THE UNIVERSITY OF SYDNEY

CANADA'S MICHAEL SMITH GENOME SCIENCES CENTRE

ESSENTIALS OF DATA VISUALIZATION

THINKING ABOUT DRAWING DATA + COMMUNICATING SCIENCE

COLOR

how to characterize and use it

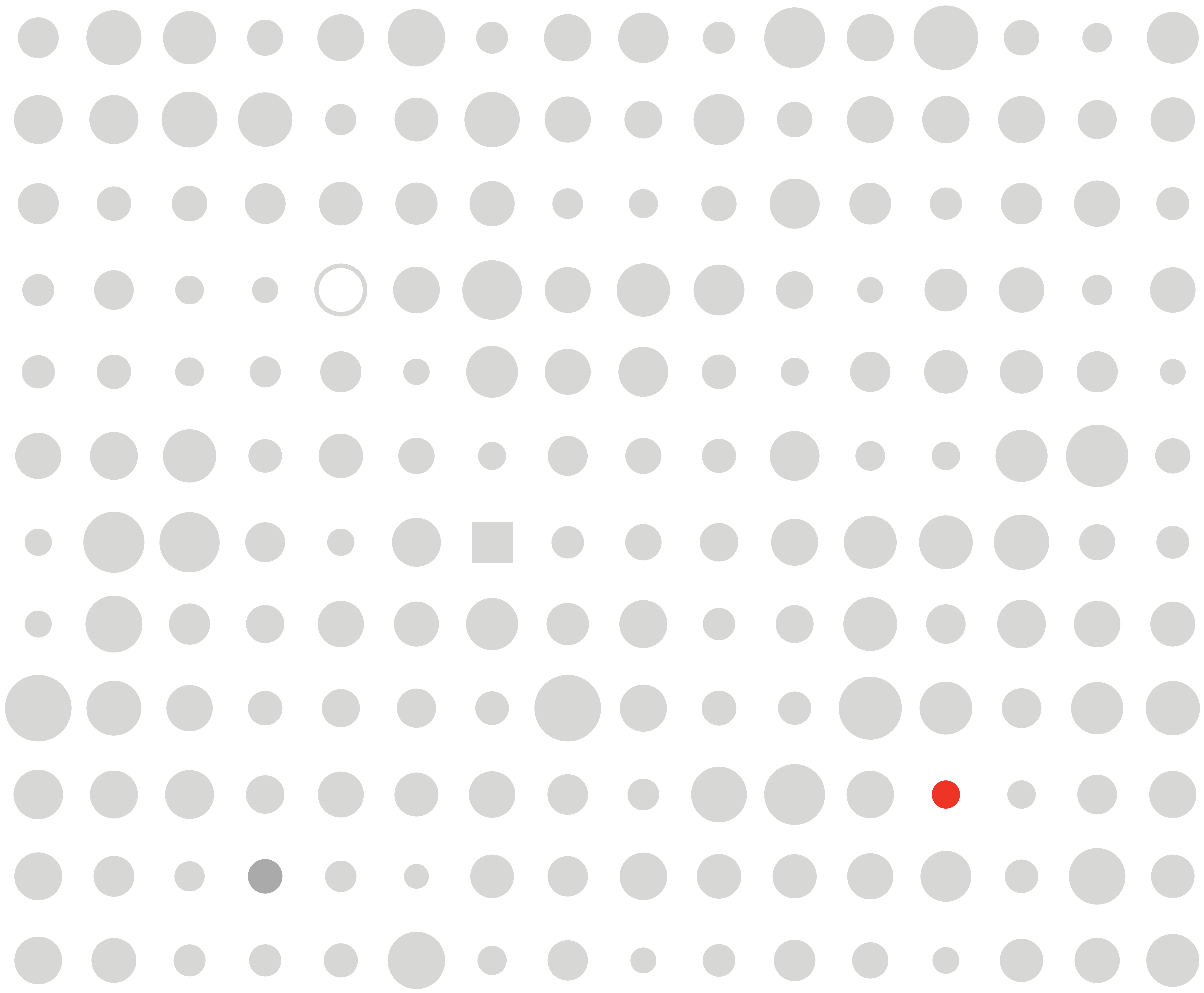
Color is one of the most exciting ways in which you can completely screw over your visualization. What can start off as a great diagram can be absolutely ruined by a lack of color judgment.

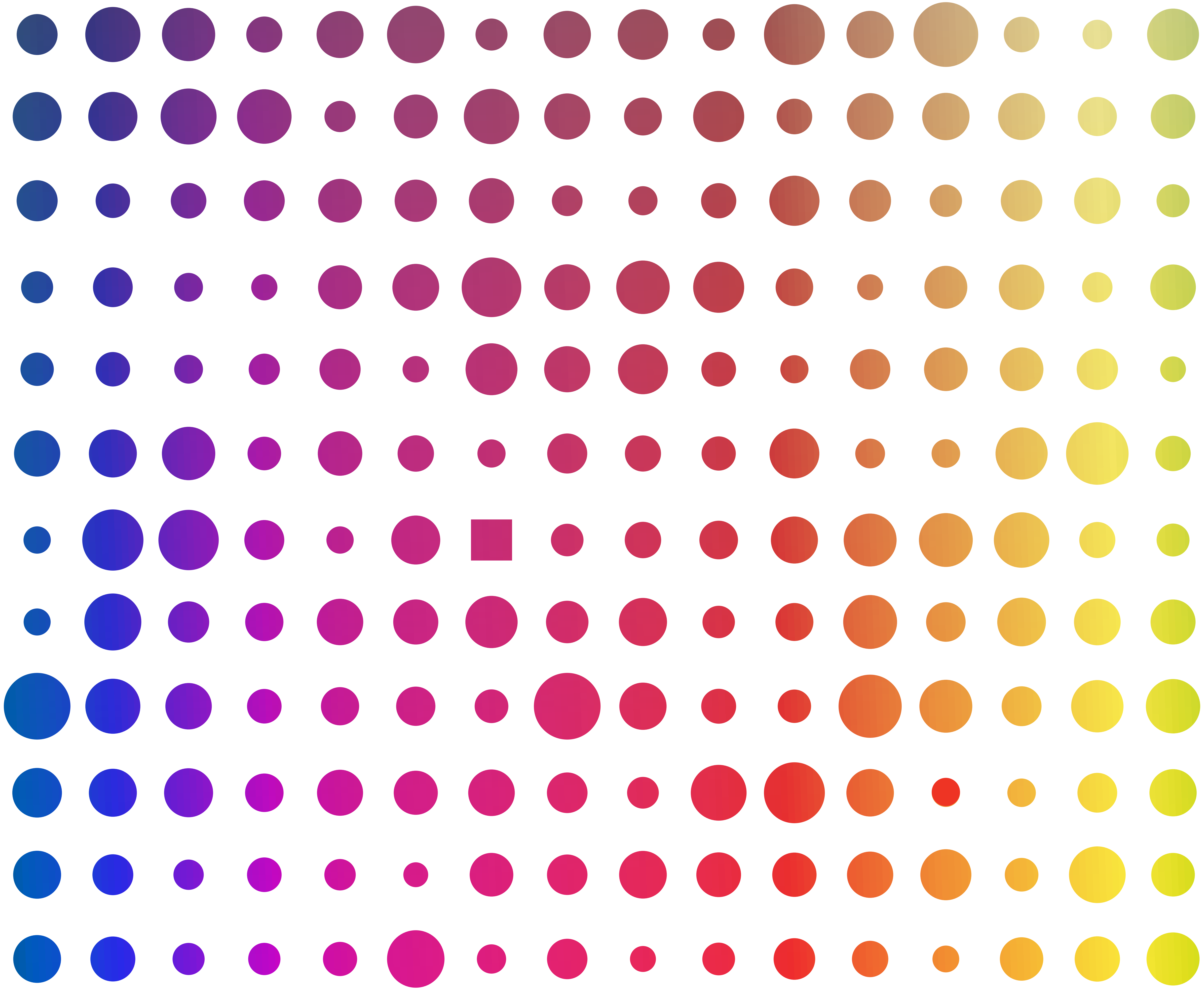
That being said, once you've selected an encoding and shapes, as we've already talked about, color is a terrific way to add layers of information. Color forms a distinct visual channel and we've already seen it act as a powerful force of grouping. This grouping can come not just from shapes of the same color, but shapes of a similar color.

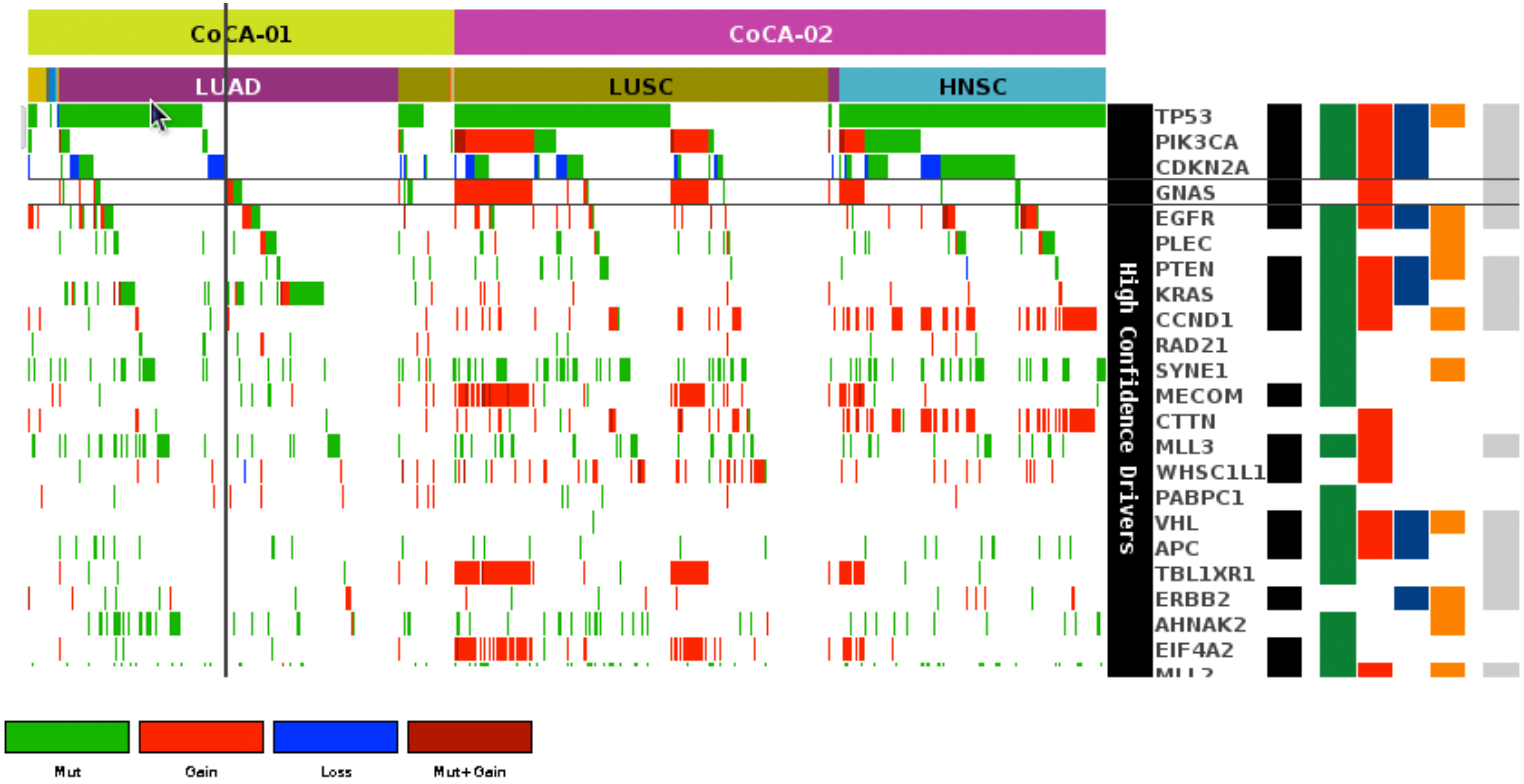
This should make you ask— how can we characterize color similarity? Can we quantify it? Can we say that one pair of colors is more similar than another pair?

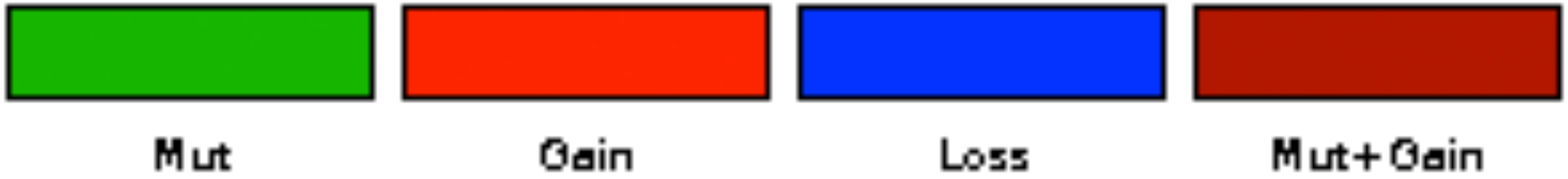
Luckily yes. And much of how color can be used effectively is based on this quantitative and perceptual characterization of it.

This is a very exciting prospect, since other senses don't have this mathematical model. Can I say that one pair of smells is twice as different as another pair? Nope.

