

CANADA'S MICHAEL SMITH
GENOME
SCIENCES
CENTRE

WWW.BCGSC.CA

visual design principles

COMMUNICATING CLEARLY

MARTIN KRZYWINSKI

CANADA'S MICHAEL SMITH GENOME SCIENCES CENTER
BC CANCER RESEARCH CENTER

**This is an extended version
of my [VIZBI 2013](#) keynote presentation.**

vizbi.org/2013

Additional redesign examples are included.

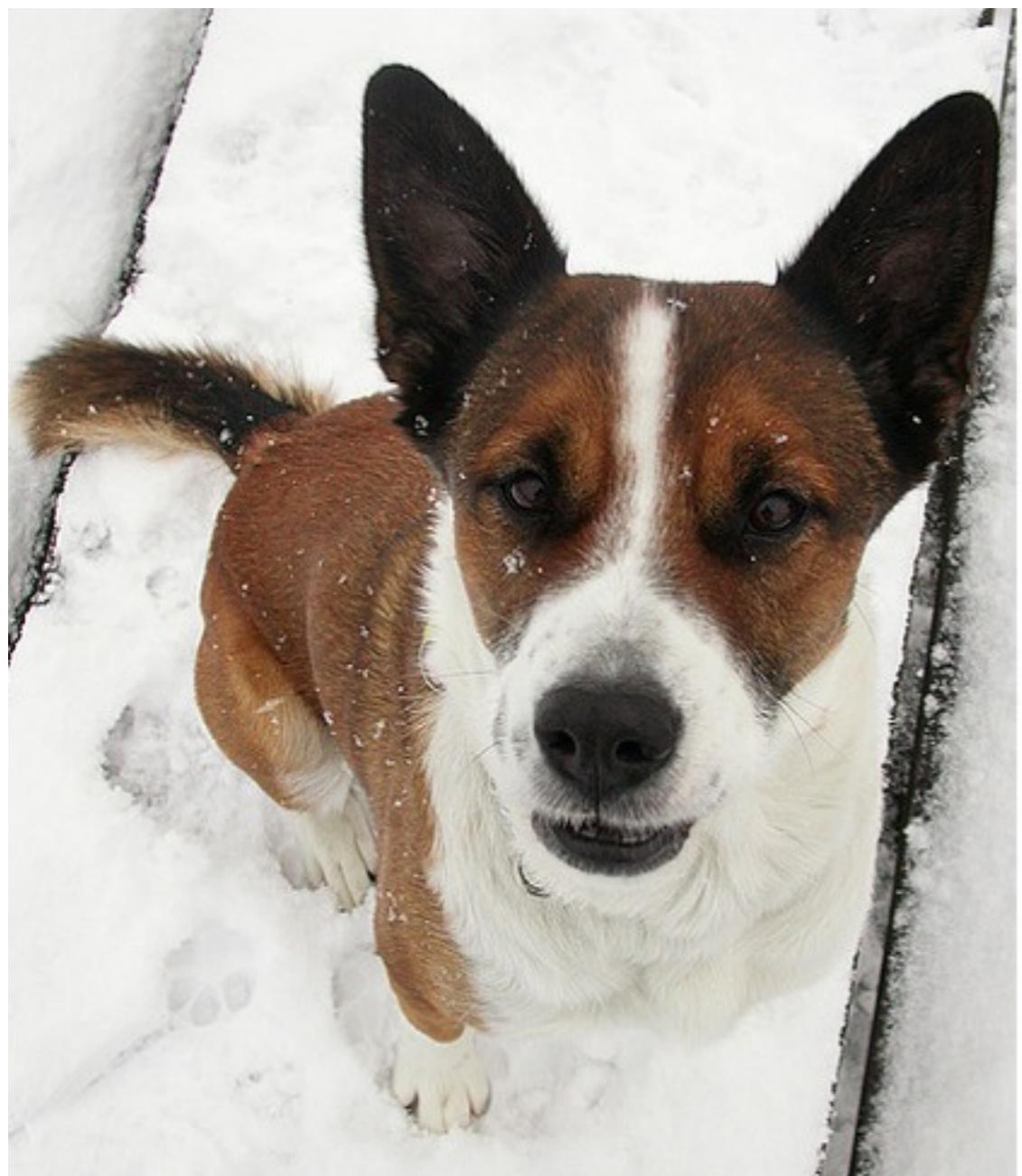
YOU WANT TO BE A BETTER COMMUNICATOR NOT A BETTER ARTIST

Do not rely solely on your personal aesthetic.

Strive for simplicity and clarity.

OBJECTIVE ASPECTS OF DESIGN

It's not all about taste.



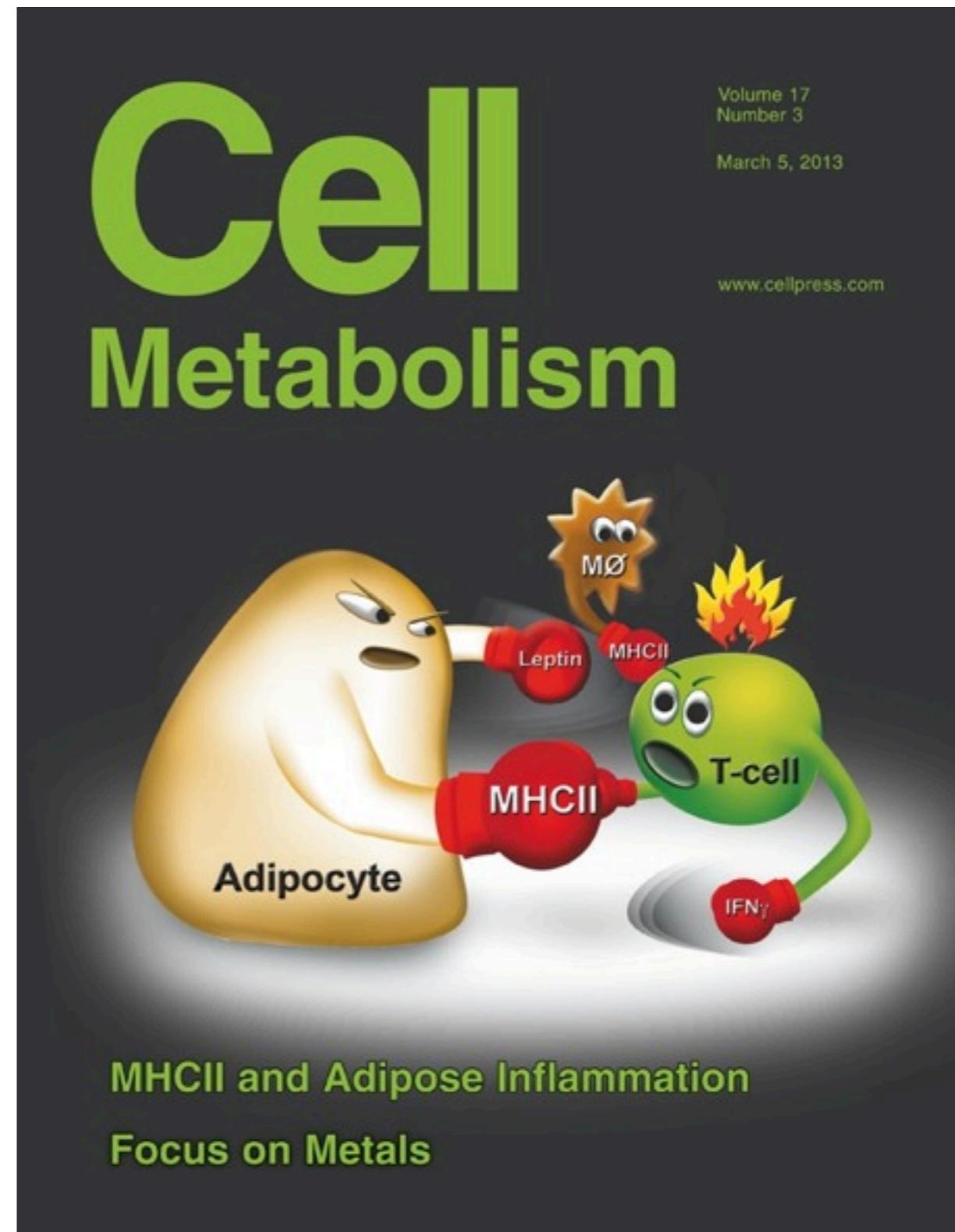
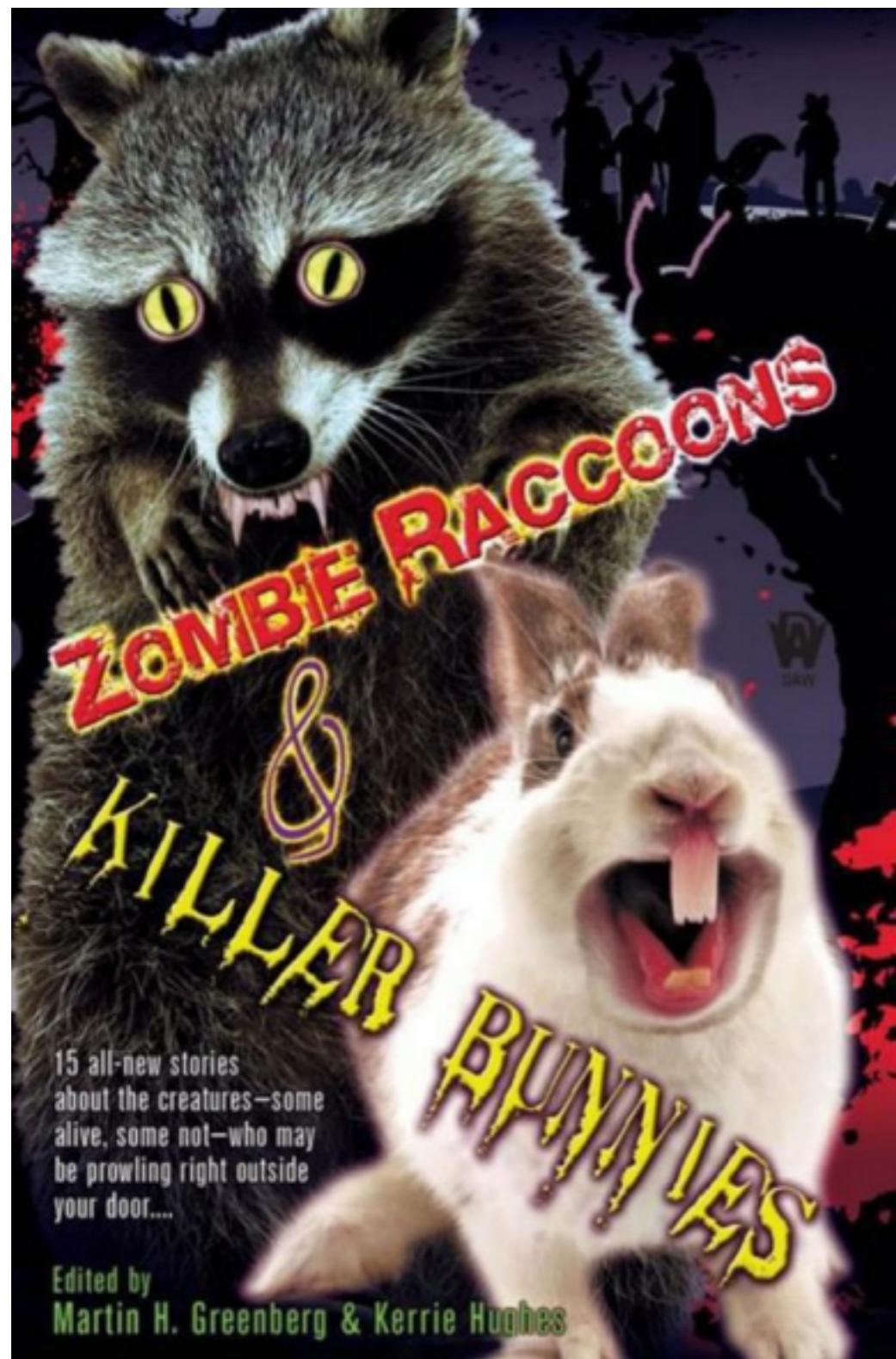


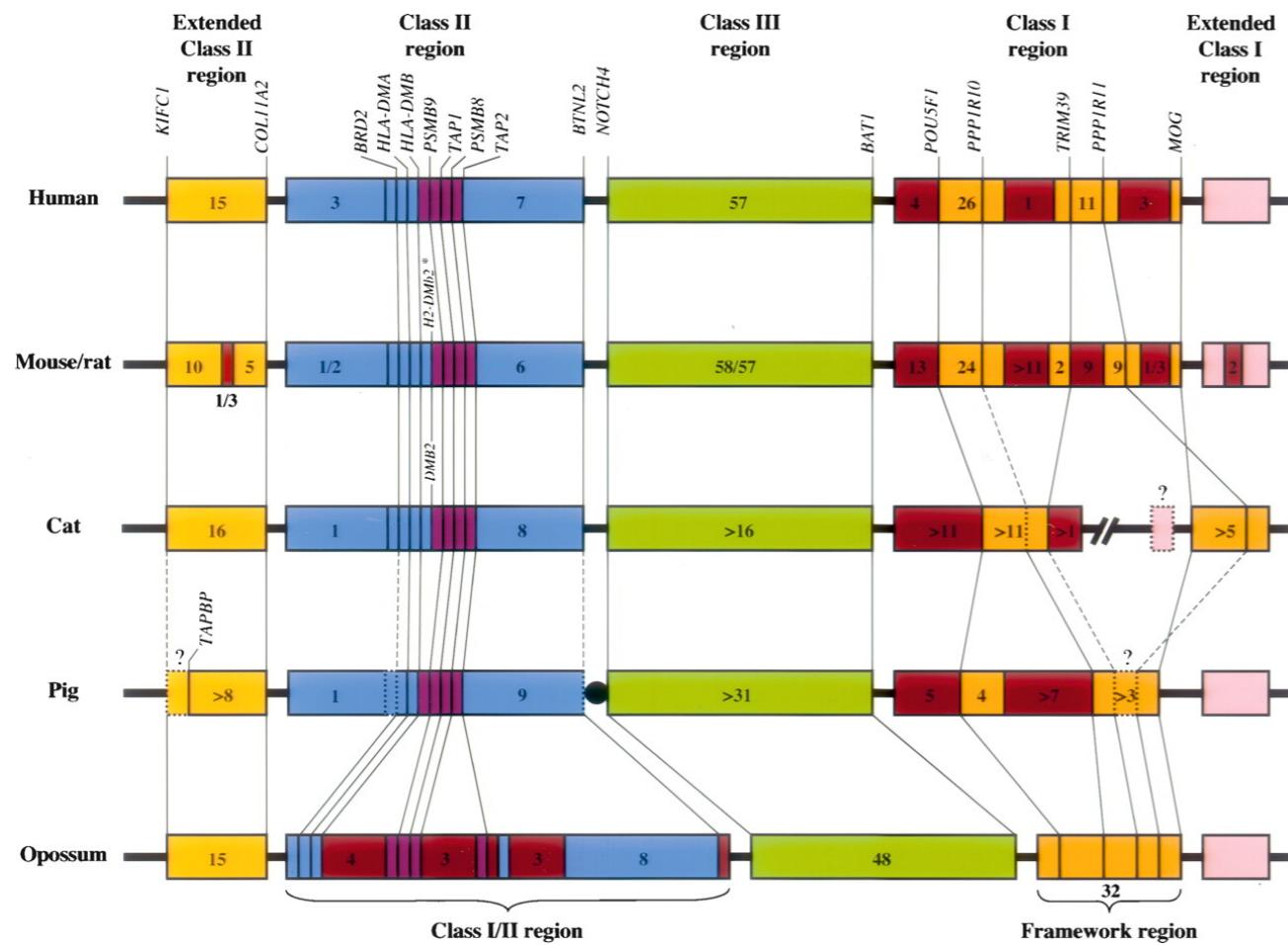
THE MUSEUM of BAD ART

MASTERWORKS



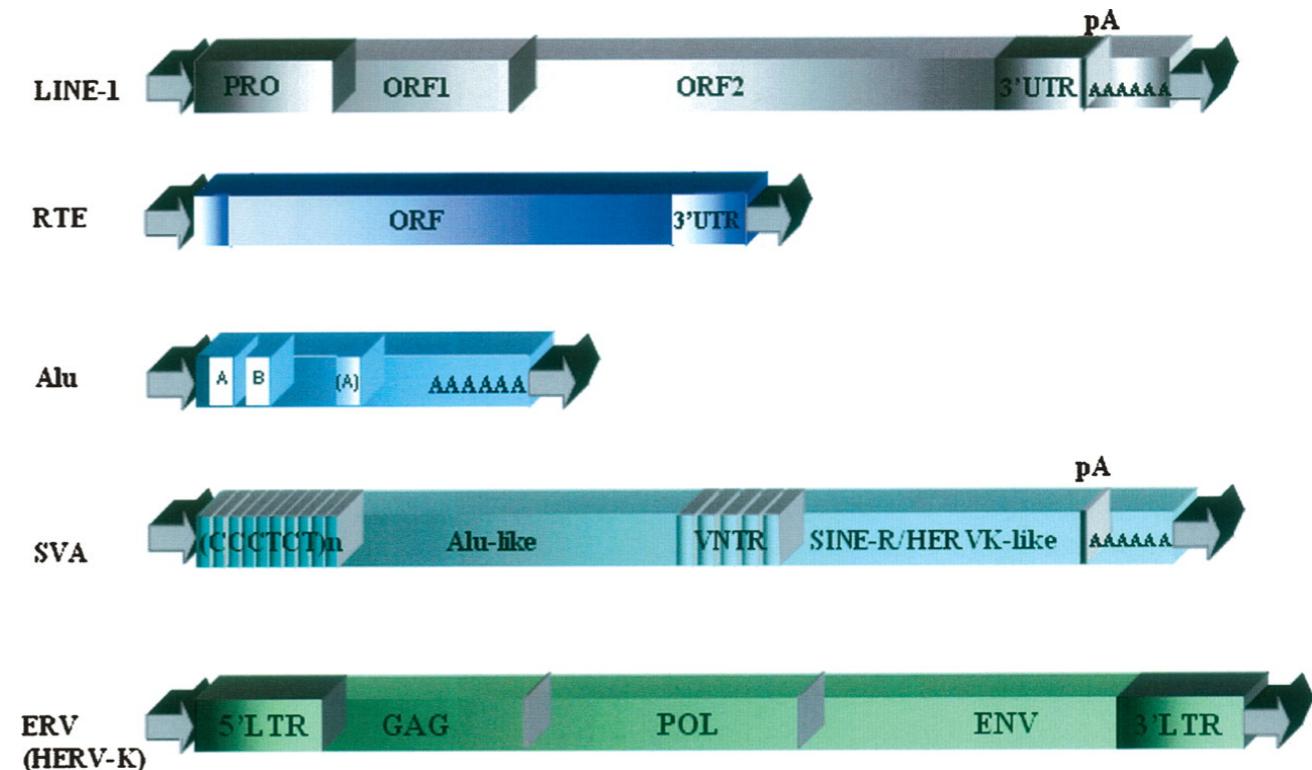
MICHAEL FRANK and LOUISE REILLY SACCO





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.

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.

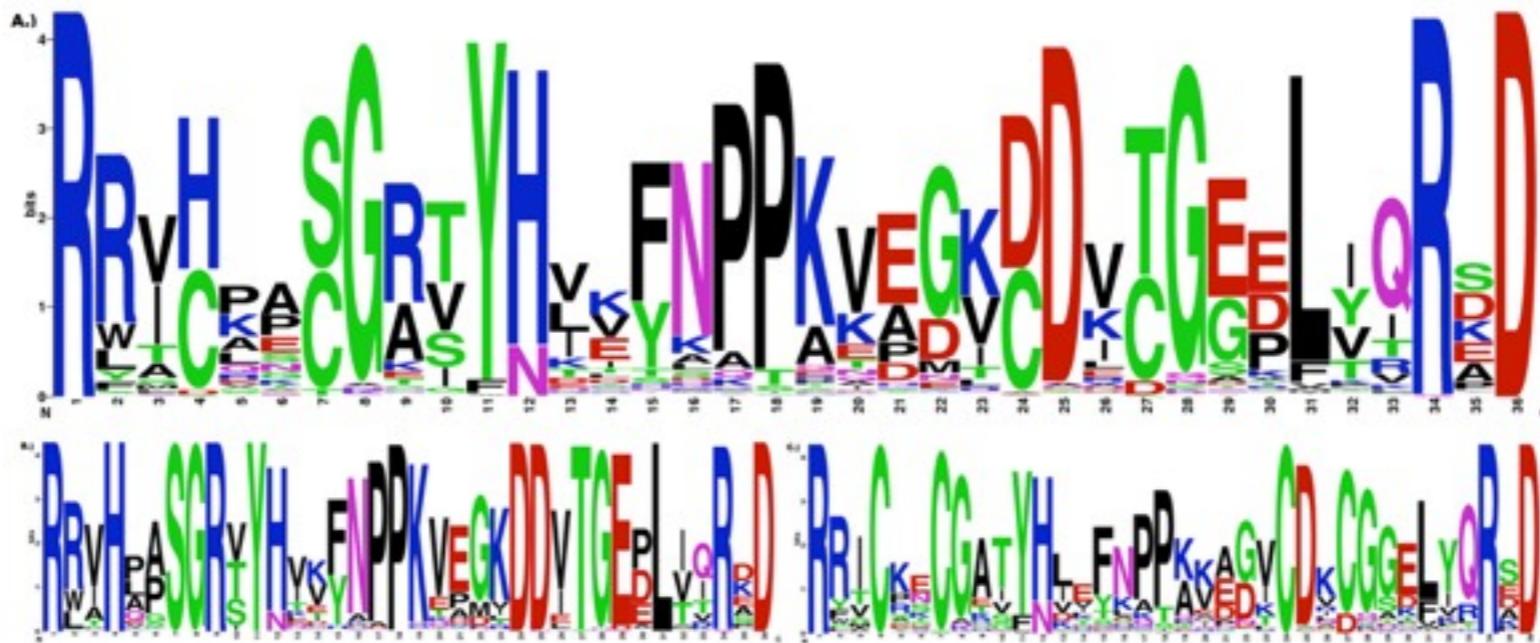
you have data

you may have a message

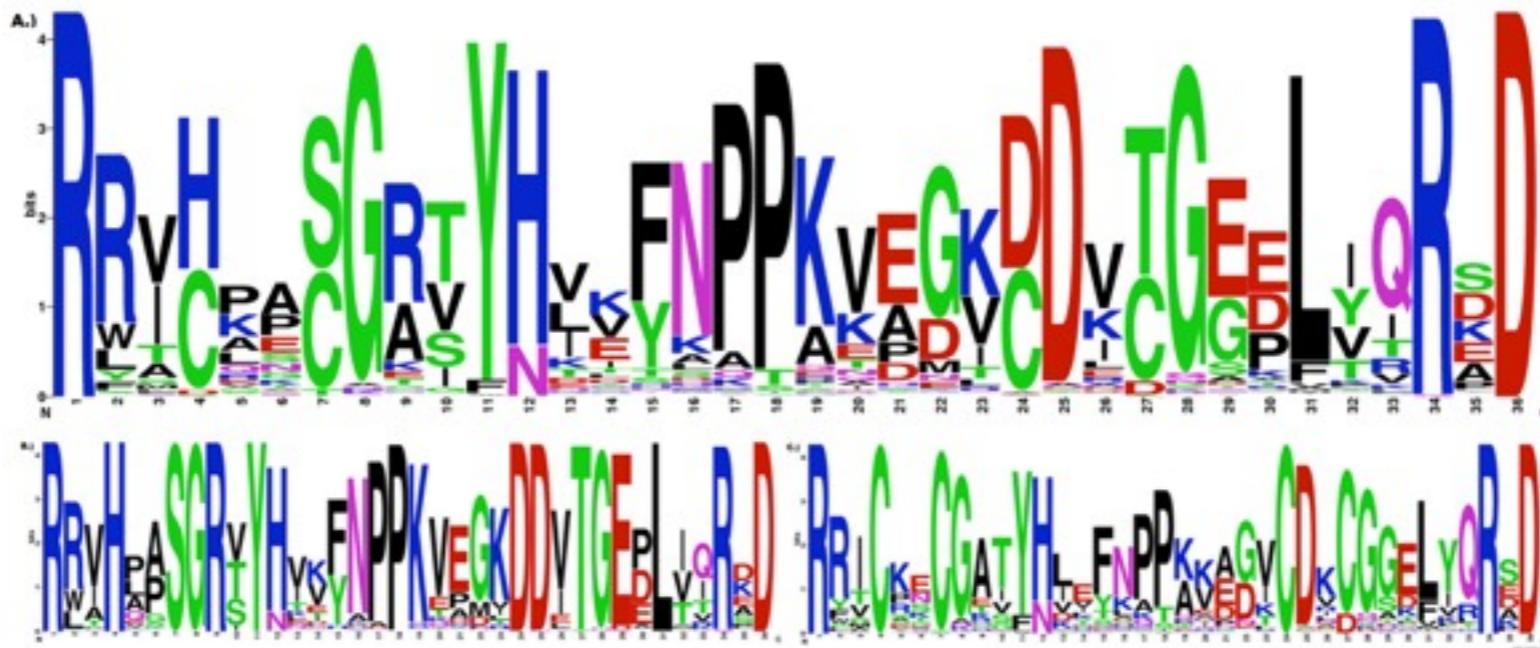
**you need to create a figure
(or think you do)**

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

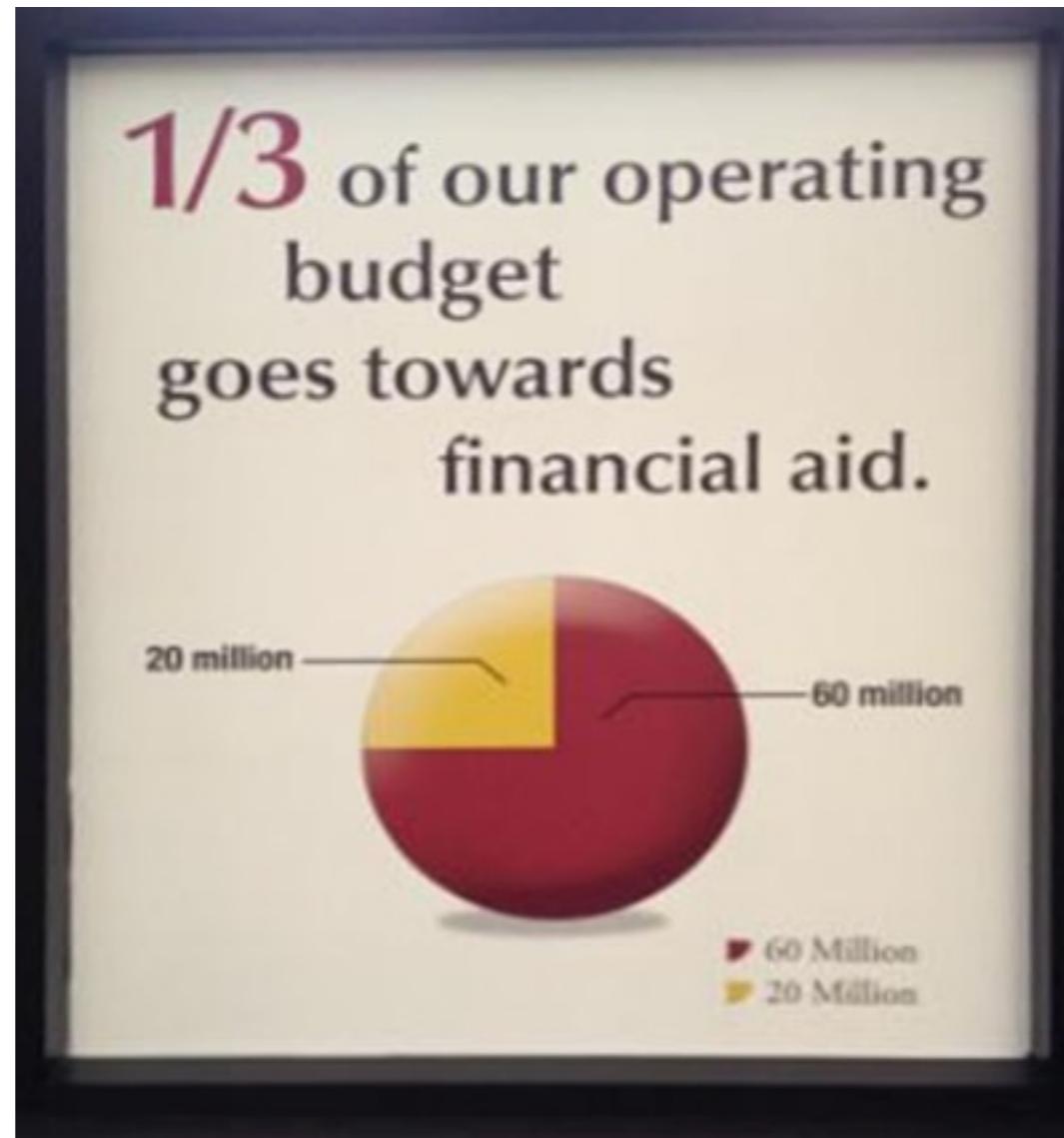
**PLEASE DO NOT
THROW STONES
AT THIS SIGN**

thank you





IS ABSOLUTE ACCURACY ALWAYS IMPORTANT?



NO



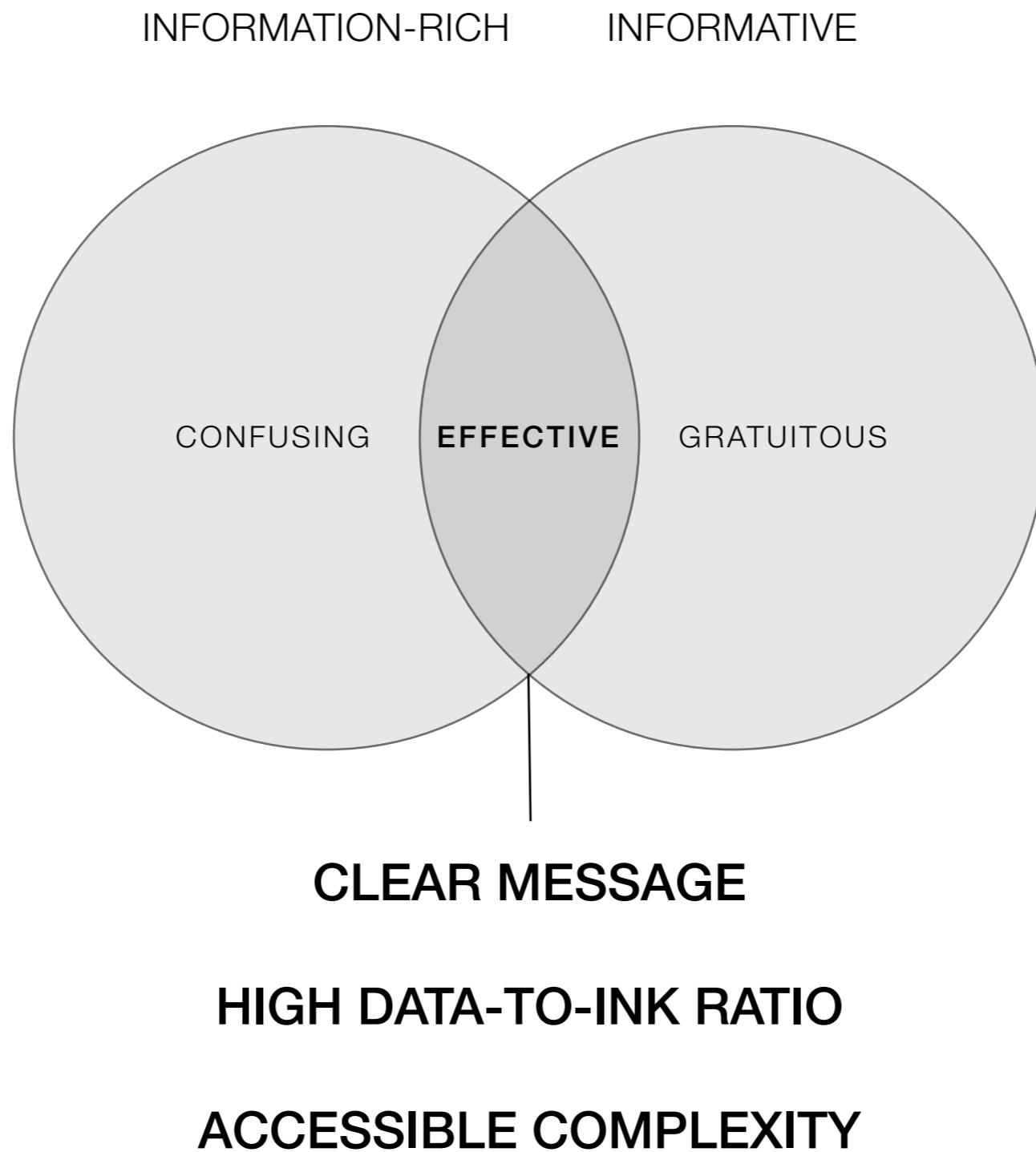
accuracy does not always add to utility

INFORMATIVE + INFORMATION-RICH

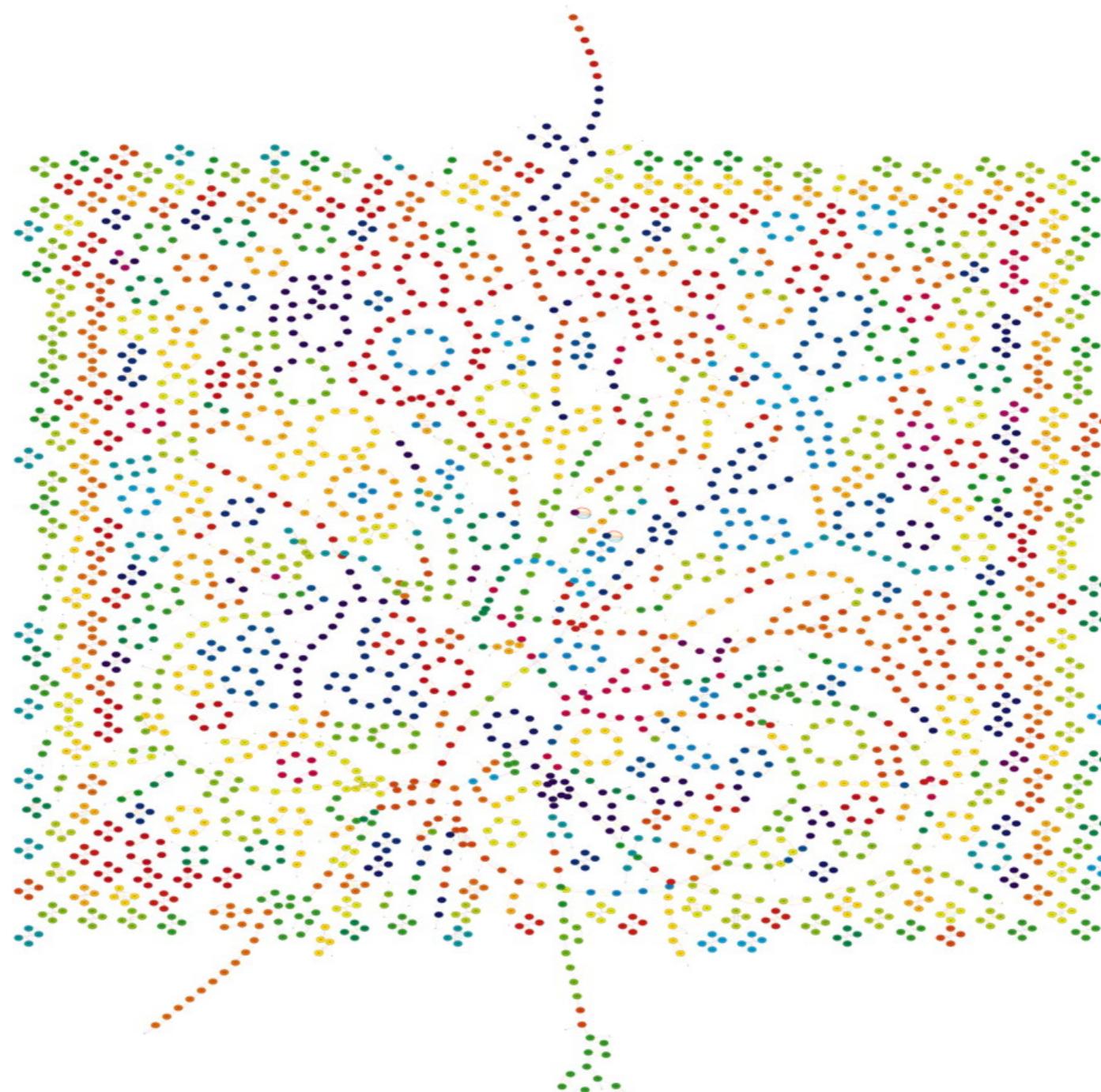
In the age of big data, figures should be worth more than 1,000 words.

**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



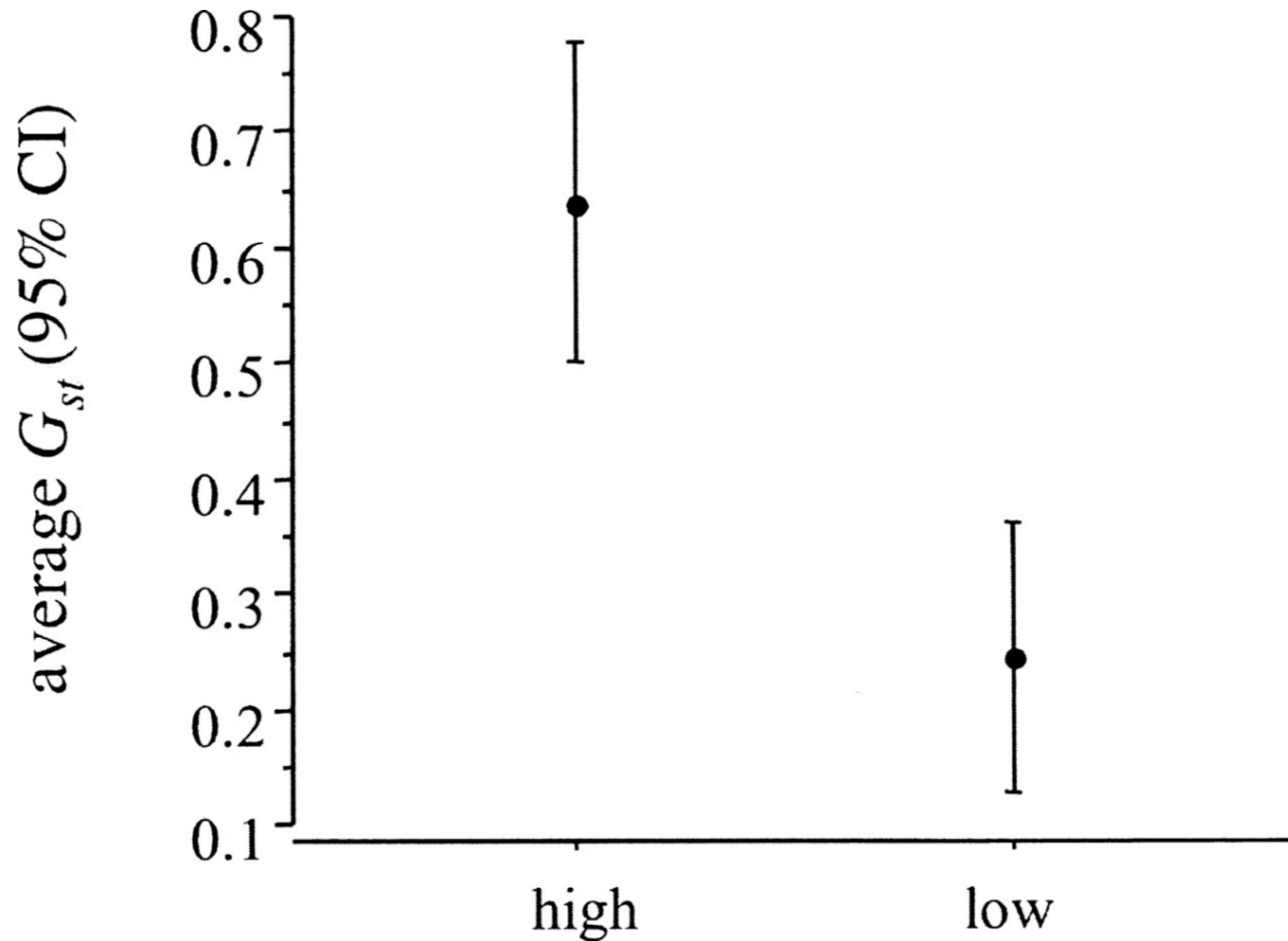
INFORMATION-RICH, NOT INFORMATIVE



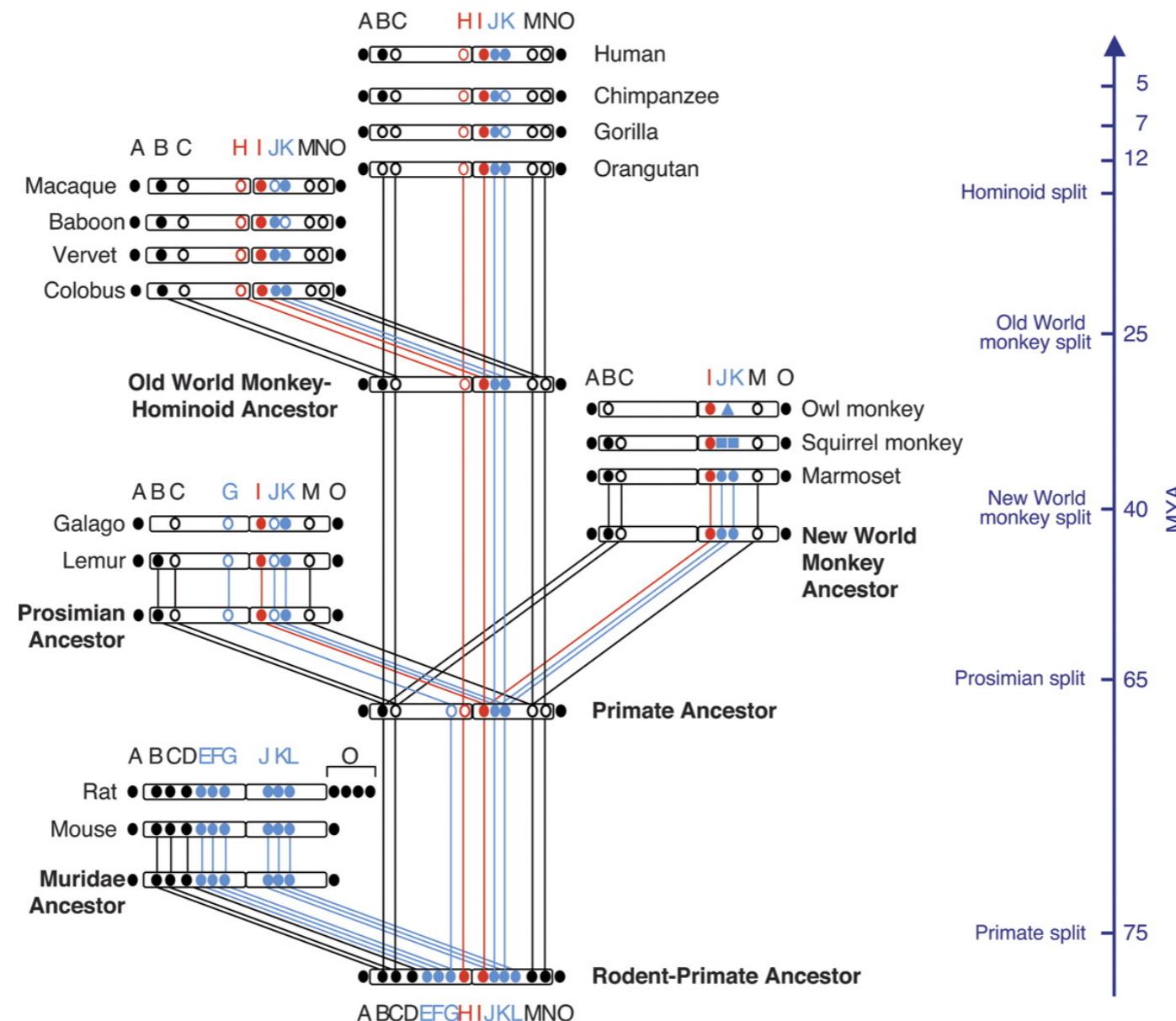
Chromosome colors:

