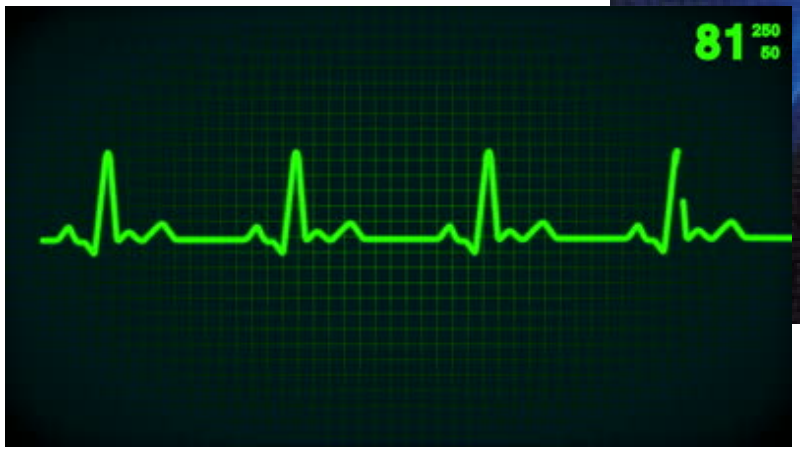
In future

Data-rich, individualised medicine poses unprecedented challenges for ICT, in hardware and software solutions.

Proposition of a data-driven, individualised medicine of the future, based on computer models (**'virtual patients'**) derived from molecular/physiological/anatomical/ environment data from individual patients.

Individualised versions of the models, produced for each patient, will then be used to identify personalised prevention/therapy schedules and side effects of drugs.





Data-rich

Run simulation on a computational Avatar to identify the most suitable treatment

Large-scale International Consortia

CNAG is a major contributor and driver in three large International Initiatives:

1.- International Cancer Genome Consortium (ICGC)

Spanish Project on Chronic Lymphocytic Leukaemia
French Project on Prostate Cancer
French Project on Ewing's Sarcoma

2.- International Human Epigenome Consortium (IHEC)

EU-funded Project Blueprint

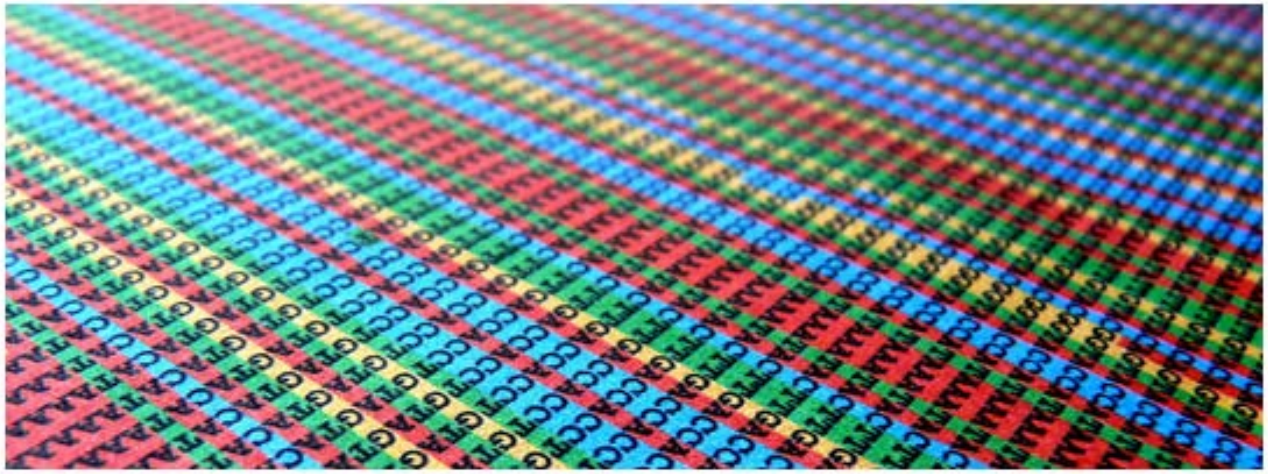
3.- International Rare Disorders Research Consortium (IRDiRC)

