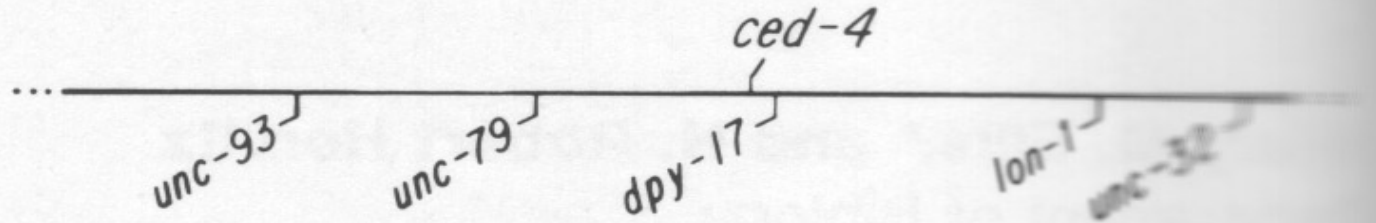


# Sample Genetic Maps of chromosomes from Nematodes (*C. elegans*)

III



IV

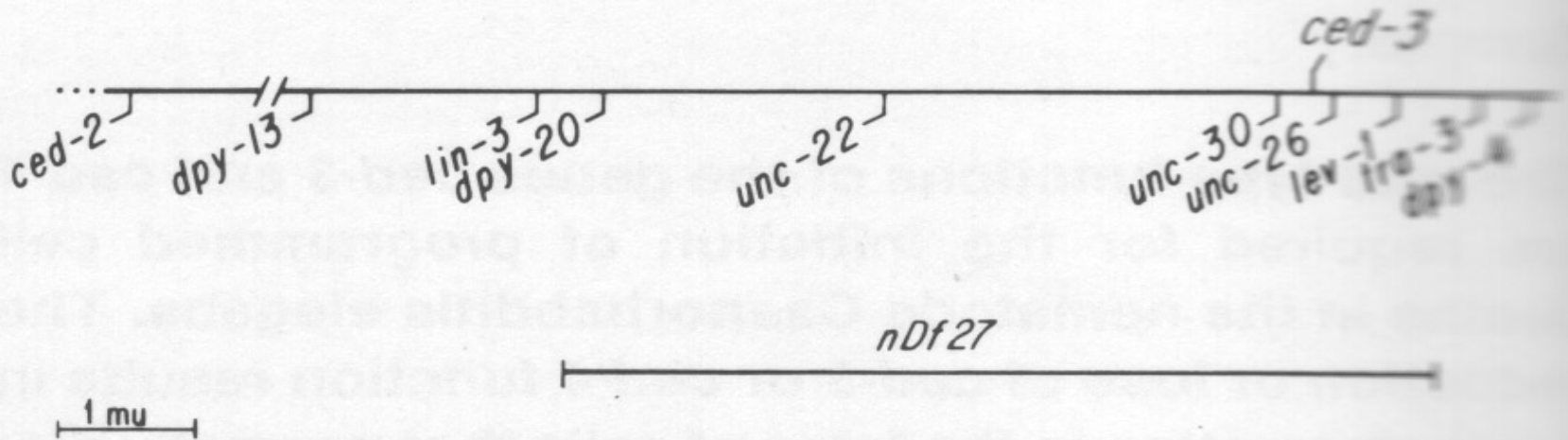
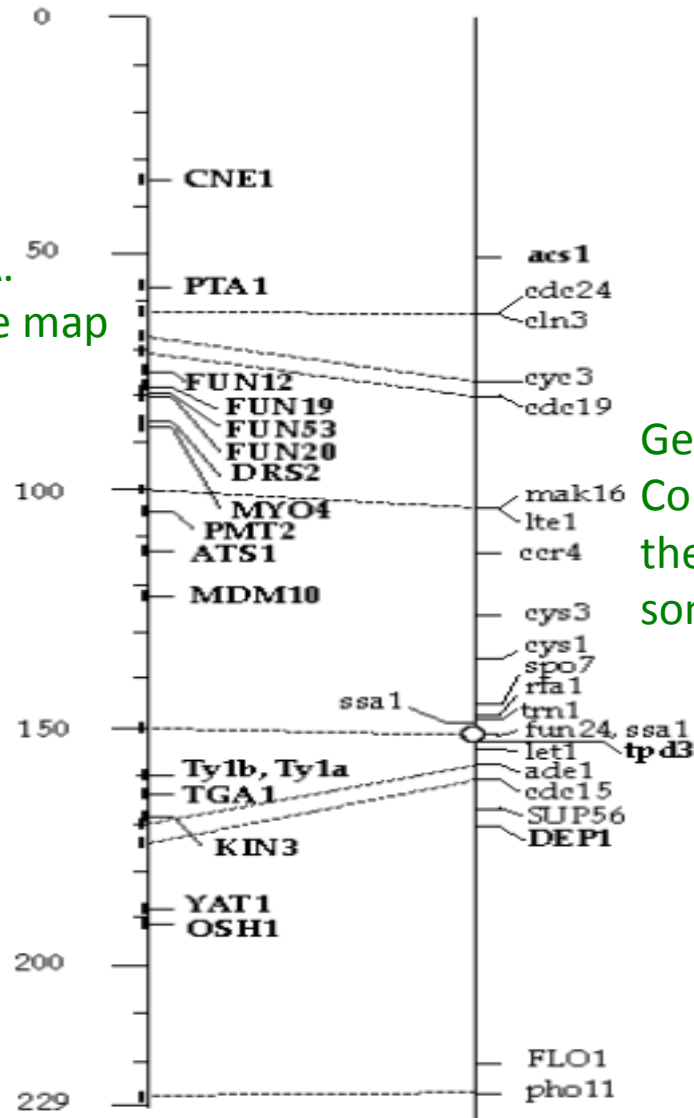


Figure 2. A Partial Genetic Map of Linkage Groups III and IV

# Physical and genetic maps of yeast chromosome I

I

Physical map. ~250kb of DNA.  
Some genes positioned on the map



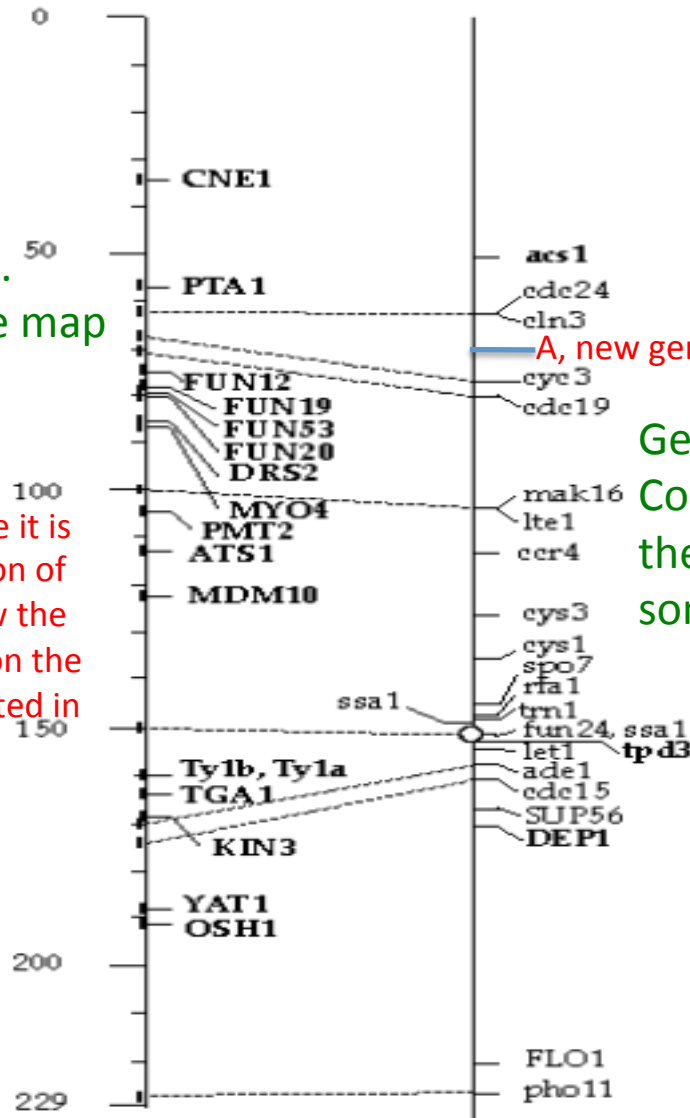
Genetic map. ~100cM.  
Correspondence between  
the two maps made for  
some genes

