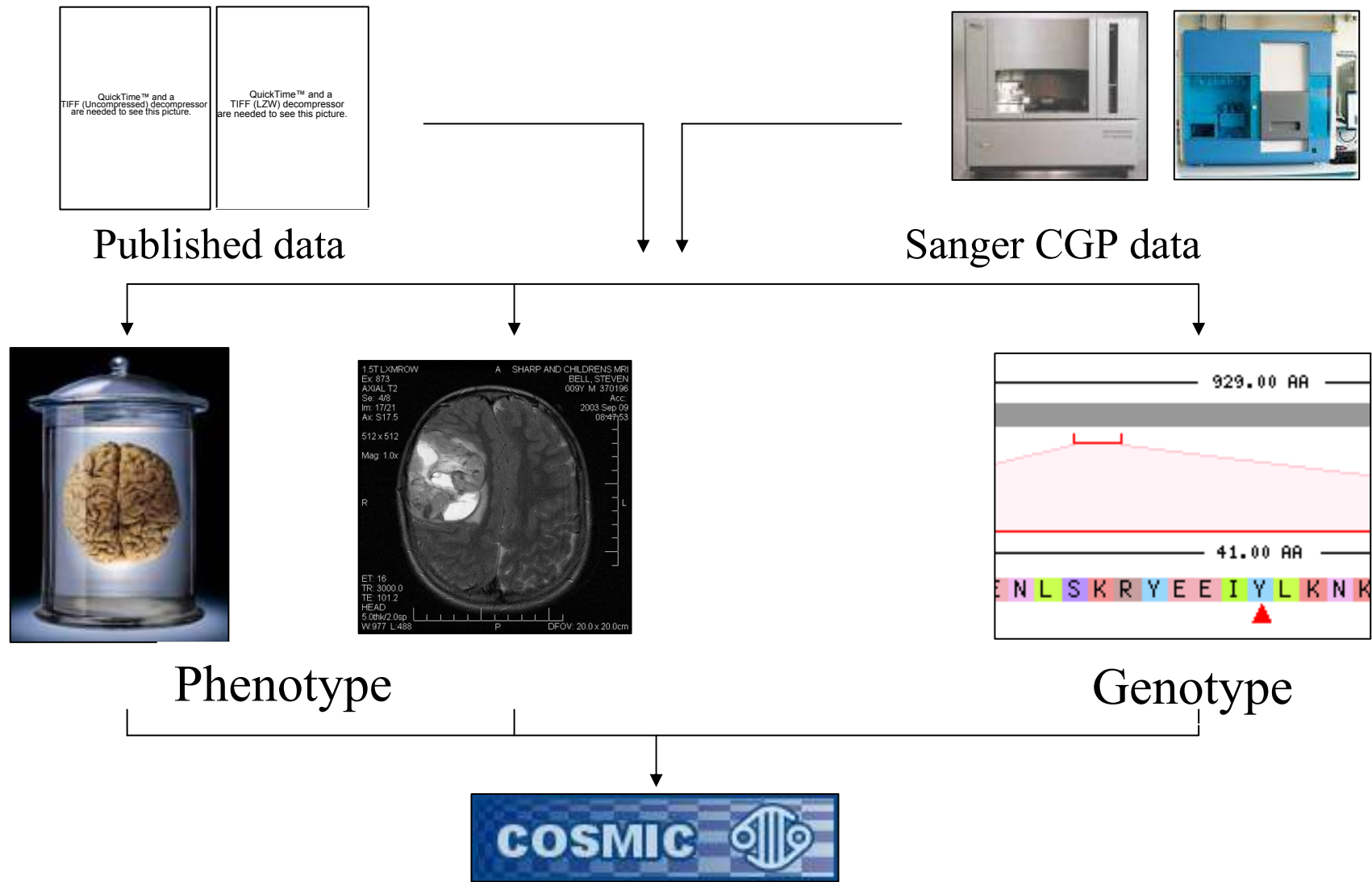


# COSMIC: Annotating cancer genomes.



# What is COSMIC ?



# 4 classes of Mutation



cDNA point mutations

X

X

X

Fusion genes

X

X

-

Whole genome annotations

X

X

-

CNV

X

X

-

COSMIC Catalogue of Somatic Mutations in Cancer	
Summary	View
<p><b>Summary</b></p> <p>The COSMIC Catalogue of Somatic Mutations in Cancer (COSMIC) is a public resource for the study of somatic mutations in cancer. It provides a comprehensive and up-to-date resource for the study of somatic mutations in cancer, including point mutations, copy number variations, and structural variants. The COSMIC Catalogue is a public resource for the study of somatic mutations in cancer, including point mutations, copy number variations, and structural variants. The COSMIC Catalogue is a public resource for the study of somatic mutations in cancer, including point mutations, copy number variations, and structural variants.</p>	<p><b>View</b></p> <p>30 genes, 20 samples, 20 mutations</p> <p><b>Point Mutations</b></p> <p>10 genes, 20 samples, 20 mutations</p> <p><b>Copy Number Variations</b></p> <p>10 genes, 20 samples, 20 mutations</p> <p><b>Structural Variants</b></p> <p>10 genes, 20 samples, 20 mutations</p>
<p><b>Advanced Search</b></p> <p>Search by Gene</p> <p>Search by Protein</p>	<p><b>Advanced Search</b></p> <p>Search by Gene</p> <p>Search by Protein</p>
<p><b>Additional Info</b></p> <p>Genes: 30</p> <p>Tissues: 20</p> <p>Alleles: 20</p> <p>Genes: 10</p>	<p><b>Additional Info</b></p> <p>Experiments: 20276</p> <p>Tissues: 12962</p> <p>Ancestors: 2829</p> <p>Alleles: 4418</p> <p>Genes: 111</p>

COSMIC Catalogue of Somatic Mutations in Cancer	
Summary	View
<p><b>Summary</b></p> <p>The COSMIC Catalogue of Somatic Mutations in Cancer (COSMIC) is a public resource for the study of somatic mutations in cancer. It provides a comprehensive and up-to-date resource for the study of somatic mutations in cancer, including point mutations, copy number variations, and structural variants. The COSMIC Catalogue is a public resource for the study of somatic mutations in cancer, including point mutations, copy number variations, and structural variants.</p>	<p><b>View</b></p> <p>30 genes, 20 samples, 20 mutations</p> <p><b>Point Mutations</b></p> <p>10 genes, 20 samples, 20 mutations</p> <p><b>Copy Number Variations</b></p> <p>10 genes, 20 samples, 20 mutations</p> <p><b>Structural Variants</b></p> <p>10 genes, 20 samples, 20 mutations</p>
<p><b>Advanced Search</b></p> <p>Search by Gene</p> <p>Search by Protein</p>	<p><b>Advanced Search</b></p> <p>Search by Gene</p> <p>Search by Protein</p>
<p><b>Additional Info</b></p> <p>Genes: 30</p> <p>Tissues: 20</p> <p>Alleles: 20</p> <p>Genes: 10</p>	<p><b>Additional Info</b></p> <p>Experiments: 20276</p> <p>Tissues: 12962</p> <p>Ancestors: 2829</p> <p>Alleles: 4418</p> <p>Genes: 111</p>

