

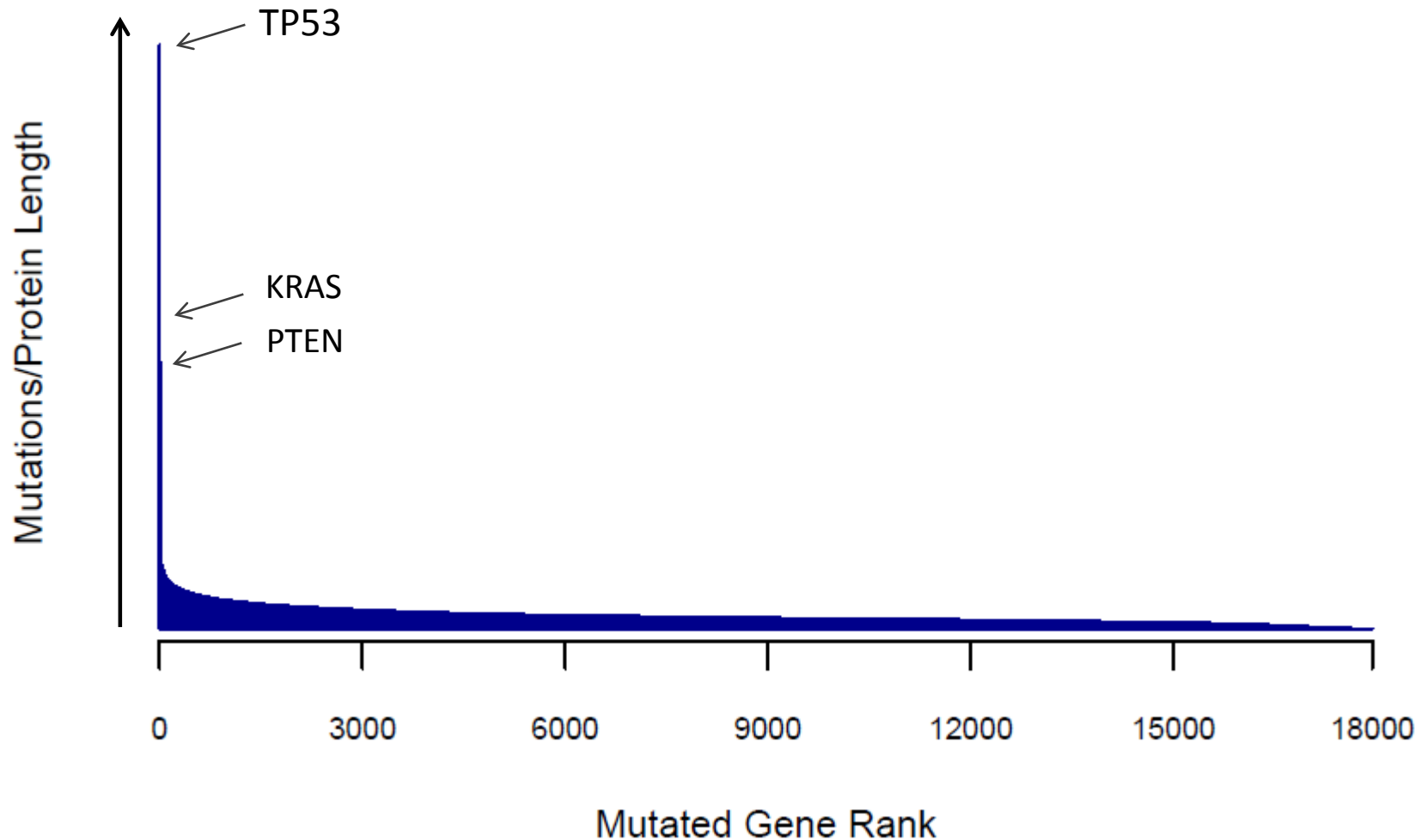
# LineUp

## Identifying Deleterious Mutations Using Protein Domain Alignment

Brady Bernard  
Shmulevich Group

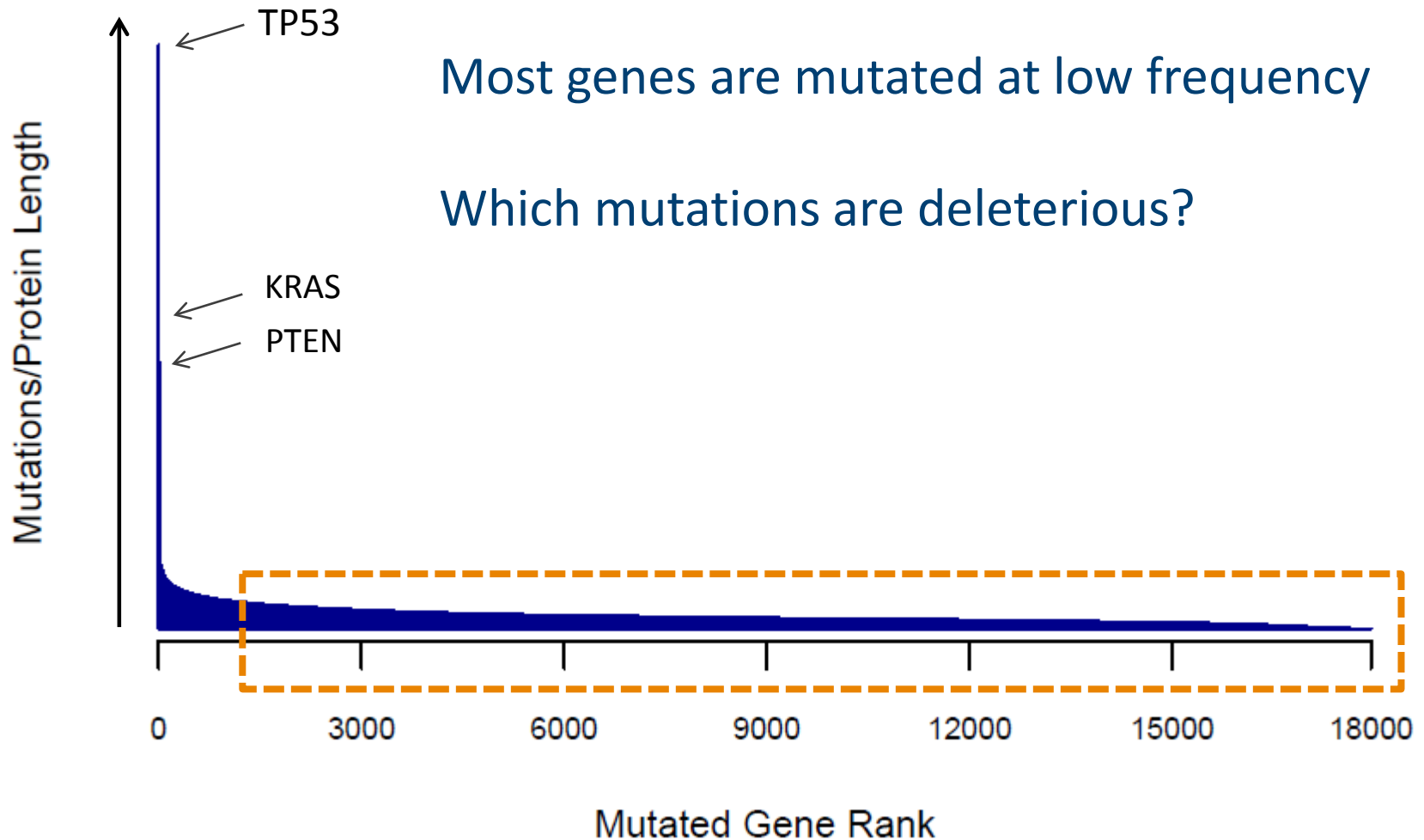
# Normalized mutation frequencies

## Twenty tumor types

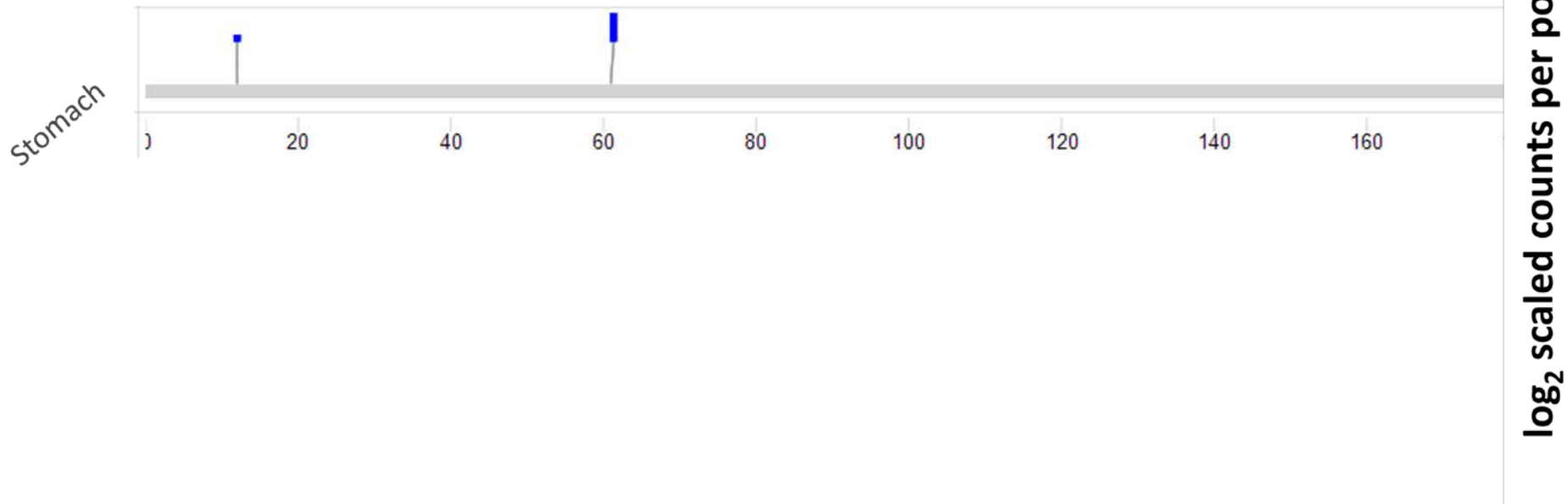


