

# Principles of data visualization

Janos Binder

[janos.binder@embl.de](mailto:janos.binder@embl.de)



# Agenda

---

- introduction to visualization principles
- group work with a figure
- presentations about group work

# Acknowledgements

---

- These slides are based on tutorials of Jessie Kennedy, Martin Krzywinski and Tamara Munzner

# **YOU WANT TO BE A BETTER COMMUNICATOR NOT A BETTER ARTIST**

Do not rely solely on your personal aesthetic.

Strive for simplicity and clarity.

# Why do we visualize?

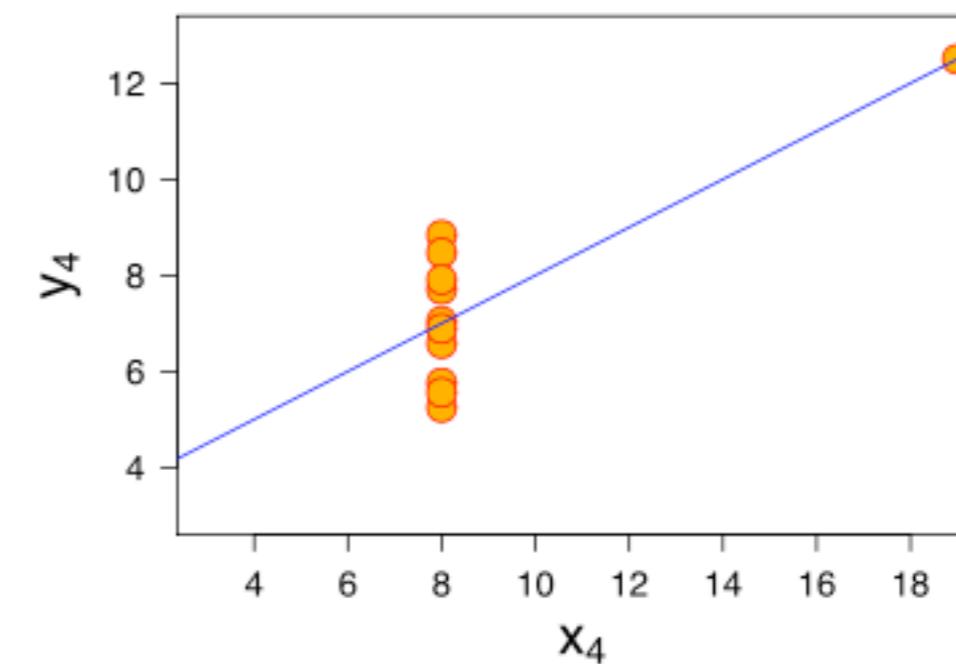
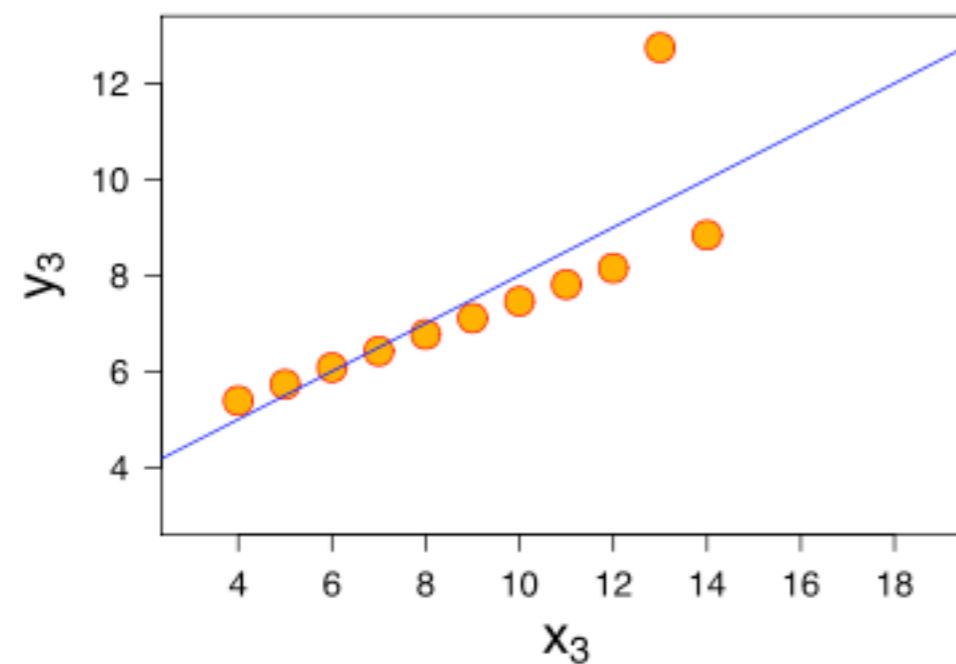
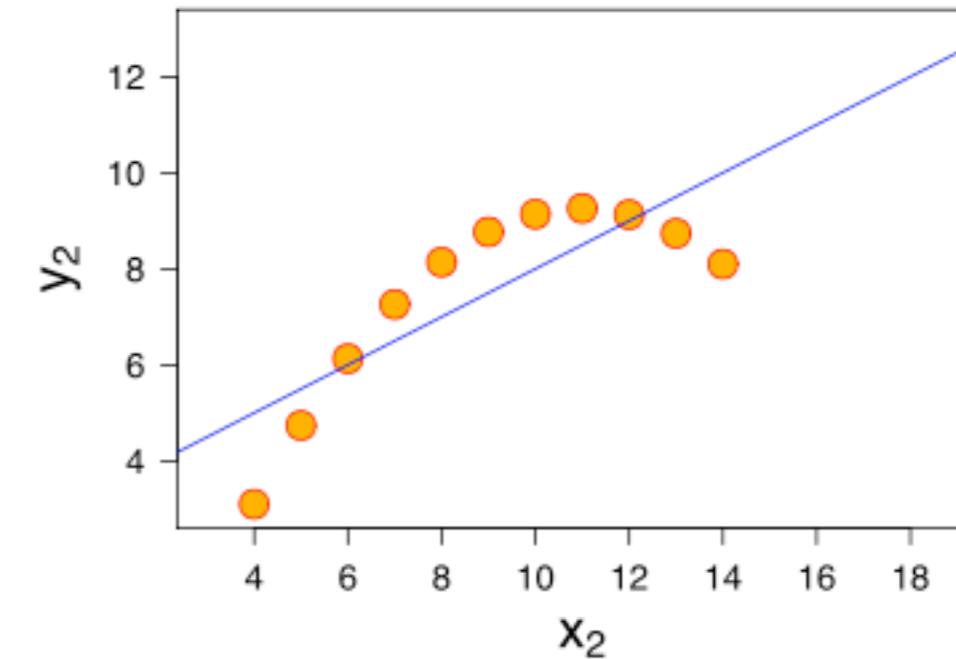
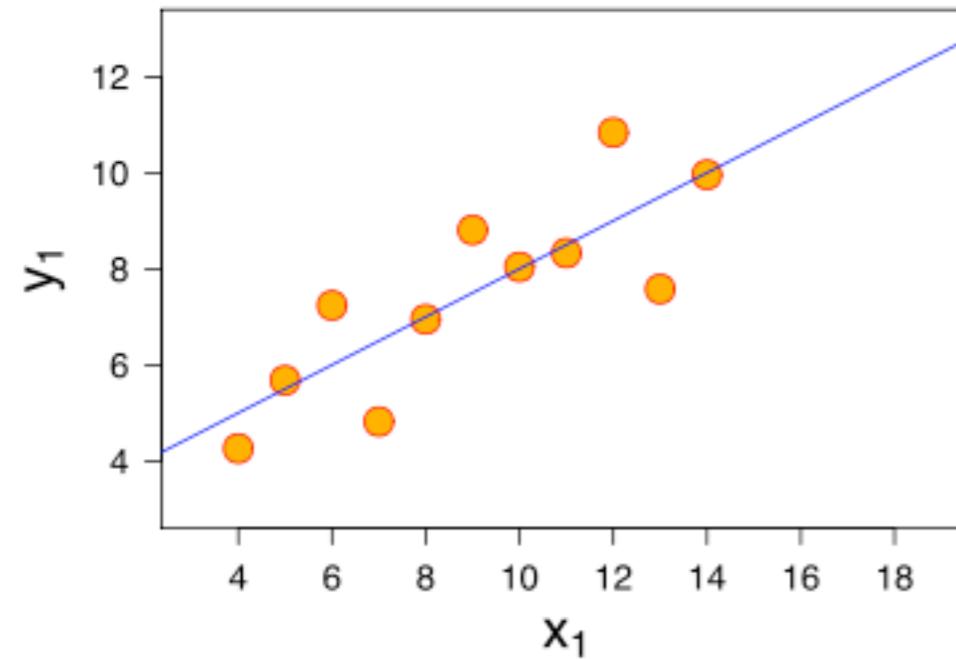
---

Anscombe's quartet								
I		II		III		IV		
X	Y	X	Y	X	Y	X	Y	
10	8,04	10	9,14	10	7,46	8	6,58	
8	6,95	8	8,14	8	6,77	8	5,76	
13	7,58	13	8,74	13	12,74	8	7,71	
9	8,81	9	8,77	9	7,11	8	8,84	
11	8,33	11	9,26	11	7,81	8	8,47	
14	9,96	14	8,1	14	8,84	8	7,04	
6	7,24	6	6,13	6	6,08	8	5,25	
4	4,26	4	3,1	4	5,39	19	12,5	
12	10,84	12	9,13	12	8,15	8	5,56	
7	4,82	7	7,26	7	6,42	8	7,91	
5	5,68	5	4,74	5	5,73	8	6,89	

Same statistical properties:  $\text{mean}(X) = 9$ ,  $\text{var}(X) = 11$ ,  $\text{mean}(Y) = 7.5$ ,  
 $\text{var}(Y) = 4.12$ ,  $\text{cor}(X,Y) = 0.816$ , linear regression line  $Y = 3 + 0.5*X$

# Anscombe's quartet

---



# OBJECTIVE ASPECTS OF DESIGN

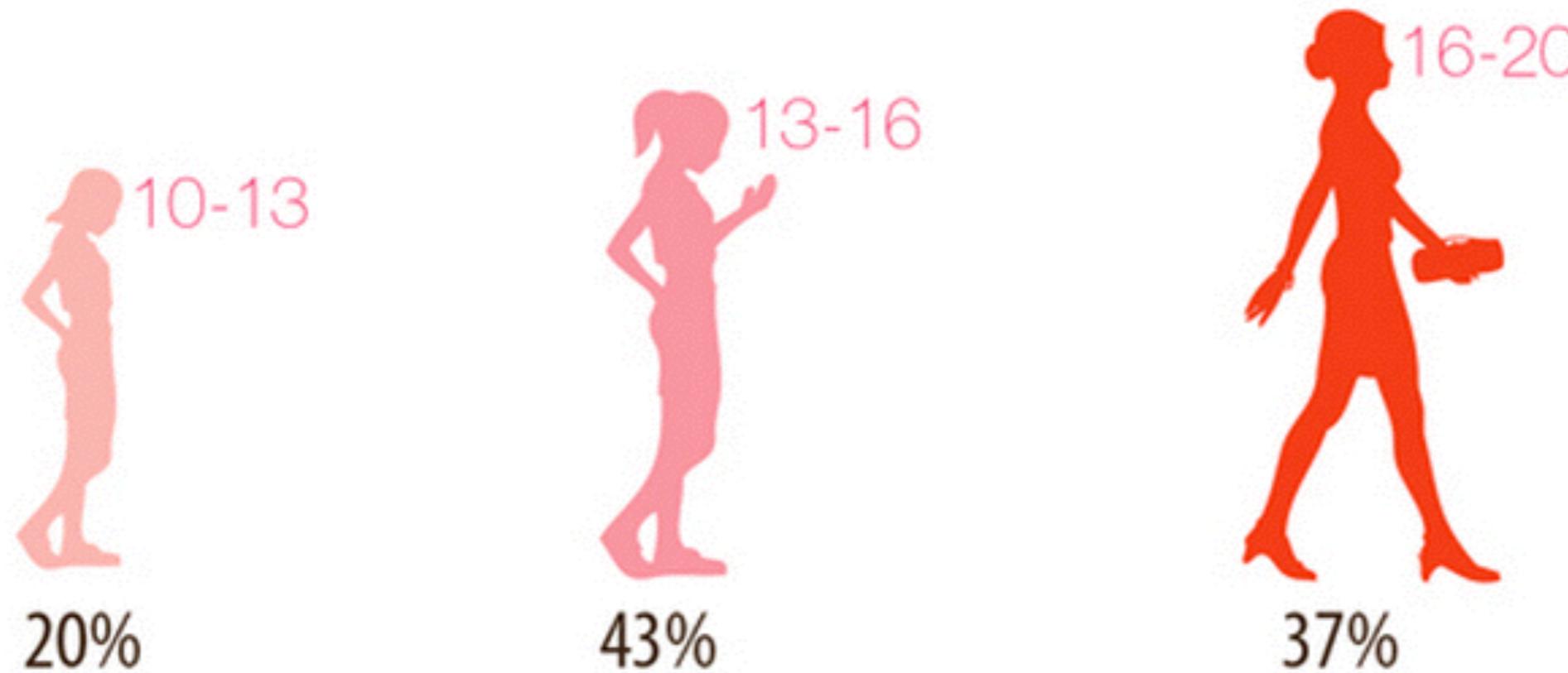
It's not all about taste.

# Is design subjective?



Attractive aspects of proportion, symmetry, color and texture have strong objective foundations.  
Few of you would choose the creature on the right as the more attractive.

## At what age did you start wearing makeup?

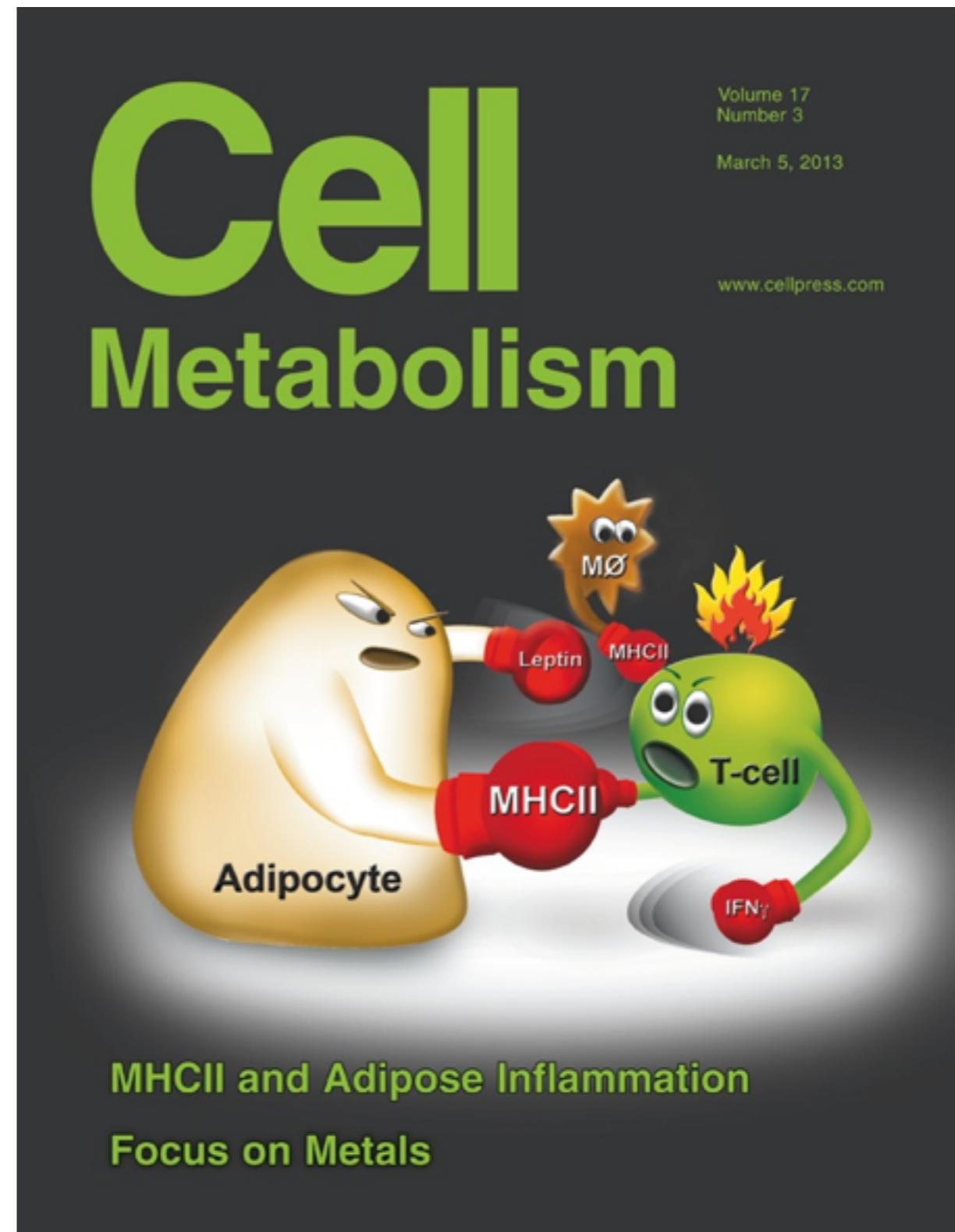
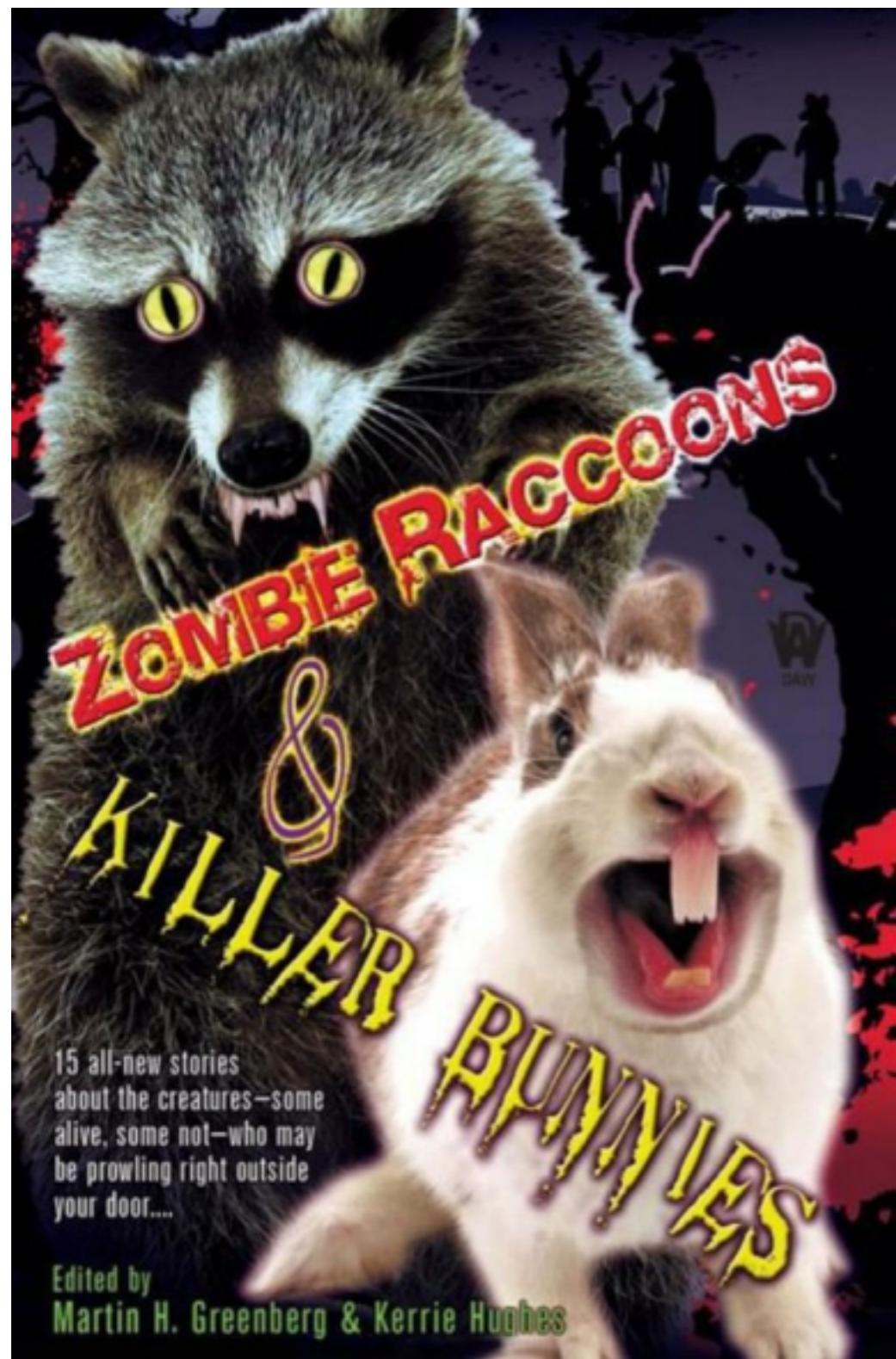


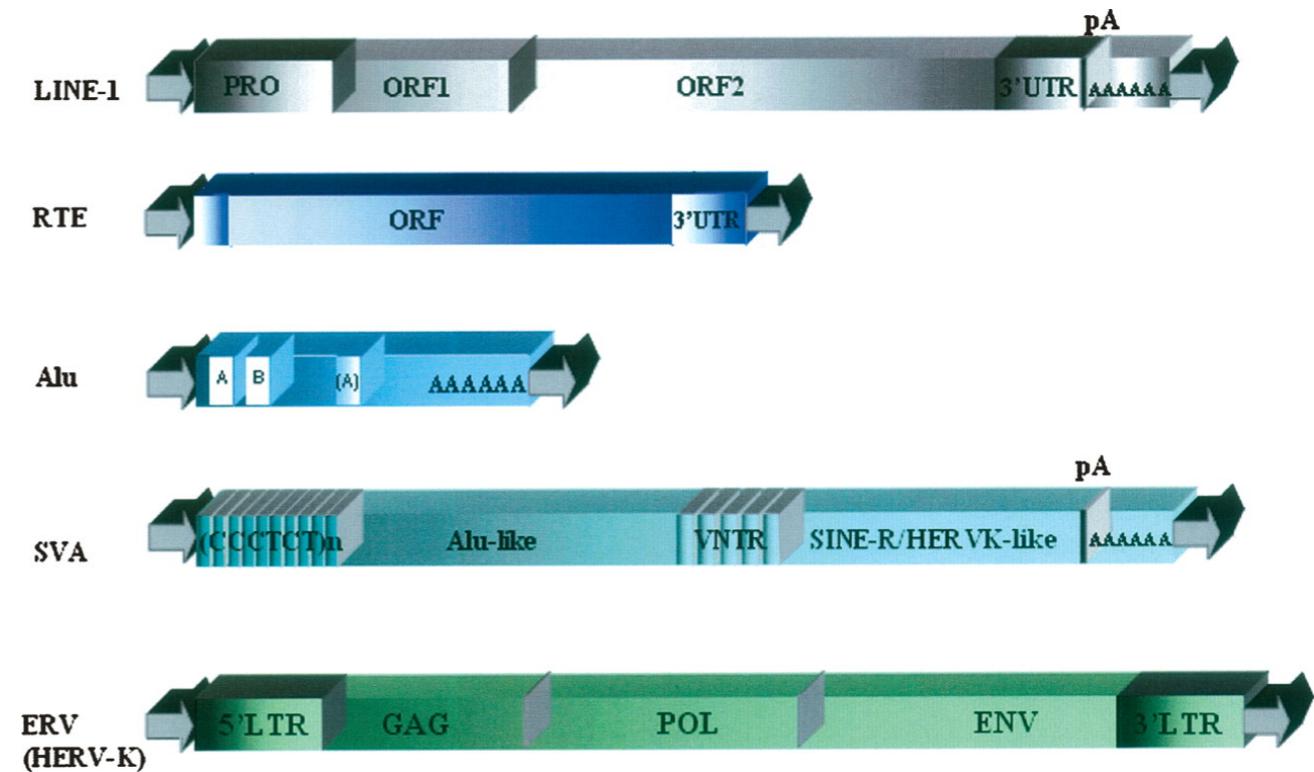
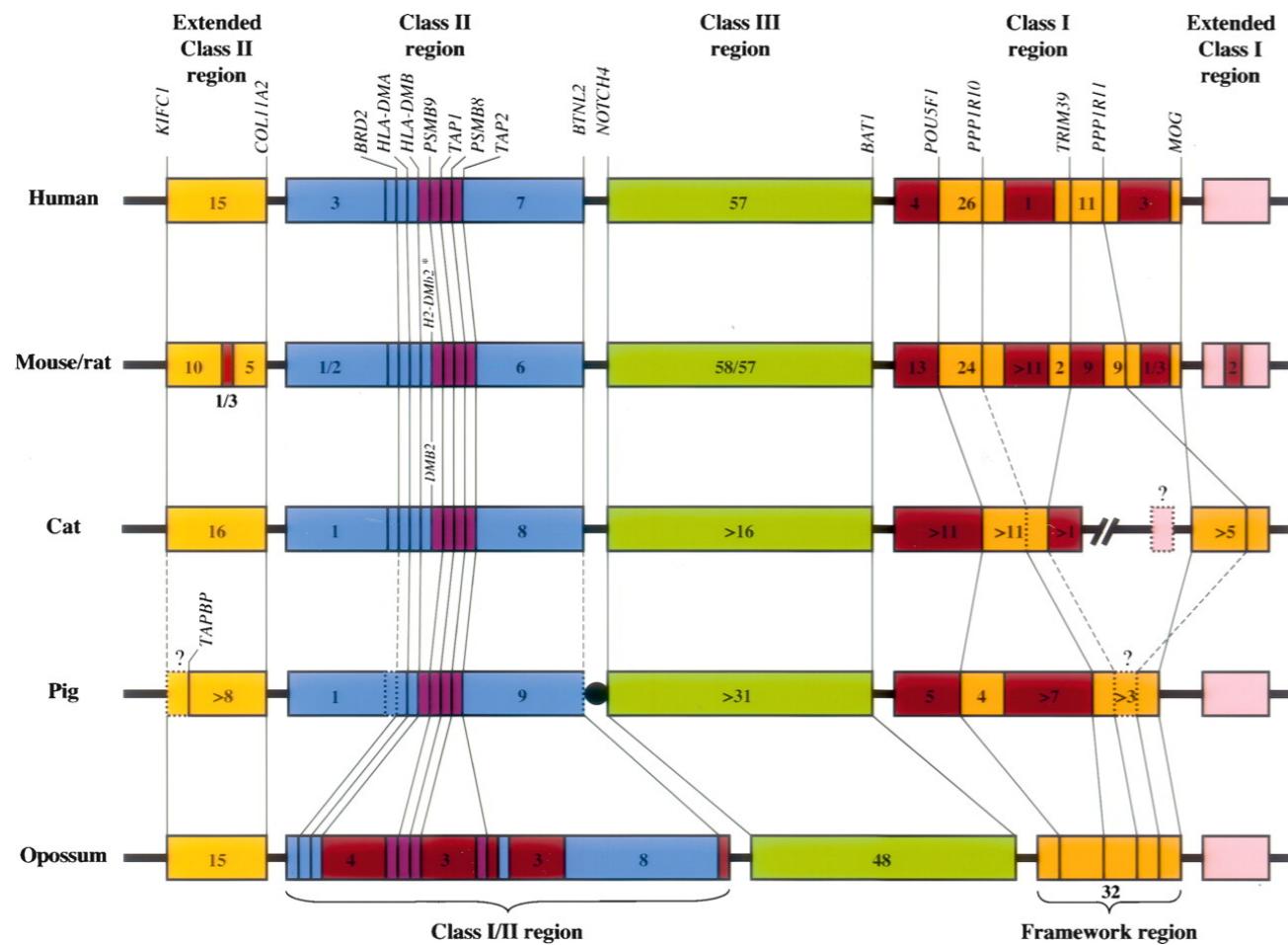
This is what happens when data wears makeup.

#WTFViz #Percentages #BarChart #Iconography

35 notes







Excellent organization and consistency. Vertical lines cue continuity. Good use of color.

Samollow, P.B., The opossum genome: insights and opportunities from an alternative mammal. *Genome Res*, 2008. 18(8): p. 1199-215.

Chartjunk plentiful. Screaming ornamental and redundant elements. Text inconsistent and illegible.

Gentles, A.J., et al., Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res*, 2007. 17(7): p. 992-1004.

# Definition of visualization

---

Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively.

Tamara Munzner

# Definition of visualization

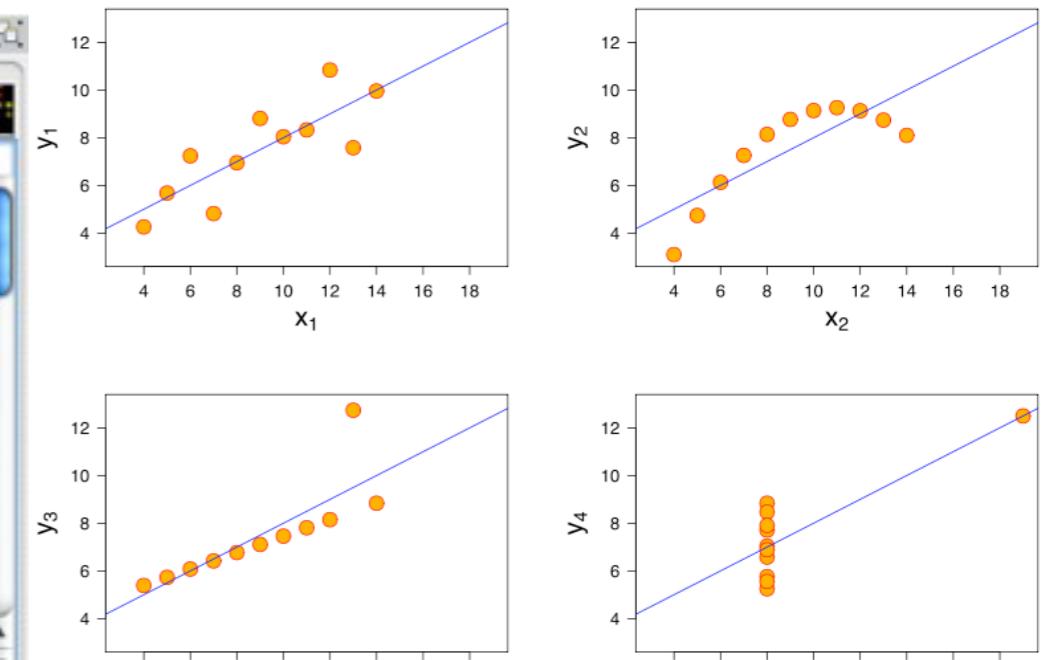
Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively.

Tamara Munzner

- human in the loop needs the details
- external representation: perception vs cognition
- intended task
- measureable definitions of effectiveness

Data Panel

ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	-1.553	
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	-1.019	0.745
CHUK	Kinase	-1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	-5.961		2.139	-1.236	0.601
RELA	Transcription factor	-1.077	0.564	-1.169	1.943	0.594
IKBKB	Kinase	<b>1.167</b>	<b>0.29</b>	<b>1.421</b>	<b>-1.907</b>	<b>0.286</b>
CCL4	Chemokine	1.254	0.878	-1.052	1.499	0.761
MAP3K7		1.01	0.956	-1.096	1.222	0.8
ICAM1	Adhesion	1.184	0.669	1.537	1.392	0.671
IRF1	Transcription factor	-1.013	0.519	1.416	1.081	0.995
CXCL3	Chemokine	1.7	0.905	1.092	-1.598	0.521
IL12B	Cytokine	-2.448	0.042	-1.473	-2.109	0.08
CCL11	Chemokine	-1.338	0.349	-1.995	-1.785	0.129
MAP3K7IP1	Adaptor					



**SCIENCE IS A PROCESS**

**DESIGN IS A PROCESS**

In either, we don't always know the end product.

But we must understand how we might get there.

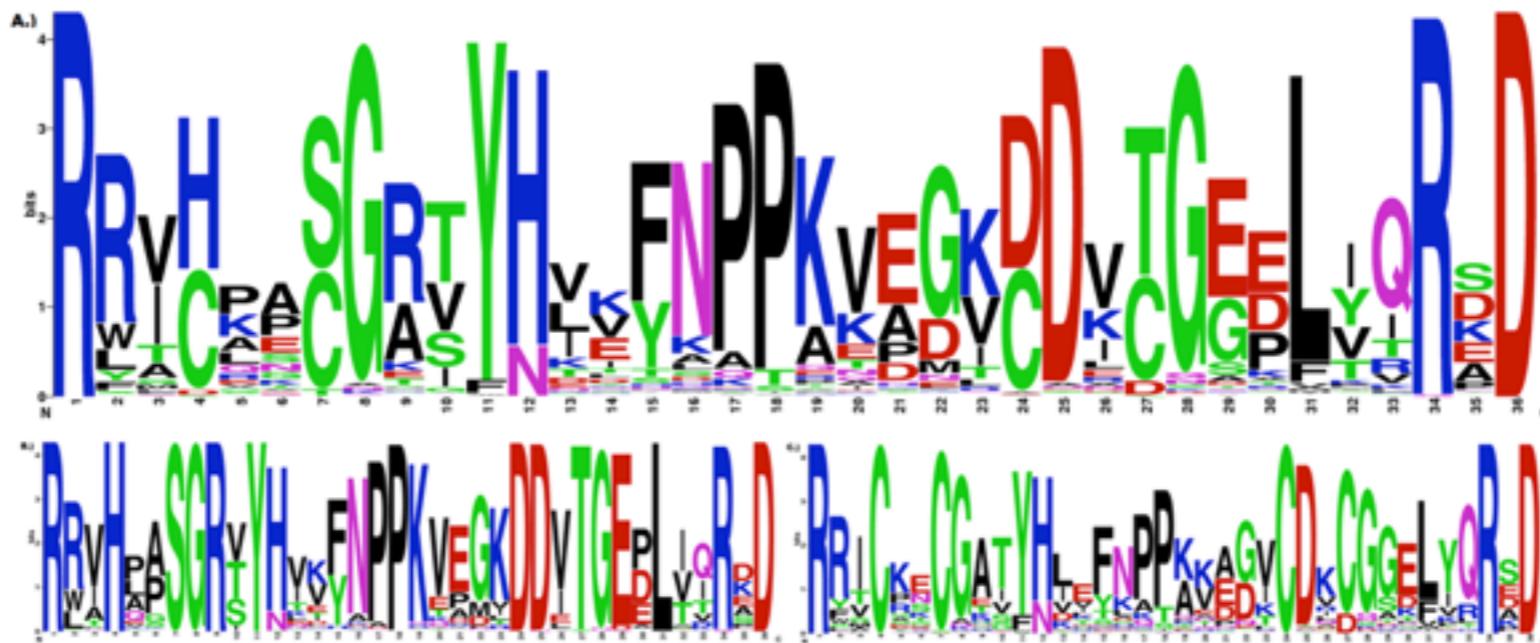
you have data

you may have a message

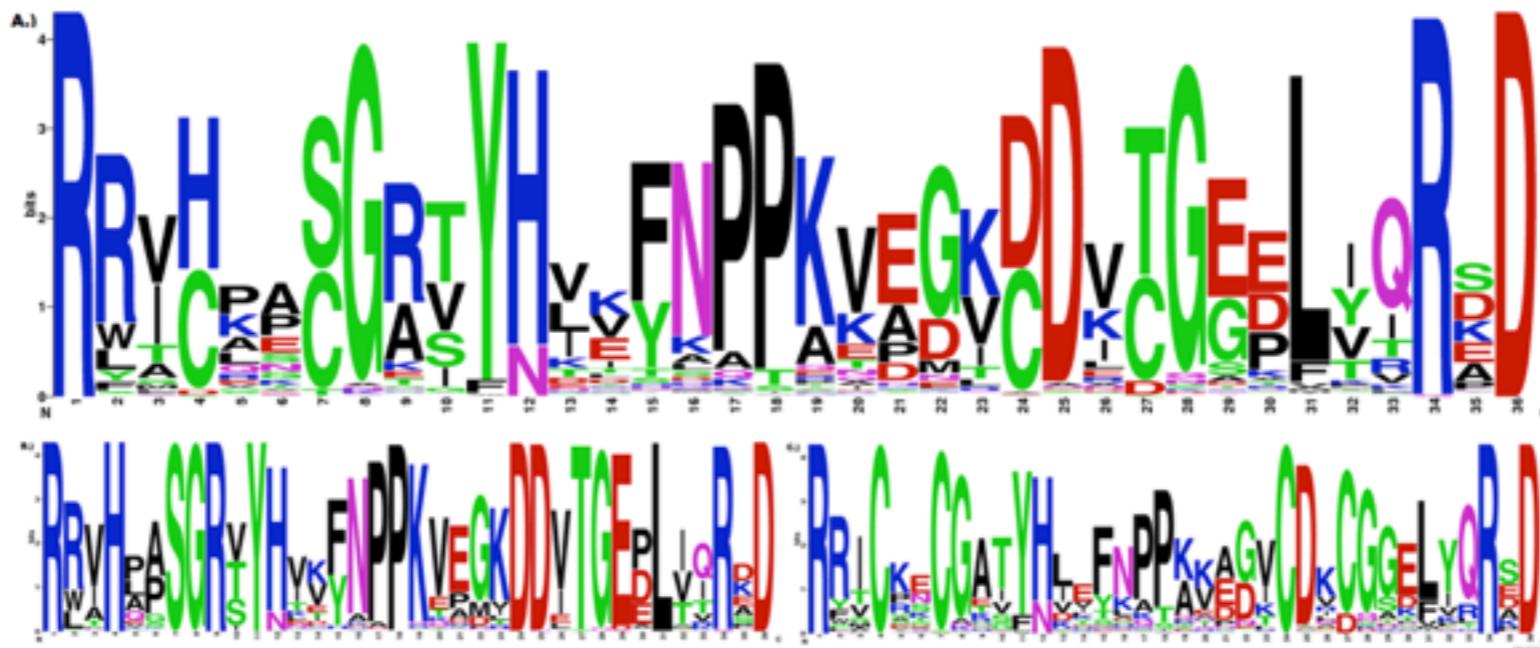
you need to create a figure  
(or think you do)

you select your encoding

e.g. sequence logo



you generate the image with an application  
that supports the encoding



you write a legend, making up  
for things that are not obvious

Sequence logos showing the amino acid usage in the adenylate kinase lid (AKL) domain. (A) Across all organisms. (B) from Gram-negative bacteria. (C) from Gram-positive bacteria.

The ADK lid domain structure is universally conserved, but is stabilized in the Gram-negatives by a hydrogen bonding network between residues 4, 7, 9, 24, 27, and 29 (and several other residues in some organisms), while the Gram-positives are stabilized by a bound metal ion, tetrahedrally coordinated by the Cysteines at 4, 7, 24 and 27. **The identities of several other positions (eg 5, 8, 30, 32) are differentially constrained** in each subfamily as well, apparently due to steric requirements of the stabilizing residues.

your figure is done  
how do you know  
whether the creation process  
was successful?

## **CREATE VISUALS WHEN NECESSARY**

The desire for a figure is not always proportional to its utility.

**NOTICE**

**PLEASE DO NOT  
THROW STONES  
AT THIS SIGN**

**thank you**



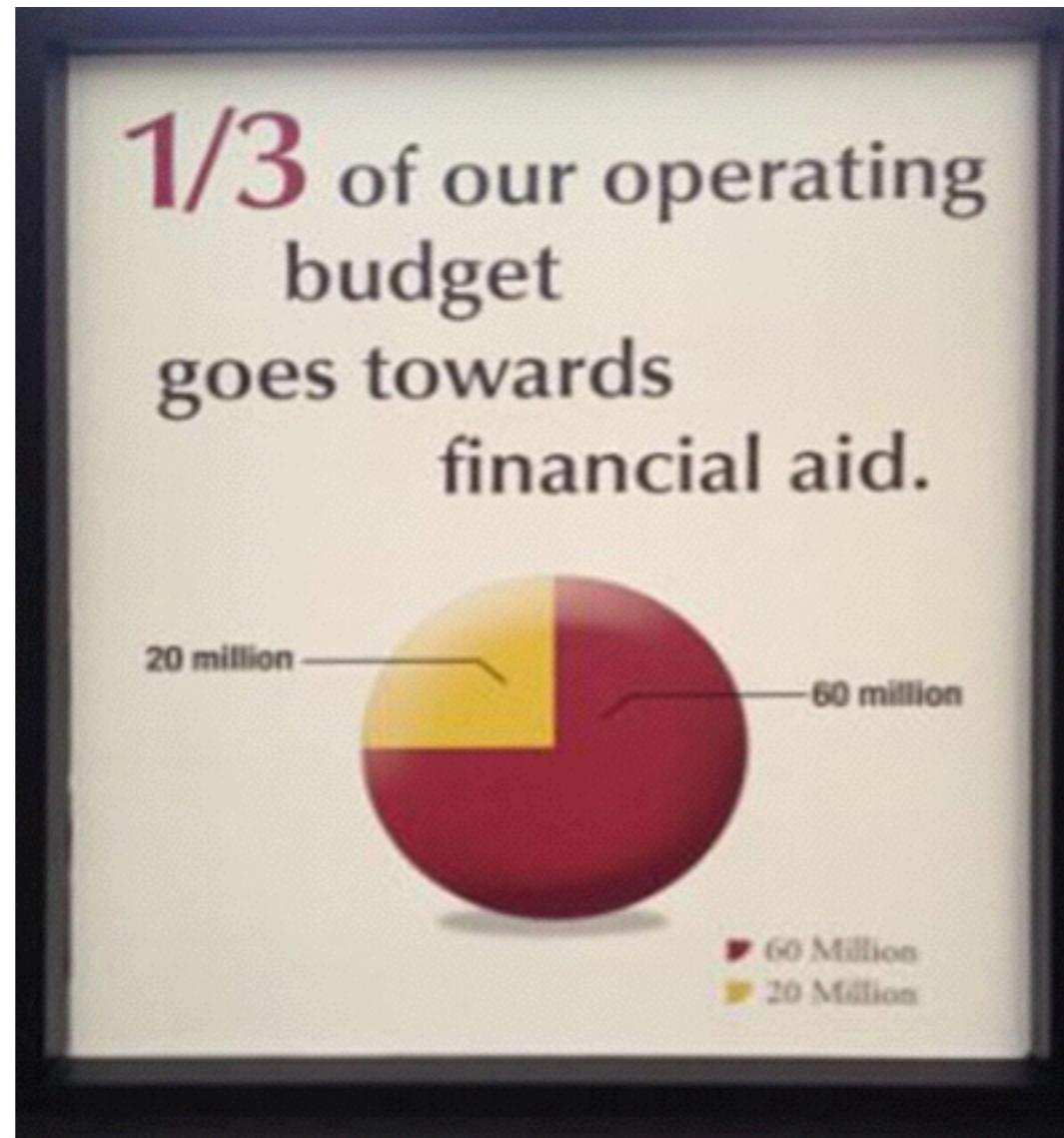


NO



accuracy does not always add to utility

# IS ABSOLUTE ACCURACY ALWAYS IMPORTANT?



# VARIOUS DATA TYPES

What encoding is the best for the data

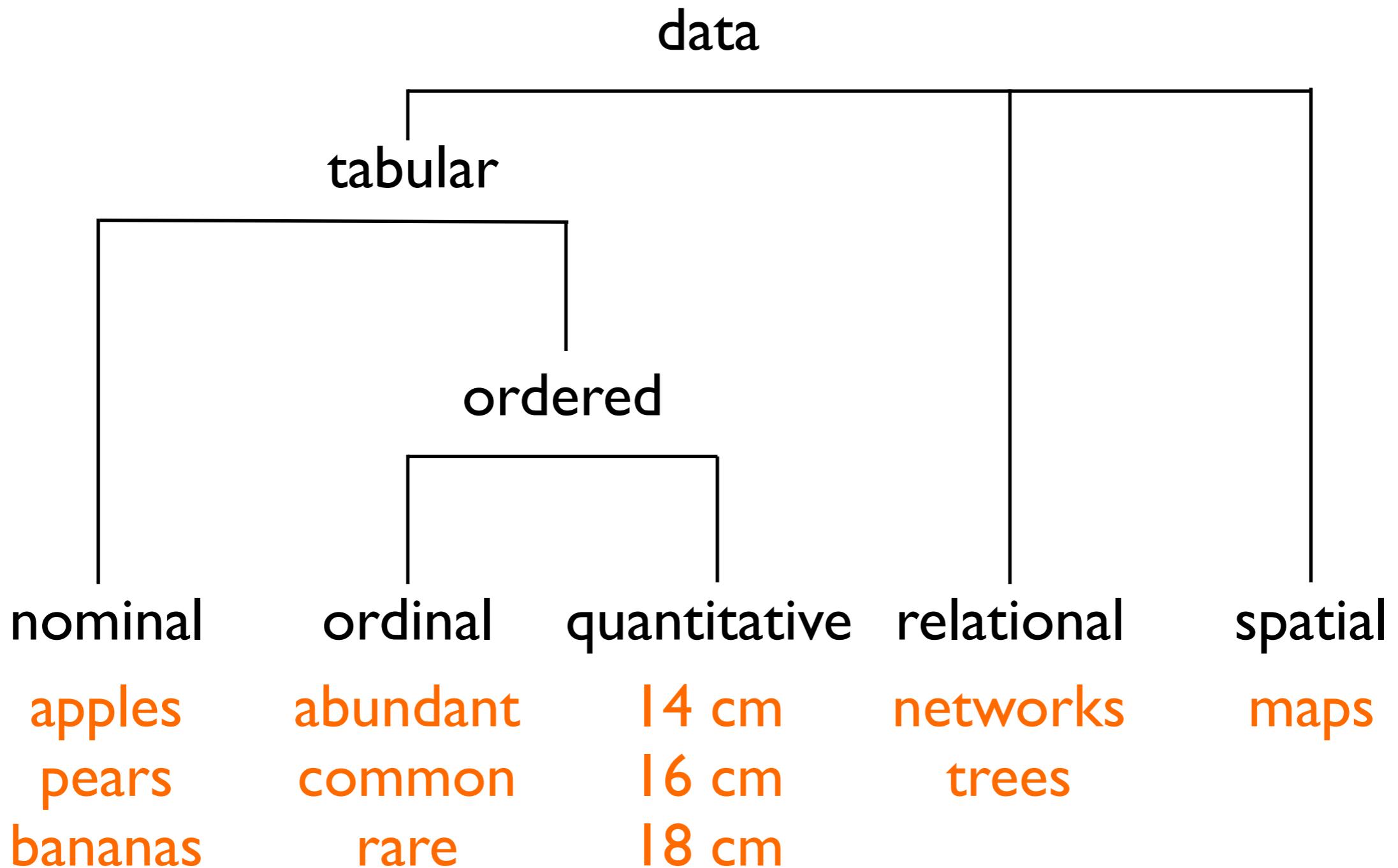
# Data types

---

apples	abundant	14 cm	networks	maps
pears	common	16 cm	trees	
bananas	rare	18 cm		

# Data types

---



# Encoding schemes

---



# Accuracy of Quantitative Perceptual Tasks

More accurate



position



length



angle



slope



area



volume



density



colour



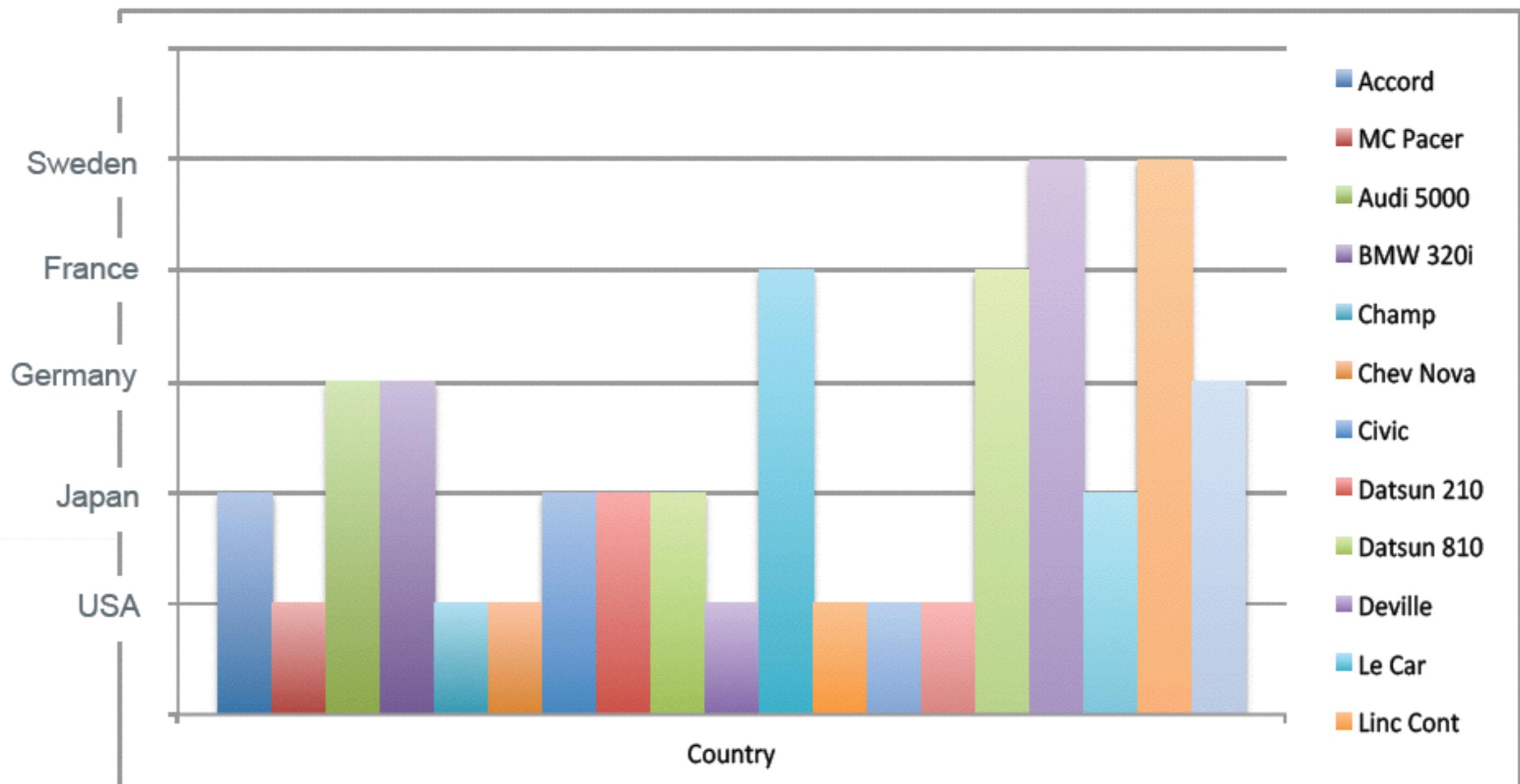
Less accurate

# Mapping data types to encoding

Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Density	Hue
Angle	Saturation	Texture
Slope	Hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density	Containment	Saturation
Saturation	Length	Shape
Hue	Angle	Length
Texture	Slope	Angle
Connection	Area	Slope
Containment	Volume	Area
Shape	Shape	Volume

# Expressiveness

- Encodes only the facts



# Language of Graphics

---

- Graphics can be thought of as forming a sign system:
  - Each mark (point, line, or area) represents a data element.
  - Choose visual variables to encode relationships between data elements
    - difference, similarity, order, proportion
    - only position supports all relationships
- Huge range of alternatives for data with many attributes
  - find images that express and effectively convey the information.

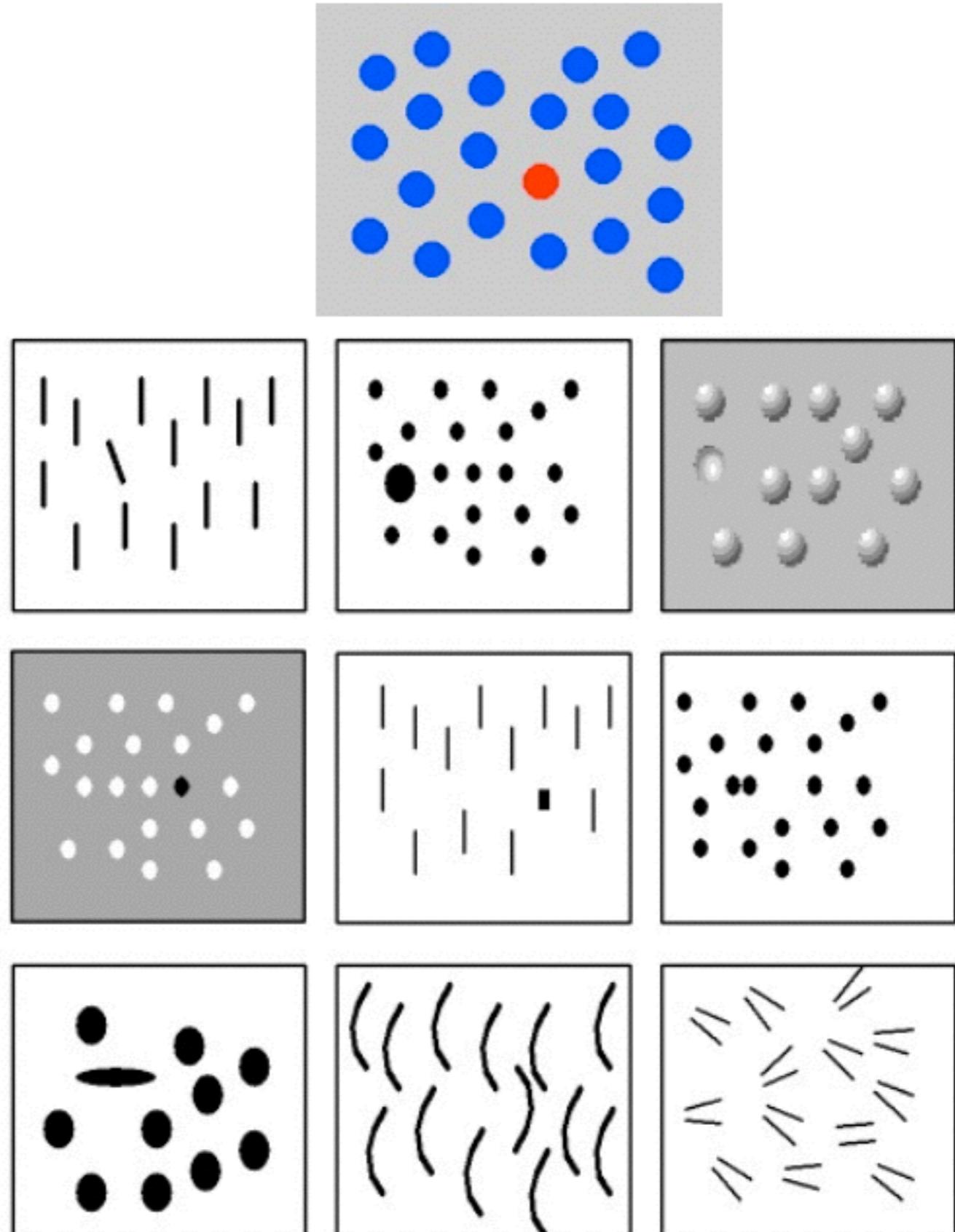
# HUMAN PERCEPTION

What features the human brain captures?

