

GLOSSARY OF TERMS

Anonymous DNA: a length of DNA of unknown gene content.

Chromosome bands: alternating light- or dark-staining sections along a chromosome that are visible by light microscopy after staining procedures are used.

Chromosome walk: a method of aligning pieces of DNA by consecutive hybridizations in which probes corresponding to one end of a cloned piece of DNA are used to identify the next clone in line.

Contig: a set of overlapping pieces of DNA that span an uninterrupted stretch of the genome.

Cosmid: a piece of DNA (35 to 45 kb) cloned into a vector that usually consists of *cos* sequences required for packaging the DNA into a bacteriophage, an origin of replication, and a drug resistance marker.

Genetic linkage map: a map that shows the relative position of loci on the basis of frequency of recombination events. Units are in centimorgans (cM) where, over small distances, 1 cM is equivalent to a 1% chance of recombination.

Physical map: a map in which the distances between landmarks such as clones, restriction endonuclease sites, or specific loci are expressed in kilobases.

Polymerase chain reaction (PCR): a technique that involves repeated cycles of DNA denaturation, renaturation with short lengths of DNA (primers) separated by up to 4 kb, and polymerase-mediated replication. This results in an exponential increase in the number of copies of the sequence between the primers.

Polymorphic marker: a locus at which there is normal sequence variation within the population that is inherited and occurs with a frequency of >1%.

Polytene chromosome: a giant chromosome consisting of many identical, slightly condensed strands of chromatin held together in parallel and in register.

GENOME

The wall chart this year is devoted to genome mapping for many of the chromosomes. This year's chart shows a summary of progress to date.

RECENT PROGRESS

Chromosome	1	2	3	4
Number of genes				
Estimated	4150	3950	3200	3050
Mapped	236	131	72	84
Disease-related	43	18	16	22
Number of polymorphic markers				
Total	158	99	333	148
PCR-based	13	15	12	14
Kilobases of sequence	450	558	104	181

This table is a depiction of the current progress on the human genome project as of December 1991. There are an estimated 3,000 megabases (Mb) of DNA sequence mapped, containing an estimated 50,000 to 100,000 genes. Since the last report in 1990 [*Science* 250, 262a (1990)], 359 new mapped genes have been added to the map.

The estimated number of genes is based on the estimated total size of the genome distributed proportionally according to the relative size of each chromosome [Lander et al., 1990, p. 237 (1990)]. The number of genes mapped is the number of genes for which a clone has been identified. The number of disease-related genes is based on locus entries in the *Gene Catalogue Project* (*Inheritance in Man* (Johns Hopkins Univ. Press, Baltimore, MD) and *Gene Catalogue Project*). The number of genes and anonymous DNA segments. They are useful in gene mapping of specific regions within a chromosome. The number of polymorphic markers is based on the number of markers mapped to date.

GENE MAPS 1999

presented an update of progress in the human genome. There have also been major advances in essential in understanding gene function. As a first example, the bottom section of the chart shows the fruit fly, an organism that is a classic model for studies in genetics and development.

GENES IN HUMAN GENE MAPPING

10	11	12	13	14	15	16	17	18	19	20	21	22	X
2200	2200	2050	1800	1750	1650	1400	1350	1250	1150	1050	900	950	2350
75	142	117	28	65	59	71	121	22	104	42	39	71	20
11	33	17	6	12	12	12	18	4	20	11	4	18	9
91	265	69	58	54	55	130	192	35	73	57	71	124	25
4	20	10	4	8	5	8	20	2	12	31	10	4	2
158	655	416	82	260	102	309	623	94	266	124	51	141	44

8 July,
me,
ome Map
s have been

me,
cience 250,
romosome.
ndelian
include
of genes to
norphic

need to distribute cloned probes for linkage or other molecular studies. Kilobases of to known loci are based on analysis of information obtained from GenBank, Los Alamos.

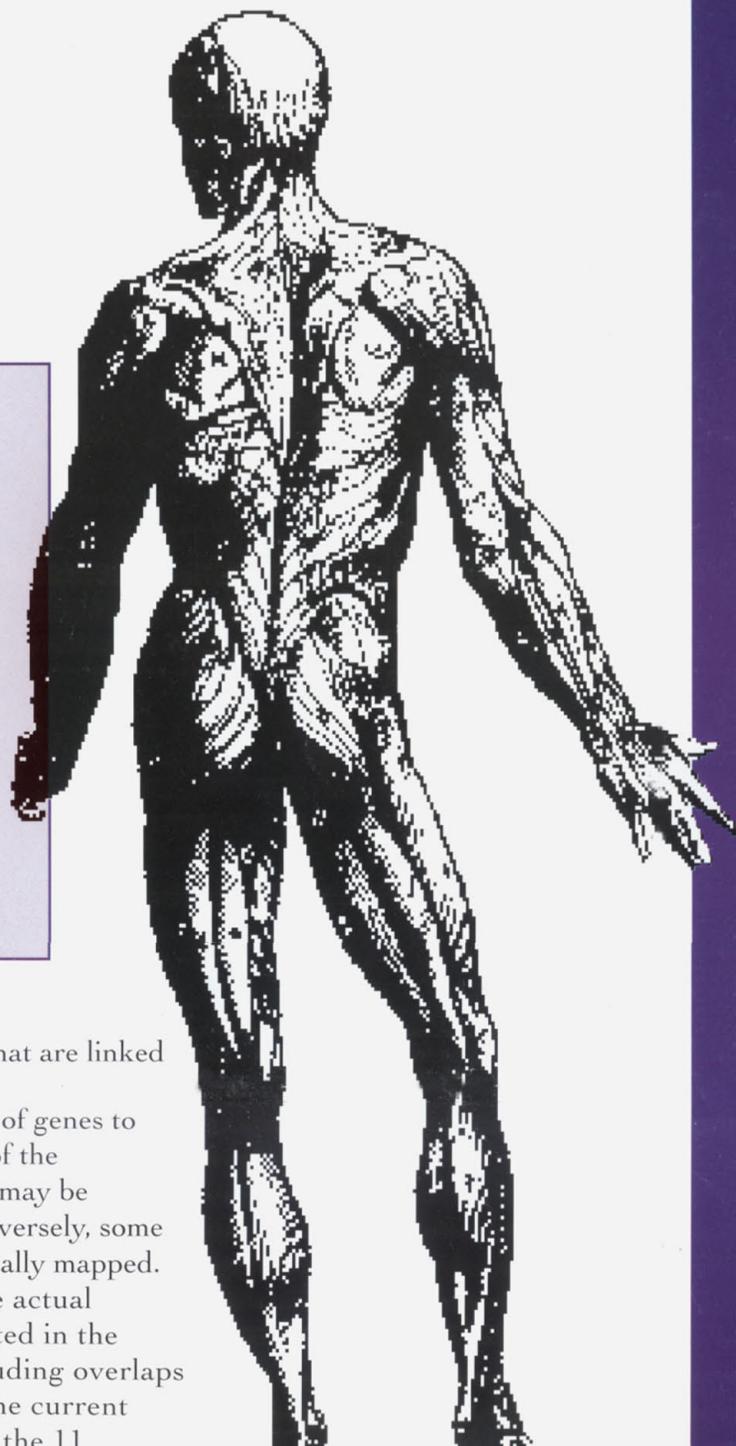
There are some caveats to these analyses. For example, comparison of the estimated number of genes currently mapped to a particular chromosome may not be a precise measure of the completeness of a map on a given chromosome. In some cases, large portions of a chromosome have been physically mapped, although the genetic content of those regions has yet to be established. Chromosomes for which a large number of genes have been described are not necessarily the most gene-rich. In addition, some chromosomes may have a higher density of genes than others. The amount of sequence available for each chromosome is likely to be less than the amount shown in the table, because the amounts given are based on the total amount of sequence reported in the database, rather than only the unique sequences. Further information regarding the status of the human genome map will be presented in the Perspective (P. Pearsons, p. 25).