# Normalized mutation frequencies

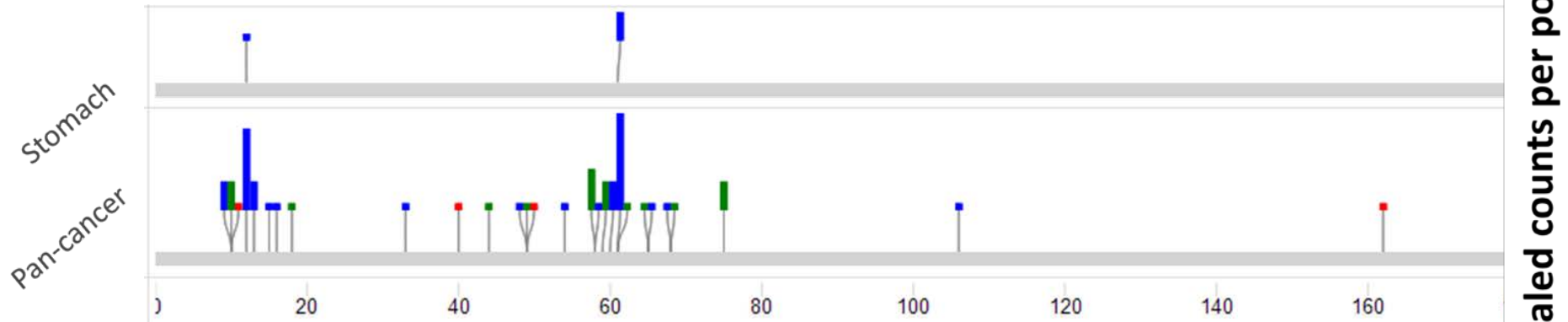
## Twenty tumor types



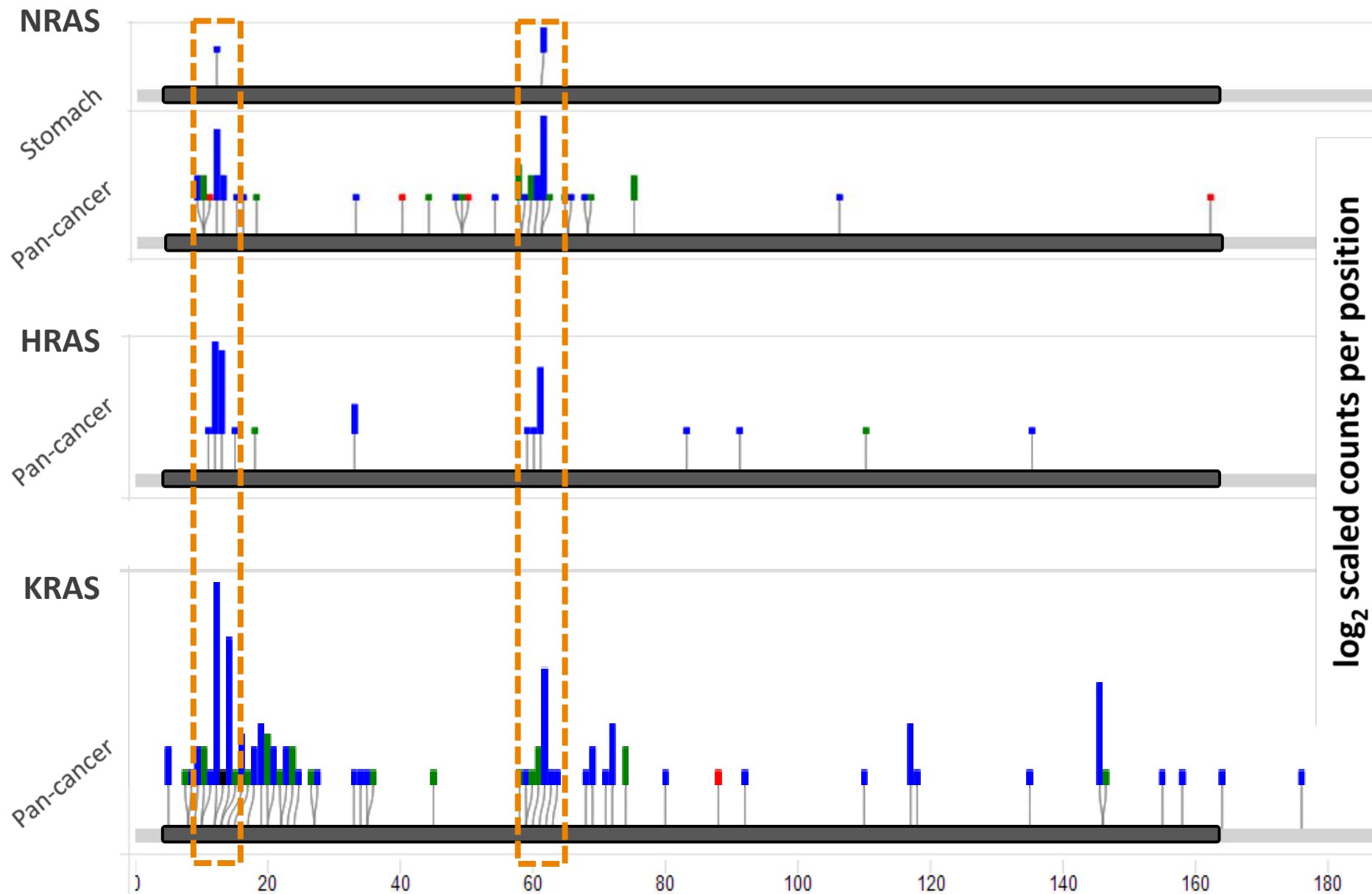
# Example protein mutation spectrum



# Pan-cancer mutation spectrum



# Pan-cancer Ras domain mutation spectra



# Ras domain structure alignment

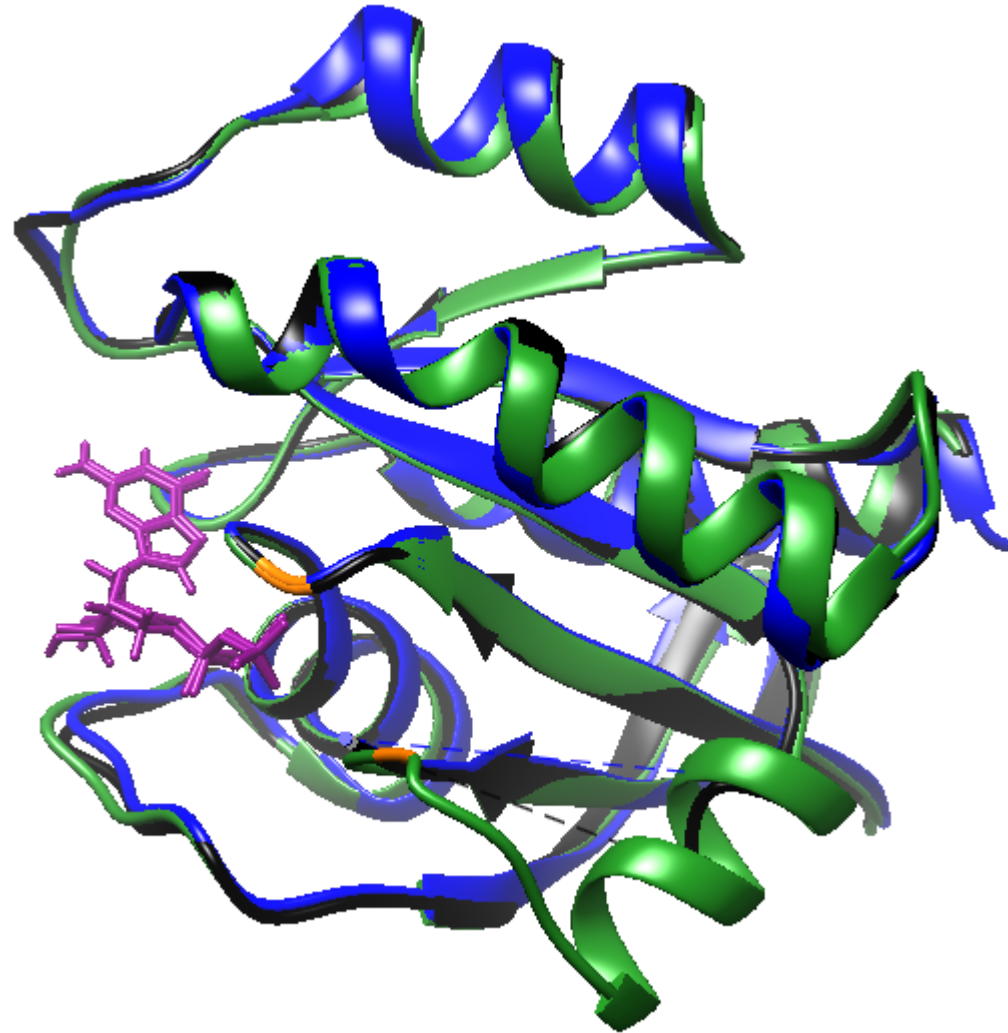