# Physical and genetic maps of yeast chromosome I

I

Physical map. ~250kb of DNA.  
Some genes positioned on the map

By mapping the gene, you know where it is on the physical map, the representation of the DNA, which has been cloned. Now the challenge is to figure out which gene on the physical map is the one you're interested in



A, new gene you've discovered

Genetic map. ~100cM.  
Correspondence between the two maps made for some genes

# Lexicon and Tools for Human Genetics

One goal is to be able to clone a disease gene of interest

To do so, we need physical and genetic maps and connections between the two

**Physical map:** Depiction of the 3 Billion bp of DNA with location of known sequences along the map.

The sites are usually anonymous, usually without any biological meaning.

They're just identified position markers, called Sequence Tagged Sites (STS), that can be assayed by PCR

**Genetic map:** Depiction of the location of sequence variation along each chromosome.

The sequence variation may be disease genes (or other physiological variation) that have been discovered. The sequence variation may also simply be at anonymous loci that happen to be polymorphic. Examples are (CA)<sub>n</sub> repeats and single nucleotide polymorphisms (SNP).

The virtues of these polymorphic markers are that they are abundant – millions in the genome – so they can be used to make a high quality genetic map and they are also STSs so they can be used to connect the genetic and physical maps

# Sequence Tagged Sites

## Standard STS

Useful for physical map.

Can ask whether site is present on any cloned DNA by PCR amplification



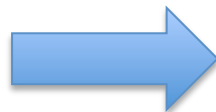
Ccatgatcgaagtctt...100-300nucleotides...ggtaactactgaata  
Ggtactagcttcagaa... ..ccattgatgacttat



## Polymorphic STS

Useful for genetic (and physical) map

Examples include SNPs (single nucleotide polymorphisms) and simple sequence repeats, such as (CA)<sub>n</sub> repeats



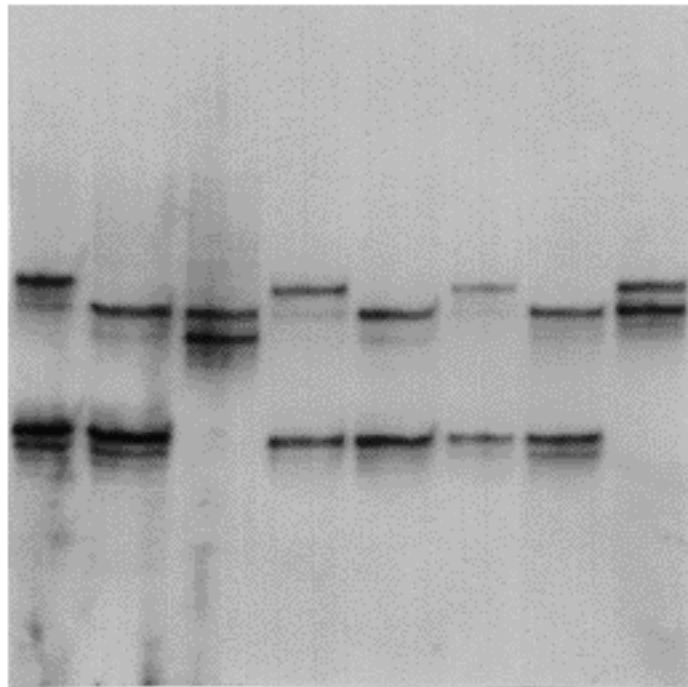
Ccatgatcgaagt<sup>g</sup>tt...100-300nucleotides...ggtaactactgaata  
Ggtactagcttca<sup>c</sup>aa... ..ccattgatgacttat



## (CA)<sub>n</sub> Repeats as examples of polymorphic loci

One allele: ccctgaaacgctt(CA)<sub>15</sub>tttgtaactaacggtt

Another allele: ccctgaaacgctt(CA)<sub>18</sub>tttgtaactaacggtt



How would you look for new CA repeat loci?

The (CA) repeats are flanked by unique DNA. PCR primers can be designed to the unique DNA. PCR reactions followed by gel electrophoresis allows the (CA) repeat alleles to be visualized.

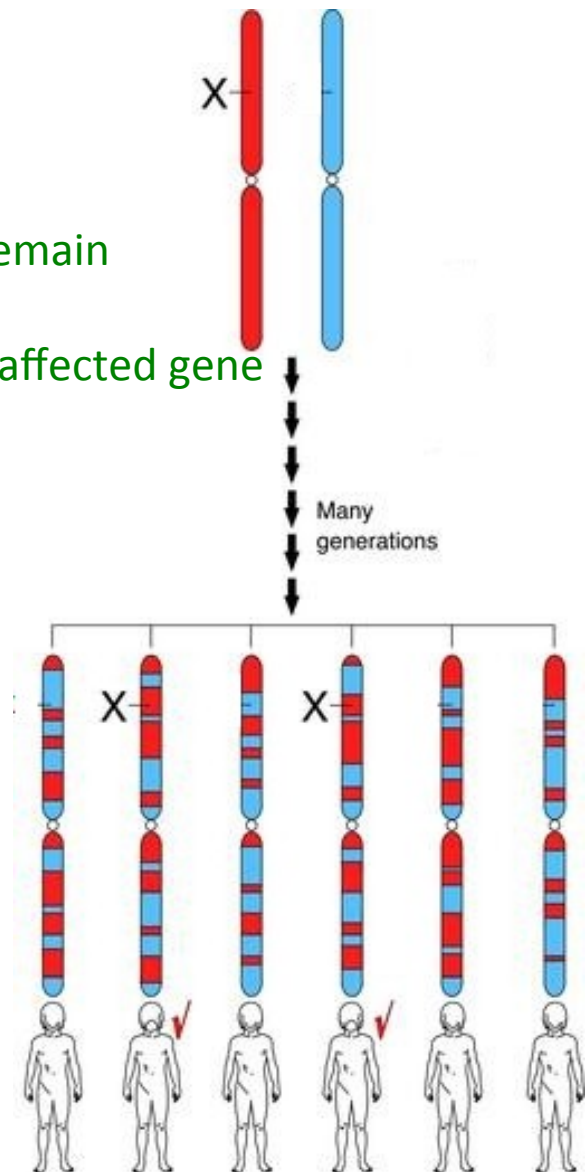


# Haplotypes

Nearby polymorphisms tend to be inherited together, creating “haplotype” blocks

Hence, a “founder” mutation that causes a disease will remain linked to nearby polymorphisms. Identification of these polymorphisms provides a way to clone and identify the affected gene

HAPLOTYPE PATTERNS	
Person A	ATTGAT <span style="border: 1px solid black; padding: 2px;">CGGAT...CCATCGGA...CTAA</span>
Person B	ATTGAT <span style="border: 1px solid black; padding: 2px;">AGGAT...CCAGCGGA...CTCA</span>
Person C	ATTGAT <span style="border: 1px solid black; padding: 2px;">CGGAT...CCATCGGA...CTAA</span>
Person D	ATTGAT <span style="border: 1px solid black; padding: 2px;">AGGAT...CCAGCGGA...CTCA</span>
Person E	ATTGAT <span style="border: 1px solid black; padding: 2px;">CGGAT...CCATCGGA...CTAA</span>





# Human Genome Project and a Physical Map

The human genome is huge (compared to anything that had been tackled previously). One strategy was to break it into smaller and smaller bits until DNA was a “comfortable” size.