# Preattentive Visual Features

---

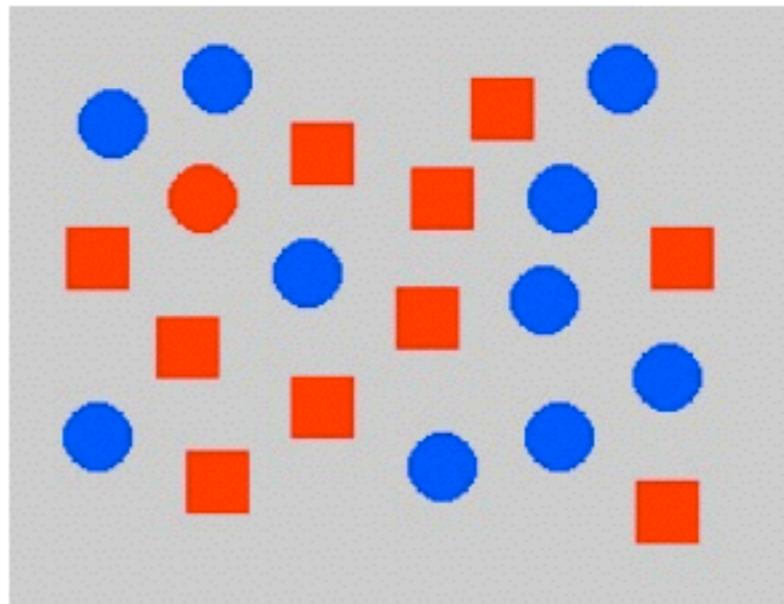
- the ability of the low-level human visual system to rapidly identify certain basic visual properties
- a sufficiently different item noticed immediately independent of distractor count



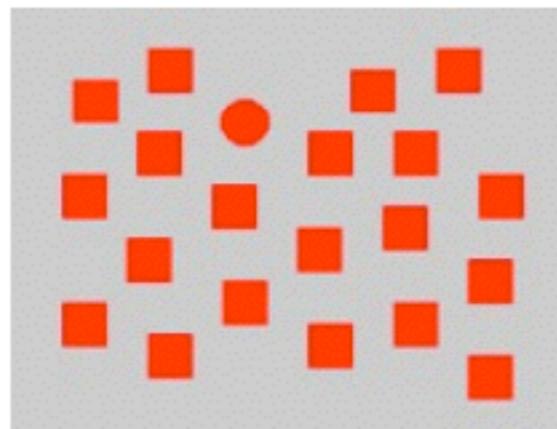
# Preattentive limits

---

- only one channel at a time



- within channel, speed depends on which channel and how different item is from the surrounding



# Use of preattentive features

---

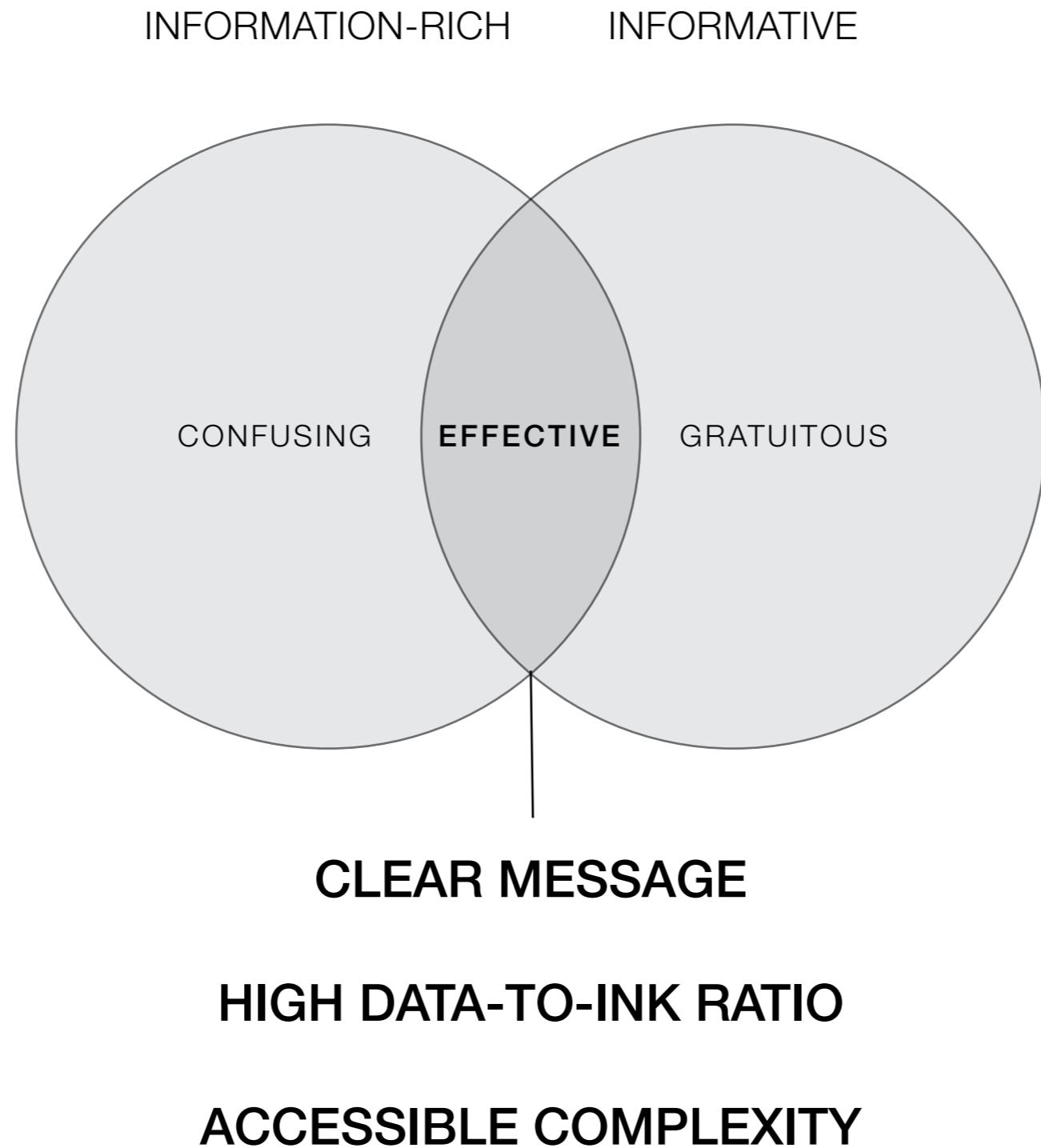
- target detection
- boundary detection
- region tracking
- counting and estimation

## **INFORMATIVE + INFORMATION-RICH**

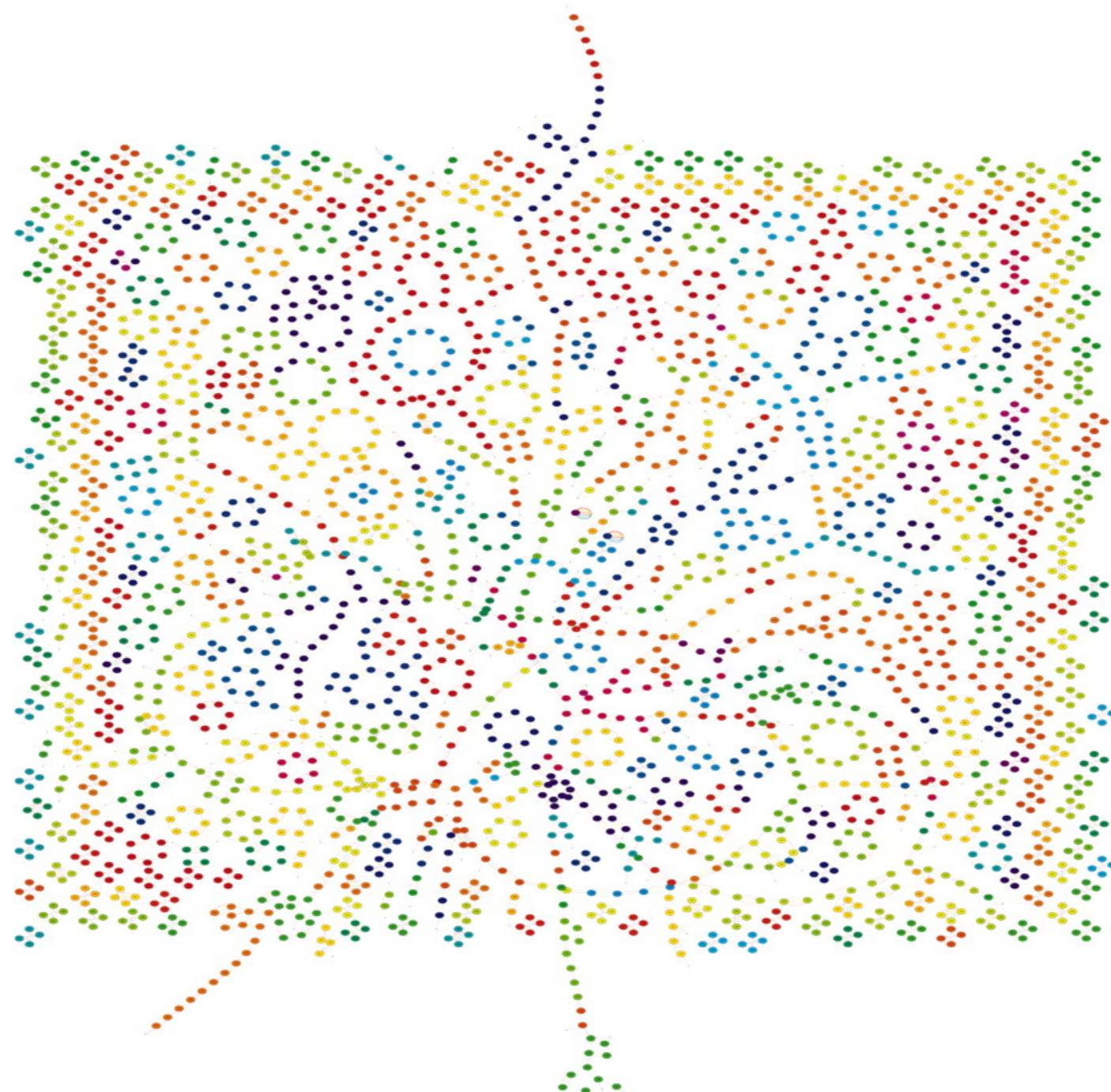
In the age of big data, figures should be worth more than 1,000 words.

strive to give your viewer  
the greatest number of useful ideas  
in the shortest time  
with the least ink  
in the smallest space

# VISUALIZATION SWEET SPOT



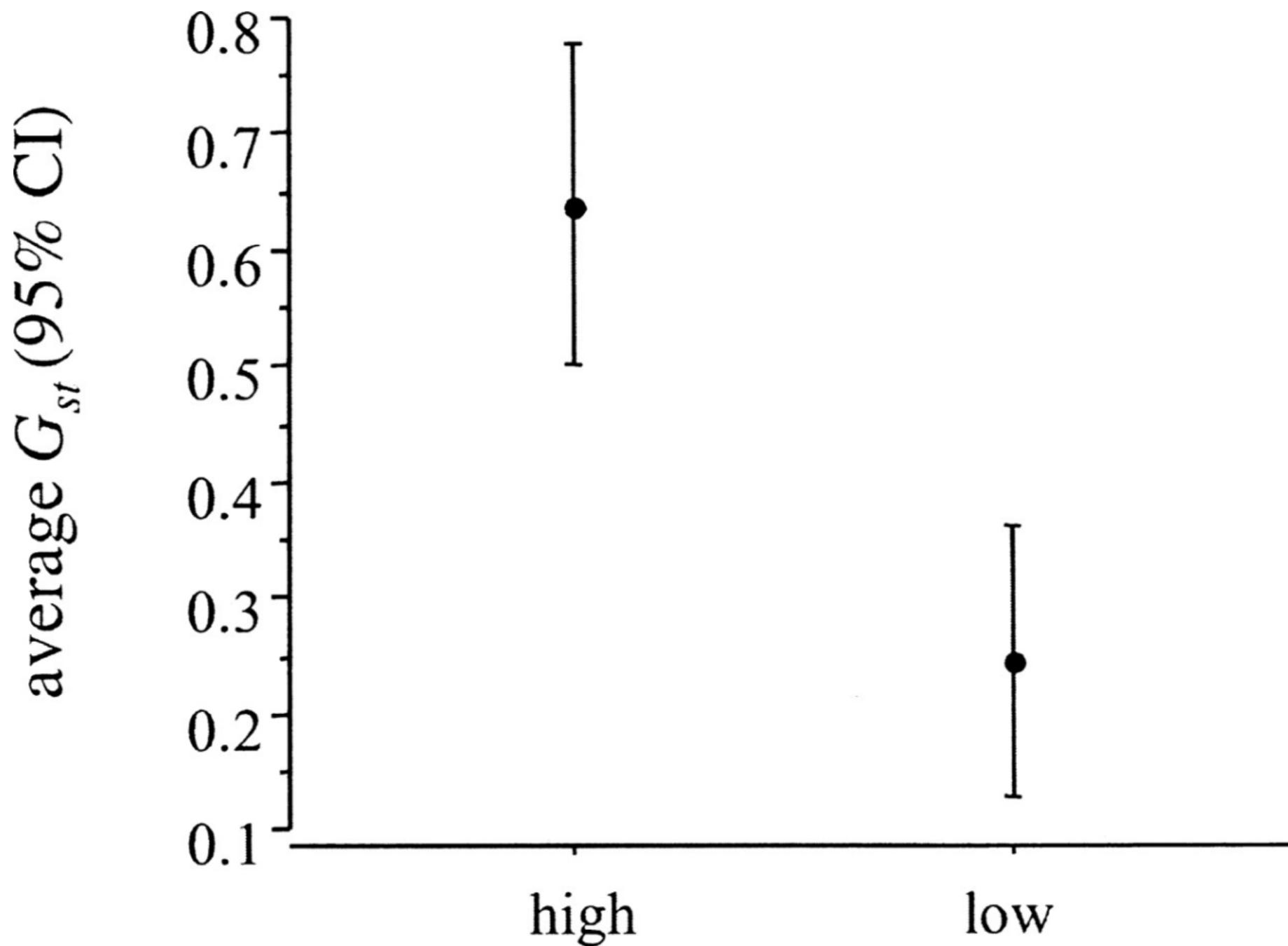
# INFORMATION-RICH, NOT INFORMATIVE



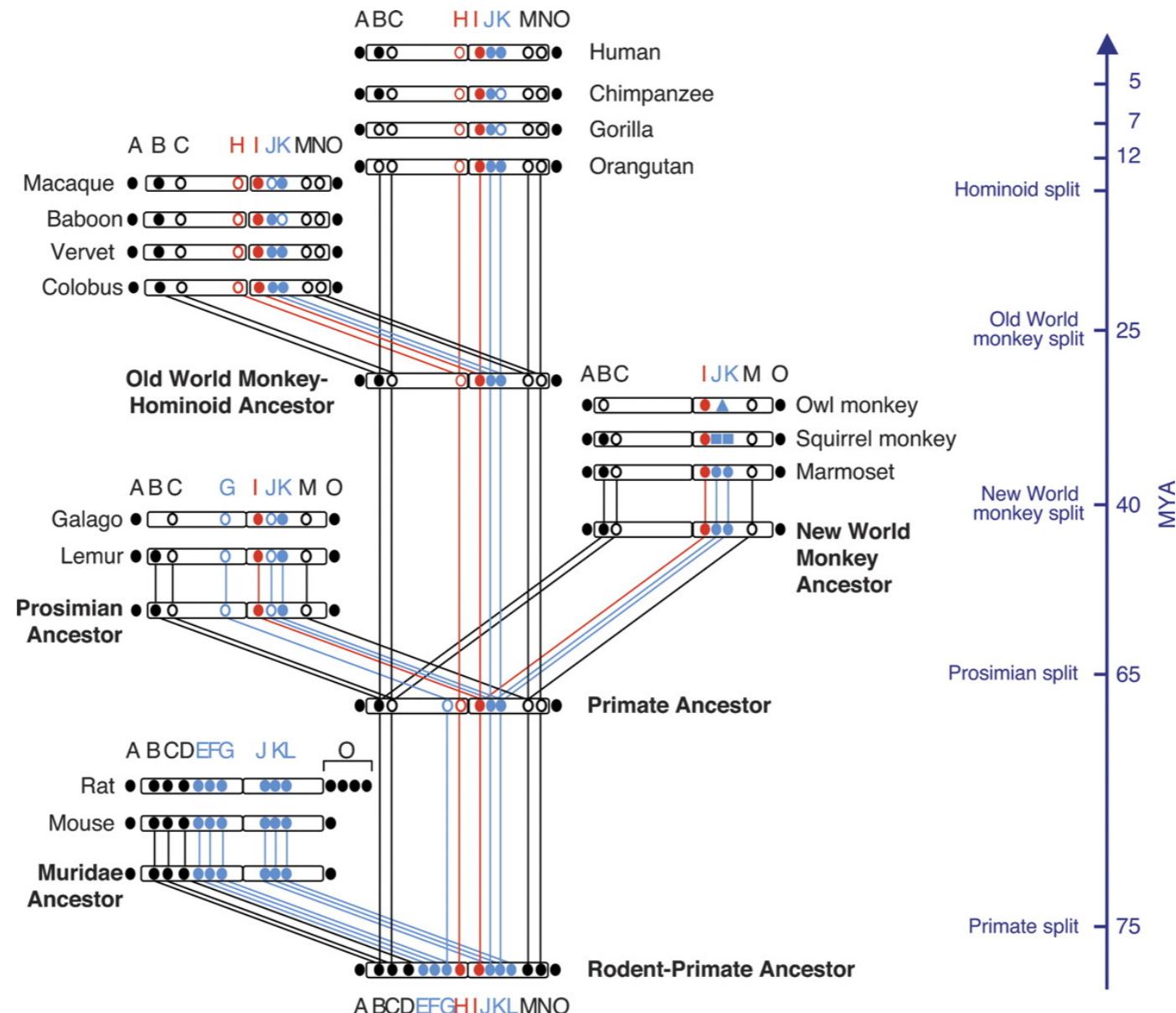
Chromosome colors:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	---

# INFORMATIVE, NOT INFORMATION-RICH

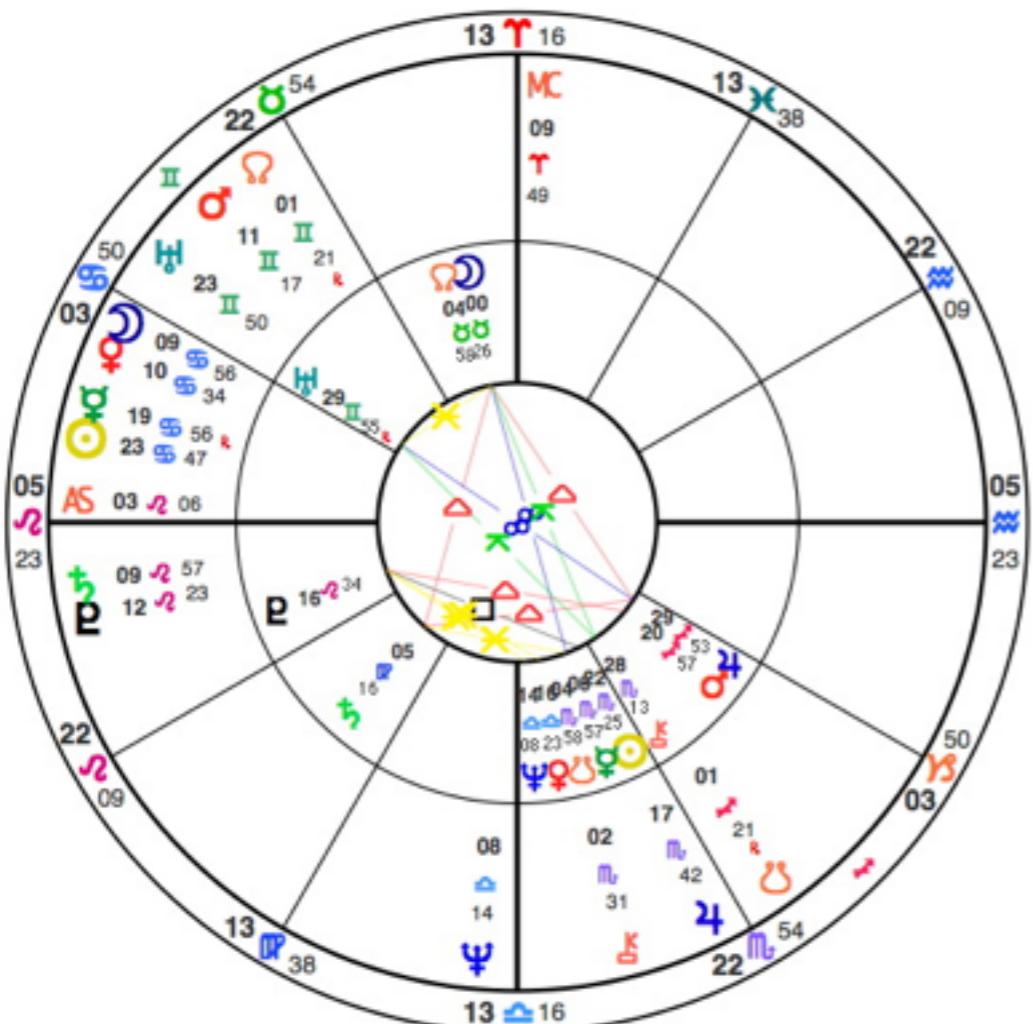


# INFORMATION-RICH AND INFORMATIVE



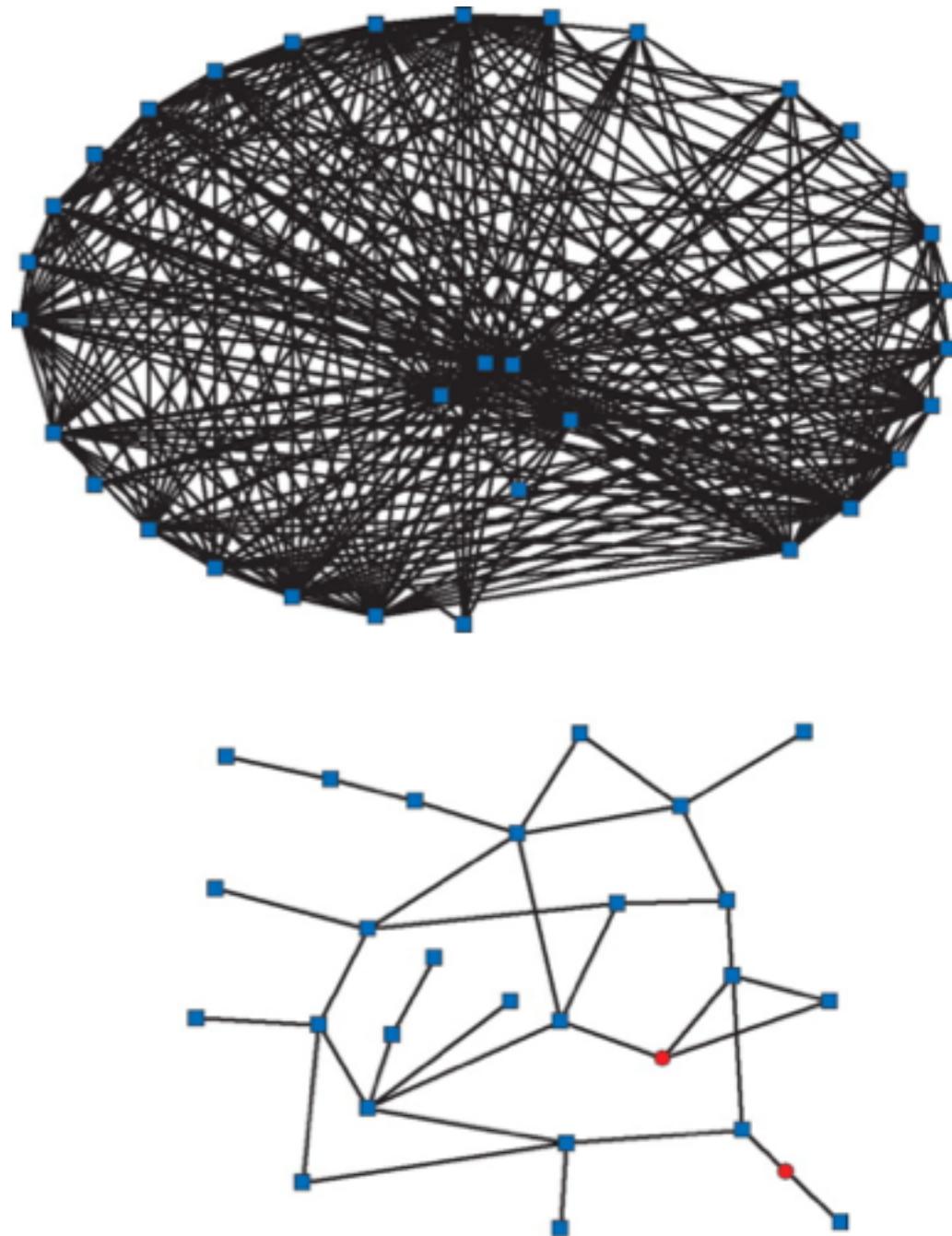
# JUNK SCIENCE

**Prince Charles**  
Reference Chart  
Natal Chart  
November 14, 1948  
9:14:00 PM GMT  
London, England



**Camilla Parker Bowles**  
2nd Chart  
Natal Chart  
July 17, 1947  
7:00:00 AM BDST  
London, England

# REAL SCIENCE



(left) Synastry chart. <http://sasastrology.com/2011/03/the-astrology-of-marriage-in-the-royal-family-a-suitable-girl-and-the-bit-on-the-side.html>  
(right) Shakhnovich, B.E. and E.V. Koonin, Origins and impact of constraints in evolution of gene families. Genome Res, 2006. 16(12): p. 1529-36.

don't merely display data  
explain it

# KNOW YOUR MESSAGE

Stick to it.

**CAUTION**  
THIS SIGN HAS  
**SHARP EDGES**

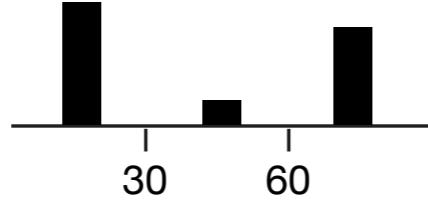
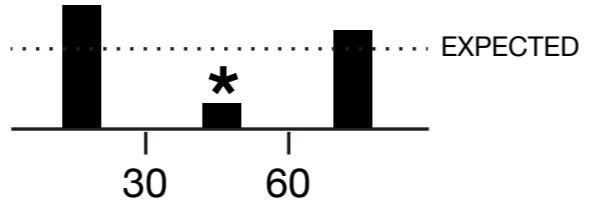
DO NOT TOUCH THE EDGES OF THIS SIGN



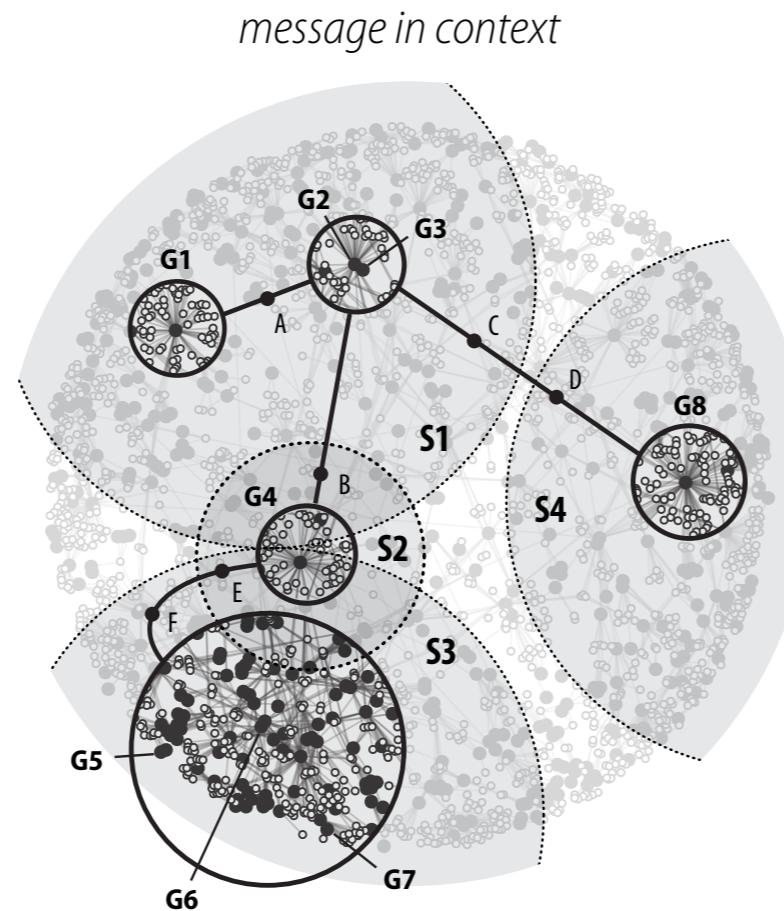
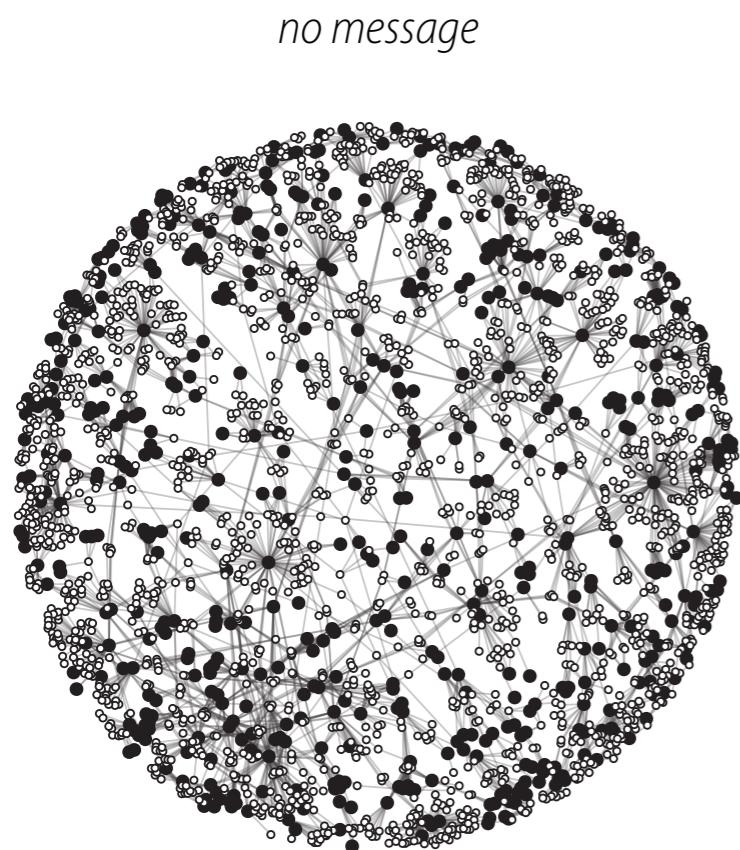
ALSO, THE BRIDGE IS OUT AHEAD



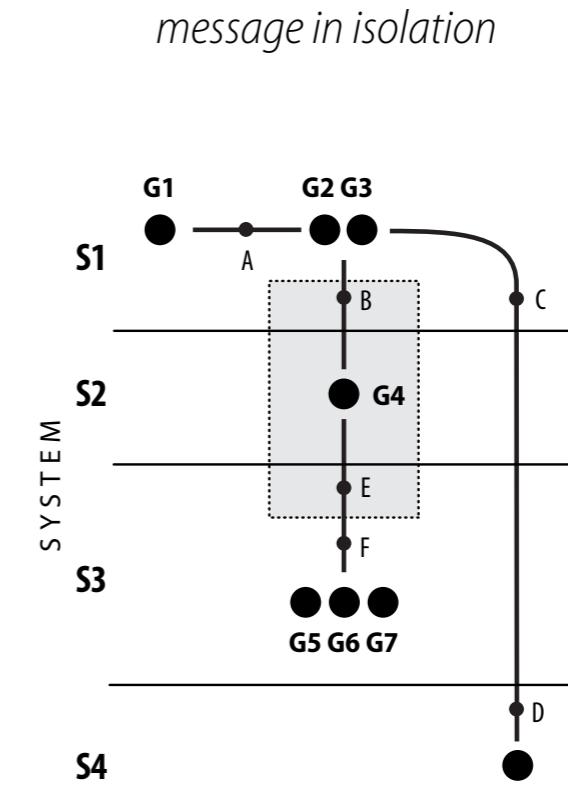
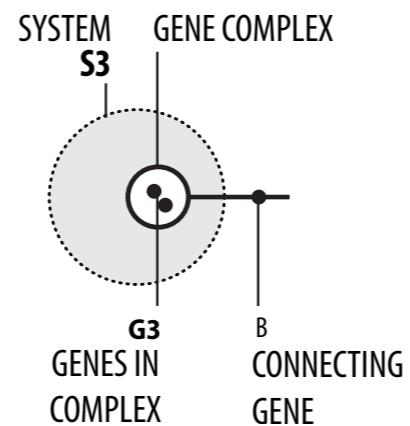
# ACHIEVE FOCUS BY AGGREGATING

WHAT IS SHOWN?	WHAT IS COMMUNICATED?	WHAT IS INTERPRETED?
RAW DATA 12 54 82 29 25 22 67 61 23 79	NO CLEAR MESSAGE.	UNKNOWN. READER IS ON THEIR OWN.
DISCRETIZED  <ul style="list-style-type: none"><li>● 0-30</li><li>● 31-60</li><li>● 61-100</li></ul>	SCALE	THREE RANGES ARE IMPORTANT. INDIVIDUAL VALUES WITHIN A RANGE ARE NOT.
BINNED 	DISTRIBUTION	THERE ARE FEWER MEDIUM-SIZED VALUES.
TREND 	SIGNIFICANCE	THERE ARE <u>SIGNIFICANTLY</u> FEWER MEDIUM-SIZED VALUES.

# CONTEXT MUST NEVER DILUTE MESSAGE



● GENE      ○ DISEASE

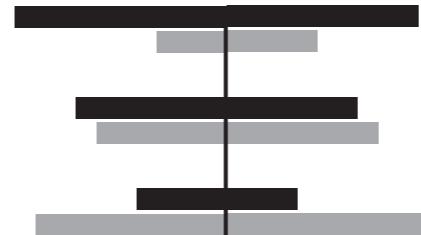
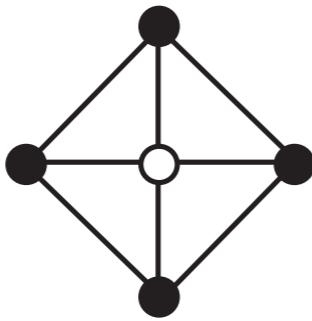


● GENE COMPLEX

● CONNECTING GENE

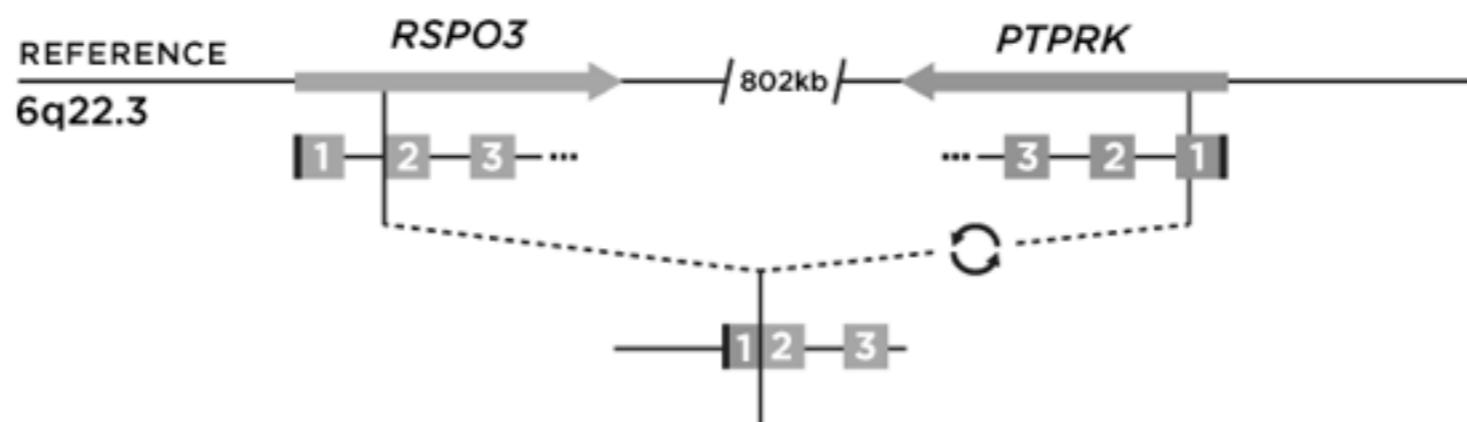
# STRIVE FOR CLEAR COMMUNICATION

Revise and redraw.



to explore data, choose effective encoding

to communicate concepts, use effective design



# HOW TO APPROACH VISUALIZATION

show the data

induce viewer to think about substance rather than methodology

encourage eye to compare different pieces of data

avoid distorting what the data represents

present many numbers in a small space

make large data sets coherent

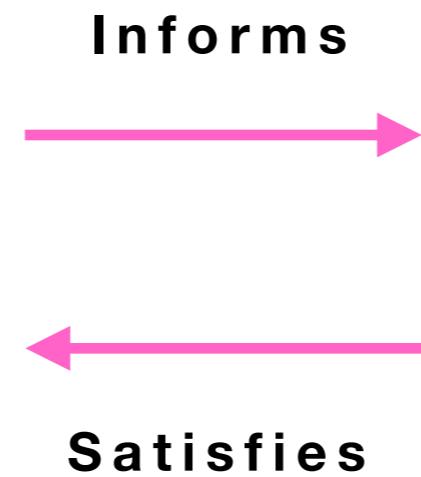
reveal data at several levels of detail – broad overview and fine structure

## TOP-DOWN

redundancy  
consistency  
conciseness  
clarity  
focus & emphasis  
salience & relevance  
truth, accuracy & detail

## BOTTOM-UP

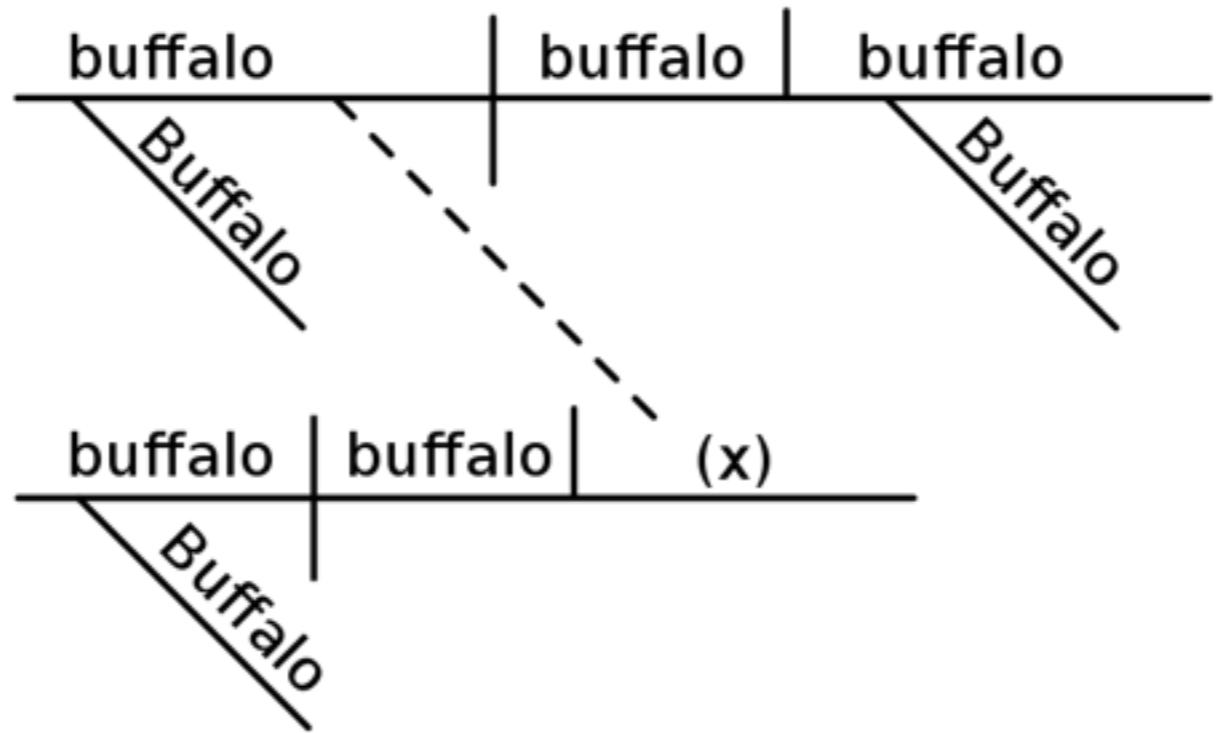
data encoding  
symbols  
color  
typeface  
arrows  
line weight  
alignment



# ACCURACY ISN'T EVERYTHING

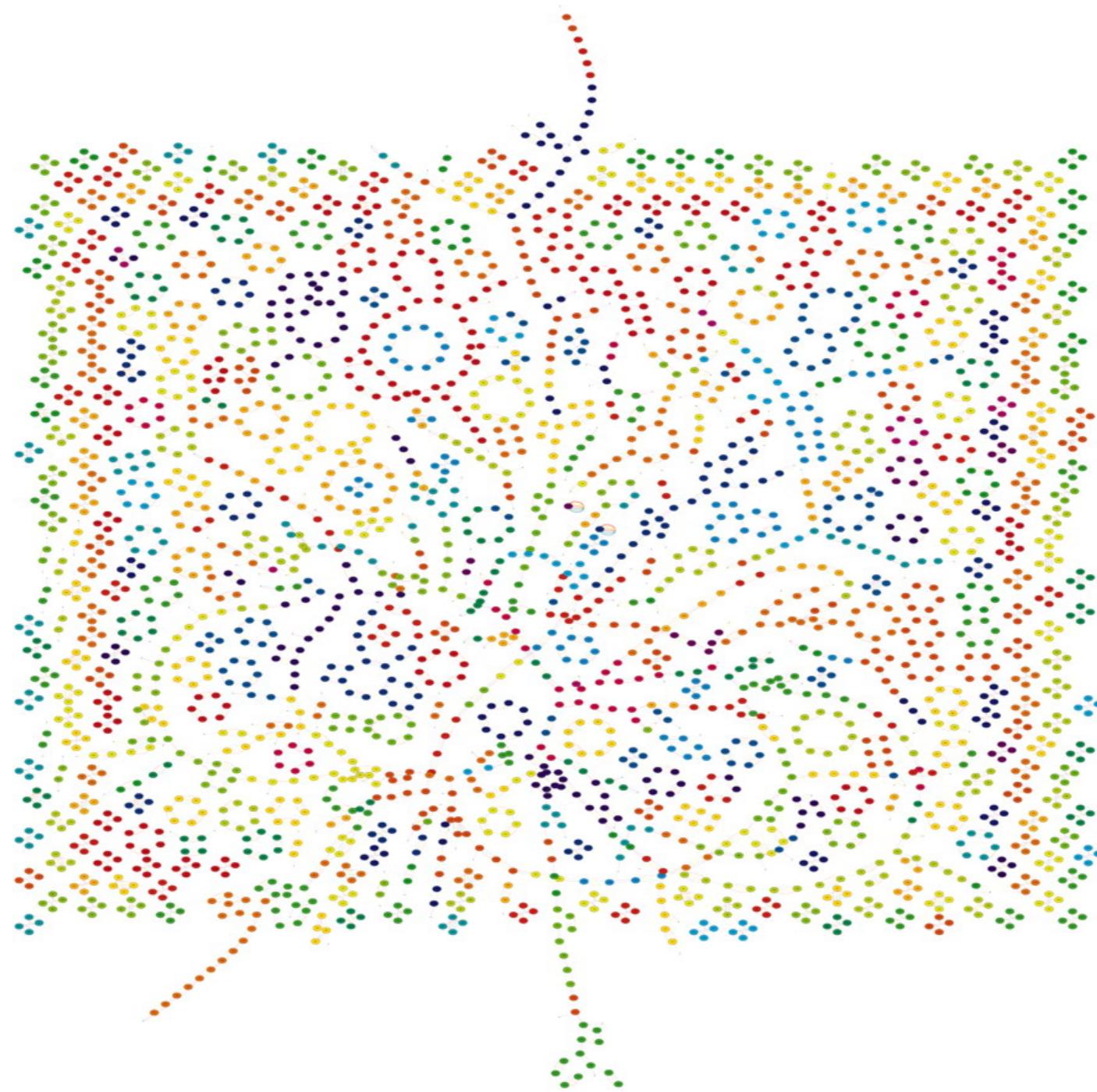
Correctness does not ensure comprehension.

Buffalo buffalo Buffalo buffalo buffalo Buffalo  
buffalo



New York bison whom other New York bison bully,  
themselves bully New York bison.

# BUFFALO BUFFALO OF VISUALIZATION



Chromosome colors:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	---

**SATISFY YOUR AUDIENCE, NOT YOURSELF.**

Be aware of bias in evaluating effectiveness of visual forms.

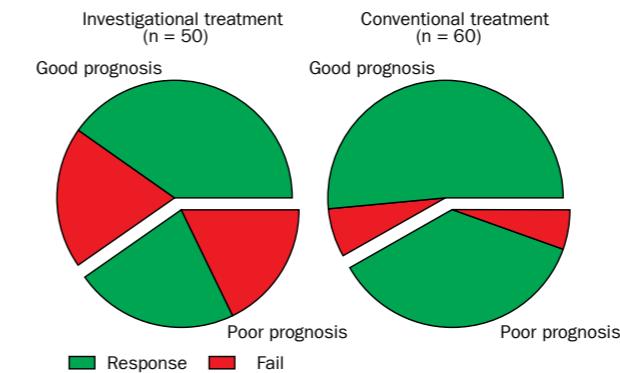
# Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures

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

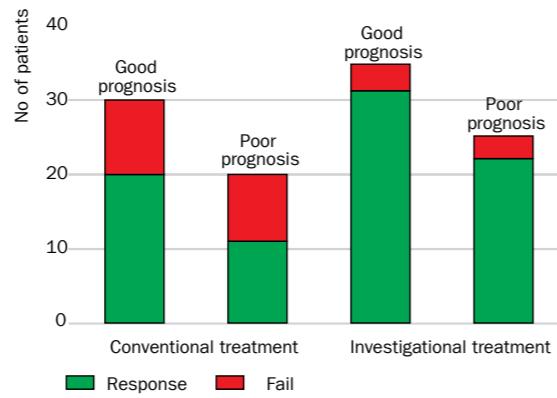
**table**

	Conventional treatment		Investigational treatment	
	Total no	% Fail	Total no	% Fail
Good prognosis	30	30	35	11
Poor prognosis	20	45	25	12
Total	50	38	60	12

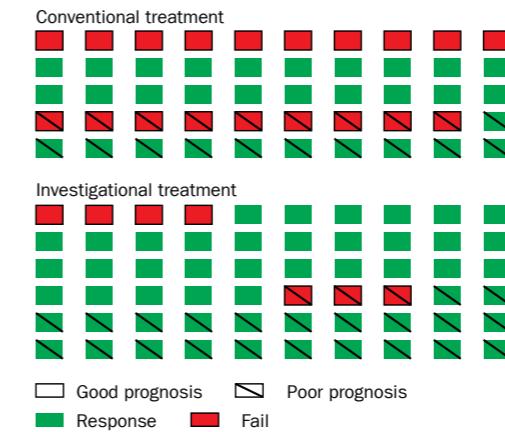
**pie chart**



**bar graph**

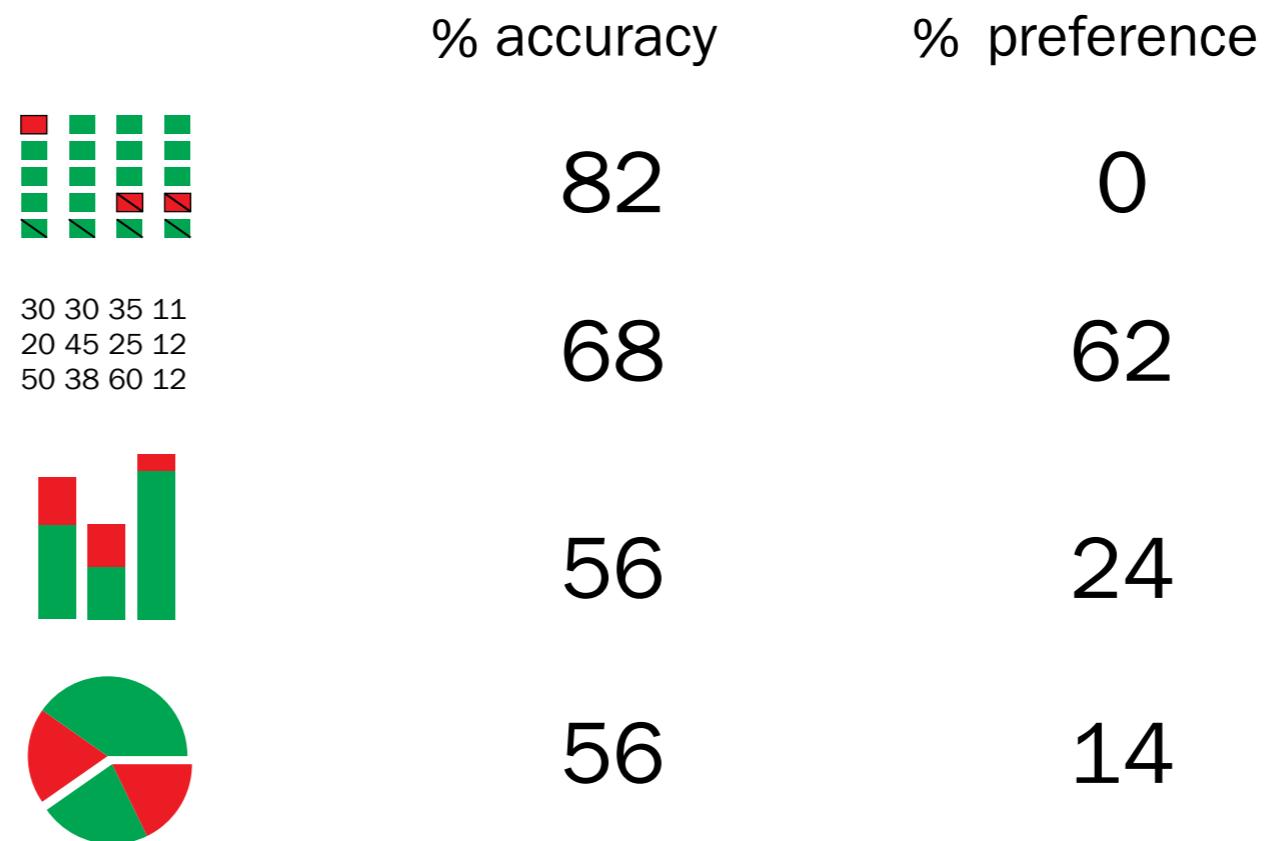


**icon graph**



# 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



# 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

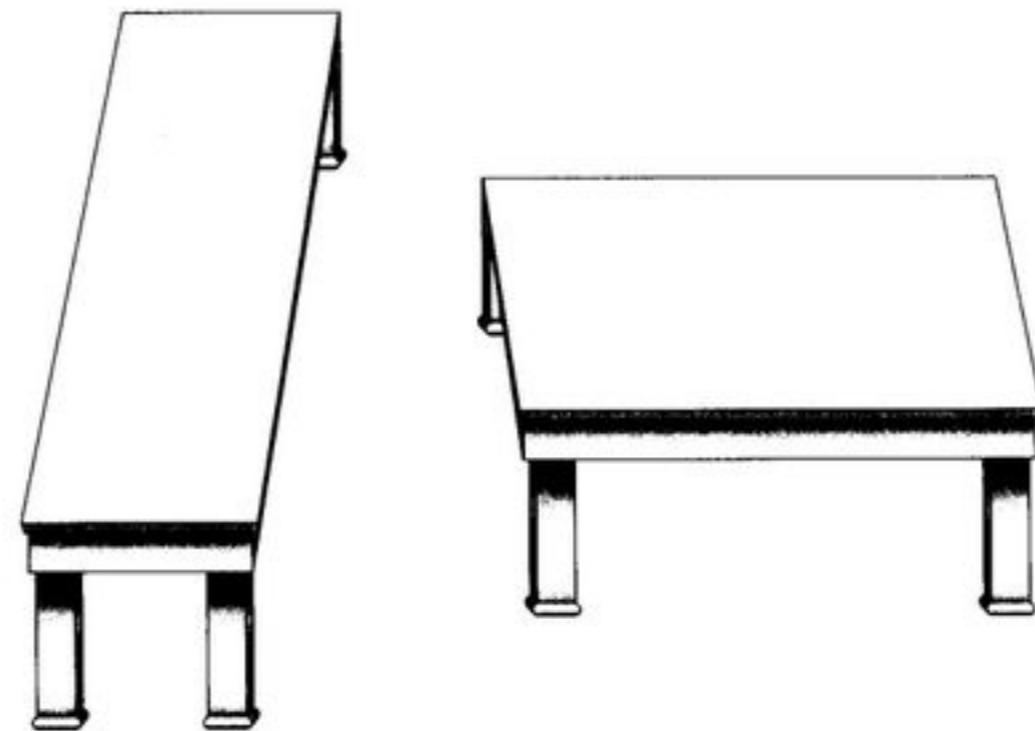
“...eight voiced  
considerable contempt  
for the [icon] display.”

# 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

“... icon displays were often preferred by nurses, students, ... but were considered unacceptable by physicians.”