**KRAS**

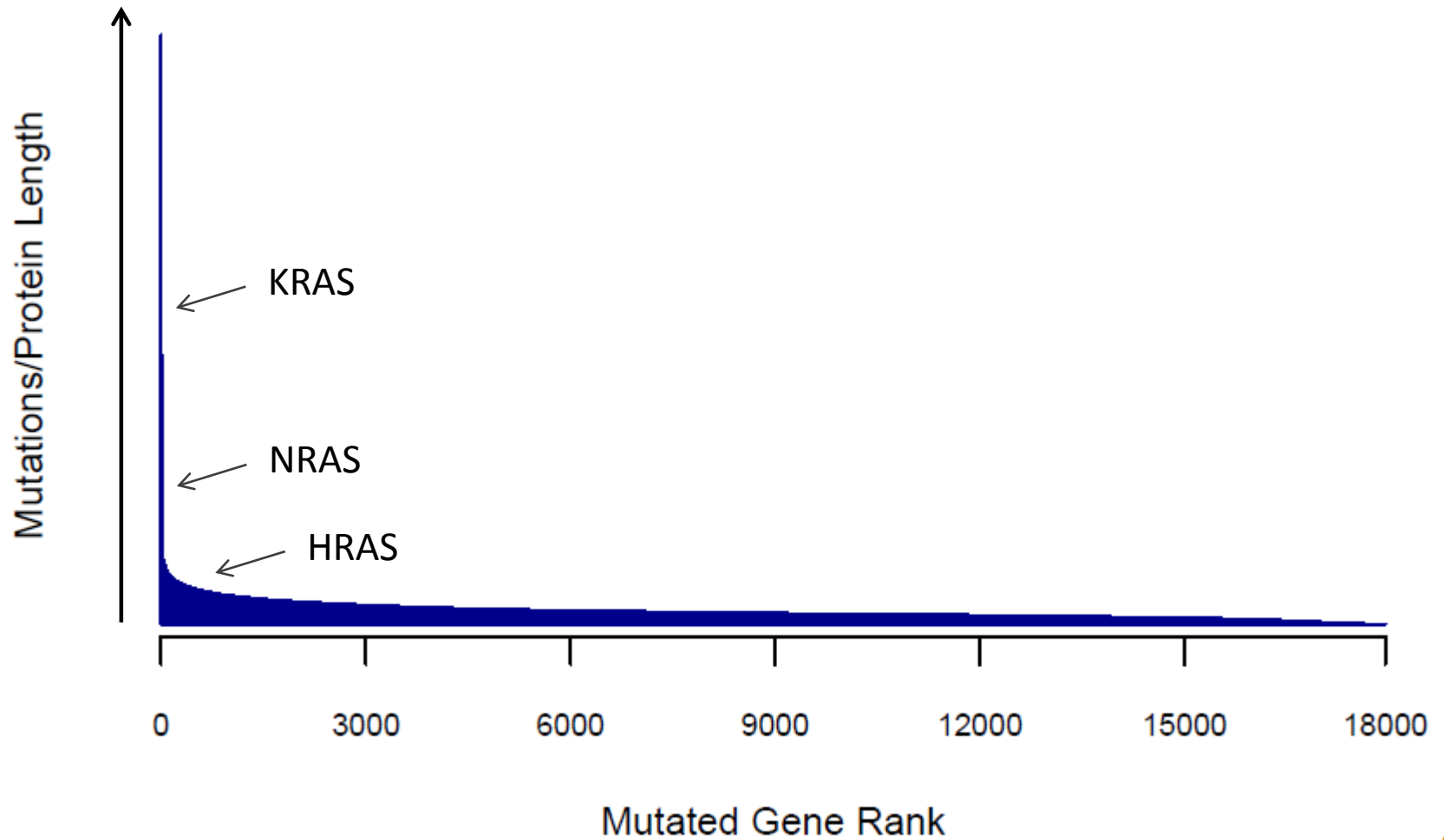
**NRAS**

**HRAS**

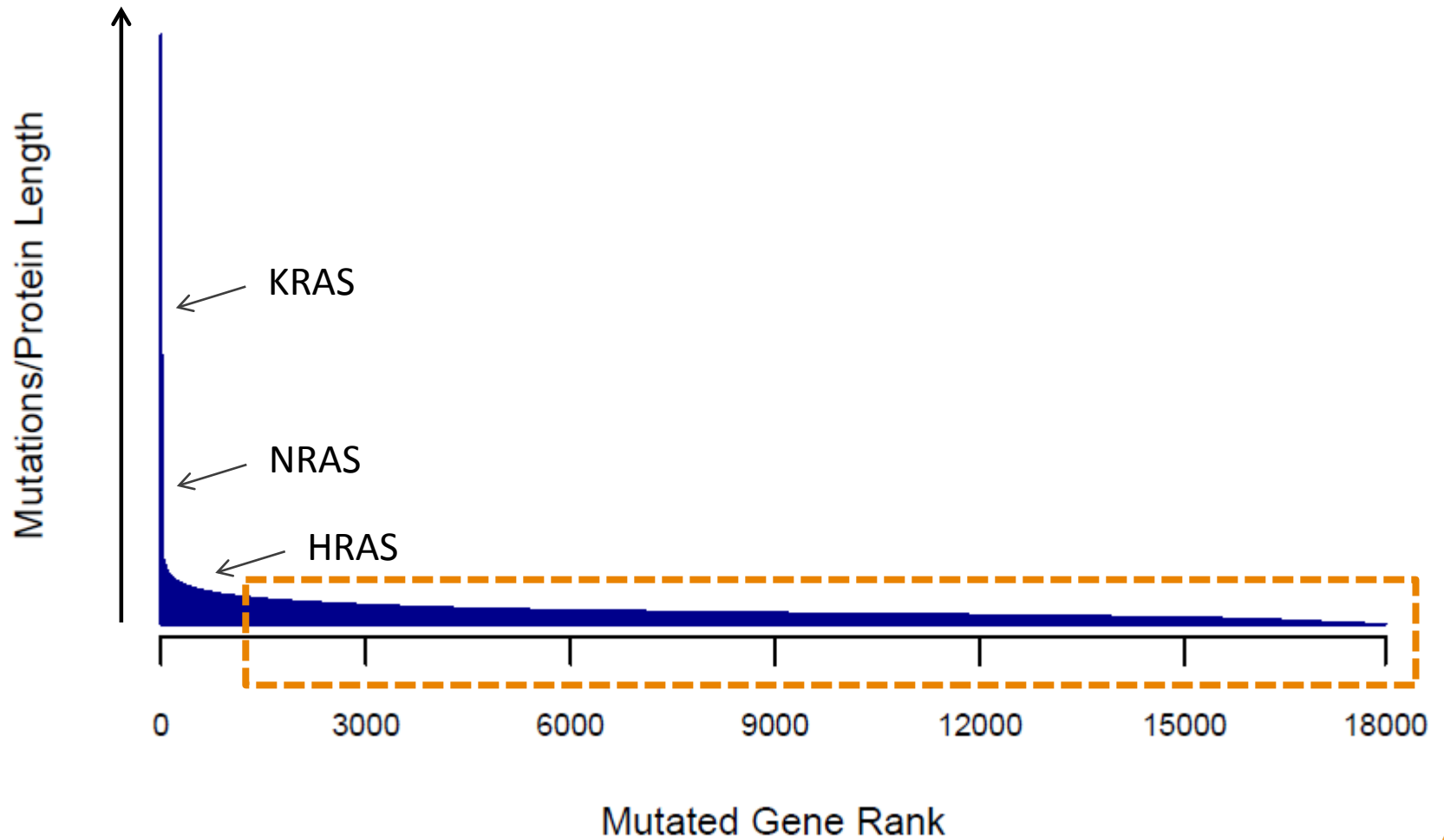
**GDP**

**Hotspots**

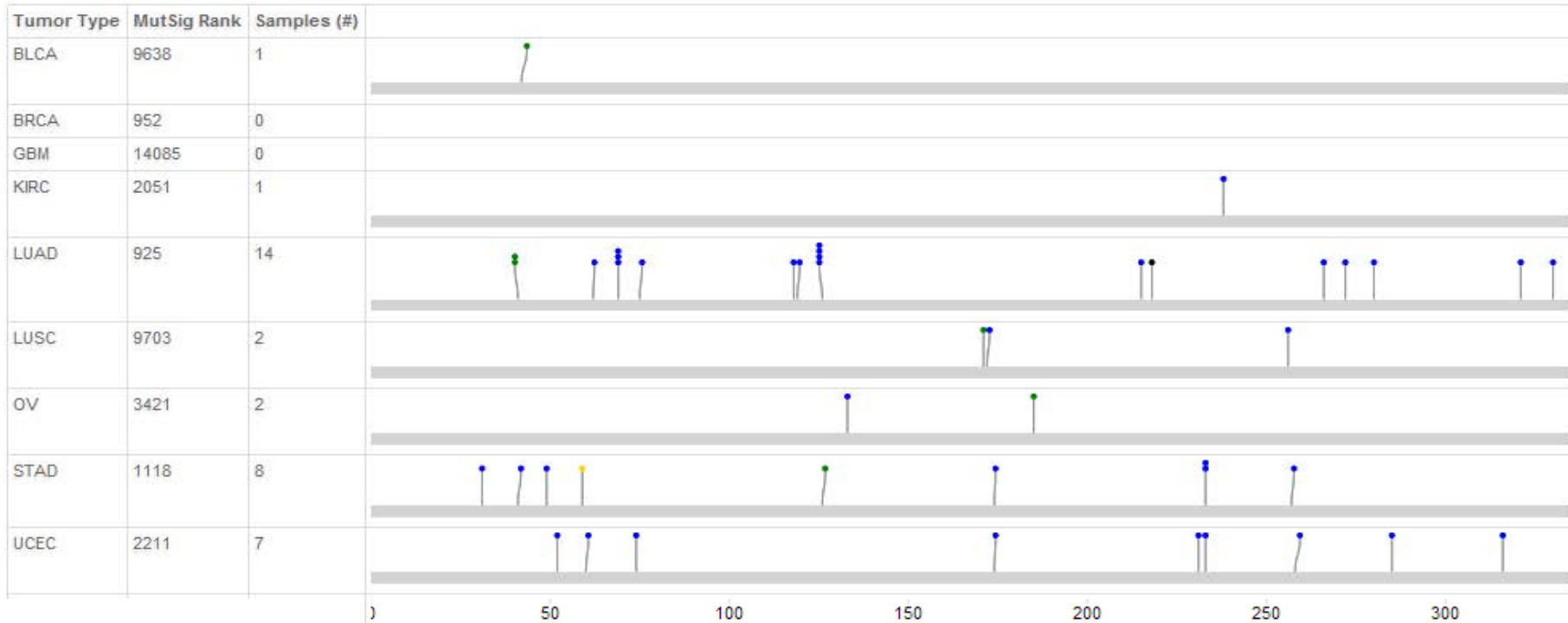






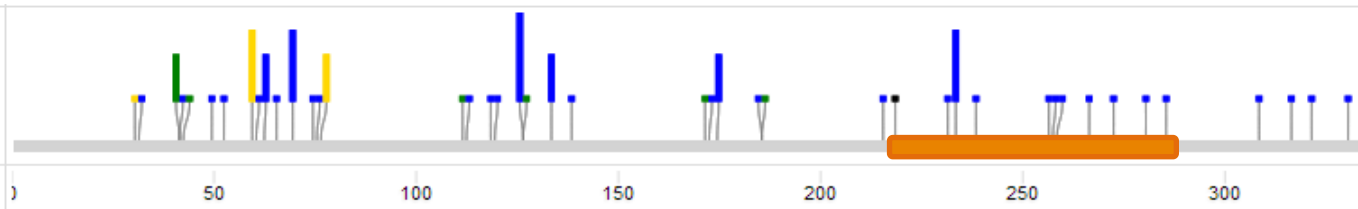
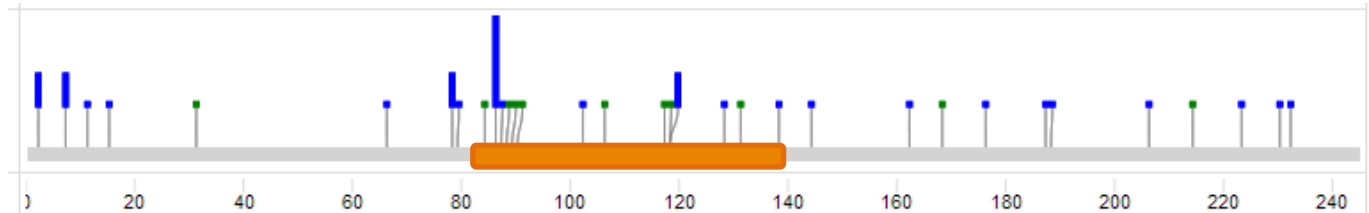