Create libraries in vectors that can accommodate different sizes of DNA

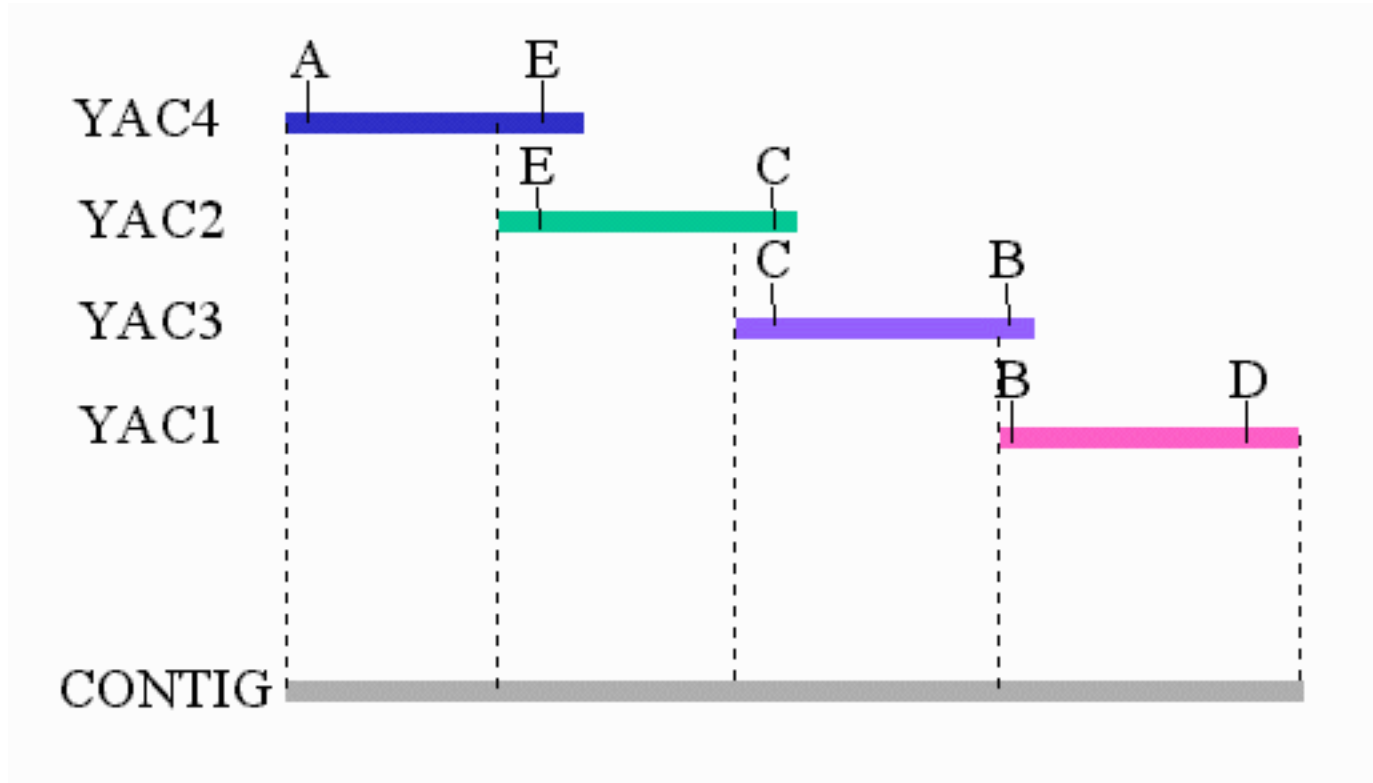
Yeast Artificial Chromosomes (YAC). 1 – 20 M bp

Bacterial Artificial Chromosomes (BAC). 100 – 500kb

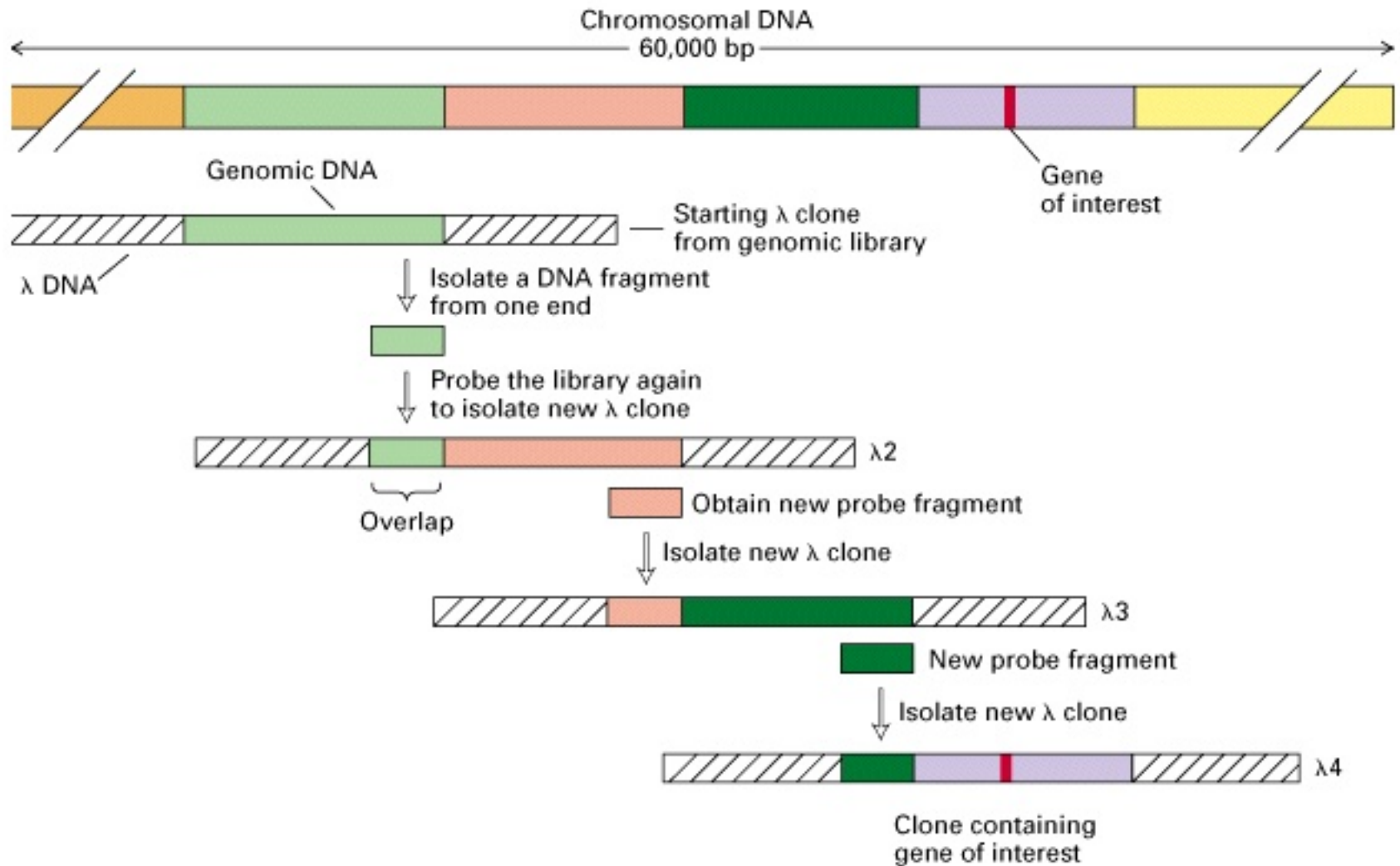
P1 Artificial Chromosomes (PAC). 50 – 100kb

