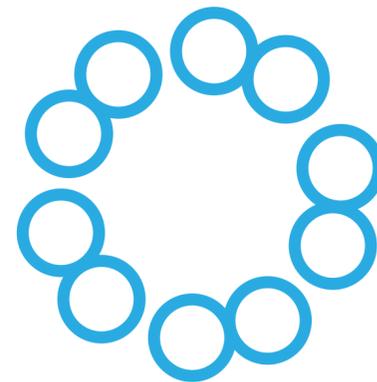


THE UNIVERSITY OF  
**SYDNEY**



CANADA'S MICHAEL SMITH  
**GENOME  
SCIENCES**  
CENTRE



# ESSENTIALS OF DATA VISUALIZATION

THINKING ABOUT DRAWING DATA + COMMUNICATING SCIENCE

# DESIGN

clarify and organize

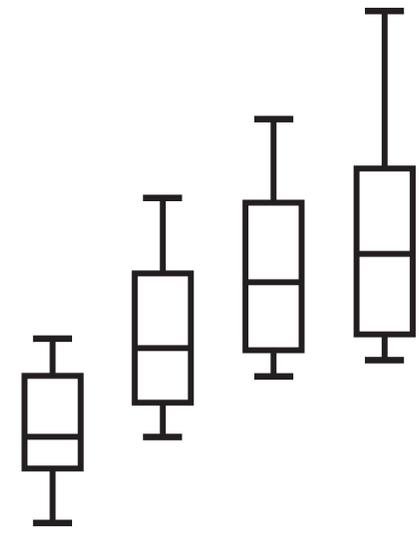
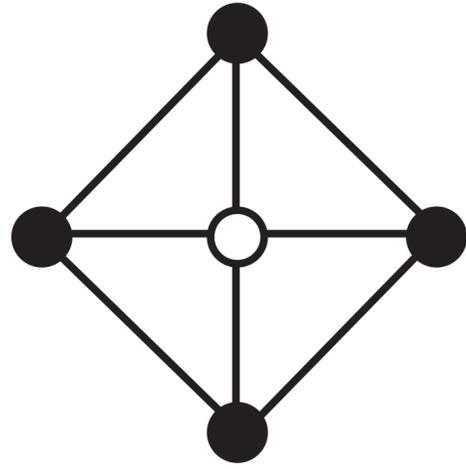
I've already mentioned the idea that design plays a role in data visualization.

Think of design as choreography for the page. In our context it's not merely driven by aesthetic, but function. Although there's always room for aesthetic—gently applied—and I really encourage you to find your own and continue to refine it.

But always remember, be understood before being articulate. Be legible before being attractive!

Your goal here isn't to make inroads on the global stage of aesthetic studies.

Become a good visual explainer. It's harder ... and more worth doing.

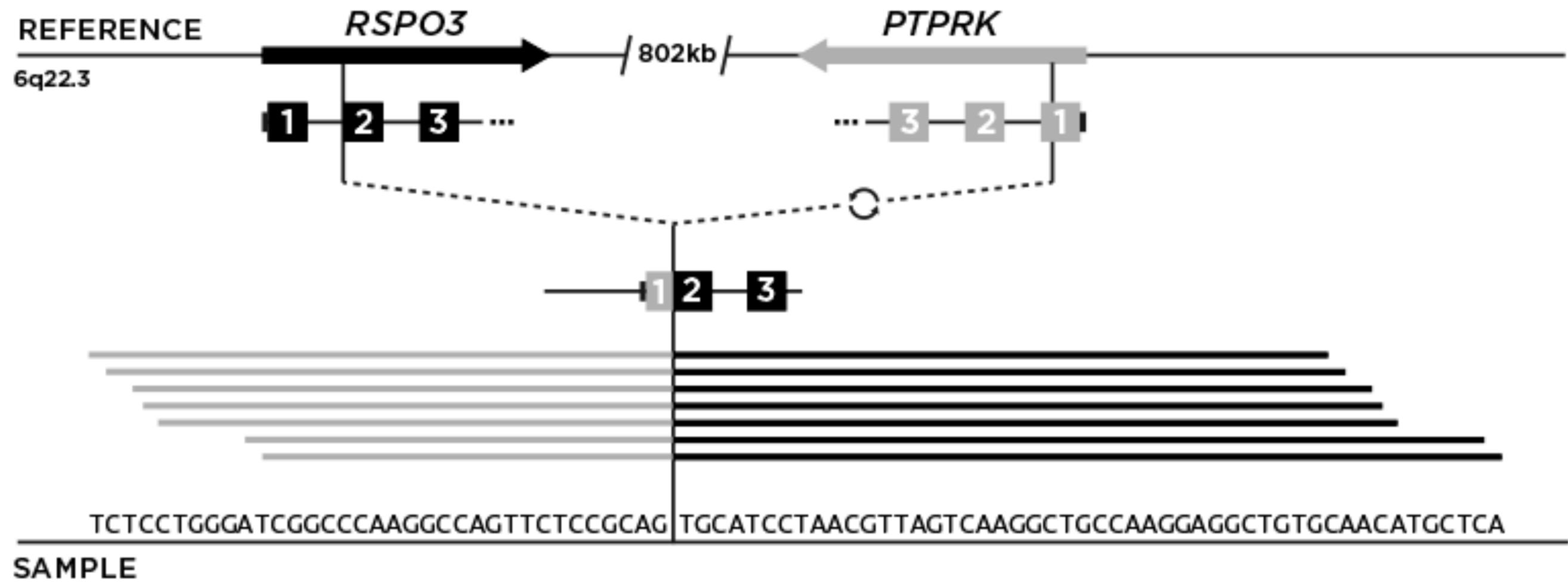


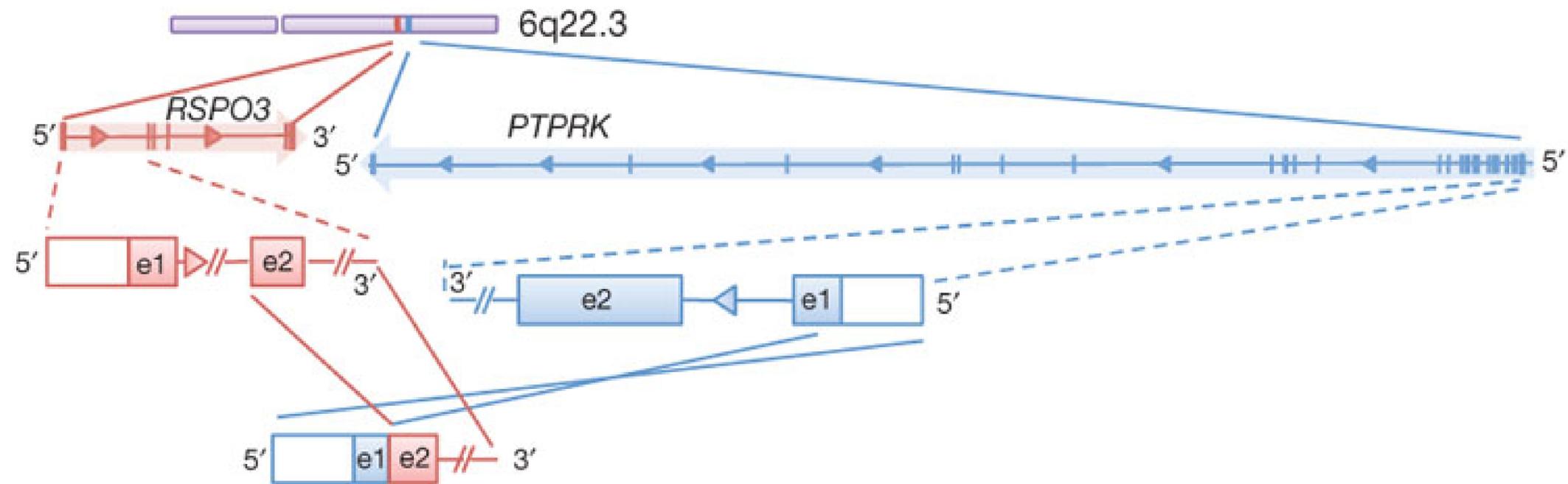
Once you have an encoding and your plot—sometimes you don't even have a plot, such as the case may be for concept or method figures—you need some way to organize the information on the page. That's the design part.

To communicate concepts and organize information, use the design force.

This is particularly important for multi-panel figures that provide the information incrementally to unfold a story.

You also need to choose the size of elements, line weights, arrow head size, colors all the while making sure everything is tightly presented so that, as much as possible, variation is due to data and relevant ideas and not merely sloppy alignment and inconsistent formatting.





TCTCCTGGGATCGGCCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGC

CTCCTGGGATCGGCCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCA

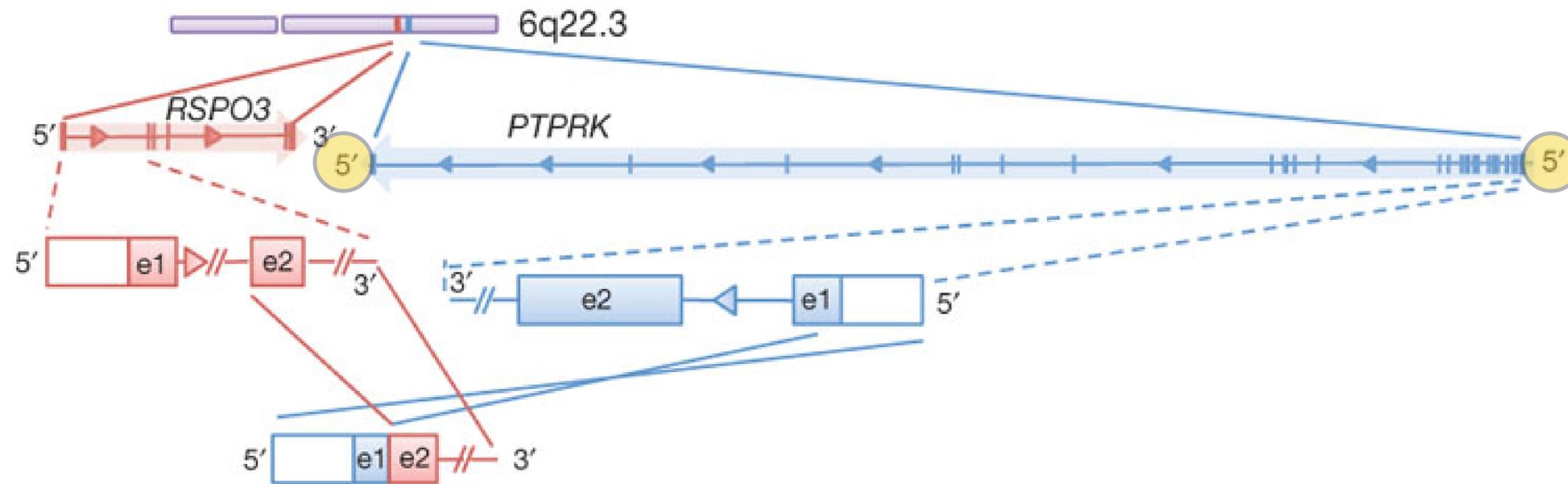
TCCTGGGATCGGCCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAA

CCTGGGATCGGCCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAAC

CTGGGATCGGCCCAAGGACAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACA

ATCGGCCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTC

TCGGGCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTCA



**TCTCCTGGGATCGGCCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGC**  
**CTCCTGGGATCGGCCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCA**  
**TCCTGGGATCGGCCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAA**  
**CCTGGGATCGGCCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAAC**  
**CTGGGATCGGCCCAAGGACAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACA**  
**ATCGGCCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTC**  
**TCGGGCCCAAGGCCAGTTCTCCGCAG | TGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTCA**

