Ooh!

(make me look)

Aah!

(make me understand)

ON

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE

PRESERVATION OF FAVOURED RACES IN THE STRUGGLE FOR LIFE.

BY CHARLES DARWIN, M.A.,

FELLOW OF THE ROYAL, GEOLOGICAL, LINNÆAN, ETC., SOCIETIES;
AUTHOR OF 'JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE'S VOYAGE ROUND THE WORLD.'

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1859.

ON

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

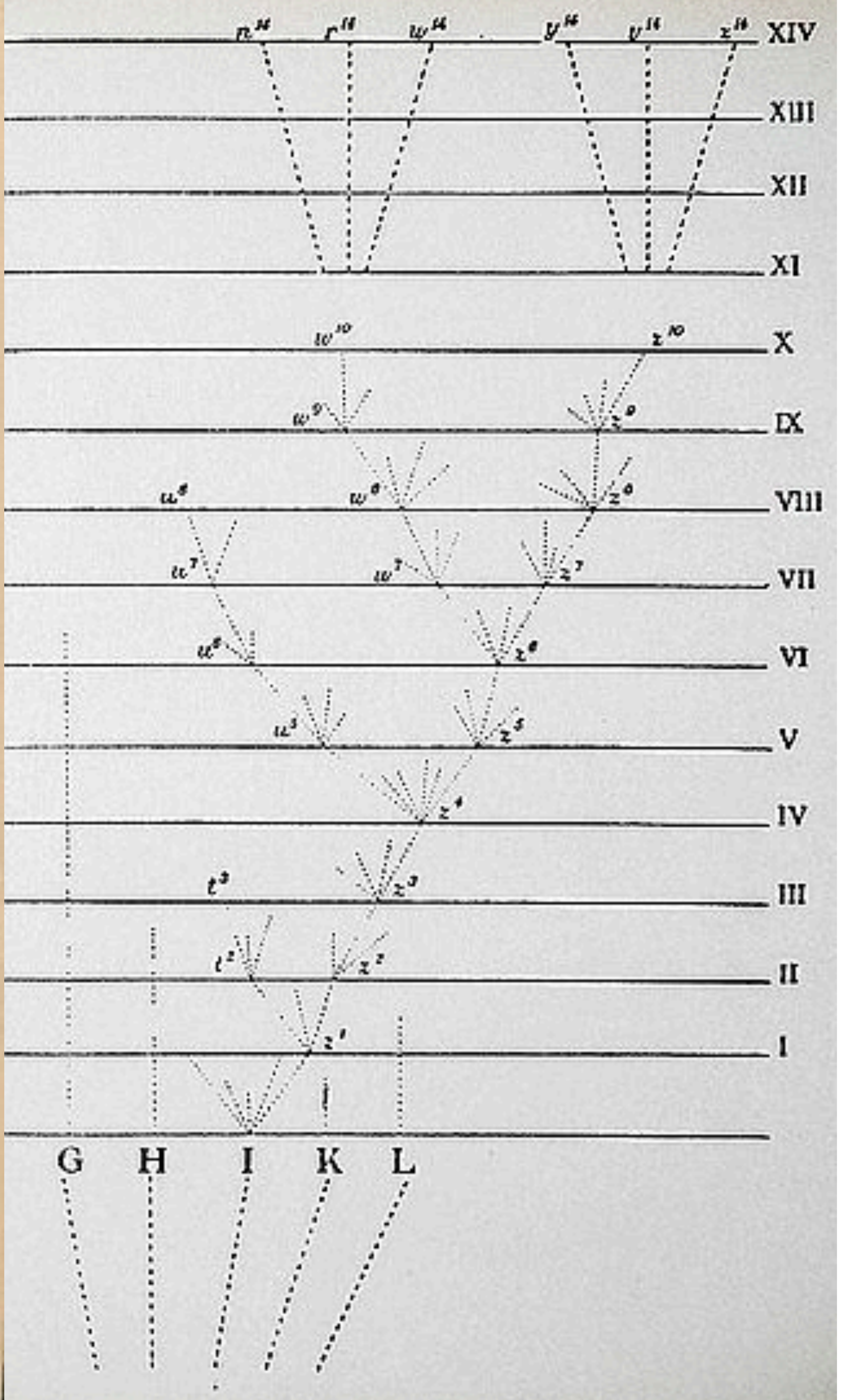
OR THE

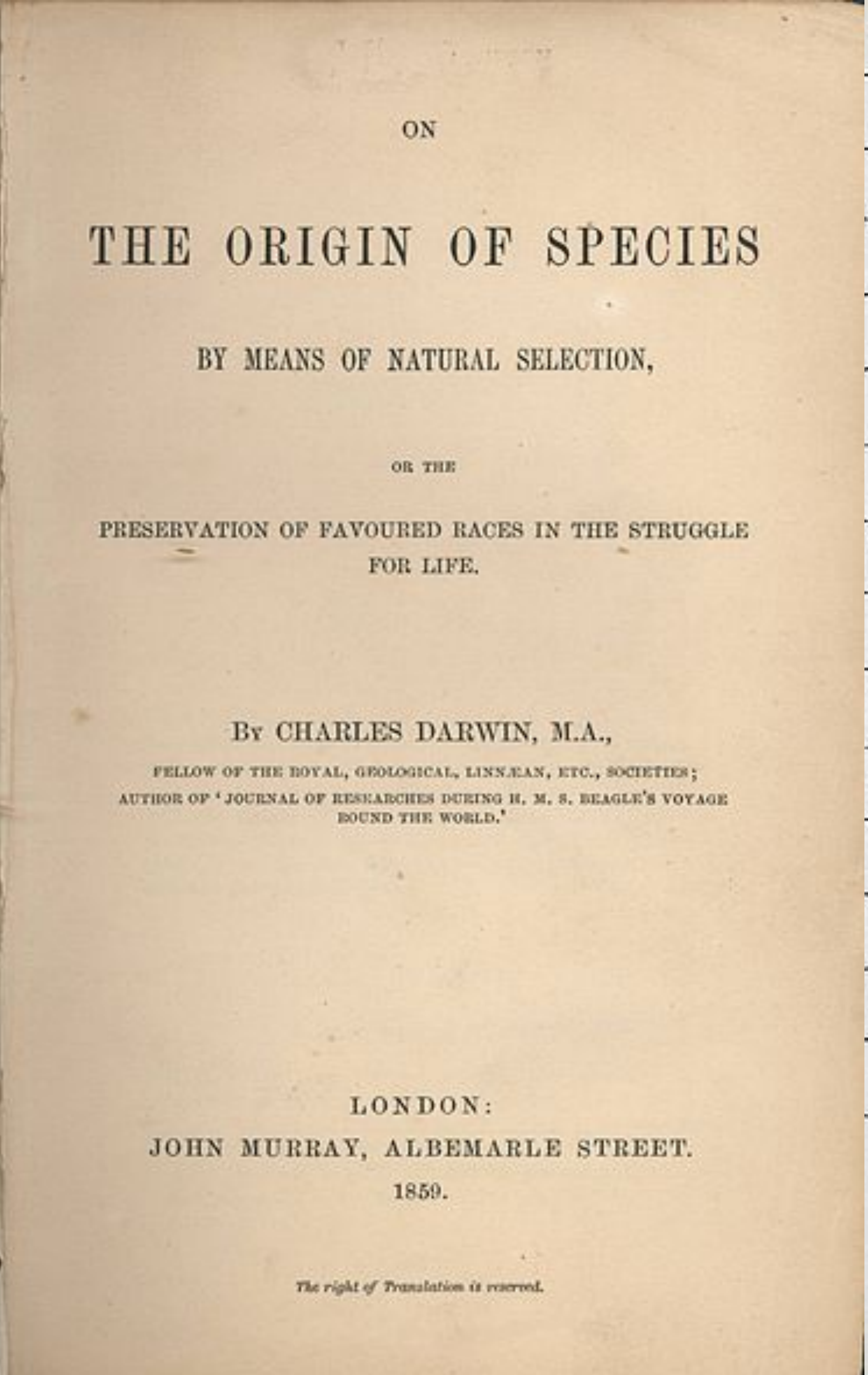
PRESERVATION OF FAVOURED RACES IN THE STRUGGLE FOR LIFE.

By CHARLES DARWIN, M.A.,

FELLOW OF THE ROYAL, GEOLOGICAL, LINNÆAN, ETC., SOCIETIES;
AUTHOR OF 'JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE'S VOYAGE ROUND THE WORLD.'

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1859.





ON

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

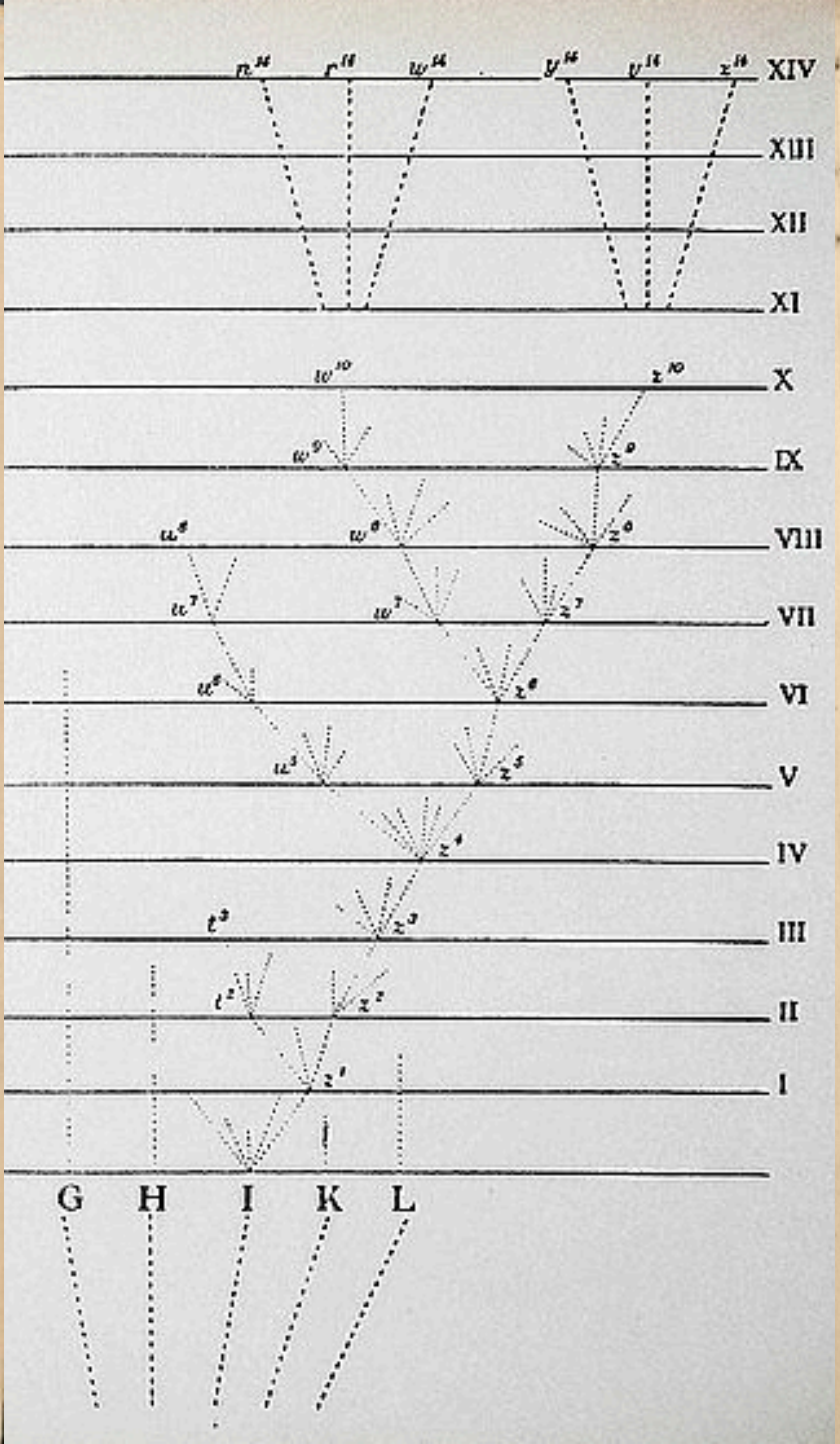
OR THE

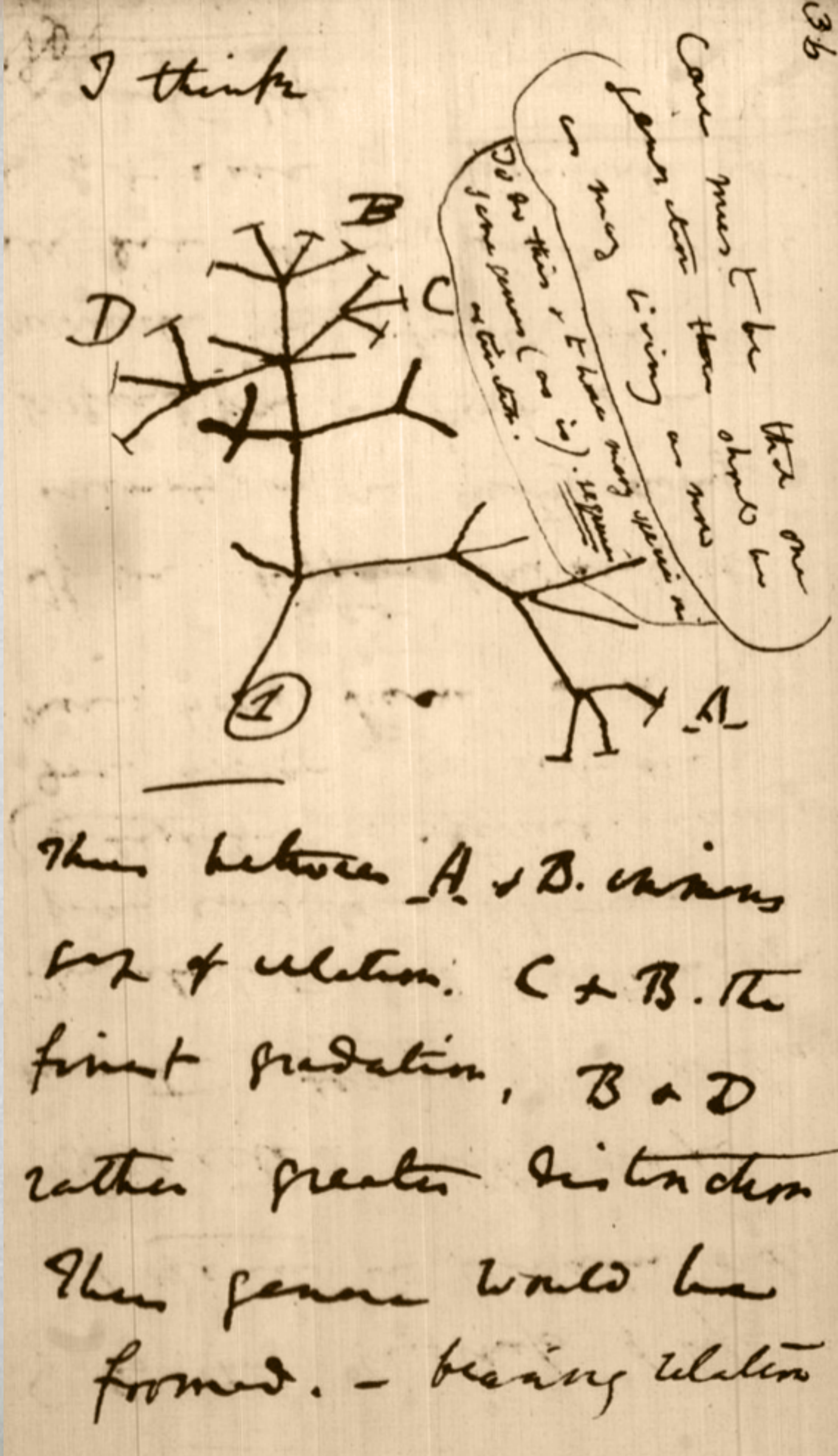
PRESERVATION OF FAVOURED RACES IN THE STRUGGLE FOR LIFE.

BY CHARLES DARWIN, M.A.,

FELLOW OF THE ROYAL, GEOLOGICAL, LINNÆAN, ETC., SOCIETIES;
AUTHOR OF 'JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE'S VOYAGE ROUND THE WORLD.'

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1859.





MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion. this structure is unsatisfactory for two reasons:
(1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions. namely, that each chain consists of phosphate diester groups joining ß-D-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow righthanded helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions.

Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3-4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain, does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain, is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in time following, communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereo-chemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at King's College, London. One of us (J.D.W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J.D. WATSON
F.H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge. April 2.

[1] Pauling, L., and Corey, R. B. nature, 171, 346 (1953); Proc, U.S. Nat Acad. Sci., 39, 84 (1953).
[2] Furberg, S., Acta Chem. Scand., 6, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., Biochim, et Biophys. Acta, 9 402 (1952).
[4] Wyatt, G.R. J. Gen. Physiol., 36 201 (1952).
[5] Astbury, W.T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947)
[6] Wilkins, M. H. F. and Randall, J. T. Biochim. et. Biophys. Acta, 10, 102 (1953).

figures 1
tables 0
references 6

figures 49
tables 27
references 452

data flood[1] tsunamis[2] deluges[3]

surging oceans[4]

avalanches[5] icebergs[6]

landslide[7] earthquakes[8]

explosions[9]

1. Andrade M et al. Curr Opin Biotechnol 8:675 (1997). 2. Wurman RS. Information Architects (1997). 3. Hess K et al. Trends Biotechnol 19:463 (2001), Editorial Nat Biotech 26:1099 (2008). 4. Dubitzky W. Brief Bioinform 10:343 (2009). 5. Antezana E et al. Brief Bioinform 10:392 (2009). 6. Hodgson C. Nat Biotechnol 19:BE44 (2001). Howe D et al. Nature 455:47 (2008). 7. Attwood T et al. Biochem J 424:317 (2009). 8. Whilbanks J. CTWatchQuarterly (2007). 9. Diehn M. et al. Nucleic Acids Res 31:219 (2003).

The New York Times

SIDE EFFECTS

‘Ome,’ the Sound of the Scientific Universe Expanding

Martin Krzywinski/Michael Smith Genome Sciences Center

Visualizations, in progressively greater detail, that show duplications within the human genome.