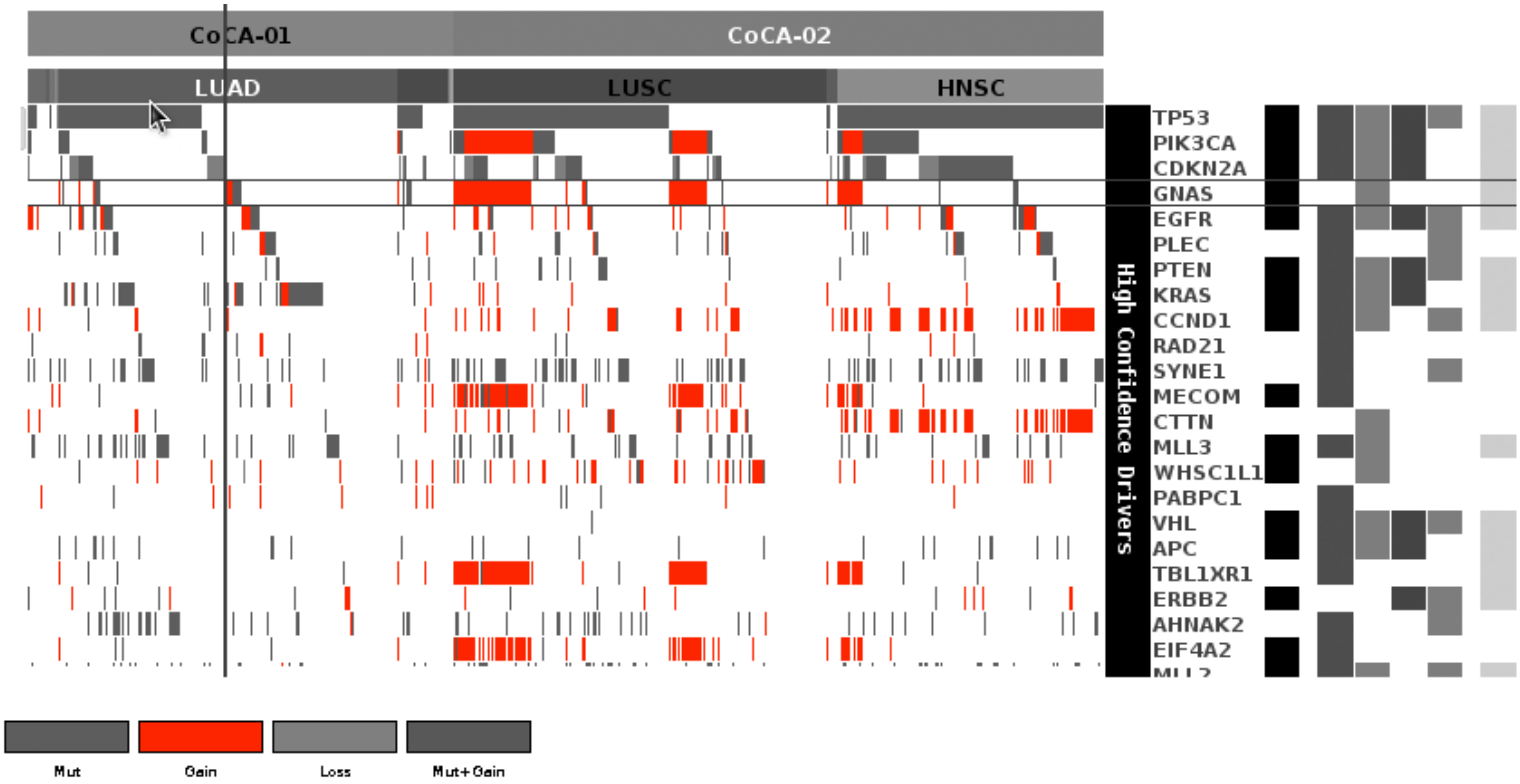




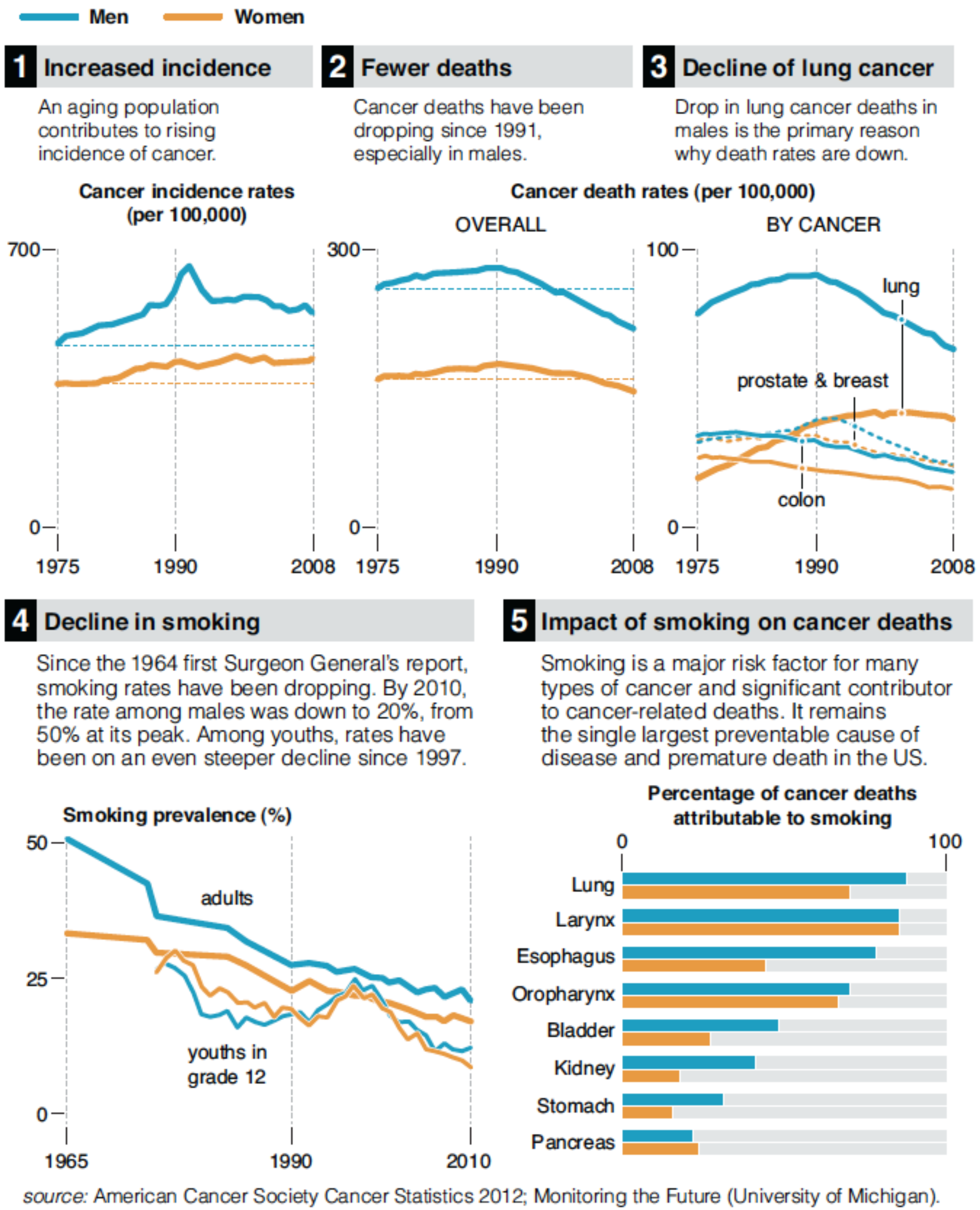




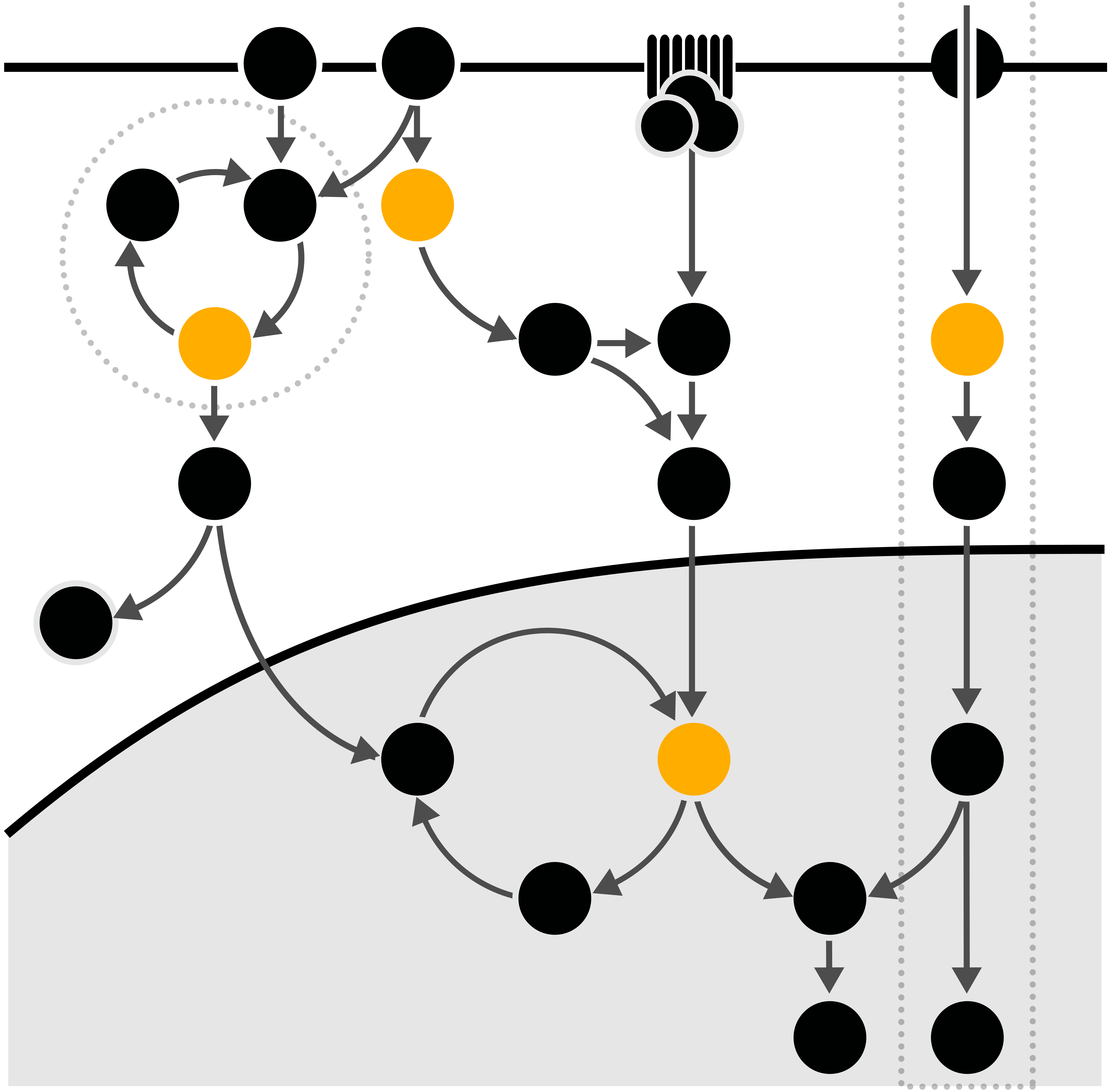
http://www.gitools.org/img/pancancer12/pancancer-drivers-gitoolsweb.screenshot.png



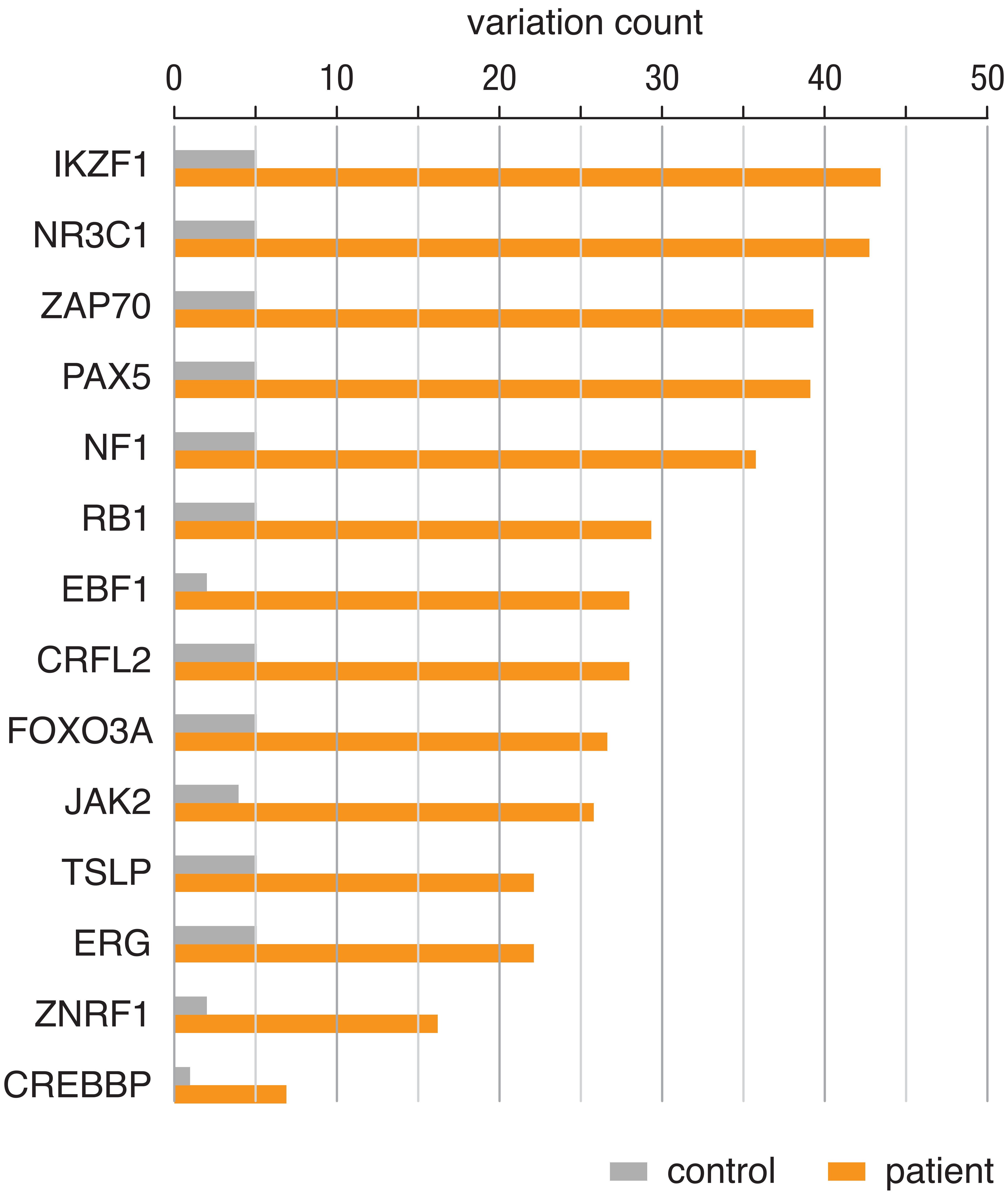
CoCA-01
CoCA-02
LUAD
LUSC
HNSC
High Confidence Drivers
TP53
PIK3CA
CDKN2A
GNAS
EGFR
PLEC
PTEN
KRAS
CCND1
RAD21
SYNE1
MECOM
CTTN
MLL3
WHSC1L1
PABPC1
VHL
APC
TBL1XR1
ERBB2
AHNAK2
EIF4A2
Mut
Gain
Loss
Mut+Gain

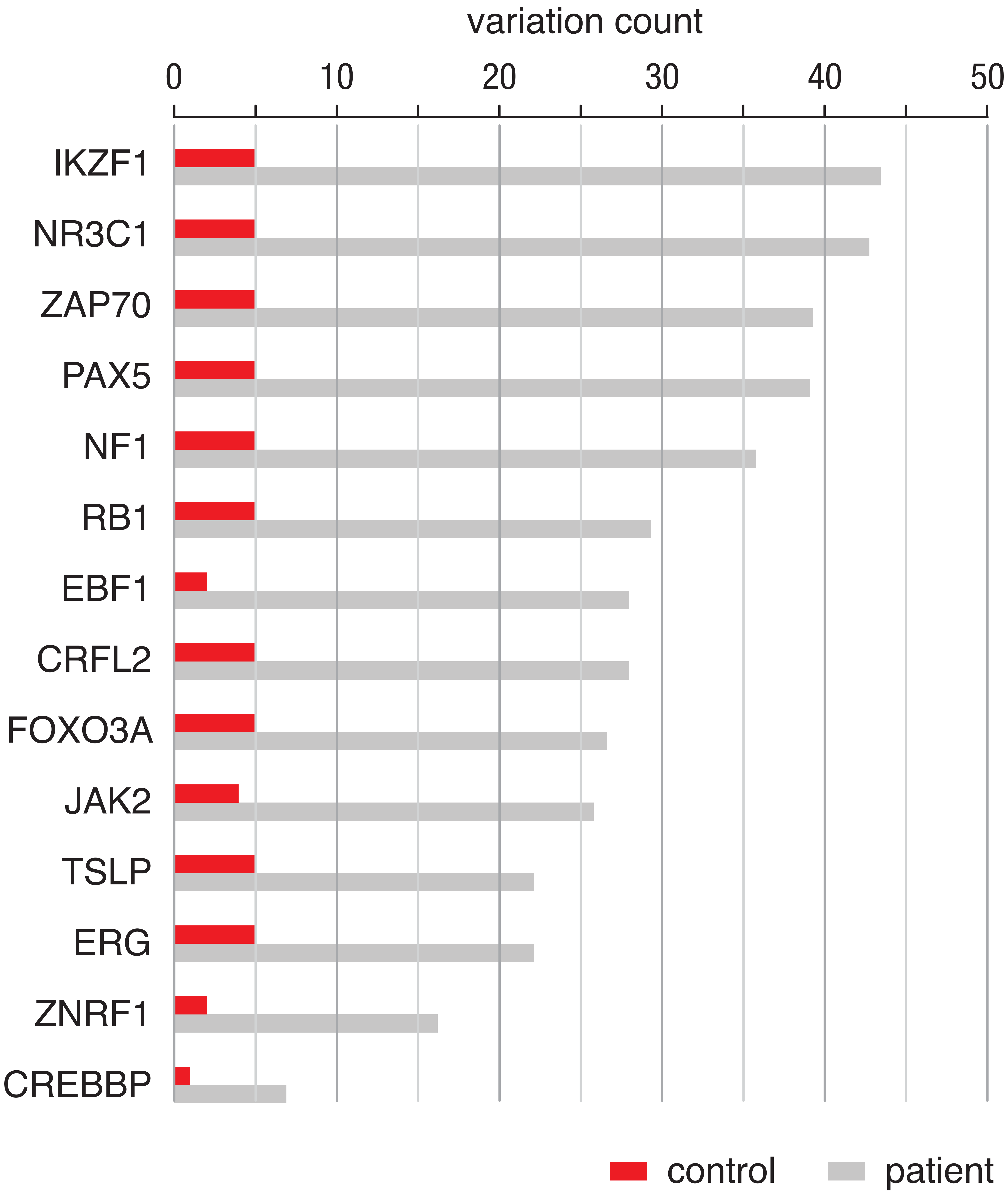


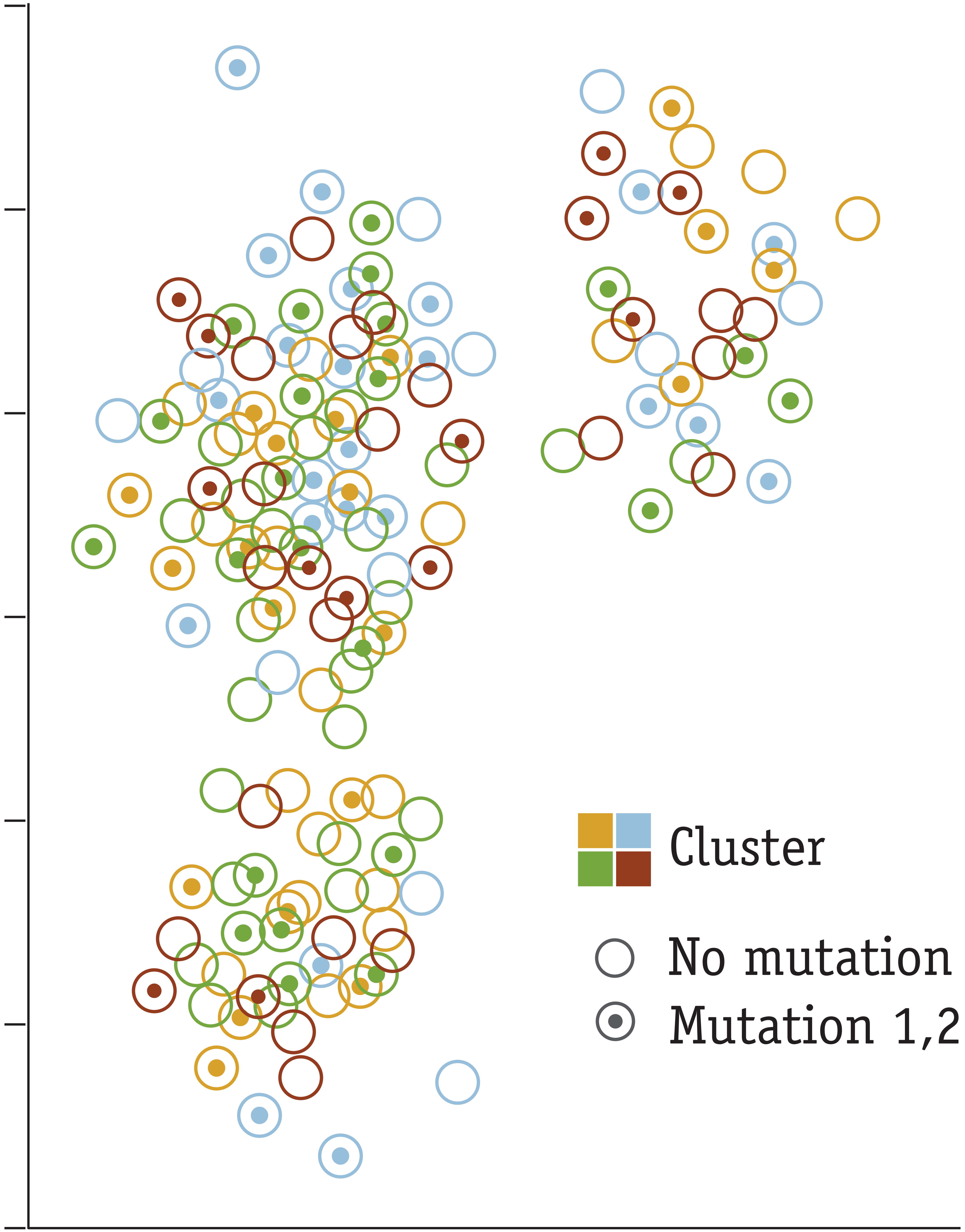
Nat Methods (2013) 10:687-687.



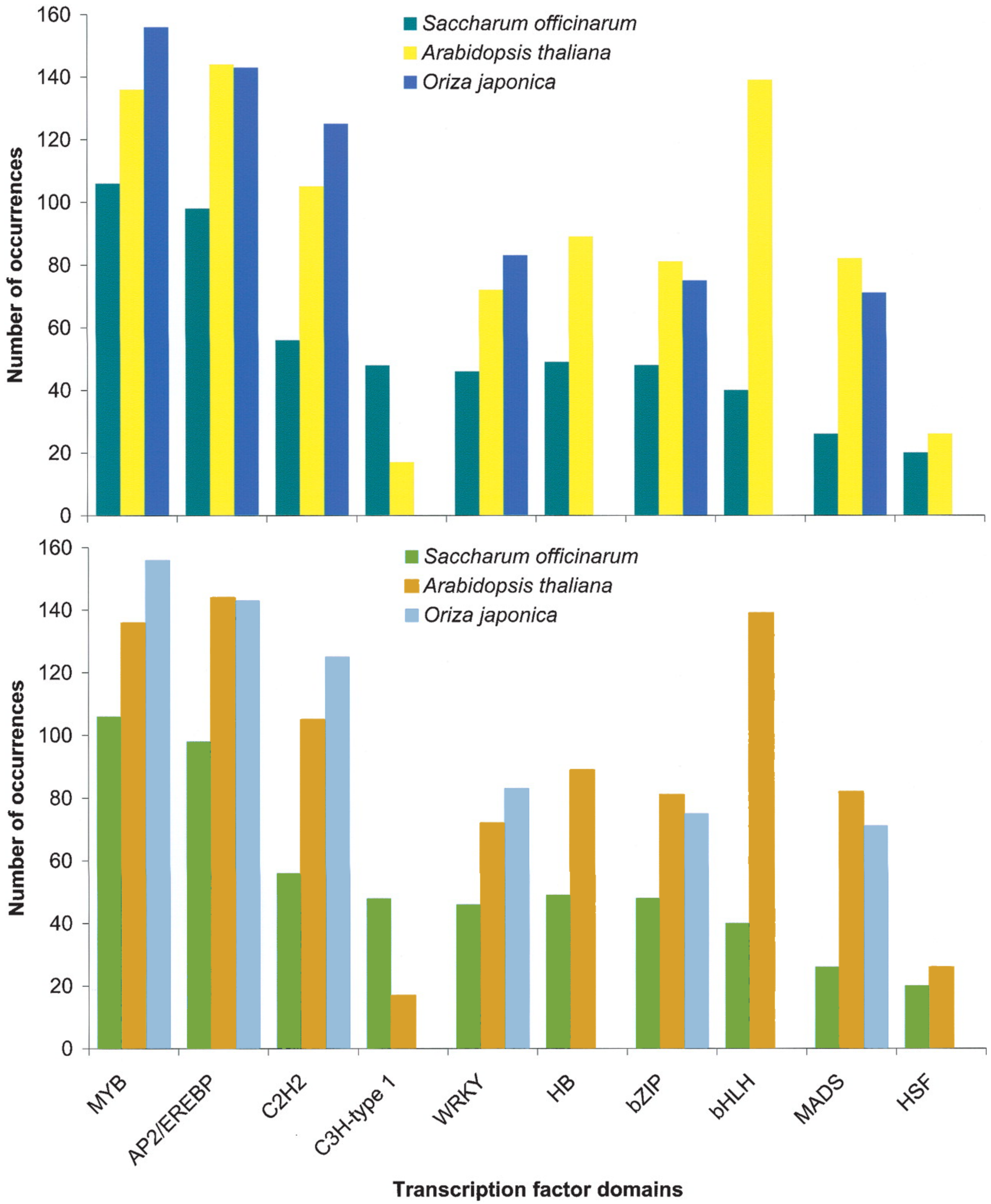
Nat Methods (2016) 13:5.


variation count
0
10
20
30
40
50
IKZF1
NR3C1
ZAP70
PAX5
NF1
RB1
EBF1
CRFL2
FOXO3A
JAK2
TSLP
ERG
ZNRF1
CREBBP
control
patient


variation count
0
10
20
30
40
50
IKZF1
NR3C1
ZAP70
PAX5
NF1
RB1
EBF1
CRFL2
FOXO3A
JAK2
TSLP
ERG
ZNRF1
CREBBP
control
patient



Nat Methods (2010) 7:773.