top-down

bottom-up

consistency

data encoding

conciseness

INFORMS

symbols



clarity

color

focus & emphasis



typeface

SATISFIES

salience & relevance

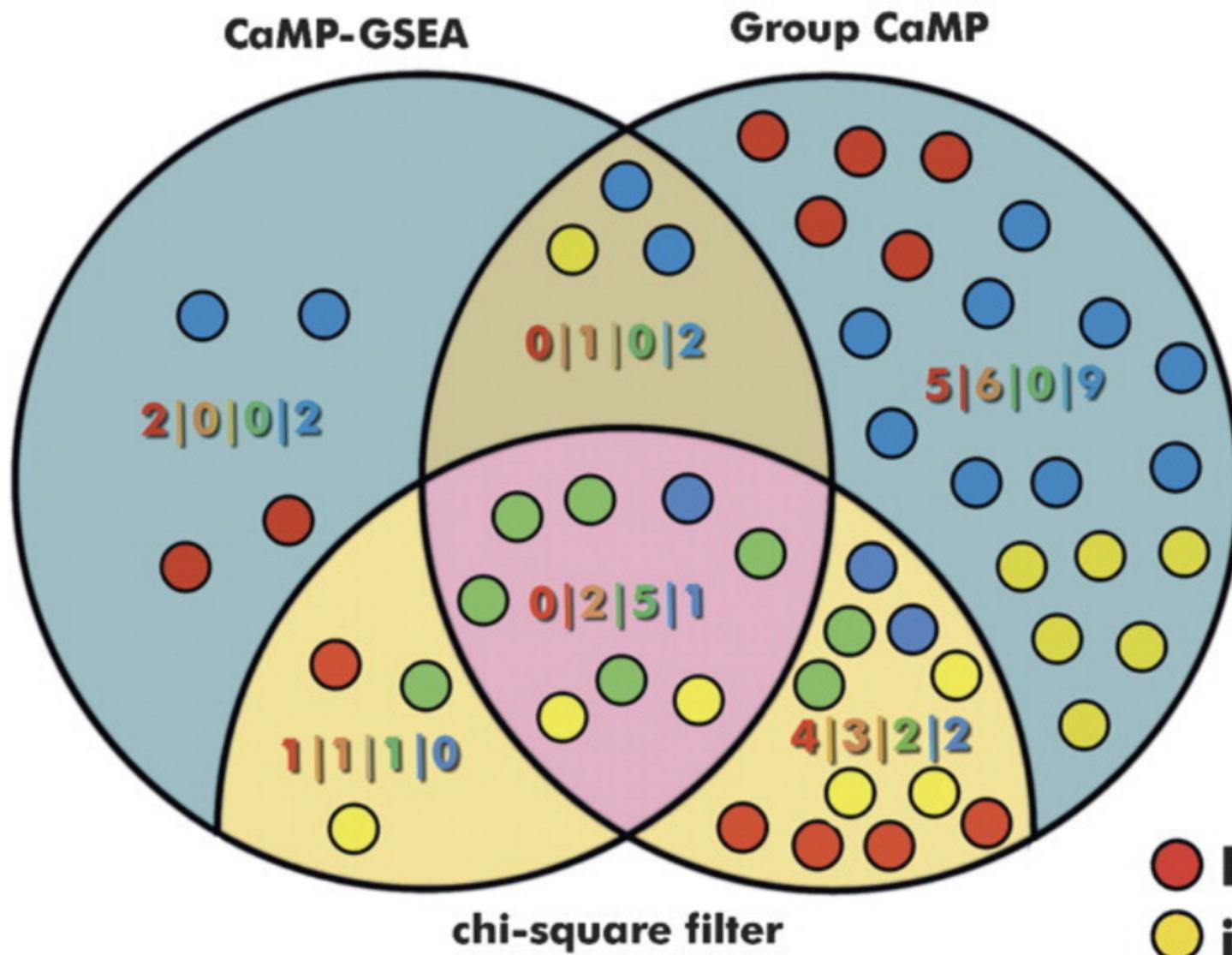
arrows

accuracy & detail

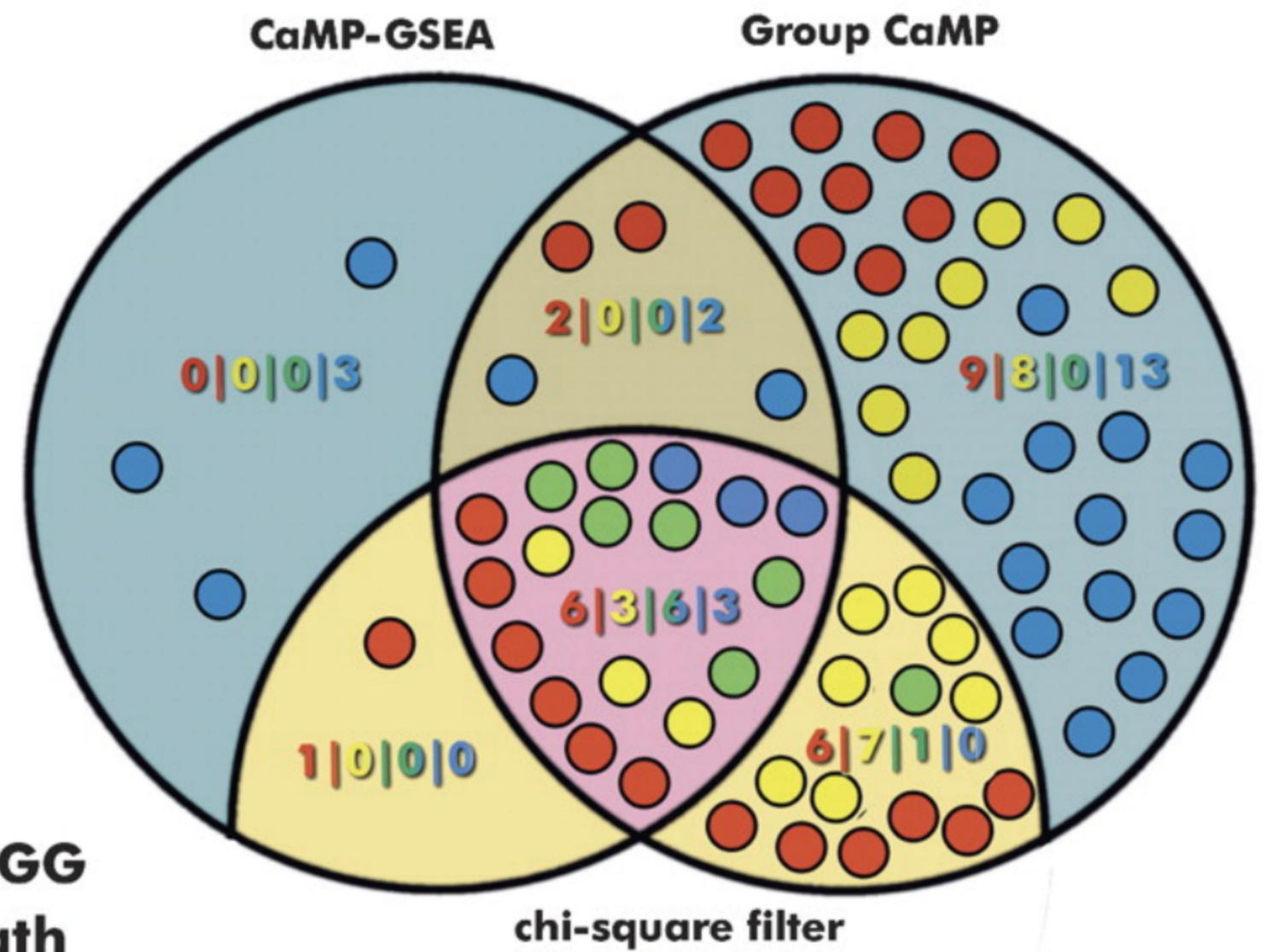
line weight

alignment

## Breast Cancer



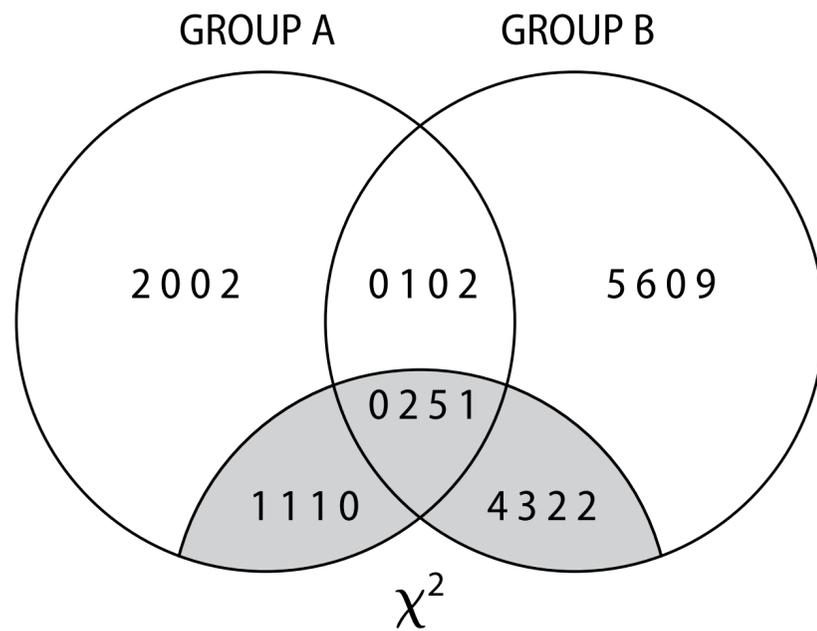
## Colorectal Cancer



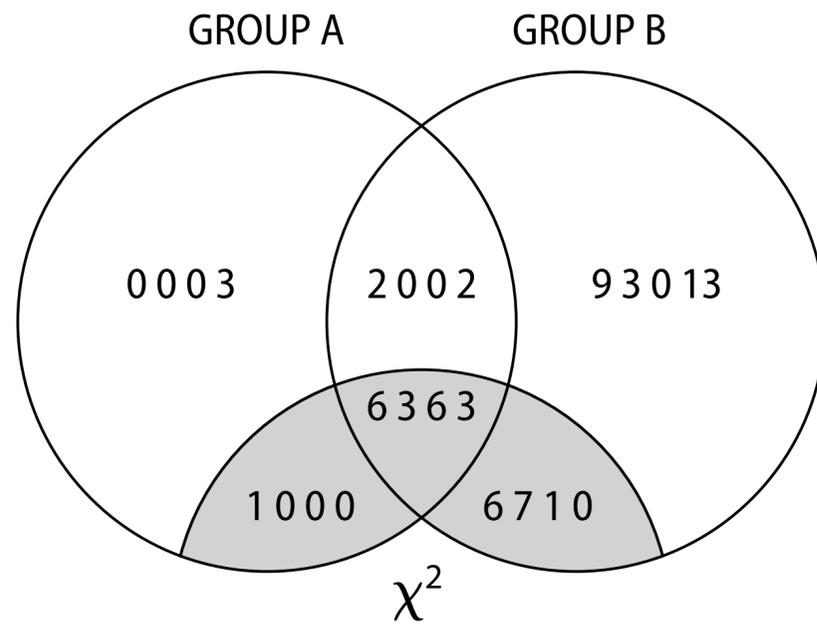
- **KEGG**
- **iPath**
- **BioCarta**
- **sigPathway**

A

### DISEASE 1

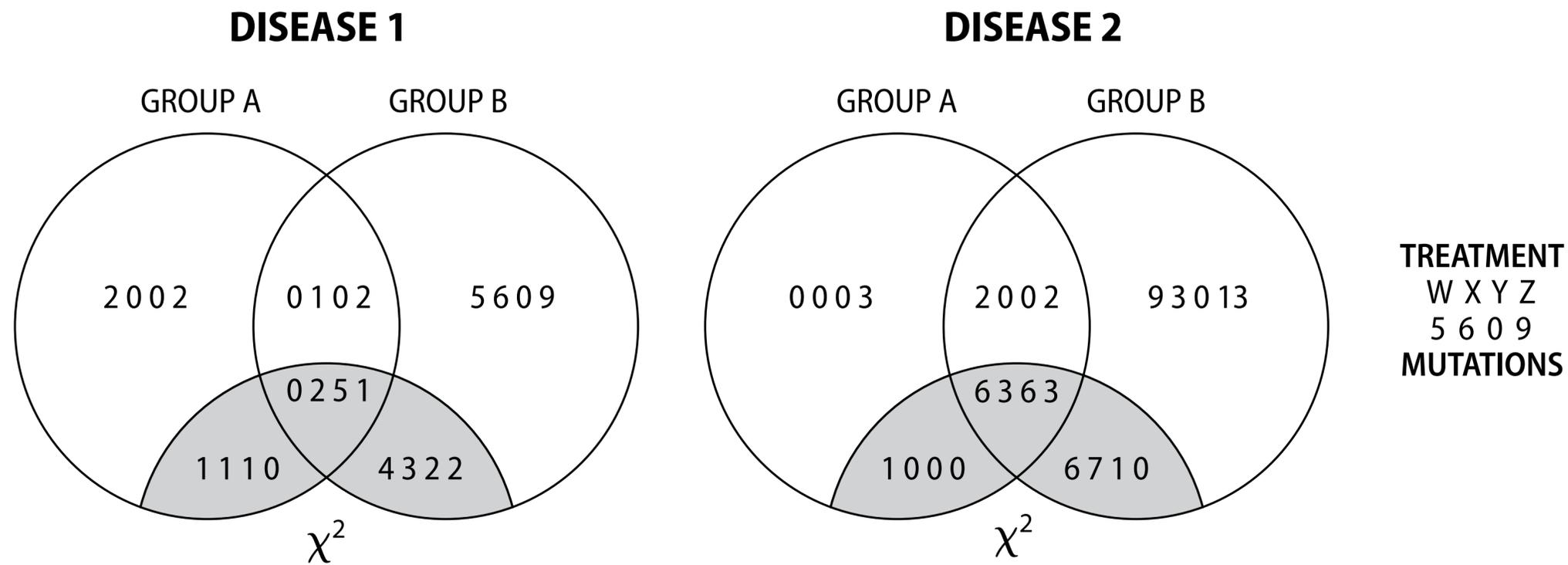


### DISEASE 2

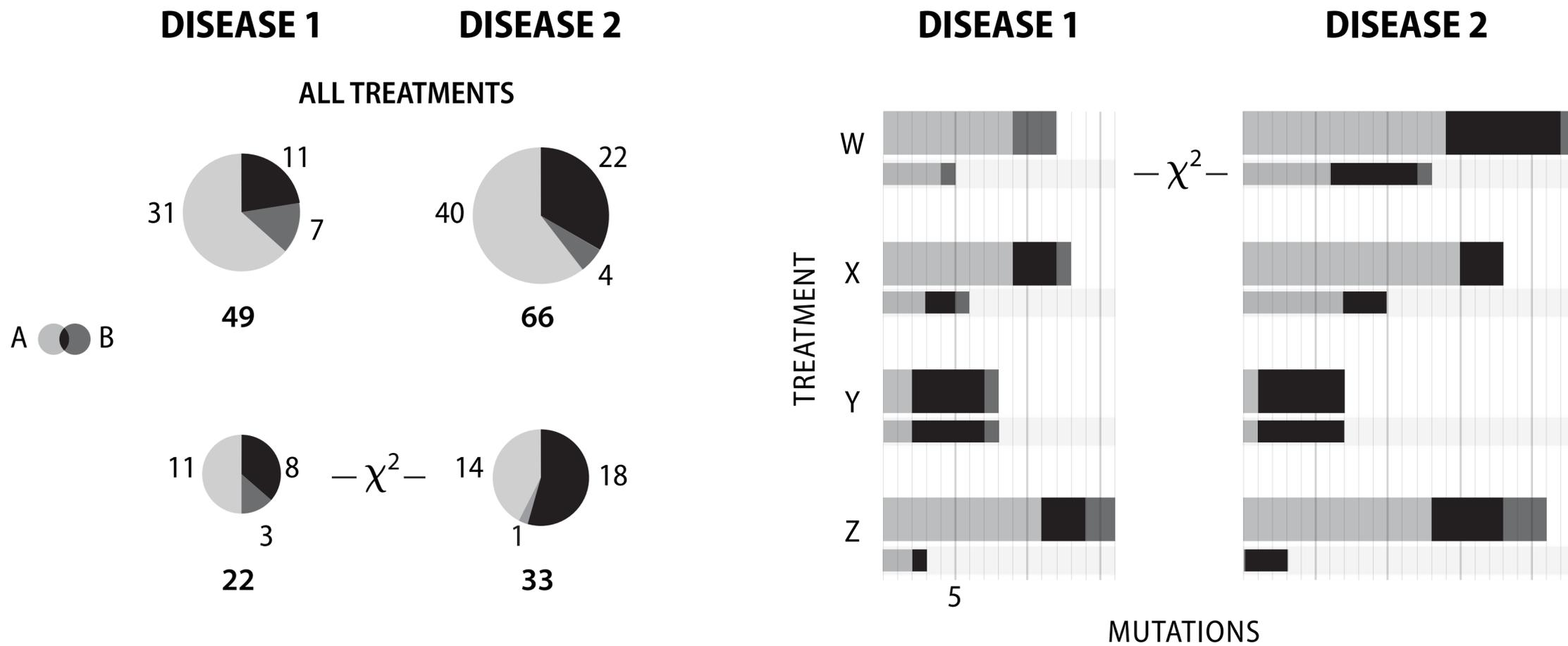


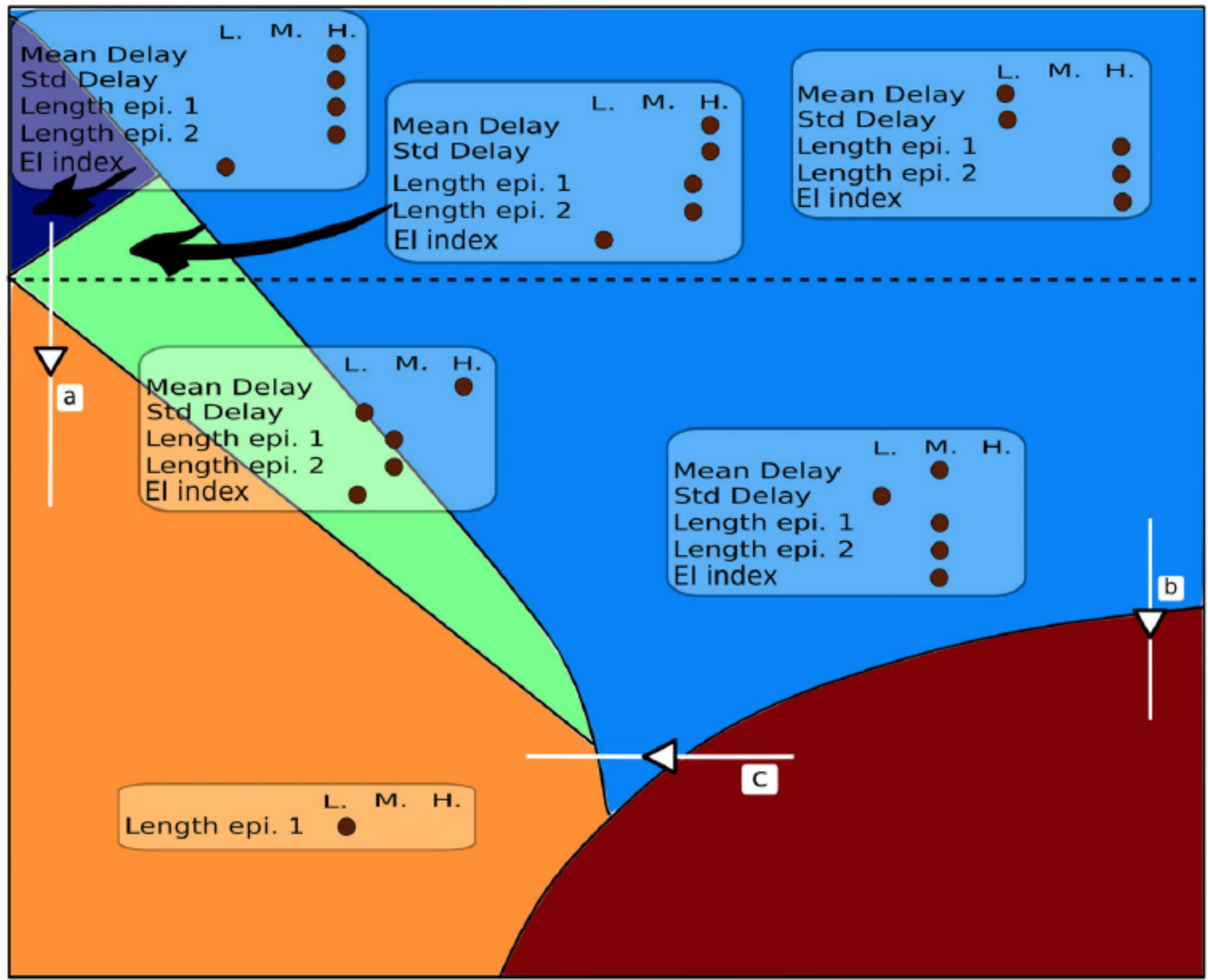
TREATMENT  
W X Y Z  
5 6 0 9  
MUTATIONS

