

Genómica Funcional en Investigación Biomédica

Pedro Jares

Pathology Department, Hospital Clinic, Barcelona
Genomics Unit, IDIBAPS



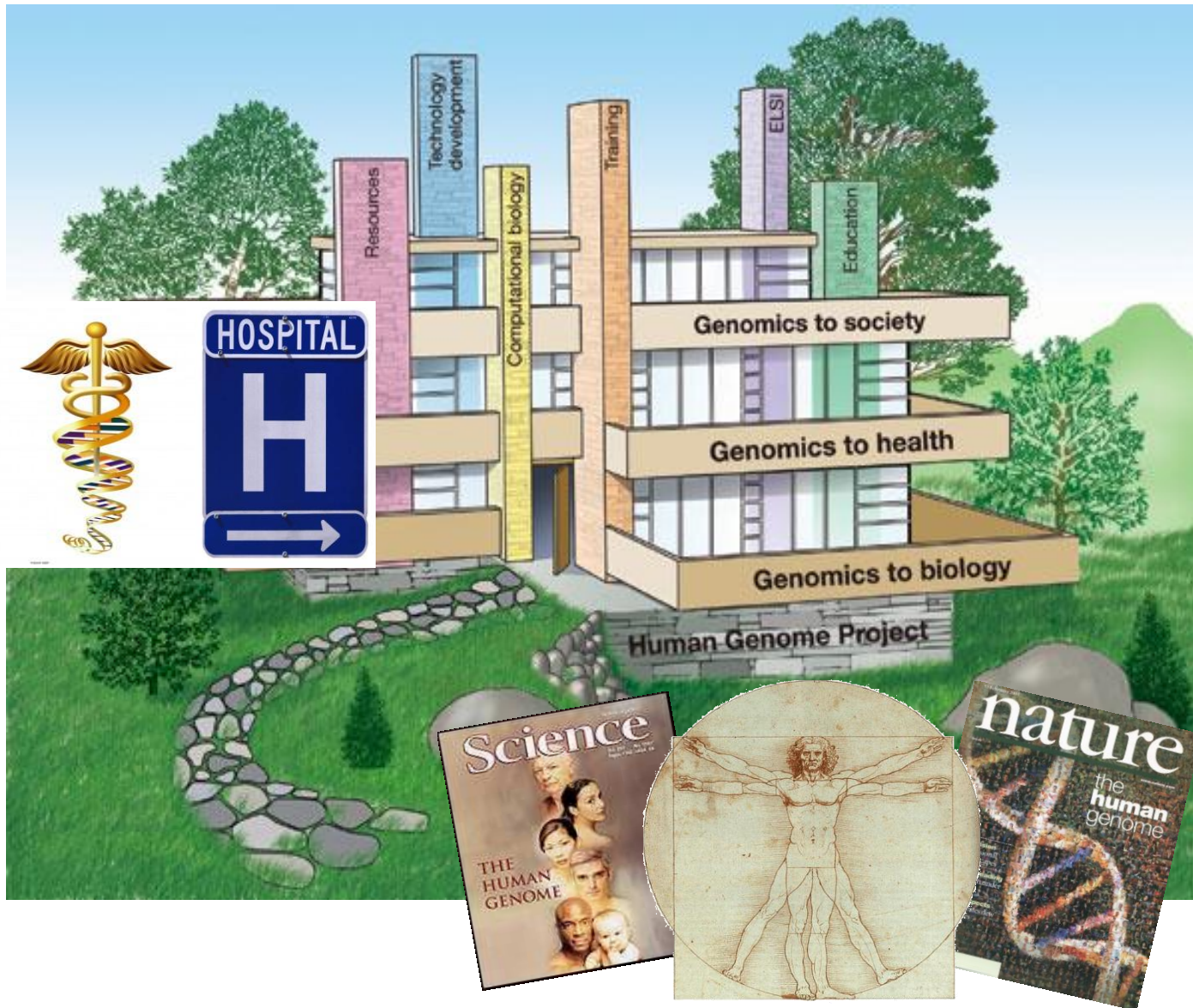
Facultat de Farmàcia
Comissió de Recerca

Avda. Joan XXIII s/n
08028 Barcelona
tel.93 402 18 81
fax.93 402 18 86



SEMINARIS TECNOLÒGICS A LA FACULTAT DE FARMÀCIA 2012

Functional Genomics–Personalized Medicine



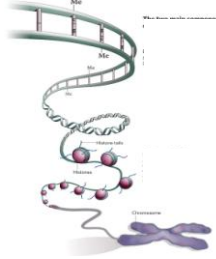
FUNCTIONAL GENOMICS

GENOMICS TOOLS

Genomics

(1) Genome

DNA sequence
Structural genomics
Epigenetics



DNA sequence (mutations)
DNA variation (SNPs, CNVs)

Structural genomics
Histone modifications
DNA methylation

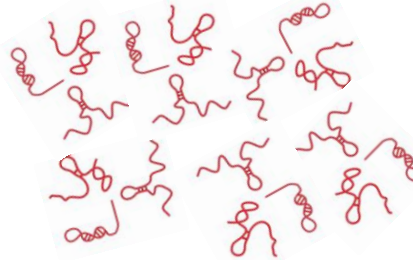
DNA microarrays



Transcriptomics

(2) Transcriptome

mRNA and
non-coding RNA

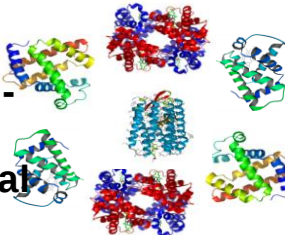


Gene expression
Splicing variants

Proteomics

(3) Proteome

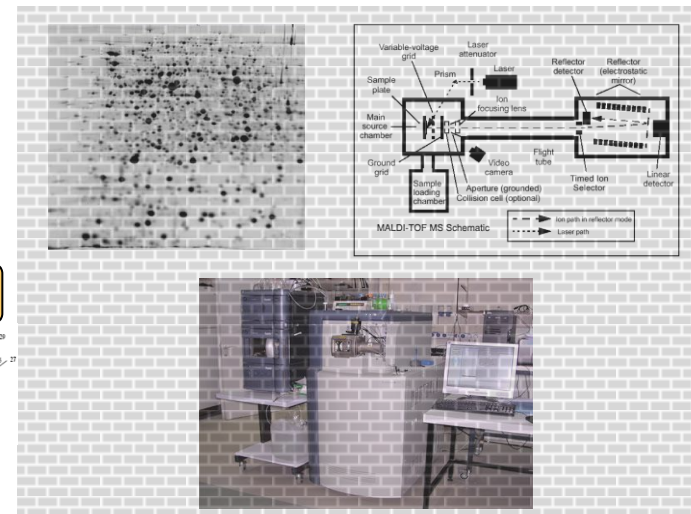
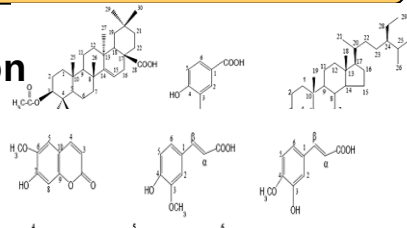
Protein expression
Post-translational modifications
Structural and functional analysis



Metabolomics

(3) Metabolome

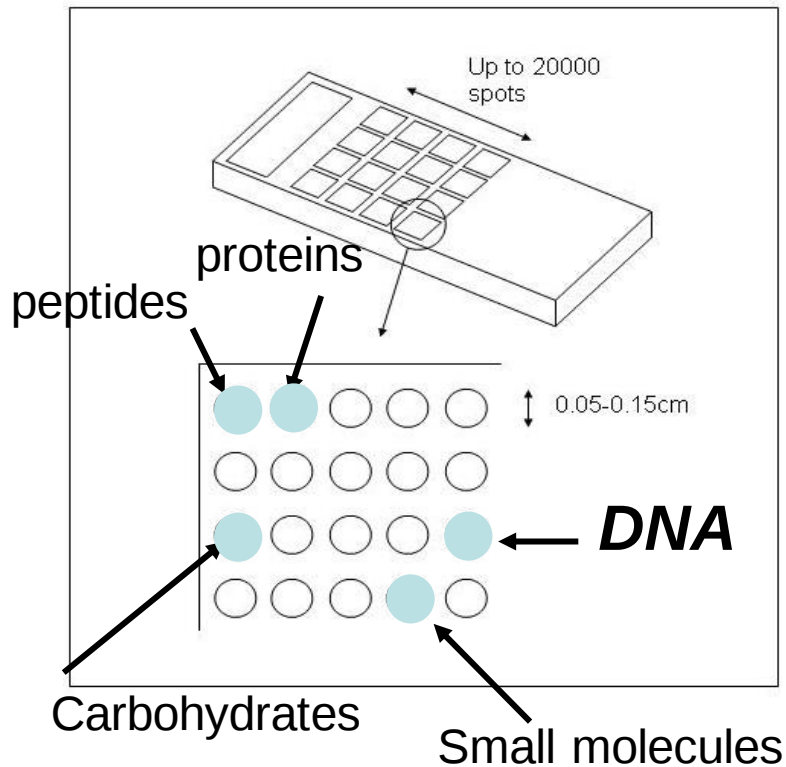
Profile and function
of metabolites



Microarray: Universal Biochemistry Platform

A microarray is a compact device that contains a large number of well-defined immobilized capture molecules assembled in an addressable format.

Synthetic oligos, PCR products, proteins, antibodies, carbohydrates etc....



a) You can expose an unknown (test) sample on it and then examine where the molecule was captured.

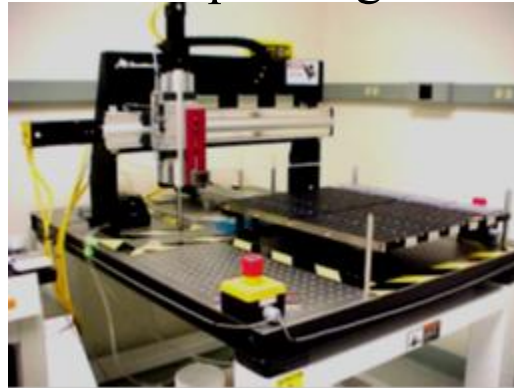
b) You can then derive information on identity and amount of captured molecule.

Array Life Cycle

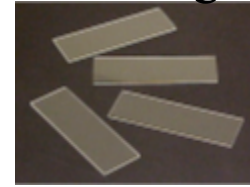
amplifying



printing



coating



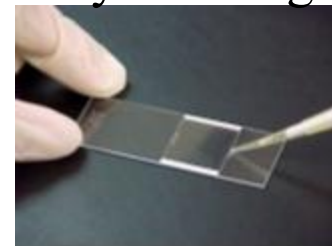
processing



labelling



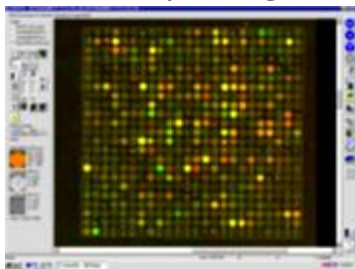
hybridizing



scanning

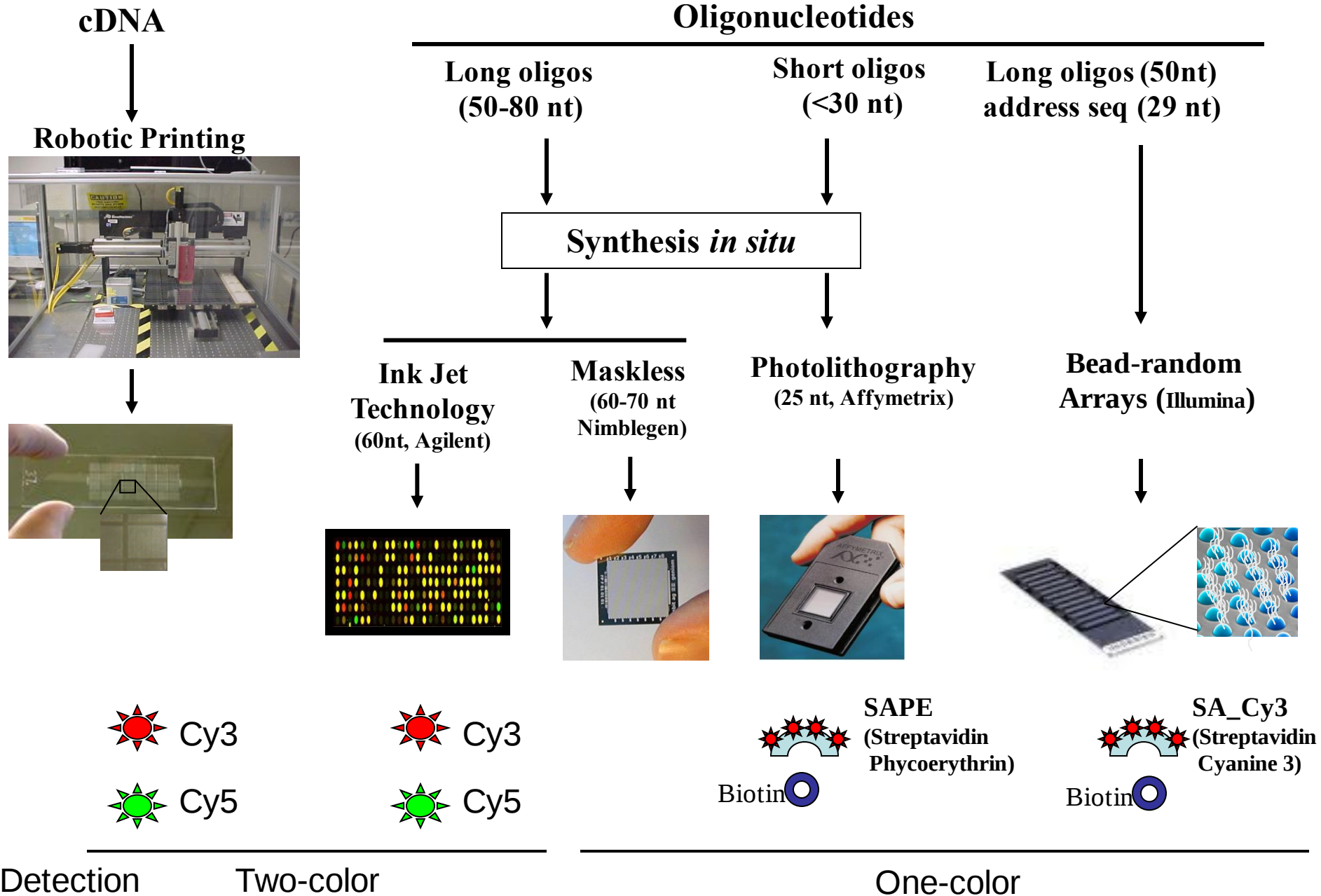


analyzing



Biological
Question ?

Types of Array

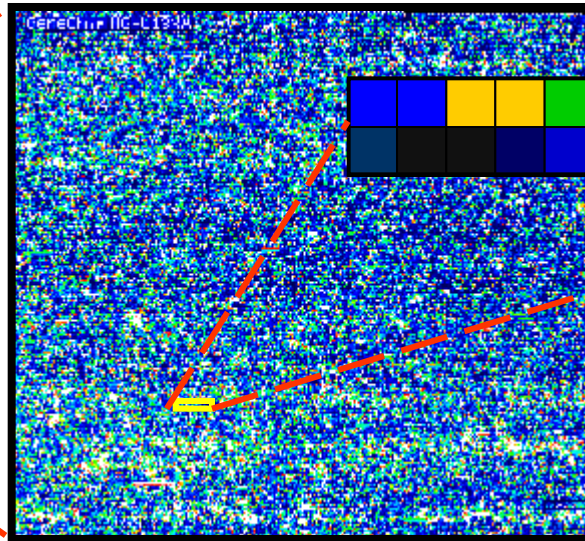


GeneChip® and Probe selection

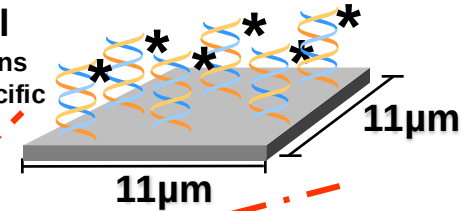
Up to ~1,400,000 features/chip



1) Probe Array



4) Probe cell
Each Probe Cell contains
 $5 \times 10^{5-6}$ copies of a specific
probe



2) Probe set

Each Probe Set contains
11 Probe Pairs (PM:MM)
of different probes

3) Probe pair

Each Perfect Match
(PM) and Mismatch
(MM) Probe Cells are
Associated by pairs

Probe selection

5' cDNA sequence 3'

Probe set

Perfect match oligo (PM)

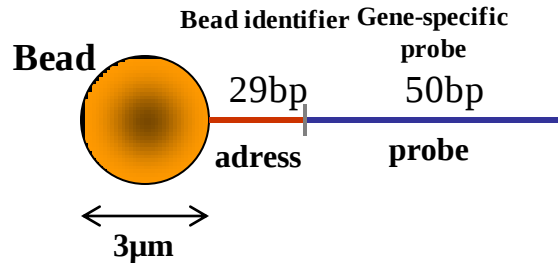
ATTAACGGGCAT**T**GCATTAGCACGT

Mismatch oligo (MM)

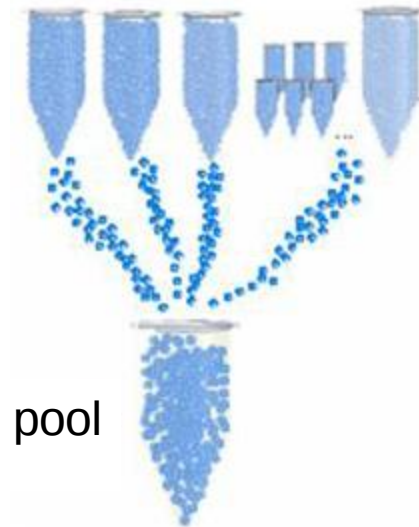
ATTAACGGGCTT**A**GCATTAGCACGT

11-20 Oligonucleotides/ gene

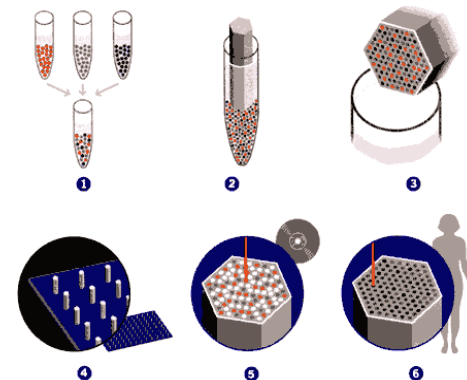
BeadArrays: assembly of a random array



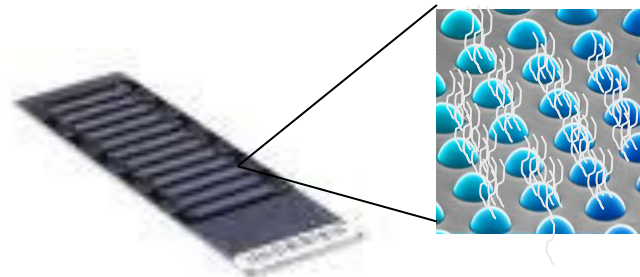
n-thousands bead types



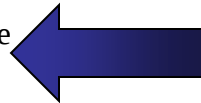
Randomly array
beads into wells



Decode each bead using
hybridisation to address sequence

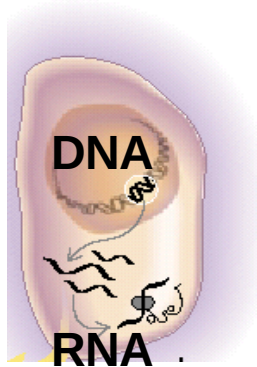


DMAF Files
The BeadChip decode
map files (.dmap)

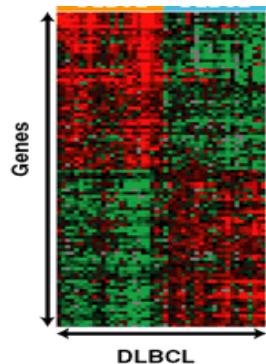


48000 probe
average 30x
99.99 % of probes are
represented in each array

DNA microarrays applications



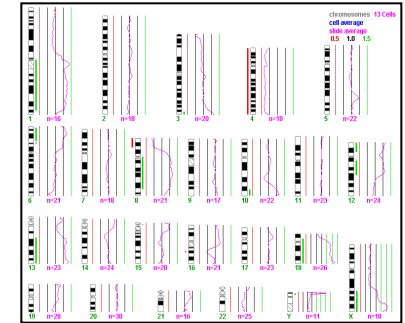
Expression arrays



CGH arrays



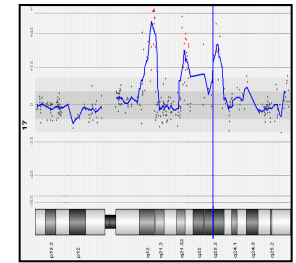
Copy number
analysis



Promoter arrays



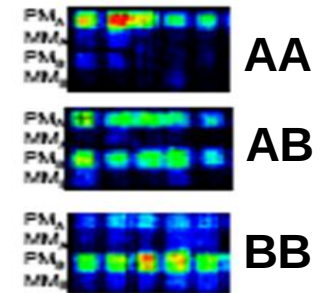
DNA regulation
Epigenetics



SNPs arrays



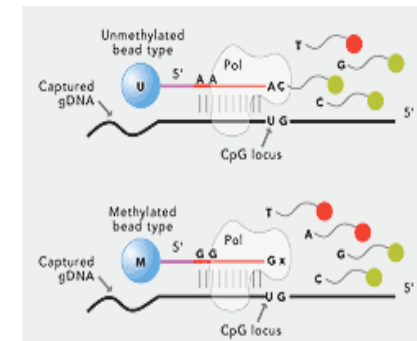
Genotyping



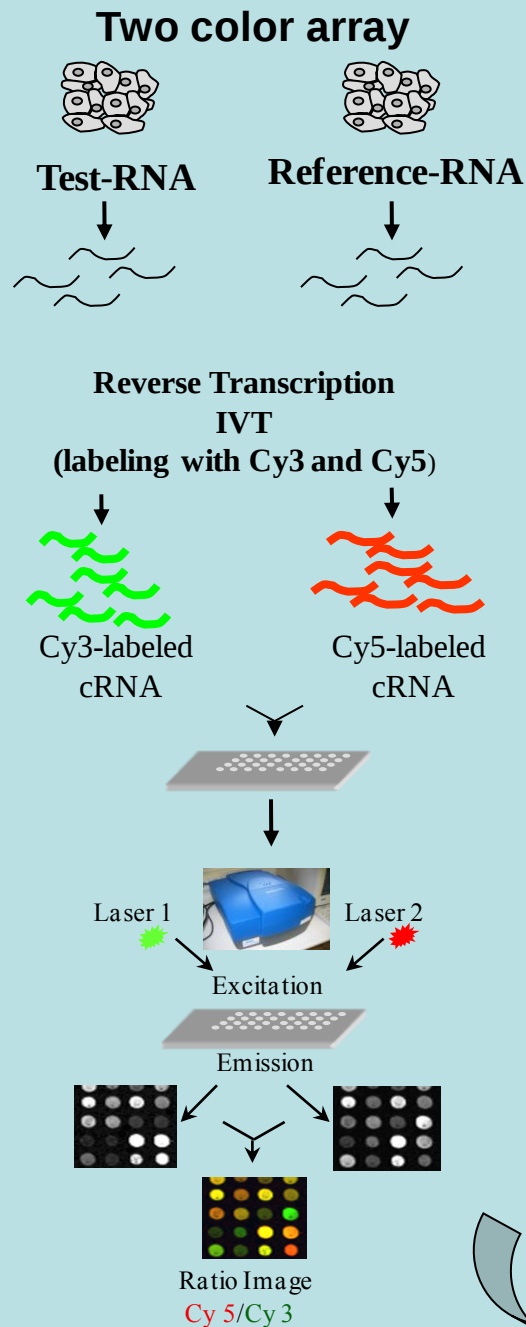
CpG arrays



DNA methylation



Gene Expression Profiling



RNA
Isolation

Target
Labeling

Hybridization
Washing
Streptavidine-PE
Staining

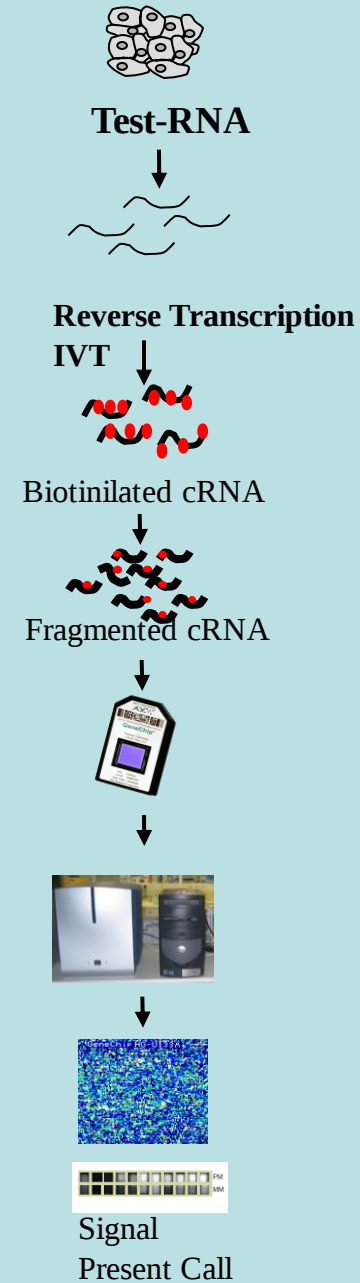
Scanning

Analysis

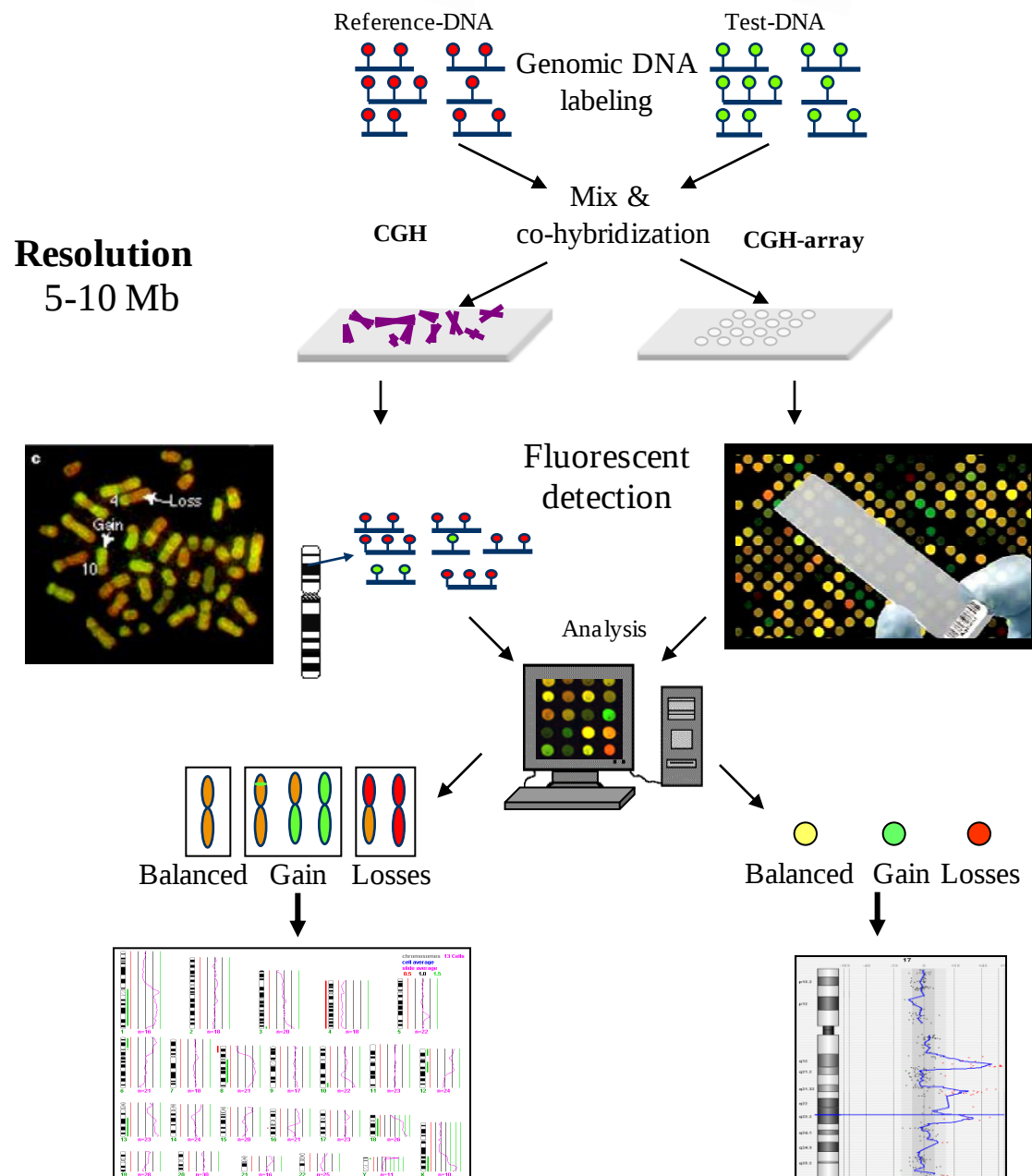


Data Mining/
Bioinformatics

Single color array



CGH-array



Resolution
5-10 Mb

Resolution

BAC-arrays:

Printing of 32,433 overlapping
BAC clones >100Kb

Agilent CGH array;

244k 60mer ;

8.9 kb median probe spacing

NimbleGen HD-arrays

Long oligos probes

~2,100,000 (50mer-85mer)

1.1kb median probe spacing

Illumina human 1Mduo

1,200,000 loci per bead chip

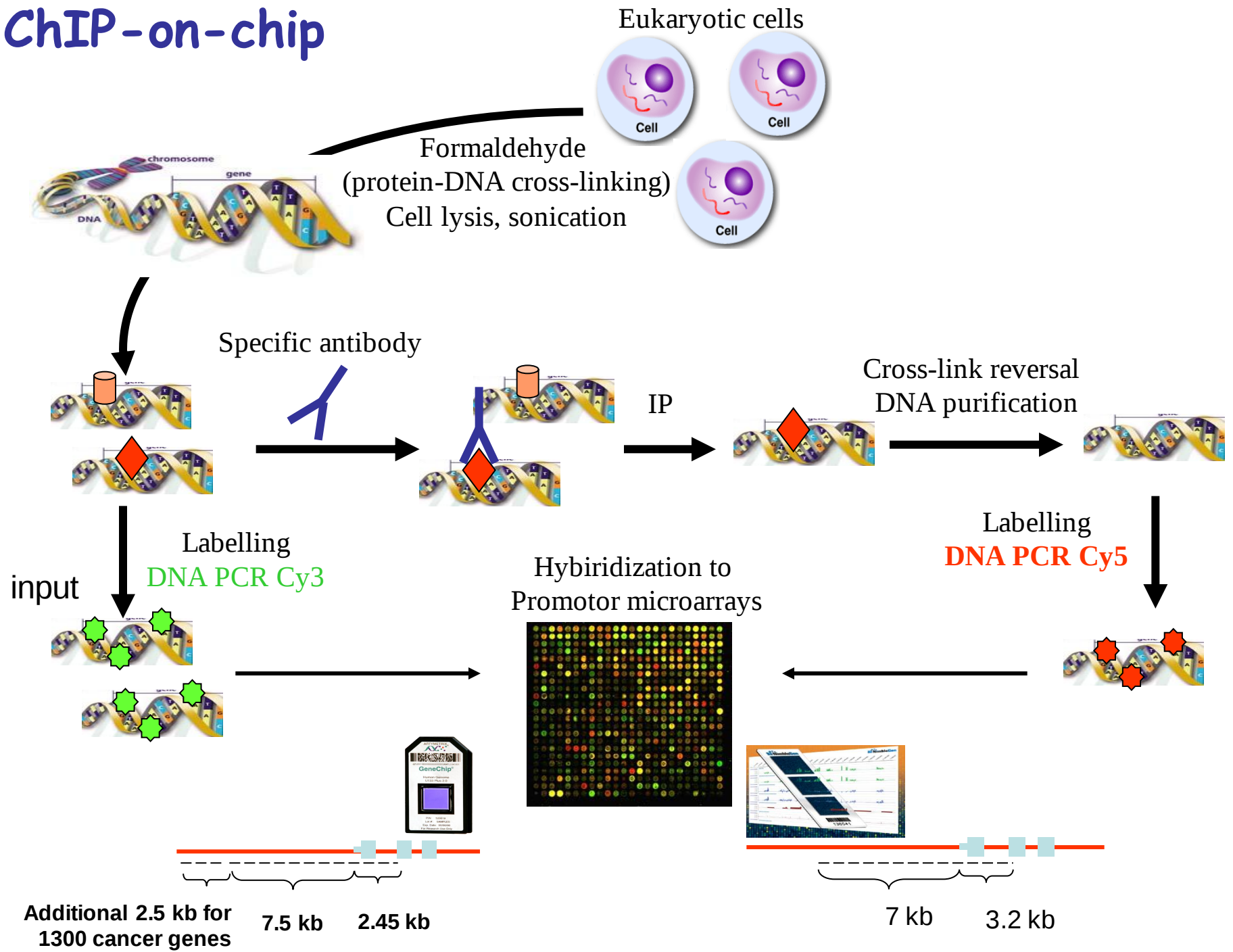
1.5 kb median marker spacing

Affymetrix SNP array 6.0

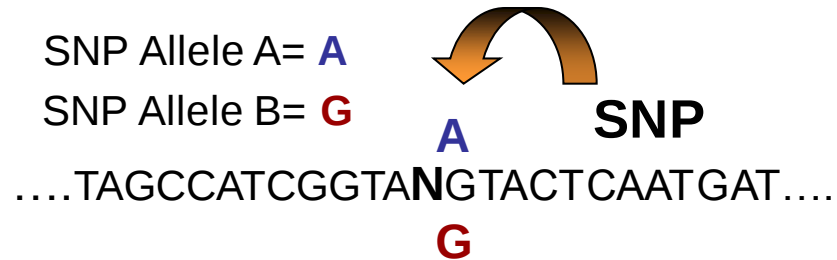
1,800,000, intermarker distance

696bp

ChIP-on-chip



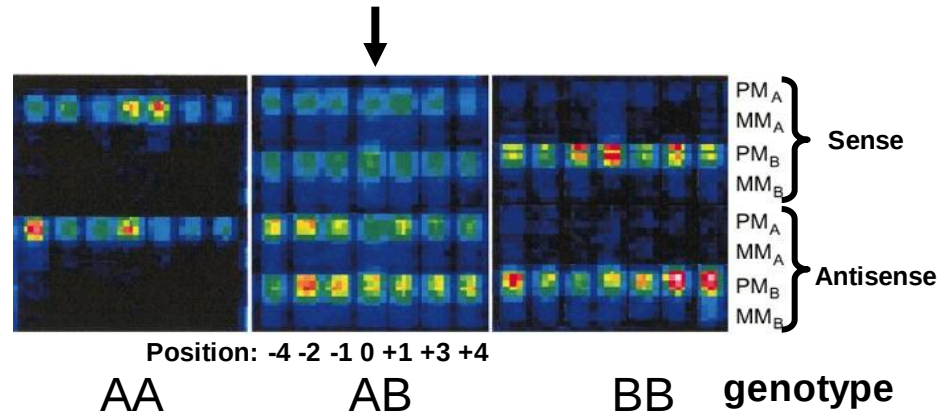
SNP Arrays



Affymetrix Genotyping Array Design

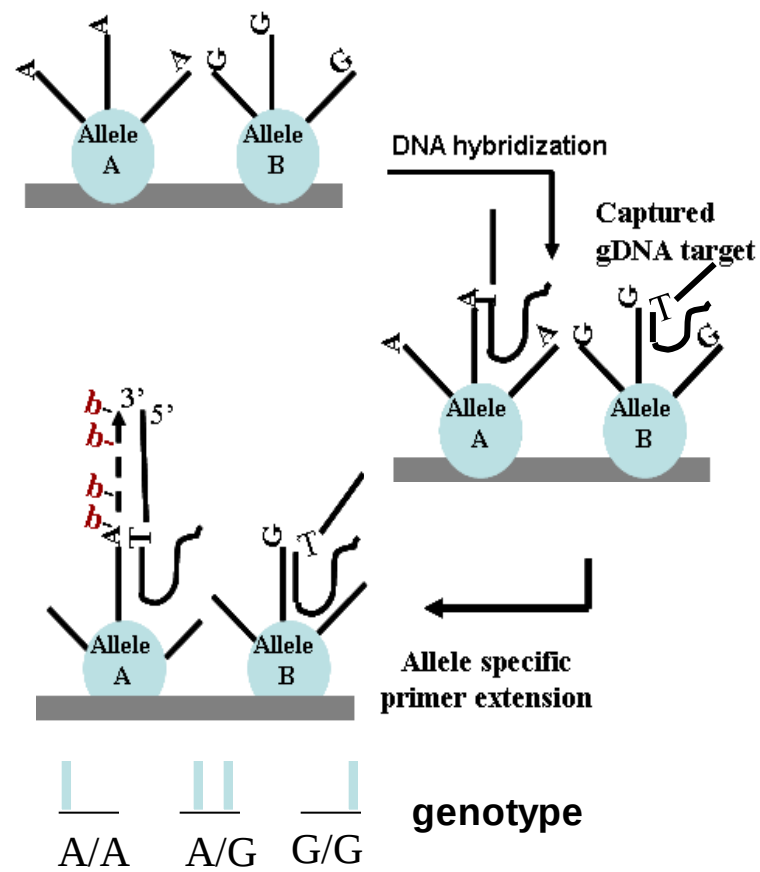
Sequence-Specific Oligonucleotide (SSO) Probes

ATCGGTAGCCAT**T**CATGAGTTACTA Allele A
 ATCGGTAGCCAT**C**CATGAGTTACTA Allele B

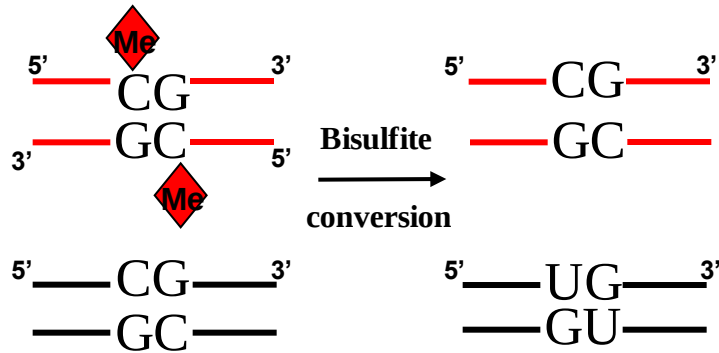


Infinium Illumina SNP

Allele-specific primer extension (ASPE)

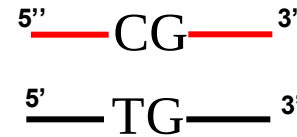


CpG Array: Infinium Methylation Assay



Whole genome amplification

Enzyme fragmentation



Hybridization



Human Methylation 27 k

N=12 samples

β values

Intensity [M]

Intensity [M] + [U]

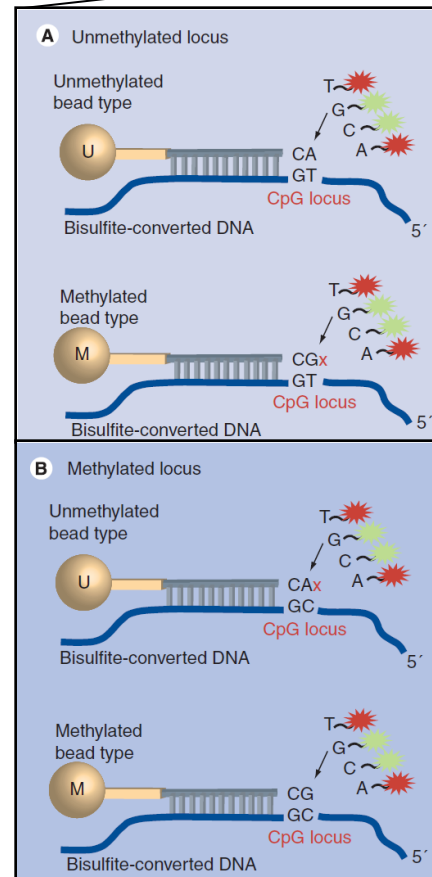
(0-1)

Stained and scanned

Unmethylated DNA



Methylated DNA



Allele-specific primer annealing.
Single base extension

B-Cell NHL Pathogenesis:

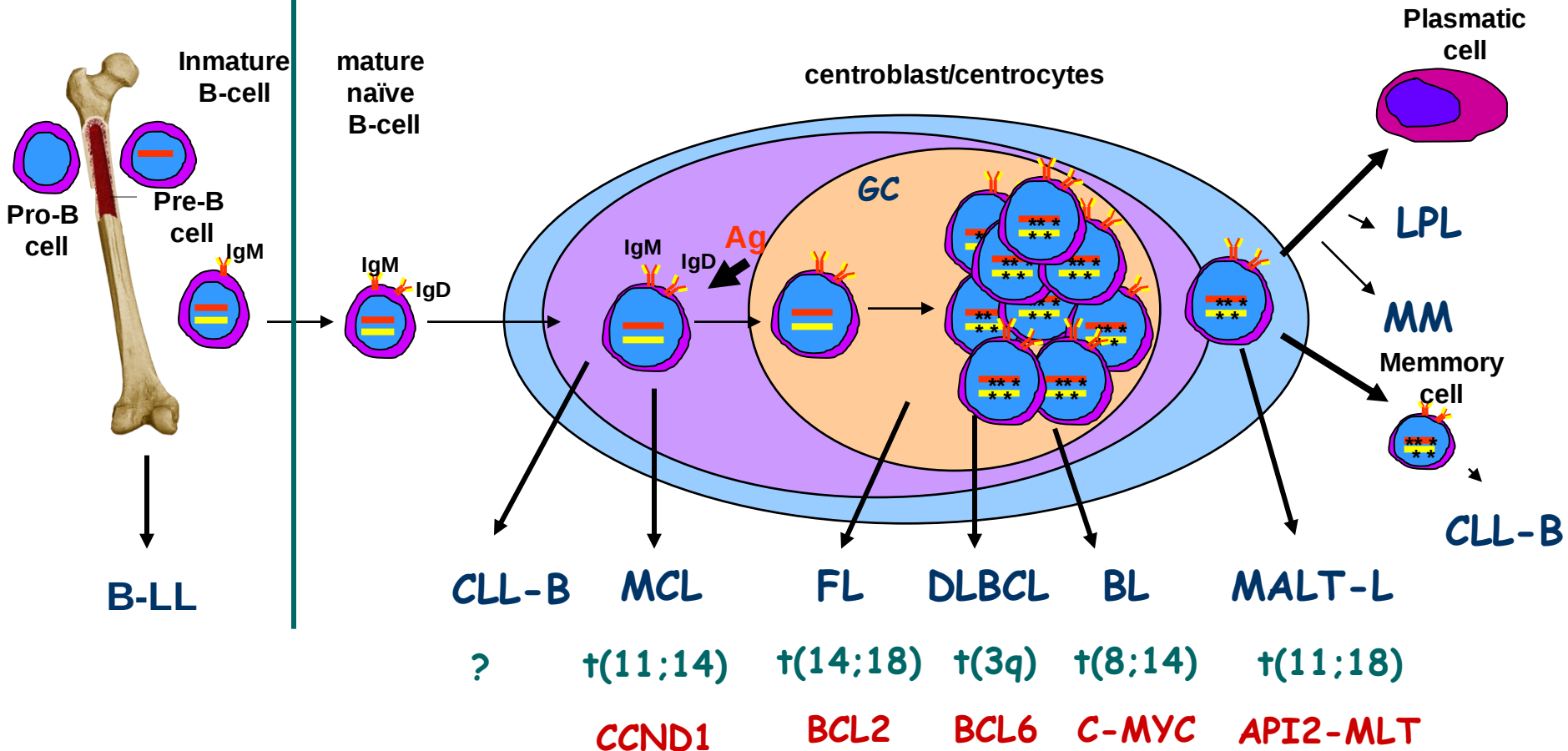
Cellular origins, primary genetic alterations and target genes

Bone Marrow

Follicular area

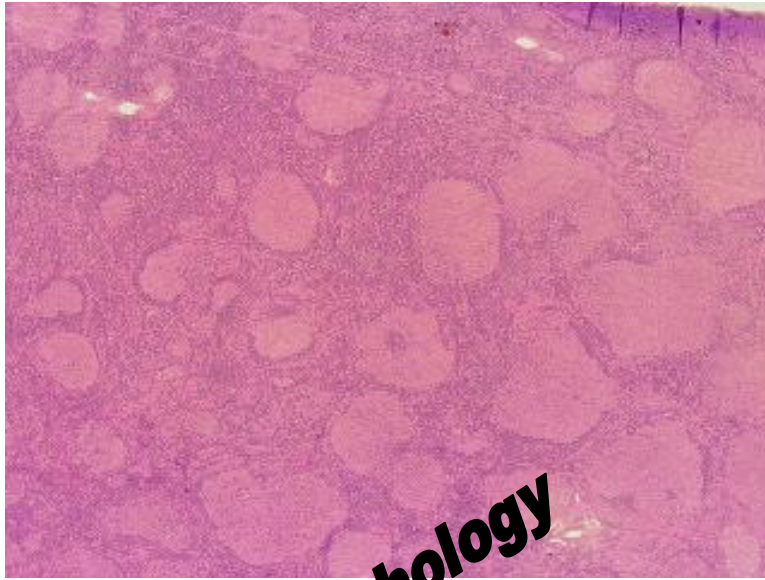
Somatic Hypermutation
Class switch recombination

V(D)J
recombination

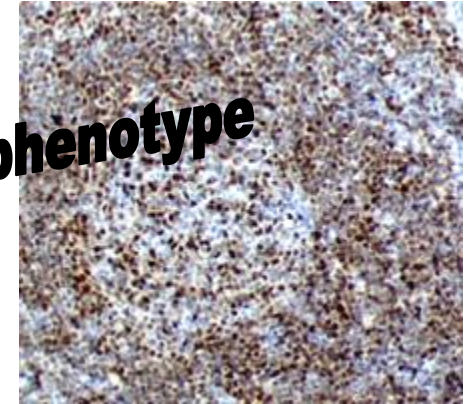
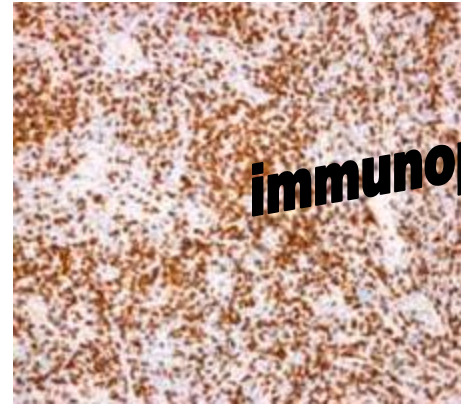
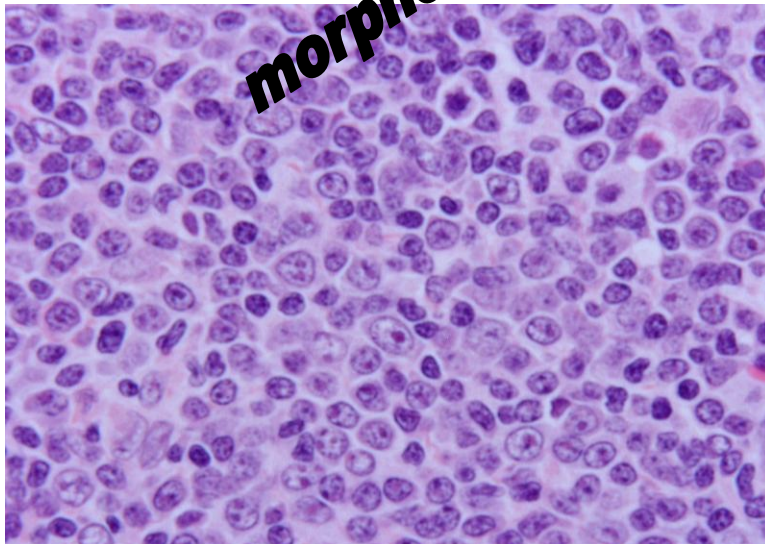


Diagnosis in Hematopathology

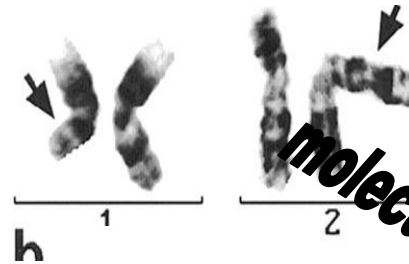
From Morphology to an Integrated Diagnosis



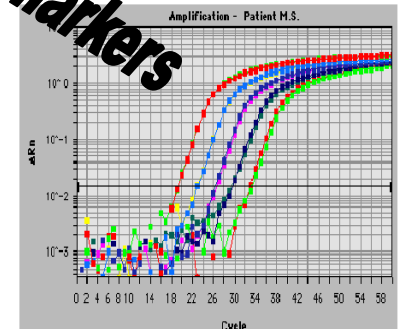
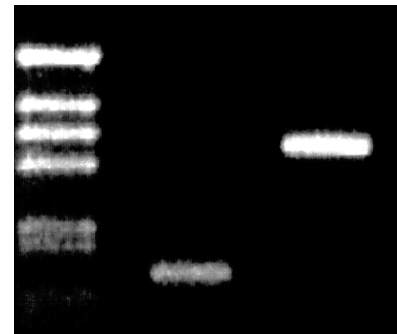
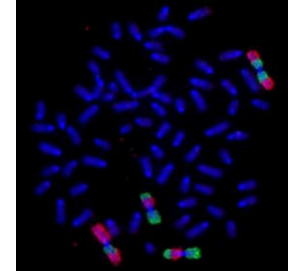
morphology