1991

There have also been major advances in the bottom section of the chart in genetics and development.

GENETICS

19	20	21	22	X	Y
150	1050	900	950	2350	1000
104	42	39	71	202	16
20	11	4	18	93	1
73	57	71	124	255	25
12	31	10	4	24	0
266	124	51	141	441	19



For molecular studies. Kilobases of sequence that are linked obtained from GenBank, Los Alamos, NM. For example, comparison of the estimated number of genes to a particular chromosome may not be a precise index of the gene density. In some cases, large portions of a chromosome may be unsequenced. In those regions has yet to be established. Conversely, some genes that have been described are not extensively physically mapped. The density of genes is higher on some chromosomes than others. Finally, the actual number of genes on a chromosome is likely to be less than the amount stated in the table. The total amount of sequence reported, including overlaps and repeats. Further information regarding the current state of the genome is provided in the Perspective (P. Pearson *et al.*) in the 11

tion of the current progress on the mapping of the human genome as of 28 July, an estimated 3,000 megabases (Mb) of DNA sequence in the human genome, estimated 50,000 to 100,000 genes. Since the publication of The Human Genome Map [Lander *et al.*, 1991], 262a (1990)], 359 new mapped genes and 673 new polymorphic markers have been

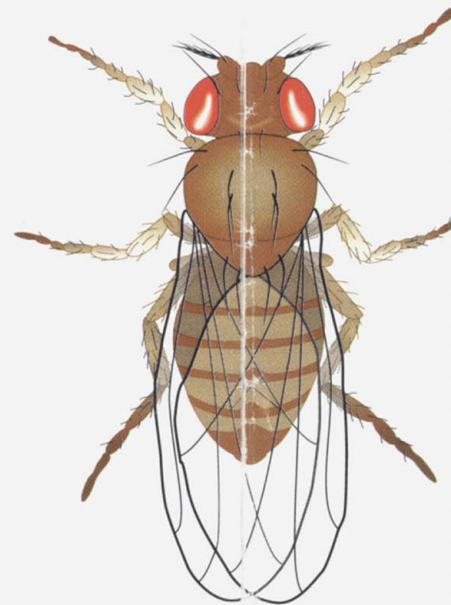
genes is based on the estimated total number of genes in the human genome, according to the relative size of each chromosome [J. C. Stephens *et al.*, *Science* 250, 1992]. The number of genes mapped is the number of genes that have been localized to the chromosome. The number of genes is based on locus entries that are referenced in V. McKusick, *Mendelian Inheritance in Man* (Johns Hopkins Univ. Press, Baltimore, MD ed. 9, 1990). Polymorphic markers include restriction enzyme sites. They are useful in genetic linkage studies for the localization of genes to a chromosome. The number of polymerase chain reaction (PCR)-based polymorphic markers is the number of polymorphic markers that can be detected by PCR. PCR-based markers obviate the

need to distill the genome to known loci. There are many reasons why the number of genes is not a complete measure of genome complexity. Physically mapped genes are not necessarily the same as genes. In addition, the number of genes is a variable amount of information. It is not a table, because it is not in sequence. The status of the genome is October, 1991.

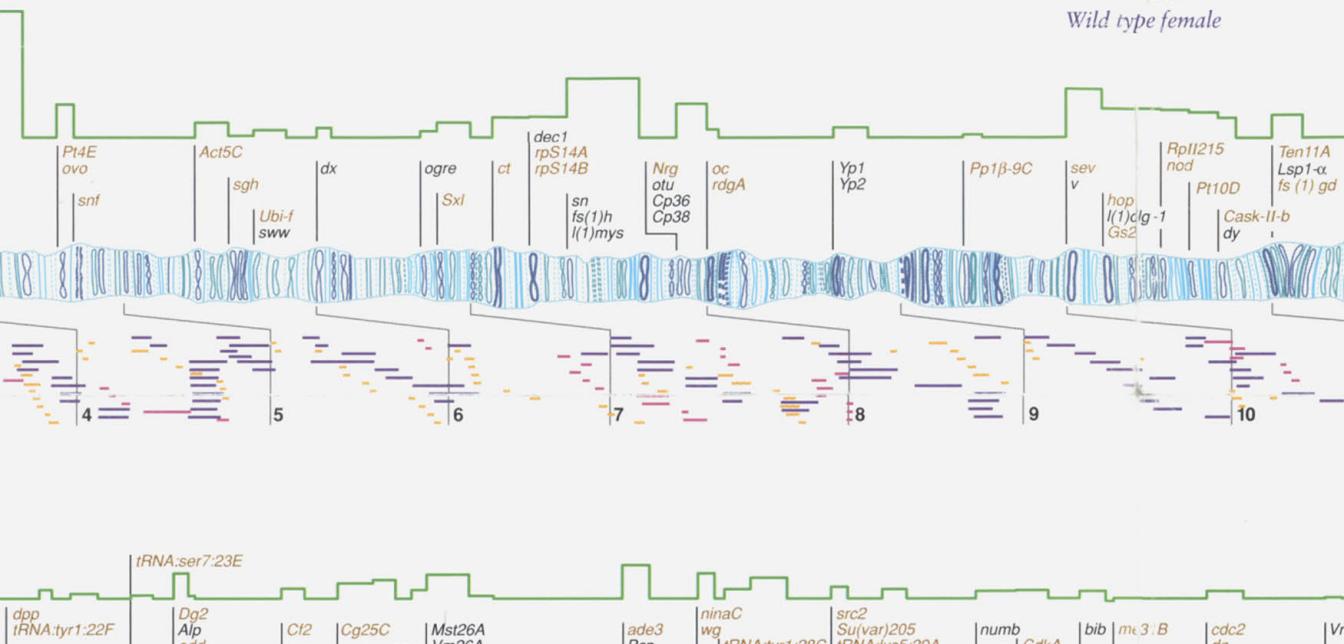
DROSOPHILA MELANOGASTER

The information, sequenced genes, chromosome bands, and DNA clones of the fruit fly, *Drosophila melanogaster*. The chromosomes pictured were taken from a wild type female (by permission); numbers 1 to 102 represent standard chromosome bands. Heterochromatic blocks, which do not polytenize, are not shown. The DNA of which 110 Mb are included in the polytene chromosomes. The number of lethal mutations is more than 15,000 (based on

1 and a graph of kilobases (kb) sequenced per chromosome subdivision [Drosophila Information Service, W. Gelbart, Ed. (Harvard, 1991), 1991]. The subdivisions. *Drosophila* genes that are similar in protein sequence to human genes. Approximately 1.6 Mb of the genome have been sequenced, including genes that overlap in sequence were not counted more than once. The number of genes (and additional genes showing such similarities that have not been



Wild type female



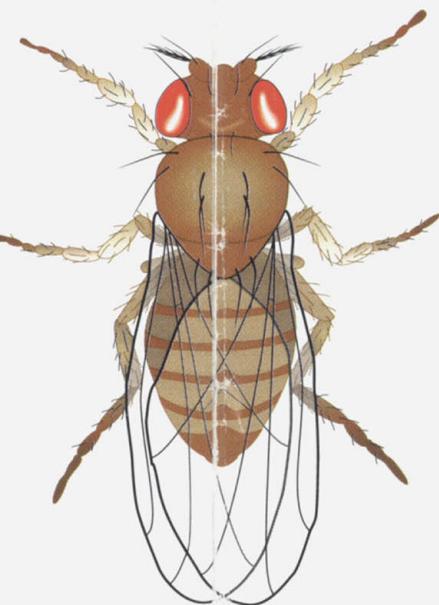
8 July,
me,
ome Map
s have been

me,
cience 250,
romosome.
ndelian
include
of genes to
morphic
viate the

need to distribute cloned probes for linkage or other molecular studies. Kilobases of
to known loci are based on analysis of information obtained from GenBank, Los Alamos

There are some caveats to these analyses. For example, comparison of the estimated
the number of genes currently mapped to a particular chromosome may not be a precise
completeness of a map on a given chromosome. In some cases, large portions of a chromosome
physically mapped, although the genetic content of those regions has yet to be established.
chromosomes for which a large number of genes have been described are not extensive.
In addition, some chromosomes may have a higher density of genes than others. The
amount of sequence available for each chromosome is likely to be less than the amount
table, because the amounts given are based on the total amount of sequence reported
in sequence data, rather than only the unique sequences. Further information on the
status of the human genome map will be presented in the Perspective (P. Pearson
October, 1991 issue of *Science*.