Genome Res (2003) 13: 2725-2735.

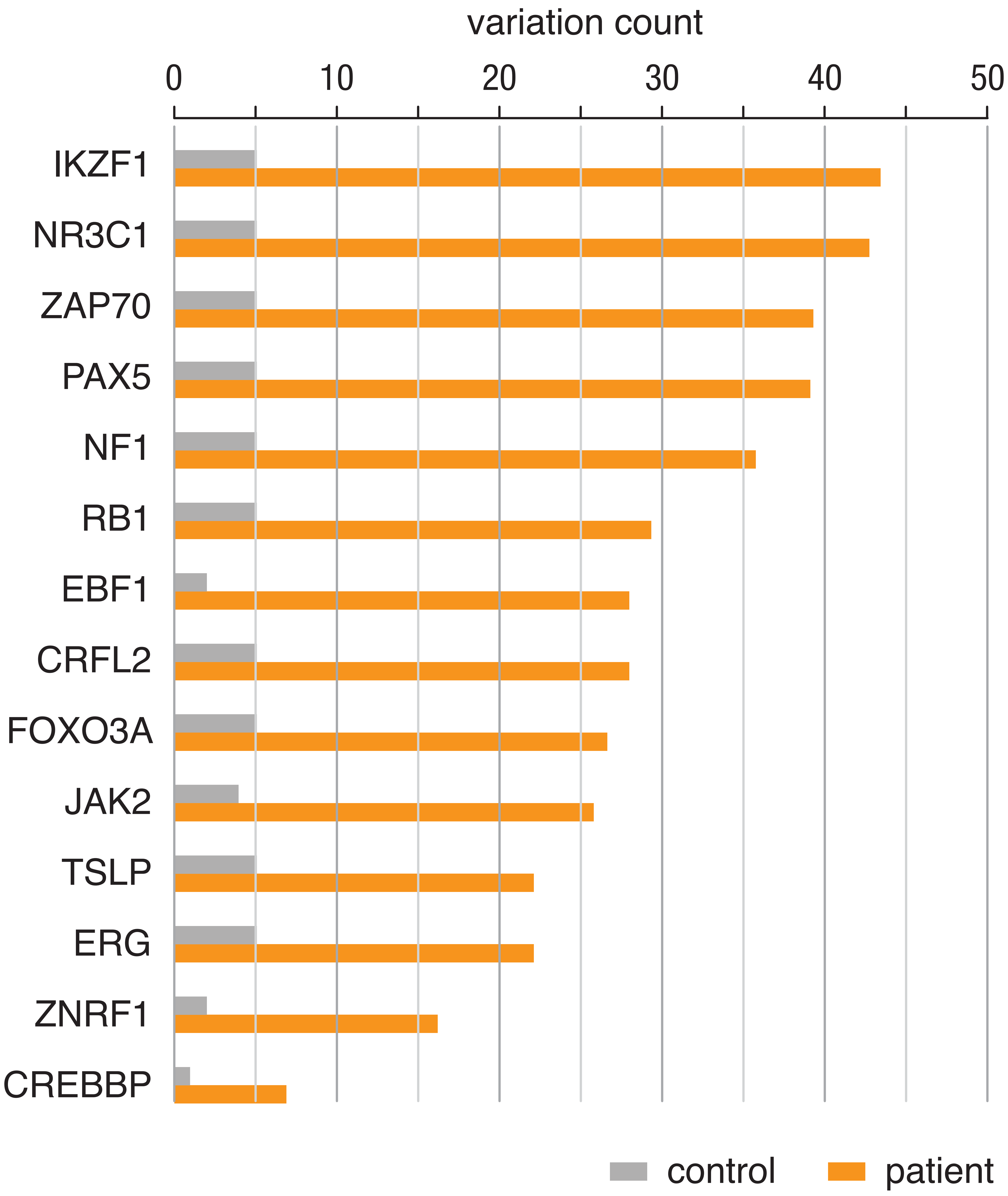
Color is a kind of visual alarm. It really makes you pay attention.

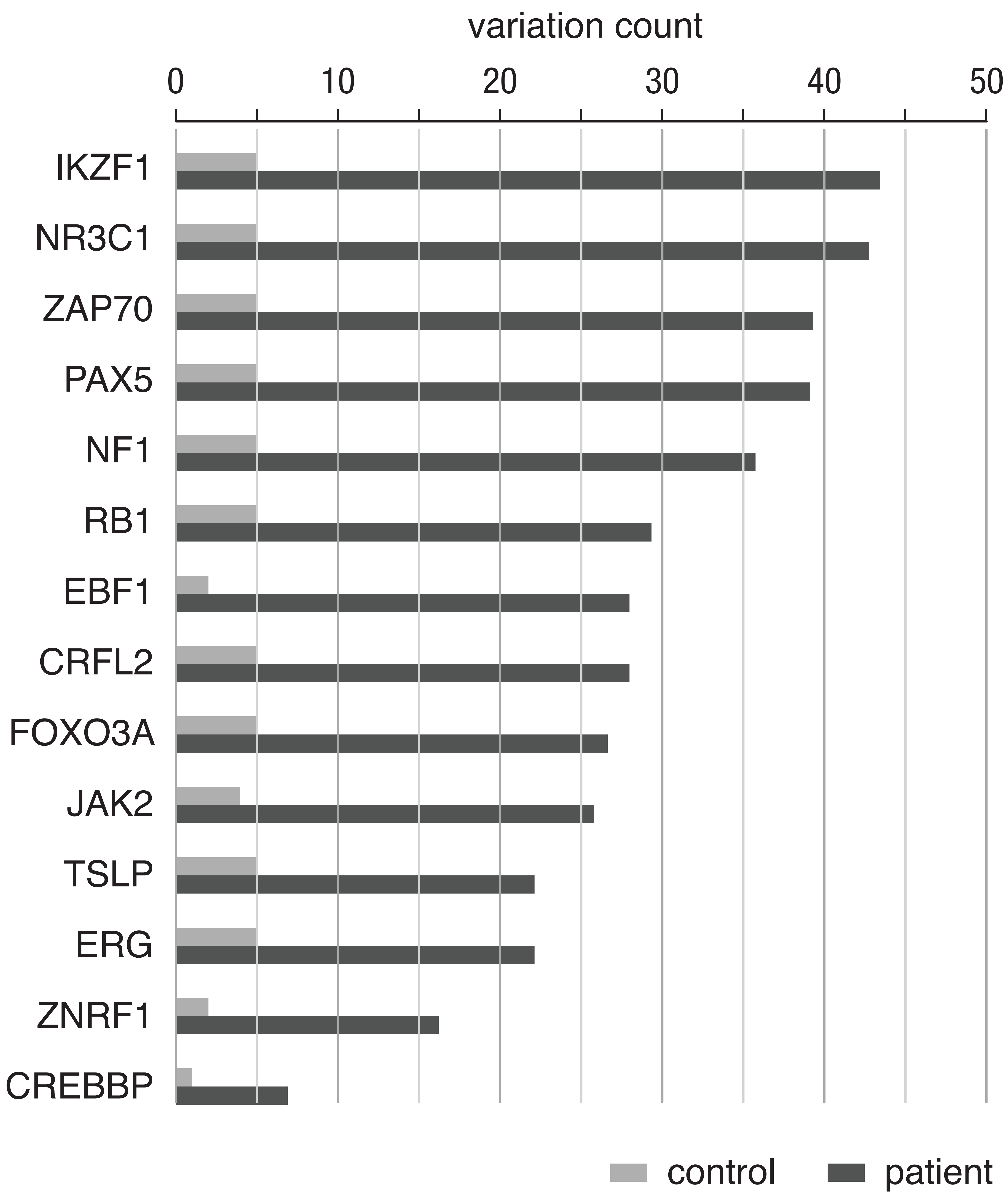
But we don’t always need an alarm—nor do we need color.

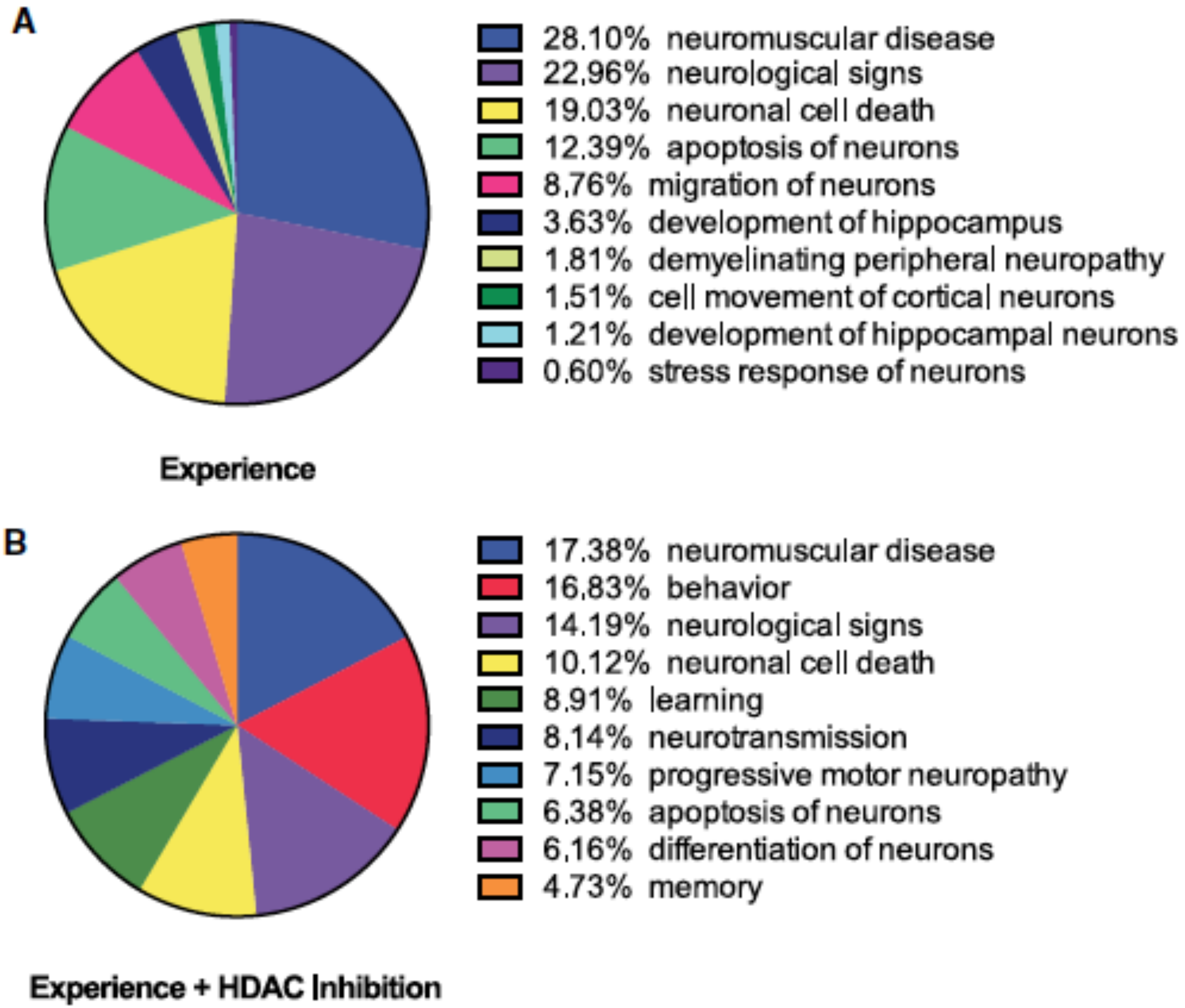
Sometimes just a little visual reminder or notice is all that is needed.

This can be achieved by tone—so, just greys.

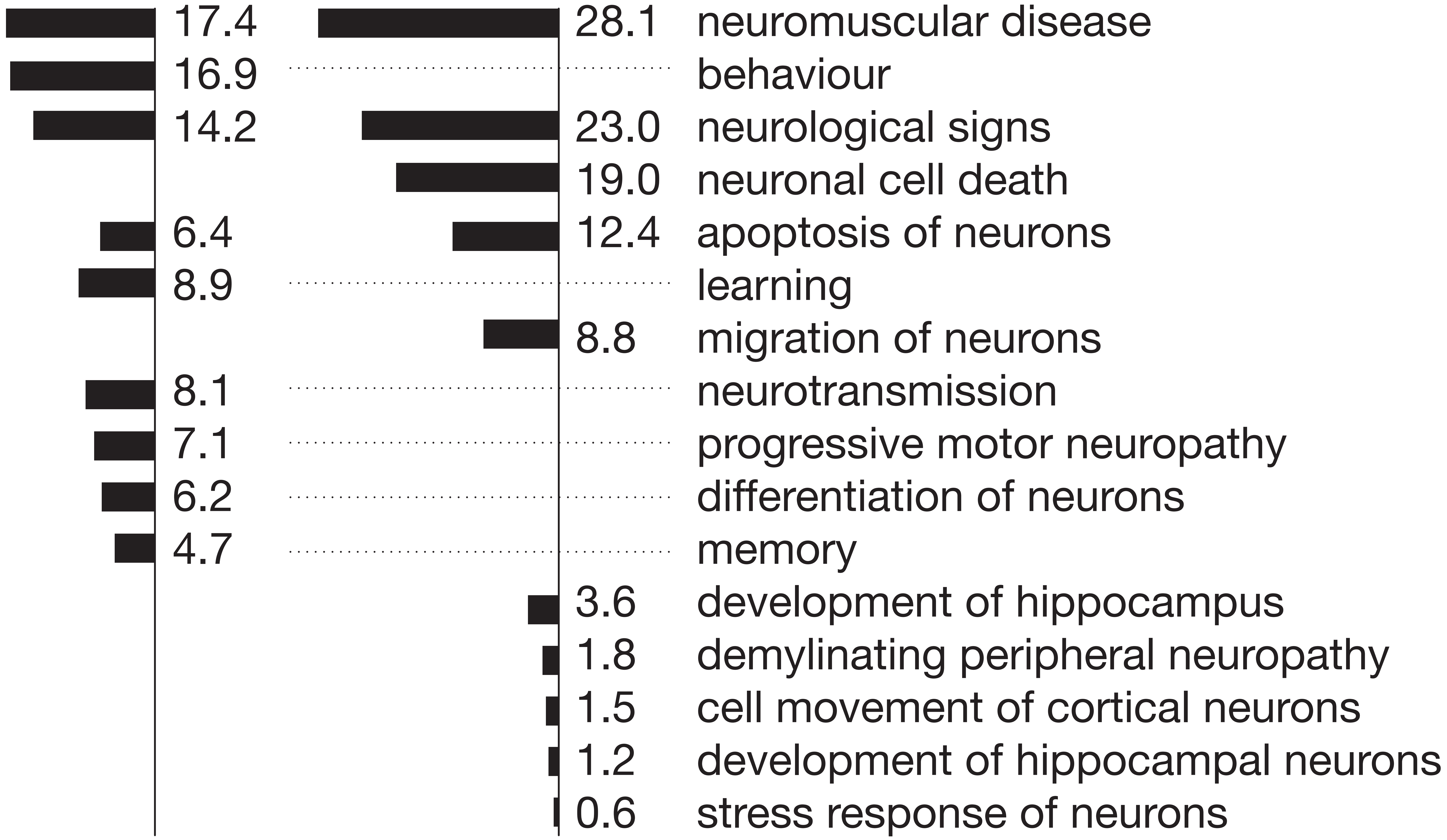
Let’s see how we can keep the attention of the eye and communicate categories and emphasis without color.

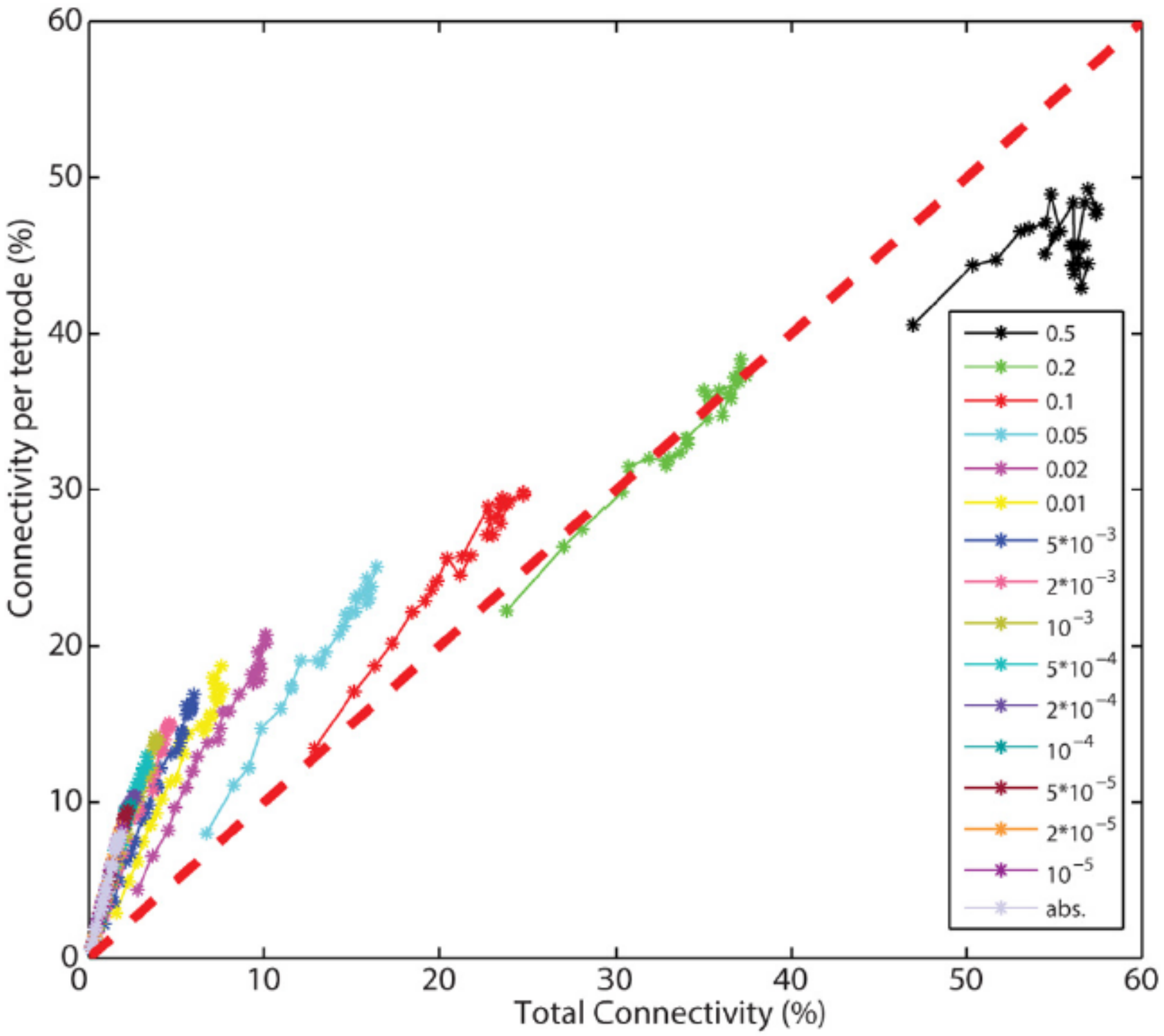

variation count
0
10
20
30
40
50
IKZF1
NR3C1
ZAP70
PAX5
NF1
RB1
EBF1
CRFL2
FOXO3A
JAK2
TSLP
ERG
ZNRF1
CREBBP
control
patient


variation count
0
10
20
30
40
50
IKZF1
NR3C1
ZAP70
PAX5
NF1
RB1
EBF1
CRFL2
FOXO3A
JAK2
TSLP
ERG
ZNRF1
CREBBP
control
patient