By JAMES GORMAN

Published: May 3, 2012

I am a specialist.

expose detail → *explore*

(inspire me)

The New York Times

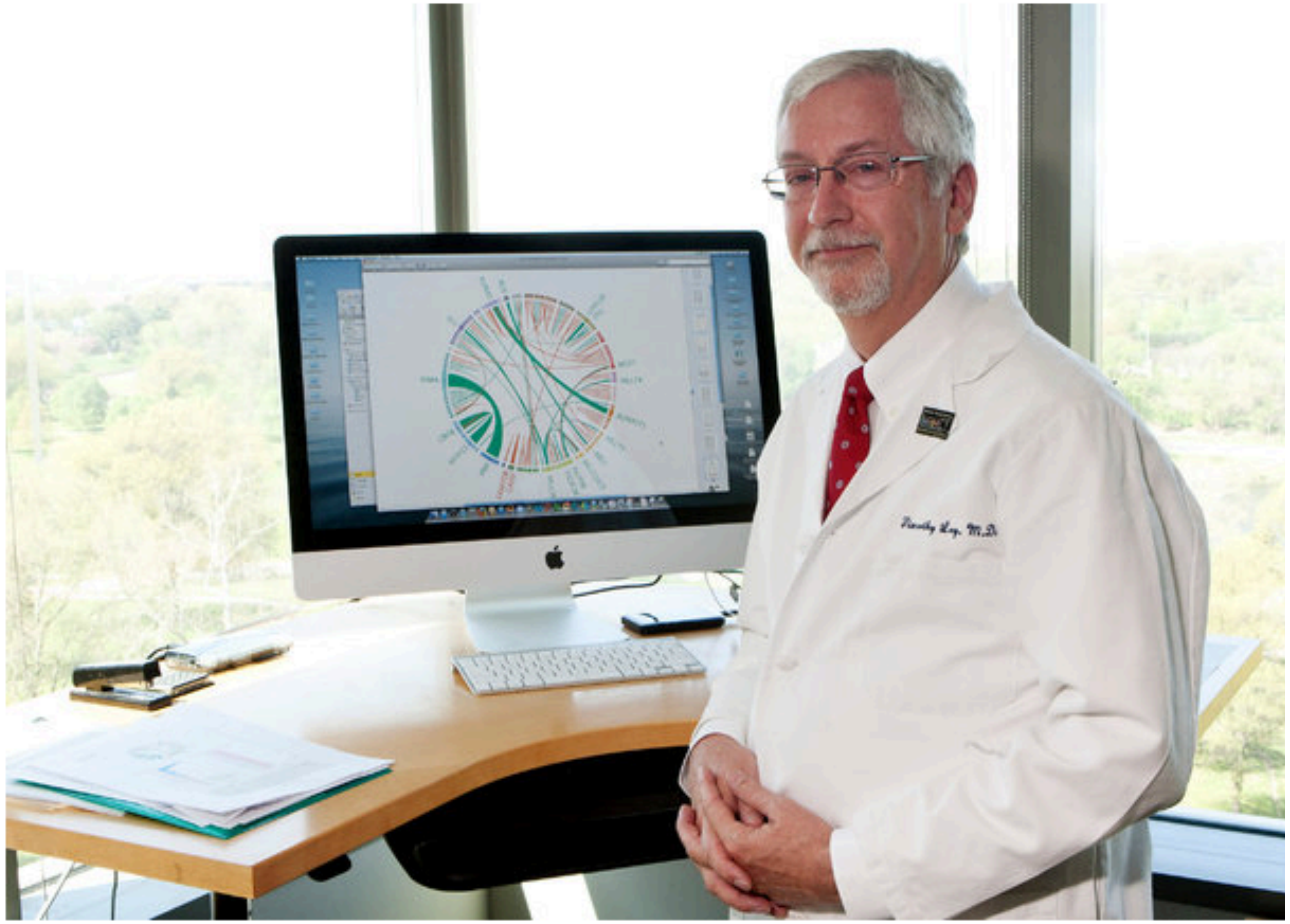
Health

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS

Search Health Go

Inside Health
Research | Fitn

Cancers Share Gene Patterns, Studies Affirm



Peter Newcomb for The New

According to Dr. Timothy Ley, traditional methods for categorizing acute myeloid leukemia were imprecise.

The New York Times

Health

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPI

Search Health Go

Inside Health
Research | Fitness

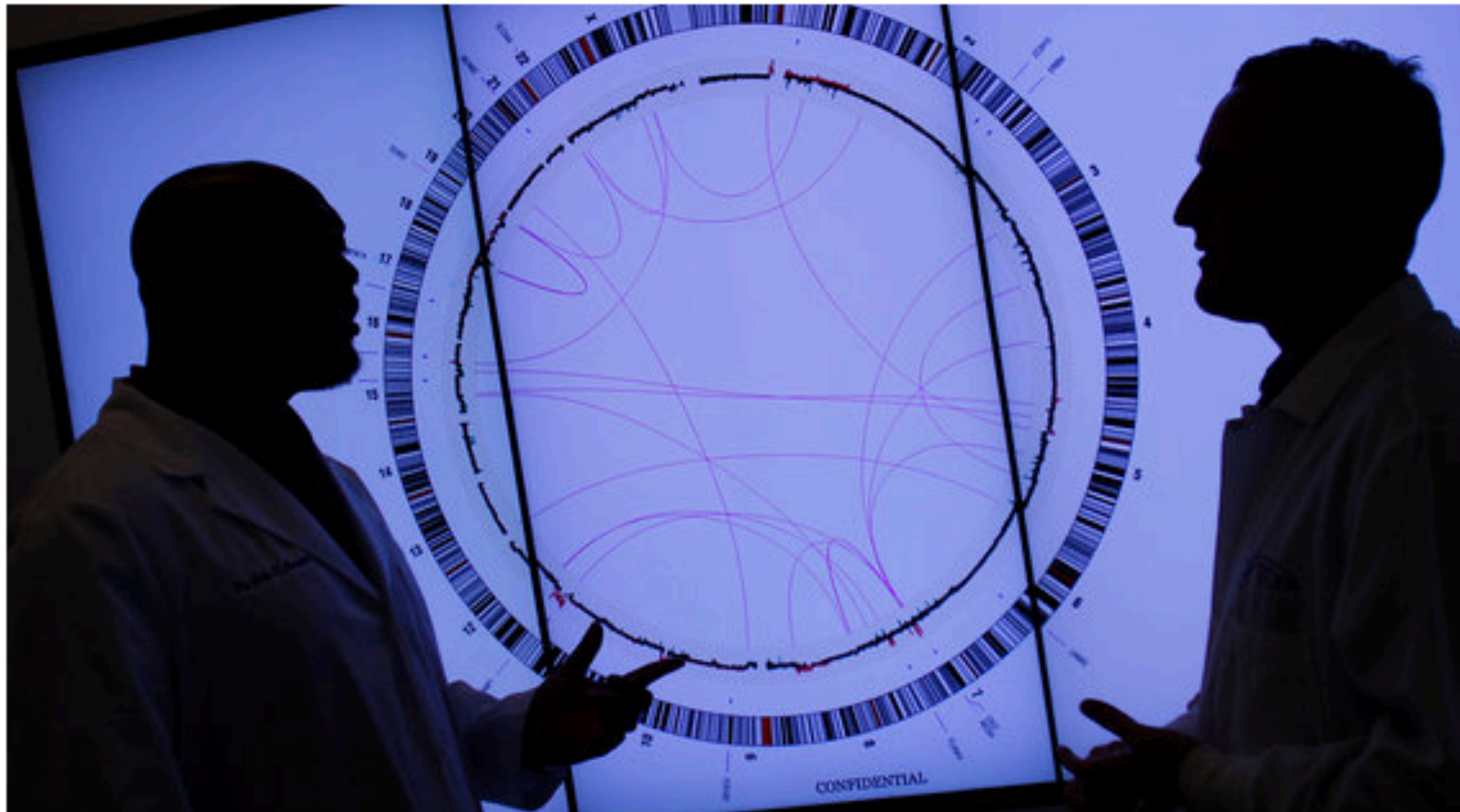
Genetic Gamble

New Approaches to Fighting Cancer

PART ONE
A Race to Leukemia's Source

PART TWO
Promise
Heartbre

A New Treatment's Tantalizing Promise Brings Heartbreaking Ups and Downs



It has become extremely hard and costly to pinpoint and understand what we *already* know.

Attwood et al. *Biochem J.* 424:317 (2009).

http://www.nature.com/nrc/calendars/2013

Good figures can encapsulate entire fields of cancer research without the need for extensive explanations.

—Nature Reviews Cancer

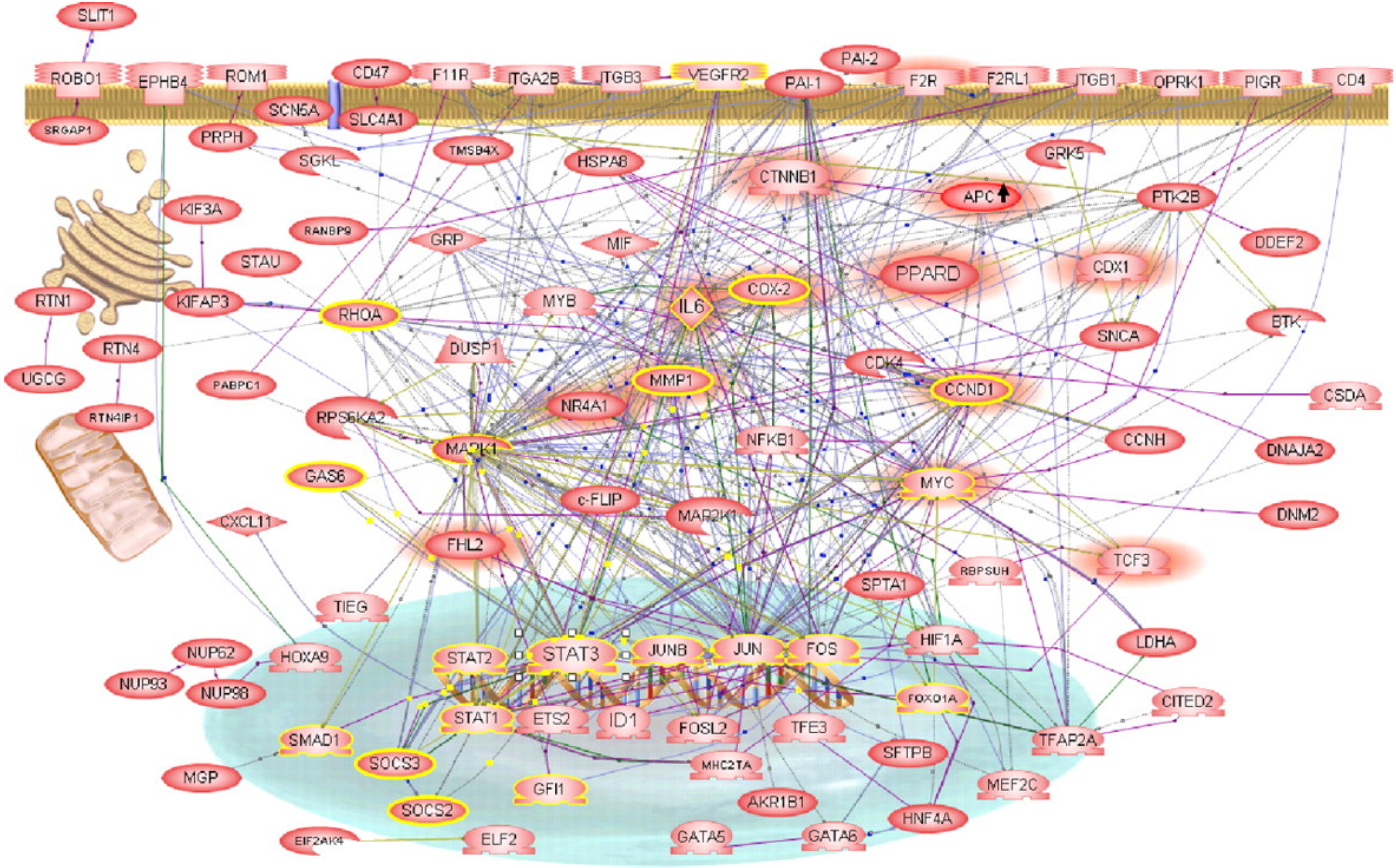
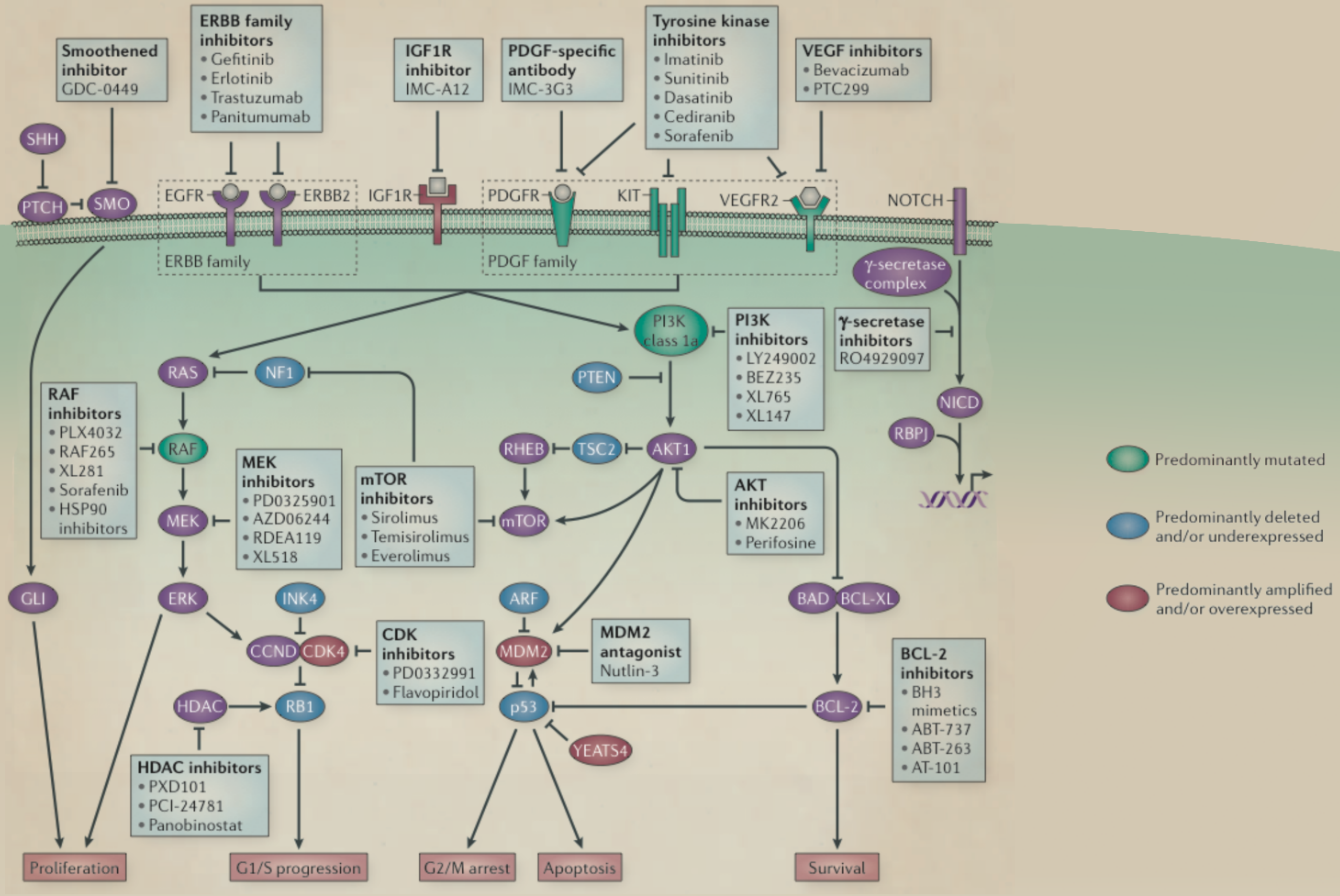
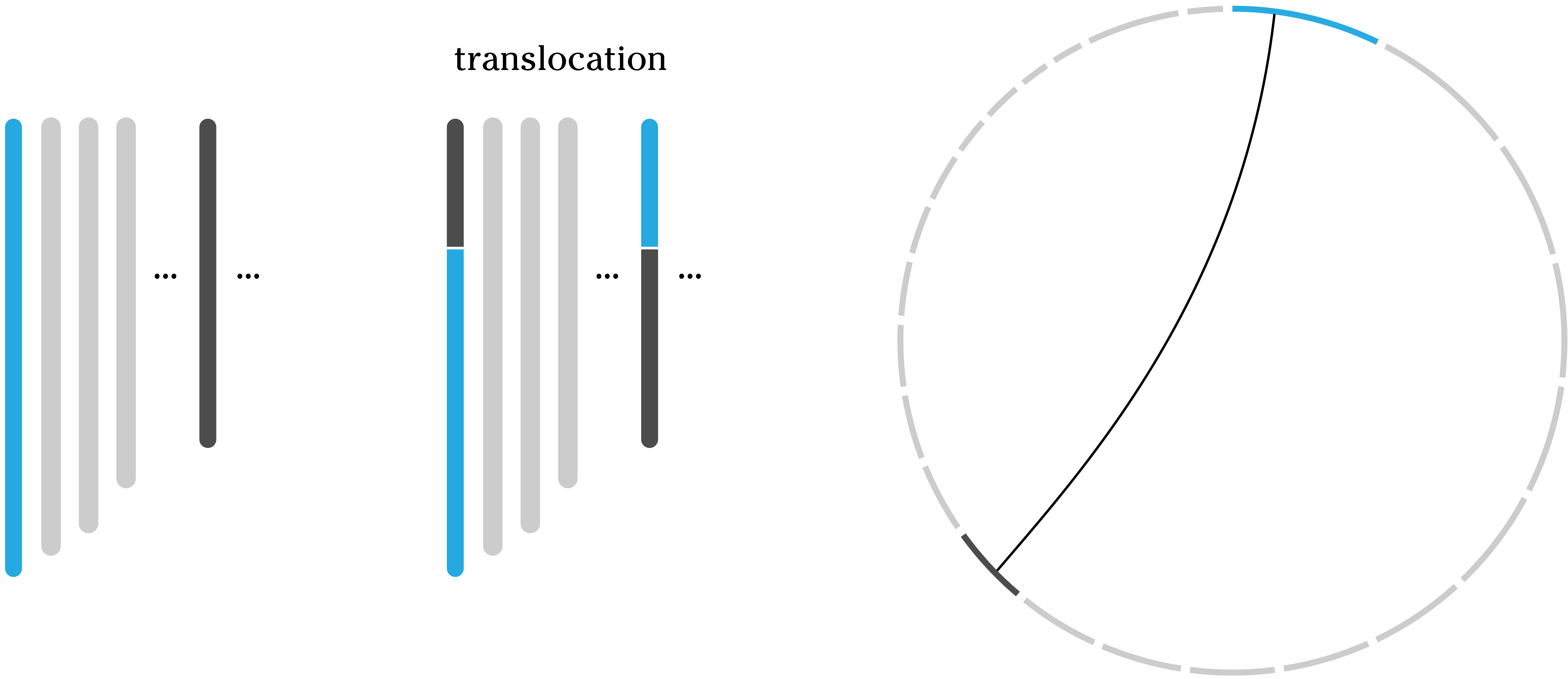


Fig 2 from Abdollahi, A. et al (2007) PNAS 104:12890-12895.

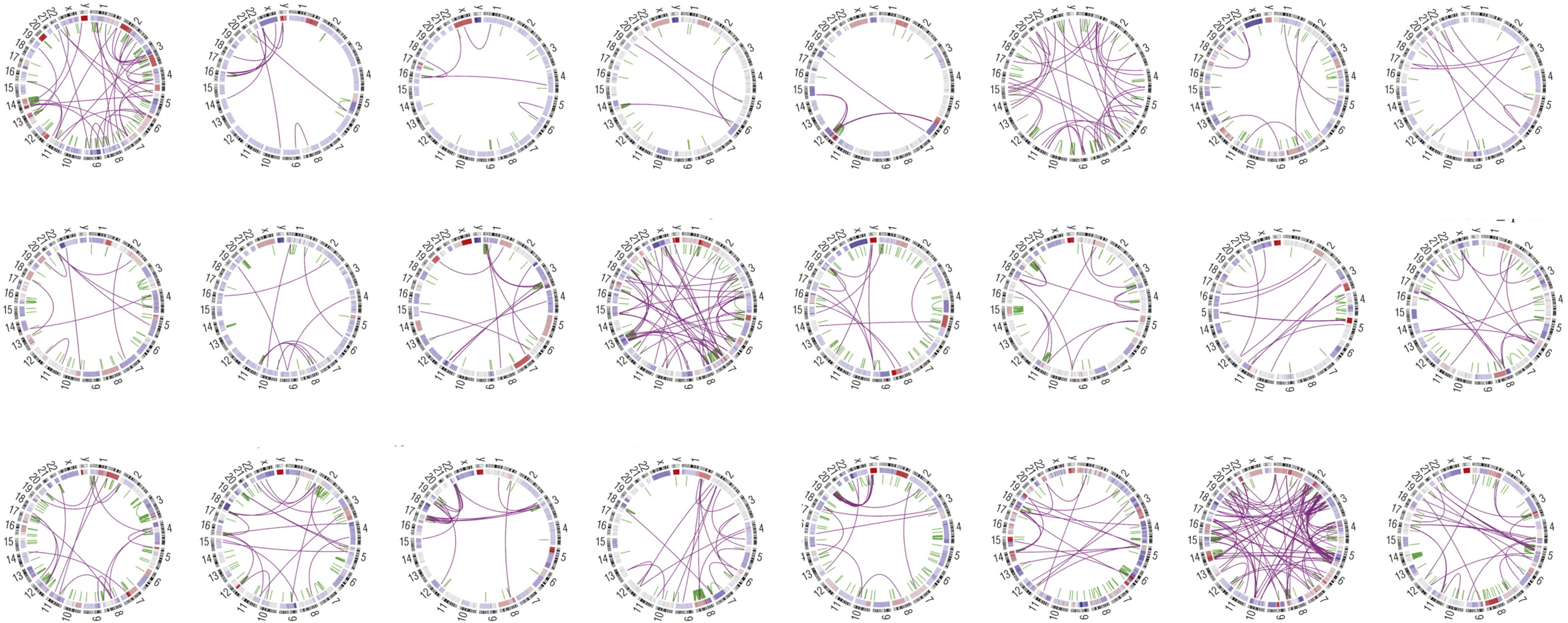


Taylor, BS et al. Nature Rev. Cancer 11, 541–557 (2011).