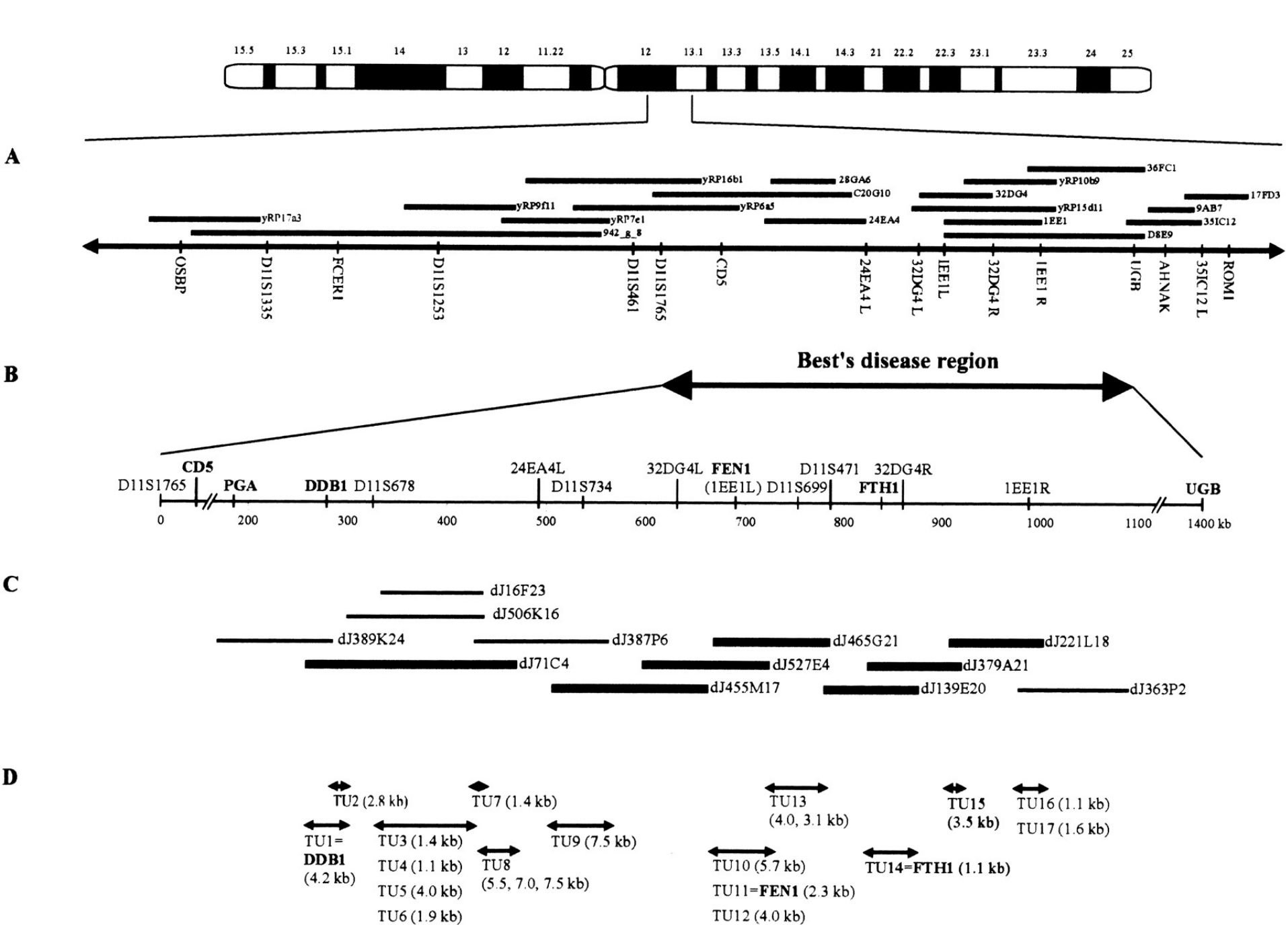
# Contigs

Use the PACs, BACs, and YACs, together with STSs, to link the cloned segments together, thereby recreating the genome, at least *in silico*



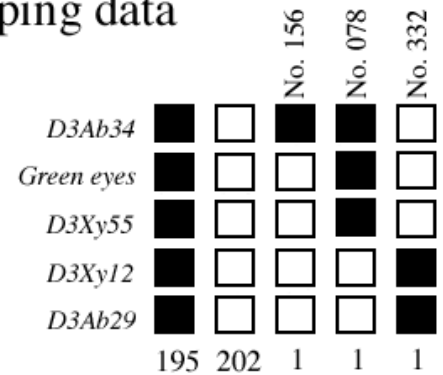
# Chromosome Walking





# The overall strategy

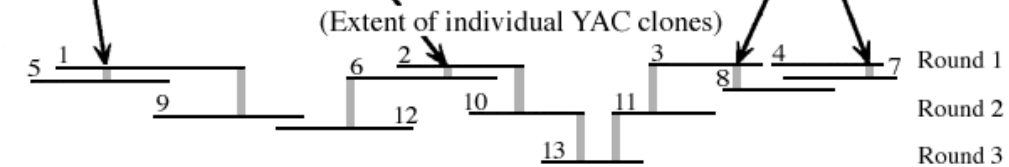
## A. Genotyping data



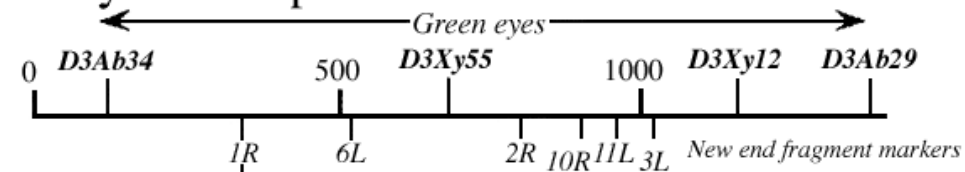
## B. Linkage Map



## C. Library screening and walking



## D. Physical Map



## E. Further mapping of crossover sites and enhanced localization of the *Green eyes* locus



# HNPCC

An inherited predisposition to colorectal cancer

Maps to chromosome 2p16

Lies between two genetic markers 25cM apart

1cM is roughly 1M bp, so clearly need more markers  
to delimit the location of HNPCC to a  
smaller, manageable region