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

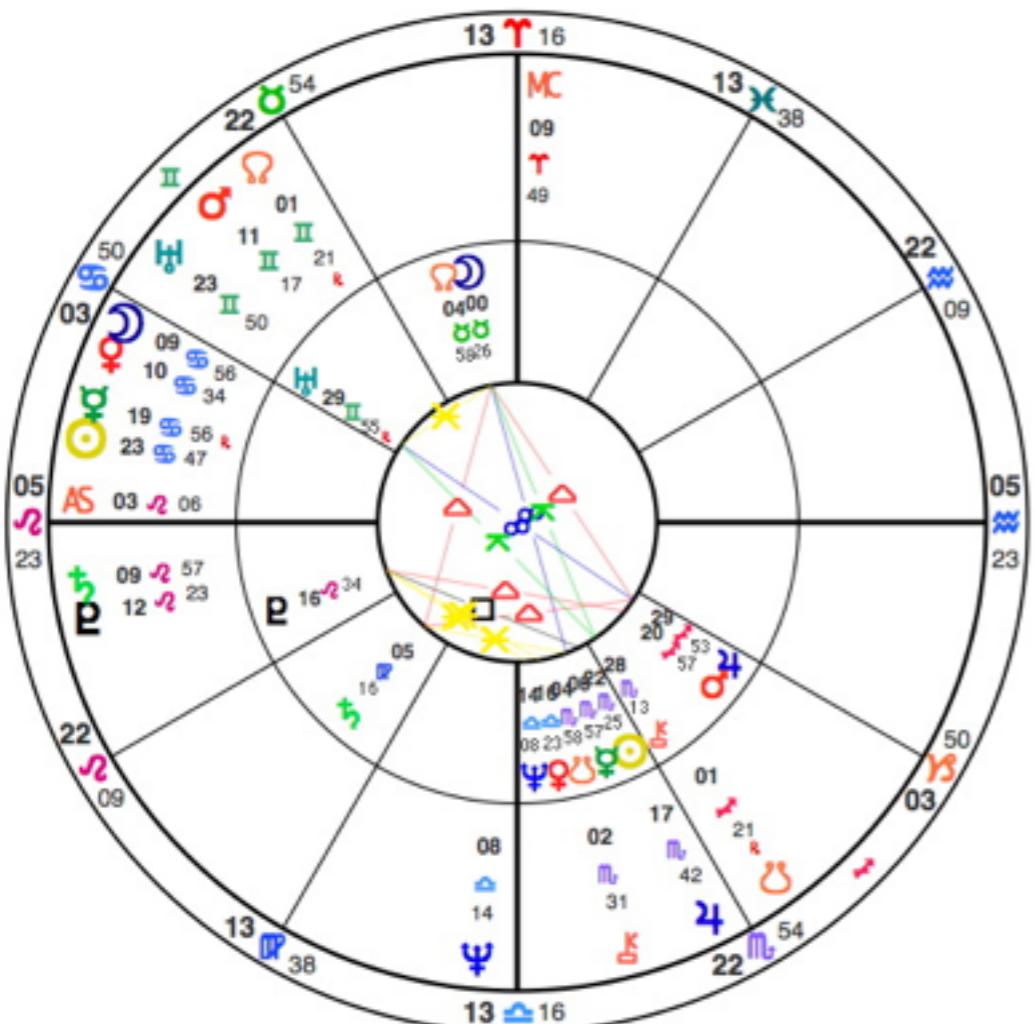


INFORMATION-RICH AND INFORMATIVE



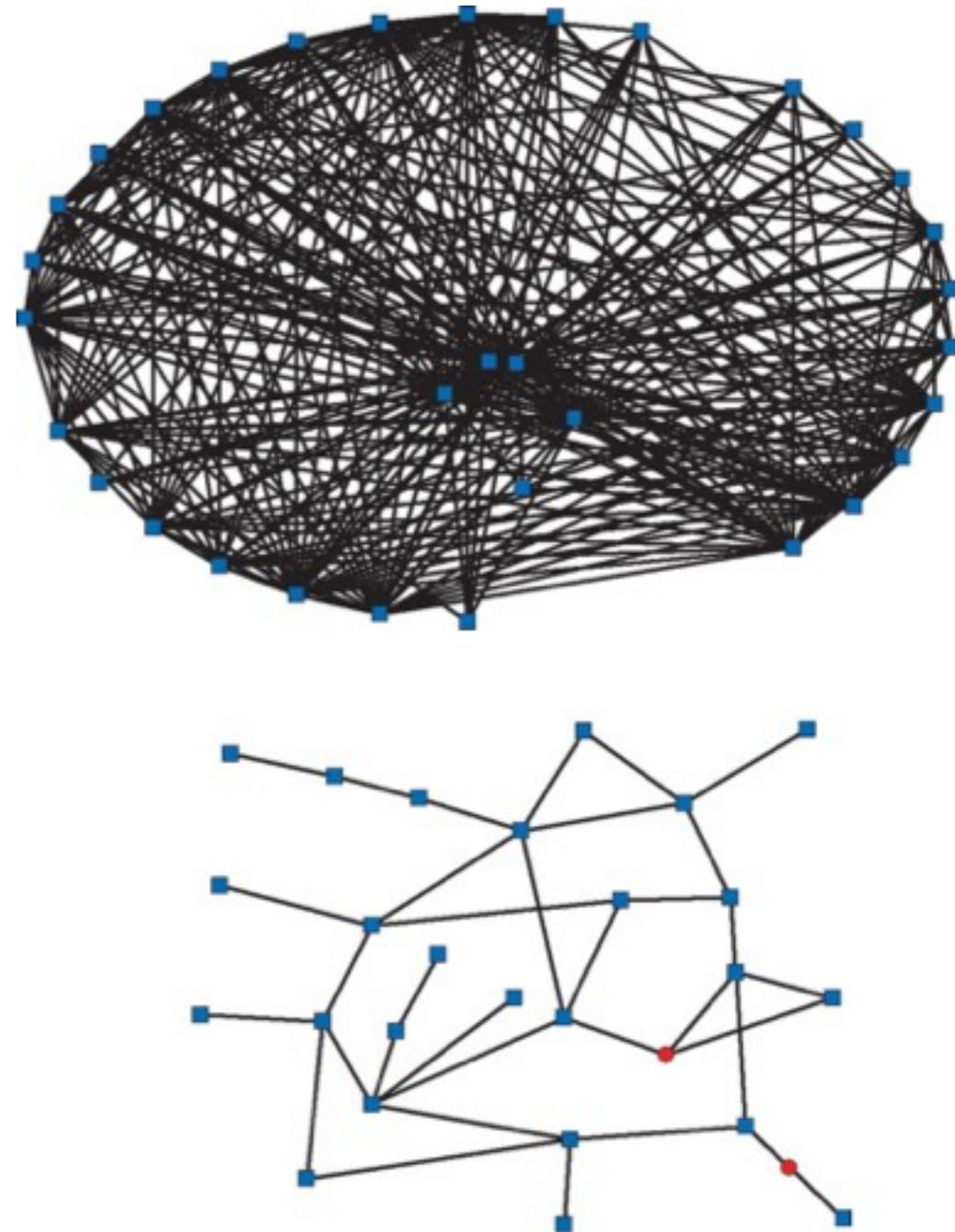
JUNK SCIENCE

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



REAL SCIENCE

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



don't merely display data
explain it

KNOW YOUR MESSAGE

Stick to it.

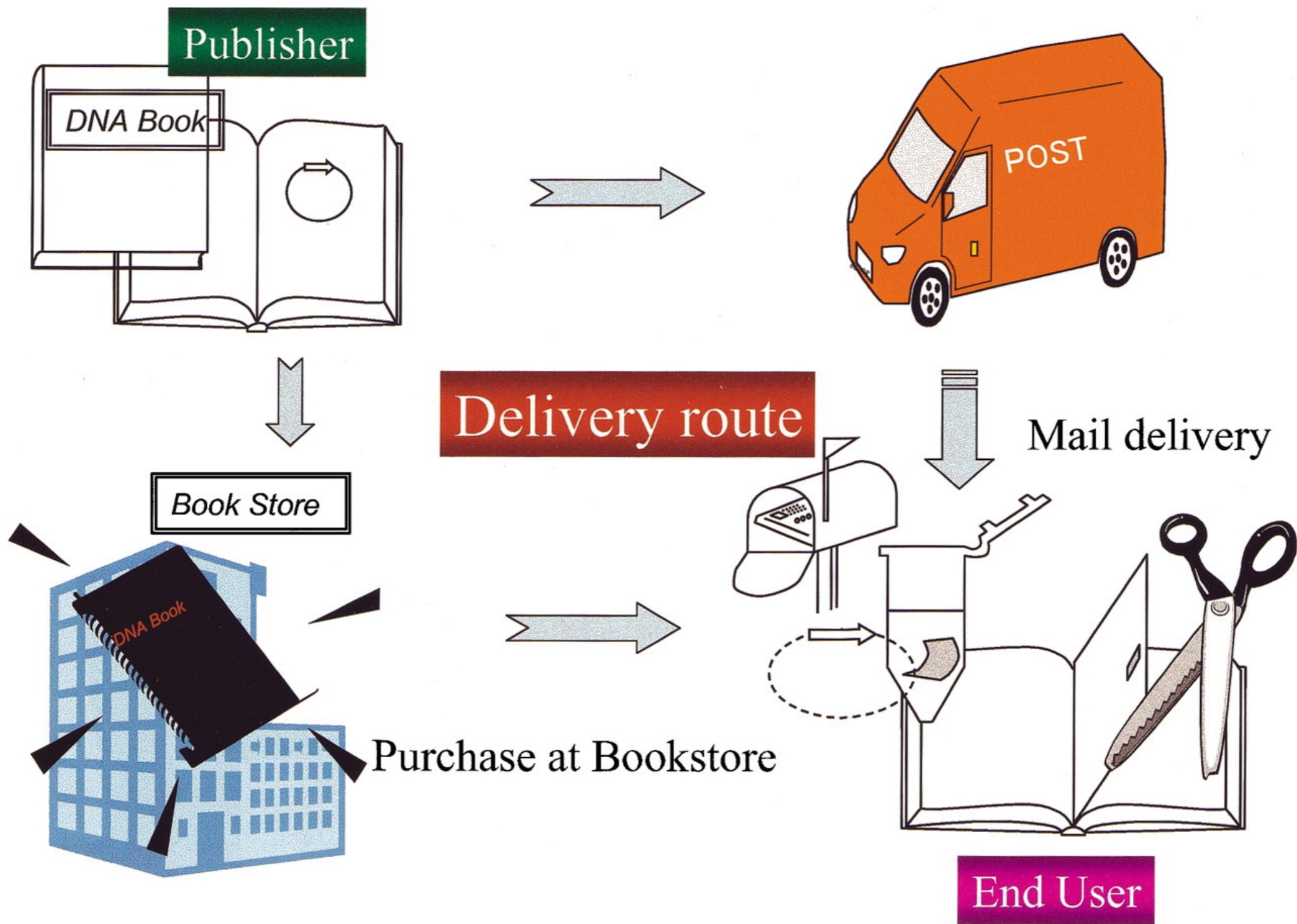
CAUTION
THIS SIGN HAS
SHARP EDGES

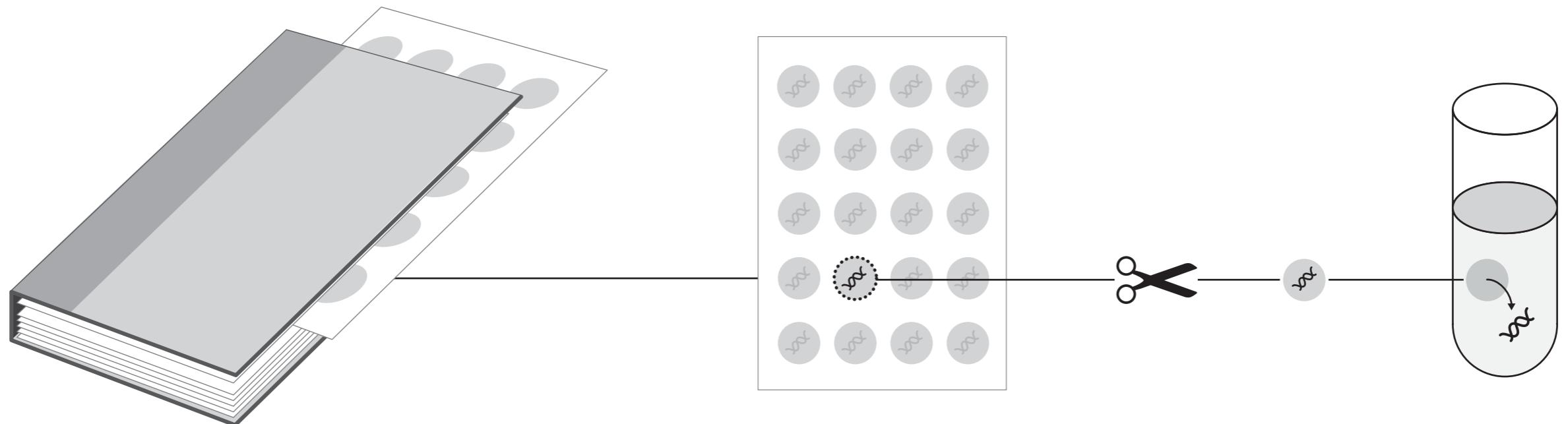
DO NOT TOUCH THE EDGES OF THIS SIGN



ALSO, THE BRIDGE IS OUT AHEAD







Kawai, J. and Y. Hayashizaki, DNA book. Genome Res, 2003. 13(6B): p. 1488-95.
Fig 1. redesigned by M Krzywinski

ACHIEVE FOCUS BY AGGREGATING

WHAT IS SHOWN?

RAW DATA

12 54 82 29 25 22 67 61 23 79

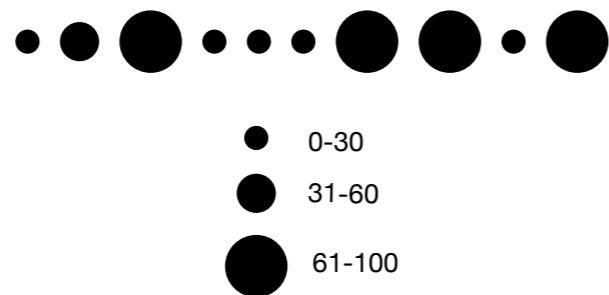
WHAT IS COMMUNICATED?

NO CLEAR MESSAGE.

WHAT IS INTERPRETED?

UNKNOWN. READER IS
ON THEIR OWN.

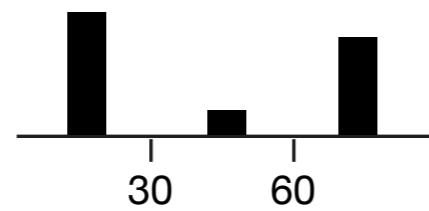
DISCRETIZED



SCALE

THREE RANGES ARE IMPORTANT.
INDIVIDUAL VALUES WITHIN A
RANGE ARE NOT.

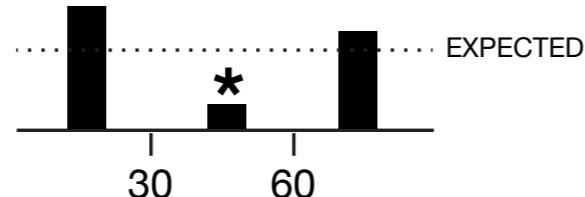
BINNED



DISTRIBUTION

THERE ARE FEWER
MEDIUM-SIZED VALUES.

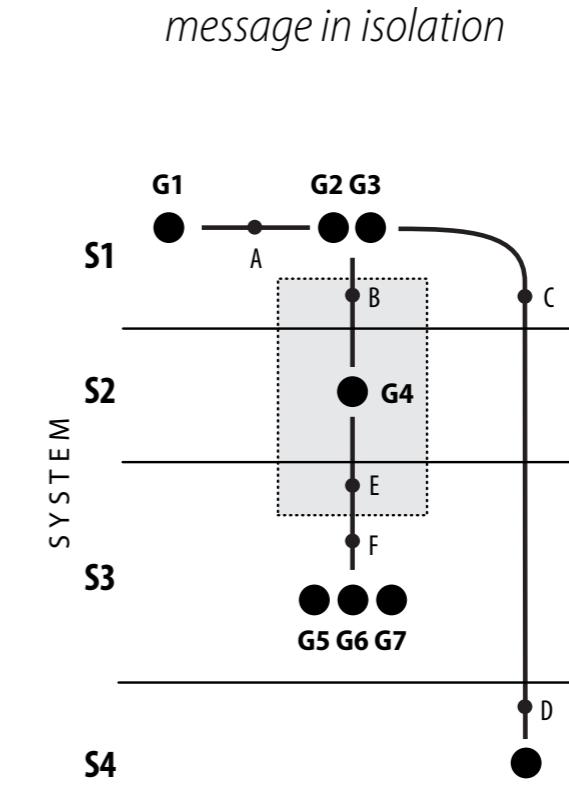
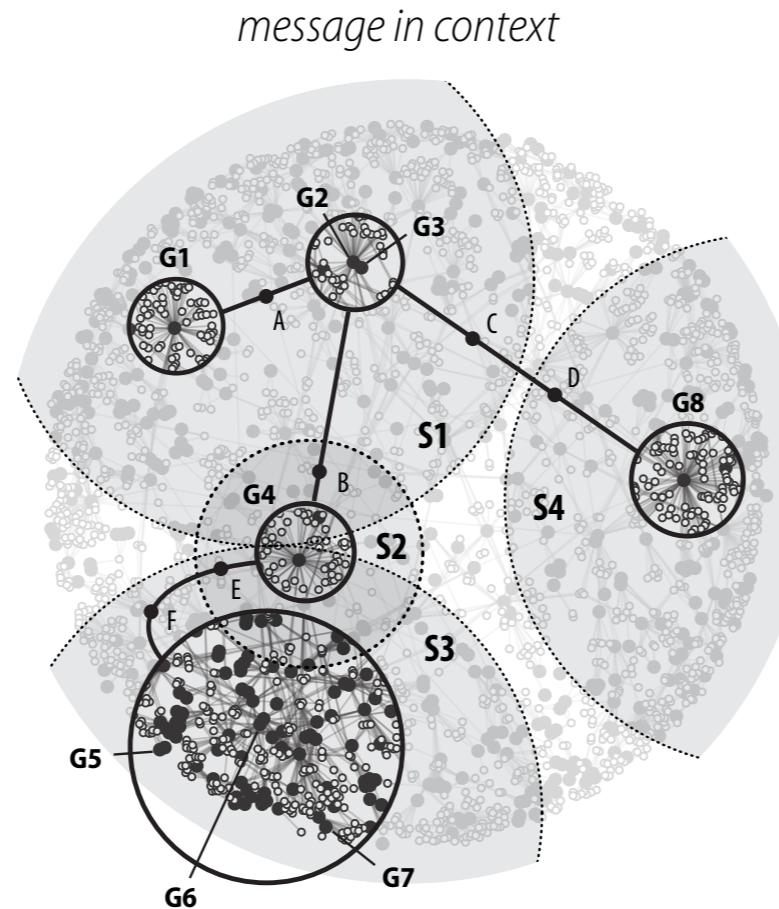
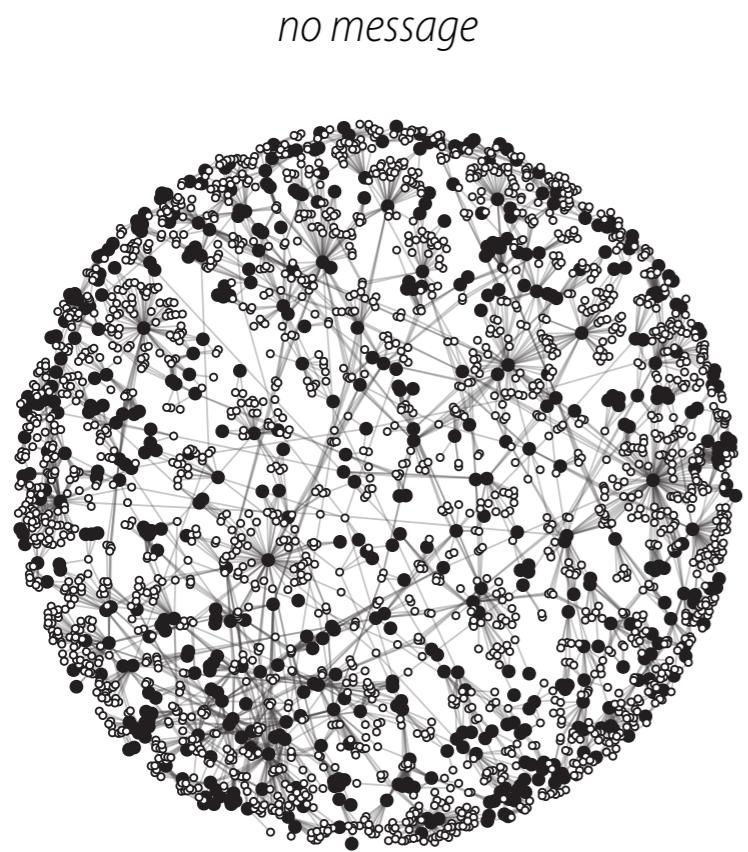
TREND



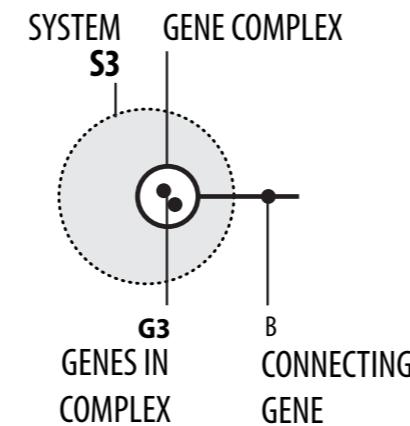
SIGNIFICANCE

THERE ARE SIGNIFICANTLY
FEWER MEDIUM-SIZED VALUES.

CONTEXT MUST NEVER DILUTE MESSAGE



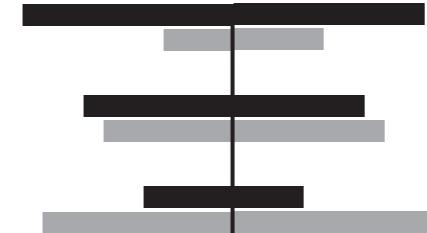
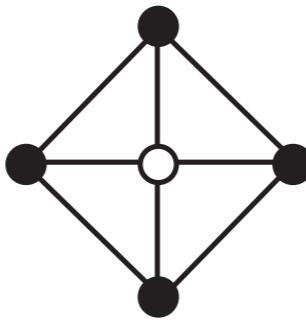
● GENE ○ DISEASE



● GENE COMPLEX
— CONNECTING GENE

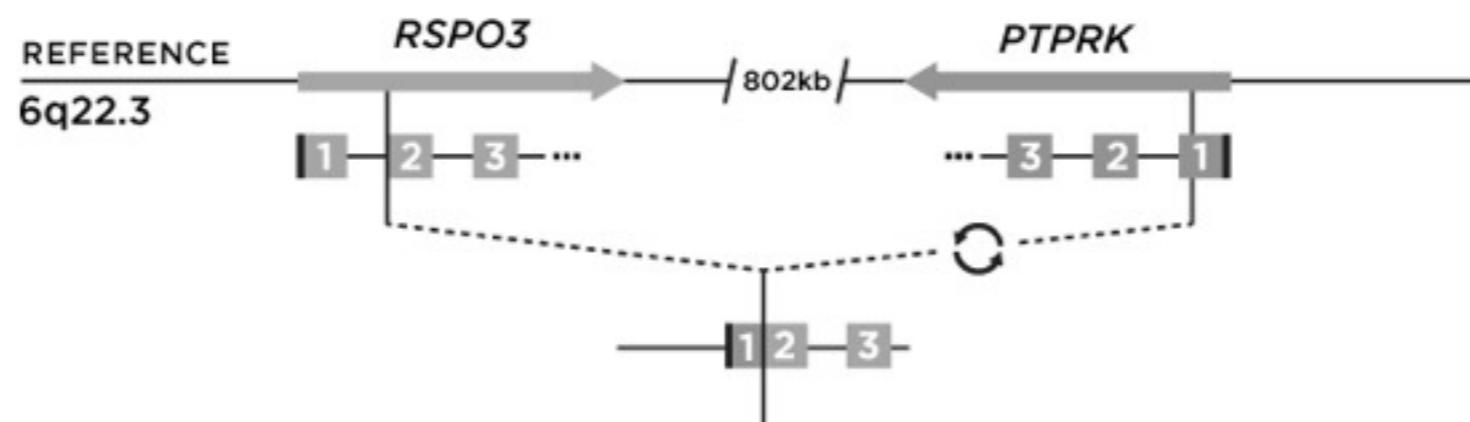
STRIVE FOR CLEAR COMMUNICATION

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

INFORMS



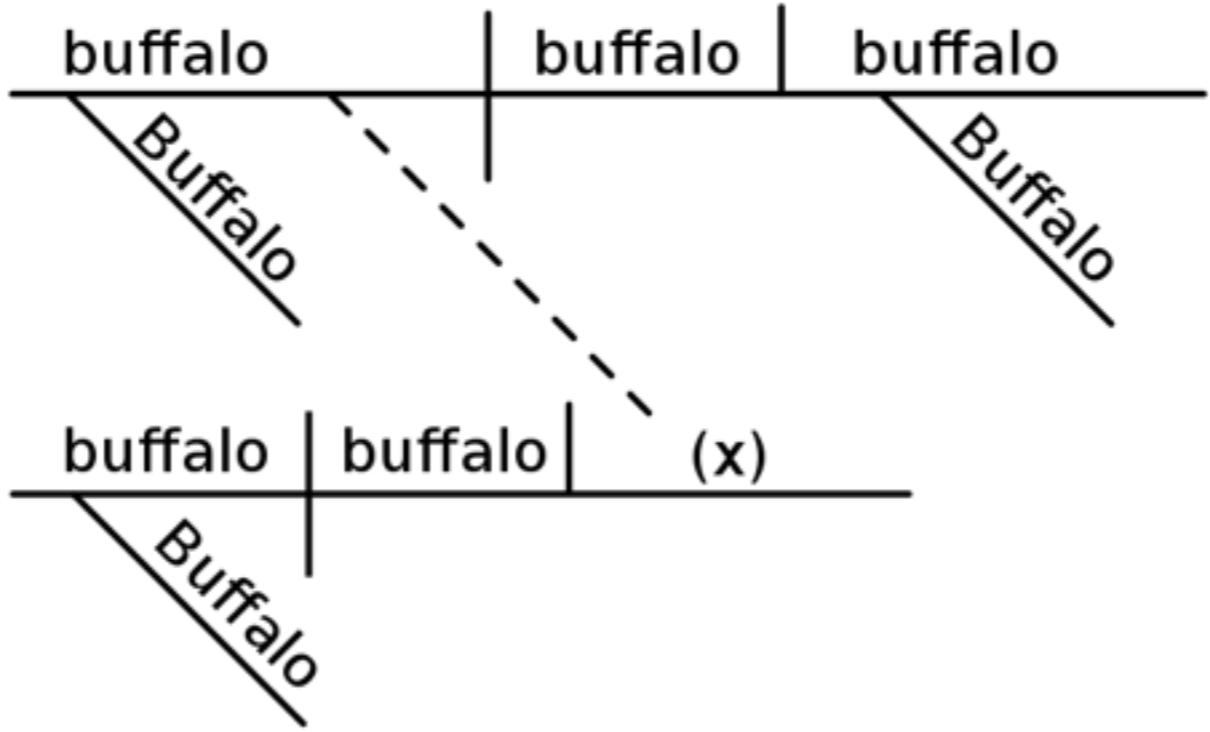
SATISFIES



ACCURACY ISN'T EVERYTHING

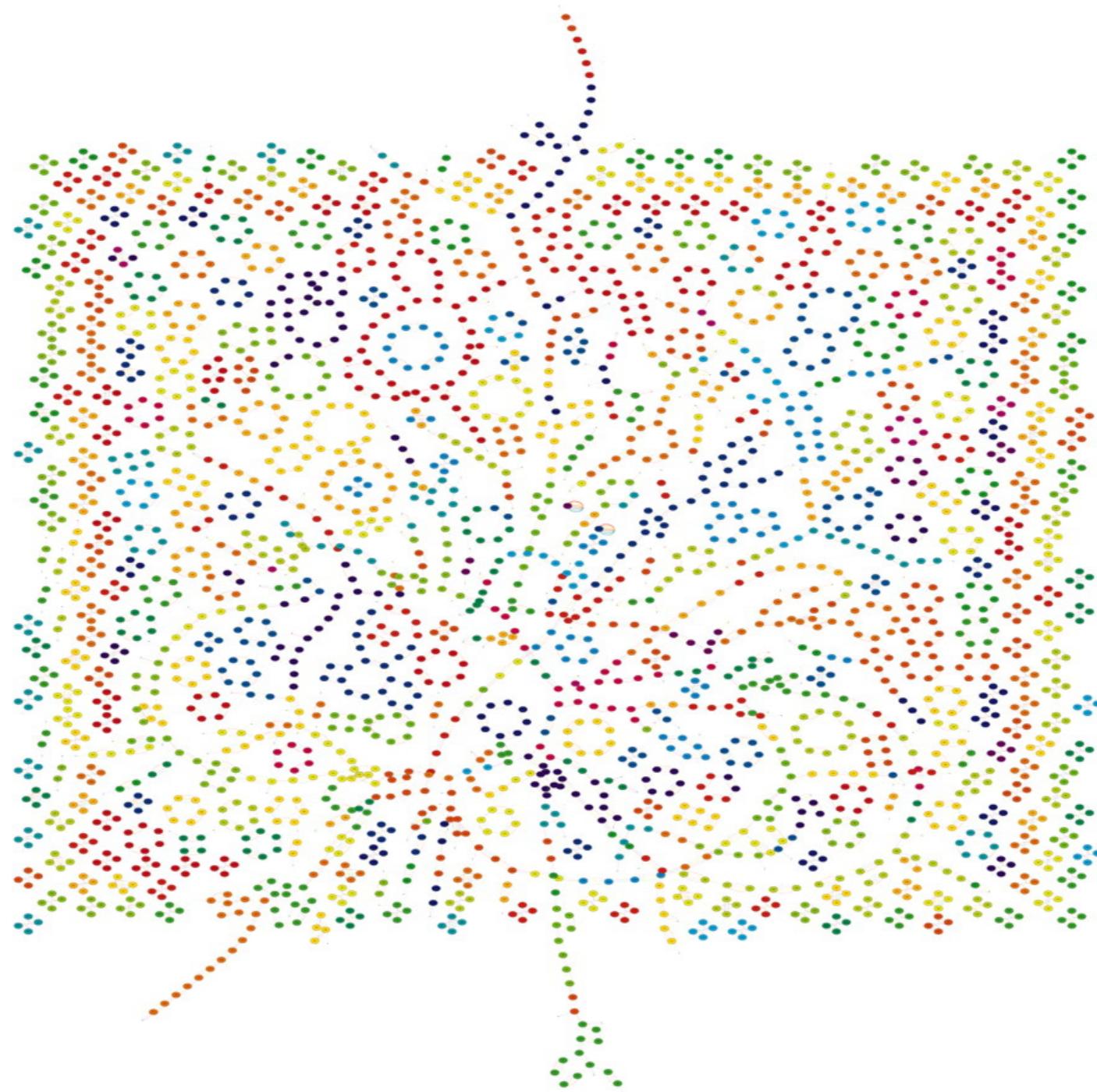
Correctness does not ensure comprehension.

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.

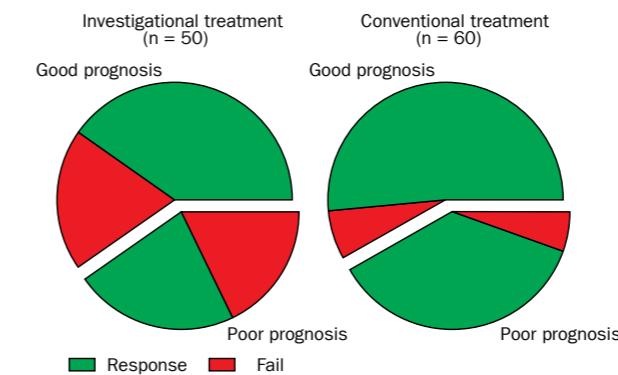
Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures

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

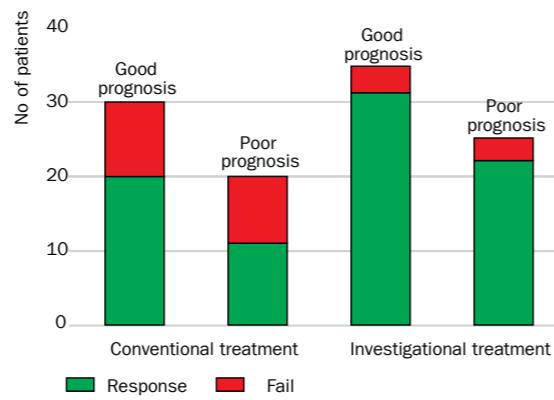
table

	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

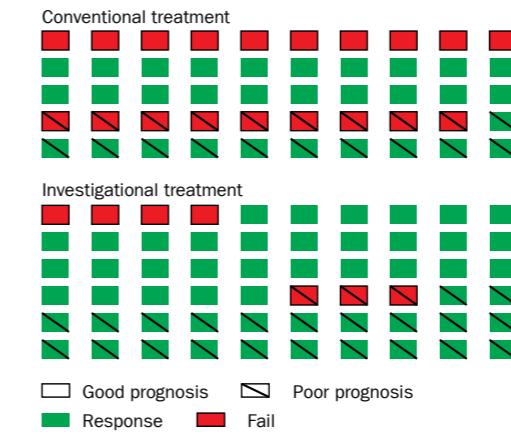
pie chart



bar graph

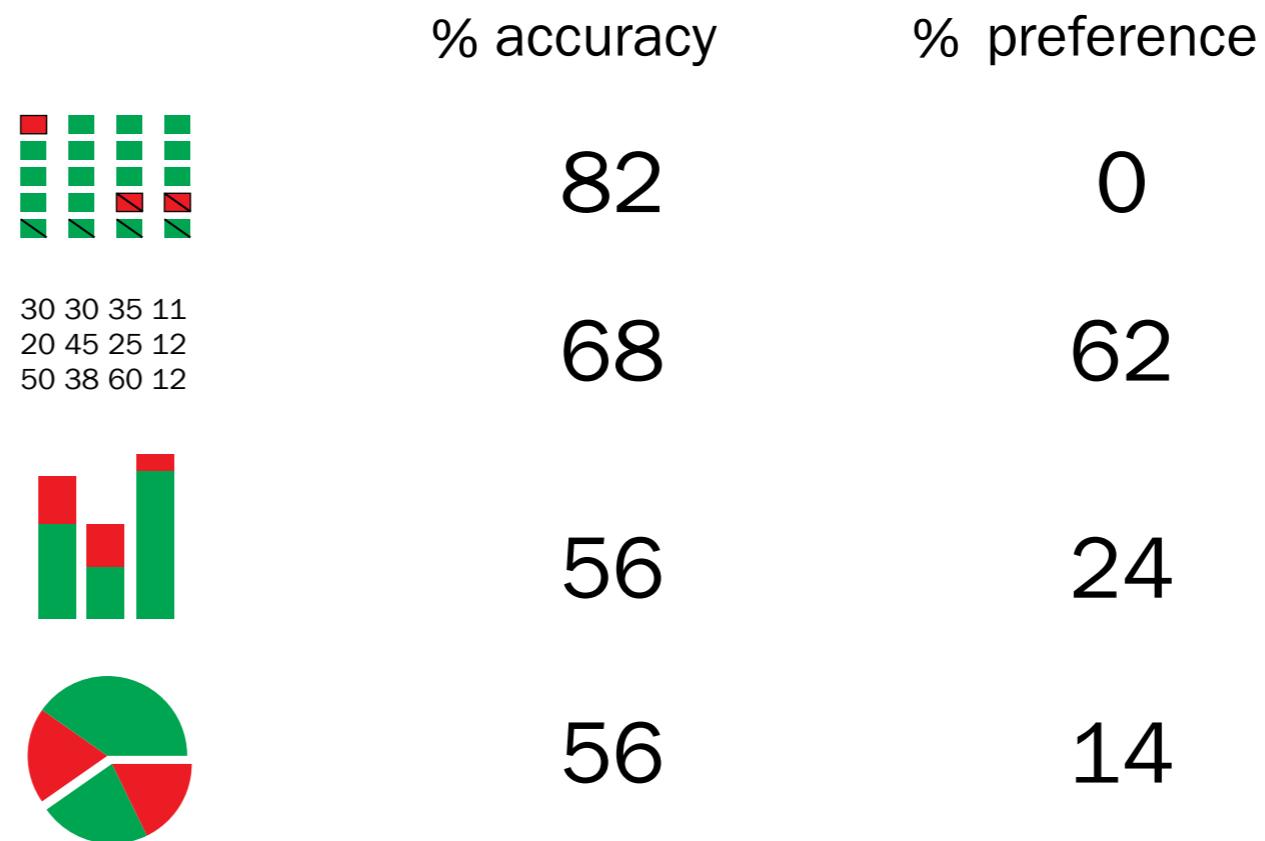


icon graph



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

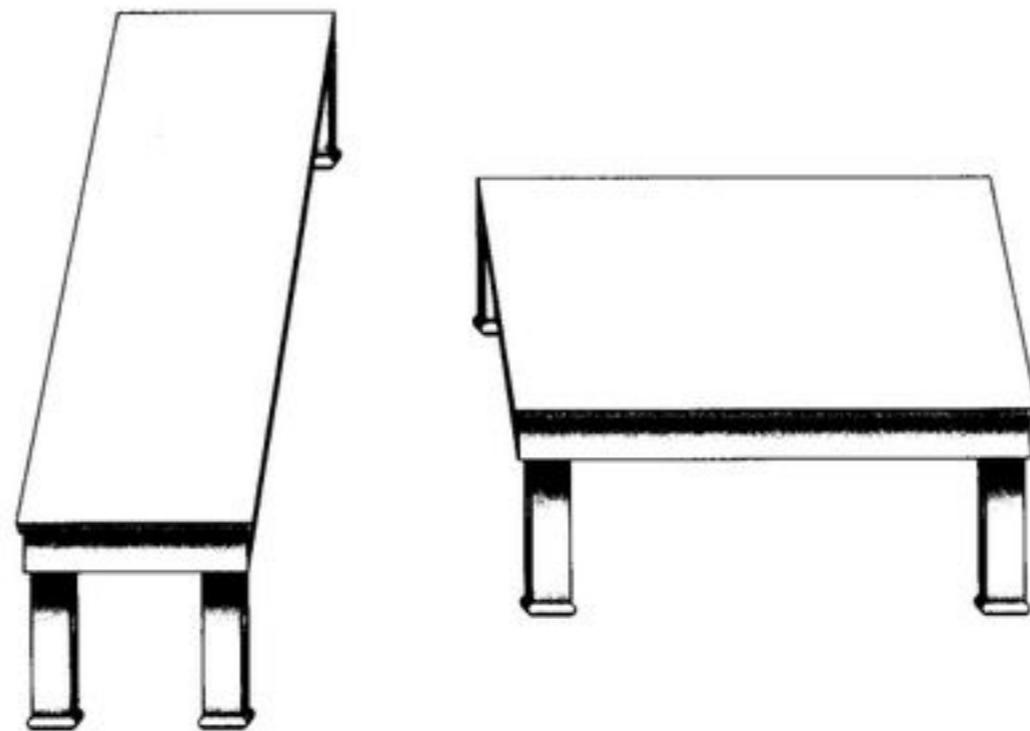
**“...eight voiced
considerable contempt
for the [icon] display.”**

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.”