Spanish Project on Charcot-Marie-Tooth Disease
EU-funded C-Project on data analysis and coordination

DNA tests to revolutionise fight against cancer and help 100,000 NHS patients

Monday 10 December 2012

Life Sciences Strategy report shows UK has generated more than £1bn on industry and private sector investment within 12 months



Prime Minister David Cameron will today announce plans to transform cancer treatment in England with new proposals to introduce high-tech DNA mapping for cancer patients and those with rare diseases, within the NHS.

The UK will be the first country in the world to introduce the technology within a mainstream health system, with up to 100,000 patients over three to five years having their whole genome – their personal DNA code –sequenced.

A word on resources

- 3000 human genomes/a – 1 cancer investigation = multiple genomes
- Alignment ~ BWA 2000 cpuh for one 30x coverage human genome
- 9.000.000 cpuh cost 200.000€ in electricity (Spain)
- Cost of electricity for the alignment of one human genome costs 45€
- Cost of electricity for all computational work **~ 100€/genome**
- Storage – data for one human genome ~ 500Gbytes (bam files)
- 1 Pbyte costs ~ 300.000€
- Cost of hardisks for storage **150€/genome**
- **Only hardisks and electricity for 100.000 human genomes would cost 25 million €**
- **Cloud cpuh ~0.11$ and 400 Gb storage ~100's $/month**

How to marry reasearch and clinic for the good of society

OICR
Ontario Institute for Cancer Research

Event registration | Member Login

Search

HOME | ABOUT | PROGRAMS | EDUCATION | CAREERS | CONTACT

Home → OICR Programs → The global alliance to enable responsible sharing of genomic and clinical data → The global alliance to enable responsible sharing of genomic and clinical data

The global alliance to enable responsible sharing of genomic and clinical data

OVERVIEW

Since its introduction, the cost of genome sequencing has plummeted, opening new opportunities in biomedical research. Research organizations worldwide use this data to advance human health, allowing us to learn more about disease and find new diagnostic tools and treatments. While these advancements are positive, a group of more than 70 organizations, including the Ontario Institute for Cancer Research (OICR), have forged a global alliance to realize the full potential of genetic and clinical data in research. The alliance aims to establish a common international framework that will allow this data to be collected, managed and shared in an effective, responsible and interpretable manner.

"There's some amazing, innovative research being conducted, but the knowledge generated is retained in separate institutions and not shared," says Dr. Tom Hudson, President and Scientific Director of OICR. "We can give this science a boost with better organization and consistent technical standards." Currently, valuable genetic and clinical data are isolated – whether it be by research organization, country or disease area - limiting its impact. The idea behind the alliance is to break down barriers so that as much data as possible can be shared and used in studies. This will also allow researchers to tackle larger, more complex research problems.

"The ability to collect and analyze large amounts of genomic and clinical data presents a tremendous opportunity to learn about the underlying causes of cancer, inherited and infectious diseases, and responses to drugs," says David Altshuler, Deputy Director of the Broad Institute of Harvard and MIT. "However, we will only realize this opportunity if we can establish effective and ethically responsible approaches to share data. We believe that by working together, and by committing to the principle that each individual has the right to decide whether and how broadly to share their personal health information, we can accelerate progress in life sciences and medicine."