# Generation of new polymorphic markers

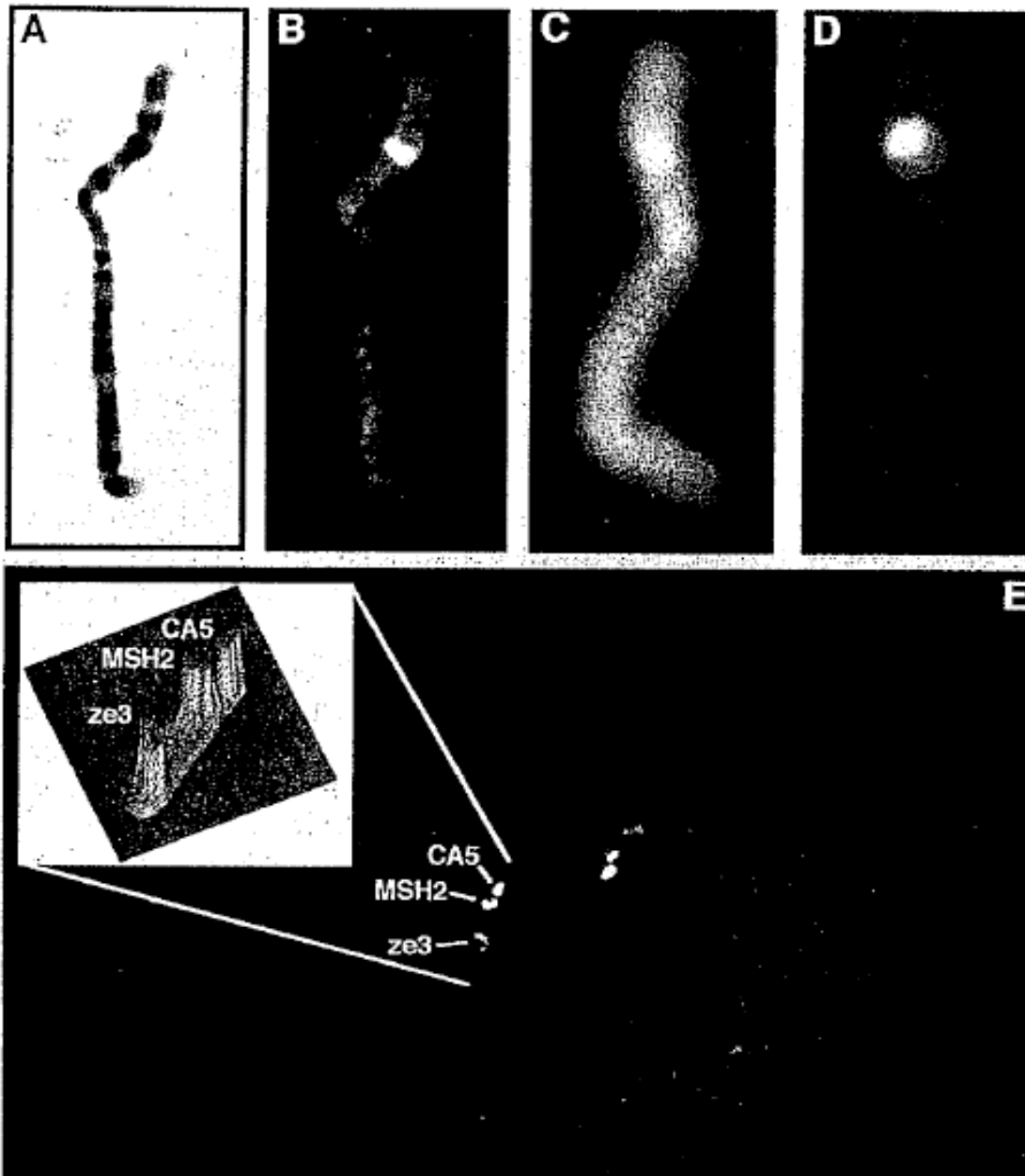


Figure 2. FISH Analysis to Determine the Proximity and Ordering of DNA Sequences within Chromosome Band 2p16

(A, B) FISH mapping of the 123 marker.

(A) G-banded metaphase chromosome 2.

(B) Identical chromosome as in (A) following FISH with a biotin-labeled P1 clone for the 123 marker. Results localize the 123 marker to chromosome band 2p16.

(C, D) Cohybridization documenting the coincident localization of a microdissection (Micro-FISH) probe from chromosome 2p16 and the 123 marker.

(C) DAPI-stained metaphase chromosome 2.

(D) Simultaneous hybridization of the biotin-labeled 123 probe (appearing as an intensely staining smaller circle) and the Spectrum-Orange labeled 2p16 Micro-FISH probe (appearing as a diffusely staining larger circle).

(E) Representative example of an interphase nucleus simultaneously hybridized with P1 clones for CA5, *hMSH2*, and *ze3*. The results were used to directly measure the distances between markers to establish the order and relative distance between markers (according to Trask et al., 1989).

Inset: The image processing program NIH Image was used to provide an average gray value displayed as a surface plot to support the length measurements and to graphically illustrate the relative order information. The surface plot presented defines the specified interphase chromosome and the relative order CA5-MSH2-ze3.

Can use cloned DNA from 2p16 to look for new polymorphic markers; challenge is to order them

# Somatic Cell Hybrids

Fuse human and mice (or other rodent) cells`

Human chromosomes are unstable in such hybrids and are lost; eventually the hybrid karyotype stabilizes

Can further alter karyotype such that only part of a particular human chromosome is present, for example by X-ray treatment



# Somatic Cell Panel for Quick Mapping of New Polymorphic Loci

## SOMATIC CELL HYBRIDS

Previous work had mapped the HNPCC locus to between markers 119 and 136

MARKERS

Z4 (V15-27)    Z11 (6CS5)    Z29 (HD(2)11)    Y3 F(2n)-21    Z12 (6CS7)    L1 (AG3A9)    L2 (AG3A10)    Z19 (XHB78)

MARKERS	Z4 (V15-27)	Z11 (6CS5)	Z29 (HD(2)11)	Y3 F(2n)-21	Z12 (6CS7)	L1 (AG3A9)	L2 (AG3A10)	Z19 (XHB78)
TEL								
177								
119								
yh5								
ze3								
CA5								
CA7								
123								
CA2								
CA18								
CA16								
yb9								
xf5								
tf1								
147								
136								

GEN

DERIVATION: M M M M M T T X  
 BACKGROUND: HA HA HA RA HA MO MO HA

New polymorphic markers can be mapped quickly to regions by hybridization to the somatic cell panel

# Linkage Analysis of Polymorphic Markers using CEPH Families

Table 1. Polymorphic Markers Used for Linkage Analysis

Marker	Derivation	cM	Lod	Heterozygosity	YAC Clones	P1 Clones
177 (AFM267zc9)	T	---		0.84		
119 (AFM077yb7)	T	6.1	5.5	0.77	11E1	406
yh5 (AFM337yh5)	T	6.4	15.4	0.76	4F4, 1E1	838, 839, 840
ze3 (AFM200ze3)	T	0.0	---	0.61	4F4, 1E1, 9H6, 4A10	836, 837
CA5 (CA5)	M	2.1	4.7	0.77	7F10, 4E2, 5A11	820
CA7 (CA7)	M	1.7	3	0.78	6B8	
123 (AFM093xh3)	T	2.4	9.9	0.76	3D11, 8C7	210, 211
CA2 (CA2)	P	0.0	---	0.75	3D11, 8C7	210, 211
CA18 (CA18)	M	4.3	17.1	0.71	8E5	
CA16 (CA16)	M	0.0	---	0.69	8E5	
yb9 (AFM320yb9)	T	1.1	3.9	0.80	264	
xf5 (AFM310xf5)	T	2.7	17.6	0.76		
tf1 (AFM348tf1)	T	0.0	---	0.79		
147 (AFM199vb6)	T	2.6	9.8	0.73		
136 (AFM172xe7)	T	2		0.73		
134 (AFM168xg11)	T	1		0.76		387, 388, 389

Each of the markers listed was found by screening a genomic library with a (CA)<sub>n</sub> probe. The libraries were made from total genomic DNA (T), microdissected human chromosome 2p16 (M), or a genomic P1 clone containing marker 123 (P). The laboratory name and the formal name (in parentheses) of each marker is listed. The genetic distance between the indicated marker and the marker listed above it was determined by linkage analysis in CEPH families, as was the heterozygosity. The odds for pairwise inversion of loci is given in all cases in which these odds were greater than 1000:1. Those YAC and P1 clones obtained in this study that contained marker sequences are also listed.

Together with the somatic cell hybrids, these data give an unambiguous order for the markers