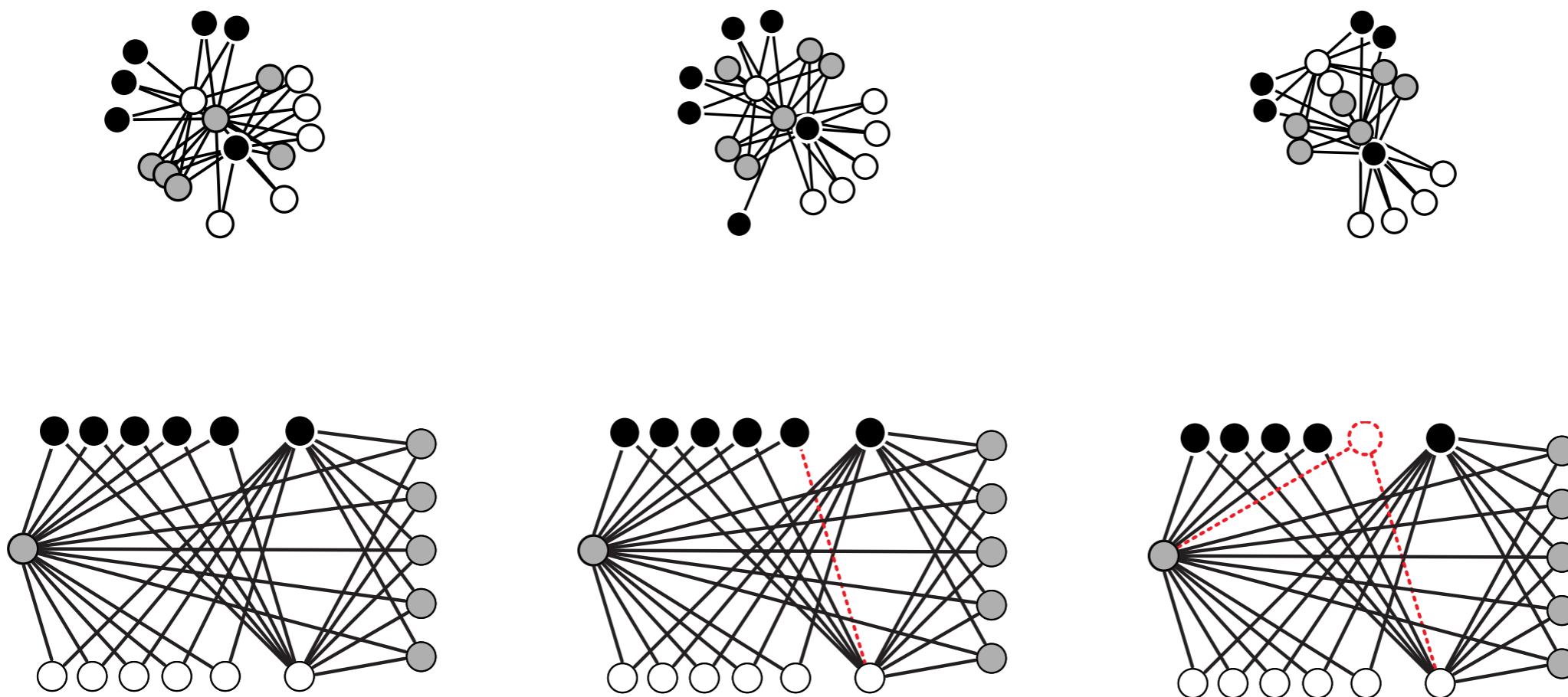
WE ARE EASILY DECEIVED



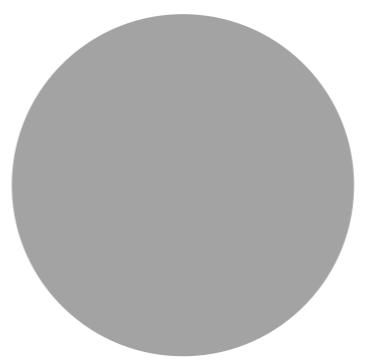
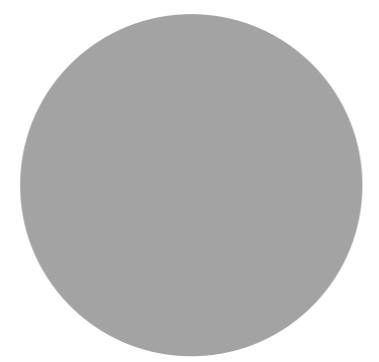
DO NOT BE CHARMED BY INEFFECTIVE FORMS

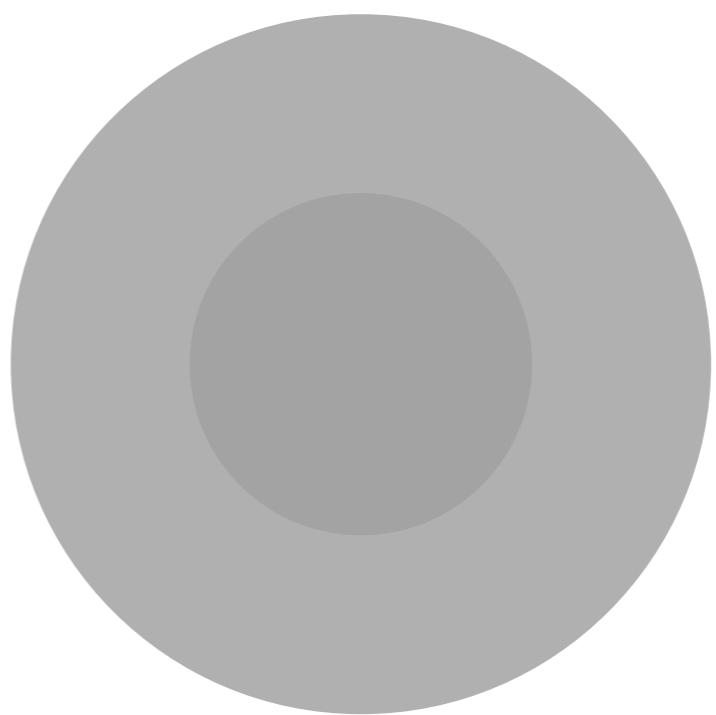
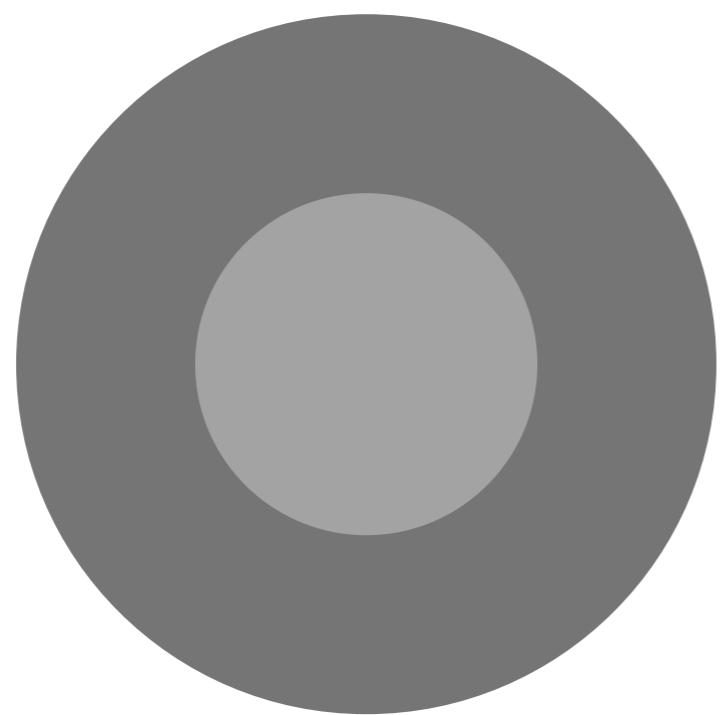


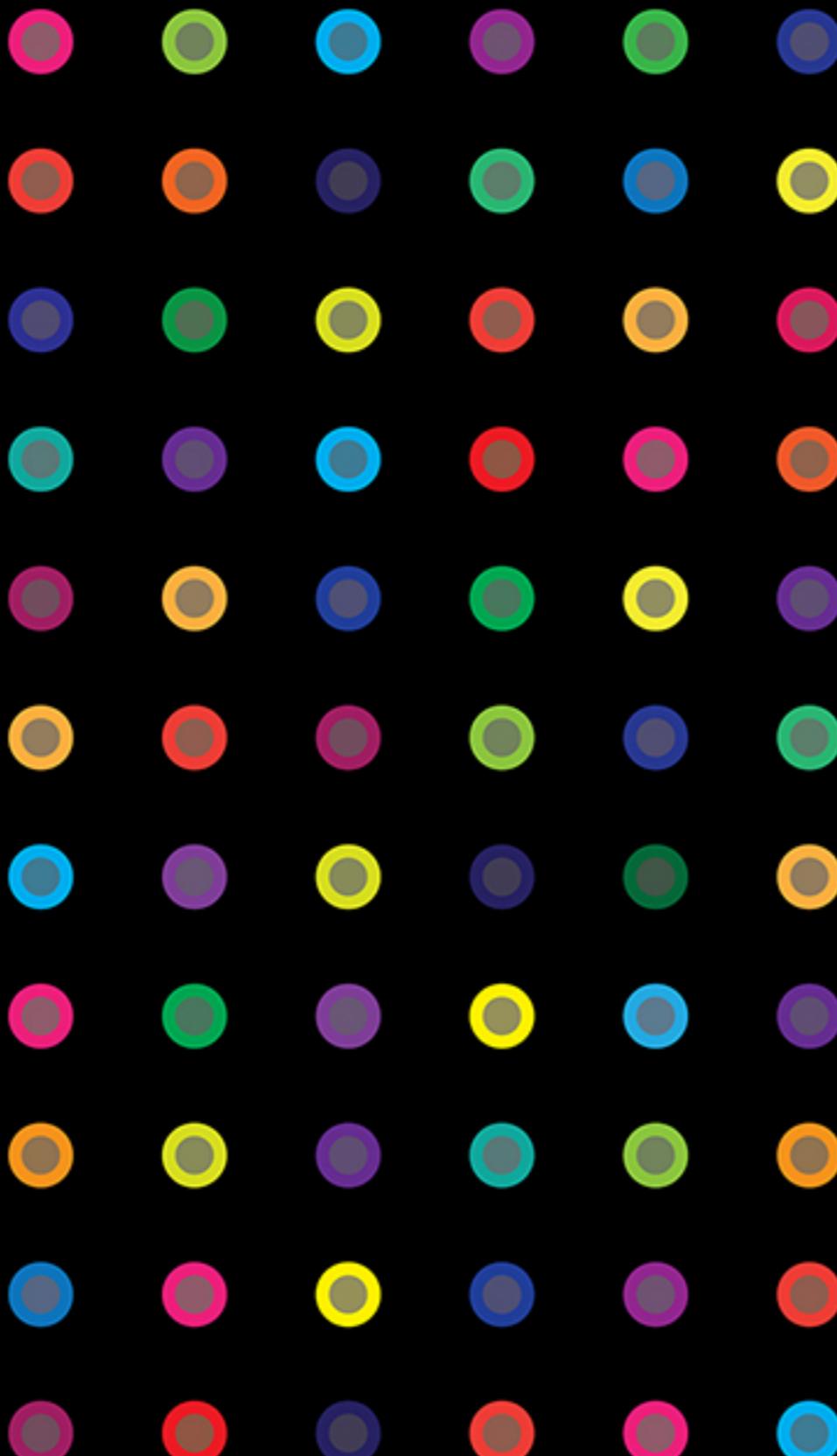
THE INCOMPARABLE NETWORK



RESPECT HUMAN VISUAL LIMITATIONS







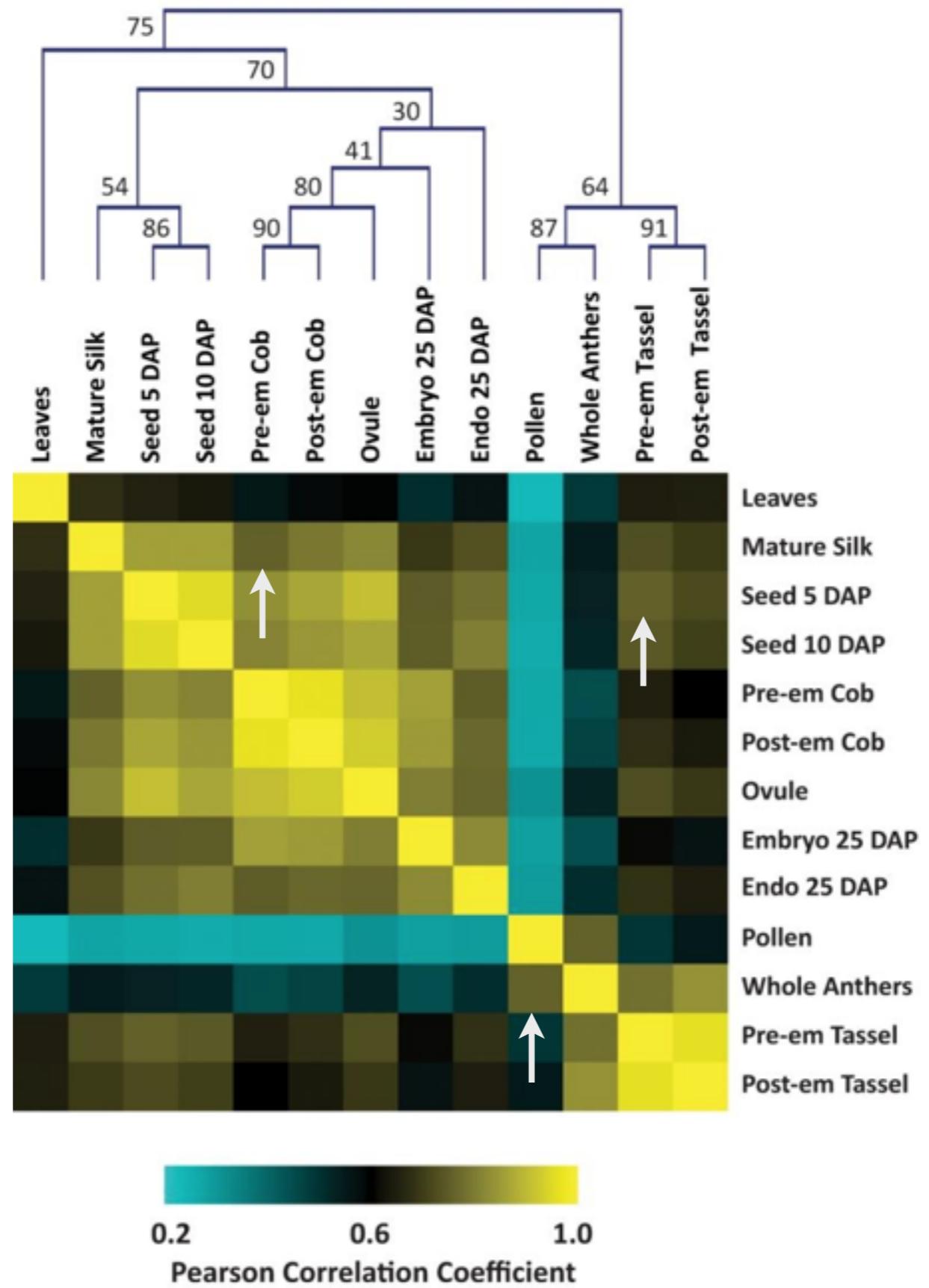
The first cases of what would later become known as AIDS were reported in the United States in June of 1981. Since then, 1.9 million people in the U.S. and around the world have been infected with HIV, including over 415,000 who have already died and approximately 1.2 million (1,174,934) adults and adolescents who were living with HIV infection at the end of 2008, 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.

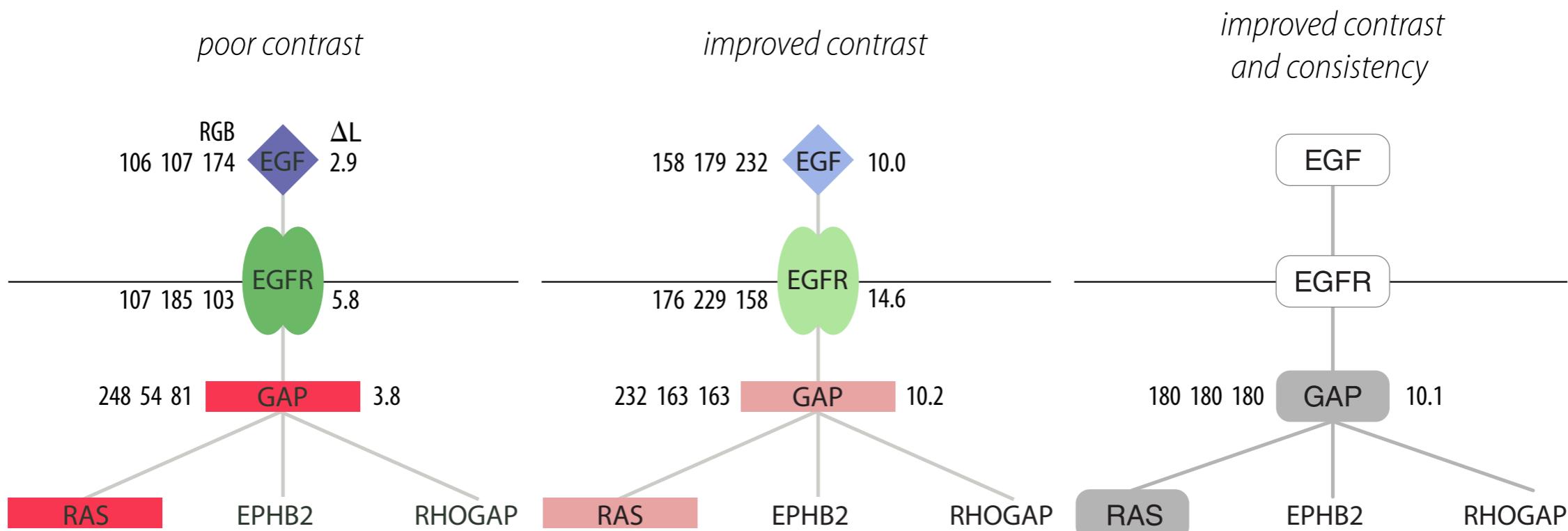
The virus can be transmitted through blood, semen, pre-cervical fluid, vaginal fluid and breast milk. It is estimated that more than one million people are living with HIV in the U.S. And even more ominously, one in five (20%) of those people living with HIV is unaware of their infection.

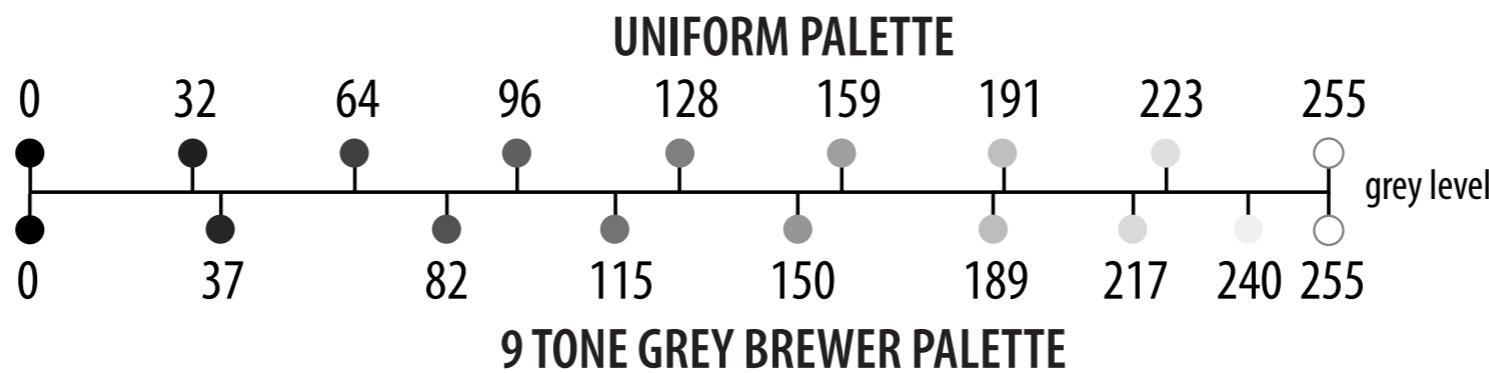
The easiest way to prevent the spread of HIV and AIDS is by wearing a condom when you have sex with your partner, especially if you don't know their sexual history. And with condoms so readily available, there's really no excuse not to wear one!

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

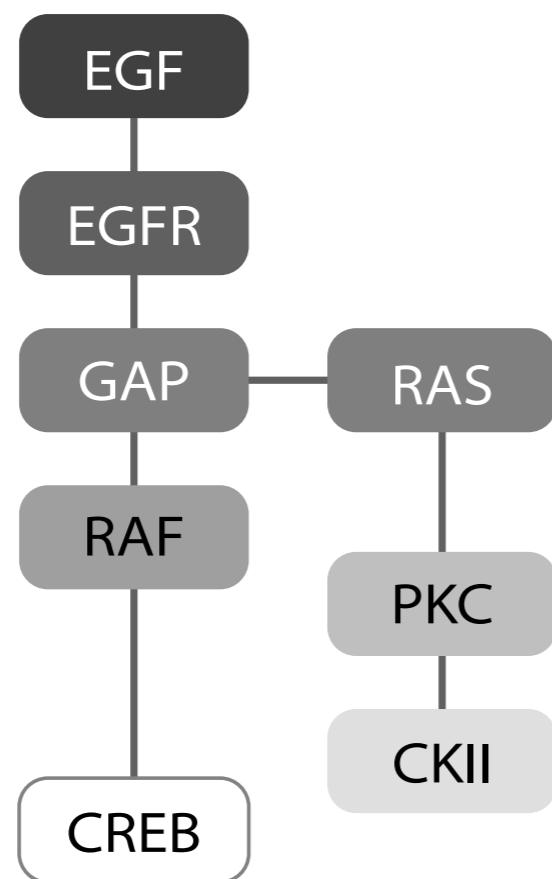
amfAR®
AIDS RESEARCH
WWW.AMFAR.GOV



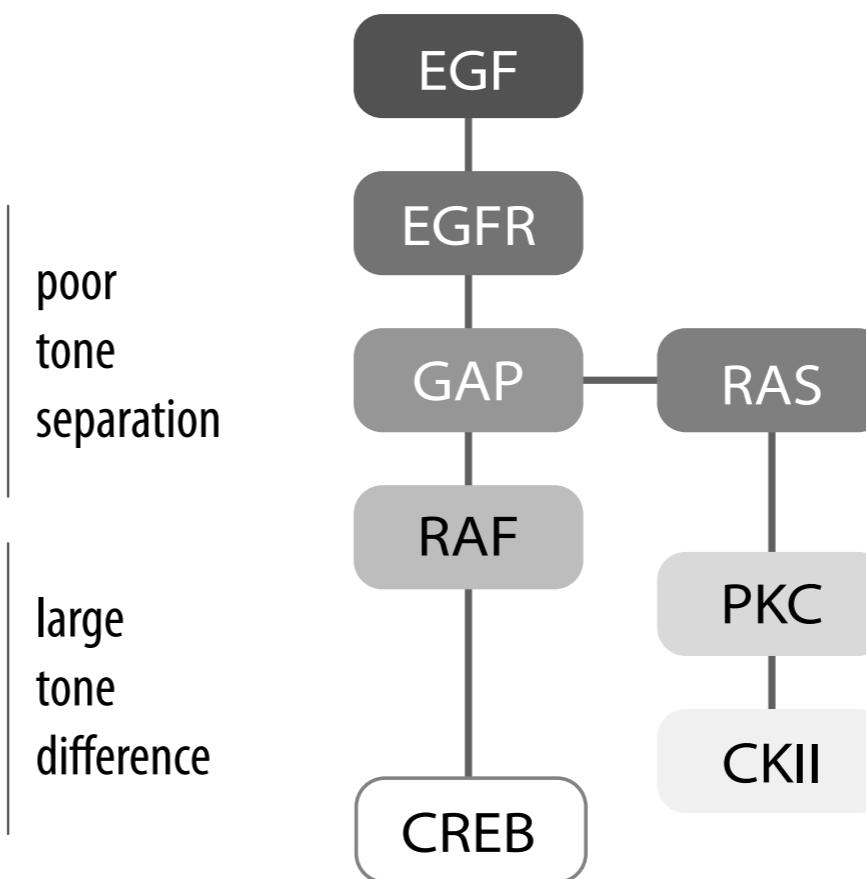




UNIFORM SPACING



BREWER SPACING



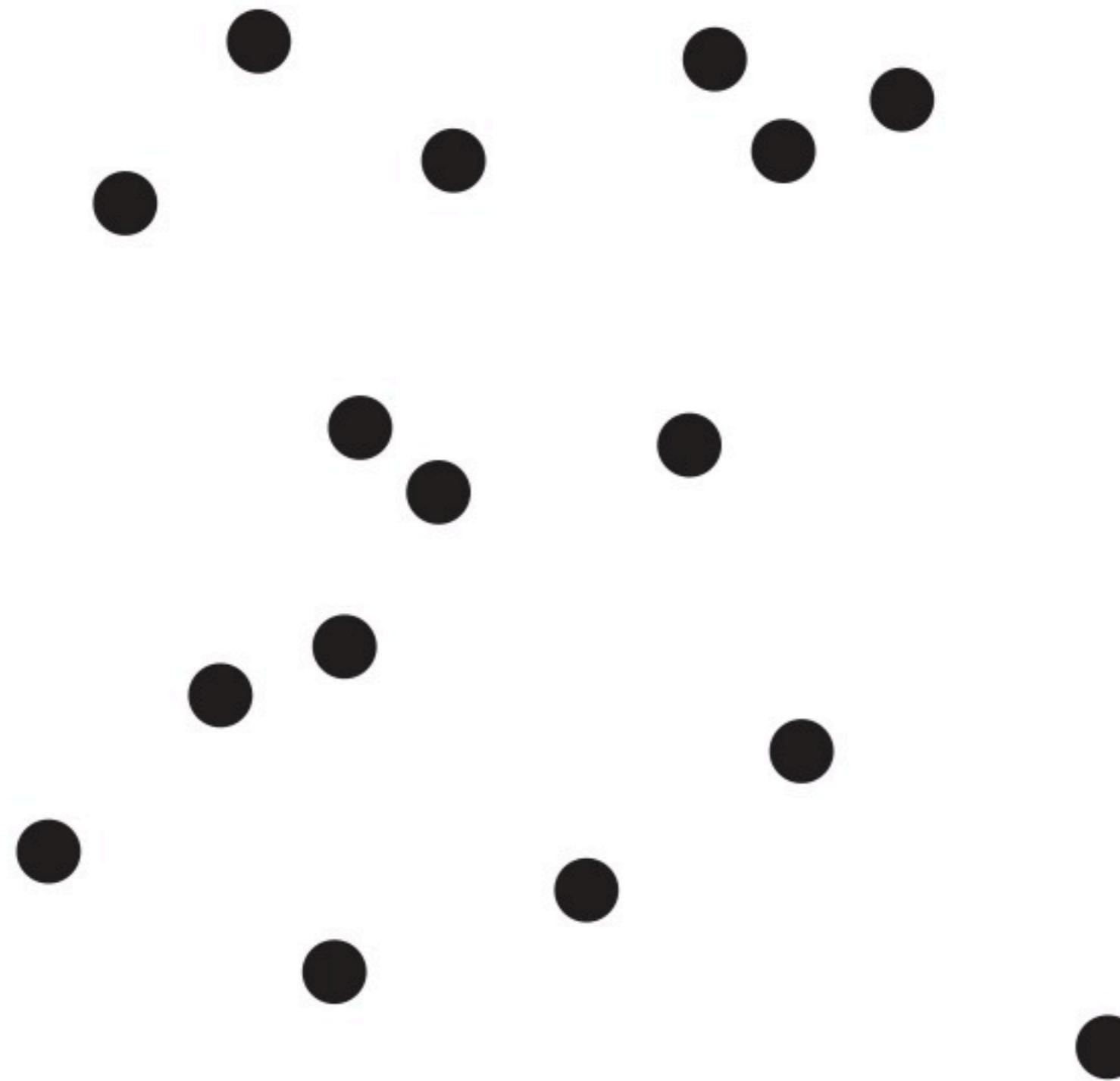
poor
tone
separation

large
tone
difference

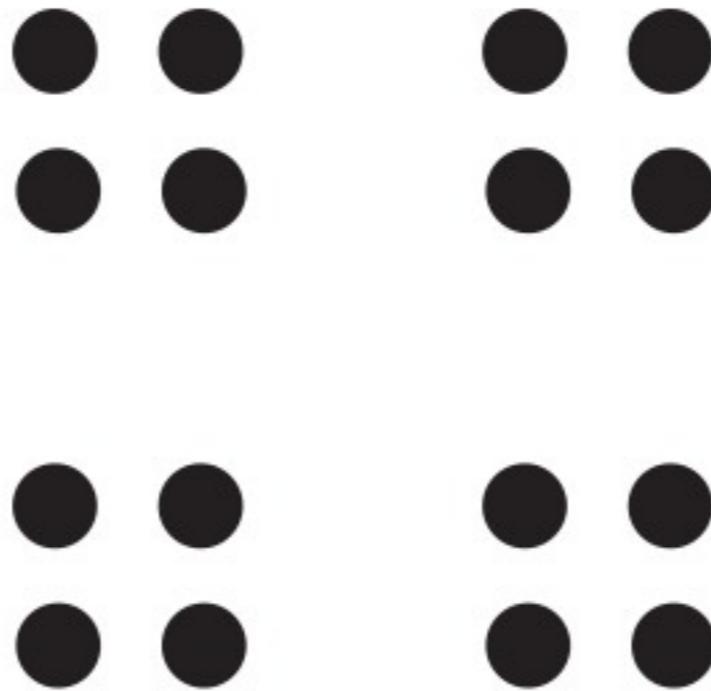
DATA INFORMS VARIATION

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

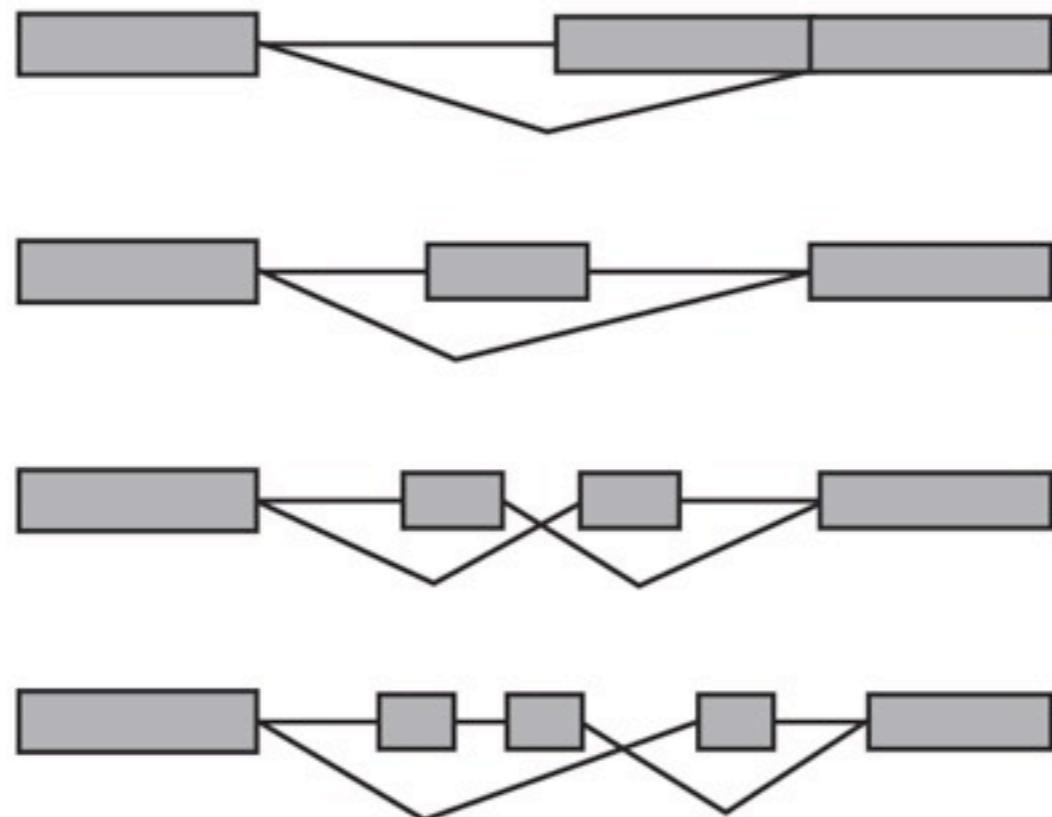
HOW MANY DOTS?



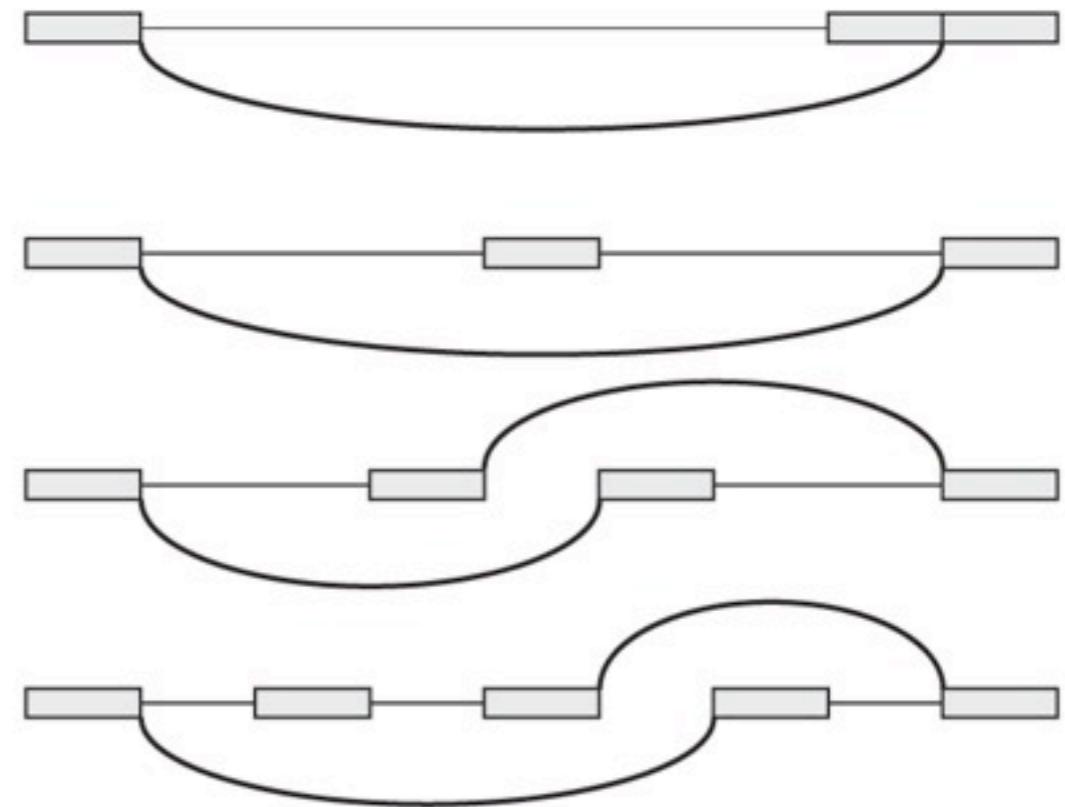
AVOID EXCESS DEGREES OF FREEDOM



spacing variation is implied

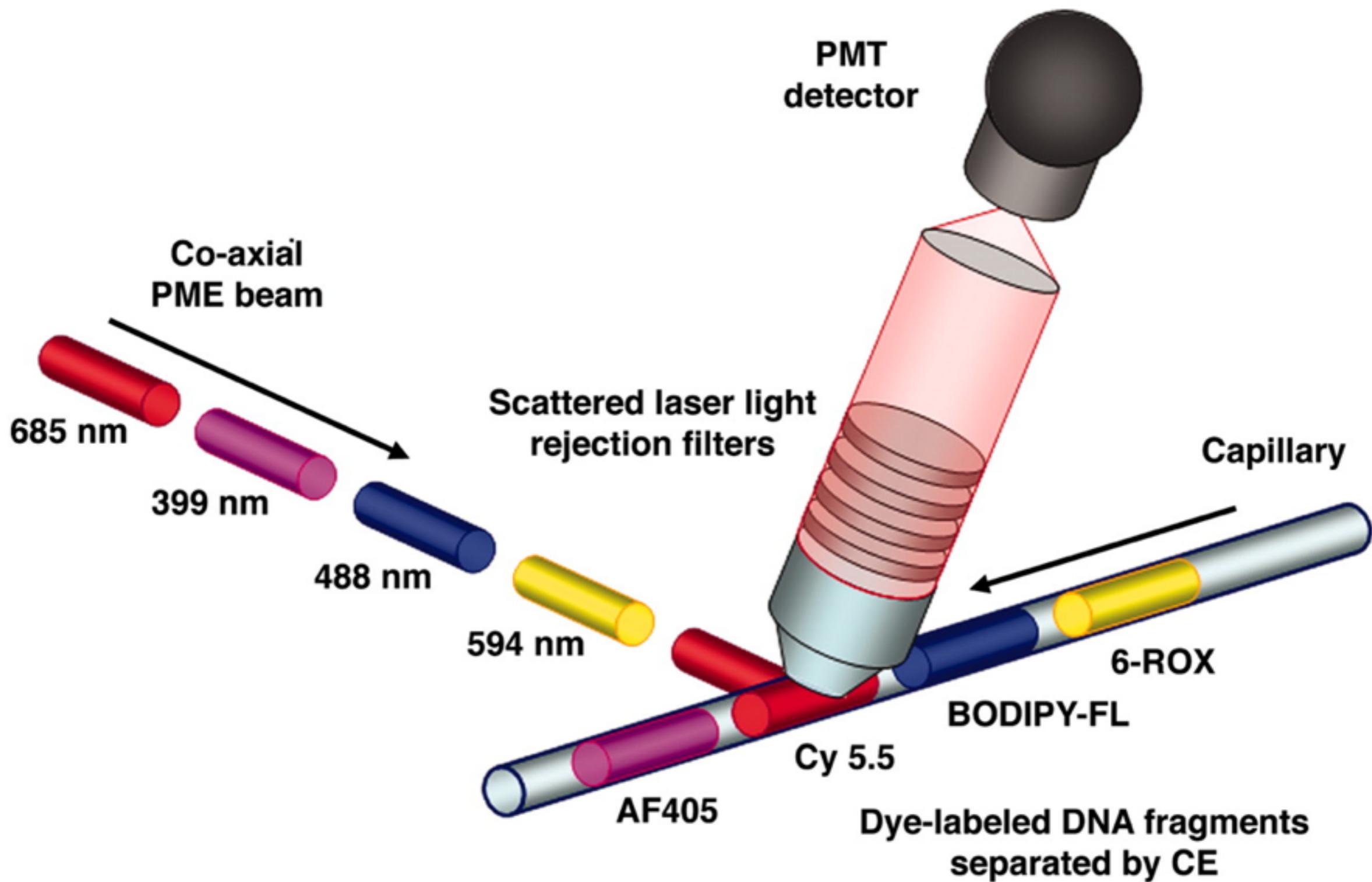


variation refactored



CONSISTENCY

Avoid the use of similar encodings for independent variables.

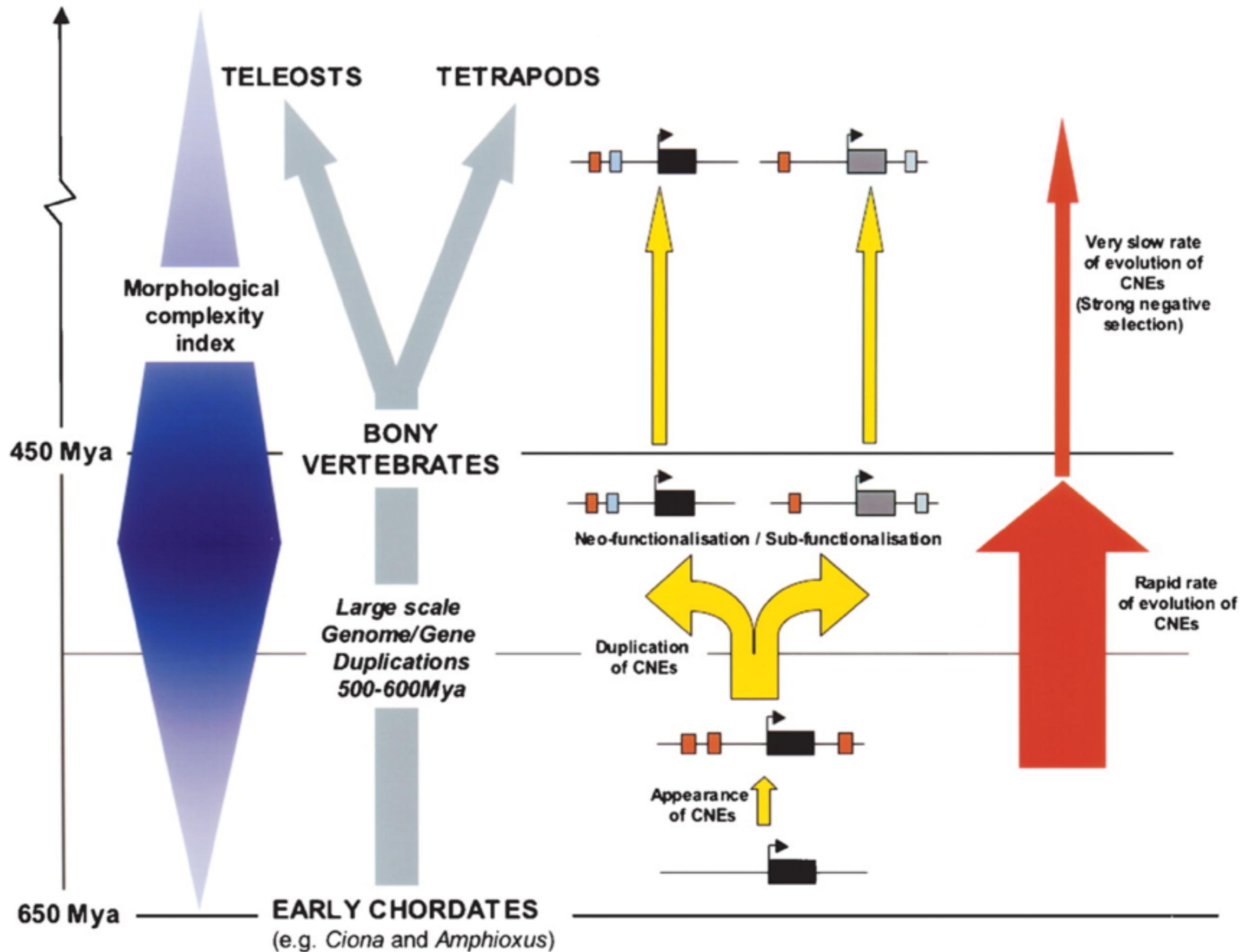


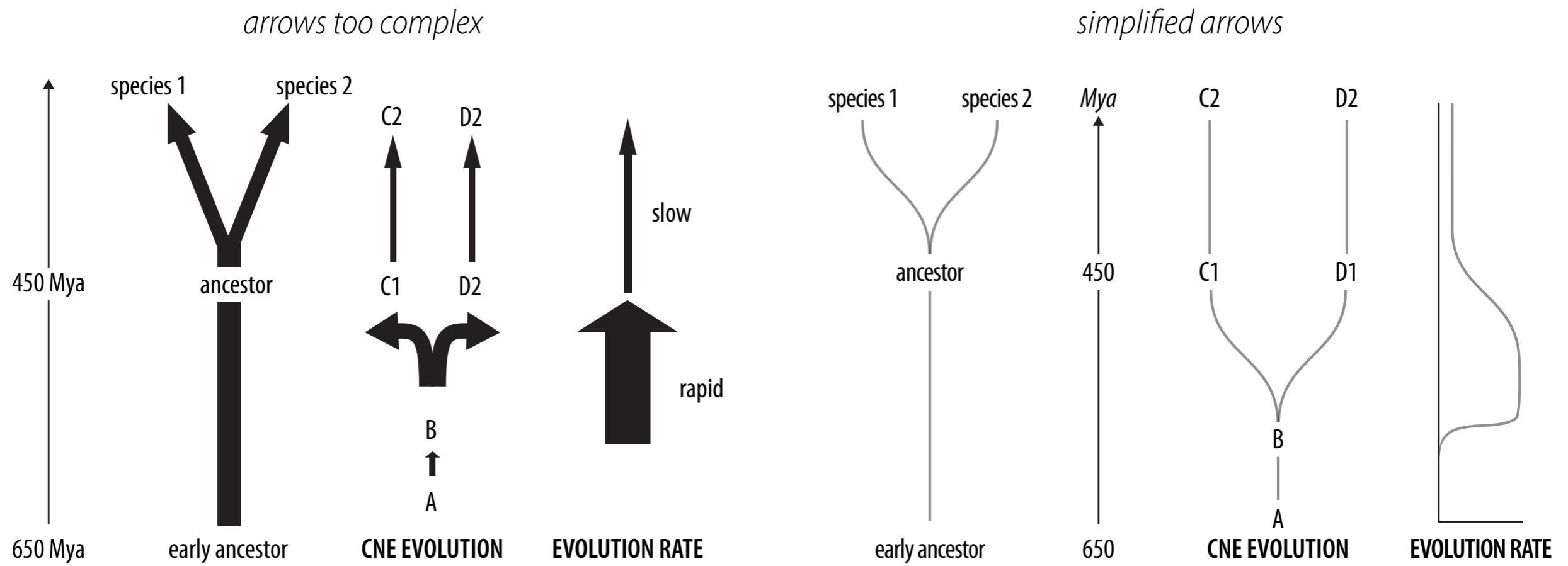
homonymous encoding



disambiguated

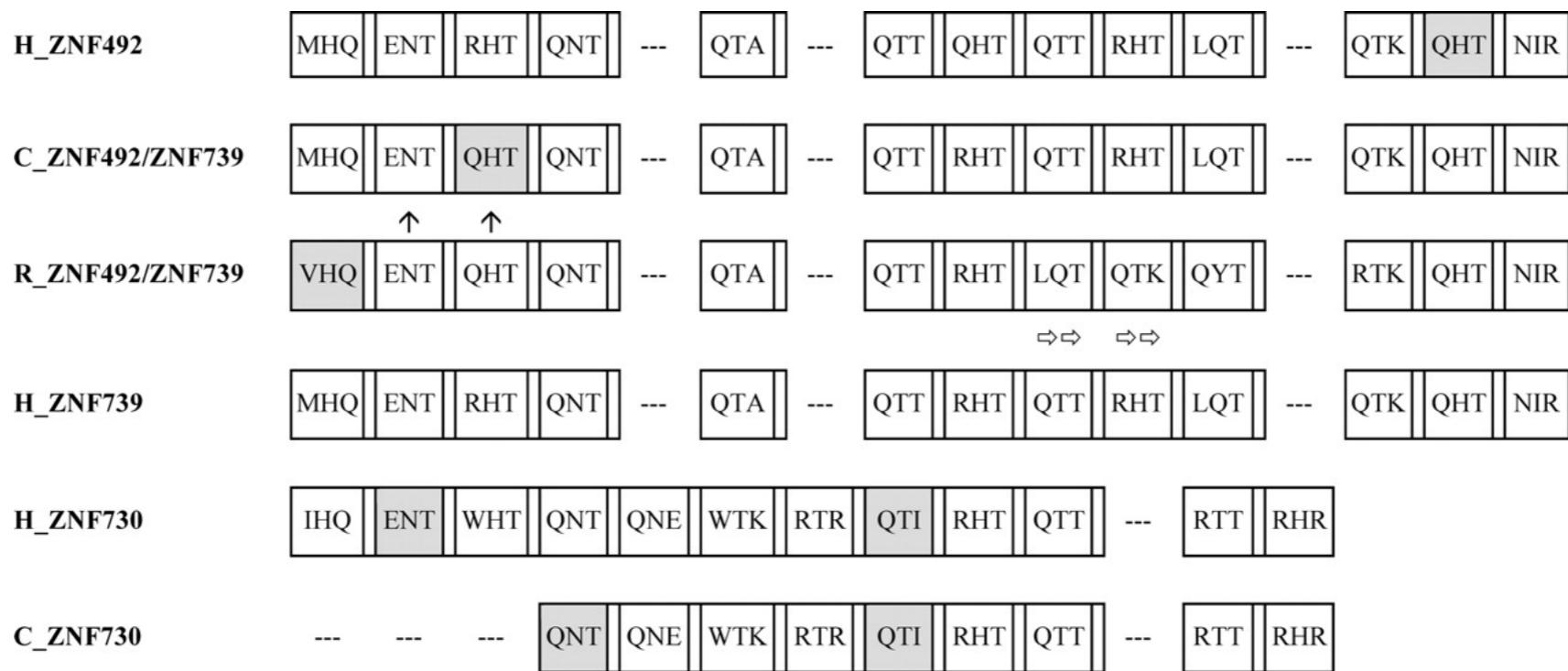






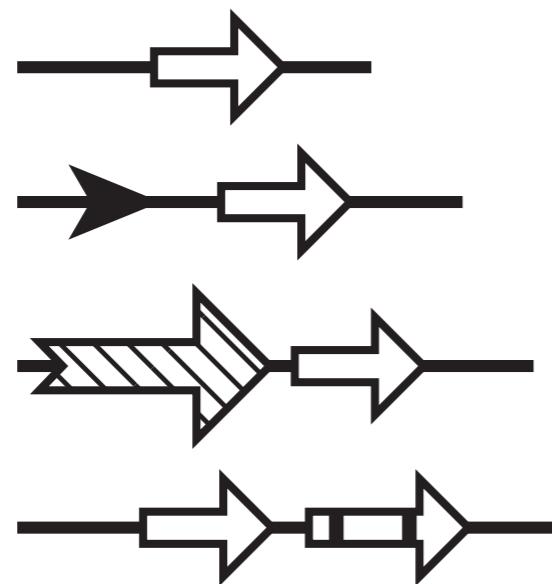
REDUNDANCY

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

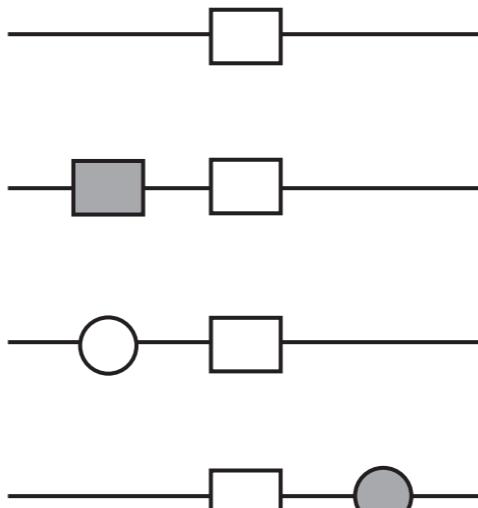


	MHQ	ENT	RHT	QNT	QNE	QTA	RTR	QTT	QHT	QTT	RHT	LQT	RHR	QTK	QHT	NIR
H492
C492/739	Q..	R..
R492/739	V..	...	Q..	R..	LQ.	QTK	QY.	...	R..
H739	M..	R..	...	QTK
H730	I..	...	W..	W.KI	R..	...	RT.
C730		W.KI	R..	...	RT.