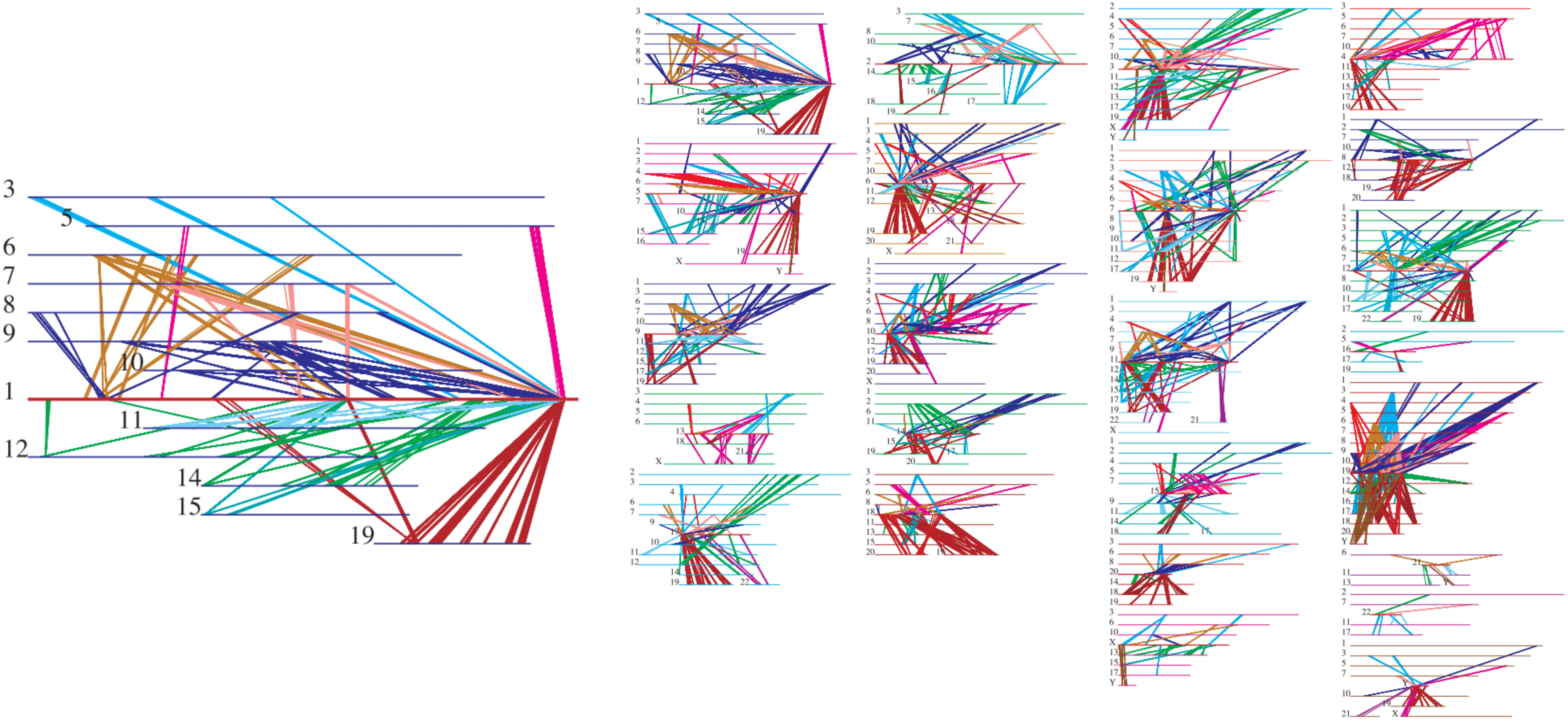

translocation

Rearrangement signatures of adenocarcinoma

(24/183 shown)



Imielinski M et al. (2012) Cell 150:1107-1120.



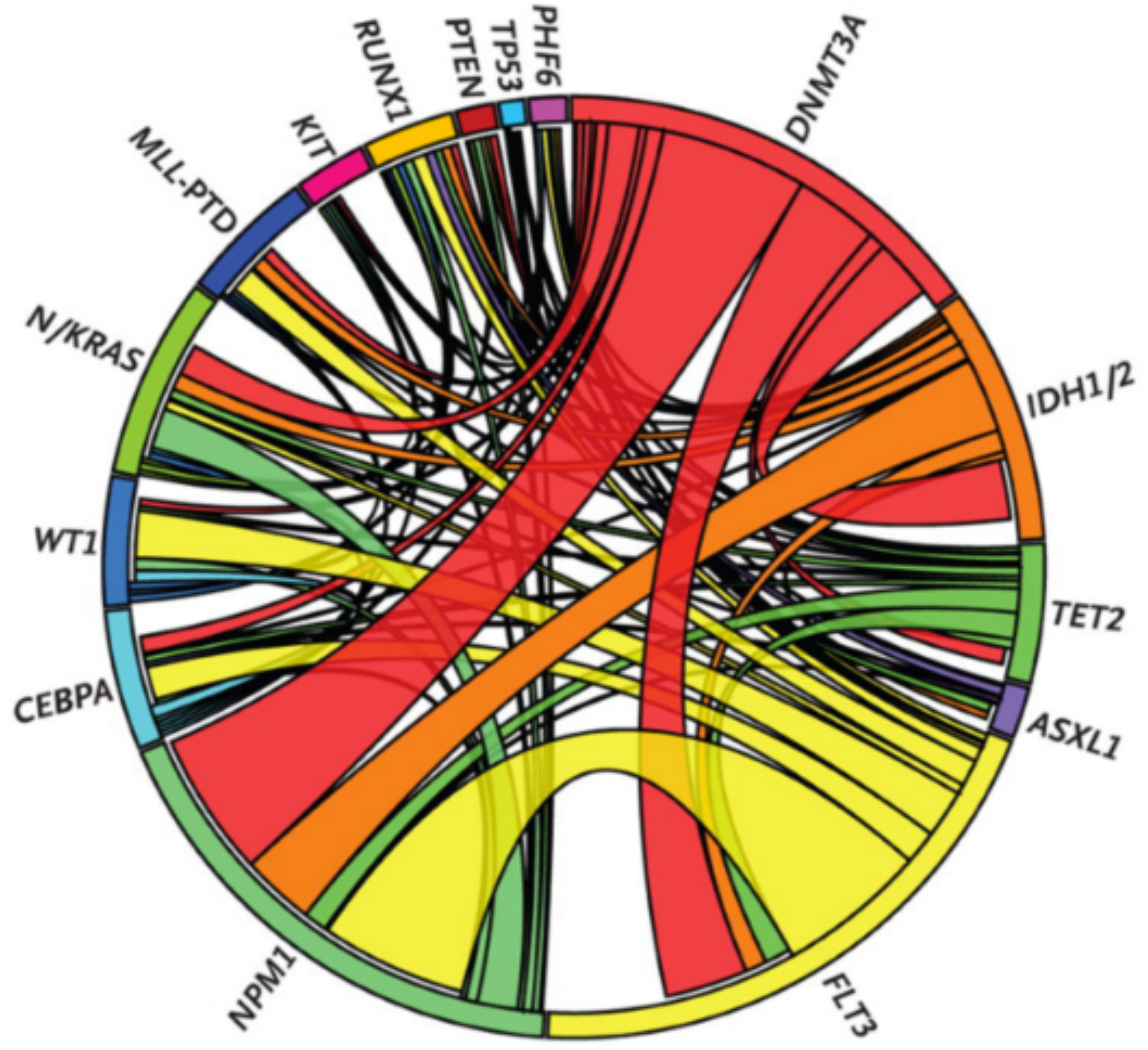
Venter JC et al. (2001) Science 291:1304-1351.


A B C D E F G ...
A
B
C
D
E
F
G
10
2
A
C
D
G
10
2

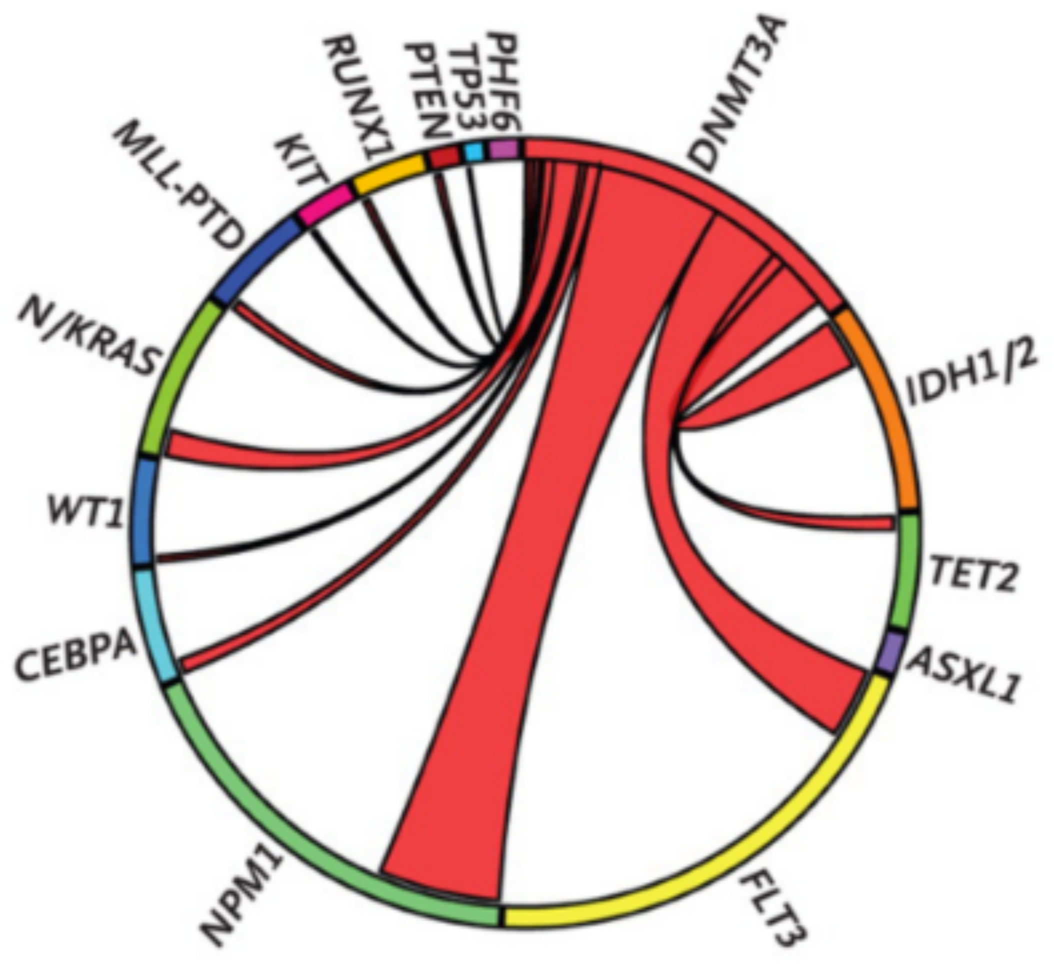
Table S5: Co-occurrences of somatic mutations and cytogenetic abnormalities in the test cohort of 398 AML patients with *de novo* AML from the ECOG E1900 trial.

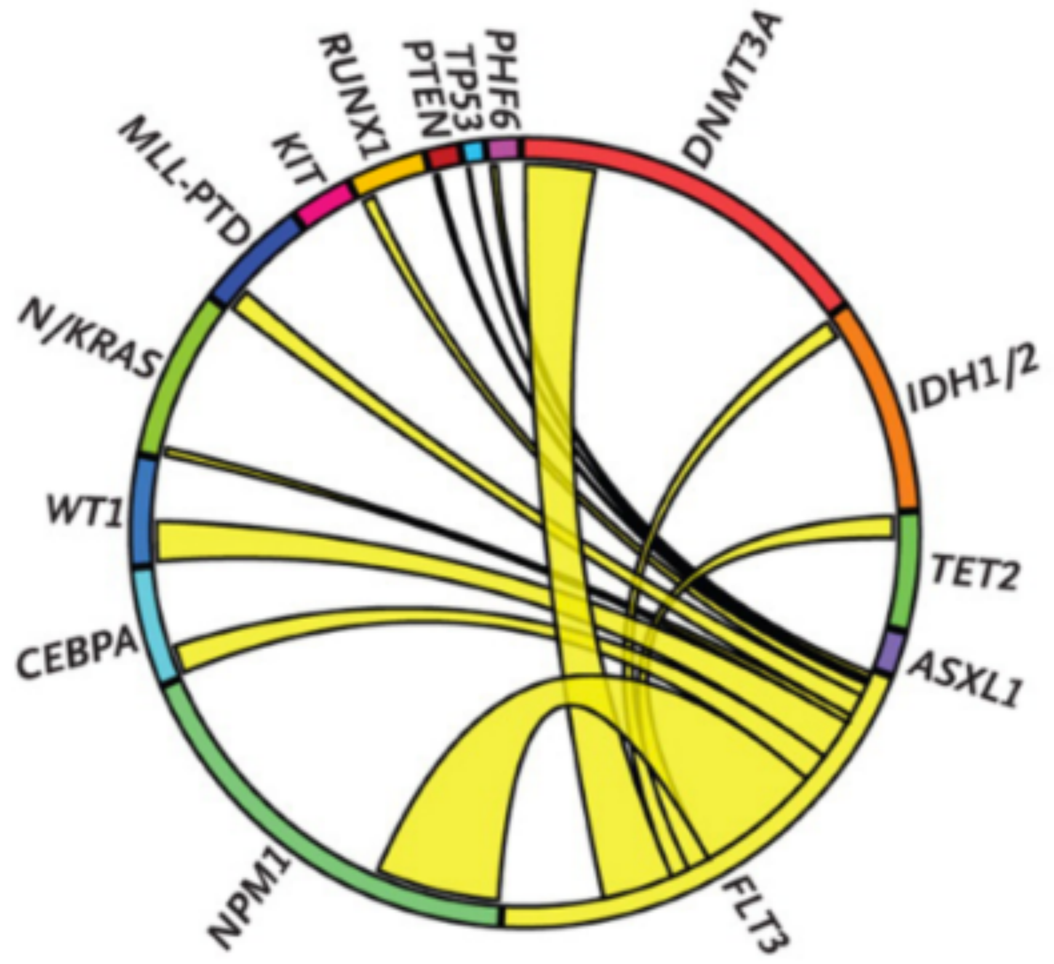
	DNMT3a	IDH1	IDH2	TET2	ASXL1	FLT3	NPM1	CEBPA	WT1	KRas	NRas	PHF6	KIT	TP53	PTEN	RUNX1	CBF	Del (5q)	EVI1	MLL-PTD	Split MLL	Monosomy (7/7q)	t(6;9)	Tri(8)	AML1-ETO
DNMT3a		3.3% (13/398)	1.5% (6/398)	1.5% (6/398)	0% (0/398)	13.3% (53/398)	14.3% (57/398)	1.75% (7/398)	0.75% (3/398)	0.75% (3/398)	2.5% (10/398)	0% (0/398)	0.5% (2/398)	0.25% (1/398)	0.75% (3/398)	0.75% (3/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	1% (4/398)	0.25% (1/398)	0.25% (1/398)	0% (0/398)	1.5% (6/398)	0% (0/398)
IDH1	3.3% (13/398)		0% (0/398)	0% (0/398)	0.25% (1/398)	1% (4/398)	1.5% (6/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	0.75% (3/398)	0.5% (2/398)	0.25% (1/398)	0% (0/398)	0.5% (2/398)	0.25% (1/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0.5% (2/398)	0.25% (1/398)	0.25% (1/398)	0% (0/398)	0.5% (2/398)	0% (0/398)
IDH2	1.5% (6/398)	0% (0/398)		0% (0/398)	0.5% (2/398)	2% (8/398)	2% (8/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.75% (3/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.75% (3/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.75% (3/398)	0.5% (2/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)
TET2	1.5% (6/398)	0% (0/398)	0% (0/398)		0.75% (3/398)	3% (12/398)	1.5% (6/398)	0.5% (2/398)	0.5% (2/398)	0% (0/398)	1% (4/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	1.3% (5/398)	0.25% (1/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	0% (0/398)
ASXL1	0% (0/398)	0.25% (1/398)	0.5% (2/398)	0.75% (3/398)		0% (0/398)	0.25% (1/398)	0.5% (2/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	1% (4/398)	1.3% (5/398)	0% (0/398)	0.25% (1/398)	0.5% (2/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	0.25% (1/398)	0% (0/398)
FLT3	13.3% (53/398)	1% (4/398)	2% (8/398)	3% (12/398)	0% (0/398)		6.8% (27/398)	3.5% (14/398)	5% (20/398)	0.25% (1/398)	0.5% (2/398)	1% (4/398)	0% (0/398)	0.25% (1/398)	0.5% (2/398)	1.5% (6/398)	1.5% (6/398)	0.25% (1/398)	0.25% (1/398)	2.5% (10/398)	0.5% (2/398)	0% (0/398)	0.25% (1/398)	2.26% (9/398)	0% (0/398)
NPM1	14.3% (57/398)	1.5% (6/398)	2% (8/398	1.5% (6/398)	0.25% (1/398)	6.8% (27/398)		0.5% (2/398)	0.25% (1/398)	0.5% (2/398)	1.3% (5/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0.5% (2/398)	0.5% (2/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)
CEBPA	1.75% (7/398)	0.25% (1/398)	0% (0/398)	0.5% (2/398)	0.5% (2/398)	3.5% (14/398)	0.5% (2/398)		1.3% (5/398)	0% (0/398)	0.5% (2/398)	0.5% (2/398)	0.5% (2/398)	0% (0/398)	0% (0/398)	0% (0/398)	1% (4/398)	0% (0/398)	0% (0/398)	0.5% (2/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	0% (0/398)
WT1	0.75% (3/398)	0% (0/398)	0% (0/398)	0.5% (2/398)	0% (0/398)	5% (20/398)	0.25% (1/398)	1.3% (5/398)		0% (0/398)	0.75% (3/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.75% (3/398)	1% (4/398)	0% (0/398)	0% (0/398)	0.5% (2/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0% (0/398)
KRas	0.75% (3/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0.5% (2/398)	0% (0/398)	0% (0/398)		0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0.5% (2/398)	0% (0/398)	0% (0/398	0% (0/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)
NRas	2.5% (10/398)	0.75% (3/398)	0.75% (3/398)	1% (4/398)	0.25% (1/398)	0.5% (2/398)	1.3% (5/398)	0.5% (2/398)	0.75% (3/398)	0% (0/398)		0% (0/398)	0.25% (1/398)	0% (0/398)	0.5% (2/398)	0.5% (2/398)	3% (12/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0.75% (3/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)
PHF6	0% (0/398)	0.5% (2/398)	0% (0/398)	0.25% (1/398)	0.25% (1/398)	1% (4/398)	0% (0/398)	0.5% (2/398)	0% (0/398)	0% (0/398)	0% (0/398)		0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0.25% (1/398)	0.25% (1/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)
KIT	0.5% (2/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0.5% (2/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)		0% (0/398)	0% (0/398)	0% (0/398)	5.3% (21/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)
TP53	0.25% (1/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)		0.25% (1/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)
PTEN	0.75% (3/398)	0.5% (2/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.5% (2/398)	0.5% (2/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.5% (2/398)	0% (0/398)	0% (0/398)	0.25% (1/398)		0% (0/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)
RUNX1	0.75% (3/398)	0.25% (1/398)	0.75% (3/398)	0.25% (1/398)	1% (4/398)	1.5% (6/398)	0.5% (2/398)	0% (0/398)	0.75% (3/398	0.25% (1/398)	0.5% (2/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)		0.5% (2/398)	0.75% (3/398)	0% (0/398)	1% (4/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)
CBF	0.25% (1/398)	0.25% (1/398)	0% (0/398)	1.3% (5/398)	1.3% (5/398)	1.5% (6/398)	0% (0/398)	1% (4/398)	1% (4/398)	0.5% (2/398)	3% (12/398)	0.25% (1/398)	5.3% (21/398)	0% (0/398)	0.25% (1/398)	0.5% (2/398)		0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)
Del (5q)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0.75% (3/398)	0% (0/398)		0% (0/398)	1% (4/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)
EVI1	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0.25% (1/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)		0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)
MLL-PTD	1% (4/398)	0.5% (2/398)	0.75% (3/398)	0% (0/398)	0.5% (2/398)	2.5% (10/398)	0% (0/398)	0.5% (2/398)	0.5% (2/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	1% (4/398)	0% (0/398)	1% (4/398)	0% (0/398)		0.5% (2/398)	0.25% (1/398)	0% (0/398)	0.25% (1/398)	0% (0/398)
Split MLL	0.25% (1/398)	0.25% (1/398)	0.5% (2/398)	0% (0/398)	0.25% (1/398)	0.5% (2/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0.75% (3/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)		0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)
Monosomy (7/7q)	0.25% (1/398)	0.25% (1/398)	0.25% (1/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)		0% (0/398)	0% (0/398)	0% (0/398)
t(6;9)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)		0% (0/398)	0% (0/398)
Tri(8)	1.5% (6/398)	0.5% (2/398)	0% (0/398)	0.25% (1/398)	0.25% (1/398)	2.26% (9/398)	0.25% (1/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)		0% (0/398)
AML1-ETO	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0.25% (1/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	0% (0/398)	