A

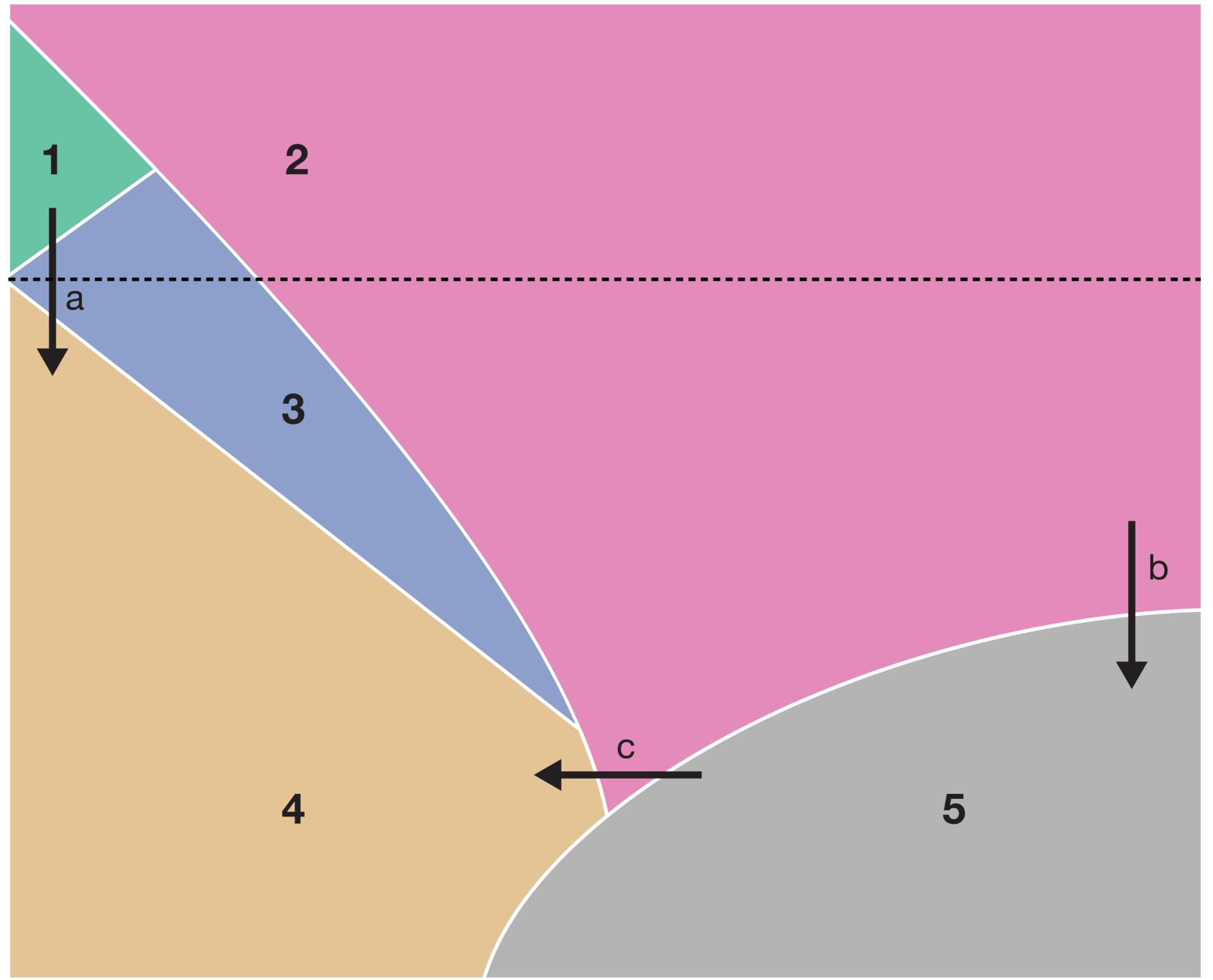


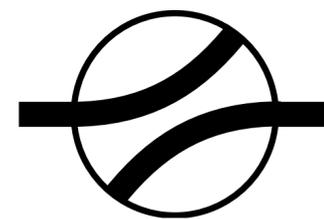
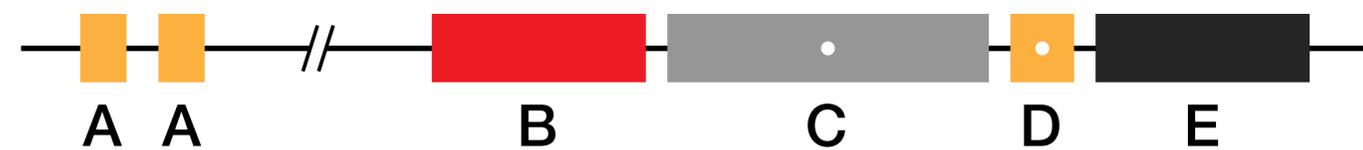
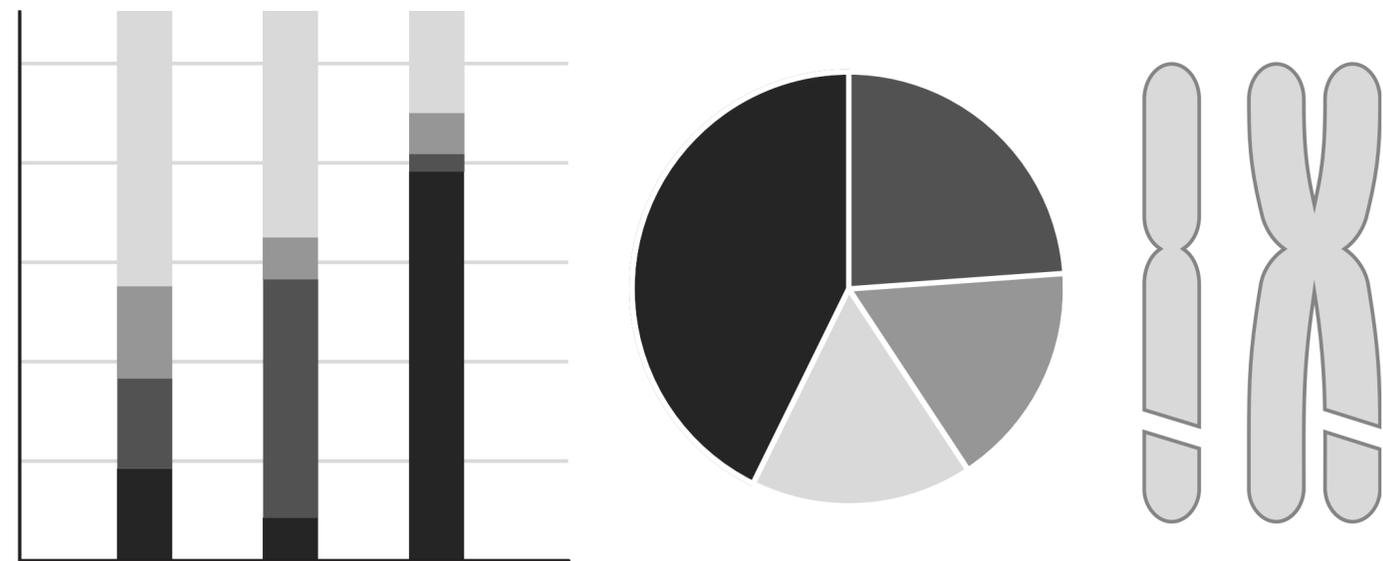
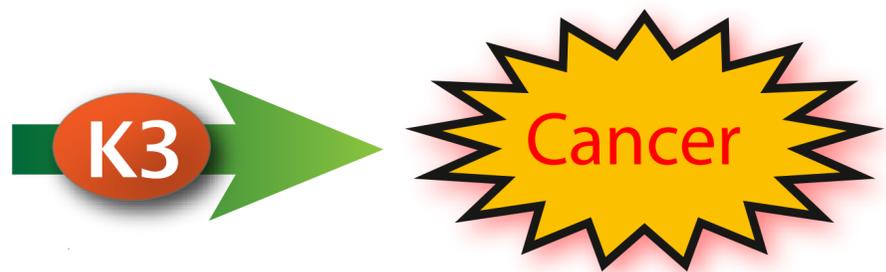
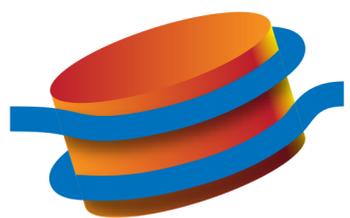
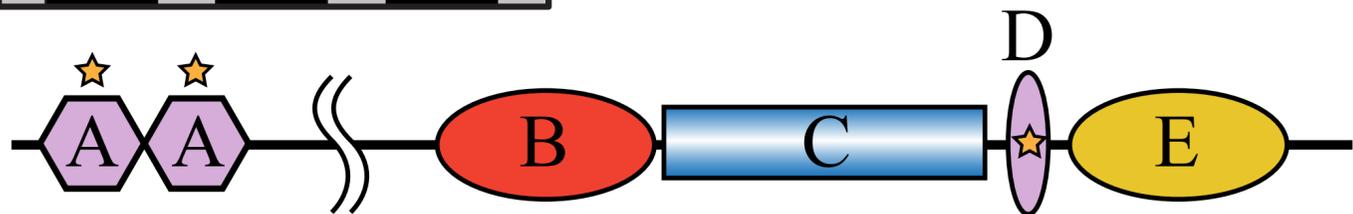
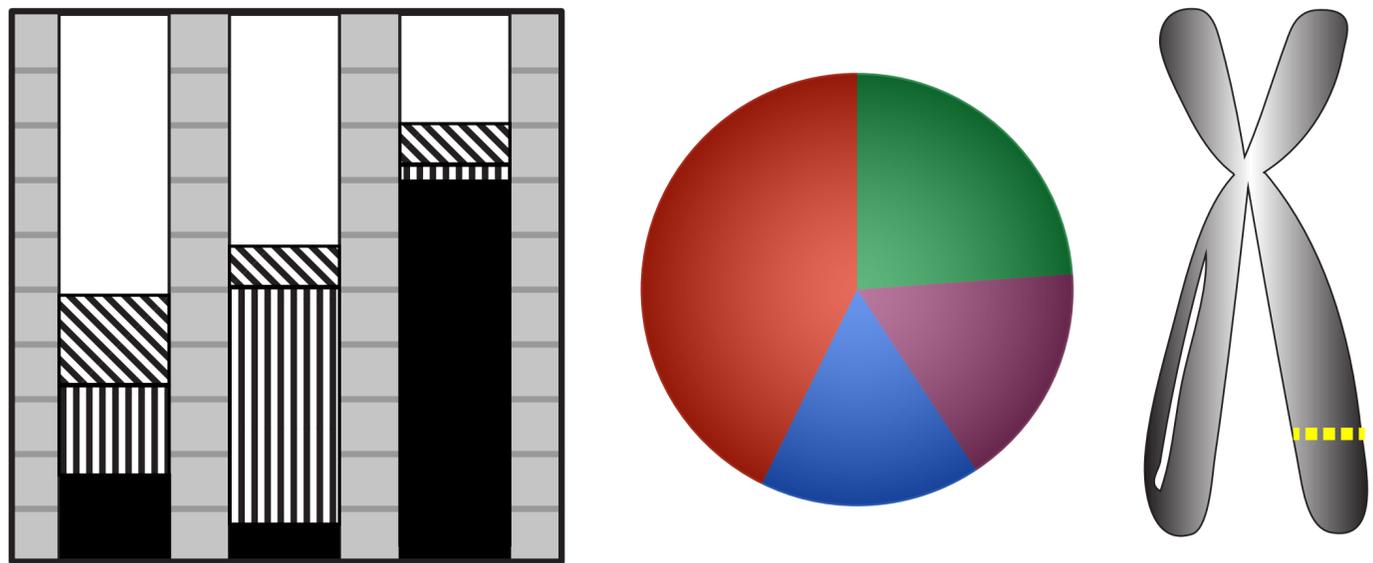
B



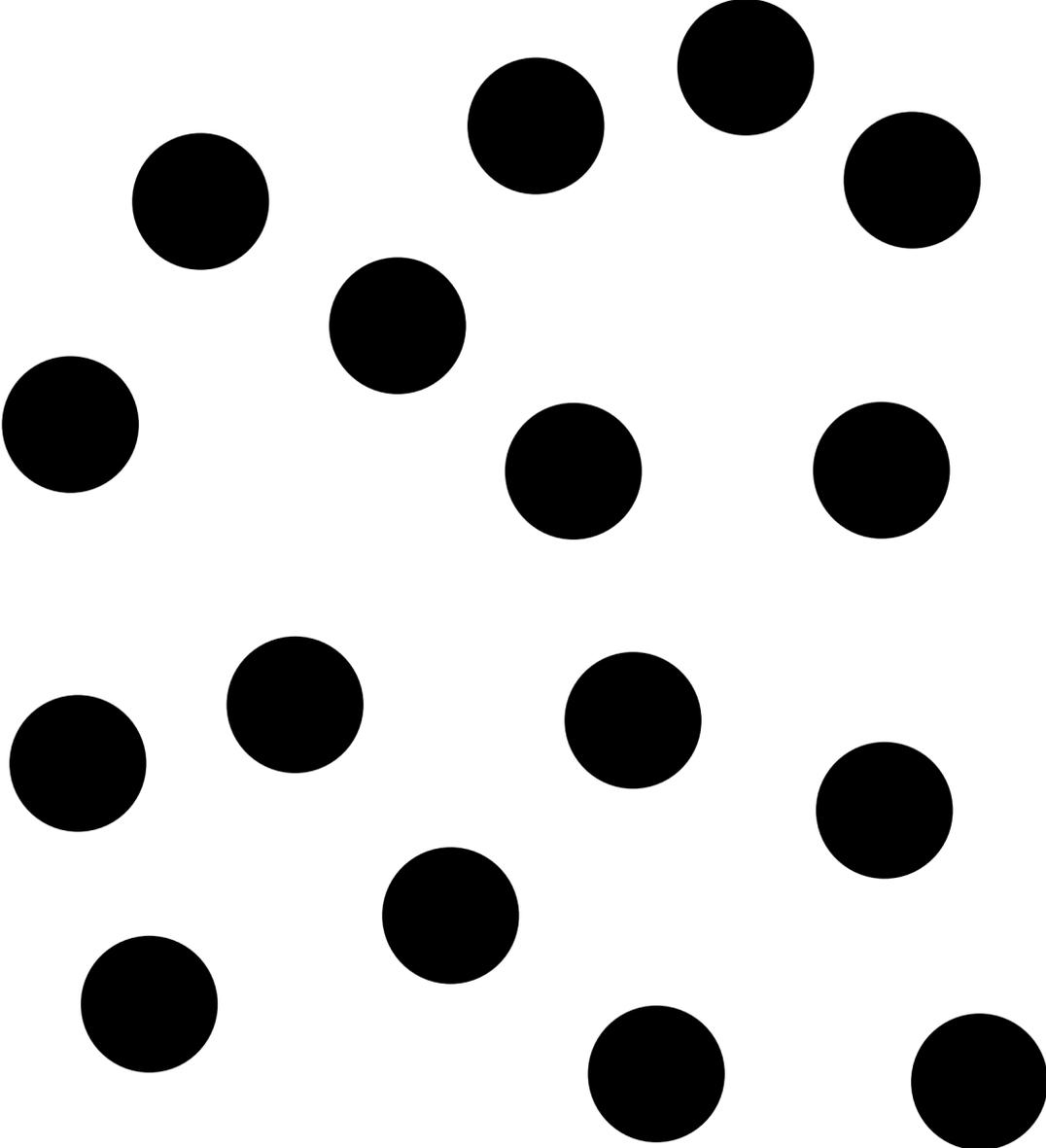


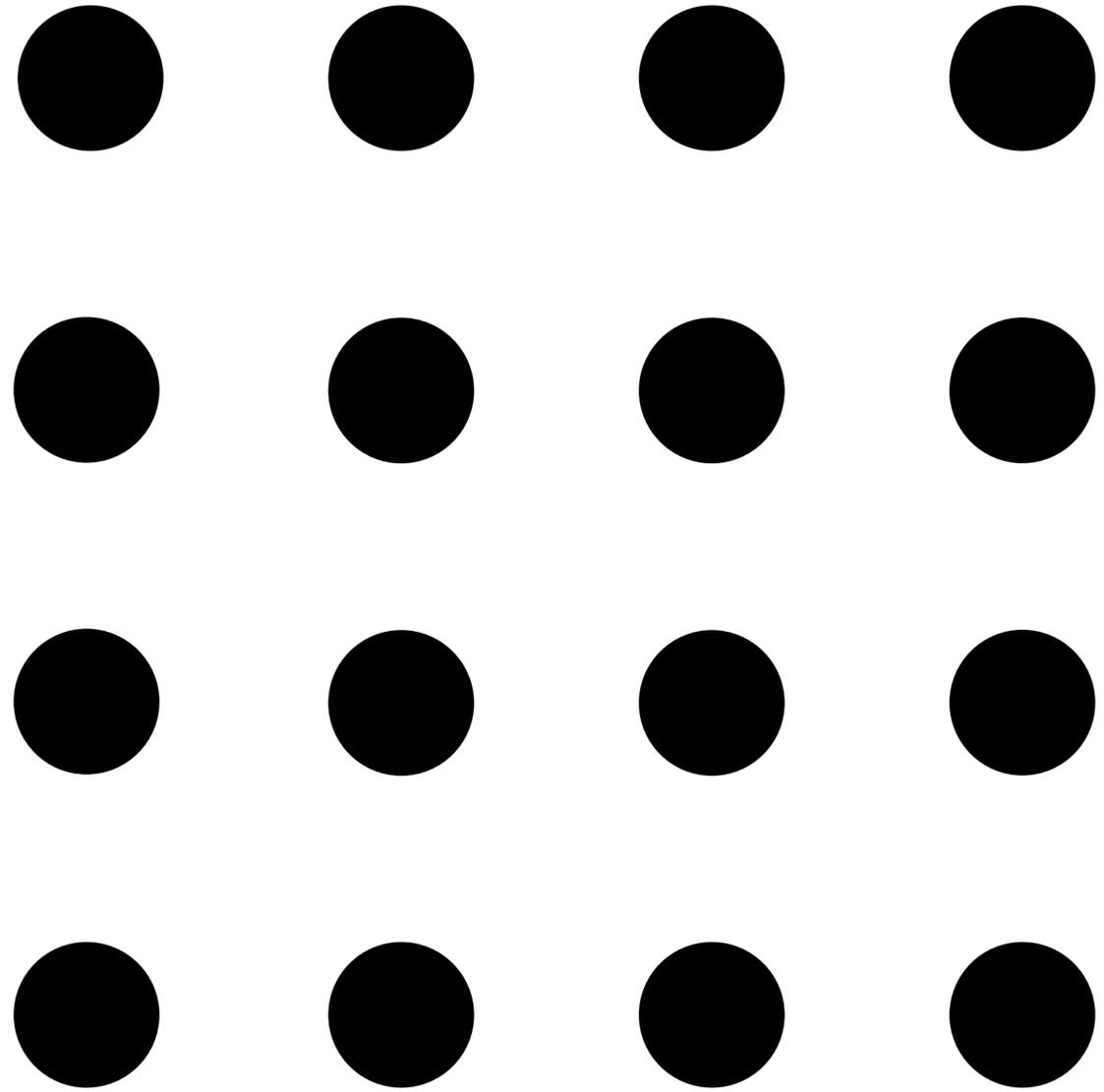
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
mean delay	●	•   •	●   ●	
std delay	●	•   •	●   •	
length epi 1	●	●   •	•   •	•
length epi 2	●	●   •	•   •	
EI index	•	●   •	•   •	





**CANCER**





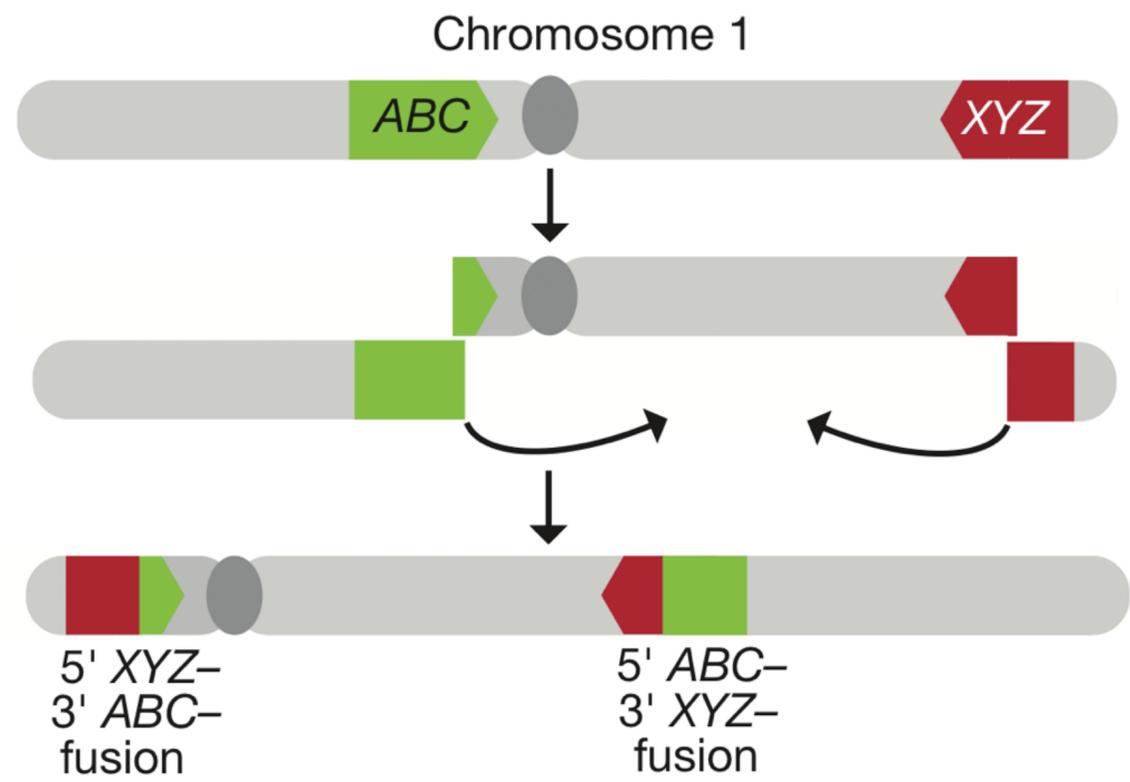
**important! important! important!**  
**important! important! important!**

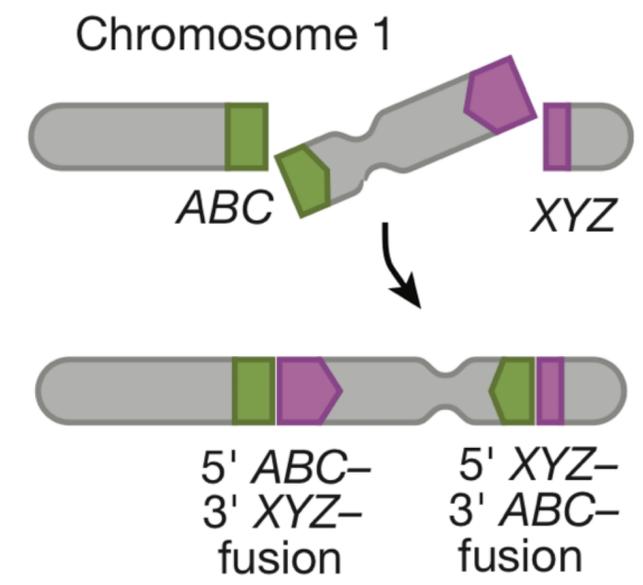
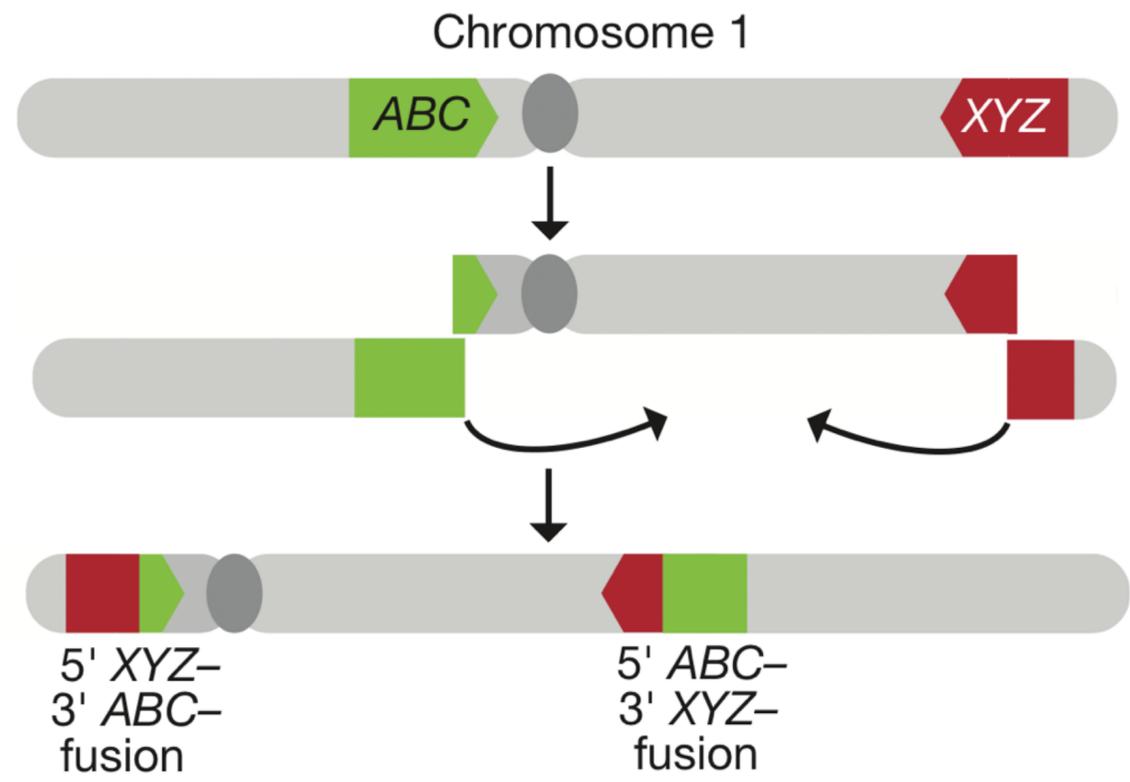
look here