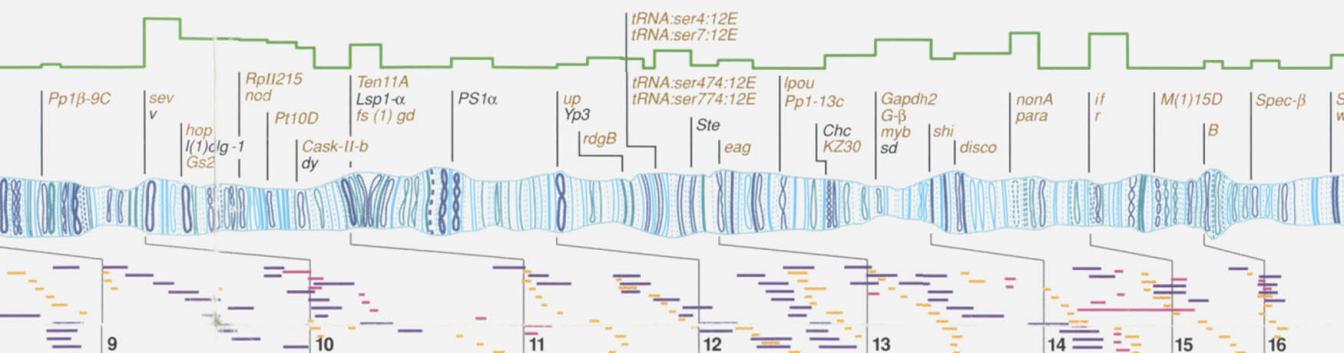
DROSOPHILA MELANOGASTER



Wild type female

mapped) are shown in the table below the chart. Numbers in parentheses
category.

Below each chromosome, the approximate locations and sizes of cloned
chromosome walks larger than 40 kb (magenta) are pictured in proportion to
horizontal scale bar at the left. Grey lines indicate the locations of the DNA
used to relate those locations back to the chromosome bands. The size of
estimated by V. Sorsa [*Chromosome Maps in Drosophila* (CRC Press, Boca Raton, FL),
I. Duncan and D. Hartl (current as of 15 June, 1991); on cosmids by D. S.
Louis, C. Savakis, and M. Ashburner (current as of 1 July, 1991); and by
Lee, and J. Johnsen (current as of 31 May, 1991). DNA localization is based on
the chromosome and the precision of in situ hybridization. The lengths of
shared restriction fragments. YACs, cosmids, and chromosome walks are
information can be found in the accompanying article (J. Merriam *et al.*)

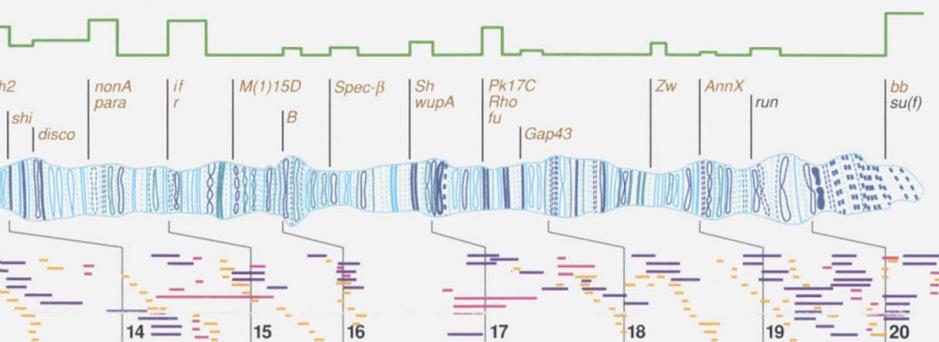


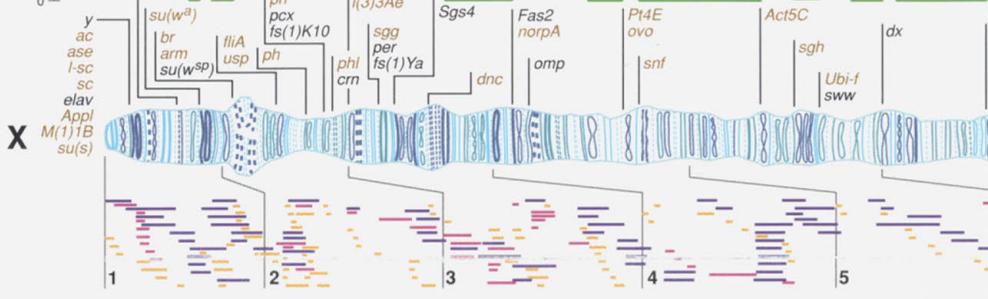
for molecular studies. Kilobases of sequence that are linked
 obtained from GenBank, Los Alamos, NM.
 example, comparison of the estimated number of genes to
 lar chromosome may not be a precise index of the
 some cases, large portions of a chromosome may be
 those regions has yet to be established. Conversely, some
 have been described are not extensively physically mapped.
 er density of genes than others. Finally, the actual
 me is likely to be less than the amount stated in the
 ne total amount of sequence reported, including overlaps
 quences. Further information regarding the current
 ed in the Perspective (P. Pearson *et al.*) in the 11



by the chart. Numbers in parentheses indicate the presence of more than one gene in that