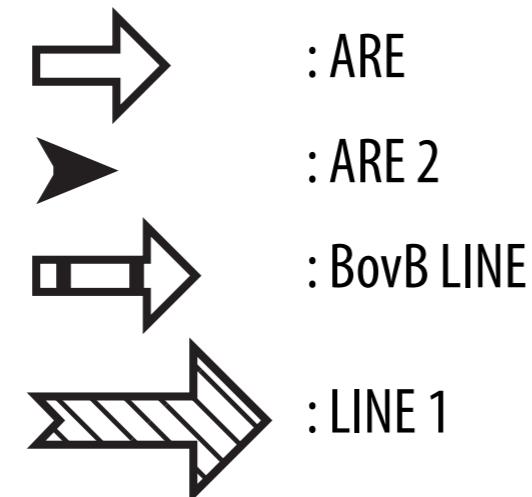
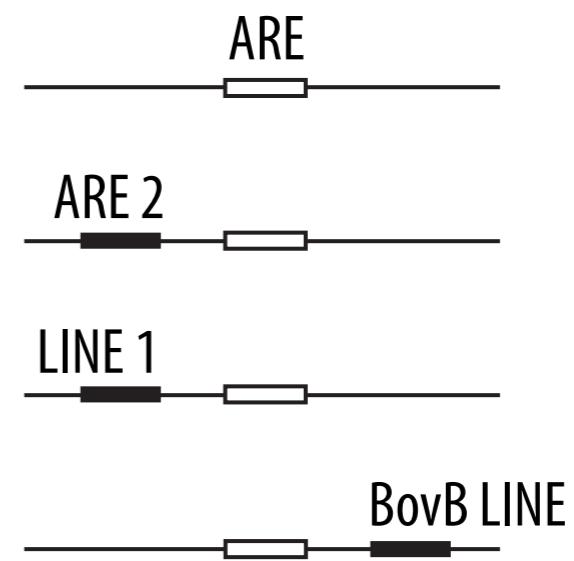
overwhelming



simplified

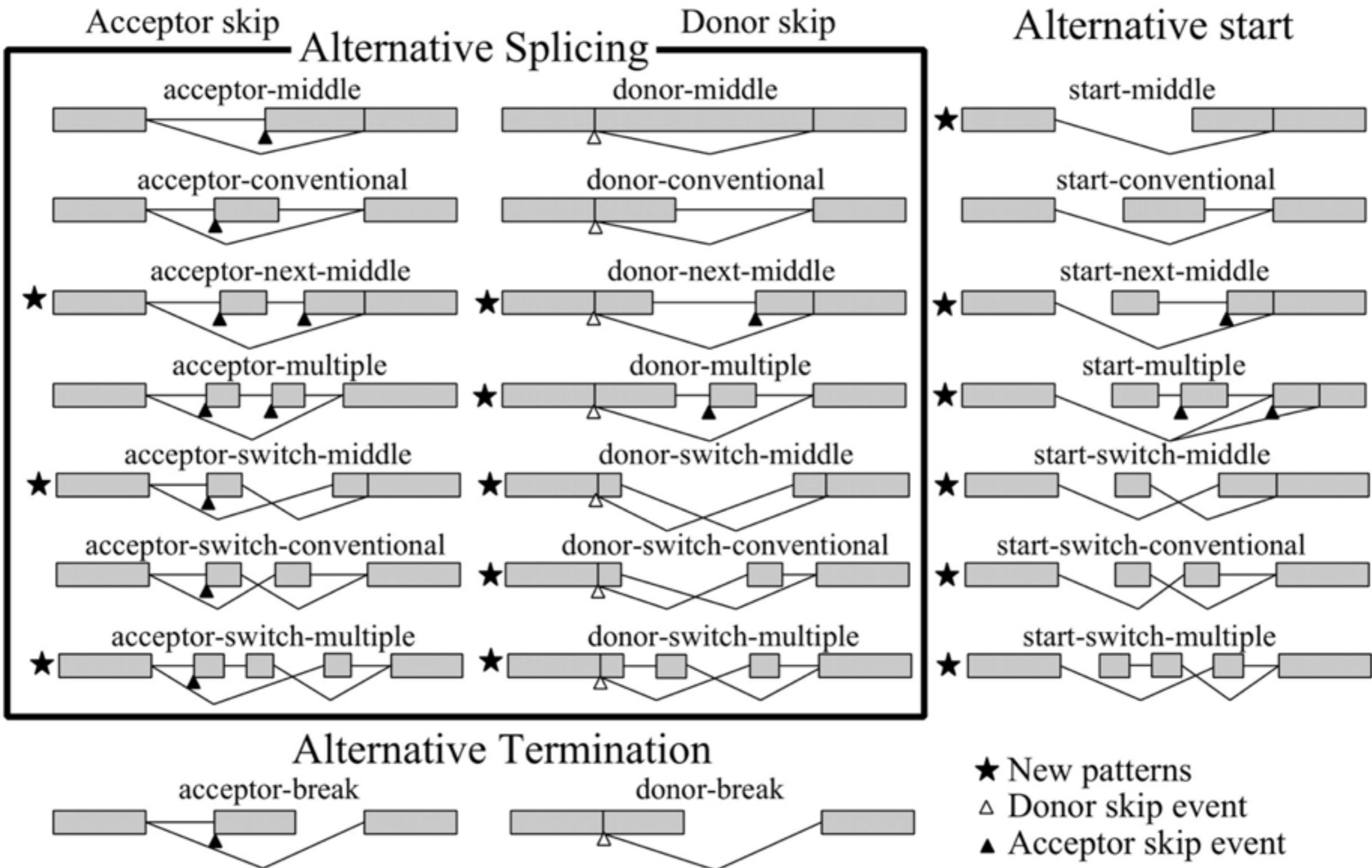


integrated key

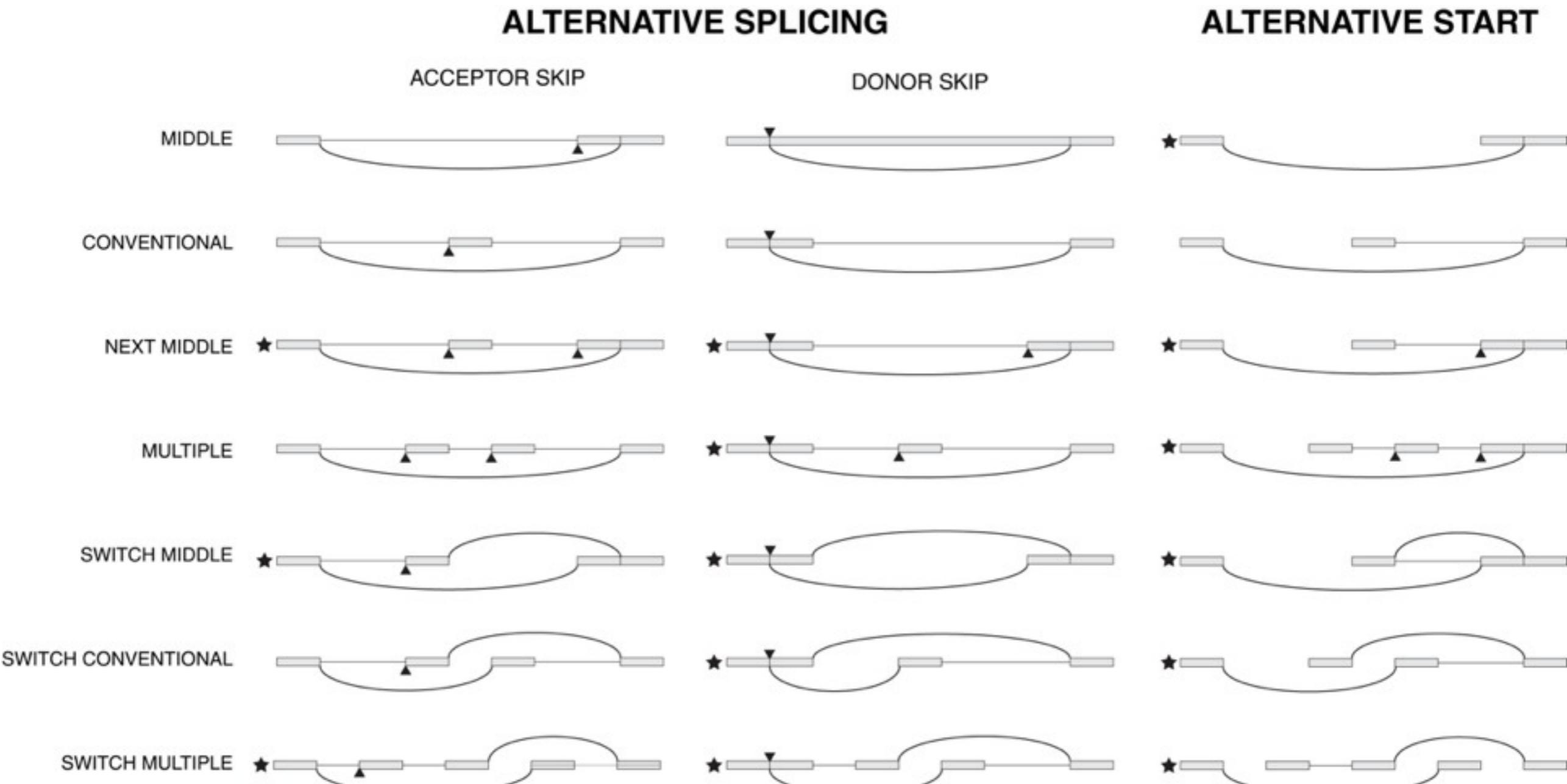


CONCISENESS

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



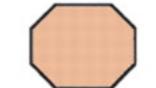
ALTERNATIVE SPlicing

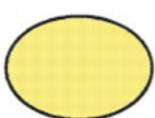


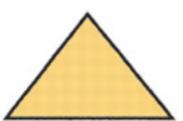
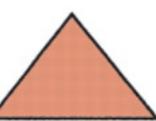
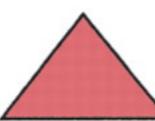
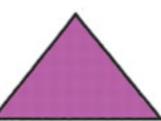
ALTERNATIVE TERMINATION

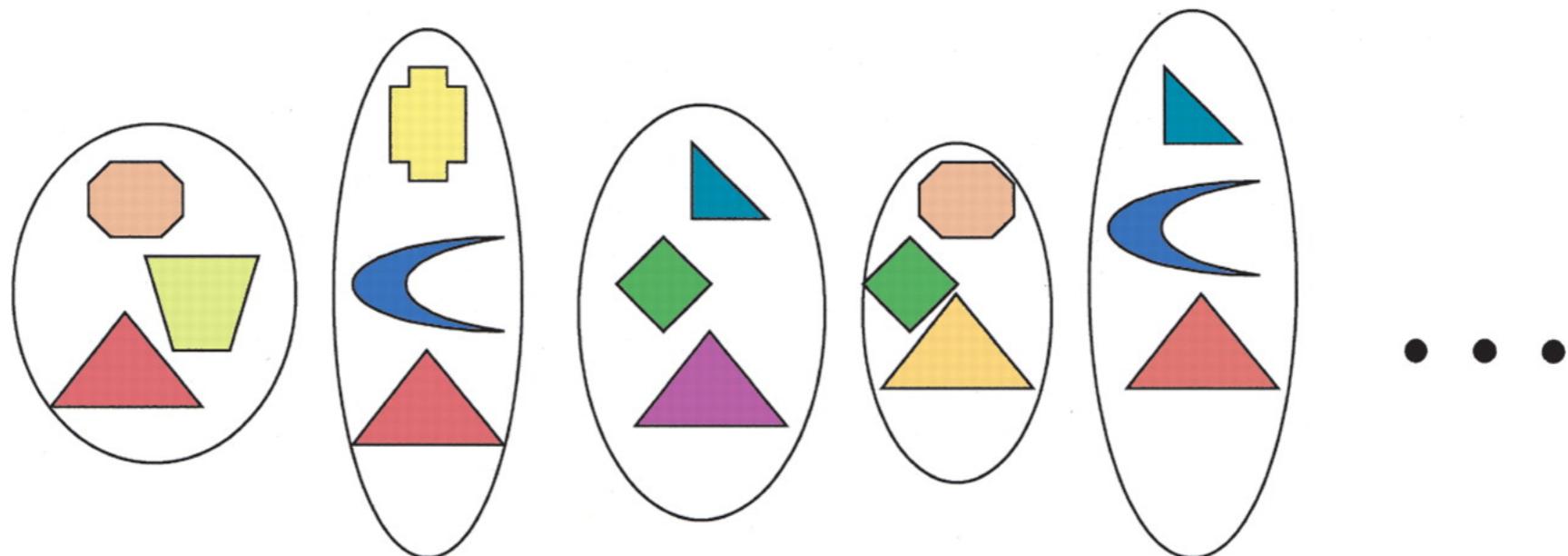


- ★ new patterns
- ▲ acceptor skip event
- ▼ donor skip event

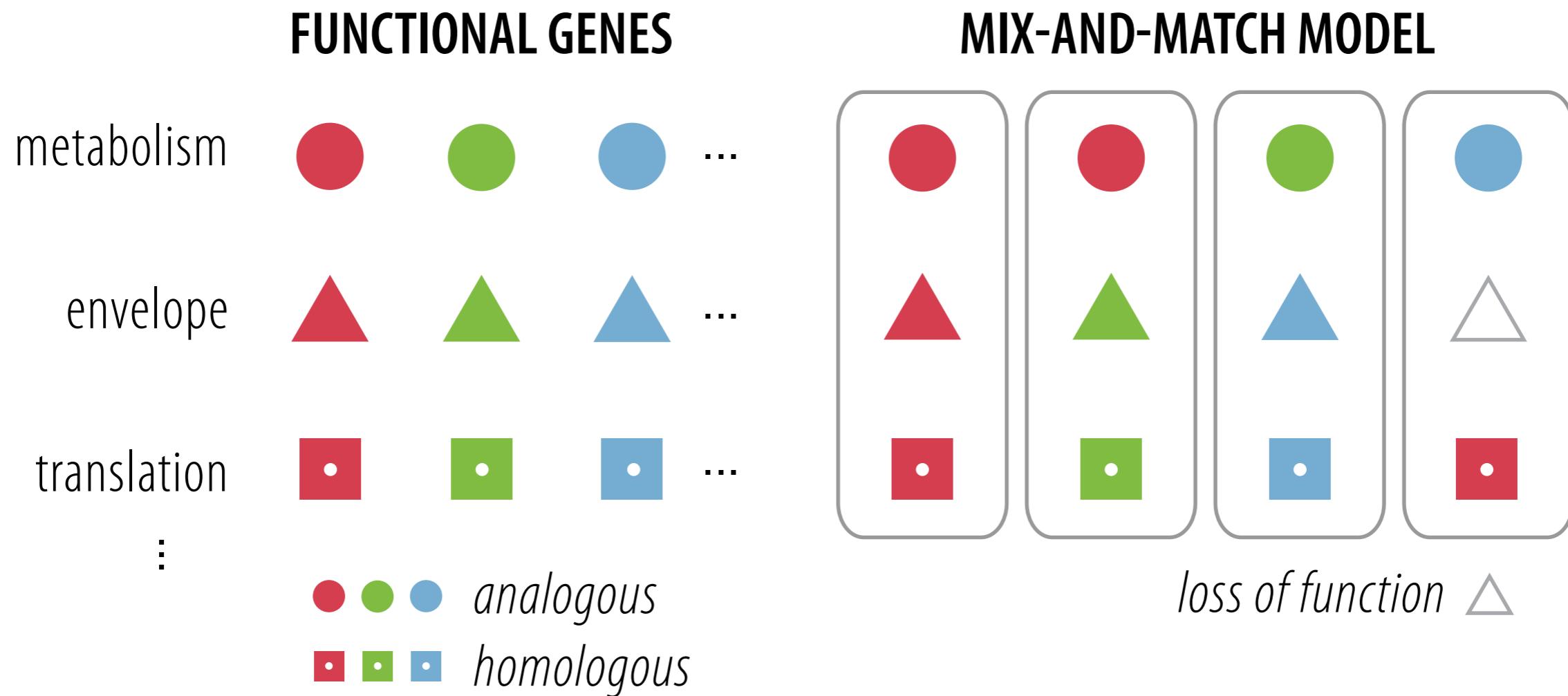
 =  =  =  ENVELOPE

 =  =  =  METABOLISM

 =  =  =  TRANSLATION



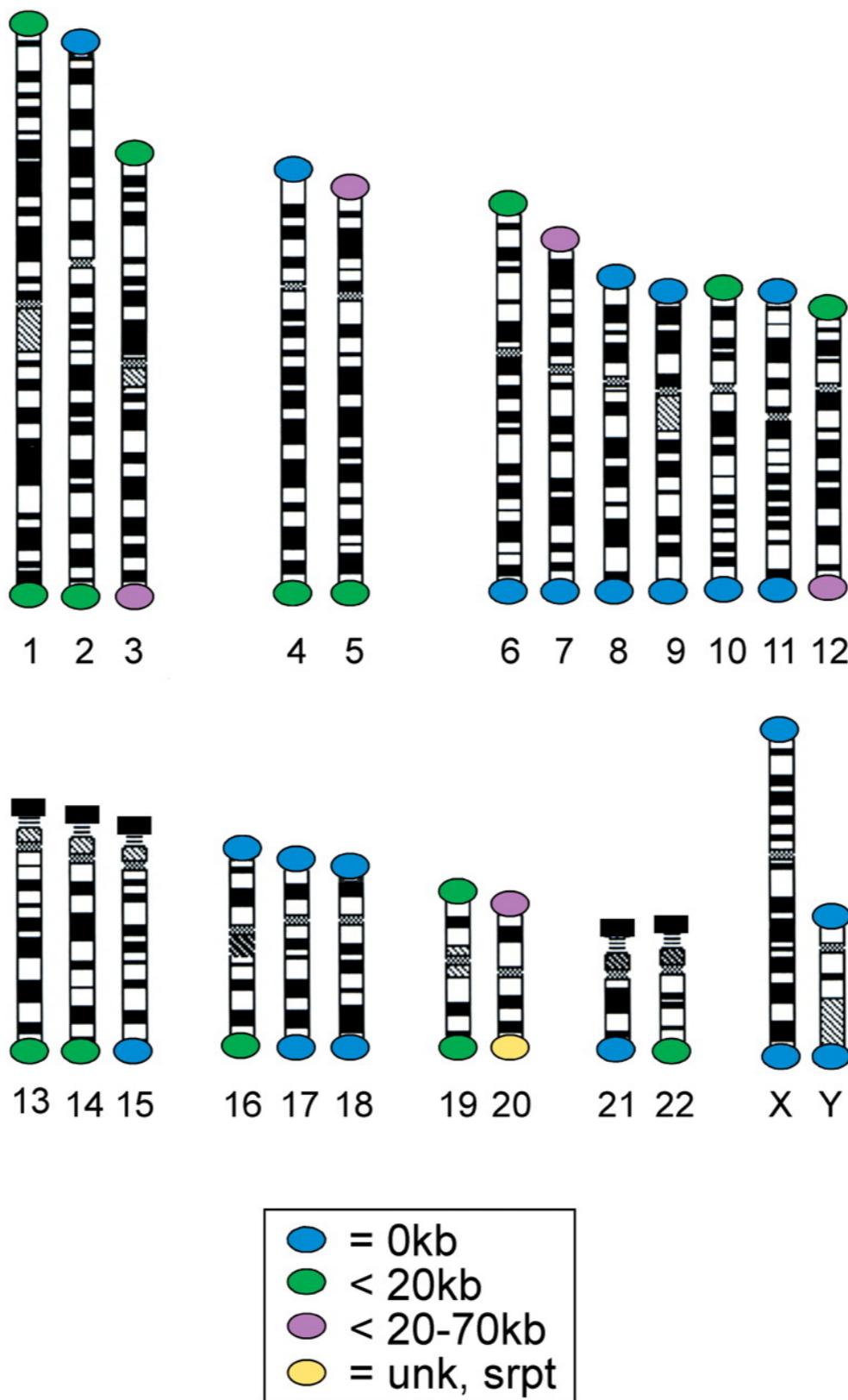
simple visual vocabulary



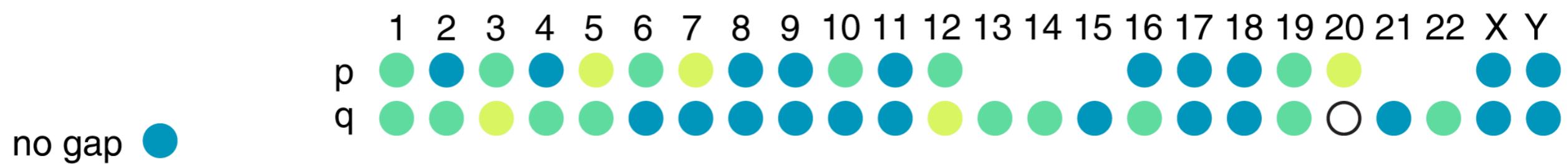
REMOVE TO IMPROVE

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

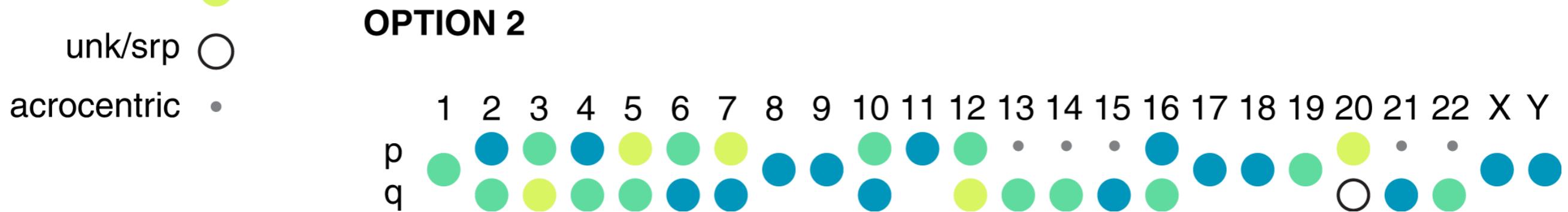
Shelter your reader from unnecessary complexity.



OPTION 1

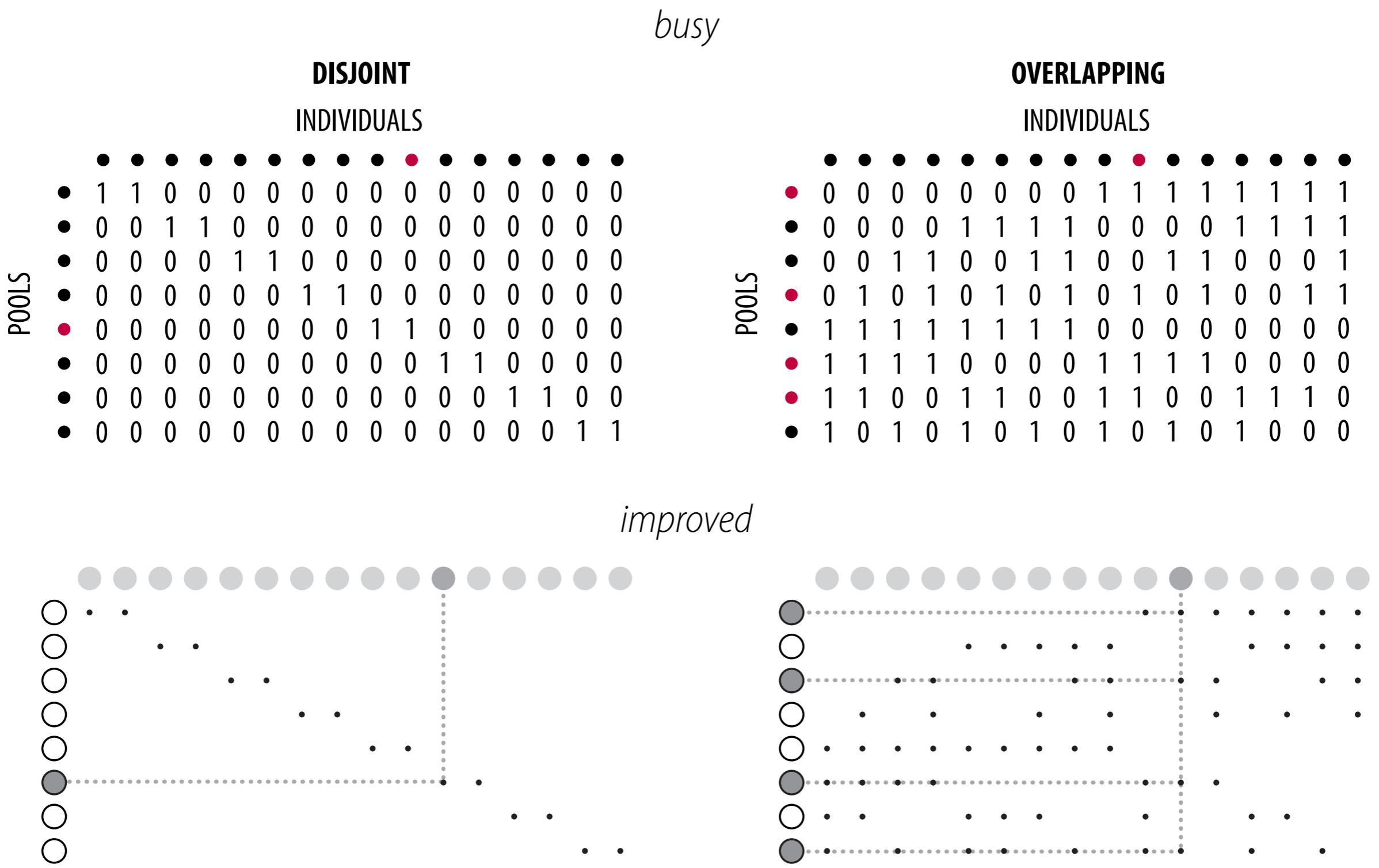


OPTION 2

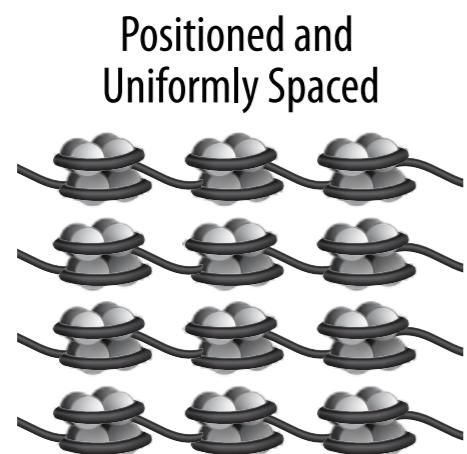


INDIVIDUALS														
Pools	1	1	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	1	1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	1	1	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	1	1	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	1	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	1	1	0	0
	0	0	0	0	0	0	0	0	0	0	0	1	1	0
	0	0	0	0	0	0	0	0	0	0	0	0	1	1

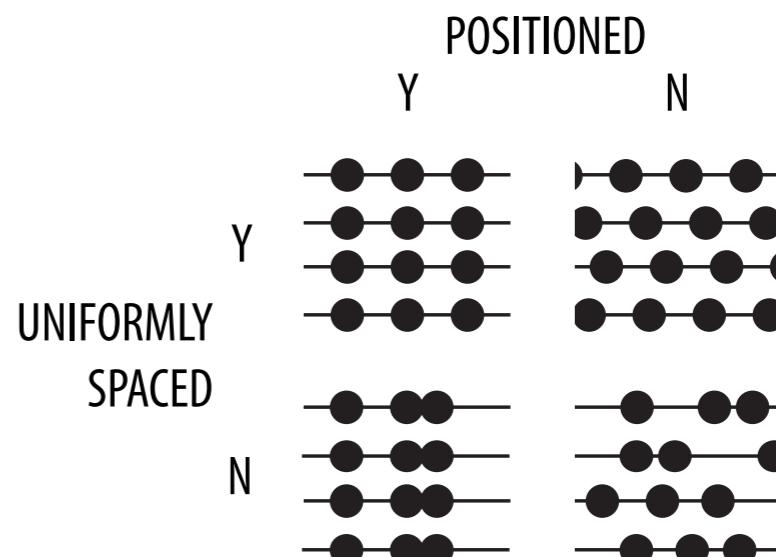
INDIVIDUALS														
Pools	1	0	0	0	0	0	0	0	1	1	1	1	1	1
	0	0	0	0	1	1	1	1	0	0	0	1	1	1
	0	0	1	1	0	0	1	1	0	0	1	0	0	1
	0	1	0	1	0	1	0	1	0	1	0	1	0	1
	1	1	1	1	1	1	1	1	0	0	0	0	0	0
	1	1	1	1	0	0	0	0	1	1	1	1	0	0
	1	1	0	0	1	1	0	0	1	1	0	0	1	1
	1	0	1	0	1	0	1	0	1	0	1	0	1	0



detail exposed



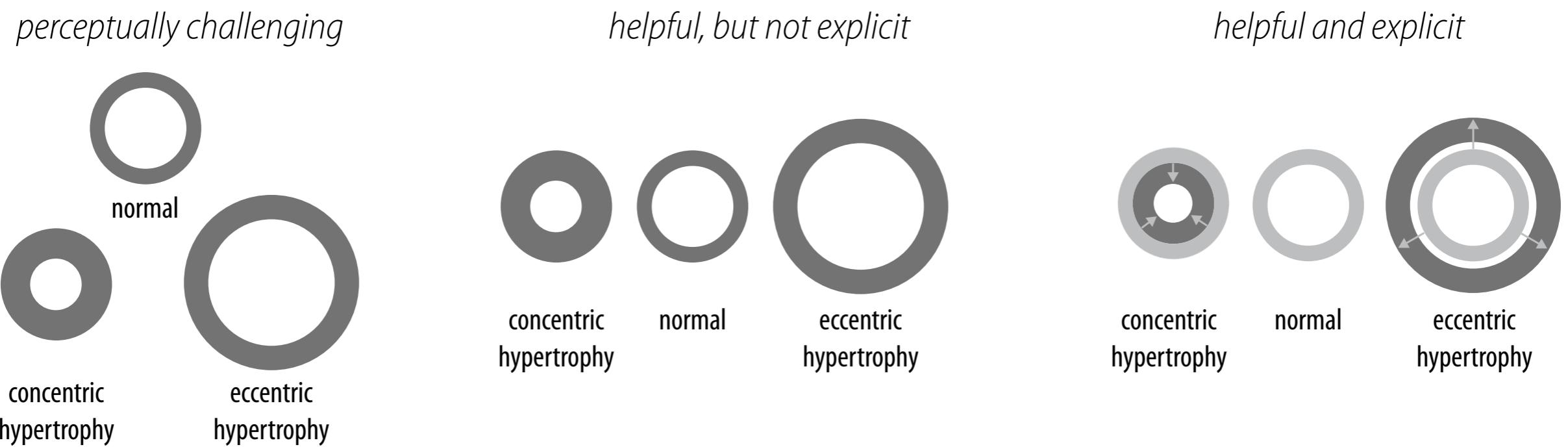
detail encapsulated



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

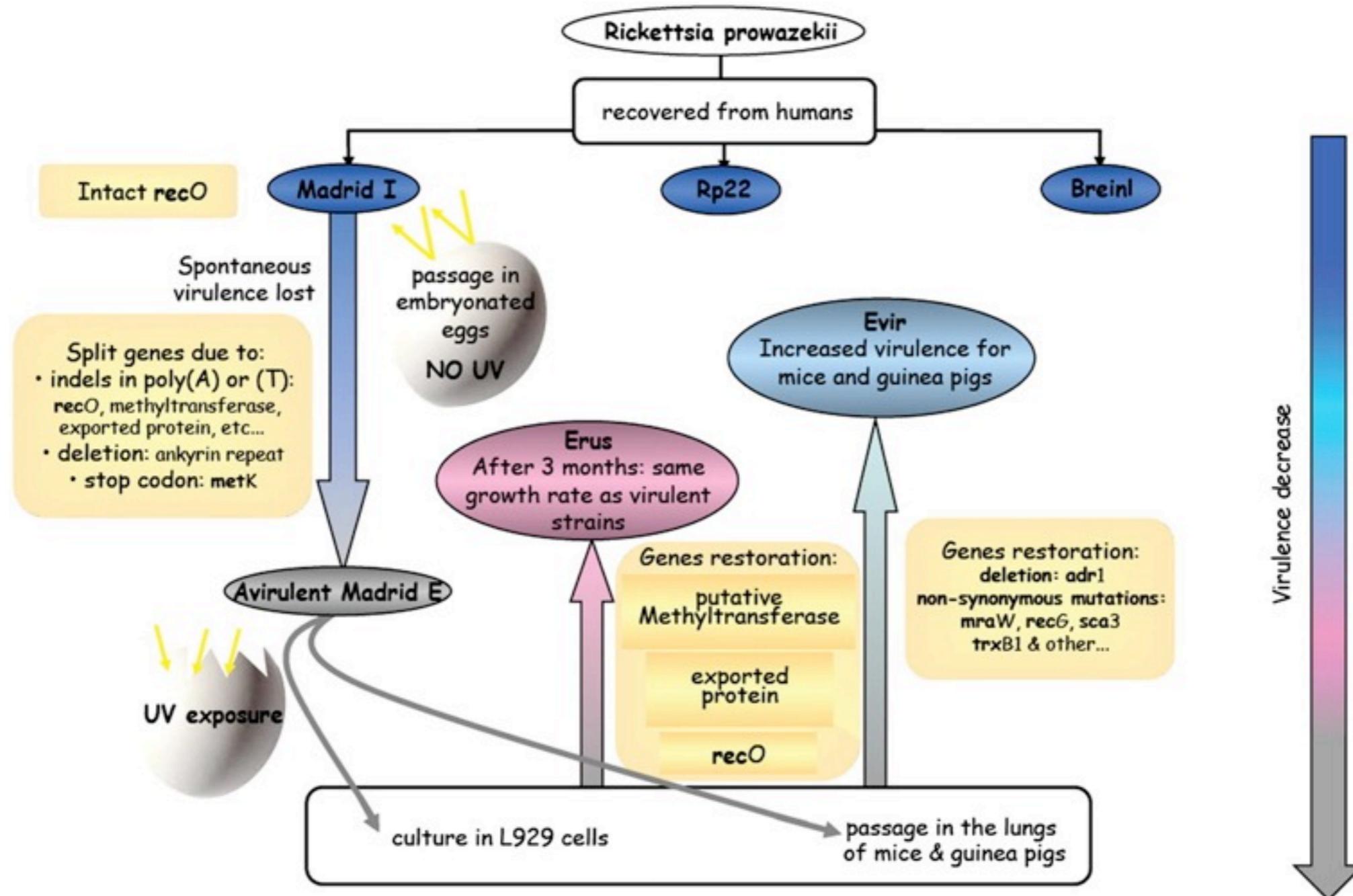
Match the pertinence of an object with its visual salience.

Apply visual organization Gestalt principles.





EVERYTHING IS EMPHASIZED



NOTHING IS EMPHASIZED

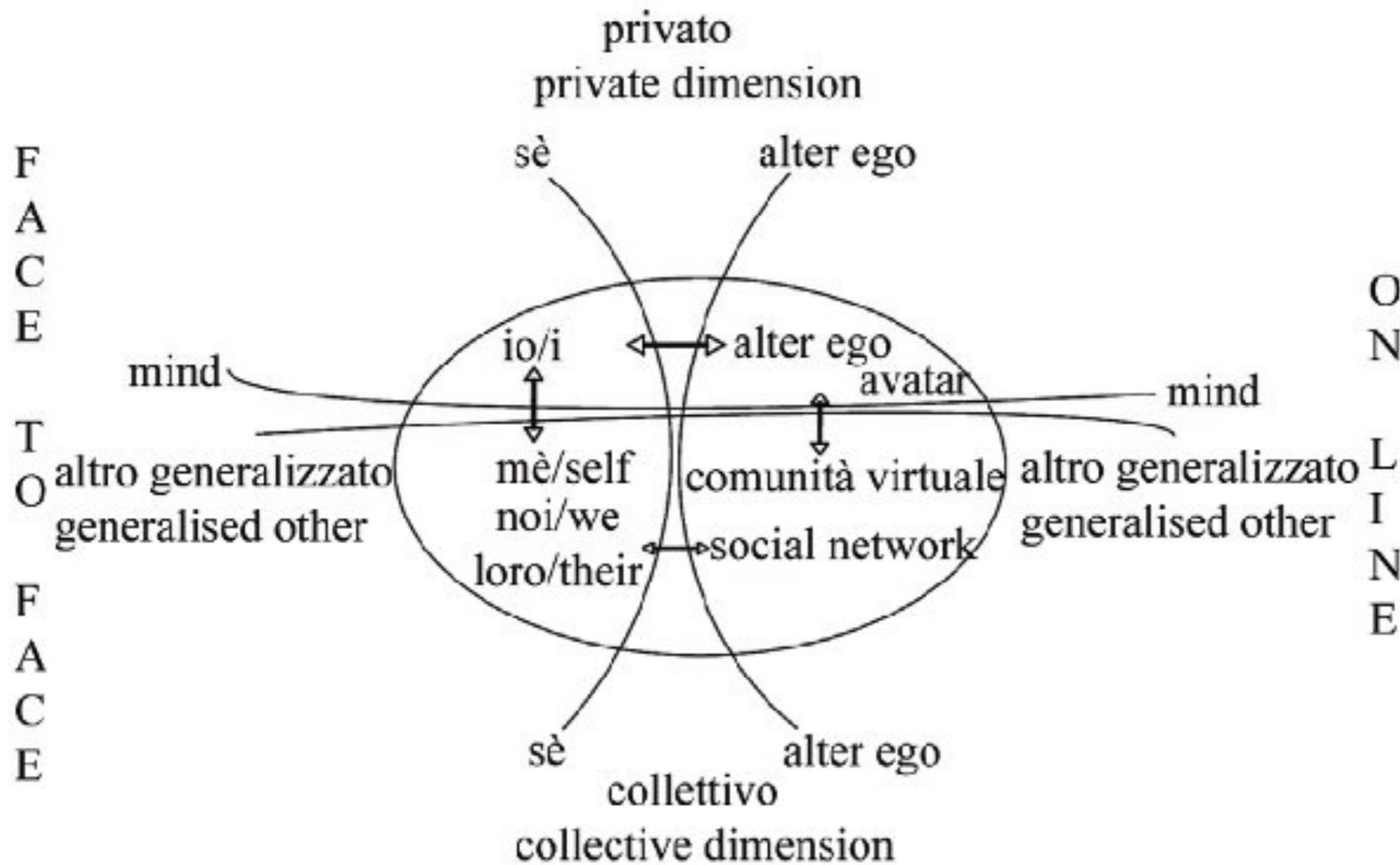
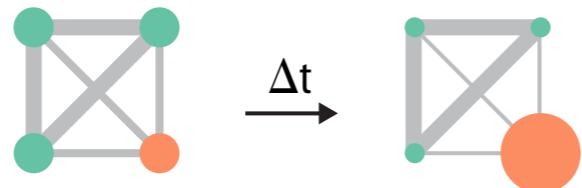


Figure 2.
Severino's flowchart.

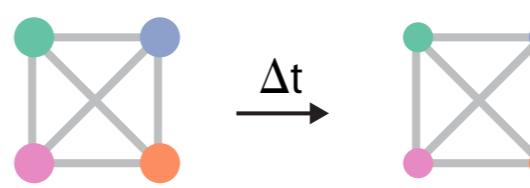
MATCH VISUAL SALIENCE TO RELEVANCE

ATTENTIONAL CAPTURE

DISTINCT



HETEROGENEOUS



inhibitory interaction



low high

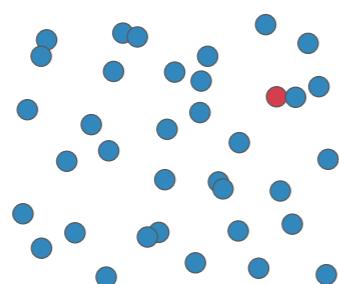
neural response



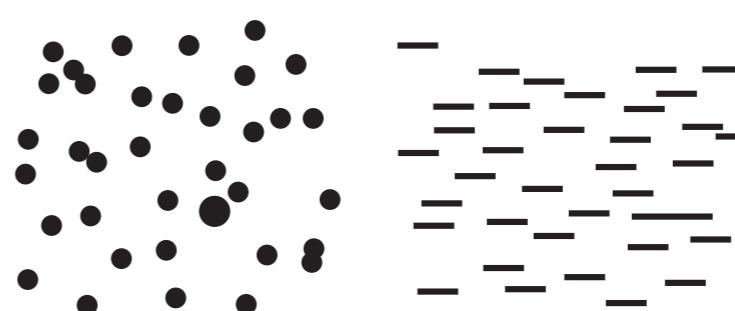
low high

SALIENCE

HIGH



LOW

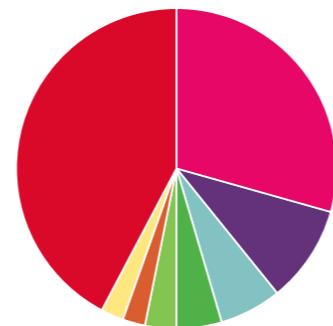


USE BREWER PALETTES

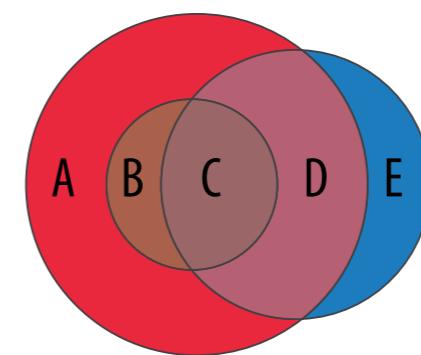
*one color
dominates*



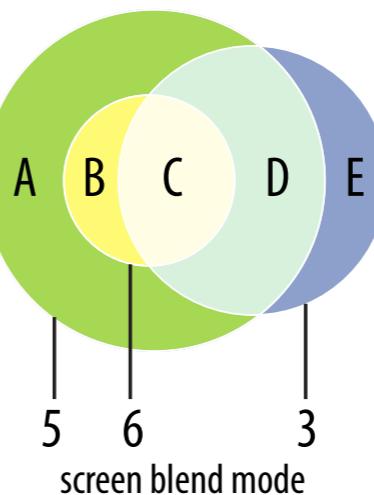
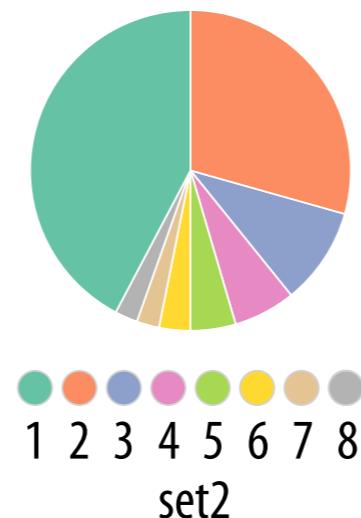
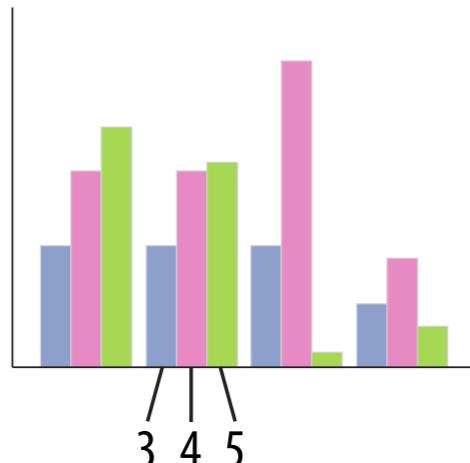
*difficult to
distinguish*



murky



recolored with Brewer palettes



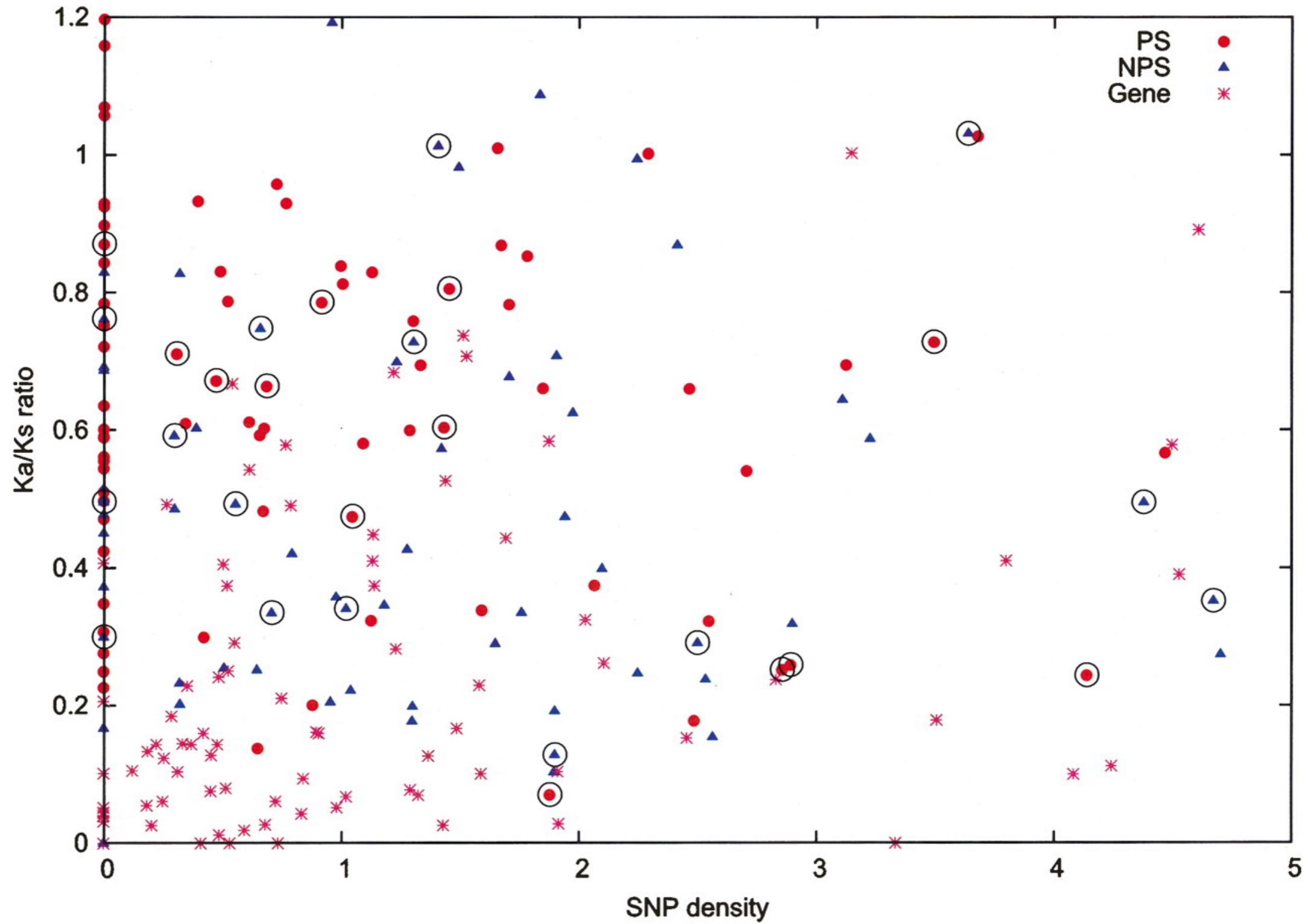
Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, et al. (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Res* 13: 2725-2735.