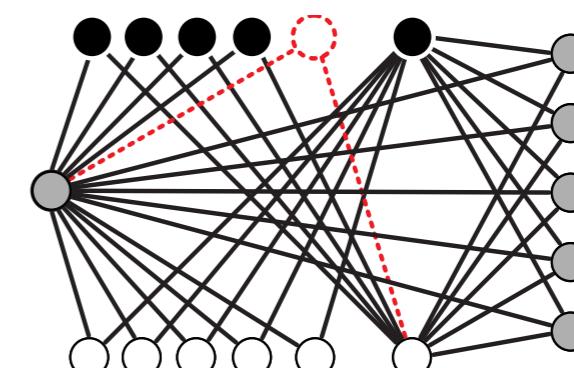
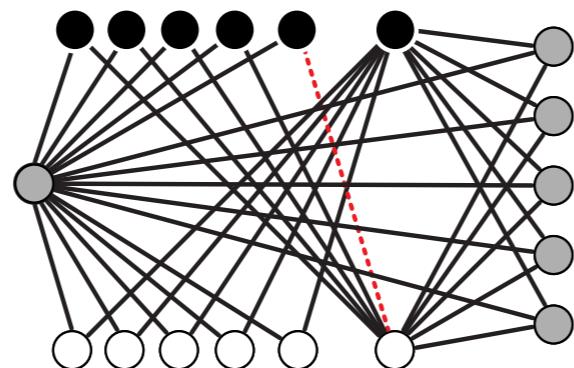
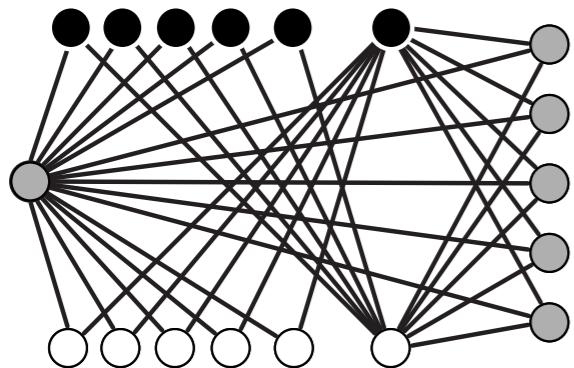
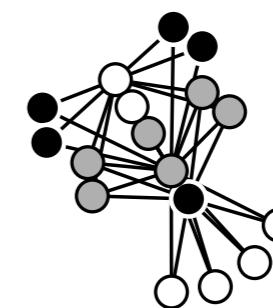
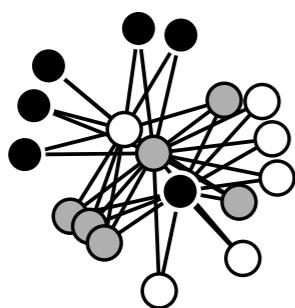
# WE ARE EASILY DECEIVED



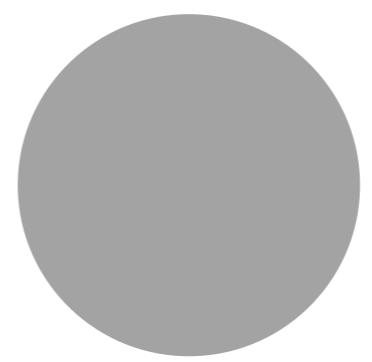
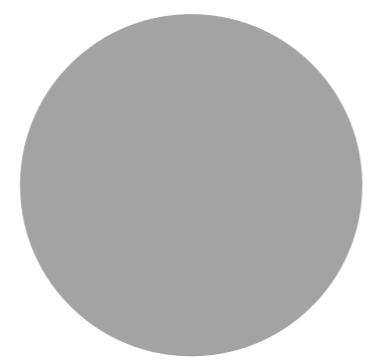
DO NOT BE CHARMED BY INEFFECTIVE FORMS

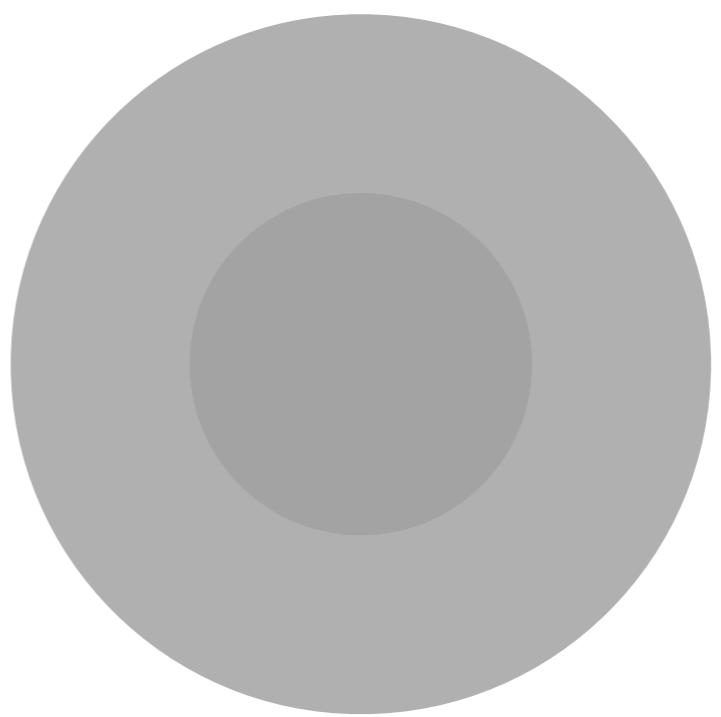
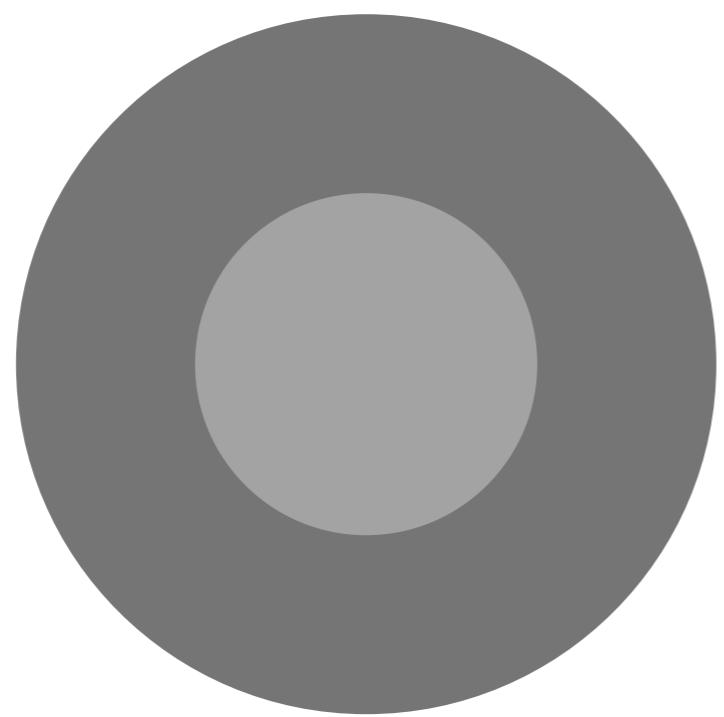


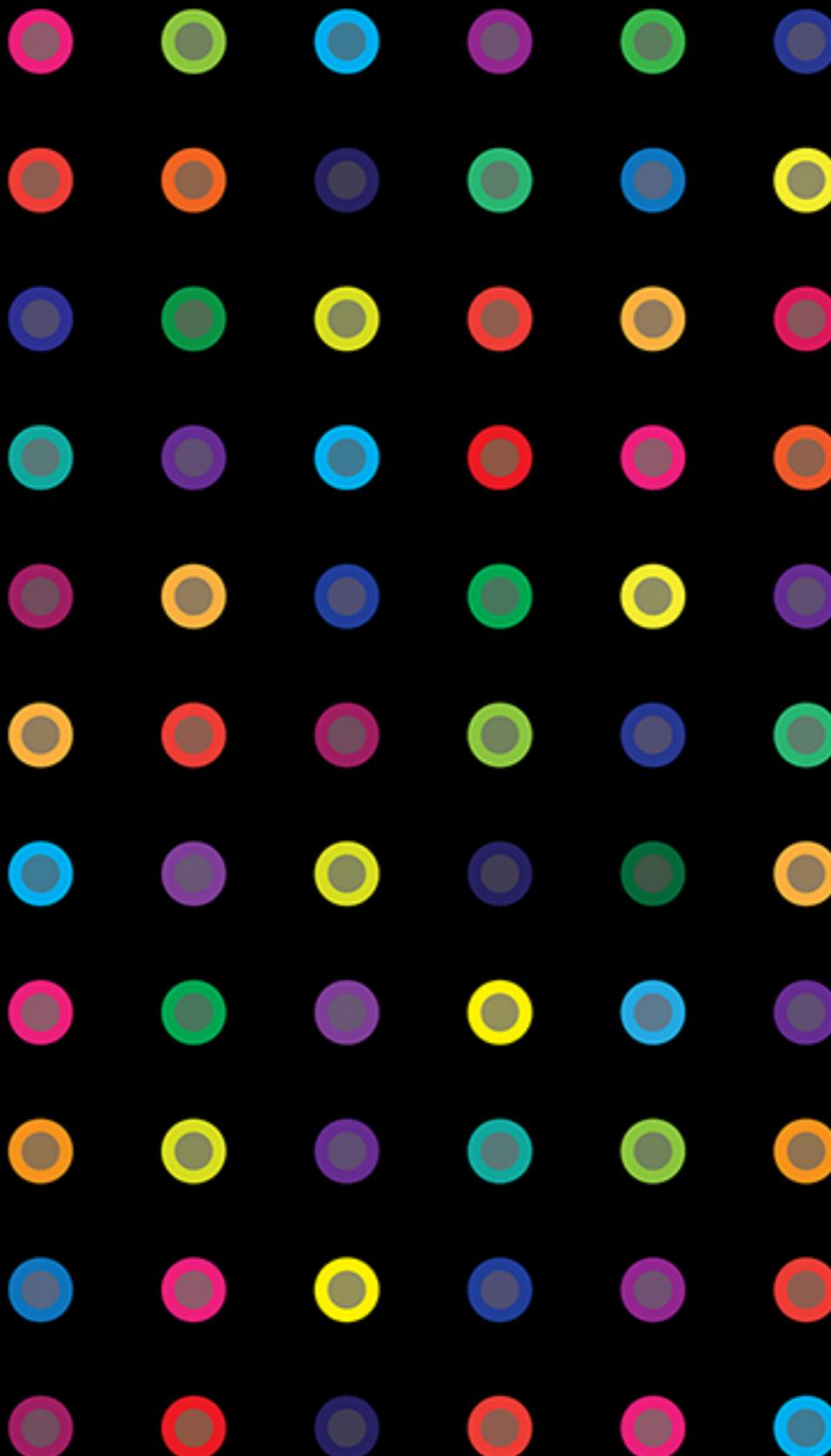
# THE INCOMPARABLE NETWORK



**RESPECT HUMAN VISUAL LIMITATIONS**







The first cases of what would later become known as AIDS were reported in the United States in June of 1981. Since then, 1.9 million people in the U.S. and around the world have been infected with HIV, including over 415,000 who have already died and approximately 1.2 million (1,174,000) adults and adolescents who were living with HIV infection at the end of 2006, the most recent year for which national prevalence estimates are available. The impact of the HIV/AIDS epidemic spans the nation with HIV diagnoses having been reported in all 50 states, the District of Columbia, and the U.S. territories, possessions, and associated nations.

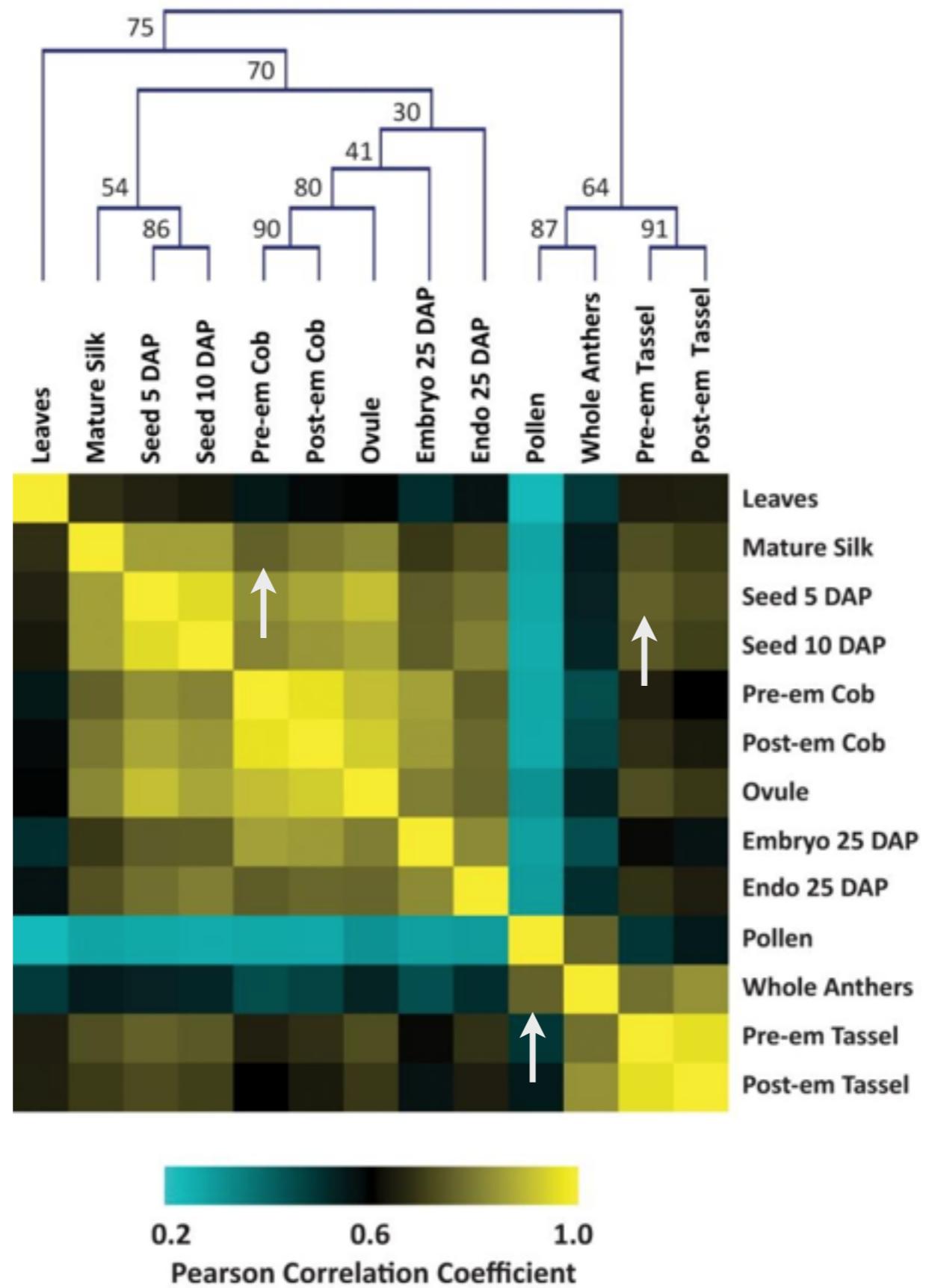
The virus can be transmitted through blood, semen, pre-cervical fluid, vaginal fluid and breast milk. It is estimated that more than one million people are living with HIV in the U.S. And even more amazingly, one in five (20%) of those people living with HIV is unaware of their infection.

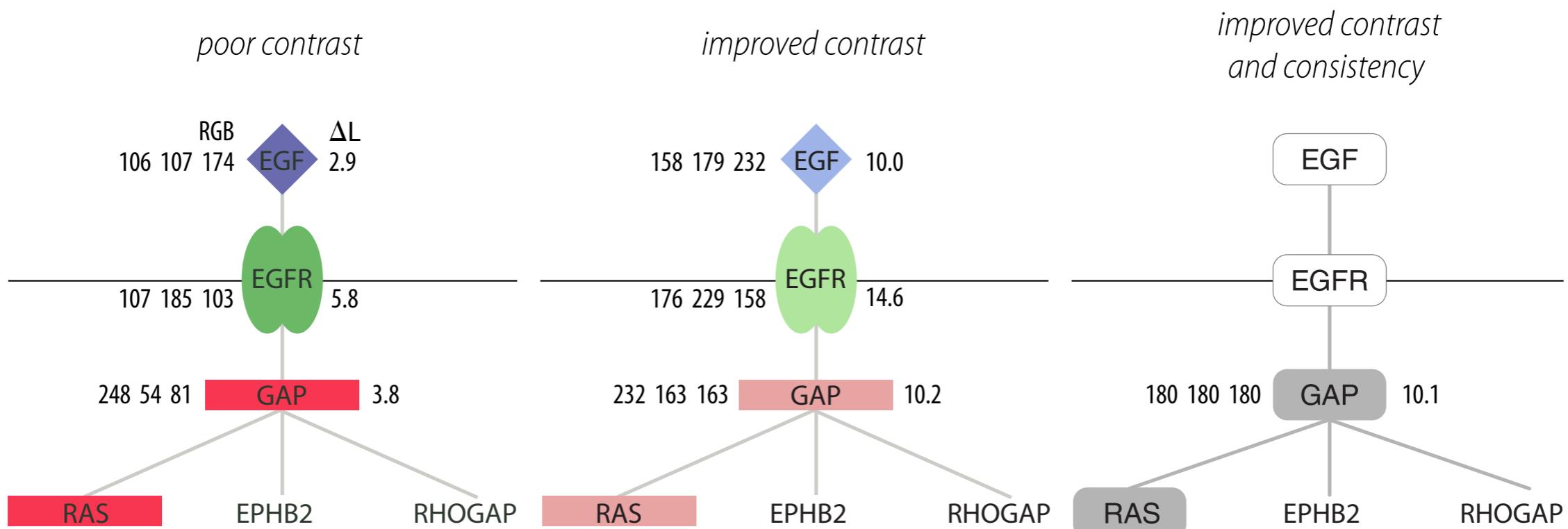
The easiest way to prevent the spread of HIV and AIDS is by wearing a condom when you have sex with your partner, especially if you don't know their sexual history. And with condoms so readily available, there's really no excuse not to wear one!

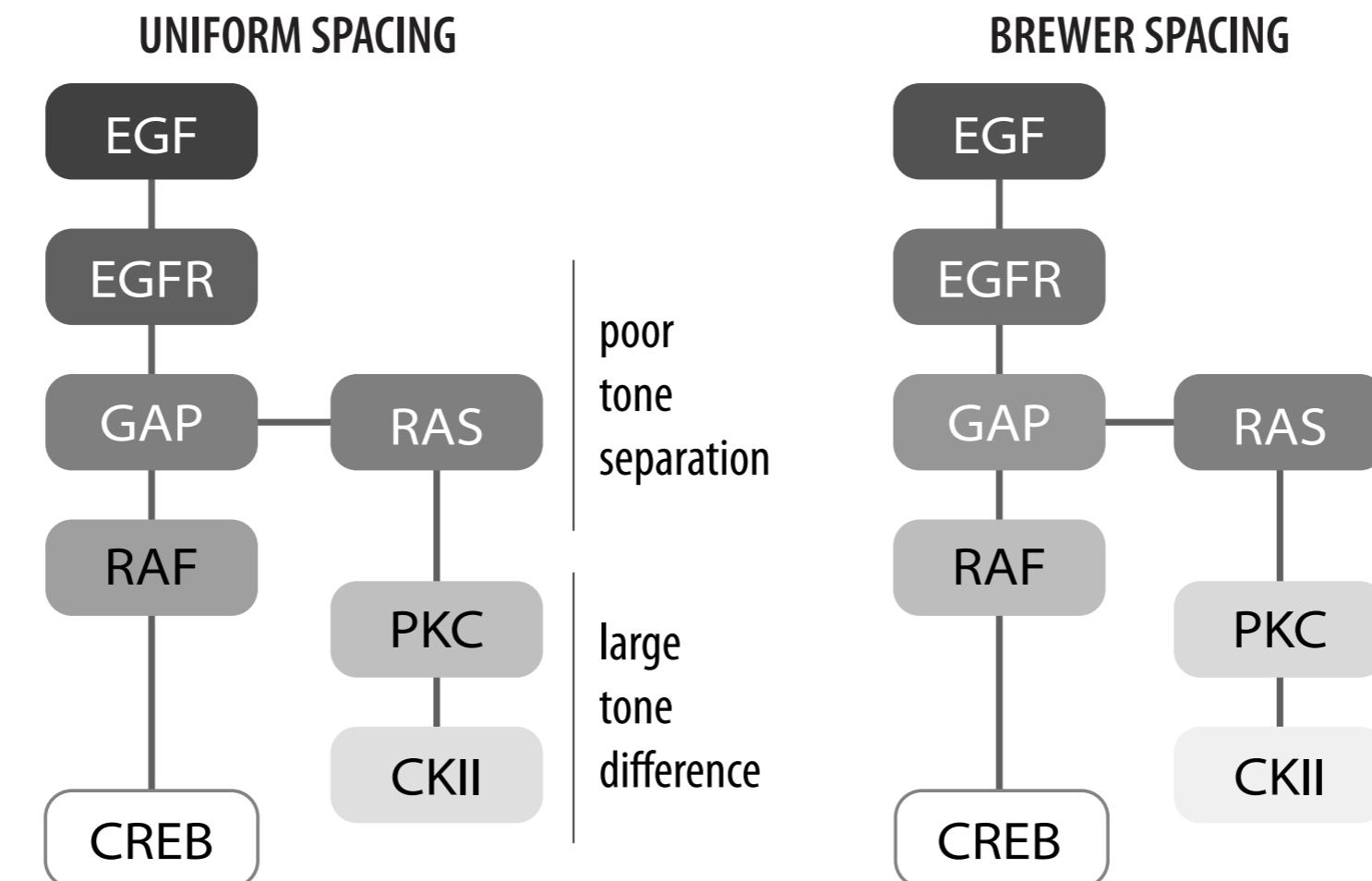
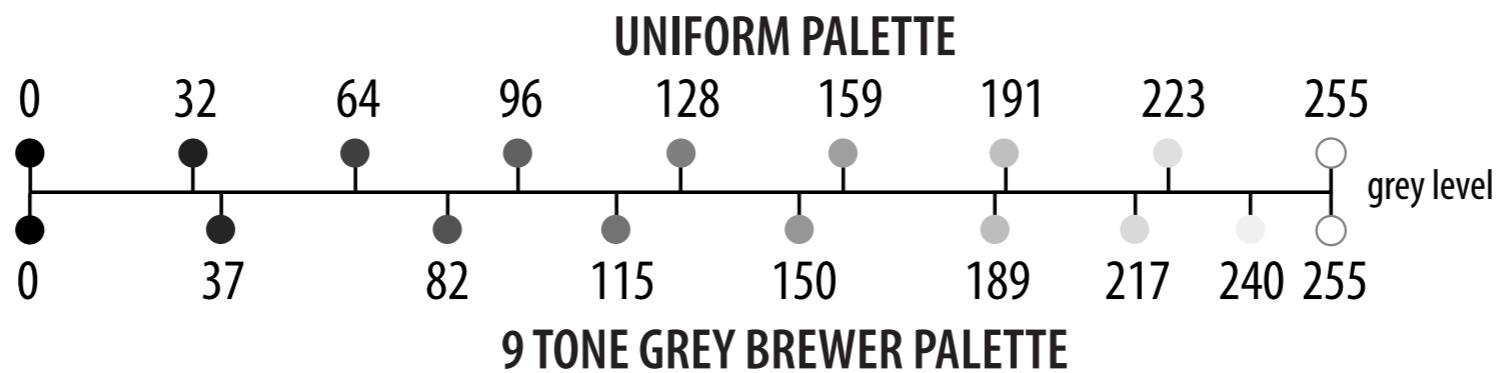


**GET LAID, NOT SCREWED.  
CONDOMS SAVE LIVES.**

amfAR®  
AIDS RESEARCH  
[WWW.AMFAR.GOV](http://WWW.AMFAR.GOV)



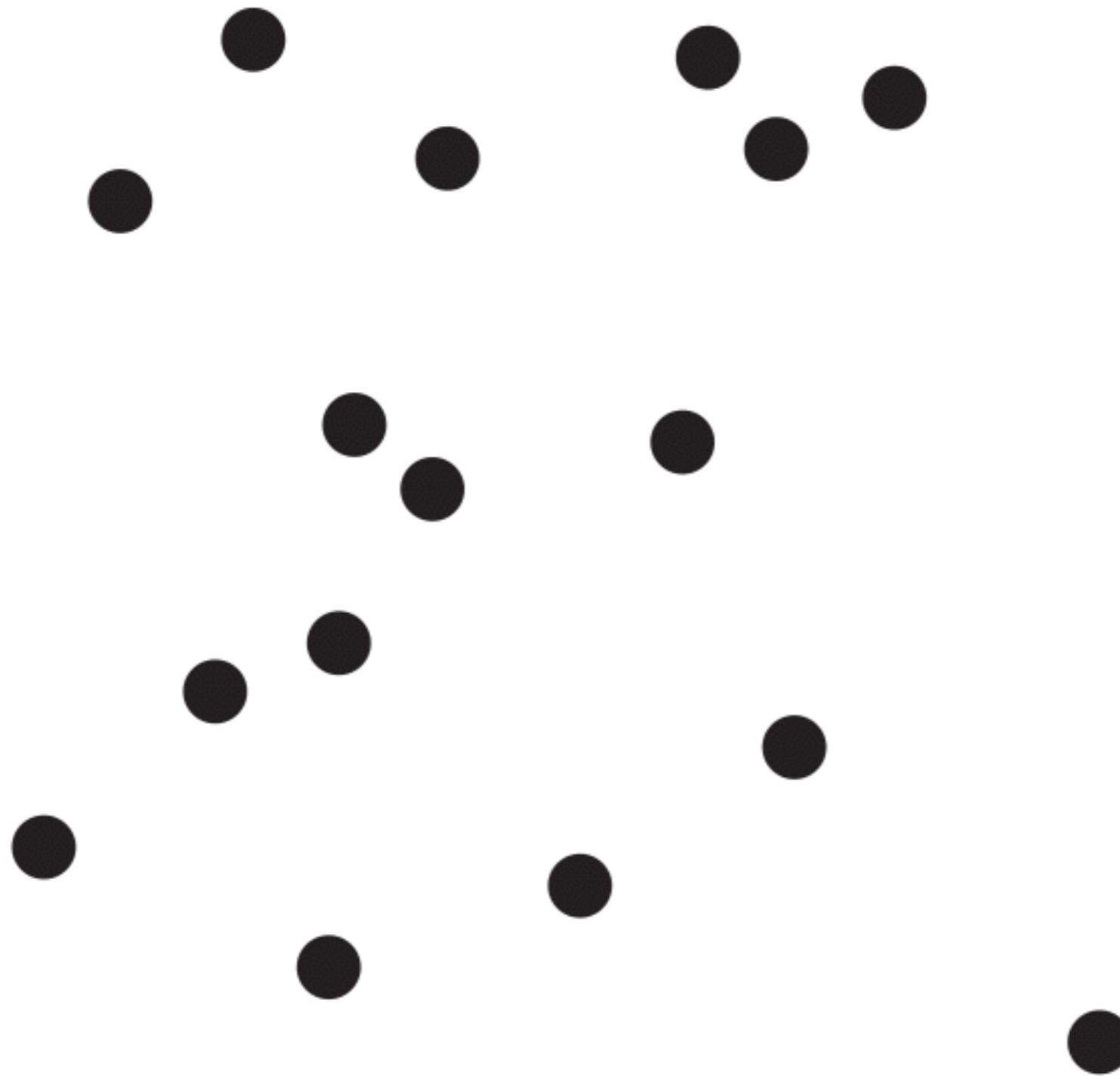




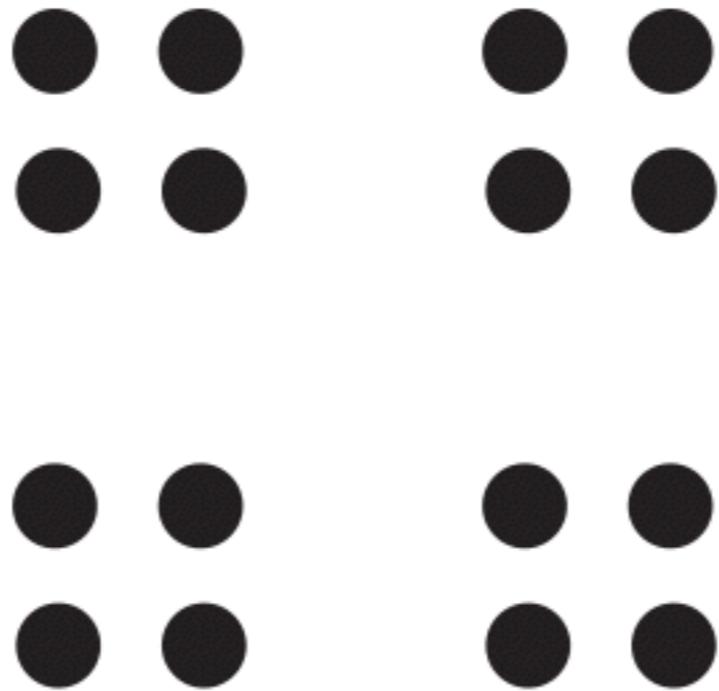
# DATA INFORMS VARIATION

Patterns are hard to see when variation is due to both data and formatting.

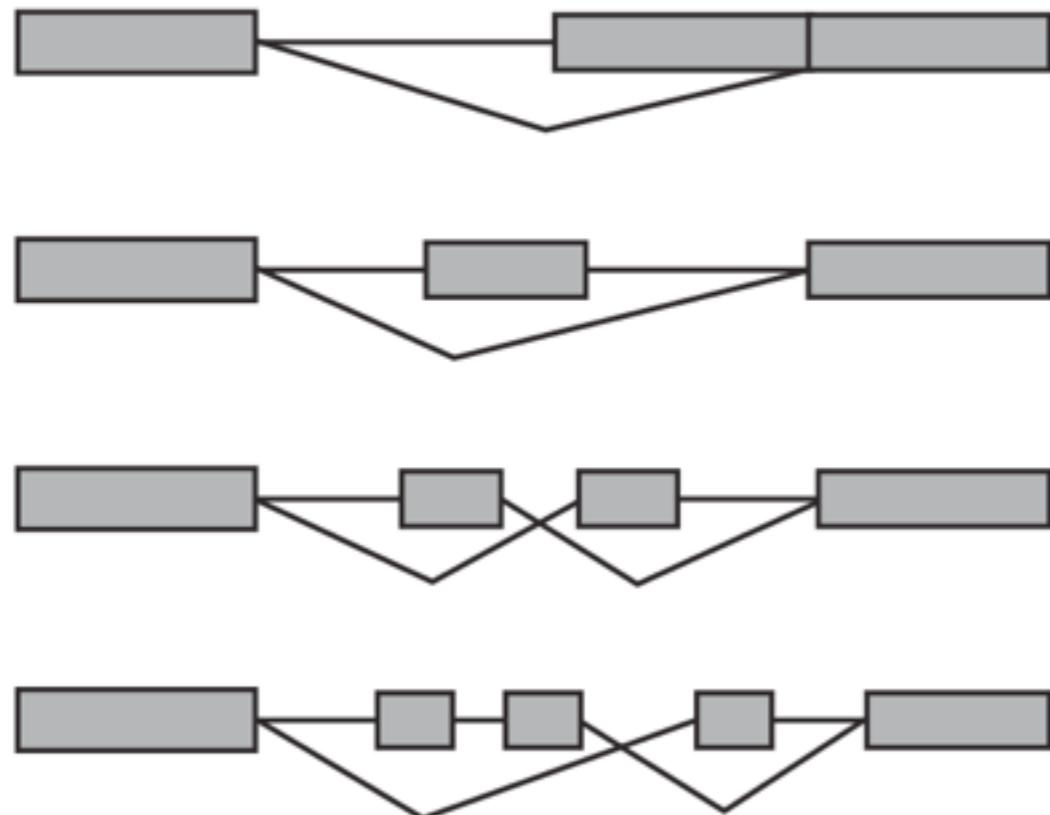
# HOW MANY DOTS?



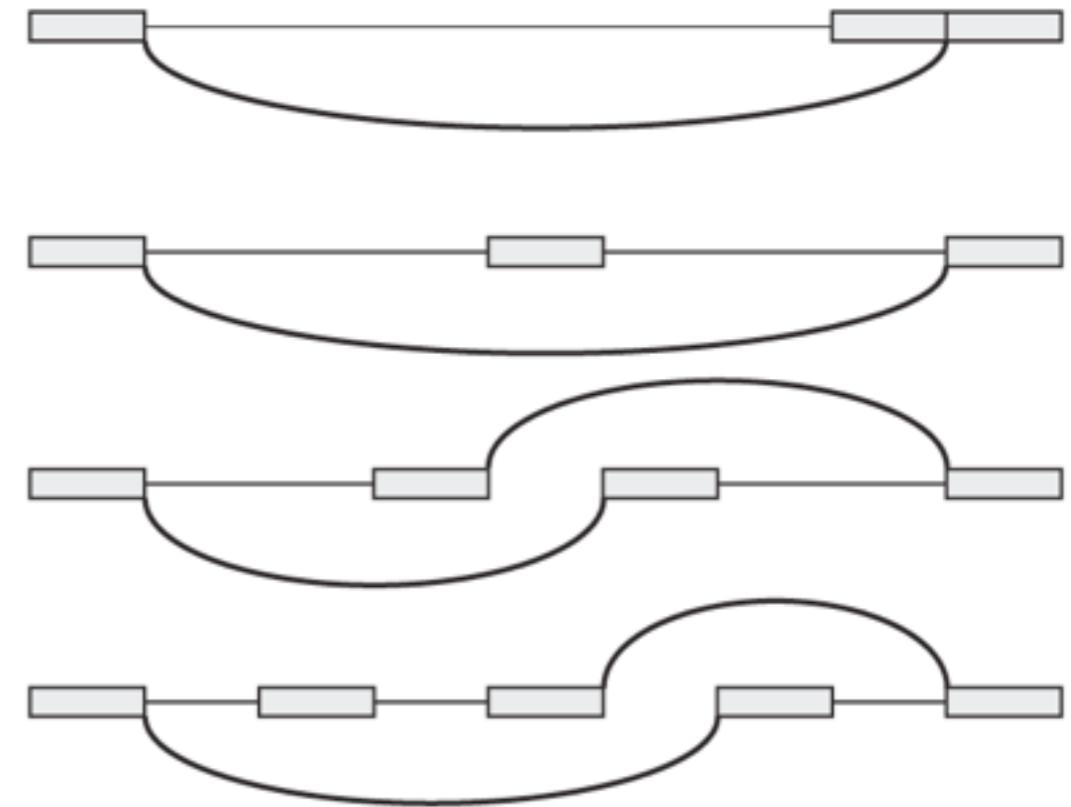
# AVOID EXCESS DEGREES OF FREEDOM



*spacing variation is implied*



*variation refactored*



# CONSISTENCY

Avoid the use of similar encodings for independent variables.

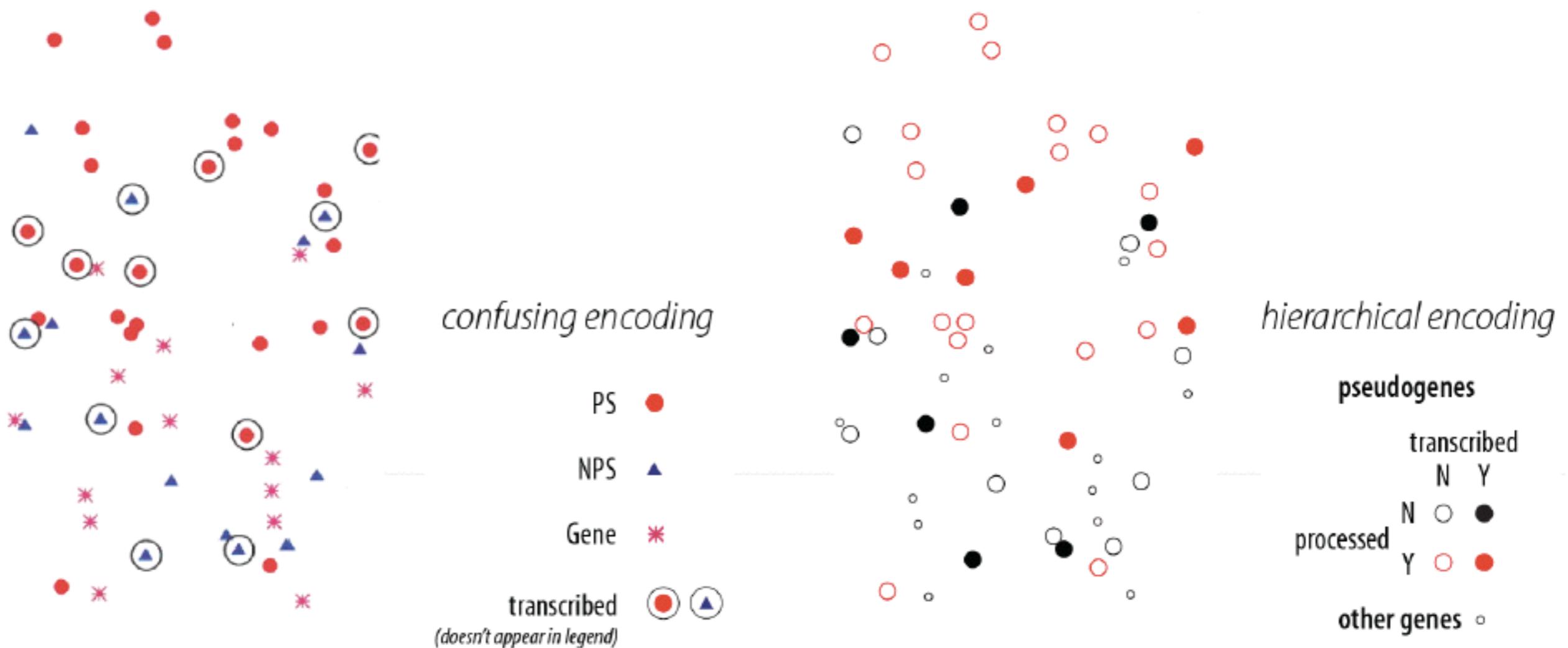
# Consistency

---

- Visual variation in a figure should always reflect and enhance any underlying variation in the data.
- Avoid using more than one encoding to communicate the same information.
- Do not use visually similar encodings for independent variables

# Consistency

- red - processed genes, but salience attenuated
- other genes encoded with competing glyph - red star.

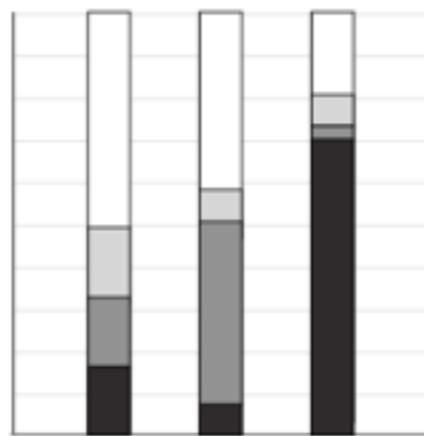


# Consistency

---

- Order items in a legend according to order of appearance in the plot

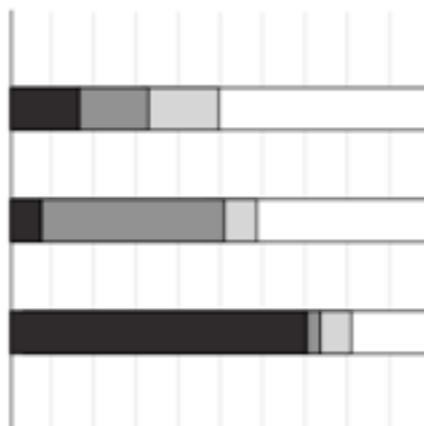
*consistent*



□ A  
□ B  
■ C  
■ D

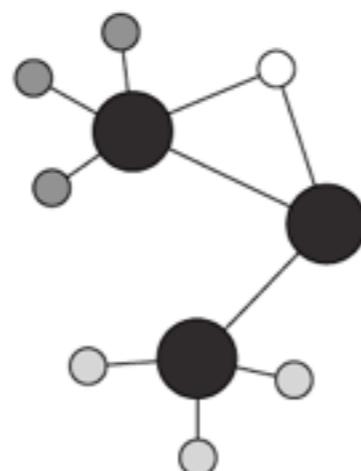
*inconsistent*

■ A  
■ B  
□ C  
□ D



■ A  
■ B  
□ C  
□ D

□ A  
□ B  
■ C  
■ D



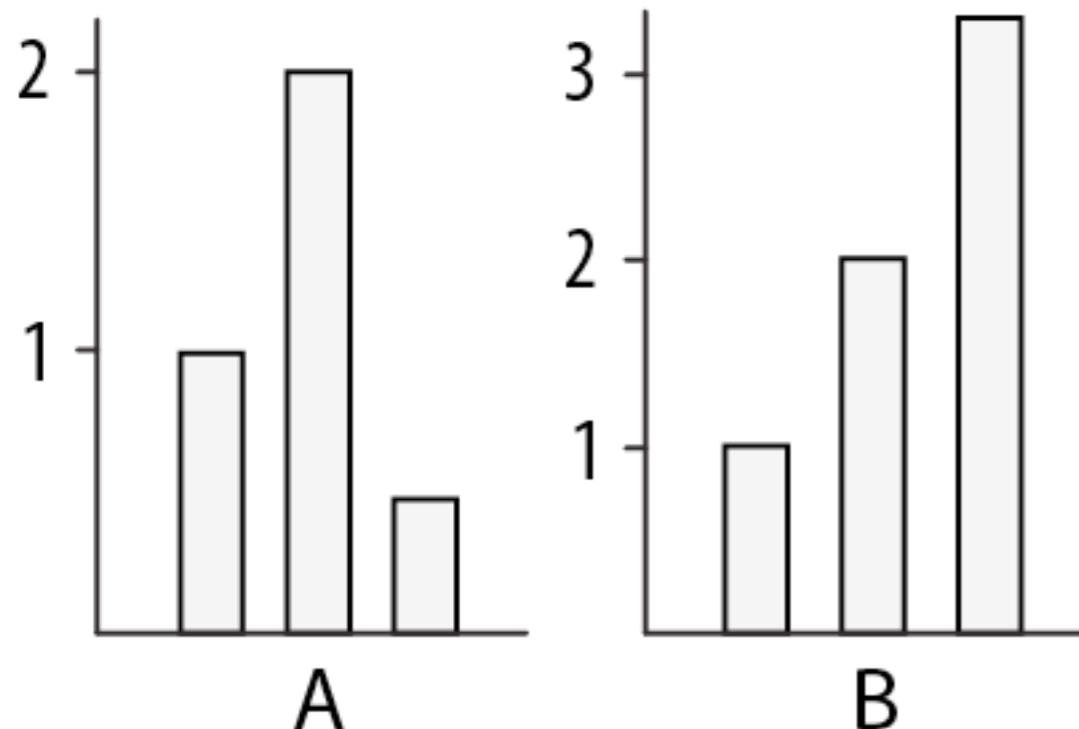
○ A  
○ B  
● C  
● D

□ A  
□ B  
■ C  
■ D

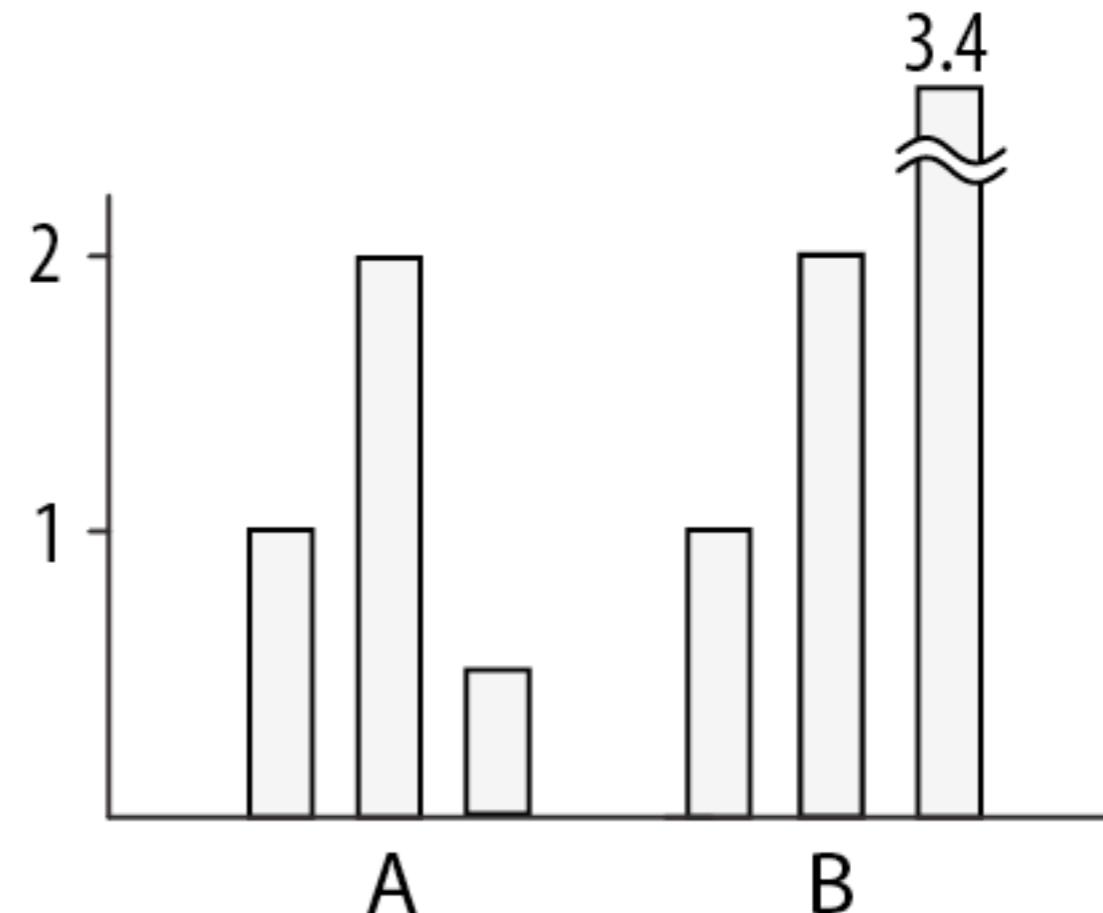
# Consistency - Navigational aids

- Use consistent axes when comparing charts

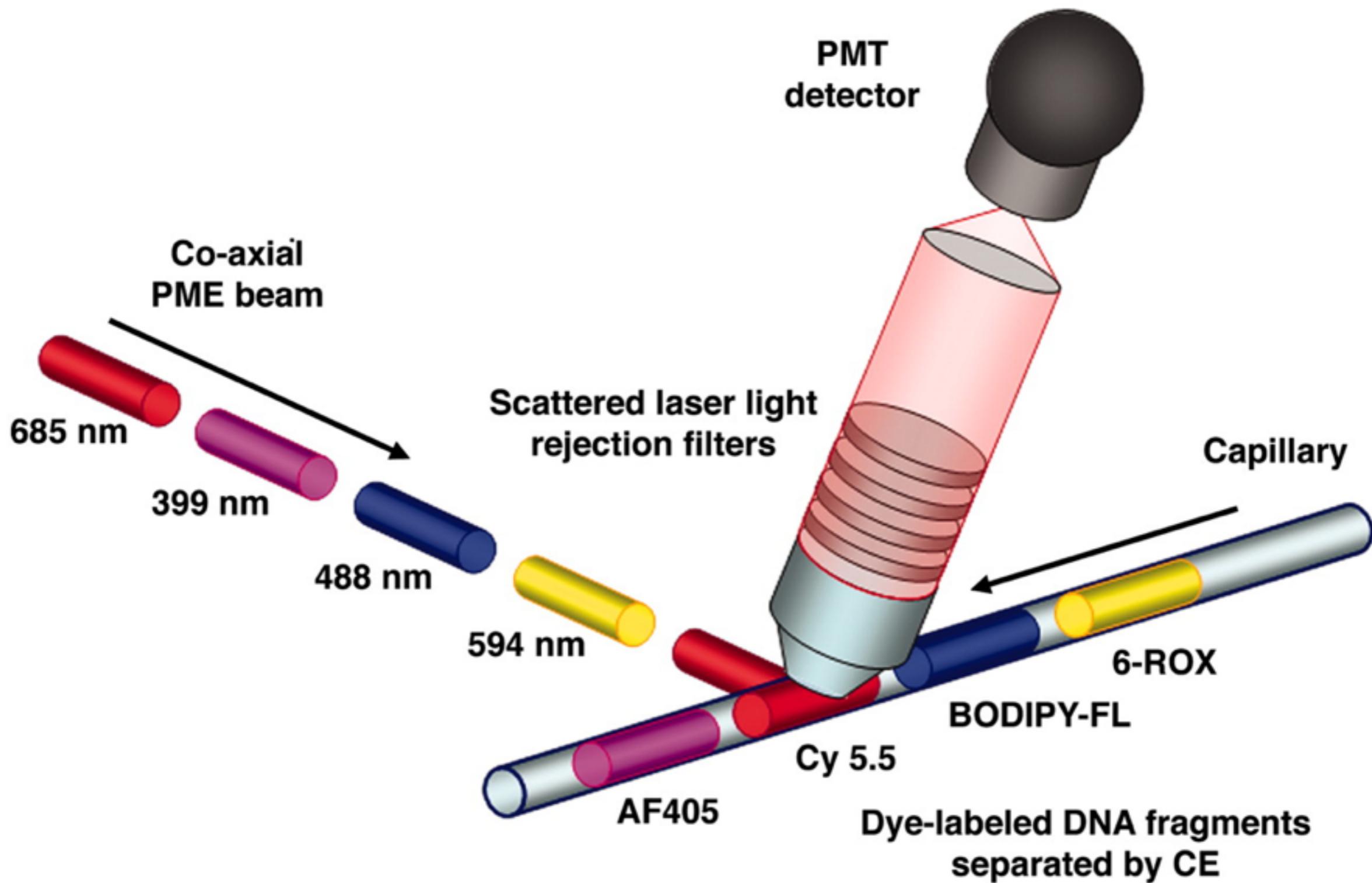
*misleading*



*improved*



e.g. Fig 1 in Raina SZ, Faith JJ, Disotell TR, Seligmann H, Stewart CB, et al. (2005) Evolution of base-substitution gradients in primate mitochondrial genomes. Genome Res 15: 665-673.

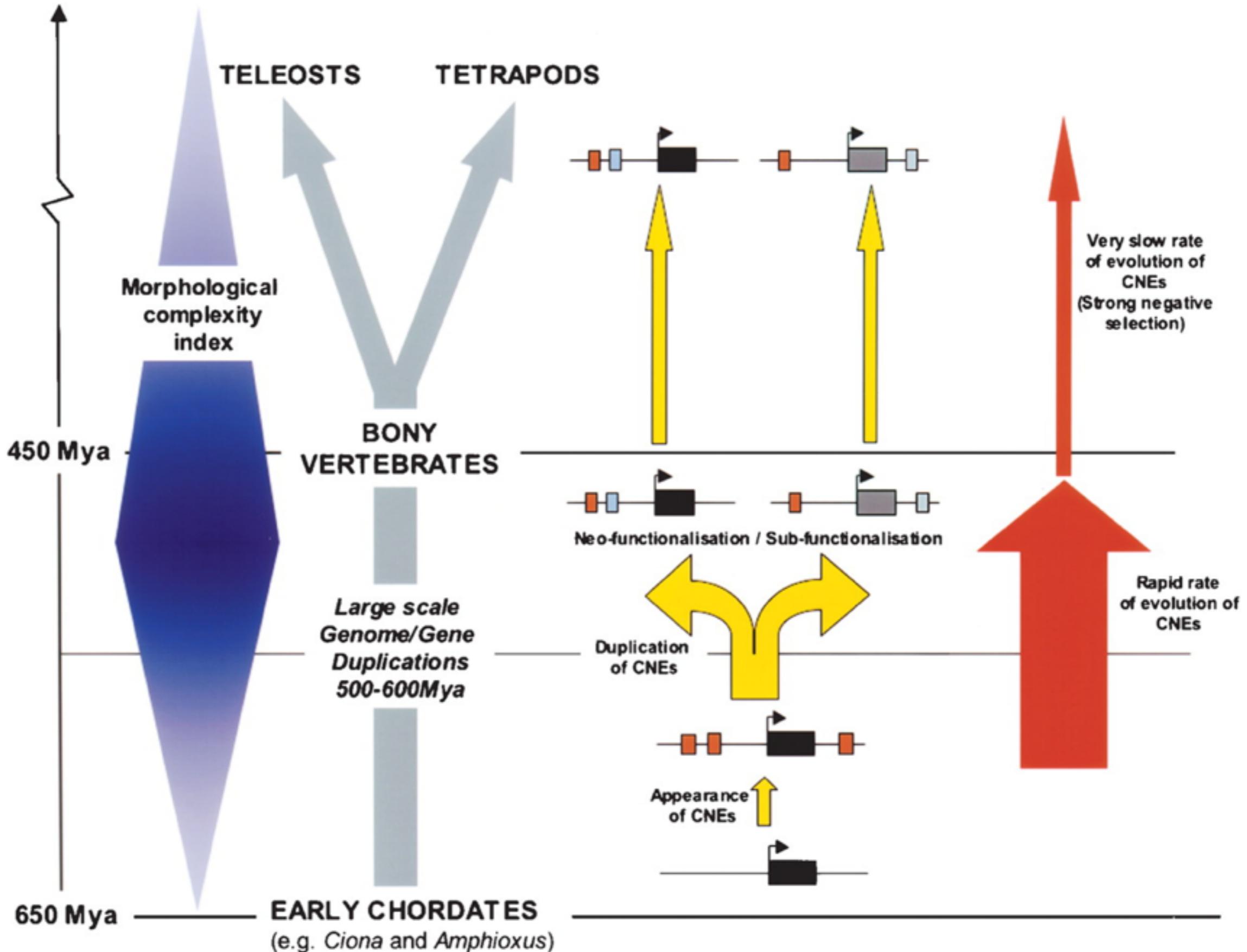


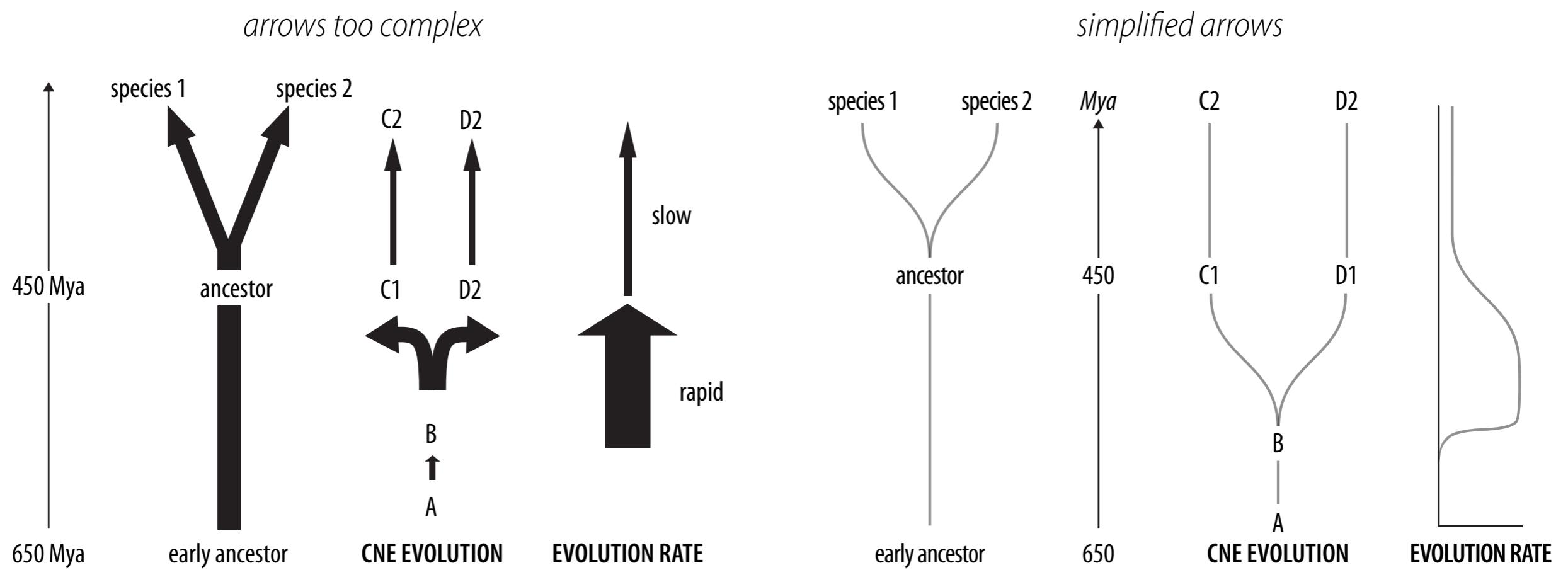
*homonymous encoding*



*disambiguated*







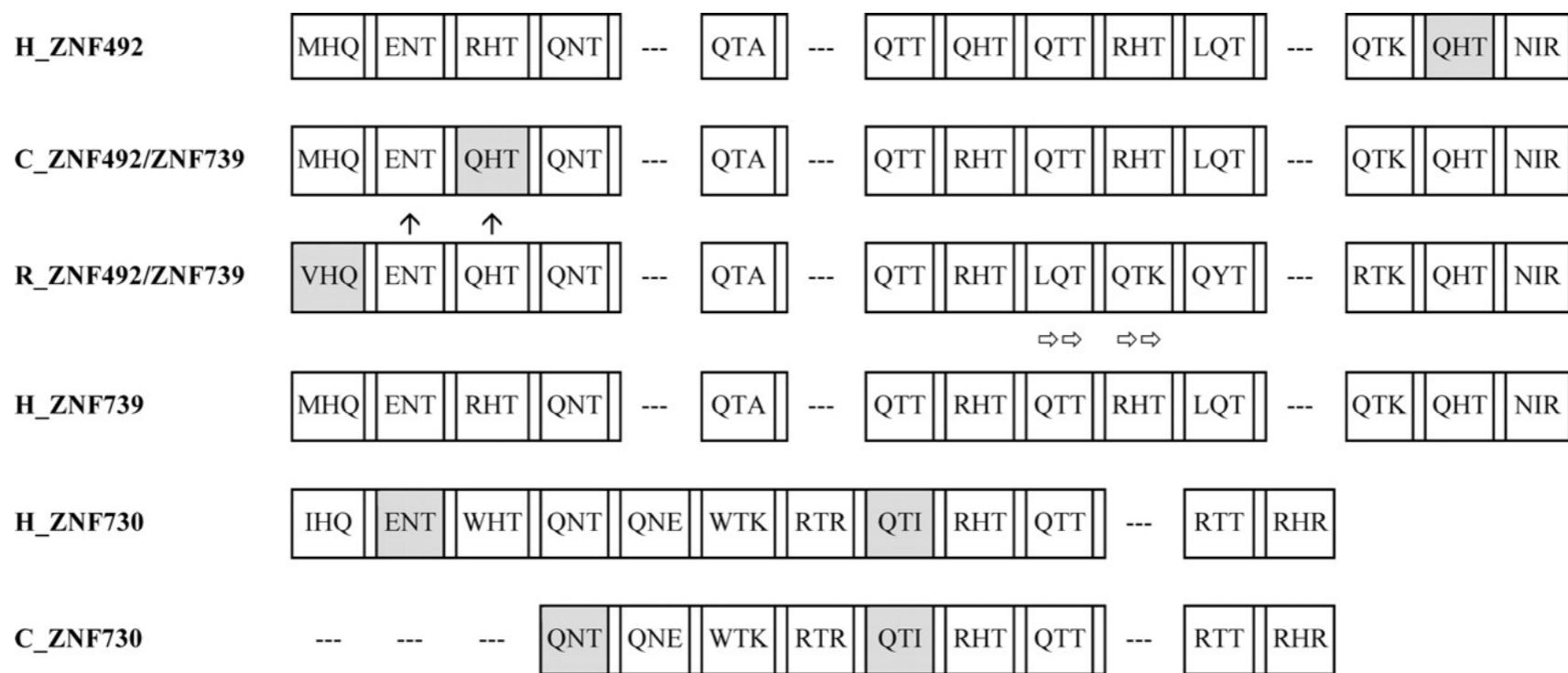
# Consistency - numbers

---

- Use the same amount of digits in the numbers - (avoid e.g. 234000, 34.567 together)
- Percentage should always sum up to 100%

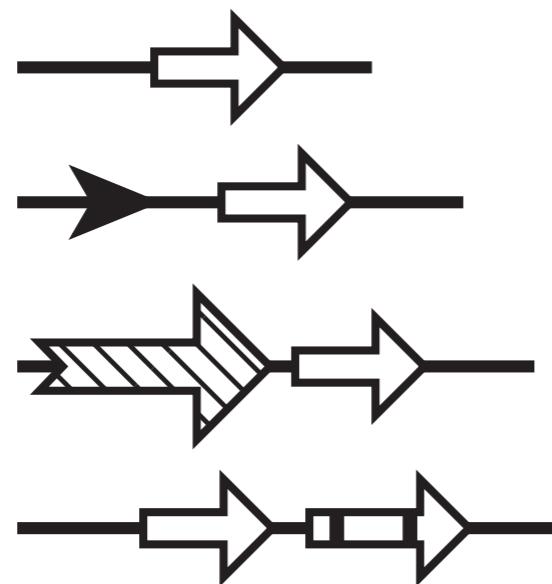
# REDUNDANCY

Don't repeat yourself. Don't repeat yourself.

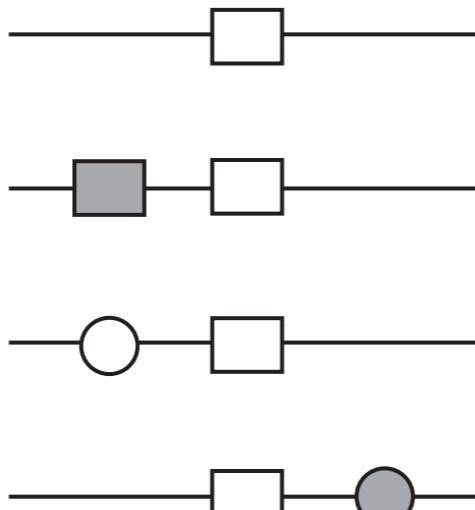


	MHQ	ENT	RHT	QNT	QNE	QTA	RTR	QTT	QHT	QTT	RHT	LQT	RHR	QTK	QHT	NIR
H492	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
C492/739	...	...	Q...	...	...	...	...	...	R..	...	...	...	...	...	...	
R492/739	V..	...	Q..	...	...	...	...	...	R..	LQ.	QTK	QY.	...	R..	...	
H739	M..	...	...	...	...	...	...	...	R..	...	QTK	...	...	...	...	
H730	I..	...	W..	...	...	W.K	...	...I	R..	...	RT.	...				
C730		...	...	W.K	...	...I	R..	...	RT.	...						