# Typical pan-cancer mutation spectrum



# LineUp Approach

Gene A  
Pan-cancer

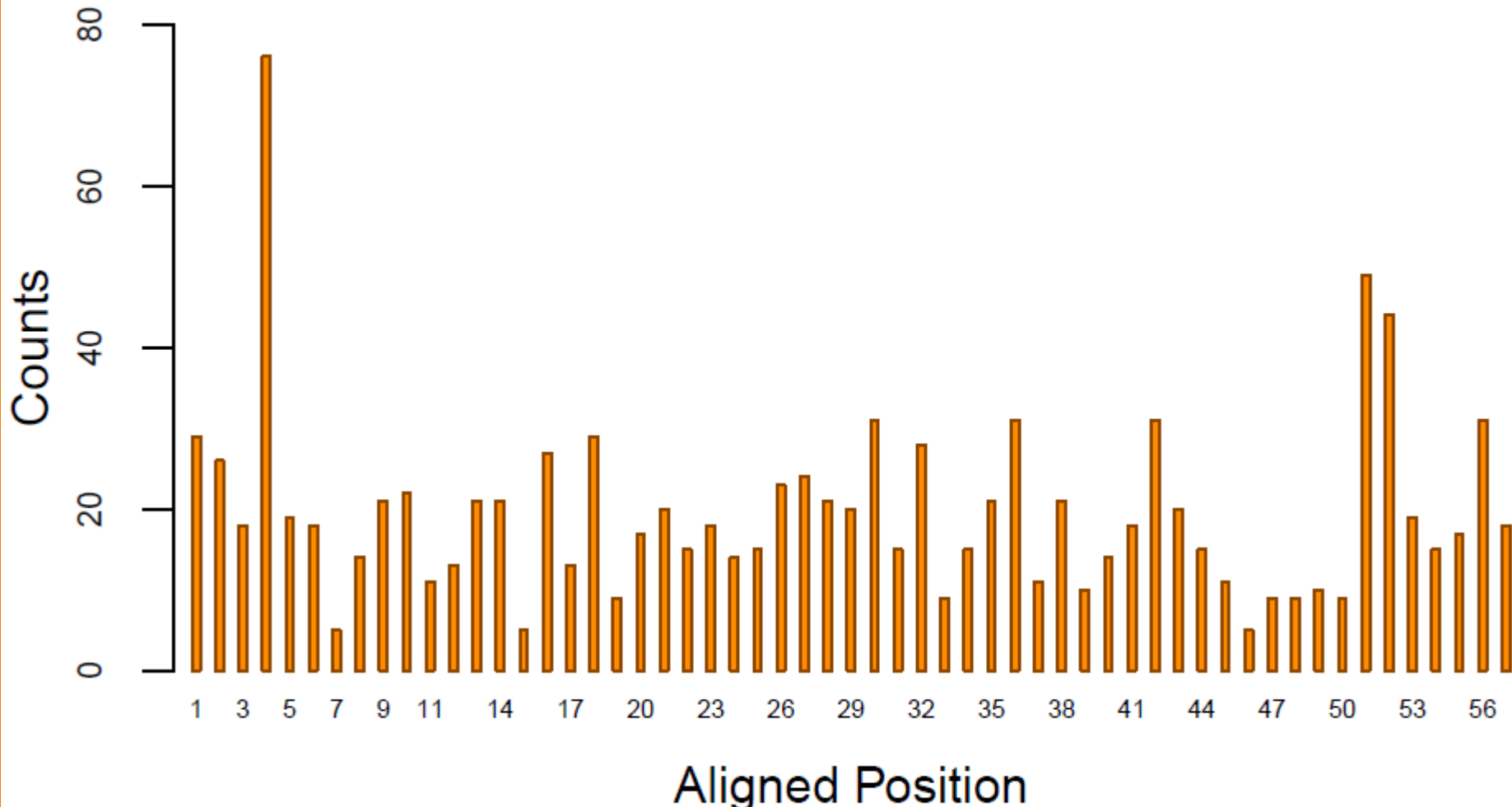


Gene B  
Pan-cancer

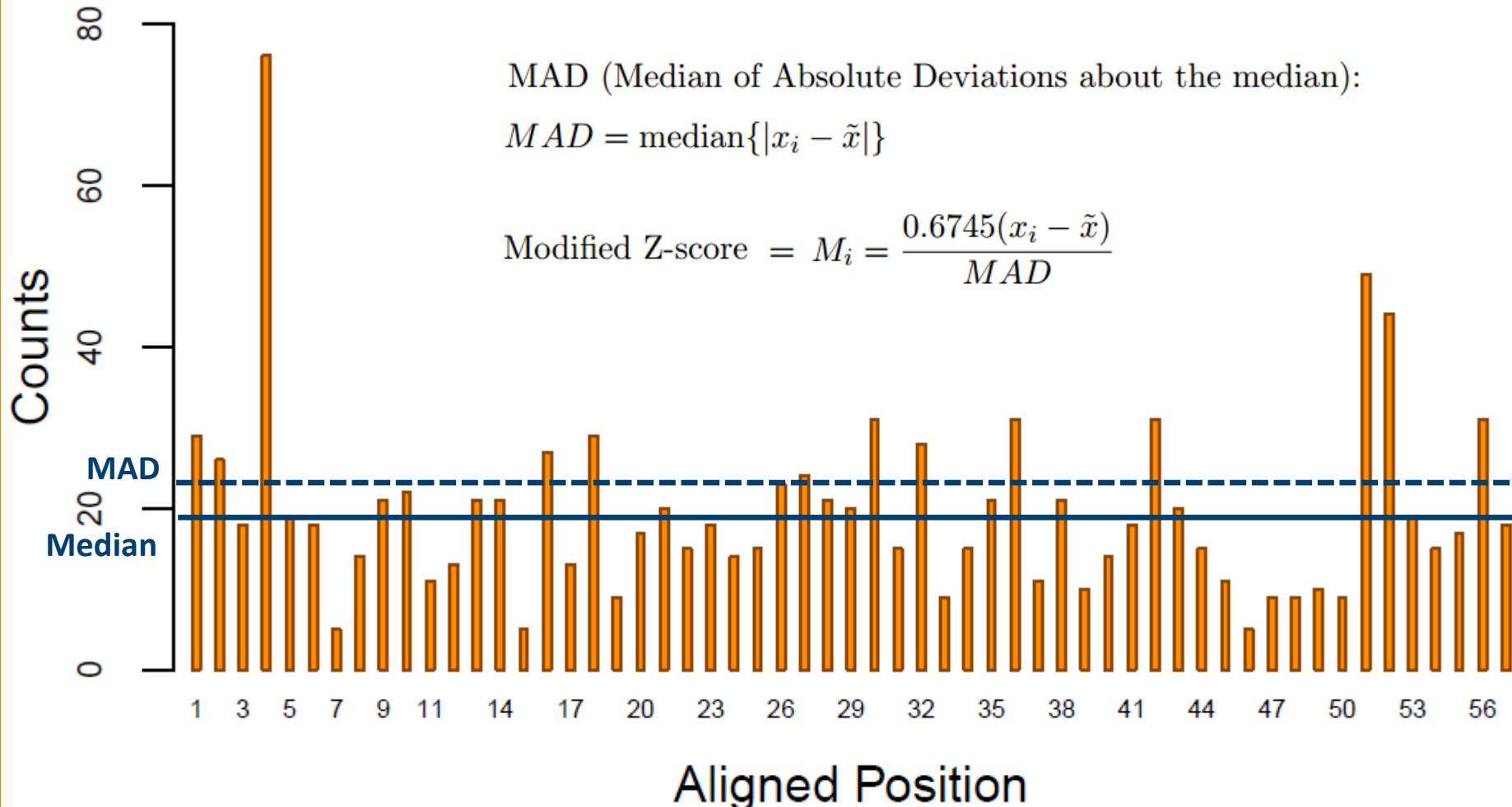


- Align sequences from all matching domains across all tumor types
- Evaluate missense mutation frequencies
- Applies to ~40% of missense mutations