COSMIC Catalogue of Somatic Mutations in Cancer	
Summary	View
<p><b>Summary</b></p> <p>The COSMIC Catalogue of Somatic Mutations in Cancer (COSMIC) is a public resource for the study of somatic mutations in cancer. It provides a comprehensive and up-to-date resource for the study of somatic mutations in cancer, including point mutations, copy number variations, and structural variants. The COSMIC Catalogue is a public resource for the study of somatic mutations in cancer, including point mutations, copy number variations, and structural variants.</p>	<p><b>View</b></p> <p>30 genes, 20 samples, 20 mutations</p> <p><b>Point Mutations</b></p> <p>10 genes, 20 samples, 20 mutations</p> <p><b>Copy Number Variations</b></p> <p>10 genes, 20 samples, 20 mutations</p> <p><b>Structural Variants</b></p> <p>10 genes, 20 samples, 20 mutations</p>
<p><b>Advanced Search</b></p> <p>Search by Gene</p> <p>Search by Protein</p>	<p><b>Advanced Search</b></p> <p>Search by Gene</p> <p>Search by Protein</p>
<p><b>Additional Info</b></p> <p>Genes: 30</p> <p>Tissues: 20</p> <p>Alleles: 20</p> <p>Genes: 10</p>	<p><b>Additional Info</b></p> <p>Experiments: 20276</p> <p>Tissues: 12962</p> <p>Ancestors: 2829</p> <p>Alleles: 4418</p> <p>Genes: 111</p>

# cDNA point mutations

<b>Mutation id</b>	13118						
<b>AA Mutation</b>	p.S567* (Substitution - Nonsense)						
<b>CDS Mutation</b>	c.1700C>A (Substitution)						
<b>Mutation id</b>	1043						
<b>AA Mutation</b>	p.T502fs*22 (Complex - frameshift)						
<b>CDS Mutation</b>	c.1504_1510ACATCTC>TCTCAAGAATCTCAAGAATCA (Complex)						
<b>Nucleotides Inserted</b>	tctcaagaatctca						
<b>Mutation id</b>	25339						
<b>AA Mutation</b>	p.? (Unknown)						
<b>CDS Mutation</b>	c.160+5G>A (Substitution - intronic)						
<b>Mutation id</b>	21239						
<b>AA Mutation</b>	p.L340fs*20 (Insertion - Frameshift)						
<b>CDS Mutation</b>	c.1019_1020insT (Insertion)						
<b>Nucleotides Inserted</b>	t						
<b>Mutation Location</b>	<p>Length (AA) Amino Acid ← 504.00 AA →</p> <p>Length (AA) Amino Acid Mutation ← 41.00 AA →</p> <p>I V G T Q G K P A I A H R D L K S K N I L V K K N G T C C I A D L G L A V R H D S</p>						
<b>Gene</b>	<a href="#">TGFB1</a>						
<b>Tissue Distribution (Top 5)</b>	<table border="1"> <tr> <td>Tissues</td> <td>1</td> </tr> <tr> <td>lung (1)</td> <td>1</td> </tr> <tr> <td>scale (Mutation #)</td> <td>1</td> </tr> </table>	Tissues	1	lung (1)	1	scale (Mutation #)	1
Tissues	1						
lung (1)	1						
scale (Mutation #)	1						
<b>Associated Samples</b>	<table border="1"> <thead> <tr> <th>Sample Name</th> <th>Primary Tissue</th> <th>Zygosity</th> </tr> </thead> <tbody> <tr> <td><a href="#">NCI-H2171</a></td> <td><a href="#">lung</a></td> <td>Heterozygous</td> </tr> </tbody> </table>	Sample Name	Primary Tissue	Zygosity	<a href="#">NCI-H2171</a>	<a href="#">lung</a>	Heterozygous
Sample Name	Primary Tissue	Zygosity					
<a href="#">NCI-H2171</a>	<a href="#">lung</a>	Heterozygous					

Small intragenic mutations putatively affecting protein product of a single gene:

- Nonsense,
- Missense
- Inframe Ins / Del
- Frameshift
- Complex replacement

# Gene-specific mutation spectrum

**COSMIC core:  
The Histogram Page  
(TP53)**

**Point mutation histogram**

QuickTime™ and a  
TIFF (Uncompressed) decompressor  
are needed to see this picture.

**Complex mutations**

**Insertions/Deletions**

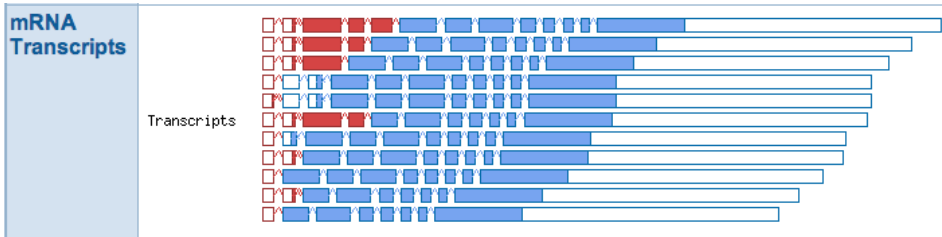
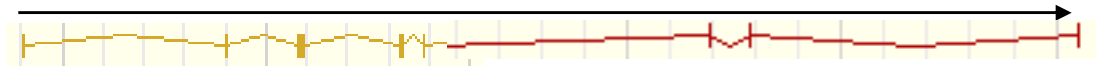
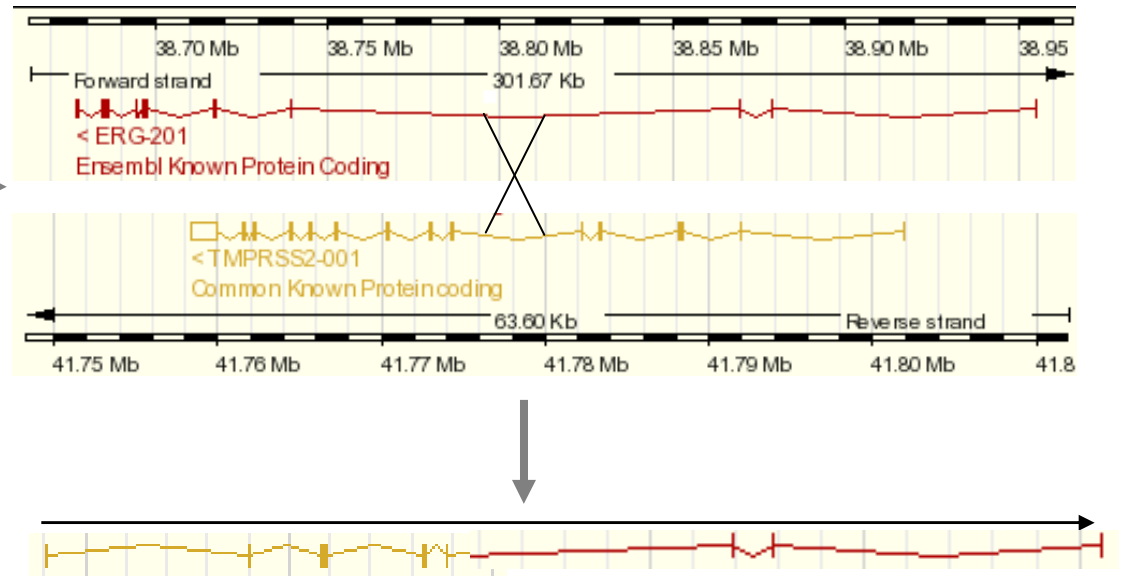
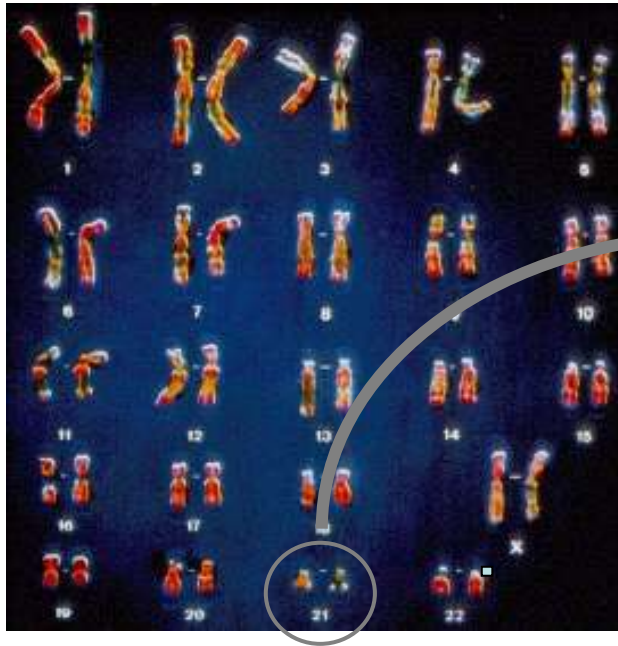
**Domain structures**