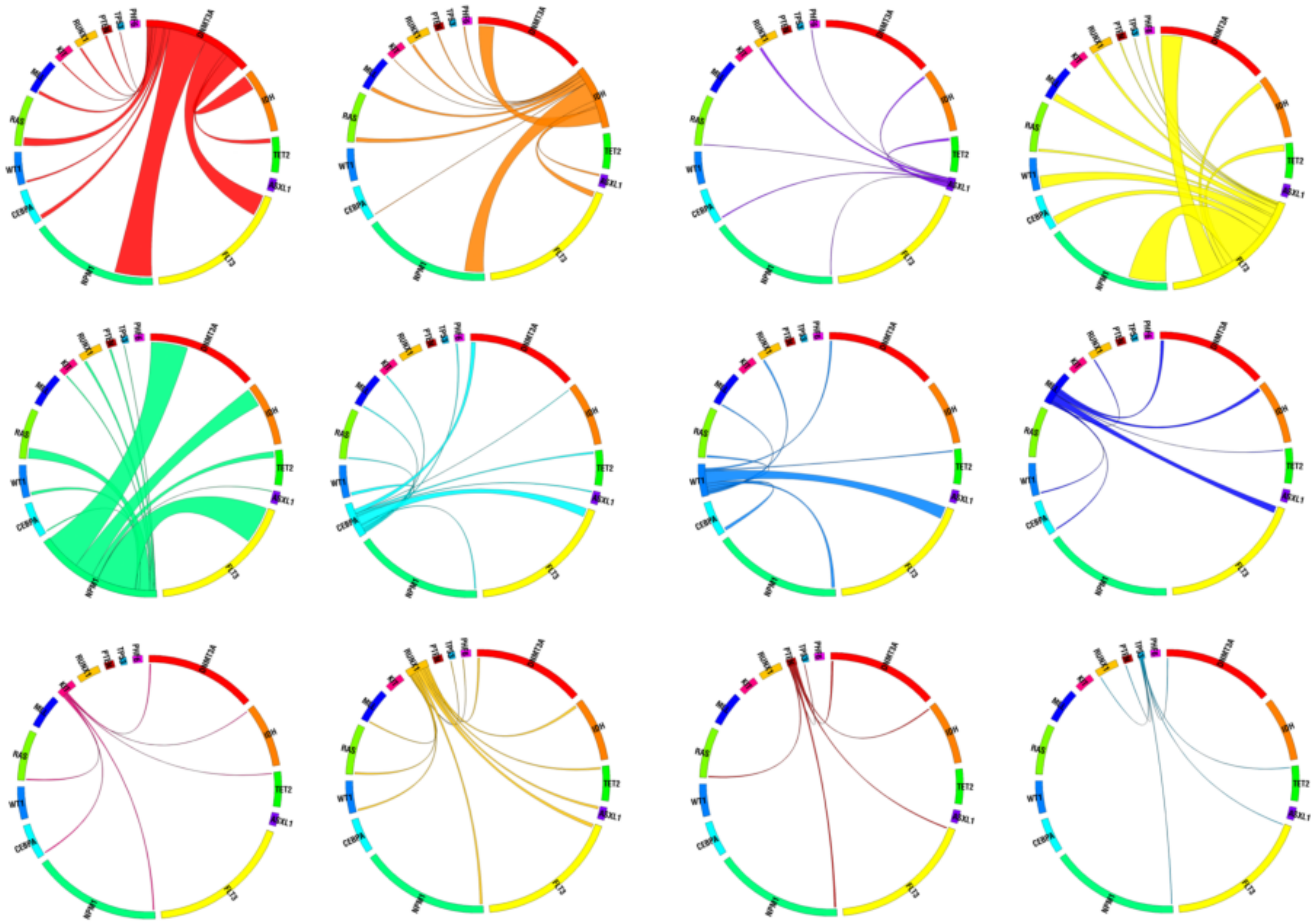
Patel JP et al (2012) N Engl J Med 366: 1079-1089.



Gene	Overall Frequency (%)
FLT3 (ITD, TKD)	37 (30, 7)
NPM1	29
DNMT3A	23
NRAS	10
CEBPA	9
TET2	8
WT1	8
IDH2	8
IDH1	7
KIT	6
RUNX1	5
MLL-PTD	5
ASXL1	3
PHF6	3
KRAS	2
PTEN	2
TP53	2
HRAS	0
EZH2	0







Patel JP et al (2012) N Engl J Med 366: 1079-1089.

I am a science enthusiast.

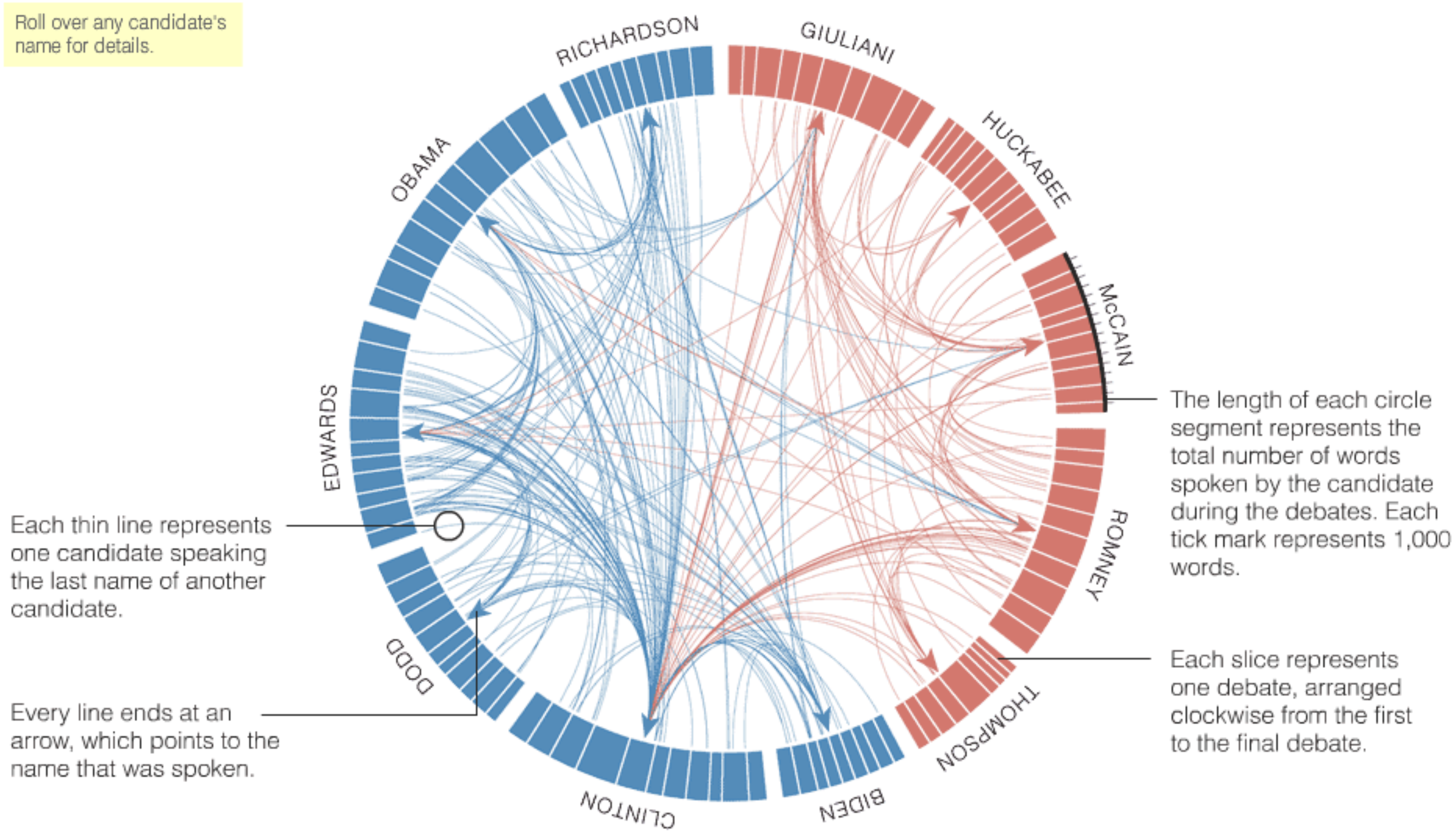
provide context → *integrate*

(inform me)

December 15, 2007

Naming Names

Names used by major presidential candidates in the series of Democratic and Republican debates leading up to the Iowa caucuses.



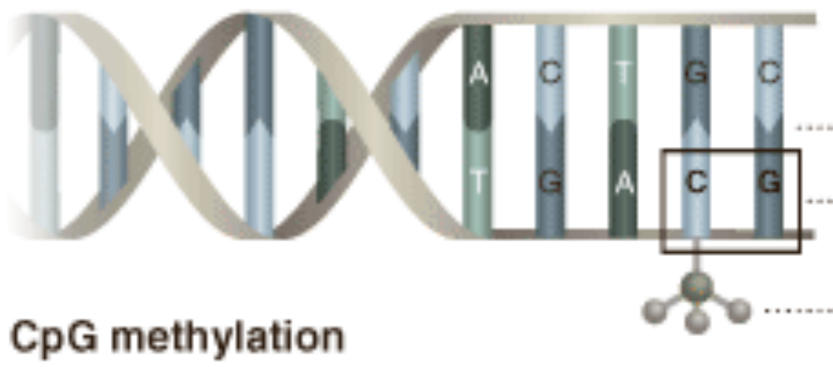
Source: Debate transcripts

Jonathan Corum and Farhana Hossain/The New York Times

The New York Times

Mapping the Epigenome

DNA contains the genetic blueprint for all human cells, but the reading and execution of the blueprint inside each cell is controlled in part by chemical markers attached to the DNA. Scientists have begun to map some of these epigenetic markers, including CpG methylation.



CpG methylation

DNA is a code written with four letters: **A**, **T**, **C** and **G**, each standing for one nucleotide.

In CpG methylation, a small marker called a methyl group attaches to the DNA at a CpG site, where a **C** and a **G** nucelotide sit next to each other.

Reading the chart

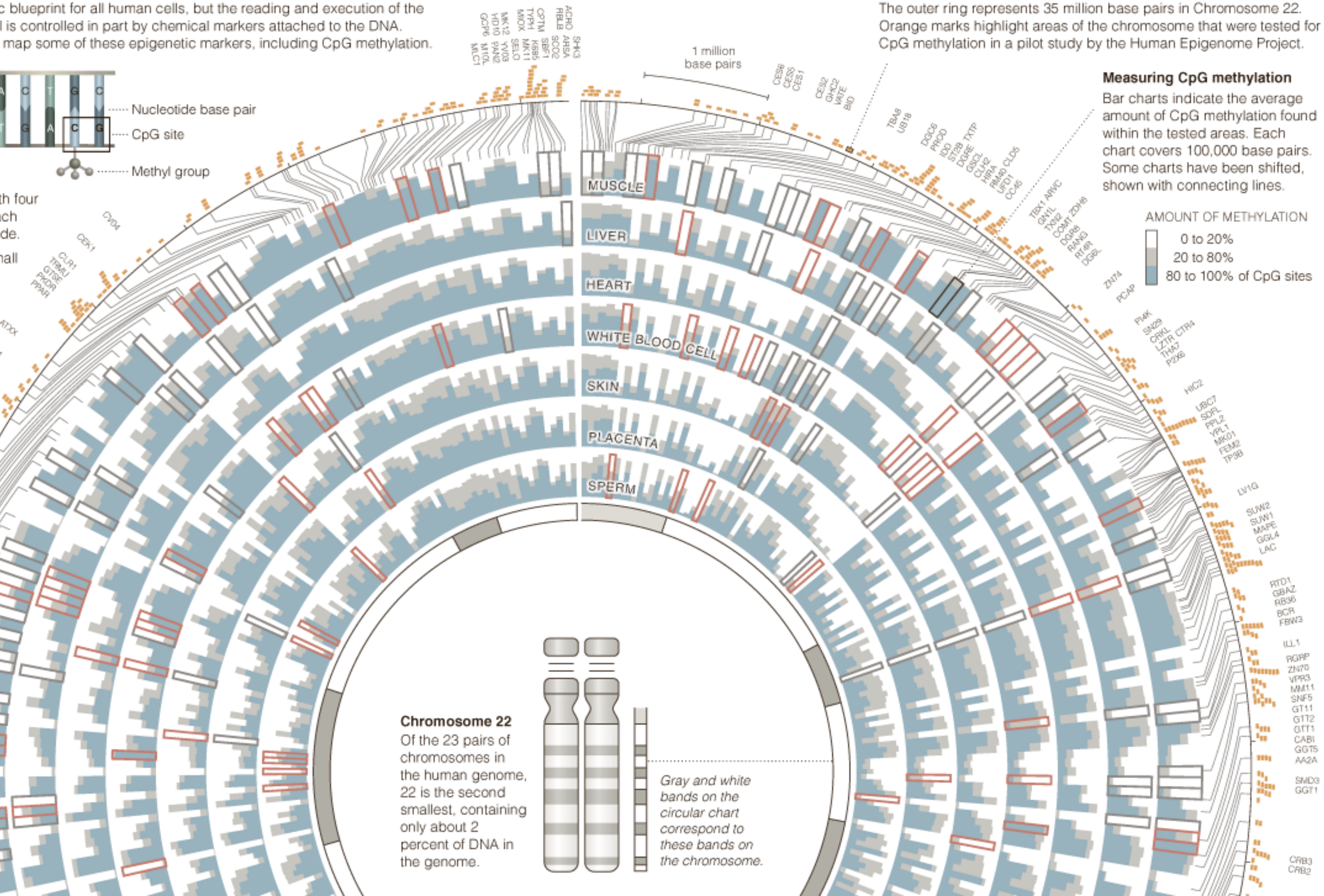
The outer ring represents 35 million base pairs in Chromosome 22. Orange marks highlight areas of the chromosome that were tested for CpG methylation in a pilot study by the Human Epigenome Project.

Measuring CpG methylation

Bar charts indicate the average amount of CpG methylation found within the tested areas. Each chart covers 100,000 base pairs. Some charts have been shifted, shown with connecting lines.

AMOUNT OF METHYLATION

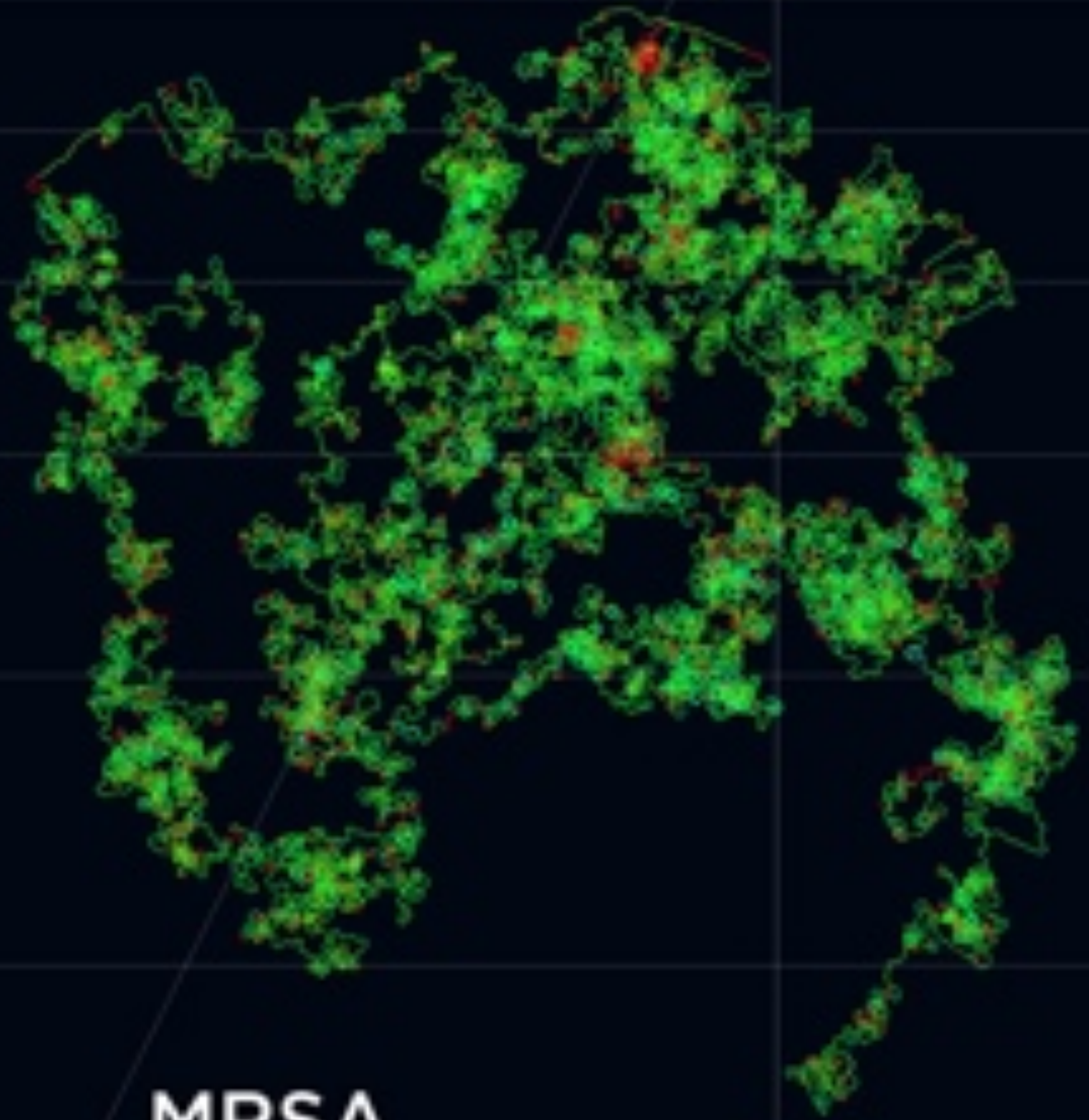
- 0 to 20%
- 20 to 80%
- 80 to 100% of CpG sites



Chromosome 22
Of the 23 pairs of chromosomes in the human genome, 22 is the second smallest, containing only about 2 percent of DNA in the genome.