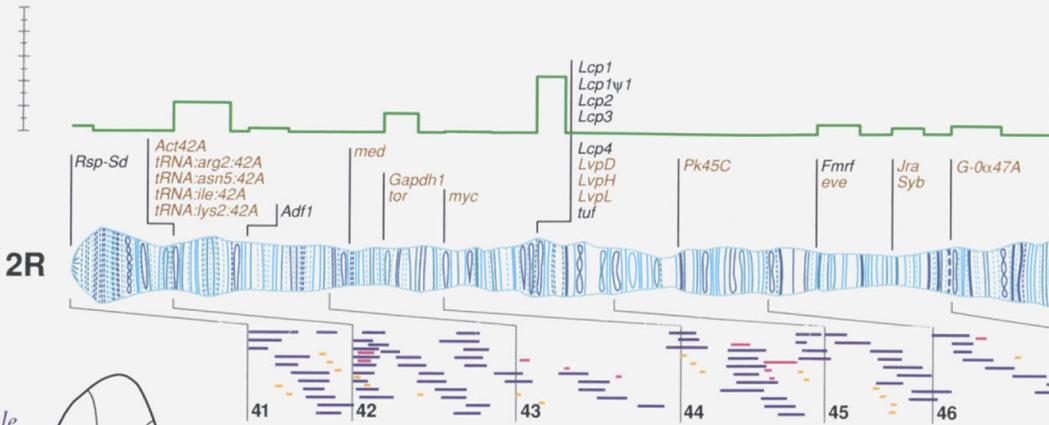
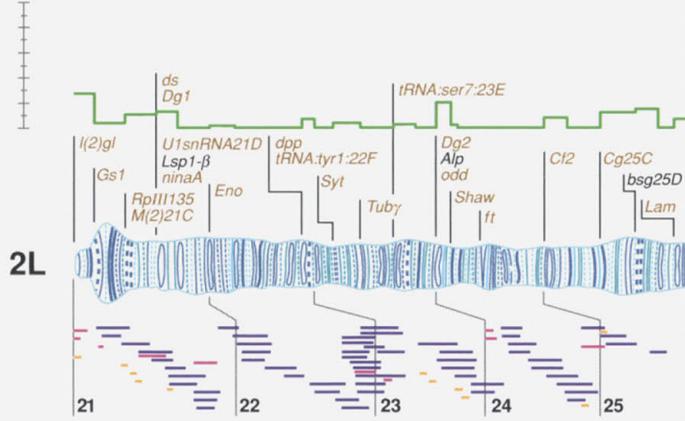
roximate locations and sizes of cloned DNA in YACs (blue), cosmids (orange), or
 (magenta) are pictured in proportion to their lengths. Cloned DNA sizes are relative to the
 lines indicate the locations of the DNAs relative to chromosome divisions and should be
 the chromosome bands. The size of each chromosome division in kilobases has been
 Maps in *Drosophila* (CRC Press, Boca Raton, FL, 1988)]. Data on YACs were supplied by
 f 15 June, 1991); on cosmids by D. Glover, R. Saunders, F. Kafatos, I. Sidén-Kiamos, C.
 (current as of 1 July, 1991); and on chromosome walks were compiled by J. Merriam, G.
 1 May, 1991). DNA localization is not exact, because of differences in DNA density along
 in situ hybridization. The lengths of cosmid contigs were estimated from the number of
 cosmids, and chromosome walks are available from the authors upon request; further
 npanying article (J. Merriam *et al.*) in the 11 October, 1991 issue of *Science*.



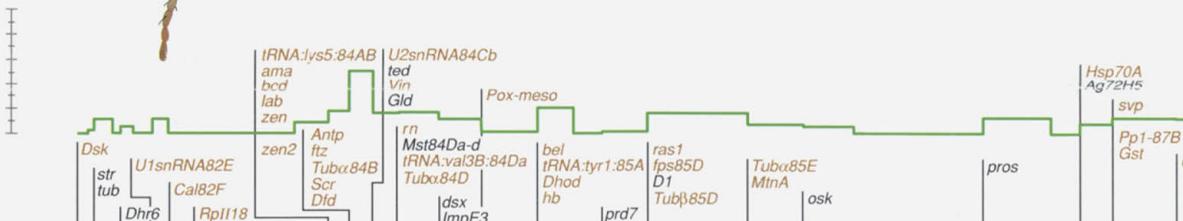
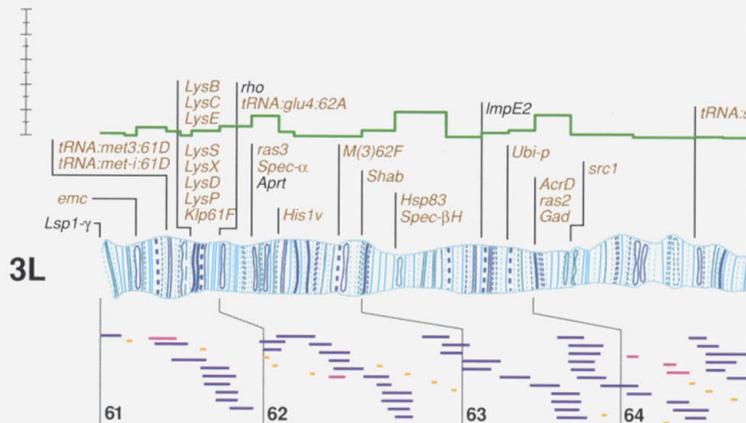
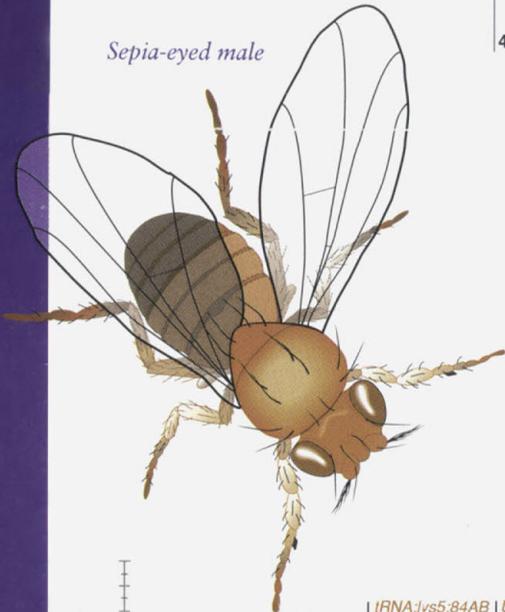