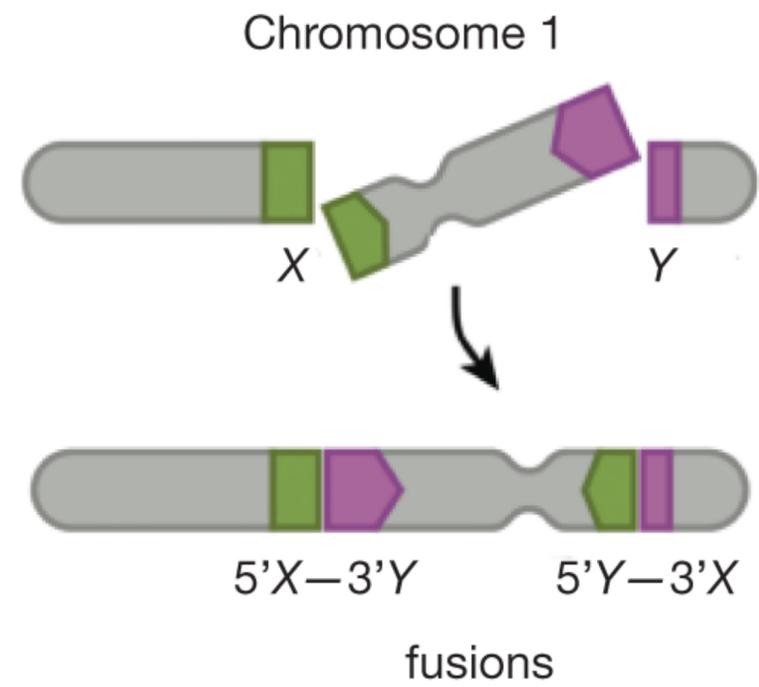
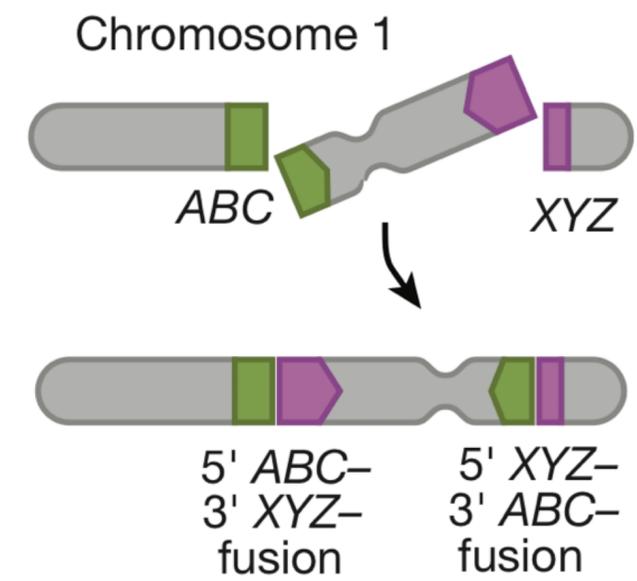
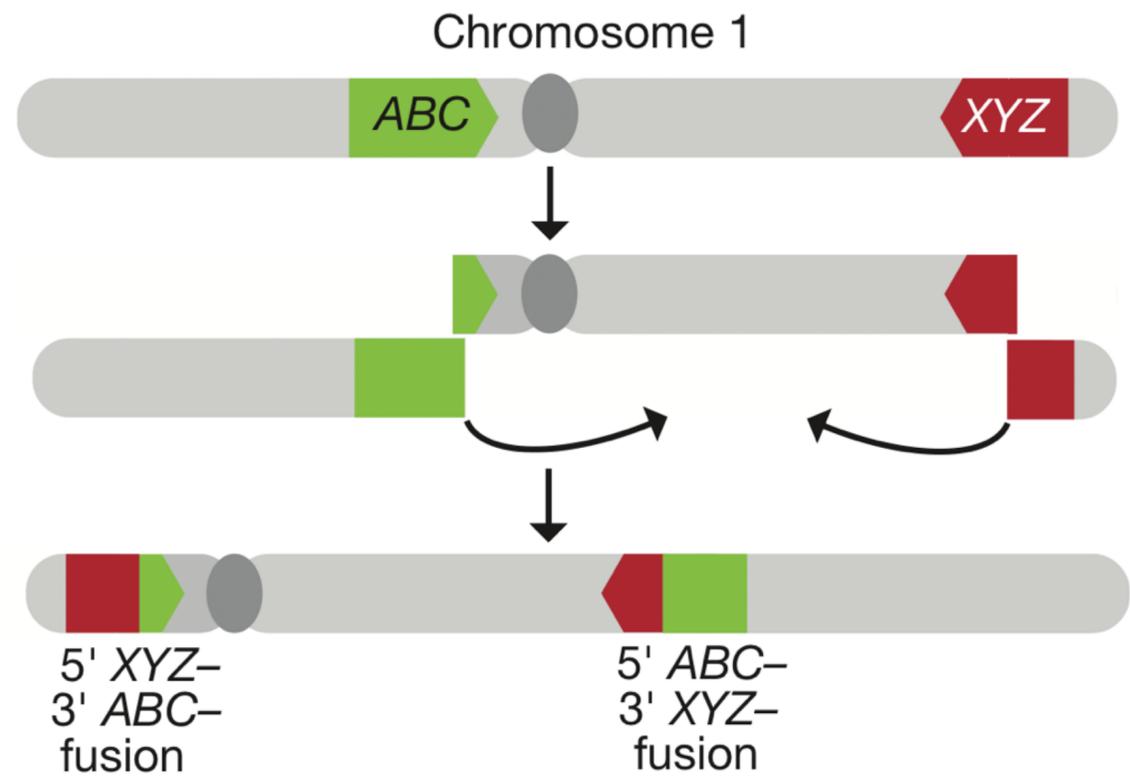
Spa ce mak es gro ups.

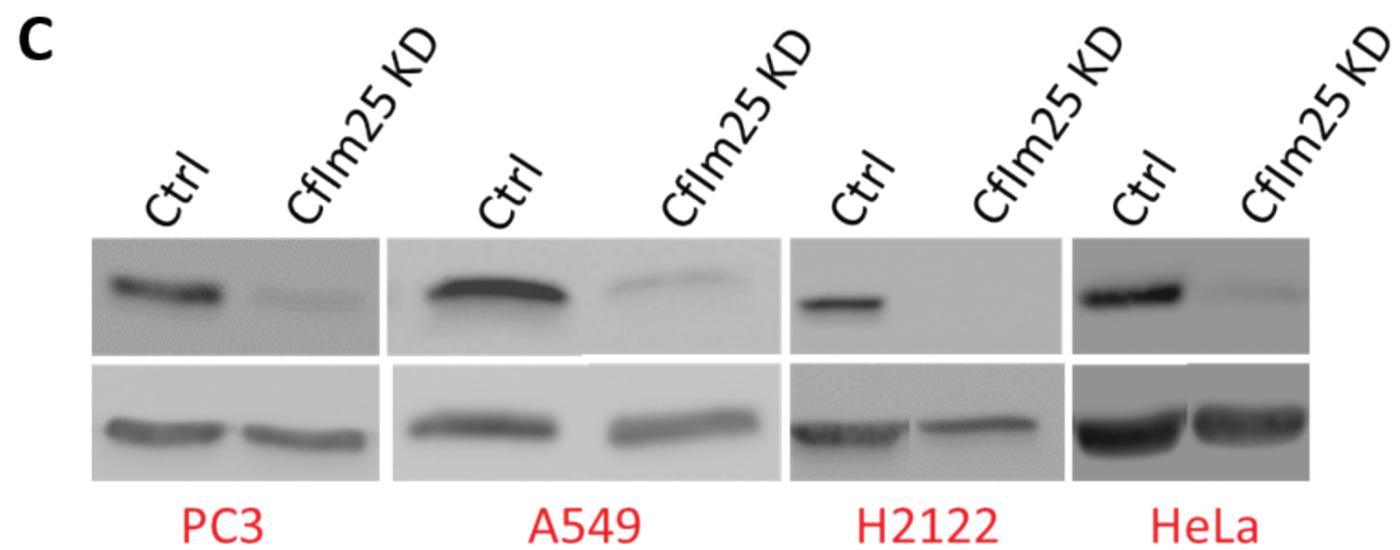
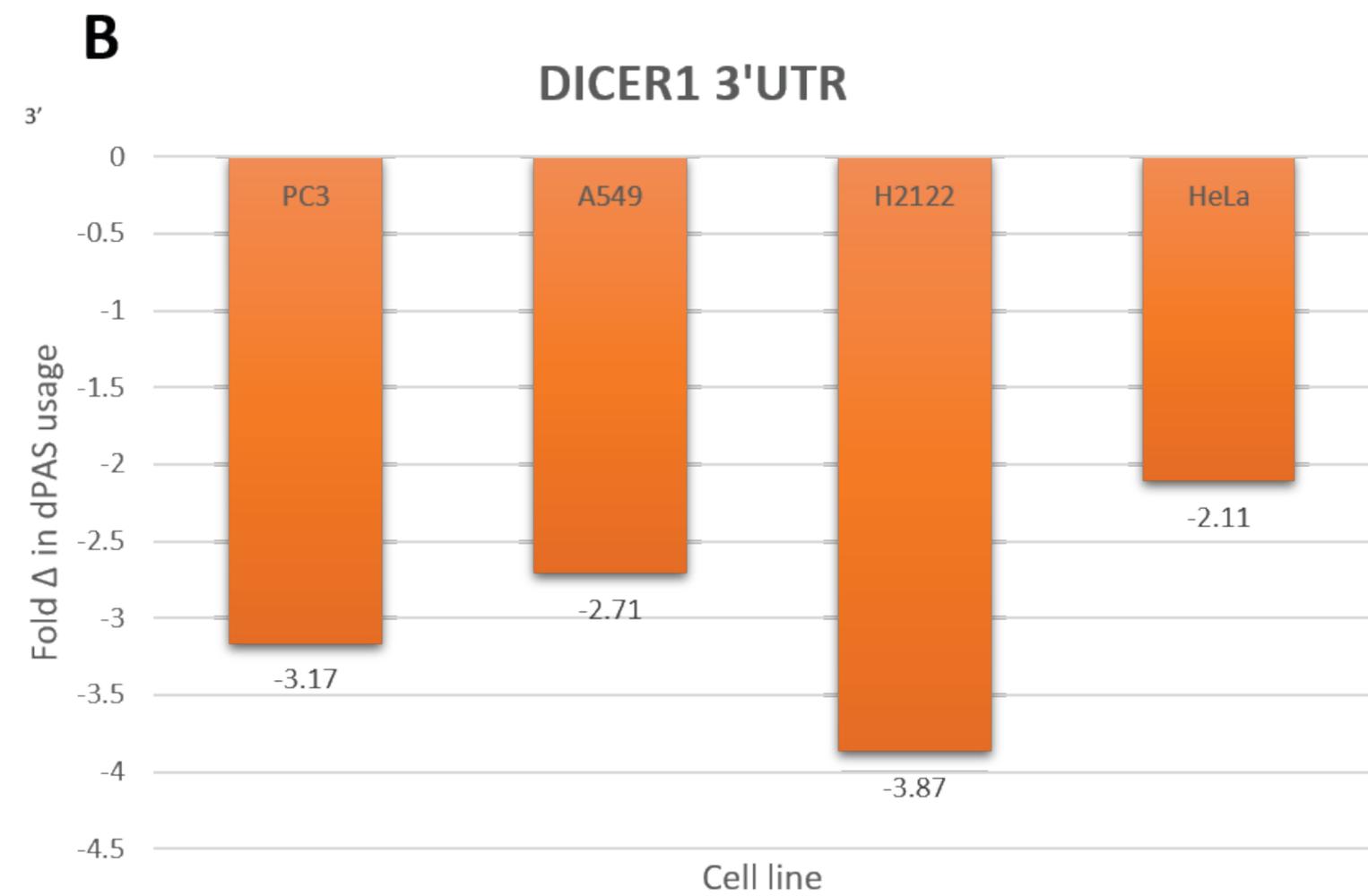
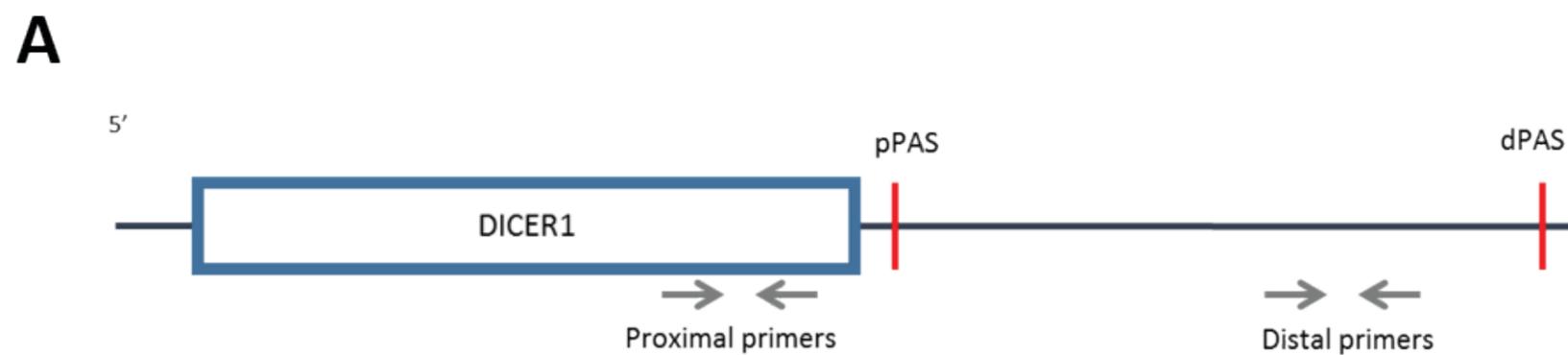
Remember, data visualization isn't just drawing numbers. It relates to communicating all aspects of the scientific endeavor including methods and concepts. Often methods and data are combined and it's imperative to clearly depict the relationship between them.

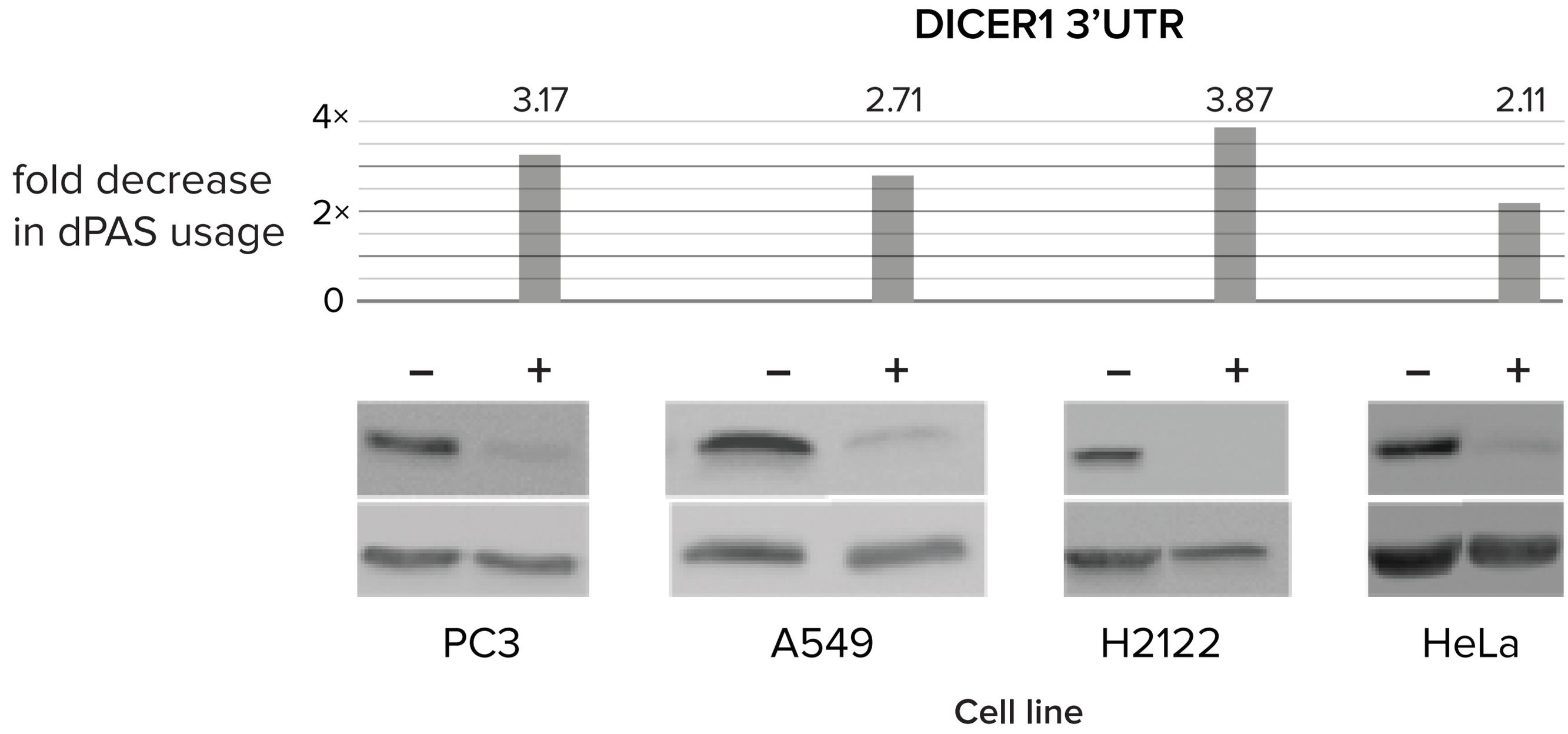
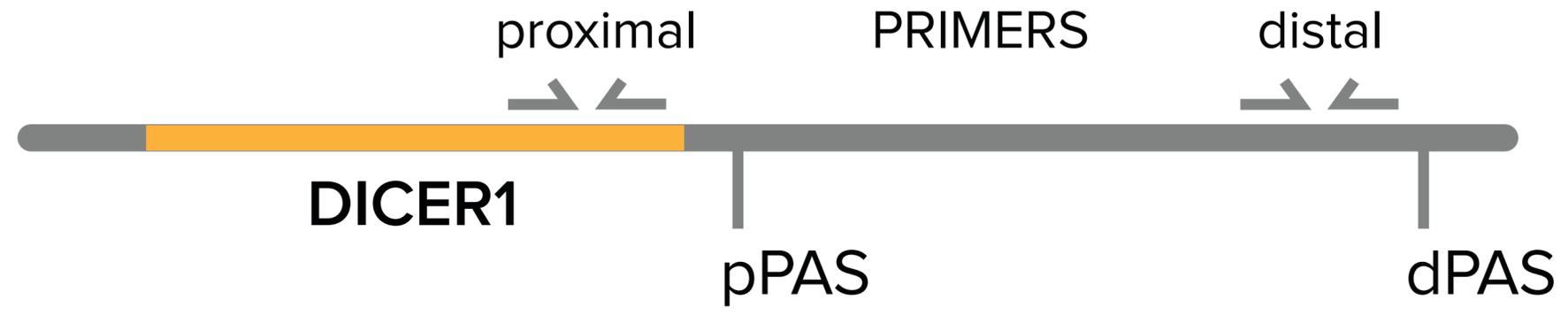
When approaching the design of a concept figure, you must understand the essence of the concept. Then, you must depict this with as few elements as possible, without sacrificing clarity.