Gray and white bands on the circular chart correspond to these bands on the chromosome.

What does a genome look like?



MRSA

Staphylococcus aureus 252

2,902,619 gc 33%

THE DEADLY GENOMES

GENOME STRUCTURE AND SIZE OF HARMFUL BACTERIA AND VIRUSES

READING THE CHART

The chart shows the genomes of a variety of bacteria ○ and viruses □ that cause human disease.

The x-axis represents the **disease burden**, the average number of worldwide deaths due to the disease.

The y-axis depicts **mortality**, the percentage of worldwide cases that result in death.

Each colored line represents the genome of the bacterium or virus that causes the disease. The size of the genome and the percentage of guanine or cytosine in the genome (GC content) are also noted. For viruses, the genomes of several strains are shown when possible.

THE GENOME AS A PATH

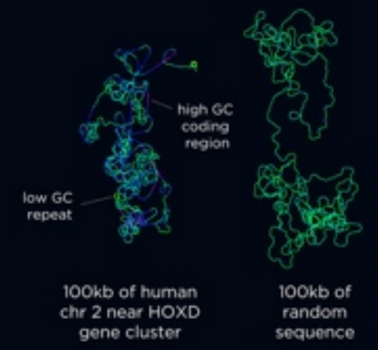
The genome of each bacterium or virus is represented by a continuous line or path. The length of the path is proportional to the size of the genome. At each point along the path, color is used to show the GC content near that position:

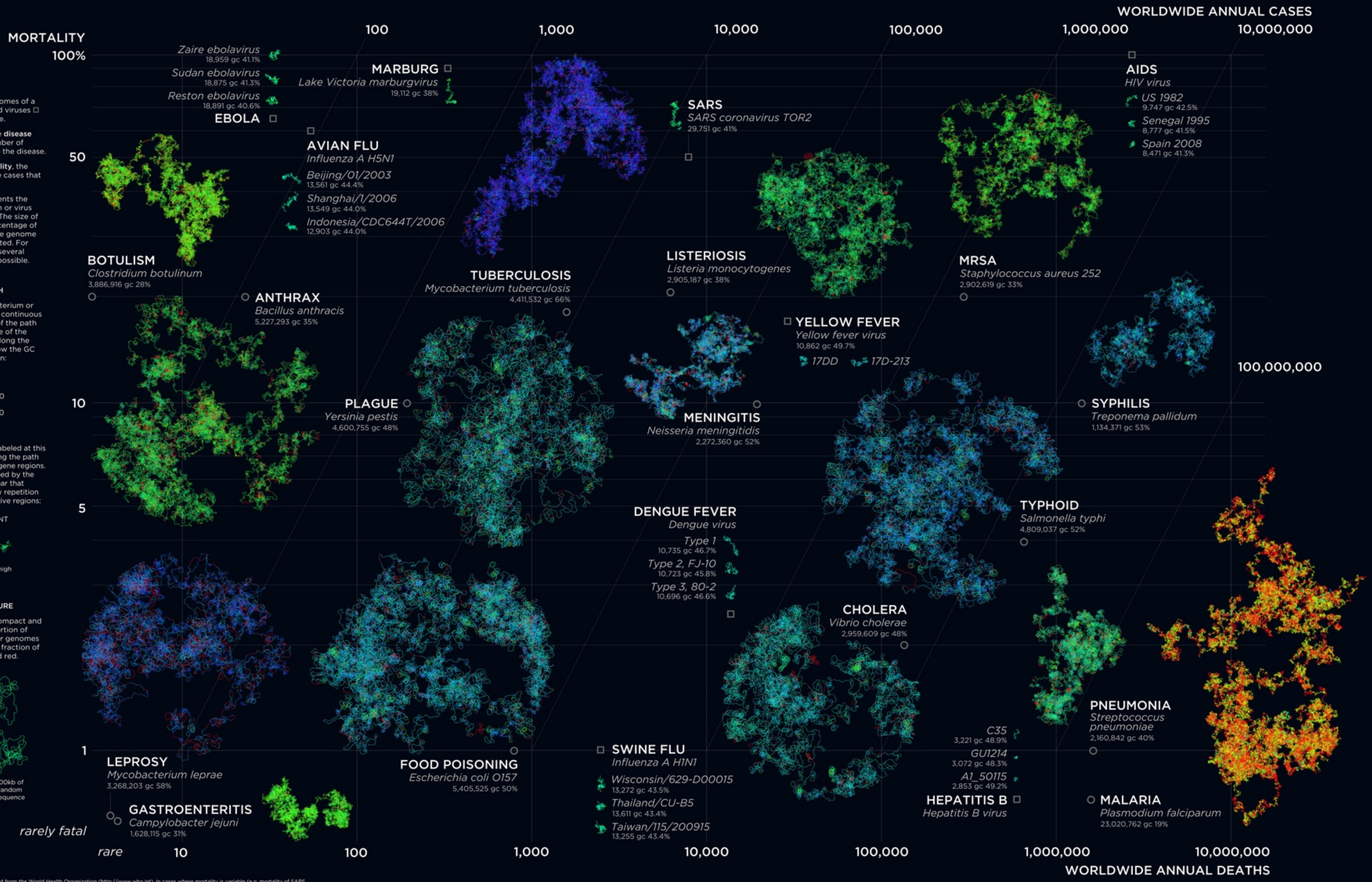
Individual genes are not labeled at this scale, but red regions along the path indicate breaks between gene regions. Path direction is determined by the repeat and GC content near that location. Regions with low repetition are straighter than repetitive regions:

REPEAT CONTENT

INTERPRETING STRUCTURE

Bacterial genomes are compact and gene rich. The low proportion of intergenic regions in their genomes can be seen by the small fraction of the genome path colored red.





Data Sources

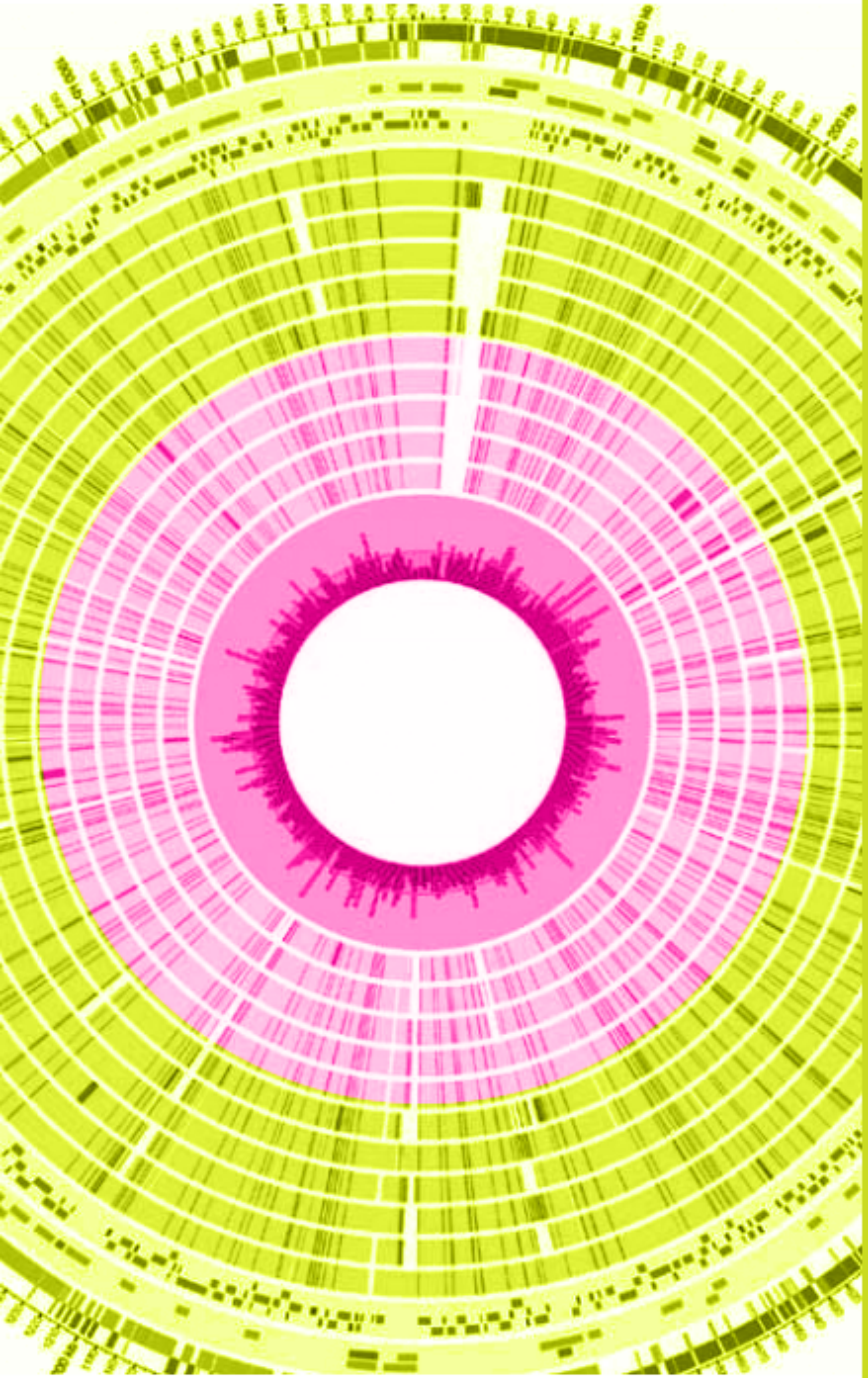
Incidence and mortality data were obtained from the World Health Organization (http://www.who.int). In cases where mortality is variable (e.g. mortality of SARS ranges from 1-50%, depending on age), the highest value is reported. For some outbreak diseases, such as Ebola, statistics were averaged over several years.

The distribution of cases and mortality across world regions for most diseases shown here is not uniform. For example, haemorrhagic fevers like Marburg, Dengue and Ebola are largely limited to Africa.

I am a layperson.

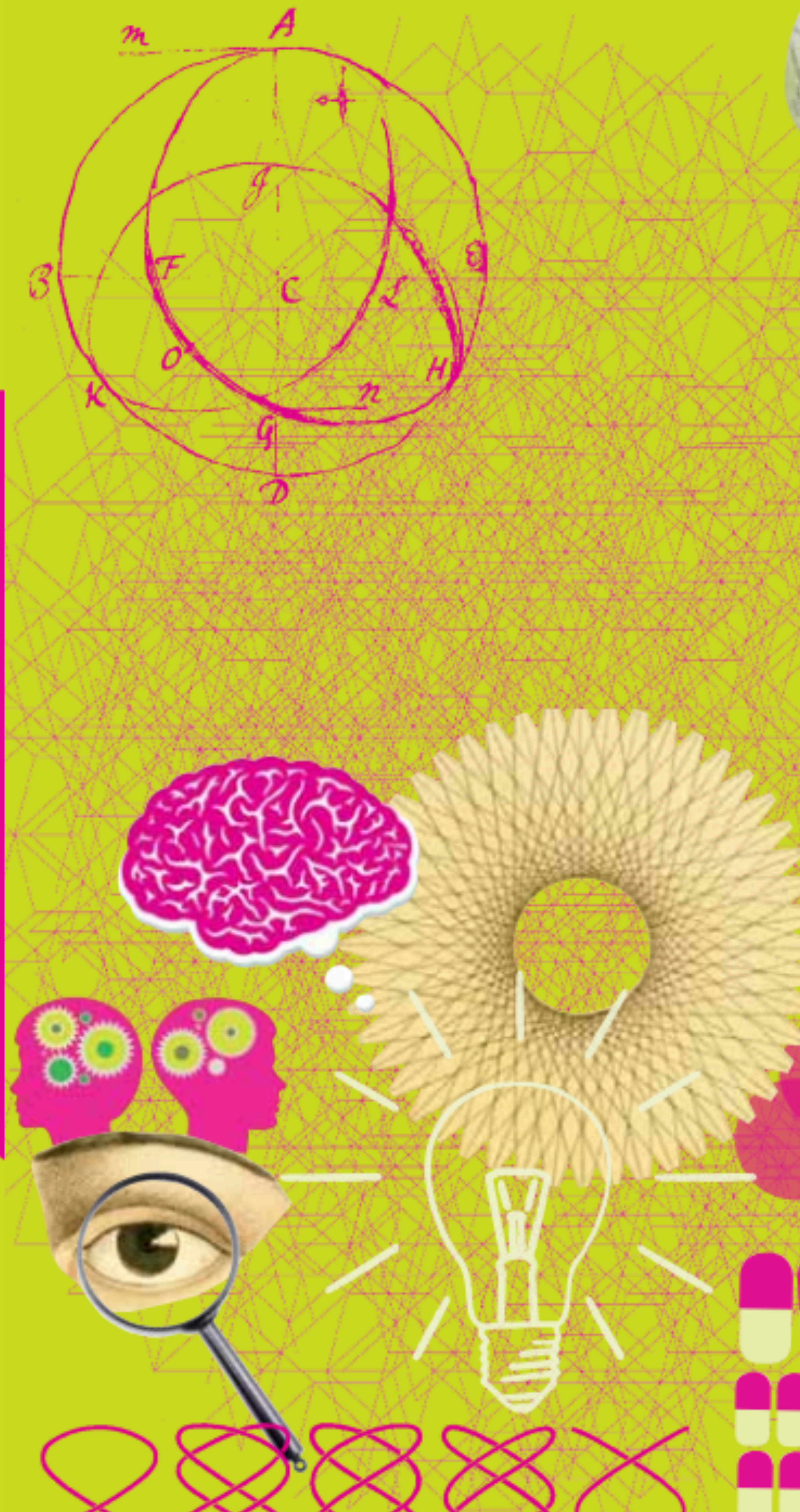
capture essence → narrate

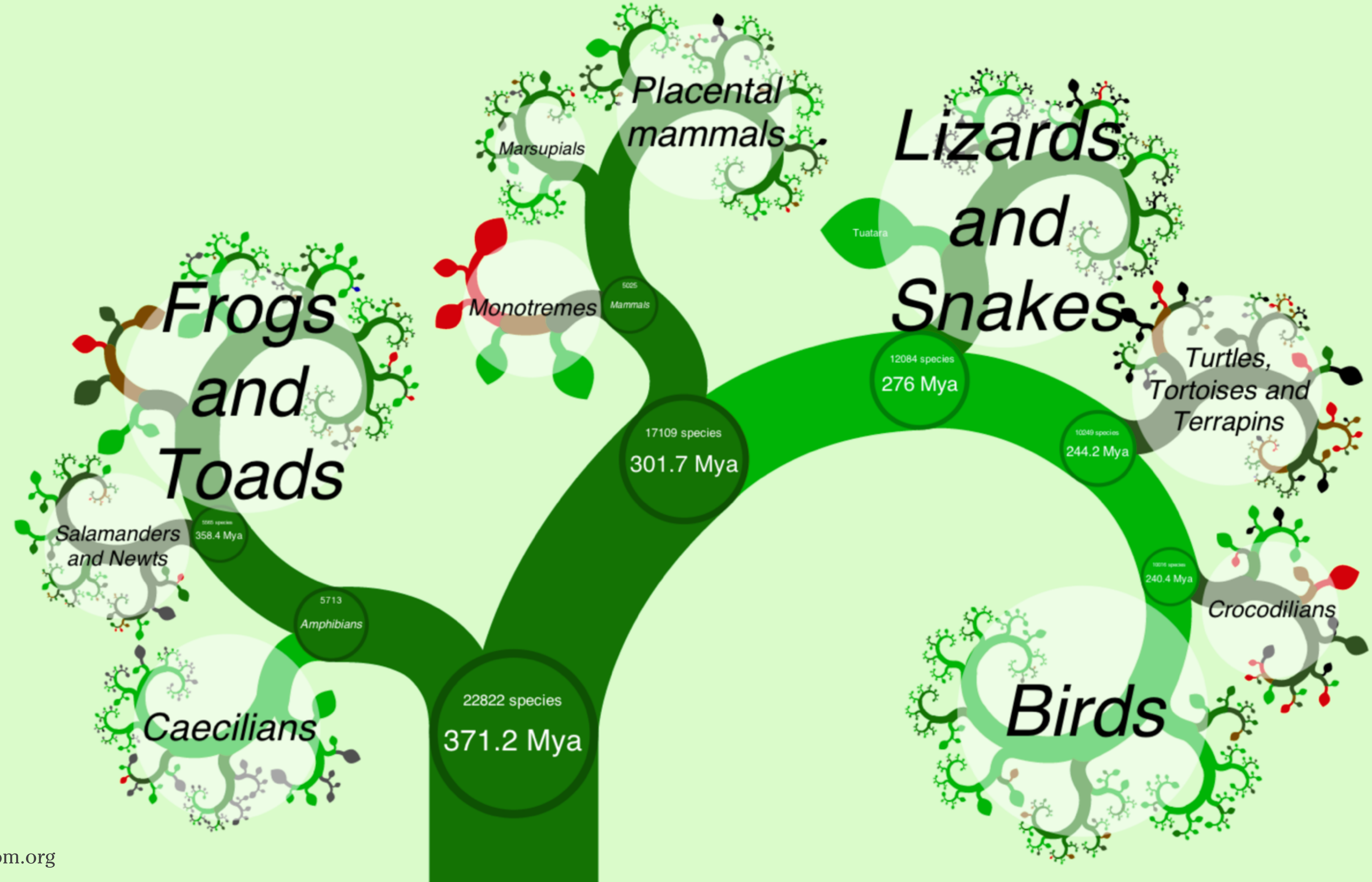
(delight me)