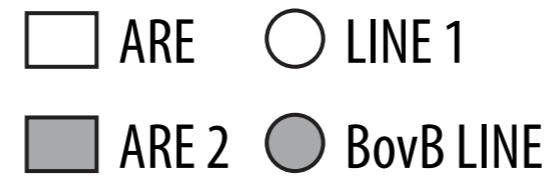
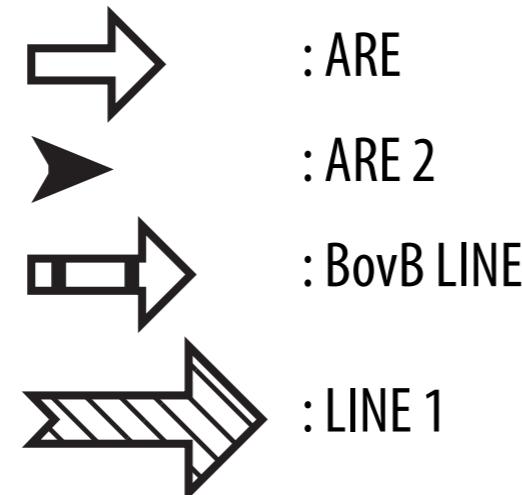
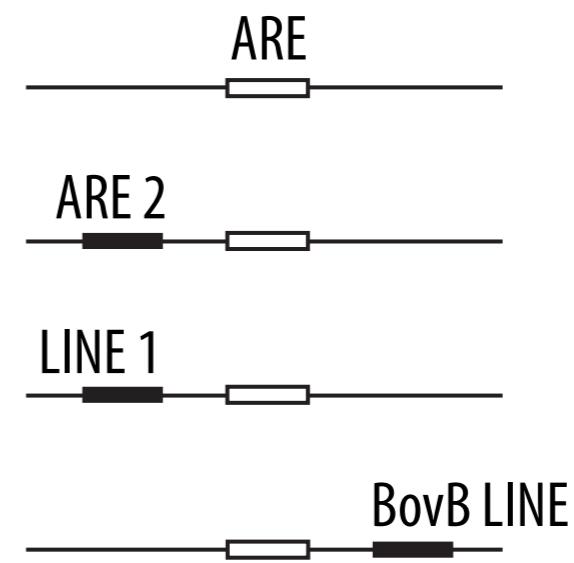
*overwhelming*



*simplified*



*integrated key*

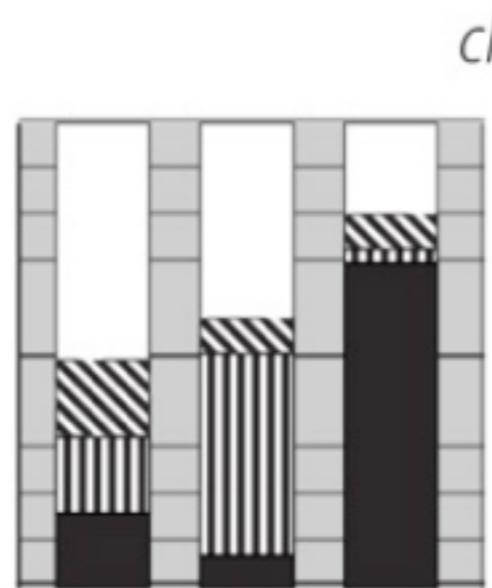


# CONCISENESS

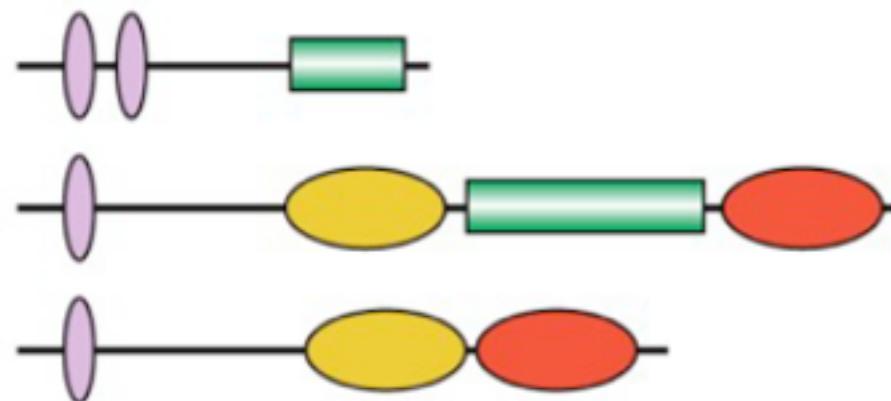
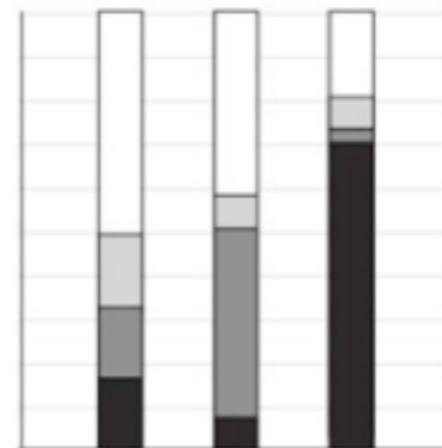
Use the fewest elements possible—keep data-to-ink ratio high

# Increase data:ink ratio

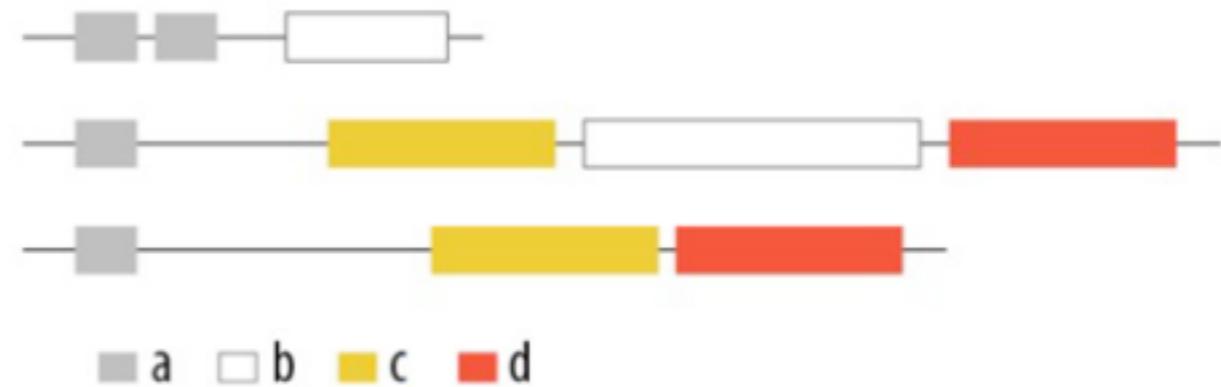
- Avoid “Chart junk”



*visually concise*



key	
■	= a
■	= b
■	= c
■	= d

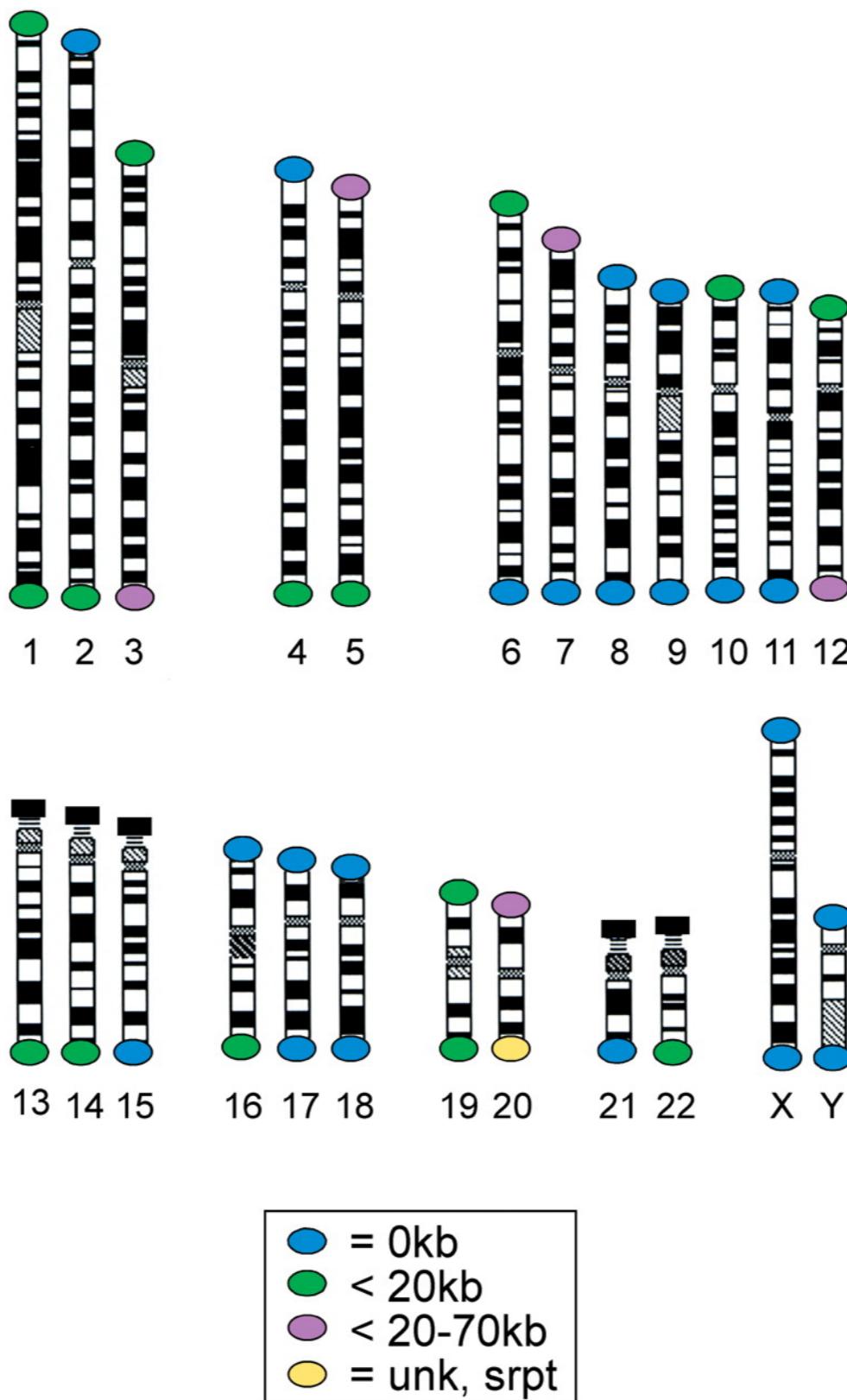


■	a
□	b
■	c
■	d

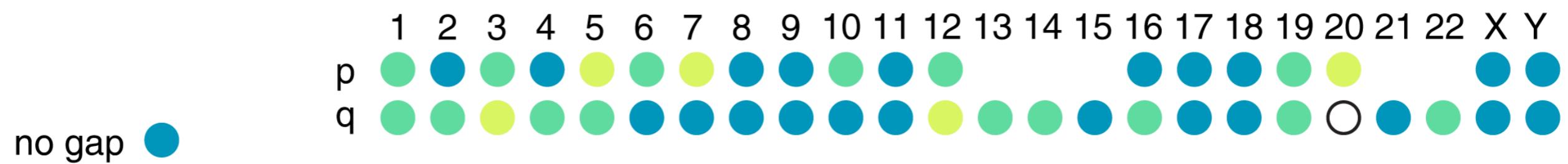
## REMOVE TO IMPROVE

Use the fewest elements possible—keep data-to-ink ratio high.

Shelter your reader from unnecessary complexity.



## OPTION 1



no gap

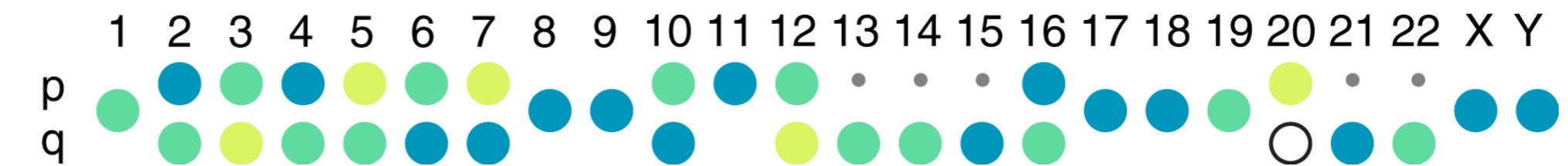
<20kb

<20-70kb

unk/srp

acrocentric

## OPTION 2



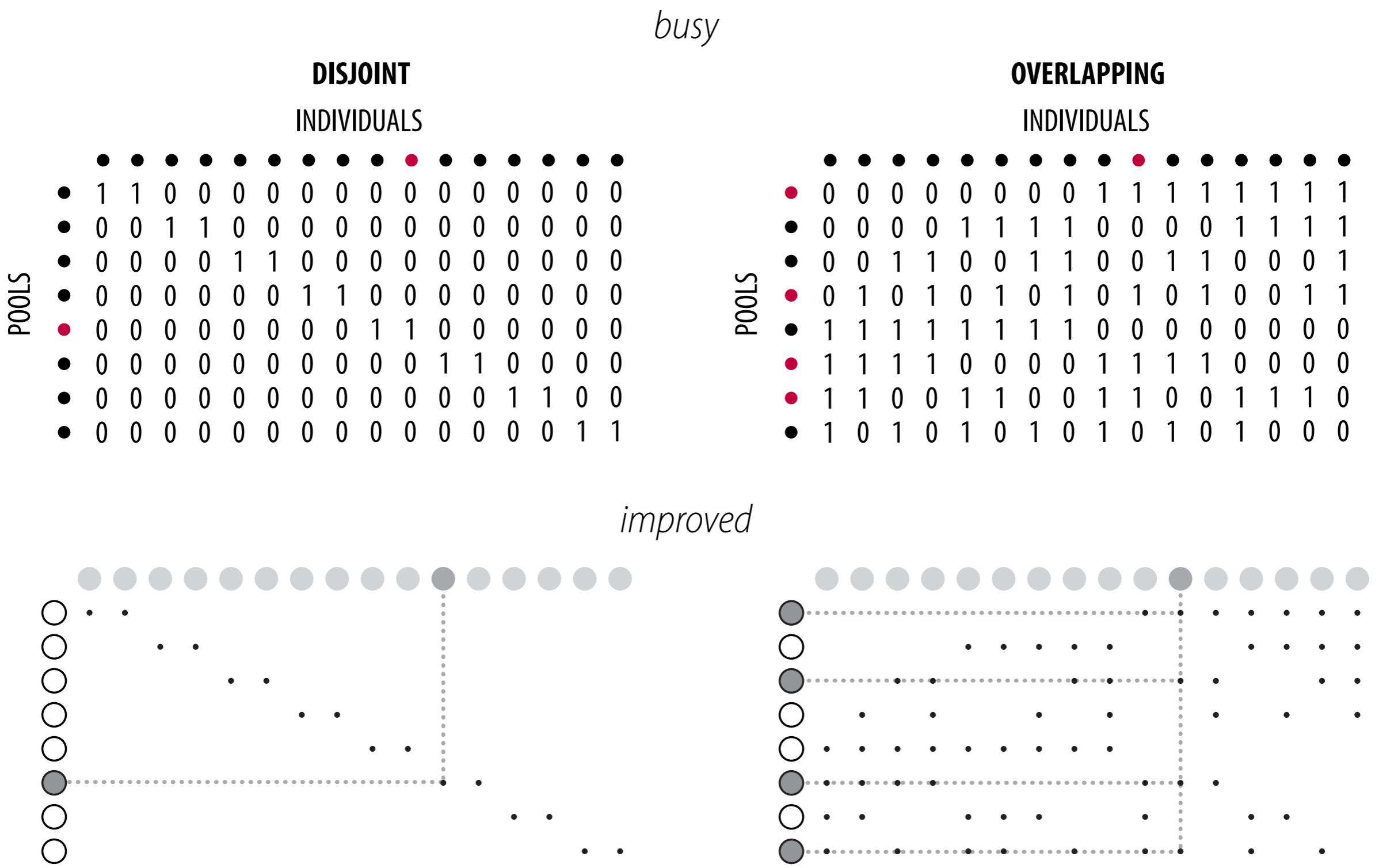
INDIVIDUALS

**Pools**

1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

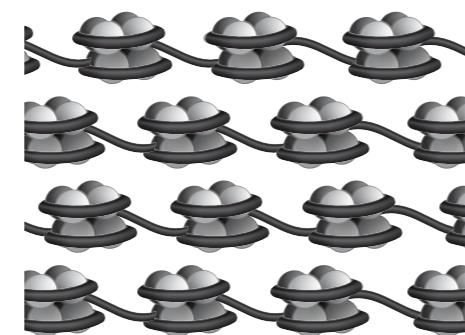
1	0	0	0	0	0	0	0	1	1	1	1	1	1	1
0	0	0	0	1	1	1	1	0	0	0	1	1	1	1
0	0	1	1	0	0	1	1	0	0	1	0	0	1	1
0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	1	1	1	1	0	0	0
1	1	0	0	1	1	0	0	1	1	0	0	1	1	0
1	0	1	0	1	0	1	0	1	0	1	0	1	0	1



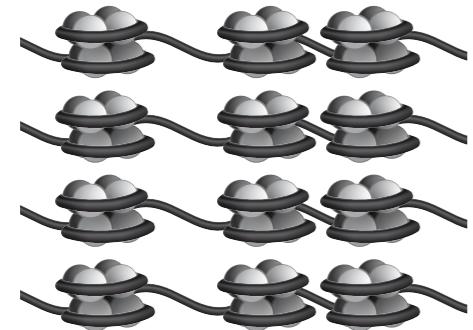
*detail exposed*



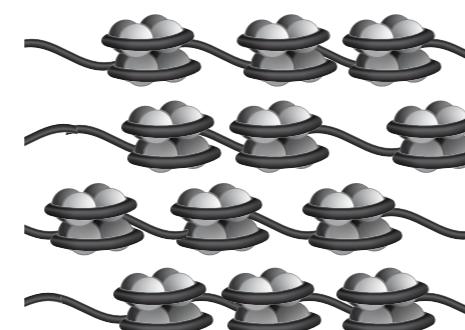
Not Positioned but  
Uniformly Spaced



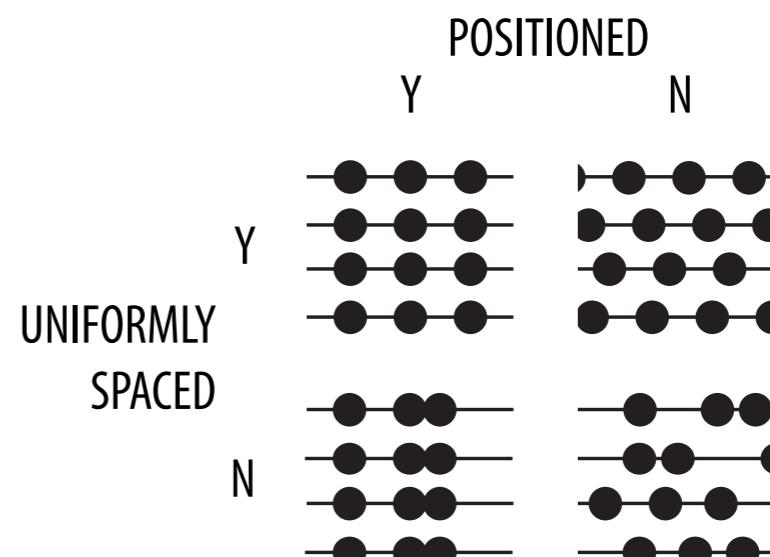
Positioned but  
Not Uniformly Spaced



Not Positioned and  
Not Uniformly Spaced



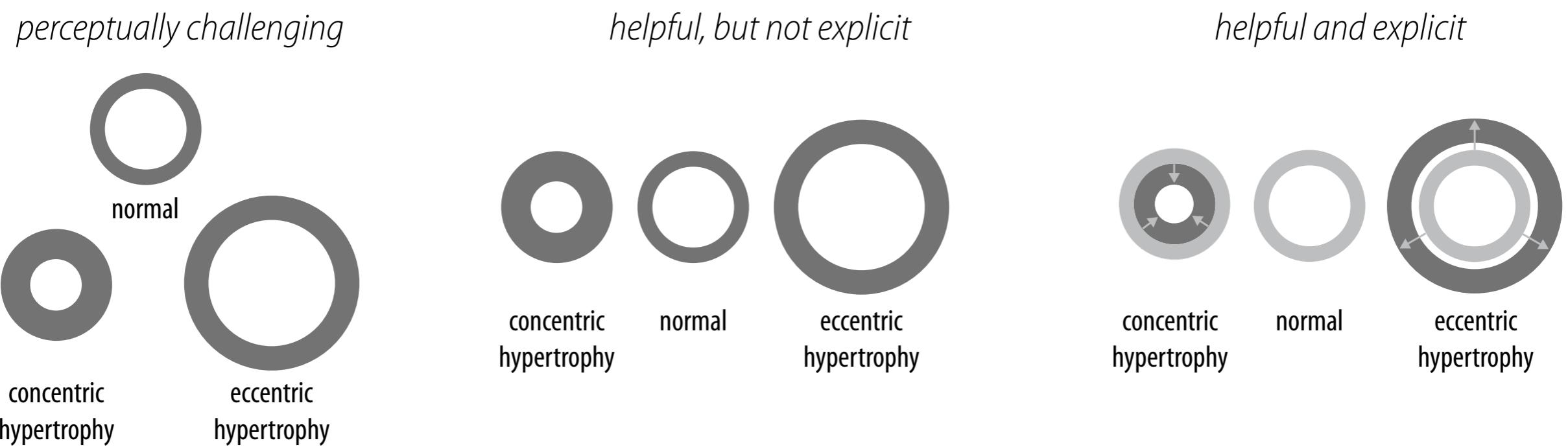
*detail encapsulated*



# CLARITY

Make sure that elements are visible and unobscured.

Don't count on your audience to figure out what you mean. Say it.



## FOCUS & EMPHASIS

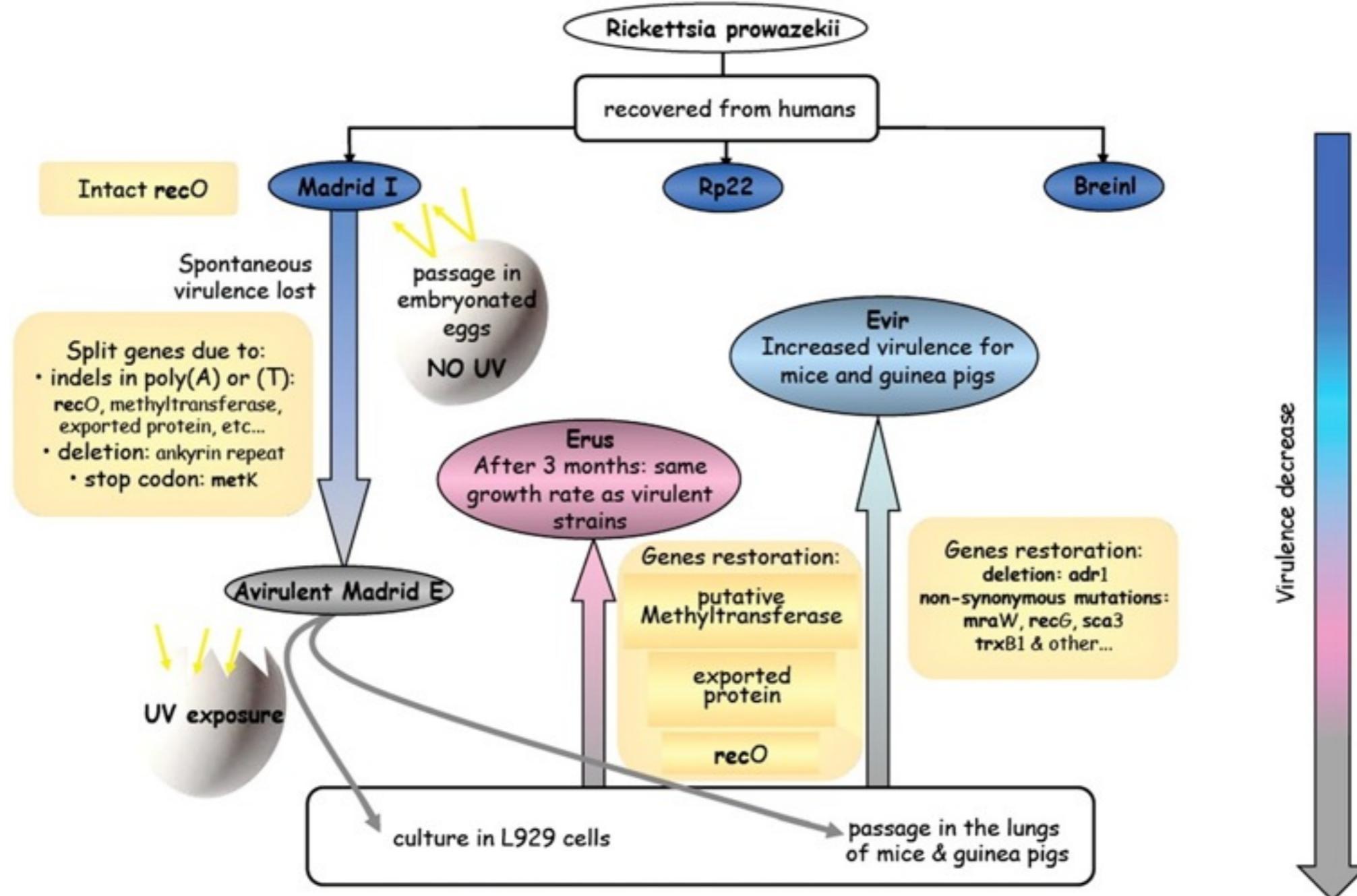
Match the pertinence of an object with its visual salience.

Apply visual organization Gestalt principles.





# EVERYTHING IS EMPHASIZED



# NOTHING IS EMPHASIZED

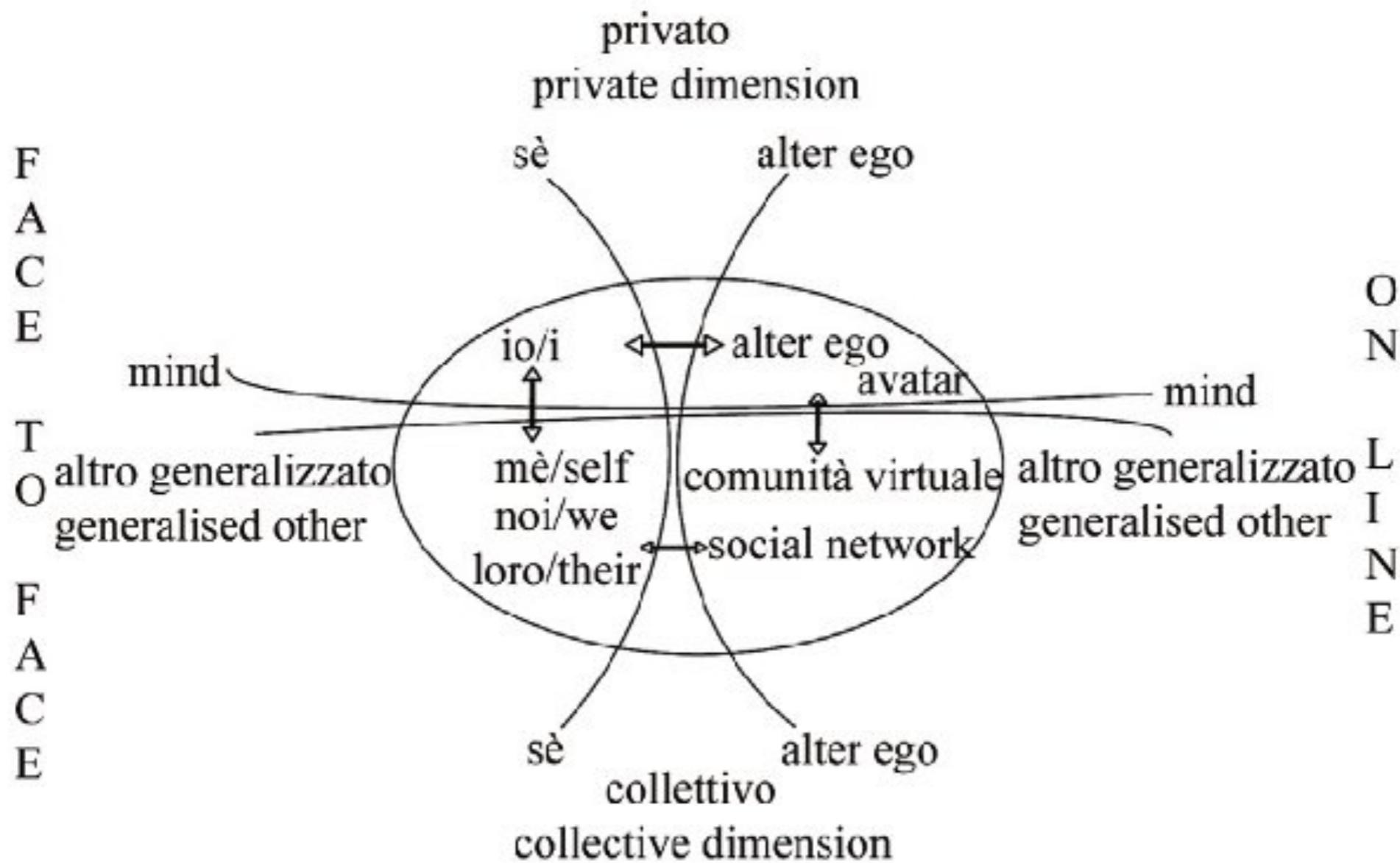
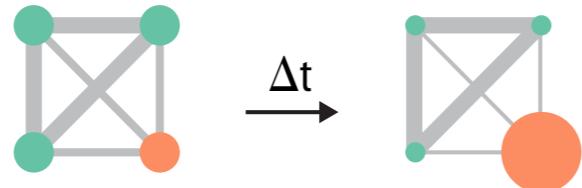


Figure 2.  
*Severino's flowchart.*

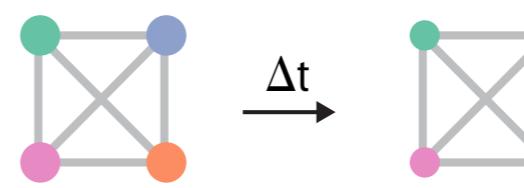
# MATCH VISUAL SALIENCE TO RELEVANCE

## ATTENTIONAL CAPTURE

DISTINCT



HETEROGENEOUS



inhibitory interaction

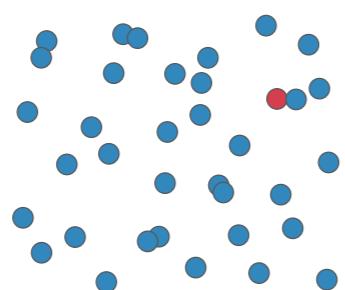
—  
low  
—  
high

neural response

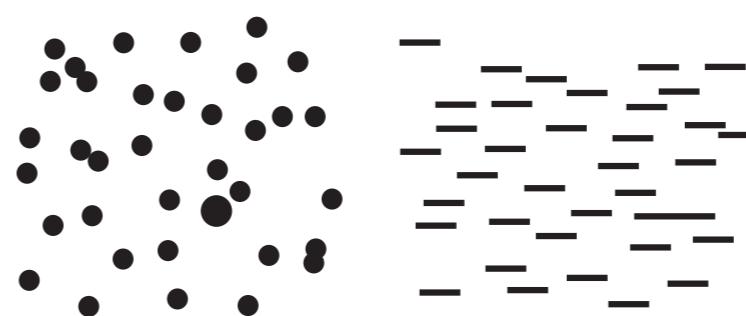
•  
low  
●  
high

## SALIENCE

HIGH



LOW





# Colours

---

- “Colour used poorly is worse than no colour at all” - Edward Tufte
- “Above all, do no harm”
- colour can cause the wrong information to stand out and make meaningful information difficult to see.

# Colour space

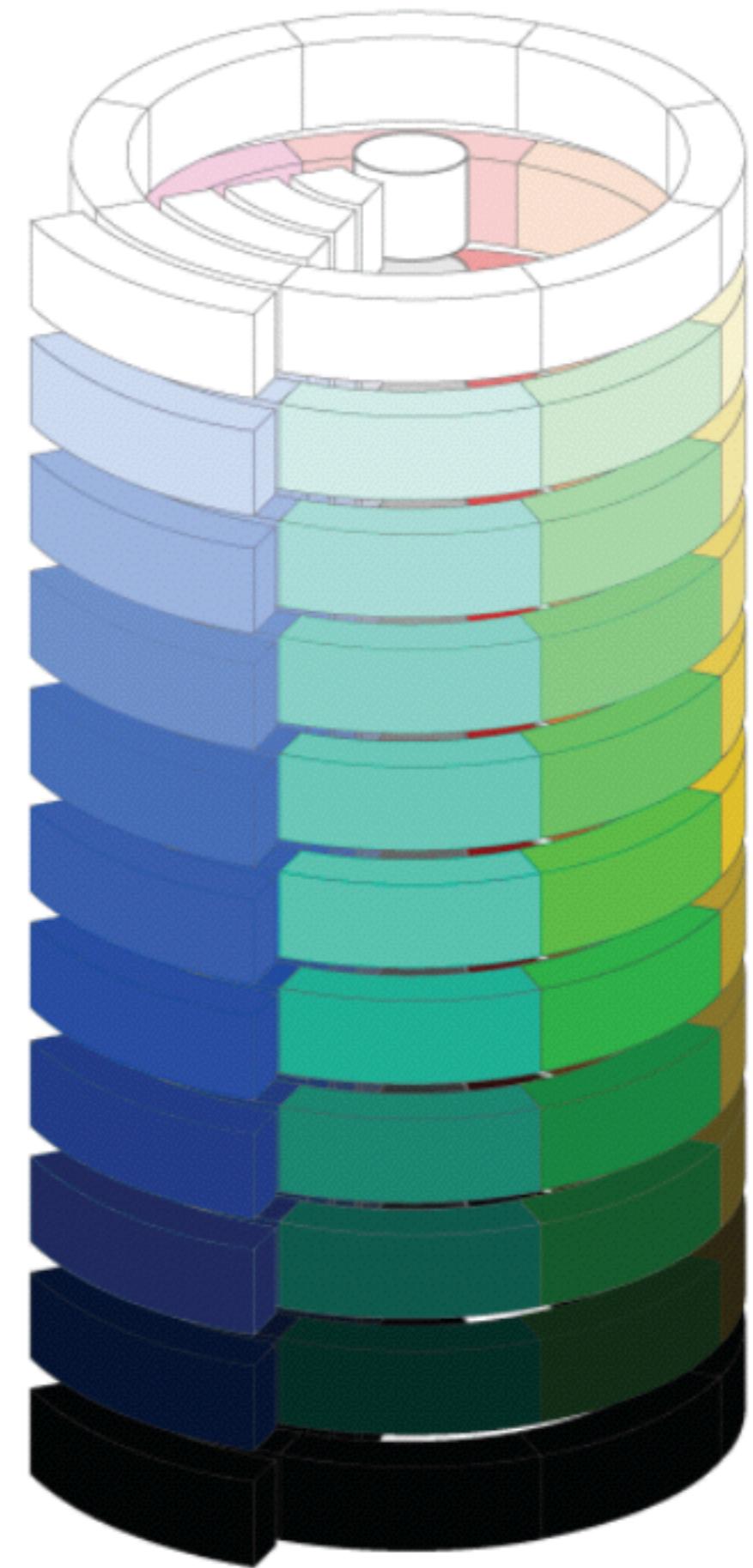
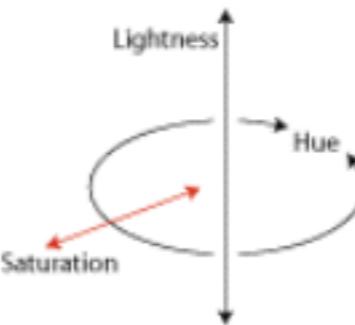
---

- A colour space is mathematical model for describing colour.
- RGB, HSB, HSL, Lab and LCH
- RGB is the most common in computer use,
  - but least useful for design
  - our eyes do not decompose colours into RGB constituents
- HSV, describes a colour in terms of its hue, saturation and value (lightness),
  - models colour based on intuitive parameters
  - more useful.

# Colourimetry

---

- Hue (colour)
  - around the circle
- Saturation
  - Inside to outside
  - Colour to grey scale
- Lightness (value)
  - top to bottom



# Brewer palettes

- Brewer palettes ([colorbrewer.org](http://colorbrewer.org)) provide a range of palettes based on HSV model which make life easier for us....

Avoid the use of hue to encode quantitative variables



Quantitative encoding  
e.g. heat maps



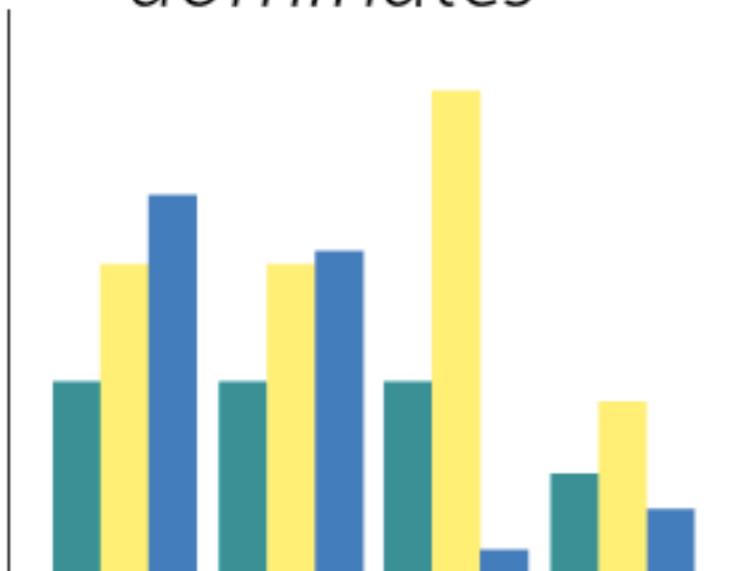
Two-sided quantitative encodings



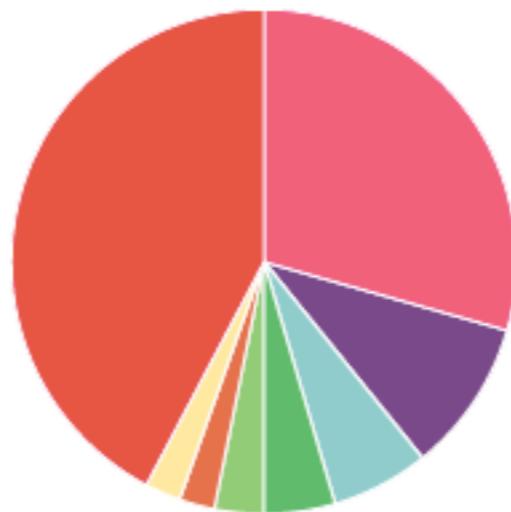
## Examples

*one color  
dominates*

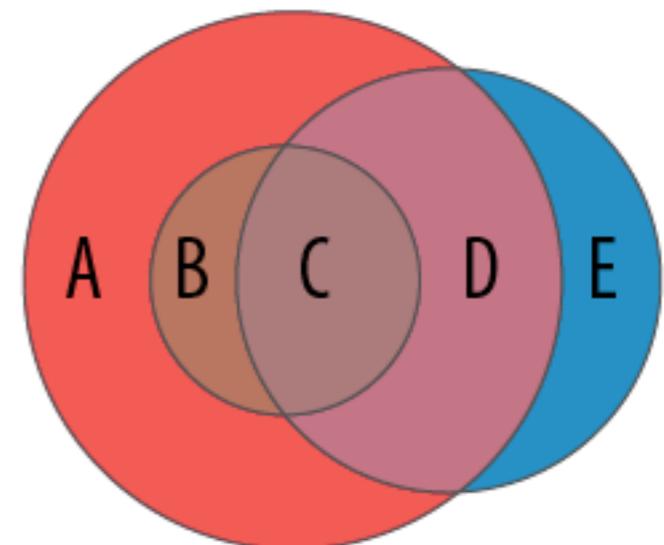
Poor use  
of colour



*difficult to  
distinguish*

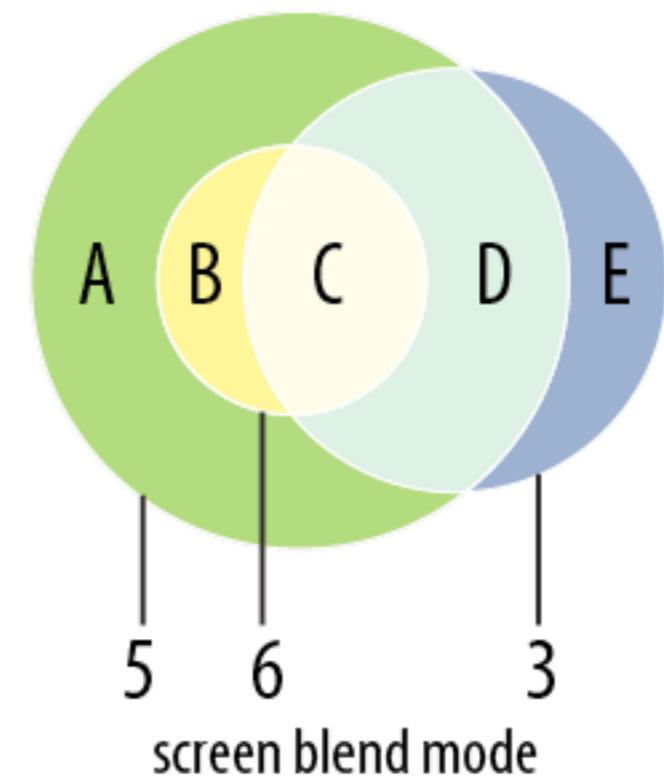
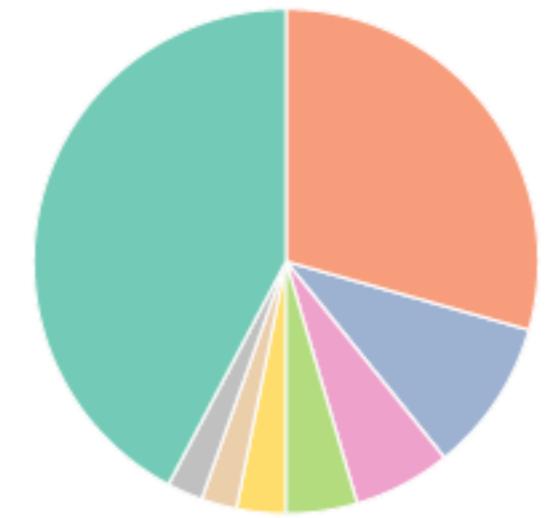
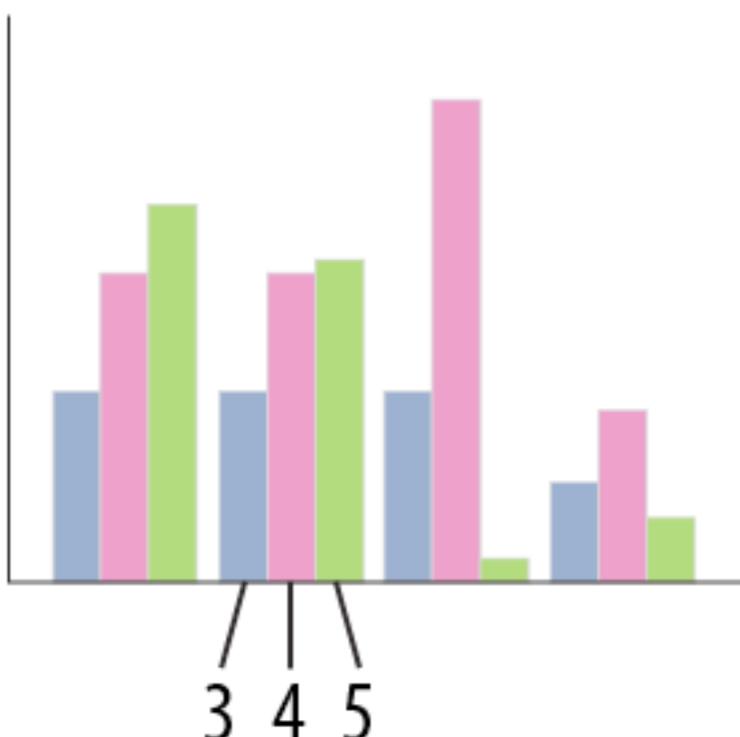


*murky*



*recolored with Brewer palettes*

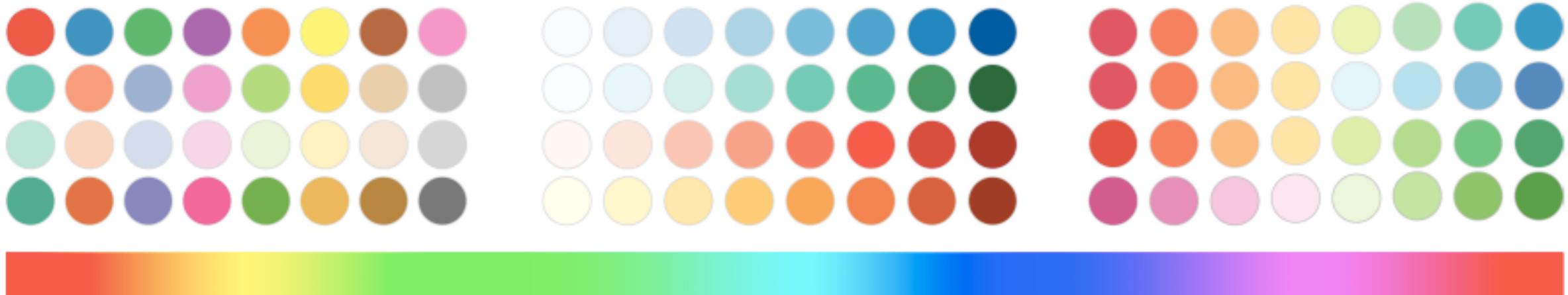
## Examples



# Conversion to Grey scale

---

- Ensure chosen colour set works well in grey scale
  - Sequential palette works well here

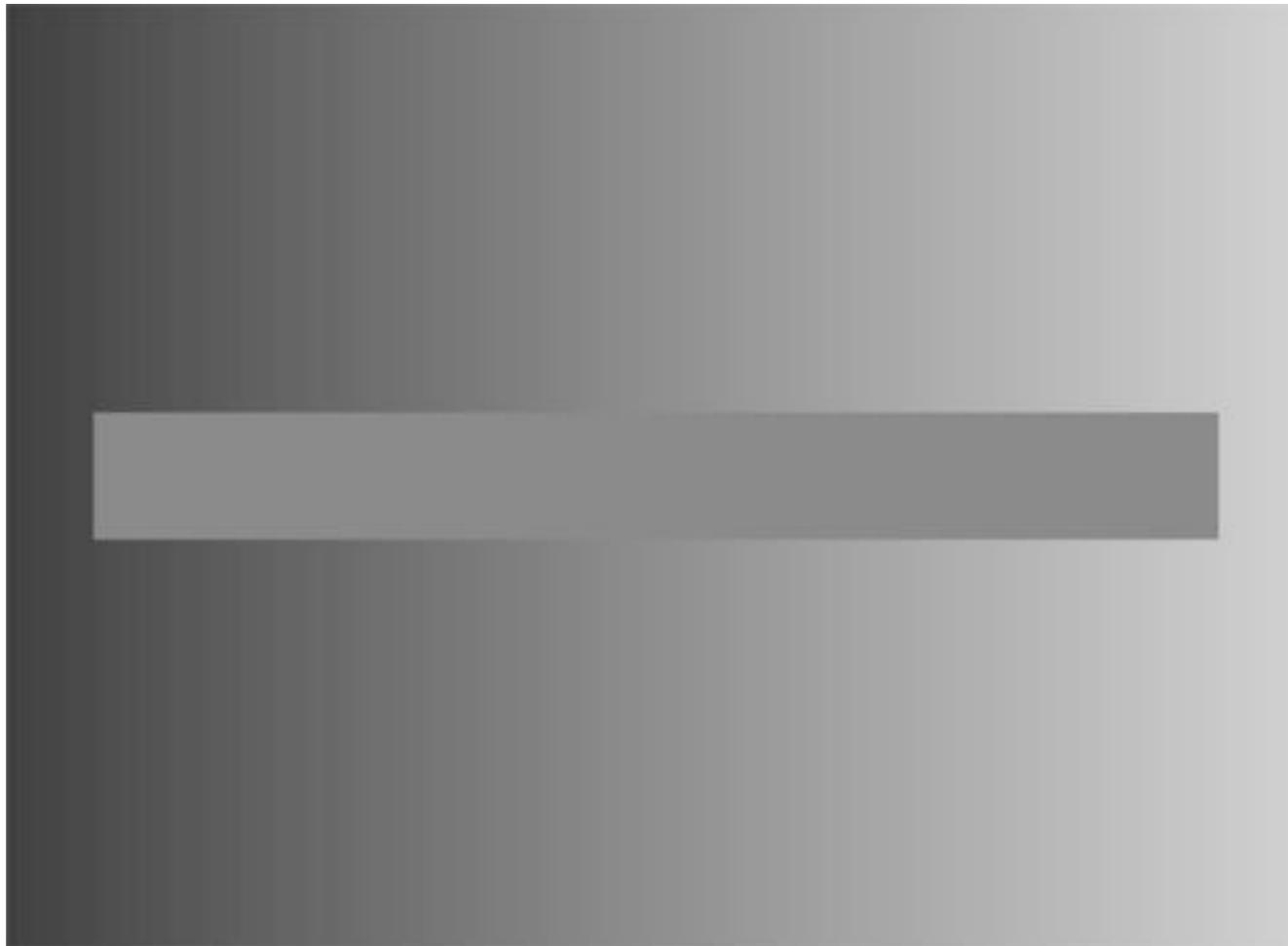


HSB DESATURATION



# Trouble with perceptual colour....

---



# Context Affects Perceived Colour

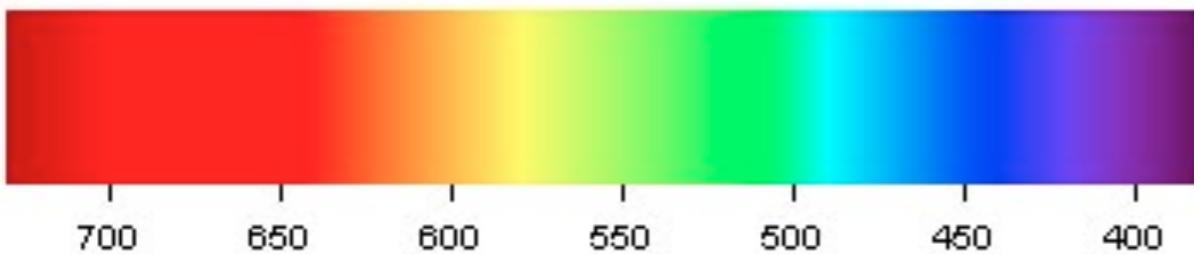
---



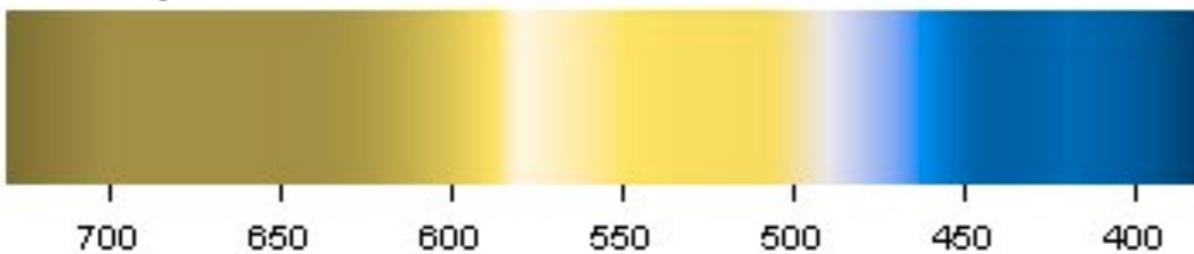
# Colour & Accessibility

---

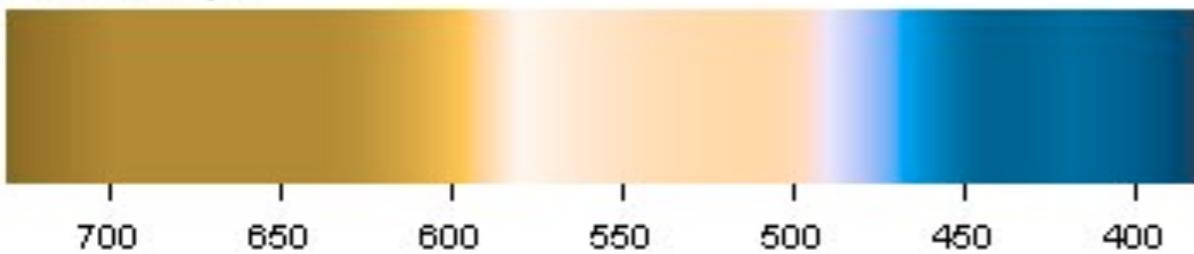
Normal



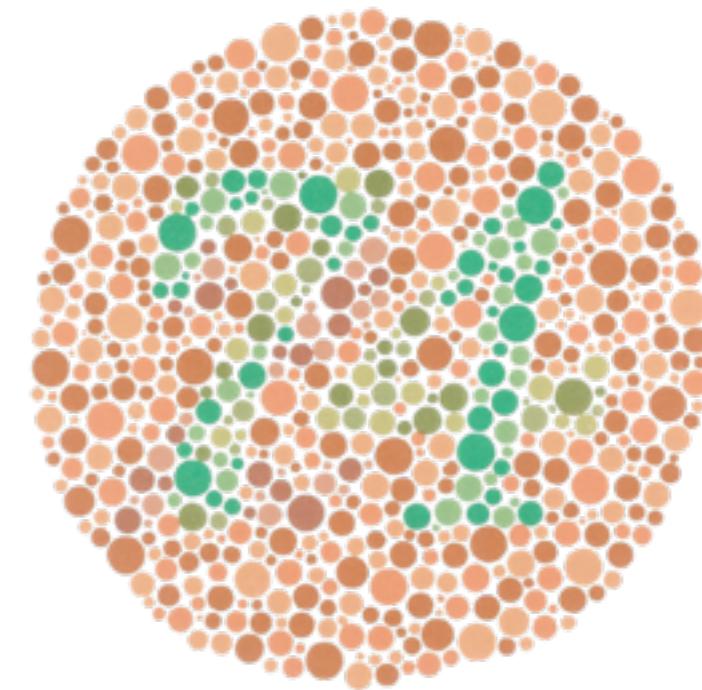
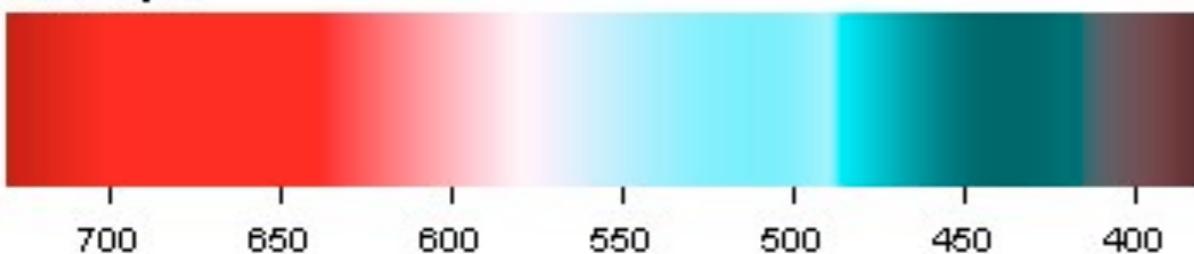
Protanopia



Deutanopia



Tritanopia



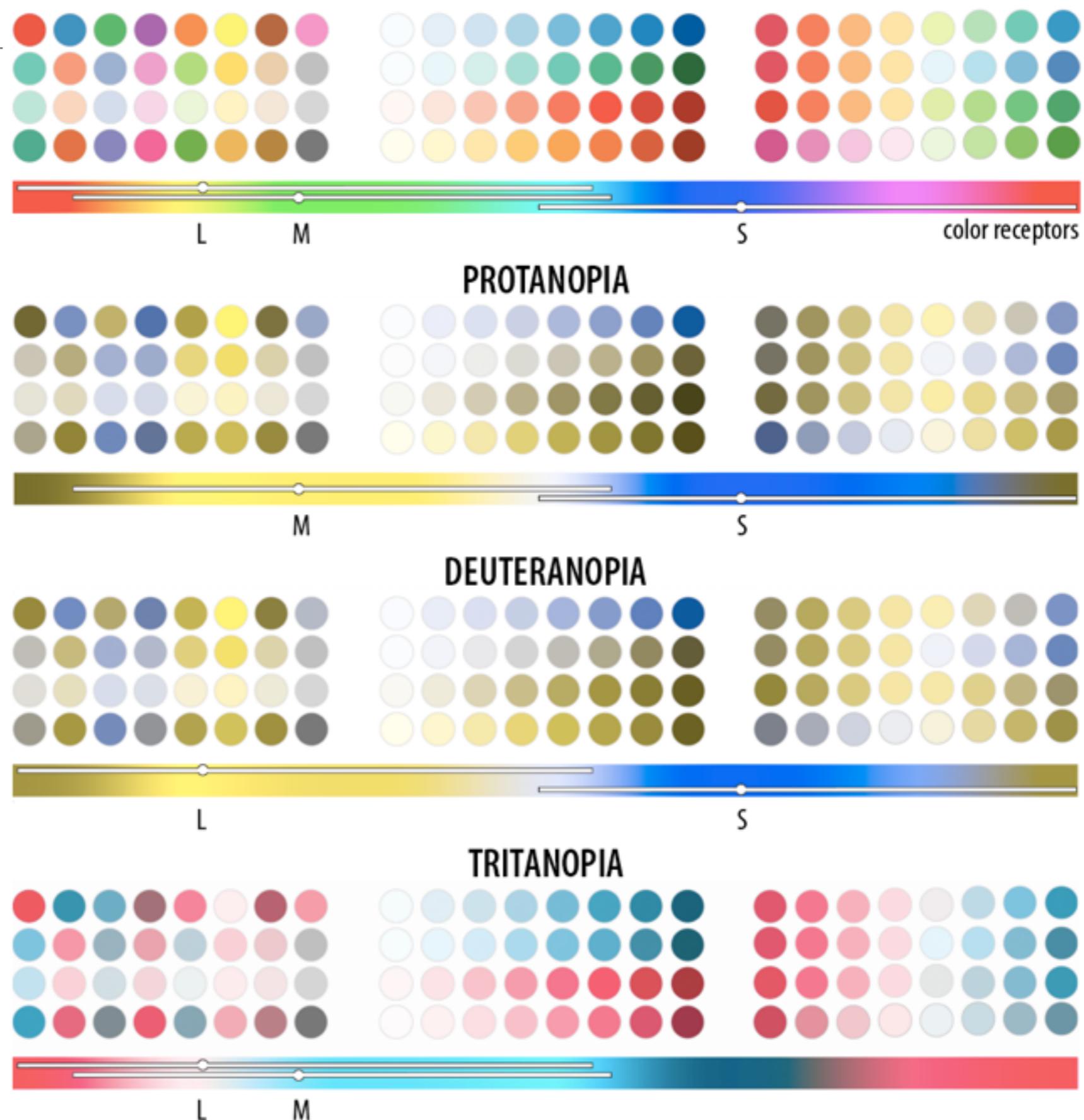
Accessibility (W3C):  
10-20% of population  
are red/green colour  
blind.