**cDNA (CDS) scale**

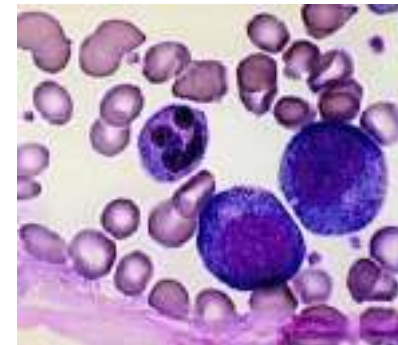
**Mutation counts/frequencies  
By tumour primary site**

Details for TP53				
Primary Tissue	Mutated Samples	% Mutated	All Samples	Mutation Data
<a href="#">bone</a>	17	53%	32	<a href="#">More Details</a>
<a href="#">breast</a>	40	50%	80	<a href="#">More Details</a>
<a href="#">central nervous system</a>	34	48%	71	<a href="#">More Details</a>
<a href="#">cervix</a>	3	23%	13	<a href="#">More Details</a>

# Fused Genes



Primary Tissue	Mutated Samples	% Mutated	All Samples	Mutation Data
NS	0	0%	20	<a href="#">More Details</a>
large intestine	0	0%	36	<a href="#">More Details</a>
prostate	712	39%	1826	<a href="#">More Details</a>
stomach	0	0%	40	<a href="#">More Details</a>
<b>Totals</b>	<b>712</b>	<b>37%</b>	<b>1922</b>	<a href="#">More Details</a>