Beautiful Science

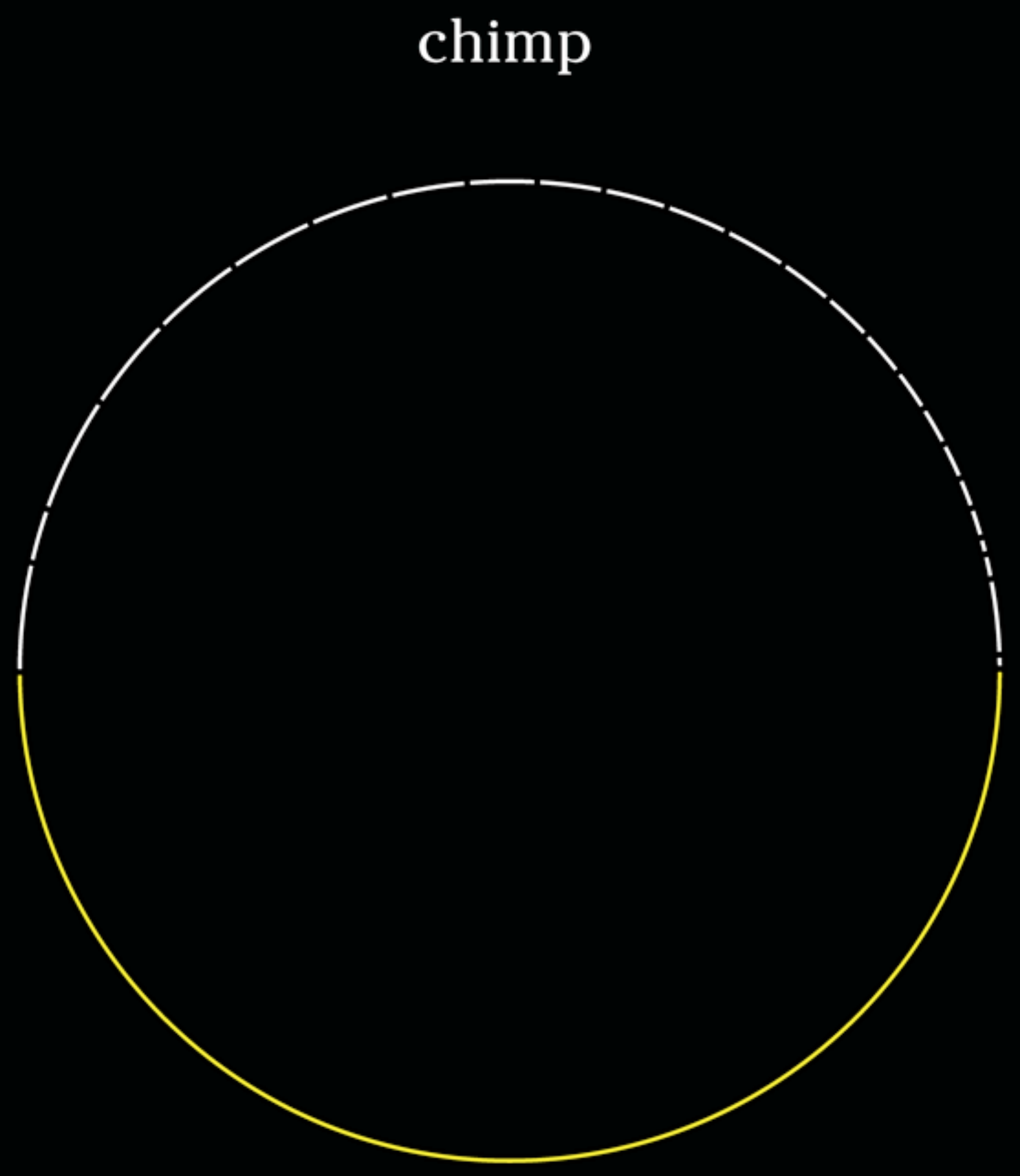
An exhibition at the British Library





http://onezoom.org

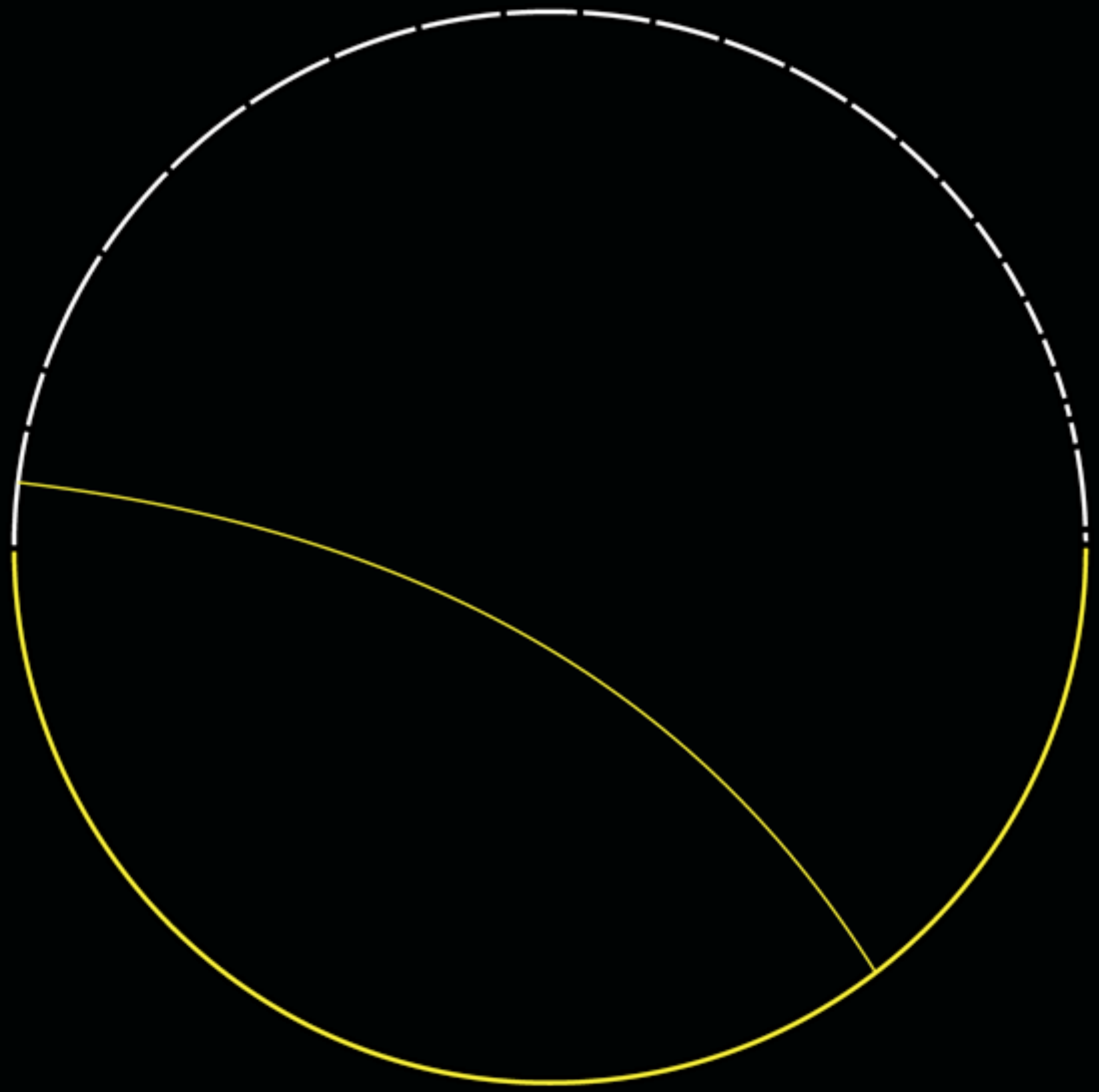
animal genome



human chromosome 1

animal genome

chimp



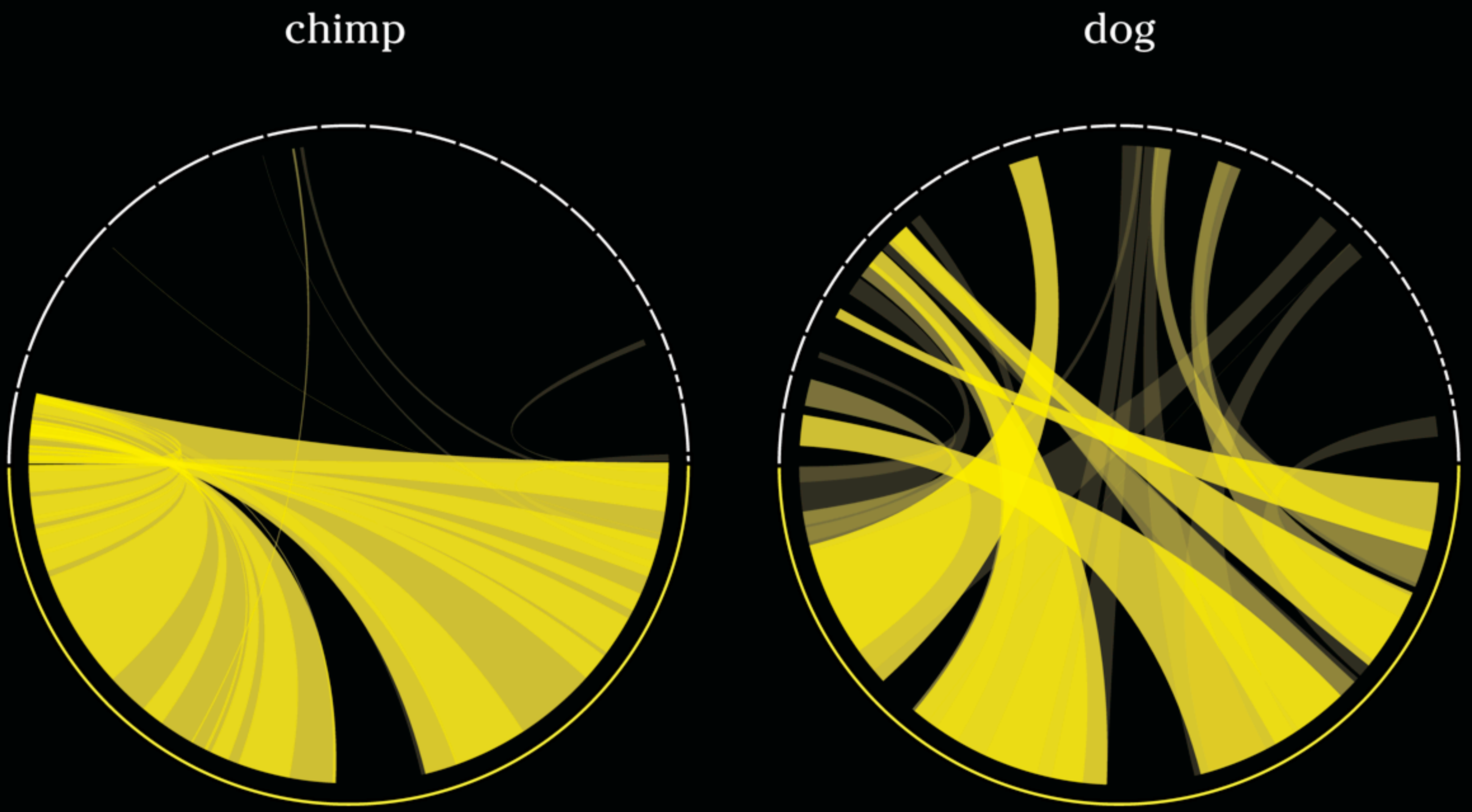
human chromosome 1

animal genome

chimp

human chromosome 1

animal genome



human chromosome 1


animal genome
chimp
dog
lizard
human chromosome 1

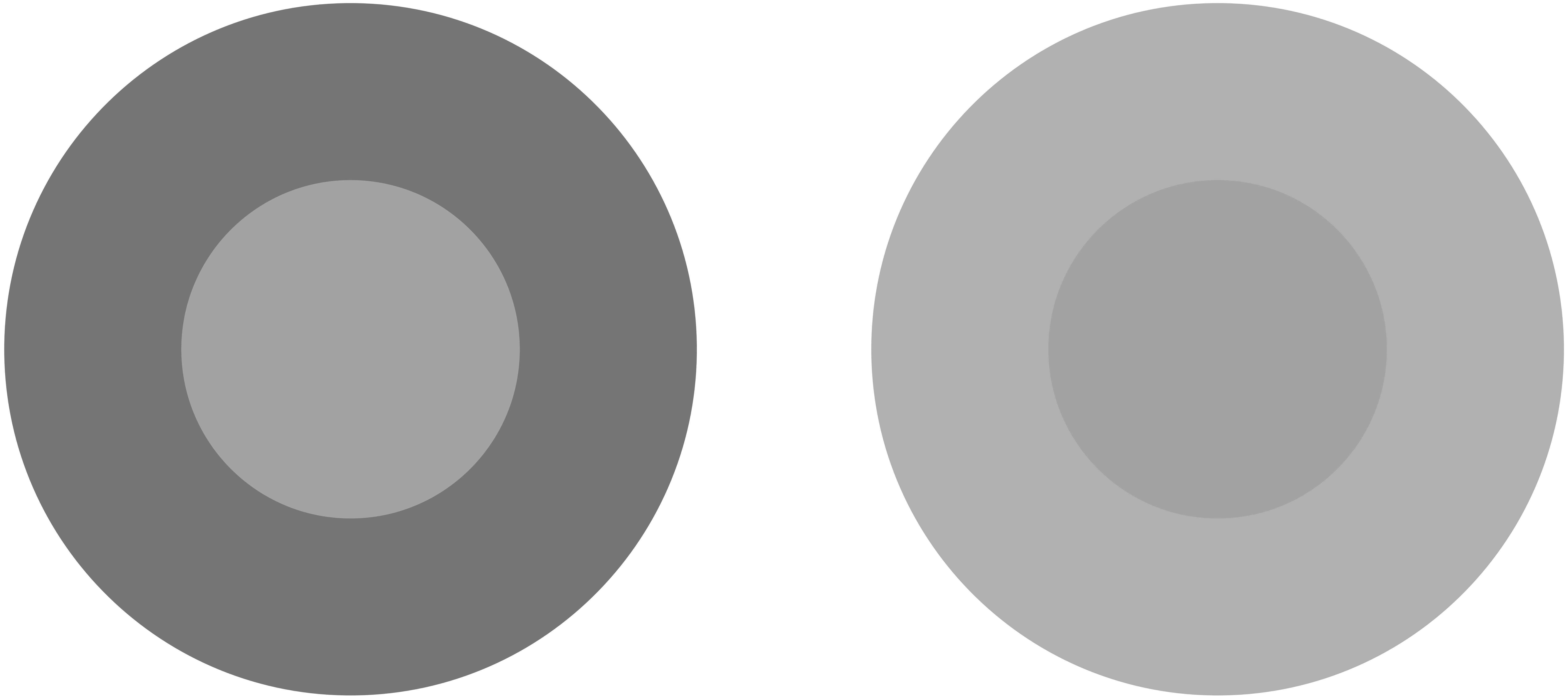

animal genome
chimp
dog
lizard
human chromosomes 1-22,X,Y

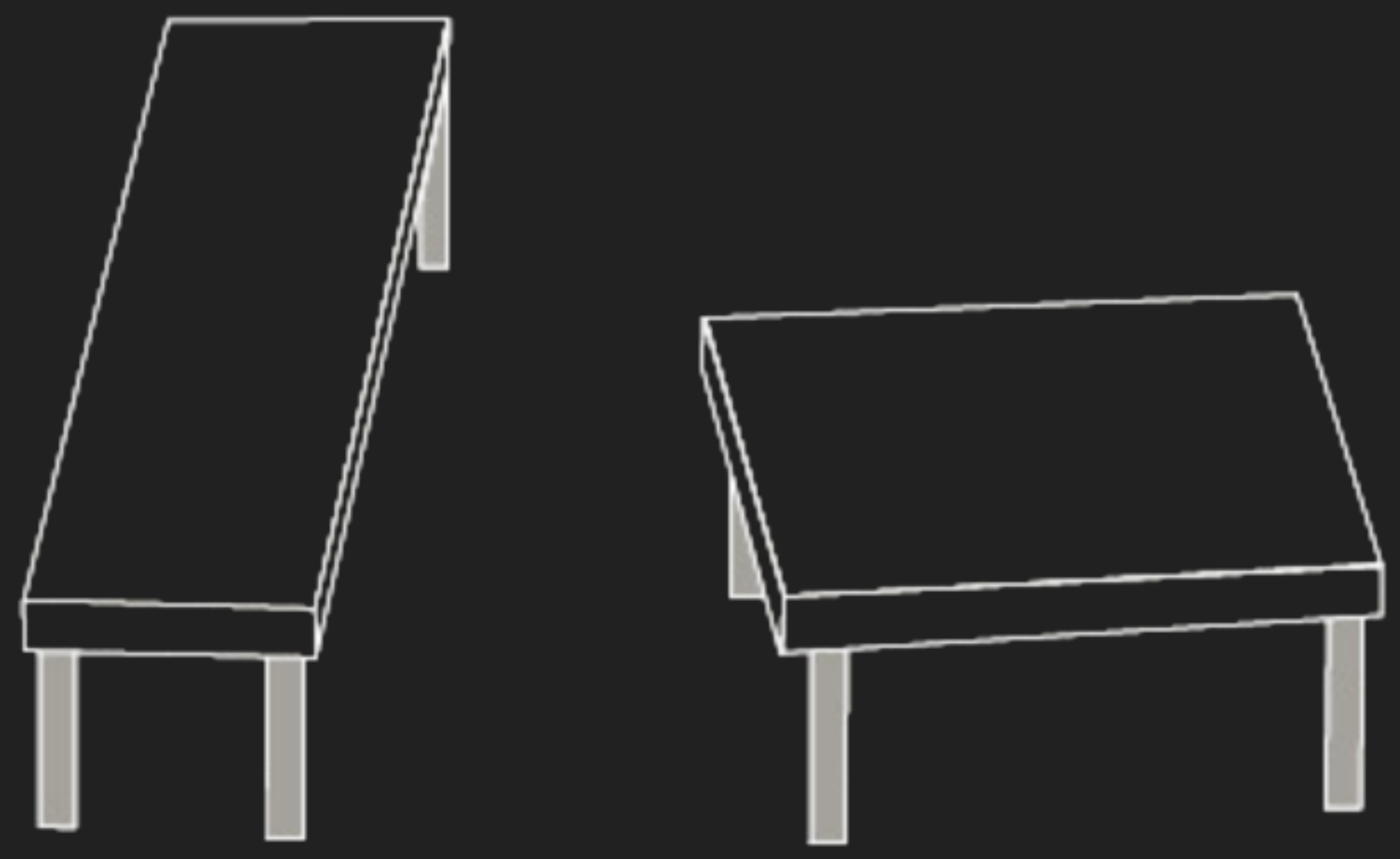
chimp
gorilla
orangutan
rhesus
marmoset
horse
cow
dog
mouse
opossum
stickleback
tetraodon
medaka
lizard
chicken
zebrafinch
platypus

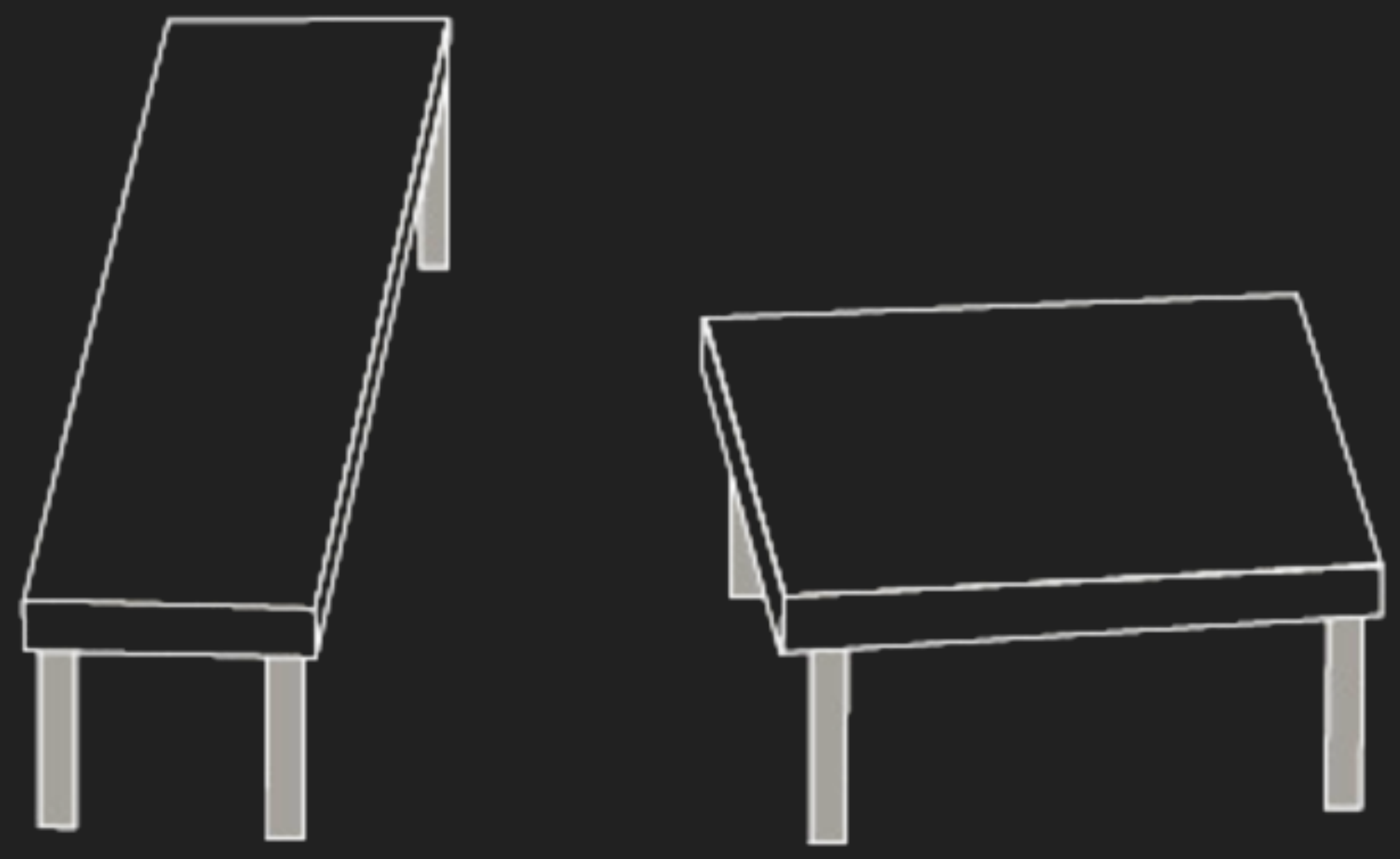
Specialists commonly overrate their ability to create and interpret scientific figures.