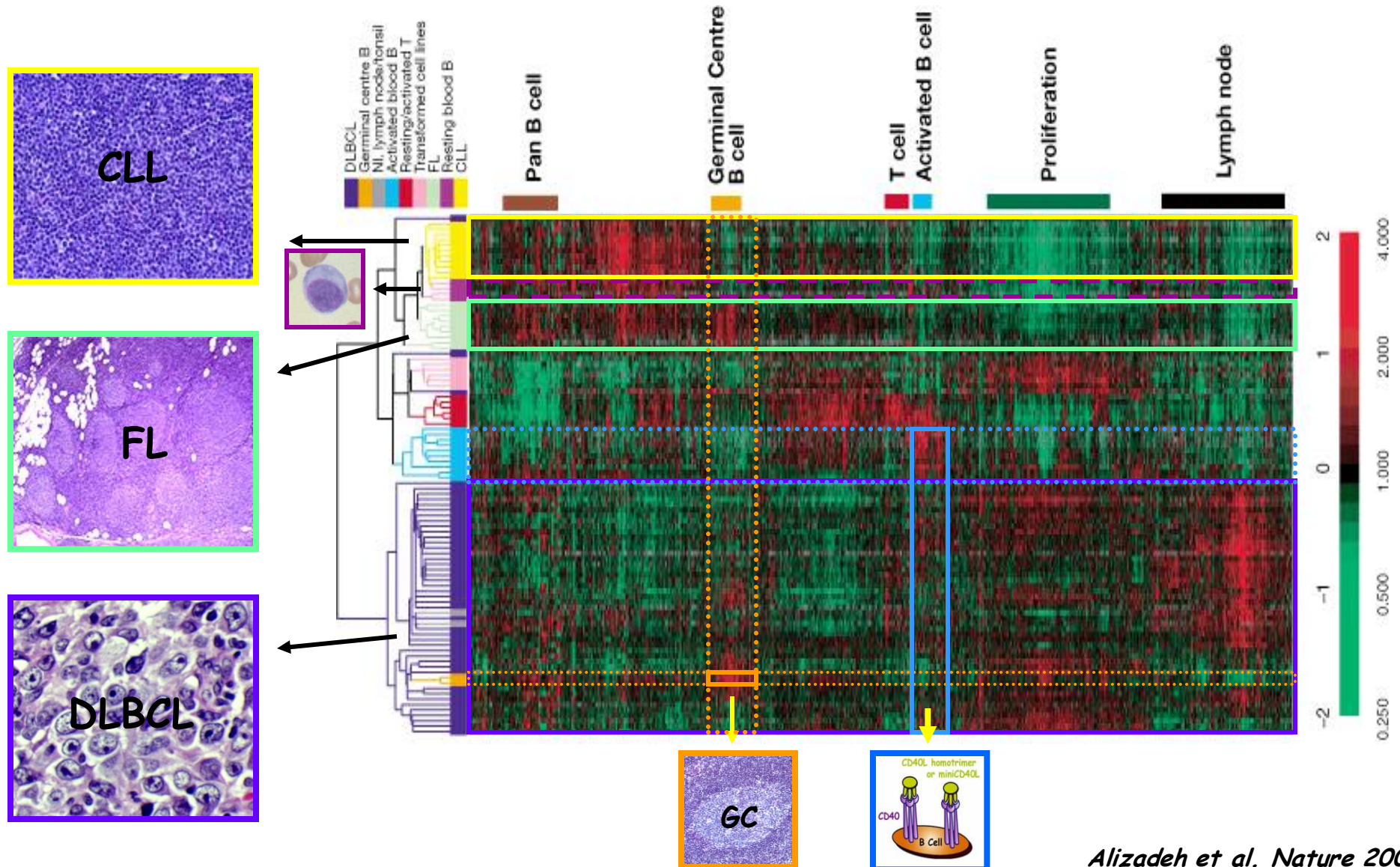
immunophenotype



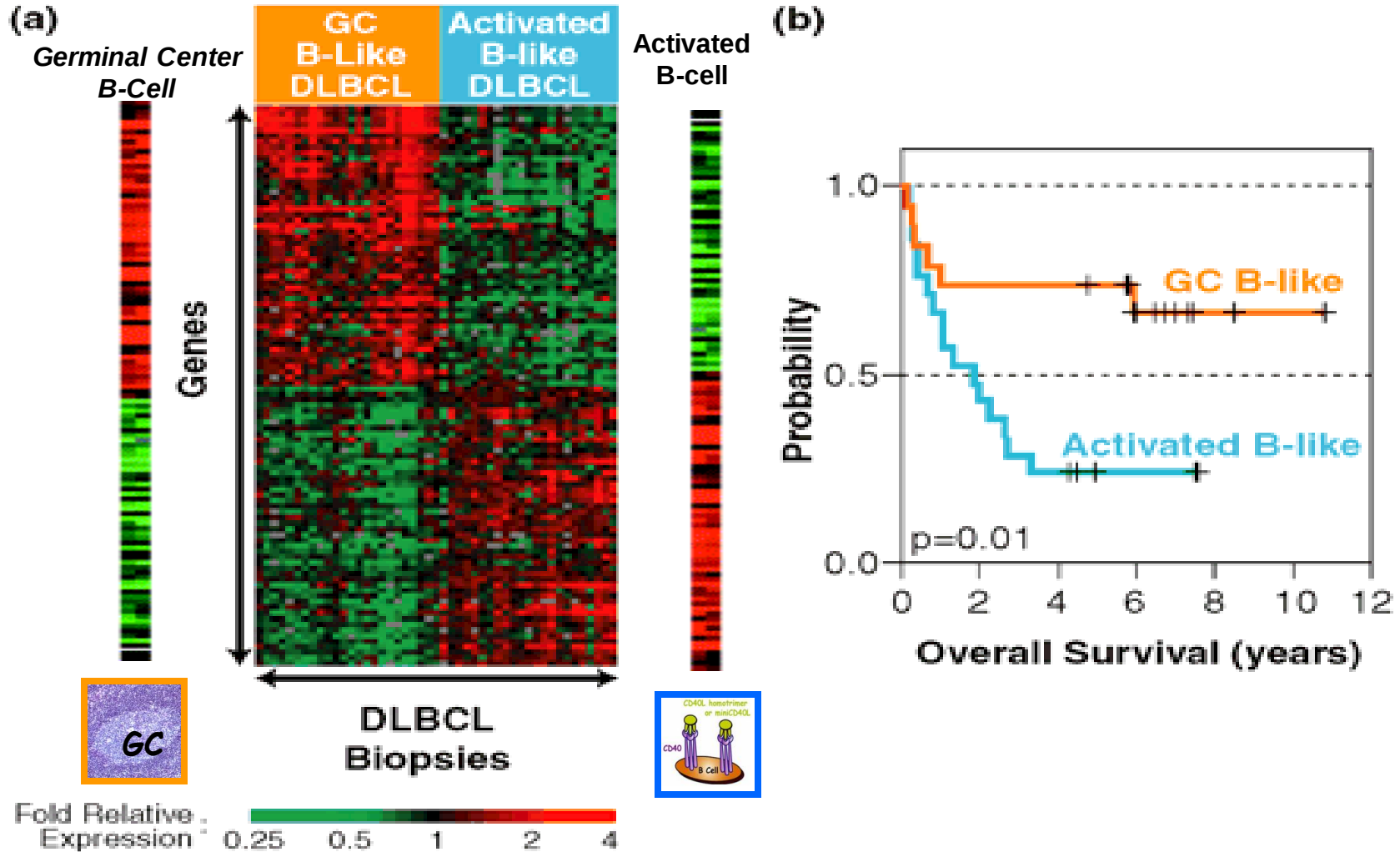
molecular markers



Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling



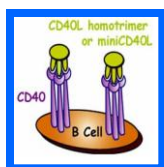
Diffuse large B cell lymphoma (DLBCL) comprises at least two distinct diseases



A



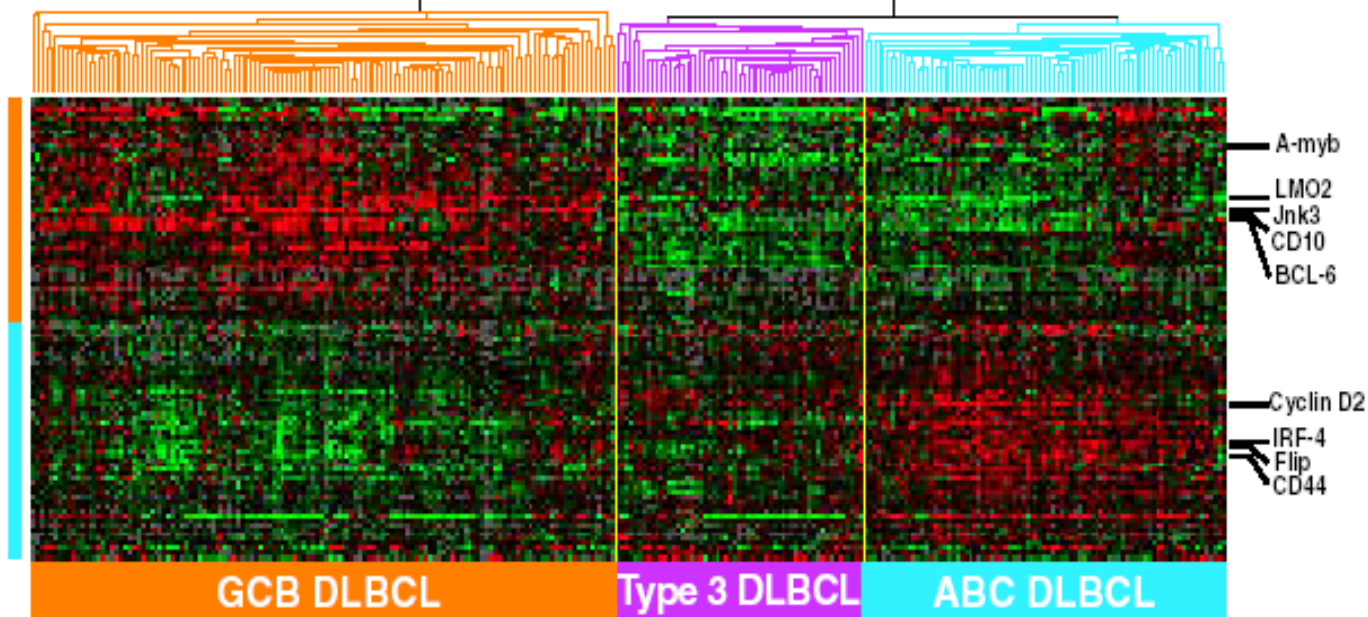
VS



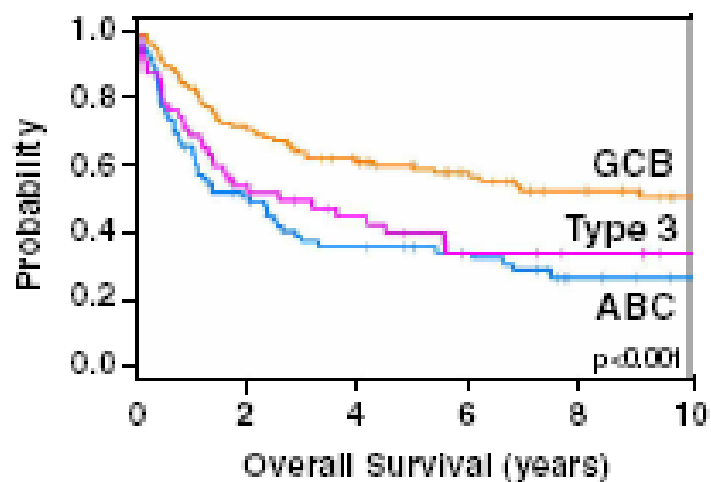
DLBCL
Subgroup
Distinction
Genes

n=100

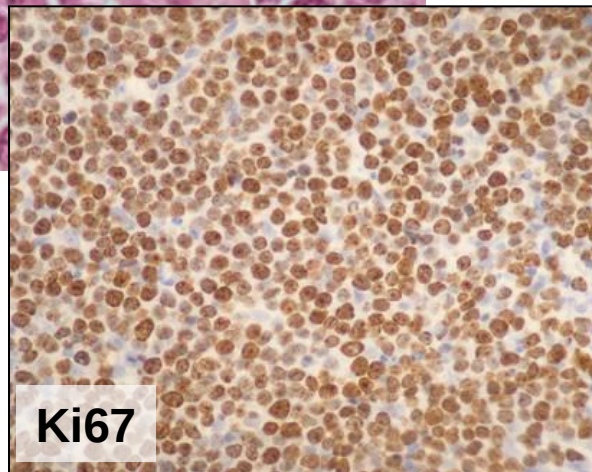
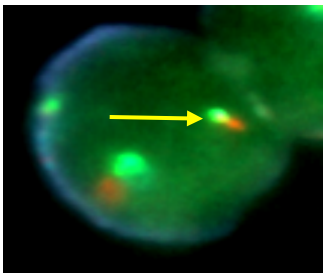
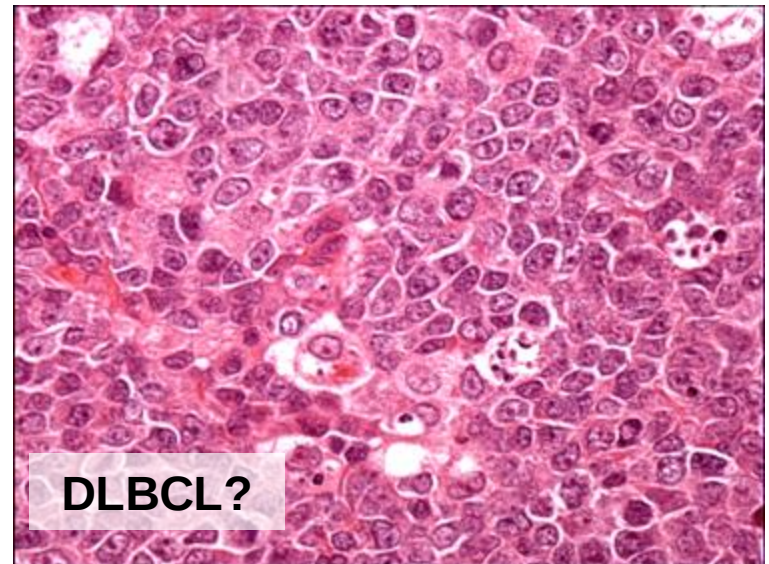
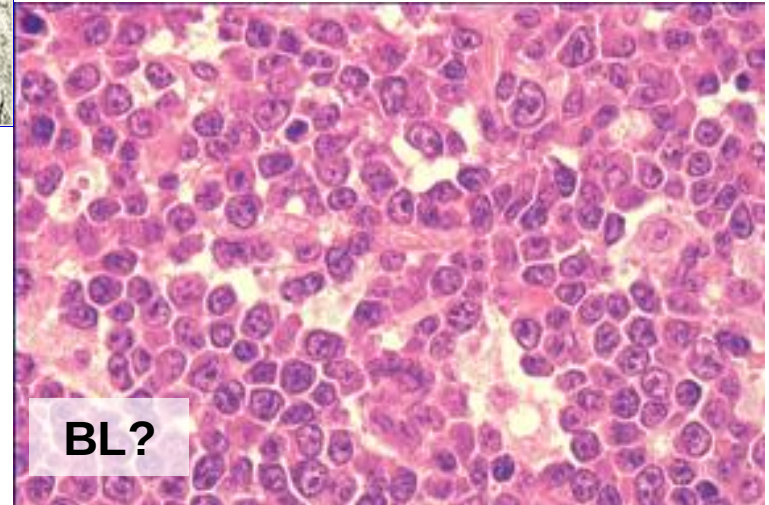
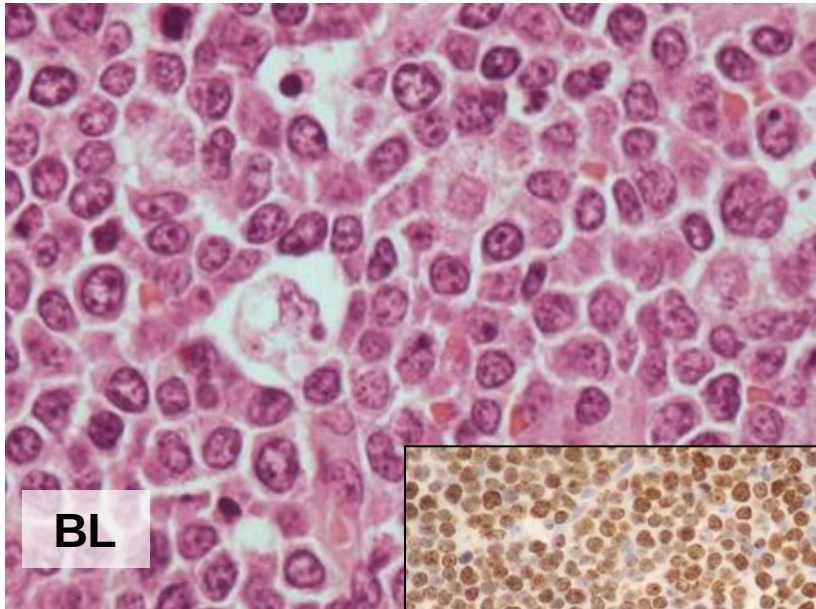
N= 240



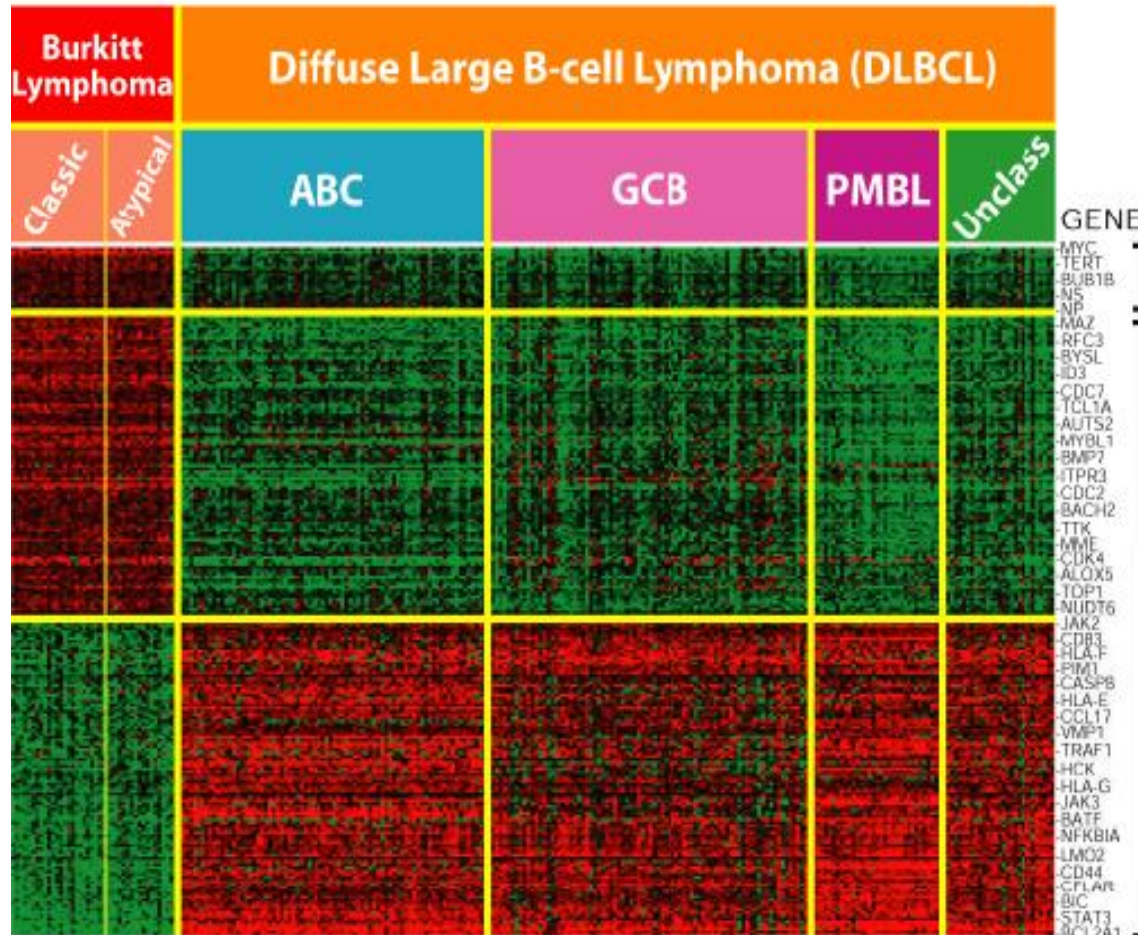
Fold
Relative
Expression 0.25 0.50 1.00 2.00 4.00



To be or not to be ...



Gene Expression Patterns in Burkitt Lymphoma



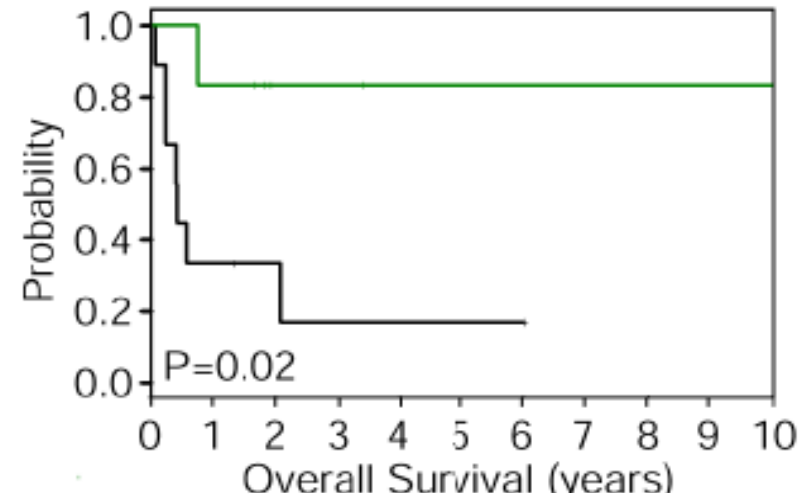
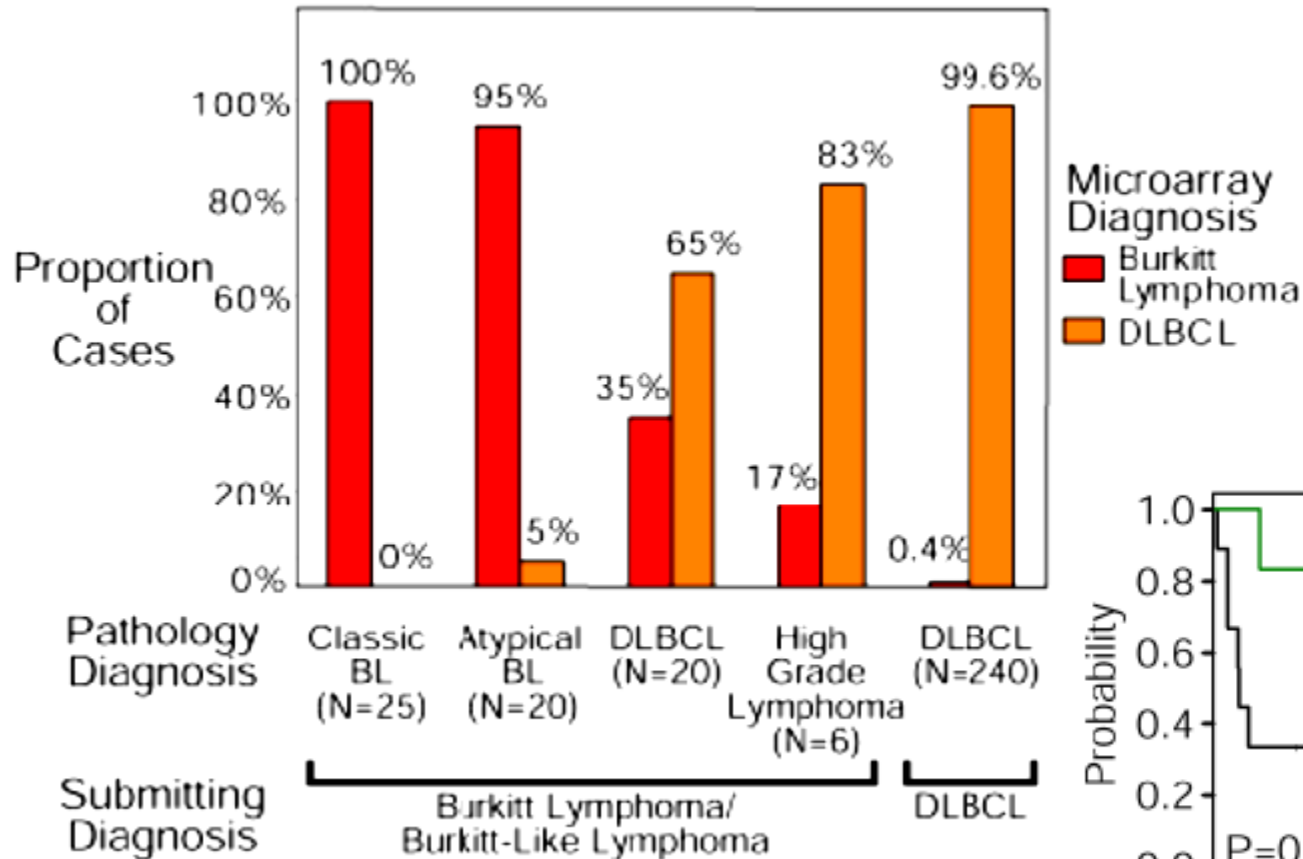
0.25 1 4
Relative Expression

Step 1: Comparison for 21 MYC target genes as defined by MYC transfection experiment in DLBCL

Step 2: Separate comparisons of BL with DLBCL-ABC, GCB and primary mediastinal B-cell lymphoma (PMBL)