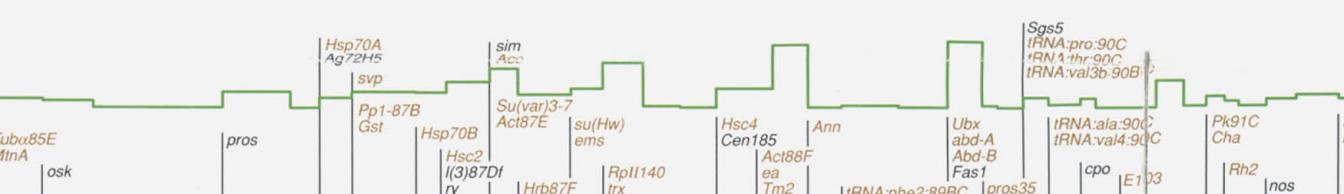
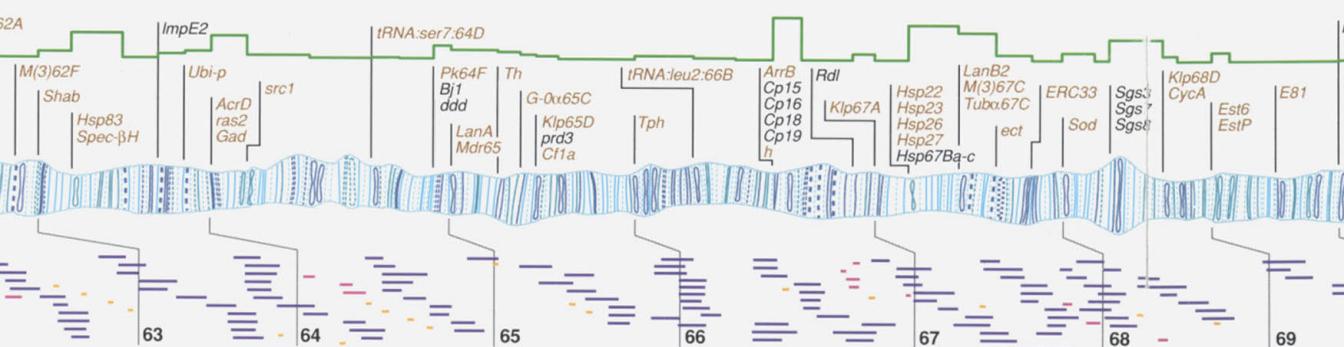
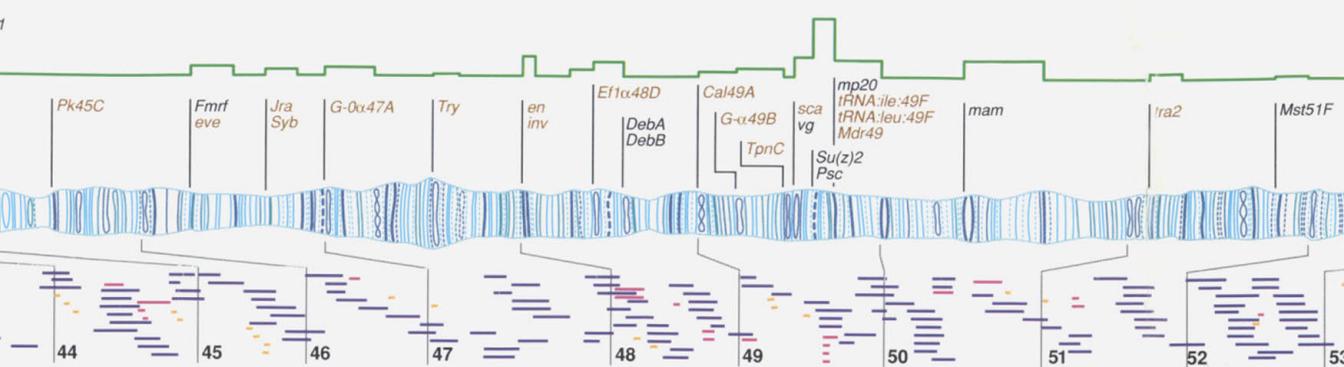
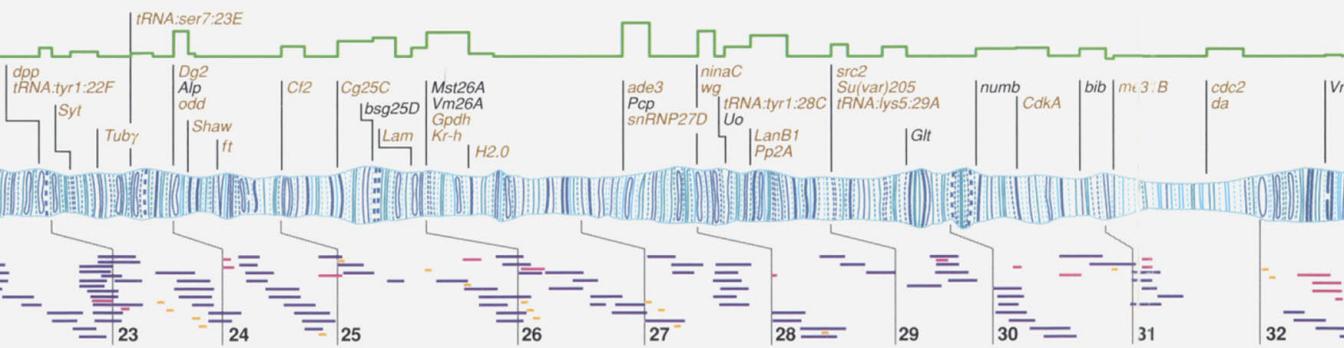
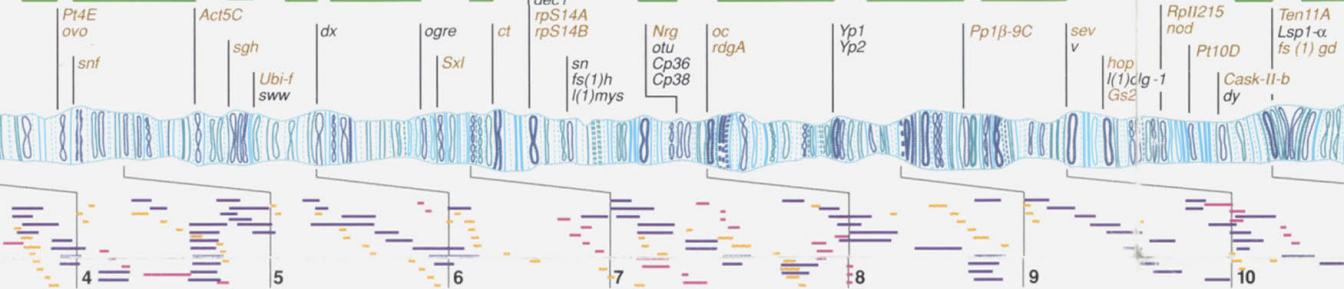
— 200 kb — For sizes of YACs, cosmids, and chromosome walks

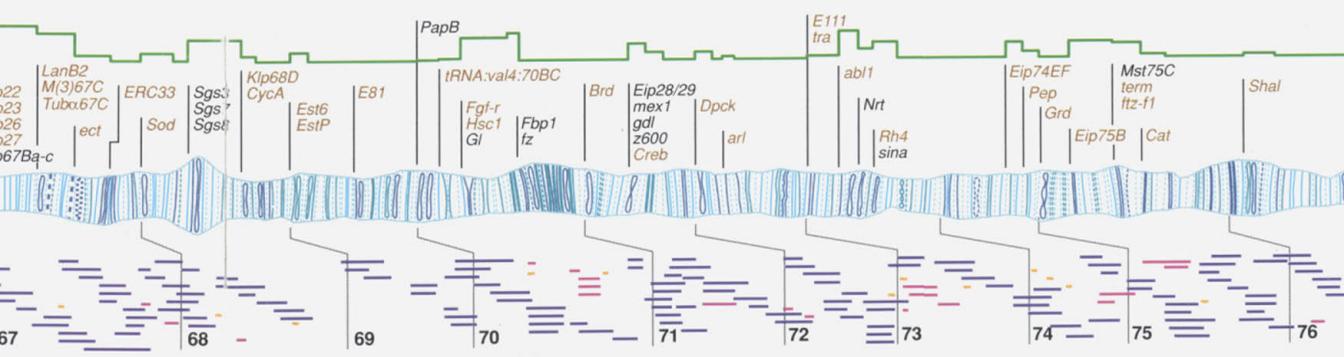
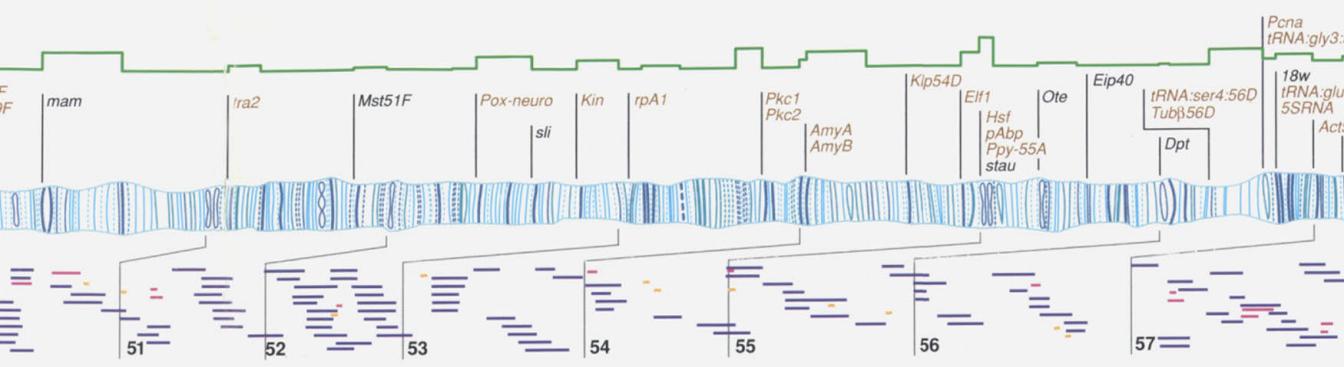
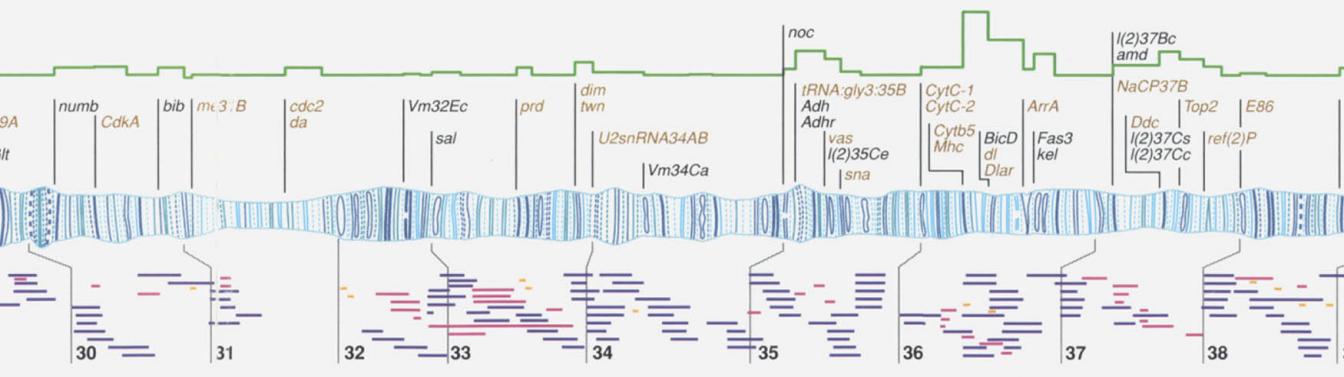
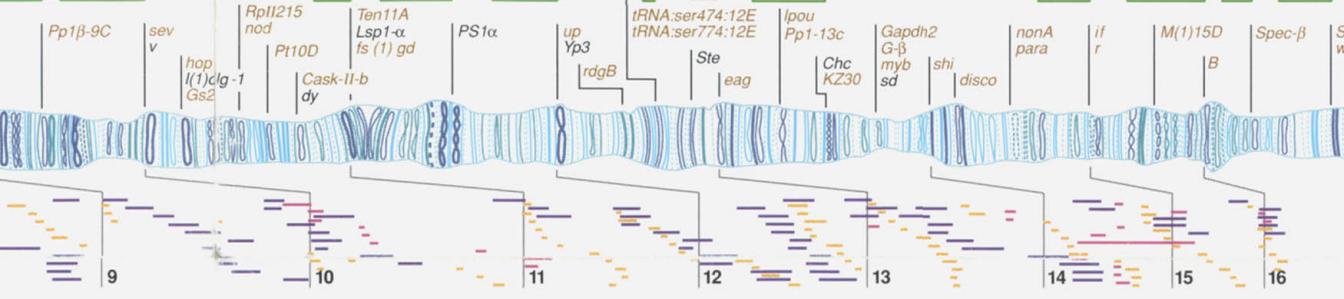
- — Kb sequenced per chromosome subdivision
- — Genes similar in *Drosophila* and vertebrates
- — YACs
- — Chromosome walks
- — Cosmids