# Color blindness

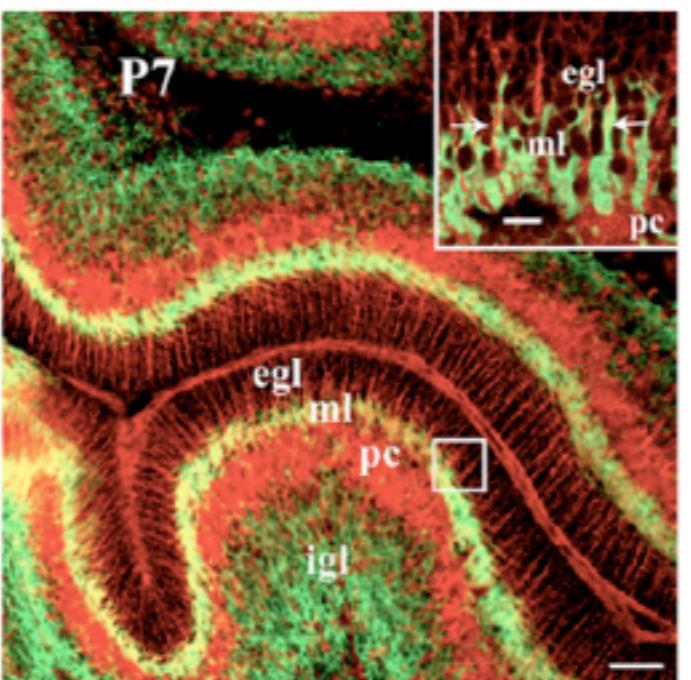
red-green

red-green

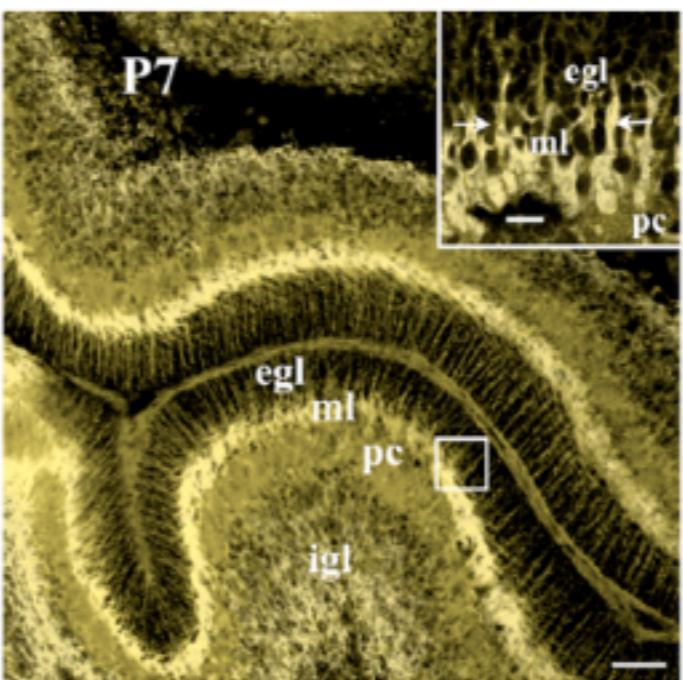
blue-yellow



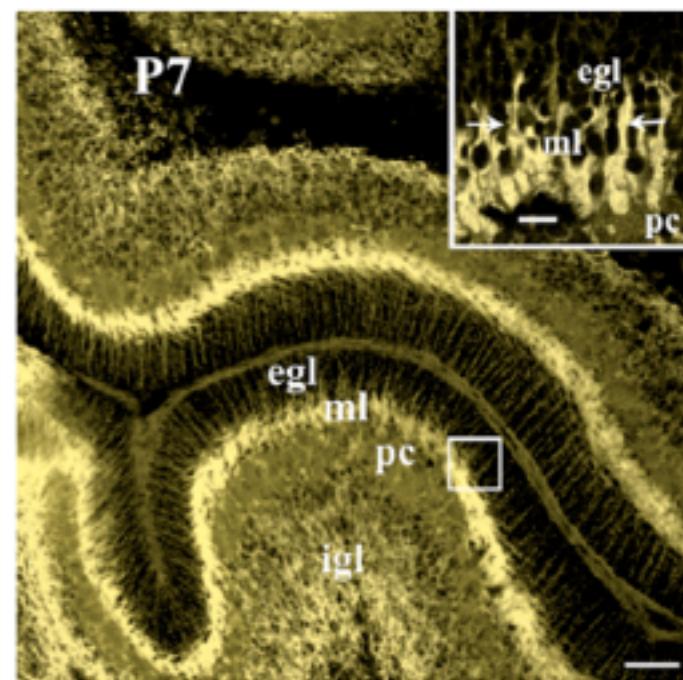
## NORMAL VISION red-green palette



## DEUTERANOPIA



## PROTANOPPIA

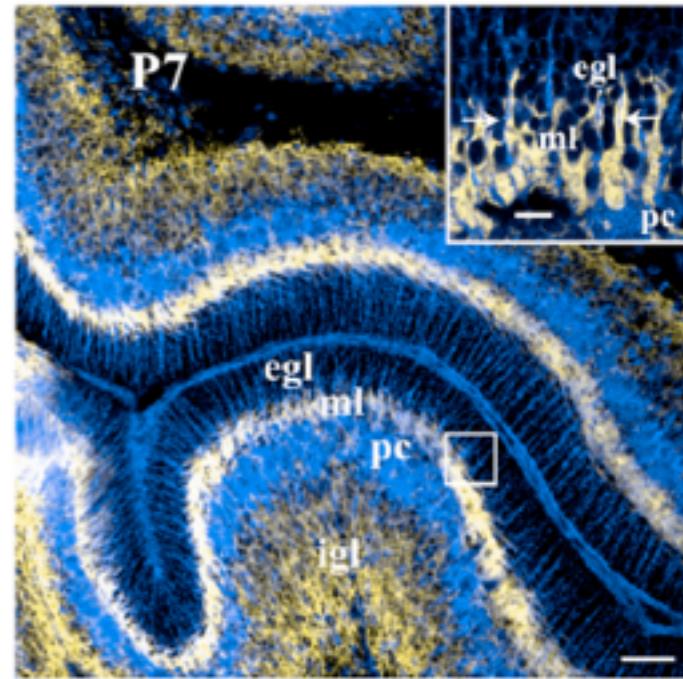
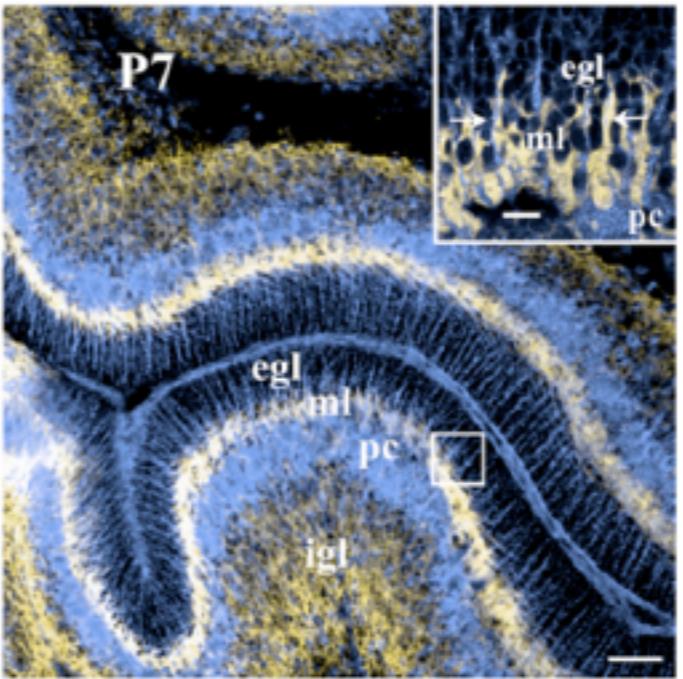
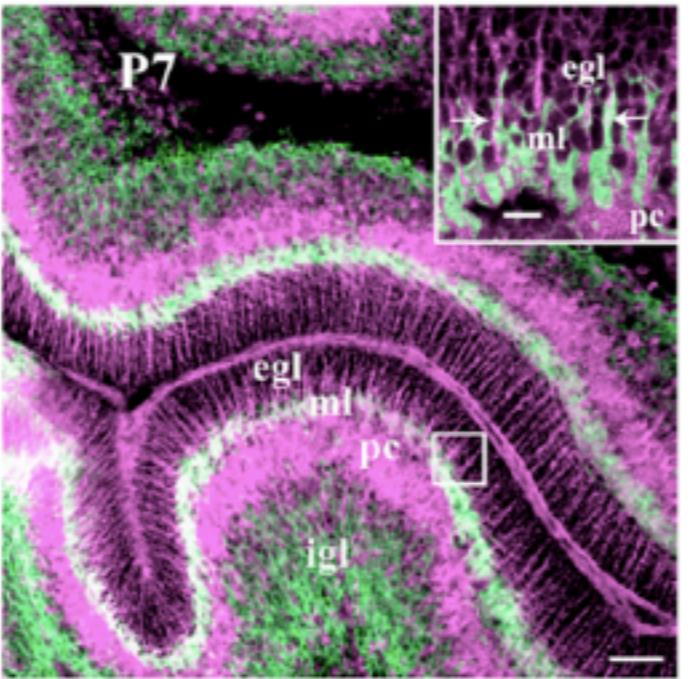


BioVis Example:  
Immunofluorescence  
images

red-green image of  
P2Y1 receptor and  
migrating granule  
neurons,

effectively remapped  
to  
magenta-green using  
the channel mixing  
method.

## magenta-green palette



# 15-COLOR PALETTE FOR COLOR BLINDNESS

NORMAL VISION

DEUTERANOPIA

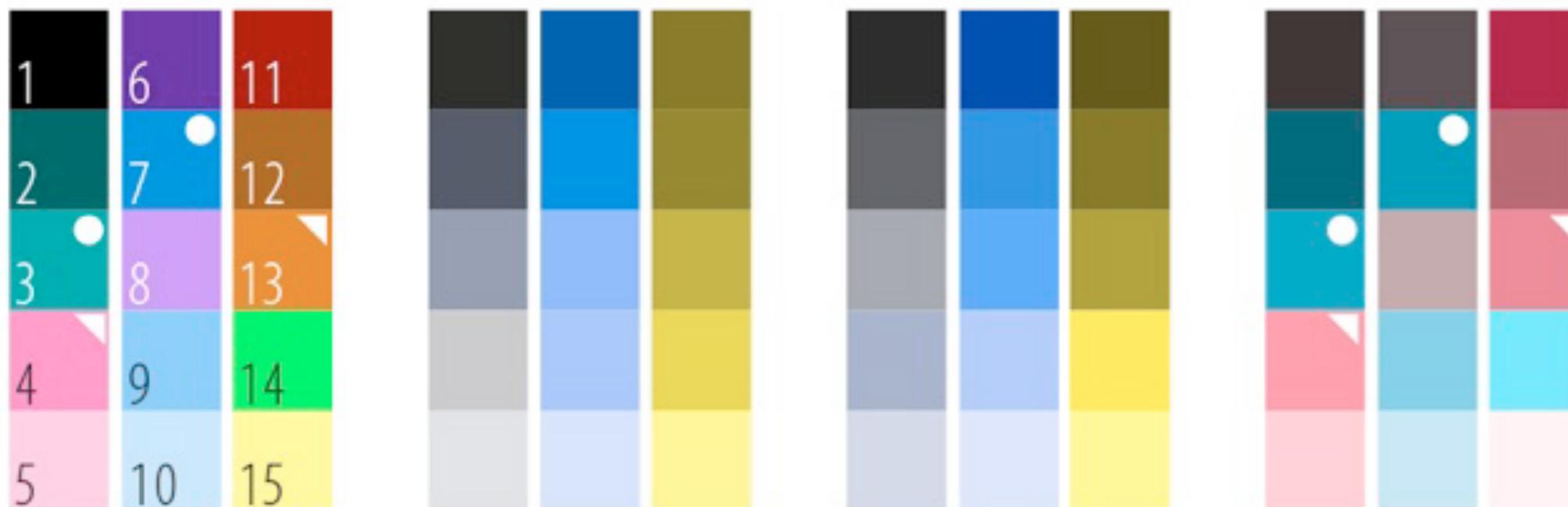
common (6%)

PROTANOPIA

rare (2%)

TRITANOPIA

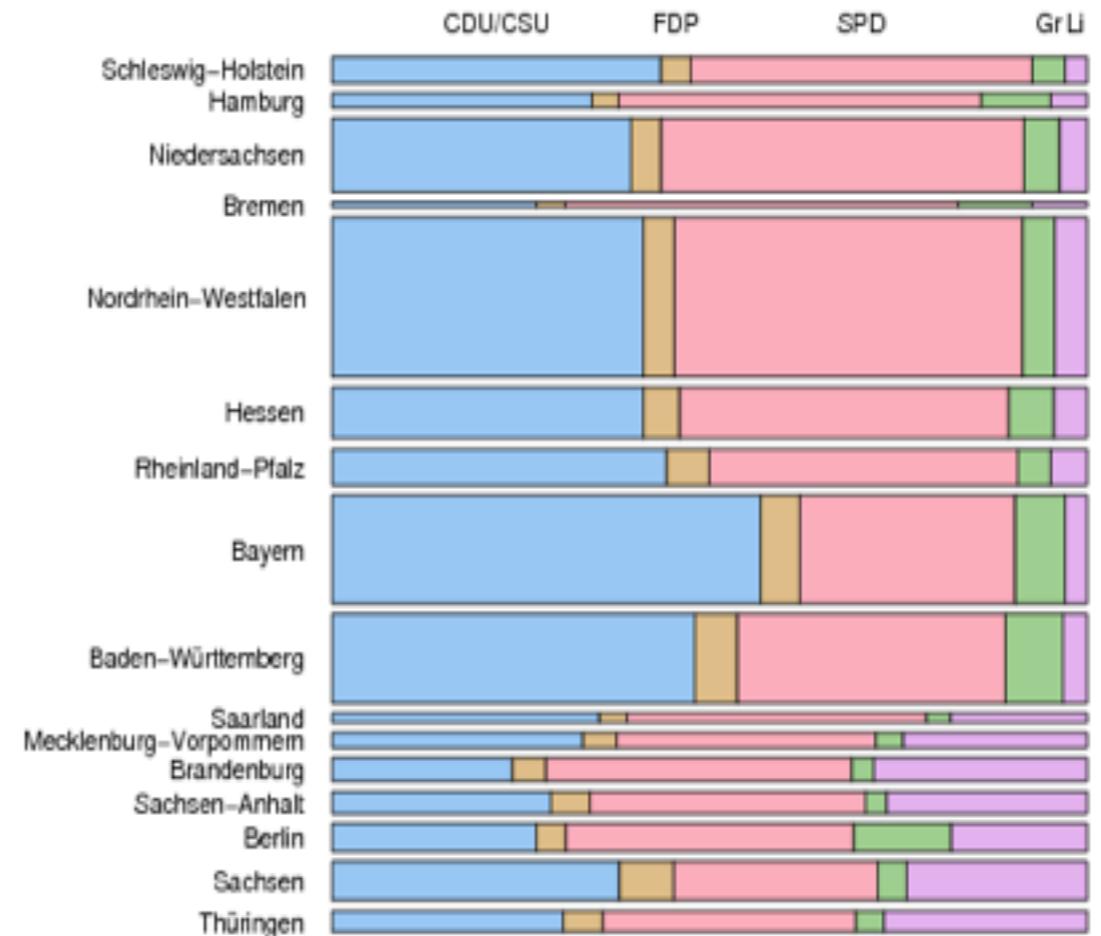
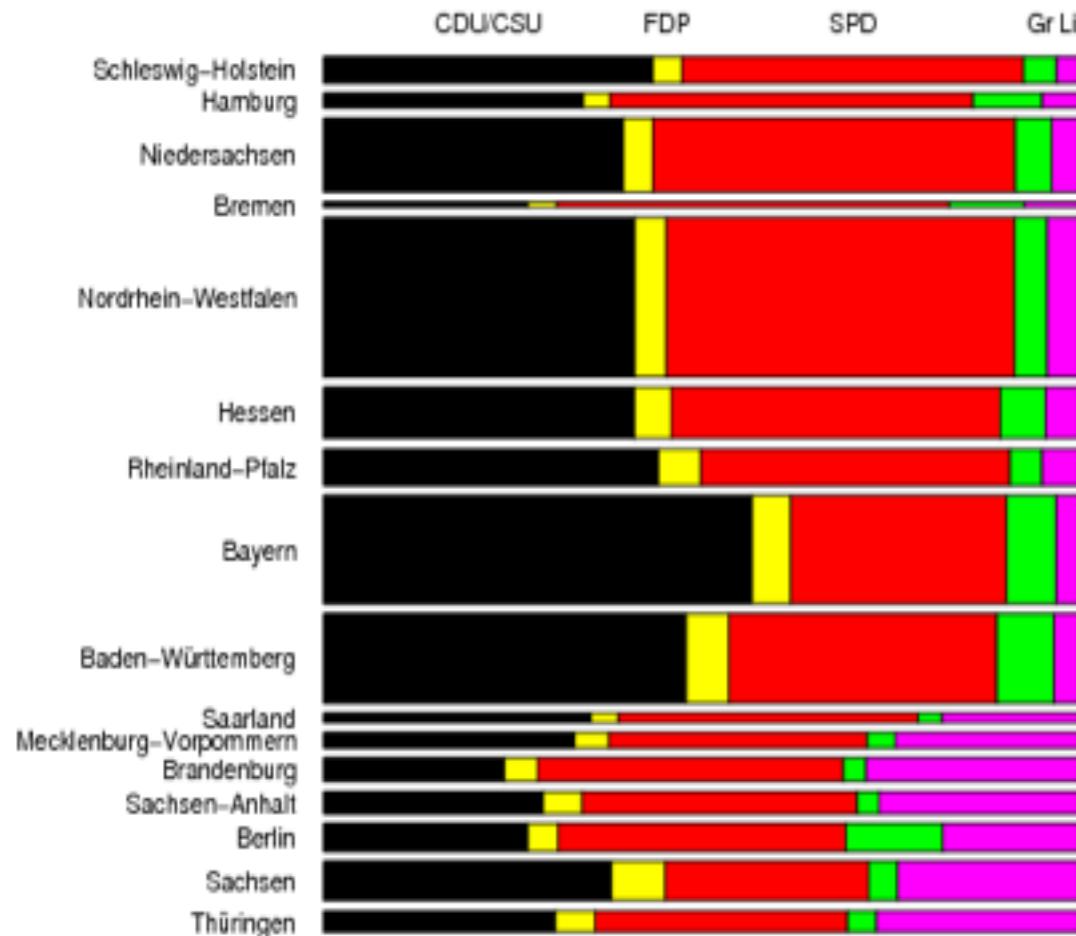
very rare (<1%)



R   G   B

1	0	0	0	6	73	0	146	11	146	0	0	
2	0	73	73	7	●	0	109	219	12	146	73	0
3	●	0	146	146	8	182	109	255	13	219	209	0
4	■	255	109	182	9	109	182	255	14	36	255	36
5	255	182	119	10	182	219	255	15	255	255	109	

# Biased and unbiased politics



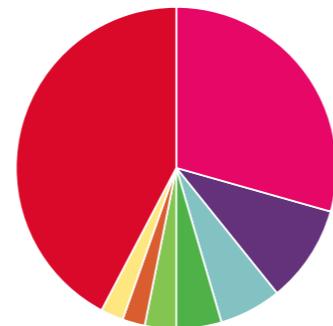
From A. Zeileis, Reisensburg 2007

# USE BREWER PALETTES

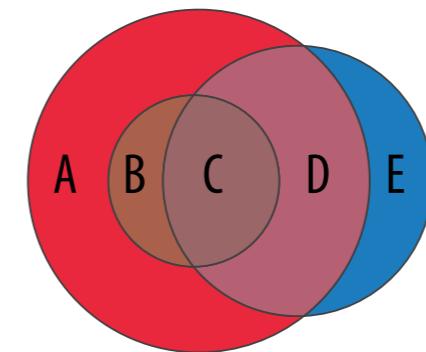
*one color  
dominates*



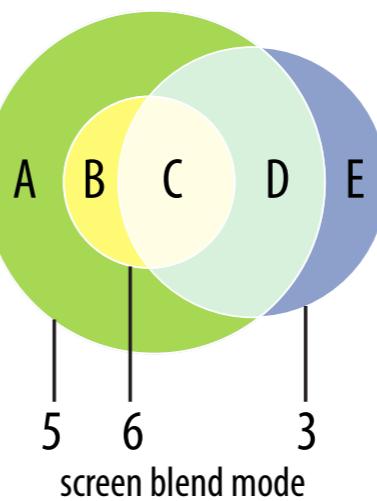
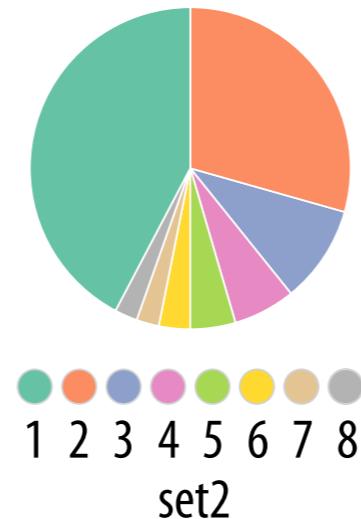
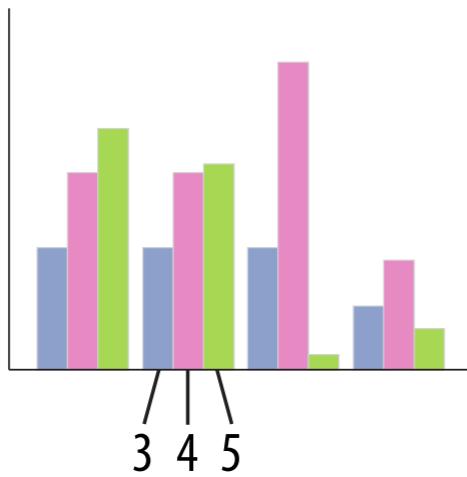
*difficult to  
distinguish*



*murky*



*recolored with Brewer palettes*



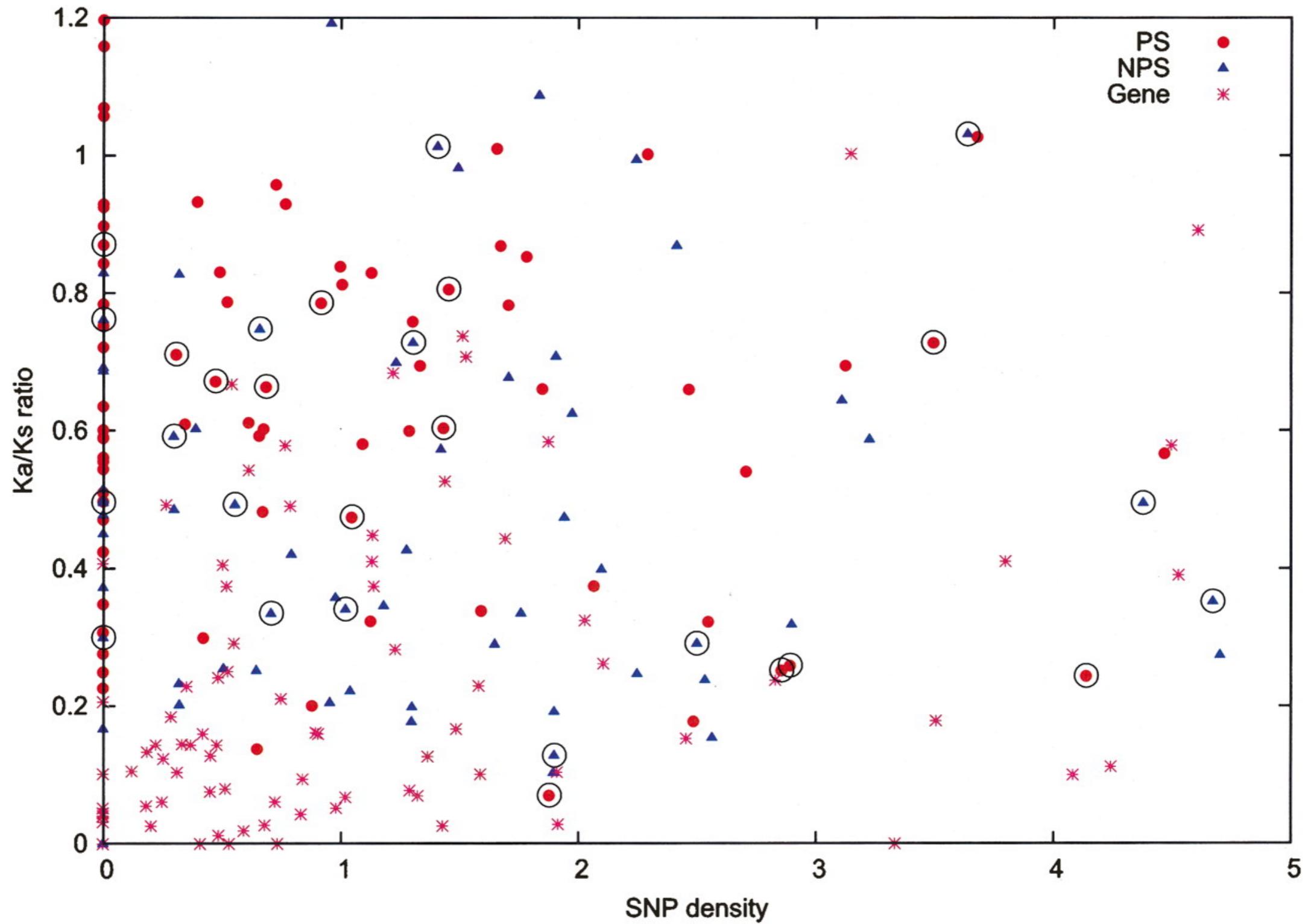
Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, et al. (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Res* 13: 2725–2735.

Bono H, Yagi K, Kasukawa T, Nikaido I, Tominaga N, et al. (2003) Systematic expression profiling of the mouse transcriptome using RIKEN cDNA microarrays. *Genome Res* 13: 1318–1323.

Tenney AE, Wu JQ, Langton L, Klueh P, Quatrano R, et al. (2007) A tale of two templates: automatically resolving double traces has many applications, including efficient PCR-based elucidation of alternative splices. *Genome Res* 17: 212–218.

# HIERARCHY AND PRIORITY

Use symbols that intuitively encode related concepts.



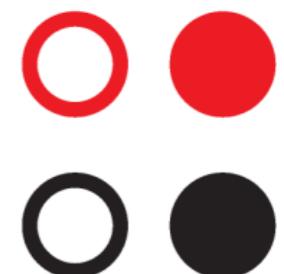
**PS**  
**NPS**  
**Gene**



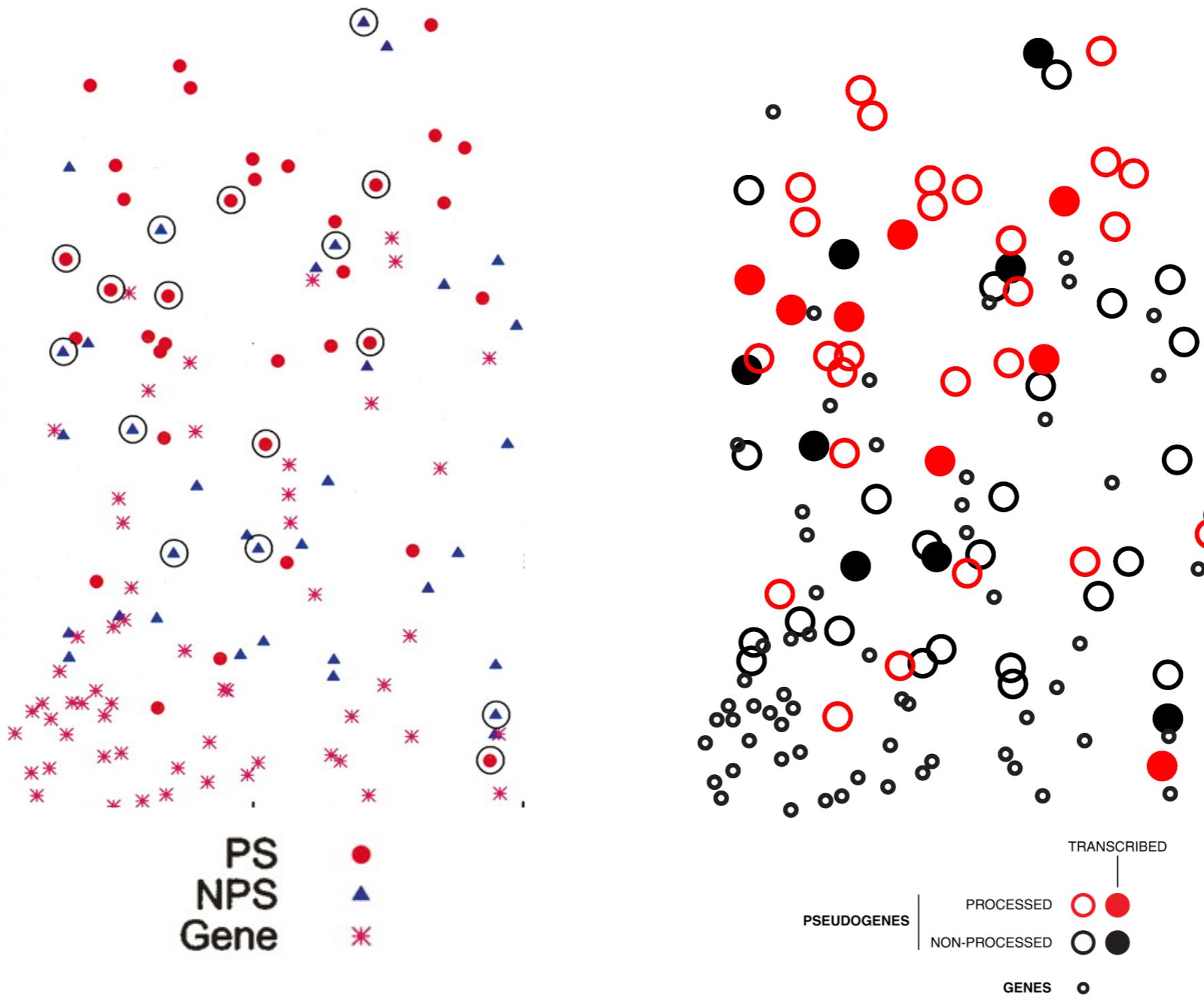
## PSEUDOGENES

PROCESSED  
NON-PROCESSED

TRANSCRIBED

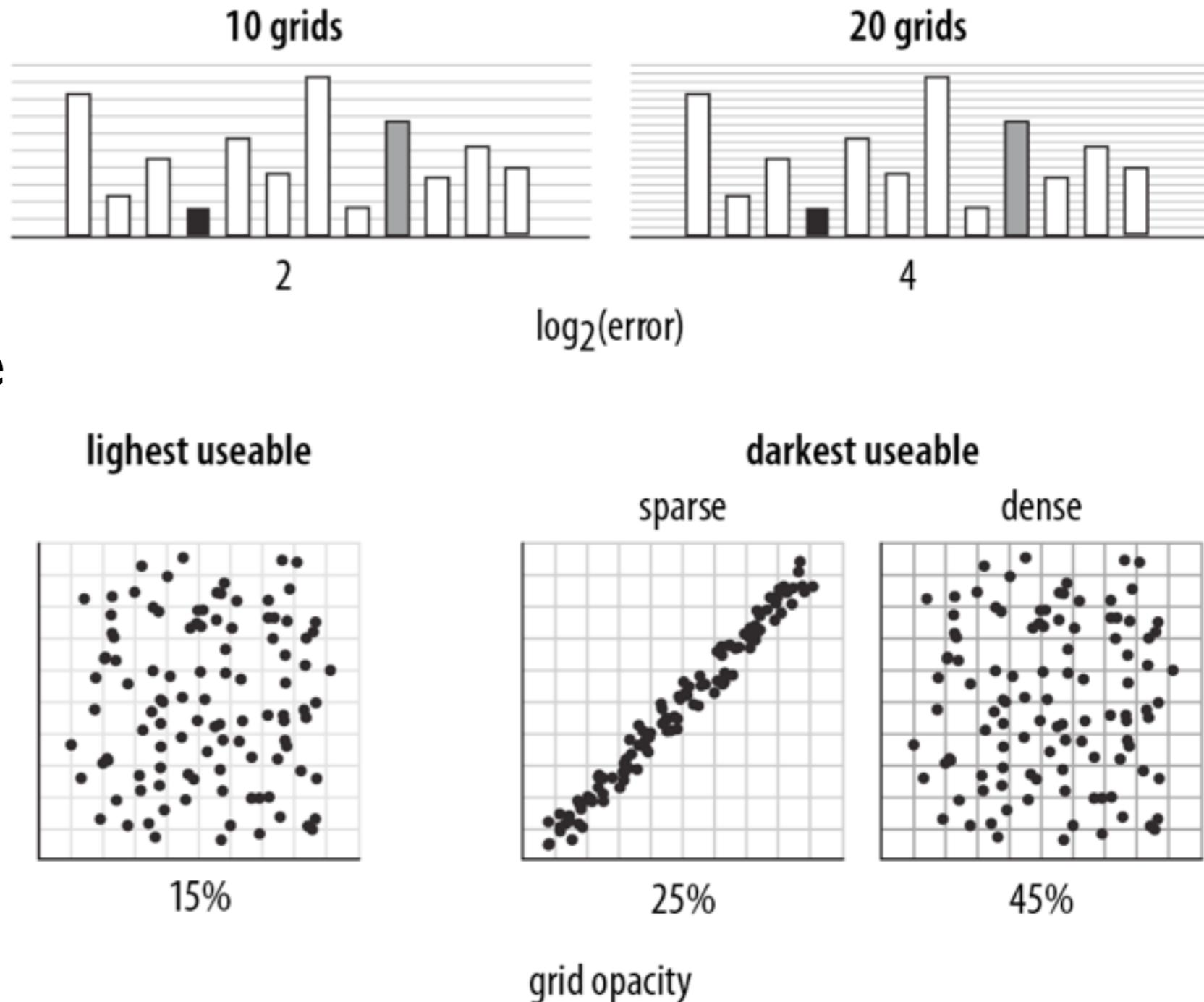


**GENES**      



# Increase data:ink ratio

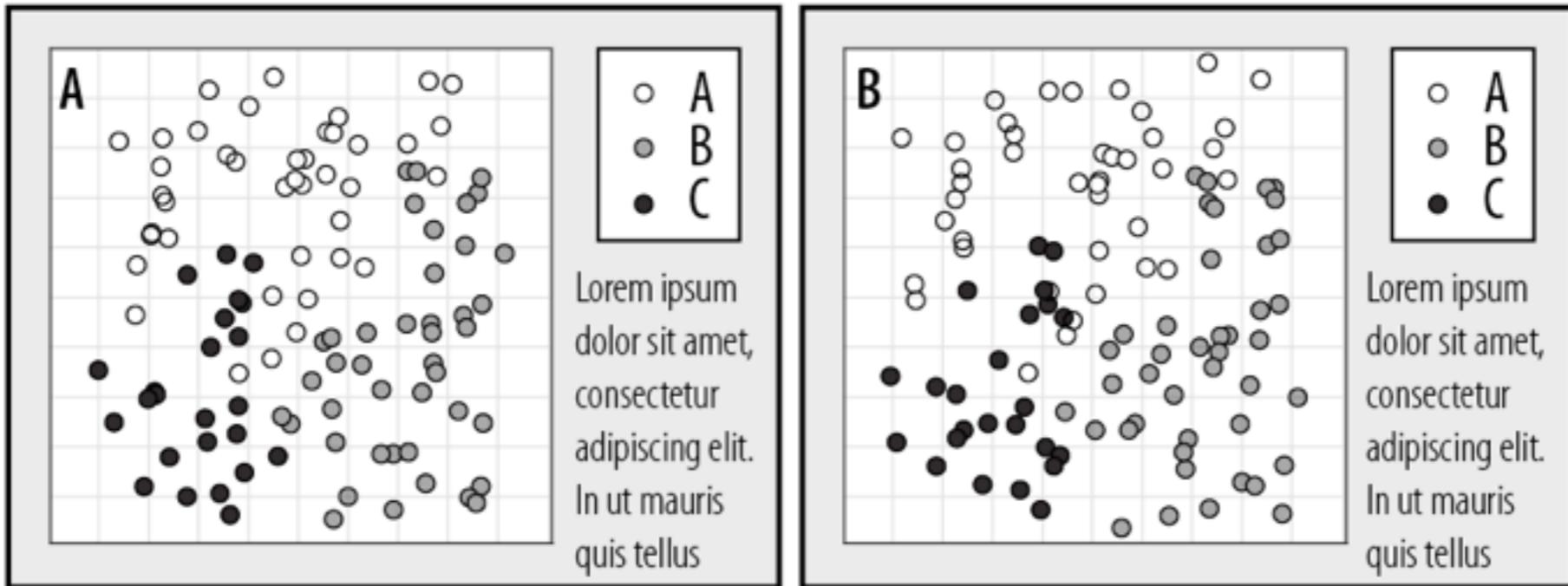
- Navigational aids
  - should not compete with the data for salience.
- Avoid
  - heavy axes
  - error bars and
  - glyphs



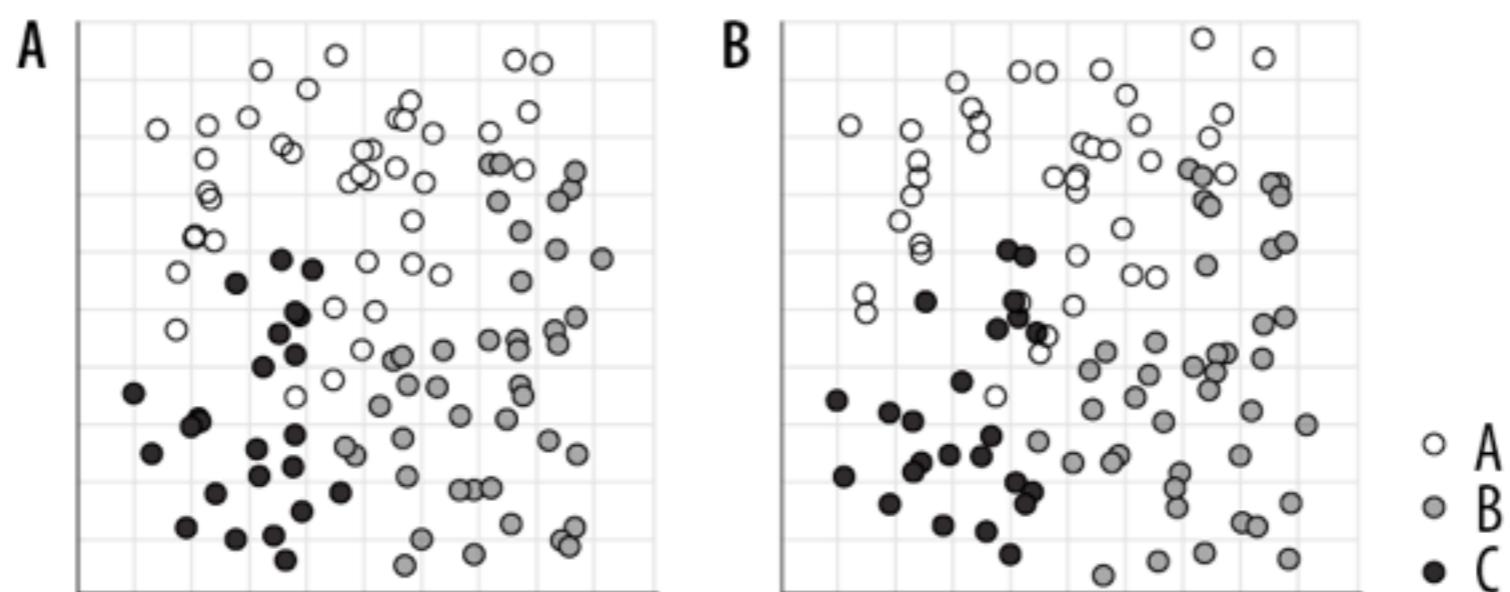
# Increase data:ink ratio

- Avoid unnecessary containment

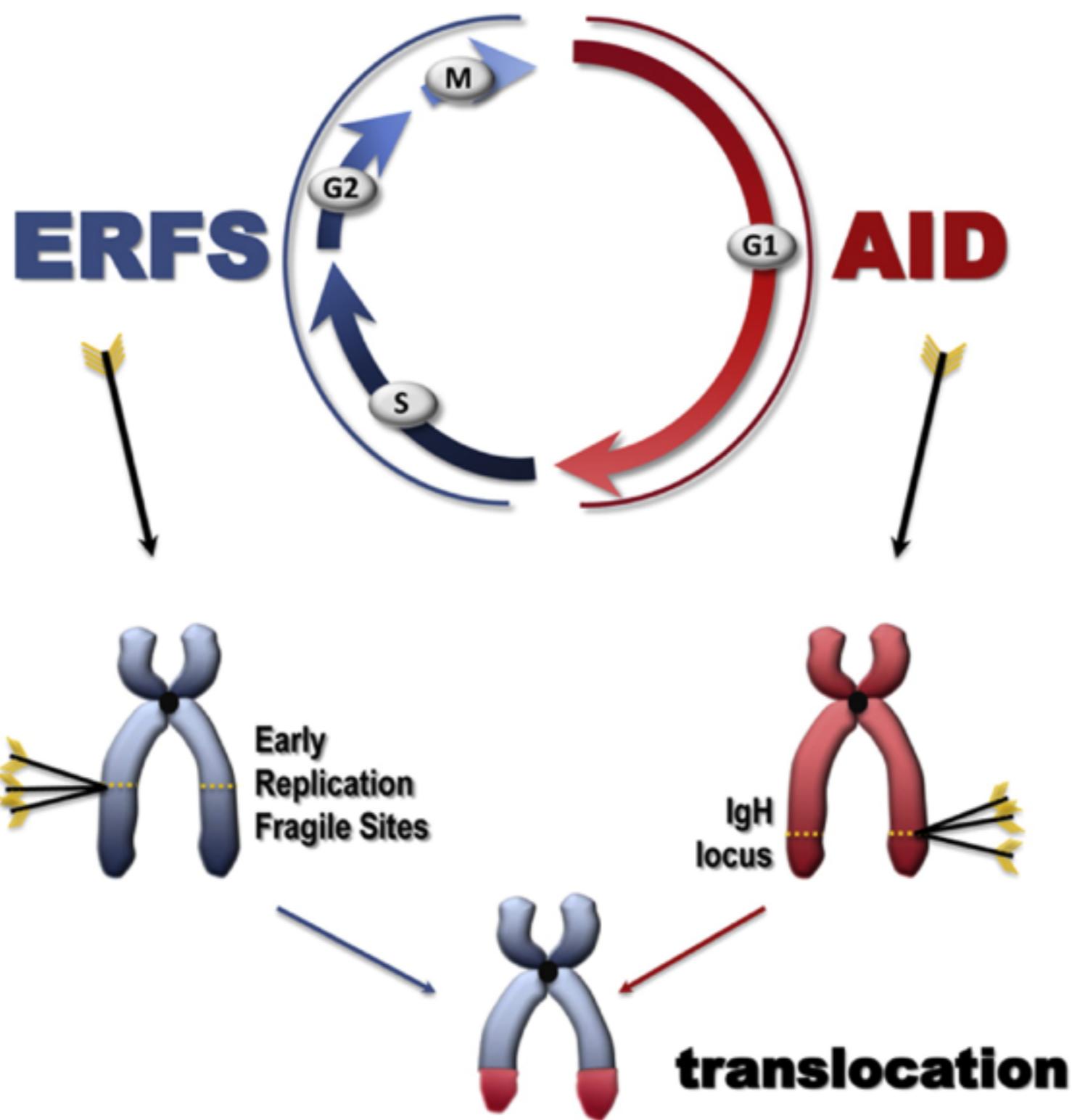
*confined*

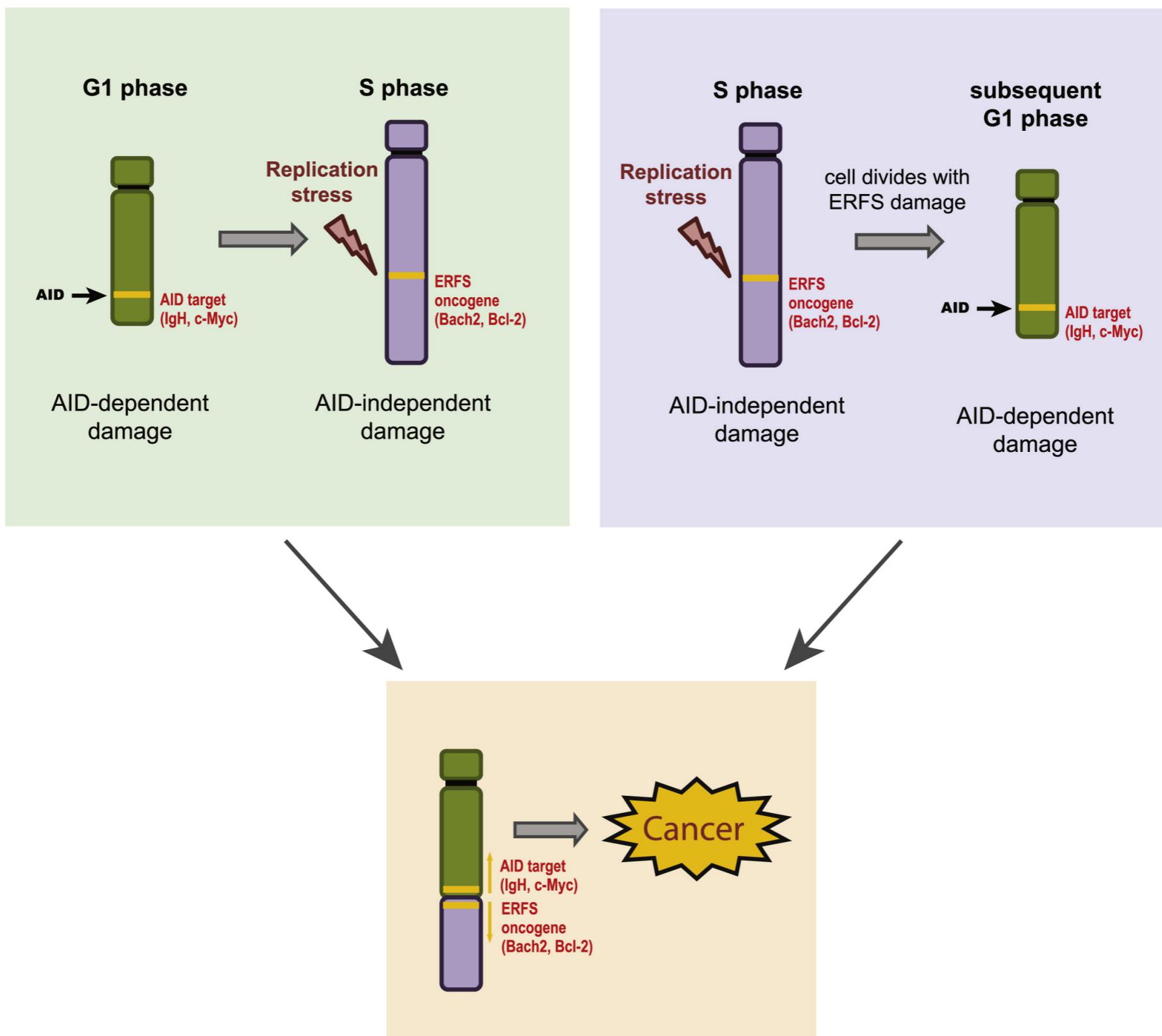


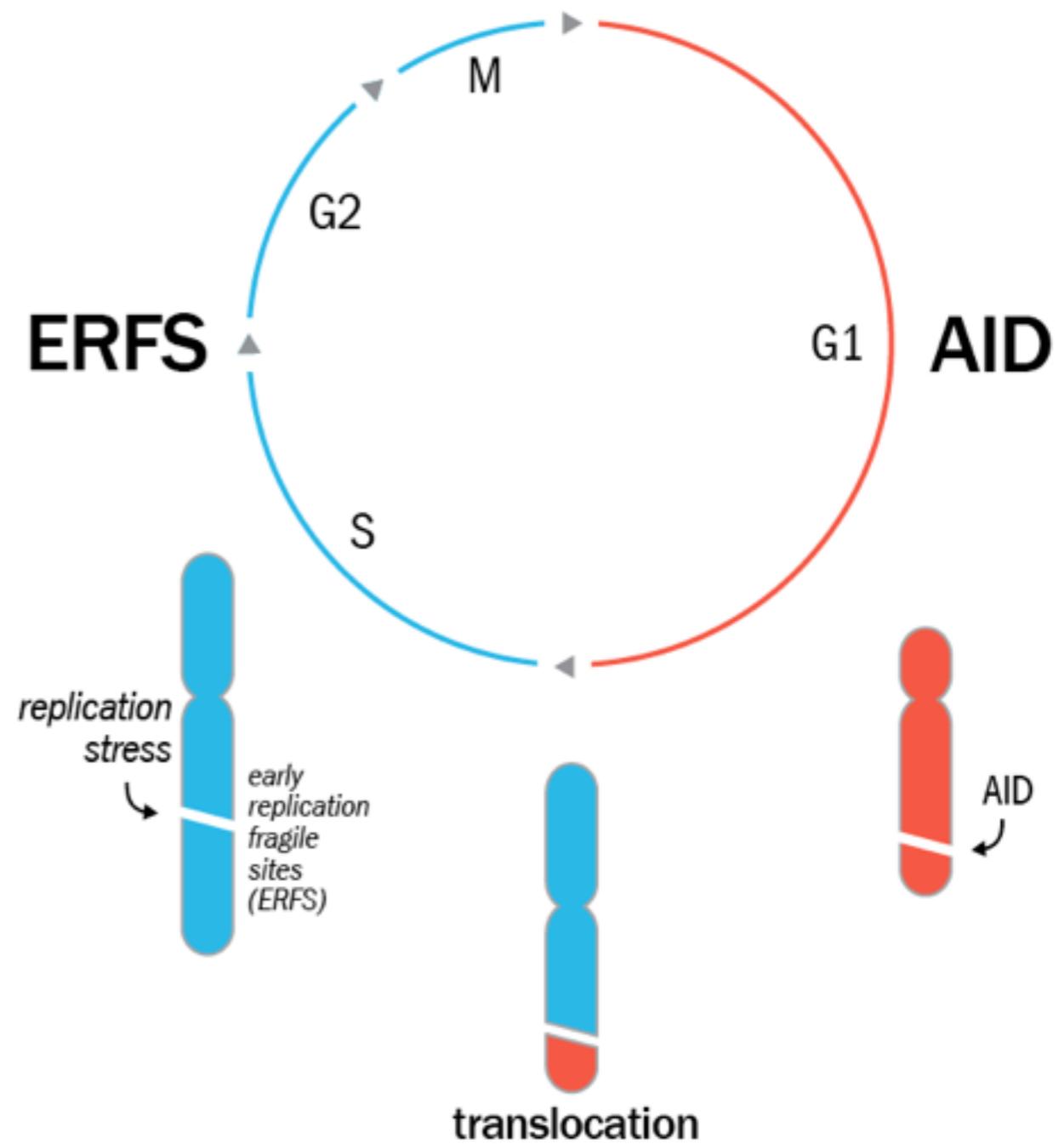
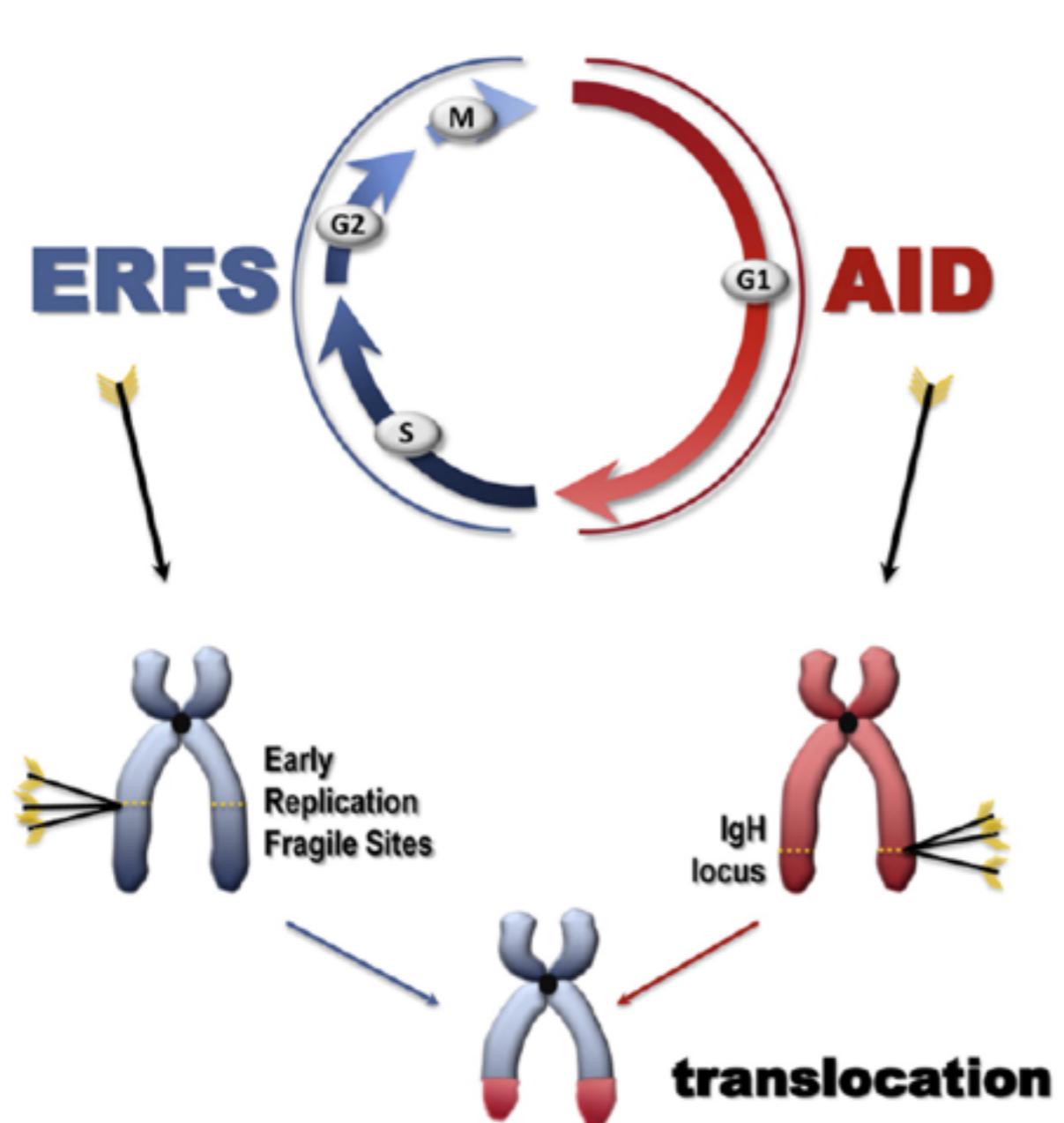
*improved*