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.

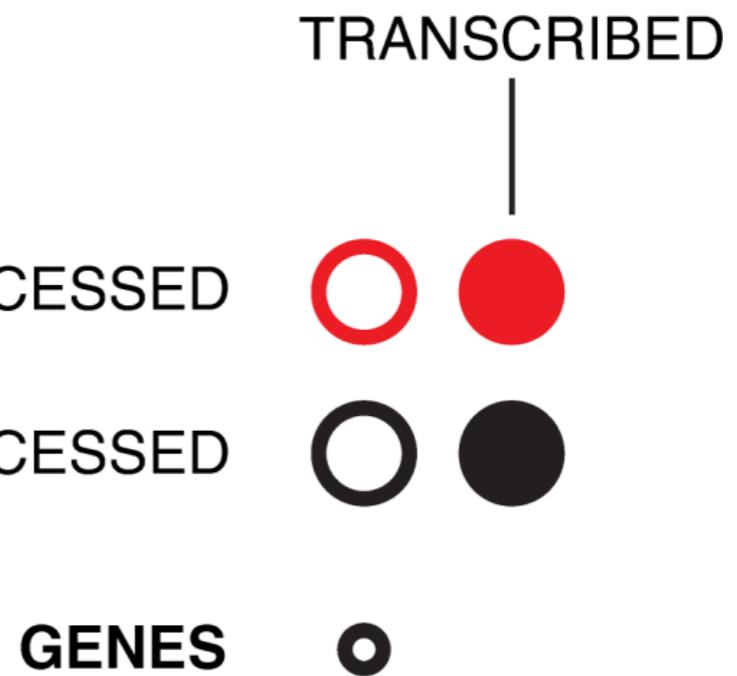


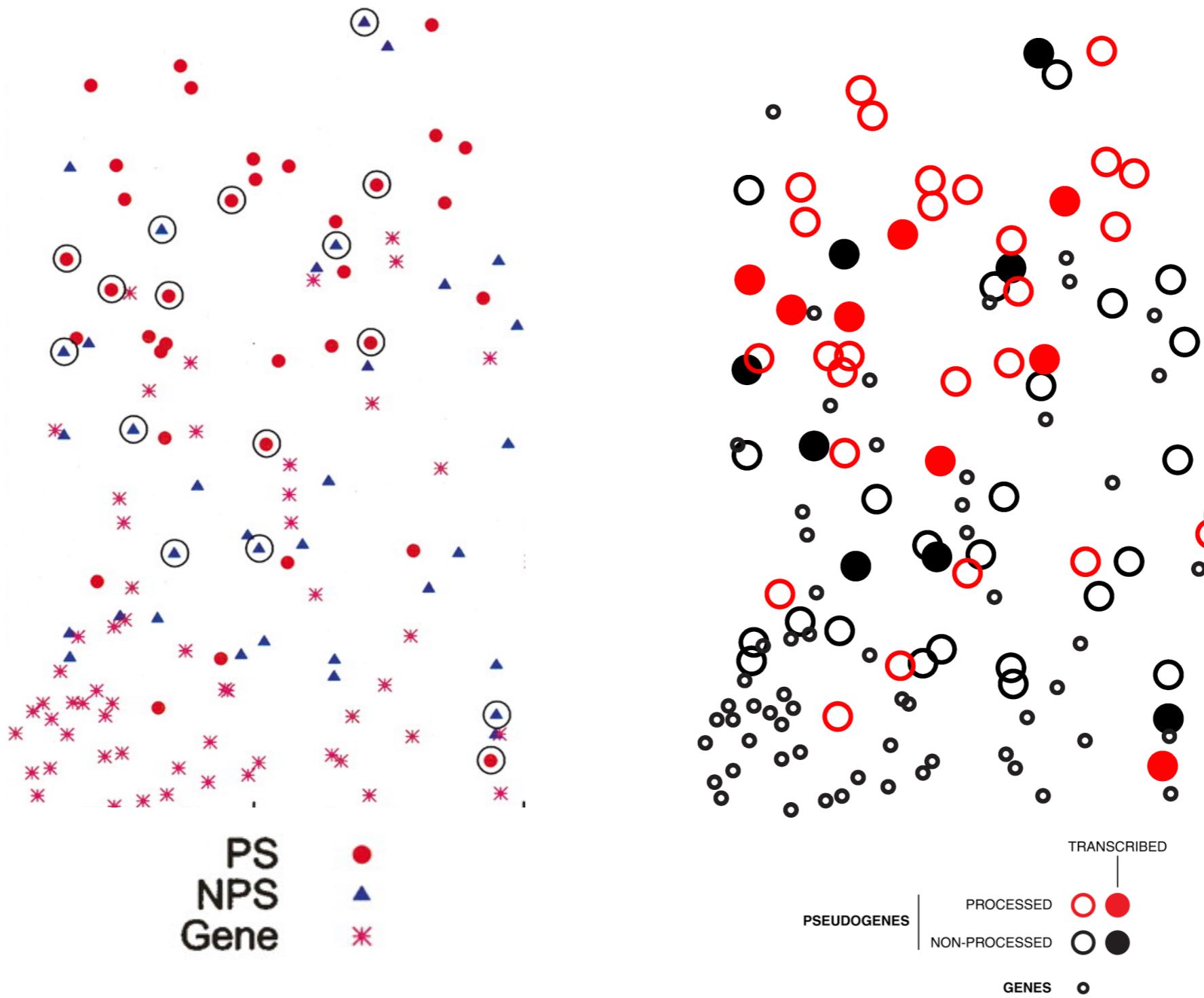
PS
NPS
Gene

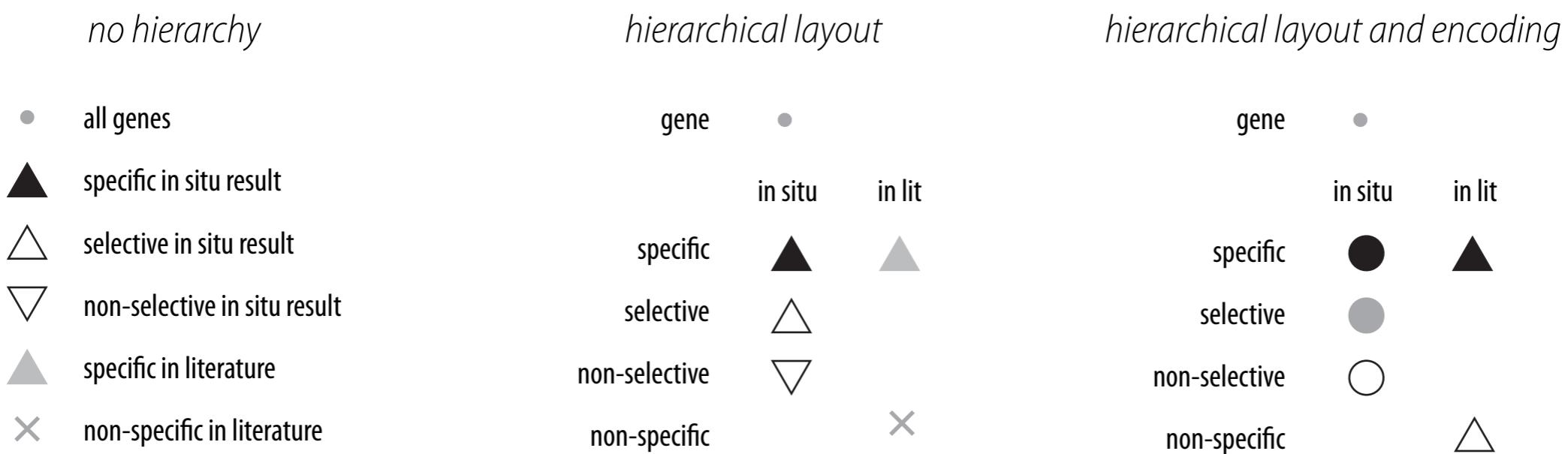


PSEUDOGENES

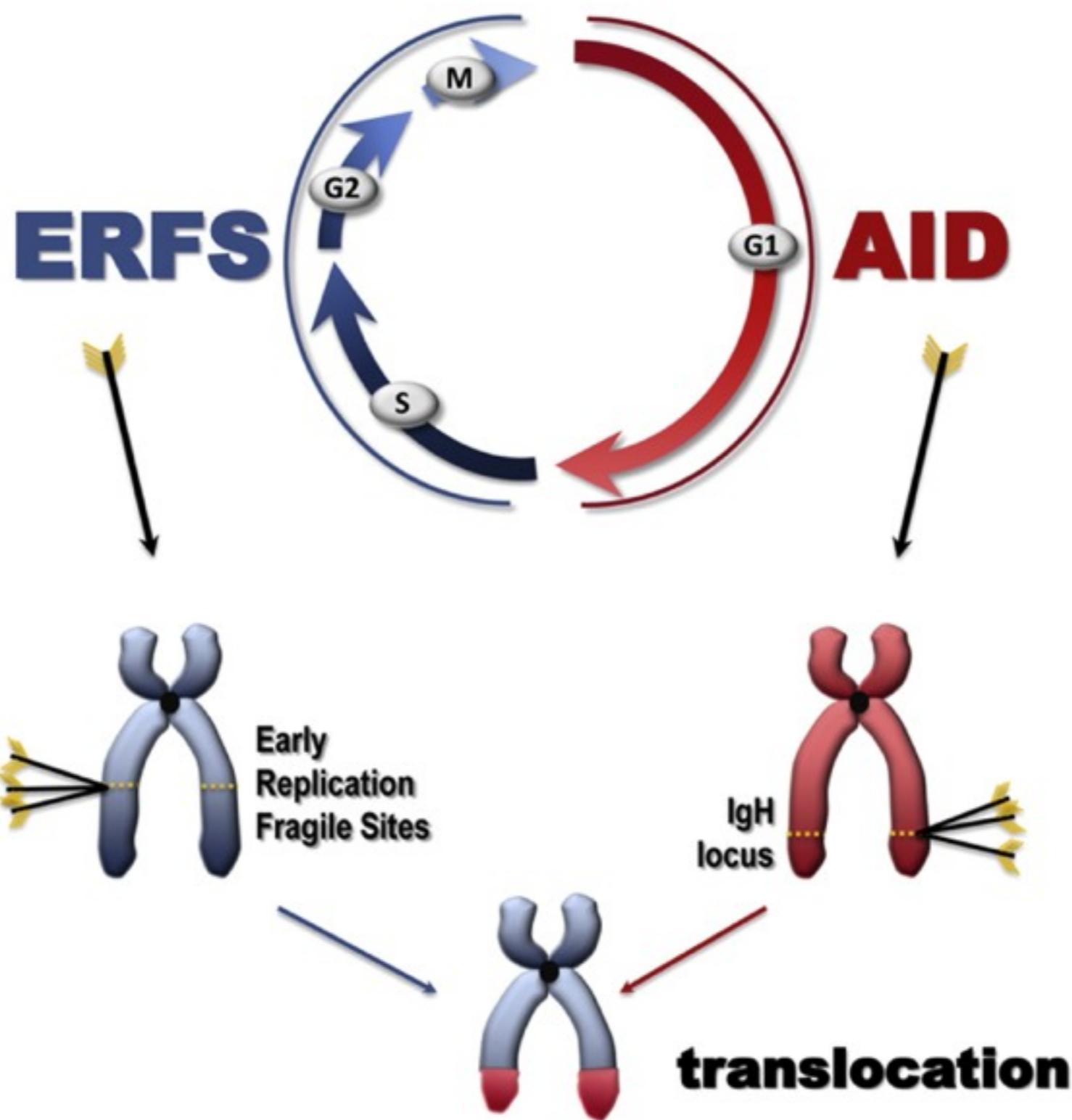
PROCESSED
NON-PROCESSED

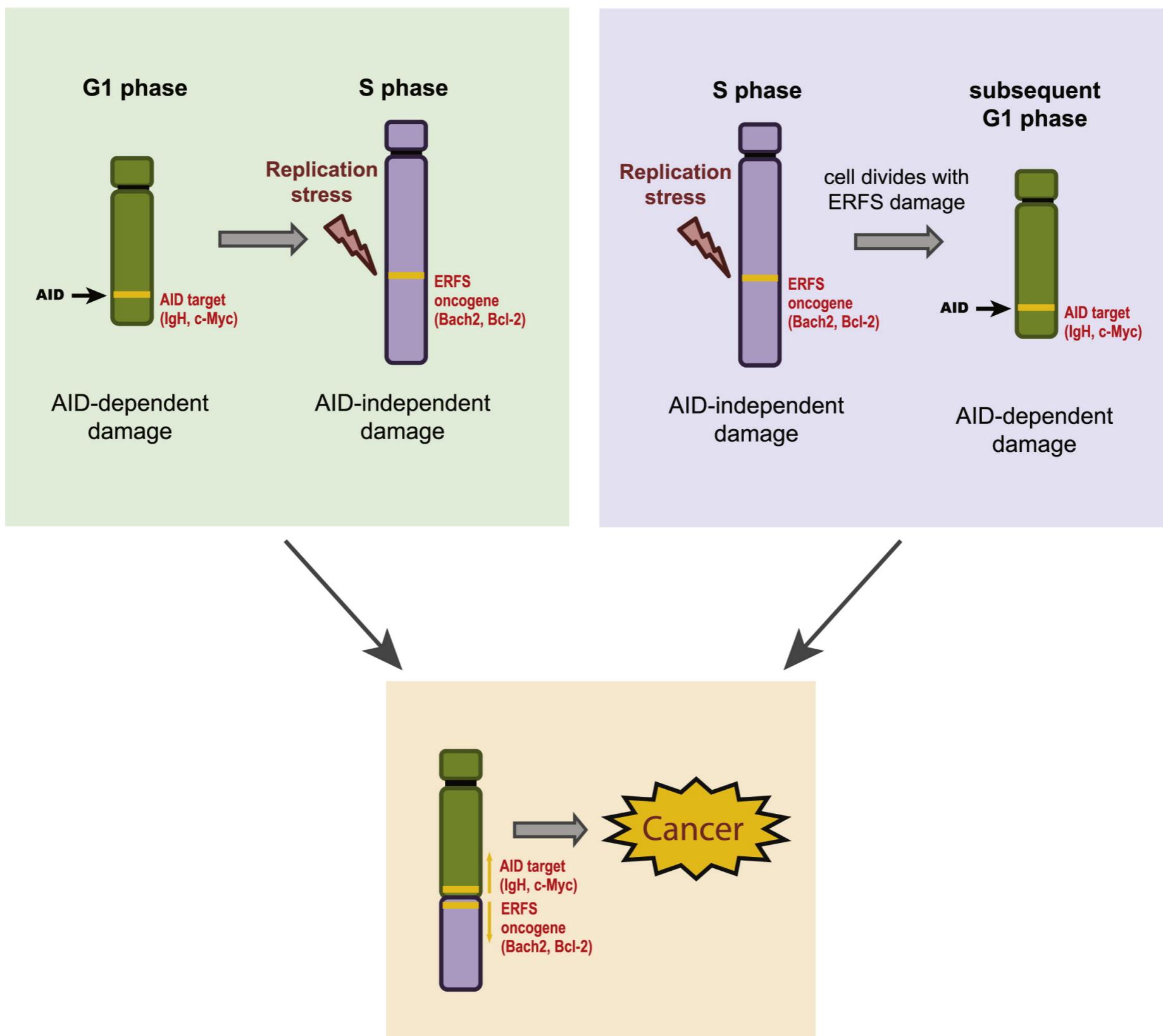


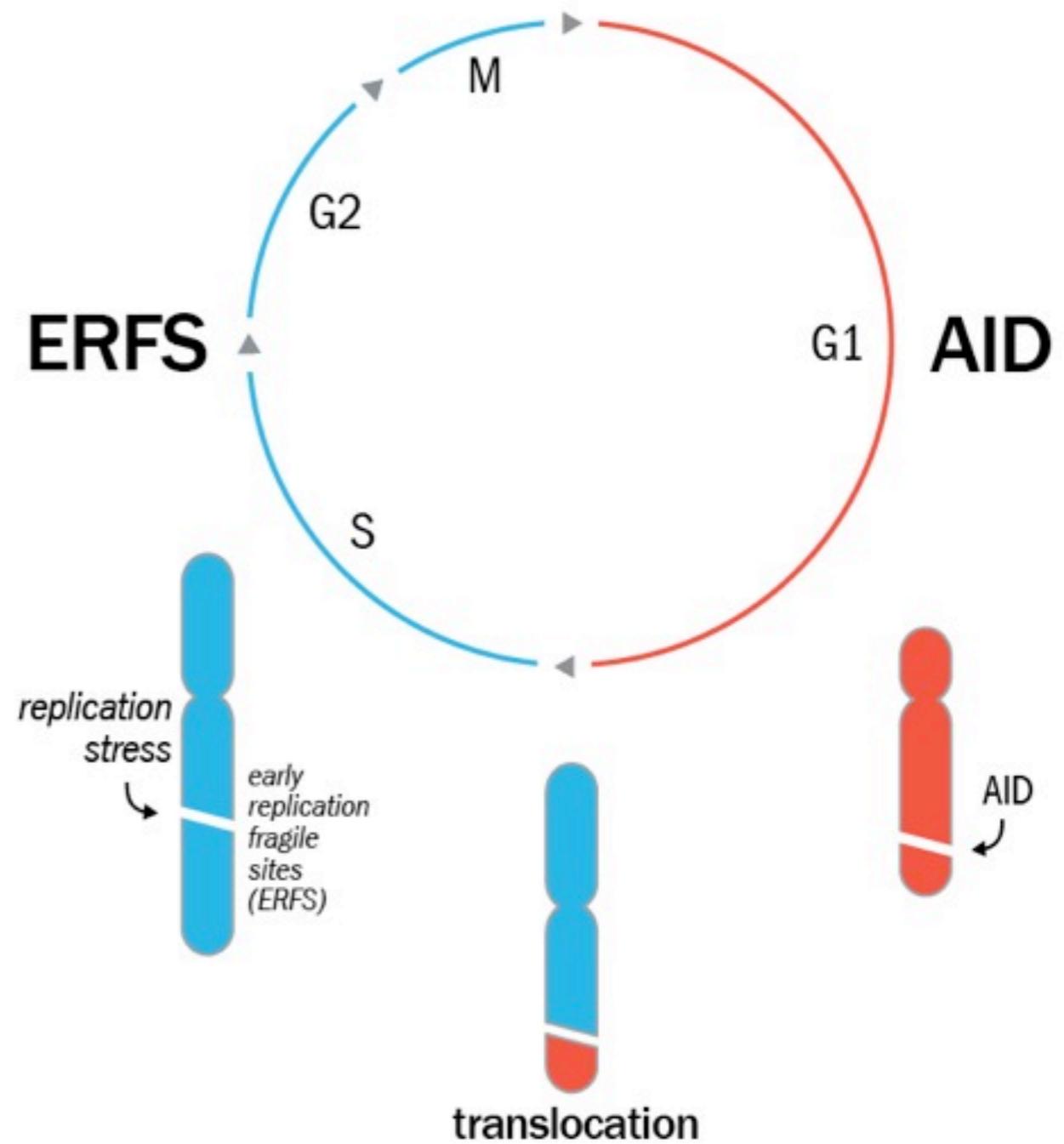
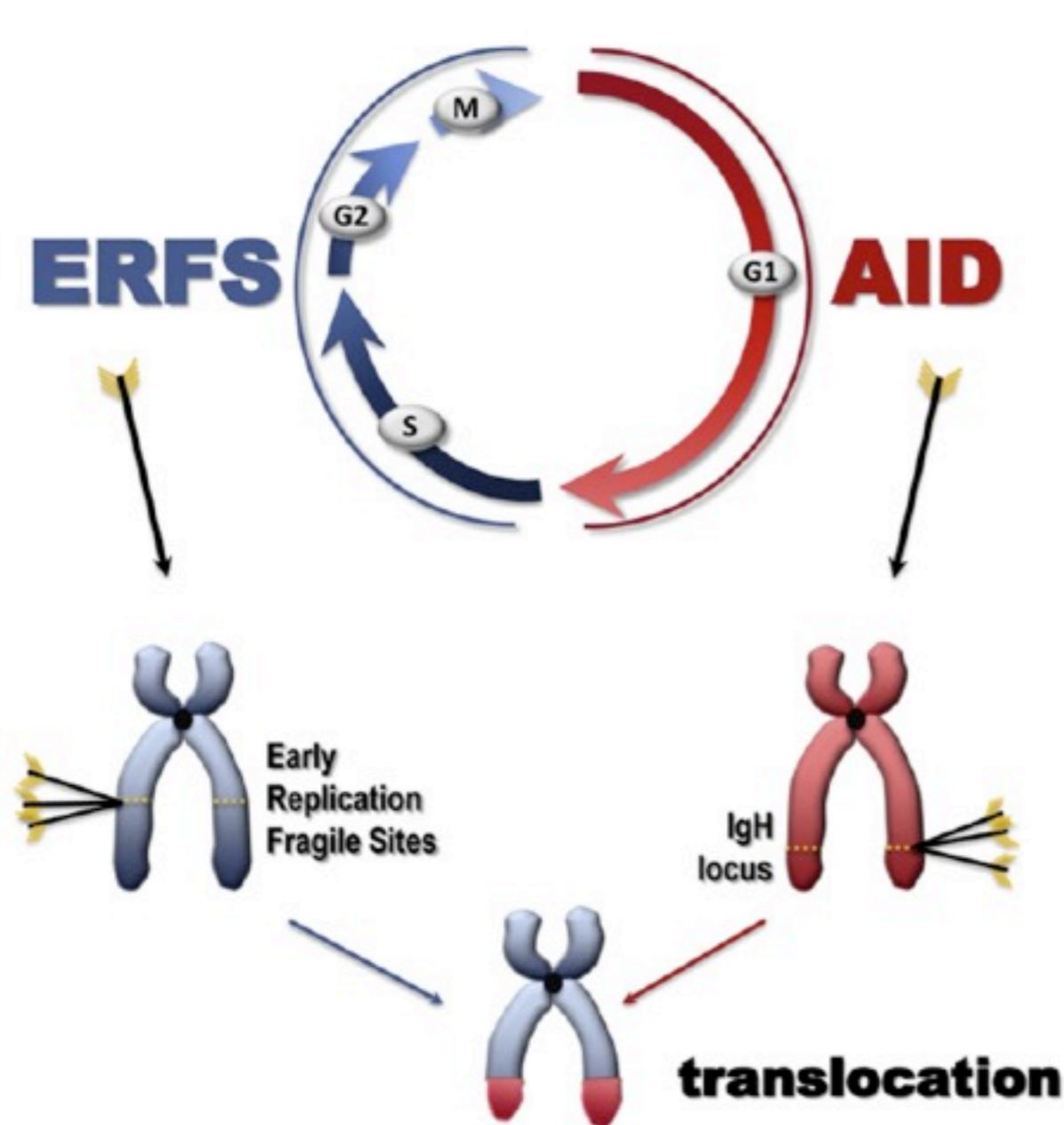


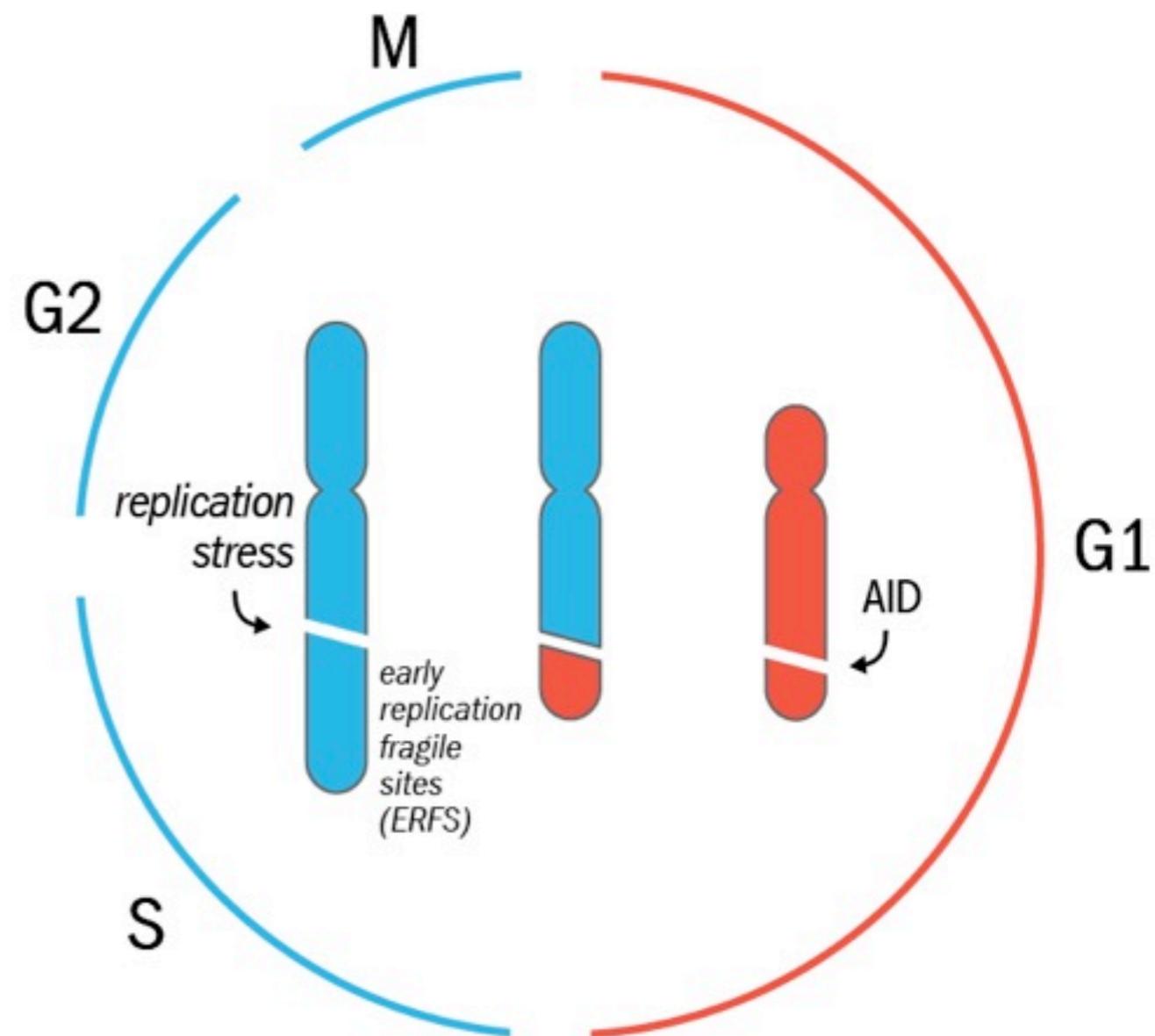
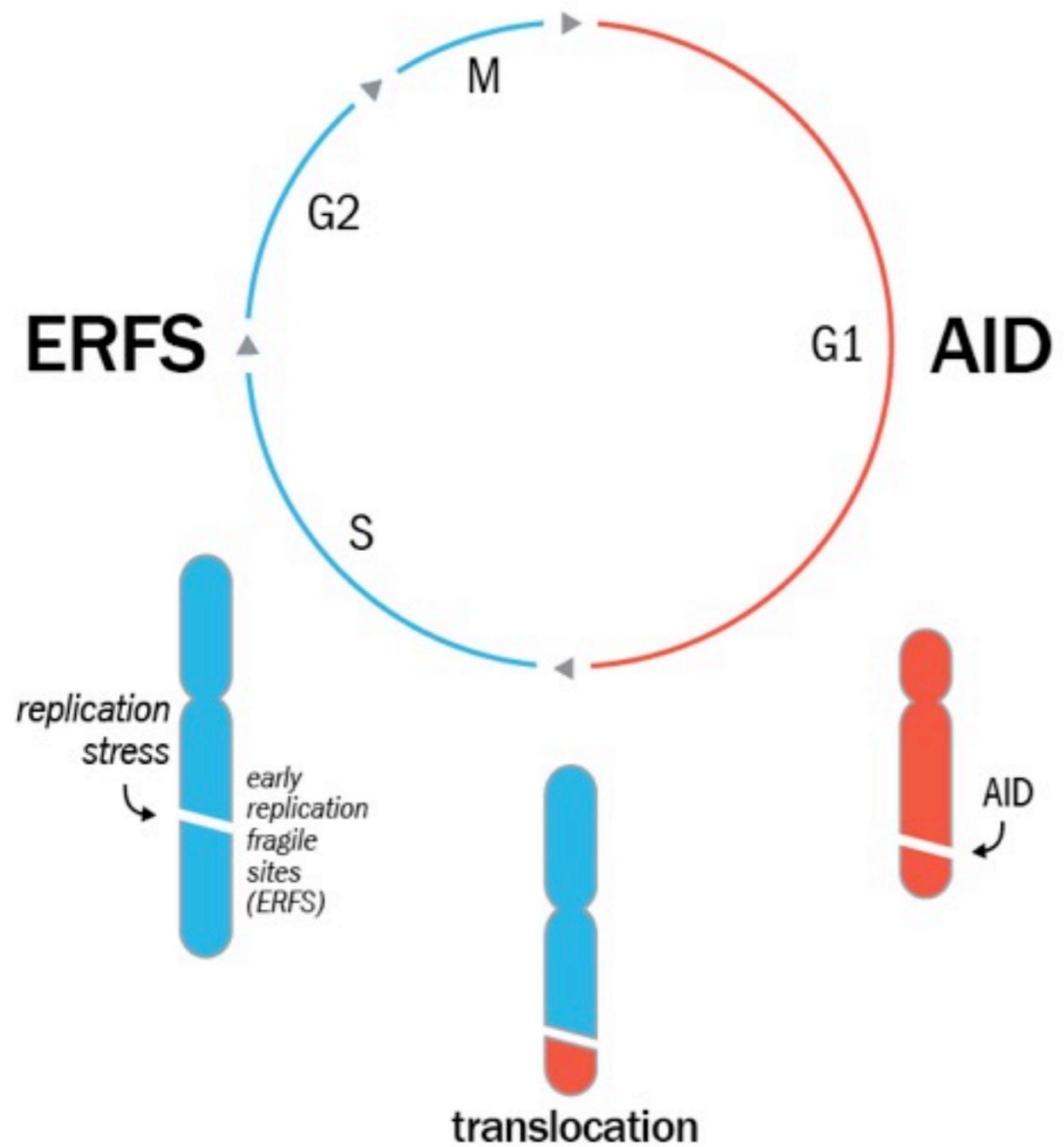


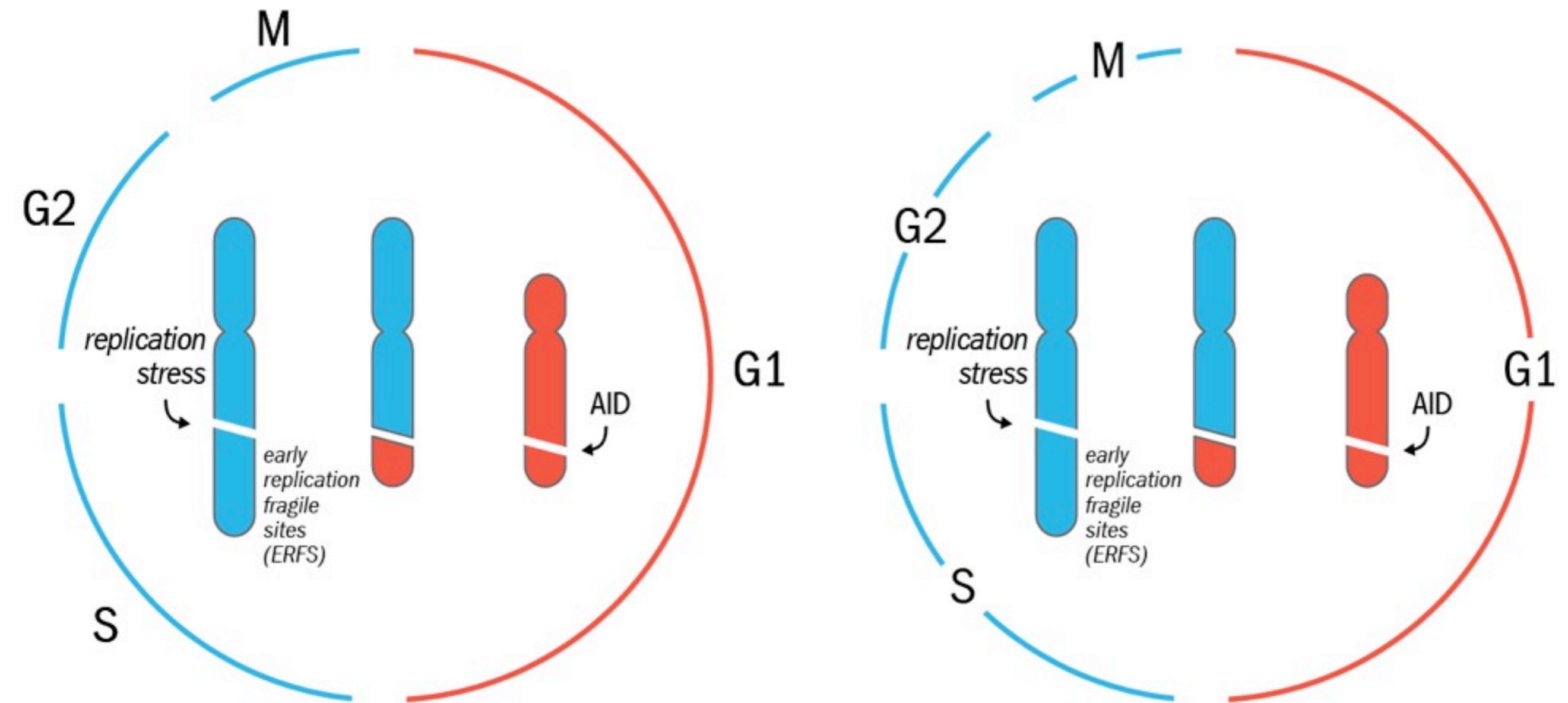
VISUAL ABSTRACT REDESIGN

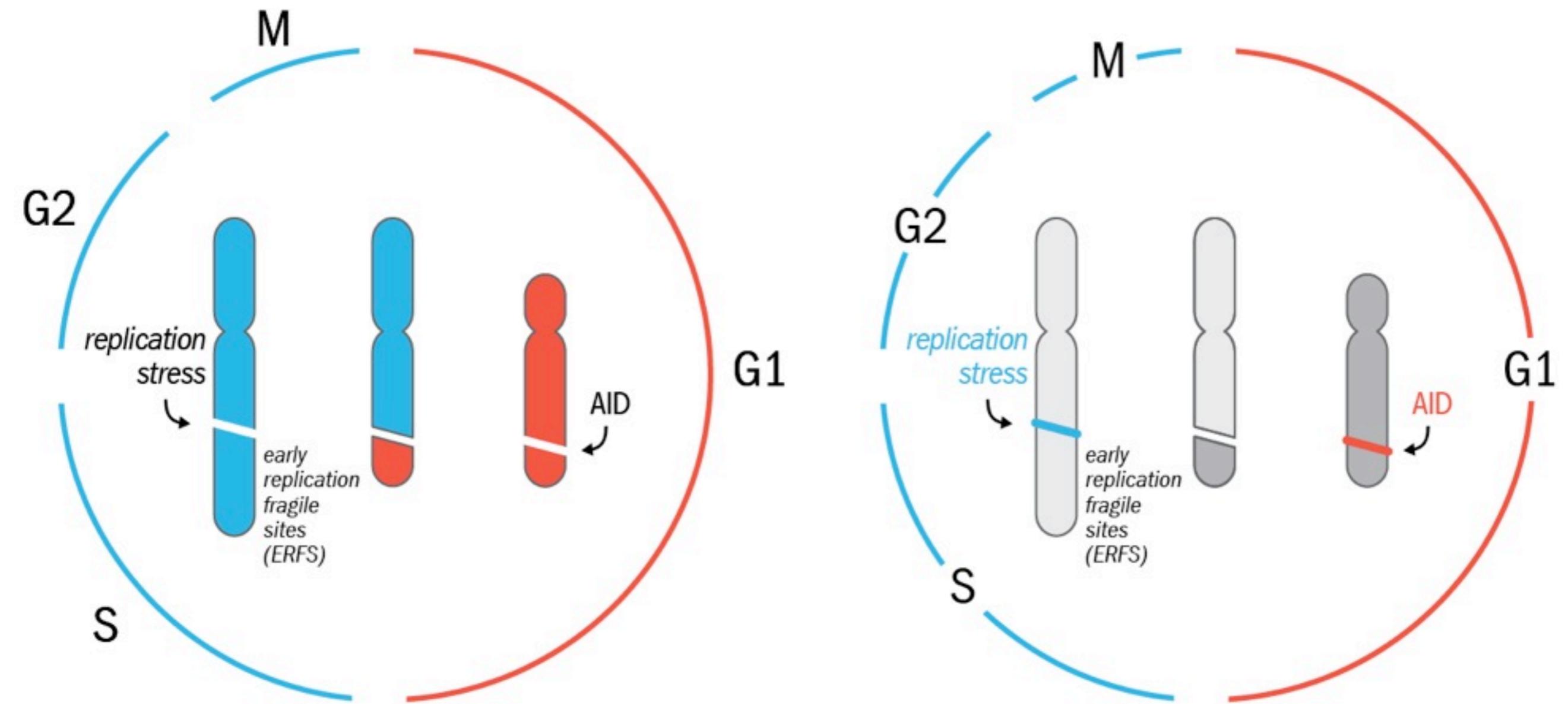


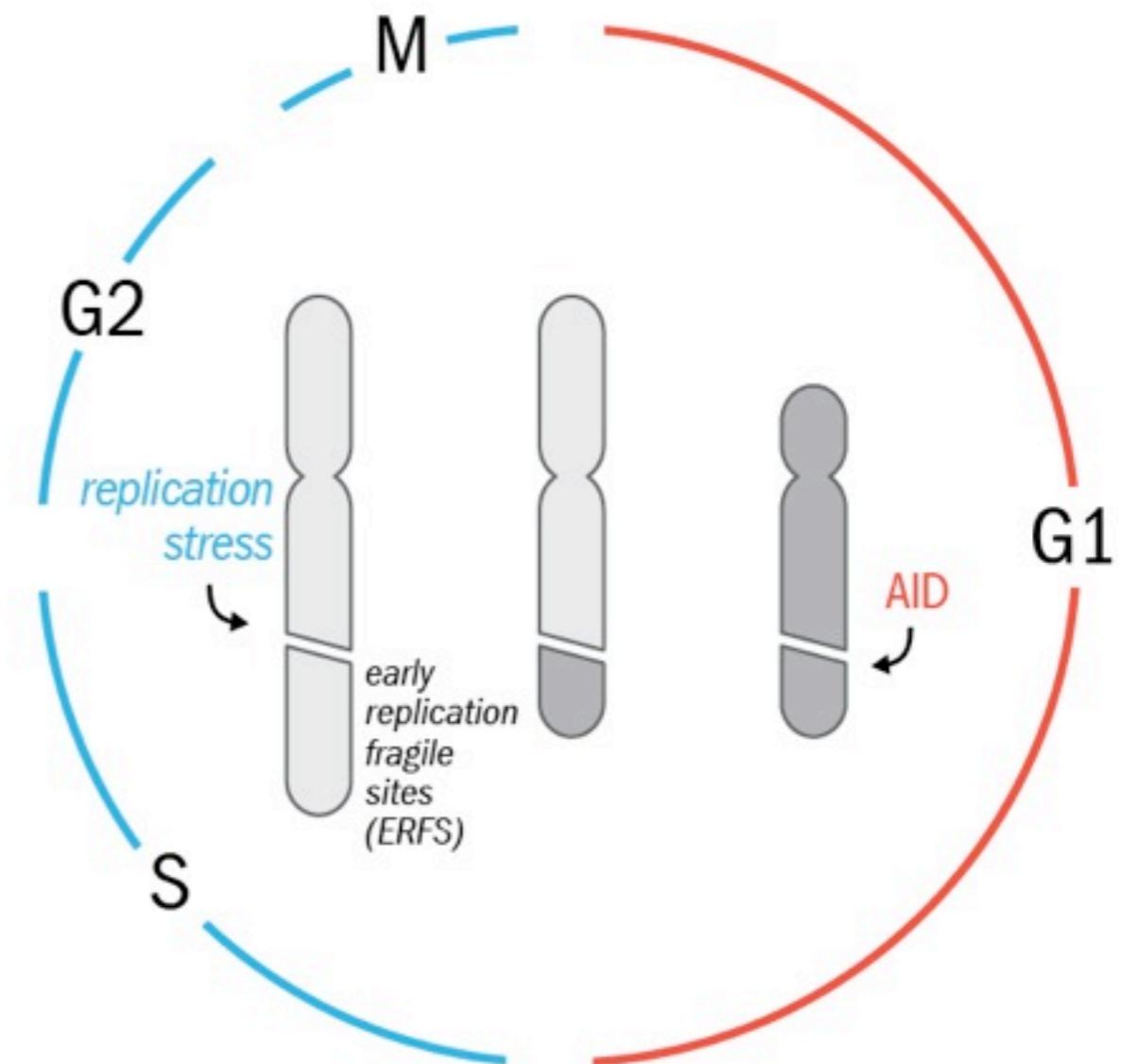
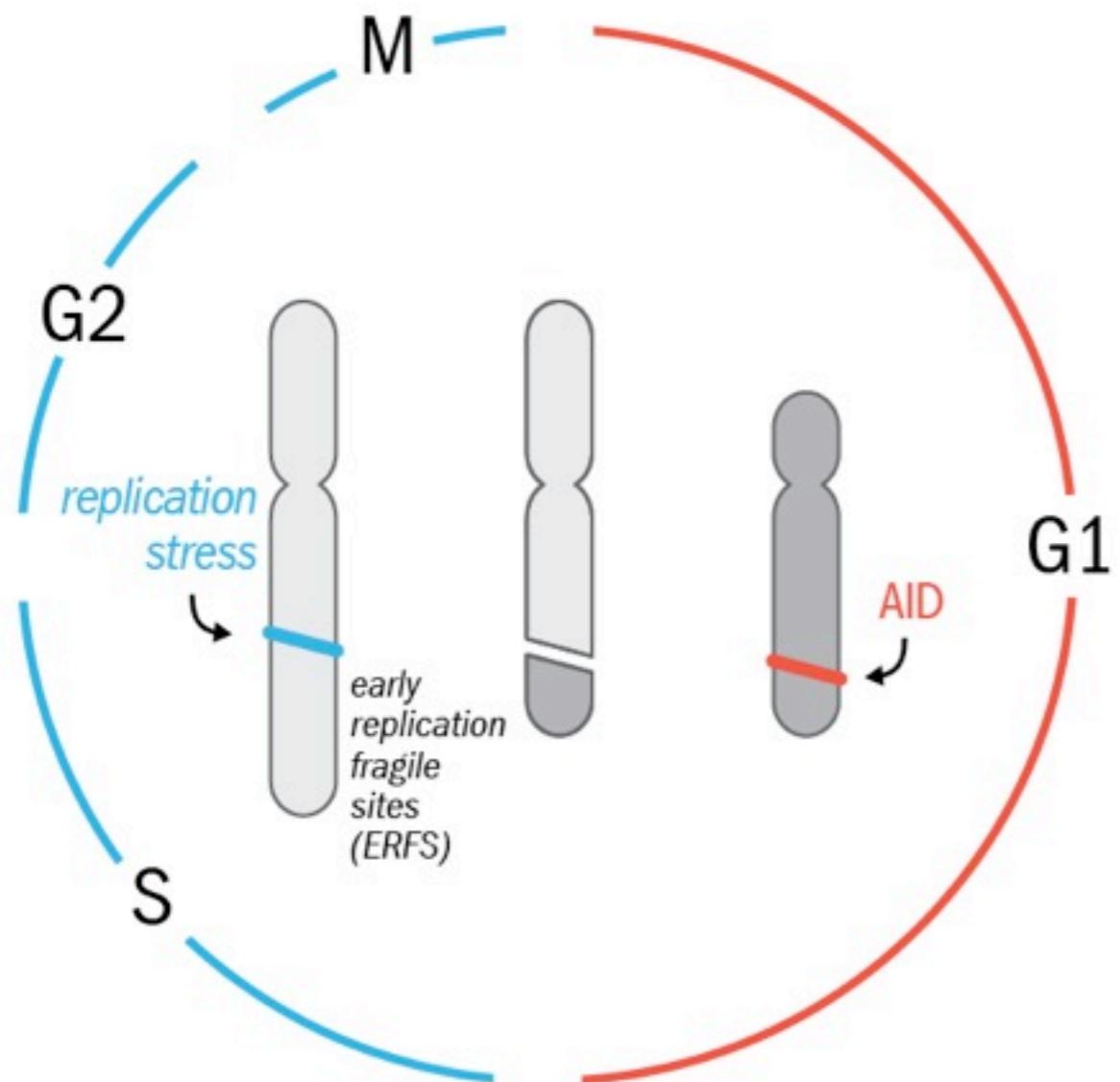