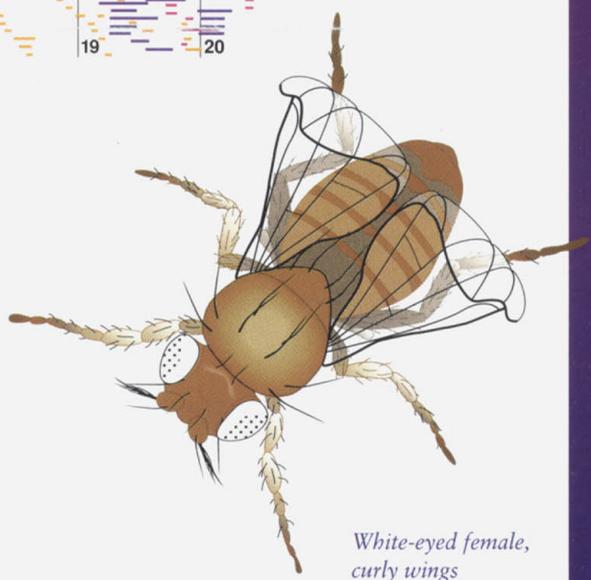
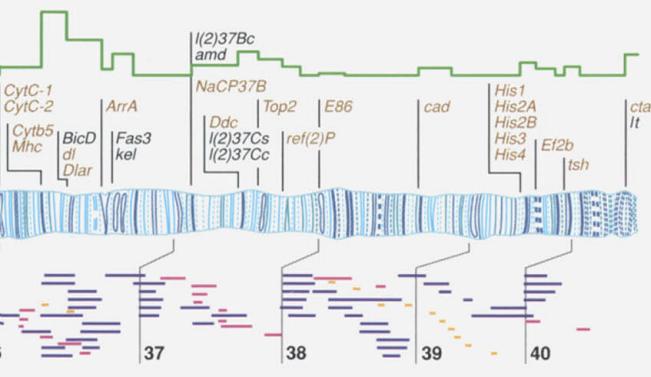
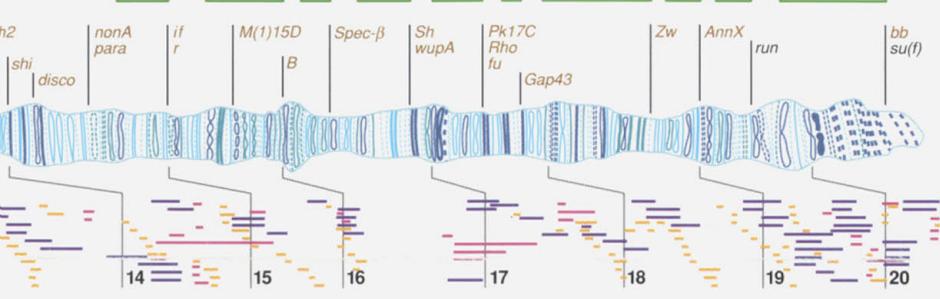


Sepia-eyed male

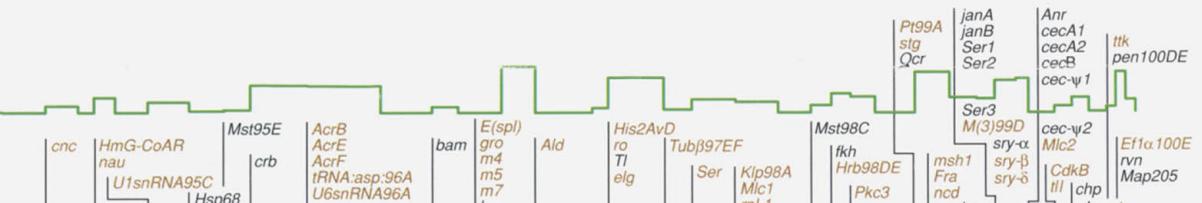
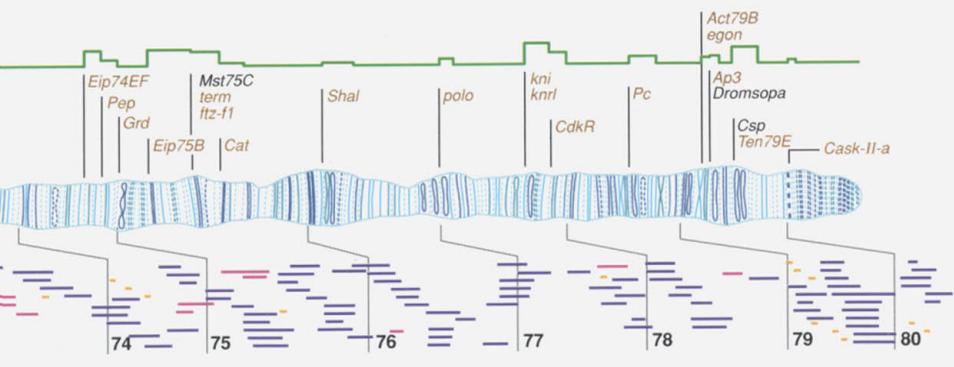
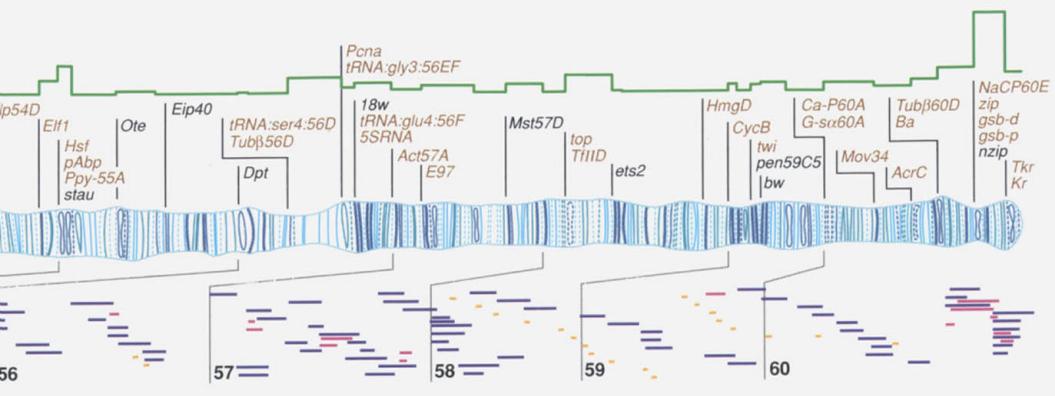