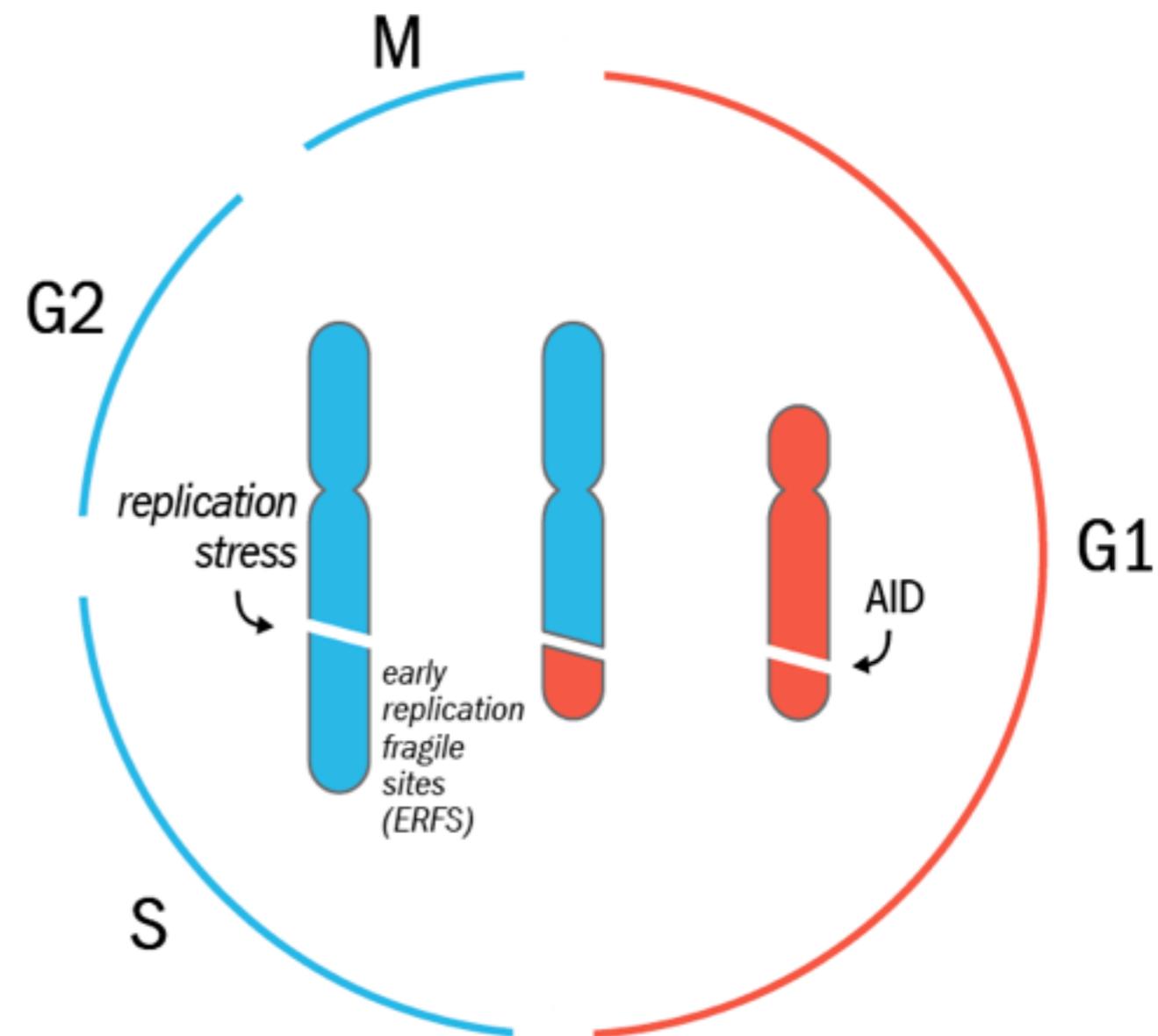
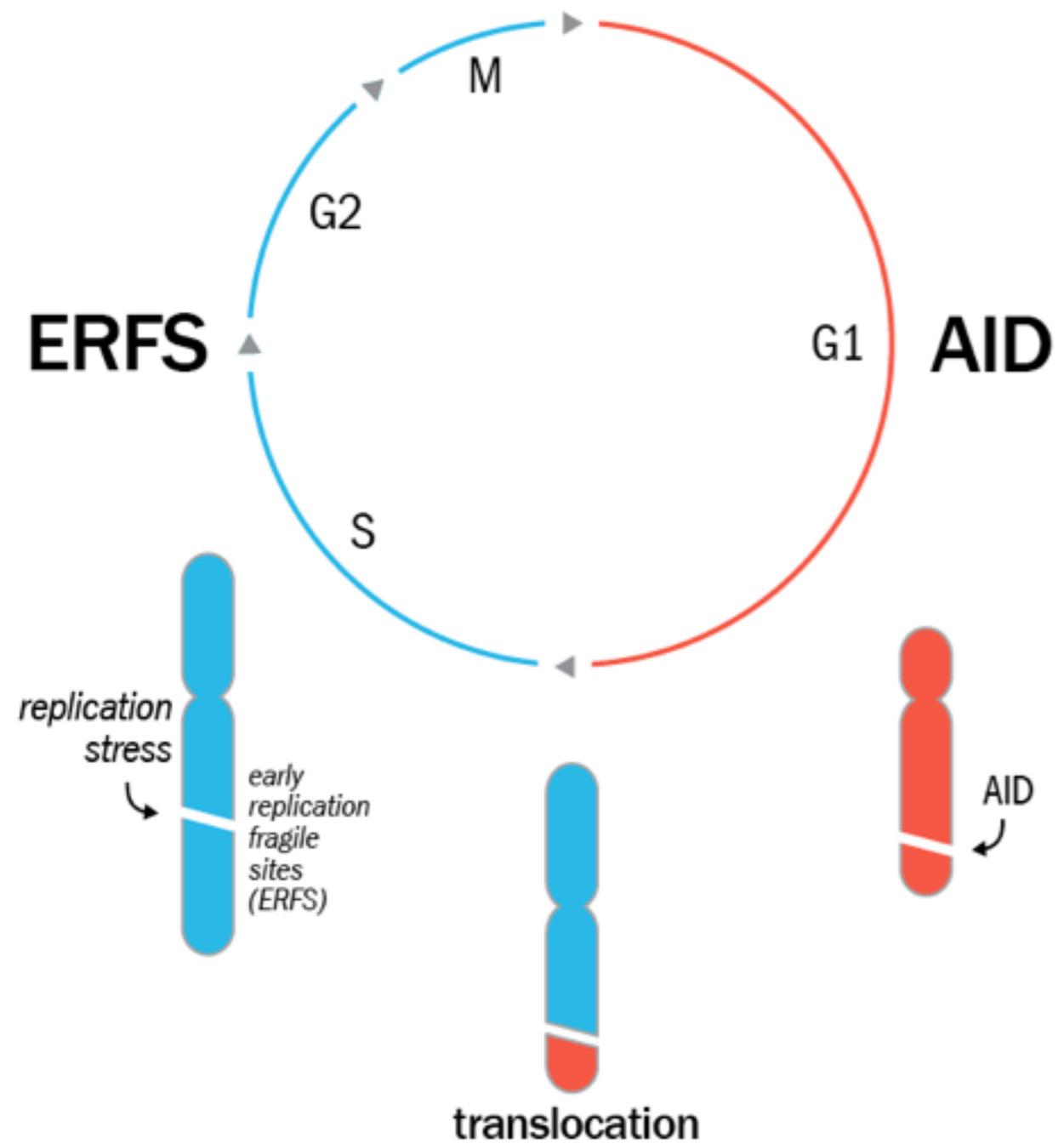


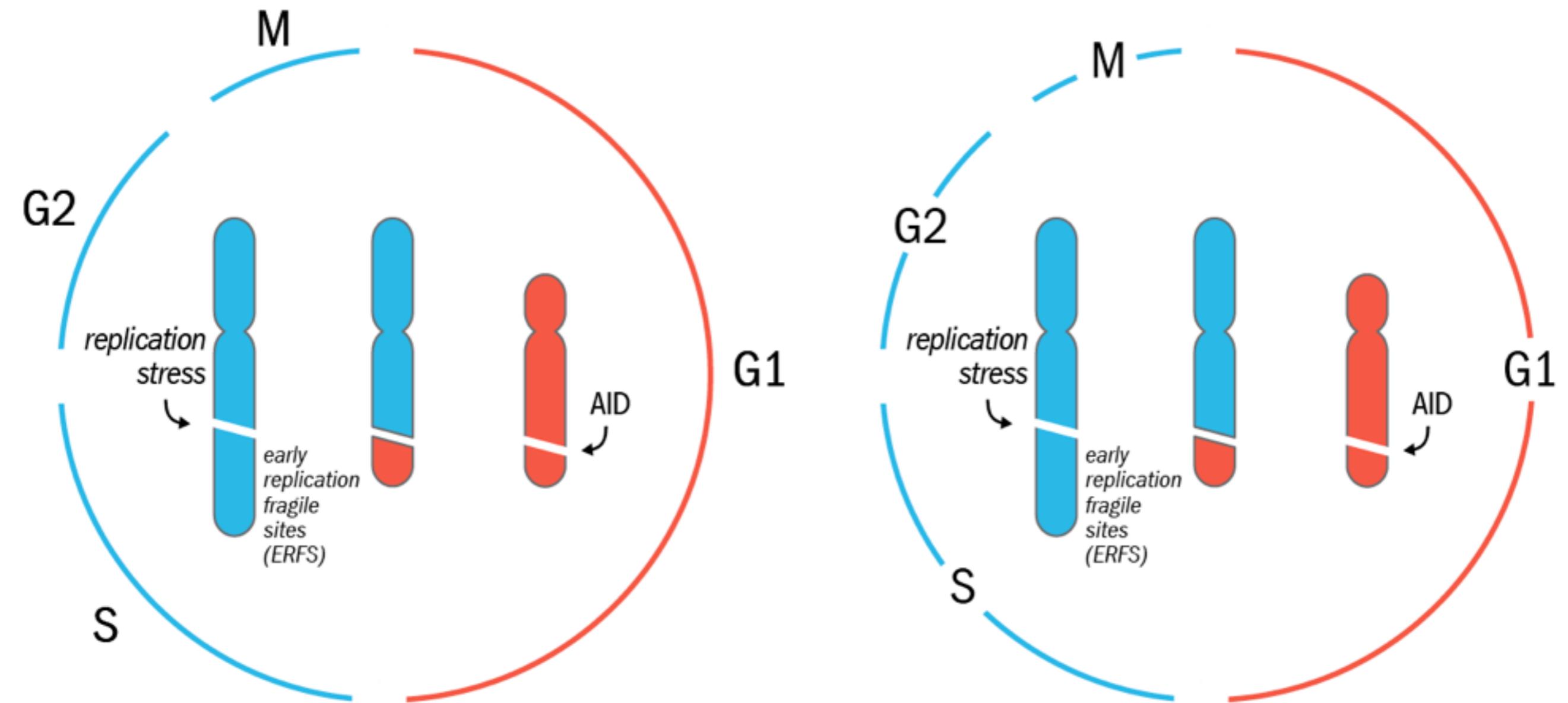
# VISUAL ABSTRACT REDESIGN

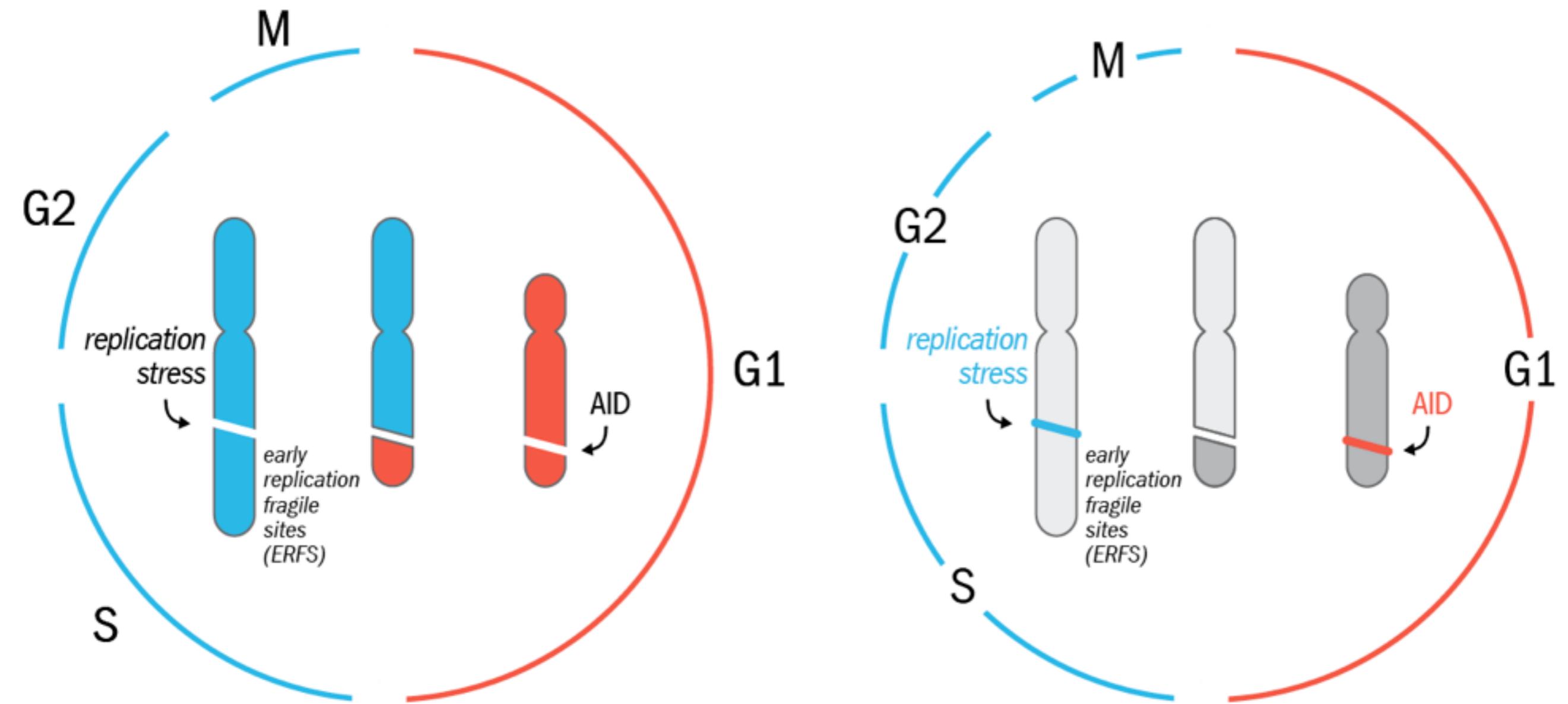


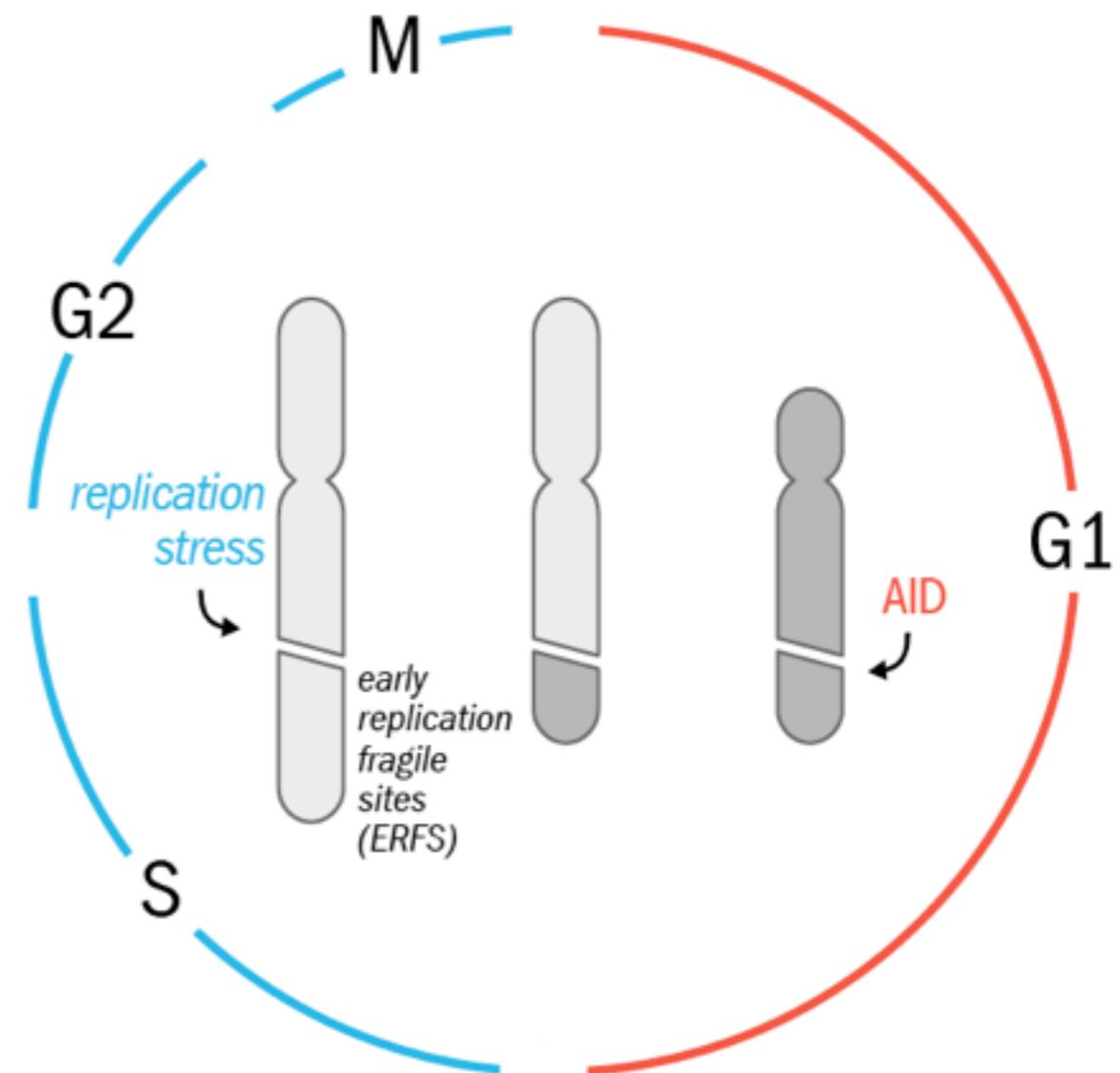
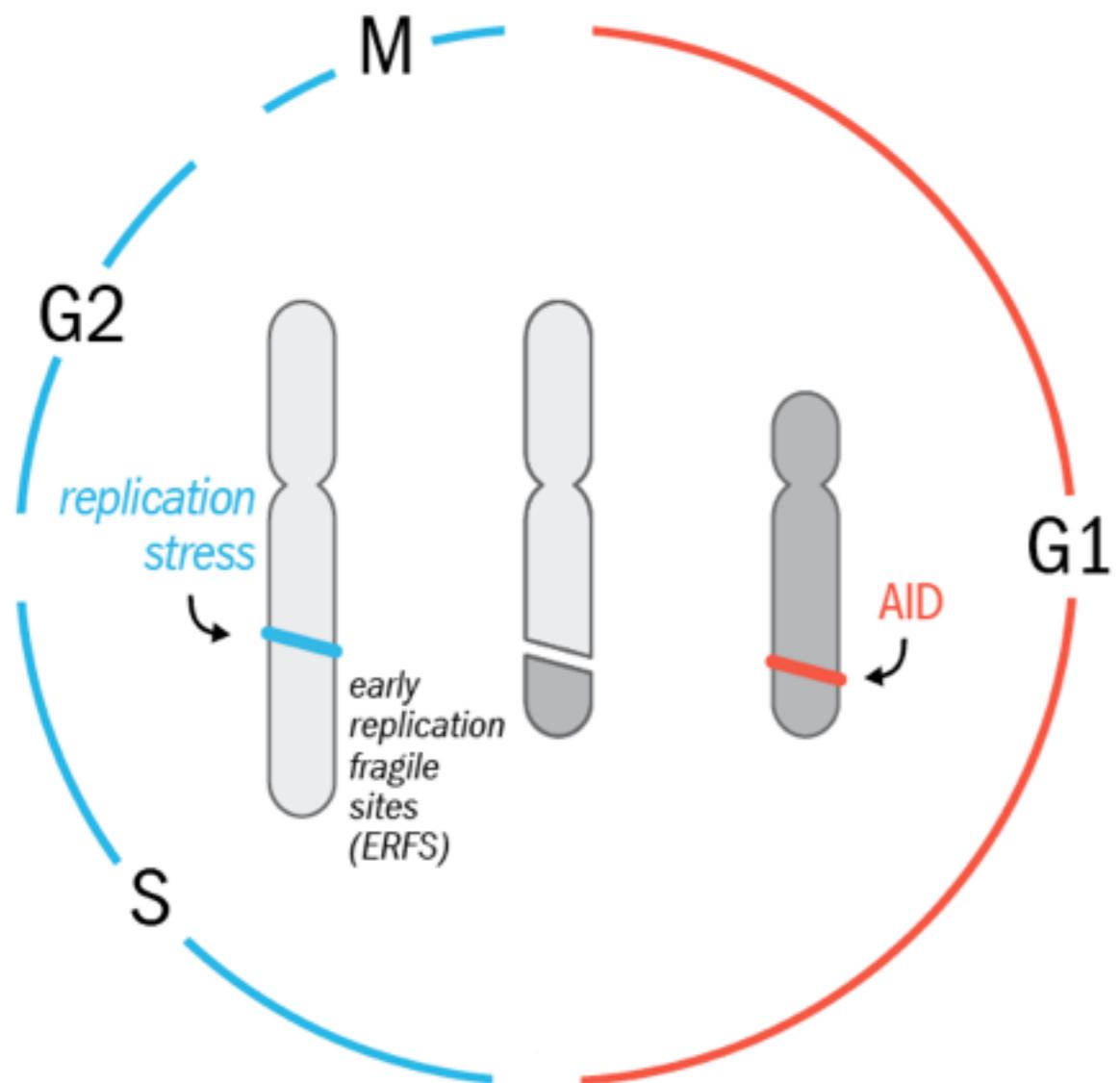


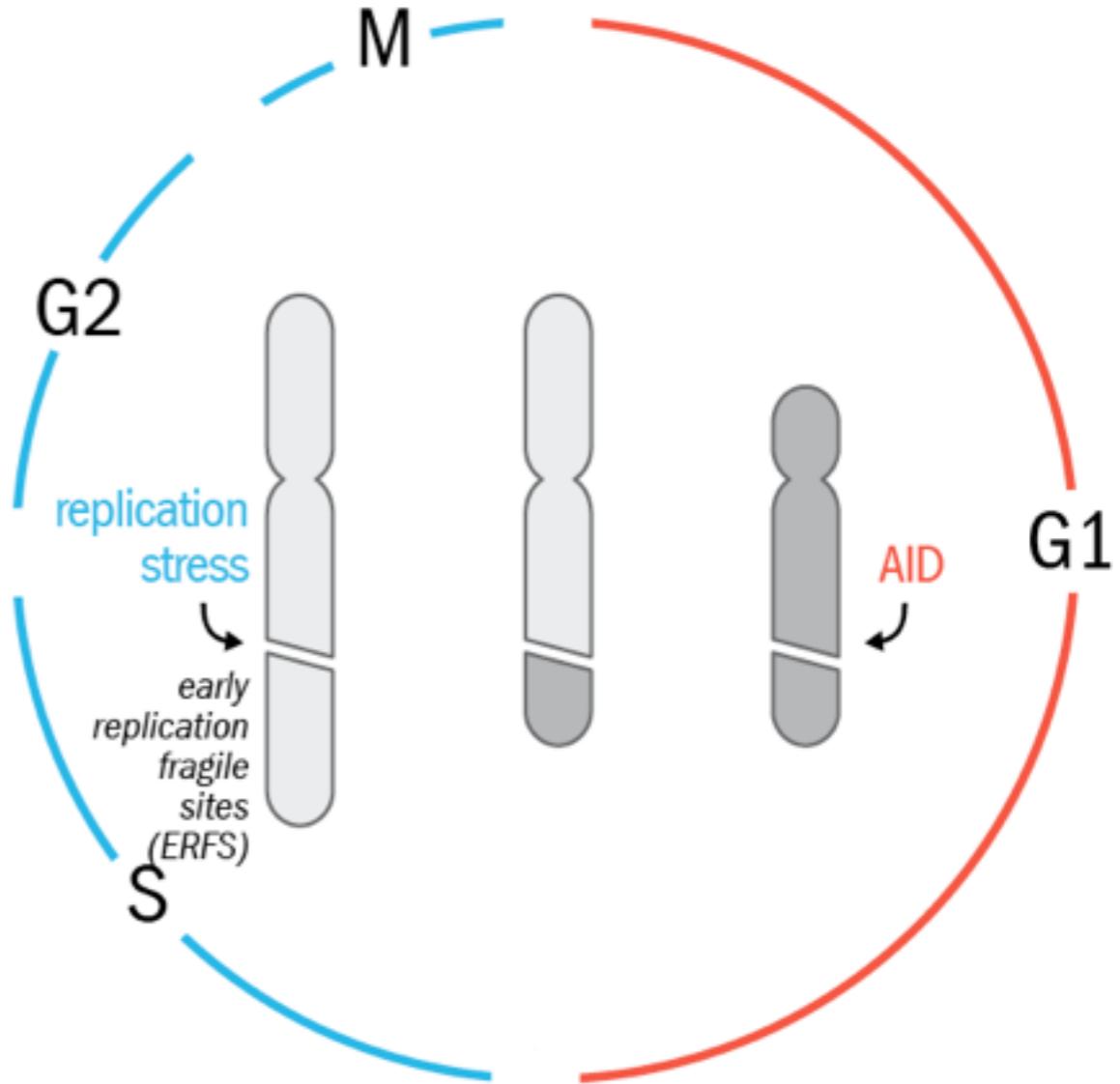
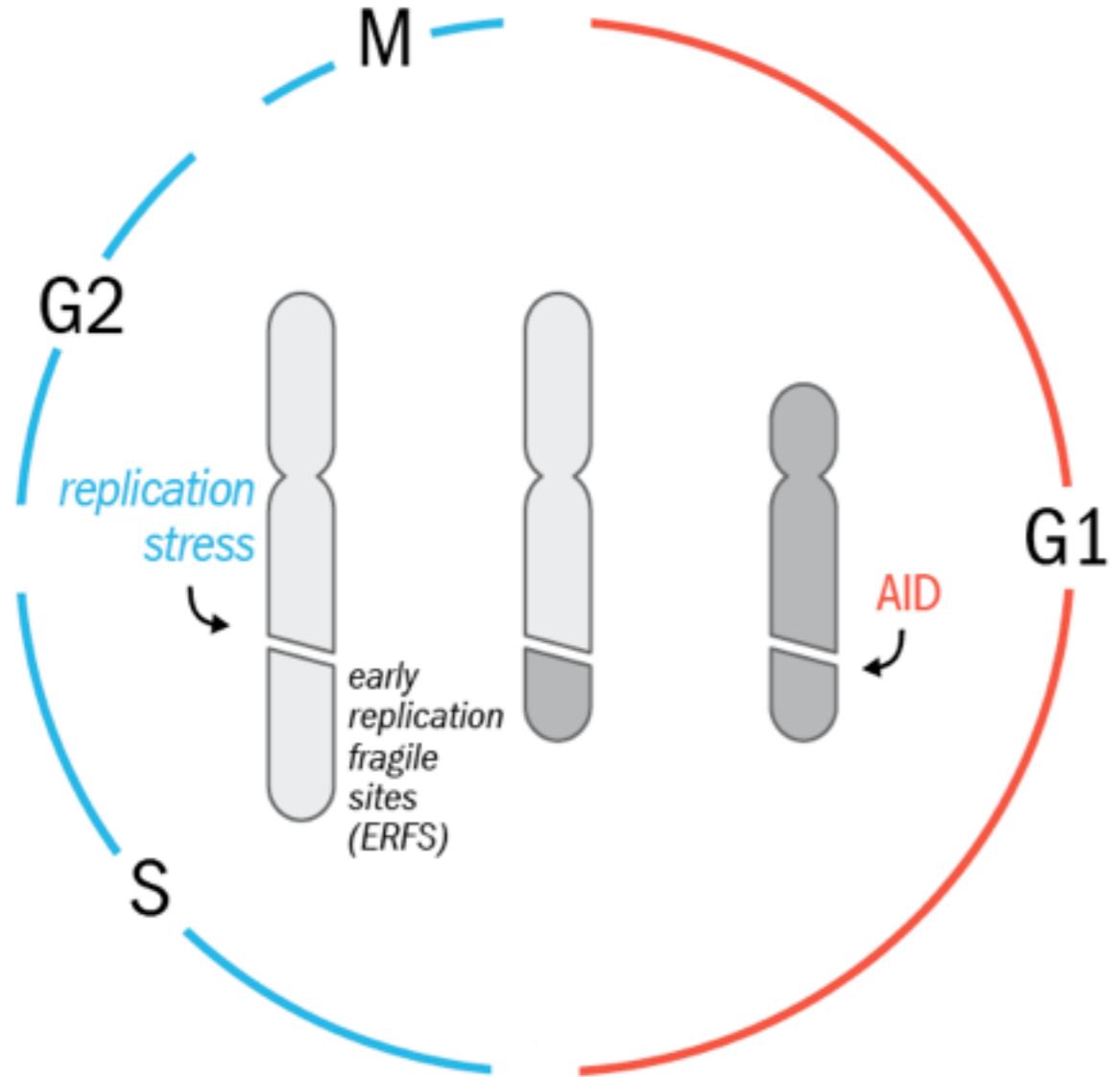


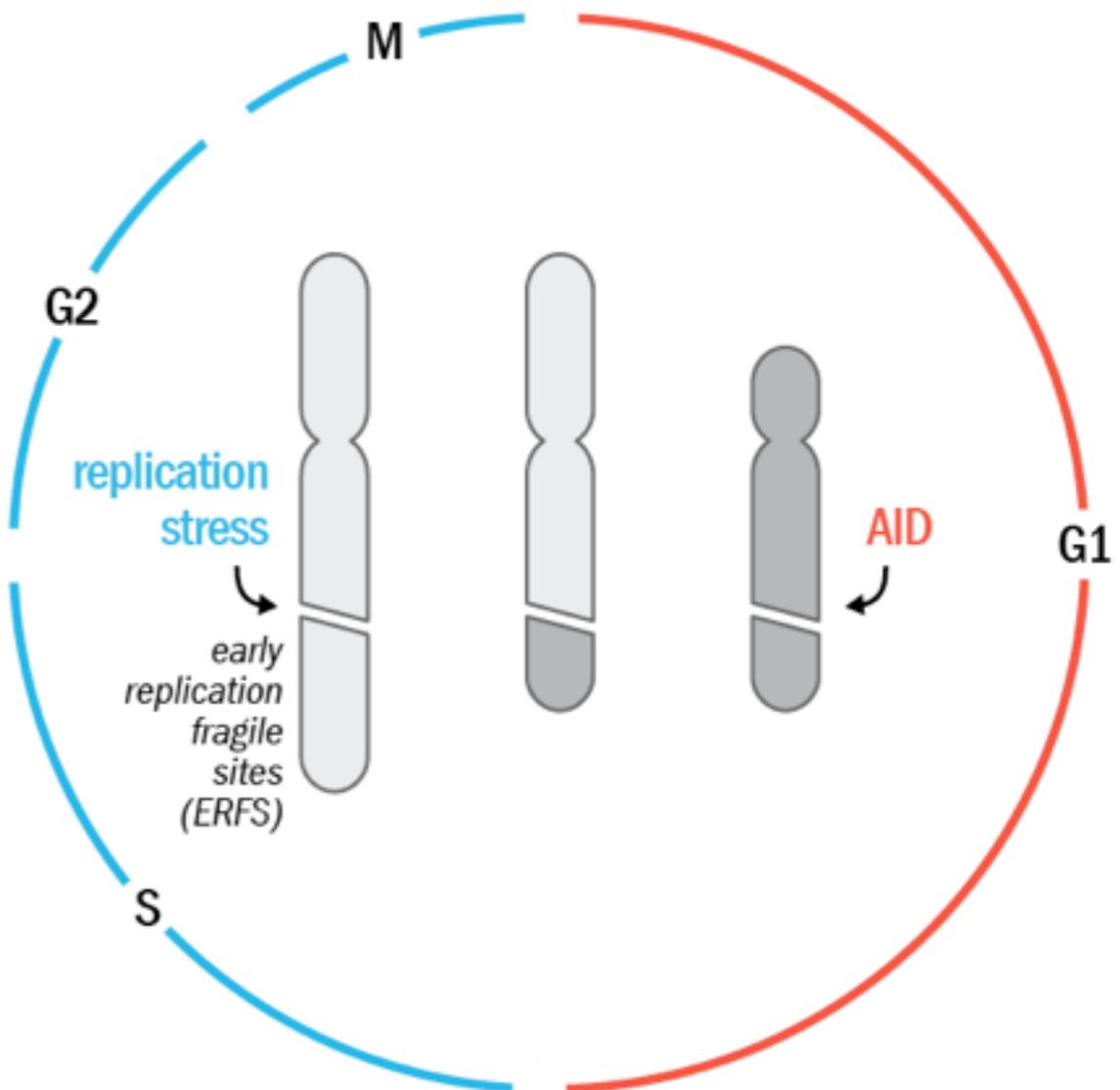
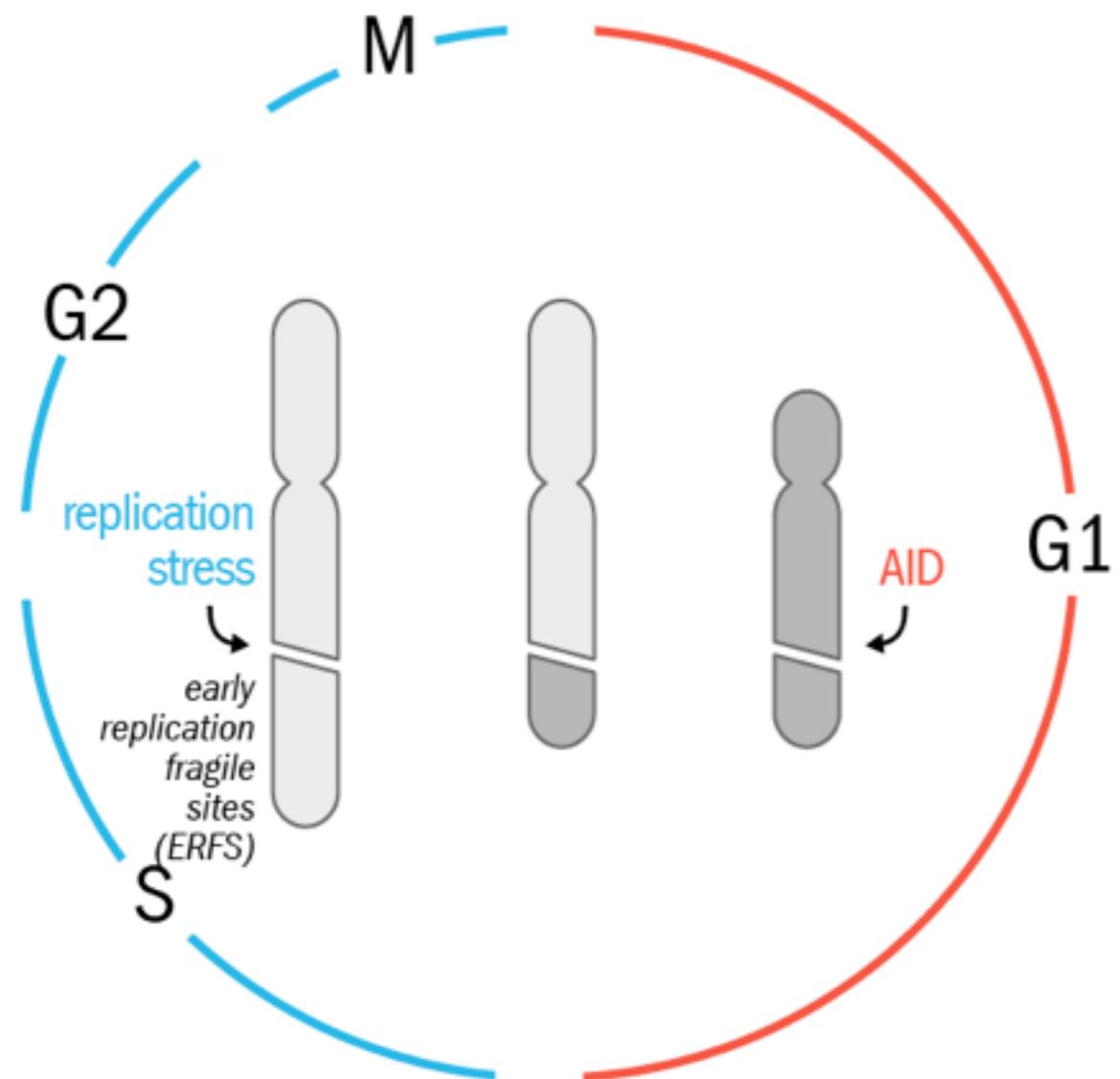


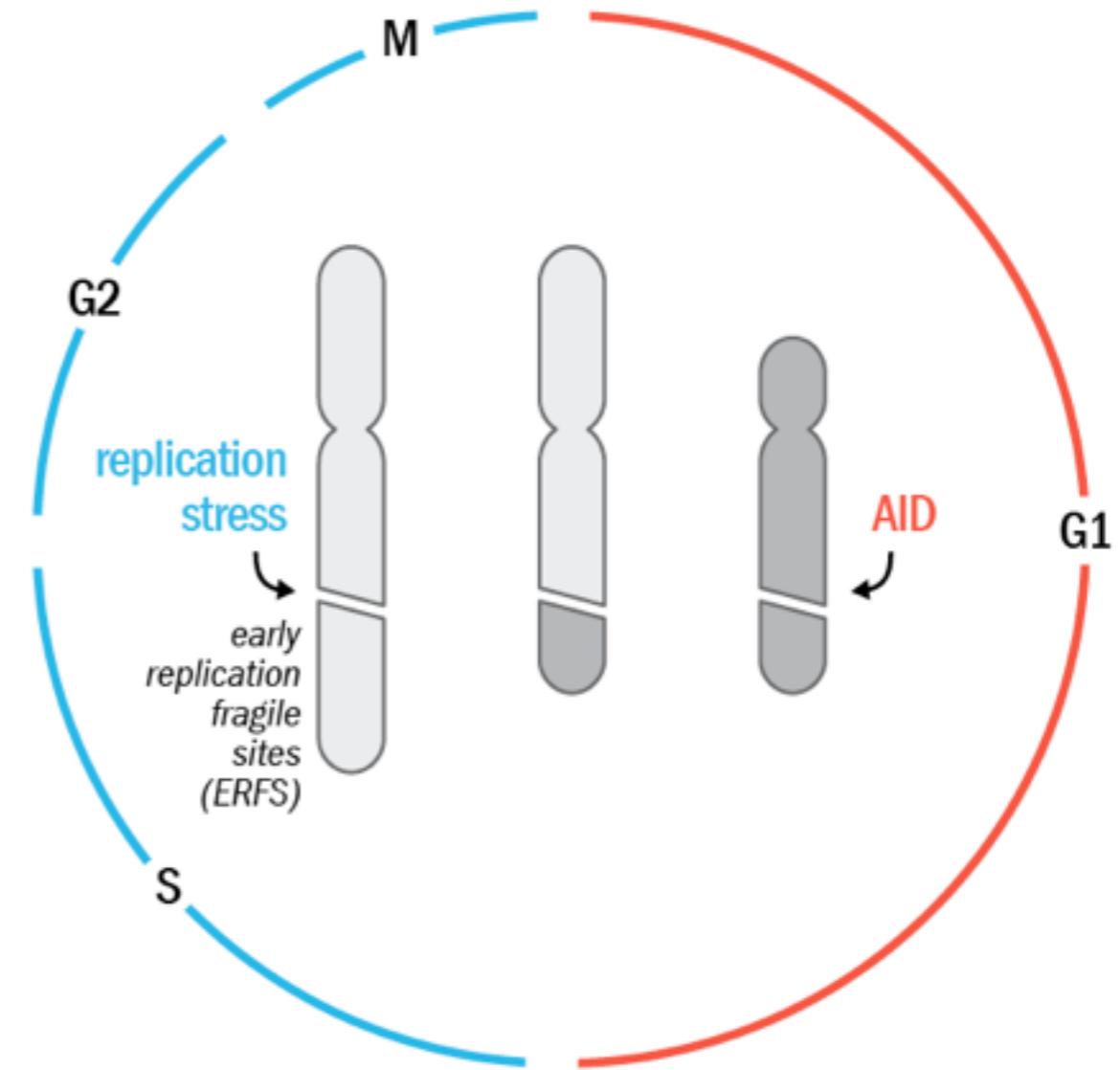
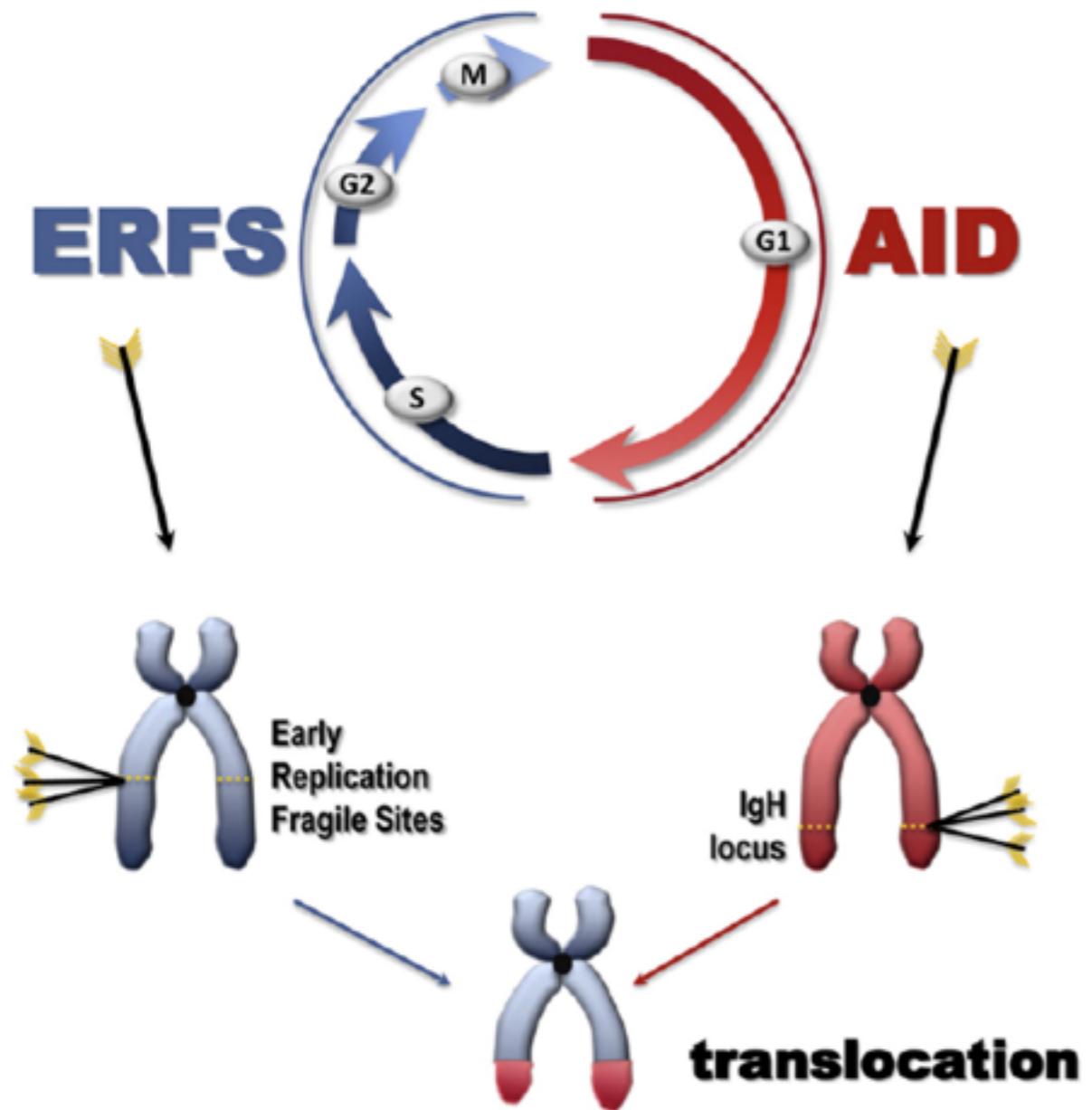












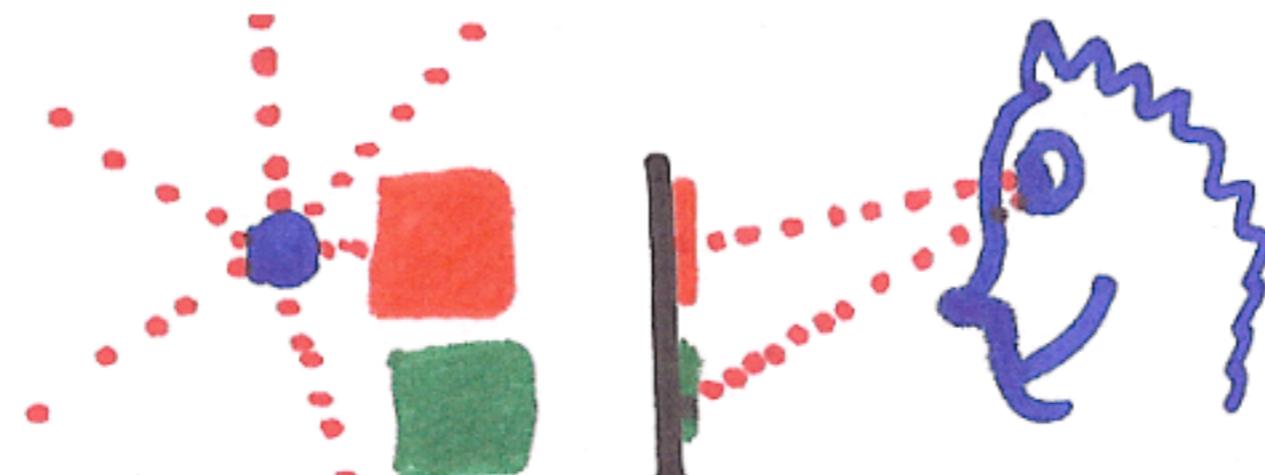
# 3D FIGURES

What you should definitely not do

# Dangers of depth

---

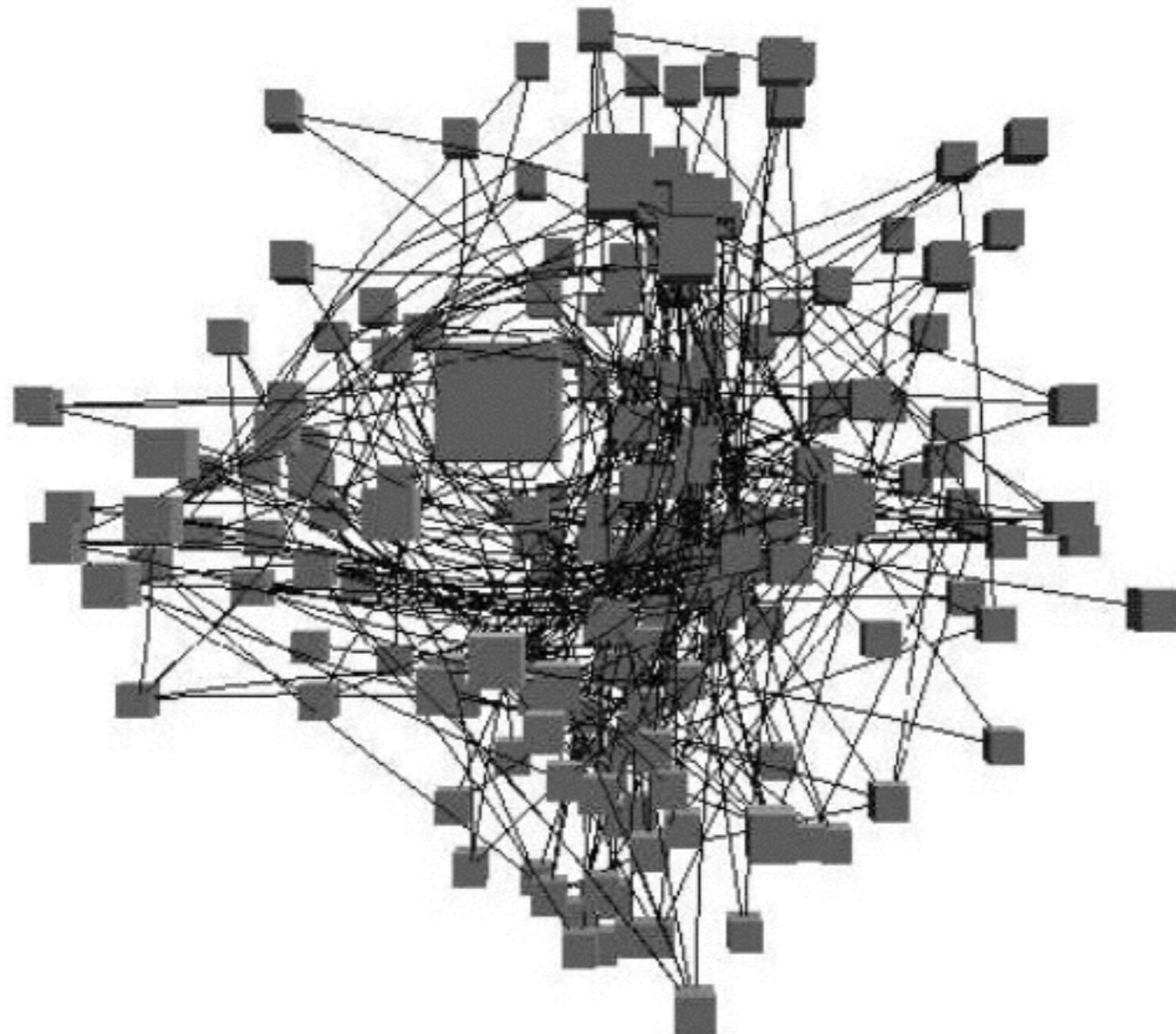
- rankings for planar spatial position, not depth!
- we don't really live in 3D: we **see** in 2.05D
- up/down and sideways: image plane
  - acquire more info quickly from eye movements
- away: depth into scene
  - only acquire more info from head/body motion



# Dangers of depth: difficulties of 3D

---

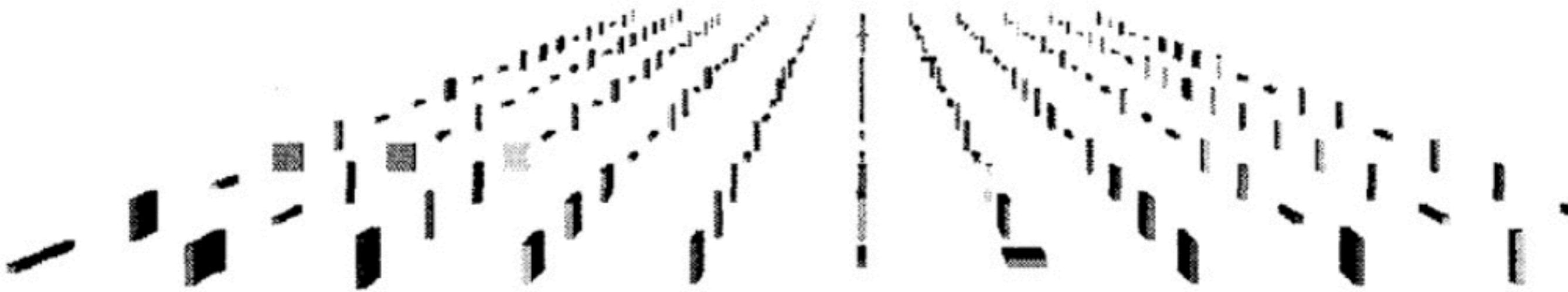
- occlusion
- interaction complexity



# Dangers of depth: difficulties of 3D

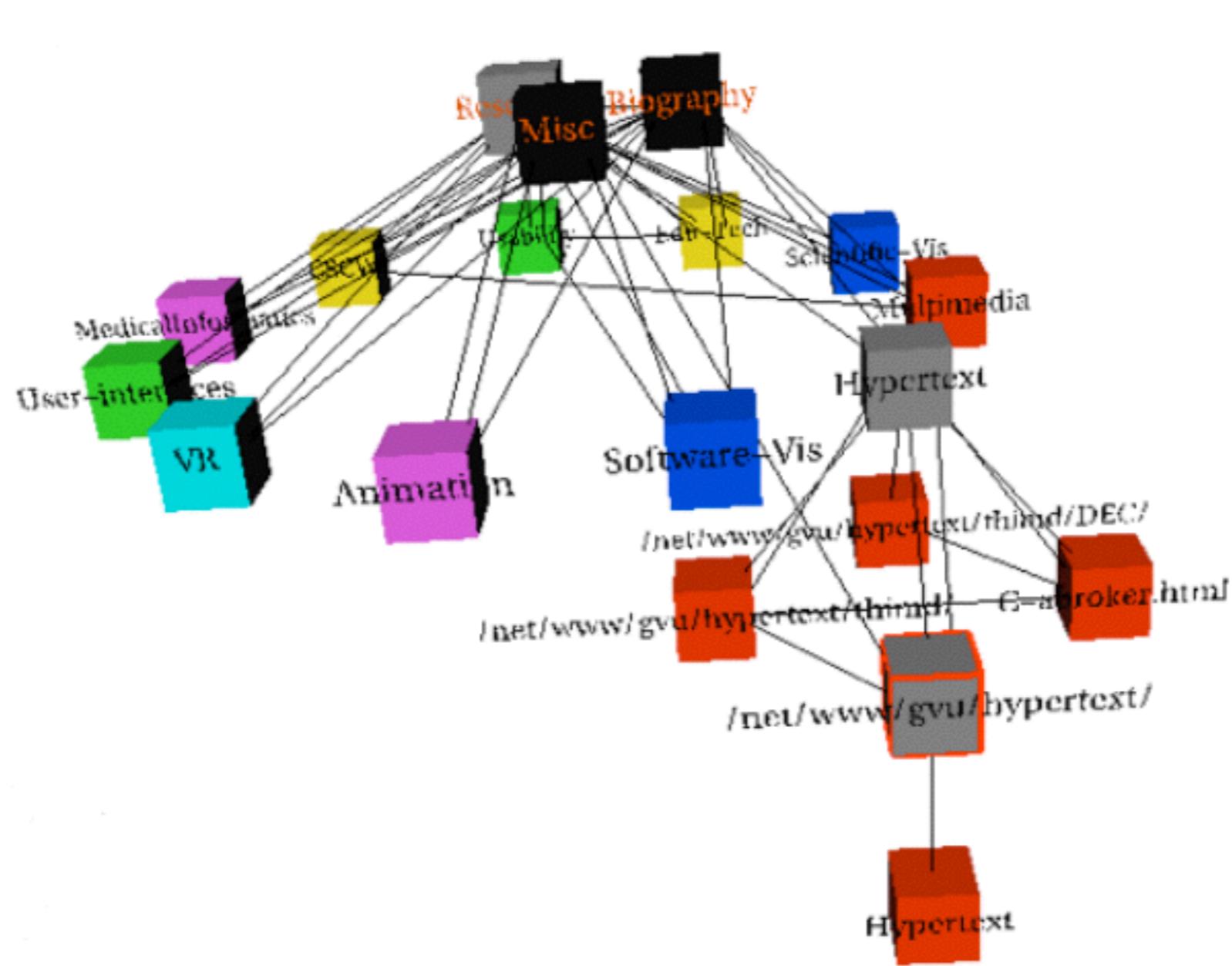
---

- perspective distortion
  - interferes with all size channel encodings
  - power of the plane is lost!