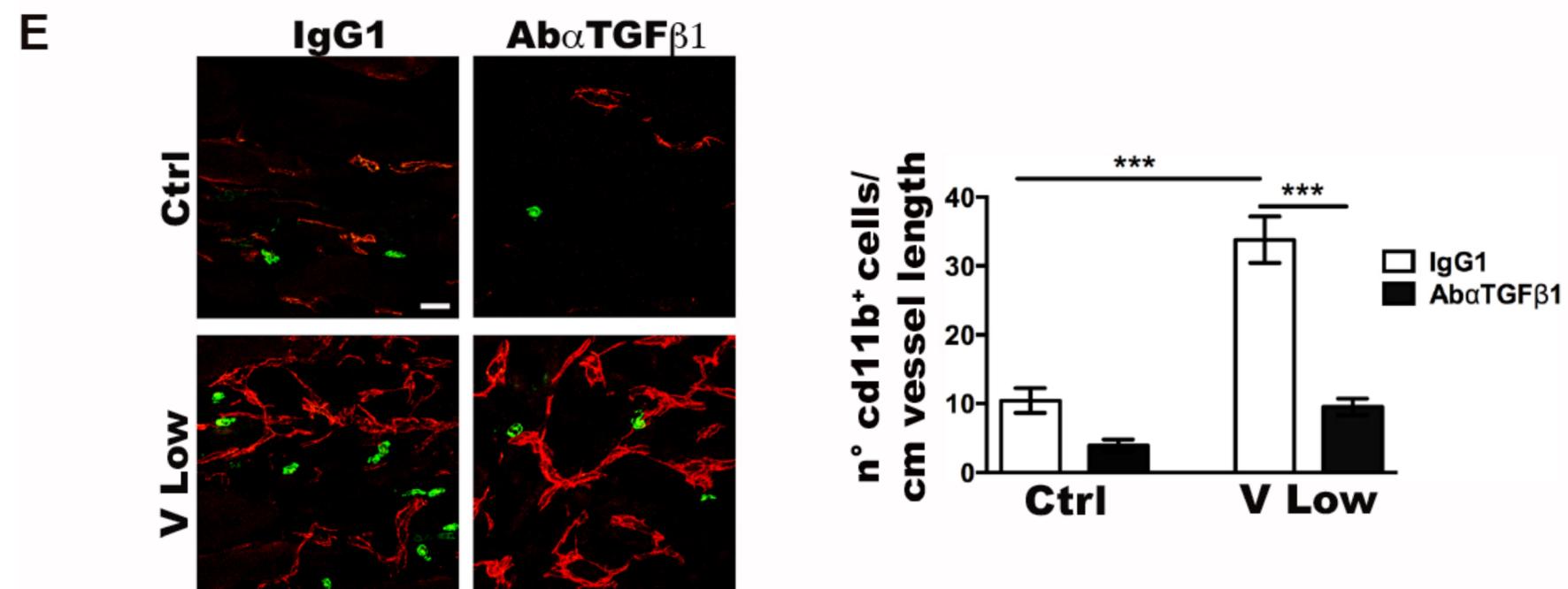
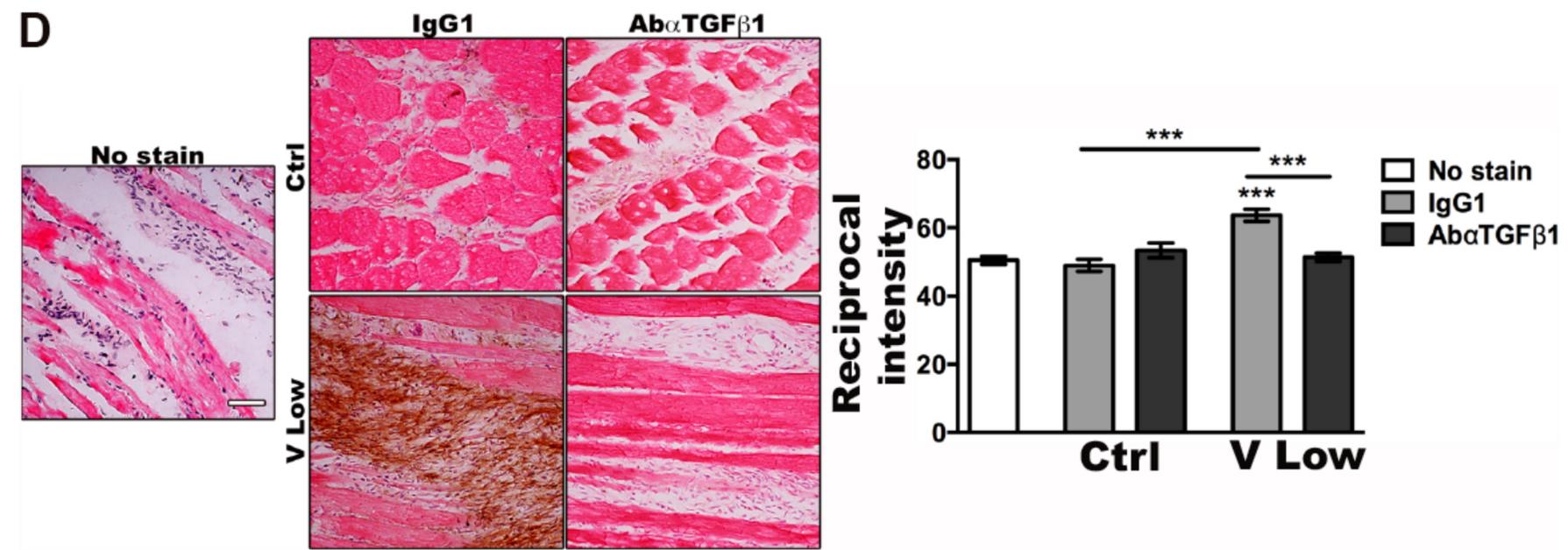
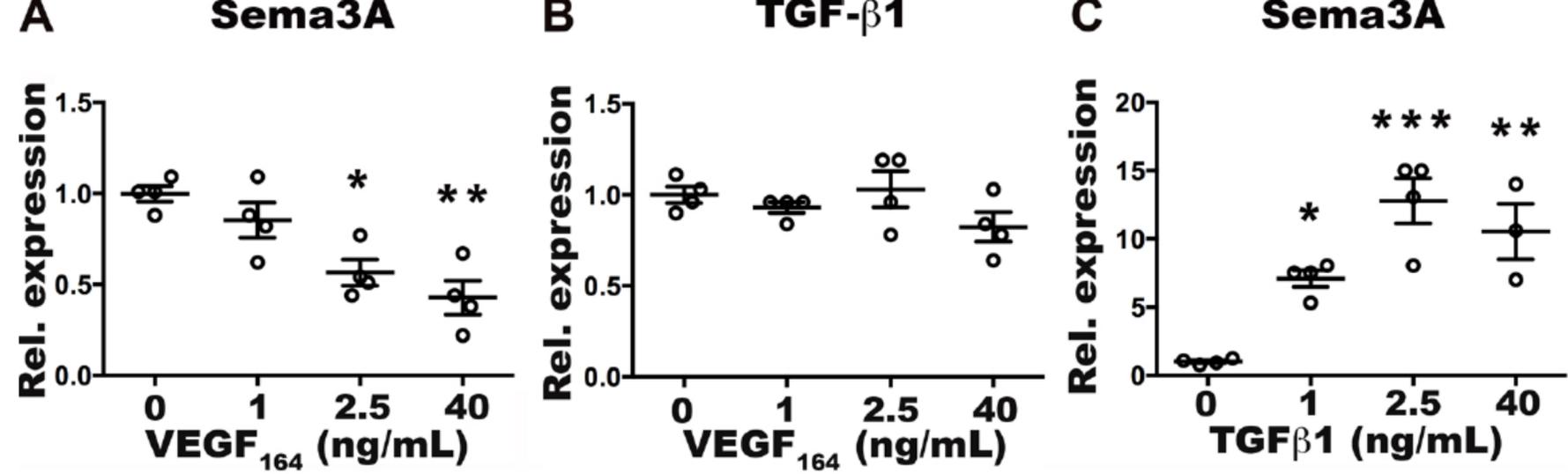












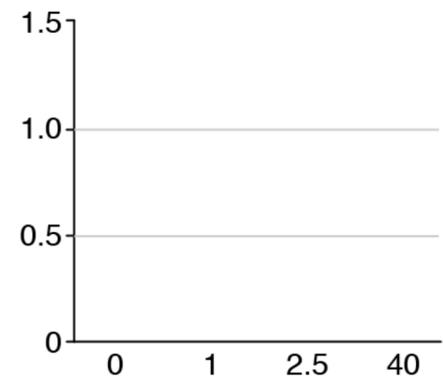
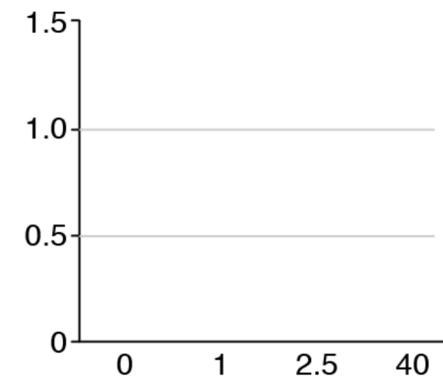
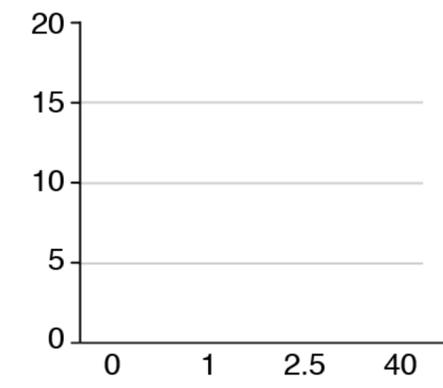
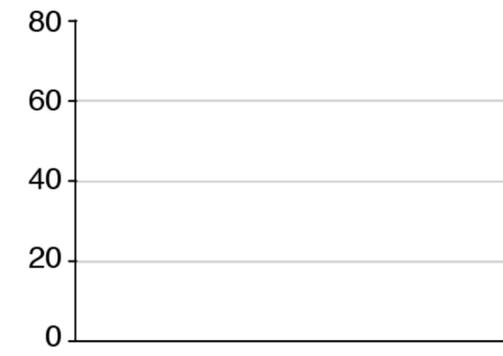
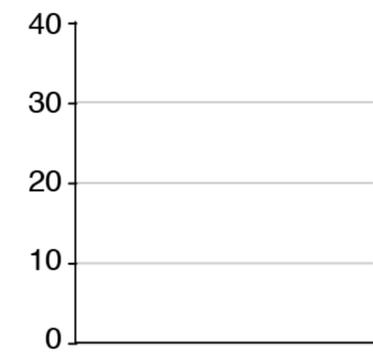
**A**

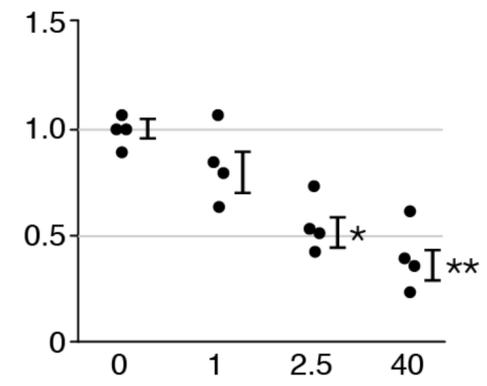
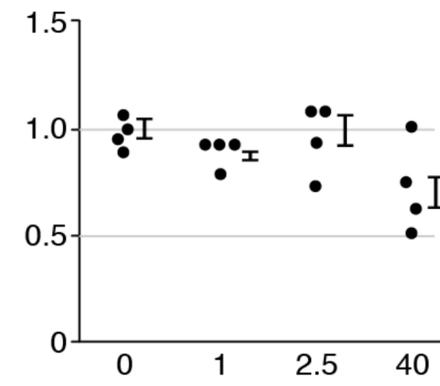
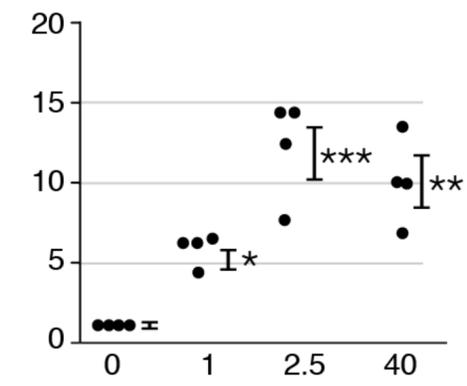
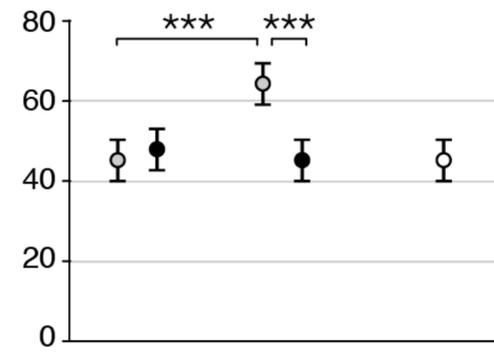
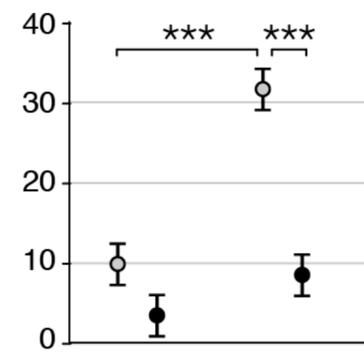
**B**

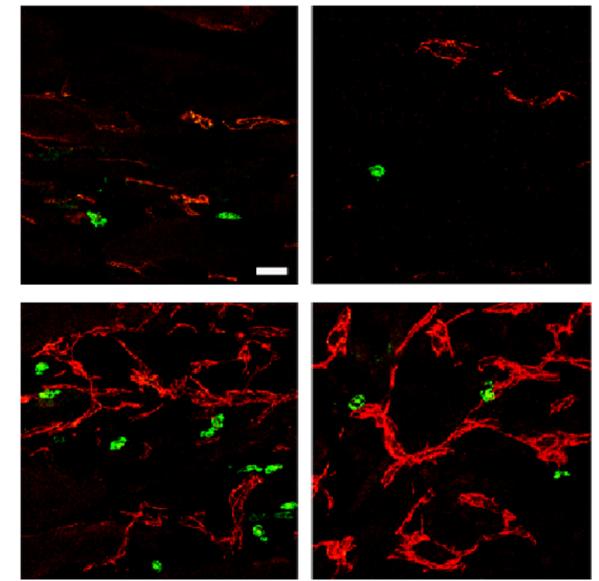
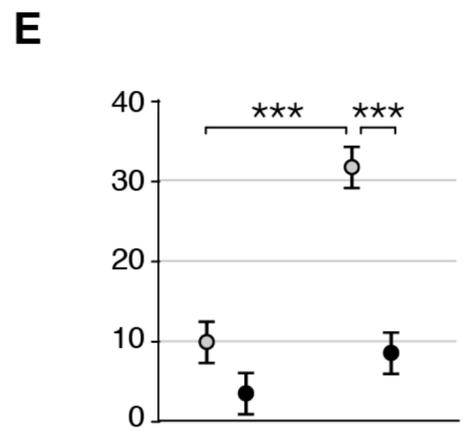
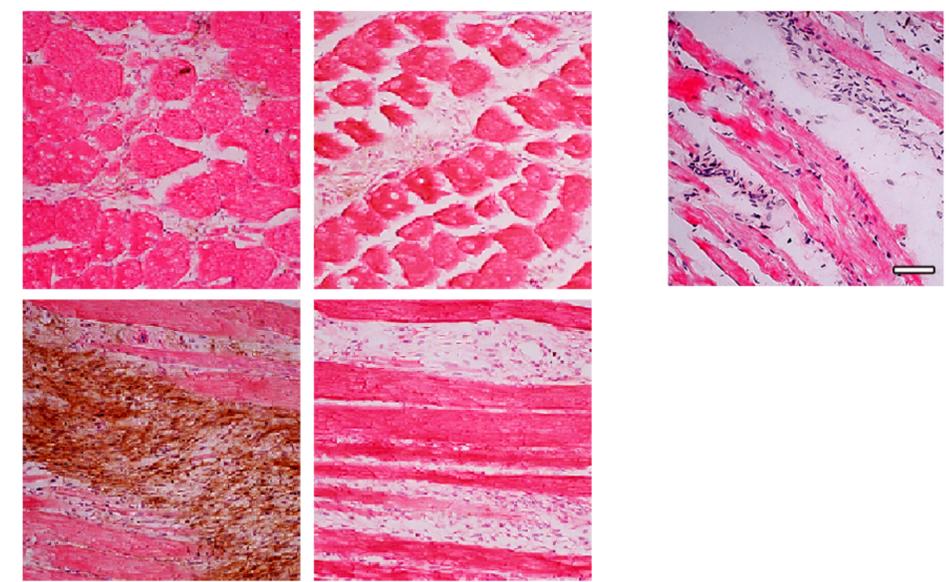
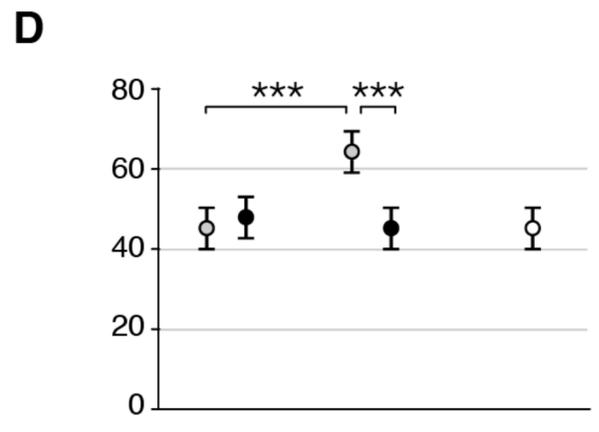
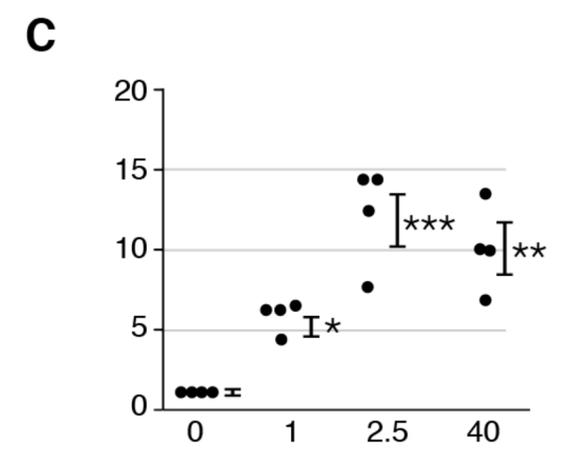
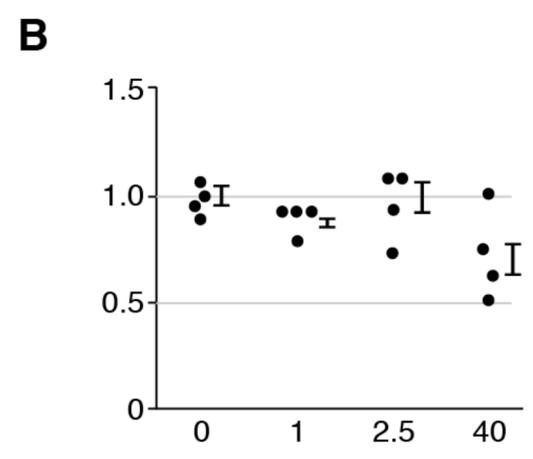
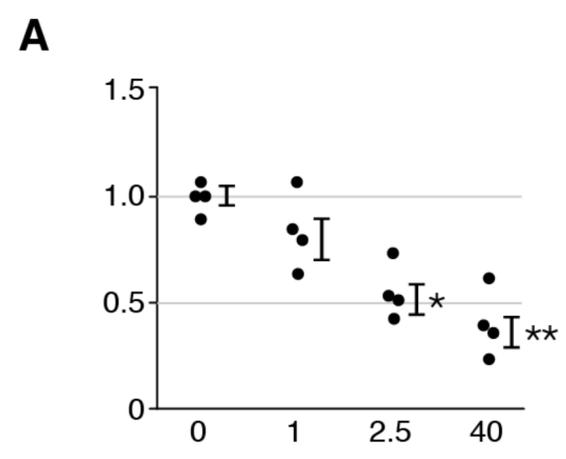
**C**