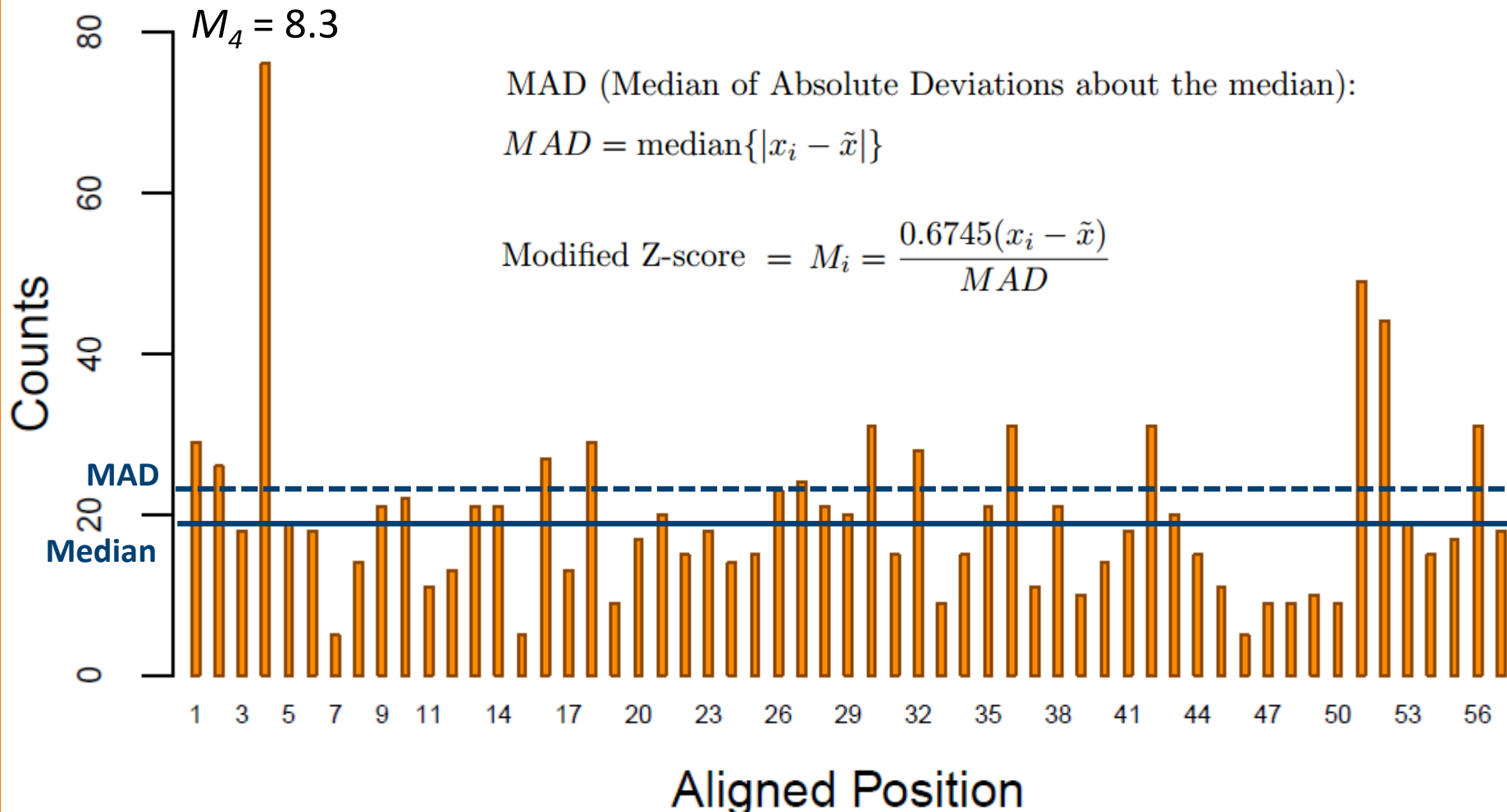
# Mutations per position for two hundred aligned Homeobox domains across twenty tumor types



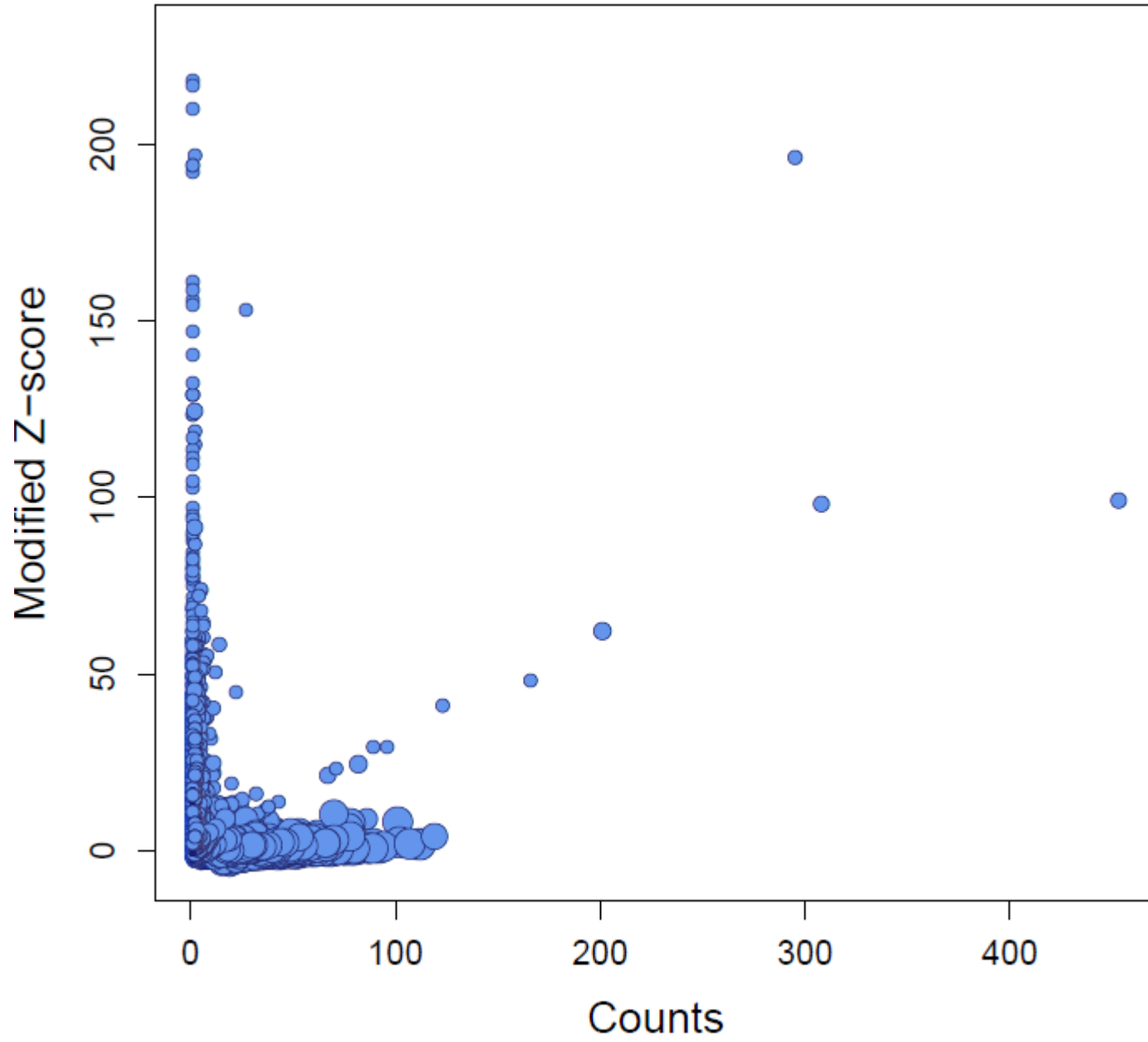
# Identifying outliers: Modified Z-score



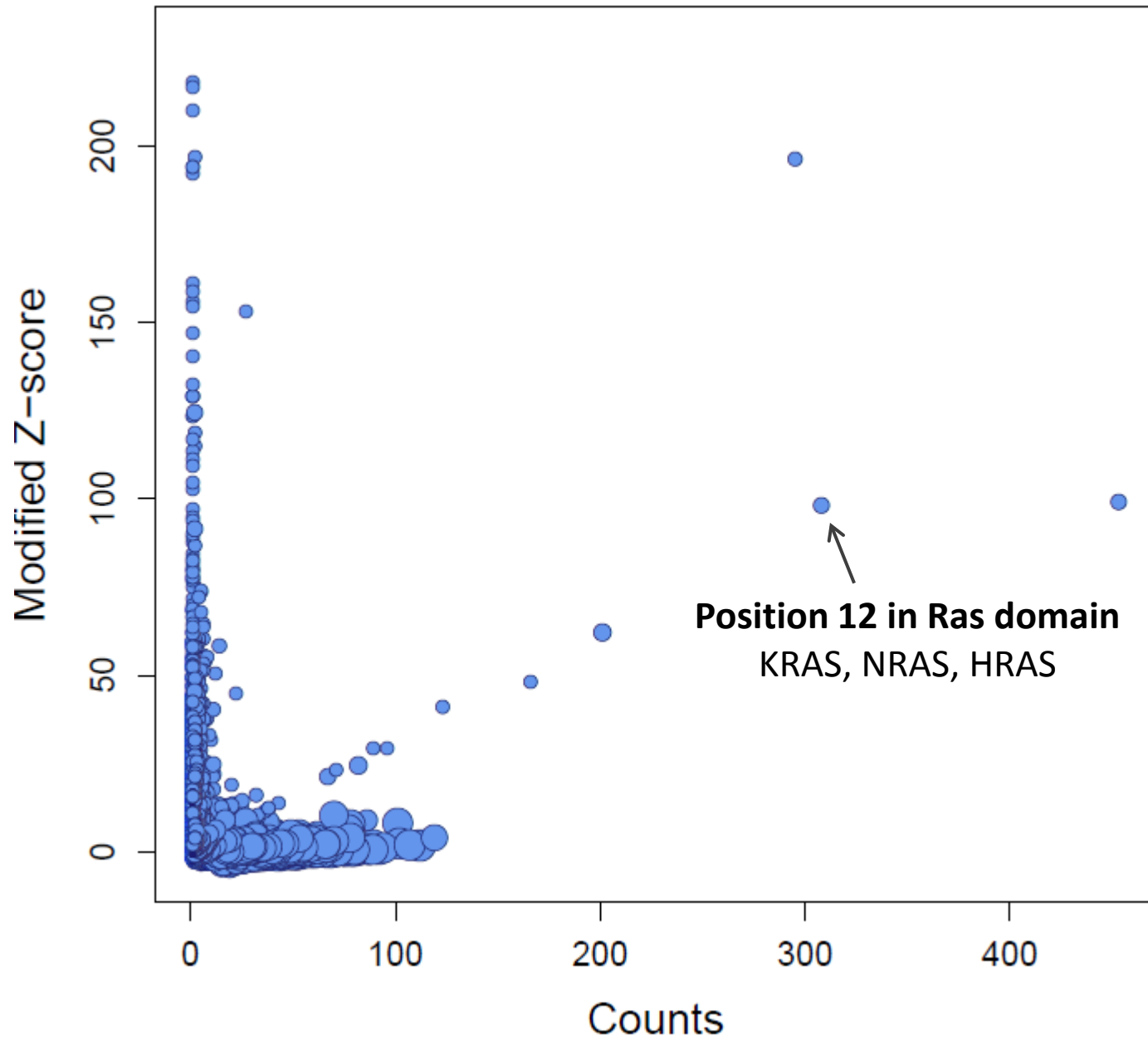
# Identifying outliers: Modified Z-score