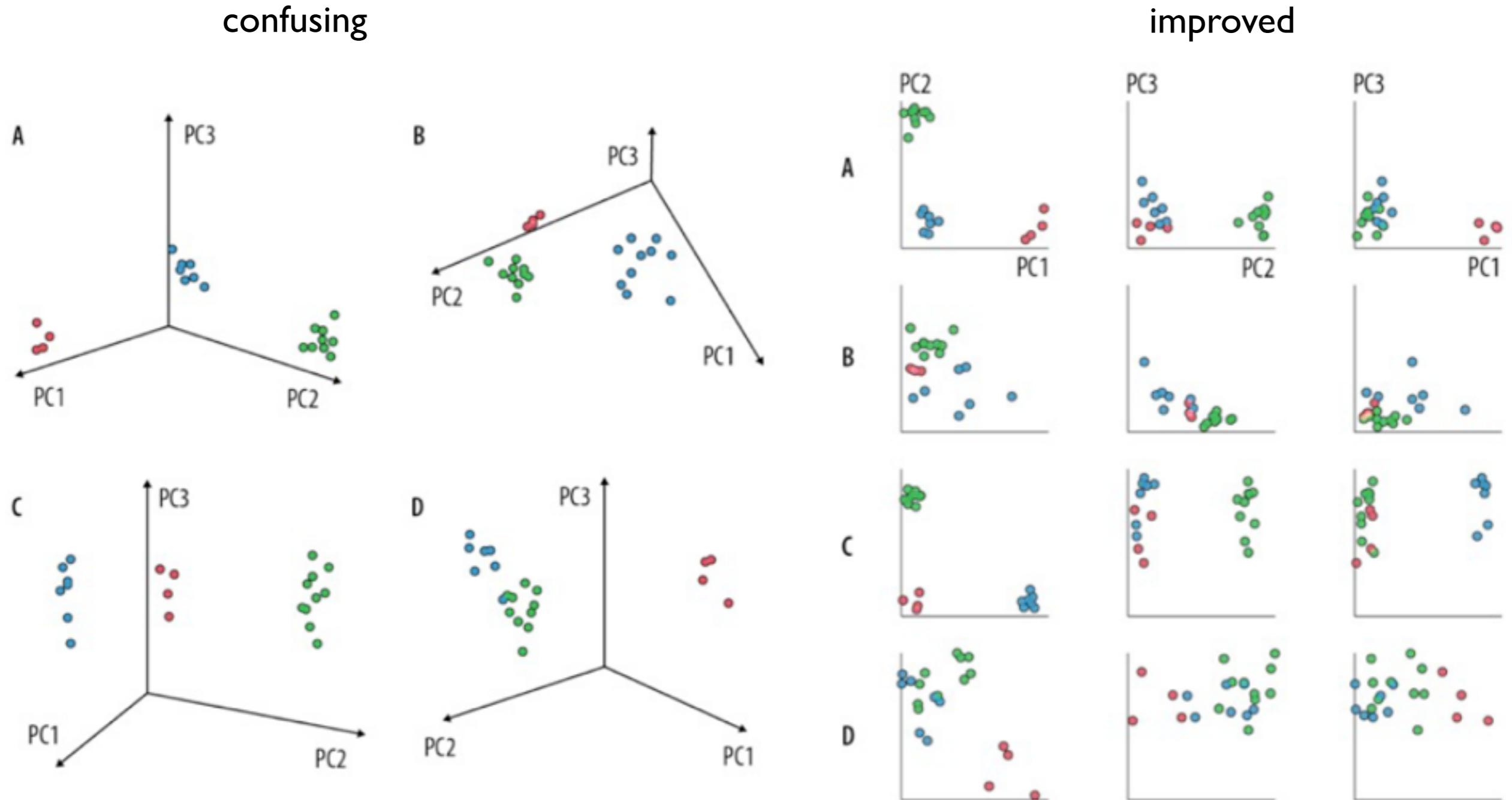


# Dangers of depth: difficulties of 3D

- text legibility
- far worse when tilted from image plane



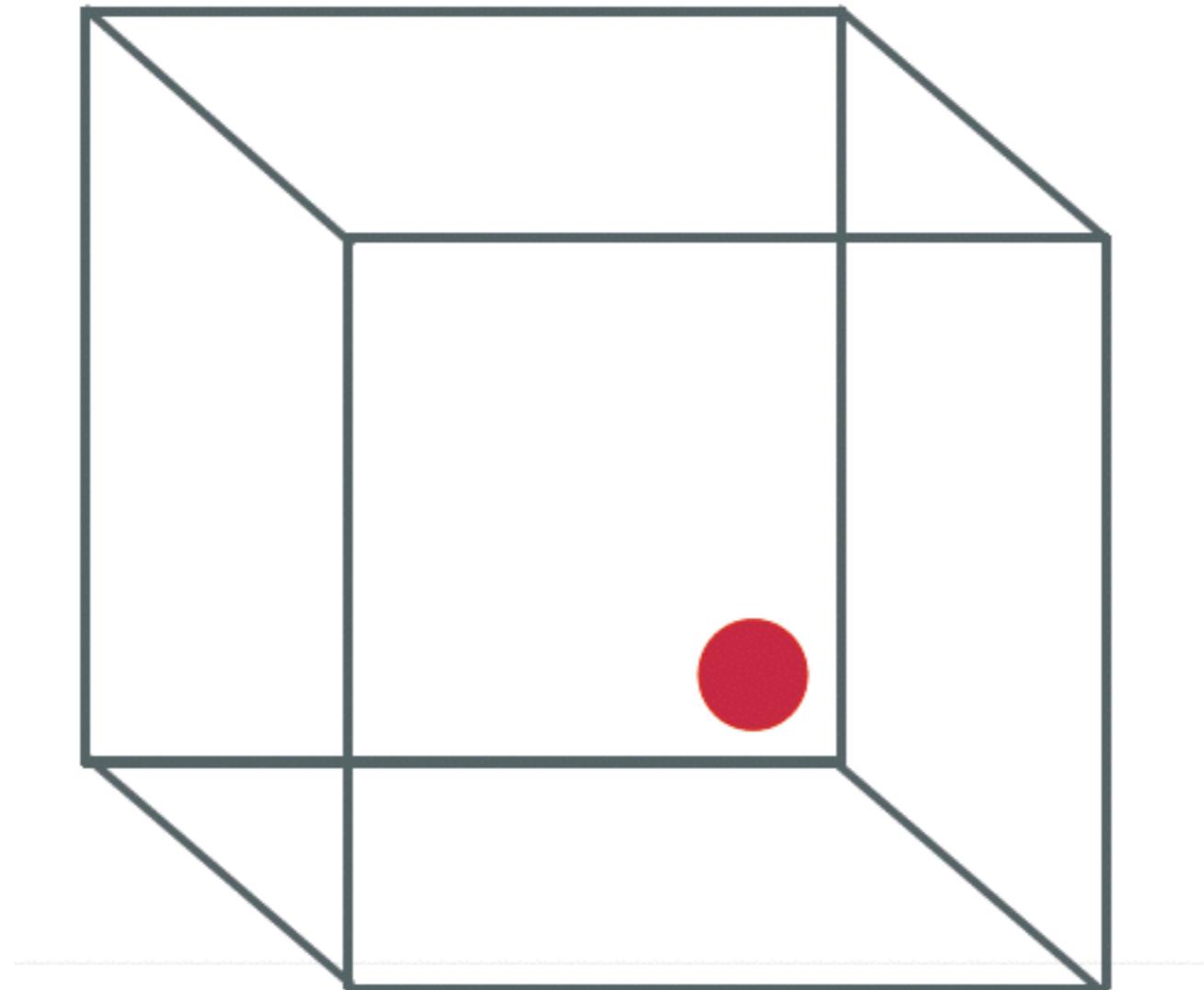
# Confusing 3D graphs



Sharov AA, Dudekula DB, Ko MS (2005) Genome-wide assembly and analysis of alternative transcripts in mouse. Genome Res 15: 748-754

# Ambiguous Information: Position in 2.05D

---



# VISUAL ABSTRACT CREATION

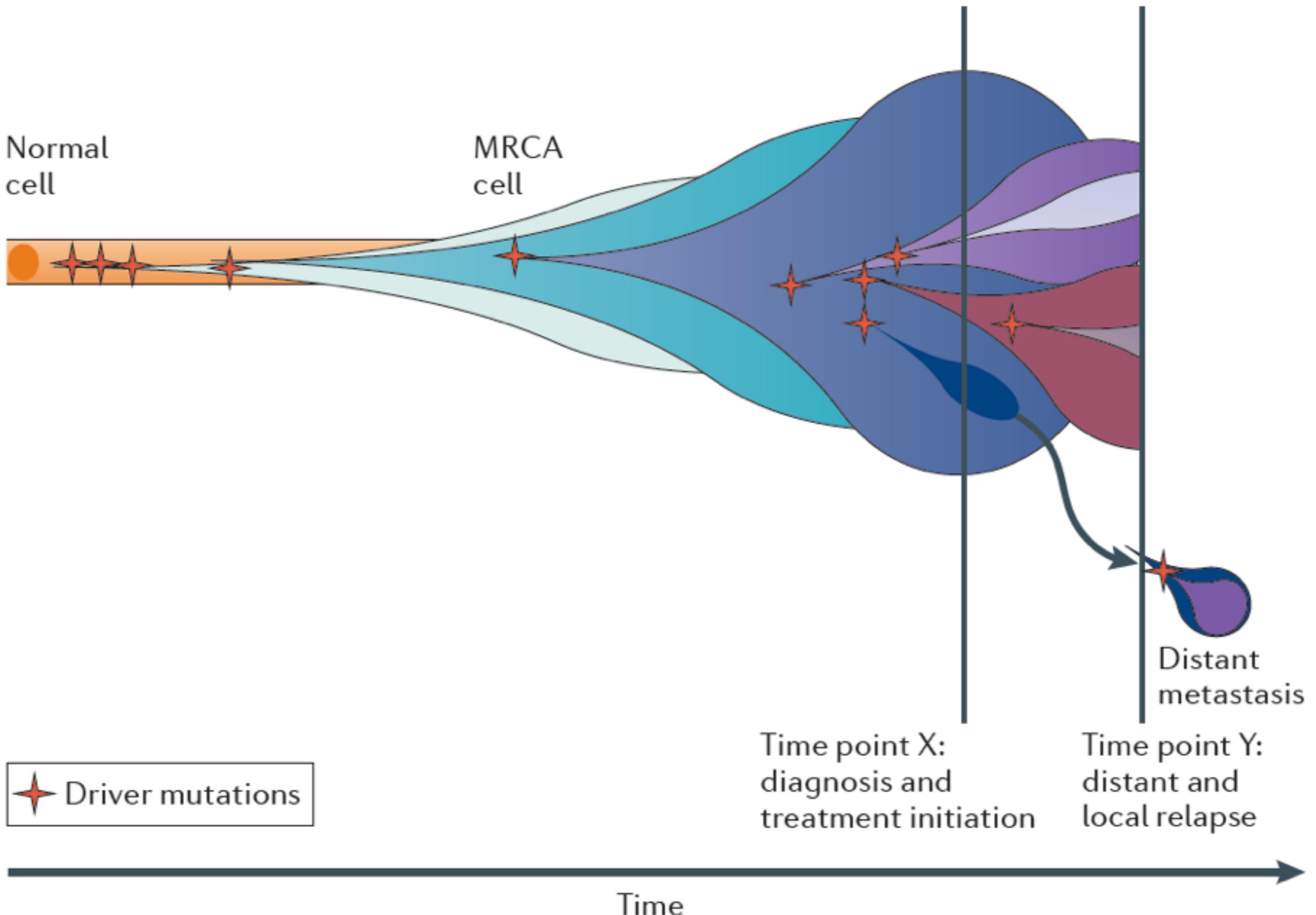
Identifying a deliverable message.

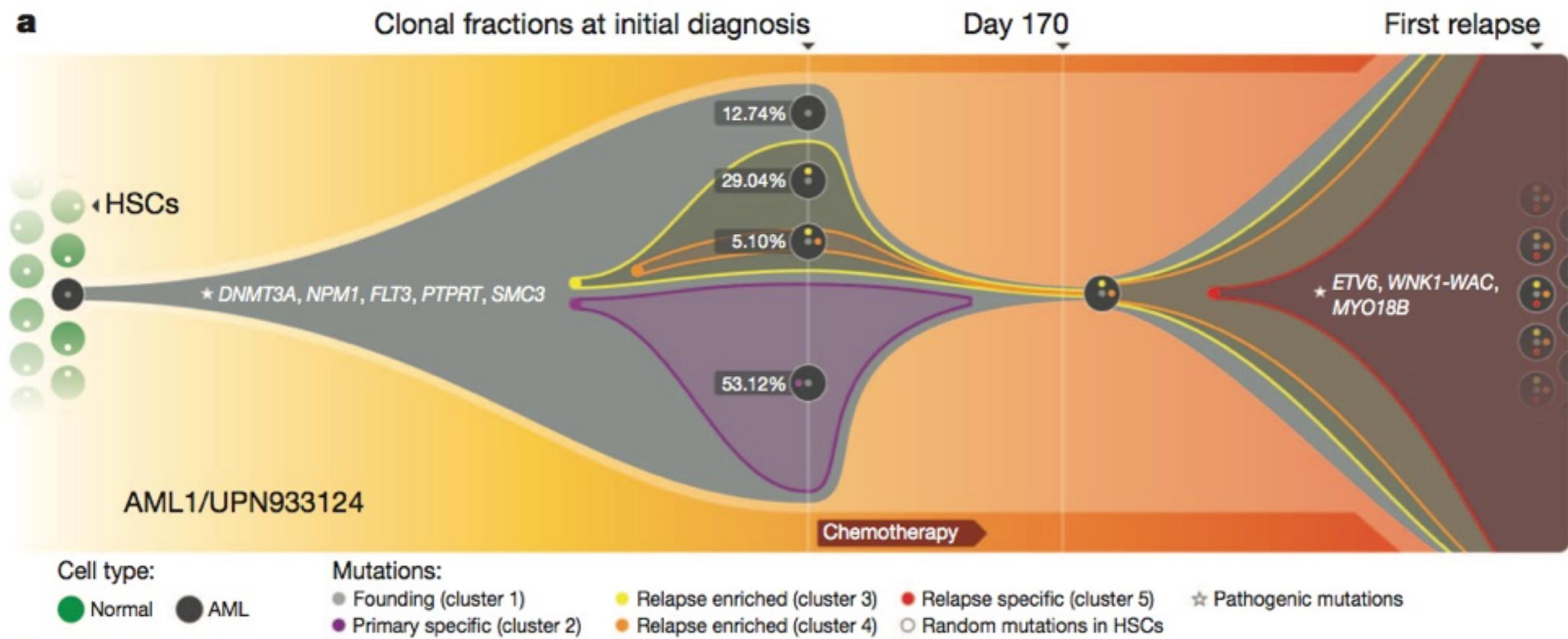
## KEYWORDS

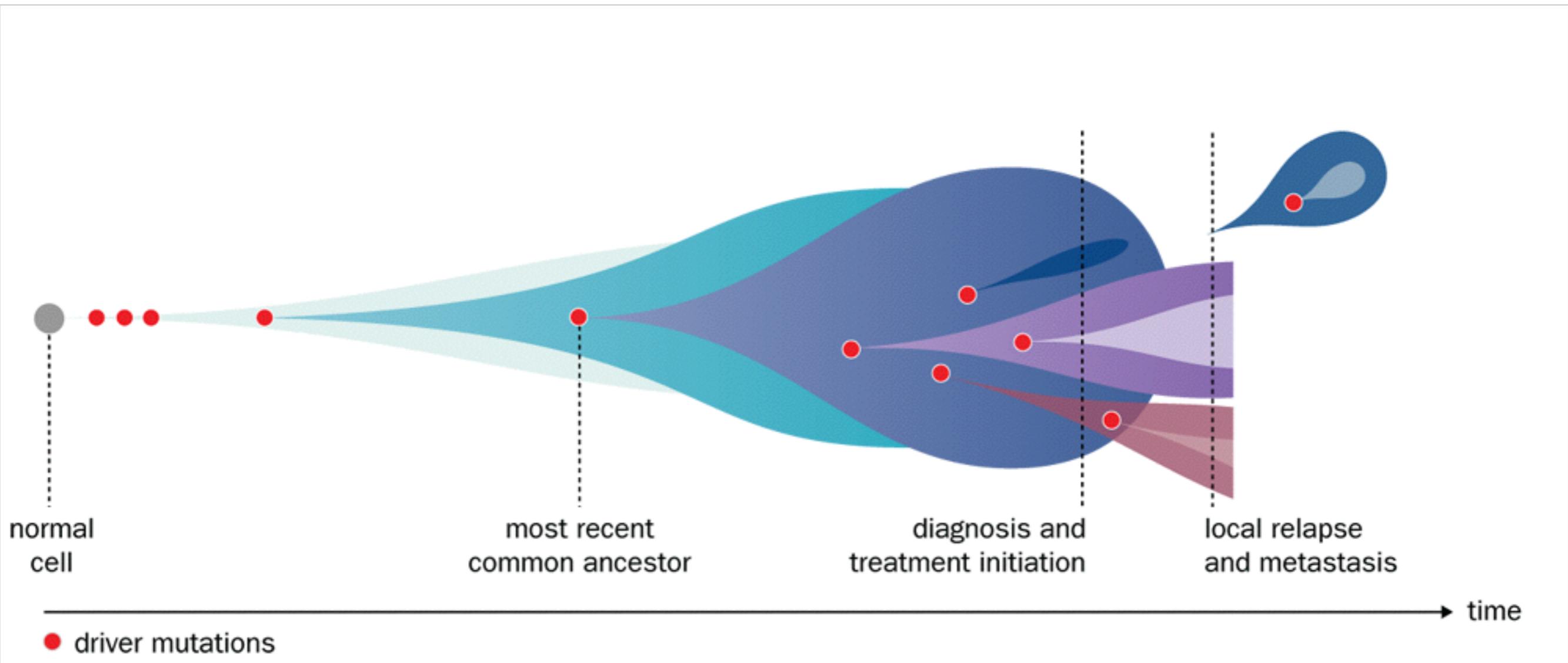
cancer  
cancer genomics  
tumor heterogeneity  
next-generation sequencing  
second-generation sequencing  
third-generation sequencing  
mutation discovery  
whole genome sequencing  
single molecule sequencing  
single cell sequencing  
personalized medicine

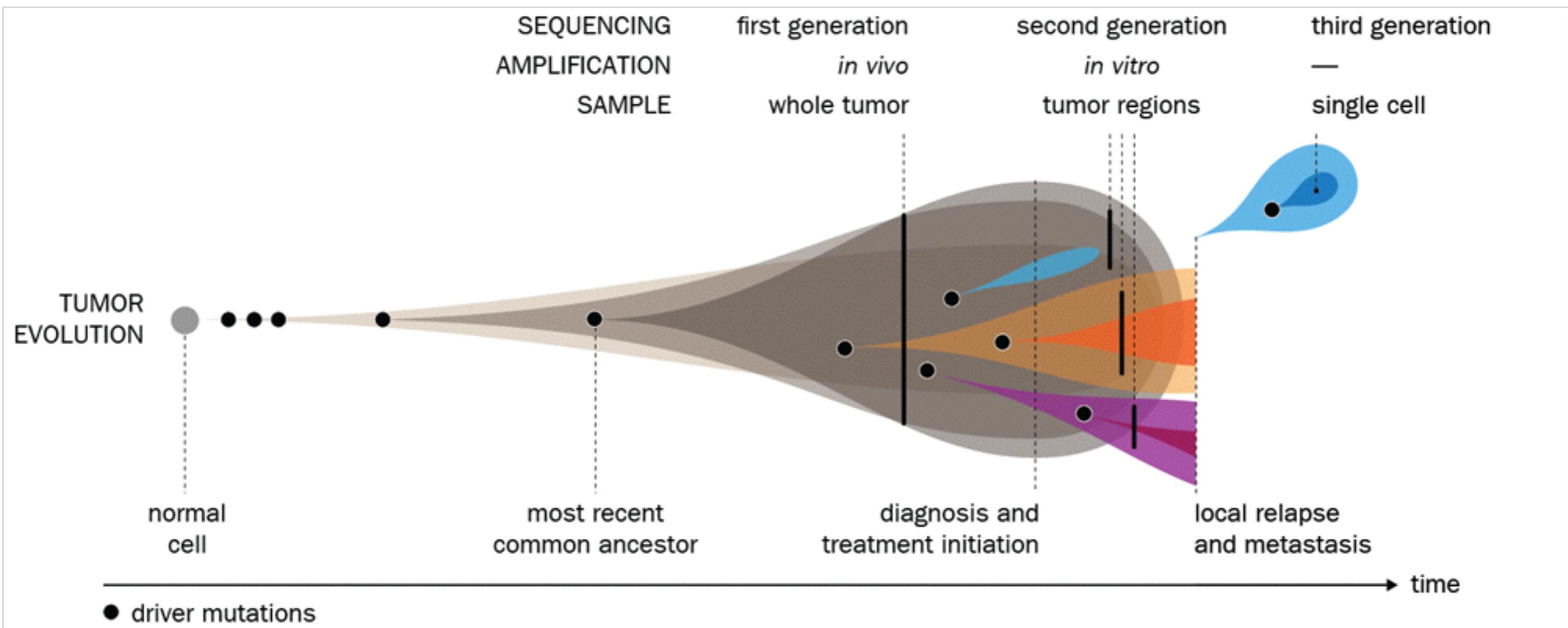
# ELSEVIER FIGURE RESTRICTION

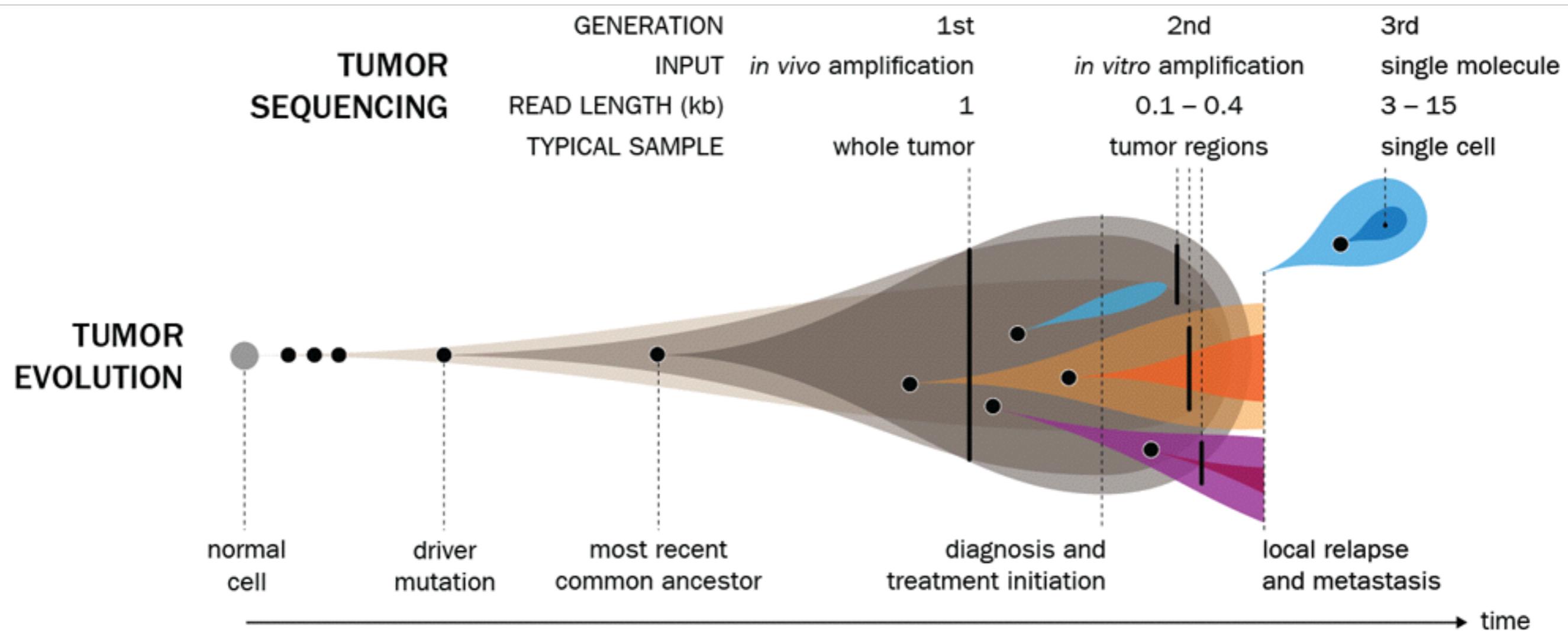
Please note that your image will be scaled proportionally to fit in the available window on ScienceDirect, a **500 by 200 pixel rectangle**.

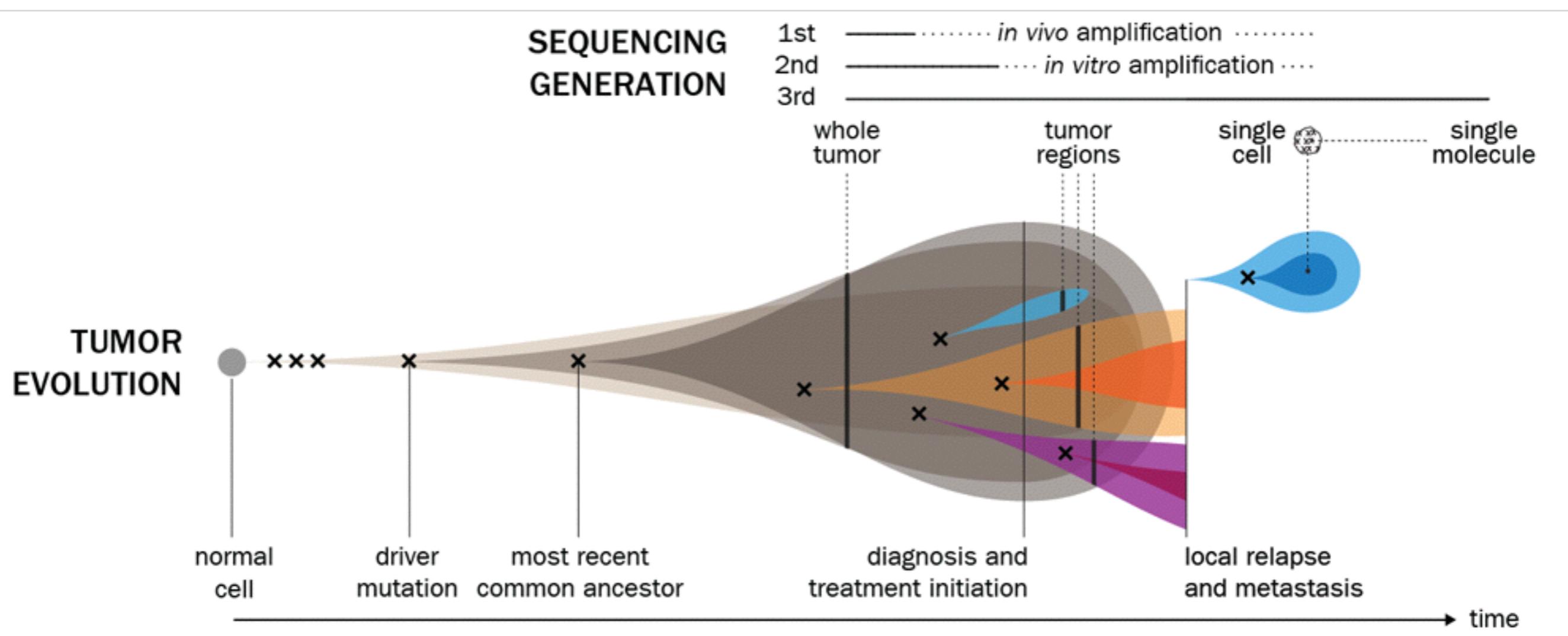


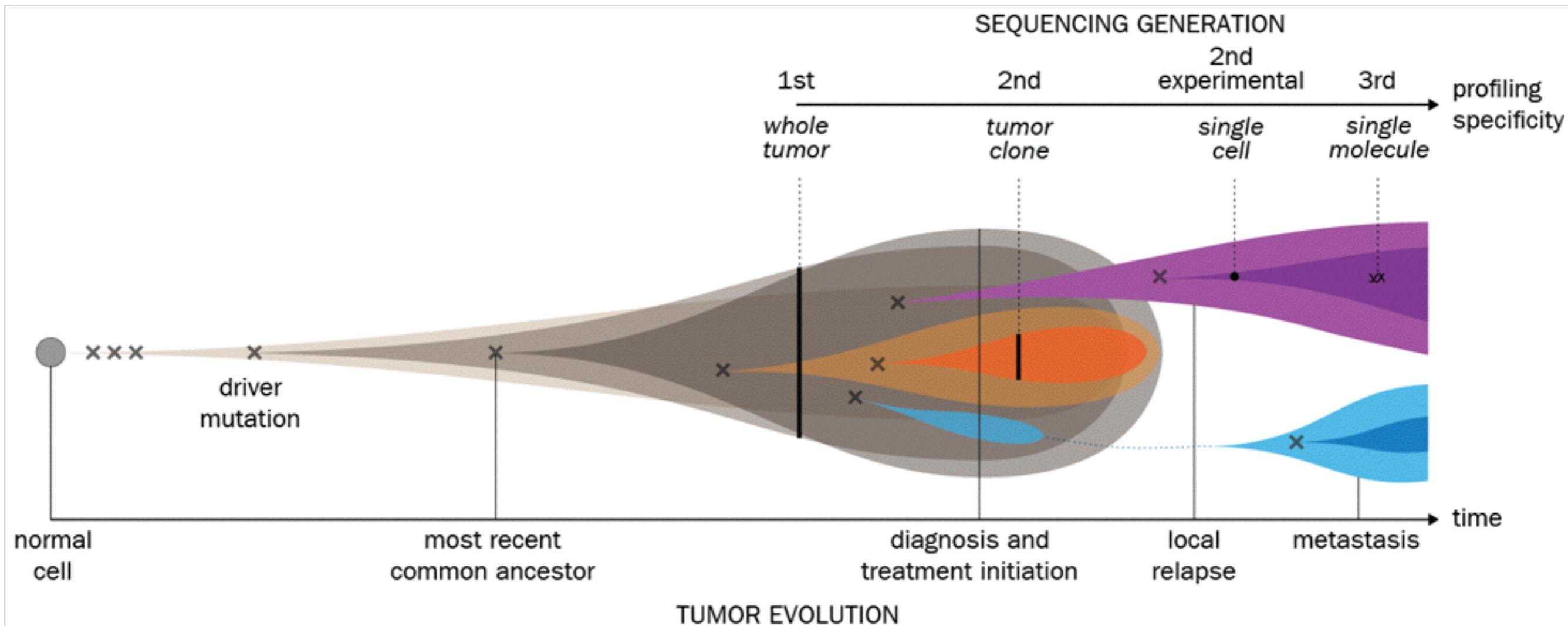
**a**

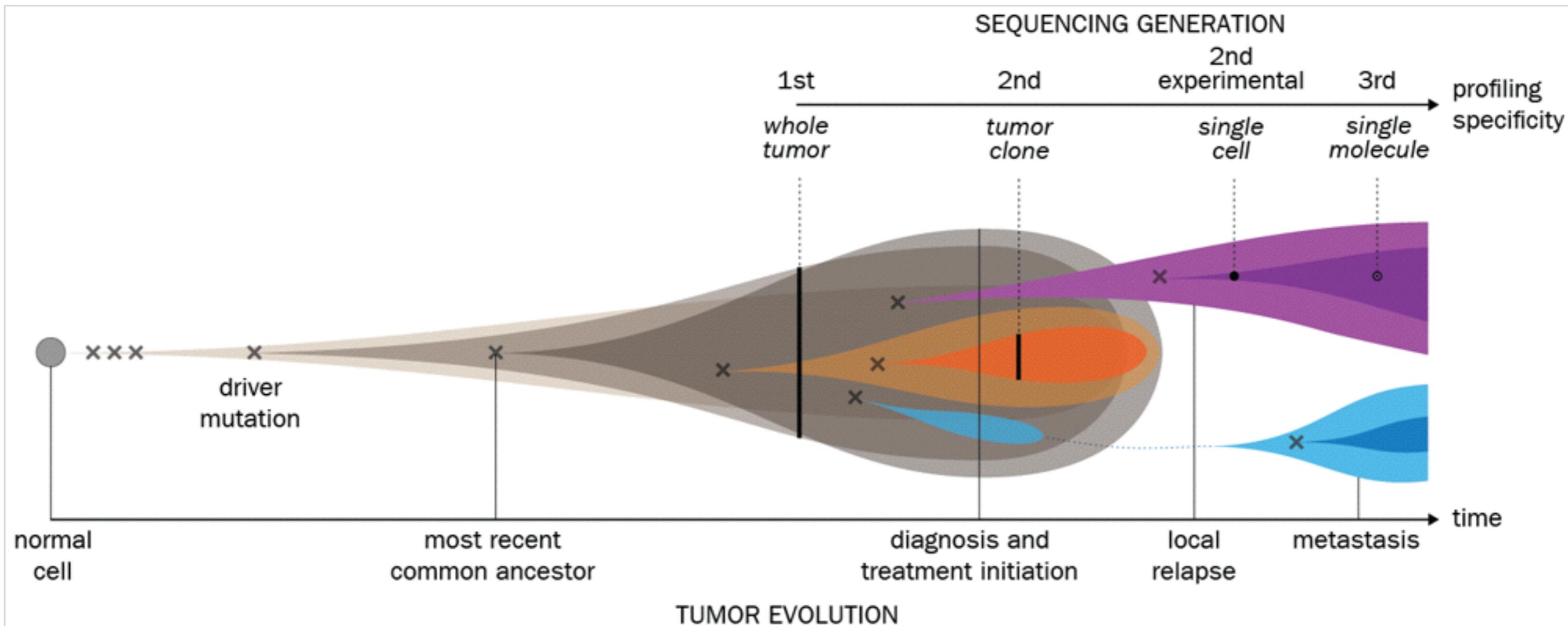


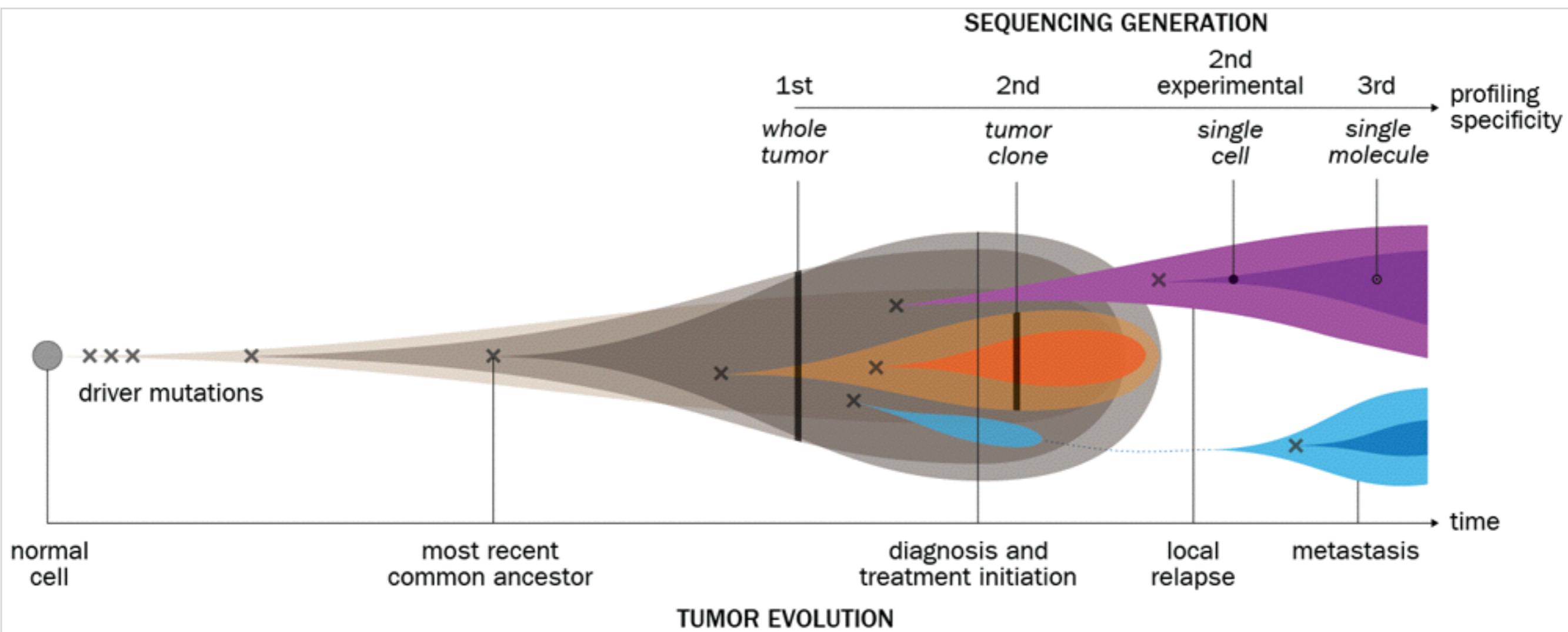


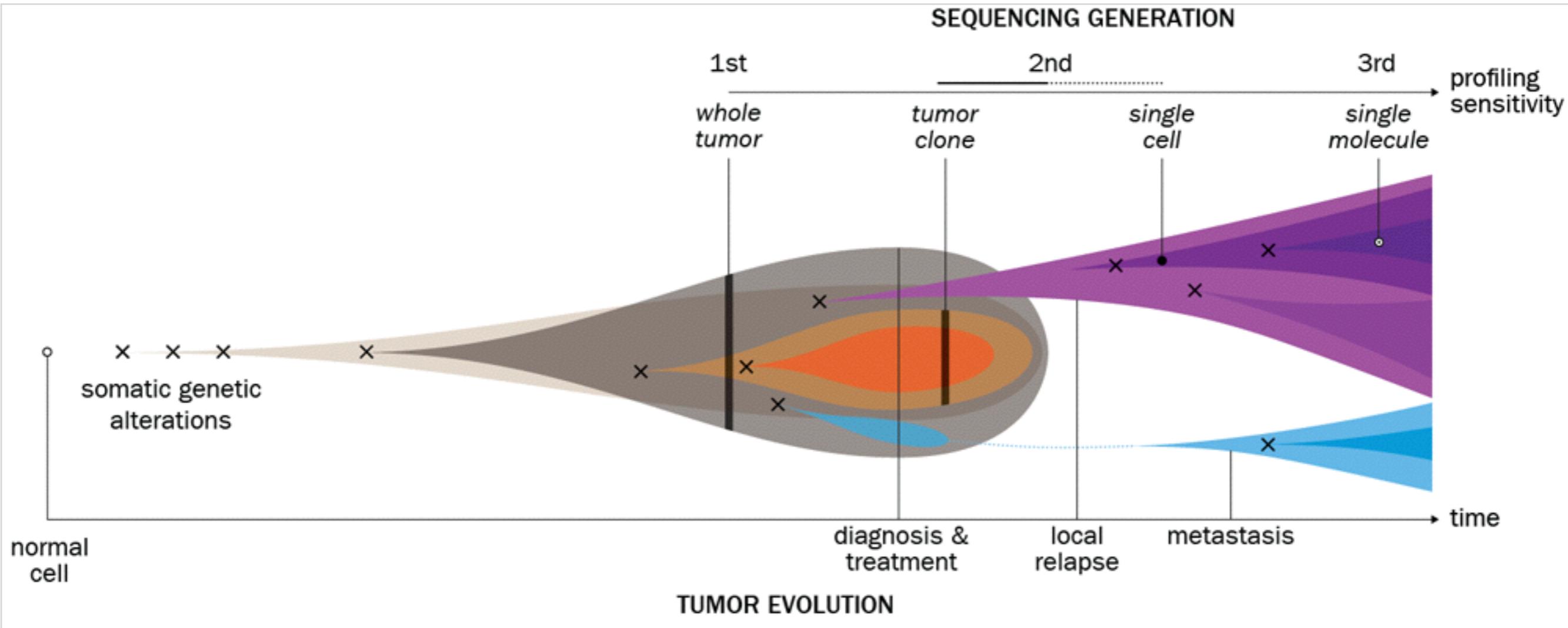




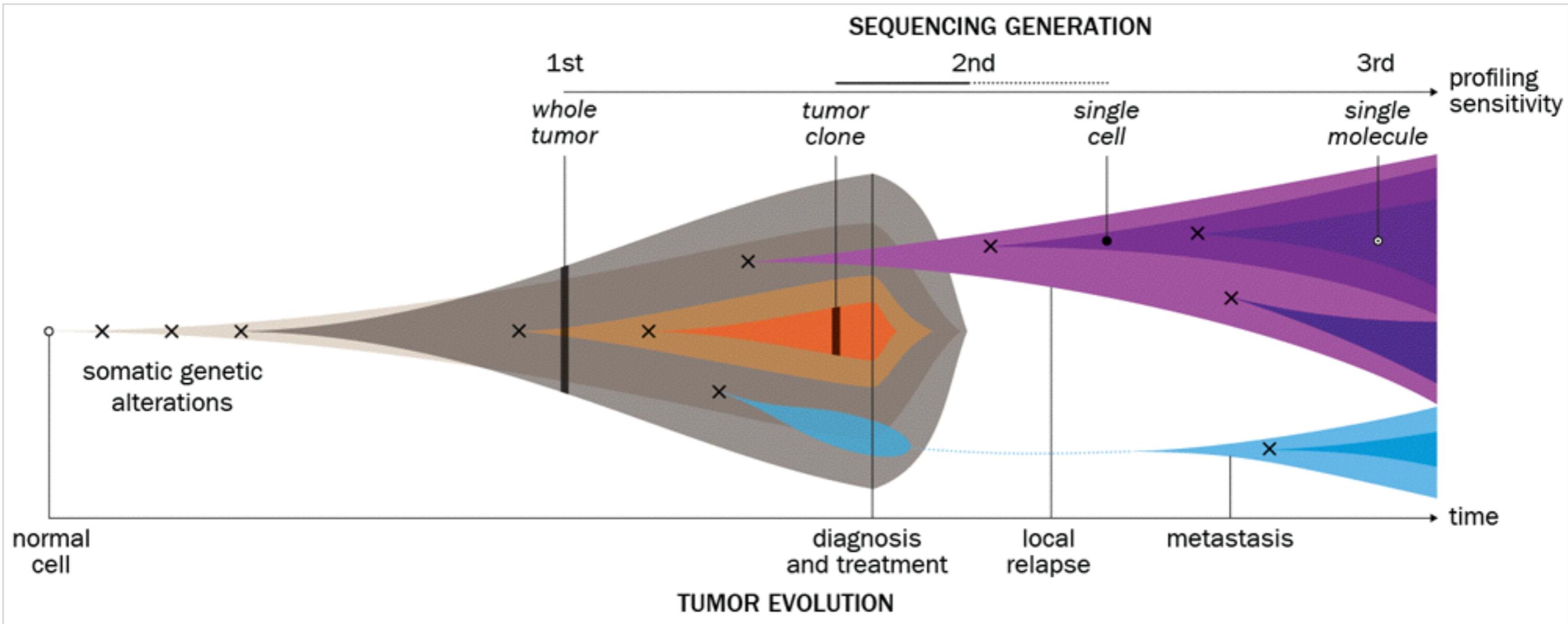


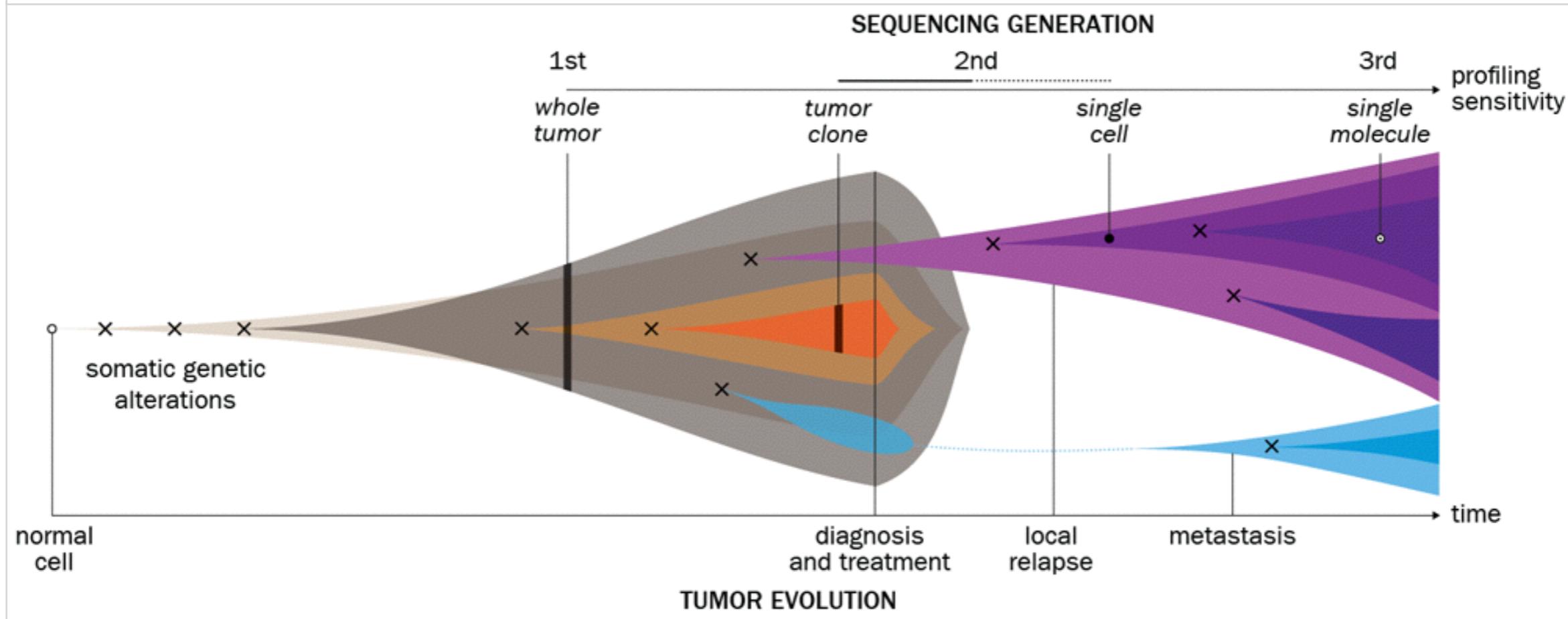
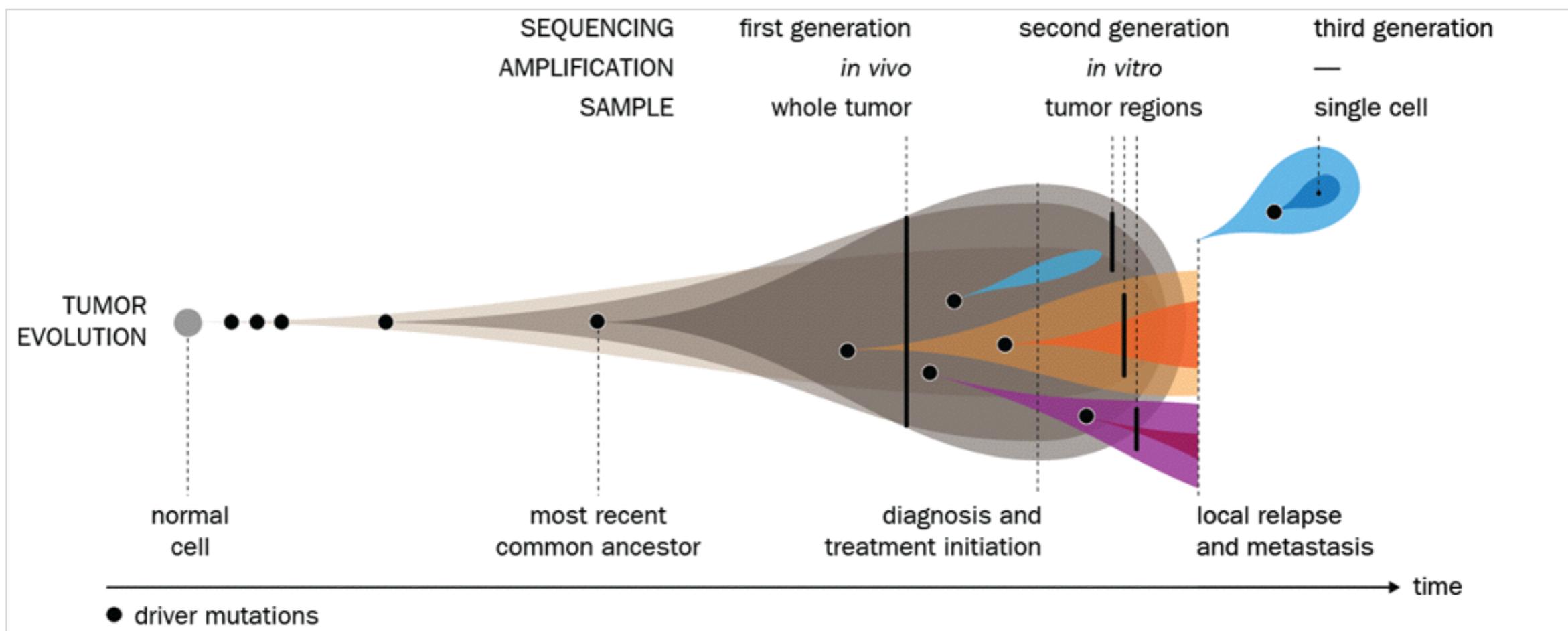






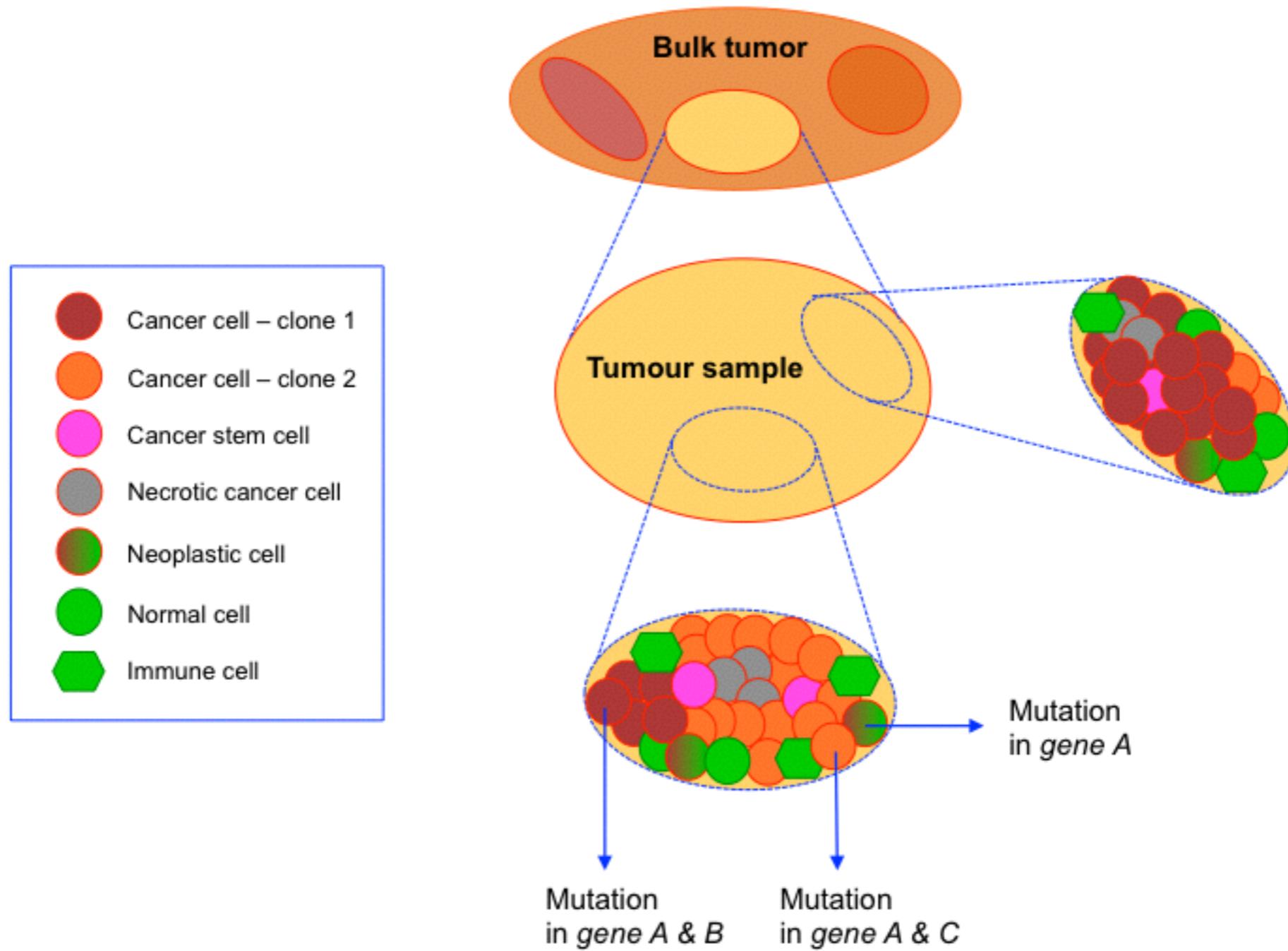
Chun HJE, Khattra J, Krzywinski M, Aparicio SA, Marra MA (Dellaire GD, Berman JN, Arceci RJ. editors). Cancer Genomics. 1st ed. Elsevier; 2013.  
Early version of visual abstract.

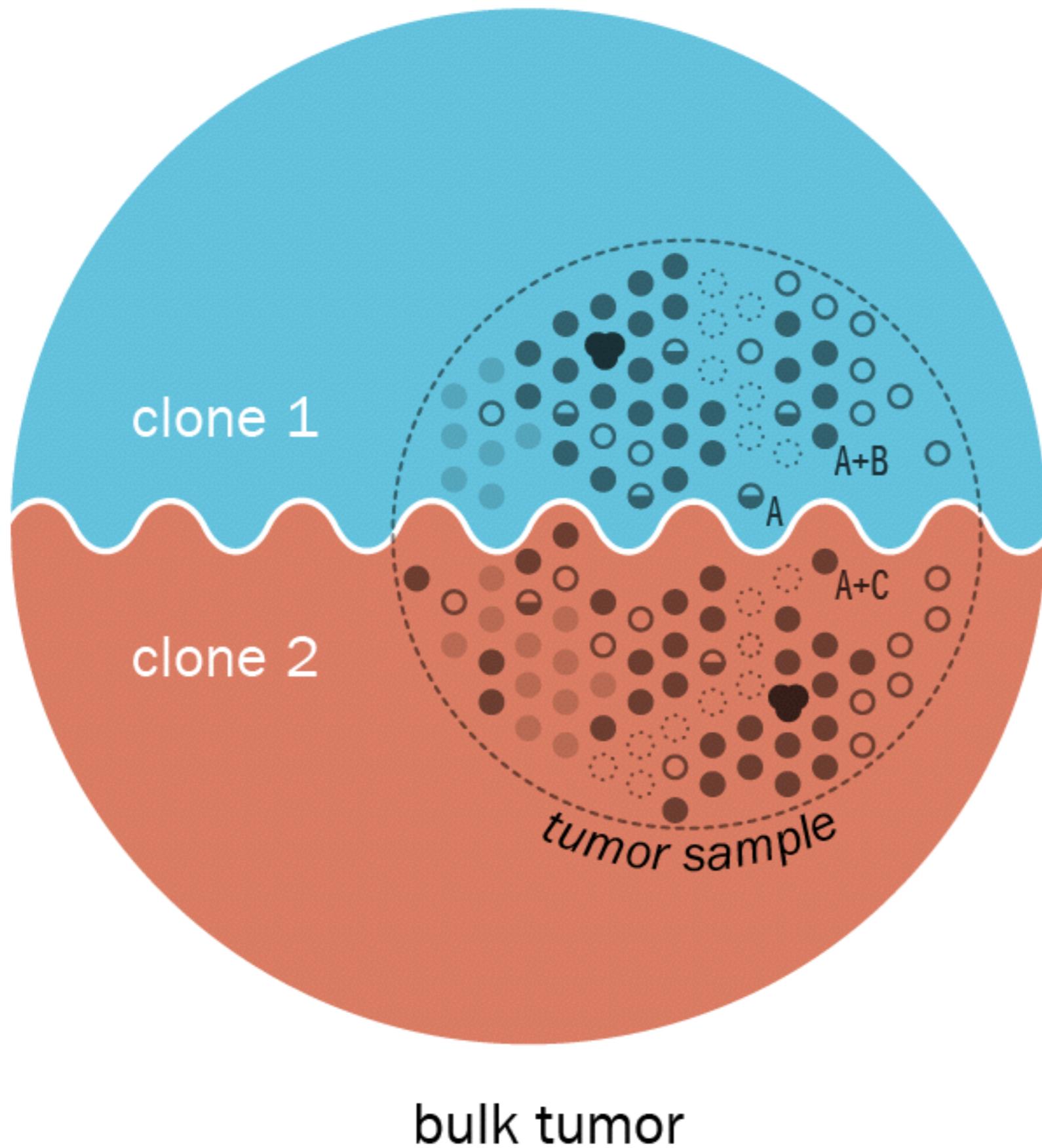




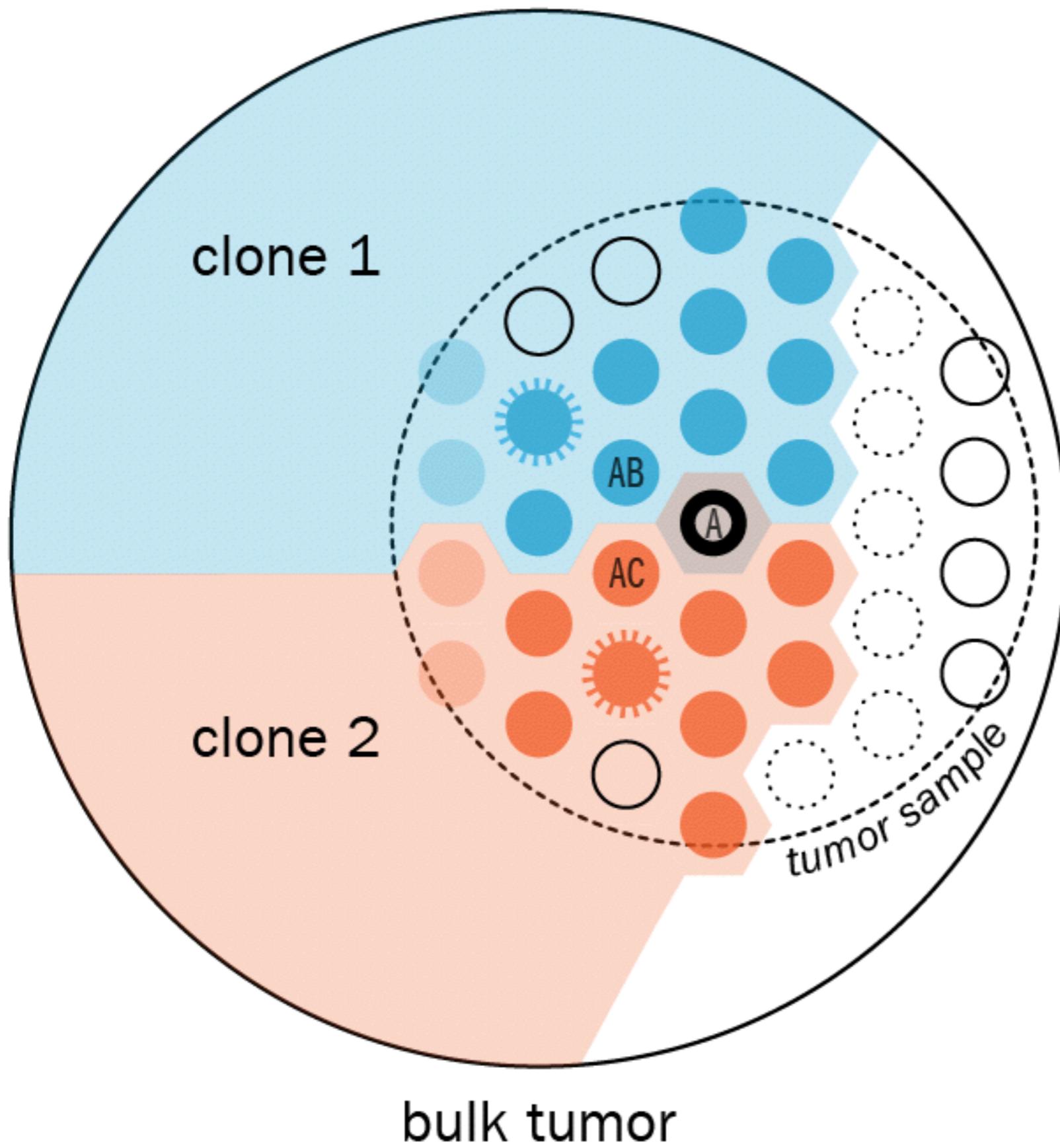
# REDESIGN—IN HOUSE

Rational visual vocabulary for cell types.