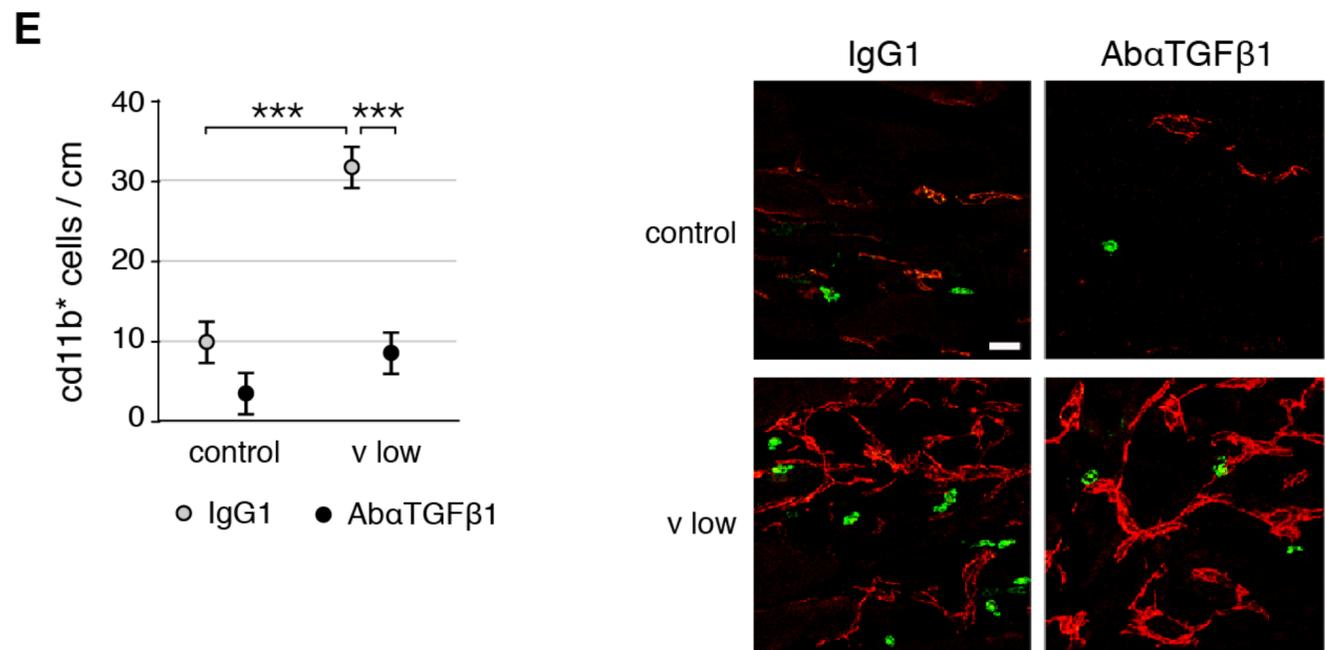
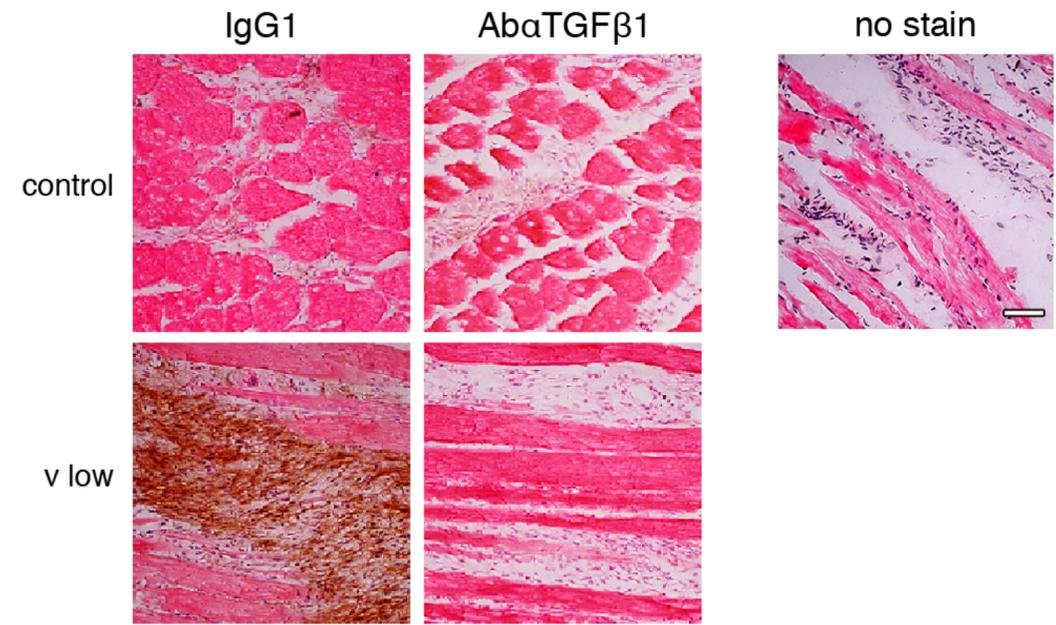
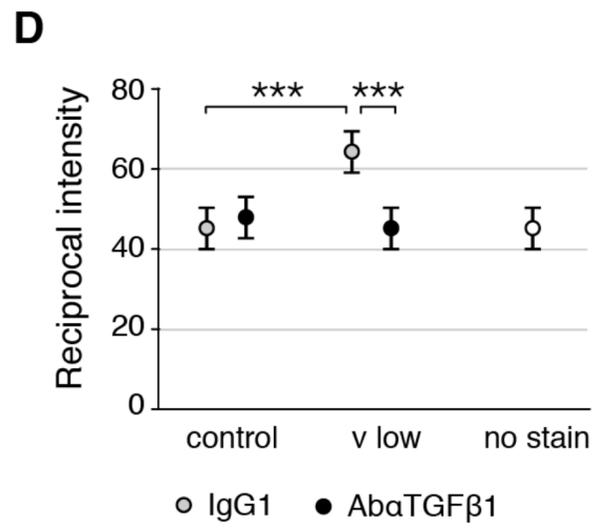
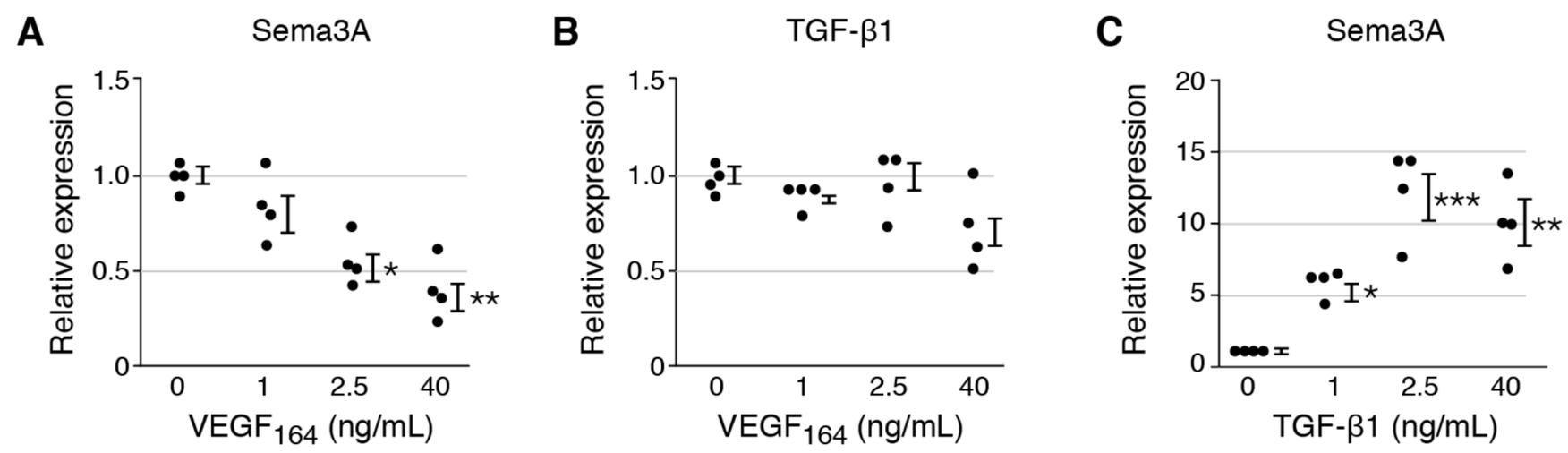
**D**