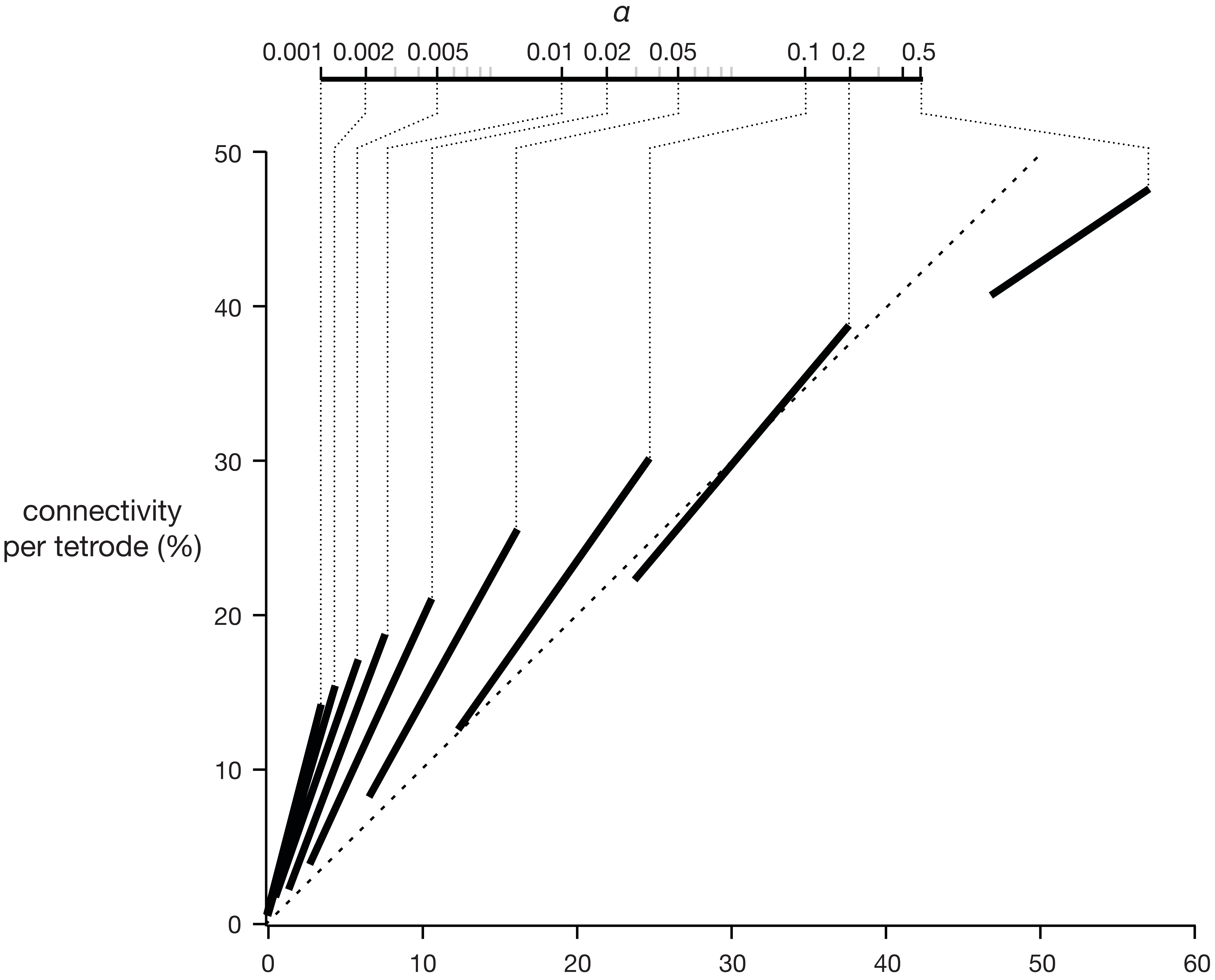


J Neurosci (2015) 35:11729–11742.

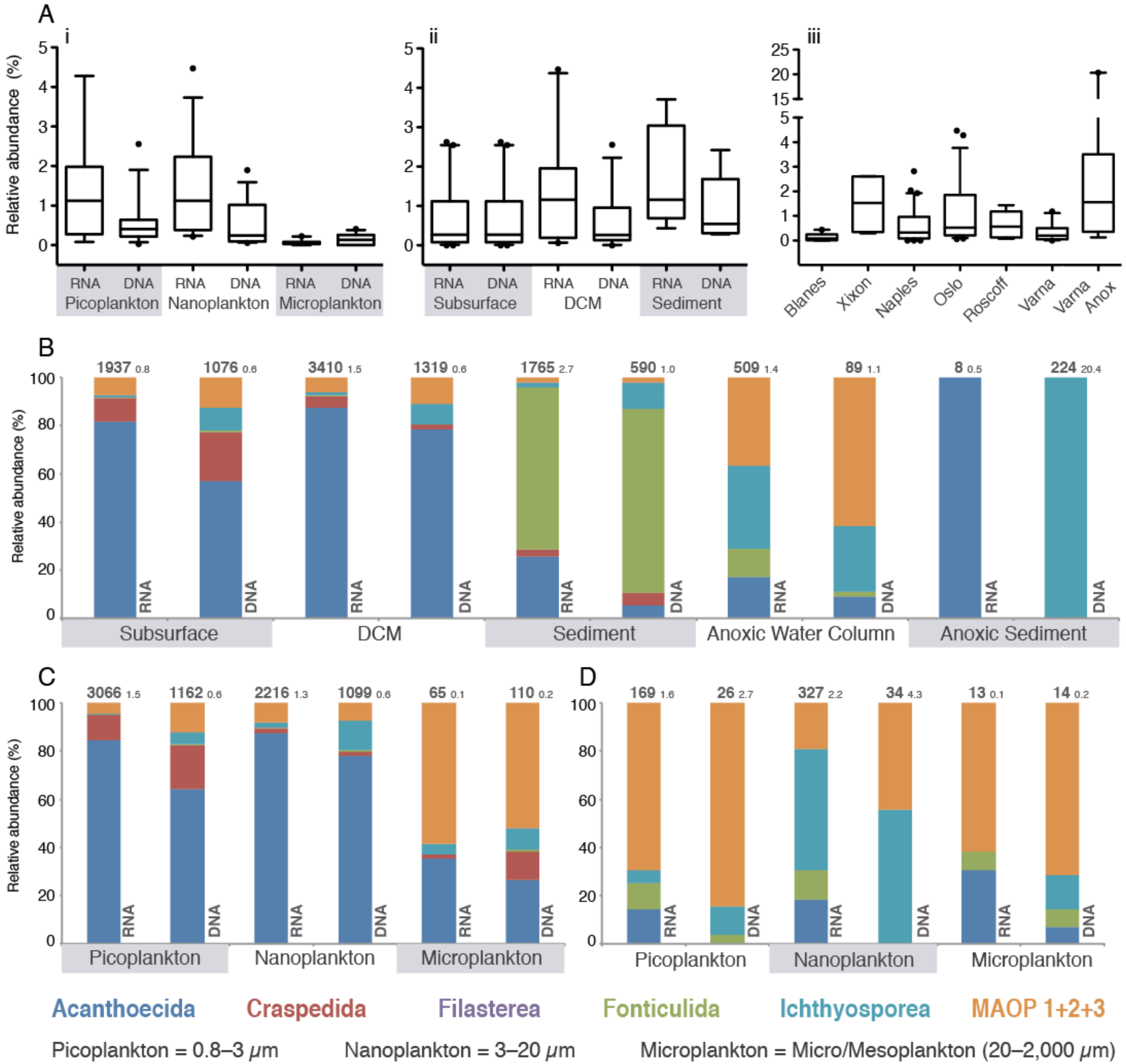

17.4
28.1
neuromuscular disease
16.9
behaviour
14.2
23.0
neurological signs
19.0
neuronal cell death
6.4
12.4
apoptosis of neurons
8.9
learning
8.8
migration of neurons
8.1
neurotransmission
7.1
progressive motor neuropathy
6.2
differentiation of neurons
4.7
memory
3.6
development of hippocampus
1.8
demylinating peripheral neuropathy
1.5
cell movement of cortical neurons
1.2
development of hippocampal neurons
0.6
stress response of neurons



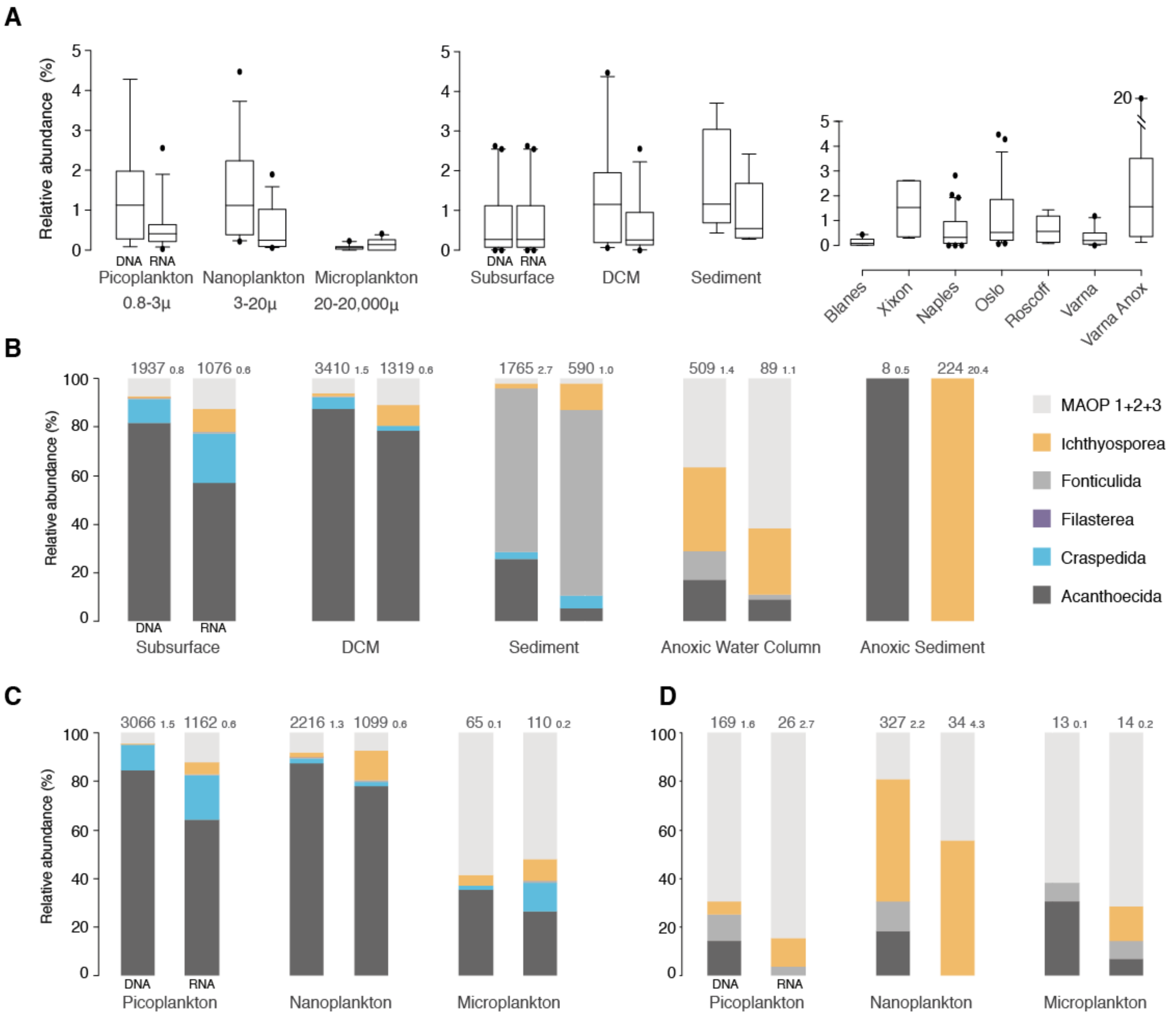
J Neurosci (2014) 34:5454–5467.



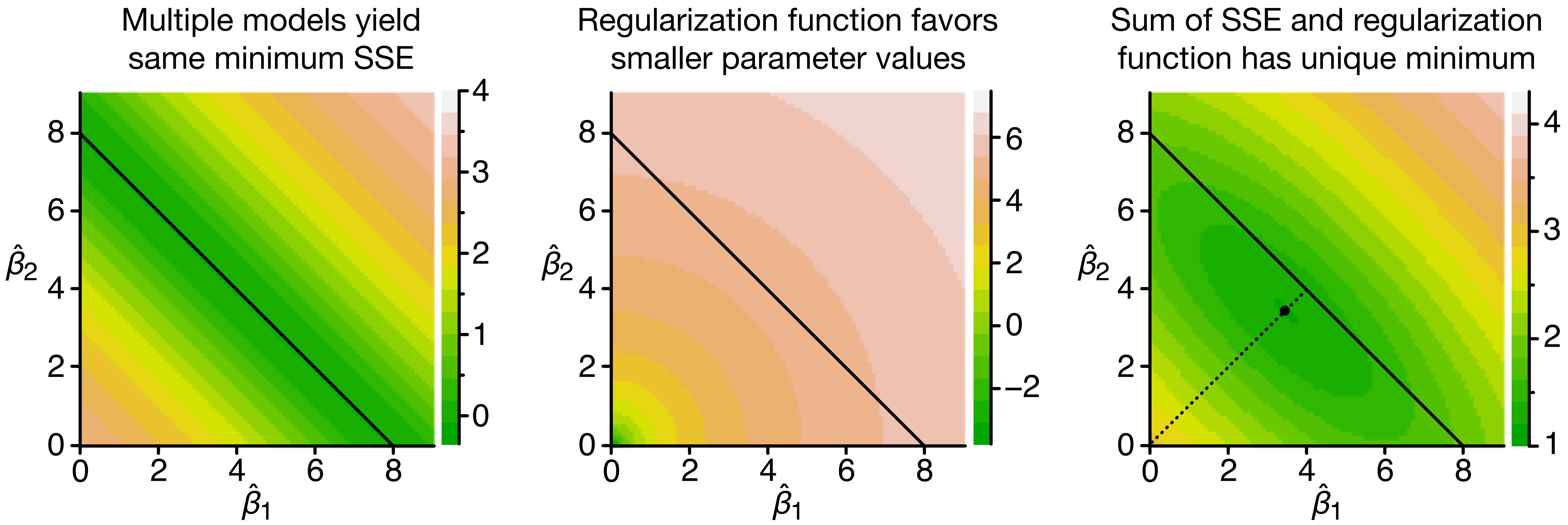
α
0.001
0.002
0.005
0.01
0.02
0.05
0.1
0.2
0.5
connectivity
per tetrode (%)
0
10
20
30
40
50
60



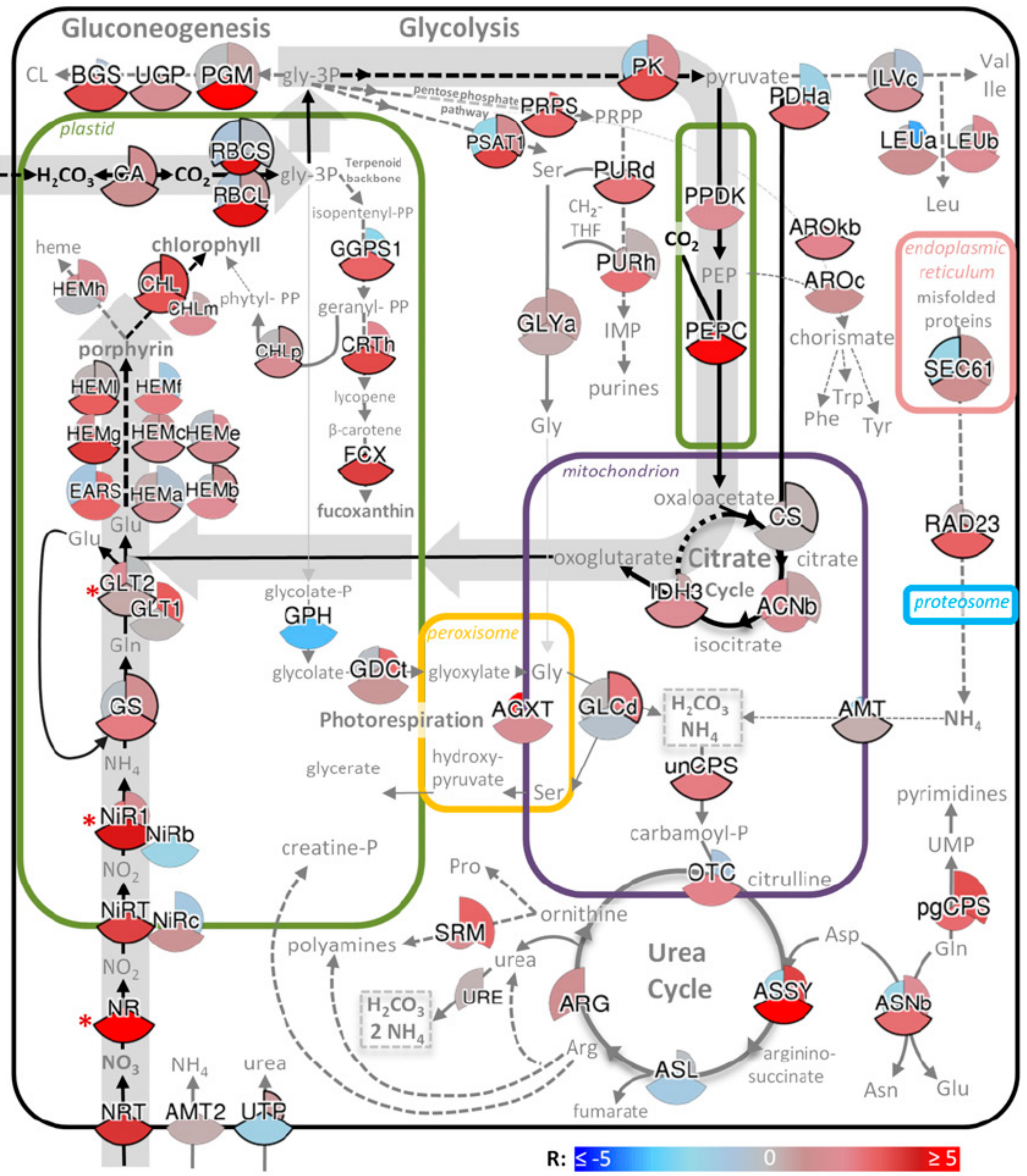
Environ Microbio (2015) 17:3195–3207.