# All positions within all domains

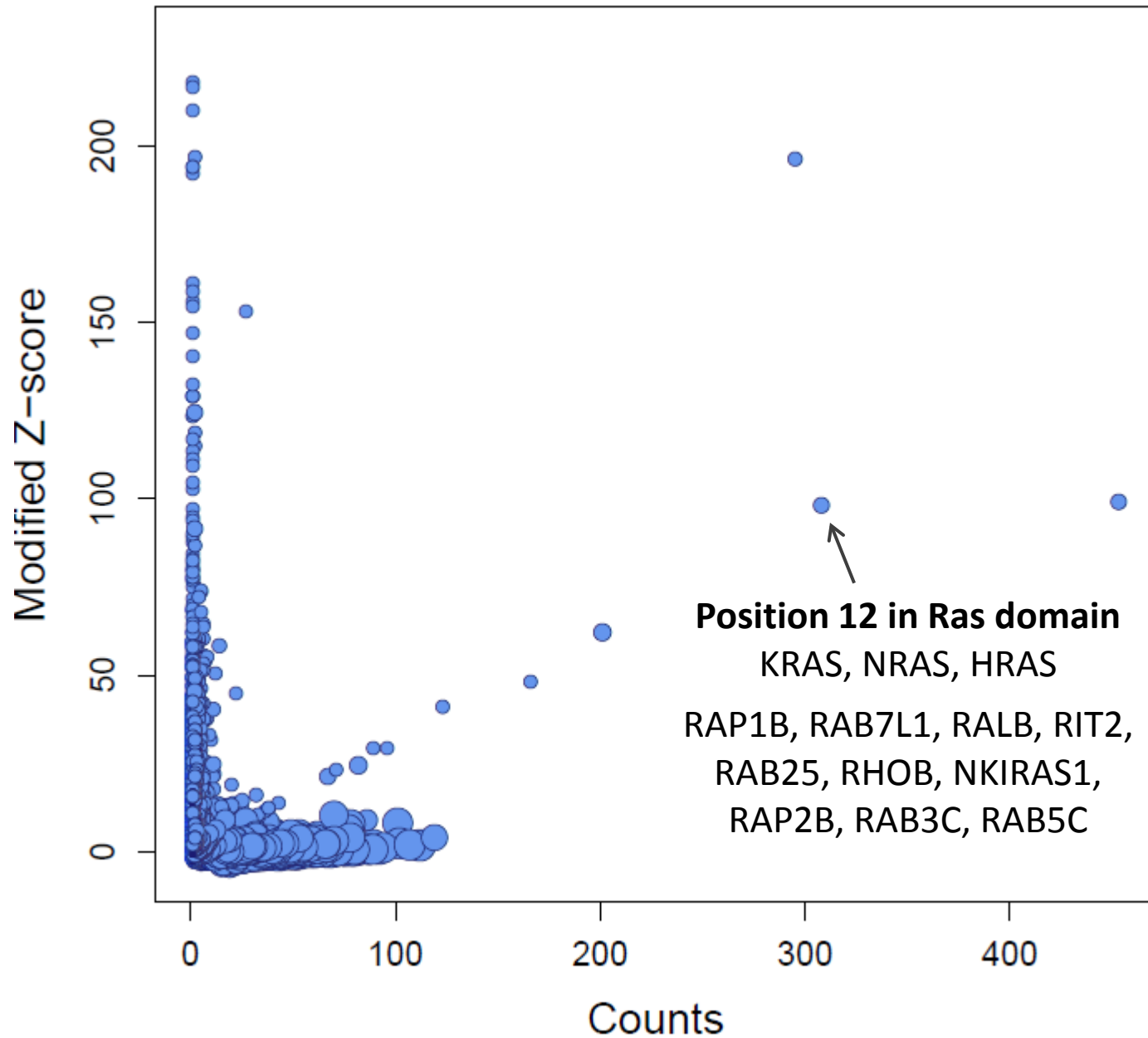


# High Modified Z-score and counts for Ras domain

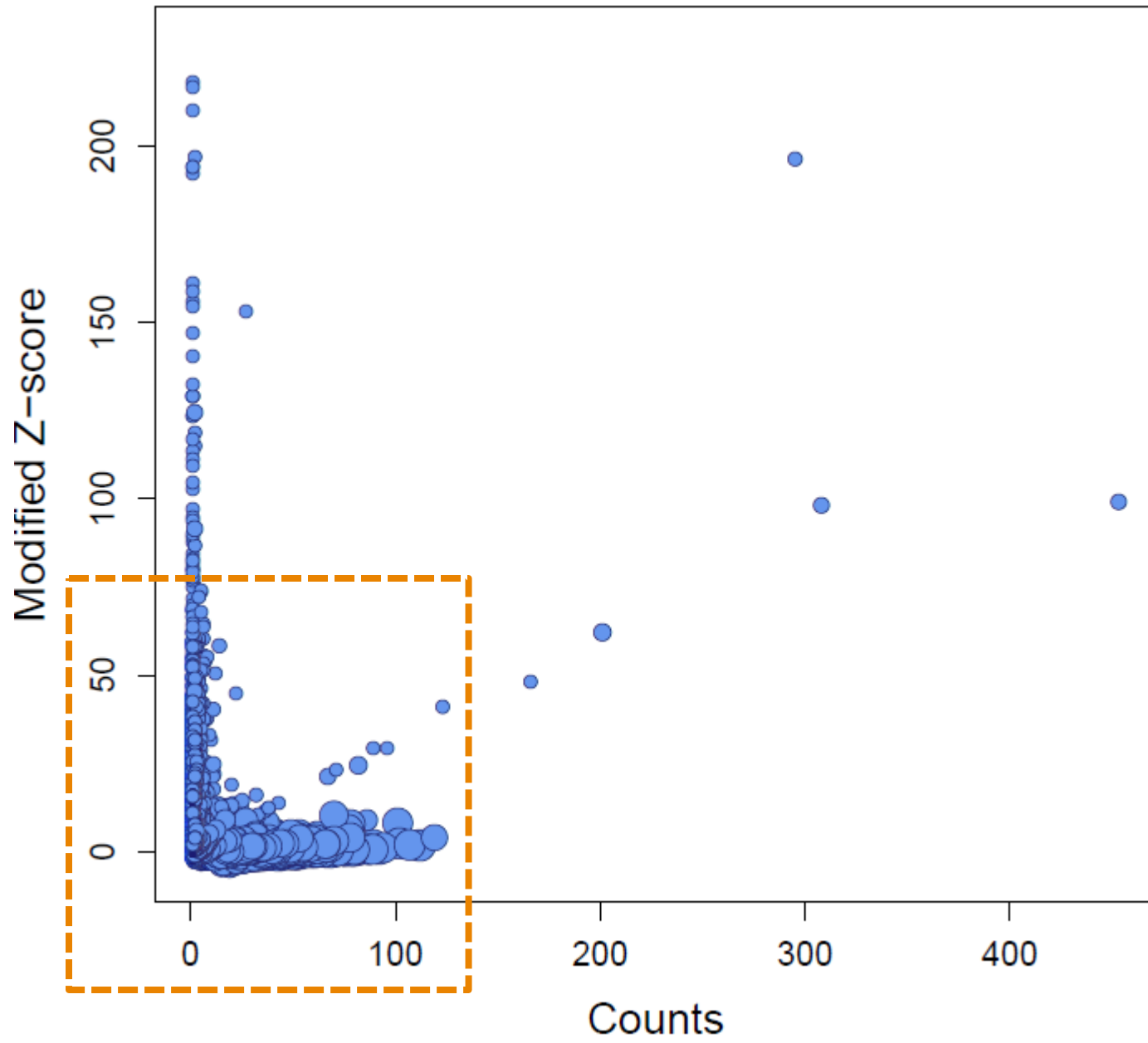




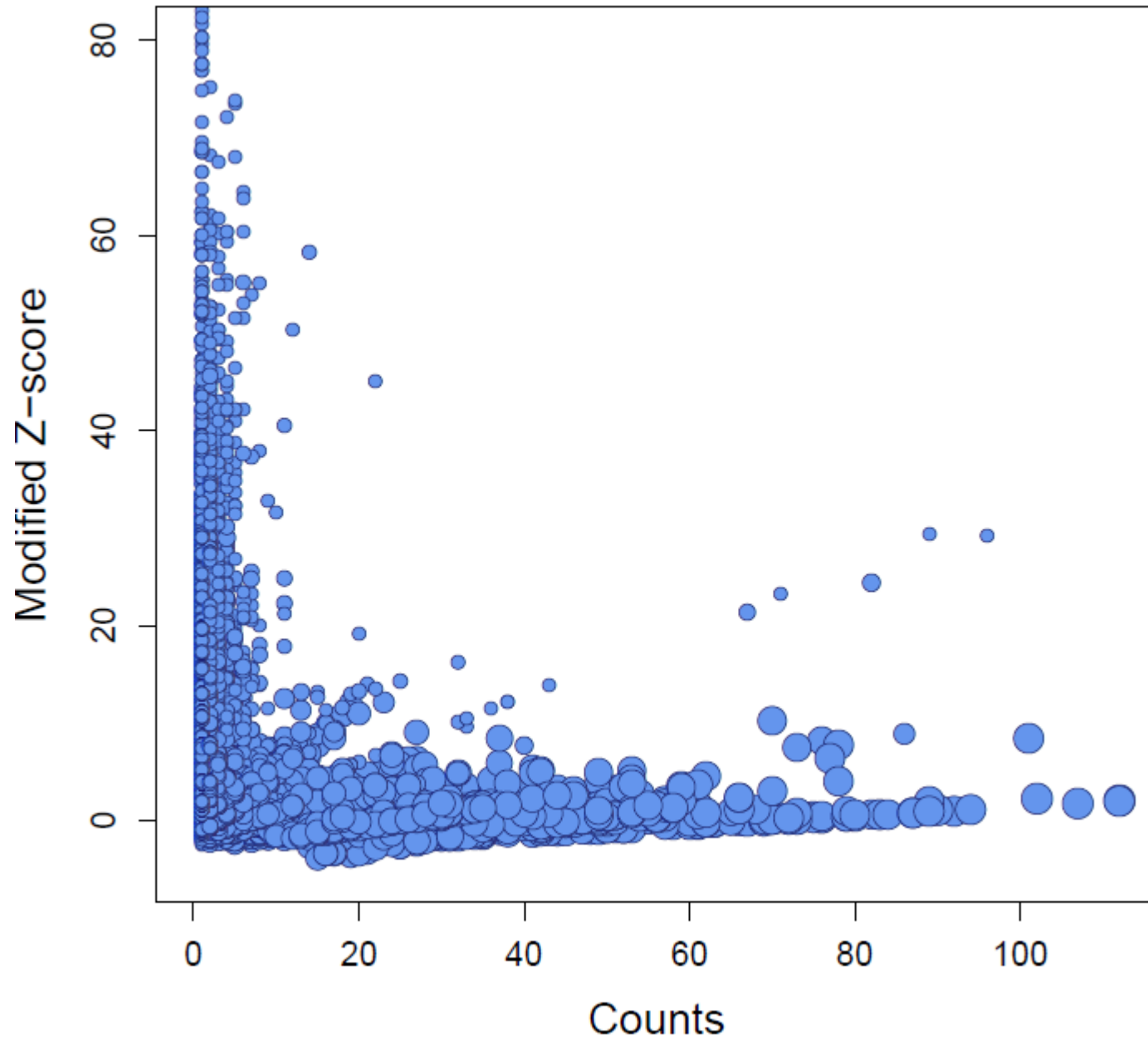