THE GLOBAL ALLIANCE TO ENABLE RESPONSIBLE SHARING OF GENOMIC AND CLINICAL DATA

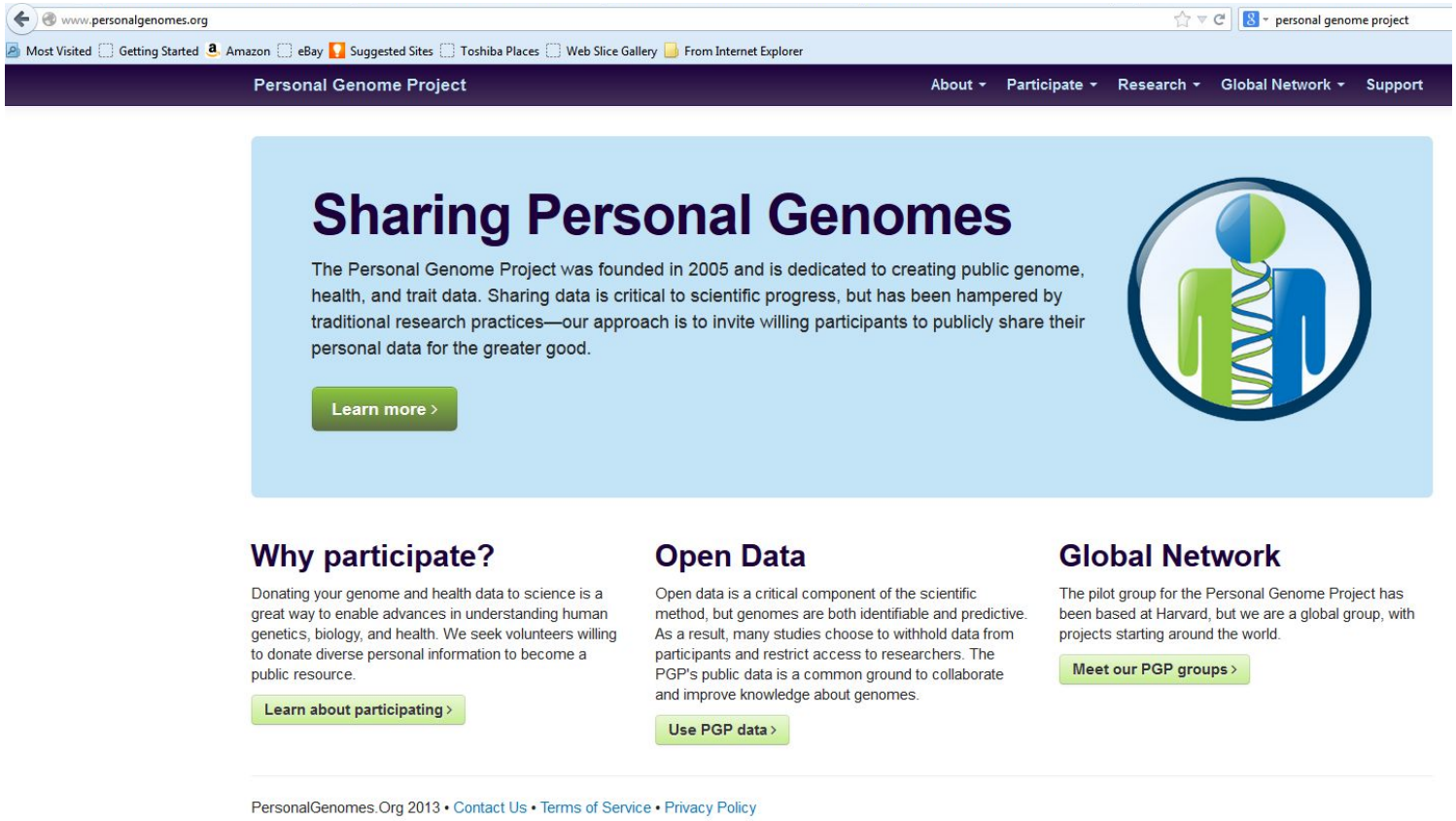
White Paper
Events
Fact Sheets
Media Coverage
Newsletters
Contact

See All OICR Programs

centro nacional de análisis genómico

cnag

The Personal Genome Project



cnag

baldiri reixac, 4
pcb - tower i, 2nd floor
08028 barcelona

t +34 93 4020542
f +34 93 4037279
www.cnag.eu

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

cnag