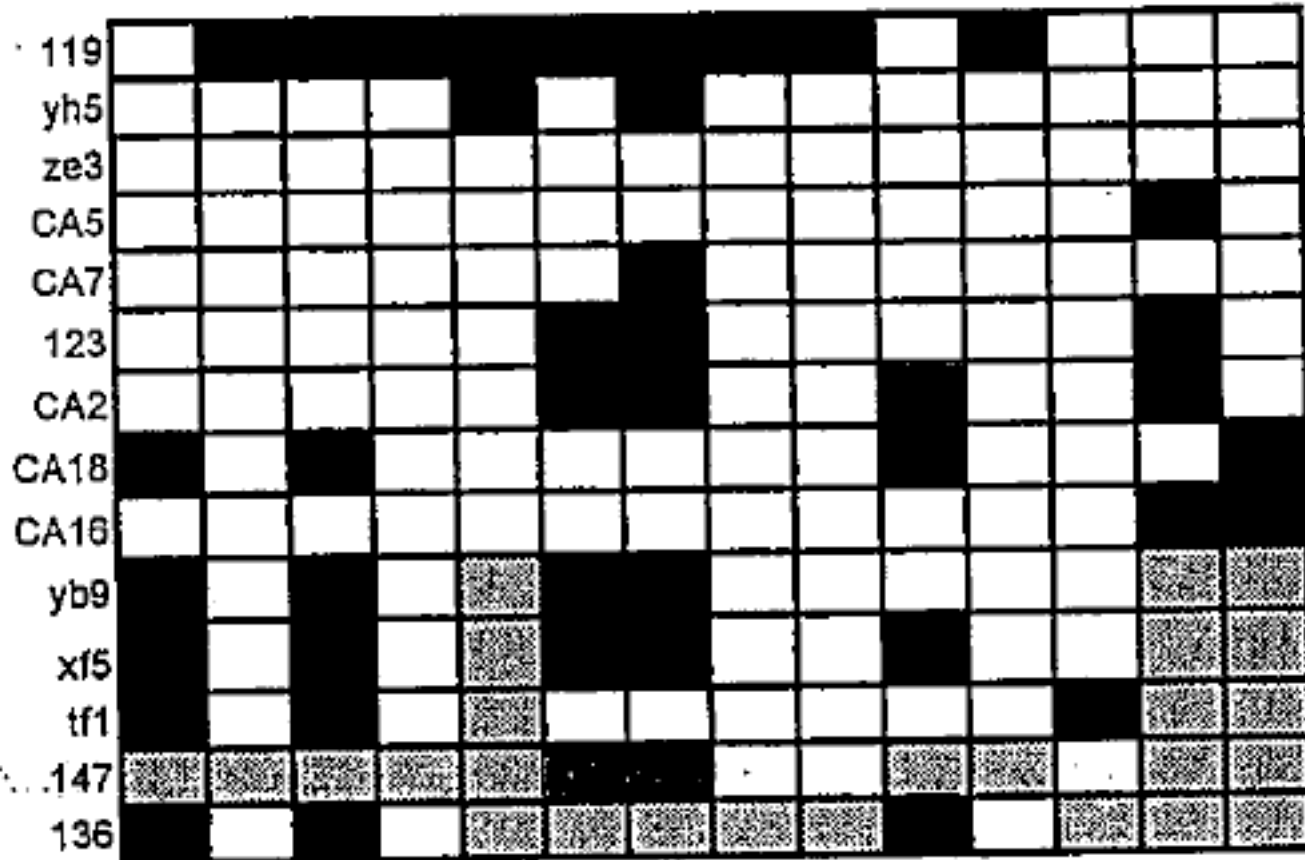
Pattern of markers in affected individuals in families suggests HNPCC locus lies in ze3 to CA5 region

KINDRED	C	C	C	J	J	4	4	8	8	K	K	621	621	621
INDIVIDUAL	74	43	202	36	148	156	92	67	309	23	37	182	44	119

\*

\*

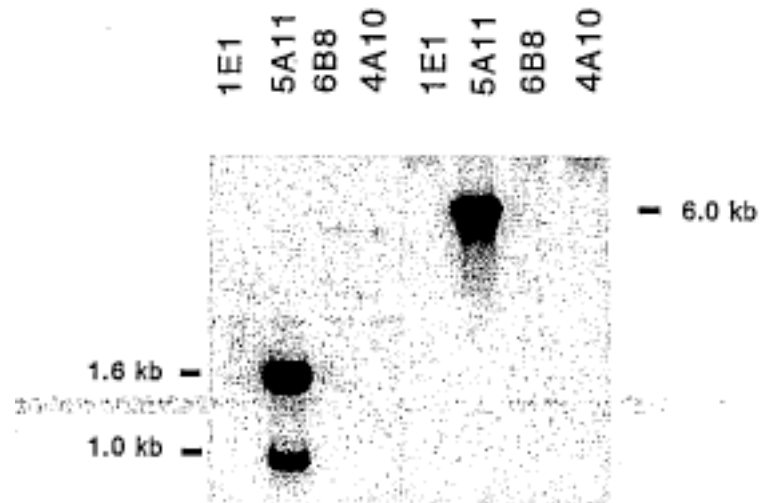
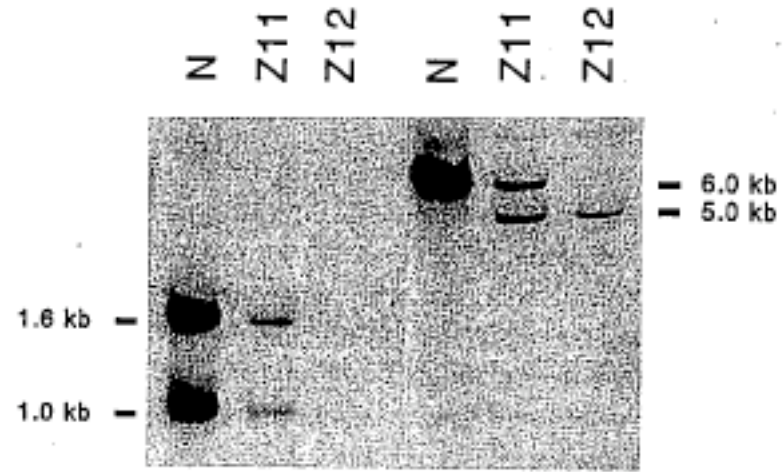
M A R K E R S



White box: presence of allele associated with disease

Black box: absence of allele associated with disease

# Test Candidate Genes: Homolog to *MSH2* hybridizes to somatic cell hybrids and YACs







## Affected individuals have mutations in the *hMSH2* gene

Table 2. Summary of Mutations

Sample	Source	Type	Codon	cDNA Nucleotide Change	Predicted Coding Change
Family J	HNPCC kindred	Germline	622	CCA to CTA	Proline to leucine
Family C	HNPCC kindred	Germline	265-314	793 to 942 deletion	In-frame deletion
Family 8	HNPCC kindred	Germline	406	CGA to TGA	Arginine to stop
Cx10	RER <sup>+</sup> tumor	Germline	639	CAT to TAT	Histidine to tyrosine
		Somatic	663	ATG to TGTG	Frameshift

Mutations were detected by sequencing PCR products, and each was confirmed by replicate PCR analyses and by examination of other affected members of the kindred (see text). The indicated germline mutations were each heterozygous; i.e., lymphocyte DNA contained one wild-type allele and one mutant allele. In Cx10, cloning of the PCR products showed that the somatic mutation at codon 663 occurred in the allele not affected by the germline mutation at codon 639.