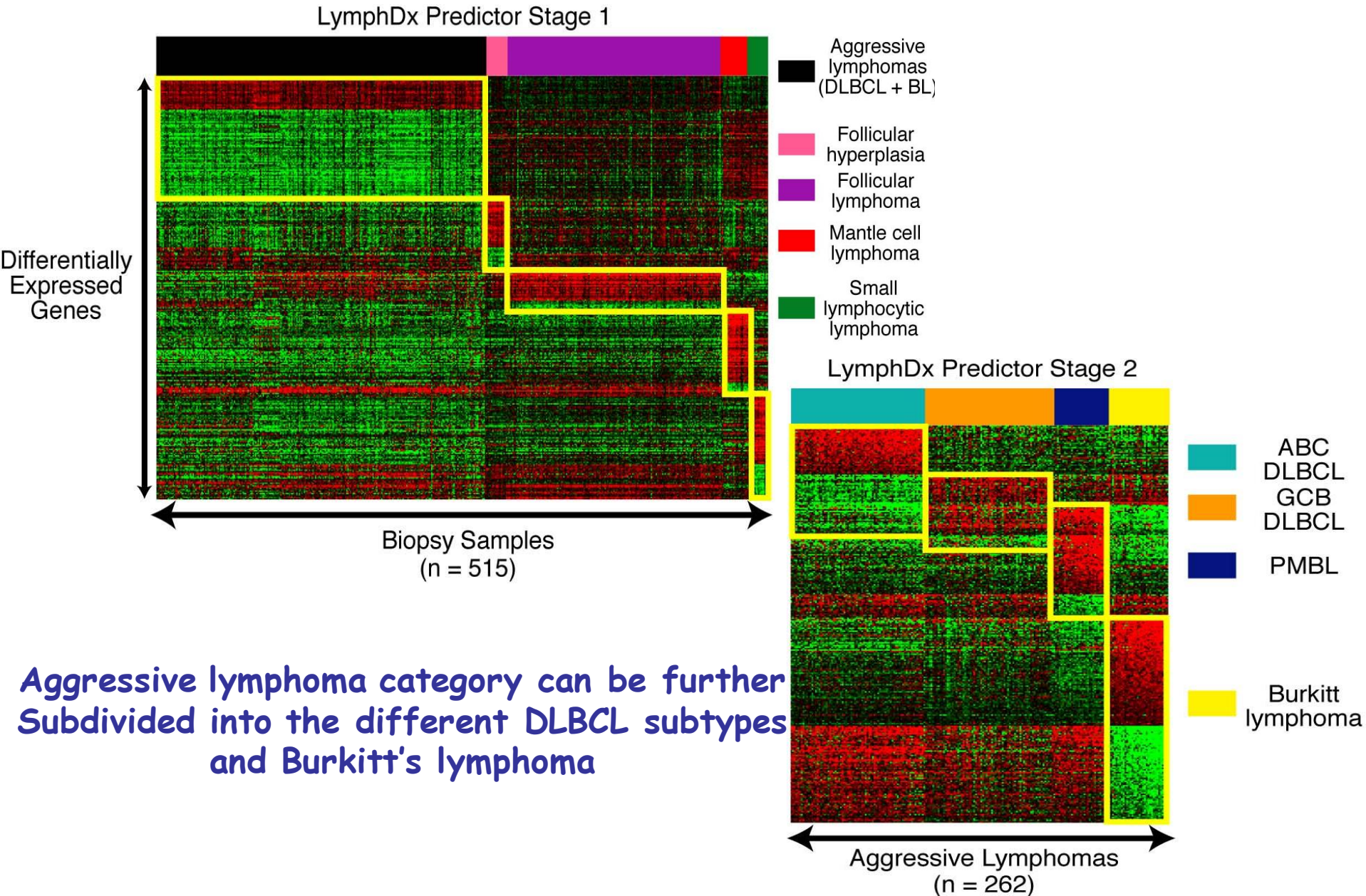
Step 3: Construction of BL predictor; also quantitative predictor (0-100%)

Molecular versus Pathology Diagnosis



Intensive Regimens
CHOP-Like Regimens

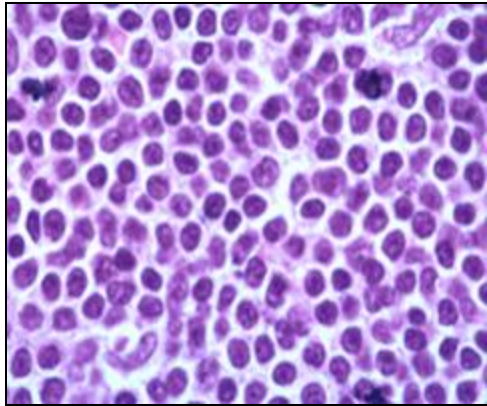
Gene expression profiles of the major classes of B-cell NHL



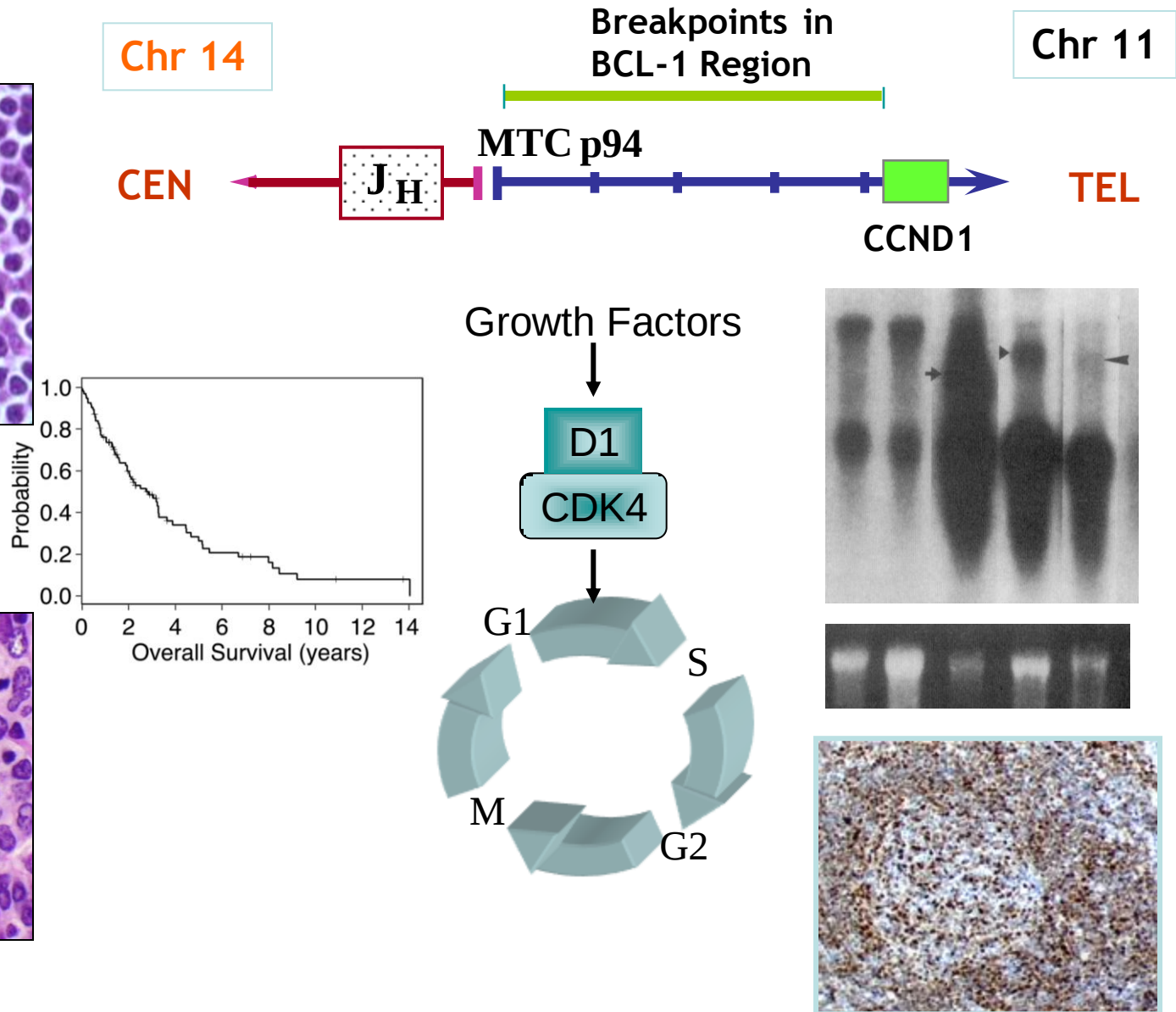
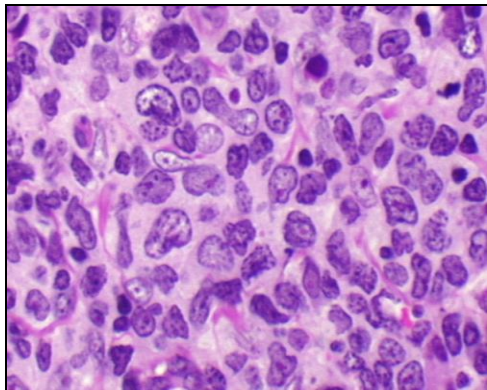
Mantle Cell Lymphoma

$t(11;14)$ Translocation and Cyclin D1

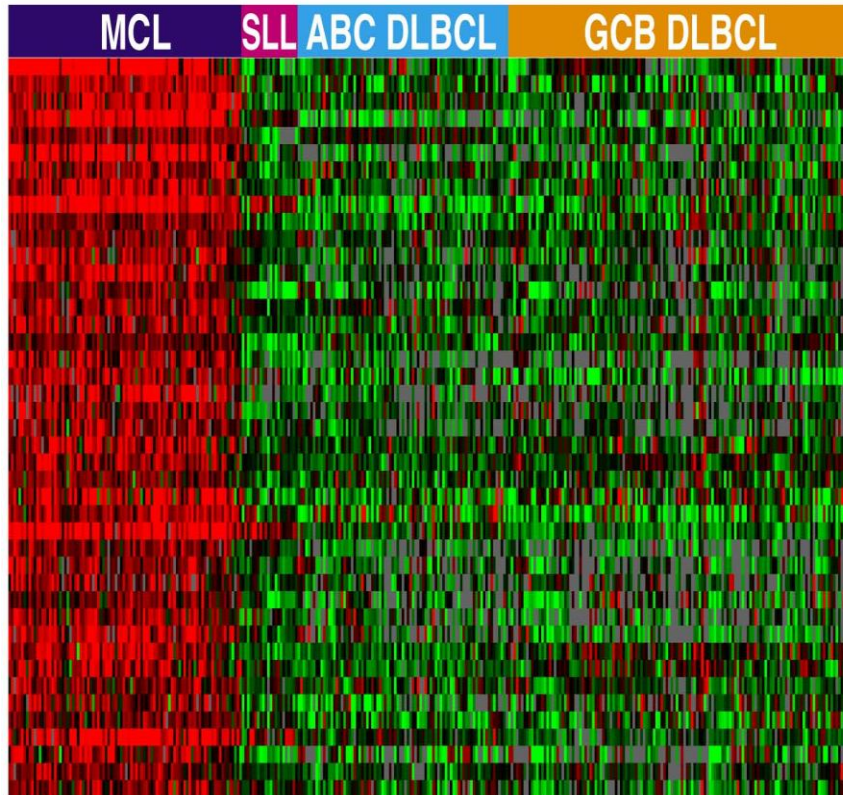
Classical



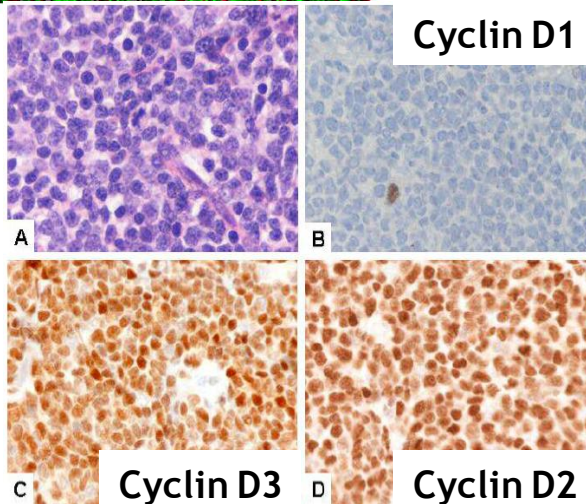
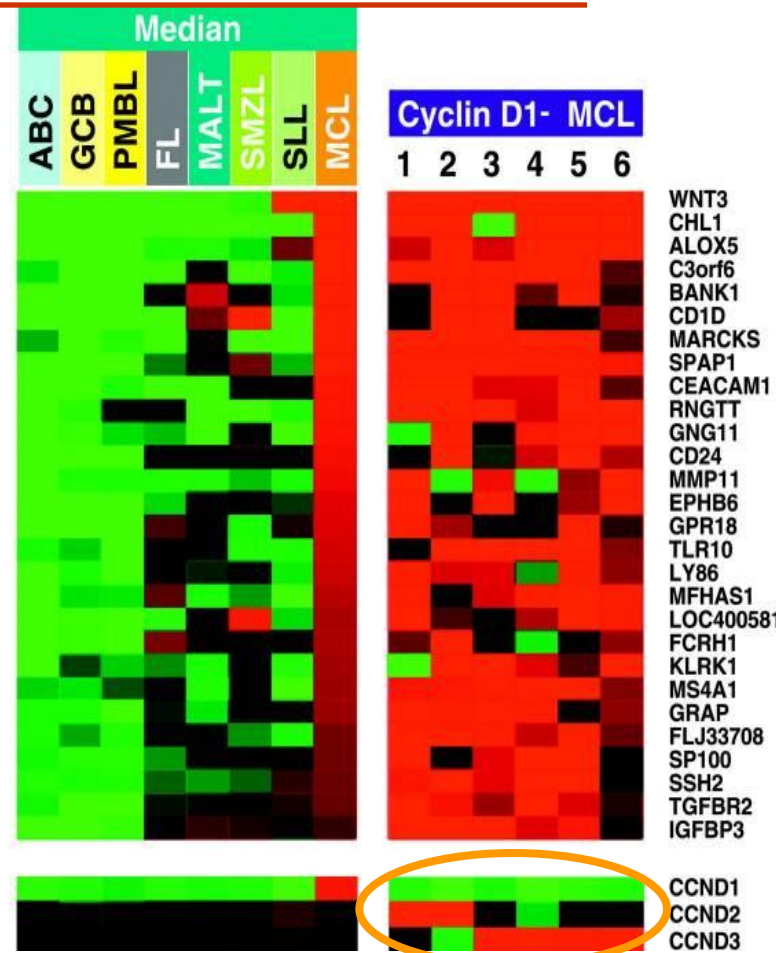
Blastic



Cyclin D1 Negative MCL Variant



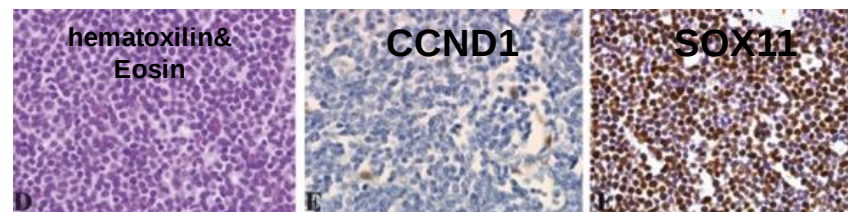
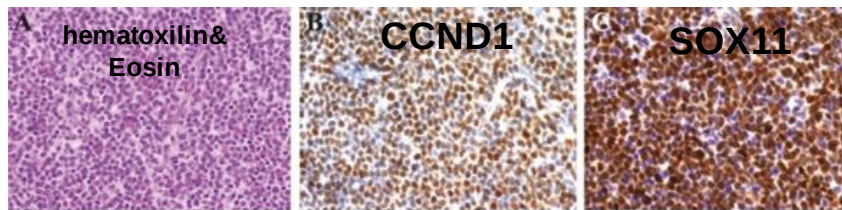
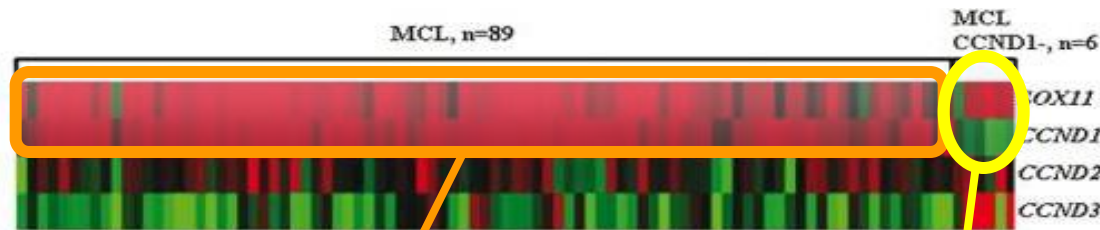
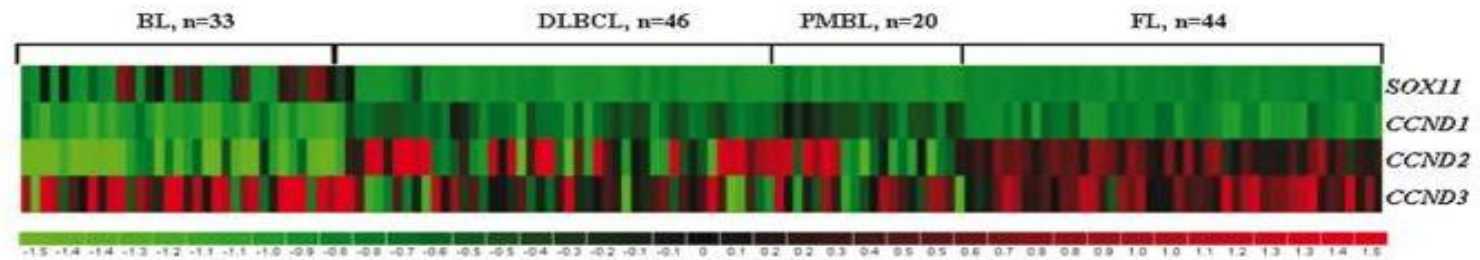
High
Relative Level of Expression
Low



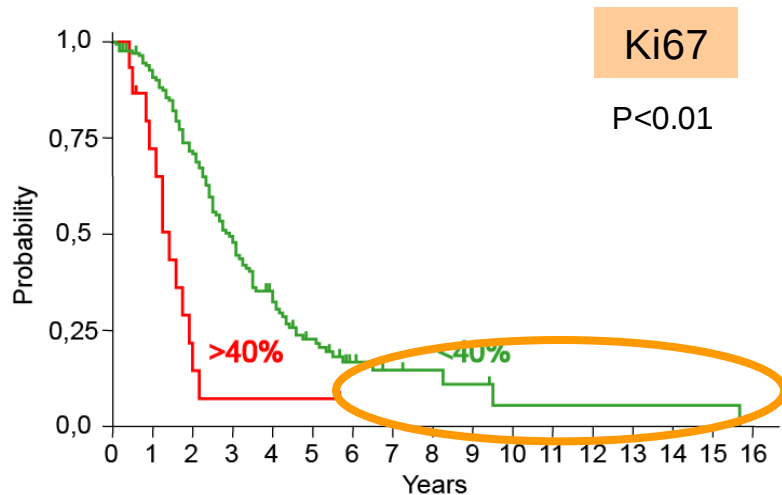
BLOOD, 15 DECEMBER 2005 • VOLUME 108, NUMBER 13

Fu K et al, Blood 2005

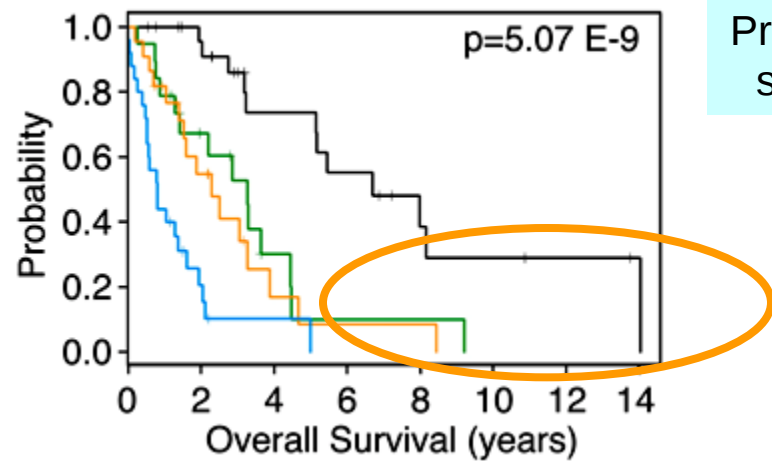
SOX11 expression is highly specific for MCL and identifies Cyclin D1-negative subtype



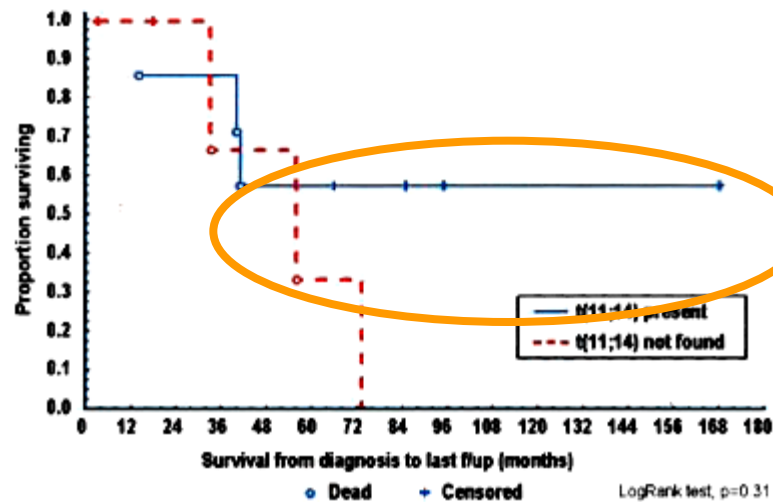
MCL prognosis



Tiemman et EMCL, Br J Haematol 2005; 131(1):29-38

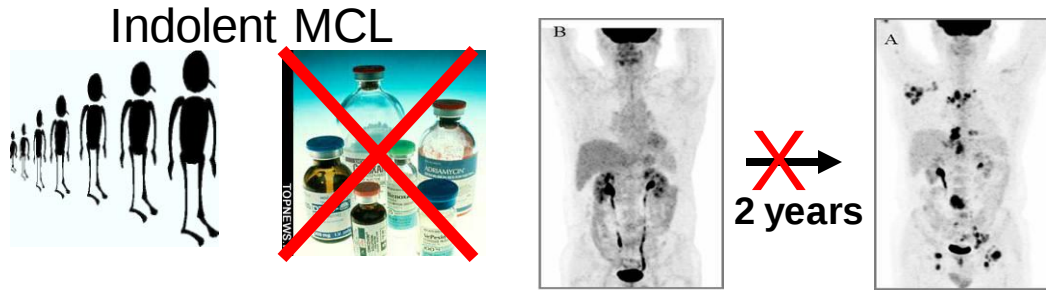


Rosenwald A et al, Cancer Cell 2003, Feb;3(2):185-97

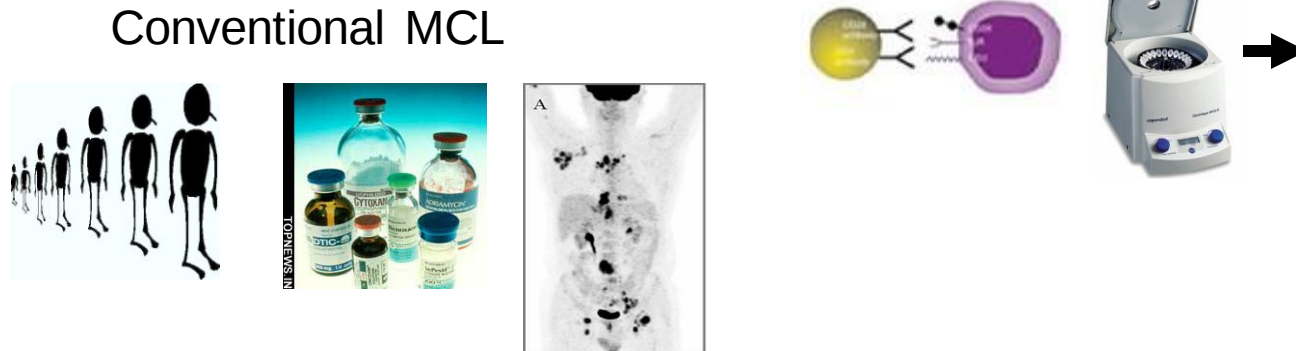


Ruchlemer R et al, Br J Haematol 2004

Study design



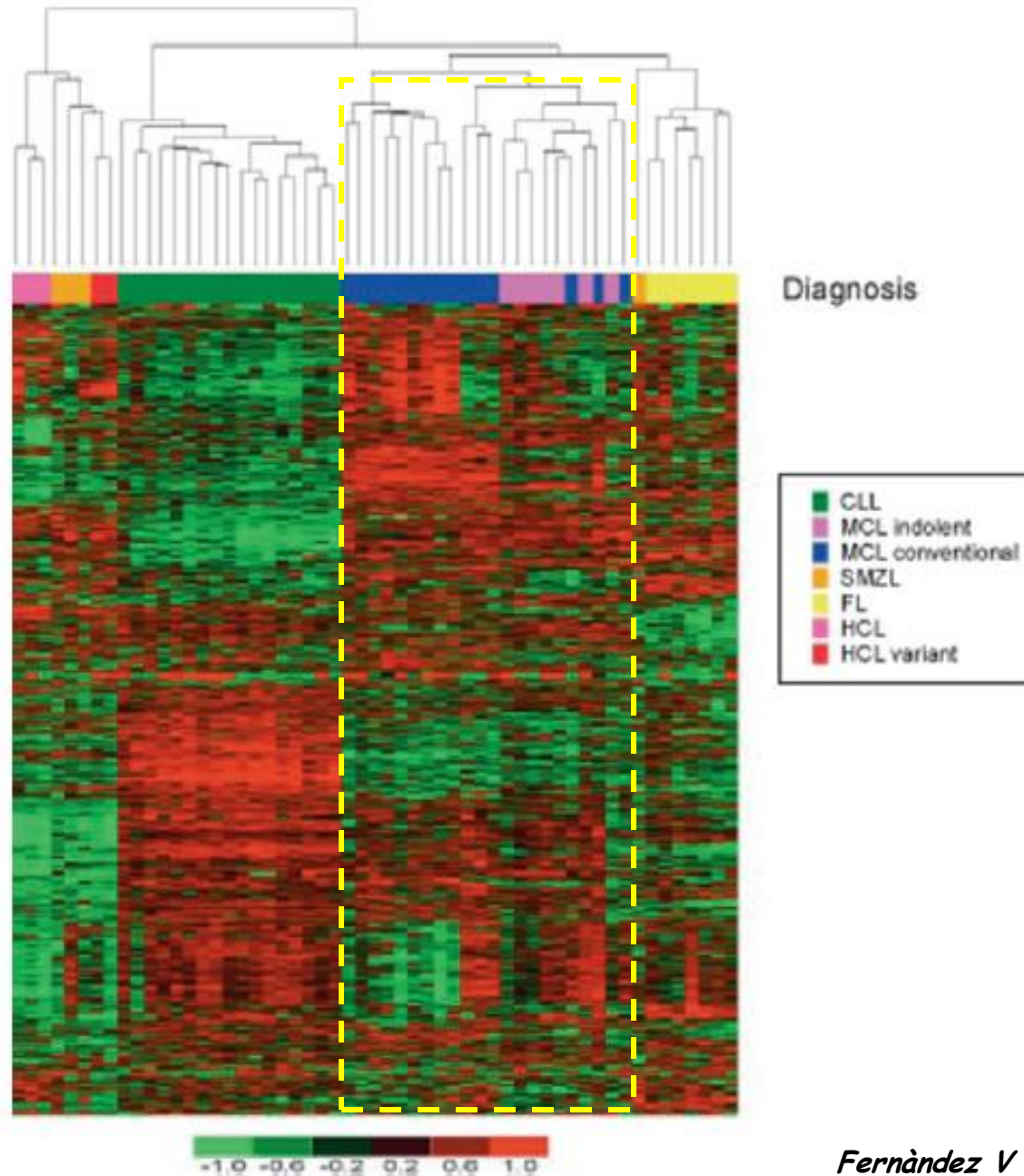
12 patients with indolent MCL (iMCL) not treated with chemotherapy and did not have evidence of clinical progression during > 2 years.
 Detection t(11;14) and overexpression CCND1
 Initial diagnosis: SMZL (4), CLL (2),
 Leukemic Lymphoid neoplasm (NOS) (4), "in situ" MCL (2)