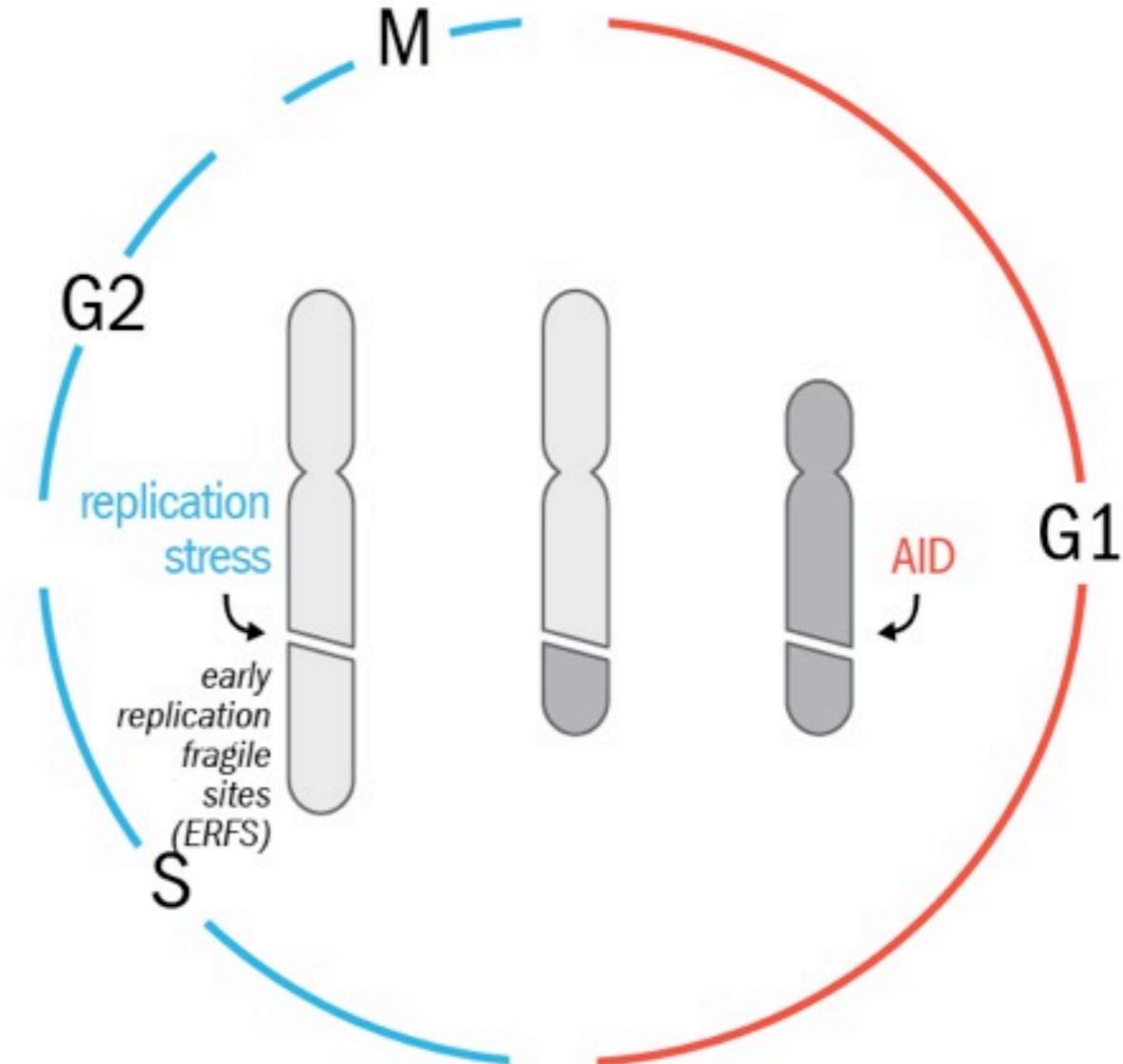
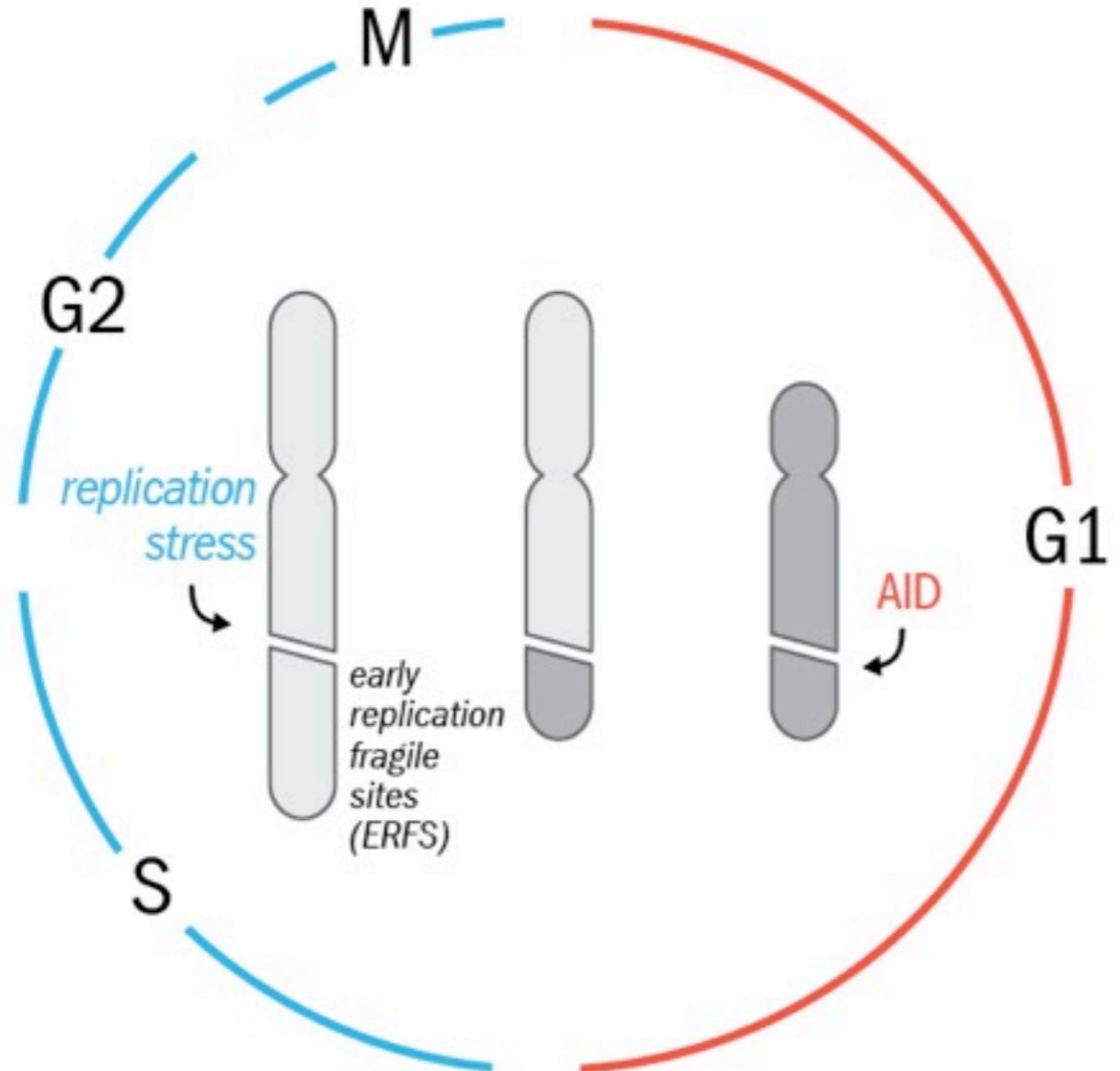


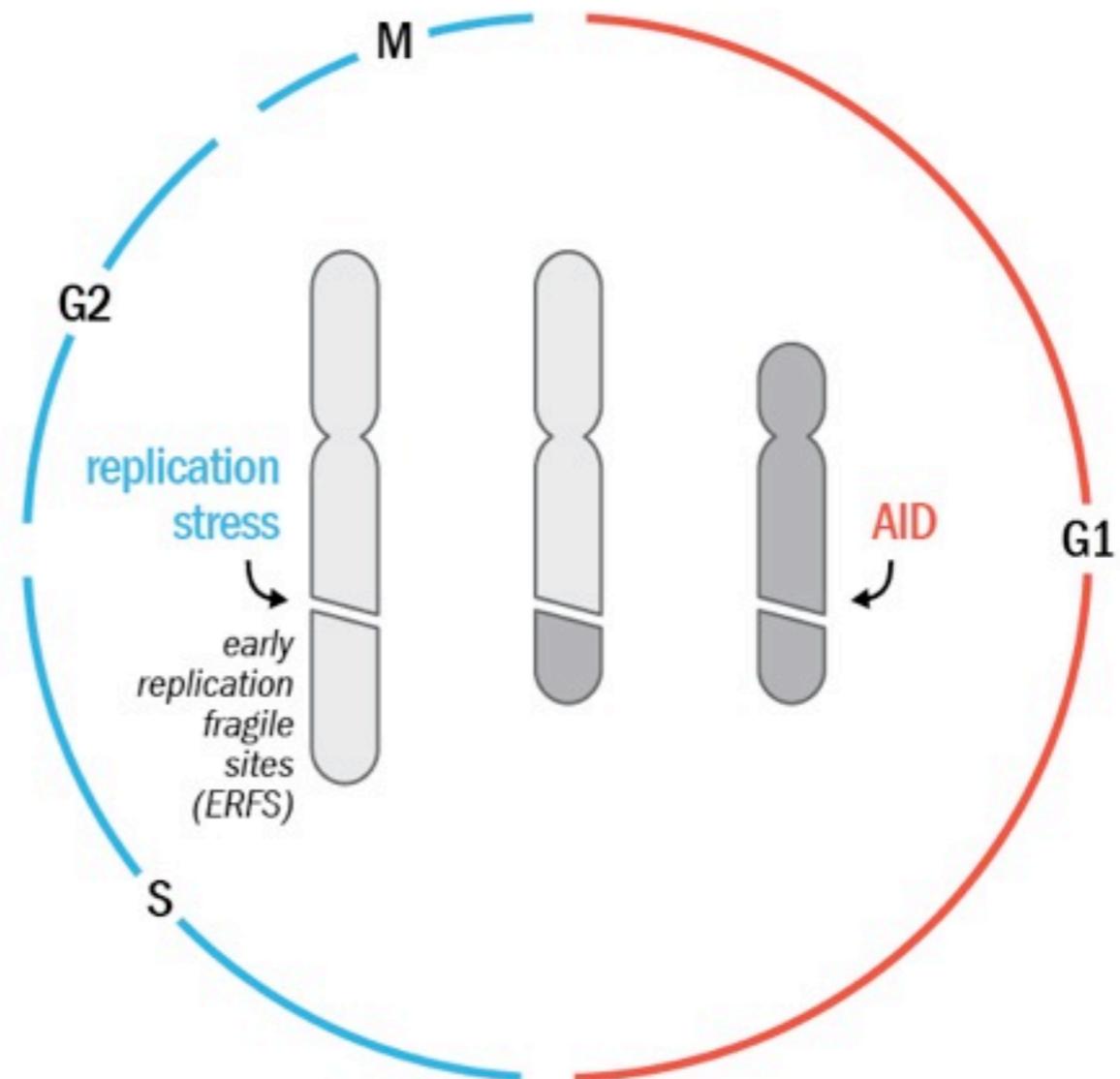
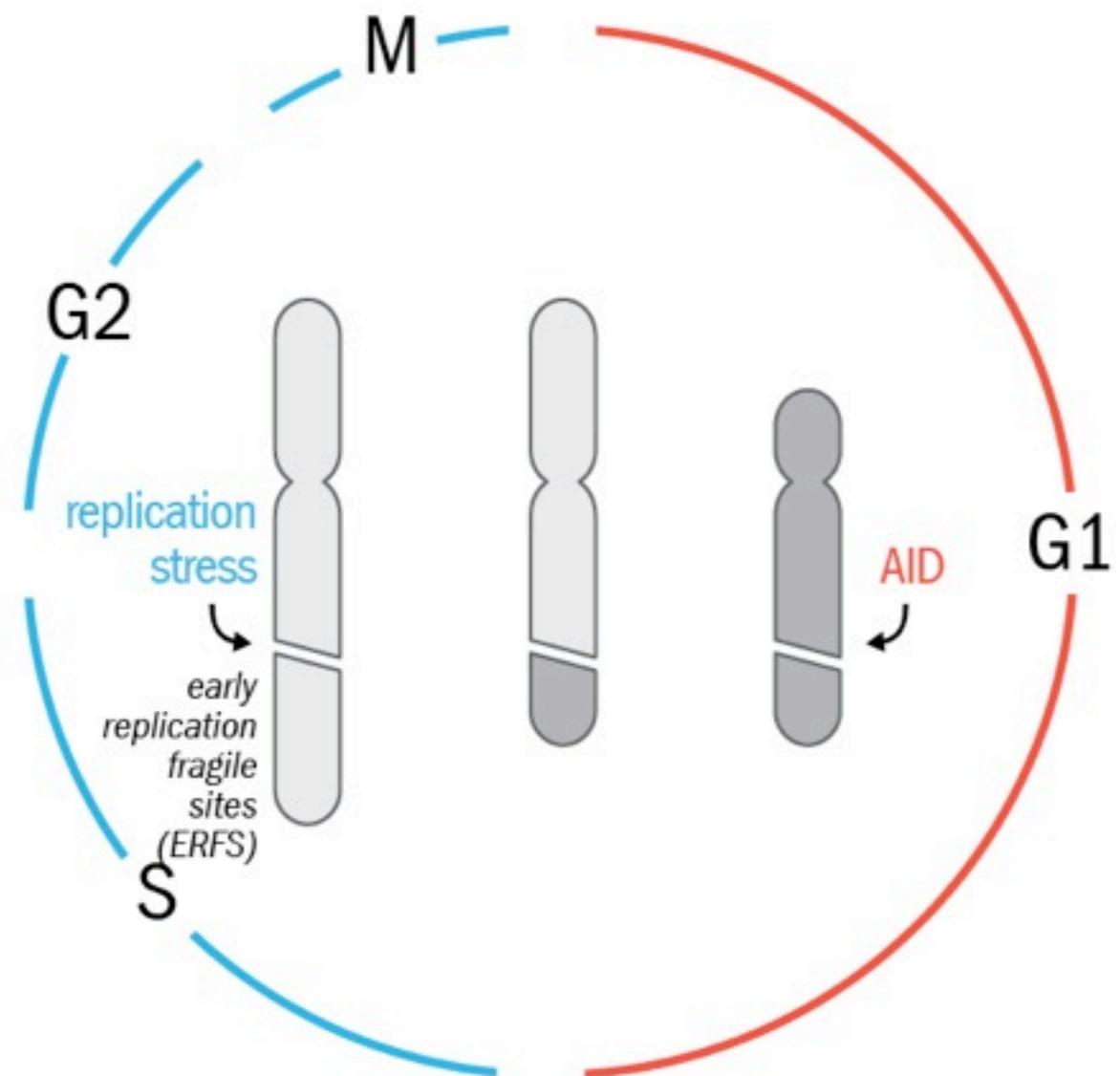


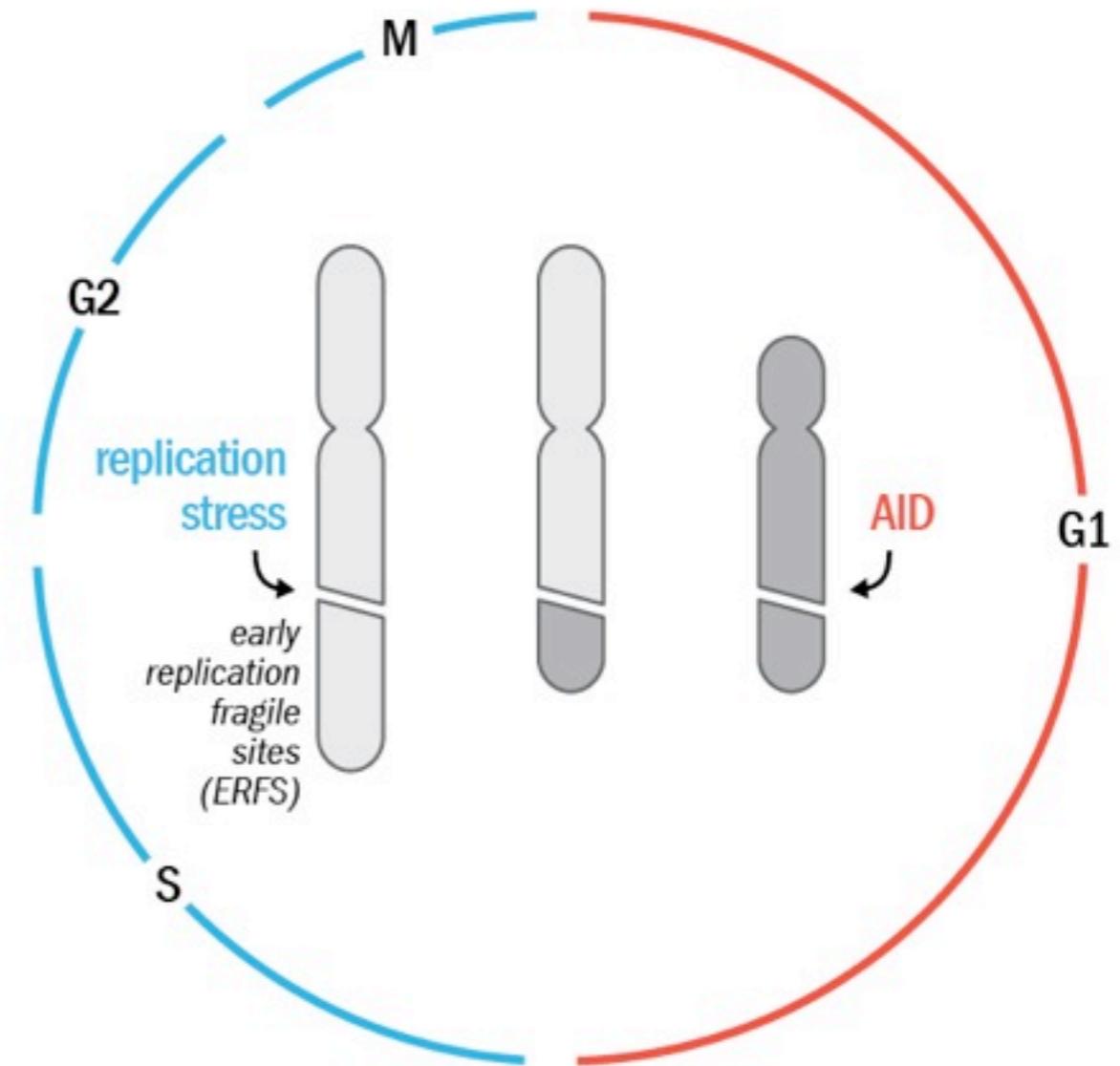
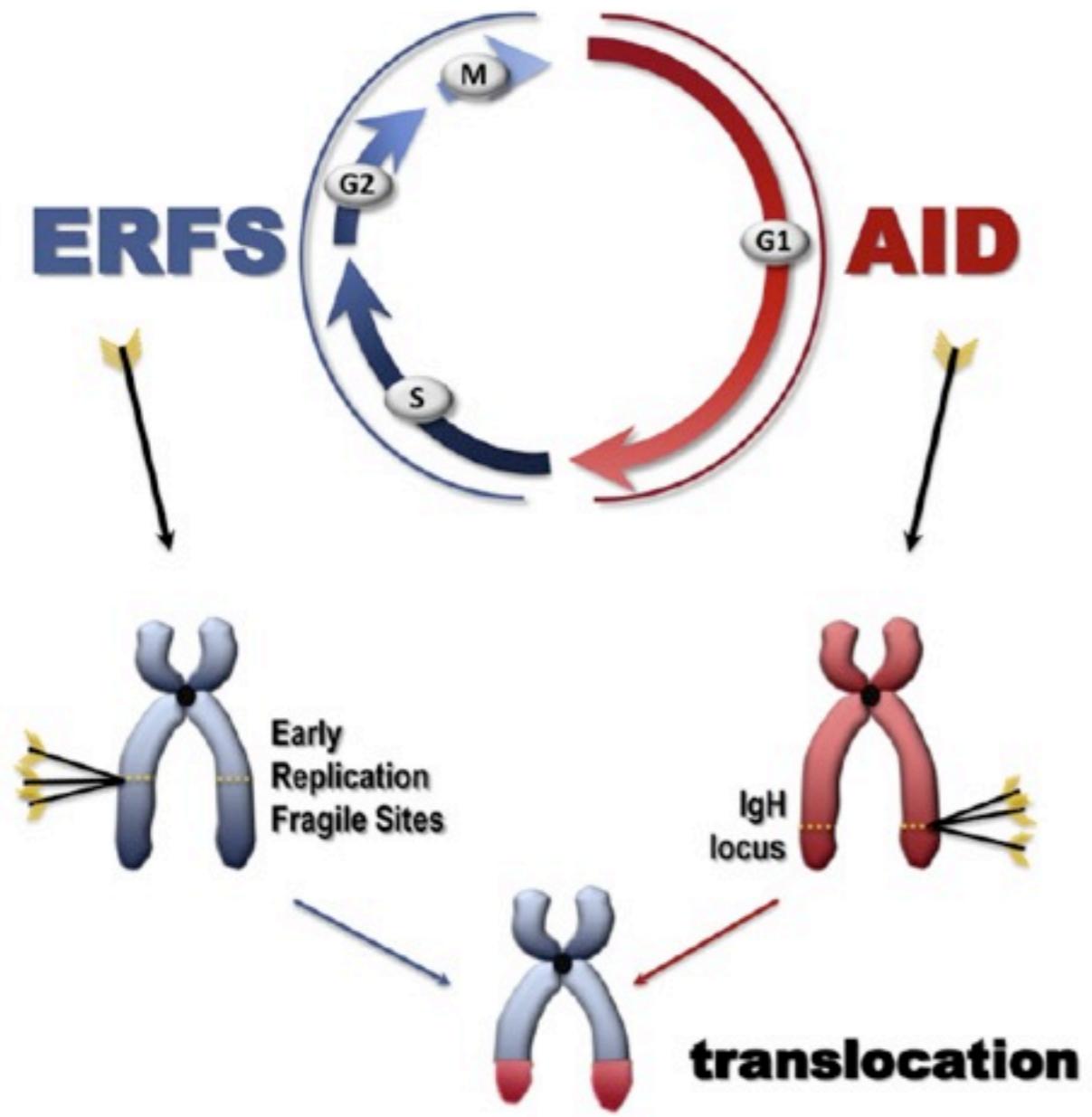












VISUAL ABSTRACT CREATION

Identifying a deliverable message.

KEYWORDS

cancer

cancer genomics

tumor heterogeneity

next-generation sequencing

second-generation sequencing

third-generation sequencing

mutation discovery

whole genome sequencing

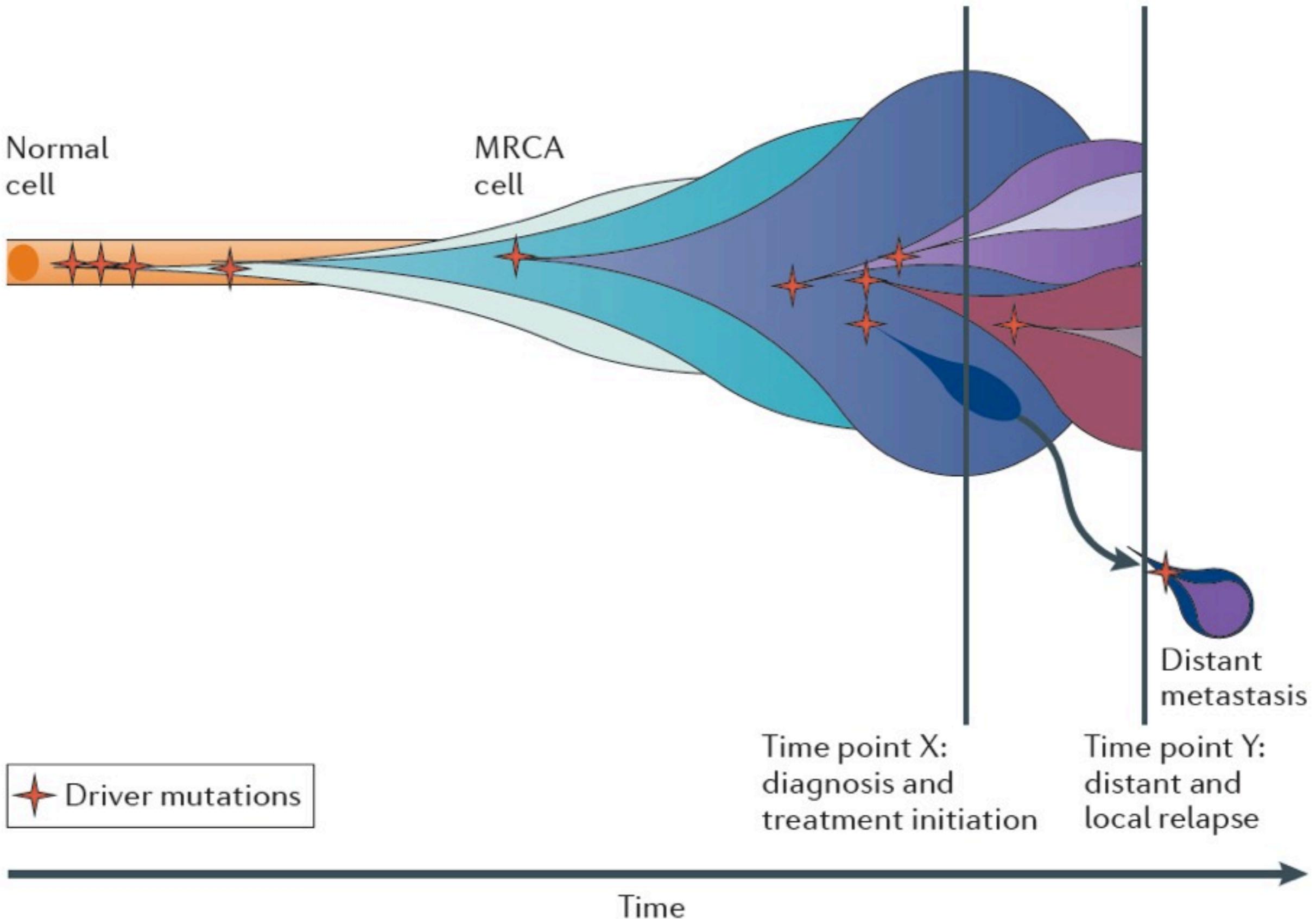
single molecule sequencing

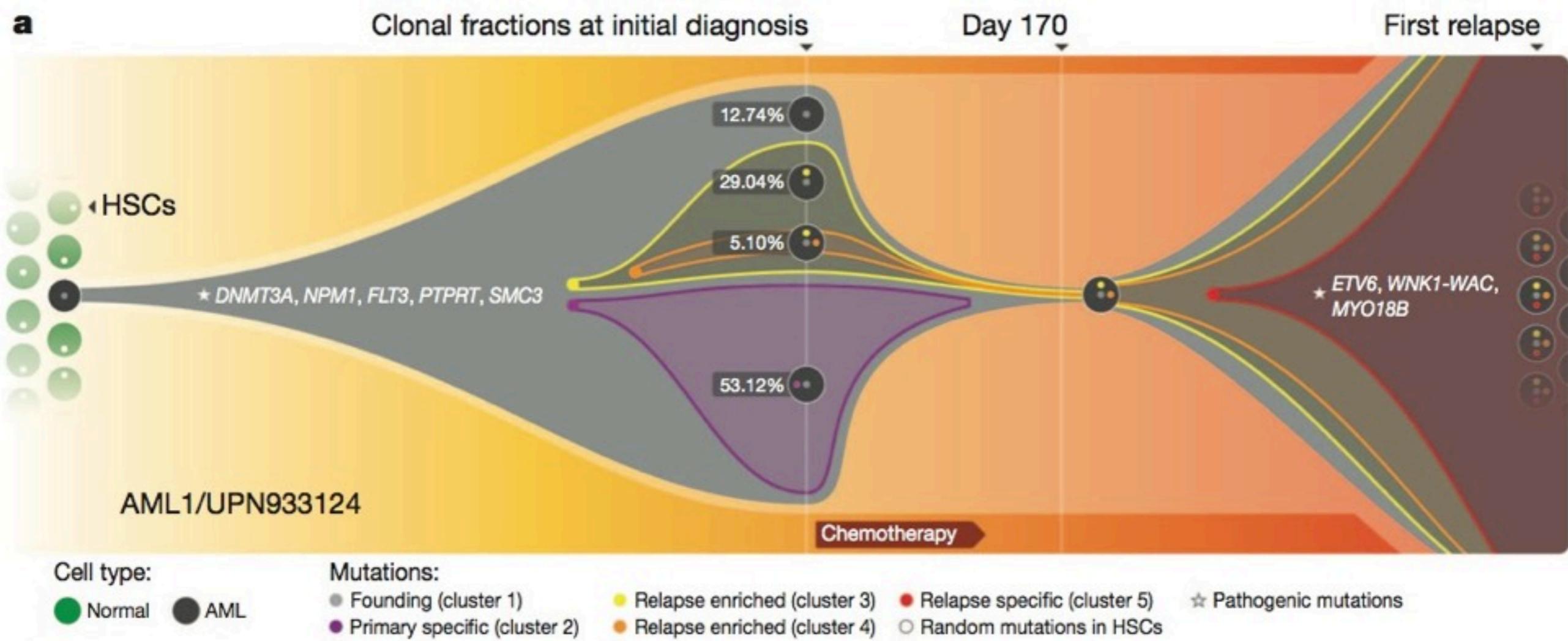
single cell sequencing

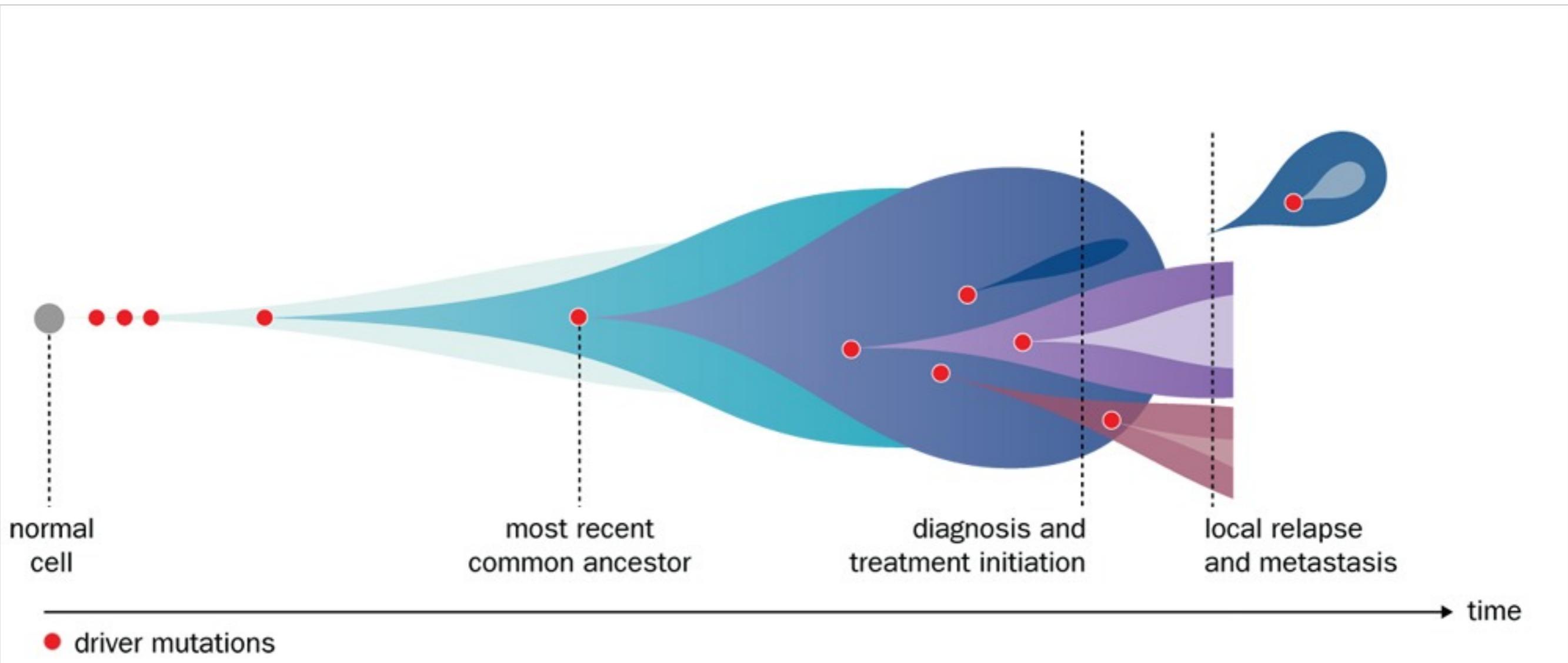
personalized medicine

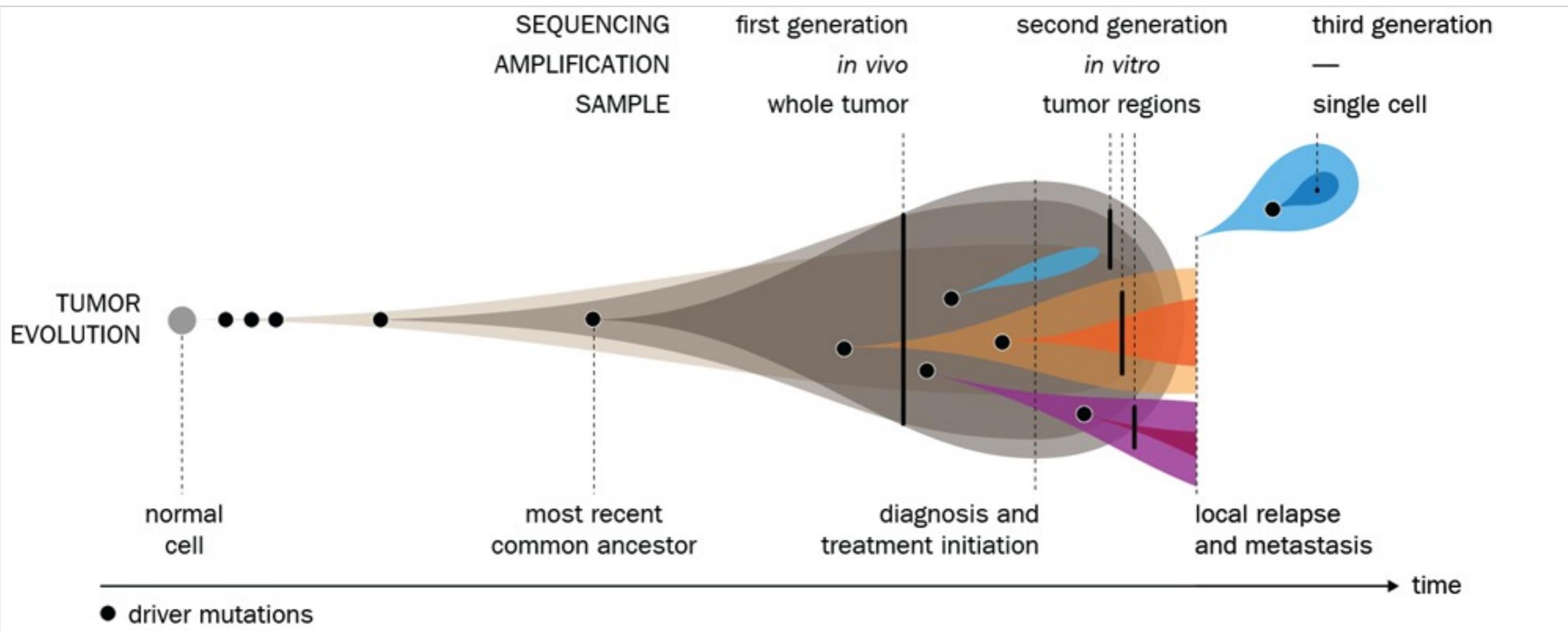
ELSEVIER FIGURE RESTRICTION

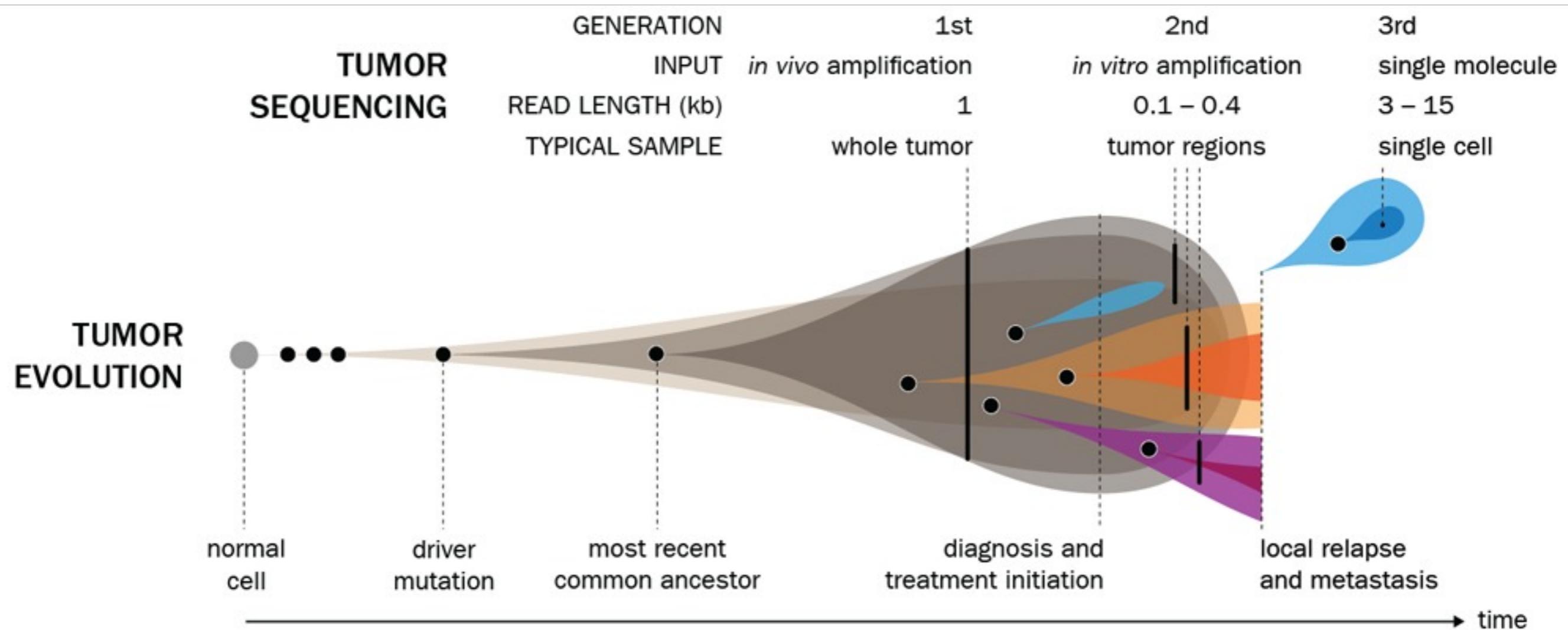
Please note that your image will be scaled proportionally to fit in the available window on ScienceDirect, **a 500 by 200 pixel rectangle.**