# Haplotype analysis of NBS patients

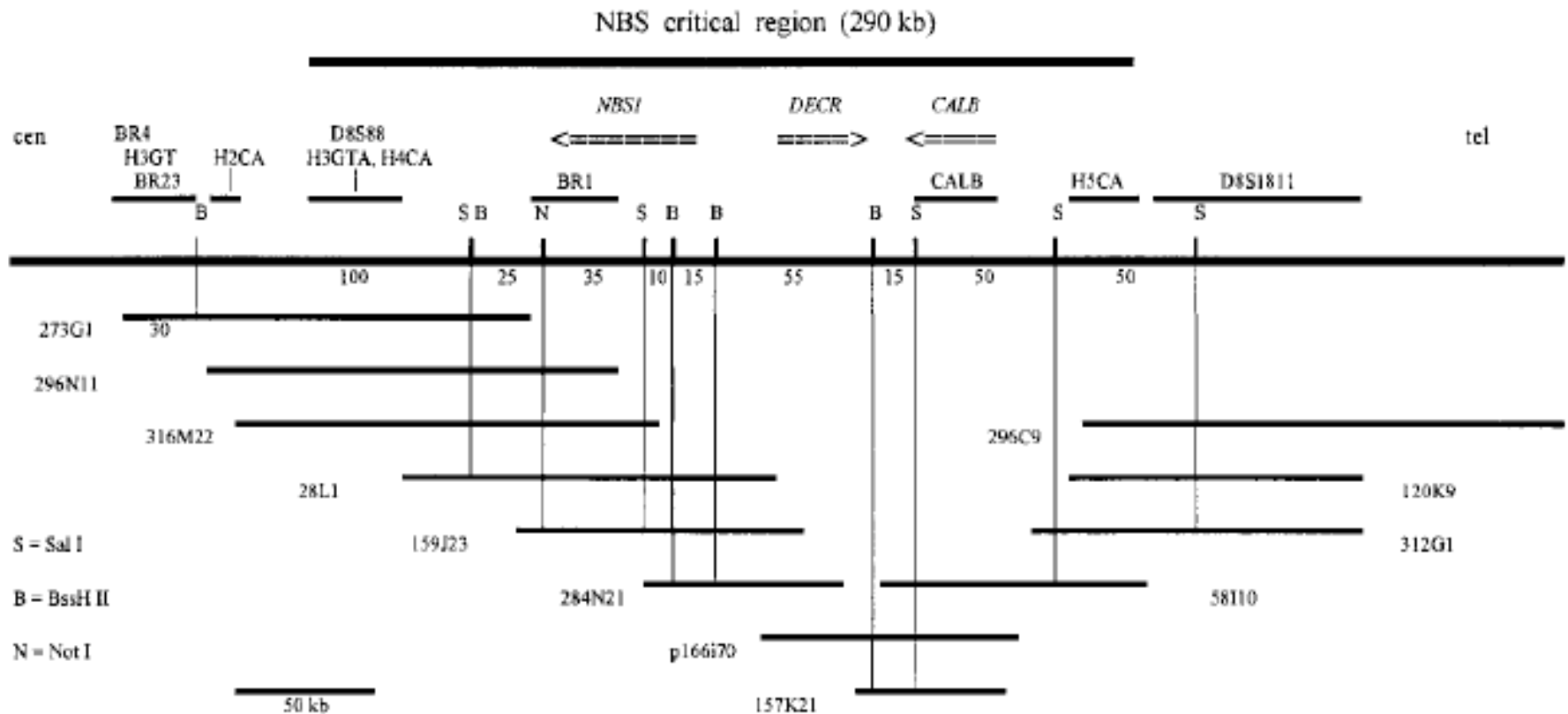
	D	D	A		D	H	D	H	D	D	D	D			
	A	S	P		S	H	S	S	S	A	S	S			
	G	T	M	H	B	H	G	G	H	C	H	S			
	T	G	R	G	T	C	A	T	C	T	L	C			
	1	2	3	4	5	6	7	8	9	10	11	12			
P1 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P1 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P13 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P13 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P21 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P25 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P26 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P17 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P2 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P21 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P14 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P13 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P20 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P25 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P32 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P32 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P19 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P4 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P24 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P4 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P3 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P2 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P9 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P16 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P18 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P46 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P43 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
GM7076 Y	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P15 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P44 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P5 M	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P12 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P42 Y	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P23 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P6 F	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P47 Y	251	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P10 M	256	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P10 F	256	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P20 C	246	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P24 Y	256	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P22 Y	256	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P29 M	256	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P26 F	256	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P34 F	256	145	180	2	216	2	232	90	174	246	2	2	108	142	200
94P126 Y	256	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P7 F	266	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P4 F	256	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P16 M	266	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P16 F	267	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P18 F	259	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P20 M	259	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P44 M	229	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P22 T	260	145	180	2	216	2	232	90	174	246	2	2	108	142	200
P18 F	255	138	180	2	216	2	232	90	174	246	2	2	108	142	200
P26 F	216	138	180	2	216	2	232	90	174	246	2	2	108	142	200
P24 M	261	138	180	2	216	2	232	90	174	246	2	2	108	142	200
P26 F	236	138	180	2	216	2	232	90	174	246	2	2	108	142	200
P24 Y	261	138	180	2	216	2	232	90	174	246	2	2	108	142	200
9230 Y	260	141	180	2	216	2	232	90	174	246	2	2	108	142	200
P1 M	270	138	180	2	216	2	232	90	174	246	2	2	108	142	200
P17 M	267	141	180	2	216	2	232	90	174	246	2	2	108	142	200
9230 T	255	138	180	2	216	2	232	90	174	246	2	2	108	142	200
P25 F	267	141	180	2	216	2	232	90	174	246	2	2	108	142	200
P41 F	268	138	180	2	216	2	232	90	174	246	2	2	108	142	200
925-1.1 C	266	132	181	2	216	2	232	90	174	246	2	2	108	142	200
P3 F	251	145	181	2	216	2	232	90	174	246	2	2	108	142	200
94P126 Y	260	141	181	2	216	2	232	90	174	246	2	2	108	142	200
P23 M	256	145	181	2	216	2	232	90	174	246	2	2	108	142	200
P46 F	251	145	181	2	216	2	232	90	174	246	2	2	108	142	200
P41 M	226	128	181	2	216	2	232	90	174	246	2	2	108	142	200
P43 M	251	145	181	2	216	2	232	90	174	246	2	2	108	142	200
GM7075 T	251	143	180	2	216	2	232	90	174	246	2	2	108	142	200
P24 F	261	141	180	2	216	2	232	90	174	246	2	2	108	142	200

# Haplotype analysis, zoomed

	D	D	A		H	B	H	D	H	<i>NBS</i>			D	D	D
	S	S	F					S	3	H	C	H	S	S	S
	2	8	2	B	3	R	2	8	G	4	B	5	8	7	2
	7	0	8	R	G	2	C	8	T	C	R	L	1	2	7
	1	0	9	4	T	3	A	8	A	A	1	B	A	4	0
F1 F	257	145	189	2	216	2	232	90	174	246	2	2	108	142	200
F7 M	257	145	189	2	216	2	232	90	174	246	2	2	189	108	142
F12 M	257	145	189	2	216	2	232	90	174	246		2	189	108	142
F13 F	257	145	189	2	216	2	232	90	174	246	2	2	108	142	
F21 M	257	145	189	2	216	2	232	90	174	246		2	108	142	200
F25 M	257	145	189		216	2	232	90			2	2	108	142	200
F39 M	257	145	189		216		232	90	174	246	2		108	142	200
F17 F	257	145	189		216	2	232	90	174	246	2	2	189	108	142
F2 F	257	145	189	2	216	2	232	90	174	246		2	110	142	200
F21 F	257	145	189	2	210	2	230	90	174	246	2	2	108	142	200
F14 M	257	145	189	2	216	2	232	82	174	246	2		108	142	200
F13 M	257	138	189	2	216	2	232	90	174	246	2	2	108	142	
F35 M			189		216	2	232	90				2	108	142	
F35 F			189		216	2	232	90				2	108	142	
F32 M			189					90					108	142	
F32 F			189					90					108	142	
F19 M	257	145	189		216	2	232	90	174	246			108	142	198



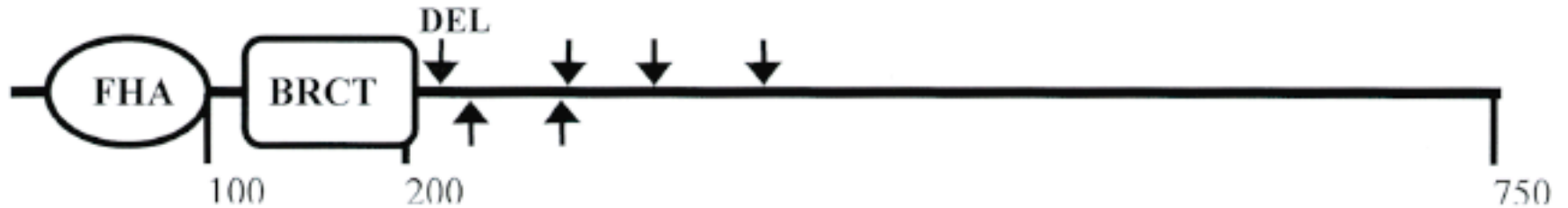
## BAC/PAC contig of NBS region



NBS patients have mutations in the *NBS1* gene.  
In unrelated patients, 6 different mutations were identified.