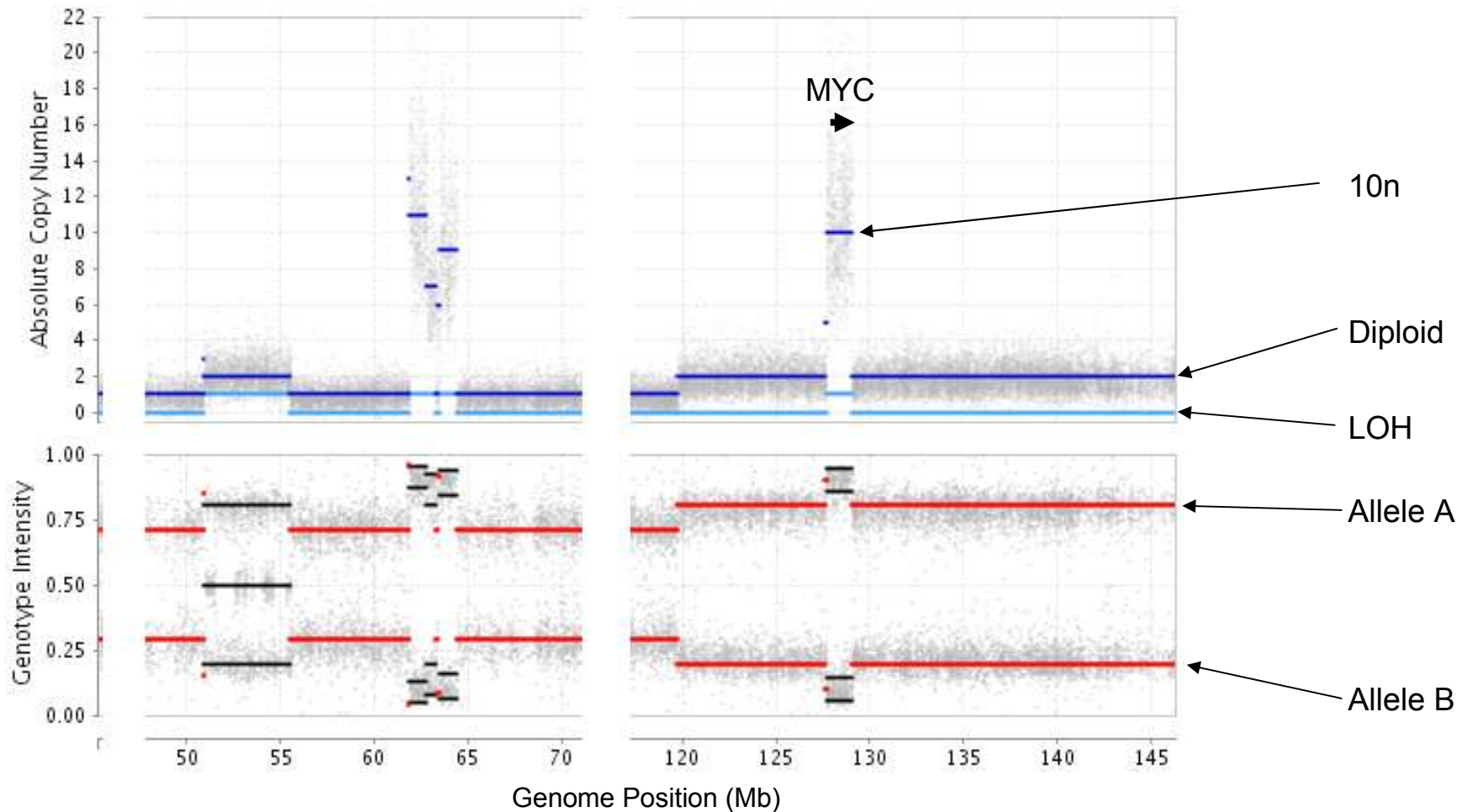


**TMPRSS2 / ERG fusions in 39% of Prostate tumours**

# Copy Number Variation (CNV)

Examining cancer aneuploidy using SNP microarrays

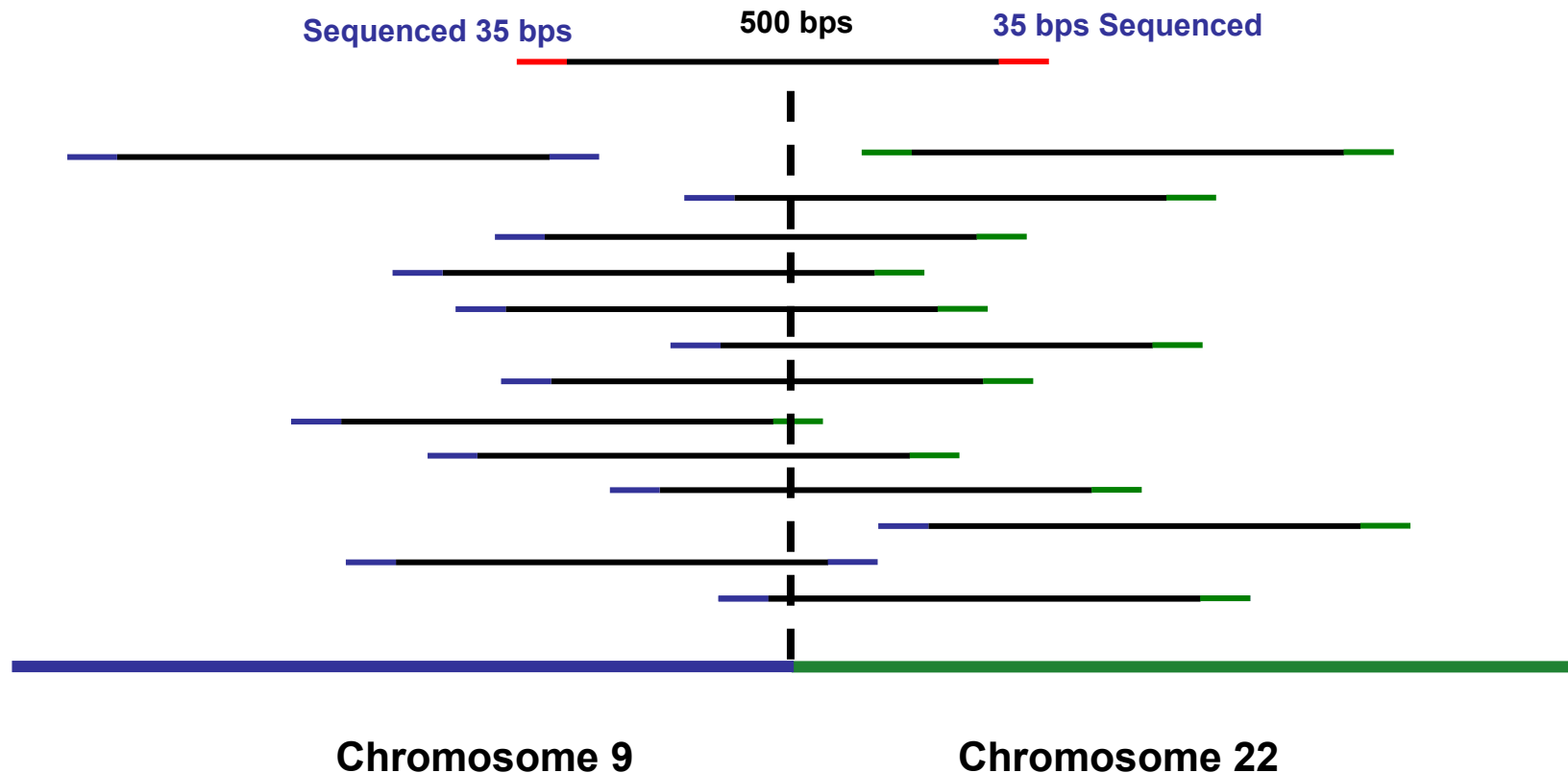
Chr. 8 amplification of MYC oncogene in NCI-H2171 Lung tumour



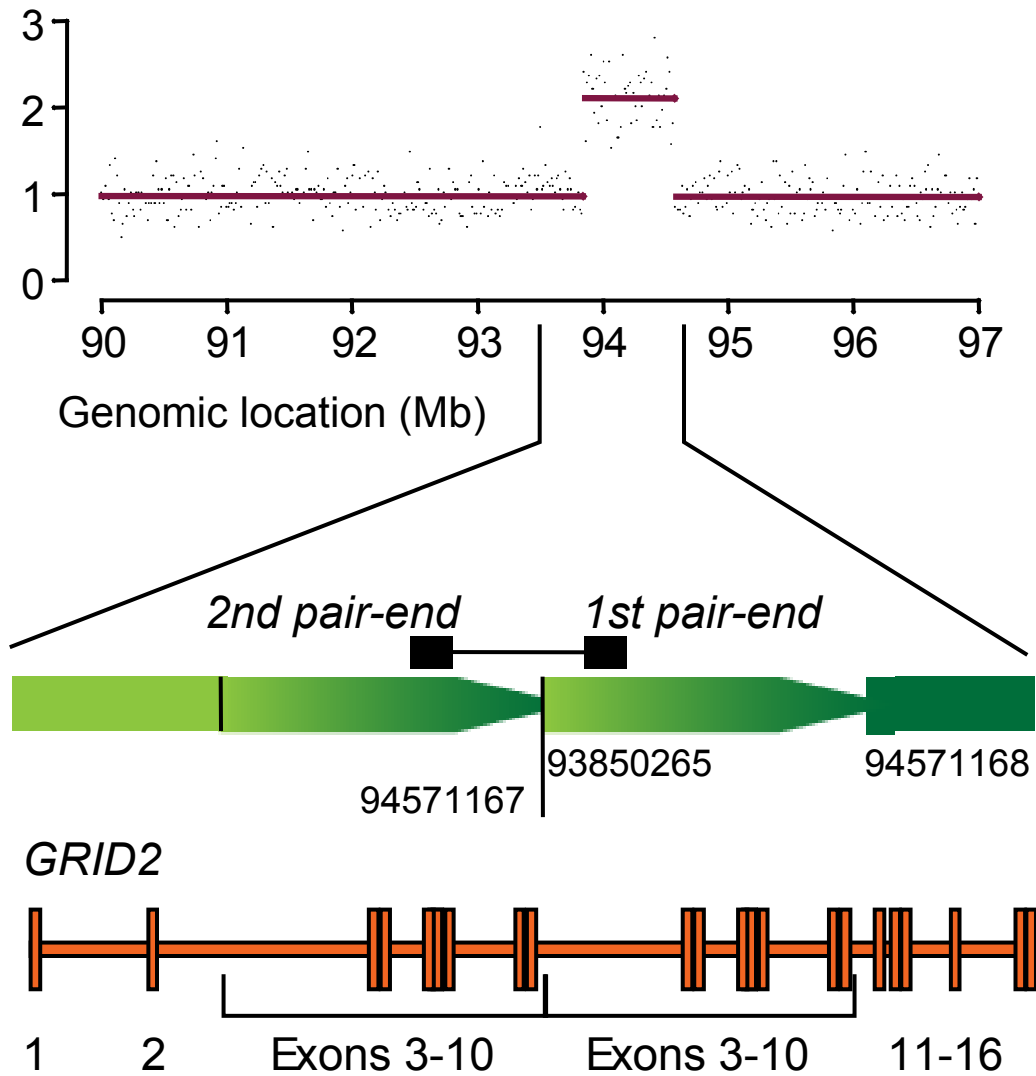
# Whole genome Solexa paired end sequencing

## Examining tumours for genomic rearrangements:

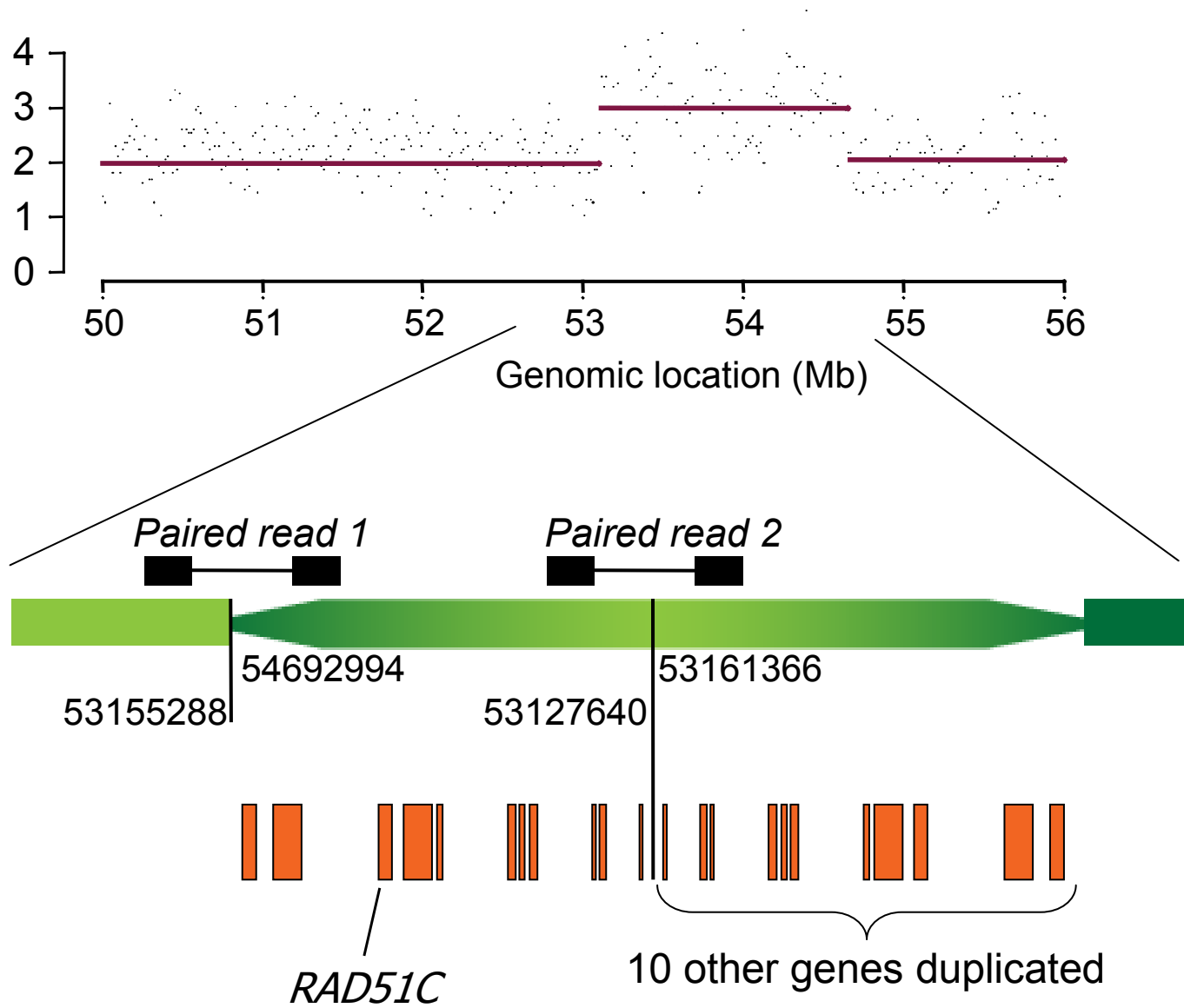
- Fragment genomic DNA to ~500bp fragments
- ligate adapters
- surface bind
- amplify
- 35x scanned single-nucleotide sequencing reactions in pairs
- align approx. 52 million sequences to reference genome
- select pairs mapping  $\leftrightarrow$  500bp apart
- capillary sequence across selected regions to define exact breakpoint