15 patients with conventional MCL (cMCL) that required chemotherapy at diagnosis

Generalized Polyadenopathy; ECOG > 2
 Evidence of disease progression at diagnosis
 Median follow-up 15 months, range 0.3-79

iMCL & cMCL clusterize together

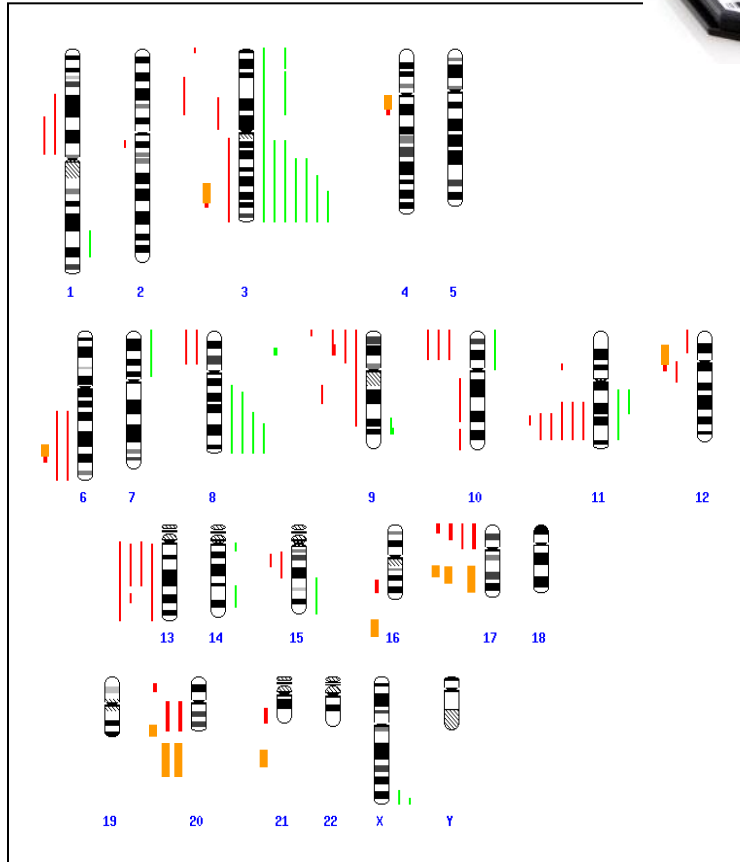


cMCL & iMCL Chromosomal Alterations by SNP Arrays

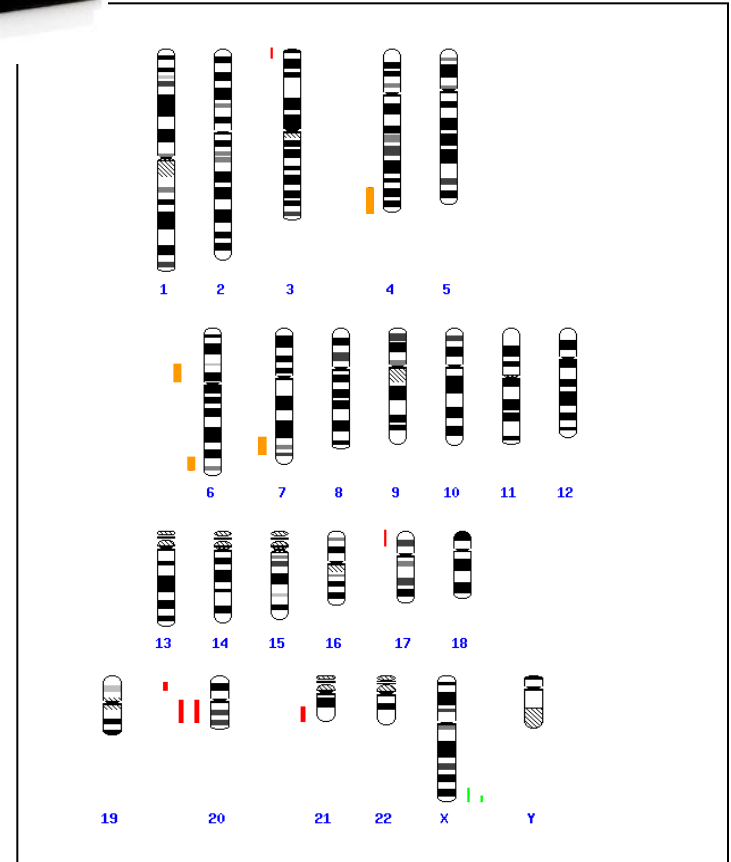
SNP 6.0



cMCL

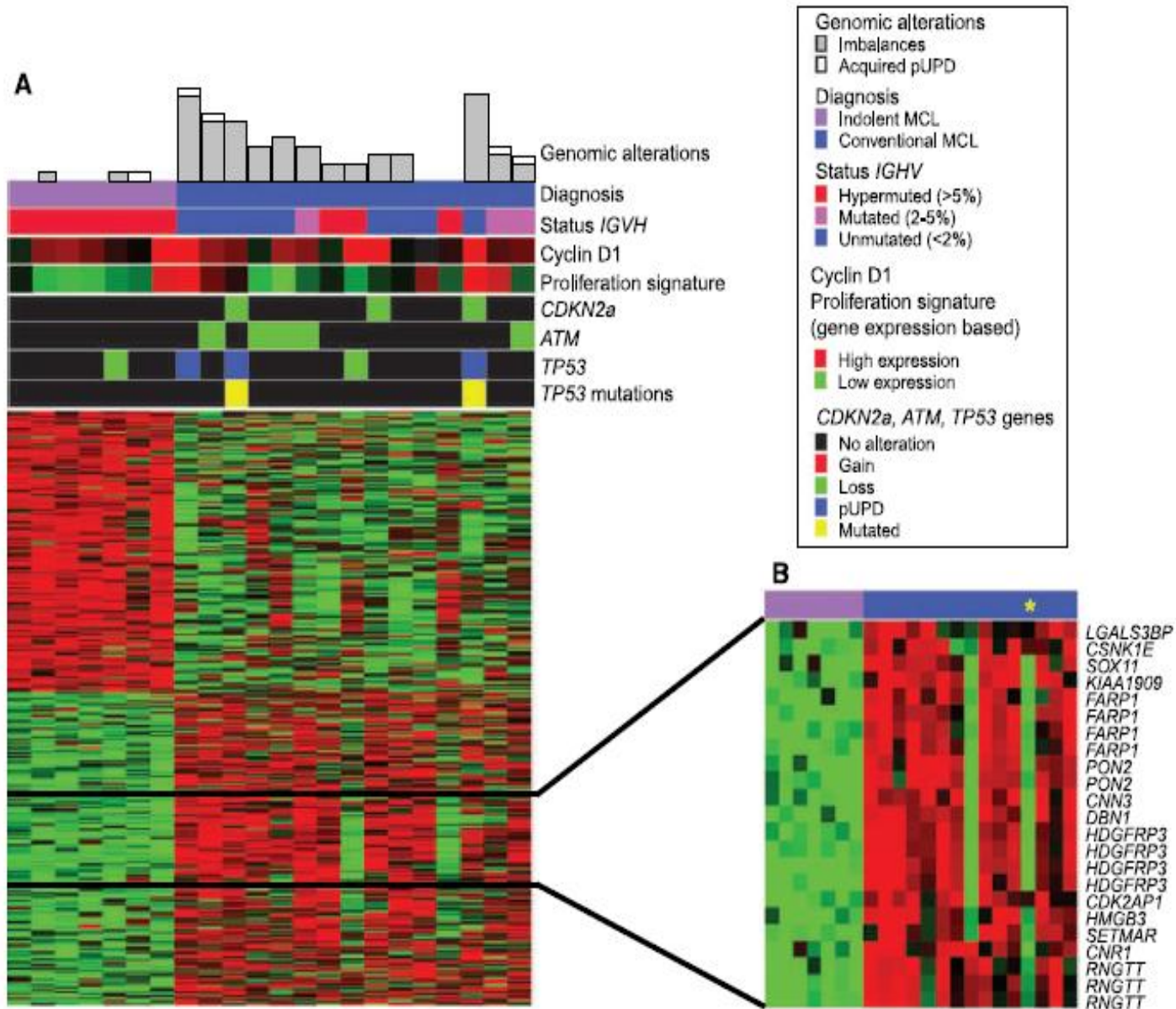


iMCL



Gain
Loss
pUPD

Gene Expression Profiling: conventional vs indolent MCL



SOX11 expression in a independent validation series

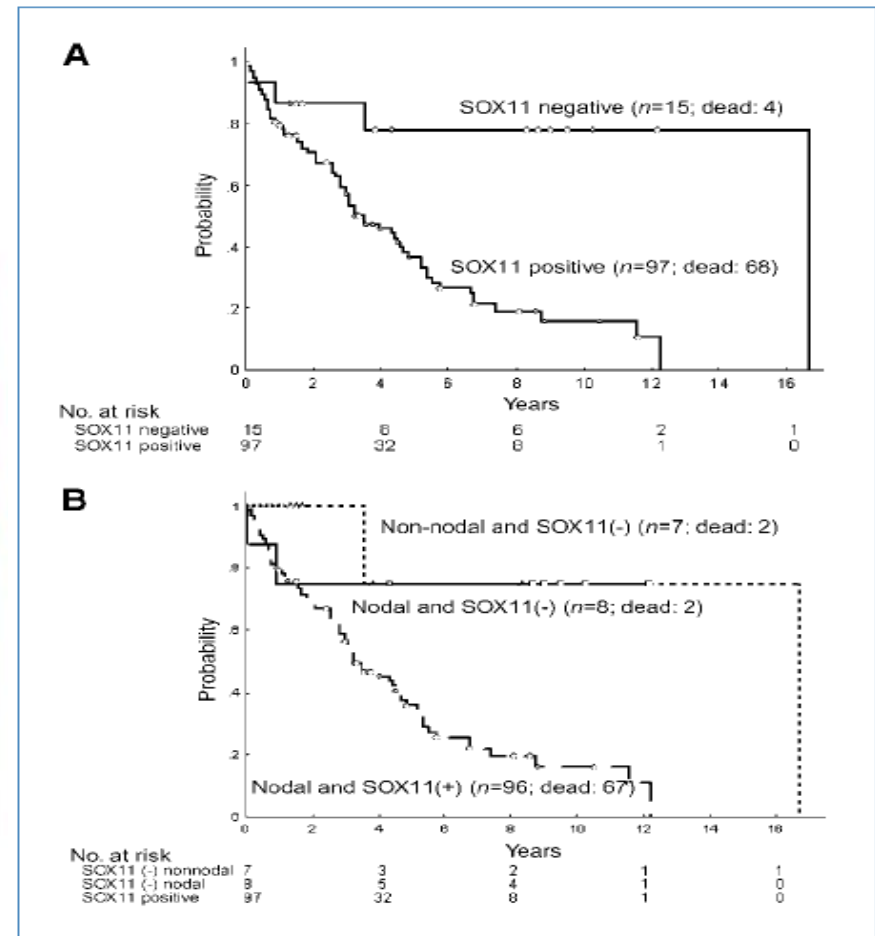
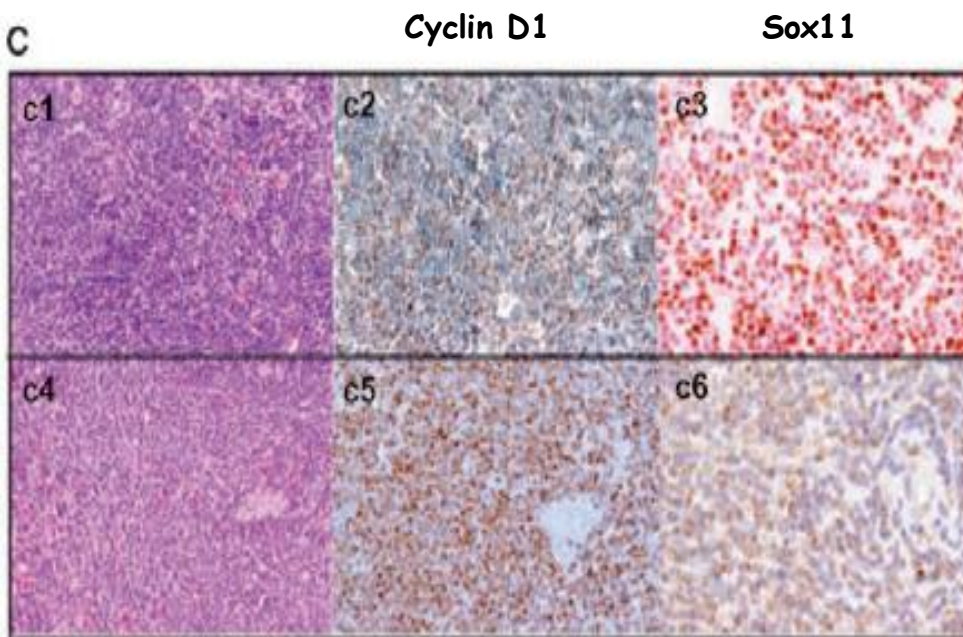


Figure 4. A, Kaplan-Meier estimates of OS of the 112 MCLs in the validation series according to SOX11 expression ($P < 0.001$). B, OS of 112 patients with MCL according to the nodal/nonnodal presentation and SOX11 expression ($P = 0.05$). The only patient with nonnodal presentation and SOX11-positive was not included.

Next Generation Sequencing

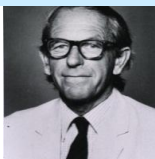


Human Genome 10th Anniversary

Chromosome

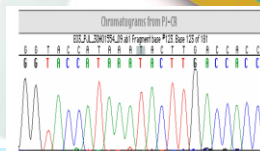
1953
DNA
double
helix
Watson
Crick
Franklin

1975.
Frederick Sanger
Sequencing method



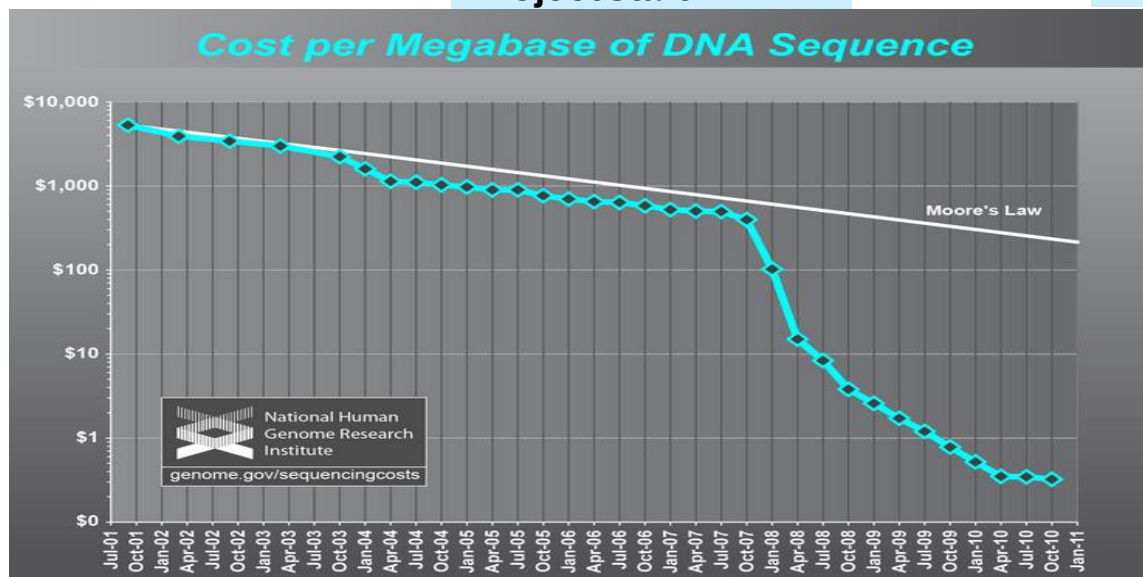
1986.
First automated
Sequencer by Applied

1995.
Capillary electrophoresis
Sequencer by Applied

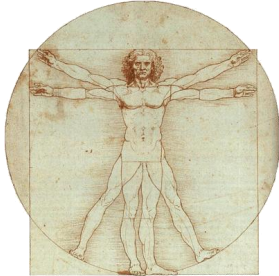


1990.
The Human Genome
Project start

2003.
The Human Genome
Project finish



Sequencing Technologies Evolution



2004 Secuenciación Sanger

300 x10⁶ \$

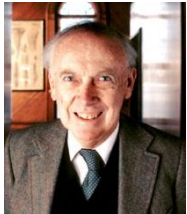
> 10 años



2007 Secuenciación Sanger

100 x10⁶ \$

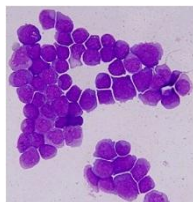
2-4 años



2008 Next-Generation 454

1 x10⁶

2 meses



2008 Next-Gen Solexa

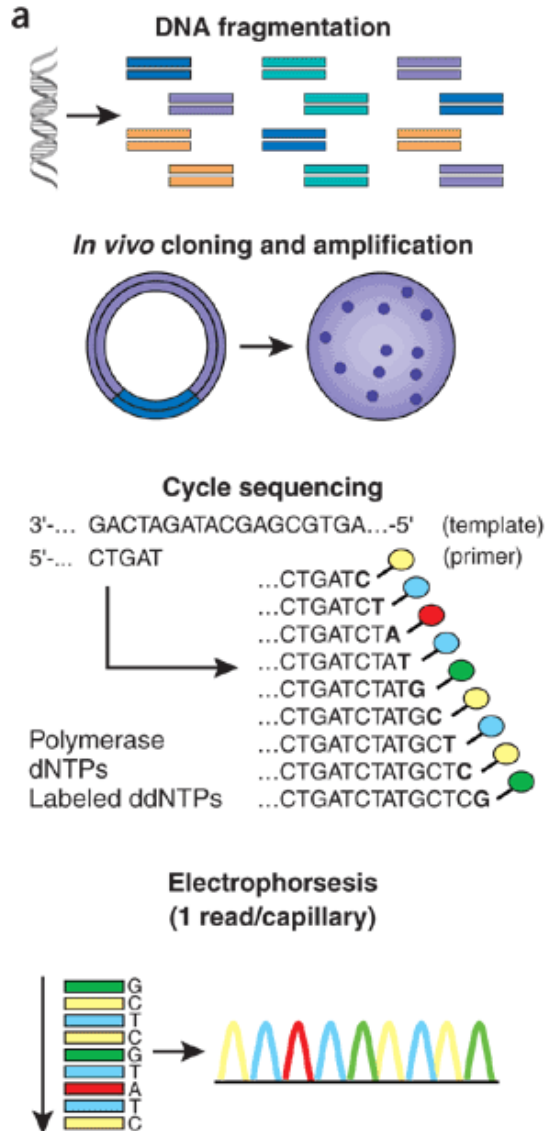
0.25 x10⁶

2 meses

Sanger Sequencing

About 2-3 weeks

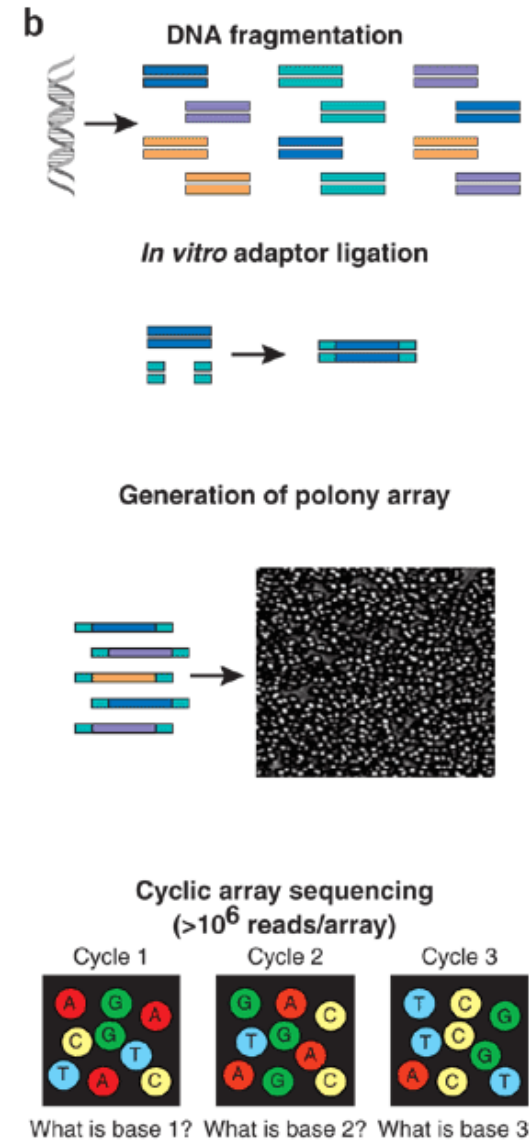
2 days



1 day

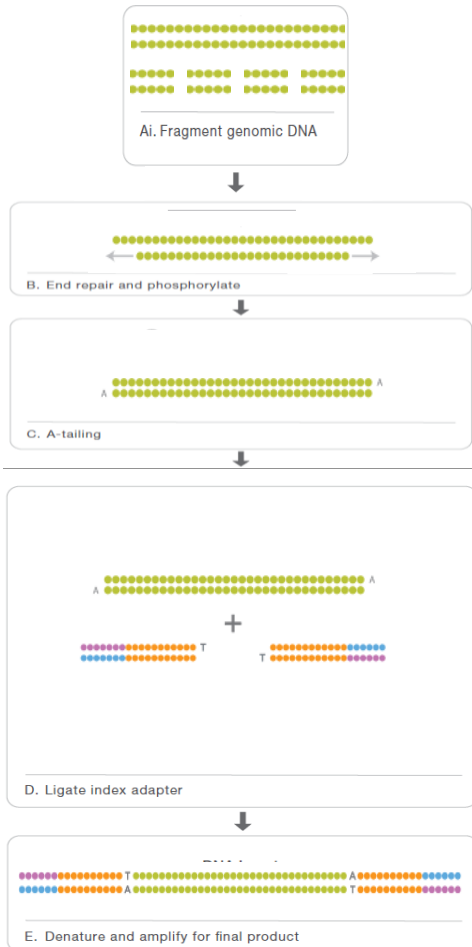
5-7 days

NGS



Next-Generation Sequencing Technology

Sample Preparation Library Generation



Target enrichment*

Cluster generation

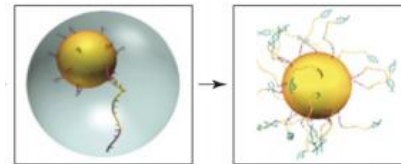
Sequencing

1. DNA Immobilization

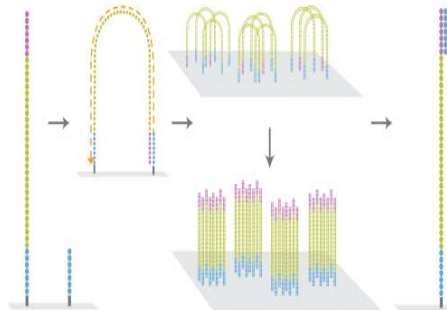
Beads
Surface

2. Amplification

- Emulsion PCR

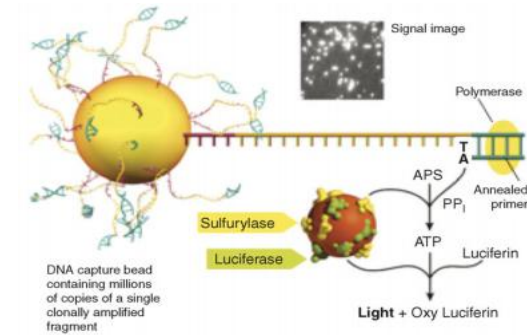


- Bridge Amplification

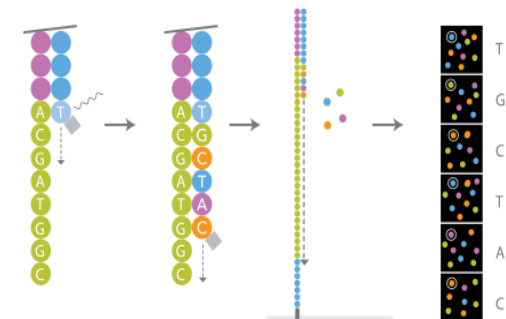


Sequencing Chemistry (SBS)