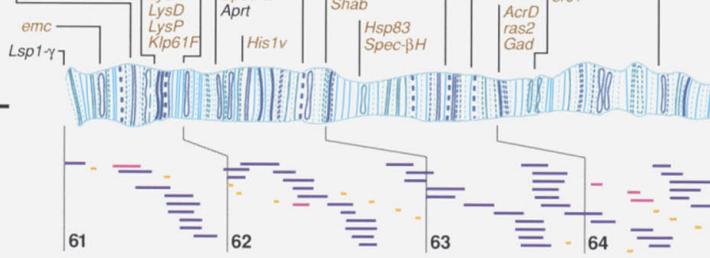


White-eyed female, curly wings

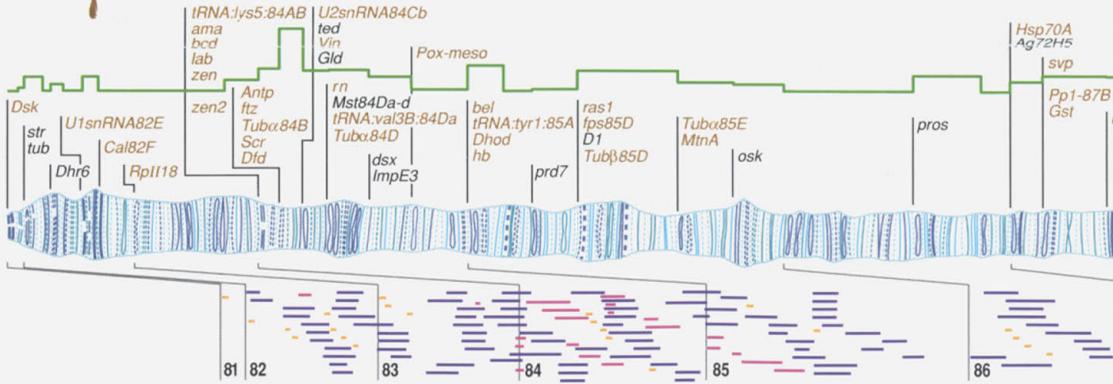




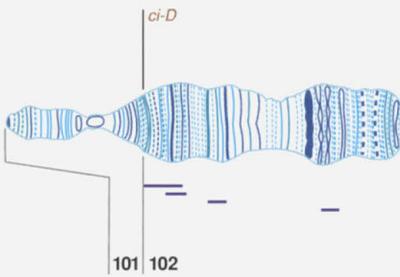
3L



3R



4



Cell cycle proteins

- Cdc2-like proteins (2)
- Cdc25-like
- Cyclin A
- Cyclin B

Cell and substrate adhesion proteins

- Cadherin (3)
- Collagen (3)
- Immunoglobulin superfamily (2)
- Integrin (3)
- Laminin (3)
- N-CAM-like
- Neural adhesion molecule L1

- Plakoglobin
- Tenascin (2)

Cytoskeleton

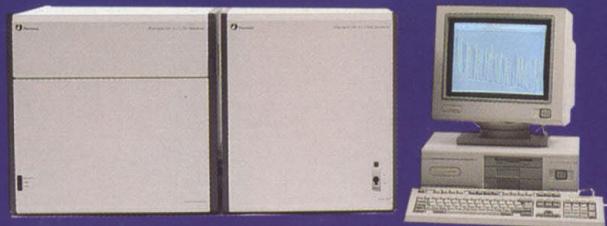
- Actin (6)
- α Actinin
- Cytoplasmic myosin I
- Cytoplasmic myosin II
- Dynamin
- Kinesin (11)
- Lamin
- Myosin heavy chain
- Myosin light chain (2)
- Paramyosin
- α Spectrin

GENOME MAPS 1991 • Barbara R. Jasny: *Science* Coordinator • *Homo sapiens* component: Authors: Michael Chipperfield, Bruce Dundee, Scotland; Ian Duncan and Daniel Hartl—Washington University, St. Louis, MO; John Merriam, Geunbae Lee, Jyong

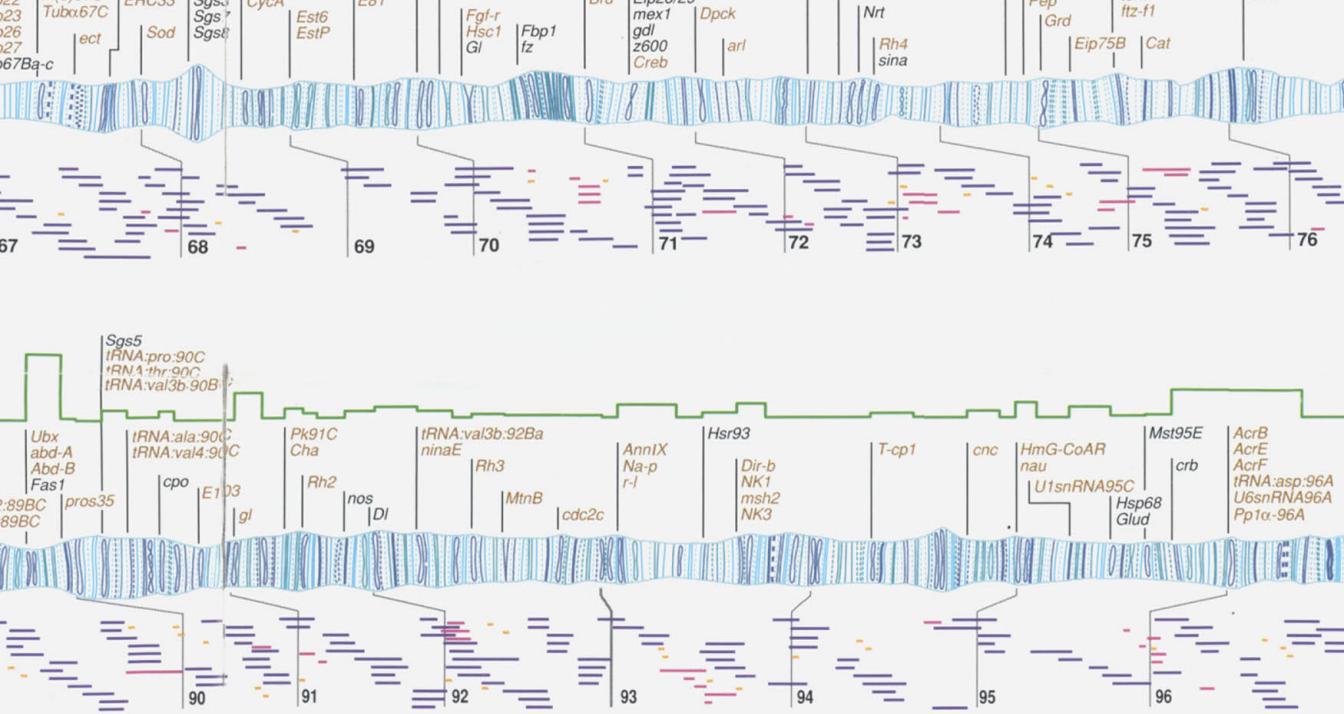
...β-thalassemia • Alport's syndrome • Fluorescent pro