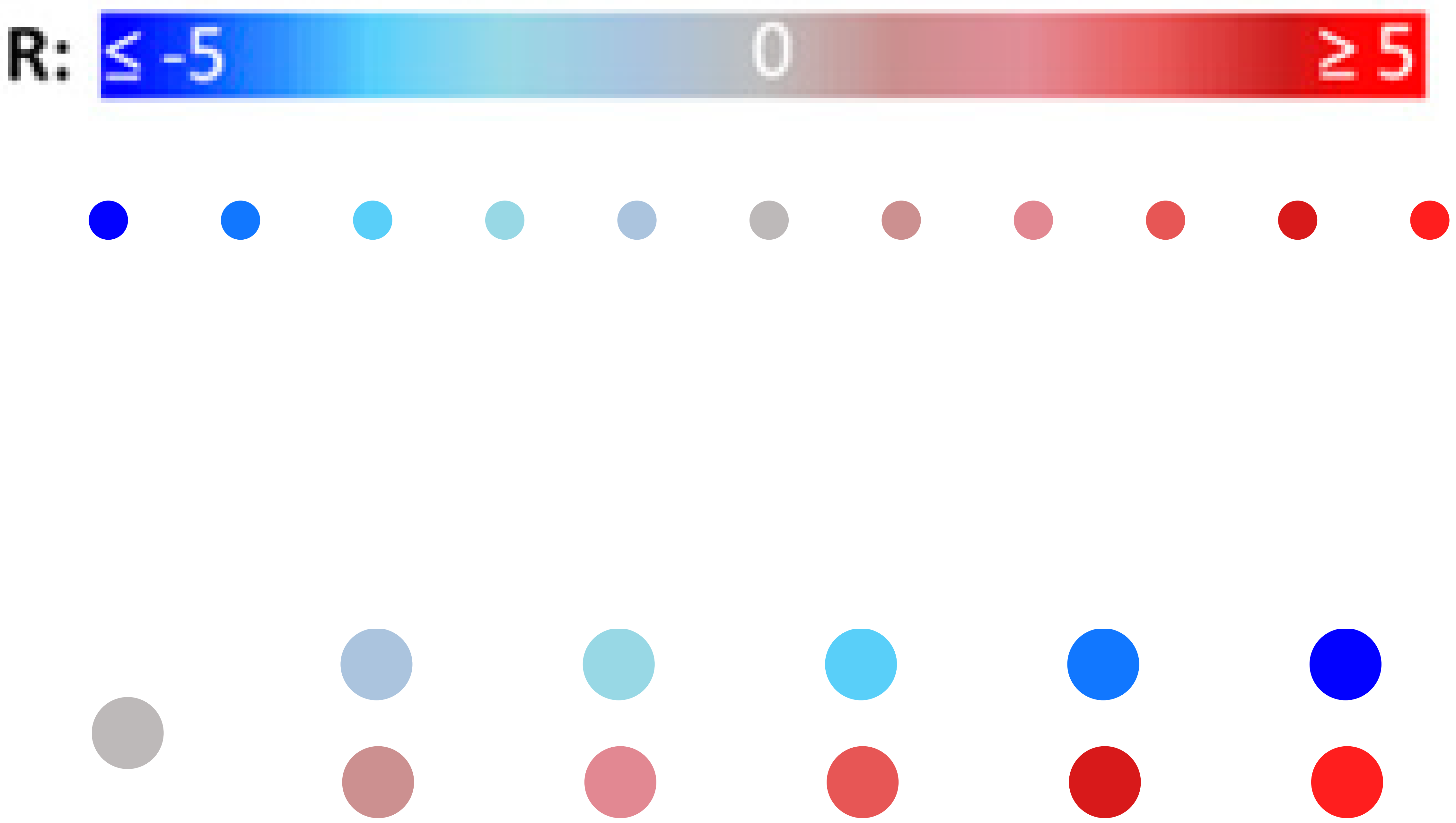
A
B
C
D
Relative abundance (%)
DNA
RNA
Picoplankton 0.8-3µ
Nanoplankton 3-20µ
Microplankton 20-20,000µ
Subsurface
DCM
Sediment
Blanes
Xixon
Naples
Oslo
Roscoff
Varna
Varna Anox
Anoxic Water Column
Anoxic Sediment
MAOP 1+2+3
Ichthyosporea
Fonticulida
Filasterea
Craspedida
Acanthoecida



Multiple models yield same minimum SSE
Regularization function favors smaller parameter values
Sum of SSE and regularization function has unique minimum
$\hat{\beta}_1$
$\hat{\beta}_2$
0
2
4
6
8
−2
1
3



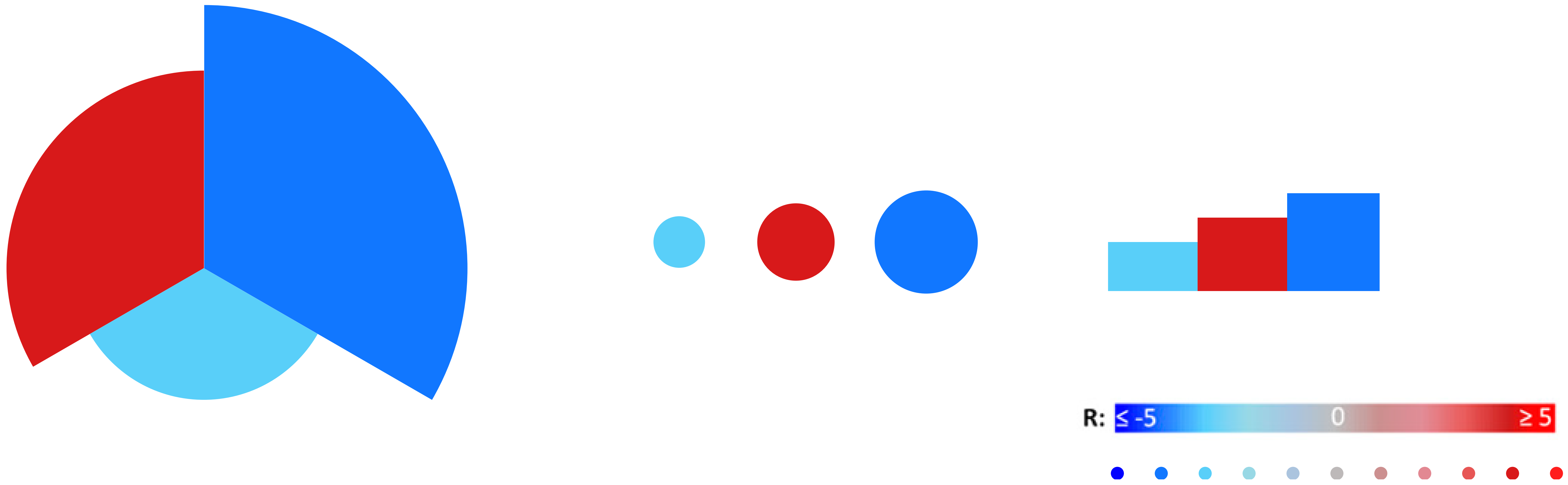
ACNb	aconitate hydratase 2
AGXT	serine--pyruvate aminotransferase
AMT, 2	**ammonium transporter**
ARG	arginase/agmatinase/formimidoglutamate hydrolase
AROc	chorismate synthase
AROkb	3-dehydroquinate synthase
ASL	argininosuccinate lyase
ASNb	asparagine synthase (glutamine-hydrolysing)
ASSY	**argininosuccinate synthase**
BGS	**callose synthase**
CA	**carbonic anhydrase**
CHL	**magnesium chelatase**
CHLm	magnesium-protoporphyrin O-methyltransferase
CHLp	geranylgeranyl reductase
CRTh	carotenoid isomerase
CS	citrate synthase
EARS	glutamyl-tRNA synthetase
FCX	**Lhcr7; fucoxanthin chl a/c light-harvesting protein**
GDTC	glycine dehydrogenase subunit 1
GGPS1	geranylgeranyl pyrophosphate synthetase
GLCD	glycolate oxidase
GLT1, 2	**glutamate synthase (NADPH/NADH, ferredoxin)**
GLYa	glycine hydroxymethyltransferase
GPH	2-phosphoglycolate phosphatase
GS	**glutamine synthetase**
HEMa	glutamyl-tRNA reductase
HEMb	porphobilinogen synthase
HEMc	hydroxymethylbilane synthase
HEMe	uroporphyrinogen decarboxylase
HEMf	coproporphyrinogen III oxidase
HEMg	protoporphyrinogen oxidase
HEMh	ferrochelatase
HEMl	glutamate-1-semialdehyde 2,1-aminomutase
IDH3	isocitrate dehydrogenase (NAD+)
ILVc	ketol-acid reductoisomerase
LEUa,b	2,3 isopropylmalate synthase, dehydrogenase
NiRb, 1	**nitrite reductase (NADH/NADPH, ferredoxin)**
NiRT, c	**nitrite transporters**
NR	**nitrate reductase**
NRT	**nitrate transporter**
OTC	**ornithine carbamoyltransferase**
PDHa	pyruvate dehydrogenase E1 component subunit alpha
PEPC	**phosphoenolpyruvate carboxylase**
pgCPS	carbamoyl-phosphate synthase / aspartate carbamoyltransferase
PGM	**phosphoglucomutase**
PK	pyruvate kinase
PPDK	pyruvate,orthophosphate dikinase
PRPS	ribose-phosphate pyrophosphokinase
PSAT1	phosphoserine aminotransferase
PURd	phosphoribosylamine--glycine ligase / phosphoribosylglycinamide
PURh	phosphoribosylaminoimidazolecarboxamide formyltransferase
RAD23	UV excision repair protein
RBCS, L	**ribulose-bisphosphate carboxylase (small , large) chain**
SEC61	transport protein subunit alpha
SRM	spermidine synthase
UGP	**precursor of phosphorylase udp-glucose diphosphorylase**
unCPS	**carbamoyl-phosphate synthase mitochondrial precursor**
URE	urease
UTP	**urea transporter**

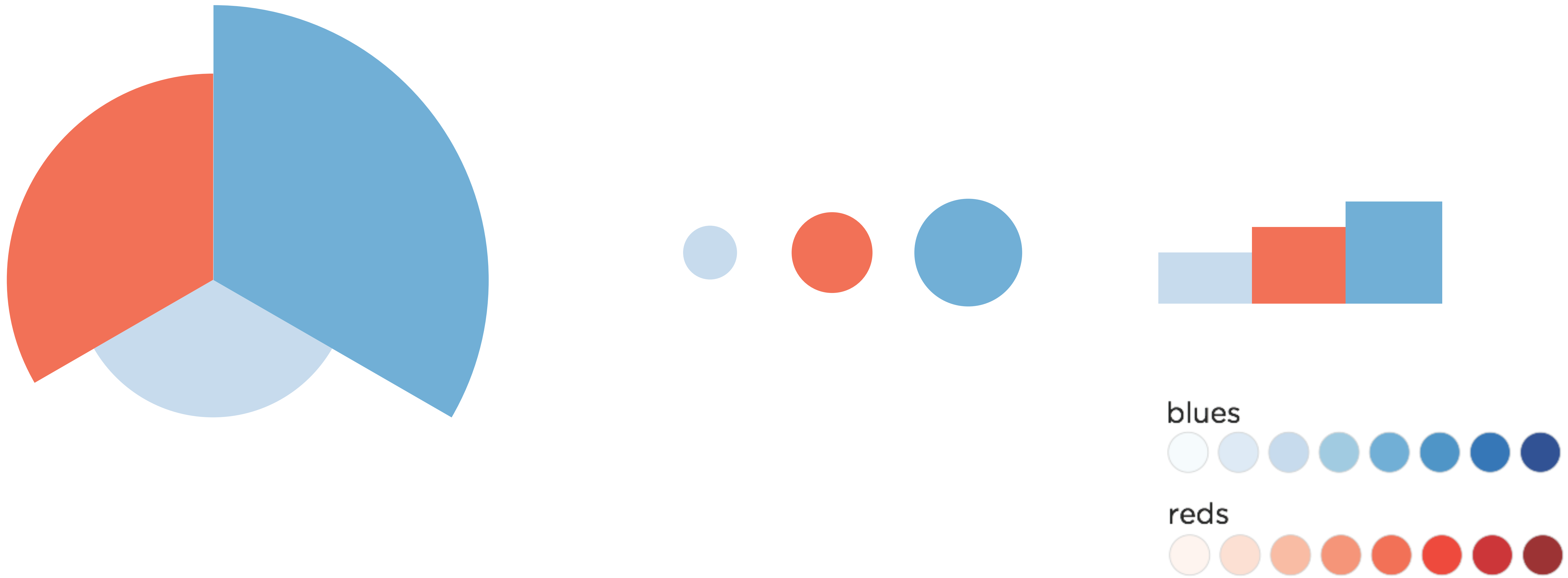


PNAS (2012) 109:E317-E325.

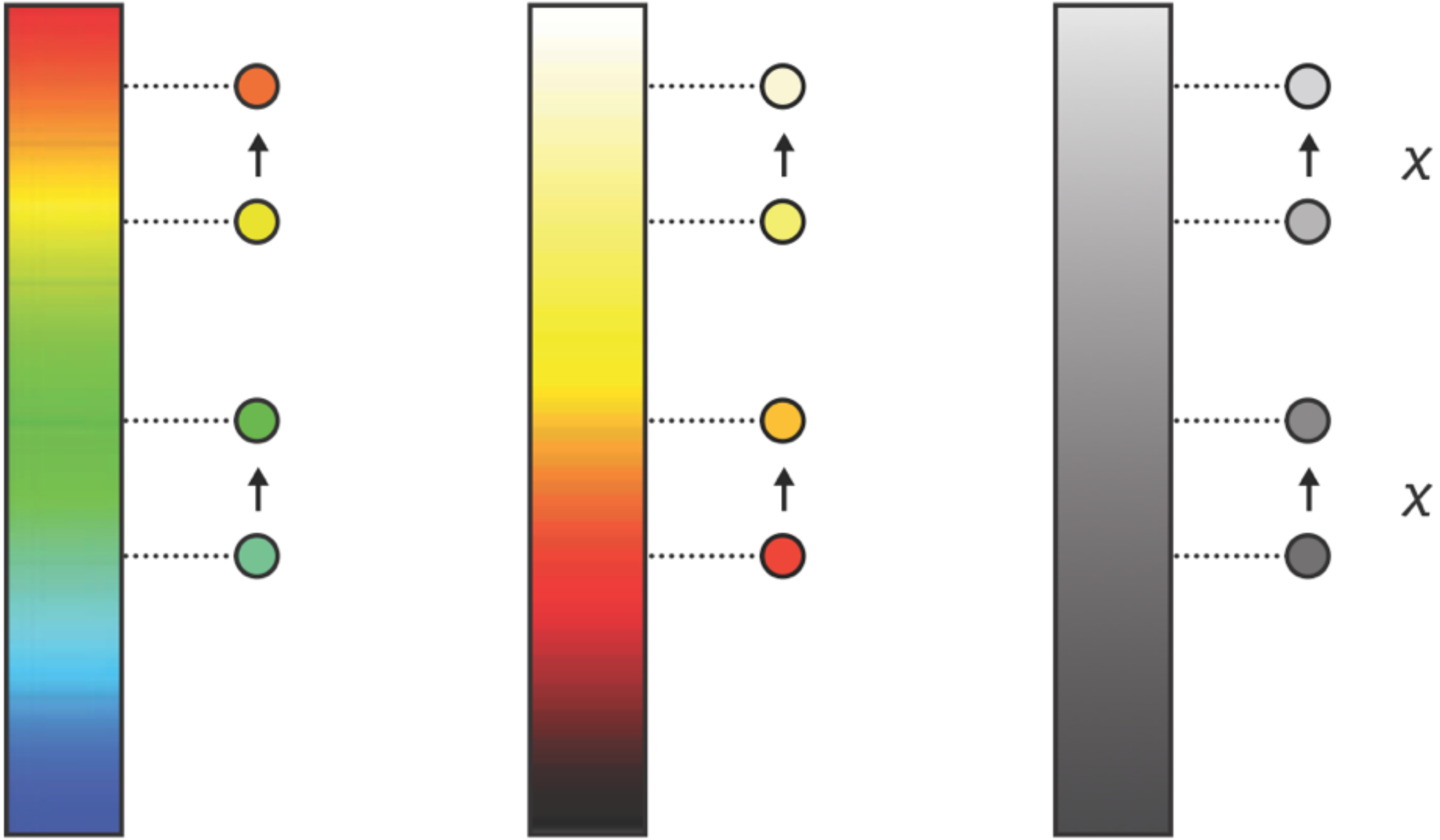
BREWER PALETTES

www.colorbrewer.org & mkweb.bcgsc.ca/brewer



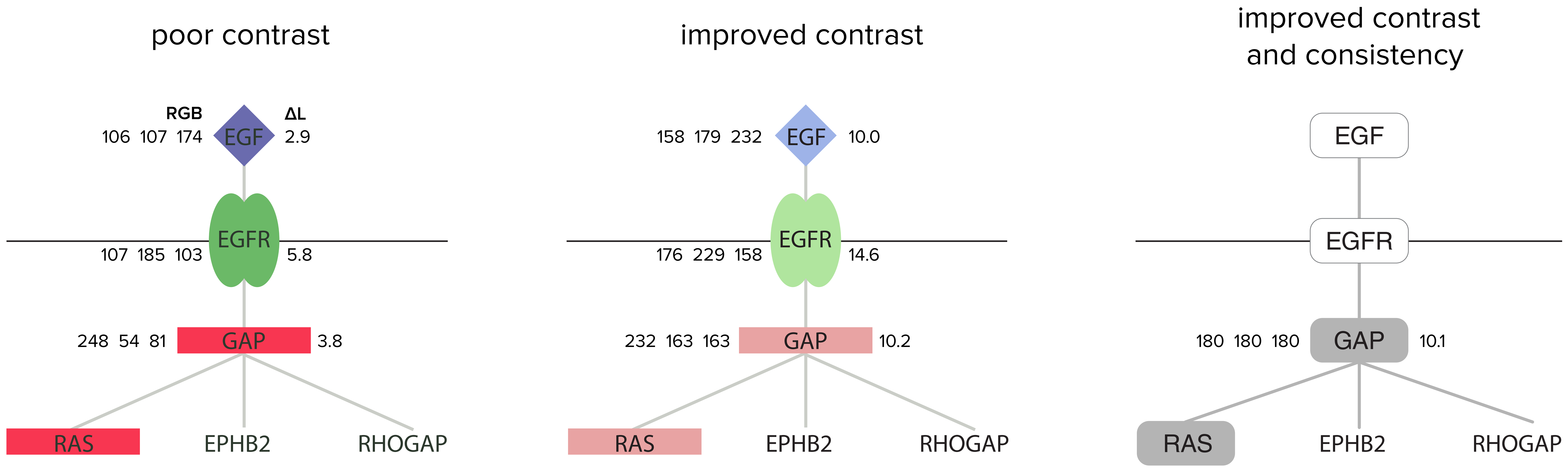


PNAS (2012) 109:E317-E325.