- Ligation-Based sequencing
- Polymerase-Pyrosequencing

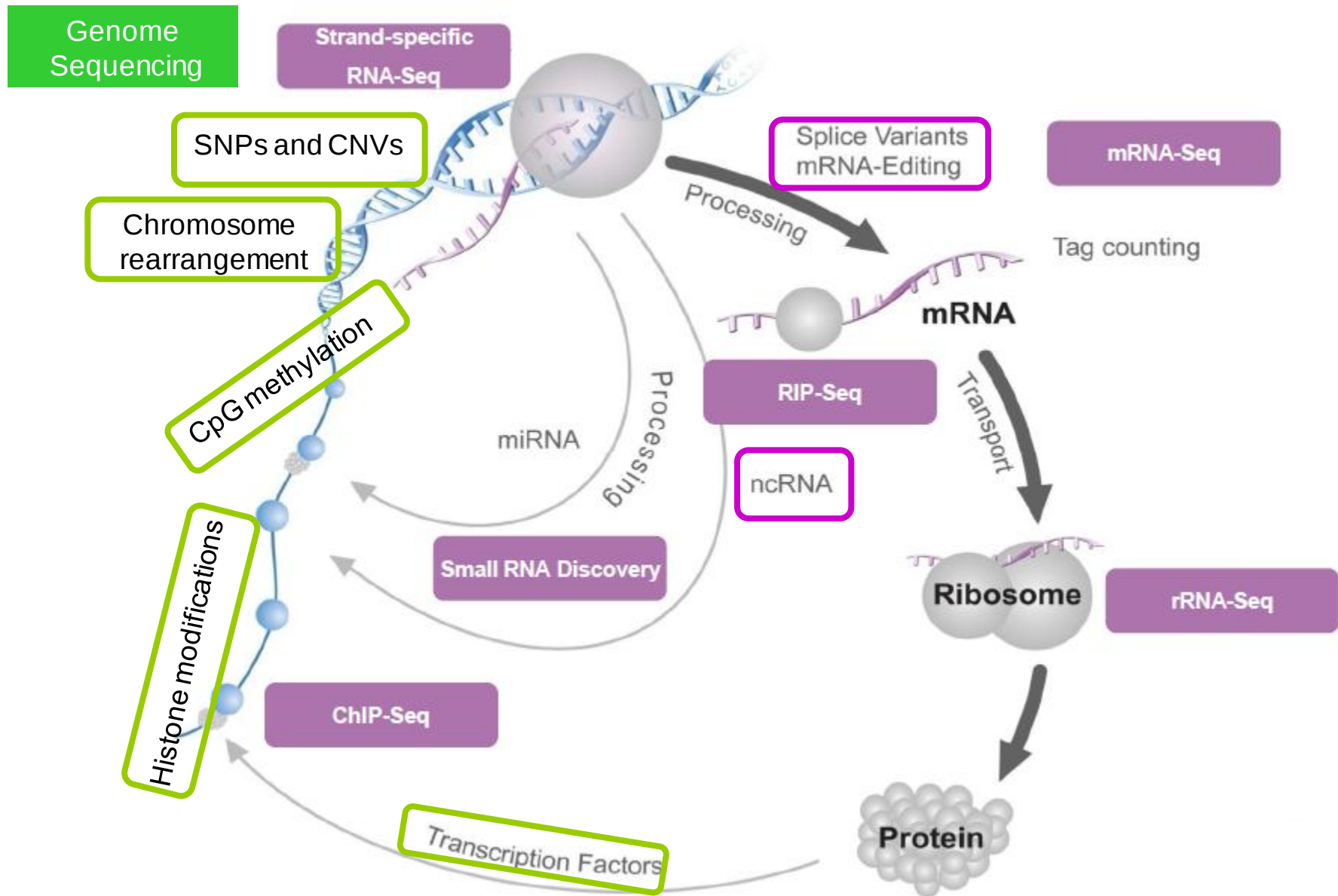


- Polymerase-Reversible terminators



Single-end / Paired-end

NGS Applications: Genomics & Transcriptomics

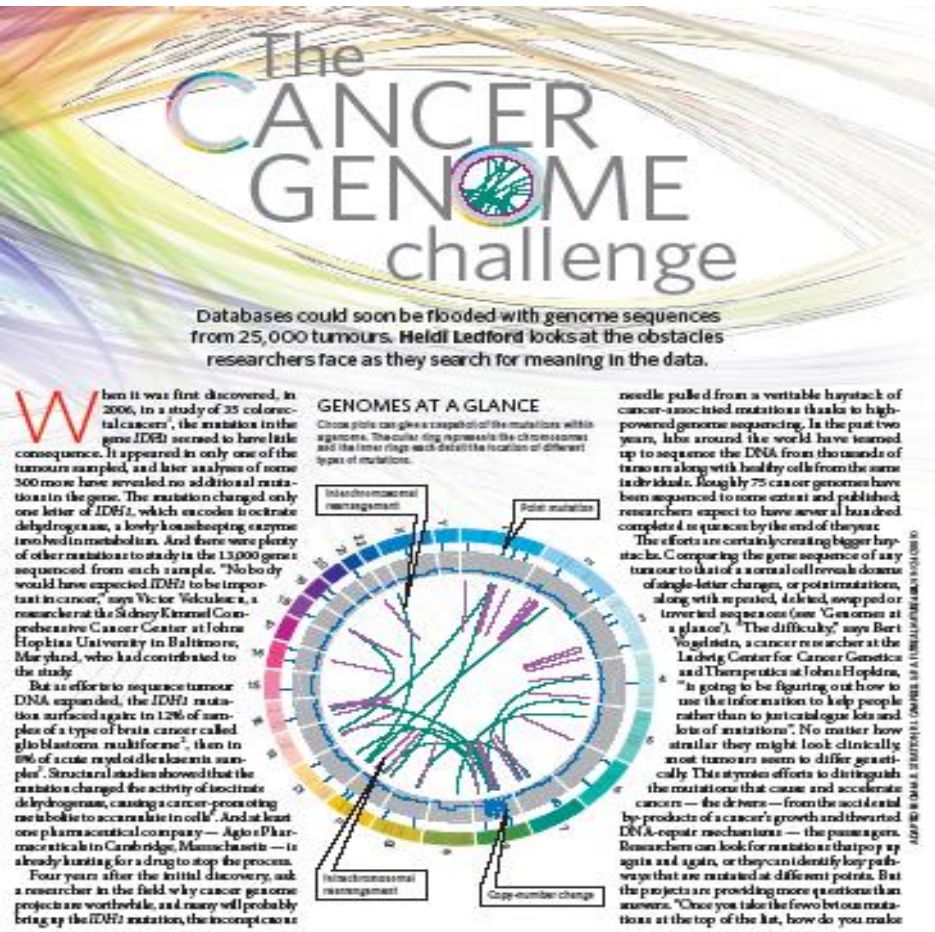


PERSPECTIVES

International network of cancer genome projects

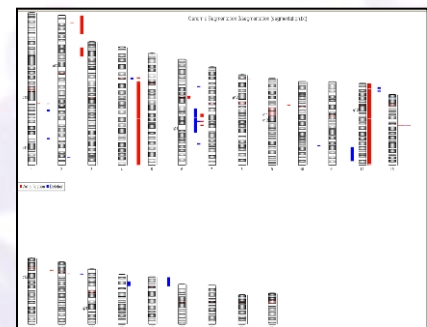
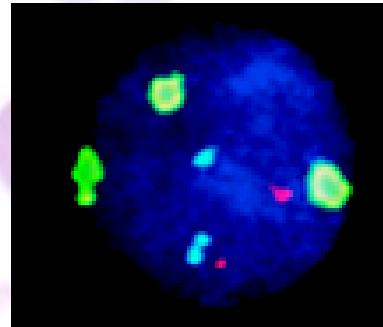
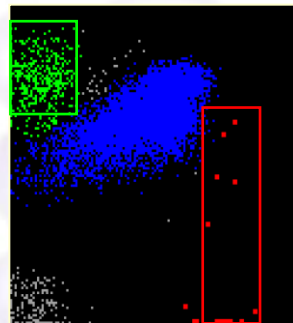
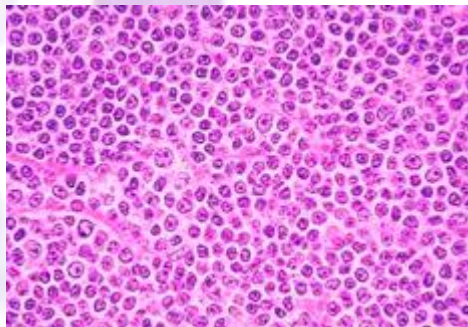
The International Cancer Genome Consortium*

The International Cancer Genome Consortium (ICGC) was launched to coordinate large-scale cancer genome studies in tumours from 50 different cancer types and/or subtypes that are of clinical and societal importance across the globe. Systematic studies of more than 25,000 cancer genomes at the genomic, epigenomic and transcriptomic levels will reveal the repertoire of oncogenic mutations, uncover traces of the mutagenic influences, define clinically relevant subtypes for prognosis and therapeutic management, and enable the development of new cancer therapies.



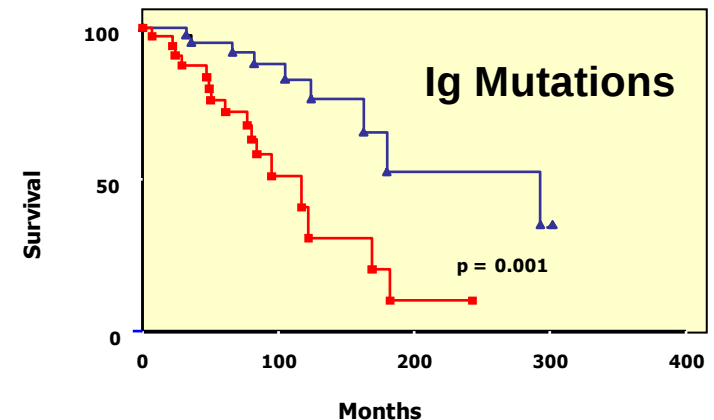
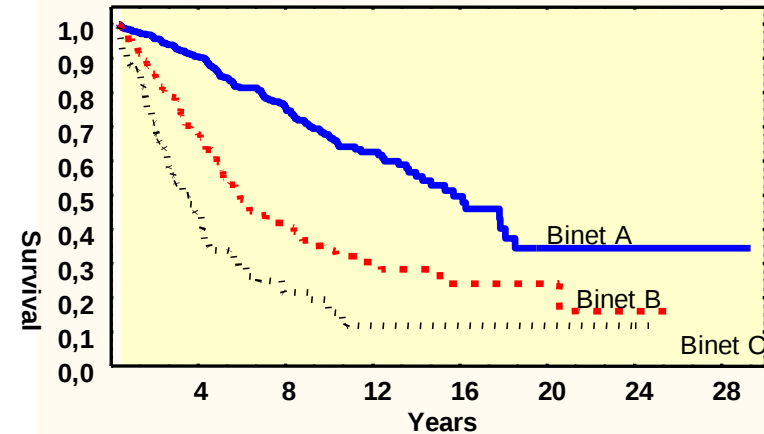
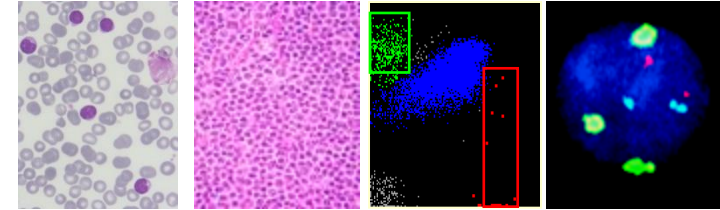
- Identification of the genetic alterations present in 50 different cancer types
- At least 500 patients for each tumor
- Uncover traces of mutagenic influences
- New targets for diagnosis and treatment
- New therapies based on genomic alterations identified in tumors

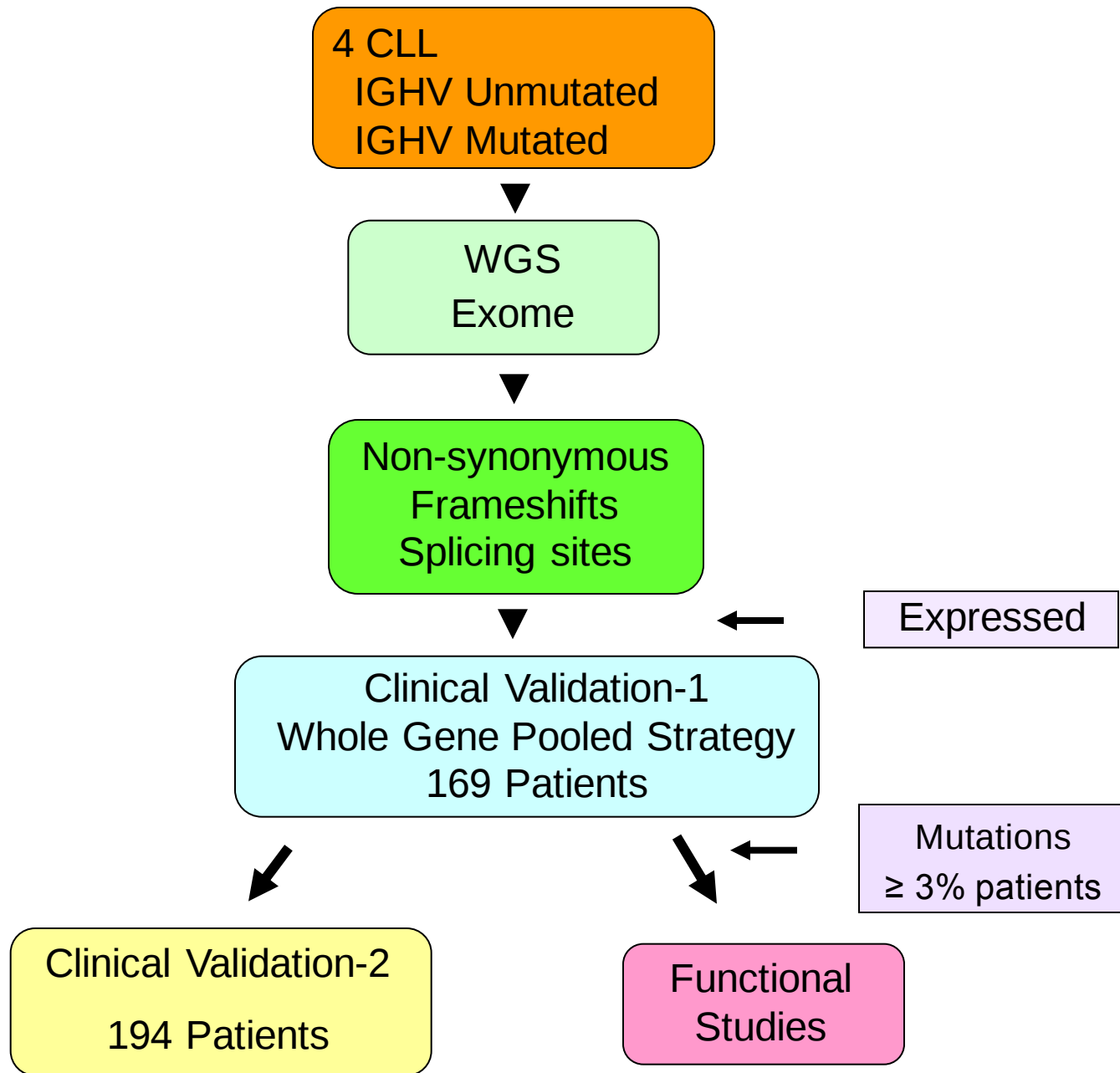
Proyecto Genoma Leucemia Linfática Crónica



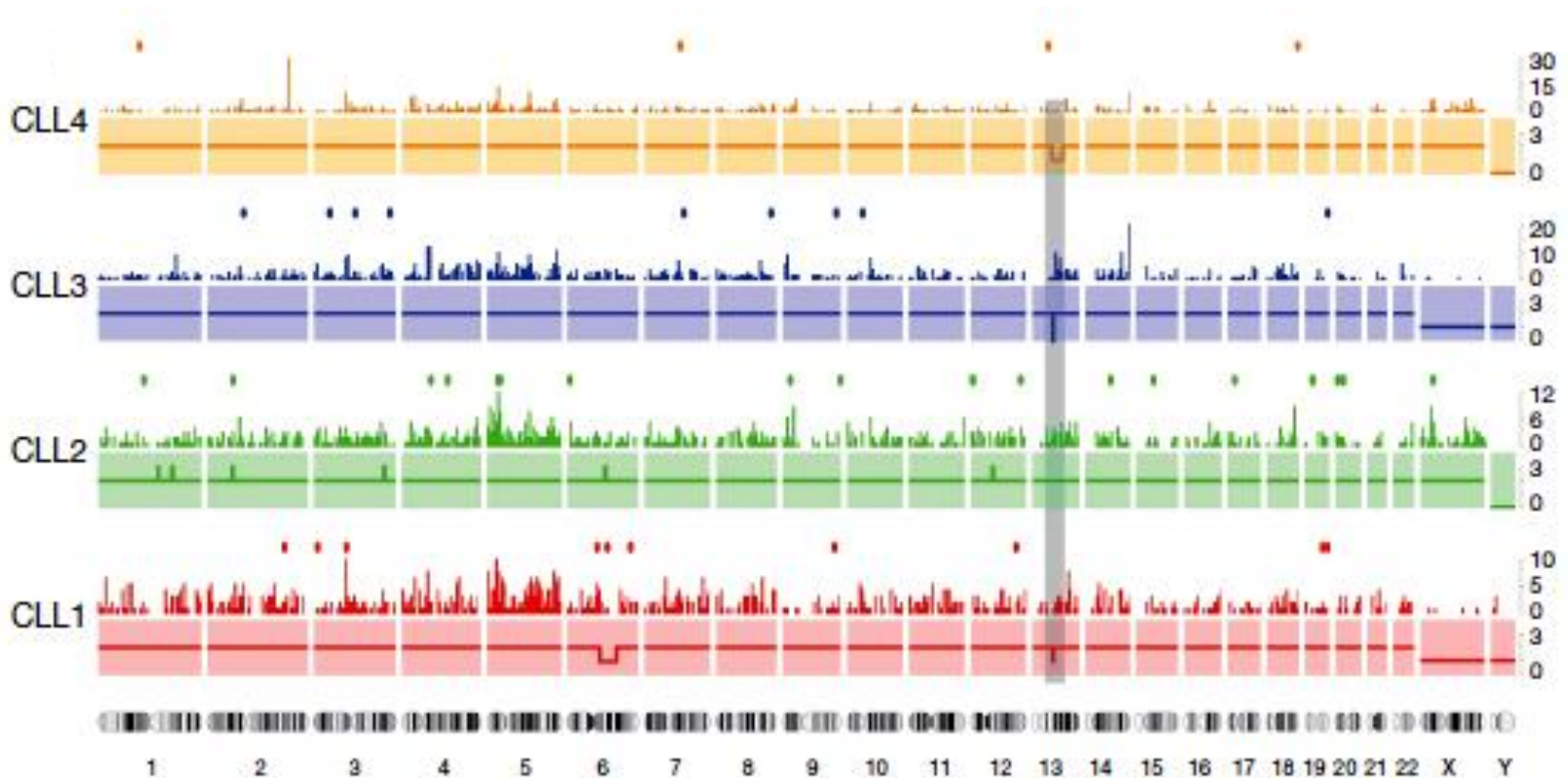
Chronic Lymphocytic Leukemia

- Most frequent leukemia in Western countries (5-7 cases /100,000 /year)
- Heterogeneous disease with different stages of progression and molecular subtypes
- No effective therapy
- Pathogenesis
 - Unknown initiating genetic alterations
 - Microenvironment
 - Evidence of genetic predisposition
 - Geographic distribution
 - Familiar clustering
 - Susceptibility loci
 - Environmental influences?





Profile of Somatic Mutations in Four CLL Genomes



Solid lines: copy number; **Bars:** mutation density per 5 Mb
Dots: class 1 mutations (non-syn., frameshifts, splicing sites)

Puente et al, Nature 2011

Genes recurrently mutated in CLL

Gene	Mutation	Mutated cases / total	Overall frequency (%)	Frequency in <i>IGHV</i> -unmutated (%)	Frequency in <i>IGHV</i> -mutated (%)
<i>NOTCH1</i>	P2515Rfs*4 Q2503* F2482Ffs*2	29/255 1/255 1/255	12.2	20.4	7
<i>MYD88</i>	L265P	9/310	2.9	0.8	5.6
<i>XPO1</i>	E571K E571G	3/165 1/165	2.4	4.6	0
<i>KLHL6</i>	F49L/L65P L90F L58P/T64A/Q81P	3/160	1.8	0	4.5

Puente et al, Nature 2011

WGS in CLL: Summary

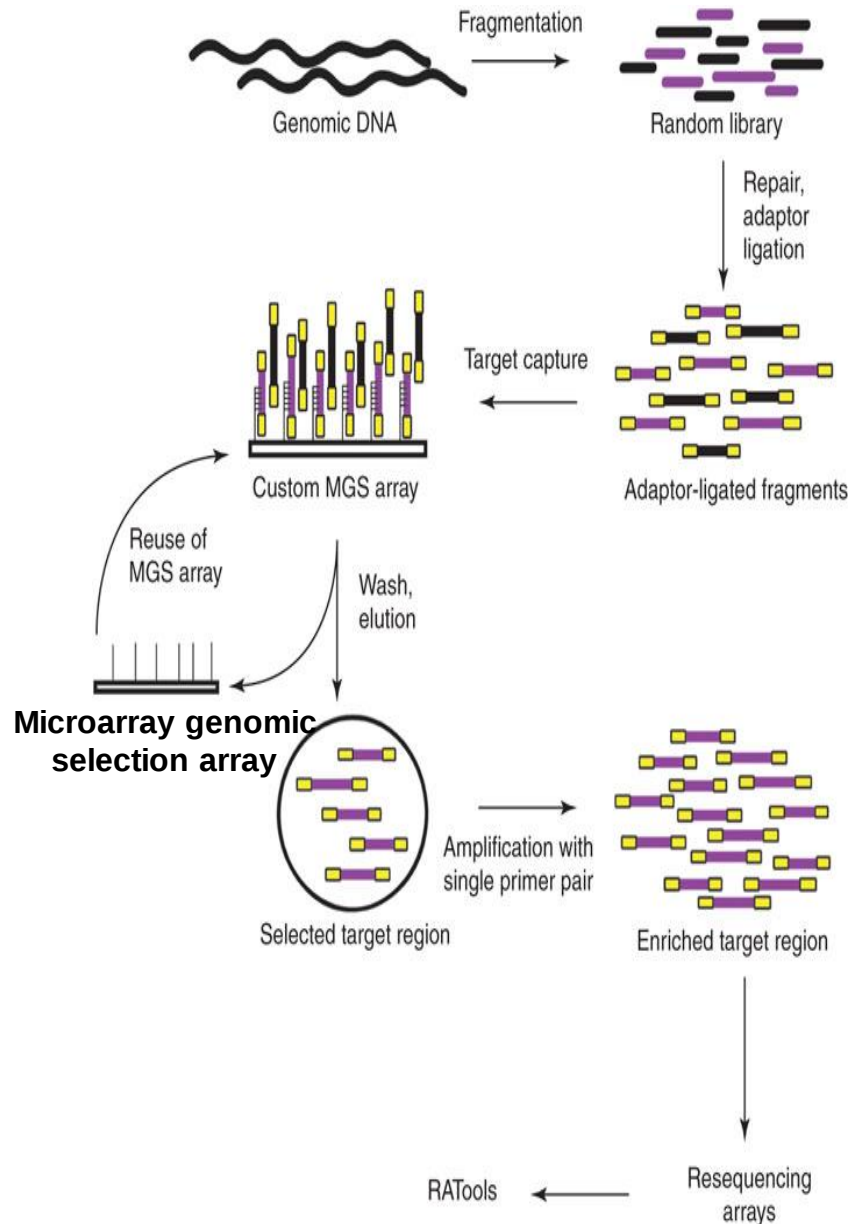
- **CLL carries approximately 1 somatic mutations /MB (1000 per case)**
- **Identification of a potential mutational mechanism in IGHV hypermutated CLL related to the activity of DNA polymerase eta**
- **Identification of 46 somatic mutations in coding regions with potential functional effect. 5-20 per case**
- **Identification of 4 recurrent mutations with clinical implication**
- **NOTCH1 mutations are a frequent event in CLL**
 - NOTCH1 mutations truncate and stabilize the protein and activate NOTCH1 pathway in CLL
 - Predominant in IGHV unmutated CLL and associated with high risk factors (ZAP70/CD38)
- **MYD88 mutations**
 - Same mutation found in ABC DLBCL
 - Activates NFkB pathway (IRAK1, STAT3)
 - Promotes a high production of several chemokines (CCL2, CCL3, CCL4, IL6, ILR1A) in response to TLR stimulation

Exome sequencing identifies recurrent mutations of the splicing factor *SF3B1* gene in chronic lymphocytic leukemia

Víctor Quesada¹, Laura Conde², Neus Villamor², Gonzalo R Ordóñez¹, Pedro Jares², Laia Bassaganyas³, Andrew J Ramsay¹, Sílvia Beà², Magda Pinyol⁴, Alejandra Martínez-Trillos⁵, Mónica López-Guerra², Dolors Colomer², Alba Navarro², Tycho Baumann⁵, Marta Aymerich², María Rozman², Julio Delgado⁵, Eva Giné⁵, Jesús M Hernández⁶, Marcos González-Díaz⁶, Diana A Puente¹, Gloria Velasco¹, José M P Freije¹, José M C Tubío³, Romina Royo⁷, Josep L Gelpí⁷, Modesto Orozco⁷, David G Pisano⁸, Jorge Zamora⁸, Miguel Vázquez⁸, Alfonso Valencia⁸, Heinz Himmelbauer⁹, Mónica Bayés¹⁰, Simon Heath¹⁰, Marta Gut¹⁰, Ivo Gut¹⁰, Xavier Estivill³, Armando López-Guillermo⁵, Xose S Puente¹, Elías Campo^{2,11} & Carlos López-Otín^{1,11}