Our eye can be easily fooled.

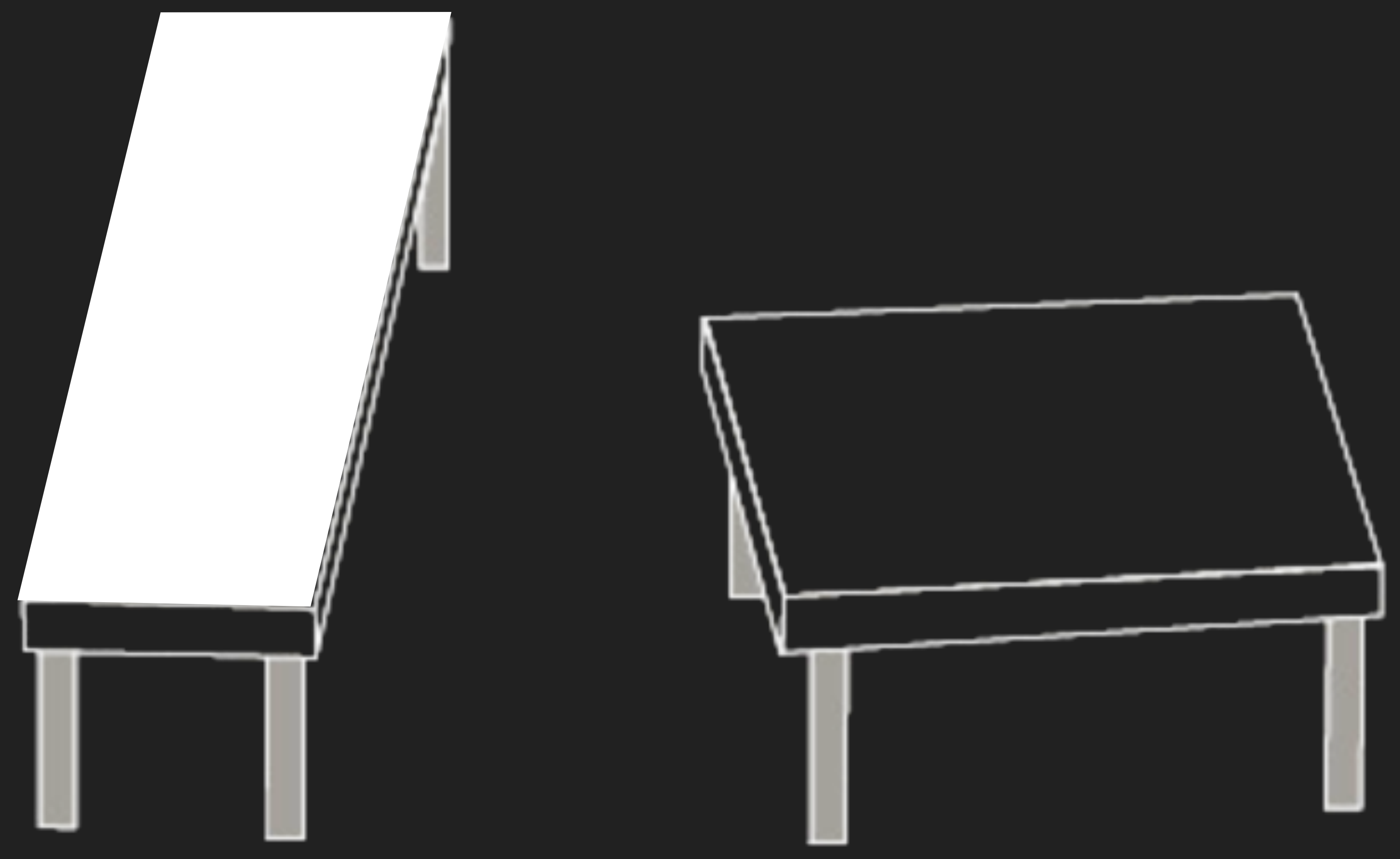
(and we don't even know it)



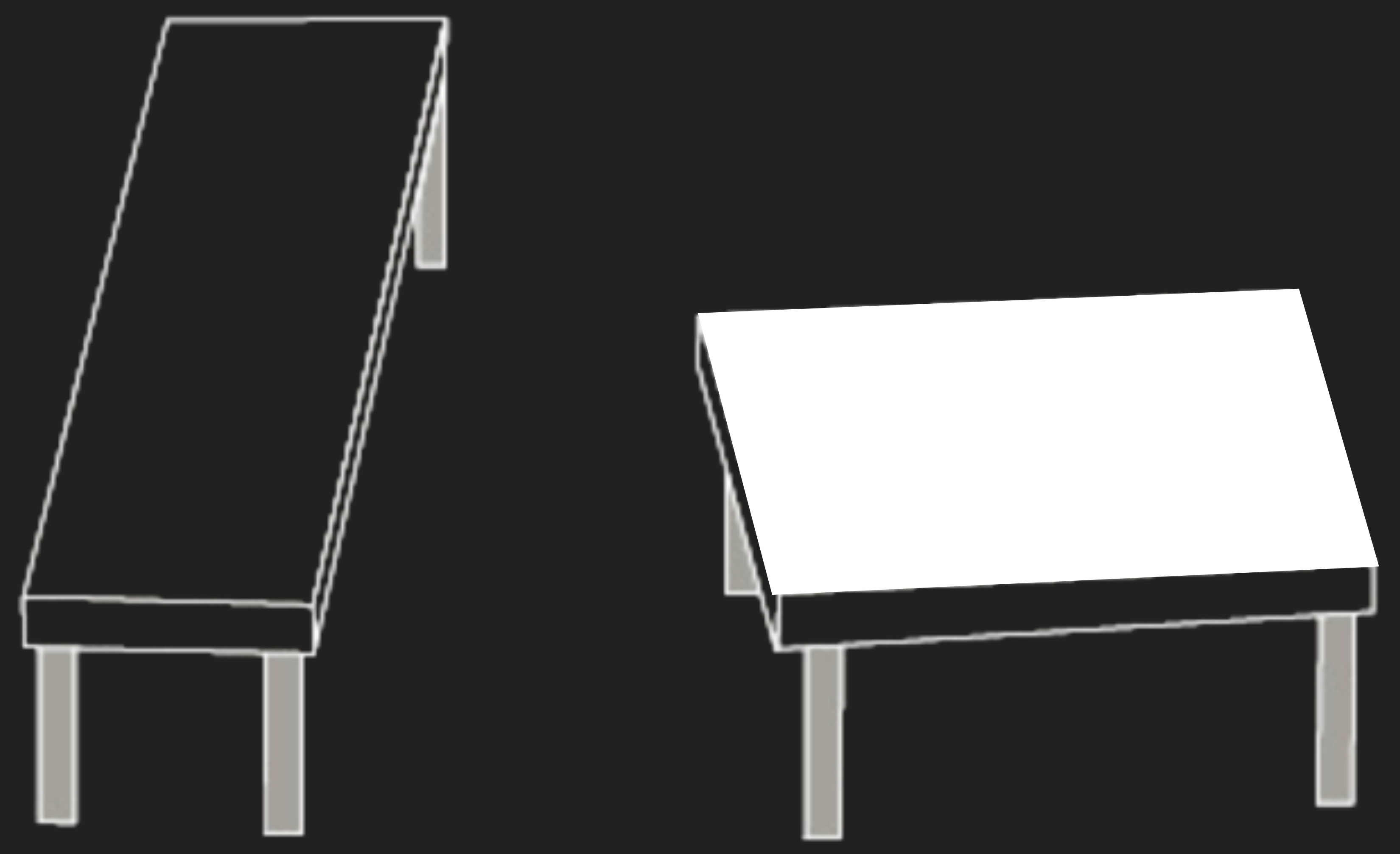




Shepard’s tables.



Shepard’s tables.



Shepard’s tables.

ISNN 1088-9051

November 2009

Genome Research

Volume 19 Number 11

Human Coexpression Network • Allele-specific DNA Methylation in Mice • Ancient Noncoding Sequences Preserve Synteny • Singapore Genome Variation Project

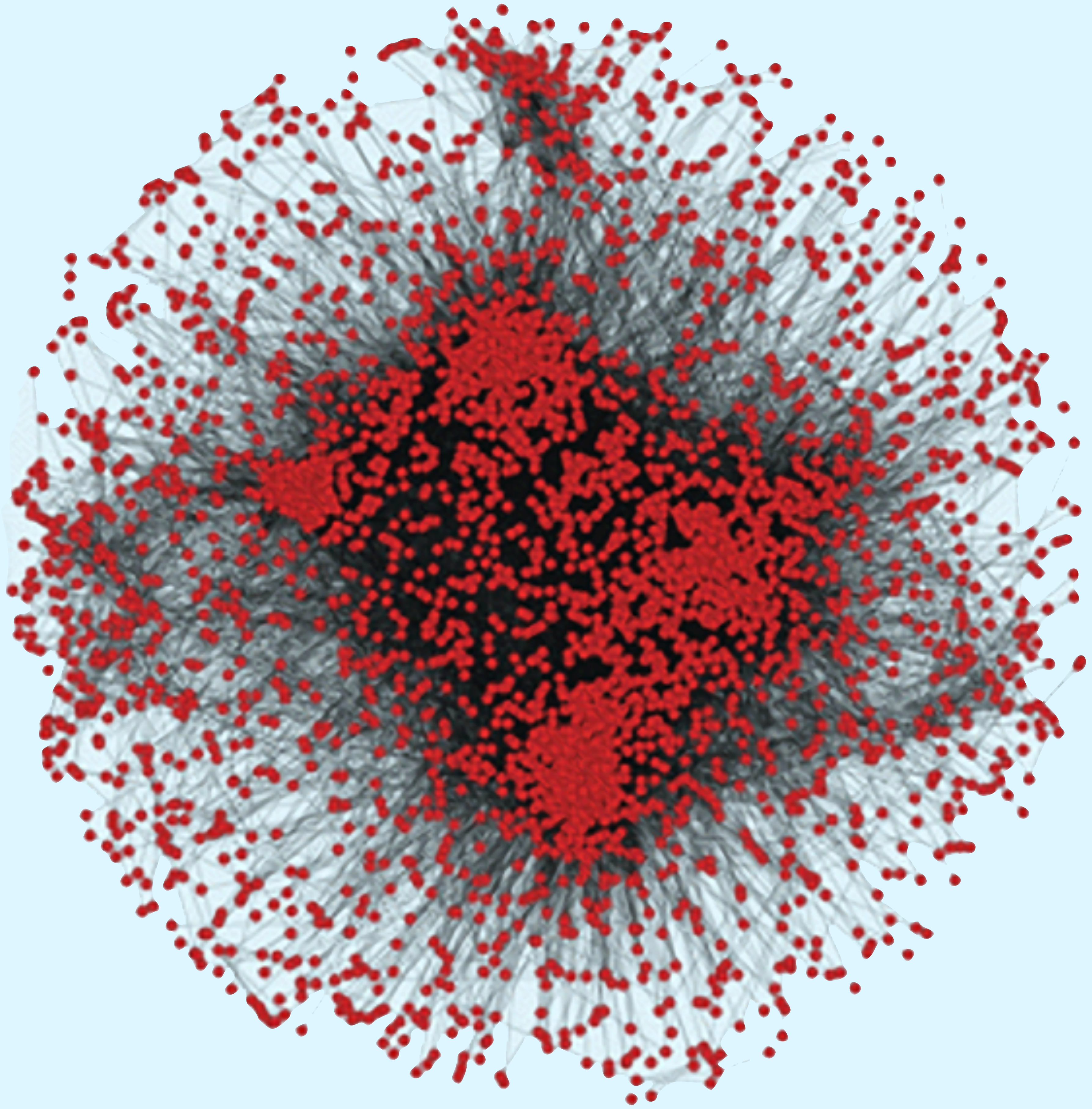
Cold Spring Harbor Laboratory Press

ISNN 1088-9051

November 2009

Genome Research

Volume 19 Number 11



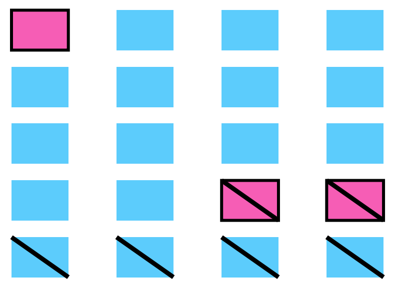
Human Coexpression Network • Allele-specific DNA Methylation in Mice • Ancient Noncoding Sequences Preserve Synteny • Singapore Genome Variation Project

Cold Spring Harbor Laboratory Press

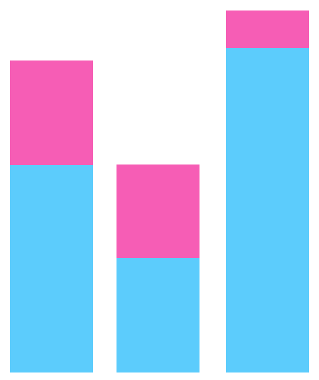
BMJ

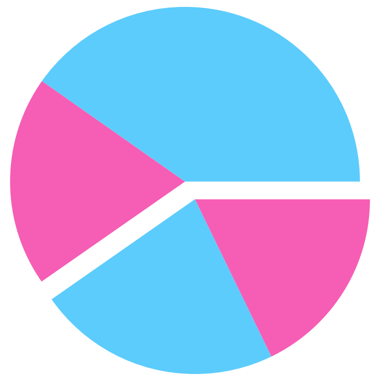
Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures

Linda S Elting, Charles G Martin, Scott B Cantor, Edward B Rubenstein



30	30	35	11
20	45	25	12
50	38	60	12





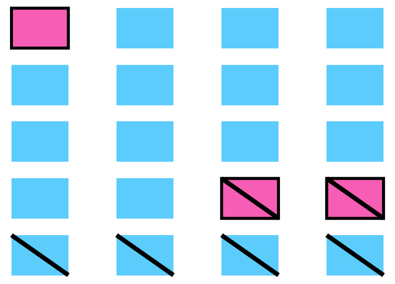
Elting et al. (1999) BMJ 316.

BMJ

Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures

Linda S Elting, Charles G Martin, Scott B Cantor, Edward B Rubenstein

accuracy

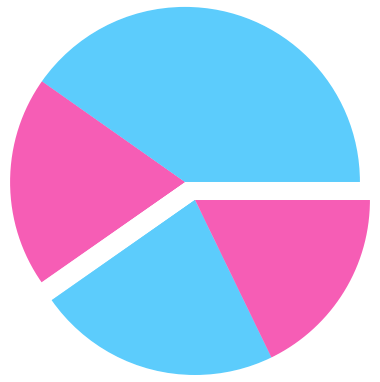


82

30	30	35	11
20	45	25	12
50	38	60	12

68

56



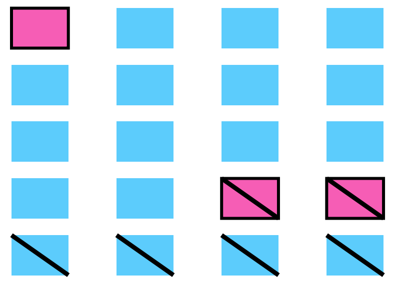
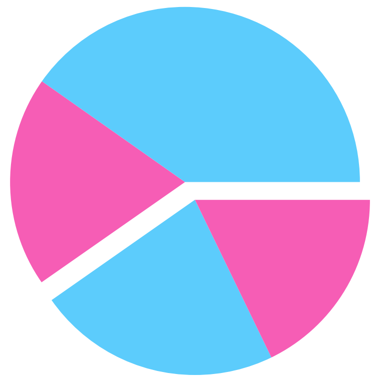
56

Elting et al. (1999) BMJ 316.

BMJ

Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures

Linda S Elting, Charles G Martin, Scott B Cantor, Edward B Rubenstein

	accuracy	preference
	82	0
30 30 35 11 20 45 25 12 50 38 60 12	68	62
	56	24
	56	14

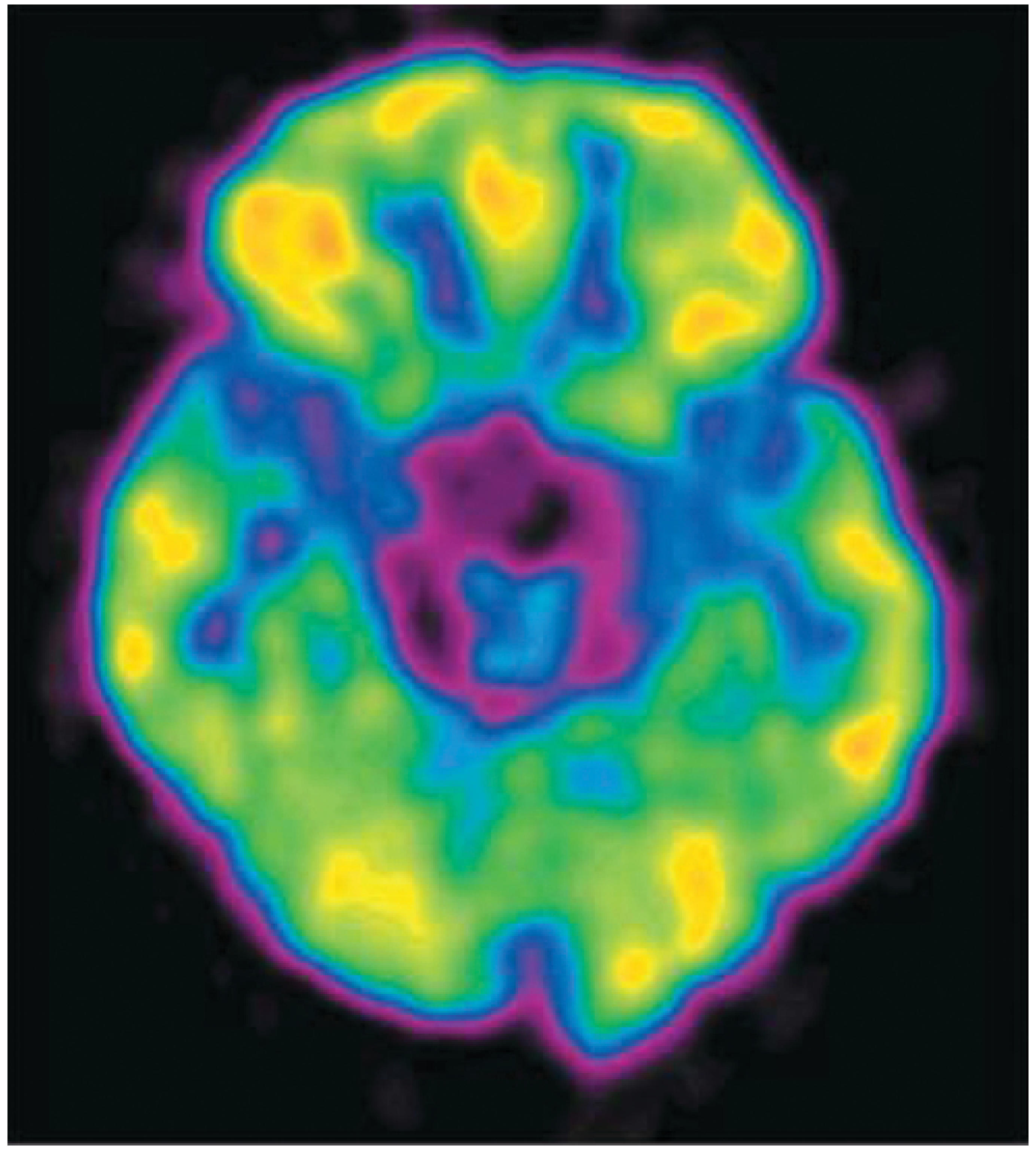
Elting et al. (1999) BMJ 316.