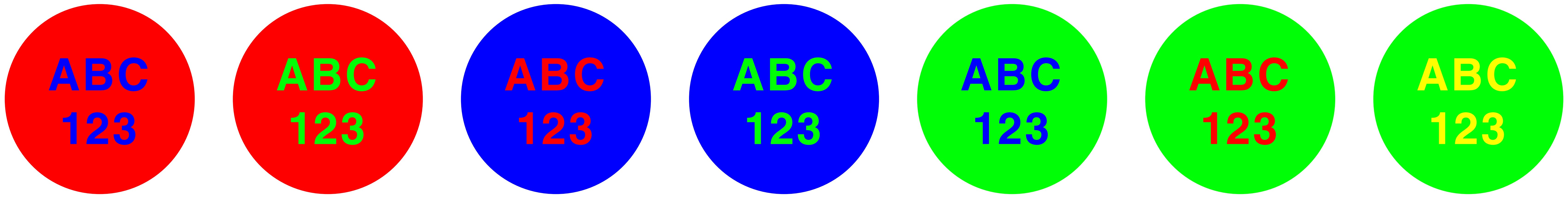
Nat Methods (2011) 8:189.


uniform HSV hue
+150%
+200%
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
luminance

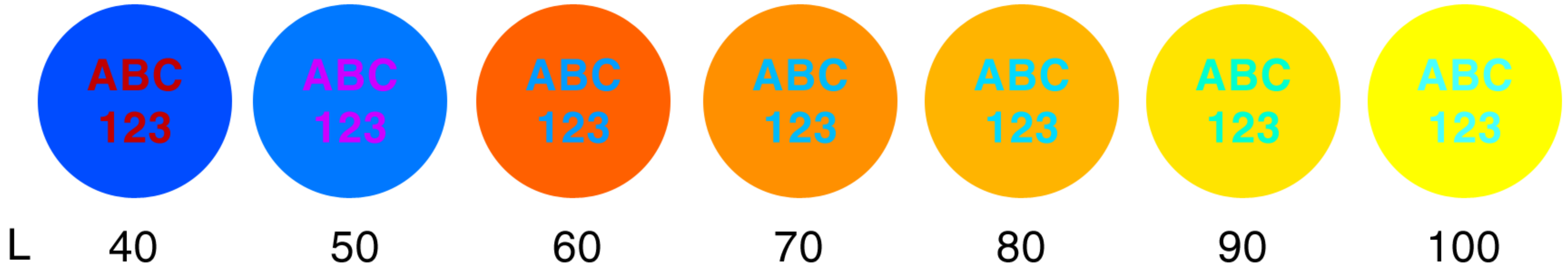


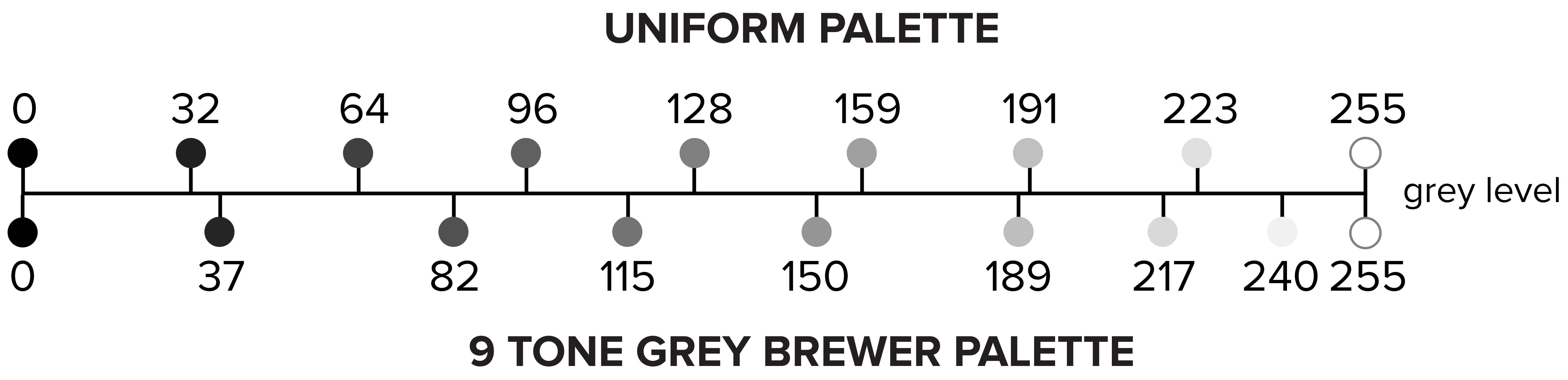
Genome Res (2003) 13: 2363-2371.

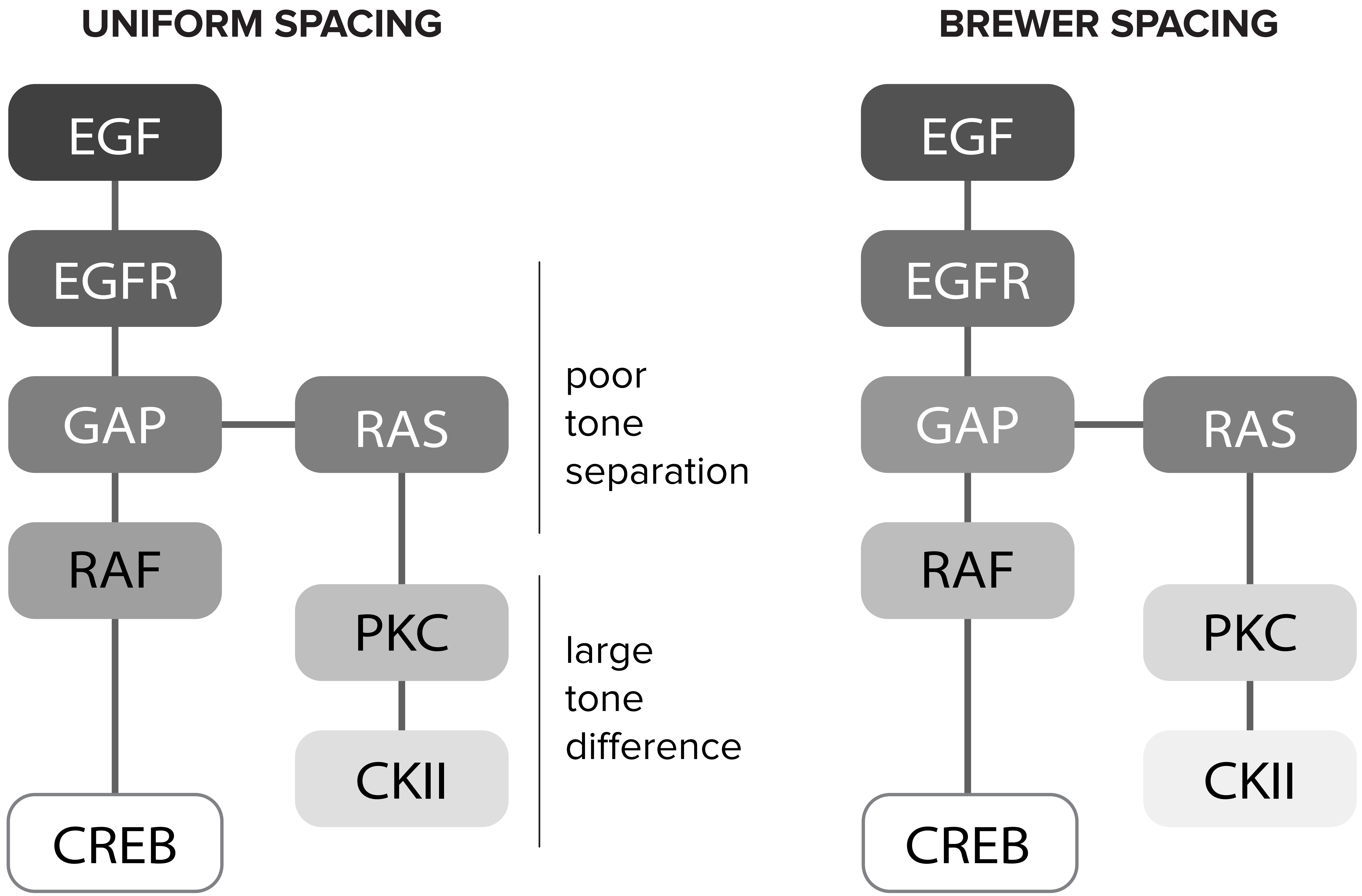
AVOID ADJACENT PURE COLORS



AVOID ADJACENT COLORS WITH SIMILAR LUMINANCE




UNIFORM PALETTE
0
32
64
96
128
159
191
223
255
grey level
0
37
82
115
150
189
217
240
255
9 TONE GREY BREWER PALETTE


UNIFORM SPACING
EGF
EGFR
GAP
RAS
RAF
PKC
CKII
CREB
poor tone separation
large tone difference
BREWER SPACING
EGF
EGFR
GAP
RAS
RAF
PKC
CKII
CREB

The moral here is that often color isn't even necessary. And even worse, often it gets in the way of legibility and clarity.

When using color, ask yourself—do I need it? Try to work around it using grey tones from Brewer palettes. If you succeed, you're in a perfect place to use spot color, sparingly, for emphasis.

Color does make things "exciting" to the eye. But then what's your goal? To excite the eye or inform the brain? Often if you just do the first one, the brain checks out because it gets satiated early. When was the last time a movie with a lot of great special effects also had a great plot?

Always be very critical of any kind of graphics that use a color ramp. If your doctor is looking at your brain scan and it's using a rainbow color map, get another doctor.

Above all, become familiar with alternative—and more useful—ways in which color is characterized. Read up on LCH and Lab color spaces—use that "L" coordinate in the Photoshop color picker!

created by

Martin Krzywinski, Kim Bell-Anderson & Philip Poronnik

written and designed by

Martin Krzywinski

production

One Ski Digital Media Productions

with financial support by

University of Sydney

filmed at

University of Sydney, Australia

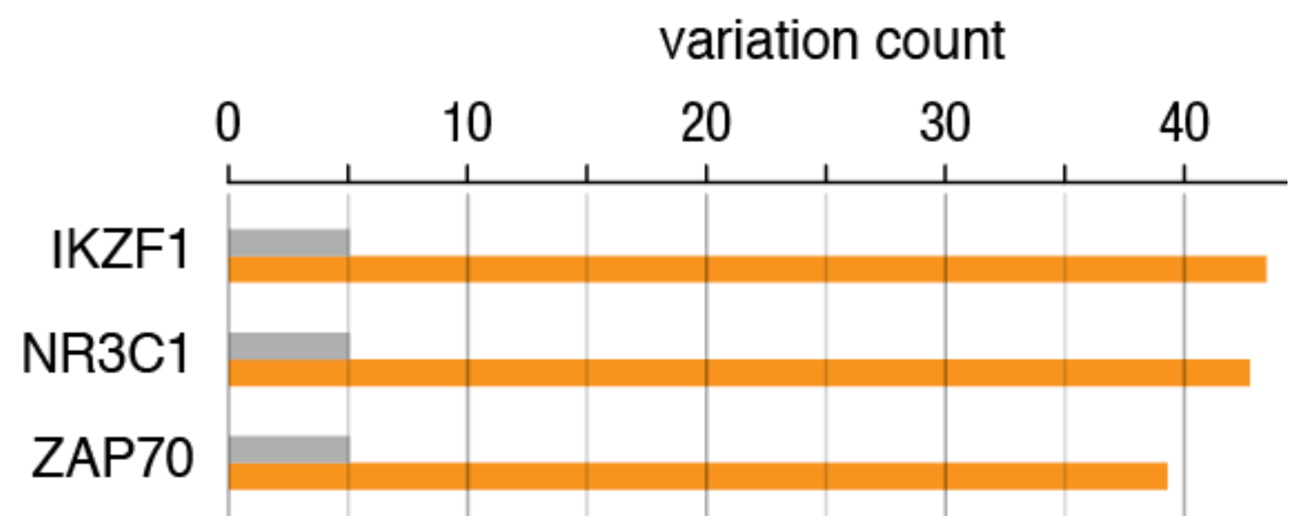
EXERCISE 1

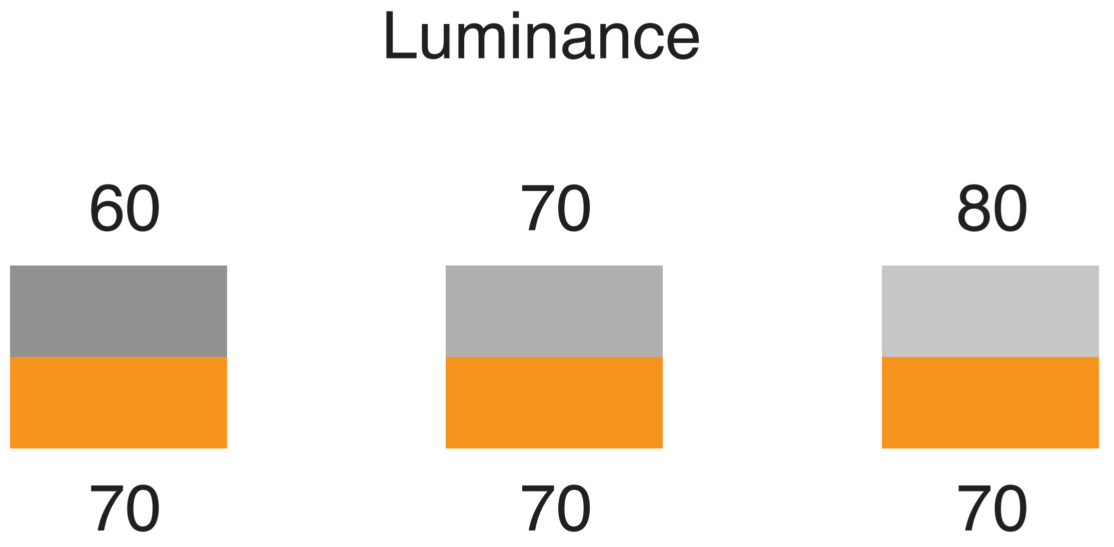
One of the most useful properties of a color is its perceived brightness. This is the “L” component in the Lab color space. It can be accessed in the Photoshop color picker.

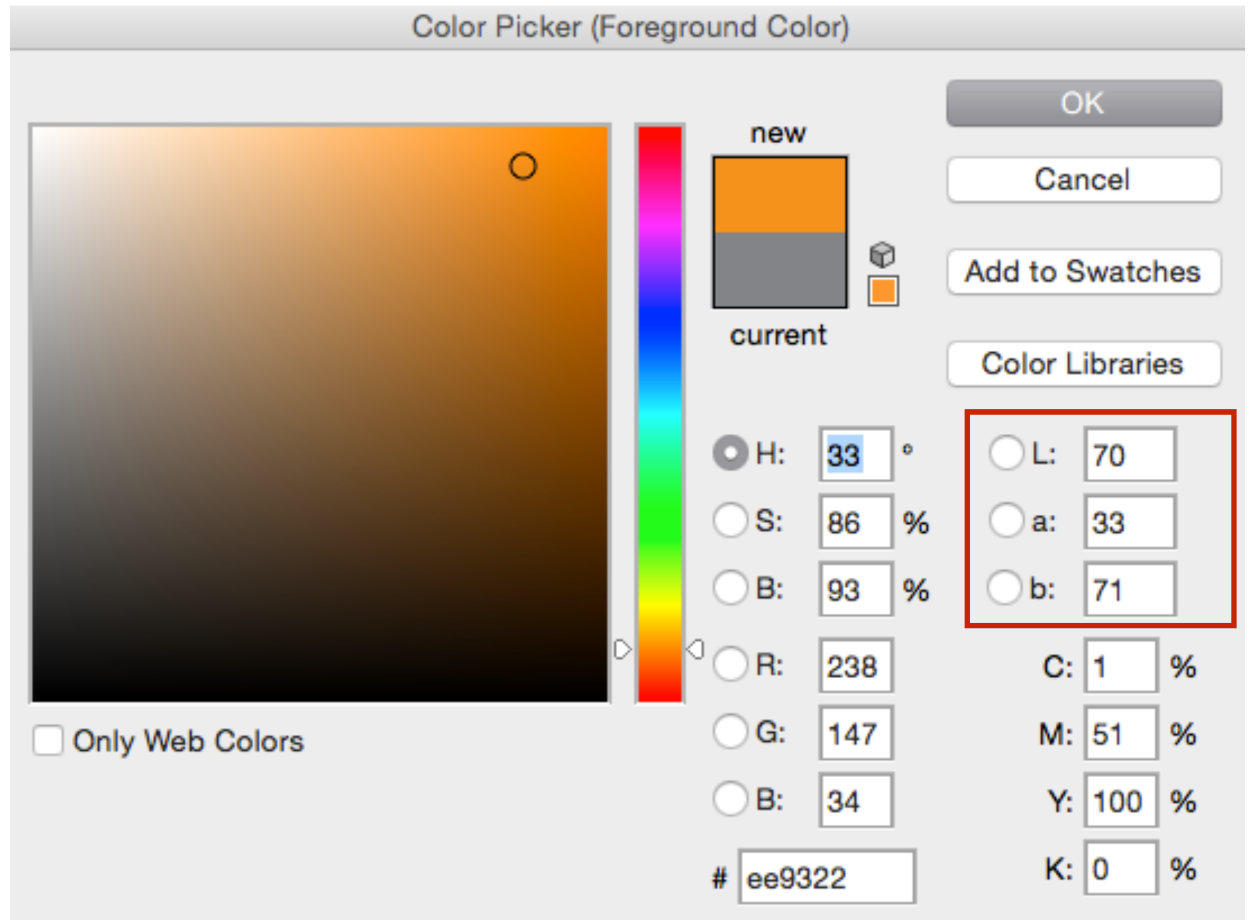
The Lab color space is perceptually uniform. This means that if two colors only vary in L, and have a difference of ΔL=5 and another pair has ΔL=15, then the second pair will be perceived roughly as three times different.

In picking the grey and orange for the bar plot, I’ve used a grey that has the same L value as the orange. This avoids unwanted luminance contrast.

For the three colors below, find their luminance. Then, find a grey that has the same luminance. Draw a bar plot that uses these pairs—are they equally effective?