Exome Sequencing: target enrichment

Microarray-based genomic selection



Target enrichment in solution

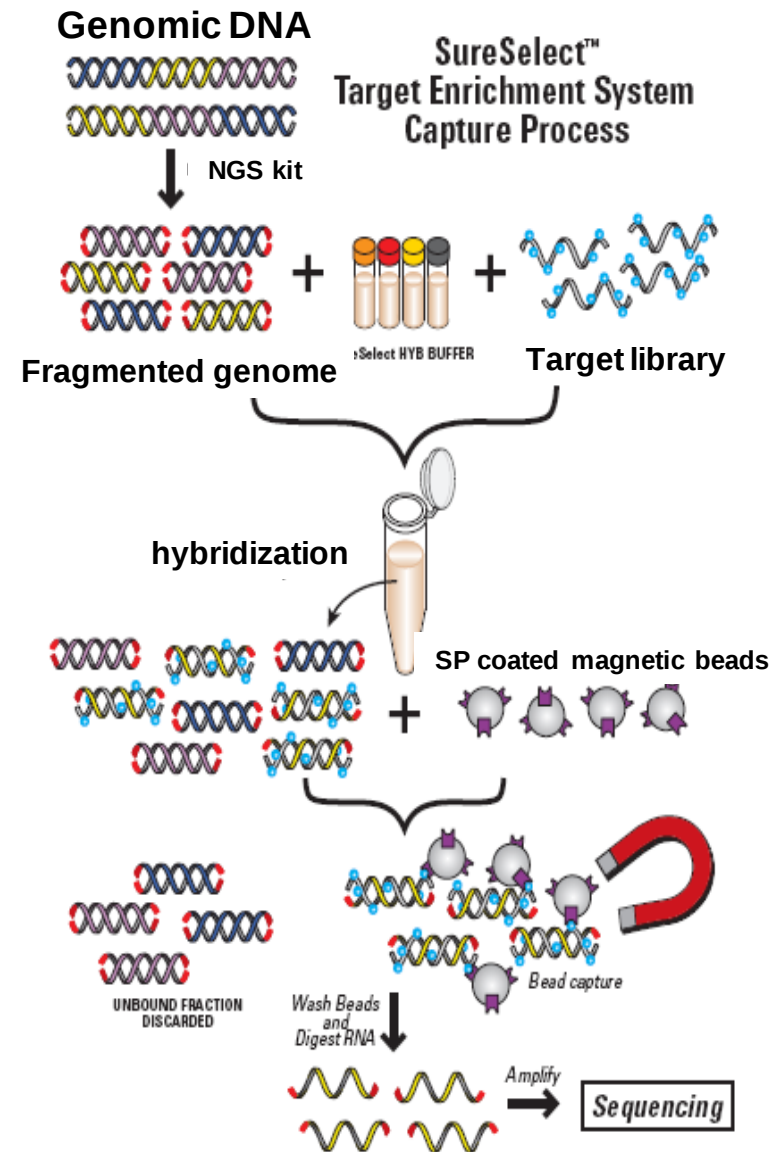
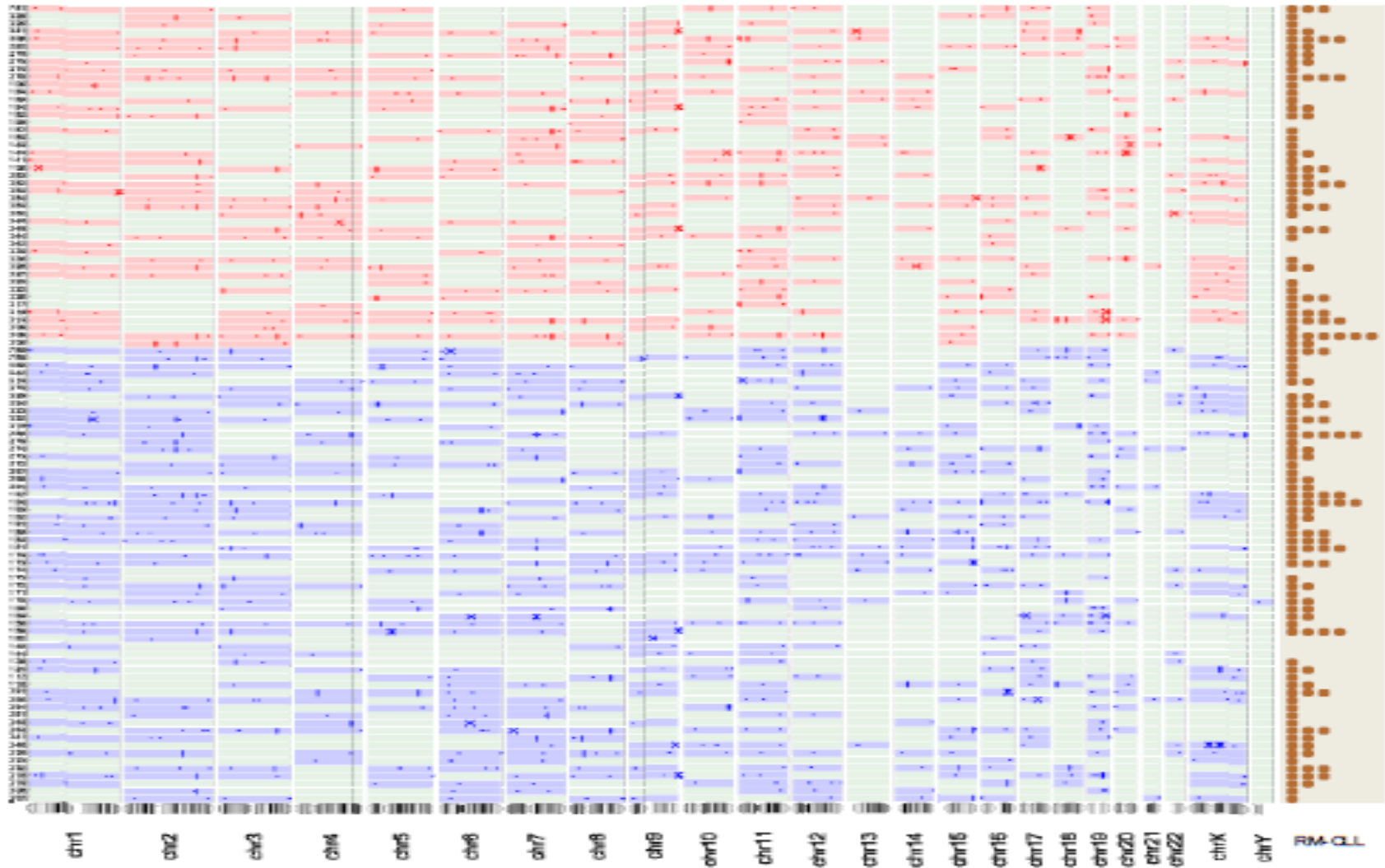


Figure 4 SureSelect Target Enrichment System Capture Process

Profile of Somatic Mutations in 105 CLL Exomes

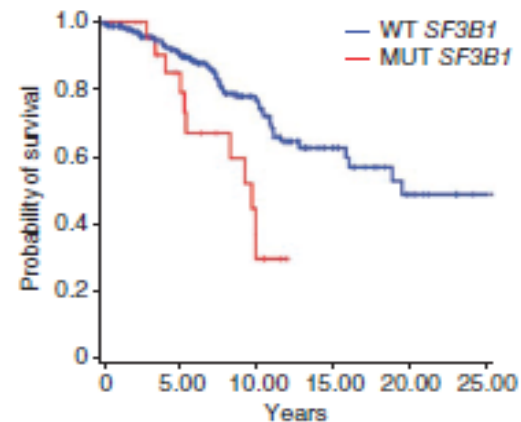
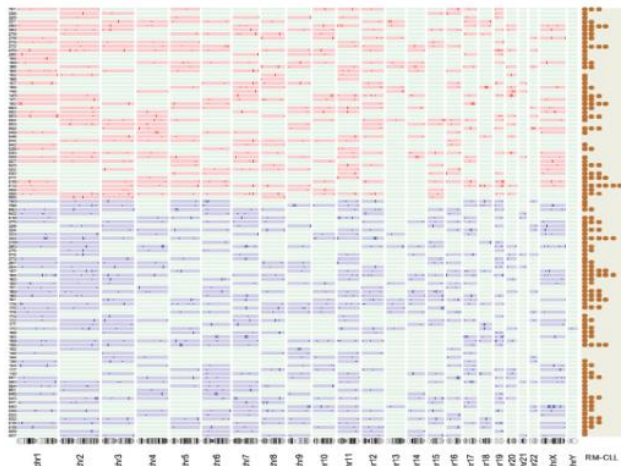


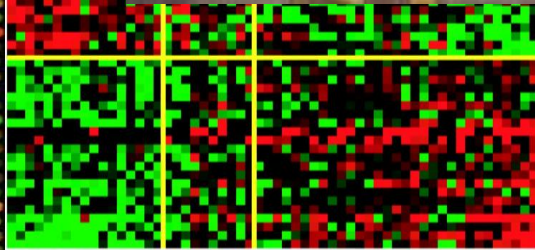
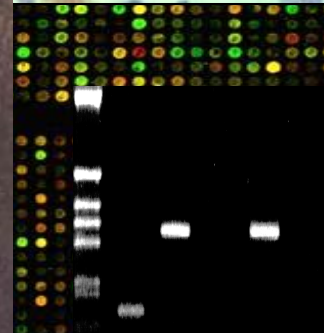
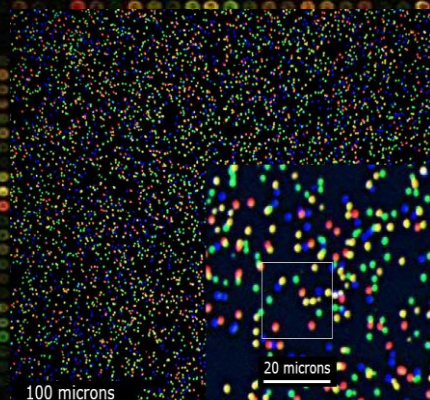
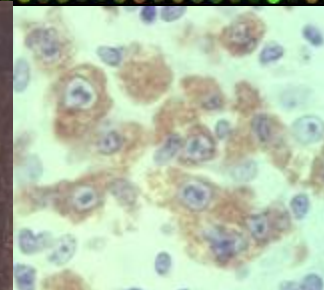
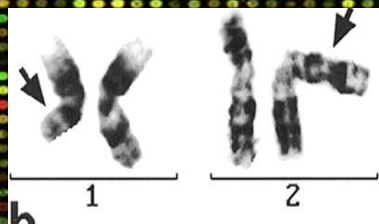
Recurrent Mutated Genes in CLL

Gene	Mutated cases / total	Overall frequency (%)	Frequency in <i>IGHV</i> -unmutated (%)	Frequency in <i>IGHV</i> -mutated (%)
<i>NOTCH1</i>	31/255	12.2	10.1	2.8
<i>SF3B1</i>	27/279	9.7	20.5	7.9
<i>POT1</i>	5/105	4.8	11.1	0
<i>CHD2</i>	5/105	4.8	0	8.3
<i>LPP1D</i>	5/105	4.8	5.0	5.0

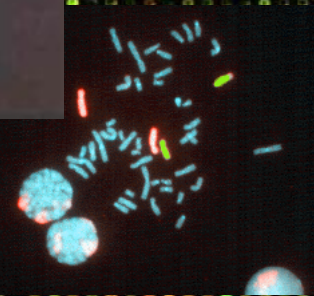
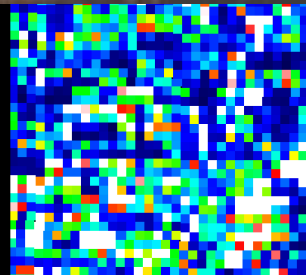
WGS in CLL: Summary

- Exome sequencing of 105 CLL patients
- Identification of 1246 somatic mutations affecting about 1100 genes.
- Mutations in 78 genes are found in more than one patient.
- The initial functional analysis showed that the genes found mutated in CLL significantly clustered in specific gene pathways including Toll-like receptor pathway, RNA splicing and processing, among others.
- The different mutations are associated to different clinical groups.
- SF3B1 mutations are identified in patients with worst prognosis.





ABC DLBCL GCB DLBCL



***Hematopathology Section
Pathology Department
Hospital Clinic,
University of Barcelona***

E Campo

S Beà

M Pinyol

L Colomo

A Martinez

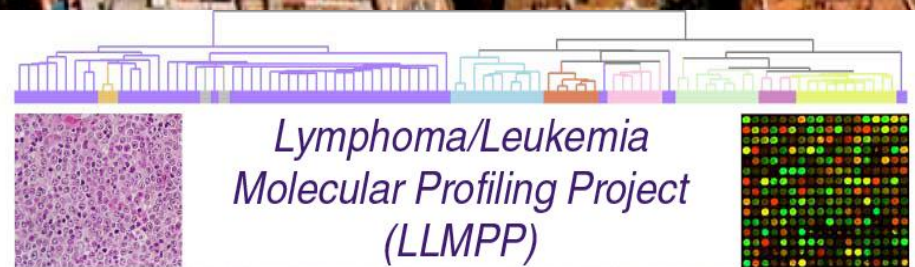
L Hernandez

V Amador

C Arroyo

A Navarro

P Jares



***Lymphoma/Leukemia Molecular
Profiling Project***

University of Nebraska, Omaha

National Cancer Institute, Bethesda, MD

South-West Oncology Group

British Columbia Cancer Agency, Vancouver

University of Wurzburg, Wurzburg

Radium Hospital, Oslo

St Bartolomew Hospital, London

Hospital Clinic, University of Barcelona

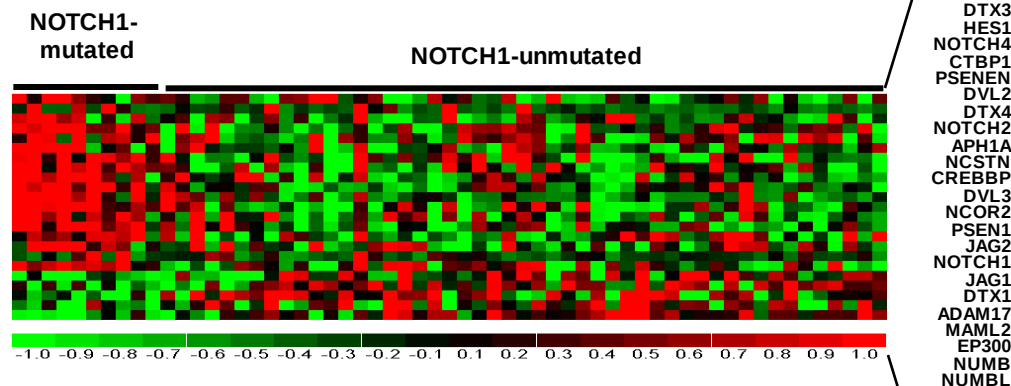
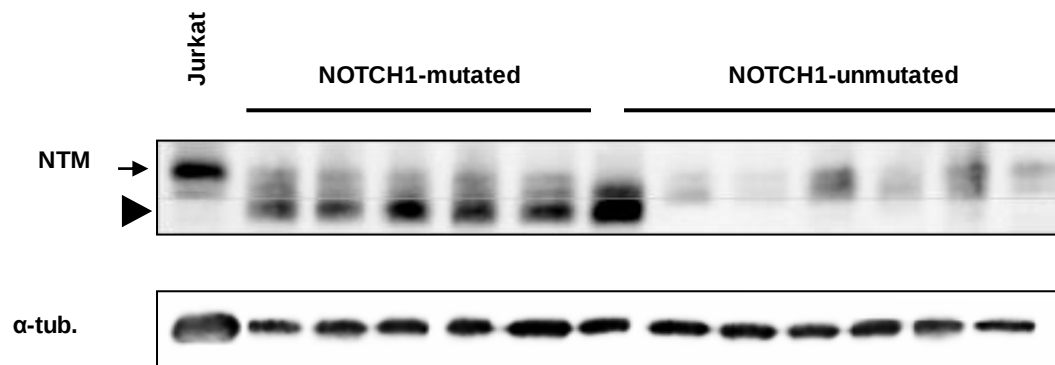
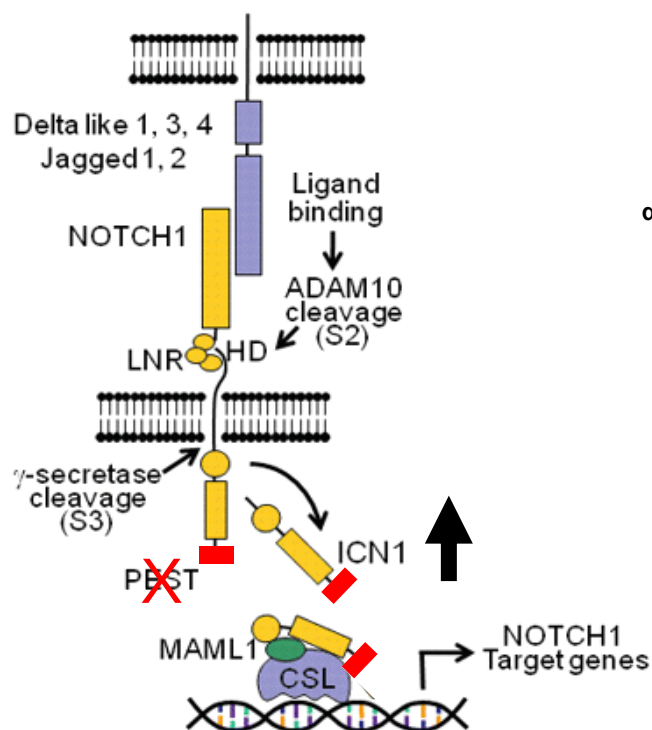
Consorcio ICGC - CLL

**Hospital Clínico, Universidad de Barcelona
Universidad de Oviedo, IUOPA
Instituto de Investigaciones Biomédicas August Pi I Sunyer
Centro de Regulación Genómica
Instituto Catalán de Oncología
Centro Investigación Cancer, Hospital Universitario (Salamanca)
Centro Nacional Investigaciones Oncológicas
Barcelona Supercomputer Center
Universidad de Santiago de Compostela
Universidad de Deusto
Universidad Pompeu Fabra
Hospital Clínico de Valencia
Hospital General de Asturias
Hospital Marques de Valde cilla**

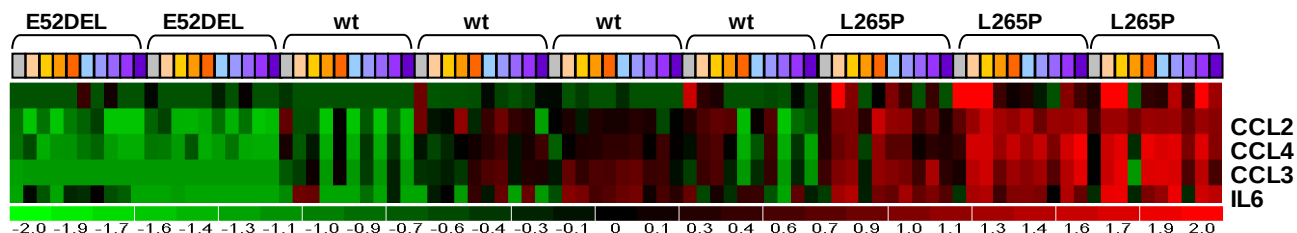
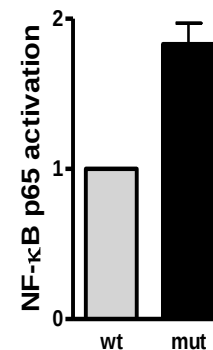
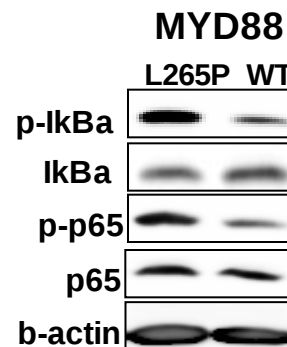
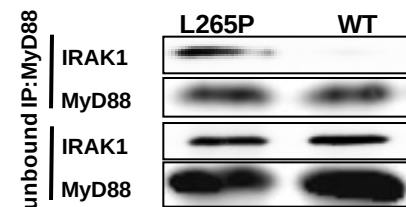
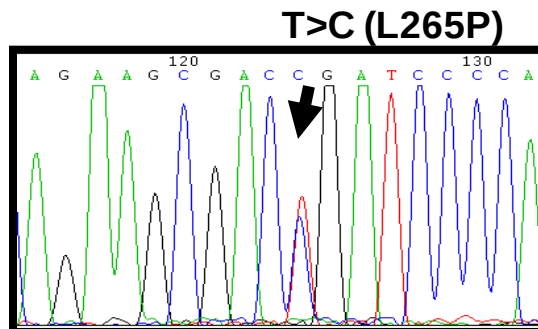
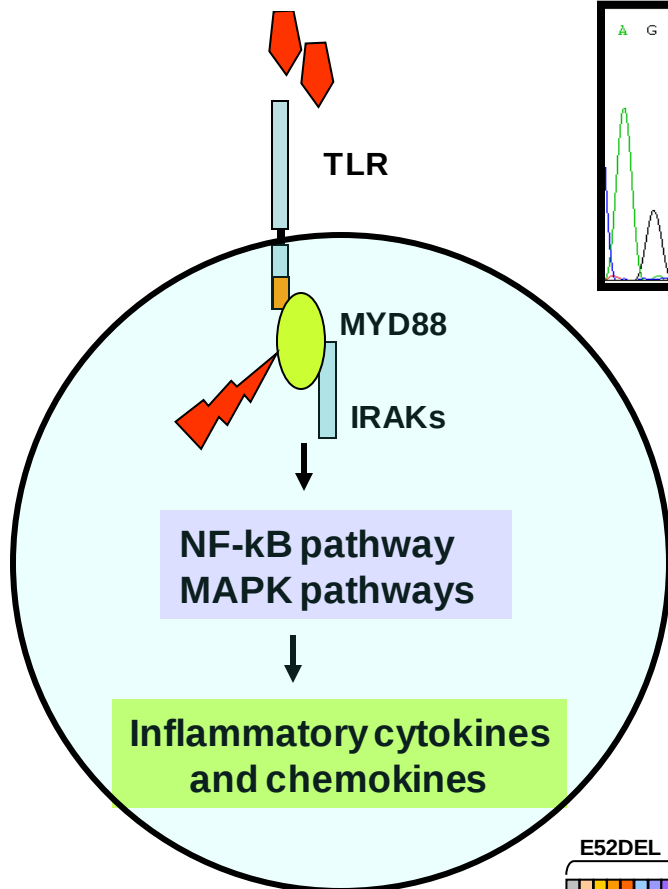
**Red de Investigación Cooperativa del Cáncer (RTICC)
Redes Nacionales del Banco de DNA y de Tumores
Instituto Nacional de Bioinformática (INB)
Centro Nacional de Analisis Genómico**

***Ministerio de Ciencia e Innovación
Instituto de Salud Carlos III***

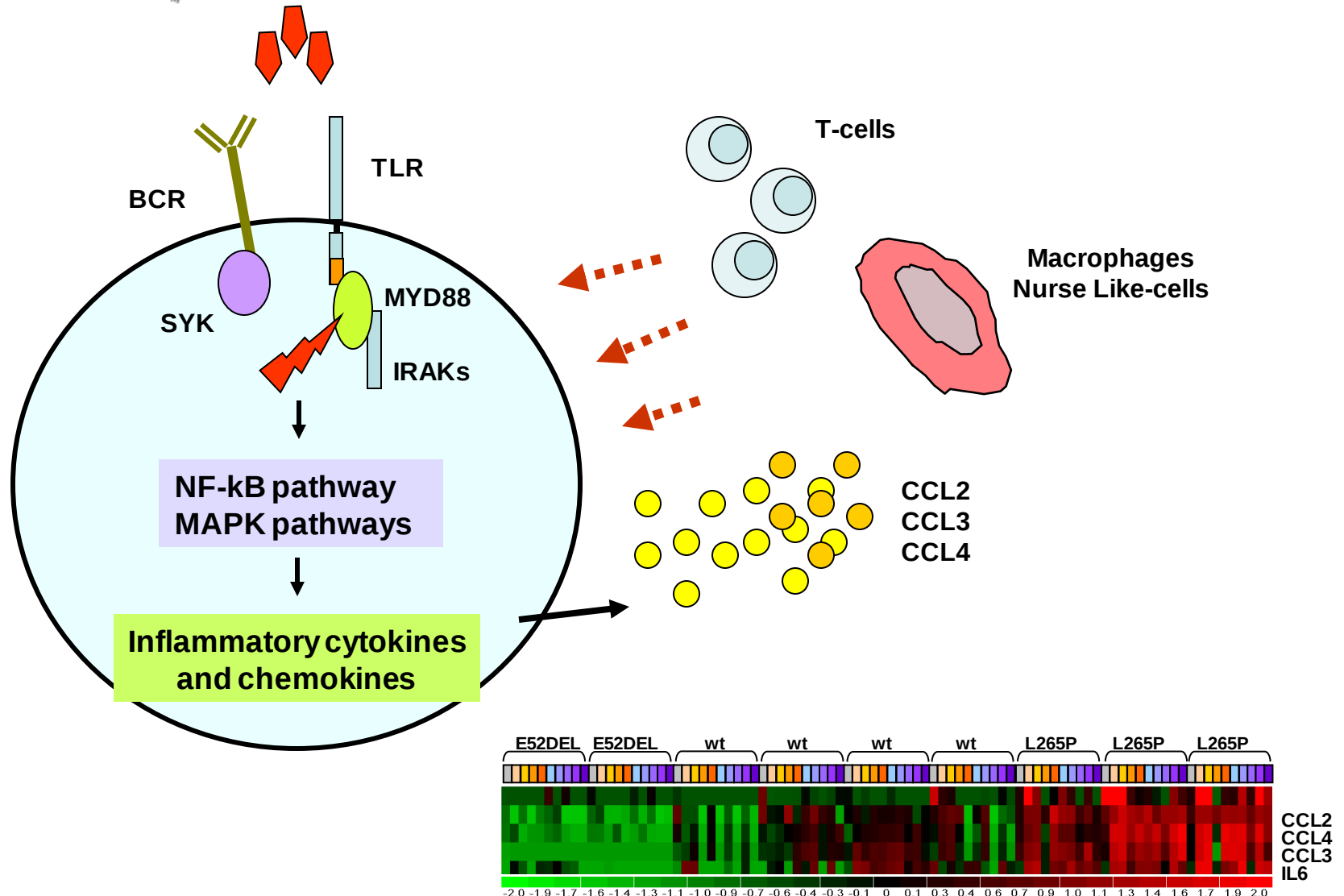
NOTCH1 oncogenic mutations in CLL



MYD88 oncogenic activation in CLL



MYD88 oncogenic activation in CLL may promote a favorable microenvironment



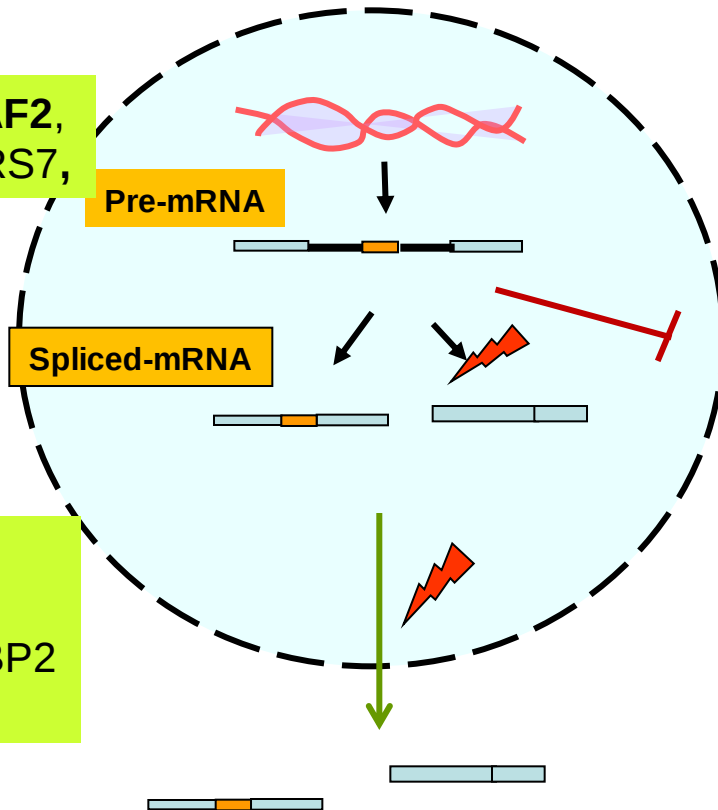
Mutacions Somàtiques en la maquinària molecular del processament del RNA en LLC