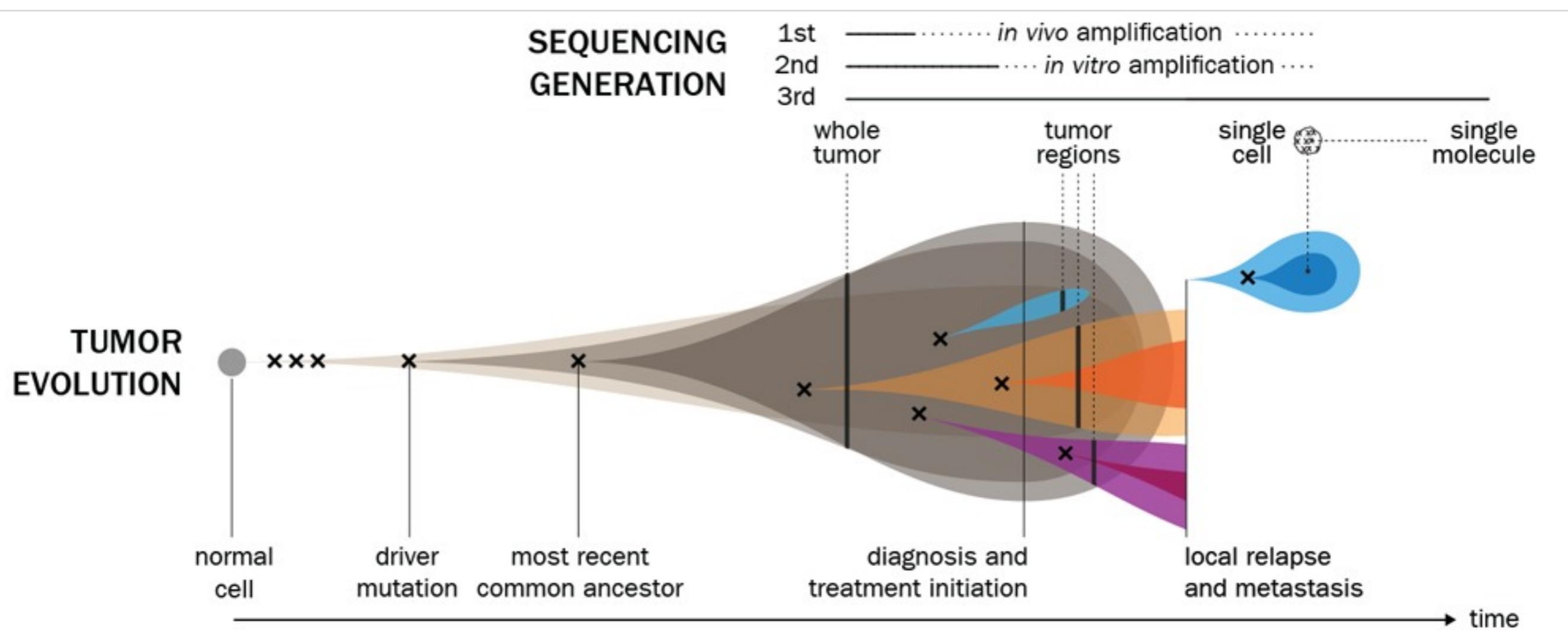


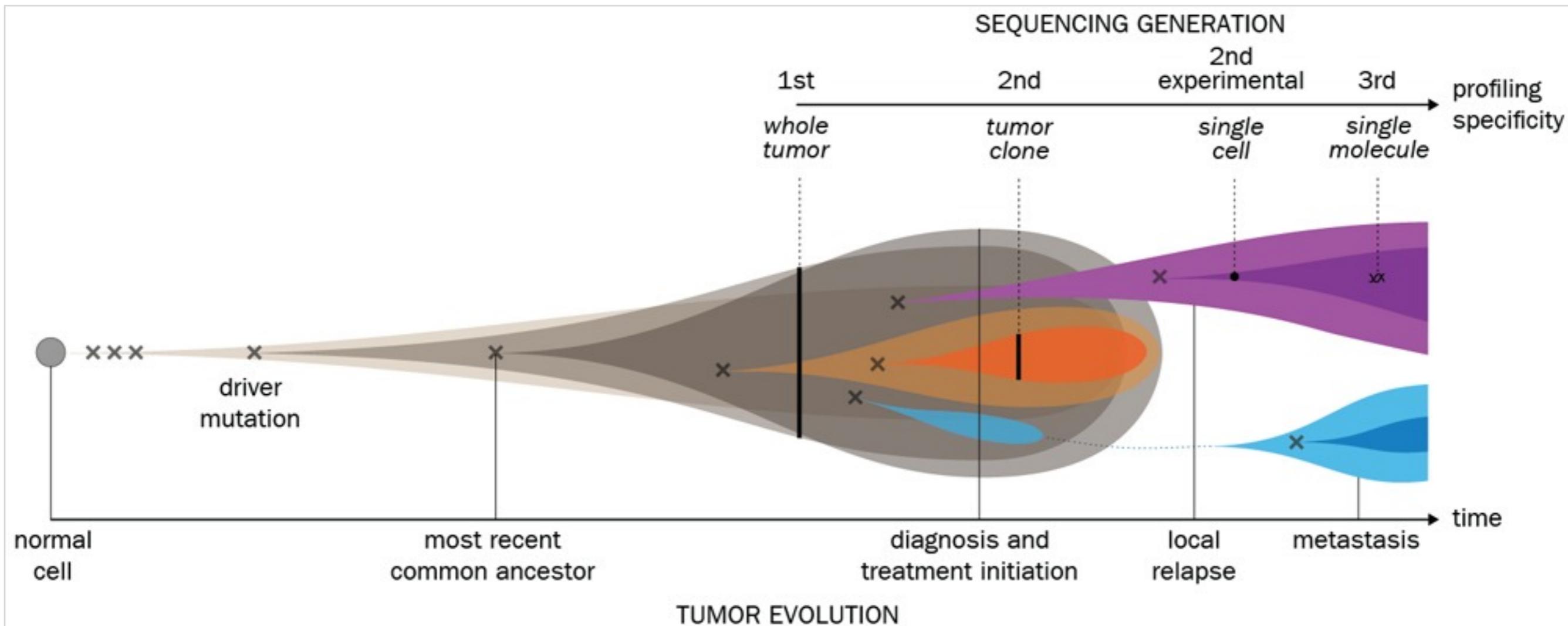
a

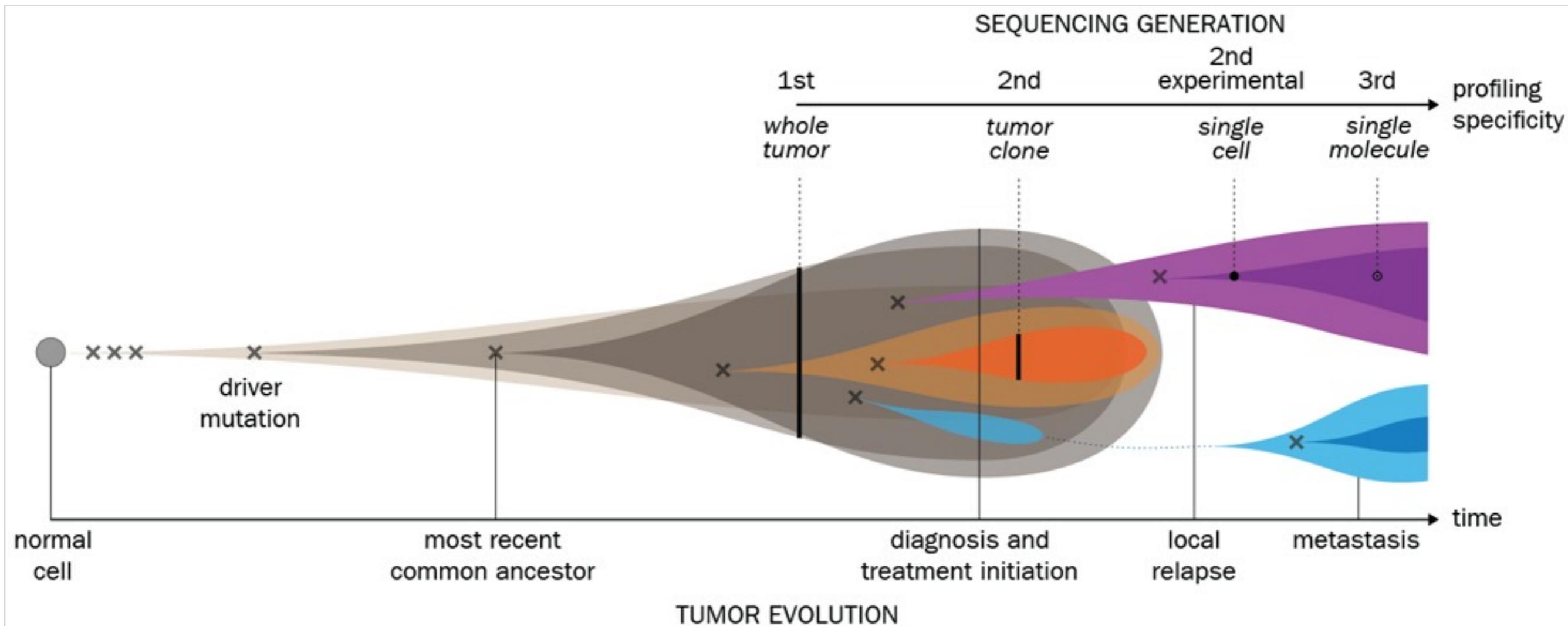


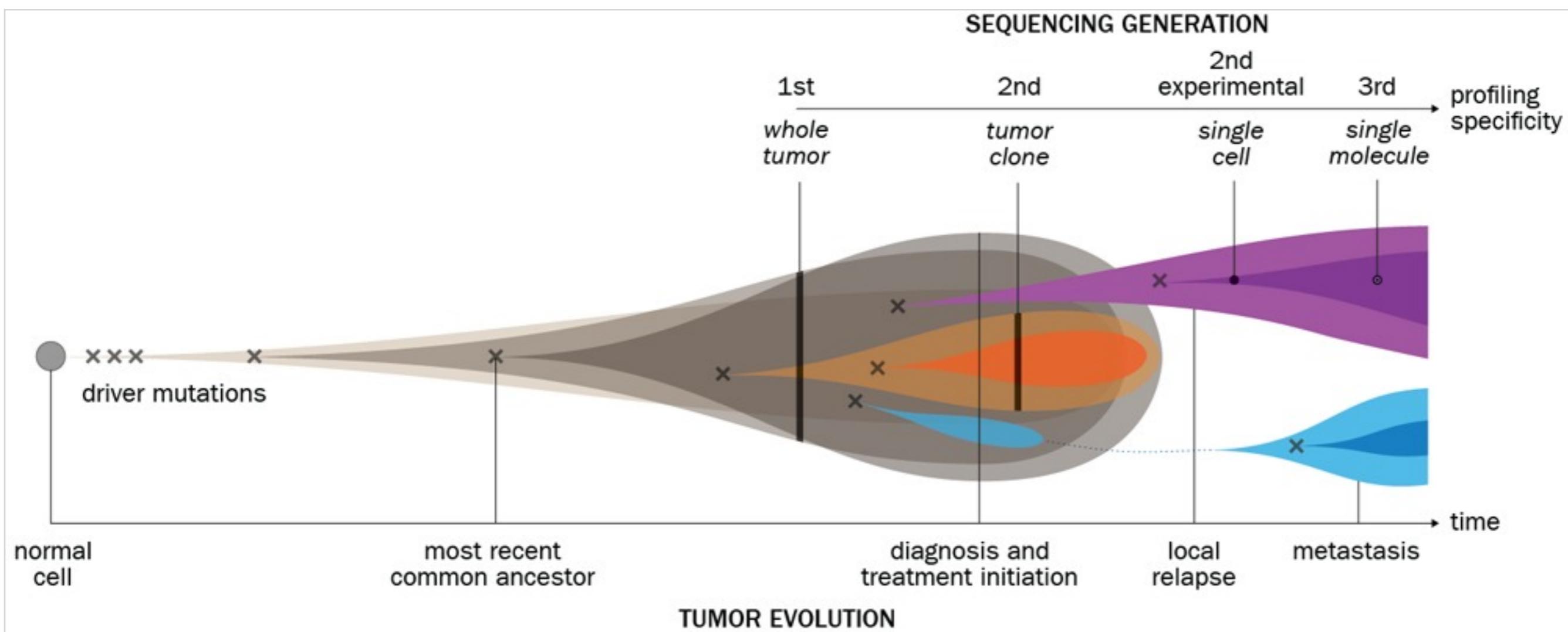


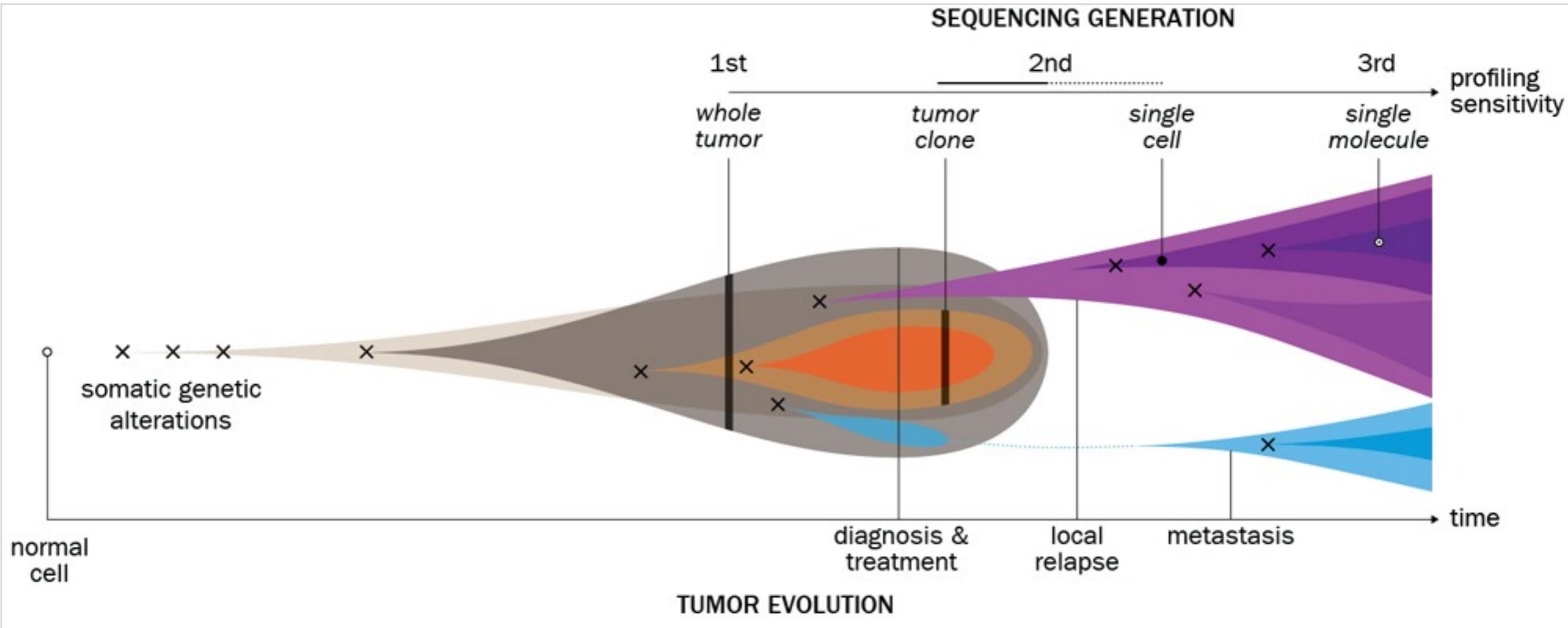


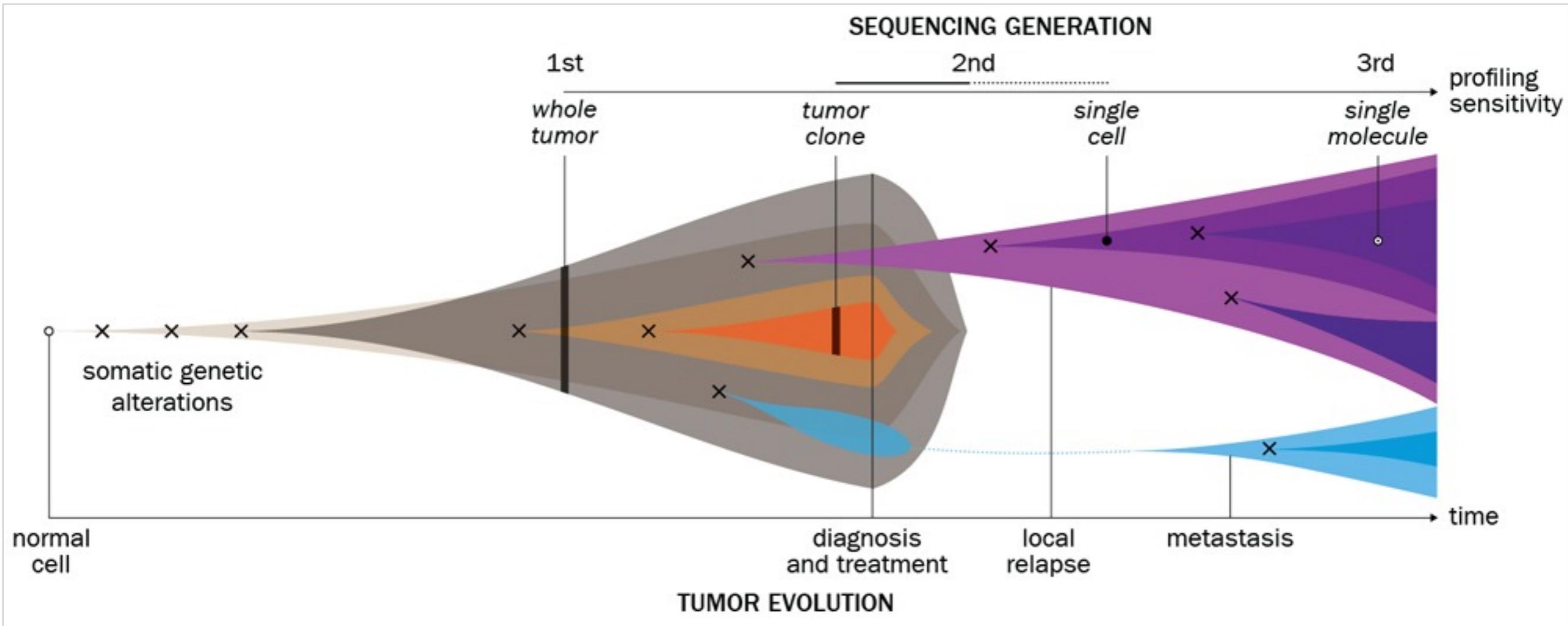


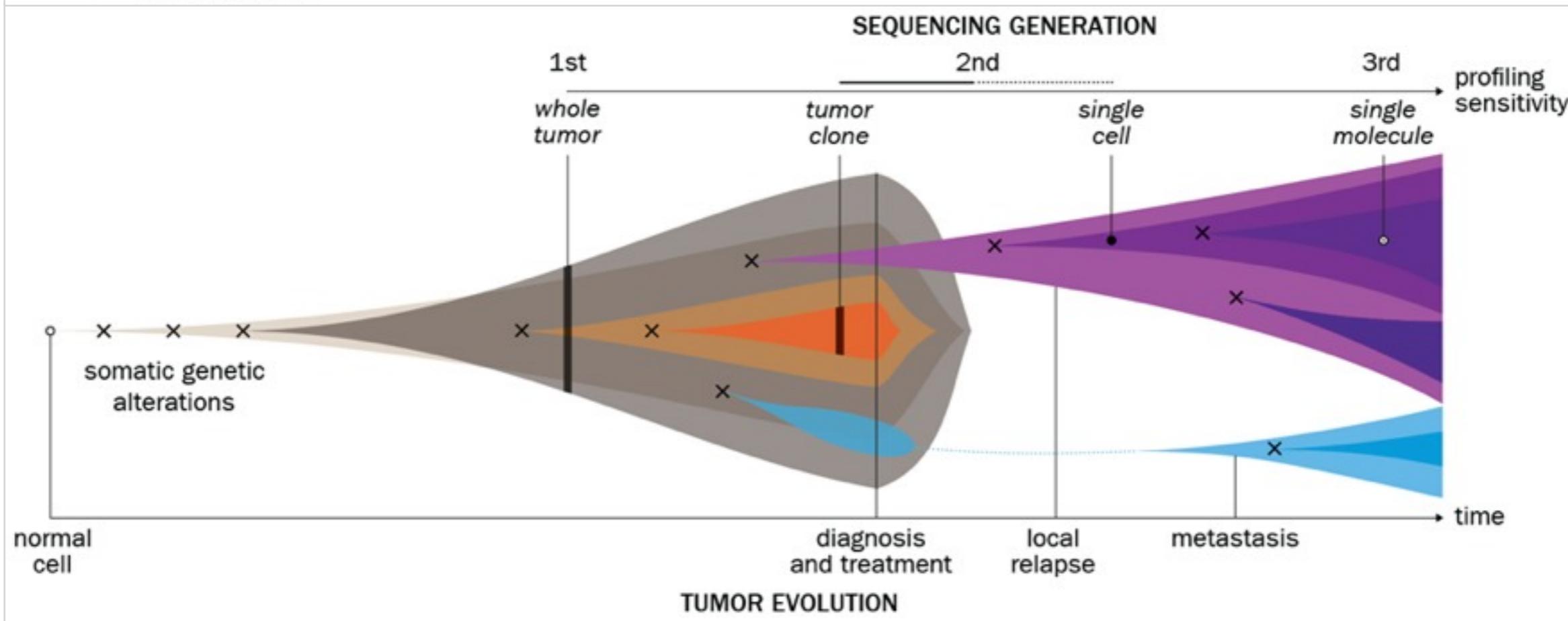
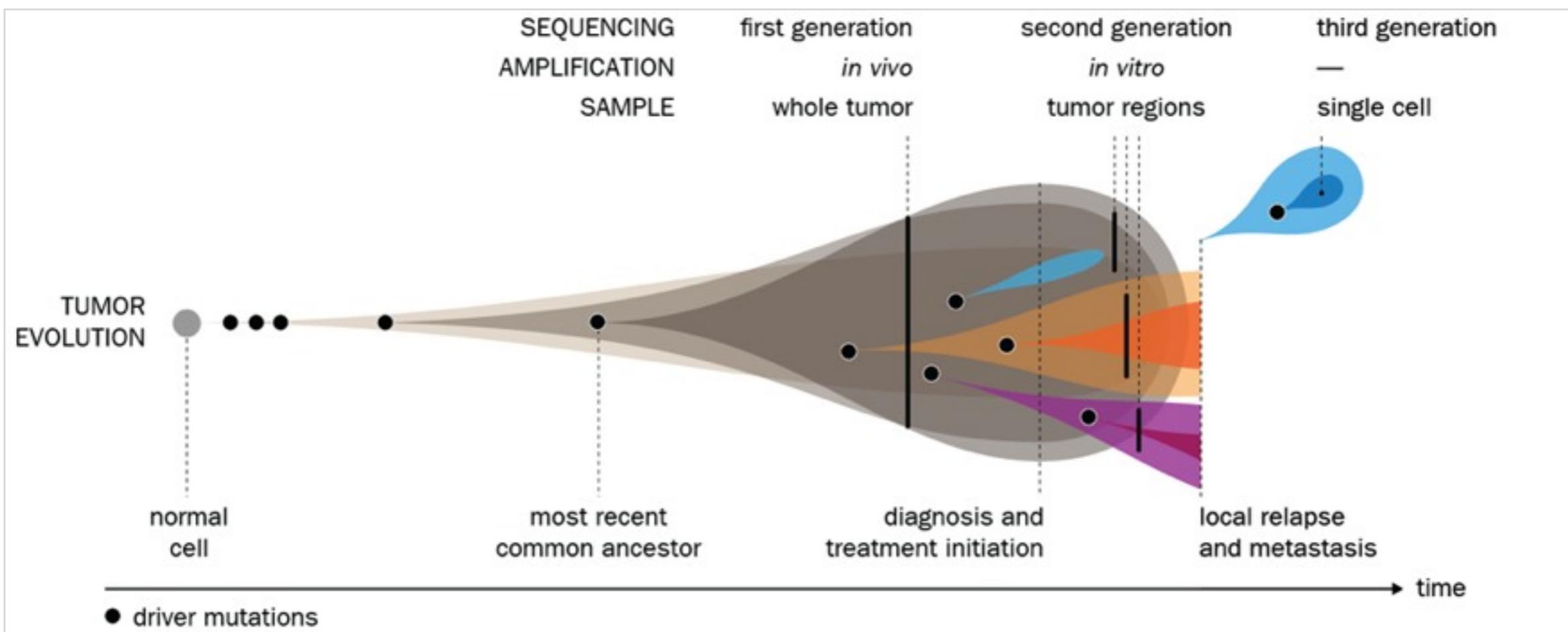






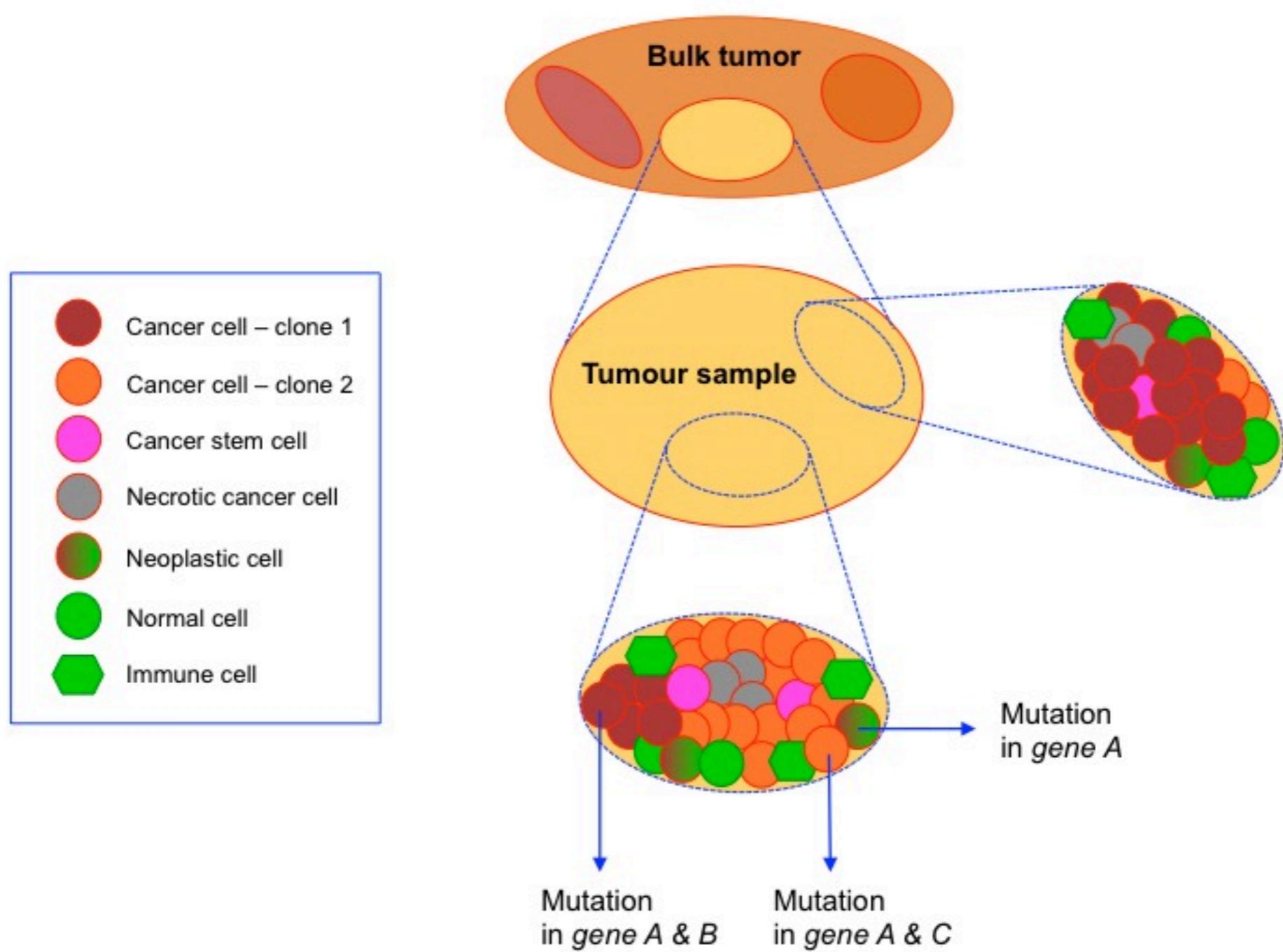


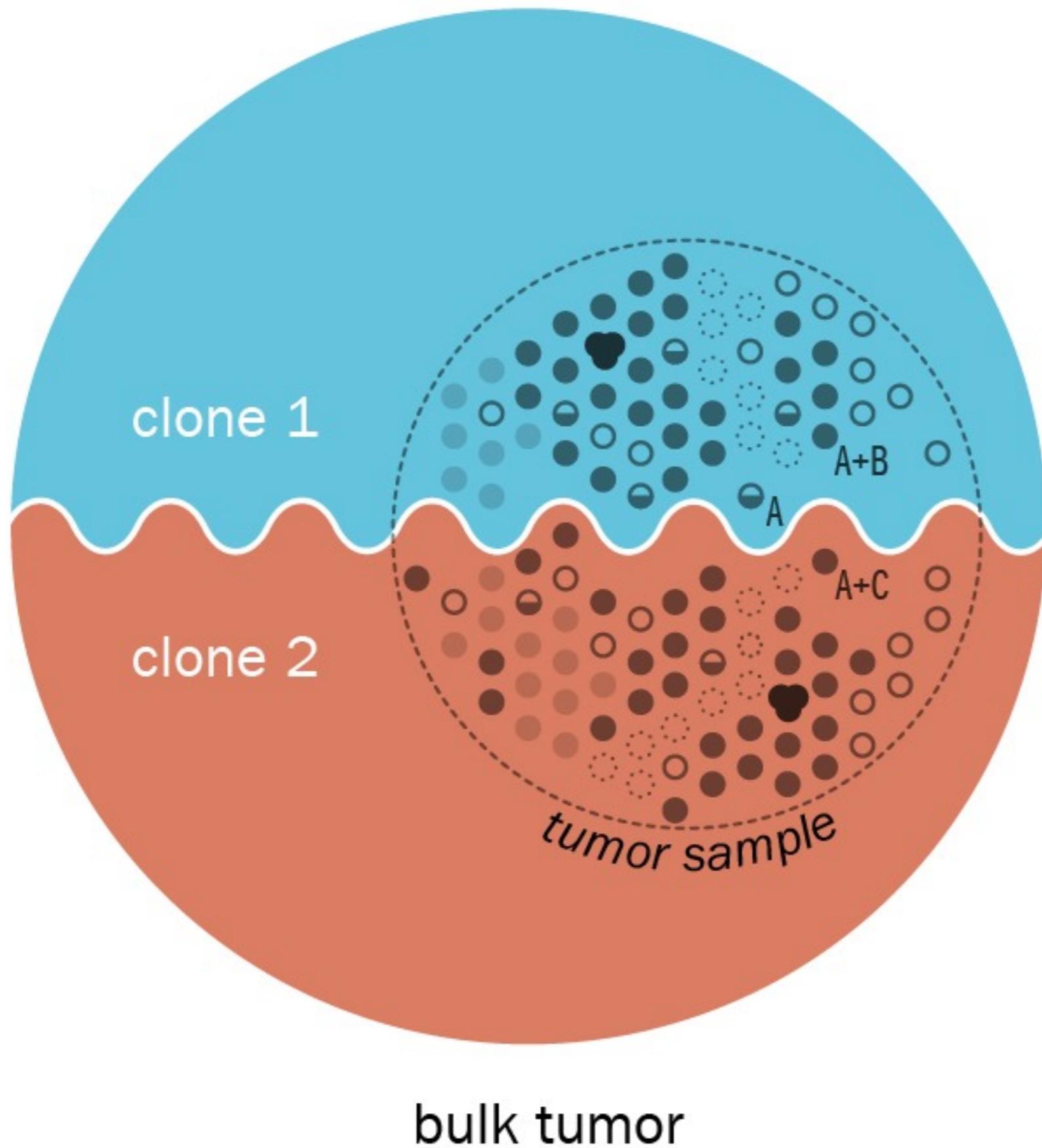




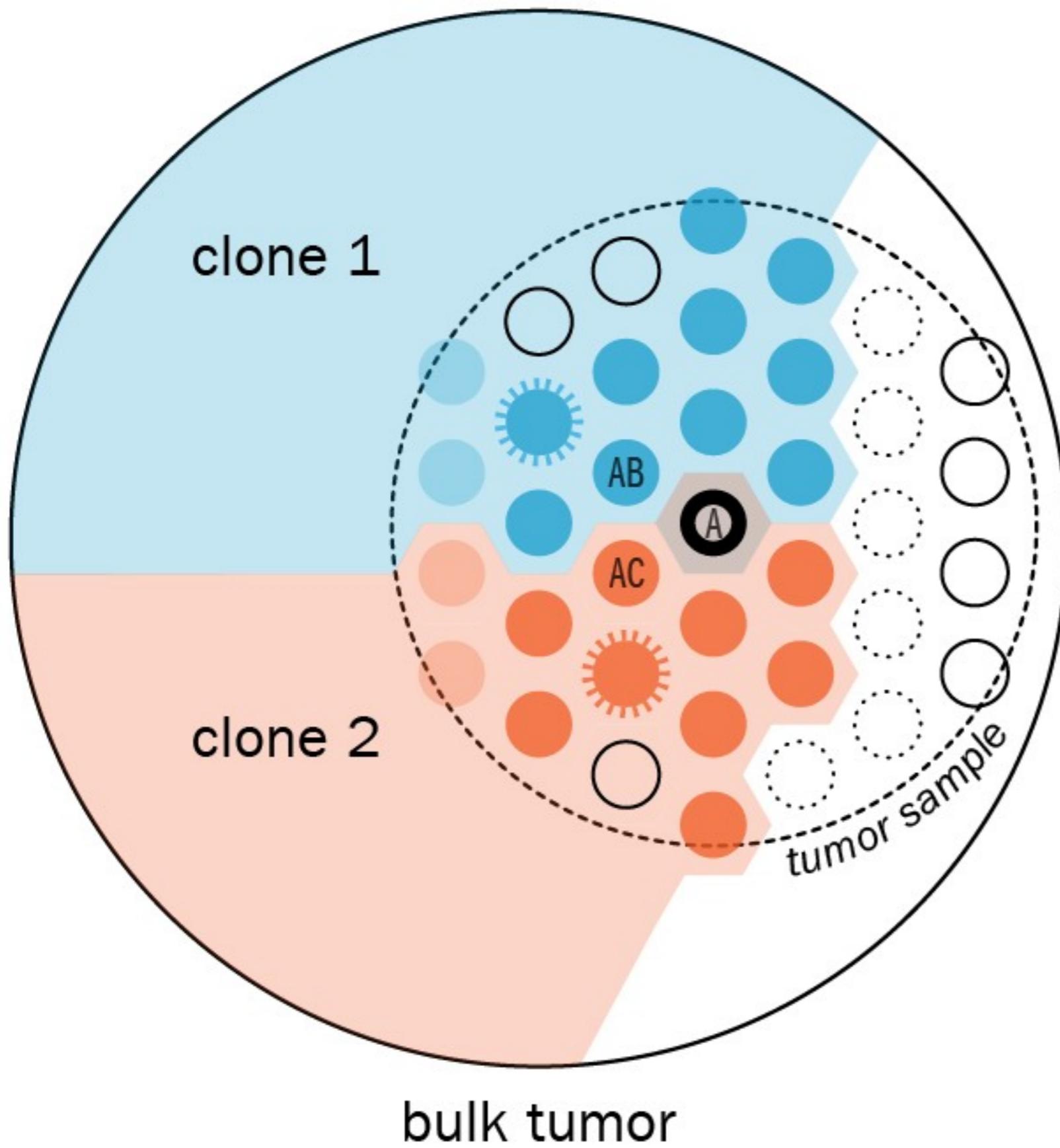
REDESIGN—IN HOUSE

Rational visual vocabulary for cell types.

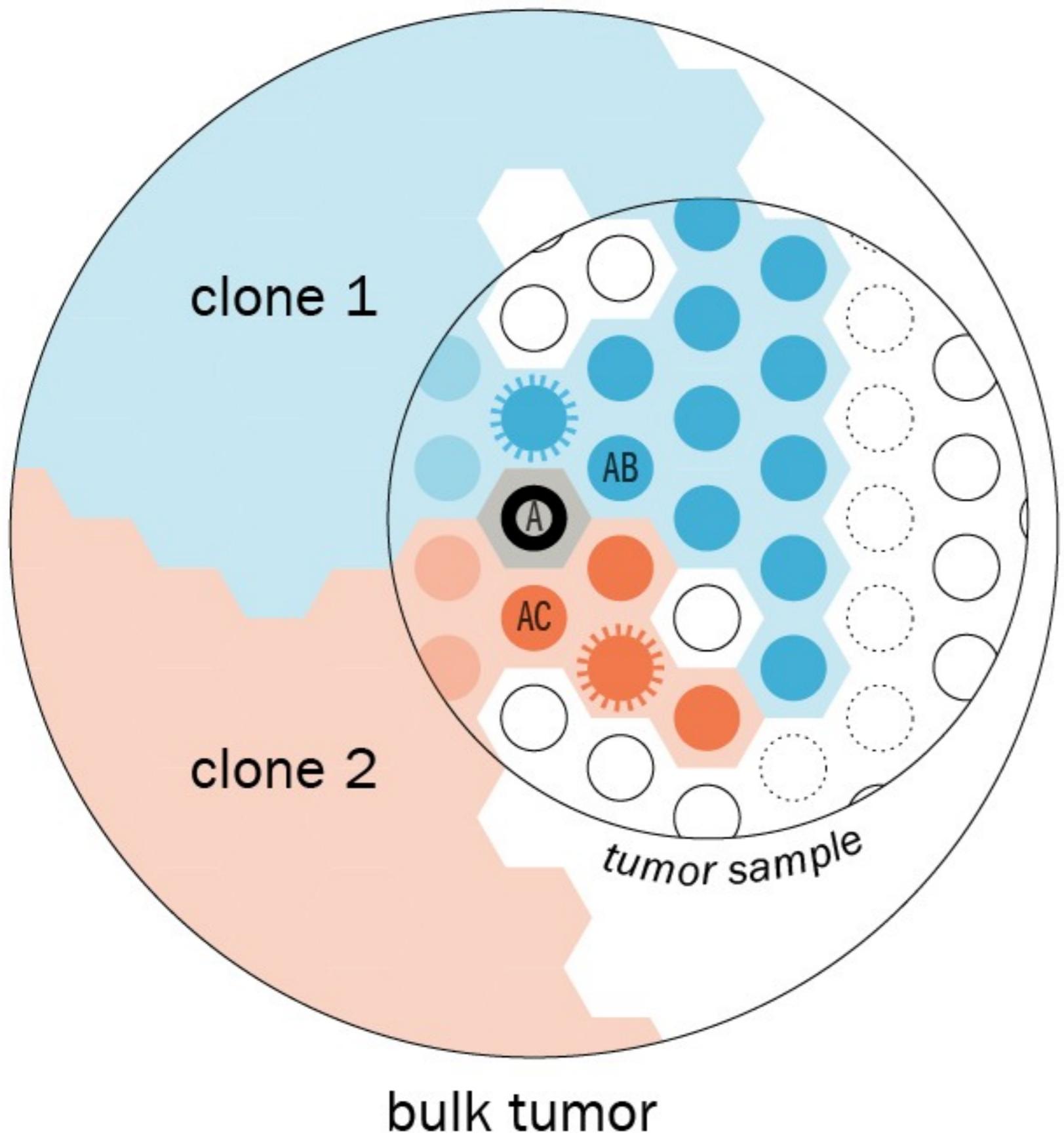




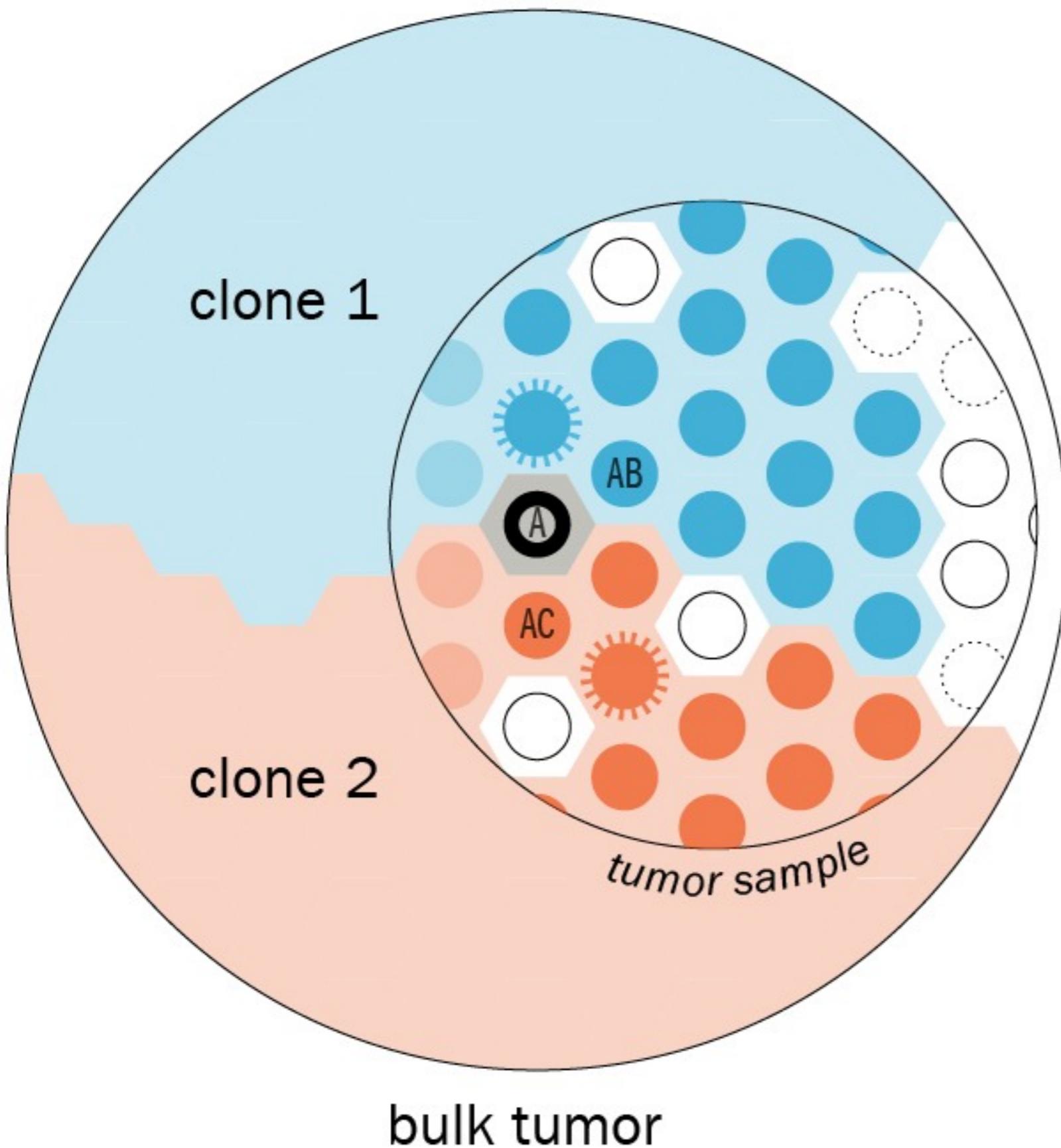
- cancer stem cell
- cancer cell
- necrotic cancer cell
- neoplastic cell
- normal cell
- stromal cell
- A+B mutated genes



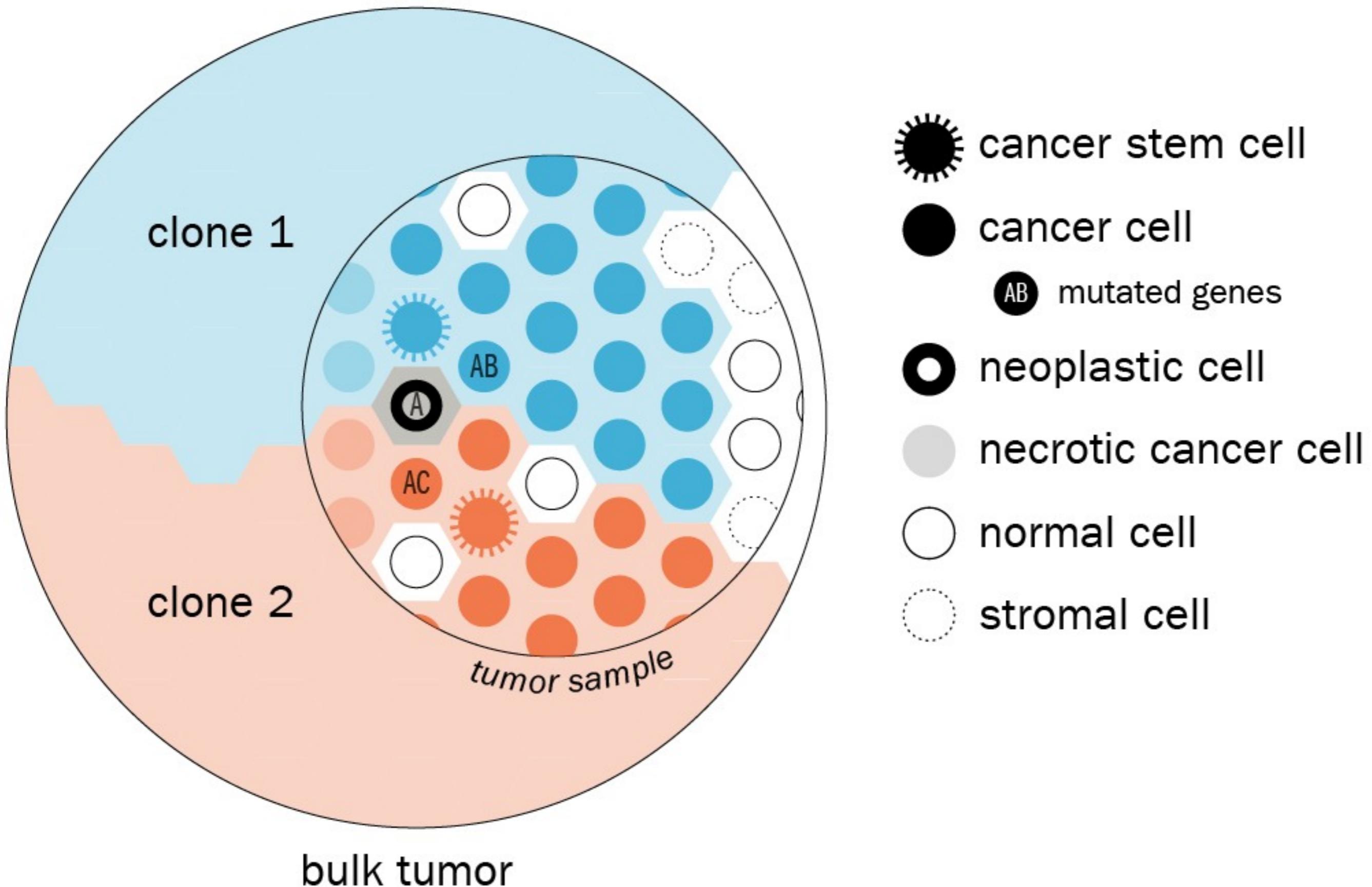
- cancer stem cell
- cancer cell
- neoplastic cell
- necrotic cancer cell
- normal cell
- stromal cell
- mutated genes