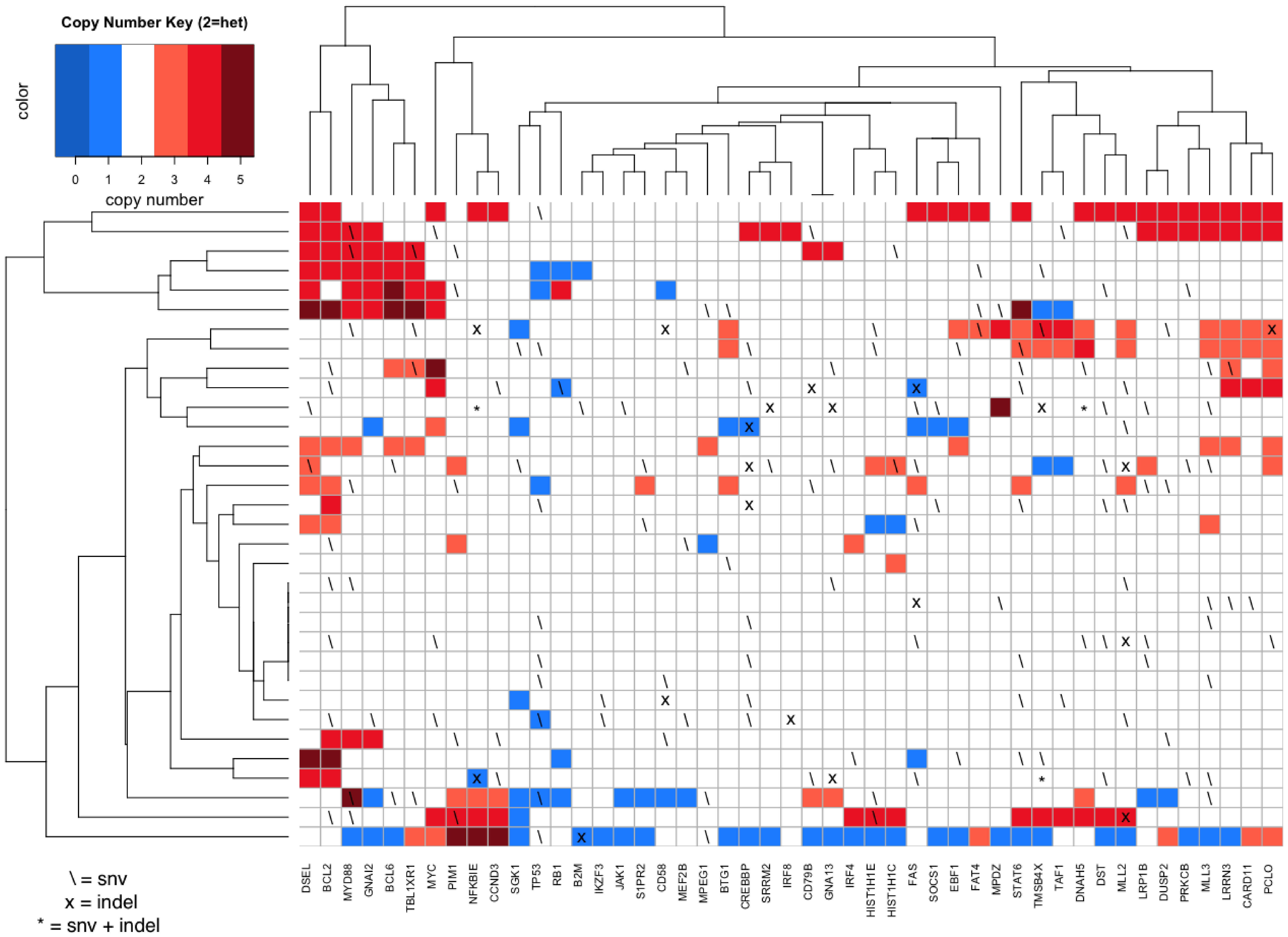
Lab color space coordinates

EXERCISE 2

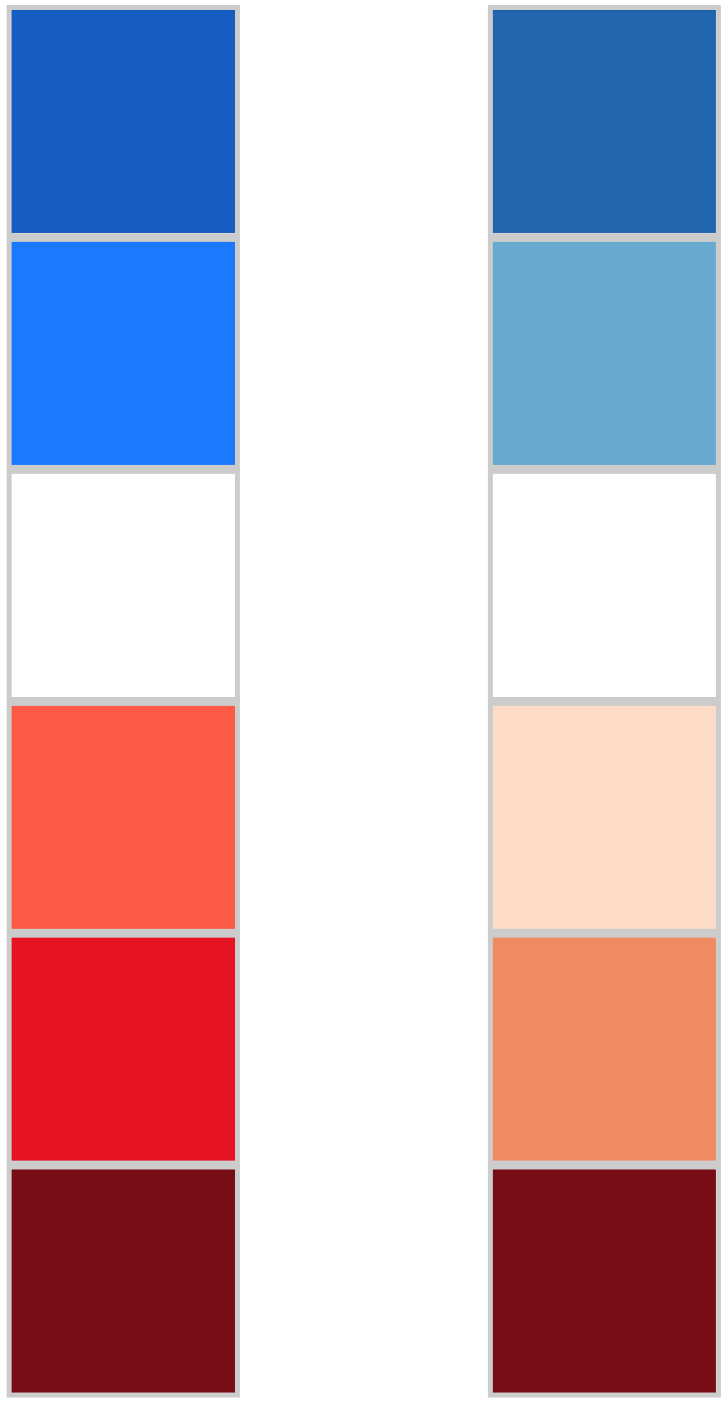
What's wrong with the color ramp in this heatmap?

Find a better one using the Brewer palettes. Pick one that is red/blue and another from the pink-yellow-green palette.

What if you were to map 0 and 1 values to grey and 3, 4, 5 to a color? Would this suggest something about the data?



CNV, SNV, and INDEL distribution for selected genes in the DLBCL cohort



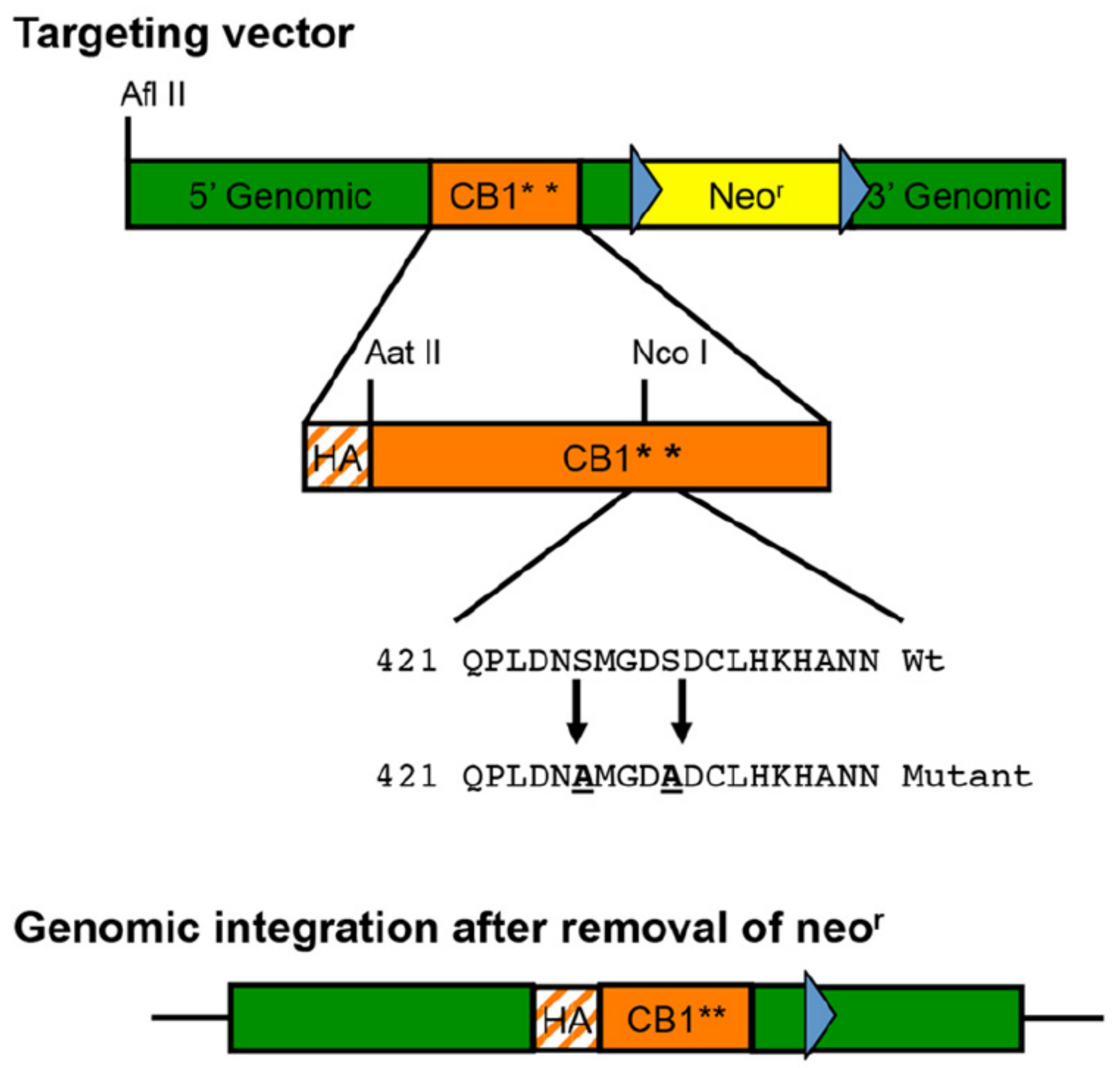
EXERCISE 3

Pick better colors for each of the elements in this figure.

What is important here? Is there a part of the genomic segment shown that can be considered as “background”?

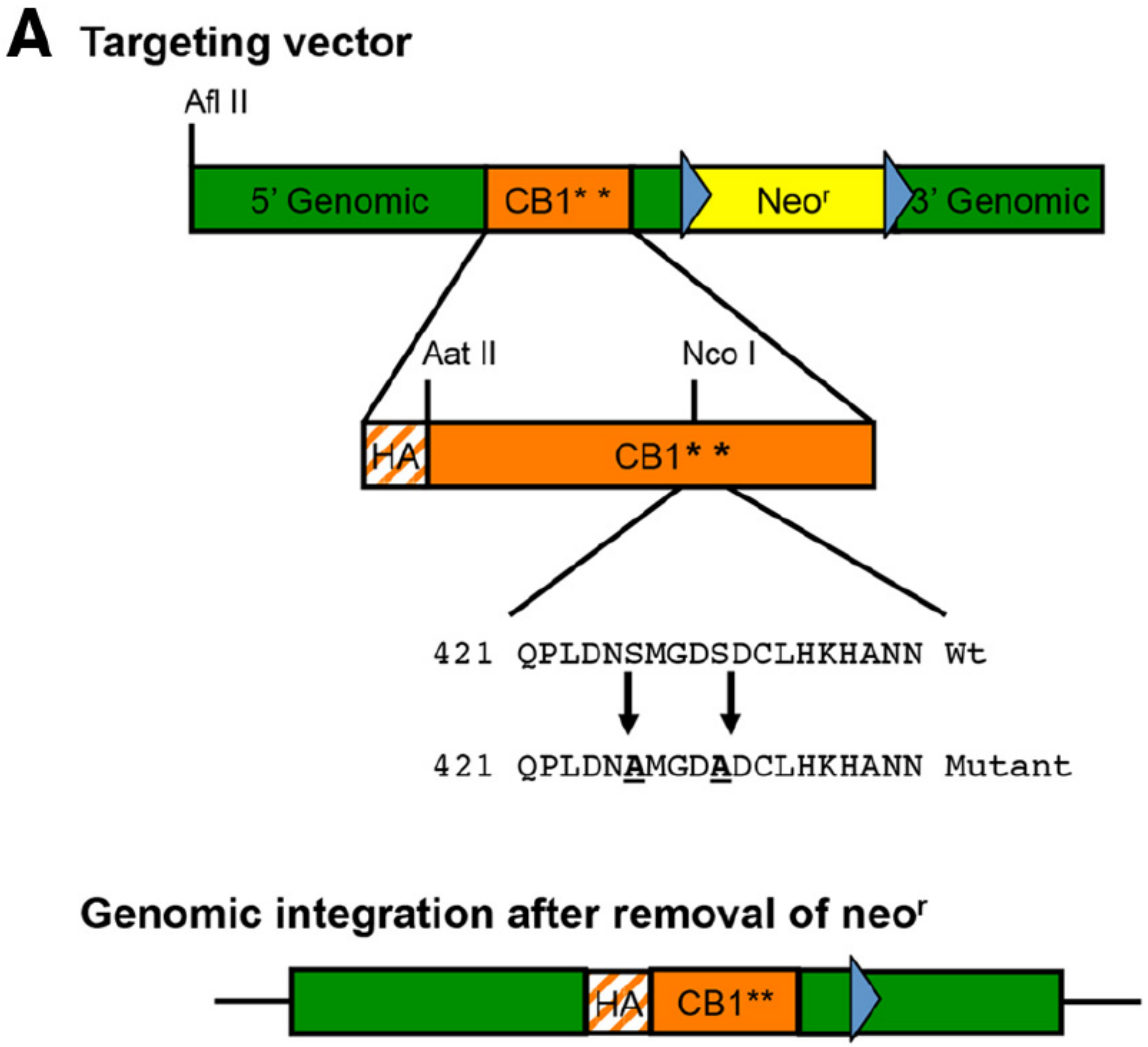
What do the blue arrows represent? Why are they blue? Are they necessary?

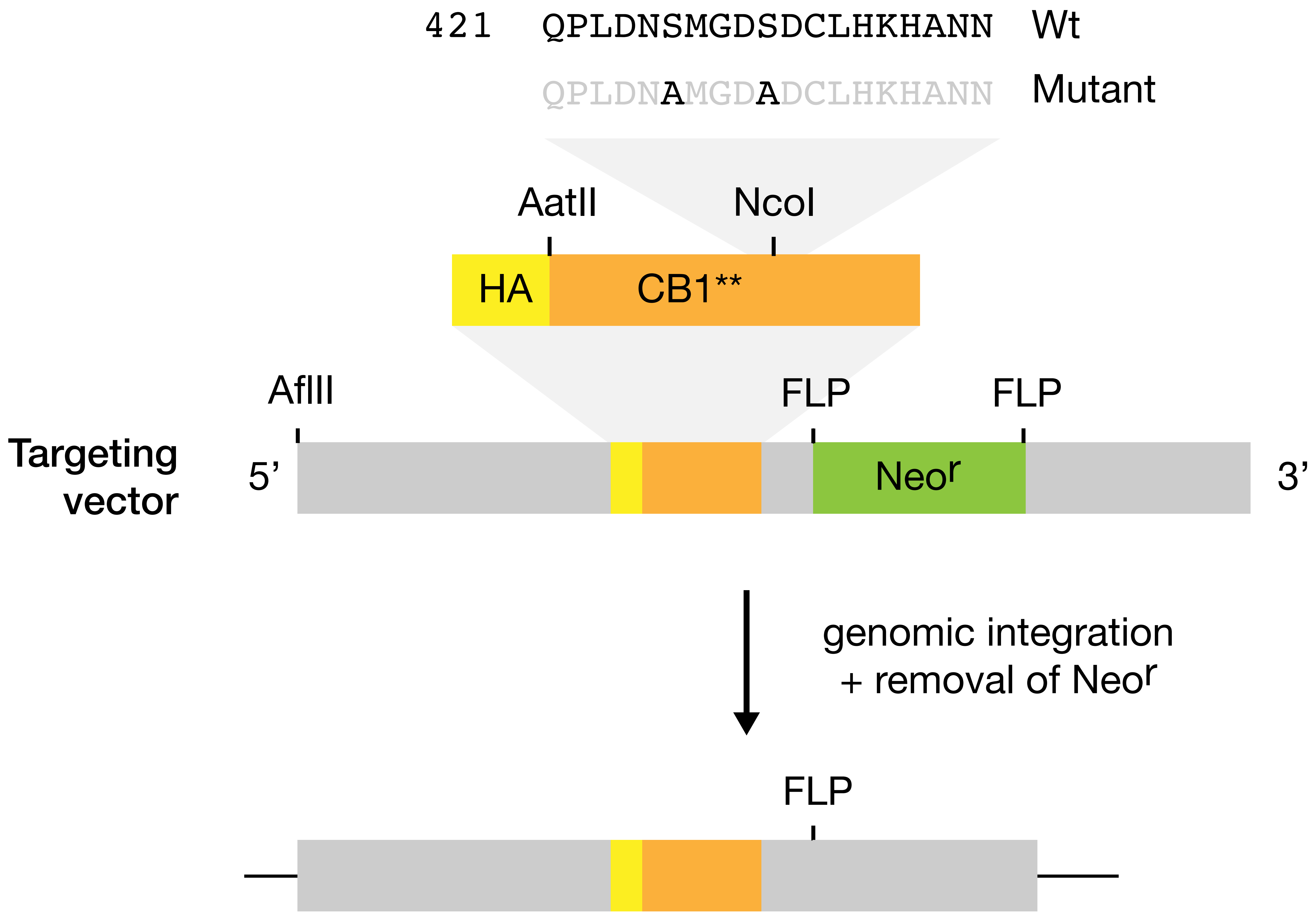
Why is HA thatched? Does it relate in any way to the color of CB1**? Can you think of a way to avoid using textures but rather similar colors to maintain their relationship?



Generation of S426A/S430A knock-in mice. Mice expressing a desensitization-resistant form of CB1R were produced using a targeting vector designed to mutate two putative GRK phosphorylation sites, serines 426 and 430, to nonphosphorylatable alanines. Additionally, the targeting vector introduced an N-terminal HA tag into CB1R and contained a NeoR gene flanked by FLP recombinase sites (blue triangles).

J Neurosci (2014) 34:5152–5163.


A
Targeting vector
Afl II
5' Genomic
CB1**
Neo^r
3' Genomic
Aat II
Nco I
HA
CB1**
421 QPLDNSMGDSDCLHKHANN Wt
421 QPLDNAMGDADCLHKHANN Mutant
Genomic integration after removal of neo^r
HA
CB1**


421 QPLDNSMGDSDCLHKHANN Wt
QPLDNAMGDADCLHKHANN Mutant
AatII
NcoI
HA
CB1**
AflII
FLP
FLP
Targeting vector
5'
Neo^r
3'
genomic integration + removal of Neo^r
FLP

EXERCISE 4

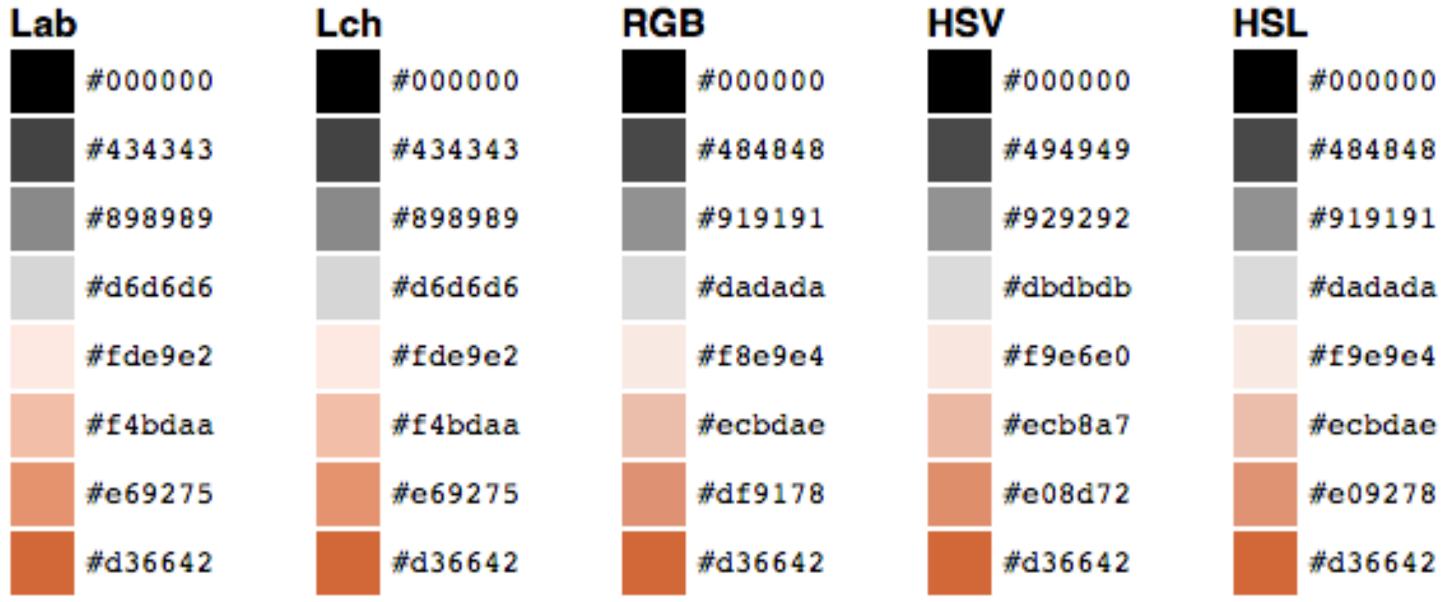
Play around a bit with the LCH color picker

http://davidjohnstone.net/pages/lch-lab-colour-gradient-picker

Make a 9-swatch ramp from pure RGB red to white to pure RGB blue (use “show all” color selection mode). What’s the difference between Lab and RGB interpolations?

Now switch to “Lch” color selection mode. Try to replicate the black-white-red color ramps. What kind of situations might call for such a ramp?