SF3B1, U2AF2,
SFRS1, SFRS7,

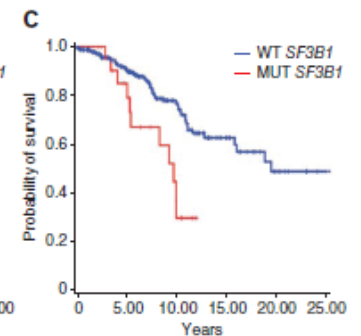
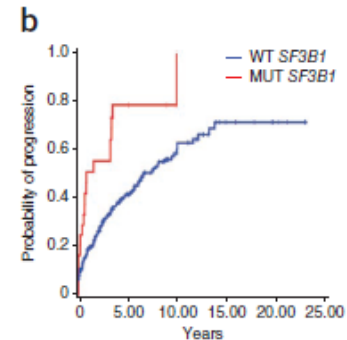
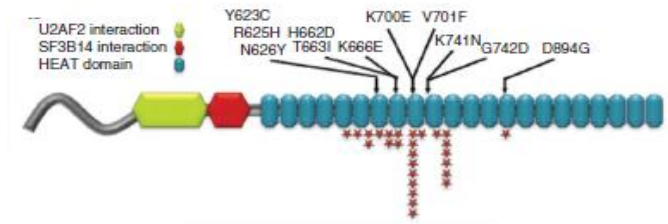
Pre-mRNA

Spliced-mRNA

XPO1, NXF1
EIF4A3,
MAGO, NCBP2
RBMX,

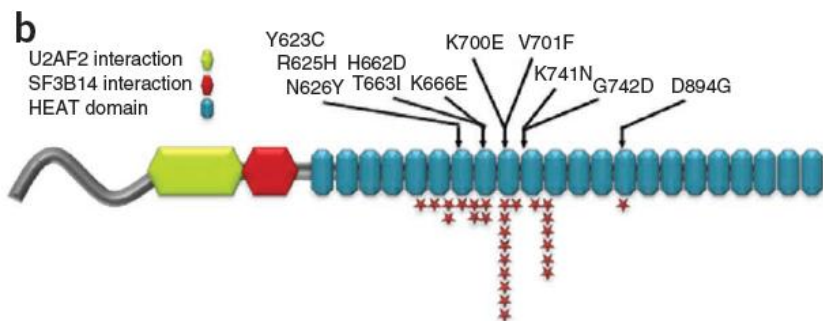


SF3B1

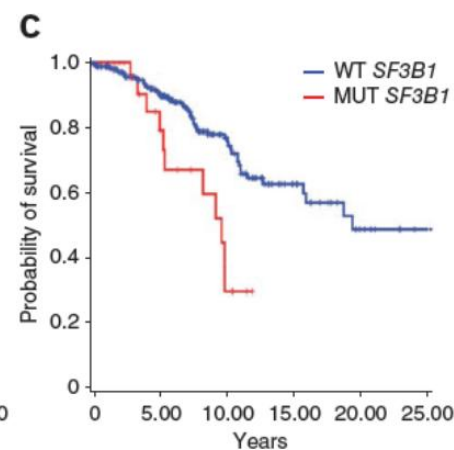
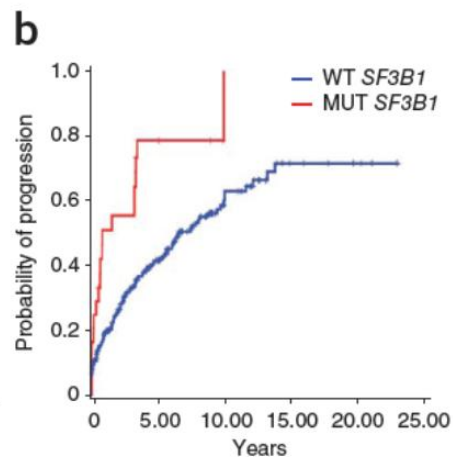
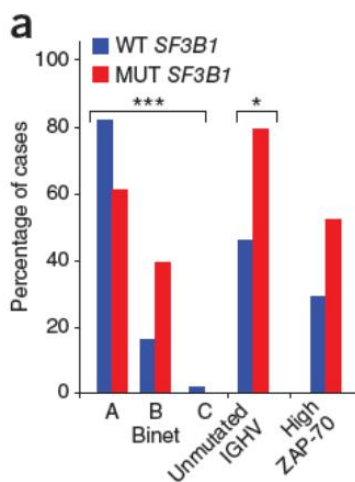
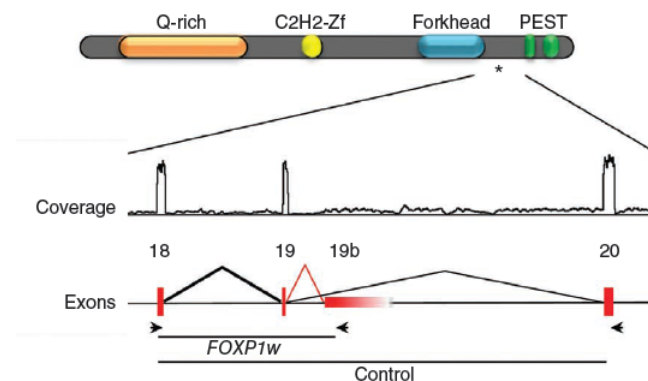


SF3B1 Mutations in CLL

SF3B1

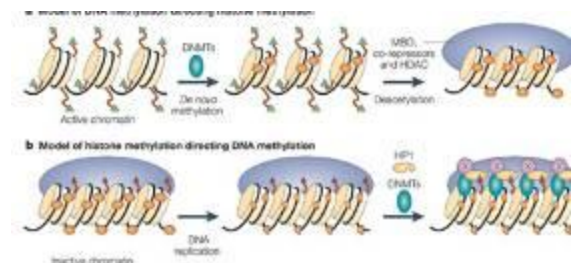
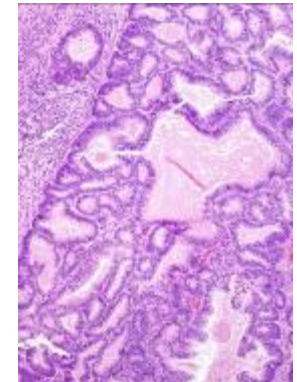
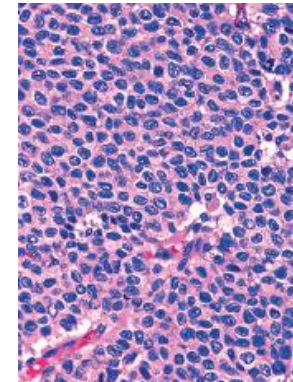
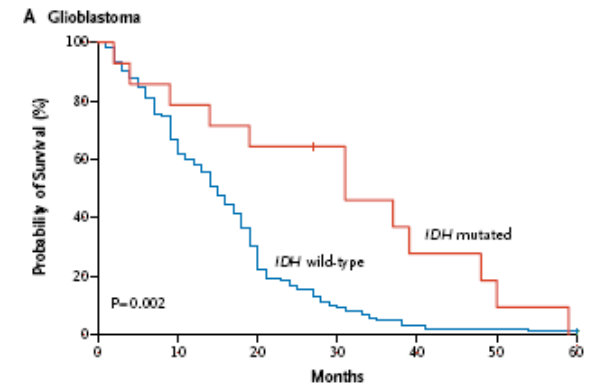


FOXP1



Somatic Mutations in Exome/Transcriptome

Tumor	Mutated Genes	Frequency
Low grade gliomas	IDH1/IDH2	70%
Glioblastoma		12%
Ovary		
• Granulosa Cell tumor	FOXL2	99%
• Clear cell Ca	ARID1A	46%
• Endometrioid Ca		30%
Clear Cell Renal Cancer	SETD2	
	JAR	15%
	UTX	
	PBRM1	44%

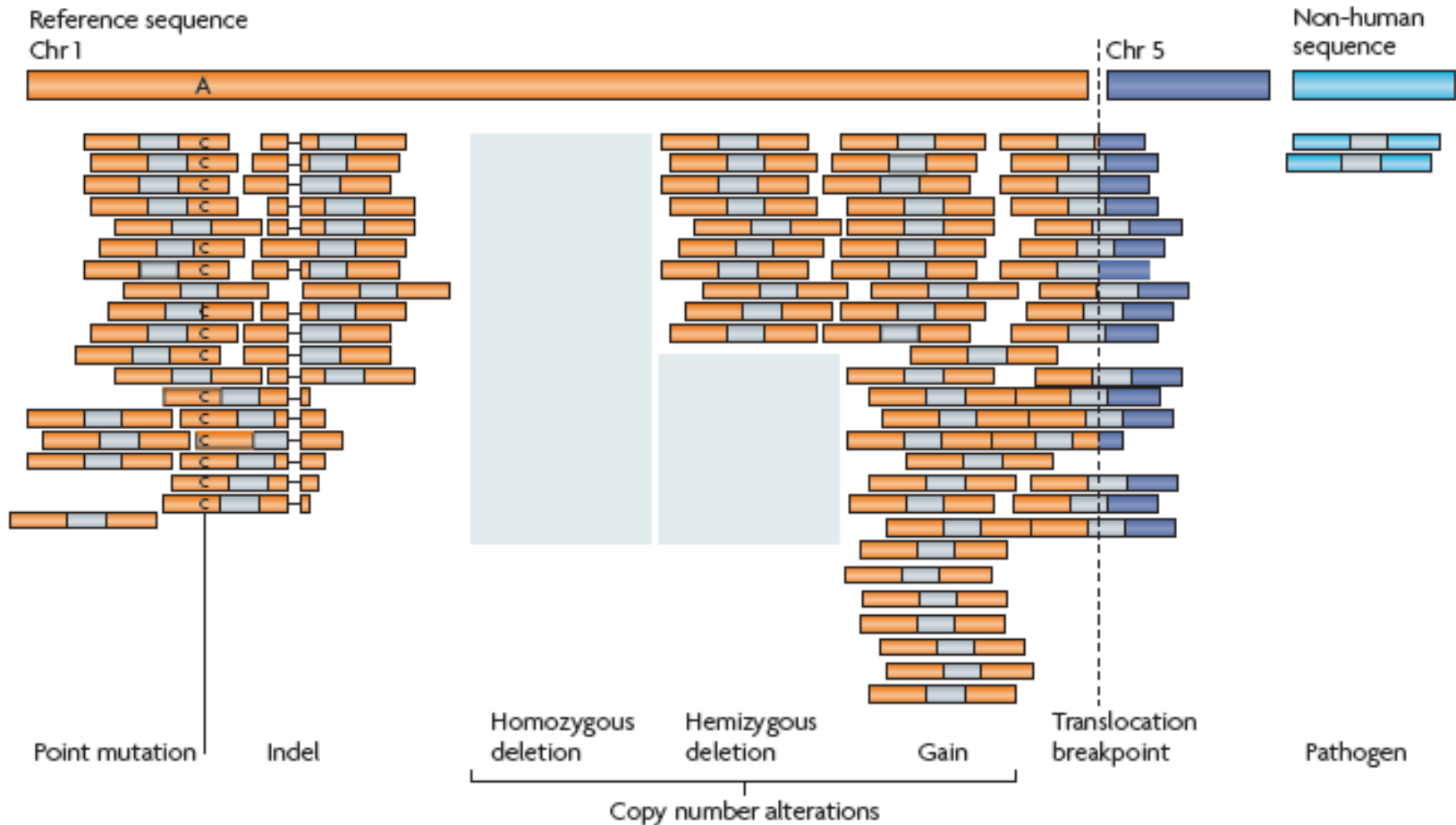


Somatic Mutations in WGS Studies

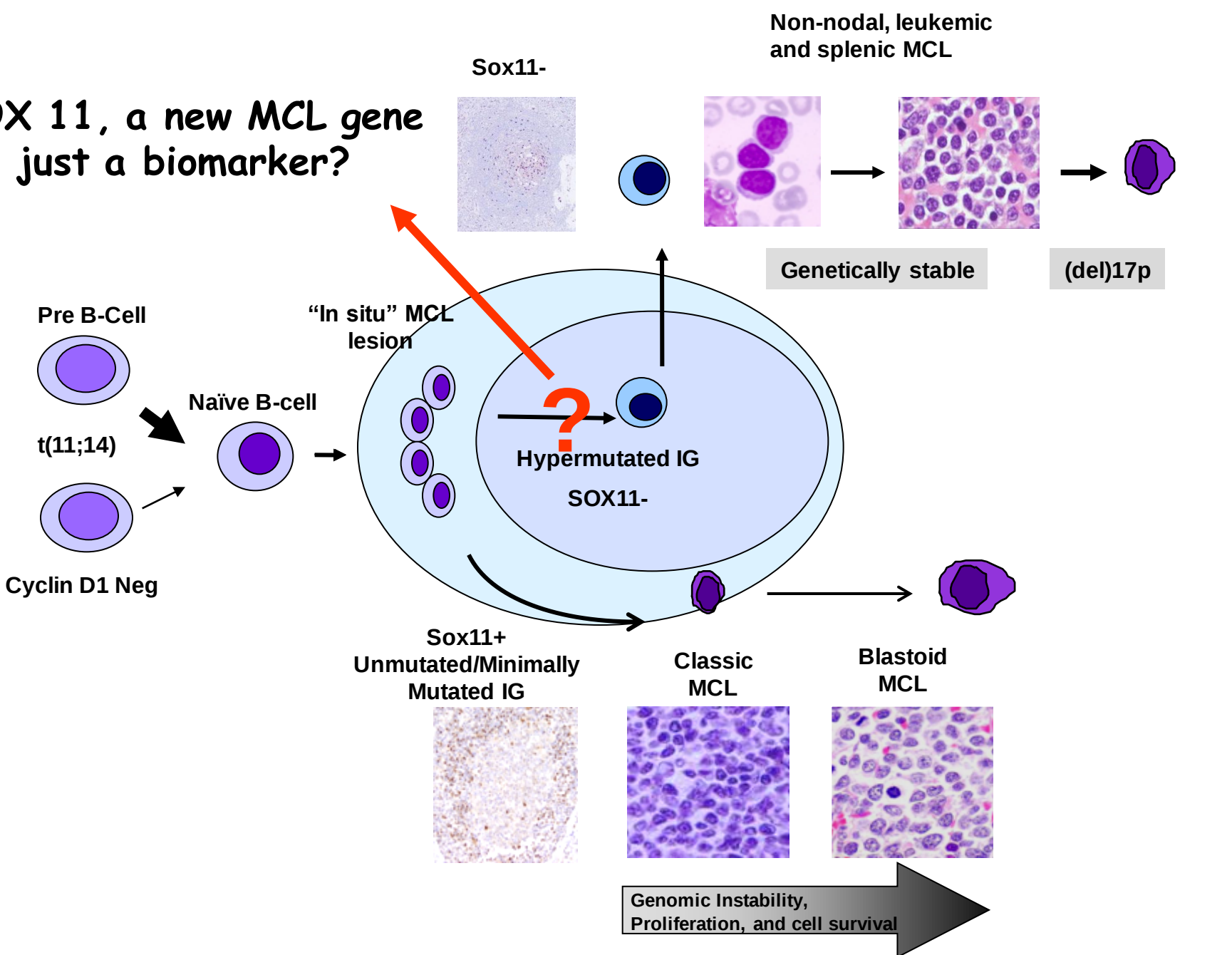
Tumor (yr)	Somatic Mutations	Non Synonymous	Clinical Validation
AML (2008,2009,2010)	750	10-12	<i>DNMT3A</i> (22%) <i>IDH1</i> (16%)
Breast			<i>ERBB2</i> (1.5%) <i>HAUS</i> (1%)
Lobular Met (2009)	-	32	
Basal-like (2009)	-	50	
Lung			Not-performed
Small-cell (cell line) (2009)	23,000	100	
Non-small cell lung (2009)	-	378	
Melanoma (cell line) (2009)	33,000	187	Not-performed
Hepatocellular Ca HVC (2011)	11,000	70	Not-performed
Multiple Myeloma (2011) 20 cases	7,500	35	10 Different pathways

NGS Applications: Genomics

Genome alterations than can be detected by NGS



SOX 11, a new MCL gene or just a biomarker?

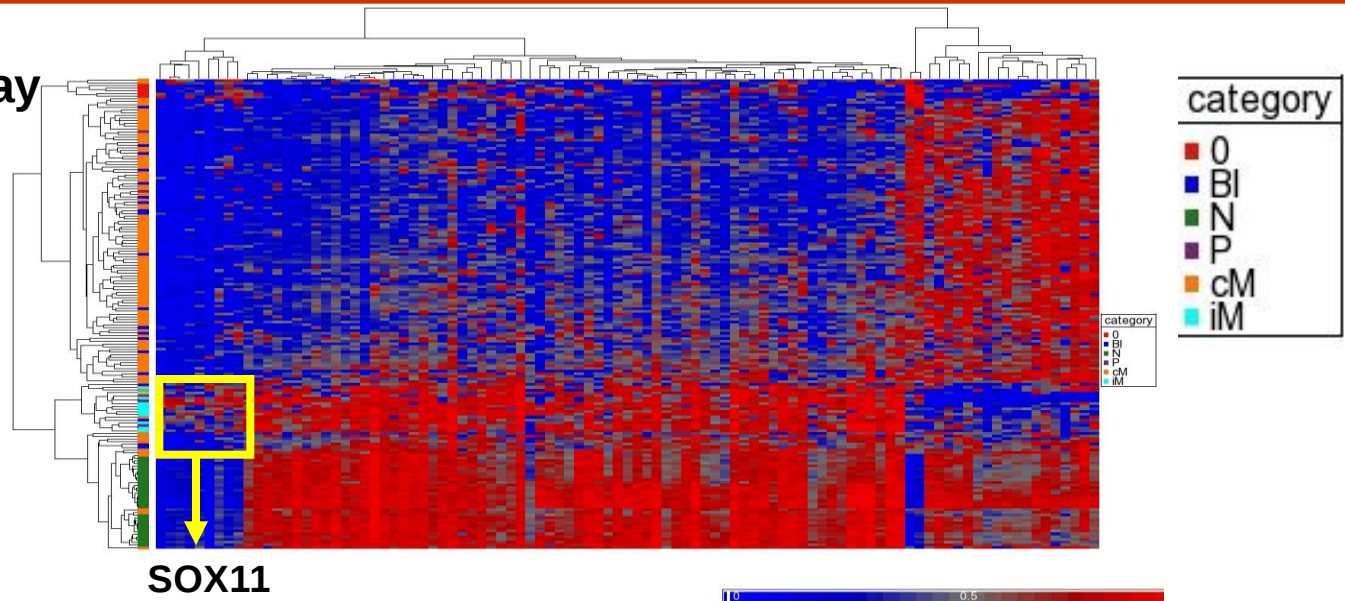


SOX 11, a new MCL gene or just a biomarker?

CpG methylation array



N=132 samples



SOX11 ChIP-on-chip

Z138 (Sox11+)

JVM2 (Sox11-)

2.1 M Promoter array

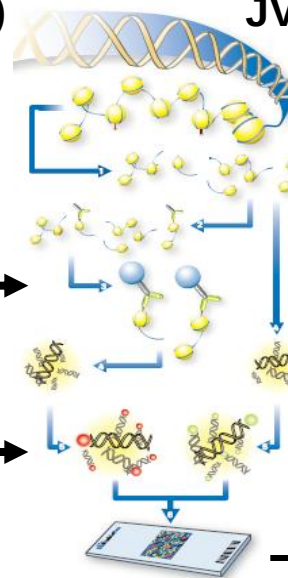


NimbleGen HD2

SOX11 Ab

Cy5

Cy3

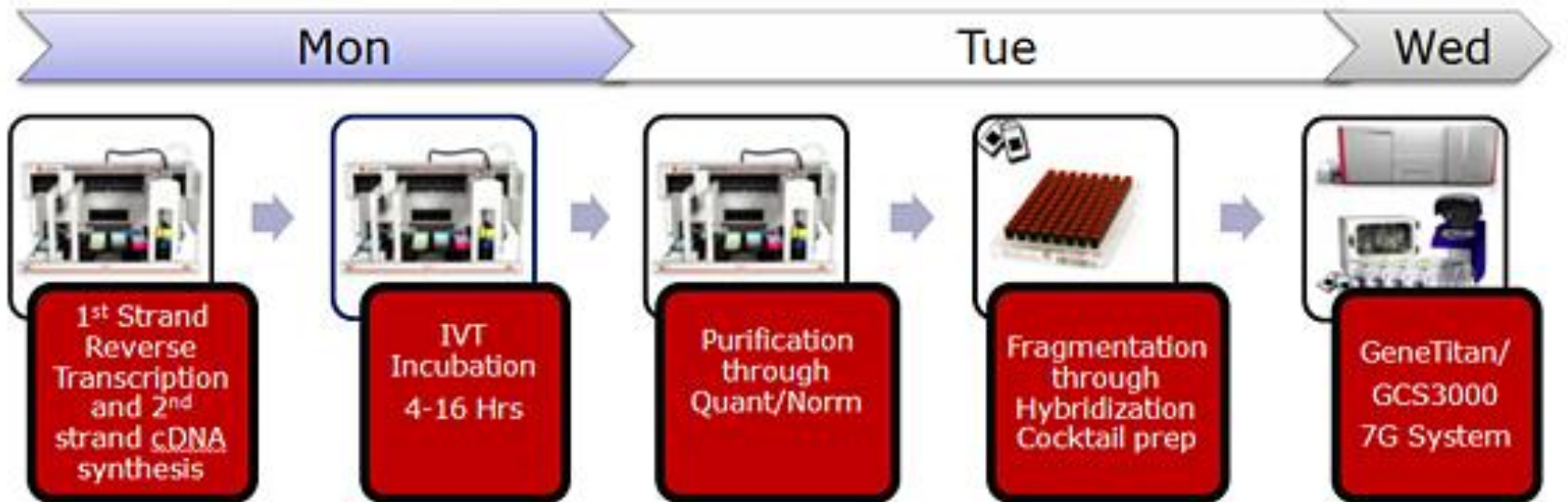


Promoters bound by
SOX11

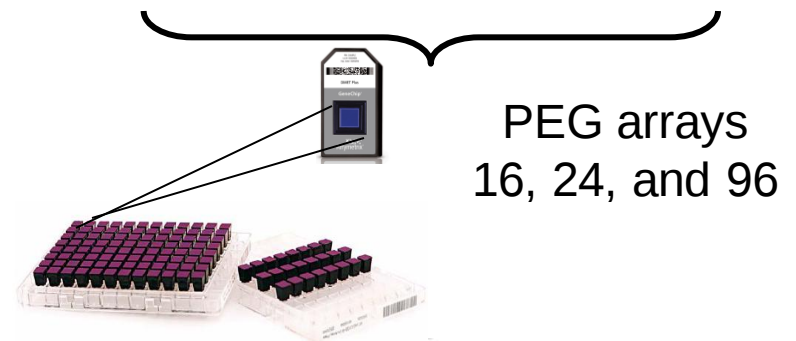
7 kb

3.2 kb

GENE TITAN MULTI-CHANNEL INSTRUMENT



**Beckman Coulter Biomek FXP[®]
Target Prep Express System**



**GeneTitan™
Multi-Channel
Instrument**

- 1. Presència d'unes 1000 mutacions somàtiques en el genoma de cada LLC**
- 2. Identificació de 46 mutacions en les regions codificants del genoma, 5-20 per cas**
- 3. Identificació de 4 mutacions recurrents amb implicacions clíniques**
- 4. Desenvolupament d'un mètode bioinformàtic altament eficaç per detectar mutacions somàtiques**