PhastSystem™



A.L.F. DNA Sequencer™



GENES SIMILAR IN *DROSOPHILA* AND VERTEBRATES

- | | | | | |
|---|---|--|---|---|
| <ul style="list-style-type: none"> receptor (6) ated ATPase (3) nel protein yltransferase boxylase id decarboxylase ceptor ated protein GAP43 roteins (5) el proteins (3) ase ceptor vin | <ul style="list-style-type: none"> Synaptotagmin Tryptophan hydrolase Tyrosine-3-hydroxylase Nucleic acid synthesis and metabolism Adenine phosphoribosyl transferase Apurinic endonuclease Aspartate transcarbamylase Calbamy phosphate synthase DEAD proteins (5) Dihydroorotase Dihydroorotate dehydrogenase Excision repair ERCC3 Nucleoside diphosphate kinase | <ul style="list-style-type: none"> PCNA Phosphoribosylamine: glycine ligase Phosphoribosylformylglycinamide cyclo-ligase Phosphoribosylglycinamide formyltransferase Poly(A)-binding protein Ribonucleoproteins (12) RNA polymerase II (3) RNA polymerase III (2) RNA polymerase elongation factor Oncogenes Abl | <ul style="list-style-type: none"> Ets (3) Fps Gli Int1 Myb Myc Raf Ras (3) Rel Rhombitin Src (2) Protein synthesis machinery Elongation factors (3) Ribosomal proteins (13) | <ul style="list-style-type: none"> Second messenger systems Annxin (3) Arrestin (2) Calcineurin Calmodulin (2) Calmodulin inhibitor cAMP-dependent kinase (4) cAMP phosphodiesterase cGMP-dependent kinase (2) Cystatin Diacylglycerol kinase G proteins (9) GTPase-activating protein (2) GTP-binding protein |
|---|---|--|---|---|

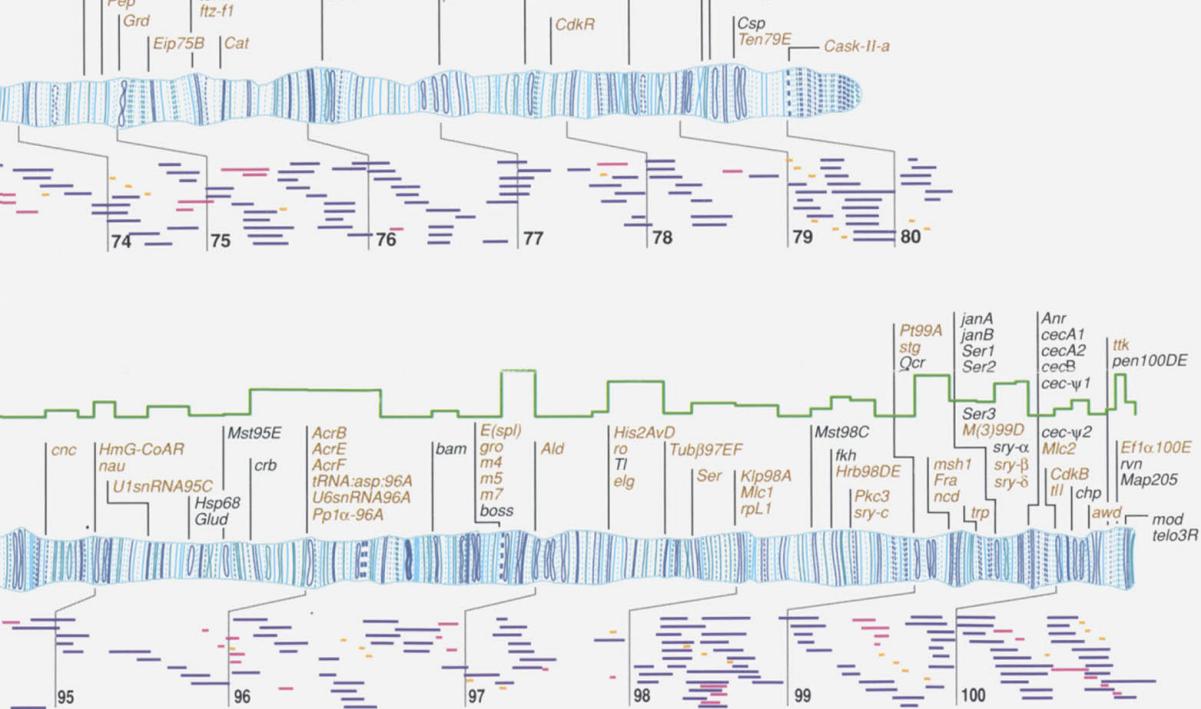
...ore, MD; Reviewer: Bertrand R. Jordan—CIML CNRS, Marseilles, France • *Drosophila melanogaster* component : Authors: Michael Ashburner—Universitatis C. Kafatos, Inga Sidén, Kiamos, Christos Louis, Charalambos Savakis, Research Center of Crete, Heraklion, Crete, Greece; Reviewer: Gerald M. Rubin
 ...ublication of The American Association for the Advancement of Science.

...gy reagents & kits • Muscular dystrophy • Cystic fibrosis



...n + Accuracy = Reproducibility

Gene Assembler®



VERTEBRATES

Second messenger systems

Annexin (3)
 Arrestin (2)
 Calcineurin
 Calmodulin (2)
 Calmodulin inhibitor
 cAMP-dependent kinase (4)
 cAMP phosphodiesterase
 cGMP-dependent kinase (2)
 Cystatin
 Diacylglycerol kinase
 G proteins (9)
 GTPase-activating protein (2)
 GTP-binding protein

Opsin (4)
 Peptidyl prolylisomerase (2)
 Phospholipase C
 Serine/threonine kinase (6)
 Serine/threonine phosphatase (6)
 Tyrosine kinase (11)
 Tyrosine phosphatase (5)

Transcription factors

cAMP response element binding protein (CREB)
 Heat-shock transcription factor
 Helix-loop-helix proteins (16)
 Homeodomain proteins (37)
 Paired-box proteins (5)

PCU domain proteins (4)
 Steroid receptor protein family (10)
 TFIIID
 Zn-finger proteins (22)
Other DNA binding proteins
 Chromobox proteins (2)
 High mobility group protein 1
 Histones (7)
 Topoisomerase II

Miscellaneous

β Amyloid protein precursor-like
 Cholecystokinin
 Clathrin heavy chain

Cytochromes (3)
 Fibrinogen-like
 Furin-like protease
 Heat-shock proteins (7)
 Lysozyme (7)
 Metallothionein (2)
 Multiple drug resistance (2)
 Proteasome subunits (3)
 Serine protease (9)
 Serine protease inhibitor
 Trypsin
 Ubiquitin (2)
 Other metabolic enzymes (25)

Component: Authors: Michael Ashburner—University of Cambridge, U.K.; David M. Glover and Robert D. C. Saunders—University of Crete, Crete, Greece; Reviewer: Gerald M. Rubin, University of California, Berkeley, CA • Graphic Artist: Susan Nowoslawski

• Cystic fibrosis • Forensic testing • Food testing...



Gene Assembler[®] Plus



Gene Navigator[™]