Can you think of cases where red would be more (or less) useful than green?



http://davidjohnstone.net/pages/lch-lab-colour-gradient-picker

EXERCISE 5

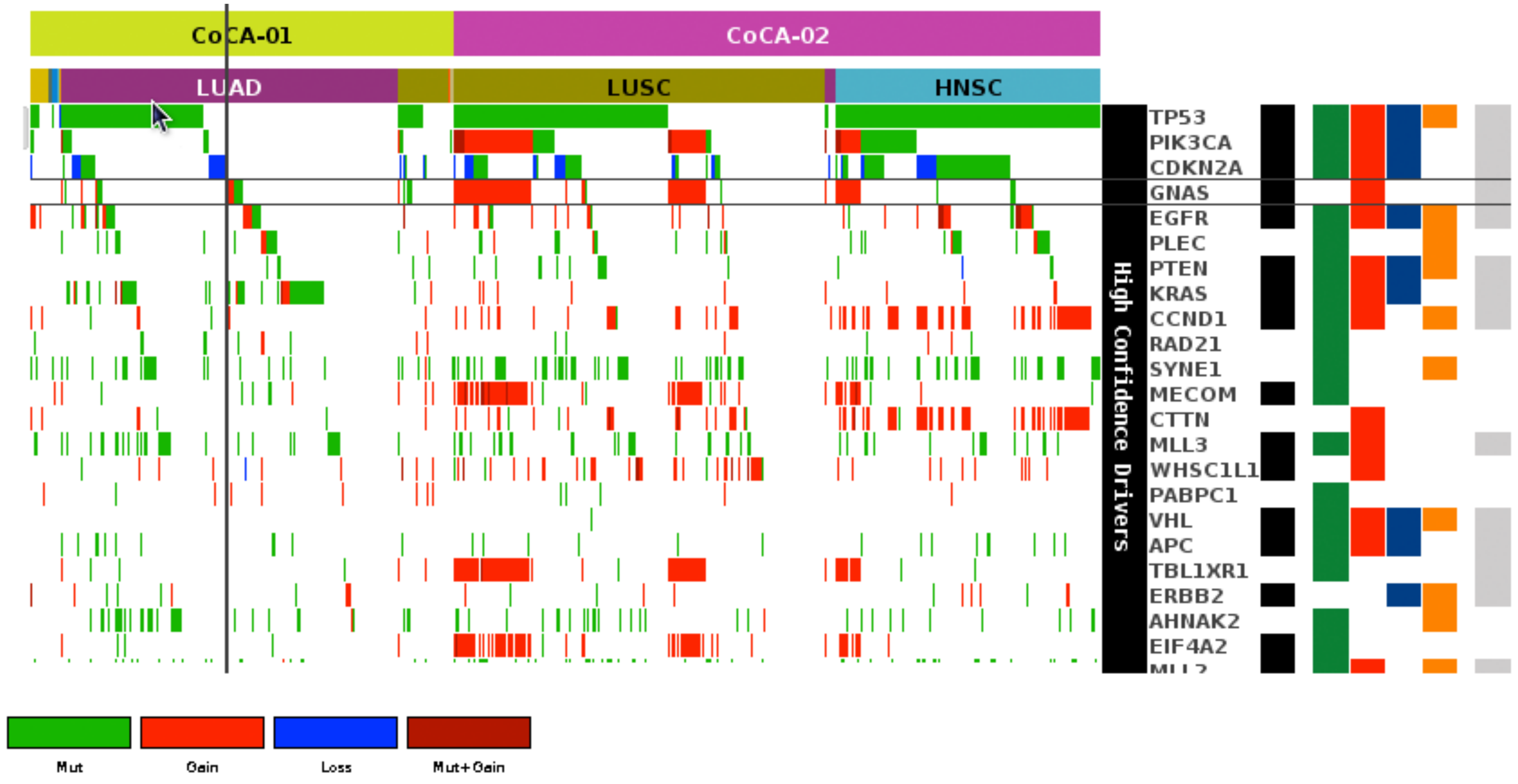
Download and install Color Oracle (colororacle.org).

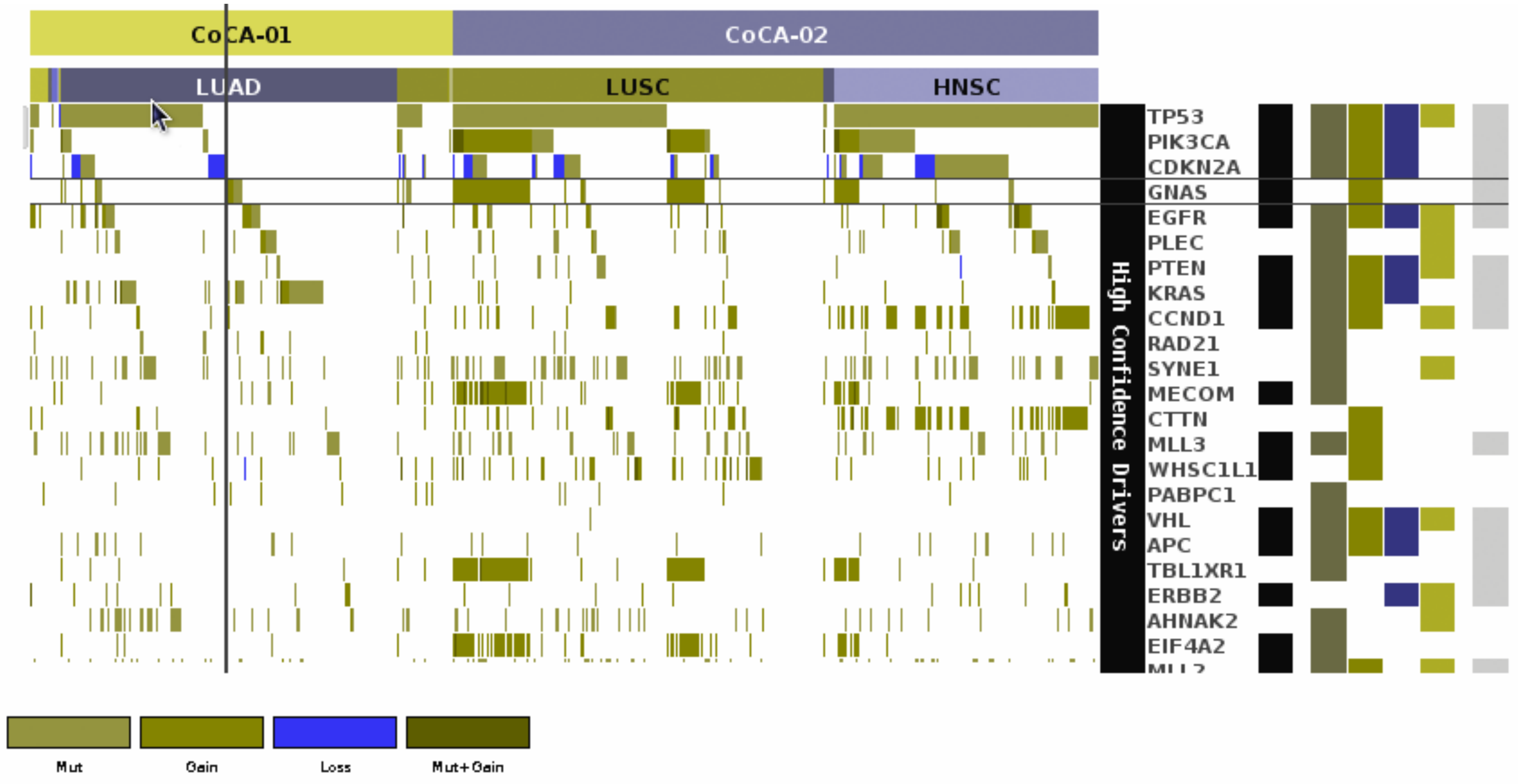
Simulate how the figure shown here appears for someone who is red-green color blind?

If color blindness affects about 8% of men, how many men do you need in your audience for at least a 50% chance that at least one person in the audience is color blind?

Visit Color Brewer (colorbrewer.org).

Look at the diverging palettes. Which are color blind safe? Use Color Oracle to see what these would look like to someone with red-green color blindness.



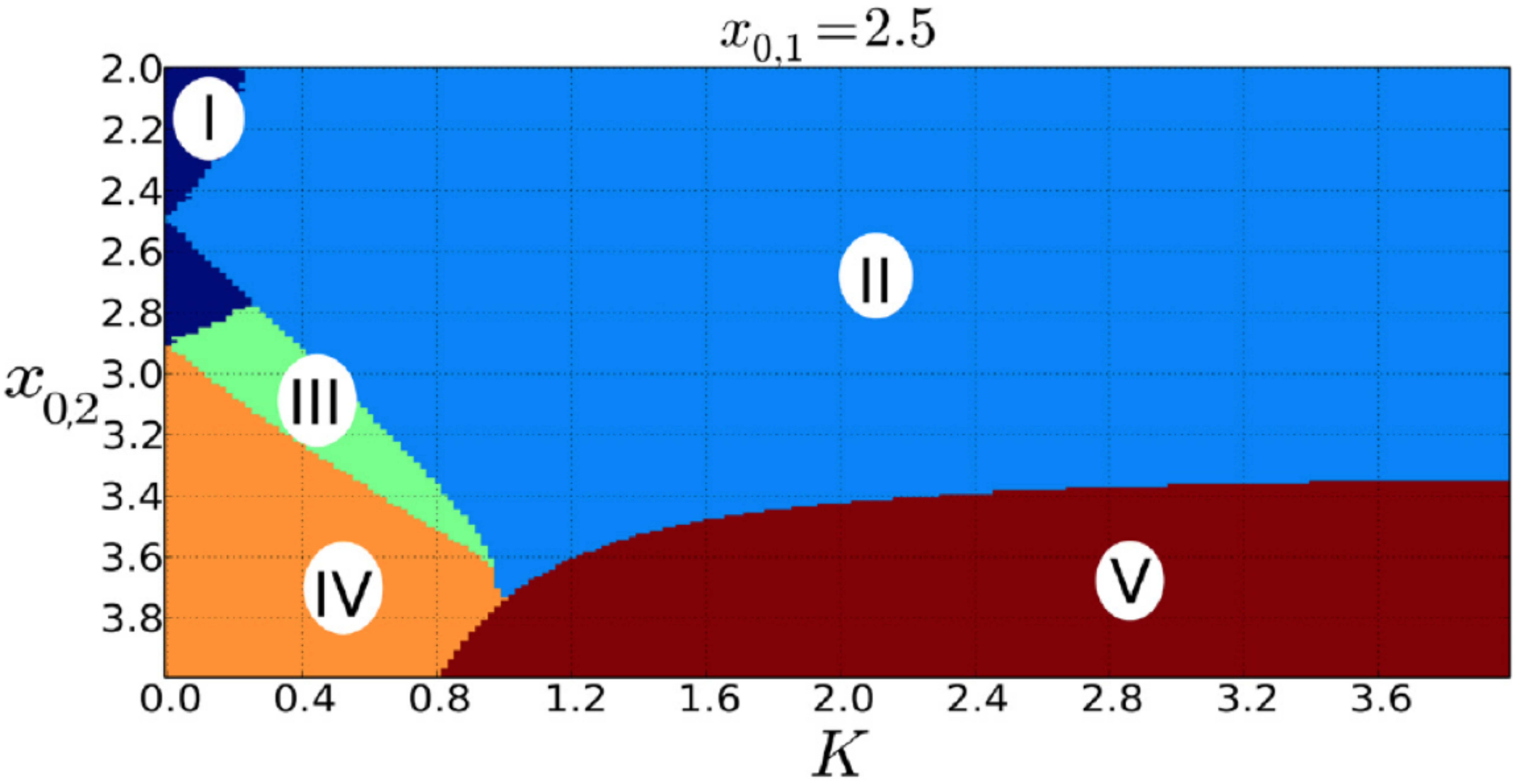


http://www.gitools.org/img/pancancer12/pancancer-drivers-gitoolsweb.screenshot.png

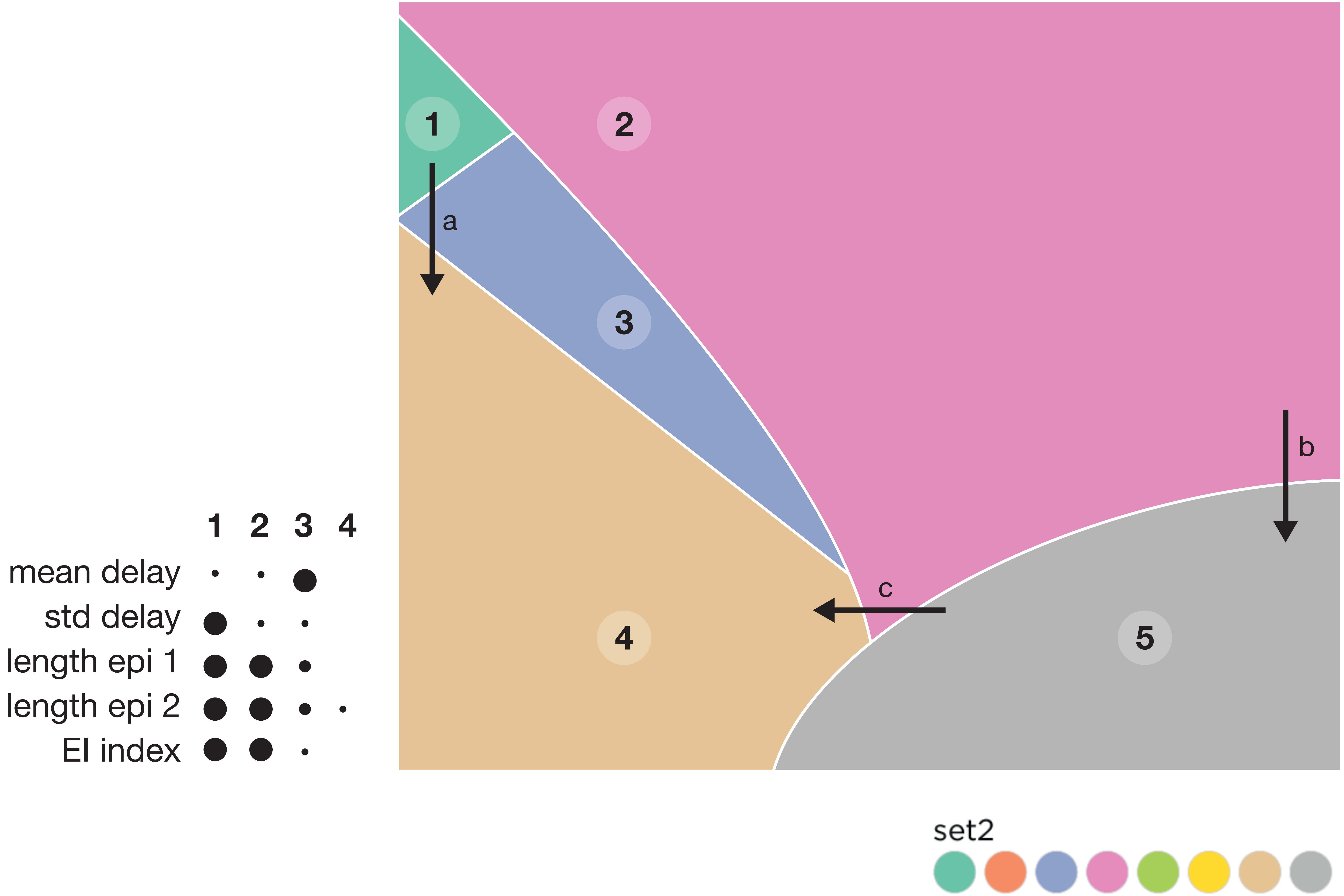
EXERCISE 6

Select better colors for this image from the Brewer palettes.

Which palette type is appropriate?



J Neurosci (2014) 34:15009–15021.


1
2
3
4
5
a
b
c
1 2 3 4
mean delay
std delay
length epi 1
length epi 2
EI index
set2

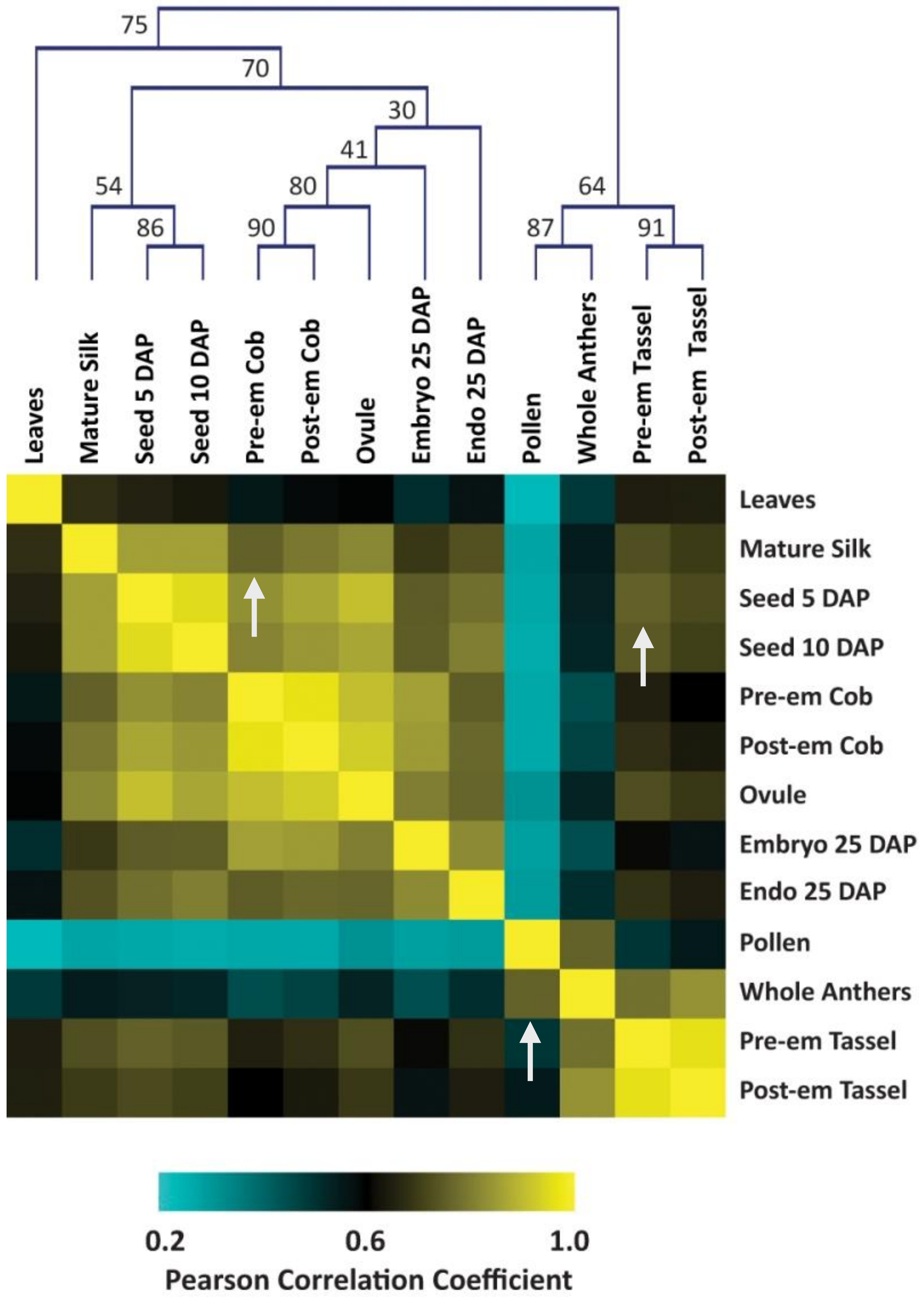
EXERCISE 7

The luminance effect changes the way a color (or tone) is perceived based on its background or neighbouring elements.

A dark background will make tones appear lighter. A light background will make them appear darker.

Convince yourself that the three squares highlighted with the white arrows have the same color (RGB approximately 100, 100, 40).

Which do you perceive to be the brightest. The darkest?



Plant Gen (2011) 4:191–203.