# Tandem duplication – Chr 4

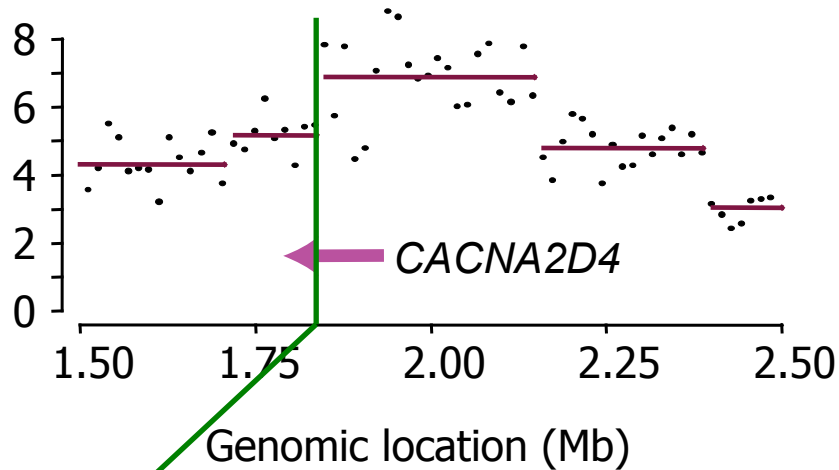


# Inverted duplication



# 2:12 Fusion gene

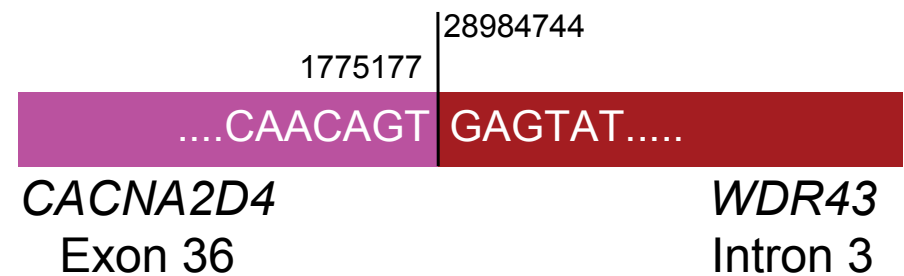
NCI-H2171: Chr 12



Chr 2

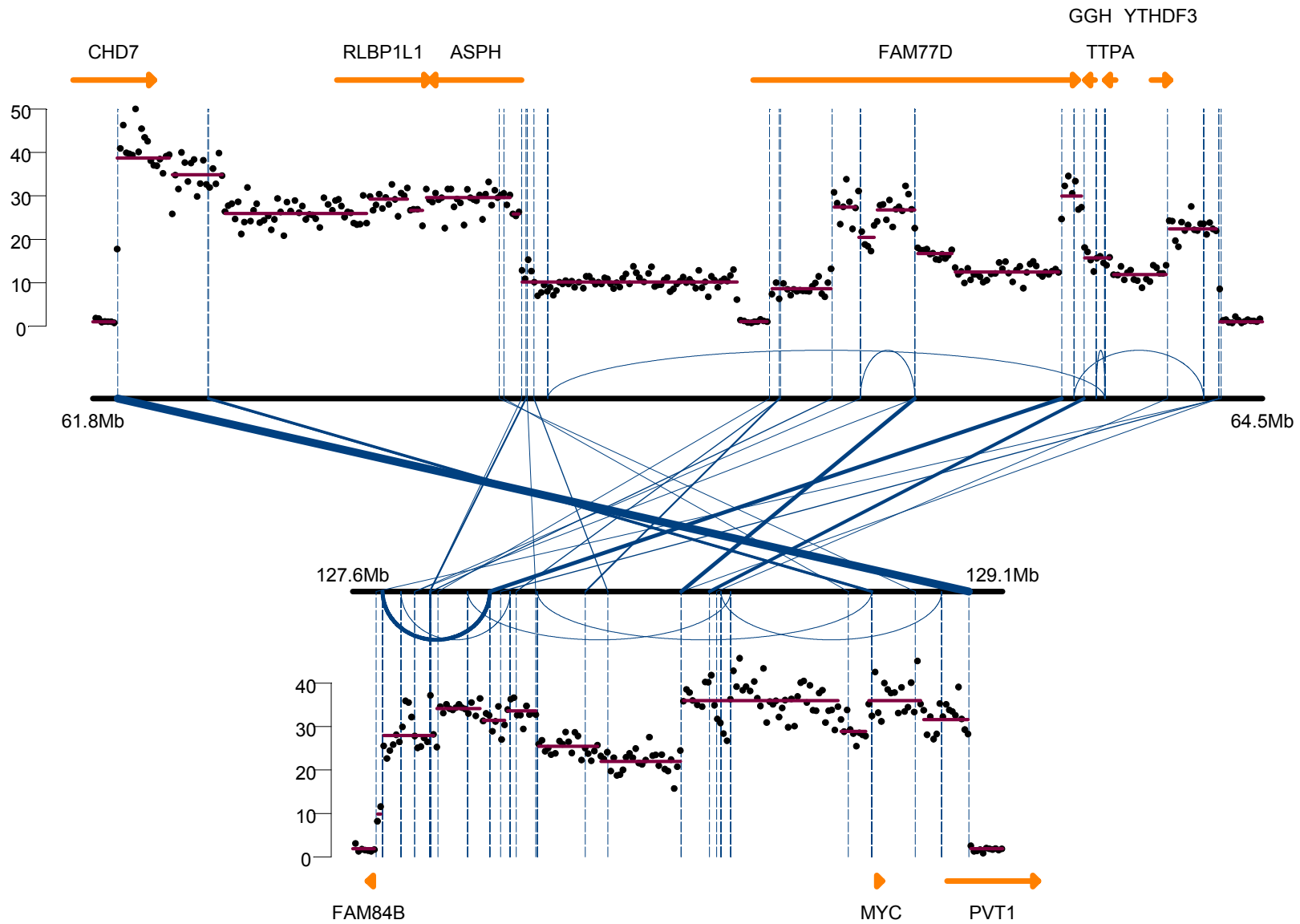
Chr 12 (- strand)