…eight voiced considerable contempt for the [icon] display.

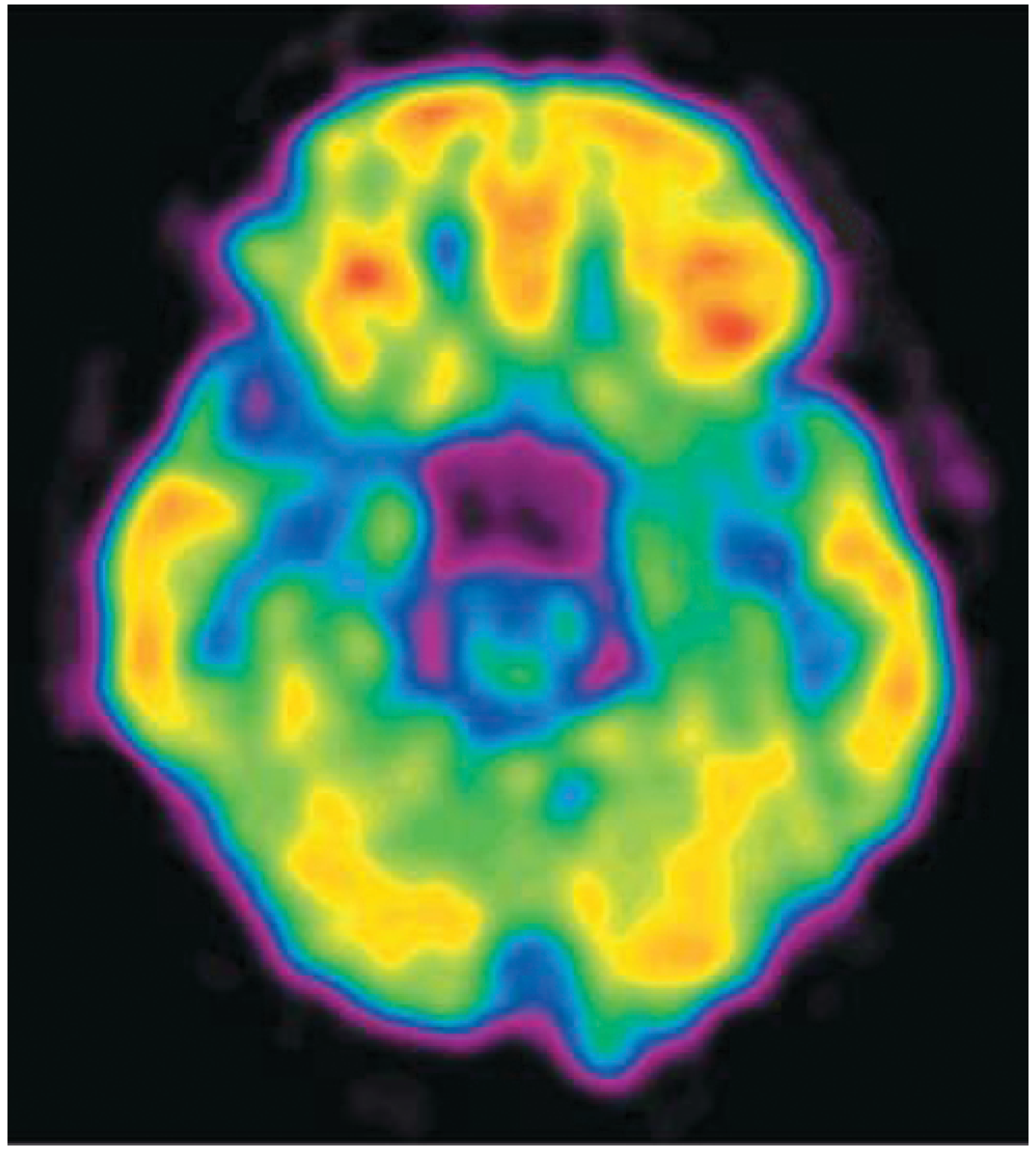
Elting et al. (1999) BMJ 316.

Cell Phones and the Brain

cellphone off



cellphone on

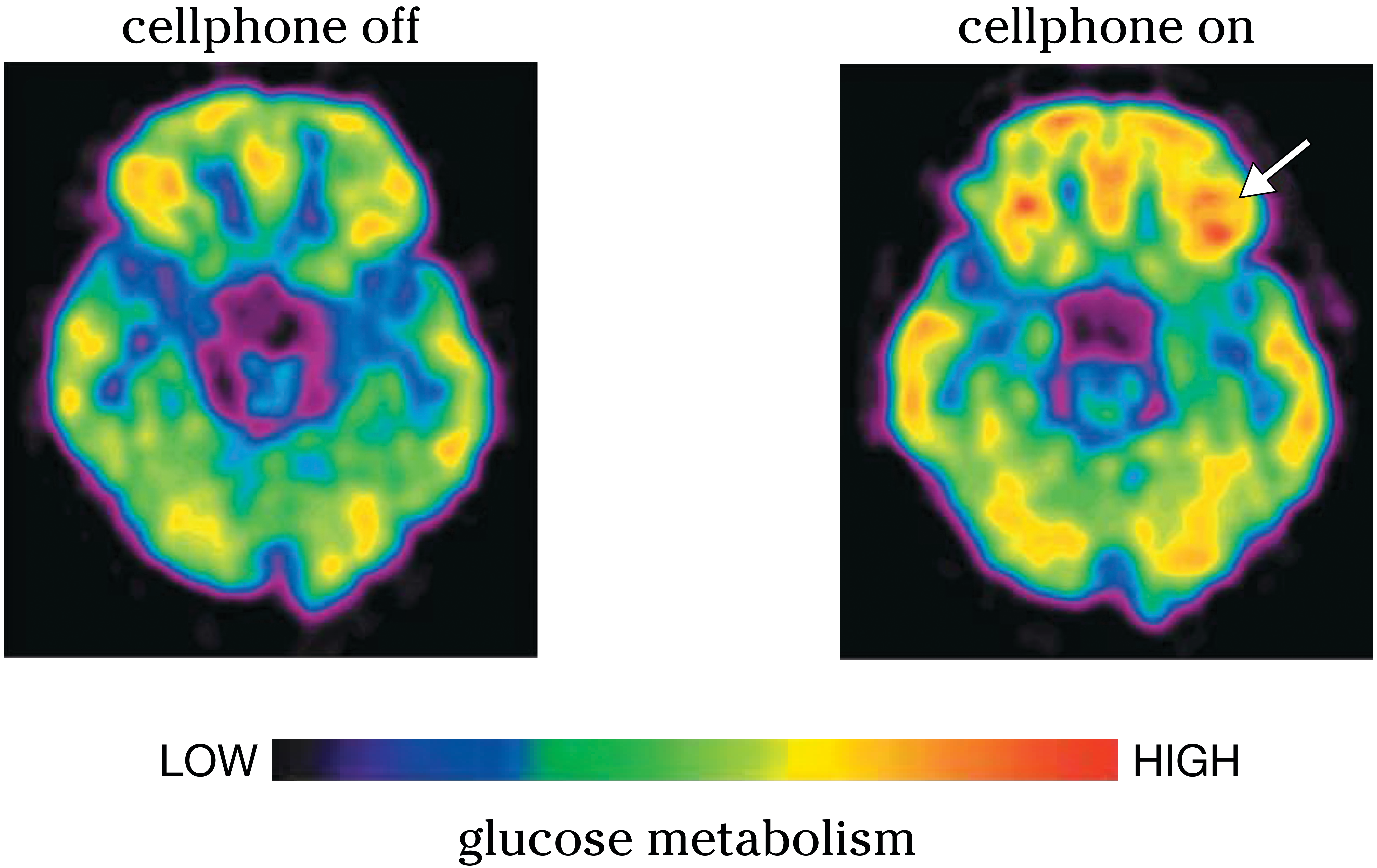


LOW  HIGH

glucose metabolism

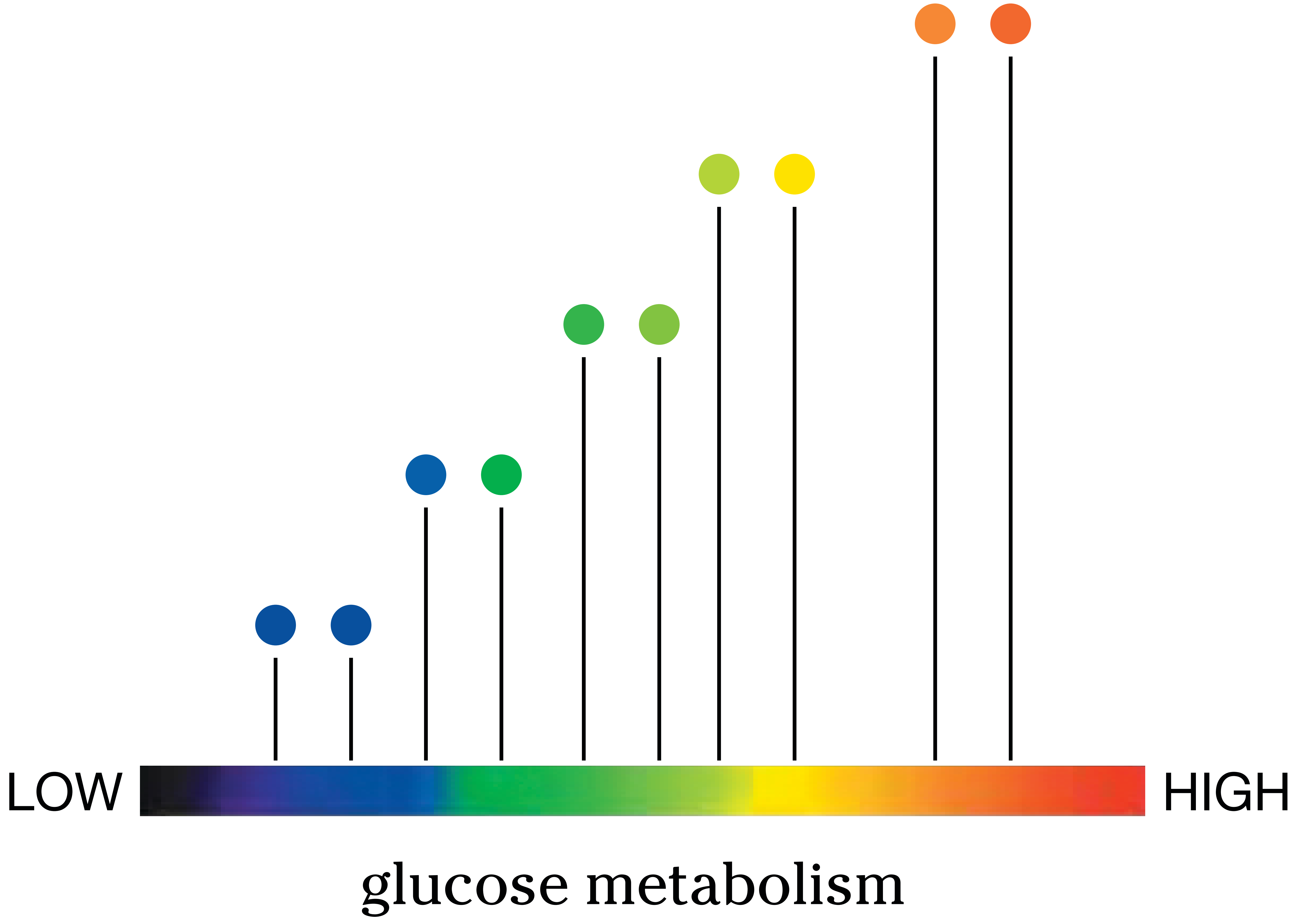
Volkov et al (2011) JAMA 305(8): 808–813.

Cell Phones and the Brain

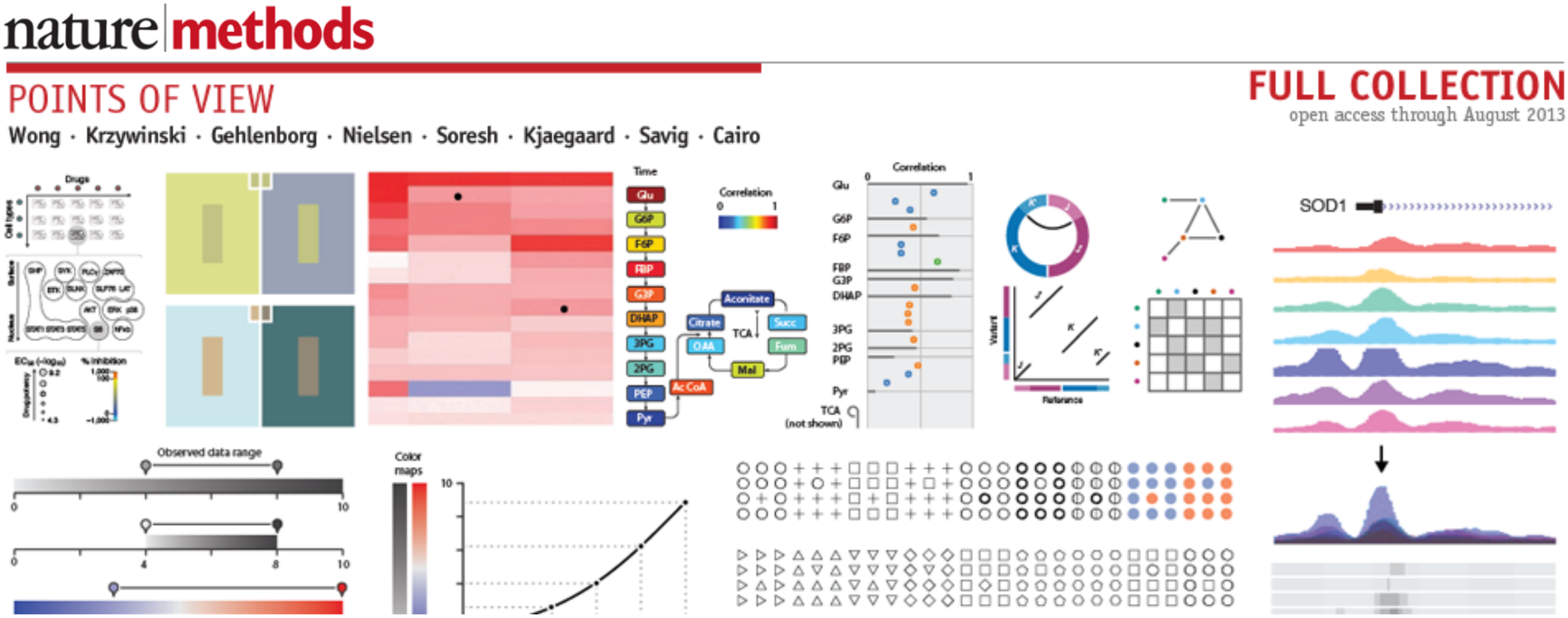


Volkov et al (2011) JAMA 305(8): 808–813.

Cell Phones and the Brain

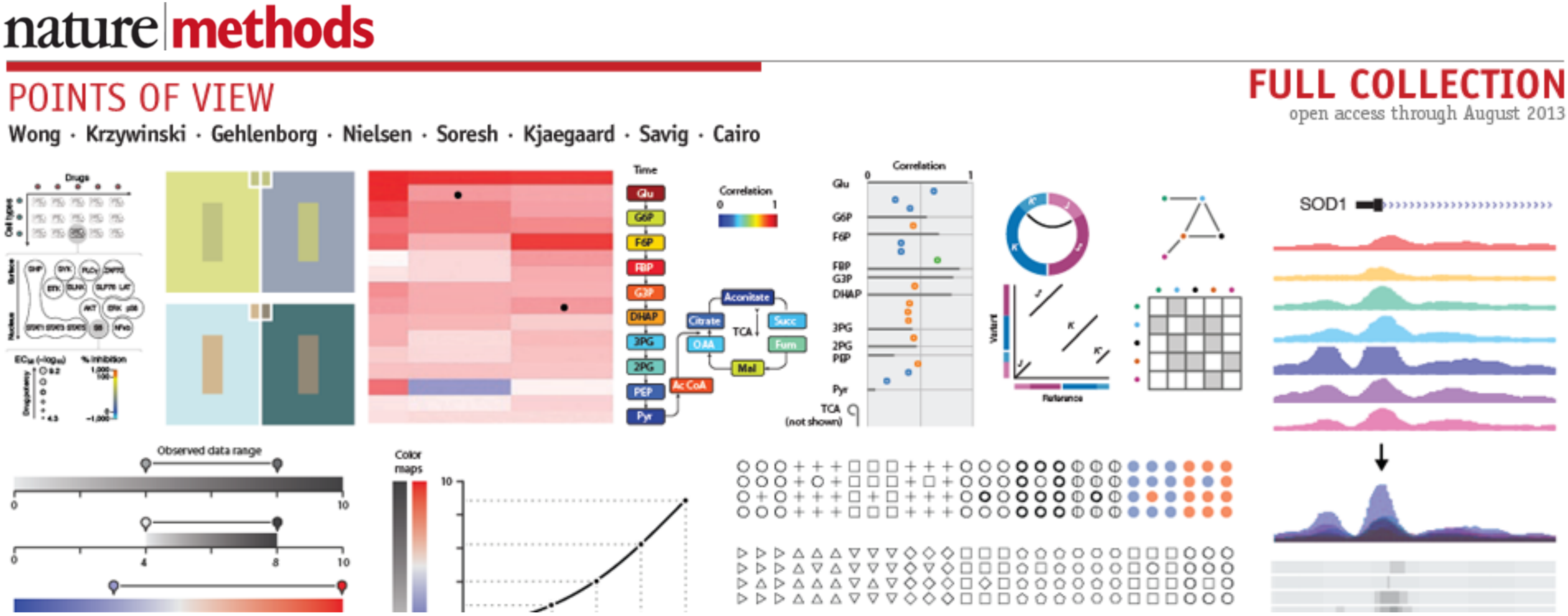


How to get started in visualization?



http://blogs.nature.com/methagora/2013/07/data-visualization-points-of-view.html

How to get started in visualization?



http://blogs.nature.com/methagora/2013/07/data-visualization-points-of-view.html