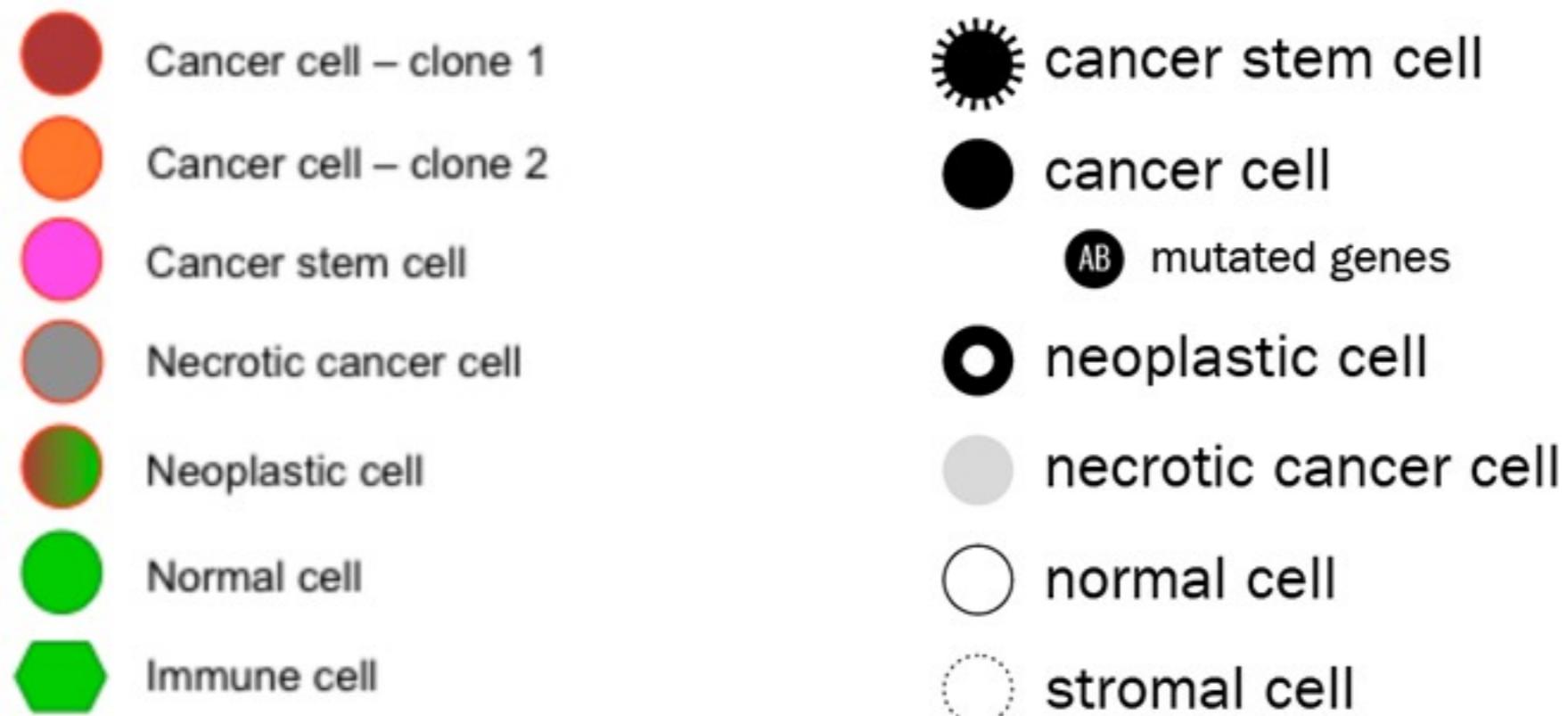


- cancer stem cell**
- cancer cell**
- neoplastic cell**
- necrotic cancer cell**
- normal cell**
- stromal cell**
- mutated genes**



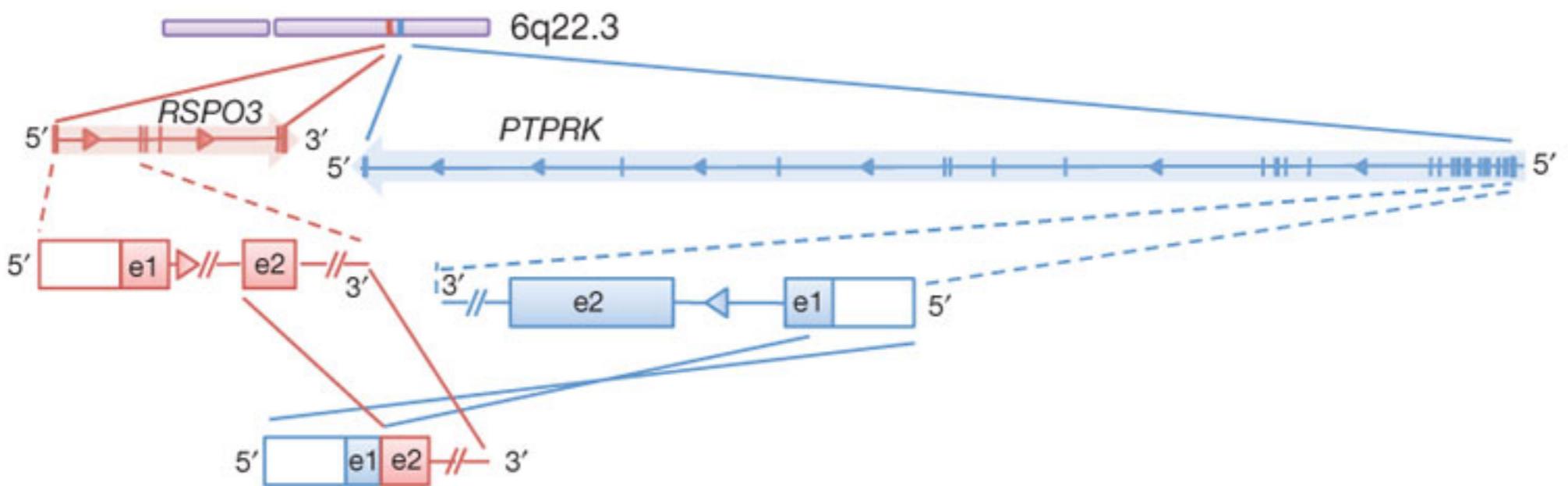
- cancer stem cell
- cancer cell
- neoplastic cell
- necrotic cancer cell
- normal cell
- stromal cell
- AB mutated genes



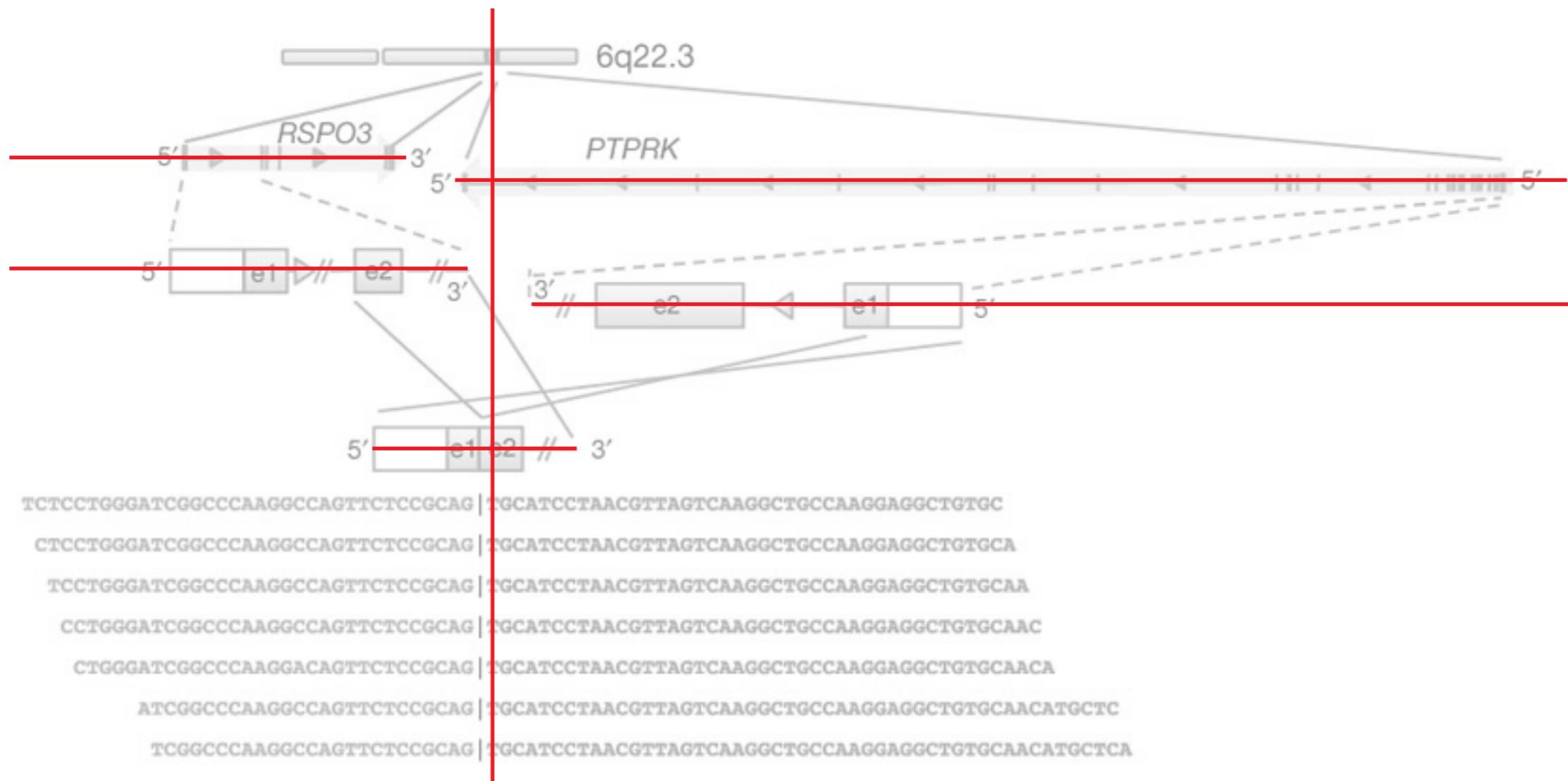


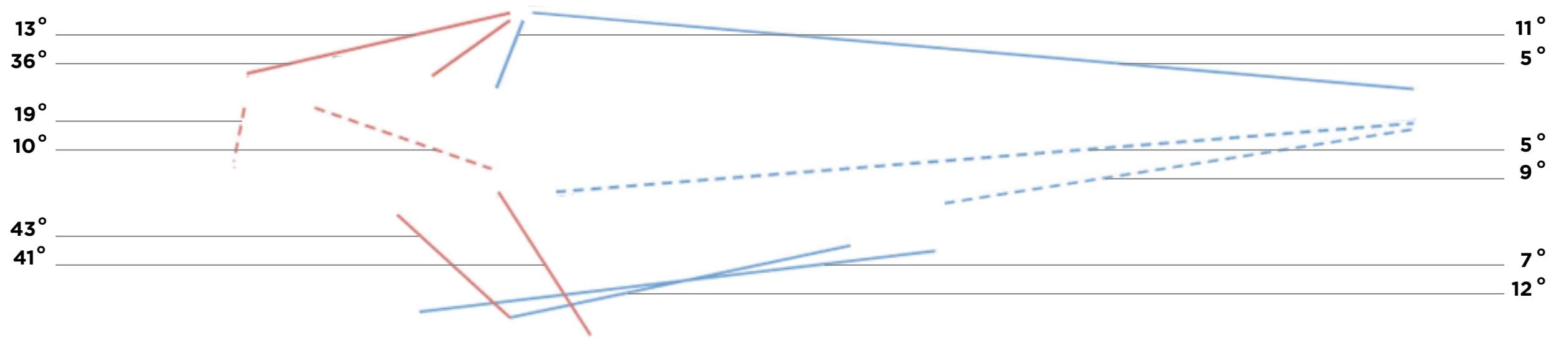
REDESIGN—FROM LITERATURE

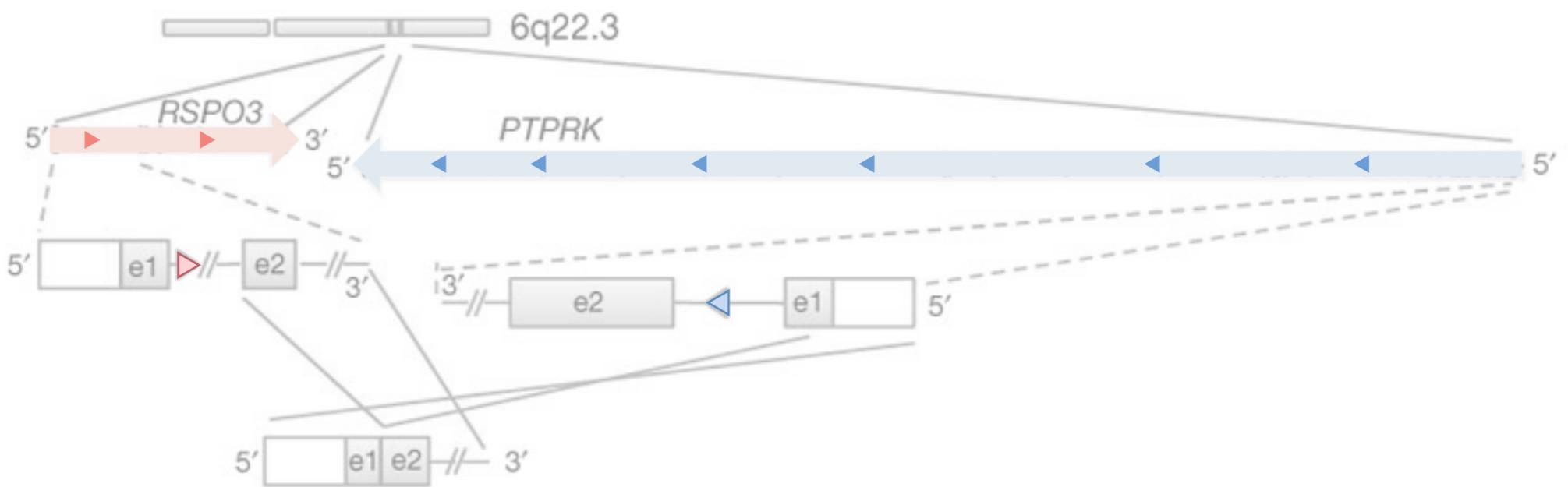
Assume your audience is smart but impatient.



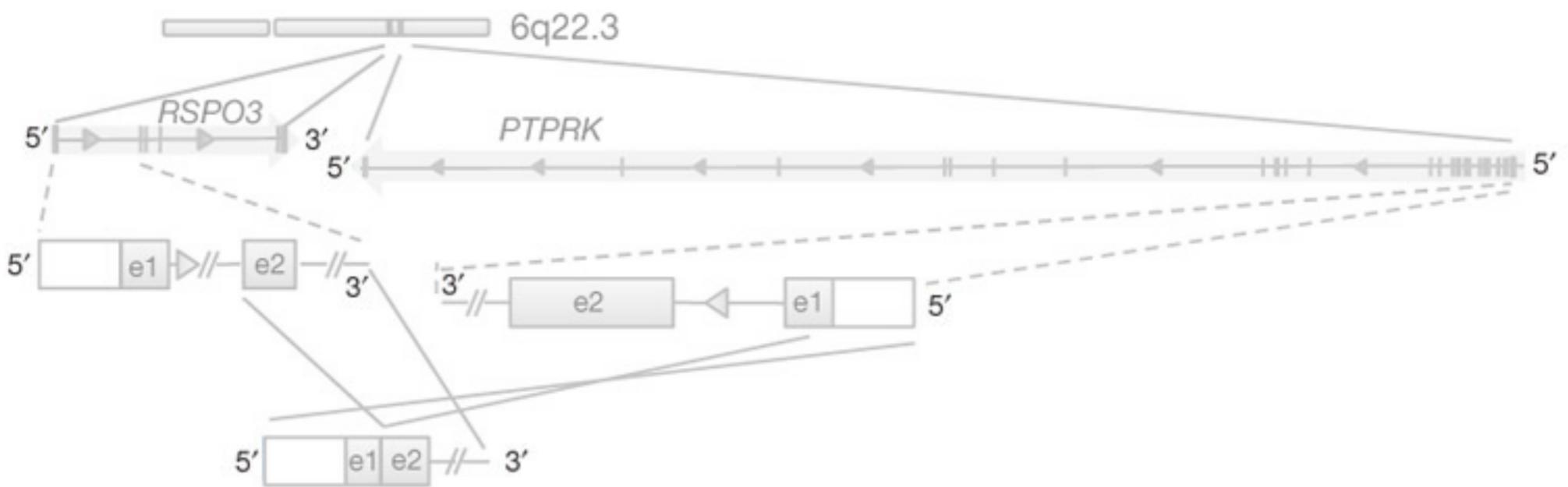
TCTCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGC
 CTCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCA
 TCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAA
 CCTGGGATCGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAAC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



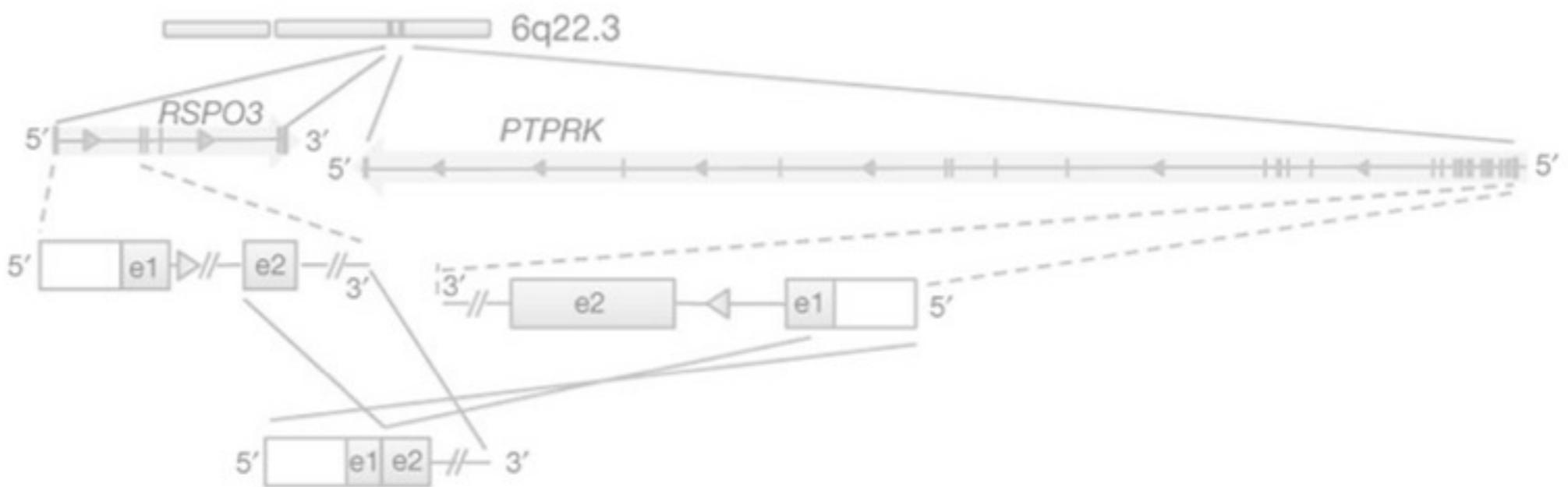




TCTCCTGGGATCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGC
 CTCCTGGGATCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGC
 TCCTGGGATCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAA
 CCTGGGATCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAAC
 CTGGGATCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACA
 ATCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTC
 TCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTCA



TCTCCTGGGATCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGC
 CTCCTGGGATCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCA
 TCCTGGGATCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAA
 CCTGGGATCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAAC
 CTGGGATCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACA
 ATCGGCCAAGGCCAGTTCTCCGAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTC
 TCGGCCAAGGCCAGTTCTCCGAG | 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

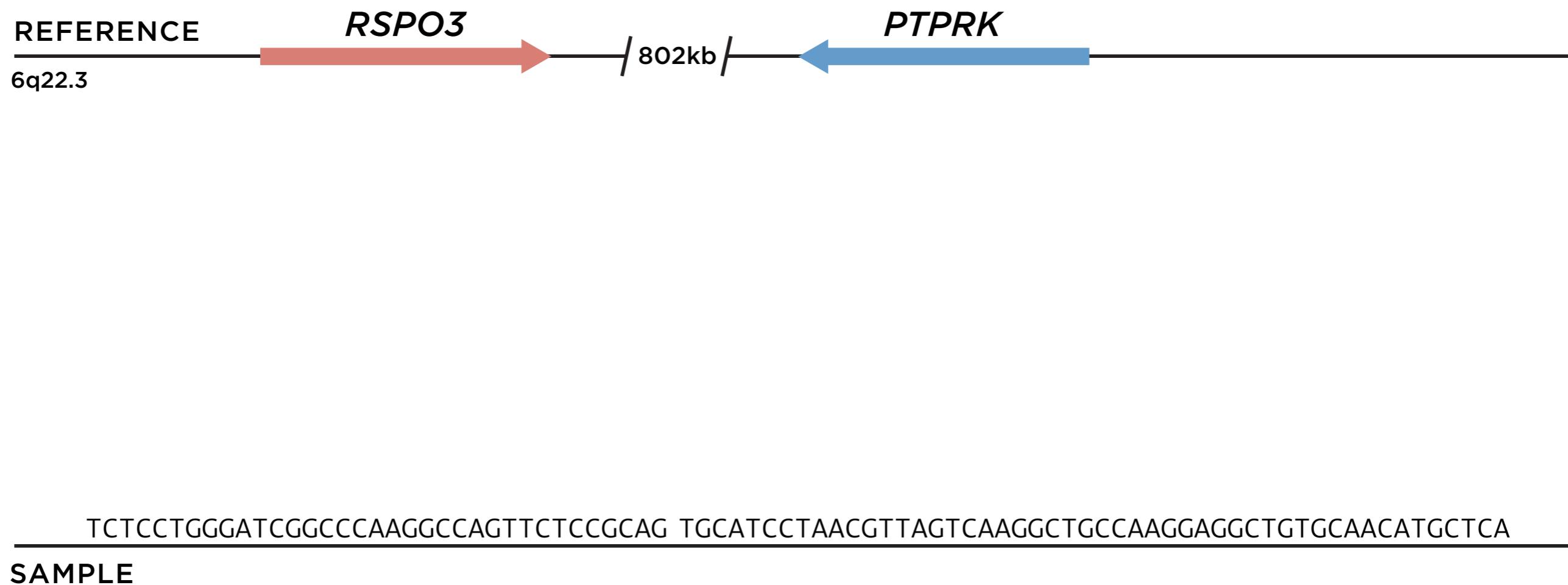
what is not important, or peripheral?
gene size
gene location
gene model (learn to let go)

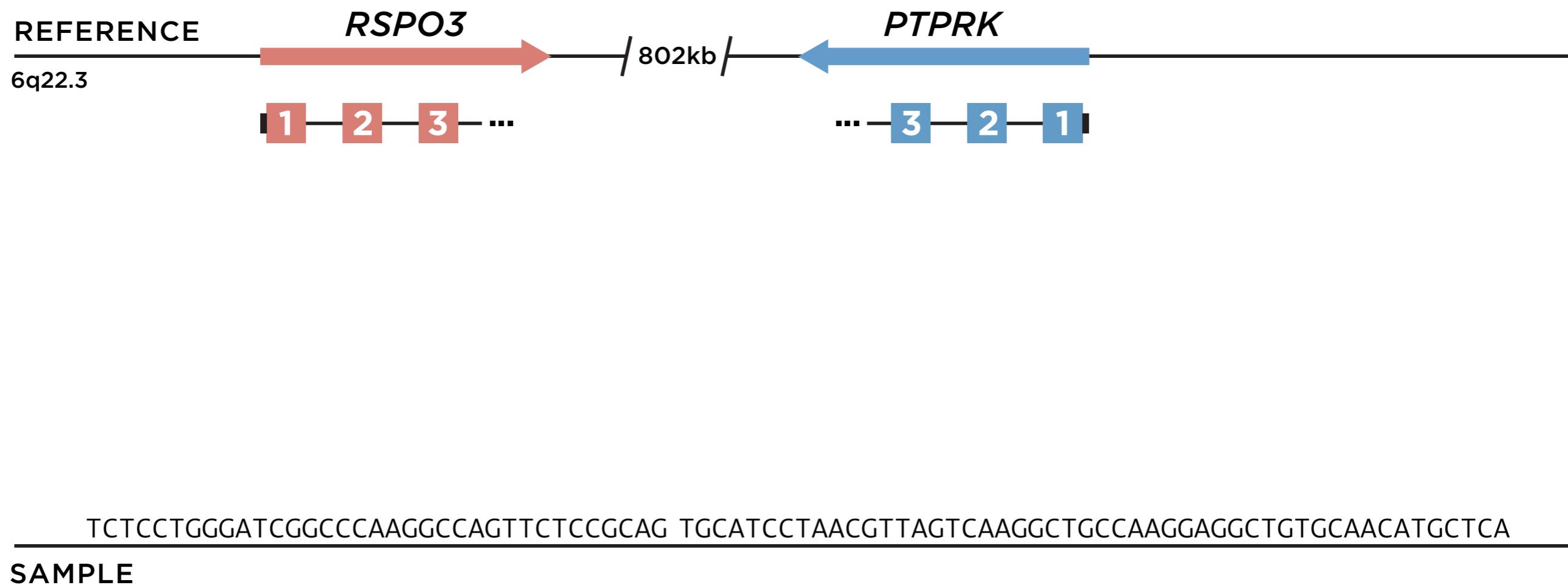
REFERENCE

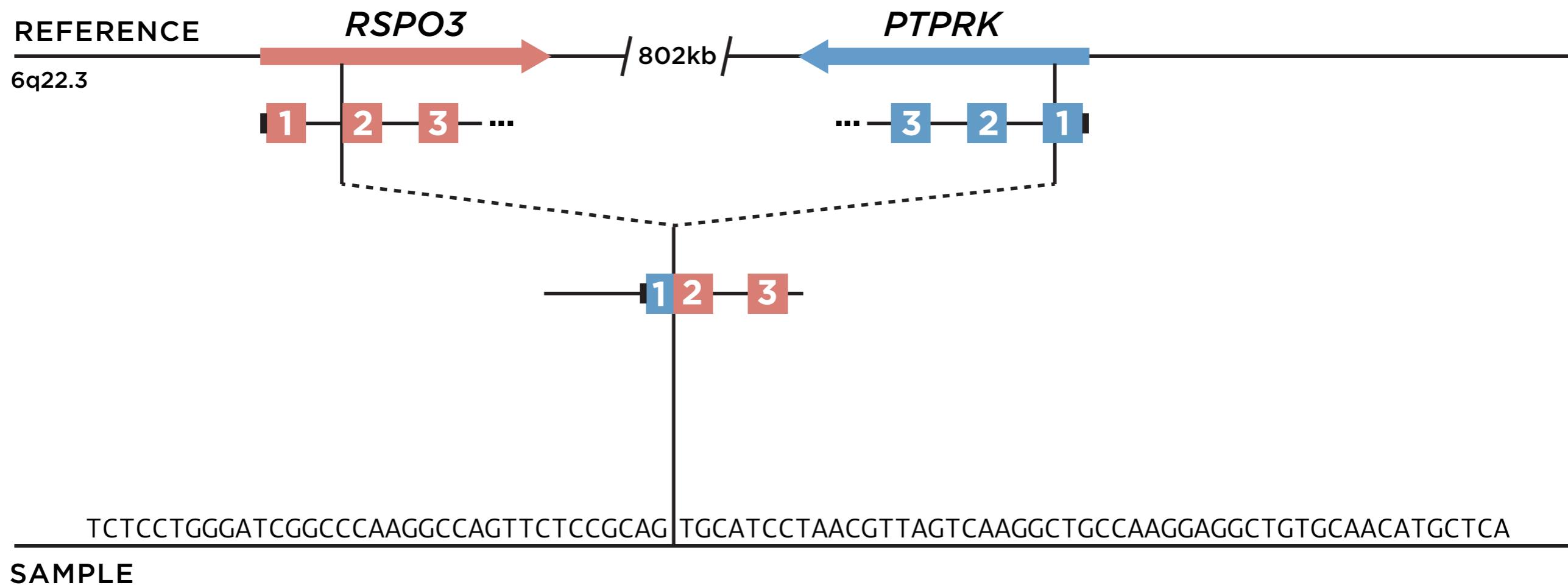
6q22.3

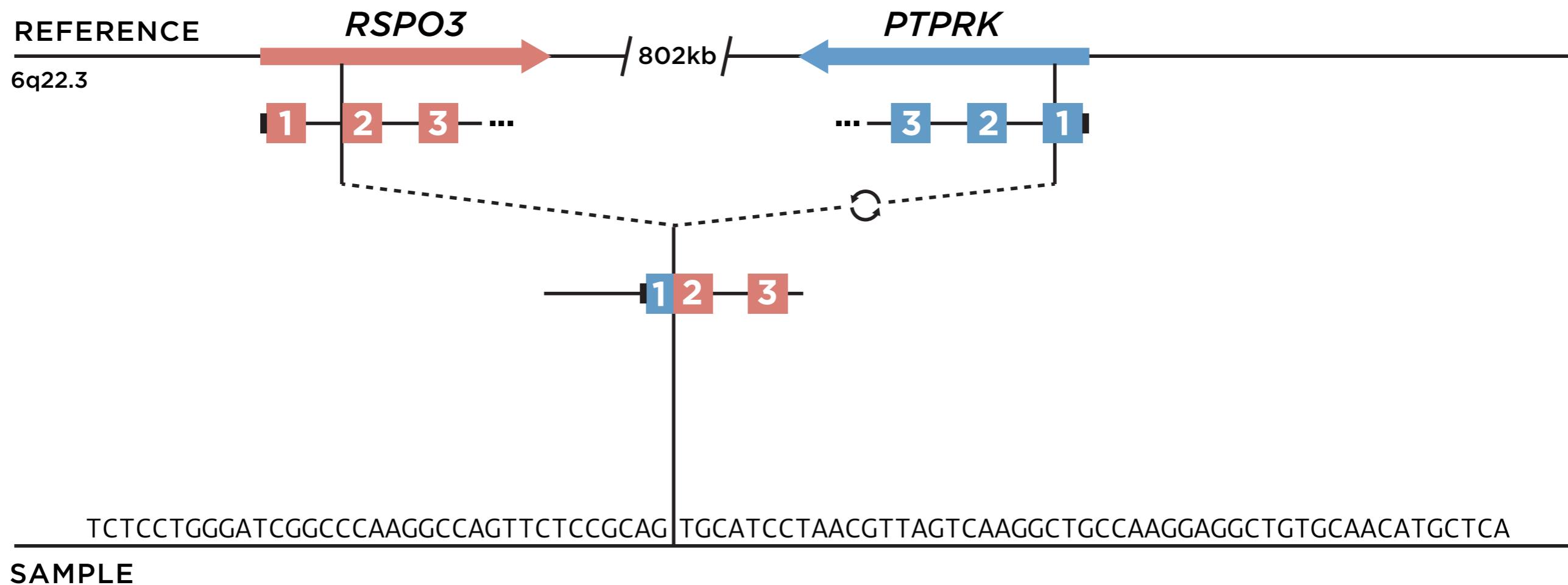
TCTCCTGGGATCGGCCAAGGCCAGTTCTCCGCAG TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTCA

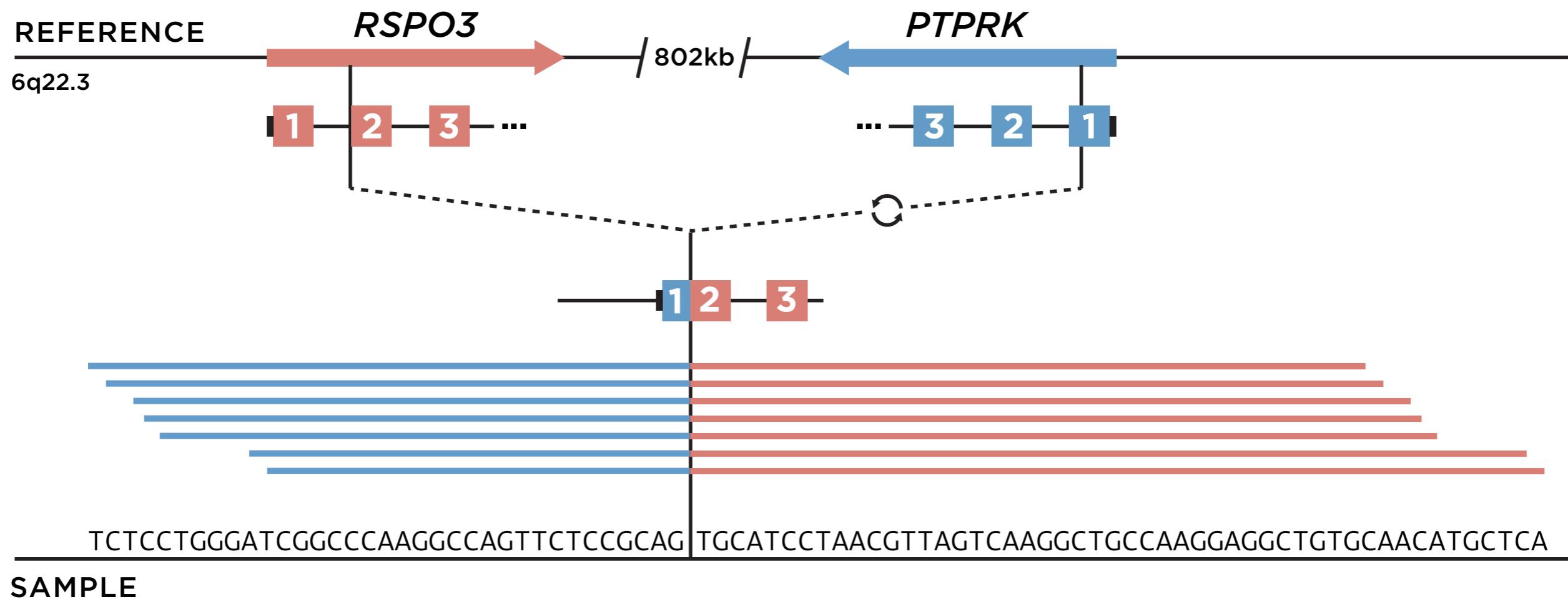
SAMPLE

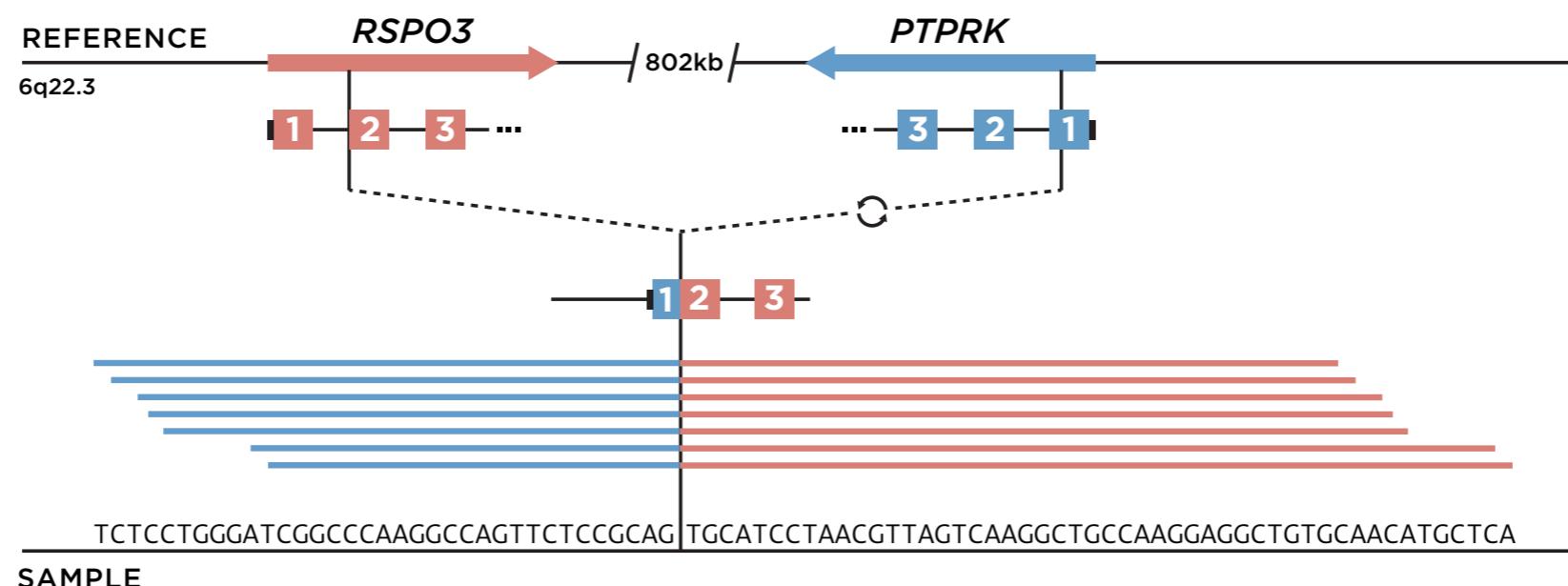
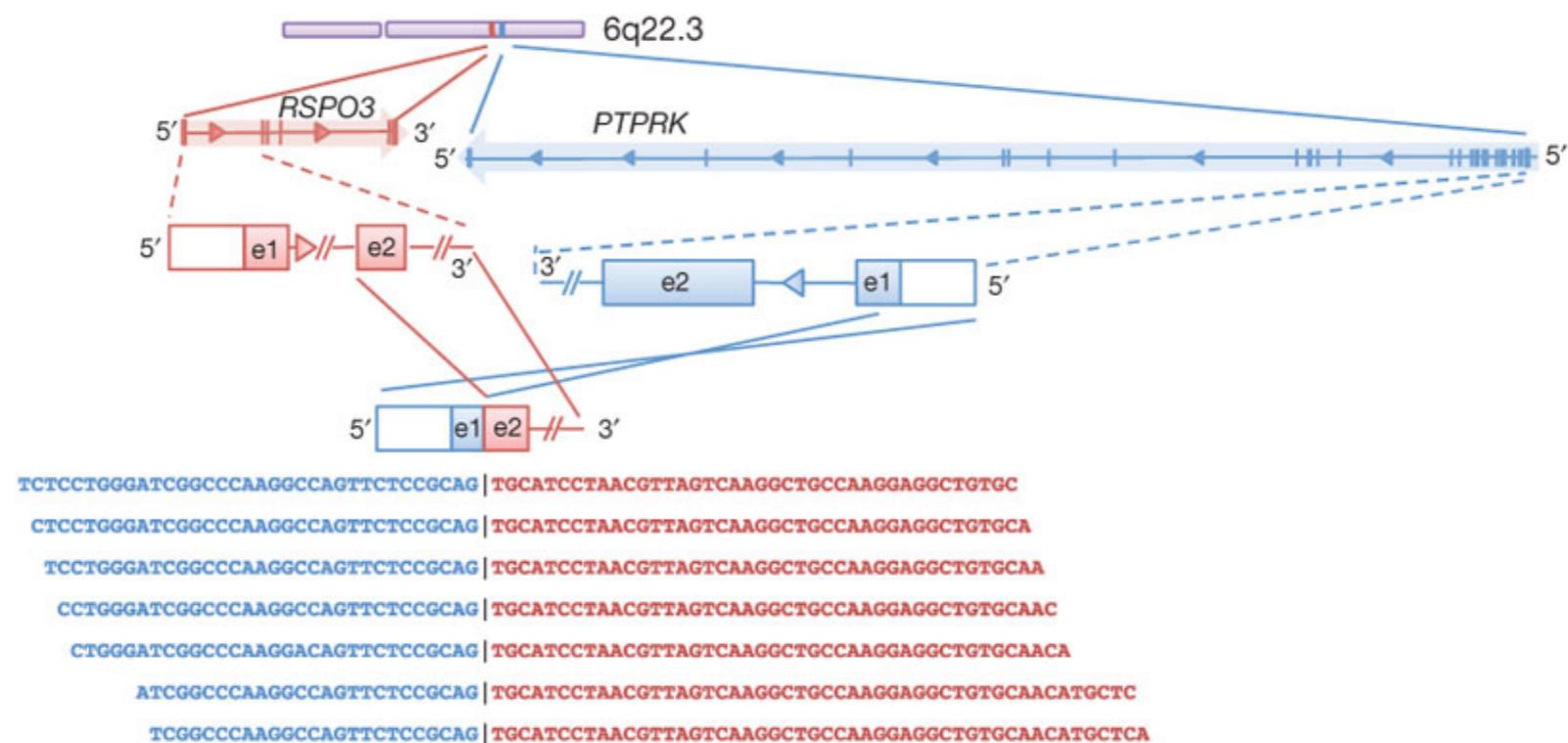


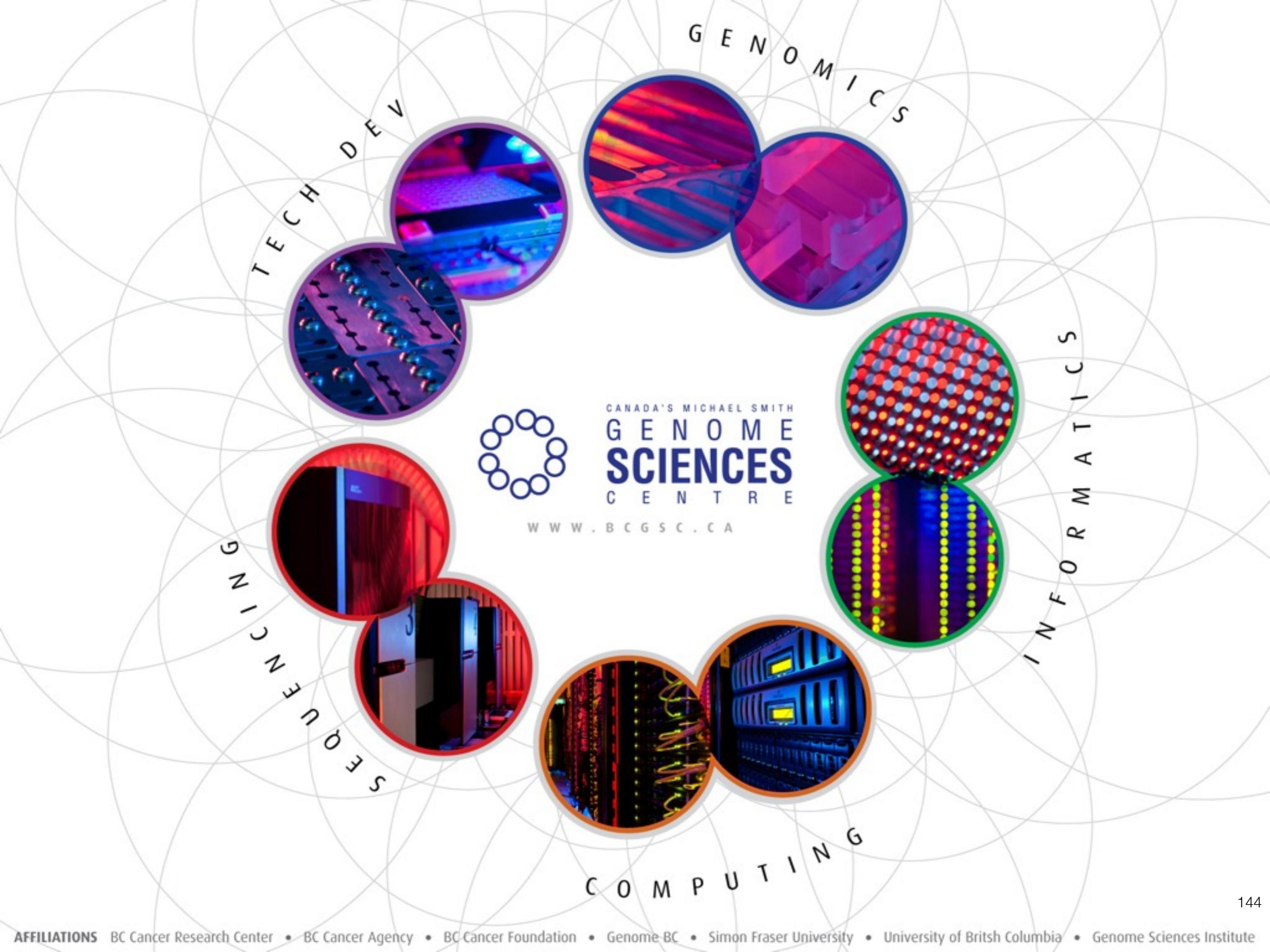












CANADA'S MICHAEL SMITH
GENOME
SCIENCES
CENTRE

WWW.BCGSC.CA