Chr 2 (+ strand)

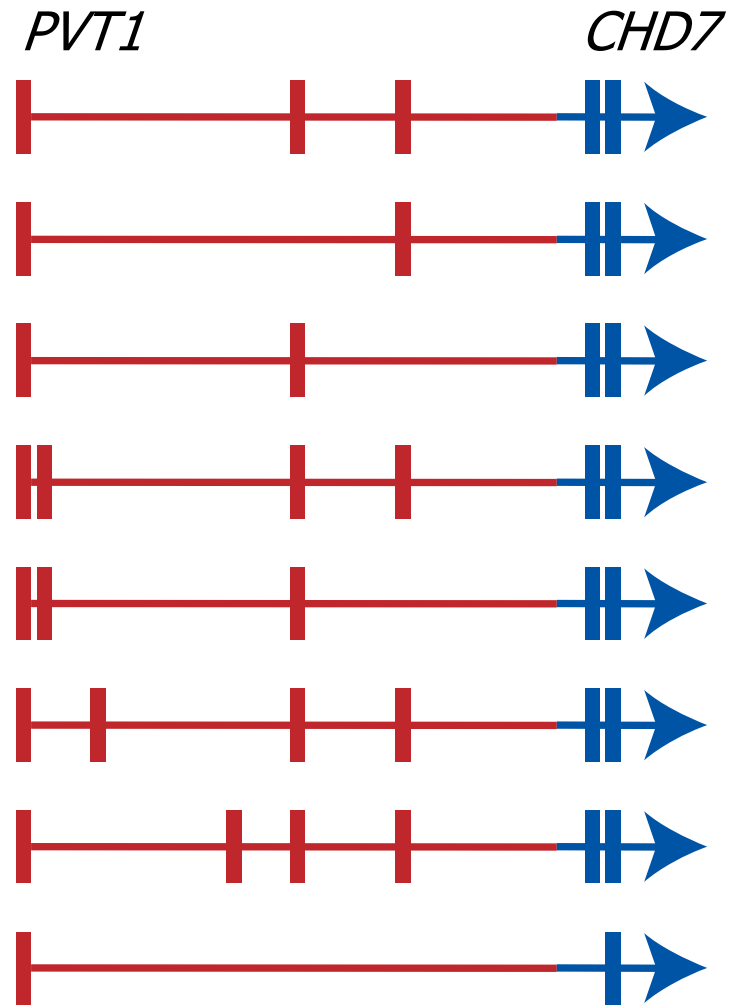
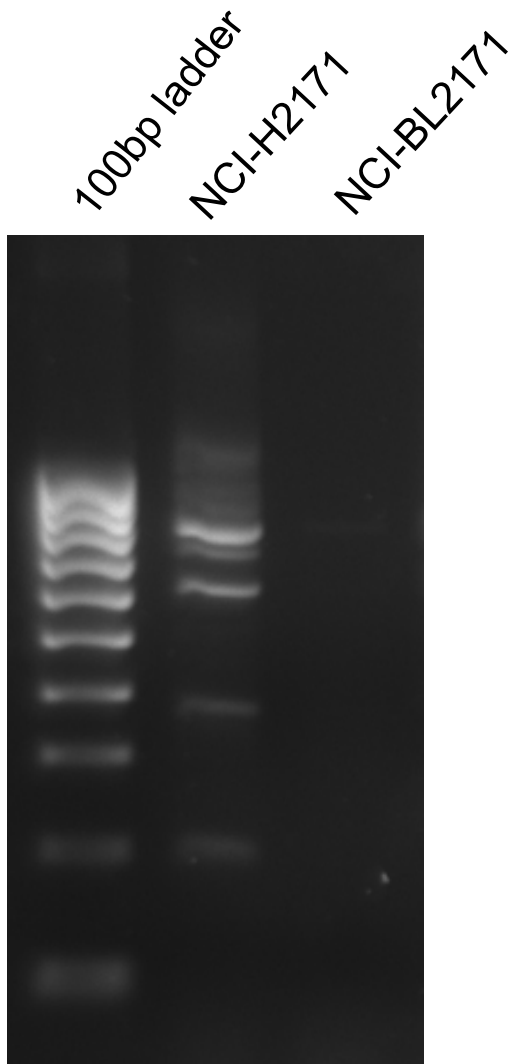


*CACNA2D4-WDR43* fusion gene

# Amplicon breakpoint detection



# *PVT1-CHD7* fusion gene



# Breast tumour summary.

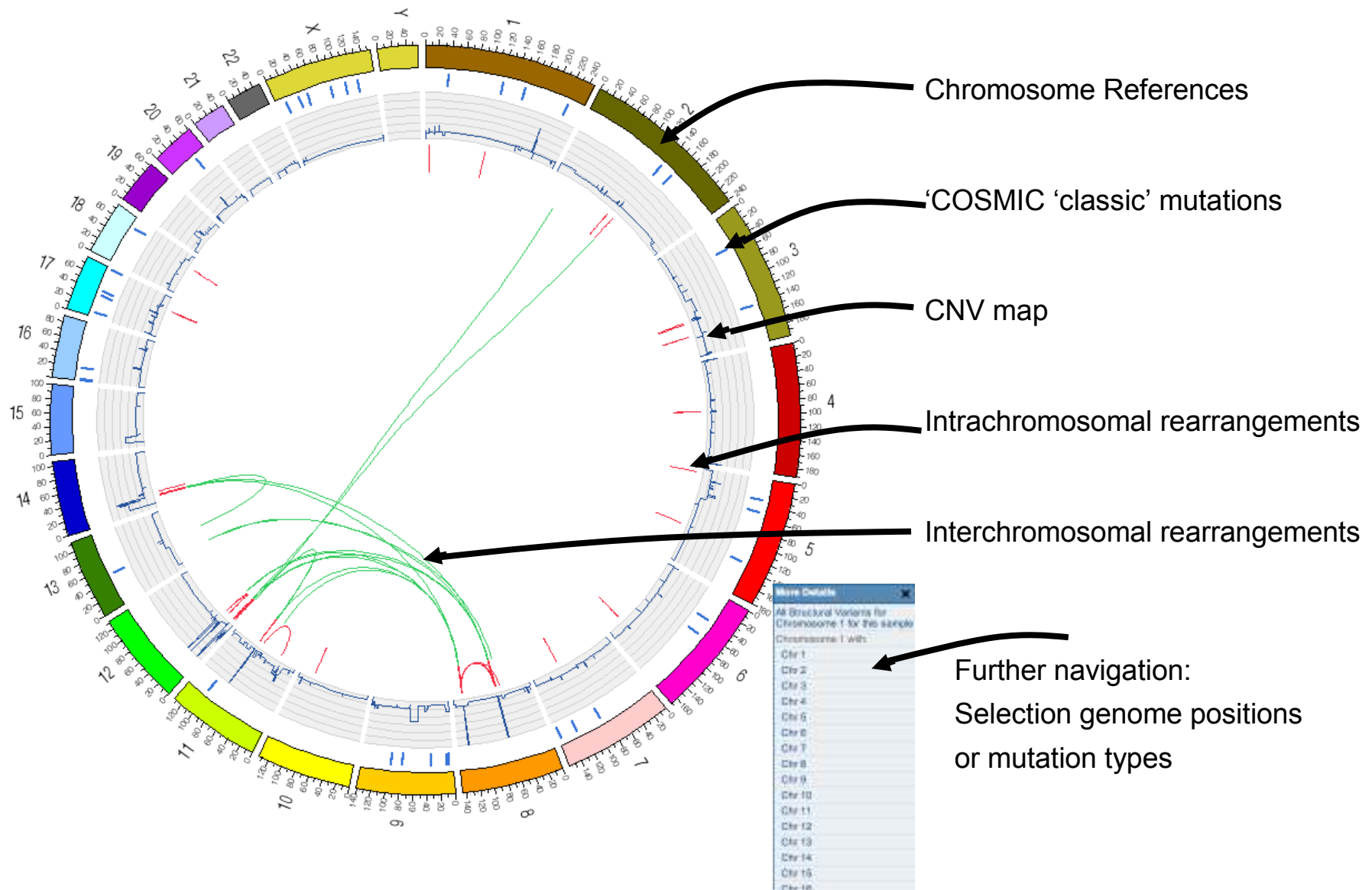
## 8 Breast cancers now fully analysed

- 888 Somatic rearrangements
- 36 Fusion genes (18 IN FRAME)
- 78 Internally rearranged genes (39 IN FRAME)
- 17 Potential Promoter fusions