# Homologs from yeasts to humans to plants

A



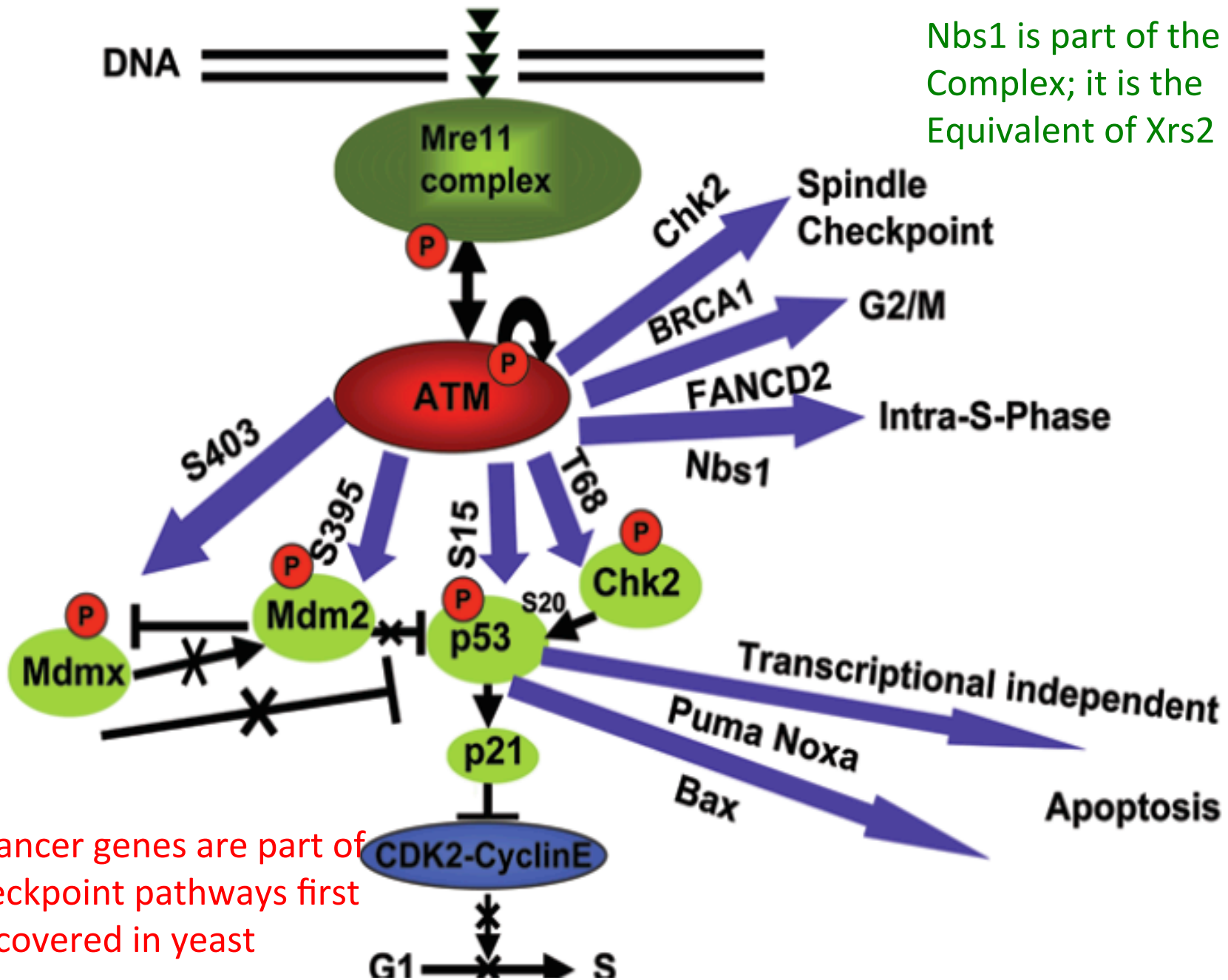
B

Arrows mark the positions of mutations in NBS patients

hNIBRIN	24	YVVGKRN...LAILLEN...DQSI SRNHA VITAN.13.VITLKN.SKYGTFVNE...EKMQNGFS...RTLKSGLGTYG
hKI67	27	CLFGRGIE...CDIRIQL...PVVSKQKCKIEH..2.EAILHNFSSTNPTQVNG..SVIDEP...VRLKHGIVTYI
ScDUN1	56	PTIGRSRS...CDVILSE...PDISTFHAEPHLL.10.LINVIDK.SRNGTFING..NRLVKKD...YILKNGDRVFG
SpCDS1	60	WGFGRHKS...CEVVLNG...PRVSNFHFYIYQG.10.VVFLHDH.SSNGTFINP..ERLAKNSR...TILSNGDEIRIG
ScMEK1	47	VKVGRNDK..1.CQLVLTN...PSISSVHCVPWCV..8.MFYVKDC.SLNGTYLNG..LLKRDKT...YLRKHCIVIELS
ScSPK1	66	WTFGRNPA...CDYHLGN...ISRLSNKHFIYLLG..3.NLLLNDI.STNGTWLNG..QKVEKNSN...QLASQGDYITVG
ScFKH1	76	VTIGRNTD.15.IDIDLGP...AKIVSRKHAAREN..4.SWELQIF.GRNGAKVNF..RRIPTGPDSPPTVLQSGCIDIG
ScFHL1	300	AITCRRSE..6.VDVNLGP...SKSISRRAQIFYN..3.RFELSII.GKNGAFVDD..IFVEKGNT...VPLRNKTKIQIG
AtKAPP	209	VKLGRVSP...SDLALKD...SEVSGKHAQITWN..4.KWELVDMGSLNGTLVNS..HSISHPDL.8.VELASDDIITLG
AsFRAH	204	VHIGKPNL..4.IDVDVSGFANSEIVSRVHADIRLE...AHYLEDVGSNGTYINN..LPLLPGR...HRLRPGDRISLG

C

hNIBRIN	109	EYEPLVAC..SSCLD.VSGKTA.NQA.LQLGGFTVNN...WTEECTHLVMS...VKVTIKTICALICRPVVKPEYTFEFKAVESKKQPPQIES
CeP37D6	972	AMNPRFLLSVSNMD.PQRAADLQETIMKLGCTIERE...FNKDVTHLIASN...MQRAPKVLCSLAAGKWCITPDYVTKSAEV.GRWLDEKSFEN
AtT10M13	1105	EHEPKFFY.VSGPR..SQRNEYQQIIRRLKKGKCRDSDHQSYSQATHFIAPE...IRRTEKFFAAAASCSWILKTDYVADSKEA.GKLLQEEPVEW
hCAGF28	539	ELTPFVLF..TGFE.FVQVQQYIKKYLIGGEVAES...AQKCTHLIASK...VTRTLKFLAAISVVKHIVTPEWLEECFRC.QKFIQEQNYIL
hBARD	566	RDGPLVLI.GSGLS.SEQQKMLSELAVILKAKKYTE...FDSTVTHVVVPG..2.VQSTLKCMLGILNGCWILKFEWVKACLRK.KVCEQEKEYEI
hXRCC1	320	LQGVVVVL..SGFQ.NPFRSELDRKALELGAIRPD...WTRDSTHLICA...FANTPKYSQVLGLGRVVRKQWLDCHRM.RRRLPSRRYLA
mECT2	1	MLNLVLCF..TGFRKKEELVKLVTLVHVMGGVIRKE...CNSKVTHLVAN...CTQGEKFRVAVSLGTPDKPFIYKAWERRNEQCFCAAVDD
hBRCA1	1647	NKRMSMVV..SGLT.PEFPMLVYKFAKHHITLNL...PTEETHVVMKT..5.CERTLKYFLGLAGCKWVSYFWVTQSIKE.RKMLNEHDFEV
ScRAD9	997	VFDKCFIV.LTSLP..ENREELRQTIESQGGIVIESGFSILFNPTHPLAKS.38.HLRSLKYLETLALQWPTLEHWKPI SACHEK.KRIVPHLIVQY
ScREV1	166	FKNCVIYI..NGYT.KPGRLOHEMIVLHGKRLHYLS.SKKTVTHIVASN...LPLKKRIEPA...NYKVVSPDWIVDSVKE.ARLLPWQNYSL



Nbs1 is part of the Mre11 Complex; it is the Equivalent of Xrs2

2 cancer genes are part of checkpoint pathways first discovered in yeast