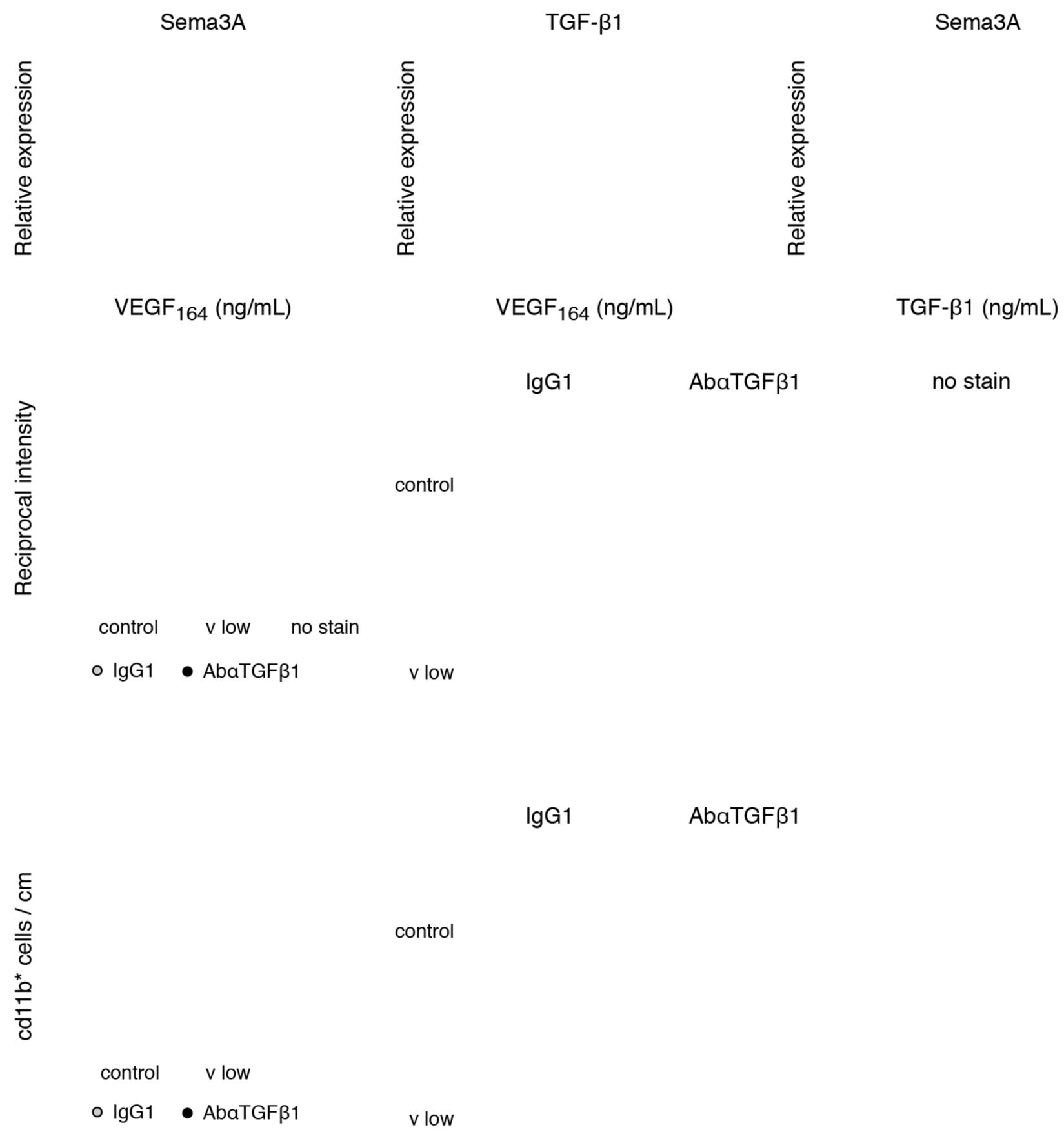
**E**

**A****B****C****D****E**

**A****B****C****D****E**



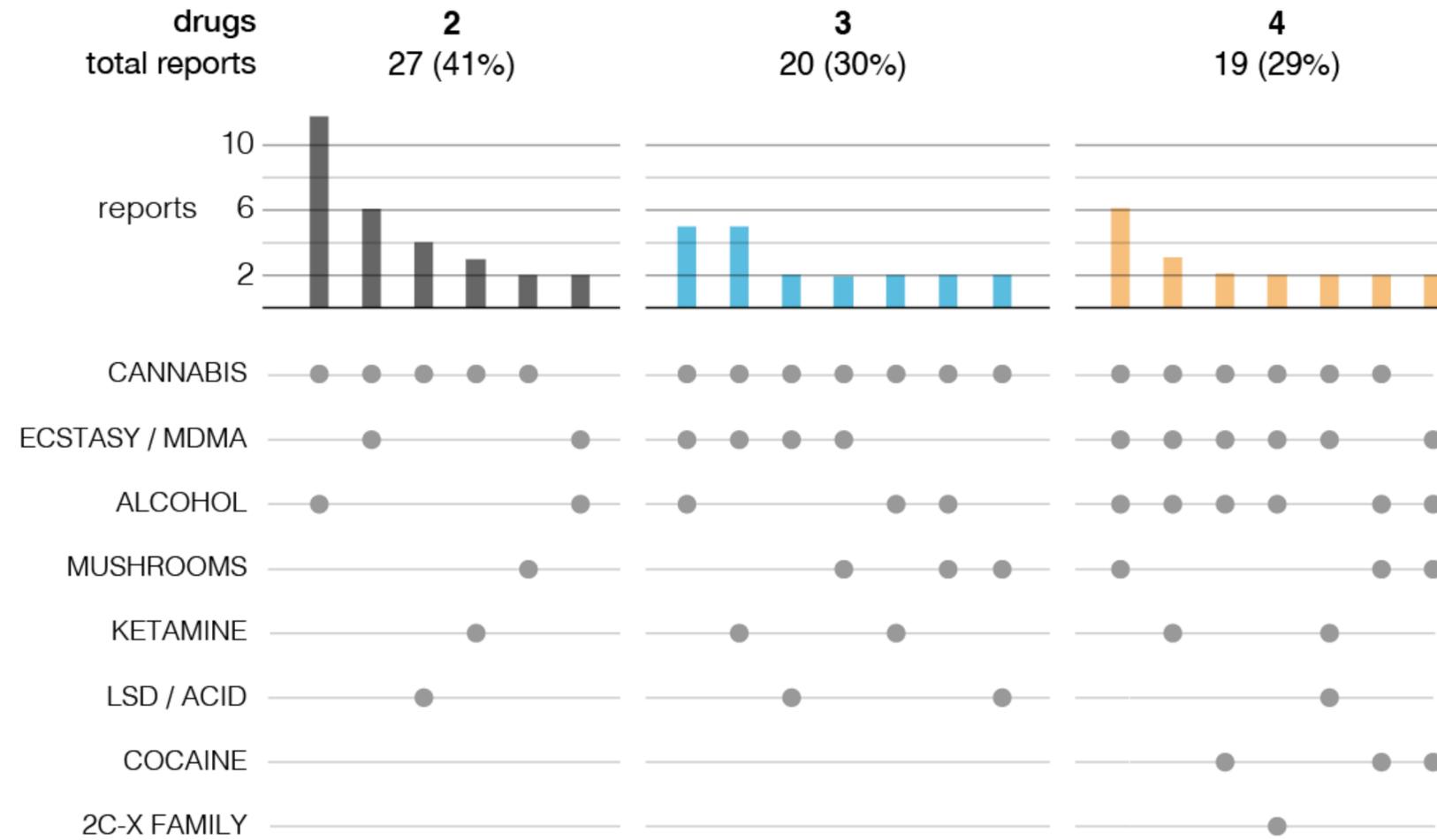




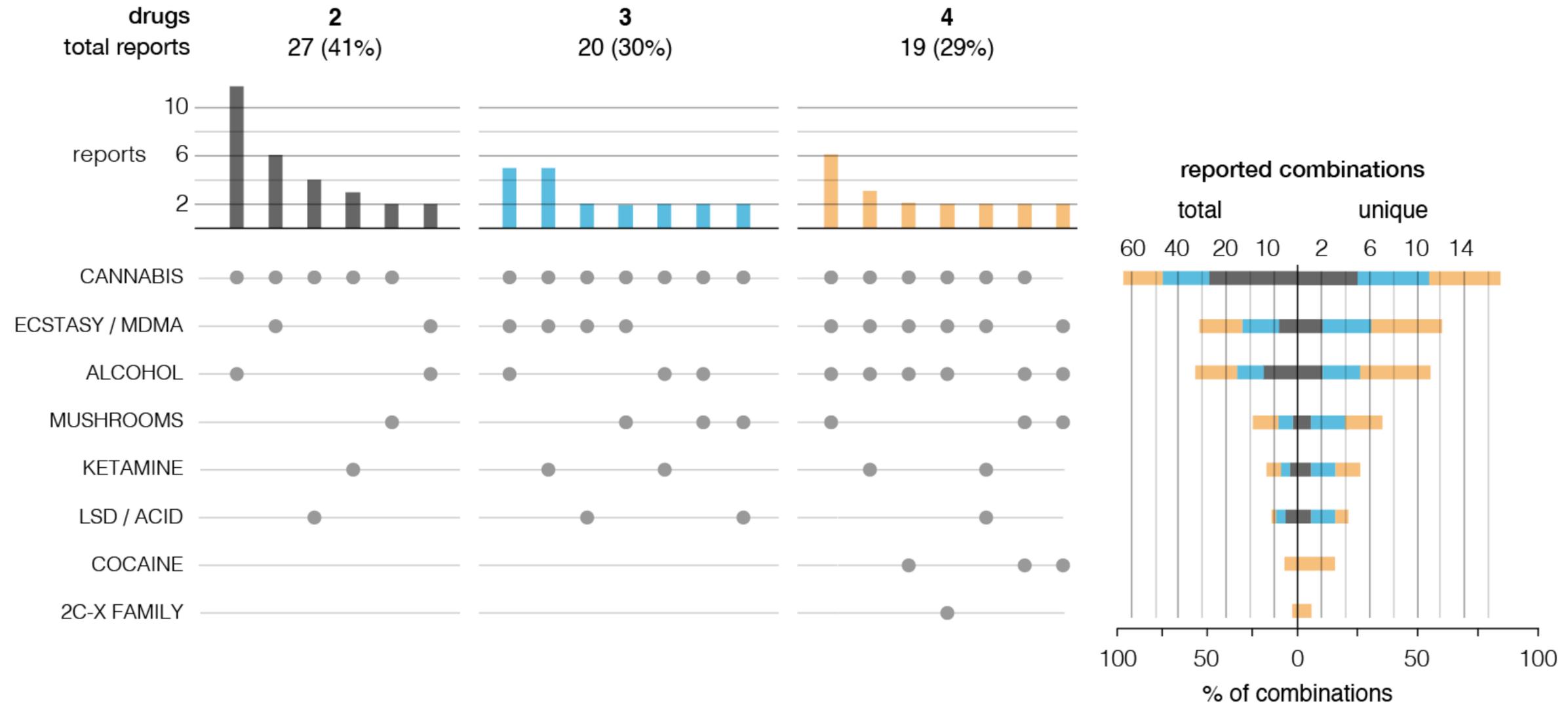
**Table 4: Composition of most prevalent 2-, 3- and 4-drug combinations and count of participants reporting each**

	2-Drug Combos						3-Drug Combos						4-Drug Combos								
Cannabis	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	Cannabis	
Ecstasy / MDMA		X				X	X	X	X	X				X	X	X	X	X		X	Ecstasy / MDMA
Alcohol	X					X	X				X	X		X	X	X	X		X	X	Alcohol
Ketamine				X				X			X				X			X			Ketamine
LSD / Acid			X						X				X					X			LSD / Acid
Cocaine (powder)																X			X	X	Cocaine (powder)
Mushrooms					X					X		X	X	X					X	X	Mushrooms
2C-x Family																	X				2C-x Family
Count of Reports	12	6	4	3	2	2	5	5	2	2	2	2	2	6	3	2	2	2	2	2	Count of Reports

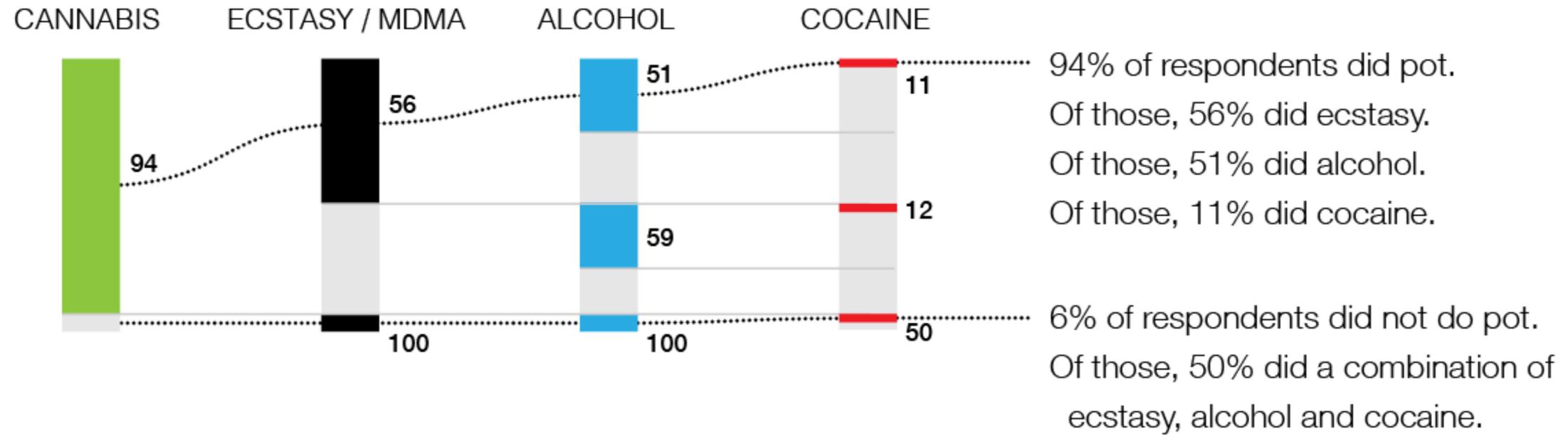
**Table 4.** Composition of the most prevalent 2-, 3- and 4-drug combinations, the count of participants reporting each and the number of times the use of a drug has been reported.



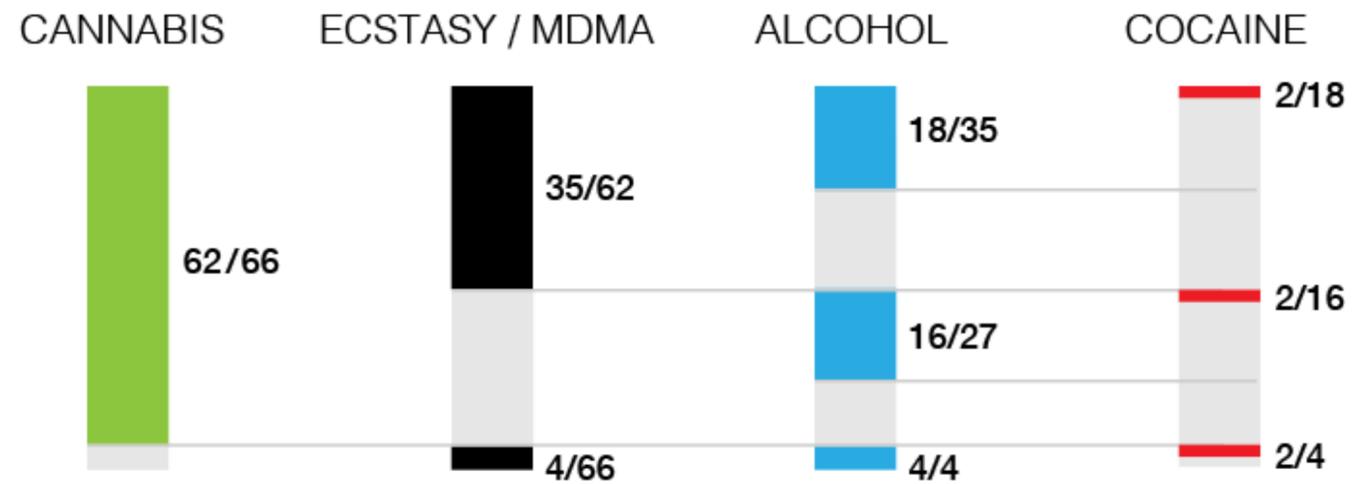
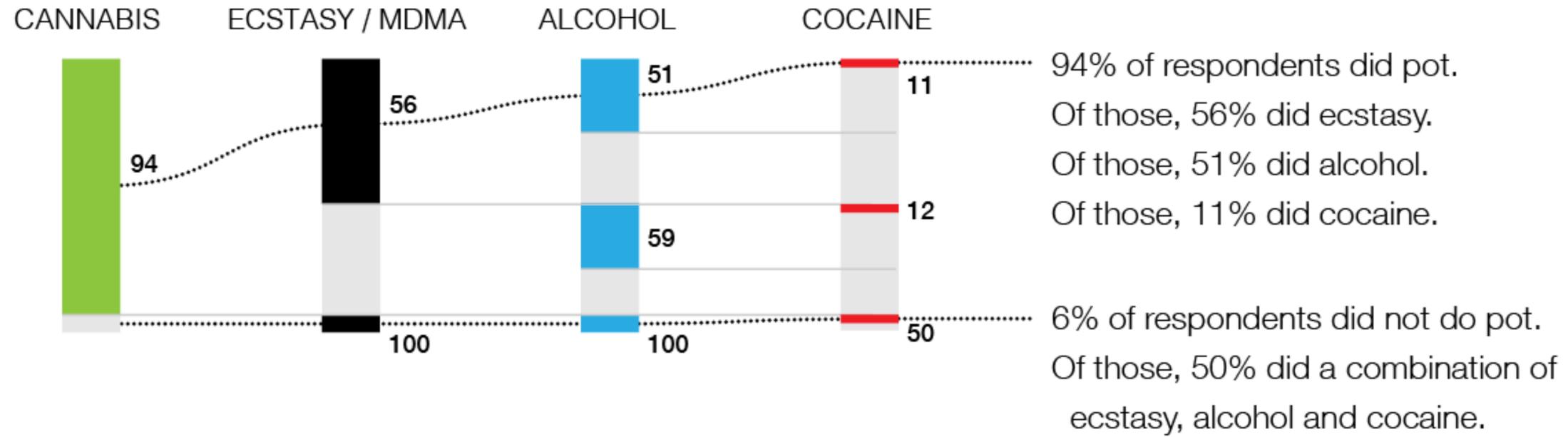
**Table 4.** Composition of the most prevalent 2-, 3- and 4-drug combinations, the count of participants reporting each and the number of times the use of a drug has been reported.



**Table 5.** Want to lower your risk of cocaine use? Smoke pot.



**Table 5.** Want to lower your risk of cocaine use? Smoke pot.



So by design I mean... don't just use your favourite color.

Don't try to scratch the itch to use that great clipart of a eppendorf tube you found online.

Don't ever think that the data will speak for themselves. They don't. They make no sound.

<hold up network diagram>

I can't hear anything!

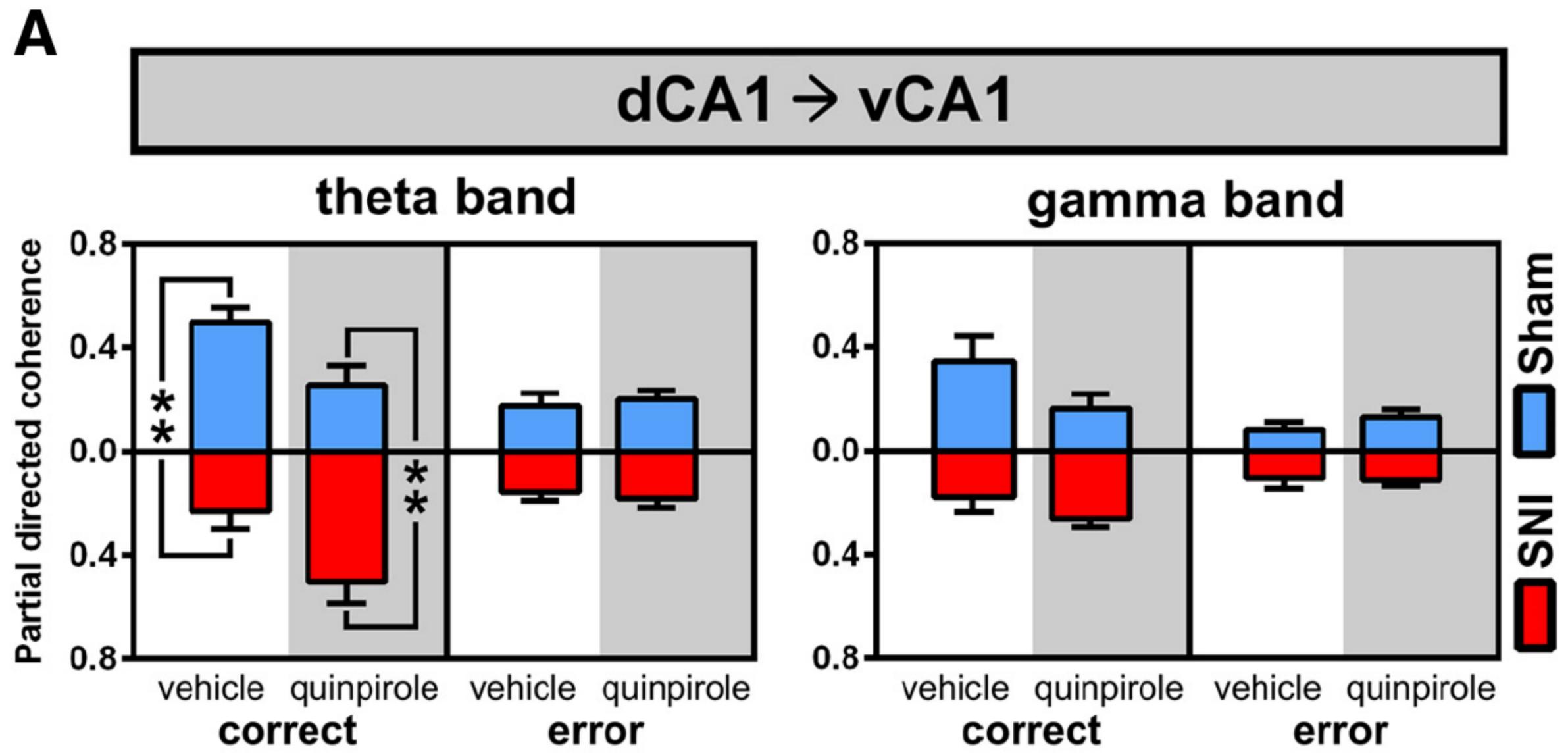
Always assume that everything you draw will influence how everything else is perceived. The weight of that line. The curvature of that arrow. It all makes a difference.

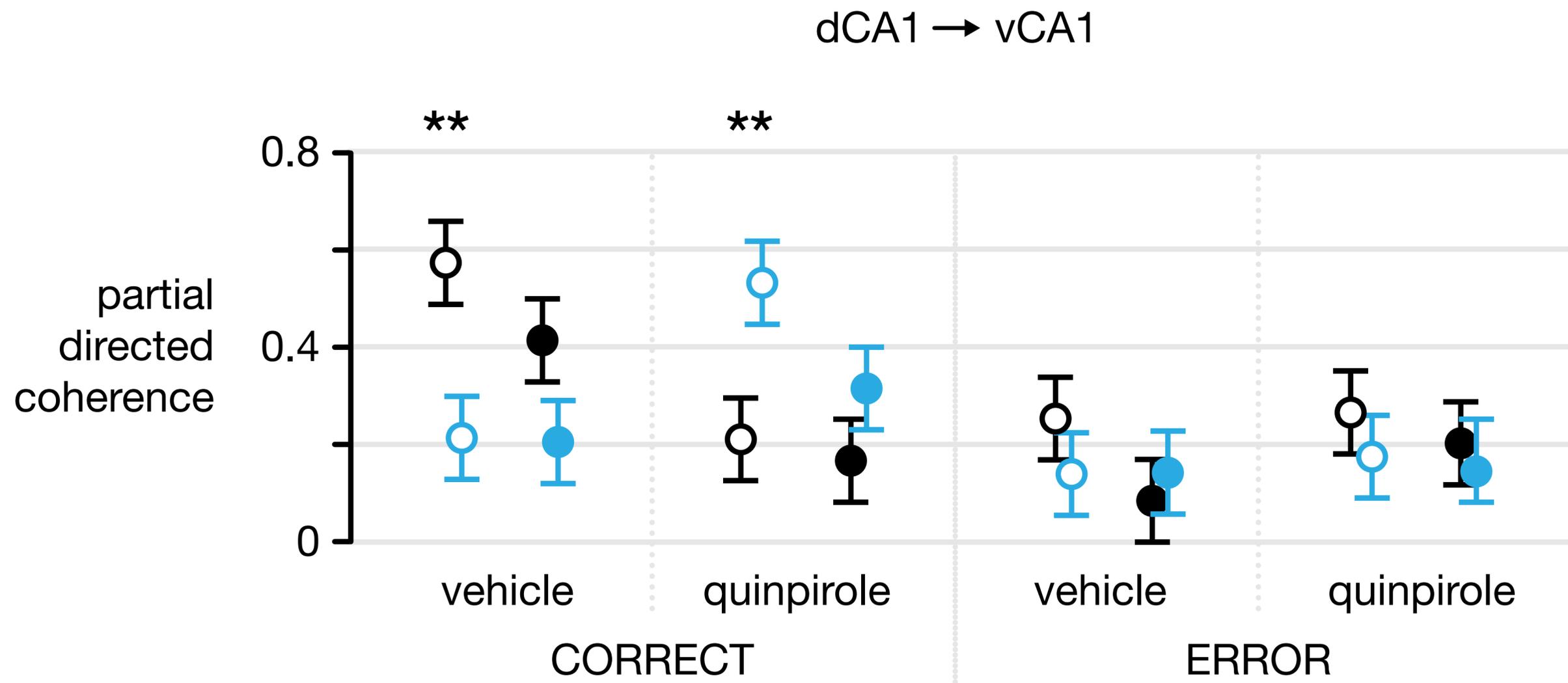
Practise the campfire rule—when you're done with your visualization always leave the data as good as or better than when you found it.



# EXERCISE 1

Redesign the figure.