Currently whole-genome-screening 94 tumours from:

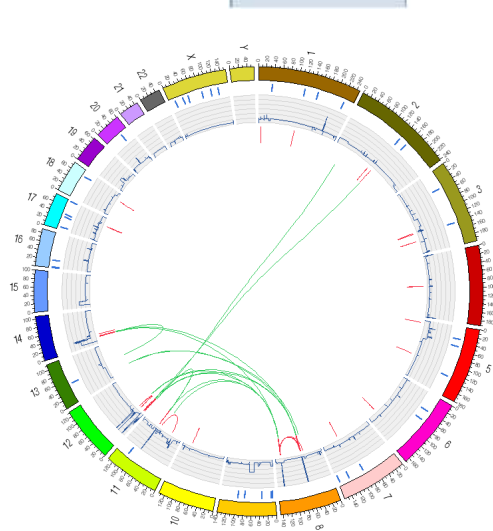
Lung, Skin, Kidney, Pancreas....

# Summarising whole-genome mutation data



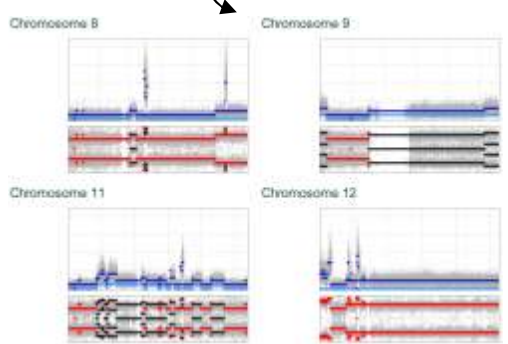
# Navigation

<b>Sample Name</b>	NCI-H2171 - Click <a href="#">here</a> for additional samples with the same name.	
<b>COSMIC id</b>	688015	
<b>Source</b>	cell-line	
<b>Tumour Classification</b>	<b>Category</b>	<b>Entry</b>
	Primary Site; Subtype1; Subtype 2; Subtype 3	lung; NS; NS; NS
	Primary Histology; Subtype1; Subtype 2; Subtype 3	carcinoma; small cell carcinoma; NS; NS
<b>Mutations</b>	<a href="#">Overview</a> <a href="#">COSMIC Point Mutations</a> <a href="#">Copy Number</a> <a href="#">Structural Variants</a>	



Chromosome	Start (kb)	End (kb)	Gene	Variant Type	Effect
1	10,000,000	10,000,000	BRCA1	Point Mutation	missense
2	20,000,000	20,000,000	EGFR	Point Mutation	missense
3	30,000,000	30,000,000	KRAS	Point Mutation	missense
4	40,000,000	40,000,000	TP53	Point Mutation	missense
5	50,000,000	50,000,000	PTEN	Point Mutation	missense
6	60,000,000	60,000,000	PIK3CA	Point Mutation	missense
7	70,000,000	70,000,000	NRAS	Point Mutation	missense
8	80,000,000	80,000,000	RAF1	Point Mutation	missense
9	90,000,000	90,000,000	BRAF	Point Mutation	missense
10	100,000,000	100,000,000	MEK1	Point Mutation	missense
11	110,000,000	110,000,000	RASGEF1B	Point Mutation	missense
12	120,000,000	120,000,000	SHC3	Point Mutation	missense
13	130,000,000	130,000,000	GRB2	Point Mutation	missense
14	140,000,000	140,000,000	SOS1	Point Mutation	missense
15	150,000,000	150,000,000	FAK1	Point Mutation	missense
16	160,000,000	160,000,000	PPP2R1A	Point Mutation	missense
17	170,000,000	170,000,000	PPP2R1B	Point Mutation	missense
18	180,000,000	180,000,000	PPP2R2	Point Mutation	missense
19	190,000,000	190,000,000	PPP2R3	Point Mutation	missense
20	200,000,000	200,000,000	PPP2R3B	Point Mutation	missense
21	210,000,000	210,000,000	PPP2R4	Point Mutation	missense
22	220,000,000	220,000,000	PPP2R5	Point Mutation	missense

Sample	Gene	AA Mutation	CDS Mutation	Description of Variant	Significance
688015	ADAM27	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM30	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM13	G>A113G	G>A113G	Confirmed somatic variant	Heterozygous
688015	ADAM9	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM12	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM10	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM11	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM14	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM15	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM16	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM17	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM18	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM19	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM20	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM21	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM22	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM23	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM24	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM25	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM26	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM27	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM28	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM29	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous
688015	ADAM30	A>G113G	A>G113G	Confirmed somatic variant	Heterozygous



# Rearrangement mutations / breakpoints

Each rearrangement can have a number of breakpoints:

Simple deletions may present only 1 breakpoint;

Rearrangements involving sequence fragments or compound amplifications can present many:

A t(12:8) translocation with 2 chromosome 12 “shards” at the interface.

104	translocation with seq fragment(s)	chr12:g.6943257_6943258_inschr12:8123579_8123600_inschr.7310260_7310594_chr.8:64286067trans[?]					
104	translocation with seq fragment(s)	12	6943257		8	64286067	
	1	12	6943257	-	12	8123600	-
	2	12	8123579	-	12	7310594	-
	3	12	7310260	-	8	64286067	-

An amplification of a t(12:8) translocation;

compound mitotic amplification events create multiple related breakpoints

105	Complex Amplicon	chr12:g.(6880974_29387067)_chr8:(63468962_128498886)trans[?]					
105	Complex Amplicon	8	63468962		12	29387067	
	1	12	28189034	+	8	128128372	-
	2	12	6880974	-	8	128498886	-
	3	12	29387067	+	8	63468962	+
	4	12	8397158	-	8	128303858	-

# A known tumour-promoting mutation dataset

COSMIC displays all mutations, not just those of known oncogenic potential.

But, COSMIC's cell-line resequencing project is examining 50 known cancer genes through 800 cell lines

All of the mutations found are manually scrutinised to exclude potential passenger mutations or SNPs

This makes it a very useful test dataset for mutation prediction software.