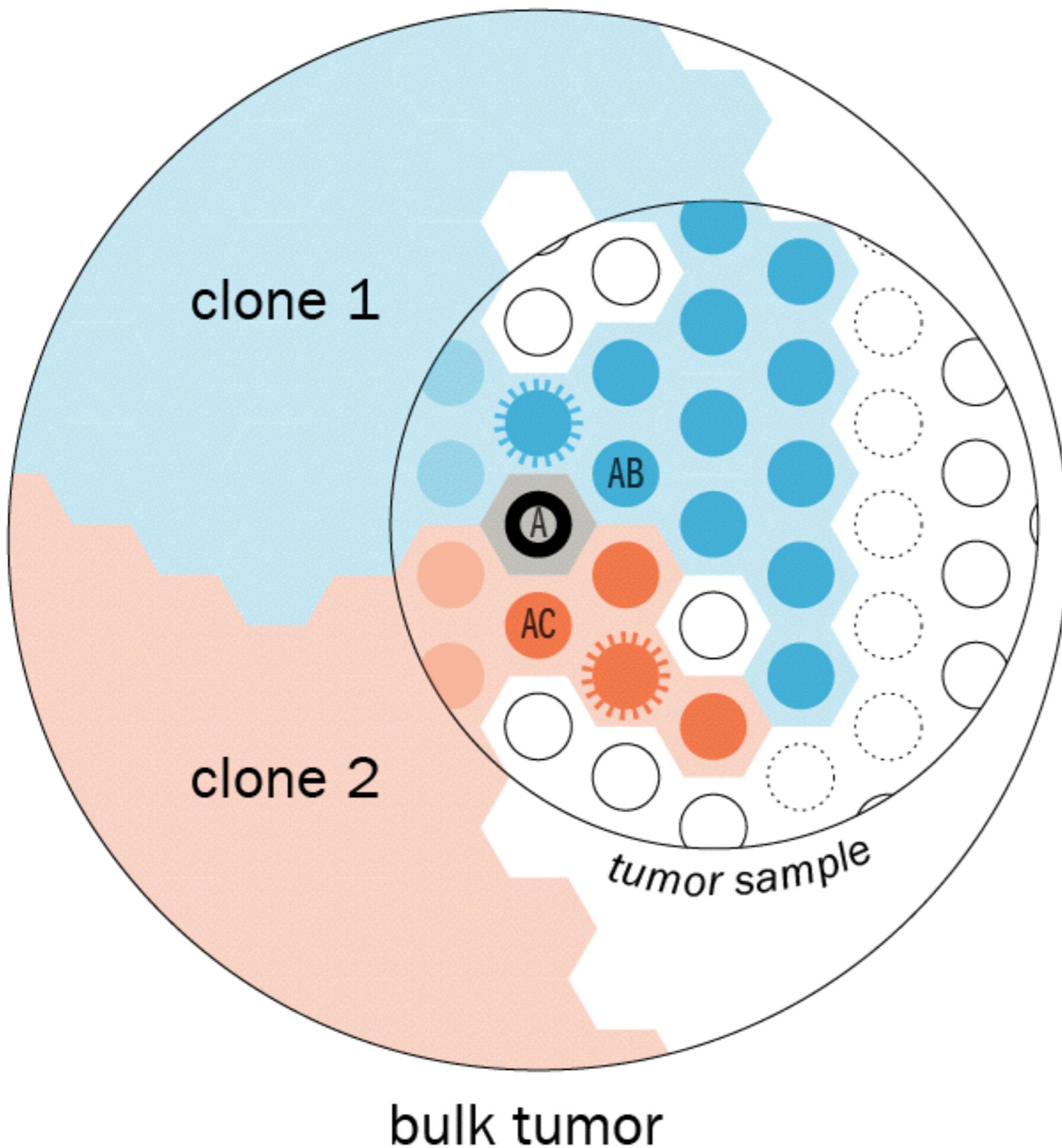




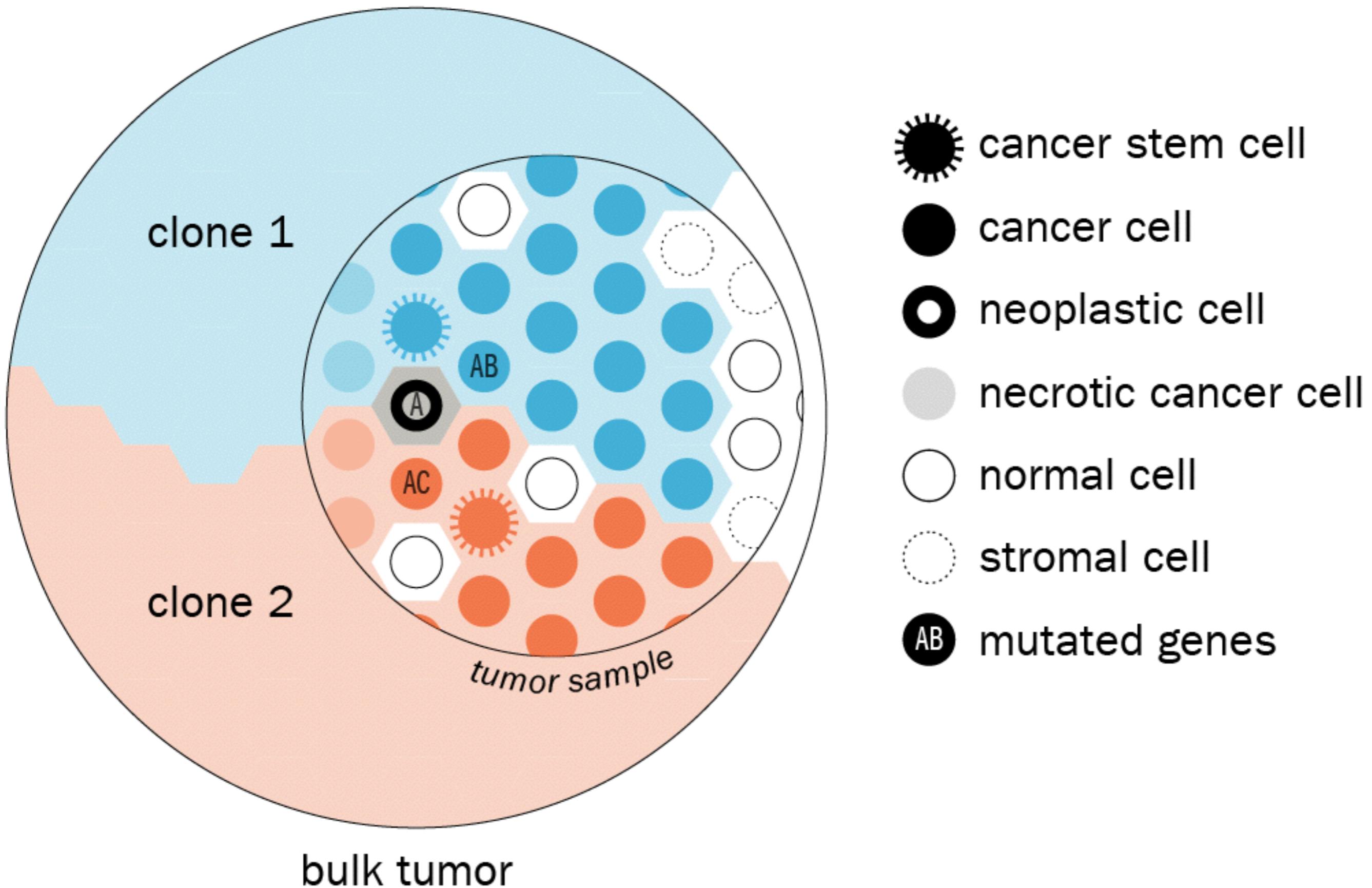
- cancer stem cell
- cancer cell
- necrotic cancer cell
- neoplastic cell
- normal cell
- stromal cell
- A+B mutated genes

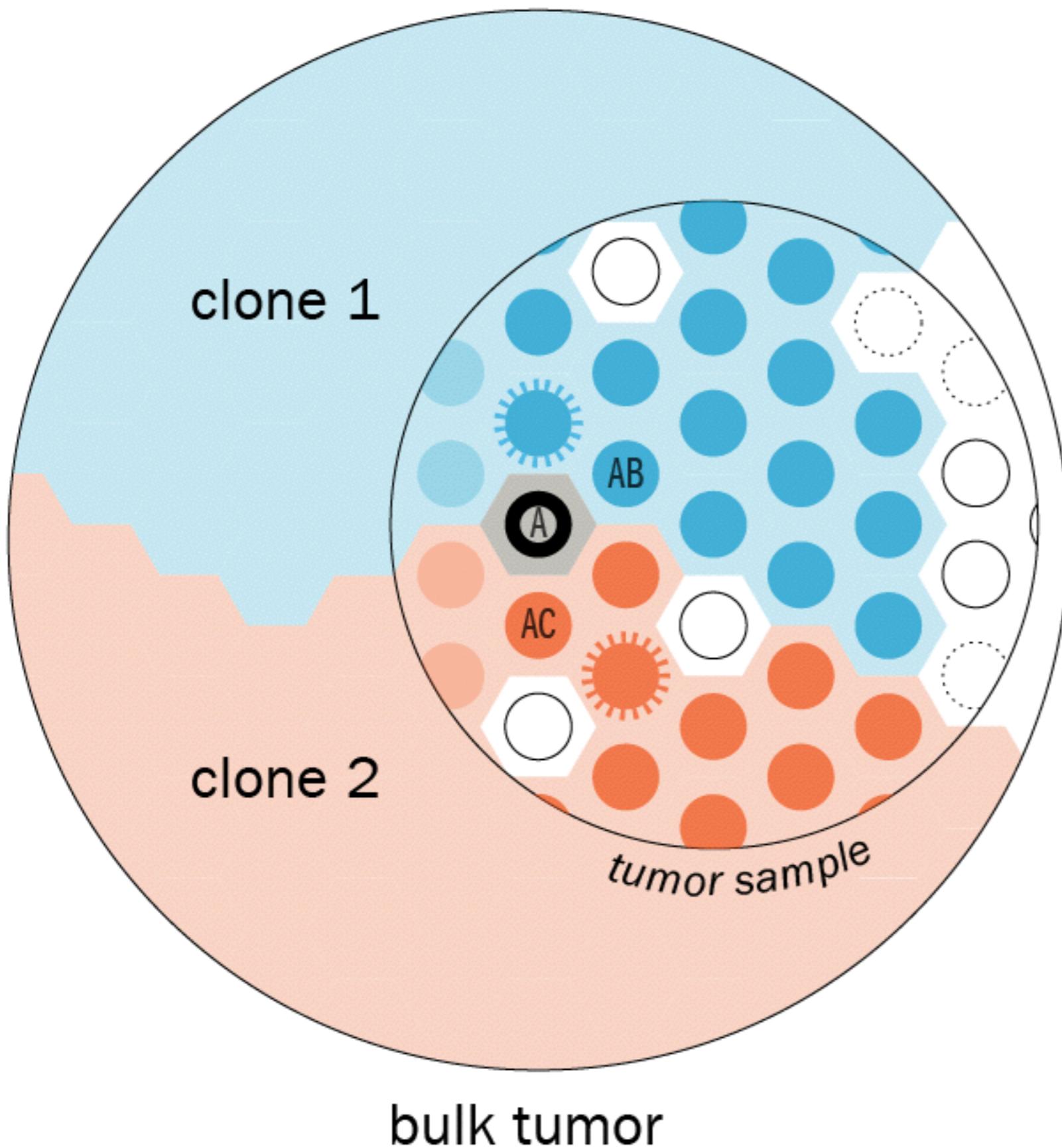


- cancer stem cell
- cancer cell
- neoplastic cell
- necrotic cancer cell
- normal cell
- stromal cell
- mutated genes

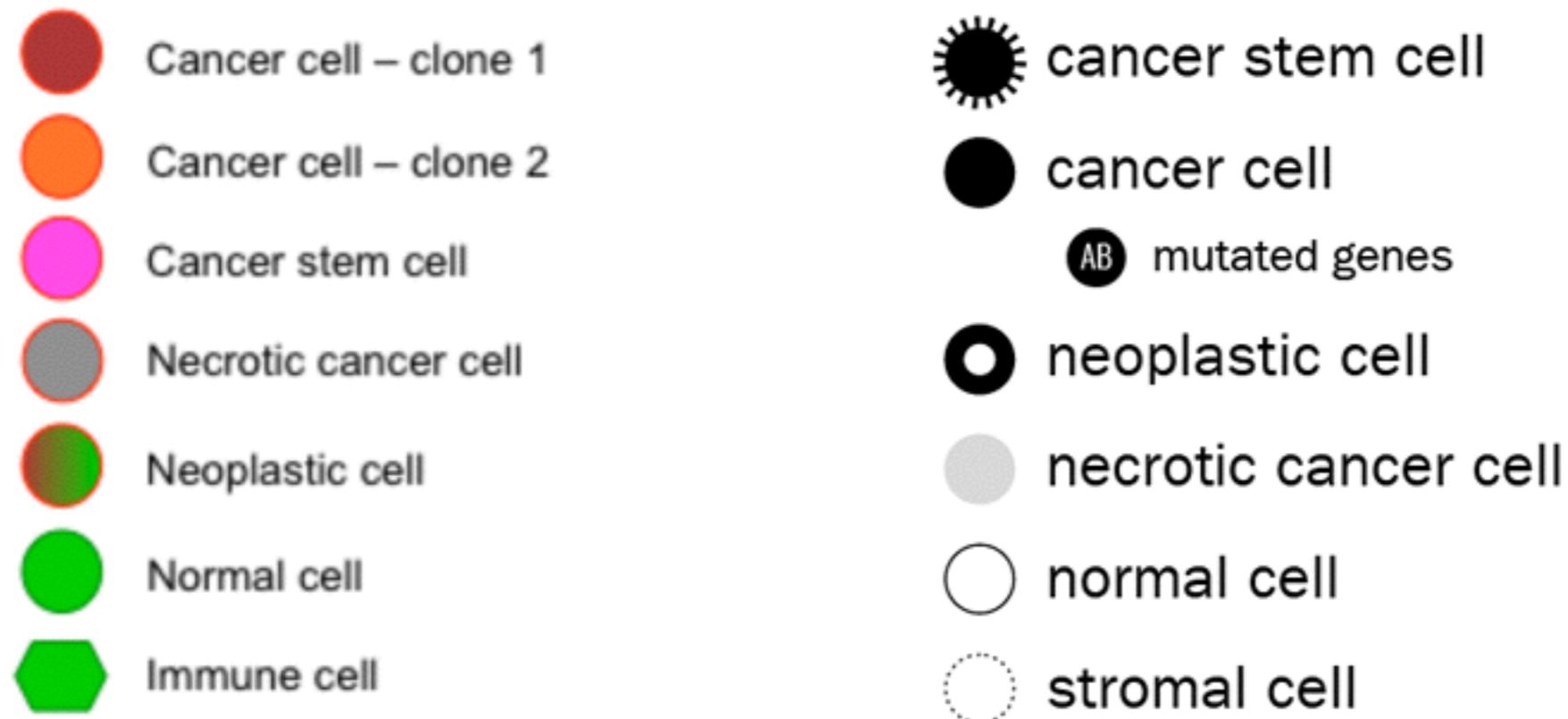


- cancer stem cell**
- cancer cell**
- neoplastic cell**
- necrotic cancer cell**
- normal cell**
- stromal cell**
- mutated genes**



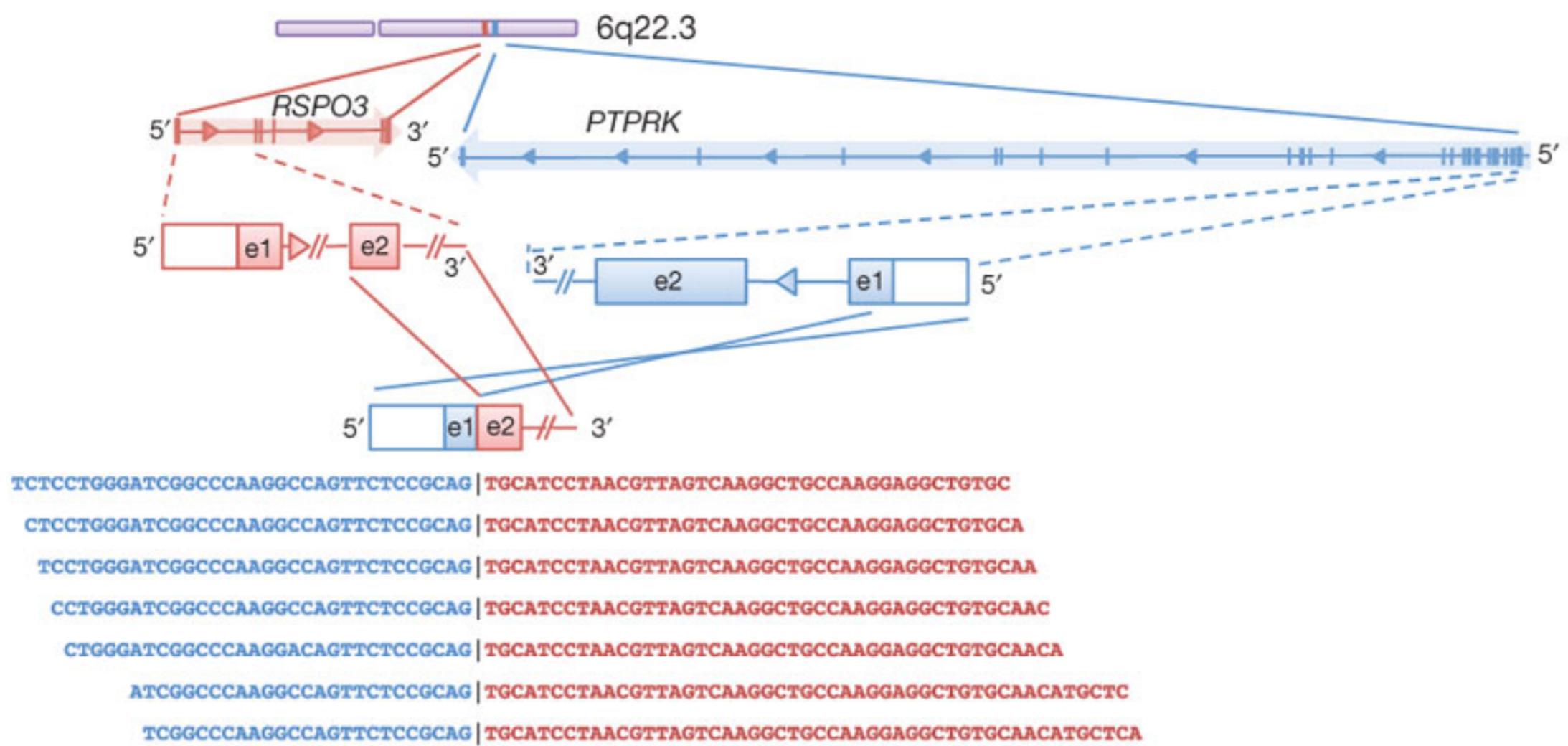


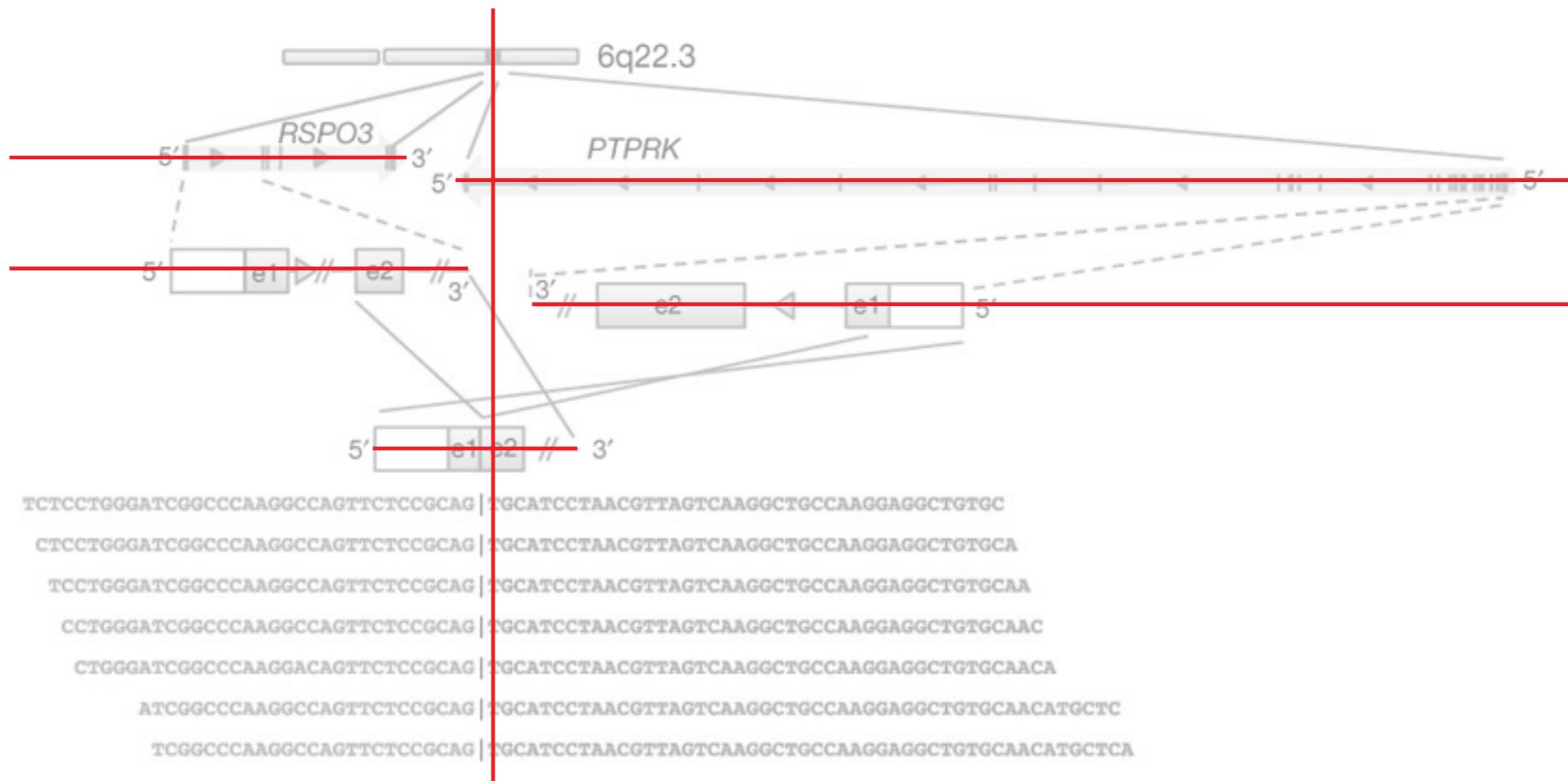
- cancer stem cell
- cancer cell
- AB mutated genes
- neoplastic cell
- necrotic cancer cell
- normal cell
- stromal cell

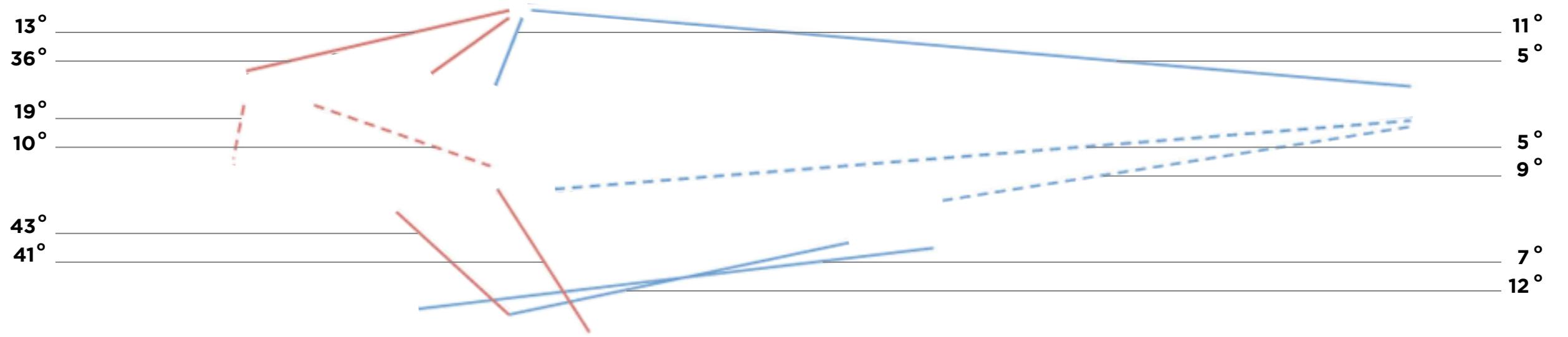


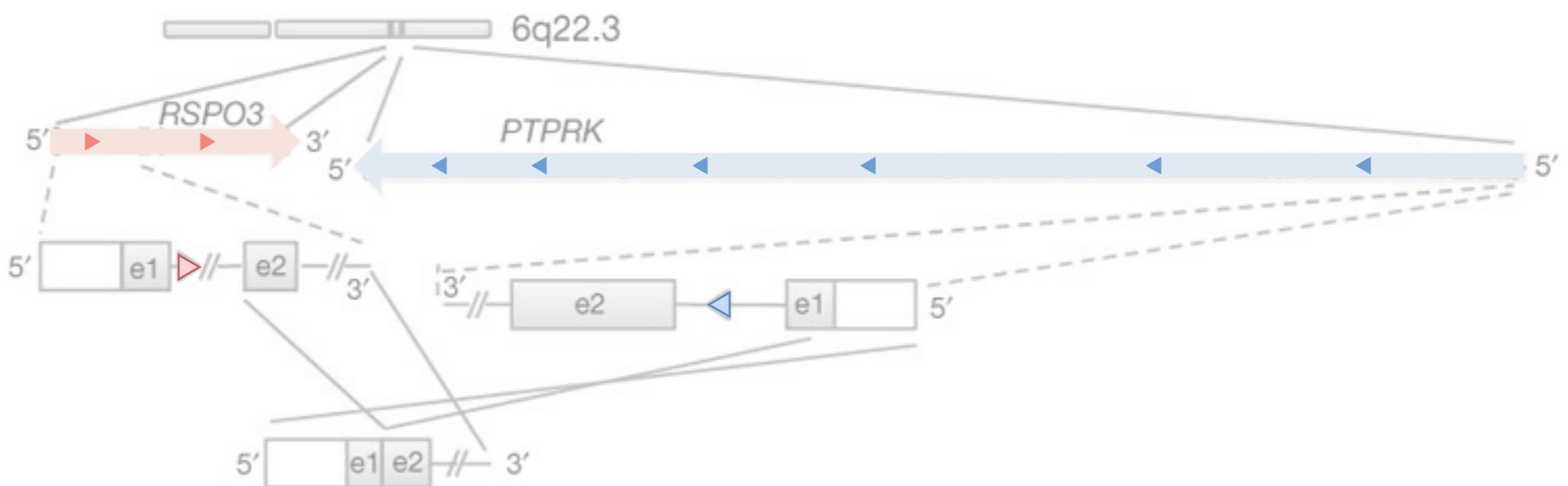
## REDESIGN—FROM LITERATURE

Assume your audience is smart but impatient.

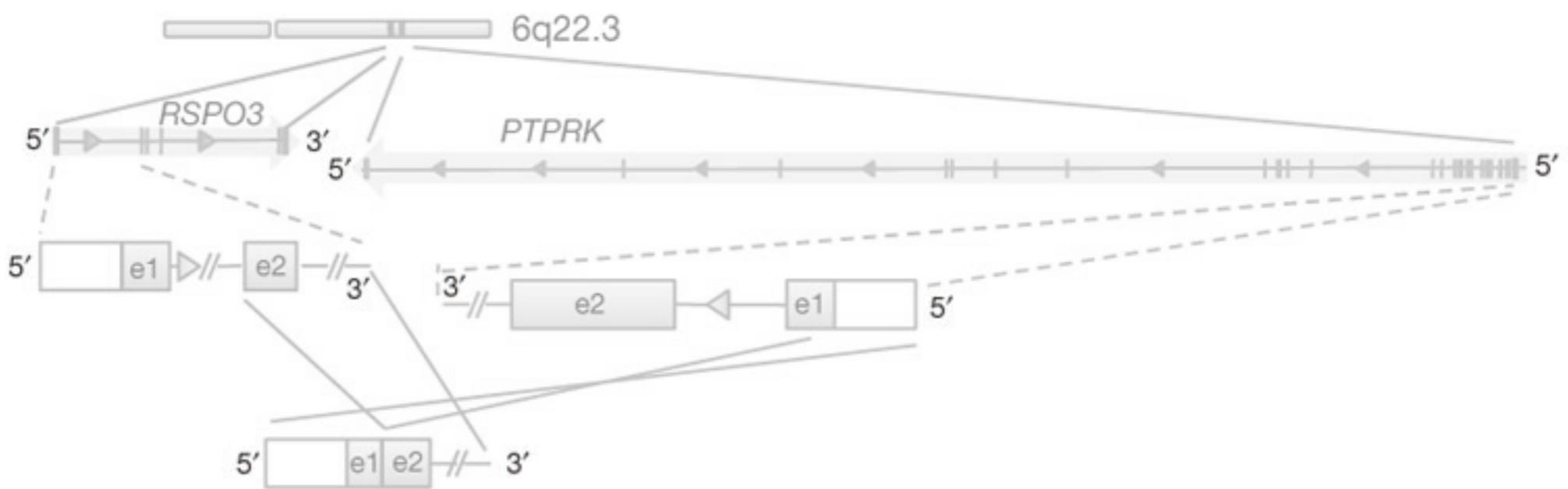




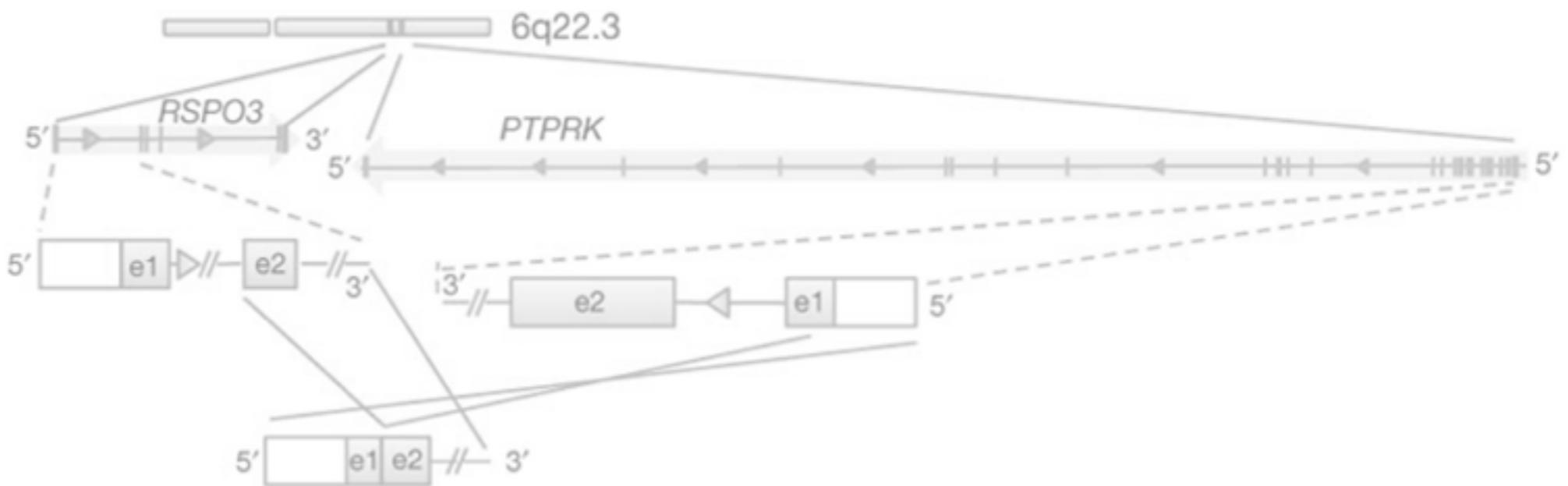




TCTCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGC  
 CTCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCA  
 TCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAA  
 CCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAAC  
 CTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACA  
 ATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTC  
 TCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTCA



TCTCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGC  
 CTCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCA  
 TCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAA  
 CCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAAC  
 CTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACA  
 ATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTC  
 TCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTCA



TCTCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGC  
 CTCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCA  
 TCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAA  
 CCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAAC  
 CTGGGATCGGCCAAGGACAGTTCTCCGCAG | TGCATCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACA  
 ATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTC  
 TCGGCCAAGGCCAGTTCTCCGCAG | TGCATCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTCA

what is the core message?  
structure and evidence of a gene fusion

what is important?  
gene name and orientation  
location of breakpoint  
change in orientation, if any  
local sequence context  
supporting evidence

what is not important, or peripheral?  
gene size  
gene location  
gene model (learn to let go)

## REFERENCE

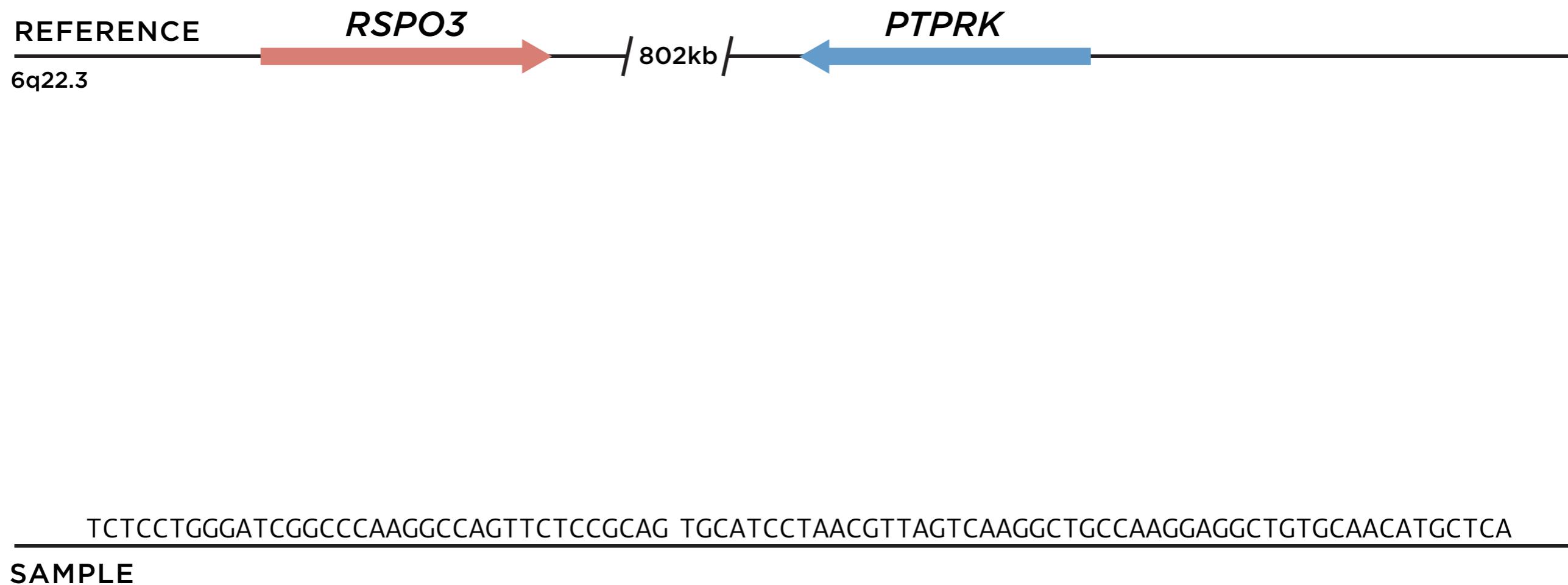
---

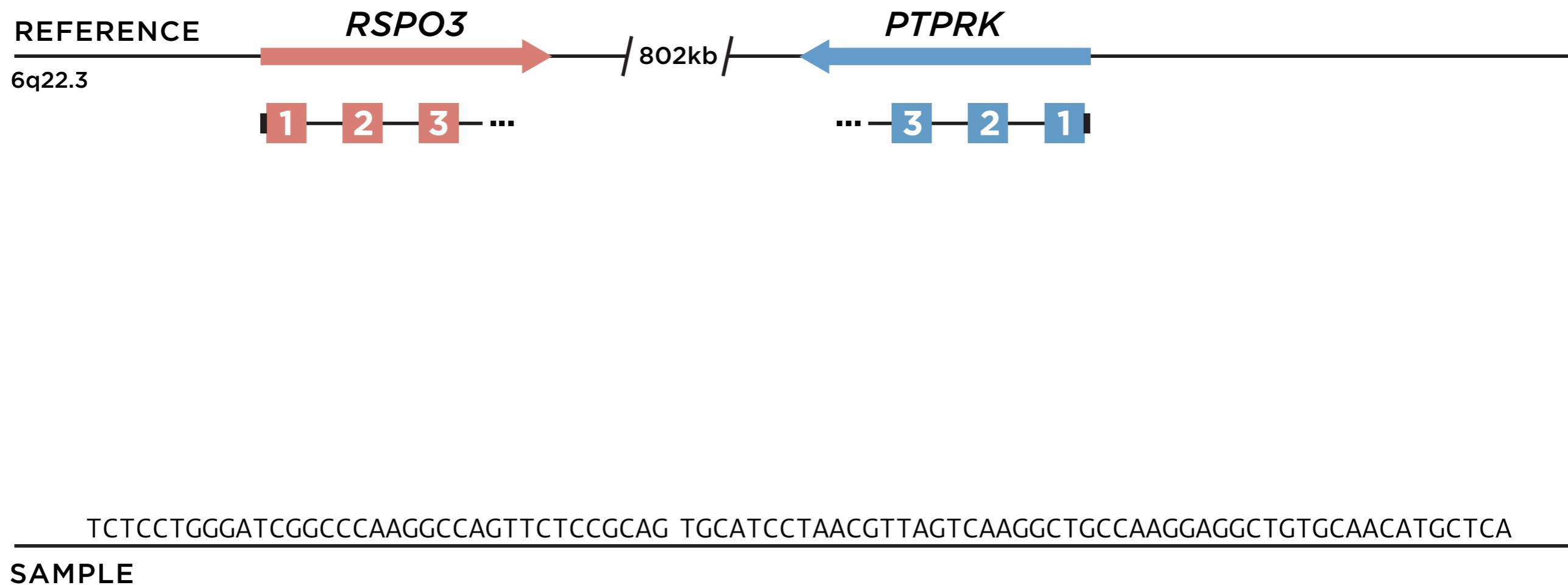
6q22.3

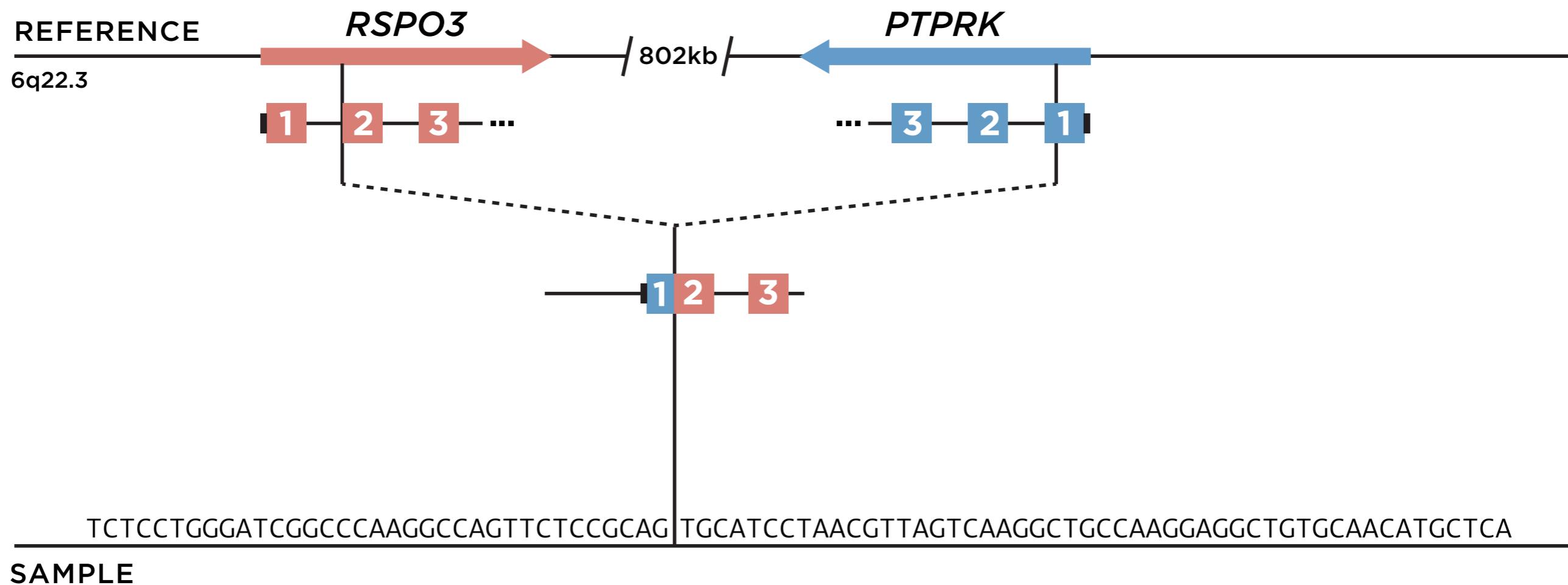
---

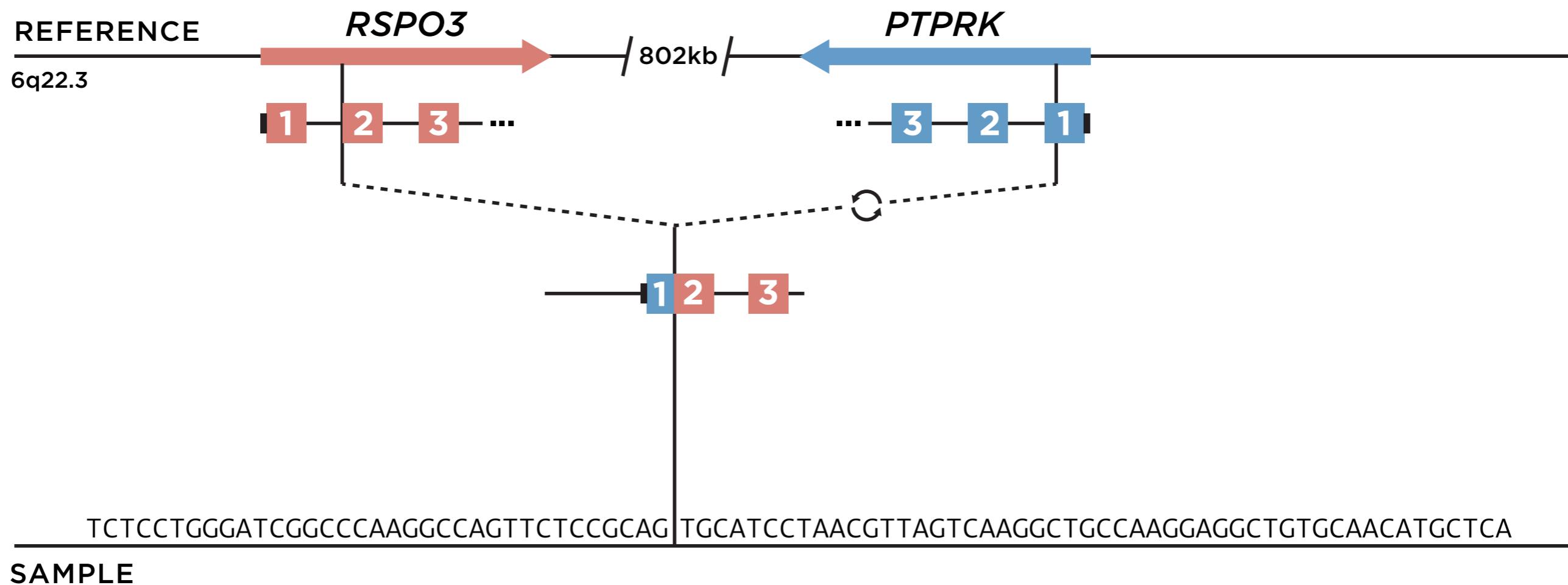
TCTCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTCA

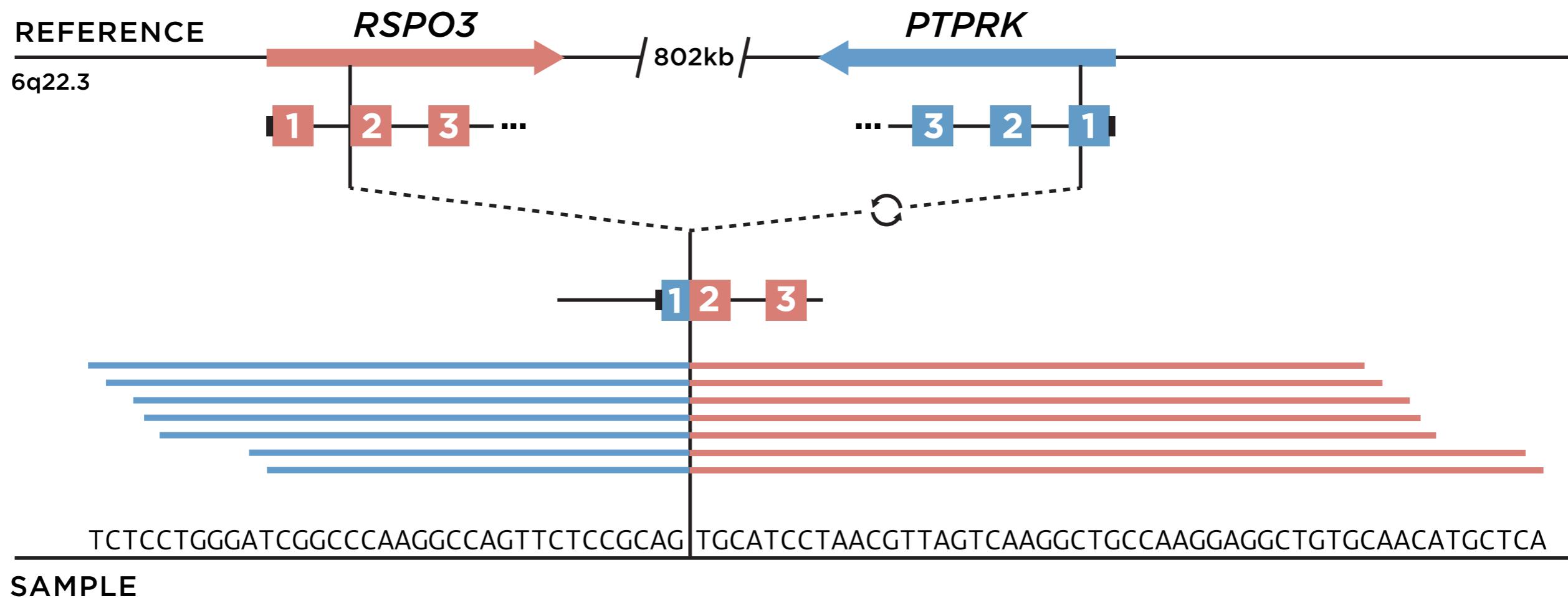
## SAMPLE

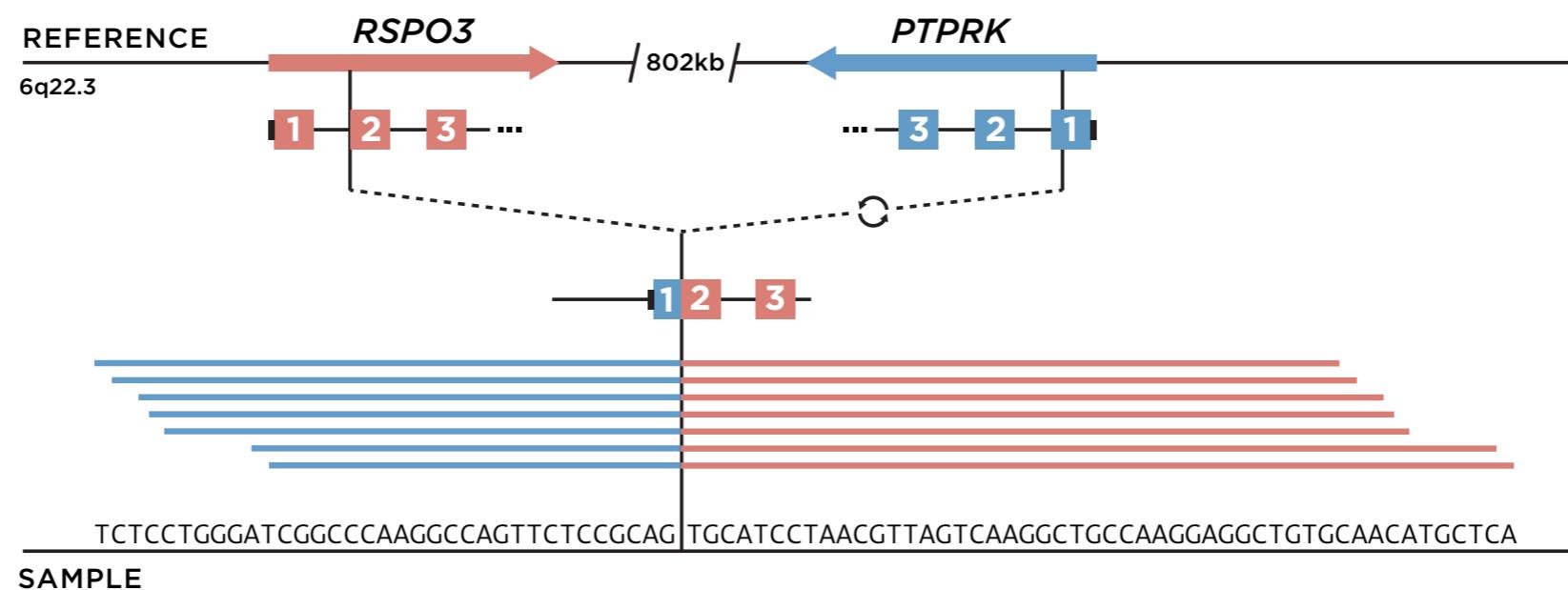
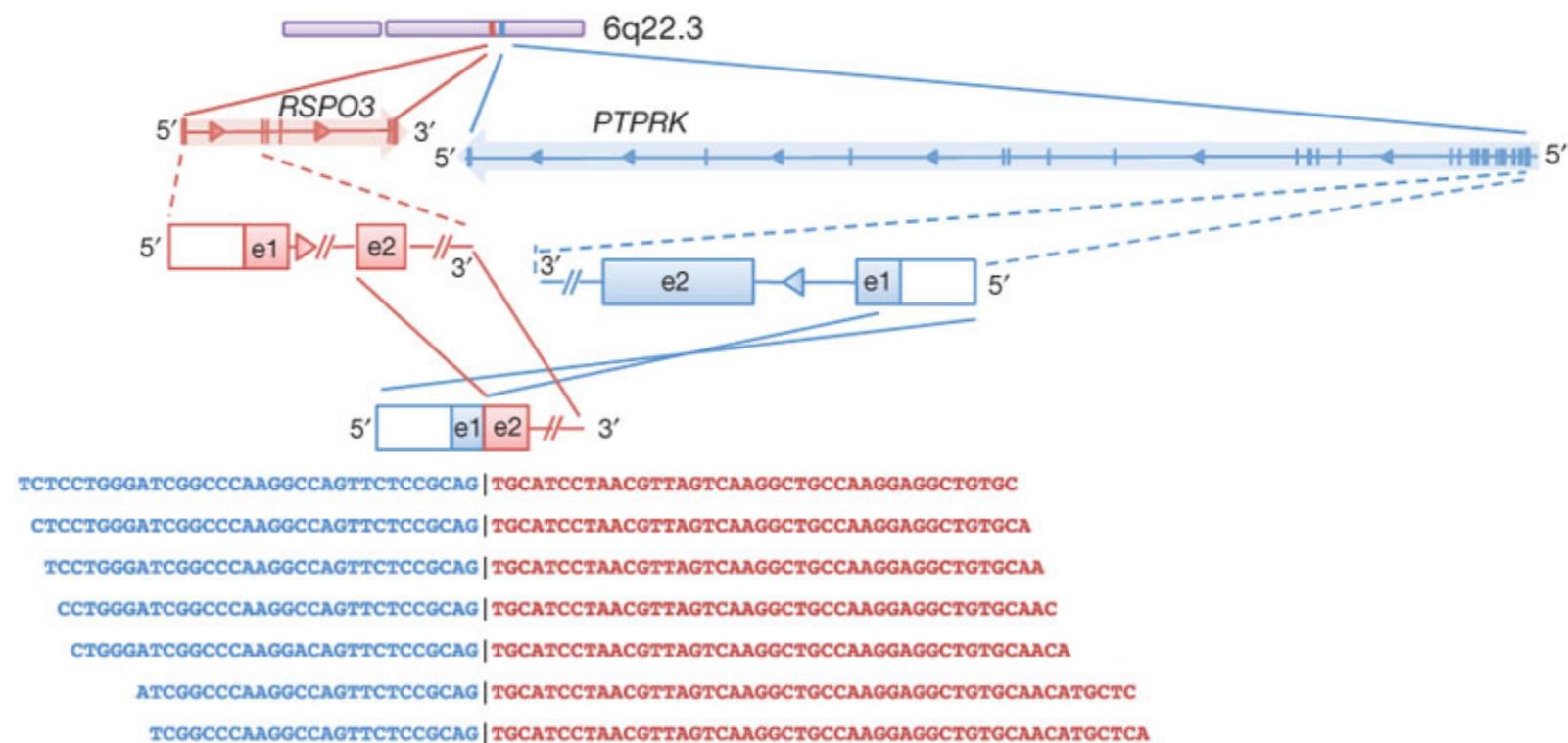












# Further reading

---

- VIZBI conference videos and tutorials:  
<http://www.vizbi.org>
- Visualization principles tutorial by Jessie Kennedy, Cydney Nielsen and Martin Krzywinski: <http://mkweb.bcgsc.ca/vizbi/2012/>
- Points of View Column in Nature Methods by Bang Wong: <http://bang.clearscience.info/?p=546>

# Exam-like questions

---

- Mention at least 3 points that you should focus on when you visualise your data.
- Mention at least 3 points that you should avoid in visualization. Also provide an example to your points.
- What is the problem with the colours? How can you tackle it?
- Why one should not use 3D figures?
- Exercise: you need to “refactor” a figure