# Multiple other Ras genes mutated at the same position



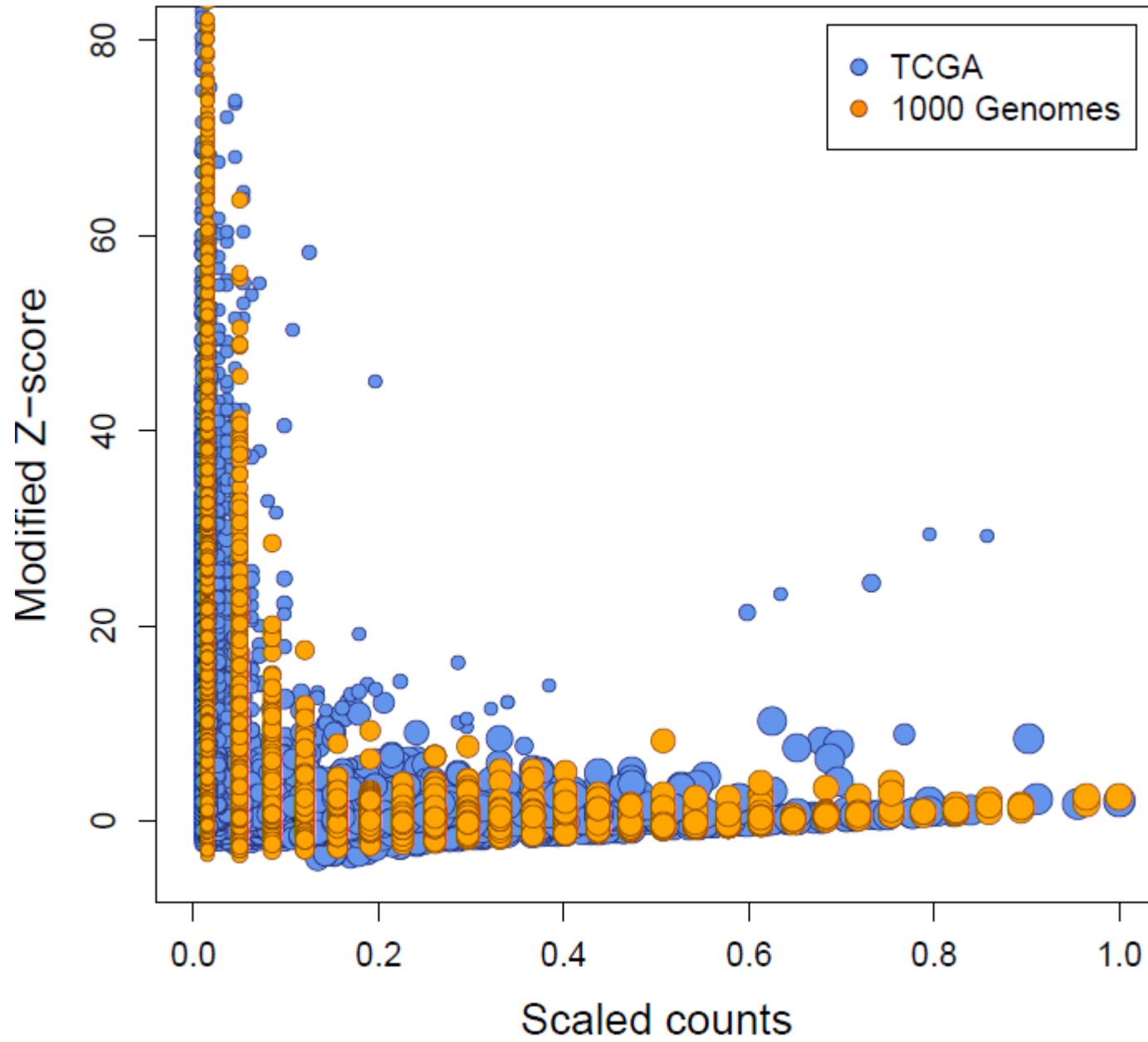
# Investigating genes with low mutation frequency



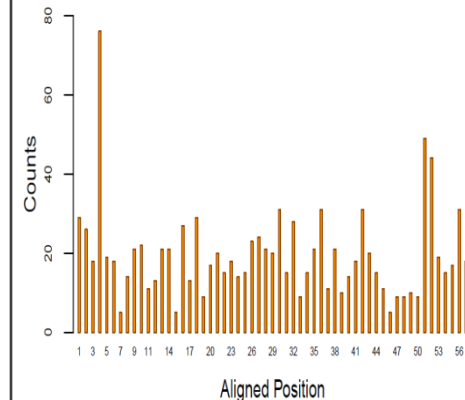
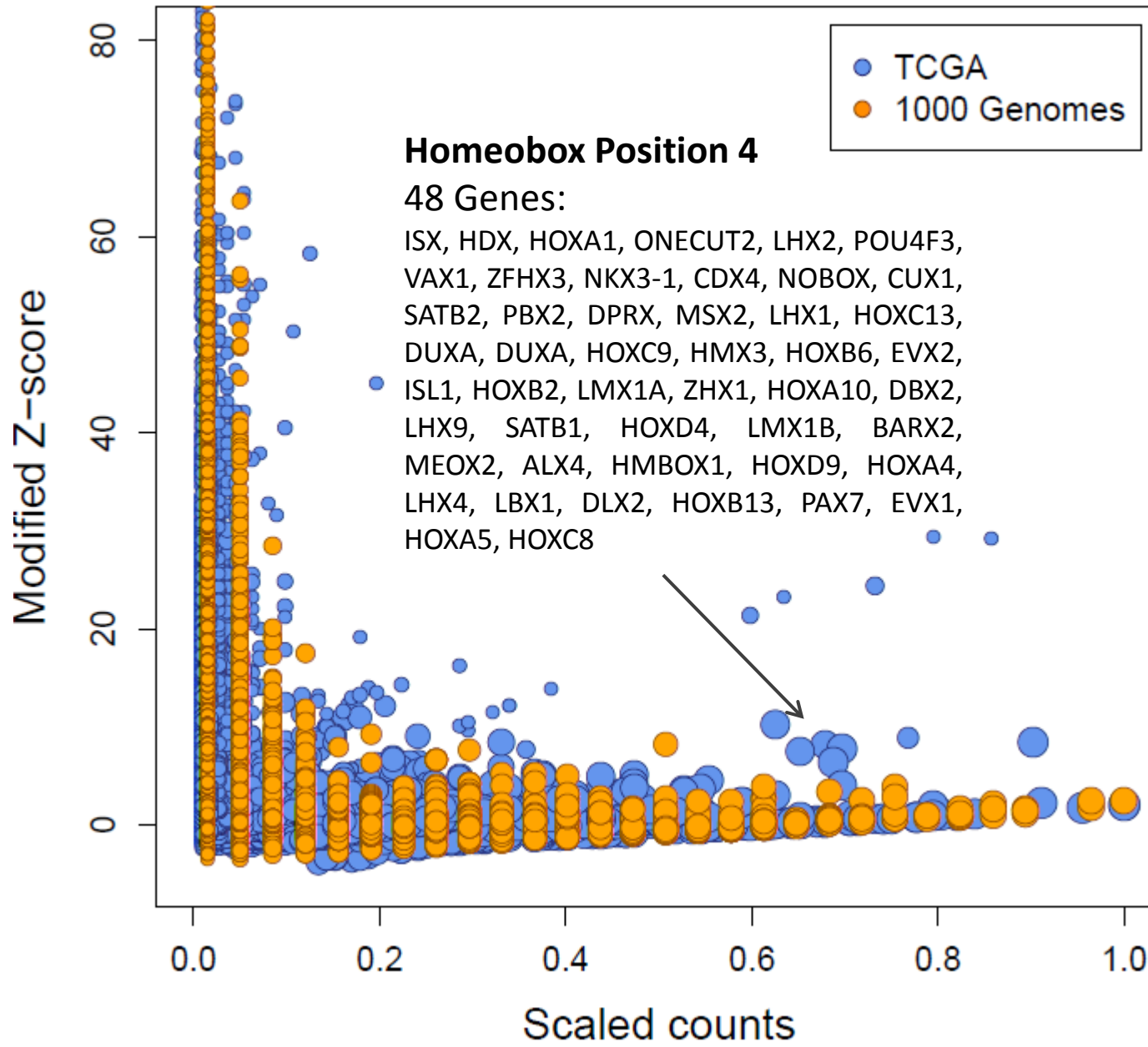
# Investigating genes with low mutation frequency