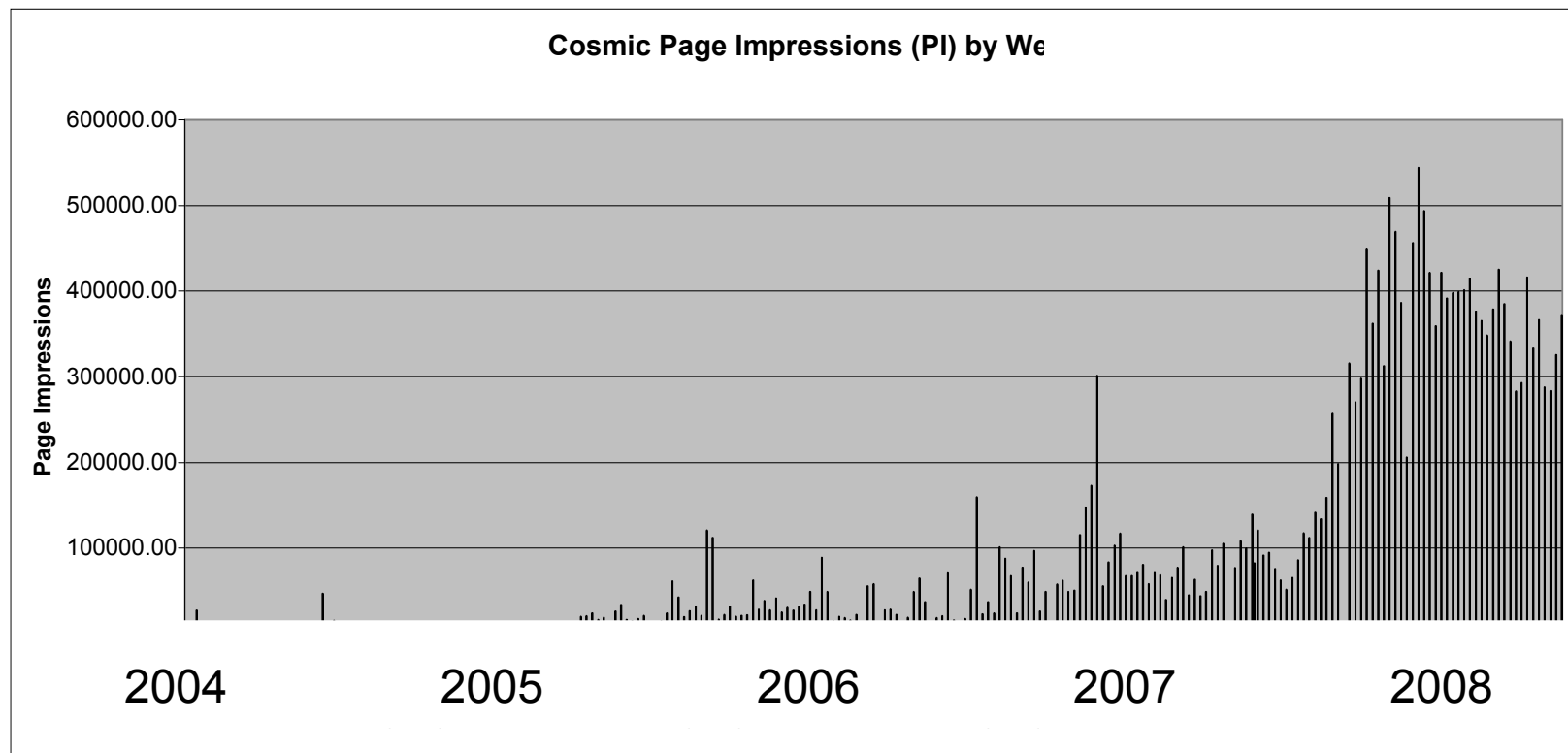
Sample Name	NCI-H2171																		
COSMIC Sample id	686015																		
Sample Source	cell-line																		
Tumour Classification	<table border="1"> <thead> <tr> <th>Category</th> <th>Entry</th> </tr> </thead> <tbody> <tr> <td>Primary Site:</td> <td>lung</td> </tr> <tr> <td>Tissue Subtype 1:</td> <td>NS</td> </tr> <tr> <td>Tissue Subtype 2:</td> <td>NS</td> </tr> <tr> <td>Tissue Subtype 3:</td> <td>NS</td> </tr> <tr> <td>Primary Histology:</td> <td>carcinoma</td> </tr> <tr> <td>Histology Subtype 1:</td> <td>small cell carcinoma</td> </tr> <tr> <td>Histology Subtype 2:</td> <td>NS</td> </tr> <tr> <td>Histology Subtype 3:</td> <td>NS</td> </tr> </tbody> </table>	Category	Entry	Primary Site:	lung	Tissue Subtype 1:	NS	Tissue Subtype 2:	NS	Tissue Subtype 3:	NS	Primary Histology:	carcinoma	Histology Subtype 1:	small cell carcinoma	Histology Subtype 2:	NS	Histology Subtype 3:	NS
Category	Entry																		
Primary Site:	lung																		
Tissue Subtype 1:	NS																		
Tissue Subtype 2:	NS																		
Tissue Subtype 3:	NS																		
Primary Histology:	carcinoma																		
Histology Subtype 1:	small cell carcinoma																		
Histology Subtype 2:	NS																		
Histology Subtype 3:	NS																		
Genes Tested With Mutations	<p>RB1, TP53</p> <p>48 genes with no mutations so far detected in this sample can be found at the bottom of this page</p>																		
Mutations	<table border="1"> <thead> <tr> <th>Sample</th> <th>Gene</th> <th>AA Mutation</th> <th>CDS Mutation</th> <th>Germine or Somatic</th> <th>Zygosity</th> </tr> </thead> <tbody> <tr> <td>686015</td> <td>RB1</td> <td>p.L356L*</td> <td>p.L1793G&gt;A</td> <td>Confirmed somatic variant</td> <td>Homozygous</td> </tr> <tr> <td>686015</td> <td>TP53</td> <td>p.Q144*</td> <td>p.L339I&gt;T</td> <td>Confirmed somatic variant</td> <td>Homozygous</td> </tr> </tbody> </table> <p>See other mutations for this sample in <a href="#">COSMIC</a></p>	Sample	Gene	AA Mutation	CDS Mutation	Germine or Somatic	Zygosity	686015	RB1	p.L356L*	p.L1793G>A	Confirmed somatic variant	Homozygous	686015	TP53	p.Q144*	p.L339I>T	Confirmed somatic variant	Homozygous
Sample	Gene	AA Mutation	CDS Mutation	Germine or Somatic	Zygosity														
686015	RB1	p.L356L*	p.L1793G>A	Confirmed somatic variant	Homozygous														
686015	TP53	p.Q144*	p.L339I>T	Confirmed somatic variant	Homozygous														

Confirmed Oncogenic

Sample	Gene	AA Mutation	CDS Mutation	Germine or Somatic	Zygosity
686015	PTPRG	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	ADAMTS12	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	NOS3	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	KIF5A	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	CH3L	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	PCRD	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	ADAM30	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	HSA1130	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	CDK5	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	ITV6	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	ALCO1	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	TRAP	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	ABGAP3	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	TGFBI	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	GAZF1	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	DCX	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	CDMD3	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	FKBP7	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	SMO1	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	TAF11	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	TRAF2	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	MYR2	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	HSPM	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	CD3A	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	RDP2	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	ADAM20	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	ITGB6	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	ND511	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	TPM3	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	RH1	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	OLSN	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	SMG2	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	RAV1	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	ITGA1	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	LOC44897	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	CDK3	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	DOCK1	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	SRFBF7	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	PLD3	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	MMP19	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	DNAHCC1	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	MPSH1	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	CH3L	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous
686015	Claf55	p.L213L	p.L333G	Confirmed somatic variant	Heterozygous

# COSMIC future

- Map mutations to Uniprot co-ords with Pfam, integrating into both websites (& distribute via DAS)
- Finalise structural rearrangement ontology & nomenclature
- Improve mining of rearrangement data; navigation by genomic positions & gene footprints
- Import enormous rearrangement & non-coding mutation datasets



# Summary

COSMIC is about to incorporate whole-genome Solexa sequencing results

- Integrate all oncogenic mutation types:
  - Point mutations
  - Fusion genes
  - Copy Number Variants
  - Genomic rearrangements
- Annotation
  - HGVS - style summarisation
  - Ensembl annotated breakpoint detail
  - Uniprot integration
- Adding meaning to the dataset
  - Increasingly important as the quantity & range of data soars
  - Mutation consequence - does the variant *promote* cancer ?
    - Software (CanPredict, SIFT etc?)
  - A known positive oncogenic mutation dataset for testing these