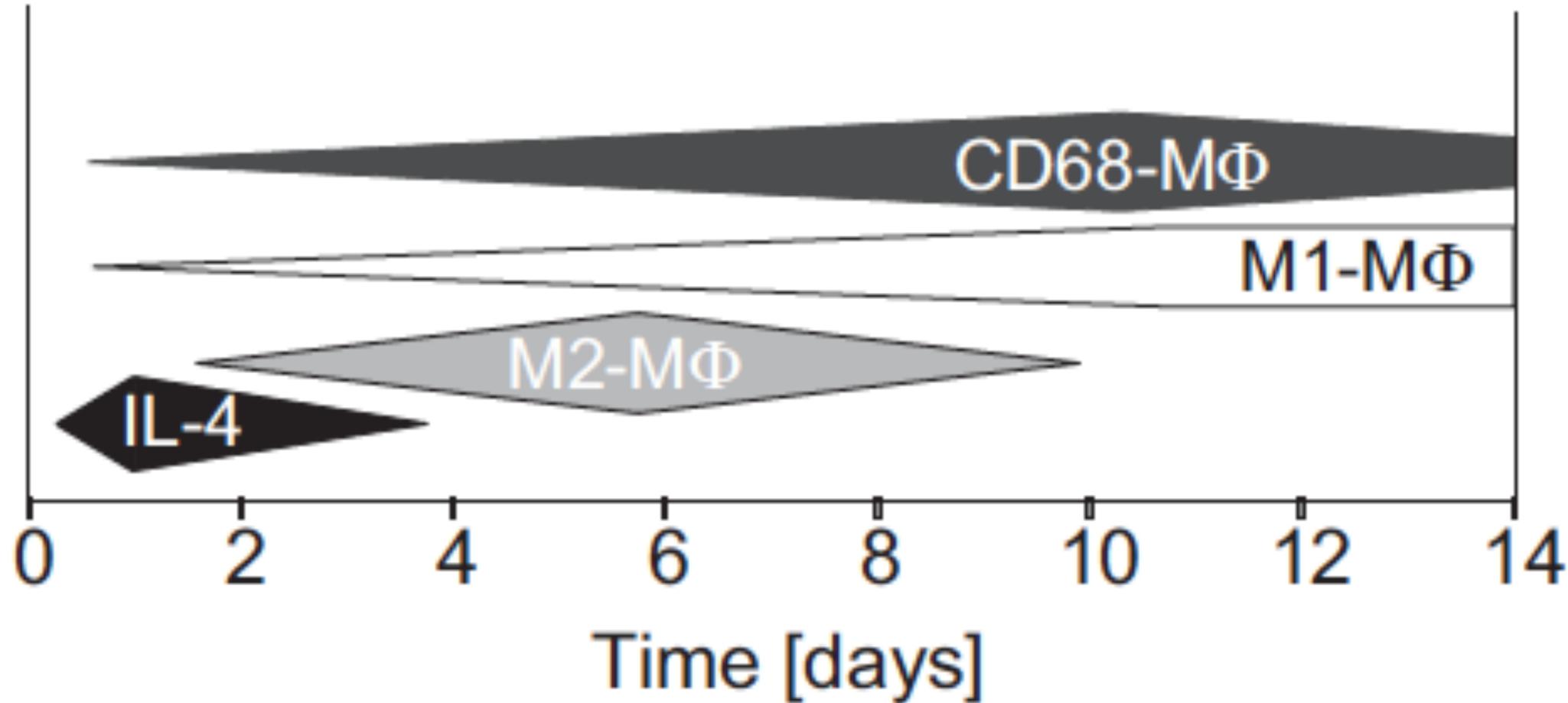


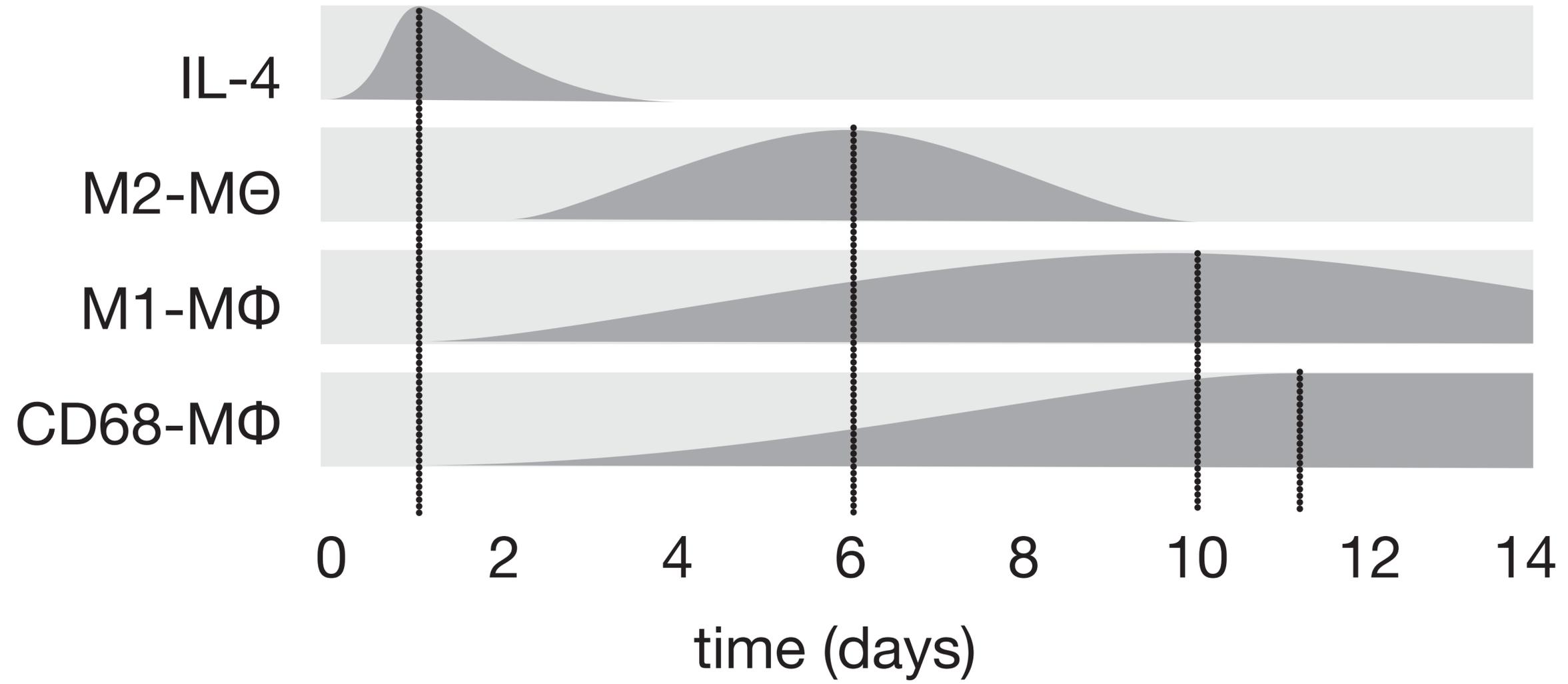


theta band ○ ○  
gamma band ● ●  
sham SNI

# EXERCISE 2

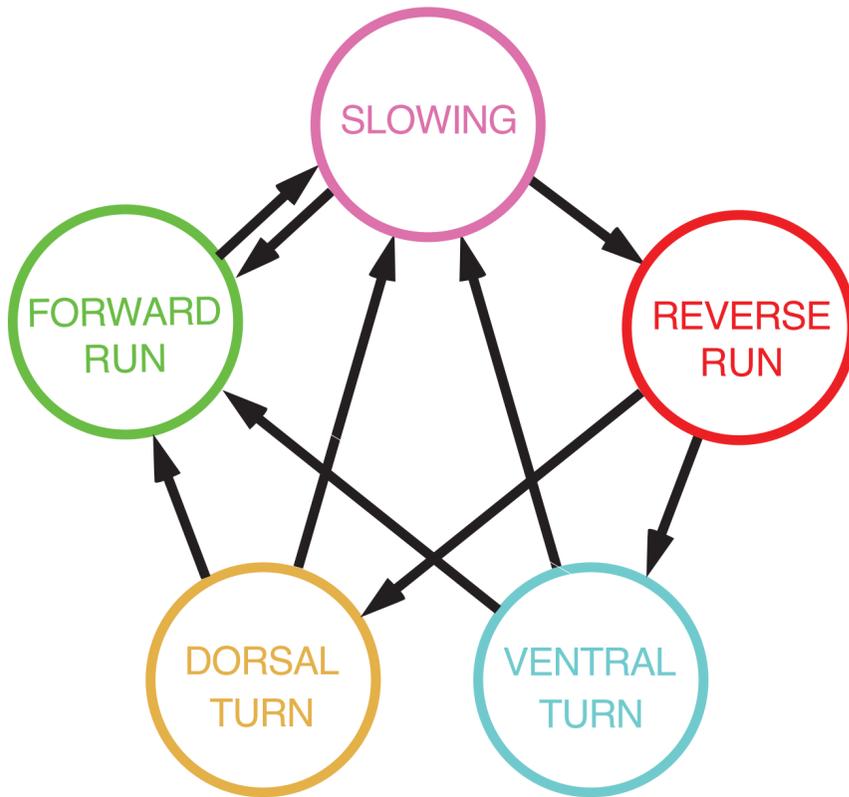
Redesign the figure.

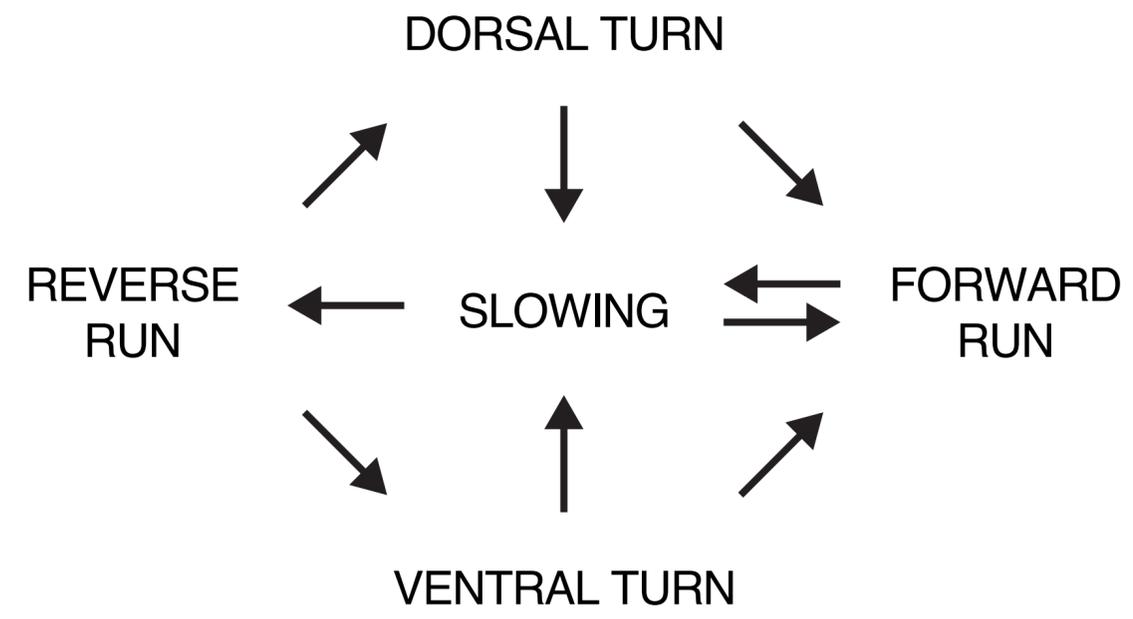
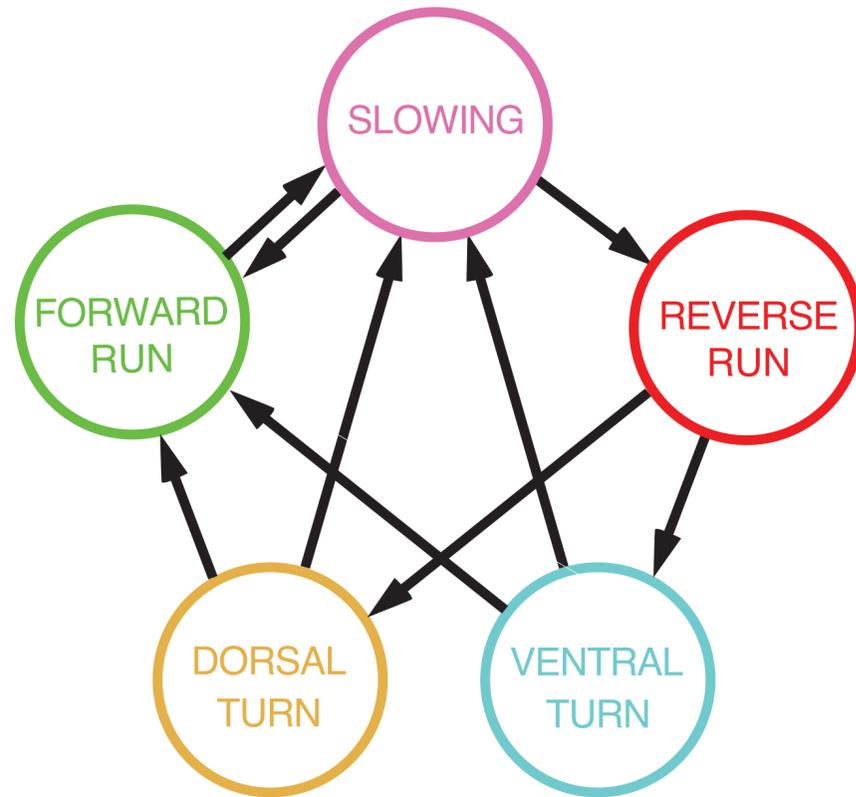




# EXERCISE 3

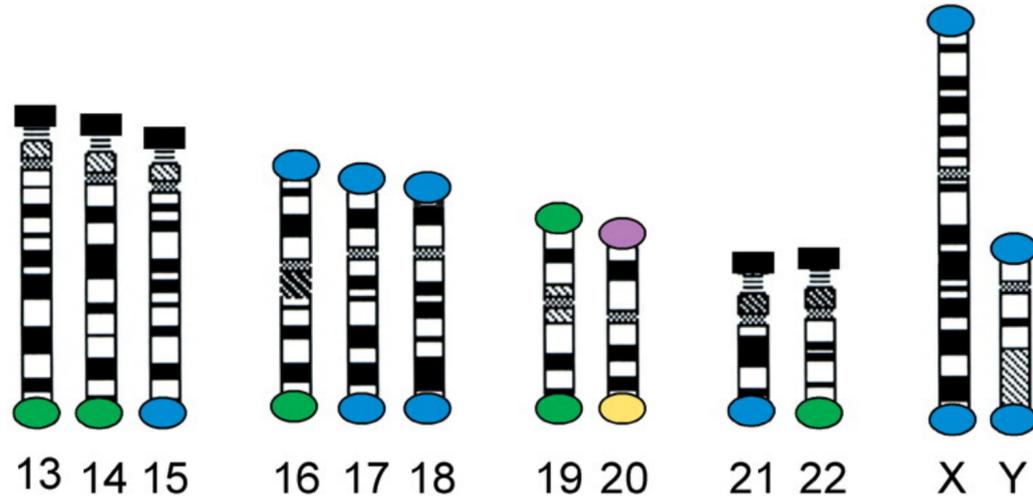
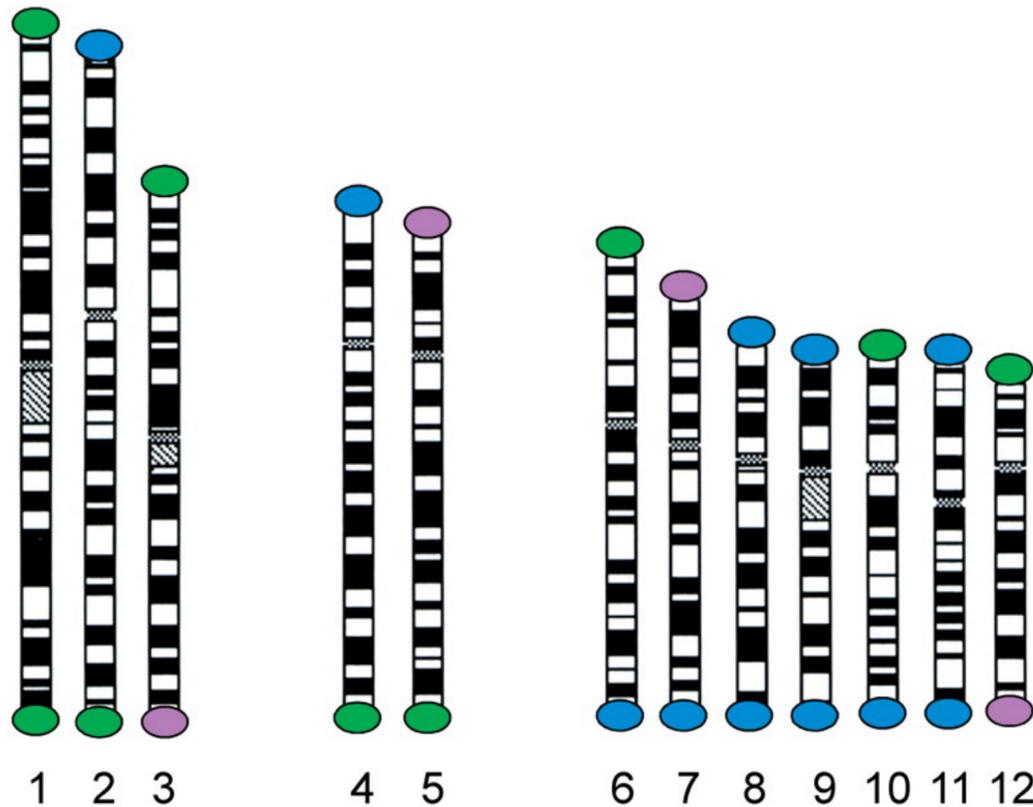
Redesign the figure.



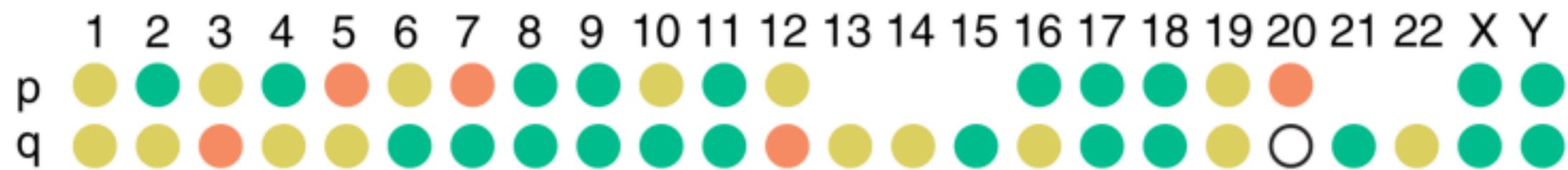


# EXERCISE 4

Redesign the figure.



### OPTION 1



- no gap ●
- <20kb ●
- <20-70kb ●
- unk/srp ○
- acrocentric ●

### OPTION 2