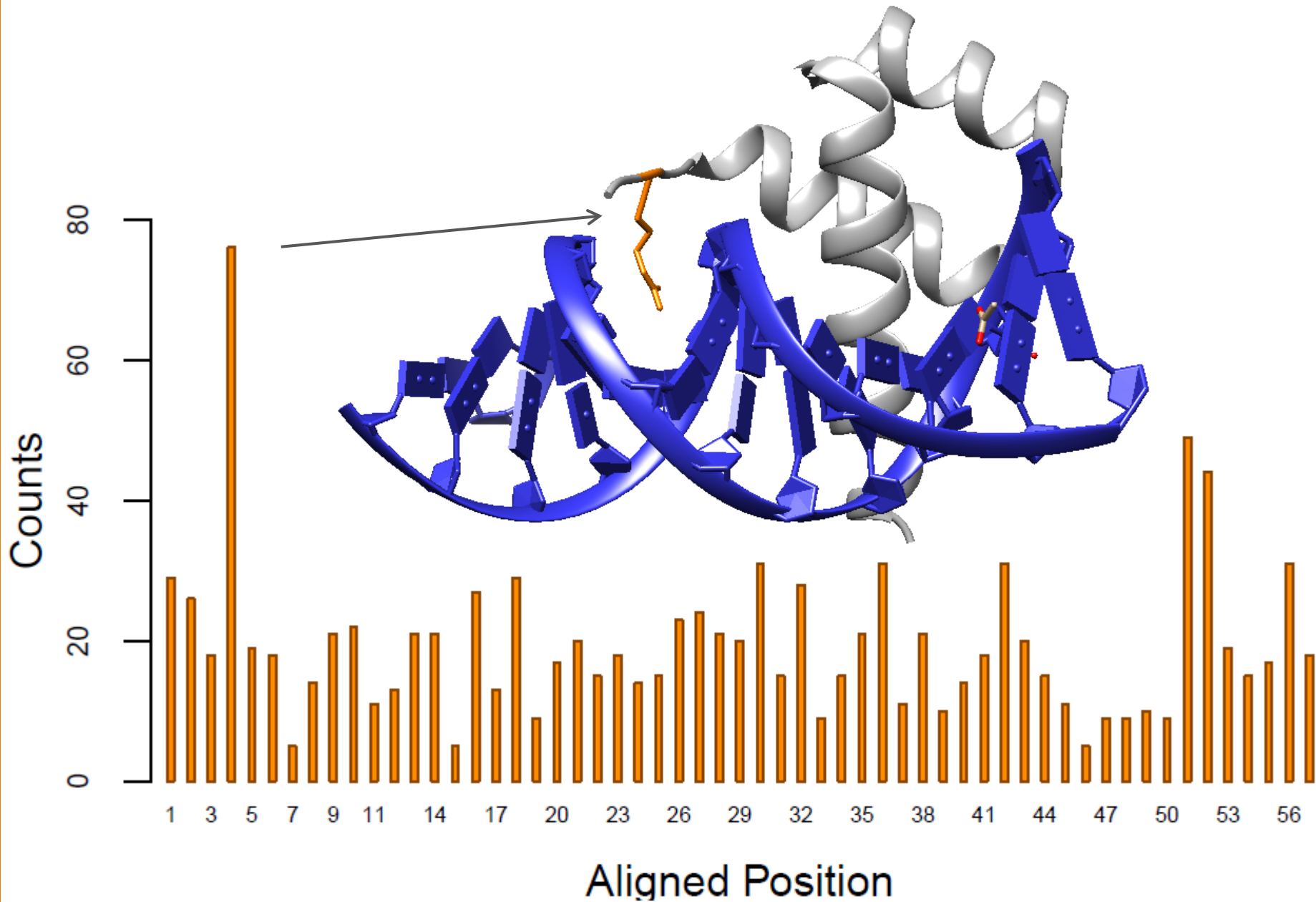
# Background expectation: Comparing TCGA to 1000 Genomes



# Outlier position within Homeobox domain



# Structural interpretation



# Summary and Next Steps

- We have comprehensively evaluated mutations at all positions within all domains to identify low frequency but likely deleterious mutations
- Hotspots outside of domains and mutations that broadly disrupt structure and function not addressed
  - Integration with other methods is essential
- Functional validation of low frequency events in such data sets remains challenging
- As cancer and normal genome data size increases, more robust normalization per position per domain can be achieved

# Acknowledgements

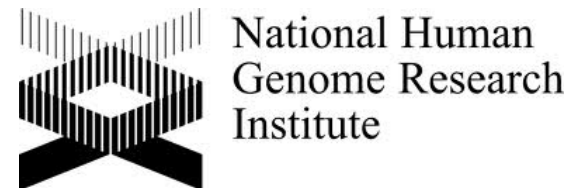


**Ilya Shmulevich**

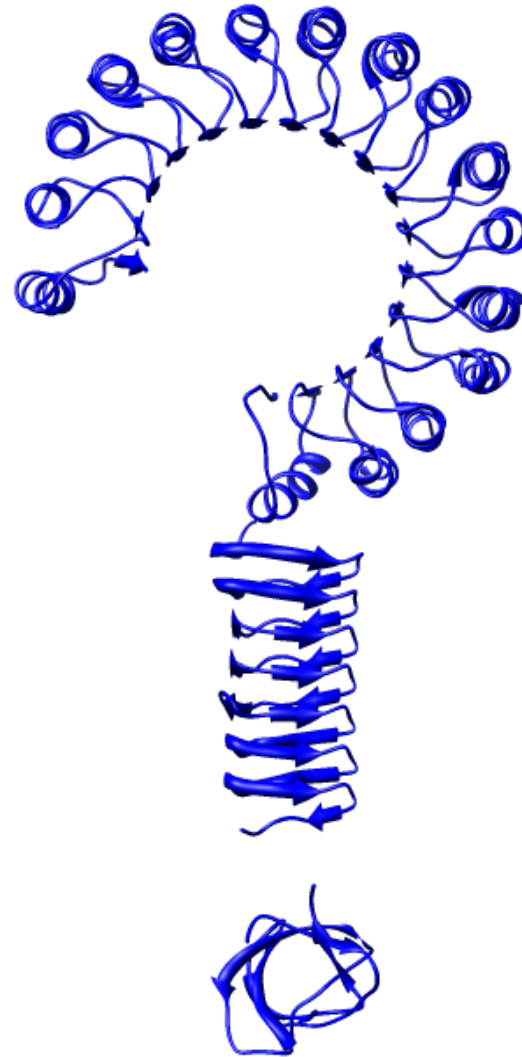
**Tae Denwongkun**

**Theo Knijnenburg**

Vesteinn Thorsson	Sheila Reynolds
Kalle Leinonen	Lisa Iype
Richard Kreisberg	Natalie Tasman
Hector Rovira	







2bnh: alpha-beta horseshoe

1hv9: left-handed beta helix

1m30: SH3-like barrel



# Example approaches and methods

- Statistical assessment of multiple samples (MutSig, MuSiC, ...)

## Mutational heterogeneity in cancer and the search for new cancer-associated genes

Michael S. Lawrence<sup>1\*</sup>, Petar Stojanov<sup>1,2\*</sup>, Paz Polakowitz<sup>1</sup>, Scott L. Carter<sup>1</sup>, Chip Stewart<sup>1</sup>, Craig H. Mermel<sup>1,5</sup>, Yotam Drier<sup>1,3,5,8</sup>, Lihua Zou<sup>1</sup>, Alex H. Ramos<sup>1</sup>, Treva M. Rodriguez<sup>1</sup>, Carrie Sougnez<sup>1</sup>, Lauren Ambrogio<sup>1</sup>, Elizabeth Nichols<sup>1</sup>, Douglas Voet<sup>1</sup>, Michael Noble<sup>1</sup>, Daniel DiCara<sup>1</sup>, Pei Wang<sup>1,5</sup>

## MuSiC: Identifying mutational significance in cancer genomes

Nathan D. Dees,<sup>1,4</sup> Qunyuan Zhang,<sup>1,4</sup> Cyriac Kandoth,<sup>1</sup> Michael C. Wendl,<sup>1,2</sup> William Schierding,<sup>1</sup> Daniel C. Koboldt,<sup>1</sup> Thomas B. Mooney,<sup>1</sup> Matthew B. Callaway,<sup>1</sup> David Dooling,<sup>1</sup> Elaine R. Mardis,<sup>1,2,3</sup> Richard K. Wilson,<sup>1,2,3</sup> and Li Ding<sup>1,2,5</sup>

- Sequence conservation (MutationAssessor, SIFT, ...)

## Predicting the functional impact of protein mutations: application to cancer genomics

Boris Reva\*, Yevgeniy Antipov

## Predicting Deleterious Amino Acid Substitutions

Pauline C. Ng<sup>1,2</sup> and Steven Henikoff<sup>1,3,4</sup>

- Machine learning and classification (PolyPhen-2, CHASM, ...)

## A method and server for predicting damaging missense mutations

Ivan A. Adzhubei<sup>1,7</sup>, Gerasimova<sup>5</sup>, Peer Bork<sup>6</sup>

## Cancer-Specific High-Throughput Annotation of Somatic Mutations: Computational Prediction of Driver Missense Mutations

Hannah Carter,<sup>1</sup> Sining Chen,<sup>2,3</sup> Leyla Isik,<sup>1</sup> Svitlana Tyekucheva,<sup>3</sup> Victor E. Velculescu,<sup>4</sup> Kenneth W. Kinzler,<sup>4</sup> Bert Vogelstein,<sup>4</sup> and Rachel Karchin<sup>1</sup>